Active enhancers strengthen insulation by RNA-mediated CTCF binding at TAD boundaries

Zubairul Islam[1,2*], Bharath Saravanan[1,2*], Kaivalya Walavalkar[1], Jitendra Thakur [3], Umer Farooq[1,4], Anurag Kumar Singh[1], Radhakrishnan Sabarinathan[1], Awadhesh Pandit[1], Steven Henikoff [5], Dimple Notani[1#]

Affiliations:

1. Genetics and Development, National Centre for Biological Sciences, Tata Institute for Fundamental Research, Bangalore, Karnataka, 560065. India
2. Sastra Deemed University, Thanjavur, Tamil Nadu, 613401. India
3. Department of Biology, Emory University, Atlanta, Georgia, 30322. USA
4. The University of Trans-Disciplinary Health Sciences and Technology, IVRI road, Bangalore, Karnataka, 560064. India
5. Basic Sciences Division, Fred Hutchinson Cancer Research Centre, 1100 N. Fairview Ave, Seattle, WA, 98109, USA

* Equal contributions

\# Corresponding author: dnotani@ncbs.res.in

Abstract

Vertebrate genomes are partitioned into Topologically Associating Domains (TADs), which are typically bound by head-to-head pairs of CTCF binding sites. Transcription at TAD boundaries correlates with better insulation, however, it is not known whether the boundary transcripts themselves contribute to boundary function. Here we characterize boundary-associated RNAs genome-wide, focusing on the disease-relevant *INK4a/ARF* TAD. Using CTCF site deletions and boundary-associated RNA knockdowns, we observe that boundary-associated RNAs facilitate recruitment and clustering of CTCF at TAD borders. The resulting CTCF enrichment enhances TAD insulation, enhancer:promoter interactions and TAD gene expression. Importantly, knockdown of boundary-associated RNAs results in loss of boundary insulation function. Using enhancer deletions and CRISPRi of promoters we show that active TAD enhancers but not promoters induce boundary-associated RNA transcription, thus defining a novel class of regulatory enhancer RNAs.

Introduction

The genomes of multicellular eukaryotes are organized into self-interacting units known as Topologically Associating Domains (TADs) (1). TADs are identified in Hi-C interaction matrices as contact domains, and contact domains that exhibit corner dots (at the apex of a loop) are known as loop domains (2). Loop and contact domains are both proposed to be formed by loop extrusion (1,3-7) and are dependent on CTCF and cohesin (6-10). The length of an extruded loop is determined by the presence and orientation of CTCF bound regions that interact within the loop (11). By preventing enhancers and promoters within the loop from interacting outside of the loops, boundary CTCF pairs insulate genes within a TAD (7-8,12-13). However, the presence of CTCF consensus motifs or their orientation at contact domains do not accurately predict TAD boundaries, suggesting that additional factors must influence CTCF binding and, therefore, TAD formation and insulation.

Mutation of CTCF sites at TAD boundaries affects gene expression, ranging from no effect to severe effect in the same and neighboring TADs (14-21). Mutation of all CTCF sites within a boundary region or depletion of CTCF may be required to display effects on transcription (20,22), suggesting redundancy among multiple CTCF sites at a TAD boundary. However, it is unclear how multiple CTCF sites within the boundary coordinate to achieve better insulation. Active transcription at TAD boundaries correlates with better insulation and CTCF occupancy (22-24) and has been speculated to increase TAD strength (25-27), consistent with observations that inhibition of transcription or Ribonuclease A (RNaseA) treatment disrupts 3D chromatin architecture (28-30). Furthermore, CTCF dimerization, important for loop extrusion, is RNA-dependent, and RNA binding of CTCF is required for TAD chromatin organization (31-32), but how CTCF:RNA interactions maintain TADs is not clear.

80 In some instances, TAD boundaries are found near sites of enhancers (33). Also, TADs containing
81 super-enhancers consist of multiple enhancers exhibit strong boundaries (25,34). However, it is
82 not known whether these enhancers play a role in boundary insulation. Moreover, boundaries
83 themselves display enhancer-like features (30,35), although the levels of H3K27ac, RNA
84 Polymerase II (PolII) and H3K4me3 on TAD boundaries are far less than are found on active
85 enhancers in general (36). These observations suggest a relationship between certain TAD
86 boundaries and enhancers, but the mechanistic basis for their putative interactions remains
87 unknown.

88 Here we show that boundaries of about half of human TADs are transcribed, either from a gene-
89 body that overlaps a boundary region or from a non-genic region that is transcribed within the
90 boundary. Compared to non-transcribed TAD boundaries, transcribed genic and non-genic
91 boundaries are more enriched for CTCF binding and are better insulators of genes within TADs,
92 which are more highly expressed. RNAs transcribed from non-genic (i.e. enhancer-like regions)
93 boundaries are more stable than RNAs expressed from enhancers at non-boundary regions.
94 Similarly, genic transcripts from boundaries are also more stable than other mRNAs. To test for
95 functional relevance of boundary-associated RNAs, we focused on the well-characterized
96 *INK4a/ARF* TAD, which is bound by CTCF sites and harbors a super-enhancer. Both deletions of
97 CTCF sites and knockdown of boundary-associated RNAs resulted in loss of enhancer:promoter
98 interactions within TADs while strengthening interactions with enhancers in regions outside the
99 TAD. These observations suggest that active enhancers within TADs promote TAD insulation by
100 triggering transcription of the boundaries. We propose that the RNAs produced at boundaries
101 recruit CTCF, resulting in robust intra-TAD transcription.
102

103 **Results**

104 **Transcribed TAD boundaries insulate better than non-transcribed boundaries**

105 To determine the frequency of transcribed TAD boundaries, we measured run-on transcription
106 with GRO-seq (37) on TAD boundaries identified by HOMER's findTADsAndLoops.pl. This
107 program identifies both contact and loop domain TADs based on intra-domain interaction
108 frequency. Both TADs and loops are thought to be formed by loop extrusion and to be dependent
109 on CTCF and cohesin (6, 8-10). HOMER identified 15,397 TAD boundaries at 5 kb resolution
110 using HiC data from HeLa cells (11). GRO-seq detected at least ten normalised counts within
111 about half of the TADs identified by HOMER (8134, 52.8%, Fig. 1A). Almost 80% of transcribed
112 boundaries are genic (Fig. 1B-C).

113 TAD boundaries are enriched for CTCF binding. We determined if strong CTCF binding at TAD
114 boundaries is associated with increased transcription. Using high-resolution CTCF CUT&RUN
115 data (38), we observed that both genic and non-genic transcribed TAD boundaries had
116 significantly higher CTCF enrichment than non-transcribed boundaries (Fig. 1D). This
117 relationship between transcription and CTCF enrichment at TAD boundaries was also observed
118 for different RNA levels (Fig. 1E). Within non-genic transcribed boundaries, CTCF enrichment

119 was also correlated with GRO-seq signal (Fig. S1A-B). Similar results were seen using CTCF
120 ChIP-seq (39) (Fig. S1C-E). We conclude that transcribed TAD boundaries are more enriched for
121 CTCF binding than non-transcribed boundaries.

122 We next asked if transcribed TAD boundaries that show high CTCF binding are stronger
123 insulators. We noted a positive correlation between the levels of RNA at the transcribed boundaries
124 and the TAD insulation (Fig. 1F). Even TADs with the least transcribed boundaries were better
125 insulated as compared to non-transcribed boundaries (Fig. 1F). We observed a similar positive
126 correlation between boundary RNA levels and TAD insulation using two different bin sizes (5 kb
127 and 20 kb) in our analyses, thereby ruling out resolution-related biases (Fig. S2). A total of 69.2%
128 (10658 out of 15397) of TADs boundaries were found to be transcribed (Fig. S2A). The CTCF
129 enrichment was stronger on transcribed genic and non-genic boundaries at both 5 kb and 20 kb
130 bin sizes (Fig. S2B-E). Insulation scores on transcribed genic and non-genic boundaries were
131 correlated with levels of RNA at the boundaries (Fig. S2F-H)

132 We also examined the relationship between transcription of TAD boundaries and CTCF binding
133 using HICCUPS, which identifies loop anchors. In total, 3525 loop anchors were identified with 5
134 kb bin size, of which 43% (1519) were genic and 19% (675) were non-genic (Fig. S3A). As was
135 the case for TAD boundaries identified by HOMER, we found a positive correlation between
136 transcription of loop anchors and CTCF enrichment using both CUT&RUN (Fig. S3B-C) and
137 ChIP-seq datasets (Fig. S3D-E). For subsequent analyses, we focused on TADs called by HOMER
138 with 5 kb boundary regions.

139 We also observed positive associations between CTCF enrichment, insulation score and
140 transcription at TAD boundaries in IMR90 fibroblasts (11) (Fig. S4A-D) and lymphoblastoid cells
141 K562 (40) (Fig. S4E-H). Together, our results suggest that CTCF enrichment and enhanced
142 insulation are general features of transcribed TAD boundaries.

143

144 **RNA increases CTCF occupancy at TAD boundaries**

145 We wondered whether elevated levels of transcription and CTCF at TAD boundaries are causally
146 related. To test this possibility, we treated HeLa cells with RNaseA and performed Western blot
147 analysis on the nuclear and chromatin fractions. We observed a significant loss of CTCF in the
148 chromatin fraction (Fig. 2A, S6A), consistent with the reported dependence of CTCF on RNA for
149 chromatin binding (29,31-32). To map the genomic sites of RNA-dependent CTCF binding, we
150 performed CUT&RUN for CTCF with RNaseA treatment (CUT&RUN.RNaseA). Upon RNA
151 depletion, we observed an overall decrease in CTCF occupancy, especially at the TAD boundaries
152 where CTCF is most enriched (Fig. 2B). Genome-wide, transcribed boundaries exhibited greater
153 loss of CTCF with RNaseA treatment as compared to non-transcribed and randomly selected
154 boundaries (Fig. 2C).

155 We next asked if the loss of CTCF at TAD boundaries is due to the digestion of RNA at direct
156 binding sites as opposed to a general effect of RNA degradation. We chose the ~500-kb

157 *INK4a/ARF* TAD, which spans the *CDKN2A*, *CDKN2B* and *MTAP* cell cycle regulatory genes and
158 a super-enhancer (Fig. 2D, highlighted region) (41-42). The super-enhancer within this TAD
159 contains several active enhancers that are enriched for H3K27ac modification and enhancer RNAs
160 (eRNAs) (Fig. 2D, mNET-seq panel) (41). We confirmed 5' TAD boundary locations on the
161 transcribed gene *MTAP* and 3' boundary on the non-genic region containing three CTCF ChIP-
162 seq peaks (B1, B2 and B3). The position of these boundary regions was persistent in HiC data
163 from other cell types (Data not shown). CTCF peaks B1 and B2 but not B3 exhibit strong nascent
164 transcription based on publicly available mNET-seq, GRO-seq and PRO-seq data (43-45) (Fig.
165 S5) and so are classified as eRNAs. In our CUT&RUN.RNaseA data, we observed a loss of CTCF
166 at the boundary regions of the *INK4a/ARF* TAD (Fig. S6B), which suggests that eRNAs help
167 maintain CTCF at B1 and B2 CTCF sites. To test this interpretation, we depleted sense and anti-
168 sense eRNAs at B1 and B2 CTCF sites using shRNAs (Fig. 2E-F) and confirmed the drop of these
169 eRNAs from chromatin fraction by qRT-PCR (Fig. S6C). Knockdowns using shRNAs to B1 and
170 B2 but not B3 and nearby TAD boundaries (T1 and T2) resulted in the loss of CTCF binding at
171 these sites (Fig. 2G-H and Fig. S6D). Notably, the loss of RNAs at B1 reduced CTCF occupancy
172 not only at B1 but also at B2 and vice versa (Fig. 2G-H). This suggests that CTCF on transcribed
173 boundaries is maintained in part by eRNAs, and different transcribed CTCF sites within a
174 boundary are potentially interdependent. Notably, shRNA-mediated knockdown or over-
175 expression of CTCF had at most minor effects on eRNA expression (Fig. 2I and Fig. S6E), which
176 suggests that CTCF enrichment at TAD boundaries depends on RNA, but not vice-versa.

177

178 **Non-coding RNAs and CTCF interact at the *INKa/ARF* TAD boundary**

179 We next asked whether RNA and CTCF interact at functional boundaries. We first performed
180 RNA pulldowns of sense and anti-sense RNAs transcribed from B1 and B2 CTCF sites in the
181 *INKa/ARF* TAD followed by immunoblotting with a CTCF antibody (Fig. 3A). We observed
182 CTCF interactions with B1 and B2 sense and anti-sense RNA and also with an eRNA from the
183 CTCF-bound E8 enhancer within the super-enhancer, but not with the Box B negative control.
184 CTCF binding to B1 and B2 sense and anti-sense RNAs was confirmed by UV-RIP, an RNA
185 immunoprecipitation method (Fig. 3B). Surprisingly, the CTCF motifs present at transcribed
186 boundaries are much sparser than at non-transcribed or randomly selected boundaries (Fig. 3C-D).
187 However, we found that transcribed boundaries are more densely packed with CTCF peaks than
188 are non-transcribed or randomly selected boundaries (Fig. 3E). These data suggest that
189 transcription or the presence of eRNA can reinforce CTCF binding to multiple weak CTCF motifs.

190 Next, we compared the stability, polyadenylation and splicing features of genic RNAs and eRNAs
191 at TAD boundaries with those that are not within boundaries. In general, eRNAs exhibit fast
192 turnover as seen by a higher ratio of nascent to mature eRNAs as compared to genic RNAs (37).
193 By this measure eRNAs on non-genic boundaries were significantly more stable than other
194 eRNAs, and RNAs on genic boundaries were slightly more stable than other genic RNAs (Fig.
195 3G).

196 The AATAAA polyadenylation motif is rare around Transcriptional Start sites (TSSs) of genic
197 transcripts, increasing in abundance downstream as expected (Fig. 3H). Non-genic eRNAs on
198 boundaries showed less enrichment of pre-mature AATAAA than other eRNAs (Fig. 3H),
199 consistent with their higher stability. As expected, 5' splice consensus sequences were strongly
200 enriched in mRNAs relative to eRNAs (Fig. 3I), consistent with their greater stability. Together,
201 enhanced stability, higher CTCF enrichment and insulation promoting activity of boundary eRNAs
202 suggest that they have evolved to facilitate CTCF binding by direct physical interaction.

203

204 **Boundary RNAs enhance TAD transcription**

205 The enhanced insulation we observed at sites of boundary eRNAs suggested that genes within the
206 TADs of transcribed boundaries might be expressed at higher levels than those in TADs of non-
207 transcribed boundaries. Accordingly, we compared the transcriptional levels of TADs with high,
208 low, and randomly selected transcribed or non-transcribed boundaries. We found that the
209 transcriptional output of entire TADs (Fig. 4A) or of individual genes within TADs (Fig. S7A) is
210 correlated with the levels of RNAs at boundaries, and we obtained similar results for non-genic
211 boundaries (Fig. S7B). To determine whether this positive association resulted from direct
212 interactions between promoters and boundaries we analysed available promoter capture-seq data
213 in HeLa cells (46). Boundaries were captured by promoters far more than by random regions (Fig.
214 4B), and genes that interacted with boundaries were relatively more active than genes that
215 interacted with other genomic regions (Fig. 4C). Together, these data suggest that transcribed
216 boundaries tend to interact with promoters and are associated with higher rates of TAD
217 transcription. We confirmed the interaction of the *CDKN2A* promoter with the nearby transcribed
218 boundary within the *INK4a/ARF* TAD using 4C-seq assays (Fig. 4D).

219 Next, we asked if boundary RNAs play a direct role in transcriptional activity of genes within this
220 TAD. Using shRNAs to sense and anti-sense B1 and B2 CTCF sites, we observed downregulation
221 of *INK4a/ARF* (transcribed from *CDKN2A*), *INK4b* (transcribed from *CDKN2B*) and *MTAP*,
222 whereas the transcription of *DDX5*, a non-*INK4a/ARF* TAD gene, remained unchanged (Fig. 4E).
223 These data suggest that boundary eRNAs positively regulate the transcription of the genes within
224 the *INK4a/ARF* TAD by direct physical interaction between promoters and boundaries.

