bioRxiv preprint first posted online May. 15, 2019; doi: http://dx.doi.org/10.1101/635722. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

1 # RNA strand invasion activity of the Polycomb complex PRC2
2

3 **Authors:** Célia Alecki[1,2], Victoria Chiwara[1], Lionel A. Sanz[3], Daniel Grau[4], Osvaldo Arias
4 Pérez[1,5], Karim-Jean Armache[4], Fréderic Chédin[3], and Nicole J. Francis[1,2,6*]

5 **Affiliations:**

6 [1]Institut de recherches cliniques de Montréal, 110 Avenue des Pins Ouest, Montréal, QC, H2W
7 1R7, Canada
8 [2]Département de biochimie et médecine moléculaire Université de Montréal, 2900 Boulevard
9 Edouard-Montpetit, Montréal, QC H3T 1J4, Canada.
10 [3]Department of Molecular and Cellular Biology and Genome Center, 1 Shields Avenue,
11 University of California, Davis, Davis, CA 95616, USA.
12 [4]Skirball Institute of Biomolecular Medicine, Department of Biochemistry and Molecular
13 Pharmacology, New York University School of Medicine, New York, NY 10016, USA.
14 [5]Natural Sciences and Engineering Postgraduate; Universidad Autonoma Metropolitana,
15 Cuajimalpa, Mexico City, Mexico
16 [6]Division of Experimental Medicine, McGill University, 1001 Decarie Boulevard, Montreal, QC
17 H4A 3J1, Canada
18 *correspondence to: nicole.francis@ircm.qc.ca
19

20 **Abstract:** Epigenetic regulation is conveyed through information encoded by specific chromatin
21 features. Non-canonical nucleic acid structures could in principle also convey biological
22 information but their role(s) in epigenetic regulation is not known. Polycomb Group (PcG)
23 proteins form memory of transient transcriptional repression events that is necessary for
24 development. In *Drosophila,* PcG proteins are recruited to specific DNA sequences, Polycomb
25 Response Elements (PREs). PREs are switchable memory elements that can exist in repressed,
26 active, or unengaged states [1,2]. How PcG activities are targeted to PREs to maintain repressed
27 states only in appropriate developmental contexts has been difficult to elucidate. Biochemically,
28 PcG protein complexes modify chromatin to maintain gene repression [1,3,4]. However, PcG
29 proteins also interact with both RNA and DNA, and RNA is implicated in the targeting of PcG
30 function. We find that R-loops, three-stranded nucleic acid structures formed when an RNA
31 hybridizes to its complementary DNA and displaces the other DNA strand [5], form at many PREs
32 in *Drosophila* embryos, and correlate with the repressive state. R-loops are recognized by the
33 PcG complex PRC1 in vitro. Unexpectedly, we find that the PcG complex PRC2 has RNA strand
34 invasion activity, which can drive formation of RNA-DNA hybrids, the key component of R-
35 loops. Our results suggest a new mechanism for targeting PcG function through R-loop
36 formation by PRC2 and recognition by PRC1. More generally, our findings suggest formation
37 and recognition [6] of non-canonical nucleic acid structures as an epigenetic mechanism.
38

39

40

41

bioRxiv preprint first posted online May. 15, 2019; doi: http://dx.doi.org/10.1101/635722. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

42 **Main Text:**

43 During *Drosophila* embryogenesis, transiently expressed transcription factors activate
44 homeotic (*Hox*) genes in certain regions of the embryo and repress them in others to dictate the
45 future body plan [7]. Polycomb Group (PcG) proteins form a memory of these early cues by
46 maintaining patterns of *Hox* gene repression for the rest of development [1,7,8]. This paradigm for
47 transcriptional memory is believed to be used by the PcG at many genes in *Drosophila*, and to
48 underlie the conserved and essential functions of PcG proteins in cell differentiation and
49 development from plants to mammals [9,10]. Polycomb Response Elements (PREs) are DNA
50 elements that can recruit PcG proteins, but they also recapitulate the memory function of the PcG—
51 when combined with early acting, region-specific enhancers in transgenes, they maintain transgene
52 repression in a PcG-dependent manner only in regions where the early enhancer was not active
53 [1,11,12]. PREs contain a high density of binding sites for transcription factors that can recruit PcG
54 proteins through physical interactions [12]. However, the widespread expression, binding pattern,
55 and properties of factors that bind PREs cannot explain how PREs can exist in alternate,
56 transcription-history dependent states to maintain restricted patterns of gene expression, or how
57 they can switch between states [1]. Furthermore, DNA sequences with PRE-like properties have
58 been difficult to identify in other species [12-14] despite the conservation of PcG complexes, their
59 biochemical activities, and their critical roles in development.

60 RNAs may provide context specificity to PcG protein recruitment and function. Some
61 PREs, and some PcG binding sites in mammalian and plant cells, are transcribed into ncRNA,
62 while others reside in gene bodies, and thus are transcribed when the gene is expressed [15,16]. Both
63 the direction and level of transcription have been correlated with the functional state of PREs [15-
64 17]. The PcG complex Polycomb Repressive Complex 2 (PRC2) has a well-described high affinity

bioRxiv preprint first posted online May. 15, 2019; doi: http://dx.doi.org/10.1101/635722. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

65 for RNA [18-22]. RNA is suggested to recruit PRC2 to specific chromatin sites [18], but RNA binding
66 can also compete for chromatin binding and inhibit PRC2 activity [16,22-24]. One way for RNA to
67 interact with the genome is by the formation of R-loops, three-stranded nucleic acid structures
68 formed when an RNA hybridizes to a complementary DNA strand, thereby displacing the second
69 DNA strand [5]. The formation of R-loops over genes with low to moderate expression is associated
70 with increased PcG binding and H3K27 trimethylation (H3K27me3) in human cells [25] and R-loops
71 have been implicated in promoting PcG recruitment in mammalian cells [6], although other evidence
72 suggests they antagonize recruitment [26]. We hypothesized that R-loop formation could
73 biochemically link RNA to silencing through PREs and tested this idea in the *Drosophila* system.

