

Cohesin-independent STAG proteins interact with RNA and localise to R-loops to promote complex loading.

Hayley Porter¹⁺, Yang Li¹⁺, Maria Victoria Neguembor², Manuel Beltran³, Wazeer Varsally¹, Laura Martin², Manuel Tavares Cornejo³, Dubravka Pezic¹, Amandeep Bhamra⁴, Silvia Surinova⁴, Richard G. Jenner³, Maria Pia Cosma^{2,5,6}, Suzana Hadjur^{1*}

1 Research Department of Cancer Biology, Cancer Institute, University College London, 72 Huntley Street, London, United Kingdom

2 Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, 08003 Barcelona, Spain.

3 Regulatory Genomics Group, Cancer Institute, University College London, London WC1E 6BT, United Kingdom

4 Proteomics Research Translational Technology Platform, Cancer Institute, University College London, 72 Huntley Street, London, United Kingdom

5 Universitat Pompeu Fabra (UPF), Dr Aiguader 88, 08003 Barcelona, Spain

6 Institució Catalana de Recerca i Estudis Avançats (ICREA), Pg. Lluís Companys 23, 08010 Barcelona, Spain.

+ These authors contributed equally

* Correspondence: s.hadjur@ucl.ac.uk

1 **ABSTRACT**

2

3 Most studies of cohesin function consider the Stromalin Antigen (STAG/SA) proteins
4 as core complex members given their ubiquitous interaction with the cohesin ring.
5 Here, we provide functional data to support the notion that the SA subunit is not a
6 mere passenger in this structure, but instead plays a key role in the localization of
7 cohesin to diverse biological processes and promotes loading of the complex at these
8 sites. We show that in cells acutely depleted for RAD21, SA proteins remain bound to
9 chromatin, cluster in 3D and interact with CTCF, as well as with a wide range of RNA
10 binding proteins involved in multiple RNA processing mechanisms. Accordingly, SA
11 proteins interact with RNA and are localised to R-loops where they contribute to R-
12 loop regulation. Our results place SA1 within R-loop domains upstream of the cohesin
13 complex and reveal a role for SA1 in cohesin loading which is independent of NIPBL,
14 the canonical cohesin loader. We propose that SA1 takes advantage of structural R-
15 loop platforms to link cohesin loading and chromatin structure with diverse functions.
16 Since SA proteins are pan-cancer targets, and R-loops play an increasingly prevalent
17 role in cancer biology, our results have important implications for the mechanistic
18 understanding of SA proteins in cancer and disease.

19

20

21

22

23 **KEY WORDS**

24 Cohesin, STAG proteins, R-loops, genome organization.

25 INTRODUCTION

26 Cohesin complexes are master regulators of chromosome structure in interphase and
27 mitosis. Accordingly, mutations of cohesin subunits lead to changes in cellular identity,
28 both during development and in cancer¹⁻³. A prevailing model is that cohesin
29 contributes to cell identity changes in large part by dynamically regulating 3D genome
30 organization and mediating communication between distal regulatory elements⁴⁻¹⁰.
31 Molecular insight into how and when cohesin subunits become associated with
32 chromatin and contribute to this function *in vivo* in human cells is still lacking under
33 this model.

34 Most studies of cohesin function consider the Stromalin Antigen (STAG/SA)
35 proteins as core complex members given their ubiquitous interaction with the tripartite
36 cohesin ring (composed of SMC1, SMC3 and SCC1/RAD21). Rarely is the SA subunit
37 considered for its roles independent of the cohesin ring, even though it is the subunit
38 most commonly mutated across a wide spectrum of cancers^{1,11,12}.

39 SA proteins contribute to cohesin's association with DNA^{13,14}. The yeast SA
40 orthologue is critical for efficient association of cohesin with DNA and its ATPase
41 activation^{13,14}. Separating interactions into SA-loader and cohesin ring-loader sub-
42 complexes still impairs cohesin loading, indicating that SA functions as more than just
43 a bridge protein¹⁴. Crystallisation studies reveal a striking similarity of SA with NIPBL
44 (the canonical cohesin loader¹⁵), in that both are highly bent, HEAT-repeat containing
45 proteins^{16,17}. Of note, NIPBL and SA interact together and wrap around both the
46 cohesin ring and DNA to position and entrap DNA¹⁸⁻²⁰, implying a potential role for SA
47 in the initial recruitment of cohesin to DNA alongside NIPBL. Further, SA proteins
48 bridge the interaction between cohesin and CTCF^{18,21,22}, and also bridge interactions
49 with specific nucleic acid structures *in vitro*^{23,24}.

50 Mammalian cells express multiple SA paralogs. SA1 binds to AT-rich telomeric
51 sequences^{23,24} and SA2 displays sequence-independent affinity for particular DNA
52 structures commonly found at sites of repair, recombination, and replication²⁵.
53 Consistent with this, results in yeast implicate non-canonical DNA structures in
54 cohesin loading in S-phase. *In vitro* experiments show that cohesin captures the
55 second strand of DNA via a single-strand intermediate²⁶, and chromatid cohesion is
56 impaired by de-stabilisation of single-strand DNA intermediates during replication²⁷.

57 Together, these suggest that SA proteins and DNA structures may play a regulatory
58 role in guiding or stabilising cohesin localisations.

59 During transcription, the elongating nascent RNA can hybridise to the template
60 strand of the DNA and form an R-loop, which is an intermediate RNA:DNA hybrid
61 conformation with a displaced single strand of DNA²⁸. A multitude of processes have
62 been linked to R-loop stability and metabolism. For example, co-transcriptional RNA
63 processing, splicing, and messenger ribonucleoprotein (mRNP) assembly counteract
64 R-loop formation^{29,30}. R-loop structures have also been shown to regulate
65 transcription of mRNA by recruitment of transcription factors, displacement of
66 nucleosomes, and preservation of open chromatin^{31,32}. Hence, like at the replication
67 fork, sites of active transcription accumulate non-canonical nucleic acid structures.

68 We set out to investigate the nature of SA proteins and cohesin loading to DNA.
69 We discovered independent functions of the SA proteins, providing critical insight into
70 the importance they play in their own right to direct cohesin's localization and loading
71 to chromatin. In cells acutely depleted of RAD21, SA proteins remain associated with
72 chromatin and CTCF where they are enriched at chromatin sites clustered in 3D.
73 Moreover, we identify numerous, diverse cohesin-independent SA1 interactors
74 involved in RNA processing, ribosome biogenesis, and translation. Consistent with
75 this, SA1 and SA2 interact with RNA and non-canonical nucleic acid structures in the
76 form of R-loops where SA1 suppresses R-loop formation. Importantly, SA proteins are
77 required for loading of cohesin to chromatin in cells deficient for NIPBL. Our results
78 highlight a central role for SA proteins in cohesin biology and the cohesin-independent
79 interaction of SA proteins with RNA processing factors opens up a new understanding
80 of how SA dysregulation can impact disease development that moves us beyond the
81 control of chromatin topology for gene expression regulation.

82 RESULTS

83

84 SA interacts with CTCF on chromatin in the absence of the cohesin trimer.

85 To determine how CTCF and cohesin assemble on chromatin, we used previously
86 described³³ human HCT116 cells engineered to carry a miniAID tag (mAID) fused to
87 monomeric Clover (mClover) at the endogenous RAD21 locus and *OsTIR1* under the
88 control of CMV (herein RAD21^{mAC}). RAD21^{mAC} cells were cultured in control EtOH
89 conditions (EtOH) or in the presence of auxin (IAA) to induce rapid RAD21 degradation
90 (Fig. S1a). Immunofluorescence (IF) was used to monitor the levels of mClover, SA1,
91 SA2 and CTCF (Fig.1a, b, S1b). As reported³³, acute IAA treatment dramatically
92 reduced mClover levels compared to control cells (mean fluorescence intensity (MFI)
93 reduction of 82%, $p=2.9E-239$). SA proteins and CTCF were also all significantly
94 reduced, however the extent of the change was notably different. We observed a small
95 but significant reduction in CTCF signal upon IAA treatment (mean reduction of 16%,
96 $p=1.6E-27$). This was similar to the mean SA1 signal which was reduced by 22%
97 compared to EtOH control ($p=3.5E-21$). However, SA2 levels mirrored more closely
98 the effect on mClover, being reduced by 63% ($p=1.9E-186$), but not completely lost
99 (Fig.1b). The retention of SA proteins despite the degradation of RAD21 was
100 surprising given the fact that they are considered to be part of a stable biochemical
101 complex.

102 We sought to validate these observations using an orthogonal technique and to
103 establish whether the residual SA proteins retained the capacity to directly interact
104 with CTCF. We prepared chromatin extracts from RAD21^{mAC} cells treated with EtOH
105 or IAA and performed chromatin co-immunoprecipitation (coIP) to probe the
106 interactions between SA proteins, RAD21 and CTCF. Both SA1 and SA2 interacted
107 with RAD21 and CTCF in control cells as expected^{34,35}, with notable differences in
108 their preferred interactions (Fig.1c). SA2 more strongly enriched RAD21 while the
109 SA1-CTCF interaction was significantly stronger than SA2-CTCF (Fig.1c), in line with
110 previous results³⁶. Upon RAD21 degradation, we again observed a stronger effect on
111 chromatin-bound SA2 levels compared to SA1, suggesting that SA2 is more sensitive
112 to cohesin loss than SA1. Residual SA proteins retained their ability to interact with
113 CTCF in the absence of RAD21 and additionally, the interaction between SA1 and

114 CTCF was further enhanced (Fig.1c). Reciprocal colPs with CTCF confirmed the
115 CTCF-SA interactions in RAD21-depleted cells, including the differences between
116 SA1 and SA2 (Fig.1d). We validated these results in a second cell line and using
117 siRNA-mediated knockdown (KD) of SMC3 (Fig. S1c). We also confirmed that
118 approximately 20% of SA1 and SA2 were bound to chromatin without RAD21 even in
119 unperturbed RAD21^{mAC} HCT116 cells (Fig S1d), reminiscent of reports of SA1 without
120 cohesin at telomeres²³.

121 Next, we performed two-color Stochastic Optical Reconstruction Microscopy
122 (STORM) to assess the nuclear distribution and co-localization of SA1, SA2 and CTCF
123 with nanometric resolution in control and RAD21-degraded cells (Fig. 1e, S1e). In
124 control RAD21^{mAC} cells, we observed clustering of CTCF, SA1 and SA2 localizations
125 as quantified by cluster analysis and nearest neighbor distance (NND) analysis of
126 protein clusters³⁷. SA1 and CTCF exhibited higher densities compared to SA2, with
127 shorter distances between clusters (mean NND of 68.9, 65.3 and 78.9nm for CTCF,
128 SA1 and SA2, respectively) (Fig 1f, g). Furthermore, we analyzed the relative
129 distribution of SA clusters to CTCF clusters by assessing the NND distribution between
130 SA1 and CTCF, and SA2 and CTCF. Both SA1 and SA2 exhibited significant co-
131 localization with CTCF at short distances in RAD21^{mAC} cells (Fig. 1h, S1f, g).

132 Upon IAA treatment, we observed a decreased density of detected SA1, SA2
133 and CTCF in two analyzed clones (Fig.1e, f, S1e), suggesting that RAD21 degradation
134 affects the stability of SA proteins and CTCF. As observed by conventional confocal
135 microscopy (Fig. 1b), SA2 localizations were more affected than SA1 (mean density
136 reduction in SA1, 32% and SA2, 42%). Accordingly, SA1, SA2 and CTCF clusters
137 were more sparsely distributed across the nucleus upon RAD21 degradation (mean
138 NND 81.1, 109.7 and 117.5 nm, respectively) with SA2 significantly more affected than
139 SA1 (mean NND increase 24.3%, 39% respectively) (Fig.1g). SA proteins and CTCF
140 remained co-localised upon RAD21 degradation as compared to both the control cells
141 and to a simulation of randomly-distributed protein clusters at the same density (Fig.
142 1h, S1f,g). Interestingly, while the probability of SA1 at CTCF is only modestly
143 affected, SA2 at CTCF is more affected in IAA treated cells (Fig. 1h), in line with our
144 previous observations. Together, our results confirm the maintained interaction and

145 spatial co-localization patterns of SA proteins with CTCF and reveal a difference in SA
146 paralog stability in the absence of the core cohesin trimer.

147

148 **Cohesin-independent SA proteins are localised at clustered regions in 3D.**

149 Previous analyses of the contribution of SA proteins to genome organization ^{7,9} were
150 performed in cells containing cohesin rings, possibly obscuring a functional role for SA
151 proteins themselves in genome organization. To determine if cohesin-independent SA
152 proteins may function at unique locations in the genome, we investigated whether the
153 residual SA-CTCF complexes (herein, SA-CTCF^{ΔCoh}) in IAA-treated RAD21^{mAC} cells
154 occupied the same chromatin locations as in control cells. Using chromatin
155 immunoprecipitation followed by sequencing (ChIP-seq), we determined the binding
156 profiles of CTCF, SA1, SA2, RAD21 and SMC3 in RAD21^{mAC} cells treated with EtOH
157 or IAA. Pairwise comparisons of CTCF ChIP-seq with RAD21 or SA in control
158 RAD21^{mAC} cells revealed the expected overlap in binding sites (Fig.1i, S1h). In
159 contrast, both global and CTCF-overlapping RAD21 and SMC3 ChIP-seq signals were
160 dramatically lost in IAA-treated cells (Fig.1i, S1h). In agreement with our microscopy
161 and biochemistry results, we detected residual SA1 and SA2 binding sites in IAA-
162 treated cells which retained a substantial overlap with CTCF (Fig.1i, S1h).
163 Furthermore, the sites co-occupied by CTCF and SA proteins in RAD21-depleted cells
164 (SA-CTCF^{ΔCoh}) were also bound in control conditions, were enriched at TSSs and
165 were characterised by active chromatin marks including phospho-S5 POLR2A,
166 H3K4me3 and H3K27ac (Fig.S1h). Thus, CTCF and SA maintain occupancy at their
167 canonical binding sites in the absence of RAD21 (Fig 1i), suggesting that SA
168 interaction with CTCF in the absence of the cohesin ring is a step in normal cohesin
169 activity.

170 While depletion of cohesin results in a dramatic loss of Topologically
171 Associated Domain (TAD) structure ⁸, the frequency of long-range inter-TAD, intra-
172 compartment contacts (LRC) is increased ^{5,8}, and enriched for CTCF ⁵ or active
173 elements ⁸. To determine whether residual, chromatin-bound SA could be associated
174 with LRCs in the absence of RAD21, we re-analysed Hi-C data from control and IAA-
175 treated RAD21^{mAC} cells ⁸. We quantified all contacts within two different scales of
176 genome organization; local TAD topology (100k-1Mb) and clustered LRCs (1-5Mb)

177 (Fig.1j). As previously shown ⁸, local TAD contacts are lost and clustered LRCs are
178 enriched in IAA conditions (Fig.1j). We probed the Hi-C datasets for contacts
179 containing the residual SA-CTCF^{ΔCoh} binding sites and observed a further enrichment
180 in IAA conditions (Fig.1j, bottom right), indicating that SA-CTCF^{ΔCoh} are enriched at
181 the clustered LRCs formed when cells are depleted of cohesin and thus implicating
182 them in 3D structural configurations. Our results suggest that cohesin-independent
183 SA, either with CTCF or alone, may itself contribute to higher-order arrangement of
184 active chromatin and regulatory features in 3D space.

185

186 **SA interacts with diverse ‘CES-binding proteins’ in cohesin-depleted cells.**

187 SA proteins contain a highly conserved domain known as the ‘stromalin conservative
188 domain’ ^{14,38}, or the ‘conserved essential surface’ (CES). Structural analysis of CTCF-
189 SA2-RAD21 has shown that a F/YXF motif in the N-terminus of CTCF engages with a
190 composite binding surface containing amino acids from RAD21 and the CES of SA2,
191 forming a tripartite interaction patch ¹⁸. Further, the authors identified a similar F/YXF
192 motif in other cohesin regulatory proteins, predicting interactions between the SA-
193 RAD21 binding surface and additional chromatin proteins. To investigate whether SA
194 proteins could associate with F/YXF-motif containing proteins beyond CTCF in the
195 absence of RAD21 *in vivo*, we performed chromatin colP with SA1 and SA2 in EtOH
196 and IAA and probed for interaction with CTCF as before, and three additional F/YXF-
197 motif containing proteins, CHD6, MCM3 and HNRNPUL2 (Fig. 2a, S2a). All of the
198 proteins interacted with SA1 in RAD21-control cells and their interaction with SA1 was
199 enriched upon RAD21-degradation, indicating that the association was not dependent
200 on the amino acids contributed by RAD21 (Fig. 2a, S2b). Interestingly, despite SA2
201 also containing the conserved CES domain, the F/YXF-motif proteins showed little
202 interaction with SA2 (Fig. 2a, S2b), pointing to the presence of additional features in
203 SA1 that stabilise its interaction with F/YXF containing proteins *in vivo*. Together, our
204 results confirm that SA, in particular SA1, can interact with proteins beyond just CTCF
205 and reveal that RAD21 is not required for the interaction between SA and F/YXF
206 proteins *in vivo*. These results prompted us to re-evaluate the role of SA in cohesin
207 activity and consider possible novel functions for these proteins.

