

Université de Montréal

**L'identification de nouvelles activités chez les complexes Polycomb les lient aux structures d'ADN non-
canoniques**

Par

Célia Alecki

Département de Biochimie

Faculté de médecine

Thèse présentée en vue de l'obtention du grade docteur en biochimie

Juin 2020

© Célia Alecki 2020

Université de Montréal
Département de biochimie, Faculté de médecine

Cette thèse intitulée

**L'identification de nouvelles activités chez les complexes Polycomb les lient aux structures d'ADN non-
canoniques**

Présentée par

Célia Alecki

A été évaluée par un jury composé des personnes suivantes

Éric Lécuyer

Président-rapporteur

Nicole Francis

Directrice de recherche

Marc Drolet

Membre du jury

Jacques Côté

Examineur externe

Résumé

Les protéines du groupe Polycomb (PcG) sont des protéines essentielles et conservées, qui forment deux complexes principaux, PRC1 et PRC2, qui sont recrutés au niveau de la chromatine et qui répriment stablement l'expression génique. Chez *Drosophila melanogaster*, les complexes Polycomb sont recrutés à des éléments d'ADN appelés éléments de réponse Polycomb (PREs) pour réprimer la transcription. PREs sont des éléments mémoires permutable qui peuvent maintenir la répression ou l'expression génique. Malgré des dizaines d'années d'étude, des questions fondamentales sur le fonctionnement du système PcG subsistent. 1) Comment les protéines PcG sont recrutées aux PREs uniquement lors du contexte développemental approprié, et comment les PREs peuvent conduire à la fois à l'activation et à la répression stable. 2) Comment les complexes PcG répriment la transcription, et si cela implique de nouvelles activités biochimiques et interactions. 3) Comment la répression dépendante des PcG peut-elle être propagé à travers le cycle cellulaire. La recherche de nouvelles activités biochimiques pour les complexes PcG pouvant répondre à ces questions fait l'objet de cette thèse.

Les PREs sont transcrits en ARN qui pourraient donner la spécificité de contexte pour recruter les protéines PcG. Nous avons supposé que des R-loops puissent se former aux PREs, et être reconnues par les complexes Polycomb, ce que nous avons testé dans le chapitre 2. Nous avons identifié les séquences formant des R-loops dans des embryons et une lignée cellulaire de *Drosophila melanogaster*, et nous avons trouvé que ~30% des PREs forment des R-loops. Nous avons découvert que les PREs ayant formé des R-loops ont une plus forte probabilité d'être liés par les protéines PcG *in vivo* et *in vitro*. PRC2 lie des milliers d'ARN *in vivo*, mais aucune fonction claire n'y a été associée. En utilisant des expériences *in vitro*, nous avons identifié une activité d'invasion de brins pour PRC2 qui induit la formation d'hybride ARN-ADN, la partie principale d'une R-loop. Dans ce chapitre, nous avons trouvé que les PREs forment des R-loops, et sont impliquées dans le recrutement des protéines PcG qui induisent la répression génique stable. Nous avons découvert une activité d'invasion de brins pour PRC2 qui pourrait impliquer ce complexe Polycomb dans la formation de R-loops *in vivo*.

Dans le chapitre 3, nous avons identifié une activité similaire à celle de la topoisomérase I associée avec PRC1 et la région C-terminale de sa sous-unité PSC (PSC-CTR). PRC1 et PSC-CTR peuvent relaxer un plasmide surenroulé négativement et ajouter des supertours négatifs à un plasmide relaxé en absence de topoisomérase. Cette activité suggère que la régulation de la topologie de l'ADN puisse être un nouveau

mécanisme utilisé par les complexes PcG. PRC1 peut résoudre les R-loops formées sur un ADN négativement surenroulé *in vitro*. Une fonction possible pour cette activité de topoisomérase peut être la régulation des R-loops, dont la stabilité dépend à la fois de la séquence d'ADN et de la topologie de l'ADN environnant, *in vivo*.

Dans cette thèse, nous avons identifié de nouvelles activités chez les complexes PcG : une activité d'invasion de brins pour PRC2 et une activité similaire à celle des topoisomérases pour PRC1. Ces deux activités impliquent les complexes PcG dans la formation et la résolution de R-loops. De plus, ces deux complexes peuvent reconnaître les R-loops et sont recrutés aux PREs ayant formé ces structures. En conclusion, nous avons identifié de nouvelles activités pour les complexes Polycomb PRC1 et PRC2 qui les lient à la formation, la reconnaissance et la résolution de R-loops.

Mots-clés

Protéine du groupe Polycomb, PRC1, PRC2, PRE, R-loop, invasion de brins, topoisomérase, PSC *Drosophila melanogaster*, chromatine

Abstract

Polycomb group (PcG) proteins are conserved, essential proteins, which assemble in two main complexes, PRC1 and PRC2, which are targeted to chromatin and stably repress gene expression. In *Drosophila melanogaster*, Polycomb complexes are targeted to DNA elements called Polycomb response elements (PREs) to repress transcription. PREs are switchable memory elements that can maintain either gene repression or gene activation. Despite decades of study, fundamental questions about how the PcG system functions remain. These include: 1) how PcG proteins are targeted to PREs only in the appropriate developmental context, and how PREs can mediate both stable activation and repression; 2) how PcG complexes repress transcription, and whether it involves novel biochemical mechanisms and interactions; 3) how PcG repression can be propagated through cell cycles. The search for new biochemical activities for PcG complexes that may answer these questions is the topic of this thesis.

PREs are transcribed into RNAs which may give the context specificity to recruit PcG proteins. We hypothesized that R-loops may form at PREs, and be recognized by PcG complexes, which we tested in Chapter 2. We identified R-loop forming sequences in *Drosophila melanogaster* embryos and tissue culture cells, and found that ~30% of the PREs form R-loops. We found that PREs which have formed R-loops are more likely to be bound by PcG proteins both *in vivo* and *in vitro*. PRC2 binds to thousand RNA *in vivo* but no clear activity has been associated with it. Using *in vitro* assays, we identified a strand exchange activity for PRC2 which induces the formation of RNA-DNA hybrid, the main part of an R-loop. In this chapter, we have found that PREs form R-loops and are involved in the targeting of PcG proteins which induce stable gene repression. We have discovered an RNA strand exchange activity for PRC2 which may involve this Polycomb complex in the formation of R-loops *in vivo*.

In Chapter 3, we identified a type I topoisomerase-like activity associated with PRC1 and the C-terminal region of its subunit PSC (PSC-CTR). PRC1 and PSC-CTR can relax a negatively supercoiled plasmid and add negative coils to a relaxed plasmid in absence of topoisomerase. This activity suggests regulation of DNA topology may be a novel mechanism used by PcG complexes. PRC1 can resolve R-loops formed on negatively supercoiled DNA *in vitro*. One role for the topoisomerase-like activity may be to regulate R-loops, whose stability depends on both the DNA sequence and the topology of the surrounding DNA, *in vivo*.

In this thesis, we identified new activities for Polycomb group complexes: an RNA strand exchange activity for PRC2 and a topoisomerase-like activity for PRC1. Both activities link PcG complexes to the formation

and resolution of R-loops. In addition, both complexes can recognize R-loops and are recruited to PREs which have formed these structures. In conclusion, we have identified new nucleic acid-based activities for the Polycomb complexes PRC1 and PRC2, which link them to the formation, recognition and resolution of R-loops.

Keywords

Polycomb group protein, PRC1, PRC2, PRE, R-loop, strand exchange, topoisomerase, PSC, *Drosophila melanogaster*, chromatin

Table des matières

| | |
|---|-------|
| Résumé..... | I |
| Abstract | III |
| Table des matières | V |
| Liste des tableaux..... | XIII |
| Liste des tableaux..... | XIII |
| Liste des tableaux supplémentaires | XIII |
| Liste des figures..... | XIV |
| Liste des figures..... | XIV |
| Liste des figures supplémentaires..... | XVI |
| Liste des abréviations..... | XVII |
| Remerciements | XXIII |
| 1 Introduction..... | 1 |
| 1.1 La chromatine..... | 1 |
| 1.1.1 La chromatine..... | 1 |
| 1.1.2 Les modifications post-traductionnelles des histones..... | 1 |
| 1.1.2.1 Les marques actives | 1 |
| 1.1.2.2 Les marques répressives | 2 |
| 1.1.2.3 Les domaines bivalents | 2 |
| 1.1.3 La propagation des modifications post-traductionnelles d’histone à travers le cycle cellulaire 3 | |
| 1.1.4 La structure de la chromatine | 3 |
| 1.2 Les protéines du groupe Polycomb..... | 4 |
| 1.2.1 Les protéines du groupe Polycomb : découverte et introduction | 4 |

| | | |
|---------|--|----|
| 1.2.1.1 | Les gènes Hox et leur répression par les protéines du groupe Polycomb durant le développement | 4 |
| 1.2.1.2 | Introduction sur les protéines du groupe Polycomb..... | 5 |
| 1.2.2 | PRC2..... | 6 |
| 1.2.2.1 | PRC2 est une méthyltransférase | 6 |
| 1.2.2.2 | PRC2 est conservé chez d'autres espèces | 8 |
| 1.2.2.3 | La reconnaissance des modifications post-traductionnelles des histones par PRC2 module son activité | 10 |
| 1.2.2.4 | Les sous-unités accessoires de PRC2 contribuent à son activité <i>in vivo</i> | 11 |
| 1.2.2.5 | La structure de PRC2 | 14 |
| 1.2.2.6 | La liaison des ARN..... | 14 |
| 1.2.3 | PRC1..... | 17 |
| 1.2.3.1 | Composition | 17 |
| 1.2.3.2 | Le rôle central de PSC dans la répression induite par PRC1..... | 19 |
| 1.2.3.3 | Ph est essentiel pour établir des interactions à longues distances..... | 20 |
| 1.2.3.4 | dRING ubiquitinye la lysine 118 de l'histone H2A..... | 21 |
| 1.2.3.5 | PRC1 reste lié aux chromosomes lors de la mitose..... | 22 |
| 1.2.3.6 | PRC1 est conservé chez d'autres espèces | 22 |
| 1.2.4 | Le complexe PhoRC | 25 |
| 1.2.5 | Les complexes Polycomb participent à leur recrutement mutuel | 26 |
| 1.3 | Les éléments de réponse Polycomb..... | 28 |
| 1.3.1 | La définition fonctionnelle des éléments de réponse Polycomb | 28 |
| 1.3.1.1 | Un élément mémoire | 29 |
| 1.3.1.2 | Un élément mémoire permutable | 29 |
| 1.3.1.3 | La répression sensible à l'appariement des PREs (<i>Pairing-sensitive silencing</i> ou PSS) ... | 32 |
| 1.3.2 | Les PREs permettent la formation de domaines Polycomb répressifs | 32 |
| 1.3.2.1 | Les PREs sont des sites de nucléation pour les protéines du groupe Polycomb..... | 32 |
| 1.3.2.2 | Les PREs sont impliqués dans la formation de domaines répressifs..... | 33 |
| 1.3.2.3 | Les PREs, entre répression et expression | 33 |
| 1.3.2.4 | Les PREs ne sont pas nécessaires pour la transmission de la répression durant un nombre limité de cycles cellulaires | 34 |
| 1.3.3 | Les propriétés de la séquence ADN..... | 35 |

| | | |
|---------|--|----|
| 1.3.3.1 | Les PREs sont liés par des facteurs de transcription | 36 |
| 1.3.3.2 | Une absence de séquence consensus | 37 |
| 1.3.4 | Les PREs ne sont pas tous identiques..... | 38 |
| 1.3.4.1 | La localisation des PREs par rapport aux gènes qu'ils régulent | 38 |
| 1.3.4.2 | Tous les PREs ne sont pas égaux | 39 |
| 1.3.5 | Les PREs sont transcrits en ARN | 39 |
| 1.3.6 | Les éléments de réponse Polycomb chez les autres espèces | 40 |
| 1.3.6.1 | Les PREs chez les mammifères | 40 |
| 1.3.6.2 | Les PREs chez les plantes..... | 44 |
| 1.4 | R-loops..... | 47 |
| 1.4.1 | Les R-loops sont source d'instabilité génomique..... | 47 |
| 1.4.2 | Les facteurs contribuant à la formation des R-loops | 48 |
| 1.4.2.1 | La séquence de l'ADN et de l'ARN..... | 48 |
| 1.4.2.2 | La topologie de l'ADN..... | 49 |
| 1.4.2.3 | Les protéines induisant la formation de R-loops | 50 |
| 1.4.3 | La résolution des R-loops | 51 |
| 1.4.3.1 | Les RNase H | 51 |
| 1.4.3.2 | Les hélicases ARN-ADN..... | 53 |
| 1.4.3.3 | Les topoisomérases | 53 |
| 1.4.4 | Les différentes techniques utilisées pour identifier les R-loops | 54 |
| 1.4.4.1 | Le séquençage au bisulfite, bis-seq..... | 54 |
| 1.4.4.2 | L'anticorps S9.6, DRIP | 55 |
| 1.4.4.3 | La RNase H1 catalytiquement inactive, R-ChIP ou DRIVE | 56 |
| 1.4.5 | Les régions du génome formant des R-loops chez différentes espèces | 56 |
| 1.4.5.1 | Chez les mammifères | 56 |
| 1.4.5.2 | Chez les levures | 57 |
| 1.4.5.3 | Chez les plantes | 58 |
| 1.4.6 | Les R-loops ne sont pas uniquement source d'instabilité génomique..... | 58 |
| 1.4.6.1 | Le rôle des R-loops se formant aux télomères..... | 59 |
| 1.4.6.2 | Le rôle des R-loops lors des cassures dans l'ADN..... | 59 |

| | | |
|---------|--|----|
| 1.4.6.3 | Les R-loops participent à la régulation de l'expression génique | 61 |
| 1.5 | La topologie de l'ADN | 62 |
| 1.5.1 | La topologie de l'ADN | 62 |
| 1.5.1.1 | Le surenroulement et la relaxation de l'ADN | 62 |
| 1.5.1.2 | Le changement du nombre de supertours ou de <i>linking number</i> | 63 |
| 1.5.1.3 | La topologie de l'ADN dans les cellules | 63 |
| 1.5.1.4 | La transcription et réplication de l'ADN altèrent la topologie de l'ADN et peuvent être la cause d'instabilité génomique | 64 |
| 1.5.2 | Les topoisomérases | 65 |
| 1.5.2.1 | La topoisomérase de type I (ou type IB) | 66 |
| 1.5.2.2 | La topoisomérase de type II | 68 |
| 1.5.2.3 | La topoisomérase de type III (ou type IA) | 68 |
| 1.5.2.4 | Les recombinaisons à site spécifique et les protéines similaires aux topoisomérases | 70 |
| 1.5.3 | Les fonctions des topoisomérases | 71 |
| 1.5.3.1 | La transcription | 71 |
| 1.5.3.2 | La réplication de l'ADN | 72 |
| 1.5.3.3 | La mitose | 73 |
| 1.5.3.4 | La recombinaison | 73 |
| 1.5.3.5 | Les topoisomérases régulent la formation et la résolution de structures d'acides nucléiques non-canoniques | 74 |
| 1.6 | Hypothèses et objectifs de la thèse | 75 |
| 1.6.1 | Objectif 1 (Chapitre 2) | 75 |
| 1.6.2 | Objectif 2 (Chapitre 2) | 76 |
| 1.6.3 | Objectif 3 (Chapitre 3) | 76 |
| 2 | Article 1: RNA-DNA strand exchange by the <i>D. melanogaster</i> Polycomb complex PRC2 | 77 |
| 2.1 | Abstract | 79 |
| 2.2 | Main Text | 80 |
| 2.2.1 | Introduction | 80 |
| 2.2.2 | Results | 81 |
| 2.2.3 | Discussion | 86 |

| | | |
|---------|---|-----|
| 2.3 | Figures | 90 |
| 2.4 | Supplementary tables | 122 |
| 2.5 | Methods | 134 |
| 2.6 | Acknowledgements | 143 |
| 2.7 | References | 144 |
| 3 | Article 2: A Topoisomerase I like activity associated with the Polycomb Group protein PSC | 152 |
| 3.1 | Abstract | 154 |
| 3.2 | Main text | 155 |
| 3.2.1 | Introduction..... | 155 |
| 3.2.2 | Results | 156 |
| 3.2.3 | Discussion | 162 |
| 3.3 | Figures | 165 |
| 3.4 | Methods | 183 |
| 3.5 | References | 187 |
| 4 | Discussion | 190 |
| 4.1 | Résumé des chapitres 2 et 3 | 190 |
| 4.2 | Les R-loops peuvent-elles déterminer le statut transcriptionnel d'un gène | 190 |
| 4.2.1 | Les séquences formant des R-loops chez <i>D. melanogaster</i> | 190 |
| 4.2.1.1 | Les R-loops se forment aux sites d'initiation et de terminaison de la transcription..... | 190 |
| 4.2.1.2 | Les marques épigénétiques..... | 191 |
| 4.2.2 | Des R-loops sont régulées pendant le développement | 193 |
| 4.2.2.1 | Chez <i>D. melanogaster</i> | 193 |
| 4.2.2.2 | Chez les mammifères | 194 |
| 4.2.2.3 | Chez les plantes | 194 |
| 4.2.3 | Les R-loops sont-elles associées avec d'autres caractéristiques..... | 195 |
| 4.2.3.1 | Les quadruplexes de G | 195 |
| 4.2.3.2 | La méthylation de l'ARN | 196 |

| | | |
|---------|---|-----|
| 4.3 | Les R-loops sont-elles une caractéristique des PREs et peuvent-elles être utilisées pour les identifier chez d'autres espèces ? | 197 |
| 4.3.1 | Les motifs aux PREs | 197 |
| 4.3.2 | Les motifs présents aux PREs ayant formés des R-loops sont différents de ceux qui n'en ont pas formés..... | 197 |
| 4.3.2.1 | Les motifs présents aux PREs qui ont ou n'ont pas formé de R-loops..... | 197 |
| 4.3.2.2 | Les facteurs de transcription aux PREs peuvent-ils reconnaître leur séquence cible quand elle est impliquée dans une R-loop ? | 198 |
| 4.4 | Les R-loops sont-elles impliquées dans le recrutement de protéines ?..... | 199 |
| 4.4.1 | Le recrutement des protéines du groupe Polycomb par les R-loops..... | 199 |
| 4.4.1.1 | PRC1 et PRC2 reconnaissent les R-loops, mais quelle partie ? | 199 |
| 4.4.1.2 | Le recrutement des complexes Polycomb par les R-loops chez les mammifères..... | 200 |
| 4.4.2 | D'autres protéines sont recrutées par les R-loops..... | 201 |
| 4.4.2.1 | Des protéines recrutées ou liant des R-loops ont des fonctions diverses | 201 |
| 4.4.2.2 | Un recrutement indirect qui ne nécessite pas la reconnaissance directe d'une R-loop..... | 202 |
| 4.4.3 | Les R-loops peuvent-elles se former <i>en trans</i> ? | 202 |
| 4.5 | Les R-loops sont-elles un marque-page du cycle cellulaire ?..... | 203 |
| 4.5.1 | L'hypothèse du marque-page mitotique..... | 203 |
| 4.5.2 | Les R-loops durant le cycle cellulaire | 204 |
| 4.5.2.1 | Les ARN et la transcription pendant la mitose..... | 204 |
| 4.5.2.2 | Les R-loops peuvent-elles participer au marque-page mitotique..... | 204 |
| 4.6 | PRC2 induit l'invasion de brins | 205 |
| 4.6.1 | Le mécanisme par lequel PRC2 pourrait induire l'invasion de brins..... | 205 |
| 4.6.2 | PRC2 nécessite un facteur manquant pour induire la formation de R-loops lorsqu'une extrémité d'ADN n'est pas présente | 209 |
| 4.6.2.1 | Les protéines accessoires du complexe PRC2 | 209 |
| 4.6.2.2 | Les protéines interagissant avec PRC2 | 210 |
| 4.6.2.3 | Autres | 210 |
| 4.6.3 | L'activité d'invasion de brins de PRC2 est-elle conservée ?..... | 210 |

| | | |
|---------|--|-----|
| 4.7 | Quelle pourrait-êre la fonction de l'activité d'invasion de brins de PRC2 <i>in vivo</i> ?..... | 211 |
| 4.7.1 | PRC2 et les hybrides ARN-ADN aux sites de cassures dans l'ADN | 211 |
| 4.7.1.1 | Les fonctions hybrides ARN-ADN dans la réparation des cassures dans l'ADN | 211 |
| 4.7.1.2 | Les protéines du groupe Polycomb sont recrutées aux sites de cassures double-brin | 211 |
| 4.7.2 | L'activité d'invasion de brins de PRC2 peut-elle participer au maintien de la transcription ? | 212 |
| 4.7.3 | L'activité d'invasion de brins de PRC2 est-il un mécanisme de surveillance permettant la détection de gènes qui échappent à la répression? | 215 |
| 4.8 | L'activité de topoisomérase de PSC | 216 |
| 4.8.1 | Le mécanisme par lequel PSC-CTR modifie la topologie de l'ADN..... | 216 |
| 4.8.1.1 | Le mécanisme par lequel PSC-CTR modifie la topologie de l'ADN..... | 216 |
| 4.8.1.2 | PSC-CTR forme-t-il un complexe covalent avec l'ADN ?..... | 217 |
| 4.8.1.3 | PSC-CTR n'utilise pas une tyrosine comme résidu catalytique | 217 |
| 4.8.1.4 | La structure de PSC pourrait nous permettre de déterminer le mécanisme par lequel elle modifie la topologie de l'ADN | 218 |
| 4.8.1.5 | Quel domaine de PSC est responsable de son activité de topoisomérase ?..... | 219 |
| 4.8.2 | L'activité de topoisomérase de PSC s'exerce-t-elle au sein du complexe PRC1 ou indépendamment de celui-ci ?..... | 219 |
| 4.8.3 | L'activité de topoisomérase de PRC1 et PSC est-elle conservée ?..... | 220 |
| 4.8.3.1 | Chez <i>Drosophila melanogaster</i> | 220 |
| 4.8.3.2 | Chez les mammifères | 220 |
| 4.8.3.3 | Chez les plantes | 221 |
| 4.9 | Quelles pourraient être les fonctions de l'activité de topoisomérase de PSC <i>in vivo</i> ?..... | 221 |
| 4.9.1 | La protéine PSC pourrait-elle compenser la perte d'une topoisomérase ? | 221 |
| 4.9.2 | PSC stimule l'activité de la topoisomérase I..... | 222 |
| 4.9.3 | La compaction de la chromatine et la relaxation de supertours sont-elles deux activités liées ? | 224 |
| 4.9.4 | Est-ce que PSC est impliqué dans la résolution des R-loops dont la stabilité dépend du surenroulement de l'ADN <i>in vivo</i> ?..... | 225 |

| | | |
|---------|--|-----|
| 4.9.4.1 | Les R-loops sont stabilisées par la présence de surenroulement négatif | 225 |
| 4.9.4.2 | Est-ce que PRC1 est impliqué dans la résolution d'autres structures d'ADN non-canoniques ?..... | 226 |
| 4.10 | Conclusions générales | 227 |
| 4.10.1 | Vers une nouvelle définition de PRE ?..... | 227 |
| 4.10.2 | La découverte de deux nouvelles activités chez les complexes Polycomb | 228 |
| 4.10.3 | Les protéines du groupe Polycomb sont impliquées dans la formation, la reconnaissance et la résolution de structures d'acides nucléiques non-canoniques..... | 228 |
| 5 | Références bibliographiques | 231 |

Liste des tableaux

Liste des tableaux

| | |
|---|----|
| Tableau 1-1 Liste des sous-unités du corps du complexe PRC2 chez <i>Drosophila melanogaster</i> , les mammifères et les plantes ainsi que leurs sous-unités accessoires | 9 |
| Tableau 1-2 Liste des sous-unités du complexe PRC1 chez <i>Drosophila melanogaster</i> , les mammifères et les plantes | 25 |

Liste des tableaux supplémentaires

| | |
|---|-----|
| Supplementary Table 2-1 Sequences of primers and oligonucleotides used | 123 |
| Supplementary Table 2-2 List of PRES used in this work | 132 |
| Supplementary Table 2-3 Data sets used for comparison with DRIP-seq data..... | 133 |

Liste des figures

Liste des figures

| | |
|--|-----|
| Figure 1-1 Le complexe PRC2 : composition et fonction..... | 7 |
| Figure 1-2 Le complexe PRC1 : composition et fonction..... | 18 |
| Figure 1-3 Les complexes Polycomb PR-DUB et dRAF | 22 |
| Figure 1-4 Le complexe Polycomb PhoRC | 26 |
| Figure 1-5 Les complexes PRC1, PRC2 et PhoRC participent à leur recrutement mutuel | 28 |
| Figure 1-6 Schémas illustrant la définition fonctionnelle des éléments de réponse Polycomb (PREs) | 31 |
| Figure 1-7 Trois modèles illustrant la phase de recrutement initial des complexes Polycomb à un PRE... 37 | |
| Figure 1-8 Exemples d'éléments pouvant jouer le rôle de PRE chez les mammifères | 43 |
| Figure 1-9 Schéma de la répression du locus FLC lors de la vernalisation impliquant des ARNnc et PRC2 46 | |
| Figure 1-10 Les facteurs contribuant à la formation d'une R-loop | 49 |
| Figure 1-11 Les protéines participant à la résolution des R-loops | 52 |
| Figure 1-12 Les différentes techniques permettant l'identification des séquences formant des R-loops . 55 | |
| Figure 1-13 Les rôles potentiels des hybrides ARN-ADN dans la réparation des cassures d'ADN double-brins | 61 |
| Figure 1-14 La topologie de l'ADN..... | 63 |
| Figure 1-15 Les topoisomérases..... | 66 |
| Figure 2-1 R-loops form at Drosophila PREs and correlate with a repressed state | 91 |
| Figure 2-2 PRC1 and PRC2 recognize R-loops | 93 |
| Figure 2-3 PRC2 has RNA-DNA strand exchange activity | 94 |
| Figure 2-4 Nuclease contamination cannot explain RNA-DNA hybrid formation by PRC2..... | 96 |
| Figure 2-5 Pattern of RNA used by PRC2 on EcoRI-digested DNA templates | 98 |
| Figure 2-6 Characteristics of PRC2 strand exchange activity | 99 |
| Figure 2-7 Model for the role of R-loop formation driven by PRC2 in PcG gene silencing | 100 |
| Figure 3-1 PRC1 and the PSC-CTR have a topoisomerase-like activity..... | 167 |
| Figure 3-2 Model depicting PRC1 induced DNA topology change by DNA wrapping associated with topoisomerase or cut and ligate activities | 168 |
| Figure 3-3 Characterization of PRC1 Topoisomerase-like activity. | 170 |

| | |
|---|-----|
| Figure 3-4 Topoisomerase-like activity co-fractionates with the PSC-CTR and topoisomerases are not detected in protein preparations. | 172 |
| Figure 3-5 PSC-CTR Topoisomerase-like activity is stimulated by KCl and doesn't require magnesium. . | 174 |
| Figure 3-6 PSC-CTR topoisomerase-like activity is inhibited by camptothecin but not by etoposide. | 176 |
| Figure 3-7 Topoisomerase-like activity of PSC truncations. | 178 |
| Figure 3-8 PSC topoisomerase-like activity doesn't required tyrosine. | 179 |
| Figure 3-9 PRC1 and PSC modify the topology of chromatin assembled plasmid. | 181 |
| Figure 3-10 DNA relaxation by PRC1 induces loss of R-loops associated with supercoiled plasmid. | 182 |
| Figure 4-1 Les R-loops, la transcription et les modifications post-traductionnelles d'histones | 193 |
| Figure 4-2 Schéma de structures d'ADN non-canoniques pouvant coexister avec une R-loop..... | 196 |
| Figure 4-3 Les motifs présents aux PREs ayant ou n'ayant pas formés de R-loops | 198 |
| Figure 4-4 Modèle des R-loops participant au marque-page mitotique..... | 205 |
| Figure 4-5 Schéma représentant des mécanismes par lesquels PRC2 peut induire la formation d'hybride ARN-ADN | 208 |
| Figure 4-6 Deux exemples possibles de la fonction de l'activité d'invasion de brins de PRC2 | 214 |
| Figure 4-7 PSC stimule la relaxation de l'ADN par la topoisomérase I..... | 223 |
| Figure 4-8 PSC compacte la chromatine et relaxe l'ADN | 224 |
| Figure 4-9 Modèle décrivant comment PRC2 et PRC1 avec leurs activités d'invasion de brins et de topoisomérase respectives peuvent participer au changement de statut d'un PRE ou d'un gène..... | 229 |

Liste des figures supplémentaires

| | |
|---|-----|
| Supplementary Figure 2-1 Overview of DRIP-seq analysis | 102 |
| Supplementary Figure 2-2 Relationship between R-loop formation, transcription, and PcG proteins..... | 104 |
| Supplementary Figure 2-3 Additional examples of R-loops..... | 105 |
| Supplementary Figure 2-4 Templates and proteins used to analyze PRC1 and PRC2 interactions with R-loops in vitro..... | 108 |
| Supplementary Figure 2-5 PRC1 and PRC2 interactions with nucleic acids..... | 110 |
| Supplementary Figure 2-6 PRC2 induces the formation of RNA-DNA hybrids that are sensitive to RNaseH and recognized by S9.6..... | 111 |
| Supplementary Figure 2-7 Time course of PRC2-mediated strand invasion and lack of requirement for nucleotides | 114 |
| Supplementary Figure 2-8 Other RNA and DNA binding proteins do not have strand exchange activity and activity co-fractionates with PRC2 | 116 |
| Supplementary Figure 2-9 PRC2 preparations do not contain detectable nuclease activity | 118 |
| Supplementary Figure 2-10 PRC2 RNA strand exchange activity on different PRE templates with different DNA ends..... | 119 |
| Supplementary Figure 2-11 Effect of order of addition and substrate concentration on PRC2-mediated strand exchange | 120 |
| Supplementary Figure 2-12 PRC2 can anneal ssDNA and RNA to form RNA:DNA hybrids..... | 121 |

Liste des abréviations

3C : chromosome conformation capture

A : adenine

ADN : acide désoxyribonucléique

ADNdb : ADN double-brin

ADNsb : ADN simple-brin

AEBP2 adipocyte enhancer-binding protein 2

ALT : alternative lengthening of telomeres

Antp : antennapedia

APOLO : auxin-regulated promoter loop

ARN : acide ribonucléique

ARNm : ARN messenger

ARNnc : ARN non codant

AS1 : asymmetric leaves 1

ASH1 : absent, small or homeotic discs 1

ASX : additional sex combs

AtBMI1 (A-C)

ATP : adenosine triphosphate

AtRING1 (a-b)

BAMBI : BMP and activin membrane-bound inhibitor homolog

BLM : Bloom syndrome protein

bx1 : bithoraxoid

C : cytosine

Cas : CRISPR associated protein

CBP : calcium-binding protein

CBX : chromobox protein homolog

ChIP : Chromatin immunoprecipitation

CLF : curly leaf

CLIP : crosslinking and immunoprecipitation

CPT : Camptothecin

CRISPR : clustered regularly interspaced short palindromic repeats

CTCF : CCCTC-binding factor
CTD : carboxy terminal domain
D. melanogaster : Drosophila melanogaster
dac : dachshund
DHX9 : DEAH box protein 9
D-loop : displacement loop
DNase : desoxyribonucléase
DNMT3B : DNA (cytosine-5)-methyltransferase 3B
DPP9 : dipeptidyl peptidase 9
dRAF : dRING-associated factor
DRIP : DNA-RNA hybrid immunoprecipitation
DSB : double-strand break
Dsp1 : dorsal switch protein 1
E(z) : enhancer of zeste
Eed : embryonic ectoderm development
EMF1 : embryonic flower 1
EMF2 : embryonic flower 2
EPOP : elongin BC and Polycomb repressive complex 2-associated protein
Esc : extra sex combs
Escl : extra sex combs-like
EZH1 : enhancer of zeste homolog 1
EZH2 : enhancer of zeste homolog 2
F : phenylalanine
Fab-7 : frontabdominal-7
FIE : fertilization independent endosperm
FIS2 : fertilization-independent seed 2
FLC : flowering locus C
FLRT2 : fibronectin leucin rich transmembrane protein 2
FOXG1 : Forkhead box protein G1
G : guanine
GADD45A : growth arrest and DNA damage inducible protein GADD45 alpha
GAF : GAGA factor

GATA3 : GATA transcription factor 3
Grh : grainy head
H : heure
H2A : histone 2A
H2AK118ub : ubiquitination de la lysine 118 de l'histone H2A
H2AK119ub : ubiquitination de la lysine 119 de l'histone H2A
H2B : histone 2B
H3 : histone 3
H3K27Ac : acétylation de la lysine 27 de l'histone 3
H3K27me1 : mono-méthylation de la lysine 27 de l'histone 3
H3K27me2 : di-méthylation de la lysine 27 de l'histone 3
H3K27me3 : tri-méthylation de la lysine 27 de l'histone 3
H3K36me2 : di-méthylation de la lysine 36 de l'histone 3
H3K36me3 : tri-méthylation de la lysine 36 de l'histone 3
H3K4me3 : tri-méthylation de la lysine 4 de l'histone 3
H3K9me3 : tri-méthylation de la lysine 9 de l'histone 3
H4 : histone 4
HOTAIR : HOX transcript antisense RNA
Hox : Homeobox
HoxC : homeobox C cluster
HoxD : homeobox D cluster
HPH1-3
HR : homologous recombination
in : engrailed
inv : invected
JMJD3 : jumonji domain-containing protein D3
kb : kilobase
KCl : chlorure de potassium
kDA : kilo Dalton
KDM2 : lysine-specific demethylase 2
KLF : kruppel-like factor
LHP1 : like heterochromatin protein 1

LiCl : chlorure de lithium
LINE-1 : long interspersed nuclear element 1
m6A : N6-methyladenosine
MDa : mega Dalton
Meg3 : maternally expressed 3
MRN : Mre11-Rad50-Nbs1
XPG : Xeroderma pigmentosum group G-complementing protein
MSI1 : multicopy suppressor of IRA 1
MSL1/2 : male-specific lethal 1/2
MTF2 : metal-response element-binding transcription factor 2
NaCl : chlorure de calcium
NHEJ : non-homologous end joining
NO66 : nucleolar protein 66
Nurf-55 : chromatin assembly factor 1, p55 subunit
ORC : origin recognition complex
ORF : open reading frame
PARP1 : Poly(ADP-ribose) polymerase 1
PARP2 : Poly(ADP-ribose) polymerase 2
Pc : Polycomb
PcG : Polycomb group
PCGF : Polycomb group RING finger protein
Pcl : Polycomblike
Ph : polyhomeotic
Ph-d : polyhomeotic distal
PHF1 : PHD finger protein 1
PHF19 : PHD finger protein 19
Pho : polyhomeotic
PhoRC : Pho repressive complex
Ph-p : polyhomeotic proximal
PICH : Plk-interacting checkpoint helicase
PRC1 : Polycomb repressive complex 1
PRC2 : Polycomb repressive complex 2

PR-DUB : Polycomb repressive deubiquitinase complex

PSC : posterior sex combs

PSC-CTR : C-terminal région de PSC

Psq : pipsqueak

qPCR : quantitative polymerase chain reaction

RBBP4/7 : retinoblastoma-binding protein 4/7

RING1 : RING finger protein

RIP : RNA immunoprecipitation

RNase : ribonucléase

RNase A : ribonuclease A

RNase H1 : ribonuclease H1

RNase H2 : : ribonuclease H2

RNase H2A : : ribonuclease H2 subunit A

RNase H2B : : ribonuclease H2 subunit B

RNase H2C : : ribonuclease H2 subunit C

RPA : replication protein A

S2 : Schneider 2

SAM : sterile alpha motif

Scx : sex combs extra or dRING

Scm : sex comb on midleg

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

SFMBT : Scm-related gene containing four mbt domains

Spo11 : meiotic recombination protein SPO11

Su(Z)12 : suppressor of zeste 12

Su(Z)2 : suppressor of zeste 2

SUZ12 : suppressor of zeste 12

SWI/SNF : SWItch/Sucrose Non-Fermentable

SWN : swinger

T : thymine

TAD : topologically associating domain

TDRD3 : Tudor domain-containing protein 3

FMRP : Fragile X mental retardation protein

TERRA : telomeric repeat-containing RNA

TET1 : methylcytosine dioxygenase TET1

T-loop : telomere loop

TOPRIM : topoisomerase-primase

TRF2 : telomeric repeat-binding factor 2

Trl : Thrithorax-like

Trx : Trithorax

TrxG : Trithorax group

TSS : transcription start site

TTS : transcription termination site

U : uracyle

UAS : upstream activating sequence

Ubx : ultrabithorax

UTX : Utx histone demethylase

vg : vestigial

VIN3 : vernalization intensive 3

VRN2 : vernalization 2

VRN5 : vernalization 5

Xist : X-inactive specific transcript

Y : tyrosine

YY1 : yin and yang 1

Remerciements

I would like to start to thank Dr Nicole Francis, my supervisor and mentor. You took a risk almost 6 years ago when you hired me as a passionate student but with whom the communication was difficult. You taught me to be rigorous in every experiment starting from its design to the interpretation of the results and to never overinterpret data. You showed me what are the requirements to be a good scientist, to know deeply your field, but to never forget to look at what is happening in others. I am especially thankful to you to let me have my own ideas and my own view of my projects, to let me design and imagine new experiments, to let me be independent but here for me when I was needing advice. You showed me that science should be a teamwork and help me be more confident in my own ideas and my work. I am grateful to have done my Ph. D. under your supervision, and despite all the difficult moments I am sure that I will only keep good memories. Thank you for pushing me to become the researcher I am today.

Je remercie les membres de mes comités de thèse : Dr Sylvie Mader, Dr Éric Lécuyer et Dr Marc Drolet pour leurs conseils tout au long de mon doctorat. Je remercie Dr Jean-François Côté pour sa porte toujours ouverte pour moi, sa gentillesse, son soutien et ses conseils.

Je remercie également tous les gens qui sont passés dans le laboratoire du Dr Francis et qui n'en ont peut-être pas eu conscience mais dont j'ai appris à leur contact. Je remercie à Nazar qui m'a montré ce qu'être un postdoc voulait dire, à Osvaldo qui m'a permis d'apprendre à superviser des étudiants, sans oublier Hannah pour sa gentillesse et Djamouna pour son soutien et son amitié lors de la fin difficile de mon doctorat.

Je remercie tous les membres de l'IRCM présents et passés, que ce soit pour leur assistance au travail que pour les moments de détente lors des 5 à 7 et en dehors.

Merci à Justine, Laura et PE, la vie et l'éloignement géographique mettent à mal de nombreuses amitiés mais pas la nôtre. Merci pour tous ces moments de détente et de joie qui ont contribué à rendre supportable le stress et la pression quotidienne.

Je remercie ma famille et spécialement mes parents. Merci d'avoir toujours été là pour moi, d'avoir répondu à mes appels quel que soit l'heure chez vous et d'avoir toujours cherché à me remonter le moral,

me rassurer et me remotiver dans tous les moments difficiles que j'ai eu à traverser. Merci de votre soutien indéfectible et d'avoir cru en moi. Merci à vous d'avoir éveillé ma curiosité et mon esprit critique quand j'étais enfant, de m'avoir donné l'envie de comprendre comment marche les choses et de ne pas me contacter des informations qui étaient données à l'école.

Pour finir, je remercie mon conjoint, Thomas. Tu as été là pour moi dès le début, dans les moments difficiles où je faisais la boule, comme dans tous les moments de joie. Je ne sais pas comment j'aurais pu réussir si tu n'avais pas été là. Je te remercie pour ces 5 années avec toi qui ne sont que le début de notre histoire.

1 Introduction

1.1 La chromatine

1.1.1 La chromatine

Dans les cellules l'ADN s'enroule autour d'un tétramère d'histone H3-H4 et de deux dimères d'histones H2A-H2B formant ainsi un nucléosome. La formation de nucléosomes permet la compaction de l'ADN et définit ainsi la base de la chromatine. La chromatine peut présenter différents niveaux de compaction déterminant ainsi l'euchromatine et l'hétérochromatine (Fedorova and Zink, 2008; Hildebrand and Dekker, 2020). L'hétérochromatine est plus compacte et tend à être transcriptionnellement inactive car plus difficilement accessible à la machinerie de transcription. À l'inverse, l'euchromatine est plus ouverte et activement transcrite (Fedorova and Zink, 2008; Hildebrand and Dekker, 2020).

1.1.2 Les modifications post-traductionnelles des histones

Le cœur d'un nucléosome est difficilement accessible, mais les queues des histones qui dépassent de celui-ci sont la cible de nombreuses protéines qui les modifient de manière post-traductionnelle en y attachant de manière covalente des groupements méthyles, acétyles ou ubiquitines. Ces modifications post-traductionnelles peuvent être reconnues par des protéines leur indiquant notamment comment l'information génétique doit être utilisée (Zhang et al., 2015). Ces modifications post-traductionnelles peuvent également être effacées afin de modifier l'expression d'un gène (Biswas and Rao, 2018).

1.1.2.1 Les marques actives

Les gènes activement transcrits se caractérisent par un niveau élevé de tri-méthylation de la lysine 4 de l'histone 3 au niveau de leur promoteur, ainsi que de tri-méthylation des lysines 36 et 79 de l'histone 3 le long des gènes (Mikkelsen et al., 2007). Ce pattern d'expression n'est pas seulement présent au niveau des gènes codant pour des protéines mais également au niveau des ARN non codants (Mikkelsen et al., 2007). L'ubiquitination de H2B par les protéines MSL1 et MSL2 est importante pour stimuler la méthylation des lysines 4 et 79 de l'histone 3 par les protéines du groupe Trithorax afin d'établir ces domaines actifs (Kingston and Tamkun, 2014; Schuettengruber et al., 2017; Wu et al., 2011). Les lysines des histones 3 et 4 des promoteurs de gènes activement transcrits sont également acétylées, ce qui faciliterait

l'assemblage du complexe d'initiation et le désassemblage des histones en neutralisant leurs charges positives.

Les amplificateurs (ou *enhancers*) sont des éléments permettant la stimulation de l'expression génique. Ils se démarquent par la mono-méthylation de la lysine 4 de l'histone 3 lorsqu'ils sont en pause. Lors de leur activation, en plus de cette mono-méthylation, ils acquièrent également un fort niveau d'acétylation de la lysine 27 de l'histone 3 (Creyghton et al., 2010).

Les mono- et di-méthylations de la lysine 36 de l'histone 3 sont présentes en dehors des régions activement transcrites, tout comme l'est la di-méthylation de la lysine 27 de l'histone 3 qui est une marque répressive. Ces marques respectivement actives et répressives en dehors de leur domaines auraient pour but d'éviter la propagation de marques antagonistes ainsi que d'empêcher une mauvaise transcription et également de prévenir l'activation d'éléments amplificateurs (Ferrari et al., 2014; Huang and Zhu, 2018).

1.1.2.2 Les marques répressives

Les marques répressives H3K27me3 et H2AK118Ub déposées respectivement par les complexes Polycomb PRC2 et PRC1 participent à leur propre recrutement ainsi qu'à leur recrutement mutuel afin de promouvoir un état répressif (voir la section 1.2).

La tri-méthylation des lysines 9 et 20 des histones 3 et 4 respectivement participent à l'établissement d'hétérochromatine constitutive, notamment dans les régions centromérique. L'hétérochromatine constitutive, contrairement à l'hétérochromatine facultative qui est définie par les protéines Polycomb, ne change pas ou peu de statut transcriptionnel au cours du développement (Rice et al., 2003; Schotta et al., 2004; Wang et al., 2016b). L'hétérochromatine constitutive liée par HP1 est plus compacte que l'hétérochromatine facultative, elle est donc moins accessible et plus réfractaire au remodelage (Kumar and Kono, 2020).

1.1.2.3 Les domaines bivalents

La présence de marques actives telles que H3K36me3 ou H3K4me3 antagonisent l'action de PRC2 et donc l'existence de marques à la fois actives et répressives dans un même domaine (Schmitges et al., 2011). Toutefois des domaines bivalents existent. Les promoteurs bivalents sont marqués par des marques à la fois actives et répressives, présentent un faible niveau de transcription et grâce à la présence de marques activatrices peuvent être induits rapidement (Mikkelsen et al., 2007)

1.1.3 La propagation des modifications post-traductionnelles d'histone à travers le cycle cellulaire

Ces marques épigénétiques permettent la propagation de l'information à travers le cycle cellulaire. Lors de la mitose, les protéines se liant à la chromatine sont en grande partie enlevées afin de permettre la compaction des chromosomes mitotiques, mais les histones restent liées à l'ADN permettant donc à la fin de la mitose de recruter à nouveau ces mêmes facteurs afin d'exprimer de manière appropriée les gènes (Bellec et al., 2018). Toutefois lors de la mitose, le niveau d'acétylation des histones diminue afin de réduire la transcription et de permettre la compaction des chromosomes (Liu et al., 2017).

La réplication de l'ADN est un évènement plus disruptif car le passage de l'ADN polymérase nécessite le désassemblage des nucléosomes (Alabert et al., 2017). Des chaperonnes d'histones permettent leur réassemblage derrière la fourche de réplication et donc à proximité de leur site d'origine. Néanmoins les histones sont ségréguées entre les deux chromatides, le niveau de modifications post-traductionnelles est donc divisé par deux à chaque fois que l'ADN est répliqué et doit être restauré afin de ne pas perdre l'information épigénétique après plusieurs divisions cellulaires.

Pour finir, de manière similaire à la réplication, la transcription force le désassemblage des nucléosomes afin de permettre le passage de l'ARN polymérase. Les chaperonnes d'histones jouent un rôle essentiel dans le repositionnement des nucléosomes qui doivent revenir à proximité de leur site d'origine car ils définissent non seulement de larges régions transcriptionnellement actives mais aussi des promoteurs qui sont des régions bien plus restreintes (Venkatesh and Workman, 2015).

1.1.4 La structure de la chromatine

La chromatine n'est pas organisée de manière linéaire dans les cellules mais forment des domaines associés topologiquement (TAD) au sein desquels de multiples interactions entre des régions génomiques distantes peuvent se produire (Sikorska and Sexton, 2020). Les bordures de ces TAD sont déterminées notamment par les protéines isolatrices CTCF qui lient la chromatine de manière séquence spécifique. La chromatine au sein d'un même TAD présente le même type de marques épigénétiques (actives ou répressives) et porte donc le même type de modifications post-traductionnelles d'histones (Rao et al., 2014). Ces TAD ne sont pas fixes, lors du développement leurs bordures sont amenées à bouger à la suite de l'activation ou à la répression d'un nouveau domaine comme cela a été observé pour les gènes Hox (Noordermeer et al., 2014).

Lors de la différenciation de cellules ou plus généralement lors du développement, les cellules doivent changer leur pattern d'expression génique donc certains gènes doivent commencer à s'exprimer alors que d'autres doivent être réprimés. Cette répression et le maintien de celle-ci durant le développement se fait notamment par le biais des protéines du groupe Polycomb. Mais les signaux indiquant qu'un gène doit commencer à s'exprimer ou à être réprimé, et donc changer de modifications post-traductionnelles d'histones afin de changer de statut transcriptionnel ne sont toujours pas bien compris.

1.2 Les protéines du groupe Polycomb

1.2.1 Les protéines du groupe Polycomb : découverte et introduction

1.2.1.1 Les gènes Hox et leur répression par les protéines du groupe Polycomb durant le développement

Durant le développement embryonnaire chez *Drosophila melanogaster* (*D. melanogaster*) des gènes sont activés et exprimés séquentiellement afin de permettre la mise en place de l'axe antéro-postérieur (Beck et al., 2010; Gilbert, 2000). Ceci commence par un gradient d'expression de *bicoid* et *nanos* qui s'expriment chacun à une extrémité de l'embryon, se poursuit par l'expression des gènes *gap*, puis *pair rules* et enfin des gènes homéotiques (ou *Hox*) qui permettent de définir chaque segment (Gilbert, 2000). Contrairement aux gènes *gap* ou *pair rules*, qui ont un pattern d'expression en gradient ou en bandes, les gènes *Hox* s'expriment en suivant une colinéarité spatio-temporelle (Beck et al., 2010). Les gènes *Hox* forment deux clusters *bithorax* et *antennapedia* dans le génome de *D. melanogaster*, avec les gènes en 3' s'exprimant dans la partie la plus antérieurs et le plus précocement, et les gènes les en 5' s'exprimant plus tardivement et dans la partie postérieure de l'embryons. Ce pattern d'expression des gènes Hox est essentiel pour le développement de l'adulte qui en cas de mauvaise expression donne lieu à des transformations homéotiques : à savoir le remplacement d'un segment par un autre, comme c'est le cas lors de la mutation du gène *bithorax*, qui conduit à la formation de deux thorax au lieu d'un abdomen et d'un thorax ou de la mutation de *antennapedia*, qui se traduit par le développement de pattes à la place des antennes (Gilbert, 2000). Les gènes décrivent précédemment et s'exprimant avant les gènes Hox sont essentiels pour la bonne expression de ces derniers mais leur expression est limitée dans le temps. Les protéines du groupe Polycomb (PcG) sont quant à elles nécessaires afin de maintenir ce pattern d'expression des gènes Hox (Gilbert, 2000). Elles vont pouvoir réprimer ces gènes de manière stable dans le temps. Ce n'est donc pas

surprenant que la mutation des gènes codant pour les protéines du groupe Polycomb conduisent aux mêmes phénotypes de transformation homéotiques que la mutation des gènes Hox puisqu'ils sont leurs répresseurs (Kassis et al., 2017).

1.2.1.2 Introduction sur les protéines du groupe Polycomb

Les protéines du groupe Polycomb s'assemblent en deux complexes principaux : PRC1 et PRC2, qui ensemble participent au maintien de la répression génique à travers le développement et donc lors de nombreuses divisions cellulaires (voir les sections 1.2.3 et 1.2.2 pour plus de détails). PRC1 et PRC2 modifient de manière post-traductionnelle les histones (PRC2 tri-méthyle la lysine 27 de l'histone 3 et PRC1 ubiquitinye la lysine 118 de l'histone H2A), de plus PRC1 compacte la chromatine empêchant ainsi sa transcription (Kuroda et al., 2020). Ces modifications post-traductionnelles d'histones déposées par PRC1 et PRC2 sont présentes au niveau de domaines réprimés au sein desquels se trouvent les sites liés par ces complexes. Chez *D. melanogaster*, PRC1 et PRC2 sont recrutés au niveau de séquences d'ADN appelées éléments de réponses Polycomb (PREs) (voir la section 1.3). Ces PREs sont nécessaires pour le recrutement des complexes Polycomb, l'établissement d'un domaine répressif et pour le maintien de la répression par les protéines du groupe Polycomb pendant de nombreuses divisions cellulaires (Coleman and Struhl, 2017; Laprell et al., 2017).

Les connaissances actuelles des activités des complexes Polycomb, de leurs sites de liaison ainsi que de leur capacité à maintenir la répression génique à travers le cycle cellulaire laissent toutefois trois grandes questions en suspens. Comment les protéines du groupe Polycomb sont recrutées aux PREs ? Une description détaillée des PREs ainsi que des connaissances actuelles du recrutement des protéines du groupe Polycomb fait l'objet de la section 1.3. Quelles sont les activités des complexes Polycomb et comment contribuent-elles à la répression génique ? Une description approfondie des activités biochimiques des protéines du groupe Polycomb font l'objet de la section 1.2. Et pour finir comment les protéines du groupe Polycomb maintiennent la répression à travers le cycle cellulaire ? Ce dernier sujet ne faisant pas l'objet de ma thèse, il ne sera abordé que brièvement.

1.2.2 PRC2

1.2.2.1 PRC2 est une méthyltransférase

Chez *D. melanogaster*, PRC2 se compose de quatre sous-unités E(Z), Su(Z)12, Esc et Nurf-55 (Figure 1-1a et Tableau 1-1) (Kassis et al., 2017; Kuroda et al., 2020; Müller et al., 2002). PRC2 est bien connue pour son activité de méthyltransférase qui repose sur le domaine SET de E(Z) et conduit à la tri-méthylation de la lysine 27 de l'histone 3 (Figure 1-1b) (Cao et al., 2002; Müller et al., 2002). La tri-méthylation de la lysine 27 de l'histone 3 ne se fait pas en une seule fois, mais de manière séquentielle. La mono-, di-, et tri-méthylation de ces lysines sont présentes dans des régions bien définies du génome et ont des fonctions différentes. H3K27me1 s'accumule au niveau des gènes activement transcrits marqués par la présence de H3K36me3 alors que H3K27me3 est au niveau de régions réprimées. Les domaines réprimés présentant la marque épigénétique H3K27me3 ne contiennent pas l'acétylation de cette même lysine, et la perte de PRC2 et donc la diminution de ce H3K27me3 conduit à l'augmentation de H3K27Ac montrant bien le caractère mutuellement exclusif de ces deux marques (Tie et al., 2009). La di-méthylation de la lysine 27 de l'histone 3 est présente dans les régions intergéniques et non activement transcrites (ou avec un faible niveau de transcription) empêchant ainsi l'acétylation de cette même lysine et donc l'activation de la transcription (Ferrari et al., 2014; Lee et al., 2015). H3K27me2 et H3K27me3 sont dans des régions distinctes, H3K27me2 présente une forte corrélation avec la marque active H3K36me2 préviendrait ainsi de l'activation de la transcription mais aussi de la répression en gardant ces régions dans un état intermédiaire ou bivalent leur permettant de changer rapidement de statut transcriptionnel (Streubel et al., 2018). La méthylation de la lysine 27 de l'histone 3 en elle-même est importante puisque le remplacement de la lysine 27 par une arginine conduit à un phénotype similaire à celui observé lors de la déplétion de PRC2 (Pengelly et al., 2013). La méthylation de la lysine 27 peut être enlevée par une déméthylase, UTX ou JMJD3, permettant alors son acétylation par CBP pouvant conduire à la transcription de ce domaine qui était alors réprimé par les protéines du groupe Polycomb (Agger et al., 2007; Tie et al., 2009). L'acétylation de la lysine 27 peut également être enlevé par RPD3 notamment, donnant la possibilité à un locus de changer de statut transcriptionnel (Tie et al., 2009).

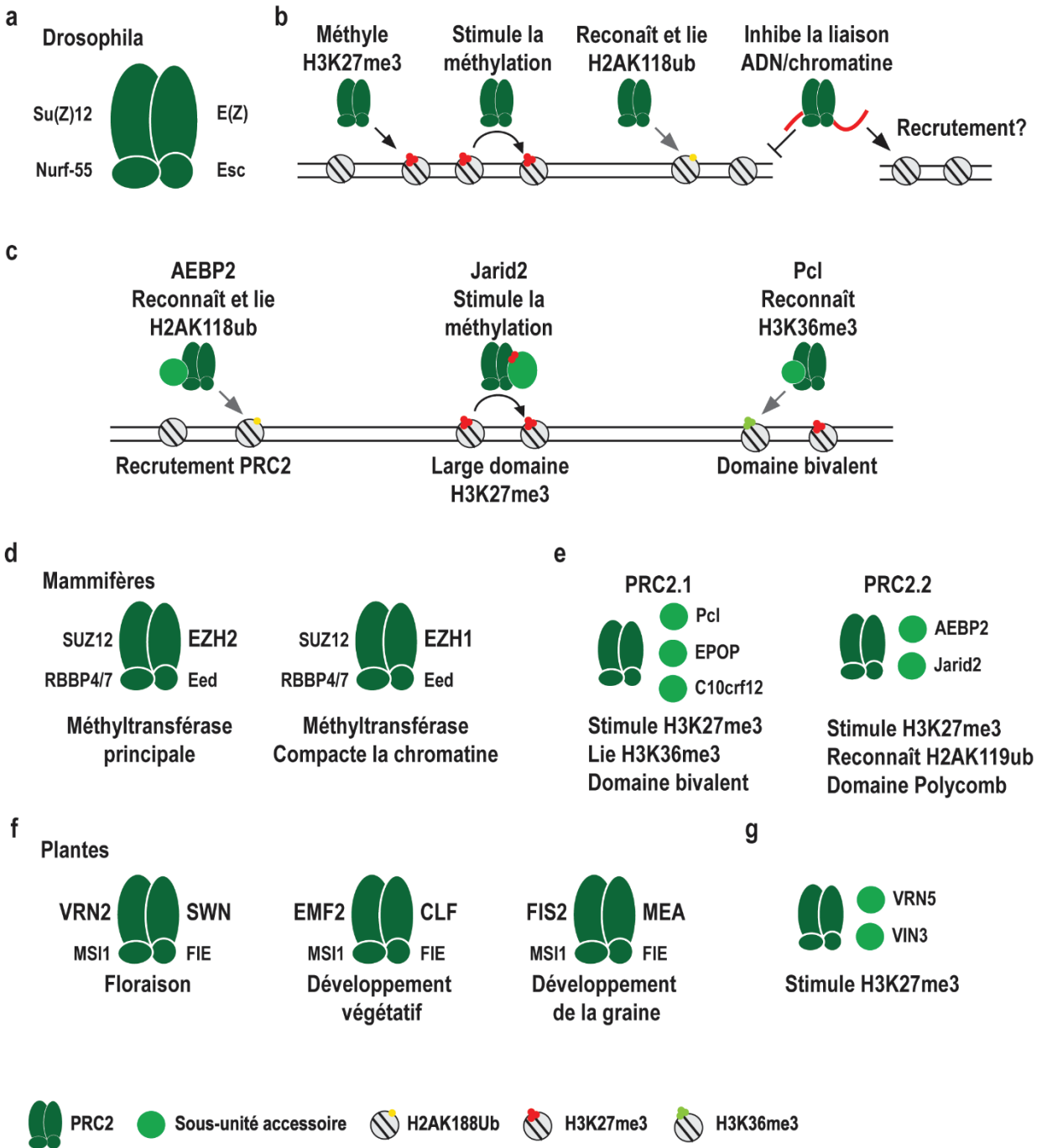


Figure 1-1 Le complexe PRC2 : composition et fonction

a. Composition du complexe PRC2 chez *Drosophila melanogaster*. b. Schéma des activités de PRC2. PRC2 tri-méthyle la lysine 27 de l'histone 3. H3K27me3 stimule l'activité méthyltransférase de PRC2. PRC2 reconnaît et lie H2AK118ub. La liaison de l'ARN par PRC2 inhibe sa liaison à l'ADN et la chromatine, mais joue également un rôle potentiellement dans son recrutement. c. Les fonctions des différentes sous-unités accessoires de PRC2. AEBP2 permet la reconnaissance et liaison de H2AK118ub. Jarid2 stimule l'activité méthyltransférase de PRC2. Pcl permet la reconnaissance de H3K36me3 par le complexe PRC2. d, f.

Composition du complexe PRC2 chez les mammifères (d) et les plantes (f). e, g. Liste des sous-unités accessoires du complexe PRC2 et de leur fonction chez les mammifères (e) et les plantes (g).

1.2.2.2 PRC2 est conservé chez d'autres espèces

1.2.2.2.1 Chez les mammifères

Chez les mammifères, PRC2 a les mêmes fonctions de méthyltransférase et de répresseur de l'expression génique que chez *D. melanogaster* (Kouznetsova et al., 2019; Yu et al., 2019). Chaque sous-unité de PRC2 possède un homologue chez les mammifères (Figure 1-1d et Tableau 1-1). Il y a deux versions de la sous-unité responsable de l'activité méthyltransférase de PRC2 : EZH1 et EZH2. EZH2 est la méthyltransférase principale et est exprimée dans les cellules en division (Margueron et al., 2008). EZH1 est exprimée dans toutes les cellules mais a une capacité à méthyler les histones plus réduite que EZH2 (Margueron et al., 2008). Le complexe PRC2 contenant EZH2 est essentiel pour l'établissement de la tri-méthylation de la lysine 27 de l'histone 3, alors que PRC2 contenant EZH1 peut compacter la chromatine et la présence de la queue des histones est essentielle pour cette dernière activité (Margueron et al., 2008). Néanmoins in vivo, EZH1 peut partiellement compenser la perte de EZH2 en maintenant le niveau de tri-méthylation à quelques gènes, même si globalement le niveau de tri- et di-méthylation est diminué (Shen et al., 2008). EZH1 suffit alors pour maintenir la répression génique dans des cellules en division mais pas lors de la différenciation (Shen et al., 2008). Cela a conduit à l'idée que PRC2-EZH2 est la méthyltransférase principale alors que PRC2-EZH1 aurait plus un rôle de maintenance d'où sa présence dans les cellules qui ne se divisent pas et donc qui n'ont besoin que de compenser la perte de méthylation due à la transcription, au renouvellement des histones et à l'action des déméthylases.

1.2.2.2.2 Chez les plantes

Chez les plantes telles que *Arabidopsis thaliana*, chaque sous-unité du complexe PRC2 possède au moins un homologue : Esc (FIE) , Nurf-55 (MSI1), Su(Z)12 (VRN2, EMF2 et FIS2) et E(Z) (SWN, CLF et MEA), qui comme chez *D. melanogaster* sont responsables de la tri-méthylation des lysines 27 de l'histone 3 et participent au maintien de la répression génique (Figure 1-1f et Tableau 1-1) (Schubert et al., 2006). La présence de plusieurs homologues pour Su(Z)12 et E(Z) permet d'assembler plusieurs versions du complexe PRC2 qui interviennent dans la régulation de différents événements (Hennig and Derkacheva, 2009; Kim and Sung, 2014; Mozgova et al., 2015).

La répression de FLC, lors de la vernalisation dépend d'un complexe PRC2 qui contient VRN2 et SWN en plus des sous-unités constantes FIE et MSI1 (Lucia et al., 2008; Wood et al., 2006). Des sous-unités accessoires telles que VRN5 et VIN3 sont également associées à ce complexe PRC2 afin de stimuler son

activité lors d'une exposition prolongée au froid, d'induire la répression du gène FLC et permettre la floraison (Lucia et al., 2008). Une autre version du complexe PRC2 contenant cette fois-ci EMF2 et CLF est essentiel lors du développement de *Arabidopsis thaliana* car en cas de mutation les plantes sont petites et avec peu de feuilles et de longues racines (Chanvivattana et al., 2004; Gu et al., 2014). EMF2 agit en réprimant l'expression de gènes impliqués dans le développement floral afin de permettre le développement végétatif de la plante avant que ces premiers ne soient activés (Moon et al., 2003). La perte de CLF peut néanmoins être partiellement compensée par l'un des deux autres homologues de E(Z), contrairement à EMF2 qui donne un phénotype plus sévère en cas de mutation (Chanvivattana et al., 2004). Pour finir, la version du complexe PRC2 contenant FIS2 et MEA est impliquée dans le maintien de la répression de MEA et FIS2 provenant du gamète paternel, lors du développement embryonnaire, alors que ces mêmes gènes sont exprimés sur l'allèle maternel (Garnier et al., 2008; Jullien et al., 2006).

| | <i>Drosophila</i> | Mammifères | Plantes |
|------------------------|---------------------|--|----------------------|
| PRC2 | E(Z) | EZH1 EZH2 | SWN CLF MEA |
| | Su(Z)12 | SUZ12 | VRN2 EMF2 FIS2 |
| | Esc | Eed | FIE |
| | Nurf-55 | RBBP4/7 | MSI1 |
| Protéine accessoire | AEBP2 (jing) | AEBP2 | |
| | Pcl | Pcl1 (PHF1) Pcl2 (MTF2) Pcl3 (PHF19) | |
| | Jarid2 (Jumonji) | Jarid2 | |
| | | EPOP | |
| | | C10crf12 | |
| | | | VRN5 |
| | | | VIN3 |

Tableau 1-1 Liste des sous-unités du corps du complexe PRC2 chez *Drosophila melanogaster*, les mammifères et les plantes ainsi que leurs sous-unités accessoires

1.2.2.3 La reconnaissance des modifications post-traductionnelles des histones par PRC2 module son activité

Concernant la reconnaissance des modifications post-traductionnelles des histones par PRC2 et la modulation de son activité, la majorité des études ont été faites sur le complexe PRC2 issu des mammifères, mais les résultats disponibles chez *D. melanogaster* montrent une certaine conservation (O'Meara and Simon, 2012; Yu et al., 2019). Toutefois il ne faut exclure la possibilité que certains de ces mécanismes puissent être spécifiques au système mammifère.

La marque répressive H3K27me3 déposée par PRC2 peut couvrir des régions larges de plusieurs kilobases qui sont liées à la fois par PRC2 et PRC1 sous forme de pics discrets au sein de ces larges domaines (Schwartz et al., 2006). PRC2 peut reconnaître la modification post-traductionnelle d'histone qu'il dispose lui-même afin de l'aider à se propager le long de la chromatine et de former ces larges régions tri-méthylées (Margueron et al., 2009; O'Meara and Simon, 2012; Schwartz et al., 2006). Cette reconnaissance de la tri-méthylation de la lysine 27 de l'histone se fait par Esc qui lie préférentiellement les nucléosomes présentant cette modification, conduisant à un changement de conformation de PRC2 et à la stimulation de l'activité méthyltransférase de E(Z) (Figure 1-1b) (Jiao and Liu, 2015; Margueron et al., 2009). Cette capacité à reconnaître H3K27me3 est essentielle *in vivo* car en cas de mutation, PRC2 n'est plus recruté aussi efficacement à ses cibles qui présentent alors un plus faible niveau de tri-méthylation (Margueron et al., 2009). Plus globalement, une mutation de Esc l'empêchant de reconnaître H3K27me3 conduit à une réduction globale du niveau de tri-méthylation mais aussi de di-méthylation chez *D. melanogaster* (Margueron et al., 2009).

Chez *D. melanogaster*, Esc est exprimé essentiellement durant le stade embryonnaire alors que la répression par les protéines du groupe Polycomb se fait tout du long du développement (Bornemann et al., 1996; Sathé and Harte, 1995). Escl, protéine homologue de Esc, s'exprime de manière plus importante en fin de stade embryonnaire et après celui-ci. Escl peut former avec un complexe fonctionnel avec les autres sous-unité de PRC2, se substituant à Esc afin de maintenir le niveau de tri-méthylation de H3K27 et de répression génique approprié tout au long du développement (Ohno et al., 2008; Wang et al., 2006).

PRC2 grâce à sa sous-unité Nurf-55 interagit avec une autre partie de la queue des histones H3, et notamment avec la lysine 4 (Schmitges et al., 2011). La présence de la tri-méthylation de la lysine 4 de l'histone 3 réduit grandement la capacité de Nurf-55 d'interagir avec celle-ci (Schmitges et al., 2011). Cette diminution d'affinité pour les histones présentant la marque active H3K4me3 ou encore H3K36me3, se traduit par une diminution de la capacité de PRC2 à di- et à tri-méthyle la lysine 27 des nucléosomes

portant déjà cette modification, alors que la mono-méthylation n'est que peu affectée (Schmitges et al., 2011). Ceci montre le caractère antagoniste des modifications post-traductionnelles des queues d'histones qui ne peuvent pas porter à la fois des marques actives et répressives au sein du même nucléosome (Schmitges et al., 2011).

1.2.2.4 Les sous-unités accessoires de PRC2 contribuent à son activité *in vivo*

1.2.2.4.1 Chez D. melanogaster

PRC2 possède des sous-unités accessoires qui ne sont pas présentes en quantité stœchiométriques par rapport aux protéines qui forment le cœur du complexe. Polycomblike (Pcl) forme un complexe stable avec PRC2, et la déplétion de Pcl dans les embryons de *D. melanogaster* conduit à une réduction du niveau de H3K27me3 aux cibles de PRC2 ainsi qu'à une dérégulation de certains gènes Hox, montrant bien que Pcl participe à la mise en place de la répression génique par les protéines du groupe Polycomb (Nekrasov et al., 2007). Pcl reconnaît la tri-méthylation de la lysine 36 de l'histone 3 (Gatchalian et al., 2015). La reconnaissance de cette marque active par le complexe PRC2-Pcl permet de recruter ce dernier au niveau de domaines bivalents ou activement transcrits afin d'induire la tri-méthylation de la lysine 27 de l'histone 3 et de réprimer ces derniers (Figure 1-1c) (Cai et al., 2013).

Jarid2, aussi appelé Jumonji, peut également interagir avec PRC2 et lier les mêmes sites *in vivo*. PRC2 méthyle Jarid2, ce qui en retour stimule l'activité méthyltransférase de PRC2 (Figure 1-1c) (Sanulli et al., 2015). Cette sous-unité accessoire aurait donc un rôle important lors du développement pour la formation de nouveaux domaines répressifs qui doivent acquérir la tri-méthylation de la lysine 27 de l'histone 3 mais pas pour le maintien de ceux-ci lorsqu'ils sont déjà formés (Herz et al., 2012; Sanulli et al., 2015; Sasai et al., 2007).

Jing (AEBP2 chez les vertébrés), Jarid2 et PRC2 peuvent former un complexe qui reconnaît la mono-ubiquitination de la lysine 118 de l'histone H2A déposée par PRC1. Ceci conduit à la stimulation de l'activité méthyltransférase de PRC2 qui dépend à la fois de AEBP2 et de l'ubiquitination de l'histone H2A (Figure 1-1c) (Ciferri et al., 2012; Kalb et al., 2014). AEBP2 possède un domaine de liaison à l'ADN qui est séquence spécifique et qui pourrait participer au recrutement initial de PRC2 à ses cibles (Kim et al., 2009).

Toutes les sous-unités accessoires de PRC2 participent donc à la régulation de l'activité de ce complexe Polycomb, en l'aidant à reconnaître d'autres modifications post-traductionnelles d'histones afin de stimuler son recrutement à de nouveaux sites et de l'aider à propager la tri-méthylation de la lysine 27 de l'histone 3 afin de former de larges domaines répressifs. Mais ces protéines accessoires semblent jouer un

rôle mineur dans le maintien de la répression une fois qu'elle a été établie par les protéines du groupe Polycomb.

1.2.2.4.2 Chez les mammifères

Chez les mammifères, EZH2 est apparue comme la méthyltransférase principale, un intérêt particulier a été porté au complexe PRC2 contenant cette sous-unité (Margueron et al., 2008). Tout comme chez *D. melanogaster*, PRC2 possède plusieurs sous-unités accessoires dont certaines sont mutuellement exclusives. PRC2-EZH2 forme deux complexes en fonction des protéines accessoires qui lui sont associés : PRC2.1 et PRC2.2 (Figure 1-1e et Tableau 1-1) (Chammas et al., 2020; van Mierlo et al., 2019).

PRC2.1 comporte les protéines accessoires Pcl et EPOP. Il y a trois homologues de Pcl chez les mammifères : PHF1, PHF19 et MTF2. Les trois homologues de Pcl lient préférentiellement un fragment d'ADN nu contenant les répétitions CpG présentent au niveau des îlots CpG et peuvent lier H3K36me3 grâce à leur domaine TUDOR (Ballaré et al., 2012; Li et al., 2017a). Même si Pcl1, Pcl2 et Pcl3 présentent les mêmes préférences pour la liaison de l'ADN et des nucléosomes, elles n'ont pas pour autant les mêmes fonctions dans les cellules. Pcl1 (PHF1) stimule l'activité méthyltransférase de PRC2 *in vitro*. La déplétion de Pcl1 conduit une diminution du recrutement de PRC2 à certaines de ses cibles, à une diminution de H3K27me3 et à leur dérégulation (Cao et al., 2008; Sarma et al., 2008). Le domaine TUDOR de Pcl1 est également responsable de son recrutement au niveau des sites de cassures dans l'ADN, PRC2 est également recruté aux cassures dans l'ADN mais sa fonction ainsi que le rôle de Pcl1 dans ce recrutement reste inconnu (Campbell et al., 2013; Musselman et al., 2012). Pcl2 (MTF2) est impliqué dans le recrutement de PRC2 au niveau de gènes ayant un rôle dans le développement et la différenciation comme le montre leur dérégulation lors de la perte de cette sous-unité accessoire (Walker et al., 2010). Pcl2 est particulièrement présent au niveau de promoteurs bivalents, expliquant pourquoi cette sous-unité accessoire est importante lors de la différenciation lorsque des gènes doivent changer rapidement de statut transcriptionnel (Walker et al., 2010). Pcl3 (PHF19) lie H3K36me3 permettant le recrutement de PRC2 et de N066 au niveau de gènes activement transcrits ou de domaines bivalents lors de la différenciation, afin de déméthyliser la lysine 36 de l'histone 3, de permettre la méthylation de la lysine 27 et donc induire leur répression (Ballaré et al., 2012; Brien et al., 2012).

In vitro, EPOP stimule l'activité méthyltransférase de PRC2, mais *in vivo* elle permet le recrutement d'un facteur d'élongation en plus de PRC2, au niveau de promoteurs bivalents afin de maintenir la transcription à ces gènes (Beringer et al., 2016; Zhang et al., 2011). La déplétion de EPOP conduit au recrutement de PRC2 avec une autre sous-unité accessoire, Jarid2, ce qui a pour conséquence d'augmenter le niveau de

tri-méthylation de la lysine 27 de l'histone 3 et de diminuer l'expression de ces gènes qui étaient lié par EPOP et PRC2 (Beringer et al., 2016).

C10orf12 peut également former un complexe avec PRC2. Le recrutement de C10orf12 à un site donné permet d'induire le recrutement de PRC2, stimule la tri-méthylation de la lysine 27 de l'histone 3 par PRC2 et donc la répression (Hauri et al., 2016; Shi et al., 2019).

En résumé le complexe PRC2.1 est essentiellement impliqué dans la liaison et régulation des domaines bivalents lors de la différenciation. Les protéines accessoires qui le constituent sont différenciellement exprimées lors de la différenciation de cellules souches, et elles sont essentielles lors de la différenciation afin de permettre le changement rapide du statut transcriptionnel des cibles de PRC2 (Kloet et al., 2016).

PRC2.2 comporte Jarid2 et AEBP2 qui ont des rôles similaires à leur homologue respectif chez *D. melanogaster*. Jarid2 lie les nucléosomes, reconnaît l'ubiquitination de la lysine 119 déposée par PRC1 participant ainsi au recrutement de PRC2, et stimule l'activité méthyltransférase de ce dernier (Healy et al., 2019; Son et al., 2013). AEBP2 lie un motif spécifique dans l'ADN ce qui avait conduit à ce qu'il soit proposé comme facteur permettant un recrutement de PRC2 qui soit séquence spécifique (Kim et al., 2009). AEBP2 stimule la liaison de PRC2 aux nucléosomes et son activité méthyltransférase *in vitro*, sans que pour autant sa présence soit nécessaire ni pour le recrutement ni pour l'activité méthyltransférase de PRC2 *in vivo* (Grijzenhout et al., 2016; Lee et al., 2018a).

Alors que des cellules qui ne sont pas en différenciation ne sont que peu affectées par la déplétion des sous-unités accessoires de PRC2.1 ou PRC2.2, la perte de Pcl1, Pcl2, Pcl3 et Jarid2 conduit à un mauvais recrutement de PRC2 à ses cibles accompagné d'une réduction de niveau de tri-méthylation et d'une augmentation de la di-méthylation de la lysine 27 de l'histone 3 (Healy et al., 2019). PRC2.1 et PRC2.2 ont des rôles redondants dans le maintien de la répression par PRC2 (Healy et al., 2019). Jarid2 et les protéines Pcl permettent le recrutement de PRC2 dépendamment ou indépendamment de PRC1 et peuvent être essentiels lors du développement afin d'établir de nouveaux domaines Polycomb (van Mierlo et al., 2019).

1.2.2.4.3 Chez les plantes

Chez les plantes comme *Arabidopsis thaliana*, le complexe PRC2 impliqué dans la vernalisation peut s'associer avec les sous-unités accessoires VRN5 et VIN3 afin de stimuler l'activité méthyltransférase de PRC2 et de participer à la répression du gène FLC (Figure 1-1g et Tableau 1-1) (Costa and Dean, 2019; Qüesta et al., 2016).

Que ce soit chez les plantes et chez les mammifères, PRC2 a la même fonction de méthyltransférase impliquée dans la répression génique que chez *D. melanogaster*. Mais la présence de multiples versions d'une même sous-unité du cœur du complexe PRC2 chez *Arabidopsis thaliana*, ainsi que la présence de nombreuses sous-unités accessoires chez les mammifères, qui ne s'expriment pas au même moment et modulent l'activité de PRC2 démontrent d'une plus grande complexité que chez *D. melanogaster*.

1.2.2.5 La structure de PRC2

La structure de PRC2 a été essentiellement étudiée chez les mammifères, mais les informations disponibles chez *D. melanogaster* montrent que cette structure est conservée. De manière synthétique, PRC2 se compose de quatre domaines : un domaine catalytique composé de E(Z), un domaine d'encrage avec Su(Z)12 qui permet l'interaction des protéines accessoires avec PRC2, et un domaine régulateur composé de Esc qui va pouvoir lier H3K27me3, entraînant un changement de conformation de E(Z) stimulant ainsi son activité méthyltransférase (Chammas et al., 2020). Pour finir un dernier domaine médian fait le lien entre les E(Z) et Su(Z)12 et participe à la stabilisation du site actif (Chammas et al., 2020).

PRC2 peut former un dimère grâce à sa sous-unité Su(Z)12 qui est stabilisée par la protéine accessoire Pcl, en revanche cette dimérisation n'est pas possible lorsque PRC2 forme un complexe avec AEBP2 (Chen et al., 2020; Davidovich et al., 2014). Cette dimérisation est importante pour la liaison des îlots CpG *in vitro* et la méthylation de quelques cibles de PRC2 testée par ChIP-qPCR *in vivo* (Chen et al., 2020). Chez *D. melanogaster* un complexe excédant 1 MDa et contenant les sous-unités de PRC2 a également été rapporté, mais dans ce cas-là ce n'est pas Su(Z)12 mais Esc qui permet la dimérisation (Tie et al., 2003, 2005). La fonction de la dimérisation de PRC2 dans la liaison de ses cibles chez les mammifères et *D. melanogaster*, ainsi que l'influence que cela a sur son activité à l'échelle du génome entier reste à élucider.

1.2.2.6 La liaison des ARN

PRC2 en plus de lier les histones et l'ADN, peut lier les ARN (Almeida et al., 2020; Yan et al., 2019). Pour cette partie la majorité des études proviennent du complexe PRC2 chez les mammifères, mais on peut assumer que des résultats similaires sont observés pour le complexe de *D. melanogaster*, qui présente une structure et fonction similaires et qui peut lui aussi interagir avec les ARN. Mais des différences peuvent toutefois exister entre le complexe PRC2 issu des mammifères et de *D. melanogaster*.

PRC2 lie des centaines d'ARN qui proviennent essentiellement de régions activement transcrites et arborant des modifications post-traductionnelles d'histones telles que H3K36me ou H3K4me3 (Beltran et al., 2016; Davidovich et al., 2013). Les régions d'où proviennent ces ARN peuvent être liées par PRC2 sans

que pour autant elles soient réprimées par celui-ci (Davidovich et al., 2013). PRC2 lie ces ARN lors de leur transcription, avant leur épissage (Beltran et al., 2016). *In vitro*, les ARN inhibent l'activité méthyltransférase de PRC2 en l'empêchant de lier l'ADN ou la chromatine (Beltran et al., 2016; Wang et al., 2017a). Cette inhibition peut être partiellement levée lorsque l'activité méthyltransférase de PRC2 est stimulée par la présence de peptides présentant la tri-méthylation de la lysine 27 (Zhang et al., 2019a).

Cette liaison globale des ARN a poussé à la formulation de l'hypothèse suivante : la production d'ARN entraîne la liaison de PRC2 qui l'empêche alors d'induire la répression de région activement transcrite en prévenant son recrutement au niveau de la chromatine et donc en inhibant indirectement son activité méthyltransférase. Cette hypothèse est soutenue notamment par des observations faites *in vivo*, où l'accumulation d'ARN dans le noyau conduit à une diminution globale de H3K27me3 ainsi qu'à l'expression de nombreux gènes qui étaient jusqu'alors réprimés (Garland et al., 2019). Mais cette hypothèse repose sur l'idée que PRC2 lie de manière non-spécifique tous les ARN, or PRC2 ne lie pas tous les ARN de la même façon. *In vitro* PRC2 montre une préférence pour les ARN riches en guanines (G) et contenant plusieurs G consécutives (Beltran et al., 2019; Wang et al., 2017b). Toujours *in vitro*, la présence de multiples G au sein d'un ARN, que ceux-ci forment un quadruplex de G ou non, suffit à inhiber sa liaison à l'ADN ; mais *in vivo* PRC2 est déplacé de la chromatine par un ARN naissant contenant de multiples G uniquement quand celui-ci forme un quadruplex de G d'après les conclusions des auteurs (Beltran et al., 2019; Wang et al., 2017b).

PRC2 lie aussi les ARN naissants produits à partir des régions qu'il réprime (Kanhere et al., 2010). Dans ce cas-là, la liaison d'ARN naissant pourrait servir de mécanisme de surveillance pour PRC2, afin de détecter rapidement lorsqu'un gène échappe à la répression. La présence de H3K27me3 et autres marques répressives stimulant de l'activité méthyltransférase de PRC2 permettraient alors de restaurer la répression génique (Zhang et al., 2019a).

PRC2 ne lie pas uniquement des ARN messagers mais également des ARN non codants (Kaneko et al., 2013; Khalil et al., 2009; Zhao et al., 2010). Un intérêt particulier a été porté à ces ARN non codants car via leur interaction avec PRC2, ils permettent le recrutement spécifique de ce dernier en *cis* ou en *trans* chez les vertébrés afin d'induire la répression génique d'un locus donné (voir la section 1.3.6.1.3).

Toutes les sous-unités de PRC2 peuvent lier les ARN, mais la liaison de l'ARN implique essentiellement le domaine régulateur de PRC2 (Zhang et al., 2019a). Ceci explique donc en partie pourquoi la présence de peptides H3K27me3 permet de restaurer l'activité méthyltransférase de PRC2 puisque ces peptides et les ARN lient le même domaine (Zhang et al., 2019a). En plus des protéines formant le noyau de PRC2, toutes les sous-unités accessoires peuvent également lier les ARN (Kaneko et al., 2014; Zhang et al., 2019a). La

liaison des ARN Meg3 ou HOTAIR par Jarid2 stimule son interaction avec PRC2 et l'absence de Meg3 dans les cellules conduit à la perte du recrutement de PRC2 aux sites qui en dépendent et à une augmentation de l'expression de ces gènes (Kaneko et al., 2014). PRC2 peut former un dimère et chaque dimère peut lier un ARN (Davidovich et al., 2014).

Finalement, la liaison des ARN par PRC2 ne donne pas lieu à un modèle simple vu que ces premiers participent et inhibent à la fois le recrutement de PRC2 (Figure 1-1b). Les données de CLIP et de RIP ont montré que PRC2 lie des centaines d'ARN codants et non codants sans spécificité de séquence, mais deux autres études ont montré que cette liaison peut se faire préférentiellement avec des ARN structurés (quadruplex de G). Le rôle des ARN dans le recrutement de PRC2 reste donc encore à étudier. Les ARN pourraient aider au recrutement de PRC2, en interagissant avec l'ADN via la formation de structures d'acides nucléiques non-canoniques comme les R-loops et les triplex, ou en adoptant eux même des structures qui vont entrainer la liaison préférentielle de PRC2, permettant alors un recrutement séquence spécifique de celui-ci. Cette idée qu'un ARN puisse adopter différentes structures et donc réguler différemment l'activité de PRC2 a été abordée dans le cas de l'étude de l'élément de réponse Polycomb du gène *vestigial* chez *D. melanogaster* (voir la section 1.3) (Herzog et al., 2014). Ce PRE est transcrit dans les deux orientations, mais l'ARN produit dans une seule orientation interagit avec PRC2 et prévient de son recrutement au niveau de la chromatine (Herzog et al., 2014). À l'inverse, l'ARN non codant produit par la transcription de ce PRE dans l'autre orientation corrèle avec le recrutement de PRC2 et la répression du gène *vestigial* qu'il régule. Cependant les auteurs n'ont pas déterminé si cet ARN est directement impliqué dans le recrutement du complexe Polycomb (Herzog et al., 2014).

PRC2 est le complexe Polycomb responsable de la tri-méthylation de la lysine 27 de l'histone 3 et est essentielle pour la répression génique. Ce complexe est conservé aussi bien en structure qu'en fonction à travers l'évolution. Depuis ces dix dernières années, la découverte de l'interaction de PRC2 avec des ARN codants et non-codants ont suscité un regain d'intérêt pour ce complexe, afin de comprendre quel est le rôle de cette liaison aux acides ribonucléiques chez un complexe protéique dont la fonction première est d'induire la répression génique.

1.2.3 PRC1

1.2.3.1 Composition

PRC1 se compose de quatre sous-unités Pc, PSC, Ph (Ph-p ou Ph-d) et dRING (Figure 1-2a et Tableau 1-2) (Kassis et al., 2017; Schuettengruber et al., 2017; Simon and Kingston, 2009). Toutes ces protéines sont indispensables *in vivo* et leur mutation conduit à des transformations homéotiques différentes qui mènent à la mort durant le développement embryonnaire chez *D. melanogaster* (Kassis et al., 2017). Même si le résultat final est le même quelle que soit la sous-unité mutée, la transformation homéotique varie d'une protéine à l'autre indiquant qu'elles ont des fonctions différentes dans le complexe au sein duquel elles s'assemblent, et potentiellement des fonctions en dehors de ce complexe PRC1 (Kassis et al., 2017).

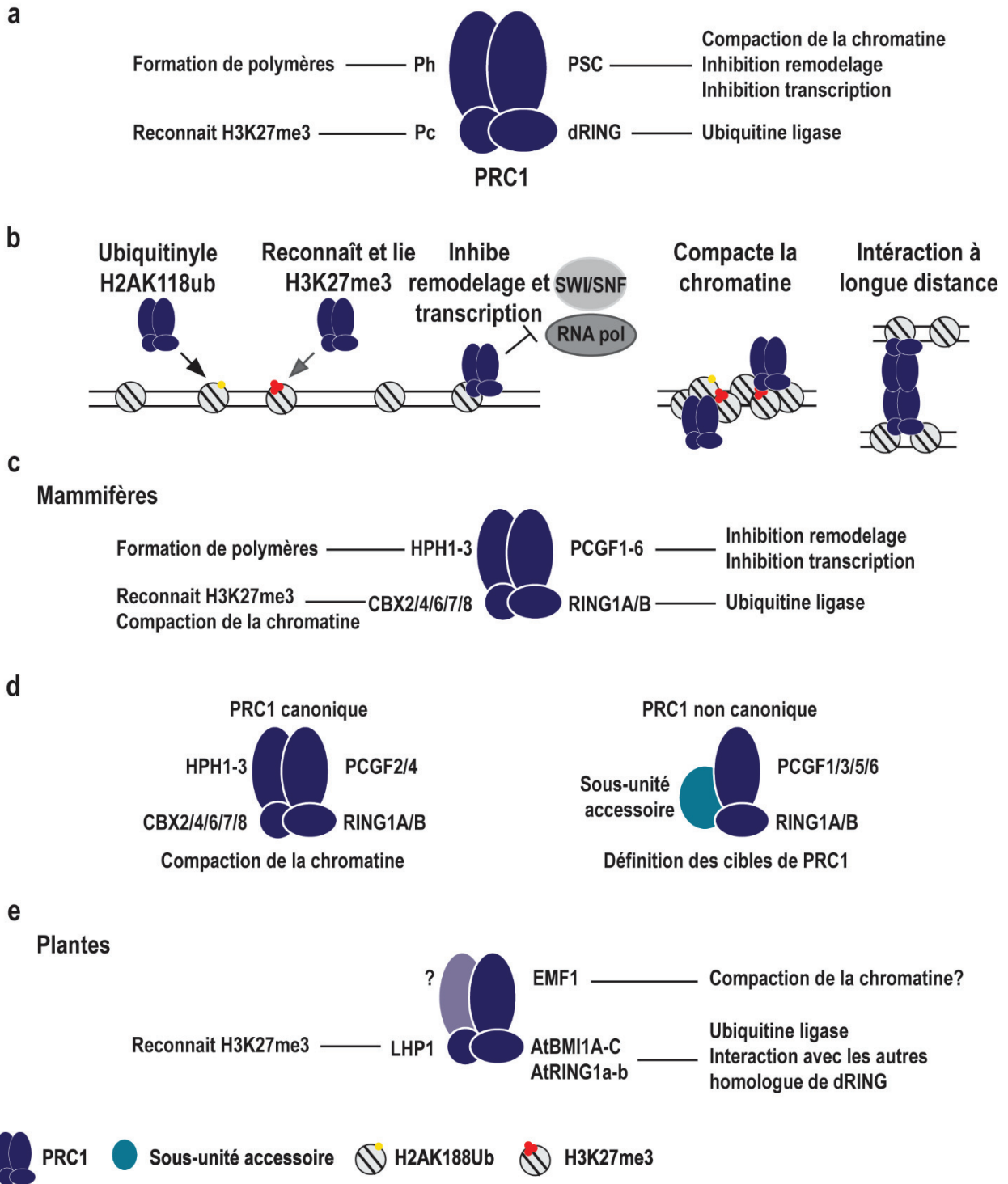


Figure 1-2 Le complexe PRC1 : composition et fonction

a. Composition du complexe PRC1 chez *Drosophila melanogaster* et liste des activités de ses sous-unités. b. Schéma des activités de PRC1. PRC1 ubiquitinylation la lysine 118 de l'histone H2A. PRC1 reconnaît et lie H3K27me3. PRC1 inhibe le remodelage de la chromatine par le complexe SWI/SNF et la transcription. PRC1 compacte la chromatine et peut lier deux fragments de chromatine ou d'ADN. c. Composition du complexe

PRC1 chez les mammifères et liste des activités des différentes sous-unités. d. Composition des complexes PRC1 canonique (gauche) et non-canonique (droite). e. Composition du complexe PRC1 chez les plantes et liste des activités des différentes sous-unités.

1.2.3.2 Le rôle central de PSC dans la répression induite par PRC1

PRC1 s'assemble en un complexe composé des protéines Pc, PSC, Ph (Ph-d ou Ph-p) et dRING, qui peuvent être co-purifiées à partir d'embryons de *D. melanogaster* (Saurin et al., 2001; Shao et al., 1999). La purification de ce complexe a permis d'identifier et de caractériser ses fonctions à l'aide d'essais *in vitro* et donc de comprendre comment les protéines du groupe Polycomb répriment l'expression génique *in vivo*. *In vitro*, PRC1 inhibe le remodelage de la chromatine par le complexe SWI/SNF lorsqu'il lie la chromatine avant ce deuxième complexe protéique (Figure 1-2b) (Francis et al., 2001; Shao et al., 1999). PRC1 ainsi que sa sous-unité PSC qui est responsable à elle seule de cette inhibition du remodelage de la chromatine, agit en empêchant le complexe SWI/SNF de lier la chromatine (Francis et al., 2001). Ils peuvent également compacter la chromatine et inhiber sa transcription en empêchant la formation du complexe de pré-initiation sans pour autant empêcher la liaison de facteurs de transcription tels que TFIID à son site d'intérêt (Figure 1-2b) (Francis et al., 2004; King et al., 2002; Lehmann et al., 2012).

PRC1 reste lié à la chromatine et à l'ADN nu durant leur réplication (Francis et al., 2009; Lengsfeld et al., 2012). *In vitro*, en utilisant deux plasmides il a été montré que durant la réplication PRC1 ne change pas de plasmide, ceci suggère que ce complexe Polycomb reste en contact avec la chromatine et ce, durant le passage de la fourche de réplication (Francis et al., 2009; Lengsfeld et al., 2012). De plus, PRC1 et PSC peuvent lier deux fragments d'ADN ou de chromatine ensemble (Figure 1-2b) (Lavigne et al., 2004; Lo et al., 2012a, 2012b). PSC peut interagir avec lui-même, permettant à chaque PSC d'établir un contact avec un nucléosome formant ainsi un lien entre deux fragments ou régions distales dans le génome (Lo and Francis, 2010; Lo et al., 2012a). Ceci a conduit au modèle suivant : PRC1 établirait des contacts avec plusieurs nucléosomes, relâcherait le nucléosome n'ayant pas été répliqué tout en maintenant le contact avec celui formé sur l'ADN ayant déjà été répliqué permettant à la fourche de réplication de progresser sans pour autant perdre le contact avec la chromatine (Francis et al., 2009; Lo et al., 2012a). Ce modèle est favorisé par rapport à celui où la persistance de PRC1 sur l'ADN ou la chromatine se ferait via une interaction avec la machinerie de réplication, car PRC1 ne nécessite pas la présence de protéines eucaryotes (Lengsfeld et al., 2012). La persistance de PRC1 sur l'ADN ou la chromatine malgré la réplication peut se faire lorsque l'ADN est répliqué avec la polymérase T7 issue de bactériophage (Lengsfeld et al., 2012). De plus, ce modèle permet à PRC1 de rester au niveau du même locus, lui permettant après la

réplication de réprimer l'expression des mêmes gènes et donc, ce mécanisme participerait à la transmission de l'information épigénétique lors de l'évènement disruptif qu'est la réplication de l'ADN.

Su(Z)2 est le paralogue de PSC. Tout comme PSC, il peut lier l'ADN et la chromatine, compacter et inhiber le remodelage de cette dernière, mais également former un complexe avec les autres sous-unités de PRC1 (Pc, Ph et dRING) (Lo et al., 2009). Su(Z)2 possède une région ne présentant pas d'homologie avec PSC ce qui peut laisser à penser que cette protéine pourrait avoir des fonctions différentes de PSC en plus de celles qui sont communes aux deux protéines (Lo et al., 2009). Su(Z)2 et PSC sont toutes deux indispensables *in vivo* et leur mutation conduit à la mort mais à des stades de développement différents (Wu and Howe, 1995). La mutation de PSC conduisant à la mort durant le développement embryonnaire et celle de Su(Z)2 durant les stades de larves et de pupes (Wu and Howe, 1995). En l'état actuel des connaissances, PSC et Su(Z)2 semblent donc assurer les mêmes fonctions au sein du complexe PRC1 mais à des stades de développement distincts.

1.2.3.3 Ph est essentiel pour établir des interactions à longues distances

Le domaine SAM de Ph, lui permet d'interagir avec lui-même et de former de longs polymères (Kim et al., 2002). Avec ce même domaine SAM, Ph peut interagir avec une autre protéine comportant elle-aussi un domaine SAM : Scm (Kim et al., 2005). Scm interagit et co-purifie avec PRC1, PRC2 et PhoRC, mais cette protéine n'est pas présente en quantité stœchiométrique avec les autres sous-unités de ces complexes (Saurin et al., 2001; Shao et al., 1999). La déplétion de Scm conduit à une diminution du recrutement de PRC2 à certains sites, mais la déplétion de PRC2 n'affecte pas le recrutement de Scm suggérant que Scm participe au recrutement de PRC2 mais pas l'inverse (Kang et al., 2015; Wang et al., 2010). Le domaine SAM de Scm pourrait également servir de lien entre PhoRC et PRC1 en formant un hétéropolymère avec les domaines SAM de Sfm (sous-unité de PhoRC, voir la section 1.2.4) et de Ph, contribuant au recrutement séquence spécifique de PRC1 (Frey et al., 2016). Mais le fait que le recrutement de PRC1 et PRC2 soit simplement réduit et non pas abolie en cas de déplétion de Scm, montre que Scm n'est pas le seul facteur contribuant à leur recrutement. Le domaine SAM de Ph, en interagissant directement avec ce même domaine de la protéine Sfm pourrait permettre à PRC1 d'être recruté au niveau de leur séquence cible de manière séquence spécifique et indépendamment de Scm (Frey et al., 2016).

Le domaine SAM de Ph est également important pour la formation d'interaction avec d'autres complexes PRC1 (Isono et al., 2013). PRC1 forment des clusters visibles en microscopie dont la taille est réduite en cas d'expression d'une version mutante de Ph incapable de former des polymères (Ph-ML) en plus de la protéine endogène (Wani et al., 2016). L'expression de Ph-ML conduit aussi à une diminution des contacts

locaux et de manières plus importantes des contacts distaux (Wani et al., 2016). PRC1 via la capacité de Ph à former des polymères participent à la formation de ces domaines Polycomb et à la formation d'interactions à longues distances dans la chromatine (Wani et al., 2016). Ce n'est donc pas surprenant que ces polymères, formés par Ph, doivent être régulés afin de permettre des interactions à longues distances sans pour autant conduire à l'agrégation de Ph ce qui conduirait à une mauvaise expression des gènes régulés par PRC1 (Gambetta and Müller, 2014; Robinson et al., 2012; Wani et al., 2016).

1.2.3.4 dRING ubiquitinyler la lysine 118 de l'histone H2A

dRING est une ubiquitine ligase E3, qui joue un rôle en dehors du complexe PRC1 en s'associant avec PSC et dKDM2 formant ainsi le complexe dRAF (Figure 1-3a) (Kassis et al., 2017; Lagarou et al., 2008; Wang et al., 2004a). dKDM2, seul ou au sein de complexe dRAF, peut enlever la méthylation de la lysine 36 de l'histone 3 (H3K36me2) déposée par les protéines du groupe Trithorax (Figure 1-3b) (Lagarou et al., 2008). dRING et PSC au sein des complexes PRC1 ou dRAF, peuvent mono-ubiquitinyler la lysine 118 de l'histone H2A, et cette ubiquitination est stimulée par la présence de dKDM2 à la fois *in vitro* et *in vivo* (Lagarou et al., 2008). *In vivo*, dKDM2, seule sous-unité de dRAF qui ne fait pas partie du complexe PRC1, participe à la répression des gènes Hox (Lagarou et al., 2008). H2AK118ub est reconnu par PRC2, induisant son recrutement et cette ubiquitination permet également la stimulation de l'activité méthyltransférase du complexe PRC2 mammalien lorsqu'il est accompagné de ses protéines accessoires Jarid2 et AEBPB2 (Kalb et al., 2014). Malgré cela, et la conservation de la reconnaissance de H2AK118ub par PRC2 chez *D. melanogaster*, l'absence d'ubiquitination de H2AK118 grâce à la substitution de la lysine 118 par une arginine n'a pas d'effet sur la répression des gènes Hox (Kalb et al., 2014; Pengelly et al., 2015).

Cette marque déposée par les complexes dRAF ou PRC1 peut être enlevée par la complexe PR-DUB, composé de Calypso et de ASX (Figure 1-3) (Kassis et al., 2017; Scheuermann et al., 2010). Tout comme dKDM2 et le complexe dRAF, le complexe PR-DUB participe également à la répression des gènes Hox (Gutiérrez et al., 2012; Lagarou et al., 2008; Scheuermann et al., 2010). Il est possible que certains gènes nécessitent l'ubiquitination des histones H2A afin de permettre le recrutement des autres complexes Polycomb ainsi que leur répression, alors que d'autres loci ne reposent pas sur cette modification post-traductionnelle d'histone comme cela a été suggéré (Gutiérrez et al., 2012; Schwartz and Pirrotta, 2014).