74 To determine whether R-loops form at PREs, we carried out two biological replicates of
75 strand-specific DNA-RNA Immunoprecipitation followed by next generation sequencing (DRIP-
76 seq) in *Drosophila* embryos (2-6 and 10-14 hour (H)) and in S2 cells (Fig. 1, Extended Data Fig.
77 1). DRIP-seq peaks called relative to both input and RNaseH-treated control samples and present
78 in both replicates were analyzed. 10 positive sites were validated by DRIP-qPCR (Extended Data
79 Fig. 1b). About two thirds of R-loops form over gene bodies (Extended Data Fig. 1). R-loops are
80 observed over genes encompassing all levels of transcription, although a majority are associated
81 with genes with no or low levels of expression (Extended Data Fig. 2a, b). Most R-loops form
82 with the strandedness expected from annotated transcripts (Fig. 1 a, c, Extended Data Fig. 2c), as
83 observed in other species [25,27,28].

84 We detect R-loops at 22-33% of PREs (Fig. 1 a-c, Extended Data Fig. 1d, 2 a-d). R-loops
85 at PREs are more likely to have an antisense orientation to annotated transcripts, or have no
86 overlapping annotated transcript, than total R-loops (Extended Data Fig. 2c). To test whether R-
87 loops are related to the functional state of PREs, we compared PcG protein binding at R-loops

bioRxiv preprint first posted online May. 15, 2019; doi: http://dx.doi.org/10.1101/635722. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

88 that do or do not form R-loops in each of our three samples. For each PcG protein tested, the
89 median read density over PREs with R-loops is higher than that for PREs without R-loops (Fig.
90 1d-f, Extended Data Fig. 2e-g). Although binding of PcG proteins to PREs is necessary for their
91 repressive function, it may not be sufficient, since analyses of PcG protein binding at a small
92 number of PREs in the ON and OFF states did not detect differences in PRC1 or PRC2 binding
93 [29,30]. Instead, histone modifications at and around PREs are correlated with the functional state so
94 that PREs in the OFF state are marked with H3K27me3 [29]. In both developing embryos and S2
95 cells, H3K27me3 density is higher at PREs with R-loops than those without R-loops (Fig. 1e).
96 H3K27Ac, a mark of the active state, is found at a small number of PREs, but correlates weakly
97 with the presence of R-loops (Extended data Fig. 2 d, h-j, 3d-f). To test whether transient
98 presence of an R-loop at PREs predicts the repressed state, we analyzed PREs that form R-loops
99 2-6H embryos that are no longer detected in 10-14H embryos for the presence of H3k27me3 in
100 later embryonic states (12-16H). PREs that formed R-loops in early embryos have a higher
101 density of H3K27me3 at subsequent developmental stages than PREs that do not form R-loops at
102 either stage (Fig. 1f), and are not enriched for H3K27Ac ($p=0.0885$).

103 To understand biochemically how R-loops could promote the repressive state of PREs,
104 we turned to in vitro assays. We prepared R-loops *in vitro* by transcribing templates containing a
105 PRE sequence (Fig. 2a). R-loops are visualized as bands containing radiolabelled RNA that co-
106 migrates with the DNA template, and their identity confirmed by their sensitivity to RNase H
107 and DNase I, and resistance to RNaseA (Fig. 2b, Extended Data Fig. 4a, b). We incubated either
108 of the two main PcG complexes, PRC1 or PRC2, (Extended Data Fig. 4c,d) with transcribed
109 templates, and fractionated the reactions by sucrose gradient sedimentation. Nucleic acids bound
110 by PcG complexes sediment near the bottom of the gradient and unbound nucleic near the top

bioRxiv preprint first posted online May. 15, 2019; doi: http://dx.doi.org/10.1101/635722. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

111 (Fig. 2a). PRC1 binds preferentially to R-loop containing templates while PRC2 shows no
112 preference (Fig. 2c-f, Extended Data Fig. 4e-g). Both complexes bind tightly to RNA (Extended
113 Data Fig. 5a-e).

114 A small increase in R-loops is observed in some experiments in which PRC2 is incubated
115 with transcribed templates, suggesting PRC2 might influence R-loop formation (Extended Data
116 Fig. 5f,g). To test this, we mixed purified radio- or fluorescently-labelled RNA with dsDNA
117 templates and titrated in PRC2 (Fig. 3a-c). We observe PRC2 dose-dependent appearance of a
118 labelled RNA that migrates at the position of dsDNA (Fig. 3d, e, g, Extended Data Fig. 6a, b).
119 These putative strand invasion products form with either the sense or anti-sense RNA, but not
120 with a non-complementary RNA, indicating that base pairing between RNA and DNA is
121 required (Fig. 3d-f). The products are sensitive to RNaseH and resistant to RNaseA, confirming
122 that they contain RNA-DNA hybrids (Fig. 3h, i, Extended Data Fig. 6b). Two control proteins,
123 the transcription factor NFY and the PcG protein Sxc, do not form strand invasion products,
124 although they bind both DNA and RNA (Extended Data Fig. 6d-h). By the end of a 60-minute
125 reaction containing 3 fmol of linear DNA and 1.9 fmol of RNA, as much as 60% of the RNA is
126 incorporated into RNA-DNA hybrids, corresponding to 38% of the DNA molecules (Fig. 4a, b).
127 Strand invasion activity of PRC2 does not require addition of nucleotides, but does require
128 $MgCl_2$ (Extended Data Fig. S6c). Finally, PRC2 strand invasion activity co-fractionates with
129 PRC2 through size exclusion chromatography (Extended Data Fig. 7a-c). We conclude that
130 PRC2 induces RNA strand invasion of dsDNA to produce RNA-DNA hybrid containing
131 structures, the key component of R-loops.
132

bioRxiv preprint first posted online May. 15, 2019; doi: http://dx.doi.org/10.1101/635722. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