225 Since the loss of specific boundary RNAs also caused loss of CTCF binding at the TAD boundary,
226 we asked if the role of boundary RNA in TAD transcription is CTCF-dependent. Depletion of
227 global CTCF using a pool of CTCF-specific siRNAs resulted in downregulation of genes within
228 and outside of the *INK4a/ARF* TAD as expected (Fig. 4F-G). To test the effects of the loss of
229 CTCF binding only at the *INK4a/ARF* TAD boundary, we deleted the B1 and B2 CTCF sites
230 individually in HeLa cells (Fig. S8A-B) and observed downregulation of *INK4a/ARF* genes within
231 the TAD (Fig. 4H) and the loss of PolII at promoters (Fig. 4I). We also observed that boundary
232 eRNA knockdown and boundary deletion reduced the levels of the p16 protein encoded by
233 *INK4a/ARF* (Fig. S8C). These data suggest that binding of CTCF mediated by boundary RNAs
234 regulates the expression of genes within the *INK4a/ARF* TAD.

235 **Active enhancers within the TAD directly regulate boundary eRNA transcription**

236 What is responsible for *INK4a/ARF* boundary transcription? Because HeLa cells are addicted to
237 high levels of proteins encoded within the TAD (47-48), instead of deleting gene promoters, we
238 used CRISPR inhibition (CRISPRi) using specific guide RNAs. We applied dCas9-KRAB on
239 promoters of *CDKN2A*, *CDKN2B* and *MTAP* and assayed transcript levels at B1 and B2 CTCF
240 sites (Fig. 5A). CRISPRi of these promoters resulted in transcriptional inhibition of these genes
241 (Fig. 5B, red bars). However, boundary RNA levels were unaffected (Fig. 5B, blue bars).
242 Consistent with unperturbed boundary RNA expression, there was no change in CTCF occupancy
243 at B1 and B2 CTCF sites. However, we did observe loss of CTCF on the 5' boundary that overlaps
244 the MTAP gene, likely due to down-regulation of *MTAP* (Fig. 5C). The boundary of flanking
245 TADs also did not exhibit any perturbations in CTCF occupancy (Fig. 5C). These data suggest
246 that initiation of transcription from *CDKN2A* and *CDKN2B* promoters is not responsible for
247 eRNAs at the boundary.

248 We next asked whether the *INK4a/ARF* TAD enhancers regulate genes within the TAD and
249 boundary eRNAs. Genome-wide, transcribed boundary TADs are enriched in H3K27ac (Fig. 6A-
250 B) and PolII (Fig. 6C) relative to non-transcribed boundaries and randomly selected genic or non-
251 genic boundaries (Fig. 6C, Fig. S9A-B). We first tested the super-enhancer using E8, E12 and E17
252 enhancer deletion lines (Fig. 6D) in which *INK4a/ARF* genes were seen to be downregulated based
253 on RNA-seq data (Fig. 6E) (41), confirming that these enhancers are functional enhancers. We
254 then asked if these enhancers indeed regulate the boundaries as well. Interestingly, we observed
255 down-regulation of sense and anti-sense boundary RNAs at the B1 and B2 CTCF sites upon E8,
256 E12 or E17 deletions (Fig. 6F), which suggests that enhancers activate boundary transcription. To
257 determine whether down-regulation might be caused by direct contact, we also performed 4C-seq
258 on the E8 enhancer. We observed strong interactions with other enhancers in the super-enhancer
259 cluster, with *CDKN2A* and *CDKN2B* promoters and with the boundary region (Fig. 6G, right
260 highlighted pink). Importantly, deletion of the E8 enhancer conspicuously reduced 4C-seq signal,
261 confirming that the interaction between the super-enhancer and distant B1 and B2 CTCF sites
262 regulates boundary transcription (Fig. 6G). Although interactions between boundaries and
263 enhancers have been previously reported (35) no regulatory relationship had been established,
264 whereas our data demonstrate that apart from activating their target promoters, functional
265 enhancers positively regulate the transcription at the boundaries by direct interactions.

266 To ascertain whether high levels of boundary RNAs enhance insulation of the boundary *in vivo*,
267 we interrogated the *INK4a/ARF* TAD upon senescence, where TAD transcription increases due to
268 its cell cycle regulatory role. Upon senescence, bidirectional eRNA expression increases several-
269 fold at the super-enhancer cluster (Fig. S10A, blue highlighted region), and *INK4a/ARF* gene
270 expression and boundary RNA expression also increases (Fig. S10A, pink highlighted region). We
271 validated these increases in gene transcription, boundary RNA expression and CTCF enrichment
272 upon X-ray induced senescence (non-replicative senescence) (Fig. S10B-D), and also observed
273 elevated p16 and p21 protein levels (Fig. S10E). Using HiC data, we observed increase in intra-

274 TAD interactions at this TAD in senescent cells as compared to proliferative cells (Fig. S10F).
275 Together these data suggest that increases in TAD enhancer activity accompany increases in their
276 interactions with boundaries to induce boundary RNA expression, which in turn recruits more
277 CTCF to enhance insulation.

278

279 **Boundary eRNAs strengthen TAD insulation**

280 We noticed that TADs with transcribed boundaries exhibited higher strength of enhancer/promoter
281 interactions than TADs with non-transcribed boundaries genome-wide (Fig. 7A). To test whether
282 boundary function of the *INK4a/ARF* TAD affects super-enhancer/promoter interactions in a
283 boundary RNA dependent manner, we performed 4C-seq assays at the E8 enhancer upon
284 knockdown of sense and anti-sense boundary RNAs at the B1 CTCF site. As expected, we
285 observed loss of boundary/enhancer interactions, but in addition, we observed weakening of
286 *INK4a/ARF* enhancer/promoter interactions (Fig. 7B). Knockdown of B1 RNAs also resulted in
287 increased interaction with the +T1 boundary of the neighboring TAD (Fig. 7B), which also showed
288 a gain in CTCF binding (Fig. 2G). To confirm that boundary RNA loss was accompanied by
289 reduced CTCF binding, we performed 4C-seq at the E8 enhancer using cells with a B1 CTCF site
290 deletion. We again observed a loss of enhancer/promoter interaction and a concomitant increase
291 in the interaction with the +1 T1 boundary of the neighboring TAD (Fig. 7B). These observations
292 demonstrate that loss of a boundary RNA reduces CTCF binding and weakens insulation, resulting
293 in aberrant enhancer interaction with an element outside of the TAD. We conclude that boundary
294 RNAs are functional components of TAD boundaries (Fig. 7C).

295

296 **Discussion**

297 Interactions between CTCF and RNA have been implicated in genome organization (31-32), but
298 the exact role of these interactions remains unknown. Here we show that eRNAs transcribed from
299 TAD boundaries function in CTCF binding, TAD insulation and gene regulation. Our data provide
300 direct evidence that boundary eRNAs create an insulated neighbourhood by associating with
301 CTCF-bound sites and facilitating enhancer:promoter interactions for robust gene activation
302 within a given TAD. We found that the transcriptional activity of promoters present within the
303 TAD does not regulate non-genic boundary transcription. Using a series of enhancer and CTCF
304 site deletions at the *INK4a/ARF* TAD, we have shown that enhancers interact with and positively
305 regulate the expression of both genic and non-genic TAD boundary RNAs. Boundary RNAs, in
306 turn, increase CTCF binding at TAD boundaries, strengthening weak CTCF motifs and increasing
307 insulation. Transcription of boundary RNAs at the *INK4a/ARF* TAD is regulated by the super-
308 enhancer within the TAD to facilitate CTCF binding and increased insulation at the boundary. This
309 may explain why super-enhancers containing TADs exhibit stronger boundaries (34-36).

310 Boundary RNAs resemble eRNAs at active enhancers in being bidirectional (49-50), and as our
311 study demonstrates, making direct contact with distant regulatory elements. However, unlike
312 eRNAs, boundary RNAs contact enhancers that also contact promoters to enhance TAD gene

313 activation. Popular chromosome conformation capture assays are strictly pairwise and cannot
314 determine if enhancer elements activate promoters and boundary elements independently or
315 together. However, more recently introduced 3D contact methods, including SPRITE (51) and
316 GAM (52) are able to capture multiple simultaneous contacts, and in the future, may be applied to
317 understand how enhancers regulate both gene promoter activity and transcription at boundaries to
318 strengthen TADs.

319 Furthermore, in the absence of the RNA binding domain of CTCF, cells exhibit a decrease in the
320 total number of TADs and an increase in the size of individual TADs due to reduced insulation
321 between neighboring TADs (31). A clear mechanistic basis for this observation comes from our
322 findings in which TAD enhancer-mediated CTCF-RNA interactions maintain strong TAD
323 boundaries by increasing insulation between neighboring TADs. Thus, the inability of CTCF
324 mutants to interact with RNA will compromise CTCF enrichment and its insulating function at
325 TAD boundaries, leading to the merging of adjacent TADs into larger TADs. In such cases, less
326 CTCF at the boundary might allow the cohesin ring to slip to the next TAD thereby compromising
327 the insulation (53-54). Thus a positive feedback loop is established whereby super-enhancers
328 maintain high specificity for promoters within the TAD by also activating boundary transcription
329 and preventing interactions with neighboring promoters.

330 TAD boundaries are enriched for H3K27ac and PolII, suggesting the presence of active regulatory
331 elements and transcription (24,34-36). However, activation of transcription at the boundary alone
332 is not sufficient to maintain a TAD boundary (23,55). Our study shows that the products of the
333 TAD boundary transcription *i.e.*, boundary RNAs, are required to strengthen TAD insulation by
334 increasing CTCF occupancy. Transcription-assisted TAD boundary strengthening may also be a
335 factor in zygotic activation-dependent boundary formation where early transcribing genes serve as
336 the nucleation sites for TAD boundaries (55). Similarly, transcription and the resulting RNA may
337 be involved in forming distinct TADs boundaries around transcribed escape genes on the inactive
338 chromosome during X chromosome inactivation (26).
339

340 **Materials and Methods**

341 **Cell Culture:**

342 HeLa cells were obtained from ATCC and were cultured in (DMEM) Dulbecco's Modified Eagle's
343 medium supplemented with 10% Fetal Bovine Serum and 1% Penicillin/Streptomycin. The cells
344 were maintained at 37°C in the presence of 5% CO_2 in a humidified incubator.

345 **Chromatin Immunoprecipitation:**

346 Cells were crosslinked with 1% formaldehyde at room temperature for 10 min with constant
347 shaking. Glycine was added to a final concentration of 125 mM to quench the formaldehyde. Cells
348 were washed three times with 1X ice-cold PBS then were scraped and pelleted down 1X PBS at 4
349 ^{o}C. Cells were gently resuspended in nuclear lysis buffer (50 mM Tris-HCl pH 7.4, 1% SDS, 10
350 mM EDTA pH 8.0) supplemented with 1X PIC and incubated on ice for 10 min. The lysate was

351 subjected to fragmentation using a Diagenode Bioruptor Pico for 20 cycles (30 s ON and 30 s OFF)
352 to generate fragments of around ~500 bp. The cell lysate was cleared at 12K rpm for 12 min. For
353 each IP, 100 µg of sheared chromatin was used. The lysates were diluted by adding Dilution Buffer
354 (DB) (20 mM Tris-HCl pH 7.4, 100 mM NaCl, 2 mM EDTA pH 8.0, 0.5% Triton X-100)
355 supplemented with 1X PIC in 1:1.5 ratio (1 volume of sheared chromatin and 1.5 volumes of
356 Dilution Buffer). One µg of antibody was added to immunoprecipitate the DNA and incubated on
357 a rocking platform overnight at 4ºC. Fifteen µl of Pre-blocked (with 1% BSA) Protein G
358 Dynabeads (10004D, Invitrogen) was added to the tubes and rotated for 4h at 4 °C. Beads were
359 collected, flow-through was discarded and the samples were washed with Wash Buffer I (20 mM
360 Tris-HCl pH 7.4, 150 mM NaCl, 0.1% SDS, 2 mM EDTA pH 8.0, 1% Triton X-100) at 4 ºC on a
361 rocking platform. Washes were sequentially repeated with Wash Buffer II (20 mM Tris-HCl pH
362 7.4, 500 mM NaCl, 2 mM EDTA pH 8.0, 1% Triton X-100), Wash Buffer III (10 mM Tris-HCl
363 pH 7.4, 250 mM LiCl, 1% NP-40, 1% Sodium Deoxycholate, 1 mM EDTA pH 8.0 and 1X TE (10
364 mM Tris pH 8.0, 1mM EDTA pH 8.0). Chromatin was eluted with 200 ul of Elution Buffer (100
365 mM NaHCO3, 1% SDS) for 30 min at 37ºC in a thermomixer. The supernatant was collected in
366 separate tubes and 14 µl NaCl (5 M) was added to eluted samples and kept at 65 °C overnight for
367 de-crosslinking. Immunoprecipitated DNA was purified by phenol:chloroform:isoamyl alcohol,
368 followed by ethanol precipitation. The final air-dried DNA pellet was dissolved in 100 µl of 1X
369 TE. These samples were then used for qRT-PCRs using SYBR Green. Mean values for all the
370 regions analyzed in different conditions were expressed as fold-enrichment compared over beads.
371 Statistically significant differences between different conditions were computed using a *T*-test (*p*-
372 value < 0.05), ChIP primers are listed in Supplementary Table. 1.

373 **RNA Isolation and cDNA synthesis:**

374 Cells were lysed in 1 ml of Trizol (Thermo Fisher Inc.). To each sample, 200 µl chloroform was
375 added, briefly vortexed and centrifuged at 12K rpm for 12 min. The aqueous phase was carefully
376 collected and transferred to the fresh tube. One volume of isopropanol was added to the sample
377 and incubated at room temperature for 10 min to precipitate the RNA. The samples were
378 centrifuged at 12K rpm for 12 min, and supernatants were discarded without disturbing the pellet.
379 The pellet was washed with 75% ethanol. The pellet was air-dried and dissolved in RNase-free
380 water. The RNA was treated with ezDNase (Invitrogen) to remove the traces of contaminating
381 DNA. One µg of RNA was used for each cDNA synthesis reaction by Superscript IV (Invitrogen)
382 and random hexamers as per the manufacturer's recommendation. The CFX96 Touch (Biorad)
383 real-time PCR instrument was used for qRT-PCRs. qRT-PCRs were performed using three
384 biological replicates for each sample. Fold change was calculated by the ΔΔCt method and
385 individual expression data were normalized to GAPDH mRNA. The *p*-values were calculated by
386 Student's unpaired two-tailed *t*-test. qRT-PCR primers are listed in Supplementary Table 1.

387 **CRISPR Cas9 mediated deletion:**

388 Guide RNAs (gRNAs) were designed with the crispr.mit.edu tool. gRNAs were selected based on
389 the highest score and the least number of off-targets. gRNAs were cloned in a pgRNA humanized

390 vector (Addgene #44248, a gift from the Stanley Qi Lab) between BstX1 and Xho1 restriction
391 sites. gRNAs were co-transduced with a lenti-Cas9 vector. Cells were placed under selection using
392 puromycin (3µg/ml) for 48 hours. Single cells were seeded in a 96 well plate. Wells with single
393 cells were marked and allowed to grow until they formed colonies. The cells were trypsinized and
394 shifted to 48 well plates and allowed to grow to confluence, and then half of the cells were taken
395 for the Surveyor assay and the other half was plated again. Surveyor assays were performed using
396 a PCR-based method using primers listed in the Supplementary Table 1.

397

398 **CRISPRi gRNA design and cloning:**

399 We used GPP sgRNA Designer (https://portals.broadinstitute.org/gpp/public/analysis-tools/sgrna-
400 design) to design the gRNAs against TSSs of *MTAP*, *CDKN2A* and *CDKN2B* in the *INK4a/ARF*
401 TAD. For each gene, two gRNAs were designed. The gRNAs were cloned in a customized PX459
402 vector (pSpCas9(BB)-2A-Puro V2.0, Addgene #62988, a gift from the Feng Zhang lab). The Cas9
403 enzyme cassette was removed from PX459 by digestion with Xba1 and Not1 followed by blunt-
404 end ligation. The cloning of the gRNAs in the customized PX459 plasmid was performed as per
405 the Zhang Lab's general cloning protocol. The dCas9 plasmid, together with gRNA specific to
406 targeted genes, were transfected in 70% confluent cells. Transfections were done using
407 Lipofectamine 2000 (Invitrogen, cat no #11668027). CRISPRi cells were harvested for ChIP and
408 gene expression analysis.

