Identification of a tissue-specific cofactor of polycomb repressive complex 2

Roberta Ragazzini

Université Pierre et Marie Curie

Ecole doctorale Complexité du Vivant, ED515

Institut Curie, INSERM U934-CNRS UMR3215

Identification of a tissue-specific cofactor of

Polycomb Repressive Complex 2

Par Roberta Ragazzini

Thèse de doctorat de Epigénétique

Dirigée par Dr. Raphaël Margueron

Présentée et soutenue publiquement le 25 septembre 2017

Devant un jury composé de :

Dr. Valérie Borde	Président
Dr. Daniel Vaiman	Examinateur
Dr. Sandra Duharcourt	Rapporteur
Dr. Adrian Bracken	Rapporteur
Dr. Raphaël Margueron	Directeur de thèse

"You look at science (or at least talk of it) as some sort of demoralising invention of man, something apart from real life, and which must be cautiously guarded and kept separate from everyday existence. But science and everyday life cannot and should not be separated. Science, for me, gives a partial explanation for life.
In so far as it goes, it is based on fact, experience and experiment."
Rosalind Franklin

TABLE OF CONTENTS

AKNOWLEDGEMENTS 3
ABSTRACT 5
RESUME 6
RESUME DES TRAVAUX 7
ABBREVIATIONS AND ACRONYMS 10
INTRODUCTION 17
CHAPTER I 18
1. Genetics mutations led to the discovery of Polycomb and Trithorax group of genes 20

1.1 Polycomb Phenotypes 20
1.2 Discovery of the Trithorax phenotype 21
2. PcG proteins form multi-subunits complexes 22
3. Towards PcG function 24
4. Trithorax multimeric complexes 26
5. PcG conservation 27
CHAPTER II 30
1. PRC1 Complexes 31

1.1 Canonical PRC1 complexes: PRC1.2 & PRC1.4 32
1.2 Non canonical PRC1: PRC1.3, PRC1.5 & PRC1.6 33
2. PRC1 Functions 34
3. PRC1 Targeting 35
4. Other PcG complexes 36
CHAPTER III 38
1. PRC2 Structure 39
2. PRC2 Regulation 40
3. PRC2 and PRC1 are intertwined 44
4. PRC2 and PRC1 Targeting 46
4.1 Transcriptin Factors 47
4.2 lncRNAs 47
4.3 Genome Wide Targeting 47
CHAPTER IV 49
1. PRC2.1 50

1.1 Polycomb Like Proteins (PCL) 50
1.1.1 PCL1 alias PHF1 51
1.1.2 PCL2 alias MTF2 51
1.1.3 PCL3 alias PHF19 51
1.2 C17orf96 alias EPOP 53
1.3 C10orf12 54
2. PRC2.2 55
2.1 JARID2 55
2.2 AEBP2 56
3. PRC2 catalytic subunit has two paralogs: *EZH1* and *EZH2* 58
CHAPTER V 60
1. Primordial Germ Cells (PGCs) onset 62

2. PGCs epigenome 63

2.1 DNA methylation 63
2.2 Repressive Histone Marks 63
3. Germ cells Fate in Females 64
3.1 Chromatin landscape in folliculogenesis 65
4. Germ cells Fate in Males 66
4.1 Chromatin landscape in spermatogenesis 67
5. PRC2 in gametogenesis 68

RESULTS 71
PHD OUTLINE AND QUESTIONS ADRESSED 72

RESULTS-CHAPTER I 73
ABSTRACT 74
INTRODUCTION 75
RESULTS 78
DISCUSSION 85
FIGURE LEGENDS 87
SUPLLEMENTARY FIGURE LEGENDS 96
MATERIALS AND METHODS 103
SUPLLEMENTAL INFO 109
AUTHORS CONTRIBUTION 110
AKNOWLEDGEMENTS 111

RESULTS-CHAPTER II 112
INTRODUCTION AND AIM OF THE PROJECT 112
1. LncRNAs functions 112
2. LncRNAs and chromatin 114
3. PRC2 and RNA: is it a preferential or rather a promiscuous binding? 118
3. lncRNA HOTAIR as a paradigm to study lncRNAs and PRC2 interplay. 119

PRC2 is dispensable for HOTAIR-mediated transcriptional repression 121

DISCUSSION AND PERSPECTIVES 147

BIBLIOGRAPHY 154

AKNOWLEDGEMENTS

First of all, I would like to thank very much my thesis director, Dr. Raphael Margueron for having given me the possibility to work in his lab on very interesting and dynamic projects during the past few years. And not only for that, since he has always been very patient with me and always available for scientific discussions and help. Particularly, I will always be grateful to him for having proposed me to take on the tissue-specific PRC2 interactome project. This brought us to an important discovery that has really motivated me my everyday life in the lab. I could not forget to thank Dr. Vermeulen, Dr. Baymaz, Dr. Pierre Fouchet et Maelle Givelet because of their great help in this project.

I am also grateful to Prof. Edith Heard and Dr. Déborah Bourc'his for scientific discussion and help in the projects.

I would like to thank also Dr. Adrien Bracken, Dr. Sandra Duharcourt, Dr. Valérie Borde and Dr. Daniel Vaimann fro kindly accepting to be part of my thesis jury.

Manuela, I cannot forget about you. You have been the first person that I worked with in the lab and that has been a great teacher and now a real friend. I would like to thank you for having pushed me to the PhD experience. Thanks to your precious advice I have always tried to pull out the best of it.

Audrey, I could just tell you that if I have grown up in these years, this is also your merit. Your honesty is priceless: it is rare to find a person that talks as directly as you and I really like it. I would just like to say thank you for everything, from a PCR you performed for me to your important presence in the difficult/joyful moments of life.

Daniel, even if you joined the lab late in my PhD, I am really glad to have met you. I really like the scientific and general discussion we had from CSH to the cell culture room. Thank you for being always nice and willing to help in every moment.

I cannot forget of Corinne and Aude, of course. You have been just wonderful. Thank you for having shared with me the difficult project of the CRISPR/Cas9 screen and also good moments outside of the lab. And Armelle…One year has already gone and now you are going so far away! I am very happy our PhDs overlapped in the lab and that we exchanged a lot of opinions and ideas.

I am also very grateful to Michel for all the smart suggestions he offered me in these years and for the interesting discussions. Eric! We only spent one year together but it has been a pleasure to meet you. Thank you for the discussions and for the great help in dealing with ESS.

Mattia, you have been my first (and last) student that I supervised in the lab. Your contribution to the screen project was really important as well all the discussions and chats that we do still have although we are separated in time by 6 hours every day.

Rafael, you have been one of the best PhD mate ever. Thank you for motivating me every day and for having listened to me in the difficult moments together with Denis and Ines.

Raquel, you are so nice and I am grateful to you since your positive attitude and your quietness transformed the fear of mice into the motivation to go and study them everyday. Arianna, you are a

great colleague and a great friend. Thank you for sharing with me your point of view and for being always there.

Flavia, you have always been there from the very beginning and we shared everything together. I hope I am going to spend new adventures with you also in your future horizons. Angelo, we start this painful trip together and we are almost at the end. Thank you for having been always there. You are a great person and I deserve you the best.

Vince, we are all missing you. Thank you for all the time you listened to me everyday even in the complaining mood and for all the chats we had. I really appreciate you.

Giulia. We met early in the Univerisity and after being separated for a while, we now share many things from the lentivirus to the dance classes. Your positive attitude enlightened my day.

I cannot forget all my family that has always supported me and motivated me to go farther during this years and Matteo who is a great support and that is always able to understand the difficult moments.

ABSTRACT

The Polycomb Repressive Complex 2 (PRC2) plays an essential role in development by maintaining gene repression through the deposition of H3K27me3. A variety of cofactors have been shown to control its function in cells of various origins however little is known about PRC2 regulation during gametogenesis. Gametogenesis entails genome wide changes in chromatin structure and is associated to important switch in gene expression profiles. The H3K27me3 mark is present throughout germ cell development and maintenance of transcriptional silencing mediated by PRC2 is required for proper gametogenesis.

During my PhD, I took advantage of murine models where Ezh2 and Ezh1 were knocked-in, I isolated nuclear extracts from whole adult testis and, identified a new polypeptide interacting with PRC2. This protein is specifically expressed in gonads, is of unknown function and does not contain any conserved domain. I have confirmed its interaction with PRC2, mapped the domain of interaction with PRC2 and shown that it could tether PRC2 to chromatin. Thanks to a knockout mouse model, I demonstrated that the protein is required for female fertility. In contrast, its ablation has little consequences on male fertility although it brings to a global increase of H3K27me3, the park deposited by PRC2.

I also contributed to the characterization of the interplay between the long non-coding RNA (lncRNA) HOTAIR and PRC2 complex. Many lncRNAs have been proposed to modulate chromatin-modifying complexes action on chromatin. With the help of novel RNA-tethering system, HOTAIR inducible expression causes transgene repression independently from PRC2. Forced overexpression of HOTAIR also has little impact on transcriptome in breast cancer cells. Generally, PRC2 binding to RNA is not required for chromatin targeting.

Taken together these results shed light on the mechanism of a newly identified cofactor regulating PRC2 in the gonads and contribute to dissect PRC2-RNA relationship at the molecular level.

KEYWORDS: Epigenetic, gene silencing, Polycomb, germ cells, gametogenesis

RESUME

Le Polycomb Repressive Complex 2 (PRC2) joue un rôle essentiel dans le développement en maintenant la répression des genes en catalysant le dépôt de la marque H3K27me3. Il a été montré que divers cofacteurs contrôlent sa fonction dans des cellules d'origines variées, mais on connait peu la régulation de PRC2 pendant la gamétogenèse. La gamétogenèse implique des changements globaux de la structure de la chromatine et est associée à d'importantes modifications des profils d'expression des gènes. La marque H3K27me3 est présente tout le long du développement des cellules germinales et la presence de PRC2 est nécessaire pour une gamétogenèse fonctionnelle.
Au cours de mon doctorat, j'ai utilisé des modèles murins ou un tag a été introduit dans les gènes Ezh2 et Ezh1, j'ai isolé des extraits nucléaires de testicules adultes entiers et identifié un nouveau polypeptide interagissant avec PRC2. Cette protéine est spécifiquement exprimée dans les gonades, elle ne contient aucun domaine conservé et sa fonction est inconnue. J'ai confirmé son interaction avec PRC2 puis caractérisé le domaine d'interaction avec PRC2. J'ai également montré que cette protéine peut recruter PRC2 à la chromatine. Grâce à un modèle de souris knock-out, j'ai démontré que cette protéine est nécessaire pour la fertilité femelle. Différement, son ablation dans les cellules germinales mâles a peu de conséquences sur la fertilité mais se traduit par une augmentation globale de H3K27me3, la marque déposée par PRC2.
J'ai également contribué à la caractérisation de l'interaction entre le long ARN non-codant (lncRNA) HOTAIR et PRC2. Il a été proposé que de nombreux ARNnc pourraient moduler l'action des complexes modifiant la chromatine. Dans un nouveau système de recrutement artificiel d'ARN, nous avons observé que l'expression d'HOTAIR provoque une répression d'un transgène indépendamment de PRC2. La surexpression forcée de HOTAIR a par ailleurs peu d'impact sur le transcriptome dans des cellules cancéreuses du sein. Généralement, la liaison PRC2 à l'ARN n'est pas requise pour le ciblage de la chromatine.
Ensemble, ces résultats nous éclairent sur le mécanisme d'action d'un nouveau cofacteur de PRC2 dans les gonades et contribuent à disséquer la relation PRC2-ARN au niveau moléculaire.

MOTS-CLE : Epigénètique, silencement des genes, Polycomb, cellules germinales, gametogénèse

RESUME DES TRAVAUX

Les enzymes qui modifient la structure de la chromatine jouent un rôle important pour la régulation transcriptionnelle au cours du développement et à l'âge adulte. Les enzymes de la machinerie Polycomb (PcG) sont cruciales pour le maintien de l'état silencieux de nombreux gènes pendant différents processus biologiques tels que la différenciation, la prolifération ou encore le maintien des cellules souches. Un fonctionnement anormal de la machinerie Polycomb est un phénomène fréquent dans les cancers. Malgré les progrès réalisés dans la compréhension de la façon dont PcG exerce sa fonction, les caractéristiques clé de la régulation de Polycomb restent à comprendre. Les protéines Polycomb ont été découvertes il y a plus de 60 ans chez la Drosophile d'après l'observation d'un mutant avec des défauts dans l'expression des gènes Hox qui contrôlent le développement de l'organisme. Tout d'abord, plusieurs mutants ont été idéntifié qui présentaient un phénotype similaire. Dans un deuxième temps, leur capacité à s'associer dans différents complexes multiprotéiques, a été découverte. Seulement plus tard, les activités enzymatiques de modification de la chromatine ont été caractérisées. Les deux complexes majeurs de la machinerie Polycomb sont les Polycomb Repressive Complex 1 et 2 (PRC1 et PRC2). Le complexe PRC1 se compose de la sous-unité cataytique E3 ubiquitine ligase Ring1 et d'une protéine PCGF Polycomb group of ring finger protein. Chez les mammifères, chaque sous-unité de PRC1 possède plusieurs homologues et peut s'associer de façon combinatoire pour donner naissance à de nombreux sous-type de complexes avec différentes propriétés. Le complexe PRC2 est composé de quatre polypeptides : EZH1 / 2, EED, SUZ12 et RbAp46 / 48. Ces sous-unités enzymatiques, EZH1 ou EZH2, sont responsables de la di et tri-méthylation de la lysine 27 de l'histone H3 (H3K27me2/3). La sous-unité catalytique du complexe n'est active qu'en association avec les autres sous-unités du complexe : leur association donne une conformation précise qui permet d'effectuer efficacement la reaction de méthylation. PRC2 est très sensible à l'état de la chromatine : il peut reconnaître sa propre modification, ce qui augmente son activité catalytique ou peut être inhibé par d'autres modifications associées à l'activation des gènes. Jusqu'à maintenant, il n'est pas encore bien défini comment le complexe est dirigé vers la chromatine. Des mécanismes différents ont été proposés, mais ils ne représentent pas une règle générale. Cependant, une question reste ouverte dans le domaine à savoir comment un complexe dont la composition ne varie pas, peut maintenir silencieux un ensemble spécifique de gènes dans différents types de cellules ? En plus de son complexe de base, PRC2 peut s'associer à des sous-unités facultatives de manière sous-stoechiométrique. Ces partenaires peuvent moduler l'activité catalytique de PRC2, favoriser son recrutement à ses cibles ou les deux. La plupart d'entre eux ont été largement étudiés individuellement, alors que récemment des analyses de spectrométrie de masse ont montré que leur association avec le complexe central peut se produire en des combinaisons particulières et

exclusives. Cela conduit à l'hypothèse que deux sous-types de complexes PRC2 existent avec une fonction potentiellement différente.

Au cours de mes travaux de thèse, nous avons voulu déterminer si des partenaires de PRC2, spécifiques à un tissu donné, existent. Nous nous sommes intéressés aux gonades (ovaires et testicules) où se produisent les gamètes (ovules et spermatozoïdes). Ces cellules sont garantes de l'hérédité et doivent protéger le matériel génétique pour la transmission à la future génération. Étant donné que la chromatine subit des changements dramatiques pendant la gamétogenèse et compte tenu du fait que la méthylation de H3K27 est présente dans le développement de cellules germinales, nous avons estimé que si PRC2 possède un (des) cofacteur (s) spécifique (s) à une cellule, il est susceptible de se produire dans les gamètes. Un tel cofacteur pourrait être important pour tenir compte du ciblage et de la régulation de la fonction de PRC2 dans les différents paysages de la chromatine. Il est à noter qu'Ezh2 est connu pour être nécessaire pour la spermatogenèse où il est proposé qu'il soit essentiel pour le silence du transcriptome somatique.

Par des méthodes de spectrométrie de masse très sensibles, nous avons purifié l'ensemble des protéines interagissant avec le complexe PRC2. Cette expérience nous a permis de découvrir un nouveau partenaire de PRC2, la protéine AU022751. Nous avons ensuite confirmé cette nouvelle interaction dans une lignée cellulaire modèle qui surexprime de façon stable AU022751. Cette protéine est retrouvée chez les mammifères placentaires (CXorf67), elle a évolué rapidement et ne comporte aucun domaine conservé en dehors d'un motif de 13 acides aminés dans la partie C-terminale. Si ce domaine est absent, l'interaction avec le complexe PRC2 est perdue. Puisque la fonction protéique de AU022751/CXorf67 est totalement inconnue, nous avons essayé d'identifier le mécanisme d'action de cette protéine *in vitro*. Nous avons observé que la protéine peut stimuler l'activité catalytique de PRC2 sur les nucléosomes recombinants alors que cet effet n'est pas observé lorsque des nucléosomes purifiés à partir de cellules de mammifères sont utilisés. Cela suggère une compétition entre l'activité stimulatrice d'AU022751 et les modifications post-traductionnelles de la chromatine. Nous avons également utilisé un système de recrutement artificiel afin de déterminer l'effet de ce nouveau cofacteur de PRC2 sur la transcription d'un gène rapporteur. Nous avons observé que la forme sauvage de la protéine est capable de réprimer l'expression du gène rapporteur et déposer la marque H3K27me3, alors que la forme mutante dépourvue du motif conservé ne le peut pas. En parallèle, nous avons réalisé des études *in vivo*. Comme cette protéine est exprimée principalement dans les gamètes, nous nous sommes intéressés particulièrement à son rôle dans la gamétogenèse. A cette fin, nous avons inactivé cette protéine dans un modèle murin transgénique avec la technologie CRISPR/Cas9. Nous avons ainsi pu montrer les conséquences de l'inactivation de notre protéine d'intérêt sur la fertilité et le développement des gamètes. L'inactivation de AU022751 se traduit par une forte réduction de la fertilité des femelles. Notre étude indique par ailleurs que la fréquence des portées diminue avec le temps et que la taille des portées est réduite. De plus, les ovaires des femelles knockout pour AU022751 sont plus petits et produisent un nombre inférieur de follicules primordiaux.

Les males AU022751 knockout sont fertiles même s'il y a une augmentation globale de la marque H3K27me3 déposée par PRC2 dans cellules germinales des testicules avec des conséquences mineures sur la fertilité des mâles. Nous évaluons actuellement quelles cibles sont affectées par l'augmentation de H3K27me3 et si des phénomènes de compensation par d'autres cofacteurs de PRC2 pourraient expliquer l'augmentation d'H3K27me3.

J'ai également activement contribué à une étude menée dans le laboratoire dont l'objectif était de caractériser l'interaction moléculaire entre PRC2 et un long ARN non-codant dans la répression des gènes. Au cours des dernières années, il a été montré que de nombreux ARN non-codants interagissent avec le complexe PRC2 et il a été proposé que cette interaction servirait à cibler le complexe PRC2 à la chromatine. Un des exemples qui a reçu le plus d'attention est le cas de l'ARN non codant HOTAIR. Grâce à des tests biochimiques et en établissant un système de recrutement artificiel de l'ARN non codant à un gène rapporteur, nous avons démontré que HOTAIR lie PRC2 avec une affinité élevée, même si le complexe lui-même n'est pas indispensable pour la répression transcriptionnelle induite par HOTAIR. Au lieu de cela, le recrutement de PRC2 semble être une conséquence de l'absence d'expression des gènes. Nous proposons que la liaison de PRC2 à l'ARN pourrait servir à d'autres fonctions que le ciblage à la chromatine. Cette étude soulève des questions importantes sur le rôle des ARN non-codants et invite à une réévaluation du paradigme fonctionnel lncRNA-PRC2.

ABBREVIATIONS AND ACRONYMS

293THEK : 293T Human Emryonic Kidney

5mC: 5’ methyl-cytosine

5hmC: 5’ hydrossi-methyl-cytosine

Abd-A: Abdominal A

Abd-B: Abdominal B

AEBP2: Adipocyte Enhancer-binding Protein2

ANT-C: Antennapedia Complex

AP-y: Activator Protein 1

ARNIL: Antisense Non-coding RNA in the INK4 Locus

Ash1: Absent, Small, or homeotic-like 1

Ash2: Absent, Small, or homeotic-like 2

Asx : Additional Sex Combs proteins

ATP: Adenosine tri-phosphate

ATRX: ATP-dependent helicase ATRX, X- linked helicase II

AU022751: unknown protein

AUTS2: Autism Susceptibility Candidate 2

BAP1: BRACA-1 associated protein 1

BCOR: BCL-6 CoRepressor

Bmi1 (Pcgf4): B lymphoma Mo-MLV insertion region 1 homolog

BMP: Bone Morphogenetic Proteins

BRDT: Bromodomain Testis Associated

brm: brahma

BX-C: Bithorax Complex

Bxd: bithoraxoid

C10orf12: Chromosome 10 Open Reading Frame 12

C17orf96/EPOP: Chromosome 17 Open Reading Frame 96 or Elongin BC And Polycomb Repressive Complex 2-Associated Protein

CAF1: chromatin assembly factor

Calypso: Ubiquitin carboxyl-terminal hydrolase calypso

CBFA2T2: Core-Binding Factor, Alpha Subunit 2 Translocation 2

CBX: Chromobox

CDH1: Cadherin-1 coding gene

CK2: Casein Kinase 2

CDKN2A (ARF): Cyclin-Dependent Kinase Inhibitor 2A

ChIP: Chromatin immunoprecipitation

ClF: cell-wall inhibitor of β-fructosidase

CpG: --C--phosphate--G—

CXorf67: Chromosome X Open Reading Frame 67

CYP26B1: Cytochrome P450 Family 26 Subfamily B

DAZL : Deleted in Azoospermia-Like

DM: Drosophila Melanogaster

DNA: DeoxyriboNucleic Acid

DNMT1-3A-3B-3L: DNA (cytosine-5)-methyltransferase

E2F6: Transcription factor E2F6

EED: Embryonic Ectoderm Development

ELOBC: Elongin B and C

EMF2: Embryonic Flower 2

EMSA: electrophoretic Mobility Shift Assay

EPH1: Enhancer Of Polycomb Homolog 1

(m) ESC: cells Embryonic Stem Cells

Esc: Extra sex combs

ESS: Endometrial Stromal Sarcoma

E(Z): Enhancer of Zeste

Ezh2: Enhancer of zeste homolog 2

FIS2: Fertilization-Independent 2

FLC: Flowering Locus C

G9A alias EHMT-2: Euchromatic histone-lysine N-methyltransferase 2

GLP alias EHMT-1: Euchromatic histone-lysine N-methyltransferase 1

H1Foo: Histone H1 Family oocyte-specific

H2A/H4R3me2: asymetrical dimethylation on Arginine 3 on histone H2A and H4

H2AK119Ub: Histone H2A monoubiquitinated on lysine 119

H3K27 (or 4/9/36): Histone H3 lysine 27 (or 4/9/36)

H3K27Ac: Histone H3 acetylated on lysine 27

H3K27me3 (or 1/2): Histone H3 trimethylated (or mono-/dimethylated) on lysine 27

H3S28p: Histone H3 phosphorylated on serine 28

H4: Histone H4

HDAC1: histone deacetylase 1

HDAC3: Histone Deacetylase 3

HILS1: Histone Linker H1 Domain, Spermatid-Specific 1

HMTase: Histone MethylTransferase

HOTAIR: Hox Transcript Antisense Intergenic RNA

HOTTIP: HOXA transcript at the distal tip

HOX: Homeotic genes

HP-y: Heterochromatic Protein y

HPH: Human Polyhomeotic Homolog

IF: Immuno Fluorescence

INK4b/INK4a: Cyclin-dependent Kinase Inhibitor 4b/4°

IP: ImmunoPrecipitation

JARID2: Jumonji AT-rich Interaction Domain

JAZF1: Juxtaposed With Another Zinc Finger Protein 1

Jing: Jing, isoform J (AEBP2 homolog in Drosophila)

dKDM2/KDM2B/FBXL10: Lysine-specific demethylase 2B

KI: KnockIn

KO: Knockout

L3MBTL2: Lethal(3)Malignant Brain Tumor-like protein 2

LCOR: Ligand-dependent corepressor

LSD1/KDM1A: Lysine-specific histone Demethylase 1A or KDM1A, lysine (K)-specific demethylase 1A

MBLR6(Pcgf6): Mel18 and Bmi1-like RING finger 6

MBTD1: Malignant Brain Tumor Domain

MEA: Histone-lysine N-methyltransferase MEDEA

MEAF6: MYST/Esa1 Associated Factor 6

MEG3: Maternally expressed gene 3

MEL-18 (Pcgf2): Melanoma nuclear protein 18

MES-2,-3,-6: Maternal-effect sterile protein 2,3,6

Mls1: Malate synthase 1

Mor: Moira

MTF2: Metal Response Element Binding Transcription Factor

MVH, DDX4: Mouse Vasa Homolog, also known as *DEAD* Box Polypeptide 4

MyoD: Myogenic Differentiation 1

MyoG: MyoGenin

MSCI: Meiotic Sex Chromosome Inactivation

NANOG2 : Homeobox transcription factor Nanog 2

NSPC1 (Pcgf1) : Nervous system Polycomb-1

NURD: Nucleosome Remodeling and Deacetylation

NURF: Nucleosome Remodeling Factor

OCT4 (Pouf5f1) : Octamer-Binding Transcription Factor 4

PCGF: PolyComb Group Factor

Pc: Polycomb

PcG: Polycomb Group of Genes

Pcl: Polycomblike

m/hPGCs: mouse/human Primordial Germ Cells

Ph: Polyhomeotic

PHF1: PHD Finger Protein 1

PHF19: PHD Finger Protein 19

Pho: Protein pleiohomeotic

pI: Isoelectric Point

PRC1: Polycomb Repressive Complex 1

PRC2: Polycomb Repressive Complex 2

PRDM1 (BLIMP)/14: PR domain zinc finger protein 1

PRM1/2: Protamine 1/2

PRE: Polycomb Responsive Elements

Psc: Posterior sex combs

RbAp48: Retinoblastoma-binding protein p48

REST/NRSF: Neuron-Restrictive Silencing Factor

dRing1/2: Really interesting new gene finger protein 1/2

RING1A/B: Really Interesting New Gene finger protein 1A/B

RYBP: RING1 and YY1-Binding Protein

Runx1/CBFβ : Runt-related transcription factor 1

RNA: RiboNucleic Acid

lncRNA: long non-coding RNA

mRNA: Messenger RNA

piRNA: PIWI interacting RNA

rRNA: ribosomal RNAs

siRNA: small inhibitory RNA

snRNA: small nuclear RNA

SAM: *S*-adenosyl-L-methionin

Scm: Sex comb on midleg

Sce: Sex comb extra

SET: Su(var)3-9, Enhancer-of-zeste and Trithorax

Sf-9 cells: Spodoptera Frugiperda 9

SL2: Schnider's Drosophila Line 2

Sfmbt: Sex comb on midleg (Scm)-like with 4 Malignant Brain Tumour domains

SHARP: SMRT and HDAC Associated Repressor Protein

SNAIL-1: Snail Family Zinc Finger 1

SOX2: Sex Determining Region Y-box 2

SRM: Stimulation-Responsive Motif

SSC: Spermatogonial Stem Cells

STRA8: Stimulated by Retinoic Acid 8

SUV39H: Suppressor of variegation 3-9 Homolog 1

Su(z)12: Suppressor of zeste 12

SWI2/SNF2: SWItch/Sucrose Non-Fermentable

SWN: SWINGER, SET DOMAIN-CONTAINING PROTEIN 10

SYCP3: Synaptonemal Complex Protein 3

TAP: Tandem Affinity Purification

TET enzymes: Ten-Eleven Translocation enzymes

TH2A/H2B/H3: testis-specific variant of histone H2A/H2B/H3

TP1/2: Transition Proteins 1/2

TrxG: Thritorax Group of Genes

U2OS: Human Bone Osteosarcoma Epithelial Cells

Ubx: Ultrabithorax

UHRF: E3 ubiquitin-protein ligase UHRF1

VRN2: vernalization protein 2

WD: a tryptophan-aspartic acid dipeptide

WD40: domain of 40 amino acids, terminating in a tryptophan-aspartic acid (W-D) dipeptide

WNT: Proto-oncogene Wnt-1

XCI: X Chromosome Inactivation

XIST: X-Inactive Specific Transcript

YAF: YY-1 Associated Factor

YY-1: Yin and Yang 1

Z: zeste gene

Zfpm2: Zinc Finger Protein, FOG Family Member 2

ZNF518B: Zinc Finger Protein 518B

INTRODUCTION

CHAPTER I

FROM DEVELOPMENTAL PHENOTYPE TO CHROMATIN MODIFIER COMPLEXES

Developmental decisions guide individual cell lineages that share an identical genetic background to follow distinct pathways of differentiation and acquire unique features. This translates into a specific pattern of gene expression that is maintained through cell divisions. One of the main goals of developmental biologists is to understand which are the molecular circuits that establish and maintain precise expression patterns in different parts of the organism. Animals as diverse as worms, flies, and humans use remarkably similar sets of molecular components to control their development. *Drosophila melanogaster*, a powerful model for genetic study, was initially the preferred organism to dissect those circuits. Homeotic cluster genes encode important transcription factors that define the identities of body segments. The expression of Hox genes is in turn driven by transiently expressed transcription factors are responsible for the establishment of developmental segments. Once established, homeotic gene expression patterns are maintained throughout subsequent developmental stages. Two main groups of proteins collaborate to fulfill this task and are respectively encoded by genes of the Polycomb Group (PcG) and the Trithorax Group (TrxG).

In Drosophila Melanogaster, early patterning originates with asymmetry in the egg, organized both by localized deposition of mRNA inside the egg and by signals from the follicles cells around it. It is then followed by a series of nuclear divisions generating a syncytium. Positional information in the multinucleate embryo is supplied by intracellular gradients that are set up by the products of egg-polarity groups of genes. These genes operate by setting up graded distributions of gene regulatory

proteins. The gradients along the anteroposterior axis initiate the orderly expression of various genes among which the homeotic genes. They become expressed in specific regions of the embryo through hierarchical interactions and they progressively divide the blastoderm into a regular series of repeating units called segments. The initial activation or repression of homeotic genes in the appropriate segments during early embryogenesis depends on the graded expression profiles of transcription factors and morphogens which initiates a cascade of zygotically activated gap and pair-rule genes (Akam, 1987).

This class of genes encodes DNA-binding proteins of the homeodomain family and is organized in two clusters. The first to be discovered was the Bithorax Complex (BX-C), a large genomic locus containing three homeotic genes, Ultrabithorax (Ubx), abdominal A (Abd-A) and Abdominal B (Abd-B), which are involved in the development of the third thoracic segment and all the abdominal segments (Lewis 1951). The other homeotic gene complex, called the Antennapedia Complex (ANT-C), contains five genes that are necessary for the development of the head and the two first thoracic segments.

Later in development, Hox genes expression patterns persist despite the disappearance of these early regulators, suggesting that other factor may be involved in maintaining their expression. Homeotic phenotypes involving specific segmental transformations were thus used to perform genetic screens for factors involved in this maintenance phase. Mutations in genes of the Hox family give rise to particular transformation termed “homeotic” which result in the conversion of the identity of one or several body segments into that of distinct segments.

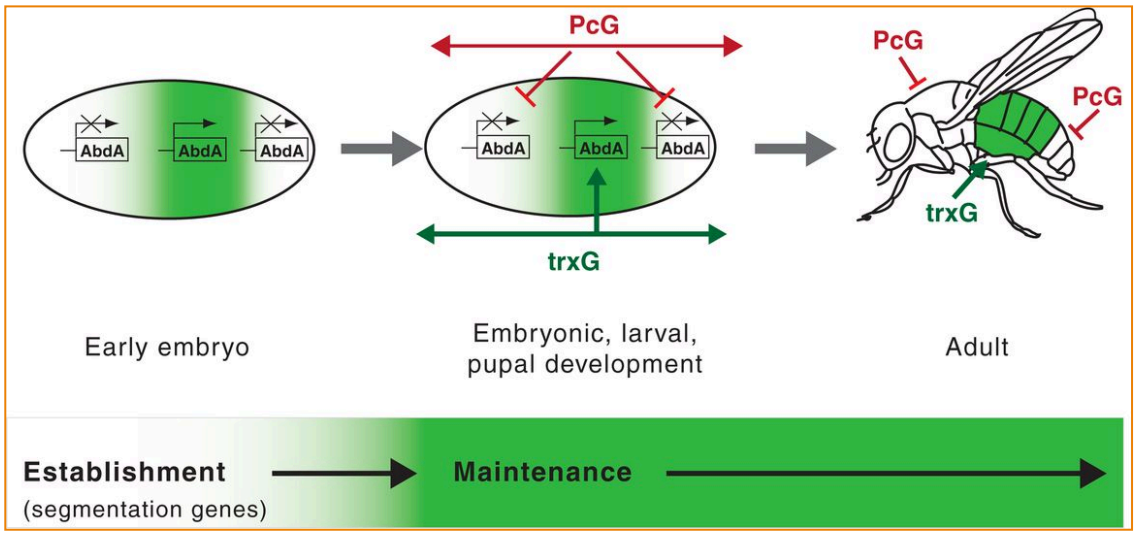


Figure 1. Regulation of Hox genes cluster transcription.

The boundaries of transcription of Abd-A and other Hox genes are established by segmentation proteins. Among them are the products of gap and pair-rule genes, which subdivide the embryo into 14 identical segments. During development, the “off” or “on” states of Hox transcription are maintained silent or active, respectively, by the ubiquitously expressed members of PcG and TrxG protein families (adapted from (Kingston & Tamkun, 2014))

1. Genetics mutations led to the discovery of Polycomb and Trithorax group of genes

1.1 Polycomb Phenotypes

The Polycomb (Pc) mutation was reported by Pam Lewis in 1947 (Lewis P, 1949). Heterozygous male Pc flies present specialized bristles called sex combs, normally restricted to the first pair of legs, on the second and third pairs of legs as well, hence the name of the mutation. In addition, heterozygous Pc flies present partial antenna to leg transformations, the characteristic phenotype of the *Antennapedia* mutant.

In his 1978 article, Edward Lewis followed up on this Polycomb mutant. From the transformation of thoracic and abdominal segment identity toward the eighth abdominal segment, he inferred that the Pc gene product must be a trans-acting negative regulator of the BX-C (Duncan and Lewis, 1982; Lewis, 1978). This hypothesis was later validated through analysis of Hox gene expression in homozygous Pc mutants, which revealed ectopic expression of Hox transcripts anterior to their normal domain of expression.

Unlike the homeotic selector loci of the ANT-C and BX-C, Pc products do not appear to "instruct" directly embryonic segment primordia to follow a particular course of development. Instead, the phenotypes observed for Pc mutants appear to result from altered patterns of ANT-C and BX-C gene expression. Duncan and Lewis (1982) postulated that the homeotic transformations seen in Pc- embryos result from indiscriminate expression in anterior abdominal segment primordia of BX-C loci that are normally expressed only in the posterior regions. In order to better understand how the ANT-C and BX-C are controlled, a systematic search was made for other genes that appear to act as regulators of these gene complexes. Another locus appears to be very similar in function to Pc and was, therefore, called Polycomblike (Pcl). Mutants in this gene cause a number of segmental transformations that resemble those caused by the known dominant gain-of-function mutants in the ANT-C and BX-C. This and the observation that the severity of particular segmental transformations in Pcl-mutant animals depends upon ANT-C and BX-C dosage suggest that Pcl, like Pc, is required for the normal control of these gene complexes. Clonal analysis experiments indicate that, at least for the BX-C, Pcl exerts this control until late in development (Duncan 1982).

Another genetic system to identify factors involved in the regulation of developmental regulators was also used: it is referred to as the *zeste-white* interaction and is based on eye pigmentation as the readout for the genetic screen. A neomorphic mutation of the zeste gene "Z^1" leads to the repression of the X-linked gene *white* only when it is present in 2 copies (i.e. in females), which may be due to physical pairing of the repressed alleles (Pirrotta, 1999). Using this system, mutations which

preferentially increase or decrease the Z^1 repressive effect (referred to as Enhancer of Zeste (E(Z)) and Suppressor of zeste (SU(Z)), respectively) were identified. For instance, E(Z) mutations result in the modulation of the Z^1 mutant phenotype through the abberant expression of *white* gene. Importantly, ectopic expression of ANT-C and BX-C genes were also reported in this mutant (Wu et al., 1989; Jones and Gelbart, 1990). Su(z)12 mutants show very strong homeotic phenotypes caused by widespread misexpression of HOX genes and are among the strongest PcG mutants. Su(z)12 function is needed for the development of germ cells and is very well conserved in vertebrates and plants (Birve et al, 2001). The homeotic gene extra sex combs (esc) is required in early embryo development to properly silence homeotic genes in the Bithorax complex (Struhl, 1981). Embryos lacking Esc show misexpression of Hox genes and the role of Esc seems to be independent from that of Pc (Struhl, 1983).

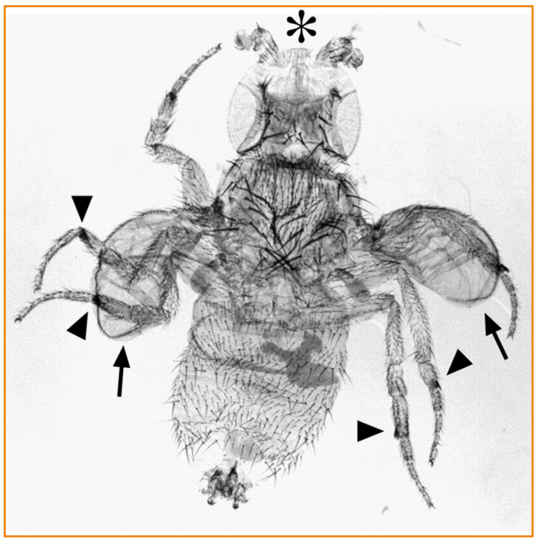


Figure 2. Su(z)12 mutations associated phenotypes in DM.

Homeotic transformations in a *Su(z)12* mutant pharate adult male. Homeotic transformations are evident in several body segments. Sex combs, a structure normally only present on the first leg, are present on the first tarsal segments of all meso- and metathoracic legs (arrowheads); the antennae are partially transformed into legs (asterisk) and wings are much smaller and partially transformed into haltere-like structures (arrows). These homeotic transformations are consistent with inappropriate activation of several ANTC and BXC genes in the imaginal disc primordia of these structures (adapted from Birve et al, 2001).

1.2 Discovery of the Trithorax phenotype

The trithorax (trx) mutation was isolated in Drosophila as a natural mutation that causes transformation of segments into more anterior ones, reflecting a loss-of-function of homeotic genes (Ingham 1980). Two more mutations, in the ash1 and ash2 genes, were isolated from animals with defective disc phenotypes indicating loss of function of Hox genes. These mutations were highly

similar to the trx mutant background, were enhanced in combination with trx alleles and suppressed in Pc mutants (Shearn A, 1989). Consistent with the phenotype, it was shown that the trithorax mutation could suppress the Polycomb phenotype (Ingham 1983), and additional tritohrax mutants were also identified as modulators of the Polycomb phenotype (Kennison 1988).

2. PcG proteins form multi-subunit complexes

Early genetic evidence suggesting that PcG proteins contribute to the same process came from the observation that inactivation of different PcG genes led to comparable patterns of aberrant gene expression outside of their normal domains (Simon et al, 1992). It was noticed that different PcG proteins were expressed at the same time and place in the developing embryo and they colocalized at chromosomal sites corresponding to silent homeotic genes. The strong colocalization of Pc and Polyhomeotic (Ph) was suggestive of direct interaction between these two proteins (DeCamillis et al, 1992). The Pc protein is a nuclear protein associated with 100 different sites on polytene chromosomes of larval salivary glands (Zink and Paro, 1989; Zink, 1990). It specifically binds to regulatory sequences of the *Antennapedia* gene (Zink et al, 1991). Importantly, reporter genes linked to homeotic cis-regulatory sequences become ectopically expressed when tested in a background lacking Pc function (Muller and Bienz, 1991; Zink et al, 1991).
Genetic analysis of the polyhomeotic gene revealed multiple developmental defects in mutant flies. While the homeotic transformations observed classified it as a PcG gene, additional defects like cell death in the ventral epidermis (Dura et al, 1987) and misrouting of central nervous system axons (Smouse et al, 1988; Smouse and Perrimon, 1990) were also reported. The Ph gene is arranged as a tandemly duplicated unit. Only lesions in both units result in null or strong hypomorphic alleles, whereas single lesion alleles display only weak hypomorphic phenotypes. The two genetic units contained within a stretch of 25kb of genomic DNA encode similar proteins that have certain sequence characteristics of proteins interacting with DNA (Deatrick et al, 1991; DeCamillis et al., 1992). In addition to four blocks of glutamine repeats and serine/threonine-rich sequences, the Ph protein displays a region with partial homology to the helix-loop-helix motif and a single C4 zinc finger. Like Pc, Ph binds to 100 polytene chromosome sites and Ph specifically recognizes regulatory sequences from the bithoraxoid region (DeCamillis et al, 1992). Of note, this binding can be recapitulated with the insertion of a construct containing a 14.5-kb fragment of bithoraxoid [bxd] regulatory DNA that is genetically sensitive to Ph as well as other PcG mutant backgrounds. This tight link between Pc and Ph was further supported by co-immunoprecipitations of both proteins and size-exclusion chromatography indicating that they are part of a soluble multimer (Franke et al, 1992).

Meanwhile, several lines of evidence were also consistent with a molecular connection between Esc and another member of the PcG, Enhancer of zeste [E(z)]. Indeed, a direct interaction between human

homologs of Esc and E(z) was reported and proposed to be evolutionarily conserved (Jones et al, 1998). Furthermore, E(Z) homozygous mutant females in Drosophila give birth to homeotic transformants embryos due to the maternal effect of that protein. Derived offspring displayed posterior hometic transformation that resembled phenoypes of embryos born from esc mutant females (Jones et al, 1990; Struhl et al, 1981). Moreover, a balance in the relative concentrations of the esc and E(z) proteins was proposed to be important for homeotic gene repression (Campbell et al, 1995). This interaction might involve the highly conserved WD region of Esc mediating binding to a conserved N-terminal sequence of E(Z). Supporting this view, mutations in the WD region that perturb ESC function in vivo also perturb binding to E(Z) in vitro (Tie et al, 1998). E(Z) had been shown to be required for the binding of other PcG proteins to chromosomes (Rastelli et al., 1993) and to co-localize with other PcG proteins at many sites (Carrington and Jones, 1996).

These observations were confirmed with mammalian homologs of PcG that are also separated into roughly two complexes, one containing homologs of Pc, Psc, and Ph (Alkema et al, 1997; Satijn et al, 1997) and which would later come to be known as PRC1 (Polycomb Repressive Complex 1), and the other containing homologs to E(Z) and Esc (Sewalt et al, 1998; van Lohuizen et al, 1998), eventually called PRC2 (Polycomb Repressive Complex 2).
PRC1 was first purified via either tagged Ph or Psc and shown to contain at least Pc, Psc, Ph and Scm (Shao 1999). Previous studies using immunoprecipitation, in vitro binding, and/or yeast two-hybrid analysis had shown that Pc, Psc, and Ph interact with each other (Strutt and Paro, 1997; Kyba and Brock, 1998), and that Scm interacts with Ph (Peterson et al., 1997). Separately, E(Z) and Esc were shown to interact with each other by similar approaches (Jones et al, 1998; Tie et al, 1998), and E(Z) separates from PRC1 by chromatography (Shao et al. 2001). Indeed, ESC and E(Z) co-elute in stable complexes of about 600 kDa, arguing against a simple heterodimer of Esc (50 kDa) and E(Z) (90 kDa). The Esc-E(Z) complex also contains the histone-binding protein p55, which is also a subunit of the CAF1 chromatin assembly factor (Tyler et al., 1996) and the nucleosome remodeling factor NURF (Martinez-Balbas et al., 1998). A direct molecular link between the complex and chromatin through p55 and a histone deacetylase suggested that histone deacetylation might be involved in the mechanism of Esc/E(Z)-mediated silencing (Tie et al. 2001).

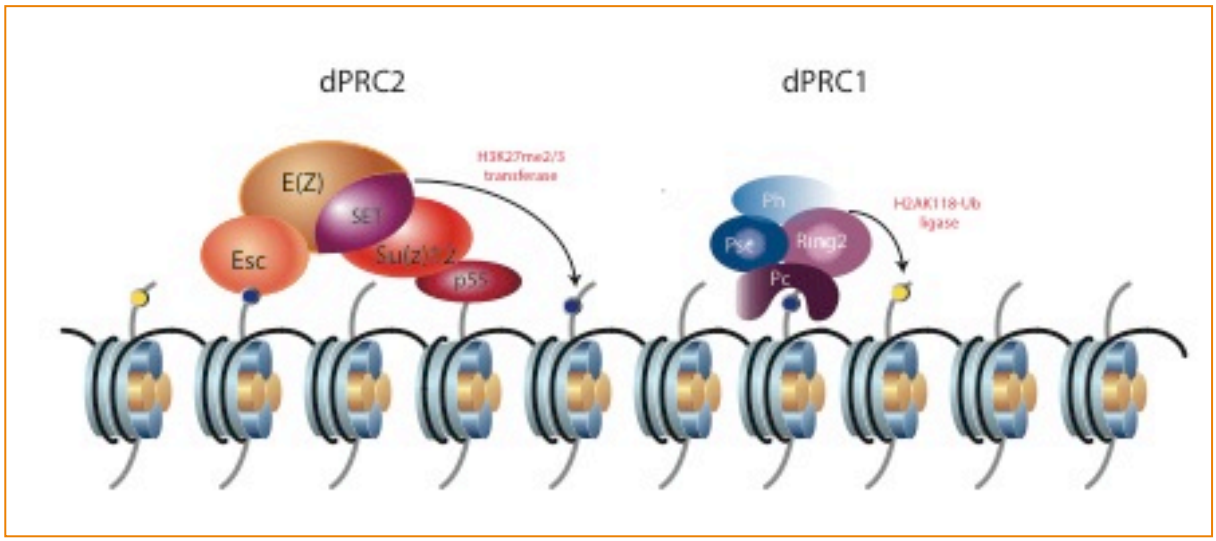


Figure 3. Composition of Polycomb Repressive Complexes in Drosophila Melanogaster

Polycomb repressive complex 2 (PRC2) core complex is composed of the catalytic subunit E(Z) with its SET methyltransferase domain, Extra Sex Combs (Esc), Suppressor of Zeste SU(Z)12 and p55. Polycomb repressive complex 1 (PRC1) core complex is composed of Polycomb (Pc), the chromodomain containing subunit, Polyhomeotic (Ph), Posterior Sex Combs (Psc) and the ubiquitinating subunit Ring2.

3. Towards PcG function

PcG proteins has been proposed to act on their target genes via binding DNA sequences capable of maintaining repressive states called Polycomb Responsive Elements (PRE). These sequences were first identified within homeotic gene complexes and Pc protein binding has been detected with specific antibodies along BX-C by Chromatin Immunoprecipitation (Orlando, Jane, Chinwalla, Harte, & Paro, 1998). PREs were extensively studied in transgenic assays to test their ability to maintain the repression of the target genes. PcG-mediated silencing has been suggested to involve the creation of altered chromatin domains (Paro & Hogness, 1991) that could render transcription factor-binding sites inaccessible (Bunker & Kingston, 1994; McCall & Bender, 1996) and/or prevent enhancer-promoter communication (Pirrotta, 1997; Pirrotta Vincenzo & Rastelli L, 1994). Disruption of PRE-mediated silencing of a reporter can be induced by a potent activator and is accompanied by local accumulation of hyperacetylated histone H4, a signature of transcriptionally active genes (Cavalli & Paro, 1999). This indirectly suggests that local histone deacetylation may accompany PcG silencing. The Pc protein has first been shown to bind to nucleosomes in vitro (Breiling, Bonte, Ferrari, Becker, & Paro, 1999), consistent with the possibility that it bears a chromatin-associated function. Purified complex PRC1, containing the PcG proteins Pc, Ph, Psc and Scm, was shown to inhibit the in vitro ATP-dependent nucleosome remodeling activity of a purified human SWI/SNF complex (Shao et al., 1999). This raises the possibility that PRC1 alters chromatin in some way that renders it refractory to the action of

nucleosome remodeling complexes, although the basis of its inhibition of remodeling was not yet understood. Studies of mammalian PcG proteins indicate that a fourth protein, RING1, interacts with homologs of Pc, Ph, and Psc (Schoorlemmer et al., 1997). Furthermore, *Drosophila* PRC1 contains apparently stoichiometric amounts of Pc, Ph, Psc (Shao et al., 1999), and dRING1, and does not contain large amounts of any other PcG proteins. These four polypeptides complex form a stable complex that results in chromatin remodeling inhibition (Francis, Saurin, Shao, & Kingston, 2001).

Importantly, dRing protein was subsequently purified and characterized as an E3 ubiquitin ligase that is specific for the K119 of histone H2A. By chromatin immunoprecipitation analysis it was observed that dRing colocalizes with ubiquitinated H2A at the PRE and promoter regions of the Drosophila *Ubx* gene in wing imaginal discs. Removal of dRing in SL2 tissue culture cells by RNA interference resulted in loss of H2A ubiquitination concomitant with derepression of Ubx. Thus, the H2A ubiquitin ligase has been identified, and H2A ubiquitination has been associated to Polycomb silencing (H. Wang et al., 2004). Consistently, reducing the expression of Ring2 results in a dramatic decrease in the level of ubiquitinated H2A (H. Wang et al., 2004).

Initial characterization of the Esc-E(Z) embryonic complex showed that it contains the PcG proteins Extra Sex Combs (ESC) and Enhancer of Zeste [E(Z)] as well as NURF-55, histone binding protein (J. Ng, Hart, Morgan, & Simon, 2000; Tie, Furuyama, Prasad-Sinha, Jane, & Harte, 2001). NURF-55 is the fly homolog of RbAp48, which copurifies with a human histone deacetylase (Taunton, Hassig, & Schreiber, 1996). This connection, together with physical interactions between Esc, E(Z), NURF-55, and fly HDAC1 have led to the suggestion that the *Drosophila* Esc-E(Z) complex might function as a histone deacetylase (Tie et al., 2001). However, HDAC activity has not been consistently found in the purified fly Esc-E(Z) complex. It was noticed that E(Z) protein shares a domain with *Drosophila* Su(var)3-9, and Trithorax proteins (Jones & Gelbart, 1993; Tschiersch et al., 1994), a domain thus refered as SET domain but those function was unknown. The real breakthrough in understanding the molecular function of the complex came from the characterization of SUV39H protein SET domain. This domain has histone methyltransferase (HMTase) activity with specificity for lysine 9 (K9) in the histone H3 tail, a histone modification implicated in targeting of heterochromatin proteins (Bannister et al., 2001; Lachner, O'Carroll, Rea, Mechtler, & Jenuwein, 2001). Subsequently, studies on several other SET domain proteins showed that they also possess HMTase activities with different lysine residue specificities (Kuzmichev, Nishioka, Edrjument-Bromage, Tempst, & Reinberg, 2002; Nakayama, Rice, Strahl, Allis, & Grewal, 2001; Strahl et al., 2002; H. Wang et al., 2001). Esc-E(Z) complex was purified from *Drosophila* embryos and human cells. It contains four core subunits: Esc, E(Z), NURF-55, and the PcG protein, SU(Z)12. A complex reconstituted from these four proteins bears H3-K27-directed HMTase activity. Residues in the E(Z) SET domain are required for HMTase activity in vitro and HOX gene repression in vivo. The lack of H3 methyltransferase activity in complexes containing E(Z) with SET domain mutations implies that E(Z) carries the catalytic function in the complex. The essential roles of other subunits may be in targeting or modulating activity of the

complex. These findings identify a catalytic activity of a PcG complex and imply that histone methylation is part of the mechanism of PcG transcriptional memory during development (Cao et al., 2002; Czermin et al., 2002; Kuzmichev et al., 2002; Müller et al., 2002).

A crosstalk between PRC1 and PRC2 was proposed based on the observation that methylation on H3-K27 facilitates binding of Polycomb (Pc), a component of the PRC1 complex to chromatin. They elaborate a model in which Esc-E(Z)–mediated H3-K27 methylation serves as a signal for the recruitment of the PRC1 complex by facilitating PC binding. Recruitment of PRC1 in turn prevents the access of nucleosome remodeling factors, such as SWI/SNF, leading to the formation of a repressive chromatin state (Cao & Zhang, 2004). This was considered as the central dogma of Polycomb maintainance of gene repression.

4. Trithorax multimeric complexes

Genetic studies have suggested that trx-G proteins may also physically interact to regulate the transcription of homeotic genes. Evidences suggesting that a complex of trx-G proteins regulates the transcription of homeotic genes have come from studies of a trx-G gene, *brahma* (*brm*). *brm* was identified in a screen for extragenic suppressors of *Pc* mutations (Kennison & Tamkunt, 1988) and subsequently shown to be highly related to a yeast protein, SWI2/SNF2, which functions as the ATPase subunit of a 2 MDa chromatin remodeling complex, the SWI/SNF complex. The BRM complex was purified from *Drosophila* embryos and the majority of the subunits of this complex is not encoded by Trithorax group genes. Furthermore, a screen for enhancers of a dominant-negative *brm* mutation identified only one Trithorax group gene, *moira* (*mor*) that appears to be essential for *brm* function in vivo. Two additional complexes containing trithorax group proteins exist: a 2 MDa ASH1 complex and a 500 kDa ASH2 complex (Papoulas et al., 1998).

5. PcG conservation

PcG proteins were first described in Drosophila Melanogaster as key factors to maintain cell-fate decisions throughout development by repressing Hox genes in a body-segment specific manner. They are now recognized as a large family of proteins conserved in most eukaryotes (Table 1). To examine the evolutionary history of those proteins, it was performed an extensive bioinformatic analysis to find homologs in diverse multicellular organisms using mains subunit of PRC1 and PRC2 complexes. To this aim, a sequence-similarity method between domains and full-length sequence was used to identify homologs.

Drosophila	**Mammals**	**Complex**
Polycomb	CBX2,4,6,7,8	PRC1
Polyhomeotic	HPH1/2/3	PRC1
Posterior Sex Comb	NSPC1 (PCGF1), MEL18 (PCGF2), PCGF3, BMI1 (PCGF4), PCGF5, MBLR6 (PCGF6)	PRC1
Sex Comb Extra/Ring	RING1A/B	PRC1
Sex Comb on midleg	SCMH1/2	PRC1 associated?
Enhancer of zeste	EZH1/2	PRC2
Suppressor of zeste 12	SUZ12	PRC2
Extra sex comb	EED	PRC2
Polycomblike	PHF1/MTF2/PCL3	PRC2 associated

Table 1. Summary table of main Polycomb Repressive Complex 1 and 2 Homologs

Representative list of main Polycomb Repressive Complex 1 and 2 Homologs in Drosophila Melanogaster (left column), Mammals (central column) and the associated complex (right column).

While the conservation of PRC2 is well admitted, the case of PRC1 is more debated. Initially, PRC1 was reported to be absent in fungi and plants genomes (reviewed in (Schuettengruber, Chourrout, Vervoort, Leblanc, & Cavalli, 2007)), whereas it is now known to exist in Plants (Bratzel, Lopez-Torrejon, Koch, Del Pozo, & Calonje, 2010; Sanchez-Pulido, Devos, Sung, & Calonje, 2008). In mammals, PRC1 components have undergone multiple duplications as compare to invertebrate such as flies. An exemple is the Pc gene, with up to five Pc homologs in vertebrates with different domains structure and biochemical properties, while just one isoform is present in the invertebrate genome. Pc homologs are known as Chromobox (Cbx), which all have highly conserved CDs and Pc boxes. Paralogs differ greatly in length and in the presence of other domains and motifs; these factors might contribute to differential functions as indicated by the distinct phenotype observed upon overexpression. Hence, the overexpression of Cbx7 or Cbx8 but not Cbx4 (Cbx2 and Cbx6 were not tested) bypasses replicative senescence in human and mouse fibroblasts (Gil, Bernard, Martínez, & Beach, 2004). Mel-18 and Bmi1 (two of six Psc homologs in mammals) are also likely to be non-redundant paralogs, despite their 63% amino acid sequence identity. Bmi1- and Mel-18-deficient mice display similar but unique phenotypes, and only 30% of Bmi1-regulated genes were found to be coregulated by Mel-18 and vice-versa (Wiederschain et al., 2007). Vertebrate homologs of dRing, Ring1A and Ring1B exhibit also functional divergence. Although they share long stretches of high conservation, Ring1A- and Ring1B-deficient mice show different phenotypes. Mice heterozygous for Ring1A exhibit classic homeotic transformations and skeletal defects, whereas Ring1B heterozygous

mice show no skeletal phenotype while it is essential for normal gastrulation, and null embryos do not survive past embryo day 10.5. The differential severity of these phenotypes correlates with the extent of H2Aub depletion in these knockouts. Global H2Aub is drastically reduced in Ring1B- but not Ring1A-null ES cells. Full-length recombinant Ring1B but not Ring1A has ubiquitin-ligase activity for H2A and Ring1B association with Ring1A enhances this activity. Although global H2Aub is drastically reduced in Ring1B-null cells, H2Aub staining is maintained on the inactive X chromosome; only in Ring1B/Ring1A double-knockout cells is H2Aub lost from this structure, revealing functional redundancy of Ring1A and Ring1B in some contexts (del Mar Lorente et al., 2000; Napoles et al., 2004; Voncken et al., 2003).

PRC2 role is classically associated with the needs of multicellular organisms to have specialized cell types. But PRC2 appears also in unicellular organisms. PRC2 is not present in distantly-related fungi like *Schyzosaccaromices Cerevisiae* and *Schyzosaccaromices Pombe*, but the pathogenic yeast *Criptococcus Neoformans* bears a PRC2 complex including Ezh2, Eed1, Mls1 (RbAp48 homolog) with two additional subunits (Fig.4). The complex is responsible for H3K27 methylation and gene silencing at all sub-telomeric regions of the genome. *C. Neoformans* genome codes also for Ccc subunit, which contains a chromodomain with affinity for H3K27 (Dumesic et al., 2015). In the unicellular alga *Chlamydomonas Reinhardtii*, PRC2 orthologs are found and knockdown of the SET-domain protein EZH causes reactivation of transgenes and retrotransposons. Moreover *Tetrahymena Thermophila* and *Paramecium Tetraurelia* eliminate repetitive sequences from the macronuclear chromosomes after sexual conjugation to maintain genomic stability involving EZH2 orthologs and their H3K27 methylation activity (Lhuillier-Akakpo et al., 2014). This led to the proposal that PRC2 first evolved in the last universal eukaryotic ancestor to maintain genome stability (Shaver, Casas-Mollano, Cerny, & Cerutti, 2010).

The nematode worm *Caenorhabditis Elegans* harbors a smaller PRC2 complex, consisting of the catalytic H3K27 methyltransferase MES-2, the EED/ESC ortholog MES-6, and the worm-specific subunit MES-3. PRC2 in C. elegans is critical for the development of the germline, in other tissus, PRC2 has been reported to play an important role in regulating cell plasticity (Fig.4).

In *Arabidopsis Thaliana* PRC2 plays major developmental roles, which are proposed to correspond to distinct complexes defined by the SUZ12-like subunit (Fig.4): VRN2, EMF2, and FIS2 and E(Z) paralogs: ClF, SWN and MEA. The VRN2 complex is required for vernalization, the process by which flowering is strictly licensed by an extended period of exposure to cold temperatures through progressive silencing of the master regulator gene FLC, accompanied by local H3K27 methylation. EMF2 prevents premature flowering through a separate genetic circuit, while FIS2 ensures that the seed development program is repressed until fertilization has taken place. Thus, the plant PRC2 complexes are responsible for maintaining transcriptional states that underpin specific cellular behaviors similarly to metazoans (Chanvivattana et al., 2004; De Lucia, Crevillen, Jones, Greb, & Dean, 2008; D. Wang, Tyson, Jackson, & Yadegari, 2006).

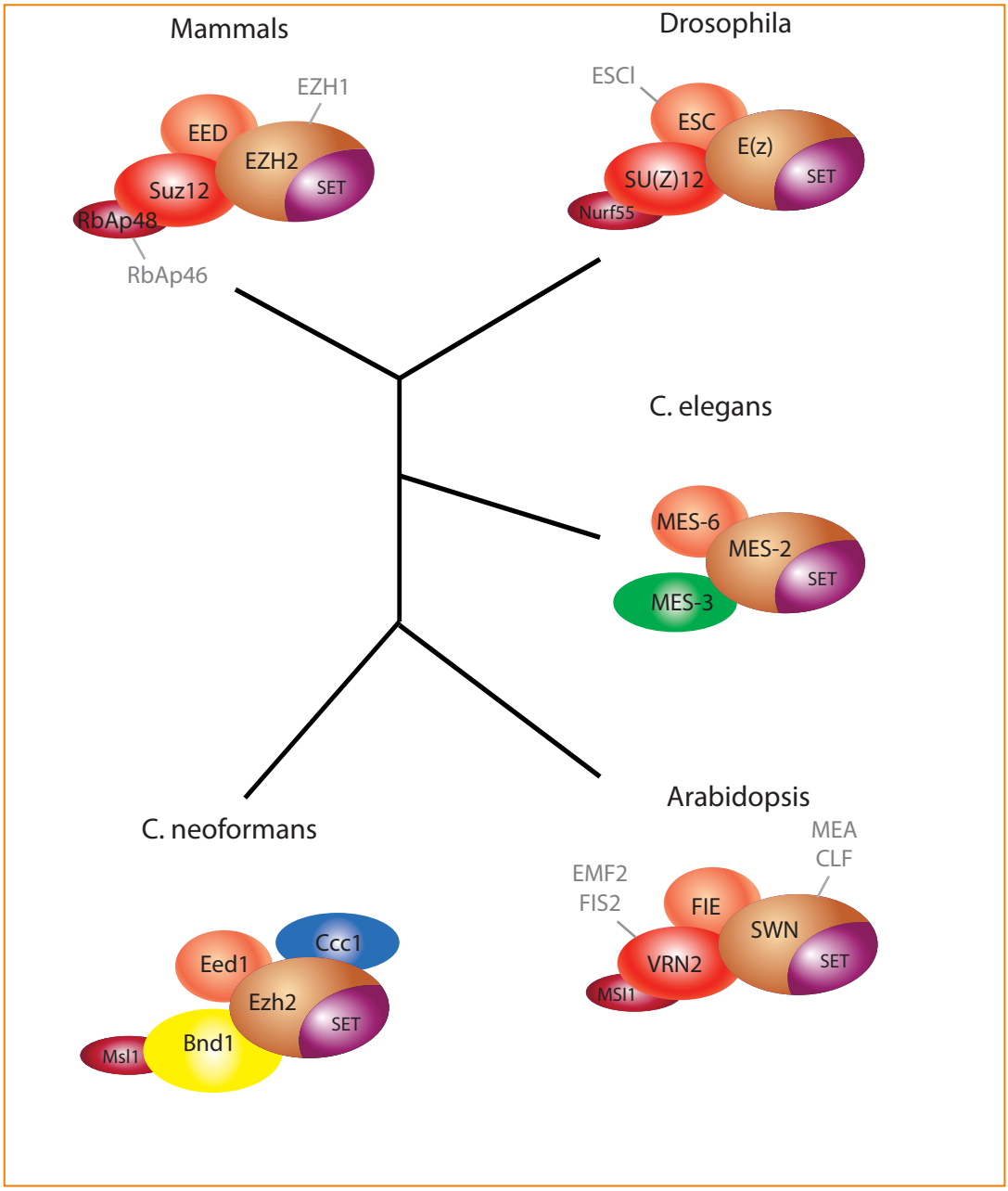


Figure 4. Evolutionary conservation and divergence of Polycomb Repressive Complex 2

Orthologs of all four core components of PRC2 have been found in mammals Drosophila and as well in plants (*Arabidopsis*) ; in the nematode *C. Elegans* a small complex similar to PRC2 is present and also the patogenic yeast *Criptococco Neoformans* bears a PRC2 complex with Ezh2, Eed, Msl1 (RbAp48 homolog), Bnd1 and Ccc1, a chromodomain containing subunit (Adapted by Holoch and Margueron Chapter IX, Vincenzo Pirrotta, 2017).

CHAPTER II

POLYCOMB GROUP OF PROTEINS

Polycomb proteins were initially identified in Drosophila Melanogaster as negative regulators of Hox gene cluster necessary for proper body segmentation during development. The characterization of mutants showing a similar Polycomb phenotype leads to the identification of 18 PcG ortholog genes. The number of PcG orthologs clearly expanded in metazoans evolution, reaching the number of 37 members in mammals.

PcG proteins associate to form functionally distinct complexes that belong to two major families: Polycomb Repressive complexes 1 and 2 (PRC1 and PRC2). Both complexes bear catalytic activity: PRC1 complexes have E3 ligase activity and their main characterized substrate is the monoubiquitinated form of histone H2A at lysine 119 (H2Aub1), while PRC2 is mainly responsible for di/tri-methylation of lysine 27 of histone H3 (H3K27me2/3). Despite association with different accessory subunits that modulate their activity, both complexes always include core components with basal catalytic activity in vitro. E3 ubiquitin ligase Ring1B (Ring1, the product of Sex Combs Extra *Sce* in Drosophila) and one Polycomb group of ring finger protein (Pcgf alias *Psc*, Posterior Sex Combs in Drosophila) represent PRC1 core complex. PRC2 core complex is constituted by suppressor of zeste (Suz12, homolog to SU(Z) protein in Drosophila) which contains a Zinc finger domain, embryonic ectoderm development (Eed or *esc*, Extra Sex Combs in Drosophila) which contains a WD40 domain recognizing tri-methylated peptides and Ezh1/2 proteins (Enhancer of zeste or E(Z) in Drosophila) which contains catalytic domain (reviewed (Aranda, Mas, & Di Croce, 2015; Schwartz & Pirrotta, 2013)).