208

209 **SA1 interacts with a diverse group of proteins in the absence of cohesin.**

210 To delineate novel protein binding partners and putative biological functions of
211 cohesin-independent SA1, we optimised our chromatin-bound, endogenous SA1 co-IP
212 protocol to be compatible with mass-spectrometry (IP-MS) and used this to
213 comprehensively characterize the SA1 protein–protein interaction (PPI) network in
214 control and RAD21-degraded RAD21^{mAC} cells. Three biological replicates were
215 prepared from RAD21^{mAC} cells that were either untreated (UT) or treated with IAA (IAA)
216 and processed for IP with both SA1 and IgG antibodies. In parallel, RAD21^{mAC} cells
217 were also treated with scrambled siRNAs or with siRNA to SA1 to confirm the
218 specificity of putative interactors. Immunoprecipitated proteins were identified by liquid
219 chromatography tandem mass spectroscopy (LC-MS/MS). SA1 peptides were
220 robustly detected in all UT and siCON samples and never detected in IgG controls,
221 validating the specificity of the antibody. We used a pairwise analysis of IAA vs UT
222 SA1 samples to generate a fold-change value for each putative interactor. These
223 candidates were changed by at least 1.5-fold compared to UT controls, and sensitive
224 to siSA1, yielding 136 high-confidence interactors whose abundance was significantly
225 altered with RAD21 loss (SA1^{ΔCoh}) (Fig. 2b, Table S1). As expected, core cohesin
226 subunits SMC1A and SMC3 were strongly depleted while no peptides were detected
227 for RAD21 (Fig. 2b). SA1 itself was significantly depleted compared to control cells,
228 as were other cohesin regulators, known to directly interact with SA1, such as PDS5B
229³⁹. In line with the enrichment we observed for the CES-binding proteins in IAA-
230 conditions (Fig. 1c, 2a), the vast majority of the SA1^{ΔCoh} interactors were enriched for
231 binding with SA1 in IAA conditions (117 of 136), and represent the cohesin-
232 independent SA1^{ΔCoh} interactors (Fig. 2b).

233 We used STRING to analyse associations in the SA1^{ΔCoh}-interactome and to
234 identify enriched biological processes and molecular functions. We calculated
235 enrichment relative to either the whole genome or to the SA1 interactome, which was
236 determined in parallel from control cells (Fig. S2c), revealing similar enrichment terms
237 to previously published cohesin interactomes^{40,41}, and validating our approach.
238 Compared to the whole genome background, processes enriched in the presence of
239 cohesin were still enriched in the SA1^{ΔCoh} PPI network and include a variety of
240 functionally diverse cellular processes such as chromosome organization,

241 transcription, RNA processing, ribosome biogenesis, and translation (Fig. 2c, d).
242 Within this group there are chromatin remodeling proteins (INO80 and SMARCAL1)
243 and several transcriptional and epigenetic regulators such as JARID2 and TAF15.
244 Similar to our ChIP and colP results, this suggests that SA1 maintains interaction with
245 proteins that localize with it in the presence of cohesin, albeit at different abundances.

246 RNA processing was the most enriched category in the SA1^{ΔCoh} PPI network
247 (FDR=3.62x10⁻³⁹) and included proteins involved in RNA modification (YTHDC1,
248 ADAR1, FTSJ3), mRNA stabilization and export (SYNCRIP, FMR1), and RNA splicing
249 regulators (SRSF1, SON). We also found a significant enrichment for DNA and RNA
250 helicases (FDR=3.54x10⁻⁰⁸) as well as RNA binding proteins (FDR=9.11x10⁻¹¹) within
251 which were many HNRNP family members (HNRNPU, aka SAF-A). We also found a
252 highly significant enrichment of proteins associated with ribosome biogenesis
253 (FDR=2.20x10⁻³⁰) including both large and small subunit components; rRNA
254 processing factors and components of the snoRNA pathway (FDR=4.39x10⁻⁰⁵).
255 Finally, translation was significantly enriched as a biological process (p=1.64x10⁻⁰⁶),
256 with several cytoplasmic translation regulators identified as SA1^{ΔCoh} interactors
257 (DHX29, GCN1L1) (Fig. 2c, d, S2d, e). Among these is ESYT2 which is primarily found
258 in the cytoplasm and contains a F/YXF-motif (Fig. 2c,d). We validated 8 of the highest-
259 ranking proteins within the enriched functional categories described above in EtOH
260 and IAA-treated RAD21^{mAC} cells (Fig. 2d). Importantly, the enrichment of these
261 proteins with SA1 in the IAA condition suggests that SA may have a role in these
262 processes independently of the core cohesin complex.

263 Comparison of the SA1^{ΔCoh} interactome with the SA1 interactome revealed that
264 the proteins involved in RNA processing (FDR= 0.0298), ribosome biogenesis
265 (0.0197), ribonucleoprotein complex biogenesis (0.0298) and rRNA processing
266 (0.0409) were enriched with SA1 following IAA treatment compared to SA1 in the
267 presence of RAD21 (Fig 2c, dotted lines). Overall, our results show that SA1^{ΔCoh} PPIs
268 contain not only transcriptional and epigenetic regulators, but are in fact predominantly
269 enriched for proteins with roles in RNA processing and modification, ribosome
270 biogenesis and translation pathways. Thus, SA1 is involved in several biological

271 processes and may facilitate an aspect of cohesin regulation at a variety of functionally
272 distinct locations.

273 **SA proteins bind RNA independently of cohesin.**

274 Since RNA binding and RNA processing were among the most enriched categories in
275 the SA1^{ΔCoh} PPI network, we hypothesized that SA proteins may also bind RNA. We
276 performed SA-crosslinking and immunoprecipitation (CLIP) in untreated RAD21^{mAC}
277 cells and found that both SA1 and SA2 directly bound RNA (Fig. 3a, b). This was
278 evidenced by detection of RNPs of the expected molecular weights, with a smear of
279 trimmed RNA, which was stronger in the +UV and +PNK conditions, increased as the
280 RNaseI concentration was reduced, and which was lost after siRNA-mediated SA KD
281 (Fig. S3a-c). We repeated the experiment in EtOH- and IAA-treated RAD21^{mAC} cells
282 to determine if the SA subunits can bind RNA in the absence of cohesin. As before,
283 RAD21 depletion reduced SA1 and SA2 (Fig. S3d) and the amount of RNA crosslinked
284 remained proportional to the amount of residual SA1 and SA2 protein (Fig. 3c, S3e),
285 demonstrating that cohesin is not required for the interaction of these proteins with
286 RNA in cells. Thus, cohesin-independent SA proteins interact with a wide array of RNA
287 binding proteins (RBPs) as well as with RNA itself.

288

289 **SA proteins localise to endogenous R-loops in the absence of cohesin.**

290 Proteins involved in RNA processing, such as splicing, modification and export, act as
291 regulators of R-loops⁴². Furthermore, R-loops accumulate at sites of multiple
292 biological processes including transcription, DNA replication and DNA repair⁴². As
293 many of these processes were enriched in the SA1 interactome, we reasoned that the
294 diversity of biological processes represented in the SA1^{ΔCoh} PPI network may be
295 reflective of a role for SA proteins in R-loop biology.

296 We performed a number of experiments to investigate the localization of SA
297 proteins at endogenous R-loops. First, we found a correlation between global SA and
298 R-loop levels. We depleted endogenous R-loops by overexpressing ppyCAG-
299 RNaseH-V5 in HCT116 cells. IF using the R-loop specific antibody S9.6 revealed that
300 nuclear S9.6 levels were significantly reduced in cells which expressed V5 (38% of
301 controls, p=0.04) and that mean SA1 signal was significantly reduced by 29% in the
302 same cells (Fig. 3d). Furthermore, RAD21^{mAC} cells treated with scramble control

303 siRNAs or Smartpool (SP) siRNAs to AQR (a known suppressor of R-loops⁴³), SA1
304 or SA2 revealed that S9.6 IF signal was significantly increased in siAQR and siSA1
305 but not siSA2 cells compared to the siScr control (mean S9.6 signal increased by 28%,
306 $p=0.0004$; 32%, $p=3.90E-8$; reduced by 10%, $p=0.17$, respectively) (Fig. S3f).
307 Although S9.6 signal was reduced by IF in RAD21^{mAC} cells treated with IAA, this did
308 not represent a significant change using this method (Fig. S3g).

309 We also performed STORM imaging on EtOH and IAA-treated RAD21^{mAC} cells
310 to assess the nuclear distribution of SA1 in the context of R-loops with and without
311 RAD21. We measured the ratio of the SA1 signal inside and outside of the S9.6 signal
312 mask. A ratio of 1 indicates a random distribution of SA1 with respect to S9.6 domains
313 while a ratio above 1 reflects enrichment within S9.6 domains. In EtOH conditions,
314 we did not detect enrichment of SA1 localizations, in fact SA1 was modestly depleted
315 (mean ratio 0.93). However, upon IAA treatment, we observed a significant enrichment
316 of SA1 localizations within S9.6 domains (mean ratio 1.24, $p<0.0001$) (Fig. 3e),
317 strongly suggesting that SA1 proteins are localized within R-loop domains
318 independently of cohesin.

319 In addition, we returned to our IP-MS experiment to analyse enrichment of R-
320 loop-associated proteins in our SA1^{ΔCoh} interactome. We overlapped the proteins
321 identified in two independent IP-MS experiments for R-loop interactors^{44,45} to create
322 a high-confidence 'R-loop interactome' and then used a hypergeometric distribution to
323 determine the significance of this category in the SA1^{ΔCoh} interactome (Methods). Both
324 the custom R-loop interactome as well as proteins from the individual studies were
325 highly over-enriched in the SA1^{ΔCoh} interactome (FDR= 1.1×10^{-15} , 1.4×10^{-47} , 7.7×10^{-19} ,
326 respectively) (Fig. 3f). To directly measure this, we optimised a colP method using the
327 S9.6 antibody in RAD21^{mAC} cells (Fig. 3g, S3h). In agreement with published results,
328 we found that S9.6 precipitated the known R-loop helicases AQR, DHX9, RNase H2
329^{43,44} as well as MCM3 and RNA Pol II (POLR2)⁴⁶. Both SA1 and SA2 precipitated with
330 S9.6 and treatment with RNase H (RNH) revealed the specificity of the S9.6-SA
331 interactions since the reduction of R-loop signal was proportional to the observed
332 reduction in colP of SA1 by S9.6 (Fig. 3g, S3h, i).

333 Finally, we used a high resolution, genome-wide method to detect R-loops in
334 HCT116 cells. RAD21^{mAC} cells were treated with RNH to confirm the specificity of our

335 method and with EtOH or IAA to assess the impact of cohesin loss on R-loops and
336 subjected to DNA-RNA Immunoprecipitation coupled with sequencing (DRIP-seq)
337 using the S9.6 antibody. We combined these datasets with our ChIP-seq for SA
338 proteins, RAD21 and CTCF in EtOH or IAA conditions to confirm the associations
339 described above. We detected 50,338 RNH-sensitive R-loop sites which were also
340 sensitive to acute degradation of RAD21, albeit not to the same extent as RNH
341 treatment (average S9.6 signal was reduced by 31.4% in RNH and 16.8% in IAA) (Fig.
342 3h, i). Among the RNH-sensitive R-loop sites, we detected two regimes of SA-R-loop
343 biology. A small proportion of R-loop sites directly overlapped with SA1/2, RAD21 and
344 CTCF in control EtOH conditions. These sites were enriched at genes and both the
345 SA1 and SA2 read density was sensitive to RAD21 loss (Fig. 3h, i, S3j, k). On the
346 other hand, a larger proportion of R-loops had SA signals adjacent (bound within 2kb
347 of the R-loop peak). Interestingly, these SA sites were enriched in repressed chromatin
348 and were not sensitive to RAD21 loss, in fact their read density was enriched
349 compared to EtOH conditions (Fig. 3h, i, S3j, k), reminiscent of the enrichment
350 observed previously by STORM imaging (Fig. 3e).

351

352 **NIPBL-independent cohesin loading mediated by SA proteins.**

353 Our results thus far revealed that SA^{ΔCoh} is localised to clustered regions, engages
354 with RNA and various RBPs and is localised to R-loops hybrids. Several lines of
355 evidence suggest that alongside the canonical NIPBL/Mau2 loading complex, SA
356 proteins contribute to cohesin's association with chromatin^{13,14} and that its functions
357 may go beyond simply acting as a bridging protein¹⁴. Thus, we hypothesized that SA
358 proteins support genome organization in their own right and herein facilitate cohesin's
359 association with chromatin.

360 The RAD21^{mAC} system has the advantage that when IAA is washed-off cells,
361 the RAD21 protein is no longer degraded and can become 're-loaded' onto chromatin.
362 We assessed this by measuring mClover signal intensity using IF and observed that it
363 was robustly lost in IAA conditions and was partially restored to EtOH levels within 4hr
364 of IAA withdrawal (Fig. 4a,b, S4a). We note the spatial distribution of RAD21 was itself
365 variable, ranging between highly compartmentalised and randomly distributed (Fig.

366 4a, c). This provided a unique opportunity to assess how SA influences cohesin
367 reloading *in vivo* and the potential role for RNA and R-loops in this process.

368 We assessed reloading using both single-cell and bulk methods, coupled with
369 siRNA-mediated KD to determine how specific proteins affected cohesin reloading *in*
370 *vivo*. We first measured the impact of the canonical cohesin loader, NIPBL. RAD21^{mAC}
371 cells were treated with scramble or NIPBL siRNAs and subsequently grown in EtOH
372 or IAA. The '0h' and '4h' post EtOH/IAA wash-off samples represent the extent of
373 cohesin *degradation* or *reloading*, respectively (Fig.S4b). Chromatin fractionation in
374 high-salt conditions followed by immunoblot analysis confirmed the loss of the loader
375 complex, NIPBL and MAU2 (known to become destabilised upon NIPBL loss⁴⁷). As
376 expected, in NIPBL KD conditions, mean RAD21 re-loading efficiency was reduced,
377 although surprisingly, this was incomplete (41% of the siRNA controls; mean re-
378 loading siNIPBL, 2.1 vs siCon, 3.6), and did not represent a statistically significant
379 difference (p=0.33) (Fig. 4d, e, S4c). This result was reproduced using IF, where mean
380 mClover signal in siNIPBL-treated cells was 45.1% of siRNA control (MFI siCON, 6563
381 vs siNIPBL, 3602) (Fig 4g), indicating that cells can still load cohesin in the absence
382 of NIPBL.

383 We reasoned that SA proteins may be contributing to the observed NIPBL-
384 independent reloading. Thus, we repeated the experiments to include siRNA to SA1
385 and SA2 together (siSA), and a siNIPBL+siSA condition. In both population and single
386 cell analysis of reloading, SA KD had a more dramatic effect on cohesin re-loading
387 efficiency than NIPBL KD, reducing RAD21 on chromatin to 51% of scramble controls
388 (mean siSA, 1.9 vs siCon, 5.1, p=0.002 for Fig. 4f, S4d and MFI siSA,2303 p<0.0001
389 for Fig. 4g). In the absence of both SA and NIPBL, cohesin reloading was reduced
390 further (mean siNIPBL+siSA, 1.4 vs siCon, 5.1, p=0.001 for Fig. 4f, S4d and MFI
391 siNIPBL+SA,1925 p<0.001 for Fig. 4g), indicating that SA performs an important and
392 complementary step to NIPBL during normal reloading. Given the differences between
393 SA1 and SA2 reported herein, we also performed the reloading experiment to separate
394 the effects of SA1 and SA2. As expected from our co-IP results (Fig. 1c), RAD21 levels
395 in RAD21^{mAC} cells were more affected by siSA2 than siSA1 (Fig. S4e). We observed
396 that cohesin reloading was more efficient in siSA2 (where SA1 is present) than in
397 siSA1 (where only SA2 is present), and that siSA1 was similar in reloading to siSA

398 (Fig. S4e). Together these observations suggest that the bulk of the reloading in IAA
399 conditions is supported by SA1.

400

401 **SA proteins stabilize nascent RNA in the absence of cohesin.**

402 Given the association of SA proteins with RNA and RBPs and the dependence of
403 cohesin reloading on SA1, we tested the requirement for RNA in cohesin reloading.
404 Cells were treated as above with a pulse of 5 ethynyl uridine (EU) prior to collection.
405 EU becomes actively incorporated into nascent RNA and can be measured by IF
406 alongside the change in RAD21-mClover. While a significant reduction in nascent
407 RNA signal was detected upon treatment with Triptolide (TRP), mClover signal was
408 not significantly changed compared to IAA washoff conditions, indicating that RNA is
409 not a key determinant of cohesin reloading *per se* (Fig 4h, left panel). However, we
410 did observe an increase in nascent RNA upon acute RAD21 degradation which
411 returned to EtOH levels when cohesin became reloaded onto chromatin (Fig 4h, right
412 panel). These results pointed to the stabilization of nascent RNA in the absence of
413 cohesin. Given the association of SA1^{ΔCoh} with RNA and RBP, we repeated the
414 experiment in siSA1 KD conditions. Again, we observed an increase in nascent RNA
415 in IAA conditions compared to EtOH control, but this was no longer detected in cells
416 treated with siRNA to SA1 (Fig. 4i). Our results point to a role for SA1 in the
417 stabilization of nascent RNA in the absence of cohesin and a possible competition
418 between SA-RNA and SA-cohesin associations.