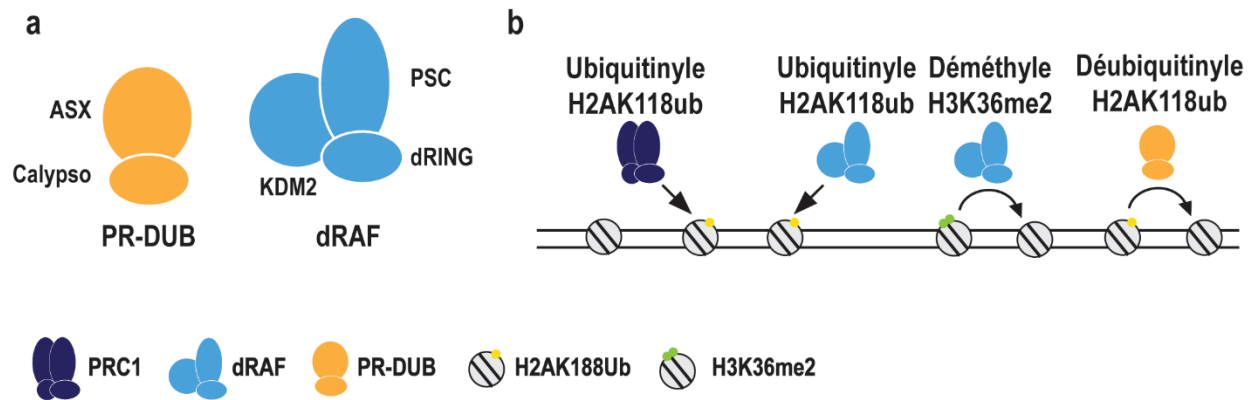


Figure 1-3 Les complexes Polycomb PR-DUB et dRAF

a. Composition des complexes PR-DUB et dRAF. b. Schéma des activités de PR-DUB et dRAF. dRAF ubiquitinye la lysine 118 de l'histone H2A et déméthyle la lysine 36 de l'histone 3. PR-DUB enlève l'ubiquitination de la lysine 118 de l'histone H2A déposée par PRC1 et dRAF.

1.2.3.5 PRC1 reste lié aux chromosomes lors de la mitose

La mitose est un évènement particulièrement disruptif du cycle cellulaire durant lequel la plupart de protéines liant la chromatine sont enlevées de celle-ci afin de lui permettre d'être compactée sous la forme de chromosomes mitotiques. Une fraction des protéines composants le complexe PRC1 reste néanmoins liée à la chromatine durant la mitose (Follmer et al., 2012). Cette diminution de la liaison de PRC1 n'est pas le résultat d'une diminution globale de sa présence à tous les sites que le complexe lie, mais est due à une rétention spécifique de PRC1 à une fraction des loci que celui-ci occupe durant les autres phases du cycle cellulaire (Follmer et al., 2012). Parmi ces sites qui conservent la liaison des protéines Polycomb, on peut voir des sites liés par le complexe PRC1 afin de réprimer l'expression d'un gène mais aussi des sites en bordures entre deux domaines topologiques (Follmer et al., 2012). L'hypothèse du marque-page mitotique a conduit à proposer que ces sites auxquels PRC1 persiste en mitose permettraient un recrutement plus rapide d'autres complexes PRC1, lorsque celle-ci se termine, grâce aux interactions possibles entre plusieurs complexes PRC1 qui les garderaient à proximité (Buchenau et al., 1998; Follmer et al., 2012; Fonseca et al., 2012).

1.2.3.6 PRC1 est conservé chez d'autres espèces

1.2.3.6.1 Chez les mammifères

La diversité du complexe PRC2 chez les mammifères provient essentiellement de la multiplicité de ses sous-unités accessoires, mais celle de PRC1 provient des sous-unités qui le compose (Figure 1-2c et Tableau 1-2). Il existe de multiples versions pour chacune des quatre sous-unités qui composent PRC1, qui

peuvent s'assembler afin de former des complexes canoniques mais également non-canoniques, où toutes les sous-unités qui composent le noyau de PRC1 ne sont pas présentes mais sont accompagnées de multiples sous-unités accessoires (Figure 1-2d) (Chittock et al., 2017; Connelly and Dykhuizen, 2017; Kuroda et al., 2020). RING1A et RING1B, homologues de dRING, peuvent ubiquitinyler la lysine 119 de l'histone H2A. CBX 2, 4, 6, 7 et 8, homologues de Pc, peuvent reconnaître H3K27me3 et Cbx2 peut compacter la chromatine (Grau et al., 2011). PCGF 1 à 6, homologues de PSC, sont recrutés dans des régions allant de complètement réprimées à transcriptionnellement actives (Scelfo et al., 2019). HPH1 à 3 sont les homologues de Ph, qui grâce à leur domaine SAM peuvent former de long polymères qui jouent un rôle dans la condensation de la chromatine (Isono et al., 2013). Ces différentes versions de chaque protéine ne s'expriment pas toutes en même temps mais certaines s'expriment dans les cellules non différenciées uniquement, d'autres dans les cellules complètement différenciées ou en différenciation.

Plus généralement les complexes canoniques et non-canoniques de PRC1 ont des fonctions différentes dans la répression génique. Les versions non-canoniques du complexe PRC1 qui contiennent RING1a ou RING1b, PCGF 1, 3, 5 ou 6 ainsi que des sous-unités accessoires, définissent les cibles du complexe PRC1 canonique et de PRC2 (Fursova et al., 2019). La perte de PCGF1, 3, 5 et 6 qui sont uniquement présents dans les versions non-canoniques du complexe PRC1, conduit à une diminution des niveaux de H2AK119ub et H3K27me3, ainsi qu'à une perte du recrutement de PRC2 à ses cibles et une dérégulation similaire à celle induite par la déplétion de RING1a et RING1b (Fursova et al., 2019). Hormis la délétion de PCGF1 qui conduit à la diminution du niveau de H2AK119ub, la perte de PCGF 3, 5 ou 6 n'a que peu d'effet montrant le caractère redondant de ces protéines (Fursova et al., 2019). Le complexe canonique PRC1, grâce aux activités de compaction locale et à longues distances des protéines CBX et HPH, établit des domaines Polycomb compacts qui ne dépendent pas des versions non-canoniques du complexe PRC1 (Kundu et al., 2018).

1.2.3.6.2 Chez les plantes

Contrairement au complexe PRC2, toutes les sous-unités de PRC1 n'ont pas d'homologues chez *Arabidopsis thaliana* (Figure 1-2e et Tableau 1-2) (Hennig and Derkacheva, 2009; Kim and Sung, 2014; Merini and Calonje, 2015; Wang and Shen, 2018). En effet, à ce jour aucun homologue n'a été trouvé pour Ph.

LHP1, analogue fonctionnel de Pc, lie H3K27me3, et la perte de LHP1 ou son incapacité à lier la modification post-traductionnelle déposée par PRC2 conduit à une dérégulation des gènes régulés par les protéines du groupe Polycomb (Exner et al., 2009).

dRING possède cinq homologues fonctionnels qui sont tous capables d'ubiquitinyler l'histone H2A : AtBMI1A, AtBMI1B, AtBMI1C, AtRING1a et AtRING1b. AtBMI1A et AtBMI1B participent à la répression des gènes impliqués dans le développement et la maintenance des cellules souches, alors que AtBMI1C agit dans l'endosperme et les étamines (Bratzel et al., 2010, 2012). Les protéines AtRING1a et AtRING1b sont quant à elles essentielles pour la régulation de la transition entre les phases végétative et reproductive (Li et al., 2017b). Les protéines AtRING1a et AtRING1b peuvent interagir avec elles-mêmes mais aussi avec les autres homologues de dRING (AtBMI1A, AtBMI1B et AtBMI1C) et avec LHP1 (Chen et al., 2010).

EMF1, homologue de PSC, interagit avec LHP1 et les différents homologues de dRING, son expression est essentielle pour l'ubiquitination de l'histone H2A (Bratzel et al., 2010; Wang et al., 2014). La perte de EMF1 conduit à un phénotype de floraison précoce similaire à celui induit par la perte de EMF2, sous-unité de PRC2 (Kim et al., 2012). PRC1 et PRC2 possèdent de nombreuses cibles communes, et la perte de EMF1 ou EMF2 conduit à une diminution du niveau de tri-méthylation de la lysine 27 de l'histone 3 ainsi qu'à une augmentation de l'expression de ces gènes (Kim et al., 2012). EMF1 semble donc servir de lien entre PRC1 et PRC2 à leurs cibles communes (Kim et al., 2012). D'autres gènes en revanche présentent un niveau de méthylation inchangé en absence de EMF1 montrant que PRC2 peut également agir indépendamment de PRC1 (Kim et al., 2012).

La multiplicité des complexes PRC1 et PRC2 chez les plantes fait que les différentes versions de ces complexes s'associent afin de réprimer l'expression de gènes différents lors de différents événements du développement ou à la suite de stimuli extérieurs (Wang et al., 2016a).

| | <i>Drosophila</i> | Mammifères | Plantes |
|------|-------------------|--|---|
| PRC1 | PSC | PCGF1 PCGF2 PCGF3 PCGF4 PCGF5 PCGF6 | EMF1 |
| | Ph | HPH1 HPH2 HPH3 | ? |
| | Pc | CBX2 CBX4 CBX6 CBX7 CBX8 | LHP1 |
| | dRING | RING1A RING1B | AtBMI1A AtBMI1B AtBMI1C AtRING1a AtRING1b |

Tableau 1-2 Liste des sous-unités du complexe PRC1 chez *Drosophila melanogaster*, les mammifères et les plantes

1.2.4 Le complexe PhoRC

Le complexe PhoRC se compose de Pho et de dSfmbt et qui, comme PRC1 et PRC2, est recruté aux PREs (Figure 1-4a) (Kassis et al., 2017; Klymenko et al., 2006). La mutation de dSfmbt ou de Pho, en combinaison avec son homologue Phol, conduit à une mauvaise expression des gènes HOX (Brown et al., 2003;

Klymenko et al., 2006). Comme les protéines des autres complexes Polycomb, leur recrutement au niveau d'un gène rapporteur induit la répression de ce dernier (Klymenko et al., 2006). Pho est la seule protéine du groupe Polycomb pouvant lier l'ADN de manière séquence spécifique (Figure 1-4b) (Brown et al., 1998), ce qui avait conduit à proposer un modèle de recrutement séquentiel des protéines Polycomb, avec PhoRC liant les PREs, induisant le recrutement de PRC2 qui tri-méthyle les lysines 27 de l'histone H3 qui induit à son tour le recrutement de PRC1 afin de conduire à la compaction de la chromatine et à la répression génique (Wang et al., 2004b). Mais ce modèle a été remis en question de nombreuses fois depuis sa publication en faveur d'un modèle de recrutement combinatoire dans lequel les différents complexes Polycomb et les modifications post-traductionnelles qu'ils déposent sur les histones vont promouvoir leur propre recrutement ainsi que celui des autres complexes Polycomb, afin d'induire la répression génique (Dorafshan et al., 2017; van Kruijsbergen et al., 2015). Sfmtb possède un domaine SAM qui lui permet d'interagir avec Scm et Ph (Frey et al., 2016). Ce domaine SAM est notamment essentiel pour le recrutement des complexes Polycomb et la répression appropriée du gène *Ubx* (Frey et al., 2016).



Figure 1-4 Le complexe Polycomb PhoRC

a. Composition du complexe PhoRC. b. PhoRC lie l'ADN de manière séquence spécifique.

1.2.5 Les complexes Polycomb participent à leur recrutement mutuel

Le premier modèle de recrutement des complexes Polycomb était un modèle hiérarchique qui reposait entièrement sur le complexe PhoRC afin de donner la spécificité de séquence (Wang et al., 2004b). Pho en interagissant avec PRC2 permettait alors le recrutement spécifique de ce dernier, qui va ensuite pouvoir modifier de manière post-traductionnelle les histones (Figure 1-5) (Wang et al., 2004b). PRC2 peut reconnaître la tri-méthylation de la lysine 27 de l'histone 3 qu'il a lui-même déposée stimulant son activité méthyltransférase (Margueron et al., 2009). Finalement, Pc en reconnaissant la modification post-traductionnelle déposée par PRC2 permet le recrutement de PRC1 (Fischle et al., 2003). Mais ce modèle hiérarchique a été remis en question. PRC1 grâce à son interaction avec Scm et le complexe PhoRC peut

être recruté spécifiquement aux PREs sans nécessiter au préalable le recrutement de PRC2 (Figure 1-5) (Frey et al., 2016). PRC2 peut également lier Scm, et la déplétion de cette dernière affecte le recrutement de PRC2 (Kang et al., 2015). PRC2, grâce à ses sous-unités accessoires Jarid2 et AEBP2, peut reconnaître H2AK118ub déposé par PRC1, permettant le recrutement de PRC2 ainsi que la stimulation de son activité méthyltransférase (Kalb et al., 2014).

PRC1 et PRC2 participent donc à leur propre recrutement mais aussi au recrutement de l'autre complexe. *In vivo*, cela s'observe par des loci auxquels le recrutement de PRC1 dépend de PRC2 et inversement, mais plus généralement la déplétion de PRC1 ou PRC2 ne se traduit pas par la perte complète de l'autre complexe mais par une diminution de son recrutement, montrant qu'un simple recrutement hiérarchique ne peut pas résumer cette complexe coopération entre les protéines du groupe Polycomb (Dorafshan et al., 2017).

L'absence de PRC1 conduit à la diminution du recrutement de PhoRC chez *D. melanogaster*, montrant que PhoRC n'est pas à la base du recrutement de tous les complexes Polycomb (Kahn et al., 2014). De plus, l'absence de complexe PhoRC chez d'autres espèces et le rôle du complexe PRC1 non-canonique dans la définition des cibles des protéines du groupe Polycomb chez les mammifères, laissent à penser que ces mécanismes de recrutement des autres complexes Polycomb reposant sur PhoRC lors de la phase d'initiation, pourraient être restreints à un nombre limité de cibles chez *D. melanogaster* (Blackledge et al., 2014; Kahn et al., 2016). Plus probablement c'est un modèle coopératif qui serait à l'origine du recrutement des complexes Polycomb, avec de multiples facteurs conduisant à leur recrutement (voir la section 1.3).

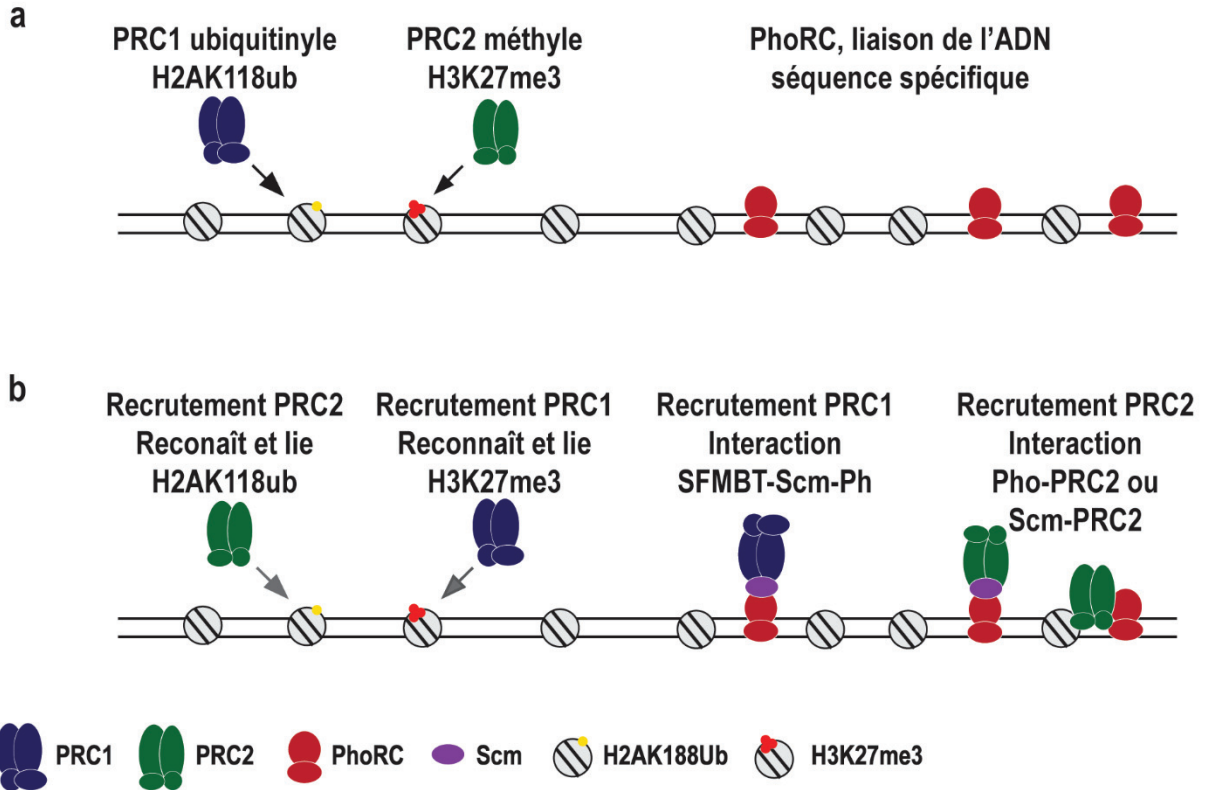


Figure 1-5 Les complexes PRC1, PRC2 et PhoRC participent à leur recrutement mutuel

a. PRC1 ubiquitinyne la lysine 118 de l'histone de H2A. PRC2 méthyle la lysine 27 de l'histone 3. PhoRC lie l'ADN de façon séquence spécifique. b. PRC1 et PRC2 reconnaissent et lient respectivement H3K27me3 et H2AK118ub participant ainsi à leur recrutement mutuel. Les domaines SAM de SFMBT, Scm et Ph permettent l'interaction de PhoRC et PRC1 participant au recrutement de ce dernier. L'interaction de PRC2 avec Pho ou de PRC2 avec le domaine SAM de Scm permet son recrutement.

1.3 Les éléments de réponse Polycomb

Chez *D. melanogaster*, les fonctions des protéines du groupe Polycomb ont été initialement découvertes en raison de leur recrutement à des sites précis appelés éléments de réponse Polycomb (PREs) qui participent au contrôle de l'expression spatio-temporelle des gènes Hox impliqués dans le développement. Les PREs, bien que découverts il y a plus de vingt ans, ont été difficiles à identifier en raison de l'absence de séquence consensus simple, mais une définition fonctionnelle bien plus précise est toutefois disponible.

1.3.1 La définition fonctionnelle des éléments de réponse Polycomb

Les éléments de réponses Polycomb furent découverts chez *D. melanogaster* pour leur capacité à induire la répression génique de manière dépendante des protéines du groupe Polycomb (Chan et al., 1994; Simon

et al., 1993; Steffen and Ringrose, 2014). Des expériences utilisant un gène rapporteur sous le contrôle d'un PRE ont permis de définir leurs fonctions.

1.3.1.1 Un élément mémoire

L'expression d'un gène rapporteur sous le contrôle d'un PRE conduit au recrutement des protéines du groupe Polycomb, qui modifient de manière post-traductionnelles les histones, compactent la chromatine conduisant ainsi à la répression de ce transgène (Figure 1-6a) (Schwartz and Pirrotta, 2008; Steffen and Ringrose, 2014). Mais lorsque ce PRE est combiné avec un amplificateur, il peut contrecarrer l'action activatrice de ce dernier afin de réprimer le gène rapporteur et de ne permettre son expression que dans les segments appropriés de *D. melanogaster* (Figure 1-6b) (Simon et al., 1993; Steffen and Ringrose, 2014). De la même façon, l'utilisation d'un amplificateur actif uniquement dans les disques imaginaux aux stades embryonnaires ou larvaires ont mis au jour le rôle mémoire des PREs qui permettent l'expression du gène rapporteur uniquement aux stades développementaux et dans les segments appropriés (Figure 1-6b) (Chan et al., 1994; Steffen and Ringrose, 2014). Le fait qu'un PRE puisse maintenir la répression ou l'expression d'un gène qu'il contrôle pendant le développement, dépendamment de son association avec différents éléments amplificateurs, montre le caractère mémoire de celui-ci. Ceci sous-entend également qu'un PRE puisse permettre la transmission des informations concernant l'expression d'un gène pendant de nombreuses divisions cellulaires et donc pendant la réplication de l'ADN ainsi que la mitose qui sont deux évènements disruptifs (Steffen and Ringrose, 2014). Ce rôle mémoire a notamment été mis en avant lors de l'excision d'un PRE durant le développement larvaire mais après établissement de la répression du gène rapporteur (Busturia et al., 1997). L'excision de ce PRE a conduit à la dérégulation du gène rapporteur après quelques heures montrant leur fonction de mémoire de la répression (Busturia et al., 1997).

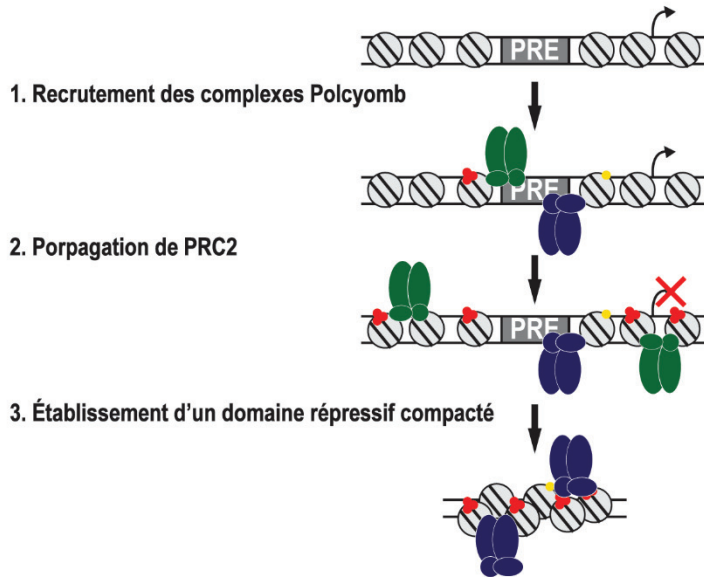
En résumé, les PREs sont des éléments mémoires qui par défaut induisent la répression génique dépendamment des protéines du groupe Polycomb mais qui, lorsqu'ils sont associés avec un élément amplificateur permettent l'expression des gènes qu'ils contrôlent dans le contexte développemental approprié uniquement (Schwartz and Pirrotta, 2008; Steffen and Ringrose, 2014).

1.3.1.2 Un élément mémoire permutable

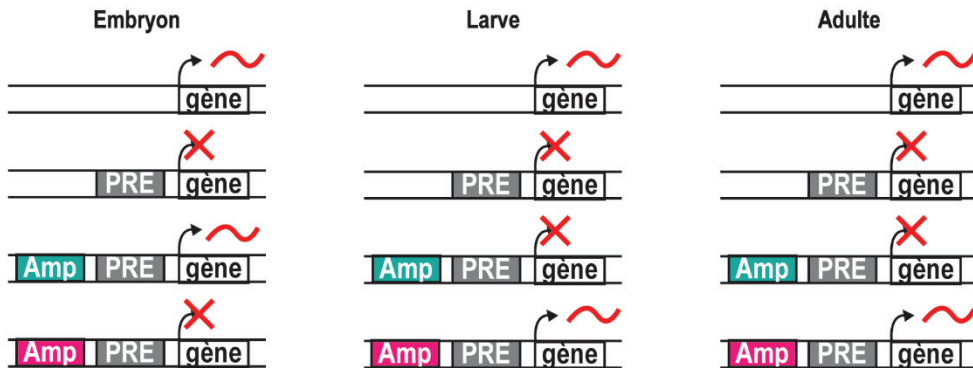
En utilisant un système similaire de gène rapporteur sous le contrôle d'un PRE et d'un UAS (*upstream activating sequence*), il a été possible de mettre au jour le rôle de mémoire de la transcription active des PREs. En absence d'activation du système UAS-GAL4, le PRE induit la répression du gène rapporteur. Mais à la suite d'un choc thermique lors du stade larvaire, le système UAS-GAL4 est induit conduisant à la transcription du gène rapporteur. Cette transcription persiste malgré l'arrêt de l'induction, signifiant que

les protéines du groupe Polycomb ont été durablement exclues de ce PRE (Cavalli and Paro, 1998; Maurange and Paro, 2002; Steffen and Ringrose, 2014). Ceci a permis la démonstration que les PREs ne sont pas uniquement capable de maintenir la répression mais également la transcription d'un gène de manière stable durant le développement. À l'opposé des protéines du groupe Polycomb qui sont nécessaires au maintien de la répression, les protéines du groupe Trithorax le sont pour l'expression, car elles antagonisent l'action de ce premier groupe de protéines en déposant des modifications post-traductionnelles sur les histones qui ne peuvent coexister au sein d'un même nucléosome avec celles déposées par les complexes Polycomb (Figure 1-6c) (Steffen and Ringrose, 2014).

a
Répression par les protéines du groupe Polycomb



b
Élément mémoire



c
Élément permutable

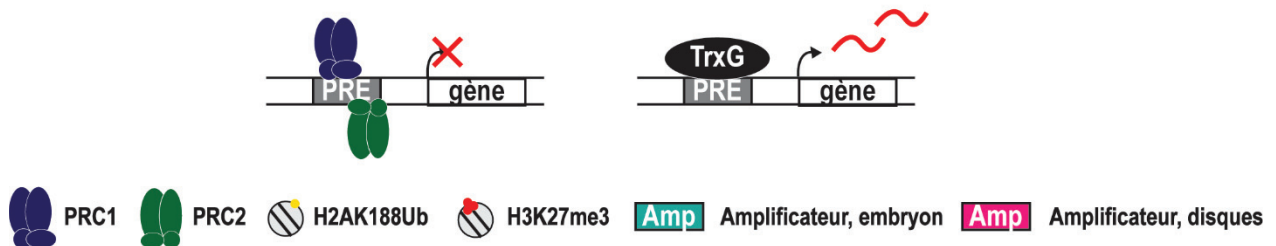


Figure 1-6 Schémas illustrant la définition fonctionnelle des éléments de réponse Polycomb (PREs)

a. Le recrutement des protéines du groupe Polycomb à un PRE conduit à l'établissement d'un domaine répressif. (1) Les complexes PRC1 et PRC2 sont recrutés au niveau d'un PRE et modifient les histones de

manières post-traductionnelles, déposition de H3K27me3 et H2AK118ub. (2) La propagation de PRC2 le long de la chromatine permet l'établissement d'un domaine marqué par H3K27me3. (3) La compaction de la chromatine par PRC1 conduit à la formation d'un domaine compacte et réprimé. b. Un PRE est un élément mémoire. Un même PRE associé à différents éléments amplificateurs permet l'expression du gène qu'il contrôle uniquement durant les stades de développement appropriés. c. Un PRE est un élément permutable. Un PRE lié par les protéines du groupe Polycomb induit la répression du gène qu'il contrôle. Le même PRE lié par les protéines du groupe Trithorax conduit à l'expression du gène.

1.3.1.3 La répression sensible à l'appariement des PREs (*Pairing-sensitive silencing* ou PSS)

L'utilisation du gène *miniwhite*, qui définit la couleur des yeux de *D. melanogaster* et donne une pigmentation différente en fonction de son niveau d'expression, comme gène rapporteur a permis de mettre en évidence la capacité des PREs à s'apparier afin de conduire à une plus forte répression (Chan et al., 1994; Fauvarque and Dura, 1993; Kassis, 1994). Sous forme hétérozygote, une copie du gène *miniwhite* sous le contrôle d'un PRE va conduire à sa répression et les Drosophile ont des yeux blanc ou jaune claire. A l'état homozygote, au lieu d'observer un œil plus sombre qui serait le résultat d'une expression plus forte due à la présence de deux copies du gène *miniwhite*, le phénomène inverse se produit et une répression plus forte est observée (Chan et al., 1994; Fauvarque and Dura, 1993; Kassis, 1994, 2002). Ceci peut être dû à PRC1 qui grâce à sa capacité à se lier à plusieurs nucléosomes à la fois, va permettre l'appariement de PREs distants observable par FISH, créant ainsi un site où les PREs et les protéines du groupe Polycomb sont regroupés permettant une répression plus forte du gène rapporteur (Bantignies et al., 2011).

1.3.2 Les PREs permettent la formation de domaines Polycomb répressifs

1.3.2.1 Les PREs sont des sites de nucléation pour les protéines du groupe Polycomb

Les PREs induisent la répression en recrutant les protéines du groupe Polycomb qui vont pouvoir modifier la chromatine et former un domaine répressif bien plus large que le PRE lui-même. Des analyses de ChIP à l'échelle du génome entier chez *D. melanogaster* ont révélé que les PREs représentent des sites de liaison pour les protéines du groupe Polycomb qui sont visibles sous forme de pics imposants. Mais ces mêmes données de ChIP-seq ont montré également la présence de plus petits pics qui ont tendance à se situer autour des PREs dans des régions pouvant être longues de plusieurs centaines de kilobases et qui se caractérisent pas un haut niveau de tri-méthylation de la lysine 27 de l'histone H3 (Nègre et al., 2006; Schwartz et al., 2006; Tolhuis et al., 2006). Les PREs sont alors pensés comme sites de nucléation initiaux des complexes PRC1 et PRC2, qui vont alors pouvoir se propager le long de la chromatine afin de former

ces domaines Polycomb compacts. Afin d'éviter une propagation trop importante de la tri-méthylation de la lysine 27 de l'histone 3 et donc de ces domaines répressifs, des protéines isolatrices telles que CTCF ont été identifiées en bordures de ces domaines Polycomb afin de contrecarrer leur propagation (Cuddapah et al., 2009; Schwartz et al., 2012; Van Bortle et al., 2012). De façon intéressante, il a été également observé une complémentarité entre les modifications post-traductionnelles d'histones répressives, telles que H3K27me3, et actives comme H3K36me3 qui sont mutuellement exclusives (Schuettengruber and Cavalli, 2013; Schuettengruber et al., 2009).

1.3.2.2 Les PREs sont impliqués dans la formation de domaines répressifs

Les protéines du groupe Polycomb ne sont pas diffuses dans le noyau mais forment des domaines Polycomb visibles sous forme de foci, dans les cellules et embryons de *D. melanogaster*, au sein desquels les protéines Polycomb sont mobiles (Buchenau et al., 1998; Ficz et al., 2005). Ces domaines Polycomb rassemblent plusieurs PREs dans un même état répressif mais excluent les PREs actifs (Bantignies et al., 2011). Les protéines du groupe Polycomb répriment la transcription dans les domaines Polycomb, notamment en empêchant leur accès aux ARN polymérases et en inhibant le remodelage de la chromatine par le complexe SWI-SNF (Fitzgerald and Bender, 2001; Shao et al., 1999). La formation de ces domaines reposent sur la capacité de Ph à former de longs polymères grâce à son domaine SAM (Isono et al., 2013; Kim et al., 2002). Ces domaines Polycomb sont présents à l'intérieur de domaines topologiquement associés (TAD) qui rassemblent des régions distantes de plusieurs mégabases (Hou et al., 2012; Lanzaolo et al., 2007; Sexton et al., 2012). À l'intérieur de ces TAD, les PREs et protéines du groupe Polycomb facilitent les contacts entre des régions distantes au sein ces domaines afin de former des compartiments répressifs réfractaires à l'activation (Ogiyama et al., 2018).

1.3.2.3 Les PREs, entre répression et expression

Les PREs peuvent induire la répression ou la transcription des gènes qu'ils contrôlent dépendamment des protéines qui y sont liés. La réduction du taux de protéines Trx chez la drosophile conduit à un plus faible niveau d'expression du gène rapporteur et à une répression plus fréquente de ce même gène durant le développement (Poux et al., 2002). En l'absence des protéines des groupes Polycomb ou Trithorax, on peut observer une exacerbation des effets du groupe de protéines restant : une plus forte répression en absence des protéines Ash1 et Trx et une plus forte expression en absence des protéines du groupe Polycomb (Klymenko and Müller, 2004; Poux et al., 2002). Il semble donc qu'il y ait un équilibre entre protéines des groupes Polycomb et Trithorax : lorsque les protéines du groupe Polycomb sont liées à un PRE ceci conduit à la répression, alors que le même PRE lié par les protéines du groupe Trithorax conduit

à la transcription, montrant le caractère antagoniste de ces deux groupes de protéines (Klymenko and Müller, 2004; Poux et al., 2002).

Les protéines du groupe Trithorax modifient de manière post-traductionnelle les histones (Yuan et al., 2011) ce qui antagonise le recrutement des protéines du groupe Polycomb et notamment du complexe PRC2 (Pasini et al., 2010; Schmitges et al., 2011) qui ne peuvent pas tri-méthyliser la lysine 27 de l'histone H3 si cette dernière est déjà acétylée (Jung et al., 2010). L'acétylation des histones conduit également à une réduction de la charge positive des histones ce qui peut déstabiliser leur interaction avec l'ADN (Hong et al., 1993) affectant potentiellement la capacité de PRC1 à lier et compacter la chromatine.

L'idée que les protéines des groupes Polycomb et Trithorax, et les marques épigénétiques qui leurs sont associées soient mutuellement exclusives est tentante mais ne représente pas la réalité. En effet, des PREs peuvent être liés par les protéines des groupes Polycomb et Trithorax en même temps et contenir des marques épigénétiques à la fois répressives et actives (qui ne sont pas présentes au sein du même nucléosome) (Langlais et al., 2012; Papp and Müller, 2006; Schwartz et al., 2010). Ces PREs dans un état bivalent penchent du côté Polycomb ou Trithorax, mais peuvent changer plus rapidement d'état (répressif ou actif) qu'un PRE monovalent (Harikumar and Meshorer, 2015; Sneppen and Ringrose, 2019). L'identification de PREs bivalents ayant été faite par ChIP-seq sur une population de cellules issues de *D. melanogaster* ou de lignées cellulaires n'exclue pas la possibilité que l'état bivalent observé pour certains PREs soient un effet de moyenne dû aux analyses faites sur une population de cellules et non pas la réalité observée dans une seule cellule.

1.3.2.4 Les PREs ne sont pas nécessaires pour la transmission de la répression durant un nombre limité de cycles cellulaires

Les protéines du groupe Polycomb assurent la répression stable des gènes pendant le développement et lors de nombreuses divisions cellulaires. Les PREs sont essentiels pour le recrutement des protéines du groupe Polycomb et l'établissement de ces domaines Polycomb répressifs. Mais une fois la répression établie, l'absence de PRE ne se traduit pas par une dérégulation immédiate puisque PRC2 et sa marque épigénétique associée H3K27me3 restent présents après plusieurs divisions (Coleman and Struhl, 2017; Laprell et al., 2017). La mitose et la réplication de l'ADN sont deux événements disruptifs du cycle cellulaire qui nécessitent d'enlever les protéines liées à l'ADN et à la chromatine. Les protéines du groupe Polycomb peuvent néanmoins continuer à assurer leurs fonctions malgré ces disruptions.

Les histones et leur modifications post-traductionnelles sont transmises lors de la réplication de l'ADN mais sont divisées entre les deux ADN, ce qui signifie qu'afin de restaurer le niveau de modifications épigénétiques qui était présent avant la réplication, de nouvelles modifications doivent être apportées. Les marques déposées par PRC2 et PRC1 participant à leur propre recrutement, et PRC1 restant lié à l'ADN durant la réplication, on peut imaginer que H3K27me3 et H2AK118ub restant sur l'ADN sont suffisants pour permettre une restauration rapide du statut épigénétique après réplication.

Lors de la mitose, les nucléosomes ne sont pas enlevés de la chromatine, mais cette dernière est fortement compactée. Les marques épigénétiques actives et répressives sont donc toujours présentes à leurs sites respectifs. La fraction des complexes PRC1 restant liée aux chromosomes lors de la mitose permettrait, selon l'hypothèse du marque-page mitotique, un recrutement rapide de PRC1 afin d'assurer la répression de leur cible dès la fin de la mitose (Follmer et al., 2012).

Toutefois, après un certain nombre de divisions, la répression finie par être perdue montrant que les PREs sont essentiels afin de permettre le maintien du statut transcriptionnel d'un gène dans le temps (Coleman and Struhl, 2017; Laprell et al., 2017).

Malgré toutes ces connaissances sur les PREs, le fait qu'ils puissent à la fois induire la répression ou permettre la transcription d'un gène dépendamment du contexte développemental, fait qu'aujourd'hui, plus de vingt ans après leur découverte nous cherchons toujours à comprendre comment les protéines du groupe Polycomb sont recrutées au niveau des PREs et qu'est-ce qui induit le changement d'un PRE d'un état répressif à permissif permettant le changement de l'expression des gènes qu'il régule.

1.3.3 Les propriétés de la séquence ADN

Afin d'assurer le bon développement de *D. melanogaster* les gènes Hox, dont l'expression dépend des PREs ainsi que du recrutement des protéines du groupe Polycomb, doivent être précisément régulés. Pour comprendre comment cette régulation se fait et essayer d'identifier des règles communes aux PREs et aux recrutements des complexes Polycomb, l'identification des PREs à l'échelle du génome entier à l'aide d'une séquence consensus a été une étape essentielle (Kassis and Brown, 2013; Ringrose and Paro, 2007).

1.3.3.1 Les PREs sont liés par des facteurs de transcription

L'étude des PREs identifiés fonctionnellement a mis en évidence plusieurs caractéristiques et notamment leur hypersensibilité à la DNase et donc leur faible densité de nucléosomes (Kahn et al., 2006). Ceci est contre-intuitif car on connaît les PREs pour leur capacité à recruter PRC1 et PRC2 qui agissent en modifiant les histones et la structure de la chromatine. L'ADN libre de nucléosome au niveau des PREs les rend accessibles à de nombreux facteurs de transcription qui peuvent les lier de manière séquence spécifique.

Parmi ces protéines liant les motifs présents au niveau des PREs, il y a Pho/PhoI (Brown et al., 2003), Dsp1 (Decoville et al., 2001), GAF (Bejarano and Busturia, 2004), Psq (Huang et al., 2002), Zeste (Déjardin and Cavalli, 2004), Grh (Blastyák et al., 2006; Tuckfield et al., 2002) et Sp1/KLF (Brown et al., 2005). À l'exception de Dsp1, la mutation de ces différentes protéines ne conduit pas à des transformations homéotiques contrairement à la mutation des protéines du groupe Polycomb. Mais lorsque ces mutations sont combinées avec celles des protéines du groupe Polycomb les phénotypes observés sont aggravés, que ce soit au niveau des transformations homéotiques (Blastyák et al., 2006; Huang et al., 2002) ou de la létalité (Brown and Kassis, 2010; Déjardin et al., 2005). De façon surprenante, la mutation de certains de ces facteurs de transcription donne des résultats qui peuvent être dépendant du PRE étudié ce qui leur suggère un rôle à la fois dans la répression et l'activation de la transcription (Bejarano and Busturia, 2004; Decoville et al., 2001; Déjardin and Cavalli, 2004).

Les protéines PcG interagissent faiblement avec ces multiples facteurs de transcription, et la mutation d'un seul d'entre eux ne suffit pas à abroger le recrutement des complexes PRC1 et PRC2 montrant leur redondance (Kassis and Brown, 2013). De plus, plusieurs de ces protéines lient le même site (GAF/Psq et Sp1/KLF) et donc ne peuvent être présentes en même temps au même site. C'est donc un mécanisme de recrutement combinatoire, où la liaison de façon spécifique de ces multiples facteurs de transcription vont ensemble contribuer au recrutement initial des complexes Polycomb qui à leur tour, vont pouvoir contribuer à leur propre recrutement via la méthylation et ubiquitination des histones H3K27 et H2AK118 (Figure 1-7a).

C'est donc une attention toute particulière qui a été portée aux motifs présents aux PREs fonctionnellement identifiés afin d'en identifier des caractéristiques communes et de permettre leur prédiction et identification à l'échelle du génome entier.

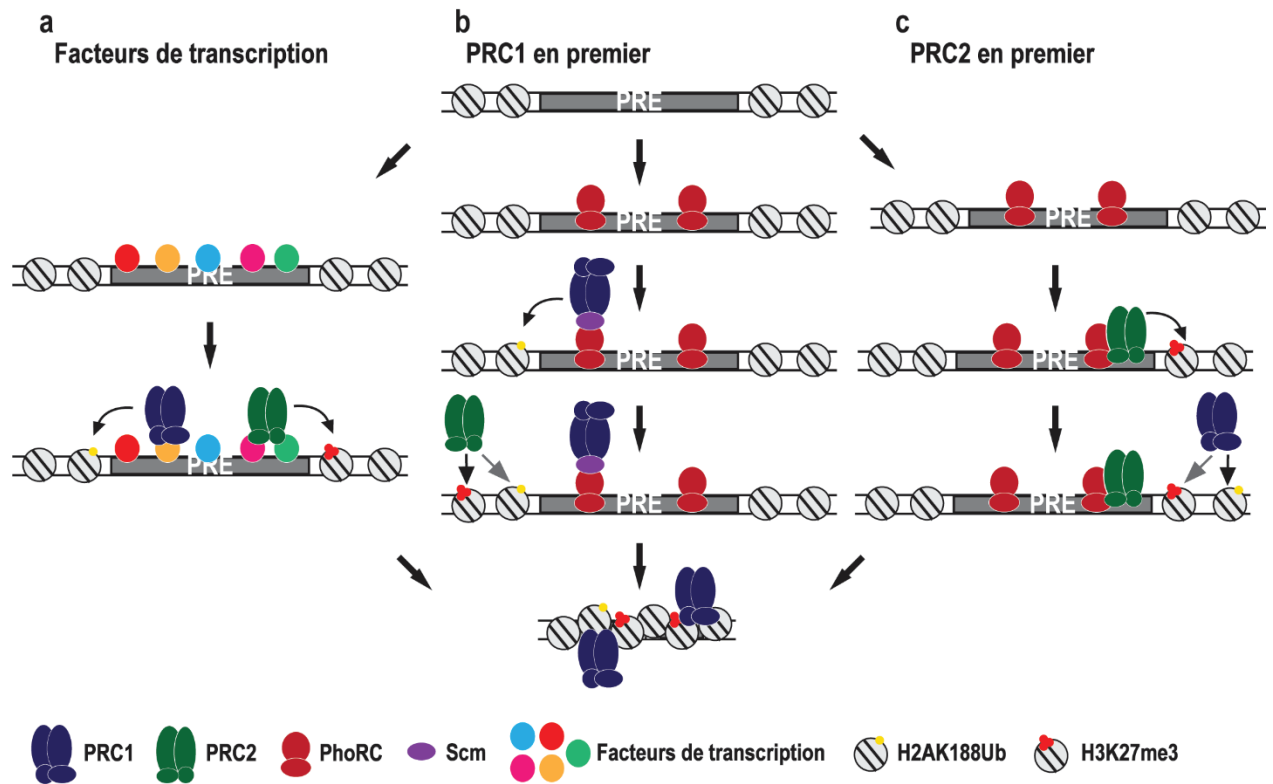


Figure 1-7 Trois modèles illustrant la phase de recrutement initial des complexes Polycomb à un PRE

a. Un PRE est lié de façon séquence spécifique par des facteurs de transcription qui participent au recrutement de PRC1 et PRC2. b. La liaison de PhoRC à son motif présent au niveau des PREs, permet le recrutement de PRC1 grâce à son interaction avec la protéine Scm. PRC1 ubiquitinye la lysine 118 de l'histone H2A, marque reconnue par PRC2, permettant le recrutement de ce dernier. c. La liaison de PhoRC à un PRE permet le recrutement de PRC2 grâce à une interaction entre PRC2 et Pho ou via le domaine SAM de Sfmt. PRC2 tri-méthyle H3K27, marque reconnu par PRC1, permettant son recrutement.

1.3.3.2 Une absence de séquence consensus

Les PREs identifiés fonctionnellement n'ont pas permis de définir une séquence consensus. C'est donc uniquement leur définition fonctionnelle décrite précédemment qui a permis leur identification. En plus des protéines des groupes Polycomb et Trithorax, de nombreux facteurs de transcription décrivent précédemment lient les PREs. C'est donc en utilisant les sites de liaison de ces facteurs de transcriptions que des PREs ont pu être identifiés *in silico*. En définissant un PRE comme une accumulation de sites de liaison de ces facteurs de transcription, qui peuvent être présents en plusieurs exemplaires et n'ont pas besoin de tous être présents pour définir un PRE (Ringrose et al., 2003). Grâce à cette définition, plus d'une centaine de PREs ont pu être identifiés (Ringrose et al., 2003). Depuis, l'identification d'autres protéines liant les PREs et les données de ChIP ont permis le développement de nouveaux outils de prédiction qui ont mis en évidence des propriétés nouvelles des PREs (Bredesen and Rehmsmeier, 2019; Zeng et al.,

2012). Parmi ces caractéristiques le fort contenu en GC est intéressante car contrairement aux facteurs de transcriptions qui sont peu conservés entre *Drosophila melanogaster*, les mammifères et les plantes, cette propriété peut l'être (Brown et al., 1998; Farkas et al., 1994).

Le fait que l'identification des PREs à l'échelle du génome ai été faite sur le même ensemble limité de PREs validés fonctionnellement, qui pour la majorité se situent aux niveaux des gènes Hox, peut avoir caché ou accentué certaines propriétés de leur séquence ADN (Ringrose et al., 2003). De plus ces facteurs de transcription tendent à être exprimé de façon ubiquitaire et se lient à de nombreux loci dans le génome sans pour autant induire le recrutement des protéines du groupe Polycomb (Alhaj Abed et al., 2018; Müller and Kassis, 2006). Ceci suggère donc que même si les facteurs de transcription jouent un rôle dans le recrutement des complexes Polycomb, ils ne peuvent pas à eux seuls expliquer leur recrutement contexte spécifique aux PREs. Des modèles de recrutement alternatifs des complexes Polycomb ne reposant pas sur les facteurs de transcription ont donc été proposés (Figure 1-7b-c) (Dorafshan et al., 2017).

1.3.4 Les PREs ne sont pas tous identiques

1.3.4.1 La localisation des PREs par rapport aux gènes qu'ils régulent

L'identification des PREs dans l'ensemble du génome de *D. melanogaster* a permis de déterminer qu'il n'y a pas une règle concernant leur localisation par rapport aux gènes qu'ils régulent. Certains comme ceux régulant *en* se situe dans la région promotrice de ce dernier, d'autres peuvent être présent dans la séquence codante d'un gène ou encore être distants de plusieurs kilobases de leur cible (Ringrose et al., 2003).

De nombreux gènes sont régulés et/ou se situent à proximité de plusieurs PREs, comme par exemple *en-inv*, les gènes *Hox*, *dac* ou encore *vg* (Kassis and Brown, 2013). En effet, l'identification des PREs par Ringrose *et al.* a permis de mettre en évidence que 90% des PREs se situent à proximité d'autres PREs ou de pics ayant des scores plus faibles mais pouvant être potentiellement des PREs (Ringrose et al., 2003), ce qui posent la question de la fonction de ces multiples éléments régulateurs. Est-ce qu'ils sont redondants et donc la suppression de l'un d'entre eux ne devrait pas affecter l'expression du gène qu'il régule ou ont-ils des fonctions différentes ?

1.3.4.2 Tous les PREs ne sont pas égaux

Les PREs se caractérisent par un recrutement important des complexes PRC1 et PRC2 qui peuvent être observé par CHIP-seq, mais il y a de nombreux autres pics de moindre importance dans les domaines Polycomb (Nègre et al., 2006; Schwartz et al., 2006; Tolhuis et al., 2006). La suppression des PREs des gènes *en* et *inv* ou encore de ceux du gène *vg* qui lorsqu'ils sont présents conduisent à un recrutement important des protéines du groupe Polycomb, n'affecte pas leur expression (De et al., 2016). En l'absence de PREs identifiés fonctionnellement ou *in silico*, les sites de liaison des protéines du groupe Polycomb, qui jusqu'alors n'étaient pas identifiés comme PREs, forment en réalité des PREs plus faibles. Le fait que les algorithmes bio-informatiques n'aient pas permis l'identification de ces PREs plus faibles montre bien que les motifs des facteurs de transcription et autres protéines liant les PREs ainsi que les répétitions GT ne sont pas suffisantes pour définir un PRE. Il est possible qu'il y ait différentes classes de PREs qui utilisent des mécanismes de recrutement différents avec les PREs plus forts reposant sur les facteurs de transcription et les PREs plus faibles sur une autre caractéristique pas encore identifiée (Figure 1-7).

Un autre exemple que tous les PREs ne sont pas strictement identiques et que leur moyen pour recruter les protéines du groupe Polycomb peuvent différer, provient de l'étude du gène *vg* (Ahmad and Spens, 2019). La suppression du PRE proximal de *vg* n'altère pas le recrutement des protéines du groupe Polycomb ni l'établissement du domaine Polycomb répressif, mais empêche l'expression appropriée du gène *vg* dans les disques imaginaux et les ailes où elle devrait se faire. À l'inverse la suppression du PRE distal de *vg* conduit à une réduction importante de la tri-méthylation de la lysine 27 de l'histone 3 dans le domaine Polycomb montrant que l'un des PREs joue un rôle plus important dans la répression et l'autre dans l'activation, expliquant peut-être ainsi pourquoi certains gènes sont régulés ou se situent à proximité de plusieurs PREs (Ahmad and Spens, 2019; De et al., 2016; Ringrose et al., 2003).

1.3.5 Les PREs sont transcrits en ARN

En dehors des PREs qui se situent dans des gènes, d'autres PREs peuvent être transcrits en ARN non codants et, une attention particulière a été portée à la transcription afin de déterminer si elle pouvait être à l'origine du changement de statut d'un PRE.

La transcription à travers le PRE *bx-d* ne suffit pas à le faire passer d'un état répressif à permissif. En forçant la transcription à travers un PRE grâce au système inductible UAS-GAL4, il y a une réduction du recrutement des protéines du groupe Polycomb, la liaison de celle du groupe Trithorax et dérégulation du

gène rapporteur tant que la transcription est induite (Erokhin et al., 2015). Mais à la suite de l'arrêt de la transcription au travers du PRE, les complexes Polycomb sont à nouveau recrutés et la répression s'installe montrant que la transcription en elle-même ne suffit pas à faire permuter durablement un PRE d'un état répressif à actif. Ces résultats sont en contradiction avec les observations faites avec le PRE *Fab-7*, pour lequel la transcription du PRE pendant le stade embryonnaire, mais pas durant le stade larvaire, conduit à la dérégulation du gène rapporteur (Cavalli and Paro, 1998). Ceci montre encore une fois que tous les PREs ne fonctionnent pas de la même façon et que la transcription ou les ARN eux-mêmes peuvent être importants pour moduler l'activité de certains PREs. Il ne faut toutefois pas exclure la possibilité que les différences observées entre ces deux études viennent du site d'insertion du transgène, puisqu'il est connu que l'environnement autour d'un PRE puisse influencer sa fonction (De et al., 2019; Okulski et al., 2011).

En étudiant l'un des PREs du gène *vg*, Herzog et al. ont mis en évidence que ce PRE est transcrit en ARN non codant dans les deux orientations, et que l'ARNnc antisens corrèle avec un gène *vg* réprimé et l'ARNnc sens avec l'expression de *vg* (Herzog et al., 2014). Ceci suggère que ce n'est peut-être la transcription mais bien l'ARN qui est à l'origine de la différence de statut du PRE. De plus, en regardant dans l'ensemble du génome, ils ont mis en évidence que ce PRE du gène *vg* n'est pas le seul à être transcrit bidirectionnellement ce qui suggère un rôle bien plus important pour les ARN et la transcription aux PREs que celui qui leur avaient été attribué jusqu'à présent.

1.3.6 Les éléments de réponse Polycomb chez les autres espèces

Malgré leur découverte chez *D. melanogaster* il y a plus de 20 ans, les PREs ont été difficiles à identifier chez d'autres espèces en raison d'une absence de séquence consensus simple. De plus l'algorithme qui a permis de l'identification de certaines PREs chez *D. melanogaster* repose sur les motifs de liaison de facteurs de transcription qui sont bien moins conservés que les protéines du groupe Polycomb. Seuls quelques PREs ont été identifiés fonctionnellement chez les plantes et mammifères (Bauer et al., 2015; Buzas et al., 2012; Kassis and Brown, 2013).

1.3.6.1 Les PREs chez les mammifères

L'étude de la régulation des gènes *Hox* chez *D. melanogaster* a permis l'identification de PREs. Les gènes *Hox* sont également présents chez les mammifères chez lesquels les clusters au sein desquels ils se regroupent ont été dupliqués. Mais malgré la conservation des protéines du groupe Polycomb et de certains de leurs gènes cibles entre *D. melanogaster* et les mammifères, les PREs ne semblent pas avoir

été conservés de la même manière (Bauer et al., 2015; Kassis and Brown, 2013). Les sites de liaison des facteurs de transcription et protéines qui ont permis l'identification des PREs à l'échelle du génome entier chez *D. melanogaster* sont peu conservés à l'exception de GAF et de Pho homologue de YY1 chez les mammifères (Brown et al., 1998; Farkas et al., 1994). D'autres facteurs de transcription tels que REST sont impliqués dans le recrutement des protéines du groupe Polycomb chez les mammifères, mais peu ont été identifiés jusqu'à maintenant freinant l'identification des PREs (Dietrich et al., 2012). Il existe toutefois la possibilité que les PREs chez les mammifères obéissent à d'autres règles et ne nécessitent pas la présence de motifs pour la liaison de facteurs de transcription comme ceux de *D. melanogaster*.

1.3.6.1.1 Les PREs identifiés de manière fonctionnelle

Quelques PREs ont pu être identifiés fonctionnellement chez les mammifères en étudiant l'expression d'un gène rapporteur sous contrôle de ce PRE potentiel chez *D. melanogaster* (Sing et al., 2009; Vasanthi et al., 2013; Woo et al., 2010). Tout comme les PREs de *D. melanogaster*, ces PREs conduisent à la répression d'un gène rapporteur dépendamment des protéines Polycomb chez *D. melanogaster* et sont nécessaires pour la bonne expression des gènes qu'ils régulent chez les vertébrés. Mais l'utilisation de transgènes chez *D. melanogaster* limite l'identification des PREs chez les mammifères, à ceux qui sont semblables à ceux de *D. melanogaster* et qui utilisent notamment des facteurs conservés entre les espèces pour assurer le recrutement des protéines du groupe du groupe Polycomb.

1.3.6.1.2 Les îlots CpG

Les motifs servant à prédire les PREs chez *D. melanogaster* ne sont pas enrichis aux sites de liaisons des protéines du groupe Polycomb chez les mammifères. Mais ces protéines sont particulièrement présentes au niveau des îlots CpG.

Les îlots CpG sont caractérisés par un plus fort enrichissement en guanines et cytosines par rapport au reste du génome, ainsi que par la présence de répétitions CG. Ils ont tendance à être situés dans les régions promotrices de gènes (Illingworth et al., 2010). Contrairement à la *D. melanogaster* dont l'ADN ne présente que peu de méthylation, l'ADN des mammifères est méthylé. Mais les îlots CpG aux promoteurs présentent peu de méthylation de l'ADN. Ces îlot CpG sont marqués par la présence de H3K4me3 et H3K27me3, modifications post-traductionnelles d'histones respectivement actives et répressives, dans les cellules souches embryonnaires (Bernstein et al., 2006). Mais dans les cellules différenciées, ces domaines bivalents deviennent monovalents et sont soit associés à H3K27me3 et aux protéines du groupe Polycomb, soit à H3K4me3 et aux protéines du groupe Trithorax (Bernstein et al., 2006; Hu et al., 2013). Tout comme les PREs, les îlots CpG au niveau des promoteurs sont donc des régions bivalentes qui lors de la

différenciation deviennent réprimées ou restent actives, c'est pour ça qu'ils ont été pensés comme étant l'équivalent des PREs chez les vertébrés (Figure 1-8a).

1.3.6.1.3 Des ARN pour recruter les protéines du groupe Polycomb chez les vertébrés

Les îlots CpG au niveau des promoteurs ne sont pas les seuls sites de liaison des protéines du groupe Polycomb. En effet, ces protéines sont recrutées au niveau de régions intergéniques qui sont transcrites en ARN non codants (ARNnc) et qui ne contiennent pas d'îlot CpG (Dietrich et al., 2012; Hekimoglu-Balkan et al., 2012). Il a été proposé que la transcription de ces sites en ARNnc participerait ou préviendrait le recrutement des protéines du groupe Polycomb (voir la section 1.2.2.6).

Xist et l'inactivation du chromosome X

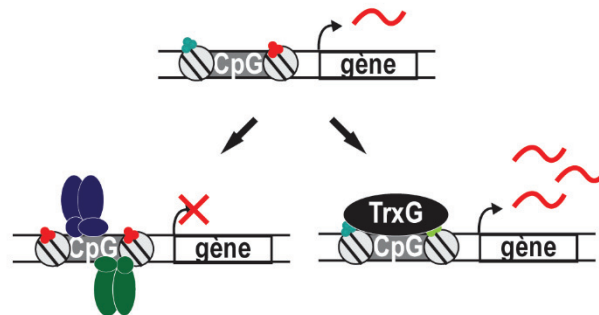
Chez les mammifères, les femelles possèdent deux chromosomes X et doivent en réprimer un à travers un mécanisme de compensation de dosage afin d'éviter une trop forte expression des gènes compris sur ce chromosome (Brockdorff, 2017; Monfort and Wutz, 2020). L'ARNnc *Xist* s'exprime à partir des chromosomes X et va former un nuage autour de ces derniers (Figure 1-8b) (Chaumeil et al., 2006; Kay et al., 1993). Après rapprochement des deux chromosomes X, le nuage d'ARN devient plus important à un des deux chromosome, entraînant sa répression notamment en permettant le recrutement de PRC2 et de la version non-canonique du complexe PRC1, en plus d'autres protéines telles que HDAC participant à l'inhibition de la transcription (Almeida et al., 2017; da Rocha et al., 2014; McHugh et al., 2015). Recruté au niveau du chromosome X, PRC2 et la version non-canonique de PRC1 peuvent modifier les histones de manière post-traductionnelle (H3K27me3 et H2AK119ub respectivement). Ces marques épigénétiques participent au recrutement d'autres complexes PRC2 et de la version canonique du complexe PRC1 qui se propagent alors du locus Xi sur l'ensemble du chromosome X conduisant à la répression stable de ce dernier.

HOTAIR et la répression des gènes Hox

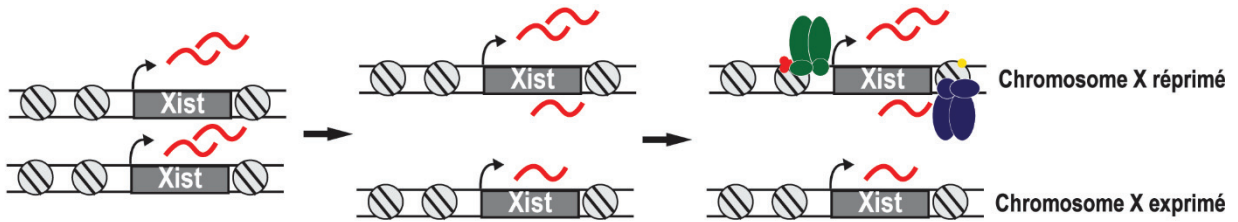
HOTAIR est un ARNnc transcrit entre les gènes *HoxC11* et *HoxC12*, et qui est conservé chez les mammifères (Rinn et al., 2007). La délétion de HOTAIR conduit à un mauvais recrutement de PRC2, une dérégulation du cluster *HoxD* et à des transformations homéotiques chez la souris (Gupta et al., 2010; Li et al., 2013; Tsai et al., 2010). PRC2 en liant directement HOTAIR va pouvoir être recruté spécifiquement au niveau du locus *HoxD*, ce qui explique la répression plus importante des gènes *Hox* observée en cas de surexpression de HOTAIR (Figure 1-8c) (Gupta et al., 2010; Li et al., 2013; Rinn et al., 2007). Ceci est un exemple saisissant d'ARNnc agissant en trans afin de recruter les protéines Polycomb à un locus spécifique. Mais la conservation de ce mécanisme chez la souris été remis partiellement en question par le groupe de Denis

Duboule qui ont présenté des résultats contradictoires avec ceux du laboratoire de Howard Chang (Amândio et al., 2016; Schorderet and Duboule, 2011).

a
Les îlots CpG



b
Xist et l'inactivation du chromosome X



c
HOTAIR et la répression des gènes Hox

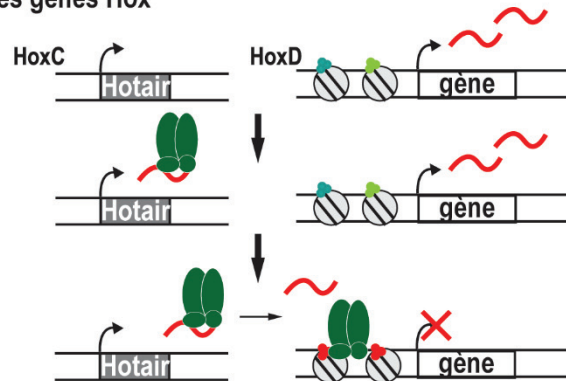


Figure 1-8 Exemples d'éléments pouvant jouer le rôle de PRE chez les mammifères

a. Les îlots CpG. Les îlots CpG non méthylés peuvent être bivalents et marqués par H3K27me3 et H3K4me3 (haut). Le recrutement des protéines du groupe Polycomb (gauche) ou Trithorax (droite) permet le changement rapide des modifications post-traductionnelles présentes à cet îlot conduisant respectivement, à la répression ou à l'expression. b. L'ARN non-codant Xist est exprimé à partir du chromosome X et forme un nuage autour de ce chromosome. Après l'appariement des chromosomes (gauche), ce nuage devient plus important au niveau de l'un des chromosomes X (milieu) conduisant au recrutement de PRC2 et à la

répression stable de ce chromosome X (droite). c. L'ARN non-codant HOTAIR transcrit à partir du locus HoxC, lie PRC2. HOTAIR participe au recrutement de PRC2 en trans au locus HoxD afin d'induire la répression de ce dernier.

1.3.6.2 Les PREs chez les plantes

1.3.6.2.1 FLC et la vernalisation

Les plantes, telles que *Arabidopsis thaliana*, afin d'entrer en floraison uniquement après l'hiver doivent réprimer l'expression du gène *Flowering Locus C (FLC)* (Costa and Dean, 2019). Cette répression démarre progressivement lors de l'exposition de la plante au froid, et devient une répression stable qui implique les protéines du groupe Polycomb après une période de froid prolongée (Kim and Sung, 2012; Song et al., 2012). Au niveau épigénétique, cette répression se traduit par un changement des modifications post-traductionnelles des histones : de H3K4me3 et H3K36me3 à H3K27me3 (Yang et al., 2014). La vernalisation et donc la répression de *FLC*, un répresseur, permettent l'expression de gènes impliqués dans la floraison.

Comme de nombreux PREs chez *D. melanogaster*, le locus *FLC* est transcrit en ARNnc sens et antisens (Swiezewski et al., 2009). Ces ARNnc sens et antisens s'expriment à des moments différents lors de la vernalisation et ont des rôles distinct dans la mise en place de cette répression progressive de *FLC* (Figure 1-9) (Kim and Sung, 2017).

COOLAIR, long ARNnc antisens forme une R-loop dans sa partie promotrice qui empêche sa propre transcription, et surtout la production de sa forme la plus longue qui s'étendrait à travers le gène *FLC* (Sun et al., 2013). En cas d'exposition prolongée au froid, COOLAIR commence à s'exprimer de manière plus importante et son expression augmente de manière graduelle au fur et à mesure que le froid s'installe pour atteindre un maximum d'expression au bout de deux semaines (Rosa et al., 2016). L'expression de COOLAIR, notamment de sa forme longue, et de *FLC* sont mutuellement exclusives à un même locus (Rosa et al., 2016). L'augmentation de l'expression de COOLAIR conduit donc à la diminution de celle de *FLC*.

COLDAIR, ARNnc sens transcrit à partir du premier intron de *FLC*, s'exprime au bout du 20^{aine} de jours (Heo and Sung, 2011). Il est essentiel pour l'établissement de la répression stable de *FLC* qui implique les protéines du groupe Polycomb afin d'éviter un délai dans la floraison (Heo and Sung, 2011). COLDAIR interagit avec le locus *FLC* (mais son moyen d'interaction n'a pas été défini) et avec PRC2 directement afin de permettre le recrutement de ce dernier (Kim et al., 2017). Le site d'interaction de COLDAIR avec *FLC* sert de site de nucléation initiale pour PRC2 qui peut méthyler la lysine 27 de l'histone 3. PRC2 peut alors se propager le long de la chromatine afin d'établir une répression stable du gène *FLC*. De façon

intéressante, la transcription de COLDAIR n'est pas requise mais c'est l'ARN lui-même qui est nécessaire pour l'établissement d'une répression stable (Kim et al., 2017).

Plus récemment, un autre ARNnc sens transcrit dans la région promotrice de *FLC* a montré pouvoir partiellement compenser l'absence de COLDAIR. L'expression de COLDWRAP, plus tardive permet la floraison et donc la répression de *FLC* mais plus tardivement. Comme COLDAIR, COLDWRAP interagit avec PRC2, l'ARN lui-même et non pas la transcription est impliqué dans le recrutement de PRC2 au locus *FLC*, et COLDWRAP est nécessaire pour la répression de *FLC* afin d'éviter un retard de la floraison (Kim and Sung, 2017). Le promoteur du gène *FLC* et la partie 3' de son premier intron forment une boucle détectable avec une expérience de 3C qui dépend de la présence de COLDAIR, de COLDWRAP et de PRC2 (Kim and Sung, 2017). Il a été suggéré que cette boucle permette une meilleur propagation de la répression induite par PRC2, ainsi que de prévenir du contact du promoteur de *FLC* avec sa région terminale qui est observée lorsque *FLC* est exprimé (Crevillén et al., 2013).

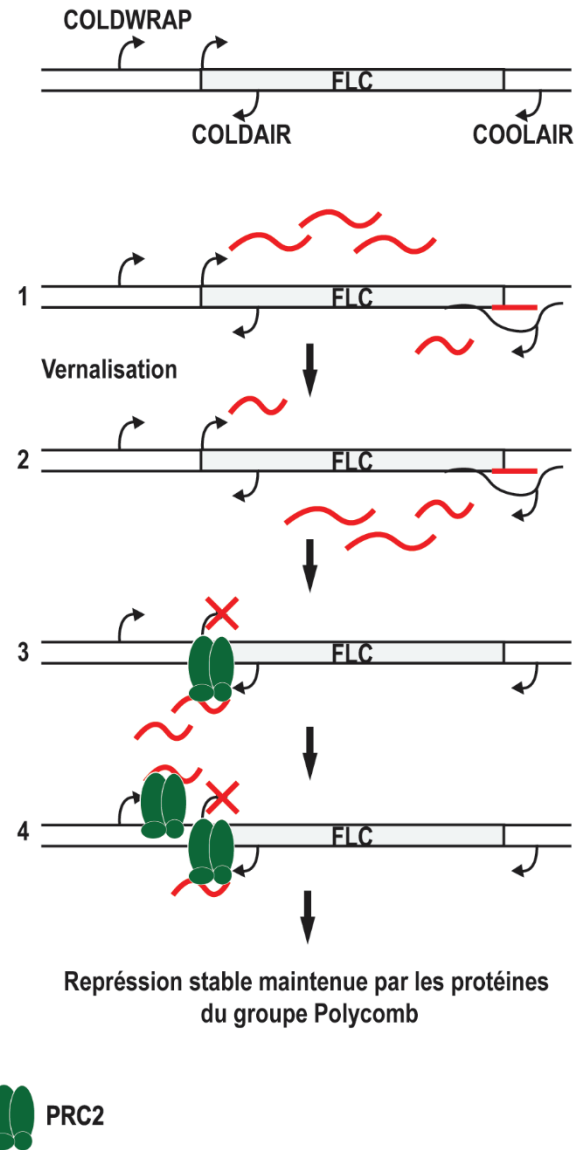


Figure 1-9 Schéma de la répression du locus FLC lors de la vernalisation impliquant des ARNnc et PRC2

En haut, schéma du locus FLC. (1) Avant vernalisation FLC est exprimé ainsi que COOLAIR dans sa forme courte formant une R-loop. (2) À la suite de la vernalisation, l'expression de COOLAIR augmente et sa forme longue est produite, réduisant l'expression de FLC. (3) L'expression de COLDAIR permet le recrutement de PRC2 et la répression stable de FLC. (4) L'expression de COLDWRAP renforce le recrutement de PRC2 et la répression de FLC.

Les exemples de PREs ou de séquences d'ADN jouant un rôle dans le recrutement des protéines du groupe Polycomb chez les autres espèces ont mis en avant le rôle prépondérant des ARNnc transcrits au niveau de ces séquences. On peut alors se demander si chez *D. melanogaster*, les ARNnc aux PREs pourraient jouer un rôle tout aussi majeur dans l'établissement de la répression et si, toute une classe de PREs n'a

pas été ignorée en raison d'une importance moindre des facteurs de transcription à ces sites et d'un rôle plus prépondérants des ARN.

1.4 R-loops

Les R-loops sont des structures d'ADN non-canoniques qui se forment lorsqu'un ARN s'hybride avec un simple-brin d'ADN entraînant le déplacement du second brin d'ADN (Figure 1-10a). Structurellement un hybride ARN-ADN se distingue d'un ADN double-brin en n'adoptant pas une forme B. L'hybride ARN-ADN adopte une structure plus proche d'une forme A alors que l'ADN double-brin de chaque côté de la R-loop adopte une forme B (Shaw and Arya, 2008; Wang et al., 1982). La forme A se distingue de la B par un grand sillon plus étroit et un petit sillon plus large (Kulkarni and Mukherjee, 2017). Les hybrides ARN-ADN sont plus stables que leur équivalent ADN-ADN (Lesnik and Freier, 1995). Ces différences structurelles et l'exposition d'un ADN simple-brin font que les R-loops peuvent être la cible de protéines pouvant entraîner l'apparition de cassures dans l'ADN et être la cause d'instabilité génomique.

1.4.1 Les R-loops sont source d'instabilité génomique

Les R-loops ont été décrites comme source d'instabilité génomique car elles exposent un simple-brin d'ADN qui peut devenir la cible de protéines (Brambati et al., 2020). C'est le cas notamment, lors de la commutation de classe au niveau du locus immunoglobuline où le simple-brin d'ADN exposé devient la cible préférentielle de la protéine AID qui va pouvoir désaminer des cytosines (Chaudhuri et al., 2003; Ramiro et al., 2003; Yu et al., 2003). Ces uraciles dans l'ADN et les mésappariements qu'elles génèrent sont alors reconnues, entraînant des cassures dans l'ADN qui sont réparées par le mécanisme de réparation de jonction des extrémités non-homologues qui conduit à la commutation des classes (Nicolas et al., 2018).

Les R-loops en formant des hybrides ARN-ADN empêchent le passage des ARN et ADN polymérases, entraînant une pause dans la progression de la fourche de réplication ou de l'ARN polymérase ce qui peut être délétère pour la cellule (Chen et al., 2017; Gan et al., 2011). Il est toutefois à noter, qu'il n'est pas encore compris si la formation de R-loops sont la cause ou la conséquence de l'arrêt de la polymérase (Chen et al., 2017).

Les R-loops sont également la cible d'endonucléases ne présentant pas une spécificité de séquence mais une spécificité de structure, qui peuvent reconnaître l'ADN simple-brin ou la jonction entre l'ADN double-

brin et l'ADN simple-brin, ainsi que potentiellement celle entre l'ADN double-brin et l'hybride ARN-ADN (Cloud et al., 1995; Dehé and Gaillard, 2017; Laa et al., 1998).

1.4.2 Les facteurs contribuant à la formation des R-loops

1.4.2.1 La séquence de l'ADN et de l'ARN

Les R-loops identifiées au locus immunoglobuline ont mis en évidence une préférence pour les séquences riches en GC avec un hybride ARN-ADN qui se compose d'un ARN enrichi en G et un ADN en C (Figure 1-10b) (Hegazy et al., 2020; Huang et al., 2006; Yu et al., 2003). Ce biais pour les séquences riches en GC, avec un ARN plus enrichi en G a depuis été confirmé chez les mammifères, les plantes et la levure *Schizosaccharomyces pombe* grâce au développement d'approches à l'échelle du génome entier (Hartono et al., 2018; Sanz et al., 2016; Xu et al., 2017). Ce biais pour les séquences riches en G s'explique par la nécessité de deux clusters de G afin de permettre l'initiation d'une R-loop ainsi qu'une forte densité de G pour son élongation (Roy and Lieber, 2009).

La levure *Saccharomyces cerevisiae* fait figure d'exception puisque dans cette espèce les R-loops se forment préférentiellement aux séquences riches en AT avec l'un des brins plus fortement enrichi en A, ainsi qu'aux sites polyA (Wahba et al., 2016). Cette différence entre les R-loops présentes chez *Saccharomyces cerevisiae* et celles des autres espèces vient peut-être de la forte densité de gènes dans ce génome. En effet *Saccharomyces cerevisiae* ne contient que très peu de séquences non codantes avec peu de régions intergéniques et des exons courts, suggérant que dans ce cas ce n'est pas la séquence d'ADN qui est à l'origine de la formation de R-loop mais peut-être la topologie de l'ADN. Les séquences formant des R-loops chez *Arabidopsis thaliana* présentent les caractéristiques des mammifères et des levures, avec un brin d'ADN à la fois enrichi en G et en A (Xu et al., 2017).

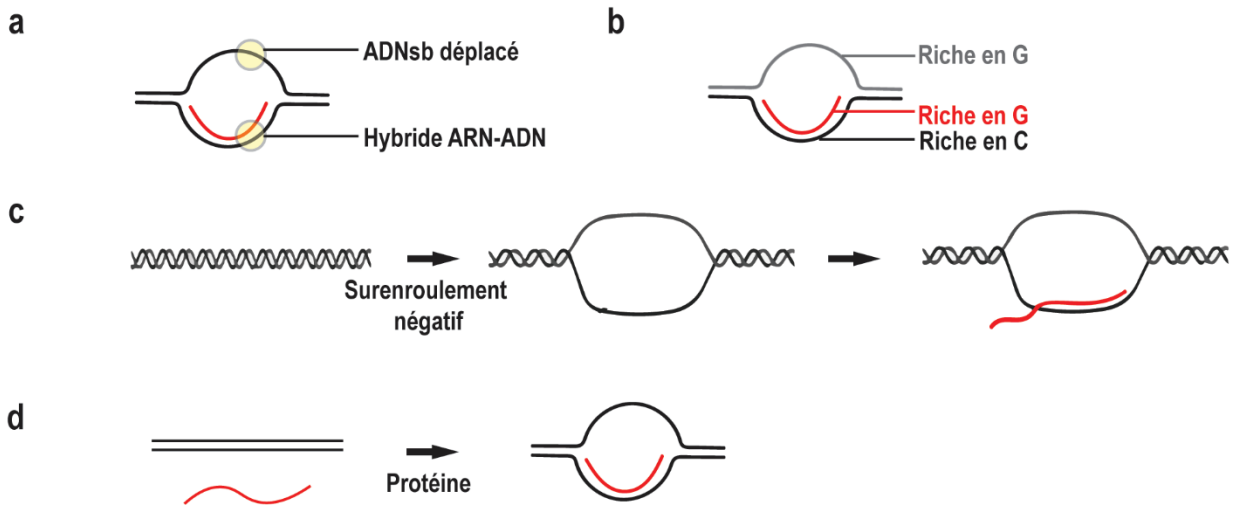


Figure 1-10 Les facteurs contribuant à la formation d'une R-loop

a. Schéma d'une R-loop. b. Un ARN enrichi en G favorise la formation de R-loops. c. L'accumulation de surenroulement négatif favorise l'ouverture de l'hélice d'ADN permettant l'hybridation d'un ARN complémentaire. d. Des protéines peuvent induire la formation d'une R-loop.

1.4.2.2 La topologie de l'ADN

Comme décrit plus en détail dans la section sur la topologie, les génomes des eucaryotes et procaryotes sont surenroulés, et ce surenroulement est sensible à tout évènement pouvant entraîner des distorsions de l'ADN (voir la section 1.5). Lors de la transcription le passage de l'ARN polymérase entraîne une accumulation de surenroulement négatif derrière la polymérase qui ne peut être complètement absorbé en raison notamment de la présence de protéines telles que CTCF qui contraignent la topologie de l'ADN (Naughton et al., 2013). En cas de surenroulement négatif extrême, la double hélice d'ADN va avoir tendance à s'ouvrir afin de dissiper ce stress, favorisant ainsi l'appariement de l'ARN nouvellement produit avec son ADN complémentaire formant ainsi une R-loop (Figure 1-10c). Les R-loops permettent donc ainsi d'absorber des supertours négatifs.

Cet équilibre entre la séquence et la topologie de l'ADN a été modélisé afin de déterminer quelles sont les propriétés nécessaires à la formation de R-loops (Stolz et al., 2019). Ce modèle, confirmé par des observations *in vitro*, a mis en évidence qu'une séquence plus favorable à la formation d'une R-loop nécessite moins d'énergie provenant de la relaxation de surenroulement négatif mais que la présence d'un surenroulement négatif plus important permet la formation de R-loops au niveau de séquences peu favorables (Stolz et al., 2019).

1.4.2.3 Les protéines induisant la formation de R-loops

Les R-loops peuvent se former en *cis* de manière co-transcriptionnelle, lorsque le passage de l'ARN polymérase génère un contexte favorable pour leur formation. Elles peuvent également se former en *trans* notamment avec l'aide de protéines telles que Rad52, qui peut induire l'invasion d'ADN double-brin par un ARN, conduisant à l'hybridation de l'ARN avec un ADN complémentaire (Figure 1-10d) (Mazina et al., 2017). *In vitro*, l'activité d'invasion de brins de Rad52 est stimulée par la présence de RPA, protéine qui lie l'ADN simple-brin. Cette stimulation se fait via une interaction directe entre RPA et Rad52 (Mazina et al., 2017). *In vivo*, il est possible d'imaginer que RPA stimule Rad52 en interagissant avec cette dernière mais aussi en favorisant la formation de l'hybride ARN-ADN via la liaison du simple-brin d'ADN déplacé (Kim et al., 1992).

Chez les bactéries et archées, les protéines Cas ont permis le développement d'outils qui permettent l'édition du génome dont le plus connu est le système CRISP-Cas9. La protéine Cas9 lie un ARN guide, lorsque Cas9 trouve une région d'homologie pour cet ARN accompagnée d'un site PAM dans l'ADN, celle-ci va pouvoir ouvrir localement la double hélice d'ADN ce qui permet l'hybridation de l'ARN avec son ADN complémentaire entraînant la formation d'une R-loop (Sternberg et al., 2014). La complémentarité entre l'ARN et l'ADN permet l'agrandissement de cette R-loop (Spies, 2014). Cas9 va pouvoir alors cliver l'ADN, cette cassure dans l'ADN pourra alors être prise en charge par les divers mécanismes de réparation de cassures dans l'ADN et notamment par le système de recombinaison homologue en cas de séquence complémentaire présente permettant ainsi l'édition du génome. L'utilisation de Cas9 incapable de couper l'ADN mais fusionner avec différents facteurs de transcription ou répresseur ont permis de nombreuses études sur les fonctions de gènes *in vivo* (Maeder et al., 2013; Qi et al., 2013). Que la Cas9 soit active ou non, elle repose toujours sur la formation de cette R-loop qui permet le recrutement ciblé de cette enzyme. Tout comme les R-loops formées indépendamment, celles formées par Cas9 dépendent également du niveau de surenroulement présent dans l'ADN. La présence d'ADN surenroulé permet à la protéine Cas9 de former des R-loop avec des mésappariements présents entre l'ARN guide et l'ADN (Ivanov et al., 2020).

La présence de protéines telles que Cas9, mais également de Cas12a ou encore CASCADE-Cas3 qui ciblent également l'ADN chez les archées et bactéries, et de Rad52 chez les eucaryotes, qui peuvent induire la formation de R-loops suggèrent un rôle bien plus important pour les R-loops qui ne leur avait été suggéré pendant de nombreuses années (Singh et al., 2018; Szczelkun et al., 2014). En effet, pendant longtemps les R-loops ont été perçues uniquement comme délétères pour le génome et étant cause d'instabilité

génomique c'est pourquoi une attention toute particulière a été portée aux mécanismes permettant leur résolution (Hegazy et al., 2020).

1.4.3 La résolution des R-loops

1.4.3.1 Les RNase H

1.4.3.1.1 La RNase H1

La RNase H1 est une enzyme monomérique qui résout les R-loops en dégradant la partie ARN de l'hybride ARN-ADN (Figure 1-11a) (Hyjek et al., 2019). Le domaine HBD de la RNase H1 reconnaît préférentiellement les hybrides ARN-ADN en établissant des liaisons hydrogènes avec le 2' OH de deux riboses et en interagissant avec un phosphate de trois désoxyriboses (Nowotny et al., 2008). Le domaine catalytique interagit avec l'hybride ARN-ADN et en présence d'ions métalliques va pouvoir cliver un phosphate entre deux ribonucléotides avant de le relâcher (Nowotny et al., 2007; Yang et al., 2006). Les domaines HBD et catalytique établissent tous deux la liaison à l'hybride ARN-ADN et sont séparés par plusieurs dizaines de résidus, expliquant donc pourquoi la RNase H1 ne peut dégrader les hybrides ARN-ADN que s'ils se forment sur plusieurs nucléotides consécutivement. La RNase H1 possède également une séquence de localisation aux mitochondries et la RNase H1 est essentiel lors de la réplication du génome mitochondrial (Cerritelli et al., 2003; Holt, 2019).

1.4.3.1.2 La RNase H2

La RNase H2 se compose de trois protéines : la RNase H2A qui est la sous-unité catalytique, et la RNase H2B et H2C qui régulent l'activité et la processivité de cette première (Chon et al., 2009; Hyjek et al., 2019). La RNase H2 contrairement à la RNase H1 peut cliver de petits hybrides ARN-ADN et est également processive en présence d'hybrides ARN-ADN plus longs (Figure 1-11a) (Chon et al., 2009). De plus, la RNase H2 est essentielle pour exciser des ribonucléotides ayant été incorporés dans l'ADN à la place de désoxyribonucléotides, ainsi que pour dégrader les fragments d'Okazaki présents en fin de phase de réplication (Reijns et al., 2012; Sparks et al., 2012). La RNase H2 clive le ribonucléotide permettant à l'ADN polymérase de placer un désoxyribonucléotide à la place, avant que des ligases ne liguent les deux fragments d'ADN (Sparks et al., 2012). Ce rôle dans l'excision et la réparation de ribonucléotides est certainement lié à la capacité de la RNase H2B à interagir avec PCNA même si, *in vivo* et *in vitro*, le motif PIP qui permet cette interaction n'apparaît pas nécessaire (Chon et al., 2009, 2013; Sparks et al., 2012).

1.4.3.1.3 La RNase H1 et la RNase H2 ont des fonctions distinctes

L'utilisation de levures dont le cycle cellulaire a été synchronisé, et le suivi de la RNase H1 et des différentes sous-unités de la RNase H2 a permis de mettre en évidence le rôle distinct de ces deux enzymes pouvant dégrader les hybrides ARN-ADN (Lockhart et al., 2019). La RNase H2, dont l'expression de la sous-unité catalytique augmente en fin de réplication, est nécessaire afin de dégrader les R-loops présentes après la réplication de l'ADN, et participe à l'éviction de ribonucléotides ayant été incorporés dans l'ADN (Lockhart et al., 2019). La RNase H1 dont l'expression est stable tout au long du cycle cellulaire, n'est présente qu'en faible quantité sur la chromatine en condition normale, mais la fraction de RNase H1 liée à la chromatine augmente lors de stress induisant l'accumulation de R-loops (Lockhart et al., 2019).

La RNase H2 est donc la RNase H principale, ayant un rôle essentiellement lors de la réplication de l'ADN alors que la RNase H1 sert uniquement en cas de stress lors d'accumulation de R-loops (Lockhart et al., 2019).

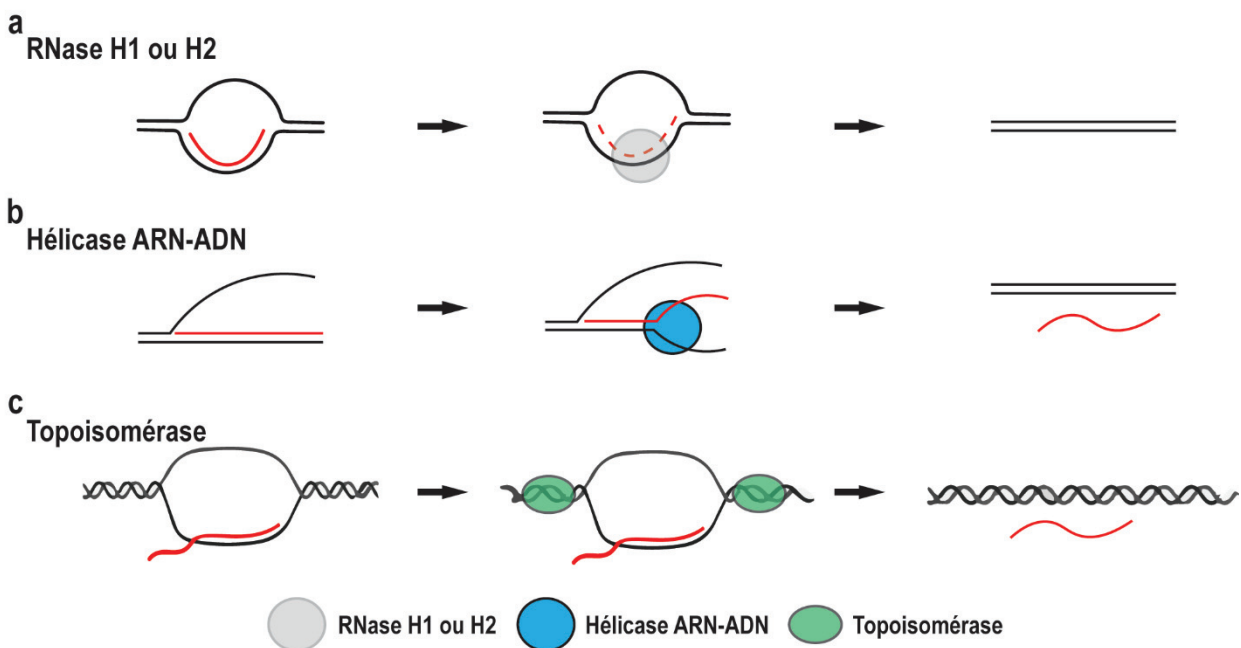


Figure 1-11 Les protéines participant à la résolution des R-loops

a. La RNase H1 ou H2 lie la partie hybride ARN-ADN d'un R-loop et dégrade l'ARN. b. Les hélicase ARN-ADN séparent les parties ARN et ADN de l'hybride, permettant l'hybridation du brin d'ADN qui était déplacé. c. Les topoisomérases réduisent le niveau de surenroulement présent au niveau de l'ADN de chaque côté de la R-loop, déstabilisant cette dernière.

1.4.3.2 Les hélicases ARN-ADN

Sen1, l'homologue de senataxin, est une hélicase pouvant résoudre les hybrides ARN-ADN et son absence chez la levure conduit à l'accumulation de R-loops et de cassures dans l'ADN (Figure 1-11b) (Kim et al., 1999; Mischo et al., 2011). Plus globalement la mutation de Sen1 conduit à l'augmentation de R-loops au niveau des gènes codants pour les ARN de transfert, au niveau de petits gènes codants pour des protéines ainsi qu'aux petits ARN nucléolaires (Chan et al., 2014). Chez l'humain le rôle de senataxin dans la résolution des hybrides ARN-ADN est moins clair, puisque sa déplétion conduit à une augmentation des R-loops après les sites de polyadénylation mais à leur diminution le long des gènes (Skourti-Stathaki et al., 2011). Cette diminution des R-loops le long des gènes peut s'expliquer par la plus faible densité d'ARN polymérase et la diminution de la transcription. Les R-loops et ARN polymérase s'accumulent après les sites de polyadénylation en absence de senataxin suggèrent que l'hélicase est impliquée dans leur résolution et que cette résolution est importante pour la terminaison de la transcription (Skourti-Stathaki et al., 2011).

Une seconde hélicase, aquarius a également été impliquée dans la résolution des R-loops. La déplétion de aquarius entraîne l'accumulation de R-loops et de dommages dans l'ADN qui peuvent être résolus par la surexpression de la RNaseH (Sollier et al., 2014).

In vitro, l'hélicase DHX9 peut résoudre les R-loops mais aussi les quadruplexes de G formés sur un ARN ou sur un ADN (Chakraborty and Grosse, 2011). *In vivo*, les résultats sont contradictoires. Cristini et al. ont observé que la déplétion de DHX9 conduit à une augmentation des R-loops après le site de polyadénylation comme cela avait été observé lors de la déplétion de la senataxin (Cristini et al., 2018). À l'opposé, Chakraborty et al. ont rapporté une diminution du niveau des R-loops à quelques gènes lors de la déplétion de DHX9 (Chakraborty et al., 2018). Le rôle de DHX9 *in vivo* n'est donc toujours défini. Mais ces contradictions peuvent en partie s'expliquer par les observations faites *in vitro* : l'activité d'hélicase des hybrides ARN-ADN de DHX9 va entraîner la résolution des R-loops tandis que sa capacité à défaire les quadruplexes de G sur un ARN pourrait le rendre disponible pour former une R-loop (Chakraborty and Grosse, 2011).

1.4.3.3 Les topoisomérases

Les topoisomérases sont des enzymes qui réduisent le niveau de surenroulement de l'ADN dans les cellules (pour plus de détails, voir la section 1.5). La capacité d'une région d'ADN à former une R-loop dépend à la fois de sa séquence et de sa topologie, il n'est donc pas surprenant que les topoisomérases aient été impliquées dans leur résolution (Figure 1-11c) (Stolz et al., 2019).

Chez les plantes telles que le riz, il a été montré que l'absence ou l'inhibition de la topoisomérase I conduit à l'accumulation de R-loops au niveau des gènes relatifs à l'auxine, diminuant leur expression et conduisant à un défaut de croissance (Shafiq et al., 2017). Chez la levure, l'absence de la topoisomérase I conduit à l'accumulation de R-loops au niveau des gènes codant pour les ribosomes et à l'inhibition de leur transcription (El Hage et al., 2010). Que ce soit chez les plantes ou les levures, ces études ont été faites en n'étudiant qu'un nombre limité de loci masquant peut-être un effet global sur les R-loops qui peut différer de ces quelques exemples. Une étude dans des cellules humaines faite à l'échelle du génome entier a mis en avant que la déplétion de la topoisomérase I ne conduit pas uniquement à l'augmentation des R-loops (Manzo et al., 2018). Comme attendue, l'absence de la topoisomérase I conduit à une augmentation globale des R-loops au niveau de régions transcriptionnellement actives marquées par la tri-méthylation des lysines 36 et 9 de l'histone 3. Mais de manière surprenante, les régions réprimées marquées par la tri-méthylation de la lysine 27 de l'histone 3 montrent une diminution du niveau de R-loops suite à la déplétion de la topoisomérase I (Manzo et al., 2018). Ceci suggère que la topoisomérase I est impliquée dans la résolution des R-loops qui se forment de manière co-transcriptionnelle aux gènes activement transcrits, là où elle est présente en temps normal; mais que cette même topoisomérase ne semble pas être impliquée dans la régulation des R-loops formées dans les régions réprimées.

Plus récemment, la topoisomérase III a elle aussi été montrée comme étant impliquée dans la régulation des R-loops (Zhang et al., 2019b). Mais il est probable que les cibles de la topoisomérase III soient différentes de celles de la topoisomérase I. La topoisomérase I réduit le niveau de surenroulement qui se forme lors de la transcription prévenant la formation co-transcriptionnelle de R-loops. La topoisomérase III est plus active lorsqu'il y a un simple-brin d'ADN accessible, suggérant qu'elle résout les R-loops en réduisant le surenroulement via la liaison du simple-brin d'ADN déplacé une fois que celles-ci ont été formées (Kim and Wang, 1992; Plank et al., 2005).

1.4.4 Les différentes techniques utilisées pour identifier les R-loops

1.4.4.1 Le séquençage au bisulfite, bis-seq

Pendant longtemps l'identification des R-loops dans le génome a reposé sur l'utilisation de bisulfite en conditions natives qui entraîne la modification des cytosines d'un ADN simple-brin en uraciles (Figure 1-12a). Les R-loops étaient alors détectées via l'identification du brin d'ADN déplacé (Yu et al., 2003). Or

cette technique est imparfaite car le bisulfite ne modifie pas les cytosines si elles sont méthylées rendant l'identification de R-loops impossible notamment au niveau d'îlot CpG méthylés.

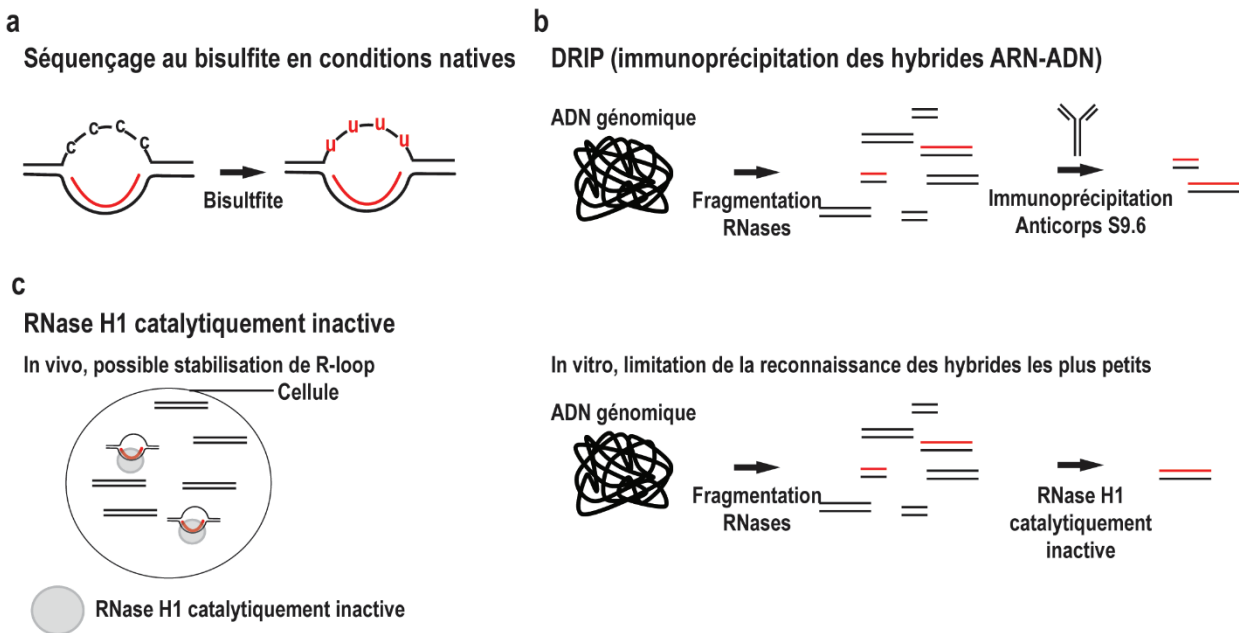


Figure 1-12 Les différentes techniques permettant l'identification des séquences formant des R-loops

a. Le séquençage au bisulfite en conditions natives. Le bisulfite désamine les cytosines en uraciles lorsqu'elles sont présentes dans le simple-brin d'ADN déplacé. b. L'immunoprécipitation des hybrides ARN-ADN (DRIP) avec l'anticorps S9.6 permet la détection des hybrides ARN-ADN après extraction de l'ADN génomique. c. La RNase H1 catalytiquement inactive permet la détection de la partie hybride d'une R-loop in vivo (gauche) ou in vitro après extraction de l'ADN génomique (droite).

1.4.4.2 L'anticorps S9.6, DRIP

L'anticorps S9.6 détecte les hybrides ARN-ADN avec une préférence pour les hybrides les plus longs mais pas de claire spécificité de séquence (Figure 1-12b) (Boguslawski et al., 1986; König et al., 2017; Phillips et al., 2013). Mais cet anticorps n'est pas parfait, même s'il ne présente pas de préférence de séquence entre les hybrides, il peut toutefois lier les ARN double-brins avec une affinité seulement 4 à 5 fois inférieure par rapport aux hybrides ARN-ADN (Phillips et al., 2013). L'utilisation de la RNase III en plus de la RNase A afin de dégrader les ARN simple-brins et double-brins semblent donc nécessaires pour obtenir un signal spécifique (Hartono et al., 2018; Sanz and Chédin, 2019).

Malgré ses imperfections, l'anticorps S9.6 a permis a permis l'identification des R-loops à l'échelle du génome entier dans de nombreuses espèces mettant en évidence les propriétés favorables à leur formation (voir la section 1.4.5) (Sanz et al., 2016; Wahba et al., 2016; Xu et al., 2017).

1.4.4.3 La RNase H1 catalytiquement inactive, R-ChIP ou DRIVE

L'utilisation de l'anticorps S9.6 ainsi que le séquençage au bisulfite se fait après extraction des acides nucléiques des cellules, et donc on peut se demander si la lyse des cellules et purification des acides nucléiques n'entraînent pas la dégradation des R-loops les moins stables. Une version catalytiquement inactive de la RNase H1 exprimée directement dans les cellules permet l'identification des R-loops *in vivo* (Figure 1-12c) (Chen et al., 2019, 2017). Mais l'utilisation de la RNase H1 a ses propres limitations. *In vitro*, la RNase H1 est bien moins efficace que l'anticorps S9.6 (Ginno et al., 2012). *In vivo*, la RNase H1 n'est pas le seul régulateur des R-loops, et son activité est plus importante lorsque la cellule subit un stress entraînant l'accumulation de R-loops (Lockhart et al., 2019). De plus, la reconnaissance de R-loops par la RNase H1 sous-entend que ces R-loops sont accessibles et non pas liées par d'autres protéines. Finalement l'utilisation d'une RNase H1 qui peut toujours lier les R-loops sans les dégrader, entraînerait en théorie la stabilisation de R-loops qui sont naturellement dégradées dans la cellule et donc qui pourraient ne pas avoir de fonction dans le recrutement d'autres protéines par exemple.

Toutes ces limitations expliquent pourquoi l'utilisation de l'anticorps S9.6 et de la RNase H1 donnent des résultats différents, mais n'excluent pas la possibilité que toutes ces R-loops identifiées par ces différentes méthodes soient réelles (Chen et al., 2019).

1.4.5 Les régions du génome formant des R-loops chez différentes espèces

Contrairement à ce qui avait été pensé pendant longtemps, les R-loops ne sont pas uniquement délétères et la cause d'instabilité génomique. Depuis plusieurs années avec le développement d'outils pour les étudier, il a été montré que les R-loops ne se forment pas de manière aléatoire dans le génome et peuvent jouer des fonctions essentielles qui nécessite encore de nombreuses investigations (Chédin, 2016; Niehrs and Luke, 2020).