1. PRC1 Complexes

The first PRC1 complex identified in the fruit fly contains: i) Pc, a protein with a chromodomain capable of binding H3K27me3 (Fischle et al., 2003; Min, Zhang, & Xu, 2003), ii) dRING, the enzymatic subunit essential for H2A monoubiquitination, iii) Posterior sex combs, which is described as important for chromatin compaction and iv) Polyhomeotic Ph. Later, a different complex called dRING-associated factors (dRAF) containing dRING, Psc and the histone demethylase Kdm2 was described (Lagarou et al., 2008). The scenario is more complicated in mammals where each Drosophila subunit of PRC1 has several homologs and they can associate in combinatorial fashion (Fig.5). In the past few years, comprehensive genomic approaches and genome-wide analysis have been undertaken in order to characterize the different existing complexes (Gao et al., 2012). All complexes contain the catalytic subunit Ring1B and one of the 6 Psc homologs, Polycomb Ring Fingers proteins (Pcgf). Together these two proteins are responsible for H2AK119 monoubiquitination (K118ub in Drosophila), although an efficient catalysis has been observed when dKdm2/KDM2B demethylase is present in the complex (Farcas et al., 2012; Lagarou et al., 2008; X. Wu, Johansen, & Helin, 2013a). Pc and Ph homologs are not required in vitro for efficient catalysis and are incompatible with dKDM2 (Lagarou et al., 2008). Additionally, dKDM2 mutant impacts on H2AK118ub level whereas Pc and Ph do not (Lagarou et al., 2008). These results underline the functional diversity among PRC1 complexes. According to the Pcgf subunit associated to the complex, six PRC1 subtypes are defined (PRC1.1-6). In PRC1.2 and PRC1.4 variants, Ring1B associates with Pcgf2/Mel-18 or Pcgf4/BMI-1 respectively together with a Cbx chromodomain containing protein and a Polyhomeotic Homolog Protein (HPH1-3). PRC1.1, PRC1.3, PRC1.5 and PRC1.6 are composed Pcgf1/NsPC1, Pcgf3, Pcgf5 and Pcgf6/MBLR6 respectively in association with Ring1 and Yy1 binding protein Rybp or its homolog Yaf1. CBX and RYBP proteins are mutually exclusive when binding to PRC1, and both complexes are present in various cell type (Gao et al., 2012; Morey et al., 2013, 2012).

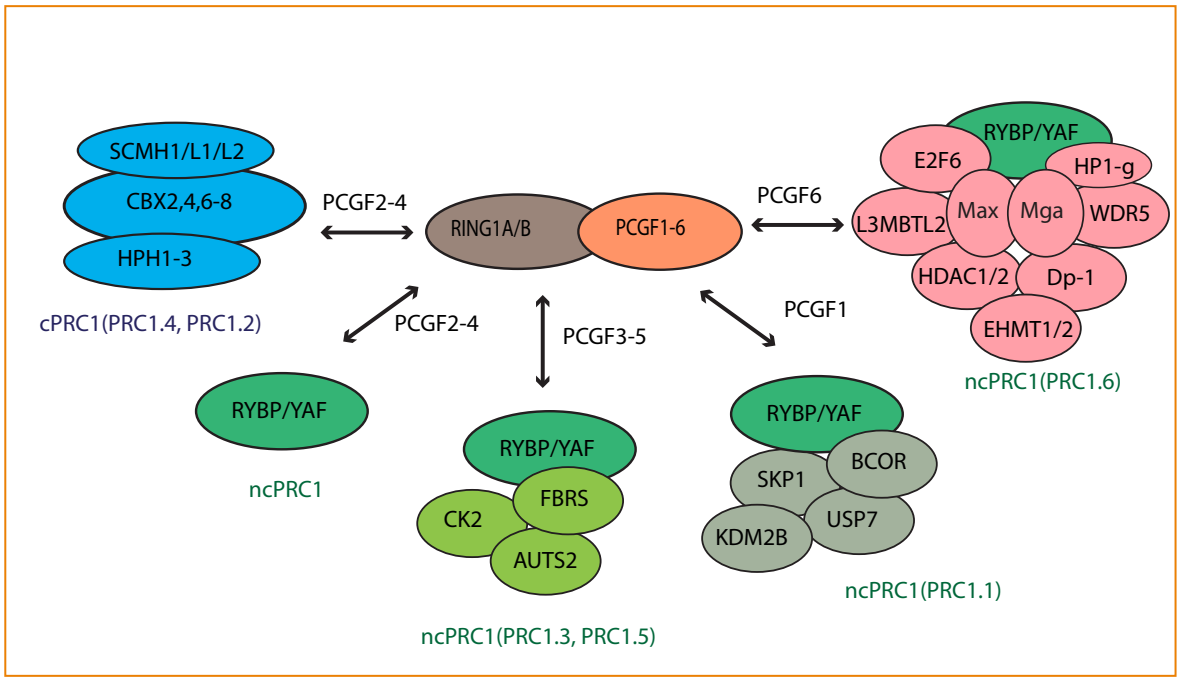


Figure 5. PRC1 complexes in mammals.

PRC1core complex can associate with distinct Pcgf proteins, which allows for an alternative nomenclature. Therefore, Pcgf2 and Pcgf4 are present in the cPRC1 complexes (PRC1.2 and PRC1.4), Pcgf2 and Pcgf4 are also associated with ncPRC1-containing Rybp or YAF proteins, Pcgf3 and Pcgf5 are present in the ncPRC1 complexes (PRC1.3 and PRC1.5), Pcgf1 is present in the ncPRC1 complex PRC1.1 (also known as BCOR), and Pcgf6 is present in the ncPRC1 complex PRC1.6 (also known as E2F6).

1.1 Canonical PRC1 complexes: PRC1.2 & PRC1.4.

The canonical PRC1 complexes are the ones containing the four different orthologs, CBX (polycomb), PCGF (polycomb group factor), HPH (human polyhomeotic homolog), and the E3-ligase protein (RING).

The CBX family members (CBX2, CBX4, CBX6, CBX7, and CBX8) physically interact with H3K27me3 via their chromodomains, promoting the recruitment and stabilization of PRC1 to specific regions of the chromatin. Mammalian CBX proteins exhibit differential bindings to methylated histone tails. CBX2 and CBX7 bind to both H3K9me3 and H3K27me3 whereas CBX4 shows stronger affinity for H3K9me3 (Bernstein et al., 2006). A DNA binding motif, AT-hook (in CBX2) or an AT-hook like motif (in the other four CBX proteins) is also present (Senthilkumar & Mishra, 2009). Live imaging revealed that a less conserved sequence in the middle of the CBX proteins might play a role in specifically directing each CBX family member to distinct regions of the chromatin (Vincenz & Kerppola, 2008). Tandem affinity purification (TAP) approach coupled with tandem mass

spectrometry (MS/MS) methodologies revealed that CBX proteins are mutually exclusive and interact with different cellular partners (Vandamme, Völkel, Rosnoblet, Le Faou, & Angrand, 2011).
It has been also proved that Cbx proteins play distinct roles during mouse embryonic stem cells pluripotency and differentiation. Cbx7 is proposed to be important in the pluripotent state, while Cbx2 and Cbx4 are required for unique aspects of early lineage specification. Cbx7 represses Cbx2, Cbx4, and Cbx8 in pluripotent cells to prevent premature differentiation, whereas during differentiation, Cbx7 and pluripotency genes are in turn repressed to enable lineage specification (Morey et al., 2012).

1.2 Non canonical PRC1: PRC1.3, PRC1.5 & PRC1.6.

In the past few years, non-canonical PRC1 complexes have been identified to repress transcription independently of H3K27me3 presence and lacking CBX protein. In fact, Ring1B association to chromatin and H2AK119Ub1 levels are preserved upon PRC2 loss. Rybp–PRC1 mediates H2A ubiquitination independently of H3K27me3 and is essential for its global maintenance in embryonic stem cells since depletion of Rybp decreases H2AK119ub in absence of Eed. Similarly, RYBP-PRC1 is recruited in response to Xist RNA expression and in H3K27me3-independent fashion (Tavares et al., 2012).
PRC1.1 complex (also named BCOR) was first purified from HEK293T and HeLa cells and it is associated with BCL6 gene silencing in B lymphocytes (Gearhart, Corcoran, Wamstad, & Bardwell, 2006). Biochemical purifications from erythroid cells confirms the interaction between the mammalian homolog KDM2B/FBX110, BCOR and PCGF1. Kdm2B bears a CXXC domain with DNA binding activity and was proposed to be important for PRC1 targeting to CG islands (Farcas et al., 2012; X. Wu et al., 2013a). Hence, KDM2B tethering results in PRC1 de novo recruitment and loss of KDM2B CXXC domain results in loss of ncPRC1 occupancy and in homeotic transformation phenotype (Blackledge et al., 2014a; Cooper et al., 2014). Of note, another study rather suggest that the function of Kdm2b is to prevent DNA methylation which could impact recruitment of the PRCs complexes (Boulard, Edwards, & Bestor, 2015).
Previous studies revealed that PCGF3 and PCGF5 form variant PRC1 complexes which contain autism susceptibility candidate 2 (AUTS2). *AUTS2* is often disrupted in patients with neuronal disorders. In neurodevelopmental processes, Casein Kinase 2 (CDK2) component of PRC1–AUTS2 neutralizes Ring1B repressive activity, while AUTS2 mediates recruitment of P300 leading to gene activation. It has been further demonstrated by ChIP-seq analysis that AUTS2 regulates neuronal gene expression through promoter association (Gao Z et al, 2014). Recently, it has been observed that PCGF3/5-PRC1 complex is able to interact with a specific sequence of Xist non-coding RNA bringing PRC1 and PRC2 on the inactive X (Almeida, Pintacuda, Masui, Koseki, Gdula, Cerase, Brown, Mould, Nakayama, et al., 2017).

PRC1.6 complex variant contains Ring1B-Pcgf6 together with E2F6 transcriptional repressor whose deletion is responsible for homeotic transformation in the axial skeleton (Storre et al., 2002). L3MBTL2, a protein whose inactivation is embryonic lethal, takes part to this complex and colocalizes with E2F6 at many target genes. In absence of L3MBTL2, H2A monoubiquitination levels are lower (Qin et al., 2012; Trojer et al., 2011). The presence of the transcription factors Max and Mga suggest the binding of the complex to the E boxes while CBX3/HP-Y could stabilize the binding of PRC1.6 to different histone modifications (Qin et al., 2012). This complex was reported to have an important function in maintaining germ-cell related genes (Endoh et al., 2017).
Further studies are needed to enlighten the exact role of each of these proteins.

2. PRC1 functions

In order to investigate H2A monoubiquitination functions, RING1B mutants have been used (Endoh et al., 2012; Eskeland et al., 2010). These studies were performed in a RING1A knockout background to avoid the redundancy between both proteins. Two classes of differentially expressed genes resulted from these analysis; one class strictly requires H2AK119Ub for silencing and another one is able to silence independently from that mark (Endoh et al., 2012). Possibly, H2A monoubiquitination is not the only available mechanism. PRC1 activity is described as important to mediate chromatin compaction. By electron microscopy, it was shown that beads-on-a-string open chromatin take a knot-like structure in presence of PRC1 (Francis et al., 2001; Grau et al., 2011). Moreover, a catalytically inactive Ring1B is still able to restore in vivo chromatin compaction suggesting that this activity is independent of the H2A ubiquitination function (Eskeland et al., 2010). The role of H2A ubiquitination is further challenged by the report that preventing H2A ubiquitination by mutated the targeted lysine does not lead to a polycomb phenotype in drosophila and does not recapitulate the phenotype observed upon interference with PRC1 components (Pengelly, Kalb, Finkl, & Muller, 2015).
Recently the panorama has been further complicated since PcG protein are thought to participate also in gene activation. A switch from CBX7 subunit to CBX8 seems to be responsible for gene activation and subsequent colocalization with H3K36me3 decorated genes, even if the mechanism is not clear yet (Creppe, Palau, Malinverni, Valero, & Buschbeck, 2014). In another report in quiescent lymphocytes PRC1 seems to be inactivated by a mechanism that requires Aurora kinase and indirectely inactivates PRC1 through phosphorylation (Frangini et al., 2013).
PRC1.5 has been also proposed to be involved in gene activation. CDK2 protein in the complex phosphorylates serine 168 of Ring1B affecting its catalytic activity. Moreover Auts2, another subunit from the same complex can recruit the histones acetylase p300 resulting in gene activation (Gao et al., 2014).

3. PRC1 Targeting

Progress toward elucidating the molecular basis of PcG protein complex targeting has come from studies in *Drosophila*, where PcG protein complexes assemble at Polycomb response elements (PREs). PREs typically comprise a few hundred base pairs and contain binding sites for PhoRC subunit Pho, the only PcG protein with sequence-specific DNA-binding activity (Oktaba et al., 2008). The molecular interactions that permit PRC1 or PRC2 to associate with specific genomic locations are currently only poorly understood. Studies in mammalian cells have suggested that targeting of these complexes to DNA may entail different mechanisms.

The YY1 transcription factor is the mammalian homolog of dPHO. Nevertheless, genome wide studies showed a poor overlap between YY1 and PcG proteins ; YY1 DNA binding elements are mostly absent at PcG targets (Mendenhall et al., 2010). Interestingly, other transcription factors have been suggested to mediate the binding of PcG to target loci. The transcription factor REST (also called Neuron-Restrictive Silencing Factor, NRSF) is required for the recruitment of PRC1 and PRC2 to a subset of its target genes in mES cells. Interestingly, the recruitment of Rnf2 to Rest binding sites can occur independently of both CpG islands and PRC2 activity (Dietrich et al., 2012). REST was originally characterized as a repressor of neuronal genes in non-neuronal cells; it can both activate and repress genes containing RE1 elements. REST and Cbx family proteins co-purified in ES cell extracts and formed complexes in cells. REST facilitates PRC1 binding at proximal RE1 elements and inhibites PRC1 binding at distal RE1 elements, suggesting a direct role in PRC1 targeting (Ren & Kerppola, 2011). Additionally in megakaryocytes, the direct physical and functional association between heterodimer Runx1/CBFβ transcription factor and PRC1 has been reported. Runx1 recruits PRC1 directly to chromatin in a PRC2-independent manner (Yu et al., 2012).

Also, long non coding RNAs may take part to PRC1 targeting. CBX7 is specifically associated with ANRIL, a long noncoding antisense RNA transcript overlapping the INK4b/ARF/INK4a locus(Pasmant, Laurendeau, Héron, et al., 2007). In vitro, CBX7 employs overlapping yet distinct regions within its chromodomain for binding to H3K27me3 and RNA. Introduction of antisense transcripts to ANRIL or structure-guided mutations of CBX7 of PRC1 that disrupt H3K27me3 or RNA binding, affects cellular life span. Thus, PRC1 and ARNIL are proposed to integrate the regulation of the INK4b/ARF/INK4a locus (Yap, Li, Muñoz-Cabello, et al., 2010).

4. Other PcG complexes

Analysis of other Drosophila PcG genes revealed other components that do not participate to PRC1 and PRC2 but form distinct complexes (Fig.6).

The Pho repressive complex (PhoRC) contains Pho (DNA-binding protein, homolog of mammalian YY1) and Sfmbt (Sex comb on midleg (Scm)-like with 4 Malignant Brain Tumour domains). This

complex has no enzymatic activity but it recognizes specific histones modifications. In Drosophila Pho is thought to facilitate the recruitment of Polycomb complexes to the PREs (Klymenko et al., 2006). Whether such a complex exists in mammals has still to be clarified. YY1, mammalian Pho homolog, has been suggested to interact with Polycomb complexes, but genomic binding profiles show little overlap between YY1 and PcG proteins binding sites in mammalian genomes (Affar, Gay, Shi, & Liu, 2006; Oktaba et al., 2008; Vella, Barozzi, Cuomo, Bonaldi, & Pasini, 2012).

The Polycomb Repressive Deubiquitinase complex (PR-DUB) includes the Calypso ubiquitin carboxy-terminl hydrolase and Additional sex combs proteins (Asx). The gene encoding the catalytic subunit Calypso was first identified as a PcG gene in a genetic screen. Calypso belongs to the ubiquitin C-terminal hydrolase (UCH) class of deubiquitinases nucleosomes in vitro (Alonso et al., 2007). It was shown that Calypso catalyses H2A deubiquitination and that this enzymatic activity requires the presence of Asx (Scheuermann et al., 2010). This result is quite unexpected, given that PR-DUB deubiquitinase activity antagonizes the H2AK118 ubiquitination activity of PRC1, which is required for HOX gene repression. It has been suggested be that PR-DUB activity is required to release ubiquitin, or H2AK118ub1-binding factors that become sequestered elsewhere in the genome and thereby limit H2AK118 ubiquitination by PRC1 (Gutierrez et al., 2012). The mammalian homologs are BAP1 (BRACA-1 associated protein, Calypso counterpart) that interacts ASXL1/2/3 through its C-terminal portion (ASX homologs). Bap1 protein interacts with additional partners in mammals such as BRCA1, host cell factor-1 (HCF-1), N-acetylglucosamine transferase, the fork- head box transcription factors FOXK1/2, MBD family proteins MBD5/6, transcription factor YY1 and ubiquitin-conjugating enzyme UBE20 (Baymaz et al., 2014; (Jensen et al., 1998); Machida, Machida, Vashisht, Wohlschlegel, & Dutta, 2009; Mashtalir et al., 2014). With the information currently available, it is difficult to define a human PR-DUB complex beyond the core components BAP1 and ASXL1–3; it is likely that BAP1 will form part of a diverse range of complexes, only some of which perform PcG functions.

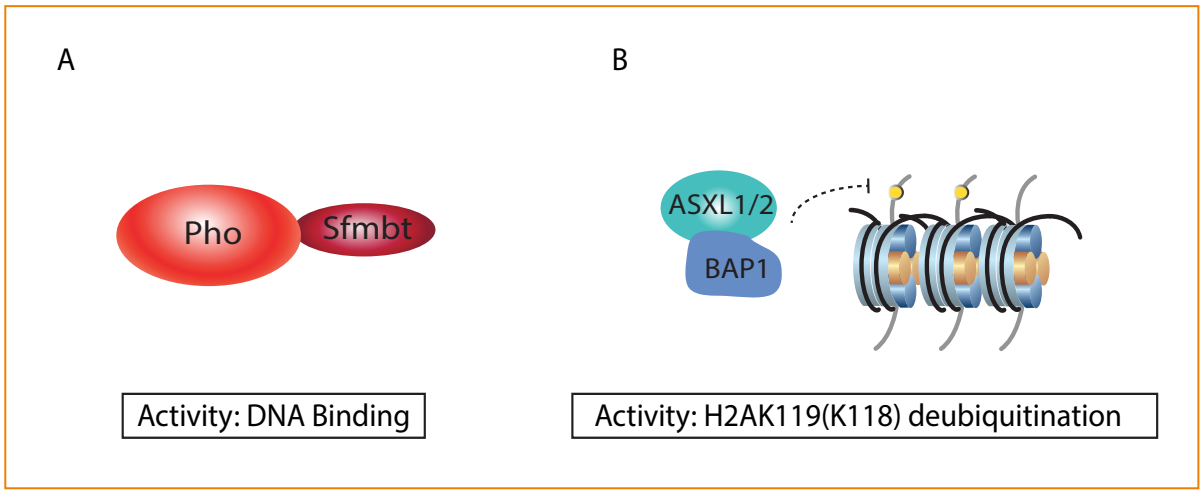


Figure 6. Other PcG complexes

(A) Pho-Rc complex in Drosophila with DNA binding activity, (B) PR-DUB complex in mammals: BAP1 deubiquitinase is the homolog of Calipso, ASXL1 and ASXL2 are the homolog of Asx

CHAPTER III

POLYCOMB REPRESSIVE COMPLEX 2 (PRC2)

In mammals, the core complex is composed by SUZ12 (Suppressor of Zeste 12), which contains a zinc finger domain; EED (embryonic ectoderm development), which contains WD40 repeats able to recognize trimethylated peptides; EZH1/2 (enhancer of zeste 1 or 2 methyltransferase, the catalytic center of the complex) and RbAp46/48, mammalian homologs of Drosophila Nurf55 (Fig.7). PRC2 enzymatic activity plays a crucial role in preserving cellular identity and PRC2 deletions cause embryonic lethality in mice and segmental transformations in Flies (Faust, Lawson, Schork, Thiel, & Magnuson, 1998; J. Ng et al., 2000; O'Carroll et al., 2001; Pasini et al., 2004). In vivo and in vitro, Lysine 27 of histone H3 (H3K27) is an undisputed methylation site for both *Drosophila* and human PRC2. Lysine 27 could be mono-, -di-, -tri-methylated in a stepwise manner by the complex (H3K27me1, H3K27me2 and H3K27me3). Oligonucleosomes were discovered as preferred substrate suggesting that PRC2 activity on a given nucleosome could be influenced by the surrounding nucleosomes (Raphaël Margueron & Reinberg, 2011).

In Drosophila, chromatin immunoprecipitation experiments associated H3K27 methylation to silent genes. Mutations impairing E(Z) activity result in the de-repression of gene silencing (Cao et al., 2002; Müller et al., 2002). Consistently, a point mutation changing lysine 27 of histone H3 into an arginine fails to repress transcription of genes that are normally repressed by PRC2. H3-K27 mutant flies were shown to recapitulate E(z) mutant. Taken together, these analyses further confirm that H3K27 is a crucial PRC2 substrate for Polycomb repression.

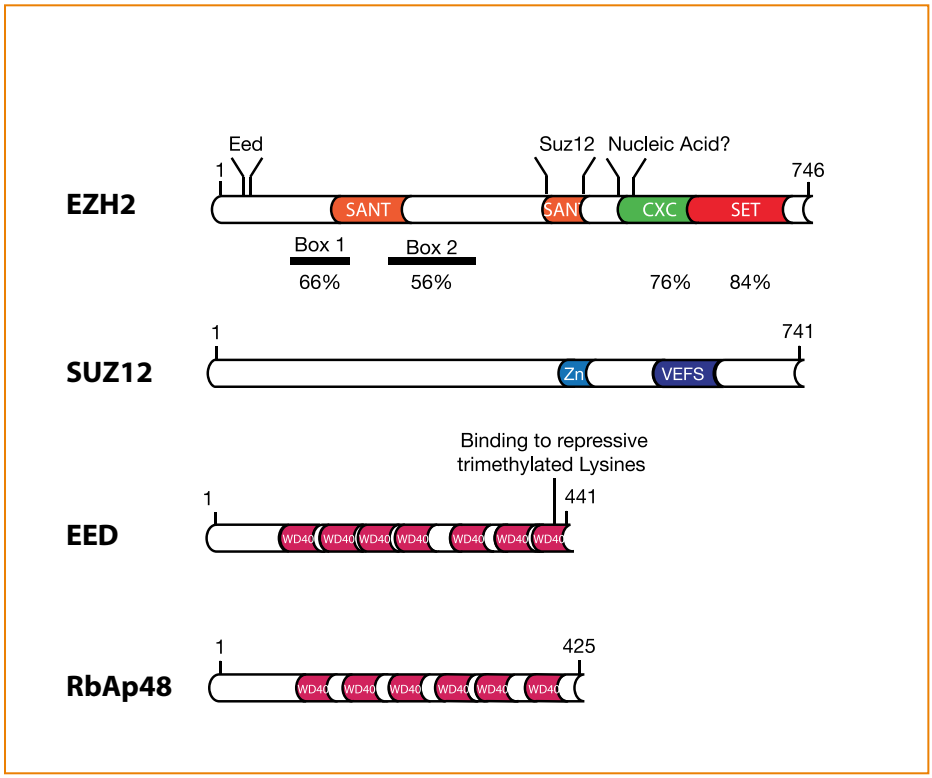


Figure 7. PRC2 core complex domains.

The four PRC2 core complex subunits are showed. The numbers indicate the percent homology between the mouse and *Drosophila* homologs (adapted by (Margueron & Reinberg, 2011)).

1. PRC2 structure

As mentioned above, EZH2 is the catalytic subunit of PRC2; however, it requires the presence of the core subunits EED and SUZ12 for its enzymatic activity. The structure of EZH2 reveals that, in its uncomplexed form, the EZH2 C-terminus folds back into the active site, blocking engagement with the substrate. The *S*-adenosyl-L-methionine (SAM) binding pocket observed in the crystal structure of homologous SET domains forms only partially. Thus, conformational changes in the EZH2 SET domain take place upon complex formation to allow binding of cofactor and substrate (Antonysamy et al, 2013). Recently, basal and activated forms of a ternary EZH2-EED-SUZ12 (VEFS domain) from the yeast *Chaetomium thermophylum* were purified and crystal structures resolved. The Eed subunit is surrounded by a belt-like structural feature of Ezh2, and Suz12 (VEFS) contacts both of these subunits to confer enzyme activity. An important SET activation loop in Ezh2 is stabilized by Eed and Suz12 (VEFS) contacts, and this could explain the indispensable role of these subunits in PRC2 catalysis. In this active state, the SAM binding pocket is accessible, thus allowing the methyltransferase reaction (Jiao et al, 2015).

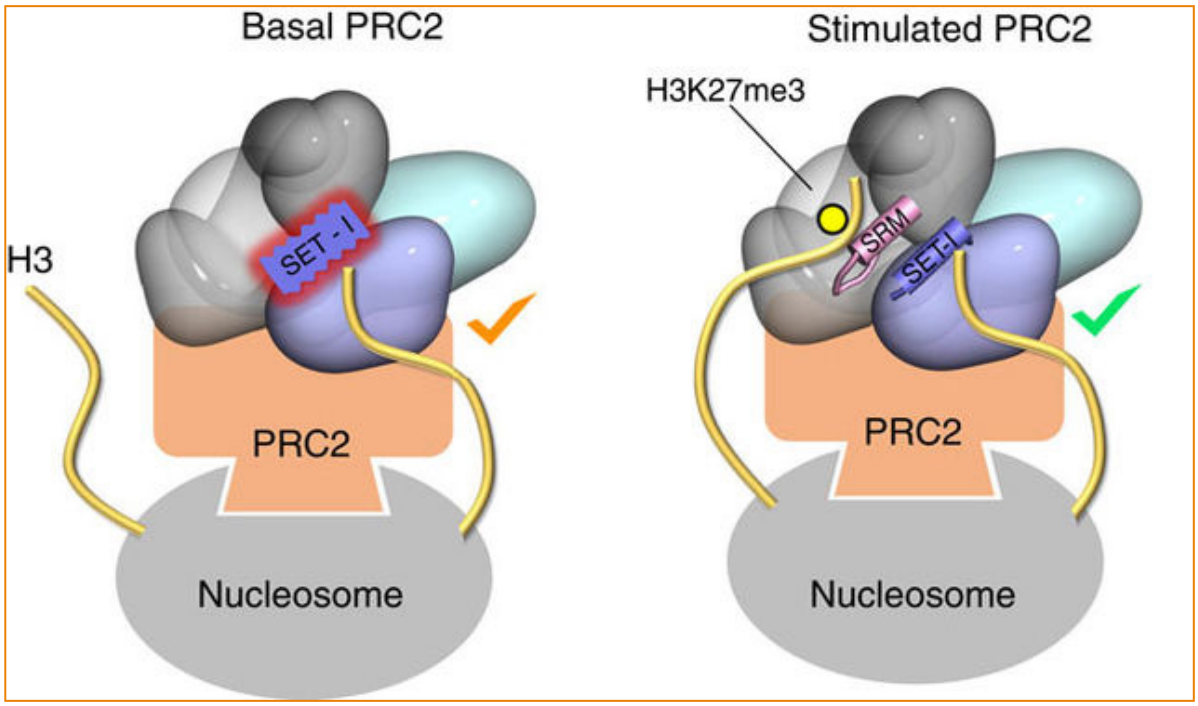


Figure 8. PRC2 architecture in the basal and activated states.

The basal activity of PRC2 towards H3K27 is shown in the left part of the figure in the absence of activating interactions, where the SET-I has more thermal mobility. This is activated by the binding of a 'H3K27me3 peptide' to EED in the right panel, mediated by stabilization of SRM and the SET-I during the propagation of repressive chromatin mark (adapted from Justin et al, 2016).

Importantly, RbAp48 subunit is also necessary for full PRC2 enzymatic activity: in the absence of RbAp48, the PRC2 complex is functional but less efficient. SUZ12 is dispensable for the association of EED with Ezh2, but it is essential for EZH2 HMT activity. SUZ12 is also required for the recruitment of the histone-binding subunit RbAp48 to PRC2 complex. Mass spectrometry analysis, both in mammals and Drosophila, is consistent with binding between a N-terminal peptide of SUZ12 and RbAp48 subunit (Ciferri et al, 2012). Thanks to the resolution of crystal structures, it is now possible to describe how SUZ12 contacts all PRC2 core complex subunits and thereby ensures a competent complex conformation capable of fulfilling its activity.

2. PRC2 regulation by the chromatin environment

The observation that PRC2 exerts higher catalytic activity on nucleosome arrays led to the hypothesis that the complex is able to sense the chromatin environment and to integrate this information to modulate its activity.

A proof of principle is the ability of the complex to recognize its own catalytic product; this further stimulates its catalysis. The catalytic activity of PRC2 on oligonucleosome substrates is increased on

arrays containing H3K27me2/3 but not H3K27me0/1. When H3K27me3 peptide binds to EED aromatic cage, PRC2 catalytic activity is further increased. Within the complex, the EED subunit assumes a donut-like B-propeller structure with a conserved aromatic cage that binds preferentially to H3K27me3 (Margueron et al, 2009). The active enzyme crystal clearly shows that an H3K27me3 peptide lies between the WD40 domain of Eed and a region of Ezh2 called the stimulation-responsive motif (SRM). An allosteric effect elicited by this interaction stabilizes the SRM region and results in increased catalytic activity (Jiao et al, 2015). Notably, the recognition by PRC2 of its own catalytic product takes on biological relevance during DNA replication, when the existing marks on histone tails become diluted two times. After replication, the complex can counteract this dilution by directly instructing the deposition of new methyl groups on histone H3 lysine 27 of newly incorporated nucleosomes, specifically along regions already bearing this mark. Developmental abnormalities of *Drosophila* embryos mutant for *esc* and *escl* (EED homologs in fly) are rescued by the introduction of the wild type form Esc but not by the esc aromatic cage mutant. Upon mutation of the aromatic cage, the levels of H3K27me2/3 and PRC2 binding at its target genes are reduced and global levels of H3K27me3 also diminish (Margueron et al, 2009).

PRC2 is also able to sense nucleosome density. This was illustrated by a systematic study in which reconstituted *Drosophila* PRC2 complex was incubated with different substrates in vitro, and which showed that nucleosome spacing modulates PRC2 activity. The effect was proposed to depend on an N-terminal region of H3 (from 31 to 42 amino acid residue), which binds to VEFS domain of SUZ12 subunit, resulting into an increase of PRC2 catalytic activity (Yuan et al, 2012). This result was interpreted to mean that chromatin compaction acts upstream of PRC2, whereas it had often been considered a consequence of PcG action. This mechanism might thus reinforce chromatin silencing after it has already been initiated.

Histone H3 can be post-translationally modified in at least 15 of the 40 residues providing cross-talk regulation (Bhaumik et al, 2007). PRC2 activity is negatively regulated by histone modifications associated with transcriptionally active chromatin. For instance, the H3K4me3 mark is associated with transcription at active gene promoters, and PRC2 catalysis is strongly reduced when H3K4me3 is present on the same histone tail both in Drosophila and in mammals. H3K4me3-mediated inhibition of PRC2 is consistent with the report that actively transcribed genes are resistant to H3K27me3 deposition (Schtmiges FW et al, 2011). Similarly, H3K36me2 and H3K36me3 also inhibit PRC2 activity on the same histone tail. These two marks are associated with active gene bodies and their inhibitory influence further supports the idea that expressed loci may limit the propagation of H3K27me3. A mass spectrometry analysis of native chromatin demonstrated that H3K4me3 or H3K36me3 are indeed mutually exclusive with H3K27me3 on the same H3 tail, yet can be found along with H3K27me3 on opposite H3 tails of the same mononucleosome. Consistently, Polycomb

Repressive Complex 2-mediated methylation of H3K27 is not inhibited when H3K4me3 or H3K36me3 lie asymmetrically on sister histone within the same nucleosome. These results uncover a potential mechanism for the incorporation of bivalent features into nucleosomes, which have been proposed to be important in cell plasticity and pluripotency (Voigt et al, 2012).

Lysine acetylation is chemically incompatible with methylation on the same residue. Indeed, in contrast to H3K27me3, H3K27 acetylation (H3K27Ac) is specifically found at active enhancers and promoters (reviewed in Heintzman and Ren, 2009). In the absence of PRC2, a global increase of acetylation is observed, suggesting some level of competition between the two marks (Tie F et al, 2009; Pasini et al, 2010). Supporting this idea, the H3K27 deacetylase NuRD seems to cooperate with PRC2 in gene silencing during mES differentiation (Kim TW et al, 2015).

H3K27 is immediately flanked by serine-28 (S28), which is phosphorylated during mitosis as well as during interphase (Dunn and Davie 2005; Dyson et al. 2005). Histone phosphorylation is mediated by external signals through transduction pathways inside the cell and is required in cellular processes such as chromosome condensation during mitosis and transcriptional regulation. H3S28 phosphorylation on promoters previously silenced by Polycomb is mediated by histone kinases MSK1 and MSK2, which are activated by different signal transductions pathways upon extracellular stimulation. The presence of H3S28 phosphorylation (H3S28p) antagonizes PRC2 activity. Indeed, the phosphate group on S28 could either prevent H3K27 recognition, or inhibit the methylation reaction by steric hindrance, or finally destabilize PRC1 and PRC2 binding to chromatin. Importantly, the H3K27me3-S28p switch might allow rapid activation and resilencing of genes by controlling PRC1 and PRC2 binding without the need for histone demethylation (Gehani SS et al, 2010; Lau PN et al, 2011; Stojic L et al, 2011).

In vertebrates, cytosine residues at CpG dinucleotides are extensively methylated, with the exception of CpG islands (CGIs), which are found in the promoter regions of more than 50% of genes. DNA methylation patterns are erased in the early embryo and developing germ cells, and then re-established by the de novo DNA methyltransferases Dnmt3a and Dnmt3b, and the accessory protein Dnmt3L (Edwards, Yarychkivska, Boulard, & Bestor, 2017). Propagation of DNA methylation patterns through DNA replication is dependent on the maintenance DNA methyltransferase, Dnmt1, and the accessory protein UHRF1. DNA methylation is necessary for normal development, but not for the maintenance of mES cultures (Lei et al, 1996).

CpG islands in which cytosine residues are methylated tend to exclude PRC2 and in mES cells disruption of DNA methylation is associated with de novo H3K27me3 deposition to previously methylated CpG islands (Wu, 2010). DNA methylation seems to antagonize PRC2 for chromatin binding rather than inhibiting its catalytic activity: in nucleosomes pull down experiments, PRC2 was

poorly recovered in presence of DNA methylation both using recombinant PRC2 (H. Wu et al., 2010) and nuclear extracts (Bartke et al, 2010). Although DNA methylation seems to be refractory to PRC2 *in vivo*, DNA methylation and H327me3 mark are nonetheless both present on the inactive X chromosome (reviewed Galupa and Heard 2015) and *in vivo* tethering of both PRC1 and PRC2 complexes to methylated DNA regions does not directly inhibit their catalytic activity (Cooper S et al, 2014). This suggests that perhaps PRC2 complex might be inhibited by methylated DNA in a context-dependent manner.

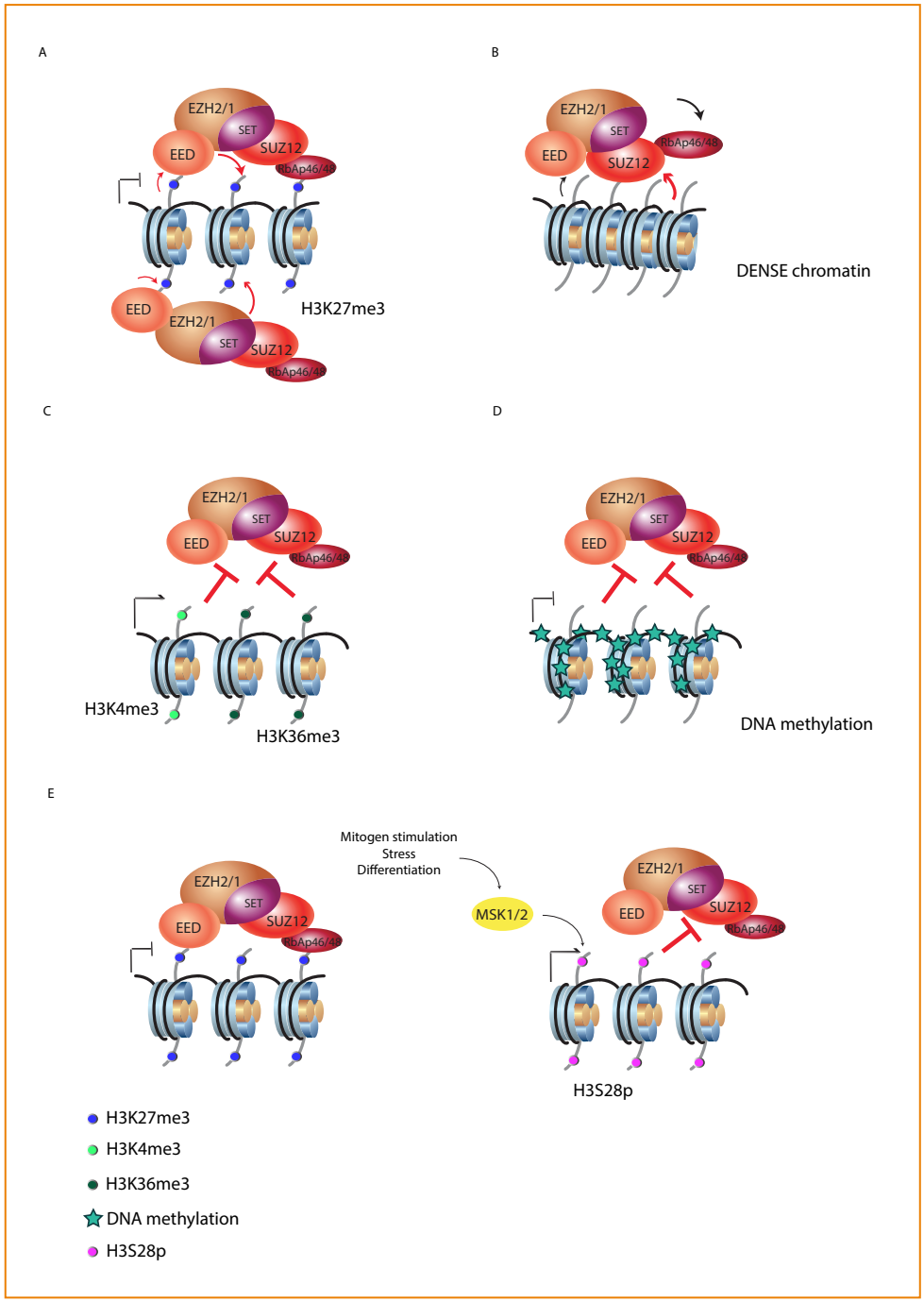


Figure 9. PRC2 activity regulation by chromatin environment

(A) Propagation of H3K27 methylation. Embryonic ectoderm development (EED) specifically binds to H3K27me2/3, and this binding allosterically activates PRC2, leading to propagation of H3K27me2/3 along the chromatin. (B) Dense chromatin activates PRC2 and facilitates the de novo establishment of H3K27 methylation. Upon cessation of gene transcription, chromatin remodelers begin to compact chromatin. PRC2 could be activated by densely packed nucleosomes, which facilitate the spreading of H3K27 methylation along the gene body. The mechanism of stimulation by dense chromatin involves the sensing of H3 from residue 31 to 42 by SUZ12. (C) Actively transcribed genes harbor H3K4me3 and H3K36me2/3 markers, both of which strongly inhibit PRC2 activity through the VEFS domain of SUZ12. This mechanism is believed to prevent the targeting of PRC2 to active genes. (D) DNA methylation seems to antagonize PRC2 for chromatin binding rather than inhibiting its catalytic activity. (E) Under certain conditions, the ERK and p38 kinase pathways will activate MSK1/2, which then phosphorylate H3S28 near the promoters of their target genes, generating an H3K27me3-S28p double marker on genes that were previously repressed by Polycomb group proteins (PcG) proteins. This double marker evicts PRC1 and PRC2, leading to open chromatin suitable for transcription. Upon release of the signal, phosphorylation is removed, leaving H3K27me3, which is again bound by PcG proteins.

3. PRC1 and PRC2 are intertwined

Genetic analyses have demonstrated that PcG target loci are often coregulated by PRC1 and PRC2, and consistently genome mapping studies in Drosophila and mouse demonstrate co-occupancy of PRC1 and PRC2 at many target loci (Boyer et al., 2006; Ku et al., 2008; Schwartz et al., 2006).

According to the conventional paradigm of PcG-mediated maintenance of gene repression both in *Drosophila* and mammals, the main biological function attributed to PRC2 and its H3K27me3 associated mark is to target PRC1 complex to chromatin, via the high affinity of the Pc/CBX chromodomain for H3K27me3 (Cao et al., 2002; Fischle et al., 2003; Min et al., 2003; Blackledge et al, 2014). Subsequently, PRC1 is thought to drive transcriptional silencing either through chromatin compaction (Francis 2001), mono-ubiquitination of H2A on lysine 119 or by interfering with transcriptional elongation (De Napoles et al, 2004; Wang H et al, 2004).

However, recent findings challenge this longstanding model. First, non-canonical PRC1 complexes have been isolated that contain no chromodomain subunits and are characterized by the distinctive presence of RING1A/B together with the RYBP protein that is required for normal development. Moreover, deletion of PRC2 components in mES does not drastically impair PRC1 binding to its target sites (Tavares et al, 2012), suggesting that PRC1 recruitment is not exclusively dependent on H3K27me3 and that PRC1 also has a role in initiating Polycomb domain formation. A further twist comes from evidence that PRC2 recruitment in mES is dependent on a functional PRC1 and that artificial recruitment of PRC1 to genomic loci devoid of PcG leads to PRC2 recruitment and H3K27me3 deposition (Blackledge et al, 2014; Cooper et al, 2014). PRC1 monoubiquitination activity

on lysine 119 of histone H2A is necessary to mediate this effect. Indeed, artificial tethering of EED to chromatin brings chromodomain-containing PRC1 complex without promoting H2AK119ub mark and the subsequent potentiation of PRC2-mediated repression (Blackledge et al, 2014). Endogenously, the variant PCGF1/PRC1 complex seems to be targeted to CpG islands by KDM2B and to be required for normal Polycomb domain formation and mouse development (Blackledge 2014). Affinity purification followed by mass spectrometry analysis revealed that PRC2 subunits are among the most highly enriched H2Aub interactors, with PRC2 interactors Jarid2 and Aebp2 also strongly enriched both from *Drosophila* and mES cell extracts (Kalb R et al, 2014). Recently, a ubiquitin interaction motif at the amino-terminus of Jarid2 was proposed to facilitate PRC2 localization to H2AK119ub both *in vivo* and *in vitro*. This attributed a critical function to Jarid2, defining a key mechanism that links PRC1 and PRC2 in the establishment of Polycomb domains (Cooper et al, 2016). H2AK119ub-mediated activation of PRC2 enzymatic activity *in vitro* is higher in the presence of AEBP2 than in the presence of JARID2 (Kalb et al, 2014). Yet, *in vivo* AEBP2 seems not to be required for H3K27me3 deposition at artificially generated H2Aub sites (Cooper et al, 2016), raising the hypothesis that AEBP2 may only be required in particular contexts.

These findings change the previous hierarchical model, but several layers of positive feedback loops could enable PRC1- and PRC2-mediated repression to reinforce one another synergistically. In a revised model, PRC1 is first recruited to CpG islands and deposits the H2AK119ub mark, which in turn favors PRC2 recruitment. H3K27me3 deposited by PRC2 might then in turn recruit canonical PRC1, promoting chromatin compaction. As an important qualification to this model, it was shown in the past decade that catalytically dead mutants of Ring1B are largely able to rescue the Ring1B-null phenotype both in Drosophila and mammals (Illingworth et al, 2015; Pengelly et al, 2015), suggesting that the critical biological role of PRC1 is separate from its ability to recruit PRC2. Furthermore, loss of H2AK119ub has no effect on H3K27me3 genome-wide enrichment, nor do mutations on monoubiquitinatinated residue on H2A trigger major developmental defects either in mouse or in fly. PRC1 thus clearly plays additional roles apart from its H2A monoubiquitination activity. Mechanistically, recent reports suggest that PRC1 subunit PHC2 is able to self-polymerize and that CBX2 can promote chromatin compaction (Isono et al, 2013; Lau et al, 2017).

Physical interactions between the two complexes have also been proposed. In *Drosophila*, a time course experiment during early development demonstrated a temporary interaction between Esc and Pc proteins at the 2-hour stage, suggesting a synergistic effect on gene repression (Poux S et al, 2001). In *Drosophila*, Sex comb on midleg (Scm) has been proposed as a functional link between PcG-repressive complexes according to cross-linked affinity purifications of BioTAP-Polycomb Pc or E(Z). These results were corroborated by *in vitro* binding and localization studies (Kang et al, 2015). Interestingly, a recent report showed that the PRC2 subunit EED is assembled into PRC1 complexes

in a prostate cancer line. It is proposed to physically interact with the PRC1, which competitively inhibits EED-EZH2 binding and PRC2 catalytic activity *in vitro*. If EED is added to H3K27me3 modified nucleosomes, it stimulates the H2A monoubiquitination activity of PRC1 (Cao Q 2014).

It is still challenging to define a comprehensive genome-wide picture of the reciprocal dynamics between PRC1 and PRC2, and it cannot be excluded that their contribution might vary considerably depending either on the cell type or on the model organism.

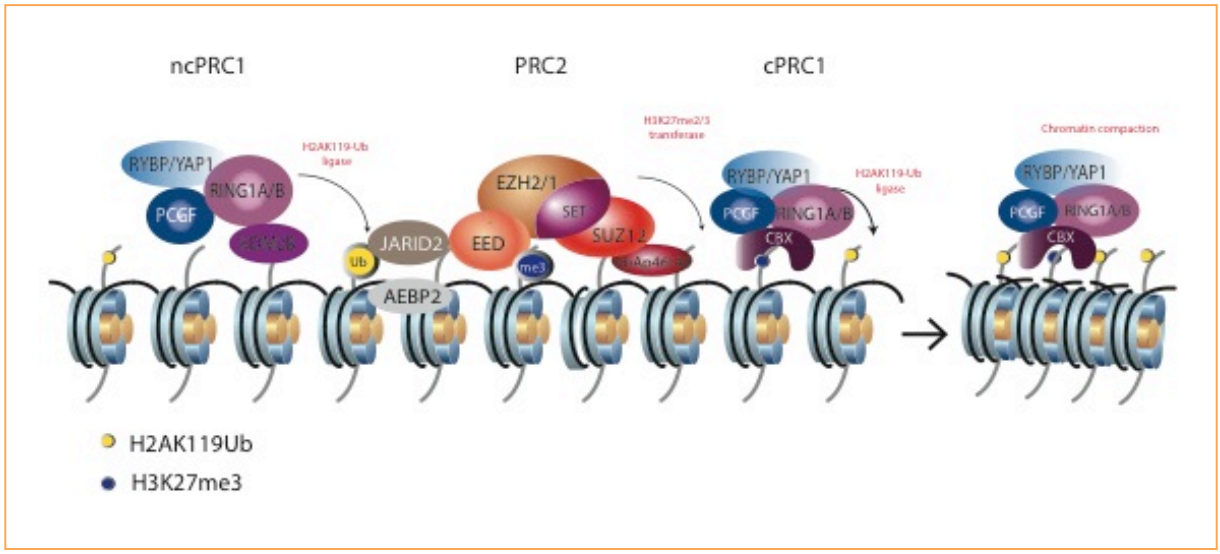


Figure 10 The Polycomb Repressive Complexes Interplay.

PRC1 repressive function has been linked on the one hand to its ability to promote chromatin compaction and on the other hand to its ability to catalyze the formation of histone H2A ubiquinated on lysine 119. Canonical PRC1 complexes contain a CBX subunit that binds with high affinity to H3K27me3 and this gave rise to a hierarchical recruitment model in which PRC2-mediated deposition of H3K27me3 serves to promote PRC1occupancy and transcriptional repression(right). Conversely, recent findings point to a role for H2AK119ub as an efficient recruiter of PRC2 complexes containing AEBP2 and JARID2.

4. PRC2 targeting

PRC2 targeting was first studied on Drosophila HOX clusters. PRE sequences can be located many kilobases away from the promoter they control, they are devoid of nucleosomes and can bind combinations of transcription factors that in turn recruit PRC1 and PRC2. One of the most characterized PcG recruiter is Pho protein (reviewed in (Ringrose & Paro, 2007)). In mammals, just few studies have tried to describe the presence of functional PRE (Sing et al., 2009; Woo, Kharchenko, Daheron, Park, & Kingston, 2010). The current view is that PRC2 recruitment might follow distinct rule in mammals.

4.1 Transcription Factors

Instructive targeting mechanism describes specific transcription factors or non-coding RNA as particular entities involved in PRC2 targeting to its loci (Klose, Cooper, Farcas, Blackledge, & Brockdorff, 2013). Several cases of transcription factors that participate in PRC2 recruitment have been reported. In one example, the transcriptional factor Snail1 is a repressor of E-cadherin (CDH1) gene expression essential for triggering epithelial-mesenchymal transition. Snail1 recruits PRC2 to the CDH1 promoter and requires the activity of this complex to repress E-cadherin expression. mES cells null for Suz12 show higher levels of Cdh1 mRNA than control cells. Moreover, Suz12 and Ezh2 interact directly with Snal1 and H3K27me3 mark decorates Cdh1 promoter (Herranz et al., 2008). Genome-wide analysis by ChIP-sequencing and biochemical interaction analysis of transcription factor Rest and PcG complexes, suggest that Rest is required for the recruitment of PRC1 and PRC2 to a subset of its target genes in mES cells (Dietrich et al., 2012). To move towards a global comprehension of PRC2 recruitment and transcription factor biding, Arnold et colleagues tried to define the relationship between genome-wide dynamics of H3K27me3 marks and the local occurrence of predicted TF binding sites during neuronal differentiation of murine stem cells. In absence of REST, a loss of H3K27me3 is observed at target promoters in trans, specifically at the neuronal progenitor state. Moreover, promoter fragments containing REST or SNAIL binding sites are sufficient to recruit H3K27me3 in *cis*, while deletion of these sites results in loss of H3K27me3. This studied aimed to give a more systematic overview of transcription factor binding and chromatin regulation at genome wide level.

In another recent report, new PRC2 partnerships have been detected through proteomic approaches in relation to G9A-GLP histone methyltransferases. Zinc finger protein ZNF518B directly interacts with EZH2 and G9A mediating their association and regulation, but it is unlikely that this protein is broadly involved in PRC2 targeting (Maier et al., 2015).

4.2 lncRNAs and PRC2 recruitment

While still a matter of debate, lncRNAs have been proposed to be important for PRC2 recruitment. This topic will be further discussed later on (Results section Chapter II).

4.3 Genome wide targeting

While instructive targeting of PRC2 might be important at some specific loci, it is considered that genome wide targeting might rely more on chromatin features. This is referred as a responsive model for PRC2 targeting (Klose et al., 2013). In mammals, gene promoters are encompassed within regions

of the genome called CpG islands that have an elevated level of non-methylated CpG dinucleotides. Specific histone modifying enzymes are recruited directly to CpG islands through recognition of non-methylated CpG dinucleotide sequence. These enzymes then impose unique chromatin architecture on CpG islands that distinguish them from the surrounding genome.

In a pioneer experiment, a series of engineered bacterial artificial chromosomes were inserted into ES cells genome and it was observed that a long region in Zfpm2 locus initiates de novo recruitment of PRC2, particularly a CpG island within this locus as both necessary and sufficient for PRC2 recruitment. Following this line of evidence, a constitutively active CpG island is able to recruit PRC2 after removing activating motifs contained in it and sequences from the *Escherichia Coli* genome with GC-contents comparable to a mammalian CpG island are capable of recruiting PRC2 when integrated into the ES cell genome. Those findings imply a causal role for GC-rich sequences, depleted of activating motifs, in PRC2 recruitment in mammalian genomes (Mendenhall et al., 2010).

Similarly, using iterative genome editing, functional Polycomb recruiter sequences have been described as 220 nucleotides invariably CpG-rich, they require protection against DNA methylation and their activity can be blocked by placement of an active promoter-enhancer pair in cis. These data support the model whereby PRC2 recruitment at specific targets in mammals is positively regulated by local CpG density yet obstructed by transcriptional activity or DNA methylation (Jermann, Hoerner, Burger, & Schübeler, 2014a). As previously mentioned, protein lysine specific demethylase 2B (KDM2B) bears a zinc finger CXXC DNA binding domain able to recognize not methylated CG islands and it takes part to some PRC1 complexes. This could explain a role for KDM2B in PRC1 targeting, but PRC1 repression occurs independently of KDM2B further evoking other mechanisms (Farcas et al., 2012; J. He et al., 2013). Interestingly, Jumonji AT-rich Interaction Domain (JARID2) is a PRC2 cofactor required to enhance its enzymatic activity with DNA binding activity. JARID2 preferentially recognizes GC-rich region and it could provide an alternative hypothesis for PRC2 recruitment to CpC rich regions (G. Li et al., 2010).

These two models in PRC2 targeting are not mutually exclusive. While the responsive model aims to define more genome-wide features peculiar for the recruitment of the complex, peculiar entities such as transcription factors or non-coding RNA could be necessary at defined loci in particular processes.

CHAPTER IV

PRC2 COFACTORS

One of the main opened questions about PRC2 activity is how does an ubiquitously expressed enzyme within its well conserved core complex maintain distinct patterns of gene repression required in different cell types?

As previously shown, PRC2 is sensitive to chromatin context, transcriptional status and it preferentially binds to for CG rich inactive promoters. Nevertheless, the basis of its cell-type target specificity is still not entirely understood.

In addition to its core complex, PRC2 can associate to facultative subunits in a substoichiometric manner. These partners are able to modulate PRC2 catalytic activity, to favor its recruitment to its targets or both (Holoch & Margueron, 2017a; Vizan, Beringer, Ballare, & Di Croce, 2015).

Most of them have been largely studied individually, while recently comprehensive mass spectrometry analysis showed that their association with the core complex can happen in particular and competitive combination (Hauri et al., 2016; Fig. 11). This lead to the hypothesis that two subtypes of PRC2 complexes exist with potentially different function.

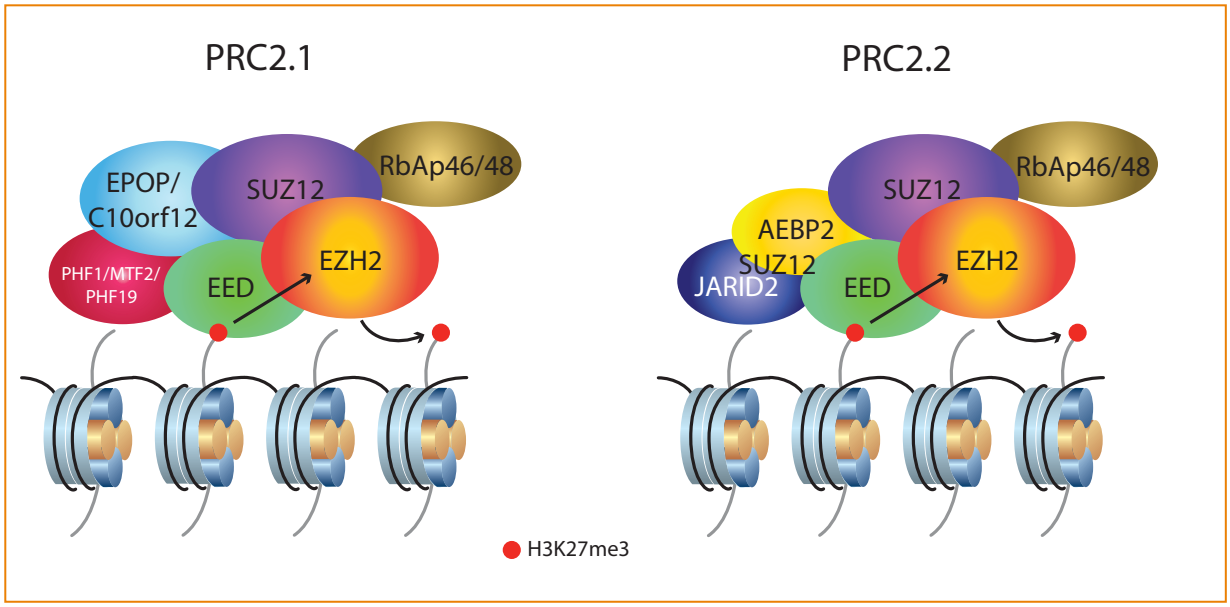


Figure 11. PRC2 core complex associates with its facultative subunits in two main subcomplexes PRC2.1 and PRC2.2.

In addition to the core PRC2 members (EZH2, EED, SUZ12, and RbAp46/48), PRC2.1 may contain one of the Polycomb-like (PCL) homologs (PHF1, MTF2, or PHF19) and either of the two proteins EPOP or C10orf12. By contrast, PRC2.2 consists of the same core PRC2 subunits in association with AEBP2 and JARID2 (adapted by Holoch and Margueron).

1. PRC2.1

One of the Polycomb-like protein homologs (PHF1, MTF2 or PHF19) and the new recently discovered C17orf96 and C10orf12 take part to this subtype of complex together with PRC2 core subunits. These components associate with different stoichiometries (Smits, Jansen, Poser, Hyman, & Vermeulen, 2013a). Their association with the core complex could also change during mES differentiation into neuronal progenitor cells: MTF2 and C17orf96 binding decreases upon differentiation, while PHF19 and C10orf12 become more enriched (Kloet et al., 2016a). Yet, what is the exact subunits contribution to PRC2 activity in different contexts has to be determined.

1.1 Polycomb Like Proteins (PCLs)

Polycomb Like gene (Pcl) was first reported in Drosophila and it was associated to E(Z) recruitment to its loci (McKenzie Duncan, 1982; Savla, Benes, Zhang, & Jones, 2008). Pcl seems necessary for H3K27me3 catalysis at PRC2 loci (Nekrasov et al., 2007). In mammals, Pcl has three homologs: PCL1/PHF1, PCL2/MTF2 and PCL3/PHF19. While Drosophila homolog bears two PHD domains and one Tudor domain without the ability to bind trimethylated lysines, all the three mammalian homologs have a functional Tudor domain together with 2 PHD fingers domain (Fig. 12). For that reason, it has been speculated their possible role in PRC2 targeting to chromatin.

1.1.1 PCL1 alias PHF1

In HeLa cells, PHF1 does not seem to be required for PRC2 occupancy nor its presence at chromatin to be dependent on PRC2. Conversely, PHF1 efficiently enhances H3K27me2 conversion into H3K27me3 that appears essential for gene silencing (Sarma, Margueron, Ivanov, Pirrotta, & Reinberg, 2008). This is in line with results obtained in Drosophila (Nekrasov et al., 2007). Contrasting hypotheses have risen from PHF1 Tudor domain. In one report, Tudor domain can bind H3K36me3, inhibiting H3K27 trimethylation in 293THEK cells and in vitro using yeast chromatin (Musselman et al., 2012). Another observation reports that PHF1 tudor domain binds H3 testis variant in vitro and colocalizes with H3K27me3 and not H3K36me3 in vivo (Kycia et al., 2014). Further analysis defining the precise context in which PHF1 overlaps with H3K36me3 should be performed. Interestingly, a recent report suggests that PHF1 plays a chromatin-independent non-redundant role through the binding to the C-terminal domain of p53 promoting cellular quiescence (Gerard L Brien et al., 2015).

1.1.2 PCL2 alias MTF2

While Phf1 does not appear to be expressed in ES cells, MTF2 has been mostly studied in this model. Mtf2 knockdown apparently increases the global levels of EZH2 and H3K27me3. However, when specific genes are studied, the scenario is different: upon MTF2 depletion, direct MTF2 targets reduce PRC2 and H3K27me3 occupancy. Upon MTF2 deletion, ES cells undifferentiated state is maintained accompanied by failure to differentiate (Walker et al., 2010). Another report observed that PRC2 members are clearly reduced at gene targets upon MTF2 depletion, even if H3K27me3 changes remain unclear. They additionally see that MTF2 mediates PRC2 recruitment on the inactive X chromosome although MTF2 presence is not required for X inactivation establishment (Casanova et al., 2011). This is in line with recent findings that support that PRC2 is not necessary for that process (McHugh et al., 2015). Importantly Tudor domain and single PHD mutants do not display a lethal phenotype as it is with the whole MTF2 locus deleted (S. Wang et al., 2007). A controversial role of this protein has been detected in mouse embryonic fibroblast where the protein seems to activate the senescence through the Cdk2na locus (X. Li et al., 2011). MTF2 role in histone recognition and in PRC2 regulation still remains elusive.

1.1.3 PCL3 alias PHF19

PHF19 is last member of the family identified in mammals. Two different isoforms, one long and one short bearing only Tudor domain with one PHD have been reported to be involved in transcriptional repression (Boulay, Rosnoblet, Guérardel, Angrand, & Leprince, 2011). PHF19 was object of extensive research enlightening its importance in the context of PRC2 functions (Ballaré et al., 2012; G L Brien et al., 2012; Hunkapiller et al., 2012). PHF19 has been proposed to take part to a PRC2

complex without JARID2 and most of its binding sites co-localize with PRC2 at genome-wide level. H3K27me3 and SUZ12 occupancy are reduced in absence of PHF19 and PRC2 knockout also results in Phf19 binding reduction in a interdependent manner (Ballaré et al., 2012). In mouse embryonic stem cells, PHF19 overexpression result in increased self-renewal capacity measured by colony formation assay (Hunkapiller et al., 2012) and PH19 depletion increases differentiation with a consequent drop in pluripotency associated markers (Ballaré et al., 2012; G L Brien et al., 2012; Hunkapiller et al., 2012). Moreover two groups reported defects in embryoid body formation (Ballaré et al., 2012; G L Brien et al., 2012), while the third one reported no differences (Hunkapiller et al., 2012). Interestingly, a complete Tudor domain is able to bind H3K36me3 and it is required for H3K27me3 deposition (Ballaré et al., 2012; G L Brien et al., 2012; Cai et al., 2013). As previously described, H3K36me3 mark inhibits PRC2 deposition on the same histone tail (Schmitges et al., 2011). In vivo colocalization of PHF19 at H3K36me3 loci is not abundant (Ballaré et al., 2012; Cai et al., 2013). An artificial tethering system in HEK293T cells promotes transcriptional repression with PRC2 and also H3K36me3 demethylase recruitment upon PHF19 induction (G L Brien et al., 2012). Conversely, another group noticed colocalization of PHF19 with KDM2B demethylase at target gene (Ballaré et al., 2012), suggesting redundancy mechanism taking place perhaps depending on the model. H3K36me3-binding pocket may be very important for PRC2 activity since only a functional Tudor domain is able to rescue PRC2 occupancy upon PHF19 loss (Cai et al., 2013). Taken together those data underline PHF19 importance in recognizing H3K36me3 mark and mediating the recruitment of PRC2 at these loci taking advantage of H3K36 demethylases (reviewed in (Holoch & Margueron, 2017b; Vizan et al., 2015)). Consistent with this model, PHF1 overexpression has been proposed to bring H3K27me3 at H3K36me3 decorated genes in a Tudor domain dependent manner (Cai et al., 2013).

