1 **ATRX promotes heterochromatin formation to protect cells from G-quadruplex DNA-**
2 **mediated stress**
3

4 **Authors**
5 Yu-Ching Teng[1], Aishwarya Sundaresan[1], Ryan O'Hara[1], Vincent U. Gant[1], Minhua Li[1], Sara
6 Martire[1], Jane N. Warshaw[1], Amrita Basu[2], Laura A. Banaszynski[1*]
7

8 **Affiliation**
9 [1]Cecil H. and Ida Green Center for Reproductive Biology Sciences, Department of Obstetrics and
10 Gynecology, Children's Medical Center Research Institute, Harold. C. Simmons Comprehensive
11 Cancer Center, Hamon Center for Regenerative Science and Medicine, University of Texas
12 Southwestern Medical Center, Dallas, Texas 75390, USA.
13

14 [2]Department of Surgery, University of California San Francisco, 550 16th St. San Francisco,
15 California 94158, USA
16

17 *Correspondence: Laura.Banaszynski@UTSouthwetsern.edu

18 **ABSTRACT**

19 ATRX is a tumor suppressor that has been associated with protection from DNA replication
20 stress, purportedly through resolution of difficult-to-replicate G-quadruplex (G4) DNA structures.
21 While several studies demonstrate that loss of ATRX sensitizes cells to chemical stabilizers of
22 G4 structures, the molecular function of ATRX at G4 regions during replication remains
23 unknown. Here, we demonstrate that ATRX associates with the MCM replication complex and
24 that loss of ATRX leads to G4 structure accumulation at newly synthesized DNA. We show that
25 both the helicase domain of ATRX and its H3.3 chaperone function are required to protect cells
26 from G4-induced replicative stress. Furthermore, these activities are upstream of
27 heterochromatin formation mediated by the histone methyltransferase, ESET, which is the
28 critical molecular event that protects cells from G4-mediated stress. In support, tumors carrying
29 mutations in either ATRX or ESET show increased mutation burden at G4-enriched DNA
30 sequences. Overall, our study provides new insights into mechanisms by which ATRX promotes
31 genome stability with important implications for understanding impacts of its loss on human
32 disease.
33

34 **INTRODUCTION**
35   ATRX is a chromatin remodeling protein of the SWI/SNF family, mutations in which cause alpha
36   thalassemia X-linked (ATRX) intellectual disability syndrome and are highly associated with a
37   number of cancers characterized by alternative lengthening of telomeres (ALT)[1]. ATRX localizes
38   to heterochromatic repetitive regions, including telomeres, pericentric repeats, rDNA repeats,
39   and endogenous retroviral elements (ERVs)[2–8]. ATRX functions with the histone chaperone
40   DAXX to facilitate deposition of the histone variant H3.3 at these regions[9], resulting in a unique
41   form of heterochromatin characterized by both trimethylation of histone H3 at lysine 9
42   (H3K9me3) and H3.3[10,11]. Loss of any one of ATRX/DAXX/H3.3 is correlated with loss of
43   H3K9me3 heterochromatin at repetitive regions[3,11,12]. Loss of ATRX has been implicated in
44   replication stress, DNA damage, and DNA repair failures that drive genome instability, and
45   aberrant homologous recombination[6,8,12–18]. However, the mechanism underlying this
46   dysfunction is unclear.
47
48   ATRX is hypothesized to prevent replication stress by resolving stable non-B form DNA
49   structures called G-quadruplexes (G4) ahead of the replication fork. These structures are
50   thought to form when double-stranded DNA is dissociated in GC-rich regions during replication
51   and transcription[19,20]. G4 structures are generally considered to block the progression of DNA
52   replication forks, ultimately leading to replication fork collapse[21]. A number of helicases,
53   including BLM, WRN, and ATRX, are proposed to protect the genome by unfolding G4
54   structures and preventing them from causing DNA breaks[22–24]. Genome-wide, ATRX is enriched
55   at GC-rich sequences with the propensity for forming G4 structures[2] although ATRX is unable to
56   unfold G4 structures in vitro[14]. However, exogenous expression of ATRX in cells lacking ATRX
57   has been shown to reduce levels of G4 structures[25] and, further, ATRX has been shown to
58   protect cells from replication stresses induced by chemical stabilizers of G4 structures[15,25,26].
59   Despite these striking observations, direct links between ATRX, G4 structures, and the
60   molecular mechanisms by which ATRX functions at these regions have not been reported.
61
62   In this study we explored the mechanisms underlying ATRX function at G4 structures. We
63   validate previous observations that ATRX is bound at G4-containing regions and further
64   demonstrate that ATRX interacts with G4 structures in cells. In addition, we find that ATRX
65   interacts with the MCM helicase and that G4 DNA accumulates at newly synthesized DNA in the
66   absence of ATRX. Genomic analysis suggests that loss of ATRX results in G4 accumulation
67   primarily at intergenic and repetitive regions. We show that ATRX requires both its helicase and
68   H3.3 chaperone activity to attenuate G4 stabilization-induced replication stress. Interestingly, we
69   find that ATRX maintains a closed chromatin state at G4-containing regions and that this
70   function is mediated by DAXX-dependent deposition of H3.3. These activities are upstream of
71   heterochromatin formation, which is ultimately required for protection from G4-mediated stress.
72   Finally, we show that pan-cancer patient samples bearing mutations in either ATRX or ESET
73   had a significantly increased mutation burden at G4-containing DNA. Overall, our findings
74   support the conclusion that ATRX protects genomic integrity at G4-containing regions by
75   maintaining these regions in a closed heterochromatic state.
76
77

78 **RESULTS**

79 **ATRX associates with G4 structures in vivo**

80 It has been reported that ATRX binds to repetitive regions of the genome such as telomeres and

81 pericentric regions[2], but also to promoters[7,8]. To confirm genomic ATRX enrichment, we

82 performed ChIP-seq using a GFP antibody and mouse embryonic stem cells (mESCs) in which

83 the endogenous ATRX locus was tagged[27] with a C-terminal GFP (**Fig. 1a** and **Supplementary**

84 **Fig. 1a, 1b**). ATRX-enriched regions were often enriched with the histone variant, H3.3, and

85 with H3K9me3, consistent with previous reports of this unique class of heterochromatin in

86 ESCs[11,28] (**Fig. 1a**). In agreement with published data[2,8], we found ATRX enriched mainly at

87 repeat regions in ESCs, including telomere and pericentric repeats that are predicted to form G-

88 quadruplex structures (G4) under physiological conditions[29,30] (**Supplementary Fig. 1c-1e**). To

89 further validate these results, we performed ChIP-seq using an ATRX antibody in both WT and

90 ATRX knockout (ATRX KO) ESCs (**Fig. 1a** and **Supplementary Fig. 1f, 1g**). For further

91 analysis, we considered our ATRX-GFP and WT ATRX ChIP-seq data sets as replicates.

92 Overall, we identified 435 regions of ATRX enrichment at uniquely mappable regions of the

93 genome. These regions contained 3.6x more (30 standard deviations more than expected by

94 chance) G4 consensus motifs ($G_3N_{1-12}G_3N_{1-12}G_3N_{1-12}G_3$)[31] compared to randomly selected

95 regions that were comparable in size and number, over 100 simulations (**Supplementary Fig.**

96 **1h**). Nearly 40% (169/435) of ATRX-enriched regions contained a G4 consensus motif (**Fig. 1b**

97 and **Supplementary Table 1**), and further, de novo motif prediction identified GC-rich motifs as

98 being overrepresented in these regions (**Supplementary Fig. 1i**). G4-containing regions

99 showed considerable ATRX enrichment compared to regions identified as enriched with ATRX

100 but without a G4 consensus motif (**Fig. 1c** and **Supplementary Fig. 1j, 1k**).

101

102 Based on the enrichment of the G4 consensus motif at sites of ATRX enrichment, we next

103 asked whether we could observe ATRX interaction with G4 structures in cells. Since ATRX is

104 reported to be recruited to telomeres during late S and associated with their replication[13], we

105 wanted to assess ATRX-G4 interaction throughout the cell cycle. We synchronized cells in

106 mitosis using a thymidine-nocodazole treatment and then released cells with EdU to visualize

107 DNA replication (**Fig. 1d** and **Supplementary Fig. 2a, 2b**). We then performed a proximity

108 ligation assay (PLA) using antibodies that recognize 1) ATRX and 2) G4 structures (e.g., the

109 BG4 antibody)[32] in both WT and ATRX KO mESCs. Interestingly, we observed ATRX and G4

110 association in cells that had not yet begun replication (**Fig. 1e,** note cells in G1 evidenced by

111 lack of EdU staining). Association was highest in G1 phase and decreased slightly in early S

112 phase (**Fig. 1f**). We did not observe an appreciable level of ATRX-G4 association on mitotic

113 chromosomes (**Supplementary Fig. 2c**). Further, we did not observe ATRX-G4 foci in single

114 antibody controls or ATRX KO mESCs (**Supplementary Fig. 3a**) or G4 association with an

115 independent chromatin-associated protein, HIRA (**Supplementary Fig. 3b, 3c**), demonstrating

116 the specificity of the interaction. Overall, these data strongly link ATRX to regions containing G4

117 consensus motifs and G4 structures.

118

119 **ATRX associates with the MCM DNA helicases**

120 Despite increasing evidence that ATRX plays a replication-associated role in resolving G4

121 structures[14–16], mechanisms that recruit ATRX to these structures remain poorly understood. We

122 therefore used immunoprecipitation-coupled to mass spectrometry (IP-MS) to identify potential
123 factors that interact with ATRX that may allow insight into its localization and function. We
124 identified many proteins related to DNA replication and transcription that were enriched by
125 ATRX pulldown in wild-type cells compared to ATRX KO cells and immunoprecipitation using a
126 control IgG antibody (**Fig. 2a** and **Supplementary Fig. 4a**, **4b**). These include previously
127 identified ATRX-interacting proteins such as Mre11[14], a member of the MRN complex involved
128 in DNA damage repair (**Supplementary Table 2**). Further, we identified a number of proteins
129 that have been genetically linked to G4 structures[33], including DDX10 and SPEN
130 (**Supplementary Table 2**). We also identified members of the MiniChromosome Maintenance
131 (MCM) complex (e.g., MCM2, MCM4, MCM6, and MCM7), a DNA helicase essential for
132 replication[34], as novel ATRX-interacting proteins (**Fig. 2a**).
133
134 We first validated interaction of ATRX and MCM proteins using co-immunoprecipitation (**Fig. 2b**
135 and **Supplementary Fig. 4c**). ATRX specifically pulled down MCM2/4/6/7 subcomplexes[34,35]
136 (**Fig 2a, b** and **Supplementary Fig. 4c**). Because the enrichment of MCM by ATRX
137 immunoprecipitation was relatively low, and to determine whether ATRX-MCM interaction
138 occurred in cells, we performed proximity ligation using ATRX and MCM antibodies. We
139 observed ATRX interaction with MCM2/3/4/6, with no PLA signal observed in ATRX KO ESCs
140 (**Supplementary Fig. 4d-4k**), suggesting that ATRX is in proximity of a complete MCM
141 complex. The majority of EdU labeling is not at sites of ATRX-MCM interaction, in line with
142 ATRX being bound at only a small number of loci genome-wide (**Fig. 1b**).
143
144 ATRX contains an ATPase/helicase domain that is often mutated in both ATRX syndrome and
145 in cancers[36,37]. Another well-characterized function of ATRX is its involvement in H3.3
146 deposition at telomeres and other types of heterochromatin through its interactions with the
147 H3.3 chaperone protein, DAXX[9]. Hypothetically, the ATRX helicase domain could unwind G4
148 structures which could be resolved by H3.3 deposition and nucleosome formation. We therefore
149 asked whether interaction with MCM requires these two functional elements of ATRX. We first
150 performed ATRX immunoprecipitation and ATRX-Mcm6 PLA in DAXX knockout (DAXX KO)
151 ESCs (**Supplementary Fig. 1f**). Interestingly, we found that ATRX maintains association with
152 Mcm6 in the absence of DAXX using both IP-WB and PLA (**Supplementary Fig. 5a-5c**). In
153 support, the ATRX-Mcm6 association was maintained in ATRX KO cells expressing an ATRX
154 mutant that attenuates DAXX binding (ATRX L1238A)[38] (**Fig. 2c** and **Supplementary Fig. 5d-
155 5f**). The ATRX/DAXX chaperone complex shares its substrate H3.3 with the HIRA complex,
156 which deposits H3.3 at regulatory elements and genes[39]. We found that HIRA did not co-
157 immunoprecipitate MCM6 in wild-type cells (**Supplementary Fig. 5g**), and that ATRX
158 interaction with Mcm6 was maintained in the absence of HIRA (**Supplementary Fig. 5a, 5h**),
159 suggesting the specificity of the ATRX-MCM interaction. Finally, we expressed an ATRX
160 helicase mutant (ATRX K1562R) in ATRX KO mESCs. Using IP-WB and PLA, we found that the
161 ATRX helicase mutant maintains association with Mcm6 (**Fig. 2c** and **Supplementary Fig. 5d-
162 5f**). Overall, these results suggest that ATRX helicase and chaperone activity is decoupled from
163 mechanisms dictating its localization on chromatin.
164

165   Recent reports demonstrate that Mcm2 itself contains chaperone function for H3-H4 dimers[40],
166   raising the intriguing possibility that the MCM-ATRX association is bridged by H3.3 itself. To test
167   the requirement of H3.3 for ATRX-Mcm6 interaction, we immunoprecipitated ATRX from H3.3
168   knockout cells (**Supplementary Fig. 6a**). We found that loss of H3.3 resulted in reduced
169   association between ATRX and Mcm6 (**Fig. 2d** and **Supplementary Fig. 6b**). ATRX-Mcm6
170   association could be restored by exogenous expression of H3.3 but not by an H3.3 mutant
171   previously shown to inhibit stable formation of nucleosomes (H3.3 L126A I130A)[41] (**Fig. 2d** and
172   **Supplementary Fig. 6c**), demonstrating the importance of H3.3 deposition for maintaining the
173   ATRX-MCM interaction.
174
175   **Loss of ATRX leads to accumulation of G4 structures at sites of DNA synthesis**
176   Our identification of an association between ATRX and the MCM complex supports a body of
177   existing literature demonstrating that ATRX plays a role in DNA replication[14–16,26,42]. While many
178   studies hypothesize that ATRX replicative stress is due to the inability to resolve G4
179   structures[2,14,15,25,26], the direct effect of loss of ATRX on G4 structures at replicating DNA has not
180   been assessed. To test whether the persistence of G4 structures at newly synthesized DNA is
181   affected by loss of ATRX, we used in situ analysis of protein interaction at DNA replication forks,
182   or SIRF[43]. Briefly, mitotic cells were released from synchronization and labeled with EdU for a
183   short period of time prior to cell fixation at early S phase. EdU incorporation into newly
184   synthesized DNA was detected using click chemistry for biotin labeling. EdU and G4 interaction
185   was then determined by proximity ligation. We found that normalized EdU-G4 signals were
186   increased in ATRX KO and DAXX KO ESCs (**Fig. 3a** and **3b**). This effect could be rescued by
187   expression of wild-type ATRX in ATRX KO ESCs, whereas we find that both the helicase
188   activity of ATRX and its ability to interact with DAXX are important to reduce G4-EdU
189   persistence (**Supplementary Fig. 7**).
190
191   While imaging experiments allow powerful observation of G4 activity at the single-cell level, they
192   fail to provide genomic identification of specific G4-containing regions that may be affected by
193   loss of ATRX. To experimentally identify G4 regions in ECSs and determine the effect of ATRX
194   loss on these regions, we performed CUT&Tag using the BG4 antibody (**Fig. 3c**). We identified
195   2,320 regions as experimentally enriched with G4 structures in wild-type cells (**Fig. 3d**). These
196   regions contained 5.3x more G4 motifs compared to randomly selected regions, with 58%
197   (1,356/2,230) containing a predicted G4 motif (**Supplementary Fig. 8a** and **8b**). Generally, G4-
198   containing regions are localized to promoters in wild-type ESCs - of note, a region that we do
199   not strongly associate with ATRX binding (**Fig. 3e**, also see **Supplementary Fig. 1c**). We next
200   wanted to determine the effect of loss of ATRX on genomic G4 accumulation. Interestingly, we
201   observe 6,301 novel G4 sites in ATRX KO ESCs that are not observed in wild-type ESCs (**Fig.
202   3d** and **Supplementary Fig. 8c** and **8d**). These regions are annotated predominantly as
203   introns, intergenic regions, LINEs, and LTRs (**Fig. 3f**), suggesting that ATRX is mainly
204   responsible for resolving G4 structures in these regions.
205
206   While the G4-EdU signal observed in our PLA assay could be due to replication, we cannot rule
207   out that this signal results from nucleotide incorporation due to DNA damage repair. To try to
208   distinguish between the two, we used two sequencing techniques - EdU-seq, in which we

209 sequenced nascently synthesized DNA after a short pulse of EdU (which cannot distinguish
210 between repair and replication), and published SNS-seq data[44], in which RNA primers
211 associated with DNA replication initiation are sequenced (mapping putative origins). We found
212 several regions where mapped origins[44] were identified upstream or downstream of ATRX-
213 enriched G4 regions, and further, that treatment with the G4 stabilizer, PhenDC3, resulted in
214 increased detection of origin activity at or near ATRX-bound G4 regions compared with
215 untreated cells (**Supplementary Fig. 9a-d**). These regions also showed increased EdU-seq
216 signal in the absence of either ATRX or DAXX (**Supplementary Fig. 9a, 9e-f**). However, when
217 we consider G4-enriched regions that are identified experimentally in the absence of ATRX, we
218 find that these regions show increased EdU-seq signal in the absence of ATRX or DAXX with
219 no evident increase in origin usage after PhenDC3 treatment[44] (**Fig. 3g** and **Supplementary**
220 **Fig. 9g**). Overall, while a subset of ATRX-dependent G4 regions may stimulate replication,
221 these data and the literature are consistent with a model in which ATRX KO cells experience
222 increased replicative stress and DNA damage due to G4 persistence.
223
224 **ATRX requires its helicase activity and interaction with DAXX to protect from G4-induced**
225 **stress**
226 Previous studies have linked loss of ATRX to increased sensitivity to chemical agents that
227 stabilize G4 structures such as pyridostatin (PDS) and CX-3543, presumably due to G4
228 persistence in cells[25,26]. Our results suggest that both ATRX and DAXX are required to reduce
229 G4 levels in cells. We therefore asked whether loss of DAXX would phenocopy loss of ATRX
230 with respect to PDS sensitivity. We treated an isogenic panel of wild-type, ATRX KO, and DAXX
231 KO ESCs with increasing concentrations of PDS. Analysis of cell viability demonstrated that
232 both DAXX KO and ATRX KO resulted in increased sensitivity to PDS (**Fig. 4a**). Importantly,
233 exogenous expression of DAXX in DAXX KO ESCs, and likewise, exogenous expression of
234 ATRX in ATRX KO ESCs, attenuated sensitivity to PDS (**Fig. 4b** and **Supplementary Fig. 10a**
235 and **10b**). To assess whether ATRX interaction with DAXX was necessary to relieve PDS-
236 induced stress, we expressed the ATRX DAXX-binding mutant (ATRX L1238A) in ATRX KO
237 ESCs. Strikingly, this mutation is unable to relieve PDS sensitivity (**Fig. 4b**), suggesting that
238 ATRX interaction with DAXX is required for this function. Further, mutations to ATRX that
239 disrupt its helicase activity (ATRX K1562R, K1612N)[45] remain sensitive to PDS (**Fig. 4b**).
240
241 The requirement for an ATRX/DAXX complex suggests that H3.3 deposition may be important
242 to protect from PDS sensitivity. In agreement, we found that loss of either ATRX or DAXX
243 results in reduced H3.3 incorporation at sites of ATRX enrichment containing G4 consensus
244 motifs (**Supplementary Fig. 10c** and **10d**). Further, H3.3 KO ESCs showed increased
245 sensitivity to PDS compared to wild-type ESCs (**Fig. 4c** and **Supplementary Fig. 10e**).
246 Sensitivity could be rescued by exogenous expression of H3.3 but not by the H3.3 deposition
247 mutant (H3.3 L126A I130A) (**Fig. 4c**). Importantly, we find that HIRA KO ESCs are highly
248 comparable with wild-type ESCs with respect to PDS sensitivity (**Fig. 4a**). Overall, these data
249 suggest that both ATRX helicase activity and ATRX/DAXX-mediated H3.3 deposition, but not
250 HIRA-mediated H3.3 deposition, are required for protection from G4 stabilizers.
251