409 **shRNA design and transfection:**

410 We used an shRNA design tool (https://portals.broadinstitute.org/gpp/public/seq/search) to design
411 the shRNAs against RNA at B1 and B2 CTCF sites within the *INK4a/ARF* TAD boundary. The
412 shRNAs were cloned in a customized pLKO.1 puro, (Addgene #8453, a gift from Bob Weinberg.
413 The cloning of the shRNAs in the pLKO.1 plasmid was performed as per the Addgene's pLKO.1
414 cloning protocol. shRNA-specific plasmid together with lentiviral packaging plasmids like VSVG
415 (a gift from Bob Weinberg, Addgene #8454), and PAX2, a gift from Didier Trono (addgene
416 #12260) were co-transfected using Lipofectamine 2000 Invitrogen (cat #11668027). Cells were
417 selected by puromycin (3µg/ml) (GIBCO cat # A11138-03) and knockdowns were confirmed
418 using qPCR oligos listed in Supplementary Table 1.

419 **siRNA transfection:**

420 SMARTpools was used to design scrambled (cat no D-001810-10-05) and CTCF siRNAs (L-
421 020165-00-0005) were purchased from GE Dharmacon. Transfections were done using
422 Lipofectamine 2000 (Invitrogen cat no 11668027).

423 **Lentiviral transduction:**

424 HEK293FT cells were grown in poly-D lysine coated culture dishes. These cells were co-
425 transfected with lentiviral packaging plasmids like VSVG (a gift from Bob Weinberg, Addgene
426 #8454) and PAX2 (a gift from Didier Trono, Addgene #12260) along with the plasmid of interest

427 using Lipofectamine 2000 Invitrogen (cat# 11668027). The medium was changed after 6 hours.
428 The viral soup was collected after 48 h and 72 hours, pooled together, filtered with a 0.44 µm
429 syringe filter and then finally added to cells along with 8 µg/ml of polybrene. Transduction was
430 stopped after 16 hours.

431 **RNA Immunoprecipitation:**

432 For RNA, primers were designed to amplify desired genomic regions that correspond to peaks of
433 boundary RNA (GRO-seq) at B1 and B2 CTCF sites of the *INK4a/ARF* TAD boundary. The
434 amplified products were cloned into the pcDNA3 BoxB plasmid (a gift from Howard Chang,
435 Addgene #29729). All clones were confirmed by Sanger sequencing and subsequently used for
436 RNA synthesis using T7 RNA polymerase (Promega Inc. P207e) and biotin RNA labeling mix
437 (Roche 11685597910). Primer sequences are listed in Supplementary Table 1.

438 RNA immunoprecipitation was adopted from (56), with some modifications. HeLa cells were
439 grown to 90–95% confluency and were harvested by scraping and washed twice with ice-cold 1X
440 PBS to gently resuspend in 2 ml of PBS. A volume of 2 ml of nuclear isolation buffer [NIB: 40
441 mM Tris–HCl pH 7.5, 20 mM $MgCl_2$, 1.28 M sucrose, 4 % Triton X-100, 1 mM PMSF, protease
442 inhibitors, and 20 U/ml SUPERase inhibitor (Thermo Fisher Scientific, catalog # AM2694)] was
443 added and the pellet was gently resuspended. Then, 6 ml of distilled water was added and kept on
444 ice for 20 min with intermittent gentle shaking. Nuclei were pelleted by centrifugation at 2500×*g*
445 for 5 min at 4°C. The pellet was resuspended in 1 ml of RIP buffer (25 mM Tri–HCl pH 7.4, 150
446 mM KCl, 0.5 mM DTT, 0.5 % NP40, 1 mM PMSF, protease inhibitors, and 20 U/ml SUPERase
447 inhibitor) and incubated on ice for 5 min. Nuclei were sheared by 10 cycles of sonication (30s on
448 and 30 s off) in a Bioruptor (Diagenode) followed by centrifugation at 12K rpm for 10 min at 4°C.
449 One µg of biotinylated RNA was incubated with 20 µl of RNA structure buffer (RSB) (10 mM
450 Tris–HCl pH 7.0, 100 mM KCl, 10 mM $MgCl_2$, 1 mM PMSF, protease inhibitors, and 20 U/ml
451 SUPERase inhibitor) at room temperature for 5 min. Folded RNA was mixed with 1 mg of nuclear
452 extract in 500 µl of RIP buffer and rotated at 4^{0}C for 1h. Fifteen µl of Dynabeads MyOne
453 Streptavidin beads was added and rotation continued for one more hour. Samples were washed
454 three times with RIP buffer and beads with proteins were boiled in 2XSDS for 10 min.
455 Immunoblotting for CTCF was performed using a CTCF antibody (Cell Signaling Technology,
456 #3418).

457 **UV RIP:**

458 Ultraviolet-RNA Immunoprecipitation was performed as in (57) with some modifications. HeLa
459 cells at a confluency of 80–90% were cross-linked by UV irradiation in Stratalinker UV cross-
460 linker. Cells were washed three times with cold 1XPBS and scraped in PBS before subjecting to
461 centrifugation at 2500 rpm. Pellets were lysed in RIP lysis buffer [25 mM HEPES-KOH at pH 7.5,
462 150 mM KCl, 0.5% NP40, 1.5 mM $MgCl_2$, 10% glycerol, 1 mM EDTA, 0.4 U RNase inhibitor
463 (Invitrogen, 18091050)], and 1X protease inhibitor cocktail (PIC) on ice for 30 min. Cleared cell
464 lysates were subjected to IP with CTCF antibody-bound Protein G Dynabeads (Invitrogen,
465 10004D) overnight. Samples were subsequently washed three times with RIP lysis buffer and RNA

466 samples were eluted using TRIzol reagent. Prior to cDNA synthesis, RNA was treated with RNase-
467 Free DNase Qiagen (cat no.79256) to remove the traces of contaminating DNA. cDNA was
468 prepared using random hexamers by Superscript IV (Invitrogen 18091050) as per the
469 manufacturer's recommendation and analyzed by PCR primers listed in Supplementary Table 1
470 and were run on 1% Agarose gel.

471 **Cell Fractionation:**

472 Cell fractionation and RNAseA treatment of nuclei was adapted from (58). Briefly, cells were
473 washed three times with cold 1XPBS and scraped in 1ml of PBS, pelleted down at 2500 rpm for 5
474 min at 4^0C. Pellets were then resuspended in Hypotonic Lysis Buffer (HLB) (10 mM Tris HCl pH
475 7.5, 10 mM NaCl, 3 mM $MgCl_2$, 0.3% NP-40, 10% glycerol, 1XPIC), the mix was incubated on
476 ice for 10 min and then centrifuged at 800g for 8 min at 4^0C. The supernatant was transferred to a
477 new tube and marked as Cytoplasmic fraction. The nuclei pellet after centrifugation was washed
478 twice with HLB and centrifuged at 200g for 2 min at 4^0C. The pellet was then resuspended in 700
479 ul of ice cold Modified Wuarin-Schiebler buffer (MWS: 10 mM Tris-HCl, pH 7.5, 300 mM NaCl,
480 4 mM EDTA, 1 M urea, 1% NP-40, 1% glycerol, 1XPIC) with or without 0.1l of 100mg/ml
481 RNaseA (Qiagen cat. no. 19101) for 15 min on ice briefly tapped few times during incubation time
482 and then centrifuged at 1000g for 5 min at 4^0C. The supernatant was collected as a nucleoplasmic
483 fraction. The chromatin pellet was washed twice with MWS buffer and centrifuged at 500g for
484 3min at 4^0C. The pellet was then dissolved in 500 µl of Nuclear Lysis Buffer (NLB: 20 mM Tris
485 HCl pH 7.5, 150 mM KCl, 3 mM $MgCl_2$, 0.3% NP-40, 10% glycerol, 1XPIC) and sonicated for
486 10 cycles (30 s on, 30 s off) in a Bioruptor (Diagenode). SDS loading dye was added and the
487 samples were boiled before immunoblotting.

488 **Chromatin-associated RNA isolation:**

489 Chromatin-associated RNA isolation was adapted from (59). Trypsinized cells were collected in
490 1.5 ml tubes and washed twice with 1xPBS and pellet down at 200xg for 2 min. To each pellet
491 was added 400 µl of Cell lysis buffer (10 mM Tris HCl ph7.4, 150 mM NaCl, 0.15% Igepal),
492 gently pipetted 3-5 times and then incubated for 5 min on ice. To the Lo-bind 1.5ml tube was
493 added 1 ml (2.5 volume) of cold sucrose buffer (10mM Tris HCl ph7.4, 150mM NaCl, 24%
494 Sucrose) and gently overlaid with the cell lysate followed by centrifugation at 3500xg for 5 min
495 to collect the nuclear pellet. Isolated nuclei were rinsed with 1 ml of ice-cold 1xPBS-EDTA
496 solution followed by a short spin at 3500xg. The nuclei were resuspended in 250 µl of glycerol
497 buffer (20 mM Tris HCl ph7.4, 75 mM NaCl, 0.5 mM EDTA pH 8.0, 50% Glycerol) and then
498 immediately 250 µl of Urea buffer (10 mM Tris HCl ph7.4, 1 M Urea, 0.3 M NaCl, 7.5 mM $MgCl_2$,
499 0.2 mM EDTA, 1% Igepal) was added, mixed by vortexing for 4 s and then incubated on ice for 2
500 min. Chromatin was centrifuged at 13000xg for 2 min to collect the Chromatin-RNA complex.
501 The chromatin-RNA pellet was washed with 1xPBS-EDTA once and RNA was isolated using
502 Trizol Reagent. RNA obtained was treated with ezDNase (Ambion) to remove the traces of
503 contaminating DNA. One µg of RNA was used for each cDNA synthesis reaction by Superscript
504 IV (Invitrogen) and random hexamers as per the manufacturer's recommendation. The CFX96

505 Touch (Biorad) real-time PCR was used for qRT-PCRs. qRT-PCRs were performed using three
506 biological replicates for each sample. Fold changes were calculated by the ΔΔCt method and
507 individual expression data was normalized to pre-c-myc. The *p*-values were calculated by
508 Student's unpaired two-tailed *t*-test. qRT-PCR primers are listed in Supplementary Table 1.

509 **X-ray induced Senescence:**
510
511 BJ fibroblasts were cultured in (DMEM) Dulbecco's Modified Eagle's medium supplemented with
512 10% Fetal Bovine Serum and 1% Penicillin/Streptomycin at 37 °C in a humidified incubator
513 containing 5% CO_2. Cells were first immortalized by expressing hTERT. Cells were then seeded
514 at around 40% confluency. The day after seeding, a single dose of 10 Gy was delivered using a
515 CIX3 cabinet x-ray irradiator (xstrahl Life Sciences) (195 kV, 10 mA). Irradiated cells were not
516 passaged, but the medium was changed on Days 4, 8,12 and 16 and cells were either harvested on
517 day 10 or day 20 and used for RNA isolation and Chip experiments respectively. Senescence was
518 confirmed by SA-β-gal staining (Cell Signaling technology #9860) and p21 expression (p21
519 Santacruz Biotechnology sc-397).

520 **Antibodies:**

521 Antibodies used were: CTCF (Cell Signaling Technologies, #3418 for immunoblotting and
522 Millipore Sigma #07-729 for CUT&RUN.RNase), DDX5 (Abcam, ab-21696), Cdkn2A, p16
523 (Abcam ab-108349), ARF, p14 (Santacruz Biotechnology, sc-53639), PolII (Santacruz
524 Biotechnology, sc-55492), GAPDH (Santacruz Biotechnology, sc-32233), Histone H3 (Sigma-
525 Aldrich H0164).

526 **ChIP-seq and CUT&RUN analysis:**

527 Sequenced reads were aligned to the hg19 assembly using default Bowtie2 options. Tag directories
528 were made from the aligned reads using the HOMER (http://homer.ucsd.edu/homer/)
529 makeTagDirectory command. A 200 bp sliding window was used to identify narrow peaks of
530 transcription factors' ChIP-seq and CTCF CUT&RUN. CUT&RUN.RNAse was performed as
531 described in (26). Artifacts from clonal amplification were neglected as only one tag from each
532 unique genomic position was considered (-tbp 1). The threshold was set at a false discovery rate
533 (FDR) of 0.001 determined by peak-finding using randomized tag positions in a genome with an
534 effective size of 2×10^9 bp. For histone marks, seed regions were initially found using a peak size
535 of 500 bp (FDR<0.001) to identify enriched loci. Enriched regions separated by <1kb were merged
536 and considered as blocks of variable lengths. The HOMER makeUCSCfile command was used to
537 generate bedgraph files of the read densities across the genome and this track was uploaded to the
538 UCSC genome browser. The tag directories generated were used to quantify signals at various
539 regions of interest using annotatePeaks.pl. The datasets used in the study are listed in
540 Supplementary Table 3.

541 **GRO-seq analysis:**

542 The sequenced reads were aligned to the hg19 assembly using default Bowtie2 options. Tag
543 directories were made from the aligned reads using the HOMER makeTagDirectory command.
544 Tag directories were used to quantify signal and gene expression. The HOMER makeUCSCfile
545 command was used to generate strand-specific bedgraph files of the read densities across the
546 genome and this track was uploaded to the UCSC genome browser.

547 **TADs and loop domain calling:**

548 Raw paired-end reads were trimmed and individually mapped to the hg19 assembly using bowtie2.
549 The HOMER makeTagDirectory was first used to create tag directories with tbp 1 option. Datasets
550 were further processed by HOMER in order to remove small fragment and self-ligations using
551 makeTagDirectory with the following options: -removePEbg -removeSpikes 10000 5. Next,
552 findTADsAndLoops.pl was used to obtain overlapping TADs (TADs and sub-TADs), produced
553 at 5 kb resolution with 10 kb windows. This program was also used to generate two bedGraph files
554 that describe the directionality index and the insulation score. tagDir2hicFile.pl was used to
555 generate .hic files. The WashU epigenome browser was used to visualize the HiC data. Boundaries
556 of the TADs were defined as 5 kb or 20 kb regions centered at the end of the TADs called.
557 HICCUPS was used to obtain loop anchors from the .hic file at a 5kb resolution. HOMER
558 annotatePeaks.pl was used to quantify signal from tagDirectories, count peaks, quantify insulation
559 and annotate features at the TAD boundaries. Non-genic TAD boundaries were identified as those
560 that did NOT intersect with an annotated gene body. GRO-seq signal was used to rank the
561 boundaries and to then select 10% of unique boundaries in each category (high transcription, low
562 transcription, no transcription and random). The .hic file was converted to .cool. The file was then
563 used to plot the averaged insulation score valleys at 5 kb resolution and 100 kb around the
564 boundary using coolpup.py.

565 **NET-CAGE and CAGE Analysis:**

566 NET-CAGE and CAGE sequenced reads were aligned to hg19 using bowtie2. The HOMER
567 program makeTagDirectory was first used to create tag directories. findPeaks with "–style tss"
568 was used to obtain the NET-CAGE TSSs. TSSs identified within 1 kb of each other on the same
569 strand were merged. NET-CAGE identified TSSs intersecting the refseq annotated TSSs were
570 called mRNA TSSs and the rest were defined as eRNA TSSs. Those TSSs which fell in boundaries
571 were identified as bmRNAs or beRNAs. A span of 2kb around each identified TSSs was used to
572 quantify NET-CAGE and CAGE signal and to calculate the $\log_2$ fold difference between the two.

573 **5' Splice site and Transcription Termination Site Sequence Analysis:**

574 To identify the 5' SS motif, HOMER denovo motif analysis was performed on the 500bp region
575 after the Refseq annotated TSSs with other H3K27ac 500 bp regions as background. The top hit
576 was observed to be the splice acceptor site. The "AATAAA" motif was used as the 3' polyA
577 sequence. HOMER scanMotifGenomeWide was used to identify "predicted" splice sites and
578 polyA terminator sites using previously identified motifs. The occurrence of predicted sites was
579 plotted at the previously identified TSSs.