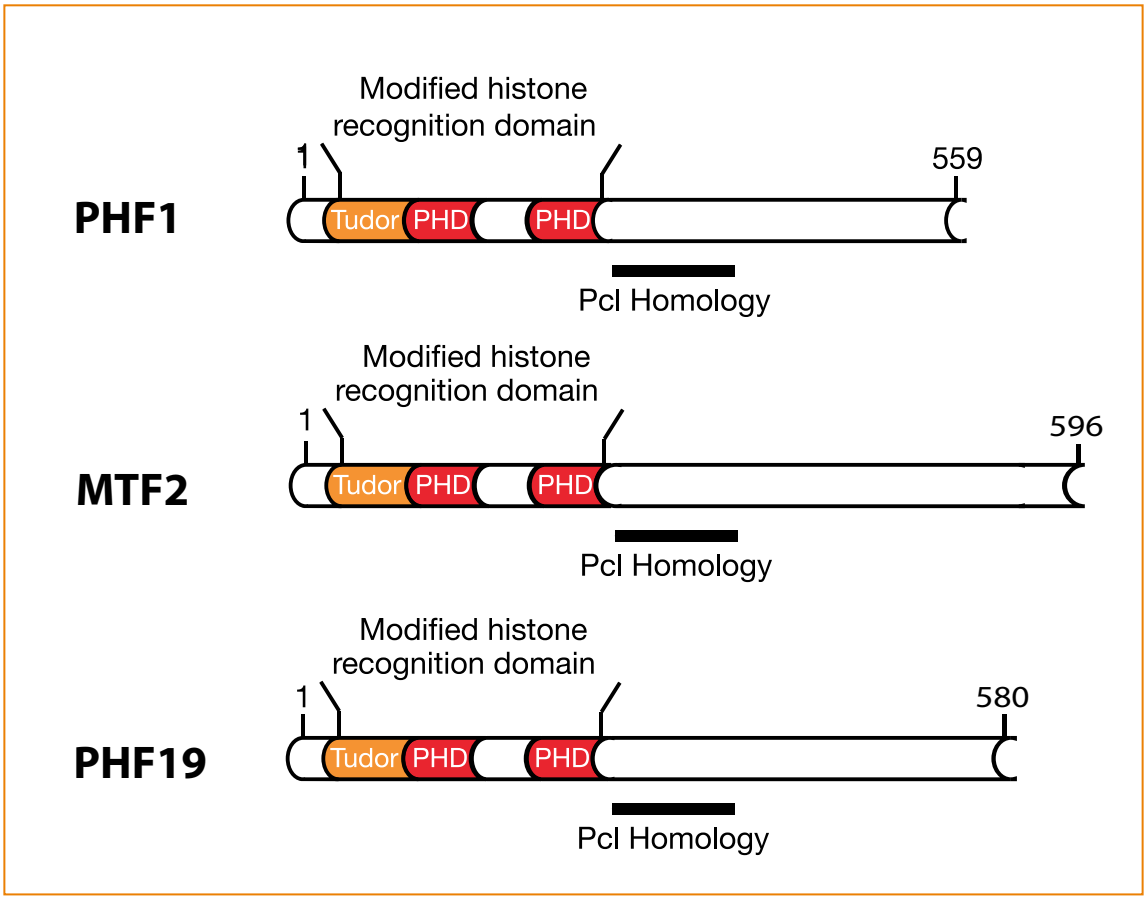


Figure 12. Mammalian Pcl homologs and their conserved domain

The mammalian homologs bear a functional Tudor domain and 2 PHD fingers domain. Thus, they ahve been proposed to target PRC2 to chromatin. Evolutionary conservation is underscored by a black line below (adapted by (Margueron & Reinberg, 2011)).

1.2 C17ORF96 alias EPOP

C17orf96 or EPOP (Elongin BC and Polycomb Repressive Complex 2 Associated Protein) is a recently discovered PRC2-associated protein that is highly expressed in mouse ES cells and in the two human-derived cell lines HeLa and HEK293 (Fig. 13, upper part). Previous efforts to identify PRC2-associated factors overlooked EPOP, presumably for several reasons: the protein was poorly annotated and does not contain any chromatin-related domains, it apparently does not have a homolog in Drosophila, and its mRNA expression during mammalian development or in adult tissue is largely unknown because of missing probes in microarrays. The first hint of its connection to PcG-mediated silencing came from a study in which it was noticed among HeLa nuclear proteins exhibiting affinity for H3K27me3-bearing nucleosomes (Bartke et al, 2010). In mouse embryonic stem cells EPOP physical association with PRC2 has been confirmed. Biochemical studies pulled down EPOP with PRC2 core components SUZ12 and EZH1/2 and with MTF2 and JARID2 cofactors (Zhang et al., 2011). Nevertheless other groups proposed a mutually exclusive binding between EPOP and JARID2 (Alekseyenko et al, 2014; Hauri et al, 2016). In vitro, C17orf96 appears to enhance the catalytic

activity of PRC2 and interestingly, this effect is additive with that of JARID2, suggesting that C17orf96 operates by a distinct and nonredundant mechanism. In HEK293T cells EPOP binding sites mostly overlap to EZH2 and H3K27me3 ones (Alekseyenko, Gorchakov, Kharchenko, & Kuroda, 2014a). Contrarily to what expected from in vitro data, C17orf96 knockdown in mouse ES cells leads to increased SUZ12 occupancy and increased levels of H3K27me3 at PRC2 targets genome-wide (Liefke & Shi, 2015). Yet, C17orf96 associates with H3K4me3 decorated genes. Very recent reports showed that EPOP interacts with Elongin BC (ELOBC) heterodimer in mouse embryonic stem cells. ELOBC is present on bivalent promoters and it is involved in transcriptional elongation complex (Beringer et al., 2016). Upon EPOP ablation, H3K4me3 levels are affected and among EPOP main partners, Elongin BC may cooperate with the deubiquitinase USP7 to alter H2B ubiquitination levels at promoters (Liefke, Karwacki-Neisius, & Shi, 2017). EPOP N-terminal region contains a BC box required for EloBC interaction (Beringer et al., 2016; Liefke et al., 2017), whereas C-terminal part is required to interact with PRC2 (Liefke & Shi, 2015). However immunopurification of EPOP on a glycerol gradient observed that EloBC is not in the same fraction as PRC2, discouraging a possible interaction. EPOP knockdown results in a decreased EloBC occupancy at PRC2 bound loci genome-wide and consistently, EPOP or ELOBC knockdown increase PRC2 occupancy while lowering residual gene expression (Beringer et al., 2016; Liefke et al., 2017). EPOP has been proposed to maintain low levels of expression at PRC2 genomic targets, perhaps contributing to maintenance of cellular plasticity (Beringer et al., 2016). Exogenous expression of Epop is found to enhance transcription factor-induced reprogramming of fibroblasts into pluripotent stem cells (Zhang et al., 2011) and its expression has been reported to decline sharply upon ES cell differentiation (De Cegli et al., 2013; Liefke & Shi, 2015).

These results suggest that despite their matching chromatin association profiles, EPOP may act to limit the PRC2 binding to target genes. Understanding how this relates to the cofactor's positive regulation of PRC2 enzymatic activity and which is EPOP impact during embryogenesis and development, will be important questions for future analysis.

1.3 C10ORF12

Another recently uncovered PRC2-interacting protein is C10orf12 (Fig.13, bottom part). It is present at a much lower stoichiometry in PRC2 purifications than C17orf96/EPOP and other known cofactors (Smits, Jansen, Poser, Hyman, & Vermeulen, 2013b), and it is found at fewer than 50% of genomic EZH2 binding sites in HEK293 cells, compared to 80% for C17orf96 . C10orf12 presence in the complex excludes C17orf96 (Alekseyenko et al., 2014; Hauri et al., 2016). C10orf12 artificially tethered to a transgene is able to repress it bringing PRC2 and the associated mark H3K27me3. C12ORF12 has been found also fused to LCOR, a ligand dependent corepressor in hormone-dependent nuclear signalling.

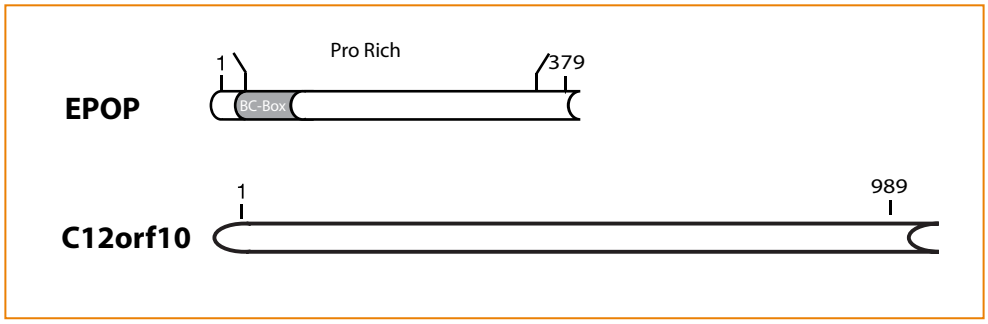


Figure 13. Recently discovered PRC2 cofactors, EPOP and C10orf12

Two recently identified factor associates with PRC2.1 complex. While EPOP bears a BC-Box domain for interaction with ELOBC, C12orf12 presents no conserved domain so far.

2. PRC2.2

Well-characterized cofactors such as AEBP2 (Adipocyte Enhancer-binding Protein2) and JARID2 (Jumonji AT-rich Interaction Domain) take part to a different type of subcomplex called PRC2.2.

2.1 JARID2

JARID2 is the founder member of the Jumonji family group of proteins. Members of the Jumonji family contain a JmjC domain, which is characterized by a-ketoglutarate-dependent histone demethylase activity, although the JmjC domain of JARID2 does not appear to be active, as it lacks the amino acids required for enzymatic activity (Klose, Kallin, & Zhang, 2006). In addition to the JmjC domain, JARID2 contains a JmjN domain, as well as two domains with DNA-binding capacity: an ARID (AT-rich interaction domain) and a zinc finger (Fig.14, upper part). JARID2 interaction with PRC2 was shown in a variety of cellular models including mES , HEK293T or HeLa (Landeira et al., 2010; Pasini et al., 2010; Peng et al., 2009a; Shen et al., 2009) as well as in mouse spleen (G. Li et al., 2010). A conserved motif in SUZ12 C-terminal portion was reported to mediate interaction with JARID2 (Peng et al., 2009), yet others found EZH2 as the strongest interactor of PRC2 (G. Li et al., 2010). The presence of a DNA binding domain within JARID2 suggest that it could be involved in PRC2 targeting. JARID2 binding sites overlap very well with PRC2 ones and JARID2 depletion impairs PRC2 enrichment at chromatin genome-wide. Conversely, PRC2 core components depletion reduces JARID2 binding at chromatin indicating that PRC2 and JARID2 recruitment is interdependent (Li et al, 2010; Peng J et al, 2009; Shen X et al, 2009; Landeira et al, 2009; Pasini et al, 2010). JARID2 artificial tethering at a transgene triggers PRC2 recruitment, H3K27me3 decoration and transcriptional repression of the reporter (Pasini D et al, 2010; Li et al. 2010; Sanulli S et al, 2015). In vitro, JARID2 binds DNA with a modest preference for GC-rich regions however this does not appear sufficient to drive selective targeting to chromatin (G. Li et al., 2010). Interestingly, N-terminal region of JARID2 has been reported to bind lncRNAs in vitro and in vivo. JARID2 interaction with an

imprinted lncRNA has been involved in PRC2 targeting in pluripotent stem cells (Kaneko, Bonasio, et al., 2014). While JARID2 was found recruited during X inactivation in a Xist dependent and PRC2 independent manner, the domain of JARID2 involved in this process seems distinct from the one previously described to interact with lncRNA (da Rocha et al., 2014).

JARID2 was shown to modulate the enzymatic activity of PRC2. The first studies reported a negative effect on PRC2 activity (Peng et al., 2009; Shen et al., 2009), but subsequent studies demonstrate that, in vitro, Jarid2 stimulates the catalytic activity of PRC2 complex (G. Li et al., 2010; Sanulli et al., 2015; Son, Shen, Margueron, & Reinberg, 2013; Zhang et al., 2011). Surprisingly, the deletion of Jarid2 in ES cells does not seem to affect the global level of H3K27me3. It was shown that JARID2 could bind nucleosomes, its stimulatory activity on PRC2 activity might reflect this ability to bridge PRC2 and chromatin (Son et al., 2013).

JARID2 was also reported to be di- and tri-methylated by PRC2 at lysine 116 both in vitro and in vivo. Under PRC2 depletion, JARID2 methylation does not occur and mutations in JARID2 lysine 116 residue does not impair JARID2 ability to tether PRC2 to a transgene, but H3K27me3 deposition is less efficient when Jarid2 cannot be methylated. It was shown that JARID2-K116me3 can promote the allosteric activation of PRC2 through the same mechanism than the one described for H3K27me3. It is therefore proposed that this mechanism would be important for de novo H3K27me3 deposition (Sanulli et al., 2015).

JARID2 function has also be studied in the context of mouse development. Jarid2 was discovered in 1995 as a regulator of neural development in a gene trap screen in mice (Takeuchi et al., 1995). Multiple phenotypes have been associated to Jarid2 loss of function: heart, neural tube formation deficiency, organ hypoplasia (Jung, Mysliwiec, & Lee, 2005; Landeira & Fisher, 2011) that appeared to vary depending on the mouse strain. Jarid2 is required for inhibition of myogenic differentiation in mammalian cells (Walters et al., 2014) and for proliferation of epidermal stem cells (Mejetta et al., 2011). In mouse embryonic stem cells, Jarid2 is abundantly expressed. Even if Jarid2-/- embryonic cells are able to survive and maintain their pluripotent status, cells depleted of Jarid2 do not properly differentiate, their growth rate is perturbed and they have a reduced capacity to form colonies in vitro (Landeira et al., 2010; G. Li et al., 2010; Pasini et al., 2010; Sanulli et al., 2015; Shen et al., 2009).

2.2 AEBP2

AEBP2 is a Gli-type zinc finger protein, which was originally identified due to its *in vitro* binding capability to the promoter region of adipose P2 (aP2) gene. This gene encodes a fatty acid-binding protein, containing three zinc finger units and a novel basic domain (Fig. 14, lower part). This protein was reported to function as a repressor based on co-transfection reporter assays (G.-P. He, Kim, & Ro,

1999). AEBP2 is evolutionarily well-conserved in all the animals ranging from fly to placental mammals. The *Drosophila* homologs, called Jing, was shown to be involved in border cell migration (Liu & Montell, 2001). Genetic studies further suggested that Jing may interact genetically with the fly Polycomb Group (PcG) protein complexes (Culi, Aroca, Modolell, & Mann, 2006). Mammalian AEBP2 has been co-purified with PRC2. It can interact with the three core components (Grijzenhout et al., 2016a). Gel shift assays using the sequences obtained from these target loci revealed a degenerated DNA-binding motif for AEBP2, CTT(N)15-23cagGCC (Kim, Kang, & Kim, 2009). Alternative splicing involving both 5′- and 3′-end exons gives rise to two major forms of AEBP2 with different protein sizes, 52 and 31 kDa. Interestingly, both isoforms keep all three zinc finger domain and have been proposed to be important for PRC2 recruitment. AEBP2 mediates stimulation of PRC2 catalytic activity however the underlying mechanism is different than for JARID2. Indeed, AEBP2 is not promoting PRC2 binding to nucleosome and act synergically with JARDI2 to stimulate PRC2 (Son et al., 2013).

Drosophila homolog of AEBP2 was reported to be required for border cell migration regulation in egg (Liu & Montell, 2001) and Aebp2 expression in mouse is detected within neural crest origin cells. Aebp2 homozigous mutant mice are lethal in early developmental phases and heterozygous mice show defects associated with neural cells migration during neural crest development (Kim, Kang, Ekram, Roh, & Kim, 2011). Later on, the same group associated the role of transcriptional repressor to the 32KDa isoform mainly expressed in the embryonic development, while the 52kDa somatic isoform has been described as a transcriptional activator enhancing cell migration properties (Kim, Ekram, Bakshi, & Kim, 2015). Recent findings observed an anterior transformation triggered by Aebp2 deletion in mice which is the opposite of the Polycomb phenotype. Additionally, in absence of AEBP2, H3K27me3 levels slightly increase at PRC2 target sites, suggesting a contrasting role toward PRC2 (Grijzenhout et al., 2016a). One possible explanation is a change in the cofactor composition of PRC2 subtypes. Hence, a transitory association of JARID2 with MTF2 (instead of AEBP2) was reported (discussed in (Holoch & Margueron, 2017)). This is further supported by in vitro data that show an increase of PRC2 catalysis adding JARID2 to C17orf96 and MTF2 (Zhang et al., 2011).

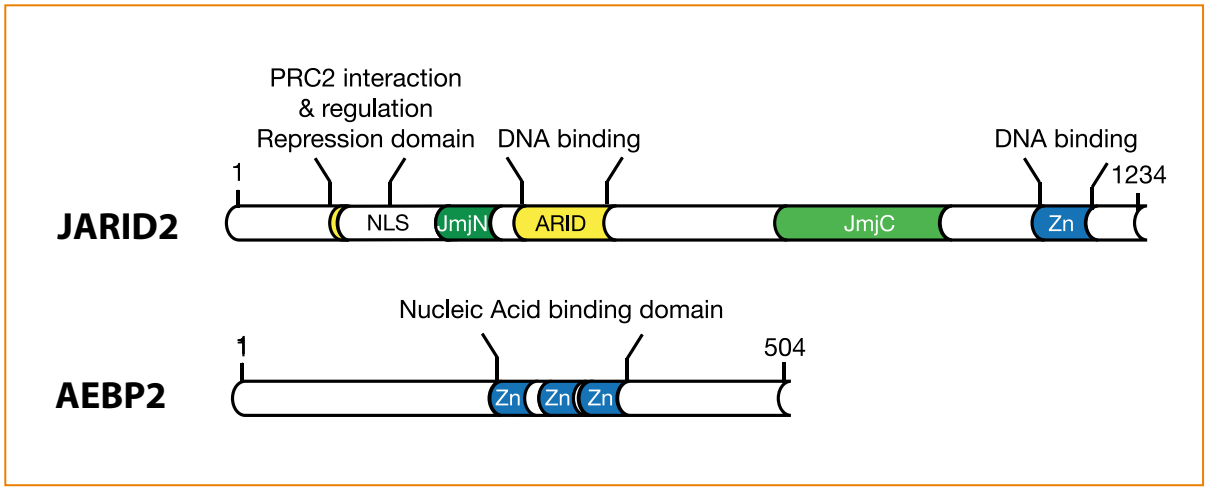


Figure 14. JARID2 and AEBP2 domains representation.

JARID2 bears an inactive JmjC domain lacking the amino acids required for enzymatic activity In addition, JARID2 contains a JmjN domain, as well as two domains with DNA-binding capacity: an ARID (AT-rich interaction domain) and a zinc finger domain. AEBP2 contains three zinc finger units (adapted by (Margueron & Reinberg, 2011)).

3. PRC2 catalytic subunit has two paralogs: *EZH1*and *EZH2*

Drosophila E(Z) gene has two paralogs in mammals: *EZH1* and *EZH2*.

Most probably, *EZH1* arises from an *EZH2* gene duplication event. These two paralogs share 63% identity and 94% identity of their catalytic domain. EZH1 associates with all the other three core components, EED, SUZ12, and RbAp46/48 (Raphael Margueron et al., 2008a; Shen et al., 2008). Despite sequence similarities between EZH1 and EZH2, their functions are quite divergent. EZH1 catalytic activity is weaker compared to EZH2, both in vitro using reconstituted complexes and in cell culture when loss of H3K27me3 due to EZH2 deletion is more efficiently rescued by Ezh2 overexpression compared to Ezh1. Interestingly, artificially tethering EZH1, even the catalytically dead version, causes target gene repression (Margueron et al., 2008). EZH1- containing PRC2 is able to compact chromatin in vitro and in vivo, as demonstrated by electron microscopy and DNase I sensitivity assay (Margueron et al., 2008) and it displays an higher affinity for chromatin (Son et al., 2013).

EZH1 and EZH2 have been seen as cooperating partners in maintaining transcriptional repression during development. EZH1 and EZH2 expression profiles are different: EZH2 is mostly expressed in proliferating cells, while EZH1 is found in both proliferating and non-proliferating tissues (Margueron et al., 2008). It has been proposed that EZH2 would be important to reestablish H3K27me3 after replication. EZH1- and EZH2- PRC2 could also work more directly together, with EZH1 counterpart being more efficient in bringing a catalytic efficient EZH2 on chromatin (Son et al., 2013). This is supported by the evidence that in myoblasts, EZH1 depletion displaces EZH2 from most of its targets, whereas EZH2 depletion could even result in slight increase of EZH1 binding.

The redundancy between Ezh1 and Ezh2 was illustrated by the worst phenotype resulting from the deletion of both genes as compare to single gene deletion in mouse skin homeostasis (Dauber et al., 2016). In hematopoiesis, Ezh1 is essential promoting slow-cycling state and preventing differentiation and senescence of adult cells (Hidalgo et al., 2012).

Recent reports proposed an unexpected role for EZH1 as an activator of genes expression. During differentiation from a myoblast into a myotube, myogenin promoter (MyoG) is activated by MyoD. This process appears to involve EZH1 for gene activation and EZH2 displacement from MyoG promoter (Stojic et al., 2011). This was supported at genome wide level in skeletal muscle differentiation, where Ezh1 complex was found to colocalize with RNAPolII and gene activation H3K4me3 mark. EZH1 depletion would lead to a global reduction in RNAPolII occupancy within gene bodies and it would cause a delay in transcriptional activation during differentiation (Mousavi, Zare, Wang, & Sartorelli, 2012). A similar hypothesis was put forward based on observation in hyppocampal development (Henriquez et al., 2013). In erythroid progenitors, EZH1 was proposed to form an independent complex with SUZ12 and without EED that is associated with active genes, suggesting a new function independent of PRC2 (Xu et al., 2015).

In conclusion, EZH1 role remains enigmatic. While it is proposed to be distinct from EZH2 function, this is difficult to conciliate with the complete lack of phenotype of its deletion and the additivity of EZH1 and EZH2 double knockout.

CHAPTER V

EPIGENETICS IN THE GERMLINE: WHICH ROLE DOES PRC2 PLAY?

"The germline is the common thread connecting the past, present and future of species"

(B J Lesch & Page, 2012)

Cells transmit information to the next generation through two different ways: genetically and epigenetically. Genetic information is based on the DNA code, whereas epigenetic information are conveyed without any modification of DNA sequence. Chromatin regulations, such as DNA methylation, histone modifications, and chromatin remodeling, have been schown to be important vectors of epigenetic information. These regulations are important at various stages of development and in particular during gametogenesis.

Gametogenesis defines the process in which germ cells develop into mature gametes, highy specialized cells capable to fuse and to give rise to the totipotent zygote. Spermatogenesis is the process of formation, development, and maturation of male germinal cells that occurs in the testis. Oogenesis is the process of formation, development, and oocyte maturation that occurs in the ovary. During gametogenesis, the formation of haploid gametes is ensured by meiosis starting from diploid cells. During meiosis, genetic exchange between parental genomes takes place through DNA recombination (meiotic recombination). In that process, homologous chromosomes are aligned in pairs

(synapsis) during prophase I and DNA double-strands breaks are made and repared to form crossover exchanges between homologous chromosomes. Chromososmes present only in one of the two homologous chromosomes such as male sexual chromosomes XY cannot be paired and are inactivated thorugh meiotic sex chromosomes inactivation process (MSCI).

The newly formed sperms and oocytes undergo epigenetic modifications during the process of differentiation into mature sperms and oocytes in order to reset their specialized programmes before allowing the formation of the totipotent zygote. High fidelity of this program has to be assured in order to prevent the transmission of aberrant epigenetic modifications to the next generation. Incorrect epigenetic settings could results either in infertility or in developmental defect (DeBaun, Niemitz, & Feinberg, 2003). In germ cells, specialized transcription factors, special histone variants or chromatin modifying complexes contribute to this process. Chromatin regulation during spermatogenesis and oogenesis, and specifically the role of the Polycomb machinery, during those processes remains underexplored. It is possibly due to the biological complexity of this developental processes and the limited availabality of the material to work with.

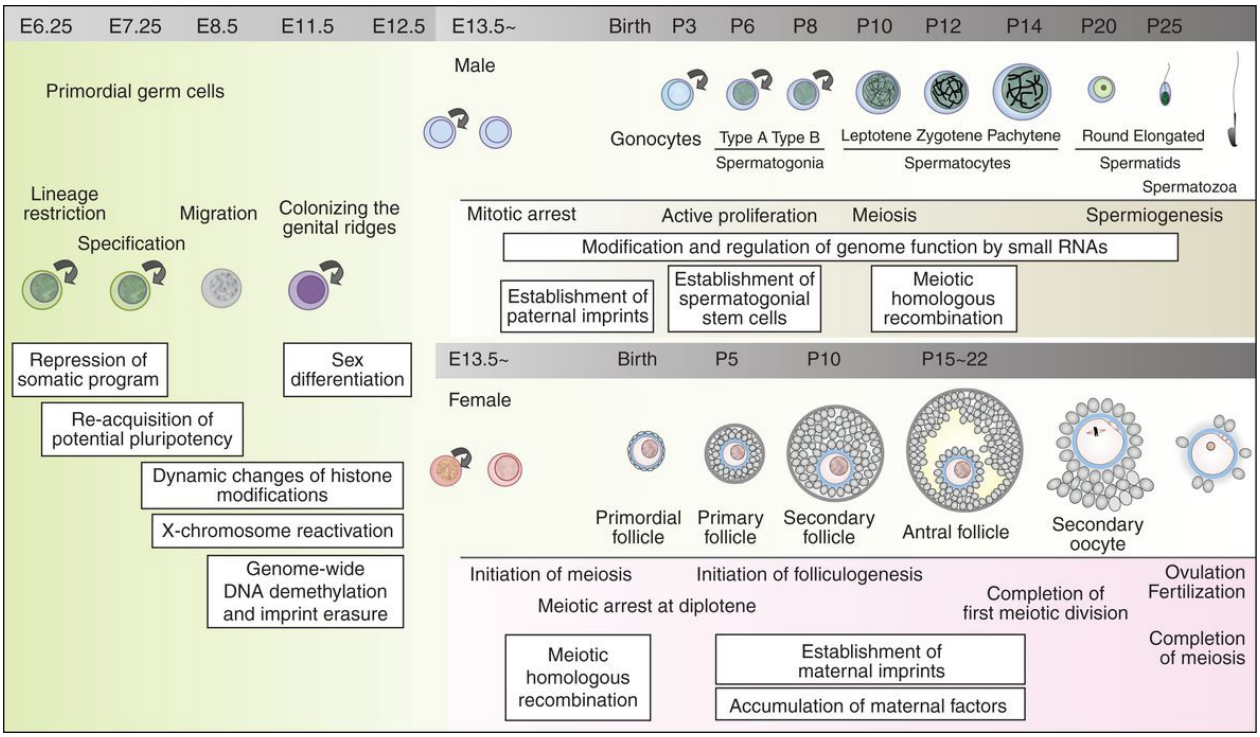


Figure 15. Schematic overview of germ cells development in mouse.

Primordial germ cells develpment from E6.25 to E12.5 (lef part); male germ cells development starting from E13.5 (right upper part) ; female germ cells development starting from E13.5 (right lower part), adapted from Saitou and Yamaji, 2012.
1. Primordial Germ Cells (PGCs) onset

Since the mouse is one of the main mammalian model systems employed in developmental biology studies, the germline development and gametogenesis will preferentially refers to this organism (Fig. 15). Human germline will be quoted once relevant biological differences have been well illustrated.

The germ cells set aside from somatic precursor early in development. PGCs specification takes place in the early post-implantation stage, at embryonic day E6.25, from pluripotent epiblast cells. Bone Morphogenetic Proteins (BMPs) and WNT signals propagate from the extra-embryonic tissues to induce the expression of PR domain zinc finger protein 1(PRDM1 or BLMP1) one of the master regulators of PGC (Ohinata et al., 2005, 2009). PRDM1 together with PRDM14 and APy constitue a tripartite transcription factor network necessary and sufficient for PGC induction in mice development (Magnúsdóttir et al., 2013). PGCs first start to form a cluster of 40 cells environ located at the base of allantois around embryonic day 7.25 (E7.25) and finally colonize genital ridge at E10.5 (Richardson & Lehmann, 2010).

Ablation of BMPs or downstream effectors in mice results in a loss or decrease of PGCs (reviewed in (Tang, Kobayashi, Irie, Dietmann, & Surani, 2016)). Also, Wnt3 mutants fail to produce PGCs even with a functional BMP signalling. Deletion of Prdm1, Prdm14 and Apy affect PGCs specification (Magnúsdóttir et al., 2013) that are subsequently lost around E8.5 and E12.5. Somatic genes are showed to be derepressed in those mutant embryos suggesting an important role of this transcription factors triumvirate in silencing the somatic program and promoting upregulation of germ cells and pluripotency genes (Grabole et al., 2013; Kurimoto, Yamaji, Seki, & Saitou, 2008). Even if these transcription factors play an independent role in other tissues, genome-wide analysis of PRDM1, PRDM14 and AP-y distribution reveals that they collaborate to provide upregulation of germ cells specific genes, to repress somatic genes and to reset epigenome. Prdm1 binds mostly at the promoters, while Prdm14 locates in distal regulatory elements. Ap-y acts downstream those two factors to activate germ cells genes or pluripotency genes like Nanog, Pou5f1 and SRY-box 2 (SOX2) (Magnúsdóttir et al., 2013; Magnusdottir & Surani, 2014). Even if no enzymatic activity has been detected for these transcription factor, it has beeen proposed that they recruit chromatin modifiers. PRDM14 has been reported to interact with the Suz12 subunit of PRC2 complex in mES, but recent reports rather indicates an interaction with CBFA2T2, protein which in turn binds to histone deacetylases. Deletion of Cbafa2t2 mimicks Prmd14 knockout condition, impairing pluripotency maintenance and PGCs specification (Nady et al., 2015; Tu et al., 2016).

2. PGCs epigenome

Global epigenetic reprogramming occuring in the early germ line erases epigenetic information coming from the parents and facilitates maturation of germ cells.

2.1 DNA methylation

DNA methylation is abundant in the post-implantation epiblast, its pattern ressembles the ones of somatic cells. PGCs need therefore to undergo genome-wide DNA demethylation to reset the epigenome (Fig.16). This process is proposed to rely both on passive and active demethylation and reaches its lower level at E13.5 (Messerschmidt, Knowles, & Solter, 2014). DNA methylation is then reestablished at after E13.5 in males and after birth for females (Kobayashi et al., 2013). During that time, the DNA methyl transferases DNMT3A and DNMT3B are repressed while DNMT1 remains active but its essential cofactor UHFR1 is repressed (Magnúsdóttir et al., 2013). For imprinted loci, active demethylation by the Ten-Eleven Translocation (TET) enzymes has been proposed to be incoved (Yamaguchi, Shen, Liu, Sendler, & Zhang, 2013).

PGCs migration through the hindgut begins on E8 and it is mainly based on ligand-receptor chemoattractant interaction. Around E.10 PGCs interact with somatic cells once they reach the gonadal ridge. Germ cells first entered the genital ridge in a "bi-potential status" before sex determination takes place. If the somatic cells have XX genotype, an ovary forms; if the somatic cells are XY, sex-determining gene Sry on the Y-chromosome is expressed around E11.5 and it favors testis formation (Koopman et al., 1991). Not only genetics makes the difference, but also somatic environment in the developing gonad. Demethylation upregulates germ cells specific genes among which deleted in azoospermia-like (*Dazl*), mouse vasa homolog, also known as *DEAD* box polypeptide 4 (*Mvh*, *Ddx4*) and synaptonemal complex protein 3 (*Sycp3*). DAZL is an RNA-binding protein, is considered a germ-cell-intrinsic competence factor: germ cells must express it to properly differentiate (reviewed in (Spiller & Bowles, 2015)).

2.2 Repressive histone marks

Loss of DNA methylation is responsible for retrotransposons deregulation and genome instability in somatic lines, while PGCs proliferate normally in the same conditions and retrotransposons remain mostly silent (Tang et al., 2015). This suggests that other silencing mechanism might compensate. Indeed, reorganization of repressive histone marks was reported (Brinkman et al., 2012; Walter, Tessandier, Perez-Palacios, & Bourc'his, 2016), in particular with an increase of repressive H3K27me3 and H2A/H4R3me2 symmetrical dimethylation. H3K9me3 mark is associated with pericentromeric chromatin and it does not vary (Seki et al., 2005). In contrast, the histone methyl transferases G9A and GLP are repressed leading to a strong reduction in H3K9me2 (Fig.16).

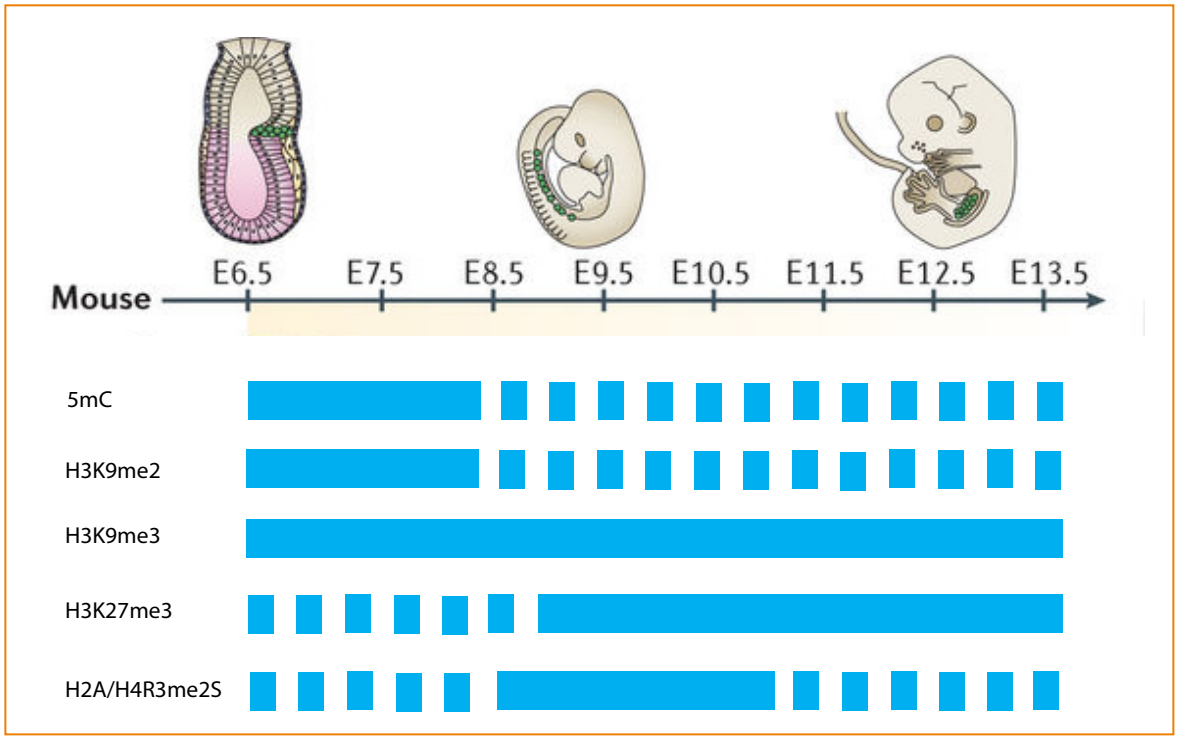


Figure 16. Epigenetic reprogramming in mouse PGCs

Main epigenetic marks are indicated on the left and high levels correspond to the continuous line whilelower levels correspond to the dashed lines. 5mC: DNA demethylation occurs in mouse primordial germ cells (PGCs) during the onset of migration and early settlement at the genital ridge (E7.5-E13.5). H3K9me2: histone H3 lysine 9 dimethylation is a repressive chromatin modification associated with DNA methylation that similarly becomes depleted. Conversely, H3K9me3 levels remai, high at pericentric chromatin, whereas global H3K27me3 levels become progressively enriched. H2A/H4R3me2s, Symmetrical dimethylation of Arginine 3 on H2A and H4 is transiently increased from E8.5 to E10.5 environ (Adapted from Tang et al, 2016).

3. Germ cells fate in Females

Female germ cells stop proliferating, begin to condense their chromosomes and enter meiosis around E.14.5. Then, they arrest in late prophase of meiosis I until ovulation. It has been proposed that retinoic acid presence boosts germ cells towards meiosis (Bowles J et al, 2006). This molecule is probably secreted either by mesonephros or by the gonad itself and triggers the expression of meiotic specific genes such stimulated by retinoic acid 8 (Stra8). Stra8 is an essential gene for meiosis and is necessary for meiosis-specific DNA replication and for DNA double-strand breaks formation late in Prophase 1. Just before entering to meiosis, pluripotency markers are down-regulated (reviewed in (Spiller & Bowles, 2015)). The Polycomb machinery was shown to be important for the timely expression of meiotic gene and to limit the response to retinoic acid signaling (Yokobayashi et al., 2013). Oogonia becomes oocytes before entering meiosis, subsequently a wave of apoptosis is proposed to eliminate the abnormal oocyte (Wear, McPike, & Watanabe, 2016). Between E16 and E18 germ cells nest form, maturing oocytes are surrounded by somatic cells. At birth, they

subsequently collapse and primordial follicles are formed with granulosa cells surrounding oocyte. Some primordial follicles from the pool proceed in to the primary follicle step with ooocyte becoming bigger and granulosa cells forming a laminar structure. A second layer of granulosa cells forms in the seconday follicle and more layer in the pre-antral one. During this maturation period, oocyte is arrested in prophase I. Once the mouse reaches maturity, meiosis further progress until Prophase II and stops at that stage untill fertilization.

3.1 Chromatin landscapes in folliculogenesis

During folliculogenesis, chromatin landscape also changes. Acetylation on Histones H3 and H4 is one of the most dominant form of modification. During meiosis, there is a decrease of acetylation levels, that could facilitate the binding of other chromatin remodelling factors such as the ATPase/helicase ATRX. ATRX function has been proposed to be important for chromosome segregation in meiosis (Kimmins & Sassone-Corsi, 2005). Histone methylation is present during primordial germ cells development. During folliculogenesis, most of the studies focused on H3 methylation on lysines 4 and lysine 9. Globally, H3K9me2 and H3K9me3 increases during oocyte maturation and it has suggested to be important for genome silencing. H3K4 methylation patterns vary during development but their role remains unexplored. The same happens for H3K27me3 whose exact role has not been defined during the different developmental steps. Different levels of H3 phosphorylation have also been detected and associated with chromatin condensation during meiosis (Pan & Keeney, 2007). Histone H1 Family oocyte-specific (H1Foo) is a maternal histone variant expressed at the secondary follicle stage and which desappears around 4-cell stage in the embryo. It seems to be required for transcriptional silencing in the oocyte (Kimmins & Sassone-Corsi, 2005).

4. Germ cells fate in Males

Following their migration toward the genital ridge, the primordial Germ Cells will interact with somatic cells (among which the Sertoli cells and peritubular myoid cells) to form the testicular cords. At E12.5, sex determination is established. Subsequently, proliferation of the PGCs (now referred as gonocytes) ceases and they entered a phase of quiescence until day 2 after birth. The control of retinoic signaling plays an important role during this process (Teletin 2017). Hence, it was shown that CYP26B1, a cytochrome P450 hydroxylase involves in the metabolism of all-trans retinoic acid is expressed in a male specific manner by ED12.5 and is required for proper mitotic arrest (Saba R et al, 2014). The first week after birth proliferation resumes and the gonocytes migrate to the basal membrane of seminiferous tubules and become the spermatogonial stem cells (SSC). These cells pool constantly maintain spermatogenesis throughout postpubertal life. Around postnatal day 8 (P8) a population of committed spermatogonia become spermatocyte. Spermatocytes undergo to a prologed meoitic prophase I further composed by 4 steps: leptotene, zygotene, pachytene and diplotene. With 2

subsequent round of meoisis, a single spermatocyte give rise to four haploid spermatids. The last steps of spermatogenesis (referred as spermiogenesis) entails the maturation of spermatids into mature spermatozoa. In the first phase they do maintain active transcription, while in the following elongating spermatid phase, their nucleus starts to elongate and transcription is repressed. Transcription levels further decrease in the condensed spermatids phase. A mature spermatozoo displays a characteristic hook-like morphology and is ready to be released in the lumen of seminiferous tubule. Spermatozoa nuclei are highly condensed and the genome is tightly compacted. Drastic chromatin changes take place during spermiogenesis: nucleosomes are displaced from chromatin and protamines populate chromatin in turn. This process seems to be required in order to promote condensation of spermatozoo head and to safeguard the paternal genome from the environment (Meistrich & RA Hess, 2013).

4.1 Chromatin landscapes in spermatogenesis

Although morphological changes during spermatogenesis have been extensively described (Fig. 17), the molecular basis of such important chromatin events has not been completely dissected. This is probably due both to the lack of suitable in vitro models that recapitulate spermatogenesis and to the lack of a specific set of cellular markers that could distinguish the different populations. So far, mouse knockout studies represent the best tools to adress this biological question.

DNA methylation is re-established quickly in gonocytes starting around E14.5. This appears to play a pivotal role in meiosis since DNMT3L knockout which impairs the de novo DNA methyltransferases leads to a general demethylation of retrotransposons and alters meiosis I (Bourc'his & Bestor, 2004). Synapsis takes place at the beginning of the meiosis, crossing overs occur at this stage. Histone modifications are also important in that process: H3K4me1/2/3, H3K9me2 and H3K9me3 marks are required for proper synapsis establishment and for post-pachytene spermatogenesis progression (Kota & Feil, 2010).

During spermiogenesis, chromatin undergo massive remodeling in a replication independent manner since cells no longer divide. This includes incorporation of histones variant, replacement by transition proteins and finaly deposition of protamine. Some histone variants are present from early meiotic spermatocytes to the elongating spermatid step, like histones H2A and H2B variants; some others, like Histone H1 variants, appear in the haploid spermatids only before the transition, most probably helping the packaging of paternal genome (Orsi, Couble, & Loppin, 2009). Testis specific histone variants TH2A and TH2B are upregulated during the first wave of spermatogenesis since the spermatogonia phase, while normal H2A and H2B expression levels tend to diminish. Double knockout for testis-specific H2A and H2B affects meiosis in spermatocytes and chromatin condensations in the later stages (Shinagawa et al., 2015). H3T incorporation in nucleosomes favors an open chromatin structure, and its knockout impairs meiosis (Ueda et al., 2017). Several variants of

histone H1 are also expressed, curiously a spermatid specific linker histone –like protein, HILS1, remains present in spermatids when core histones are mostly depleted raising the question of its function (Rathke, Baarends, Awe, & Renkawitz-Pohl, 2014). An increase of histone acetylation occurs prior to the hitone-to-protamine transition, it has been hypothesized to be important for unpacking the high order chromatin structure in order to help nucleosomes eviction and basic proteins incorporation. Acetylated H4 disapperars progressively in the condensed spermatid phase. Hitsone acetylation can also be recognized by specific "readers", hence, BRDT, a contains two bromodomains able to recognize acetylated lysines, its deletion results in abnormal elongatid spermatids and mice sterility (reviewed in(Bao & Bedford, 2016)).

Transition Proteins 1 and 2 (TP1 and TP2) are arginines and lysines rich proteins that move to chromatin in condensing spermatids. TP1 is well conserved in mammals and aboundantly expressed, while TP2 is less conserved. In vitro studies demonstrate that TP1 tends to relax DNA in the nucleosome by reducing DNA melting temperature and therefore promoting histones eviction. TP2 can act as a condensing chromatin, compacting nucleosomal DNA (Kolthur-Seetharam, Pradeepa, Gupta, Narayanaswamy, & Rao, 2009). Although their function in vitro is different, single ablation of either TP1 or TP2 results in subtle defects, probably reflecting some degree of redundancy. Consistently, double knockout mice have a more pronounced phenotype with an high histones retention (reviewed in (Bao & Bedford, 2016)).

In the last phases of spermatogenesis, the protamine 1 and 2 (PRM1 and PRM2), basic proteins enriched in lysines and cysteines, constitue most of the chromatin. Genetic deletion of only one allele of either Prm1 or Prm2 severly impairs in mice fertility, further underling the importance of these components (Cho et al., 2001). More than 90% of histones in humans and 99% in mouse are replaced by protamines very late in spermatogenesis. Nucleoprotamines assume a super-coiled toroid structure in the sperm, leading to high compaction in the sperm head. This structure inside the sperm nucleus is 6 – 20 times more condensed than nucleosome-based chromatin structure, and is believed to be essential for fertility.

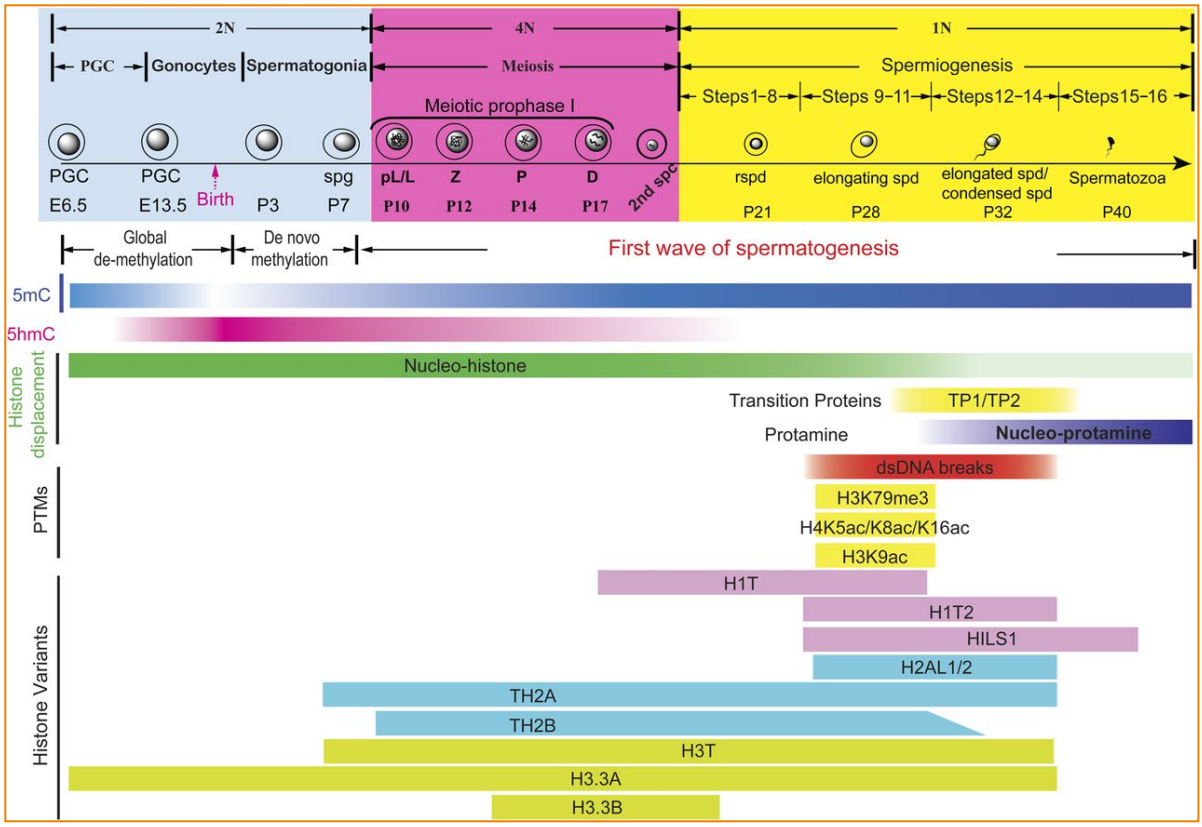


Figure 17. Schematic representation of chromatin landscape during Histones to Protamines transition in mouse spermatogenesis

The three main colors in the uppert part of the figure refer to the main phases of this process. In light blue, the diploid stage (primordial germ cells and spermatogonia); in pink, the tetraploid stage (spermatocytes) and in yellow the haploid stage (spermatids maturation). In the lower part all the epigenetic events: DNA methylation (5mC and 5hmC), histones gradual replacement with protamines on chromatin and all the histone variants and histone post-translational modifications taking part in that process (Adapted from Bao and Bedford, 2016).

As mentioned above, most histones are evicted from sperm chromatin in mammals. Yet, a small fraction remains, recent studies tried to determine whether these nucleosomes are enriched at specific location and whether they carry specific posttranslational modifications. Nucleosomes retention has been associated with imprinted genes, microRNA clusters and Hox genes (Hammoud et al., 2009). H3K4me3 has been detected on paternally expressed imprinted loci and on spermatogenesis important genes, while H3K27me3 was found on developmental genes (Brykczynska et al., 2010). Importantly, other groups suggested that sperm nucleosomes retention is not enriched for regulatory region (Brykczynska et al., 2010; Carone et al., 2014) or takes place in genes desert regions (Samans et al., 2014). The impact of sperm nucleosomes reention and their associated histones modifications on the offspring require further investingation.

While our knowledge of spermatogenesis regulation has progressed, many questions remain open in

particular regarding the role of chromatin regulation. This is of particular interest when considering the potential for epigenetic transmission to the offspring. Hence, a recent report claims that overexpression of one histone H3 lysine 4 demethylase reduces H3K4 dimethylation in the developing sperm and that transmission to the next generations affects offspring development and survival. Transmitted epigenetic signature still persists in the following generations in the absence of the H3K4 demethylating enzyme and results in aberrant gene expression patterns in the sperm (Siklenka et al., 2015).

5. PRC2 in gametogenesis

The mark deposited by PRC2, H3K27me3, is detected at all the stages of spermatogenesis (Lesch & Page, 2014; Ng et al., 2013; Sachs et al., 2013). A recent study proposed that H3K27me3 mark peaks around E10.5 and slight decrease towards E13.5. Super resolution microscopy analysis indicate that H3K27me3 move from central nucleus to nuclear lamina. The authors further suggest that PRC2 presence with a small H3K27me3 could be necessary for protecting genome from aberrant expression while DNA methylation levels are very low (Prokopuk, Stringer, Hogg, Elgass, & Western, 2017). A more direct role for PRC2 was revealed by the conditional knockout of either Eed or Suz12 in mice testis. It was shown that PRC2 is required for spermatogenesis and male fertility. In absence of a functional PRC2, the functional pool of spermatogonial stem cells exhaust rapidly probably due to the aberrant expression of somatic developmental genes (Mu, Starmer, Fedoriw, Yee, & Magnuson, 2014). Suprisingly, the same group reports that Ezh2 selective ablation in germ cells mice are still fertile and H3K27me3 mark is only 50% reduced in spermatogonial stem cells. They explained this result by the expression of Ezh1 during spermatogenesis and the cooperation between EZH1 and EZH2 to maintain H3K27me3 patterns in spermatogenesis. Consistently, double Ezh1 and Ezh2 knockout in germ cells recapitulates Eed or Suz12 knockout phenotype characterized by spermatogonial population loss and meiotic arrest (Mu, Starmer, Shibata, Yee, & Magnuson, 2017).

In contrast to the spermatogenesis, little is known about PRC2 function in the oogenesis. During folliculogenesis, PRC2 is present throughout all the developmental steps (Hinkins, Huntriss, Miller, & Picton, 2005). EZH2 ablation in growing oocytes does not impair fertility although offspring results underweight, proposing PRC2 as a maternal factor in oocytes (Erhardt et al., 2003).

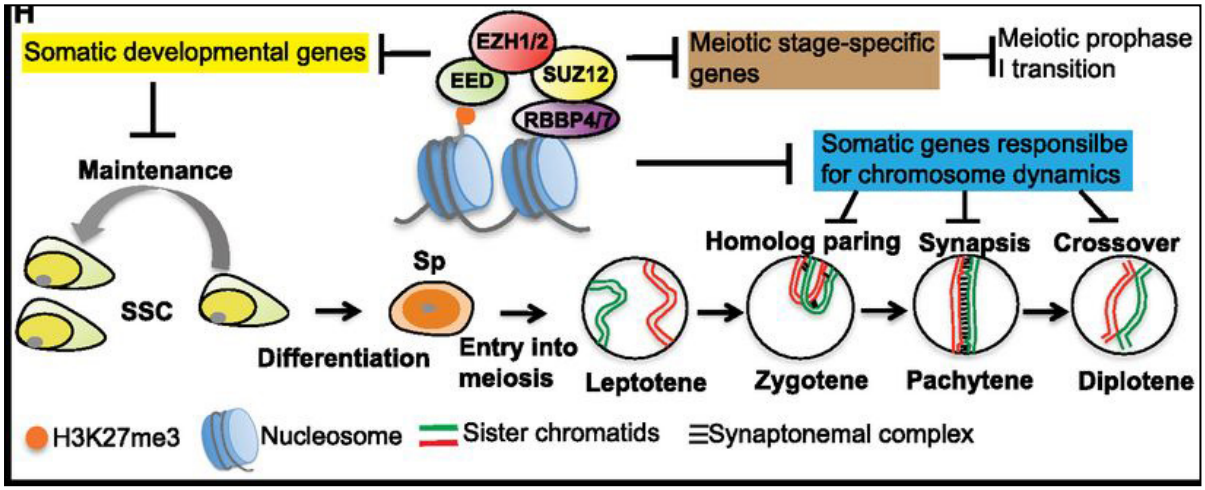


Figure 18. Proposed modelfor PRC2 role in spermatogenesis by Mu et al, 2014.

PRC2 represses different set of genes (somatic developmental genes, somatic genes responsible for chromosomes dnamics and meiotic stag specific-genes) in order to safeguard spermatogonial stem cells maintenance and meiotic progression. Upon PRC2 ablation, the expression of somatic genes impairs SSC homeostasis, ectopic expression of somatic genes interferes with homologous chromosome rearrangments and finally the upregulation of meiotic stage-specific genes affects the proper transition to Prophase I (Adapted from Mu et al, 2014)

RESULTS

PHD OUTLINE AND QUESTIONS ADRESSED

I joined Dr. Margueron's laboratory for my PhD to work on the Polycomb Repressive Complex 2 (PRC2) transacting factors. I set up a CRISPR/Cas9 genome-wide screen to identify new factors involved in PRC2 regulation. This project has yield unclear result and is still ongoing, it is therefore not discussed in that manuscript.

I also seek to identify tissue-specific cofactor of PRC2 through a mass spectrometry analysis focusing on PRC2 interactome in mice gonads (testis). In these organs, highly specialized haploid cells (gametes) are produced to ensure the continuity of life. Beside the already known cofactors of PRC2, I discovered a new potential partner. To confirm the interaction, I performed reverse experiment mass spectrometry and conventional co-immunoprecipitation. I pursued the characterization of this protein thanks to in vitro assays and by generating model cell lines to monitor its consequences on transcription and chromatin regulation. Then, I established a mouse model to investigate its function in a physiological context. Since this new protein is highly expressed in gonads, I focused on the impact of its knockout on fertility. These results are discussed in Chapter I.

Moreover I participated to an important study in the lab in which the molecular interplay between PRC2 and lncRNAs in genes silencing were dissected. We addressed this question following the case of the Long non-coding RNA HOTAIR. This lncRNA is overexpressed in several types of cancers and has been proposed to target PRC2 to its target loci in previous studies. Through biochemical assays, we defined the role of HOTAIR interaction in genes silencing and we showed that in model cell lines where we artificially tethered HOTAIR to chromatin, PRC2 is dispensable for transcriptional regulation. Moreover, HOTAIR overexpression in a breast cancer cell line model leads to subtle transcriptomic changes independently of PRC2.

These results, that we recently published in EMBO Journal (Portoso et al, 2017), are developed in Results-Chapter II.

CXorf67 interacts with Polycomb Repressive Complex 2 in gonads and is required for female fertility

Roberta Ragazzini[1], Maelle Givelet[3], Irem Baymazt[2], Raquel Perez Palacios[1], Carole Kesrouani[1], Déborah Bourc'his, Pierre Fouchet[3], Michiel Vermeulen[2], Raphaël Margueron[1,*]

[1] Institut Curie, PSL research University, INSERM U934/ CNRS UMR3215, 11-13 Rue Pierre et Marie Curie, 75005 Paris (France),
[2] Institute for Molecular Life Sciences Radboud umc, Route 274 M850.03.87 Geert Grooteplein 28, Nijmegen 6525 GA (Netherlands),
[3] Commissariat à l'énergie atomique et aux énergies alternatives, Institut de radiobiologie cellulaire et moléculaire, INSERM U967, 92260 Fontenay-aux-Roses (France).
[*]Corresponding author:
Institut Curie : 11-13, Rue Pierre et Marie Curie, 75005, Paris (e-mail: raphael.margueron@curie.fr)
Tel: +33 (0)156246551
Fax: +33 (0)156246939

ABSTRACT

The Polycomb Repressive Complex 2 (PRC2) plays an essential role in development by maintaining gene repression through the deposition of H3K27me3. A variety of cofactors have been shown to control its function in cells of various origins however little is known about PRC2 regulation during gametogenesis. This process entails genome wide changes in chromatin structure and is associated to important switch in gene expression profiles. The H3K27me3 mark is present throughout germ cell development and PRC2 function is required for spermatogenesis. Taking advantage of murine models where Ezh2 and Ezh1 were knocked in, we characterized PRC2 interactome from adult testis and identified a new polypeptide interacting with PRC2. This protein is specifically expressed in gonads, is of unknown function and does not contain any conserved domain. We have confirmed its interaction with PRC2, identified the domain of interaction with PRC2 and shown that it could tether PRC2 to chromatin. Its genetic inactivation results in aberrant deposition of H3K27me3 in male germ cells and impairs female fertility.

In conclusion, we have identified a new cofactor of PRC2 required for proper functioning of this complex in gonads.

INTRODUCTION

Early in development, cells commit to specific lineage and acquire precise identity that need subsequently to be maintained throughout the entire lifespan of the organism. Polycomb group (PcG) proteins play an important role in this process by maintaining transcriptional repression through the regulation of chromatin structure (Raphaël Margueron & Reinberg, 2011; Schwartz & Pirrotta, 2013; Simon & Kingston, 2013). In mammals, this machinery is composed of two main complexes: Polycomb Repressive Complex 1 and 2 (PRC1 and 2). PRC2 is composed of four subunits that form the core complex: the catalytic subunit EZH1/2, SUZ12, Eed and RbAp46/48 (Raphaël Margueron & Reinberg, 2011). PRC2 catalyzes the di- and tri- methylation of lysine 27 on histone H3 (H3K27me2/3). H3K27me3 is generally enriched around the promoter of transcriptionally silent genes, this mark is supposed to stabilize the recruitment of PRC1. The role of H3K27me2 is less defined; it is a widely distributed modification (covering 50-70% of all histones) whose function could be to prevent aberrant enhancer activation (Ferrari et al., 2014). PRC1 deposits H2AK119 mono-ubiquitination and promotes the compaction of chromatin (Simon & Kingston, 2013). The role of H2A ubiquitination remains debated. On the one hand, recent reports indicate that this mark could enhance the enzymatic activity of PRC2 and, at least in some context, could be sufficient to trigger PRC2 recruitment at chromatin (Blackledge et al., 2014b; Cooper et al., 2014, 2016; Kalb et al., 2014a). On the other hand, the developmental defects due to a knockout of RING1B can be partially rescued by a catalytically dead mutant (Illingworth et al., 2015; Pengelly et al., 2015). Furthermore, abrogating H2A ubiquitination by point mutation of H2A/H2Av targeted lysines does not recapitulate a Polycomb phenotype (Pengelly, Kalb, Finkl, And, et al., 2015). In contrast, the mark deposited by the PRC2 complex was shown to be required for its action, since the mutation of lysine 27 of histone H3 into an arginine leads to loss of gene repression and mutant flies that display Polycomb phenotype.

The question of how PRC2 is targeted to chromatin and what controls its enzymatic activity has retained lots of attention (Holoch & Margueron, 2017). Accumulating evidences suggest that PRC2 senses the chromatin and that an active recruitment is not necessary involved. For instance, PRC2 activity is promoted by the recognition of its own mark H3K27me3 (Raphael Margueron et al., 2009), by ubiquitination of lysine 119 of H2A (Blackledge et al., 2014; Kalb et al., 2014), by GC-rich region (Jermann, Hoerner, Burger, & Schübeler, 2014; Mendenhall et al., 2010) or by a condensed chromatin (Yuan et al., 2012). Conversely, some histone modifications negatively influence its function, in particular the modifications associated with active transcription (H3K4me3 and H3K36me3; (Schmitges et al., 2011; Voigt et al., 2012; Yuan et al., 2011)). PRC2 binding to chromatin might also be restrained by DNA methylation (Bartke et al., 2010; Jermann et al., 2014; X. Wu, Johansen, & Helin, 2013), although other reports suggest that PRC2 is compatible with DNA methylation (Cooper et al., 2014). In support of the former, DNA methylation and H3K27me3 co-localize on the inactive X suggesting that this antagonism might be context dependent. The contribution of trans-acting factors to

PRC2 recruitment remains a matter of debate. For instance, long non-coding RNA were reported to interact with PRC2 and proposed to help targeting PRC2 to chromatin (Cao et al., 2002; Pandey et al., 2008; Rinn et al., 2007). However, recent evidences suggest that PRC2 binds promiscuously to RNA (Davidovich & Cech, 2015), that PRC2 binding to RNA and chromatin might be exclusive (Beltran et al., 2016) and that PRC2 binding to RNA does not necessary mean PRC2-dependent transcriptional regulation (Portoso et al., 2017).

A number of accessory subunits have now been shown to influence PRC2 function (Holoch & Margueron, 2017; Vizàn, Beringer, Ballaré, & Di Croce, 2015). Recent comprehensive proteomic analyses suggest that they might form around two main PRC2 subtypes, PRC2.1 and PRC2.2 (Hauri et al., 2016c). In this model, PRC2.1 includes one of the three Polycomb-like proteins (PHF1, MTF2 or PHF19) together with two recently identified PRC2 partners EPOP and C10ORF12 (Alekseyenko, Gorchakov, Kharchenko, & Kuroda, 2014b). The three Polycomb-like proteins harbor one Tudor domain and two PHD fingers domain each (Holoch & Margueron, 2017b). Despite some degree of redundancy, they might also insure distinct function (Gerard L. Brien & Bracken, 2016). Their Tudor domain is able to recognize H3K36me3 decorated genes, which could be important for PRC2 association with transcribed targets (Ballaré et al., 2012; G L Brien et al., 2012; Cai et al., 2013). EPOP and C10ORF12 were purified with PRC2 through proteomic approaches (Alekseyenko et al., 2014a).Yet, the function of EPOP remains ambiguous since, in vitro, it stimulates PRC2 catalytic activity while, in vivo, it limits PRC2 binding probably through interaction with Elongin BC (Beringer et al., 2016; Liefke et al., 2017; Liefke & Shi, 2015; Zhang et al., 2011). The other complex, PRC2.2, includes JARID2 and AEBP2 subunits in equal stoichiometry (Kloet et al., 2016a; Smits et al., 2013a). Both are able to stimulate PRC2 catalytic activity in vitro with JARID2 being also able to bind nucleosomes (Son et al., 2013). JARID2 seems also required for PRC2 targeting at its loci possibly through its DNA binding domain or alternatively thanks to the methylation by PRC2 (G. Li et al., 2010; Pasini et al., 2010; Peng et al., 2009b; Sanulli et al., 2015) . AEBP2 also binds to DNA in vitro, but in vivo it appears to negatively modulate PRC2 (Grijzenhout et al., 2016b; Kim et al., 2009).

While we now have a good picture of the main accessory subunits interacting with PRC2, their precise functions are only partially understood. This might be due to compensatory mechanisms that are likely to occur when deleting a given cofactor to investigate its biological role. Also, it is not known whether additional co-factors, modulating PRC2 function in a tissue specific manner, remain to be uncovered. This prompted us to analyze PRC2 interactome in gonads. We reasoned that specialized cofactor might be needed to mediate PRC2 interaction with chromatin while its structure undergoes massive rearrangement during gametogenesis. Focusing initially on testis, we showed that PRC2 interacts with AU022751, an uncharacterized protein with no known protein domain. This protein is expressed mostly in ovaries and testis, and its expression drops sharply after fertilization. We identified a conserved stretch of 13 amino acids necessary for the interaction with PRC2. In vitro, AU022751 has a stimulatory effect on PRC2 catalytic activity and its forced recruitment at a reporter transgene leads

to transcriptional repression and H3K27me3 deposition. Constitutive inactivation of AU022751 in mice has no overt developmental phenotype, however, females have a dramatically reduced fertility. This is a correlated to aberrant deposition of H3K27me3 which appears to increase in the absence of this cofactor. Precise investigation of the underlying molecular mechanisms is ongoing.

RESULTS

The uncharacterized protein AU022751 interacts with PRC2 complex in mice testis.

A variety of cofactors were shown to modulate the enzymatic activity and the binding to chromatin of PRC2 (Holoch & Margueron, 2017; Vizàn et al., 2015). Mass spectrometry analysis suggests that all the cofactors are not binding simultaneously to PRC2 but, that instead specific combinations form around two main PRC2 subtypes (Hauri et al., 2016). Whether we have now a complete list of cofactors or whether additional cofactor, potentially functioning in specific context, remains to be investigated. The case of gonads is of particular interest considering the peculiar chromatin regulations occurring during gametogenesis. It has been shown that testis-specific PRC2 ablation results in pleotropic effects on male gonad (Mu et al., 2014, 2017) however specific regulation of PRC2 in this context has not been reported to our knowledge. To address this question, we took advantage of knock-in mouse models expressing either Flag-Tagged version of Ezh1 or Ezh2 at their respective endogenous locus (Fig. 1A). Ezh2-Flag line has been previously reported (Raphael Margueron et al., 2008a). We verified the expression of the tagged-EZH1, which is confirmed by the presence of a slowly migrating band above the signal detected in wild-type extract by western blot probed with an antibody specific for EZH1. The tag is functional since Flag-immunoprecipitation enables to pull down EZH1 from extract of the knockin line but not from extract from the corresponding wild-type line (Fig. 1B). We therefore performed Flag-immunoprecipitation followed by Mass Spectrometry analysis from testis nuclear extracts from WT, Ezh2-Flag and Ezh1-Flag lines. The results of three independents IP are represented in Volcano plots (Figure 1C). As expected, both EZH1 and EZH2 proteins interact with core complex subunits and with some of the already known accessory subunits such as AEBP2, JARID2 and two Polycomb-like proteins, PHF1 and MTF2 (Fig. 1C, left and right panel). Both PRC2 subtypes appear to be present in these extracts as illustrated by the detection Polycomb Like proteins and JARID2/AEBP2. Interestingly, our experiments also suggest the existence a new interactor, the uncharacterized protein AU022751 (ENSMUST00000117544; NM_001166433.1), which is retrieved both in EZH1 and EZH2 pull-downs. Of note, this protein was also identified as a substoichiometric PRC2 cofactor in previous mass spectrometry but its relevance was not further investigated (Kalb et al., 2014b; Kloet et al., 2016; Maier et al., 2015) and its function is so far unknown. In order to confirm this interaction, we overexpressed flag tagged versions of AU022751 and its human homolog CXorf67 in HeLa S3 cells infected with retroviral plasmid pREV enabling the expression of Au022751 and CXorf67 respectively (Fig. S1A). We performed Flag immunoprecipitation on WT, CXorf67-Flag and Au022751-Flag HeLa S3 followed by Mass Spectrometry analysis. These reverse immunoprecipitations confirmed that PRC2 is a specific interactor Au022751/CXorf67 (Fig. 1D, left panel). It also suggests that Au022751/CXorf67 does not

interact with a specific subtype of PRC2 since cofactors of both complexes are identified in the mass spec (e.g. Polycomb Like and JARID2).