252 **ATRX maintains closed chromatin states upstream of heterochromatin formation to**
253 **protect cells from G4-mediated stress**
254 We next wanted to understand how ATRX protects cells from G4-mediated stress. If ATRX is
255 involved in nucleosome assembly at G4 DNA, it follows that these regions should be more open
256 in the absence of ATRX. In support, recent studies demonstrate that loss of ATRX induces
257 chromatin de-compaction at telomeres and repetitive elements[3,10,12,46]. To explore whether
258 ATRX deficiency altered chromatin accessibility at G4 regions, we performed ATAC-seq in wild-
259 type and ATRX KO ESCs. Strikingly, we found that ATRX depletion elevated chromatin
260 accessibility at ATRX-enriched regions containing G4 consensus motifs while paradoxically
261 resulting in reduced H3.3 deposition, a histone variant long associated with open chromatin
262 (**Fig. 5a, 5b** and **Supplementary Fig. 10c**). Analogous results were observed in DAXX KO
263 ESCs and H3.3 KO ESCs (**Fig. 5c** and **5d and Supplementary Fig. 10d**). Further, we find that
264 expression of exogenous ATRX in ATRX KO ESCs rescues this effect, while neither the ATRX
265 helicase mutant nor the DAXX-interaction mutant in ATRX KO ESCs results in reduced
266 accessibility at these regions (**Fig. 5e**). Maintaining these regions in a closed state requires
267 H3.3 deposition, as only expression of wild-type H3.3, and not that of the deposition mutant,
268 was able to rescue this effect in H3.3 KO ESCs (**Fig. 5f**). Importantly, chromatin accessibility at
269 G4 regions was not altered in HIRA KO ESCs (**Supplementary Fig. 11a**). Additionally,
270 increased chromatin accessibility did not result in concomitant increases in histone acetylation
271 or transcription from nearby genes (**Supplementary Fig. 11b-f**).
272
273 Previous studies have shown that ATRX facilitates the establishment and maintenance of a
274 subset of H3.3- and H3K9me3-marked heterochromatin[3,5,11]. We therefore asked how loss of
275 the ATRX/DAXX/H3.3 complex would influence the chromatin state at ATRX-enriched G4
276 regions. First, we re-analyzed published H3K9me3 ChIP-seq data for ATRX KO, DAXX KO and
277 H3.3 KO ESCs[11]. We found that H3K9me3 enrichment at ATRX-bound G4 regions was reduced
278 in ATRX KO, DAXX KO and H3.3 KO cells (**Fig. 6a-6d**). Further, we find that expression of
279 exogenous ATRX in ATRX KO ESCs rescues this effect, while neither the ATRX helicase
280 mutant nor the DAXX-interaction mutant in ATRX KO ESCs rescues heterochromatin formation
281 at these regions (**Fig. 6e**). Maintaining these regions in a closed state requires H3.3 deposition,
282 as only expression of wild-type H3.3, and not that of the deposition mutant, was able to rescue
283 this effect in H3.3 KO ESCs (**Fig. 6f**). These data demonstrate that the ATRX/DAXX/H3.3
284 pathway are key factors for H3K9me3 maintenance at G-quadruplex regions.
285
286 Previous studies demonstrate that the histone methyltransferase ESET has a critical role in the
287 establishment of H3K9me3 at H3.3-enriched repetitive elements[11,47]. Accordingly, H3K9me3
288 enrichment on H3.3-containing nucleosomes at ATRX-bound G4 regions was reduced in the
289 absence of ESET (**Fig. 7a, 7b and Supplementary Fig. 12a**). We next asked whether there
290 exists a molecular hierarchy (ATRX helicase and histone chaperone activity vs ATRX promotion
291 of heterochromatin formation) in protecting cells from G4-mediated stress. To this end, we used
292 ESET conditional KO ESCs to determine the effect of loss of heterochromatin at G4 regions.
293 First, we found that H3.3 deposition at ATRX-bound G4 regions is unaffected in the absence of
294 ESET (**Fig. 7c**), even though heterochromatin at these regions is reduced (**Supplementary Fig.**
295 **12b and 12c**). Interestingly, ATRX-bound G4 regions become more open in the absence of

296 ESET (**Fig. 7d**). Further, like ATRX/DAXX/H3.3 KO ESCs, ESET KO ESCs are sensitive to
297 PDS treatment (**Fig. 7e**). Taken together, these results suggest that heterochromatin formation
298 is the ultimate molecular event that protects cells from G4-mediated stress.
299
300 It has been shown that G4 structures are correlated with increased DNA polymerization error
301 rates[48], and that G4 structures harbor a high incidence of disease-causing point mutations and
302 indels[49–51]. Further, studies demonstrate that ATRX mutant tumors carry a high mutation rate
303 either at the single-nucleotide variant level or copy number alterations[25,52]. Given these
304 observations, we asked whether regions identified as enriched with G4 structures had increased
305 mutation burden in human cancers. Strikingly, we found that patient samples containing either
306 ATRX or ESET mutations contained a significantly higher number of substitutions at G4-
307 containing DNA compared to an iteratively sampled random selection of donors (**Fig. 8a and**
308 **8b**). These substitutions were more likely to involve mutation of G or C bases compared with A
309 or T (**Fig. 8c and 8d**). Further, this trend was not observed from donors carrying mutations in
310 oncogenes such as KRAS and IDH1 (**Supplementary Fig. 13**), lesions that we do not expect to
311 impact G4 DNA regulation.
312
313 **DISCUSSION**
314 Previous studies have linked ATRX to replication fork progression, presumably through
315 resolution of G4 structures[2,14,15,25,26]. Here, we provide a direct link between ATRX and the
316 replication machinery. We find that ATRX interacts with the MCM helicase and that G4
317 structures show increased coincidence with newly synthesized DNA in the absence of ATRX.
318 We find that ATRX maintains G4-containing regions in a closed heterochromatic state through
319 DAXX-mediated deposition of the histone variant H3.3, and that establishment of
320 heterochromatin is the critical molecular event protecting cells from G4-mediated stress (**Fig. 9**).
321
322 Previous studies using immunofluorescence have not observed colocalization of ATRX and G4
323 structures in cells, presumably because ATRX activity results in the resolution of G4s[25]. One
324 advantage of PLA is the ability to capture these brief interactions. Interestingly, we find that
325 ATRX colocalizes with G4 structures throughout both G1 and S phase. This result suggests that
326 ATRX recruitment to these regions is not a result of replication fork stalling, and that it is instead
327 a normal feature of ATRX function. While we observe co-localization between ATRX and the
328 MCM complex, the exact molecular nature of this interaction remains to be seen and is unlikely
329 to drive ATRX recruitment to chromatin, given that MCM-bound regions vastly outnumber
330 ATRX-bound regions. To date, binding of G4 structures has not been attributed to a particular
331 region of ATRX. ATRX-bound regions thought to contain G4 structures, such as telomeres, are
332 generally enriched with H3K9me3, and ATRX has been shown to bind H3K9me3 through its
333 ADD domain[53,54]. However, ESCs contain many regions of H3K9me3 heterochromatin that are
334 not bound by ATRX or enriched with H3.3[11], so this cannot be the sole determinant of
335 recruitment and instead may stabilize ATRX association with these regions. Interestingly, we do
336 not observe appreciable ATRX enrichment at the thousands of new G4 structures that we
337 identify genomically in the absence of ATRX, in support of the hypothesis that ATRX recruitment
338 to these regions may be transitory and that ATRX is released once the G4 has been resolved. It
339 is also formally possible that ATRX may in some way regulate other helicases responsible for

9

340   regulating G4 structures in cells. Detailed molecular understanding of the kinetics of ATRX
341   recruitment and function at G4 regions remains an important point of future study.
342
343   Strikingly, we observe a dramatic increase in chromatin accessibility at ATRX-enriched G4
344   regions in the absence of ATRX/DAXX/H3.3, accompanied by a loss of H3K9me3
345   heterochromatin at these regions. These demonstrate that deposition of H3.3 plays a role in
346   maintaining a closed chromatin state in these regions. This is a somewhat surprising
347   observation given that many regions of H3.3 deposition experience high rates of nucleosome
348   turnover[55–57], and that in general, H3.3 is associated with open chromatin states[58,59]. These data
349   do, however, support that keeping G4 regions heterochromatinized is likely a defining functional
350   role of ATRX. Given the published literature[3,12,46], and our results in this study, it is likely that
351   ATRX protects cells from replicative stress by doing so. Additionally, we observe increased
352   chromatin accessibility and decreased H3K9me3 heterochromatin at ATRX-enriched G4 regions
353   in ATRX helicase mutant-expressing cells. Our data show an increase in G4 at EdU-
354   synthesized DNA and G4-mediated replicative stress in helicase-mutant expressing cells, as
355   well. Based on these observations, it is likely that ATRX unwinds G4 in cells. This is in contrast
356   to a previous study showing that ATRX is not able to unwind G4 structures in vitro[14]. This could
357   be due to lack of nearby chromatin on the G4 DNA in vitro or perhaps because ATRX has a
358   preference for a specific fold of G4 structure. Additional experiments are needed to test whether
359   chromatinized DNA or other varieties of G4 DNA structures are requirements for ATRX helicase
360   activity.
361
362   It is widely reported that ATRX mutation is strongly linked to human ALT cancers such as
363   pancreatic neuroendocrine tumors and glioma[39]. Many works support the hypothesis that ATRX-
364   deficient tumors show genome instability due dysregulation at specific repetitive DNA regions
365   (e.g., telomere or ribosomal DNA)[39]. Our analysis of mutation at observed G4 regions in ATRX-
366   deficient tumors are the first to show the direct link between ATRX mutation and G4 in tumor
367   samples by whole genome analysis, with C/G to A/T mutation being common in ATRX-deficient
368   tumors. This result is phenocopied in tumors containing mutation in ESET, supporting our model
369   that heterochromatinization protects G4 DNA from replicative stress. In the future, it will be
370   interesting to determine the molecular mechanism by which ATRX loss promotes substitutions
371   at G4 regions through extensive biochemical and molecular biology approaches including
372   careful analysis of the replication machinery and whole genome analysis in tumors.
373
374   Overall, we conclude that ATRX/DAXX protects G4 regions by maintaining a heterochromatin
375   state marked by both H3.3 deposition and reduced chromatin accessibility. These data provide
376   new insights into molecular mechanisms by which this complex supports genome integrity with
377   important implications for our understanding of human disease.
378
379
380
381

METHODS

Plasmids. To tag the endogenous ATRX locus with GFP, pCAS9-mCherry empty, pCAS9-mCherry-Frame +1, and pCRISPaint-TagGFP2-PuroR plasmids were purchased from Addgene. To generate HA-tagged ATRX expression plasmid (PB-ATRX-HA-Neo), human ATRX cDNA were amplified from HeLa and assembled into PB-EF1α-MCS-IRES-Neo vector (System Biosciences) by Gibson assembly. The ATRX sequence corresponded to isoform 2 (NM_138270.4) and was verified by standard Sanger sequencing. Site-directed mutagenesis was used to generate mutations in PB-ATRX-HA-Neo. The ATRX plasmids were propagated using ElectroMAX^{TM} Stbl4^{TM} Competent cells (Life Technologies). To generate Flag-tagged DAXX expression plasmid (pCDH-Flag-DAXX-Puro), Flag-tagged DAXX DNA were amplified from plasmid Flag-Daxx/pRK5 (Addgene) and assembled into pCDH-EF1α-MCS-IRES-Puro vector (System Biosciences).

Cell culture. ESCs were cultured on gelatin-coated plates under standard conditions (KO-DMEM, 2 mM Glutamax, 15% ES grade fetal bovine serum, 0.1 mM 2-mercaptoethanol, 1x Pen/Strep, 1x NEAA and leukemia inhibitory factor) at 37°C with 5% CO_2. ATRX KO, DAXX KO, HIRA KO (HIRA KO2), H3.3 KO and ESET cKO cells have been described previously[3,47,60,61]. ESET deletion was induced by 1 µM 4-OHT (Sigma) treatment in ESET cKO cells. HeLa cells were purchased from ATCC and cultured in medium (DMEM, 10% fetal bovine serum, 1x Pen/Strep and 2 mM Glutamax) at 37°C with 5% CO_2. Cells were tested routinely for mycoplasma.

Stable cell lines.

ATRX-GFP ESCs. GFP-tagged ATRX ESCs were generated using CRISPR/Cas9-mediated homology directed repair[27]. 1×10^6 ESCs were mixed with 2.5 µg target selector plasmid (pCAS9-mCherry harboring ATRX guide RNAs (AAATCACATTGATTTCCCTG)), 2.5 µg frame selector plasmid (pCAS9-mCherry-Frame +1) and 5 µg donor plasmid (pCRISPaint-TagGFP2-PuroR). Electroporation was performed using the Neon transfection system (Thermo Fisher) with the settings: 1400V, 10ms pulse width, and 3 pulses. Cells were selected with 1 µgml^{-1} puromycin for three days. Single-cell clones were analyzed by genotyping to confirm GFP in-frame tagging of the ATRX locus and cell lysates of single clones were subjected to immunoprecipitation to confirm that GFP-tagged ATRX interacts with its complex partner, DAXX.

Addback ESCs. To generate exogenous wild-type or mutants HA-tagged ATRX in ATRX KO ESCs, 1×10^5 ESCs were co-transfected with 1.5 µg wild-type or mutants of PB-ATRX-HA-Neo and 1.5 µg Super PiggyBac transposase using Lipofectamine 3000 (Thermo Fisher). After two days of transfection, cells were selected with 300 µgml^{-1} Geneticin for 5 days and subjected to downstream analysis. Generation of stable H3.2, H3.3, and H3.3 mutant (L126AI130A, LI>AA) addback cell lines was described previously[10]. To generate exogenous DAXX in DAXX KO ESCs, cells were transduced with lentivirus encoding Flag-tagged DAXX and 8 µgml^{-1} polybrene overnight followed by selection with 1 µgml^{-1} puromycin for 3 days. To generate lentivirus, pCDH-Flag-DAXX-Puro plasmid (5 µg) and lentiviral components (5 µg psPAX2 and 1 µg VSVG) were transfected into 2×10^6 293T cells in a 10 cm^2 dish using Lipofectamine 3000.

11

423 Lentivirus-containing supernatants from 48h and 72h post-transfection were concentrated with
424 Lenti-X (Clontech).

425 *ATRX KO HeLa cells.* To generate ATRX KO HeLa cells, 1×10^5 cells were transfected with
426 SpCas9-2A-EGFP plasmid containing a single guide (CAGGATCGTCACGATCAAAGAGG)
427 targeting exon 4 of the human ATRX gene using Lipofectamine 3000. Two days post-
428 transfection, single cells were sorted onto gelatin-coated 96-well plates using a FACSAria
429 Fusion cell sorter according to the instruction of the UT Southwestern Flow Cytometry Facility.
430 Clones were expanded, genotyped, assessed for protein levels, and subject to downstream
431 analysis.

432 **Co-immunoprecipitation.** Nuclear extracts were prepared as previously described[11]. For
433 immunoprecipitation-coupled to mass spectrometry, 5 mg of pre-cleared HeLa nuclear extracts
434 were incubated overnight at 4°C with either 40 µg ATRX antibody or rabbit IgG cross-linked to
435 Dynabeads Protein G (Invitrogen). Beads were washed three times with 5 ml wash buffer (20
436 mM HEPES, pH 7.9, 150 mM KCl, 0.01% NP-40, 10% glycerol, 1 mM DTT, 0.4 mM PMSF) and
437 once with TE buffer. Samples were eluted in 0.1 M glycine, pH 2 and neutralized with 1 M Tris,
438 pH 8. To verify ATRX binding specificity, 10% of immunoprecipitated samples were subjected to
439 immunoblot assay. Remaining samples were run resolved by SDS-PAGE and stained with
440 Coomassie Blue R-250. The desired bands were excised and subjected to Thermo Orbitrap
441 Fusion Lumos mass spectrometer. Protein identification was based on counts and sequences of
442 peptides according to pipelines of UT Southwestern Proteomics Core Facility. For standard
443 immunoprecipitation, 0.5 mg of nuclear extracts were incubated with 2 µg antibody bound to 10
444 µl Dynabeads Protein A or Protein G at 4°C for 2 hours. Beads were washed three times with 1
445 ml wash buffer and eluted in 0.1 M glycine, pH 2. Samples were neutralized with 1 M Tris, pH 8,
446 denatured in SDS loading buffer and subjected to immunoblot assay.

447 **Mass spectrometry analysis.** Raw mass spectrometry data files were converted to a peak list
448 format and analyzed using the central proteomics facilities pipeline (CPFP), version 2.2.0.
449 Peptide identification was performed using the X!Tandem and open MS search algorithm
450 (OMSSA) search engines against the human protein database from Uniprot, with stable
451 contaminants and reversed decoy sequences appended. Fragment and precursor tolerances of
452 20 ppm and 0.5 Da were specified, and three missed cleavages were allowed.
453 Carbamidomethylation of Cys was set as a fixed modification and oxidation of Met was set as a
454 variable modification. Label-free quantitation of proteins across samples was performed using
455 SINQ normalized spectral index software.

456 **Proximity ligation assay (PLA).** Cells were seeded on glass coverslips coated with 8 µgml^{-1}
457 fibronectin in12-well plates. Cell cycle synchronization was performed as previously described[62].
458 Cells were incubated with 2 mM thymidine for 14h and following 50 ngml^{-1} nocodazole treatment
459 for 7h. Mitotic cells were released and newly synthesized DNA was labeled with 2 µM EdU for
460 20min in prior to fixation. Cells were fixed with 4% paraformaldehyde in PBS pH 7.4 for 10min at
461 room temperature then permeabilized in 0.5% Triton X-100 in PBS for 15min.

462   *For ATRX and G4 foci*, cells were fixed and permeabilized as described previously. To detect
463   newly synthesized DNA, EdU was clicked with 488-azide using a Click-iT™ Cell Reaction kit
464   (Thermo Fisher, C10269) according to the manufacturer's protocols. Cells were incubated with
465   100 µgml^{-1} RNase A (Sigma, R4642) for 1h at 37°C and then washed in PBS four times for 5min
466   each. Cells were blocked in Duolink blocking solution for 1h at 37°C and then incubated with
467   0.06 µg BG4 antibody for 1h at 37°C, then with mouse anti-ATRX antibody diluted 1:100 and
468   rabbit anti-DYKDDDDK antibody diluted 1:1,000 overnight at 4°C. All primary antibodies were
469   diluted with Duolink Antibody Diluent. PLA foci were developed with Duolink In Situ Red Starter
470   Kit (Sigma, DUO92101) according to the manufacturer's protocols. The EdU and PLA
471   fluorescence signals in representative images for Figure 1e, Figure 3a, Supplementary Figure
472   2c and Supplementary Figure 7a were pseudocolored red and green, respectively, to clearly
473   show PLA foci.

474   *For detecting G4 at newly synthesized DNA*, *in situ* analysis of protein interactions at DNA
475   replication forks (SIRF) was used[43]. Cells were fixed and permeabilized as described previously.
476   Biotin-azide added to the Click-iT reaction cocktail and incubated with cells. RNase A treatment,
477   cell blocking and BG4 antibody incubation were the same as described previously. Cells were
478   incubated with 1:1000 of mouse biotin antibody and 1:1,000 of rabbit anti-DYKDDDDK antibody
479   in Duolink Antibody Diluent overnight at 4°C. For normalization of G4-EdU foci, cells were
480   separately incubated with 1:1,000 of mouse biotin antibody and 1:1,000 of rabbit biotin antibody.
481   The PLA foci were developed with Duolink In Situ Red Starter Kit. The cover glass was mounted
482   with the mounting medium containing DAPI in the dark overnight and kept at 4°C or -20°C for
483   imaging.

484   **Imaging and processing.** Cells were z-stack imaged using a DeltaVision Elite Deconvolution
485   microscopy (GE Healthcare). Images were further processed by deconvolution using softWoRx
486   software. Mean fluorescence intensities (MFI) of PLA foci between G4 and EdU were counted
487   using maximum intensity z-projection of ImageJ. At least 30 nuclei were counted for each
488   condition. Then the data were normalized to MFI of PLA foci between EdU and EdU in which we
489   used mouse and rabbit antibodies against biotin-clicked EdU.

490   **Cell viability assay.** 0.5×10^3 cells were treated with serial dilutions of PDS (Sigma, SML0678)
491   for five days on a 96-well plate. Cell viability was analyzed by MTT assay (Sigma, M2128).

492   **Chromatin immunoprecipitation (ChIP).** Crosslinking ChIP was performed according to
493   published methods[2,63]. Native ChIP was performed according to published methods[11]. Details
494   are described in supplementary methods.

495   **ChIP-seq and data analysis.** ChIP-seq libraries were prepared from 5 ng ChIP DNA following
496   the Illumina TruSeq protocol. The size of libraries was determined using a D5000 ScreenTape
497   on a 2200 TapeStation (Agilent) and the amount of libraries was quantified using a Qubit dsDNA
498   HS Assay kit (Thermo Fisher). Libraries were paired-end 33-base sequenced on the Illumina
499   NextSeq 500. Typical sequencing depth was at least 20 million reads per sample.

500 *ChIP-seq data quality control, alignment and spike-in normalization.* Quality of ChIP-seq data
501 sets was assessed using the FastQC tool (v.0.11.2). ChIP-seq raw reads were aligned
502 separately to the mouse reference genome (mm10) and the spike-in Drosophila reference
503 genome (dm3) using BOWTIE2 (v.2.2.8). Only one alignment is reported for each read (either
504 the single best alignment or, if more than one equivalent best alignment was found, one of those
505 matches selected randomly). Duplicate reads were filtered using the MarkDuplicates tool of
506 Picard (v.1.127). Uniquely mapped Drosophila reads were counted in the sample containing the
507 least Drosophila mapped reads and used to generate a normalization factor for random
508 downsampling. Reads were converted into bedgraph files using BEDTools (v.2.29.0) and then
509 converted to bigwig using bedGraphToBigWig utility of UCSC kent tools (v317) for visualization
510 in Integrative Genomics Viewer (v.2.3).

511 *Peak calling.* ATRX ChIP and ATRX-GFP ChIP samples were merged using MergeSamFiles
512 tool of Picard (v.2.10.3). Peak calling was performed on the merged files using MACS software
513 (v.1.4.2) using cut-off values '--pvalue 1e-5 --mfold 10, 30' and GFP ChIP in ESCs (without
514 GFP) as input.

515 *Box plots.* Box plot representations were used to quantitatively assess the read distribution. Box
516 plots are defined by the median, box limits at upper and lower quartiles of 75% and 25%, and
517 whiskers at 1.5x interquartile range. The read distribution of the peaks was calculated and
518 plotted using custom R scripts.

519 *Average profiles.* BigWig files were used to generate average ChIP-seq profiles using
520 deepTools (v.3.3.0). The peaks were scaled to the average length of peaks (1300 bp). Read
521 densities surrounding 4 kb (±2 kb) of the scaled peaks were determined and visualized.

522 *G-quadruplex prediction.* G4 motif was predicted by modifying the Quadparser available at:
523 https://github.com/dariober/bioinformatics-cafe/tree/master/fastaRegexFinder. Sequences within
524 peak regions were analyzed by regular expression matching for the G4 motif defined below. The
525 peaks were then assigned to one of two groups: either a G4-containing peak (presence of at
526 least one motif) or a non G4 containing peak (absence of a G4 motif). G4 motif is defined as:
527 '([gG]{3,}\w{1,12}){3,}[gG]{3,}'. The regex looks for 3 or more runs of guanines followed by 1 to
528 12 of any other bases. This is repeated 3 or more times, ending with 3 runs of guanines.