580 **Circular Chromatin Conformation Capture-seq:**

581 4C was performed as per the protocol described in van de Werken et al., 2012 with minor
582 modifications. HeLa cells were fixed with fresh formaldehyde (1.5%) and quenched with glycine
583 (125mM) followed by washes with ice-cold PBS (2x) and scraped, pelleted and stored at -80^{0}C.
584 Lysis buffer [Tris-Cl pH 8.0 (10mM), NaCl (10mM), NP-40 (0.2%), PIC (1x)] was added to the
585 pellets and homogenized with a Dounce homogenizer (20 stroked with pestle A followed by pestle
586 B). The 3C digestion was performed with HindIII (400units, NEB) and ligation was performed
587 using the T4 DNA ligase and ligation mix (1% Triton X-100, 0.105 mg/ml BSA, 1.05 mM ATP,
588 1x Ligation buffer, where 10x Ligation buffer is 500 mM Tris-Cl pH 7.5, 100 mM $MgCl_2$, 100
589 mM DTT). The ligated samples were then purified by PCI (Phenol/Chloroform/Isoamyl alcohol)
590 extraction, subjected to ethanol precipitation, and the pellet was eluted in TE (pH 8.0) to obtain
591 the 3C library. The 4C digestion was performed using 50 units DpnII (NEB) and the samples were
592 ligated, purified and precipitated similar to the 3C library to obtain the 4C library. The 4C library
593 was subjected to RNaseA treatment and purified by the QIAquick PCR purification kit. The
594 concentration of the library was then measured by Nanodrop and subjected to PCR using the oligos
595 for the enhancer viewpoint. The samples were next PCR purified using the same kit and subjected
596 to Illumina HiSeq2500 sequencing using 50 bp single-end reads.

597 **4C method and analysis:**

598 The sequenced reads were aligned to the hg19 assembly as described in R.4Cker
599 (https://github.com/rr1859/R.4Cker) with Bowtie2. Data analysis was further performed using
600 4C-ker with default parameters. 4Cseqpipe
601 (http://compgenomics.weizmann.ac.il/tanay/?page_id=367) was also used to process the
602 sequenced data. 4C-seq images were generated using truncated mean at a 10kb resolution.

603 **Motif analysis:**

604 annotatePeaks.pl was used with –hist option with the CTCF motif probability matrix to plot the
605 motif occurrence in the peaks.

606 **Promoter capture analysis:**

607 Processed CHiCAGO data was obtained from Thiecke et al. (https://osf.io/brzuc/). The BEDTools
608 suite was used to obtain interactions of interest and perform further analysis. The promoter
609 interacting region of the Bedpe files of pcHiC interactions were then intersected with boundaries
610 to obtain promoters that interact with boundaries. The strength of the interactions obtained from
611 the processed CHiCAGO data was plotted for interactions from boundary interacting promoters to
612 enhancers (H3K27ac regions).

613 **RNA-seq analysis:**

614 The raw reads were mapped to the hg19 assembly using hisat2 in a strand-specific manner. Tag
615 directories were made from the aligned reads using the HOMER makeTagDirectory command.

The HOMER makeUCSCfile command was used to generate strand-specific bedgraph files of the read densities across the genome and this track was uploaded to the UCSC genome browser.

Acknowledgments:

We are grateful to Henrietta Lacks and to her family members for their contributions to biomedical research. We thank DN lab members, Ranveer Jayani, Amanjot Singh for the discussions.

Funding:

The Department of Atomic Energy, Government of India, under project no.12-R&D-TFR-5.04-0800. Intramural funds from NCBS-TIFR (to DN)
DN is a EMBO Global Investigator.
Welcome-IA: IA/1/14/2/501539 (to DN).
UF is supported by CSIR (Govt. of India) senior research fellowship.
KW is supported by the TIFR-NCBS graduate program and CSIR-SPM (Govt. of India) fellowship.

Author contributions:

Conceptualization and design: ZI, BS and DN
NGS analysis: BS
Methodology: ZI, BS, UF and KW
CTCF CUT&RUN: JT and SH
NGS library preparation and sequencing: AP
Writing—original draft: DN, SH, ZI and BS
Reviewing and editing: DN, ZI, BS, KW, UF, JT, AKS, AP, RS

Competing interests: All other authors declare they have no competing interests.

Data and materials availability: The accession codes for data used in the study are mentioned in Table. 3. The CUT&RUN raw and processed data for control and RNaseA- treated HeLa cells have been deposited in the GEO repository with accession code GSE180659. The data can be reviewed using token ehwnowswbbcdxyl into the box.
Goto https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE180659

References

1. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B. 2012. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*. doi:10.1038/nature11082
2. Beagan JA, Phillips-Cremin JE. 2020. On the existence and functionality of topologically associating domains. *Nat Genet.* doi: 10.1038/s41588-019-0561-1.
3. Dekker J, Mirny L. 2016. The 3D Genome as Moderator of Chromosomal Communication. *Cell*. doi:10.1016/j.cell.2016.02.007
4. Fudenberg G, Imakaev M, Lu C, Goloborodko A, Abdennur N, Mirny LA. 2016. Formation of Chromosomal Domains by Loop Extrusion. *Cell Rep*. doi:10.1016/j.celrep.2016.04.085
5. Nuebler J, Fudenberg G, Imakaev M, Abdennur N, Mirny LA. 2018. Chromatin organization by an interplay of loop extrusion and compartmental segregation. *Proc Natl Acad Sci U S A*. doi:10.1073/pnas.1717730115
6. Rao SSP, Huang SC, Glenn St Hilaire B, et al. 2017. Cohesin loss eliminates all loop domains. *Cell*.;171(2):305–320.e24.
7. Szabo Q, Donjon A, Jerković I, Papadopoulos GL, Cheutin T, Bonev B, Nora EP, Bruneau BG, Bantignies F, Cavalli G. 2020. Regulation of single-cell genome organization into TADs and chromatin nanodomains. *Nat Genet*. doi:10.1038/s41588-020-00716-8
8. Nora EP, Goloborodko A, Valton AL, Gibcus JH, Uebersohn A, Abdennur N, Dekker J, Mirny LA, Bruneau BG. 2017. Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell*. doi:10.1016/j.cell.2017.05.004
9. Schwarzer W, Abdennur N, Goloborodko A, et al. 2017. Two independent modes of chromatin organization revealed by cohesin removal. *Nature*. 2017;551(7678):51–56
10. Wutz G, Várnai C, Nagasaka K, et al. 2017. Topologically associating domains and chromatin loops depend on cohesin and are regulated by CTCF, WAPL, and PDS5 proteins. *Embo J*. 2017;36(24):3573–3599
11. Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES, Aiden EL. 2014. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*. doi:10.1016/j.cell.2014.11.021
12. Krivega I, Dean A. 2017. CTCF fences make good neighbours. *Nat Cell Biol*. doi:10.1038/ncb3584
13. Rowley MJ, Nichols MH, Lyu X, Ando-Kuri M, Rivera ISM, Hermetz K, Wang P, Ruan Y, Corces VG. 2017. Evolutionarily Conserved Principles Predict 3D Chromatin Organization. *Mol Cell*. doi:10.1016/j.molcel.2017.07.022
14. Despang A, Schöpflin R, Franke M, Ali S, Jerković I, Paliou C, Chan WL, Timmermann B, Wittler L, Vingron M, Mundlos S, Ibrahim DM. 2019. Functional dissection of the Sox9–Kcnj2 locus identifies nonessential and instructive roles of TAD architecture. *Nat Genet*. doi:10.1038/s41588-019-0466-z

15. Flavahan WA, Drier Y, Liau BB, Gillespie SM, Venteicher AS, Stemmer-Rachamimov AO, Suvà ML, Bernstein BE. 2016. Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature*. doi:10.1038/nature16490

16. Hanssen LLP, Kassouf MT, Oudelaar AM, Biggs D, Preece C, Downes DJ, Gosden M, Sharpe JA, Sloane-Stanley JA, Hughes JR, Davies B, Higgs DR. 2017. Tissue-specific CTCF-cohesin-mediated chromatin architecture delimits enhancer interactions and function in vivo. *Nat Cell Biol*. doi:10.1038/ncb3573

17. Kloetgen A, Thandapani P, Ntziachristos P, Ghebrechristos Y, Nomikou S, Lazaris C, Chen X, Hu H, Bakogianni S, Wang J, Fu Y, Boccalatte F, Zhong H, Paietta E, Trimarchi T, Zhu Y, Van Vlierberghe P, Inghirami GG, Lionnet T, Aifantis I, Tsirigos A. 2020. Three-dimensional chromatin landscapes in T cell acute lymphoblastic leukemia. *Nat Genet*. doi:10.1038/s41588-020-0602-9

18. Narendra V, Rocha PP, An D, Raviram R, Skok JA, Mazzoni EO, Reinberg D. 2015. CTCF establishes discrete functional chromatin domains at the Hox clusters during differentiation. *Science (80-)*. doi:10.1126/science.1262088

19. Sima J, Chakraborty A, Dileep V, Michalski M, Klein KN, Holcomb NP, Turner JL, Paulsen MT, Rivera-Mulia JC, Trevilla-Garcia C, Bartlett DA, Zhao PA, Washburn BK, Nora EP, Kraft K, Mundlos S, Bruneau BG, Ljungman M, Fraser P, Ay F, Gilbert DM. 2019. Identifying cis Elements for Spatiotemporal Control of Mammalian DNA Replication. *Cell*. doi:10.1016/j.cell.2018.11.036

20. Williamson I, Kane L, Devenney PS, Flyamer IM, Anderson E, Kilanowski F, Hill RE, Bickmore WA, Lettice LA. 2019. Developmentally regulated Shh expression is robust to TAD perturbations. *Dev*. doi:10.1242/dev.179523

21. Yang M, Vesterlund M, Siavelis I, Moura-Castro LH, Castor A, Fioretos T, Jafari R, Lilljebjörn H, Odom DT, Olsson L, Ravi N, Woodward EL, Harewood L, Lehtiö J, Paulsson K. 2019. Proteogenomics and Hi-C reveal transcriptional dysregulation in high hyperdiploid childhood acute lymphoblastic leukemia. *Nat Commun*. doi:10.1038/s41467-019-09469-3

22. Khoury A, Achinger-Kawecka J, Bert SA, Smith GC, French HJ, Luu PL, Peters TJ, Du Q, Parry AJ, Valdes-Mora F, Taberlay PC, Stirzaker C, Statham AL, Clark SJ. 2020. Constitutively bound CTCF sites maintain 3D chromatin architecture and long-range epigenetically regulated domains. *Nat Commun*. doi:10.1038/s41467-019-13753-7

23. Bonev B, Mendelson Cohen N, Szabo Q, Fritsch L, Papadopoulos GL, Lubling Y, Xu X, Lv X, Hugnot JP, Tanay A, Cavalli G. 2017. Multiscale 3D Genome Rewiring during Mouse Neural Development. *Cell*. doi:10.1016/j.cell.2017.09.043

24. Hsieh THS, Cattoglio C, Slobodyanyuk E, Hansen AS, Rando OJ, Tjian R, Darzacq X. 2020. Resolving the 3D Landscape of Transcription-Linked Mammalian Chromatin Folding. *Mol Cell*. doi:10.1016/j.molcel.2020.03.002

25. Zhang D, Huang P, Sharma M, Keller CA, Giardine B, Zhang H, Gilgenast TG, Phillips-Cremin JE, Hardison RC, Blobel GA. 2020. Alteration of genome folding via contact domain boundary insertion. *Nat Genet*. doi:10.1038/s41588-020-0680-8

26. Giorgetti L, Lajoie BR, Carter AC, Attia M, Zhan Y, Xu J, Chen CJ, Kaplan N, Chang HY, Heard E, Dekker J. 2016. Structural organization of the inactive X chromosome in the mouse. *Nature*. doi:10.1038/nature18589

27. Zhang Y, Li T, Preissl S, Amaral ML, Grinstein JD, Farah EN, Destici E, Qiu Y, Hu R, Lee AY, Chee S, Ma K, Ye Z, Zhu Q, Huang H, Fang R, Yu L, Izpisua Belmonte JC, Wu J, Evans SM, Chi NC, Ren B. 2019. Transcriptionally active HERV-H retrotransposons demarcate topologically associating domains in human pluripotent stem cells. *Nat Genet*. doi:10.1038/s41588-019-0479-7

28. Barutcu AR, Blencowe BJ, Rinn JL. 2019. Differential contribution of steady-state RNA and active transcription in chromatin organization . *EMBO Rep*. doi:10.15252/embr.201948068

29. Thakur J, Fang H, Llagas T, Disteche CM, Henikoff S. 2019. Architectural RNA is required for heterochromatin organization. *bioRxiv*. doi: https://doi.org/10.1101/784835

30. Swygert SG, Lin D, Portillo-Ledesma S, Lin PY, Hunt DR, Kao CF, Schlick T, Noble WS, Tsukiyama T. 2020. Chromatin Fiber Folding Represses Transcription and Loop Extrusion in Quiescent Cells. *bioRxiv*. doi: https://doi.org/10.1101/2020.11.24.396713

31. Hansen AS, Hsieh THS, Cattoglio C, Pustova I, Saldaña-Meyer R, Reinberg D, Darzacq X, Tjian R. 2019. Distinct Classes of Chromatin Loops Revealed by Deletion of an RNA-Binding Region in CTCF. *Mol Cell*. doi:10.1016/j.molcel.2019.07.039

32. Saldaña-Meyer R, Rodriguez-Hernaez J, Escobar T, Nishana M, Jácome-López K, Nora EP, Bruneau BG, Tsirigos A, Furlan-Magaril M, Skok J, Reinberg D. 2019. RNA Interactions Are Essential for CTCF-Mediated Genome Organization. *Mol Cell*. doi:10.1016/j.molcel.2019.08.015

33. Nandy Mazumdar M, Yin S, Paranjapye A, Kerschner JL, Swahn H, Ge A, Leir SH, Harris A. 2020. Looping of upstream cis-regulatory elements is required for CFTR expression in human airway epithelial cells. *Nucleic Acids Res*. doi:10.1093/nar/gkaa089

34. Gong Y, Lazaris C, Sakellaropoulos T, Lozano A, Kambadur P, Ntziachristos P, Aifantis I, Tsirigos A. 2018. Stratification of TAD boundaries reveals preferential insulation of super-enhancers by strong boundaries. *Nat Commun*. doi:10.1038/s41467-018-03017-1

35. Sun F, Chronis C, Kronenberg M, Chen XF, Su T, Lay FD, Plath K, Kurdistani SK, Carey MF. 2019. Promoter-Enhancer Communication Occurs Primarily within Insulated Neighborhoods. *Mol Cell*. doi:10.1016/j.molcel.2018.10.039

36. Matthews BJ, Waxman DJ. 2018. Computational prediction of CTCF/cohesin-based intra-TAD loops that insulate chromatin contacts and gene expression in mouse liver. *Elife*. 2018 May 14;7:e34077. doi: 10.7554/eLife.34077

37. Andersson R, Gebhard C, Miguel-Escalada I. et al. 2014. An atlas of active enhancers across human cell types and tissues. *Nature* 507, 455–461 (2014). https://doi.org/10.1038/nature12787

38. Skene PJ, Henikoff S. 2017. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *Elife*. 2017 Jan 16;6:e21856. doi: 10.7554/eLife.21856

39. Zhang J, Lee D, Dhiman V, Jiang P, Xu Jie, McGillivray P, Yang H, Liu J, Meyerson W, Clarke D, Gu M, Li S, Lou S, Xu Jinrui, Lochovsky L, Ung M, Ma L, Yu S, Cao Q,

Harmanci A, Yan KK, Sethi A, Gürsoy G, Schoenberg MR, Rozowsky J, Warrell J, Emani P, Yang YT, Galeev T, Kong X, Liu S, Li X, Krishnan J, Feng Y, Rivera-Mulia JC, Adrian J, Broach JR, Bolt M, Moran J, Fitzgerald D, Dileep V, Liu T, Mei S, Sasaki T, Trevilla-Garcia C, Wang S, Wang Y, Zang C, Wang D, Klein RJ, Snyder M, Gilbert DM, Yip K, Cheng C, Yue F, Liu XS, White KP, Gerstein M. 2020. An integrative ENCODE resource for cancer genomics. *Nat Commun*. doi:10.1038/s41467-020-14743-w

40. Zirkel A, Nikolic M, Sofiadis K, Mallm JP, Brackley CA, Gothe H, Drechsel O, Becker C, Altmüller J, Josipovic N, Georgomanolis T, Brant L, Franzen J, Koker M, Gusmao EG, Costa IG, Ullrich RT, Wagner W, Roukos V, Nürnberg P, Marenduzzo D, Rippe K, Papantonis A. 2018. HMGB2 Loss upon Senescence Entry Disrupts Genomic Organization and Induces CTCF Clustering across Cell Types. *Mol Cell*. doi:10.1016/j.molcel.2018.03.030

41. Farooq U, Saravanan B, Islam Z, Walavalkar K, Singh AK, Jayani RS, Meel S, Swaminathan S, Notani D. 2021. An interdependent network of functional enhancers regulates transcription and EZH2 loading at the INK4a/ARF locus. *Cell Rep*. doi: 10.1016/j.celrep.2021.108898.