Notably, other proteins seem to belong to CXorf67 and AU022751 interactome but most of them were pulled down by one homolog and not by the other thus questioning their relevance. The only exception is the deubiquitinase USP7. USP7 has been proposed to interact with SCML2, a PRC1 components important for chromatin regulation in meiotic sex chromosome inactivation (MSCI) during male meiosis (Luo M et al, 2015). Yet, since it does appear neither in Ezh1-Flag nor in Ezh2-Flag lines interactome, we decided to not pursue this hit and further investigations will be required to confirm its interaction.

In conclusion, we identified AU022751/CXorf67 as new partner of PRC2 in male gonad and showed that it does not interact with a specific PRC2 subtypes.

AU022751 is a poorly conserved protein expressed in mammalian gonads

To get some insight on the potential role of Au022751/CXorf67, we first interrogate the properties Au022751/CXorf67. Au022751 gene is located on X chromosome and two transcripts have been characterized in mouse (UCSC Mm10): one coding only for a single exon and the other for two shorter exons. Using basic alignments tool (Phylogenetic Analysis by Maximum Likelihood (PAML)), we observed that Au022751/CXorf67 homologs are present across *Eutheria* (placental mammals) but we were unable to identify any homologs outside of this clade based either on sequence conservation or on syntheny (Au022751 and *CXorf67* homologs occupy the same genetic locus between Nudt10 and Nudt11 genes on the X chromosomes). In Fig.2A phlogenetic tree of AU022751 and EZH2 are showed. Au022751/CXorf67 homologs are characterized by a fast evolution both at the nucleotide and amino acids levels, the rodent homologs being particularly distant from the rest of the homologs (Fig. 2A, upper part). This contrast with the other PRC2 components such as EED or EZH2 that are well conserved in mammals (Fig. 2A, lower part). Not any known protein domain was predicted for Au022751/CXorf67 and the only specific feature is the existence of amino acids highly enriched in serine.

We then investigated Au022751/CXorf67 gene expression. We extracted RNA from various mice tissues (adult, females and males) and analyzed Au022751/CXorf67 expression by RT-qPCR. Au022751 mRNA expression seems to be particularly high in ovary, it is also expressed in testis and brain but barely detectable in the other tissues (Fig. 2B). This pattern of expression is distinct from the one of EZH2, which is present in all tissues, the strongest expression occurring in spleen. To further confirm this specific expression of Au022751, we took advantage of publically available transcriptomic data comparing gene expression in male and female primordial germ cells versus somatic cells at E13.5 (the day by which female PGCs enter meiosis and male PGCs undergo mitotic arrest; (Percharde, Wong, & Ramalho-Santos, 2017)). This revealed that Au022751 is preferentially

expressed in primordial germ cells compared to somatic cells both in females and males following the trend of germ cells markers such as Piwil2 or Prdm14 (Fig 2C). In contrast, PRC2 core components such as Ezh2 are expressed at similar levels both in somatic and germ cells. To expand this result to the human homolog, we analyzed Tang et colleagues' transcriptomic data during human primordial germ cells development (Tang et al., 2015). We noticed that CXorf67 expression is almost absent in ESC and in somatic cells whereas, again, it is highly transcribed in male and female primordial germ cells from the onset at week 5 until week 9 of gestation (Fig. S2A).

To confirm that gene expression translates into protein accumulation, we performed immunohistochemistry on section of ovaries and testis of human origin. CXorf67 protein is detected in male germ cells inside the seminiferous tubule especially spermatogonia and round spermatids (Fig. 2D, left panel). EZH2 expression in human testis seems to follow CXorf67 pattern of expression (Fig. 2D Right panel). In ovaries, CXorf67 antibody stained primordial follicles and oocyte while it does not for external follicle cells such as thecal cells in primary follicles (Fig. 2D, right panel; Fig. S2B, left panel). EZH2 protein appears more ubiquitous with IHC signal also present in external theca cells in primary follicles (S2C, right panel).

In conclusion, AU022751/CXorf67 genes appear specific to placental mammals genomes. These fast-evolving homologs have no known protein domain. They are expressed early in germ cells development and remain present in adult gonads.

A small stretch of amino acids is necessary for PRC2 interaction

Despite the lack of overall conservation among CXorf67 homologs, sequence alignment revealed the presence of a stretch of 13 amino acids in the C-terminal part of Au022751/CXorf67, which is identical in all analyzed homologs (Fig. 3A). Considering the conserved interaction between CXorf67 homologs and PRC2, we speculated that this 13 identical amino acids sequence could be important for this interaction. To address this question, we generated plasmids allowing the expression of Cxorf67 full length or deletion mutants progressively chopping the C-terminal of the protein (Figure 3B). We transfected these constructs in HEK293 and detected their expression thanks to a N-terminal Flag Tag (figure 3B). All proteins were expressed at similar levels, we therefore performed co-immunoprecipitation to assess their capacity to interact with PRC2. Deleting the C-terminal of CXorf67 appears to weaken the interaction with PRC2 (comparing mutant lane 3 and 4 to lane 1) but does not prevent it. However, deleting the 25 amino acids comprising the conserved stretch totally abrogates CXorf67 interaction with PRC2 indicating that this sequence is necessary for the interaction (Fig. 3C).

Having established that Au022751/CXorf67 interact with PRC2, next question was to determine whether it could regulate its enzymatic activity. First, we purified hCXorf67 from Sf-9 cells infected with Baculovirus expressing either the full-length protein or the deletion mutant unable to interact

with PRC2. A strep-tag was used to purify the recombinant proteins and we checked on sizing column that they elute as expected for monomeric protein (Fig. S3). We then quantified PRC2 enzymatic activity by monitoring the transfer of methyl from radioactive S-Adenosyl Methionine (^{3}H-SAM) to recombinant histone. Prior to the addition of SAM, we incubated PRC2 complex with increasing amount of CXorf67 WT, its mutant version as negative control or Jarid2 1-530 fragment as positive control (Sanulli et al., 2015). First, we performed this assay using recombinant nucleosomes as substrate (purified from bacteria and that do not carry post-translational modification). As expected, JARID2 robustly enhances PRC2 activity, while CXorf67 has a milder but nonetheless pronounced stimulatory activity (Figure 3D left panel). We then repeated this assay on nucleosome reconstituted with octamer purified from Hela cells. These octamers carry all sort of post-translational modifications present in proliferating cells including H3K27me3. As expected, PRC2 has a more robust basal activity on this substrate, probably due to H3K27me3-mediated stimulation of PRC2 activity (Figure 3D right panel). In this context, JARID2 retains some stimulatory activity on PRC2 however the addition of CXorf67 is ineffectual.

Altogether, we identified a short sequence of CXorf67, which is necessary for the interaction with PRC2 and we showed that CXorf67 can stimulate *in vitro* the enzymatic activity of PRC2 when added to its basal but not to its already activated forms.

CXorf67 artificial recruitment mediates transcriptional repression and H3K27me3 deposition.

Considering the role of Au022751/CXorf67 in the regulation of PRC2, it was tempting and in order to further investigate the relevance of CXorf67 interaction with PRC2 complex, we generated model cell lines enabling to artificially tethered either CXorf67 WT or its mutated version to a stably integrated luciferase reporter gene (Fig. 4A and (Sanulli et al., 2015)). We first checked by western blot that both fusion proteins are expressed upon treatment with doxocycline and that they have a similar level of expression (Fig. 4B). We then analyzed the consequences of the artificial recruitment of CXorf67 on the reporter transgene. We estimated the reporter expression based on the luciferase activity quantification. This assay revealed that the recruitment of CXorf67 has a robust repressive activity on the transgene (Fig. 4C). Importantly, the mutant of CXorf67 unable to interact with PRC2 has a much milder, if any, repressive potential. To determine whether the chromatin modifying activity of PRC2 could be involved, we performed Chromatin Immunoprecipitation to measure H3K27me3 enrichment at the transgene. As a control, we checked that both fusion proteins are recruited in a doxycycline dependent manner at the transgene (Fig 4D). Upon recruitment of the CXorf67, we observed a robust increase of H3K27me3, while its enrichment is not affected upon recruitment of the mutant version of CXorf67 (Fig 4E).

We conclude that CXorf67 is able to mediate transcriptional silencing and to promote H3K27me3 PRC2-mediated deposition when artificially tethered at a reporter gene.

AU022751 Knockout males are viable and fertile despite a global increase in H3K27me3.

To determine the contribution of Au022751/CXorf67 in a more physiological context, we generated a knockout mouse model in which both AU0022751 genetic isoforms have been completely deleted (Fig. S3A). Genetic deletion was realized taking advantage of CRISPR/Cas9 system in which 2 guides RNA, one targeting position +28 after TSS and the other +1511, and recombinant hCas9 were injected into mouse embryo at one cell stage and blastocyst implanted into a surrogate mother. Thirteen pups have been harvested and eight were founders carrying 1.5Kb deletion, 6 males and 2 females (data not shown). Two pups carrying the same genetic deletion were further backcrossed with WT C57B6N mice in order to obtain non-mosaic homozygous knockout mice for AU022751. Mice phenotype has been analyzed starting from F1 generation in which KOY males have been obtained. We confirmed that Au0022751 is absent from both testis and ovaries from knockout mice at the mRNA level while the expression of the flanking genes (Nudt10 and Nudt11) appears unaffected (Fig. S5C). We confirmed this result at the protein level thanks to antibodies that we rose against AU022751, showing that a band corresponding to AU022751 disappear from testis of knockout mice (Fig. S5B). As expected considering Au022751 specific expression in gonads, the knockout mice did not show any obvious sign of developmental defect, we therefore focused on gametogenesis.

First, we analyzed more precisely AU022751 expression in male germ cells. Toward this end, we sorted different subpopulations of cells by Flow Cytometry based on staining for α6-integrin, for the tyrosine kinase receptor c-Kit and for DNA content as previously described (Corbineau et al., 2016). We extracted the RNA from these sorted cells and quantified gene expression by RT-qPCR. This experiment indicated that AU0022751 is mostly expressed in spermatogonia (α6-integrin positive), particularly in the differentiating spermatogonia (c-Kit positive). Its expression is very low in spermatocyte I and II, which is consistent with the global X transcriptional inactivation at that stages (Lifschytz & Lindsley, 1972). In contrast, EZH2 transcript increases at the final stages of differentiation (Fig. 5A, spermatogenesis scheme below).

Considering the link between AU022751 and PRC2, we investigated whether the deletion of AU022751 could affect H3K27me3 global level. We evaluated several histone modifications probing nuclear extracts from whole testis by western blot. Surprisingly, H3K27me3 levels are almost 2 fold higher in absence of AU022751, while none of the other histone marks analyzed vary noticeably (Fig. 5B). Importantly, this upregulation does not appear to reflect a direct regulation of PRC2 core components since EZH2 and SUZ12 levels were unaffected as gauged by western blot (Fig. 5B, bottom panel). To further confirm the upregulation of H3K27me3 and to determine in which cell population it occurs, we stained testis section for H3K27me3 (Fig. 5C). As previously reported, we observed that round-shaped somatic Sertoli cells (Fig. 5C, yellow arrows) are strongly positive for H3K27me3 both in WT and KO condition (Iwamori, Iwamori, & Matzuk, 2013). In contrast, the germ

cells, identified by the positive stain for the nuclear germ cells marker TRA98, display a much stronger staining H3K27me3.

We then investigated whether this alteration of H3K27me3 impact on spermatogenesis. We noticed that KOY males are fertile and that the testis-over-body weight ratio is similar in KOY and WTY mice suggesting a lack of major defect (Fig.S2C). Consistently, we did not change any change in the spermatogenesis sub-populations that we separated by FACS (Fig S4A, S4B and S4C). We then evaluated sperm quality with an apposite machine. Visually sperm motility appears mostly in the normal standards, although AU022751 knockout sperm could be slightly less progressive and a bit more static (Fig. 5E, left and right panel respectively).

Taken together our results indicate that AU022751 regulates PRC2 activity during spermatogenesis and that its deletion result in a global increase in H3K27me3. Yet, this aberrant regulation of H3K27me3 has limited consequences, if any, on spermatogenesis.

AU022751 deletion impairs female fertility

In view of the high expression level of Au022751/CXorf67 in ovaries, we then investigated the consequences of its deletion in this organ. We first determine its expression during oocyte development (GEO: GSE70116) thanks to available RNA-seq (Veselovska et al., 2015). The data indicates that Au022751 (Fig.6A) and CXorf67 (Fig. S5A) are highly expressed at all developmental phases, from non-growing oocyte to a fully developed oocyte, but their expression drops during post-fertilization stages around 2-4 cell stages embryo (Fig. 6B and S6A). As observed in male gonads, this pattern of expression does not resemble what we observed for EZH2 (Fig. 6A, 6B, S6A).

While we obtained Au022751 knockout mice that appear normal in terms of development, we noticed the rare number of pups originating from these mice. This prompted us to carefully monitor their fertility and we therefore compare the progeny of wild type versus knockout females. In brief, 6-weeks-old wild type and knockout females were mated with a known fertile male (previously tested) in the same cage and followed for 20 weeks (14 cages in total) on a daily basis. All born litters were genotyped in order to insure the mother genotype. The results revealed a substantial reduction in the number of pups, which become even more dramatic while females aged (Fig. 6C). This reflects a reduced number of litter, WT females give birth to around 3 litters over a period of 20 weeks while knockout females give on average birth to only one litter during the same time. This effect is further amplified by the size of the litter: wild type females give birth to an average 8 pups per litter while AU022751 knockout average litter size is around 3/4 pups per litter (Fig. 6D). We then analyzed whether this result could be due to an abnormal reproductive system. The uterus and uterine horn appear normal, yet the ovaries seem to have a reduced size (Fig S6B and S6C). This observation was confirmed when we weighted the ovaries: ovaries from wild type and heterozygous mice have an identical weight while homozygous AU022751 knockout ovaries had a weight reduced of about 30%

(Fig S6D). To further investigate this defect, we prepared ovaries sections stained with either hematoxylin/eosin or immunostained against VASA (DEAD box RNA helicase), a cytoplasmic germ cells marker. We did not observe any reduction in the number of secondary and antral follicles in the knockout females. However, the mutant females display a reduced number of primary follicles and even more pronounced depletion of primordial follicles (Fig. 6D). These data might reflect a decrease in the initial pool of primordial follicles or instead a progressive exhaustion of this pool due to an altered control of differentiation.

Altogether, these results highlight the crucial role of AU022751 during oogenesis, which could be due to an aberrant deposition of PRC2-mediated H3K27me3.

DISCUSSION

Here, we report the identification of a new partner of PRC2 expressed mostly in gonads. We show that artificially targeting this factor to chromatin is sufficient to promote PRC2 recruitment, transcriptional repression and H3K27me3 deposition. Yet, inactivating this factor result in global increase of H3K27me3. While this aberrant regulation of H3K27me3 appear to have limited consequences on spermatogenesis, it seems to jeopardize oogenesis leading to a dramatically reduced fertility.

Several cofactors of PRC2 were identified, they were proposed to assemble around two subtypes of PRC2 implying that some combination of cofactors do not exist in a normal context (Hauri et al., 2016). For instance, it was shown that the immunoprecipitation of AEBP2 does not pull down the Polycomb Like cofactors (Grijzenhout et al., 2016). The case of Au022751/CXorf67 seems different as it pulls down cofactors belonging to both PRC2 subtypes when overexpressed in Hela cells. It remains to be demonstrated that this result hold true in its physiological environment. Indeed, CXorf67 is normally absent from Hela cells, its expression appears more specific to the gonads. Of note, neither EPOP (C17orf96) nor C10orf12 were identified in the EZH1/2 IPs of testis suggesting of a specific cofactor composition in this context. This question whether these cofactors and Au022751/CXorf67 are normally found in the same cells. Indeed, looking at oogenesis and early mouse development Au022751 and Epop have a rather exclusive pattern of expression (our result and (Beringer et al., 2016)). Incidentally, it is also striking that both cofactors appear to interact with the deubiquitinase USP7 when not associated to PRC2 (our result and (Liefke et al., 2017).

An intriguing result of our study is the substantial increase in H3K27me3 global level when Au022751 is deleted from mouse testis. This appears counterintuitive with our *in vitro* observations that titrating Au022751 in a histone methyl transferase assay monitoring PRC2 activity on recombinant histone enhances H3K27 methylation. Surprisingly, similar observations were reported with other cofactors of PRC2. For instance, AEBP2 knockout leads to slight increase of H3K27me3 (Grijzenhout et al., 2016), knock down of Epop promotes H3K27me3 deposition at CpG island (Liefke & Shi, 2015), and, to a lesser extend, JARID2 deletion has limited consequences on H3K27me3 maintenance (G. Li et al., 2010; Peng et al., 2009; Sanulli et al., 2015). Yet, those three factors were reported to stimulate PRC2 enzymatic activity in vitro (Son et al., 2013; Zhang et al., 2011). This apparent discrepancy might be due to the fact that the in vitro assays generally assess PRC2 activity in particular context. Hence, often the assays are performed on recombinant histone thus evaluating only de novo deposition. In agreement, we showed that the stimulatory activity of Au022751 on PRC2 activity is lost when native histones are used as substrate. Another source of discrepancy is related to the cofactors. On the one hand, it is difficult to take into consideration all the cofactors of PRC2 in vitro. On the other hand, some of the in vivo analysis might be mitigated by redundancy or compensatory

mechanism between the cofactors. For instance, it has been shown that the deletion of AEBP2 promotes the formation of a new flavor of PRC2 complex including JARID2 and PCLs (Grijzenhout et al., 2016). In the future, it will be very informative to determine the composition of PRC2 in testis depleted of Au022751.

Our study shed a new light on PRC2 regulation during gametogenesis. For the first time to our knowledge, we identified a cofactor whose expression seems to be mostly restricted to specific organs. Although it was also identified in a few studies analyzing PRC2 interactome in mES cells, gene expression profile suggest that its level is very low in this context (as compared to gonads) and the lack of clear developmental defect suggest that it may not play a major role in this context. Since Au022751 has no clear protein domain, its molecular function remains elusive. We speculate that its basic amino acids composition (Isoelectric point around 9.5-10) makes it a good candidate to interact with nucleic acid yet further investigations are required. Importantly, Au022751 enable us to manipulate H3K27me3 deposition during spermatogenesis. While it was previously shown that the inactivation of PRC2 was deleterious (Mu et al., 2014), our results suggest that an increase dosage of H3K27me3 does not impair spermatogenesis substantially. It will now be interesting to determine whether this excess of H3K27me3 is removed during spermiogenesis or whether some of it is retained and could potentially impact on the early development of the embryo. In contrast, a right dosage of H3K27me3 seems to be important during oogenesis considering that the deletion of Au022751 impairs the control of follicle differentiation. Transcriptomic analysis will be necessary to dissect the precise contribution of Au022751. It will also be informative to determine whether the early development of the embryo is affected by the deletion of Au022751.

Alterations of the Polycomb machinery are frequently observed in cancer. The genetic mutations that impact PRC2 result in a broad range of outcome ranging from complete loss of function to gain of function (Wassef & Margueron, 2017). A less studied case is the translocations that have been reported in endometrial stromal sarcoma, a rare malignant tumor of the uterus (Puliyath & Nair, 2012). Previous reports revealed frequent fusion between the transcriptional repressor JAZF1 and the PRC2 core component SUZ12 (Hrzenjak, 2016) and it was proposed that this fusion would alter PRC2 function (Ma et al., 2017). Interestingly, other fusions have already been reported to implicate PRC2 through the fusion between the PRC2 cofactor PHF1 and various transcriptional regulators (JAZF1, MEAF6, EPC1 (Dewaele et al., 2014) . Our study extends this observation by showing that the fusion between the nuclear protein malignant brain tumor domain-containing 1 (MBTD1) and CXorf67 is also likely to result in the aberrant targeting of PRC2 (Dewaele et al., 2014). It will be important to investigate how these fusions contribute to tumor progression and whether the inhibition of PRC2 could constitute a therapeutical strategy.

FIGURE LEGENDS

Figure 1. AU022751 is an uncharacterized protein mainly interacting with PRC2 complex.

(A) Schematic representation of Ezh1- and Ezh2- Flag mice lines knock-in (KI) (B) Anti-EZH1 WB analysis after Flag-IP from wild type or EZH1-KI nuclear extracts. (C) Volcano Plot representation of EZH1 (left panel) and EZH2 (right panel) interactome from mice testis. In red are all the core complex subunits, in green the cofactors and in blue a new interactor. (D) Volcano Plot representation of CXorf67 (left panel) or Au02275 (right panel) interactome after Flag-IP on HeLa S3 overexpressing a tagged version of the previously mentioned proteins. Same color codes as in (C).

Figure 2. AU022751 is present across placental mammal genomes and is expressed mainly in germ cells.

(A) Phylogenetic tree representing AU022751/CXorf67 (upper part) and EZH2 (lower part)across representative placental mammals. The branches length (red number) is proportional to the number of substitutions per site. Analysis performed with LIRMM (Laboratoire d'Informatique, de Robotique et de Microélectronique de Montpellier) (B) AU022751 and EZH2 mRNA relative abundancy normalized to TBP in various mice tissues (N=2). (C) RNA-seq data (GSE89711) from murine male PGCs at E13.5, male somatic cells (mSOMA), female PGCs at E13.5 and female somatic cells (fSOMA) showing Au022751, Piwil2, Ezh2, Jarid2, Aebp and Gapdh expression. FPKM log2 values (mean ±SD, n≥ 3), (D) CXorf67 and EZH2 IHC staining on human ovaries sections (E) CXorf67 and EZH2 IHC staining on human seminiferous tubules sections.

Figure 3. CXorf67 contains a conserved region required for interaction with PRC2 and regulation of its enzymatic activity.

(A) Protein Sequence aligments (ClustalW) from Dolphin, human, Dog, Mouse and Rat. The color of the letters represents the amino acid charge; asterisks and black dot below the letters indicate identical or similar amino acids respectively. The 12 amino acids conserved stretch is circled in red. (B) Different CXorf67 cDNA truncation cloned in pREV vector and transiently overexpressed in 293HEK cells: CXorf67 full length (a.a. 1-503), CXorf67 mutant (a.a. 1-395), CXorf67 mutant (a.a. 1-420) and CXorf67 mutant (a.a. 1-450) (C) WB Analysis of EZH2, EED, SUZ12, RbAp48, HDAC1 after Flag-IP in HEK 293 transiently overexpressing the different versions of CXorf67 described above (D) HKMT assay performed with PRC2-Ezh2 on recombinant nucleosomes in presence of either increasing doses of CXorf67 WT or CXorf67 (a.a. 1-395). Jarid2 fragment 1-530 is used as positive control for PRC2 regulation. (E) HKMT assay performed with PRC2-Ezh2 on native nucleosomes purified from HeLa as in (D). The upper panel in (D) and (E) are autoradiography and the lower panel is the corresponding coomassie staining.

Figure 4. Artificially tethering of CXorf67 leads to transcriptional repression of reporter gene and H3K27me3 deposition.

(A) Schematic representation of doxycyclin inducible system to express Gal4-CXorf67 full length or deletion mutant (a.a. 1-395) and of the luciferase reporter transgene. The reporter is downstream the minimal TK promoter controlled by 5X UAS sequence. Black arrows indicate primers used for ChIP, Upstream LUCA and downstream LUCD. (B) WB probed with CXorf67 antibody on nuclear extracts of T-Rex 293T cells model cell lines described above with or without 48h doxocyclin induction. (C) Luciferase activity in response to dox treatment in the cell lines reported on the X axis. Values represent the Relative Luciferase Activity normalized to the amount of protein (mean ±SD, n≥ 3). (D) ChIP in the same condition as above. Antibodies used for ChIP are indicated on the X axis. Y axis represent percent of input recovered on LUCD region (mean ±SD, n≥ 3). (E) Same as (D). Y axis represent percent of input recovered on LUCA region (mean ±SD, n≥ 3).

Figure 5. AU022751 knockout males' analysis.

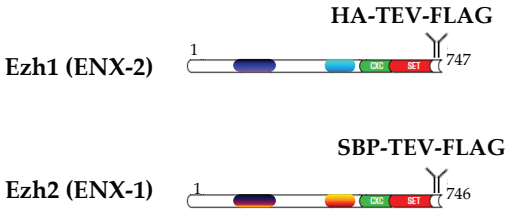
(A) RT-qPCR analysis of Au022751 and Ezh2 mRNA expression during spermatogenesis. Au022751 and Ezh2 mRNA amount is normalized to TBP. All the different spermatogenic populations (undifferentiated spermatogonia kit-, differentiating spermatogonia kit+, 4N, 2N, N) have been sorted and mRNA extracted (mean ±SD, n≥ 2). (B) Western blot analysis of histone marks H3K27ac, H3K4me3, H3K27m3, H3K27me3, H3K9me2 and H3 (loading control) on whole testis extracts Au022751 WT and knockout. In the bottom part, western blot analysis robing EZH2 and SUZ12 on whole testis extracts WT and Au022751 knockout. (C) Immunofluorescence analysis of H3K27me3 in testis sections (6 months old mice). DAPI staining, H3K27me3 staining, TRA98 staining (germ cells marker) and merge. (D) Sperm quality analysis of wt and Au022751 knockout males. Left panel: % sperm progressivity. Right panel: % sperm rapidity (mean ±SD, n≥ 2).

Figure 6. AU022751 knockout females' analysis.

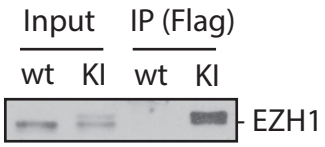
(A) Strand-specific RNA-Seq libraries using ribosomal RNA depletion on oocytes isolated at different stages of follicular growth (i.e., non-growing oocytes (NGOs); growing oocytes (GOs; GO1 for mice aged 8–14 days post-partum (dpp), GO2 for mice aged 15 dpp); fully grown oocytes (FGOs)) (GEO:GSE70116). (B) Single-cell RNA-seq Ezh2 and Au022751 expression data on early embryo developmental phases (oocyte, pronucleus, 2 cells, 4 cell, 8 cells and morula). (C), (D) and (E) comparative study of WT and Au022751 knockout fertility monitored during 5 months. (C) total number of pups per each genotype is represented on Y axis, month time points on X axis, (D) total number of litters in 5 months is represented on Y axis, genotype is represented on the X axis and (E) size of each litter is represented on Y axis, genotype is represented on the X axis. (F) Follicles

counting on WT and Au022751 knockout of 5 months old female slides (30 slides counted for each genotype). Each panel corresponds to a different folliculogenesis step: primordial follicles, primary follicles, secondary follicles and antral follicles number per genotype with respective schematic illustration below.

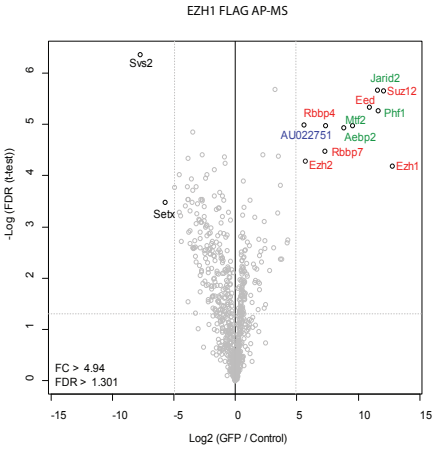
A

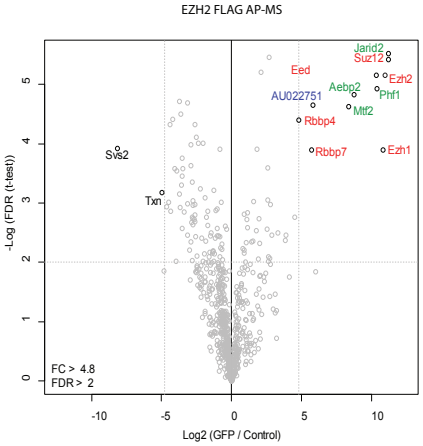


B

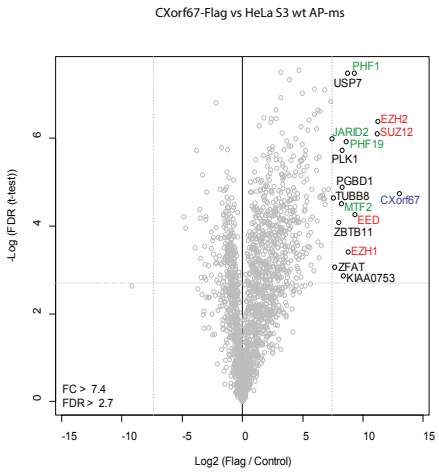


C





D



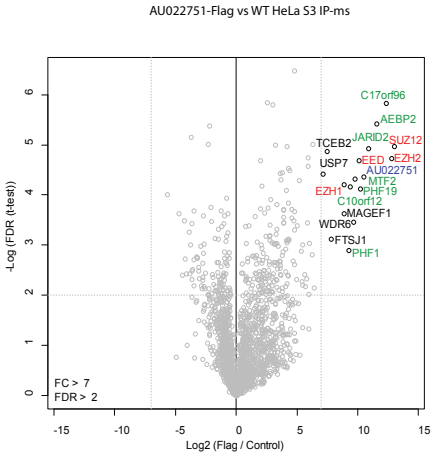
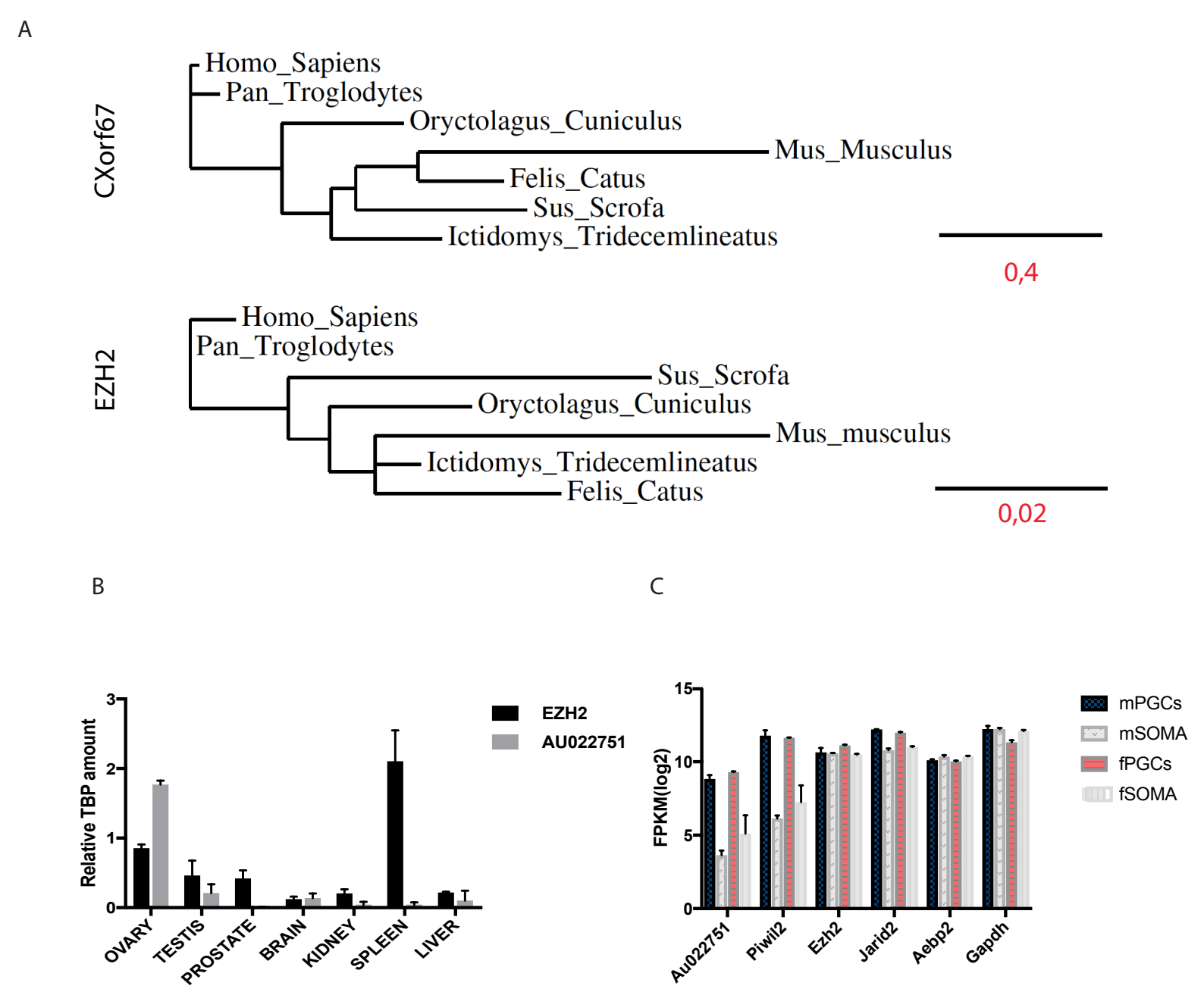


Fig. 1 Ragazzini et al



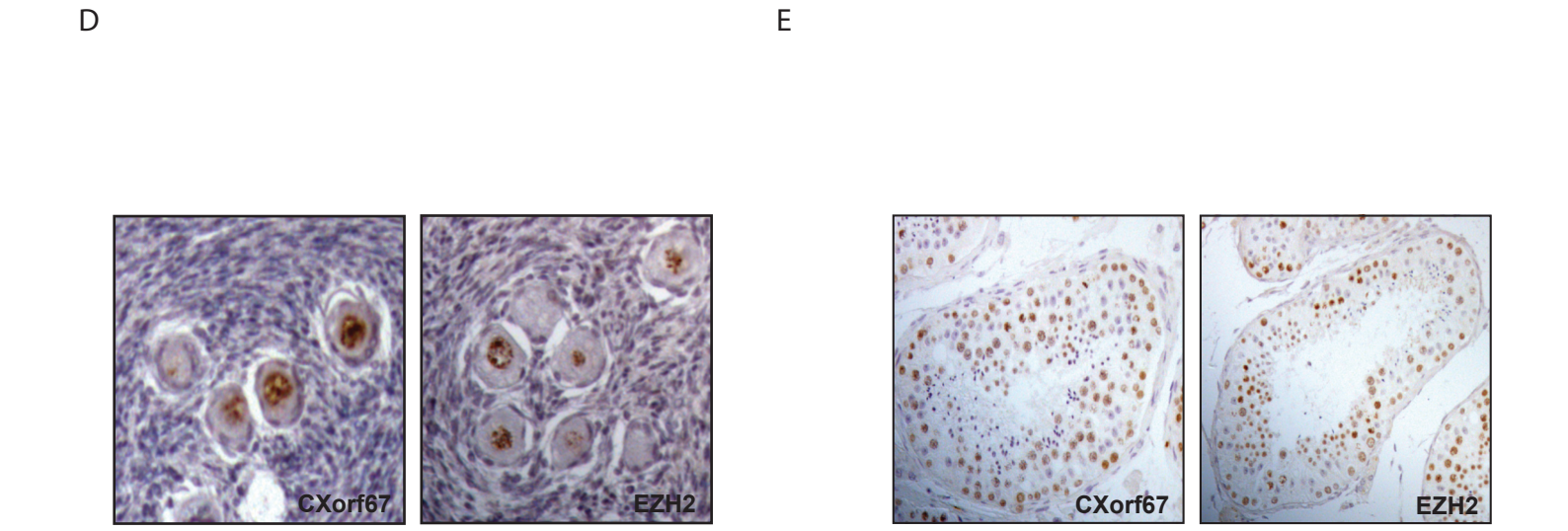


Fig. 2 Ragazzini et al

Fig. 3 Ragazzini et al

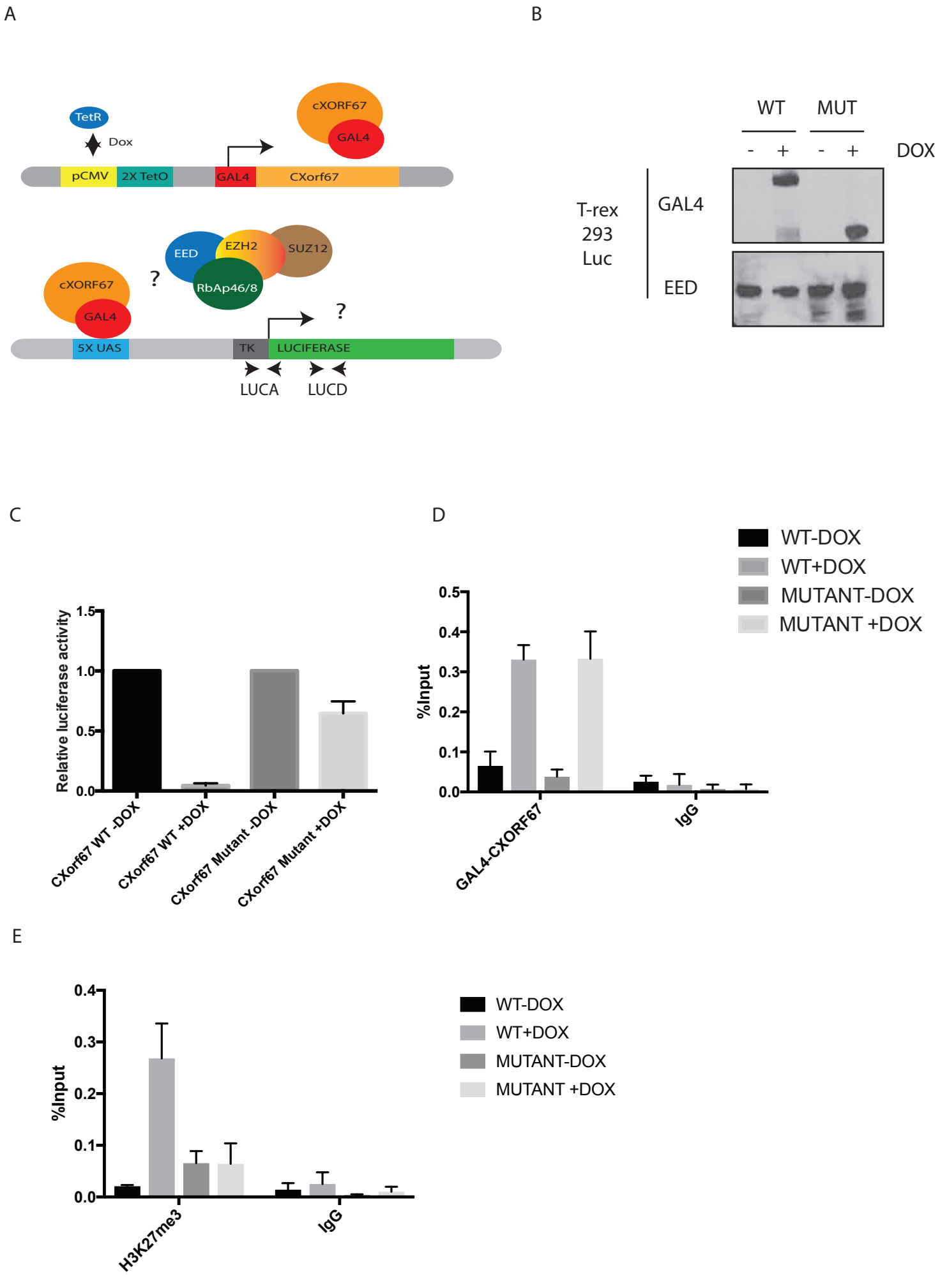


Fig. 4 Ragazzini et al

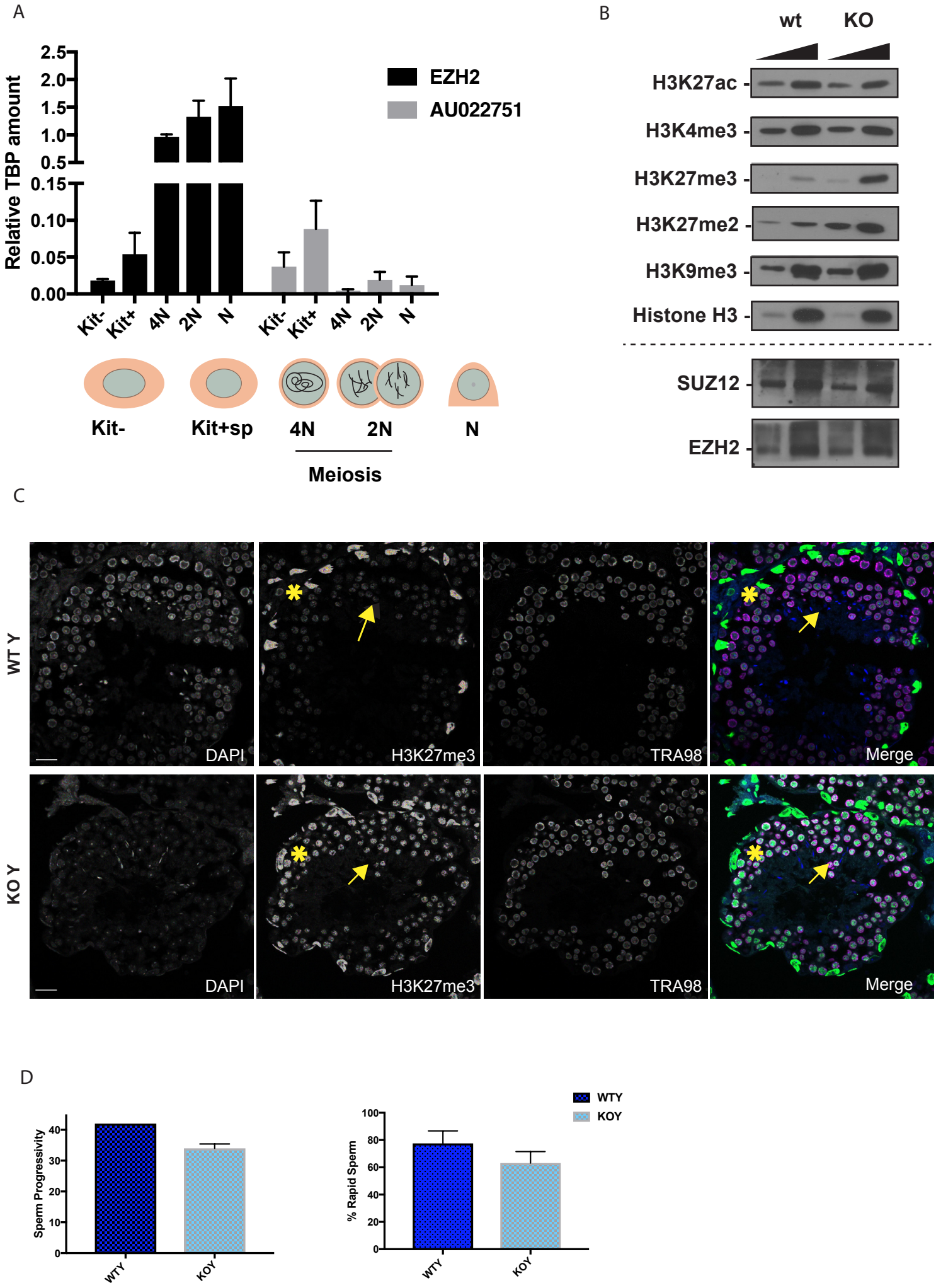


Fig. 5 Ragazzini et al

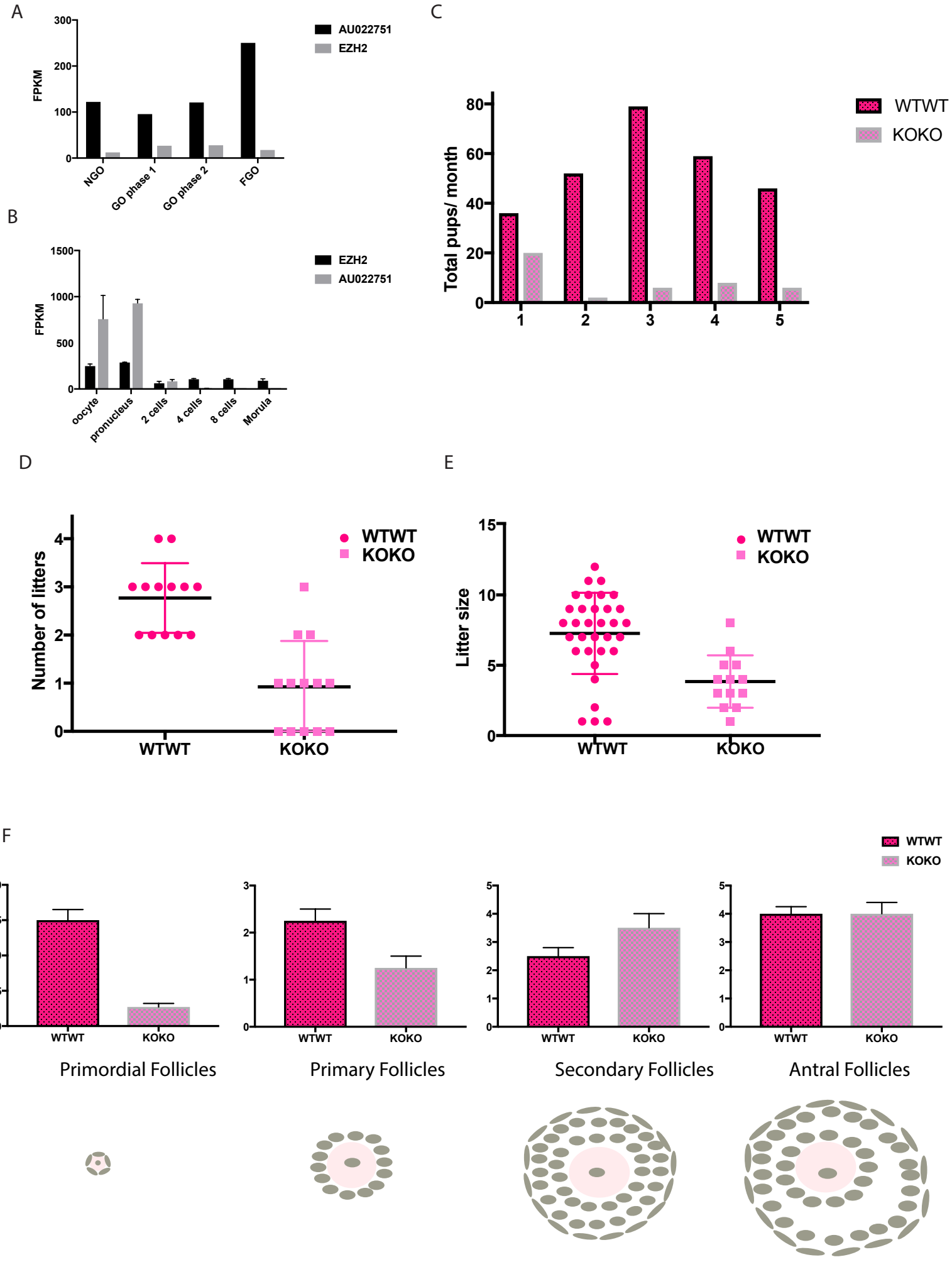


Fig. 6 Ragazzini et al

SUPPLEMENTARY FIGURE LEGENDS

Figure S1. (A) Anti Flag WB analysis of HeLa S3 nuclear extracts WT or stably overexpressing either CXorf67- or AU022751- Flag ; HDAC-1 has been used as loading control. (B) Left : Scheme for CXorf67 WT and mutant (a.a. 1-395) purification in Sf-9 insect cells. Right : Coomassie staining of eluted fractions from the Shepadex Increase 200 size-exclusion column (left panel). Top : CXorf67 WT, bottom : CXorf67 mutant.

Figure S2. (A) RNA-seq data from Percharde et al, 2017 (GSE89711) on E13.5 embryos PGCs and Somatic cells (males and females). Analyzed genes are shown on the X axis and FPKM log2 values are shown on the Y axis (mean ±SD, n≥ 3). (B) CXorf67 and EZH2 IHC staining on human ovaries sections showing primary follicle (at the center) and a group of primordial follicles (black arrows). (C) Mice testis absolute weight (gr) and testis/body weight ratio values from adult WT and AU022751 knockout mice (mean ±SD, n≥ 3).

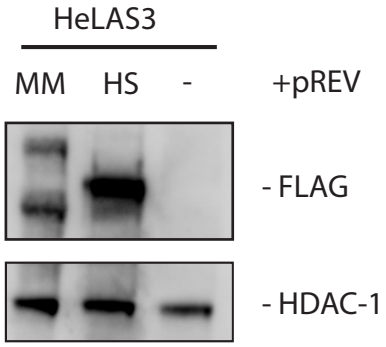
Figure S3. (A) Au022751 locus deletion in mice with schematic representation of founders sequencing. (B) WB analysis anti-AU022751 purified serum on WT and knockout testis nuclear extracts. (C) NUDT10, AU022751, NUDT11 and EZH2 mRNA relative abundancy normalized to GAPDH in adult males mice WT and knockout (mean ±SD, n≥ 3). (D) NUDT10, AU022751 and NUDT11 mRNA relative abundancy normalized to GAPDH in adult males mice WT and knockout (mean ±SD, n≥ 3).

Figure S4. FACS analysis of WT and Au022751 knockout adult mice cell testicular supensions. (A) Whole cellular suspension has been stained for PLZF (undifferentiated spermatogonia): PLZF is on the X axis and SSC on the Y axis (Side Scatter), (B) B-microglobulin 2 selected population has been analyzed for DNA content stained with Hoechst (Red and Blue Hoechst on the X and Y axis, respectively) in order to distinguish spermatogonial population (SP), primary spermatocytes (4N), secondary spermatocytes (2N) and round spermatids (N). (C) Spermatogonial Population (SP) has been stained for alpha-integrin (Y axis) and KIT (X axis) in order to distinguish undifferentiated spermatogonia (Kit-) from more differentiated spermatogonia (Kit+). Percentage of different cell population is similar in WT and AU022751 knockout (n≥ 3).

Figure S5. (A) CXorf67 and EZH2 analysis from transcriptomic data set in the human early embryo development (GSE18290): 1-cell stage, 2-cells stage, 4-cells stage, 8-cells stage, Morula and Blastocyst (mean ±SD, n≥ 3) (B), (C) and (D) Adult females WT, Au022751 Heterozygous and knockout reproductive tractus analysis: in (B) the all reproductive tractus is showed, in (C)

each dissected ovaries pair is showed and in (D) absolute ovaries weight (gr) is reported on the Y axis and the respective genotypes indicated in the legend.

A



B

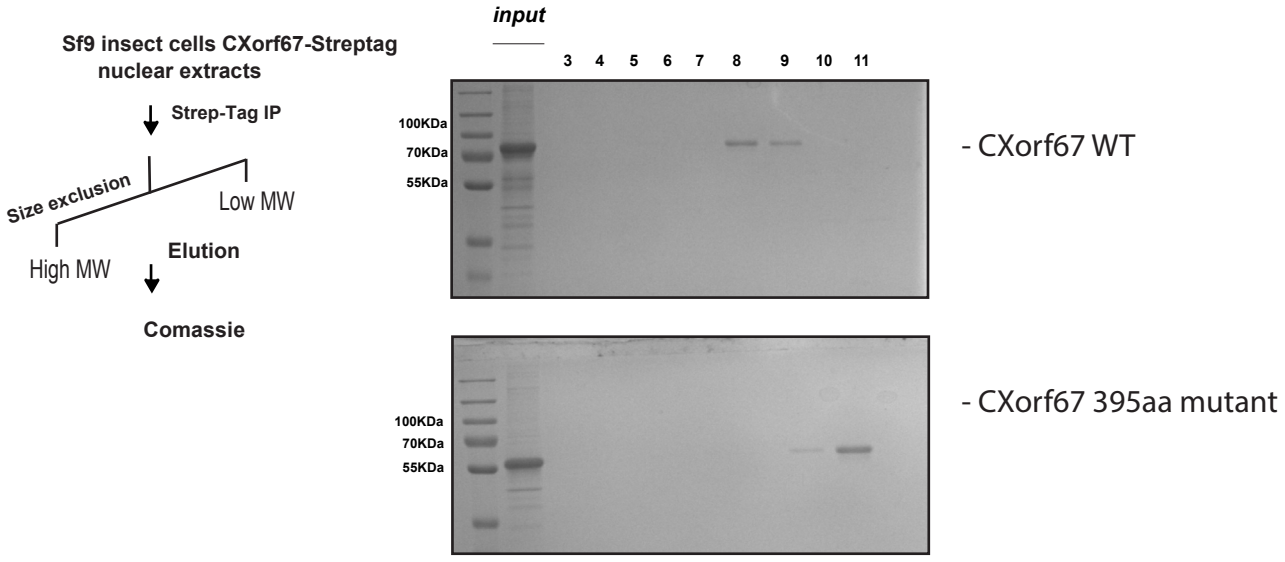
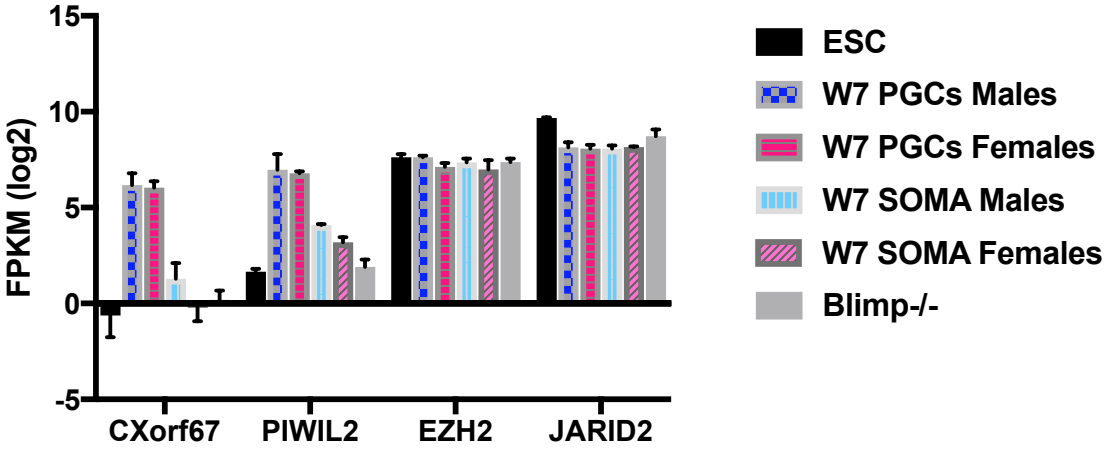
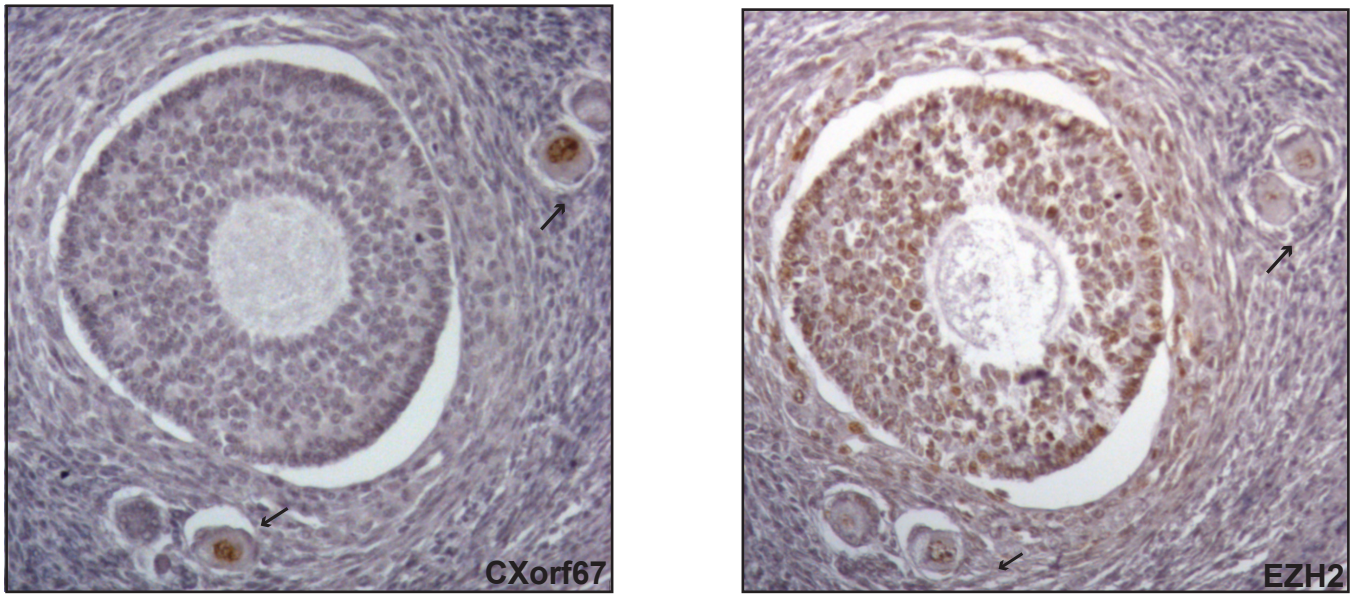


Fig. S1 Ragazzini et al

A



B



C

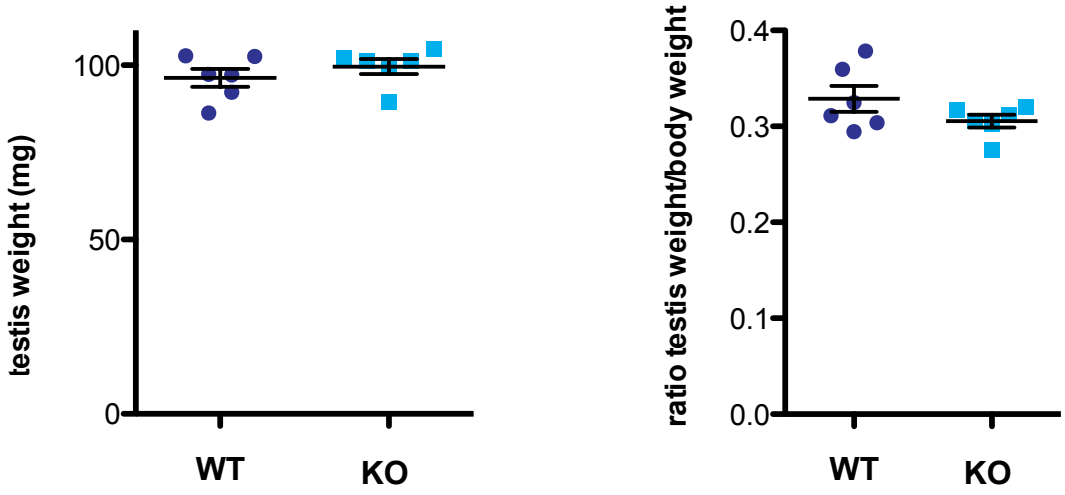


Fig. S2 Ragazzini et al

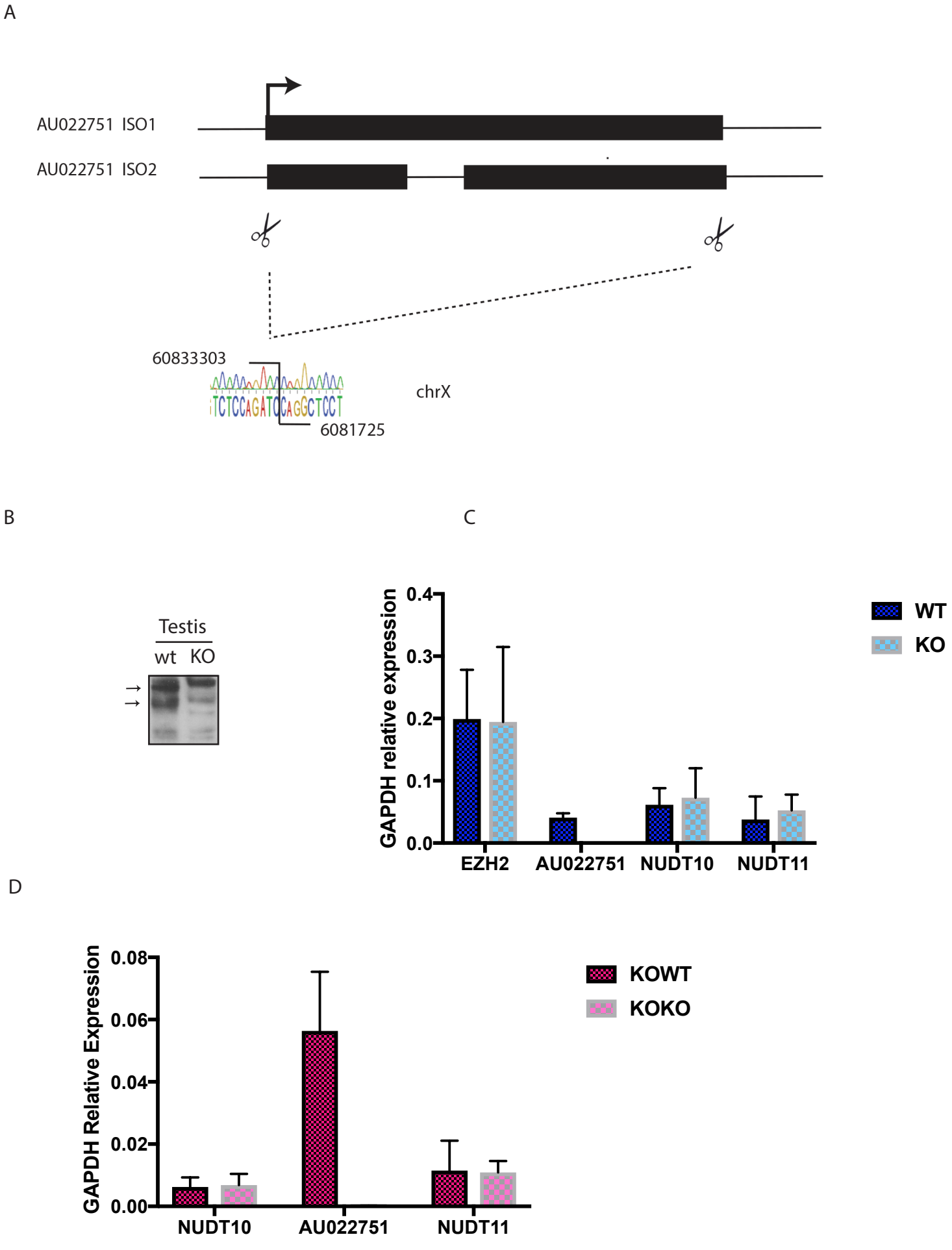


Fig. S3 Ragazzini et al

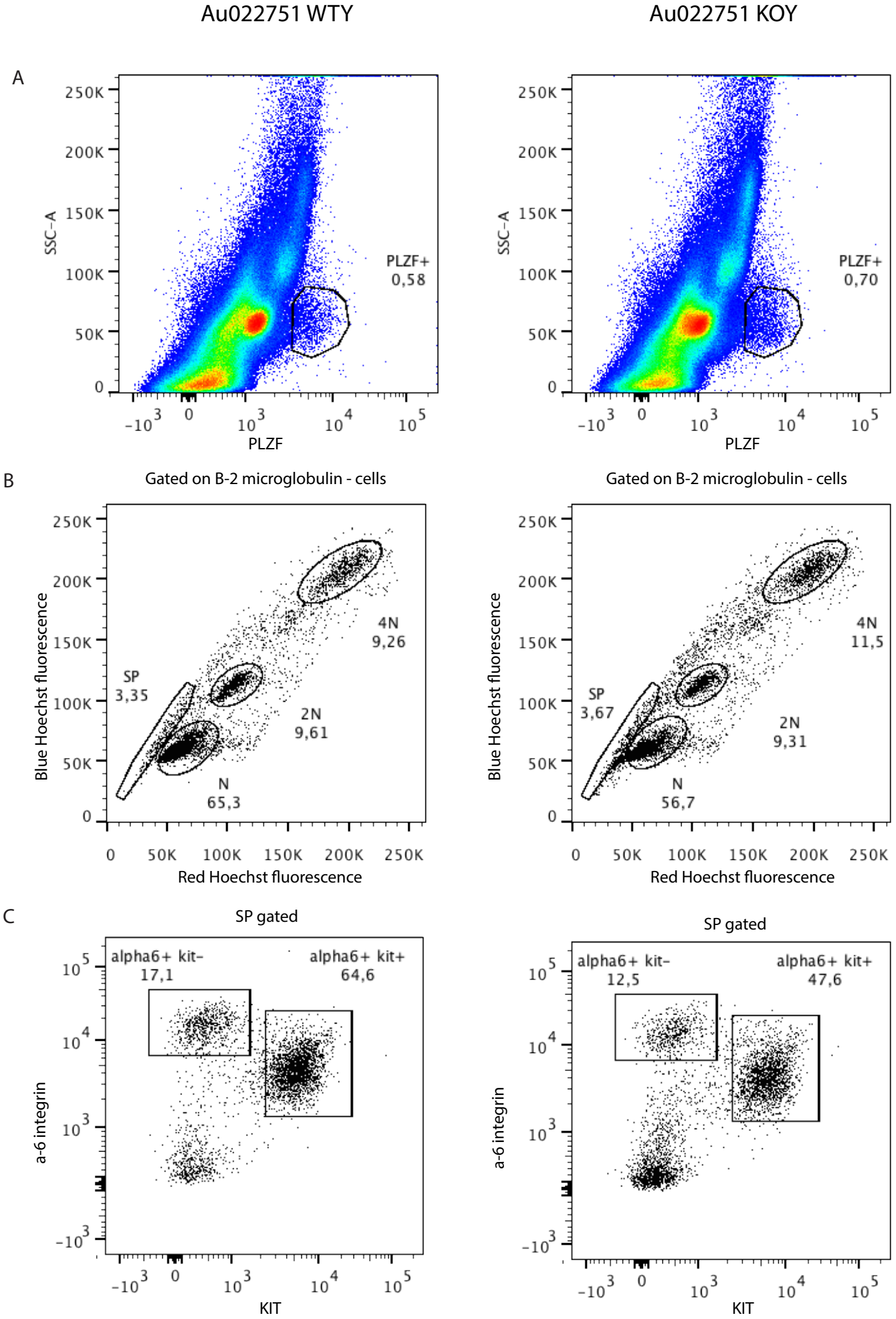
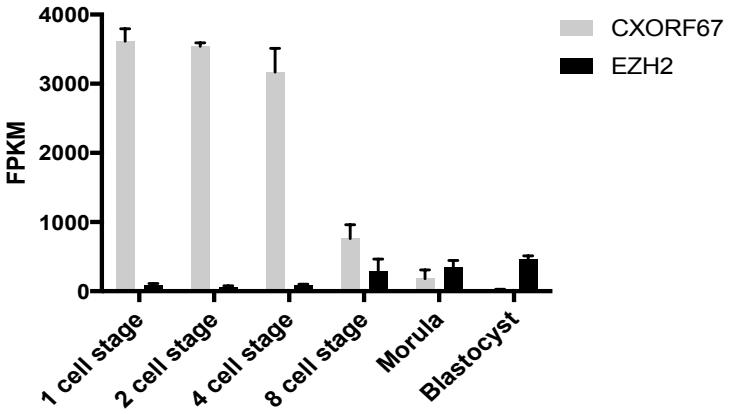
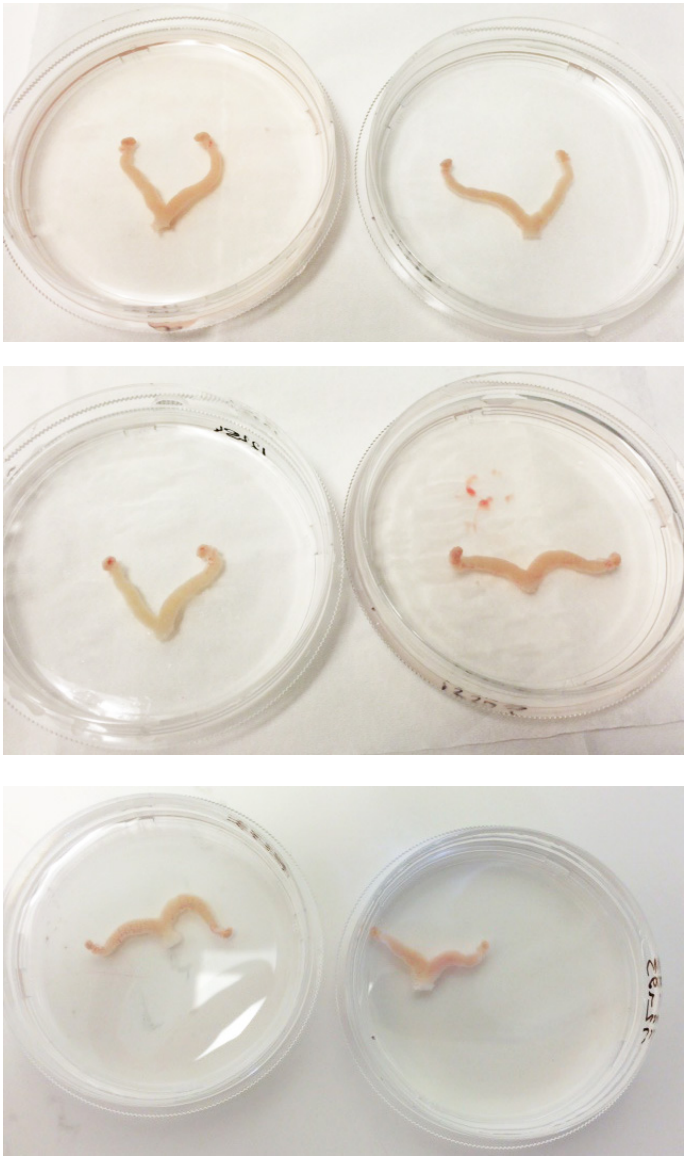


Fig. S4 Ragazzini et al

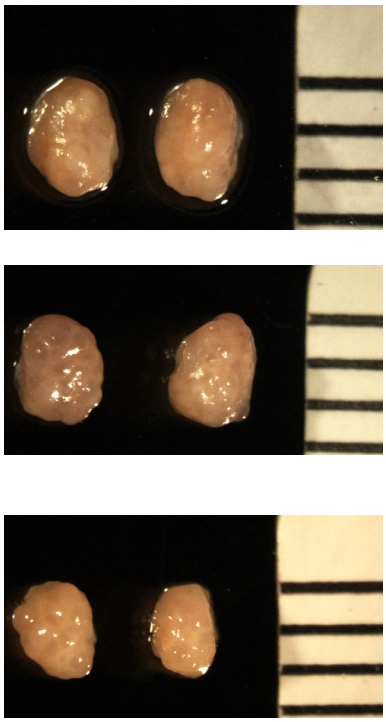
A



B



C



D

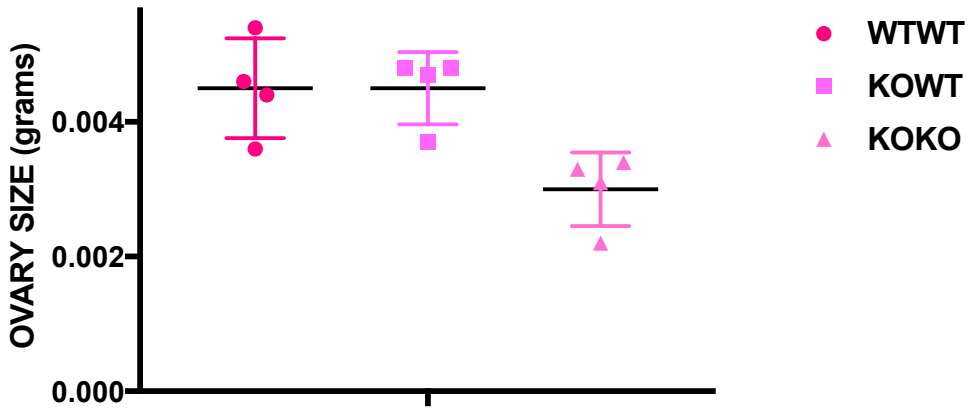


Fig. S5 Ragazzini et al

MATERIALS AND METHODS

Cloning

mAU022751 cDNA clone was obtained from the ORIGENE. hCXorf67 cDNA clone was amplified from HEK293T genomic DNA. hCXorf67 395aa mutant, 420 aa and 450 aa cDNA were generated by PCR. cDNAs were subcloned into pREV retroviral vector (gift form S. Ait-Si-Ali), downstream a 2X-Flag-2X-HA sequence and upstream IRES followed by CD25 cDNA or in pCMV4-HA.