529 *G-quadruplex enrichment analysis.* Simulation was used to estimate if the number of G-
530 quadruplex motif is considerably enriched in the ATRX peaks. The peaks were shuffled using
531 the BEDTools shuffle command at random throughout the genome while maintaining the
532 number of peaks and their size and the number of G-quadruplexes computed. The fold-
533 enrichment analysis was calculated by comparing the actual count of G-quadruplex motif within
534 ATRX peak regions to the average count of the same peak regions after hundred
535 randomizations.
536
537 *Motif Analysis.* Multiple EM for motif elicitation (MEME) was used to interrogate the peak
538 sequence data set for recurring motifs across G4 and non-G4 peaks.
539

14

540   *Peak Annotation.* Peaks were annotated to nearest genes using HOMER annotatePeaks.pl with
541   default settings.

542   **ATAC-seq and analysis.** ATAC-seq was performed and analyzed as previously described[64].
543   Additional details are described in supplementary methods.

544   **CUT&Tag and analysis.** CUT&Tag was performed and analyzed as previously described[6566].
545   Additional details can be found in supplementary methods.

546   **Analysis of mutation rates from patient data.** Observed G-quadruplex regions in the human
547   genome were downloaded from GSE110582 in the form of bedfiles. G4 regions enriched after
548   PDS treatment were identified using G4-seq2 methodology as previously described[67]. For each
549   gene, donors were selected from the International Cancer Genome Consortium Data Portal[68]
550   who were classified as having high impact functional mutations in that gene. All simple somatic
551   mutations (SSMs) within that donor set were then intersected with the above G4 regions using
552   bedtools intersect -u. The total number of substitutions and INDELs for each gene's donor set
553   were counted. Likewise, for each gene, null donor sets were created by randomly sampling the
554   same number of donors from the total pool of 19,729 ICGC donors with available SSM data and
555   counting G4-intersecting substitutions and INDELs within that set. This was repeated 504 times
556   for each gene in order to create a null distribution for that gene.

557

558   Null distribution for each gene was standardized by centering at the mean and scaling by
559   standard deviation (s.d.). For the test observation itself (red line), we standardized by
560   subtracting the mean of the null and scaling by the s.d. of the null. We calculated the p value
561   from first definitions by calculating the area under the curve (AUC) to the right of the observed
562   point using the null distribution.

563

564   **Statistical analysis.** Data analysis was performed using Prism 8 for PLA, MTT assay and
565   ChIP-qPCR. Quantification of PLA data is presented as scatter dots on box-and-whisker plots
566   marking a horizontal median line. In box and whisker plots, box and whiskers indicate 25-75 and
567   10-90 percentiles, respectively. Statistical significance is determined by One-way ANOVA test
568   or two-tailed Mann-Whitney U test, where noted. Number of individual values for each
569   experiment are reported in the figure legends. For MTT data, mock-treated cells at day 0 were
570   taken as 100% survival. Quantification of MTT data is presented as a XY graph or column graph
571   with mean ± SD. Statistical significance is determined by Two-way ANOVA test. Data are
572   representative of three independent experiments unless indicated in the figure legends. Half-
573   maximal inhibitory concentration (IC50) was determined by nonlinear regression. Statistical
574   significance is determined by One-way ANOVA test for ChIP-qPCR. Wilcoxon rank sum test
575   was used to calculate p values for all comparisons in NGS datasets. *p < 0.05; **p < 0.01; ***p <
576   0.001; ns, not significant.

577

578   **Data sets.** The following published next-generation sequencing data sets were meta-analysed
579   in this study: (1) ChIP-ATRX in ESCs[2], H3K9me3 for WT, ATRX KO, DAXX KO, H3.3 KO,
580   ESET KO[69], and ESET KO re-ChIP[11], H3.3 for WT, ATRX KO, DAXX KO, ESET KO[11,60] and
581   H3K27ac for WT, ATRX KO and DAXX KO[60]; (2) RNA-seq in WT, ATRX KO, DAXX KO and

15

582   H3.3 KO[60,70]; (3) ATAC-seq in HIRA KO1, H3.3 KO[60] and H3.3 addback[10]; (4) SNS-seq in

583   ESCs[71]; (5) G4-seq in human cells[67] and (6) Pan-cancer analysis of whole genome[72]. Also,

584   **Supplementary Table 3**.

585

586   **Code Availability.** Code to generate figures is available upon request.

587

588   **Data Availability.** Data files have been deposited in the Gene Expression Omnibus database

589   under accession number GSE---. The mass spectrometry proteomics data have been deposited

590   to MassIVE with the dataset identifier MSV---.

591

592   **Supplementary Data.** Supplementary data is available online.

593

594   **FUNDING**

602

603   **ACKNOWLEDGMENTS**

608

609   **AUTHOR CONTRIBUTIONS**

610   Y.-C.T. and L.B. conceived and designed the study; Y.-C.T. performed experiments with help

611   from V.G., M.L., S.M., and J.W.; A.S. and R.O. performed computational analysis; R.O. and

612   A.B. performed patient mutation analysis; L.A.B. supervised and provided funding for the

613   project; Y.-C.T and L.A.B. wrote the manuscript with input from all authors.

614

615

616

617

618

619

620

Teng et al., Figure 1
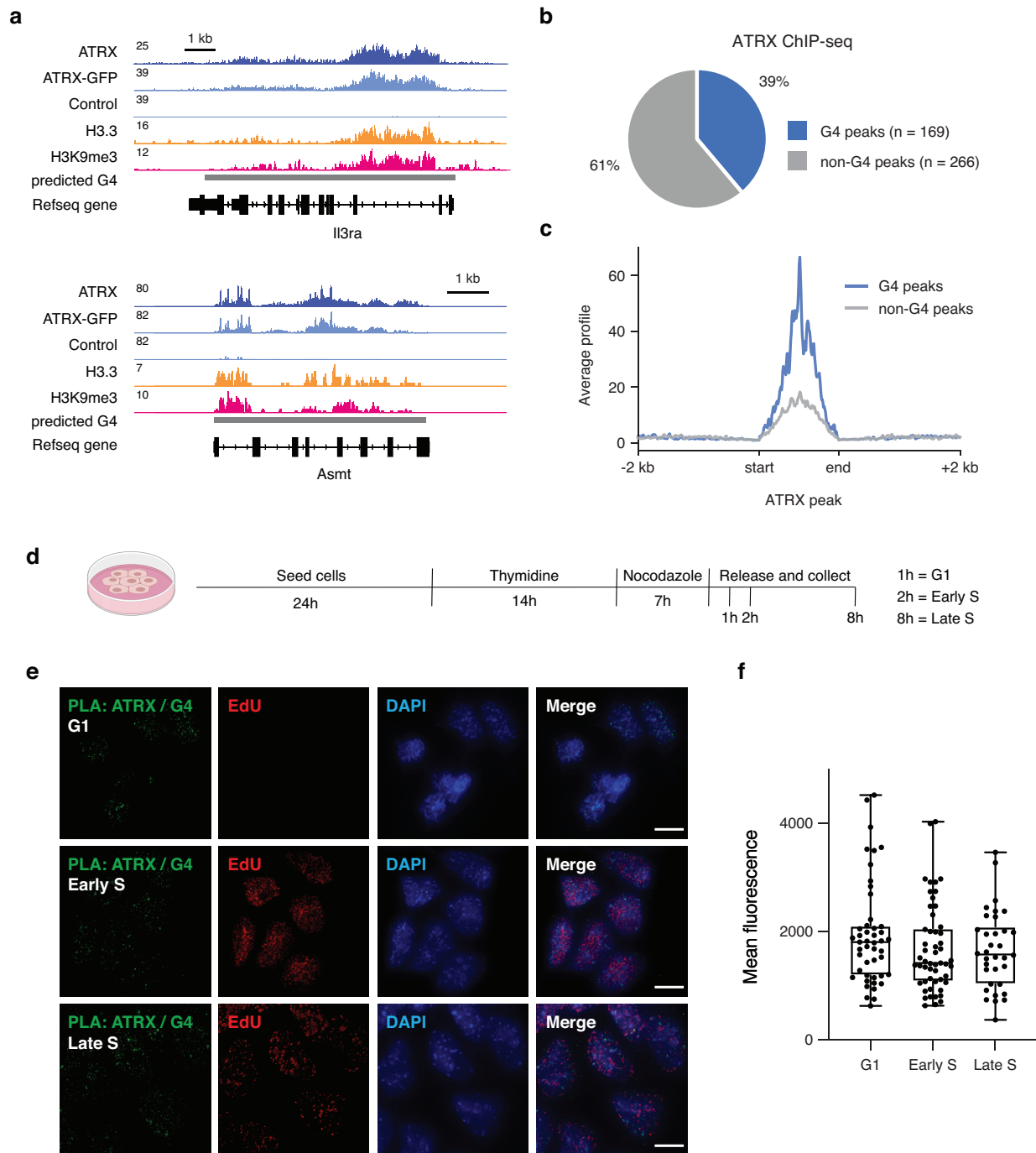
Figure 1 ATRX associates with G4 structures *in vivo*. (**a**) Genome browser representations of ATRX, ATRX-GFP, H3.3, and H3K9me3 ChIP-seq at predicted G4 regions in ESCs. Data represented as read density in reads per kilobase per million mapped reads (RPKM) normalized to an external standard for each data set. Gray boxes indicate predicted G4 regions. (**b**) ChIP-seq analysis of ATRX enrichment in ESCs. Pie chart represents the percentage of ATRX-enriched regions containing G4 consensus motifs (169/435, 39%). (**c**) ATRX ChIP-seq average

17

629   profiles in ESCs at ATRX-enriched regions containing G4 consensus motifs (G4) compared with

630   ATRX-enriched regions that do not contain a G4 consensus motif (non-G4). (**d**) Schematic of

631   ESCs synchronization protocol. Cells are incubated with thymidine for 14h, washed, and treated

632   in medium with nocodazole for 7h. After washing, mitotic cells are released in medium and

633   incubated with EdU in prior cell fixation for downstream experiments. Cells in G1, early S, and

634   late S phase were analyzed 1h, 2h, and 8h after release, respectively. (**e**) Representative

635   images demonstrating ATRX and G4 colocalization by proximity ligation assay (PLA) in

636   synchronized ESCs. Green - PLA (ATRX-G4). Red - EdU-labeling, indicative of newly

637   synthesized DNA. Blue - DAPI nuclear stain. Scale bar equals 10 μm. (**f**) Quantification of signal

638   intensity from ATRX-G4 PLA foci in G1 (n=47), early S (n=53) and late S (n=34) phases of

639   ESCs. No statistical significance was determined by One-way ANOVA test.

640

641

Teng et al., Figure 2
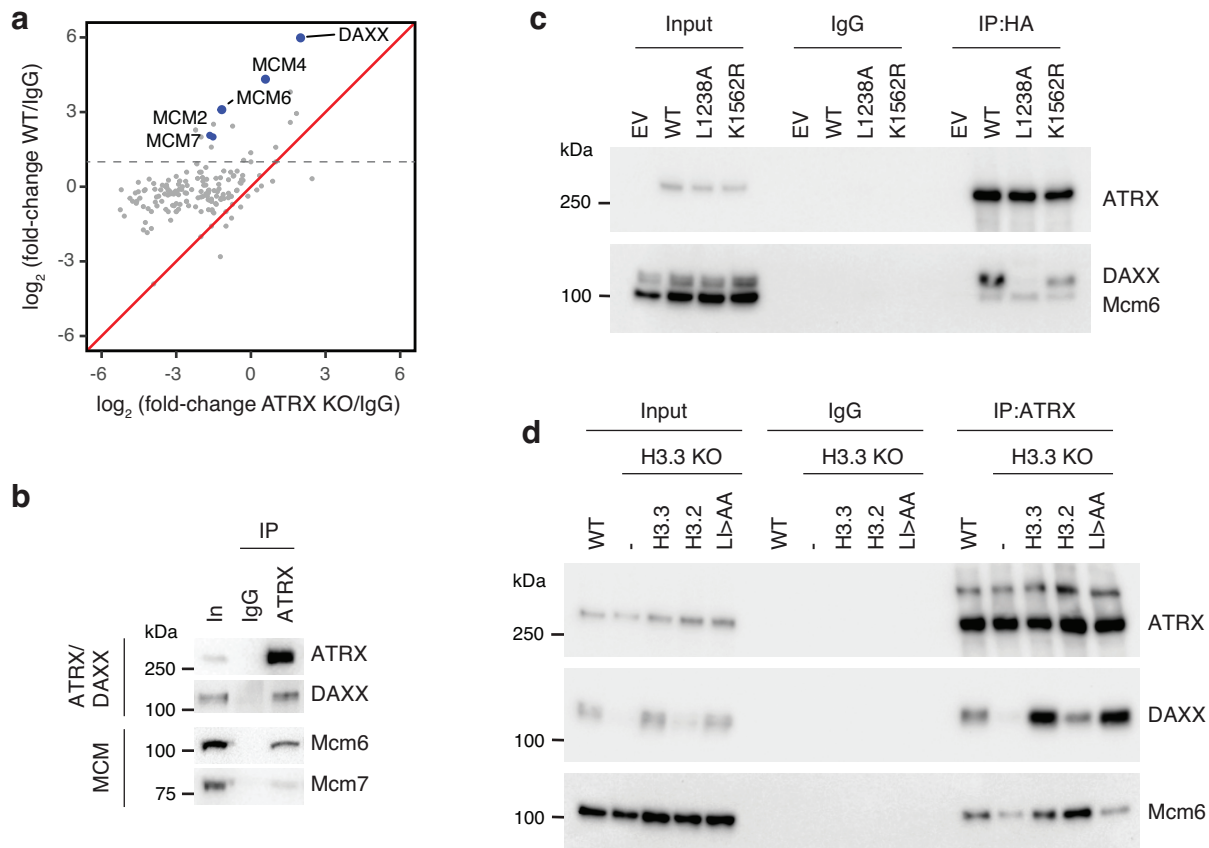
Figure 2 ATRX associates with the MCM DNA helicase. (**a**) Proteomic analysis of ATRX-interacting proteins. Mass spectrometry data represented a scatter plot of the log2 abundance ratios of ATRX-enriched proteins compared to IgG control in both wild-type and ATRX KO HeLa cells. Gray dotted line indicates a fold-change > 2 in ATRX-enriched proteins compared to IgG control. (**b**) Co-immunoprecipitation from wild-type ESC nuclear extracts showing ATRX interaction with DAXX, Mcm6 and Mcm7. (**c**) Co-immunoprecipitation from ATRX KO addback ESC nuclear extracts showing HA-tagged ATRX interaction with DAXX and Mcm6. K1562R - ATRX mutation in helicase domain; L1238A - ATRX mutation in DAXX-binding motif. (**d**) Co-immunoprecipitation from H3.3 KO addback ESC nuclear extracts showing ATRX interaction with DAXX and Mcm6. H3.3 LI/AA - H3.3 deposition mutant.

Teng et al., Figure 3
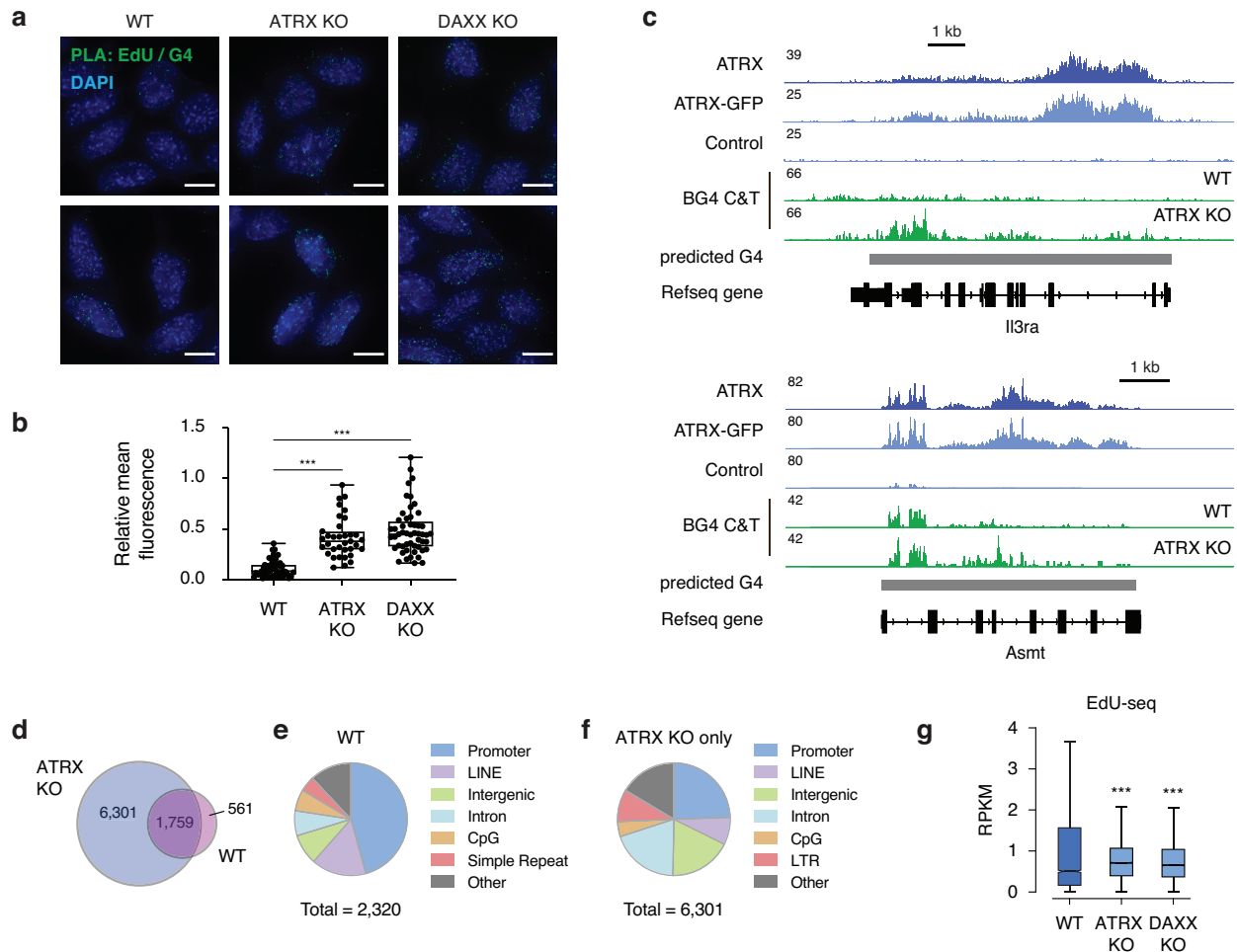
Figure 3 ATRX prevents the accumulation of G4 structures at sites of DNA synthesis. (**a**) Representative images demonstrating EdU and G4 co-localization by proximity ligation assay (PLA) in synchronized ESCs at early S phase. Green - PLA (EdU-G4). Blue - DAPI nuclear stain. Scale bar equals 10 μm. (**b**) Quantification of signal intensity from EdU-G4 PLA foci in wild-type (n=48), ATRX KO (n=35) and DAXX KO (n=52) ESCs. Statistical significance determined by a One-way ANOVA test. ***P < 0.001. (**c**) Genome browser representations of ATRX, ATRX-GFP ChIP-seq and BG4 CUT&Tag at observed G4 regions in ESCs. (**d-f**) CUT&Tag analysis of G4 enrichment in ESCs. (**d**) Venn diagram showing the unique and overlaps between wild-type and ATRX KO ESCs. (**e** and **f**) Pie chart of genomic region annotation in wild-type **e**) and ATRX KO ESCs (**f**). (**g**) Box plots representing EdU-seq read counts of early S phase in wild-type, ATRX KO and DAXX KO ESCs at observed G4 regions that are only identified in ATRX KO (n=6,301). Data are representative of two independent experiments. The bottom and the top of the boxes correspond to the 25th and 75th percentiles, and the internal band is the 50th percentile (median). The plot whiskers correspond to 1.5 interquartile range. Statistical significance determined by Wilcoxon Mann Whitney test. ***p < 0.001.

20

Teng et al., Figure 4

674
675
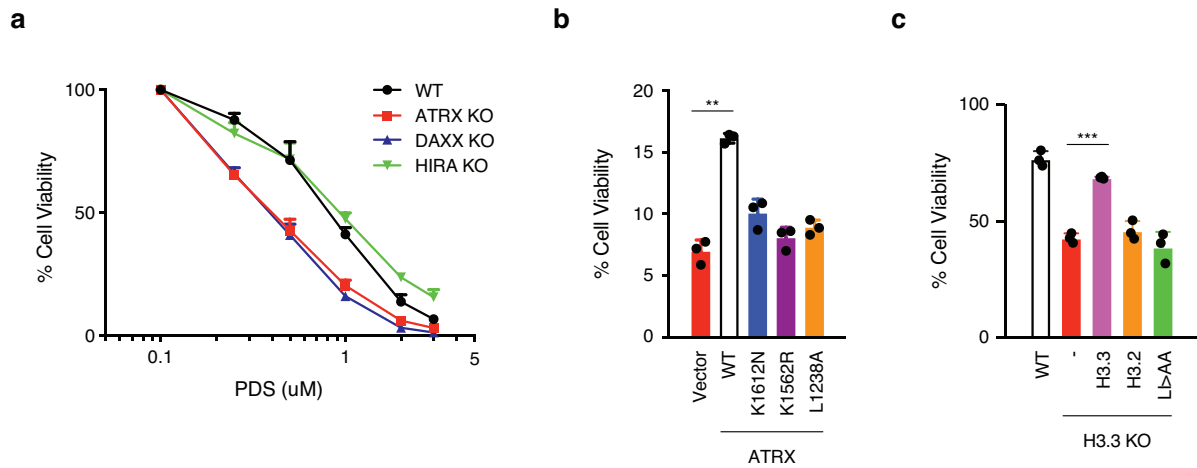
Figure 4 ATRX requires its helicase and chaperone activity to protect from G4-induced stress. (**a**) Cell viability of wild-type, ATRX KO, DAXX KO, and HIRA KO ESCs treated with PDS for 5 days. IC50: wild-type - 0.804 µM, ATRX KO - 0.396 µM, DAXX KO - 0.384 µM, HIRA KO - 0.909 µM. (**b**) Cell viability of ATRX KO ESCs exogenously expressing either wild-type or mutant ATRX constructs treated with 2 µM PDS for 5 days. K1612N and K1562R - ATRX mutations in helicase domain; L1238A - ATRX mutation in DAXX-binding motif. (**c**) Cell viability of H3.3 KO ESCs exogenously expressing either wild-type or mutant H3.3 (L126A I130A, LI/AA) or H3.2 constructs treated with 1 µM PDS for 5 days. For both panels, mock-treated cells at day 0 were taken as 100% survival. Data represented as mean ± SD (n=3). Statistical significance determined by a Two-way ANOVA test. **p < 0.01; ***p < 0.001.