42. Harismendy O, Notani D, Song X, Rahim NG, Tanasa B, Heintzman N, Ren B, Fu XD, Topol EJ, Rosenfeld MG, Frazer KA. 2011. 9p21 DNA variants associated with coronary artery disease impair interferon-γ 3 signalling response. *Nature*. doi:10.1038/nature09753

43. Bouvy-Liivrand M, De Sande AH, Pölönen P, Mehtonen J, Vuorenmaa T, Niskanen H, Sinkkonen L, Kaikkonen MU, Heinäniemi M. 2017. Analysis of primary microRNA loci from nascent transcriptomes reveals regulatory domains governed by chromatin architecture. *Nucleic Acids Res*. doi:10.1093/nar/gkx680

44. Nilson KA, Lawson CK, Mullen NJ, Ball CB, Spector BM, Meier JL, Price DH. 2017. Oxidative stress rapidly stabilizes promoter-proximal paused Pol II across the human genome. *Nucleic Acids Res*. doi:10.1093/nar/gkx724

45. Schlackow M, Nojima T, Gomes T, Dhir A, Carmo-Fonseca M, Proudfoot NJ. 2017. Distinctive Patterns of Transcription and RNA Processing for Human lincRNAs. *Mol Cell*. doi:10.1016/j.molcel.2016.11.029

46. Thiecke MJ, Wutz G, Muhar M, Tang W, Bevan S, Malysheva V, Stocsits R, Neumann T, Zuber J, Fraser P, Schoenfelder S, Peters JM, Spivakov M. 2020. Cohesin-Dependent and -Independent Mechanisms Mediate Chromosomal Contacts between Promoters and Enhancers. *Cell Rep*. doi:10.1016/j.celrep.2020.107929

47. McLaughlin-Drubin ME, Park D, Munger K. 2013. Tumor suppressor p16INK4A is necessary for survival of cervical carcinoma cell lines. *Proc Natl Acad Sci U S A*. 2013;110(40):16175-16180. doi:10.1073/pnas.1310432110

48. Pauck A, Lener B, Hoell M, Kaiser A, Kaufmann AM, Zwerschke W, Jansen-Dürr P. 2014. Depletion of the cdk inhibitor p16INK4a differentially affects proliferation of established cervical carcinoma cells. *J Virol*. 2014 May;88(10):5256-62. doi: 10.1128/JVI.03817-13.

49. Li W, Notani D, Ma Q, Tanasa B, Nunez E, Chen AY, Merkurjev D, Zhang J, Ohgi K, Song X, Oh S, Kim HS, Glass CK, Rosenfeld MG. 2013. Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. *Nature*. doi:10.1038/nature12210

50. Li W, Notani D, Rosenfeld MG. 2016. Enhancers as non-coding RNA transcription units: Recent insights and future perspectives. *Nat Rev Genet*. doi:10.1038/nrg.2016.4

51. Quinodoz SA, Ollikainen N, Tabak B, Palla A, Schmidt JM, Detmar E, Lai MM, Shishkin AA, Bhat P, Takei Y, Trinh V, Aznauryan E, Russell P, Cheng C, Jovanovic M, Chow A, Cai L, McDonel P, Garber M, Guttman M. 2018. Higher-Order Inter-chromosomal Hubs Shape 3D Genome Organization in the Nucleus. *Cell*. 2018 Jul 26;174(3):744-757.e24. doi: 10.1016/j.cell.2018.05.024.

52. Beagrie RA, Scialdone A, Schueler M, Kraemer DC, Chotalia M, Xie SQ, Barbieri M, de Santiago I, Lavitas LM, Branco MR, Fraser J, Dostie J, Game L, Dillon N, Edwards PA, Nicodemi M, Pombo A. 2017. Complex multi-enhancer contacts captured by genome architecture mapping. *Nature*. 2017 Mar 23;543(7646):519-524. doi: 10.1038/nature21411.

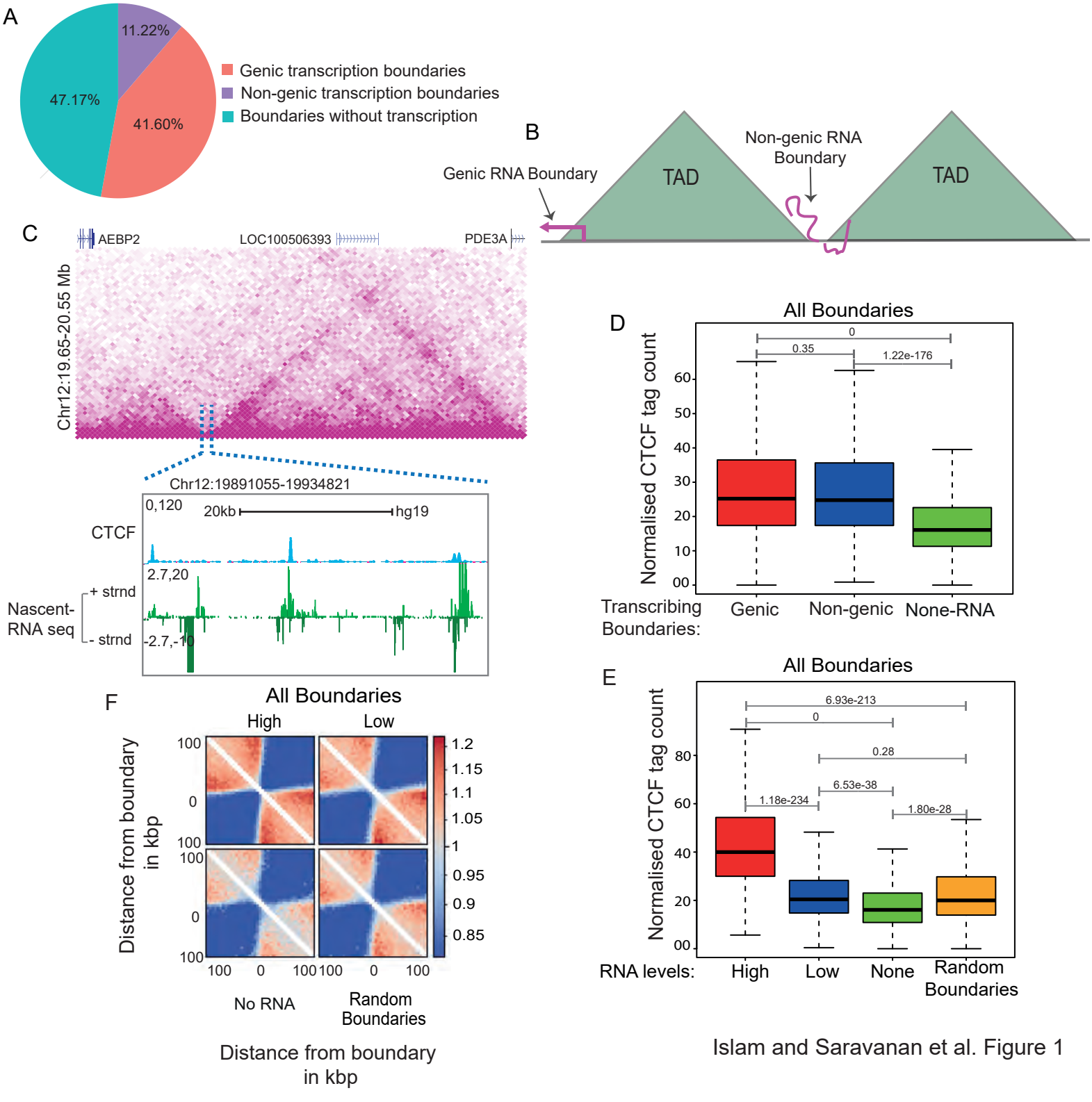
53. Hansen AS. 2020. CTCF as a boundary factor for cohesin-mediated loop extrusion: evidence for a multi-step mechanism. *Nucleus*. doi:10.1080/19491034.2020.1782024

54. Kentepozidou E, Aitken SJ, Feig C, Stefflova K, Ibarra-Soria X, Odom DT, Roller M, Flicek P. 2020. Clustered CTCF binding is an evolutionary mechanism to maintain topologically associating domains. *Genome Biol*. doi:10.1186/s13059-019-1894-x

55. Hug CB, Grimaldi AG, Kruse K, Vaquerizas JM. 2017. Chromatin Architecture Emerges during Zygotic Genome Activation Independent of Transcription. *Cell*. doi:10.1016/j.cell.2017.03.024

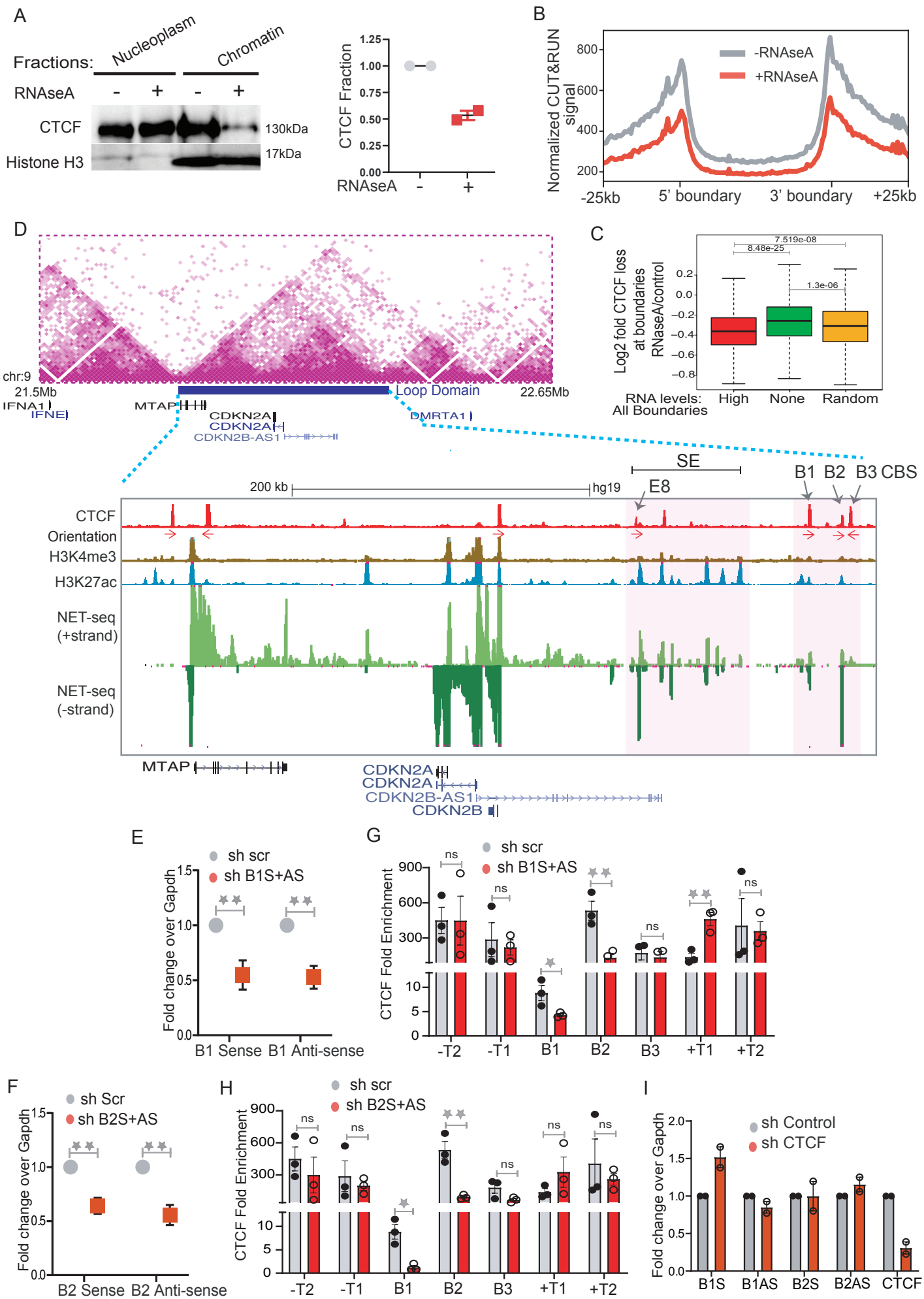
56. Jayani RS, Singh A, Notani D. 2017. Isolation of nuclear RNA-associated protein complexes. *Methods in Molecular Biology*. doi:10.1007/978-1-4939-6716-2_9

57. Rahnamoun H, Lee J, Sun Z, Lu H, Ramsey KM, Komives EA, Lauberth SM. 2018. RNAs interact with BRD4 to promote enhanced chromatin engagement and transcription activation. *Nat Struct Mol Biol*. doi:10.1038/s41594-018-0102-0