Cell lines

T-Rex 293 cells (Invitrogen) were grown according to the manufacturer's instructions. First, 5XGal4RE-tk-Luc-Neo plasmid was stably integrated into the cells and selected by G418. Subsequently, the selected clone was generated by stable transfection of pCDNA4-T0-Gal4-hCXorf67 WT and 395AA mutant. Single clones were selected with zeocin (300 μg/ml) and screened for the expression of Gal4-Jarid2 48 hours after doxycycline induction at 1μg/ml (Sigma). Stable clones of HEK 293T cell over-expressing pCMV4-HA-Flag-CXorf67 WT and deletion mutant (a.a. 1-395) selected with G418 and screened for the expression of the Flag-tagged proteins by WB.

HeLa-S3 cells (gift form S. Ait-Si-Ali) were grown according to the manufacturer's instructions. pREV retroviruses are produced by transfecting of 293T Phoenix cell line (gift form S. Ait-Si-Ali) and collecting supernatant after 60 hours. HeLa S3 cells were infected by incubation with viral supernatants for 3h at 37°C. Infected cells were then selected by FACS sorting using an anti-CD25-FITC conjugated antibody and following manufacturer instructions (BD Biosciences 553866). Expression of the recombinant proteins were assessed by WB analysis of nuclear extracts.

Mice lines

EZH1- and EZH2-flag knock-in mice generation

Knock-in mice were generated by homologous recombination at the Institute Clinique de la Souris (ICS). Targeted ES cells were screened by PCR and injected into blastocysts. After germline transmission, deletion of the neomycin cassette was achieved by breeding with a CMV-Cre transgenic line (ICS).

AU022751 knockout mouse model generation

Mice were hosted on a 12-h light/12-h dark cycle with free access to food and water in the pathogen-free Animal Care Facility of the Institut Curie (agreement under evaluation). All experimentation was

approved by the Institut Curie Animal Care and Use Committee and adhered to European and national regulation for the protection of vertebrate animals used for experimental and other scientific purposes (directives 86/609 and 2010/63). For tissue and collection, mice were euthanized by cervical dislocation. The *AU022751* mouse line was derived by CRISPR/Cas9 engineering of a 1.5-kb deletion spanning the AU022751 locus (Supplementary Figure) in embryos at the one-cell stage, according to published protocols (Wang H et al, 2013). Briefly, the *Cas9* DNA sequence was amplified by PCR from the pX330 plasmid and cloned into the T7-promoter-containing pCR2.1-XL plasmid. After plasmid linearization and phenol-chloroform purification, *in vitro* transcription was performed using the mMESSAGE mMACHINE T7 ULTRA kit (Life Technologies). The *Cas9* mRNA was then purified with the RNeasy Mini kit (Qiagen). The two deletion-promoting sgRNAs (Supplementary Table) were amplified by PCR with primers containing a 5′ T7 promoter sequence. After gel purification, the T7-sgRNA PCR products were used as the template for *in vitro* transcription with the MEGAshortscript T7 kit (Life Technologies) and the products were purified using the MEGAclear kit (Life Technologies). *Cas9* mRNA and the sgRNAs were eluted in DEPC-treated RNase-free water, and their quality was assessed by electrophoresis on a denaturing agarose gel. Zygote injection with the CRISPR/Cas9 system was performed by the Transgenesis Platform of the Institut Curie. Eight-week-old superovulated C57B6N females were mated to males of the same background. Cytoplasmic injection with *Cas9* mRNA and sgRNAs (at 100 ng/µl and 50 ng/µl, respectively) was performed in zygotes collected in M2 medium (Sigma) at E0.5, with well-recognized pronuclei. Injected embryos were cultured in M16 medium (Sigma) at 37 °C under 5% CO2, until transfer at the one-cell stage the same day or at the two-cell stage the following day to the infudibulum of the oviduct of a pseudogestant female at E0.5. Twenty-five to 30 embryos were transferred per female. Of the 13 pups generated, 8 carried at least one modified allele. Two founders (N0) carrying the expected 1.5-kb deletion were selected (female 10992 and male 10989). After the absence of *in silico*–predicted off-target mutations was verified by Sanger sequencing, these two founders were bred with wild-type C57B6N mice. Two additional backcrosses were performed to segregate out undesired genetic events, following a systematic breeding scheme of crossing *AU022751* heterozygous females with wild-type C55B6N males to promote transmission of the deletion. Cohorts of female and male mice were then mated to study maternal- and paternal- knockout progeny.

Luciferase Assay

Luciferase reporter activities were measured on whole cell lysates using the Luciferase Assay System (Promega, #E15020) and Fluostar Optima BMG labtech luminometer. All experiments were done in biological and technical triplicates and normalized for protein concentration (Bradford).

Histological sections and immunofluorescence microscopy

For histological sections, testis and ovary from either human patients from Curie Institute Pathology Platform or mice were dissected, fixed for 6h in 4% paraformaldheide (Sigma) and washed with 70% ethanol according to pathology platoform standard protocols. Organs were paraffin-embedded, sectioned (8 μm) and stained with Hematoxylin using standard protocols.

For cryosections, testes and ovaries from adult mice (6 months old males; 2,5 and 5 months old females) were dissected, fixed overnight in 4% paraformaldehyde at 4°C, washed in 1X PBS, followed by two consecutive overnight incubations in 15% and 30% sucrose at 4°C respectively. Testes were embedded in O.C.T. compound (Tissue-Tek), 8-10 μm thick sections were cut and spotted onto Superfrost Plus slides (Thermo Fisher Scientific). For immunofluorescence detection, slides were brought to room temperature (RT), and then blocked and permeabilized for 1h (10% horse serum, 3% BSA and 0.2% Triton). Sections were incubated with primary antibodies (Supplementary table) at 4°C overnight, followed by 1X PBS washes and 2h incubation with Alexa Fluor-conjugated secondary antibodies. Slides were mounted with ProLong Gold Antifade with DAPI (Molecular Probes).

Immunomagnetic, flow cell sorting and flow cytometry analysis of mice testis cell populations

Testicular single-cell suspensions were prepared from 2–3-months-oldfrom WT and KO mice as described previously (Lassalle et al., 2004). The albuginea was removed and the seminiferous tubules were dissociated using enzymatic digestion by collagenase type I at 100 U/ml for 25 minutes at 32°C in Hanks' balanced salt solution (HBSS) supplemented with 20 mM HEPES pH 7.2, 1.2 mM $MgSO_4 7H_2O$, 1.3 mM $CaCl_2 2H_2O$, 6.6 mM sodium pyruvate, 0.05% lactate. Next, a filtration step was performed with a 40 μm nylon mesh to discard the interstitial cells. After HBSS wash, tubules were further incubated in Cell Dissociation Buffer (Invitrogen) for 25 minutes at 32°C. The resulting whole cell suspension was successively filtered through a 40 μm nylon mesh and through a 20 μm nylon mesh to remove cell clumps. After an HBSS wash, the cell pellet was resuspended in incubation buffer (HBSS supplemented with 20 mM HEPES pH 7.2, 1.2 mM $MgSO_4 7H_2O$, 1.3 mM $CaCl_2 2H_2O$, 6.6 mM sodium pyruvate, 0.05% lactate, glutamine and 1% fetal calf serum). Cell concentrations were estimated using Trypan Blue staining (>95% viable cells).

The immunomagnetic selection of $\alpha 6^+$ cells was performed using anti-alpha-6 integrin-PE (GoH3) (BD Pharmingen) and anti-PE microbeads (Miltenyi Biotec) according to the manufacturer's protocol. Hoechst staining (5 microglobulin 2/ml) of the cell suspensions was performed as described previously (Lassalle et al., 2004). Cells were then labelled with β2m-FITC (Santa Cruz) and anti-CD117-APC (2B8) antibodies (BD Pharmingen). Propidium iodide (Sigma) was added before cell sorting to exclude dead cells. Analyses and cell sorting were respectively performed on LSR II and

ARIA flow cytometers (Becton Dickinson).

Mice sperm quality test

Adult male mice were euthanized by dislocation and cauda epididymis has been collected post-mortem carefully removing fat pad. Epididimis has been opened and sperm released in IVF media (Vitrolife). 1:100 sperm dilution is loaded in Ivos (Hamilton Thorne machine) and sperm parameters are evaluated by Remote Capture software.

Mice fertility evaluation

6 weeks old wild type and AU022751 knockout females (N=14 each genotype) are subjected to a continuous mating study. One WT female and one KO female mice are housed with an adult breeder male tested before. Cages were monitored daily; pup numbers and litters was constantly registered.

Follicles counting

Sections have been prepared as described above. Follicles number have been counted from at least 15 sections each organ/genotype (N=4 each genotype). Frozen sections have been stained with VASA/dsRED antibody (cytoplasmic germ cells marker) and counterstained with DAPI. Red spots corresponding to germ cells are counted and cell shape visualized by Leica Epifluorescence microscope in order to distinguish different follicle shapes.

Nuclear Extract and Immunoprecipitation from cell extracts

For nuclear extract preparation cells were incubated with buffer A (10mM Hepes pH 7.9, 2.5mM MgCl2, 0.25M sucrose, 0.1% NP40, 0.5mM DTT, 1mM PSMF) for 10 min on ice, centrifuged at 8000 rpm for 10 min, resuspended in buffer B (25mM Hepes pH 7.9, 1.5mM MgCl2, 700 mM NaCl, 0.5mM DTT, 0.1 mM EDTA, 20% glycerol), sonicated and centrifuged at 14000 rpm 15min.

For immunoprecipitation 1mg of nuclear extract was incubated with 250 microl of Flag M2 Beads (SIGMA-ALDRICH-A4596). Then beads were washed three times with BC250 (50mM Tris pH7.9, 250mM KCl, 2mM EDTA, 10% Glycerol, and protease inhibitors), and eluted with 0.2 M glycine pH 2.6.

Nuclei isolation and extraction from tissues

Mice tissues are rapidly extracted in PBS 1X and homogenized with dounce (cut into small pieces with scissors, then 6x up-down with loose and 4x with tight pestle) adding sucrose 2.2M solution (Sucrose 2.2 M, Hepes 1M pH7.6, KCl 3M, EDTA 0.5M, Spermine 0.1M, Spermidine 1M and complete set of protease inhibitors added last minute). Homogenized is mixed and added very gently

onto Sucrose 2.05 (Sucrose 2.05 M, Hepes 1M pH7.6, KCl 3M, EDTA 0.5M, Spermine 0.1M, Spermidine 1M and complete set of protease inhibitors added last minute) in ultracentrifuge Beckemann tubes. Spin 45(-60) min 24k at 1°C in SW28. Supernatant was removed and nuclei at the bottom of the tube resuspended in an equal volume of Nuclear Lysis Buffer (Hepes 1M, pH7.6, KCl 3M, EDTA 0.5M, Glycerol 87%, Spermine 0.1M, Spermidine 1M, NaF 0.5M, Na_2VO_4 0.5M, $ZnSO_4$ 50mM and complete set of protease inhibitors added last minute). Samples were snap-frozen in liquid nitrogen.

Mass Spectrometry analysis

Mass spectrometry analysis has been performed as described in (Baymaz HI et al, 2014) starting from tissue nuclear extracts or cellular extracts.

Size Exclusion Chromatography

Recombinant Purified Step-tagged CXorf67 WT and mutant proteins profile stored in BC100 (50mM Tris pH8,0 100mM KCl, 1 mM EDTA, 10% glycerol and protease inhibitors complete set) and clarified by high-speed centrifugation. Samples were loaded on a size exclusion chromatography (Sephadex Increase 200, Column Volume 2.4ml, GE healthcare). Elution was analyzed by SDS-PAGE coomassie staining.

Recombinant proteins purification, baculoviruses.

Recombinant EZH2, SUZ12, EED, RbAp48, JARID2 1-530, CXorf67 full length and mutant proteins were produced in SF-9 insect cells after infection with the corresponding baculoviruses as described previously (Li et al., 2010; Margueron et al., 2009; Margueron et al., 2008). Lysates were resuspended in BC300, sonicated and clarified by centrifugation before incubation with either with Flag-beads (M2-beads, SIGMA_ALDRICH 4596) and eluted with Flag peptide or Streptactin-50% sepharose suspension (IBA, 2-1201-010) and eluted with 2,5mM Desthiobiotin in BC300.

hCXorf67 and mutant form baculoviruses were produced accordingly to Bac-to-Bac Baculovirus Expression Systems (Invitrogen) after cDNA cloning into pFASTbac vectors.

KMT assay

KMT assay were performed as described previously (Margueron et al., 2009). Briefly, the reaction was performed with 200ng of PRC2 alone or in presence of either CXorf67, CXorf67 mutant or JARID2, 1 µg of substrates, 4mM DTT in methylation reaction buffer (50mM Tris-HCl pH 8.5, 2.5mM MgCl2) in presence of 3H-SAM and incubated at 30°C for 15 min or 30 min. Reactions were stopped by boiling 5 min in SDS Laemmli buffer, run on acrylamide gels and transferred into PVDF

membranes. Nucleosomes were generated by salt dialysis.

RT-qPCR from cells or tissues

Total RNA was isolated using the Rneasy Mini Kit (Qiagen). cDNA was synthetized 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 three biological independent experiments were performed for each assay and negative controls RT are always included. Primers sequences are provided below.

ChIP

ChIP are performed as previously described (Margueron et al, 2008). Cell confluence and amount of starting material are kept constant by plating defined number of cells two day before cross-linking. Quantification as done as previously described for the RT-qPCR. Primers sequences are provided in the Supplementary Table.

Antibodies

Antibodies against EZH1/2, SUZ12 and EED have been previously described (Margueron et al, 2008); total H3 (39163) has been purchased from Active Motif; H3K9me2 (Ab1220) and H3K27Ac(Ab 4729) have been purchased from Abcam; H3K27me2 (D18C8), H3K27me3 (C36B11), H3K4me3(C42D8) antibodies from Ozyme (Cell signaling Technologies). CXorf67 (HPA006128) and mAb Flag-M2 (F1804) purchased from SIGMA-ALDRICH; mAb GAL4 Antibody DBD (sc-577) from Santa Cruz Biotechnologies; GAL4 (06-262) for ChIP and HDAC-1 (mAb clone 2E10) from Merck-Millipore; Anti-Germ cell-specific antigen antibody, TRA98 (Ab82527) from Abcam. Antibody against AU022751 was raised against the two following synthetic peptides: CAESSRAESDQSSPAG (corresponding to aminoacids residues 91-106) and CAQSAGRNLRPRPRSS (corresponding to aminoacids residues192-206). Rabbit sera were affinity purified on affigel (Biorad) immobilized peptides.

SUPPLEMENTAL INFORMATION

Supplemental Information includes five supplementary figures and one supplementary table.

Table S1. Primers

ChIP	ATGGAAGACGCCAAAAACAT	Portoso et al, 2017
ChIP	GCCTTATGCAGTTGCTCTCC	Portoso et al, 2017
ChIP	GTGTTGGGCGCGTTATTTAT	Portoso et al, 2017
ChIP	TACGGTAGGCTGCGAAATGT	Portoso et al, 2017
RT-qPCR	AACAGCAACTCCCACTCTTC	
RT-qPCR	TGGTCCAGGGTTTCTTACTC	
RT-qPCR	AATACATGTGCAGCTTTCTGTTC	
RT-qPCR	ACGAATTTTGTTGCCCTTTC	
RT-qPCR	TTCCGGAGTTGTACCTTTCG	
RT-qPCR	ACGTAAATTCCAGCCTGTGC	
RT-qPCR	AGAGAGCGAGCCCTAGTGAATGGA	
RT-qPCR	GAGCTCACCTGTGCTTCACAATTCC	
RT-qPCR	ACCGAGGCATGCTCAAGATCACA	
RT-qPCR	TGAGCGGTCTCCTTGGCAACCTTA	

AUTHORS CONTRIBUTIONS

R.R. and R.M. designed the experiments. R.R. performed all cell culture lines establishment, in vitro experiments and in vivo experiments. R.R. designed AU022751 knockout mice line. I.B. and M.V. performed IP and Mass Spectrometry analysis of Ezh1/Ezh2-flag testis and AU022751/CXorf67-Flag HeLa cell lines. M.G. and P.F. sorted spermatogonial stem cells population and kept them in culture. R.R. and R.M. prepared the manuscript. R.P.P and R.R. performed mice ovaries dissection and analysis.

AKNOWLEDGEMENTS

We thank very much Fatima El Marjou and the Mice Transgenic Facility in Curie Institute for the creation of AU022751 mouse model, Joan Barau, Katia Ancelin and Rafael Galupa for valuable inputs on the in vivo part of the project. R.R. fellowship is funded by the Ministère de l'enseignement supérieur et de la recherche (MSER) et la Fondation de la Recherche Médicale (FRM). Work in R.M. was funded by the ERC-stg (REPODDID) and labelization from the ARC (Association pour la Recherche sur le Cancer).

RESULTS-CHAPTER II

INTRODUCTION AND AIM OF THE PROJECT

Far away from the Beadle and Tatum paradigm "One gene-one enzyme", in the past few decades ENCODE project further confirmed that only a small fraction of the human genome encodes proteins and most 60% is composed by processed transcripts that are supposed to lack protein-coding capacity (Djebali et al., 2012). Yet, these non-protein-coding RNA molecules exert important cellular functions. For example, ribosomal RNAs (rRNAs) are essential elements of the translation machinery and small nuclear RNAs (snRNAs) are required for splicing of nascent RNA transcripts. Additionally, various classes of small (around 20–30 nucleotides) noncoding RNAs such as micro (mi)RNAs, small inhibitory (si)RNAs or PIWI interacting (pi)RNAs are well known as gene silencers (Okazaki et al., 2002)A large proportion of detected RNA transcripts are structurally indistinguishable from protein-coding and processed messenger RNAs (mRNAs). They tend to be expressed at a low level and have little protein-coding potential. This subclass of noncoding transcripts of variable length and function is referred to as long noncoding RNAs (lncRNAs).

LncRNAs are defined as longer than 200 nucleotides, and like mRNAs, they are transcribed by RNA polymerase II, capped, spliced and polyadenylated. Yet, in contrast to mRNAs, lincRNAs have little or no protein-coding potential (Guttman et al., 2009; Khalil et al., 2009). Several thousand of lincRNAs have been identified in a variety of genomes (Guttman et al., 2011) raising the question of their biological relevance. Although the functions of most lincRNAs are unknown, there is a growing repertoire of cellular processes that implicate lincRNAs as key regulators: their described functions include X-chromosome inactivation, imprinting, maintenance of pluripotency, and transcriptional regulation (Mercer, Dinger, & Mattick, 2009)

1. LncRNA functions

LncRNAs associated functions have been often illustrated by knockdown or knockout of RNA expression and subsequent phenotypic and molecular analysis. Many lncRNA have been related to developmental processes such as apoptosis or metastasis formation (Gupta et al., 2010). Nevertheless, lncRNAs deletion does not always result in major phenotype but rather in subtle changes like behavioral changes that are important in a normal environment but more difficult to demonstrate (Morris & Mattick, 2014).

RNA is a versatile polymer, prone to perform different roles in a physiological context. At molecular level, they fold into complex 3D-conformations: this enables them to form specific interactions with proteins. Moreover, they can interact with other RNA or DNA molecules via base pairing, even with double-stranded DNA, and form networks with DNA, protein complexes and RNA molecules,

suggesting their large potential as an important player for many biological functions (Geisler & Coller, 2013).

LncRNAs can modulate protein-coding gene expression either in cis- if they control the locus from which they are transcribed or in trans- if they act on another locus (Rinn & Chang, 2012). They can act directly on chromatin template through chromatin modifying complexes and consequently targeting their activity to specific DNA loci. Depending on the type of complex they bind, lncRNAs directed chromatin modifiers could result in gene expression or gene activation (Khalil et al., 2009; Zhao et al., 2010). Yet, lncRNAs could also interfere with transcriptional machinery in various ways. LncRNAs have been proposed to bind RNAPolII directly inhibiting transcription (Mariner et al., 2008) or to hybridize with DNA forming a triple helix that hinders Pre Initiation Complex assembly (Martianov, Ramadass, Barros, Chow, & Akoulitchev, 2007). Also, they are able to hoodwink transcription factor assuming particular conformations able to mimick their DNA binding sites (Kino, Hurt, Ichijo, Nader, & Chrousos, 2010). Finally, they may also control transcription by binding specific transport factors to inhibit the nuclear localization of transcription factors (Willingham et al., 2005).

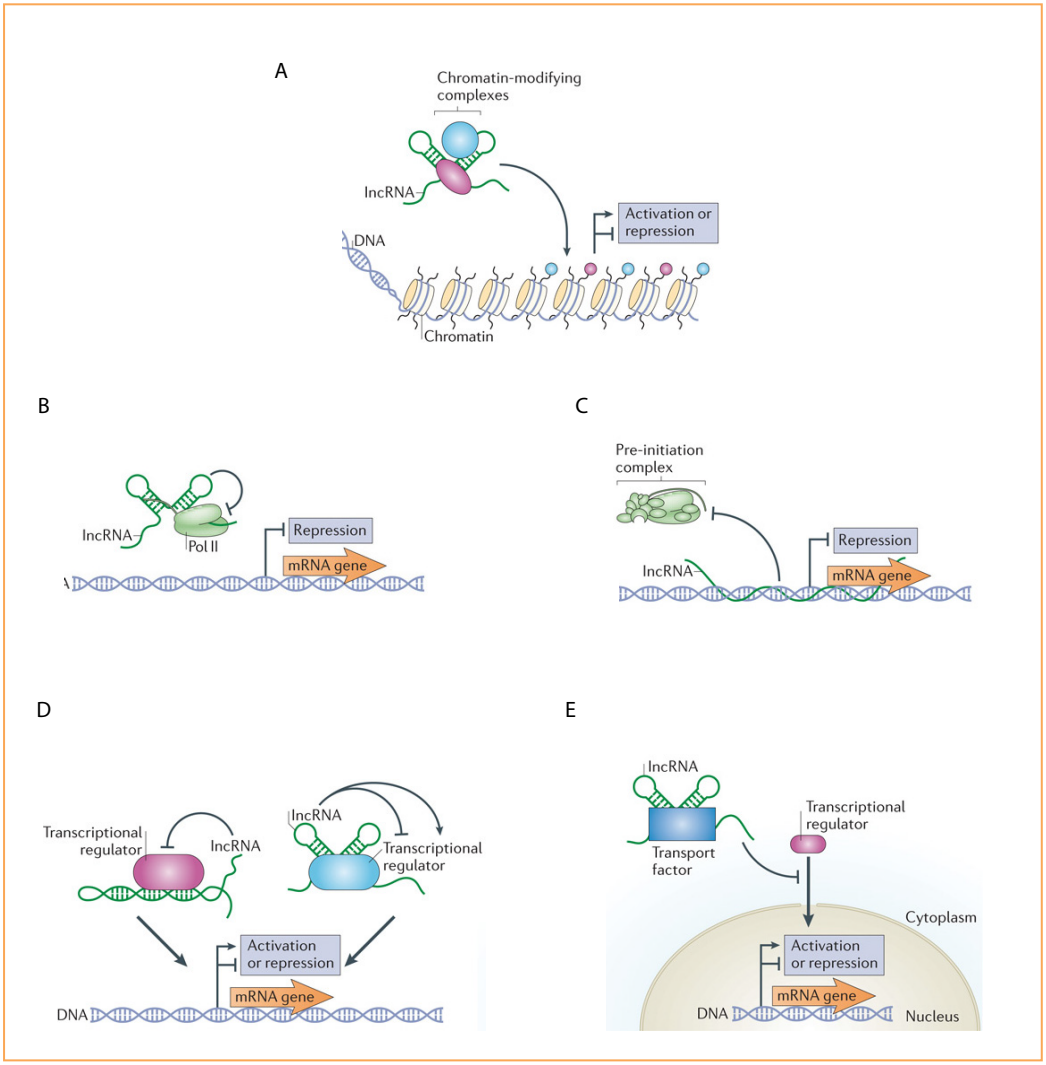


Figure 1. Transcriptional regulation by lncRNAs.

(A) lncRNAs can bind one or more chromatin-modifying complexes and direct their activities to specific target genes, (B) lncRNAs can bind Pol II directly to block transcription, (C) ncRNA-DNA triplex structures can be formed in order to inhibit the assembly of the pre-initiation complex, (D) lncRNAs can fold into structures that mimic DNA-binding sites ort hat generally block or enhance the activity of specific transcription factors and (F) lncRNAs can also regulate gene expression by binding specific transport factors to inhibit the nuclear localization of specific transcription factors (adapted from Geisler and Coller, 2013).

2. LncRNAs and chromatin modifiers: Polycomb case

Prominent examples of histone-modifying complexes interacting with lncRNAs are the two polycomb repressive complexes, PRC1 and, in particular PRC2, which mediates H3K27me3, a histone mark associated with repressed or poised genetic loci.

For example, lncRNA ANRIL is transcribed from the INK4 gene and regulates in cis INK4b-ARF-INK4a. This transcript physically is reported to interact with both Suz12 and Cbx7 consequently recruiting PRC2 and PRC1 thus mediating gene repression (Pasmant, Laurendeau, He, et al., 2007;

Yap, Li, Munoz-Cabello, et al., 2010). PRC2 has been reported to interact with many other lncRNAs involved in various biological processes, yet PRC2-lncRNAs relationship has become a topic of intense debate, more and more controversial.

Imprinting

In mammals, one gene is present in two copies, one inherited from the father and the other one from the mother. Imprinted genes are expressed in a mono-allelic manner depending on the father or mother inheritance (Barton, MA, & Norris, 1984); lncRNAs can regulate transcription in cis and recruite in turn chromatin modifying complexes. Kcnq1ot1 is an RNA polymerase II-encoded, 91 kb-long nuclear transcript and its stability is important for bidirectional silencing of genes in the Kcnq1 domain. Kcnq1ot1 was reported to interact with chromatin and with G9a and PRC2, the histone methyltransferases depositing H3K9- and H3K27, in order to repress Kcnq1 paternal allele in the placental genes (Pandey et al., 2008; Redrup et al., 2009). In another example, the accessory PRC2 cofactor JARID2 presents RNA binding activity in vivo and in vitro towards MEG3, an lncRNA encoded by the imprinted *DLK1–DIO3* locus. This lncRNA is necessary for proper recruitment and ativity of PRC2 at a subset of target genes in pluripotent stem cells (Kaneko, Bonasio, et al., 2014).

X chromosome inactivation

X Inactive Specific Transcript (Xist) is a 17 Kb non-coding RNA and a key player in X chromosome inactivation (XCI). In this process one copy of X chromosome is transcriptionally silent in female mammalian cells ensuring dosage compensation on X chromosome (Lyon MF, 1961). XCI is promoted by the action of non-coding RNA Xist that spreads in cis and triggers the recruitment of chromatin modifying complexes including Polycomb proteins (Plath et al., 2003; Zhao, Sun, Erwin, Song, & Lee, 2008).

During X inactivation, PRC2 has been reported to colocalize with Xist non coding RNA on the inactive X (Plath et al., 2003; Silva et al., 2003). A conserved 2 stem-loop structure on Xist RNA called RepA has been suggested to interact with PRC2 via Suz12 (Maenner et al., 2010; Zhao et al., 2008). Nonetheless, deletions in the RepA region still induces H3K27me3 mark (Kohlmaier et al., 2004). Furthermore, immediately after inducing Xist expression in mouse embryonic stem cells, it has been observed by super resolution microscopy that silenced genes do not localize with H3K27me3. These results separate the initial silencing and PRC2 recruitment (Cerase et al., 2013).

In the past few years, RNA immunoprecipitation techniques coupled to mass spectrometry aimed to characterize Xist interacting proteins. Factors directly linked to PRC2 complexes were not retrieved (Chu et al, 2015; Minajigi et al, 2015). SMRT and HDAC associated repressor protein (SHARP) has been pulled-down from Xist and associated to transcriptional silencing. This factor was proposed to

help Xist-mediated recruitment of the polycomb repressive complex 2 (PRC2) across the X chromosome (Mchugh et al., 2015).

Interestingly, PRC1 subunit Ring1B was identified among Xist binding partners in one of this screen (Minajigi et al., 2015). Recently, non-canonical PRC1 containing subunits Polycomb group RING finger 3/5 (PCGF3/5) have been proposed to initiate recruitment of both PRC1 and PRC2 in response to Xist RNA expression. Pcgf3/5 knockout results in female embryos lethality and eliminates genes repression mediated by Xist (Almeida, Pintacuda, Masui, Koseki, Gdula, Cerase, Brown, Mould, Innocent, et al., 2017).

HOTAIR: a paradigm for PRC2 and lncRNA interaction

HOX genes are known for their essential roles in defining identities of body segments in fruit flies. In mammals, 39 HOX genes are encoded across 4 loci (from A to D) on different chromosomes. The position of each gene in the cluster reflects its spatial and temporal expression along proximal-distal and anterior-posterior axis. Their gene expression maintenance is supposed to result from the balanced action between PcG and trithorax (Trx) proteins. Several lncRNAs are located within the HOX clusters and exhibit similar spatial and temporal expression pattern to the coding genes (Gaell Mainguy, Koster, Woltering1, Jansen, & Durston, 2007; Rinn et al., 2007).

Hox Transcript Antisense Intergenic RNA (Hotair) is a 2148-base long lncRNA that is transcribed from the intergenic region between Hoxc11 and Hoxc12 within the HoxC gene cluster. HOTAIR is expressed in posterior and distal sites from development to adulthood. Hotair was suggested to play a repressive role in trans at the HoxD locus by interacting with PRC2 in human cells as it is shown in Fig.2 (Rinn et al., 2007).

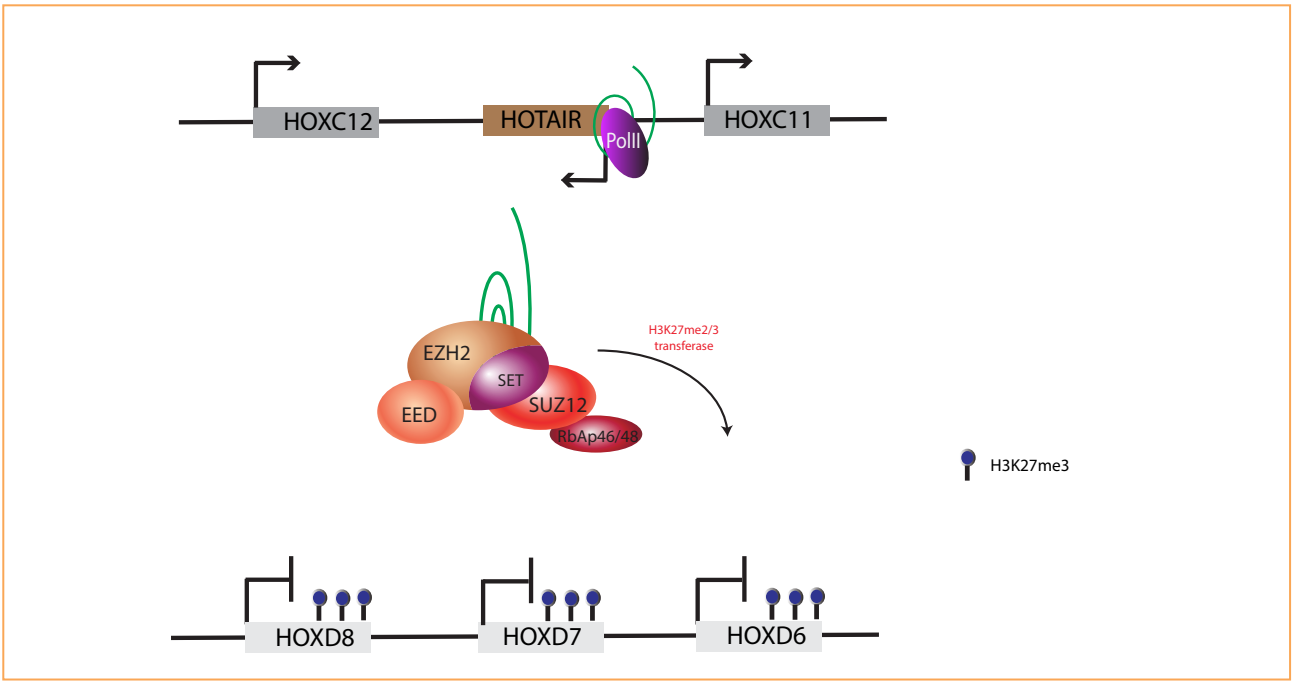


Figure 2. lncRNA Hotair transcription

Schematic representation of human lncRNA HOTAIR transcription and function following J. Rinn's findings (Rinn et al., 2007)

Chromatin Isolation by RNA Purification indicate that, upon overexpression, this lncRNA preferentially occupies a GA-rich DNA motif to nucleate broad domains of Polycomb occupancy and histone H3 lysine 27 trimethylation (Chu, Qu, Zhong, Artandi, & Chang, 2011). It has been demonstrated that HOTAIR RNA adopts a defined secondary structure at its 5′ end, which is proposed to be critical for its interaction with the PRC2 in vitro (Tsai et al., 2010). The 3' region HOTAIR would also interacts with another repressive chromatin modifier, the LSD1/coREST/REST complex, catalyzing H3K4me2 demethylation. Hence, it has been proposed to act as a scaffold coordinating the recruitment of both the PRC2 and LSD1/coREST/REST complexes onto chromatin (Tsai et al., 2010). Interestingly, HOTAIR expression is increased in primary breast tumours and metastases, and is considered as a predictor of metastatic evolution. Enforced expression of HOTAIR in epithelial cancer cells is reported to induce genome-wide re-targeting of PRC2, leading to altered histone H3 lysine 27 methylation, gene repression, and increased cancer invasiveness and metastasis in a PRC2-depedent manner. Conversely, loss of HOTAIR can inhibit cancer invasiveness, particularly in cells with excessive PRC2 activity (Gupta et al., 2010).

Looking at the possible HOTAIR physiological roles in a mouse model, the scenario further complicates. In mouse genome, Hotair sequence is poorly conserved. In order to assess Hotair regulation during development, the entire HoxC locus, containing Hotair, was deleted. Surprisingly, HoxD target genes expression patterns, transcriptional activity and H3K27me3 decoration are not

affected (Schorderet & Duboule, 2011). It was initially speculated that Hotair could play a different biological role from humans to mice since its poor conservation or rather that compensating mechanisms might take place in its absence. Since the HoxC deletion considered removes not only 8 Hox genes but also microRNAs and lncRNAs, it could have masked a potential alteration due to the absence of Hotair alone. In order to test this hypothesis, the Hotair gene was specifically targeted by deletion. A deletion in the two main exons brought to malformations in the wrist and homeotic transformation of the spine, either in the lumbar or post-sacral region. This was associated to a derepression of HoxD targets. No difference was noticed in HoxC genes, further supporting Hotair role in trans (L. Li et al., 2013). Recently, another group performed two additional deletion of Hotair in mouse and besides subtle alteration of one caudal vertebra, they detected no apparent defects in wrist and the spine, no homeotic transformations (Lai et al., 2015). In order to clarify this issue, Hotair deletion previously described was reconsidered during mouse embryonic development. Strikingly, no obvious phenotype was reported. Moreover, transcriptome analysis revealed that HoxD expression is not affected upon Hotair knockout in any embryonic tissue, whereas just subtle changes were detected in Hoxc11 and Hoxc12 expression, Hotair's flanking genes. These last studies support a little effect mediated by Hotair during mouse development and together with other recent reports about promiscuous binding of PRC2 to long non coding RNAs, put Hotair-PRC2 paradigm into discussion (Amândio, Necsulea, Joye, Mascrez, & Duboule, 2016) .

3. PRC2 and RNA: is it a preferential or rather a promiscuous binding?

The above-mentioned models suggest that PRC2 binds preferentially to some RNAs, but recently PRC2 has also been reported to bind to RNAs in a non-specific manner. A first indication that PRC2 poorly discriminate RNA came from the report that short RNAs (50-200bp) transcribed from Polycomb target genes in mES and PolII associated globally interact with PRC2. These transcripts were proposed to associate with Suz12 thanks to a stem-loop conformation, and to cause gene repression in cis (Kanhere et al., 2010). Subsequently, Davidovich and colleagues as well as Kaneko and others confirmed this promiscuous binding in vitro using different RNAs and in vivo as, comparing genome-wide data from different mouse cell lines (Davidovich, Zheng, Goodrich, & Cech, 2013; Kaneko, Son, Shen, Reinberg, & Bonasio, 2013). It had been previously noticed that RNA immunoprecipitation experiments pulled-down many lncRNAs and mRNAs together with PRC2 (Khalil et al., 2009; Zhao et al., 2010). In this regard, it is worthy to mention that RNA immunoprecipitation stringency is not extremely high since it does not discriminate between direct or indirect binding and lysis conditions could create artificial protein-RNA interaction. Ultraviolet Crosslinking followed by sequencing has been employed to detect EZH2-bound RNA more specifically. In mESC, PRC2 is detected on the active transcribed genes and report to bind to nascent transcripts. This correlates with H3K27me3 decrease at these promoters and it was proposed as

mechanism by which PRC2 senses active transcription state (Kaneko, Son, Bonasio, Shen, & Reinberg, 2014) .

A more recent report confirmed PRC2 promiscuously binding to nascent RNA of all transcribed genes, but observed that the binding is enriched at exon-intron boundaries and at 3' untranslated regions and that Suz12 subunit is sufficient for RNA binding. Moreover, PRC2 binding to RNA and chromatin is reported to be mutually exclusive (Beltran et al., 2016). This model differs from the one of Kaneko and others in which PRC2 still resides at chromatin even if RNA inhibits its catalytic properties and from the one of Davidovich in which RNA binding recruits PRC2 to promoters that escapes silencing. Finally, PRC2 was reported to specifically sense four G stretch binding both on RNA single strand and on chromatin (X. Wang et al., 2017).

4. lncRNA HOTAIR as a paradigm to study lncRNAs and PRC2 interplay.

Globally, promiscuous binding models do not rule out that some lncRNAs function to specifically target PRC2 to chromatin. Perhaps, the targeting ability depends on the particular genomic localization of the lncRNAs rather than bearing a real preference for PRC2. Indeed, PRC2 seems to specifically sense four G stretch binding both on RNA and on chromatin. Apart from the example of Xist, the specific role of others non-coding RNA in targeting chromatin modifiers remains unclear. As previously mentioned, lncRNA HOTAIR has been associated with gene repression in trans and a direct interaction has been proven in vitro with PRC2 complex (Rinn et al., 2007; Tsai et al., 2010). HOTAIR role in vivo appears more controversial: mouse phenotype derived by different deletions is not clear, while HOTAIR deregulation in cancer cells is reported to drive PRC2 mis-targeting possibly contributing to tumoral progression.

We aimed to set up a study in order to minutely dissect the molecular interplay between PRC2 and HOTAIR. *HOTAIR* is a 2,2Kb lncRNA and thus represents an easy to handle lncRNA, in contrast to Xist, for genetic manipulation and artificial recruitment. First of all, we aimed to characterize transcriptome of breast cancer cell lines overexpressing HOTAIR both in presence and on absence of PRC2. Interestingly, moderate changes affect transcriptome of this tumoral cell line overexpressing HOTAIR, mostly in a PRC2 independent manner. Therefore we set up an artificial tethering system expressing inducible HOTAIR in order to characterize HOTAIR effects on transcription and the possible PRC2 recruitment to a transgene. We observed that artificial recruitment of HOTAIR sense transcript results in transgene repression associated with subtle gain and loss of active and repressive chromatin at the transgene, respectively. In contrast, no transcriptional repression occurs once recruiting the antisense transcript. Since recent advances in the field put specificity of PRC2-lncRNAs interactions into discussion, we aimed to clarify that in vitro and in vivo context using HOTAIR antisense transcript as a control. Sucrose gradient experiments and Electrophoretic Mobility Shift Assay (EMSA) consistently showed that PRC2 binds RNA with high affinity and low specificity. In

vivo RNA Immuno Precipitation (RIP) of EZH2 pulled down both HOTAIR sense and antisense transcripts confirming the low specificity of this interaction. Finally, taking advantage of CRISPR/Cas9 technology, we performed knockout of two PRC2 subunits in our artificial tethering system lines. While H3K27me3 levels are abolished, we noticed the same level of transcriptional repression as it occurs in a PRC2 wild type context, leading to the hypothesis that alternative mechanisms drives to transcriptional repression.

I personally contributed to this work by performing validations of the artificial tethering systems: I performed part of the cloning and most of stable cell lines generation. I also contributed to the analysis of different chromatin marks by ChIP-qPCR. I took part to the sucrose gradient experiments in order to assess HOTAIR-PRC2 binding specificity and I set up RIP conditions to specifically immunoprecipitate EZH2 subunit. Last, I performed ChIP-qPCR experiment to assess H3K27me3 upon PRC2 knockout in our tethering model cell line.

Published online: February 6, 2017

Article

SOURCE DATA | TRANSPARENT PROCESS | OPEN ACCESS

THE EMBO JOURNAL

PRC2 is dispensable for *HOTAIR*-mediated transcriptional repression

Manuela Portoso[1,2], Roberta Ragazzini[1,2], Živa Brenčič[1,2], Arianna Moiani[1,2], Audrey Michaud[1,2], Ivaylo Vassilev[1,2], Michel Wassef[1,2], Nicolas Servant[1,3], Bruno Sargueil[4] & Raphaël Margueron[1,2,*]

Abstract

Long non-coding RNAs (lncRNAs) play diverse roles in physiological and pathological processes. Several lncRNAs have been suggested to modulate gene expression by guiding chromatin-modifying complexes to specific sites in the genome. However, besides the example of Xist, clear-cut evidence demonstrating this novel mode of regulation remains sparse. Here, we focus on *HOTAIR*, a lncRNA that is overexpressed in several tumor types and previously proposed to play a key role in gene silencing through direct recruitment of Polycomb Repressive Complex 2 (PRC2) to defined genomic loci. Using genetic tools and a novel RNA-tethering system, we investigated the interplay between *HOTAIR* and PRC2 in gene silencing. Surprisingly, we observed that forced overexpression of *HOTAIR* in breast cancer cells leads to subtle transcriptomic changes that appear to be independent of PRC2. Mechanistically, we found that artificial tethering of *HOTAIR* to chromatin causes transcriptional repression, but that this effect does not require PRC2. Instead, PRC2 recruitment appears to be a consequence of gene silencing. We propose that PRC2 binding to RNA might serve functions other than chromatin targeting.

Keywords chromatin; lincRNA; Polycomb; transcription
Subject Categories Chromatin, Epigenetics, Genomics & Functional Genomics; RNA Biology; Transcription
DOI 10.15252/embj.201695335 | Received 25 July 2016 | Revised 23 December 2016 | Accepted 5 January 2017 | Published online 6 February 2017
The EMBO Journal (2017) 36: 981–994

See also: **MR Blanco & M Guttman** (April 2017)

Introduction

Polycomb group proteins (PcG) are highly conserved factors that mainly act in the context of multi-subunit nuclear complexes to maintain transcriptional repression. Their disruption interferes with various processes, ranging from genomic imprinting to cell identity and differentiation. The functions of PcG proteins rely on the regulation of chromatin structure, either through histone modifications or through chromatin compaction (Simon & Kingston, 2009). In *Drosophila*, four PcG complexes have been identified, while in mammals, only two complexes are well characterized so far: Polycomb Repressive Complex 2 (PRC2) and Polycomb Repressive Complex 1 (PRC1). The PRC2 is responsible for histone H3 lysine 27 (H3K27) di- and tri-methylation (Margueron & Reinberg, 2011).

Although our understanding of how PRC2 contacts chromatin has improved, how it is specifically recruited to defined genomic loci is still only partially understood. The core PRC2 has no known sequence-specific DNA-binding domain. In *Drosophila,* DNA sequences known as Polycomb responsive elements (PREs) mediate PcG recruitment through a combination of specific transcription factors. Although similar mechanisms have been proposed in mammals (Arnold *et al*, 2013; Sing *et al*, 2009; Woo *et al*, 2010), they do not appear to be the general rule. Indeed, the specific transcription factors found to bind these putative mammalian PREs do not act consistently as PRC2 genomewide recruiters. Importantly, GC-rich regions are frequently bound by PRC2 components (Ku *et al*, 2008) and they are, in some instances, sufficient to mediate PRC2 recruitment (Mendenhall *et al*, 2010; Jermann *et al*, 2014), although once again this cannot account for the specificity and dynamics of Polycomb recruitment in diverse developmental contexts.

It has been long known that RNAs carry out many functions independent of their protein-coding potential (Cech & Steitz, 2014). Non-coding RNAs are divided into various subclasses, one of which comprises the lncRNAs. LncRNAs are defined as RNA molecules longer than 200 nucleotides that are transcribed by RNA polymerase II, capped, spliced, and polyadenylated. Hence, with the exception of their lack of coding potential, lncRNAs fully resemble messenger RNAs. Many cellular functions have been ascribed to lncRNAs, although genetic inactivation has not always substantiated the initial observations (Rutenberg-Schoenberg *et al*, 2016). Nonetheless, several lncRNAs are reported to influence transcription in the nucleus, in particular through the regulation of chromatin modifiers (Schmitz *et al*, 2016). The variety of lncRNAs and their tissue-specific patterns of expression point toward potential functions in development.

Maybe not surprisingly, lncRNAs have been proposed to play an important role in the recruitment of PRC2 to specific

1 Institut Curie, PSL Research University, Paris, France
2 INSERM U934, CNRS UMR3215, Paris, France
3 INSERM U900, Mines ParisTech, Paris, France
4 CNRS UMR 8015, Université Paris Descartes, Paris, France
*Corresponding author. Tel: +33 156246551; Fax: +33 156246939; E-mail: raphael.margueron@curie.fr

chromatin regions, both in *cis* and in *trans* (Koziol & Rinn, 2010). The best-studied example of lncRNA-dependent *cis*-targeting of chromatin modifiers is the localization of PRC2 to the inactive chromosome X (Xi), downstream of Xist RNA (Plath *et al*, 2003). A direct interaction between PRC2 and the conserved Xist A-repeat region has been suggested to mediate this effect (Zhao *et al*, 2008). However, H3K27me3 deposition is still induced when Xist RNA is deleted for the A-repeats (Kohlmaier *et al*, 2004; da Rocha *et al*, 2014), and recent studies aimed at characterizing the Xist interactome did not retrieve factors unambiguously linked to PRC2 (Chu *et al*, 2015; McHugh *et al*, 2015; Minajigi *et al*, 2015). While other domains of Xist could be involved in PRC2 targeting (da Rocha *et al*, 2014), direct physical interaction between Xist and PRC2 still remains to be proven. Importantly, PRC2 is not required for establishement of transcriptional silencing of the future inactive X; instead, it prevents aberant gene re-activation in specific tissues (Kalantry *et al*, 2006).

The best-known example of PRC2 targeting in *trans* by a lncRNA comes from the HOX antisense intergenic RNA *HOTAIR*. This is a 2,148-nucleotide-long RNA, originating from the *HOXC* locus, that has been reported to be necessary to target PRC2 in *trans* to the *HOXD* locus and additional genomic loci (Rinn *et al*, 2007; Chu *et al*, 2011). *HOTAIR* RNA adopts a defined secondary structure at its 5′ end, which is proposed to be critical for its interaction with the PRC2 *in vitro* (Tsai *et al*, 2010; Somarowthu *et al*, 2015). *HOTAIR* RNA also interacts with another repressive chromatin modifier, the LSD1/coREST/REST complex that catalyzes H3K4me2 demethylation. Hence, it has been proposed to act as a scaffold to coordinate recruitment of both the PRC2 and LSD1/coREST/REST complexes onto chromatin (Tsai *et al*, 2010). However, in mice, genetic deletion of the entire *HOXC* cluster (including *HOTAIR*) does not seem to impair H3K27me3 at the *HOXD* locus in any major way (Schorderet & Duboule, 2011). On the other hand, a more localized deletion of several kb including *HOTAIR* is reported to do so (Li *et al*, 2013). Deregulation of *HOTAIR* has also been observed in cancer cells (Gupta *et al*, 2010). Overexpression studies performed in a cell line model of triple-negative breast cancer have linked elevated *HOTAIR* levels to a re-targeting of PRC2 to several hundred genes, and it has been proposed that *HOTAIR* deregulation might contribute to tumor progression (Gupta *et al*, 2010).

Given the defined interaction suggested to occur between PRC2 and *HOTAIR*, it is surprising to note that *in vitro* and *in vivo* studies investigating the interplay between PRC2 and lncRNAs have reported a rather promiscuous binding of PRC2 and its cofactor JARID2 to RNA (Davidovich *et al*, 2013; Kaneko *et al*, 2013, 2014; Beltran *et al*, 2016). Altogether, the functional specificity of this interaction remains highly debated in the field (Brockdorff, 2013).

In the present study, we set out to further investigate the link between lncRNAs and PRC2 using *HOTAIR* as paradigm. To this end, we first evaluated the transcriptomic consequences of *HOTAIR* overexpression in breast cancer cells in the context of a functional or inactivated PRC2. The lack of substantial changes prompted us to study the role of *HOTAIR* at a local level in model cell lines enabling artificial tethering of *HOTAIR* at a reporter transgene. Our study provides evidence that *HOTAIR* RNA can indeed repress transcription in this context, but that this local effect is PRC2 independent.

Results

Overexpression of *HOTAIR* RNA leads to subtle, PRC2-independent transcriptional changes in the MDA-MB-231 breast cancer cell line

To investigate the link between *HOTAIR* RNA and PRC2, we took advantage of an MDA-MD-231 breast cancer cell line in which we had knocked out *EZH2* by genome editing (Wassef *et al*, 2015). *HOTAIR* RNA overexpression was previously reported to lead to the transcriptional repression of hundreds of genes in the same model, presumably in *trans* (Gupta *et al*, 2010). We overexpressed *HOTAIR* RNA in MDA-MB-231 EZH2+++ (wild-type, original cell pool), MDA-MB-231 EZH2++− (clone with one *EZH2* allele targeted, behaving as wild type), and MDA-MB-231 EZH2−−− breast cancer cells (subclone derived from the MDA-MB-231 EZH2++−, Fig 1A upper and lower panels). Transcript quantification by qRT–PCR revealed that *HOTAIR* RNA is expressed at similar levels in all three conditions (Fig 1A) and that its level of overexpression is comparable to a previous study (Gupta *et al*, 2010). As expected, overexpression of *HOTAIR* RNA has no effect on H3K27me3 global level (Fig 1A) and we did not detect any obvious change of cellular phenotypes such as cell proliferation (Fig EV1A). To get a global picture of *HOTAIR*-mediated transcriptional regulation and of the contribution of PRC2 to this process, we performed RNA sequencing on MDA-MB-231 EZH2++− and MDA-MB-231 EZH2−−− cell types, both in the control condition and upon overexpression of *HOTAIR* RNA. We obtained good correlation between replicates as shown by the Pearson correlation value matrix (Appendix Table S1 and Fig EV1B). We subsequently focused on transcripts displaying the highest dispersion (higher interquartile) across the four conditions. Heatmap representing relative gene expression revealed that the

Figure 1. Limited transcriptomic changes upon *HOTAIR* RNA overexpression in MDA-MB-231 breast cancer cells.

A Western blot analysis of nuclear extracts from indicated cell lines with antibodies for EZH2 and H3K27m2/3 mark. Lamin B1 and H3 are shown as loading controls (upper panel). qRT–PCR to test *HOTAIR* overexpression in the corresponding cell lines. *Y*-axis represents *HOTAIR* expression relative to U1 RNA (individual experiments and mean, $n = 2$) (lower panel).

B Heatmap showing expression intensity of the 1,000 genes with the higher interquartile range. Genes up- or downregulated from MDA-MB-231 EZH2++− and MDA-MB-231 EZH2−−− with (+*HOTAIR*) or without *HOTAIR* samples are shown. Red indicates high expression, and green indicates low expression.

C Volcano plots representing gene expression change upon overexpression of *HOTAIR* in MDA-MB-231 EZH2++−, or MDA-MB-231 EZH2−−− (*y*-axis: $\log_{10}$ *P*-value, *x*-axis: $\log_2$ fold change). Red dots represent genes whose expression changes by more than twofold with a *P*-value < 0.05. *P*-values: moderated *t*-statistics.

D Left panel: Gene expression correlation between cells overexpressing *HOTAIR* or not. Expression is quantified as FPKM; red dots are differentially expressed genes (DEG). Right panel: Average FPKM for non-DEG genes (> 1 FPKM in at least one of the four conditions) or DEG as defined in (C).

Source data are available online for this figure.

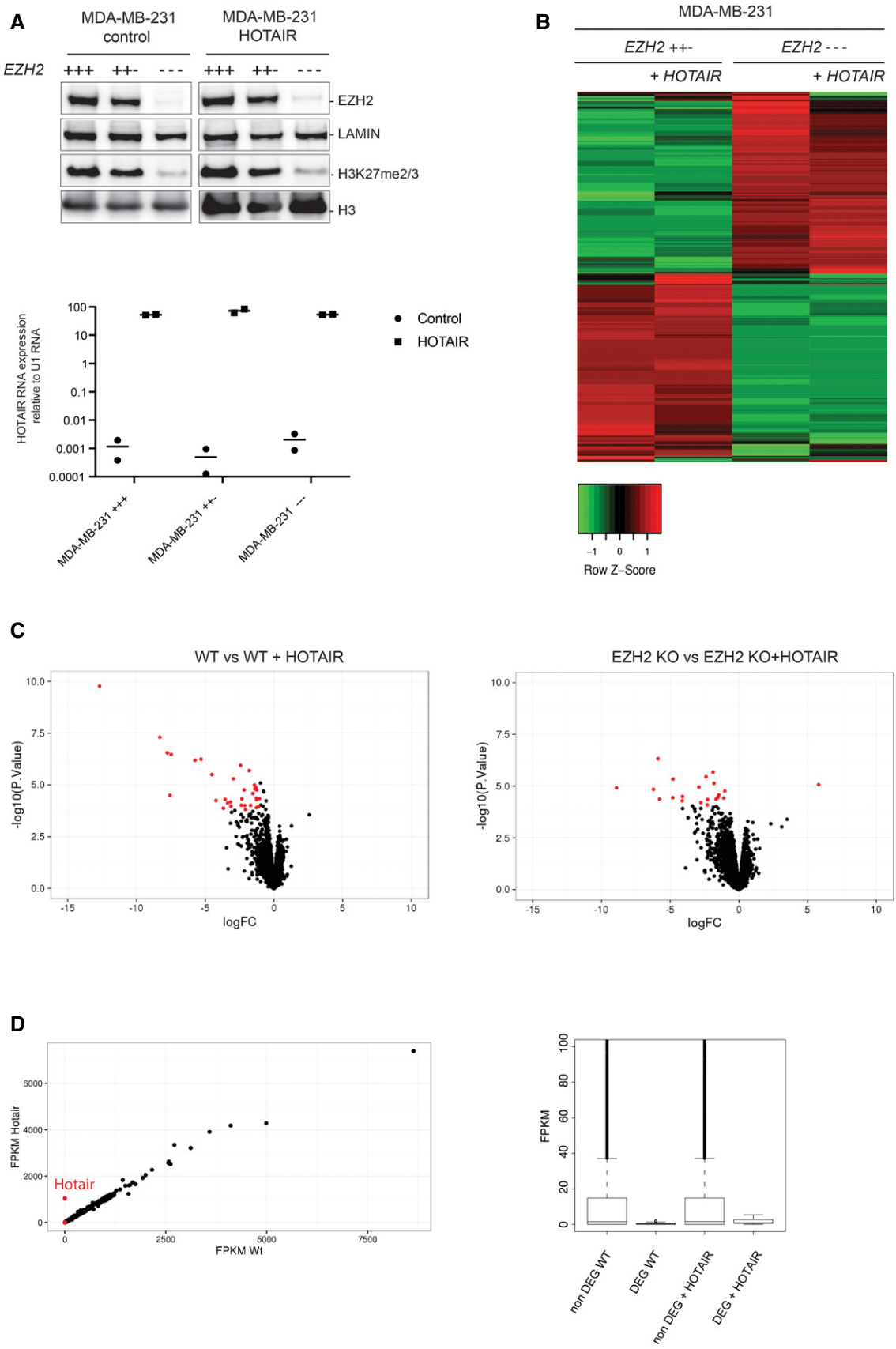

A
MDA-MB-231 control
MDA-MB-231 HOTAIR
EZH2
+++ ++- --- +++ ++- ---
EZH2
LAMIN
H3K27me2/3
H3
HOTAIR RNA expression relative to U1 RNA
100
10
1
0.1
0.01
0.001
0.0001
Control
HOTAIR
MDA-MB-231 +++
MDA-MB-231 ++-
MDA-MB-231 ---
B
MDA-MB-231
EZH2 ++-
+ HOTAIR
EZH2 ---
+ HOTAIR
−1 0 1
Row Z−Score
C
WT vs WT + HOTAIR
EZH2 KO vs EZH2 KO+HOTAIR
-log10(P.Value)
logFC
D
FPKM Hotair
FPKM Wt
Hotair
FPKM
non DEG WT
DEG WT
non DEG + HOTAIR
DEG + HOTAIR

Figure 1.

two main clusters of differentially expressed genes are defined by *EZH2* mutation and not by *HOTAIR* RNA overexpression (Fig 1A). In fact, we observed very similar correlation levels between duplicates and upon overexpression of *HOTAIR* (Fig EV1B). Nonetheless, we selected transcripts that were differentially expressed upon overexpression of *HOTAIR* RNA with an absolute expression fold change superior to 2 and a *P*-value lower than 0.05. Within these criteria, very few transcripts were differentially expressed and most of them were upregulated regardless of whether PRC2 was functional or deficient (red dots on volcano plot, Figs 1C and EV1C). In addition, close examination of this set of genes revealed that they are characterized by a very low read count (red dots, Fig 1D). Of note, transcripts of genes that were previously reported to gain H3K27me3 upon overexpression of *HOTAIR* RNA (Gupta *et al*, 2010) or that are located within 100 kb of its binding sites identified by ChIRP (Chu *et al*, 2011) revealed a similar trend (Fig EV1D).

Altogether, these experiments suggest that *HOTAIR* RNA overexpression only marginally affects gene expression in the MDA-MB-231 breast cancer cell line and that this function does not critically require PRC2.

In vivo tethering of *HOTAIR* RNA induces gene silencing

The lack of a substantial effect of *HOTAIR* overexpression on gene expression profiles prompted us to develop a method to assess whether *HOTAIR* could have a more local impact on transcription. To this end, we set up an RNA-tethering system to force the recruitment of *HOTAIR* at a reporter transgene.