419

420 Our results thus far showed that SA proteins remain chromatin associated in the
421 absence of cohesin (Fig. 1), when they bind RBP (Fig. 2) and RNA and are localized
422 to R-loops (Fig. 3). We also report that SA1 proteins contribute to cohesin's re-
423 association with chromatin and that this involves nascent RNA (Fig. 4). Thus, we
424 reasoned that SA may facilitate cohesin reloading at R-loops. It was technically
425 challenging to measure reloading upon over-expression of ppyCAG-RNaseH. As an
426 alternative, we used STORM imaging to assess the nuclear distribution of the reloaded
427 cohesin in the context of R-loop clusters by comparing EtOH- and IAA-treated to IAA-
428 washoff RAD21^{mAC} cells (Fig. 4j). As before, we measured the ratio of signal (this time
429 RAD21-mClover) inside and outside of the S9.6 mask. Interestingly, in EtOH

430 conditions, RAD21 localizations were depleted from the S9.6 domain (mean ratio 0.95)
431 (Fig. 4j) similar to what we observed for SA1 (Fig 3h). Since STORM is such a
432 sensitive approach, trace localizations of mClover will always be detected, even in IAA
433 conditions when the bulk of the signal is lost. The few localizations we observed were
434 indeed modestly enriched within the S9.6 mask, although these were not significantly
435 different from EtOH (mean ratio 1.08, $p=0.10$). These localizations may represent
436 either extremely stable or freshly loaded cohesin. Upon IAA washoff, new RAD21-
437 mClover molecules are readily detected, became significantly enriched within S9.6
438 domains compared to EtOH treated cells (mean ratio 1.19, $p=0.029$) and were
439 sensitive to treatment with RNase H (mean ratio 0.98) (Fig 4j). Overall, our results
440 point to a role for SA1 proteins in mediating reloading of cohesin at R-loops.

441

442 **A basic exon in the C-terminus of SA2 tunes interactions with RBPs.**

443 While both SA1 and SA2 played a role in cohesin's reloading, SA1 was the dominant
444 paralog (Fig S4e). In addition, SA2 was not able to compensate for SA1 in R-loop
445 stability (Fig S3f), despite its interaction with RNA (Fig. 3a, b) and R-loops (Fig. S3h).
446 Previous publications have described association of RBP from SA2 MS-IP in HCT
447 cells⁴⁰. Indeed, several of these RBPs overlap with the proteins described here as
448 SA1 interactors (Fig. 2b, c) and are enriched in SA1 IP in IAA conditions (Fig. 2a, d).
449 However, we did not observe robust enrichment of RBPs compared to input in SA2
450 IP, in either EtOH or IAA conditions. This was reminiscent of the differential
451 interactions between SA1 and SA2 with F/YXF containing proteins (Fig. 2a). These
452 results thus raised the question of whether additional features in SA2 may be required
453 to stabilize these interactions and functions.

454 SA1 and SA2 express transcript variants in RAD21^{mAC} cells. We re-analysed
455 publicly available RNA-seq datasets and quantified alternative splicing profiles using
456 VAST-tools analysis⁴⁸. One prominent variant [which is conserved between human
457 and mouse \(Fig. S5a, b\)](#), arises from the alternative splicing of a single C-terminal
458 exon, exon 31 in SA1 (SA1^{e31Δ}) and exon 32 in SA2 (SA2^{e32Δ}) (Fig. 5a). The
459 significance of this is unknown. We found that in human HCT cells, the majority of
460 SA1 mRNAs *include* e31 (average 'percent spliced in' (PSI) 97.7%), while the majority
461 of SA2 mRNAs *exclude* e32 (average PSI 20.4%) (Fig. 5b, S5a, b). We confirmed
462 this at the protein level by designing custom esiRNAs to specifically target SA1 e31 or

463 SA2 e32 (Methods). Smartpool (SP) KD reduced the levels of SA1 and SA2 to similar
464 extents compared to scrambled controls (87% and 94%, respectively) (Fig. 5c).
465 Specific targeting of SA1 e31 led to a reduction of 85% of SA1 compared to esiRNA
466 control (which was comparable to SP KD). In contrast, SA2 e32 targeting had a
467 minimal effect on SA2 protein levels compared to its esiRNA control (reduction of 2%)
468 (Fig. 5c), in line with the PSI data (Fig. 5b) and indicating that the dominant SA2
469 isoform does not contain e32.

470 These results imply that cells ‘tune’ the availability of e31/32 in SA proteins,
471 prompting us to investigate the nature of these exons. Interestingly, the amino acid
472 (aa) sequence of the spliced SA exons encode a highly basic domain within an
473 otherwise acidic C-terminus (Fig. 5a, zoom-in). Overall, the SA paralogs are highly
474 homologous, however the N- and C-termini diverge in their aa sequence. Despite this,
475 e31 and e32 have retained their basic properties (pI=10.4 and 9.9, respectively)
476 (Fig.5a, zoom-in). Basic patches can act as regulatory domains and bind nucleic acids
477 prompting us to ask whether these alternatively spliced basic exons contribute to the
478 association of SA proteins with RNA (Fig 3a). We cloned cDNAs from HCT116 cells
479 representing the exon32-containing SA2 (SA2^{e32+}) and the canonical exon32-lacking
480 SA2 (SA2^{e32Δ}), tagged them with YFP, expressed them in HCT116 cells and purified
481 the tagged isoforms to compare their ability to interact with RNA (Fig. 5d) using CLIP.
482 While the presence of e32 did not change the ability of SA2 to interact with RNA (Fig.
483 5d, blue arrows), cells expressing the alternative exon routinely enriched RBPs with
484 molecular weights ~110-140kDa (Fig 5d, black arrow), strongly suggesting that the
485 e32 domain may act to stabilize the association of SA2 with RBPs.

486 To identify the proteins stabilized by the presence of e32, we coupled YFP-SA2
487 isoform CLIP with Mass Spectrometry. Three biological replicate IPs were prepared
488 from RAD21^{mAC} cells that were transfected with either YFP-SA2^{e32+} or YFP-SA2^{e32Δ}.
489 YFP IP efficiency for SA2^{e32+} or SA2^{e32Δ} was similar and both isoforms interacted with
490 core cohesin subunits (Fig S5c). We identified a total of 238 proteins, the majority of
491 which overlap in the two SA IPs and with a previously published SA2 IP⁴⁰ (Fig S5c,
492 d). We used a pairwise analysis of SA2^{e32+} vs SA2^{e32Δ} samples to generate a fold-
493 change value for each putative interactor (Fig 5e). GO analysis of proteins changed
494 by at least 1.5-fold, and absent in Mock IP revealed a mild enrichment for post-

495 translational modification category from the SA2^{e32Δ} IP (FDR=0.0234, p=1.35e-06),
496 and conversely an enrichment of the RNA Binding category from SA2^{e32+} (FDR=
497 3.43E-05, p=6.56E-09). Interestingly, the enriched proteins included YTHDC1 and
498 YTHDF3 (previously identified in the SA1^{ΔCoh} interactome, Fig 2b), DIS3 and POLR2B,
499 all known to play key roles in RNA-protein complexes and stability, have molecular
500 weights ~110-140kDa and thus likely represent the specifically enriched band in the
501 CLIP experiments (Fig 5d, black arrow). Finally, the observation that a basic exon 32
502 domain in SA2 supports the stability of RNA-RBP interactions led us to investigate if
503 exon 32 also stabilized SA2 at R-loops. We repeated the S9.6 IP in RAD21^{mAC} cells
504 expressing either YFP-SA2^{e32+} or YFP-SA2^{e32Δ}. As before, AQR and MCM3 were
505 enriched by S9.6 IP (Fig. 5f) and we found that SA2^{e32+} was more enriched in the S9.6
506 IP compared to SA2^{e32Δ} (enrichment of 1.8x and 1.24x respectively, relative to
507 endogenous SA2) (Fig 5f, g). Taken together, our results support a role for the
508 alternatively spliced C-terminal basic domain of SA in stabilizing interactions with
509 RBPs and R-loops.

510

511

512 **DISCUSSION**

513 Whether SA proteins function in their own right outside of the cohesin complex is rarely
514 considered. Consequently, our understanding of how these proteins contribute to
515 cohesin function and disease is incomplete. In this study, we shed light on this
516 question by uncovering a diverse repertoire of SA1 interactors in cells acutely depleted
517 for the cohesin ring. This ranges from proteins associated with translation and
518 ribosome biogenesis to RNA processing factors and regulators of
519 the epitranscriptome. These observations suggest that SA1 has a previously
520 unappreciated role in post-transcriptional regulation of gene expression which offers
521 much-needed new insight into its roles in disease and cancer.

522 Acute depletion of the cohesin ring has allowed us to capture a moment in the
523 normal life cycle of cohesin–DNA associations and unveiled a previously unknown
524 step for SA proteins herein. We show that SA proteins, independent of the cohesin
525 trimer, bind to DNA and RNA, either in the context of RNA:DNA hybrid structures, as
526 we have shown here, or perhaps sequentially, and use this platform for the loading of

527 cohesin to chromatin. Our results are supportive of biophysical observations of SA
528 proteins and R-loops⁴⁹ and *in vitro* assessment of cohesin loading at DNA
529 intermediates²⁶. These results point to the importance of DNA *structure*, as opposed
530 to sequence, in the targeting of cohesin to chromatin. Furthermore, structural studies
531 suggest that NIPBL and SA1 together bend DNA and cohesin to guide DNA entering
532 into the cohesin ring^{19,20,50}. Our work shows that in cells lacking either the canonical
533 NIPBL/MAU2 loader complex or the SA proteins, cohesin can still associate with
534 chromatin, suggesting that loading can occur with either component alone, albeit most
535 effectively together.

536 Our results represent a new view onto the role of SA proteins in cohesin biology.
537 Since SA paralogs have distinct terminal ends and nucleic acid targeting mechanisms
538^{24,25}, their recruitment to chromatin may be specified by unique DNA, RNA or protein-
539 interactions, or indeed all three. Such diversification of loading platforms would be
540 important in large mammalian genomes to ensure sufficient amounts of cohesin on
541 chromatin or to stabilize cell-type specific chromatin structures⁵¹. Indeed, SA1 and
542 SA2 show clear differences in interaction with F/YXF-motif containing proteins, despite
543 the fact that both paralogs contain a CES domain⁵², underscoring the importance of
544 investigating interactions *in vivo* and arguing that additional features play an important
545 role in complex stabilization. In this context, RNA-associated protein interaction has
546 previously been shown to support cohesin stabilisation at CTCF at the *IGF2/H19* locus
547⁵³. These results are in line with our findings that a basic domain in the unstructured
548 C-terminal portion of SA supports RNA-associated protein interactions.

549 This study also reveals SA1 as a novel regulator of R-loop homeostasis. It is
550 noteworthy that other suppressors of R-loop formation include RNA processing
551 factors, chromatin remodellers and DNA repair proteins²⁸ which all function in the
552 context of nuclear bodies⁵⁴. We find that SA1 proteins are enriched at very distal
553 chromatin interactions in cohesin-depleted Hi-C data, interact with numerous RBPs
554 known to condense in 3D^{55,56} and are enriched in S9.6 domains in cells where we find
555 cohesin becomes associated with chromatin. Harnessing such condensates would
556 provide an efficient loading platform for cohesin at sites of similar biological function.
557 Yeast cohesin has been shown to mediate phase separated condensate structures⁵⁷.
558 Our results support this view and further suggest that it is SA (and possibly

559 predominantly SA1 in HCT116 cells), with its propensity for intrinsically disordered
560 domains ⁵¹ that contribute to this formation, thereby linking cohesin loading to
561 biological functions. We note that if SA paralogs or isoforms direct different localization
562 of cohesin loading or stability of its association, this could have important implications
563 in our understanding of disease where SA proteins are commonly mutated, such as in
564 cancers.

565 **ACKNOWLEDGMENTS**

566 This work was supported by a Senior Research Fellowship from the Wellcome Trust
567 awarded to S.H. (106985/Z/15/Z) and a CRUK PhD studentship awarded to H.P. The
568 Proteomics work was supported by the CRUK–UCL Centre Award [C416/A25145].
569 We are grateful to Jernej Ule for his support with DRIP-sequencing and to Julian
570 Zagalak and the CRICK sequencing facility for reagents, advice and assistance. We
571 thank Stanimir Dulev for his contributions at the early stages of the project and Jiten
572 Manji for his support with microscopy. We also thank Konstantina Skourti-Stathaki for
573 advice about S9.6 antibody, IFs and R-loops. We are grateful to the members of the
574 Hadjur lab for critical discussions and reading of the manuscript.

575

576 **AUTHOR CONTRIBUTIONS**

577 H.P. and S.H. conceived the project. H.P. designed and performed the colIP, Mass
578 spectrometry, DRIP-sequencing and cohesin re-loading experiments, analysed the
579 ChIP, DRIP and Hi-C data and performed the statistical analysis for mass
580 spectrometry with the support of A.B. and S.S. Y.L. performed and analysed all
581 imaging experiments (apart from STORM), derived clonal lines of RAD21^{mAC} cells,
582 cloned YFP-tagged SA2 cDNAs and performed CLIP and CLIP-MS together with
583 M.T.C. W.V. performed Hi-C, ChIP-seq and splicing analyses. M.V.N., L.M. and
584 M.P.C. performed and analysed STORM imaging. D.P. discovered splicing features
585 of the SA isoforms. H.P., Y.L. and M.B. prepared cellular materials for CLIP, which
586 was carried out by M.B. and M.T.C. and supervised by R.G.J. A.B. and S.S. performed
587 mass spectrometric and proteomic analysis. H.P., Y.L. and S.H. formatted all figures
588 and wrote the manuscript with input from all authors.

589

590 **Accession Numbers**

591 Genomic data generated in this study (ChIP-seq) was submitted to GEO with the
592 Accession GSE167887. The mass spectrometry proteomics data was deposited to the
593 ProteomeXchange Consortium via the PRIDE partner repository with the dataset
594 identifier PXD024354.

595

596 **Declaration of Interests**

597 The authors declare no competing interests.

598 **FIGURE LEGENDS**

599

600 **Figure 1. SA interacts with CTCF in the absence of cohesin.**

601 a) Representative confocal images of SA1 and CTCF IF in RAD21^{mAC} cells treated
602 with EtOH (EtOH) as a control or Auxin (IAA) for 4hrs. Nuclei were counterstained with
603 DAPI.

604

605 b) Imaris quantification of the relative mean fluorescence intensity (MFI) of mClover,
606 CTCF, SA1 and SA2 in EtOH and IAA-treated RAD21^{mAC} cells. Whiskers and boxes
607 indicate all and 50% of values, respectively. Central line represents the median.
608 Asterisks indicate a statistically significant difference as assessed using two-tailed t-
609 test. **** p<0.0001. n>50 cells/condition from 3 biological replicates.

610

611 Chromatin coIP of (c) SA1, SA2 and IgG with RAD21 and CTCF or (d) CTCF and
612 IgG with RAD21, SA1 and SA2 in RAD21^{mAC} cells treated with EtOH or IAA for
613 4hrs. Input represents (c) 2.5% and (d) 1.25% of the material used for
614 immunoprecipitation.

615

616 e) Dual-color STORM images of SA1 (green) and CTCF (magenta) in EtOH and IAA-
617 treated RAD21^{mAC} cells. Representative full nuclei and zoomed nuclear areas are
618 shown. Line denotes 2 microns and 200nm for full nuclei and zoomed areas
619 respectively. See Supplementary Figures for SA2 STORM images.

620

621 f) Mean CTCF, SA1 and SA2 localization densities (localizations normalized with
622 nuclear area) in EtOH and IAA-treated RAD21^{mAC} cells (n = >30, >17 and >15 nuclei
623 for CTCF, SA1 and SA2 respectively). Mean and SD are plotted, Mann Whitney test.
624 ** p<0.005, *** p<0.0005, **** p<0.0001.

625

626 g) Mean Nearest Neighbor Distance (NND) of CTCF, SA1 and SA2 clusters in
627 nanometers in EtOH and IAA-treated cells (n = >38, >14 and >23 nuclei for CTCF,
628 SA1 and SA2 respectively). Mean and SD are plotted, Mann Whitney test. ****
629 p<0.0001.

630

631 h) NND distribution plot of the distance between CTCF and SA1 (left panel) or SA2
632 (right panel) clusters in EtOH and IAA-treated cells. Experimental data are shown as
633 continuous lines, random simulated data are displayed as dotted lines.

634

635 i) ChIP-seq deepTools heat map of CTCF, SA1, SA2, Rad21 and
636 SMC3 binding profiles in control (EtOH) and IAA-treated RAD21^{mAC} cells. Selected
637 regions are bound by CTCF in control conditions.

638

639 j) Analysis of contact frequency hotspots from Hi-C libraries generated from EtOH-
640 treated (top row) and IAA-treated (bottom row) RAD21^{mAC} cells. Contact frequencies
641 were calculated in two distance ranges of 100kb – 1Mb and 1-5Mb. The last column
642 includes contact frequencies specifically at SA-CTCF^{ΔCoh} binding sites.

643

644

645 **Figure 2. Characterization of SA1 protein-protein interaction network in RAD21-**
646 **depleted cells.**

647 a) Chromatin coIP of SA1, SA2 and IgG with 4 predicted CES-binding proteins in
648 RAD21^{mAC} cells treated with EtOH or IAA for 4hrs. Input represents 1.25% of the
649 material used for immunoprecipitation.