1.4.5.1 Chez les mammifères

Au niveau de certains promoteurs, les îlots CpG non méthylés forment des R-loops de manière co-transcriptionnelle et ces R-loops préviendraient la méthylation de l'ADN (Ginno et al., 2012). La formation de ces R-loops dépend notamment de la surreprésentation de G sur un des brins d'ADN plutôt que l'autre (Ginno et al., 2013). Les R-loops, en empêchant la méthylation des îlots CpG, préviendraient ainsi de la répression génique (Ginno et al., 2012). De plus ces promoteurs ayant formé des R-loops sont associés avec modifications post-traductionnelles d'histones telles que H3K9me3, H3K4me3, H3K27Ac et ont une

plus grande sensibilité à la DNase montrant le caractère actif de ces gènes (Sanz et al., 2016), les R-loops et nucléosomes ne sont probablement pas présentes à un même loci à un moment donné. Des R-loops se forment également aux sites de terminaison de la transcription des gènes possédant des sites de polyadénylation (Ginno et al., 2013; Sanz et al., 2016). Ces R-loops à l'extrémité 3' des gènes participeraient à la terminaison de la transcription (Sanz et al., 2016; Skourti-Stathaki et al., 2011). À l'exception des gènes courts ou enrichis en G, les R-loops ne se forment que peu le long des gènes (Sanz et al., 2016).

Les amplificateurs peuvent également former des R-loops mais nous ne savons pas si ces amplificateurs sont actifs et engagés dans un contact avec un gène et donc si ces R-loops jouent un rôle positif ou négatif dans la stimulation de l'expression génique par un élément amplificateur (Nojima et al., 2018; Tan-Wong et al., 2019).

In vitro, les R-loops elles-mêmes peuvent agir comme un promoteur et induire la transcription grâce au simple-brin d'ADN qu'elles exposent qui facilite l'assemblage de l'ARN polymérase et donc la transcription (Tan-Wong et al., 2019). *In vivo*, environ 50% des ARNnc antisens et 25% des ARNnc produits à partir des amplificateurs semblent l'être à la suite de la formation de R-loops comme le montre leur sensibilité à la surexpression de la RNase H1 (Tan-Wong et al., 2019). Il est difficile d'imaginer que ces nombreux ARNnc ne soient que la conséquence de la formation de R-loops non-résolues et qu'ils n'ont aucune fonction *in vivo*.

1.4.5.2 Chez les levures

Chez la levure *Saccharomyces cerevisiae* les R-loops se forment principalement aux petits ARN de transferts, ARNnc, petits ARN nucléaires et nucléolaires ainsi qu'aux télomères (Wahba et al., 2016). Les R-loops, là encore, s'accumulent aux sites d'initiation et de terminaison de la transcription, ainsi que le long des gènes fortement exprimés ce qui est contre intuitif puisqu'on s'attendrait à ce que la présence d'une R-loop empêche le passage de l'ARN polymérase et donc réduise le niveau de transcription (Wahba et al., 2016). De plus, il a été observé que des R-loops se forment au niveau de sites de terminaison à proximité de séquences polyA et qu'elles dépendent de ces sites polyA pour se former (Wahba et al., 2016).

Chez la levure *Schizosaccharomyces pombe*, les R-loops s'accumulent également dans les régions promotrices et terminatrices et elles sont associées avec des protéines impliquées dans l'élongation de la

transcription (Hartono et al., 2018). De plus les R-loops se forment aussi aux ARN de transfert (Hartono et al., 2018).

1.4.5.3 Chez les plantes

Chez *Arabidopsis thaliana*, les R-loops s'accumulent essentiellement aux promoteurs et le long des gènes mais très peu se situent aux régions terminatrices (Xu et al., 2017). Les plantes présentent également de nombreuses R-loops antisens qui se forment dans les régions promotrices. Comme chez les mammifères la présence de R-loops est corrélée négativement avec le méthylation de l'ADN CG, positivement avec la tri-méthylation des lysines 4 et 36 de l'histone 3 et ces gènes présentent une expression élevée. Les gènes ayant formés des R-loops dans les deux orientations ont le niveau d'expression le plus élevé, ce qui peut s'expliquer par les observations faites par Tan-Wong et al. où la formation d'une R-loop antisens faciliterait l'assemblage du complexe d'initiation et donc la transcription (Tan-Wong et al., 2019; Xu et al., 2017). Les R-loops ne sont que peu enrichies avec la marque épigénétique répressive H3K27me3 déposée par PRC2, mais très fortement enrichies avec H3K9me2 et H3K27me1 (Xu et al., 2017). Finalement, il a été observé que des R-loops se forment également aux gènes codants pour les ARN ribosomiaux, ARN de transferts, aux petits ARN nucléolaires, aux éléments transposables et aux télomères (Xu et al., 2017).

De manière générale, les R-loops se forment préférentiellement au niveau de gènes et dans des régions du génome activement transcrites laissant à penser qu'elles se forment de manière co-transcriptionnelle et participeraient à l'expression génique. Toutefois certaines R-loops se forment dans des régions réprimées et la fonction ainsi que l'origine de ces dernières restent plus énigmatiques.

1.4.6 Les R-loops ne sont pas uniquement source d'instabilité génomique

Contrairement à ce qui a été pensé pendant de nombreuses années, les R-loops n'ont pas uniquement un rôle délétère, mais peuvent avoir des fonctions positives comme c'est le cas chez les plantes lors de la vernalisation où la formation de R-loop participe au maintien de l'expression du gène FLC, ou encore chez les mammifères au niveau des promoteurs contenant des îlots CpG où les R-loops préviennent de la méthylation de l'ADN et donc de la répression (Ginno et al., 2012; Sun et al., 2013). Deux autres exemples de fonctions de R-loops dans la maintenance des télomères et la réparation des cassures double-brins sont présentés ci-dessous.

1.4.6.1 Le rôle des R-loops se formant aux télomères

Les extrémités des chromosomes doivent être cachées afin d'éviter d'être reconnues comme des cassures d'ADN doubles brins pouvant entraîner la fusion de chromosomes et le raccourcissement des télomères. Les télomères se composent de répétitions riches en G dont la séquence varie selon les espèces. Un des deux brins d'ADN, plus long et enrichis en G, va pouvoir envahir l'ADN double-brin du télomère grâce à l'action de TRF2, formant une boucle (T-loop) protégeant ainsi l'extrémité du chromosome (Doksani et al., 2013). Les extrémités d'ADN sont difficiles à répliquer et chaque cycle cellulaire conduit au raccourcissement des télomères, qui doivent être agrandis afin d'éviter la sénescence de la cellule. Les télomères sont transcrits en un ARNnc TERRA contenant ces répétitions riches en G. Cet ARN sert de matrice à la télomérase, une transcriptase inverse, qui va pouvoir ainsi allonger les télomères.

TRF2, tout comme elle le fait avec l'ADN télomérique simple-brin, peut induire l'invasion de l'ADN double-brin télomérique par l'ARN TERRA formant ainsi une R-loop (Lee et al., 2018c; Toubiana and Selig, 2018). Le rôle exact de ces R-loops qui se forment aux télomères n'est pas encore défini. Chez la levure, il a été montré que les R-loops s'accumulent lorsque les télomères sont courts permettant l'activation des mécanismes de réparation des cassures d'ADN doubles brins donnant ainsi la possibilité à la cellule de rallonger les télomères (Balk et al., 2013; Graf et al., 2017; Pérez-Martínez et al., 2020). Les cellules humaines déficientes pour la méthyltransférase DNMT3B, présentent un fort niveau d'ARNnc TERRA et de R-loops aux télomères, ainsi qu'un niveau de dommages dans l'ADN plus important dans ces régions, suggérant alors un rôle négatif pour ces R-loops (Sagie et al., 2017). Une observation similaire a été faite dans des cellules humaines ALT où une accumulation de R-loops aux télomères corrèle avec des recombinaisons plus importantes entre les télomères (Arora et al., 2014). Ce phénotype peut être restauré par la surexpression de la RNase H1 (Arora et al., 2014). Ces observations montrent que les R-loops aux télomères jouent un rôle positif pour le maintien de ces derniers mais que le niveau de ces R-loops doit être régulé afin de ne pas devenir délétères.

1.4.6.2 Le rôle des R-loops lors des cassures dans l'ADN

Il existe deux mécanismes de réparation des cassures doubles brins. Lors de la réparation par la jonction des extrémités non homologues (NHEJ), les extrémités d'ADN sont liées par les protéines Ku70 et Ku80, qui vont pouvoir orienter ces extrémités afin de permettre leur ligation entraînant la perte d'information. La recombinaison homologue (HR), mécanisme de réparation plus fidèle nécessite l'invasion d'ADN double-brin présentant une homologie de séquence par l'un des brins d'ADN présent au site de cassures, permettant ainsi de réparer la cassure dans l'ADN sans entraîner de perte d'information génétique. Ce

mécanisme de réparation repose notamment sur les protéines Rad51 ou Rad52 qui peuvent induire l'invasion de brins.

Chez la levure *Saccharomyces cerevisiae*, l'utilisation d'un système inductible de cassures double-brins a permis de mettre au jour que les ARN ayant été produit avant le bris de l'ADN, peuvent être utilisés pour la réparation soit sous forme d'ADN après rétrotranscription soit directement sous forme d'ARN comme le montre leur souche mutante pour les RNase H1 et H2 ainsi que *spt3* (Figure 1-13b-c) (Keskin et al., 2014). Keskin et al. ont proposé que cet ARN produit en *cis* ou en *trans*, envahisse l'ADN double-brins grâce à l'action de Rad52 et puisse maintenir les deux extrémités ensemble ou être utilisé directement comme matrice afin de réparer la séquence perdue suite à la cassure de l'ADN (Figure 1-13a) (Keskin et al., 2014; Mazina et al., 2017).

Chez la levure *Schizosaccharomyces pombe*, c'est un autre mécanisme impliquant lui aussi les ARN qui a été montré (Ohle et al., 2016). À la suite d'une cassure double-brin et après résection d'un brin d'ADN par le complexe MRN laissant une extrémité d'ADN simple-brin accessible, l'ARN polymérase est recrutée au site de cassure et produit un ARN pouvant s'hybrider avec son brin matrice en absence d'un second brin d'ADN compétiteur (Figure 1-13d) (Ohle et al., 2016). Ces hybrides sont alors dégradés par les RNase H1 et H2, permettant la liaison de RPA à l'ADN simple-brin de nouveau accessible, laissant la possibilité à Rad51 de remplacer RPA et de poursuivre la réparation de la cassure par recombinaison homologue (Ohle et al., 2016).

Cette relation entre les R-loops, Rad52 et la réparation des cassures dans l'ADN semble être en partie conservée chez les mammifères (Yasuhara et al., 2018). À la suite de cassures, Rad52 est recrutée après la formation d'hybrides ARN-ADN au site de bris, ce qui conduit au recrutement de l'endonucléase XPG qui peut dégrader la partie ADN de l'hybride, libérant ainsi un simple-brin d'ADN qui va pouvoir envahir un ADN double-brin permettant la recombinaison homologue (Yasuhara et al., 2018). Cette étude montre donc la relation inverse entre Rad52 et les hybrides ARN-ADN où la formation d'hybrides entraîne le recrutement de Rad52 (Yasuhara et al., 2018). Alors que *in vitro* il avait été montré que la protéine humaine Rad52, tout comme la protéine de levure, pouvait induire la formation d'hybride ARN-ADN grâce à son activité d'invasion de brins (Mazina et al., 2017).

La fonction des hybrides ARN-ADN aux sites de cassures dans l'ADN n'est pas encore complètement élucidée. Finalement, ces hybrides pourraient également permettre le recrutement d'autres facteurs qui interviennent dans la répression des gènes à proximité du site de cassure en plus de contrôler la résection

des brins comme cela a été suggéré (Figure 1-13e) (Caron et al., 2019; Chen et al., 2015; Hegazy et al., 2020; Ohle et al., 2016).

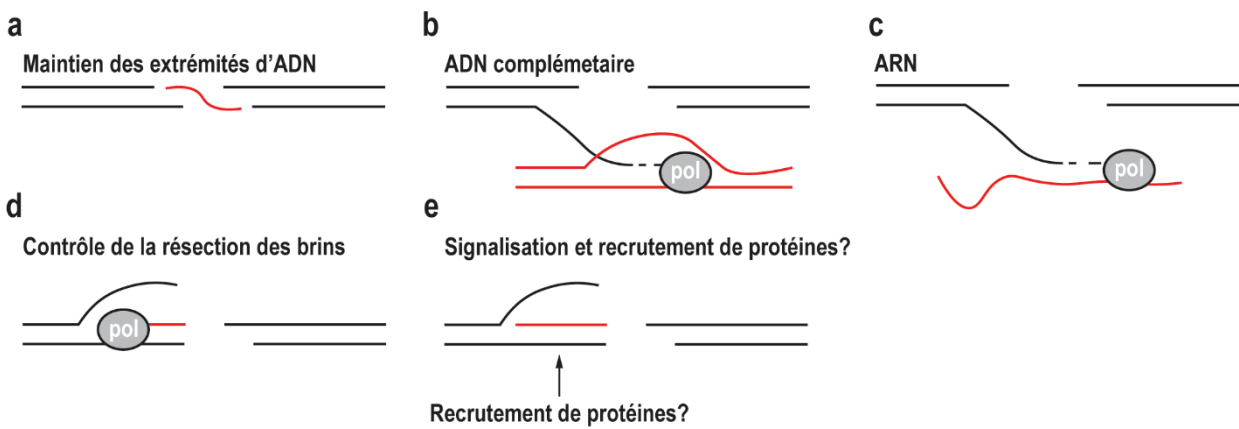


Figure 1-13 Les rôles potentiels des hybrides ARN-ADN dans la réparation des cassures d'ADN double-brins

a. Un ARN produit avant le bris peut servir à maintenir ensemble les deux extrémités d'ADN. b-c. L'ARN après transcription inverse (b) ou directement (c) peut servir de matrice permettant de restaurer l'information perdue à la suite du bris de l'ADN. d. L'ARN produit à partir du site de cassures, s'hybride avec l'ADN et participe au contrôle de la résection des brins. e. La formation d'un hybride ARN-ADN au site de cassure peut participer au recrutement de protéines impliquées dans la signalisation et/ou la réparation du bris.

1.4.6.3 Les R-loops participent à la régulation de l'expression génique

Tout comme l'ARNnc COOLAIR qui forme une R-loop empêchant la mise en place de la répression stable du gène *FLC*, d'autres ARNnc participent à la régulation de l'expression d'autres gènes. L'ARNnc antisens GATA3-AS1 forme une R-loop et participe au recrutement de MLL, qui par la suite méthyle la lysine 4 de l'histone 3 stimulant l'expression de GATA3 (Gibbons et al., 2018). Ces R-loops formées de manière co-transcriptionnelle permettent de réguler l'expression d'un gène qui se situe à proximité en recrutant des protéines pouvant induire l'expression ou la répression génique.

Les R-loops permettraient donc de recruter spécifiquement des protéines qui lient les acides nucléiques sans spécificité de séquence en exposant une structure d'acides nucléiques non-canonique qui peut favoriser leur recrutement (Chédin, 2016). De façon inverse ces R-loops pourraient prévenir le recrutement de certains facteurs de transcription qui pourraient ne pas reconnaître leur séquence cible si celle-ci est comprise dans un hybride ARN-ADN ou sur un ADN simple-brin. Actuellement la majorité des exemples proviennent de R-loops se formant co-transcriptionnellement et ayant un effet à proximité de leur site de transcription, il ne faut toutefois pas exclure la possibilité que certaines de ces R-loops puissent se former en *trans*, entre un ARN et un ADN présentant une homologie de séquence (Chédin, 2016).

Les R-loops ont des rôles dans la régulation de l'expression génique en recrutant ou prévenant du recrutement de certains facteurs, mais également dans d'autres mécanismes comme la réparation des cassures de l'ADN et la maintenance des télomères. L'identification des nombreuses R-loops récemment, le fait qu'elles ne soient pas affectées de la même façon par la surexpression ou l'absence des protéines qui les résolvent, ainsi que l'expression différentielle de nombreux gènes lorsque la stabilité de ces R-loops est affectée laissent à penser qu'elles ont de nombreuses fonctions qui restent encore inexplorées.

1.5 La topologie de l'ADN

1.5.1 La topologie de l'ADN

1.5.1.1 Le surenroulement et la relaxation de l'ADN

L'ADN excède de beaucoup la taille de la cellule, et à l'instar de la chromatine qui permet sa compaction, le surenroulement y contribue également. Contrairement aux bactéries, les organismes eucaryotes sont dénués d'ADN gyrase pouvant introduire des supertours. Le surenroulement est le résultat de la liaison à l'ADN de protéines telles que les histones autour desquelles s'enroulent l'ADN entraînant la formation de supertours négatifs (Simpson et al., 1985). La transcription et la réplication de l'ADN sont également deux acteurs majeurs du surenroulement présent dans l'ADN des eucaryotes. En effet l'ouverture de la double hélice d'ADN, le déplacement des polymérase le long de l'ADN entraîne la formation de supertours positifs et négatifs qui doivent être résolus afin d'éviter l'apparition de cassures dans l'ADN, la recombinaison entre les chromosomes ou encore l'arrêt des polymérase qui sont causes d'instabilité génomique (Figure 1-14).

Les approches à l'échelle du génome entier tels que le CHIP-seq ont permis d'identifier les sites de liaison des protéines régulant la topologie de l'ADN et les psoralènes, molécules qui s'intercalent dans l'ADN relaxé, ont permis l'identification des régions où l'ADN est relaxé dans les cellules (Corless and Gilbert, 2017).

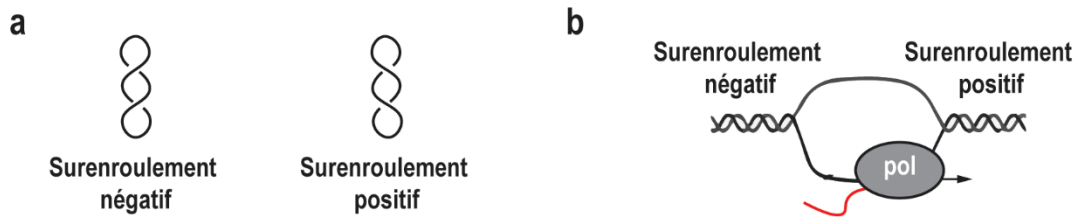


Figure 1-14 La topologie de l'ADN

a. Schéma d'un ADN surenroulé négativement (gauche) ou positivement (droite). b. La transcription et la réplication de l'ADN conduit à l'accumulation de surenroulement négative en amont et de surenroulement positif en avant de la bulle de réplication ou de transcription.

1.5.1.2 Le changement du nombre de supertours ou de *linking number*

Chez les eucaryotes l'ADN est linéaire et ne devrait pas en principe être affecté par le surenroulement, mais la taille de chaque chromosome et la proximité des molécules dans le noyau font que les extrémités ne sont pas libres de tourner afin d'éliminer le surenroulement.

Le *linking number* définit le nombre de fois qu'un brin de la double hélice d'ADN passe au-dessus de l'autre (Mirkin, 2001). Ce nombre ne peut être changé à moins qu'au moins l'un des deux brins de la molécule d'ADN ne soit coupé, qu'il ne tourne autour ou ne traverse l'autre brin d'ADN avant d'être reliqué. Mathématiquement ce nombre est défini par la torsion et l'entortillement (*writhe* et *twist*), qui expliquent les différentes modifications que la double hélice d'ADN peut subir lors de surenroulement. Le surenroulement peut être positif si le brin du dessus tourne vers la droite et négatif s'il va vers la gauche (Figure 1-14a).

1.5.1.3 La topologie de l'ADN dans les cellules

Contrairement aux bactéries, les eucaryotes ne possèdent pas de gyrase pouvant introduire des supertours dans l'ADN. Le surenroulement est le résultat de la liaison à l'ADN de protéines telles que les histones, mais également de la transcription et de la réplication de l'ADN. Les histones, assemblées en nucléosomes sont entourées d'ADN ce qui conduit à l'introduction d'un tour négatif pour chaque nucléosome assemblé (Simpson et al., 1985). La chromatine n'est pas pour autant une barrière rigide qui contraint le surenroulement, en effet elle peut absorber en partie les supertours générés lors de la transcription et réplication de l'ADN (Bancaud et al., 2006).

Le développement des psoralènes, molécules qui s'intercalent dans l'ADN négativement enroulé (*underwound*), combiné à des approches de séquençage à l'échelle du génome entier ont permis l'identification des régions positivement ou négativement surenroulées (Corless and Gilbert, 2017). Ces

différents domaines positivement ou négativement surenroulés, longs d'environ 100 kilobases se distinguent par un enrichissement différent en GC et peuvent être bordés par la protéine isolatrice CTCF (Dixon et al., 2012; Naughton et al., 2013). Leur formation dépend à la fois de la transcription et de l'activité des topoisomérases (Naughton et al., 2013). A l'intérieur de ces domaines, la transcription altère localement la topologie de l'ADN. En amont du site d'initiation de la transcription et en aval du site de terminaison, l'ADN est négativement surenroulé et ce niveau de surenroulement dépend notamment du niveau d'expression (Achar et al., 2020; Naughton et al., 2013). À l'inverse, le cadre de lecture transcrit présente un surenroulement positif qui dépend peu du niveau d'expression du gène (Achar et al., 2020; Naughton et al., 2013). Ceci a conduit à la formulation de l'hypothèse suivante : dans le cas des gènes faiblement transcrits les nucléosomes sont déplacés au fur et à mesure de la progression de l'ARN polymérase celle-ci entraîne la formation de supertours positifs devant elle et de supertours négatifs derrière qui sont en partie réduits par les topoisomérases de type I et II (voir les sections 1.5.2 et 1.5.3). Dans le cas des gènes fortement transcrits la succession d'ARN polymérase entraîne l'accumulation de surenroulements négatifs que les Topoisomérases n'arrivent pas à réduire assez rapidement. Dans ce cas la formation de surenroulement négatif extrême en 5' de l'ARN polymérase peut potentiellement favoriser la transcription elle-même puisque ça peut entraîner l'ouverture de la double hélice d'ADN, et l'exposition d'un ADN simple-brin pouvant faciliter l'assemblage de l'ARN polymérase et du complexe de pré-initiation comme cela a été proposé (Tan-Wong et al., 2019). L'accumulation de supertours positifs quel que soit le niveau de la transcription déstabilise les nucléosomes, aidant ainsi la progression de l'ARN polymérase (Achar et al., 2020; Gupta et al., 2009; Teves and Henikoff, 2014).

1.5.1.4 La transcription et réplication de l'ADN altèrent la topologie de l'ADN et peuvent être la cause d'instabilité génomique

La transcription et la réplication nécessitent l'ouverture de la double hélice d'ADN et le déplacement de l'ARN polymérase et du réplisome respectivement le long de l'ADN. Ceci entraîne la formation de surenroulement positif devant la polymérase et du surenroulement négatif derrière qui ne peuvent être absorbés en raison de protéines liées à l'ADN et l'incapacité des extrémités de l'ADN à tourner librement (Pommier et al., 2016). Ces surenroulements positifs et négatifs doivent néanmoins être réduits dans le but de permettre la progression de la polymérase et d'éviter l'apparition de cassures dans l'ADN qui pourraient survenir en cas de stress topologique trop important. Les topoisomérases et les mécanismes par lesquels elles réduisent le surenroulement sont présentés dans les parties suivantes (Pommier et al., 2016).

Néanmoins, à certains sites les topoisomérases ne suffisent pas à réduire le surenroulement pouvant entraîner l'arrêt des machineries de réplication ou de transcription, leur collision, l'accumulation du stress topologique ainsi que l'exposition d'un ADN simple-brin, conduisant à l'apparition de cassures dans l'ADN (Keszthelyi et al., 2016; Magdalou et al., 2014). Ces cassures ne se produisent pas de manière complètement aléatoire mais se produisent plus fréquemment à des sites dit fragiles qui corrèlent avec un niveau de transcription plus élevé, la présence de structure secondaire dans l'ADN ou encore des sites de liaison pour des protéines qui contribuent à l'augmentation du surenroulement ou empêchent sa réduction (Dhar and Schildkraut, 1991; Helmrich et al., 2011; Szilard et al., 2010; Thys et al., 2015). Ces cassures entraînent l'arrêt du cycle cellulaire et sont alors prises en charge par les machineries de réparation homologue ou non homologue qui pourront conduire à des pertes d'informations ou recombinaisons lors de leur réparation (Magdalou et al., 2014; Zeman and Cimprich, 2014).

1.5.2 Les topoisomérases

Les topoisomérases sont des protéines conservées à travers l'évolution qui peuvent relaxer l'ADN surenroulé de manière positive et/ou négative mais qui ne peuvent pas introduire de surenroulement qui est une propriété qui les font diverger de l'ADN gyrase bactérienne. Les topoisomérases se divisent en différentes catégories basées sur le mécanisme par lequel elles agissent. Dans les parties suivantes les fonctions des topoisomérases des eucaryotes supérieurs sont décrites.

Les topoisomérases ont un mécanisme d'action similaire (Bizard and Hickson, 2020; Capranico et al., 2017; Riccio et al., 2020). Elles reposent toutes sur une tyrosine comme résidu catalytique qui va pouvoir attaquer l'ADN et briser un phosphate afin de former une liaison covalente avec le ou les brins d'ADN coupés. Puis elles permettent la réduction du nombre de supertour soit par rotation des brins d'ADN ou par passage du brin d'ADN intact à travers celui qui a été coupé, avant la ligation de l'ADN coupé, ce qui conduit à une réduction du niveau de surenroulement. Quelle que soit la topoisomérase celle-ci reste lié à l'ADN de manière covalente afin de réduire rapidement le niveau de surenroulement et d'éviter la reconnaissance de cette cassure par les mécanismes de réparation de l'ADN qui pourrait conduire à des recombinaisons entre les chromosomes ce qui serait dommageable pour la cellule (Bizard and Hickson, 2020; Capranico et al., 2017; Riccio et al., 2020).

Une description plus approfondie du mécanisme d'action des topoisomérases ainsi que de leur structure suit dans les parties suivantes. En raison de l'implication des topoisomérases lors de multiples évènements

du cycle cellulaire et de leurs rôles complémentaires ou partiellement redondants, leurs fonctions sont présentées dans une section commune après présentation des différentes topoisomérases (Figure 1-15).

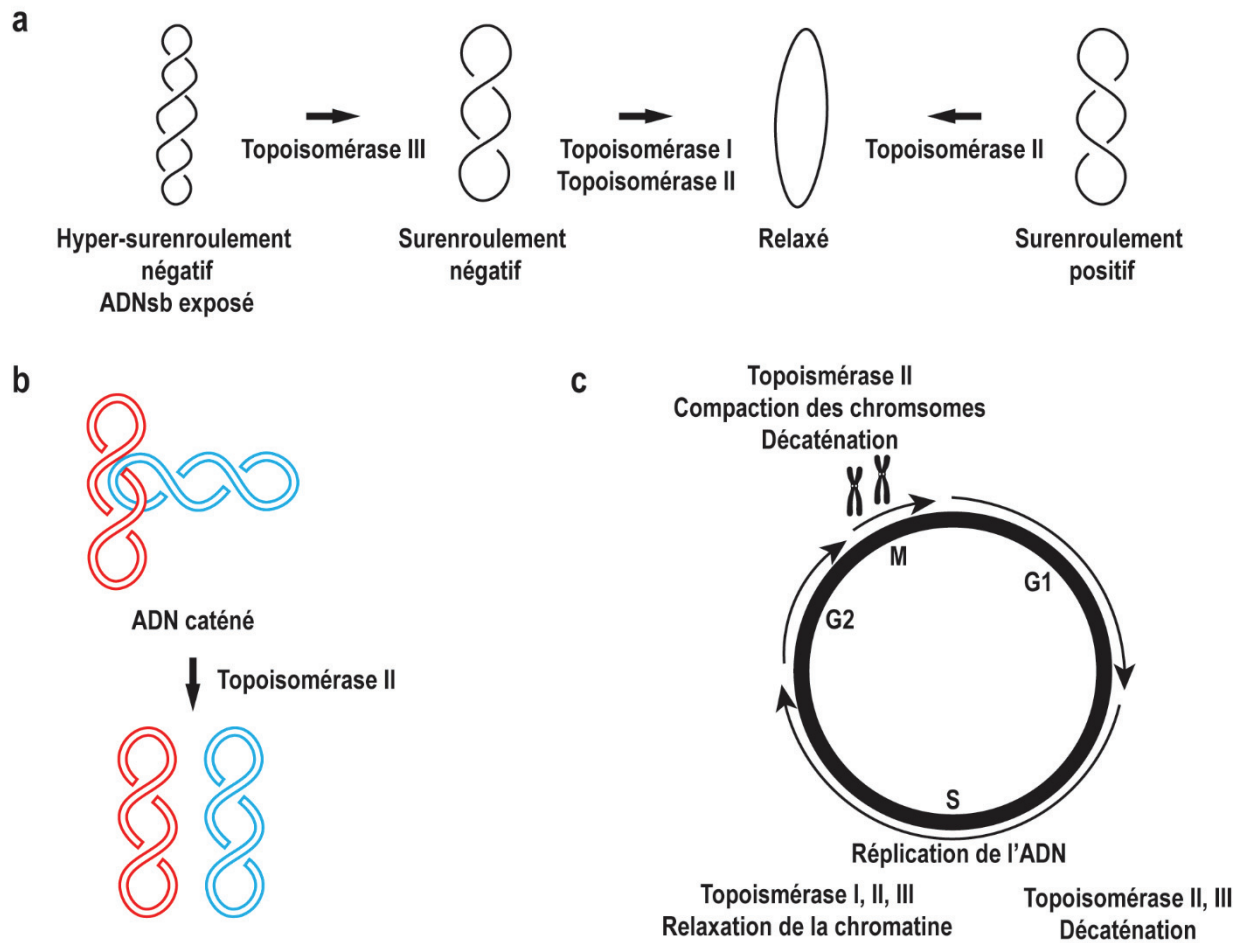


Figure 1-15 Les topoisomérases

a. Schéma résumant les préférences des topoisomérases de type I, II et III pour l'ADN surenroulé positivement, négativement ou extensivement enroulé. b. La décaténation de deux molécules d'ADN par la topoisomérase de type II. c. Schéma résumant les fonctions des topoisomérases de type I, II et III lors de la mitose et de la phase S du cycle cellulaire.

1.5.2.1 La topoisomérase de type I (ou type IB)

La topoisomérase de type I coupe un seul brin d'ADN et forme un lien covalent avec la partie 3' de l'ADN avant de faire pivoter les brins afin de réduire le niveau de surenroulement négatif (Figure 1-15a) (Capranico et al., 2017). Contrairement aux autres topoisomérases elle ne nécessite pas d'hydrolyse d'ATP ni de liaison de cation divalent tel que le magnésium. La Topoisomérase I se compose de 4 domaines : la partie N-terminale comprend la séquence de localisation nucléaire, le domaine central qui contient les résidus catalytiques, la partie C-terminale où se trouve la tyrosine catalytique, et pour finir le domaine de

liaison entre le domaine central et la partie C-terminale qui est responsable de la processivité de la Topoisomérase (Stewart et al., 1997). La structure de la topoisomérase I ayant formée un complexe covalent ou non est connue, mais cette structure ne contient pas le domaine N-terminal (Leshner et al., 2002; Redinbo et al., 1998; Staker et al., 2002a). La topoisomérase forme deux lobes qui font office de pince et encerclent l'ADN. La première partie se compose des sous-domaines 1 et 2 du domaine central, et la deuxième partie de la pince de la 3^{ème} partie du domaine central avec le domaine C-terminal. Le domaine de liaison entre ces deux lobes est flexible et permet de réguler la processivité de la topoisomérase (Stewart et al., 1997). Malgré l'absence de la structure du domaine N-terminal, il a été suggéré que celui-ci, flexible comme le domaine de liaison, régulerait la rotation de la topoisomérase et la liaison de l'ADN (Chillemi et al., 2004; Frøhlich et al., 2004). La partie N-terminale de la topoisomérase I interagit avec la CTD de la plus grosse sous-unité de l'ARN polymérase II (Merino et al., 1993) et peut ainsi stimuler l'ARN polymérase (Baranello et al., 2016).

La tyrosine catalytique est essentielle puisque son remplacement par une phénylalanine entraîne la perte de l'activité de la topoisomérase, les résidus qui composent le site catalytique jouent également un rôle prépondérant pouvant entraîner une réduction de l'activité de la topoisomérase en cas de mutation (Baker et al., 2009; Eng et al., 1989). En plus des résidus en tant que tels, les modifications post-traductionnelles de la topoisomérase telles que la phosphorylation ou SUMOylation sont également essentielles pour que celle-ci soit efficace (Durban et al., 1983; Li et al., 2015; Pommier et al., 1990). Ces modifications régulent également sa capacité à être la cible d'inhibiteurs tels que la Camptothecin (Yu et al., 2004). La Camptothecin est un inhibiteur spécifique de la topoisomérase I qui s'intercale dans l'ADN lorsque la topoisomérase a formé un complexe covalent avec ce dernier (Eng et al., 1989; Nitiss and Wang, 1988; Pommier, 2006; Staker et al., 2002a).

Il y a deux topoisomérases de type I, une mitochondriale qui ne sera pas discutée ici (Zhang et al., 2001) et une nucléaire qui se localise à la fois dans le nucléoplasme et le nucléole. Cette localisation nucléolaire est en partie le résultat de l'interaction du domaine N-terminal de la topoisomérase I avec l'ARN polymérase I, qui transcrit les ARN ribosomiaux, et la nucléoline, protéine nucléolaire (Bharti et al., 1996; Christensen et al., 2002); même si une délétion du domaine N-terminale qui contient les domaines de localisation nucléaire et nucléolaire affecte peu sa localisation cellulaire (Alsner et al., 1992; Christensen et al., 2003; Mo et al., 2000).

1.5.2.2 La topoisomérase de type II

Chez les mammifères il y a deux topoisomérases II α et β alors que chez la plupart des autres eucaryotes tels que *D. melanogaster* ou la levure il n'en existe qu'une. Elles diffèrent notamment par leur fonction dans la cellule, l'une ayant un rôle plus important dans la relaxation du surenroulement et l'autre dans la décaténation (Roca, 2009). Les topoisomérases de type II agissent sous forme d'homodimères et coupent les deux brins d'ADN laissant quelques bases débordantes. Elles restent associées avec les deux extrémités d'ADN générées afin de réduire le niveau de surenroulement positif ou négatif en passant les brins d'ADN au travers l'un de l'autre (Figure 1-15a) (Roca, 2009). Contrairement aux topoisomérases de type I ou III, la topoisomérase de type II nécessite la liaison et l'hydrolyse d'ATP ainsi que la liaison de magnésium. La topoisomérase de type II repose sur une tyrosine qui va pouvoir attaquer l'ADN et sur un domaine TOPRIM qui lie l'ion métallique nécessaire à son activité. Puisqu'elle peut couper les deux brins qui composent la molécule d'ADN elle peut également décaténer deux plasmides ou deux molécules d'ADN (Figure 1-15b) (Lee and Berger, 2019; Riccio et al., 2020).

La topoisomérase de type II se divise en 3 domaines, le domaine N-terminal qui permet la liaison et l'hydrolyse d'ATP, le domaine central responsable de la liaison et de la coupure de l'ADN puisqu'il contient la tyrosine catalytique et le domaine TOPRIM, et pour finir le domaine C-terminal qui est différent entre les topoisomérases II α et β expliquant les différences observées dans leur activités (Austin and Marsh, 1998; Berger et al., 1996). La structure de la topoisomérase II de levure est connue mais en absence du domaine C-terminal (Berger et al., 1996; Classen et al., 2003; Dong and Berger, 2007). Le domaine N-terminal lie l'ATP et contribue à la dimérisation de la topoisomérase, il se situe au-dessus du domaine central qui adopte une forme de cœur avec une large cavité centrale dans laquelle l'ADN va pouvoir subir l'attaque nucléophile de la tyrosine (Berger et al., 1996; Classen et al., 2003; Dong and Berger, 2007). A l'extrémité de ce cœur se situe une porte par laquelle l'ADN va pouvoir entrer et sortir.

La topoisomérase II peut réduire le nombre de supertours d'une molécule d'ADN surenroulée positivement ou négativement et peut également décaténer deux molécules d'ADN en faisant passer l'une à travers l'autre (Fernández et al., 2014; McClendon et al., 2005). Tout comme pour la topoisomérase de type I, des inhibiteurs de la topoisomérase de type II tels que *etoposide* et ses dérivés inhibent spécifiquement la topoisomérase II lorsqu'elle a formé un lien covalent avec l'ADN empêchant sa ligation.

1.5.2.3 La topoisomérase de type III (ou type IA)

La topoisomérase de type III possède, comme les topoisomérases de type II, une tyrosine catalytique et un domaine TOPRIM qui nécessite la liaison de cations divalents mais pas celle d'ATP (Bizard and Hickson,

2020; Srivenugopal et al., 1984). Chez les eucaryotes, il y a deux topoisomérases de type III α et β qui ont des fonctions différentes mais une structure similaire.

La structure de la topoisomérase III se compose de 3 parties, une première partie qui contient le domaine TOPRIM et forme la base de la pince, une seconde partie qui permet la formation de la cavité dans laquelle se fait le passage de brins et qui assure la flexibilité de la topoisomérase et une troisième qui forme la partie supérieure de la pince et qui contient la tyrosine catalytique (Changela et al., 2001; Lima et al., 1994)

La topoisomérase de type III réduit le niveau de surenroulement non pas en faisant pivoter les brins d'ADN comme la topoisomérase I mais en passant le brin d'ADN intact à travers celui qu'elle a coupé (Tse et al., 1980). De manière similaire à la topoisomérase II, ce passage de brin s'effectue dans la cavité de la structure en forme de cadenas qu'adopte la topoisomérase III (Changela et al., 2001; Lima et al., 1994). Puisque cette topoisomérase peut faire passer un brin d'ADN à travers un autre, elle peut non seulement relaxer un ADN surenroulé mais également désemmêler des héli-caténats ou pré-caténats qui peuvent se former lors de la transcription ou lors de réplication de l'ADN (DiGate and Marians, 1988; Wang, 1971). La topoisomérase III se distingue des topoisomérases de types I et II par sa préférence pour l'ADN présentant un surenroulement négatif extrême ce qui peut générer l'apparition d'ADN simple-brin qui est le substrat préféré de cette topoisomérase (Figure 1-15a) (Kim and Wang, 1992; Plank et al., 2005). L'affinité de la topoisomérase III pour les régions simple-brin ne se limite pas à l'ADN mais s'étend jusqu'aux ARN pour la topoisomérase III β (Wang et al., 1996; Xu et al., 2013). La capacité de la topoisomérase III β à décaténer les acides nucléiques simples brins peut lui permettre de modifier la structure d'ARN et de défaire de potentiels nœuds afin de permettre leur traduction (Wang et al., 1996; Xu et al., 2013).

La dépendance de la topoisomérase de type III à un ADN simple-brin fait qu'elle est dépendante d'autres protéines ou évènement cellulaire qui conduisent à leur exposition. La topoisomérase III α coopère avec des hélicases telles que BLM ou avec PICH afin de résoudre des pré-caténats ou des fourches de Holliday, de faciliter la ségrégation des chromosomes ou encore réduire le niveau de surenroulement causé par l'extrusion de l'ADN dans une boucle par les cohésines (Bizard et al., 2019; Chen et al., 2012, 2014; Fabre et al., 2002; Kaur et al., 2015; Suski and Marians, 2008; Tang et al., 2015). La topoisomérase III β est la seule capable d'interagir avec les ARN. Ce n'est donc pas surprenant qu'elle ne soit pas seulement présente dans le noyau mais aussi dans le cytoplasme où elle interagit avec des ARN messagers, la protéine TDRD3 et le complexe FMRP afin de permettre leur traduction (Ahmad et al., 2017; Xu et al., 2013). La perte de la topoisomérase III β conduit à l'accumulation de R-loops et à une augmentation de l'instabilité génomique,

mais également à un défaut dans l'établissement de l'hétérochromatine et une dérégulation de certains gènes lui suggérant un rôle dans la régulation de la transcription (Lee et al., 2018b; Zhang et al., 2019b).

1.5.2.4 Les recombinaisons à site spécifique et les protéines similaires aux topoisomérases

1.5.2.4.1 Les recombinaisons à site spécifique

Les recombinaisons à sites spécifiques permettent la recombinaison entre deux séquences spécifiques pouvant entraîner l'insertion, la délétion ou l'échange d'une séquence ADN mais aussi la modification du nombre de supertours présent dans l'ADN. Ces recombinaisons sont présentes chez les archées, les bactéries et certains microbes des eucaryotes (Olorunniji et al., 2016; Yang, 2010).

Elles ne présentent pas de similarité de structure et se divisent en deux catégories dépendamment de si elles reposent sur une sérine ou sur une tyrosine comme résidu catalytique (Bizard and Hickson, 2020; Capranico et al., 2017; Riccio et al., 2020). Les recombinaisons telles que les recombinaisons CRE ou FLP, ou encore l'intégrase ϕ C31 et la résolvasse $\gamma\delta$, vont pouvoir, grâce à leur résidu catalytique, couper l'ADN et y rester lié de manière covalente, entraîner la recombinaison avant de liguer l'ADN et tout cela se fait sans avoir besoin d'ATP. Contrairement aux topoisomérases, la nécessité de deux sites de recombinaison restreint leur activité, et montre la nécessité n'ont pas d'un mais de deux éléments d'ADN (les deux sites pouvant être présents sur deux molécules d'ADN différentes ou sur la même mais éloignés l'un de l'autre).

Les tyrosines recombinaisons s'oligomérisent afin de former un tétramère pour induire la recombinaison. Une première paire de recombinaisons va être active, chacune va entraîner la scission d'un phosphate, l'échange du brin d'ADN qui n'est pas lié de manière covalente et pour finir la ligation. Puis c'est une seconde paire de recombinaisons qui est active afin de recombiner le second brin d'ADN et de résoudre la fourche de Holliday qui a été formée (Biswas et al., 2005; Burgin and Nash, 1995; Nunes-Düby et al., 1995). Les sites de recombinaisons sont longs de quelques paires de bases et peuvent s'accompagner de séquences accessoires qui réguleraient l'activité des recombinaisons et indiqueraient le sens préférentiel dans lequel doit se faire la recombinaison (Biswas et al., 2005). L'intégration, délétion ou recombinaison entre deux séquences entraînent également un changement pair du nombre de supertours présents dans les molécules d'ADN (Crisona et al., 1999).

Tout comme les tyrosines recombinaisons, les sérines recombinaisons s'assemblent également en complexe synaptique. Mais c'est par un mécanisme de cassure double-brins avec l'ADN présent sur l'extérieur de ce complexe synaptique que se fait la recombinaison (Leschziner and Grindley, 2003; Reed and Grindley, 1981). Les quatre recombinaisons coupent l'ADN en même temps et restent liées avec l'extrémité 5' de

manière covalente laissant deux bases débordantes, puis les brins d'ADN sont inversés par rotation. Une complémentarité entre les bases débordantes générées par le coupure double-brin est nécessaire pour permettre la ligation (Cozzarelli et al., 1984; Kanaar et al., 1990; Li et al., 2005; Stark et al., 1991; Wasserman et al., 1985). Le fait que la recombinaison se fasse par rotation des extrémités modifie la topologie de l'ADN et peut entraîner la formation de nœuds (Kanaar et al., 1990; Stark et al., 1991).

1.5.2.4.2 *Les protéines similaires aux topoisomérases (ou topoisomerase-like)*

Spo11 entraîne des cassures double-brins dans l'ADN lors de la méiose et reste lié de manière covalente à la partie 5' de l'ADN (Keeney et al., 1997). Tout comme la topoisomérase de type II, elle nécessite une tyrosine comme résidu catalytique et un domaine TOPPRIM (Bergerat et al., 1997). Une analyse de séquence a montré que Spo11 est l'homologue de la topoisomérase VIA présente chez les archéobactéries (Malik et al., 2007). Ces cassures générées par Spo11 sont essentielles afin d'induire des recombinaisons entre les chromosomes lors de la méiose (Boateng et al., 2013).

Une fois que Spo11 a coupé l'ADN et y est lié de manière covalente, afin de générer une extrémité 3' débordante qui va pouvoir envahir une région présentant une homologie, le complexe MRN clive l'ADN entraînant le départ de Spo11 avec un oligonucléotide (Neale et al., 2005). Cette extrémité débordante est reconnue par la machinerie de réparation par recombinaison homologue, ce qui conduit à l'invasion d'ADN double-brins par cette extrémité débordante.

Spo11 forme un complexe avec TopoVI qui régule son activité afin d'induire des cassures double-brins uniquement lors de la méiose, suggérant que Spo11 et TopoVI forment ensemble une topoisomérase de type II (Robert et al., 2016; Vrielynck et al., 2016). Il reste encore à déterminer si Spo11 et TopoVI peuvent elles-mêmes résoudre les cassures d'ADN qu'elles induisent (Robert et al., 2016; Vrielynck et al., 2016).

1.5.3 Les fonctions des topoisomérases

1.5.3.1 La transcription

Contrairement à ce qui se passe chez la levure *Saccharomyces cerevisiae*, les topoisomérases I, II et III α sont indispensables chez les eucaryotes supérieurs (Lee et al., 1993; Morham et al., 1996; Plank et al., 2005). Elles jouent un rôle essentiel lors de la réplication et la transcription de l'ADN en réduisant les surenroulements générés par les machineries de réplication et transcription. La topoisomérase I est recruté au niveau des gènes activement transcrits (Gilmour et al., 1986; Zhang et al., 1988) mais il n'est

pas clair si elle sert uniquement à permettre à la polymérase de progresser en résolvant les surpertours que sa progression génère ou si elle a une fonction dans la régulation de la transcription comme son interaction avec l'ARN polymérase suggère (Shaiu and Hsieh, 1998; Shykind et al., 1997). L'utilisation de la Camptothecin avec l'immunoprécipitation de la chromatine associée à la topoisomérase de type I, a permis de mettre en lumière que la topoisomérase est inactive au niveau des promoteurs mais active dans les gènes, montrant qu'elle est essentielle pour réduire le surenroulement généré par l'élongation (Baranello et al., 2016). Mais aux promoteurs, la topoisomérase I est présente, attendant d'être stimulée par la CTD phosphorylée de l'ARN polymérase, afin de permettre la progression de cette dernière comme le propose les auteurs (Baranello et al., 2016). À l'inverse de la topoisomérase I qui s'accumule le long des gènes, la topoisomérase II est présente en amont du site d'initiation et en aval du site de terminaison de la transcription afin de résoudre les supertours positifs générés (Achar et al., 2020).

1.5.3.2 La réplication de l'ADN

Lors de la réplication de l'ADN les topoisomérases ne se contentent pas de diminuer les surenroulements causés par la machinerie de réplication mais sont importantes dès l'initiation de la réplication et jusqu'à la séparation des chromatides sœurs (Figure 1-15c). L'assemblage du complexe de reconnaissance de l'origine de réplication (ORC) ne se fait pas de manière complètement aléatoire, chez *Saccharomyces cerevisiae* il y a une claire spécificité de séquence (Broach et al., 1983; Fujita et al., 1998), chez d'autres espèces telles que *D. melanogaster* ou chez les mammifères il y a une préférence pour les régions riches en AT et surenroulées négativement qui sont donc plus favorables à l'ouverture de la double hélice d'ADN (Chuang et al., 2002; Houchens et al., 2008; Remus et al., 2004; Vashee et al., 2003). Il serait faux de penser que les topoisomérases contribuent de manière passive à l'assemblage du complexe de réplication, en effet la topoisomérase II est essentielle pour l'assemblage de ce complexe et la topoisomérase I pour l'initiation de la réplication (Abdurashidova et al., 2007; Hu et al., 2009; Rampakakis and Zannis-Hadjopoulos, 2009). Tout comme lors de la transcription, la réplication de l'ADN génère l'apparition de supertours, en raison de l'ouverture de l'hélice d'ADN par les hélicases. Ce surenroulement doit être réduit afin de permettre la progression de la machinerie de réplication. Les topoisomérases I et II α résolvent les supertours positifs générés en aval de la fourche de réplication (Capranico et al., 2017; McClendon et al., 2005; Pommier et al., 2016). La résolution par la topoisomérase I qui réduit le surenroulement en faisant pivoter les brins d'ADN conduit à l'entrelacement des molécules d'ADN (ADN caténé ou héli-caténé) qui sont résolues par les topoisomérases II et III afin de permettre la séparation des chromatides sœurs lors de la mitose (Lee et al., 2019, 2013; Lucas et al., 2001; Suski and Marians, 2008). Les topoisomérases II et

III α sont également importantes lors de la fusion de fourche de réplication (Fachinetti et al., 2010; Mundbjerg et al., 2015).

1.5.3.3 La mitose

La mitose requiert la condensation des chromosomes, suivi de leur séparation entre les deux cellules filles. La condensation se fait par l'action des condensines qui vont compacter les chromosomes créant ainsi une accumulation de surenroulements positifs. La topoisomérase II, présente sur les chromosomes de manière active, est nécessaire pour la séparation des chromatides sœurs grâce à sa capacité à décaténer l'ADN (Figure 1-15c) (Giménez-Abián et al., 2000; Swedlow et al., 1993). Étonnamment la perte de la topoisomérase II conduit également à une décondensation des chromosomes (Lupo et al., 2001; Salceda et al., 2006; Vas et al., 2007). En compactant les chromosomes, les condensines induisent l'accumulation de supertours positifs, il a été proposé que ces supertours soient reconnues par la topoisomérase II qui les réduit afin de permettre une compaction plus importante des chromosomes (Baxter and Aragón, 2012; Baxter et al., 2011). La topoisomérase II joue donc deux fonctions lors de la mitose : en réduisant le surenroulement positif elle permet une compaction plus importante des chromosomes et sa fonction de décaténase permet leur séparation. Chez les organismes possédant deux topoisomérases II, c'est la topoisomérase II α qui agit sur les chromosomes, mais en cas de mutation, la topoisomérase II β peut partiellement se substituer (Sakaguchi and Kikuchi, 2004).

1.5.3.4 La recombinaison

Spo11 en complexe avec la topoisomérase VI induisent des cassures double-brins lors de la méiose, qui sont reconnues par les mécanismes de réparation des cassures dans l'ADN, et notamment de réparation par recombinaison homologue. Spo11 reste lié de manière covalente à l'extrémité 5', à la suite d'un clivage endonucléique par le complexe MRN, Spo11 part avec un oligonucléotide libérant une extrémité 3' débordante. La recombinaison se fait grâce aux protéines comme Rad51 qui induisent l'invasion d'ADN double-brins présentant une homologie de séquence par cette extrémité simple-brin. La fourche de Holliday ainsi formée peut migrer grâce à l'action d'hélicases telle que BLM, et pourra être résolue par la topoisomérase III qui peut décaténer deux molécules d'ADN simple-brin permettant ainsi la recombinaison entre deux chromosomes homologues (Cejka et al., 2012; Chen et al., 2012)

1.5.3.5 Les topoisomérases régulent la formation et la résolution de structures d'acides nucléiques non-canoniques

Le surenroulement négatif est à la fois une cause et une conséquence de la formation de R-loops (Drolet et al., 1994; Phoenix et al., 1997; Stolz et al., 2019). Ces dernières sont stabilisées en présence d'un ADN extrêmement négativement surenroulé, mais également par des séquences plus favorables que d'autres (Stolz et al., 2019). Les R-loops maintiennent la double hélice d'ADN ouverte en empêchant l'appariement du brin d'ADN déplacé avec celui impliqué dans l'hybride ARN-ADN. Cette ouverture de l'ADN double-brin, tout comme la réplication ou la transcription, induit la formation de supertours de chaque côté de la R-loop qui ne peuvent être complètement absorbés par l'ADN, en raison de la présence de protéines liant l'ADN qui empêchent la dissipation de supertours (Drolet et al., 1994). La topoisomérase I participe à réduire les R-loops pendant la transcription en réduisant le niveau de surenroulement et donc en prévenant leur formation (Manzo et al., 2018; Phoenix et al., 1997), tandis que la Topoisomérase III tout comme la RNase H1 ou RNase H2 participent à leur résolution une fois qu'elles ont été formées (Broccoli et al., 2000).

Les quadruplexes de G peuvent se former entre plusieurs molécules ou au sein d'une même molécule d'ADN, lorsque plusieurs guanines se suivent et forment des liaisons de Hoogsteen. Elles se forment notamment dans un contexte topologique favorable comme en présence de surenroulement négatif qui contribue à leur formation (Li et al., 2019; Sekibo and Fox, 2017; Sun and Hurley, 2009). Il a donc été proposé que la topoisomérase I joue un rôle essentiel en réduisant le niveau de surenroulement afin de prévenir de l'instabilité génomique qui peut découler de la formation de quadruplexes de G (Yadav et al., 2014). La topoisomérase I peut lier directement les quadruplexes de G, l'empêchant ainsi d'être recrutée au niveau de l'ADN surenroulé et donc inhibant son activité de topoisomérase (Arimondo et al., 2000).

Les topoisomérases, protéines conservées à travers l'évolution, jouent un rôle essentiel en réduisant le stress topologique et en décaténant des molécules d'ADN aussi bien lors de la transcription et de la réplication de l'ADN, de la mitose ou encore de la méiose. Il était connu depuis longtemps que les topoisomérases participent à la résolution de structures d'ADN non-canoniques, mais la découverte de nouvelles fonctions pour ces structures qui ne sont pas uniquement la cause d'instabilité génomique, ainsi que de la topoisomérase de type III β qui peut interagir avec des ARN, a participé au regain d'intérêt pour ces protéines.

1.6 Hypothèses et objectifs de la thèse

Les protéines du groupe Polycomb sont recrutées au niveau des PREs afin de maintenir la répression génique. Malgré la découverte des protéines du groupe Polycomb et des PREs il y a plusieurs dizaines d'années, des questions demeurent quant à leur fonctionnement. Les mécanismes permettant de déterminer le statut d'un PRE et comment ils peuvent maintenir à la fois la répression et l'activation de l'expression génique dépendamment du contexte développemental ne sont toujours pas élucidés. De plus, de nombreuses questions subsistent quant au recrutement des protéines du groupe Polycomb au niveau des PREs, qui ne doit se faire que dans le contexte développemental approprié, à la répression de la transcription par ces mêmes protéines et à la propagation de cette répression à travers le cycle cellulaire. Le fait que les PREs puissent être transcrits en ARN, et que la transcription d'un même PRE puisse corrélérer avec l'expression ou la répression génique dépendamment de l'orientation de la transcription (Herzog et al., 2014), nous a poussé à nous intéresser aux ARN produits aux PREs. Nous avons formulé l'hypothèse que ces ARN puissent donner la spécificité de contexte afin d'induire le recrutement des protéines du groupe Polycomb uniquement dans le contexte développemental approprié. L'orientation de la transcription d'un PRE pourrait déterminer si celui-ci forme une R-loop lorsqu'il est transcrit dans un sens et pas dans l'autre, puisque leur stabilité dépend en partie de la séquence d'ADN, et ces R-loops pourraient alors participer au recrutement des protéines du groupe Polycomb. La possible formation de R-loops au niveau des PREs et l'étude de leur fonction dans le recrutement des protéines du groupe Polycomb fait l'objet du premier objectif. La recherche de nouvelles activités biochimiques chez les complexes Polycomb PRC2 et PRC1 afin de mieux comprendre comment ces complexes protéiques sont recrutés aux PREs uniquement dans le contexte approprié et comment ils maintiennent la répression génique font l'objet des objectifs 2 et 3.

1.6.1 Objectif 1 (Chapitre 2)

Dans l'objectif 1, nous avons identifié les séquences formant des R-loops par DRIP-seq chez *D. melanogaster* dans deux stades de développement embryonnaire ainsi que dans une lignée cellulaire. Nous avons trouvé que les R-loops s'accumulent aux PREs et à des sites liés par les protéines du groupe Polycomb en dehors des PREs. Nous avons également montré que les PREs ayant formés des R-loops ont une plus forte probabilité d'être liés par PRC1 et PRC2 ainsi que d'être tri-méthylé par ce dernier. Nous avons également observé que les PREs ayant formé des R-loops lors du stade de développement

embryonnaire plus précoce ont une plus forte probabilité d'être liés par les protéines du groupe Polycomb au stade de développement embryonnaire plus tardif. Les R-loops semblent donc être impliqués dans le recrutement des protéines du groupe Polycomb aux PREs *in vivo*. Nous avons validé cette observation *in vitro*, en montrant que PRC1 et PRC2 lie préférentiellement une R-loop par rapport à l'ADN double-brin.

1.6.2 Objectif 2 (Chapitre 2)

Dans l'objectif 2, nous avons voulu déterminer si PRC2 pouvait influencer la formation des R-loops. Nous avons identifié une nouvelle activité d'invasion de brins pour le complexe PRC2 qui conduit à la formation d'hybride ARN-ADN, la partie principale d'une R-loop, lorsque nous lui avons fourni un ADN double-brin et son ARN complémentaire. La suite de l'objectif 2 a consisté à la caractérisation de cette activité.

1.6.3 Objectif 3 (Chapitre 3)

Nous avons observé que PRC1 modifie la topologie d'un plasmide *in vitro*. Ceci nous a poussé à déterminer si PRC1 possède une activité lui permettant d'agir comme une topoisomérase dans l'objectif 3. Nous avons donc identifié une nouvelle activité similaire à celle de la topoisomérase de type I chez PRC1 et sa sous-unité PSC. Nous avons trouvé que PRC1 ainsi que la région C-terminale de la protéine PSC, peuvent relaxer un plasmide surenroulé négativement et ajouter des supertours négatifs à un plasmide relaxé. Nous avons déterminé que cette activité de topoisomérase ne se limite pas à l'ADN nu mais que PRC1 peut relaxer un plasmide assemblé en chromatine. Nous avons également observé que *in vitro* la relaxation d'un ADN négativement surenroulé et ayant formé des R-loops, par PRC1 conduit à la résolution de ces structures d'ADN non-canoniques.

2 Article 1: RNA-DNA strand exchange by the *D. melanogaster* Polycomb complex PRC2

RNA-DNA strand exchange by the *D. melanogaster* Polycomb complex PRC2

Authors:

Célia Alecki^{1,2}, Victoria Chiwara¹, Lionel A. Sanz³, Daniel Grau⁴, Osvaldo Arias Pérez^{1,5}, Elodie L. Boulier¹, Karim-Jean Armache⁴, Frédéric Chédin³, and Nicole J. Francis^{1,2,6*}

Affiliations:

¹Institut de recherches cliniques de Montréal, 110 Avenue des Pins Ouest, Montréal, QC, H2W 1R7, Canada

²Département de biochimie et médecine moléculaire Université de Montréal, 2900 Boulevard Edouard-Montpetit, Montréal, QC H3T 1J4, Canada.

³Department of Molecular and Cellular Biology and Genome Center, 1 Shields Avenue, University of California, Davis, Davis, CA 95616, USA.

⁴Skirball Institute of Biomolecular Medicine, Department of Biochemistry and Molecular Pharmacology, New York University School of Medicine, New York, NY 10016, USA.

⁵Natural Sciences and Engineering Postgraduate; Universidad Autonoma Metropolitana, Cuajimalpa, Mexico City, Mexico

⁶Division of Experimental Medicine, McGill University, 1001 Decarie Boulevard, Montreal, QC H4A 3J1, Canada

*correspondence to: nicole.francis@ircm.qc.ca

Author contributions: Conceptualization: N.J.F. & C.A.; Investigation: C.A., O.A. & N.J.F.; Formal analysis: C.A. & N.J.F.; Resources: V.C., E.L.B., D.G., K.J.A., L.A.S., F.C.; Writing: N.J.F., C.A., K.J.A., F. C., Supervision: N.J.F., Funding acquisition: N.J.F., F.C., K.J.A.

2.1 Abstract

Polycomb Group (PcG) proteins form memory of transient transcriptional repression that is necessary for development. In *Drosophila*, DNA elements termed Polycomb Response Elements (PREs) recruit PcG proteins. How PcG activities are targeted to PREs to maintain repressed states only in appropriate developmental contexts has been difficult to elucidate. PcG complexes modify chromatin, but also interact with both RNA and DNA, and RNA is implicated in PcG targeting and function. Here we show that R-loops form at many PREs in *Drosophila* embryos, and correlate with repressive states. In vitro, both PRC1 and PRC2 can recognize R-loops and open DNA bubbles. Unexpectedly, we find that PRC2 drives formation of RNA-DNA hybrids, the key component of R-loops, from RNA and dsDNA. Our results identify R-loop formation as a feature of *Drosophila* PREs that can be recognized by PcG complexes, and RNA-DNA strand exchange as a PRC2 activity that could contribute to R-loop formation.

2.2 Main Text

2.2.1 Introduction

During *Drosophila* embryogenesis, transiently expressed transcription factors activate homeotic (*Hox*) genes in certain regions of the embryo and repress them in others to dictate the future body plan (Maeda and Karch, 2009). Polycomb Group (PcG) proteins form a memory of these early cues by maintaining patterns of *Hox* gene repression for the rest of development (Maeda and Karch, 2009; Simon, 1995; Steffen and Ringrose, 2014). This paradigm for transcriptional memory is believed to be used by the PcG at many genes in *Drosophila*, and to underlie the conserved and essential functions of PcG proteins in cell differentiation and development from plants to mammals (Grossniklaus and Paro, 2014; Prezioso and Orlando, 2011). Polycomb Response Elements (PREs) are DNA elements that can recruit PcG proteins, but they also recapitulate the memory function of the PcG—when combined with early acting, region-specific enhancers in transgenes, they maintain transgene repression in a PcG-dependent manner only in regions where the early enhancer was not active (Kassis and Brown, 2013; Poux et al., 1996; Steffen and Ringrose, 2014). PREs contain a high density of binding sites for transcription factors that can recruit PcG proteins through physical interactions (Kassis and Brown, 2013). However, the widespread expression, binding pattern, and properties of factors that bind PREs cannot explain how PREs can exist in alternate, transcription-history dependent states to maintain restricted patterns of gene expression, or how they can switch between states (Steffen and Ringrose, 2014). Furthermore, DNA sequences with PRE-like properties have been difficult to identify in other species (Bauer et al., 2016; Kassis and Brown, 2013; Xiao et al., 2017) despite the conservation of PcG complexes, their biochemical activities, and their critical roles in development.

RNAs may provide context specificity to PcG protein recruitment and function. Some PREs, and some PcG binding sites in mammalian and plant cells, are transcribed into ncRNA, while others reside in gene bodies, and thus are transcribed when the gene is expressed (Hekimoglu and Ringrose, 2009; Herzog et al., 2014). Both the direction and level of transcription have been correlated with the functional state of PREs (Hekimoglu and Ringrose, 2009; Herzog et al., 2014; Ringrose, 2017). The PcG complex Polycomb Repressive Complex 2 (PRC2) has a well-described high affinity for RNA (Davidovich and Cech, 2015; Davidovich et al., 2015; Davidovich et al., 2013; Wang et al., 2017a; Wang et al., 2017b). RNA is suggested to recruit PRC2 to specific chromatin sites (Davidovich and Cech, 2015), but RNA binding can also compete for chromatin binding and inhibit PRC2 activity (Beltran et al., 2019; Beltran et al., 2016; Herzog et al.,

2014; Wang et al., 2017b; Zhang et al., 2019). One way for RNA to interact with the genome is by the formation of R-loops, three-stranded nucleic acid structures formed when an RNA hybridizes to a complementary DNA strand, thereby displacing the second DNA strand (Santos-Pereira and Aguilera, 2015). R-loops have been linked to regulation of transcription and chromatin previously, through a variety of mechanisms (reviewed in (Crossley et al., 2019; Garcia-Muse and Aguilera, 2019)). This includes links to PcG regulation in mammalian cells. The formation of R-loops over genes with low to moderate expression is associated with increased PcG binding and H3K27 trimethylation (H3K27me3) in human cells (Sanz et al., 2016) and R-loops have recently been implicated in promoting PRC1 and PRC2 recruitment in mammalian cells (Skourti-Stathaki et al., 2019), although other evidence suggests they antagonize recruitment of PRC2 (Chen et al., 2015). We hypothesized that R-loop formation could biochemically link RNA to PcG-mediated silencing through PREs and tested this idea in the *Drosophila* system.

We identify R-loop forming sequencing in *Drosophila* embryos and S2 cells and observed ~25% of PREs form R-loops. Interestingly PREs which have form R-loop are more likely to be bound by PcG proteins compared to PRE which have not form R-loop, suggesting that R-loop may be involve in PcG targeting. In vitro PRC1 and PRC2 recognize R-loop and open DNA-bubble. In vitro, when provided dsDNA and RNA PRC2 induces the formation of RNA-DNA hybrid, the main component of an R-loop. In conclusion, it suggests a mechanism in which RNA are involved in the targeting of PcG proteins via the formation of R-loop induced by PRC2 RNA-DNA strand exchange activity.

2.2.2 Results

R-loops form at many PREs in *Drosophila* embryos and cells

To determine whether R-loops form at PREs, we carried out two biological replicates of strand-specific DNA-RNA Immunoprecipitation followed by next generation sequencing (DRIP-seq) in *Drosophila* embryos (2-6 and 10-14 hour (H)) and in S2 cells (Fig. 2-1, Supplementary Figure 2-1). DRIP-seq peaks called relative to both input and RNase H-treated control samples and present in both replicates were analyzed. 10 positive sites were validated by DRIP-qPCR (Supplementary Figure 2-1b). Nearly 3/4 of R-loops formed over annotated genes (Supplementary Figure 2-1). R-loops were observed over genes encompassing all levels of transcription, although a majority were associated with genes with no or low levels of expression (Supplementary Figure 2-2a, b). Most R-loops formed with the strandedness expected from annotated

transcripts (Fig. 2-1 a, c, Supplementary Figure 2-2c), as observed in other species (Hartono et al., 2018; Sanz et al., 2016; Xu et al., 2017).

We detected R-loops at 22-33% of PREs (Fig. 2-1 a-c, Supplementary Figure 2-1d, 2-2 a-d, 2-3a-c). R-loops at PREs in embryos were more likely to have an antisense orientation to annotated transcripts, than total R-loops (Supplementary Figure 2-2c). PREs which form R-loops were also more likely to overlap with RNA polymerase II than PREs which do not form R-loops (Supplementary Figure 2-2d).

To test whether R-loops are related to the functional state of PREs, we used publicly available ChIP-seq datasets to compare PcG protein binding over PREs as a function of R-loop formation in each of our three samples. For each PcG protein tested, the median read density for PcG proteins was higher over PREs with R-loops than that for PREs without R-loops (Fig. 2-1d-f, Supplementary Figure 2-2f-h). The binding of two proteins implicated in PcG recruitment, Dsp1 and GAF, and both repressive and active states of PREs (Decoville et al., 2001; Dejardin et al., 2005; Kassis et al., 2017), were also higher at R-loop-positive PREs. Although binding of PcG proteins to PREs is necessary for their repressive function, it may not be sufficient, since analyses of PcG protein binding at a small number of PREs in the ON and OFF states did not detect differences in PRC1 or PRC2 binding (Langlais et al., 2012; Papp and Muller, 2006). Instead, histone modifications at and around PREs are correlated with the functional state so that PREs in the OFF state are marked with H3K27me3 (Papp and Muller, 2006). In both developing embryos and S2 cells, H3K27me3 density was higher at PREs with R-loops than at those without R-loops (Fig. 2-1e). H3K27Ac, a mark of the active state, was found at a small number of PREs, but correlated weakly with the presence of R-loops (Supplementary Fig. 2-2e, i-k, 2-3d-f). A small fraction of R-loops that are present at PREs in early stage embryos are absent in the later stage (Fig. 2-1c), suggesting that some R-loops at PREs are developmentally regulated. Developmental dynamics of R-loops at PREs are likely underestimated in our experiments because we used whole embryos. To test whether transient presence of an R-loop at PREs predicts the repressed state, we identified PREs that form R-loops early (2-6H) but not late (10-14H) in embryos and interrogated the levels of H3K27me3 in later embryonic states (12-16H). PREs that formed R-loops in early embryos had a higher density of H3K27me3 at subsequent developmental stages than PREs that did not form R-loops at either stage (Fig. 2-1f); these PREs were not enriched for H3K27Ac ($p=0.0885$).

PRC1 and PRC2 bind R-loops and open DNA bubbles *in vitro*

To understand biochemically how R-loops could promote the repressive state of PREs, we turned to *in vitro* assays. Recruitment of PRC1 and PRC2 to some sites in mammalian cells has recently been linked to the presence of R-loops (Skourti-Stathaki et al., 2019) so that we wondered if either complex might recognize R-loops. To measure the relative affinities of PRC1 and PRC2 for different nucleic acid substrates, we prepared dsDNA, R-loop or open DNA bubble templates from synthetic oligonucleotides corresponding to a sequence in the vg PRE (Supplementary Table 2-1), PRC2, and PRC1 (Supplementary Figure 2-4a, b) and used them in EMSA experiments. PRC1 lacking the Ph subunit (PRC1 Δ Ph) was used for these assays because this complex can be isolated in larger amounts. Our previous work indicates that PRC1 with and without Ph behave similarly in DNA and chromatin binding experiments, where binding is largely dependent on the C-terminal region of PSC (Francis et al., 2004). EMSA experiments showed that PRC2 and PRC1 bind more tightly to R-loops or open DNA bubble templates than dsDNA (Fig. 2-2a, b). The fact that binding of PcG complexes produce a “well-shift” rather than a discrete band make quantification of these experiments imprecise. We therefore used filter binding with the same templates to measure binding (Fig. 2-2c, d, Supplementary Figure 2-5a, b). Under our conditions, both PRC1 Δ Ph and PRC2 bind more strongly (~8x and ~23X, respectively) to an R-loop or an open DNA bubble (~10X and ~40X, respectively) than to dsDNA. For PRC1 Δ Ph the K_d measured with R-loop and open DNA bubble substrates are close to the probe concentration. We were unable to lower the probe concentration due to limitations on the sensitivity of detection. Therefore, these K_ds should be regarded as upper limits. Because of this limitation, the difference between the R-loop and open DNA bubble could be larger than what we measure. We conclude that PRC2 and PRC1 recognize R-loops, as well as an open DNA bubble, suggesting that PcG complexes recognize structured DNA or ssDNA and not the RNA-DNA hybrid part of an R-loop.

PRC2 induces RNA-DNA strand exchange

Because interactions between PRC2 and RNA are widely implicated in its regulation and function, we wondered if PRC2 might influence R-loop formation. We titrated PRC2 into reactions with radio- or fluorescently-labelled RNA and the corresponding linear dsDNA (Fig. 2-3a-c). We observed a PRC2 dose-dependent appearance of an RNA species that migrates at the position of dsDNA (Fig. 3d, e, g, Supplementary Figure 2-6a, b). These putative RNA-DNA hybrids formed with either the sense or anti-sense RNA, but not with a non-complementary RNA, indicating that base pairing between RNA and DNA is required (Fig. 2-3d-g).

To confirm that the PRC2 reaction products indeed contain RNA-DNA hybrids, we tested their nuclease sensitivity: the RNA band that migrates at the position of dsDNA was fully degraded by RNase H and resistant to RNase A (Fig. 2-3h, i, Supplementary Figure 2-6b). We also tested if the S9.6 antibody can recognize the PRC2 reaction products, as expected if they contain RNA-DNA hybrids. RNA-DNA strand exchange assays were carried out with or without RNA, the purified products were incubated with S9.6 antibody-coupled magnetic beads, and the isolated DNA analyzed on agarose gels. We observe that nucleic acids were efficiently immunoprecipitated by S9.6 antibody only when RNA was included in the reaction (Supplementary Figure 2-6c, d). We conclude that PRC2 mediates RNA-DNA strand exchange when incubated with RNA and dsDNA

Using fluorescently labelled RNAs, we estimated the extent of RNA-DNA hybrid formation. By the end of a 60-minute reaction containing 3 fmol of linear DNA and 1.9 fmol of RNA, close to 40% of the DNA had undergone strand exchange with the RNA (Supplementary Figure 2-7a, b). RNA-DNA strand exchange induced by PRC2 required $MgCl_2$ but not nucleotide addition (Supplementary Figure 2-7c, d). To confirm that the ATP-independence of the reaction did not reflect contaminating ATP in the RNA or PRC2 preparations, we treated each with apyrase prior to carrying out RNA-DNA strand exchange assays and the results were unchanged (Supplementary Figure 2-7e-g).

To confirm that RNA-DNA hybrid forming activity is specific to PRC2, we tested two control proteins, the transcription factor NFY and the PcG protein Sxc. Neither of these proteins induced formation of RNA-DNA hybrids, although they bind both DNA and RNA (Supplementary Figure 2-8a-e). RNA-DNA hybrid formation activity also co-fractionated with PRC2 through size exclusion chromatography (Supplementary Figure 2-8f-h).

Cellular nucleases are common contaminants when purifying chromatin-associated enzymes; the presence of nuclease contaminants in preparations of PRC2 could promote RNA-DNA hybrid formation by exposing single-stranded DNA and enabling spontaneous annealing with complementary RNA. To address this possibility, we performed three experiments. First, we incubated phosphorylated ds and ssDNA oligonucleotides with PRC2, or a series of commercially available endo- and exonucleases. Oligonucleotides were then analyzed on denaturing acrylamide gels, which were stained with SYBRGold to visualize degradation products. While oligonucleotides were fully degraded across the nuclease titrations, we did not detect degradation products after incubation with PRC2 (Supplementary Figure 2-9a-f). We also tested whether exonuclease treatment can lead to RNA-DNA hybrid formation under experimental conditions used for PRC2 (Supplementary Figure 2-9g-j). Exonuclease III treatment led to

RNA-DNA hybrid formation, but this required enzyme concentrations that clearly degrade ds and ssDNA oligonucleotides (Supplementary Figure 2-10f, j).

Second, we reasoned that if our PRC2 preparations contained nuclease activity, treatment of dsDNA with PRC2 should expose long stretches of ssDNA that could form filaments with single strand DNA binding protein (SSB) that would be visible in electron microscopy (EM) (Fig 2-4a-c). We incubated linear DNA with PRC2 or exonuclease III using the same experimental conditions leading to RNA-DNA hybrid formation, purified the DNA, incubated it with SSB, and visualized the samples by negative stain EM. SSB-coated ssDNA filaments were clearly visible in DNA samples pre-treated with exonuclease III (Fig. 2-4b) but not in DNA samples treated with PRC2 (Fig. 2-4c).

Finally, to functionally test whether PRC2 nuclease contaminants in PRC2 could account for RNA-DNA hybrid formation, we pre-treated DNA templates with PRC2, and used them in RNA-DNA hybrid forming assays in the presence or absence of PRC2 (Fig. 2-4d). If nuclease activity in PRC2 preparations exposes ssDNA that allows formation of RNA-DNA hybrids, the pre-treated templates should form RNA-DNA hybrids without further requirement for PRC2. However, we detected RNA-DNA hybrids only when PRC2 was added during the RNA-DNA strand exchange reaction, and not in samples in which the DNA was pre-treated with PRC2 (Fig. 2-4e-f). We conclude that contaminating nuclease activity in PRC2 preparations cannot explain PRC2 induced RNA-DNA hybrid formation and therefore that formation of these RNA-DNA hybrids was directly catalyzed by PRC2.

Substrate requirements for RNA-DNA strand exchange

To determine the DNA and RNA substrate requirements for RNA-DNA strand exchange, we tested DNA templates with different ends (4 or 2 base pair (bp) 5' or 4 bp 3' overhang, or blunt end), prepared by digestion with different restriction enzymes. In all cases, the RNAs used overlap the end of the DNA (Fig. 2-5a-c). Similar levels of RNA-DNA strand exchange were observed with all types of DNA ends (Supplementary Figure 2-10a, b, e). However, templates for which the RNA was internal to the DNA did not lead to hybrid formation (Fig. 2-5a-c and Supplementary Figure 2-10c, d).

The above experiments suggested that the overlap of the RNA with the ends of the dsDNA is important for PRC2-mediated RNA-DNA strand exchange. To further analyze this, we digested DNA templates by EcoRI such that the resulting products have 8 non-complementary bases prior to the start of the RNA on the bottom DNA strand, and 4 on the top strand (Fig. 2-5 d-j). For each of 3 pairs of sense-antisense RNA-

DNA combinations, RNA-DNA strand exchange is only observed when the RNA is complementary to the top strand (i.e. with 4 unmatched bases rather than 8). Together, these experiments suggest near complete overlap between the RNA and DNA ends is important for PRC2-mediated RNA-DNA strand exchange, implying that the reaction initiates at the DNA end, but that a specific DNA end structure is not required.

PRC2-mediated RNA-DNA strand exchange could require binding to DNA, to RNA, or to both. Detailed analyses of PRC2 binding to nucleic acids and chromatin are consistent with PRC2 making multiple contacts with both substrates (Long et al., 2017; Zhang et al., 2019), while functional assays are consistent with a single binding site that can bind chromatin, DNA or RNA, but has highest affinity for RNA so that RNA can compete for binding to DNA or chromatin (Beltran et al., 2016; Herzog et al., 2014; Wang et al., 2017b). To understand the role of RNA and DNA interactions in PRC2-mediated RNA-DNA strand exchange, we titrated each substrate and changed the order of addition in the reaction. Addition of RNA prior to DNA inhibited the reaction at low concentrations of PRC2. RNA-DNA strand exchange increased with increasing RNA concentration and decreased with increasing DNA concentration (Fig. 2-6a-f, Supplementary Figure 2-11a-c).

While the exact mechanism by which PRC2 induces RNA-DNA strand exchange remains unknown, one step must be the annealing of RNA and DNA. We tested whether PRC2 enhances annealing of RNA and ssDNA and find that PRC2 can induce RNA-DNA hybrids by annealing of single-stranded oligonucleotides. PRC2 can also anneal ssDNA to form dsDNA (Supplementary Figure 2-12). Unlike RNA-DNA strand exchange, RNA-DNA oligo annealing did not require $MgCl_2$.

2.2.3 Discussion

The demonstration that PRC2 induces the formation of RNA-DNA hybrids in vitro, that PRC2 and PRC1 recognize R-loops in vitro, and that R-loops are present at PREs in vivo suggest a mechanistic model for how RNAs could induce or maintain the OFF state of PREs (Fig. 2-7). If PREs (or the gene they control and in many cases are embedded in) are highly transcribed, the RNA could compete for PRC2 binding to chromatin, as has been demonstrated in vitro and in vivo (Beltran et al., 2019; Beltran et al., 2016; Wang et al., 2017b) (Fig. 2-7d). However, a lower level of transcription through a PRE (or transcription in an orientation that is favorable for R-loop formation) could allow R-loops to form, possibly via the RNA-DNA hybrid forming activity of PRC2 (Fig. 2-7a, c). R-loop formation will repress additional RNA production by

preventing RNA polymerase passage (Garcia-Muse and Aguilera, 2019) allowing recruitment of additional PRC2 (by PRE-binding transcription factors or interactions with other PcG proteins (Kang et al., 2015; Sneppen and Ringrose, 2019)) and its retention on the chromatin. PRC2 could then modify histones to maintain a repressive chromatin state (Fig. 2-7c). The R-loop, in conjunction with H3K27me3 and PRE-binding transcription factors, would also promote binding of PRC2 and PRC1 (Fig. 2-7b, c). R-loops may also interfere with binding or function of proteins that promote the active state of PREs, although this remains to be tested. Our data indicate that both coding and ncRNAs form R-loops. The regulation of these RNAs and therefore of R-loops could provide transcriptional memory and developmental context specificity to PcG recruitment by transcription factors that constitutively recognize PREs. A conceptually similar model for how high levels of RNA production at PREs could promote the ON state and low levels the OFF state was proposed previously (Ringrose, 2017); R-loop formation provides one mechanism by which it can occur. Although this model is highly speculative at this time, it integrates many observations, and provides testable hypotheses.

Observations in *Drosophila* are also consistent with a possible connection between R-loops and PcG function. The helicase Rm62 interacts genetically with both PcG and TrxG genes, and colocalizes with the PRE binding protein Dsp1 on polytene chromosomes (Lamiable et al., 2010). Rm62 is the *Drosophila* homologue of the DDX5 helicase, which can unwind RNA-DNA hybrids in vitro and is implicated in R-loop resolution in vivo (Mersaoui et al., 2019). A recent genome-wide RNAi screen for TrxG interacting genes (which should antagonize PcG function) identified the gene for RNaseH1 (Umer et al., 2019). RNA has been suggested to be important in switching PREs between OFF and ON states (Herzog et al., 2014; Rank et al., 2002; Ringrose, 2017; Schmitt et al., 2005), although this has been contested by experiments aiming to test whether transcription through a PRE can switch it to the active state (Erokhin et al., 2015; Kassis and Muller, 2015). Resolution of R-loops by cellular RNases or RNA-DNA helicases could contribute to switching PRE states, which will be intriguing to test. It is also likely that even in the simple model suggested in Fig. 2-7, the levels of RNA corresponding to “low” and “high”, and the strength of the effect will depend both on the genomic context and the sequences of the RNAs that are produced.

R-loop formation is observed at ~30% of PREs; these may represent a specific class of PREs. Most R-loops are believed to form co-transcriptionally, so that R-loops would be predicted to depend on PRE transcription. Indeed, >70% of R-loops formed at PREs overlap an annotated coding or non-coding RNA, and PREs with R-loops are more likely to have RNA Pol II signal in ChIP-seq experiments. However, ~67% of PREs where we did not observe R-loops also overlap an annotated transcript. Further, a fraction of PREs

with R-loops (and a fraction of total R-loops) either do not overlap any annotated transcripts, or overlap a transcript in the opposite orientation as the R-loop. While some of these discrepancies likely reflect incomplete annotation of rare transcripts, they raise the intriguing possibility that the RNA used to form the R-loops could be supplied in trans. Careful analysis of the RNA component of R-loops at PREs will be needed to resolve this. Although speculative at this time, the ability of PRC2 to induce RNA-DNA hybrids could contribute to non-co-transcriptional R-loop formation.

We find that PRC2 can induce RNA-DNA strand exchange from RNA and linear dsDNA *in vitro*. A small number of other proteins have been shown to have similar activity, using various types of substrates. These include the repair proteins Rad52/RecA (Kasahara et al., 2000; Mazina et al., 2017; McDevitt et al., 2018; Zaitsev and Kowalczykowski, 2000) and PALB2 (Deveryshetty et al., 2019), the human Capping Enzyme (CE) (Kaneko et al., 2007), the viral protein ICP8 (Boehmer, 2004) and the telomere binding protein TRF2 (Lee et al., 2018). Like the activity of PRC2, none of these reactions require ATP hydrolysis (although R-loop formation by RecA is stimulated by ATPγS (Kasahara et al., 2000)), and most use linear DNA substrates (Deveryshetty et al., 2019; Kaneko et al., 2007; Mazina et al., 2017; McDevitt et al., 2018) or an unpaired or ssDNA region (Kasahara et al., 2000; Zaitsev and Kowalczykowski, 2000). The exceptions are TRF2 and ICP8. ICP8 can mediate R-loop formation from an RNA and a supercoiled plasmid (Boehmer, 2004). TRF2 stimulates invasion of RNA oligos into a supercoiled plasmid encoding a telomeric DNA array (Lee et al., 2018), but the mechanism is believed to be induction of positive supercoiling by TRF2 that facilitates DNA unwinding and RNA invasion (Amiard et al., 2007). RNA-DNA strand exchange has been investigated most closely for Rad52, and its homologue RecA (Kasahara et al., 2000; Zaitsev and Kowalczykowski, 2000). Rad52 has been shown both to carry out “inverse strand exchange” where Rad52 first binds the dsDNA, allowing RNA strand exchange (Mazina et al., 2017), and to use an RNA-bridging mechanism, in which Rad52 first binds the RNA, and can bridge two dsDNA fragments by forming RNA-DNA hybrids with segments of each of them (McDevitt et al., 2018). Both of these mechanisms are candidates to mediate RNA-mediated repair of DSBs (Mazina et al., 2017; McDevitt et al., 2018). PRC2 requires a DNA end for RNA-DNA strand exchange *in vitro*; for this activity to occur *in vivo*, either a DNA break would be required, or PRC2 would need to be able to use DNA opened by (an)other factors, or by transcription. These requirements may limit PRC2 strand exchange activity at PREs. In order to fully understand the impact of this activity *in vivo* and to what extent PRC2 contributes to R-loop formation at PREs, additional experiments will be necessary. Interestingly, Topoisomerase II interacts with a subunit of PRC1, colocalizes with PcG proteins in the BX-C, and is implicated in PRE-mediated silencing (Lupo et al., 2001); transient Topo II induced breaks have been implicated in regulation of transcription and chromatin

compaction (Calderwood, 2016; Wong et al., 2018), and could also be used by PRC2. It is also possible that the activity of PRC2 contributes to RNA-DNA strand exchange at DNA breaks where RNA-DNA hybrids have been shown to form (Ohle et al., 2016) and where PRC2 is recruited (Campbell et al., 2013; Chou et al., 2010).

The connection between RNA and PRC2 has been recognized for some time, in species from plants to humans (Berry and Dean, 2015; Davidovich and Cech, 2015; Herzog et al., 2014; Ringrose, 2017), but mechanisms beyond RNA binding by PRC2 have not previously been described. Our discovery of PRC2-mediated RNA-DNA strand exchange, suggests one mechanism to connect RNA to PcG targeting and function.

2.3 Figures

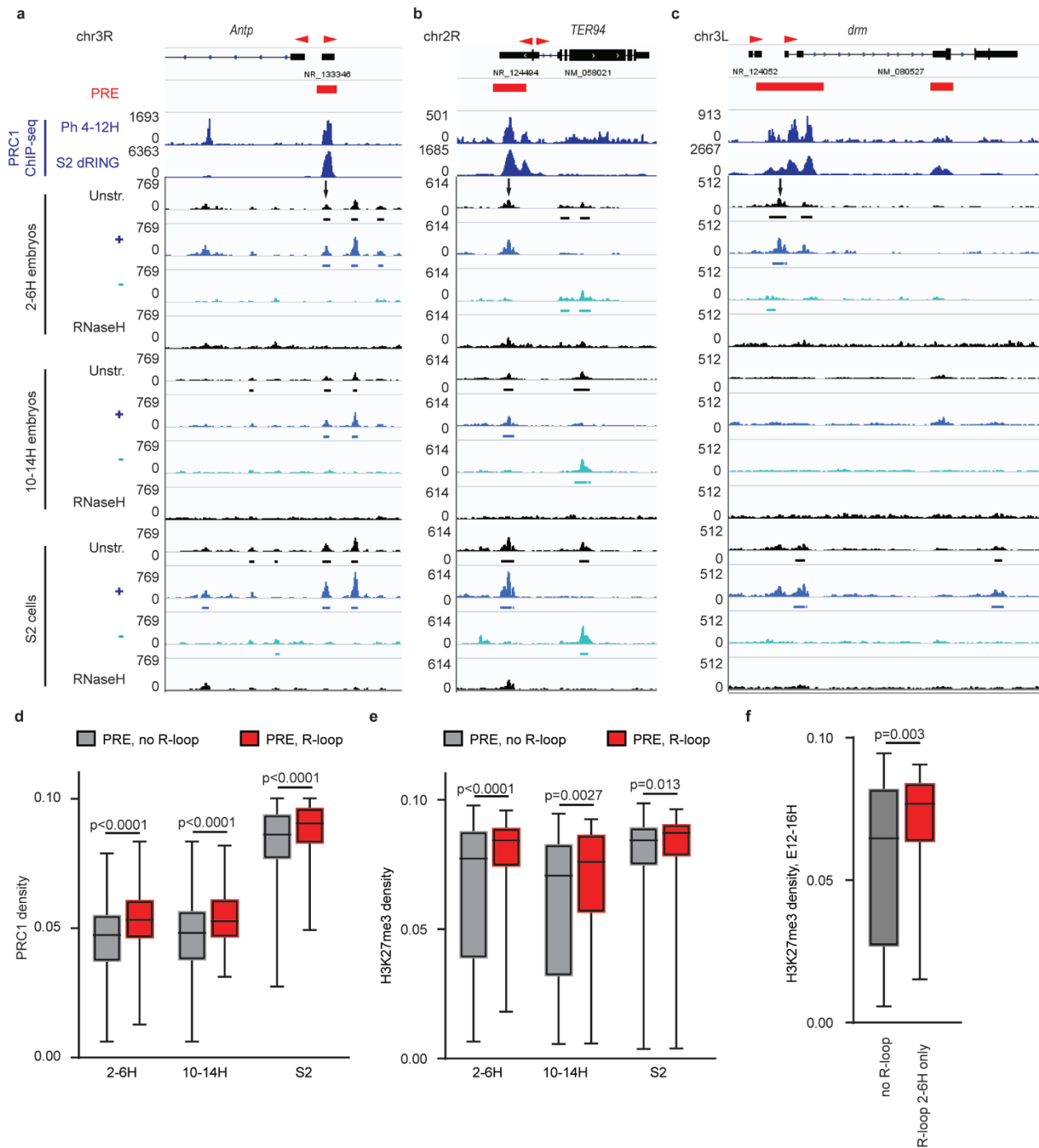


Figure 2-1 R-loops form at *Drosophila* PREs and correlate with a repressed state

a-c. DRIP-seq traces showing R-loop formation at PREs bound by PRC1 components (arrows) in 2-6H and 10-14H *Drosophila* embryos (Ph), and in S2 cells (dRING). RNaseH-treated samples are negative controls. "Unstr" indicates all R-loops, while + and - indicate strand specific tracks; direction refers to the DNA in the RNA-DNA hybrid. Called peaks are indicated under the traces. Red arrowheads above genes indicate direction of annotated transcripts. d, e. Median normalized intensity of PRC1 components (d), or H3K27me3 (e) over PREs with or without R-loops. 2-6H and 10-14H R-loop data are compared with Ph at 4-12H and

H3K27me3 at 4-8H and 12-16H respectively. S2 cell R-loop data are compared with dRING. Whiskers show min. to max. f. Median normalized intensity of H3K27me3 at 12-16H over PREs where R-loop formation is detected in 2-6H but not in 10-14H compared to PREs with no R-loops detected at either stage. p-values are for two-tailed Mann-Whitney tests. See also Supplementary Figure 2.

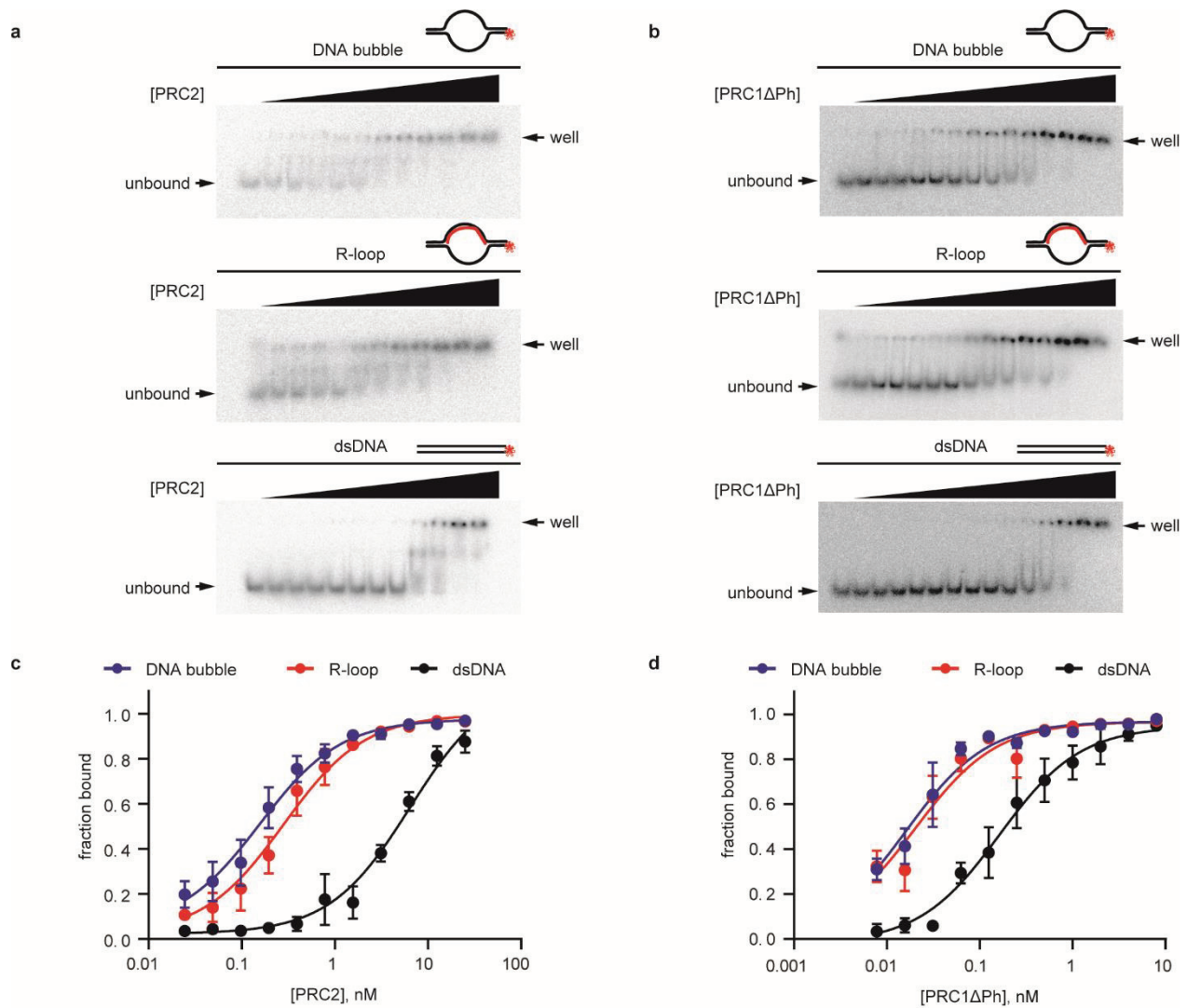


Figure 2-2 PRC1 and PRC2 recognize R-loops

a, b. PRC2 (a) and PRC1ΔPh (b) binding to open DNA bubble, R-loop and dsDNA oligonucleotides measured by EMSA. [PRC2]=0.02-25 nM for DNA bubble, R-loop and dsDNA and [PRC1ΔPh]=0.02-16 nM (active concentration). c. PRC2 binding to DNA bubble, R-loop and dsDNA oligonucleotides measured by filter binding assay. [PRC2]=0.02-25 nM, n=3. Points show the mean +/- S.E.M. Apparent Kd(95% CI): DNA bubble=0.16 nM (0.11-0.23); R-loop=0.28 nM (0.20-0.41); dsDNA=6.44 (4.26-9.99). d. PRC1ΔPh binding to DNA bubble, R-loop and dsDNA oligonucleotides measured by filter binding assay. [PRC1ΔPh]=0.08-8 nM, n=3. Points show the mean +/- S.E.M. Apparent Kd(95% CI): DNA bubble=0.016 nM (0.012-0.022); R-loop=0.021 nM (0.014-0.031); dsDNA=0.164 (0.106-0.254). Source data are provided in Source Data file.

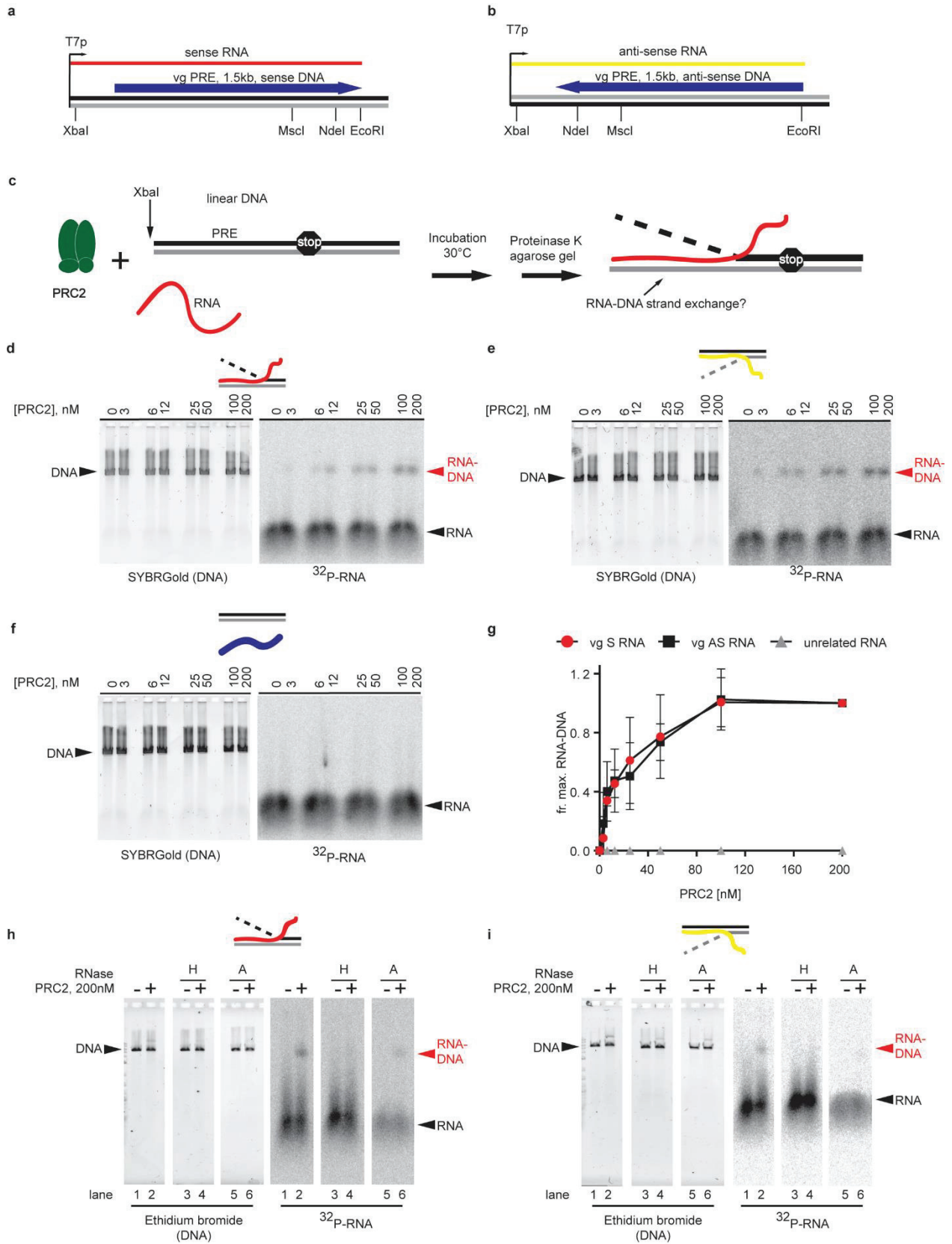


Figure 2-3 PRC2 has RNA-DNA strand exchange activity

a, b. Nucleic acids used for RNA strand exchange. Plasmids with the vestigial (vg) PRE in either orientation are transcribed to produce sense (a) or anti-sense RNAs (b). Linearized plasmid is used as the DNA template. "Stop" indicates where the RNA ends relative to the DNA. c. Strand exchange assay scheme. d-f. Titration of PRC2 with either the sense (d) or anti-sense (e) vg PRE RNA, or a non-complementary RNA (f). g. Quantification of 3 titrations; graphs show mean +/- S.D of the fraction of signal with 200 nM PRC2. h, i. RNA strand invasion products are sensitive to RNaseH (lane 4) but resistant to RNaseA (lane 6). See also Supplementary Figure 6. Source data are provided in Source Data file.