686
687

Teng et al., Figure 5
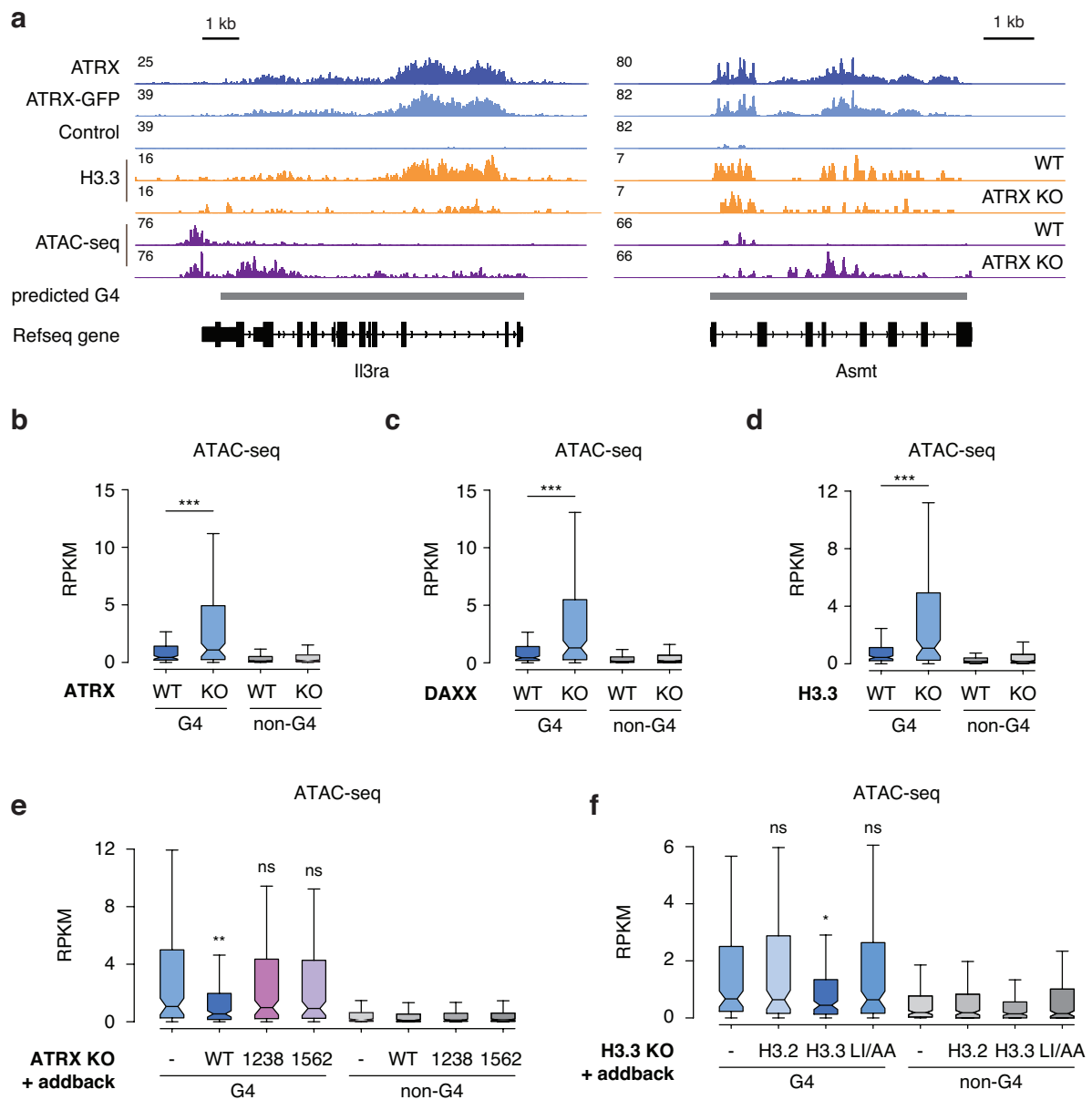
Figure 5 ATRX maintains closed chromatin states at G4 structures. (**a**) Genome browser representations of ATRX, ATRX-GFP, H3.3 ChIP-seq and ATAC-seq at predicted G4 regions in ESCs. Box plots representing ATAC-seq read counts at ATRX-enriched G4 and non-G4 regions in wild-type ESCs compared to (**b**) ATRX KO, (**c**) DAXX KO and (**d**) H3.3 KO ESCs. (**e**) Box plots representing ATAC-seq read counts at ATRX-enriched G4 and non-G4 regions in ATRX KO ESCs exogenously expressing either wild-type or mutant ATRX constructs. (**f**) Box plots representing ATAC-seq read counts at ATRX-enriched G4 and non-G4 regions in H3.3 KO ESCs exogenously expressing either wild-type H3.2, H3.3 or mutant H3.3 constructs. Data are representative of two independent experiments. The bottom and the top of the boxes correspond to the 25th and 75th percentiles, and the internal band is the 50th percentile

22

700   (median). The plot whiskers correspond to 1.5 interquartile range. Statistical significance

701   determined by Wilcoxon Mann Whitney test compared to ATRX KO (**e**) and H3.3 KO in (**f**). *p <

702   0.05; **p < 0.01; ***p < 0.001. ns, not significant.

703

704

Teng et al., Figure 6
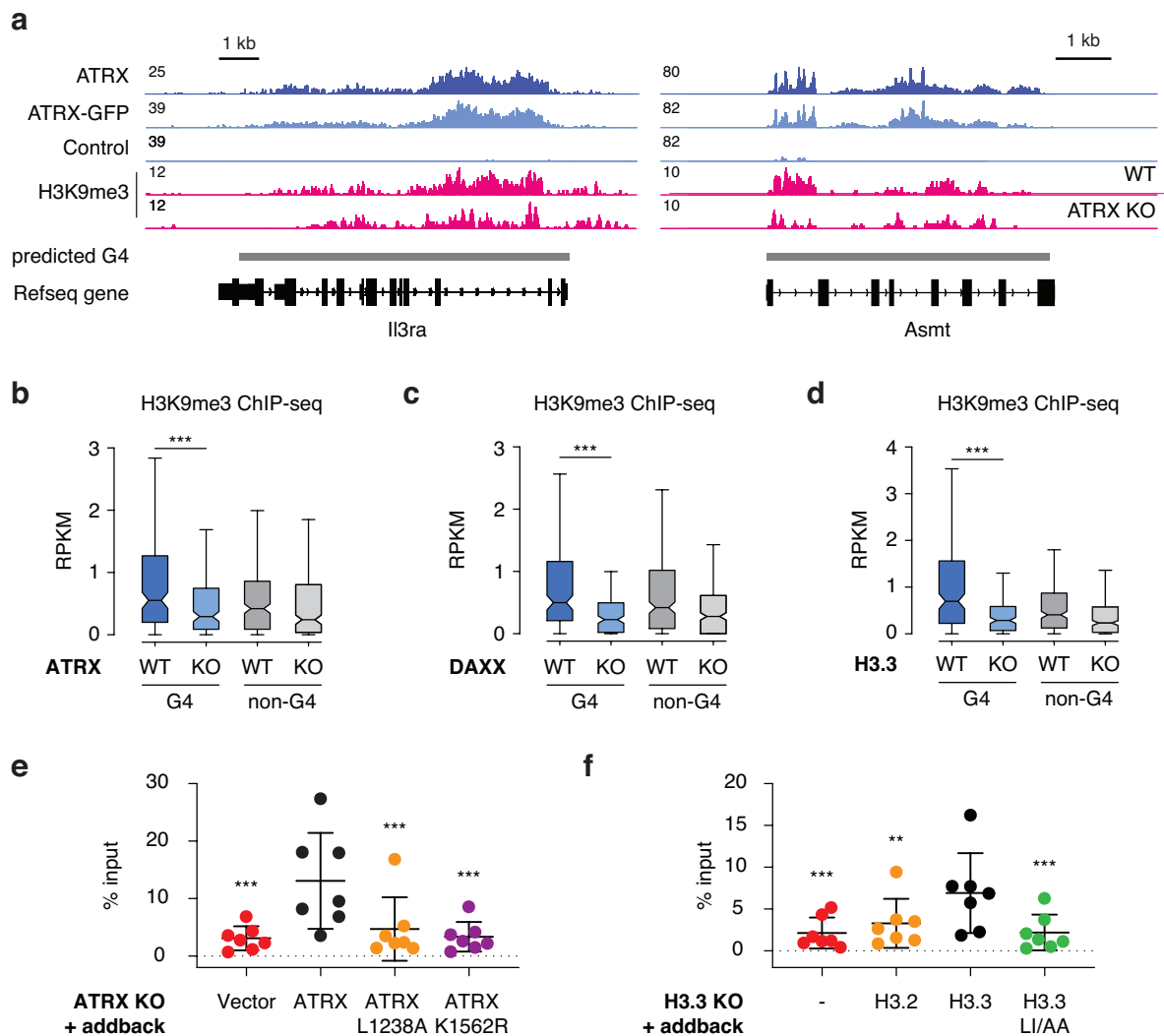
Figure 6 ATRX/DAXX/H3.3 are key factors for H3K9me3 maintenance at G4 regions. (**a**) Genome browser representations of ATRX, ATRX-GFP, and H3K9me3 ChIP-seq at predicted G4 regions in ESCs. Box plots representing ChIP-seq read counts for H3K9me3[11] at ATRX-enriched G4 and non-G4 regions in wild-type ESCs compared to (**b**) ATRX KO, (**c**) DAXX KO and (**d**) H3.3 KO cells. Data are representative of two independent experiments. The bottom and the top of the boxes correspond to the 25th and 75th percentiles, and the internal band is the 50th percentile (median). The plot whiskers correspond to 1.5 interquartile range. Statistical significance determined by Wilcoxon Mann Whitney test. ***p < 0.001. (**e**) ChIP–qPCR of H3K9me3 at ATRX-enriched G4 regions ($n = 7$) in ATRX KO ESCs exogenously expressing either wild-type or mutant ATRX constructs. (**f**) ChIP–qPCR of H3K9me3 at ATRX-enriched G4 regions ($n = 7$) in H3.3 KO ESCs exogenously expressing either wild-type H3.2, H3.3 or mutant H3.3 constructs. Data represent mean ± SD. Statistical significance determined by one-way

24

719   ANOVA compared to ATRX KO + ATRX in (**e**) and H3.3 KO + H3.3 in (**f**). **p < 0.01; ***p <

720   0.001.

721

722

25
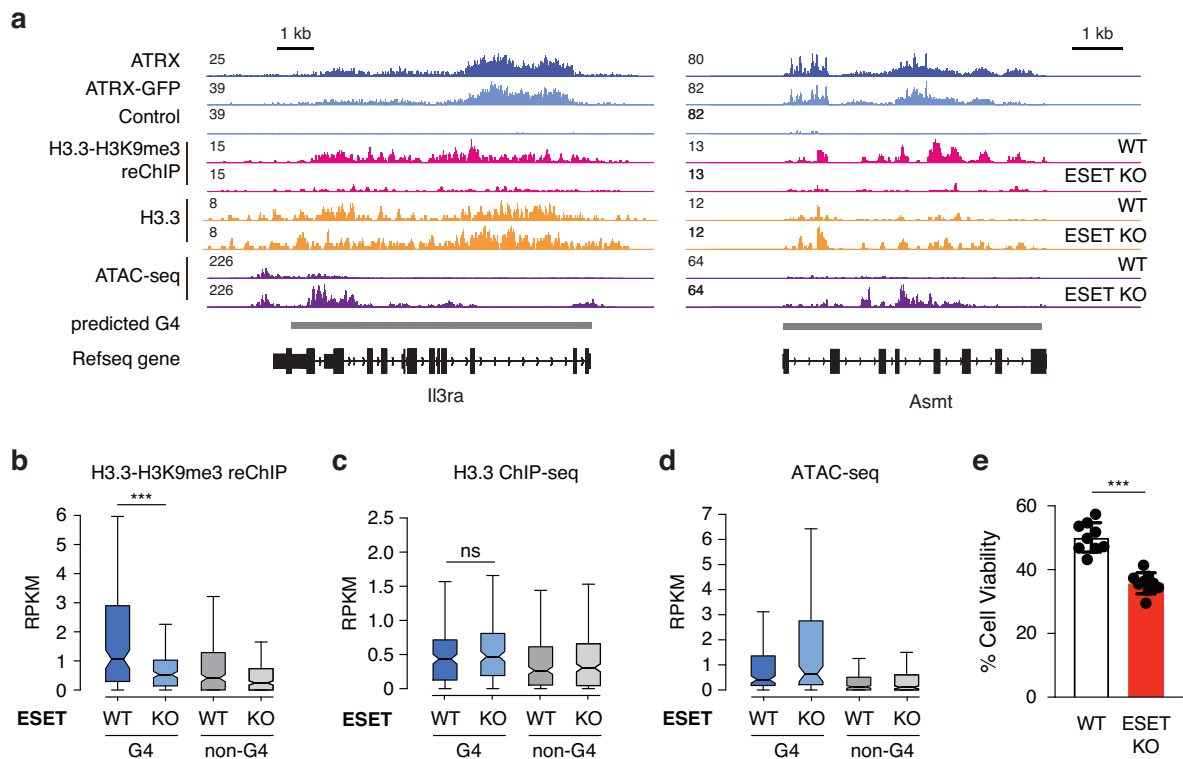
Teng et al., Figure 7

Figure 7 ESET facilitates heterochromatin at G4 regions and protects cells from G4-mediated stress. (**a**) Genome browser representations of ATRX, ATRX-GFP, H3.3-H3K9me3 reChIP-seq, H3.3 ChIP-seq, and ATAC-seq at predicted G4 regions in ESCs. Box plots representing (**b**) H3.3-H3K9me3 reChIP-seq[11], (**c**) H3.3 ChIP-seq[11], and (**d**) ATAC-seq at ATRX-enriched G4 and non-G4 regions in wild-type ESCs compared to ESET KO ESCs. The bottom and the top of the boxes correspond to the 25th and 75th percentiles, and the internal band is the 50th percentile (median). The plot whiskers correspond to 1.5 interquartile range. Statistical significance determined by Wilcoxon Mann Whitney test. ***$p < 0.001$; ns: not significant. (**e**) Cell viability of wild-type and ESET KO ESCs treated with 1 µM of PDS for 3 days. Mock-treated cells at day 0 were taken as 100% survival. Data represented as mean ± SD (n=9). Statistical significance determined by Two-way ANOVA test. ***$p < 0.001$.

26
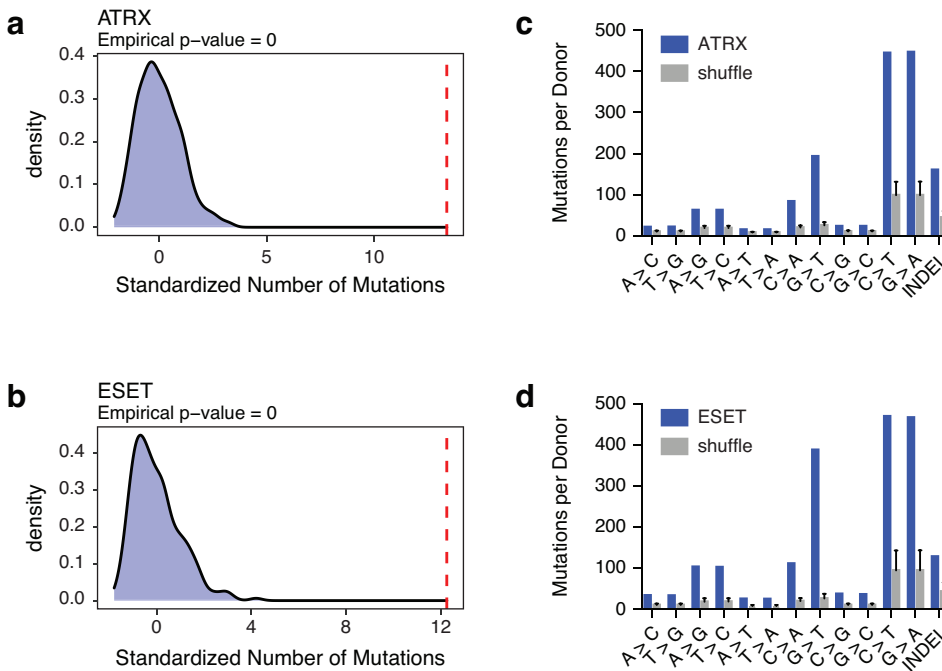
Teng et al., Figure 8

Figure 8 Mutations at G4 regions are highly correlated with ATRX and ESET mutations in human tumors. (**a** and **b**) Red line shows standardized number of mutations at G4 regions in ATRX (**a**) or ESET (**b**) mutant tumors. Histograms show the mutation density at observed G4 regions in an iteratively and randomly selected patient cohort of the same size. p-value calculated as described in methods. (**c** and **d**) Analysis of single-nucleotide mutations and insertion-deletion (INDEL) mutations in the ATRX (**c**) or ESET (**d**) mutant tumors compared with the shuffled patient cohort described above.

Teng et al., Figure 9
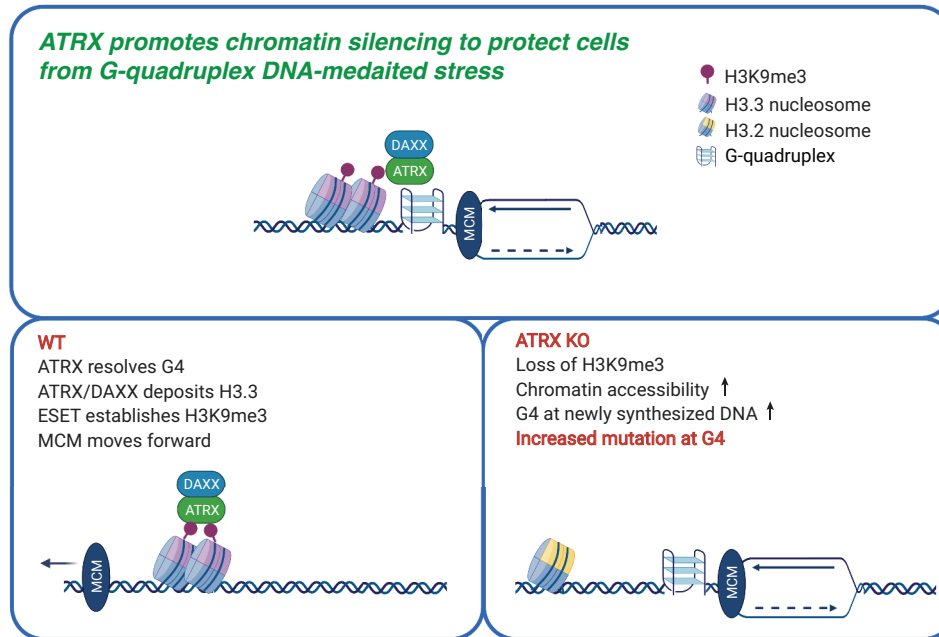
Figure 9 ATRX is a critical chromatin remodeler at G-quadruplex regions during DNA replication. When the replicative MCM helicase complex encounters G4 DNA, ATRX resolves G4 DNA through its helicase activity and H3.3 deposition activity to facilitate MCM progression. These activities are ultimately upstream of ESET-mediated heterochromatin formation to protect cells from G4-mediated replicative stress. Thus, ATRX/DAXX/H3.3 and ESET cooperate to prevent physical G-quadruplex stress and maintain genome stability.