650
651 b) Volcano plot displaying the statistical significance (-log₁₀ p-value) versus
652 magnitude of change (log₂ fold change) from SA1 IP-MS data produced
653 from untreated or IAA-treated RAD21^{mAC} cells (n=3). Vertical dashed lines represent
654 changes of 1.5-fold. Horizontal dashed line represents a pvalue of 0.1.
655 Cohesin complex members and validated high-confidence proteins have been
656 highlighted.

657
658 c) SA1^{ΔCoh} interaction network of protein-protein interactions identified in
659 RAD21^{mAC} cells using STRING. Node colours describe the major
660 enriched categories, with squares denoting helicases. Proteins within each enrichment
661 category were subset based on p-value change in B). See supplemental figures for
662 full network. Dashed boxes indicate the proteins and categories which were
663 specifically enriched in IAA-treatment compared to the SA1 interactome.

664
665 d) Chromatin IP of SA1 and IgG in RAD21^{mAC} cells treated with EtOH or IAA and
666 immunoblotted with antibodies to validate the proteins identified by IP-MS. Input
667 represents 1.25% of the material used for immunoprecipitation. * We note that ESYT2
668 is a F/YXF-motif containing protein.

669
670

671 **Figure 3. SA proteins bind to RNA and localise to R-loops.**

672 a) CLIP for SA1, b) SA2 and IgG controls. Autoradiograms of crosslinked ³²P-labelled
673 RNA are shown at the top and the corresponding immunoblots, below. CLIP was
674 performed with and without UV crosslinking and polynucleotide kinase (PNK) and with
675 high (H; 1/50 dilution) or low (L; 1/500 dilution) concentrations of RNase I.

676
677 c) CLIP for SA1, SA2 and IgG control in EtOH (-) or IAA-treated (+) Rad21^{mAC} cells.
678 ³²P-labelled RNA and the corresponding immunoblots are shown as above.

679
680 d) Top, Representative confocal images of S9.6 and SA1 IF in RAD21^{mAC} cells
681 untreated or overexpressing ppyCAG-RNaseH-v5. Expressing cells were identified with
682 v5 staining. Nuclear outlines (white) are derived from DAPI counterstain. Bottom,
683 Imaris quantification of the relative mean fluorescence Intensity (MFI) of S9.6 and
684 SA1. Data are from two biological replicates with >75 cells
685 counted/condition. Quantifications and statistical analysis were done as above.

686
687 e) STORM analysis of localization density for SA1 in S9.6 masks in EtOH and IAA.
688 Ratio of SA1 localizations inside and outside S9.6 masks is shown. Ratio of above 1
689 represents an enrichment within the S9.6 domain. Mean and SD are plotted, statistics
690 based on One-Way Anova test. Data are from two biological replicates.

691

692 f) $-\log_{10}$ transformed adjusted p-value (FDR) for enrichment of S9.6 interactome data
693 from Cristini et al. and Wang et al., with the SA1^{ΔCoh} interactome. Overlap indicates
694 the proteins identified in both of the S9.6 interactome datasets, representing a high
695 confidence R-loop interactome list.

696
697 g) Chromatin colP of S9.6 and IgG in RAD21^{mAC} cells treated with RNase H and
698 immunoblotted with antibodies representing known R-loop proteins, as well as
699 SA1. Input represents 1.25% of the material used for
700 immunoprecipitation. Bottom, S9.6 dot blot of lysates used in colP.

701
702 h) deepTools heatmap of DRIP-seq and ChIP-seq from RAD21^{mAC} cells. DRIP-seq
703 was carried out in control (EtOH), RNase H (RNH) and IAA-treated cells. ChIP-seq
704 was carried out for SA1, SA2, RAD21, CTCF and IgG in EtOH and IAA treated cells.
705 Regions were selected based on DRIP-seq sensitivity to RNH and proximity with SA1
706 ChIP-seq. BEDTools identified regions of overlap or adjacent SA1 co-binding.

707
708 i) Summary plots showing mean DRIP-seq (top) or ChIP-seq (bottom) read density
709 across the regions from h), including sites of R-loop and SA 'overlap' (Left) or
710 'adjacent' (right) regions. Input samples are indicated with dotted lines.

711
712

713 **Figure 4. SA proteins contribute to cohesin loading.**

714 a) Representative confocal images of immunofluorescence for mClover in EtOH, IAA
715 and IAA washoff conditions. White lines denote nuclei based on DAPI staining.

716
717 b) Imaris quantification of the mean fluorescence intensity (MFI) of mClover in EtOH,
718 IAA-treated and IAA washoff RAD21^{mAC} cells. Analysis and statistics as before. n>50
719 cells/condition from 2 biological replicates.

720
721 c) Examples of individual cells 4hr post IAA washoff showing different distributions of
722 mClover signal within the nucleus. White lines denote nuclei based on DAPI staining.

723
724 d) Representative immunoblot analysis of chromatin-bound RAD21, MAU2 and NIPBL
725 levels in RAD21^{mAC} cells treated with scramble control siRNA (si Con) or siRNA to
726 NIPBL followed by EtOH or IAA treatment. 0h and 4h represent no wash-off of IAA or
727 a sample taken 4 hours after washoff of IAA (the 'reloading timepoint'). H3 was used
728 as a loading control. *NB* The full blots are in Fig. S4c.

729
730 e) Western blot densitometry quantification. RAD21 fold change relative
731 to siCon samples at the 0h timepoint in siCon 4h (grey), siNIPBL 0hr (light blue)
732 and siNIPBL 4hr (dark blue). Whiskers and boxes indicate all and 50% of values
733 respectively. Central line represents the median. Statistical analysis as assessed
734 using a two-tailed t-test. Data is from 8 biological replicates.

735
736 f) Representative immunoblot analysis of chromatin-bound RAD21, SA1, SA2, MAU2
737 and NIPBL levels in RAD21^{mAC} cells treated according to the schematic described in
738 Fig. S4b and including samples treated with siRNA to SA1 and SA2 together (siSA)

739 and siRNA to NIPBL + siSA. H3 was used as a loading control. Quantifications can
740 be seen in in Fig. S4c.

741

742 g) Imaris quantification of the mClover MFI in EtOH, IAA-treated and IAA
743 washoff RAD21^{mAC} cells treated with siRNA to NIPBL, SA1/2 and siRNA to NIPBL
744 + siSA. Asterisks indicate a statistically significant difference as assessed using 2-
745 tailed T-test. Data is from 2 biological replicates with >50 cells per experiment.

746

747 h) Imaris quantification of the mClover MFI (left) and RNA (based on EU incorporation
748 (right) in EtOH and IAA-treated RAD21^{mAC} cells. Whiskers and boxes indicate all and
749 50% of values, respectively. Central line represents the median. Asterisks indicate a
750 statistically significant difference as assessed using one-way ANOVA. n>50
751 cells/condition from 2 biological replicates.

752

753 i) Analysis of mClover MFI as in h) above, this time treated with siRNA to SA1/2 or
754 scrambled controls. Analysis and statistics as above. n>50 cells/condition from 2
755 biological replicates.

756

757 j) STORM analysis of localization density for RAD21-mClover in S9.6 masks in EtOH
758 (black), IAA (grey) and IAA washoff (green) conditions. Ratio of mClover localizations
759 inside and outside S9.6 masks is shown. Ratio of above 1 represents an enrichment
760 within the S9.6 domain. Mean and SD are plotted, statistics based on One-Way Anova
761 test. Data are from two biological replicates.

762

763

764 **Figure 5. A basic exon in SA2 influences RBP stability.**

765 a) Schematic of the SA1 and SA2 proteins showing the SA1-specific AT-hook, the
766 conserved CES domain (blue) and the acidic C-terminus (green) which contains
767 the basic alternatively spliced exon (red). Right-hand zoom-in indicates the spliced
768 exons for SA1 (top) and SA2 (bottom) and the pI for each. The conservation scores for
769 the divergent N- and C-termini and the middle portion of the proteins which contains
770 the CES domain are shown.

771

772 b) Percent Spliced In (PSI) calculations for SA1 exon 31 (black) and SA2 exon 32
773 (grey) based on VAST-Tools analysis of RNA-seq from multiple datasets (see
774 Methods).

775

776 c) (top) Immunoblot analysis of SA1 levels in chromatin lysates after treatment
777 with scrambled siRNAs (siCon), SmartPool SA1 siRNAs (siSA1 SP),
778 control esiRNAs (esiCon) and esiRNA designed to target SA1 exon
779 31 for 48hrs in RAD21^{mAC} cells. (bottom) Immunoblot analysis of SA2 levels in
780 chromatin lysates after treatment with scrambled siRNAs (siCon), SmartPool SA2
781 siRNAs (siSA2 SP), control esiRNAs (esiCon) and siRNA designed to target SA2 exon
782 32 for 48hrs in RAD21^{mAC} cells. H3 serves as a loading control. The percentage of
783 knockdown (KD) after SA signal is normalised to H3 is shown.

784

785 d) CLIP with endogenous SA2, IgG control and cells where either YFP-tagged SA2
786 containing exon 32 (e32+) or YFP-tagged SA2 lacking exon 32 are expressed for

787 48hrs. CLIP reveals RNA associated with SA2 (blue arrows) and RBPs
788 which specifically associate with exon-32 containing SA2 (black arrow).

789

790 e) Volcano plot displaying the statistical significance ($-\log_{10}$ p-value) versus
791 magnitude of change (\log_2 fold change) from IP-MS of HCT116 cells expressing
792 either YFP-SA2^{e32+} or YFP-SA2^{e32Δ} (n=3 biological replicate IP).
793 Cohesin complex members are highlighted in green and the two most enriched
794 functional categories of RNA-binding proteins in blue or Post-translational modification
795 in red.

796

797 f) Chromatin coIP of S9.6 and IgG in RAD21^{mAC} cells expressing the YFP-SA2
798 isoforms and immunoblotted with antibodies representing known R-loop proteins, as
799 well as endogenous SA2 (eSA2). Input represents 1.25% of the material used for
800 immunoprecipitation. Bottom, S9.6 dot blot of lysates used in coIP. *NB* the shift in
801 SA2 signal representing overexpressed protein (SA2-YFP).

802

803 g) Quantification of the immunoblot signal from f) of SA2 in the YFP-SA2 isoform band
804 relative to input and to eSA2 signal. YFP-SA2^{e32+} is more enriched by S9.6 IP
805 compared to YFP-SA2^{e32Δ}.

806 REFERENCES

- 807 1. Leiserson, M. D. M. *et al.* Pan-cancer network analysis identifies combinations of rare
808 somatic mutations across pathways and protein complexes. *Nat. Genet.* **47**, 106–114
809 (2015).
- 810 2. Horsfield, J. A. *et al.* Cohesin-dependent regulation of Runx genes. *Development* **134**,
811 2639–2649 (2007).
- 812 3. Viny, A. D. & Levine, R. L. Cohesin mutations in myeloid malignancies made simple.
813 *Curr. Opin. Hematol.* **25**, 61–66 (2018).
- 814 4. Hadjur, S. *et al.* Cohesins form chromosomal cis-interactions at the developmentally
815 regulated IFNG locus. *Nature* **460**, 410–413 (2009).
- 816 5. Sofueva, S. *et al.* Cohesin-mediated interactions organize chromosomal domain
817 architecture. *EMBO J.* **32**, 3119–3129 (2013).
- 818 6. Zuin, J. *et al.* Cohesin and CTCF differentially affect chromatin architecture and gene
819 expression in human cells. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 996–1001 (2014).
- 820 7. Kojic, A. *et al.* Distinct roles of cohesin-SA1 and cohesin-SA2 in 3D chromosome
821 organization. *Nature Publishing Group* **25**, 496–504 (2018).
- 822 8. Rao, S. S. P. *et al.* Cohesin Loss Eliminates All Loop Domains. *Cell* **171**, 305–309.e24
823 (2017).
- 824 9. Wutz, G. *et al.* Topologically associating domains and chromatin loops depend on
825 cohesin and are regulated by CTCF, WAPL, and PDS5 proteins. *EMBO J.* **36**, 3573–
826 3599 (2017).
- 827 10. Fudenberg, G., Abdennur, N., Imakaev, M., Goloborodko, A. & Mirny, L. Emerging
828 Evidence of Chromosome Folding by Loop Extrusion. *bioRxiv* 1–24 (2018).
829 doi:10.1101/264648
- 830 11. Balbás-Martínez, C. *et al.* Recurrent inactivation of STAG2 in bladder cancer is not
831 associated with aneuploidy. *Nat. Genet.* **45**, 1464–1469 (2013).
- 832 12. Solomon, D. A. *et al.* Frequent truncating mutations of STAG2 in bladder cancer. *Nat.*
833 *Genet.* **45**, 1428–1430 (2013).
- 834 13. Murayama, Y. & Uhlmann, F. Biochemical reconstitution of topological DNA binding by
835 the cohesin ring. *Nature* **505**, 367–371 (2014).
- 836 14. Orgil, O. *et al.* A conserved domain in the scc3 subunit of cohesin mediates the
837 interaction with both mcd1 and the cohesin loader complex. *PLoS Genet.* **11**, e1005036
838 (2015).
- 839 15. Ciosk, R. *et al.* Cohesin's Binding to Chromosomes Depends on a Separate Complex
840 Consisting of Scc2 and Scc4 Proteins. *Molecular Cell* **5**, 243–254 (2000).
- 841 16. Kikuchi, S., Borek, D. M., Otwinowski, Z., Tomchick, D. R. & Yu, H. Crystal structure of
842 the cohesin loader Scc2 and insight into cohesinopathy. *Proc. Natl. Acad. Sci. U.S.A.*
843 **113**, 12444–12449 (2016).
- 844 17. Hara, K. *et al.* Structure of cohesin subcomplex pinpoints direct shugoshin-Wapl
845 antagonism in centromeric cohesion. *Nature Publishing Group* **21**, 864–870 (2014).
- 846 18. Li, Y. *et al.* Structural basis for Scc3-dependent cohesin recruitment to chromatin. *eLIFE*
847 **7**, 352 (2018).
- 848 19. Shi, Z., Gao, H., Bai, X.-C. & Yu, H. Cryo-EM structure of the human cohesin-NIPBL-
849 DNA complex. *Science* **368**, 1454–1459 (2020).
- 850 20. Higashi, T. L. *et al.* A Structure-Based Mechanism for DNA Entry into the Cohesin Ring.
851 *Molecular Cell* **79**, 917–933.e9 (2020).
- 852 21. Xiao, T., Wallace, J. & Felsenfeld, G. Specific sites in the C terminus of CTCF interact
853 with the SA2 subunit of the cohesin complex and are required for cohesin-dependent
854 insulation activity. *Mol. Cell. Biol.* **31**, 2174–2183 (2011).
- 855 22. Saldaña-Meyer, R. *et al.* CTCF regulates the human p53 gene through direct interaction
856 with its natural antisense transcript, Wrap53. *Genes Dev.* **28**, 723–734 (2014).
- 857 23. Bisht, K. K., Daniloski, Z. & Smith, S. SA1 binds directly to DNA through its unique AT-
858 hook to promote sister chromatid cohesion at telomeres. *J. Cell. Sci.* **126**, 3493–3503
859 (2013).
- 860 24. Lin, J. *et al.* Functional interplay between SA1 and TRF1 in telomeric DNA binding and
861 DNA-DNA pairing. *Nucleic Acids Research* **44**, 6363–6376 (2016).