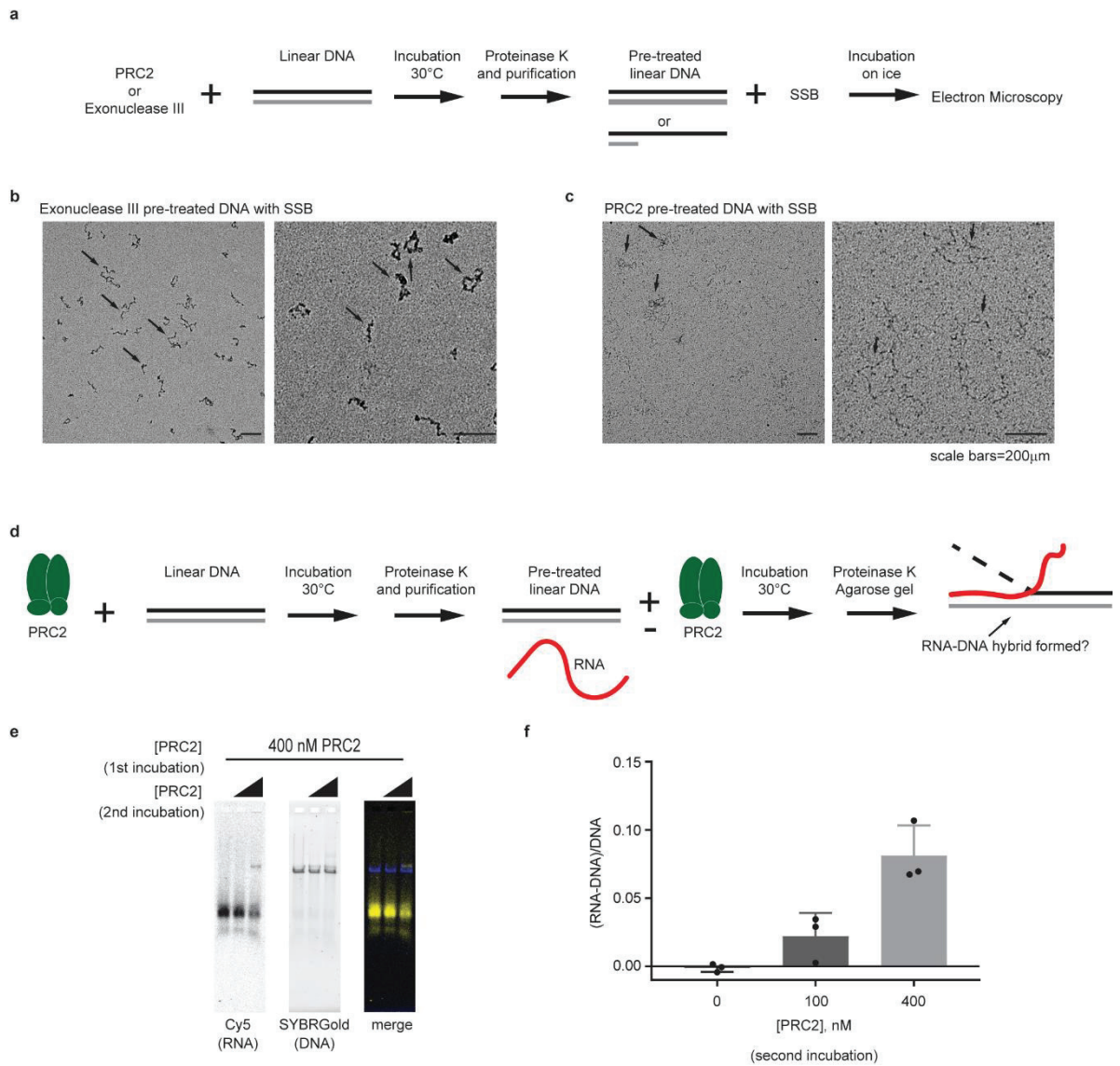


Figure 2-4 Nuclease contamination cannot explain RNA-DNA hybrid formation by PRC2

a. Scheme of experiment to test whether PRC2 generate a long ssDNA filament. Linear DNA was incubated with PRC2 or exonuclease III and purified. SSB protein was added to pre-treated DNA and samples were visualized by electron microscopy. *b, c.* Representative EM pictures of DNA pre-treated with Exonuclease III (*b*) or with PRC2 (*c*). Arrows indicate DNA with (*b*) or without (*c*) SSB coating. *d.* Scheme of experiment to test whether pre-incubation of DNA with PRC2 allows RNA-DNA hybrid formation in the absence of PRC2.is sufficient. *e, f.* Representative gel (*e*) and quantification (*f*) of strand exchange reactions using DNA that was pre-treated with PRC2 (400 nM) as the substrate. $n=3$. Graph shows the mean \pm S.E.M. Source data are provided in Source Data file.

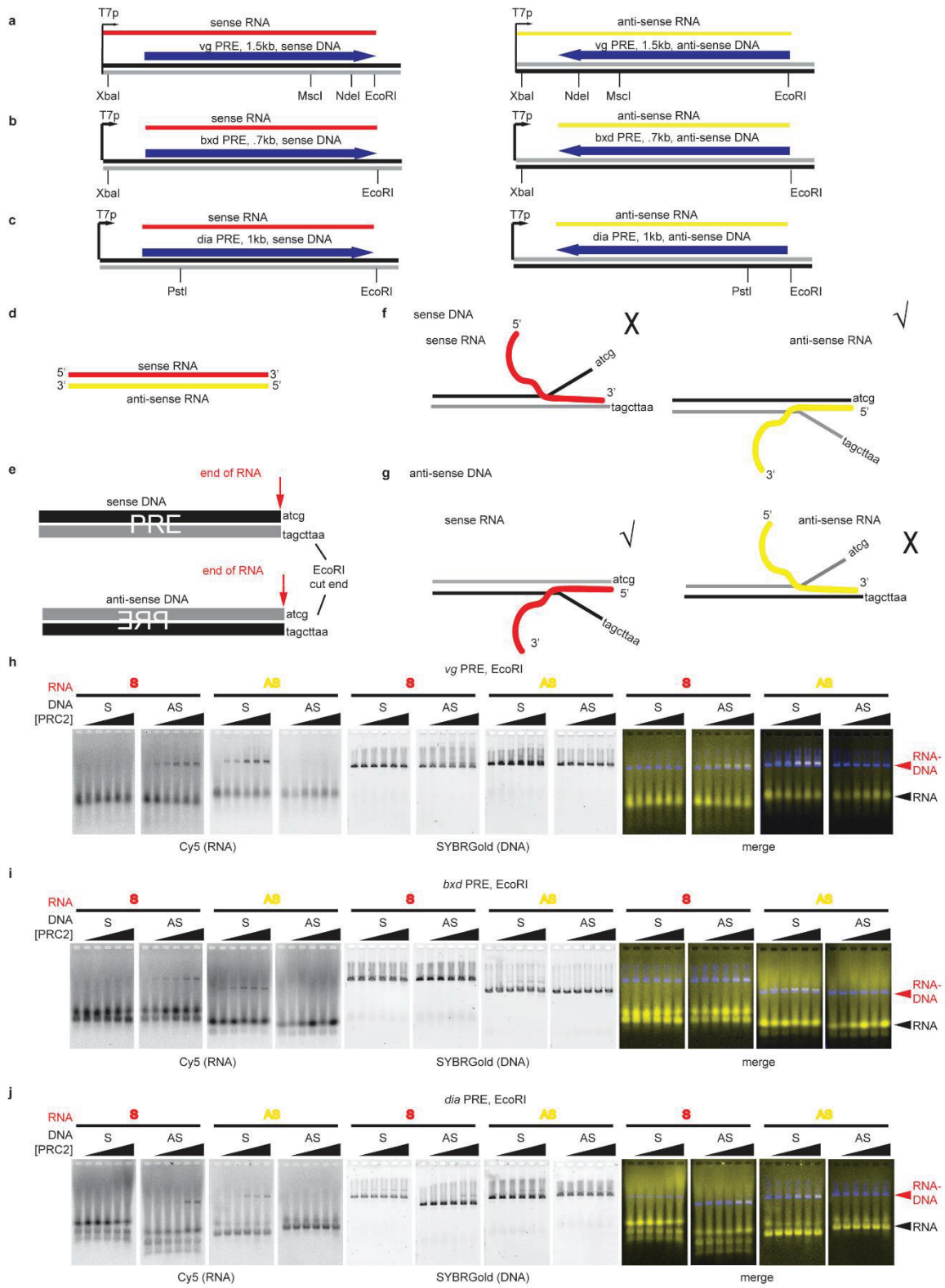


Figure 2-5 Pattern of RNA used by PRC2 on EcoRI-digested DNA templates

a-c. Schematic of DNA and RNA templates for the vg (a), bxd (b) and dia (c) PREs. DNA template were linearized at one of the indicated sites. d-e. Schematic of RNA and DNA templates. RNA transcripts were produced from linear templates that end 4 bases before the EcoRI site so that the sequence indicated is not present in the RNAs. f-g. Schematics of the 4 possible hybrids formed using the 2 RNAs and the sense (f) or anti-sense (g) DNA, indicating which two are observed. h-j show representative gels for each data set. Experiments were repeated 3 times, and all 4 sets for each PRE were always run in the same experiment. [PRC2] is 25-400 nM.

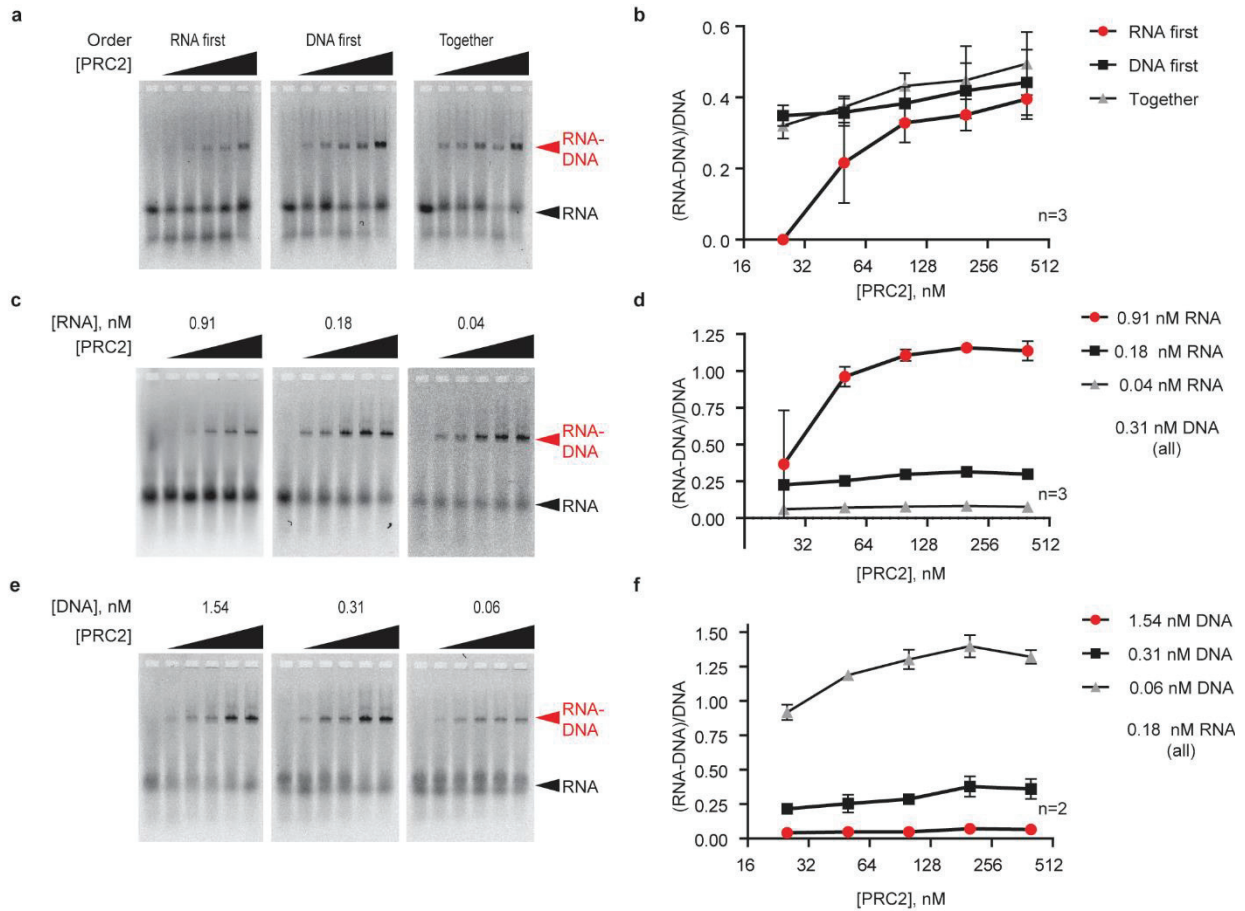


Figure 2-6 Characteristics of PRC2 strand exchange activity

In a-f, left panel shows a representative gel of Cy5-labelled RNA, and graph (right panel) summarizes multiple experiments. a, b. Effect of order of addition of nucleic acids. c-f. Effect of increasing RNA (c,d) or DNA (e,f). Asterisks: red=0.91 nM vs. 0.04 nM; black=0.91 nM vs. 0.18 nM. |- All PRC2 titrations are 25-400 nM. All graphs show the mean +/- S.E.M. Source data are provided in Source Data file.

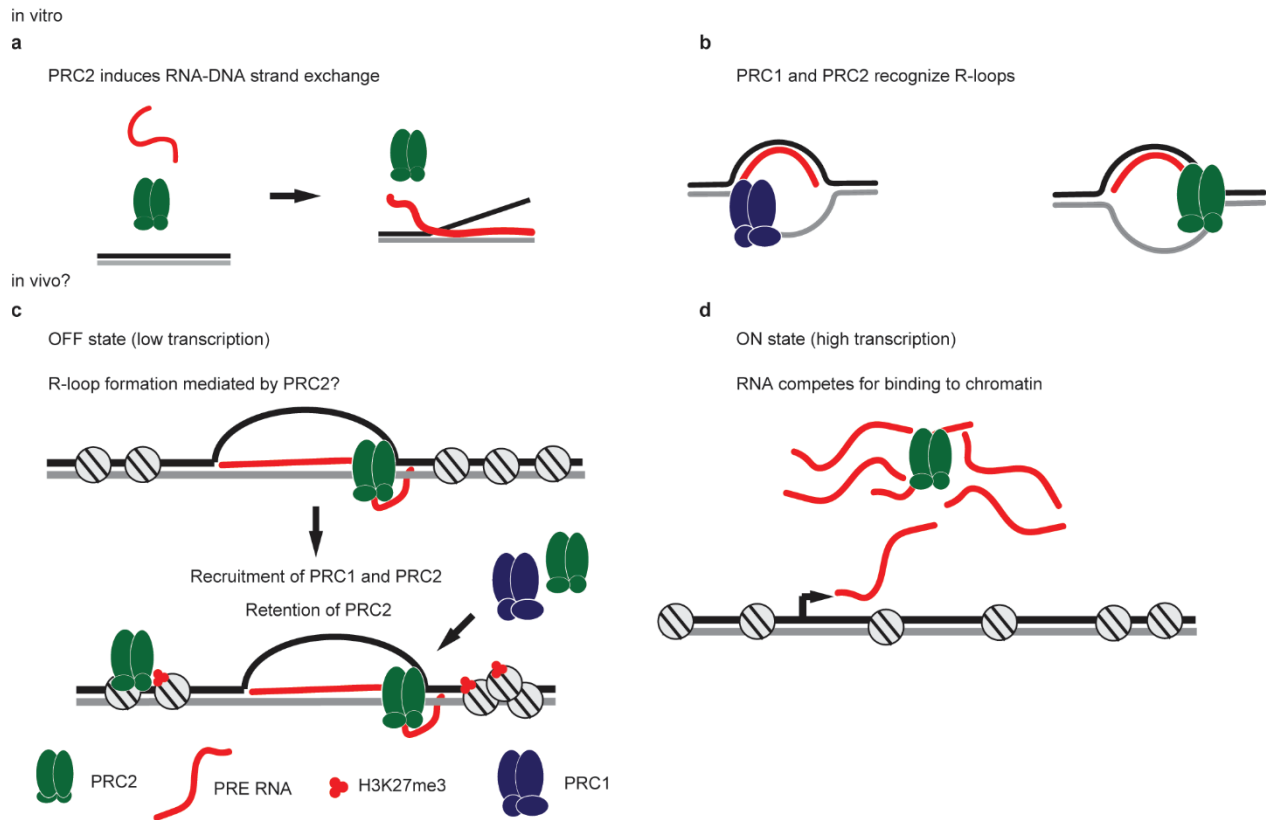
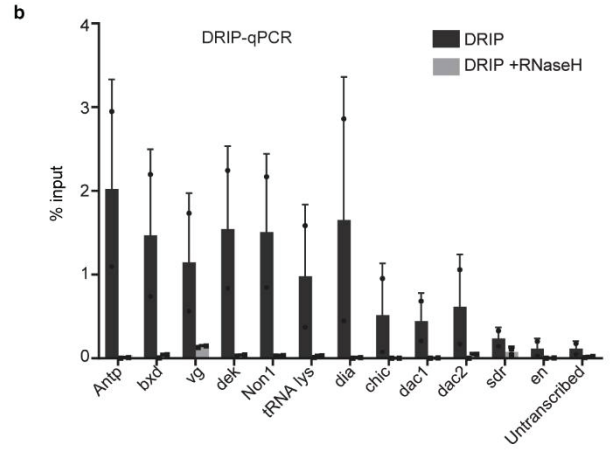
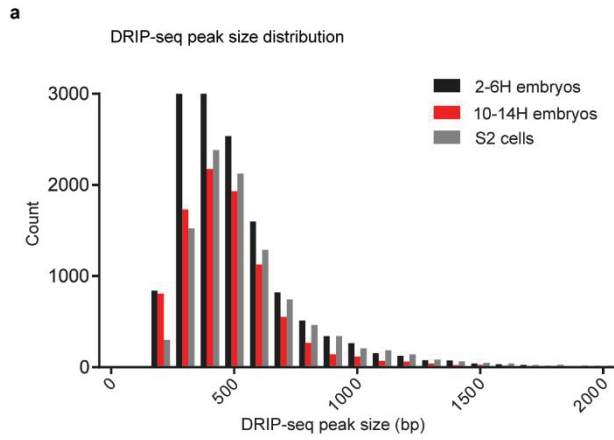


Figure 2-7 Model for the role of R-loop formation driven by PRC2 in PcG gene silencing

Top panels (a,b) summarize our in vitro results with PRC2 (a, b) and PRC1 (b). Bottom panels (c,d) are a speculative model for how these activities could contribute to establishing or maintaining the OFF state at PREs. The ability of RNA to compete for PRC2 binding to chromatin is based on extensive observations in the literature. a. PRC2 mediates RNA-DNA strand exchange, creating RNA-DNA hybrids. b. PRC1 and PRC2 recognize R-loops. c. Low levels of transcription through PREs could allow R-loop formation, which could depend on PRC2 activity (but could also be independent). R-loop formation sequesters the RNA and inhibits new RNA production, allowing PRC2 to be retained on chromatin. Recruitment of PRC1 and PRC2 (and other factors) may occur through recognition of the R-loop. d. RNA produced by high levels of transcription competes for PRC2 binding to chromatin. See Discussion for details. This model is inspired by that of Ringrose (Ringrose, 2017), but incorporates our new in vitro observations.

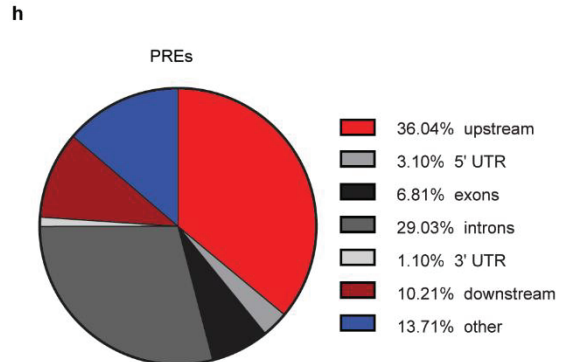
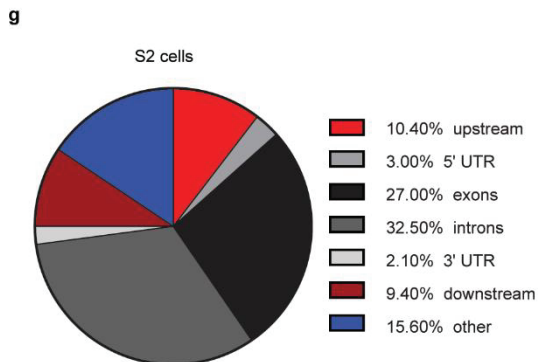
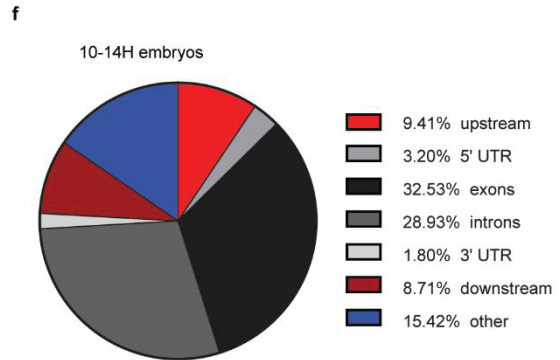
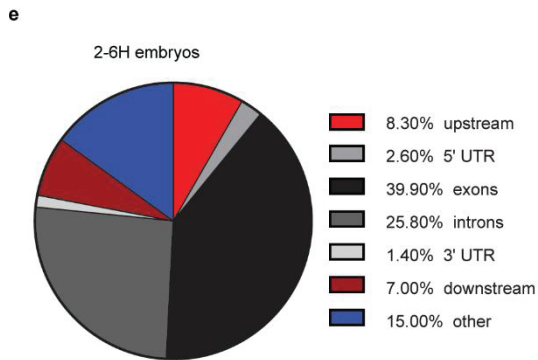


c

| | R-loop peaks | Overlap with gene body | Overlap with EST |
|----------------|--------------|------------------------|------------------|
| 2-6H embryos | 11233 | 8479 (75%) | 9090 (81%) |
| 10-14H embryos | 7471 | 5378 (72%) | 5675 (76%) |
| S2 cells | 9605 | 7258 (76%) | 7380 (77%) |

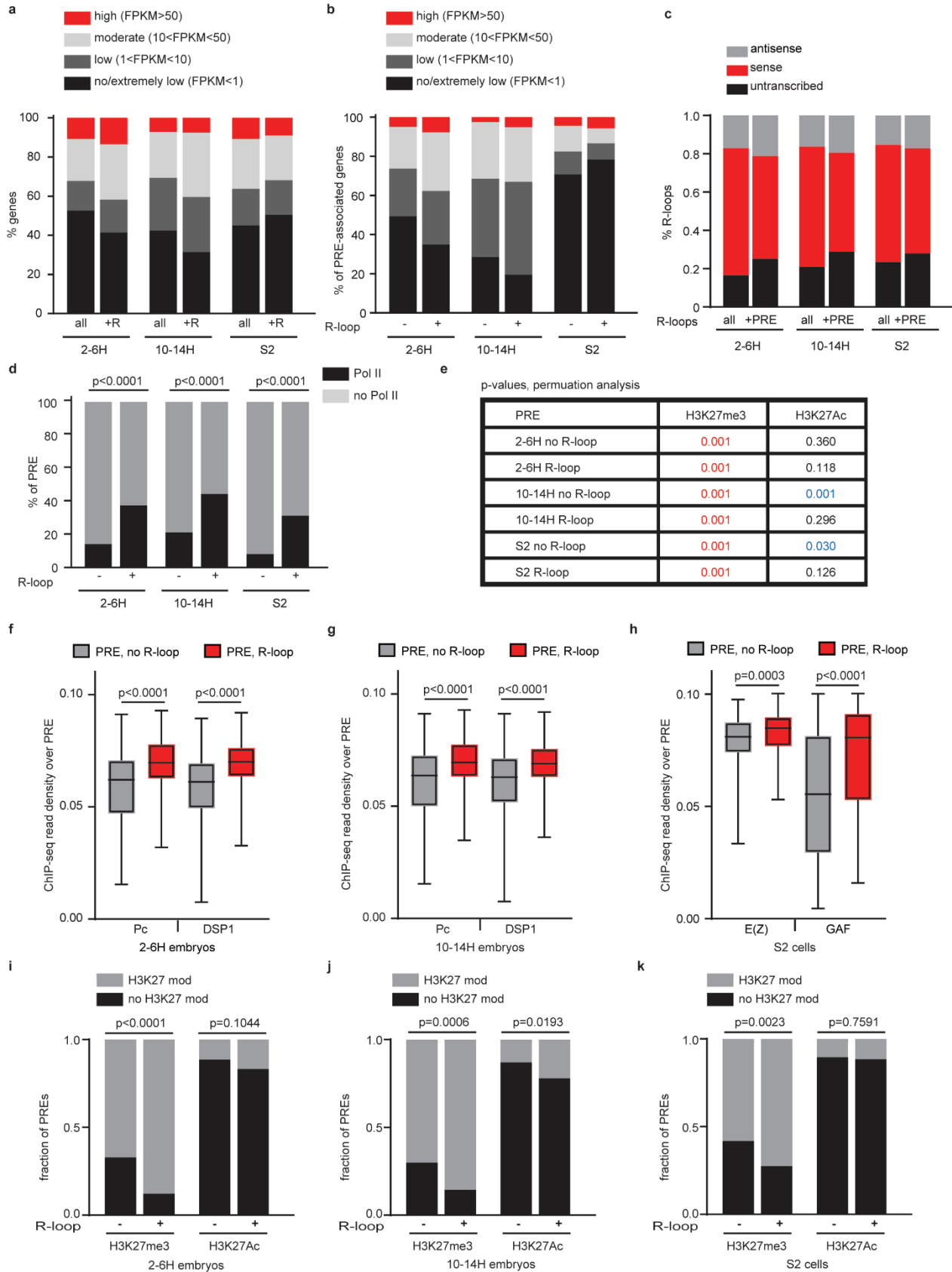
d

| | PREs overlapping R-loops | p-value (permutation analysis) |
|----------------|--------------------------|--------------------------------|
| 2-6H embryos | 180/541 (33%) | 0.001 |
| 10-14H embryos | 118/541 (22%) | 0.001 |
| S2 cells | 156/541 (29%) | 0.001 |



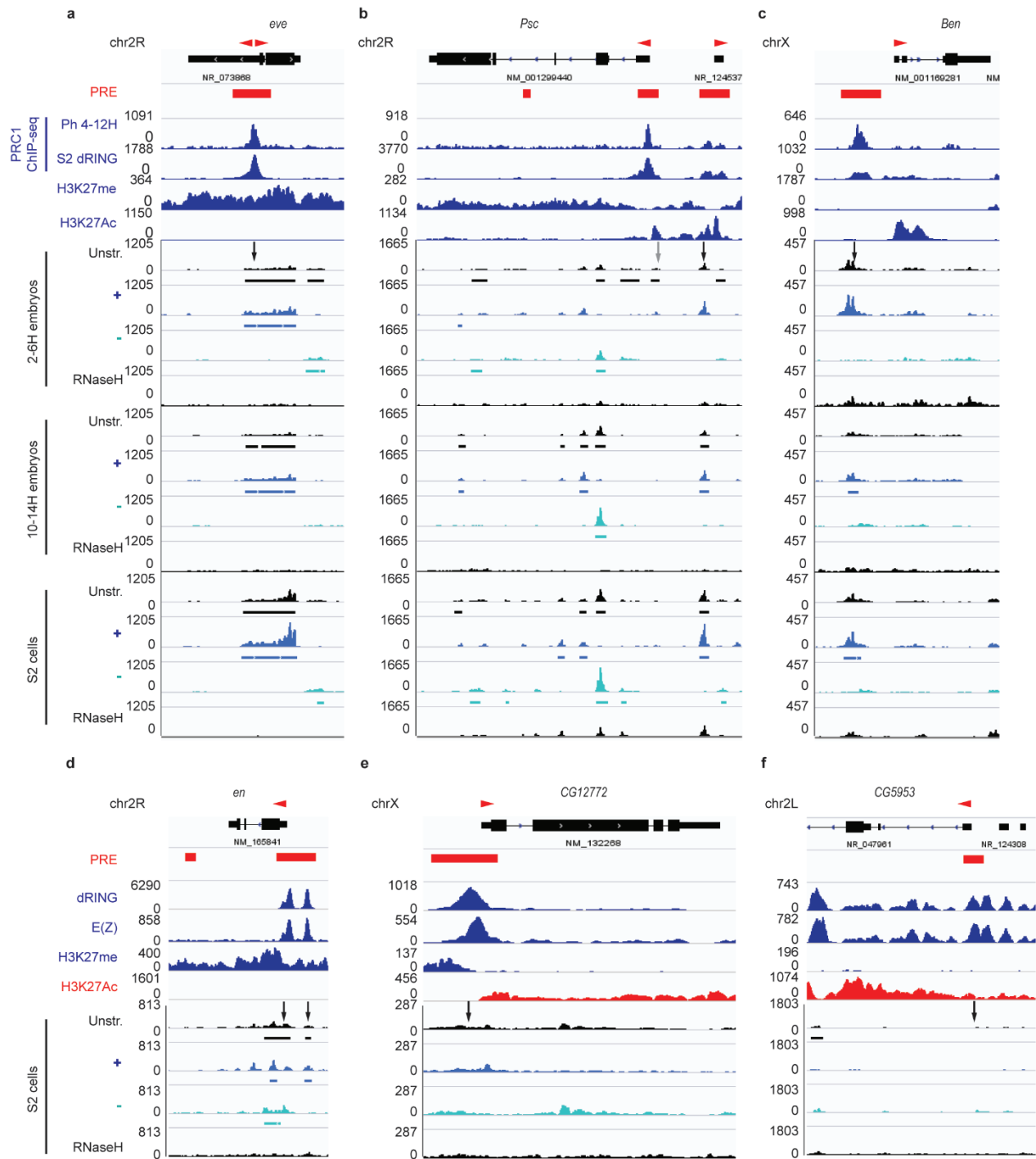
Supplementary Figure 2-1 Overview of DRIP-seq analysis

a. Distribution of DRIP-seq peak sizes. b. 10 positive sites and 3 negative sites were confirmed by DRP-qPCR using S2 cell genomic DNA with or without RNaseH treatment. Graph show the mean +/- S.D. c. Summary of overlaps of R-loops with genes. Note that heterochromatin, ChrU, and mitochondrial genome sequences were removed for this analysis. d. Summary of overlaps of R-loops with PREs. e-h. Pie charts of distribution of R-loops (e-g) and PREs (h) across the genome. Source data are provided in Source Data file.



Supplementary Figure 2-2 Relationship between R-loop formation, transcription, and PcG proteins

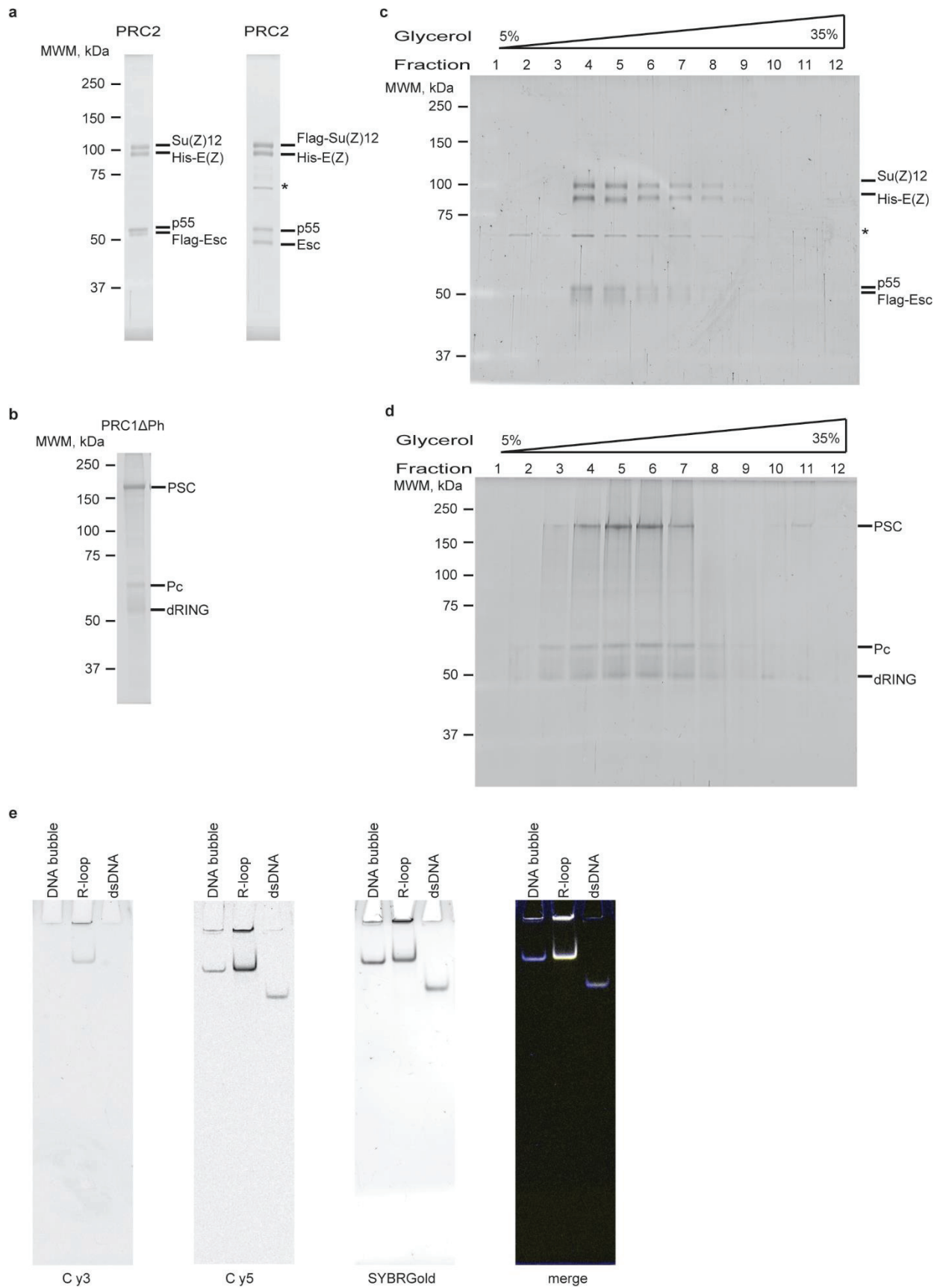
a. Expression levels of all genes as compared with genes with R-loops. b. Expression level of genes associated with PREs that do not or do form R-loops. c. Strandedness of all R-loops or R-loops formed at PREs relative to annotated transcripts. R-loops at PREs are more likely to only overlap antisense transcripts in 2-6H ($p=0.002$) and 10-14H ($p=0.0117$) data sets by Fisher's exact test. Note that "antisense" means that transcription is only annotated in the antisense orientation, but many R-loops that overlap a "sense" transcript that could explain the R-loop also have annotated antisense transcripts. d. Fraction of PREs that do or do not form R-loops that overlap RNA polymerase II peaks. p -values are for Fisher's exact test comparing the overlap for PREs that do or do not form R-loops. e. Summary of overlaps between PREs that do or do not form R-loops and H3K27me3 or H3K27Ac ChIP-seq peaks. Red numbers indicate that overlap is higher than random and blue numbers that it is lower than random. Overlap between PREs with or without R-loops is not significant, and is in fact anticorrelated in some cases (blue numbers). f-h. Median density of normalized ChIP-seq reads over PREs that do or do not form R-loops. Whiskers show min. to max. i-k. Fraction of PREs that do or do not form R-loops that overlap H3K27me3 or H3K27Ac peaks. p -values are for Fishers exact test comparing the overlap for PREs that do or do not have R-loops.



Supplementary Figure 2-3 Additional examples of R-loops

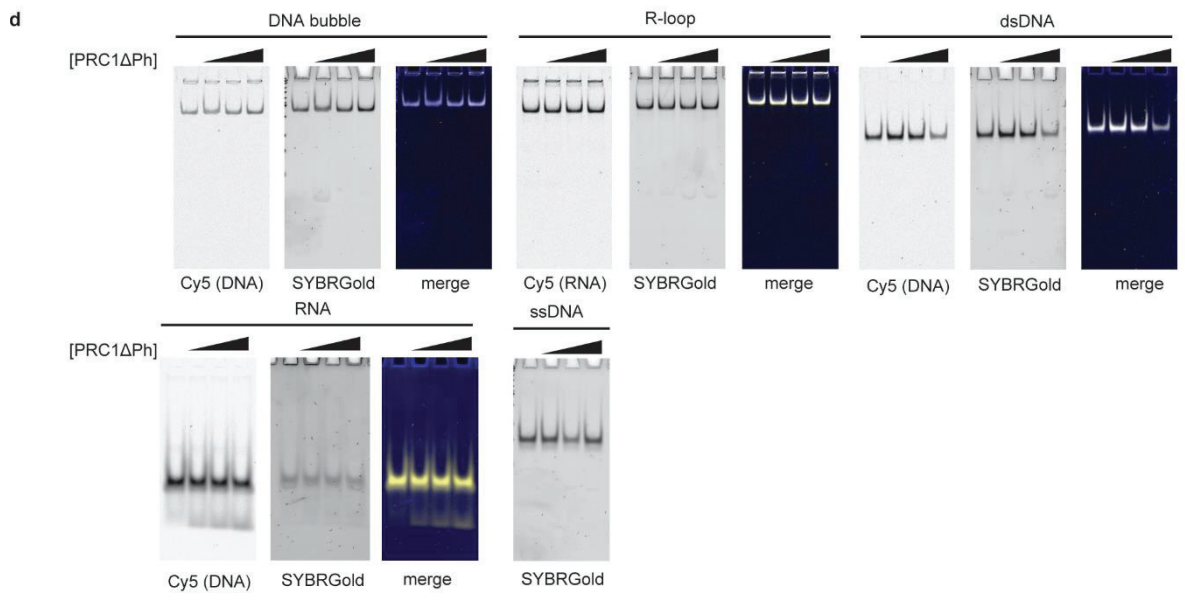
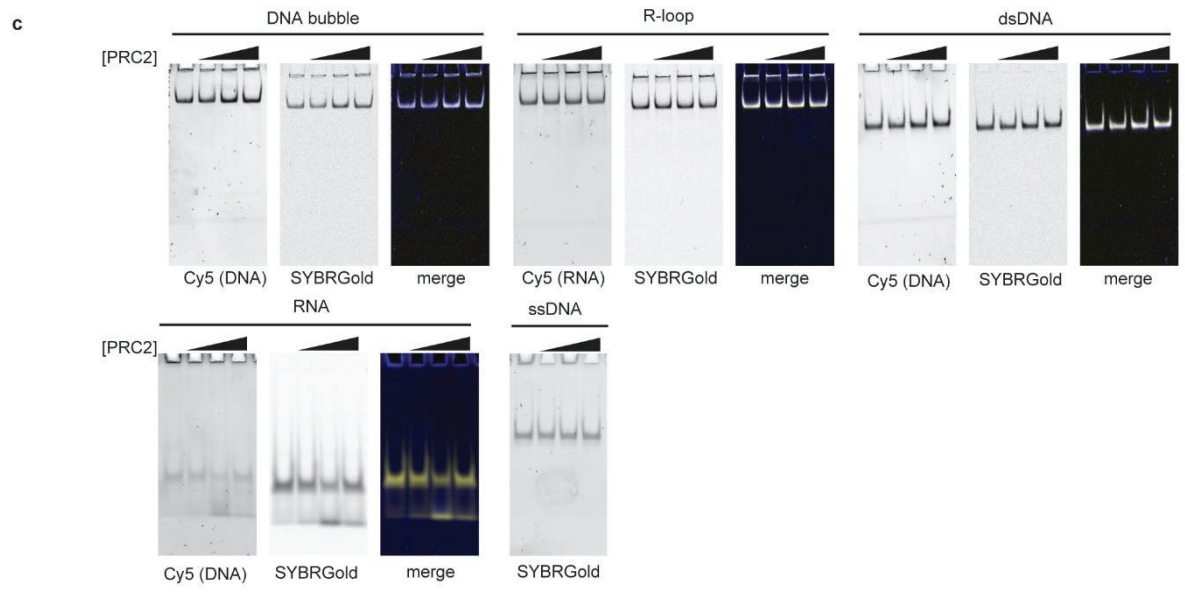
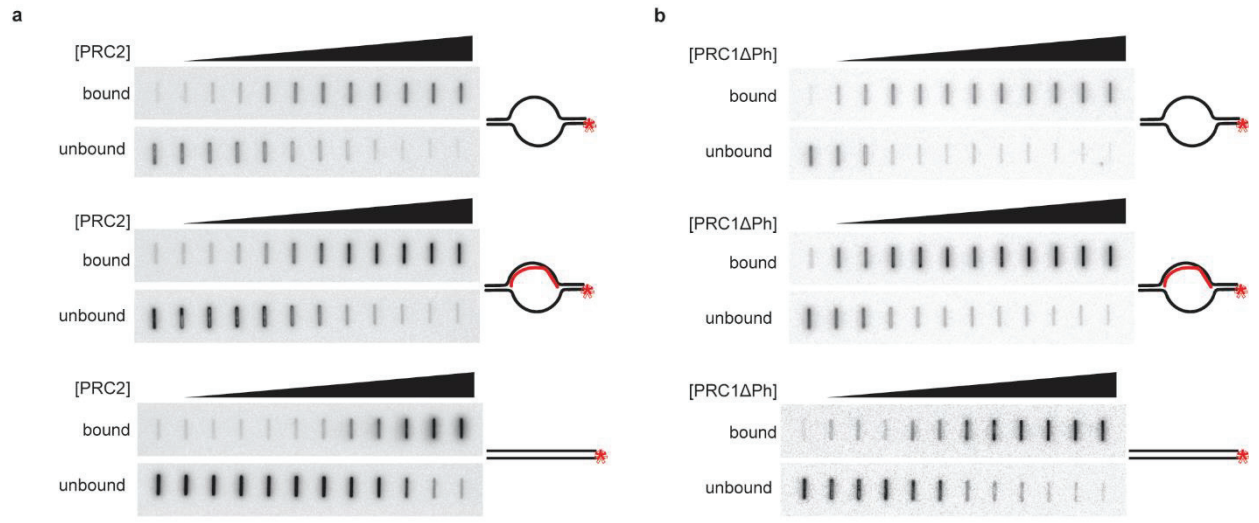
a. An R-loop forms over the PRC1-bound PRE with the same strandedness as *eve* transcription (arrow), but opposite to the direction of the overlapping ncRNA. b. An R-loop forms at a PRC1-bound PRE at the *Psc* gene. Note that an R-loop is detected at the PRE overlapping the promoter in 2-6H embryos (grey arrow), while R-loop signal is detected in all samples at the upstream PRE that overlaps the non-coding RNA (black arrow). c. Arrow indicates R-loop that overlaps a PRC1-bound PRE upstream of the *ben* gene but has no annotated transcript. d-f. Relationship between H3K27 modification, R-loops, and PREs in S2 cells. d. The

en PRE has characteristics of the OFF state: the PRE has high levels of H3K27me3, both PRC1 and PRC2 binding, and forms R-loops (arrows). e. The PRE at the promoter for CG12772 has characteristics of a balanced state, with both H3K27Ac and H3K27me3 modifications, PRC1 and PRC2 binding, and no R-loop (arrow). f. The PRE at CG5953 has characteristics of the ON state, with PRC1 binding, high levels of H3K27Ac, no H3K27me3, and no R-loop formation. RNA-seq data (FPKM) are consistent with these states: *en*=0.016, CG12772=6.3, CG5953=3.6, 15.6 (2 isoforms).



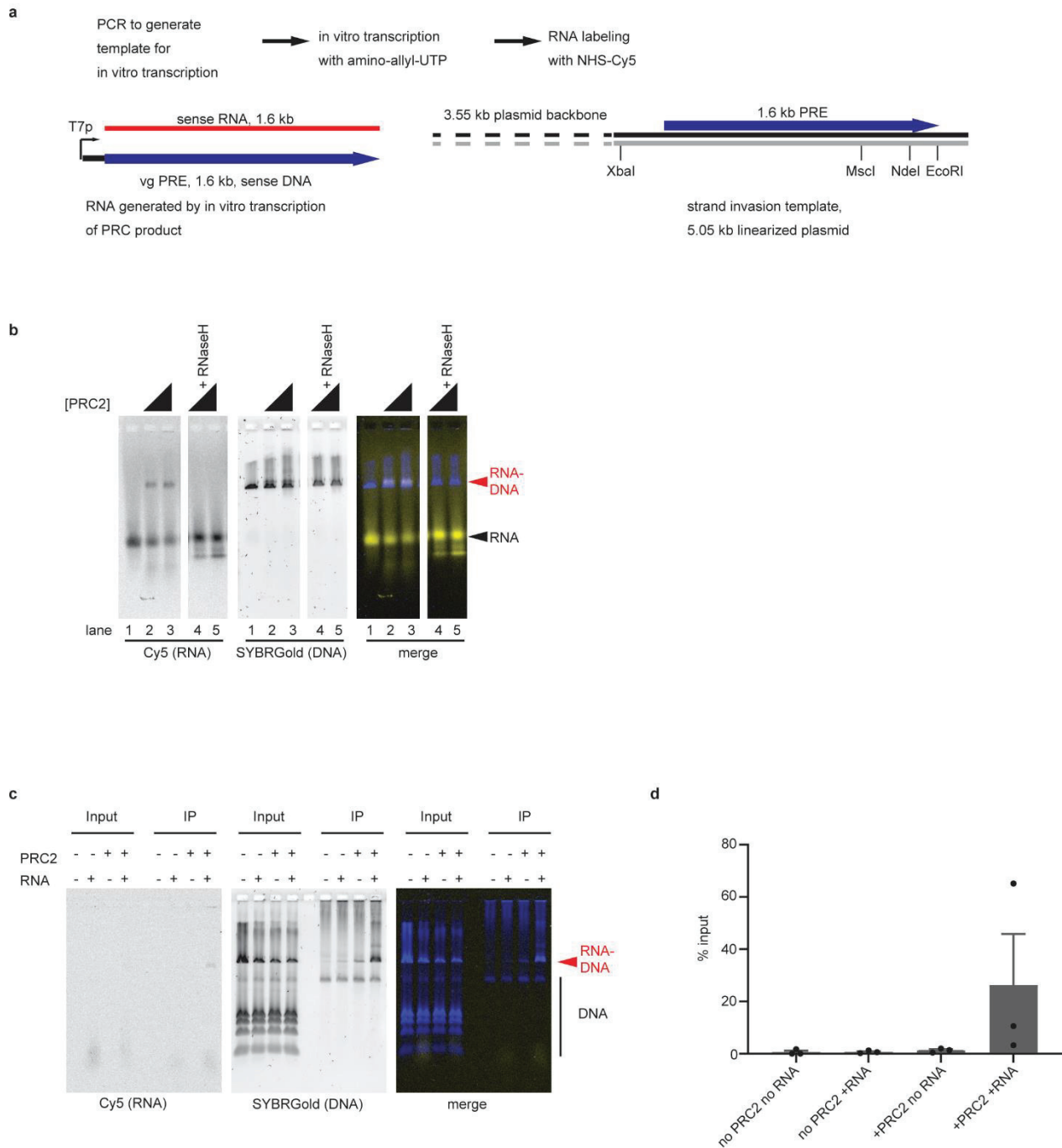
Supplementary Figure 2-4 Templates and proteins used to analyze PRC1 and PRC2 interactions with R-loops in vitro

*a, b. SYPRO Ruby-stained SDS-PAGE of PRC2 (a) (10% acrylamide), and PRC1ΔPh (b) (8% acrylamide) used in this study. PRC2 prepared with Flag-Su(Z)12 or Flag-Esc were used interchangeably. *=HSC70 contaminant. c, d. SYPRO Ruby-stained SDS-PAGE of glycerol gradient fractions of PRC2 (c) or PRC1ΔPh (d). The predicted molecular weight of PRC2 is 284 kDa, and of PRC1ΔPh 261 kDa. e. Oligonucleotide substrates used for EMSA and filter binding. Note that these substrates were prepared with fluorescently labelled oligonucleotides and the unlabelled 5' ends subsequently labelled with ³²P for binding assays.*



Supplementary Figure 2-5 PRC1 and PRC2 interactions with nucleic acids

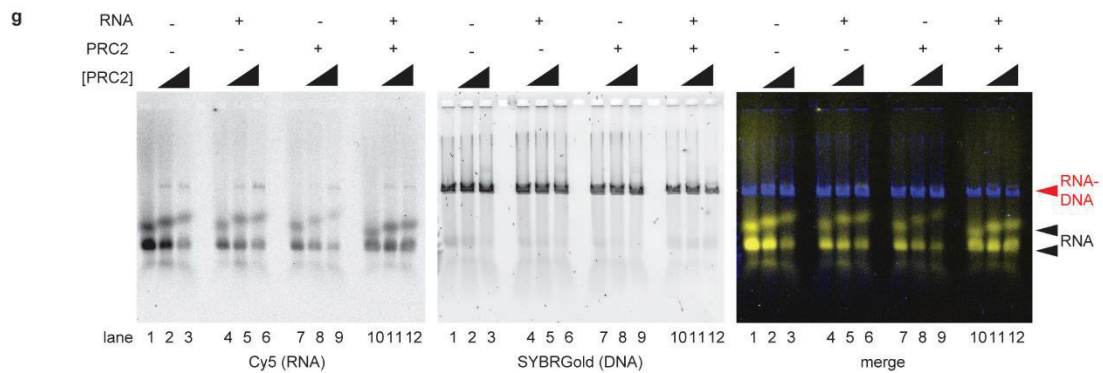
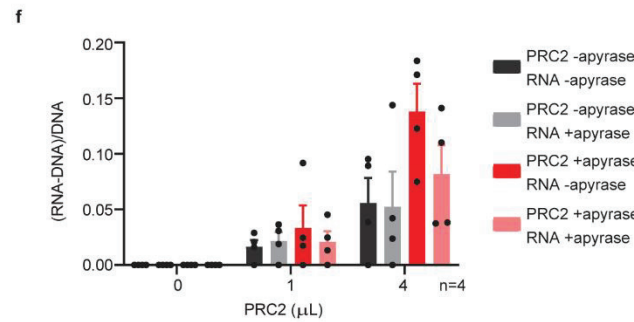
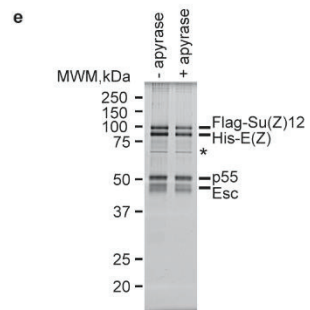
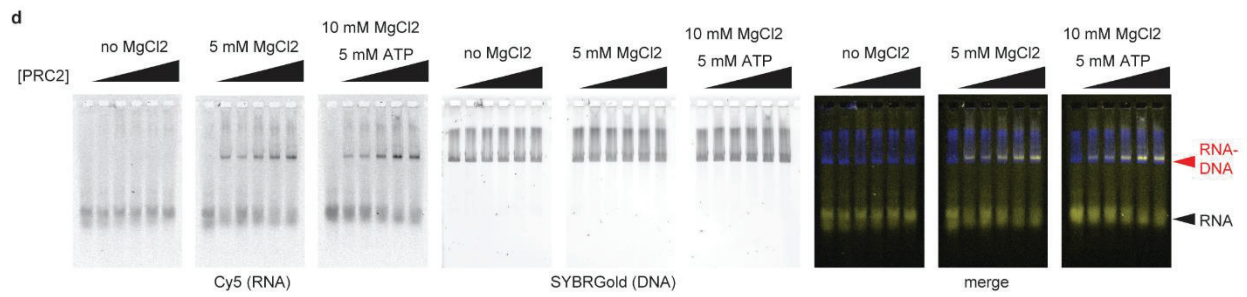
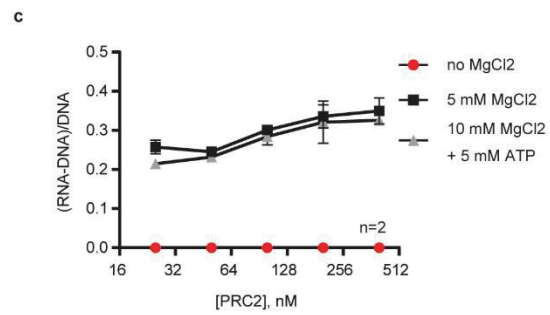
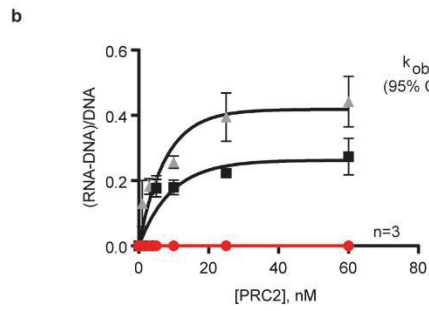
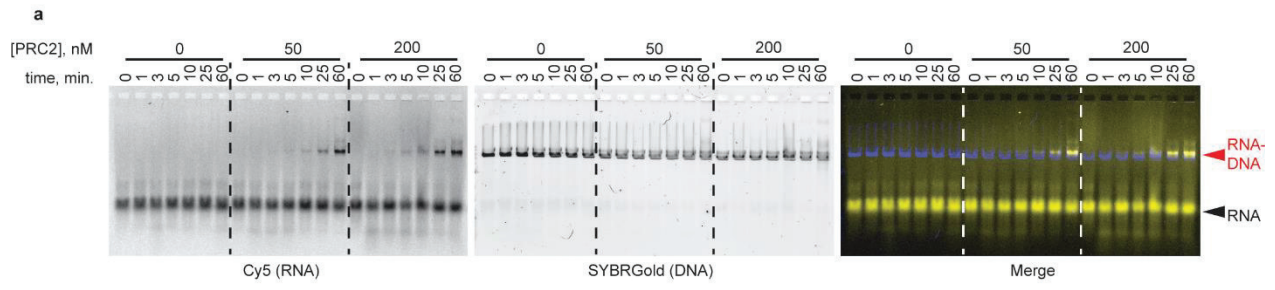
a, b. Representative filter binding assay for the data presented in Fig. 2a, b. c, d. Native acrylamide gel analysis of oligonucleotide substrates after incubation with PRC2 (c) or PRC1 Δ Ph (d) under conditions used for filter binding assays, demonstrating that the substrates remain intact. PRC2 or PRC1 Δ Ph were incubated with 20 nM probe followed by Proteinase K digestion.



Supplementary Figure 2-6 PRC2 induces the formation of RNA-DNA hybrids that are sensitive to RNaseH and recognized by S9.6

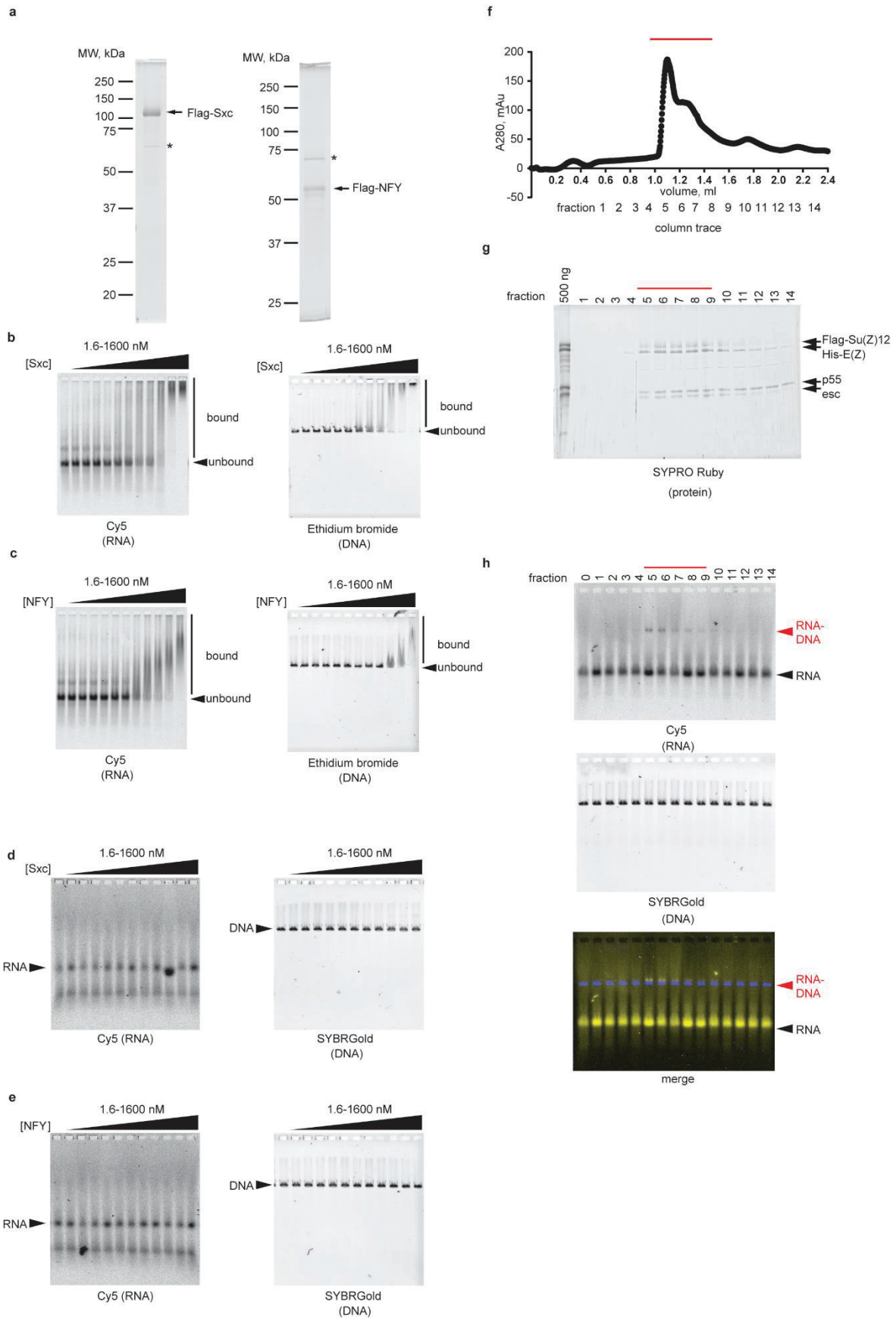
a. Scheme for generating fluorescently labelled RNAs. DNA corresponding to the PRE sequence is amplified with primers, one of which contains the T7 promoter sequence for in vitro transcription. RNAs are produced by in vitro transcription in the presence of amino-allyl UTP, purified, and labelled with NHS-Cy5. These RNAs correspond precisely to the PRE sequence in the DNA template, and are used for strand exchange with linearized, PRE-containing plasmid DNA. Dashed lines indicate plasmid backbone sequence, which is not drawn to scale. Note that the radio-labelled RNAs used in Fig. 3 were generated by in vitro transcription of

the circular plasmid template, so they all share ~200 bases of plasmid sequence between the T7 promoter and start of the PRE. b. Representative gels of PRC2 strand exchange activity with Cy5-labelled RNA. PRC2 strand invasion products (lanes 2, 3) are sensitive to RNaseH (lanes 4, 5). [PRC2]=50 and 200 nM, [DNA]=0.3 nM, and [RNA]=0.18 nM. c. Representative gel of DNA incubated with PRC2 and Cy5-labelled RNA after purification and immunoprecipitation with S9.6 antibody. Only the sample containing PRC2, RNA and DNA forms RNA-DNA hybrids that are immunoprecipitated with S9.6. [PRC2]=100 nM, [DNA]=0.3 nM and [RNA]=0.18 nM. d. Quantification of three experiments; error bars are S.E.M. Source data are provided in Source Data file.



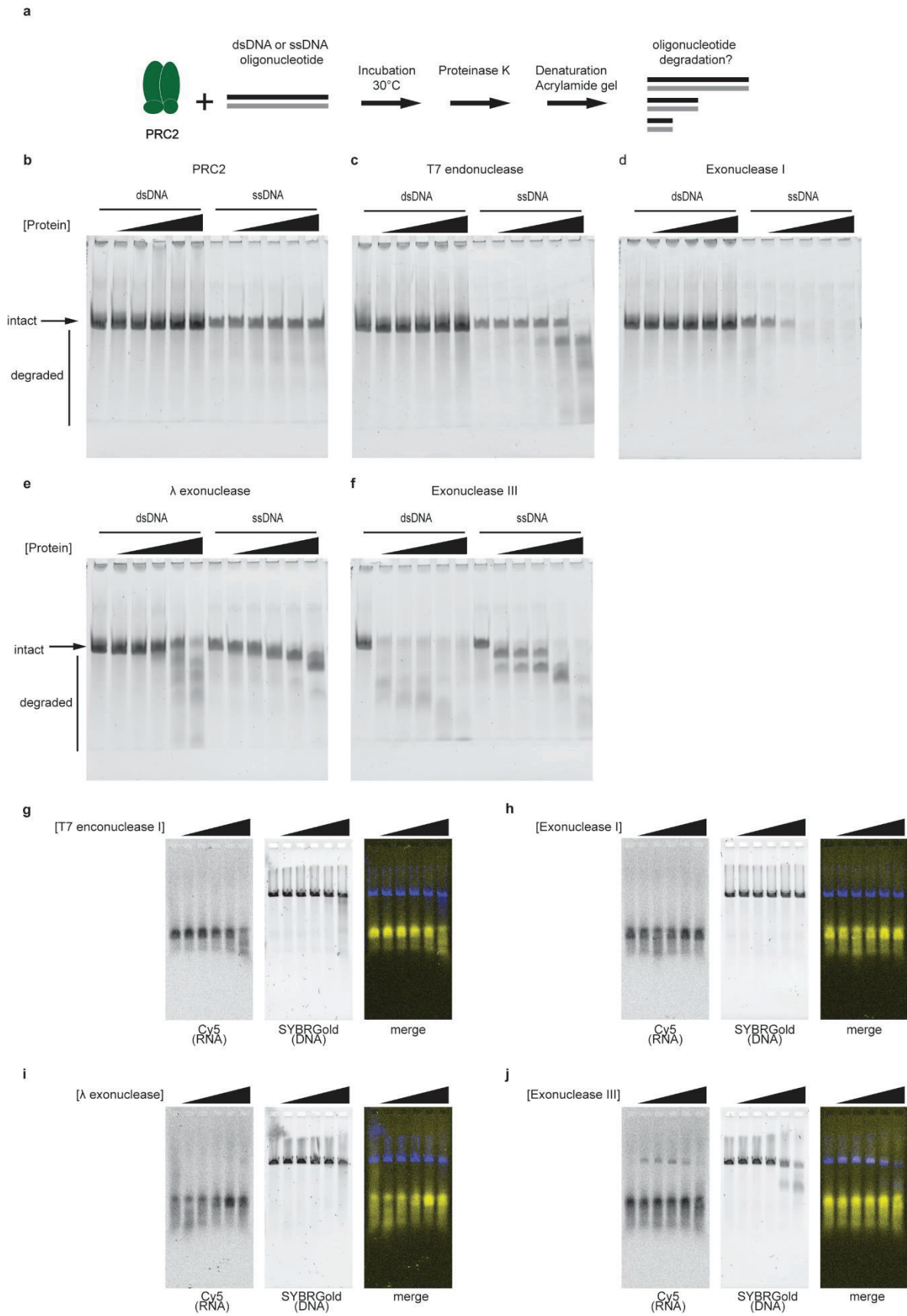
Supplementary Figure 2-7 Time course of PRC2-mediated strand invasion and lack of requirement for nucleotides

*a, b. Representative gels (a) and quantification (b) of time course experiment for strand exchange activity. c, d. Quantification (c) and representative gels (d) of RNA-DNA strand exchange activity demonstrating that strand exchange requires MgCl₂ but not ATP. PRC2 titration is 25-400 nM. e. SYPRO Ruby-stained SDS-PAGE gel of PRC2 treated with apyrase. *HSC-70 contaminant. f, g. Quantification (f) and representative gels (g) of strand exchange assays carried out with PRC2 and/or RNA that was pre-treated with apyrase to remove any contaminating ATP. Lanes 1-3 PRC2 and RNA not incubated with apyrase, lanes 4-6 RNA pre-treated with apyrase, lanes 7-9 PRC2 pre-treated with apyrase, lanes 10-12 PRC2 and RNA pre-treated with apyrase. Error bars in all graphs are S.E.M. Source data are provided in Source Data file.*



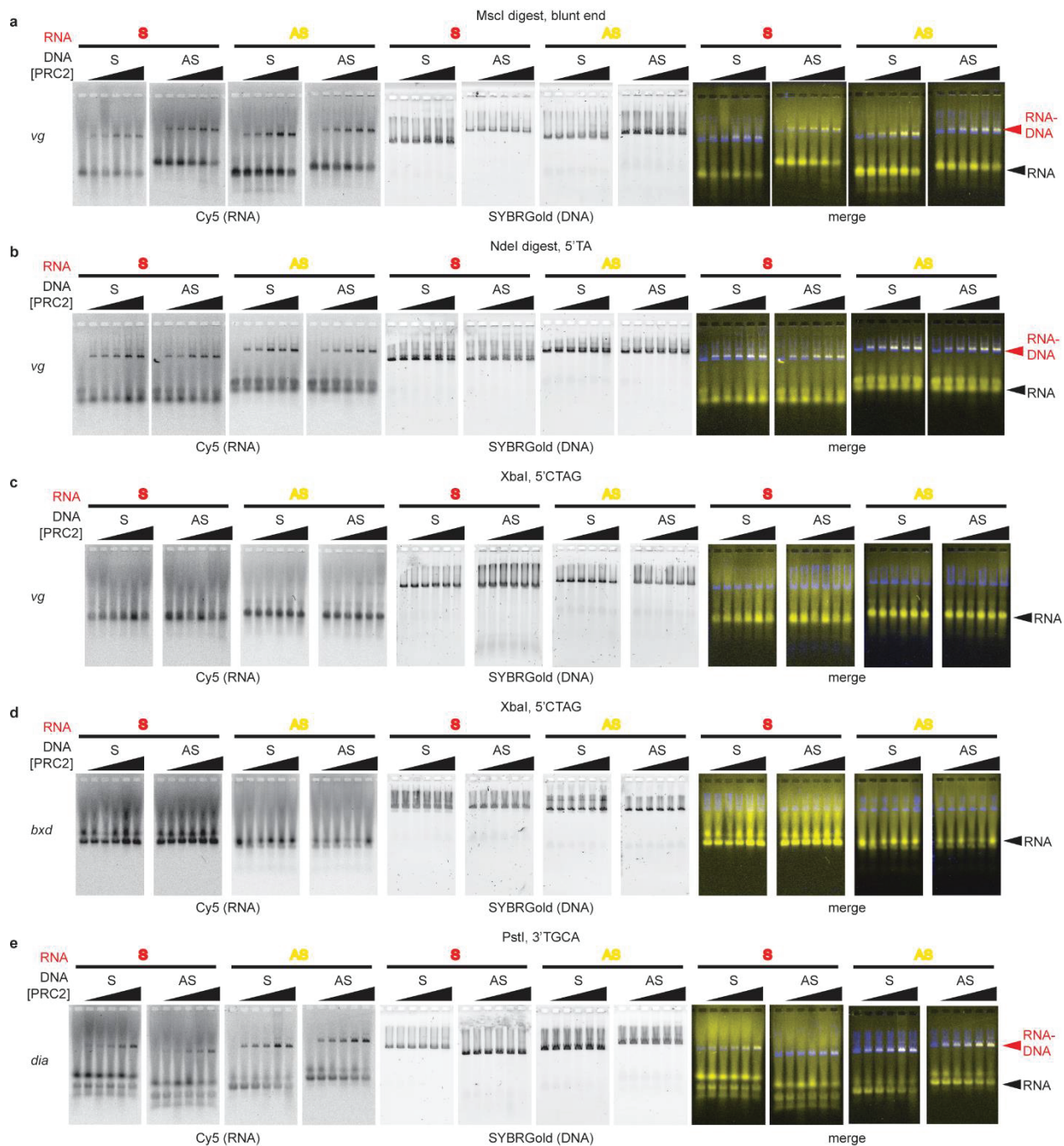
Supplementary Figure 2-8 Other RNA and DNA binding proteins do not have strand exchange activity and activity co-fractionates with PRC2

*a. SYPRO Ruby-stained SDS-PAGE gels of Flag-Sxc and Flag-NFY. *=HSC70 contaminant. b,c. EMSA demonstrating that both Sxc (b) and NFY (c) bind RNA (left) and linear DNA (right), the substrates for strand exchange. d, e. Neither Sxc (d) nor NFY (e) have strand exchange activity at concentrations where they can bind to both substrates. f. 45 μ g of PRC2-Flag-Su(Z)12 were fractionated on a Superdex 200 PC3.2/300 column. g. SDS-PAGE (10%) of fractions from size exclusion column. 7 μ l of each fraction were loaded on the gel, which was stained with SYPRO Ruby. h. RNA strand exchange assay of the size column fractions. 4 μ l of each fraction was used in this assay with 0.3 nM DNA and 0.18 nM RNA. Similar co-fractionation of activity with PRC2 was observed with two additional preparations of PRC2 using glycerol gradient sedimentation, and a second PRC2 preparation using size exclusion chromatography.*



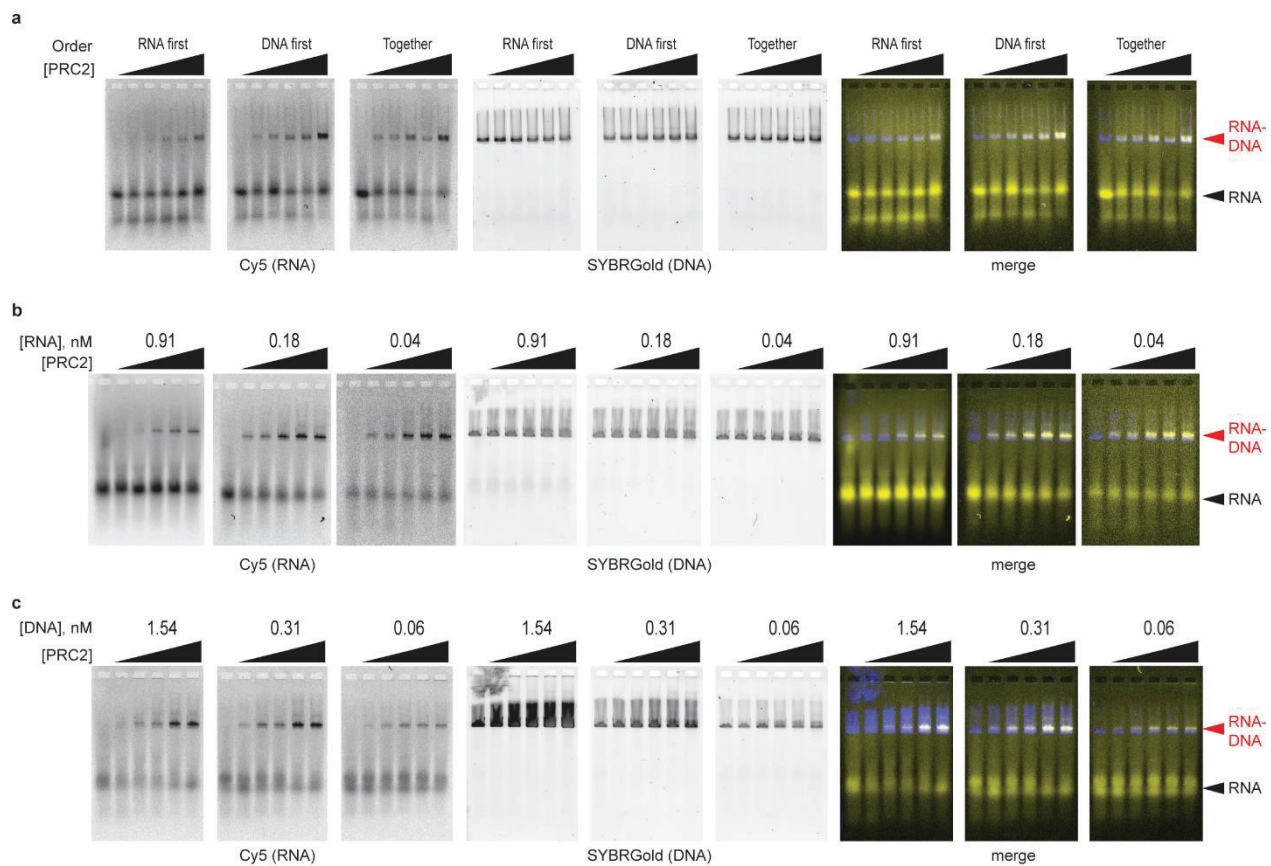
Supplementary Figure 2-9 PRC2 preparations do not contain detectable nuclease activity

a. Scheme of the experiment testing whether PRC2 preparations contain contaminant nuclease activity. b-f. Representative gels of phosphorylated dsDNA or ssDNA oligonucleotides after incubation with PRC2 (b, n=3), T7 endonuclease I (c, n=3), exonuclease I (d, n=3), λ exonuclease (e, n=3) or exonuclease III (f, n=3) incubated using strand exchange reaction conditions. g-j. Effects of different nucleases on oligonucleotide substrates under the same conditions used for PRC2-induced RNA-DNA strand exchange. Representative gels of DNA and Cy5-labelled RNA incubated with T7 endonuclease I (g), exonuclease I (h), λ exonuclease (i) or exonuclease III (j) (n=2). In all panels the titrations are [PRC2]: 25-400 nM, [T7 endonuclease I]: 0.00004-4 units, [exonuclease I]: 0.00008-8 units, [λ exonuclease]: 0.00002-2 units, [exonuclease III]: 0.0004-40 units.



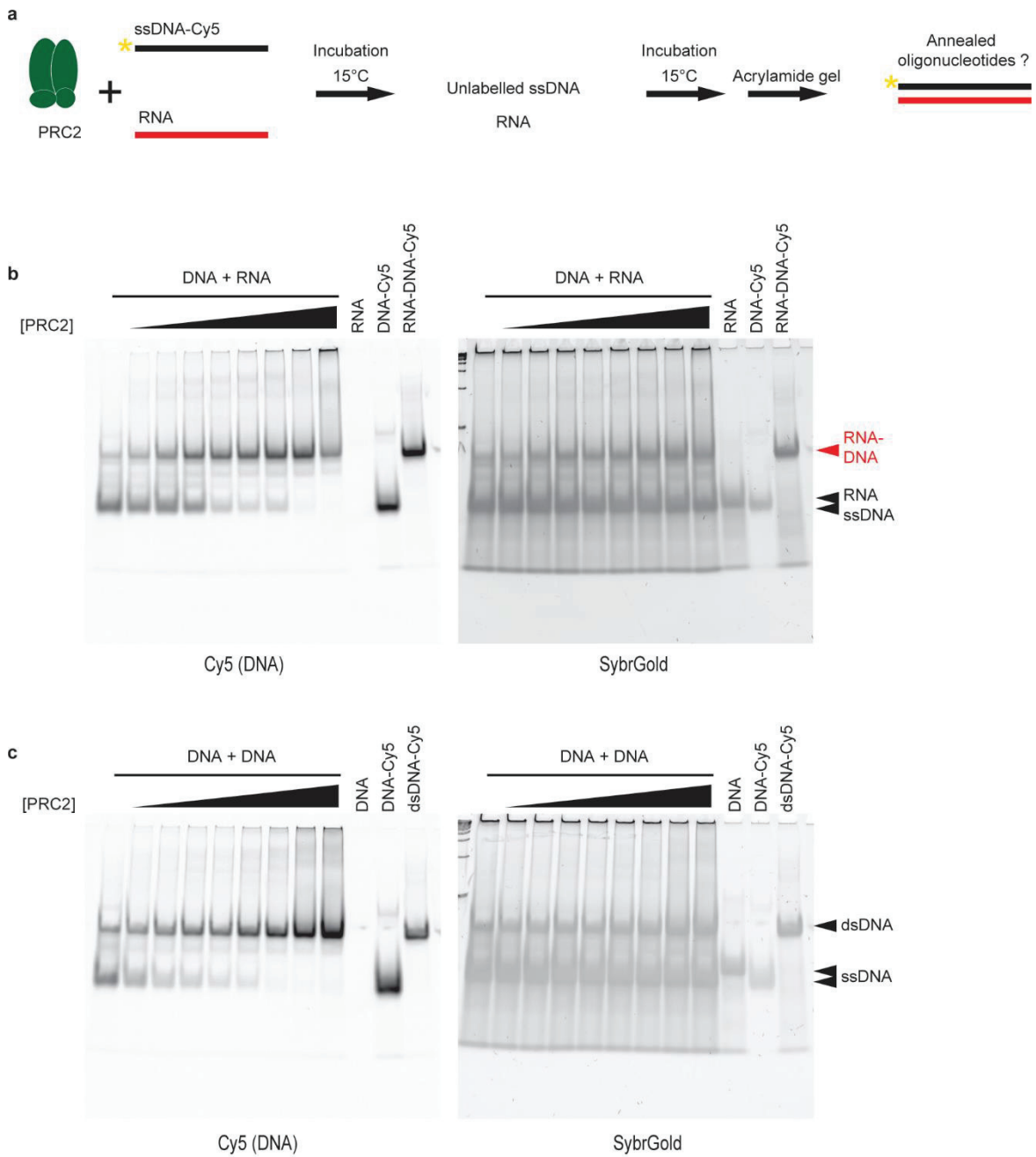
Supplementary Figure 2-10 PRC2 RNA strand exchange activity on different PRE templates with different DNA ends

a-e. Representative gels of RNA strand exchange with different DNA templates and RNAs. PRC2 can induce the formation of RNA:DNA hybrids when provide blunt (a), 5' overhang (b) or 3' overhang (e) DNA end. RNA strand exchange on the *vg* (c), or *bxd* (d) PRE template linearized with *XbaI*, which is 231 base pairs upstream of where the RNA initiates. Note that both *bxd* RNAs are active in strand exchange with other DNA templates (Fig. 5). In all panels, [PRC2] is 25-400 nM. Experiments were repeated 3 times.



Supplementary Figure 2-11 Effect of order of addition and substrate concentration on PRC2-mediated strand exchange

a. Gels from Fig. 6a (order of addition of DNA and RNA). b. Gels from Fig. 6b (RNA titration). c. Gels from Fig. 6c (DNA titration).



Supplementary Figure 2-12 PRC2 can anneal ssDNA and RNA to form RNA:DNA hybrids

a. Scheme for testing annealing activity of PRC2. b. PRC2 can anneal ssDNA and RNA leading to the formation of RNA:DNA hybrid (n=2). c. PRC2 can anneal two ssDNA to form dsDNA (n=2).

2.4 Supplementary tables

| Sequences of primers used for DRIP-qPCR | |
|---|-------------------------|
| Primer name | Sequence |
| Antp F | CAGCATCGCTCTCACTCTCT |
| Antp R | CAGAAGCAGCGGCTGCAAAA |
| bx1 F | CCATAAGAAATGCCACTTTGC |
| bx1 R | CTCTCACTCTCTCACTGTGAT |
| vg F | GGCCTCACTTAAGCGCGAAA |
| vg R | GAGGCGGATCAGCGTTCGAT |
| dek F | GCGATGAGCCAGAAGATGAG |
| dek R | CTTGGACTCATCAGTGGCAT |
| Non1 F | GCGTCCTGTACTTCATGGAT |
| Non1 R | CTTGTTGATGGCCAGGATGA |
| tRNA lys F | GCCAAGCTCATTTTCTACGATCT |
| tRNA lys R | GTCCGACAACGCCGATGATA |
| dia F | CGGCTCATCTGGAGGAGAAA |
| dia R | GCGGTGGCATGGGTATGTTT |
| chic F | AGAGATGGAGAGAGAGTGGC |
| chic R | GAACGGGTGACTCACGGCTT |
| dac1 F | GAGAGAGCGGGAGGATTTGA |
| dac1 R | TTTCTGTGTGCAACCCCGAA |
| dac2 F | GATACCGAGGGAGCGAAAAGA |
| dac2 R | TGCAAATTTCACTACTGCGC |
| sdr F | ACAGCTGATGTCGCTCACAT |
| sdr R | CGCTGAATGATCACCAGGTGA |
| en F | GGGTGGTGCACCTTGCAGTT |
| en R | CGCTGCGATGGGCCAAATAA |
| Untranscribed F | TCAAGCCGAACCCTCTAAAAT |
| Untranscribed R | AACGCCAACAAACAGAAAATG |
| | |

| Sequences of oligonucleotides used to test for nuclease activity and in annealing assay | |
|--|--|
| Primer name | Sequence |
| bxd PHO core U, DNA or RNA | GAGGCCATCTCAGTCGCACTTAAAACGGCCATTACGAACGACAGTTATGGCGACGGAGCT GCAGA |
| bxd PHO core L | TCTGCAGCTCCGTCGCCATAACTGTCGTTTCGTAATGGCCGTTTTAAGTGCGACTGAGATGG CCTC |
| | |
| Sequences of oligonucleotides assembled into R-loop and used in filter binding assay and EMSA | |
| Primer name | Sequence |
| vgcorePRE RNA F Cy5 | TGTGAGTGTGTGCGAGCTGCCAAGTGTGTGTGAGTGTGTT |
| vgcorePRE DNA R Cy5 or Cy3 | GAGCACTTACATGAGCTTGCCTTTGTCACA AACACACTCACACACTTGGCAGCTCGCACACTCACA GAAATGGCCGTTAAAATTCTTGACAACCCG |
| vgcorePRE DNA F | CGGGTTGTCAAGAATTTAACGGCCATTTTC TGTGAGTGTGTGCGAGCTGCCAAGTGTGTGTGAGTGTGTT TGTGACAAACGCAAGCTCATGTAAGTGCTC |
| vgcorePRE DNA random F | CGGGTTGTCAAGAATTTAACGGCCATTTTC ATAGCTCACCATCATTTGTGACAATTAAGTGTCTACTAAT TGTGACAAACGCAAGCTCATGTAAGTGCTC |

Supplementary Table 2-1 Sequences of primers and oligonucleotides used

| Chromosome | start | end |
|------------|---------|---------|
| chr2L | 5410 | 6069 |
| chr2L | 368280 | 369380 |
| chr2L | 377320 | 378520 |
| chr2L | 585040 | 585960 |
| chr2L | 594180 | 595780 |
| chr2L | 597440 | 598540 |
| chr2L | 817270 | 817869 |
| chr2L | 1954140 | 1955540 |
| chr2L | 1966960 | 1967960 |
| chr2L | 1969980 | 1971360 |
| chr2L | 1972160 | 1973500 |
| chr2L | 2093620 | 2094269 |
| chr2L | 2449340 | 2450460 |
| chr2L | 2454260 | 2455360 |
| chr2L | 2754860 | 2755780 |
| chr2L | 2828680 | 2829309 |
| chr2L | 2900140 | 2900779 |
| chr2L | 3538120 | 3540780 |
| chr2L | 3545000 | 3545900 |
| chr2L | 3571080 | 3572160 |
| chr2L | 3580520 | 3581560 |
| chr2L | 3606060 | 3607460 |
| chr2L | 3657160 | 3658580 |
| chr2L | 3661720 | 3663140 |
| chr2L | 3824900 | 3826140 |
| chr2L | 3834800 | 3836020 |
| chr2L | 3836080 | 3837280 |
| chr2L | 4809740 | 4812340 |
| chr2L | 5403080 | 5404460 |
| chr2L | 5421240 | 5422360 |

| | | |
|-------|----------|----------|
| chr2L | 5460640 | 5462100 |
| chr2L | 5923280 | 5924400 |
| chr2L | 5924760 | 5925940 |
| chr2L | 5979930 | 5980659 |
| chr2L | 6019440 | 6020540 |
| chr2L | 6535680 | 6536860 |
| chr2L | 6545070 | 6545849 |
| chr2L | 6821160 | 6822400 |
| chr2L | 6833200 | 6834480 |
| chr2L | 7302040 | 7303060 |
| chr2L | 7305740 | 7307600 |
| chr2L | 8862880 | 8863509 |
| chr2L | 9581260 | 9582460 |
| chr2L | 9607720 | 9608720 |
| chr2L | 10076810 | 10077499 |
| chr2L | 11358000 | 11359180 |
| chr2L | 11445160 | 11446940 |
| chr2L | 11447060 | 11448480 |
| chr2L | 12003280 | 12005020 |
| chr2L | 12005240 | 12008020 |
| chr2L | 12009620 | 12011919 |
| chr2L | 12085360 | 12086500 |
| chr2L | 12586520 | 12587640 |
| chr2L | 12617520 | 12619080 |
| chr2L | 12655420 | 12656520 |
| chr2L | 12677920 | 12679040 |
| chr2L | 12986910 | 12987539 |
| chr2L | 13098100 | 13098639 |
| chr2L | 14409920 | 14410980 |
| chr2L | 14489740 | 14491300 |
| chr2L | 15110920 | 15112100 |

| | | |
|-------|----------|----------|
| chr2L | 15333085 | 15334320 |
| chr2L | 15339280 | 15340320 |
| chr2L | 16422117 | 16423740 |
| chr2L | 16485320 | 16488220 |
| chr2L | 16532500 | 16533380 |
| chr2L | 16982270 | 16983139 |
| chr2L | 18735160 | 18736320 |
| chr2L | 18736960 | 18737920 |
| chr2L | 18782360 | 18784040 |
| chr2L | 18795500 | 18797180 |
| chr2L | 18880740 | 18882260 |
| chr2L | 18935740 | 18936860 |
| chr2L | 19108200 | 19109200 |
| chr2L | 19259290 | 19259909 |
| chr2L | 19768700 | 19770440 |
| chr2L | 20764200 | 20765119 |
| chr2L | 20769940 | 20771240 |
| chr2L | 20775303 | 20778340 |
| chr2L | 20782720 | 20783840 |
| chr2L | 21789160 | 21790180 |
| chr2L | 21827700 | 21828960 |
| chr2L | 21851100 | 21852260 |
| chr2L | 21866100 | 21867100 |
| chr2L | 21868380 | 21869860 |
| chr2L | 21898800 | 21899840 |
| chr2L | 22024020 | 22025220 |
| chr2R | 1613860 | 1615120 |
| chr2R | 2993620 | 2994500 |
| chr2R | 3230830 | 3231549 |
| chr2R | 3305560 | 3307320 |
| chr2R | 3316300 | 3317440 |

| | | |
|-------|---------|---------|
| chr2R | 3533450 | 3533959 |
| chr2R | 3905880 | 3907160 |
| chr2R | 3908280 | 3909640 |
| chr2R | 3911520 | 3912620 |
| chr2R | 3917780 | 3919180 |
| chr2R | 3938360 | 3939960 |
| chr2R | 4159660 | 4160720 |
| chr2R | 5081960 | 5083160 |
| chr2R | 5490190 | 5491169 |
| chr2R | 5866000 | 5867320 |
| chr2R | 5875400 | 5876600 |
| chr2R | 6690880 | 6691539 |
| chr2R | 7043840 | 7044419 |
| chr2R | 7078830 | 7079359 |
| chr2R | 7356580 | 7358045 |
| chr2R | 7362342 | 7363954 |
| chr2R | 7364840 | 7365900 |
| chr2R | 7408390 | 7409159 |
| chr2R | 7414940 | 7417800 |
| chr2R | 7429180 | 7430220 |
| chr2R | 7439140 | 7440160 |
| chr2R | 7446560 | 7447600 |
| chr2R | 7448200 | 7449100 |
| chr2R | 7450060 | 7451100 |
| chr2R | 7452680 | 7453860 |
| chr2R | 7469160 | 7470080 |
| chr2R | 8420990 | 8421599 |
| chr2R | 8705580 | 8706359 |
| chr2R | 8765500 | 8766209 |
| chr2R | 8771360 | 8772600 |
| chr2R | 8792050 | 8793631 |

| | | |
|-------|----------|----------|
| chr2R | 8842420 | 8848140 |
| chr2R | 8859860 | 8860419 |
| chr2R | 8867680 | 8869100 |
| chr2R | 8871820 | 8873900 |
| chr2R | 8877900 | 8879860 |
| chr2R | 8881080 | 8881779 |
| chr2R | 8883280 | 8885220 |
| chr2R | 8924771 | 8933108 |
| chr2R | 9209830 | 9210669 |
| chr2R | 10095080 | 10096160 |
| chr2R | 10343260 | 10344320 |
| chr2R | 10345020 | 10346440 |
| chr2R | 10364760 | 10365740 |
| chr2R | 10473680 | 10474479 |
| chr2R | 10686120 | 10687660 |
| chr2R | 10845181 | 10848201 |
| chr2R | 11216300 | 11217460 |
| chr2R | 11722160 | 11723920 |
| chr2R | 11793210 | 11793899 |
| chr2R | 12794540 | 12795099 |
| chr2R | 12817340 | 12818079 |
| chr2R | 13644260 | 13645240 |
| chr2R | 13709300 | 13711440 |
| chr2R | 13824250 | 13825109 |
| chr2R | 14405190 | 14405879 |
| chr2R | 14481120 | 14482059 |
| chr2R | 14793480 | 14794009 |
| chr2R | 14807810 | 14808569 |
| chr2R | 14853393 | 14856461 |
| chr2R | 15150680 | 15151820 |
| chr2R | 15162500 | 15163640 |

| | | |
|-------|----------|----------|
| chr2R | 15953920 | 15954439 |
| chr2R | 16220960 | 16221609 |
| chr2R | 16442560 | 16443189 |
| chr2R | 16443800 | 16444740 |
| chr2R | 16451040 | 16451589 |
| chr2R | 16738130 | 16738749 |
| chr2R | 16790560 | 16791800 |
| chr2R | 16800680 | 16803040 |
| chr2R | 16803340 | 16804740 |
| chr2R | 16847980 | 16849520 |
| chr2R | 16861560 | 16862379 |
| chr2R | 16950710 | 16951279 |
| chr2R | 17266540 | 17267520 |
| chr2R | 17565210 | 17565799 |
| chr2R | 18129020 | 18131920 |
| chr2R | 18157740 | 18158840 |
| chr2R | 18506210 | 18506809 |
| chr2R | 18747300 | 18748400 |
| chr2R | 18754800 | 18755960 |
| chr2R | 18930180 | 18931200 |
| chr2R | 18933200 | 18934320 |
| chr2R | 18981500 | 18982460 |
| chr2R | 19467880 | 19468880 |
| chr2R | 19508500 | 19509660 |
| chr2R | 19519340 | 19520480 |
| chr2R | 19534200 | 19535280 |
| chr2R | 19805730 | 19806469 |
| chr2R | 19954770 | 19955469 |
| chr2R | 20353010 | 20353629 |
| chr2R | 20699220 | 20700200 |
| chr2R | 20700800 | 20702840 |

| | | |
|-------|----------|----------|
| chr2R | 20739120 | 20740240 |
| chr2R | 20939540 | 20940700 |
| chr2R | 20948720 | 20949900 |
| chr2R | 20987240 | 20988860 |
| chr2R | 21002560 | 21003840 |
| chr2R | 21101880 | 21103420 |
| chr3L | 361380 | 362119 |
| chr3L | 370880 | 371960 |
| chr3L | 376800 | 378860 |
| chr3L | 432240 | 433320 |
| chr3L | 699220 | 700120 |
| chr3L | 1110470 | 1111329 |
| chr3L | 1176900 | 1178880 |
| chr3L | 1349370 | 1349969 |
| chr3L | 1919800 | 1921680 |
| chr3L | 1962500 | 1963109 |
| chr3L | 2904980 | 2905639 |
| chr3L | 3624480 | 3625540 |
| chr3L | 3625620 | 3626740 |
| chr3L | 3845360 | 3846720 |
| chr3L | 3855160 | 3856420 |
| chr3L | 3876520 | 3877400 |
| chr3L | 3878420 | 3879460 |
| chr3L | 5874080 | 5875680 |
| chr3L | 6090680 | 6091459 |
| chr3L | 6782380 | 6783740 |
| chr3L | 6788690 | 6789399 |
| chr3L | 6877860 | 6878589 |
| chr3L | 6896720 | 6898000 |
| chr3L | 6981580 | 6983200 |
| chr3L | 7198990 | 7199489 |

| | | |
|-------|----------|----------|
| chr3L | 7230560 | 7231169 |
| chr3L | 7827940 | 7828749 |
| chr3L | 7921670 | 7922189 |
| chr3L | 7951440 | 7952380 |
| chr3L | 8997420 | 8998720 |
| chr3L | 9011920 | 9013120 |
| chr3L | 9033620 | 9034960 |
| chr3L | 9930040 | 9930759 |
| chr3L | 10218740 | 10219980 |
| chr3L | 10282000 | 10283300 |
| chr3L | 10713710 | 10714469 |
| chr3L | 11001540 | 11004360 |
| chr3L | 11447200 | 11447849 |
| chr3L | 11920000 | 11921180 |
| chr3L | 12321440 | 12322640 |
| chr3L | 12426480 | 12428300 |
| chr3L | 12460680 | 12461920 |
| chr3L | 12572840 | 12574200 |
| chr3L | 12583620 | 12584600 |
| chr3L | 12601940 | 12603240 |
| chr3L | 12620380 | 12621420 |
| chr3L | 12645860 | 12647020 |
| chr3L | 12683420 | 12684560 |
| chr3L | 12685340 | 12687180 |
| chr3L | 13393840 | 13394840 |
| chr3L | 14124120 | 14125560 |
| chr3L | 14125580 | 14127261 |
| chr3L | 14142920 | 14143980 |
| chr3L | 14170200 | 14171720 |
| chr3L | 14545630 | 14546239 |
| chr3L | 14571580 | 14572820 |

| | | |
|-------|----------|----------|
| chr3L | 14592220 | 14594420 |
| chr3L | 14761660 | 14762539 |
| chr3L | 15036620 | 15037780 |
| chr3L | 17094070 | 17094869 |
| chr3L | 17100230 | 17100989 |
| chr3L | 17123050 | 17123569 |
| chr3L | 17499420 | 17500219 |
| chr3L | 17650650 | 17651389 |
| chr3L | 18392620 | 18394080 |
| chr3L | 19499760 | 19500609 |
| chr3L | 20629150 | 20629879 |
| chr3L | 21469180 | 21470620 |
| chr3L | 21472320 | 21473540 |
| chr3L | 22249520 | 22250419 |
| chr3R | 166560 | 167580 |
| chr3R | 171980 | 173140 |
| chr3R | 175620 | 176800 |
| chr3R | 659060 | 660060 |
| chr3R | 660400 | 661740 |
| chr3R | 677600 | 679000 |
| chr3R | 1108260 | 1108759 |
| chr3R | 2356480 | 2357099 |
| chr3R | 2503560 | 2505040 |
| chr3R | 2508520 | 2509620 |
| chr3R | 2551420 | 2552340 |
| chr3R | 2561620 | 2575018 |
| chr3R | 2616320 | 2618040 |
| chr3R | 2631280 | 2632240 |
| chr3R | 2636020 | 2637040 |
| chr3R | 2656551 | 2661848 |
| chr3R | 2673820 | 2675140 |

| | | |
|-------|---------|---------|
| chr3R | 2675220 | 2676380 |
| chr3R | 2706860 | 2708280 |
| chr3R | 2711640 | 2713052 |
| chr3R | 2718380 | 2721410 |
| chr3R | 2728500 | 2729980 |
| chr3R | 2738120 | 2739760 |
| chr3R | 2747020 | 2747980 |
| chr3R | 2755920 | 2756840 |
| chr3R | 2763040 | 2764240 |
| chr3R | 2825820 | 2827340 |
| chr3R | 2862660 | 2863509 |
| chr3R | 2961360 | 2962300 |
| chr3R | 3414450 | 3415189 |
| chr3R | 3788100 | 3789860 |
| chr3R | 4006660 | 4008220 |
| chr3R | 4016160 | 4017360 |
| chr3R | 4103160 | 4104180 |
| chr3R | 4159080 | 4161400 |
| chr3R | 4246540 | 4247309 |
| chr3R | 4528660 | 4529189 |
| chr3R | 4898960 | 4899559 |
| chr3R | 5058020 | 5058839 |
| chr3R | 5124740 | 5125379 |
| chr3R | 6001750 | 6002339 |
| chr3R | 6193600 | 6194760 |
| chr3R | 6361180 | 6362180 |
| chr3R | 6439600 | 6441160 |
| chr3R | 6442020 | 6443520 |
| chr3R | 6447500 | 6448600 |
| chr3R | 6490340 | 6491480 |
| chr3R | 7128280 | 7129360 |

| | | |
|-------|----------|----------|
| chr3R | 7137620 | 7138520 |
| chr3R | 7178800 | 7179960 |
| chr3R | 7296770 | 7297279 |
| chr3R | 8069800 | 8070840 |
| chr3R | 8104300 | 8105420 |
| chr3R | 8106100 | 8107280 |
| chr3R | 8837630 | 8838559 |
| chr3R | 8882680 | 8883820 |
| chr3R | 8889060 | 8891840 |
| chr3R | 9638040 | 9639180 |
| chr3R | 9700240 | 9701500 |
| chr3R | 9702440 | 9703620 |
| chr3R | 9720980 | 9722160 |
| chr3R | 9726800 | 9728040 |
| chr3R | 9748620 | 9749980 |
| chr3R | 9755640 | 9756720 |
| chr3R | 9973680 | 9974249 |
| chr3R | 10547800 | 10548559 |
| chr3R | 10604850 | 10605369 |
| chr3R | 10854680 | 10855399 |
| chr3R | 11222160 | 11222899 |
| chr3R | 11818420 | 11819620 |
| chr3R | 11834460 | 11836380 |
| chr3R | 11839020 | 11840020 |
| chr3R | 11851270 | 11852049 |
| chr3R | 11863320 | 11864500 |
| chr3R | 11947050 | 11947549 |
| chr3R | 12210900 | 12212080 |
| chr3R | 12228800 | 12230220 |
| chr3R | 12239140 | 12240240 |
| chr3R | 12248660 | 12249920 |

| | | |
|-------|----------|----------|
| chr3R | 12487100 | 12488160 |
| chr3R | 12503400 | 12504400 |
| chr3R | 12527147 | 12529718 |
| chr3R | 12543920 | 12544880 |
| chr3R | 12545340 | 12546520 |
| chr3R | 12589020 | 12590917 |
| chr3R | 12619900 | 12620900 |
| chr3R | 12636230 | 12638320 |
| chr3R | 12641040 | 12642120 |
| chr3R | 12644800 | 12645920 |
| chr3R | 12655380 | 12656500 |
| chr3R | 12662960 | 12664020 |
| chr3R | 12672560 | 12673700 |
| chr3R | 12674100 | 12675060 |
| chr3R | 12677440 | 12678400 |
| chr3R | 12682960 | 12684200 |
| chr3R | 12694584 | 12695860 |
| chr3R | 12707720 | 12709000 |
| chr3R | 12711800 | 12712960 |
| chr3R | 12722831 | 12726610 |
| chr3R | 12738860 | 12739980 |
| chr3R | 12744380 | 12746860 |
| chr3R | 12748920 | 12750040 |
| chr3R | 12759500 | 12760700 |
| chr3R | 12763480 | 12764520 |
| chr3R | 12765960 | 12766980 |
| chr3R | 12773860 | 12775080 |
| chr3R | 12778380 | 12779300 |
| chr3R | 12783240 | 12784780 |
| chr3R | 12785560 | 12786780 |
| chr3R | 12789000 | 12789920 |

| | | |
|-------|----------|----------|
| chr3R | 12794760 | 12796620 |
| chr3R | 12800100 | 12801020 |
| chr3R | 12808380 | 12809500 |
| chr3R | 13382860 | 13385060 |
| chr3R | 13392160 | 13393160 |
| chr3R | 13665160 | 13665709 |
| chr3R | 13769340 | 13770280 |
| chr3R | 14061260 | 14062340 |
| chr3R | 14196820 | 14197920 |
| chr3R | 14524740 | 14526129 |
| chr3R | 14538050 | 14538809 |
| chr3R | 15662120 | 15663120 |
| chr3R | 16672200 | 16673220 |
| chr3R | 17204060 | 17205260 |
| chr3R | 17259420 | 17260740 |
| chr3R | 17271600 | 17272940 |
| chr3R | 17324640 | 17326000 |
| chr3R | 17382440 | 17384280 |
| chr3R | 17386860 | 17388060 |
| chr3R | 18751520 | 18752039 |
| chr3R | 18957657 | 18960119 |
| chr3R | 18967140 | 18968380 |
| chr3R | 18968720 | 18969920 |
| chr3R | 19018720 | 19019820 |
| chr3R | 19022740 | 19023640 |
| chr3R | 19026380 | 19027620 |
| chr3R | 19032620 | 19033680 |
| chr3R | 19036960 | 19038360 |
| chr3R | 19118340 | 19119520 |
| chr3R | 19120680 | 19121780 |
| chr3R | 19122580 | 19123540 |

| | | |
|-------|----------|----------|
| chr3R | 20274110 | 20274859 |
| chr3R | 20412020 | 20412759 |
| chr3R | 20563690 | 20564199 |
| chr3R | 20906070 | 20906699 |
| chr3R | 21514670 | 21515599 |
| chr3R | 21846160 | 21847860 |
| chr3R | 21924480 | 21925480 |
| chr3R | 22376610 | 22377489 |
| chr3R | 22415140 | 22415789 |
| chr3R | 22612790 | 22613499 |
| chr3R | 23343810 | 23344579 |
| chr3R | 23543670 | 23544339 |
| chr3R | 23989520 | 23990089 |
| chr3R | 24069750 | 24070279 |
| chr3R | 24074310 | 24074839 |
| chr3R | 24133610 | 24134229 |
| chr3R | 24372480 | 24373700 |
| chr3R | 24410580 | 24411940 |
| chr3R | 25378140 | 25379260 |
| chr3R | 25381260 | 25382600 |
| chr3R | 25397270 | 25398620 |
| chr3R | 25515680 | 25518740 |
| chr3R | 25598120 | 25599000 |
| chr3R | 25855500 | 25856580 |
| chr3R | 26098320 | 26100420 |
| chr3R | 26441780 | 26443020 |
| chr3R | 26543020 | 26543759 |
| chr3R | 26562800 | 26563559 |
| chr3R | 26589380 | 26590640 |
| chr3R | 26590920 | 26591920 |
| chr3R | 26592360 | 26593560 |

| | | |
|-------|----------|----------|
| chr3R | 26598160 | 26599240 |
| chr3R | 26606380 | 26608340 |
| chr3R | 26677180 | 26678440 |
| chr3R | 26734440 | 26735480 |
| chr3R | 26737460 | 26738960 |
| chr3R | 26887620 | 26888840 |
| chr3R | 27363230 | 27363909 |
| chr3R | 27897320 | 27898599 |
| chrX | 482620 | 483720 |
| chrX | 484100 | 485380 |
| chrX | 1465980 | 1466709 |
| chrX | 1751910 | 1752599 |
| chrX | 1832030 | 1832699 |
| chrX | 1846700 | 1847580 |
| chrX | 2016880 | 2021172 |
| chrX | 2030422 | 2033330 |
| chrX | 2036680 | 2037760 |
| chrX | 2322722 | 2324624 |
| chrX | 2327298 | 2330636 |
| chrX | 2764990 | 2765779 |
| chrX | 3153980 | 3154819 |
| chrX | 4271780 | 4272289 |
| chrX | 4304880 | 4306240 |
| chrX | 4307790 | 4308509 |
| chrX | 4317420 | 4319900 |
| chrX | 4320260 | 4321640 |
| chrX | 4401760 | 4403100 |
| chrX | 4420540 | 4421520 |
| chrX | 4448460 | 4449440 |
| chrX | 4483500 | 4484900 |
| chrX | 4490140 | 4491660 |

| | | |
|------|---------|---------|
| chrX | 4510340 | 4512020 |
| chrX | 4556370 | 4556929 |
| chrX | 4881030 | 4881839 |
| chrX | 5104440 | 5105159 |
| chrX | 5157100 | 5157819 |
| chrX | 5454060 | 5455400 |
| chrX | 5457980 | 5458900 |
| chrX | 5473000 | 5474360 |
| chrX | 5487500 | 5488700 |
| chrX | 5894400 | 5895860 |
| chrX | 5906800 | 5908140 |
| chrX | 6353870 | 6354789 |
| chrX | 6379120 | 6379979 |
| chrX | 7004520 | 7005620 |
| chrX | 7119740 | 7120700 |
| chrX | 7449730 | 7450249 |
| chrX | 7520260 | 7521740 |
| chrX | 8041710 | 8042469 |
| chrX | 8171430 | 8172269 |
| chrX | 8324250 | 8324869 |
| chrX | 8348840 | 8349720 |
| chrX | 8530040 | 8531600 |
| chrX | 8534320 | 8535580 |
| chrX | 8544240 | 8546060 |
| chrX | 8570280 | 8571360 |
| chrX | 8651440 | 8652560 |
| chrX | 8699380 | 8702020 |
| chrX | 8769530 | 8770209 |
| chrX | 9430870 | 9431389 |
| chrX | 9586340 | 9588680 |
| chrX | 9622680 | 9624080 |

| | | |
|------|----------|----------|
| chrX | 9691870 | 9692389 |
| chrX | 9767360 | 9768219 |
| chrX | 10500130 | 10500759 |
| chrX | 10862540 | 10863149 |
| chrX | 10948390 | 10949239 |
| chrX | 11097230 | 11098069 |
| chrX | 11560680 | 11561359 |
| chrX | 11974910 | 11975689 |
| chrX | 12316580 | 12317209 |
| chrX | 12449570 | 12450279 |
| chrX | 12483460 | 12484069 |
| chrX | 12509420 | 12510560 |
| chrX | 12671680 | 12672439 |
| chrX | 12755880 | 12756409 |
| chrX | 13820470 | 13821059 |
| chrX | 13871990 | 13872759 |
| chrX | 13888440 | 13889880 |
| chrX | 14337960 | 14338739 |
| chrX | 14804700 | 14805199 |
| chrX | 15249980 | 15251200 |
| chrX | 16041160 | 16042580 |
| chrX | 16042640 | 16043820 |
| chrX | 16044160 | 16045380 |

| | | |
|------|----------|----------|
| chrX | 16050930 | 16051639 |
| chrX | 16110200 | 16112300 |
| chrX | 17037000 | 17037809 |
| chrX | 17191780 | 17192529 |
| chrX | 17207020 | 17209000 |
| chrX | 17216120 | 17216839 |
| chrX | 17239520 | 17240740 |
| chrX | 17290740 | 17291900 |
| chrX | 17601930 | 17602789 |
| chrX | 17658620 | 17659640 |
| chrX | 17661660 | 17663100 |
| chrX | 17682700 | 17684960 |
| chrX | 17696060 | 17697080 |
| chrX | 18077030 | 18077889 |
| chrX | 18202900 | 18204220 |
| chrX | 18757770 | 18758559 |
| chrX | 19289800 | 19290599 |
| chrX | 19429300 | 19430280 |
| chrX | 19439560 | 19441460 |
| chrX | 20560060 | 20562260 |
| chrX | 20564560 | 20565920 |
| chrX | 20747910 | 20748549 |
| chrX | 20825720 | 20826239 |

Supplementary Table 2-2 List of PRES used in this work

| Data set | Source |
|---|--|
| RNA-seq embryos | SRA009364 |
| RNA-seq S2 cells | GSE72830 |
| All_EST | UCSC (<i>D. melanogaster</i> ESTs - <i>D. melanogaster</i> ESTs Including Unspliced), dm3 |
| PREs | Predicted PREs (Ringrose et al., 2003) |
| | Conserved PcG binding sites (Schuettengruber et al., 2014) |
| | vg PRE (Herzog et al., 2014) |
| | list of PREs (Erceg et al., 2017) |
| | dac PRE (Ogiyama et al., 2018) |
| ChIP-seq, 4-12 hour embryos (Ph, PC, H3K27me3, DSP1) (merge of 2 replicates) (18) | GSE60428 (used to compare to both 2-6H and 10-14H R-loop data) |
| ChIP-seq, H3K27Ac, H3K27me3, RNA pol II 0-4H, 4-8H, 8-12H, 12-16H embryos | GSE16013 |
| ChIP-seq, S2 cells (E(Z), dRING, GAF) | GSE101554 |
| ChIP-seq, S2 cells H3K27Ac, H3K27me3 | modENCODE_326; GSE41440 |
| ChIP-chip, S2 cells, RNA pol II | modENCODE_3295 |

Supplementary Table 2-3 Data sets used for comparison with DRIP-seq data

2.5 Methods

S2 cell culture: *Drosophila* S2 cells were purchased from Invitrogen, and grown in Schneider's media (Invitrogen) with 10% heat inactivated, insect cell tested FBS (Invitrogen). Cells were cultured at 27°C in suspension in shaking flasks.