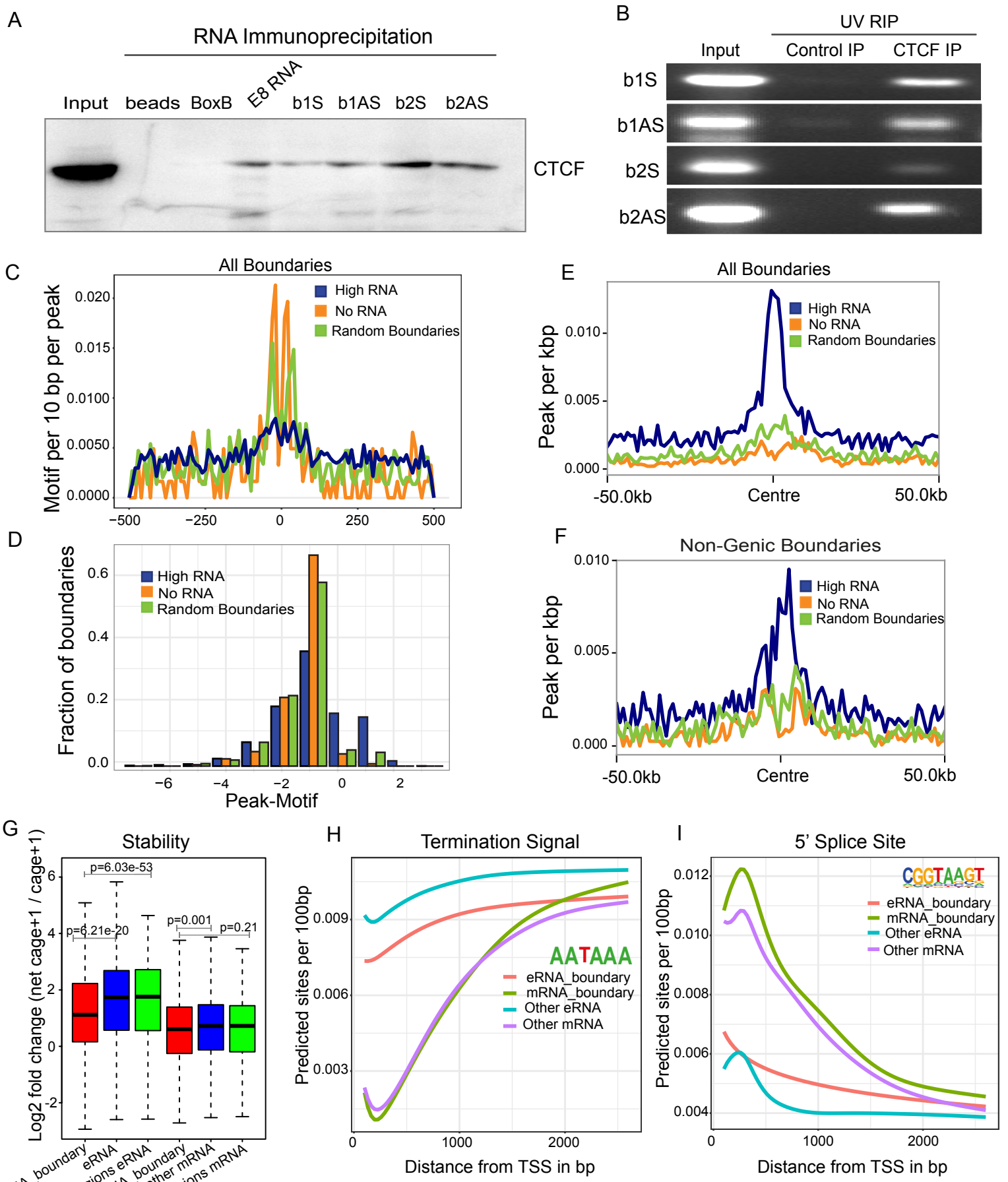
Islam and Saravanan et al. Figure 1

Figure 1. Transcribed TAD boundaries insulate better than non-transcribed boundaries: A. Pie chart displaying the percentage of boundaries that show genic transcription, non-genic de novo transcription and no transcription. **B.** Schematic depicting genic and non-genic transcribed boundaries. **C.** Browser shot displaying a TAD structure from HiC data. The zoomed in box shows a *de novo* non-genic transcribed boundary region that is overlaid with CTCF and mNET-seq signal. **D.** Box plots showing CTCF enrichment on genic, non-genic and non-transcribed boundaries. **E.** Box plots showing CTCF enrichment on all boundaries exhibiting varying levels of RNAs versus non-transcribed and random boundaries. **F.** Pile up (Aggregated normalized HiC interactions) plot centered at high transcribed, low transcribed, non-transcribed and random boundaries at a 5 kb resolution. The 100 kb distances are taken from the boundary region. The p-values in boxplots were calculated using the Wilcoxon rank-sum test. The boxplots depict the minimum (Q1-1.5*IQR), first quartile, median, third quartile and maximum (Q3+1.5*IQR) without outliers.



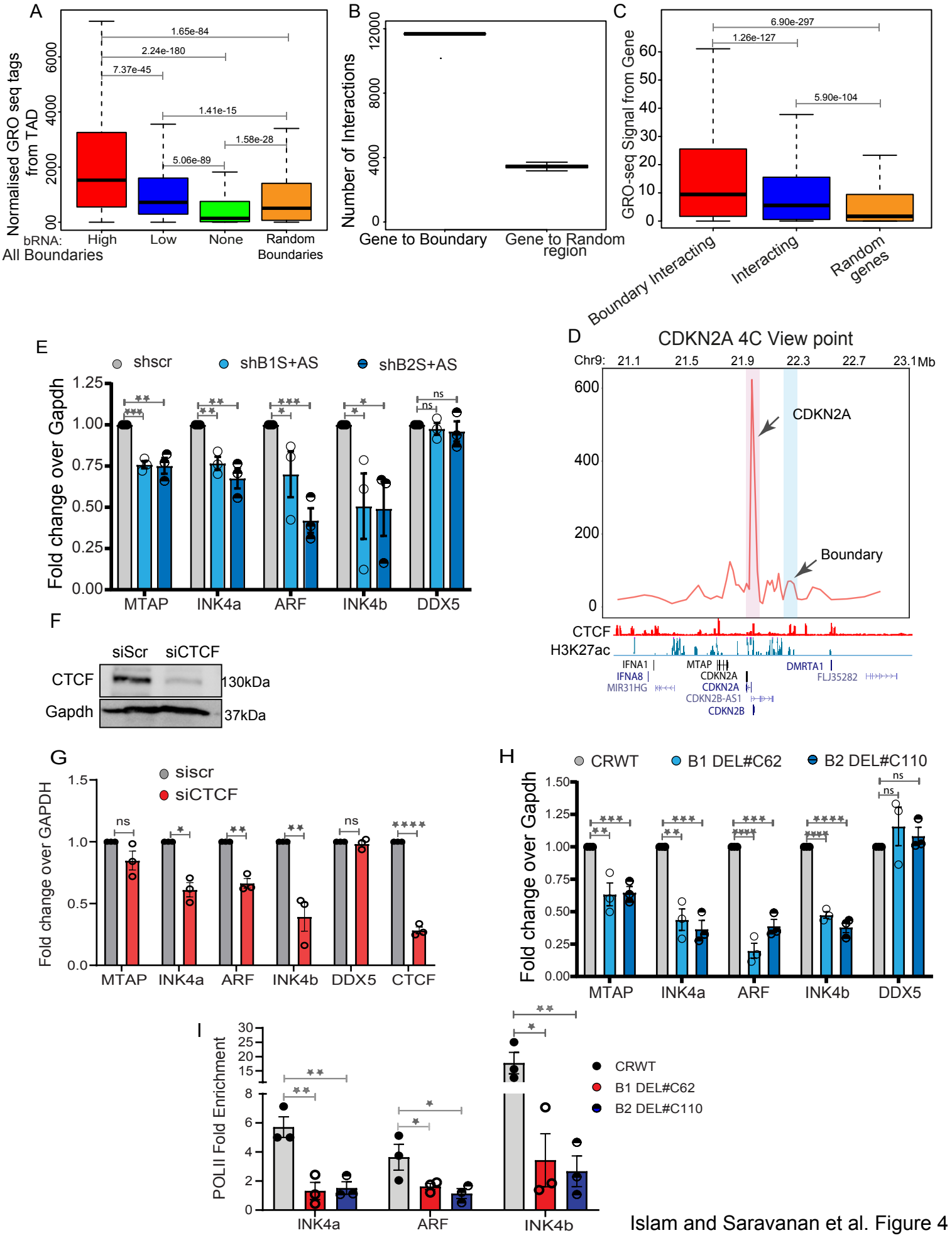
Islam and Saravanan et al Figure 2

Figure 2. RNA increases CTCF occupancy at TAD boundaries: A. Immunoblotting with CTCF and Histone H3 on soluble nucleoplasm and chromatin bound fractions of the nucleus treated with or without RNaseA (Left panel). The right panel shows the quantified loss of CTCF from chromatin fractions from two replicates. **B.** Normalised CTCF CUT&RUN signal in control and RNAseA-treated conditions plotted at equally scaled TADs of 50 kb length each and 25 kb flanks**. C.** Box plot showing the fold loss in CTCF upon RNAseA treatment at boundaries with high RNA levels, non-transcribed and random boundaries. **D.** TAD structure at *INK4a/ARF* locus on 9p21 region (Rao et al., 2014). The loop domain (Blue bar) is overlaid by the position of CTCF peaks, CTCF motif orientation, H3K4me3, H3K27ac (enhancers), mNET-seq tracks and gene annotations. The highlighted region shows the B1, B2 and B3 CTCF sites on the 3' boundary and super-enhancer with the position of enhancer E8 marked by an arrow. **E.** qRT-PCRs showing the levels of sense and anti-sense non-coding RNA at B1 CTCF site upon B1 RNA knockdown. **F.** qRT-PCRs showing the levels of sense and anti-sense RNA at the B2 CTCF site upon B2 RNA knockdown. **G.** CTCF ChIP enrichment before and after shRNA mediated knockdown of sense and anti-sense RNA at B1 CTCF sites. The bars show CTCF fold enrichment on three CTCF sites (B1, B2 and B3) at the 3' boundary, on -T1 (5' boundary) of the *INK4a/ARF* TAD, on -T2 on the adjacent TAD boundary upstream, +T1 (on the adjacent TAD boundary downstream) and +T2, the boundary of the following TAD downstream. **H.** CTCF ChIP enrichment before and after shRNA mediated knockdown of sense and anti-sense RNA at B2 CTCF sites. The bars show CTCF on three CTCF sites (B1, B2 and B3) at the 3' boundary, on -T1 (5' boundary) of the *INK4a/ARF* TAD, on -T2 of the adjacent TAD boundary upstream, on +T1 of the adjacent TAD boundary downstream, and on +T2 of the boundary of the following TAD downstream. **I.** qRT-PCRs of sense and anti-sense RNAs from B1 and B2 CTCF sites upon CTCF knockdown. The drops in CTCF levels are shown in the last two bars. Error bars denote SEM from three biological replicates. *p*-values were calculated by Student's two-tailed unpaired *t*-test in **E, F, G and H .** $^{**}p < 0.01$, $^{*}p < 0.05$, $^{ns}p > 0.05$.



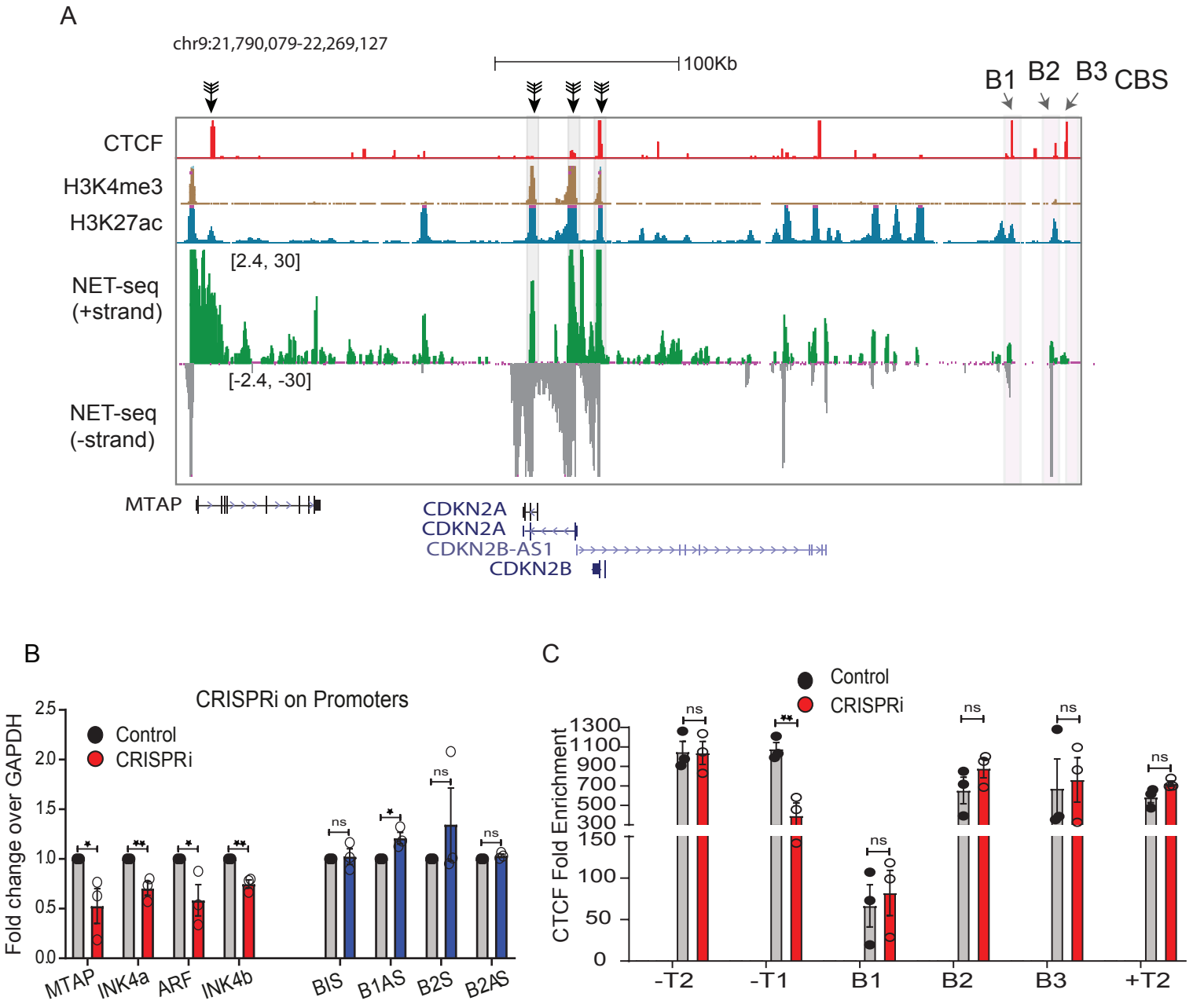
Islam and Saravanan et al. Figure 3

Figure 3. Non-coding RNAs and CTCF interact at the *INKa/ARF* TAD boundary:
A. Immunoblot for CTCF on RNA pulldowns performed on various RNAs (n=2). **B.** Agarose gel image showing the UV-RIP RT-PCRs for sense and anti-sense RNA arising from B1 and B2 CTCF sites (n=2). **C.** Line plots show the CTCF motif per 10 base pair per peak on various classes of boundaries. **D.** Fraction of boundaries with differential occupancy of CTCF peaks over the presence of the CTCF motif**.** **E.** CTCF peaks per kb on all transcribed boundaries with high RNA, no RNAs and random boundaries (50 kb flanks on both side). **F.** CTCF peaks per kb on non-genic transcribed boundaries with high RNAs, no RNAs and random boundaries. **G.** Log_2FC of NET-CAGE+1/CAGE+1, indicating stability of RNA at TSSs identified from NET-CAGE at boundaries and TSSs not at boundaries. eRNA denotes the eRNA from non-genic boundaries, mRNA denotes the mRNA from genic boundaries **H.** Average predicted TTS sites from the TSSs identified in different classes of RNA by NET-CAGE. **I.** Average predicted 5'SS sites from the TSSs identified by NET-CAGE. The p-values in boxplots were calculated by the Wilcoxon rank-sum test. The boxplots depict the minimum (Q1-1.5*IQR), first quartile, median, third quartile and maximum (Q3+1.5*IQR) without outliers.