- 862 25. Countryman, P. *et al.* Cohesin SA2 is a sequence-independent DNA-binding protein that
863 recognizes DNA replication and repair intermediates. *J. Biol. Chem.* **293**, 1054–1069
864 (2018).
- 865 26. Murayama, Y., Samora, C. P., Kurokawa, Y., Iwasaki, H. & Uhlmann, F. Establishment
866 of DNA-DNA Interactions by the Cohesin Ring. *Cell* 1–29 (2018).
867 doi:10.1016/j.cell.2017.12.021
- 868 27. Zheng, G., Kanchwala, M. & Xing, C. MCM2–7-dependent cohesin loading during S
869 phase promotes sister-chromatid cohesion. *eLIFE* 1–25 (2018).
870 doi:10.7554/eLife.33920.001
- 871 28. García-Muse, T. & Aguilera, A. R Loops: From Physiological to Pathological Roles. *Cell*
872 1–15 (2019). doi:10.1016/j.cell.2019.08.055
- 873 29. Crossley, M. P., Bocek, M. & Cimprich, K. A. R-Loops as Cellular Regulators and
874 Genomic Threats. *Molecular Cell* **73**, 398–411 (2019).
- 875 30. Li, X. & Manley, J. L. Inactivation of the SR protein splicing factor ASF/SF2 results in
876 genomic instability. *Cell* **122**, 365–378 (2005).
- 877 31. Boque-Sastre, R. *et al.* Head-to-head antisense transcription and R-loop formation
878 promotes transcriptional activation. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 5785–5790
879 (2015).
- 880 32. Powell, W. T. *et al.* R-loop formation at Snord116 mediates topotecan inhibition of
881 Ube3a-antisense and allele-specific chromatin decondensation. *Proc. Natl. Acad. Sci.*
882 *U.S.A.* **110**, 13938–13943 (2013).
- 883 33. Natsume, T., Kiyomitsu, T., Saga, Y. & Kanemaki, M. T. Rapid Protein Depletion in
884 Human Cells by Auxin-Inducible Degron Tagging with Short Homology Donors. *Cell Rep*
885 **15**, 210–218 (2016).
- 886 34. Parelho, V. *et al.* Cohesins functionally associate with CTCF on mammalian
887 chromosome arms. *Cell* **132**, 422–433 (2008).
- 888 35. Wendt, K. S. *et al.* Cohesin mediates transcriptional insulation by CCCTC-binding factor.
889 *Nature* **451**, 796–801 (2008).
- 890 36. Wutz, G. *et al.* ESCO1 and CTCF enable formation of long chromatin loops by
891 protecting cohesin-STAG1 from WAPL. *eLIFE* **9**, (2020).
- 892 37. Ricci, M. A., Manzo, C., García-Parajo, M. F., Lakadamyali, M. & Cosma, M. P.
893 Chromatin fibers are formed by heterogeneous groups of nucleosomes in vivo. *Cell* **160**,
894 1145–1158 (2015).
- 895 38. Roig, M. B. *et al.* Structure and function of cohesin’s Scc3/SA regulatory subunit.
896 *FEBS Letters* **588**, 3692–3702 (2014).
- 897 39. Hons, M. T. *et al.* Topology and structure of an engineered human cohesin complex
898 bound to Pds5B. *Nature Communications* **7**, 1–11 (2017).
- 899 40. Kim, J.-S. *et al.* Systematic proteomics of endogenous human cohesin reveals an
900 interaction with diverse splicing factors and RNA-binding proteins required for mitotic
901 progression. *J. Biol. Chem.* **294**, 8760–8772 (2019).
- 902 41. Panigrahi, A. K., Zhang, N., Otta, S. K., Journal, D. P. B. 2012. A cohesin–RAD21
903 interactome. doi:10.1042/BJ20111745
- 904 42. Santos-Pereira, J. M. & Aguilera, A. R loops: new modulators of genome dynamics and
905 function. *Nat. Rev. Genet.* **16**, 583–597 (2015).
- 906 43. Sollier, J. *et al.* Transcription-coupled nucleotide excision repair factors promote R-loop-
907 induced genome instability. *Molecular Cell* **56**, 777–785 (2014).
- 908 44. Cristini, A., Groh, M., Kristiansen, M. S. & Gromak, N. RNA/DNA Hybrid Interactome
909 Identifies DXH9 as a Molecular Player in Transcriptional Termination and R-Loop-
910 Associated DNA Damage. *Cell Rep* **23**, 1891–1905 (2018).
- 911 45. Wang, I. X. *et al.* Human proteins that interact with RNA/DNA hybrids. *Genome Res.* **28**,
912 1405–1414 (2018).
- 913 46. Skourti-Stathaki, K., Kamieniarz-Gdula, K. & Proudfoot, N. J. R-loops induce repressive
914 chromatin marks over mammalian gene terminators. *Nature* **516**, 436–439 (2014).
- 915 47. Watrin, E. *et al.* Human Scc4 is required for cohesin binding to chromatin, sister-
916 chromatid cohesion, and mitotic progression. *Curr. Biol.* **16**, 863–874 (2006).
- 917 48. Irimia, M. *et al.* A Highly Conserved Program of Neuronal Microexons Is Misregulated in
918 Autistic Brains. *Cell* **159**, 1511–1523 (2014).
- 919 49. Pan, H. *et al.* Cohesin SA1 and SA2 are RNA binding proteins that localize to RNA

- 920 containing regions on DNA. *Nucleic Acids Research* **24**, 105–17 (2020).
- 921 50. Chao, W. C. H. *et al.* Structural Studies Reveal the Functional Modularity of the Scc2-
922 Scc4 Cohesin Loader. *Cell Rep* **12**, 719–725 (2015).
- 923 51. Pezic, D. *et al.* The cohesin regulator Stag1 promotes cell plasticity through
924 heterochromatin regulation. *bioRxiv* 1–60 (2021). doi:10.1101/2021.02.14.429938
- 925 52. Li, Y. *et al.* The structural basis for cohesin-CTCF-anchored loops. *Nature* **578**, 472–476
926 (2020).
- 927 53. Yao, H. *et al.* Mediation of CTCF transcriptional insulation by DEAD-box RNA-binding
928 protein p68 and steroid receptor RNA activator SRA. *Genes Dev.* **24**, 2543–2555
929 (2010).
- 930 54. Misteli, T. Beyond the Sequence: Cellular Organization of Genome Function. *Cell* **128**,
931 787–800 (2007).
- 932 55. Nozawa, R.-S. *et al.* SAF-A Regulates Interphase Chromosome Structure through
933 Oligomerization with Chromatin-Associated RNAs. *Cell* **169**, 1214–1227.e18 (2017).
- 934 56. Huo, X. *et al.* The Nuclear Matrix Protein SAFB Cooperates with Major Satellite RNAs to
935 Stabilize Heterochromatin Architecture Partially through Phase Separation. *Molecular*
936 *Cell* **77**, 368–383.e7 (2020).
- 937 57. Ryu, J.-K. *et al.* Phase separation induced by cohesin SMC protein complexes. *bioRxiv*
938 **58**, 142–40 (2020).
- 939 58. Skourti-Stathaki, K. *et al.* R-Loops Enhance Polycomb Repression at a Subset of
940 Developmental Regulator Genes. *Molecular Cell* **73**, 930–945.e4 (2019).
- 941 59. Nadel, J. *et al.* RNA:DNA hybrids in the human genome have distinctive nucleotide
942 characteristics, chromatin composition, and transcriptional relationships. *Epigenetics &*
943 *Chromatin* **8**, 46–19 (2015).
- 944 60. Barrington, C., Georgopoulou, D., Nature, D. P.2019. Enhancer accessibility and CTCF
945 occupancy underlie asymmetric TAD architecture and cell type specific genome
946 topology. *nature.com* doi:10.1038/s41467-019-10725-9
- 947 61. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized
948 p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol*
949 **26**, 1367–1372 (2008).
- 950 62. Choi, M. *et al.* MSstats: an R package for statistical analysis of quantitative mass
951 spectrometry-based proteomic experiments. *Bioinformatics* **30**, 2524–2526 (2014).
- 952 63. Beltran, M. *et al.* The interaction of PRC2 with RNA or chromatin is mutually
953 antagonistic. *Genome Res.* **26**, 896–907 (2016).
- 954

955 METHODS

956 Cell culture and IAA-mediated degradation of Rad21.

957 HCT116 cells with engineered RAD21-miniAID-mClover (RAD21^{mAC}), or OsTIR1-
958 only, or both (RAD21^{mAC}-OsTIR) were obtained from Masato T. Kanemaki.
959 Throughout this study we used RAD21^{mAC}-OsTIR cells, and for simplicity we refer to
960 them in the text as RAD21^{mAC}. The cells were maintained in McCoy's 5A medium
961 with Glutamax (Thermo Fisher Scientific) supplemented with 10% Heat-inactivated
962 FBS (Gibco), 700 μ g/ml Geneticin, 100 μ g/ml Hygromycin B Gold
963 and 100 μ g/ml Puromycin as described. We clonally selected the RAD21^{mAC}-OsTIR
964 cells by sorting green fluorescence positive single cells on a FACS Aria Fusion cell
965 sorter (BD Bioscience). Single cells were individually seeded into one well of a 96-well
966 plate, expanded for 10 days into 6cm culture dishes and selected with Geneticin,
967 Hygromycin B Gold and Puromycin as indicated above in McCoy's medium for another
968 10 days. Each clone was assessed for efficiency of Rad21 degradation using FACS
969 analysis and western blotting (WB) using mClover, mAID and OsTIR antibodies. Two
970 clones (H2 and H11) were taken forward and used throughout this study. To deplete
971 RAD21, RAD21^{mAC}-OsTIR cells were grown in adherent conditions for 3 days and
972 treated with 500 μ M Indole-3-acetic acid (IAA, Auxin, diluted in EtOH) for 4 hours. For
973 IAA withdrawal, IAA treated cells were washed with PBS and replaced with
974 fresh supplemented McCoy's medium for another 4 hours. Cells were washed twice
975 with ice-cold PBS before being harvested for later experimental procedures.

976

977 siRNA-mediated knockdowns.

978 For siRNA transfections, RAD21^{mAC}-OsTIR cells were reverse transfected with
979 scramble siRNA (siCon) or siRNAs targeting SA1, SA2, NIPBL, or AQR (Dharmacon,
980 Horizon Discovery). A final concentration of 10 nM of siSA1, siSA2, or siNIPBL or
981 5 nM of siAQR was reverse transfected into the cells using
982 Lipofectamine RNAiMAX reagent (Invitrogen), as per the manufacturer's
983 instructions. Cells were plated at a density of 1 – 1.25 x 10⁶ cells per 10 cm dish and
984 harvested 72hrs post-transfection, at a confluency of ~70%. The Lipofectamine-
985 containing media was replaced with fresh media 12-16 hrs post-transfection to avoid
986 toxicity. For Figure 5f/g, incubation time was reduced to 40 hrs. To account for the
987 reduced growth time, cells were plated at a density of 2-3 x 10⁶ cells per 10 cm dish.
988 Here siCon- and siNIPBL-transfected cells were plated at a lower cell number
989 than siAQR-transfected cells to ensure equalised confluence (~70%) at the time of
990 collection. When IAA-treatment was combined with siRNA mediated KD, the IAA was
991 added at the end of the normal KD condition so that total KD time was not changed
992 compared to UT cells. For esiRNA treatment, RAD21^{mAC}-OsTIR cells were reverse-
993 transfected with 20 μ M FLUC control esiRNA or esiRNA custom designed to SA1
994 exon31 or SA2 exon32 (MISSION® siRNA, Sigma Aldrich)
995 using RNAiMAX (Invitrogen). Cells were incubated in transfection mixture for 7-8

996 hours before being replaced with fresh supplemented McCoy's medium and left for
997 another 40h until harvest. Efficiency of KD was assessed by WB. siRNA information
998 can be found in Table 1.

999

1000 **Immunofluorescence**

1001 Cells were adhered onto poly L-lysine coated glass coverslips in 6 well
1002 culture dishes and were washed twice with ice-cold PBS before IF procedures. For
1003 RAD21-depletion analysis, cells were fixed for 10 mins at room temperature with 3.7%
1004 paraformaldehyde (Alfa Aesar) in PBS, washed 3 times with PBS and
1005 then permeabilized at room temperature for 10 mins with 0.25% Triton X-100 in PBS
1006 (Sigma Aldrich). For R-loop imaging, cells were fixed and permeabilised with ice-cold
1007 ultra-pure mEtOH (Sigma Aldrich) for 10 mins at -20°C. After 3 washes with PBS, cells
1008 were blocked for 45 mins at room temperature with 10% FCS-PBS. For RNASEH1
1009 enzyme treatment, cells were incubated with blocking solution supplemented with 1x
1010 RNASEH1 reaction buffer alone (50 mM Tris-HCl, 75 mM KCl, 3 mM MgCl₂, 10 mM
1011 DTT) or 5 units of RNASEH1 enzyme (M0297, New England Biolabs) for 30 mins at
1012 37 °C, PBS-washed twice, before blocking. Cells were washed twice with PBS before
1013 incubation with primary antibodies diluted in 5% FCS-PBS at 4 °C overnight. Anti-SA1,
1014 anti-SA2 and anti-AQR were used at 1:3000 dilutions; anti-CTCF was used at 1:2500
1015 dilution; anti-s9.6 was used at 1:1000 dilution; anti-V5 was used at 1:1000. After 4
1016 washes with PBS, cells were incubated with secondary antibodies (donkey anti-Goat
1017 AF555 or AF647 for SA1/2 used at 1:3000; donkey anti-Rabbit AF647 for CTCF used
1018 at 1:2500; donkey anti-Mouse AF555 for s9.6 used at 1:2000; donkey anti-Rabbit
1019 AF647 for AQR used at 1:3000; donkey anti-Rabbit AF488 for V5 used at 1:2000)) in
1020 5% FCS-PBS for 1 hour at room temperature, and washed 4 times with PBS before
1021 being mounted onto glass slides with ProLong™ Diamond Antifade Mountant with
1022 DAPI (Thermo Fisher Scientific) to stabilise overnight in dark before
1023 imaging. See Table 2 for details of where antibodies were purchased.

1024 Imaging was performed on Zeiss LSM confocal microscopes using 63x/1.40
1025 NA Oil Plan-Apochromat objective lens (Carl Zeiss, Inc.). Images were captured as z-
1026 stacks and under consistent digital gain, laser intensity and resolution for each
1027 experiment. Numerical analysis was carried out using Imaris software (Oxford
1028 Instruments, version 9.5.1) and representative images are shown as maximum z-
1029 projected views generated using Fiji Image J. In brief, z-stack images were imported
1030 into Imaris, cells were identified using DAPI and only those located 1 μm away from
1031 image boundary and sized between 120-800 μm³ were selected. A seed-split function
1032 of 7.5um was used to separate closely situated cells. Fluorescence intensities of
1033 individual DAPI-selections in each channel were determined by Imaris and exported
1034 into Excel for further analysis. Distribution plots were generated from >50 cells of each
1035 replicate with 3 biological replicates per experiment. Student's *t*-test was performed
1036 between control and experimental conditions and statistical significance was
1037 determined by detecting the difference between means (unequal variance, two-tailed).

1038 Significance is denoted as $p > 0.05$ = not significant (ns), $p \leq 0.05$ = *, $p \leq 0.005$ = **,
1039 $p \leq 0.0005$ = *** and $p \leq 0.0001$ = ****.

1040

1041 **Chromatin Fractionation and colmunoprecipitation.**

1042 Cells were washed twice with ice-cold PBS (Sigma Aldrich) and lysed in Buffer A (10
1043 mM HEPES, 10mM KCl, 1.5 mM MgCl₂, 0.34 M Sucrose, 10% Glycerol, 1mM DTT,
1044 1mM PMSF/Pefabloc, protease inhibitor), supplemented with 0.1% T-X100, for 10 min
1045 on ice. Lysed cells were collected by scraping. Nuclei and cytoplasmic
1046 material was separated by centrifugation for 4 min at 1300 g at 4oC. The supernatant
1047 was collected as the cytoplasmic fraction and cleared of any insoluble material with
1048 further centrifugation for 15 min at 20,000 g at 4oC. The nuclear pellet was washed
1049 once with buffer A before lysis in buffer B (3mM EDTA, 0.2mM EGTA, 1mM DTT, 1mM
1050 PMSF/Pefabloc, protease inhibitor) with rotation for 30 min at 4oC. Insoluble nuclear
1051 material was spun down for 4 min at 1700 g at 4oC and the supernatant taken as
1052 nuclear soluble fraction. The insoluble material was wash once with buffer B and then
1053 resuspended in high-salt chromatin solubilization buffer (50mM Tris-HCl pH 7.5, 1.5
1054 mM MgCl₂, 500mM KCl, 1mM EDTA, 20% Glycerol, 0.1% NP-40, 1mM
1055 PMSF/Pefabloc, protease inhibitor). The lysate was vortexed for 2 min to aid
1056 solubilization. Nucleic acids were digested with 85U benzonase (Sigma-Aldrich) per
1057 100 x 10⁶ cells, with incubation for 10 min at 37oC and 20 min at 4oC. Chromatin was
1058 further solubilized with ultra-sonication for 3 x 10 sec at an amplitude of 30. The lysate
1059 was diluted to 200 mM KCl and insoluble material was removed by centrifugation at
1060 15,000 RPM for 30 min at 4oC.

1061 For coIP, antibodies were bound to Dynabead Protein A/G beads
1062 (ThermoFisher Scientific) for 10 min at room temperature and ~ 5 hr at 4oC. For mock
1063 IgG IPs, beads were incubated with serum from the same host type as the antibody of
1064 interest. 1mg of chromatin extract was incubated with the antibody-bead conjugate per
1065 IP for approximately 16 hr at 4oC. IPs were washed x5 with IP buffer (200mM
1066 chromatin solubilization buffer) and eluted by boiling in either 2x Laemmli sample
1067 buffer (BioRad) or 4x NuPAGE LDS sample buffer (ThermoFisher Scientific). Proteins
1068 ≤ 250 kDa were separated by SDS-PAGE electrophoresis using 4–20% Mini-
1069 PROTEAN® TGX™ Precast Protein Gels (BioRad) and transferred to Immobilon-P
1070 PVDF Membrane (Merck Millipore) for detection. Proteins ≥ 250 kDa were separated
1071 by SDS-PAGE electrophoresis using Invitrogen NuPAGE 3-8% Tris-Acetate precast
1072 protein gels. Transfer was extended to overnight with low voltage (20V) to aid in
1073 transfer of the high-molecular weight proteins. Membranes were incubated in primary
1074 antibody solution overnight at 4oC and images were detected using chemiluminescent
1075 fluorescence. Densitometry was carried out using ImageStudio Lite software with
1076 statistical significance calculated by unpaired t test, unless otherwise specified. Fold
1077 enrichment quantifications were performed by first normalising the raw densitometry
1078 value to its corresponding Histone H3 quantification and the comparing between the
1079 samples indicated. See Table 2 for details of antibodies.