133 PRC2-mediated strand invasion could require binding to DNA, to RNA, or to both.
134 Detailed analyses of PRC2 binding to nucleic acids and chromatin are consistent with PRC2
135 making multiple contacts with both substrates [24,31], while functional assays are consistent with a
136 single binding site that can bind chromatin, DNA or RNA but has highest affinity for RNA
137 [16,22,23]. To understand the role of RNA and DNA interactions in PRC2-mediated strand invasion,
138 we manipulated the reaction conditions. Addition of RNA prior to DNA, or increasing the
139 amount of RNA, inhibits the reaction, while adding DNA prior to RNA or increasing DNA has
140 little effect (Fig. 4c-h, Extended Data Fig. 8a-d). These data are consistent with PRC2-DNA
141 interactions being critical for strand invasion, and higher affinity PRC2-RNA interactions
142 competing for them. This proposed mechanism, in which the protein binds the dsDNA template,
143 resembles "inverse RNA strand invasion" described for the repair proteins Rad52 and RecA [32-34]
144 (See Extended Data Fig. 9a-c for models).
145 PRC2 and its methyltransferase activity are conserved from plants to human, as are
146 connections between PRC2 and RNA [9,16-18,35]. We therefore anticipated that strand invasion
147 activity would be conserved. We find that PRC2-EZH1 has strand invasion activity at
148 concentrations that coincide with its DNA binding activity, while PRC2-EZH2 has little activity
149 under these conditions (Fig. 4i, j, Extended Data Fig. 8e).
150 The demonstration that PRC2 has RNA strand invasion activity, that PRC1 can recognize
151 R-loops, and that R-loops are present at PREs in vivo suggest a mechanistic model for how
152 RNAs can promote the off state of PREs through PRC2-driven R-loop formation. R-loops could
153 synergize with PRE-binding proteins to recruit PRC1 and would also sequester the RNA and
154 tether it to the genome, preventing it from competing with the chromatin substrate for PRC2
155 binding. Stabilization of PcG proteins at PREs through R-loop formation would promote

bioRxiv preprint first posted online May. 15, 2019; doi: http://dx.doi.org/10.1101/635722. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

156 chromatin modification through the well-known activities of PRC1 and PRC2 [1,4]. R-loops may
157 also interfere with binding or function of proteins that promote the active state of PREs. Our data
158 indicate that both coding and ncRNAs form R-loops. The regulation of these RNAs and therefore
159 of R-loops could provide transcription history and developmental context specificity to PcG
160 recruitment by transcription factors that constitutively recognize PREs. A conceptually similar
161 model for how high levels of RNA production at PREs could promote the ON state and low
162 levels the OFF state was proposed previously [17], but R-loop formation provides a mechanism by
163 which it can occur.

164 The connection between RNA and PRC2 has been recognized for some time, in species
165 from plants to humans [16-18,35], but mechanisms beyond RNA binding by PRC2 have not
166 previously been described. Our discovery of PRC2-mediated RNA strand invasion, and R-loop
167 formation at PREs, suggests a mechanism to connect RNA to PcG targeting and function, and
168 the formation and recognition [6] of non-canonical nucleic acid structures to epigenetics (Extended
169 Data 10).

170

bioRxiv preprint first posted online May. 15, 2019; doi: http://dx.doi.org/10.1101/635722. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

171 **References:**

172 1 Steffen, P. A. & Ringrose, L. What are memories made of? How Polycomb and Trithorax
173 proteins mediate epigenetic memory. *Nat Rev Mol Cell Biol* **15**, 340-356,
174 doi:10.1038/nrm3789 (2014).

175 2 Schwartz, Y. B. *et al.* Alternative epigenetic chromatin states of polycomb target genes.
176 *PLoS Genet* **6**, e1000805, doi:10.1371/journal.pgen.1000805 (2010).

177 3 Aranda, S., Mas, G. & Di Croce, L. Regulation of gene transcription by Polycomb
178 proteins. *Sci Adv* **1**, e1500737, doi:10.1126/sciadv.1500737 (2015).

179 4 Simon, J. A. & Kingston, R. E. Occupying chromatin: Polycomb mechanisms for getting
180 to genomic targets, stopping transcriptional traffic, and staying put. *Mol Cell* **49**, 808-
181 824, doi:10.1016/j.molcel.2013.02.013 (2013).

182 5 Santos-Pereira, J. M. & Aguilera, A. R loops: new modulators of genome dynamics and
183 function. *Nat Rev Genet* **16**, 583-597, doi:10.1038/nrg3961 (2015).

184 6 Skourti-Stathaki, K. *et al.* R-Loops Enhance Polycomb Repression at a Subset of
185 Developmental Regulator Genes. *Mol Cell* **73**, 930-945 e934,
186 doi:10.1016/j.molcel.2018.12.016 (2019).

187 7 Maeda, R. K. & Karch, F. The bithorax complex of Drosophila an exceptional Hox
188 cluster. *Curr Top Dev Biol* **88**, 1-33, doi:10.1016/S0070-2153(09)88001-0 (2009).