This system exploits two well-known heterologous tools: the bacteriophage MS2 coat protein (MS2BP), which binds to the MS2 stem loop RNA (MS2 loop), and the UAS/Gal4 tethering system. Both systems have been successfully used in eukaryotic cells, the former to tether MS2BP-fused proteins to MS2 loop hybrid RNAs (Keryer-Bibens *et al*, 2008) and the latter to target transcription factors or chromatin modifiers. The parental cell line (labeled 1) is T-Rex HEK293 stably transfected with a luciferase reporter gene, the expression of which is controlled by the *tk* minimal promoter. UAS/Gal4-binding sites enable the recruitment of a Gal4-DNA binding domain fused to a protein of interest (Fig 2A). We derived a subclone constitutively expressing a Gal4-DNA binding domain MS2 coat-protein fusion protein (labeled 2). From this clone, we subsequently derived cells expressing either *MS2 loop-HOTAIR* (labeled 3) or *MS2 loop-HOTAIR-Rev* (RNA antisense to *HOTAIR*, labeled 4) hybrid RNAs (Fig 2A). We checked the expression of the fused Gal4-MS2BP protein by Western blot (Fig EV2A). Both *MS2-HOTAIR* RNAs were expressed at similar levels (Fig EV2B). We confirmed the recruitment of Gal4-MS2BP to the transgene by performing chromatin immunoprecipitation (Wang *et al*, 2005), using an antibody recognizing the Gal4-binding domain (Fig 2B). RNA immunoprecipitation (RIP) with the same antibody further indicated that the fusion protein indeed interacts with the *MS2-HOTAIR* or *MS2-HOTAIR-Rev* RNAs (Fig 2C).

Having established the functionality of our system, we tested the transcriptional consequences of tethering *HOTAIR* RNA on the activity of the luciferase reporter. We observed a 75% reduction in luciferase activity in the presence of *MS2-HOTAIR* RNA, a reduction that was not seen in the presence of *MS2-HOTAIR-Rev* RNA (Fig 2D) or in cells overexpressing the MS2 loops alone (Fig EV2B). To verify that this effect was not clone specific and that it required continuous tethering of *MS2-HOTAIR* RNA, we used three different strategies. First, we confirmed the repression of the luciferase reporter in another clone expressing equal levels of *MS2-HOTAIR* (Fig EV2B and C). Then, we checked that *HOTAIR*-mediated transcriptional repression is relieved when preventing its recruitment by knocking down the Gal4-MS2BP protein. Indeed, upon effective knockdown of the Gal4-MS2BP protein by RNA interference (shGAL4) in the MS2BP and MS2BP *MS2-HOTAIR* cell models (Fig 2E, lower panel), we observed a release of luciferase repression as compared to a scramble shRNA construct (Fig 2E, upper panel). Finally, we verified that abrogating *HOTAIR* expression by knocking out the *MS2-HOTAIR* construct in the MS2BP *MS2-HOTAIR* cell model also releases luciferase repression. In two clones knocked out for the *MS2-HOTAIR* construct as shown by qRT–PCR (Fig 2F, lower panel, labeled 3-*HOTAIR* K.O. cl.1 and 3-*HOTAIR* K.O. cl.2), we could confirm a consistent increase in luciferase activity (Fig 2F, upper panel).

Altogether, these results extensively validate our approach to tether RNA to chromatin. More importantly, we demonstrate that forced recruitment of *HOTAIR* specifically leads to transcriptional repression.

Artificial tethering of *HOTAIR* RNA is associated with changes in chromatin structure

Given the observed gene silencing effect of *HOTAIR* RNA, we wished to explore its underlying mechanisms and, in particular,

Figure 2. *MS2-HOTAIR* RNA causes repression when tethered to the luciferase transgene.

A Schematic representation of the RNA-tethering system to chromatin. *LUC A*, *LUC D* and *LUC E* indicate primer sets along the luciferase transgene used for ChIP qPCR. Each cell model is labeled by a number which is used in all figure legends hereafter.

B ChIP experiments with Gal4 antibody in the cell lines indicated on the *x*-axis. *Y*-axis represents percent of input (mean ± SD, $n = 3$).

C RIP experiments with Gal4 antibody in the cell lines indicated on the *x*-axis. MS2 loop and U1 primers were used in qRT–PCR. *Y*-axis represents fold enrichment to IgG (individual experiments and mean, $n = 2$). Input (In) and IP were loaded and probed with Gal4 antibody (lower panel).

D Relative luciferase activity in the cell lines indicated on the *x*-axis. Values represent the relative luciferase activity normalized to the amount of protein (mean ± SD, $n = 4$). Statistical analysis: unpaired *t*-test, $^{***}P < 0.001$.

E Upper panel: Relative luciferase activity in the cell lines indicated in the right legend (mean ± SD, $n = 4$). Statistical analysis: unpaired *t*-test, $^{***}P < 0.001$. Cells were infected either with scramble (scr) or shRNA targeting Gal4-MS2BP (shGAL4). Lower panel: Western blot analysis with anti-Gal4 antibody in the different cell models; SUZ12 was used as a loading control.

F Upper panel: Relative luciferase activity in the cell lines indicated in the right legend (mean ± SD, $n \geq 2$). Lower panel: qRT–PCR to detect *MS2-HOTAIR* RNA in the cell models indicated in the right legend. *Y*-axis represents MS2 loop RNA levels normalized to actin and calculated over the parental cells (individual experiments and mean, $n = 2$).

Source data are available online for this figure.

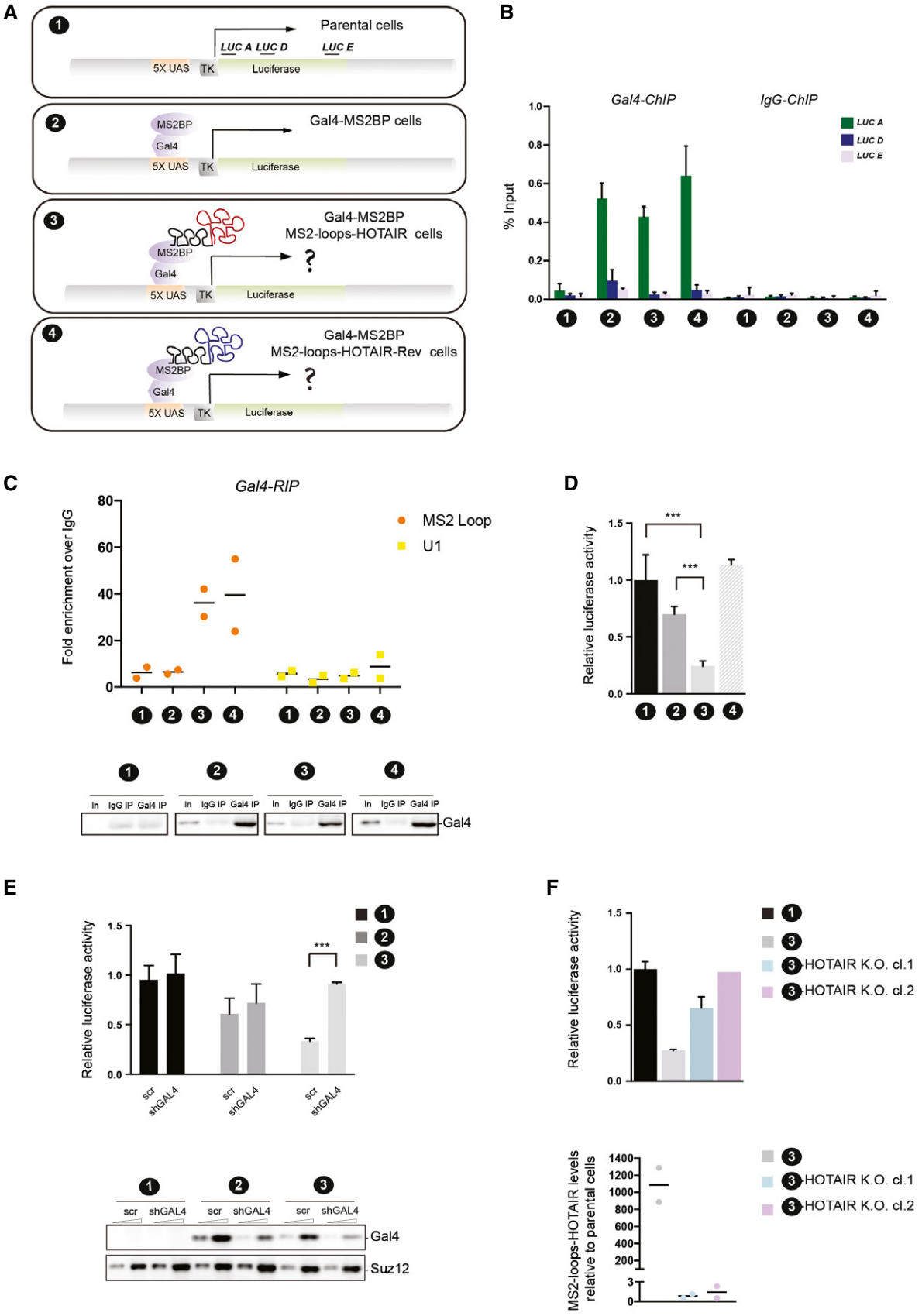

A
Parental cells
LUC A LUC D LUC E
5X UAS TK Luciferase
Gal4-MS2BP cells
MS2BP
Gal4
Gal4-MS2BP
MS2-loops-HOTAIR cells
Gal4-MS2BP
MS2-loops-HOTAIR-Rev cells
B
Gal4-ChIP
IgG-ChIP
LUC A
LUC D
LUC E
% Input
C
Gal4-RIP
MS2 Loop
U1
Fold enrichment over IgG
In IgG IP Gal4 IP
Gal4
D
Relative luciferase activity
E
Relative luciferase activity
scr
shGAL4
Gal4
Suz12
F
Relative luciferase activity
HOTAIR K.O. cl.1
HOTAIR K.O. cl.2
MS2-loops-HOTAIR levels relative to parental cells

Figure 2.

whether it involves specific chromatin regulatory activities. Therefore, we performed ChIP experiments in our MS2-*HOTAIR* RNA-tethered system to evaluate H3K27me3 enrichment upon recruitment of *HOTAIR*. This assay revealed increased enrichment of H3K27me3 downstream of the 5× UAS (*LUC D* and *LUC E* primers) specifically in cells expressing MS2-*HOTAIR* RNA (Fig 3A). Importantly, not all repressive signatures were increased, since we did not observe any change when probing DNA methylation and

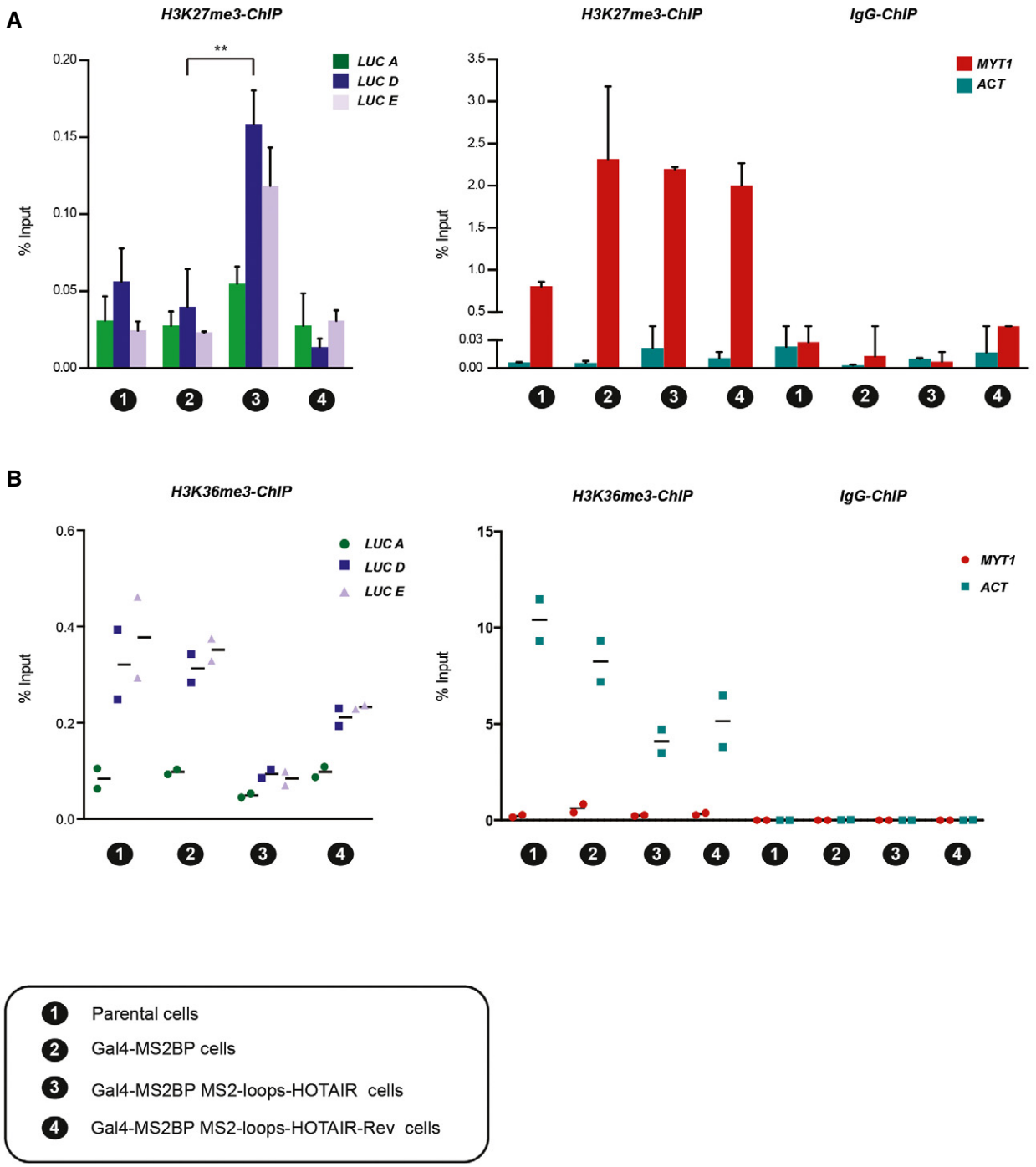


Figure 3. *MS2-HOTAIR* RNA modulates chromatin structure.

A, B ChIP experiments with H3K27me3 (A) or H3K36me3 (B) antibody in the cell models numbered on the *x*-axis of each graph; corresponding legend is at the bottom of the figure. Enrichment for primers located along the luciferase reporter (left) and enrichment for control regions (MYT1 and ACT) (right). *Y*-axis represents percent of input (mean ± SD, $n = 3$ in A; individual experiments and mean, $n = 2$ in B). Statistical analysis: unpaired *t*-test, $^{**}P < 0.01$.

Source data are available online for this figure.

H3K9me2 enrichment at the reporter transgene (Fig EV3B and C). Also, this effect required *HOTAIR* RNA recruitment as it was lost upon knockdown of the Gal4-MS2BP protein (Fig EV3A). Of note, the gain of H3K27me3 in *MS2-HOTAIR* cell line was relatively mild as compared to endogenous PRC2 target such as *MYT1* (Figs 3A and EV3A). To determine whether other chromatin changes occur, we tested the enrichment of the H3K36me3 chromatin mark, which maps to the gene body of transcribed genes. In Gal4-MS2BP and Gal4-MS2BP *MS2-HOTAIR-Rev* cells, H3K36me3 levels in the gene body of the luciferase reporter (*LUC D* and *LUC E* primers) are 10 times lower than in the highly transcribed gene *ACT* (Fig 3B). Nonetheless, we could detect a reduction in H3K36me3 enrichment in *MS2-HOTAIR* cells (Fig 3B). We observed similar trends when analyzing the enrichment for RNA polymerase II (Fig EV3B) but not for H3K27ac, which seems to have the same level of enrichment in all the model cell lines (Fig EV3B).

We conclude from these experiments that *HOTAIR*-mediated transcriptional repression correlates with mild losses and gains of a subset of active and repressive chromatin marks, respectively.

Gain of H3K27me3 upon tethering of *HOTAIR* RNA does not reflect a specific interaction between *HOTAIR* and PRC2

The results described above could fit with the hypothesis that *HOTAIR* RNA recruits PRC2 to chromatin. However, recent studies lead to contrasting conclusions regarding the specificity of PRC2 binding to RNA (Davidovich *et al*, 2013; Kaneko *et al*, 2013, 2014; Beltran *et al*, 2016). These discrepancies might be due in part to the control used to determine whether an interaction is specific. Having established that *HOTAIR-Rev* transcript does not lead to an increased enrichment of H3K27me3 when tethered at a transgene, but considering that it is identical in size to *HOTAIR* transcript and that it is also predicted to form secondary structures, we reasoned that it represents an ideal control for interaction assays. We probed PRC2-*HOTAIR* RNA interaction through two methods: first by sucrose density gradient and then by electrophoretic mobility shift assays (EMSA). We used highly purified PRC2 (Fig EV4A) and *in vitro* transcribed full-length *HOTAIR* or *HOTAIR-Rev* (Fig EV4B) for these assays. Results from the two approaches were mutually consistent and showed that PRC2 binds RNA with high affinity and little specificity; PRC2 interacts equally well with *HOTAIR* and *HOTAIR-Rev* but displays a slightly higher affinity for MS2 loop RNA in EMSA (Fig 4A and B). Next, we analyzed whether adding chromatin to the assay could impact the PRC2–*HOTAIR* interaction, as one might expect if *HOTAIR* acted as a bridge between PRC2 and chromatin. When we incubated chromatin with full-length *HOTAIR* RNAs, the elution pattern of chromatin moved one fraction toward the RNA (Fig EV4C). This event is not specific, as both *HOTAIR* and *HOTAIR-Rev* similarly affect the chromatin elution pattern. We then determined the effect of incubating all three partners together at an equimolar concentration: RNA, PRC2, and chromatin (Fig EV4D). We did not detect any obvious synergy between the three partners. Indeed, the elution pattern of PRC2 in the presence of *HOTAIR* and chromatin was similar to that observed with *HOTAIR* alone. Similarly, chromatin in the presence of *HOTAIR* and PRC2 shifts by one fraction as previously observed with *HOTAIR* alone. Of note, a recent report proposed that the interaction of PRC2 with RNA or chromatin is mutually exclusive, a conclusion which could be consistent with our observations (Beltran *et al*, 2016).

To exclude the possibility that the lack of specificity of PRC2 binding to RNA *in vitro* could be due to inappropriate folding of the RNA under our experimental settings, we probed *HOTAIR* RNA structure by SHAPE-MaP (selective 2′-hydroxyl acylation analyzed by primer extension) (Siegfried *et al*, 2014; Smola *et al*, 2015). Briefly, the RNA is incubated with small molecules that react with single-stranded nucleotides, and high-throughput sequencing is then used to identify the extent of mutations for each position. We used two different chemicals for this assay (NMIA and 1M7) and obtained SHAPE reactivity, which showed a good correlation between the two chemical probings as well as with the previously published data (see source data for Fig 4C). We then focused on results obtained with 1M7 (Figs 4C and EV5) for direct comparison with the previously published structure model of *HOTAIR* (Somarowthu *et al*, 2015). Our reactivity map was used as constrains to model *HOTAIR* secondary structure using the software RNAstructure (Deigan *et al*, 2009). The most stable secondary structure model predicted based on our 1M7 reactivities is slightly distinct from the previous report (Fig EV5); nonetheless, we observed a good overlap between the two structures as shown for the D1 domain (Fig 4C). In summary, the consistency with Somarowthu's thorough *HOTAIR* structure probing makes us confident that *HOTAIR* RNA is folded in a similar structure in both studies.

Finally, we determined whether our *in vitro* results hold true in a cellular context. To address this question, we performed RIP pulling down RNAs interacting with EZH2 in our different cell models. As expected, we observed that EZH2 RIP is enriched for *HOTAIR* over

Figure 4. PRC2 interacts with RNA with low specificity.

A PRC2 was incubated with or without biotinylated *HOTAIR* or *HOTAIR* reverse-complement RNAs and analyzed by density gradient centrifugation on a linear sucrose gradient (10–30%). Individual fractions collected from sucrose gradient were probed by Western blot for EZH2 (upper panel) or by dot blot for biotinylated RNA (lower panel).

B Representative EMSA experiments showing binding of PRC2 to full-length *HOTAIR*, *HOTAIR-Rev* and MS2 loop RNA probes. Equilibrium dissociation constant (K_d) values and Hill slope are calculated on biological replicates (n = 2). Corresponding binding curves of biological duplicate EMSA experiments (bottom panel).

C Predictive secondary structure for the first 530 bp of *HOTAIR* RNA from the RNAstructure software and VARNA visualization software. *HOTAIR* D1 domain as modeled by Somarowthu *et al* (2015) according to SHAPE-CE probing is shown. SHAPE reactivities from Somarowthu *et al* (2015) are depicted by colored nucleotides; 1M7 SHAPE reactivity obtained in our experiment is represented by colored dots over the nucleotides. Highly reactive nucleotides are displayed in red and orange, and low reactive nucleotides are displayed in black or blue according to the values reported in the legend.

D EZH2 binds both *MS2-HOTAIR* and *MS2-HOTAIR-Rev* RNAs *in vivo*. RIP experiments with EZH2 antibody in cell models indicated on the *x*-axis. MS2 loop and U1 primers were used in qRT–PCR. *Y*-axis represents relative enrichment (individual experiments and mean, n = 2). Input (In) and IP were loaded and probed with EZH2 antibody (lower panel). Correspondence between numbers and model cell lines is indicated at the bottom.

Source data are available online for this figure.

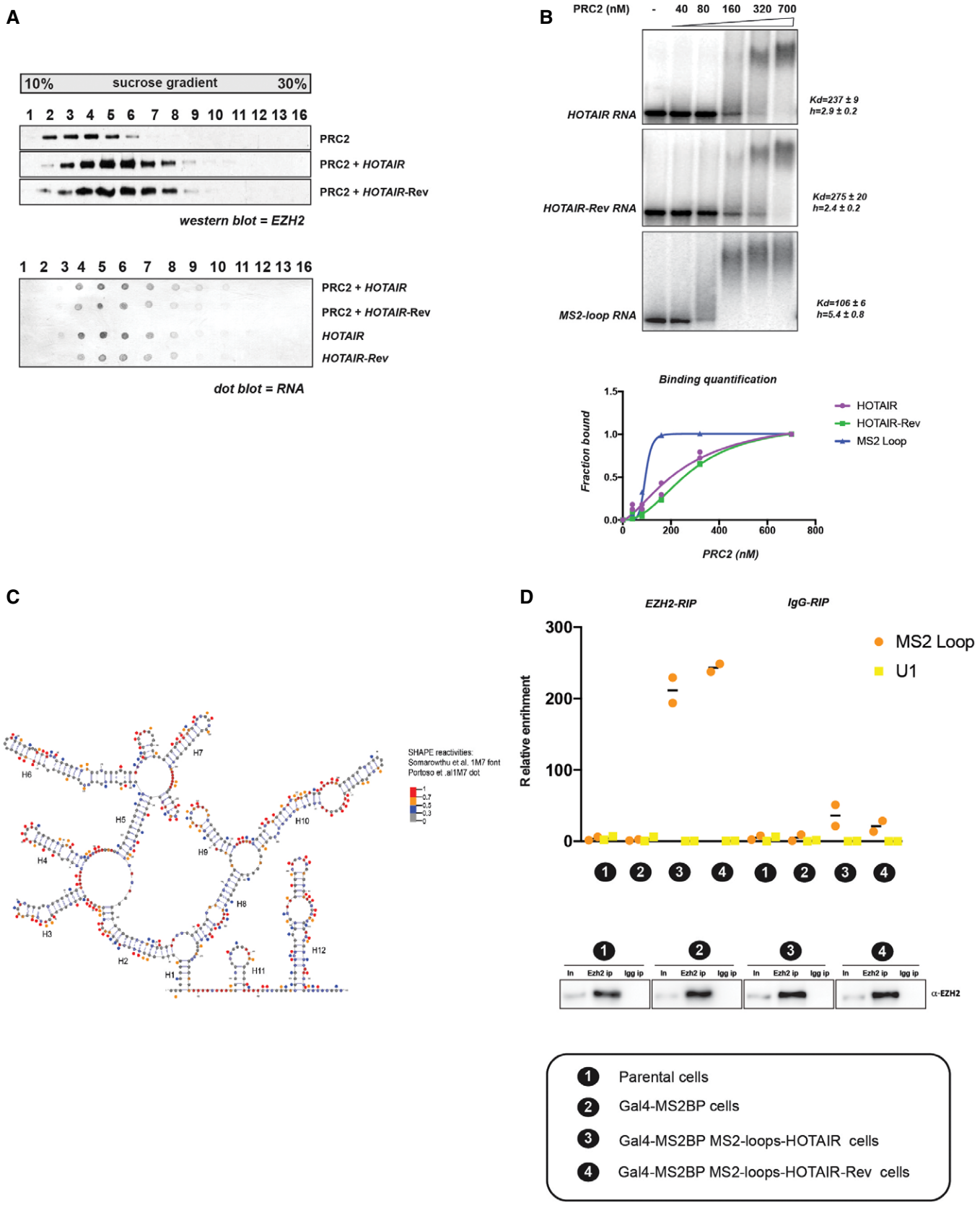


Figure 4.

the IgG control. However, we obtained a similar enrichment for *HOTAIR-Rev,* thus corroborating the *in vitro* findings (Fig 4D). Although we cannot exclude the possibility that the MS2 loops interfere with *HOTAIR* structure in the artificial tethering assay, the similarity between our *in vitro* interaction experiments (*HOTAIR* without MS2 loops) and the RIP experiment in a cellular context (*HOTAIR* with MS2 loops) suggests that it is not the case.

Altogether, our experiments confirm the lack of specificity in the interaction of PRC2 with RNAs both *in vitro* and in cultured cells.

HOTAIR-mediated transcriptional repression does not require PRC2

A previous study reported that simply inhibiting transcription is sufficient to trigger the recruitment of PRC2 to many loci across the genome (Riising *et al*, 2014). In light of those findings and the results of our interaction assays, we considered the possibility that the observed increased H3K27me3 enrichment subsequent to *HOTAIR* tethering might not be caused by direct *HOTAIR*-mediated recruitment of PRC2, but might rather occur as a consequence of reduced transcription. To clarify this point, we employed CRISPR/Cas9 to knock out two essential PRC2 components (EED and SUZ12, Fig 5A). Deleting either *EED* or *SUZ12* led to a complete loss of H3K27me3 both at the global level (Fig 5A, top panel) and at the local level (Fig 5B, lower panel). Yet, in two different *MS2-HOTAIR*-expressing subclones deleted for *EED* (cl.1 and cl.2) or *SUZ12* (cl.1 and cl.2) proteins, we observed the same transcriptional repression as in the *MS2-HOTAIR* parental cell line with wild-type PRC2 components (Fig 5C).

Altogether, our results demonstrate that the silencing of the luciferase reporter requires the continuous presence of *MS2-HOTAIR*

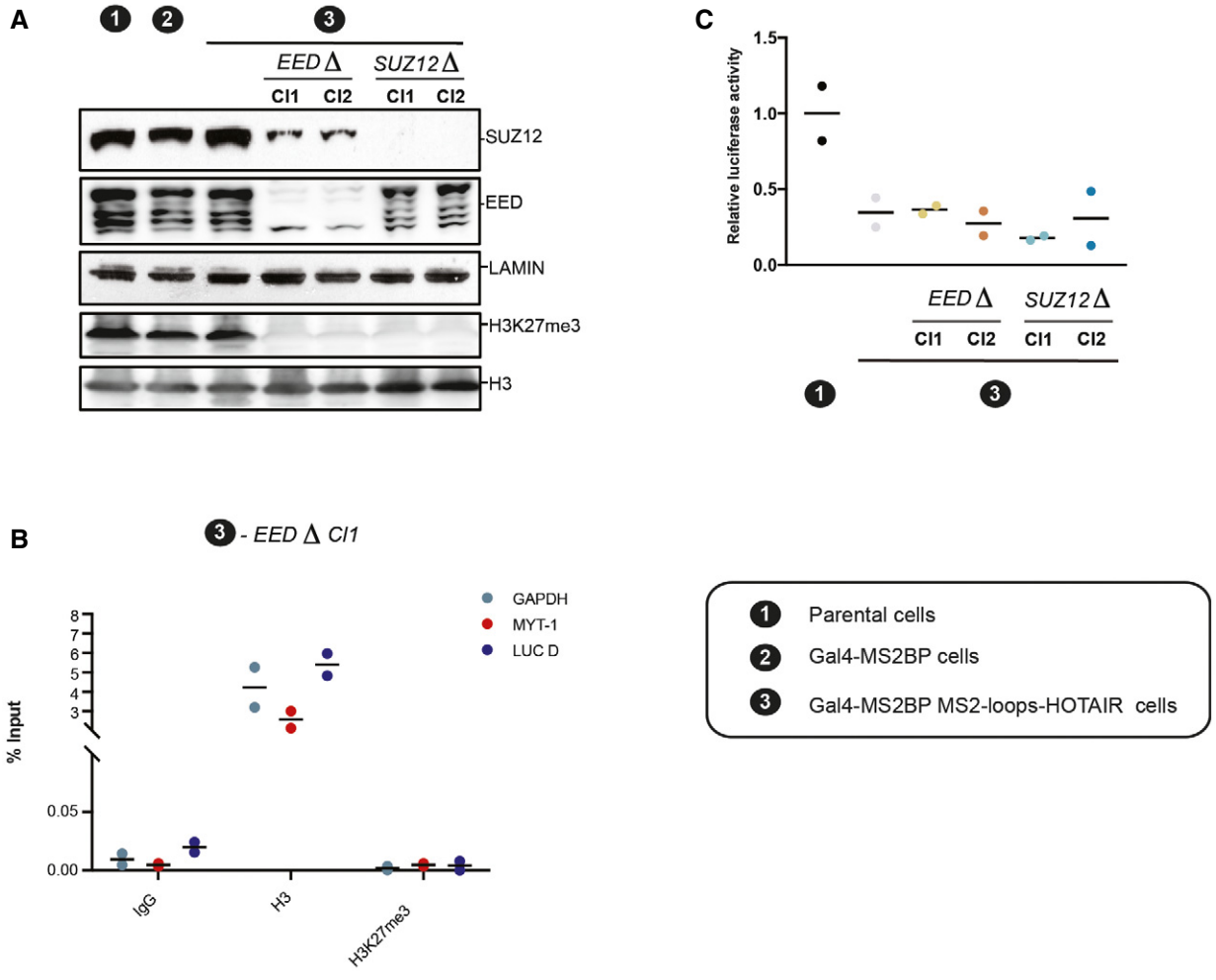


Figure 5. PRC2 is dispensable for *HOTAIR*-mediated transcriptional repression.

A Western blot analysis of nuclear extract from indicated cell lines with antibodies for SUZ12, EED and H3K27m2/3 mark. Lamin and H3 are shown as loading controls.

B ChIP experiments with IgG, histone H3, or H3K27me3 antibodies in the cell model indicated on top. Enrichment for the primers indicated on the right (individual experiments and mean, $n = 2$)

C Relative luciferase activity in the cell lines indicated on the *x*-axis. Values represent the relative luciferase activity normalized to the amount of protein (individual experiments and mean, $n = 2$).

Data information: Correspondence between numbers and model cell lines is indicated at the bottom right.
Source data are available online for this figure.

RNA but not H3K27me3 deposition. It suggests therefore that the *MS2-HOTAIR* transcript modulates transcription independently of PRC2.

Discussion

While an unexpected proportion of eukaryotic genomes is transcribed, many of the resulting transcripts are non-coding RNAs. Among them, the subclass of lncRNAs has been implicated in the regulation of a variety of cellular functions. In particular, nuclear lncRNAs have been found to modulate transcription through the targeting of chromatin modifiers to specific genomic regions. One such example is *HOTAIR*, a lncRNA which was reported to promote breast cancers through the aberrant targeting of PRC2 and consequently inappropriate gene silencing (Gupta *et al*, 2010). However, when we overexpressed *HOTAIR* RNA in MDA-MB-231 breast cancer cells, either in the presence or absence of PRC2, we detected few transcriptomic changes. While the reasons for the discrepancy with the previous report remain unclear, it underscores the need for caution when considering the potential contribution of *HOTAIR* transcript to tumorigenesis.

Considering the lack of broad *trans* effects of overexpressing *HOTAIR* RNA in MDA-MB-231 cell line, we sought a more direct way to gauge whether and how *HOTAIR* RNA regulates transcription. To address this question, we established cell models enabling to artificially tether *HOTAIR* RNA at a stably integrated reporter transgene. This approach revealed that, at least in this specific context and assuming that the MS2 loops do not interfere with *HOTAIR* structure, *HOTAIR* RNA can repress transcription. Several mechanisms could mediate this repressive activity. The most trivial model would be that *HOTAIR* RNA recruitment directly interferes with the RNA polymerase machinery, that is, by steric hindrance. Although we cannot formally exclude this hypothesis, it is undermined by the fact that the recruitment of the transcript antisense to *HOTAIR* (an RNA of identical size) did no impact on reporter activity. An alternative hypothesis is that *HOTAIR* recruits chromatin modifiers, which in turn modulate transcription. Accordingly, we observed that *HOTAIR* artificial recruitment is paralleled by mild changes in chromatin structure (histone methylation). However, when we deleted essential components of PRC2 that abrogate its methyltransferase activity, this did not affect the repressive activity of *HOTAIR* RNA. This shows that at least some of the major changes in chromatin composition upon recruitment of *HOTAIR* RNA are a secondary consequence of changes in transcription in our model. Last, *HOTAIR* may interfere with transcription by interacting with yet unknown factors. Unfortunately, our attempt to use yeast three-hybrid system to identify such factors was unsuccessful (data not shown). Further investigation will therefore be required to address this point.

The interaction between PRC2 and RNA has retained a great deal of attention; however, different studies have reached contrasting conclusions (Davidovich *et al*, 2013; Cifuentes-Rojas *et al*, 2014; Beltran *et al*, 2016). Our results strongly support the weak specificity but strong affinity of PRC2 for RNAs (Davidovich *et al*, 2013; Cifuentes-Rojas *et al*, 2014; Beltran *et al*, 2016). The authors of the latest study proposed that chromatin and RNA might compete for binding to PRC2. In agreement with this idea, we did not find evidence for a complex between RNA, chromatin, and PRC2. We also observed that an excess of RNA could reduce PRC2 enzymatic activity on chromatin but not on another substrate, JARID2 (data not shown). This result is consistent with a specific competition between chromatin and RNA for interaction with PRC2. It is proposed that this antagonism could explain why active transcription prevents PRC2 recruitment (Beltran *et al*, 2016). Intriguingly, the inhibitory activity of RNA on chromatin-modifying enzymes is not exclusive to the PRC2, but appears to be a rather common property, since it has also been observed for SET9 (Kaneko *et al*, 2014), G9A (data not shown), BRG1-BAF (Cajigas *et al*, 2015), and DNMT1 (Di Ruscio *et al*, 2013), even though these enzymes have very distinct functions in transcriptional regulation. It is possible that the affinity of chromatin modifiers for RNA is important to compete with and therefore prevent low affinity and random binding to chromatin. Further studies will be required to test this hypothesis.

Materials and Methods

Recombinant proteins and PRC2 purification

hPRC2 production in SF9 insect cells was performed upon co-infection with EZH2-His, SUZ12-His, RBBP4-Strep-TAG, and EED-Flag-tagged baculoviruses. Cells were lysed in BC300 buffer (300 mM KCl, 10% glycerol, 25 mM Tris–HCl pH 8, 1 mM EDTA), sonicated, and clarified by centrifugation before incubation with Flag beads (M2 beads SIGMA). PRC2 was eluted with Flag peptide and further purified on a MiniQ column to assure homogeneity and complete removal of nucleic acid contaminants. Fraction content was verified on Coomassie.

In vitro RNA transcription with biotinylated or radiolabeled UTP

One microgram of linearized pBluescript plasmid expressing *HOTAIR* reverse-complement or MS2 loop RNA was *in vitro* transcribed for 3 h at 37°C using the MEGAscript T7 transcription kit (AM1334). After DNase treatment, when biotinylated, samples were purified over the MEGAclear™ transcription clean-up kit (AM1908) and checked for full length on agarose gel. When radiolabeled, samples were cleaned with acid phenol/chloroform and precipitated at −20°C with 2.5 vol EtOH and 1/10 3 M NaAcet pH 5.3, 70% EtOH-washed, and resuspended in DEPC H_2O. RNAs were successively quantified by UV absorbance at 260 nm. Purity and integrity of all RNA batches were examined on a 0.8% agarose gel.

Refolding of *in vitro* transcribed RNA

In vitro transcribed RNA was heated at 95°C for 3 min, then immediately placed on ice for 2 min, added 2× refolding buffer (20 mM Tris pH 7; 200 mM KCl; 20 mM $MgCl_2$), and refolded at RT for 20 min.

Nucleosome reconstitution

Nucleosomes were assembled from 5S 12 repeat DNA (Dorigo *et al*, 2004) and purified HeLa cell histone octamers by salt dialysis through a linear gradient (2.2 M NaCl to 0.4 M NaCl) for 20 h, followed by a step dialysis against TE.

Sucrose gradient sedimentation analysis

Equimolar PRC2, *in vitro* reconstituted chromatin, and *in vitro* transcribed and refolded biotinylated RNAs were incubated together prior to sucrose gradients in HEB buffer (25 mM Hepes, 40 mM KCl, 0.2 mM EDTA, 1 mM DTT) for 1 h at RT.

Sucrose gradients were prepared using a gradient maker (Biocomp) according to the manufacturer's instruction and centrifuged for 16 h at 55,000 *g* in a Beckmann 60Ti rotor. Fractions were collected manually (250 μl each fraction), and 30 μl of each sample was loaded on NuPAGE Novex 4–12% Bis–Tris protein for Western blot analysis; 20 μl of each sample added with 0.5% SDS was loaded on 0.8% agarose gel, and 1 μl for each sample was spotted on positively charged nylon membrane nylon for dot blot.

Dot blot

One micoliter from each sucrose gradient centrifugated fraction was spotted on positively charged nylon membrane, let dry for 30 min, and UV-cross-linked. RNA was revealed using the biotin chromogenic detection kit (KO661 Thermo Fischer Scientific) following the manual instructions.

Electrophoretic mobility shift assay

Five nanomolar refolded RNA was incubated with increasing concentration of PRC2 in binding buffer (50 mM Tris–HCl, pH 7.5 at 25°C, 100 mM KCl, 5 mM $MgCl_2$, 0.5 mM $ZnCl_2$, 0.1 mM $CaCl_2$, 2 mM 2-mercaptoethanol, 0.1 mg/ml BSA, 0.1 mg/ml fragmented yeast tRNA, 5% v/v glycerol, 0.025% w/v bromophenol blue, and 0.025% w/v xylene cyanol) at 30°C for 30 min. Samples were cooled to 4°C for 10 min and loaded on a 0.7% agarose gel in 1× TBE buffer at 4°C. Gels were vacuum-dried for 45 min at 80°C on a nylon membrane and two sheets of Whatman 3-mm chromatography paper. Dried gels were exposed to phosphorimaging plates, and signal acquisition was performed with a Typhoon Trio phosphorimager (GE Healthcare). Densitometry was carried out with ImageJ software and data fitted to a sigmoidal binding curve with Prism Software. Data ranges for both dissociation constants and Hill coefficients were calculated on the basis of two replicates.

Cloning

Construction of templates for *in vitro* transcription: *HOTAIR* cDNAs from LZRS-HOTAIR (purchased from Addgene, plasmid #26110, deposited by Howard Chang) were cloned into pBluescript plasmid digested with BamHI. pCR4-24XMS2SL-stable (Addgene plasmid #31865, deposited by Robert Singer) was used to produce *in vitro* MS2 loop RNA. Fusion protein vector: pFLAG_NLS_MS2-MS2 plasmid was a kind gift from Richard Breathnach (Gesnel *et al*, 2009). Gal4-DBD was cloned upstream MS2-MS2 dimer coat protein with EcoRV/XbaI sites. MS2 loop-RNA hybrid constructs: The MS2 loop repeat fragment was digested BamHI/BglII from pCR4-24XMS2SL-stable and inserted in the modified mammalian expression vector pCDNA4/TO linearized with BamHI restriction enzyme. To this plasmid, *HOTAIR* cDNA, digested BamHI from LZRS-*HOTAIR* was ligated 5′ to the MS2 loop repeats. Both orientations were checked by restriction enzyme digestion and sequencing. Cloning the shGal4 sequence into the pLKO.1 hygro vector (Addgene plasmid #24150, deposited by Bob Weinberg) was performed according to the pKLO-TRC cloning vector procedure at http://www.addgene.org/tools/protocols/plko/. The target sequence for Gal4 was ATCGAACAAG CATGCGATATT.

The deletion cassette for *HOTAIR* was built by compatible restriction enzyme digestion and ligation and verified both by restriction enzyme digestion and sequencing at all steps. Briefly, 500 bp was amplified from the pCDNA4/TO HOTAIR-MS2 loop plasmid comprising the promoter region and ligated to the hygromycin B resistance cassette followed by 640 bp of *HOTAIR* cDNA fragment, located 300 bp downstream the J. Rinn *HOTAIR* start site in a pBluescript plasmid. The gRNA target site, designed using the http://crispor.tefor.net/ Web site, GAGAGCACCTCCGGGATATT was comprised within the first 300 bp of *HOTAIR* cDNA and cloned into the gRNA vector (Addgene plasmid #41824, deposited by George Church) according to the Addgene procedure.

The deletion cassette for hEED was built cloning hygromycin B resistance cassette between left and right region homologs to EED exon 2. The gRNA target site GCACCTGGAAGGAAAAGTTG was cloned into the gRNA vector according to the Addgene procedure. The deletion cassette for hSUZ12 was done as for EED with left and right arm homologs to SUZ12 exon 10. The gRNA target site GAGACTCTCTGAATTTCTAG was cloned into the gRNA vector according to the Addgene procedure.

Cell culture and transfections

T-Rex 293 cells (Invitrogen) were grown according to the manufacturer's instructions. MDA-MB-231-derived cell lines were previously described (Wassef *et al*, 2015). All cell lines were tested for the absence of mycoplasma on a monthly basis. All transfections were performed using PEI (polyethylenimine) at 3:1 ratio to DNA.

First, 5XGal4RE-tk-Luc-Neo plasmid was stably integrated into the cells and selected for G418 resistance (0.5 μg/ml). One highly expressing luciferase clone was stably transfected and selected for pFLAG_Gal4DBD_NLS_MS2-MS2 bearing puromycin resistance (10 μg/ml). Subsequently a single clone verified by Western blot for the expression of the fused Gal4DBD_NLS_MS2-MS2 protein was transfected with each MS2 loop-RNA hybrid plasmid bearing Zeocin resistance. Resistant clones selected for Zeocin (0.4 μg/ml) were tested for expression of the different MS2 loop-RNA hybrid constructs by strand-specific RT–PCR and qRT–PCR.

Co-transfection with gRNA targeting *HOTAIR*, hCas9, and the targeted *HOTAIR* construct was performed with PEI. Hygromycin B selection was performed at 0.3 μg/ml.

Co-transfections with gRNAs targeting EED or SUZ12, hCas9, and each of the EED and SUZ12 targeted constructs were performed with PEI. Hygromycin B selection was performed at 0.3 μg/ml.

Retroviral vector production and transduction

Production of shGal4 lentiviral vector was performed in 293T cells. Transduction and selection of target cells were performed according to the online Addgene procedure. Hygromycin B was added at 0.3 μg/ml. Production of overexpressing *HOTAIR* retroviral vectors was performed in 293T cells. Transduction of target cells was performed as for the lentiviral vector.

Quantification of mRNA levels by qRT–PCR

Total RNA was isolated following TRIzol reagent (Invitrogen) extraction instructions. cDNA was synthetized using SuperScript III reverse transcriptase kit (18080044 Invitrogen), and quantitative PCR was performed with technical triplicate using SYBR green reagent (Roche) on a ViiA7 equipment (Applied Biosystems). At least three biological independent experiments were performed for each assay.

Luciferase assay

Luciferase reporter activities were measured in whole-cell lysates using the Luciferase Assay System (Promega, #E15020) and Fluostar Optima BMG Labtech luminometer. All experiments were done in triplicate and normalized for protein concentration (Bradford).

Chromatin immunoprecipitation

ChIPs were performed as previously described (Sanulli *et al*, 2015). 1.2×10^7 cells were plated in 15-cm plates 2 days before cross-linking. Quantification was done as previously described for the qRT–PCR. Primers sequences and antibodies used are provided in Appendix Tables S2 and S3.

DNA methylation analysis

Genomic DNA was treated with bisulfite using EpiTect Bisulfite Kit (Qiagen) and PCR-amplified using a nested PCR strategy. The PCR products were cloned using New England Biolab PCR cloning kit, and individual clones were analyzed by Sanger sequencing. DNA methylation analysis was performed using Quma (Kumaki *et al*, 2008) (http://quma.cdb.riken.jp/).

RNA immunoprecipitation

RNA immunoprecipitation experiments were performed as previously described (Rinn *et al*, 2007) with the following modifications. Two 15-cm plates with 1.5×10^7 cells each were plated 2 days before the experiment, a pre-clearing step for 1–2 h at 4°C before the IP was performed with ON blocked beads (BSA 10 mg/ml as 100× and salmon sperm 10 mg/ml as 10×) in PBS or RIP buffer (150 mM KCl, 25 mM Tris pH 7.4, 5 mM EDTA, 0.5 mM DTT, 0.5% NP-40 added with protease inhibitors, PMSF, and RNase inhibitors). One-fourth of the immunoprecipitated material was tested in Western blot analysis, and the rest was resuspended in TRIzol. Co-precipitated RNAs were isolated, and qRT–PCR for MS2 loop RNA and U1 RNA was performed as described above. Primer sequences and antibodies used are provided in Appendix Tables S2 and S3.

RNA-seq analysis

Total RNA from MDA-MB-231 cell lines was isolated by TRIzol extraction and quality-verified by Bionalyzer. Isolated RNA was used to prepare cDNA libraries and amplified with primers containing sequences required for the Illumina platform. PCR products were cleaned and subjected to 100-bp paired-end sequencing on an Illumina Hi-seq 2500. Sequenced reads from duplicate samples were assembled on the human genome hg19, using tophat_2.0.6 (Kim *et al*, 2013).

The Htseq software (v0.6.0.) was used to define the number of reads associated with each gene. TMM normalization from the edgeR package v3.6.2 (Robinson & Oshlack, 2010) was first applied. As described in the guideline of limma R package v3.20.4, normalized counts were processed by the voom method (Law *et al*, 2014) to convert them into $\log_2$ counts per million with associated precision weights. The differential expression was estimated with the limma package. The *P*-values were adjusted for multiple testing using the Benjamini–Hochberg procedure. Finally, differentially expressed genes with a log fold change > 1, FPKM > 1, and adjusted *P*-value < 0.05 were used for downstream analysis. Genes' FPKM was estimated using the Cuffquant and Cuffnorm tools of the Cufflinks suite (v2.2.1).

The hierarchical clustering was performed using a Pearson correlation distance and a Ward linkage (R v3.2.0, hclust function).

Baculoviruses production

RBBP4-Strep-TAG baculovirus was produced according to the Bac-to-Bac Baculovirus Expression Systems (Invitrogen) starting from pFASTbac vectors.

SHAPE-MaP

SHAPE-MaP structure probing was performed as described by Smola *et al* (2015). Refolded RNA was incubated with 10 mM 1M7(+) (1-methyl-7-nitroisatoic anhydride), 10 mM NMIA (+) (*N*-methylisatoic anhydride) or an equal amount of pure DMSO as a control (−) for 3 min or 22 min at 37°C, respectively, due to the different half-lives of the SHAPE reagents. The samples were then purified by G50 columns and subsequently fragmented to obtain 300-bp RNA fragments. Reverse transcription was then performed in the presence of Mn^{2+} using SuperScript III reverse transcriptase kit (Invitrogen); finally, samples (+) and (−) were purified using G50 columns. In parallel, an RNA-denatured sample was treated following the same steps as the (+) samples as a second negative control (−). SHAPE reactions (+) and (−) were then sequenced using an Ion Torrent sequencing platform, and sequencing data were taken into a bioinformatics pipeline to obtain SHAPE reactivities for 1M7 or NMIA for each RNA nucleotide and normalized for DMSO negative control and RNA-denatured negative control. The bioinformatics script provided by the Weeks laboratory was adapted for Ion Torrent output files by A. Saadi and Y. Ponty (manuscript in preparation).

To generate the *HOTAIR* secondary structure maps using the software RNAstructure, SHAPE 1M7 reactivity was used to provide pseudo-energy constraints, while VARNA software was used to visualize the predicted structure (Darty *et al*, 2009). Resulting structures were manually evaluated for match with NMIA probing data. SHAPE reactivities are listed in the source data for Fig 4C.

Nuclear extract

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, 0.5 mM DTT, 1 mM PSMF) for 10 min on ice, centrifuged at

7,000 *g* for 10 min, resuspended in buffer B (25 mM Hepes pH 7.9, 1.5 mM $MgCl_2$, 700 mM NaCl, 0.5 mM DTT, 0.1 mM EDTA, 20% glycerol), sonicated, and centrifuged at 21,000 *g* for 15 min.

Data access

The data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus and are accessible through GEO series accession number GSE72524 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE72524).

Expanded View for this article is available online.

Acknowledgements
We thank Dr. Breathnach Richard for providing the pFLAG_NLS_MS2-MS2 plasmid, Margaux Charruel for technical help, Armelle Luscan for providing EED and SUZ12 targeting vectors, Anne-Catherine Dock-Bregeon for advices on RNA SHAPE, Afaf Saadi and Yann Ponty for help with processing the raw SHAPE-MaP Data, and Delphine Allouche for advices on the SHAPE-MaP experiments. We also thank Drs. Edith Heard, Antoine Graindorge, Serena Sanulli, Daniel Holoch, and Mythili Ganapathi for critical reading of the manuscript. Work in RM laboratory is funded by an ERC-Stg (REPODDID) grant and the Institut National du Cancer (INCa, grant 2012-1-PLBIO). High-throughput sequencing has been performed by the NGS platform of the Institut Curie and supported by the grants ANR-10-EQPX-03 and ANR10-INBS-09-08 from the Agence Nationale de le Recherche (investissements d'avenir) and by the Canceropôle Ile-de-France.

Author contributions
MP, ZB, RR, AMo, AMi, and MW performed experiments. MP, IV, NS, BS, and RM analyzed data. MP and RM wrote the manuscript. All authors edited the manuscript.

Conflict of interest
The authors declare that they have no conflict of interest.

References

Arnold P, Scholer A, Pachkov M, Balwierz PJ, Jorgensen H, Stadler MB, van Nimwegen E, Schubeler D (2013) Modeling of epigenome dynamics identifies transcription factors that mediate Polycomb targeting. *Genome Res* 23: 60 – 73

Beltran M, Yates CM, Skalska L, Dawson M, Reis FP, Viiri K, Fisher CL, Sibley CR, Foster BM, Bartke T, Ule J, Jenner RG (2016) The interaction of PRC2 with RNA or chromatin is mutually antagonistic. *Genome Res* 26: 896 – 907

Brockdorff N (2013) Noncoding RNA and Polycomb recruitment. *RNA* 19: 429 – 442

Cajigas I, Leib DE, Cochrane J, Luo H, Swyter KR, Chen S, Clark BS, Thompson J, Yates JR III, Kingston RE, Kohtz JD (2015) Evf2 lncRNA/BRG1/DLX1 interactions reveal RNA-dependent inhibition of chromatin remodeling. *Development* 142: 2641 – 2652

Cech TR, Steitz JA (2014) The noncoding RNA revolution-trashing old rules to forge new ones. *Cell* 157: 77 – 94

Chu C, Qu K, Zhong FL, Artandi SE, Chang HY (2011) Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. *Mol Cell* 44: 667 – 678

Chu C, Zhang QC, da Rocha ST, Flynn RA, Bharadwaj M, Calabrese JM, Magnuson T, Heard E, Chang HY (2015) Systematic discovery of Xist RNA binding proteins. *Cell* 161: 404 – 416

Cifuentes-Rojas C, Hernandez AJ, Sarma K, Lee JT (2014) Regulatory interactions between RNA and Polycomb repressive complex 2. *Mol Cell* 55: 171 – 185

Darty K, Denise A, Ponty Y (2009) VARNA: interactive drawing and editing of the RNA secondary structure. *Bioinformatics* 25: 1974 – 1975

Davidovich C, Zheng L, Goodrich KJ, Cech TR (2013) Promiscuous RNA binding by Polycomb repressive complex 2. *Nat Struct Mol Biol* 20: 1250 – 1257

Deigan KE, Li TW, Mathews DH, Weeks KM (2009) Accurate SHAPE-directed RNA structure determination. *Proc Natl Acad Sci USA* 106: 97 – 102

Di Ruscio A, Ebralidze AK, Benoukraf T, Amabile G, Goff LA, Terragni J, Figueroa ME, De Figueiredo Pontes LL, Alberich-Jorda M, Zhang P, Wu M, D'Alo F, Melnick A, Leone G, Ebralidze KK, Pradhan S, Rinn JL, Tenen DG (2013) DNMT1-interacting RNAs block gene-specific DNA methylation. *Nature* 503: 371 – 376

Dorigo B, Schalch T, Kulangara A, Duda S, Schroeder RR, Richmond TJ (2004) Nucleosome arrays reveal the two-start organization of the chromatin fiber. *Science* 306: 1571 – 1573

Gesnel MC, Del Gatto-Konczak F, Breathnach R (2009) Combined use of MS2 and PP7 coat fusions shows that TIA-1 dominates hnRNP A1 for K-SAM exon splicing control. *J Biomed Biotechnol* 2009: 104853

Gupta RA, Shah N, Wang KC, Kim J, Horlings HM, Wong DJ, Tsai MC, Hung T, Argani P, Rinn JL, Wang Y, Brzoska P, Kong B, Li R, West RB, van de Vijver MJ, Sukumar S, Chang HY (2010) Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* 464: 1071 – 1076

Jermann P, Hoerner L, Burger L, Schubeler D (2014) Short sequences can efficiently recruit histone H3 lysine 27 trimethylation in the absence of enhancer activity and DNA methylation. *Proc Natl Acad Sci USA* 111: E3415 – E3421

Kalantry S, Mills KC, Yee D, Otte AP, Panning B, Magnuson T (2006) The Polycomb group protein EED protects the inactive X-chromosome from differentiation-induced reactivation. *Nat Cell Biol* 8: 195 – 202

Kaneko S, Son J, Shen SS, Reinberg D, Bonasio R (2013) PRC2 binds active promoters and contacts nascent RNAs in embryonic stem cells. *Nat Struct Mol Biol* 20: 1258 – 1264

Kaneko S, Bonasio R, Saldana-Meyer R, Yoshida T, Son J, Nishino K, Umezawa A, Reinberg D (2014) Interactions between JARID2 and noncoding RNAs regulate PRC2 recruitment to chromatin. *Mol Cell* 53: 290 – 300

Keryer-Bibens C, Barreau C, Osborne HB (2008) Tethering of proteins to RNAs by bacteriophage proteins. *Biol Cell* 100: 125 – 138

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

Kohlmaier A, Savarese F, Lachner M, Martens J, Jenuwein T, Wutz A (2004) A chromosomal memory triggered by Xist regulates histone methylation in X inactivation. *PLoS Biol* 2: E171

Koziol MJ, Rinn JL (2010) RNA traffic control of chromatin complexes. *Curr Opin Genet Dev* 20: 142 – 148

Ku M, Koche RP, Rheinbay E, Mendenhall EM, Endoh M, Mikkelsen TS, Presser A, Nusbaum C, Xie X, Chi AS, Adli M, Kasif S, Ptaszek LM, Cowan CA, Lander ES, Koseki H, Bernstein BE (2008) Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. *PLoS Genet* 4: e1000242

Kumaki Y, Oda M, Okano M (2008) QUMA: quantification tool for methylation analysis. *Nucleic Acids Res* 36: W170 – W175

Law CW, Chen Y, Shi W, Smyth GK (2014) Voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol* 15: R29

Li L, Liu B, Wapinski OL, Tsai MC, Qu K, Zhang J, Carlson JC, Lin M, Fang F, Gupta RA, Helms JA, Chang HY (2013) Targeted disruption of Hotair leads to homeotic transformation and gene derepression. *Cell Rep* 5: 3 – 12

Margueron R, Reinberg D (2011) The Polycomb complex PRC2 and its mark in life. *Nature* 469: 343 – 349

McHugh CA, Chen CK, Chow A, Surka CF, Tran C, McDonel P, Pandya-Jones A, Blanco M, Burghard C, Moradian A, Sweredoski MJ, Shishkin AA, Su J, Lander ES, Hess S, Plath K, Guttman M (2015) The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature* 521: 232 – 236

Mendenhall EM, Koche RP, Truong T, Zhou VW, Issac B, Chi AS, Ku M, Bernstein BE (2010) GC-rich sequence elements recruit PRC2 in mammalian ES cells. *PLoS Genet* 6: e1001244

Minajigi A, Froberg JE, Wei C, Sunwoo H, Kesner B, Colognori D, Lessing D, Payer B, Boukhali M, Haas W, Lee JT (2015) Chromosomes. A comprehensive Xist interactome reveals cohesin repulsion and an RNA-directed chromosome conformation. *Science* 349: aab2276

Plath K, Fang J, Mlynarczyk-Evans SK, Cao R, Worringer KA, Wang H, de la Cruz CC, Otte AP, Panning B, Zhang Y (2003) Role of histone H3 lysine 27 methylation in X inactivation. *Science* 300: 131 – 135

Riising EM, Comet I, Leblanc B, Wu X, Johansen JV, Helin K (2014) Gene silencing triggers polycomb repressive complex 2 recruitment to CpG islands genome wide. *Mol Cell* 55: 347 – 360

Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Brugmann SA, Goodnough LH, Helms JA, Farnham PJ, Segal E, Chang HY (2007) Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* 129: 1311 – 1323

Robinson MD, Oshlack A (2010) A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol* 11: R25

da Rocha ST, Boeva V, Escamilla-Del-Arenal M, Ancelin K, Granier C, Matias NR, Sanulli S, Chow J, Schulz E, Picard C, Kaneko S, Helin K, Reinberg D, Stewart AF, Wutz A, Margueron R, Heard E (2014) Jarid2 is implicated in the initial Xist-induced targeting of PRC2 to the inactive X chromosome. *Mol Cell* 53: 301 – 316

Rutenberg-Schoenberg M, Sexton AN, Simon MD (2016) The Properties of long noncoding RNAs that regulate chromatin. *Annu Rev Genomics Hum Genet* 17: 69 – 94

Sanulli S, Justin N, Teissandier A, Ancelin K, Portoso M, Caron M, Michaud A, Lombard B, da Rocha ST, Offer J, Loew D, Servant N, Wassef M, Burlina F, Gamblin SJ, Heard E, Margueron R (2015) Jarid2 methylation via the PRC2 complex regulates H3K27me3 deposition during cell differentiation. *Mol Cell* 57: 769 – 783

Schmitz SU, Grote P, Herrmann BG (2016) Mechanisms of long noncoding RNA function in development and disease. *Cell Mol Life Sci* 73: 2491 – 2509

Schorderet P, Duboule D (2011) Structural and functional differences in the long non-coding RNA hotair in mouse and human. *PLoS Genet* 7: e1002071

Siegfried NA, Busan S, Rice GM, Nelson JA, Weeks KM (2014) RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP). *Nat Methods* 11: 959 – 965

Simon JA, Kingston RE (2009) Mechanisms of polycomb gene silencing: knowns and unknowns. *Nat Rev Mol Cell Biol* 10: 697 – 708

Sing A, Pannell D, Karaiskakis A, Sturgeon K, Djabali M, Ellis J, Lipshitz HD, Cordes SP (2009) A vertebrate Polycomb response element governs segmentation of the posterior hindbrain. *Cell* 138: 885 – 897

Smola MJ, Rice GM, Busan S, Siegfried NA, Weeks KM (2015) Selective 2′-hydroxyl acylation analyzed by primer extension and mutational profiling (SHAPE-MaP) for direct, versatile and accurate RNA structure analysis. *Nat Protoc* 10: 1643 – 1669

Somarowthu S, Legiewicz M, Chillon I, Marcia M, Liu F, Pyle AM (2015) HOTAIR forms an intricate and modular secondary structure. *Mol Cell* 58: 353 – 361

Tsai MC, Manor O, Wan Y, Mosammaparast N, Wang JK, Lan F, Shi Y, Segal E, Chang HY (2010) Long noncoding RNA as modular scaffold of histone modification complexes. *Science* 329: 689 – 693

Wang S, Cowan CA, Chipperfield H, Powers RD (2005) Gene expression in the preimplantation embryo: in-vitro developmental changes. *Reprod Biomed Online* 10: 607 – 616

Wassef M, Rodilla V, Teissandier A, Zeitouni B, Gruel N, Sadacca B, Irondelle M, Charruel M, Ducos B, Michaud A, Caron M, Marangoni E, Chavrier P, Le Tourneau C, Kamal M, Pasmant E, Vidaud M, Servant N, Reyal F, Meseure D *et al* (2015) Impaired PRC2 activity promotes transcriptional instability and favors breast tumorigenesis. *Genes Dev* 29: 2547 – 2562

Woo CJ, Kharchenko PV, Daheron L, Park PJ, Kingston RE (2010) A region of the human HOXD cluster that confers polycomb-group responsiveness. *Cell* 140: 99 – 110

Zhao J, Sun BK, Erwin JA, Song JJ, Lee JT (2008) Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* 322: 750 – 756

Expanded View Figures

Figure EV1. *HOTAIR* overexpression has marginal phenotypic and transcriptomic consequences on MDA-MB-231.

A Proliferation assays in two different cell culture conditions. Cells were counted with an automatic cell counter at the time indicated on the plot (individual experiments and mean, $n = 2$).

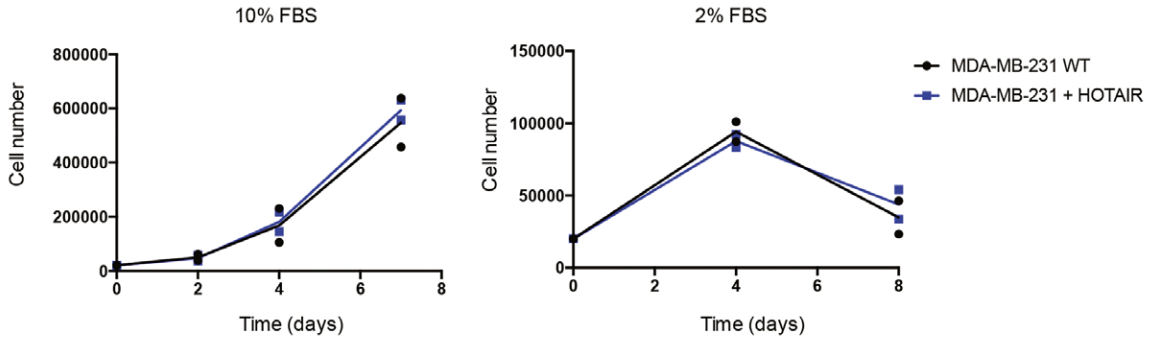
B Pearson correlation matrix between each single-sequenced sample. Value varies between 1 and 0.85.

C Number of differentially expressed genes depending on cutoff parameter.

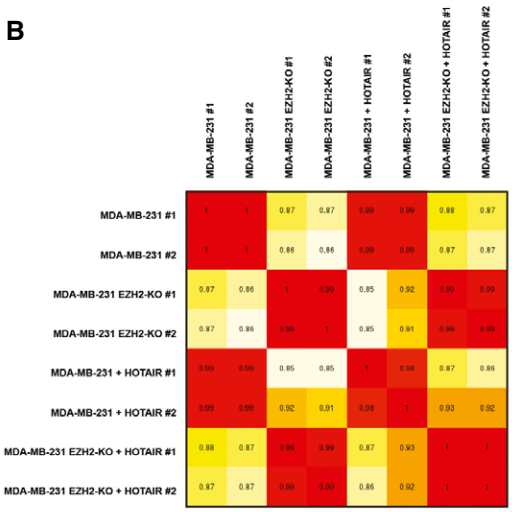
D Volcano plots representing gene expression change upon overexpression of *HOTAIR* in MDA-MB-231 EZH2++– and focusing on genes either previously reported to gain H3K27me3 upon overexpression of *HOTAIR* (left panel) or to be located with 100 kb of a binding sites for *HOTAIR* (right panel) (*y*-axis: $\log_{10}$ *P*-value, *x*-axis: $\log_2$ fold change). Red dots represent genes whose expression changes by more than twofold with a *P*-value < 0.05.

Source data are available online for this figure.