1080

1081 **S9.6 IP and Dot Blot.**

1082 Cells were fractioned and processed for S9.6 IP as described above, with the following
1083 modifications. To avoid digestion of RNA:DNA hybrids, samples were not treated
1084 with benzonase during chromatin solubilization and sonication was carried out for 10
1085 min (Diagenode Biorupter) as in ⁴⁴. Where indicated, chromatin samples were treated
1086 with Ribonuclease H enzyme (NEB) overnight at 37°C to digest RNA:DNA hybrids in
1087 the extract. To avoid detection of single-stranded RNA by the S9.6 antibody, all S9.6
1088 IP samples were pre-treated with Purelink RNase A (Thermo Fisher Scientific) at
1089 0.25ug/1mg chromatin extract for 1 hr 30 min at 4°C. The reaction was stopped with
1090 addition of 143U Invitrogen SUPERase•In RNase Inhibitor (Thermo Fisher
1091 Scientific). RNA:DNA hybrid levels were assessed in chromatin samples by dot blot.
1092 Specifically, the chromatin lysate was directly wicked
1093 onto Amersham Protran nitrocellulose membrane (Merck) by pipetting small volumes
1094 above the membrane. Membranes were blocked in 5% (w/v) non-fat dry milk in PBS-
1095 0.1% Tween and incubated with S9.6 antibody overnight as for standard western blot.
1096 As above, detection was carried out using chemiluminescent fluorescence. RNase A-
1097 mediated digestion of RNA:DNA hybrids was performed using a non-ssRNA-specific
1098 enzyme (Thermo Scientific) at 1.5ug/25ug chromatin extract at 37°C.

1099

1100 **ChIP-sequencing, library preparation and analysis.**

1101 ChIP lysates were prepared from RAD21^{mAC} cells treated with EtOH or IAA for 6hrs in
1102 two biological replicates. Formaldehyde (1%) was added to the culture medium for
1103 10min at room temperature. Fixation was blocked with 0.125M glycine and cells were
1104 washed in cold PBS. Nuclear extracts were prepared by douncing (20 strokes, medium
1105 pestle) in swelling buffer (25 mM HEPES pH8, 1.5 mM MgCl₂, 10mM KCL, 0.1%
1106 NP40, 1 mM DTT and protease inhibitors) and centrifuged for 5min at 2000rpm at
1107 4C. Nuclear pellets were resuspended in Sucrose buffer I (15mM Hepes pH 8, 340
1108 mM Sucrose, 60mM KCL, 2mM EDTA, 0.5 mM EGTA, 0.5% BSA, 0.5 mM DTT and
1109 protease inhibitors) and dounced again with 20 strokes. The lysate was carefully laid
1110 on top of an equal volume of Sucrose buffer II (15mM Hepes pH 8, 30% Sucrose,
1111 60mM KCL, 2mM EDTA, 0.5 mM EGTA, 0.5 mM DTT and protease inhibitors) and
1112 centrifuged for 15min at 4000rpm at 4C. Nuclei were washed twice to remove
1113 cytoplasmic proteins, centrifuged and the pellet was resuspended in
1114 Sonication/RIPA buffer (50mM Tris, pH 8.0, 140 mM NaCl, 1 mM EDTA, 1% Triton X-
1115 100, 0.1% Na-deoxycholate, 0.1% SDS and protease inhibitors) at a concentration of
1116 5 x10⁶ nuclei in 130ul buffer. This was transferred to a sonication tube (AFA Fiber
1117 Pre-Slit Snap-Cap 6x16mm) and sonicated in a Covaris S2 (settings; 4 cycles of 60
1118 seconds, 10% duty cycle, intensity: 5, 200 cycles per burst). Soluble chromatin was
1119 in the range of 200 - 400 bp. Triton X100 was added (final concentration 1%) to the
1120 sonicated chromatin and moved to a low-retention tube (Eppendorf) before

1121 centrifugation at 14,000 rpm for 15min at 4C and pellets were discarded. 1/100th of the
1122 chromatin lysate was retained as the Input sample.

1123 For Immunoprecipitation, 200ug chromatin aliquots/IP were precleared with a
1124 slurry of Protein A/G (50:50) (Dynabeads) and incubated for 4hr at 4C. Meanwhile,
1125 washed Protein A/G beads (40ul per IP) were mixed with primary antibodies and
1126 incubated for 4hrs at 4C. The following amounts of antibodies were used: anti-
1127 CTCF, 5ug/ChIP; anti-SA1, 15ug/ChIP; anti-SA2, 10ug of the mixed antibody
1128 pack/ChIP; anti-Smc3, 5ug/ChIP and anti-IgG, 10ug/ChIP. See Table 2 for information
1129 about the antibodies. Washed, pre-bound Protein A/G beads+antibody were mixed
1130 with pre-cleared chromatin lysates and incubated overnight with rotation at 4C. The
1131 next day, the supernatant was removed and the beads were washed 9 times with
1132 increasing salt concentrations. Protein-DNA crosslinks were reversed in ChIP elution
1133 buffer (1% SDS, 5 mM EDTA, 10 mM Tris HCl pH 8) + 2.5 ul of Proteinase K and
1134 incubated for 1 hour at 55°C and overnight at 65°C. Samples were phenol-chloroform
1135 extracted, resuspended in TE buffer and assessed by qPCR as a quality
1136 control. Libraries were prepared from 5-10ng of purified DNA, depending on
1137 availability of material, using NEBNext Ultra II DNA Library Prep Kit for Illumina kit and
1138 using NEBNext Multiplex Oligos for Illumina (Index Primers Set 2) according to
1139 manufacturer's instructions using 6-8 cycles of PCR. ChIP-seq libraries from one
1140 biological set (all ChIP libraries for both EtOH and
1141 IAA) were multiplexed and sequenced on the Illumina HiSeq2500 platform, 80bp
1142 single-end reads. Each biological set was sequenced on a separate run.

1143 Quality control of reads was performed using FASTQC. Reads were aligned to
1144 the hg19 reference genome using Bowtie with 3 mismatches. PCR duplicates were
1145 detected and removed using SAMTOOLS. Bam files were imported into MISHA (v
1146 3.5.6) and peaks were identified using a 0.995 percentile. Peaks that overlapped in
1147 both replicates were retained. Only replicate 1 of the SA1 library was used. Correlation
1148 plots of peaks across the genome from different ChIP libraries were compared with
1149 log-transformed percentiles plotted as a smoothed scatter plot. Comparison of peaks
1150 at regions of interest were carried out using deepTools (Version 3.1.0-2). For input
1151 into deepTools, peak data was converted to bigwig format, with a bin size of 500, using
1152 the UCSC bedGraphToBigWig package. The signal matrix was calculated for a window
1153 2,000 bp up- and down-stream of the region of interest, missing data was treated as
1154 zero, and all other parameters were as default. Heatmaps were generated
1155 within deepTools, with parameters as default. Read density profile plots were plotted
1156 in ggplot using deepTools profilePlot -perGroup data and smoothed using
1157 geom_smooth default 'gam' settings.

1158

1159 **DRIP-sequencing.**

1160 DRIP lysates were prepared from chromatin. Chromatin was fractionated as described
1161 for ChIP samples above, with the following changes. Samples were not fixed and were
1162 collected from the plate by scraping in ice-cold PBS. Sonication was performed to

1163 solubilise the chromatin using a picorupter with 10 cycles of 30 sec on, 30 sec off.
1164 Following sonication, 60U of RNase I (Ambion) was incubated with each sample for
1165 1.5hrs at 37°C to reduce off-target RNA binding by S9.6. Protein was digested with
1166 184ug of proteinase K incubated with each sample for 2.5hrs at 45°C and 3.5hrs at
1167 55°C. Samples were spun briefly and the supernatant taken. Nucleic acid material was
1168 isolated from the sample by phenol-chloroform extraction and isopropanol
1169 precipitation. Purified nucleic acid material was IP'd as in ⁵⁸, including resuspension
1170 nuclear lysis buffer and fragmentation by sonication using a picorupter for 4 cycles of
1171 30sec on, 30 sec off. Following elution, reverse crosslinking, and proteinase K
1172 treatment the immunoprecipitated nucleic acid was purified by phenol-chloroform
1173 extraction and isopropanol precipitation. Second strand synthesis was carried out
1174 according to ⁵⁹ with minor changes. Namely, the reaction was set up in a PCR tube
1175 as: ~28ul eluted DRIP sample, ~27ul nuclease-free water (depending on DRIP
1176 sample, for a total of 75ul), 15ul 5X ss buffer, 2ul 10mM dNTP, 0.5ul DNA ligase (E.
1177 coli, NEB cat. M0205S), 2ul DNA polymerase I (E.coli, NEB cat. M0209S), 0.5ul
1178 RNase H (2.5 units). This was incubated for 2hrs at 16°C. dsDNA was purified using
1179 a 1.8X ratio of SPRI beads and eluted in ultrapure water. Libraries were prepared
1180 using NEB Ultra II DNA library prep kit according to manufacturer's instructions and
1181 sequenced on the Illumina NovaSeq platform, 100bp paired-end reads (30M per
1182 sample). Reads were aligned and processed as for the ChIP-seq samples above with
1183 two changes; i) reads were quality trimmed using trimfq -b 14 -e 16, and ii) a maximum
1184 insert size of 1000 was set for bowtie. DeepTools analysis was carried out as for
1185 ChIP-seq samples and with binSize set to 2bp.

1186

1187 **ChromHMM.**

1188 ChIP-seq data for YY1, CBX3, SIN3A, POLR2A, POLR2AphosphoS5, H3K27ac,
1189 H3K4me3, H3K4me1, H3K27me3, EZH2, and H3K9me3 from HCT116 cells were
1190 obtained from ENCODE and processed as above. ChIP-seq, DRIP-seq, and ENCODE
1191 ChIP-seq. BAM files were binarized in ChromHMM using a bin size of 200 bp and a
1192 shift of 150 bp. Where input files were available on ENCODE they were used in
1193 ChromHMM to determine the binarization threshold, otherwise the ChromHMM default
1194 of a uniform background was assumed. The chromatin state model was generated for
1195 15 states and compared to the hg19 genome assembly. All other parameters were as
1196 default.

1197

1198 **Hi-C data and contact hotspots analysis.**

1199 Generating hotspots - Previously published Hi-C datasets derived from
1200 RAD21^{mAC} cells treated with EtOH or IAA ⁸ were analyzed as previously described
1201 ⁶⁰. Custom R scripts were written to identify Hi-C hotspots, i.e. regions of Hi-C maps
1202 with high contact frequency. To begin, for each chromosome, all contacts were
1203 extracted and subsetted for only high scoring (>=60) contacts between a band of 10e3
1204 – 70e6. Using KNN, for each high scoring contact, the 250 nearest neighbour contacts

1205 were identified and subset for only the high-scoring neighbours. This created a list of
1206 high scoring neighbours for each high scoring contact, where the first neighbour is the
1207 contact itself with a distance of 0. This allowed the neighbour information to be
1208 converted into edge information, thereby allowing high score fend contacts to be
1209 grouped into cluster hotspots using the R package 'igraph'. Hotspots that contained
1210 less than the minimum number of high scoring fends (<100) were removed. The output
1211 list of hotspots were represented as 2D intervals which contained high scoring
1212 contacts. In total, 5539 hotspots were identified in EtOH and 759 in IAA Hi-C data.

1213 Creating aggregate plots - To calculate and visualise the contact enrichment at
1214 hotspots in the EtOH and IAA Hi-C, we used the R package 'shaman'. Firstly, we used
1215 the function 'shaman_generate_feature_grid' to calculate the enrichment profile at
1216 EtOH and IAA hotspots. Using the weighted centre for each hotspot, represented as
1217 a 2D interval we used the function to build grids for the EtOH and IAA hotspots in the
1218 HiC data at 3 specific bands, 100k – 1MB, 1MB – 5MB, 5MB – 10MB. A range of
1219 250kb was visualised around the weighted centre. The grid was built by taking all
1220 combinations interval1 and interval2 of the EtOH and IAA hotspot centres, with each
1221 combination termed a 'window'. Hotspots were not filtered for size or shape. A score
1222 threshold of 60 was used to focus on enriched pairs, those windows that did not
1223 contain at least one point with a score of 60 were discarded. Each window was then
1224 split into 1000nt bins and the windows were summed together to generate a grid
1225 containing the observed and expected contacts. We visualised the grid using
1226 'shaman_plot_feature_grid' using 'enrichment' mode and a plot_resolution value of
1227 6000, due to the large range being visualised.

1228

1229 **STORM – Immunolabelling and imaging.**

1230 Two clones of RAD21^{mAC}-OsTIR cells were seeded at a density of 30,000 cells per
1231 well per 400ul) onto poly-L-lysine coated 8-well chamber slides (Lab-Tek™ 155411)
1232 overnight. Each clone was treated with EtOH, IAA or IAA washoff and then fixed with
1233 PFA 4% (Alfa Aesar) for 10 min at room temperature and rinsed with PBS three times
1234 for 5 min each. The cells were shipped to the Cosma Lab after fixation for STORM
1235 processing and imaging. Cells were permeabilized with 0.3% Triton X-100 in PBS
1236 and blocked in blocking buffer (10% BSA – 0.01 % Triton X-100 in PBS) for one hour
1237 at room temperature. Cells were incubated with primary antibodies (see Table 2) in
1238 blocking buffer at 1:50 dilution. For combined S9.6/STAG1 and S9.6/RAD21-GFP
1239 imaging, cells were incubated with primary antibodies in blocking buffer (dilutions
1240 1:100 for S9.6 and STAG1, 1:250 for GFP). Cells were washed three times for 5 min
1241 each with wash buffer (2% BSA – 0.01 % Triton X-100 in PBS) and incubated in
1242 secondary antibody. For STORM imaging, home-made (Bates et al., 2007) dye pair
1243 labeled secondary antibodies were added at a 1:50 dilution in blocking buffer and were
1244 incubated for 45 min at room temperature or single fluorophore labeled commercial
1245 antibodies were added at a 1:250 dilution in blocking buffer and were incubated for 45

1246 min at room temperature (see Table 2). Cells were washed three times for 5 min each
1247 with wash buffer.

1248 STORM imaging was performed on an N-STORM 4.0 microscope (Nikon)
1249 equipped with a CFI HP Apochromat TIRF 100x 1.49 oil objective and an iXon Ultra
1250 897 camera (Andor) and using Highly Inclined and Laminated Optical sheet
1251 illumination (HILO). Dual color STORM imaging was performed with a double activator
1252 and single reporter strategy by combining AF405_AF647 anti-Goat secondary with
1253 Cy3_AF647 anti-Rabbit secondary antibodies. Sequential imaging acquisition was
1254 performed (1 frame of 405 nm activation followed by 3 frames of 647 nm reporter and
1255 1 frame of 560 nm activation followed by 3 frames of 647 nm reporter) with
1256 10 ms exposure time for 120000 frames. 647 nm laser was used at constant ~ 2
1257 kW/cm^2 power density and 405 nm and 560 nm laser powers were gradually increased
1258 over the imaging. For S9.6 experiments, before STORM imaging, conventional images
1259 were taken for s9.6 signal (AF568 labeled) with a TRITC filter, for endogenous
1260 mClover signal with a FITC filter and for STAG1 or GFP (AF647 labeled) with a
1261 Quadband filter. STORM imaging for STAG1 or GFP was performed with continuous
1262 imaging acquisition (i.e. simultaneous stimulation with 405 and 647 nm lasers) with 10
1263 ms exposure time for 60000 frames. 647 nm laser was used at constant ~ 2 kW/cm^2
1264 power density and 405 nm was gradually increased over the imaging. Imaging buffer
1265 composition for STORM imaging was 100 mM Cysteamine MEA (Sigma-Aldrich,
1266 #30070) - 5% Glucose (Sigma-Aldrich, #G8270) – 1% Glox Solution (0.5 mg/ml
1267 glucose oxidase, 40 mg/ml catalase (Sigma-Aldrich, #G2133 and #C100)) in PBS.

1268

1269 **STORM imaging analysis and quantifications.**

1270 STORM images were analyzed and rendered with Insight3 software (kind gift of Bo
1271 Huang, UCSF, Huang et al., 2008) as previously described (Bates et al., 2007; Rust
1272 et al., 2006). Localizations were identified based on a threshold and fit to a simple
1273 Gaussian to determine the x and y positions. Cluster analysis of CTCF, SA1 and SA2
1274 STORM signal was performed as previously described (Ricci et al., 2015) to obtain
1275 cluster size and positions and to measure Nearest
1276 Neighbour distributions (NND) between clusters of the same protein in individual
1277 nuclei. NND between clusters' centroids of two different proteins (i.e. CTCF-SA1 and
1278 CTCF-SA2) was calculated by knnsearch.m Matlab function and the NND histogram
1279 of experimental data was obtained by considering all the NNDs of individual nuclei
1280 (histogram bin, from 0 to 500 nm, 5 nm steps). Simulated NNDs recapitulating random
1281 spatial distribution of cluster centroids were first obtained for each nucleus separately
1282 and then merged to calculate the simulated NND histogram (histogram bin, from 0 to
1283 500 nm, 5 nm steps). The difference plot reports the difference between experimental
1284 NND and simulated NND. Quantification and analysis of STORM
1285 images was performed in Matlab and statistical analysis was performed
1286 in Graphpad Prism (v7.0e). The type of statistical test is specified in each case.
1287 Statistical significance is represented as indicated above.