Teng et al., Supplementary Fig. 1

759
760
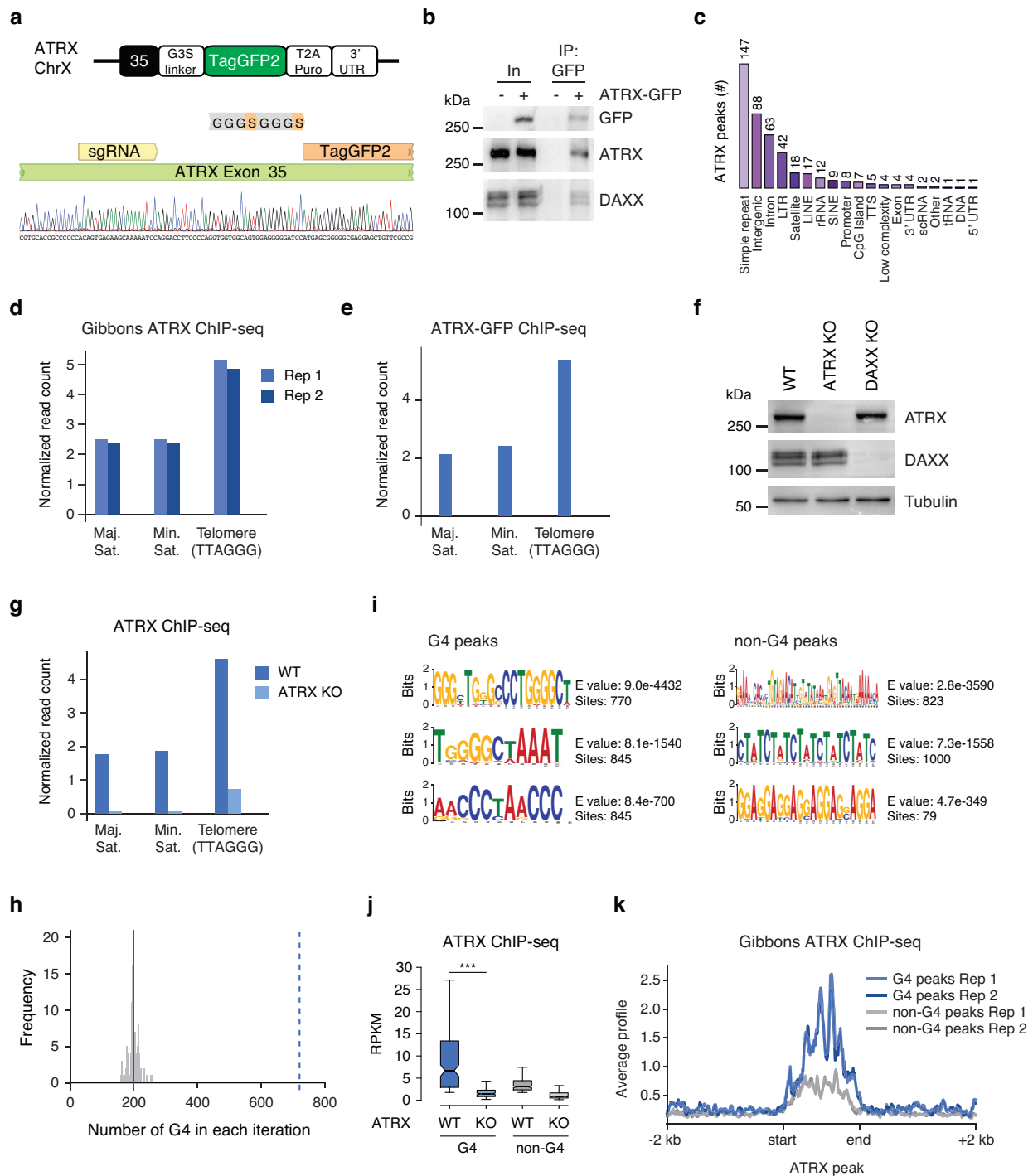
761 **Supplementary Figure 1. ATRX is enriched at repetitive elements and predicted G4**
762 **elements.** Related to Figure 1. (**a**) Schematic of GFP tagging of the endogenous mouse ATRX
763 gene. An sgRNA targets exon 35 of ATRX gene. A CRISPR/Cas9-mediated homology directed
764 repair system allows selection for positively tagged cells based on expression of a puromycin

29

765   resistance gene separated from the GFP sequence by a T2A peptide. Genotyping shows a
766   single-cell clone with the GFP gene inserted in-frame into the ATRX gene. (**b**) Co-
767   immunoprecipitation from ESC nuclear extracts showing ATRX-GFP interaction with DAXX. (**c**)
768   Peak Annotation. The x axis represents the annotation category and the number above the bar
769   graph represents the number of peaks in the particular annotation category. (**d**) Published
770   ATRX ChIP-seq in ESCs shows ATRX enrichment at telomeres and satellite sequences[2]. (**e**)
771   ATRX-GFP ChIP-seq in ESCs shows ATRX-GFP enrichment at telomeres and satellite
772   sequences. (**f**) Immunoblot from ESC whole-cell lysates showing the expression of ATRX and
773   DAXX in wild-type, ATRX KO and DAXX KO ESCs. Tubulin as loading control. (**g**) ATRX ChIP-
774   seq in ESCs shows ATRX enrichment at telomeres and satellite sequences. This enrichment is
775   reduced in ATRX KO ESCs. (**h**) The histogram shows the number of a G-quadruplex motif in
776   random sequences compared with the 720 observed number of G-quadruplex motif (dashed
777   line). Simulation was used to estimate the number of G-quadruplex motif by chance. For 100
778   times, the peaks were randomly shuffled throughout the genome while maintaining the number
779   of peaks and their size and the number of G-quadruplexes computed. On average,198 G-
780   quadruplexes (sd= 17) (solid line) were observed therefore the observed number of G-
781   quadruplex motifs is significantly larger (by 30 standard deviations) than expected by chance. (**i**)
782   Motif enrichment for ATRX-enriched G4 (left) and non-G4 (right) peaks. (**j**) Box plots
783   representing ATRX ChIP-seq in ESCs at ATRX-enriched G4 and non-G4 regions in wild-type
784   and ATRX KO cells. The bottom and the top of the boxes correspond to the 25th and 75th
785   percentiles, and the internal band is the 50th percentile (median). The plot whiskers correspond
786   to 1.5 interquartile range. Statistical significance is determined by Wilcoxon Mann Whitney test.
787   ***$P < 0.001$. (**k**) Published ATRX ChIP-seq average profiles in ESCs for ATRX-enriched G4
788   and non-G4 regions[2].
789
790

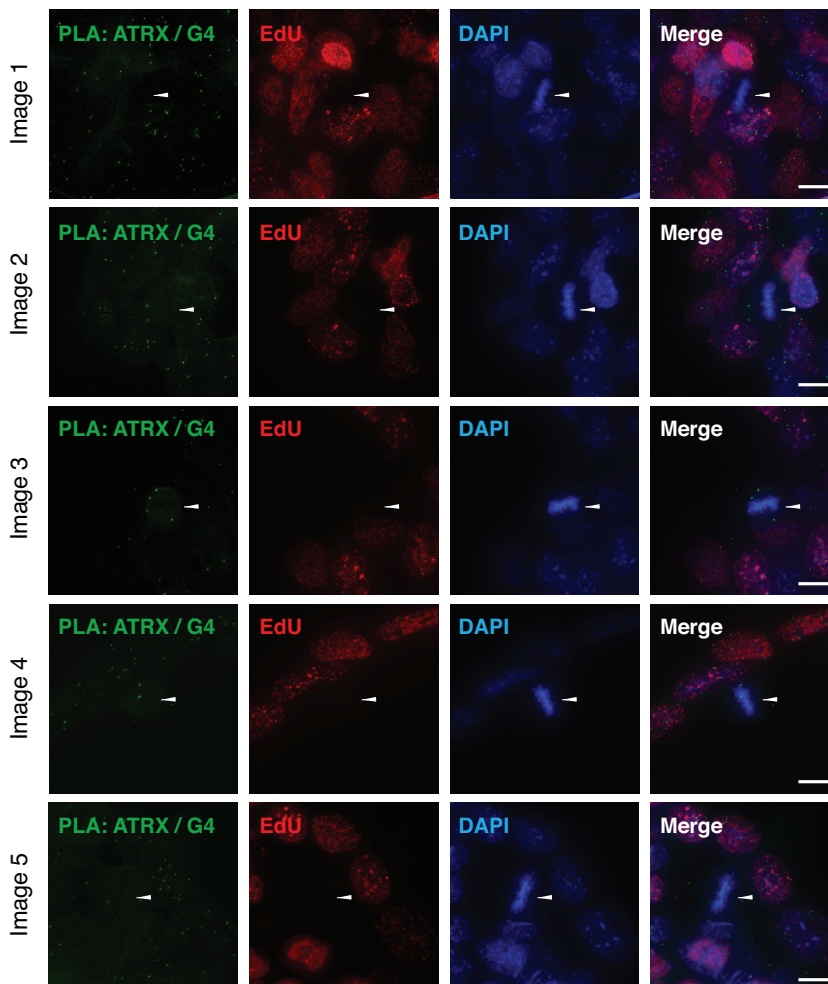# Teng et al., Supplementary Fig. 2

a

b

c

791

31
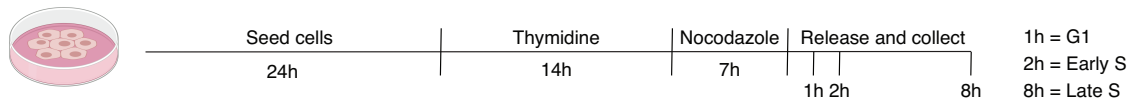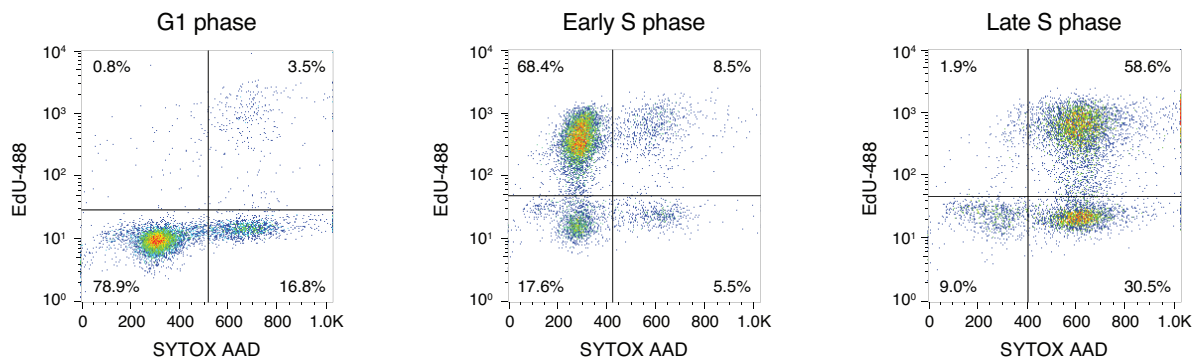
792

793 **Supplementary Figure 2. Cell cycle analysis of synchronized ESCs.** Related to Figure 1. (**a**)

794 Schematic of ESCs synchronization protocol. Cells are incubated with thymidine for 14h,

795 washed, and treated in medium with nocodazole for 7h. After washing, mitotic cells are released

796 in medium and incubated with EdU in prior cell fixation for downstream experiments. Cells in

797 G1, early S, and late S phase were analyzed 1h, 2h, and 8h after release, respectively. (**b**) Cell

798 cycle profiles of ESCs in G1, early S and late S phases. (**c**) Representative images assessing

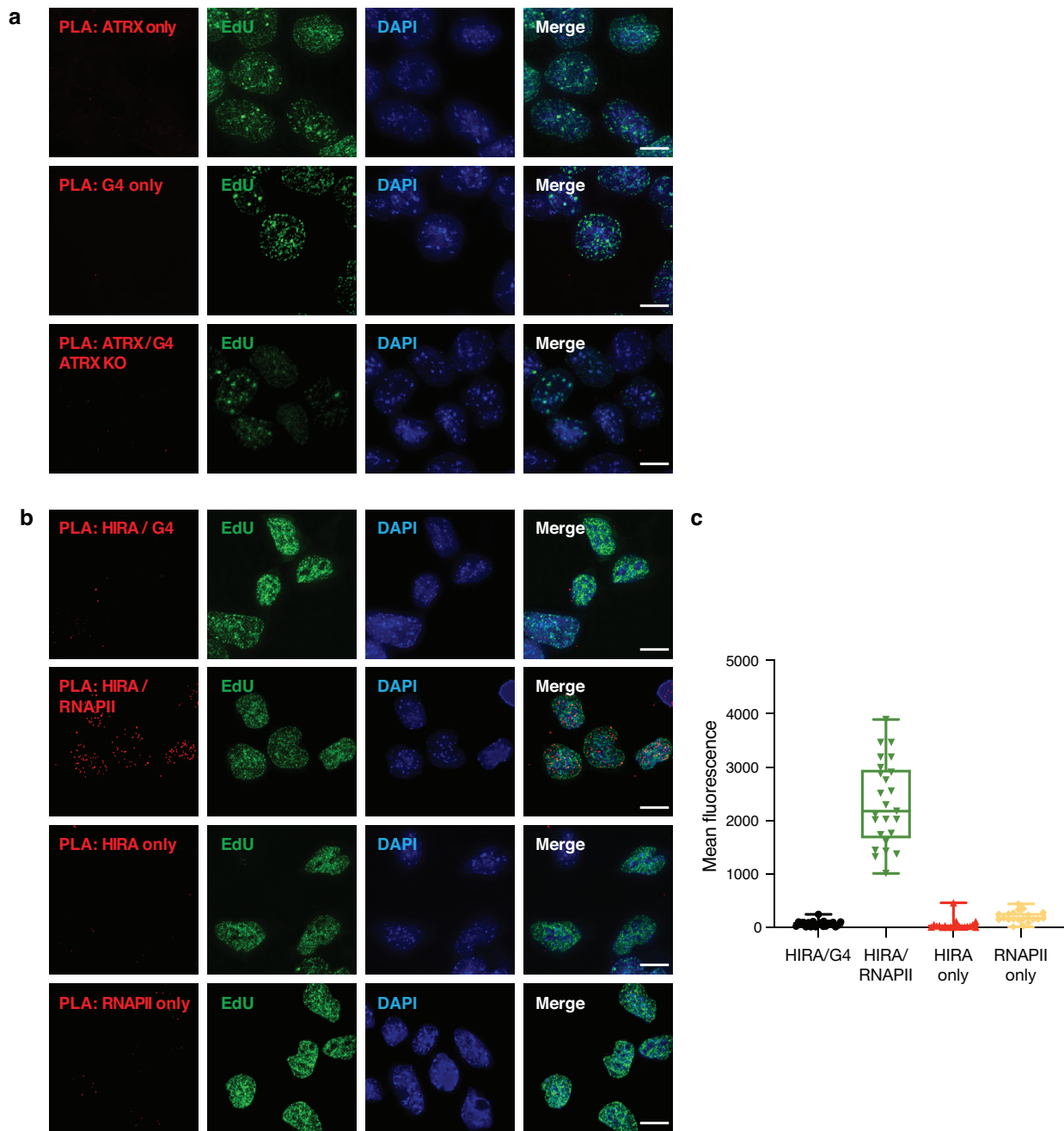799 ATRX and G4 colocalization by proximity ligation assay (PLA) in asynchronized ESCs. Green -

800 PLA (ATRX-G4). Red - EdU-labeling, indicative of newly synthesized DNA. Blue - DAPI nuclear

801 stain. Scale bar equals 10 μm. Arrows indicate metaphase chromosomes.

802

803

Teng et al., Supplementary Fig. 3

804
805
806 **Supplementary Figure 3. Negative controls for proximity ligation assay.** Related to Figure
807 1. (**a**) Representative images displaying PLA foci when ATRX or BG4 antibody alone was
808 applied in ESCs (top and middle panels) and when both ATRX and BG4 antibodies were
809 applied in ATRX KO ESCs (bottom panels). (**b**) Representative images showing the PLA foci in
810 ESCs when both HIRA and BG4 antibodies were applied (top panels), when both HIRA and
811 RNA polymerase II (RNAPII) antibodies were applied (second panels) and when HIRA or
812 RNAPII antibody alone was applied (third and bottom panels). For a single antibody experiment,

813   all procedures of PLA were described in methods except only one antibody used. HIRA and

814   RNAPII antibodies used 1:1,000 dilution. Red - PLA foci. Green - EdU-labeling, indicative of

815   newly synthesized DNA. Blue -DAPI nuclear stain. Scale bar equals 10 μm. (**c**) Quantification of

816   signal intensity from PLA foci in ESCs shown in **Supplementary Figure 3b**. HIRA/G4 (n=22),

817   HIRA/RNAPII (n=25), HIRA only (n=21), and RNAPII only (n=20).

818

819

Teng et al., Supplementary Fig. 4
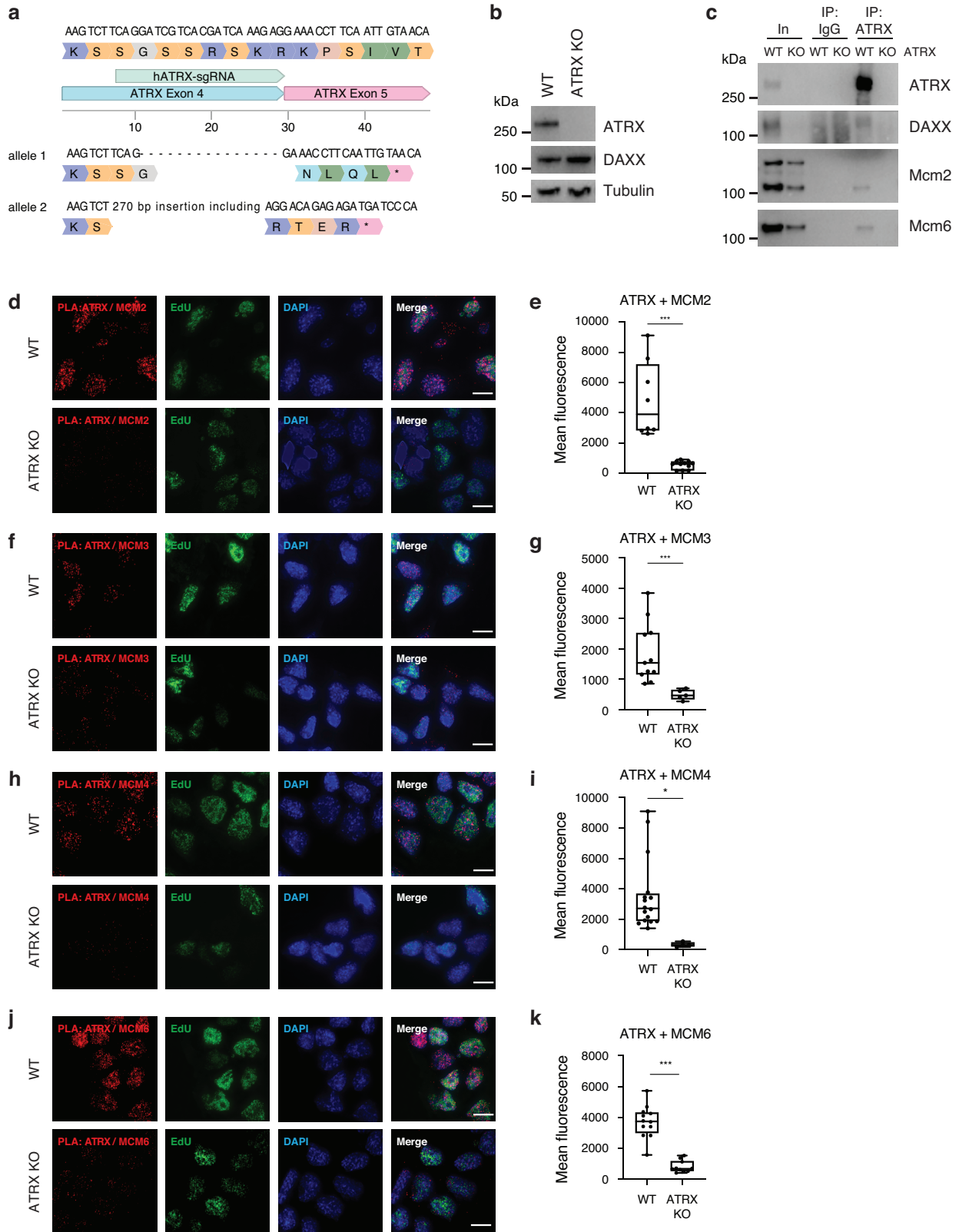
822 **Supplementary Figure 4. Validation of interaction between ATRX and MCM proteins.**
823 Related to Figure 2. (**a**) Generation of CRISPR/Cas9 mediated ATRX KO cells. HeLa cells were
824 transfected with a Cas9 expression plasmid containing an sgRNA targeting exon 4 of the ATRX
825 gene. A single clone was isolated and validated by Sanger sequencing. (**b**) Immunoblot from
826 HeLa whole-cell lysates showing the expression of ATRX and DAXX in wild-type and ATRX KO
827 cells. Tubulin as loading control. (**c**) Co-immunoprecipitation from HeLa nuclear extracts
828 showing ATRX interaction with DAXX, MCM2 and MCM6. (**d, f, h, and j**) Representative images
829 demonstrating ATRX and Mcm2, Mcm3, Mcm4, and Mcm6 co-localization by proximity ligation
830 assay (PLA) in early S phase of wild-type and ATRX KO ESCs. Red - PLA (ATRX-Mcm
831 proteins). Green - EdU-labeling, indicative of newly synthesized DNA. Blue - DAPI nuclear stain.
832 Scale bar equals 10 μm. (**e, g, i, and k**) Quantification of signal intensity from ATRX-Mcm PLA
833 foci in early S phase of ESCs in **Supplementary Figure 4d, 4f, 4h, and 4j**. Statistical
834 significance is determined by Mann-Whitney U-test. *P < 0.05; ***P < 0.001. In (**e**), wild-type
835 (n=8) and ATRX KO (n=11); In (**g**), wild-type (n=11) and ATRX KO (n=5); In (**i**), wild-type (n=16)
836 and ATRX KO (n=4); In (**k**), wild-type (n=12) and ATRX KO (n=10).
837
838