189 8 Simon, J. Locking in stable states of gene expression: transcriptional control during
190 Drosophila development. *Curr Opin Cell Biol* **7**, 376-385 (1995).

191 9 Grossniklaus, U. & Paro, R. Transcriptional silencing by polycomb-group proteins. *Cold*
192 *Spring Harb Perspect Biol* **6**, a019331, doi:10.1101/cshperspect.a019331 (2014).

bioRxiv preprint first posted online May. 15, 2019; doi: http://dx.doi.org/10.1101/635722. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

193 10 Prezioso, C. & Orlando, V. Polycomb proteins in mammalian cell differentiation and
194 plasticity. *FEBS Lett* **585**, 2067-2077, doi:10.1016/j.febslet.2011.04.062 (2011).
195 11 Poux, S., Kostic, C. & Pirrotta, V. Hunchback-independent silencing of late Ubx
196 enhancers by a Polycomb Group Response Element. *EMBO J* **15**, 4713-4722 (1996).
197 12 Kassis, J. A. & Brown, J. L. Polycomb group response elements in Drosophila and
198 vertebrates. *Adv Genet* **81**, 83-118, doi:10.1016/B978-0-12-407677-8.00003-8 (2013).
199 13 Bauer, M., Trupke, J. & Ringrose, L. The quest for mammalian Polycomb response
200 elements: are we there yet? *Chromosoma* **125**, 471-496, doi:10.1007/s00412-015-0539-4
201 (2016).
202 14 Xiao, J. *et al.* Cis and trans determinants of epigenetic silencing by Polycomb repressive
203 complex 2 in Arabidopsis. *Nat Genet* **49**, 1546-1552, doi:10.1038/ng.3937 (2017).
204 15 Hekimoglu, B. & Ringrose, L. Non-coding RNAs in polycomb/trithorax regulation. *RNA*
205 *Biol* **6**, 129-137 (2009).
206 16 Herzog, V. A. *et al.* A strand-specific switch in noncoding transcription switches the
207 function of a Polycomb/Trithorax response element. *Nat Genet* **46**, 973-981,
208 doi:10.1038/ng.3058 (2014).
209 17 Ringrose, L. Noncoding RNAs in Polycomb and Trithorax Regulation: A Quantitative
210 Perspective. *Annu Rev Genet* **51**, 385-411, doi:10.1146/annurev-genet-120116-023402
211 (2017).
212 18 Davidovich, C. & Cech, T. R. The recruitment of chromatin modifiers by long noncoding
213 RNAs: lessons from PRC2. *RNA* **21**, 2007-2022, doi:10.1261/rna.053918.115 (2015).
214 19 Davidovich, C. *et al.* Toward a consensus on the binding specificity and promiscuity of
215 PRC2 for RNA. *Mol Cell* **57**, 552-558, doi:10.1016/j.molcel.2014.12.017 (2015).

bioRxiv preprint first posted online May. 15, 2019; doi: http://dx.doi.org/10.1101/635722. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

216 20 Davidovich, C., Zheng, L., Goodrich, K. J. & Cech, T. R. Promiscuous RNA binding by
217 Polycomb repressive complex 2. *Nat Struct Mol Biol* **20**, 1250-1257,
218 doi:10.1038/nsmb.2679 (2013).
219 21 Wang, X. *et al.* Targeting of Polycomb Repressive Complex 2 to RNA by Short Repeats
220 of Consecutive Guanines. *Mol Cell* **65**, 1056-1067 e1055,
221 doi:10.1016/j.molcel.2017.02.003 (2017).
222 22 Wang, X. *et al.* Molecular analysis of PRC2 recruitment to DNA in chromatin and its
223 inhibition by RNA. *Nat Struct Mol Biol* **24**, 1028-1038, doi:10.1038/nsmb.3487 (2017).
224 23 Beltran, M. *et al.* The interaction of PRC2 with RNA or chromatin is mutually
225 antagonistic. *Genome Res* **26**, 896-907, doi:10.1101/gr.197632.115 (2016).
226 24 Zhang, Q. *et al.* RNA exploits an exposed regulatory site to inhibit the enzymatic activity
227 of PRC2. *Nat Struct Mol Biol* **26**, 237-247, doi:10.1038/s41594-019-0197-y (2019).
228 25 Sanz, L. A. *et al.* Prevalent, Dynamic, and Conserved R-Loop Structures Associate with
229 Specific Epigenomic Signatures in Mammals. *Mol Cell* **63**, 167-178,
230 doi:10.1016/j.molcel.2016.05.032 (2016).
231 26 Chen, P. B., Chen, H. V., Acharya, D., Rando, O. J. & Fazzio, T. G. R loops regulate
232 promoter-proximal chromatin architecture and cellular differentiation. *Nat Struct Mol*
233 *Biol* **22**, 999-1007, doi:10.1038/nsmb.3122 (2015).
234 27 Hartono, S. R. *et al.* The Affinity of the S9.6 Antibody for Double-Stranded RNAs
235 Impacts the Accurate Mapping of R-Loops in Fission Yeast. *J Mol Biol* **430**, 272-284,
236 doi:10.1016/j.jmb.2017.12.016 (2018).
237 28 Xu, W. *et al.* The R-loop is a common chromatin feature of the Arabidopsis genome. *Nat*
238 *Plants* **3**, 704-714, doi:10.1038/s41477-017-0004-x (2017).

bioRxiv preprint first posted online May. 15, 2019; doi: http://dx.doi.org/10.1101/635722. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