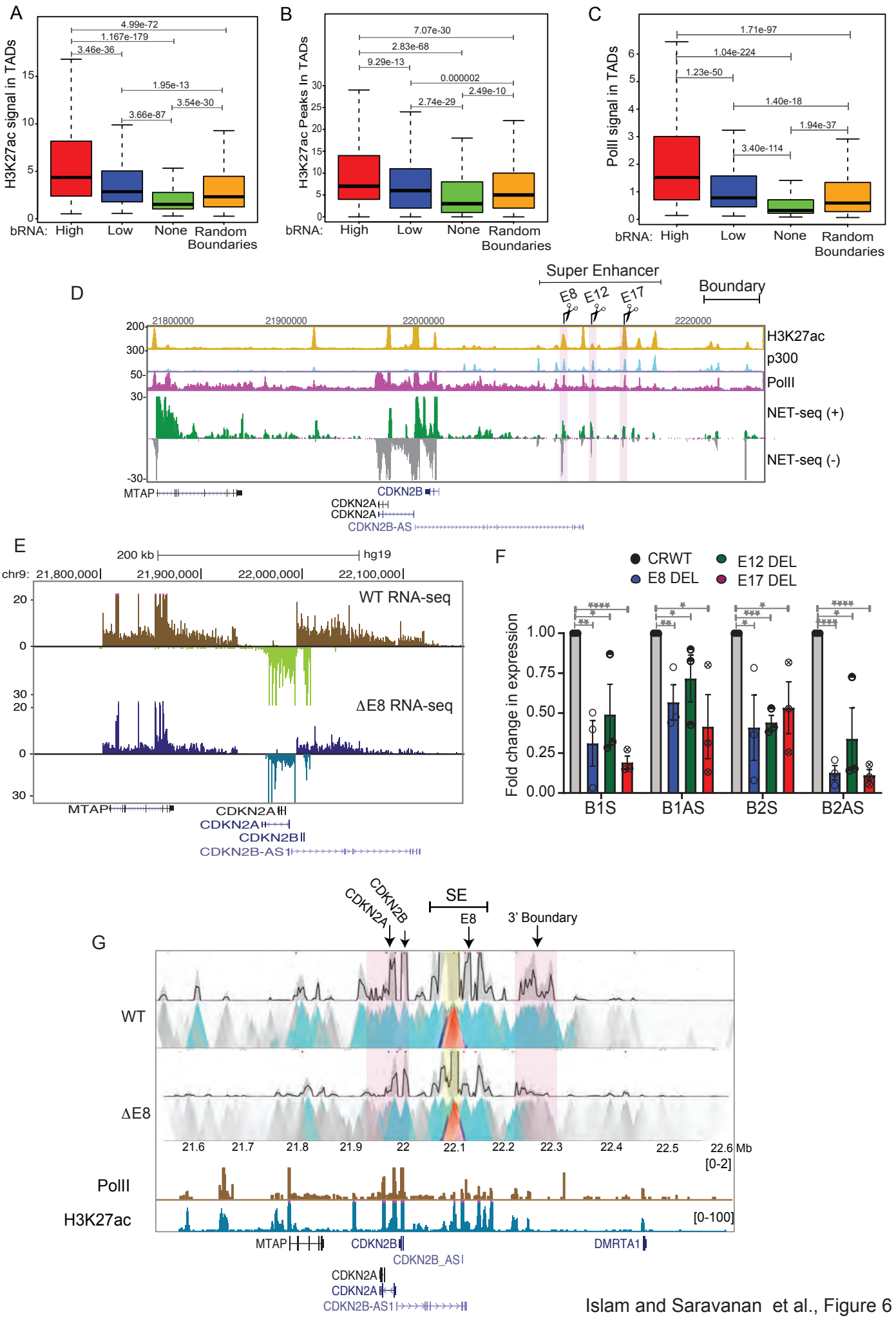
Islam and Saravanan et al. Figure 4

Figure 4. Boundary RNAs enhance TAD transcription: A. Box plots showing GRO-seq tags from the entire TAD with high, low, non-transcribed and random boundaries. **B.** Graph showing the number of interactions between genes and boundaries, and 100 randomizations of interactions between boundaries and random regions. **C.** Box plots showing the GRO-seq tag counts from genes when they interact with boundaries or any region vs. random interactions. **D.** 4C plot at the *CDKN2A* promoter viewpoint, where the interaction with the boundary is highlighted in blue. Below are H3K27ac, CTCF and gene annotations for HeLa cells. **E.** qRT-PCR plot shows the change in gene expression upon shRNA mediated knockdowns of boundary RNAs arising from B1 and B2 CTCF sites at the *INK4a/ARF* TAD boundary. **F.** Immunoblot showing the drop in CTCF protein levels upon its siRNA mediated knockdown. **G**. qRT-PCRs showing down-regulation of *INK4a/ARF* genes and levels of DDX5 and CTCF upon CTCF knockdown. **H.** qRT-PCR showing the changes in expression of different genes in the *INK4a/ARF* TAD and DDX5 upon Cas9-mediated independent deletion of B1 and B2 CTCF sites. **I.** PolII ChIP signal at *INK4a*, *ARF* and *INK4b* promoters in WT and in cells with B1 and B2 CTCF sites deletions. Error bars denote SEM from three biological replicates. *p*-values were calculated by Student's two-tailed unpaired *t*-test in **E, G, H** and **I.** ****$p < 0.0001$, ***$p < 0.001$, **$p < 0.01$, *$p < 0.05$, $^{ns}p > 0.05$. p-values in boxplots were calculated by the Wilcoxon two-sided rank-sum test. The boxplots depict the minimum (Q1-1.5*IQR), first quartile, median, third quartile and maximum (Q3+1.5*IQR) without outliers.



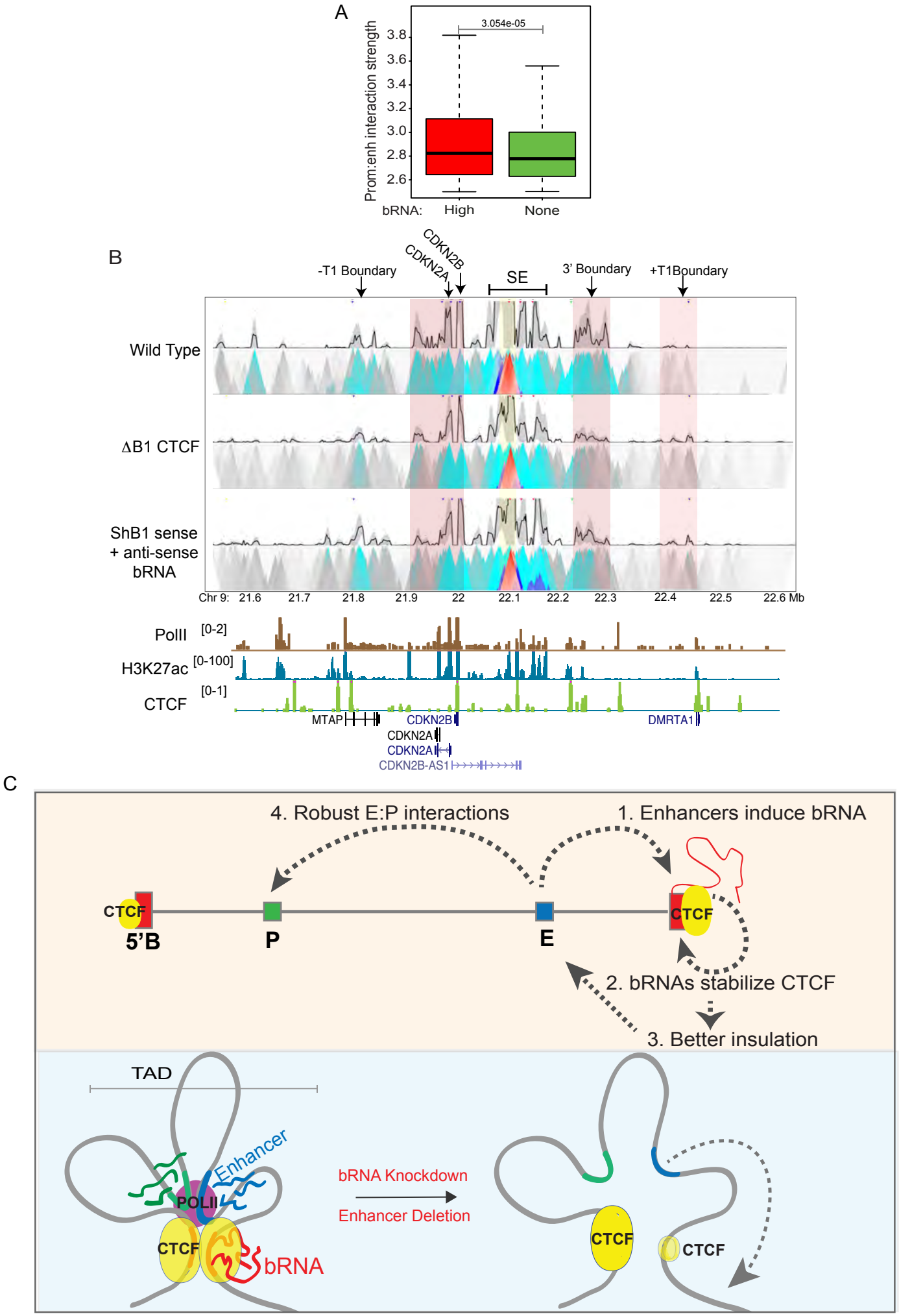
Islam and Saravanan et al. Figure 5

Figure 5. Transcriptional activity of promoters does not regulate non-genic transcription at boundaries: A. Browser shots displaying the promoters targeted by CRISPRi (Black arrows), the track is overlaid with CTCF, H3K4me3, H3K27ac and mNET-seq tracks. **B.** qRT-PCR assessing the changes in *INK4a/ARF* genes, sense and anti-sense boundary RNA arising from B1 and B2 CTCF sites upon CRISPRi on promoters. **C.** CTCF enrichment on boundaries of *INK4a/ARF* and its flanking TADs upon CRISPRi on *INK4a/ARF* promoters. Error bars denote SEM from three biological replicates. *p*-values were calculated by Student's two-tailed unpaired *t*-test in **B** and **C.** $^{**}p < 0.01$, $^{*}p < 0.05$, $^{ns}p > 0.05$.



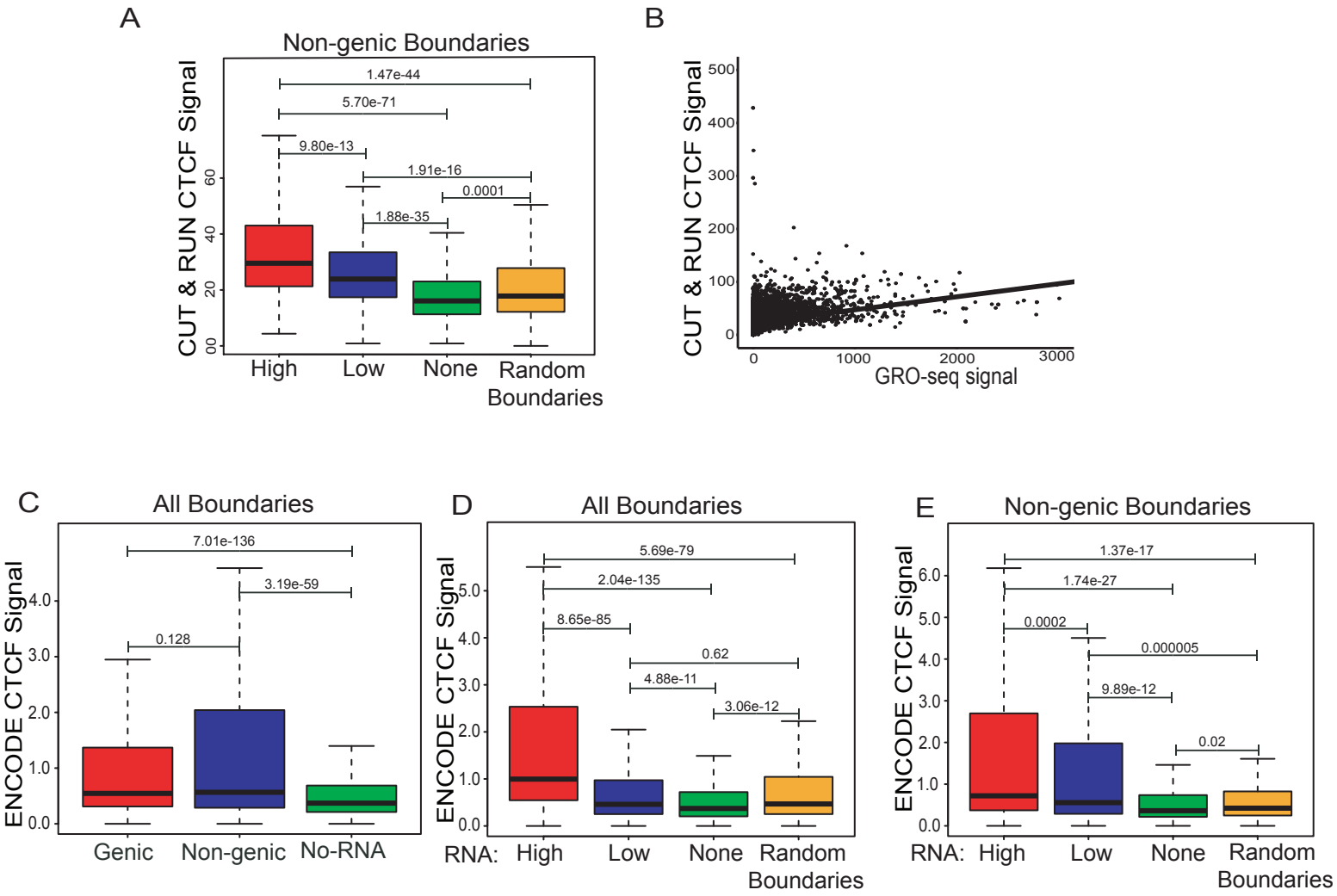
Islam and Saravanan et al., Figure 6

Figure 6. Active enhancers within a TAD directly regulate boundary eRNA transcription : A. Boxplots displaying enrichment of H3K27ac in TADs with high, low, non-transcribed and random boundaries. **B.** Boxplot showing number of H3K27ac peaks in TADs with various boundaries. **C.** Levels of PolII in TADs with different boundaries. **D.** UCSC browser shot showing H3K27ac, p300, PolII ChIP-seq and mNET-seq on the *INK4a/ARF* TAD. The tracks are overlaid by the gene annotations. The highlighted regions represent the enhancers (E8, E12 and E17) that were deleted. **E.** Browser shot shows RNA-seq at *INK4a/ARF* TAD in WT and E8 enhancer delete lines. **F.** qRT-PCR shows levels of sense and anti-sense boundary RNA arising from B1 and B2 CTCF sites upon E8, E12 and E17 enhancer deletions. **G.** 4C-seq plot in WT and in E8 deletion lines on an enhancer with the E8 enhancer as viewpoint (Yellow highlight) exhibiting interactions with promoters and 3' boundary region (Pink highlight). Below are PolII, H3K27ac ChIP-seq tracks and gene annotations. Error bars denote SEM from three biological replicates. *p*-values were calculated by Student's two-tailed unpaired *t*-test in **F.** ****$p < 0.0001$, ***$p < 0.001$, **$p < 0.01$, *$p < 0.05$. p-values in boxplots were calculated by the Wilcoxon rank-sum test. The boxplots depict the minimum (Q1-1.5*IQR), first quartile, median, third quartile and maximum (Q3+1.5*IQR) without outliers.