Teng et al., Supplementary Fig. 5
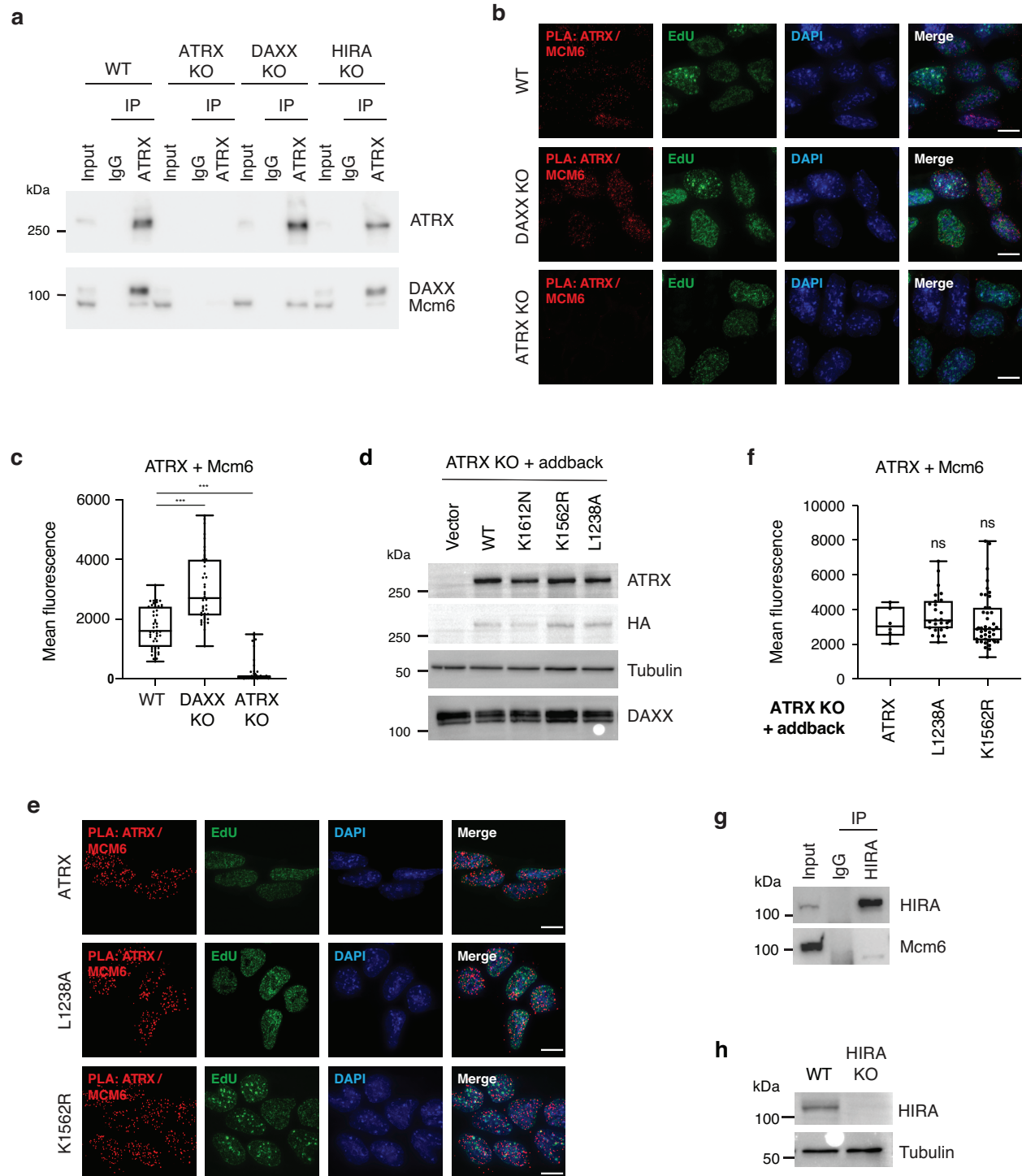
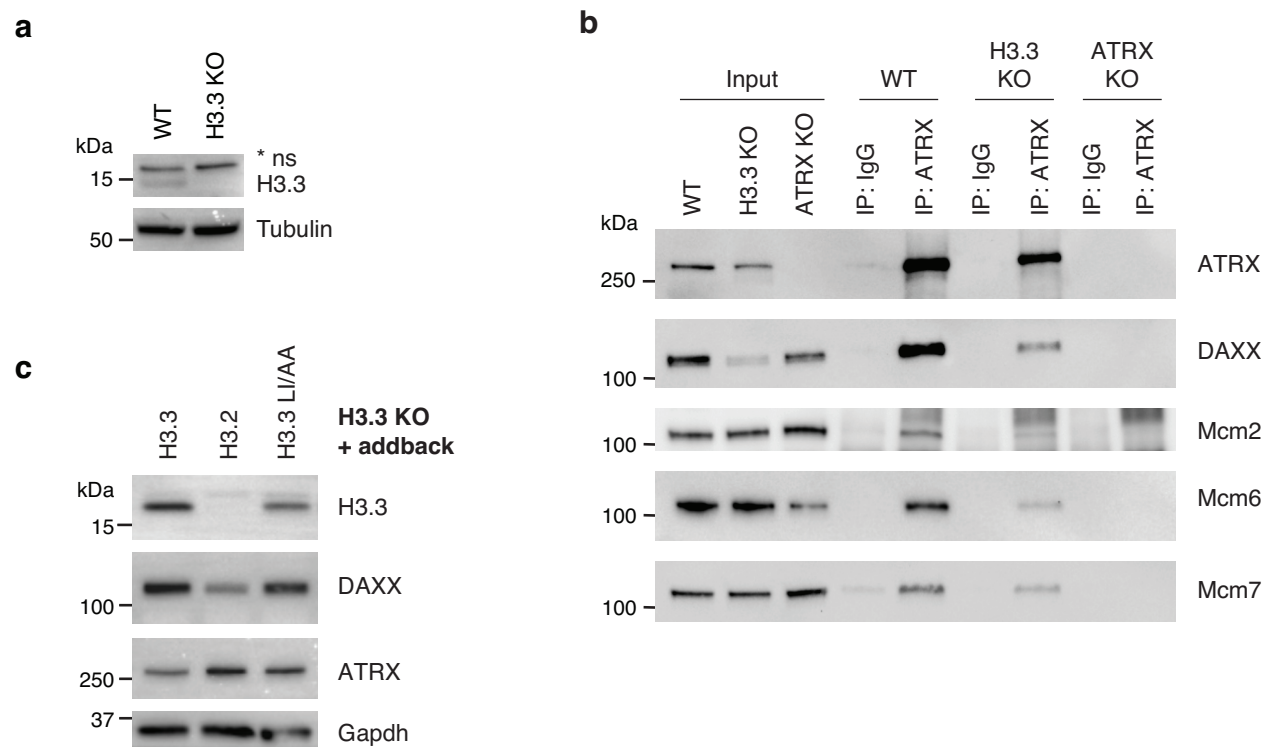**839**
840
841 **Supplementary Figure 5. ATRX-Mcm6 interaction does not require DAXX.** Related to
842 Figure 2. (**a**) Co-immunoprecipitation from ESC nuclear extracts showing ATRX interaction with
843 DAXX and Mcm6 in wild-type, ATRX KO, DAXX KO, and HIRA KO ESCs. (**b**) Representative
844 images demonstrating ATRX-Mcm6 co-localization by proximity ligation assay (PLA) in

845   asynchronized wild-type, ATRX KO (n=47), and DAXX KO ESCs (n=42). (**c**) Quantification of

846   signal intensity from ATRX-Mcm6 PLA foci in **Supplementary Figure 5b**. (**d**) Immunoblot from

847   ESC whole-cell lysates showing the expression of wild-type and mutants of HA-tagged ATRX in

848   ATRX KO cells. DAXX expression remains the same within each cell lysate. (**e**) Representative

849   images demonstrating ATRX-Mcm6 co-localization by proximity ligation assay (PLA) in early S

850   phase of ATRX KO ESCs that expressing exogenous wild-type ATRX and ATRX mutants

851   (L1238A and K1562R). (**f**) Quantification of signal intensity from ATRX-Mcm6 PLA foci in

852   **Supplementary Figure 5e**. ATRX (n=6), L1238A (n=26), and K1562R (n=43). (**g**) Co-

853   immunoprecipitation from HeLa nuclear extracts showing HIRA does not interact with MCM6.

854   (**h**) Immunoblot from ESC whole-cell lysates showing the expression of HIRA in wild-type and

855   HIRA KO cells. For **d** and **h**, tubulin as loading control. For **b** and **e**, Red - PLA (ATRX-Mcm6).

856   Green - EdU-labeling, indicative of newly synthesized DNA. Blue - DAPI nuclear stain. Scale bar

857   equals 10 μm. Statistical significance determined by One-way ANOVA test. ***$p < 0.001$. ns; not

858   significant.

859

860

Teng et al., Supplementary Fig. 6

861
862
863 **Supplementary Figure 6. ATRX-Mcm6 interaction is partially reduced in H3.3 KO.** Related
864 to Figure 2. (**a**) Immunoblot from ESC whole-cell lysates showing H3.3 expression levels in wild-
865 type and H3.3 KO cells. ns, non-specific band. (**b**) Co-immunoprecipitation from ESC nuclear
866 extracts showing ATRX interaction with DAXX, Mcm2, Mcm6 and Mcm7 in wild-type, H3.3 KO,
867 and ATRX KO cells. (**c**) Immunoblot form ESC whole-cell lysates showing the expression of
868 H3.3, DAXX, and ATRX in H3.3 KO cells exogenously expressing H3.3, H3.2 or H3.3 L126A
869 L130A (H3.3 Ll/AA). For **a** and **c**, tubulin and Gapdh as loading control.
870
871

Teng et al., Supplementary Figure 7
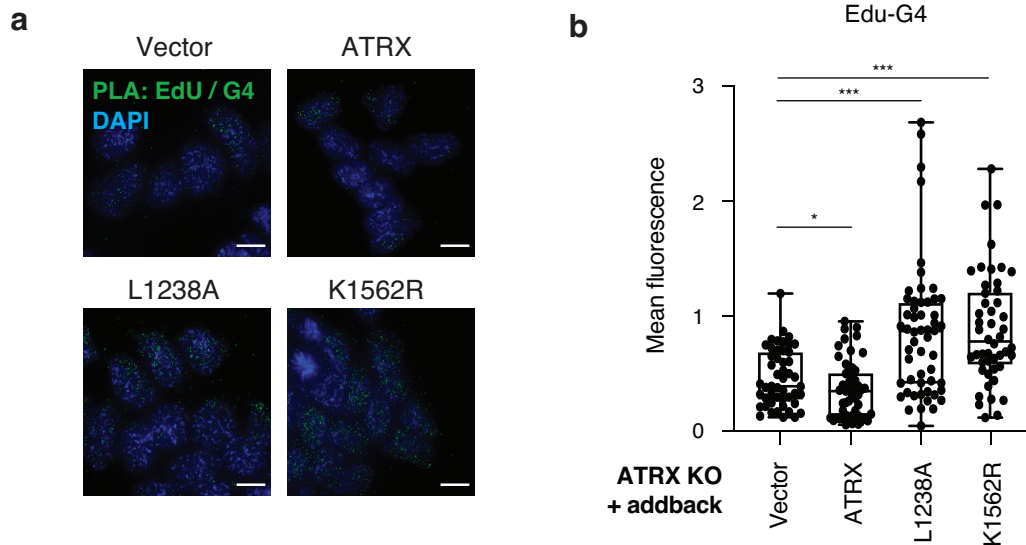
872
873
874 **Supplementary Figure 7. ATRX requires its helicase and chaperone activity for**
875 **preventing G4 formation at sites of DNA synthesis.** Related to Figure 3. (**a**) Representative
876 images demonstrating EdU and G4 co-localization by proximity ligation assay (PLA) in early S
877 phase of ATRX KO ESCs exogenously expressing wild-type ATRX (n=53) and ATRX mutants
878 (L1238A, n=54 and K1562R, n=48). Green - PLA (EdU-G4). Blue - DAPI nuclear stain. Scale
879 bar equals 10 μm. (**b**) Quantification of signal intensity from EdU-G4 PLA foci in
880 **Supplementary Figure 7a**. Statistical significance determined by a One-way ANOVA test. *p <
881 0.05; ***p < 0.001.
882
883

Teng et al., Supplementary Figure 8
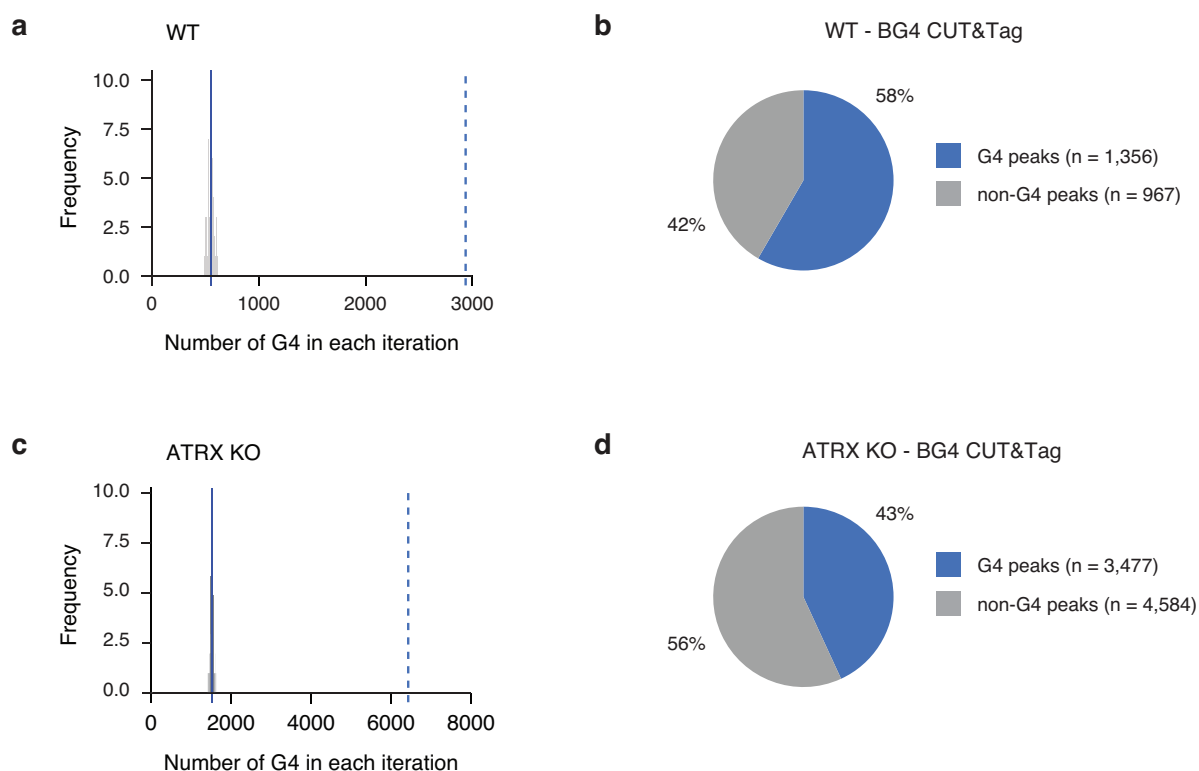

Supplementary Figure 8. G4 prediction on observed G4 peaks of CUT&Tag. Related to Figure 3. (**a**) The histogram shows the number of G4 motifs in random sequences compared with the 2,936 observed number of G4 motifs (dashed line) identified in WT ESCs. Simulation was used to estimate the number of G4 motifs by chance. For 100 times, the peaks were randomly shuffled throughout the genome while maintaining the number of peaks and their size and the number of G-quadruplexes computed. On average, 550 G-quadruplexes (sd= 28) (solid line) were observed therefore the observed number of G-quadruplex motifs is significantly larger (by 85 standard deviations) than expected by chance. (**b**) BG4 CUT&Tag analysis in WT ESCs. Pie chart represents the percentage of BG4-enriched regions containing G4 consensus motifs (1,356/2,323, 58%). (**c**) As described in panel a, but for ATRX KO ESCs. 6,428 G4 motifs were observed compared to simulation average of 1,511 (sd=45). Observed G4 motifs in ATRX KO ESCs were 110 standard deviations larger than expected by chance. (**d**) BG4 CUT&Tag analysis in ATRX KO ESCs. Pie chart represents the percentage of BG4-enriched regions containing G4 consensus motifs (3,477/8,061, 43%).

Teng et al., Supplementary Figure 9
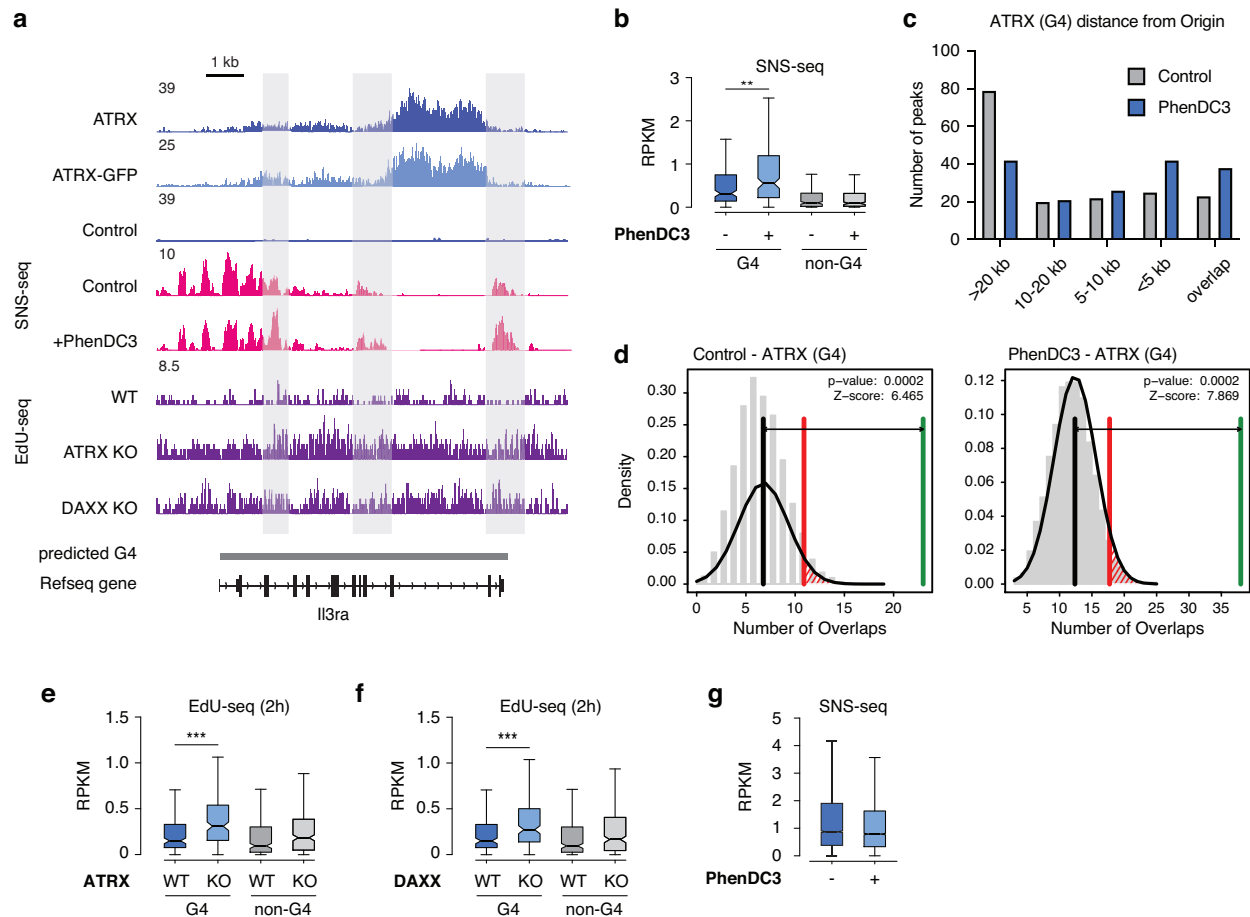
Supplementary Figure 9. Analysis of G4-related DNA synthesis and origin activity.
Related to Figure 3. (**a**) Genome browser representations of ATRX, ATRX-GFP ChIP-seq, SNS-seq, and EdU-seq at predicted G4 regions in ESCs. The SNS-seq in ESCs treated with DMSO (control) or G4 stabilizer PhenDC3 (10 µM) for 48h was described previously (Prorok, P. et al. 2019). Data represented as read density in reads per kilobase per million mapped reads (RPKM) normalized to an external standard for each data set. Gray boxes indicate predicted G4 regions. (**b**) Box plots representing SNS-seq (Prorok, P. et al. 2019) in ESCs showing origin enrichment at ATRX-enriched G4 and non-G4 regions in the presence and absence of PhenDC3. (**c**) The distance between ATRX G4 peaks and origins defined in either SNS control or SNS PhenDC3 were determined using bedtools closest. In case of a tie, the first value is reported. The peaks were classified based on the distance to the origin in the following groups: overlap, less than 5kb, 5-10kb, 10-20kb, and more than 20 kb. The x-axis represents the distance category, and the y-axis represents the number of peaks in the particular distance category. (**d**) Statistical analysis of origin overlap shown in panel c. The number of overlaps (origins and ATRX sites) in random sequences compared with the observed number of overlaps (green line) identified in ESCs either control or PhenDC3. Simulation was used to estimate the number of overlaps by chance. For 100 times, the peaks were randomly shuffled throughout the

921   genome while maintaining the number of peaks and their size and the number of origins
922   computed. After simulation, the number of overlaps (black line) were observed and the statistic
923   was tested at α = 0.05 (red line). (**e** and **f**) Box plots representing EdU-seq read counts at early
924   S phase at ATRX-enriched G4 and non-G4 regions in wild-type ESCs compared to (**e**) ATRX
925   KO and (**f**) DAXX KO ESCs. Data are representative of two independent experiments. (**g**) Box
926   plots representing SNS-seq read counts in control or PhenDC3-treated ESCs at G4 regions
927   experimentally identified after ATRX KO. The bottom and the top of the boxes correspond to the
928   25th and 75th percentiles, and the internal band is the 50th percentile (median). The plot
929   whiskers correspond to 1.5 interquartile range. Statistical significance determined by Wilcoxon
930   Mann Whitney test. **$p < 0.01$; ***$p < 0.001$.
931
932

Teng et al., Supplementary Fig. 10
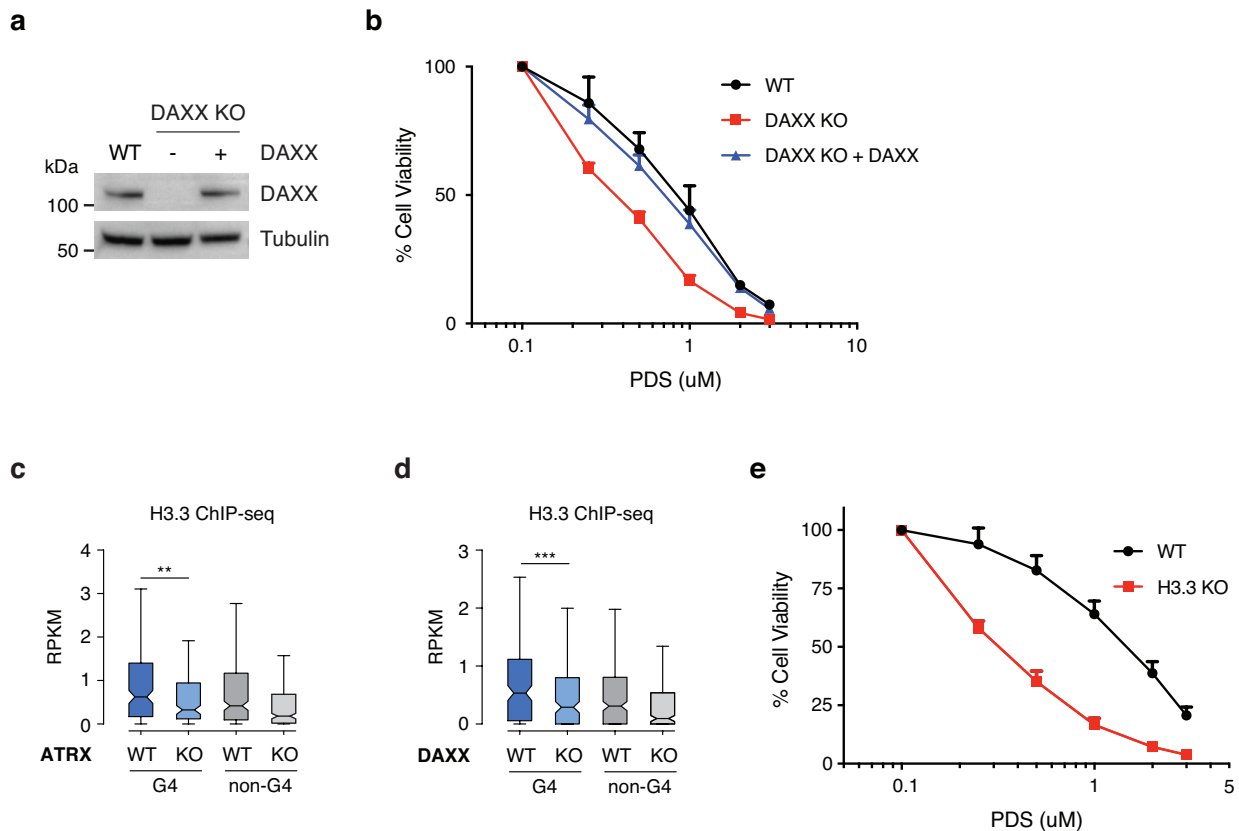
**Supplementary Figure 10. DAXX KO and H3.3 KO cells are sensitive to the G4 stabilizer, PDS.** Related to Figure 4. (**a**) Immunoblot from ESC whole-cell lysates showing the expression of DAXX in wild-type ESCs and DAXX KO ESCs with or without exogenous DAXX expression constructs. Tubulin as loading control. (**b**) Cell viability of wild-type, DAXX KO, and DAXX KO ESCs expressing exogenous DAXX treated with PDS for 5 days. Data is one independent experiment with three technical replicates. (**c** and **d**) Box plots representing H3.3 ChIP-seq in ESCs[11] at ATRX-enriched G4 and non-G4 regions in wild-type ESCs compared to (**c**) ATRX KO and (**d**) DAXX KO ESCs. The bottom and the top of the boxes correspond to the 25th and 75th percentiles, and the internal band is the 50th percentile (median). The plot whiskers correspond to 1.5 interquartile range. Statistical significance is determined by Wilcoxon Mann Whitney test. **$p < 0.01$; ***$p < 0.001$. (**e**) Cell viability of wild-type and H3.3 KO ESCs treated with PDS for 5 days. Data are representative of two independent experiments. For **b** and **e**, mock-treated cells at day 0 were taken as 100% survival. Data represented as mean ± SD.