239 29 Papp, B. & Muller, J. Histone trimethylation and the maintenance of transcriptional ON
240 and OFF states by trxG and PcG proteins. *Genes Dev* **20**, 2041-2054,
241 doi:10.1101/gad.388706 (2006).
242 30 Langlais, K. K., Brown, J. L. & Kassis, J. A. Polycomb group proteins bind an engrailed
243 PRE in both the "ON" and "OFF" transcriptional states of engrailed. *PLoS One* **7**,
244 e48765, doi:10.1371/journal.pone.0048765 (2012).
245 31 Long, Y. *et al.* Conserved RNA-binding specificity of polycomb repressive complex 2 is
246 achieved by dispersed amino acid patches in EZH2. *Elife* **6**, doi:10.7554/eLife.31558
247 (2017).
248 32 Kasahara, M., Clikeman, J. A., Bates, D. B. & Kogoma, T. RecA protein-dependent R-
249 loop formation in vitro. *Genes Dev* **14**, 360-365 (2000).
250 33 Mazina, O. M., Keskin, H., Hanamshet, K., Storici, F. & Mazin, A. V. Rad52 Inverse
251 Strand Exchange Drives RNA-Templated DNA Double-Strand Break Repair. *Mol Cell*
252 **67**, 19-29 e13, doi:10.1016/j.molcel.2017.05.019 (2017).
253 34 Zaitsev, E. N. & Kowalczykowski, S. C. A novel pairing process promoted by
254 Escherichia coli RecA protein: inverse DNA and RNA strand exchange. *Genes Dev* **14**,
255 740-749 (2000).
256 35 Berry, S. & Dean, C. Environmental perception and epigenetic memory: mechanistic
257 insight through FLC. *Plant J* **83**, 133-148, doi:10.1111/tpj.12869 (2015)
258

bioRxiv preprint first posted online May. 15, 2019; doi: http://dx.doi.org/10.1101/635722. The copyright holder for this preprint
(which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license.

259 **Acknowledgements:**

260 We thank E. Lécuyer's lab for assistance collecting *Drosophila* embryos, O. Neyret for advice
261 on preparation of NGS samples, J. Mallette for technical assistance, C. Gentile for advice on
262 bioinformatics, M. Wilson for S9.6 antibody, M. Reijns for plasmid to express hRNaseH2, M.
263 Drolet for intellectual support, and F. Robert, M. Drolet, and members of the Francis lab for
264 comments on the manuscript. This research was enabled in part by support provided by Calcul
265 Québec (www.calculquebec.ca) and Compute Canada (www.computecanada.ca). **Funding:**
266 Work in the N.J.F. lab was funded by CIHR 311557, in the K.J.A. lab by a grant from the David
267 and Lucile Packard Foundation, NIH 5R01GM115882-03, and 5T32HL007151-40 (to D.G.), and
268 in the F.C. lab by NIH R01-GM120607. **Author contributions:** Conceptualization: N.J.F. &
269 C.A.; Investigation: C.A., O.A. & N.J.F.; Formal analysis: C.A. & N.J.F.; Resources: V.C. &
270 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
271 acquisition: N.J.F., F.C., K.J.A. **Competing interests:** The authors declare no competing
272 interests. **Materials & Correspondence:** All correspondence and material requests should be
273 directed to N.J.F.

274

bioRxiv preprint first posted online May. 15, 2019; doi: http://dx.doi.org/10.1101/635722. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

275 **Fig. 1 R-loops form at *Drosophila* PREs and correlate with a repressed state.** a-c. DRIP-seq
276 traces showing R-loop formation at PREs bound by PRC1 components (arrows) in 2-6H and 10-
277 14H *Drosophila* embryos (Ph), and in S2 cells (dRING). RNaseH-treated samples are negative
278 controls. "Unstr" indicates all R-loops, while + and – indicate strand specific tracks; direction
279 refers to the DNA in the RNA-DNA hybrid. Called peaks are indicated under the traces. Red
280 arrowheads above genes indicate direction of annotated transcripts. d, e. Median normalized
281 intensity of PRC1 components (d), or H3K27me3 (e) over PREs with or without R-loops. 2-6H
282 and 10-14H R-loop data are compared with Ph at 4-12H and H3K27me3 at 4-8H and 12-16H
283 respectively. S2 cell R-loop data are compared with dRING. Whiskers show min. to max. f.
284 Median normalized intensity of H3K27me3 at 12-16H over PREs where R-loop formation is
285 detected in 2-6H but not in 10-14H compared to PREs with no R-loops detected at either stage.
286 See also Extended Data Fig. 2.