***Drosophila* collection:** Oregon R flies were grown at 25°C. Embryos were collected on apple juice plates and dechorionated for 2 min. in 50% bleach before being washed with H₂O and stored at -80°C.

Total nucleic acid extraction from S2 cells: 8*10⁷ S2 cells were washed with 1X PBS and resuspended in 10 mL TE. Cells were lysed O.N. at 37°C in presence of 0.5% SDS and 62.5 µg/mL of proteinase K. After phenol-chloroform-isoamyl alcohol extraction, total nucleic acids were precipitated in the presence of 0.3 M sodium acetate pH 5.2 and 2.4 volume of 100% ethanol. Nucleic acids were washed carefully 5 times with 70% ethanol, and resuspended in TE.

Total nucleic acid extraction from *Drosophila* embryos: Total nucleic acids were extracted from 500 µL of Oregon R embryos as described in Ejsmont et al., 2009 (Ejsmont et al., 2009) with the omission of RNaseA. After precipitation the nucleic acids were washed carefully 5 times with 70% ethanol, and resuspended in TE. This material was subsequently processed for DRIP analysis as described below.

DRIP-seq and DRIP-qPCR: The DRIP protocol was adapted from Ginno et al. (Ginno et al., 2012). 500 µg of total nucleic acid were divided in 3 and each treated with 150 µg of RNaseA in presence of 0.5 M NaCl for 3 hours at 37°C. gDNA was purified by phenol-chloroform extraction followed by ethanol precipitation and sonicated to 300 bp using a Covaris. Fragmented gDNA was treated with 2 U of RNaseIII (Hartono et al., 2018) (Thermo Fisher) +/- 10 µg each of homemade RNaseH I and RNaseH II overnight at 37°C. Immunoprecipitation was performed as described in Ginno et al. (Ginno et al., 2012). After elution, samples were purified with a PCR clean-up column (Macherey-Nagel) with NTB buffer to get rid of SDS followed by a DNA clean and concentrator column (Zymo Research). For sequencing library preparation, material from three immunoprecipitations were pooled. Libraries were prepared using the NEB next Ultra II kit for a directional library for Illumina (NEB). For strand specific DNA sequencing of the RNA-DNA hybrids, we started at the second strand synthesis step and ligated with NEB-next multiplex oligonucleotides for Illumina (NEB). Paired-end sequencing was performed on an Illumina HiSeq 2500 at Genome Quebec.

For qPCR, input was diluted 10-fold and IPs 2-fold in water. PCR was carried out in 5 µL reactions consisting of 2 µL DNA, 2.5 µL PowerUp SYBR Green master mix (Thermo Fisher) and 0.25 µL of a 1 µM stock of each

primer diluted in water. Standard curves were generated using a log titration of *Drosophila* genomic DNA purified from S2 cells (25 to 0.025 ng). Data were collected using a Vii7 PCR system (Thermo Fisher) with 40 cycles. The standard curve was used to calculate DNA amounts. All standard curves had R^2 values of 0.9 or higher. Oligonucleotides used for qPCR (Ogiyama et al., 2018; Papp and Muller, 2006) are listed in Supplementary Table 2-1.

DRIP-seq analysis: FastQ files of DRIP-seq reads were trimmed with Trimmomatic (PE -phred33), using the GenPipes ChIP-seq pipeline (steps 1-3) (Bourgey et al., 2019). Reads with both mate pairs were aligned to the dm3 version of the *Drosophila* genome using Bowtie2/2.3.1(--fr --no-mixed --no-unal) (Langmead and Salzberg, 2012). Sam files generated by Bowtie2 were converted to bam, sorted and indexed (samtools (v. 1.4.1) (Li et al., 2009) and Picard (<http://broadinstitute.github.io/picard>) MarkDuplicates (default parameters) was used to remove duplicates. To generate strand specific bam files, samtools was used as follows:

Forward strand: samtools view -f 99; samtools view -f 147, followed by samtools merge

Reverse strand: samtools view -f 83; samtools view -f 163, followed by samtools merge.

Peaks were called for DRIP versus input and DRIP versus RNaseH treated using MACS2(Zhang et al., 2008) (v. 2.1.1) (-f BAMPE --bw 250 -g dm --mfold 10 30 -q 0.01). For strand specific peaks, strand specific files were used (e.g. F-strand DRIP, F-strand input, F-strand RNaseH). Peaks present in both DRIP vs. input and DRIP vs. RNaseH were retained (BEDTools intersect) (Quinlan and Hall, 2010) for each duplicate. Finally, BEDTools (intersect) was used to retain only peaks present in both duplicates, which were used for further analysis. The correlation between the replicates was examined using multiBigwigSummary on Galaxy (bin size: 1000 bp) followed by plotCorrelation using the Pearson correlation method. Correlations for replicates were: 2-6H 0.97, 10-14H 0.87, S2 0.99. Bigwig files were generated using DeepTools (Ramirez et al., 2014) v 2.5.3 (--binSize 10\ --normalizeUsingRPKM)

A list of PREs (Table S2) was generated by combining predicted PREs (Ringrose et al., 2003), PcG binding sites conserved through *Drosophila* species (Schuettengruber et al., 2014), and additional PREs from recent reports (Erceg et al., 2017; Herzog et al., 2014; Ogiyama et al., 2018). Multiple PREs predicted in the repeated histone gene clusters were removed, although ChIP-seq peaks for PcG proteins are observed at these sites. Finally, overlapping or touching PREs were merged (using BEDTools). The list of genomic coordinates for PREs is in Supplementary Table 2-2.

To analyze overlaps between R-loops and PREs or other genomic elements, bed files of peak calls of unstranded, forward, and reverse strand peaks were merged to produce a consolidated set of R-loops. Overlap of R-loops or PREs with different genomic elements (Supplementary Fig. 1e-h) were generated with Pavis, with upstream and downstream regions both set at 5000 bp (Huang et al., 2013). To correlate gene expression levels with R-loop formation (Supplementary Fig. 2a, b) the overlapping or closest gene to each R-loop was identified using BEDTools, ClosestBed on Galaxy. Level of gene expression were determined using RNA-seq data from embryos or S2 cells and genes were divided into categories based on their FPKM level (no to extremely low expression: $FPKM < 1$, low expression: $1 < FPKM < 10$, moderate expression: $10 < FPKM < 50$ and high expression: $FPKM > 50$). To compare R-loop orientation to annotated transcripts, the “all EST” track was downloaded from UCSC, and BEDTools was used (intersect intervals, only overlaps occurring on the same strand).

To analyze overlap of PREs with PcG protein, RNA Pol II, or H3K27me3 ChIP-seq peaks (Fig. 1d-f, Supplementary Fig. 2d-k), previously processed bed files were used with BEDTools (intersect). To analyze ChIP-seq signal intensity over PREs with and without R-loops, raw data (FASTQ files) were downloaded using the SRA toolkit (v2.9.6) (<http://ncbi.github.io/sra-tools/>, SRA Toolkit Development Team), aligned with Bowtie2 as described above, duplicates removed (Picard), and RPKM-normalized bigwig files generated (DeepTools bamCoverage). BEDOPS (Neph et al., 2012) (v2.4.34) was used to convert bigwig files to wig and then bed files, and read densities quantified using BEDOPS bedmap (bedmap -count -echo-ref-name). Read densities over each PRE were divided by the PRE length to obtain the final values. All data sets used to analyze R-loops are listed in Supplementary Table 2-3.

To analyze overlap with annotated genes and RNA, we first converted DRIP-seq peaks to the dm6 genome using the UCSC genome browser liftOver tool. DRIP-seq peaks mapping to heterochromatin, ChrU, and ChrMT were removed. Unstranded DRIP-seq peaks were overlapped with either all ESTs (“mRNA and EST”) or all genes (“Genes and Gene Predictions”, UCSC Table Browser) using bedtools -intersect to produce the table shown in Extended Data Figure 1c.

To analyze the overlap with annotated RNAs in a strand specific manner, we used the strand-specific peak calls. Files were prepared as described above. To remove peaks with R-loops called on both strands, we first intersected the F and R strand files, and removed DRIP-seq peaks that were called on both strands. The remaining peaks were overlapped with all ESTs using bedtools intersect with the -s or -S options to obtain overlaps with sense and anti-sense transcripts. The same process was carried out for PREs, except

that PREs were first intersected with F and R DRIP-seq peaks. In the analysis shown in Extended Data Figure 2c, “sense” indicates that the peak overlaps an annotated RNA in the sense orientation, but may also overlap an annotated RNA in the antisense orientation; “antisense” indicates overlap only with an antisense transcript.

Protein expression and purification:

Human RNaseH1: A 6X-His tag was added to MBP-hRNaseH1, which was expressed in and purified from *E. coli* based on a previously described protocol (Loomis et al., 2014; Reijns et al., 2011), except that Ni-NTA beads were used for the first step instead of amylose beads.

hRNaseH2: The RNaseH2 trimer was produced using the multi-cistronic pMAR22 vector essentially as described (Reijns et al., 2011).

PRC1ΔPh, PRC2, dSxc, hNFY: PRC1ΔPh and PRC2 were expressed in and purified from Sf9 cells, with the following modifications to previously published protocols for anti-FLAG affinity purification (Abmayr et al., 2006; Francis et al., 2001; Muller et al., 2002). For PRC1ΔPh, nuclear extracts were prepared from Sf9 cells infected with viruses for the 4 subunits (Abmayr et al., 2006) but nuclei were purified through a sucrose cushion prior to nuclear extraction. During the purification, the 2 M KCl wash in the published protocol was replaced with a wash consisting of BC200N + 1 M Urea (20 mM Hepes, pH 7.9, 2 0.4 mM EDTA, 2 M KCl, 1 M deionized urea, 0.05% NP40, no glycerol). Additionally, prior to eluting the protein, anti-FLAG beads were incubated 3-5 volumes of BC300N with 4 mM ATP + 4 mM MgCl₂ for 30 min. at room temperature. This step reduces the amount of HSC-70 that co-purifies with PRC1ΔPh.

For PRC2 expression and purification, E(Z) was tagged with 6-His, and either Esc or Su(Z)12 with FLAG, and baculovirus infected Sf9 cells were harvested after 3 days. PRC2 was purified by anti-FLAG affinity as described (Muller et al., 2002) followed by Ni-NTA. FLAG peptide elutions were carried out in BC300 without EDTA or DTT. FLAG elutions were passed over Ni-NTA beads twice, beads were washed with 30 volumes of BC300 (without EDTA or DTT) and eluted in BC300 + 250 mM Imidazole. Eluted protein was pooled and dialyzed through 3 changes of BC300 with EDTA, PMSF, and DTT. PRC2 was concentrated to ~1 mg/ml, NP40 was added to 0.05%, and protein was stored at -80° C.

Extract preparation and anti-Flag purification of F-Sxc and F-NFY were as described for PRC2.

For glycerol gradient fractionation of PRC1ΔPh or PRC2 (Supplementary Fig. 4e, f), 5-10 μg of protein were loaded on a 280 μl step gradient (35/30/25/20/15/10/5% glycerol) in BC300 buffer. Gradients were centrifuged for 3 hours at 367,600 g at 4°C using an SW55Ti rotor and resolved into 50 μl fractions.

Oligonucleotides assembly and labelling: DNA and RNA oligonucleotides described in Supplementary Table 2-1 were diluted to 1 μM in TE supplemented with 50 mM NaCl (TE-50), boiled and cooled O.N. with the exception of ssDNA which was snap frozen. All substrates were gel purified on 8% acrylamide 0.5 X TBE gels and nucleic acids were extracted O.N. by incubated the band of interest in 20 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.1% SDS, 10 mM EDTA. Nucleic acids were precipitated with ethanol, washed with 70% ethanol and resuspended in TE-50. Substrates were labelled with T4 PNK (New England Biolabs, NEB) and ³²P-γATP (PerkinElmer). Samples were purified by phenol-chloroform extraction, followed by purification through a G-25 spin column equilibrated with TE-50.

Filter binding assay: Filter binding was carried out essentially as described (Wong and Lohman, 1993), using a nitrocellulose filter (to capture protein-nucleic acid complexes) (Biotrace) stacked on a charged nylon membrane (to capture free nucleic acids) (HYBOND membrane, GE Healthcare) with a slot blot apparatus. Nitrocellulose filters were prepared by incubating in 0.4 M KOH for 10 min, washing extensively with H₂O, and equilibrating at least one hour in binding buffer. Nylon filters were equilibrated at least 10 min. in 0.4 M Tris, pH 8.0. We first measured the active concentration of two preparations of PRC1ΔPh; concentrations reported in Fig. 2 are the active concentration. To measure binding to oligonucleotide substrates, PRC1ΔPh and PRC2 were titrated into reactions with 0.01 nM of DNA bubble, R-loop or dsDNA in 20 μl reactions containing 12 mM Hepes, pH 7.9, 0.12 mM EDTA, 120 nM KCl, 1 mM DTT, 0.01% NP40, and 12% glycerol and incubated 30 min. at 30° C. To apply reactions to the filters, each well was washed with 100 μL binding buffer, sample was applied, and wells were washed twice with 100 μL of binding buffer. Membranes were exposed to a phosphor imager screen, scanned on a Typhoon imager (GE Healthcare), and quantified with ImageQuant.

EMSA: PRC1ΔPh and PRC2 were titrated into reactions with 0.01 nM of DNA bubble, R-loop or dsDNA in the same reaction conditions as for filter binding except that 50 ng/μL BSA were included. Reactions were resolved on 6% acrylamide 0.5 X TBE gel. Gels were dried and exposed to a phosphor imager screen and scanned on a Typhoon imager.

PRC1ΔPh and PRC2 incubation with filter binding probes: PRC1ΔPh and PRC2 were titrated into reactions with 20 nM of DNA bubble, R-loop, dsDNA, ssDNA and RNA in the same conditions as for filter binding. Proteins were digested with 3 μL of DSB-PK (6.7 μg/μL of proteinase K (Biobasic), 1% SDS, 50 mM Tris-HCl

pH 8.0, 25% glycerol and 100mM EDTA) for 30 min. at 50°C, nucleic acids were resolved on 8% acrylamide 0.5X TBE gels and stained with SYBRGold (Thermo Fisher).

Plasmids for strand exchange: PRE sequences were amplified by PCR from *Drosophila* genomic DNA and cloned into the pET-Blue1 vector (Millipore) downstream of the T7 promoter. Detailed maps are available on request. For strand exchange, plasmids were digested with a single restriction enzyme, purified by phenol-chloroform extraction, and ethanol precipitated.

RNA production and labelling: RNAs were produced from linear templates using the Ampliscribe T7-flash transcription kit (Lucigen) using the manufacturer's protocol in the presence of 25 mM of amino-allyl UTP (Sigma). After purification, RNAs were labelled with NHS-Cyanin-5 (Kerafast) in 70 mM NaHCO₃ pH 8.8 with murine RNase inhibitor (NEB) for 2 hours at RT. RNAs were then precipitated with 0.3 M sodium acetate pH 5.3, glycogen and ethanol, washed with 70% ethanol, and resuspended in TE. RNAs were passed through a G50 column equilibrated with TE. The quality of labelled RNA and efficient removal of free dye were determined by loading the RNA on agarose gels.

Radiolabelled RNAs were produced from circular templates by transcribing 600 ng of DNA in RNA polymerase buffer (NEB), 1 mM DTT, 625 μM of rNTP (NEB), 6.5 nmol of radiolabelled UTP, 200 U of RNase inhibitor, and 250 U of T7 or T3 RNA polymerases (NEB) in 100 μL O.N. at 37°C. DNA was removed from the reaction by adding 4 U of DNaseI (NEB) and incubating 2 hours at 37°C. RNAs were extracted with phenol-chloroform, ethanol precipitated, washed with 70% ethanol, resuspended in TE, and stored at -20°C.

RNA strand exchange assay: PRC2, diluted in BC300N was incubated with the indicated amount of DNA and fluorescent- or radio- labelled RNA for 25 min. at 30°C in 180 mM KCl, 5 mM MgCl₂, 1 mM DTT and with 50 ng/ μL BSA in 10 μL reaction. After incubation, samples were treated with 3 μL of DSB-PK for 30 min. at 50°C and resolved on 0.8% agarose 0.5X TBE gel. Gels were stained with SYBRGold or ethidium bromide, and imaged on a Typhoon Imager. For experiments with radio-labelled RNA, gels were transferred to HYBOND membrane and exposed to a phosphor imager screen.

For nuclease treatment of strand invasion products without phenol-chloroform extraction, after incubation with PRC2, samples were treated immediately with nucleases. For RNaseH treatment, 10X RNaseH buffer was added to a final concentration of 1X, followed by 2.5 U (radio-labelled RNA) or 1.25 U (fluorescently labelled RNA) of RNaseH. For RNaseA treatment, reactions were supplemented with 500 mM NaCl and 50 pg of RNaseA (Qiagen) were added. Reactions were incubated for 30 min. at 30°C. For

phenol-chloroform extracted samples, reactions were stopped with 3 μ L of DSB-PK and incubated for 30 min. at 50°C. Nucleic acids were extracted with phenol-chloroform followed by ethanol precipitation and resuspension in TE. Nuclease digestion was carried out as described above. Nuclease digestions were stopped by the addition of 3 μ L of DSB-PK, and samples were incubated 30 min. at 50°C before analyzing on agarose gels. When the order of DNA and RNA addition was tested, the first nucleic acid was added to PRC2 for 10 min. at 30°C before the addition of the second.

For the pre-incubation assay shown in Fig. 4d-f, DNA was incubated with PRC2 under conditions described above, proteins were removed by digesting with DSB-PK O.N. at 50°C. DNA was purified using a PCR clean up column (Macherey-Nagel) and eluted in 10 mM Tris. This DNA was used in RNA strand exchange reactions.

RNA strand exchange after apyrase treatment: 10 μ g of PRC2 and 330 ng of Cy5-labelled RNA were incubated 30 min. at 30°C with respectively 0.5 and 1 unit of apyrase (NEB) in 15 or 30 μ L reaction. Apyrase was removed by passing PRC2 and RNA on G50 column equilibrated respectively with BC300N and TE. PRC2 and RNA treated with apyrase were used in strand exchange reactions.

RNA strand exchange followed by DRIP in vitro: DNA, RNA and PRC2 was incubated as describe for RNA-DNA strand exchange in 100 μ L reactions. Reactions were stopped with 30 μ L of DSB-PK and digested for 30 min. at 50°C. Nucleic acids were extracted with phenol-chloroform, ethanol precipitated, washed with 70% ethanol and resuspended in TE. RNA-DNA hybrid containing fragment were immunoprecipitated by incubated O.N. at 4°C with protein G beads pre-incubated with 2.5 μ g S9.6 antibody (Kerafast ENH001 and M. Wilson) and competitor DNA (pUC19 plasmid digested with DrdI, AlwNI and Scal-HF). Beads were washed three times with 10 mM NaPO₄ pH 7.0, 140 mM NaCl and 0.05% Triton X-100. Nucleic acids were eluted from the beads by incubating them 45 min. at 50°C in presence of 50 mM Tris pH 8.0, 10 mM EDTA, 0.5% SDS and 56 μ g proteinase K. Samples were resolved on 0.8% agarose 0.5X TBE gels. Gels were stained with SYBRGold and imaged on a Typhoon Imager.

Oligonucleotide Annealing assay: Phosphorylated ssDNA or RNA oligos corresponding to a sequence in the bxd PRE (Supplementary Table 2-1) were used at a final concentration of 40 nM. One ssDNA oligo is 5' labelled with Cy5. Annealing was carried out with the same reaction conditions as RNA strand exchange except that no MgCl₂ was added and the [KCl] was 60 mM. Reactions were incubated for 25 min. at 15°C, and stopped by adding 200 nM of unlabelled ssDNA oligonucleotides and 25 ng of *vg* RNA and incubation for 10 min at 15°C. Reactions were loaded on 8% acrylamide 0.5 X TBE gels. Gels were stained with SYBRGold and imaged for Cy5 and SYBRGold on a Typhoon Imager.

Nuclease activity assay on oligonucleotides: Oligonucleotides were phosphorylated with T4 PNK (NEB). dsDNA oligonucleotides were annealed by incubating equal amounts of ssDNA in TE-50, boiling for 5 min. and slow cooling over several hours. ssDNA oligonucleotides were boiled and transferred immediately to ice. PRC2, T7 endonuclease (NEB), Exonuclease I (NEB), λ exonuclease (NEB) or Exonuclease III (NEB) were incubated with phosphorylated ssDNA or dsDNA oligonucleotides under RNA-DNA hybrid forming conditions. Reactions were stopped by adding DSB-PK and incubated 1 hour at 50° C. Samples were denatured by addition of 26% formamide, 0.3 mM EDTA, 3.3 mM NaOH and boiled before loading on denaturing gels (10% acrylamide, 1X TBE, 7 M urea). Gels were stained with SYBRGold and imaged on Typhoon Imager.

Gel quantification: For quantification of DNA and RNA-DNA hybrids from phosphor imager and SYBR gold scans using ImageQuant, RNaseA treated samples were used. In cases where gel flaws obscured quantification of a lane, the gradient was excluded from analysis. Background subtraction was done using the rolling ball method. For band selection, the smallest possible “fixed width” bands that capture the whole signal were set for each gradient. These bands were placed in each lane so that every fraction was quantified. The signal from the bottom three fractions was divided by that for the total of the gradient for the fraction bound.

RNA strand exchange assay gels of Cy5 labelled RNA were imaged using Typhoon Imager (GE Healthcare) were quantified using ImageQuant (GE Healthcare). Lanes were created manually, then background was removed using minimum profile method and bands were identified manually.

Electron microscopy: 200 nM of PRC2 or 20 units of exonuclease III were incubated with 1.9 nM of pFC53 DNA linearized with HindIII using RNA-DNA hybrid formation conditions. Reactions were stopped with 15 μ L of DSB-PK and incubated O.N. at 50° C. DNA was purified on PCR clean-up columns (Macherey-Nagel) with NTB buffer. DNA was incubated with *E. coli* SSB (NEB) at a ratio of 6 μ g SSB per μ g DNA on ice for 10 min. (Thresher and Griffith, 1992) Glutaraldehyde (Electron Microscopy Sciences) was added to a final concentration of 0.6% and samples incubated 10 min. on ice. Samples were purified through a G-50 column equilibrated in TE. Samples were diluted 1: 5 in 2 mM Spermidine, 150 mM NaCl, 1 mM MgCl₂, and applied to glow-discharged 400-mesh grids coated with thin carbon (#Cu-400CN, Pacific Grid Tech) for 5 min. Grids were washed through two drops of 100 mM MgOAc, and stained with 3 drops of 2% Uranyl Acetate (Electron Microscopy Sciences), blotted and air-dried. Grids were photographed at 120 kV using a FEI Technai G2 Spirit BioTwin Cryo-TEM at the McGill Facility for Electron Microscopy Research. Note that

both positively and negatively stained regions were observed on the grids; positive staining is shown in Fig. 4 because it facilitates visualization of SSB coated DNA filaments.

Statistics & curve fitting: Graphpad Prism was used for statistics and curve fitting. For time course data, the equation $Y=AB_{max}*(1-e^{-k*X})$ was used; for binding data, $Y=(AB_{max}*X)/(X+K_d)+b$ was used. Fisher's exact test was used to compare curve fits. For students t-tests (Fig. 2), we used Holm-Sidak correction for multiple comparisons, $\alpha=0.05$, with all points assumed to come from populations with the same S.D. For comparing distribution of ChIP-seq peaks or transcripts (Supplementary Fig. 2), data were organized as a contingency table (i.e. columns=peak/no peak, rows=R-loop/no R-loop) and compared using Fisher's exact test, reporting two-sided P-values. To compare ChIP-seq read intensities over PREs, Mann-Whitney tests were used, with 2-tailed p-values reported.

Regioner (Gel et al., 2016) was used to conduct permutation tests of the overlaps between PREs and R-loops, or ChIP-seq peaks and PREs with and without R-loops (1000 permutations, `randomize.function=randomizeRegions`, `evaluate.function=numOverlaps`, `count.once=TRUE`, `genome="dm3"`).

Data Availability: Sequence (DRIP-seq) data that support the findings of this study have been deposited in NCBI GEO with the accession code GSE127329. Other data that support the findings of this study are available from the corresponding author upon reasonable request. The source data underlying Figures 2a, 2b, 2c, 2d, 3g, 4f, 6b, 6d, and 6f and Supplementary Figures 1b, 6d, 7b, 7c and 7c are provided as a Source Data file.

2.6 Acknowledgements

We thank K. Sears at the McGill Facility for Electron Microscopy for assistance with sample preparation, E. Lécuyer's lab for assistance collecting *Drosophila* embryos, O. Neyret for advice on preparation of NGS samples, J. Mallette for technical assistance, C. Gentile for advice on bioinformatics, M. Wilson for S9.6 antibody, M. Reijns for plasmid to express hRNaseH2, M. Drolet for intellectual support, J.-Y. Masson for suggesting the EM experiment, Y-S Kang for help with bioinformatics, and F. Robert, M. Drolet, and members of the Francis lab for comments on the manuscript. This research was enabled in part by support provided by Calcul Québec (www.calculquebec.ca) and Compute Canada (www.computecanada.ca).

Work in the N.J.F. lab was funded by CIHR 311557, in the K.J.A. lab by a grant from the David and Lucile Packard Foundation, NIH 5R01GM115882-03, and 5T32HL007151-40 (to D.G.), and in the F.C. lab by NIH R01-GM120607. **Competing interests:** The authors declare no competing interests. **Materials & Correspondence:** All correspondence and material requests should be directed to N.J.F.

2.7 References

- Abmayr, S.M., Yao, T., Parmely, T., and Workman, J.L. (2006). Preparation of nuclear and cytoplasmic extracts from mammalian cells. *Curr Protoc Pharmacol Chapter 12*, Unit12 13.
- Amiard, S., Doudeau, M., Pinte, S., Poulet, A., Lenain, C., Faivre-Moskalenko, C., Angelov, D., Hug, N., Vindigni, A., Bouvet, P., *et al.* (2007). A topological mechanism for TRF2-enhanced strand invasion. *Nat Struct Mol Biol* 14, 147-154.
- Bauer, M., Trupke, J., and Ringrose, L. (2016). The quest for mammalian Polycomb response elements: are we there yet? *Chromosoma* 125, 471-496.
- Beltran, M., Tavares, M., Justin, N., Khandelwal, G., Ambrose, J., Foster, B.M., Worlock, K.B., Tvardovskiy, A., Kunzelmann, S., Herrero, J., *et al.* (2019). G-tract RNA removes Polycomb repressive complex 2 from genes. *Nat Struct Mol Biol* 26, 899-909.
- Beltran, M., Yates, C.M., Skalska, L., Dawson, M., Reis, F.P., Viiri, K., Fisher, C.L., Sibley, C.R., Foster, B.M., Bartke, T., *et al.* (2016). The interaction of PRC2 with RNA or chromatin is mutually antagonistic. *Genome Res* 26, 896-907.
- Berry, S., and Dean, C. (2015). Environmental perception and epigenetic memory: mechanistic insight through FLC. *Plant J* 83, 133-148.
- Boehmer, P.E. (2004). RNA binding and R-loop formation by the herpes simplex virus type-1 single-stranded DNA-binding protein (ICP8). *Nucleic Acids Res* 32, 4576-4584.
- Bourgey, M., Dali, R., Eveleigh, R., Chen, K.C., Letourneau, L., Fillon, J., Michaud, M., Caron, M., Sandoval, J., Lefebvre, F., *et al.* (2019). GenPipes: an open-source framework for distributed and scalable genomic analyses. *Gigascience* 8.
- Calderwood, S.K. (2016). A critical role for topoisomerase IIb and DNA double strand breaks in transcription. *Transcription* 7, 75-83.
- Campbell, S., Ismail, I.H., Young, L.C., Poirier, G.G., and Hendzel, M.J. (2013). Polycomb repressive complex 2 contributes to DNA double-strand break repair. *Cell Cycle* 12, 2675-2683.
- Chen, P.B., Chen, H.V., Acharya, D., Rando, O.J., and Fazzio, T.G. (2015). R loops regulate promoter-proximal chromatin architecture and cellular differentiation. *Nat Struct Mol Biol* 22, 999-1007.

Chou, D.M., Adamson, B., Dephoure, N.E., Tan, X., Nottke, A.C., Hurov, K.E., Gygi, S.P., Colaiacovo, M.P., and Elledge, S.J. (2010). A chromatin localization screen reveals poly (ADP ribose)-regulated recruitment of the repressive polycomb and NuRD complexes to sites of DNA damage. *Proc Natl Acad Sci U S A* *107*, 18475-18480.

Crossley, M.P., Bocek, M., and Cimprich, K.A. (2019). R-Loops as Cellular Regulators and Genomic Threats. *Mol Cell* *73*, 398-411.

Davidovich, C., and Cech, T.R. (2015). The recruitment of chromatin modifiers by long noncoding RNAs: lessons from PRC2. *RNA* *21*, 2007-2022.

Davidovich, C., Wang, X., Cifuentes-Rojas, C., Goodrich, K.J., Gooding, A.R., Lee, J.T., and Cech, T.R. (2015). Toward a consensus on the binding specificity and promiscuity of PRC2 for RNA. *Mol Cell* *57*, 552-558.

Davidovich, C., Zheng, L., Goodrich, K.J., and Cech, T.R. (2013). Promiscuous RNA binding by Polycomb repressive complex 2. *Nat Struct Mol Biol* *20*, 1250-1257.

Decoville, M., Giacomello, E., Leng, M., and Locker, D. (2001). DSP1, an HMG-like protein, is involved in the regulation of homeotic genes. *Genetics* *157*, 237-244.

Dejardin, J., Rappailles, A., Cuvier, O., Grimaud, C., Decoville, M., Locker, D., and Cavalli, G. (2005). Recruitment of Drosophila Polycomb group proteins to chromatin by DSP1. *Nature* *434*, 533-538.

Deveryshetty, J., Peterlini, T., Ryzhikov, M., Brahiti, N., Dellaire, G., Masson, J.Y., and Korolev, S. (2019). Novel RNA and DNA strand exchange activity of the PALB2 DNA binding domain and its critical role for DNA repair in cells. *Elife* *8*.

Ejsmont, R.K., Sarov, M., Winkler, S., Lipinski, K.A., and Tomancak, P. (2009). A toolkit for high-throughput, cross-species gene engineering in Drosophila. *Nat Methods* *6*, 435-437.

Erceg, J., Pakozdi, T., Marco-Ferreres, R., Ghavi-Helm, Y., Girardot, C., Bracken, A.P., and Furlong, E.E. (2017). Dual functionality of cis-regulatory elements as developmental enhancers and Polycomb response elements. *Genes Dev* *31*, 590-602.

Erokhin, M., Elizar'ev, P., Parshikov, A., Schedl, P., Georgiev, P., and Chetverina, D. (2015). Transcriptional read-through is not sufficient to induce an epigenetic switch in the silencing activity of Polycomb response elements. *Proc Natl Acad Sci U S A* *112*, 14930-14935.

Francis, N.J., Kingston, R.E., and Woodcock, C.L. (2004). Chromatin compaction by a polycomb group protein complex. *Science* 306, 1574-1577.

Francis, N.J., Saurin, A.J., Shao, Z., and Kingston, R.E. (2001). Reconstitution of a functional core polycomb repressive complex. *Mol Cell* 8, 545-556.

Garcia-Muse, T., and Aguilera, A. (2019). R Loops: From Physiological to Pathological Roles. *Cell* 179, 604-618.

Gel, B., Diez-Villanueva, A., Serra, E., Buschbeck, M., Peinado, M.A., and Malinverni, R. (2016). regioneR: an R/Bioconductor package for the association analysis of genomic regions based on permutation tests. *Bioinformatics* 32, 289-291.

Ginno, P.A., Lott, P.L., Christensen, H.C., Korf, I., and Chedin, F. (2012). R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. *Mol Cell* 45, 814-825.

Grossniklaus, U., and Paro, R. (2014). Transcriptional silencing by polycomb-group proteins. *Cold Spring Harb Perspect Biol* 6, a019331.

Hartono, S.R., Malapert, A., Legros, P., Bernard, P., Chedin, F., and Vanoosthuyse, V. (2018). The Affinity of the S9.6 Antibody for Double-Stranded RNAs Impacts the Accurate Mapping of R-Loops in Fission Yeast. *J Mol Biol* 430, 272-284.

Hekimoglu, B., and Ringrose, L. (2009). Non-coding RNAs in polycomb/trithorax regulation. *RNA Biol* 6, 129-137.

Herzog, V.A., Lempradl, A., Trupke, J., Okulski, H., Altmutter, C., Ruge, F., Boidol, B., Kubicek, S., Schmauss, G., Aumayr, K., *et al.* (2014). A strand-specific switch in noncoding transcription switches the function of a Polycomb/Trithorax response element. *Nat Genet* 46, 973-981.

Huang, W., Loganantharaj, R., Schroeder, B., Fargo, D., and Li, L. (2013). PAVIS: a tool for Peak Annotation and Visualization. *Bioinformatics* 29, 3097-3099.

Kaneko, S., Chu, C., Shatkin, A.J., and Manley, J.L. (2007). Human capping enzyme promotes formation of transcriptional R loops in vitro. *Proc Natl Acad Sci U S A* 104, 17620-17625.

Kang, H., McElroy, K.A., Jung, Y.L., Alekseyenko, A.A., Zee, B.M., Park, P.J., and Kuroda, M.I. (2015). Sex comb on midleg (Scm) is a functional link between PcG-repressive complexes in *Drosophila*. *Genes Dev* 29, 1136-1150.

Kasahara, M., Clikeman, J.A., Bates, D.B., and Kogoma, T. (2000). RecA protein-dependent R-loop formation in vitro. *Genes Dev* 14, 360-365.

Kassis, J.A., and Brown, J.L. (2013). Polycomb group response elements in Drosophila and vertebrates. *Adv Genet* 81, 83-118.

Kassis, J.A., Kennison, J.A., and Tamkun, J.W. (2017). Polycomb and Trithorax Group Genes in Drosophila. *Genetics* 206, 1699-1725.

Kassis, J.A., and Muller, J. (2015). Transcription through Polycomb response elements does not induce a switch from repression to activation. *Proc Natl Acad Sci U S A* 112, 14755-14756.

Lamiable, O., Rabhi, M., Peronnet, F., Locker, D., and Decoville, M. (2010). Rm62, a DEAD-box RNA helicase, complexes with DSP1 in Drosophila embryos. *Genesis* 48, 244-253.

Langlais, K.K., Brown, J.L., and Kassis, J.A. (2012). Polycomb group proteins bind an engrailed PRE in both the "ON" and "OFF" transcriptional states of engrailed. *PLoS One* 7, e48765.

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

Lee, Y.W., Arora, R., Wischnewski, H., and Azzalin, C.M. (2018). TRF1 participates in chromosome end protection by averting TRF2-dependent telomeric R loops. *Nat Struct Mol Biol* 25, 147-153.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078-2079.

Long, Y., Bolanos, B., Gong, L., Liu, W., Goodrich, K.J., Yang, X., Chen, S., Gooding, A.R., Maegley, K.A., Gajiwala, K.S., *et al.* (2017). Conserved RNA-binding specificity of polycomb repressive complex 2 is achieved by dispersed amino acid patches in EZH2. *Elife* 6.

Loomis, E.W., Sanz, L.A., Chedin, F., and Hagerman, P.J. (2014). Transcription-associated R-loop formation across the human FMR1 CGG-repeat region. *PLoS Genet* 10, e1004294.

Lupo, R., Breiling, A., Bianchi, M.E., and Orlando, V. (2001). Drosophila chromosome condensation proteins Topoisomerase II and Barren colocalize with Polycomb and maintain Fab-7 PRE silencing. *Mol Cell* 7, 127-136.

Maeda, R.K., and Karch, F. (2009). The bithorax complex of *Drosophila* an exceptional Hox cluster. *Curr Top Dev Biol* 88, 1-33.

Mazina, O.M., Keskin, H., Hanamshet, K., Storic, F., and Mazin, A.V. (2017). Rad52 Inverse Strand Exchange Drives RNA-Templated DNA Double-Strand Break Repair. *Mol Cell* 67, 19-29 e13.

McDevitt, S., Rusanov, T., Kent, T., Chandramouly, G., and Pomerantz, R.T. (2018). How RNA transcripts coordinate DNA recombination and repair. *Nat Commun* 9, 1091.

Mersaoui, S.Y., Yu, Z., Coulombe, Y., Karam, M., Busatto, F.F., Masson, J.Y., and Richard, S. (2019). Arginine methylation of the DDX5 helicase RGG/RG motif by PRMT5 regulates resolution of RNA:DNA hybrids. *EMBO J* 38, e100986.

Muller, J., Hart, C.M., Francis, N.J., Vargas, M.L., Sengupta, A., Wild, B., Miller, E.L., O'Connor, M.B., Kingston, R.E., and Simon, J.A. (2002). Histone methyltransferase activity of a *Drosophila* Polycomb group repressor complex. *Cell* 111, 197-208.

Neph, S., Kuehn, M.S., Reynolds, A.P., Haugen, E., Thurman, R.E., Johnson, A.K., Rynes, E., Maurano, M.T., Vierstra, J., Thomas, S., *et al.* (2012). BEDOPS: high-performance genomic feature operations. *Bioinformatics* 28, 1919-1920.

Ogiyama, Y., Schuettengruber, B., Papadopoulos, G.L., Chang, J.M., and Cavalli, G. (2018). Polycomb-Dependent Chromatin Looping Contributes to Gene Silencing during *Drosophila* Development. *Mol Cell* 71, 73-88 e75.

Ohle, C., Tesorero, R., Schermann, G., Dobrev, N., Sinning, I., and Fischer, T. (2016). Transient RNA-DNA Hybrids Are Required for Efficient Double-Strand Break Repair. *Cell* 167, 1001-1013 e1007.

Papp, B., and Muller, J. (2006). Histone trimethylation and the maintenance of transcriptional ON and OFF states by trxG and PcG proteins. *Genes Dev* 20, 2041-2054.

Poux, S., Kostic, C., and Pirrotta, V. (1996). Hunchback-independent silencing of late Ubx enhancers by a Polycomb Group Response Element. *EMBO J* 15, 4713-4722.

Prezioso, C., and Orlando, V. (2011). Polycomb proteins in mammalian cell differentiation and plasticity. *FEBS Lett* 585, 2067-2077.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841-842.

Ramirez, F., Dundar, F., Diehl, S., Gruning, B.A., and Manke, T. (2014). deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res* 42, W187-191.

Rank, G., Prestel, M., and Paro, R. (2002). Transcription through intergenic chromosomal memory elements of the *Drosophila* bithorax complex correlates with an epigenetic switch. *Mol Cell Biol* 22, 8026-8034.

Reijns, M.A., Bubeck, D., Gibson, L.C., Graham, S.C., Baillie, G.S., Jones, E.Y., and Jackson, A.P. (2011). The structure of the human RNase H2 complex defines key interaction interfaces relevant to enzyme function and human disease. *J Biol Chem* 286, 10530-10539.

Ringrose, L. (2017). Noncoding RNAs in Polycomb and Trithorax Regulation: A Quantitative Perspective. *Annu Rev Genet* 51, 385-411.

Ringrose, L., Rehmsmeier, M., Dura, J.M., and Paro, R. (2003). Genome-wide prediction of Polycomb/Trithorax response elements in *Drosophila melanogaster*. *Dev Cell* 5, 759-771.

Santos-Pereira, J.M., and Aguilera, A. (2015). R loops: new modulators of genome dynamics and function. *Nat Rev Genet* 16, 583-597.

Sanz, L.A., Hartono, S.R., Lim, Y.W., Steyaert, S., Rajpurkar, A., Ginno, P.A., Xu, X., and Chedin, F. (2016). Prevalent, Dynamic, and Conserved R-Loop Structures Associate with Specific Epigenomic Signatures in Mammals. *Mol Cell* 63, 167-178.

Schmitt, S., Prestel, M., and Paro, R. (2005). Intergenic transcription through a polycomb group response element counteracts silencing. *Genes Dev* 19, 697-708.

Schuettengruber, B., Oded Elkayam, N., Sexton, T., Entrevan, M., Stern, S., Thomas, A., Yaffe, E., Parrinello, H., Tanay, A., and Cavalli, G. (2014). Cooperativity, specificity, and evolutionary stability of Polycomb targeting in *Drosophila*. *Cell Rep* 9, 219-233.

Simon, J. (1995). Locking in stable states of gene expression: transcriptional control during *Drosophila* development. *Curr Opin Cell Biol* 7, 376-385.

Skourti-Stathaki, K., Torlai Triglia, E., Warburton, M., Voigt, P., Bird, A., and Pombo, A. (2019). R-Loops Enhance Polycomb Repression at a Subset of Developmental Regulator Genes. *Mol Cell* 73, 930-945 e934.

Sneppen, K., and Ringrose, L. (2019). Theoretical analysis of Polycomb-Trithorax systems predicts that poised chromatin is bistable and not bivalent. *Nat Commun* 10, 2133.

Steffen, P.A., and Ringrose, L. (2014). What are memories made of? How Polycomb and Trithorax proteins mediate epigenetic memory. *Nat Rev Mol Cell Biol* 15, 340-356.

Thresher, R., and Griffith, J. (1992). Electron microscopic visualization of DNA and DNA-protein complexes as adjunct to biochemical studies. *Methods Enzymol* 211, 481-490.

Umer, Z., Akhtar, J., Khan, M.H.F., Shaheen, N., Haseeb, M.A., Mazhar, K., Mithani, A., Anwar, S., and Tariq, M. (2019). Genome-wide RNAi screen in *Drosophila* reveals Enok as a novel trithorax group regulator. *Epigenetics Chromatin* 12, 55.

Wang, X., Goodrich, K.J., Gooding, A.R., Naeem, H., Archer, S., Paucek, R.D., Youmans, D.T., Cech, T.R., and Davidovich, C. (2017a). Targeting of Polycomb Repressive Complex 2 to RNA by Short Repeats of Consecutive Guanines. *Mol Cell* 65, 1056-1067 e1055.

Wang, X., Paucek, R.D., Gooding, A.R., Brown, Z.Z., Ge, E.J., Muir, T.W., and Cech, T.R. (2017b). Molecular analysis of PRC2 recruitment to DNA in chromatin and its inhibition by RNA. *Nat Struct Mol Biol* 24, 1028-1038.

Wong, I., and Lohman, T.M. (1993). A double-filter method for nitrocellulose-filter binding: application to protein-nucleic acid interactions. *Proc Natl Acad Sci U S A* 90, 5428-5432.

Wong, M.M., Belew, M.D., Kwieraga, A., Nhan, J.D., and Michael, W.M. (2018). Programmed DNA Breaks Activate the Germline Genome in *Caenorhabditis elegans*. *Dev Cell* 46, 302-315 e305.

Xiao, J., Jin, R., Yu, X., Shen, M., Wagner, J.D., Pai, A., Song, C., Zhuang, M., Klasfeld, S., He, C., *et al.* (2017). Cis and trans determinants of epigenetic silencing by Polycomb repressive complex 2 in *Arabidopsis*. *Nat Genet* 49, 1546-1552.

Xu, W., Xu, H., Li, K., Fan, Y., Liu, Y., Yang, X., and Sun, Q. (2017). The R-loop is a common chromatin feature of the *Arabidopsis* genome. *Nat Plants* 3, 704-714.

Zaitsev, E.N., and Kowalczykowski, S.C. (2000). A novel pairing process promoted by *Escherichia coli* RecA protein: inverse DNA and RNA strand exchange. *Genes Dev* 14, 740-749.

Zhang, Q., McKenzie, N.J., Warneford-Thomson, R., Gail, E.H., Flanigan, S.F., Owen, B.M., Lauman, R., Levina, V., Garcia, B.A., Schittenhelm, R.B., *et al.* (2019). RNA exploits an exposed regulatory site to inhibit the enzymatic activity of PRC2. *Nat Struct Mol Biol* 26, 237-247.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., *et al.* (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biol* 9, R137.

3 Article 2: A Topoisomerase I like activity associated with the Polycomb Group protein PSC

A Topoisomerase I like activity associated with the Polycomb Group protein PSC

Authors:

Célia Alecki^{1,2}, Victoria Chiwara¹, and Nicole J. Francis^{1,2,3*}

Affiliations:

¹Institut de recherches cliniques de Montréal, 110 Avenue des Pins Ouest, Montréal, QC, H2W 1R7, Canada

²Département de biochimie et médecine moléculaire Université de Montréal, 2900 Boulevard Edouard-Montpetit, Montréal, QC H3T 1J4, Canada.

³Division of Experimental Medicine, McGill University, 1001 Decarie Boulevard, Montreal, QC H4A 3J1, Canada

*correspondence to: nicole.francis@ircm.qc.ca

Author contributions: Conceptualization: N.J.F. & C.A.; Investigation: C.A.; Resources: V.C; Writing: C.A. & N.J.F., Supervision: N.J.F., Funding acquisition: N.J.F.

3.1 Abstract

The Polycomb complex PRC1 induces chromatin compaction and inhibits chromatin remodeling in order to induce a stable repressive chromatin state and inhibit gene expression. PRC1 chromatin compaction activity requires the binding and bridging of nucleosomes, but PRC1 has also a strong affinity for naked DNA. We noticed that PRC1 preparations induce a change in DNA topology when incubated with supercoiled plasmid DNA and we characterized this activity. PRC1 preparations have topoisomerase I-like activity. This activity relaxes a negatively supercoiled plasmid and adds negative supercoils to a relaxed plasmid without requiring the addition of topoisomerase, ATP or Mg^{2+} . At least one PRC1 per plasmid is required to change topology. The large, intrinsically disordered C-terminal region of the PRC1 subunit PSC (PSC-CTR), also has associated topoisomerase-like activity, but requires ~ 100 copies per plasmid. We were unable to biochemically separate PRC1 or the PSC-CTR from topoisomerase activity, and we did not detect topoisomerase I in active PRC1 preparations by mass spectrometry. Topoisomerases use a catalytic tyrosine to form a covalent protein-DNA intermediate during cleavage and re-ligation. However, mutation of all tyrosines in a minimal active fragment of the PSC-CTR did not change the associated topoisomerase activity. We conclude that PRC1 and its PSC-CTR has an associated topoisomerase activity that is either tightly associated and highly active, or intrinsic to the PSC-CTR. Both PRC1 and the PSC-CTR can relax supercoils on chromatinized plasmids and can disrupt R-loops formed on supercoiled DNA. Our results identify control of DNA topology as a potential novel mechanism used by PRC1 to modify chromatin organization. Topoisomerase activity may also be involved in the regulation of non-canonical DNA structures such as R-loops, which are stabilized by supercoiling and accumulate at Polycomb binding sites.

3.2 Main text

3.2.1 Introduction

The Polycomb group complex PRC1 represses gene expression and is essential during development to ensure the appropriate gene expression pattern of an organism (Ringrose and Paro, 2004). PRC1 compacts chromatin by contacting multiple nucleosomes and bridging them together (Francis et al., 2004; Lo et al., 2012c), and inhibits chromatin remodeling and transcription to ensure stable gene repression (King et al., 2002; Lehmann et al., 2012; Lo and Francis, 2010). PRC1 binds DNA and chromatin in a non-sequence specific manner but its targeting is not random. It is recruited to specific DNA elements which are nucleosome-free (Kahn et al., 2006; Mohd-Sarip et al., 2005; Papp and Müller, 2006). PRC1 does not act only on chromatin but also on DNA. PRC1 has been shown to bind tightly DNA (Francis et al., 2001), to bridge DNA fragments (Lo et al., 2012c), to remain associated with DNA through DNA replication (Francis et al., 2009) and to inhibit transcription without having to interact with nucleosomes (King et al., 2002).

In cells, DNA is not relaxed but supercoiled. The level of supercoiling or linking number is defined by the twist and writhe of the DNA molecule. In order to change the linking number, one or both DNA strands must be cleaved, and the strands allowed to rotate around one another before being religated (Brahmachari and Marko, 2018; Mirkin, 2001). Multiple events occurring in cells can locally increase supercoiling. These include both transcription and DNA replication, during which the unwinding of the DNA double helix is opened it leads to local compensatory increases in supercoiling before and after the transcription or DNA replication bubble (Wang, 2002; Wu et al., 1988). The torsional stress created by these DNA distortions can halt polymerase passage, and result in DNA breaks. To allow polymerase passage, and to avoid DNA breaks and potential harmful recombination between chromosomes, and more generally genomic instability, cells have developed proteins that can modify DNA topology: topoisomerases.

Topoisomerases are conserved from prokaryotes to eukaryotes and are the main proteins involved in the regulation of supercoiling in cells. Type I and type II topoisomerases differ by the number of DNA strands they cut and their utilization of ATP, and target different cellular processes (Pommier et al., 2016). Topoisomerases are highly expressed in cells, and involved in all cellular DNA transactions, including transcription, DNA replication, and chromosome compaction. The regulation of DNA topology local and large scale chromatin organization, and its interplay with DNA and chromatin modifying proteins are not fully understood.

We observed that preparations of PRC1 induce a striking change in DNA plasmid topology. Here, we characterize this activity. We find that PRC1 and the C-terminal region of its PSC subunit (PSC-CTR) can relax a supercoiled plasmid and add supercoils to a relaxed plasmid. We were unable to biochemically separate the topoisomerase activity from PRC1 or the PSC-CTR, and we did not identify topoisomerases in active PRC1 preparations by mass spectrometry. Biochemically, this activity mirrors that of eukaryotic topoisomerase I, including being independent of ATP and magnesium but stimulated by KCl. However, mutation of all tyrosines in a minimal active fragment of the PSC-CTR did not reduce topoisomerase activity. PRC1 and the PSC-CTR can relax unconstrained supercoils on nucleosome-containing plasmids, and can disrupt R-loops formed on supercoiled templates.

3.2.2 Results

Changes in DNA topology introduced by incubation of plasmid DNA with the Polycomb complex PRC1

We previously noticed unexplained changes in DNA plasmid topology after incubation with PRC1. To directly test if PRC1 modifies plasmid topology, we incubated PRC1 with negatively supercoiled plasmid DNA (pUC19, 2900 bp) in the absence of any added topoisomerase. We removed PRC1 by digestion with proteinase K, and analyzed the protein-free plasmids on agarose gels in the presence or absence of chloroquine to resolve topoisomers (Fig. 3-1a). Titration of PRC1 into reactions with supercoiled DNA results in near complete plasmid relaxation (Fig. 3-1c left). To test whether PRC1 can do the opposite reaction—add supercoils to a relaxed plasmid—we treated supercoiled DNA plasmids with *E. coli* Topoisomerase I and purified the resulting relaxed plasmid (Fig. 3-1b). When PRC1 is titrated with relaxed DNA, a small number of supercoils are introduced (Fig. 3-1c right). The topoisomer distribution of pUC19 plasmids incubated with PRC1 is similar, comprised of 4 prominent bands, irrespective of whether a supercoiled or relaxed plasmid is used as the substrate. For both reactions, the reaction is complete with 0.62 nM PRC1 and 0.6 nM plasmid.

The PSC-CTR, a subunit of PRC1, encodes many PRC1 activities. This large, intrinsically disordered region is sufficient to compact chromatin (Francis et al., 2004) and to bridge naked DNA fragments (Lo et al., 2012c). We wondered whether the PSC-CTR can also modify DNA topology. In titrations of the PSC-CTR with supercoiled or relaxed DNA, we observe near complete relaxation of supercoiled plasmids, and addition of a small number of supercoils to relaxed plasmids (Fig. 3-1d). The effect of PRC1 and the PSC-

CTR on plasmid topology is similar, but PRC1 is ~100 times more efficient than the PSC-CTR, which requires 62.5 nM protein to complete reactions with 0.6 nM plasmid.

To determine whether the supercoils in the final state induced by PRC1 and the PSC-CTR are positive or negative, we performed the same topology assays and ran the sample on two-dimensional gels (Fig. 3-1e-f). The final state starting from either a supercoiled (left panels in Fig. 3-1e, f) or relaxed (right panels in Fig. 3-1e, f) plasmid contains a small number of negative supercoils. Thus, PRC1 and the PSC-CTR are able to partially relax a negatively supercoiled plasmid and add a small number of negative supercoils onto a relaxed plasmid, thereby maintaining the DNA in a slightly negatively supercoiled state.

In order to change the linking number of a plasmid, at least one DNA strand must be cleaved (Mirkin, 2001). The main enzymes that are able to carry out this activity are the topoisomerases (Pommier et al., 2016). Topoisomerases change linking number in four steps: DNA cleavage, protein-DNA covalent complex formation, addition or loss of supercoils by strand rotation or strand passage, and DNA religation. Proteins like the histone octamer introduce toroidal supercoils by wrapping the DNA around them (Simpson et al., 1985). However, in the absence of a topoisomerase (i.e. a strand cleavage and ligation reaction) these protein-induced DNA distortions do not change the linking number because compensatory distortions (changes in twist or plectonemal supercoils) occur. When the histone proteins are removed, the level of supercoiling returns to what it was before nucleosome assembly. If a topoisomerase is added to the nucleosomal plasmids, the compensatory plectonemes are removed, so that after removal of the histone proteins, a change in linking number (addition of one negative supercoil per nucleosome) is observed.

There are four possibilities for how PRC1 and the PSC-CTR induce changes in DNA topology: 1) the protein preparations contain a topoisomerase contaminant; 2) PRC1 and the PSC-CTR encode topoisomerase-like activity; 3) PRC1 and the PSC-CTR may introduce toroidal supercoils by wrapping the DNA around them as shown before for PRC1 (Mohd-Sarip et al., 2006), and a contaminant is responsible for cutting and ligating the DNA, resulting in the observed change in linking number; 4) PRC1 introduces supercoils by wrapping the DNA around itself *and* contains a topoisomerase-like activity, which results in the change in linking number. The combination of protein-constrained supercoils and topoisomerase-like activity can explain how the final reaction product contains a small number of supercoils (Fig. 3-2).

At least one PRC1 per plasmid is required to change DNA topology

If the PRC1-induced topology change involves wrapping the DNA around the protein complex, it should require at least one complex per plasmid, which is consistent with our initial characterization (Fig. 3-1). To analyze this more carefully, we tested PRC1:DNA ratios of 0.5:1, 2:1, and 8:1 over an extended time course (Fig. 3-3a). At 8:1, essentially all of the plasmid is converted to its final form within 30 minutes. At 2:1, between 60 and 120 minutes are required to convert essentially all of the plasmid. At 0.5:1, only a fraction of the plasmid is converted, even after 5 hours. The time course also demonstrates that changes in topology are observed within the first minutes of the reaction.

To confirm that we need at least one molecule of PRC1 per plasmid to change its topology, we titrated the DNA. In 30 min., 8 molecules of PRC1 per plasmid are needed to observe relaxation of all of the plasmids (Figure 3-3b). Conversely, with a ratio of 8 plasmids per PRC1 only a small number of plasmids is relaxed (Figure 3-3b). These data suggest the activity of PRC1 is not limited by binding affinity in these experiments, but instead by the number of complexes available. This argues in favor of the wrapping (model 3) or wrapping-topoisomerase (model 4) models.

Topoisomerase-like activity co-fractionates with the PSC-CTR through additional purification

To test whether the topoisomerase-like activity is the result of a protein contaminant, we tested whether we could biochemically separate the PSC-CTR from the activity. PRC1 is already isolated from nuclear extracts using a stringent affinity purification protocol that includes washing with 2 M KCl, and 1 M Urea (Alecki et al., 2020). To carry out a more stringent purification with the PSC-CTR, we introduced a hexahistidine tag in addition to our standard Flag tag. We carried out a two step purification in which the PSC-CTR was first isolated by anti-Flag affinity, and then further purified on Ni-NTA beads. We washed the Ni-NTA bound PSC-CTR with up to 6 M urea to remove (and denature) even tightly associated contaminants. The specific activity of the PSC-CTR purified with this stringent protocol (Fig. 3-4a) is unchanged, with full plasmid modification (starting from either a supercoiled or relaxed substrate) requiring 62.5nM protein (Fig. 3-4b). As a second means to fractionate the PSC-CTR, we used glycerol gradient sedimentation. The topoisomerase-like activity (on both supercoiled and relaxed plasmids) co-fractionates with the PSC-CTR through the gradient (Figure 3-4c, d). Finally, we analyzed an extensive set of mass spectrometry data generated with PRC1 lacking its Ph subunit or the PSC-CTR (15 samples from a protein footprinting experiment, (Kang et al., 2020) for the presence of Sf9 cell topoisomerases, and find no evidence for them (Fig. 3-4e). Thus, we are unable to biochemically separate the PSC-CTR from

topoisomerase-like activity, and we do not find mass spectrometry evidence for topoisomerase contamination in our protein preparations.

PRC1 associated topoisomerase activity is type I

The three experiments presented above argue against contaminating topoisomerase accounting for the activity of PRC1 and the PSC-CTR. To understand if the activity observed is consistent with a topoisomerase, we further characterized it biochemically. Type I and type II topoisomerases use distinct reaction mechanisms and have different salt, magnesium and ATP requirements. We tested whether the activity associated with the PSC-CTR is a type I or type II activity. All of our topology assays are performed in absence of ATP, arguing against a type II topoisomerase activity. The PSC-CTR relaxes negatively supercoiled plasmid (Fig. 3-5a left) and adds negative supercoils to a relaxed plasmid (Fig. 3-5a right) in both the presence and absence of magnesium. This magnesium independence is shared with eukaryotic Topoisomerase I (Chen and Wang, 1998), consistent with a type I activity. Our topology assays are carried out in 180 mM KCl, similar to the 150 mM KCl previously described as optimal for human Topoisomerase I (Christensen et al., 2003). We titrated KCl down in experiments with the PSC-CTR and supercoiled or relaxed plasmid substrates, and observe that activity decreases with decreasing KCl, similar to human Topoisomerase I (Fig. 3-5b).

Finally, we tested the sensitivity of the PSC-CTR activity to type I and type II Topoisomerase inhibitors (Fig. 3-6a-b). The PSC-CTR activity is not affected by etoposide, a type II inhibitor. This, in conjunction with the lack of ATP-dependence, rules out a type II activity (either intrinsic the PSC-CTR, or contaminating the protein preparations). CPT inhibits Topoisomerase I by capturing the covalent complex formed between it and the DNA, preventing the re-ligation (Staker et al., 2002b). The activity of the PSC-CTR is reduced by Camptothecin (CPT), a type I Topoisomerase inhibitor, although full inhibition is not observed, even at 200 μ M.

In summary, the PSC-CTR changes DNA topology in the absence of ATP and Mg^{2+} , is active at higher concentrations of KCl, and is inhibited by CPT. We conclude that the PSC-CTR associates with a type I topoisomerase-like activity, which may be intrinsic to the PSC-CTR or a tightly bound, highly active contaminant.

Identification of minimal active PSC-CTR sequences

In order to identify the minimal fragment of PSC responsible for the change in topology, we tested a series of truncated versions of the PSC-CTR (Fig. 3-7). The activities of the PSC-CTR (PSC⁴⁵⁶⁻¹⁶⁰³) and PSC⁴⁵⁶⁻¹³³⁵ are similar, with the relaxation of all supercoiled plasmids requiring ~100 molecules per plasmid. Two shorter truncations, PSC⁴⁵⁶⁻⁹⁰⁹ PSC⁶⁵⁶⁻⁹⁰⁹ are active at ~800 molecules per plasmid. Finally, PSC⁴⁵⁶⁻⁷⁵⁹ was not able to completely convert either supercoiled or relaxed templates, even at ratios of 6670 molecules/plasmid. These data point to an important function for residues between 759 and 909, but clearly they are not sufficient to explain the activity of the PSC-CTR. The increasing amounts of protein required for progressively shorter truncations (ranging from 120 kDa for the PSC-CTR down to 27 kDa) could indicate that an oligomer of a certain size is required to wrap DNA. PSC⁴⁵⁶⁻⁹⁰⁹ also binds DNA 10-fold weaker than the full PSC-CTR (Lo et al., mol cell), so that decreased DNA affinity may contribute to decreased activity. We recently identified two candidate DNA binding regions in the PSC-CTR, one from aa631-867, and the second from aa1225-1342 (Kang et al., 2020). The minimal fully active fragment (PSC⁴⁵⁶⁻¹³³⁵) contains both of these regions, while the less active truncations contain only the first region, and PSC⁴⁵⁶⁻⁷⁵⁹ has only part of the first region. A combination of protein footprinting and cross-linking mass spectrometry are consistent with the two DNA binding regions coming together, which may facilitate DNA wrapping.

The mutation of all tyrosine residues in an active PSC-CTR fragment does not affect its activity

Topoisomerases use a catalytic tyrosine to form the covalent intermediate with DNA. The Topoisomerase I catalytic site is composed of conserved RKRH residues and a tyrosine as the catalytic residue. The mutation of this tyrosine to phenylalanine abolishes topoisomerase activity (Baker et al., 2009; Eng et al., 1989). To test whether such an activity could be encoded in the PSC-CTR, we made Y-->F mutations of the first three tyrosines in PSC⁴⁵⁶⁻¹³³⁵ and tested the activity of the mutant protein. We observe that PSC^{456-1335(3YF)} activity on DNA topology is not affected (Figure 8a-c). We therefore mutated all 15 tyrosines of PSC⁴⁵⁶⁻¹³³⁵ to phenylalanine and observe that PSC^{456-1335YF} is also fully active (Fig. 3-8d). Thus, if the PSC-CTR encodes a topoisomerase activity, it does not depend on a catalytic tyrosine. Some site-specific DNA recombinases use serine in place of tyrosine to form covalent complexes with DNA as part of their reaction mechanism (Stark, 2014). It is possible that the PSC-CTR uses serine as a catalytic residue, which will require further investigation.

PRC1 and the PSC-CTR can modify the topology of nucleosomal plasmids

The primary described functions of PRC1 are on chromatin. PRC1 can compact the chromatin by bridging nucleosomes (Francis et al., 2004; Lo et al., 2012c) and can inhibit nucleosome remodeling (Shao et al., 1999). We therefore tested whether PRC1 can affect the topology of plasmids assembled into chromatin. We assembled the pUC19 plasmid into chromatin with various histone/DNA ratios (Fig. 3-9a) and carried out topology assays with increasing amounts of PRC1 on these templates. We observe that PRC1 relaxes chromatin similarly to *E. coli* Topoisomerase I for the lower density of nucleosomes and more effectively than Topoisomerase I when more nucleosomes are present (most evident on the agarose gel containing 4 µg/mL of chloroquine) (Fig. 3-9c). The PSC-CTR is also able to relax chromatinized plasmids (Figure 3-9d). To ensure that PRC1 has not removed nucleosomes we performed a Micrococcal nuclease assay on the chromatinized plasmid DNA in absence or in presence of PRC1. We observe protection of ~150 bp DNA fragments, consistent with nucleosome assembly (Fig. 3-9b). PRC1 and PSC-CTR can relax chromatin assembled plasmid but can't remove the nucleosome-constrained supercoils.

Relaxation of supercoiled DNA by PRC1 correlates with R-loop loss

In vivo, topoisomerases resolve non canonical DNA structures that involve DNA unwinding by relaxing the DNA around them (Chedin and Benham, 2020). In *Drosophila melanogaster*, we showed that many of the sites bound by PRC1 form R-loops (Alecki et al., 2020). We therefore tested whether the activity associated with PRC1 can resolve R-loops. R-loops were prepared by in vitro transcription of supercoiled plasmids in the presence of radiolabeled nucleotides (Fig. 3-10a). R-loops are visualized as bands containing radiolabelled RNA that comigrates with the DNA template on agarose gels. Their identity is confirmed by their sensitivity to RNaseH and DNaseI, and resistance to RNaseA (Fig. 3-10a). When these transcribed templates are incubated with PRC1 we observe loss of R-loops from supercoiled plasmids, along with plasmid relaxation (Fig. 3-10b). The level of R-loops associated with relaxed/nicked plasmids remains unchanged (Fig. 3-10b). We conclude that PRC1 can disrupt R-loops on supercoiled templates, likely by relaxing the supercoils.

3.2.3 Discussion

We have found that PRC1 and the PSC-CTR can change DNA topology, and we have characterized this activity. PRC1 and the PSC-CTR can relax a negatively supercoiled plasmid and add a small number of supercoils to a relaxed plasmid (Fig. 3-10c), in the absence of ATP, Mg^{2+} , or added topoisomerase. They can relax supercoils between nucleosomes without removing them (Fig. 3-10d) and resolve R-loops formed on supercoiled plasmids (Fig. 3-10e).

Mechanisms behind the change in plasmid topology by PRC1 and the PSC-CTR

PRC1 wraps DNA around itself

We have not been able to identify fully the mechanism that PRC1 and the PSC-CTR use to modify plasmid topology. The addition of supercoils on a relaxed plasmid is more likely to be the result a Polycomb group proteins wrapping DNA around them (Fig. 3-9f), because PRC1 was already shown to be able to wrap DNA (Mohd-Sarip et al., 2006) and no eukaryotic type I topoisomerase can introduce supercoils. The addition of supercoils is an activity only found in bacterial gyrase or in site specific recombinases but this last group of proteins require specific DNA sequences to perform recombination between them (Kampranis et al., 1999; Yang, 2010). It is likely that the observed introduction of a small number of negative supercoils on a relaxed plasmid is the result of PRC1 or the PSC-CTR wrapping DNA around themselves, combined with another activity which is responsible for cutting and ligating the DNA (Fig. 3-9f).

Is topoisomerase-like activity intrinsic to PSC or the result of a contaminant protein?

The characterization of the activity involved in the relaxation and addition of supercoils by cutting and ligating the DNA, has the characteristics of a type I topoisomerase activity (Fig. 5 and 6). This topoisomerase-like activity can be the result of a contaminant protein or it can be intrinsic to PSC. Although the results presented figure 4 show that we have not been able to biochemically separate the topoisomerase activity from the PSC-CTR, and that we do not find evidence for contaminating topoisomerases, the sensitivity to the topoisomerase I inhibitor CPT may suggest the opposite. However, CPT does not fully inhibit the activity of the PSC-CTR, even at 200 μ M, although 1 μ M is sufficient to inhibit eukaryotic Topoisomerase I (Wu et al., 2010). It is therefore possible that inhibition is an indirect effect of CPT intercalating into DNA (Wu et al., 2010).

Tyrosines in the PSC-CTR are not required for topoisomerase-like activity

If PSC has an intrinsic type I topoisomerase activity, we would have expected that it relies on a tyrosine to perform the nucleophilic attack (Pommier et al., 2016). However, mutating all tyrosines in a fully active fragment of the PSC-CTR did not change the associated topoisomerase-like activity. The serine recombinases use serine and not tyrosine to modify DNA topology (Yang, 2010). These enzymes were originally described by their role in site-specific recombination which can lead to the exclusion or insertion of DNA fragment between two sites, as for example the phage integrase which allows recombination between attP and attB sites (Groth and Calos, 2004). On a plasmid, recombination or inversion by serine or tyrosine recombinase between two sites on the same plasmid without the removal of the sequence in between, was shown to lead to a change in topoisomers (Johnson, 2015; Stark, 2014). The key outstanding question is whether the PSC-CTR forms a covalent complex with DNA to modify topology, and if so, which residue is involved.

The change in DNA topology by the PSC-CTR and PRC1 may rely on three steps: wrapping, cutting and ligating the DNA.

If the change in plasmid topology is not due to a topoisomerase contaminant or an intrinsic topoisomerase activity of the PSC-CTR, it could be explained by PSC or another contaminant protein being able to cut and ligate the DNA. PRC1 or the PSC-CTR, by wrapping the DNA around itself will introduce a transient change in topology, which will be converted to a change in linking number that persists after protein removal only if the DNA has been cut and re-ligated while the protein was bound (Fig. 2). If this hypothetical endonuclease and ligase are involved, our biochemical analysis indicates that neither enzyme requires ATP or Mg^{2+} which is not typical (Taylor et al., 2011; Yang and Chan, 1992). The mechanism by which PRC1 and PSC alter DNA topology remains unclear. We believe it is likely to rely on PRC1 wrapping DNA around itself to add a small number of negative supercoils to a relaxed plasmid and maintain a small number of negative supercoils after relaxing a supercoiled plasmid, combined with a topoisomerase-like activity (Fig. 2).

Is PRC1 topoisomerase-like activity involved in R-loop resolution in vivo?

R-loops are formed when RNA hybridizes with DNA forming RNA-DNA hybrid and a displaced ssDNA strand. R-loops, in common with all processes that open the dsDNA helix, lead to local accumulation of DNA supercoils. R-loop formation is determined by the sequence of the RNA-DNA hybrid but also by the supercoiling of the region: a high level of negative supercoiling will stabilize R-loop formation (Brochu et al., 2020; Chedin and Benham, 2020). Reduction of topoisomerase activity is therefore expected to stabilize R-loops by increasing supercoiling. In human cells Topoisomerase I depletion indeed leads to R-loop accumulation at sites enriched for H3K36me3/H3K9me3 (Manzo et al., 2018). Unexpectedly, however, the same treatment results in R-loop depletion at sites enriched for H3K27me3 and PcG proteins. These data suggest distinct topology-related mechanisms control R-loops operate in these repressed domains. In *Drosophila*, we found that some Polycomb binding sites form R-loop and the ones which form R-loops are more likely to be bound by Polycomb group proteins (Alecki et al., 2020). Given that the activity associated with PRC1 can resolve R-loops in vitro (Fig. 3-10b, e), it will be interesting to determine whether PRC1 controls R-loop dynamics in vivo.

3.3 Figures

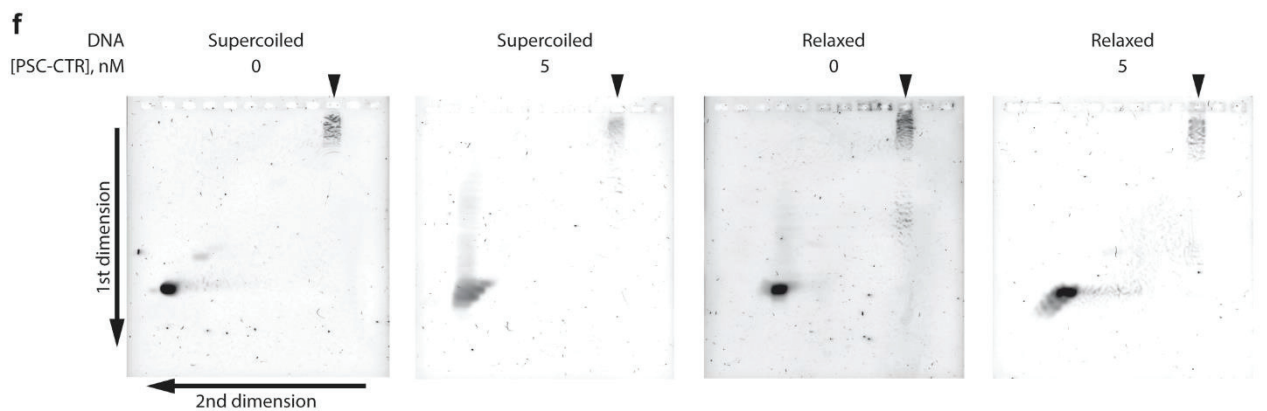
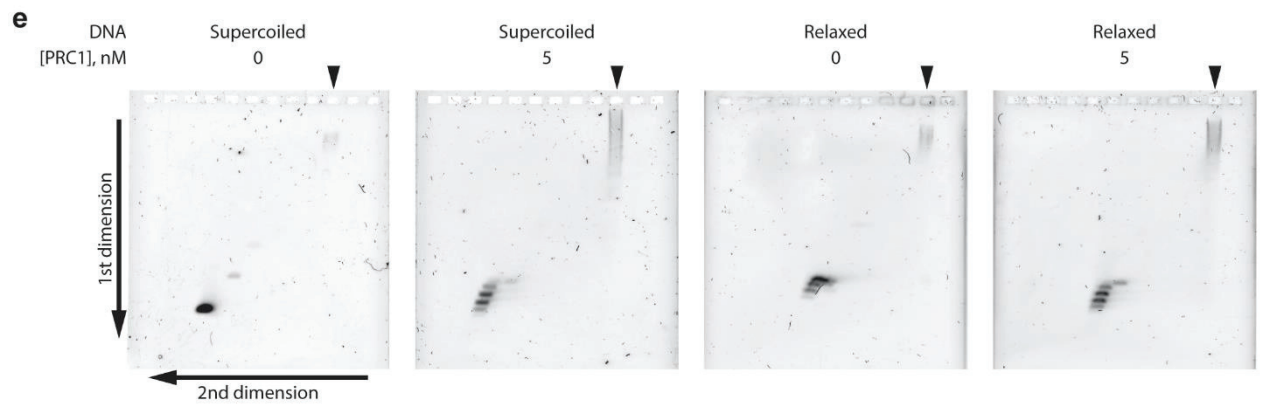
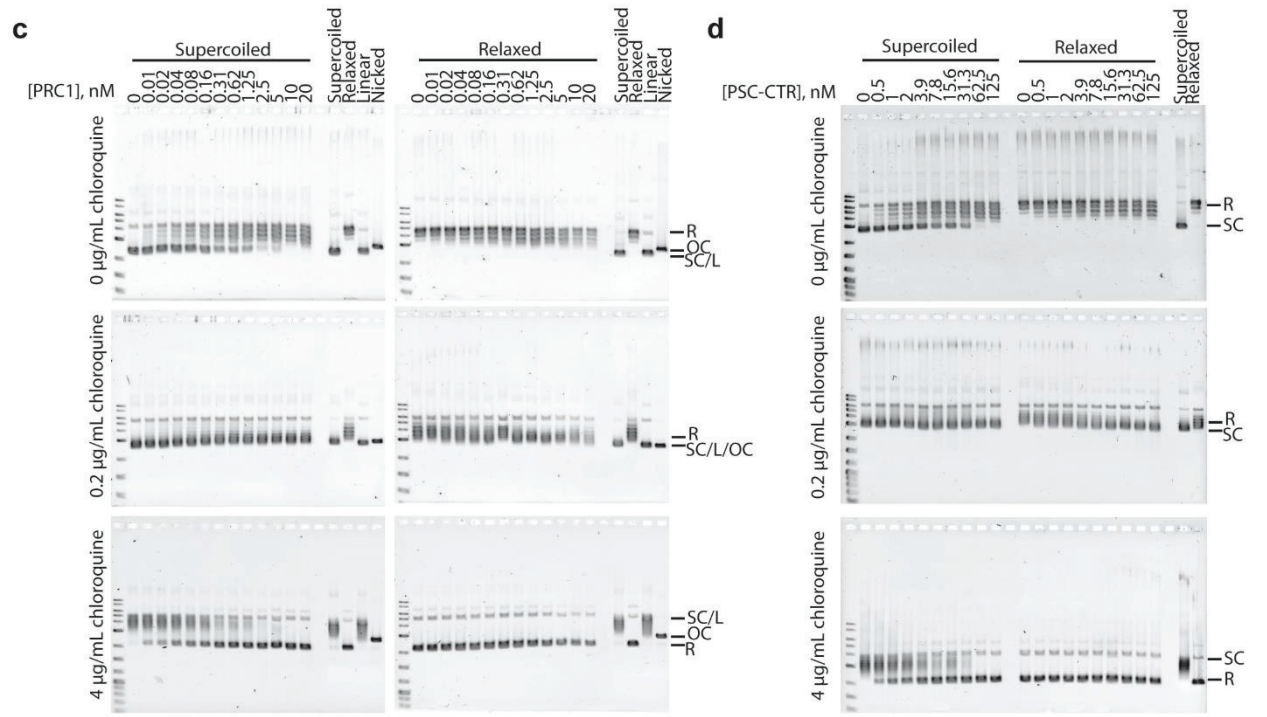
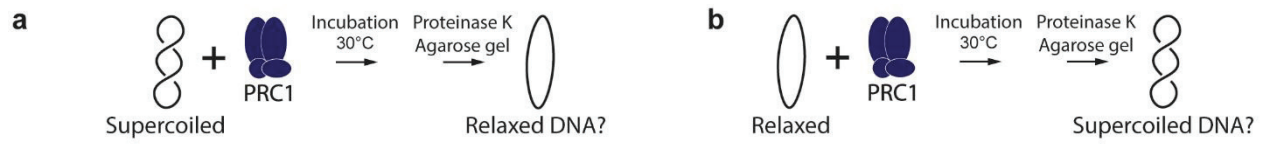


Figure 3-1 PRC1 and the PSC-CTR have a topoisomerase-like activity.

a-b. Schematic of the topology assay starting from negatively supercoiled DNA (a) or relaxed DNA (b). c. Representative gels of PRC1-induced relaxation of negatively supercoiled DNA and addition of supercoils to relaxed DNA. [PRC1]: 0.01-10 nM. d. Representative gels of the PSC-CTR-induced relaxation of negatively supercoiled DNA and addition of supercoils to relaxed DNA [PSC-CTR]: 0.5-125 nM e-f. Representative 2 dimensional gels of supercoiled and relaxed plasmid after incubation with PRC1 (e) or the PSC-CTR (f). Black arrow indicates where the samples were loaded.

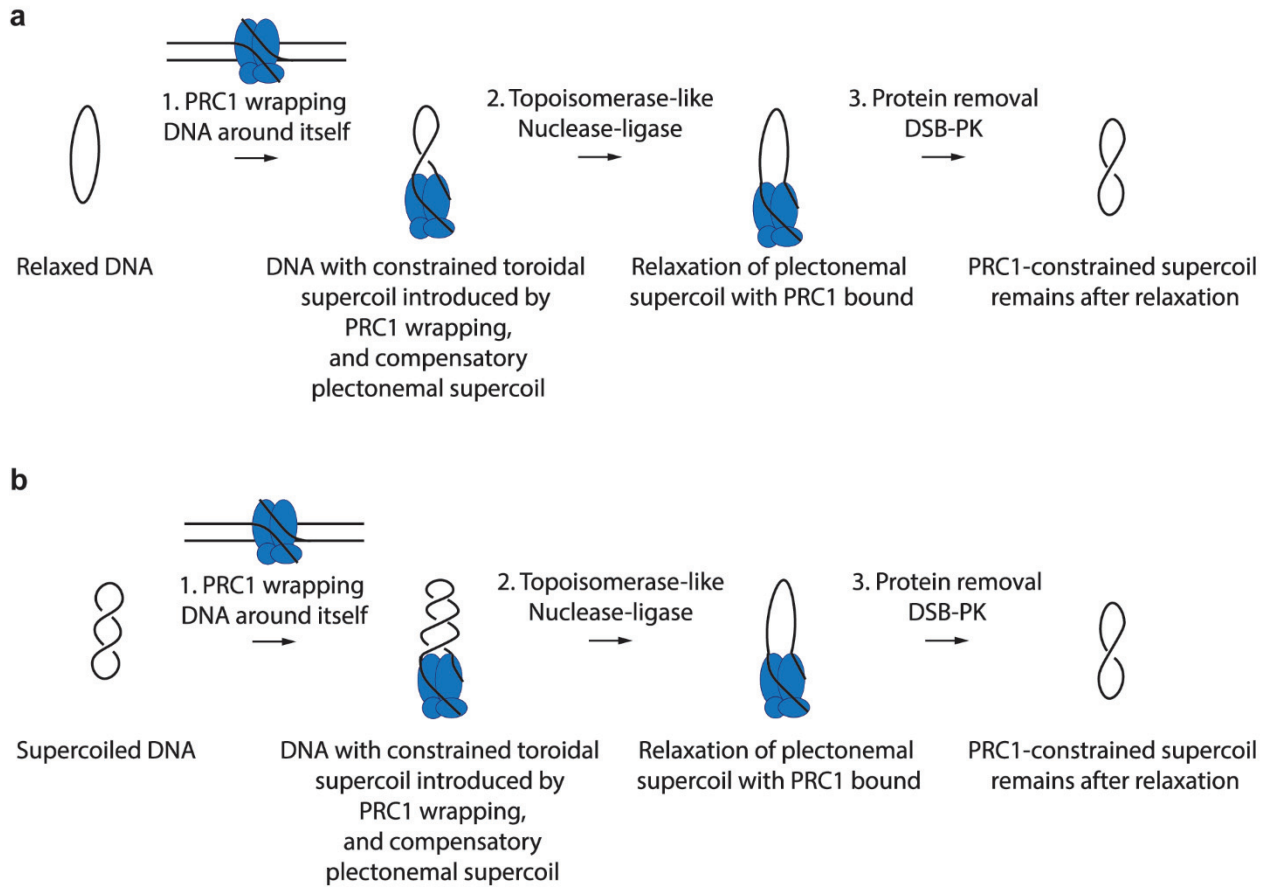


Figure 3-2 Model depicting PRC1 induced DNA topology change by DNA wrapping associated with topoisomerase or cut and ligate activities

a. Model describing how PRC1 can introduce negative coils in a relaxed plasmid. (1) PRC1 introduces a negative supercoil by wrapping the DNA around itself. (2) Complete DNA relaxation by a topoisomerase-like or a nuclease-ligase activity which can be intrinsic to PRC1 or the action of a contaminant protein. (3) Protein removal reveals the constrained supercoil that was due to the PRC1 DNA binding and wrapping. b. Model describing how PRC1 can partially relax a supercoiled plasmid. (1) PRC1 introduces a supercoil by wrapping the DNA around itself. This step can be preceded by DNA relaxation. (2) and (3) are the same as in the model of the introduction of coils by PRC1.

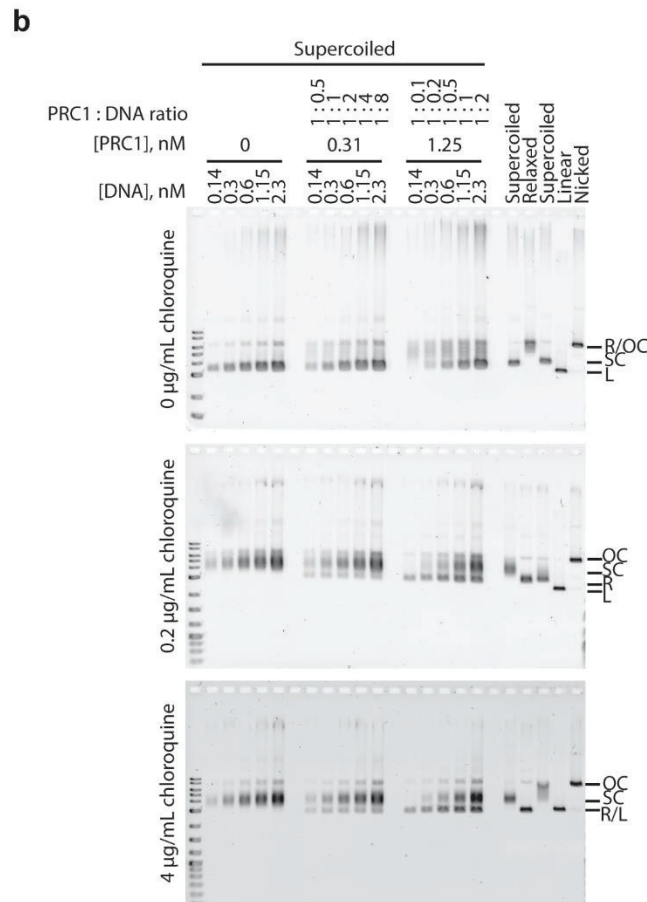
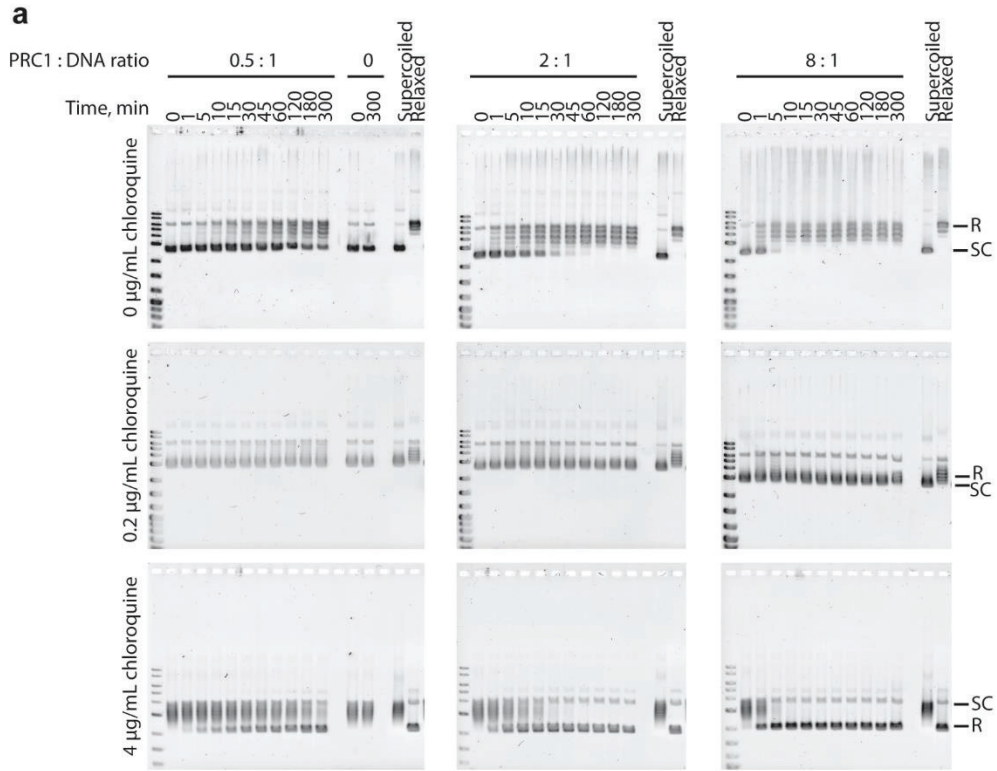
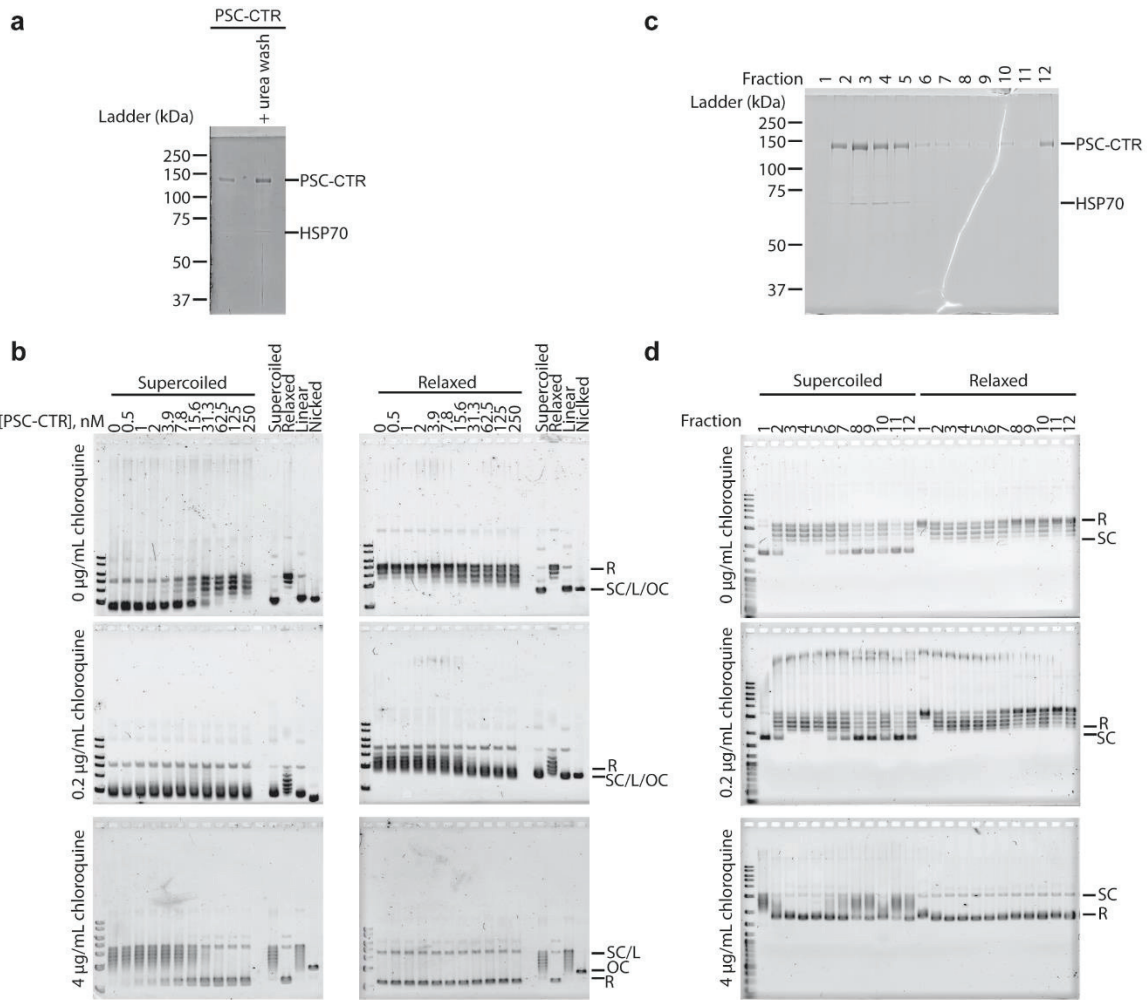


Figure 3-3 Characterization of PRC1 Topoisomerase-like activity.

a. Time course of PRC1 Topoisomerase-like activity. Time: 0 min., 1 min., 5 min., 10 min., 15 min., 30 min., 45 min., 1 H, 2 H, 3 H and 5 H. [DNA] = 0.6 nM. b. Representative gels of PRC1 topology assay in presence of increasing amount of plasmid. DNA: 5-80 ng.