1288 Analysis for S9.6 experiments was performed in the following way. After
1289 generating localization lists for each STORM image, nuclear masks and S9.6 were
1290 generated to segment the obtained localizations belonging to nuclear areas inside or
1291 outside s9.6 enriched regions. Masks generation, quantification of masks' areas and
1292 segmentation of STORM localizations were performed in Fiji/ImageJ. Nuclear masks
1293 were manually designed based on STAG1/GFP signal. S9.6 masks were generated
1294 by applying an automatic threshold on s9.6 images based on nuclear intensity signal.
1295 Masks were visually inspected individually and adjusted manually in cases where dim
1296 signal or noise from cytosolic signal compromised the identification of the mask. The
1297 area of all masks was calculated. Masks were applied to the STORM localization lists
1298 to generate segmented lists with the localizations belonging to the entire nucleus and
1299 to the s9.6 enriched areas. Finally, the density of localizations (number of
1300 localizations/area of the mask) for the areas inside and outside s9.6 masks was
1301 calculated and the ratio between both values was plotted.

1302 Graphpad Prism software used for statistical analysis can be found
1303 at: <https://www.graphpad.com/scientific-software/prism/> MatLab software used for
1304 imaging data analysis can be found
1305 at: <https://www.mathworks.com/products/matlab.html>

1306

1307 **Mass spectrometry sample preparation and running.**

1308 SA1 immunoprecipitation samples were analysed by liquid chromatography–tandem
1309 mass spectrometry (LC-MS/MS). Three biological replicate experiments were carried
1310 out for MS and each included four samples, untreated (UT), treated with IAA for
1311 4hrs, siCon, or siSA1, generated as described above. Cells were fractionated to purify
1312 chromatin-bound proteins as above and immunoprecipitated with IgG- or SA1-bead
1313 conjugates. To maximise IP material for the MS, the antibody amount was increased
1314 to 15ug and the chromatin amount was increased to 2mg.

1315 The IP eluates were loaded into a pre-cast SDS-PAGE gel (4–20% Mini-
1316 PROTEAN® TGX™ Precast Protein Gel, 10-well, 50 μ L) and proteins were run
1317 approximately 1 cm to prevent protein separation. Protein bands were excised and
1318 diced, and proteins were reduced with 5 mM TCEP in 50 mM triethylammonium
1319 bicarbonate (TEAB) at 37°C for 20 min, alkylated with 10 mM 2-chloroacetamide in 50
1320 mM TEAB at ambient temperature for 20 min in the dark. Proteins were then digested
1321 with 150ng trypsin, at 37°C for 4 h followed by a second trypsin addition for 4 h, then
1322 overnight at room temperature. After digestion, peptides were extracted with
1323 acetonitrile and 50 mM TEAB washes. Samples were evaporated to dryness at 30°C
1324 and resolubilised in 0.1% formic acid.

1325 nLC-MS/MS was performed on a Q Exactive Orbitrap Plus interfaced to a
1326 NANOSPRAY FLEX ion source and coupled to an Easy-nLC 1200 (Thermo Scientific).
1327 25% (first, second and fourth biological replicate) or 50% (third biological replicate) of
1328 each sample was loaded as 5 or 10 μ L injections. Peptides were separated on a 27cm
1329 fused silica emitter, 75 μ m diameter, packed in-house with Reprosil-Pur 200 C18-AQ,

1330 2.4 μm resin (Dr. Maisch) using a linear gradient from 5% to 30% acetonitrile/ 0.1%
1331 formic acid over 60 min, at a flow rate of 250 nL/min. Peptides were ionised by
1332 electrospray ionisation using 1.8 kV applied immediately prior to the analytical column
1333 via a microtee built into the nanospray source with the ion transfer tube heated to
1334 320°C and the S-lens set to 60%. Precursor ions were measured in a data-dependent
1335 mode in the orbitrap analyser at a resolution of 70,000 and a target value of 3e6 ions.
1336 The ten most intense ions from each MS1 scan were isolated, fragmented in the HCD
1337 cell, and measured in the orbitrap at a resolution of 17,500.

1338

1339 **Mass spectrometry analysis**

1340 Raw data was analysed with MaxQuant⁶¹ version 1.5.5.1 where they were searched
1341 against the human UniProtKB database using default settings
1342 (<http://www.uniprot.org/>). Carbamidomethylation of cysteines was set as fixed
1343 modification, and oxidation of methionines and acetylation at protein N-termini were
1344 set as variable modifications. Enzyme specificity was set to trypsin with maximally 2
1345 missed cleavages allowed. To ensure high confidence identifications, PSMs, peptides,
1346 and proteins were filtered at a less than 1% false discovery rate (FDR). Label-free
1347 quantification in MaxQuant was used with LFQ minimum ratio count set to 2 with
1348 'FastLFQ' (LFQ minimum number of neighbours = 3, and LFQ average number of
1349 neighbours = 6) and 'Skip normalisation' selected. In Advanced identifications,
1350 'Second peptides' was selected and the 'match between runs' feature was not
1351 selected. Statistical protein quantification analysis was done in MSstats⁶²(version
1352 3.14.0) run through RStudio. Contaminants and reverse sequences were removed
1353 and data was log₂ transformed. To find differential abundant proteins across
1354 conditions, paired significance analysis consisting of fitting a statistical model and
1355 performing model-based comparison of conditions. The group comparison function
1356 was employed to test for differential abundance between conditions. Unadjusted p-
1357 values were used to rank the testing results and to define regulated proteins between
1358 groups.

1359 Proteins with peptides discovered in the IgG samples were disregarded from
1360 downstream analyses. Significantly depleted/enriched proteins were considered with
1361 an absolute log₂foldchange > 0.58 (1.5-fold change) and a p-value < 0.1. SA1
1362 interactome analysis was performed in STRING. The network was generated as a full
1363 STRING network with a minimum interaction score of 0.7 required. Over-enrichment
1364 of GO biological process and molecular function terms was calculated with the human
1365 genome as background. Network analysis of the SA1 interactome in IAA-treated
1366 samples was generated from the significantly depleted/enriched proteins, with a
1367 minimum interaction score of 0.4 required. Two conditions for functional enrichments
1368 were considered; i) enrichment was calculated with the human genome as background
1369 to determine the full SA1 interactome in the absence of cohesin, and ii) enrichment
1370 was calculated with the untreated SA1 interactome as background, to determine the
1371 statistical effect of cohesin loss of the SA1 interactome itself. The network developed

1372 in i) was manually rearranged in Cytoscape for visual clarity, enriched categories were
1373 visualized using the STRING pie chart function and half of the proteins within each
1374 category were subset from the network based on pvalue change between UTR and
1375 IAA samples.

1376 Over-enrichment of the s9.6 interactome was calculated separately using the
1377 hypergeometric distribution for comparison with ^{44,45}. Significance was calculated
1378 using the dhyper function in R and multiple testing was corrected for using the p.adjust
1379 Benjamini & Hochberg method. To compare with a minimal background protein list,
1380 <http://www.humanproteomemap.org> was analysed on the Expression Atlas database
1381 to determine a list of proteins expressed in one or more of three tissue types
1382 corresponding to the cell types used across the different studies.

1383

1384 **SLiMSearch analysis**

1385 The SLiMSearch tool <http://slim.icr.ac.uk/slimsearch/>, with default parameters was
1386 used to search the human proteome for additional proteins that contained the FGF-
1387 like motif determined in ¹⁸ to predict binding to SA proteins. The motif was input as
1388 [PFCAVIYL][FY][GDEN]F.{0,1}[DANE].{0,1}[DE]. Along with CTCF, four proteins
1389 found to contain the FGF-like motif, CHD6, MCM3, HNRNPUL2 and ESYT2 were
1390 validated for interaction with SA.

1391

1392 **CLIP**

1393 Crosslinking immunoprecipitation (CLIP) was performed as previously described⁶³.
1394 Briefly, mESC or HCT116 cells were irradiated with 0.2 J/cm² of 254 nm UV light in a
1395 Strat linker 2400 (Stratagene). Cells were lysed in 1 ml of lysis buffer with Complete
1396 protease inhibitor (Roche). Lysates were passed through a 27 G needle, 1.6 U DNase
1397 Turbo (ThermoFisher) per 10⁶ cells and 0.8 (low) or 8 U (high) U RNase I (Ambion) per
1398 10⁶ cells added, and incubated in a thermomixer at 37°C and 1100 rpm for 3 minutes.
1399 Lysates were then cleared by centrifugation and using Proteus clarification spin
1400 column, according to the manufacturer's instructions. Endogenous SA1 and SA2 were
1401 immunoprecipitated with 10 µg SA1 and SA2 antibodies or non-specific IgG control
1402 (Sigma) conjugated to protein G dynabeads (Dyna) for 4 hrs at 4°C. Tagged SA2
1403 proteins were immunoprecipitated from HCT116 cells 40 hours after transfection with
1404 30 µl GFP-Trap beads. IPs were washed three times with high salt buffer (containing
1405 1M NaCl and 1M urea) and once with PNK buffer and RNA labelled with 8 µl
1406 radioactive ³²P-gamma-ATP (Hartmann Analytic) for 5 mins at 37°C. RNPs were
1407 eluted in LDS loading buffer (Invitrogen) and resolved on a 4-12% gradient NuPAGE
1408 Bis-Tris gel (Invitrogen) and transferred onto 0.2 µm diameter pore nitrocellulose
1409 membrane. After blocking with PBST+milk, membranes were washed and exposed
1410 overnight to phosphorimager screen (Fuji) and RNA-³²P visualized using a Typhoon
1411 phosphorimager (GE) and ImageQuant TL (GE). Membranes were then
1412 immunoblotted for SA1, SA2, and RAD21 and visualized using an ImageQuantLAS
1413 4000 imager (GE). See Table 2 for details on antibodies.

1414

1415 **GFP-TRAP + Cloning of STAG2 isoforms and YFP constructs.**

1416 SA2 cDNAs were cloned directly from HCT116 cells by PCR using KAPA
1417 HiFi HotStart PCR kit (Roche) (Fwd: ATGATAGCAGCTCCAGAAAACCAACTG; Rev:
1418 TTA AACATTGACACTCCAAGAACTGATTCATCC). Two major isoforms were
1419 detected, SA2^{Δex32} where exon32 has been spliced out and SA2^{+ex32} where exon
1420 32 has been spliced in. Both SA2 cDNAs were cloned into pENTR/D vector
1421 (Invitrogen) and then into an N-terminal YFP-tagged Gateway cloning vector (a kind
1422 gift from Endre Kiss-Toth, University of Sheffield). Sequences were confirmed by
1423 restriction enzyme digestion and Sanger sequencing. Recombinant YFP-SA2^{Δex32} or
1424 YFP-SA2^{+ex32} were transfected into adherent HCT116 cells for 40 hours before being
1425 harvested. Cells were lysed and fractionated as indicated for CLIP. One third of
1426 the whole cell lysate was pre-cleared with a 50:50 mixture of protein A/G magnetic
1427 beads and GFP-Trap (Chromotek, gtd-20) was pre-blocked with 1mg/mL ultra-pure
1428 BSA (AM2616, Invitrogen) for 2h at 4°C. After blocking, GFP-Trap was washed twice
1429 with CLIP lysis buffer and added to pre-cleared lysates to immunoprecipitate proteins
1430 for 1h at 4°C. Samples were washed in high salt buffer (50mM Tris-HCL pH 7.4, 1M
1431 NaCl, 1mM EDTA, 1% NP-40, 0.1% SDS, 0.5% sodium deoxycholate, 1M Urea) and
1432 low salt PNK buffer (20mM Tris-HCL pH 7.4, 10mM MgCl₂, 0.2% Tween-20), and
1433 eluted in 2x Laemmli buffer (Bio-Rad). Proteins were separated by SDS-PAGE on a
1434 4-20% gradient mini-PROTEAN® Precast Gel (Bio-Rad) and transferred onto PVDF
1435 membrane for visualization.

1436

1437 **VAST-TOOLS**

1438 VAST-TOOLS was used to generate Percent Spliced In (PSI) scores, a statistic which
1439 represents how often a particular exon is spliced into a transcript using the ratio
1440 between reads which include and exclude said exon. Paired-end RNA-seq datasets
1441 were submitted to VAST-TOOLS (v2.1.3) using the Mmu genome (Tapial J et al, Gen
1442 Res 2017). Briefly, reads are split into 50nt words with a 25nt sliding window. The 50nt
1443 words are aligned to a reference genome using Bowtie to obtain unmapped reads.
1444 These unmapped reads are then aligned to a set of predefined exon-exon junction
1445 (EJJ) libraries allowing for the quantification of alternative exon events. The output
1446 was further interrogated using a script which searches all hypothetical EEJ
1447 combinations between potential donors and acceptors within Stag1. PSI scores could
1448 be obtained providing there was at least a single read within the RNAseq data that
1449 supported the event, although we only considered events supported by a minimum
1450 of 50 reads. Calculated PSI values for each alternatively spliced exon as well as
1451 the average PSI reported in the text are shown below. See Supp Fig 5 for names of
1452 published datasets used in this analysis.

1453 **Table 1. siRNAs used in this study.**

siRNA name	Company	Target	Catalogue no.	custom siRNA sequence
si scramble control	Dharmacon	Smartpool	D-001810-10-05	
siSA1	Dharmacon	Smartpool	L-010638-01-0010	
siSA2	Dharmacon	Smartpool	L-021351-00-0010	
siNIPBL	Dharmacon	Smartpool	L-012980-00-0010	
siAQR	Dharmacon	Smartpool	L-022214-01-0005	
esi control	Sigma	Luciferase	EHUFLUC	
esi SA1	Sigma	exon 31	custom esiRNA	TCCTCAGATGCAGATCTCTTGGTTAGGCC AGCCGAAGTTAGAAGACTTAAATCGGAAG GACAGAACAGGAATGAACTACATGAAAGTG AGAAGTGGAGTGAGGCATGCTGT
esi SA2	Sigma	exon 32	custom esiRNA	CACGCAGGTAACATGGATGTTAGCTCAAAG ACAACAAGAGGAAGCAAGGCAACAGCAGG AGAGAGCAGCAATGAGCTATGTTAAACTG CGAACTAATCTTCAGCATGCCAT

1454

1455

1456

Table 2. Antibodies used in this study.

Protein	Company	Catalogue No.	Species	Figure Reference
SA1	Abcam	ab4455	mouse	1a, c, d, e, 2a, b, d, 3a, b, e, 4g, 5b, f, h, i
SA1	Abcam	ab4457	mouse	1i
SA2	Bethyl	A300-159	goat	1b, c, d, e, 2a, 3a, b, f, 4g, e, 5c, g
SA2	Bethyl, AbVantage Pack	A310-941A	goat	1i
CTCF	Diagenode	C15410210	rabbit	1c, d, i, 2a
CTCF	Cell signalling	2899s	rabbit	1a, e
RAD21	Abcam	ab992	rabbit	1c, d, i, 2a, d, 4e, g, j
GFP-TRAP	ChromoTek	gtd-20		1i, 3g
GFP	Invitrogen	A11122	rabbit	1a, e
mAID	MBL	M214-3	mouse	S1a
OsTIR	MBL	PD048	rabbit	S1a
SMC3	Abcam	ab9263	rabbit	1i
CHD6	Bethyl	A301-221A	rabbit	2a
MCM3	Bethyl	A300-124A	goat	2a, 5b
HNRNPUL2	Abcam	ab195338	rabbit	2a
YTHDC1	Abcam	ab122340	rabbit	2d
FTSJ3	Bethyl	A304-199A-M	rabbit	2d
FANCI	Bethyl	A301-254A-M	rabbit	2d
TAF15	Abcam	ab134916	rabbit	2d
DHX9	Abcam	Ab26271	rabbit	2d, 5b
SSRP1	Abcam	ab26212	mouse	2d
INO80	Proteintech	18810-1-AP	rabbit	2d
ESYT2	Sigma-Aldrich	HPA002132	rabbit	2d
S9.6	Kerafast	ENH001	mouse	5b-j
RNASE H2	Novus	NBP1-76981	rabbit	5b
AQR	Bethyl	A302-547A	rabbit	5b, e
POLR2	Covance	MMS-1289	mouse	5b
MAU2	Abcam	ab183033	rabbit	4e, g, f
NIPBL	Abbiotec	250133	rat	4e, g, f, S5
H3	Abcam	ab1791	rabbit	3e, f, 4e, g, f

Name (Secondary Abs)	Fluorophore	Company	Catalogue No.	Figure Reference
Donkey anti-Rabbit	Cy3_AF647	Home made from Jackson ImmunoResearch IgG	Home made from 711-005-152	1e, S1
Donkey anti-Goat	AF405_AF647	Home made from Jackson ImmunoResearch IgG	Home made from 705-005-147	1e, S1
Donkey anti-mouse	AF647	Invitrogen	A31570	1a, d, e
Donkey anti-rabbit	AF488	Invitrogen	A21206	1a, d, e
Donkey anti-rabbit	AF647	Invitrogen	A31573	1a, d, e
Donkey anti-goat	AF555	Invitrogen	A21432	1a, d, e
Donkey anti-goat	AF647	Invitrogen	A21447	1a, d, e
Goat anti-Mouse	AF568	ThermoFisher Scientific	A-11031	
Goat anti-Rabbit	AF647	ThermoFisher Scientific	A-21244	
Rabbit anti-Goat	AF647	ThermoFisher Scientific	A-21446	

1457
1458
1459

Table 3. Published datasets used in this study.