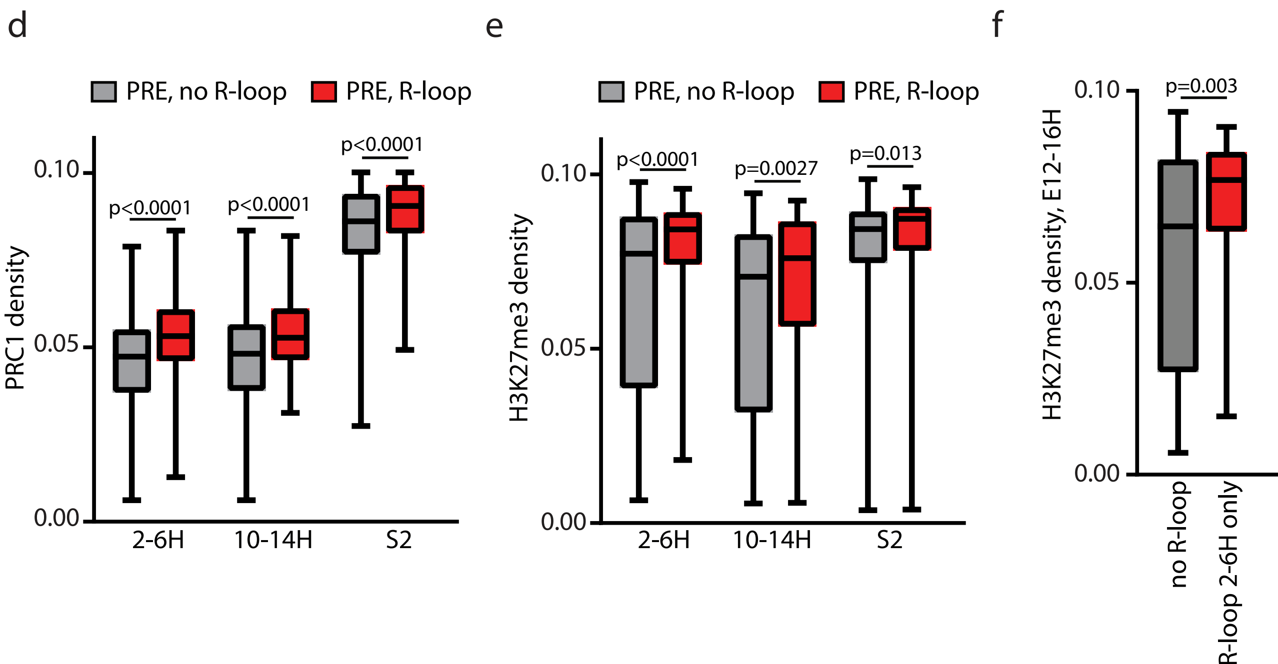
287

bioRxiv preprint first posted online May. 15, 2019; doi: http://dx.doi.org/10.1101/635722. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

a
chr3R *Antp*
PRE
PRC1 ChIP-seq
Ph 4-12H
S2 dRING
2-6H embryos
10-14H embryos
S2 cells
Unstr.
+
-
RNaseH

b
chr2R *TER94*

c
chr3L *drm*



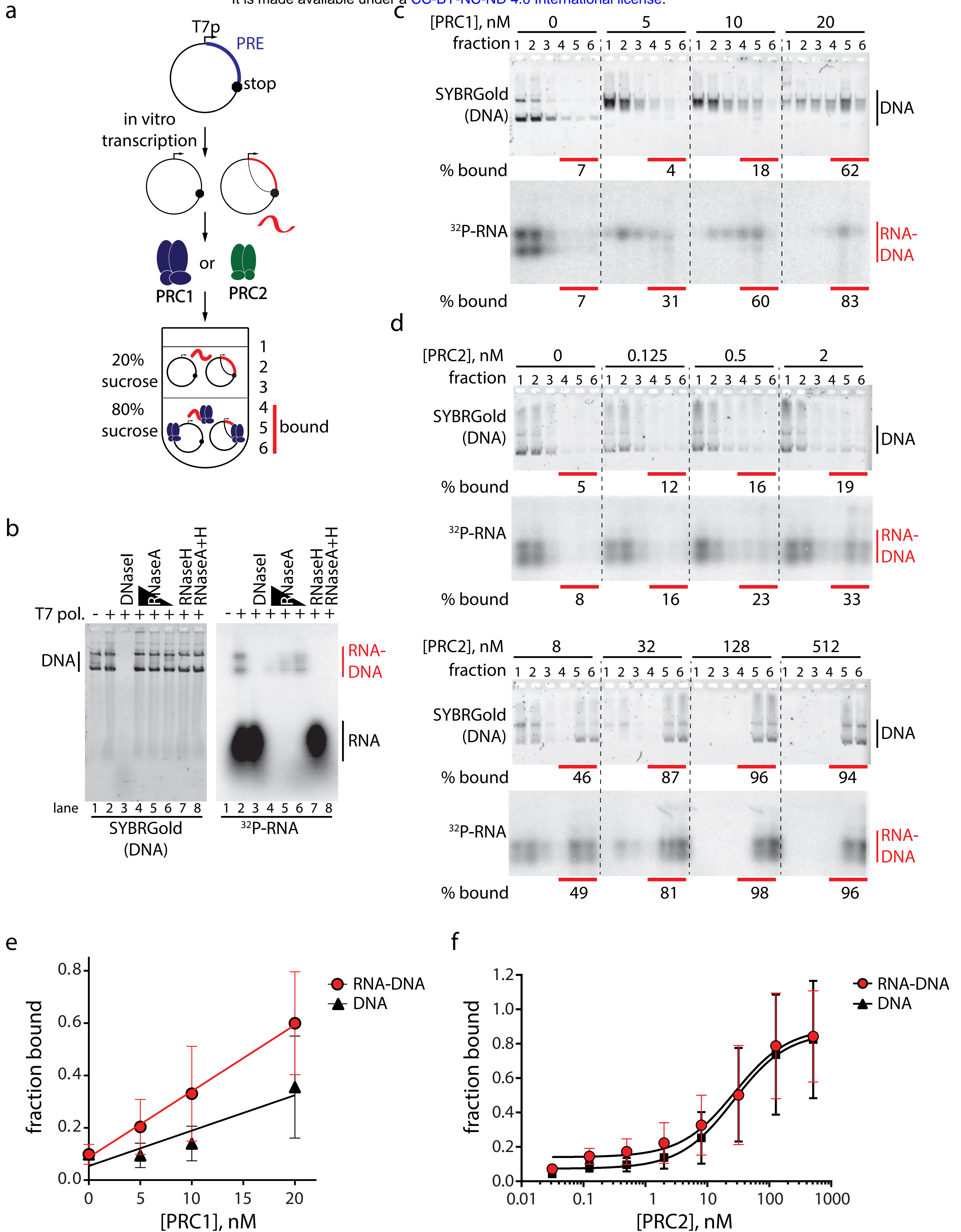
Alecki et al. Figure 1

bioRxiv preprint first posted online May. 15, 2019; doi: http://dx.doi.org/10.1101/635722. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