44

Teng et al., Supplementary Fig. 11
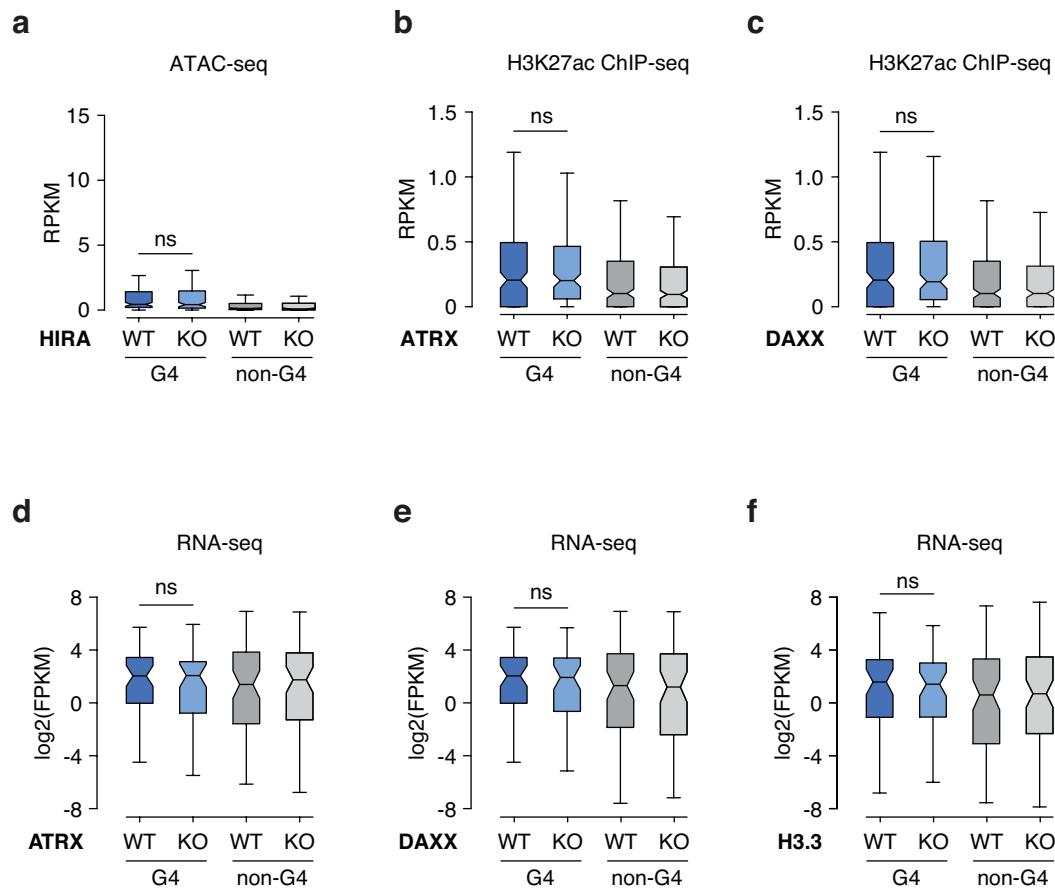
950
951

Supplementary Figure 11. ATRX/DAXX complex does not contribute uniformly to transcriptional activity of genes containing ATRX-enriched G4 regions. Related to Figure 5. (**a**) Box plots representing ATAC-seq read counts at ATRX-enriched G4 and non-G4 regions in wild-type ESCs compared to HIRA KO ESCs. (**b** and **c**) Box plots representing ChIP-seq read counts for H3K27ac[60] at ATRX-enriched G4 and non-G4 regions in wild-type ESCs compared to (**b**) ATRX KO and (**c**) DAXX KO cells. (**d**, **e**, and **f**) Box plots representing RNA-seq demonstrate that expression from ATRX-enriched G4 and non-G4 genes does not change in (**d**) ATRX KO (n=58), (**e**) DAXX KO (n=56), and (**f**) H3.3 KO ESCs (n=62), compared to wild-type ESCs[60]. Genic ATRX peak defined as within -3kb to +15kb from promoter (n=69). The bottom and the top of the boxes correspond to the 25th and 75th percentiles, and the internal band is the 50th percentile (median). The plot whiskers correspond to 1.5 interquartile range. Statistical significance is determined by Wilcoxon Mann Whitney test. ns, not significant.

964
965

Teng et al., Supplementary Fig. 12
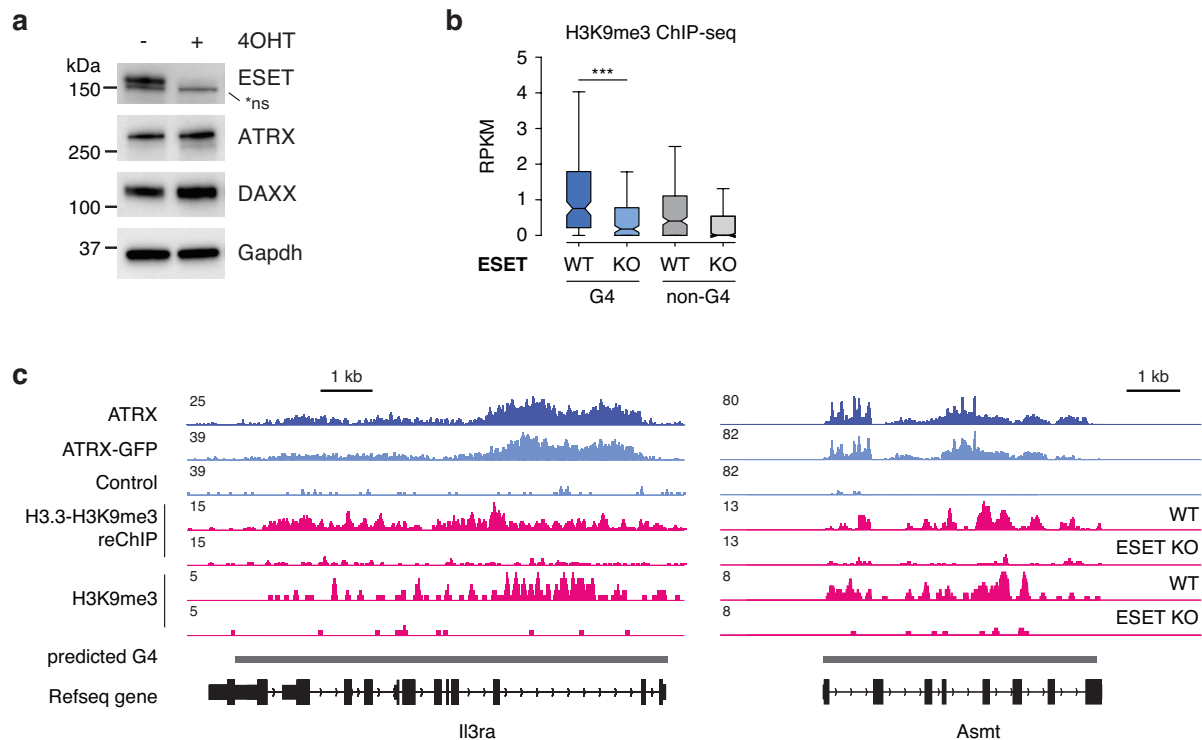
Supplementary Figure 12. ESET-mediated heterochromatin formation at ATRX-enriched G4 regions. Related to Figure 7. (**a**) Immunoblot from ESC whole-cell lysates showing the expression of ESET, ATRX, and DAXX in wild-type and ESET KO (+ 4OHT) ESCs. Gapdh as loading control. (**b**) Box plots representing ChIP-seq read counts for H3K9me3[69] at ATRX-enriched G4 and non-G4 regions in wild-type ESCs compared to ESET KO cells. The bottom and the top of the boxes correspond to the 25th and 75th percentiles, and the internal band is the 50th percentile (median). The plot whiskers correspond to 1.5 interquartile range. Statistical significance is determined by Wilcoxon Mann Whitney test. ***p < 0.001. (**c**) Genome browser representations of ATRX, ATRX-GFP, H3K9me3 ChIP-seq[69] and H3.3-H3K9me3 reChIP-seq[11] at predicted G4 regions in ESCs.

Teng et al., Supplementary Figure 13
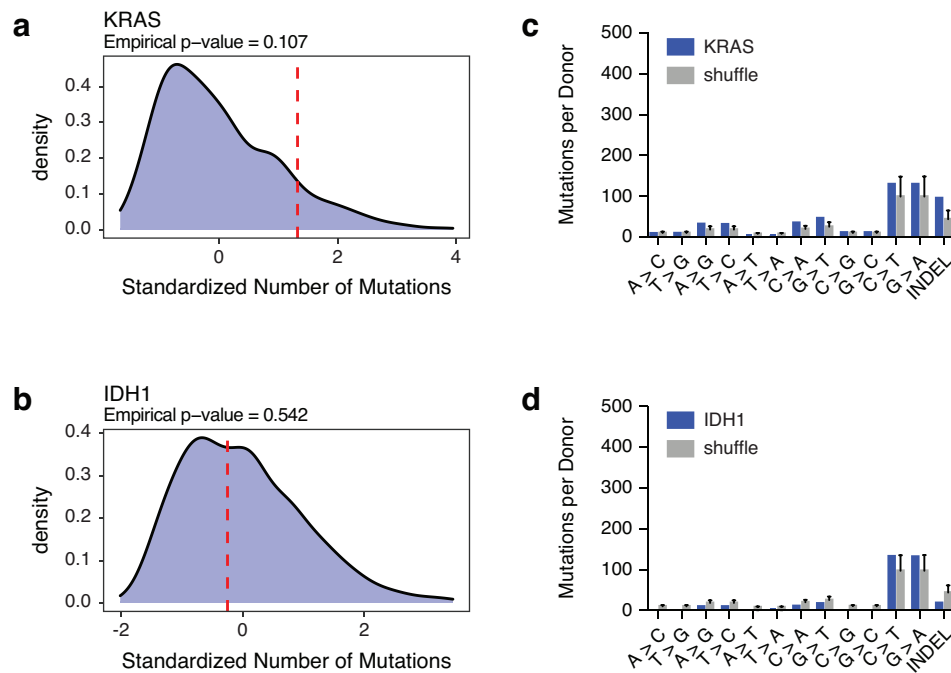
Supplementary Figure 13. Mutations at G4 regions are not correlated with KRAS or IDH1 mutations in human tumors. Related to Figure 8. (**a** and **b**) Red line shows standardized number of mutations at G4 regions in KRAS (**a**) or IDH1 (**b**) mutant tumors. Histograms show the mutation density at observed G4 regions in an iteratively and randomly selected patient cohort of the same size. (**c** and **d**) Analysis of single-nucleotide mutations and insertion-deletion (INDEL) mutations in the KRAS (**c**) or IDH1 (**d**) mutant tumors compared with the shuffled patient cohort described above.

Supplementary Tables

Supplementary Table 1. List of ATRX-bound G4 and ATRX-bound non-G4 regions.
Supplementary Table 2. List of ATRX-interacting proteins in HeLa WT and ATRX KO cells.
Supplementary Table 3. List of published genomic data sets analyzed in this study.
Supplementary Table 4. List of antibodies used in this study.
Supplementary Table 5. List of primers used in H3K9me3 ChIP.

Supplementary Methods

Antibodies. All antibodies used in this study are listed in Supplementary Table 4.

Immunoblot. Cell lysates were generated using digestion buffer (50 mM Tris-HCl, pH 7.6, 1 mM $CaCl_2$, 0.2% Triton X-100 and protease inhibitor cocktail (Roche)) with micrococcal nuclease for 5min at 37°C and denatured in SDS loading buffer. 1.5×10^4 cell lysates were run on NuPAGE 4-12% Bis-Tris gel at 180V. Then, the protein on the gel was transferred onto a

47

1004   PVDF membrane (Millipore). The membrane was incubated with antibodies in 5% milk/TBST
1005   (0.1% Tween 20) overnight at 4°C, washed, incubated with secondary antibody and developed
1006   on the ChemiDoc MP camera system (Bio-Rad). Blot is representative of three independent
1007   experiments.

1008   **Cell cycle profiling in ESCs.** Cell cycle phase analysis was performed using the Click-iT EdU
1009   Alexa Fluor 488 Flow Cytometry Assay kit (Thermo Fisher, C10425) according to the
1010   manufacturer's instructions. ESCs were grown overnight in a plate. Cells were incubated in
1011   medium with 2 mM thymidine for 14h, washed, and incubated in medium with 50 ngml^{-1}
1012   nocodazole for 7h[62]. After wash, mitotic cells were labeled with 10 µM EdU for 30 min prior to
1013   fixation. 1x 10^6 cells were trypsinized, washed with 1% BSA/PBS, fixed with Click-it fixative,
1014   permeabilized with Click-it saponin-based permeabilization and wash reagent and incubated
1015   with Click-it reaction (see kit manual). Cells were washed and resuspended in Click-it saponin-
1016   based permeabilization and wash reagent with 2 drops of SYTOX AADvanced^{TM} Ready Flow
1017   Reagent^{TM} (Thermo Fisher, R37173). Flow cytometry performed on a BD FACSCanto^{TM} II (BD
1018   Biosciences). SYTOX AADvanced dye fluorescence was excited by 488 nm laser light and
1019   detected in the far red range of the spectrum. Alexa Fluor 488 fluorescence was excited by 488
1020   nm laser light and detected in the green range of the spectrum. Cell cycle analysis was
1021   performed using FlowJo software.

1022   **Proximity ligation assay (PLA).** Cells were seeded on 8 µgml^{-1} fibronectin-coated round cover
1023   glass on a 12-well plate. Cell cycle synchronization was performed as previously described[62].
1024   Cells were incubated with 2 mM thymidine for 14h and following 50 ngml^{-1} nocodazole treatment
1025   for 7h. Mitotic cells were released and incubated with 2 µM EdU for 20min in prior to time point.
1026   Cells were fixed with 4% paraformaldehyde in PBS pH 7.4 for 10min at room temperature then
1027   permeabilized in 0.5% Triton X-100 in PBS for 15min.

1028   *For ATRX and MCM foci*, cells were fixed and permeabilized as described previously. Click-iT
1029   reaction and 488-azide were used for detection of EdU-labeling cells. PLA was the same as
1030   described previously but using ATRX (ab97508, 1:2500) and Mcm (1:2500) antibodies.

1031   **Chromatin immunoprecipitation.** Crosslinking ChIP was performed according to published
1032   methods[2,63]. Cells were fixed with 2 mM EGS (Thermo Fisher, 21565) for 45min at room
1033   temperature, 1% paraformaldehyde for 10min and quenched with 0.125 M glycine. Lysate was
1034   sonicated in lysis buffer (1% SDS, 10 mM EDTA, 50 mM Tris, pH 8 and protease inhibitor
1035   cocktail) to get 150-500 bp chromatin using a Covaris M220 Focused-ultrasonicator and were
1036   immunoprecipitated with ATRX antibody or GFP antibody bound to Dynabeads Protein G
1037   overnight at 4°C, with 5% kept as input DNA. Magnetic beads were washed, chromatin was
1038   eluted and ChIP DNA was purified by Qiagen PCR purification column. Native ChIP was
1039   performed according to published methods[11]. Cells were collected, washed, and lysed in
1040   digestion buffer (50 mM Tris, pH 7.4, 1mM CaCl$_2$, 0.2% Triton X-100, and protease inhibitor
1041   cocktail) with micrococcal nuclease for 5min at 37°C. Nuclei were sonicated briefly and dialyzed
1042   into RIPA buffer (10 mM Tris, pH 7.6, 1 mM EDTA, 0.1% SDS, 0.1% Na-deoxycholate, 1%
1043   Triton X-100) for 2h at 4°C. Soluble materials containing mono- to tri-nucleosomes were
1044   incubated with H3K9me3 antibody bound to Dynabeads Protein G overnight at 4°C, with 5%

1045   kept as input DNA. Magnetic beads were washed, chromatin was eluted and ChIP DNA was

1046   purified by Qiagen PCR purification column. The Spike-in chromatin (Active Motif, 53083) and

1047   Spike-in antibody were used in all ChIP experiments according to the manufacturer's

1048   instructions.

1049   **ChIP-qPCR.** qPCR was performed in triplicate using a LightCycler 480 Instrument II system and

1050   Power SYBR Green PCR master mix. ChIP DNA samples were diluted 1:100 in water, with 5 µl

1051   used per reaction. ChIP–qPCR signal is represented as percent input. All qPCR primer

1052   sequences used in this study are listed in Supplementary Table 5.

1053   **ATAC-seq.** ATAC-seq was performed as previously described[64]. Cells were collected, washed

1054   and lysed in lysis buffer (10 mM Tris, pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$, and 0.1%NP-40).

1055   Nuclei were collected and subjected to transposase reaction cocktail (25 µl 2X TD buffer, 2.5 µl

1056   transposase (Illumina) and 22.5 µl nuclease-free water) at 37°C for 30min. DNA was

1057   immediately collected using Qiagen MinElute kit and eluted in 10 mM Tris, pH 8. Eluted DNA

1058   was amplified using a KAPA non-hot-start PCR kit with Nextera PCR primer 1 and 2. Libraries

1059   were amplified no more than 11 cycles and purified using AMPure XP beads. The quality of

1060   libraries was determined using a D5000 ScreenTape on a 2200 TapeStation and a Qubit

1061   dsDNA HS Assay kit. Libraries were paired-end 33-base sequenced on the Illumina NextSeq

1062   500. Typical sequencing depth was at least 40 million reads per sample.