Accession no.	Analysis description	Publication DOI or Ref	Figure Reference
GSE104334	Long-range contact analysis of Hi-C datasets	10.1016/j.cell.2017.09.026	1i
GSE89729	Percent Spliced In (PSI) analysis of RNA-seq datasets	10.1172/jci.insight.91419	4d, "HCT Zuo"
GSM958749	Percent Spliced In (PSI) analysis of RNA-seq datasets	ENCODE HCT116 RNAseq	4d, "HCT ENCODE"
GSM958735	Percent Spliced In (PSI) analysis of RNA-seq datasets	ENCODE HeLa RNAseq	4d, "HeLa"

1460

1461 **Table 4. Published ChIP-seq datasets used for ChromHMM.**

Protein	Accession	Publication	Matched input
NIPBL (EtOH- and IAA-treated)	GSE104334	(Rao <i>et al.</i> , 2017)	-
CBX1	GSM1010758	(Gertz <i>et al.</i> , 2013)	
EZH2	GSM3498250	(Dunham <i>et al.</i> ,	GSM2308475;GSM2308476
POLR2A	GSM935426	(Dunham <i>et al.</i> ,	GSM2308422
POLR2AphosphoS5	GSM803474	(Gertz <i>et al.</i> , 2013)	GSM803475
SIN3A	GSM1010905	(Gertz <i>et al.</i> , 2013)	
YY1	GSM803354	(Gertz <i>et al.</i> , 2013)	GSM803475
H3K4me1	GSM945858		GSM2308475; GSM2308476
H3K4me1	GSM2527549	(Dunham <i>et al.</i> ,	GSM2308422
H3K4me3	GSM2533929	(Dunham <i>et al.</i> ,	GSM2308475; GSM2308476
H3K4me3	GSM945304	(Thurman <i>et al.</i> ,	GSM945287
H3K9me3	GSM2527565	(Dunham <i>et al.</i> ,	

H3K9me3	GSM2308431	(Dunham <i>et al.</i> ,	
H3K27ac	GSM2534277	(Dunham <i>et al.</i> ,	GSM2308422
H3K27me3	GSM2308612	(Dunham <i>et al.</i> ,	

1462

Figure 1.

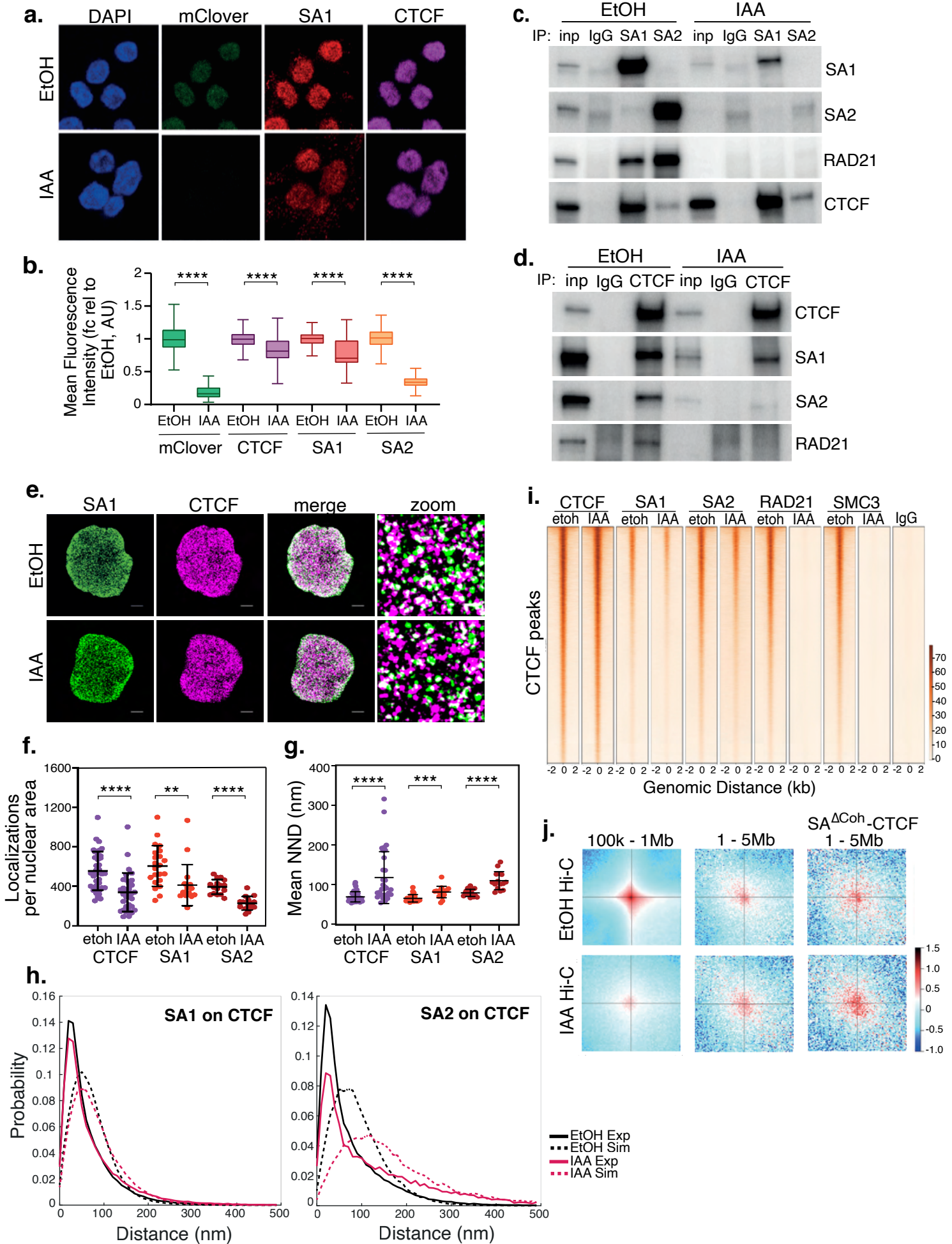


Figure 2.

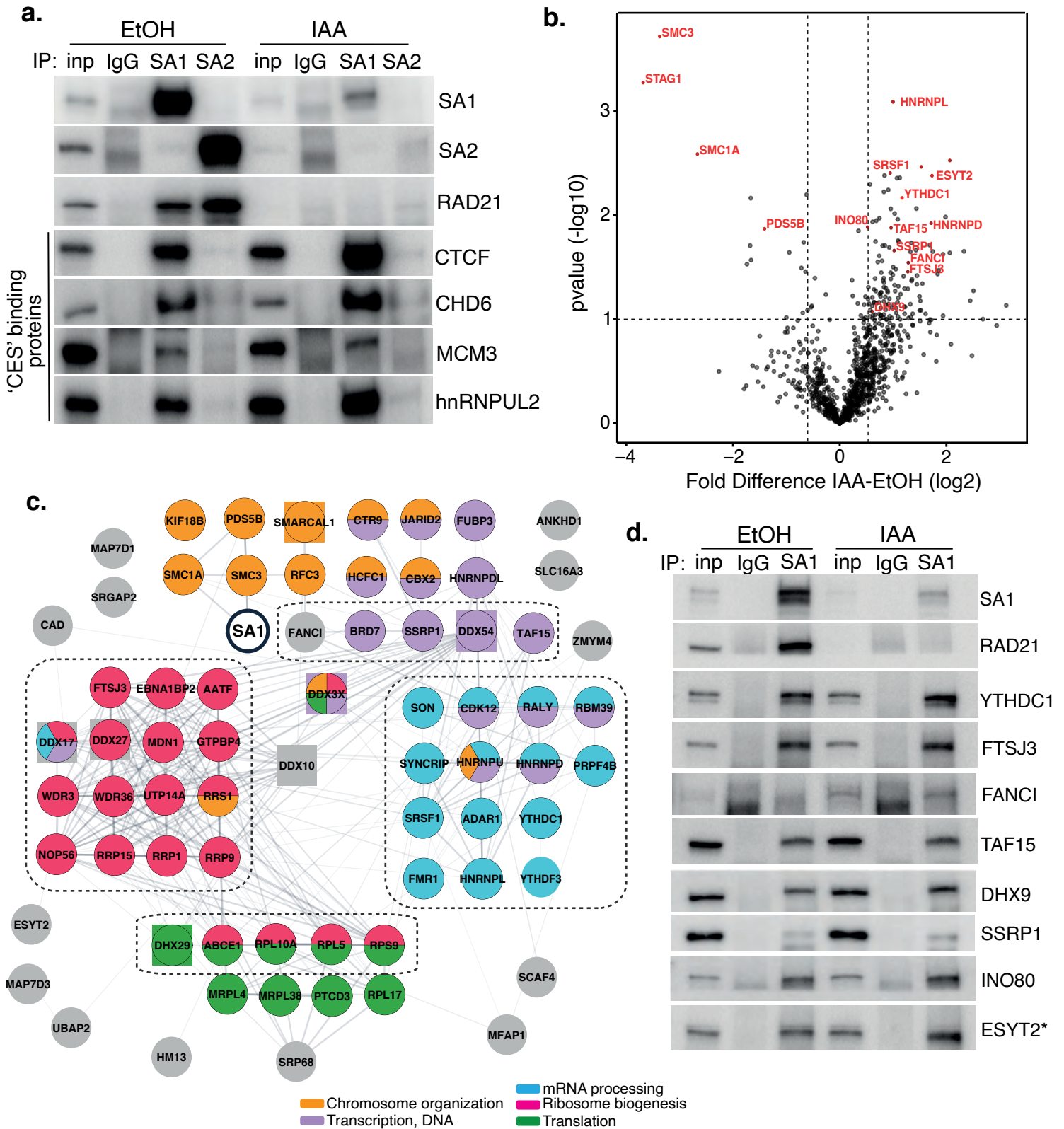


Figure 3.

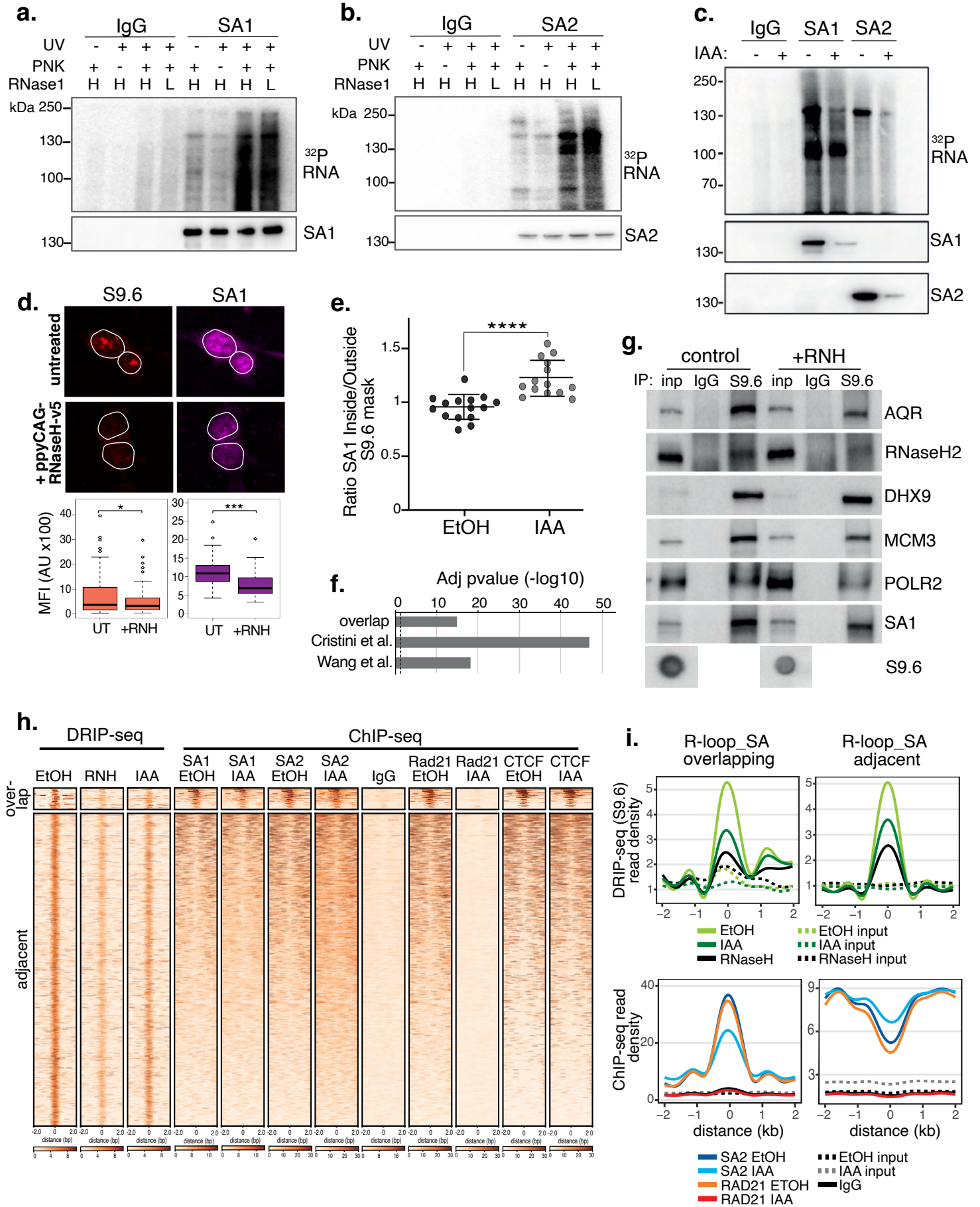


Figure 4

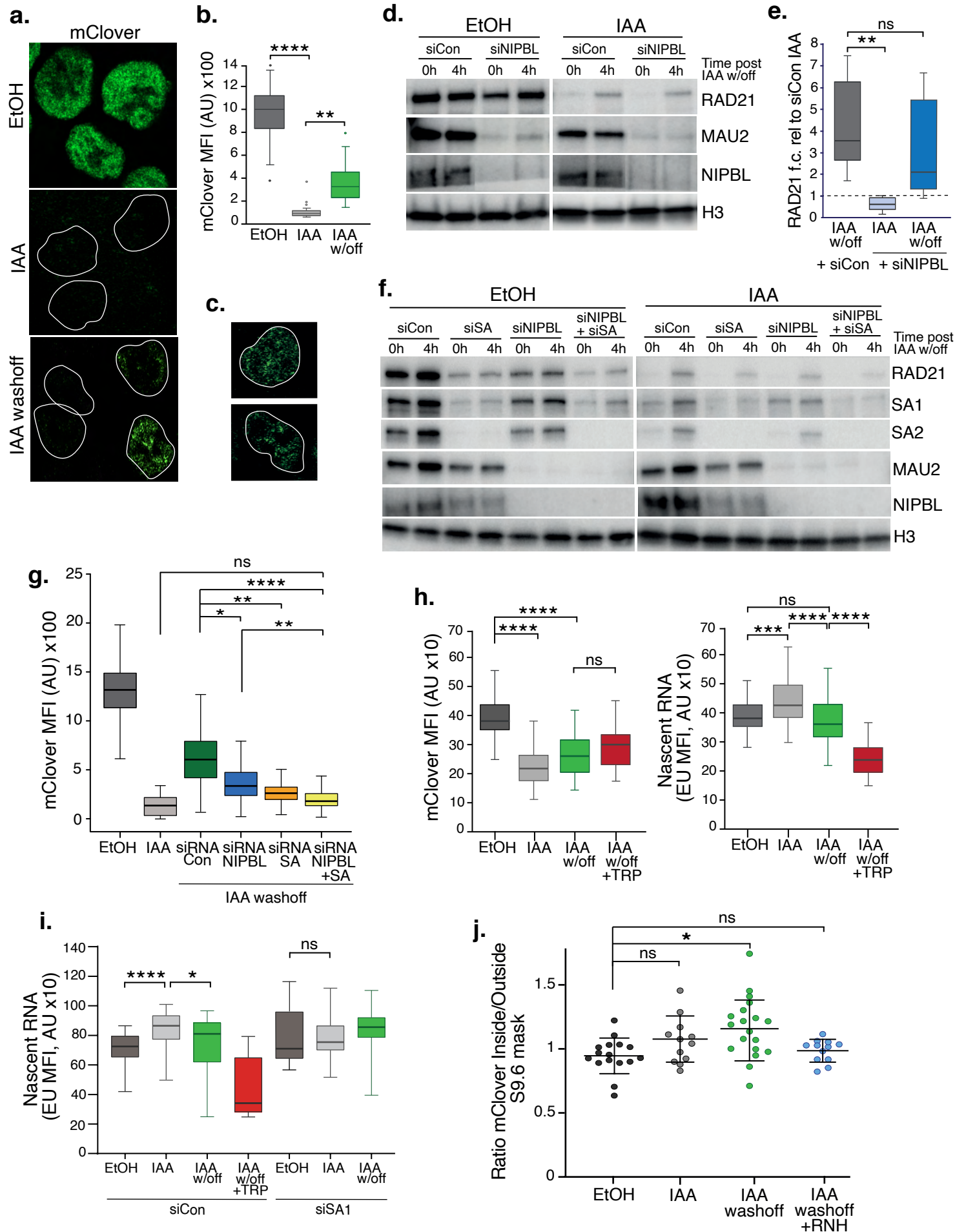


Figure 5.