288 **Fig. 2 PcG complexes recognize R-loop containing templates.** a. Scheme for testing binding of
289 PcG complexes to R-loop containing plasmids. b. Radiolabelled RNAs that co-migrate with the
290 DNA plasmid (lane 2) are confirmed as R-loops by their sensitivity to DNaseI (lane3) and
291 RNaseH (lane 4), and resistance to RNaseA (lanes 5-7). c, d. Representative gradients of PRC1
292 (c) and PRC2 (d) binding. ^{32}P-RNA panel shows RNaseA digested RNA-DNA hybrids. See
293 Extended Data Fig. 4 for full gels. E, F. Quantification for PRC1 (e, n=8) and PRC2 (f, n=6).
294 Points show the mean +/- S.D. RNA-DNA and DNA curve fits are different for PRC1
295 ($p<0.0001$, exact sum-of-squares F-test).

296

bioRxiv preprint first posted online May. 15, 2019; doi: http://dx.doi.org/10.1101/635722. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.



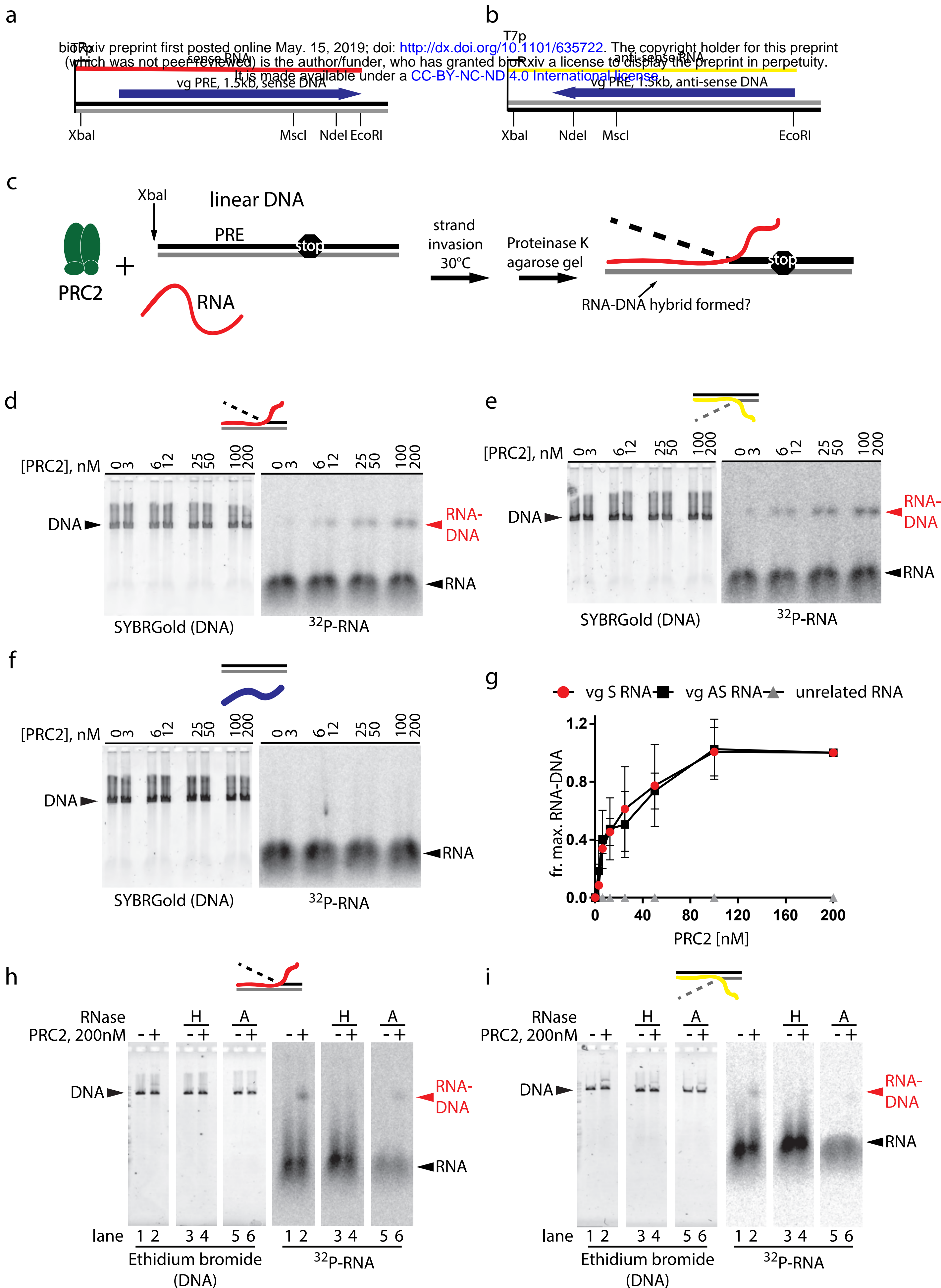
Alecki et al. Figure 2

bioRxiv preprint first posted online May. 15, 2019; doi: http://dx.doi.org/10.1101/635722. The copyright holder for this preprint
(which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license.