SC: supercoiled, R: relaxed, L: Linearized, OC: nicked



e

| Protein name | Protein IDs | Unique peptides | Mol. weight [kDa] | Sequence length | Q-value | Score | MS/MS count | Only identified by site |
|--------------------------------|--------------------------------|-----------------|-------------------|-----------------|---------|--------|-------------|-------------------------|
| PRC1 subunit PSC | sp P35820 PSC_DROME | 665 | 169.83 | 1601 | 0 | 323.31 | 16629 | |
| PRC1 subunit dRING | sp Q9VB08 RING1_DROME | 184 | 47.255 | 435 | 0 | 323.31 | 3787 | |
| PRC1 subunit Pc | sp P26017 PC_DROME | 133 | 43.976 | 390 | 0 | 323.31 | 1323 | |
| Histone H2A | SP P06897 Histone_H2A | 64 | 13.966 | 130 | 0 | 323.31 | 1306 | |
| Histone H2B | SP Q92130 HistoneH2B | 111 | 13.894 | 126 | 0 | 323.31 | 871 | |
| Histone H3 | SP Q92133 Histone_H3 | 89 | 15.374 | 136 | 0 | 323.31 | 516 | |
| Histone H4 | SP P62799 Histone_H4 | 77 | 11.367 | 103 | 0 | 323.31 | 185 | |
| S. frugiperda Topoisomerase II | tr A0A2H1VFE3 A0A2H1VFE3_SPOFR | 105 | 176.85 | 1564 | 1 | -2 | 14 | + |
| S. frugiperda Topoisomerase I | tr A0A2H1WTL6 A0A2H1WTL6_SPOFR | 57 | 108.04 | 929 | 1 | -2 | 1 | + |
| PSC, reverse | REV_sp P35820 PSC_DROME | 22 | 169.83 | 1601 | 1 | -2 | 2 | + |

Figure 3-4 Topoisomerase-like activity co-fractionates with the PSC-CTR and topoisomerases are not detected in protein preparations.

a. SyproRuby stained SDS-PAGE gel of PSC-CTR purified from Sf9 cells with and without urea wash. b. Representative gels of topology assays with PSC-CTR washed with urea. [PSCΔN]: 0.5-250 nM. c. SyproRuby stained SDS-PAGE gel of PSC-CTR fractionated on glycerol gradient. d. Topology assay of PSC-CTR after fractionation on glycerol gradient. 4 μL of each fraction was used in this assay. e. Table summarising the peptide count of PRC1 subunits, histones and topoisomerases found in mass spectrometry data of PRC1 lacking Ph subunit.

For all panels [DNA] = 0.6 nM. SC: supercoiled, R: relaxed, L: Linearized, OC: nicked

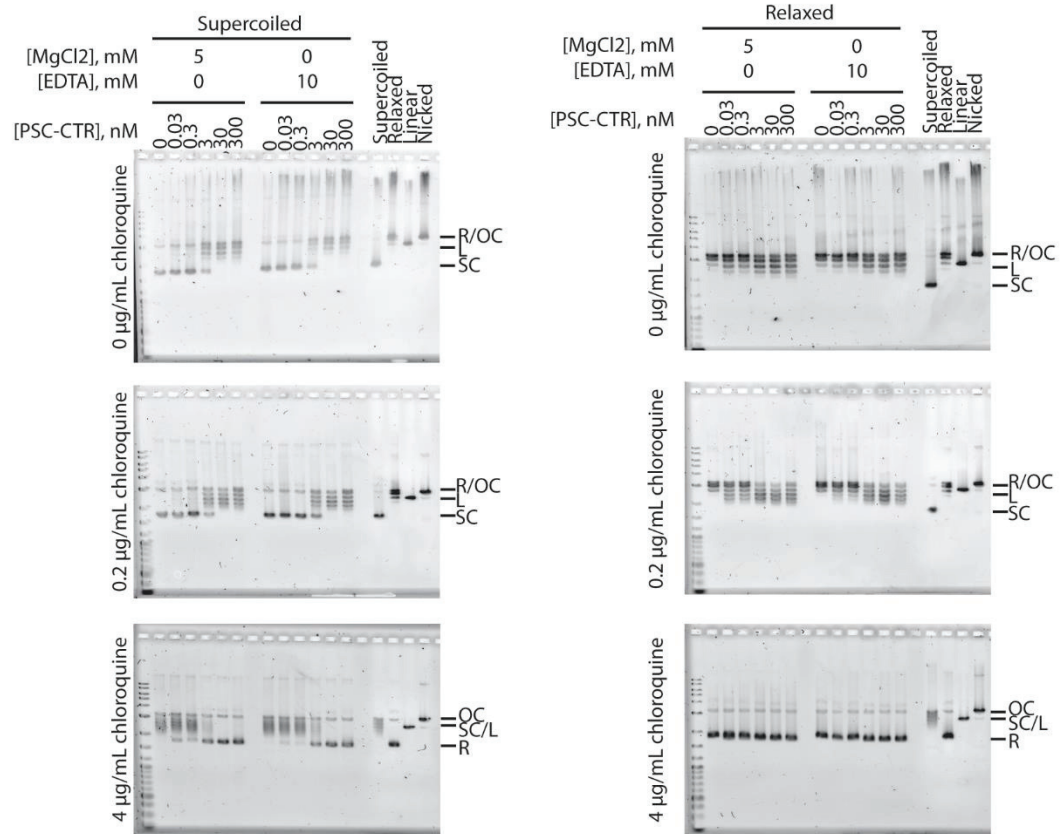
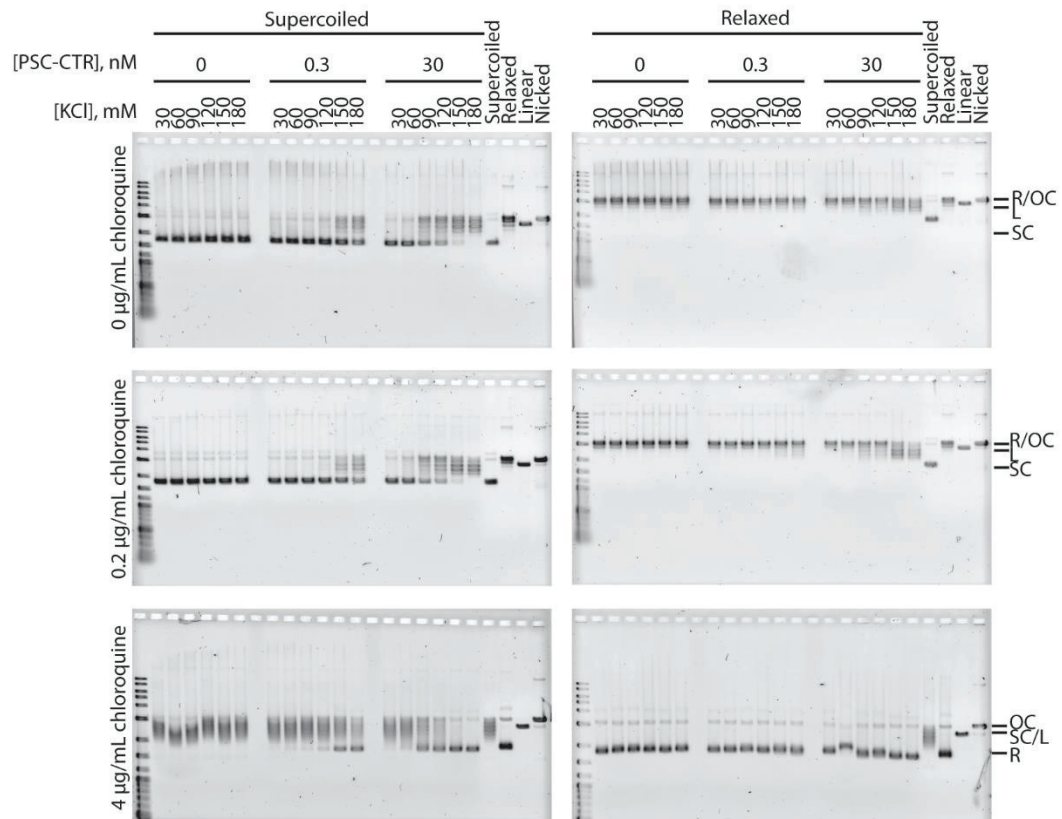
a**b**

Figure 3-5 PSC-CTR Topoisomerase-like activity is stimulated by KCl and doesn't require magnesium.

a. Representative gels of topology assay with PSCΔN in presence of 5 mM MgCl₂ or 10 mM EDTA. [PSC-CTR]: 0.03-300 nM. b. Representative gels of topology assay with PSC-CTR in presence of increasing amount of KCl. [KCl]: 30-180 mM. c. Representative gels of PSC-CTR topology assay in presence of ethidium bromide. [Ethidium bromide]: 25-400 ng/μL.

For all panels [DNA] = 0.6 nM. SC: supercoiled, R: relaxed, L: Linearized, OC: nicked

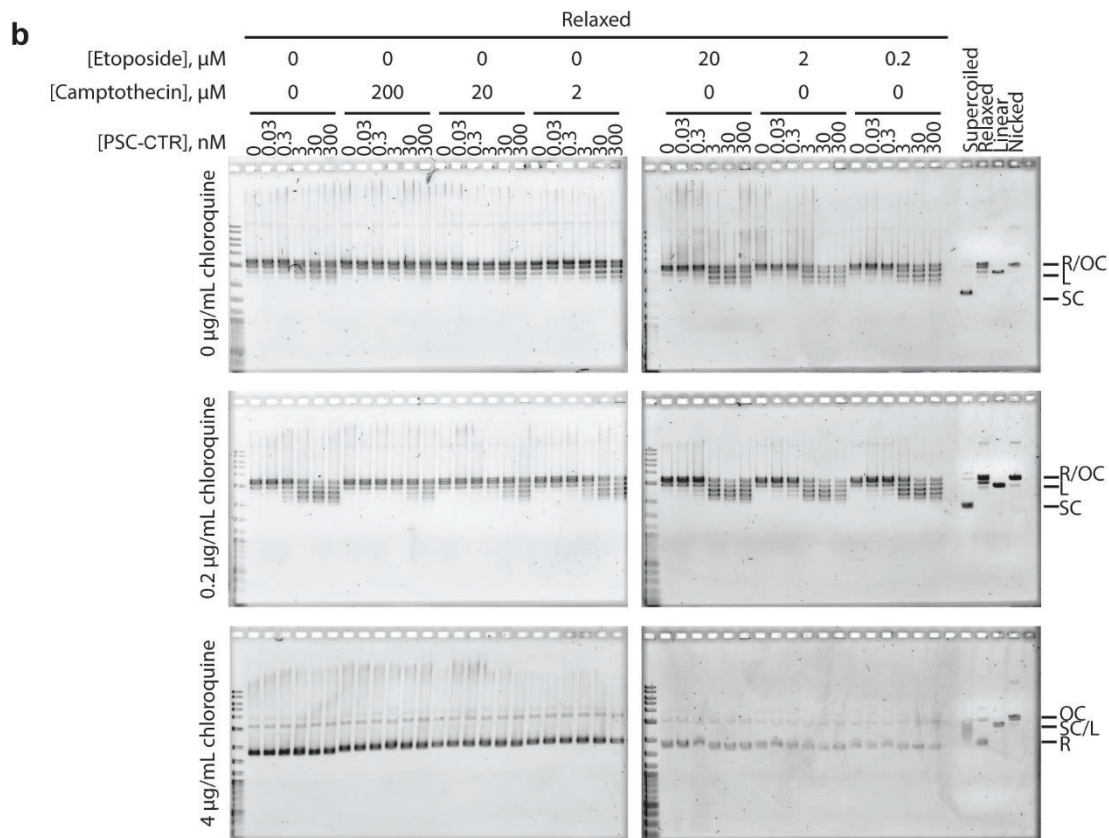
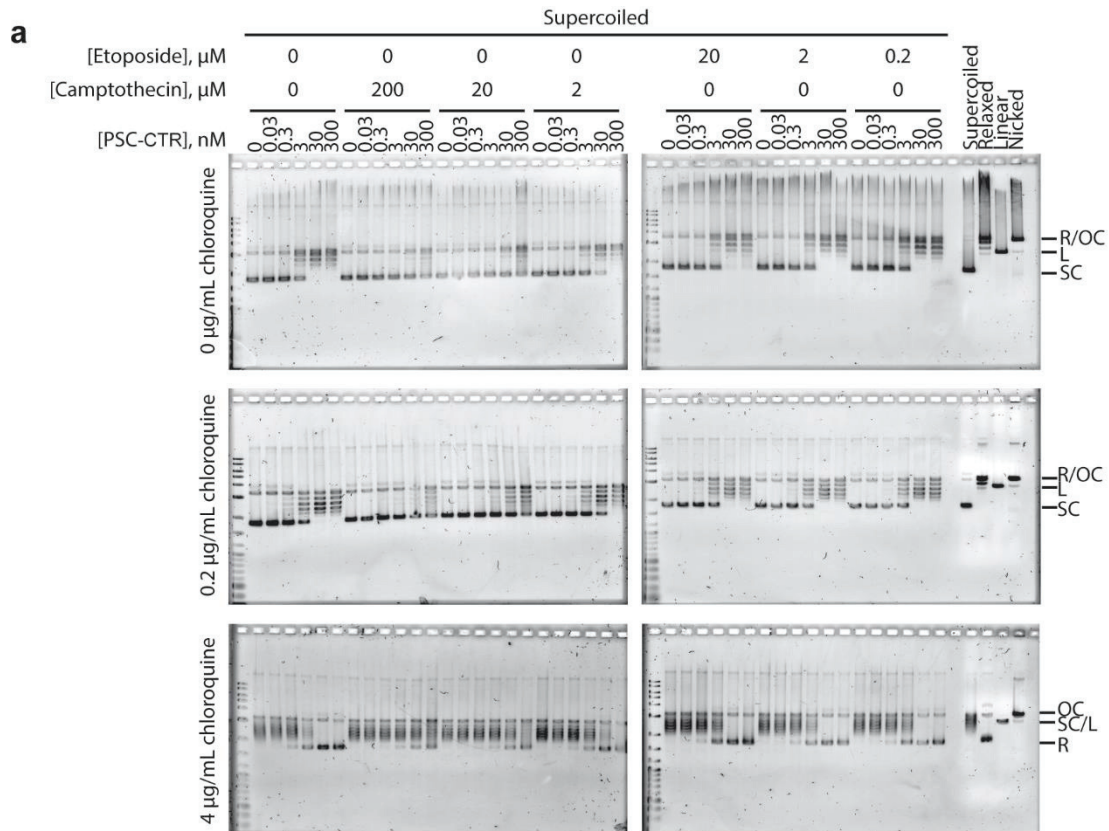


Figure 3-6 PSC-CTR topoisomerase-like activity is inhibited by camptothecin but not by etoposide.

a-b. Representative gels of PSC-CTR topology assay starting from negatively supercoiled (a) or relaxed (b) plasmids in presence of 2-200 μM of Camptothecin or 0.2-20 μM of etoposide dissolved in DMSO, or the equivalent amount of DMSO. [PSC-CTR]: 0.03-300 nM.

For all panels [DNA] = 0.6 nM. SC: supercoiled, R: relaxed, L: Linearized, OC: nicked

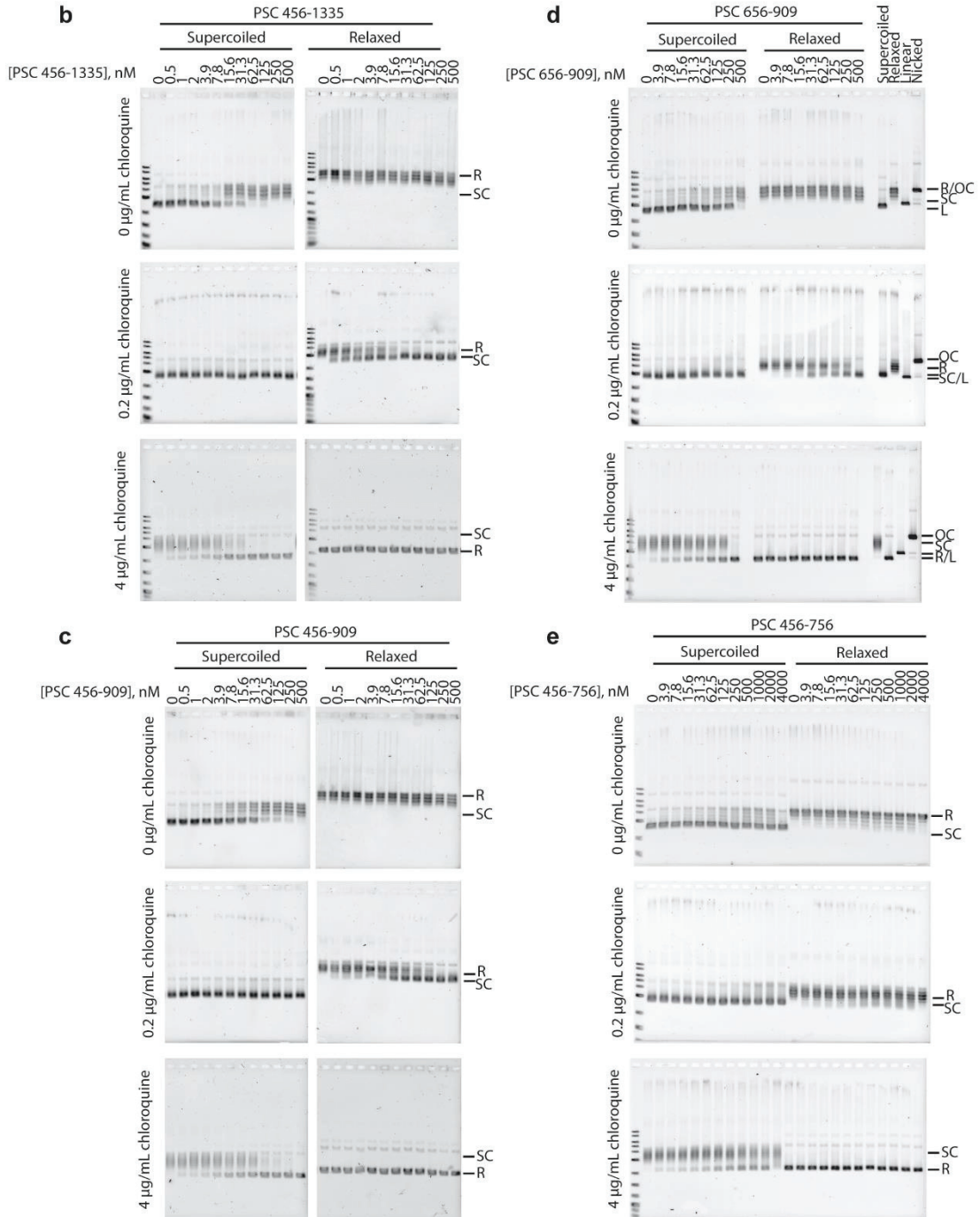
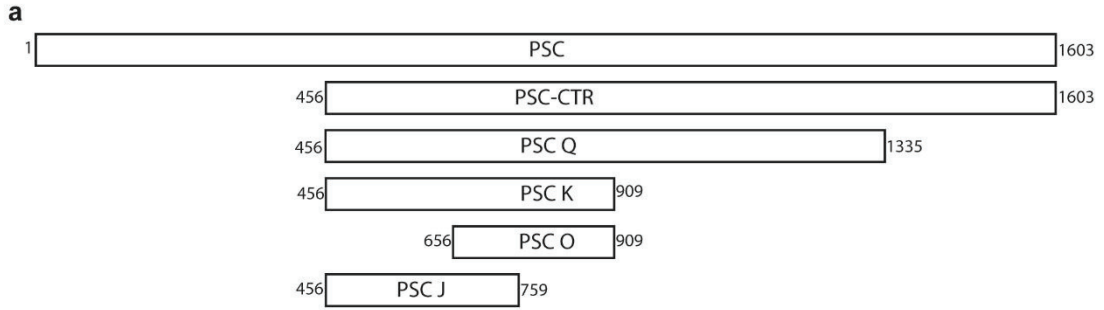


Figure 3-7 Topoisomerase-like activity of PSC truncations.

a. Scheme of PSC truncations. b-e. Representative gels of topology assay with PSC⁴⁵⁶⁻¹³³⁵ (b), PSC⁴⁵⁶⁻⁹⁰⁹ (c), PSC⁶⁵⁶⁻⁹⁰⁹ (d) and PSC⁴⁵⁶⁻⁷⁵⁶ (e). [PSC⁴⁵⁶⁻¹³³⁵]: 0.5-500 nM, [PSC⁴⁵⁶⁻⁹⁰⁹]: 0.5-500 nM, [PSC⁶⁵⁶⁻⁹⁰⁹]: 3.9-500 nM, [PSC⁴⁵⁶⁻⁷⁵⁶]: 3.9-4000 nM.

For all panels [DNA] = 0.6 nM. SC: supercoiled, R: relaxed, L: Linearized, OC: nicked

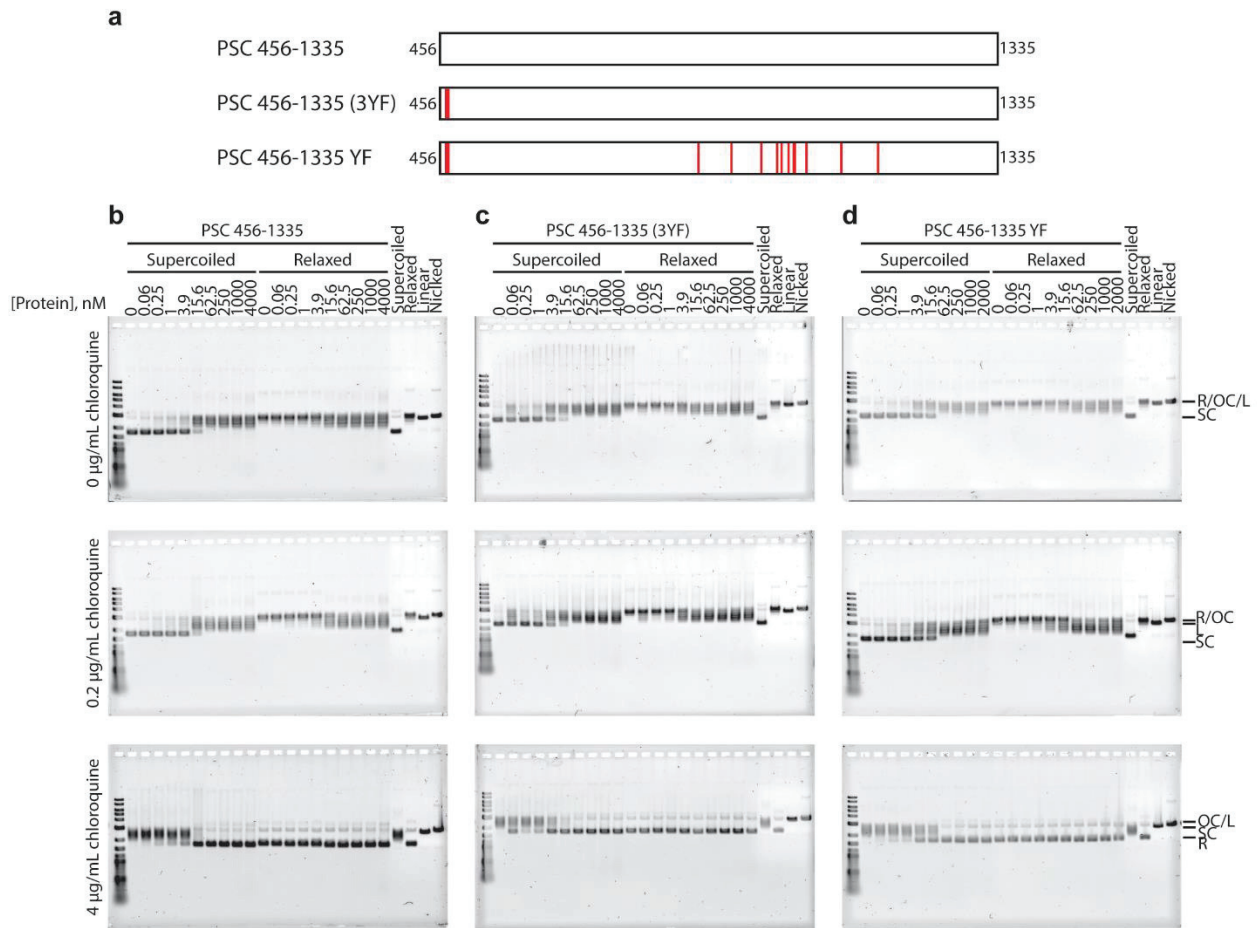


Figure 3-8 PSC topoisomerase-like activity doesn't require tyrosine.

a. Schematic of PSC⁴⁵⁶⁻¹³³⁵ protein with in red the tyrosine mutated to phenylalanine. *b-d.* Representative gels of topology assay with PSC⁴⁵⁶⁻¹³³⁵ (*b*), PSC⁴⁵⁶⁻¹³³⁵ Y2Y1 (*c*) and PSC⁴⁵⁶⁻¹³³⁵ YF (*d*). [PSC⁴⁵⁶⁻¹³³⁵]: 0.06-4000 nM, [PSC⁴⁵⁶⁻¹³³⁵ Y2Y1]: 0.06-4000 nM, [PSC⁴⁵⁶⁻¹³³⁵ YF]: 0.06-2000 nM.

For all panels [DNA] = 0.6 nM. SC: supercoiled, R: relaxed, L: Linearized, OC: nicked

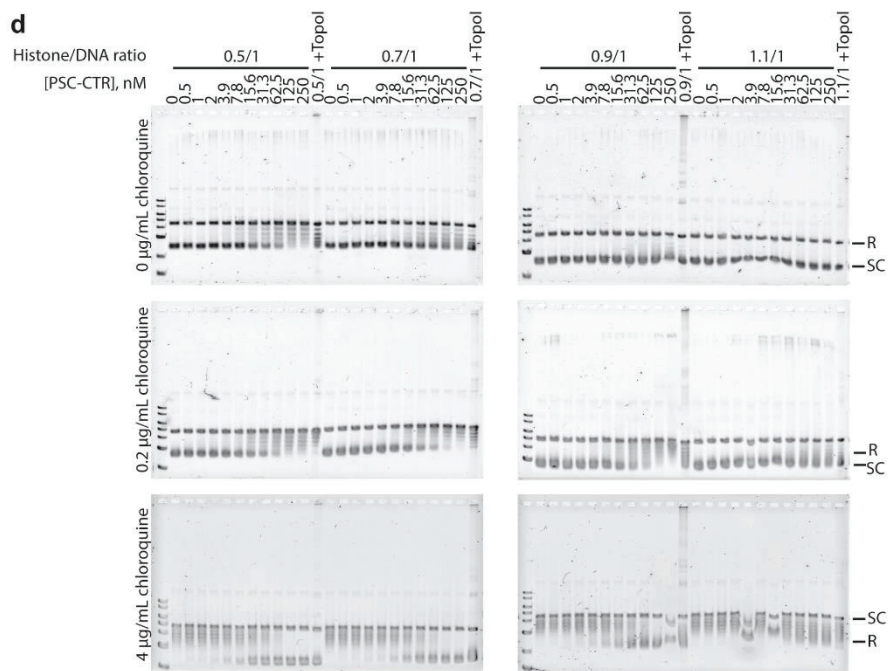
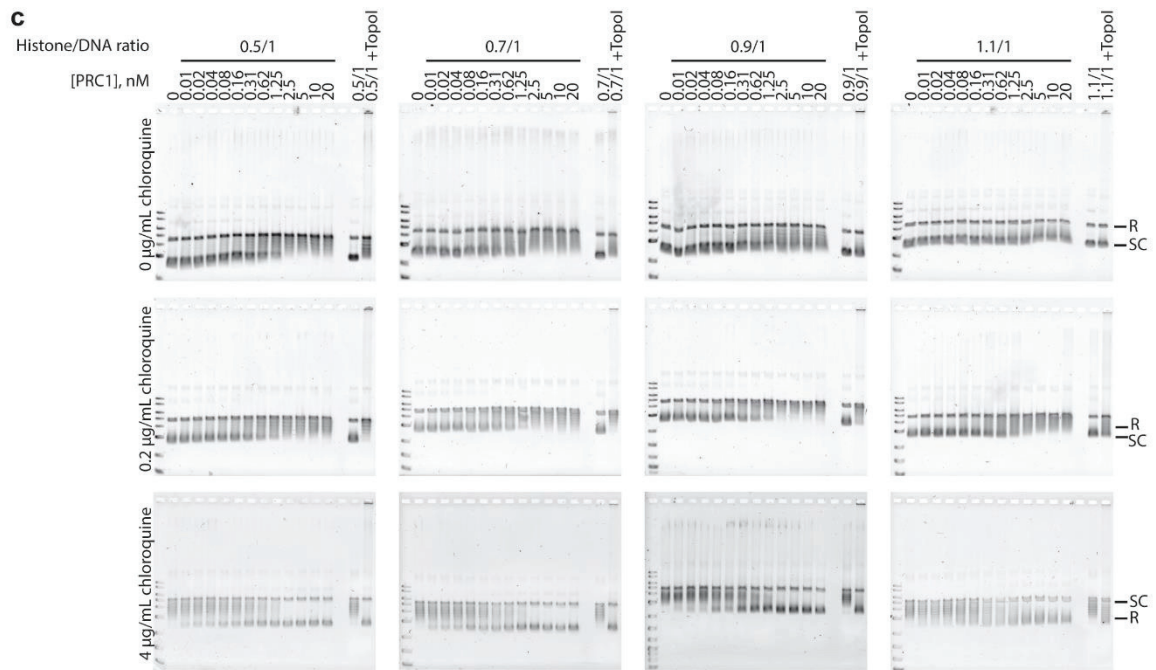
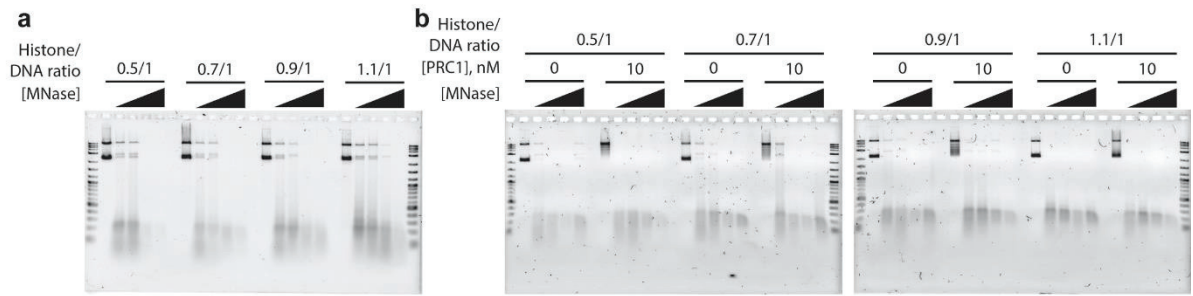


Figure 3-9 PRC1 and PSC modify the topology of chromatin assembled plasmid.

a. Representative gel of MNase digestion of chromatin assembled plasmid with various histone/DNA ratios.

b. Representative gel of MNase digestion of chromatin assembled plasmid after incubation with PRC1. c-d.

Representative gels of topology assay with PRC1 (c) and PSC-CTR (d) and chromatin assembled plasmid.

[PRC1]: 0.01-20 nM, [PSC-CTR]: 0.5-250 nM.

For all panels [chromatin] = 0.6 nM. SC: supercoiled, R: relaxed

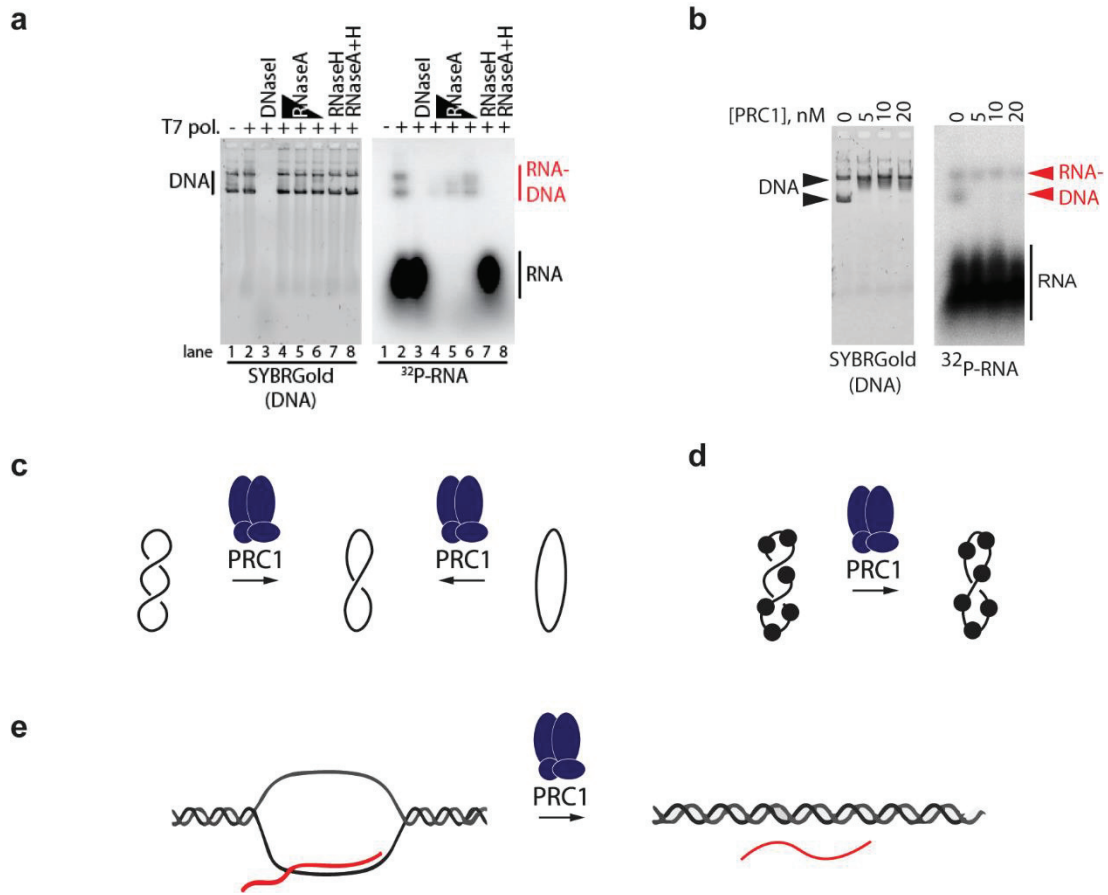


Figure 3-10 DNA relaxation by PRC1 induces loss of R-loops associated with supercoiled plasmid.

a. Radiolabelled RNAs that co-migrate with the DNA plasmid (lane 2) are confirmed as R-loops by their sensitivity to DNaseI (lane 3) and RNaseH (lane 4), and resistance to RNaseA (lanes 5-7). *b.* Representative gel showing of transcribed plasmid after incubation with PRC1. *c.* PRC1 relaxes negatively supercoiled plasmid and adds supercoils to a relaxed plasmid to keep the DNA into an intermediate level of relaxation. *d.* PRC1 modifies chromatin topology without removing nucleosomes. *e.* PRC1 relaxes supercoiled DNA which leads to loss of the R-loops associated with this plasmid.

3.4 Methods

Expression and purification of protein from Sf9 cells

PRC1, and the PSC-CTR, and other PSC truncations were expressed in Sf9 cells and purified from nuclear extracts as described for PRC1 (Alecki et al., 2020). PSC-CTR truncations were expressed in Sf9 cells and purified from whole cell extracts prepared as described for PRC2 (Alecki et al., 2020). Anti-Flag affinity purifications were carried out as described (Alecki et al., 2020). In some cases, the 1 M urea-2 M KCl was of PRC1 was not performed, but this did not affect the activity.

To purify the PSC-CTR under high stringency conditions, an N-terminal hexahistidine tag was added before the Flag tag. Nuclear extracts were prepared from baculovirus infected Sf9 cells, and anti-Flag purification carried out as described (Alecki et al., 2020), except with washing to 2 M KCl, without urea. Protein eluted from the Flag resin with Flag peptide in BC300N (20 mM Hepes, pH 7.9, 0.4 mM EDTA, 20% glycerol, 300 mM KCl, 0.05% NP40) was pooled and imidazole added to 5 mM. Ni-NTA beads were added and binding allowed to proceed overnight at 4°C. The next day, beads were loaded into a column, which was brought to room temperature. Ni-NTA beads were washed with 20 CV BC300N without glycerol but with 4M deionized urea. Beads were then washed with 5 CV BC300N without glycerol but with 6 M urea, allowed to incubate for 15min., and washed with an additional 15CV of BC300N-6 M urea, and 5 CV of BC300N-4 M urea. The column was returned to the cold room, washed with 30CV of BC300N and eluted twice with BC300 (no NP40) with 250 mM Imidazole by incubating 30 min. and collecting fractions. Fractions were dialyzed through three changes of BC300N with 0.01% NP40 to remove the imidazole, centrifuged to remove precipitates, concentrated 5-fold, flash frozen and stored at -80°C.

Protein fractionation on a glycerol gradient

7.5 µg of the PSC-CTR were loaded onto a 560 µL BC300N glycerol gradient (5%, 10%, 15%, 20%, 25%, 30% and 35% glycerol) and centrifuge 5 H at 55000 rpm in SW55-Ti swinging bucket. The gradient was resolve in twelve fractions of 47 µL each by pipetting from the top. 10 µL of each fraction were resolved on an 8% SDS-PAGE gel, which was stained with SyproRuby and imaged on a Typhoon imager (GE Healthcare). Activity was tested as described above, with 4 µL of each fraction.

Preparation of plasmid templates

Supercoiled pUC19 was prepared using a Qiagen maxiprep kit. To prepare relaxed plasmid, 5 µg of pUC19 was incubated 30 min. at 37 °C with 15 units of *E. coli* Topoisomerase I (NEB). To prepare linearized plasmid, 5 µg of pUC19 was incubated 30 min. at 37 °C with 60 units of Scal-HF restriction enzyme (NEB). Enzymes were heat inactivated for 20 min. at 65 °C. To prepare nicked plasmid, 5 µg of pUC19 was incubated 1 H at 50 °C with 30 units of Nt. BspQI nicking enzyme (NEB). DNA was purified by phenol-chloroform extraction followed by ethanol precipitation and resuspended in TE.

Topology assays

Standard conditions for topology assays are: 12 mM Hepes, pH 7.9, 180 mM KCl, 5 mM MgCl₂, 1 mM DTT, 50 µg/µL BSA (NEB), and 20 ng of negatively supercoiled or relaxed pUC19 plasmid. Reactions were initiated by addition of the protein and incubated 30 min. at 30 °C unless stated otherwise. Reactions were stopped by adding of 10 µL of DSB-PK (6.7 µg/µL of proteinase K (Biobasic), 1% SDS, 50 mM Tris-HCl pH 8.0, 25% glycerol and 100mM EDTA) and digested O.N. at 50 °C. Samples were split and resolved on 0.8% agarose 0.5X TBE gels without chloroquine or with 0.2 µg/mL or 4 µg/mL chloroquine. Gels were stained with SYBRGold (Thermo Fisher) and imaged on a Typhoon Imager. Topology assay with chromatin assembled DNA was carried out exactly the same way but with 20 ng chromatin per reaction.

2D gel

For two-dimensional gels, samples were prepared as for the topology assay and were resolved on 0.8% agarose 1X TAE gel. Gels were soaked for 4 H in 1X TAE 0.2 µg/mL chloroquine. Gels were turned 90°, run in the second dimension in a buffer containing 1X TAE 0.2 µg/mL chloroquine, stained with SYBRGold and imaged on a Typhoon Imager.

Chromatin assembly

Chromatin templates were prepared as described (Seif et al., submitted). Briefly, 10 µg of pUC19 plasmid was mixed with 5, 7, 9 or 11 µg of HeLa histone octamers in presence of 2 M NaCl, 20 mM Tris pH 7.7, 1 mM EDTA, 10 mM DTT, 0.5 mM Benzamidine HCl. Histones were assembled onto the DNA by gradient dialysis from high salt buffer (2 M NaCl, 20 mM Tris pH 7.7, 1 mM EDTA, 1 mM DTT, 0.5 mM Benzamidine HCl) to low salt buffer (250 mM NaCl, 20 mM Tris pH 7.7, 1 mM EDTA, 1 mM DTT, 0.5 mM Benzamidine

HCl). Chromatin was finally dialyzed O.N. in HEN buffer (10 mM HEPES pH 7.9, 0.25 mM EDTA, 2.5 mM NaCl) and stored at 4 °C. To prepare relaxed chromatin, 3 µg of chromatin assembled plasmid was treated with 10 units of *E. coli* Topoisomerase I for 30 min. at 37 °C. Topoisomerase was inactivated by incubation the relaxed chromatin for 20 min. at 65 °C. Chromatin was purified by phenol-chloroform extraction followed by ethanol precipitation and resuspended in TE.

MNase digestion

MNase assay and analysis of chromatin assembly was performed by incubating 20 ng of chromatin for 7 min. at RT with a titration of MNase in presence of 60% BC100 (60 mM KCl), 2 mM MgCl₂, 5 mM Tris pH 8.0, 1 mM NaCl, 12.6 mM CaCl₂. Reactions were stopped with 2.5 µL of DSB-PK and incubated 30 min. at 50 °C. Nucleic acids were resolved on 1.5% agarose 1X TBE gel. Gels were stained with SYBRGold and imaged on Typhoon imager.

MNase assay after incubation with PRC1 was carried out exactly the same way but the chromatin was incubated before with 10 nM PRC1 as in topology assay, but without the digestion with DSB-PK prior MNase assay.

R-loop formation in vitro

In vitro transcription was carried out in 50 µL reactions with 300 ng of DNA in 40 mM Tris-HCl pH 8.0, 8 mM MgCl₂, 25 mM NaCl, 2 mM spermidine, 30 mM DTT, 40 nM ATP, 40 nM CTP, 40 nM GTP, 8 nM UTP (NEB), 2.6 nmol of radiolabelled UTP (PerkinElmer) and 6 U of T7 RNA polymerase (NEB) for 30 min at 30° C. Polymerase was heat inactivated for 10 min. at 65° C. To analyze R-loop formation, each sample was split and incubated with 2 U of DNaseI in DNaseI buffer (NEB), 5 U of RNaseH in RNaseH buffer (NEB), 1 ng, 100 pg or 10 pg of RNaseA (Qiagen) or 5 U of RNaseH and 100 pg of RNaseA for 1 H at 30° C. After digestion with 3 µL of DSB-PK for 1 H at 50° C, nucleic acids were resolved on 1% agarose 0.5X TBE gels and stained with SYBRGold followed by transfer to HYBOND membranes (GE Healthcare) and exposure to a phosphor imager screen.

PRC1 incubation with transcribed template

A 5 μ l transcription reaction containing 25 ng of DNA template was carried out as described above. PRC1 was added and the reaction brought to a final composition of 180 mM KCl, 5 mM MgCl₂, 1 mM DTT and 50 ng/ μ l BSA in a 20 μ L final volume. Reactions were incubated 1 hour at 30⁰ C. Reactions were treated with 5 μ L of DSB-PK for at least 2 Hat 55° C. Samples were then resolved on 1% agarose 0.5X TBE gels, stained with SYBRGold, transferred to a HYBOND membrane, and exposed to a phosphor imager screen.

3.5 References

- Alecki, C., Chiwara, V., Sanz, L.A., Grau, D., Arias Pérez, O., Boulier, E.L., Armache, K.-J., Chédin, F., and Francis, N.J. (2020). RNA-DNA strand exchange by the *Drosophila* Polycomb complex PRC2. *Nat. Commun.* 11, 1–14.
- Baker, N.M., Rajan, R., and Mondragón, A. (2009). Structural studies of type I topoisomerases. *Nucleic Acids Res.* 37, 693–701.
- Brahmachari, S., and Marko, J.F. (2018). DNA Mechanics and Topology. In *Biomechanics in Oncology*, C. Dong, N. Zahir, and K. Konstantopoulos, eds. (Cham: Springer International Publishing), pp. 11–39.
- Brochu, J., Breton, É.-V., and Drolet, M. (2020). Supercoiling, R-Loops, Replication and the Functions of Bacterial Type 1A Topoisomerases. *Genes* 11, 249.
- Chedin, F., and Benham, C.J. (2020). Emerging roles for R-loop structures in the management of topological stress. *J. Biol. Chem.* jbc.REV119.006364.
- Chen, S.-J., and Wang, J.C. (1998). Identification of Active Site Residues in *Escherichia coli* DNA Topoisomerase I. *J. Biol. Chem.* 273, 6050–6056.
- Christensen, M.O., Barthelmes, H.U., Boege, F., and Mielke, C. (2003). Residues 190–210 of human topoisomerase I are required for enzyme activity in vivo but not in vitro. *Nucleic Acids Res.* 31, 7255–7263.
- Eng, W.K., Pandit, S.D., and Sternglanz, R. (1989). Mapping of the active site tyrosine of eukaryotic DNA topoisomerase I. *J. Biol. Chem.* 264, 13373–13376.
- Francis, N.J., Saurin, A.J., Shao, Z., and Kingston, R.E. (2001). Reconstitution of a Functional Core Polycomb Repressive Complex. *Mol. Cell* 8, 545–556.
- Francis, N.J., Kingston, R.E., and Woodcock, C.L. (2004). Chromatin Compaction by a Polycomb Group Protein Complex. *Science* 306, 1574–1577.
- Francis, N.J., Follmer, N.E., Simon, M.D., Aghia, G., and Butler, J.D. (2009). Polycomb proteins remain bound to chromatin and DNA during DNA replication in vitro. *Cell* 137, 110–122.
- Groth, A.C., and Calos, M.P. (2004). Phage Integrases: Biology and Applications. *J. Mol. Biol.* 335, 667–678.
- Johnson, R.C. (2015). Site-specific DNA Inversion by Serine Recombinases. 61.

Kahn, T.G., Schwartz, Y.B., Dellino, G.I., and Pirrotta, V. (2006). Polycomb Complexes and the Propagation of the Methylation Mark at the *Drosophila* Ubx Gene. *J. Biol. Chem.* 281, 29064–29075.

Kampranis, S.C., Bates, A.D., and Maxwell, A. (1999). A model for the mechanism of strand passage by DNA gyrase. *Proc. Natl. Acad. Sci.* 96, 8414–8419.

Kang, J.J., Faubert, D., Boulais, J., and Francis, N.J. (2020). DNA binding reorganizes the intrinsically disordered C-terminal region of PSC in *Drosophila* PRC1. *BioRxiv* 2020.06.02.130492.

King, I.F.G., Francis, N.J., and Kingston, R.E. (2002). Native and Recombinant Polycomb Group Complexes Establish a Selective Block to Template Accessibility To Repress Transcription In Vitro. *Mol. Cell. Biol.* 22, 7919–7928.

Lehmann, L., Ferrari, R., Vashisht, A.A., Wohlschlegel, J.A., Kurdistani, S.K., and Carey, M. (2012). Polycomb Repressive Complex 1 (PRC1) Disassembles RNA Polymerase II Preinitiation Complexes. *J. Biol. Chem.* 287, 35784–35794.

Lo, S.M., and Francis, N.J. (2010). Inhibition of chromatin remodeling by Polycomb Group protein Posterior Sex Combs is mechanistically distinct from nucleosome binding. *Biochemistry* 49, 9438–9448.

Lo, S.M., Follmer, N.E., Lengsfeld, B.M., Madamba, E.V., Seong, S., Grau, D.J., and Francis, N.J. (2012). A bridging model for persistence of a Polycomb Group protein complex through DNA replication in vitro. *Mol. Cell* 46, 784–796.

Manzo, S.G., Hartono, S.R., Sanz, L.A., Marinello, J., De Biasi, S., Cossarizza, A., Capranico, G., and Chedin, F. (2018). DNA Topoisomerase I differentially modulates R-loops across the human genome. *Genome Biol.* 19.

Mirkin, S.M. (2001). DNA Topology: Fundamentals. In *ELS*, (American Cancer Society), p.

Mohd-Sarip, A., Cléard, F., Mishra, R.K., Karch, F., and Verrijzer, C.P. (2005). Synergistic recognition of an epigenetic DNA element by Pleiohomeotic and a Polycomb core complex. *Genes Dev.* 19, 1755–1760.

Mohd-Sarip, A., Knaap, J.A. van der, Wyman, C., Kanaar, R., Schedl, P., and Verrijzer, C.P. (2006). Architecture of a Polycomb Nucleoprotein Complex. *Mol. Cell* 24, 91–100.

Papp, B., and Müller, J. (2006). Histone trimethylation and the maintenance of transcriptional ON and OFF states by trxG and PcG proteins. *Genes Dev.* 20, 2041–2054.

- Pommier, Y., Sun, Y., Huang, S.N., and Nitiss, J.L. (2016). Roles of eukaryotic topoisomerases in transcription, replication and genomic stability. *Nat. Rev. Mol. Cell Biol.* 17, 703–721.
- Ringrose, L., and Paro, R. (2004). Epigenetic Regulation of Cellular Memory by the Polycomb and Trithorax Group Proteins. *Annu. Rev. Genet.* 38, 413–443.
- Shao, Z., Raible, F., Mollaaghababa, R., Guyon, J.R., Wu, C., Bender, W., and Kingston, R.E. (1999). Stabilization of Chromatin Structure by PRC1, a Polycomb Complex. *Cell* 98, 37–46.
- Simpson, R.T., Thoma, F., and Brubaker, J.M. (1985). Chromatin reconstituted from tandemly repeated cloned DNA fragments and core histones: A model system for study of higher order structure. *Cell* 42, 799–808.
- Staker, B.L., Hjerrild, K., Feese, M.D., Behnke, C.A., Burgin, A.B., and Stewart, L. (2002). The mechanism of topoisomerase I poisoning by a camptothecin analog. *Proc. Natl. Acad. Sci. U. S. A.* 99, 15387–15392.
- Stark, W.M. (2014). The Serine Recombinases. *Microbiol. Spectr.* 2.
- Taylor, M.R., Conrad, J.A., Wahl, D., and O'Brien, P.J. (2011). Kinetic Mechanism of Human DNA Ligase I Reveals Magnesium-dependent Changes in the Rate-limiting Step That Compromise Ligation Efficiency. *J. Biol. Chem.* 286, 23054–23062.
- Wang, J.C. (2002). Cellular roles of DNA topoisomerases: a molecular perspective. *Nat. Rev. Mol. Cell Biol.* 3, 430–440.
- Wu, H.-Y., Shyy, S., Wang, J.C., and Liu, L.F. (1988). Transcription generates positively and negatively supercoiled domains in the template. *Cell* 53, 433–440.
- Wu, N., Wu, X.-W., Agama, K., Pommier, Y., Du, J., Li, D., Gu, L.-Q., Huang, Z.-S., and An, L.-K. (2010). A novel DNA topoisomerase I inhibitor with different mechanism from camptothecin induces G2/M phase cell cycle arrest to K562 cells. *Biochemistry* 49, 10131–10136.
- Yang, W. (2010). Topoisomerases and site-specific recombinases: similarities in structure and mechanism. *Crit. Rev. Biochem. Mol. Biol.* 45, 520–534.
- Yang, S.W., and Chan, J.Y. (1992). Analysis of the formation of AMP-DNA intermediate and the successive reaction by human DNA ligases I and II. *J. Biol. Chem.* 267, 8117–8122.

4 Discussion

4.1 Résumé des chapitres 2 et 3

Dans le chapitre 2, nous avons identifié les séquences formant des R-loops chez *D. melanogaster*. Nous avons observé que les PREs forment des R-loops *in vivo*, et que PRC1 et PRC2 sont recrutés préférentiellement au niveau de PREs ayant formé des R-loops *in vivo* et *in vitro*. Dans les chapitres 2 et 3, nous avons également identifiés deux nouvelles activités pour les complexes Polycomb : une activité d'invasion de brins pour le complexe PRC2 ainsi qu'une activité similaire à celle de la topoisomérase I pour PRC1. Dans cette section, nous discuterons de ces nouvelles activités découvertes chez les complexes Polycomb et de leurs fonctions potentielles. De plus, nous nous attarderons sur le rôle des complexes Polycomb dans la formation, reconnaissance et résolution de structures d'acides nucléiques non-canoniques telles que les R-loops.

4.2 Les R-loops peuvent-elles déterminer le statut transcriptionnel d'un gène

4.2.1 Les séquences formant des R-loops chez *D. melanogaster*

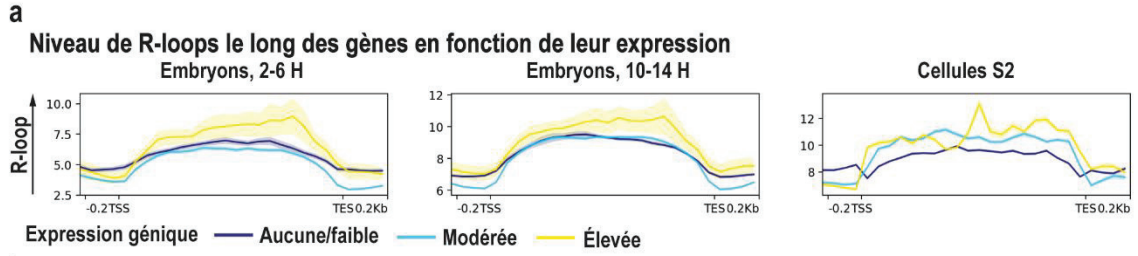
Dans le chapitre 2 nous avons pour la première fois identifié les séquences formant des R-loops dans les embryons de *D. melanogaster*. Lors de l'analyse de ces R-loops, nous nous sommes particulièrement intéressés à celles se formant au niveau des PREs. Dans cette section, nous discuterons succinctement l'ensemble des séquences formant des R-loops ainsi que leurs associations avec différentes marques épigénétiques en comparaison avec ce qui est connu chez d'autres espèces.

4.2.1.1 Les R-loops se forment aux sites d'initiation et de terminaison de la transcription

Dans le chapitre 2, nous avons identifié que les séquences formant des R-loops chez *D. melanogaster* se trouvent dans des régions transcrites. Environ 10% des séquences formant des R-loops se forment au niveau des promoteurs, 10% au niveau des sites de terminaison de transcription et 60% le long des gènes chez *D. melanogaster* (Figure 4-1a). Cette répartition est légèrement différente par rapport à ce qui a été observé chez les mammifères qui présentent un plus fort enrichissement de R-loops dans les régions promotrices et terminatrices et moins important le long du transcrit.

4.2.1.2 Les marques épigénétiques

Les R-loops, que ce soit chez les mammifères ou chez les plantes, sont essentiellement associées avec des marques actives telles que H3K4me3, H3K36me3 ou H3K27Ac, mais certaines corréntent avec la marque répressive H3K27me3 déposée par PRC2 (Fang et al., 2019; Sanz et al., 2016). À l'opposé, aucune corrélation positive n'a été observée entre les R-loops et H3K9me3 (Fang et al., 2019; Sanz et al., 2016). À l'exception des régions péri-centromériques de *S. pombe*, les R-loops sont exclues de l'hétérochromatine constitutive, car probablement trop compacte et ne permettant pas l'accès à l'ARN polymérase ou à des protéines induisant la formation de R-loop, mais toutes les autres régions du génome pouvant être transcrites et étant plus accessibles peuvent former des R-loops. Les séquences formant des R-loops chez *D. melanogaster* ne se démarquent pas de celles des mammifères et des plantes. Les R-loops sont enrichies en histones acétylées ou tri-méthylées pour la lysine 4 de l'histone 3 par rapport aux séquences à proximité de ces R-loops (Figure 4-1b). Elles ne présentent pas un enrichissement particulier pour les marques épigénétiques H3K27me3 et H3K9me3 qui sont associées à l'hétérochromatine facultative et constitutive, avec un niveau de méthylation constant au niveau d'une R-loop et à proximité de celle-ci lorsque visualisée sur une région couvrant 10 kilobases.



b
R-loop et les modifications post-traductionnelles d'histones

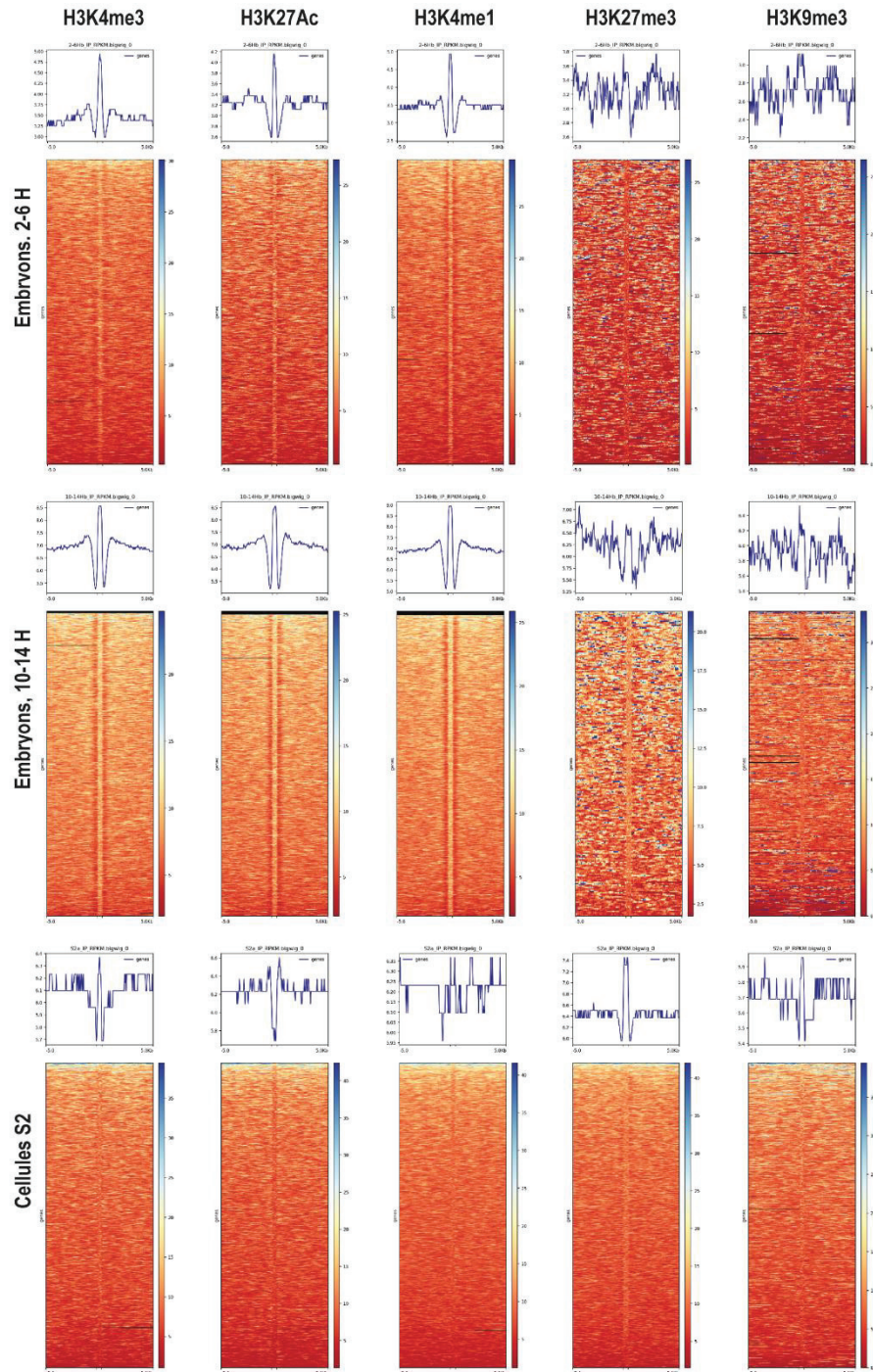


Figure 4-1 Les R-loops, la transcription et les modifications post-traductionnelles d'histones

a. Analyse métagène du niveau de R-loop en fonction de l'expression du gène au sein duquel elles se forment ou du gène le plus proche. Les données d'expression de modencode ont été utilisées. Aucune/faible expression = 0-10 FPKM, expression modérée = 10-50 FPKM, expression élevée = >50 FPKM. TSS : site d'initiation de la transcription, TES : site de terminaison de la transcription. 200 paires de bases sont présents de chaque des côtés des sites d'initiation et de terminaison de la transcription. b. Heatmap et tracé récapitulatif (summary plot) des R-loops +/- 5 kb par rapport à différentes modifications post-traductionnelles d'histones. Les données de R-loops proviennent du chapitre 2. Pour chaque échantillon les données de CHIP provenant d'embryons âgés de 0 à 4 H, de 8 à 10 H ou de cellules S2 disponibles sur modencode ont été utilisées.

4.2.2 Des R-loops sont régulées pendant le développement

4.2.2.1 Chez *D. melanogaster*

Dans les chapitre 2, nous avons montré pour la première fois que des R-loops se forment durant le développement embryonnaire de *D. melanogaster*, et qu'une partie de ces R-loops sont présentes lors du stade de développement embryonnaire précoce ne le sont plus dans le stade de développement embryonnaire plus tardif. Toutefois ces dernières R-loops représentent un nombre restreint par rapport au nombre d'entre elles qui pourrait être formées différenciellement suivant le stade de développement. En effet, bien plus souvent qu'une absence complète de R-loop au stade embryonnaire plus tardif, nous avons observé une diminution du signal. Cette diminution peut représenter soit une diminution globale de leur présence de manière aléatoire dans les différentes cellules de l'embryon, soit cette diminution est le résultat de la différenciation des cellules. Il est possible que les embryons maintiennent la présence de ces R-loops à un locus donné dans certains segments de l'embryon mais pas dans d'autres. La diminution observée serait alors le résultat de DRIP-seq fait sur une population de cellule non uniforme masquant des différences d'une cellule à l'autre. Le développement de technique pour identifier les R-loops dans une seule cellule ou performer le DRIP-seq après isolation de tissus tels que les disques imaginaux pourraient nous en apprendre bien plus sur le rôle de ces R-loops durant le développement.

Ces R-loops qui se forment différenciellement en fonction du stade de développement pourraient faire partie des étapes qui nous manquaient jusqu'à présent pour comprendre ce qui est à l'origine du recrutement des protéines du groupe Polycomb à un PRE. Ces R-loops pourrait être ce qui fait pencher un PRE du côté Polycomb et répression, plutôt que de celui des protéines du groupe Trithorax et de la transcription.

4.2.2.2 Chez les mammifères

De façon similaire aux R-loops se formant différentiellement aux PREs durant le développement, chez les mammifères les îlots CpG non méthylés peuvent former des R-loops (Ginno et al., 2012). L'identification des séquences formant des R-loops dans différentes lignées cellulaires humaines ont mis en évidence que 40 à 60% de ces R-loops étaient conservées, notamment dans les régions promotrices et terminatrices de gènes (Sanz et al., 2016). Intéressamment, les auteurs ont mis en avant deux gènes FLRT2 et FOXG1, pour lesquels la formation de R-loop varie avec l'expression de ce même gène (Sanz et al., 2016). Cette corrélation entre R-loop et expression génique peut supposer un rôle régulateur de ces premières mais aucune expérience n'a été faite pour déterminer si la formation de R-loop est la cause ou la conséquence de l'expression de ces deux gènes.

4.2.2.3 Chez les plantes

Deux études ont mis au jour que la formation de R-loops chez les plantes varie en fonction des tissus et des conditions de développement chez *Arabidopsis thaliana* et le riz (Fang et al., 2019; Xu et al., 2020). Le profil de ces R-loops est relativement constant pendant les différentes conditions de croissance, mais des variations sont néanmoins visibles pendant le développement (Xu et al., 2020). Chez *Arabidopsis thaliana*, les racines et les feuilles présentent une certaine similarité quant au profil des séquences formant des R-loops alors que les échantillons provenant de fleurs ou de graines en germination sont plus différents (Xu et al., 2020). Des résultats similaires ont été obtenus chez le riz, où 70% des séquences formant des R-loops sont communes entre la graine et un cal issu de feuilles, 30% se forment différentiellement entre ces deux échantillons (Fang et al., 2019).

Parmi les exemples de R-loops qui sont régulés à la suite de différents stimuli chez les plantes ayant fait l'objet d'étude plus approfondie, il y a celle se formant au locus FLC qui a été décrite précédemment dans l'introduction (voir la section 1.3.6.2.1) mais également celle impliquant l'ARN non codant APOLO. En présence d'auxine, une hormone de croissance végétale, l'ARN non codant APOLO est transcrit et va pouvoir en *trans*, former des R-loops à proximité de gènes (Ariel et al., 2020). Cette formation de R-loop conduit à la perte de contact entre la région ayant formée cette structure d'acides nucléiques non-canonique et une région distale qui conduisaient jusqu'alors à l'inhibition de l'expression de ces gènes (Ariel et al., 2020). Cet ARN non codant dont l'expression est induite en présence d'auxine, forme des R-loops durant le développement végétatif de la plante afin de permettre l'expression de gènes jusqu'alors réprimés et stimuler la croissance verticale (Ariel et al., 2014, 2020).

Des plantes aux mammifères en passant par les insectes, des R-loops se forment différemment durant le développement ou en réponse à divers signaux. Quelques-unes de ces R-loops ayant fait l'objet d'une attention particulière chez les plantes ont mis au jour qu'elles pouvaient participer au maintien ou au changement de statut transcriptionnel d'un gène. De plus la partie hybride de ces R-loops se forme avec un ARN non codant soulevant la possibilité que l'une des fonctions de certains ARN non codants serait de réguler l'expression génique via la formation de R-loops. Si tel est le cas y-a-t-il une différence entre ces R-loops formées par des ARN non codants, potentiellement régulatrices et les R-loops se formant de manière co-transcriptionnelle au niveau de gènes codants pour des protéines ? Ces dernières pouvant hypothétiquement être un sous-produit de la transcription devant être résolues.

4.2.3 Les R-loops sont-elles associées avec d'autres caractéristiques

4.2.3.1 Les quadruplexes de G

A l'exception de *Saccharomyces cerevisiae*, les séquences formant des R-loops sont enrichies en GC et présentent un biais avec plus de G dans le brin d'ARN formant l'hybride ARN-ADN et donc plus de G également dans le brin d'ADN déplacé. Ces G peuvent être prônes à former des structures secondaires tels que les quadruplexes de G qui dans ce cas stabiliseraient la formation de R-loops en ne laissant pas exposé l'ADN simple-brin qui est susceptible d'être la cible de nucléase (Figure 4-2a). *In vitro*, la présence de quadruplexes de G sur le brin d'ADN déplacé d'une R-loop détermine en partie la forme qu'adopte la R-loop lorsqu'elle est visualisée à l'aide d'un microscope à force atomique (Carrasco-Salas et al., 2019). *In vivo*, les télomères peuvent former des R-loops ainsi que des quadruplexes de G, ces derniers peuvent être liés par la télomérase (Moye et al., 2015). Plus généralement, l'utilisation d'outils de prédiction a mis au jour que 75% des séquences prédites pour former des R-loops au niveau de gènes comprennent des quadruplexes de G, ce qui laisse présumer que les R-loops et les quadruplexes de G pourraient être associés à de nombreux loci à l'échelle du génome (Jenjaroenpun et al., 2015). Ces quadruplexes de G, tout comme les R-loops, pourraient avoir un potentiel régulateur en permettant le recrutement de certaines protéines ou en excluant d'autres.

Nous, ainsi que d'autres groupes, avons mis en évidence que la partie ARN de l'hybride ARN-ADN d'une R-loop est enrichie en G, la partie ADN étant donc enrichie en C. Si cet hybride ARN-ADN au sein d'une R-loop n'est pas continu et possède quelques dizaines de paires de bases où ces deux acides nucléiques ne sont pas hybridés l'un avec l'autre, des structures d'acides nucléiques non-canoniques pourraient apparaître. L'ARN riche en G pourrait lui-même former un quadruplexe de G, seul ou de façon hybride avec

un ADN présentant des propriétés similaires (Fay et al., 2017). L'ADN riche en C pourrait quant à lui former un motif I qui repose sur des liaisons entre deux cystéines ou parfois deux thymines (Figure 4-2) (Abou Assi et al., 2018)

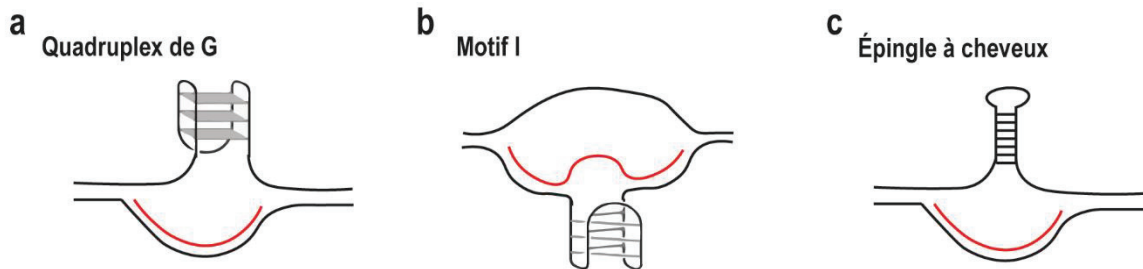


Figure 4-2 Schéma de structures d'ADN non-canoniques pouvant coexister avec une R-loop

a. Schéma d'un quadruplex de G. b. Schéma d'un motif I. c. Schéma d'une structure en épingle à cheveux.

4.2.3.2 La méthylation de l'ARN

La méthylation de l'adénine de l'ARN (m6A) est une modification réversible qui peut être reconnue par des protéines lectrices qui participent à la maturation des ARN ainsi qu'à leur localisation (Zhu et al., 2020). Dans les cellules humaines il y a une corrélation positive entre la présence d'une R-loop et la méthylation de l'ARN (Abakir et al., 2020; Yang et al., 2019). Abakir et al. ont montré que ces pics communs aux R-loops et à la méthylation de l'ARN sont sensibles à la RNase H mettant en évidence que ces premières peuvent être méthylées. Néanmoins, l'anticorps utilisé afin d'identifier les adénines méthylées n'est pas spécifique aux ARN et peut détecter la méthylation d'ADN, soulevant donc la possibilité que m6A puisse être présent sur l'ADN ou sur l'ARN. Que ce soit sur la partie ARN, ADN ou les deux, ces hybrides ARN-ADN présentent des méthylations au niveau d'adénine. Cependant il est à regretter que les auteurs n'aient pas cherché à connaître le degré de méthylation des R-loops car cette marque est très courante dans le génome à en juger par quelques 20 000 pics que les auteurs ont détecté par séquençage (Abakir et al., 2020). Chez les plantes, c'est la méthylation des adénines de l'ADN qui a été corrélée positivement avec la présence de R-loops (Fang et al., 2019).

S'il est certain que les R-loops peuvent être méthylés au niveau de l'ADN et/ou de l'ARN, il reste toutefois à déterminer le rôle de cette méthylation : si cela peut affecter la stabilité des R-loops, permettre ou prévenir le recrutement de certaines protéines.

4.3 Les R-loops sont-elles une caractéristique des PREs et peuvent-elles être utilisées pour les identifier chez d'autres espèces ?

4.3.1 Les motifs aux PREs

Comme décrit dans l'introduction, les PREs ont été identifiés fonctionnellement chez *D. melanogaster*. La séquence définissant un PRE est bien moins précise que celle d'un facteur de transcription par exemple. Les séquences définissant les PREs ont été initialement décrites comme une accumulation de sites de liaison pour des facteurs de transcription, depuis d'autres motifs tels que des répétitions GT ont été ajoutés (Bredesen and Rehmsmeier, 2019; Ringrose et al., 2003; Zeng et al., 2012). Dans cette partie, nous discuterons si la formation de R-loops est une caractéristique des PREs.

4.3.2 Les motifs présents aux PREs ayant formés des R-loops sont différents de ceux qui n'en ont pas formés

4.3.2.1 Les motifs présents aux PREs qui ont ou n'ont pas formé de R-loops

L'identification des R-loops aux PREs nous a permis de mettre en évidence que les PREs ayant formés des R-loops sont enrichis avec des motifs différents que ceux qui n'en ont pas formés. (Figure 4-3). Que ce soient les embryons (aux deux stades de développement embryonnaires) ou la lignée cellulaire S2, les PREs ayant formé des R-loops sont enrichis avec le motif du facteur de transcription Trl. Ce facteur de transcription fait partie du groupe de protéine Trithorax qui sont antagonistes aux protéines du groupe Polycomb. Parmi les motifs présents aux PREs n'ayant pas formé de R-loop on peut observer celui de Pho, protéine du groupe Polycomb et sous-unité du complexe PhoRC.

Ces différences de motifs entre les PREs ayant et n'ayant pas formé de R-loops peut laisser penser qu'il y a plusieurs catégories de PREs. Cette idée de différentes catégories de PREs est supportée notamment par l'exemple du gène *vg*. La délétion de PREs du gène *vg* dont l'expression est contrôlée par plusieurs d'entre eux, a montré que l'un de ces PREs est impliqué dans la mise en place de la répression de *vg* alors que l'autre dans celle de l'expression (Ahmad and Spens, 2019). La capacité à former une R-loop et donc d'être enrichi avec des motifs différents par rapport aux PREs qui ne forment pas de telles structures non-canoniques, pourraient indiquer d'une fonction de ces PREs plus dans la mise en place de la répression que dans l'activation de la transcription.

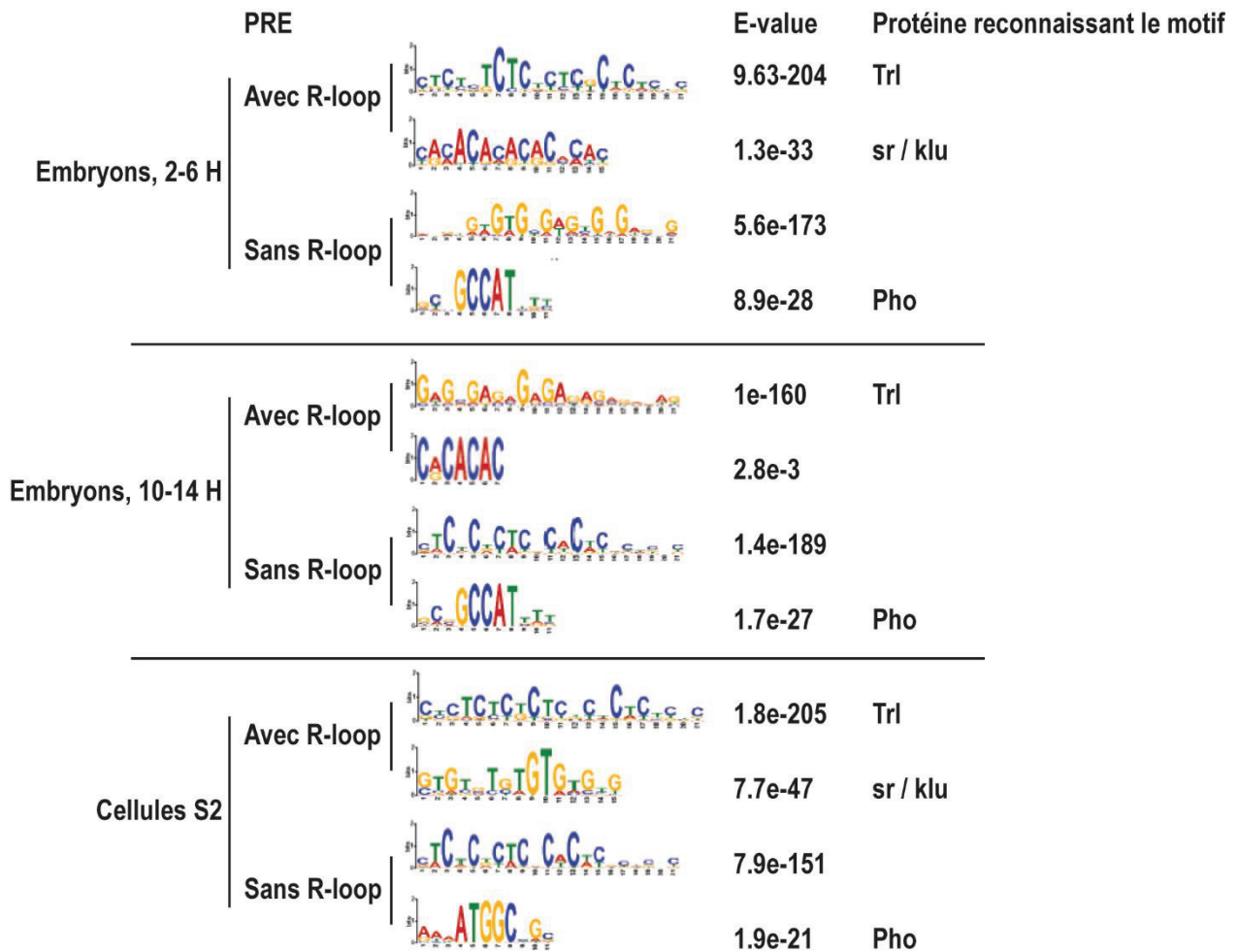


Figure 4-3 Les motifs présents aux PREs ayant ou n'ayant pas formés de R-loops

Identification des deux motifs les plus enrichis au niveau des PREs ayant ou n'ayant pas formés de R-loop avec MEME-suite. Pour les motifs au niveau d'un PRE ayant formés des R-loops, le motif peut être présent dans la partie hybride d'une R-loop, l'ADN simple-brin ou sur une partie du PRE n'étant pas comprise dans la R-loop. Lorsque cela a été possible, la ou les protéines reconnaissant le motif identifié a été indiquée(s).

4.3.2.2 Les facteurs de transcription aux PREs peuvent-ils reconnaître leur séquence cible quand elle est impliquée dans une R-loop ?

Il est possible également que ce soit la présence d'une R-loop qui fasse changer un même PRE de permissif à répressif ou inversement, en affectant la reconnaissance de ces motifs par les protéines qui les ciblent. Si comme nous l'avons observé *in vitro* dans le chapitre 2, les R-loops promeuvent le recrutement des protéines Polycomb, la formation d'une R-loop à ce même site pourrait prévenir le recrutement de Trl en l'empêchant de reconnaître son site de liaison. Il sera donc intéressant de déterminer si Trl, et les autres facteurs de transcription et protéines recrutés au niveau des PREs peuvent encore reconnaître leur motif s'il est présent au sein d'un hybride ARN-ADN ou sur un simple-brin d'ADN, et si leur affinité pour leur site

de liaison augmente lorsqu'il est compris dans une R-loop. De plus, plusieurs des protéines liant les PREs reconnaissent les mêmes motifs et donc ne peuvent pas co-exister sur le même site. La présence d'une R-loop déterminerait peut-être le recrutement préférentiel de GAF par rapport à Psq ou de Sp1 par rapport à KLF.

La présence d'une R-loop en permettant ou en prévenant la liaison des différentes protéines pouvant être recrutées aux PREs, pourrait déterminer le statut de ce dernier et donc de l'expression ou de la répression du gène qu'il contrôle. Les R-loops permettraient donc de comprendre comment un même PRE peut induire la répression ou permettre l'expression d'un même gène dépendamment du stade de développement ou du segment de l'animal. Si tel est le cas, il restera à déterminer l'origine de ces R-loops, si elle se forme en *cis* de manière co-transcriptionnelle lors de la transcription d'un PRE, ou en *trans* avec l'assistance de protéines pouvant induire leur formation.

4.4 Les R-loops sont-elles impliquées dans le recrutement de protéines ?

4.4.1 Le recrutement des protéines du groupe Polycomb par les R-loops

4.4.1.1 PRC1 et PRC2 reconnaissent les R-loops, mais quelle partie ?

Dans le chapitre 2 nous avons montré que PRC1 et PRC2 lient préférentiellement les R-loops *in vitro*, et sont recrutés aux PREs ayant formé des R-loops *in vivo*. Mais nous ne savons pas quelle partie d'une R-loop ils lient. *In vitro* nous avons observé que PRC1 et PRC2 lient de manière similaire une R-loop et l'ADN double-brin avec une bulle au milieu. Il est possible que les complexes Polycomb reconnaissent l'ADN simple-brin ou la jonction entre ADN double-brin et l'ADN simple-brin. La reconnaissance de la R-loop pourrait également se faire sur la jonction entre l'hybride ARN-ADN et l'ADN double-brin. Il est aussi possible que les complexes Polycomb lient préférentiellement la partie hybride ARN-ADN en raison de sa structure qui est différente de celle de l'ADN double-brin. D'autres expériences seront nécessaires afin de déterminer quelle structure est reconnue par les complexes Polycomb ainsi que s'ils reconnaissent des parties différentes leur permettant alors théoriquement de lier une même R-loop simultanément.

In vivo, les R-loops se composent d'un ARN qui peut, ne pas être dans sa totalité impliqué dans la formation d'une R-loop. Il se peut que la partie d'ARN simple-brin adopte des structures secondaires pouvant elles

aussi jouer un rôle dans le recrutement de protéines. De plus l'ADN simple-brin déplacé par l'hybride peut lui aussi adopter des structures secondaires. Nous n'avons testé ni le rôle de l'ARN, ni celui de potentielles structures secondaires dans le recrutement des protéines Polycomb par les R-loops. *In vitro*, l'assemblage de R-loop en présence de KCl, NaCl et LiCl qui permettent ou préviennent la formation de structures secondaires telles que les quadruplex de G, pourrait nous permettre de déterminer si ces structures contribuent au recrutement des complexes Polycomb.

Finalement, *in vitro* nous avons uniquement testé la séquence du PRE au niveau du promoteur de *Antennapedia* qui forme une R-loop sur le brin - et est lié par les protéines du groupe Polycomb *in vivo*. Il est possible que la préférence des protéines du groupe Polycomb pour les R-loops que nous avons observé, soit spécifique aux séquences des PREs ou à une fraction d'entre eux. *In vivo*, les protéines Polycomb liant de nombreux loci en dehors des PREs dont certains forment des R-loops, peut laisser penser que les protéines du groupe Polycomb présentent une préférence pour les R-loops quel que soit leur séquence. Seul des expériences complémentaires nous permettront de comprendre le rôle des R-loops dans le recrutement des protéines du groupe Polycomb.

4.4.1.2 Le recrutement des complexes Polycomb par les R-loops chez les mammifères

Les expériences utilisant la surexpression de la RNase H1 dans des cellules afin de déterminer la contribution des R-loops au recrutement de protéines ou à l'établissement de marques épigénétiques ont donné des résultats mitigés (Skourti-Stathaki et al., 2011, 2014, 2019). Certes Skourti-Stathaki et al. ont conclu que les R-loops participaient au recrutement de PRC1 et PRC2 à certaines de leurs cibles (Skourti-Stathaki et al., 2019). Néanmoins, dans ces articles l'efficacité de la surexpression de la RNase H1 dans les cellules n'a été démontrée qu'à un nombre très limité de loci (un plasmide ainsi que cinq gènes présents dans le génome de leur lignée cellulaire). En temps normal, la RNase H1 n'est que peu recrutée au niveau de la chromatine pour résoudre les R-loops, son rôle est plus important en cas de stress ayant entraîné l'accumulation de R-loops (Lockhart et al., 2019). La surexpression de la RNase H1 conduit donc à la dégradation de R-loops qu'elle ne résout pas en conditions physiologiques, pouvant avoir de nombreuses conséquences sur la cellule en dehors de la perte potentielle du recrutement de la protéine étudiée. De plus, la RNase H1 a besoin d'un hybride ARN-ADN assez long afin que ses deux domaines puissent la R-loop et induire sa dégradation. La résolution de ces R-loops par la RNase H1 sous-entend donc qu'elle est accessible et donc pas liée par d'autres protéines. Finalement, la surexpression de la RNase H1 conduit à une augmentation globale du niveau de transcription, il est donc impossible de déterminer si les effets observés sont dus à la diminution du niveau de R-loops ou à une augmentation de la transcription (Skourti-

Stathaki et al., 2019). Plutôt que de surexprimer la RNase H1 afin de déterminer si une R-loop est impliquée dans le recrutement d'autre protéine *in vivo*, la fusion de celle-ci avec une Cas9 inactive permettant alors le recrutement ciblé de la protéine de fusion grâce à un ARN guide et donc de dégrader seulement la R-loop ciblée semble plus adapté.

4.4.2 D'autres protéines sont recrutées par les R-loops

En plus des protéines qui vont reconnaître les R-loops afin d'induire leur dégradation, d'autres protéines peuvent lier celles-ci et donc potentiellement se servir des R-loops pour être recrutées spécifiquement au niveau de la chromatine.

4.4.2.1 Des protéines recrutées ou liant des R-loops ont des fonctions diverses

Deux expériences de pull down utilisant un hybride ARN-ADN comme appât ont mis au jour que de nombreuses protéines interagissent avec ceux-ci (Cristini et al., 2018; Wang et al., 2018). Parmi les protéines identifiées il n'est pas surprenant de retrouver en premier des protéines impliquées dans la dégradation et résolution des R-loops, mais aussi dans la maturation des ARN. Quelques protéines liant et modifiant la chromatine reconnaissent ces hybrides ARN-ADN, laissant supposer que le recrutement de ces protéines, capables de modifier de manière post-traductionnelle les histones et la structure de la chromatine, pourrait donc se faire via la liaison de la partie hybride d'une R-loop (Cristini et al., 2018). Ce recrutement via une R-loop donnerait la spécificité à des protéines bien souvent dénuées de capacité à lier l'ADN ou la chromatine de manière séquence spécifique (Cristini et al., 2018).

Il est néanmoins à noter que ces deux études présentent des discordances. Parmi le protéome des hybrides ARN-ADN identifié par Cristini et al., figurent des protéines impliquées dans la réparation des cassures d'ADN telles que PARP1 et PARP2, renforçant l'idée qu'au niveau des cassures dans l'ADN la formation d'hybrides pourrait être également impliqué dans le recrutement des protéines qui participent à la signalisation des cassures et à leur réparation. Cependant cette observation n'est pas partagée par Wang et al. qui ont montré que PARP1 est repulsé par les hybrides ARN-ADN mais attiré par l'ADN double-brin. Cette différence peut venir de la sonde utilisée, Cristini et al. ont utilisé les hybrides ARN-ADN purifiés à partir de cellules HeLa, alors que Wang et al., ont eu recours aux hybrides formés aux gènes BAMBI et DPP9 et ont étudié le protéome dans les lymphocytes B.

Les hybrides ARN-ADN et plus généralement les structures d'acides nucléiques non-canoniques pourraient permettre un recrutement préférentiel de certaines protéines en raison de leur structure qui diffère de celle d'un ADN double-brin ou d'un ARN. Pour autant, des différences doivent exister entre deux structures

d'acides nucléiques non-canoniques d'un même type, afin de permettre un recrutement sélectif, et non pas une liaison aveugle de ces structures par toutes les protéines pouvant les lier.

4.4.2.2 Un recrutement indirect qui ne nécessite pas la reconnaissance directe d'une R-loop

Le recrutement de protéines au niveau de séquences ayant formées des R-loops pourrait se faire de manière indirecte sans nécessiter la reconnaissance ni la liaison directe de celle-ci. GADD45A lie des R-loops qui se forment aux promoteurs des îlots CpG. GADD45A interagit également avec TET1, permettant le recrutement de cette dernière et promouvant ainsi la déméthylation au niveau de ces îlots CpG où TET1 a été recrutée (Arab et al., 2019). Cet exemple montre que les R-loops peuvent également permettre le recrutement de protéines qui n'ont pas la capacité de les lier directement grâce à des interactions protéine-protéine. Est-ce que ce genre de recrutement indirect au niveau d'une R-loop pourrait se faire par le biais d'un ARN qui pourrait former un ARN double-brin en s'hybridant avec une partie libre de l'ARN impliqué dans la formation d'une R-loop ? Cette interaction ARN-ARN pourrait permettre le recrutement de protéines à des loci distants du site de production de l'ARN lié. C'est notamment ce qui a été proposé pour PRC2 et HOTAIR (Balas et al., 2020). PRC2 lie HOTAIR, ARN non codant produit au niveau du cluster HoxC, afin d'induire la répression du cluster HoxD. HOTAIR peut s'hybrider avec un ARN produit au niveau de HoxD : JAM2 (Balas et al., 2020). Cette formation d'un ARN double-brin entraînant un changement de conformation de HOTAIR, qui se traduit par une augmentation de l'activité méthyltransférase de PRC2 certainement dû à une diminution de l'affinité pour l'ARN HOTAIR ayant adopté cette nouvelle conformation (Balas et al., 2020). Même s'il n'a pas été montré que JAM2 forme une R-loop, cette interaction entre deux ARN permettant le recrutement spécifique de PRC2 laisse supposer qu'une interaction ARN-ARN puisse également être à l'origine du recrutement de certaines protéines au niveau de R-loops.

4.4.3 Les R-loops peuvent-elles se former *en trans* ?

Dans le chapitre 2, nous avons mis en évidence qu'environ 20% des R-loops se formant chez *D. melanogaster* ne correspondent à aucun transcrit. Cette absence d'ARN peut laisser penser que cela peut être en raison d'un manque d'annotations du génome de *D. melanogaster*, du faible niveau de ces transcrits, en effet un seul ARN suffit à former une R-loop ce qui peut les rendre difficile à identifier, ou encore de la formation de ces R-loops *en trans* avec un ARN provenant d'un autre locus. Cette dernière hypothèse suggère que des facteurs vont promouvoir la formation de ces R-loops. L'accumulation de supertours négatifs entraînant l'ouverture de l'hélice d'ADN peut permettre la formation de telles R-loops. Elles peuvent être également le résultat de protéines induisant l'invasion de brins telles que Rad52, TRF2

ou PRC2 (Lee et al., 2018c; Mazina et al., 2017). *In vitro*, la capacité de Rad52, TRF2 ou de PRC2 à former des hybrides ARN-ADN avec un ARN et un ADN double-brin pas entièrement complémentaire n'a pas été testée. L'immunoprécipitation des hybrides ARN-ADN suivie du séquençage d'à la fois la partie ARN et de la partie ADN pourrait permettre de déterminer le pourcentage de R-loops qui résulteraient potentiellement d'une formation avec un ARN produit en *trans*.

4.5 Les R-loops sont-elles un marque-page du cycle cellulaire ?

4.5.1 L'hypothèse du marque-page mitotique

Lors de la mitose, la compaction des chromosomes oblige l'éviction de nombreuses protéines qui jusqu'alors liaient la chromatine afin d'assurer l'expression appropriée des gènes. Néanmoins, certaines de ces protéines restent liées à la chromatine. Il y a différents moyens pour des protéines de lier les chromosomes mitotiques : grâce à des interactions séquences spécifiques, grâce aux charges négatives présentes à la fois sur l'ADN et les histones en raison la présence de certaines de leur modifications post-traductionnelles telles que la phosphorylation qui permettent la liaison de protéines ayant des domaines chargés positivement, ou en périphérie de la chromatine de façon non-séquence spécifique où sont présents protéines et ARN (Festuccia et al., 2017).

Comme décrit dans l'introduction, une fraction de PSC, sous-unité de PRC1, reste liée à la chromatine pendant la mitose (Follmer et al., 2012). Le complexe PRC2 reste également lié aux chromosomes mitotiques mais l'observation ayant été faite par microscopie et Western Blot il n'est pas possible de savoir si cette rétention se fait comme dans le cas de PRC1 à une fraction ou à tous ses sites de liaison, ou bien à de nouveaux sites qui n'avaient pas été mis en évidence jusqu'à présent en raison du faible taux de cellules en mitose dans une population non synchronisée (Aoto et al., 2008; Follmer et al., 2012). La liaison de PRC1 et de PRC2, ainsi que d'autres protéines aux chromosomes mitotiques qui reconnaissent les R-loops, peut-elle se faire par le biais de R-loops qui persisteraient malgré la compaction de la chromatine ? Alternativement, la présence de R-loop durant la mitose pourrait-elle permettre, lors de la fin de celle-ci, le recrutement rapide de protéines, telles que PRC1 et PRC2, qui les lient préférentiellement (Figure 4-4) ?

4.5.2 Les R-loops durant le cycle cellulaire

4.5.2.1 Les ARN et la transcription pendant la mitose

Contrairement à ce qui a été pensé pendant longtemps, la transcription ne s'arrête pas complètement durant la mitose (Palozola et al., 2017). Un faible niveau de transcription est présent aux gènes impliqués dans la transcription et aux structures extracellulaires ; lors de la sortie de la mitose leur niveau de transcription ainsi que la transcription d'autres gènes augmentent rapidement (Palozola et al., 2017). De plus des ARN restent liés à l'ADN durant les différentes phases de la mitose chez *D. melanogaster* et chez l'humain (L Black et al., 2016; Petruk et al., 2016). Ceci suggère que la transcription et les ARN peuvent être importants lors de la mitose afin de maintenir l'expression génique et peuvent peut-être avoir d'autres fonctions inexplorées en adoptant par exemple des structures d'acides nucléiques non-canoniques.

4.5.2.2 Les R-loops peuvent-elles participer au marque-page mitotique

En mitose, des R-loops restent présentes aux centromères et aux cadres ouverts de lecture en levures (Castellano-Pozo et al., 2013). Chez *C. elegans* et dans des cellules humaines des R-loops ont également été observées en mitose par microscopie (Abakir et al., 2020; Castellano-Pozo et al., 2013). Par qPCR, il a été montré que chez l'humain ces R-loops présentes en mitose sont au niveau de certains gènes mais aussi du rétrotransposon LINE-1 (Abakir et al., 2020). Malheureusement aucune des deux études n'ayant observé la présence de R-loop en mitose n'a eu recours au séquençage afin de déterminer à l'échelle du génome entier les séquences auxquelles elles restent présentes. Il est donc possible que des R-loops soient présentes en dehors des centromères et des quelques cadres de lectures testés par qPCR. La fonction de ces R-loops présentes en mitose reste également à déterminer.

Ces R-loops pourraient se former au niveau des gènes qui restent transcrits en mitose afin de les garder ouverts et de permettre l'augmentation rapide de leur transcription dès la fin de la mitose, dans ce cas-là ces R-loops devraient être rapidement résolues en fin de mitose. À l'inverse, les R-loops présentes au niveau de rétrotransposon pourraient empêcher leur transcription afin d'éviter qu'ils ne soient rétro-transcrits par la suite et ne viennent s'insérer à d'autres loci dans le génome. Les R-loops pourraient également participer directement au marque-page mitotique, soit en permettant la liaison de protéines à ces R-loops durant la mitose, soit en favorisant le recrutement de protéines ayant une affinité pour celles-ci dès la fin de mitose afin de permettre la persistance de l'expression génique (Figure 4-4).

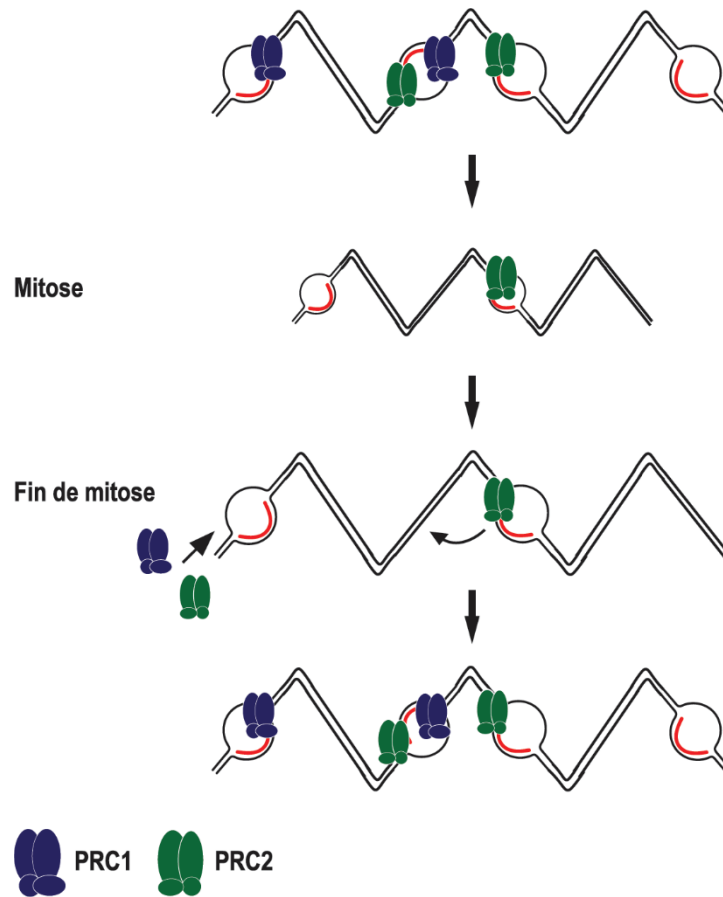


Figure 4-4 Modèle des R-loops participant au marque-page mitotique

La compaction des chromosomes lors de la mitose conduit à l'exclusion de nombreuses protéines et potentiellement de certaines R-loops. Les R-loops restantes peuvent permettre la rétention de protéines pendant la mitose, leur permettant de lier rapidement de nouveaux loci et d'agir dès la fin de celle-ci. D'autres R-loops persistants en mitose mais n'étant pas liées par des protéines, peuvent permettre le recrutement rapide de protéines dès la fin de la mitose si leur liaison est favorisée par rapport à l'ADN double ou la chromatine. Sur le schéma sont présentés les complexes Polycomb, mais d'autres protéines pourraient reposer sur les R-loops afin de permettre leur liaison aux chromosomes mitotiques et/ou de lier leurs cibles rapidement dès la fin de la mitose.

4.6 PRC2 induit l'invasion de brins

4.6.1 Le mécanisme par lequel PRC2 pourrait induire l'invasion de brins

Dans le chapitre 2, nous avons identifié et caractérisé une nouvelle activité d'invasion de brins pour PRC2 qui conduit à la formation d'hybrides ARN-ADN. Cependant, nous n'avons pas déterminé le mécanisme par lequel PRC2 peut induire la formation d'hybrides ARN-ADN. Nous avons montré que PRC2 peut

hybrider un ADN simple-brin avec un ARN complémentaire, qui est la seconde étape de l'invasion de brins mais nous ne savons pas comment PRC2 induit l'ouverture de l'hélice d'ADN. Quatre possibles modèles pouvant expliquer comment PRC2 induit la formation d'hybrides ARN-ADN sont décrit ci-dessous (Figure 4-5).

PRC2 lie l'ADN, induit l'ouverture de la double hélice, et par simple complémentarité de séquence entre l'ARN et l'ADN l'hybride se forme (Figure 4-5a). Dans ce cas-là, PRC2 pourrait également promouvoir l'invasion d'ADN double-brin par un ADN simple-brin, chose que nous n'avons pas testé. Mais puisque nous savons que PRC2 peut hybrider un ARN avec un simple-brin d'ADN, il est donc probable que ce complexe Polycomb participe activement à la formation de l'hybride après ouverture de l'ADN double-brin.