A



B



C

P value <	0.05	0.05	0.01	0.01
LogFC >	1	2	1	2
WT vs WT + HOTAIR	32	20	9	8
KO EZH2 vs KO EZH2 + HOTAIR	20	13	0	0
WT vs KO EZH2	1272	465	1209	463

D

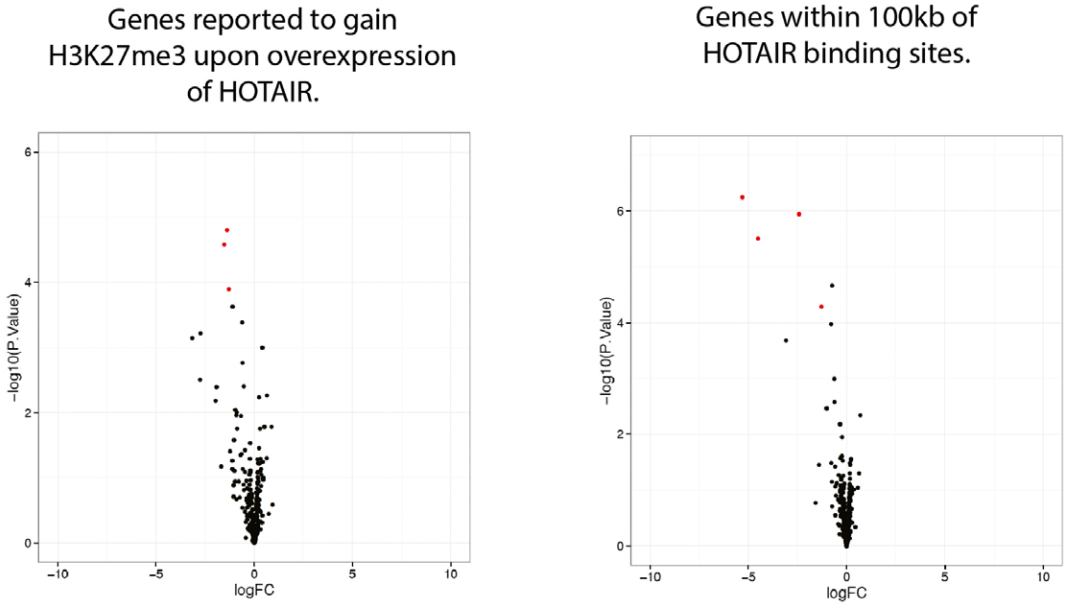


Figure EV1.

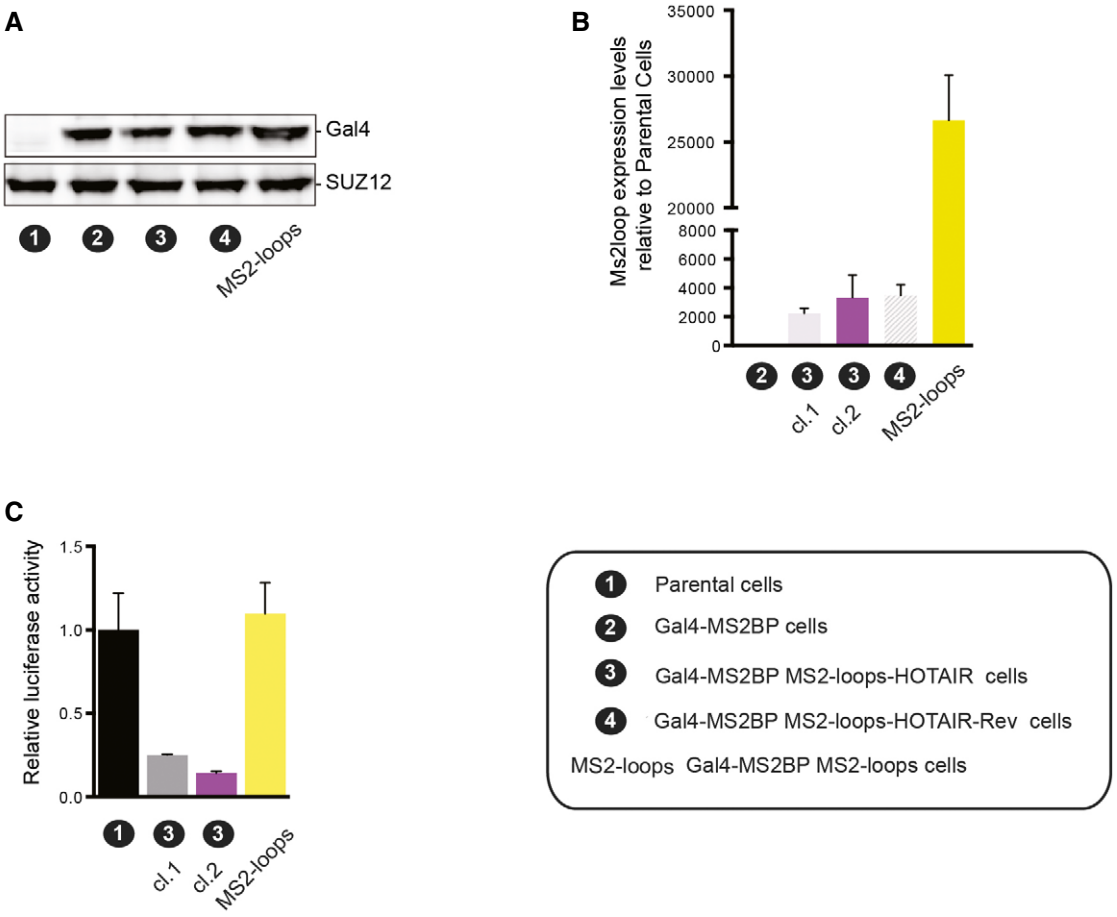


Figure EV2. Expression of Gal4-MS2CP fusion proteins and luciferase activity in clones overexpressing *HOTAIR*.

A Western blot analysis with Gal4 antibody on nuclear extracts from parental cell line (1) MS2BP, (2) MS2BP *MS2-HOTAIR* (3) or MS2BP *MS2-HOTAIR-Rev* (4) and MS2 loop cell lines. SUZ12 was used as a loading control.

B qRT–PCR to detect MS2-hybrid RNAs in the cell lines indicated on the *x*-axis. *Y*-axis represents MS2 loop RNA levels normalized to actin and calculated over the parental cells (mean ± SD, $n = 2$).

C Relative luciferase activity in the cell lines indicated on the *x*-axis. Values represent the relative luciferase activity normalized to the amount of protein (mean ± SD, $n = 3$).

Data information: Correspondence between numbers and model cell lines is indicated at the bottom right.
Source data are available online for this figure.

Figure EV3. Chromatin regulation upon recruitment of *HOTAIR*.

A ChIP experiments with H3K27me3 and primers located along the luciferase reporter (left) or control regions (MYT1 and ACT) (right) in the cell lines indicated on the *x*-axis of each graph (individual experiments and mean, $n = 2$). A H3K27me3 antibody of different origin was used in this assay as compared to Fig 3.

B ChIP experiments with H3K9me2 (top), H3K27ac (middle), or RNA polymerase II (bottom) in the cell lines indicated on the *x*-axis of each graph. Primers used are indicated in the color legend (individual experiments and mean, $n = 2$).

C Analysis of DNA methylation. Schematic representation of the portion of the reporter construct analyzed for DNA methylation is displayed on top. It comprises the Gal4-binding sites (5× UAS), the TK minimal promoter and part of the luciferase gene. DNA methylation analysis by bisulfite cloning is shown below in three different cell lines as indicated on the right. Black circles indicate methylated CpGs, and white circles indicate unmethylated CpGs. The percentage of methylated CpG is indicated below each condition.

Data information: Correspondence between numbers and model cell lines is indicated at the bottom right.
Source data are available online for this figure.

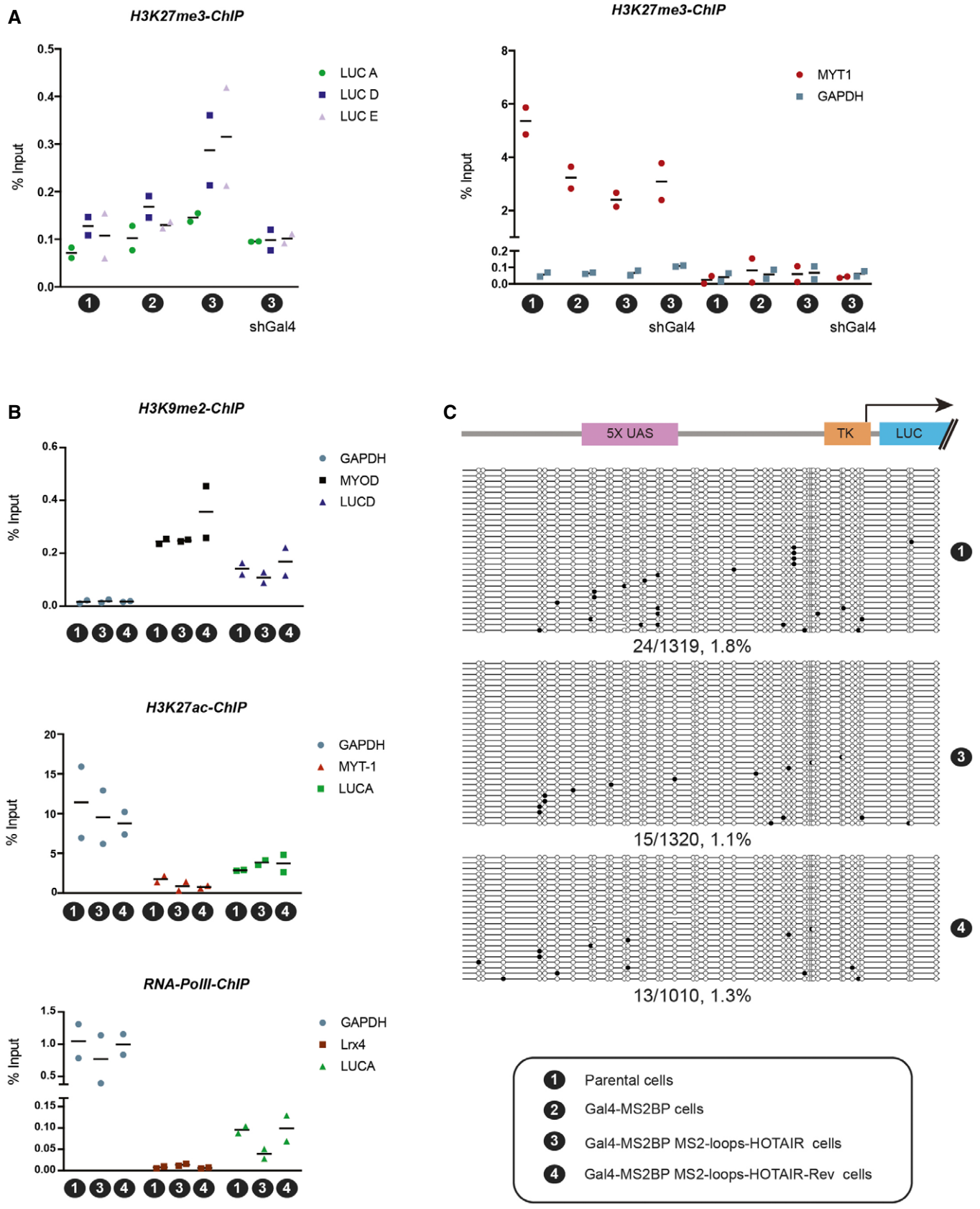


Figure EV3.

Figure EV4. PRC2 and *HOTAIR in vitro* production.

A Scheme for PRC2 production in Sf9 insect cells and purification over MiniQ column (left panel). Coomassie of eluted fractions. EZH2, SUZ12, EED-FLAG and RbAp48 proteins are indicated (right panel).

B Agarose gel showing 2.2-kb full-length *in vitro* transcribed *HOTAIR* and *HOTAIR-Rev* RNAs.

C Left: Scheme representing the interacting partners (chromatin and RNA) analyzed by sucrose gradient sedimentation on the right. Right: Native chromatin was incubated with or without biotinylated *HOTAIR/HOTAIR-Rev* RNAs, and chromatin conformation was analyzed by density gradient centrifugation.

D Left: Scheme representing the interacting partners (chromatin, RNA and PRC2) analyzed by sucrose gradient on the right. Right: PRC2 and native chromatin were incubated with or without biotinylated *HOTAIR/HOTAIR-Rev* RNAs and their profile analyzed by density gradient centrifugation.

Source data are available online for this figure.

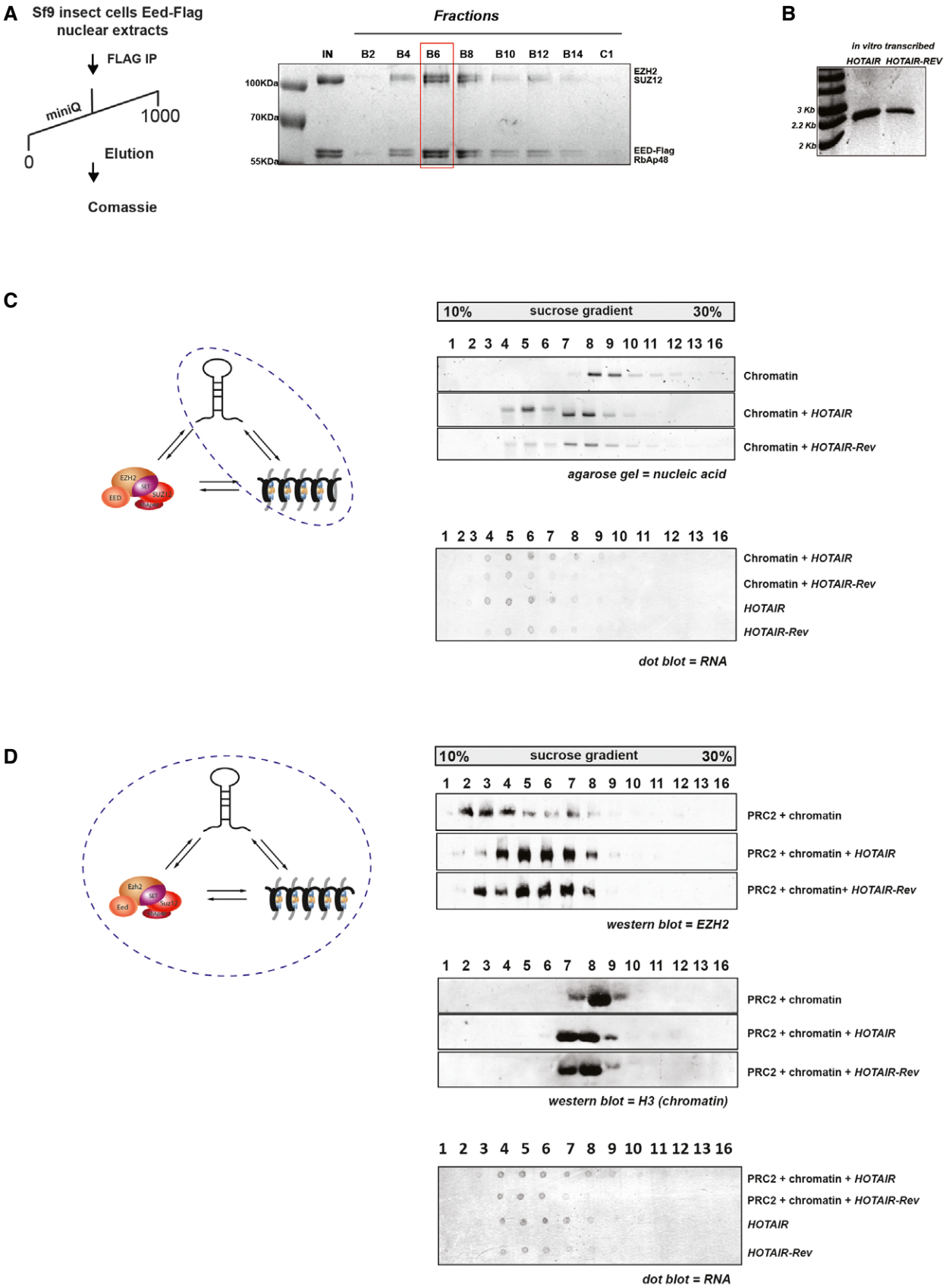


Figure EV4.

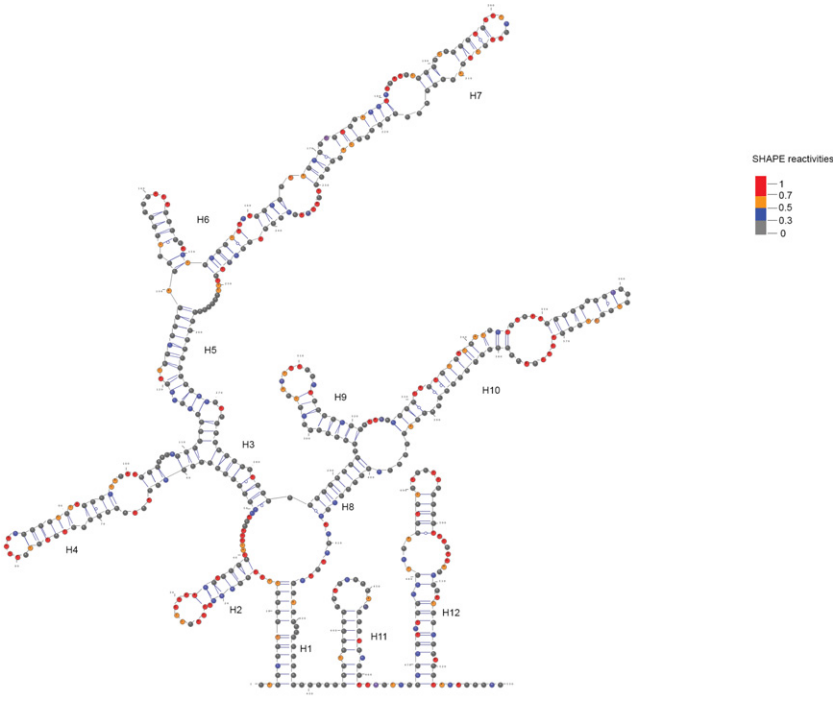


Figure EV5. *HOTAIR* domain 1 secondary structure.

Secondary structure model for the first 530 bp of *HOTAIR* RNA obtained from the RNAstructure software using the 1M7 reactivity map obtained in this study and displayed with predictive secondary structure. *HOTAIR* D1 domain as mapped with 1M7 SHAPE probing from our experiment is shown. SHAPE reactivity is depicted by colored nucleotides. Highly reactive nucleotides are displayed in red and orange, and low reactive nucleotides are displayed in black or blue according to the values reported in the legend.

Source data are available online for this figure.

PRC2 is dispensable for *HOTAIR*-mediated transcriptional repression.

Manuela Portoso, Roberta Ragazzini, Živa Brenčič, Arianna Moiani, Audrey Michaud, Ivaylo Vassilev, Michel Wassef, Nicolas Servant, Bruno Sargueil and Raphaël Margueron.

Appendix

Content:

Tables S1 to S3 and respective legend.

Sample	Library	Total reads	Mapped reads	% Mapped
B130T3_wt1	100 bp PE	103,335,602	77,922,669	75.41%
B130T4_wt2	100 bp PE	115,487,294	88,459,074	76.60%
B130T5_ezh2-/-1	100 bp PE	101,006,669	75,153,717	74.40%
B130T6_ezh2-/-2	100 bp PE	91,670,800	69,708,128	76.04%
B130T7_wt +hotair1	100 bp PE	90,196,578	67,304,498	74.62%
B130T8_wt +hotair2	100 bp PE	74,599,977	56,660,718	75.95%
B130T9_ezh2-/- +hotair1	100 bp PE	91,955,265	67,997,257	73.95%
B130T10_ezh2-/- + hotair 2	100 bp PE	88,585,080	66,982,306	75.61%

Appendix Table S1. RNA- sequencing library dataset.

Table indicating sequenced sample names, type of library, number of total sequenced reads, number of mapped reads to human hg19 genome and percentage of mapped reads.

qChIP primers	Forward	Reverse
LUC A	ATGGAAGACGCCAAAAACAT	GCCTTATGCAGTTGCTCTCC
LUC D	GTGTTGGGCGCGTTATTTAT	TACGGTAGGCTGCGAAATGT
LUC E	AACACCCCAACATCTTCGAC	TCGCGGTTGTTACTTGACTG
ACTIN	TGCTATCCCTGTACGCCTCT	CTCCTTAATGTCACGCACGA
MYT1	AGGCACCTTCTGTTGGCCGA	AGGCAGCTGCCTCCCGTACA
IRX4	AGCGTGTCCTTTCATAGGTC	TTTAGTCACCCGAGAACGAG
GAPDH	CGGGATTGTCTGCCCTAATTAT	GCACGGAAGGTCACGATGT
MYOD	TGCCCCTGACGCTCTAAGCTTTCT	AGCGGCTGTAGAAATCAGGTCGT

qRT-PCR primers	Forward	Reverse
HOTAIR	GGTAGAAAAAGCAACCACGAAGC	ACATAAACCTCTGTCTGTGAGTGCC
U1	ATACTTACCTGGCAGGGGAG	CAGGGGGAAAGCGCGAACGCA
MS2LOOP	TTCTGCAGATATCCAGCACAGT	CAACAGATGGCTGGCAACTAGAAG

bisulfite primers	Forward	Reverse
tkluc-bis-1 (1st PCR)	TAATTAATGTGAGTTAGTTTATTTATTAGG	TATAAAAACAATTATTCCAAAAACCAAAAC
tkluc-bis-2 (2nd PCR)	TGAGTTAGTTTATTTATTAGGTATTTTAGG	TCTTCATAACCTTATACAATTACTCTCC

Appendix Table S2. List of primers used.

Antibody	Reference	Assay
mGAL4 (DBD) RK5C1	Santa Cruz sc-510	WB
H3	Active motif 39163	WB
H3K36me3	Abcam ab9050	ChIP
EZH2	homemade	WB
GAL4	Millipore 06-262	ChIP, RIP
H3K27me3	Abcam ab6002	ChIP, WB
H3K27me3	Cell signalling 9733	ChIP
FLAG M2 beads	Sigma Aldrich A2220	IP
Lamin B1	Abcam ab16048	WB
SUZ12	homemade	WB
EED	homemade	WB
RNA-PolII (N20)	Santa Cruz Sc-899	ChIP
H3K9me2	Abcam ab1220	ChIP
H3K27ac	Abcam ab4729	ChIP

Appendix Table S3. List of antibodies used.

DISCUSSION

DISCUSSION AND PERSPECTIVES

The complexity of a multicellular organism comes from the presence of specialized cell types that paradoxically share an identical genome. Different cell types adopt a specific gene expression patterns that is maintained during many cell divisions. Polycomb Repressive Complex 2 plays a pivotal role in maintaining genes repressed through the deposition of a tri-methyl group on lysine 27 of Histone H3. The lack of a functional complex results in embryonic lethality in mice and in Drosophila (Raphaël Margueron & Reinberg, 2011; Pengelly, Copur, Jackle, Herzi, & Muller, 2013).

Intriguingly, PRC2 complex is rather ubiquitously expressed and its composition appears to not vary substantially while it maintains different set of genes repressed in distinct cell types. To date, it is still unclear how these different patterns are established in the first place and what controls PRC2 recruitment at different loci. These questions have retained the attention of many laboratories recently. It has been shown that facultative subunits associate with PRC2 and regulate it, that PRC2 is able to sense chromatin and that the complex can associate with RNA. However, several pieces are still missing to the puzzle.

In my main PhD project, I contributed to this field by looking for association of tissue-specific subunit to PRC2 in the gonads.

In addition to the core complex subunits, it has been shown that PRC2 associates to different accessory subunits in a substoichiometric manner. These subunits are important to regulate the catalytic activity of the complex and/or to guide its recruitment to target loci. These subunits have been studied separately, but recent comprehensive mass spectrometry analyses suggest that they can associate in mutually exclusive combinations forming two different sub-complexes (Hauri et al., 2016b; Kloet et al., 2016a). Yet, it is still difficult to conciliate the rather ubiquitous composition of PRC2 complex (core complex and cofactors) with the specificity of its targeting (cell specific gene targeting) in a variety of chromatin environment (stem cells, somatic cells, gametes…).

We speculated that tissue-specific subunits could associate with the complex in a particular context. To test this hypothesis, the aim of my PhD project was to characterize PRC2 interactome in mammalian tissues. Brain and gonads are two organs governed by finely-tuned epigenetic mechanisms. Recent work in the field of neurobiology has revealed that epigenetic processes are essential for complex brain functions. For example, formation of long-term memory requires epigenetic processes inducing lasting changes in gene expression. Mice lacking functional epigenetic component that contribute to these changes are impaired in long-term memory (Feng et al., 2010; Korzus, Rosenfeld, & Mayford, 2004). Hence, it has been proposed that PRC2 supports neuronal differentiation contributing to silencing of a transcriptional program that is necessary for adult neuron functions and survival (von Schimmelmann et al., 2016). Similarly, important chromatin regulations take place during male and female gametogenesis (Sasaki & Matsui, 2008). PRC2 function is crucial for this process since its ablation results in male infertility. In fact, Ezh1-and Ezh2- PRC2 role are

crucial to maintain spermatogonial stem cells and to repress somatic program (Mu et al., 2014, 2017). In contrast, the role of PRC2 in female gametogenesis is less clear to date.

Therefore, we explored PRC2 interactome in mice testis. Both EZH1 and EZH2 pulled-down the expected core complex subunits and most of the already known accessory subunits without distinguishing the two PRC2 subtypes. Considering that we looked at the whole testis, supposedly PRC2 complex could associate either with all of them or with specific combinations in different cell populations. More importantly, we identified an uncharacterized protein (AU022751) as interacting with both EZH1- and EZH2-PRC2. PRC2 stands out as the main partner of this new protein and of its human homolog even in a cellular context very different from the testis. This strengthen the importance of the interaction although other partners of AU022751 might also be important in the context of gonads. An interesting complementary approach would be to tag AU022751 protein in mice and perform a reverse IP in a physiological context to look for any other potential tissue-specific partner that could help to define is molecular function.

Au022751 gene is located on the X chromosome and codes for a fast evolving, poorly conserved protein with no characterized domains. The protein seems to be mainly found in *Eutheria*, placental mammals. We addressed question of its conservation through basic alignment in silico analysis. It would be very interesting to expend the analysis focusing on the origin and rapid divergence of AU022751/CXorf67 homolog across mammals. Interestingly, it seems that rapid evolution occurs in reproductive proteins sequences of several taxonomic groups (Swanson & Vacquier, 2002). The protein sequence is rich in serine and it has been predicted to be phosphorylated, suggesting the presence of post-translational regulation. Interestingly, alignment of the protein sequences revealed a motif of 13 amino acid in the C-terminal part of the protein which is conserved across species: We showed that this stretch is necessary to mediate PRC2 interaction. It suggests that the interaction with PRC2 is critical for the protein and that interaction could be relevant for transcriptional regulation.

In order to gain insight into the molecular mechanism of the protein, we first purified CXorf67 human homolog from bacterial system and we analyzed it on size-exclusion chromatography. The protein mainly formed oligomers and was quite unstable thus preventing us from further experiment. Therefore, we moved to baculovirus mediated expression in insect cells and we obtained a more stable monomeric protein to work with. We performed HKMT assay on both recombinant and native histones. While we observed stimulation of PRC2 catalytic activity incubating the wild type form of the protein with PRC2 on recombinant nucleosomes template, this effect is no more detectable when native chromatin was used as substrate. Therefore we aimed to determine how the protein can stimulate PRC2 catalytic activity. One possibility is that the protein interacts with nucleosomes. We attempted to perform nucleosome binding assay to test this hypothesis. We added the protein to a tagged-PRC2 on both recombinant and native nucleosomes and we immunoprecipitated in order to see if histones were differentially recovered once CXorf67 is added. The result of this experiment is not clear to date, an alternative approach might be required. It may not be excluded that the protein, which

is rich in basic aminoacid residues (pI= 10), could bind to a region that is depleted by nucleosomes. Sucrose gradient experiments and Electrophoretic Mobility Shif Assays (EMSA) could enable to determine whether the protein is able to bind DNA and whether it elutes either alone or in the same fraction as PRC2.

We set up also an artificial tethering model to assess CXorf67 effects on transcription in vivo as previously described (Margueron et al., 2008; Sanulli et al., 2015). The protein is able to drive transgene repression and, we showed by ChIP that EZH2 (data not yet shown) and H3K27me3 are brought to the transgene. Conversely, the mutated CXorf67 unable to interact with PRC2 does not induce transcriptional repression nor H3K27me3 transgene It appears consistent with the *in vitro* studies suggesting that AU022751/CXorf67 positively regulates PRC2 activity. Nevertheless, it should be considered that both approaches have limitations. For instance, the GAL4-system forces the recruitment of CXorf67 on a transgene, a situation that might not occur in an *in vivo* context. It therefore prompts for further investigations to determine what happen in a physiological context.

AU022751 and its human homolog CXorf67 are mainly expressed in gonads. We checked for their expression during early embryo development around the days in which primordial germ cells (PGCs) colonize the genital ridge starting their specification. Au022751/CXorf7 expression appears specifically upregulated in PGCs compared to somatic cells, while PRC2 expression remains almost unvaried. In the adult gonad, Au022751 is still expressed in the male germ line in the differentiated spermatogonia population and among developing oocyte in females. To bypass the lack of model cell line expressing Au022751/Cxorf67 (with the enigmatic exception of U2OS, human osteosarcoma derived cells), we generated a knockout model in order to investigate its function in a physiological context. We took advantage of the CRISPR/Cas9 system to delete the entire locus without affecting the neighbors genes. Knockout mice are viable with no major developmental defects. Since Au022751 is expressed in gonads, we mostly investigate its role in this physiological context. Male mice knockout for Au022751 display normal fertility although they have a global increase of H3K27me3 in germ cells as detected by western blot and by IF. To date, we still do not know which loci gained this mark. Ongoing RNA-seq and H3K27me3 ChIP-seq analysis will address these questions. We performed RNA-seq in the sorted spermatogenic populations that seems to express AU022751 protein, undifferentiated and differentiating spermatogonial cells. In the ongoing analysis, we will determine whether transcription is altered at any particular set of genes. In parallel, ChIP-seq analysis will allow us to determine on which loci H3K27me3 mark does increase. Considering the substantial increase of the mark (about 2 fold at the global level), it is unlikely to affect only a very specific set of genes, and it would not be surprising to observe that the mark spread into intergenic regions. As previously discussed, histones are gradually replaced by protamines during the last phases of spermatogenesis and only around 1% of nucleosomes remain on sperm chromatin. While nucleosomes retention has been associated with imprinted genes, microRNA clusters and Hox genes (Hammoud et al., 2009), other report suggested that sperm nucleosomes retention may take place in gene desert regions

(Samans et al., 2014). Since male mice are fertile, we expect that H3K27me3 increase only slightly impacts transcription during spermatogenesis. Nonetheless, if this excess H3K27me3 level is retained in sperm, it could have an impact on the next generations.

Interestingly, these data suggest that Au022751/Cxorf67 might behave as a negative regulator of PRC2 since in its absence PRC2 associated mark levels are globally higher than in WT conditions. Notably, H3K27me3 increase appears to be associated specifically with germ cells as shown by IF and by WB from in vitro derived spermatogonial culture nuclear extracts (data not shown). In contrast, H3K27me3 staining does not appear to vary in Sertoli cells. We are generating a germ-free male line Au022751-Dnmt3L knockout in order to confirm that H3K27me3 levels does not vary outside of germ cells.

AU022751 is largely expressed in all oocyte developmental phases and accordingly we reasoned that the protein could play an important role in female gametogenesis. Au022751 deletion has a strong impact on female fertility. We monitored fertility for over 20 weeks and showed that Au022751 knockout females are able to give birth only occasionally to one litter. Moreover, the size of the litter is however generally half the size of the wild type mice ones. Accordingly, the number of primordial follicles appears lower compared to the wild-type cells in 5-months old females. It we be interesting to define whether primordial follicles levels are lower at birth or whether the pool reduces at or after puberty in a hormone-dependent manner. We are also currently setting up the conditions to perform H3K27me3 staining in the oocytes in the late stages of oogenesis of Au022751 knockout females to assess if the absence of the protein leads to aberrant H3K27me3 deposition. Of note, we also noticed pregnancies prematurely terminating with apparent resorption of the embryos. We found a case where Au022751 knockout female display sign of aberrant embryo implantation. Cxorf67 homolog seems to strongly stain in trophoblastic cells in human placenta. These cells forms the outer layer of blastocyst which develop into placenta and they develop early being the first ones differentiating from the fertilized egg. Abnormal increase of H3K27me3 might be affecting a proper differentiation. If it is the case, in a more long-term perspective it would be crucial to characterize the transcriptome and H3K27me3 targets in absence of AU022751.

Unexpectedly, the human osteosarcoma cell line (U2OS) expresses CXorf67. Taking advantage of CRISPR/Cas9 technology, we knocked out the protein and we observed an increase of H3K27me3. Even if this cell line does not represent the best physiological model, the role of CXorf67 appears consistent with our *in vivo* observations and suggests that it is a relevant model for study. We aim to define PRC2 localization upon CXorf67 knockout and to determine whether any of the two PRC2 subcomplexes recently identified could be affected by the inactivation of CXorf67.

Altogether, our ongoing experiments should give a clear picture of how CXorf67/AU022751 regulate PRC2 function.

Another aspect of the protein should be further considered in a long term perspective. CXorf67 rearranges with MBTD1 gene in a chromosomal translocation in low-grade Endometrial stromal

sarcoma (ESS) (Dewaele et al., 2014). ESS is a gynaecological sarcoma that is composed of cells that resemble those of proliferative-phase endometrial stroma. Low-grade pathology frequently reports chromosomal translocations in which PRC2 genes rearrange with other proteins ((Lee & Nucci, 2014): 80% JAZF1-SUZ12, 6% JAZF1-PHF1, 4% EPC1-PHF1 and 3% MEAF6-PHF1). To date, this observation is limited to the clinical report of these translocations and it should be investigated what functional role it plays in the tumorigenesis. *JAZF1-SUZ12* translocation is one of the most abundant translocation and even if few reports attempted to unveil its function, it has been challenging to work in a cellular model close to ESS model and to have a clear picture of the translocation function (Koontz et al., 2001; Ma et al., 2017). We generated a fluorescent immortalized ESS model in vitro in which JAZF1-SUZ12 and MBTD1-CXORF67 fusion are stably expressed. It will be possible to play with this model in order to assess any ESS-associated phenotype as well as to characterize the molecular role of these translocations through proteomic, transcriptomic and genomic data.

Last, but not least in our studies, another important aspect that still remains elusive is the role of RNA in PRC2 regulation. LncRNAs take part to important cellular processes and some of them have been proposed to modulate chromatin modifying complexes activity. HOTAIR is expressed from the HoxC locus in human and mouse and it has been shown to repress HoxD locus in trans through a specific interaction with PRC2 by RNA immunoprecipitation and in vitro biochemical assay (Rinn J et al, 2007). However, discrepant phenotypes have been reported upon Hotair deletion in mice. One group reported dramatic effects deleting the two main exons of Hotair in mice (L. Li et al., 2013), while the same deletion analyzed by others did not show deleterious effects on mouse development (Amândio et al., 2016). Different phenotypical outputs have been attributed to either to the different mouse genetic background employed in the studies (proper backcrossing) or also to the fact that in one case, histologic sections at multiple stages have been analyzed whereas whole mounts at one time point was evaluated in the other. In addition, differences in transcriptomic changes have been associated with different models and tools used in the analysis. Li et colleagues looked at tail tip fibroblasts, an homogeneous cell population expressing Hotair. Instead, Armandio and collegues focused on E12.5 embryo fragments that is composed by a mix heterogeneous population expressing or not Hotair. Intriguingly, Duboule's group observed alterations in *Hoxc11* expression domain and found out new transcribed regions at the junction of the *Hotair* deletion. This further complicates the panorama of Hox genes regulation and may prompt for further studies to sort out the discrepancies (L. Li, Helms, & Chang, 2016). Our data also contrast with the phenotype previously described for HOTAIR overexpression in a human breast cancer cell line MDA-MB-231 in which it was reported to promote aberrant targeting of PRC2 and consequently inappropriate gene silencing (Gupta et al., 2010). In our hands, HOTAIR overexpression triggers few transcriptomic changes either in the presence or absence of PRC2. While the underlying mechanism explaining these discrepancies remain elusive, it prompts for further investigations. For that reason, we set up a HOTAIR specific tethering system in HEK293T

reporter cell line in order to underpin how *HOTAIR* RNA regulates transcription. When HOTAIR sense is tethered, it causes the repression of the transgene which is not observed upon HOTAIR antisense tethering. This excludes the possibility of a general interference with transcription machinery (i.e. steric hindrance). While specific tethering of HOTAIR sense slightly increases H3K27me3 deposition at the transgene, transgene repression is maintained even upon PRC2 deletion. This suggest that repression of the transgene occurs independently of H3K27me3 deposition, and that this chromatin modification may be a secondary consequence of transcriptional repression as previously suggested (Riising et al, 2014). In this model, other factors are likely involved in HOTAIR mediated transcriptional repression, it would be interesting to set up new approaches to identify them. At a methodological level, our approach constitutes a powerful tool to facilitate the analysis of a particular lncRNA effect to a single target site with defined components (Blanco & Guttman, 2017). In the past few years a similar approach was set up for ncRNA HOTTIP, a lincRNA transcribed from the 5′ tip of the *HOXA* locus that targets WDR5/MLL complexes leading to genes activation. Wang and colleagues set up a tethering system in which *BoxB*–HOTTIP specifically binds λN fused to GAL4 DNA binding domain, recruiting it to a UAS-luciferase reporter gene and leading to transgene transcriptional activation. Nevertheless, this approach shows some limitations: those constructs are transiently transfected and a proper negative control such as a non-specific RNA is missing (Wang et al., 2011). In our work, we could show that PRC2 binds to HOTAIR with similar affinity to other RNAs and that these interactions are not required to modulate gene expression. For instance, HOTAIR antisense RNA interacts in RNA immunoprecipitation (RIP) equally well with PRC2 but its artificial recruitment does not promote H3K27me3 deposition at the transgene. This observation actually raises question about the interpretation of RIP experiments. It would be interesting to set up a similar approach to investigate Xist mechanism of action and in particular the role of SPEN (Blanco & Guttman, 2017). It has been shown that PRC2 bind RNA promiscuously (Davidovich et al., 2013; Kaneko et al., 2013). It has also been observed that chromatin and RNA might compete for binding to PRC2 (Beltran et al, 2016). Supporting this hypothesis, we also observed that incubating PRC2 with an excess of RNA reduces its enzymatic activity on chromatin but not on another substrate, JARID2 (data not shown). Interestingly, other chromatin modifying complexes with distinct functions have been reported to be inhibited by RNA. Thus, it is tempting to speculate that it might represent a more general mechanism whereby the interaction between RNA and chromatin modifiers avoid non-specific binding to chromatin. In conclusion, a full understanding of the relationship between PRC2 and RNA will still require further investigation. Nonetheless, our work and the studies from several other labs call for caution regarding the simple model of lncRNA interacting with PRC2 to promote its targeting to chromatin and subsequent transcriptional silencing.

BIBLIOGRAPHY

Affar, E. B., Gay, F., Shi, Y., & Liu, H. (2006). Essential Dosage-Dependent Functions of the Transcription Factor Yin Yang 1 in Late Embryonic Development and Cell Cycle Progression Essential Dosage-Dependent Functions of the Transcription Factor Yin Yang 1 in Late Embryonic Development and Cell Cycle P. *Molecular and Cellular Biology*, *26*(9), 3565–3581. https://doi.org/10.1128/MCB.26.9.3565

Akam, M. (1987). The molecular basis for metameric pattern in the Drosophila embryo. *Development*, *101*(1), 1–22. https://doi.org/citeulike-article-id:10064978

Alekseyenko, A. a, Gorchakov, A. a, Kharchenko, P. V, & Kuroda, M. I. (2014a). Reciprocal interactions of human C10orf12 and C17orf96 with PRC2 revealed by BioTAP-XL cross-linking and affinity purification. *Proc Natl Acad Sci USA*, *111*(7), 2488–93. https://doi.org/10.1073/pnas.1400648111

Alekseyenko, A. a, Gorchakov, A. a, Kharchenko, P. V, & Kuroda, M. I. (2014b). Reciprocal interactions of human C10orf12 and C17orf96 with PRC2 revealed by BioTAP-XL cross-linking and affinity purification. *Proc Natl Acad Sci USA*, *111*(7), 2488–93. https://doi.org/10.1073/pnas.1400648111

Almeida, M., Pintacuda, G., Masui, O., Koseki, Y., Gdula, M., Cerase, A., … Brockdorff, N. (2017). PCGF3/5–PRC1 initiates Polycomb recruitment in X chromosome inactivation. *Science*, *356*(June), 1081–1084. https://doi.org/10.1126/science.aal2512

Almeida, M., Pintacuda, G., Masui, O., Koseki, Y., Gdula, M., Cerase, A., … Nesterova, T. B. (2017). PCGF3/5 – PRC1 initiates Polycomb recruitment in X chromosome inactivation, *1084*(June), 1081–1084.

Alonso, A. G. D. A., Gutiérrez, L., Fritsch, C., Papp, B., Beuchle, D., & Müller, J. (2007). A genetic screen identifies novel polycomb group genes in drosophila. *Genetics*, *176*(4), 2099–2108. https://doi.org/10.1534/genetics.107.075739

Amândio, A. R., Necsulea, A., Joye, E., Mascrez, B., & Duboule, D. (2016). Hotair Is Dispensible for Mouse Development. *PLOS Genetics*, *12*(12), e1006232. https://doi.org/10.1371/journal.pgen.1006232

Aranda, S., Mas, G., & Di Croce, L. (2015). Regulation of gene transcription by Polycomb proteins. *Science Advances*, *1*(11), e1500737. https://doi.org/10.1126/sciadv.1500737

Ballaré, C., Lange, M., Lapinaite, A., Martin, G. M., Morey, L., Pascual, G., … Di Croce, L. (2012). Phf19 links methylated Lys36 of histone H3 to regulation of Polycomb activity. *Nature Structural & Molecular Biology*, *19*(12), 1257–1265. https://doi.org/10.1038/nsmb.2434

Bannister, A. J., Zegerman, P., Partridge, J. F., Miska, E. a, Thomas, J. O., Allshire, R. C., & Kouzarides, T. (2001). Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature*, *410*(6824), 120–124. https://doi.org/10.1038/35065138

Bao, J., & Bedford, M. T. (2016). Epigenetic regulation of the histone-to-protamine transition during spermiogenesis. *Reproduction*, *151*(5), R55–R70. https://doi.org/10.1530/REP-15-0562

Bartke, T., Vermeulen, M., Xhemalce, B., Robson, S. C., Mann, M., & Kouzarides, T. (2010). Nucleosome-interacting proteins regulated by DNA and histone methylation. *Cell*, *143*(3), 470–484. https://doi.org/10.1016/j.cell.2010.10.012

Barton, S., MA, S., & Norris, M. (1984). Role of paternal and maternal genomes in mouse development. *Nature*, *311*(5984), 374–376.

Baymaz, H. I., Fournier, A., Laget, S., Ji, Z., Jansen, P. W. T. C., Smits, A. H., … Vermeulen, M. (2014). MBD5 and MBD6 interact with the human PR-DUB complex through their methyl-CpG-binding domain. *Proteomics*, *14*(19), 2179–2189. https://doi.org/10.1002/pmic.201400013

Beltran, M., Yates, C. M., Skalska, L., Dawson, M., Reis, F. P., Viiri, K., … Jenner, R. G. (2016). The interaction of PRC2 with RNA or chromatin is mutually antagonistic, 896–907. https://doi.org/10.1101/gr.197632.115.5

Beringer, M., Pisano, P., Di Carlo, V., Blanco, E., Chammas, P., Viz??n, P., … Di Croce, L. (2016). EPOP Functionally Links Elongin and Polycomb in Pluripotent Stem Cells. *Molecular Cell*, *64*(4), 645–658. https://doi.org/10.1016/j.molcel.2016.10.018

Bernstein, E., Duncan, E. M., Masui, O., Gil, J., Heard, E., & Allis, C. D. (2006). Mouse polycomb proteins bind differentially to methylated histone H3 and RNA and are enriched in facultative heterochromatin. *Molecular and Cellular Biology*, *26*(7), 2560–9.

https://doi.org/10.1128/MCB.26.7.2560-2569.2006
Blackledge, N. P., Farcas, A. M., Kondo, T., King, H. W., McGouran, J. F., Hanssen, L. L. P., … Klose, R. J. (2014a). Variant PRC1 complex-dependent H2A ubiquitylation drives PRC2 recruitment and polycomb domain formation. *Cell*, *157*(6), 1445–1459. https://doi.org/10.1016/j.cell.2014.05.004
Blackledge, N. P., Farcas, A. M., Kondo, T., King, H. W., McGouran, J. F., Hanssen, L. L. P., … Klose, R. J. (2014b). Variant PRC1 complex-dependent H2A ubiquitylation drives PRC2 recruitment and polycomb domain formation. *Cell*, *157*(6), 1445–1459. https://doi.org/10.1016/j.cell.2014.05.004
Blanco, M. R., & Guttman, M. (2017). lncRNA – Polycomb function, *36*(8), 964–966. https://doi.org/10.15252/embj.201796796
Boulard, M., Edwards, J. R., & Bestor, T. H. (2015). FBXL10 protects Polycomb-bound genes from hypermethylation. *Nature Genetics*, *47*(5), 1–9. https://doi.org/10.1038/ng.3272
Boulay, G., Rosnoblet, C., Guérardel, C., Angrand, P.-O., & Leprince, D. (2011). Functional characterization of human Polycomb-like 3 isoforms identifies them as components of distinct EZH2 protein complexes. *The Biochemical Journal*, *434*(2), 333–342. https://doi.org/10.1042/BJ20100944
Bourc'his, D., & Bestor, T. H. (2004). Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature*, *431*(7004), 96–99. https://doi.org/10.1038/nature02886
Bratzel, F., Lopez-Torrejon, G., Koch, M., Del Pozo, J. C., & Calonje, M. (2010). Keeping cell identity in arabidopsis requires PRC1 RING-finger homologs that catalyze H2A monoubiquitination. *Current Biology*, *20*(20), 1853–1859. https://doi.org/10.1016/j.cub.2010.09.046
Breiling, A., Bonte, E., Ferrari, S., Becker, P. B., & Paro, R. (1999). The Drosophila polycomb protein interacts with nucleosomal core particles In vitro via its repression domain. *Molecular and Cellular Biology*, *19*(12), 8451–60. https://doi.org/10.1128/MCB.19.12.8451
Brien, G. L., & Bracken, A. P. (2016). The PCL1-p53 axis promotes cellular quiescence. *Cell Cycle*, *15*(3), 305–306. https://doi.org/10.1080/15384101.2015.1124701
Brien, G. L., Gambero, G., O'Connell, D. J., Jerman, E., Turner, S. a, Egan, C. M., … Bracken, a P. (2012). Polycomb PHF19 binds H3K36me3 and recruits PRC2 and demethylase NO66 to embryonic stem cell genes during differentiation. *Nat Struct Mol Biol*, *19*(12), 1273–1281. https://doi.org/10.1038/nsmb.2449
Brien, G. L., Healy, E., Jerman, E., Conway, E., Fadda, E., Donovan, D. O., … Bracken, A. P. (2015). A chromatin-independent role of Polycomb-like 1 to stabilize p53 and promote cellular quiescence. *Genes & Development*, *29*(21), 1–14. https://doi.org/10.1101/gad.267930.115.7
Brinkman, A. B., Gu, H., Bartels, S. J. J., Zhang, Y., Matarese, F., Simmer, F., … Stunnenberg, H. G. (2012). Sequential ChIP-bisulfite sequencing enables direct genome-scale investigation of chromatin and DNA methylation cross-talk. *Genome Research*, *22*(6), 1128–1138. https://doi.org/10.1101/gr.133728.111
Brykczynska, U., Hisano, M., Erkek, S., Ramos, L., Oakeley, E. J., Roloff, T. C., … Peters, A. H. (2010). Repressive and active histone methylation mark distinct promoters in human and mouse spermatozoa. *Nat Struct Mol Biol*, *17*(6), 679–687. https://doi.org/10.1038/nsmb.1821
Bunker, C. A., & Kingston, R. E. (1994). Transcriptional repression by Drosophila and mammalian Polycomb group proteins in transfected mammalian cells. *Mol Cell Biol*, *14*(3), 1721–1732. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7906858%5Cnhttp://www.ncbi.nlm.nih.gov/pmc/articles/PMC358530/pdf/molcellb00003-0203.pdf
Cai, L., Rothbart, S. B., Lu, R., Xu, B., Chen, W. Y., Tripathy, A., … Wang, G. G. (2013). An H3K36 Methylation-Engaging Tudor Motif of Polycomb-like Proteins Mediates PRC2 Complex Targeting. *Molecular Cell*, *49*(3), 571–582. https://doi.org/10.1016/j.molcel.2012.11.026
Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., … Yi Zhang1. (2002). Role of Histone H3 Lysine 27 Methylation in X Inactivation. *Science*, *298*(November), 1039–1. https://doi.org/10.1126/science.1084274
Cao, R., & Zhang, Y. (2004). The functions of E(Z)/EZH2-mediated methylation of lysine 27 in histone H3. *Current Opinion in Genetics and Development*, *14*(2), 155–164.

https://doi.org/10.1016/j.gde.2004.02.001
Carone, B. R., Hung, J. H., Hainer, S. J., Chou, M. Te, Carone, D. M., Weng, Z., … Rando, O. J. (2014). High-resolution mapping of chromatin packaging in mouse embryonic stem cells and sperm. *Developmental Cell*, *30*(1), 11–22. https://doi.org/10.1016/j.devcel.2014.05.024
Casanova, M., Preissner, T., Cerase, A., Poot, R., Yamada, D., Li, X., … Brockdorff, N. (2011). Polycomblike 2 facilitates the recruitment of PRC2 Polycomb group complexes to the inactive X chromosome and to target loci in embryonic stem cells. *Development (Cambridge, England)*, *138*(8), 1471–1482. https://doi.org/10.1242/dev.053652
Cavalli, G., & Paro, R. (1999). Epigenetic Inheritance of Active Chromatin After Removal of the Main Transactivator. *Science*, *286*(5441), 955–958. https://doi.org/10.1126/science.286.5441.955
Cerase, A., Smeets, D., Tang, Y. A., Gdula, M., Kraus, F., Spivakov, M., & Moindrot, B. (2013). Spatial separation of Xist RNA and polycomb proteins revealed by superresolution microscopy, 1–6. https://doi.org/10.1073/pnas.1312951111
Chanvivattana, Y., Bishopp, A., Schubert, D., Stock, C., Moon, Y.-H., Sung, Z. R., & Goodrich, J. (2004). Interaction of Polycomb-group proteins controlling flowering in Arabidopsis. *Development (Cambridge, England)*, *131*(21), 5263–76. https://doi.org/10.1242/dev.01400
Cho, C., Willis, W. D., Goulding, E. H., Jung-Ha, H., Choi, Y. C., Hecht, N. B., & Eddy, E. M. (2001). Haploinsufficiency of protamine-1 or -2 causes infertility in mice. *Nature Genetics*, *28*(may), 82–86. https://doi.org/10.1038/ng0501-82
Chu, C., Qu, K., Zhong, F. L., Artandi, S. E., & Chang, H. Y. (2011). Resource Genomic Maps of Long Noncoding RNA Occupancy Reveal Principles of RNA-Chromatin Interactions. *Molecular Cell*, *44*(4), 667–678. https://doi.org/10.1016/j.molcel.2011.08.027
Cooper, S., Dienstbier, M., Hassan, R., Schermelleh, L., Sharif, J., Blackledge, N., … Brockdorff, N. (2014). Targeting Polycomb to Pericentric Heterochromatin in Embryonic Stem Cells Reveals a Role for H2AK119u1 in PRC2 Recruitment. *Cell Reports*, *7*(5), 1456–1470. https://doi.org/10.1016/j.celrep.2014.04.012
Cooper, S., Grijzenhout, A., Underwood, E., Ancelin, K., Zhang, T., Nesterova, T. B., … Bartke, T. (2016). Jarid2 binds mono-ubiquitylated H2A lysine 119 to mediate crosstalk between Polycomb complexes PRC1 and PRC2. *Nature Communications*, *7*, 13661. https://doi.org/10.1038/ncomms13661
Corbineau, S., Lassalle, B., Givelet, M., Souissi-Sarahoui, I., Firlej, V., Henri Romeo, P., … Fouchet, P. (2016). Spermatogonial stem cells and progenitors are refractory to reprogramming to pluripotency by the transcription factors Oct3/4, c-Myc, Sox2 and Klf4. *Oncotarget*, *8*(6), 1–14. https://doi.org/10.18632/oncotarget.14327
Creppe, C., Palau, A., Malinverni, R., Valero, V., & Buschbeck, M. (2014). A Cbx8-Containing Polycomb Complex Facilitates the Transition to Gene Activation during ES Cell Differentiation. *PLoS Genetics*, *10*(12). https://doi.org/10.1371/journal.pgen.1004851
Culi, J., Aroca, P., Modolell, J., & Mann, R. S. (2006). jing is required for wing development and to establish the proximo-distal axis of the leg in Drosophila melanogaster. *Genetics*, *173*(1), 255–266. https://doi.org/10.1534/genetics.106.056341
Czermin, B., Melfi, R., McCabe, D., Seitz, V., Imhof, A., & Pirrotta, V. (2002). Drosophila enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites. *Cell*, *111*(2), 185–196. https://doi.org/10.1016/S0092-8674(02)00975-3
da Rocha, S. T., Boeva, V., Escamilla-Del-Arenal, M., Ancelin, K., Granier, C., Matias, N. R., … Heard, E. (2014). Jarid2 Is Implicated in the Initial Xist-Induced Targeting of PRC2 to the Inactive X Chromosome. *Molecular Cell*, *53*(2), 301–316. https://doi.org/10.1016/j.molcel.2014.01.002
Dauber, K. L., Perdigoto, C. N., Valdes, V. J., Santoriello, F. J., Cohen, I., & Ezhkova, E. (2016). Dissecting the Roles of Polycomb Repressive Complex 2 Subunits in the Control of Skin Development. *Journal of Investigative Dermatology*, *136*(8), 1647–1655. https://doi.org/10.1016/j.jid.2016.02.809
Davidovich, C., & Cech, T. R. (2015). The recruitment of chromatin modifiers by long noncoding RNAs : lessons from PRC2, 2007–2022. https://doi.org/10.1261/rna.053918.115.Kuzmichev
Davidovich, C., Zheng, L., Goodrich, K. J., & Cech, T. R. (2013). Promiscuous RNA binding by Polycomb repressive complex 2. *Nature Structural & Molecular Biology*, *20*(11), 1250–

1257. https://doi.org/10.1038/nsmb.2679
De Cegli, R., Iacobacci, S., Flore, G., Gambardella, G., Mao, L., Cutillo, L., … Di Bernardo, D. (2013). Reverse engineering a mouse embryonic stem cell-specific transcriptional network reveals a new modulator of neuronal differentiation. *Nucleic Acids Research*, *41*(2), 711–726. https://doi.org/10.1093/nar/gks1136
De Lucia, F., Crevillen, P., Jones, A. M., Greb, T., & Dean, C. (2008). A PHD-polycomb repressive complex 2 triggers the epigenetic silencing of FLC during vernalization. *Proc Natl Acad Sci U S A*, *105*(44), 16831–16836. https://doi.org/10.1073/pnas.0808687105
DeBaun, M. R., Niemitz, E. L., & Feinberg, A. P. (2003). Association of In Vitro Fertilization with Beckwith-Wiedemann Syndrome and Epigenetic Alterations of LIT1 and H19. *The American Journal of Human Genetics*, *72*(1), 156–160. https://doi.org/10.1086/346031
del Mar Lorente, M., Marcos-Gutierrez, C., Perez, C., Schoorlemmer, J., Ramirez, A., Magin, T., & Vidal, M. (2000). Loss- and gain-of-function mutations show a polycomb group function for Ring1A in mice. *Development*, *127*(23), 5093–5100. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=11060235%5Cnhttp://dev.biologists.org/content/127/23/5093.full.pdf
Dewaele, B., Przybyl, J., Quattrone, A., Finalet Ferreiro, J., Vanspauwen, V., Geerdens, E., … Debiec-Rychter, M. (2014). Identification of a novel, recurrent MBTD1-CXorf67 fusion in low-grade endometrial stromal sarcoma. *International Journal of Cancer*, *134*(5), 1112–1122. https://doi.org/10.1002/ijc.28440
Dietrich, N., Lerdrup, M., Landt, E., Agrawal-Singh, S., Bak, M., Tommerup, N., … Hansen, K. (2012). REST-mediated recruitment of polycomb repressor complexes in mammalian cells. *PLoS Genetics*, *8*(3). https://doi.org/10.1371/journal.pgen.1002494
Djebali, S., Davis, C. A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., … Skancke, J. (2012). Landscape of transcription in human cells, 2–9. https://doi.org/10.1038/nature11233
Dumesic, P. A., Homer, C. M., Moresco, J. J., Pack, L. R., Shanle, E. K., Coyle, S. M., … Madhani, H. D. (2015). Product binding enforces the genomic specificity of a yeast Polycomb repressive complex. *Cell*, *160*(1–2), 204–218. https://doi.org/10.1016/j.cell.2014.11.039
Edwards, J. R., Yarychkivska, O., Boulard, M., & Bestor, T. H. (2017). DNA methylation and DNA methyltransferases. *Epigenetics & Chromatin*, *10*, 1–10. https://doi.org/10.1186/s13072-017-0130-8
Endoh, M., Endo, T. A., Endoh, T., Isono, K. ichi, Sharif, J., Ohara, O., … Koseki, H. (2012). Histone H2A mono-ubiquitination is a crucial step to mediate PRC1-dependent repression of developmental genes to maintain ES cell identity. *PLoS Genetics*, *8*(7). https://doi.org/10.1371/journal.pgen.1002774
Endoh, M., Endo, T. A., Shinga, J., Hayashi, K., Farcas, A., Ma, K. W., … Koseki, H. (2017). PCGF6-PRC1 suppresses premature differentiation of mouse embryonic stem cells by regulating germ cell-related genes. *eLife*, *6*, 1–26. https://doi.org/10.7554/eLife.21064
Erhardt, S., Su, I.-H., Schneider, R., Barton, S., Bannister, A. J., Perez-Burgos, L., … Surani, M. A. (2003). Consequences of the depletion of zygotic and embryonic enhancer of zeste 2 during preimplantation mouse development. *Development*, *130*(18), 4235–4248. https://doi.org/10.1242/dev.00625
Eskeland, R., Leeb, M., Grimes, G. R., Kress, C., Boyle, S., Sproul, D., … Bickmore, W. A. (2010). Ring1B Compacts Chromatin Structure and Represses Gene Expression Independent of Histone Ubiquitination. *Molecular Cell*, *38*(3), 452–464. https://doi.org/10.1016/j.molcel.2010.02.032
Farcas, A. M., Blackledge, N. P., Sudbery, I., Long, H. K., McGouran, J. F., Rose, N. R., … Klose, R. J. (2012). KDM2B links the polycomb repressive complex 1 (PRC1) to recognition of CpG islands. *eLife*, *2012*(1), 1–26. https://doi.org/10.7554/eLife.00205
Faust, C., Lawson, K. A., Schork, N. J., Thiel, B., & Magnuson, T. (1998). The Polycomb-group gene eed is required for normal morphogenetic movements during gastrulation in the mouse embryo. *Development*, *125*(22), 4495–4506. Retrieved from http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=9778508&retmode=ref&cmd=prlinks%5Cnpapers3://publication/uuid/C9508FC5-94D5-4C7C-9B73-98964F0E8AEB
Feng, J., Zhou, Y., Campbell, S. L., Le, T., Li, E., Sweatt, J. D., … Fan, G. (2010). Dnmt1 and Dnmt3a maintain DNA methylation and regulate synaptic function in adult forebrain neurons.