1063   **Analysis of ATAC-seq data.** Quality of the ATAC-seq data sets was assessed using the

1064   FastQC tool (v.0.11.2). The ATAC-seq reads were then aligned to the mouse reference genome

1065   (mm10) using BWA (v.0.7.5). For unique alignments, duplicate reads were filtered out. The

1066   resulting uniquely mapped reads were normalized to the same read depth across all samples

1067   and converted into bigWig files using BEDTools (v.2.29.0) and then converted to bigwig using

1068   bedGraphToBigWig utility of UCSC kent tools (v317) for visualization in Integrative Genomics

1069   Viewer (v.2.3).

1070   **CUT&Tag.** CUT&Tag was performed as previously described[65]. Cells were collected, washed

1071   and lysed in nuclear extraction buffer (20 mM HEPES-KOH, pH 7.9, 10 mM KCl, 0.1% Triton X-

1072   100, 20% glycerol, 0.5 mM spermidine, and proteinase inhibitor cocktail) on ice for 10min. While

1073   cells were lysed, concanavalin A coated magnetic beads (Epicypher, 21-1401) were activated in

1074   binding buffer (20 mM HEPES-KOH, pH 7.9, 10 mM KCl, 1 mM $CaCl_2$, and 1 mM $MnCl_2$). 5 x

1075 10^5 nuclei per sample were washed with PBS once, resuspended in 100 µl wash buffer (20 mM

1076   HEPES-KOH, pH 7.5, 150 mM NaCl, 0.5 mM spermidine, and proteinase inhibitor cocktail) and

1077   incubated with 10 µl activated beads for 10min at room temperature. Nuclei-bound beads were

1078   resuspended in 50 µl antibody buffer (2 mM EDTA and 0.1% BSA in wash buffer) containing 2

1079   µl BG4 antibody and incubated at 4°C overnight. Unbound antibodies were removed and beads

1080   were resuspended in 50 µl antibody buffer containing 1 µl anti-DYKDDDK antibody and

1081   incubated at room temperature for 2h. After removing unbound antibodies, beads were

1082   resuspended in 100 µl wash buffer containing 1 µl anti-rabbit secondary antibody and incubated

1083   at room temperature for 30min. Beads were washed in 1 ml wash buffer three times,

1084   resuspended in 100 µl wash-300 buffer (20 mM HEPES-KOH, pH 7.5, 300 mM NaCl, 0.5 mM

1085   spermidine, and proteinase inhibitor cocktail) containing 2.5 µl pAG-Tn5 (Epicypher, 15-1017)

49

1086   and incubated at room temperature for 1h. Beads were washed in 1 ml wash-300 buffer three
1087   times to remove unbound pAG-Tn5. Next, beads were resuspended in 300 µl tagmentation
1088   buffer (10 mM MgCl$_2$ in wash-300 buffer) and incubated at 37°C for 1h. To stop tagmentation,
1089   10 µl of 0.5 M EDTA, 3 µl of 10% SDS, and 2.5 µl of Proteinase K were added to 300 µl of
1090   sample, which was incubated at 55°C for 1h. To extract DNA, samples were transferred to
1091   Phase Lock Gel tube (QuantaBio, 2302820) and equal volume of Phenol:Chloroform:Isoamyl
1092   Alcohol was added into the tube and spin at 13000 rpm for 5min. Supernatant containing DNA
1093   was precipitated in 2.5 volumes of absolute alcohol and incubated at -20°C for 30min to
1094   overnight. DNA was pelleted after centrifuge and dissolved in 25 µl 0.1x TE. To amplify libraries,
1095   DNA was amplified using a KAPA non-hot-start PCR kit with Nextera PCR primer 1 and 2.
1096   Libraries were amplified for 12 cycles and purified using AMPure XP beads. The quality of
1097   libraries was determined using a D5000 ScreenTape on a 2200 TapeStation and a Qubit
1098   dsDNA HS Assay kit. Libraries were paired-end 33-base sequenced on the Illumina NextSeq
1099   500. Typical sequencing depth was at least 100 million reads per sample.

1100   **Analysis of CUT&Tag data.** CUT&Tag was analyzed following the methods available at
1101   https://yezhengstat.github.io/CUTTag_tutorial/ . Briefly, quality of CUT&Tag data sets was
1102   assessed using the FACTQC tool (v.0.11.2). CUT&Tag raw reads were trimmed using
1103   TrimGalore(v.0.6.4). The trimmed reads were aligned to the mouse reference genome (mm10)
1104   using BOWTIE2 (v.2.3.2). The optical duplicate reads were filtered using the MarkDuplicates
1105   tool of Picard (v.2.10.3). The resulting uniquely mapped reads were normalized to the same
1106   read depth across all samples. Reads were converted into bedgraph files using BEDTools
1107   (v.2.29.0) and the. Converted to bigwig using bedGraphToBigWig utility of UCSC kent tools
1108   (v317) for visualization in Integrative Genomics Viewer (v.2.3).

1109   **EdU-seq.** EdU-seq was performed as previously described[73]. Mitotic ESCs were released in
1110   medium and incubated in medium with 10 µM EdU for 30min before the cells were collected. At
1111   the indicated time, cells were washed and immediately fixed with cold 90% methanol overnight
1112   at -20°C. Cells were permeabilized with 0.2% Triton X-100 in PBS for 30min at room
1113   temperature. After washing, EdU was coupled to a cleavable biotin-azide (Azide-SS-biotin,
1114   BroadPharm, BP-22877) using the Click-iT reaction cocktail (Thermo Fisher, C10269). The DNA
1115   was purified by phenol/chloroform extraction in Phase Lock Gel, Light, 1.5-ml tube and ethanol
1116   precipitation for EdU-labeled DNA isolation. Total 15 µg DNA was sonicated to a size range of
1117   150-500 bp with a Covaris M220 Focused-ultrasonicator. EdU-labeled DNA fragments were
1118   pulled down by Dynabeads MyOne streptavidin C1 (Invitrogen) for 15min at room temperature,
1119   washed, and eluted in elution solution (10 mM Tris, pH 8) containing fresh 2% β-
1120   mercaptoethanol for 1h at room temperature. The eluted DNA was directly used for library
1121   preparation. Asynchronized EdU-labeled HeLa DNA fragments were generated as described
1122   previously for a spike-in normalization and used in EdU-seq. The EdU-seq libraries were
1123   prepared from 5 ng DNA following the Illumina TruSeq protocol. The quality of libraries was
1124   determined using a D5000 ScreenTape on a 2200 TapeStation and a Qubit dsDNA HS Assay
1125   kit. Libraries were paired-end 33-base sequenced on the Illumina NextSeq 500. Typical
1126   sequencing depth was at least 50 million reads per sample.

1127   **Analysis of EdU-seq data.** Quality of EdU-seq data sets was assessed using the FastQC tool
1128   (v.0.11.2). EdU-seq raw reads were aligned separately to the mouse reference genome (mm10)
1129   and the spike-in human reference genome (hg19) using BOWTIE2 (v.2.2.8). Only one alignment
1130   is reported for each read (either the single best alignment or, if more than one equivalent best
1131   alignment was found, one of those matches selected randomly). Duplicate reads were filtered
1132   using the MarkDuplicates tool of Picard (v.1.127). The resulting uniquely mapped reads were
1133   normalized to the same read depth across all samples. Reads were converted into bedgraph
1134   files using BEDTools (v.2.29.0) and then converted to bigwig using bedGraphToBigWig utility of
1135   UCSC kent tools (v317) for visualization in Integrative Genomics Viewer (v.2.3).

1136   **Analysis of SNS-seq data.** Quality of SNS-seq data sets was assessed using the FastQC tool
1137   (v.0.11.2). SNS-seq raw reads were aligned to the mouse reference genome (mm10) using
1138   BOWTIE2 (v.2.2.8). Only one alignment is reported for each read (either the single best
1139   alignment or, if more than one equivalent best alignment was found, one of those matches
1140   selected randomly). Duplicate reads were filtered using the MarkDuplicates tool of Picard
1141   (v.1.127). The resulting uniquely mapped reads were normalized to the same read depth across
1142   all samples. Reads were converted into bedgraph files using BEDTools (v.2.29.0) and then
1143   converted to bigwig using bedGraphToBigWig utility of UCSC kent tools (v317) for visualization
1144   in Integrative Genomics Viewer (v.2.3).

1145   **Analysis of RNA-seq data.** Quality of the RNA-seq raw reads was assessed using the FastQC
1146   tool (v.0.11.2). The reads were then aligned to the mouse reference genome (mm10) using the
1147   spliced read aligner TopHat version v.2.0.12, transcriptome assembly was carried out using
1148   Cufflinks v.2.2.1 with default parameters, filtered transcripts were merged into distinct
1149   nonoverlapping sets using Cuffmerge, and Cuffdiff was used to calculate the differential
1150   expression genes between the conditions.

1151

REFERENCES

1. Dyer, M. A., Qadeer, Z. A., Valle-Garcia, D. & Bernstein, E. ATRX and DAXX: Mechanisms and Mutations. *Cold Spring Harb. Perspect. Med.* **7**, (2017).

2. Law, M. J. *et al.* ATR-X syndrome protein targets tandem repeats and influences allele-specific expression in a size-dependent manner. *Cell* **143**, 367–378 (2010).

3. Sadic, D. *et al.* Atrx promotes heterochromatin formation at retrotransposons. *EMBO reports* vol. 16 836–850 (2015).

4. He, Q. *et al.* The Daxx/Atrx Complex Protects Tandem Repetitive Elements during DNA Hypomethylation by Promoting H3K9 Trimethylation. *Cell Stem Cell* **17**, 273–286 (2015).

5. Voon, H. P. J. *et al.* ATRX Plays a Key Role in Maintaining Silencing at Interstitial Heterochromatic Loci and Imprinted Genes. *Cell Reports* vol. 11 405–418 (2015).

6. Udugama, M. *et al.* Ribosomal DNA copy loss and repeat instability in ATRX-mutated cancers. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 4737–4742 (2018).

7. Sarma, K. *et al.* ATRX directs binding of PRC2 to Xist RNA and Polycomb targets. *Cell* **159**, 869–883 (2014).

8. Danussi, C. *et al.* Atrx inactivation drives disease-defining phenotypes in glioma cells of origin through global epigenomic remodeling. *Nat. Commun.* **9**, 1057 (2018).

9. Goldberg, A. D. *et al.* Distinct factors control histone variant H3.3 localization at specific genomic regions. *Cell* **140**, 678–691 (2010).

10. Navarro, C., Lyu, J., Katsori, A.-M., Caridha, R. & Elsässer, S. J. An embryonic stem cell-specific heterochromatin state promotes core histone exchange in the absence of DNA accessibility. *Nat. Commun.* **11**, 5095 (2020).

11. Elsässer, S. J., Noh, K.-M., Diaz, N., Allis, C. D. & Banaszynski, L. A. Histone H3.3 is required for endogenous retroviral element silencing in embryonic stem cells. *Nature* **522**, 240–244 (2015).

12. Li, F. *et al.* ATRX loss induces telomere dysfunction and necessitates induction of alternative lengthening of telomeres during human cell immortalization. *The EMBO Journal* vol. 38 (2019).

13. Wong, L. H. *et al.* ATRX interacts with H3.3 in maintaining telomere structural integrity in pluripotent embryonic stem cells. *Genome Research* vol. 20 351–360 (2010).

14. Clynes, D. *et al.* ATRX Dysfunction Induces Replication Defects in Primary Mouse Cells. *PLoS ONE* vol. 9 e92915 (2014).

15. Clynes, D. *et al.* Suppression of the alternative lengthening of telomere pathway by the chromatin remodelling factor ATRX. *Nature Communications* vol. 6 (2015).

16. Huh, M. S. *et al.* Stalled replication forks within heterochromatin require ATRX for protection. *Cell Death & Disease* vol. 7 e2220–e2220 (2016).

17. Juhász, S., Elbakry, A., Mathes, A. & Löbrich, M. ATRX Promotes DNA Repair Synthesis and Sister Chromatid Exchange during Homologous Recombination. *Mol. Cell* **71**, 11–24.e7 (2018).

18. Lovejoy, C. A., Takai, K., Huh, M. S., Picketts, D. J. & de Lange, T. ATRX affects the repair of telomeric DSBs by promoting cohesion and a DAXX-dependent activity. *PLoS Biol.* **18**, e3000594 (2020).

19. Varshney, D., Spiegel, J., Zyner, K., Tannahill, D. & Balasubramanian, S. The regulation and functions of DNA and RNA G-quadruplexes. *Nat. Rev. Mol. Cell Biol.* (2020) doi:10.1038/s41580-020-0236-x.

20. Rhodes, D. & Lipps, H. J. G-quadruplexes and their regulatory roles in biology. *Nucleic Acids Res.* **43**, 8627–8637 (2015).

21. Técher, H., Koundrioukoff, S., Nicolas, A. & Debatisse, M. The impact of replication stress on replication dynamics and DNA damage in vertebrate cells. *Nat. Rev. Genet.* **18**, 535–550 (2017).

22. van Wietmarschen, N. *et al.* BLM helicase suppresses recombination at G-quadruplex

52

1203   motifs in transcribed genes. *Nat. Commun.* **9**, 271 (2018).

1204   23.   Wu, W. *et al.* HERC2 Facilitates BLM and WRN Helicase Complex Interaction with RPA to
1205   Suppress G-Quadruplex DNA. *Cancer Res.* **78**, 6371–6385 (2018).

1206   24.   Pladevall-Morera, D. *et al.* Proteomic characterization of chromosomal common fragile site
1207   (CFS)-associated proteins uncovers ATRX as a regulator of CFS stability. *Nucleic Acids*
1208   *Res.* **47**, 8332 (2019).

1209   25.   Wang, Y. *et al.* G-quadruplex DNA drives genomic instability and represents a targetable
1210   molecular abnormality in ATRX-deficient malignant glioma. *Nat. Commun.* **10**, 943 (2019).

1211   26.   Watson, L. A. *et al.* Atrx deficiency induces telomere dysfunction, endocrine defects, and
1212   reduced life span. *J. Clin. Invest.* **123**, 2049–2063 (2013).

1213   27.   Schmid-Burgk, J. L., Höning, K., Ebert, T. S. & Hornung, V. CRISPaint allows modular
1214   base-specific gene tagging using a ligase-4-dependent mechanism. *Nat. Commun.* **7**,
1215   12338 (2016).

1216   28.   Navarro, C. & Elsasser, S. J. An embryonic stem cell-specific heterochromatin state allows
1217   core histone exchange in the absence of DNA accessibility. *bioRxiv* (2020).

1218   29.   Sen, D. & Gilbert, W. Formation of parallel four-stranded complexes by guanine-rich motifs
1219   in DNA and its implications for meiosis. *Nature* **334**, 364–366 (1988).

1220   30.   Sundquist, W. I. & Klug, A. Telomeric DNA dimerizes by formation of guanine tetrads
1221   between hairpin loops. *Nature* **342**, 825–829 (1989).

1222   31.   Huppert, J. L. & Balasubramanian, S. Prevalence of quadruplexes in the human genome.
1223   *Nucleic Acids Res.* **33**, 2908–2916 (2005).

1224   32.   Biffi, G., Tannahill, D., McCafferty, J. & Balasubramanian, S. Quantitative visualization of
1225   DNA G-quadruplex structures in human cells. *Nat. Chem.* **5**, 182–186 (2013).

1226   33.   Zyner, K. G. *et al.* Genetic interactions of G-quadruplexes in humans. *Elife* **8**, (2019).

1227   34.   Bochman, M. L. & Schwacha, A. The Mcm complex: unwinding the mechanism of a
1228   replicative helicase. *Microbiol. Mol. Biol. Rev.* **73**, 652–683 (2009).

1229   35.   Tan, B. C.-M., Liu, H., Lin, C.-L. & Lee, S.-C. Functional cooperation between FACT and
1230   MCM is coordinated with cell cycle and differential complex formation. *J. Biomed. Sci.* **17**,
1231   11 (2010).

1232   36.   Gibbons, R. J. *et al.* Mutations in transcriptional regulator ATRX establish the functional
1233   significance of a PHD-like domain. *Nat. Genet.* **17**, 146–148 (1997).

1234   37.   Schwartzentruber, J. *et al.* Driver mutations in histone H3.3 and chromatin remodelling
1235   genes in paediatric glioblastoma. *Nature* **482**, 226–231 (2012).

1236   38.   Hoelper, D., Huang, H., Jain, A. Y., Patel, D. J. & Lewis, P. W. Structural and mechanistic
1237   insights into ATRX-dependent and -independent functions of the histone chaperone DAXX.
1238   *Nat. Commun.* **8**, 1193 (2017).

1239   39.   Martire, S. & Banaszynski, L. A. The roles of histone variants in fine-tuning chromatin
1240   organization and function. *Nature Reviews Molecular Cell Biology* (2020)
1241   doi:10.1038/s41580-020-0262-8.

1242   40.   Huang, H. *et al.* A unique binding mode enables MCM2 to chaperone histones H3--H4 at
1243   replication forks. *Nat. Struct. Mol. Biol.* **22**, 618 (2015).

1244   41.   Ramachandran, S., Vogel, L., Strahl, B. D. & Dokholyan, N. V. Thermodynamic stability of
1245   histone H3 is a necessary but not sufficient driving force for its evolutionary conservation.
1246   *PLoS Comput. Biol.* **7**, e1001042 (2011).

1247   42.   Alabert, C. *et al.* Nascent chromatin capture proteomics determines chromatin dynamics
1248   during DNA replication and identifies unknown fork components. *Nat. Cell Biol.* **16**, 281–
1249   293 (2014).

1250   43.   Roy, S., Luzwick, J. W. & Schlacher, K. SIRF: Quantitative in situ analysis of protein
1251   interactions at DNA replication forks. *Journal of Cell Biology* vol. 217 1553–1553 (2018).

1252   44.   Prorok, P. *et al.* Involvement of G-quadruplex regions in mammalian replication origin
1253   activity. *Nat. Commun.* **10**, 3274 (2019).

1254 45. Mitson, M., Kelley, L. A., Sternberg, M. J. E., Higgs, D. R. & Gibbons, R. J. Functional
1255   significance of mutations in the Snf2 domain of ATRX. *Human Molecular Genetics* vol. 20
1256   2603–2610 (2011).
1257 46. Liang, J. *et al.* Global changes in chromatin accessibility and transcription following ATRX
1258   inactivation in human cancer cells. *FEBS Lett.* **594**, 67–78 (2020).
1259 47. Matsui, T. *et al.* Proviral silencing in embryonic stem cells requires the histone
1260   methyltransferase ESET. *Nature* **464**, 927–931 (2010).
1261 48. Guiblet, W. M. *et al.* Long-read sequencing technology indicates genome-wide effects of
1262   non-B DNA on polymerization speed and error rate. *Genome Res.* **28**, 1767–1778 (2018).
1263 49. Kamat, M. A., Bacolla, A., Cooper, D. N. & Chuzhanova, N. A Role for non-B DNA forming
1264   sequences in mediating microlesions causing human inherited disease. *Hum. Mutat.* **37**,
1265   65–73 (2016).
1266 50. Du, X. *et al.* Potential non-B DNA regions in the human genome are associated with higher
1267   rates of nucleotide mutation and expression variation. *Nucleic Acids Res.* **42**, 12367–12379
1268   (2014).
1269 51. Guiblet, W. M. *et al.* Non-B DNA: a major contributor to small- and large-scale variation in
1270   nucleotide substitution frequencies across the genome. *Nucleic Acids Res.* **49**, 1497–1516
1271   (2021).
1272 52. Koschmann, C. *et al.* ATRX loss promotes tumor growth and impairs nonhomologous end
1273   joining DNA repair in glioma. *Sci. Transl. Med.* **8**, 328ra28 (2016).
1274 53. Eustermann, S. *et al.* Combinatorial readout of histone H3 modifications specifies
1275   localization of ATRX to heterochromatin. *Nat. Struct. Mol. Biol.* **18**, 777–782 (2011).
1276 54. Iwase, S. *et al.* ATRX ADD domain links an atypical histone methylation recognition
1277   mechanism to human mental-retardation syndrome. *Nat. Struct. Mol. Biol.* **18**, 769–776
1278   (2011).
1279 55. Ha, M., Kraushaar, D. C. & Zhao, K. Genome-wide analysis of H3.3 dissociation reveals
1280   high nucleosome turnover at distal regulatory regions of embryonic stem cells. *Epigenetics*
1281   *& Chromatin* vol. 7 (2014).
1282 56. Deaton, A. M. *et al.* Enhancer regions show high histone H3.3 turnover that changes during
1283   differentiation. *Elife* **5**, (2016).
1284 57. Schlesinger, S. *et al.* A hyperdynamic H3.3 nucleosome marks promoter regions in
1285   pluripotent embryonic stem cells. *Nucleic Acids Research* vol. 45 12181–12194 (2017).
1286 58. Ahmad, K. & Henikoff, S. The Histone Variant H3.3 Marks Active Chromatin by Replication-
1287   Independent Nucleosome Assembly. *Molecular Cell* vol. 9 1191–1200 (2002).
1288 59. Ray-Gallet, D. *et al.* HIRA is critical for a nucleosome assembly pathway independent of
1289   DNA synthesis. *Mol. Cell* **9**, 1091–1100 (2002).
1290 60. Martire, S. *et al.* Phosphorylation of histone H3.3 at serine 31 promotes p300 activity and
1291   enhancer acetylation. *Nat. Genet.* **51**, 941–946 (2019).
1292 61. Banaszynski, L. A. *et al.* Hira-dependent histone H3.3 deposition facilitates PRC2
1293   recruitment at developmental loci in ES cells. *Cell* **155**, 107–120 (2013).
1294 62. Ballabeni, A. *et al.* Cell cycle adaptations of embryonic stem cells. *Proc. Natl. Acad. Sci. U.*
1295   *S. A.* **108**, 19252–19257 (2011).
1296 63. Truch, J., Telenius, J., Higgs, D. R. & Gibbons, R. J. How to Tackle Challenging ChIP-Seq,
1297   with Long-Range Cross-Linking, Using ATRX as an Example. *Methods Mol. Biol.* **1832**,
1298   105–130 (2018).
1299 64. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition
1300   of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-
1301   binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).
1302 65. Kaya-Okur, H. S. *et al.* CUT&Tag for efficient epigenomic profiling of small samples and
1303   single cells. *Nat. Commun.* **10**, 1930 (2019).
1304 66. Henikoff, S., Henikoff, J. G., Kaya-Okur, H. S. & Ahmad, K. Efficient chromatin accessibility

1305   mapping in situ by nucleosome-tethered tagmentation. *Elife* **9**, e63274 (2020).

1306   67. Marsico, G. *et al.* Whole genome experimental maps of DNA G-quadruplexes in multiple
1307   species. *Nucleic Acids Res.* **47**, 3862–3874 (2019).

1308   68. Zhang, J. *et al.* The International Cancer Genome Consortium Data Portal. *Nat. Biotechnol.*
1309   **37**, 367–369 (2019).

1310   69. Wu, K. *et al.* SETDB1-Mediated Cell Fate Transition between 2C-Like and Pluripotent
1311   States. *Cell Rep.* **30**, 25–36.e6 (2020).

1312   70. Elsässer, S. J., Noh, K.-M., Diaz, N., Allis, C. D. & Banaszynski, L. A. Elsässer et al. reply.
1313   *Nature* **548**, E7–E9 (2017).

1314   71. Prorok, P. *et al.* Involvement of G-quadruplex regions in mammalian replication origin
1315   activity. *Nature Communications* vol. 10 (2019).

1316   72. ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium. Pan-cancer analysis of
1317   whole genomes. *Nature* **578**, 82–93 (2020).

1318   73. Macheret, M. & Halazonetis, T. D. Monitoring early S-phase origin firing and replication fork
1319   movement by sequencing nascent DNA from synchronized cells. *Nat. Protoc.* **14**, 51–67
1320   (2019).

1321