PRC2 peut lier l'ADN et l'ARN, toutefois il a été montré qu'un même complexe ne lie qu'une molécule d'acide nucléique à la fois (Davidovich et al., 2014). Il est donc possible que PRC2 induise l'invasion de brins de manière séquentielle (Figure 4-5b). PRC2 lie l'ADN, ouvre l'hélice avant de promouvoir l'hybridation de l'ARN avec le simple-brin d'ADN disponible. Mais PRC2 peut former un dimère (Davidovich et al., 2014), un complexe PRC2 peut donc lier la partie ADN tandis que l'autre pourrait interagir avec la partie ARN. L'ouverture de la double hélice permettrait à PRC2 de promouvoir l'hybridation entre l'ARN et l'ADN qui sont liés par un même dimère (Figure 4-5c).

PRC2 n'est pas la première protéine pour laquelle il a été montré qu'elle peut induire l'invasion d'ADN double-brin par un ARN. Rad52, protéine impliquée dans la réparation des cassures d'ADN double-brins peut également induire la formation d'hybrides ARN-ADN et nécessite une extrémité d'ADN *in vitro* comme PRC2 (Mazina et al., 2017). L'invasion de brins par Rad52 repose sur la formation d'un filament nucléoprotéique visible par microscopie électronique qui permet l'ouverture de la double hélice d'ADN et l'invasion de l'ADN par un ADN ou ARN complémentaire (Kagawa et al., 2008; Van Dyck et al., 1998). Il est possible que PRC2 forme un filament nucléoprotéique afin de pouvoir induire l'invasion de brins (Figure 4-5d). Le ratio de molécules de PRC2 par paire de bases d'ADN que nous avons utilisé dans le chapitre 2 est proche de celui de Rad 52 : une molécule de PRC2 pour 8 paires de bases, une molécule de Rad52 pour 5 paires de bases (Mazina et al., 2017). Cette invasion de brins par PRC2 via la formation d'un filament nucléoprotéique est donc possible.

La mutation des acides aminés permettant la formation d'un dimère, mais aussi la visualisation de PRC2 lié à l'ADN devraient permettre de mieux comprendre le mécanisme par lequel PRC2 peut induire l'invasion de brins (Chen et al., 2020).

Quel que soit le mécanisme par lequel PRC2 procède à l'invasion de brins, cette nouvelle activité permet de réconcilier deux visions de la relation entre PRC2 et les ARN qui étaient jusqu'alors incompatibles (Yan et al., 2019). Un excès d'ARN, tout comme la liaison d'un ARN avant l'ADN empêche la liaison de PRC2 à l'ADN et la chromatine, inhibant ses activités méthyltransférase et d'invasion de brin. À l'inverse, moins d'ARN ou une affinité plus faible de PRC2 pour ces ARN, peut lui permettre d'interagir avec d'autres acides nucléiques lui donnant l'opportunité d'induire l'invasion de brins mais aussi de méthyler les histones.

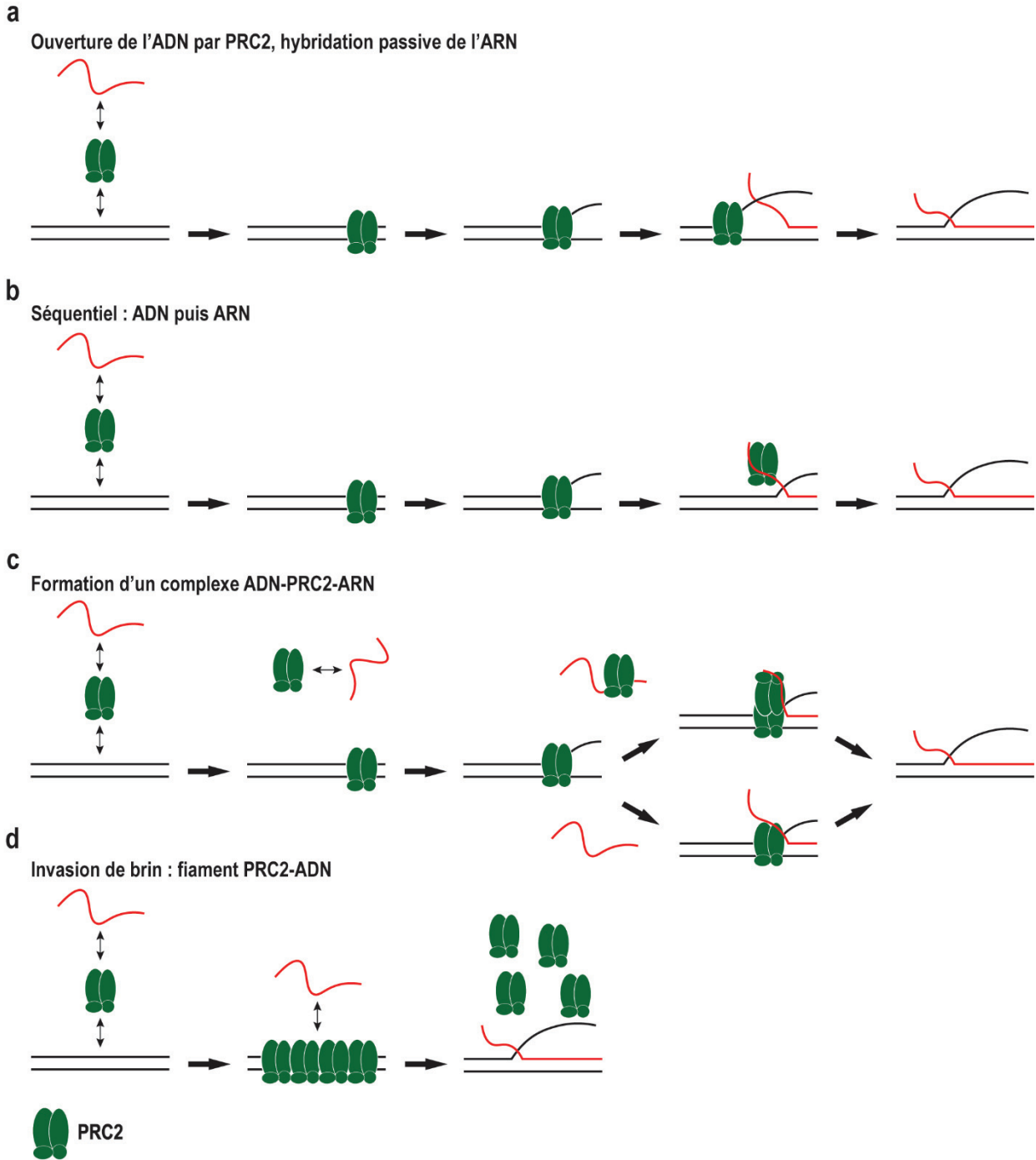


Figure 4-5 Schéma représentant des mécanismes par lesquels PRC2 peut induire la formation d'hybride ARN-ADN

a. La liaison de PRC2 conduit à l'ouverture de l'ADN double-brin permettant l'hybridation d'un ARN complémentaire de façon passive. b. PRC2 lie et ouvre l'ADN double-brin, puis interagit avec l'ARN et induit l'hybridation entre l'ARN lié et l'ADN simple-brin. La liaison de l'ARN par PRC2 peut également participer au déplacement de ce dernier de l'ADN. c. PRC2 lie et ouvre l'ADN double-brin. Le même complexe (bas) ou un second complexe PRC2 formant ainsi un dimère (haut) interagit avec la partie ARN en même temps que

la partie ADN, conduisant à la formation d'un hybride ARN-ADN. d. PRC2 forme un filament nucléoprotéique avec l'ADN afin d'induire la formation de l'hybride ARN-ADN par un mécanisme similaire à celui de Rad52.

4.6.2 PRC2 nécessite un facteur manquant pour induire la formation de R-loops lorsqu'une extrémité d'ADN n'est pas présente

Dans le chapitre 2, nous avons montré qu'*in vitro* PRC2 nécessite une extrémité d'ADN afin de pouvoir induire la formation d'hybrides ARN-ADN. Le fait que nous n'ayons pas été capables d'observer la formation d'hybride sur un plasmide circulaire qu'il soit surenroulé négativement ou relaxé ne signifie pas forcément que *in vivo*, l'activité d'invasion de brins de PRC2 soit restreinte aux sites présentant une extrémité d'ADN. En effet, il est possible que PRC2 ait besoin de protéines accessoires ou d'interagir avec d'autres protéines afin d'induire l'invasion de brins.

De plus, parmi les protéines capables d'induire la formation d'hybrides ARN-ADN, seules deux protéines ont été montrées comme pouvant promouvoir l'invasion d'ADN double-brin par un ARN sur un plasmide circulaire : TRF2 et ICP8. Pour TRF2, le plasmide utilisé est particulier, en effet il comporte des répétitions télomériques qui vont avoir tendance à induire l'ouverture de la double hélice d'ADN en absence de protéines (Amiard et al., 2007; Lee et al., 2018c). C'est pour cela qu'avec ce plasmide particulier, la présence de TRF2 n'est pas nécessaire pour observer la formation de D-loop ou de R-loop, mais TRF2 induit une augmentation de l'apparition de ces produits (Amiard et al., 2007; Lee et al., 2018c). Dans le cas de ICP8, la présence d'une D-loop sur le plasmide est nécessaire afin de pouvoir observer l'invasion de brins (Boehmer, 2004; Nimonkar and Boehmer, 2003). Le fait que *in vitro* PRC2 nécessite une extrémité d'ADN ne signifie donc pas obligatoirement c'est aussi le cas que *in vivo*.

4.6.2.1 Les protéines accessoires du complexe PRC2

Le complexe PRC2 chez *D. melanogaster*, les mammifères ou les plantes possèdent des sous-unités accessoires qui s'associent avec le cœur du complexe. Ces sous-unités accessoires participent à la régulation de l'activité de PRC2 en stimulant son activité méthyltransférase et en permettant sa liaison à des histones modifiées de manière post-traductionnelle (voir la section 1.2.2.4). Ces protéines accessoires lient également les ARN (Zhang et al., 2019a). Les protéines PHF19 et MTF2, homologues de la protéine Pcl de *D. melanogaster*, stabilisent la formation d'un dimère (Chen et al., 2020). Si la formation d'un dimère de PRC2 est importante pour l'invasion de brins, ces protéines accessoires pourraient donc moduler cette activité. Ces sous-unités accessoires pourraient permettre la stimulation de l'activité d'invasion de brins

de PRC2 et/ou lui permettre d'induire l'invasion d'un ADN double-brin sans nécessiter la présence d'une extrémité.

4.6.2.2 Les protéines interagissant avec PRC2

Pour Rad52, la formation de D-loop sur un plasmide circulaire se fait par le changement de la topologie de ce plasmide (Kagawa et al., 2008). Est-il possible que PRC1 et PRC2 agissent ensemble afin de promouvoir l'invasion de brins ? Ces complexes pourraient interagir ensemble via la liaison de la sous-unité accessoire Scm ou en liant le même plasmide/domaine de l'ADN (Kang et al., 2015). PRC1 pourrait alors modifier la topologie de l'ADN grâce à son activité de topoisomérase permettant alors à PRC2 de promouvoir la formation d'hybride ARN-ADN.

4.6.2.3 Autres

La transcription pourrait faciliter la formation de R-loop par PRC2 en raison de l'accumulation de supertours négatifs derrière la polymérase qui s'ils ne sont pas résolus, vont entraîner l'ouverture de l'hélice d'ADN afin de réduire le stress topologique (Naughton et al., 2013). De la même façon la réplication de l'ADN qui nécessite l'ouverture de l'ADN et entraîne l'accumulation de supertours, pourraient également permettre la formation de R-loops par PRC2 en absence d'extrémité d'ADN. Plus généralement, tout évènement pouvant conduire à l'ouverture l'hélice d'ADN, que ce soit la transcription ou la réplication via la modification de la topologie de l'ADN ou par l'action d'hélicases, peut permettre à PRC2 d'induire la formation d'hybride ARN-ADN. La présence de RPA peut également faciliter l'activité d'invasion de brins de PRC2 en liant le simple-brin d'ADN déplacé et en l'empêchant de rentrer en compétition avec l'ARN lors de la formation d'un hybride ARN-ADN (Mazina et al., 2017). RPA peut aussi favoriser l'invasion de brins par PRC2 en induisant la formation initiale d'une R-loop sur un plasmide surenroulé que PRC2 pourrait agrandir ou coopérer avec RPA afin d'en favoriser la formation (Mazina et al., 2020)

4.6.3 L'activité d'invasion de brins de PRC2 est-elle conservée ?

Une étude dans des cellules souches de souris a mis en évidence que la déplétion de EZH2 n'affecte pas le niveau de R-loops aux quelques loci testés par DRIP-qPCR *in vivo*, et les auteurs en ont conclu que PRC2 ne joue aucun rôle dans la formation et la résolution de R-loops (Skourti-Stathaki et al., 2019). Hors, EZH1 et EZH2 ont des fonctions partiellement redondantes et la déplétion de EZH2 n'a que peu d'effet dans les lignées cellulaires car EZH1 compense pour sa perte (Shen et al., 2008). Skourti-Stathaki et al. ont également utilisé l'inhibiteur unc1999, qui inhibe l'activité méthyltransférase de EZH2 et EZH1 en ciblant le domaine SET de ces protéines. Cependant cet inhibiteur n'affecte pas la formation du complexe PRC2

(Arbuckle et al., 2017). Donc si l'activité d'invasion de brins de PRC2 est distincte de son activité méthyltransférase, il n'est pas surprenant que les auteurs n'aient observé aucun effet sur la formation de R-loops lors de l'utilisation de celui-ci. Il est donc encore possible que chez les mammifères, les complexes PRC2 contenant EZH2 ou EZH1 soient tous deux capables d'induire la formation d'hybrides ARN-ADN, ou que cette activité soit présente uniquement chez EZH1. Mais si elle est conservée, que cette activité d'invasion de brins soit présente dans une seule version du complexe PRC2 ou dans les deux, l'utilisation de l'inhibiteur a révélé qu'elle est indépendante de son activité méthyltransférase.

Contrairement à EZH2, EZH1 est exprimé dans toutes les cellules, et peut compacter la chromatine (Margueron et al., 2008). Est-il possible que le complexe PRC2 contenant EZH1 soit également le seul capable d'induire l'invasion de brins ? Si tel est le cas, le complexe PRC2 contenant EZH1 a-t-il d'autres fonctions qui le distinguent de celui contenant EZH2 qui ont été ignorées en raison de l'intérêt qui s'est porté sur EZH2 principalement.

4.7 Quelle pourrait-être la fonction de l'activité d'invasion de brins de PRC2 *in vivo* ?

4.7.1 PRC2 et les hybrides ARN-ADN aux sites de cassures dans l'ADN

4.7.1.1 Les fonctions hybrides ARN-ADN dans la réparation des cassures dans l'ADN

Comme décrit précédemment dans l'introduction, les hybrides ARN-ADN interviennent dans la réparation de cassures doubles brins, que ce soit en maintenant les deux extrémités ensemble afin d'éviter la recombinaison due à la ligation de deux extrémités, comme matrice sous la forme d'hybride ou après transcription inverse afin de permettre la restauration de l'information perdue, ou encore afin de contrôler la résection des brins. Les hybrides ARN-ADN pourraient également servir au recrutement de protéines impliquées dans la signalisation et/ou la réparation des cassures d'ADN.

4.7.1.2 Les protéines du groupe Polycomb sont recrutées aux sites de cassures double-brin

À la suite de cassures dans l'ADN, les complexes Polycomb PRC2 et PRC1 sont recrutés à ces sites de dommages quelques secondes après leur apparition (Campbell et al., 2013; Ismail et al., 2010). Le recrutement de PRC1 conduit à l'ubiquitination des histones H2A, mais celui de PRC2 n'induit pas la méthylation de la lysine 27 de l'histone 3 (Campbell et al., 2013; Ginjala et al., 2011; Ismail et al., 2010).

De plus, PRC2 n'est pas nécessaire au recrutement de PRC1 à ces cassures (Ismail et al., 2010). Pour autant, la déplétion de PRC1 ou PRC2 conduit à une diminution de l'efficacité de la réparation de ces cassures (Campbell et al., 2013; Ginjala et al., 2011; Ismail et al., 2010). Si PRC2 n'est pas présent aux sites de cassures afin de promouvoir la tri-méthylation de la lysine 27 de l'histone 3, ni dans le but de permettre le recrutement de PRC1, on peut se poser la question sur la fonction de PRC2 à ces sites de dommages dans l'ADN. PRC2 est-il présent aux sites de cassures afin de permettre la formation d'hybrides ARN-ADN? Dans le chapitre 2, *in vitro* nous avons démontré que PRC2 pouvait induire la formation d'hybrides ARN-ADN en présence d'une extrémité d'ADN et de son ARN complémentaire. PRC2 pourrait donc favoriser la formation d'hybride formé de manière co-transcriptionnelle lorsqu'une ARN polymérase lie et transcrit l'ADN à partir de l'extrémité, mais aussi favoriser la formation d'un hybride entre les deux fragments d'ADN et l'ARN produit avant l'apparition de la cassure. Ainsi grâce à son activité d'invasion de brins PRC2 pourrait prendre part à la signalisation de cassures dans l'ADN en participant au recrutement des protéines PARP, et/ou dans leur réparation comme le contrôle de la résection des brins par un hybride ARN-ADN le suggère (Cristini et al., 2018; Keskin et al., 2014; Ohle et al., 2016).

4.7.2 L'activité d'invasion de brins de PRC2 peut-elle participer au maintien de la transcription ?

Lors de la transcription, la double hélice d'ADN est ouverte par les hélicases et l'ARN polymérase. La transcription conduit à l'accumulation de surenroulement négatif derrière la polymérase, surenroulement qui s'il n'est pas résolu peut entraîner l'ouverture de la double hélice d'ADN afin d'aider à dissiper le stress topologique, pouvant ainsi favoriser la formation de R-loops. Est-ce que PRC2 pourrait profiter de l'ouverture de l'ADN afin de promouvoir la formation d'un hybride ARN-ADN? Dans le chapitre 2 nous n'avons pas observé un fort niveau de R-loops au niveau du site d'initiation de transcription mais un niveau constant de R-loop le long des gènes. Le complexe PRC2 murin interagit avec les facteurs d'élongation Elongin B et C via sa sous-unité accessoire EPOP, cette interaction permet le maintien d'un faible niveau d'expression pour certaines cibles de PRC2 (Beringer et al., 2016; Liefke et al., 2016). PRC2 peut méthyler un autre facteur d'élongation, Elongin A (Ardehali et al., 2017). La mutation de la lysine de Elongin A qui est la cible de PRC2, et la déplétion de Elongin B ou C affectent l'expression de nombreuses cibles de PRC2 (Ardehali et al., 2017; Beringer et al., 2016).

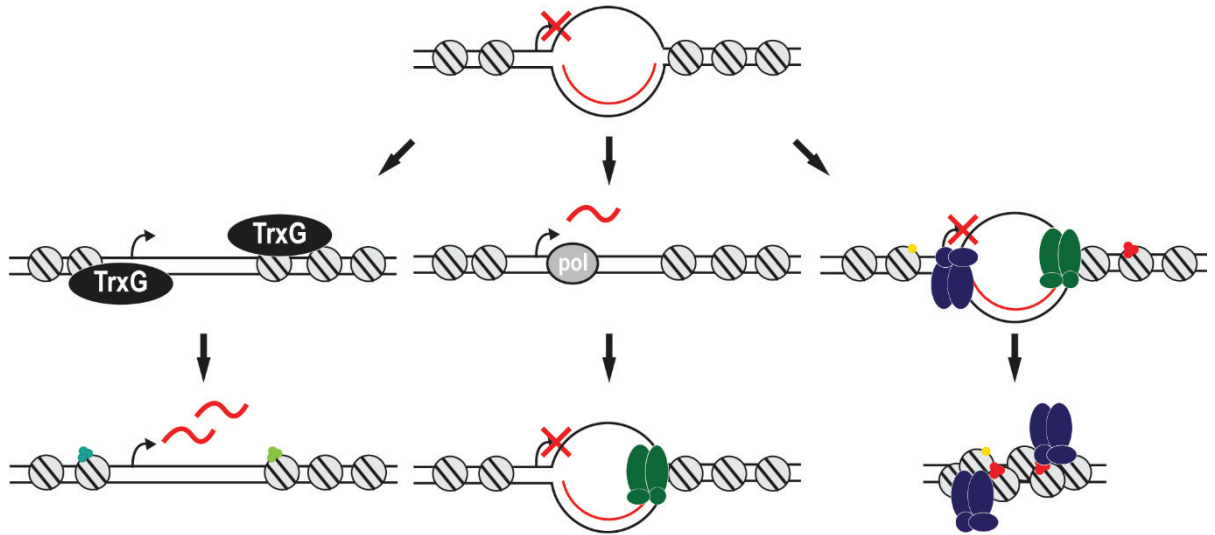
L'interaction de PRC2 avec les protéines Elongin B et C ne conduit pas à la répression, mais permet le maintien d'un faible niveau d'expression soulevant la possibilité que cette interaction puisse permettre une autre activité de PRC2 qui n'est pas la répression comme par exemple la formation d'un hybride ARN-

ADN. En formant un hybride ARN-ADN, PRC2 conduirait à la réduction de la transcription en empêchant le passage de l'ARN polymérase mais aussi en prévenant l'établissement d'un fort niveau de H3K27me3 en raison de l'absence de nucléosome au niveau de la R-loop. PRC2 seul, en coopération avec d'autres protéines, ou encore lors d'un contexte topologique favorable pourrait induire la formation de R-loop le long de gènes afin de diminuer leur niveau d'expression dans le but de maintenir un niveau d'expression minimum (Figure 4-6a milieu).

La formation de R-loop pourrait aussi conduire au recrutement des complexes Polycomb PRC1 et PRC2 et à la répression génique (Figure 4-6 à droite). À l'inverse, la résolution de cette R-loop pourrait permettre le recrutement des protéines du groupe Trithorax qui pourraient reconnaître leur site de liaison dans cette zone déplétée en nucléosomes (Figure 4-6 à gauche). L'augmentation rapide de la transcription au niveau de ce gène pourrait être également facilitée par l'absence de nucléosomes dans cette région à la suite de la résolution de la R-loop. Dans ce cas, la formation de R-loop par PRC2 ainsi que sa résolution pourrait permettre le changement rapide du niveau de transcription d'un gène.

a

Maintien de la transcription ou changement du statut transcriptionnel d'un gène à la suite de la formation d'une R-loop par PRC2



b

Détection d'un gène échappant à la répression

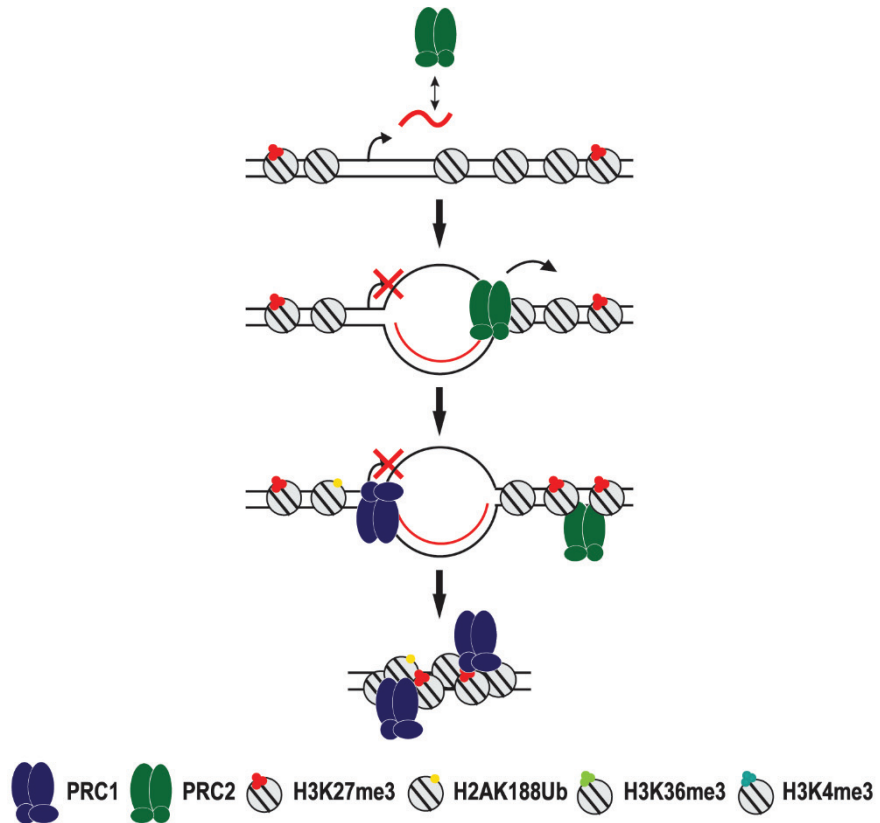


Figure 4-6 Deux exemples possibles de la fonction de l'activité d'invasion de brins de PRC2

a. La présence d'une R-loop ayant été formée par PRC2 conduit au recrutement des complexes Polycomb et donc à la répression (droite). La résolution de la R-loop peut permettre la transcription par l'ARN polymérase donnant l'opportunité à PRC2 de former une nouvelle R-loop (milieu). La résolution de cette R-loop, libère une région sans nucléosome pouvant permettre la liaison de protéines reconnaissant un motif dans l'ADN et/ou le recrutement des protéines du groupe Trithorax conduisant ainsi à l'expression (gauche). b. La transcription d'un gène échappant à la répression par les complexes Polycomb permet à PRC2 d'induire la formation d'une R-loop avec l'ARN produit. Cette R-loop participe au recrutement d'autres complexes Polycomb pouvant modifier la chromatine et induire sa compaction afin de restaurer la répression de la transcription.

4.7.3 L'activité d'invasion de brins de PRC2 est-il un mécanisme de surveillance permettant la détection de gènes qui échappent à la répression?

Les PREs et autres sites de liaison des protéines du groupe Polycomb peuvent être transcrits en ARN, et ce même lorsqu'ils sont liés par les protéines du groupe Polycomb et présentent un fort niveau de H3K27me3 (Herzog et al., 2014). Ces ARN produits à leurs sites de liaison pourraient avoir deux fonctions dépendamment du niveau de transcription. La production d'un haut niveau de transcrits pourrait prévenir le recrutement de PRC2. L'affinité de PRC2 pour les ARN étant plus importante que pour la chromatine ou l'ADN, ces ARN entreraient alors en compétition avec la chromatine empêchant le recrutement de PRC2 (Beltran et al., 2016; Wang et al., 2017a). Dans ce cas-là les ARN auraient un rôle inhibiteur pour PRC2. Mais la transcription moins importante de ces mêmes loci et donc la plus faible production d'ARN, pourrait alors permettre à PRC2 de lier l'ARN et l'ADN, et d'induire la formation de R-loop grâce à son activité d'invasion de brins. Les R-loops, parce que nous avons observé qu'elles sont liées préférentiellement par PRC1 et PRC2, pourraient alors promouvoir le recrutement des complexes Polycomb et la répression (Figure 4-6b).

La formation de R-loop, grâce à l'activité d'invasion de brins de PRC2, pourrait donc servir de mécanisme de surveillance. En effet si un PRE ou un gène échappe à la répression par les protéines du groupe Polycomb, la transcription va être lente avant de potentiellement augmenter. Cette phase de faible transcription pourrait être favorable à la formation de R-loop par PRC2, comme décrit précédemment. Cette R-loop en entraînant un recrutement plus important d'autres complexes PRC2 ainsi que de PRC1, permettrait ainsi le rétablissement de la répression à laquelle ce locus avait échappé.

L'utilisation d'un gène rapporteur sous le contrôle d'un PRE lui-même associé avec différents promoteurs permettant de moduler le niveau de transcription pourrait permettre de déterminer si le niveau de transcription est plus important que l'orientation de la transcription pour la régulation de l'activité des

protéines du groupe Polycomb, contrairement à ce qui avait été proposé par Herzog *et al* (Herzog et al., 2014).

4.8 L'activité de topoisomérase de PSC

Dans le chapitre 3 nous avons décrit une activité similaire à celle de la topoisomérase I associée au complexe Polycomb PRC1 ainsi qu'à la région C-terminale de sa sous-unité PSC (PSC-CTR). PRC1 et PSC-CTR peuvent relaxer un plasmide surenroulé négativement et ajouter des supertours à un plasmide relaxé *in vitro*. Les expériences présentées dans le chapitre 3 ne nous ont pas permis d'identifier d'élucider entièrement le mécanisme par lequel ces protéines modifient la topologie de l'ADN, mais elle repose vraisemblablement sur deux activités distinctes qui font l'objet de cette section.

4.8.1 Le mécanisme par lequel PSC-CTR modifie la topologie de l'ADN

4.8.1.1 Le mécanisme par lequel PSC-CTR modifie la topologie de l'ADN

Comme présenté dans la discussion du chapitre 3, le mécanisme que nous avons proposé par lequel PSC relaxe de manière incomplète un ADN surenroulé et ajoute des supertours à un ADN relaxé repose sur deux activités distinctes. L'addition d'un faible nombre de supertours, ainsi que la persistance de quelques supertours repose sur la capacité de PRC1 et de PSC-CTR à enrouler l'ADN, tandis que le changement du nombre de surenroulement sur la présence d'une activité similaire à celle des topoisomérases, qui est soit intrinsèque à PSC comme le suggère les expériences du chapitre 3 ou présente chez une protéine contaminante qui co-purifie avec PSC.

La capacité de PRC1 à enrouler l'ADN autour de lui-même n'est pas une surprise puisque cela avait été montré auparavant en utilisant la microscopie à force atomique (Mohd-Sarip et al., 2006). De plus, sur un plasmide contenant une coupure d'ADN simple-brin, PRC1 accompagné de la protéine Pho, peuvent enrouler l'ADN ce qui crée un supertour visible sur gel après la ligation de l'ADN en présence des protéines du groupe Polycomb (Mohd-Sarip et al., 2006). En enroulant l'ADN autour d'eux PRC1 et PSC-CTR introduisent donc un supertour toroïdal, sans pour autant changer le nombre de supertours présent dans cette molécule d'ADN, mais qui pourra être relaxé par l'activité de topoisomérase présente dans les préparations de ces protéines.

Les expériences du chapitre 3 ont montré que PSC-CTR possède une activité similaire à celle de la topoisomérase I. Si les expériences présentées ne nous ont pas permis de séparer biochimiquement l'activité de topoisomérase de la protéine PSC, ni d'identifier de topoisomérase contaminante par spectrométrie de masse, nous ne sommes pas certains que cette activité soit intrinsèque à PSC. En assumant que l'activité de topoisomérase de PRC1 et PSC-CTR soit intrinsèque à ces protéines, elle fait l'objet des sections ci-dessous.

4.8.1.2 PSC-CTR forme-t-il un complexe covalent avec l'ADN ?

Après avoir coupé l'ADN, les topoisomérases forment un complexe covalent avec celui-ci avant de réduire le nombre de supertours présents dans l'ADN et de le liguer (Champoux, 2001). Cependant, nous n'avons pas été capables de détecter la formation d'un complexe covalent entre PSC-CTR et l'ADN. Ceci peut être dû à la faible proportion de protéines formant un complexe covalent avec l'ADN à un moment donné et donc rendant leur détection difficile. Alternativement, le changement de nombre de supertours par PSC-CTR et PRC1, pourrait ne pas nécessiter la formation d'un complexe covalent contrairement à ce qui a été montré pour les topoisomérases ainsi que les recombinaisons à site spécifique (Champoux, 2001; Grindley et al., 2006). Que le changement du niveau de surenroulement implique ou n'implique pas la formation d'un complexe covalent, la coupure d'au minimum un brin d'ADN est nécessaire. Des expériences supplémentaires seront nécessaires afin d'identifier la présence d'une extrémité d'ADN durant la modification de la topologie de l'ADN par la protéine PSC-CTR. Afin d'augmenter la sensibilité de nos expériences, l'utilisation de radioactivité apparaît nécessaire afin de pouvoir détecter la présence d'une extrémité d'ADN et/ou la formation d'un complexe covalent, qui sont deux produits intermédiaires et dont la présence est transitoire lors de la relaxation de l'ADN.

En assumant que PSC-CTR change la topologie d'un plasmide de manière similaire aux topoisomérases, en coupant l'ADN, puis en formant complexe covalent avec ce dernier avant de le religuer, des questions concernant cette activité de topoisomérase sont discutées dans les sections suivantes.

4.8.1.3 PSC-CTR n'utilise pas une tyrosine comme résidu catalytique

Les topoisomérases utilisent une tyrosine lors de l'attaque nucléophile de l'ADN et pour la formation d'un complexe covalent (Champoux, 2001), mais la présence de tyrosine au sein de la protéine PSC n'est pas nécessaire pour son activité de topoisomérase. D'autres protéines peuvent modifier la topologie de l'ADN sans reposer sur une tyrosine : les sérines recombinaisons (Olorunniji et al., 2016). Il reste toutefois à déterminer si PSC-CTR utilise une sérine pour procéder à l'attaque nucléophile ou si elle repose sur un autre résidu n'ayant pas été décrit jusqu'à présent.

Si PSC-CTR utilise une sérine comme certaines recombinaisons, est-ce que la modification de la topologie de l'ADN se fait de la même façon ? Les sérines recombinaisons induisent la recombinaison entre deux sites spécifiques conduisant à la délétion, l'insertion, ou à la recombinaison pouvant ainsi entraîner la modification de la topologie. Dans le chapitre 3, la caractérisation de l'activité de topoisomérase de PRC1 et de PSC-CTR a été faite à l'aide de deux plasmides (pUC19 et pET-blue1 au sein duquel a été inséré la séquence du PRE *bx_d*) qui ne contiennent pas de sites de recombinaison. PSC-CTR ne nécessite pas la présence de sites spécifiques afin de modifier la topologie d'un plasmide, son mécanisme d'action apparaît donc différent de celui des recombinaisons à site spécifique.

4.8.1.4 La structure de PSC pourrait nous permettre de déterminer le mécanisme par lequel elle modifie la topologie de l'ADN

Les topoisomérases réduisent le niveau de surenroulement d'un plasmide en coupant un ou deux brins d'ADN avant de modifier le nombre de supertours et de religuer les brins coupés. Malgré ce mécanisme général similaire, des différences existent notamment au niveau du nombre de brins d'ADN coupés et de la façon dont se fait la réduction du nombre de supertours (Champoux, 2001). Les structures des topoisomérases diffèrent également. La topoisomérase de type I forme une pince qui va sectionner un seul brin d'ADN et réduire le nombre de supertours présents en faisant tourner l'ADN. La topoisomérase de type II agit sous forme de dimère qui lie les deux brins d'ADN qu'elle coupe, sa structure présente une cavité centrale au sein de laquelle elle fait passer l'ADN intact à travers cette cassure qu'elle a générée. La topoisomérase de type III coupe un seul brin et adopte une structure qui possède une cavité dans laquelle elle peut faire passer le brin intact à travers celui qu'elle a coupé. La structure de la topoisomérase reflète donc son mécanisme : la présence d'une cavité permet de réduire le surenroulement par passage de brins, la formation d'un dimère ou tétramères pour les recombinaisons permet la coupure de plus d'un brin d'ADN.

La structure du complexe PRC1 et celle de PSC ne sont pas connues. PSC est une protéine intrinsèquement désordonnée pouvant interagir avec elle-même (Lo and Francis, 2010). L'absence de structure disponible et le fait que PSC puisse former un dimère soulève plusieurs questions quant au mécanisme par lequel PSC modifie la topologie. Est-ce que PSC modifie la topologie de l'ADN de manière similaire aux topoisomérases, par rotation ou par passage de brin ? L'activité de topoisomérase de PSC nécessite-t-elle la formation d'un dimère ? PSC coupe-t-elle un ou deux brins d'ADN, et le nombre de brins coupés dépend-il de la dimérisation de PSC ?

4.8.1.5 Quel domaine de PSC est responsable de son activité de topoisomérase ?

Malgré l'absence de structure de PSC, de nombreuses études fonctionnelles ont permis de déterminer quels sont les domaines de PSC nécessaires à ses différentes activités *in vitro*. Les différentes tronctions utilisées dans le chapitre 3 ont mis en évidence que toutes pouvaient modifier la topologie d'un plasmide mais que les plus petites tronctions ont une activité bien plus faible. Cette diminution d'activité pourrait être en partie expliquée par la délétion d'un domaine impliqué dans la liaison de l'ADN, la plus faible affinité de ces tronctions pour l'ADN de ces tronctions, mais aussi la capacité à former un dimère qui est abolie pour la tronction PSC⁴⁵⁶⁻⁹⁰⁹ et la possible délétion d'acides aminés responsables de la stimulation de son activité de topoisomérase (Kang et al., 2020; Lo et al., 2012a). Tout cela pouvant à la fois affecter la capacité de PSC à enrouler l'ADN et son activité de topoisomérase. L'activité de topoisomérase de PSC-CTR et notamment son site catalytique semblent donc être compris entre les acides aminés 656 et 759, même si nous n'avons pas directement testé cette tronction. L'utilisation de cette forme minimale de PSC, si elle est active, pourrait faciliter l'identification du site catalytique de cette dernière et du mécanisme par lequel PSC réduit le nombre de supertours présents dans l'ADN.

4.8.2 L'activité de topoisomérase de PSC s'exerce-t-elle au sein du complexe PRC1 ou indépendamment de celui-ci ?

Dans le chapitre 3, nous avons mis en évidence que PRC1 est environ 100 fois plus efficace que PSC-CTR pour induire un changement de la topologie de l'ADN. PRC1 est un complexe de 431 kDa et PSC-CTR une protéine faisant 120 kDa. Si un seul complexe PRC1 peut enrouler l'ADN et introduire un supertour, PSC-CTR qui est environ 4 fois plus petit, pourrait nécessiter la formation d'un oligomère afin de produire le même changement topologique que PRC1. Il a été montré que l'interaction de PSC-CTR avec lui-même est également stimulé par la présence d'ADN (Lo et al., 2012a). Ceci pouvant donc contribuer à la différence de ratios de protéines par plasmide observés entre PRC1 et PSC-CTR dans le chapitre 3.

Si *in vitro*, PRC1 est environ 100 fois plus efficace que PSC-CTR pour modifier la topologie de l'ADN, cela ne signifie pas pour autant que PSC ne puisse pas exercer cette activité similaire à celle de la topoisomérase I indépendamment de PRC1. *In vivo*, la présence de PSC avec son domaine N-terminal, mais aussi de possible interaction avec d'autres protéines pourraient contribuer à stimuler l'activité de topoisomérase de PSC-CTR.

PSC fait partie du complexe PRC1 au sein duquel les différentes sous-unités sont présentes en quantité stœchiométrique (Francis et al., 2004). Cependant une analyse du protéome de *D. melanogaster* a montré

que durant le développement embryonnaire la protéine PSC est présente en quantité plus abondante que les autres sous-unités de PRC1 et notamment la protéine dRING avec laquelle PSC s'associe afin de former le complexe dRAF (Casas-Vila et al., 2017; Lagarou et al., 2008; Wang et al., 2004a). Cette surabondance relative de PSC ouvre la possibilité à une fonction pour celle-ci en dehors des complexes Polycomb PRC1 et dRAF (Li et al., 2010). La surexpression de PSC pourrait peut-être compenser pour la perte de fonction d'une topoisomérase en constituant un nouvel ensemble de protéines capables de modifier la topologie de l'ADN. Si *in vitro* PSC-CTR et PRC1 peuvent tous deux modifier la topologie d'un plasmide, il reste à déterminer si *in vivo*, cette activité de topoisomérase est exercée par les complexes PRC1, dRAF ou encore par la protéine PSC seule.

4.8.3 L'activité de topoisomérase de PRC1 et PSC est-elle conservée ?

4.8.3.1 Chez *Drosophila melanogaster*

Chez *D. melanogaster*, PSC possède un homologue Su(Z)2 qui possède des activités similaires à PSC mais est exprimé plus tardivement, en fin de stade embryonnaire et au stade larvaire lorsque l'expression de PSC diminue (Lo et al., 2009; Wu and Howe, 1995). L'activité de topoisomérase des protéines du groupe Polycomb pourrait donc ne pas être seulement restreinte aux stades de développement où la protéine PSC est exprimée.

4.8.3.2 Chez les mammifères

L'une des premières activités caractérisées pour la protéine PSC est sa capacité à compacter la chromatine (Francis et al., 2004). Or chez les mammifères cette activité n'est pas présente chez les protéines homologues de PSC mais chez les homologues de Pc (Grau et al., 2011). Si cette activité de topoisomérase est conservée chez les mammifères est-elle présente chez les protéines homologues de PSC ou celles de Pc?

Cbx2, un des homologues de Pc capable de compacter la chromatine, comporte de nombreux résidus chargés positivement ou négativement. Ces résidus chargés positivement sont essentiels pour sa capacité à compacter la chromatine alors que ceux chargés négativement ne sont pas indispensables (Grau et al., 2011). Ces mêmes acides aminés sont-ils importants pour l'activité de topoisomérase de Cbx2 si elle est présente chez cette protéine?

4.8.3.3 Chez les plantes

Chez les plantes, PSC et Pc possèdent chacun un seul homologue, EMF1 et LHP1 respectivement. Si cette activité de topoisomérase est conservée, est-elle présente chez l'homologue de PSC ou celui de Pc? *In vitro* la relaxation d'un plasmide par PRC1 nécessite moins de protéine par rapport à PSC seul. Cette différence d'efficacité pourrait être dû à la différence d'affinité pour l'ADN mais aussi à la présence des autres sous-unités de PRC1 qui pourraient stimuler l'activité de topoisomérase de PSC. Est-il possible que l'absence d'homologue de Ph rende le complexe PRC1 d'*Arabidopsis thaliana* moins efficace pour la relaxation d'ADN? De plus, la multiplicité des homologues de dRING pourrait également influencer l'activité de topoisomérase des différents complexes PRC1.

4.9 Quelles pourraient être les fonctions de l'activité de topoisomérase de PSC *in vivo* ?

4.9.1 La protéine PSC pourrait-elle compenser la perte d'une topoisomérase ?

Chez les eucaryotes supérieurs la perte de la topoisomérase I, II ou III est létale (Lee et al., 1993; Morham et al., 1996; Plank et al., 2005). La présence de PSC, protéine possédant une activité similaire à celles des topoisomérases ne suffit donc pas à compenser leur perte *in vivo*. Mais chez la levure ces protéines ne sont pas indispensables mais leur absence entraîne tout de même un ralentissement de la croissance. L'expression de PSC chez les levures pourrait nous permettre de déterminer si PSC peut compenser la perte de topoisomérase chez la levure, et donc si *in vivo* PSC a bien une activité de topoisomérase mais dont l'activité pourrait être masquée par celles des topoisomérases.

Les différents mécanismes utilisés par les topoisomérases de type II et III leur permettent, en plus de diminuer le niveau de surenroulement présent dans les cellules, de résoudre des caténats et héli-caténats (Champoux, 2001). Une meilleure connaissance du mécanisme par lequel PSC et PRC1 modifient la topologie de l'ADN pourrait mettre en évidence s'ils peuvent également participer à la résolution de ces caténats qui sont notamment présents sur les chromosomes mitotiques où une fraction de PRC1 restent liés (Buchenau et al., 1998; Follmer et al., 2012).

PRC1 ne lie pas exclusivement l'ADN, le complexe humain peut lier des ARN *in vivo* (Rosenberg et al., 2017) et *in vitro* nous avons observé que c'était également le cas du complexe PRC1 de *D. melanogaster*. Est-il possible que PRC1, tout comme la topoisomérase III, puisse modifier la topologie d'un ARN ?

4.9.2 PSC stimule l'activité de la topoisomérase I

Une version de PSC⁴⁵⁶⁻⁹⁰⁹ purifiée à partir de bactérie et donc dénuée de modification post-traductionnelle, est incapable de modifier la topologie l'ADN (Figure 4-7b), montrant l'importance des modifications post-traductionnelles dans la capacité à modifier la topologie de l'ADN de PSC. Cette version inactive de PSC⁴⁵⁶⁻⁹⁰⁹ peut néanmoins stimuler l'activité de la topoisomérase de type I issu de thymus de veau (Figure 4-7a-c). Cette stimulation pourrait se faire par une interaction directe entre PSC et la topoisomérase, mais aucune interaction entre PSC et la topoisomérase I n'a été rapporté. Par contre, une autre protéine du complexe PRC1, Ph, interagit avec la topoisomérase de type II (Lupo et al., 2001), mais la fonction de cette interaction n'a pas été explorée. Le fait qu'une interaction n'a pas été détectée entre PSC et la topoisomérase par co-immunoprécipitation ou par spectrométrie de masse, ne signifie pas obligatoirement que ces protéines n'interagissent pas entre elles. Il est possible que cette interaction ne soit que transitoire et/ou pas assez stable pour être observée avec les expériences utilisées (Figure 4-7d). Si la stimulation de l'activité de la topoisomérase par PSC ne se fait pas par le biais d'une interaction entre les deux protéines, il est possible qu'elle se fasse via l'ADN. PSC en liant l'ADN et en enroulant l'ADN, peut altérer la structure de ce dernier augmentant l'affinité de la topoisomérase I pour cet ADN lié par PSC ou en permettant une relaxation plus efficace du surenroulement par la topoisomérase (Figure 4-7d). Si cette stimulation ne nécessite pas une interaction directe, on peut dans ce cas-là s'attendre à ce que PSC puisse stimuler l'activité des autres topoisomérases de type II et III.

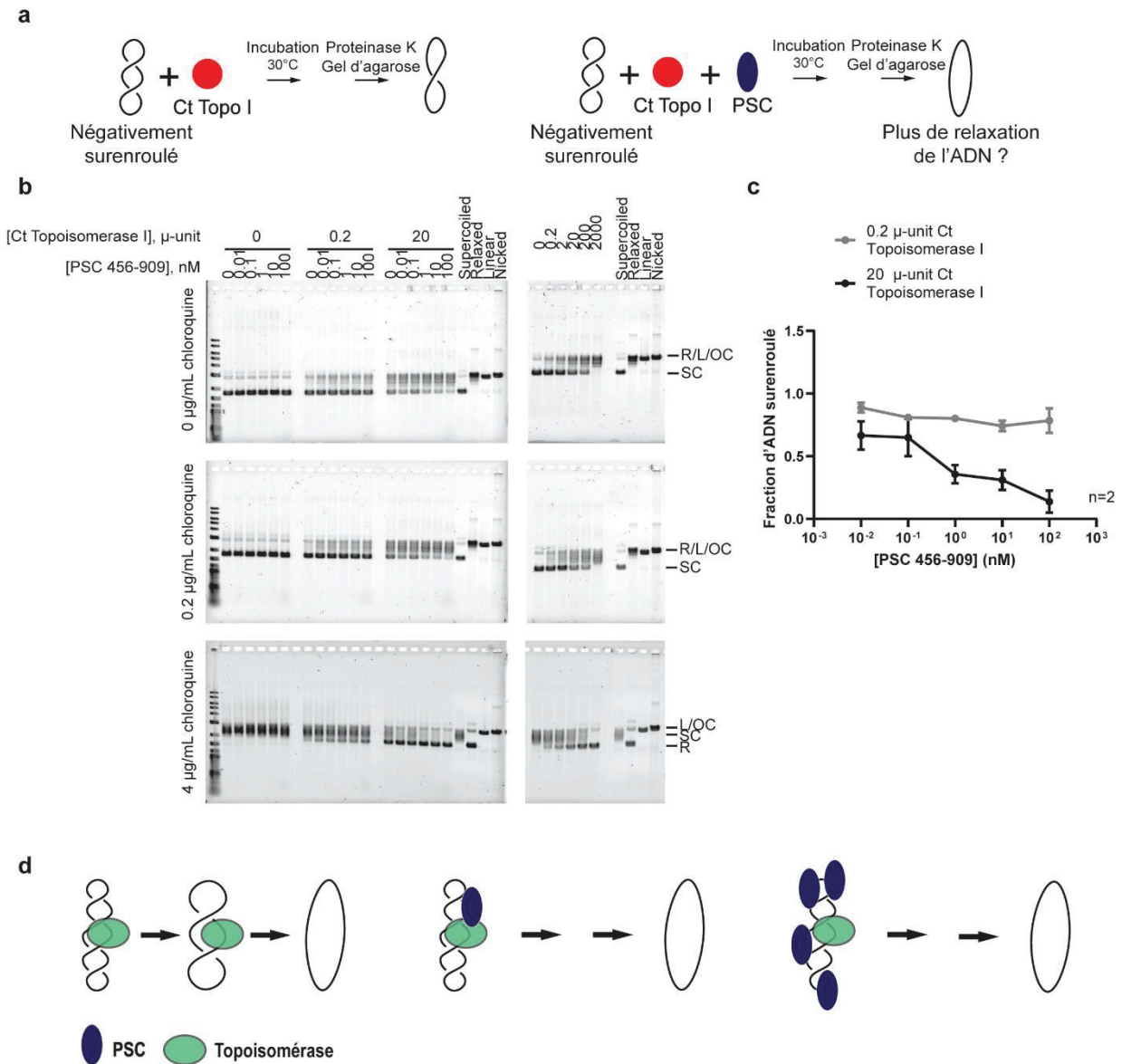


Figure 4-7 PSC stimule la relaxation de l'ADN par la topoisomérase I

a. Schéma de l'expérience de topologie avec PSC⁴⁵⁶⁻⁹⁰⁹ purifié à partir de bactérie ainsi que la Topoisomérase de thymus de veau (Thermo Fisher). b-c. Gels représentatifs (b) et quantifications (c) de l'expérience de topologie en présence d'une quantité constante de topoisomérase I et une quantité croissante de PSC⁴⁵⁶⁻⁹⁰⁹. [PSC⁴⁵⁶⁻⁹⁰⁹] = 0.01-100 nM, [Topoisomérase I] = 0.2-2000 μ -unit, [DNA] = 0.6 nM. SC: surenroulé, R: relaxé, L: Linéaire, OC: cercle ouvert. Le graphique présente la moyenne \pm SEM. d. Modèles de la stimulation de la topoisomérase I par PSC. La stimulation de la topoisomérase I par PSC se fait par une interaction transitoire entre les deux protéines (milieu). La liaison de l'ADN par PSC stimule la liaison ou la relaxation de l'ADN par la topoisomérase I (droite).

PRC1 n'est pas uniquement impliqué dans la répression, mais le complexe est recruté à des gènes activement transcrits (Pherson et al., 2017). La déplétion de dRING, de Ph ou de Pc (les auteurs n'ont pas

testé PSC) conduit à une altération de l'expression de ces gènes, à une diminution du niveau d'ARN polymérase II dans la région promotrice ainsi qu'une diminution de sa phosphorylation (Pherson et al., 2017). Est-il possible qu'à ces gènes activement transcrits PRC1 soit impliqué dans la résolution des supertours générés par la transcription, directement ou indirectement via la stimulation de la topoisomérase, afin de permettre la progression de l'ARN polymérase ?

4.9.3 La compaction de la chromatine et la relaxation de supertours sont-elles deux activités liées ?

La topoisomérase de type II agit avec les protéines condensines afin de compacter les chromosomes lors de la mitose, et la perte de l'une ou de l'autre conduit à un défaut compaction (Baxter and Aragón, 2012). Ceci suppose une coopération entre la compaction et la relaxation de surenroulement. La compaction de la chromatine introduit des supertours positifs, il est donc possible que la topoisomérase II soit indispensable pour leur relaxation au fur et à mesure de la compaction afin de permettre la progression de la condensation des chromosomes sans entrainer de bris de l'ADN en cas de stress topologique trop important (Figure 4-8). Il est intéressant de noter que PRC1 ainsi que sa sous-unité PSC seule, possèdent tous deux la capacité à compacter la chromatine et à modifier la topologie de l'ADN et de la chromatine. Est-il possible que la compaction induite par PRC1 afin de réprimer l'expression génique nécessite aussi la relaxation du surenroulement que cela crée? L'identification du résidu catalytique de PSC permettant d'abolir son activité de topoisomérase sans affecter ses autres activités pourrait permettre de comprendre le rôle de la relaxation de l'ADN dans la compaction de la chromatine.

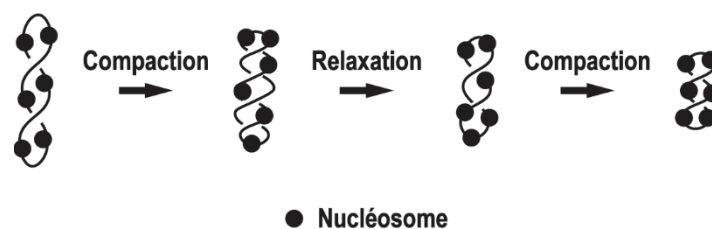


Figure 4-8 PSC compacte la chromatine et relaxe l'ADN

Schéma reliant les activités de compaction de la chromatine et de relaxation de l'ADN. La compaction de la chromatine conduit à l'apparition de surenroulement qui doit être réduit afin de permettre la compaction plus importante de la chromatine. PSC ou PRC1 pourraient agir séquentiellement en compactant la chromatine, puis réduisant le surenroulement avant de la compacter à nouveau.

4.9.4 Est-ce que PSC est impliqué dans la résolution des R-loops dont la stabilité dépend du surenroulement de l'ADN *in vivo* ?

4.9.4.1 Les R-loops sont stabilisées par la présence de surenroulement négatif

Comme présenté dans l'introduction, l'un des facteurs contribuant à la formation de R-loops est le surenroulement négatif qui permet la formation de ces structures d'acides nucléiques non-canoniques au niveau de séquences d'ADN moins favorables (Stolz et al., 2019). La déplétion de la topoisomérase I conduit à une augmentation globale du niveau de R-loops (Manzo et al., 2018). Cette augmentation est notamment présente dans les régions activement transcrites mais aussi au niveau de l'hétérochromatine constitutive marquée par H3K9me3 (Manzo et al., 2018). À l'opposé, des régions réprimées, marquées par H3K27me3 et la présence des protéines du groupe Polycomb présentent une diminution du niveau de R-loop à la suite de la déplétion de la topoisomérase I (Manzo et al., 2018), laissant supposer que toutes les R-loops ne sont pas régulées de la même façon par la topoisomérase I. Dans d'autres organismes comme chez les bactéries tels que *E. coli*, la perte de la topoisomérase I et III conduit aussi à une augmentation globale du niveau de R-loops détectées par *Dot Blot* (Brochu et al., 2018). Malheureusement, l'absence d'identification des séquences formant ces R-loops n'a pas permis de déterminer si cette augmentation globale s'accompagne d'une diminution à certains loci comme cela a été observée dans les cellules humaines (Brochu et al., 2018; Manzo et al., 2018).

Il se pourrait que les R-loops dans des régions réprimées par les protéines du groupe Polycomb, dont la formation ne se fait pas à la suite de l'accumulation d'ADN surenroulé négativement généré par la transcription, et donc n'étant pas régulées par la topoisomérase I, deviennent la cible de protéines qui jusqu'à lors ne participaient pas à leur résolution. La déplétion de la topoisomérase I stresse les cellules en raison de l'accumulation de R-loops, la RNase H1 qui n'est que peu présente au niveau de la chromatine en conditions physiologiques, pourrait alors les dégrader (Lockhart et al., 2019). Dans le chapitre 3, nous avons montré que *in vitro* PRC1 induit la résolution de R-loops qui étaient formées sur un ADN négativement surenroulé en relaxant ce dernier. *In vivo*, PSC ou PRC1 pourraient participer à la résolution de ces R-loops formées au niveau de ces domaines réprimés et dont la stabilité dépend au moins partiellement de la présence de supertours dans l'ADN, car les protéines du groupe Polycomb y sont déjà présentes. Dans ce cas-ci, il serait intéressant de déterminer ce qui conduit à l'activation de l'activité de topoisomérase de PSC. Est-ce que le stress généré par l'accumulation globale de R-loops lors de la déplétion de la topoisomérase I conduit à la modification post-traductionnelle de PSC ce qui pourrait entraîner l'activation de son activité de topoisomérase? La purification de PSC à partir de bactéries a mis

en évidence que des modifications post-traductionnelles sont essentielles pour l'activité de topoisomérase de PSC (Figure 4-7). De plus, PSC n'est pas la seule sous-unité de PRC1 pouvant être modifiée de manière post-traductionnelle (Niessen et al., 2009). Certaines de ces modifications présentes sur les différentes sous-unités de PRC1 ne peuvent coexister car elles sont présentes sur un même résidu ou deux résidus proches l'un de l'autre (Niessen et al., 2009). Ceci suggère que ces modifications post-traductionnelles puissent être régulées durant le cycle cellulaire ou à la suite de divers signaux afin de moduler les différentes activités de PRC1 dans la cellule.

4.9.4.2 Est-ce que PRC1 est impliqué dans la résolution d'autres structures d'ADN non-canoniques ?

Les R-loops ne sont pas les seules structures d'acides nucléiques non-canoniques dont la formation et la stabilité dépendent en partie du niveau de surenroulement présent dans l'ADN. En effet toutes les structures secondaires dans l'ADN nécessitant l'ouverture de l'ADN double-brin sont logiquement favorisées par la présence de surenroulement négatif qui favorise cette ouverture dans le but de dissiper ce stress topologique. Les quadruplex de G, les motifs I ou les structures en épingle à cheveux se forment préférentiellement en présence de surenroulement négatif sur un plasmide ou localement dans le génome (Dai et al., 1997; Sun and Hurley, 2009).

Tout comme les R-loops, la formation de ces structures non-canoniques est régulée par les topoisomérases, et leur niveau augmente à la suite de la déplétion ou de l'inhibition des topoisomérases (Abou Assi et al., 2018; Wright et al., 2016; Yadav et al., 2014). Est-ce que PRC1 pourrait participer à la résolution de ces structures d'ADN non-canoniques directement grâce à son activité de topoisomérase ou indirectement en stimulant la topoisomérase I ?

Les triplex ou triples hélices, structures à trois brins pouvant être composées exclusivement d'ADN ou d'ARN, ou bien à la fois d'ADN et d'ARN sont devenues l'objet d'un intérêt grandissant ces dernières années (Kunkler et al., 2019). Ces triplex lorsqu'ils sont composés de deux brins d'ADN autour desquels est enroulé un brin d'ARN peuvent servir, de manière similaire aux R-loops, au recrutement spécifique de protéines grâce à la reconnaissance par ces dernières de la structure du triplex ou de la liaison d'une partie d'ARN libre (Li et al., 2016). Comme toutes structures ne reposant pas sur des liaisons de Watson et Crick, la formation de triplex dépend de la séquence mais aussi de facteurs extérieurs tels que le magnésium, le pH ou encore la présence de surenroulement négatif (Li et al., 2016; Maxwell et al., 2006). Ces triplex composés d'ARN et d'ADN se forment dans des régions activement transcrites et sont déplétées dans celles marquées par H3K27me3 et donc réprimées par les protéines du groupe Polycomb (Sentürk Cetin et al., 2019). Néanmoins la présence de PRC1 dans des régions activement transcrites et la liaison de PRC2

à certains de ces ARN formant des triplex n'excluent pas la possibilité que les protéines du groupe Polycomb soient présentes au niveau de triplex sans pour autant modifier de manière post-traductionnelle les histones ou induire la répression (Kalwa et al., 2016; Mondal et al., 2015; Pherson et al., 2017). Dans cette situation, PRC1 pourrait-il participer à la résolution de ces triplex grâce à son activité de topoisomérase ou via la stimulation de la topoisomérase I ?

4.10 Conclusions générales

L'hypothèse initiale de ma thèse était de déterminer si les PREs forment des R-loops, et si ces R-loops peuvent déterminer le statut de ces PREs en participant au recrutement des protéines du groupe Polycomb. En partant de cette hypothèse nous avons identifié de nouvelles activités chez les complexes Polycomb PRC2 et PRC1 les impliquant dans la formation de ces R-loops, leur reconnaissance ainsi que dans leur résolution. De plus, l'identification de R-loops aux PREs régulées durant le développement laisse à penser qu'elles sont une caractéristique qui était inconnu des PREs.

4.10.1 Vers une nouvelle définition de PRE ?

La propension des PREs et autres sites liés par les protéines du groupe Polycomb à former des R-loops présentent deux idées qui peuvent faire l'objet d'une attention particulière dans le futur. Premièrement, la formation de R-loops est une caractéristique qui était jusqu'alors inconnu des PREs et elle pourrait permettre l'identification de PREs chez les mammifères et les plantes. De plus, il y a potentiellement plusieurs classes de PREs, certains dont leur fonction dépend uniquement de la présence de motifs pour des facteurs de transcription afin de recruter les protéines des groupes Polycomb ou Trithorax, et d'autres qui dépendent de la présence ou de l'absence de R-loops bien plus que de ces facteurs de transcription. Dans cette seconde catégorie, il est possible que des loci liés par les protéines Polycomb et formant des R-loops n'ayant pas été considérés comme PREs jusqu'à présent, doivent y être inclus.

Deuxièmement, les R-loops, pouvant être reconnues par les complexes Polycomb, ainsi que la formation d'autres structures d'ADN non-canoniques comme les quadruplex de G ou les motifs I pouvant coexister avec une R-loop, peuvent constituer une catégorie de marque épigénétique. Cette seconde catégorie n'implique pas la déposition de modifications post-traductionnelles, et peuvent réguler le recrutement de nombreuses protéines en inhibant ou en permettant leur recrutement à un locus donné dépendamment de l'affinité de chaque protéine pour ces structures d'acides nucléiques non-canoniques.

4.10.2 La découverte de deux nouvelles activités chez les complexes Polycomb

La formation et la reconnaissance de R-loops, structures d'acides nucléiques non-canoniques, par les protéines du groupe Polycomb offrent la toute nouvelle perspective que PRC2 pourrait activement participer à la définition du statut d'un PRE : la formation d'une R-loop grâce à l'activité d'invasion de brins de PRC2, entraînerait le recrutement d'autres complexes Polycomb et donc la répression. En dehors des PREs, l'activité d'invasion de brins par PRC2 conduisant à la formation d'hybrides ARN-ADN *in vitro*, et sa potentielle conservation chez d'autres espèces, ouvrent des perspectives nouvelles pour ce complexe qui a pour longtemps été réduit à sa capacité à méthyler les histones. De plus, cette activité permet de mettre en évidence que la liaison des ARN par PRC2 n'est pas uniquement inhibitrice pour ce complexe, mais que ces ARN peuvent à la fois inhiber son activité méthyltransférase mais aussi stimuler son activité d'invasion de brins.

L'activité de topoisomérase de PRC1 et de PSC, leur suggèrent une capacité à résoudre des structures d'ADN non-canoniques, mais aussi à participer activement à la transcription en éliminant le stress topologique que cela génère. De plus, la coexistence au sein d'une même protéine de deux activités permettant la compaction de la chromatine et la relaxation de surenroulement laisse présumer que ces deux mécanismes sont bien plus liés que ce que l'étude de la compaction des chromosomes mitotiques n'avait laissé apparaître.

4.10.3 Les protéines du groupe Polycomb sont impliquées dans la formation, la reconnaissance et la résolution de structures d'acides nucléiques non-canoniques

L'activité d'invasion de brins de PRC2 conduisant à la formation d'hybrides ARN-ADN et celle de topoisomérase de PRC1 qui peut conduire à la résolution de ces hybrides, peuvent sembler surprenantes car elles sont potentiellement antagonistes. Mais placées dans un contexte développemental, elles peuvent faire du sens (Figure 4-9). La formation d'une R-loop à un PRE par PRC2 conduit au recrutement des complexes Polycomb et à la répression, mais cette R-loop peut être résolue par PRC1 si sa stabilité dépend de la présence de surenroulements dans l'ADN. Cette résolution donne alors l'opportunité à d'autres protéines de lier ce PRE, ou à l'ARN polymérase de le transcrire à nouveau permettant soit le maintien soit le changement du statut de ce dernier. Les activités d'invasion de brins de PRC2 et de topoisomérase de PRC1 pourraient donc contribuer au changement de statut d'un PRE en participant à la formation et à la résolution de structures d'acides nucléiques non-canoniques.

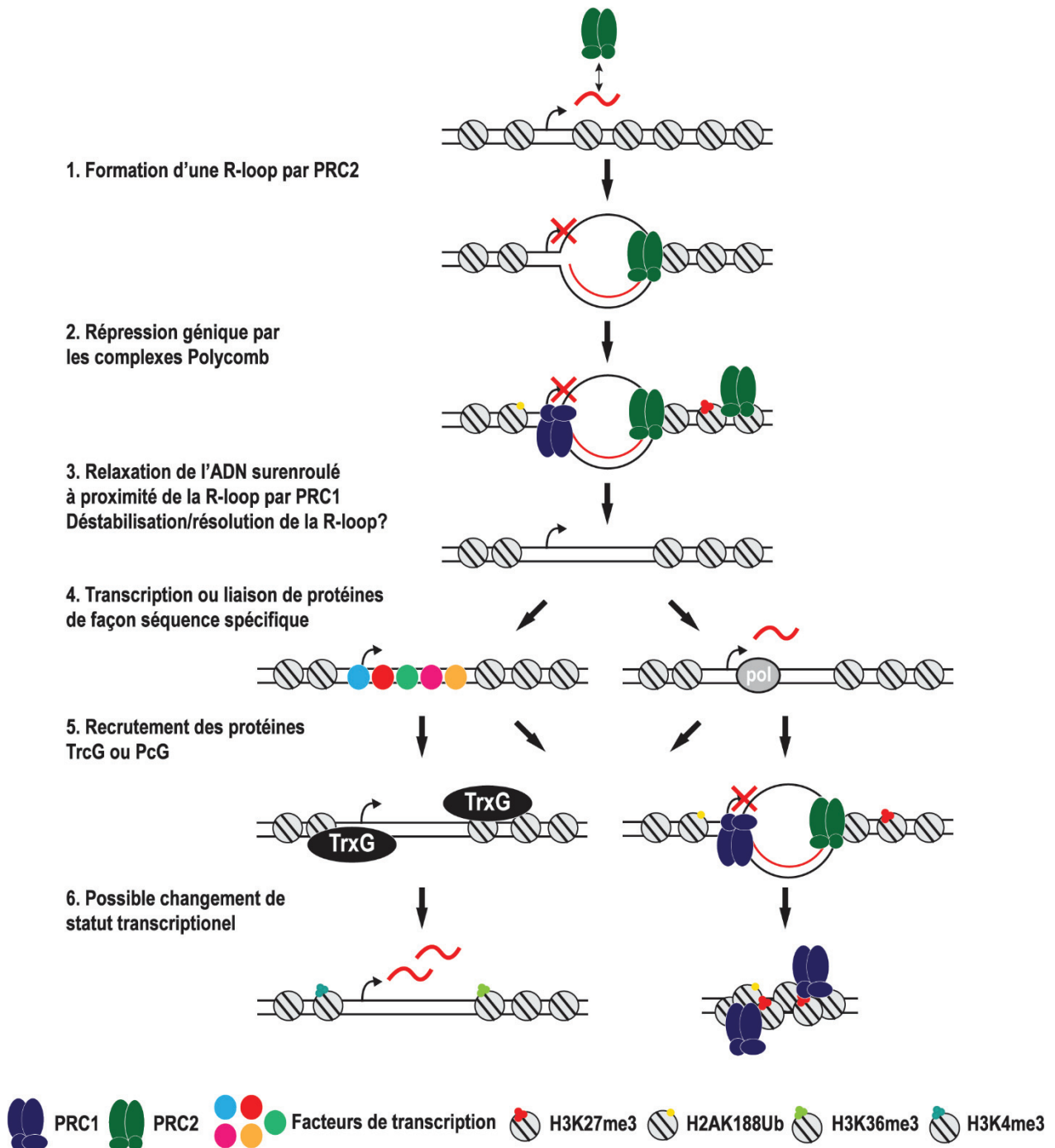


Figure 4-9 Modèle décrivant comment PRC2 et PRC1 avec leurs activités d'invasion de brins et de topoisomérase respectives peuvent participer au changement de statut d'un PRE ou d'un gène

(1) PRC2 induit la formation d'une R-loop. (2) Cette R-loop participe au recrutement des complexes Polycomb qui induisent la répression. (3) La relaxation du surenroulement de l'ADN à proximité de la R-loop, grâce à l'activité de topoisomérase de PRC1 ou via la stimulation de la topoisomérase, conduit à la résolution de la R-loop. (4 gauche) L'ADN double-brin peut être lié par des facteurs de transcription (gauche) participant au recrutement des protéines du groupe Trithorax (5 gauche) ou de celles du groupe

Polycomb (5 droite). (4 droite) L'ADN double peut être également transcrit par l'ARN polymérase (droite) permettant à PRC2 de former à nouveau une R-loop (5 droite) et d'induire la répression (6 droite). La transcription et les ARN peuvent inhiber la liaison de PRC2 et permettre le recrutement des protéines du groupe Trithorax (5 gauche) conduisant à l'augmentation de la transcription (6 gauche).

Plus généralement, les complexes Polycomb par la formation, la reconnaissance et la résolution de structures d'acides nucléiques non-canoniques pourraient être impliquées dans de nombreux mécanismes allant potentiellement de la régulation de l'expression génique, à la réparation des cassures dans l'ADN en passant par la diminution du stress topologique, expliquant ainsi pourquoi les protéines du groupe Polycomb sont détectées en dehors des régions qu'elles répriment. Des études futures seront nécessaires afin de comprendre toute l'étendue des fonctions des protéines du groupe Polycomb, pas uniquement chez *D. melanogaster* mais aussi chez les mammifères et les plantes chez lesquels ces protéines sont conservées.

5 Références bibliographiques

Abakir, A., Giles, T.C., Cristini, A., Foster, J.M., Dai, N., Starczak, M., Rubio-Roldan, A., Li, M., Eleftheriou, M., Crutchley, J., et al. (2020). N⁶-methyladenosine regulates the stability of RNA:DNA hybrids in human cells. *Nat. Genet.* 52, 48–55.

Abdurashidova, G., Radulescu, S., Sandoval, O., Zahariev, S., Danailov, M.B., Demidovich, A., Santamaria, L., Biamonti, G., Riva, S., and Falaschi, A. (2007). Functional interactions of DNA topoisomerases with a human replication origin. *EMBO J.* 26, 998–1009.

Abou Assi, H., Garavís, M., González, C., and Damha, M.J. (2018). i-Motif DNA: structural features and significance to cell biology. *Nucleic Acids Res.* 46, 8038–8056.

Achar, Y.J., Adhil, M., Choudhary, R., Gilbert, N., and Foiani, M. (2020). Negative supercoil at gene boundaries modulates gene topology. *Nature* 577, 701–705.

Agger, K., Cloos, P.A.C., Christensen, J., Pasini, D., Rose, S., Rappsilber, J., Issaeva, I., Canaani, E., Salcini, A.E., and Helin, K. (2007). UTX and JMJD3 are histone H3K27 demethylases involved in HOX gene regulation and development. *Nature* 449, 731–734.

Ahmad, K., and Spens, A.E. (2019). Separate Polycomb Response Elements control chromatin state and activation of the vestigial gene. *PLOS Genet.* 15, e1007877.

Ahmad, M., Shen, W., Li, W., Xue, Y., Zou, S., Xu, D., and Wang, W. (2017). Topoisomerase 3 β is the major topoisomerase for mRNAs and linked to neurodevelopment and mental dysfunction. *Nucleic Acids Res.* 45, 2704–2713.

Alabert, C., Jasencakova, Z., and Groth, A. (2017). Chromatin Replication and Histone Dynamics. In *DNA Replication: From Old Principles to New Discoveries*, H. Masai, and M. Foiani, eds. (Singapore: Springer), pp. 311–333.

Alecki, C., Chiwara, V., Sanz, L.A., Grau, D., Arias Pérez, O., Boulrier, E.L., Armache, K.-J., Chédin, F., and Francis, N.J. (2020). RNA-DNA strand exchange by the *Drosophila* Polycomb complex PRC2. *Nat. Commun.* 11, 1–14.

Alhaj Abed, J., Ghotbi, E., Ye, P., Frolov, A., Benes, J., and Jones, R.S. (2018). De novo recruitment of Polycomb-group proteins in *Drosophila* embryos. *Dev. Camb. Engl.* 145.

Almeida, M., Pintacuda, G., Masui, O., Koseki, Y., Gdula, M., Cerase, A., Brown, D., Mould, A., Innocent, C., Nakayama, M., et al. (2017). PCGF3/5–PRC1 initiates Polycomb recruitment in X chromosome inactivation. *Science* 356, 1081–1084.

Almeida, M., Bowness, J.S., and Brockdorff, N. (2020). The many faces of Polycomb regulation by RNA. *Curr. Opin. Genet. Dev.* 61, 53–61.

Alsner, J., Svejstrup, J.Q., Kjeldsen, E., Sørensen, B.S., and Westergaard, O. (1992). Identification of an N-terminal domain of eukaryotic DNA topoisomerase I dispensable for catalytic activity but essential for in vivo function. *J. Biol. Chem.* 267, 12408–12411.

Amândio, A.R., Necsulea, A., Joye, E., Mascrez, B., and Duboule, D. (2016). Hotair Is Dispensable for Mouse Development. *PLOS Genet.* 12, e1006232.

Amiard, S., Doudeau, M., Pinte, S., Poulet, A., Lenain, C., Faivre-Moskalenko, C., Angelov, D., Hug, N., Vindigni, A., Bouvet, P., et al. (2007). A topological mechanism for TRF2-enhanced strand invasion. *Nat. Struct. Mol. Biol.* 14, 147–154.

Aoto, T., Saitoh, N., Sakamoto, Y., Watanabe, S., and Nakao, M. (2008). Polycomb Group Protein-associated Chromatin Is Reproduced in Post-mitotic G1 Phase and Is Required for S Phase Progression. *J. Biol. Chem.* 283, 18905–18915.

Arab, K., Karaulanov, E., Musheev, M., Trnka, P., Schäfer, A., Grummt, I., and Niehrs, C. (2019). GADD45A binds R-loops and recruits TET1 to CpG island promoters. *Nat. Genet.* 51, 217–223.

Arbuckle, J.H., Gardina, P.J., Gordon, D.N., Hickman, H.D., Yewdell, J.W., Pierson, T.C., Myers, T.G., and Kristie, T.M. (2017). Inhibitors of the Histone Methyltransferases EZH2/1 Induce a Potent Antiviral State and Suppress Infection by Diverse Viral Pathogens. *MBio* 8.

Ardehali, M.B., Anselmo, A., Cochrane, J.C., Kundu, S., Sadreyev, R.I., and Kingston, R.E. (2017). Polycomb Repressive Complex 2 methylates Elongin A to regulate transcription. *Mol. Cell* 68, 872-884.e6.

Ariel, F., Jegu, T., Latrasse, D., Romero-Barrios, N., Christ, A., Benhamed, M., and Crespi, M. (2014). Noncoding Transcription by Alternative RNA Polymerases Dynamically Regulates an Auxin-Driven Chromatin Loop. *Mol. Cell* 55, 383–396.

Ariel, F., Lucero, L., Christ, A., Mammarella, M.F., Jegu, T., Veluchamy, A., Mariappan, K., Latrasse, D., Blein, T., Liu, C., et al. (2020). R-Loop Mediated trans Action of the APOLO Long Noncoding RNA. *Mol. Cell* 77, 1055-1065.e4.

Arimondo, P.B., Riou, J.-F., Mergny, J.-L., Tazi, J., Sun, J.-S., Garestier, T., and Hélène, C. (2000). Interaction of human DNA topoisomerase I with G-quartet structures. *Nucleic Acids Res.* 28, 4832–4838.

Arora, R., Lee, Y., Wischnewski, H., Brun, C.M., Schwarz, T., and Azzalin, C.M. (2014). RNaseH1 regulates TERRA-telomeric DNA hybrids and telomere maintenance in ALT tumour cells. *Nat. Commun.* 5.

Austin, C.A., and Marsh, K.L. (1998). Eukaryotic DNA topoisomerase II β . *BioEssays* 20, 215–226.

Baker, N.M., Rajan, R., and Mondragón, A. (2009). Structural studies of type I topoisomerases. *Nucleic Acids Res.* 37, 693–701.

Balas, M.M., Hartwick, E.W., Barrington, C., Roberts, J.T., Wu, S.K., Bettcher, R., Griffin, A.M., Kieft, J.S., and Johnson, A.M. (2020). RNA matchmaking remodels lncRNA structure and promotes PRC2 activity. *BioRxiv* 2020.04.13.040071.

Balk, B., Maicher, A., Dees, M., Klermund, J., Luke-Glaser, S., Bender, K., and Luke, B. (2013). Telomeric RNA-DNA hybrids affect telomere-length dynamics and senescence. *Nat. Struct. Mol. Biol.* 20, 1199.

Ballaré, C., Lange, M., Lapinaite, A., Martin, G.M., Morey, L., Pascual, G., Liefke, R., Simon, B., Shi, Y., Gozani, O., et al. (2012). Phf19 links methylated Lys36 of histone H3 to regulation of Polycomb activity. *Nat. Struct. Mol. Biol.* 19, 1257–1265.

Bancaud, A., Conde e Silva, N., Barbi, M., Wagner, G., Allemand, J.-F., Mozziconacci, J., Lavelle, C., Croquette, V., Victor, J.-M., Prunell, A., et al. (2006). Structural plasticity of single chromatin fibers revealed by torsional manipulation. *Nat. Struct. Mol. Biol.* 13, 444–450.

Bantignies, F., Roure, V., Comet, I., Leblanc, B., Schuettengruber, B., Bonnet, J., Tixier, V., Mas, A., and Cavalli, G. (2011). Polycomb-Dependent Regulatory Contacts between Distant Hox Loci in *Drosophila*. *Cell* 144, 214–226.

Baranello, L., Wojtowicz, D., Cui, K., Devaiah, B.N., Chung, H.-J., Chan-Salis, K.Y., Guha, R., Wilson, K., Zhang, X., Zhang, H., et al. (2016). RNA Polymerase II Regulates Topoisomerase 1 Activity to Favor Efficient Transcription. *Cell* 165, 357–371.

Bauer, M., Trupke, J., and Ringrose, L. (2015). The quest for mammalian Polycomb response elements: are we there yet? *Chromosoma* 125, 471–496.

Baxter, J., and Aragón, L. (2012). A model for chromosome condensation based on the interplay between condensin and topoisomerase II. *Trends Genet.* 28, 110–117.

Baxter, J., Sen, N., Martínez, V.L., De Carandini, M.E.M., Schwartzman, J.B., Diffley, J.F.X., and Aragón, L. (2011). Positive supercoiling of mitotic DNA drives decatenation by topoisomerase II in eukaryotes. *Science* 331, 1328–1332.

Beck, S., Faradji, F., Brock, H., and Peronnet, F. (2010). Maintenance of Hox gene expression patterns. *Adv. Exp. Med. Biol.* 689, 41–62.

Bejarano, F., and Busturia, A. (2004). Function of the Trithorax-like gene during *Drosophila* development. *Dev. Biol.* 268, 327–341.

Bellec, M., Radulescu, O., and Lagha, M. (2018). Remembering the past: Mitotic bookmarking in a developing embryo. *Curr. Opin. Syst. Biol.* 11, 41–49.

Beltran, M., Yates, C.M., Skalska, L., Dawson, M., Reis, F.P., Viiri, K., Fisher, C.L., Sibley, C.R., Foster, B.M., Bartke, T., et al. (2016). The interaction of PRC2 with RNA or chromatin is mutually antagonistic. *Genome Res.* 26, 896–907.

Beltran, M., Tavares, M., Justin, N., Khandelwal, G., Ambrose, J., Foster, B.M., Worlock, K.B., Tvardovskiy, A., Kunzelmann, S., Herrero, J., et al. (2019). G-tract RNA removes Polycomb repressive complex 2 from genes. *Nat. Struct. Mol. Biol.* 26, 899–909.

Berger, J.M., Gamblin, S.J., Harrison, S.C., and Wang, J.C. (1996). Structure and mechanism of DNA topoisomerase II. *Nature* 379, 225–232.

Bergerat, A., de Massy, B., Gadelle, D., Varoutas, P.-C., Nicolas, A., and Forterre, P. (1997). An atypical topoisomerase II from archaea with implications for meiotic recombination. *Nature* 386, 414–417.

Beringer, M., Pisano, P., Di Carlo, V., Blanco, E., Chammas, P., Vizán, P., Gutiérrez, A., Aranda, S., Payer, B., Wierer, M., et al. (2016). EPOP Functionally Links Elongin and Polycomb in Pluripotent Stem Cells. *Mol. Cell* 64, 645–658.

Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., et al. (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* 125, 315–326.

Bharti, A.K., Olson, M.O.J., Kufe, D.W., and Rubin, E.H. (1996). Identification of a Nucleolin Binding Site in Human Topoisomerase I. *J. Biol. Chem.* 271, 1993–1997.

Biswas, S., and Rao, C.M. (2018). Epigenetic tools (The Writers, The Readers and The Erasers) and their implications in cancer therapy. *Eur. J. Pharmacol.* 837, 8–24.

Biswas, T., Aihara, H., Radman-Livaja, M., Filman, D., Landy, A., and Ellenberger, T. (2005). A structural basis for allosteric control of DNA recombination by λ integrase. *Nature* 435, 1059–1066.

Bizard, A.H., and Hickson, I.D. (2020). The many lives of type IA topoisomerases. *J. Biol. Chem.*

Bizard, A.H., Allemand, J.-F., Hassenkam, T., Paramasivam, M., Sarlós, K., Singh, M.I., and Hickson, I.D. (2019). PICH and TOP3A cooperate to induce positive DNA supercoiling. *Nat. Struct. Mol. Biol.* 26, 267–274.

Blackledge, N.P., Farcas, A.M., Kondo, T., King, H.W., McGouran, J.F., Hanssen, L.L.P., Ito, S., Cooper, S., Kondo, K., Koseki, Y., et al. (2014). Variant PRC1 Complex-Dependent H2A Ubiquitylation Drives PRC2 Recruitment and Polycomb Domain Formation. *Cell* 157, 1445–1459.

Blastyák, A., Mishra, R.K., Karch, F., and Gyurkovics, H. (2006). Efficient and specific targeting of Polycomb group proteins requires cooperative interaction between Grainyhead and Pleiohomeotic. *Mol. Cell. Biol.* 26, 1434–1444.

Boateng, K.A., Bellani, M.A., Gregoret, I.V., Pratto, F., and Camerini-Otero, R.D. (2013). Homologous Pairing Preceding SPO11-Mediated Double-Strand Breaks in Mice. *Dev. Cell* 24, 196–205.

Boehmer, P.E. (2004). RNA binding and R-loop formation by the herpes simplex virus type-1 single-stranded DNA-binding protein (ICP8). *Nucleic Acids Res.* 32, 4576–4584.

Boguslawski, S.J., Smith, D.E., Michalak, M.A., Mickelson, K.E., Yehle, C.O., Patterson, W.L., and Carrico, R.J. (1986). Characterization of monoclonal antibody to DNA:RNA and its application to immunodetection of hybrids. *J. Immunol. Methods* 89, 123–130.

Bornemann, D., Miller, E., and Simon, J. (1996). The *Drosophila* Polycomb group gene *Sex comb on midleg* (*Scm*) encodes a zinc finger protein with similarity to polyhomeotic protein. *Development* 122, 1621–1630.

- Brahmachari, S., and Marko, J.F. (2018). DNA Mechanics and Topology. In *Biomechanics in Oncology*, C. Dong, N. Zahir, and K. Konstantopoulos, eds. (Cham: Springer International Publishing), pp. 11–39.
- Brambati, A., Zardoni, L., Nardini, E., Pellicoli, A., and Liberi, G. (2020). The dark side of RNA:DNA hybrids. *Mutat. Res. Mutat. Res.* *784*, 108300.
- Bratzel, F., López-Torrejón, G., Koch, M., Del Pozo, J.C., and Calonje, M. (2010). Keeping Cell Identity in Arabidopsis Requires PRC1 RING-Finger Homologs that Catalyze H2A Monoubiquitination. *Curr. Biol.* *20*, 1853–1859.
- Bratzel, F., Yang, C., Angelova, A., López-Torrejón, G., Koch, M., del Pozo, J.C., and Calonje, M. (2012). Regulation of the New Arabidopsis Imprinted Gene AtBMI1C Requires the Interplay of Different Epigenetic Mechanisms. *Mol. Plant* *5*, 260–269.
- Bredesen, B.A., and Rehmsmeier, M. (2019). DNA sequence models of genome-wide *Drosophila melanogaster* Polycomb binding sites improve generalization to independent Polycomb Response Elements. *Nucleic Acids Res.* *47*, 7781–7797.
- Brien, G.L., Gambero, G., O’Connell, D.J., Jerman, E., Turner, S.A., Egan, C.M., Dunne, E.J., Jurgens, M.C., Wynne, K., Piao, L., et al. (2012). Polycomb PHF19 binds H3K36me3 and recruits PRC2 and demethylase NO66 to embryonic stem cell genes during differentiation. *Nat. Struct. Mol. Biol.* *19*, 1273–1281.
- Broach, J.R., Li, Y.Y., Feldman, J., Jayaram, M., Abraham, J., Nasmyth, K.A., and Hicks, J.B. (1983). Localization and sequence analysis of yeast origins of DNA replication. *Cold Spring Harb. Symp. Quant. Biol.* *47 Pt 2*, 1165–1173.
- Broccoli, S., Phoenix, P., and Drolet, M. (2000). Isolation of the topB gene encoding DNA topoisomerase III as a multicopy suppressor of topA null mutations in *Escherichia coli*. *Mol. Microbiol.* *35*, 58–68.
- Brochu, J., Vlachos-Breton, É., Sutherland, S., Martel, M., and Drolet, M. (2018). Topoisomerases I and III inhibit R-loop formation to prevent unregulated replication in the chromosomal Ter region of *Escherichia coli*. *PLoS Genet.* *14*.
- Brochu, J., Breton, É.-V., and Drolet, M. (2020). Supercoiling, R-Loops, Replication and the Functions of Bacterial Type 1A Topoisomerases. *Genes* *11*, 249.
- Brockdorff, N. (2017). Polycomb complexes in X chromosome inactivation. *Philos. Trans. R. Soc. B Biol. Sci.* *372*.

Brown, J.L., and Kassis, J.A. (2010). Spps, a Drosophila Sp1/KLF family member, binds to PREs and is required for PRE activity late in development. *Dev. Camb. Engl.* *137*, 2597–2602.

Brown, J.L., Mucci, D., Whiteley, M., Dirksen, M.-L., and Kassis, J.A. (1998). The Drosophila Polycomb Group Gene pleiohomeotic Encodes a DNA Binding Protein with Homology to the Transcription Factor YY1. *Mol. Cell* *1*, 1057–1064.

Brown, J.L., Fritsch, C., Mueller, J., and Kassis, J.A. (2003). The Drosophila pho-like gene encodes a YY1-related DNA binding protein that is redundant with pleiohomeotic in homeotic gene silencing. *Development* *130*, 285–294.

Brown, J.L., Grau, D.J., DeVido, S.K., and Kassis, J.A. (2005). An Sp1/KLF binding site is important for the activity of a Polycomb group response element from the Drosophila engrailed gene. *Nucleic Acids Res.* *33*, 5181–5189.

Buchenau, P., Hodgson, J., Strutt, H., and Arndt-Jovin, D.J. (1998). The Distribution of Polycomb-Group Proteins During Cell Division and Development in Drosophila Embryos: Impact on Models for Silencing. *J. Cell Biol.* *141*, 469–481.

Burgin, J., Alex B., and Nash, H.A. (1995). Suicide substrates reveal properties of the homology-dependent steps during integrative recombination of bacteriophage λ . *Curr. Biol.* *5*, 1312–1321.

Busturia, A., Wightman, C.D., and Sakonju, S. (1997). A silencer is required for maintenance of transcriptional repression throughout Drosophila development. *Development* *124*, 4343–4350.

Buzas, D.M., Tamada, Y., and Kurata, T. (2012). FLC: A Hidden Polycomb Response Element Shows Up in Silence. *Plant Cell Physiol.* *53*, 785–793.

Cai, L., Rothbart, S.B., Lu, R., Xu, B., Chen, W.-Y., Tripathy, A., Rockowitz, S., Zheng, D., Patel, D.J., Allis, C.D., et al. (2013). An H3K36 methylation engaging Tudor motif of polycomb-like proteins mediates PRC2 complex targeting. *Mol. Cell* *49*, 571–582.

Campbell, S., Ismail, I.H., Young, L.C., Poirier, G.G., and Hendzel, M.J. (2013). Polycomb repressive complex 2 contributes to DNA double-strand break repair. *Cell Cycle* *12*, 2675–2683.

Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., Jones, R.S., and Zhang, Y. (2002). Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science* *298*, 1039–1043.

- Cao, R., Wang, H., He, J., Erdjument-Bromage, H., Tempst, P., and Zhang, Y. (2008). Role of hPHF1 in H3K27 Methylation and Hox Gene Silencing. *Mol. Cell. Biol.* *28*, 1862–1872.
- Capranico, G., Marinello, J., and Chillemi, G. (2017). Type I DNA Topoisomerases. *J. Med. Chem.* *60*, 2169–2192.
- Caron, P., van der Linden, J., and van Attikum, H. (2019). Bon voyage: A transcriptional journey around DNA breaks. *DNA Repair* *82*, 102686.
- Carrasco-Salas, Y., Malapert, A., Sulthana, S., Molcrette, B., Chazot-Franguiadakis, L., Bernard, P., Chédin, F., Faivre-Moskalenko, C., and Vanoosthuysse, V. (2019). The extruded non-template strand determines the architecture of R-loops. *Nucleic Acids Res.* *47*, 6783–6795.
- Casas-Vila, N., Bluhm, A., Sayols, S., Dinges, N., Dejung, M., Altenhein, T., Kappei, D., Altenhein, B., Roignant, J.-Y., and Butter, F. (2017). The developmental proteome of *Drosophila melanogaster*. *Genome Res.* *27*, 1273–1285.
- Castellano-Pozo, M., Santos-Pereira, J.M., Rondón, A.G., Barroso, S., Andújar, E., Pérez-Alegre, M., García-Muse, T., and Aguilera, A. (2013). R Loops Are Linked to Histone H3 S10 Phosphorylation and Chromatin Condensation. *Mol. Cell* *52*, 583–590.
- Cavalli, G., and Paro, R. (1998). The *Drosophila* Fab-7 Chromosomal Element Conveys Epigenetic Inheritance during Mitosis and Meiosis. *Cell* *93*, 505–518.
- Cejka, P., Plank, J.L., Dombrowski, C.C., and Kowalczykowski, S.C. (2012). Decatenation of DNA by the *S. cerevisiae* Sgs1-Top3-Rmi1 and RPA Complex: A Mechanism for Disentangling Chromosomes. *Mol. Cell* *47*, 886–896.
- Cerritelli, S.M., Frolova, E.G., Feng, C., Grinberg, A., Love, P.E., and Crouch, R.J. (2003). Failure to Produce Mitochondrial DNA Results in Embryonic Lethality in Rnaseh1 Null Mice. *Mol. Cell* *11*, 807–815.
- Chakraborty, P., and Grosse, F. (2011). Human DHX9 helicase preferentially unwinds RNA-containing displacement loops (R-loops) and G-quadruplexes. *DNA Repair* *10*, 654–665.
- Chakraborty, P., Huang, J.T.J., and Hiom, K. (2018). DHX9 helicase promotes R-loop formation in cells with impaired RNA splicing. *Nat. Commun.* *9*, 1–14.
- Chammas, P., Mocavini, I., and Di Croce, L. (2020). Engaging chromatin: PRC2 structure meets function. *Br. J. Cancer* *122*, 315–328.

- Champoux, J.J. (2001). DNA Topoisomerases: Structure, Function, and Mechanism. *Annu. Rev. Biochem.* *70*, 369–413.
- Chan, C.S., Rastelli, L., and Pirrotta, V. (1994). A Polycomb response element in the *Ubx* gene that determines an epigenetically inherited state of repression. *EMBO J.* *13*, 2553–2564.
- Chan, Y.A., Aristizabal, M.J., Lu, P.Y.T., Luo, Z., Hamza, A., Kobor, M.S., Stirling, P.C., and Hieter, P. (2014). Genome-Wide Profiling of Yeast DNA:RNA Hybrid Prone Sites with DRIP-Chip. *PLoS Genet.* *10*.
- Changela, A., DiGate, R.J., and Mondragón, A. (2001). Crystal structure of a complex of a type IA DNA topoisomerase with a single-stranded DNA molecule. *Nature* *411*, 1077–1081.
- Chanvivattana, Y., Bishopp, A., Schubert, D., Stock, C., Moon, Y.-H., Sung, Z.R., and Goodrich, J. (2004). Interaction of Polycomb-group proteins controlling flowering in Arabidopsis. *Development* *131*, 5263–5276.
- Chaudhuri, J., Tian, M., Khuong, C., Chua, K., Pinaud, E., and Alt, F.W. (2003). Transcription-targeted DNA deamination by the AID antibody diversification enzyme. *Nature* *422*, 726–730.
- Chaumeil, J., Baccon, P.L., Wutz, A., and Heard, E. (2006). A novel role for Xist RNA in the formation of a repressive nuclear compartment into which genes are recruited when silenced. *Genes Dev.* *20*, 2223–2237.
- Chédin, F. (2016). Nascent connections: R-loops and chromatin patterning. *Trends Genet.* *TIG 32*, 828–838.
- Chedin, F., and Benham, C.J. (2020). Emerging roles for R-loop structures in the management of topological stress. *J. Biol. Chem.* jbc.REV119.006364.
- Chen, S.-J., and Wang, J.C. (1998). Identification of Active Site Residues in Escherichia coli DNA Topoisomerase I. *J. Biol. Chem.* *273*, 6050–6056.
- Chen, D., Molitor, A., Liu, C., and Shen, W.-H. (2010). The Arabidopsis PRC1-like ring-finger proteins are necessary for repression of embryonic traits during vegetative growth. *Cell Res.* *20*, 1332–1344.
- Chen, J.-Y., Zhang, X., Fu, X.-D., and Chen, L. (2019). R-ChIP for genome-wide mapping of R-loops by using catalytically inactive RNASEH1. *Nat. Protoc.* *14*, 1661.

Chen, L., Chen, J.-Y., Zhang, X., Gu, Y., Xiao, R., Shao, C., Tang, P., Qian, H., Luo, D., Li, H., et al. (2017). R-ChIP Using Inactive RNase H Reveals Dynamic Coupling of R-loops with Transcriptional Pausing at Gene Promoters. *Mol. Cell* 68, 745-757.e5.

Chen, P.B., Chen, H.V., Acharya, D., Rando, O.J., and Fazio, T.G. (2015). R loops regulate promoter-proximal chromatin architecture and cellular differentiation. *Nat. Struct. Mol. Biol.* 22, 999–1007.

Chen, S., Jiao, L., Liu, X., Yang, X., and Liu, X. (2020). A Dimeric Structural Scaffold for PRC2-PCL Targeting to CpG Island Chromatin. *Mol. Cell* 77, 1265-1278.e7.

Chen, S.H., Wu, C.-H., Plank, J.L., and Hsieh, T. (2012). Essential Functions of C Terminus of *Drosophila* Topoisomerase III α in Double Holliday Junction Dissolution. *J. Biol. Chem.* 287, 19346–19353.

Chen, S.H., Plank, J.L., Willcox, S., Griffith, J.D., and Hsieh, T. (2014). Top3 α Is Required during the Convergent Migration Step of Double Holliday Junction Dissolution. *PLoS ONE* 9.

Chillemi, G., Redinbo, M., Bruselles, A., and Desideri, A. (2004). Role of the linker domain and the 203-214 N-terminal residues in the human topoisomerase I DNA complex dynamics. *Biophys. J.* 87, 4087–4097.

Chittock, E.C., Latwiel, S., Miller, T.C.R., and Müller, C.W. (2017). Molecular architecture of polycomb repressive complexes. *Biochem. Soc. Trans.* 45, 193–205.

Chon, H., Vassilev, A., DePamphilis, M.L., Zhao, Y., Zhang, J., Burgers, P.M., Crouch, R.J., and Cerritelli, S.M. (2009). Contributions of the two accessory subunits, RNASEH2B and RNASEH2C, to the activity and properties of the human RNase H2 complex. *Nucleic Acids Res.* 37, 96–110.

Chon, H., Sparks, J.L., Rychlik, M., Nowotny, M., Burgers, P.M., Crouch, R.J., and Cerritelli, S.M. (2013). RNase H2 roles in genome integrity revealed by unlinking its activities. *Nucleic Acids Res.* 41, 3130–3143.

Christensen, M.O., Barthelmes, H.U., Boege, F., and Mielke, C. (2002). The N-terminal Domain Anchors Human Topoisomerase I at Fibrillar Centers of Nucleoli and Nucleolar Organizer Regions of Mitotic Chromosomes. *J. Biol. Chem.* 277, 35932–35938.

Christensen, M.O., Barthelmes, H.U., Boege, F., and Mielke, C. (2003). Residues 190–210 of human topoisomerase I are required for enzyme activity in vivo but not in vitro. *Nucleic Acids Res.* 31, 7255–7263.

Chuang, R.-Y., Chretien, L., Dai, J., and Kelly, T.J. (2002). Purification and characterization of the *Schizosaccharomyces pombe* origin recognition complex: interaction with origin DNA and Cdc18 protein. *J. Biol. Chem.* 277, 16920–16927.

Ciferri, C., Lander, G.C., Maiolica, A., Herzog, F., Aebersold, R., and Nogales, E. (2012). Molecular architecture of human polycomb repressive complex 2. *ELife* 1.

Classen, S., Olland, S., and Berger, J.M. (2003). Structure of the topoisomerase II ATPase region and its mechanism of inhibition by the chemotherapeutic agent ICRF-187. *Proc. Natl. Acad. Sci. U. S. A.* 100, 10629–10634.

Cloud, K.G., Shen, B., Strniste, G.F., and Park, M.S. (1995). XPG protein has a structure-specific endonuclease activity. *Mutat. Res. Lett.* 347, 55–60.

Coleman, R.T., and Struhl, G. (2017). Causal role for inheritance of H3K27me3 in maintaining the OFF state of a *Drosophila* HOX gene. *Science* 356.

Connelly, K.E., and Dykhuizen, E.C. (2017). Compositional and functional diversity of canonical PRC1 complexes in mammals. *Biochim. Biophys. Acta BBA - Gene Regul. Mech.* 1860, 233–245.

Corless, S., and Gilbert, N. (2017). Investigating DNA supercoiling in eukaryotic genomes. *Brief. Funct. Genomics* 16, 379–389.

Costa, S., and Dean, C. (2019). Storing memories: the distinct phases of Polycomb-mediated silencing of *Arabidopsis* FLC. *Biochem. Soc. Trans.* 47, 1187–1196.

Cozzarelli, N.R., Krasnow, M.A., Gerrard, S.P., and White, J.H. (1984). A Topological Treatment of Recombination and Topoisomerases. *Cold Spring Harb. Symp. Quant. Biol.* 49, 383–400.

Crevillén, P., Sonmez, C., Wu, Z., and Dean, C. (2013). A gene loop containing the floral repressor FLC is disrupted in the early phase of vernalization. *EMBO J.* 32, 140–148.

Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. U. S. A.* 107, 21931–21936.

Crisona, N.J., Weinberg, R.L., Peter, B.J., Sumners, D.W., and Cozzarelli, N.R. (1999). The Topological Mechanism of Phage λ Integrase. *J. Mol. Biol.* 289, 747–775.