Nature Neuroscience, *13*(4), 423–30. https://doi.org/10.1038/nn.2514
Ferrari, K. J., Scelfo, A., Jammula, S., Cuomo, A., Barozzi, I., Stützer, A., … Pasini, D. (2014). Polycomb-Dependent H3K27me1 and H3K27me2 Regulate Active Transcription and Enhancer Fidelity. *Molecular Cell*, *53*(1), 49–62. https://doi.org/10.1016/j.molcel.2013.10.030
Fischle, W., Wang, Y., Jacobs, S. A., Kim, Y., Allis, C. D., & Khorasanizadeh, S. (2003). Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by polycomb and HP1 chromodomains. *Genes and Development*, *17*(15), 1870–1881. https://doi.org/10.1101/gad.1110503
Francis, N. J., Saurin, A. J., Shao, Z., & Kingston, R. E. (2001). Reconstitution of a functional core polycomb repressive complex. *Molecular Cell*, *8*(3), 545–556. https://doi.org/10.1016/S1097-2765(01)00316-1
Frangini, A., Sjöberg, M., Roman-Trufero, M., Dharmalingam, G., Haberle, V., Bartke, T., … Dillon, N. (2013). The Aurora B Kinase and the Polycomb Protein Ring1B Combine to Regulate Active Promoters in Quiescent Lymphocytes. *Molecular Cell*, *51*(5), 647–661. https://doi.org/10.1016/j.molcel.2013.08.022
Gaell Mainguy, Koster, J., Woltering1, J., Jansen, H., & Durston, A. (2007). Extensive Polycistronism and Antisense Transcription in the Mammalian Hox Clusters, (4), 1–7. https://doi.org/10.1371/journal.pone.0000356
Gao, Z., Lee, P., Stafford, J. M., von Schimmelmann, M., Schaefer, A., & Reinberg, D. (2014). An AUTS2-Polycomb complex activates gene expression in the CNS. *Nature*, *516*(7531), 349–354. https://doi.org/10.1038/nature13921
Gao, Z., Zhang, J., Bonasio, R., Strino, F., Sawai, A., Parisi, F., … Reinberg, D. (2012). PCGF Homologs, CBX Proteins, and RYBP Define Functionally Distinct PRC1 Family Complexes. *Molecular Cell*, *45*(3), 344–356. https://doi.org/10.1016/j.molcel.2012.01.002
Gearhart, M. D., Corcoran, C. M., Wamstad, J. A., & Bardwell, V. J. (2006). Polycomb group and SCF ubiquitin ligases are found in a novel BCOR complex that is recruited to BCL6 targets. *Mol Cell Biol*, *26*(18), 6880–6889. https://doi.org/10.1128/MCB.00630-06
Geisler, S., & Coller, J. (2013). RNA in unexpected places: long non-coding RNA functions in diverse cellular contexts. *Nature Reviews. Molecular Cell Biology*, *14*(11), 699–712. https://doi.org/10.1038/nrm3679
Gil, J., Bernard, D., Martínez, D., & Beach, D. (2004). Polycomb CBX7 has a unifying role in cellular lifespan. *Nature Cell Biology*, *6*(1), 67–72. https://doi.org/10.1038/ncb1077
Grabole, N., Tischler, J., Hackett, J. A., Kim, S., Tang, F., Leitch, H. G., … Surani, M. A. (2013). Prdm14 promotes germline fate and naive pluripotency by repressing FGF signalling and DNA methylation. *Nature Publishing Group*, *14*(7), 629–63767. https://doi.org/10.1038/embor.2013.67
Grau, D. J., Chapman, B. A., Garlick, J. D., Borowsky, M., Francis, N. J., & Kingston, R. E. (2011). Compaction of chromatin by diverse polycomb group proteins requires localized regions of high charge. *Genes and Development*, *25*(20), 2210–2221. https://doi.org/10.1101/gad.17288211
Grijzenhout, A., Godwin, J., Koseki, H., Gdula, M., Szumska, D., McGouran, J. F., … Cooper, S. (2016a). Functional analysis of AEBP2, a PRC2 Polycomb protein, reveals a Trithorax phenotype in embryonic development and in ES cells. *Development (Cambridge, England)*, 2716–2723. https://doi.org/10.1242/dev.123935
Grijzenhout, A., Godwin, J., Koseki, H., Gdula, M., Szumska, D., McGouran, J. F., … Cooper, S. (2016b). Functional analysis of AEBP2, a PRC2 Polycomb protein, reveals a Trithorax phenotype in embryonic development and in ES cells. *Development (Cambridge, England)*, 2716–2723. https://doi.org/10.1242/dev.123935
Gupta, R. a, Shah, N., Wang, K. C., Kim, J., Horlings, H. M., Wong, D. J., … Chang, H. Y. (2010). Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature*, *464*(7291), 1071–1076. https://doi.org/10.1038/nature08975
Gutierrez, L., Oktaba, K., Scheuermann, J. C., Gambetta, M. C., Ly-Hartig, N., & Muller, J. (2012). The role of the histone H2A ubiquitinase Sce in Polycomb repression. *Development*, *139*(1), 117–127. https://doi.org/10.1242/dev.074450
Guttman, M., Amit, I., Garber, M., French, C., Lin, M. F., Feldser, D., … Lander, E. S. (2009). Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in

mammals. *Nature*, *457*(7235), 223–227. https://doi.org/10.1038/nature07672
Guttman, M., Donaghey, J., Carey, B. W., Garber, M., Grenier, J. K., Munson, G., … Lander, E. S. (2011). lincRNAs act in the circuitry controlling pluripotency and differentiationfile:///Users/RMargueron/Desktop/mercer mattick dinger 2010.pdf. *Nature*, *477*(7364), 295–300. https://doi.org/10.1038/nature10398
Hammoud, S. S., Nix, D. A., Zhang, H., Purwar, J., Carrell, D. T., & Cairns, B. R. (2009). Distinctive chromatin in human sperm packages genes for embryo development. *Nature*, *460*(7254), 473–478. https://doi.org/10.1038/nature08162
Hauri, S., Comoglio, F., Seimiya, M., Gerstung, M., Glatter, T., Hansen, K., … Beisel, C. (2016a). A High-Density Map for Navigating the Human Polycomb Complexome. *Cell Reports*, *17*(2), 583–595. https://doi.org/10.1016/j.celrep.2016.08.096
Hauri, S., Comoglio, F., Seimiya, M., Gerstung, M., Glatter, T., Hansen, K., … Beisel, C. (2016b). A High-Density Map for Navigating the Human Polycomb Complexome. *Cell Reports*, *17*(2), 583–595. https://doi.org/10.1016/j.celrep.2016.08.096
Hauri, S., Comoglio, F., Seimiya, M., Gerstung, M., Glatter, T., Hansen, K., … Beisel, C. (2016c). A high density map for navigating the human Polycomb complexome. *bioRxiv*, *17*(2), 59964. https://doi.org/10.1101/059964
He, G.-P., Kim, S., & Ro, H. S. (1999). Cloning and characterization of a novel zinc finger gene Transcriptional repressor. *Journal of Biological Chemistry*, *274*(21), 14678–14684. https://doi.org/10.1006/geno.1995.0031
He, J., Shen, L., Wan, M., Taranova, O., Wu, H., & Zhang, Y. (2013). Kdm2b maintains murine embryonic stem cell status by recruiting PRC1 complex to CpG islands of developmental genes. *Nat Cell Biol*, *15*(4), 373–384. https://doi.org/10.1038/ncb2702
Henriquez, B., Bustos, F. J., Aguilar, R., Becerra, A., Simon, F., Montecino, M., & van Zundert, B. (2013). Ezh1 and Ezh2 differentially regulate PSD-95 gene transcription in developing hippocampal neurons. *Molecular and Cellular Neuroscience*, *57*, 130–143. https://doi.org/10.1016/j.mcn.2013.07.012
Herranz, N., Pasini, D., Díaz, V. M., Francí, C., Gutierrez, A., Dave, N., … Peiró, S. (2008). Polycomb complex 2 is required for E-cadherin repression by the Snail1 transcription factor. *Molecular and Cellular Biology*, *28*(15), 4772–81. https://doi.org/10.1128/MCB.00323-08
Hidalgo, I., Herrera-Merchan, A., Ligos, J. M., Carramolino, L., Nuñez, J., Martinez, F., … Gonzalez, S. (2012). Ezh1 is required for hematopoietic stem cell maintenance and prevents senescence-like cell cycle arrest. *Cell Stem Cell*, *11*(5), 649–662. https://doi.org/10.1016/j.stem.2012.08.001
Hinkins, M., Huntriss, J., Miller, D., & Picton, H. M. (2005). Expression of Polycomb-group genes in human ovarian follicles, oocytes and preimplantation embryos. *Reproduction*, *130*(6), 883–888. https://doi.org/10.1530/rep.1.00675
Holoch, D., & Margueron, R. (2017a). Mechanisms Regulating PRC2 Recruitment and Enzymatic Activity. *Trends in Biochemical Sciences*, *42*(7), 531–542. https://doi.org/10.1016/j.tibs.2017.04.003
Holoch, D., & Margueron, R. (2017b). Mechanisms Regulating PRC2 Recruitment and Enzymatic Activity. *Trends in Biochemical Sciences*, *42*(7), 531–542. https://doi.org/10.1016/j.tibs.2017.04.003
Hrzenjak, A. (2016). JAZF1/SUZ12 gene fusion in endometrial stromal sarcomas. *Orphanet Journal of Rare Diseases*, *11*, 15. https://doi.org/10.1186/s13023-016-0400-8
Hunkapiller, J., Shen, Y., Diaz, A., Cagney, G., McCleary, D., Ramalho-Santos, M., … Reiter, J. F. (2012). Polycomb-like 3 promotes polycomb repressive complex 2 binding to CpG islands and embryonic stem cell self-renewal. *PLoS Genetics*, *8*(3). https://doi.org/10.1371/journal.pgen.1002576
Illingworth, R. S., Moffat, M., Mann, A. R., Read, D., Hunter, C. J., Pradeepa, M. M., … Bickmore, W. A. (2015). The E3 ubiquitin ligase activity of RING1B is not essential for early mouse development1. Illingworth, R. S. et al. The E3 ubiquitin ligase activity of RING1B is not essential for early mouse development. Genes Dev. 1897–1902 (2015). doi:10.1101/gad.26815. *Genes and Development*, 1897–1902. https://doi.org/10.1101/gad.268151.115.
Iwamori, N., Iwamori, T., & Matzuk, M. M. (2013). H3K27 Demethylase , JMJD3 , Regulates Fragmentation of Spermatogonial Cysts, *8*(8), 1–14.

https://doi.org/10.1371/journal.pone.0072689
Jensen, D., Proctor, M., Marquis, S., Gardner, H., Ha, S., Chodosh, L., … FJ, R. (1998). BAP1: a novel ubiquitin hydrolase which binds to the BRCA1 RING finger and enhances BRCA1-mediated cell growth suppression. *Oncogene*, *16*(9), 1097–112.
Jermann, P., Hoerner, L., Burger, L., & Schübeler, D. (2014a). Short sequences can efficiently recruit histone H3 lysine 27 trimethylation in the absence of enhancer activity and DNA methylation. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(33), E3415-21. https://doi.org/10.1073/pnas.1400672111
Jermann, P., Hoerner, L., Burger, L., & Schübeler, D. (2014b). Short sequences can efficiently recruit histone H3 lysine 27 trimethylation in the absence of enhancer activity and DNA methylation. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(33), E3415-21. https://doi.org/10.1073/pnas.1400672111
Jones, R. S., & Gelbart, W. M. (1993). The Drosophila Polycomb-group gene Enhancer of zeste contains a region with sequence similarity to trithorax. *Molecular and Cellular Biology*, *13*(10), 6357–66. https://doi.org/10.1128/MCB.13.10.6357.Updated
Jung, J., Mysliwiec, M. R., & Lee, Y. (2005). Roles of JUMONJI in mouse embryonic development. *Developmental Dynamics*, *232*(1), 21–32. https://doi.org/10.1002/dvdy.20204
Kalb, R., Latwiel, S., Baymaz, H. I., Jansen, P. W. T. C., Müller, C. W., Vermeulen, M., & Müller, J. (2014a). Histone H2A monoubiquitination promotes histone H3 methylation in Polycomb repression. *Nature Structural & Molecular Biology*, *21*(6), 569–71. https://doi.org/10.1038/nsmb.2833
Kalb, R., Latwiel, S., Baymaz, H. I., Jansen, P. W. T. C., Müller, C. W., Vermeulen, M., & Müller, J. (2014b). Histone H2A monoubiquitination promotes histone H3 methylation in Polycomb repression. *Nature Structural & Molecular Biology*, *21*(6), 569–71. https://doi.org/10.1038/nsmb.2833
Kaneko, S., Bonasio, R., Saldaña-Meyer, R., Yoshida, T., Son, J., Nishino, K., … Reinberg, D. (2014). Interactions between JARID2 and Noncoding RNAs Regulate PRC2 Recruitment to Chromatin. *Molecular Cell*, *53*(2), 290–300. https://doi.org/10.1016/j.molcel.2013.11.012
Kaneko, S., Son, J., Bonasio, R., Shen, S. S., & Reinberg, D. (2014). Nascent RNA interaction keeps PRC2 activity poised and in check, 1983–1988. https://doi.org/10.1101/gad.247940.114.GENES
Kaneko, S., Son, J., Shen, S. S., Reinberg, D., & Bonasio, R. (2013). PRC2 binds active promoters and contacts nascent RNAs in embryonic stem cells. *Nature Structural & Molecular Biology*, *20*(11), 1258–64. https://doi.org/10.1038/nsmb.2700
Kanhere, A., Viiri, K., Araujo, C. C., Rasaiyaah, J., Bouwman, R. D., Whyte, W. A., … Jenner, R. G. (2010). Article Short RNAs Are Transcribed from Repressed Polycomb Target Genes and Interact with Polycomb Repressive Complex-2, 675–688. https://doi.org/10.1016/j.molcel.2010.03.019
Kennison, J. A., & Tamkunt, J. W. (1988). Dosage-dependent modifiers of Polycomb and Antennapedia mutations in Drosophila (homoeotic). *Genetics*, *85*(November), 8136–8140. https://doi.org/10.1073/pnas.85.21.8136
Khalil, A. M., Guttman, M., Huarte, M., Garber, M., Raj, A., Rivea Morales, D., … Rinn, J. L. (2009). Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(28), 11667–11672. https://doi.org/10.1073/pnas.0904715106
Kim, H., Ekram, M. B., Bakshi, A., & Kim, J. (2015). AEBP2 as a transcriptional activator and its role in cell migration. *Genomics*, *105*(2), 108–115. https://doi.org/10.1016/j.ygeno.2014.11.007
Kim, H., Kang, K., Ekram, M. B., Roh, T. Y., & Kim, J. (2011). Aebp2 as an epigenetic regulator for neural crest cells. *PLoS ONE*, *6*(9). https://doi.org/10.1371/journal.pone.0025174
Kim, H., Kang, K., & Kim, J. (2009). AEBP2 as a potential targeting protein for Polycomb Repression Complex PRC2. *Nucleic Acids Research*, *37*(9), 2940–2950. https://doi.org/10.1093/nar/gkp149
Kimmins, S., & Sassone-Corsi, P. (2005). Chromatin remodelling and epigenetic features of germ cells. *Nature*, *434*(7033), 583–589. https://doi.org/10.1038/nature03368
Kingston, R. E., & Tamkun, J. W. (2014). Transcriptional regulation by trithorax-group proteins. *Cold Spring Harbor Perspectives in Biology*, *6*(10). https://doi.org/10.1101/cshperspect.a019349
Kino, T., Hurt, D. E., Ichijo, T., Nader, N., & Chrousos, G. P. (2010). Noncoding RNA Gas5 Is a

Growth Arrest – and Starvation-Associated Repressor of the Glucocorticoid Receptor, *3*(107), 1–16.
Kloet, S. L., Makowski, M. M., Baymaz, H. I., van Voorthuijsen, L., Karemaker, I. D., Santanach, A., … Vermeulen, M. (2016a). The dynamic interactome and genomic targets of Polycomb complexes during stem-cell differentiation. *Nature Structural & Molecular Biology*, *23*(7), 682–690. https://doi.org/10.1038/nsmb.3248
Kloet, S. L., Makowski, M. M., Baymaz, H. I., van Voorthuijsen, L., Karemaker, I. D., Santanach, A., … Vermeulen, M. (2016b). The dynamic interactome and genomic targets of Polycomb complexes during stem-cell differentiation. *Nature Structural & Molecular Biology*, *23*(7), 682–690. https://doi.org/10.1038/nsmb.3248
Klose, R. J., Cooper, S., Farcas, A. M., Blackledge, N. P., & Brockdorff, N. (2013). Chromatin Sampling-An Emerging Perspective on Targeting Polycomb Repressor Proteins. *PLoS Genetics*, *9*(8). https://doi.org/10.1371/journal.pgen.1003717
Klose, R. J., Kallin, E. M., & Zhang, Y. (2006). JmjC-domain-containing proteins and histone demethylation. *Nature Reviews. Genetics*, *7*(9), 715–27. https://doi.org/10.1038/nrg1945
Klymenko, T., Papp, B., Fischle, W., Kocher, T., Schelder, M., Fritsch, C., … Müller, J. (2006). A Polycomb group protein complex with sequence-specificDNA-binding and selective methyl-lysine-binding activities. *Genes & Development*, *20*(9), 1110–1122. https://doi.org/10.1101/gad.377406
Kobayashi, H., Sakurai, T., Miura, F., Imai, M., Mochiduki, K., Yanagisawa, E., … Kono, T. (2013). High-resolution DNA methylome analysis of primordial germ cells identifies gender-specific reprogramming in mice. *Genome Research*, *23*(4), 616–627. https://doi.org/10.1101/gr.148023.112
Kohlmaier, A., Savarese, F., Lachner, M., Martens, J., Jenuwein, T., & Wutz, A. (2004). PLoS BIOLOGY A Chromosomal Memory Triggered by Xist Regulates Histone Methylation in X Inactivation, *2*(7). https://doi.org/10.1371/journal.pbio.0020171
Kolthur-Seetharam, U., Pradeepa, M. M., Gupta, N., Narayanaswamy, R., & Rao, M. R. S. (2009). Spatiotemporal organization of AT- and GC-rich DNA and their association with transition proteins TP1 and TP2 in rat condensing spermatids. *The Journal of Histochemistry and Cytochemistry : Official Journal of the Histochemistry Society*, *57*(10), 951–962. https://doi.org/10.1369/jhc.2009.953414
Koontz, J. I., Soreng, a L., Nucci, M., Kuo, F. C., Pauwels, P., van Den Berghe, H., … Sklar, J. (2001). Frequent fusion of the JAZF1 and JJAZ1 genes in endometrial stromal tumors. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(11), 6348–53. https://doi.org/10.1073/pnas.101132598
Koopman Peter, Gubbay, J., Vivian, N., Goodfellow, P., & Lovell-Badge, R. (1991). Male development of chromosomally female mice transgenic for Sry. *Letters To Nature*, *351*(6322), 117–121.
Korzus, E., Rosenfeld, M. G., & Mayford, M. (2004). CBP histone acetyltransferase activity is a critical component of memory consolidation. *Neuron*, *42*(6), 961–972. https://doi.org/10.1016/j.neuron.2004.06.002
Kota, S. K., & Feil, R. (2010). Epigenetic Transitions in Germ Cell Development and Meiosis. *Developmental Cell*, *19*(5), 675–686. https://doi.org/10.1016/j.devcel.2010.10.009
Kurimoto, K., Yamaji, M., Seki, Y., & Saitou, M. (2008). Specification of the germ cell lineage in mice: A process orchestrated by the PR-domain proteins, Blimp1 and Prdm14. *Cell Cycle*, *7*(22), 3514–3518. https://doi.org/10.4161/cc.7.22.6979
Kuzmichev, A., Nishioka, K., Edrjument-Bromage, H., Tempst, P., & Reinberg, D. (2002). Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. *Genes & …*, 2893–2905. https://doi.org/10.1101/gad.1035902.repression
Kycia, I., Kudithipudi, S., Tamas, R., Kungulovski, G., Dhayalan, A., & Jeltsch, A. (2014). The tudor domain of the PHD finger protein 1 is a dual reader of lysine trimethylation at lysine 36 of histone H3 and lysine 27 of histone variant H3t. *Journal of Molecular Biology*, *426*(8), 1651–1660. https://doi.org/10.1016/j.jmb.2013.08.009
Lachner, M., O'Carroll, D., Rea, S., Mechtler, K., & Jenuwein, T. (2001). Methylation of histone H3

lysine 9 creates a binding site for HP1 proteins. *Nature*, *410*(6824), 116–20. https://doi.org/10.1038/35065132
Lagarou, A., Mohd-Sarip, A., Moshkin, Y. M., Chalkley, G. E., Bezstarosti, K., Demmers, J. A. A., & Verrijzer, C. P. (2008). dKDM2 couples histone H2A ubiquitylation to histone H3 demethylation during Polycomb group silencing. *Genes and Development*, *22*(20), 2799–2810. https://doi.org/10.1101/gad.484208
Lai, K. V., Gong, G., Atanasio, A., Rojas, J., Quispe, J., Miloscio, L., … Auerbach, W. (2015). Diverse Phenotypes and Specific Transcription Patterns in Twenty Mouse Lines with Ablated LincRNAs, 1–21. https://doi.org/10.1371/journal.pone.0125522
Landeira, D., & Fisher, A. G. (2011). Inactive yet indispensable: The tale of Jarid2. *Trends in Cell Biology*, *21*(2), 74–80. https://doi.org/10.1016/j.tcb.2010.10.004
Landeira, D., Sauer, S., Poot, R., Dvorkina, M., Mazzarella, L., Jørgensen, H. F., … Fisher, A. G. (2010). Jarid2 is a PRC2 component in embryonic stem cells required for multi-lineage differentiation and recruitment of PRC1 and RNA Polymerase II to developmental regulators. *Nature Cell Biology*, *12*(6), 618–24. https://doi.org/10.1038/ncb2065
Lassalle, B., Bastos, H., Louis, J. P., Riou, L., Testart, J., Dutrillaux, B., … Allemand, I. (2004). " Side Population " cells in adult mouse testis express Bcrp1 gene and are enriched in spermatogonia and germinal stem cells. https://doi.org/10.1242/dev.00918
Lee, C.-H., & Nucci, M. R. (2014). Endometrial stromal sarcoma - the new genetic paradigm. *Histopathology*, 1–19. https://doi.org/10.1111/his.12594
Lesch, B. J., & Page, D. C. (2012). Genetics of germ cell development. *Nat Rev Genet*, *13*(11), 781–794. https://doi.org/10.1038/nrg3294
Lesch, B. J., & Page, D. C. (2014). Poised chromatin in the mammalian germ line. *Development (Cambridge, England)*, *141*(19), 3619–26. https://doi.org/10.1242/dev.113027
Lhuillier-Akakpo, M., Frapporti, A., Denby Wilkes, C., Matelot, M., Vervoort, M., Sperling, L., & Duharcourt, S. (2014). Local Effect of Enhancer of Zeste-Like Reveals Cooperation of Epigenetic and cis-Acting Determinants for Zygotic Genome Rearrangements. *PLoS Genetics*, *10*(9). https://doi.org/10.1371/journal.pgen.1004665
Li, G., Margueron, R., Ku, M., Chambon, P., Bernstein, B. E., & Reinberg, D. (2010). Jarid2 and PRC2, partners in regulating gene expression. *Genes and Development*, *24*(4), 368–380. https://doi.org/10.1101/gad.1886410
Li, L., Helms, J. A., & Chang, H. Y. (2016). Comment on " Hotair Is Dispensable for Mouse Development ," 10–11. https://doi.org/10.1101/gad.205701
Li, L., Liu, B., Wapinski, O. L., Tsai, M., Qu, K., Zhang, J., … Chang, H. Y. (2013). Report Targeted Disruption of Hotair Leads to Homeotic Transformation and Gene Derepression. *CellReports*, *5*(1), 3–12. https://doi.org/10.1016/j.celrep.2013.09.003
Li, X., Isono, K.-I., Yamada, D., Endo, T. A., Endoh, M., Shinga, J., … Koseki, H. (2011). Mammalian polycomb-like Pcl2/Mtf2 is a novel regulatory component of PRC2 that can differentially modulate polycomb activity both at the Hox gene cluster and at Cdkn2a genes. *Molecular and Cellular Biology*, *31*(2), 351–64. https://doi.org/10.1128/MCB.00259-10
Liefke, R., Karwacki-Neisius, V., & Shi, Y. (2017). EPOP Interacts with Elongin BC and USP7 to Modulate the Chromatin Landscape. *Molecular Cell*, *64*, 659–672. https://doi.org/10.1016/j.molcel.2016.12.006
Liefke, R., & Shi, Y. (2015). The PRC2-associated factor C17orf96 is a novel CpG island regulator in mouse ES cells. *Cell Discovery*, *1*(MAY), 15008. https://doi.org/10.1038/celldisc.2015.8
Lifschytz, E., & Lindsley, D. L. (1972). The role of X-chromosome inactivation during spermatogenesis (Drosophila-allocycly-chromosome evolution-male sterility-dosage compensation). *Proceedings of the National Academy of Sciences of the United States of America*, *69*(1), 182–186. https://doi.org/10.1073/pnas.69.1.182
Liu, Y., & Montell, D. J. (2001). Jing: a downstream target of slbo required for developmental control of border cell migration. *Development (Cambridge, England)*, *128*(3), 321–30. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11152631
Ma, X., Wang, J., Wang, J., Ma, C. X., Gao, X., Patriub, V., … Sklar, J. L. (2017). The JAZF1-SUZ12 Fusion Protein Disrupts PRC2 Complexes and Impairs Chromatin Repression During Human Endometrial Stromal Tumorogenesis. *Oncotarget*, *8*(3), 4062–4078.

https://doi.org/10.18632/oncotarget.13270
Machida, Y. J., Machida, Y., Vashisht, A. A., Wohlschlegel, J. A., & Dutta, A. (2009). The deubiquitinating enzyme BAP1 regulates cell growth via interaction with HCF-1. *Journal of Biological Chemistry*, *284*(49), 34179–34188. https://doi.org/10.1074/jbc.M109.046755
Maenner, S., Blaud, M., Fouillen, L., Savoye, A., Marchand, V., Dubois, A., … Branlant, C. (2010). 2-D Structure of the A Region of Xist RNA and Its Implication for PRC2 Association, *8*(1), 1–16. https://doi.org/10.1371/journal.pbio.1000276
Magnúsdóttir, E., Dietmann, S., Murakami, K., Günesdogan, U., Tang, F., Bao, S., … Azim Surani, M. (2013). A tripartite transcription factor network regulates primordial germ cell specification in mice. *Nature Cell Biology*, *15*(8), 905–915. https://doi.org/10.1038/ncb2798
Magnusdottir, E., & Surani, M. A. (2014). How to make a primordial germ cell. *Development*, *141*(2), 245–252. https://doi.org/Doi 10.1242/Dev.098269
Maier, V. K., Feeney, C. M., Taylor, J. E., Creech, A. L., Qiao, J. W., Szanto, A., … Lee, J. T. (2015). Functional Proteomic Analysis of Repressive Histone Methyltransferase Complexes Reveals ZNF518B as a G9A Regulator. *Mol Cell Proteomics*, *14*(6), 1435–1446. https://doi.org/10.1074/mcp.M114.044586
Margueron, R., Justin, N., Ohno, K., Sharpe, M. L., Son, J., Drury III, W. J., … Gamblin, S. J. (2009). Role of the polycomb protein EED in the propagation of repressive histone marks. *Nature*, *461*(7265), 762–767. https://doi.org/10.1038/nature08398
Margueron, R., Li, G., Sarma, K., Blais, A., Zavadil, J., Woodcock, C. L., … Reinberg, D. (2008a). Ezh1 and Ezh2 Maintain Repressive Chromatin through Different Mechanisms. *Molecular Cell*, *32*(4), 503–518. https://doi.org/10.1016/j.molcel.2008.11.004
Margueron, R., Li, G., Sarma, K., Blais, A., Zavadil, J., Woodcock, C. L., … Reinberg, D. (2008b). Ezh1 and Ezh2 Maintain Repressive Chromatin through Different Mechanisms. *Molecular Cell*, *32*(4), 503–518. https://doi.org/10.1016/j.molcel.2008.11.004
Margueron, R., & Reinberg, D. (2011). The Polycomb complex PRC2 and its mark in life. *Nature*, *469*(7330), 343–349. https://doi.org/10.1038/nature09784
Mariner, P. D., Walters, R. D., Espinoza, C. A., Drullinger, L. F., Wagner, S. D., Kugel, J. F., & Goodrich, J. A. (2008). Article Human Alu RNA Is a Modular Transacting Repressor of mRNA Transcription during Heat Shock, 499–509. https://doi.org/10.1016/j.molcel.2007.12.013
Martianov, I., Ramadass, A., Barros, A. S., Chow, N., & Akoulitchev, A. (2007). Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript, *445*(February). https://doi.org/10.1038/nature05519
Mashtalir, N., Daou, S., Barbour, H., Sen, N. N., Gagnon, J., Hammond-Martel, I., … Affar, E. B. (2014). Autodeubiquitination protects the tumor suppressor BAP1 from cytoplasmic sequestration mediated by the atypical ubiquitin ligase UBE2O. *Molecular Cell*, *54*(3), 392–406. https://doi.org/10.1016/j.molcel.2014.03.002
McCall, K., & Bender, W. (1996). Probes of chromatin accessibility in the Drosophila bithorax complex respond differently to Polycomb-mediated repression. *The EMBO Journal*, *15*(3), 569–580. https://doi.org/10.1016/0168-9525(96)81432-6
McHugh, C. A., Chen, C.-K., Chow, A., Surka, C. F., Tran, C., McDonel, P., … Guttman, M. (2015). The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature*, *521*(7551), 232–6. https://doi.org/10.1038/nature14443
Mchugh, C. A., Chen, C., Chow, A., Surka, C. F., Tran, C., Mcdonel, P., … Guttman, M. (2015). silence transcription through HDAC3. https://doi.org/10.1038/nature14443
McKenzie Duncan, I. (1982). Polycomblike: A gene that appears to be required for the normal expression of the bithorax and antennapedia gene complexes of Drosophila melanogaster. *Genetics*, *102*(1), 49–70.
Meistrich, M., & RA Hess. (2013). Assessment of spermatogenesis through staging of seminiferous tubules. *Methods Mol Biol*, *927*, 299–307.
Mejetta, S., Morey, L., Pascual, G., Kuebler, B., Mysliwiec, M. R., Lee, Y., … Benitah, S. A. (2011). Jarid2 regulates mouse epidermal stem cell activation and differentiation. *The EMBO Journal*, *30*(17), 3635–46. https://doi.org/10.1038/emboj.2011.265
Mendenhall, E. M., Koche, R. P., Truong, T., Zhou, V. W., Issac, B., Chi, A. S., … Bernstein, B. E. (2010). GC-rich sequence elements recruit PRC2 in mammalian ES cells. *PLoS Genetics*, *6*(12),

1–10. https://doi.org/10.1371/journal.pgen.1001244
Mercer, T. R., Dinger, M. E., & Mattick, J. S. (2009). Long non-coding RNAs:insights into functions. *Nature Reviews Genetics*, *10*, 155–159.
Messerschmidt, D. M., Knowles, B. B., & Solter, D. (2014). DNA methylation dynamics during epigenetic reprogramming in the germline and preimplantation embryos. *Genes and Development*, *28*(8), 812–828. https://doi.org/10.1101/gad.234294.113
Min, J., Zhang, Y., & Xu, R. (2003). Structural basis for specific binding of Polycomb chromodomain to histone H3 methylated at Lys 27 Structural basis for specific binding of Polycomb chromodomain to histone H3 methylated at Lys 27, 1823–1828. https://doi.org/10.1101/gad.269603
Minajigi, A., Froberg, J. E., Wei, C., Sunwoo, H., Kesner, B., Colognori, D., … Lee, J. T. (2015). directed chromosome conformation, *2276*. https://doi.org/10.1126/science.
Morey, L., Pascual, G., Cozzuto, L., Roma, G., Wutz, A., Benitah, S. A., & Di Croce, L. (2012). Nonoverlapping functions of the polycomb group Cbx family of proteins in embryonic stem cells. *Cell Stem Cell*, *10*(1), 47–62. https://doi.org/10.1016/j.stem.2011.12.006
Morey, Lluis, Aloia, L., Cozzuto, L., Benitah, S. A., & Di Croce, L. (2013). RYBP and Cbx7 Define Specific Biological Functions of Polycomb Complexes in Mouse Embryonic Stem Cells. *Cell Reports*, *3*(1), 60–69. https://doi.org/10.1016/j.celrep.2012.11.026
Morris, K. V, & Mattick, J. S. (2014). The rise of regulatory RNA. *Nature Publishing Group*, *15*(6), 423–437. https://doi.org/10.1038/nrg3722
Mousavi, K., Zare, H., Wang, A. H., & Sartorelli, V. (2012). Polycomb Protein Ezh1 Promotes RNA Polymerase II Elongation. *Molecular Cell*, *45*(2), 255–262. https://doi.org/10.1016/j.molcel.2011.11.019
Mu, W., Starmer, J., Fedoriw, A. M., Yee, D., & Magnuson, T. (2014). Repression of the soma-specific transcriptome by Polycomb-repressive complex 2 promotes male germ cell development, 2056–2069. https://doi.org/10.1101/gad.246124.114.gonial
Mu, W., Starmer, J., Shibata, Y., Yee, D., & Magnuson, T. (2017). EZH1 in germ cells safeguards the function of PRC2 during spermatogenesis. *Developmental Biology*. https://doi.org/10.1016/j.ydbio.2017.02.017
Müller, J., Hart, C. M., Francis, N. J., Vargas, M. L., Sengupta, A., Wild, B., … Simon, J. A. (2002). Histone methyltransferase activity of a Drosophila Polycomb group repressor complex. *Cell*, *111*(2), 197–208. https://doi.org/10.1016/S0092-8674(02)00976-5
Musselman, C. A., Avvakumov, N., Watanabe, R., Abraham, C. G., Lalonde, M.-E., Hong, Z., … Kutateladze, T. G. (2012). Molecular basis for H3K36me3 recognition by the Tudor domain of PHF1. *Nature Structural & Molecular Biology*, *19*(12), 1266–1272. https://doi.org/10.1038/nsmb.2435
Nady, N., Gupta, A., Ma, Z., Swigut, T., Koide, A., Koide, S., & Wysocka, J. (2015). ETO family protein Mtgr1 mediates Prdm14 functions in stem cell maintenance and primordial germ cell formation. *eLife*, *4*(NOVEMBER2015), 1–28. https://doi.org/10.7554/eLife.10150
Nakayama, J., Rice, J. C., Strahl, B. D., Allis, C. D., & Grewal, S. I. (2001). Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. *Science (New York, N.Y.)*, *292*(5514), 110–3. https://doi.org/10.1126/science.1060118
Napoles, M. De, Mermoud, J. E., Wakao, R., Tang, Y. A., Endoh, M., Appanah, R., … Brockdorff, N. (2004). Polycomb Group Proteins Ring1A / B Link Ubiquitylation of Histone H2A to Heritable Gene Silencing and X Inactivation. *Proteins*, *7*, 663–676.
Nekrasov, M., Klymenko, T., Fraterman, S., Papp, B., Oktaba, K., Köcher, T., … Müller, J. (2007). Pcl-PRC2 is needed to generate high levels of H3-K27 trimethylation at Polycomb target genes. *The EMBO Journal*, *26*(18), 4078–88. https://doi.org/10.1038/sj.emboj.7601837
Ng, J.-H., Kumar, V., Muratani, M., Kraus, P., Yeo, J.-C., Yaw, L.-P., … Ng, H.-H. (2013). In vivo epigenomic profiling of germ cells reveals germ cell molecular signatures. *Developmental Cell*, *24*(3), 324–33. https://doi.org/10.1016/j.devcel.2012.12.011
Ng, J., Hart, C. M., Morgan, K., & Simon, J. A. (2000). A Drosophila ESC-E(Z) protein complex is distinct from other polycomb group complexes and contains covalently modified ESC. *Molecular and Cellular Biology*, *20*(9), 3069–3078. https://doi.org/10.1128/MCB.20.9.3069-3078.2000

O'Carroll, D., Erhardt, S., Pagani, M., Barton, S. C., Surani, M. A., & Jenuwein, T. (2001). The polycomb-group gene Ezh2 is required for early mouse development. *Mol Cell Biol*, *21*(13), 4330–4336. https://doi.org/10.1128/MCB.21.13.4330

Ohinata, Y., Ohta, H., Shigeta, M., Yamanaka, K., Wakayama, T., & Saitou, M. (2009). A Signaling Principle for the Specification of the Germ Cell Lineage in Mice. *Cell*, *137*(3), 571–584. https://doi.org/10.1016/j.cell.2009.03.014

Ohinata, Y., Payer, B., O'Carroll, D., Ancelin, K., Ono, Y., Sano, M., … Surani, M. A. (2005). Blimp1 is a critical determinant of the germ cell lineage in mice. *Nature*, *436*(July), 207–213. https://doi.org/10.1038/nature03813

Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., … Zimmer, & A. (2002). Analysis of the mouse transcriptome based on functional annotation of. *Nature*, *420*, 563–573.

Oktaba, K., Gutiérrez, L., Gagneur, J., Girardot, C., Sengupta, A. K., Furlong, E. E. M., & Müller, J. (2008). Dynamic Regulation by Polycomb Group Protein Complexes Controls Pattern Formation and the Cell Cycle in Drosophila. *Developmental Cell*, *15*(6), 877–889. https://doi.org/10.1016/j.devcel.2008.10.005

Orlando, V., Jane, E. P., Chinwalla, V., Harte, P. J., & Paro, R. (1998). Binding of Trithorax and Polycomb proteins to the bithorax complex: Dynamic changes during early Drosophila embryogenesis. *EMBO Journal*, *17*(17), 5141–5150. https://doi.org/10.1093/emboj/17.17.5141

Orsi, G. A., Couble, P., & Loppin, B. (2009). Epigenetic and replacement roles of histone variant H3.3 in reproduction and development. *International Journal of Developmental Biology*, *53*(2–3), 231–243. https://doi.org/10.1387/ijdb.082653go

Pan, J., & Keeney, S. (2007). Molecular cartography: Mapping the landscape of meiotic recombination. *PLoS Biology*, *5*(12), 2774–2777. https://doi.org/10.1371/journal.pbio.0050333

Pandey, R. R., Mondal, T., Mohammad, F., Enroth, S., Redrup, L., Komorowski, J., … Kanduri, C. (2008). Article Kcnq1ot1 Antisense Noncoding RNA Mediates Lineage-Specific Transcriptional Silencing through Chromatin-Level Regulation. *Molecular Cell*, *32*(2), 232–246. https://doi.org/10.1016/j.molcel.2008.08.022

Papoulas, O., Beek, S. J., Moseley, S. L., McCallum, C. M., Sarte, M., Shearn, A., & Tamkun, J. W. (1998). The Drosophila trithorax group proteins BRM, ASH1 and ASH2 are subunits of distinct protein complexes. *Development (Cambridge, England)*, *125*(20), 3955–3966. Retrieved from http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=9735357&retmode=ref&cmd=prlinks

Paro, R., & Hogness, D. S. (1991). The Polycomb protein shares a homologous domain with a heterochromatin-associated protein of Drosophila. *Proceedings of the National Academy of Sciences*, *88*(1), 263–267. https://doi.org/10.1073/pnas.88.1.263

Pasini, D., Bracken, A. P., Jensen, M. R., Denchi, E. L., Helin, K., Birve, A., … Sasazuki, T. (2004). Suz12 is essential for mouse development and for EZH2 histone methyltransferase activity. *The EMBO Journal*, *23*(20), 4061–4071. https://doi.org/10.1038/sj.emboj.7600402

Pasini, D., Cloos, P. A. C., Walfridsson, J., Olsson, L., Bukowski, J.-P., Johansen, J. V, … Helin, K. (2010). JARID2 regulates binding of the Polycomb repressive complex 2 to target genes in ES cells. *Nature*, *464*(7286), 306–310. https://doi.org/10.1038/nature08788

Pasmant, E., Laurendeau, I., He, D., Vidaud, M., Vidaud, D., & Bie, I. (2007). Characterization of a Germ-Line Deletion , Including the Entire INK4 / ARF Locus , in a Melanoma-Neural System Tumor Family : Identification of ANRIL , an Antisense Noncoding RNA Whose Expression Coclusters with ARF, (8), 3963–3970. https://doi.org/10.1158/0008-5472.CAN-06-2004

Pasmant, E., Laurendeau, I., Héron, D., Vidaud, M., Vidaud, D., & Bièche, I. (2007). Characterization of a germ-line deletion, including the entire INK4/ARF locus, in a melanoma-neural system tumor family: Identification of ANRIL, an antisense noncoding RNA whose expression coclusters with ARF. *Cancer Research*, *67*(8), 3963–3969. https://doi.org/10.1158/0008-5472.CAN-06-2004

Peng, J. C., Valouev, A., Swigut, T., Zhang, J., Zhao, Y., Sidow, A., & Wysocka, J. (2009a). Jarid2/Jumonji Coordinates Control of PRC2 Enzymatic Activity and Target Gene Occupancy in Pluripotent Cells. *Cell*, *139*(7), 1290–1302. https://doi.org/10.1016/j.cell.2009.12.002

Peng, J. C., Valouev, A., Swigut, T., Zhang, J., Zhao, Y., Sidow, A., & Wysocka, J. (2009b). Jarid2/Jumonji Coordinates Control of PRC2 Enzymatic Activity and Target Gene Occupancy in

Pluripotent Cells. *Cell*, *139*(7), 1290–1302. https://doi.org/10.1016/j.cell.2009.12.002
Pengelly, A. R., Copur, O., Jackle, H., Herzi, A., & Muller, J. (2013). A Histone Mutant Reproduces the Phenotype Caused by Loss of Histone-Modifying Factor Polycomb, *339*(February), 698–700.
Pengelly, A. R., Kalb, R., Finkl, K., And, & Muller, J. (2015). Transcriptional repression by PRC1 in the absence of H2A monoubiquitylation. *Genes and Development*, *29*(14), 1487–1492. https://doi.org/10.1101/gad.265439.115
Pengelly, A. R., Kalb, R., Finkl, K., & Muller, J. (2015). Transcriptional repression by PRC1 in the absence of H2A monoubiquitylation. *Genes and Development*, *29*(14), 1487–1492. https://doi.org/10.1101/gad.265439.115
Percharde, M., Wong, P., & Ramalho-Santos, M. (2017). Global Hypertranscription in the Mouse Embryonic Germline. *Cell Reports*, *19*(10), 1987–1996. https://doi.org/10.1016/j.celrep.2017.05.036
Pirrotta, V. (1997). Chromatin-silencing mechanisms in Drosophila maintain patterns of gene expression. *Trends in Genetics*, *13*(8), 314–318. https://doi.org/10.1016/S0168-9525(97)01178-5
Pirrotta Vincenzo, & Rastelli L. (1994). White gene expression, repressive chromatin domains and homeotic gene regulation in Drosophila. *BioEssays : News and Reviews in Molecular, Cellular and Developmental Biology*, *16*(8), 549–556. https://doi.org/10.1002/bies.950160808
Plath, K., Fang, J., Mlynarczyk-evans, S. K., Cao, R., Worringer, K. A., Wang, H., … Panning, B. (2003). Role of Histone H3 Lysine 27 Methylation in X Inactivation, *300*(April), 131–136.
Portoso, M., Ragazzini, R., Brenčič, Ž., Moiani, A., Michaud, A., Vassilev, I., … Margueron, R. (2017). PRC2 is dispensable for HOTAIR-mediated transcriptional repression. *EMBO Journal*, *36*(8). https://doi.org/10.15252/embj.201695335
Prokopuk, L., Stringer, J. M., Hogg, K., Elgass, K. D., & Western, P. S. (2017). PRC2 is required for extensive reorganization of H3K27me3 during epigenetic reprogramming in mouse fetal germ cells. *Epigenetics & Chromatin*, *10*(1), 7. https://doi.org/10.1186/s13072-017-0113-9
Puliyath, G., & Nair, M. K. (2012). Endometrial stromal sarcoma: A review of the literature. *Indian Journal of Medical and Paediatric Oncology : Official Journal of Indian Society of Medical & Paediatric Oncology*, *33*(1), 1–6. https://doi.org/10.4103/0971-5851.96960
Qin, J., Whyte, W. A., Anderssen, E., Apostolou, E., Chen, H. H., Akbarian, S., … Hock, H. (2012). The polycomb group protein L3mbtl2 assembles an atypical PRC1-family complex that is essential in pluripotent stem cells and early development. *Cell Stem Cell*, *11*(3), 319–332. https://doi.org/10.1016/j.stem.2012.06.002
Rathke, C., Baarends, W. M., Awe, S., & Renkawitz-Pohl, R. (2014). Chromatin dynamics during spermiogenesis. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, *1839*(3), 155–168. https://doi.org/10.1016/j.bbagrm.2013.08.004
Redrup, L., Branco, M. R., Perdeaux, E. R., Krueger, C., Lewis, A., Santos, F., … Reik, W. (2009). The long noncoding RNA Kcnq1ot1 organises a lineage- specific nuclear domain for epigenetic gene silencing, *530*, 525–530. https://doi.org/10.1242/dev.031328
Ren, X., & Kerppola, T. K. (2011). REST interacts with Cbx proteins and regulates polycomb repressive complex 1 occupancy at RE1 elements. *Molecular and Cellular Biology*, *31*(10), 2100–10. https://doi.org/10.1128/MCB.05088-11
Richardson, B. E., & Lehmann, R. (2010). Mechanisms guiding primordial germ cell migration: strategies from different organisms. *Nature Reviews. Molecular Cell Biology*, *11*(1), 37–49. https://doi.org/10.1038/nrm2815
Ringrose, L., & Paro, R. (2007). Polycomb/Trithorax response elements and epigenetic memory of cell identity. *Development (Cambridge, England)*, *134*(2), 223–32. https://doi.org/10.1242/dev.02723
Rinn, J. L., & Chang, H. Y. (2012). Genome Regulation by Long Noncoding RNAs. *Annual Reviews Biochemistry*, *81*, 145–66. https://doi.org/10.1146/annurev-biochem-051410-092902
Rinn, J. L., Kertesz, M., Wang, J. K., Squazzo, S. L., Xu, X., Brugmann, S. A., … Chang, H. Y. (2007). Functional Demarcation of Active and Silent Chromatin Domains in Human HOX Loci by Noncoding RNAs, 1311–1323. https://doi.org/10.1016/j.cell.2007.05.022
Sachs, M., Onodera, C., Blaschke, K., Ebata, K., Song, J., & Ramalho-Santos, M. (2013). Bivalent Chromatin Marks Developmental Regulatory Genes in the Mouse Embryonic Germline InVivo. *Cell Reports*, *3*(6), 1777–1784. https://doi.org/10.1016/j.celrep.2013.04.032

Samans, B., Yang, Y., Krebs, S., Sarode, G. V., Blum, H., Reichenbach, M., … Schagdarsurengin, U. (2014). Uniformity of nucleosome preservation pattern in mammalian sperm and Its connection to repetitive DNA elements. *Developmental Cell*, *30*(1), 23–35. https://doi.org/10.1016/j.devcel.2014.05.023

Sanchez-Pulido, L., Devos, D., Sung, Z., & Calonje, M. (2008). RAWUL: A new ubiquitin-like domain in PRC1 Ring finger proteins that unveils putative plant and worm PRC1 orthologs. *BMC Genomics*, *9*(1), 308. https://doi.org/10.1186/1471-2164-9-308

Sanulli, S., Justin, N., Teissandier, A., Ancelin, K., Portoso, M., Caron, M., … Margueron, R. (2015). Jarid2 Methylation via the PRC2 Complex Regulates H3K27me3 Deposition during Cell Differentiation. *Molecular Cell*, *57*(5), 769–783. https://doi.org/10.1016/j.molcel.2014.12.020

Sarma, K., Margueron, R., Ivanov, A., Pirrotta, V., & Reinberg, D. (2008). Ezh2 requires PHF1 to efficiently catalyze H3 lysine 27 trimethylation in vivo. *Mol Cell Biol*, *28*(8), 2718–31. https://doi.org/10.1128/MCB.02017-07

Sasaki, H., & Matsui, Y. (2008). Epigenetic events in mammalian germ-cell development: reprogramming and beyond. *Nat Rev Genet*, *9*(2), 129–140. https://doi.org/10.1038/nrg2295

Savla, U., Benes, J., Zhang, J., & Jones, R. S. (2008). Recruitment of Drosophila Polycomb-group proteins by Polycomblike, a component of a novel protein complex in larvae. *Development*, *135*(5), 813–817. https://doi.org/dev.016006 [pii]\r10.1242/dev.016006

Scheuermann, J. C., de Ayala Alonso, A. G., Oktaba, K., Ly-Hartig, N., McGinty, R. K., Fraterman, S., … Müller, J. (2010). Histone H2A deubiquitinase activity of the Polycomb repressive complex PR-DUB. *Nature*, *465*(7295), 243–7. https://doi.org/10.1038/nature08966

Schmitges, F. W., Prusty, A. B., Faty, M., Stützer, A., Lingaraju, G. M., Aiwazian, J., … Thomä, N. H. (2011). Histone Methylation by PRC2 Is Inhibited by Active Chromatin Marks. *Molecular Cell*, *42*(3), 330–341. https://doi.org/10.1016/j.molcel.2011.03.025

Schoorlemmer, J., Marcos-Gutierrez, C., Were, F., Martinez, R., Garcia, E., Satijn, D. P. E., … Vidal, M. (1997). Ring1A is a transcriptional repressor that interacts with the Polycomb-M33 protein and is expressed at rhombomere boundaries in the mouse hindbrain. *EMBO Journal*, *16*(19), 5930–5942. https://doi.org/10.1093/emboj/16.19.5930

Schorderet, P., & Duboule, D. (2011). Structural and Functional Differences in the Long Non- Coding RNA Hotair in Mouse and Human, *7*(5), 1–10. https://doi.org/10.1371/journal.pgen.1002071

Schuettengruber, B., Chourrout, D., Vervoort, M., Leblanc, B., & Cavalli, G. (2007). Genome Regulation by Polycomb and Trithorax Proteins. *Cell*, *128*(4), 735–745. https://doi.org/10.1016/j.cell.2007.02.009

Schwartz, Y. B., & Pirrotta, V. (2013). A new world of Polycombs: unexpected partnerships and emerging functions. *Nature Reviews. Genetics*, *14*(12), 853–64. https://doi.org/10.1038/nrg3603

Seki, Y., Hayashi, K., Itoh, K., Mizugaki, M., Saitou, M., & Matsui, Y. (2005). Extensive and orderly reprogramming of genome-wide chromatin modifications associated with specification and early development of germ cells in mice. *Developmental Biology*, *278*(2), 440–458. https://doi.org/10.1016/j.ydbio.2004.11.025

Senthilkumar, R., & Mishra, R. K. (2009). Novel motifs distinguish multiple homologues of Polycomb in vertebrates: expansion and diversification of the epigenetic toolkit. *BMC Genomics*, *10*(1), 549. https://doi.org/10.1186/1471-2164-10-549

Shao, Z., Raible, F., Mollaaghababa, R., Guyon, J. R., Wu, C. T., Bender, W., & Kingston, R. E. (1999). Stabilization of chromatin structure by PRC1, a polycomb complex. *Cell*, *98*(1), 37–46. https://doi.org/10.1016/S0092-8674(00)80604-2

Shaver, S., Casas-Mollano, J. A., Cerny, R. L., & Cerutti, H. (2010). Origin of the polycomb repressive complex 2 and gene silencing by an E(z) homolog in the unicellular alga Chlamydomonas. *Epigenetics*, *5*(4), 301–312. https://doi.org/10.4161/epi.5.4.11608

Shen, X., Kim, W., Fujiwara, Y., Simon, M. D., Liu, Y., Mysliwiec, M. R., … Orkin, S. H. (2009). Jumonji Modulates Polycomb Activity and Self-Renewal versus Differentiation of Stem Cells. *Cell*, *139*(7), 1303–1314. https://doi.org/10.1016/j.cell.2009.12.003

Shen, X., Liu, Y., Hsu, Y. J., Fujiwara, Y., Kim, J., Mao, X., … Orkin, S. H. (2008). EZH1 Mediates Methylation on Histone H3 Lysine 27 and Complements EZH2 in Maintaining Stem Cell Identity and Executing Pluripotency. *Molecular Cell*, *32*(4), 491–502. https://doi.org/10.1016/j.molcel.2008.10.016

Shinagawa, T., Huynh, L. M., Takagi, T., Tsukamoto, D., Tomaru, C., Kwak, H.-G., … Ishii, S. (2015). Disruption of Th2a and Th2b genes causes defects in spermatogenesis. *Development*, *142*(7), 1287–1292. https://doi.org/10.1242/dev.121830

Siklenka, K., Erkek, S., Godmann, M., Lambrot, R., McGraw, S., Lafleur, C., … Kimmins, S. (2015). Disruption of histone methylation in developing sperm impairs offspring health transgenerationally. *Science (New York, N.Y.)*, *350*(6261), aab2006. https://doi.org/10.1126/science.aab2006

Silva, J., Mak, W., Zvetkova, I., Appanah, R., Nesterova, T. B., Webster, Z., … Brockdorff, N. (2003). Establishment of Histone H3 Methylation on the Inactive X Chromosome Requires Transient Recruitment of Eed-Enx1 Polycomb Group Complexes, *4*, 481–495.

Simon, J. a., & Kingston, R. E. (2013). Occupying Chromatin: Polycomb Mechanisms for Getting to Genomic Targets, Stopping Transcriptional Traffic, and Staying Put. *Molecular Cell*, *49*(5), 808–824. https://doi.org/10.1016/j.molcel.2013.02.013

Sing, A., Pannell, D., Karaiskakis, A., Sturgeon, K., Djabali, M., Ellis, J., … Cordes, S. P. (2009). A Vertebrate Polycomb Response Element Governs Segmentation of the Posterior Hindbrain. *Cell*, *138*(5), 885–897. https://doi.org/10.1016/j.cell.2009.08.020

Smits, A. H., Jansen, P. W. T. C., Poser, I., Hyman, A. a., & Vermeulen, M. (2013a). Stoichiometry of chromatin-associated protein complexes revealed by label-free quantitative mass spectrometry-based proteomics. *Nucleic Acids Research*, *41*(1), 1–8. https://doi.org/10.1093/nar/gks941

Smits, A. H., Jansen, P. W. T. C., Poser, I., Hyman, A. A., & Vermeulen, M. (2013b). Stoichiometry of chromatin-associated protein complexes revealed by label-free quantitative mass spectrometry-based proteomics. *Nucleic Acids Research*, *41*(1), 1–8. https://doi.org/10.1093/nar/gks941

Son, J., Shen, S. S., Margueron, R., & Reinberg, D. (2013). Nucleosome-binding activities within JARID2 and EZH1 regulate the function of PRC2 on chromatin. *Genes and Development*, *27*(24), 2663–2677. https://doi.org/10.1101/gad.225888.113

Spiller, C., & Bowles, J. (2015). Germ Cell Biology Sex determination in mammalian germ cells. *Asian Journal of Andrology*, *17*(November 2014), 427–432. https://doi.org/10.4103/1008-682X.150037

Stojic, L., Jasencakova, Z., Prezioso, C., Stützer, A., Bodega, B., Pasini, D., … Orlando, V. (2011). Chromatin regulated interchange between polycomb repressive complex 2 (PRC2)-Ezh2 and PRC2-Ezh1 complexes controls myogenin activation in skeletal muscle cells. *Epigenetics & Chromatin*, *4*(1), 16. https://doi.org/10.1186/1756-8935-4-16

Storre, J., Elsasser, H. P., Fuchs, M., Ullmann, D., Livingston, D. M., & Gaubatz, S. (2002). Homeotic transformations of the axial skeleton that accompany a targeted deletion of E2f6. *EMBO Reports*, *3*(7), 695–700. https://doi.org/10.1093/embo-reports/kvf141

Strahl, B. D., Grant, P. A., Briggs, S. D., Sun, Z.-W., Bone, J. R., Caldwell, J. A., … Allis, C. D. (2002). Set2 Is a Nucleosomal Histone H3-Selective Methyltransferase That Mediates Transcriptional Repression. *Molecular and Cellular Biology*, *22*(5), 1298–1306. https://doi.org/10.1128/MCB.22.5.1298-1306.2002

Swanson, W. J., & Vacquier, V. D. (2002). The rapid evolution of reproductive proteins. *Genetics*, *3*(February), 137–144. https://doi.org/10.1038/nrg/733

Takeuchi, T., Takeuchi, T., Katoh-fukui, Y., Katoh-fukui, Y., Tsuchiya, R., Tsuchiya, R., … Sciences, L. (1995). Gene trap capture of a novel mouse gene,. *Genes & Development*, 1211–1222. https://doi.org/10.1101/gad.9.10.1211

Tang, W. W. C., Dietmann, S., Irie, N., Leitch, H. G., Floros, V. I., Bradshaw, C. R., … Surani, M. A. (2015). A unique gene regulatory network resets the human germline epigenome for development. *Cell*, *161*(6), 1453–1467. https://doi.org/10.1016/j.cell.2015.04.053

Tang, W. W. C., Kobayashi, T., Irie, N., Dietmann, S., & Surani, M. A. (2016). Specification and epigenetic programming of the human germ line. *Nature Reviews Genetics*, *17*(10), 585–600. https://doi.org/10.1038/nrg.2016.88

Taunton, J., Hassig, C. A., & Schreiber, S. L. (1996). A mammalian histone deacetylase related to the yeast transcriptional regulator Rpd3p. *Science*, *272*(April), 408–411. https://doi.org/10.1126/science.272.5260.408

Tavares, L., Dimitrova, E., Oxley, D., Webster, J., Poot, R., Demmers, J., … Brockdorff, N. (2012).

RYBP-PRC1 complexes mediate H2A ubiquitylation at polycomb target sites independently of PRC2 and H3K27me3. *Cell*, *148*(4), 664–678. https://doi.org/10.1016/j.cell.2011.12.029

Tie, F., Furuyama, T., Prasad-Sinha, J., Jane, E., & Harte, P. J. (2001). The Drosophila Polycomb Group proteins ESC and E(Z) are present in a complex containing the histone-binding protein p55 and the histone deacetylase RPD3. *Development (Cambridge, England)*, *128*, 275–286.

Trojer, P., Cao, A. R., Gao, Z., Li, Y., Zhang, J., Xu, X., … Reinberg, D. (2011). L3MBTL2 Protein Acts in Concert with PcG Protein-Mediated Monoubiquitination of H2A to Establish a Repressive Chromatin Structure. *Molecular Cell*, *42*(4), 438–450. https://doi.org/10.1016/j.molcel.2011.04.004

Tsai, M.-C., Manor, O., Wan, Y., Mosammaparast, N., Wang, J. K., Lan, F., … Chang, H. Y. (2010). Long noncoding RNA as modular scaffold of histone modification complexes. *Science (New York, N.Y.)*, *329*(5992), 689–693. https://doi.org/10.1126/science.1192002

Tschiersch, B., Hofmann, A., Krauss, V., Dorn, R., Korge, G., & Reuter, G. (1994). The protein encoded by the Drosophila position-effect variegation suppressor gene Su(var)3-9 combines domains of antagonistic regulators of homeotic gene complexes. *The EMBO Journal*, *13*(16), 3822–3831.

Tu, S., Narendra, V., Yamaji, M., Vidal, S. E., Rojas, L. A., Wang, X., … Reinberg, D. (2016). Co-repressor CBFA2T2 regulates pluripotency and germline development. *Nature*, *534*(7607), 1–11. https://doi.org/10.1038/nature18004

Ueda, J., Harada, A., Urahama, T., Machida, S., Maehara, K., Hada, M., … Yamagata, K. (2017). Testis-Specific Histone Variant H3t Gene Is Essential for Entry into Spermatogenesis. *Cell Reports*, *18*(3), 593–600. https://doi.org/10.1016/j.celrep.2016.12.065

Vandamme, J., Völkel, P., Rosnoblet, C., Le Faou, P., & Angrand, P.-O. (2011). Interaction proteomics analysis of polycomb proteins defines distinct PRC1 complexes in mammalian cells. *Molecular & Cellular Proteomics : MCP*, *10*(4), M110.002642. https://doi.org/10.1074/mcp.M110.002642

Vella, P., Barozzi, I., Cuomo, A., Bonaldi, T., & Pasini, D. (2012). Yin Yang 1 extends the Myc-related transcription factors network in embryonic stem cells. *Nucleic Acids Research*, *40*(8), 3403–3418. https://doi.org/10.1093/nar/gkr1290

Veselovska, L., Smallwood, S. A., Saadeh, H., Stewart, K. R., Krueger, F., Maupetit-Méhouas, S., … Kelsey, G. (2015). Deep sequencing and de novo assembly of the mouse oocyte transcriptome define the contribution of transcription to the DNA methylation landscape. *Genome Biology*, *16*, 209. https://doi.org/10.1186/s13059-015-0769-z

Vincenz, C., & Kerppola, T. K. (2008). Different polycomb group CBX family proteins associate with distinct regions of chromatin using nonhomologous protein sequences. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(43), 16572–7. https://doi.org/10.1073/pnas.0805317105

Vincenzo Pirrotta. (2017). *Polycomb Group Proteins*. (S. Tenney, Ed.).

Vizan, P., Beringer, M., Ballare, C., & Di Croce, L. (2015). Role of PRC2-associated factors in stem cells and disease. *FEBS Journal*, *282*(9), 1723–1735. https://doi.org/10.1111/febs.13083

Vizàn, P., Beringer, M., Ballaré, C., & Di Croce, L. (2015). Role of PRC2-associated factors in stem cells and disease. *FEBS Journal*, *282*(9), 1723–1735. https://doi.org/10.1111/febs.13083

Voigt, P., LeRoy, G., Drury, W. J., Zee, B. M., Son, J., Beck, D. B., … Reinberg, D. (2012). Asymmetrically modified nucleosomes. *Cell*, *151*(1), 181–193. https://doi.org/10.1016/j.cell.2012.09.002

von Schimmelmann, M., Feinberg, P. A., Sullivan, J. M., Ku, S. M., Badimon, A., Duff, M. K., … Schaefer, A. (2016). Polycomb repressive complex 2 (PRC2) silences genes responsible for neurodegeneration. *Nature Neuroscience*, *2*(August), 1–14. https://doi.org/10.1038/nn.4360

Voncken, J. W., Roelen, B. A., Roefs, M., de Vries, S., Verhoeven, E., Marino, S., … van Lohuizen, M. (2003). Rnf2 (Ring1b) deficiency causes gastrulation arrest and cell cycle inhibition. *Proc Natl Acad Sci U S A*, *100*(5), 2468–2473. https://doi.org/10.1073/pnas.0434312100

Walker, E., Chang, W. Y., Hunkapiller, J., Cagney, G., Garcha, K., Torchia, J., … Stanford, W. L. (2010). Polycomb-like 2 Associates with PRC2 and Regulates Transcriptional Networks during Mouse Embryonic Stem Cell Self-Renewal and Differentiation. *Cell Stem Cell*, *6*(2), 153–166. https://doi.org/10.1016/j.stem.2009.12.014

Walter, M., Tessandier, A., Perez-Palacios, R., & Bourc'his, D. (2016). An epigenetic switch ensures transponson repressioin upon dynamic loss of DNA methylation in embryonic stem cells. *eLife*. https://doi.org/10.1017/CBO9781107415324.004

Walters, Z. S., Villarejo-Balcells, B., Olmos, D., Buist, T. W. S., Missiaglia, E., Allen, R., … Shipley, J. (2014). JARID2 is a direct target of the PAX3-FOXO1 fusion protein and inhibits myogenic differentiation of rhabdomyosarcoma cells. *Oncogene*, *33*(9), 1148–57. https://doi.org/10.1038/onc.2013.46

Wang, D., Tyson, M. D., Jackson, S. S., & Yadegari, R. (2006). Partially redundant functions of two SET-domain polycomb-group proteins in controlling initiation of seed development in Arabidopsis. *Proc Natl Acad Sci U S A*, *103*(35), 13244–13249. https://doi.org/10.1073/pnas.0605551103

Wang, H., Cao, R., Xia, L., Erdjument-Bromage, H., Borchers, C., Tempst, P., & Zhang, Y. (2001). Purification and functional characterization of a histone H3-lysine 4-specific methyltransferase. *Molecular Cell*, *8*(6), 1207–1217. https://doi.org/10.1016/S1097-2765(01)00405-1

Wang, H., Wang, L., Erdjument-Bromage, H., Vidal, M., Tempst, P., Jones, R. S., & Zhang, Y. (2004). Role of histone H2A ubiquitination in Polycomb silencing. *Letters to Nature*, *431*(October), 873–877. https://doi.org/10.1038/nature02966.1.

Wang, K. C., Yang, Y. W., Liu, B., Sanyal, A., Corces-Zimmerman, R., Chen, Y., … Chang, H. Y. (2011). A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature*, *472*(7341), 120–124. https://doi.org/10.1038/nature09819

Wang, S., He, F., Xiong, W., Gu, S., Liu, H., Zhang, T., … Chen, Y. (2007). Polycomblike-2-deficient mice exhibit normal left-right asymmetry. *Developmental Dynamics*, *236*(3), 853–861. https://doi.org/10.1002/dvdy.21070

Wang, X., Goodrich, K. J., Gooding, A. R., Youmans, D. T., Cech, T. R., Wang, X., … Davidovich, C. (2017). Targeting of Polycomb Repressive Complex 2 to RNA by Short Repeats of Consecutive Guanines Article Targeting of Polycomb Repressive Complex 2 to RNA by Short Repeats of Consecutive Guanines. *Molecular Cell*, *65*(6), 1056–1067.e5. https://doi.org/10.1016/j.molcel.2017.02.003

Wassef, M., & Margueron, R. (2017). The Multiple Facets of PRC2 Alterations in Cancers. *Journal of Molecular Biology*, *429*(13), 1978–1993. https://doi.org/10.1016/j.jmb.2016.10.012

Wear, H. M., McPike, M. J., & Watanabe, K. H. (2016). From primordial germ cells to primordial follicles: a review and visual representation of early ovarian development in mice. *Journal of Ovarian Research*, *9*(1), 36. https://doi.org/10.1186/s13048-016-0246-7

Wiederschain, D., Chen, L., Johnson, B., Bettano, K., Jackson, D., Taraszka, J., … Benson, J. D. (2007). Contribution of polycomb homologues Bmi-1 and Mel-18 to medulloblastoma pathogenesis. *Molecular and Cellular Biology*, *27*(13), 4968–79. https://doi.org/10.1128/MCB.02244-06

Willingham, A. T., Orth, A. P., Batalov, S., Peters, E. C., Wen, B. G., Hogenesch, J. B., & Schultz, P. G. (2005). A Strategy for Probing the Function of Noncoding RNAs Finds a Repressor of NFAT, *309*(September), 1570–1574.

Woo, C. J., Kharchenko, P. V., Daheron, L., Park, P. J., & Kingston, R. E. (2010). A Region of the Human HOXD Cluster that Confers Polycomb-Group Responsiveness. *Cell*, *140*(1), 99–110. https://doi.org/10.1016/j.cell.2009.12.022

Wu, H., Coskun, V., Tao, J., Xie, W., Ge, W., Yoshikawa, K., … Sun, Y. E. (2010). Dnmt3a-dependent nonpromoter DNA methylation facilitates transcription of neurogenic genes. *Science*, *329*(5990), 444–448. https://doi.org/329/5990/444 [pii]\r10.1126/science.1190485

Wu, X., Johansen, J. V., & Helin, K. (2013a). Fbxl10/Kdm2b Recruits Polycomb Repressive Complex 1 to CpG Islands and Regulates H2A Ubiquitylation. *Molecular Cell*, *49*(6), 1134–1146. https://doi.org/10.1016/j.molcel.2013.01.016

Wu, X., Johansen, J. V., & Helin, K. (2013b). Fbxl10/Kdm2b Recruits Polycomb Repressive Complex 1 to CpG Islands and Regulates H2A Ubiquitylation. *Molecular Cell*, *49*(6), 1134–1146. https://doi.org/10.1016/j.molcel.2013.01.016

Xu, J., Shao, Z., Li, D., Xie, H., Kim, W., Huang, J., … Orkin, S. H. (2015). Developmental control of Polycomb subunit composition by GATA factors mediates a switch to non-canonical functions. *Mol Cell*, *57*(2), 304–316. https://doi.org/10.1523/JNEUROSCI.5473-10.2011.Loss

Yamaguchi, S., Shen, L., Liu, Y., Sendler, D., & Zhang, Y. (2013). Role of Tet1 in erasure of genomic imprinting. *Nature*, *504*(7480), 460–4. https://doi.org/10.1038/nature12805

Yap, K. L., Li, S., Munoz-Cabello, A. M., Raguz, S., Zeng, L., Mujtaba, S., … Zhou, M. (2010). Article Molecular Interplay of the Noncoding RNA ANRIL and Methylated Histone H3 Lysine 27 by Polycomb CBX7 in Transcriptional Silencing of INK4a, 662–674. https://doi.org/10.1016/j.molcel.2010.03.021

Yap, K. L., Li, S., Muñoz-Cabello, A. M., Raguz, S., Zeng, L., Mujtaba, S., … Zhou, M. M. (2010). Molecular Interplay of the Noncoding RNA ANRIL and Methylated Histone H3 Lysine 27 by Polycomb CBX7 in Transcriptional Silencing of INK4a. *Molecular Cell*, *38*(5), 662–674. https://doi.org/10.1016/j.molcel.2010.03.021

Yokobayashi, S., Liang, C. Y., Kohler, H., Nestorov, P., Liu, Z., Vidal, M., … Peters, a. H. (2013). Re: PRC1 coordinates timing of sexual differentiation of female primordial germ cells. *Nature*, *495*, 236–240. https://doi.org/10.1016/j.juro.2013.07.027

Yu, M., Mazor, T., Huang, H., Huang, H. T., Kathrein, K. L., Woo, A. J., … Cantor, A. B. (2012). Direct Recruitment of Polycomb Repressive Complex 1 to Chromatin by Core Binding Transcription Factors. *Molecular Cell*, *45*(3), 330–343. https://doi.org/10.1016/j.molcel.2011.11.032

Yuan, W., Wu, T., Fu, H., Dai, C., Wu, H., Liu, N., … Zhu, B. (2012). Dense chromatin activates Polycomb repressive complex 2 to regulate H3 lysine 27 methylation. *Science (New York, N.Y.)*, *337*(6097), 971–5. https://doi.org/10.1126/science.1225237

Yuan, W., Xu, M., Huang, C., Liu, N., Chen, S., & Zhu, B. (2011). H3K36 methylation antagonizes PRC2-mediated H3K27 methylation. *Journal of Biological Chemistry*, *286*(10), 7983–7989. https://doi.org/10.1074/jbc.M110.194027

Zhang, Z., Jones, A., Sun, C. W., Li, C., Chang, C. W., Joo, H. Y., … Wang, H. (2011). PRC2 complexes with JARID2, MTF2, and esPRC2p48 in ES cells to modulate ES cell pluripotency and somatic cell reprograming. *Stem Cells*, *29*(2), 229–240. https://doi.org/10.1002/stem.578

Zhao, J., Ohsumi, T. K., Kung, J. T., Ogawa, Y., Grau, D. J., Sarma, K., … Lee, J. T. (2010). Article Genome-wide Identification of Polycomb-Associated RNAs by RIP-seq. *Molecular Cell*, *40*(6), 939–953. https://doi.org/10.1016/j.molcel.2010.12.011

Zhao, J., Sun, B. K., Erwin, J. A., Song, J., & Lee, J. T. (2008). Polycomb Proteins Targeted by a Short Repeat RNA to the Mouse X Chromosome, *215*(October).