297 **Fig. 3 PRC2 has RNA strand invasion activity.** a, b. Nucleic acids used for RNA strand
298 invasion. Plasmids with the *vestigial* (*vg*) PRE in either orientation are transcribed to produce
299 sense (a) or anti-sense RNAs (b). Linearized plasmid is used as the DNA template. c. Strand
300 invasion assay scheme. d-f. Titration of PRC2 with either the sense (d) or anti-sense (e) *vg* PRE
301 RNA, or a non-complementary RNA (f). g. Quantification of 3 titrations; graphs show mean +/-
302 S.D of the the fraction of signal with 200 nM PRC2. h, i. RNA strand invasion products are
303 sensitive to RNaseH (lane 4) but resistant to RNaseA (lane 6). See also Extended Data Fig. 6.

304

bioRxiv preprint first posted online May. 15, 2019; doi: http://dx.doi.org/10.1101/635722. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

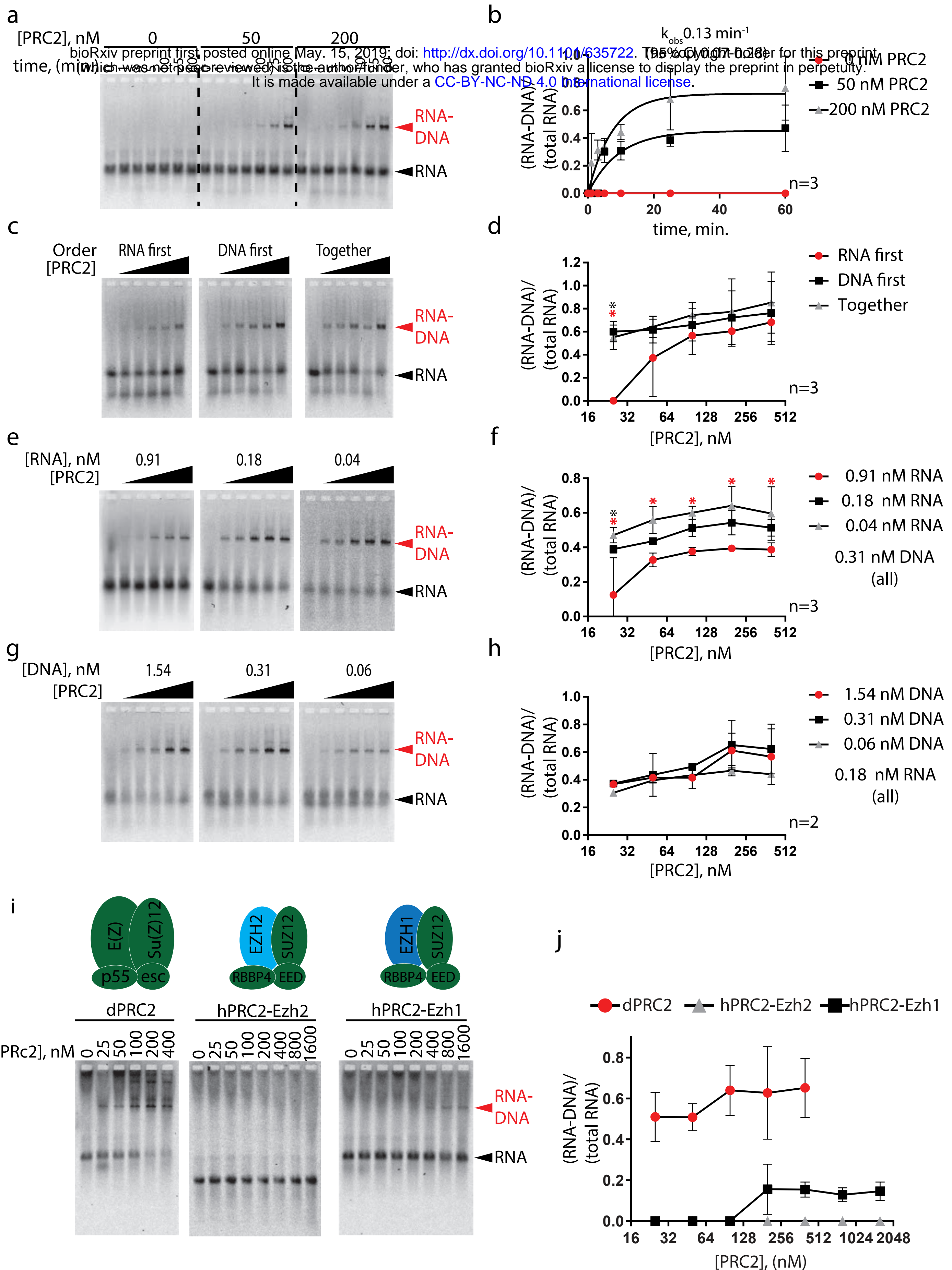


Alecki et al. Figure 3

bioRxiv preprint first posted online May. 15, 2019; doi: http://dx.doi.org/10.1101/635722. The copyright holder for this preprint
(which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license.

305 **Fig. 4 Characteristics and conservation of PRC2 strand invasion activity.** In a-h, left panel
306 shows a representative gel of Cy5-labelled RNA, and graph (right panel) summarizes multiple
307 experiments. a, b. Time course of RNA strand invasion. c, d. Effect of order of addition of
308 nucleic acids. Asterisks in all panels indicate $p<0.05$ (unpaired student's t-test). red=RNA vs.
309 DNA; black=RNA vs. together. c-g. Effect of increasing RNA (e, f) or DNA (g, h). Asterisks:
310 red=0.91 nM vs. 0.04 nM; black=0.91 nM vs. 0.18 nM. I- All PRC2 titrations are 25-400 nM. i,
311 j. PRC2-EZH1 has strand invasion activity. See also Extended Data Fig. 8.

bioRxiv preprint first posted online May. 15, 2019; doi: http://dx.doi.org/10.1101/635722. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.



Alecki et al. Figure 4