Cristini, A., Groh, M., Kristiansen, M.S., and Gromak, N. (2018). RNA/DNA Hybrid Interactome Identifies DXH9 as a Molecular Player in Transcriptional Termination and R-Loop-Associated DNA Damage. *Cell Rep.* 23, 1891–1905.

Cuddapah, S., Jothi, R., Schones, D.E., Roh, T.-Y., Cui, K., and Zhao, K. (2009). Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains. *Genome Res.* *19*, 24–32.

Dai, X., Greizerstein, M.B., Nadas-Chinni, K., and Rothman-Denes, L.B. (1997). Supercoil-induced extrusion of a regulatory DNA hairpin. *Proc. Natl. Acad. Sci. U. S. A.* *94*, 2174–2179.

da Rocha, S.T., Boeva, V., Escamilla-Del-Arenal, M., Ancelin, K., Granier, C., Matias, N.R., Sanulli, S., Chow, J., Schulz, E., Picard, C., et al. (2014). Jarid2 Is Implicated in the Initial Xist-Induced Targeting of PRC2 to the Inactive X Chromosome. *Mol. Cell* *53*, 301–316.

Davidovich, C., Zheng, L., Goodrich, K.J., and Cech, T.R. (2013). Promiscuous RNA binding by Polycomb repressive complex 2. *Nat. Struct. Mol. Biol.* *20*, 1250–1257.

Davidovich, C., Goodrich, K.J., Gooding, A.R., and Cech, T.R. (2014). A dimeric state for PRC2. *Nucleic Acids Res.* *42*, 9236–9248.

De, S., Mitra, A., Cheng, Y., Pfeifer, K., and Kassis, J.A. (2016). Formation of a Polycomb-Domain in the Absence of Strong Polycomb Response Elements. *PLoS Genet.* *12*.

De, S., Cheng, Y., Sun, M., Gehred, N.D., and Kassis, J.A. (2019). Structure and function of an ectopic Polycomb chromatin domain. *Sci. Adv.* *5*.

Decoville, M., Giacomello, E., Leng, M., and Locker, D. (2001). DSP1, an HMG-like protein, is involved in the regulation of homeotic genes. *Genetics* *157*, 237–244.

Dehé, P.-M., and Gaillard, P.-H.L. (2017). Control of structure-specific endonucleases to maintain genome stability. *Nat. Rev. Mol. Cell Biol.* *18*, 315–330.

Déjardin, J., and Cavalli, G. (2004). Chromatin inheritance upon Zeste-mediated Brahma recruitment at a minimal cellular memory module. *EMBO J.* *23*, 857–868.

Déjardin, J., Rappailles, A., Cuvier, O., Grimaud, C., Decoville, M., Locker, D., and Cavalli, G. (2005). Recruitment of *Drosophila* Polycomb group proteins to chromatin by DSP1. *Nature* *434*, 533–538.

Dhar, V., and Schildkraut, C.L. (1991). Role of EBNA-1 in arresting replication forks at the Epstein-Barr virus oriP family of tandem repeats. *Mol. Cell. Biol.* *11*, 6268–6278.

Dietrich, N., Lerdrup, M., Landt, E., Agrawal-Singh, S., Bak, M., Tommerup, N., Rappsilber, J., Södersten, E., and Hansen, K. (2012). REST-Mediated Recruitment of Polycomb Repressor Complexes in Mammalian Cells. *PLoS Genet.* *8*.

DiGate, R.J., and Mariani, K.J. (1988). Identification of a potent decatenating enzyme from *Escherichia coli*. *J. Biol. Chem.* *263*, 13366–13373.

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological Domains in Mammalian Genomes Identified by Analysis of Chromatin Interactions. *Nature* *485*, 376–380.

Doksani, Y., Wu, J.Y., de Lange, T., and Zhuang, X. (2013). Super-resolution fluorescence imaging of telomeres reveals TRF2-dependent T-loop formation. *Cell* *155*, 345–356.

Dong, K.C., and Berger, J.M. (2007). Structural basis for gate-DNA recognition and bending by type IIA topoisomerases. *Nature* *450*, 1201–1205.

Dorafshan, E., Kahn, T.G., and Schwartz, Y.B. (2017). Hierarchical recruitment of Polycomb complexes revisited. *Nucleus* *8*, 496–505.

Drolet, M., Bi, X., and Liu, L.F. (1994). Hypernegative supercoiling of the DNA template during transcription elongation in vitro. *J. Biol. Chem.* *269*, 2068–2074.

Durban, E., Mills, J.S., Roll, D., and Busch, H. (1983). Phosphorylation of purified Novikoff hepatoma topoisomerase I. *Biochem. Biophys. Res. Commun.* *111*, 897–905.

El Hage, A., French, S.L., Beyer, A.L., and Tollervey, D. (2010). Loss of Topoisomerase I leads to R-loop-mediated transcriptional blocks during ribosomal RNA synthesis. *Genes Dev.* *24*, 1546–1558.

Eng, W.K., Pandit, S.D., and Sternglanz, R. (1989). Mapping of the active site tyrosine of eukaryotic DNA topoisomerase I. *J. Biol. Chem.* *264*, 13373–13376.

Erokhin, M., Elizar'ev, P., Parshikov, A., Schedl, P., Georgiev, P., and Chetverina, D. (2015). Transcriptional read-through is not sufficient to induce an epigenetic switch in the silencing activity of Polycomb response elements. *Proc. Natl. Acad. Sci. U. S. A.*

Exner, V., Aichinger, E., Shu, H., Wildhaber, T., Alfarano, P., Caflisch, A., Grussem, W., Köhler, C., and Hennig, L. (2009). The Chromodomain of LIKE HETEROCHROMATIN PROTEIN 1 Is Essential for H3K27me3 Binding and Function during Arabidopsis Development. *PLoS ONE* *4*.

Fabre, F., Chan, A., Heyer, W.-D., and Gangloff, S. (2002). Alternate pathways involving Sgs1/Top3, Mus81/Mms4, and Srs2 prevent formation of toxic recombination intermediates from single-stranded gaps created by DNA replication. *Proc. Natl. Acad. Sci. U. S. A.* *99*, 16887–16892.

Fachinetti, D., Bermejo, R., Cocito, A., Minardi, S., Katou, Y., Kanoh, Y., Shirahige, K., Azvolinsky, A., Zakian, V.A., and Foiani, M. (2010). Replication termination at eukaryotic chromosomes is mediated by Top2 and occurs at genomic loci containing pausing elements. *Mol. Cell* *39*, 595–605.

Fang, Y., Chen, L., Lin, K., Feng, Y., Zhang, P., Pan, X., Sanders, J., Wu, Y., Wang, X., Su, Z., et al. (2019). Characterization of functional relationships of R-loops with gene transcription and epigenetic modifications in rice. *Genome Res.* *29*, 1287–1297.

Farkas, G., Gausz, J., Galloni, M., Reuter, G., Gyurkovics, H., and Karch, F. (1994). The Trithorax-like gene encodes the Drosophila GAGA factor. *Nature* *371*, 806–808.

Fauvarque, M.O., and Dura, J.M. (1993). polyhomeotic regulatory sequences induce developmental regulator-dependent variegation and targeted P-element insertions in Drosophila. *Genes Dev.* *7*, 1508–1520.

Fay, M.M., Lyons, S.M., and Ivanov, P. (2017). RNA G-quadruplexes in biology: principles and molecular mechanisms. *J. Mol. Biol.* *429*, 2127–2147.

Fedorova, E., and Zink, D. (2008). Nuclear architecture and gene regulation. *Biochim. Biophys. Acta BBA - Mol. Cell Res.* *1783*, 2174–2184.

Fernández, X., Díaz-Ingelmo, O., Martínez-García, B., and Roca, J. (2014). Chromatin regulates DNA torsional energy via topoisomerase II-mediated relaxation of positive supercoils. *EMBO J.* *33*, 1492–1501.

Ferrari, K.J., Scelfo, A., Jammula, S., Cuomo, A., Barozzi, I., Stützer, A., Fischle, W., Bonaldi, T., and Pasini, D. (2014). Polycomb-Dependent H3K27me1 and H3K27me2 Regulate Active Transcription and Enhancer Fidelity. *Mol. Cell* *53*, 49–62.

Festuccia, N., Gonzalez, I., Owens, N., and Navarro, P. (2017). Mitotic bookmarking in development and stem cells. *Development* *144*, 3633–3645.

Ficz, G., Heintzmann, R., and Arndt-Jovin, D.J. (2005). Polycomb group protein complexes exchange rapidly in living Drosophila. *Development* *132*, 3963–3976.

Fischle, W., Wang, Y., Jacobs, S.A., Kim, Y., Allis, C.D., and Khorasanizadeh, S. (2003). Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. *Genes Dev.* *17*, 1870–1881.

Fitzgerald, D.P., and Bender, W. (2001). Polycomb Group Repression Reduces DNA Accessibility. *Mol. Cell Biol.* *21*, 6585–6597.

Follmer, N.E., Wani, A.H., and Francis, N.J. (2012). A Polycomb Group Protein Is Retained at Specific Sites on Chromatin in Mitosis. *PLOS Genet.* *8*, e1003135.

Fonseca, J.P., Steffen, P.A., Müller, S., Lu, J., Sawicka, A., Seiser, C., and Ringrose, L. (2012). In vivo Polycomb kinetics and mitotic chromatin binding distinguish stem cells from differentiated cells. *Genes Dev.* *26*, 857–871.

Francis, N.J., Saurin, A.J., Shao, Z., and Kingston, R.E. (2001). Reconstitution of a Functional Core Polycomb Repressive Complex. *Mol. Cell* *8*, 545–556.

Francis, N.J., Kingston, R.E., and Woodcock, C.L. (2004). Chromatin Compaction by a Polycomb Group Protein Complex. *Science* *306*, 1574–1577.

Francis, N.J., Follmer, N.E., Simon, M.D., Aghia, G., and Butler, J.D. (2009). Polycomb proteins remain bound to chromatin and DNA during DNA replication in vitro. *Cell* *137*, 110–122.

Frey, F., Sheahan, T., Finkl, K., Stoehr, G., Mann, M., Benda, C., and Müller, J. (2016). Molecular basis of PRC1 targeting to Polycomb response elements by PhoRC. *Genes Dev.* *30*, 1116–1127.

Frøhlich, R.F., Andersen, F.F., Westergaard, O., Andersen, A.H., and Knudsen, B.R. (2004). Regions within the N-terminal Domain of Human Topoisomerase I Exert Important Functions During Strand Rotation and DNA Binding. *J. Mol. Biol.* *336*, 93–103.

Fujita, M., Hori, Y., Shirahige, K., Tsurimoto, T., Yoshikawa, H., and Obuse, C. (1998). Cell cycle dependent topological changes of chromosomal replication origins in *Saccharomyces cerevisiae*. *Genes Cells Devoted Mol. Cell. Mech.* *3*, 737–749.

Fursova, N.A., Blackledge, N.P., Nakayama, M., Ito, S., Koseki, Y., Farcas, A.M., King, H.W., Koseki, H., and Klose, R.J. (2019). Synergy between Variant PRC1 Complexes Defines Polycomb-Mediated Gene Repression. *Mol. Cell* *74*, 1020-1036.e8.

Gambetta, M.C., and Müller, J. (2014). O-GlcNAcylation Prevents Aggregation of the Polycomb Group Repressor Polyhomeotic. *Dev. Cell* 31, 629–639.

Gan, W., Guan, Z., Liu, J., Gui, T., Shen, K., Manley, J.L., and Li, X. (2011). R-loop-mediated genomic instability is caused by impairment of replication fork progression. *Genes Dev.* 25, 2041–2056.

Garland, W., Comet, I., Wu, M., Radzsheuskaya, A., Rib, L., Vitting-Seerup, K., Lloret-Llinares, M., Sandelin, A., Helin, K., and Jensen, T.H. (2019). A Functional Link between Nuclear RNA Decay and Transcriptional Control Mediated by the Polycomb Repressive Complex 2. *Cell Rep.* 29, 1800-1811.e6.

Garnier, O., Laouiellé-Duprat, S., and Spillane, C. (2008). Genomic imprinting in plants. *Adv. Exp. Med. Biol.* 626, 89–100.

Gatchalian, J., Kingsley, M.C., Moslet, S.D., Rosas Ospina, R.D., and Kutateladze, T.G. (2015). An aromatic cage is required but not sufficient for binding of Tudor domains of the Polycomblike protein family to H3K36me3. *Epigenetics* 10, 467–473.

Gibbons, H.R., Shaginurova, G., Kim, L.C., Chapman, N., Spurlock, C.F., and Aune, T.M. (2018). Divergent lncRNA GATA3-AS1 Regulates GATA3 Transcription in T-Helper 2 Cells. *Front. Immunol.* 9.

Gilbert, S.F. (2000). *The Origins of Anterior-Posterior Polarity.* Dev. Biol. 6th Ed.

Gilmour, D.S., Pflugfelder, G., Wang, J.C., and Lis, J.T. (1986). Topoisomerase I interacts with transcribed regions in *Drosophila* cells. *Cell* 44, 401–407.

Giménez-Abián, J.F., Clarke, D.J., Devlin, J., Giménez-Abián, M.I., De la Torre, C., Johnson, R.T., Mullinger, A.M., and Downes, C.S. (2000). Premitotic chromosome individualization in mammalian cells depends on topoisomerase II activity. *Chromosoma* 109, 235–244.

Ginjala, V., Nacerddine, K., Kulkarni, A., Oza, J., Hill, S.J., Yao, M., Citterio, E., van Lohuizen, M., and Ganesan, S. (2011). BMI1 Is Recruited to DNA Breaks and Contributes to DNA Damage-Induced H2A Ubiquitination and Repair. *Mol. Cell. Biol.* 31, 1972–1982.

Ginno, P.A., Lott, P.L., Christensen, H.C., Korf, I., and Chédin, F. (2012). R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. *Mol. Cell* 45, 814–825.

Ginno, P.A., Lim, Y.W., Lott, P.L., Korf, I., and Chédin, F. (2013). GC skew at the 5' and 3' ends of human genes links R-loop formation to epigenetic regulation and transcription termination. *Genome Res.* 23, 1590–1600.

- Graf, M., Bonetti, D., Lockhart, A., Serhal, K., Kellner, V., Maicher, A., Jolivet, P., Teixeira, M.T., and Luke, B. (2017). Telomere Length Determines TERRA and R-Loop Regulation through the Cell Cycle. *Cell* 170, 72-85.e14.
- Grau, D.J., Chapman, B.A., Garlick, J.D., Borowsky, M., Francis, N.J., and Kingston, R.E. (2011). Compaction of chromatin by diverse Polycomb group proteins requires localized regions of high charge. *Genes Dev.* 25, 2210–2221.
- Grijzenhout, A., Godwin, J., Koseki, H., Gdula, M.R., Szumska, D., McGouran, J.F., Bhattacharya, S., Kessler, B.M., Brockdorff, N., and Cooper, S. (2016). Functional analysis of AEBP2, a PRC2 Polycomb protein, reveals a Trithorax phenotype in embryonic development and in ESCs. *Dev. Camb. Engl.* 143, 2716–2723.
- Grindley, N.D.F., Whiteson, K.L., and Rice, P.A. (2006). Mechanisms of Site-Specific Recombination. *Annu. Rev. Biochem.* 75, 567–605.
- Groth, A.C., and Calos, M.P. (2004). Phage Integrases: Biology and Applications. *J. Mol. Biol.* 335, 667–678.
- Gu, X., Xu, T., and He, Y. (2014). A Histone H3 Lysine-27 Methyltransferase Complex Represses Lateral Root Formation in *Arabidopsis thaliana*. *Mol. Plant* 7, 977–988.
- Gupta, P., Zlatanova, J., and Tomschik, M. (2009). Nucleosome assembly depends on the torsion in the DNA molecule: a magnetic tweezers study. *Biophys. J.* 97, 3150–3157.
- Gupta, R.A., Shah, N., Wang, K.C., Kim, J., Horlings, H.M., Wong, D.J., Tsai, M.-C., Hung, T., Argani, P., Rinn, J.L., et al. (2010). Long noncoding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* 464, 1071–1076.
- Gutiérrez, L., Oktaba, K., Scheuermann, J.C., Gambetta, M.C., Ly-Hartig, N., and Müller, J. (2012). The role of the histone H2A ubiquitinase Sce in Polycomb repression. *Dev. Camb. Engl.* 139, 117–127.
- Harikumar, A., and Meshorer, E. (2015). Chromatin remodeling and bivalent histone modifications in embryonic stem cells. *EMBO Rep.* 16, 1609–1619.
- Hartono, S.R., Malapert, A., Legros, P., Bernard, P., Chédin, F., and Vanoosthuyse, V. (2018). The Affinity of the S9.6 Antibody for Double-Stranded RNAs Impacts the Accurate Mapping of R-Loops in Fission Yeast. *J. Mol. Biol.* 430, 272–284.

Hauri, S., Comoglio, F., Seimiya, M., Gerstung, M., Glatter, T., Hansen, K., Aebersold, R., Paro, R., Gstaiger, M., and Beisel, C. (2016). A High-Density Map for Navigating the Human Polycomb Complexome. *Cell Rep.* *17*, 583–595.

Healy, E., Mucha, M., Glancy, E., Fitzpatrick, D.J., Conway, E., Neikes, H.K., Monger, C., Mierlo, G.V., Baltissen, M.P., Koseki, Y., et al. (2019). PRC2.1 and PRC2.2 Synergize to Coordinate H3K27 Trimethylation. *Mol. Cell* *0*.

Hegazy, Y.A., Fernando, C.M., and Tran, E.J. (2020). The balancing act of R-loop biology: The good, the bad, and the ugly. *J. Biol. Chem.* *295*, 905–913.

Hekimoglu-Balkan, B., Aszodi, A., Heinen, R., Jaritz, M., and Ringrose, L. (2012). Intergenic Polycomb target sites are dynamically marked by non-coding transcription during lineage commitment. *RNA Biol.* *9*, 314–325.

Helmrich, A., Ballarino, M., and Tora, L. (2011). Collisions between replication and transcription complexes cause common fragile site instability at the longest human genes. *Mol. Cell* *44*, 966–977.

Hennig, L., and Derkacheva, M. (2009). Diversity of Polycomb group complexes in plants: same rules, different players? *Trends Genet. TIG* *25*, 414–423.

Heo, J.B., and Sung, S. (2011). Vernalization-Mediated Epigenetic Silencing by a Long Intronic Noncoding RNA. *Science* *331*, 76–79.

Herz, H.-M., Mohan, M., Garrett, A.S., Miller, C., Casto, D., Zhang, Y., Seidel, C., Haug, J.S., Florens, L., Washburn, M.P., et al. (2012). Polycomb Repressive Complex 2-Dependent and -Independent Functions of Jarid2 in Transcriptional Regulation in *Drosophila*. *Mol. Cell. Biol.* *32*, 1683–1693.

Herzog, V.A., Lempradl, A., Trupke, J., Okulski, H., Altmutter, C., Ruge, F., Boidol, B., Kubicek, S., Schmauss, G., Aumayr, K., et al. (2014). A strand-specific switch in noncoding transcription switches the function of a Polycomb/Trithorax response element. *Nat. Genet.* *46*, 973–981.

Hildebrand, E.M., and Dekker, J. (2020). Mechanisms and Functions of Chromosome Compartmentalization. *Trends Biochem. Sci.* *45*, 385–396.

Holt, I.J. (2019). The Jekyll and Hyde character of RNase H1 and its multiple roles in mitochondrial DNA metabolism. *DNA Repair* *84*, 102630.

Hong, L., Schroth, G.P., Matthews, H.R., Yau, P., and Bradbury, E.M. (1993). Studies of the DNA binding properties of histone H4 amino terminus. Thermal denaturation studies reveal that acetylation markedly reduces the binding constant of the H4 “tail” to DNA. *J. Biol. Chem.* *268*, 305–314.

Hou, C., Li, L., Qin, Z.S., and Corces, V.G. (2012). Gene Density, Transcription, and Insulators Contribute to the Partition of the *Drosophila* Genome into Physical Domains. *Mol. Cell* *48*, 471–484.

Houchens, C.R., Lu, W., Chuang, R.-Y., Frattini, M.G., Fuller, A., Simancek, P., and Kelly, T.J. (2008). Multiple mechanisms contribute to *Schizosaccharomyces pombe* origin recognition complex-DNA interactions. *J. Biol. Chem.* *283*, 30216–30224.

Hu, D., Garruss, A.S., Gao, X., Morgan, M.A., Cook, M., Smith, E.R., and Shilatifard, A. (2013). The Mll2 branch of the COMPASS family regulates bivalent promoters in mouse embryonic stem cells. *Nat. Struct. Mol. Biol.* *20*, 1093–1097.

Hu, H., Baack, M., and Knippers, R. (2009). Proteins of the origin recognition complex (ORC) and DNA topoisomerases on mammalian chromatin. *BMC Mol. Biol.* *10*, 36.

Huang, C., and Zhu, B. (2018). Roles of H3K36-specific histone methyltransferases in transcription: antagonizing silencing and safeguarding transcription fidelity. *Biophys. Rep.* *4*, 170–177.

Huang, D.-H., Chang, Y.-L., Yang, C.-C., Pan, I.-C., and King, B. (2002). pipsqueak encodes a factor essential for sequence-specific targeting of a polycomb group protein complex. *Mol. Cell. Biol.* *22*, 6261–6271.

Huang, F.-T., Yu, K., Hsieh, C.-L., and Lieber, M.R. (2006). Downstream boundary of chromosomal R-loops at murine switch regions: Implications for the mechanism of class switch recombination. *Proc. Natl. Acad. Sci. U. S. A.* *103*, 5030–5035.

Hyjek, M., Figiel, M., and Nowotny, M. (2019). RNases H: Structure and mechanism. *DNA Repair* *84*, 102672.

Illingworth, R.S., Gruenewald-Schneider, U., Webb, S., Kerr, A.R.W., James, K.D., Turner, D.J., Smith, C., Harrison, D.J., Andrews, R., and Bird, A.P. (2010). Orphan CpG islands identify numerous conserved promoters in the mammalian genome. *PLoS Genet.* *6*, e1001134.

Ismail, I.H., Andrin, C., McDonald, D., and Hendzel, M.J. (2010). BMI1-mediated histone ubiquitylation promotes DNA double-strand break repair. *J. Cell Biol.* *191*, 45–60.

Isono, K., Endo, T.A., Ku, M., Yamada, D., Suzuki, R., Sharif, J., Ishikura, T., Toyoda, T., Bernstein, B.E., and Koseki, H. (2013). SAM Domain Polymerization Links Subnuclear Clustering of PRC1 to Gene Silencing. *Dev. Cell* 26, 565–577.

Ivanov, I.E., Wright, A.V., Cofsky, J.C., Aris, K.D.P., Doudna, J.A., and Bryant, Z. (2020). Cas9 interrogates DNA in discrete steps modulated by mismatches and supercoiling. *Proc. Natl. Acad. Sci.* 117, 5853–5860.

Jenjaroenpun, P., Wongsurawat, T., Yenamandra, S.P., and Kuznetsov, V.A. (2015). QmRLFS-finder: a model, web server and stand-alone tool for prediction and analysis of R-loop forming sequences. *Nucleic Acids Res.*

Jiao, L., and Liu, X. (2015). Structural basis of histone H3K27 trimethylation by an active polycomb repressive complex 2. *Science* 350, aac4383.

Johnson, R.C. (2015). Site-specific DNA Inversion by Serine Recombinases. 61.

Jullien, P.E., Katz, A., Oliva, M., Ohad, N., and Berger, F. (2006). Polycomb Group Complexes Self-Regulate Imprinting of the Polycomb Group Gene MEDEA in Arabidopsis. *Curr. Biol.* 16, 486–492.

Jung, H.R., Pasini, D., Helin, K., and Jensen, O.N. (2010). Quantitative Mass Spectrometry of Histones H3.2 and H3.3 in Suz12-deficient Mouse Embryonic Stem Cells Reveals Distinct, Dynamic Post-translational Modifications at Lys-27 and Lys-36. *Mol. Cell. Proteomics MCP* 9, 838–850.

Kagawa, W., Kagawa, A., Saito, K., Ikawa, S., Shibata, T., Kurumizaka, H., and Yokoyama, S. (2008). Identification of a Second DNA Binding Site in the Human Rad52 Protein. *J. Biol. Chem.* 283, 24264–24273.

Kahn, T.G., Schwartz, Y.B., Dellino, G.I., and Pirrotta, V. (2006). Polycomb Complexes and the Propagation of the Methylation Mark at the Drosophila Ubx Gene. *J. Biol. Chem.* 281, 29064–29075.

Kahn, T.G., Stenberg, P., Pirrotta, V., and Schwartz, Y.B. (2014). Combinatorial Interactions Are Required for the Efficient Recruitment of Pho Repressive Complex (PhoRC) to Polycomb Response Elements. *PLoS Genet.* 10.

Kahn, T.G., Dorafshan, E., Schultheis, D., Zare, A., Stenberg, P., Reim, I., Pirrotta, V., and Schwartz, Y.B. (2016). Interdependence of PRC1 and PRC2 for recruitment to Polycomb Response Elements. *Nucleic Acids Res.* gkw701.

Kalb, R., Latwiel, S., Baymaz, H.I., Jansen, P.W.T.C., Müller, C.W., Vermeulen, M., and Müller, J. (2014). Histone H2A monoubiquitination promotes histone H3 methylation in Polycomb repression. *Nat. Struct. Mol. Biol.* *21*, 569–571.

Kalwa, M., Hänzelmann, S., Otto, S., Kuo, C.-C., Franzen, J., Jousen, S., Fernandez-Rebollo, E., Rath, B., Koch, C., Hofmann, A., et al. (2016). The lncRNA HOTAIR impacts on mesenchymal stem cells via triple helix formation. *Nucleic Acids Res.* *44*, 10631–10643.

Kampranis, S.C., Bates, A.D., and Maxwell, A. (1999). A model for the mechanism of strand passage by DNA gyrase. *Proc. Natl. Acad. Sci.* *96*, 8414–8419.

Kanaar, R., Klippel, A., Shekhtman, E., Dungan, J.M., Kahmann, R., and Cozzarelli, N.R. (1990). Processive recombination by the phage Mu Gin system: Implications for the mechanisms of DNA strand exchange, DNA site alignment, and enhancer action. *Cell* *62*, 353–366.

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

Kang, H., McElroy, K.A., Jung, Y.L., Alekseyenko, A.A., Zee, B.M., Park, P.J., and Kuroda, M.I. (2015). Sex comb on midleg (*Scm*) is a functional link between PcG-repressive complexes in *Drosophila*. *Genes Dev.* *29*, 1136–1150.

Kang, J.J., Faubert, D., Boulais, J., and Francis, N.J. (2020). DNA binding reorganizes the intrinsically disordered C-terminal region of PSC in *Drosophila* PRC1. *BioRxiv* 2020.06.02.130492.

Kanhere, A., Viiri, K., Araújo, C.C., Rasaiyaah, J., Bouwman, R.D., Whyte, W.A., Pereira, C.F., Brookes, E., Walker, K., Bell, G.W., et al. (2010). Short RNAs are transcribed from repressed Polycomb target genes and interact with Polycomb Repressive Complex-2. *Mol. Cell* *38*, 675–688.

Kassis, J.A. (1994). Unusual properties of regulatory DNA from the *Drosophila* engrailed gene: three “pairing-sensitive” sites within a 1.6-kb region. *Genetics* *136*, 1025–1038.

Kassis, J.A. (2002). 14 - Pairing-Sensitive Silencing, Polycomb Group Response Elements, and Transposon Homing in *Drosophila*. In *Advances in Genetics*, J.C. Dunlap, and C. -ting Wu, eds. (Academic Press), pp. 421–438.

Kassis, J.A., and Brown, J.L. (2013). Polycomb Group Response Elements in *Drosophila* and Vertebrates. *Adv. Genet.* *81*, 83–118.

Kassis, J.A., Kennison, J.A., and Tamkun, J.W. (2017). Polycomb and Trithorax Group Genes in *Drosophila*. *Genetics* *206*, 1699–1725.

Kaur, H., De Muyt, A., and Lichten, M. (2015). Top3-Rmi1 DNA single-strand decatenase is integral to the formation and resolution of meiotic recombination intermediates. *Mol. Cell* *57*, 583–594.

Kay, G.F., Penny, G.D., Patel, D., Ashworth, A., Brockdorff, N., and Rastan, S. (1993). Expression of Xist during mouse development suggests a role in the initiation of X chromosome inactivation. *Cell* *72*, 171–182.

Keeney, S., Giroux, C.N., and Kleckner, N. (1997). Meiosis-Specific DNA Double-Strand Breaks Are Catalyzed by Spo11, a Member of a Widely Conserved Protein Family. *Cell* *88*, 375–384.

Keskin, H., Shen, Y., Huang, F., Patel, M., Yang, T., Ashley, K., Mazin, A.V., and Storici, F. (2014). Transcript RNA-templated DNA recombination and repair. *Nature* *515*, 436–439.

Keszthelyi, A., Minchell, N.E., and Baxter, J. (2016). The Causes and Consequences of Topological Stress during DNA Replication. *Genes* *7*.

Khalil, A.M., Guttman, M., Huarte, M., Garber, M., Raj, A., Rivea Morales, D., Thomas, K., Presser, A., Bernstein, B.E., van Oudenaarden, A., et al. (2009). Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc. Natl. Acad. Sci. U. S. A.* *106*, 11667–11672.

Kim, D.-H., and Sung, S. (2012). Environmentally coordinated epigenetic silencing of FLC by protein and long noncoding RNA components. *Curr. Opin. Plant Biol.* *15*, 51–56.

Kim, D.-H., and Sung, S. (2014). Polycomb-Mediated Gene Silencing in *Arabidopsis thaliana*. *Mol. Cells* *37*, 841–850.

Kim, D.-H., and Sung, S. (2017). Vernalization-Triggered Intragenic Chromatin Loop Formation by Long Noncoding RNAs. *Dev. Cell*.

- Kim, R.A., and Wang, J.C. (1992). Identification of the yeast TOP3 gene product as a single strand-specific DNA topoisomerase. *J. Biol. Chem.* *267*, 17178–17185.
- Kim, C.A., Gingery, M., Pilpa, R.M., and Bowie, J.U. (2002). The SAM domain of polyhomeotic forms a helical polymer. *Nat. Struct. Biol.* *9*, 453–457.
- Kim, C.A., Sawaya, M.R., Cascio, D., Kim, W., and Bowie, J.U. (2005). Structural Organization of a Sex-comb-on-midleg/Polyhomeotic Copolymer. *J. Biol. Chem.* *280*, 27769–27775.
- Kim, D.-H., Xi, Y., and Sung, S. (2017). Modular function of long noncoding RNA, COLDAIR, in the vernalization response. *PLoS Genet.* *13*.
- Kim, H., Kang, K., and Kim, J. (2009). AEBP2 as a potential targeting protein for Polycomb Repression Complex PRC2. *Nucleic Acids Res.* *37*, 2940–2950.
- Kim, H.-D., Choe, J., and Seo, Y.-S. (1999). The *sen1+* Gene of *Schizosaccharomyces pombe*, a Homologue of Budding Yeast SEN1, Encodes an RNA and DNA Helicase. *Biochemistry* *38*, 14697–14710.
- Kim, S.Y., Lee, J., Eshed-Williams, L., Zilberman, D., and Sung, Z.R. (2012). EMF1 and PRC2 Cooperate to Repress Key Regulators of Arabidopsis Development. *PLoS Genet.* *8*.
- King, I.F.G., Francis, N.J., and Kingston, R.E. (2002). Native and Recombinant Polycomb Group Complexes Establish a Selective Block to Template Accessibility To Repress Transcription In Vitro. *Mol. Cell. Biol.* *22*, 7919–7928.
- Kingston, R.E., and Tamkun, J.W. (2014). Transcriptional Regulation by Trithorax-Group Proteins. *Cold Spring Harb. Perspect. Biol.* *6*.
- Kloet, S.L., Makowski, M.M., Baymaz, H.I., van Voorthuisen, L., Karemaker, I.D., Santanach, A., Jansen, P.W.T.C., Di Croce, L., and Vermeulen, M. (2016). The dynamic interactome and genomic targets of Polycomb complexes during stem-cell differentiation. *Nat. Struct. Mol. Biol.* *23*, 682–690.
- Klymenko, T., and Müller, J. (2004). The histone methyltransferases Trithorax and Ash1 prevent transcriptional silencing by Polycomb group proteins. *EMBO Rep.* *5*, 373–377.
- Klymenko, T., Papp, B., Fischle, W., Köcher, T., Schelder, M., Fritsch, C., Wild, B., Wilm, M., and Müller, J. (2006). A Polycomb group protein complex with sequence-specific DNA-binding and selective methyl-lysine-binding activities. *Genes Dev.* *20*, 1110–1122.

König, F., Schubert, T., and Längst, G. (2017). The monoclonal S9.6 antibody exhibits highly variable binding affinities towards different R-loop sequences. *PLOS ONE* 12, e0178875.

Kouznetsova, V.L., Tchekanov, A., Li, X., Yan, X., and Tsigelny, I.F. (2019). Polycomb repressive 2 complex—Molecular mechanisms of function. *Protein Sci.* 28, 1387–1399.

van Kruijsbergen, I., Hontelez, S., and Veenstra, G.J.C. (2015). Recruiting polycomb to chromatin. *Int. J. Biochem. Cell Biol.* 67, 177–187.

Kulkarni, M., and Mukherjee, A. (2017). Understanding B-DNA to A-DNA transition in the right-handed DNA helix: Perspective from a local to global transition. *Prog. Biophys. Mol. Biol.* 128, 63–73.

Kumar, A., and Kono, H. (2020). Heterochromatin protein 1 (HP1): interactions with itself and chromatin components. *Biophys. Rev.*

Kundu, S., Ji, F., Sunwoo, H., Jain, G., Lee, J.T., Sadreyev, R.I., Dekker, J., and Kingston, R.E. (2018). Polycomb Repressive Complex 1 Generates Discrete Compacted Domains that Change during Differentiation. *Mol. Cell* 71, 191.

Kunkler, C.N., Hulewicz, J.P., Hickman, S.C., Wang, M.C., McCown, P.J., and Brown, J.A. (2019). Stability of an RNA•DNA–DNA triple helix depends on base triplet composition and length of the RNA third strand. *Nucleic Acids Res.* 47, 7213–7222.

Kuroda, M.I., Kang, H., De, S., and Kassis, J.A. (2020). Dynamic Competition of Polycomb and Trithorax in Transcriptional Programming. *Annu. Rev. Biochem.*

L Black, K., Petruk, S., Fenstermaker, T.K., Hodgson, J.W., Caplan, J.L., Brock, H.W., and Mazo, A. (2016). Chromatin proteins and RNA are associated with DNA during all phases of mitosis. *Cell Discov.* 2, 16038.

Laat, W.L. de, Appeldoorn, E., Jaspers, N.G.J., and Hoeijmakers, J.H.J. (1998). DNA Structural Elements Required for ERCC1-XPF Endonuclease Activity. *J. Biol. Chem.* 273, 7835–7842.

Lagarou, A., Mohd-Sarip, A., Moshkin, Y.M., Chalkley, G.E., Bezstarosti, K., Demmers, J.A.A., and Verrijzer, C.P. (2008). dKDM2 couples histone H2A ubiquitylation to histone H3 demethylation during Polycomb group silencing. *Genes Dev.* 22, 2799–2810.

Langlais, K.K., Brown, J.L., and Kassis, J.A. (2012). Polycomb group proteins bind an engrailed PRE in both the “ON” and “OFF” transcriptional states of engrailed. *PLoS One* 7, e48765.

Lanzuolo, C., Roure, V., Dekker, J., Bantignies, F., and Orlando, V. (2007). Polycomb response elements mediate the formation of chromosome higher-order structures in the bithorax complex. *Nat. Cell Biol.* *9*, 1167–1174.

Laprell, F., Finkl, K., and Müller, J. (2017). Propagation of Polycomb-repressed chromatin requires sequence-specific recruitment to DNA. *Science* *356*, 85–88.

Lavigne, M., Francis, N.J., King, I.F.G., and Kingston, R.E. (2004). Propagation of silencing; recruitment and repression of naive chromatin in trans by polycomb repressed chromatin. *Mol. Cell* *13*, 415–425.

Lee, J.H., and Berger, J.M. (2019). Cell Cycle-Dependent Control and Roles of DNA Topoisomerase II. *Genes* *10*.

Lee, C.-H., Holder, M., Grau, D., Saldaña-Meyer, R., Yu, J.-R., Ganai, R.A., Zhang, J., Wang, M., LeRoy, G., Dobenecker, M.-W., et al. (2018a). Distinct stimulatory mechanisms regulate the catalytic activity of Polycomb Repressive Complex 2 (PRC2). *Mol. Cell* *70*, 435-448.e5.

Lee, C.M., Wang, G., Pertsinidis, A., and Mariani, K.J. (2019). Topoisomerase III Acts at the Replication Fork To Remove Precatenanes. *J. Bacteriol.* *201*.

Lee, H.-G., Kahn, T.G., Simcox, A., Schwartz, Y.B., and Pirrotta, V. (2015). Genome-wide activities of Polycomb complexes control pervasive transcription. *Genome Res.* *25*, 1170–1181.

Lee, M.P., Brown, S.D., Chen, A., and Hsieh, T.S. (1993). DNA topoisomerase I is essential in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci.* *90*, 6656–6660.

Lee, S.-H., Siaw, G.E.-L., Willcox, S., Griffith, J.D., and Hsieh, T.-S. (2013). Synthesis and dissolution of hemicatenanes by type IA DNA topoisomerases. *Proc. Natl. Acad. Sci. U. S. A.* *110*, E3587-3594.

Lee, S.K., Xue, Y., Shen, W., Zhang, Y., Joo, Y., Ahmad, M., Chinen, M., Ding, Y., Ku, W.L., De, S., et al. (2018b). Topoisomerase 3 β interacts with RNAi machinery to promote heterochromatin formation and transcriptional silencing in *Drosophila*. *Nat. Commun.* *9*, 1–17.

Lee, Y.W., Arora, R., Wischnewski, H., and Azzalin, C.M. (2018c). TRF1 participates in chromosome end protection by averting TRF2-dependent telomeric R-loops. *Nat. Struct. Mol. Biol.* *25*, 147–153.

Lehmann, L., Ferrari, R., Vashisht, A.A., Wohlschlegel, J.A., Kurdistani, S.K., and Carey, M. (2012). Polycomb Repressive Complex 1 (PRC1) Disassembles RNA Polymerase II Preinitiation Complexes. *J. Biol. Chem.* *287*, 35784–35794.

Lengsfeld, B.M., Berry, K.N., Ghosh, S., Takahashi, M., and Francis, N.J. (2012). A Polycomb complex remains bound through DNA replication in the absence of other eukaryotic proteins. *Sci. Rep.* 2.

Leschziner, A.E., and Grindley, N.D.F. (2003). The Architecture of the $\gamma\delta$ Resolvase Crossover Site Synaptic Complex Revealed by Using Constrained DNA Substrates. *Mol. Cell* 12, 775–781.

Leshner, D.-T.T., Pommier, Y., Stewart, L., and Redinbo, M.R. (2002). 8-Oxoguanine rearranges the active site of human topoisomerase I. *Proc. Natl. Acad. Sci.* 99, 12102–12107.

Lesnik, E.A., and Freier, S.M. (1995). Relative Thermodynamic Stability of DNA, RNA, and DNA:RNA Hybrid Duplexes: Relationship with Base Composition and Structure. *Biochemistry* 34, 10807–10815.

Li, D., Peng, P., Yang, Z., and Lv, B. (2019). Formation of G-quadruplex structure in supercoiled DNA under molecularly crowded conditions. *RSC Adv.* 9, 26248–26251.

Li, H., Liefke, R., Jiang, J., Kurland, J.V., Tian, W., Deng, P., Zhang, W., He, Q., Patel, D.J., Bulyk, M.L., et al. (2017a). Polycomb-like proteins link the PRC2 complex to CpG islands. *Nature* 549, 287–291.

Li, J., Wang, Z., Hu, Y., Cao, Y., and Ma, L. (2017b). Polycomb Group Proteins RING1A and RING1B Regulate the Vegetative Phase Transition in Arabidopsis. *Front. Plant Sci.* 8.

Li, L., Liu, B., Wapinski, O.L., Tsai, M.-C., Qu, K., Zhang, J., Carlson, J.C., Lin, M., Fang, F., Gupta, R.A., et al. (2013). Targeted disruption of Hotair leads to homeotic transformation and gene de-repression. *Cell Rep.* 5, 3–12.

Li, M., Pokharel, S., Wang, J.-T., Xu, X., and Liu, Y. (2015). RECQ5-dependent SUMOylation of DNA topoisomerase I prevents transcription-associated genome instability. *Nat. Commun.* 6, 1–13.

Li, W., Kamtekar, S., Xiong, Y., Sarkis, G.J., Grindley, N.D.F., and Steitz, T.A. (2005). Structure of a Synaptic $\gamma\delta$ Resolvase Tetramer Covalently Linked to Two Cleaved DNAs. *Science* 309, 1210–1215.

Li, X., Han, Y., and Xi, R. (2010). Polycomb group genes Psc and Su(z)2 restrict follicle stem cell self-renewal and extrusion by controlling canonical and noncanonical Wnt signaling. *Genes Dev.* 24, 933–946.

Li, Y., Syed, J., and Sugiyama, H. (2016). RNA-DNA Triplex Formation by Long Noncoding RNAs. *Cell Chem. Biol.* 23, 1325–1333.

Liefke, R., Karwacki-Neisius, V., and Shi, Y. (2016). EPOP interacts with Elongin BC and USP7 to modulate the chromatin landscape. *Mol. Cell* 64, 659–672.

Lima, C.D., Wang, J.C., and Mondragón, A. (1994). Three-dimensional structure of the 67K N-terminal fragment of *E. coli* DNA topoisomerase I. *Nature* *367*, 138–146.

Liu, Y., Pelham-Webb, B., Di Giammartino, D.C., Li, J., Kim, D., Kita, K., Saiz, N., Garg, V., Doane, A., Giannakakou, P., et al. (2017). Widespread mitotic bookmarking by histone marks and transcription factors in pluripotent stem cells. *Cell Rep.* *19*, 1283–1293.

Lo, S.M., and Francis, N.J. (2010). Inhibition of chromatin remodeling by Polycomb Group protein Posterior Sex Combs is mechanistically distinct from nucleosome binding. *Biochemistry* *49*, 9438–9448.

Lo, S.M., Ahuja, N.K., and Francis, N.J. (2009). Polycomb Group Protein Suppressor 2 of Zeste Is a Functional Homolog of Posterior Sex Combs. *Mol. Cell. Biol.* *29*, 515–525.

Lo, S.M., Follmer, N.E., Lengsfeld, B.M., Madamba, E.V., Seong, S., Grau, D.J., and Francis, N.J. (2012a). A bridging model for persistence of a Polycomb Group protein complex through DNA replication in vitro. *Mol. Cell* *46*, 784–796.

Lo, S.M., McElroy, K.A., and Francis, N.J. (2012b). Chromatin Modification by PSC Occurs at One PSC per Nucleosome and Does Not Require the Acidic Patch of Histone H2A. *PLoS ONE* *7*.

Lo, S.M., Follmer, N.E., Lengsfeld, B.M., Madamba, E.V., Seong, S., Grau, D.J., and Francis, N.J. (2012c). A bridging model for persistence of a Polycomb Group protein complex through DNA replication in vitro. *Mol. Cell* *46*, 784–796.

Lockhart, A., Pires, V.B., Bento, F., Kellner, V., Luke-Glaser, S., Yakoub, G., Ulrich, H.D., and Luke, B. (2019). RNase H1 and H2 Are Differentially Regulated to Process RNA-DNA Hybrids. *Cell Rep.* *29*, 2890-2900.e5.

Lucas, I., Germe, T., Chevrier-Miller, M., and Hyrien, O. (2001). Topoisomerase II can unlink replicating DNA by precatenane removal. *EMBO J.* *20*, 6509–6519.

Lucia, F.D., Crevillen, P., Jones, A.M.E., Greb, T., and Dean, C. (2008). A PHD-Polycomb Repressive Complex 2 triggers the epigenetic silencing of FLC during vernalization. *Proc. Natl. Acad. Sci.* *105*, 16831–16836.

Lupo, R., Breiling, A., Bianchi, M.E., and Orlando, V. (2001). *Drosophila* Chromosome Condensation Proteins Topoisomerase II and Barren Colocalize with Polycomb and Maintain Fab-7 PRE Silencing. *Mol. Cell* *7*, 127–136.

Maeder, M.L., Linder, S.J., Cascio, V.M., Fu, Y., Ho, Q.H., and Joung, J.K. (2013). CRISPR RNA-guided activation of endogenous human genes. *Nat. Methods* *10*, 977–979.

Magdalou, I., Lopez, B.S., Pasero, P., and Lambert, S.A.E. (2014). The causes of replication stress and their consequences on genome stability and cell fate. *Semin. Cell Dev. Biol.* *30*, 154–164.

Malik, S.-B., Ramesh, M.A., Hulstrand, A.M., and Logsdon, J.M. (2007). Protist homologs of the meiotic Spo11 gene and topoisomerase VI reveal an evolutionary history of gene duplication and lineage-specific loss. *Mol. Biol. Evol.* *24*, 2827–2841.

Manzo, S.G., Hartono, S.R., Sanz, L.A., Marinello, J., De Biasi, S., Cossarizza, A., Capranico, G., and Chedin, F. (2018). DNA Topoisomerase I differentially modulates R-loops across the human genome. *Genome Biol.* *19*.

Margueron, R., Li, G., Sarma, K., Blais, A., Zavadil, J., Woodcock, C.L., Dynlacht, B.D., and Reinberg, D. (2008). Ezh1 and Ezh2 Maintain Repressive Chromatin through Different Mechanisms. *Mol. Cell* *32*, 503–518.

Margueron, R., Justin, N., Ohno, K., Sharpe, M.L., Son, J., Drury, W.J., Voigt, P., Martin, S., Taylor, W.R., De Marco, V., et al. (2009). Role of the polycomb protein Eed in the propagation of repressive histone marks. *Nature* *461*, 762–767.

Maurange, C., and Paro, R. (2002). A cellular memory module conveys epigenetic inheritance of hedgehog expression during *Drosophila* wing imaginal disc development. *Genes Dev.* *16*, 2672–2683.

Maxwell, A., Burton, N.P., and O’Hagan, N. (2006). High-throughput assays for DNA gyrase and other topoisomerases. *Nucleic Acids Res.* *34*, e104.

Mazina, O.M., Keskin, H., Hanamshet, K., Storici, F., and Mazin, A.V. (2017). Rad52 Inverse Strand Exchange Drives RNA-Templated DNA Double-Strand Break Repair. *Mol. Cell* *67*, 19-29.e3.

Mazina, O.M., Somarowthu, S., Kadyrova, L.Y., Baranovskiy, A.G., Tahirov, T.H., Kadyrov, F.A., and Mazin, A.V. (2020). Replication protein A binds RNA and promotes R-loop formation. *BioRxiv* 2020.02.11.943977.

McClendon, A.K., Rodriguez, A.C., and Osheroff, N. (2005). Human topoisomerase IIalpha rapidly relaxes positively supercoiled DNA: implications for enzyme action ahead of replication forks. *J. Biol. Chem.* *280*, 39337–39345.

McHugh, C.A., Chen, C.-K., Chow, A., Surka, C.F., Tran, C., McDonel, P., Pandya-Jones, A., Blanco, M., Burghard, C., Moradian, A., et al. (2015). The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature* *521*, 232–236.

Merini, W., and Calonje, M. (2015). PRC1 is taking the lead in PcG repression. *Plant J.* *83*, 110–120.

Merino, A., Madden, K.R., Lane, W.S., Champoux, J.J., and Reinberg, D. (1993). DNA topoisomerase I is involved in both repression and activation of transcription. *Nature* *365*, 227–232.

van Mierlo, G., Veenstra, G.J.C., Vermeulen, M., and Marks, H. (2019). The Complexity of PRC2 Subcomplexes. *Trends Cell Biol.* *29*, 660–671.

Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.-K., Koche, R.P., et al. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* *448*, 553–560.

Mirkin, S.M. (2001). DNA Topology: Fundamentals. In *ELS*, (American Cancer Society), p.

Mischo, H.E., Gómez-González, B., Grzechnik, P., Rondón, A.G., Wei, W., Steinmetz, L., Aguilera, A., and Proudfoot, N.J. (2011). Yeast Sen1 Helicase Protects the Genome from Transcription-Associated Instability. *Mol. Cell* *41*, 21–32.

Mo, Y.-Y., Wang, C., and Beck, W.T. (2000). A Novel Nuclear Localization Signal in Human DNA Topoisomerase I. *J. Biol. Chem.* *275*, 41107–41113.

Mohd-Sarip, A., Cléard, F., Mishra, R.K., Karch, F., and Verrijzer, C.P. (2005). Synergistic recognition of an epigenetic DNA element by Pleiohomeotic and a Polycomb core complex. *Genes Dev.* *19*, 1755–1760.

Mohd-Sarip, A., Knaap, J.A. van der, Wyman, C., Kanaar, R., Schedl, P., and Verrijzer, C.P. (2006). Architecture of a Polycomb Nucleoprotein Complex. *Mol. Cell* *24*, 91–100.

Mondal, T., Subhash, S., Vaid, R., Enroth, S., Uday, S., Reinius, B., Mitra, S., Mohammed, A., James, A.R., Hoberg, E., et al. (2015). MEG3 long noncoding RNA regulates the TGF- β pathway genes through formation of RNA–DNA triplex structures. *Nat. Commun.* *6*, 7743.

Monfort, A., and Wutz, A. (2020). The B-side of Xist. *F1000Research* *9*.

Moon, Y.-H., Chen, L., Pan, R.L., Chang, H.-S., Zhu, T., Maffeo, D.M., and Sung, Z.R. (2003). EMF Genes Maintain Vegetative Development by Repressing the Flower Program in Arabidopsis. *Plant Cell* *15*, 681–693.

Morham, S.G., Kluckman, K.D., Voulomanos, N., and Smithies, O. (1996). Targeted disruption of the mouse topoisomerase I gene by camptothecin selection. *Mol. Cell. Biol.* *16*, 6804–6809.

Moye, A.L., Porter, K.C., Cohen, S.B., Phan, T., Zyner, K.G., Sasaki, N., Lovrecz, G.O., Beck, J.L., and Bryan, T.M. (2015). Telomeric G-quadruplexes are a substrate and site of localization for human telomerase. *Nat. Commun.* *6*.

Mozgova, I., Köhler, C., and Hennig, L. (2015). Keeping the gate closed: functions of the polycomb repressive complex PRC2 in development. *Plant J.* *83*, 121–132.

Müller, J., and Kassis, J.A. (2006). Polycomb response elements and targeting of Polycomb group proteins in *Drosophila*. *Curr. Opin. Genet. Dev.* *16*, 476–484.

Müller, J., Hart, C.M., Francis, N.J., Vargas, M.L., Sengupta, A., Wild, B., Miller, E.L., O'Connor, M.B., Kingston, R.E., and Simon, J.A. (2002). Histone Methyltransferase Activity of a *Drosophila* Polycomb Group Repressor Complex. *Cell* *111*, 197–208.

Mundbjerg, K., Jørgensen, S.W., Fredsøe, J., Nielsen, I., Pedersen, J.M., Bentsen, I.B., Lisby, M., Bjergbaek, L., and Andersen, A.H. (2015). Top2 and Sgs1-Top3 Act Redundantly to Ensure rDNA Replication Termination. *PLoS Genet.* *11*.

Musselman, C.A., Avvakumov, N., Watanabe, R., Abraham, C.G., Lalonde, M.-E., Hong, Z., Allen, C., Roy, S., Nuñez, J.K., Nickoloff, J., et al. (2012). Molecular basis for H3K36me3 recognition by the Tudor domain of PHF1. *Nat. Struct. Mol. Biol.* *19*, 1266–1272.

Naughton, C., Avlonitis, N., Corless, S., Prendergast, J.G., Mati, I.K., Eijk, P.P., Cockroft, S.L., Bradley, M., Ylstra, B., and Gilbert, N. (2013). Transcription forms and remodels supercoiling domains unfolding large-scale chromatin structures. *Nat. Struct. Mol. Biol.* *20*, 387–395.

Neale, M.J., Pan, J., and Keeney, S. (2005). Endonucleolytic processing of covalent protein-linked DNA double-strand breaks. *Nature* *436*, 1053–1057.

Nègre, N., Hennetin, J., Sun, L.V., Lavrov, S., Bellis, M., White, K.P., and Cavalli, G. (2006). Chromosomal distribution of PcG proteins during *Drosophila* development. *PLoS Biol.* *4*, e170.

Nekrasov, M., Klymenko, T., Fraterman, S., Papp, B., Oktaba, K., Köcher, T., Cohen, A., Stunnenberg, H.G., Wilm, M., and Müller, J. (2007). Pcl-PRC2 is needed to generate high levels of H3-K27 trimethylation at Polycomb target genes. *EMBO J.* *26*, 4078–4088.

Nicolas, L., Cols, M., Choi, J.E., Chaudhuri, J., and Vuong, B. (2018). Generating and repairing genetically programmed DNA breaks during immunoglobulin class switch recombination. *F1000Research* *7*.

- Niehrs, C., and Luke, B. (2020). Regulatory R-loops as facilitators of gene expression and genome stability. *Nat. Rev. Mol. Cell Biol.* *21*, 167–178.
- Niessen, H.E., Demmers, J.A., and Voncken, J.W. (2009). Talking to chromatin: post-translational modulation of polycomb group function. *Epigenetics Chromatin* *2*, 10.
- Nimonkar, A.V., and Boehmer, P.E. (2003). The Herpes Simplex Virus Type-1 Single-strand DNA-binding Protein (ICP8) Promotes Strand Invasion. *J. Biol. Chem.* *278*, 9678–9682.
- Nitiss, J., and Wang, J.C. (1988). DNA topoisomerase-targeting antitumor drugs can be studied in yeast. *Proc. Natl. Acad. Sci. U. S. A.* *85*, 7501–7505.
- Nojima, T., Tellier, M., Foxwell, J., Ribeiro de Almeida, C., Tan-Wong, S.M., Dhir, S., Dujardin, G., Dhir, A., Murphy, S., and Proudfoot, N.J. (2018). Deregulated Expression of Mammalian lncRNA through Loss of SPT6 Induces R-Loop Formation, Replication Stress, and Cellular Senescence. *Mol. Cell* *72*, 970-984.e7.
- Noordermeer, D., Leleu, M., Schorderet, P., Joye, E., Chabaud, F., and Duboule, D. (2014). Temporal dynamics and developmental memory of 3D chromatin architecture at Hox gene loci. *ELife* *3*, e02557.
- Nowotny, M., Gaidamakov, S.A., Ghirlando, R., Cerritelli, S.M., Crouch, R.J., and Yang, W. (2007). Structure of Human RNase H1 Complexed with an RNA/DNA Hybrid: Insight into HIV Reverse Transcription. *Mol. Cell* *28*, 264–276.
- Nowotny, M., Cerritelli, S.M., Ghirlando, R., Gaidamakov, S.A., Crouch, R.J., and Yang, W. (2008). Specific recognition of RNA/DNA hybrid and enhancement of human RNase H1 activity by HBD. *EMBO J.* *27*, 1172–1181.
- Nunes-Düby, S.E., Azaro, M.A., and Landy, A. (1995). Swapping DNA strands and sensing homology without branch migration in λ site-specific recombination. *Curr. Biol.* *5*, 139–148.
- Ogiyama, Y., Schuettengruber, B., Papadopoulos, G.L., Chang, J.-M., and Cavalli, G. (2018). Polycomb-Dependent Chromatin Looping Contributes to Gene Silencing during *Drosophila* Development. *Mol. Cell* *71*, 73-88.e5.
- Ohle, C., Tesorero, R., Schermann, G., Dobrev, N., Sinning, I., and Fischer, T. (2016). Transient RNA-DNA Hybrids Are Required for Efficient Double-Strand Break Repair. *Cell* *167*, 1001-1013.e7.
- Ohno, K., McCabe, D., Czermin, B., Imhof, A., and Pirrotta, V. (2008). ESC, ESCL and their roles in Polycomb Group mechanisms. *Mech. Dev.* *125*, 527–541.

Okulski, H., Druck, B., Bhalerao, S., and Ringrose, L. (2011). Quantitative analysis of polycomb response elements (PREs) at identical genomic locations distinguishes contributions of PRE sequence and genomic environment. *Epigenetics Chromatin* 4, 4.

Olorunniji, F.J., Rosser, S.J., and Stark, W.M. (2016). Site-specific recombinases: molecular machines for the Genetic Revolution. *Biochem. J.* 473, 673–684.

O'Meara, M.M., and Simon, J.A. (2012). Inner workings and regulatory inputs that control Polycomb repressive complex 2. *Chromosoma* 121, 221–234.

Palozola, K.C., Donahue, G., Liu, H., Grant, G.R., Becker, J.S., Cote, A., Yu, H., Raj, A., and Zaret, K.S. (2017). Mitotic transcription and waves of gene reactivation during mitotic exit. *Science* 358, 119–122.

Papp, B., and Müller, J. (2006). Histone trimethylation and the maintenance of transcriptional ON and OFF states by trxG and PcG proteins. *Genes Dev.* 20, 2041–2054.

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

Pengelly, A.R., Copur, Ö., Jäckle, H., Herzig, A., and Müller, J. (2013). A Histone Mutant Reproduces the Phenotype Caused by Loss of Histone-Modifying Factor Polycomb. *Science* 339, 698–699.

Pengelly, A.R., Kalb, R., Finkl, K., and Müller, J. (2015). Transcriptional repression by PRC1 in the absence of H2A monoubiquitylation. *Genes Dev.* 29, 1487–1492.

Pérez-Martínez, L., Öztürk, M., Butter, F., and Luke, B. (2020). Npl3 stabilizes R-loops at telomeres to prevent accelerated replicative senescence. *EMBO Rep.* 21, e49087.

Petruk, S., Fenstermaker, T.K., Black, K.L., Brock, H.W., and Mazo, A. (2016). Detection of RNA-DNA association by a proximity ligation-based method. *Sci. Rep.* 6.

Pherson, M., Misulovin, Z., Gause, M., Mihindikulasuriya, K., Swain, A., and Dorsett, D. (2017). Polycomb repressive complex 1 modifies transcription of active genes. *Sci. Adv.* 3.

Phillips, D.D., Garboczi, D.N., Singh, K., Hu, Z., Leppla, S.H., and Leysath, C.E. (2013). The sub-nanomolar binding of DNA-RNA hybrids by the single chain Fv fragment of antibody S9.6. *J. Mol. Recognit. JMR* 26, 376–381.

- Phoenix, P., Raymond, M.-A., Massé, É., and Drolet, M. (1997). Roles of DNA Topoisomerases in the Regulation of R-loop Formation in Vitro. *J. Biol. Chem.* *272*, 1473–1479.
- Plank, J.L., Chu, S.H., Pohlhaus, J.R., Wilson-Sali, T., and Hsieh, T. (2005). *Drosophila melanogaster* Topoisomerase III α Preferentially Relaxes a Positively or Negatively Supercoiled Bubble Substrate and Is Essential during Development. *J. Biol. Chem.* *280*, 3564–3573.
- Pommier, Y. (2006). Topoisomerase I inhibitors: camptothecins and beyond. *Nat. Rev. Cancer* *6*, 789–802.
- Pommier, Y., Kerrigan, D., Hartman, K.D., and Glazer, R.I. (1990). Phosphorylation of mammalian DNA topoisomerase I and activation by protein kinase C. *J. Biol. Chem.* *265*, 9418–9422.
- Pommier, Y., Sun, Y., Huang, S.N., and Nitiss, J.L. (2016). Roles of eukaryotic topoisomerases in transcription, replication and genomic stability. *Nat. Rev. Mol. Cell Biol.* *17*, 703–721.
- Poux, S., Horard, B., Sigrist, C.J.A., and Pirrotta, V. (2002). The *Drosophila* Trithorax protein is a coactivator required to prevent re-establishment of Polycomb silencing. *Development* *129*, 2483–2493.
- Qi, L.S., Larson, M.H., Gilbert, L.A., Doudna, J.A., Weissman, J.S., Arkin, A.P., and Lim, W.A. (2013). Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. *Cell* *152*, 1173–1183.
- Qüesta, J.I., Song, J., Geraldo, N., An, H., and Dean, C. (2016). *Arabidopsis* transcriptional repressor VAL1 triggers Polycomb silencing at FLC during vernalization. *Science* *353*, 485–488.
- Ramiro, A.R., Stavropoulos, P., Jankovic, M., and Nussenzweig, M.C. (2003). Transcription enhances AID-mediated cytidine deamination by exposing single-stranded DNA on the nontemplate strand. *Nat. Immunol.* *4*, 452–456.
- Rampakakis, E., and Zannis-Hadjopoulos, M. (2009). Transient dsDNA breaks during pre-replication complex assembly. *Nucleic Acids Res.* *37*, 5714–5724.
- Rao, S.S.P., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., et al. (2014). A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell* *159*, 1665–1680.
- Redinbo, M.R., Stewart, L., Kuhn, P., Champoux, J.J., and Hol, W.G.J. (1998). Crystal Structures of Human Topoisomerase I in Covalent and Noncovalent Complexes with DNA. *Science* *279*, 1504–1513.

Reed, R.R., and Grindley, N.D.F. (1981). Transposon-mediated site-specific recombination in vitro: DNA cleavage and protein-DNA linkage at the recombination site. *Cell* 25, 721–728.

Reijns, M.A.M., Rabe, B., Rigby, R.E., Mill, P., Astell, K.R., Lettice, L.A., Boyle, S., Leitch, A., Keighren, M., Kilanowski, F., et al. (2012). Enzymatic removal of ribonucleotides from DNA is essential for mammalian genome integrity and development. *Cell* 149, 1008–1022.

Remus, D., Beall, E.L., and Botchan, M.R. (2004). DNA topology, not DNA sequence, is a critical determinant for *Drosophila* ORC-DNA binding. *EMBO J.* 23, 897–907.

Riccio, A.A., Schellenberg, M.J., and Williams, R.S. (2020). Molecular mechanisms of topoisomerase 2 DNA–protein crosslink resolution. *Cell. Mol. Life Sci.* 77, 81–91.

Rice, J.C., Briggs, S.D., Ueberheide, B., Barber, C.M., Shabanowitz, J., Hunt, D.F., Shinkai, Y., and Allis, C.D. (2003). Histone Methyltransferases Direct Different Degrees of Methylation to Define Distinct Chromatin Domains. *Mol. Cell* 12, 1591–1598.

Ringrose, L., and Paro, R. (2004). Epigenetic Regulation of Cellular Memory by the Polycomb and Trithorax Group Proteins. *Annu. Rev. Genet.* 38, 413–443.

Ringrose, L., and Paro, R. (2007). Polycomb/Trithorax response elements and epigenetic memory of cell identity. *Development* 134, 223–232.

Ringrose, L., Rehmsmeier, M., Dura, J.-M., and Paro, R. (2003). Genome-Wide Prediction of Polycomb/Trithorax Response Elements in *Drosophila melanogaster*. *Dev. Cell* 5, 759–771.

Rinn, J.L., Kertesz, M., Wang, J.K., Squazzo, S.L., Xu, X., Bruggmann, S.A., Goodnough, H., Helms, J.A., Farnham, P.J., Segal, E., et al. (2007). Functional Demarcation of Active and Silent Chromatin Domains in Human HOX Loci by Non-Coding RNAs. *Cell* 129, 1311–1323.

Robert, T., Nore, A., Brun, C., Maffre, C., Crimi, B., Guichard, V., Bourbon, H.-M., and Massy, B. de (2016). The TopoVIB-Like protein family is required for meiotic DNA double-strand break formation. *Science* 351, 943–949.

Robinson, A.K., Leal, B.Z., Chadwell, L.V., Wang, R., Ilangovan, U., Kaur, Y., Junco, S.E., Schirf, V., Osmulski, P.A., Gaczynska, M., et al. (2012). The Growth-Suppressive Function of the Polycomb Group Protein Polyhomeotic Is Mediated by Polymerization of Its Sterile Alpha Motif (SAM) Domain. *J. Biol. Chem.* 287, 8702.

- Roca, J. (2009). Topoisomerase II: a fitted mechanism for the chromatin landscape. *Nucleic Acids Res.* 37, 721–730.
- Rosa, S., Duncan, S., and Dean, C. (2016). Mutually exclusive sense–antisense transcription at FLC facilitates environmentally induced gene repression. *Nat. Commun.* 7, 1–7.
- Rosenberg, M., Blum, R., Kesner, B., Maier, V.K., Szanto, A., and Lee, J.T. (2017). Denaturing CLIP (dCLIP) pipeline identifies discrete RNA footprints on chromatin-associated proteins and reveals that CBX7 targets 3'UTRs to regulate mRNA expression. *Cell Syst.* 5, 368–385.e15.
- Roy, D., and Lieber, M.R. (2009). G Clustering Is Important for the Initiation of Transcription-Induced R-Loops In Vitro, whereas High G Density without Clustering Is Sufficient Thereafter. *Mol. Cell. Biol.* 29, 3124–3133.
- Sagie, S., Toubiana, S., Hartono, S.R., Katzir, H., Tzur-Gilat, A., Havazelet, S., Francastel, C., Velasco, G., Chédin, F., and Selig, S. (2017). Telomeres in ICF syndrome cells are vulnerable to DNA damage due to elevated DNA:RNA hybrids. *Nat. Commun.* 8.
- Sakaguchi, A., and Kikuchi, A. (2004). Functional compatibility between isoform α and β of type II DNA topoisomerase. *J. Cell Sci.* 117, 1047–1054.
- Salceda, J., Fernández, X., and Roca, J. (2006). Topoisomerase II, not topoisomerase I, is the proficient relaxase of nucleosomal DNA. *EMBO J.* 25, 2575–2583.
- Sanulli, S., Justin, N., Teissandier, A., Ancelin, K., Portoso, M., Caron, M., Michaud, A., Lombard, B., da Rocha, S.T., Offer, J., et al. (2015). Jarid2 Methylation via the PRC2 Complex Regulates H3K27me3 Deposition during Cell Differentiation. *Mol. Cell* 57, 769–783.
- Sanz, L.A., and Chédin, F. (2019). High-resolution, strand-specific R-loop mapping via S9.6-based DNA–RNA immunoprecipitation and high-throughput sequencing. *Nat. Protoc.* 14, 1734–1755.
- Sanz, L.A., Hartono, S.R., Lim, Y.W., Steyaert, S., Rajpurkar, A., Ginno, P.A., Xu, X., and Chédin, F. (2016). Prevalent, Dynamic, and Conserved R-Loop Structures Associate with Specific Epigenomic Signatures in Mammals. *Mol. Cell* 63, 167–178.
- Sarma, K., Margueron, R., Ivanov, A., Pirrotta, V., and Reinberg, D. (2008). Ezh2 Requires PHF1 To Efficiently Catalyze H3 Lysine 27 Trimethylation In Vivo. *Mol. Cell. Biol.* 28, 2718–2731.

Sasai, N., Kato, Y., Kimura, G., Takeuchi, T., and Yamaguchi, M. (2007). The *Drosophila* jumonji gene encodes a JmjC-containing nuclear protein that is required for metamorphosis. *FEBS J.* 274, 6139–6151.

Sathe, S.S., and Harte, P.J. (1995). The *Drosophila* extra sex combs protein contains WD motifs essential for its function as a repressor of homeotic genes. *Mech. Dev.* 52, 77–87.

Saurin, A.J., Shao, Z., Erdjument-Bromage, H., Tempst, P., and Kingston, R.E. (2001). A *Drosophila* Polycomb group complex includes Zeste and dTAFII proteins. *Nature* 412, 655–660.

Scelfo, A., Fernández-Pérez, D., Tamburri, S., Zanotti, M., Lavarone, E., Soldi, M., Bonaldi, T., Ferrari, K.J., and Pasini, D. (2019). Functional Landscape of PCGF Proteins Reveals Both RING1A/B-Dependent-and RING1A/B-Independent-Specific Activities. *Mol. Cell* 74, 1037.

Scheuermann, J.C., de Ayala Alonso, A.G., Oktaba, K., Ly-Hartig, N., McGinty, R.K., Fraterman, S., Wilm, M., Muir, T.W., and Müller, J. (2010). Histone H2A deubiquitinase activity of the Polycomb repressive complex PR-DUB. *Nature* 465, 243–247.

Schmitges, F.W., Prusty, A.B., Faty, M., Stützer, A., Lingaraju, G.M., Aiwazian, J., Sack, R., Hess, D., Li, L., Zhou, S., et al. (2011). Histone Methylation by PRC2 Is Inhibited by Active Chromatin Marks. *Mol. Cell* 42, 330–341.

Schorderet, P., and Duboule, D. (2011). Structural and Functional Differences in the Long Non-Coding RNA *Hotair* in Mouse and Human. *PLoS Genet.* 7.

Schotta, G., Lachner, M., Sarma, K., Ebert, A., Sengupta, R., Reuter, G., Reinberg, D., and Jenuwein, T. (2004). A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. *Genes Dev.* 18, 1251–1262.

Schubert, D., Primavesi, L., Bishopp, A., Roberts, G., Doonan, J., Jenuwein, T., and Goodrich, J. (2006). Silencing by plant Polycomb-group genes requires dispersed trimethylation of histone H3 at lysine 27. *EMBO J.* 25, 4638–4649.

Schuettengruber, B., and Cavalli, G. (2013). Polycomb Domain Formation Depends on Short and Long Distance Regulatory Cues. *PLoS ONE* 8.

Schuettengruber, B., Ganapathi, M., Leblanc, B., Portoso, M., Jaschek, R., Tolhuis, B., van Lohuizen, M., Tanay, A., and Cavalli, G. (2009). Functional anatomy of polycomb and trithorax chromatin landscapes in *Drosophila* embryos. *PLoS Biol.* 7, e13.

- Schuettengruber, B., Bourbon, H.-M., Di Croce, L., and Cavalli, G. (2017). Genome Regulation by Polycomb and Trithorax: 70 Years and Counting. *Cell* *171*, 34–57.
- Schwartz, Y.B., and Pirrotta, V. (2008). Polycomb complexes and epigenetic states. *Curr. Opin. Cell Biol.* *20*, 266–273.
- Schwartz, Y.B., and Pirrotta, V. (2014). Ruled by Ubiquitylation: A New Order for Polycomb Recruitment. *Cell Rep.* *8*, 321–325.
- Schwartz, Y.B., Kahn, T.G., Nix, D.A., Li, X.-Y., Bourgon, R., Biggin, M., and Pirrotta, V. (2006). Genome-wide analysis of Polycomb targets in *Drosophila melanogaster*. *Nat. Genet.* *38*, 700–705.
- Schwartz, Y.B., Kahn, T.G., Stenberg, P., Ohno, K., Bourgon, R., and Pirrotta, V. (2010). Alternative Epigenetic Chromatin States of Polycomb Target Genes. *PLoS Genet.* *6*.
- Schwartz, Y.B., Linder-Basso, D., Kharchenko, P.V., Tolstorukov, M.Y., Kim, M., Li, H.-B., Gorchakov, A.A., Minoda, A., Shanower, G., Alekseyenko, A.A., et al. (2012). Nature and function of insulator protein binding sites in the *Drosophila* genome. *Genome Res.* *22*, 2188–2198.
- Sekibo, D.A.T., and Fox, K.R. (2017). The effects of DNA supercoiling on G-quadruplex formation. *Nucleic Acids Res.* *45*, 12069–12079.
- Sentürk Cetin, N., Kuo, C.-C., Ribarska, T., Li, R., Costa, I.G., and Grummt, I. (2019). Isolation and genome-wide characterization of cellular DNA:RNA triplex structures. *Nucleic Acids Res.* *47*, 2306–2321.
- Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A., and Cavalli, G. (2012). Three-Dimensional Folding and Functional Organization Principles of the *Drosophila* Genome. *Cell* *148*, 458–472.
- Shafiq, S., Chen, C., Yang, J., Cheng, L., Ma, F., Widemann, E., and Sun, Q. (2017). DNA Topoisomerase 1 Prevents R-loop Accumulation to Modulate Auxin-Regulated Root Development in Rice. *Mol. Plant* *10*, 821–833.
- Shaiu, W.-L., and Hsieh, T. (1998). Targeting to Transcriptionally Active Loci by the Hydrophilic N-Terminal Domain of *Drosophila* DNA Topoisomerase I. *Mol. Cell. Biol.* *18*, 4358–4367.
- Shao, Z., Raible, F., Mollaaghababa, R., Guyon, J.R., Wu, C., Bender, W., and Kingston, R.E. (1999). Stabilization of Chromatin Structure by PRC1, a Polycomb Complex. *Cell* *98*, 37–46.

Shaw, N.N., and Arya, D.P. (2008). Recognition of the unique structure of DNA:RNA hybrids. *Biochimie* 90, 1026–1039.

Shen, X., Liu, Y., Hsu, Y.-J., Fujiwara, Y., Kim, J., Mao, X., Yuan, G.-C., and Orkin, S.H. (2008). EZH1 mediates methylation on histone H3 lysine 27 and complements EZH2 in maintaining stem cell identity and executing pluripotency. *Mol. Cell* 32, 491–502.

Shi, Y., Ma, H., Zhuang, Y., Wang, X., Jiang, Y., and Xu, H.E. (2019). C10ORF12 modulates PRC2 histone methyltransferase activity and H3K27me3 levels. *Acta Pharmacol. Sin.* 40, 1457–1465.

Shykind, B.M., Kim, J., Stewart, L., Champoux, J.J., and Sharp, P.A. (1997). Topoisomerase I enhances TFIIID-TFIIA complex assembly during activation of transcription. *Genes Dev.* 11, 397–407.

Sikorska, N., and Sexton, T. (2020). Defining Functionally Relevant Spatial Chromatin Domains: It is a TAD Complicated. *J. Mol. Biol.* 432, 653–664.

Simon, J.A., and Kingston, R.E. (2009). Mechanisms of Polycomb gene silencing: knowns and unknowns. *Nat. Rev. Mol. Cell Biol.* 10, 697–708.

Simon, J., Chiang, A., Bender, W., Shimell, M.J., and O'Connor, M. (1993). Elements of the *Drosophila* bithorax complex that mediate repression by Polycomb group products. *Dev. Biol.* 158, 131–144.

Simpson, R.T., Thoma, F., and Brubaker, J.M. (1985). Chromatin reconstituted from tandemly repeated cloned DNA fragments and core histones: A model system for study of higher order structure. *Cell* 42, 799–808.

Sing, A., Pannell, D., Karaiskakis, A., Sturgeon, K., Djabali, M., Ellis, J., Lipshitz, H.D., and Cordes, S.P. (2009). A Vertebrate Polycomb Response Element Governs Segmentation of the Posterior Hindbrain. *Cell* 138, 885–897.

Singh, D., Mallon, J., Poddar, A., Wang, Y., Tippana, R., Yang, O., Bailey, S., and Ha, T. (2018). Real-time observation of DNA target interrogation and product release by the RNA-guided endonuclease CRISPR Cpf1 (Cas12a). *Proc. Natl. Acad. Sci. U. S. A.* 115, 5444–5449.

Skourti-Stathaki, K., Proudfoot, N.J., and Gromak, N. (2011). Human Senataxin Resolves RNA/DNA Hybrids Formed at Transcriptional Pause Sites to Promote Xrn2-Dependent Termination. *Mol. Cell* 42, 794–805.

Skourti-Stathaki, K., Kamieniarz-Gdula, K., and Proudfoot, N.J. (2014). R-loops induce repressive chromatin marks over mammalian gene terminators. *Nature* 516, 436–439.

Skourti-Stathaki, K., Torlai Triglia, E., Warburton, M., Voigt, P., Bird, A., and Pombo, A. (2019). R-Loops Enhance Polycomb Repression at a Subset of Developmental Regulator Genes. *Mol. Cell* 73, 930-945.e4.

Sneppen, K., and Ringrose, L. (2019). Theoretical analysis of Polycomb-Trithorax systems predicts that poised chromatin is bistable and not bivalent. *Nat. Commun.* 10, 1–18.

Sollier, J., Stork, C.T., García-Rubio, M.L., Paulsen, R.D., Aguilera, A., and Cimprich, K.A. (2014). Transcription-Coupled Nucleotide Excision Repair Factors Promote R-Loop-Induced Genome Instability. *Mol. Cell* 56, 777–785.

Son, J., Shen, S.S., Margueron, R., and Reinberg, D. (2013). Nucleosome-binding activities within JARID2 and EZH1 regulate the function of PRC2 on chromatin. *Genes Dev.* 27, 2663–2677.

Song, J., Angel, A., Howard, M., and Dean, C. (2012). Vernalization – a cold-induced epigenetic switch. *J. Cell Sci.* 125, 3723–3731.

Sparks, J.L., Chon, H., Cerritelli, S.M., Kunkel, T.A., Johansson, E., Crouch, R.J., and Burgers, P.M. (2012). RNase H2-Initiated Ribonucleotide Excision Repair. *Mol. Cell* 47, 980–986.

Spies, M. (2014). Fulfilling the dream of a perfect genome editing tool. *Proc. Natl. Acad. Sci.* 111, 10029–10030.

Srivenugopal, K.S., Lockshon, D., and Morris, D.R. (1984). Escherichia coli DNA topoisomerase III: purification and characterization of a new type I enzyme. *Biochemistry* 23, 1899–1906.

Staker, B.L., Hjerrild, K., Feese, M.D., Behnke, C.A., Burgin, A.B., and Stewart, L. (2002a). The mechanism of topoisomerase I poisoning by a camptothecin analog. *Proc. Natl. Acad. Sci.* 99, 15387–15392.

Staker, B.L., Hjerrild, K., Feese, M.D., Behnke, C.A., Burgin, A.B., and Stewart, L. (2002b). The mechanism of topoisomerase I poisoning by a camptothecin analog. *Proc. Natl. Acad. Sci. U. S. A.* 99, 15387–15392.

Stark, W.M. (2014). The Serine Recombinases. *Microbiol. Spectr.* 2.

Stark, W. m., Grindley, N. d., Hatfull, G. f., and Boocock, M. r. (1991). Resolvase-catalysed reactions between res sites differing in the central dinucleotide of subsite I. *EMBO J.* 10, 3541–3548.

Steffen, P.A., and Ringrose, L. (2014). What are memories made of? How Polycomb and Trithorax proteins mediate epigenetic memory. *Nat. Rev. Mol. Cell Biol.* 15, 340–356.

Sternberg, S.H., Redding, S., Jinek, M., Greene, E.C., and Doudna, J.A. (2014). DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* *507*, 62–67.

Stewart, L., Ireton, G.C., and Champoux, J.J. (1997). Reconstitution of human topoisomerase I by fragment complementation. *J. Mol. Biol.* *269*, 355–372.

Stolz, R., Sulthana, S., Hartono, S.R., Malig, M., Benham, C.J., and Chedin, F. (2019). Interplay between DNA sequence and negative superhelicity drives R-loop structures. *Proc. Natl. Acad. Sci. U. S. A.* *116*, 6260–6269.

Streubel, G., Watson, A., Jammula, S.G., Scelfo, A., Fitzpatrick, D.J., Oliviero, G., McCole, R., Conway, E., Glancy, E., Negri, G.L., et al. (2018). The H3K36me2 Methyltransferase Nsd1 Demarcates PRC2-Mediated H3K27me2 and H3K27me3 Domains in Embryonic Stem Cells. *Mol. Cell* *70*, 371-379.e5.

Sun, D., and Hurley, L.H. (2009). The importance of negative superhelicity in inducing the formation of G-quadruplex and i-motif structures in the c-Myc promoter: implications for drug targeting and control of gene expression. *J. Med. Chem.* *52*, 2863–2874.

Sun, Q., Csorba, T., Skourti-Stathaki, K., Proudfoot, N.J., and Dean, C. (2013). R-Loop Stabilization Represses Antisense Transcription at the Arabidopsis FLC Locus. *Science* *340*, 619–621.

Suski, C., and Marians, K.J. (2008). Resolution of Converging Replication Forks by RecQ and Topoisomerase III. *Mol. Cell* *30*, 779–789.

Swedlow, J.R., Sedat, J.W., and Agard, D.A. (1993). Multiple chromosomal populations of topoisomerase II detected in vivo by time-lapse, three-dimensional wide-field microscopy. *Cell* *73*, 97–108.

Swiezewski, S., Liu, F., Magusin, A., and Dean, C. (2009). Cold-induced silencing by long antisense transcripts of an Arabidopsis Polycomb target. *Nature* *462*, 799–802.

Szczelkun, M.D., Tikhomirova, M.S., Sinkunas, T., Gasiunas, G., Karvelis, T., Pschera, P., Siksnys, V., and Seidel, R. (2014). Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes. *Proc. Natl. Acad. Sci. U. S. A.* *111*, 9798–9803.

Szilard, R.K., Jacques, P.-E., Laramée, L., Cheng, B., Galicia, S., Bataille, A.R., Yeung, M., Mendez, M., Bergeron, M., Robert, F., et al. (2010). Systematic identification of fragile sites via genome-wide location analysis of gamma-H2AX. *Nat. Struct. Mol. Biol.* *17*, 299–305.

Tang, S., Wu, M.K.Y., Zhang, R., and Hunter, N. (2015). Pervasive and essential roles of the Top3-Rmi1 decatenase orchestrate recombination and facilitate chromosome segregation in meiosis. *Mol. Cell* 57, 607–621.

Tan-Wong, S.M., Dhir, S., and Proudfoot, N.J. (2019). R-Loops Promote Antisense Transcription across the Mammalian Genome. *Mol. Cell* 76, 600-616.e6.

Taylor, M.R., Conrad, J.A., Wahl, D., and O'Brien, P.J. (2011). Kinetic Mechanism of Human DNA Ligase I Reveals Magnesium-dependent Changes in the Rate-limiting Step That Compromise Ligation Efficiency. *J. Biol. Chem.* 286, 23054–23062.

Teves, S.S., and Henikoff, S. (2014). Transcription-generated torsional stress destabilizes nucleosomes. *Nat. Struct. Mol. Biol.* 21, 88–94.

Thys, R.G., Lehman, C.E., Pierce, L.C.T., and Wang, Y.-H. (2015). DNA Secondary Structure at Chromosomal Fragile Sites in Human Disease. *Curr. Genomics* 16, 60–70.

Tie, F., Prasad-Sinha, J., Birve, A., Rasmuson-Lestander, Å., and Harte, P.J. (2003). A 1-Megadalton ESC/E(Z) Complex from *Drosophila* That Contains Polycomblike and RPD3. *Mol. Cell. Biol.* 23, 3352–3362.

Tie, F., Siebold, A.P., and Harte, P.J. (2005). The N-terminus of *Drosophila* ESC mediates its phosphorylation and dimerization. *Biochem. Biophys. Res. Commun.* 332, 622–632.

Tie, F., Banerjee, R., Stratton, C.A., Prasad-Sinha, J., Stepanik, V., Zlobin, A., Diaz, M.O., Scacheri, P.C., and Harte, P.J. (2009). CBP-mediated acetylation of histone H3 lysine 27 antagonizes *Drosophila* Polycomb silencing. *Dev. Camb. Engl.* 136, 3131–3141.

Tolhuis, B., de Wit, E., Muijers, I., Teunissen, H., Talhout, W., van Steensel, B., and van Lohuizen, M. (2006). Genome-wide profiling of PRC1 and PRC2 Polycomb chromatin binding in *Drosophila melanogaster*. *Nat. Genet.* 38, 694–699.

Toubiana, S., and Selig, S. (2018). DNA:RNA hybrids at telomeres – when it is better to be out of the (R) loop. *FEBS J.* 285, 2552–2566.

Tsai, M.-C., Manor, O., Wan, Y., Mosammamaparast, N., Wang, J.K., Lan, F., Shi, Y., Segal, E., and Chang, H.Y. (2010). Long Noncoding RNA as Modular Scaffold of Histone Modification Complexes. *Science* 329, 689–693.

Tse, Y.C., Kirkegaard, K., and Wang, J.C. (1980). Covalent bonds between protein and DNA. Formation of phosphotyrosine linkage between certain DNA topoisomerases and DNA. *J. Biol. Chem.* 255, 5560–5565.

Tuckfield, A., Clouston, D.R., Wilanowski, T.M., Zhao, L.-L., Cunningham, J.M., and Jane, S.M. (2002). Binding of the RING polycomb proteins to specific target genes in complex with the grainyhead-like family of developmental transcription factors. *Mol. Cell. Biol.* 22, 1936–1946.

Van Bortle, K., Ramos, E., Takenaka, N., Yang, J., Wahi, J.E., and Corces, V.G. (2012). *Drosophila* CTCF tandemly aligns with other insulator proteins at the borders of H3K27me3 domains. *Genome Res.* 22, 2176–2187.

Van Dyck, E., Hajibagheri, N.M.A., Stasiak, A., and West, S.C. (1998). Visualisation of human rad52 protein and its complexes with hrad51 and DNA11 Edited by J. Karn. *J. Mol. Biol.* 284, 1027–1038.

Vas, A.C.J., Andrews, C.A., Kirkland Matesky, K., and Clarke, D.J. (2007). In vivo analysis of chromosome condensation in *Saccharomyces cerevisiae*. *Mol. Biol. Cell* 18, 557–568.

Vasanthi, D., Nagabhushan, A., Matharu, N.K., and Mishra, R.K. (2013). A functionally conserved Polycomb response element from mouse HoxD complex responds to heterochromatin factors. *Sci. Rep.* 3, 1–11.

Vashee, S., Cvetic, C., Lu, W., Simancek, P., Kelly, T.J., and Walter, J.C. (2003). Sequence-independent DNA binding and replication initiation by the human origin recognition complex. *Genes Dev.* 17, 1894–1908.

Venkatesh, S., and Workman, J.L. (2015). Histone exchange, chromatin structure and the regulation of transcription. *Nat. Rev. Mol. Cell Biol.* 16, 178–189.

Vrielynck, N., Chambon, A., Vezon, D., Pereira, L., Chelysheva, L., Muyt, A.D., Mézard, C., Mayer, C., and Grelon, M. (2016). A DNA topoisomerase VI-like complex initiates meiotic recombination. *Science* 351, 939–943.

Wahba, L., Costantino, L., Tan, F.J., Zimmer, A., and Koshland, D. (2016). S1-DRIP-seq identifies high expression and polyA tracts as major contributors to R-loop formation. *Genes Dev.* 30, 1327–1338.

Walker, E., Chang, W.Y., Hunkapiller, J., Cagney, G., Garcha, K., Torchia, J., Krogan, N.J., Reiter, J.F., and 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, 153–166.

Wang, J.C. (1971). Interaction between DNA and an *Escherichia coli* protein ω . *J. Mol. Biol.* 55, 523–IN16.

- Wang, J.C. (2002). Cellular roles of DNA topoisomerases: a molecular perspective. *Nat. Rev. Mol. Cell Biol.* *3*, 430–440.
- Wang, Q., and Shen, W.-H. (2018). Chromatin modulation and gene regulation in plants: insight about PRC1 function. *Biochem. Soc. Trans.* *46*, 957–966.
- Wang, A.H.-J., Fujii, S., van Boom, J.H., van der Marel, G.A., van Boeckel, S.A.A., and Rich, A. (1982). Molecular structure of r(GCG)_n(TATACGC): a DNA–RNA hybrid helix joined to double helical DNA. *Nature* *299*, 601–604.
- Wang, H., Di Gate, R.J., and Seeman, N.C. (1996). An RNA topoisomerase. *Proc. Natl. Acad. Sci. U. S. A.* *93*, 9477–9482.
- Wang, H., Wang, L., Erdjument-Bromage, H., Vidal, M., Tempst, P., Jones, R.S., and Zhang, Y. (2004a). Role of histone H2A ubiquitination in Polycomb silencing. *Nature* *431*, 873–878.
- Wang, H., Liu, C., Cheng, J., Liu, J., Zhang, L., He, C., Shen, W.-H., Jin, H., Xu, L., and Zhang, Y. (2016a). Arabidopsis Flower and Embryo Developmental Genes are Repressed in Seedlings by Different Combinations of Polycomb Group Proteins in Association with Distinct Sets of Cis-regulatory Elements. *PLoS Genet.* *12*.
- Wang, I.X., Grunseich, C., Fox, J., Burdick, J., Zhu, Z., Ravazian, N., Hafner, M., and Cheung, V.G. (2018). Human proteins that interact with RNA/DNA hybrids. *Genome Res.* *28*, 1405–1414.
- Wang, J., Jia, S.T., and Jia, S. (2016b). New insights into the regulation of heterochromatin. *Trends Genet.* *TIG 32*, 284–294.
- Wang, L., Brown, J.L., Cao, R., Zhang, Y., Kassis, J.A., and Jones, R.S. (2004b). Hierarchical recruitment of polycomb group silencing complexes. *Mol. Cell* *14*, 637–646.
- Wang, L., Jahren, N., Vargas, M.L., Andersen, E.F., Benes, J., Zhang, J., Miller, E.L., Jones, R.S., and Simon, J.A. (2006). Alternative ESC and ESC-Like Subunits of a Polycomb Group Histone Methyltransferase Complex Are Differentially Deployed during Drosophila Development. *Mol. Cell. Biol.* *26*, 2637–2647.
- Wang, L., Jahren, N., Miller, E.L., Ketel, C.S., Mallin, D.R., and Simon, J.A. (2010). Comparative Analysis of Chromatin Binding by Sex Comb on Midleg (SCM) and Other Polycomb Group Repressors at a Drosophila Hox Gene. *Mol. Cell. Biol.* *30*, 2584–2593.

Wang, X., Paucek, R.D., Gooding, A.R., Brown, Z.Z., Ge, E.J., Muir, T.W., and Cech, T.R. (2017a). Molecular analysis of PRC2 recruitment to DNA in chromatin and its inhibition by RNA. *Nat. Struct. Mol. Biol.*

Wang, X., Goodrich, K.J., Gooding, A.R., Naeem, H., Archer, S., Paucek, R.D., Youmans, D.T., Cech, T.R., and Davidovich, C. (2017b). Targeting of Polycomb Repressive Complex 2 to RNA by Short Repeats of Consecutive Guanines. *Mol. Cell* 65, 1056-1067.e5.

Wang, Y., Gu, X., Yuan, W., Schmitz, R.J., and He, Y. (2014). Photoperiodic Control of the Floral Transition through a Distinct Polycomb Repressive Complex. *Dev. Cell* 28, 727–736.

Wani, A.H., Boettiger, A.N., Schorderet, P., Ergun, A., Münger, C., Sadreyev, R.I., Zhuang, X., Kingston, R.E., and Francis, N.J. (2016). Chromatin topology is coupled to Polycomb group protein subnuclear organization. *Nat. Commun.* 7.

Wasserman, S.A., Dungan, J.M., and Cozzarelli, N.R. (1985). Discovery of a predicted DNA knot substantiates a model for site-specific recombination. *Science* 229, 171–174.

Woo, C.J., Kharchenko, P.V., Daheron, L., Park, P.J., and Kingston, R.E. (2010). A Region of the Human HOXD Cluster that Confers Polycomb-Group Responsiveness. *Cell* 140, 99–110.

Wood, C.C., Robertson, M., Tanner, G., Peacock, W.J., Dennis, E.S., and Helliwell, C.A. (2006). The *Arabidopsis thaliana* vernalization response requires a polycomb-like protein complex that also includes VERNALIZATION INSENSITIVE 3. *Proc. Natl. Acad. Sci.* 103, 14631–14636.

Wright, E.P., Day, H.A., Ibrahim, A.M., Kumar, J., Boswell, L.J.E., Huguin, C., Stevenson, C.E.M., Pors, K., and Waller, Z.A.E. (2016). Mitoxantrone and Analogues Bind and Stabilize i-Motif Forming DNA Sequences. *Sci. Rep.* 6.

Wu, C.T., and Howe, M. (1995). A genetic analysis of the Suppressor 2 of zeste complex of *Drosophila melanogaster*. *Genetics* 140, 139–181.

Wu, H.-Y., Shyy, S., Wang, J.C., and Liu, L.F. (1988). Transcription generates positively and negatively supercoiled domains in the template. *Cell* 53, 433–440.

Wu, L., Zee, B.M., Wang, Y., Garcia, B.A., and Dou, Y. (2011). The RING Finger Protein MSL2 in the MOF Complex is an E3 Ubiquitin Ligase for H2B K34 and is Involved in Crosstalk with H3 K4 and K79 Methylation. *Mol. Cell* 43, 132–144.

Wu, N., Wu, X.-W., Agama, K., Pommier, Y., Du, J., Li, D., Gu, L.-Q., Huang, Z.-S., and An, L.-K. (2010). A novel DNA topoisomerase I inhibitor with different mechanism from camptothecin induces G2/M phase cell cycle arrest to K562 cells. *Biochemistry* 49, 10131–10136.

Xu, D., Shen, W., Guo, R., Xue, Y., Peng, W., Sima, J., Yang, J., Sharov, A., Srikantan, S., Yang, J., et al. (2013). Top3 β is an RNA topoisomerase that works with Fragile X syndrome protein to promote synapse formation. *Nat. Neurosci.* 16, 1238–1247.

Xu, W., Xu, H., Li, K., Fan, Y., Liu, Y., Yang, X., and Sun, Q. (2017). The R-loop is a common chromatin feature of the Arabidopsis genome. *Nat. Plants* 3, 704–714.

Xu, W., Li, K., Li, S., Hou, Q., Zhang, Y., Liu, K., and Sun, Q. (2020). The R-Loop Atlas of Arabidopsis Development and Responses to Environmental Stimuli[OPEN]. *Plant Cell* 32, 888–903.

Yadav, P., Harcy, V., Argueso, J.L., Dominska, M., Jinks-Robertson, S., and Kim, N. (2014). Topoisomerase I Plays a Critical Role in Suppressing Genome Instability at a Highly Transcribed G-Quadruplex-Forming Sequence. *PLoS Genet.* 10.

Yan, J., Dutta, B., Hee, Y.T., and Chng, W.-J. (2019). Towards understanding of PRC2 binding to RNA. *RNA Biol.* 16, 176–184.

Yang, W. (2010). Topoisomerases and site-specific recombinases: similarities in structure and mechanism. *Crit. Rev. Biochem. Mol. Biol.* 45, 520–534.

Yang, S.W., and Chan, J.Y. (1992). Analysis of the formation of AMP-DNA intermediate and the successive reaction by human DNA ligases I and II. *J. Biol. Chem.* 267, 8117–8122.

Yang, H., Howard, M., and Dean, C. (2014). Antagonistic Roles for H3K36me3 and H3K27me3 in the Cold-Induced Epigenetic Switch at Arabidopsis FLC. *Curr. Biol.* 24, 1793–1797.

Yang, W., Lee, J.Y., and Nowotny, M. (2006). Making and Breaking Nucleic Acids: Two-Mg²⁺-Ion Catalysis and Substrate Specificity. *Mol. Cell* 22, 5–13.

Yang, X., Liu, Q.-L., Xu, W., Zhang, Y.-C., Yang, Y., Ju, L.-F., Chen, J., Chen, Y.-S., Li, K., Ren, J., et al. (2019). m⁶A promotes R-loop formation to facilitate transcription termination. *Cell Res.* 29, 1035–1038.

Yasuhara, T., Kato, R., Hagiwara, Y., Shiotani, B., Yamauchi, M., Nakada, S., Shibata, A., and Miyagawa, K. (2018). Human Rad52 Promotes XPG-Mediated R-loop Processing to Initiate Transcription-Associated Homologous Recombination Repair. *Cell* 175, 558-570.e11.

Yu, D., Khan, E., Khaleque, M.A., Lee, J., Laco, G., Kohlhagen, G., Kharbanda, S., Cheng, Y.-C., Pommier, Y., and Bharti, A. (2004). Phosphorylation of DNA Topoisomerase I by the c-Abl Tyrosine Kinase Confers Camptothecin Sensitivity. *J. Biol. Chem.* *279*, 51851–51861.

Yu, J.-R., Lee, C.-H., Oksuz, O., Stafford, J.M., and Reinberg, D. (2019). PRC2 is high maintenance. *Genes Dev.* *33*, 903–935.

Yu, K., Chedin, F., Hsieh, C.-L., Wilson, T.E., and Lieber, M.R. (2003). R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells. *Nat. Immunol.* *4*, 442–451.

Yuan, W., Xu, M., Huang, C., Liu, N., Chen, S., and Zhu, B. (2011). H3K36 Methylation Antagonizes PRC2-mediated H3K27 Methylation. *J. Biol. Chem.* *286*, 7983–7989.

Zeman, M.K., and Cimprich, K.A. (2014). Causes and consequences of replication stress. *Nat. Cell Biol.* *16*, 2–9.

Zeng, J., Kirk, B.D., Gou, Y., Wang, Q., and Ma, J. (2012). Genome-wide polycomb target gene prediction in *Drosophila melanogaster*. *Nucleic Acids Res.* *40*, 5848–5863.

Zhang, H., Wang, J.C., and Liu, L.F. (1988). Involvement of DNA topoisomerase I in transcription of human ribosomal RNA genes. *Proc. Natl. Acad. Sci.* *85*, 1060–1064.

Zhang, H., Barceló, J.M., Lee, B., Kohlhagen, G., Zimonjic, D.B., Popescu, N.C., and Pommier, Y. (2001). Human mitochondrial topoisomerase I. *Proc. Natl. Acad. Sci.* *98*, 10608–10613.

Zhang, Q., McKenzie, N.J., Warneford-Thomson, R., Gail, E.H., Flanigan, S.F., Owen, B.M., Lauman, R., Levina, V., Garcia, B.A., Schittenhelm, R.B., et al. (2019a). RNA exploits an exposed regulatory site to inhibit the enzymatic activity of PRC2. *Nat. Struct. Mol. Biol.* *26*, 237–247.

Zhang, T., Cooper, S., and Brockdorff, N. (2015). The interplay of histone modifications – writers that read. *EMBO Rep.* *16*, 1467–1481.

Zhang, T., Wallis, M., Petrovic, V., Challis, J., Kalitsis, P., and Hudson, D.F. (2019b). Loss of TOP3B leads to increased R-loop formation and genome instability. *Open Biol.* *9*.

Zhang, Z., Jones, A., Sun, C.-W., Li, C., Chang, C.-W., Joo, H.-Y., Dai, Q., Mysliwiec, M.R., Wu, L.-C., Guo, Y., et al. (2011). PRC2 Complexes with JARID2, MTF2, and esPRC2p48 in ES Cells to Modulate ES Cell Pluripotency and Somatic Cell Reprogramming. *Stem Cells Dayt. Ohio* *29*, 229–240.

Zhao, J., Ohsumi, T.K., Kung, J.T., Ogawa, Y., Grau, D.J., Sarma, K., Song, J.J., Kingston, R.E., Borowsky, M., and Lee, J.T. (2010). Genome-wide identification of Polycomb-associated RNAs by RIP-seq. *Mol. Cell* *40*, 939–953.

Zhu, Z.-M., Huo, F.-C., and Pei, D.-S. (2020). Function and evolution of RNA N6-methyladenosine modification. *Int. J. Biol. Sci.* *16*, 1929–1940.