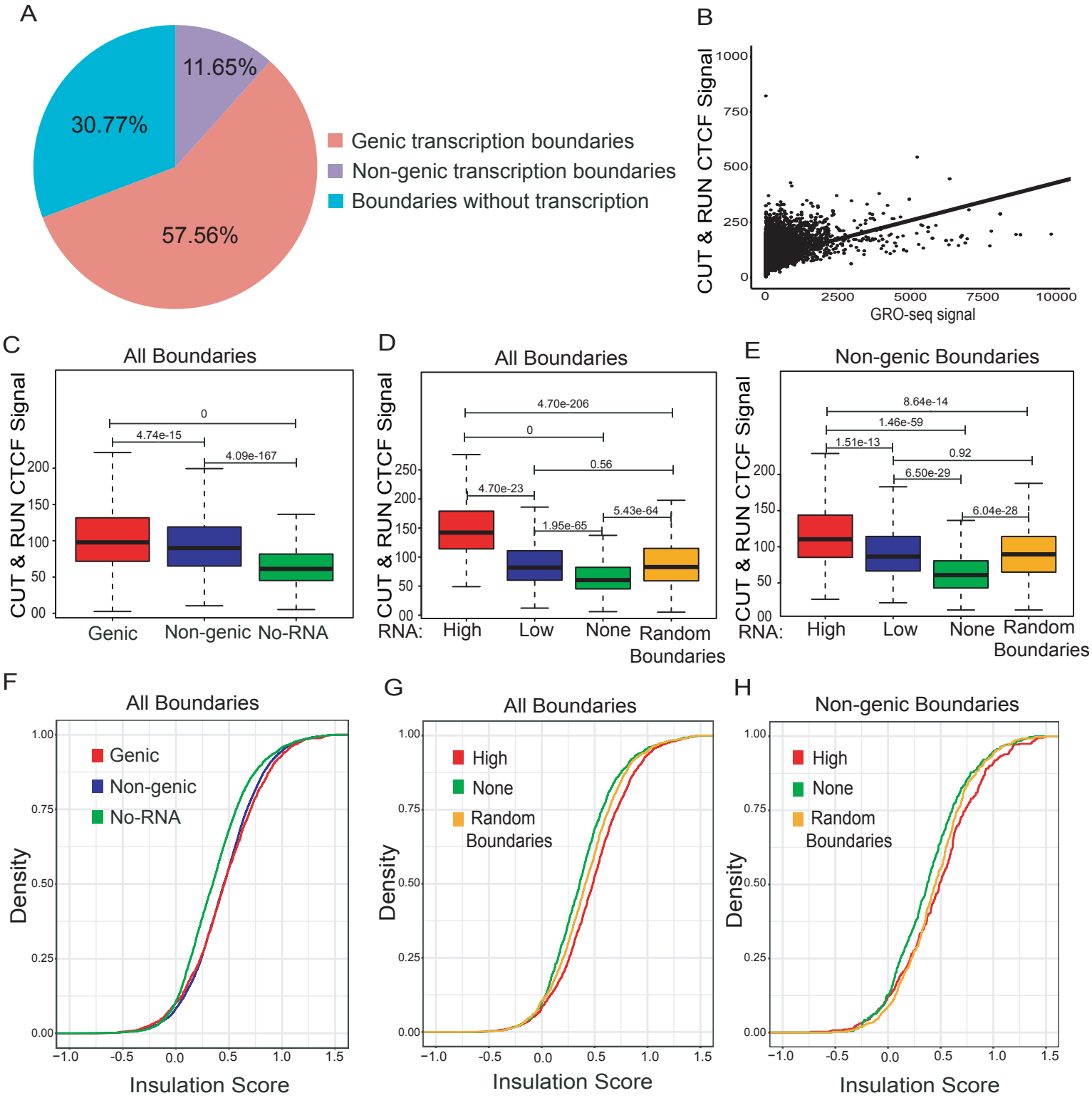
Islam and Saravanan et al. Figure 7

Figure 7. Boundary eRNAs strengthen TAD insulation: A. Enhancer:promoter interaction strength within TADs with high and non-transcribed boundaries. **B.** 4C-seq plot with the E8 enhancer as viewpoint (Yellow highlight) in WT and upon B1 CTCF site deletion (ΔB1) and knockdown of B1 sense and anti-sense boundary RNAs. The anchored region exhibit interactions with promoters and the 3' and +T1 boundaries (Pink highlights). Below are PolII, H3K27ac and CTCF ChIP-seq tracks and gene annotations. **C.** Model: (1) An active enhancer physically interacts with a boundary and activates boundary RNA transcription. (2) The boundary RNA in return stabilizes the CTCF at these boundaries, thereby (3) strengthening insulation of these TADs. (4) This favors intra-TAD enhancer: promoter interactions to facilitate robust gene transcription. The loss of boundary RNA/enhancer deletion reduces the boundary RNA levels which triggers the loss of CTCF and insulation of TADs. Enhancer:promoter interactions weaken in these scenarios which causes concomitant loss of gene transcription. p-values in boxplots were calculated by the Wilcoxon rank-sum test. The boxplots depict the minimum (Q1-1.5*IQR), first quartile, median, third quartile and maximum (Q3+1.5*IQR) without outliers.



Islam and Saravanan et al Sup Fig. 1

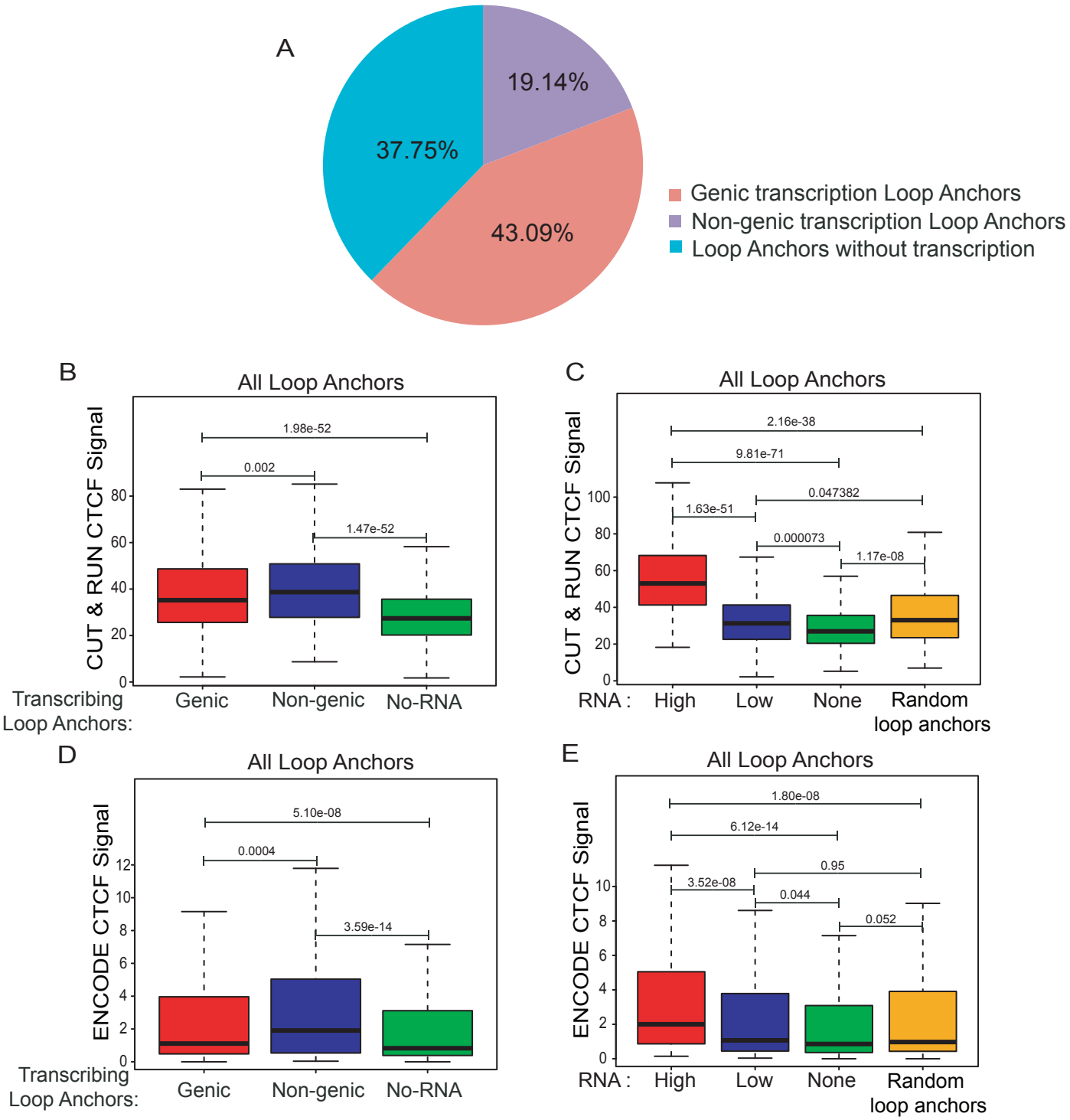
Supp Figure 1. Transcribed boundaries exhibit better insulation: A. CTCF enrichment (CTCF CUT&RUN) at non-genic transcribed boundaries with high, low RNA versus non-transcribed and random boundaries **B.** levels of RNA and CTCF occupancy at TAD boundaries are positively correlated. **C.** CTCF enrichment (ChIP-seq) at transcribed genic, non-genic and non-transcribed boundaries. **D.** CTCF enrichment at all transcribed boundaries with high, low RNA versus non-transcribed and random boundaries. **E.** CTCF enrichment at non-genic transcribed boundaries with high, low non-coding RNA versus non-transcribed and random boundaries. p-values in boxplots were calculated by the Wilcoxon rank-sum test. The boxplots depict the minimum (Q1-1.5*IQR), first quartile, median, third quartile and maximum (Q3+1.5*IQR) without outliers.



Islam and Saravanan et al Sup Fig. 2

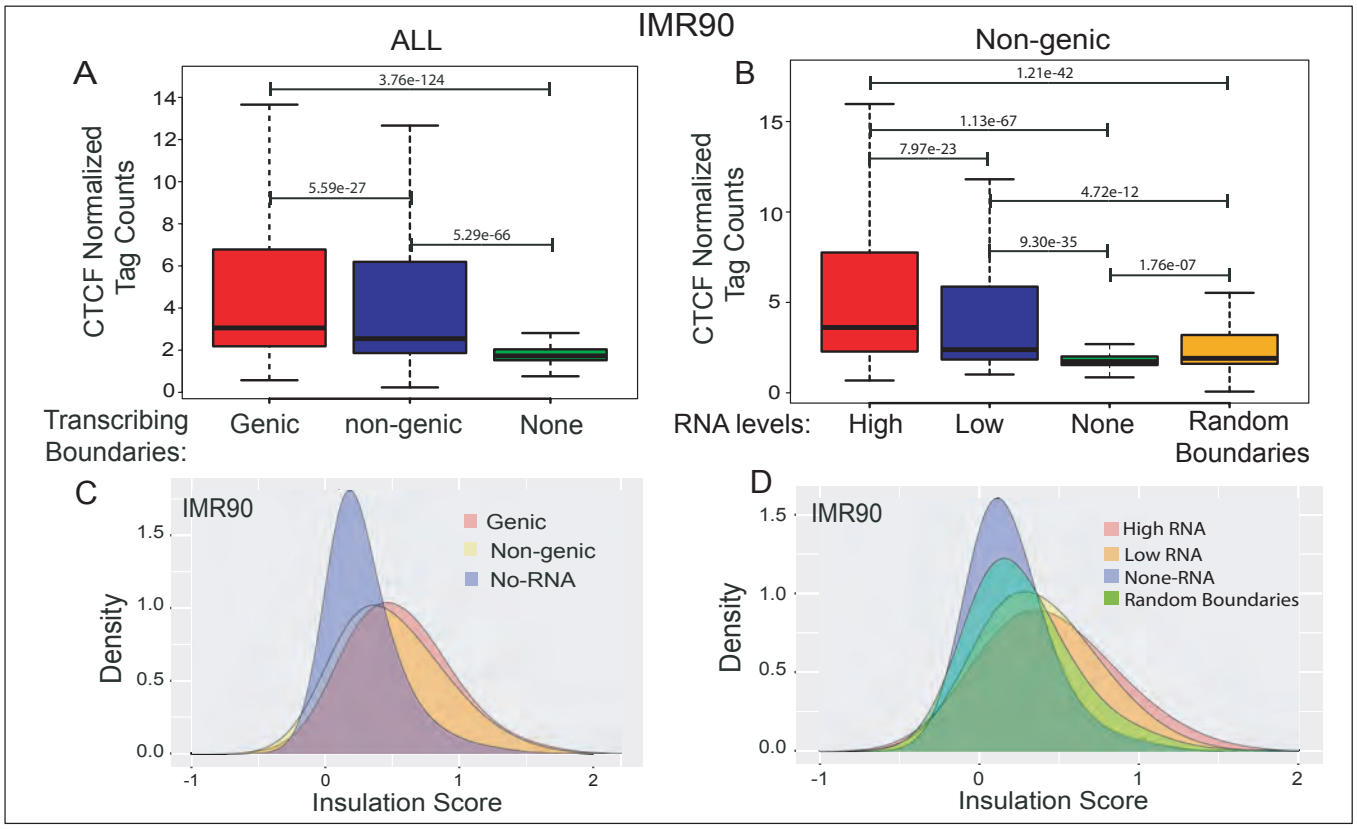
Supp Figure 2. boundaries defined at 20 kb resolution show the similar features as at 5 kb resolution:

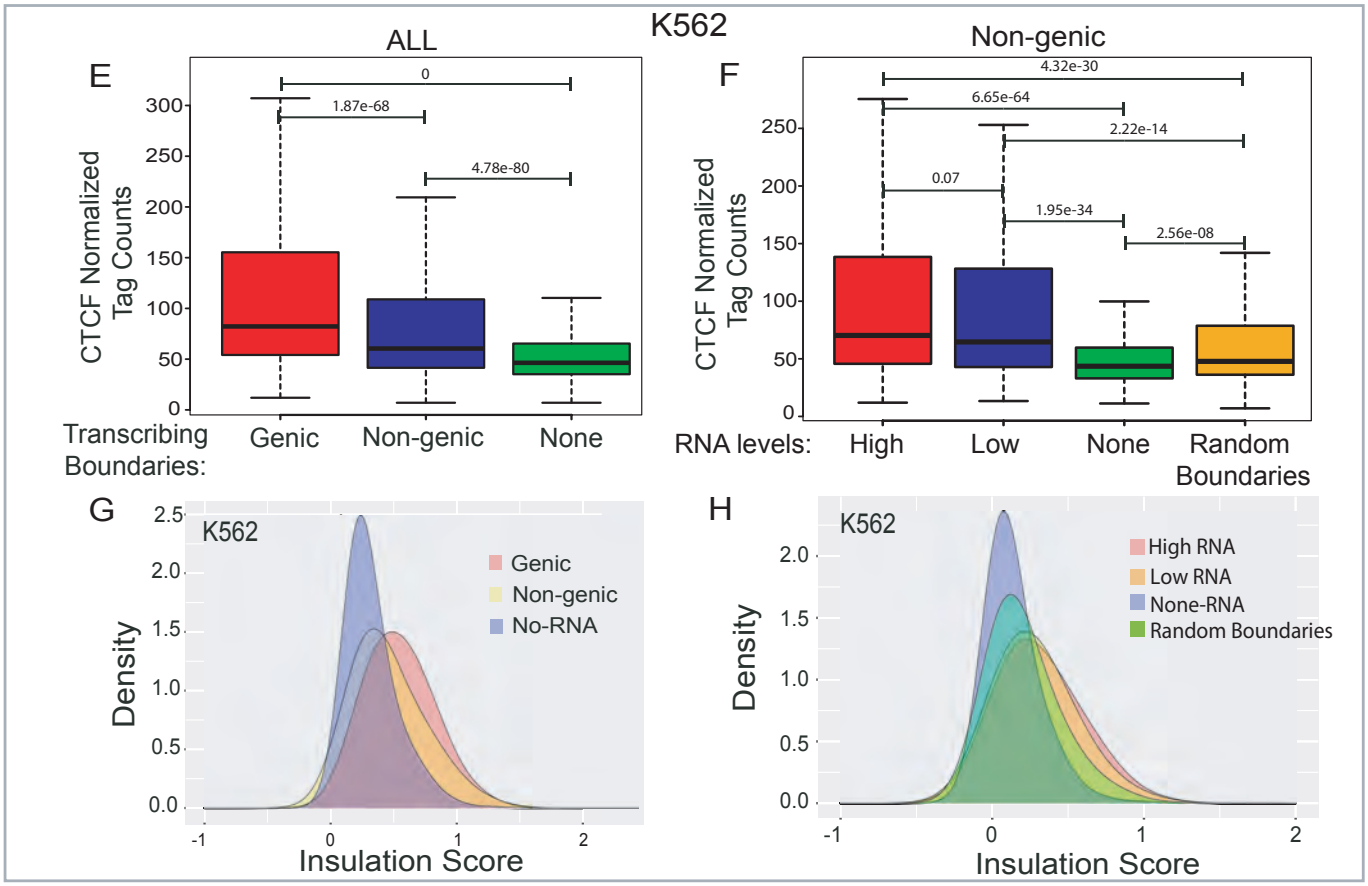
A. Pie chart displaying the percentage of boundaries that show genic and non-genic transcribed boundaries and not transcribed boundaries. **B.** Levels of RNA and CTCF occupancy at TAD boundaries are positively correlated. **C.** CTCF enrichment on transcribed genic, non-genic and non-transcribed boundaries. **D.** CTCF enrichment on all transcribed boundaries with high and low RNA versus non-transcribed and random boundaries. **E.** CTCF enrichment on non-genic transcribed boundaries with high and low non-coding RNA versus non-transcribed and random boundaries. **F.** Density plot showing insulation scores for transcribed genic, non-genic and non-transcribed boundaries. **G.** Density plot showing insulation scores for all transcribed boundaries exhibiting varying levels of RNA vs. non-transcribed and random boundaries. **H.** Density plot showing insulation scores of non-genic transcribed boundaries exhibiting varying levels of RNA versus non-transcribed and random boundaries. p-values in boxplots were calculated by Wilcoxon rank-sum test. The boxplots depict the minimum (Q1-1.5*IQR), first quartile, median, third quartile and maximum (Q3+1.5*IQR) without outliers.



Islam and Saravanan et al Sup Fig. 3

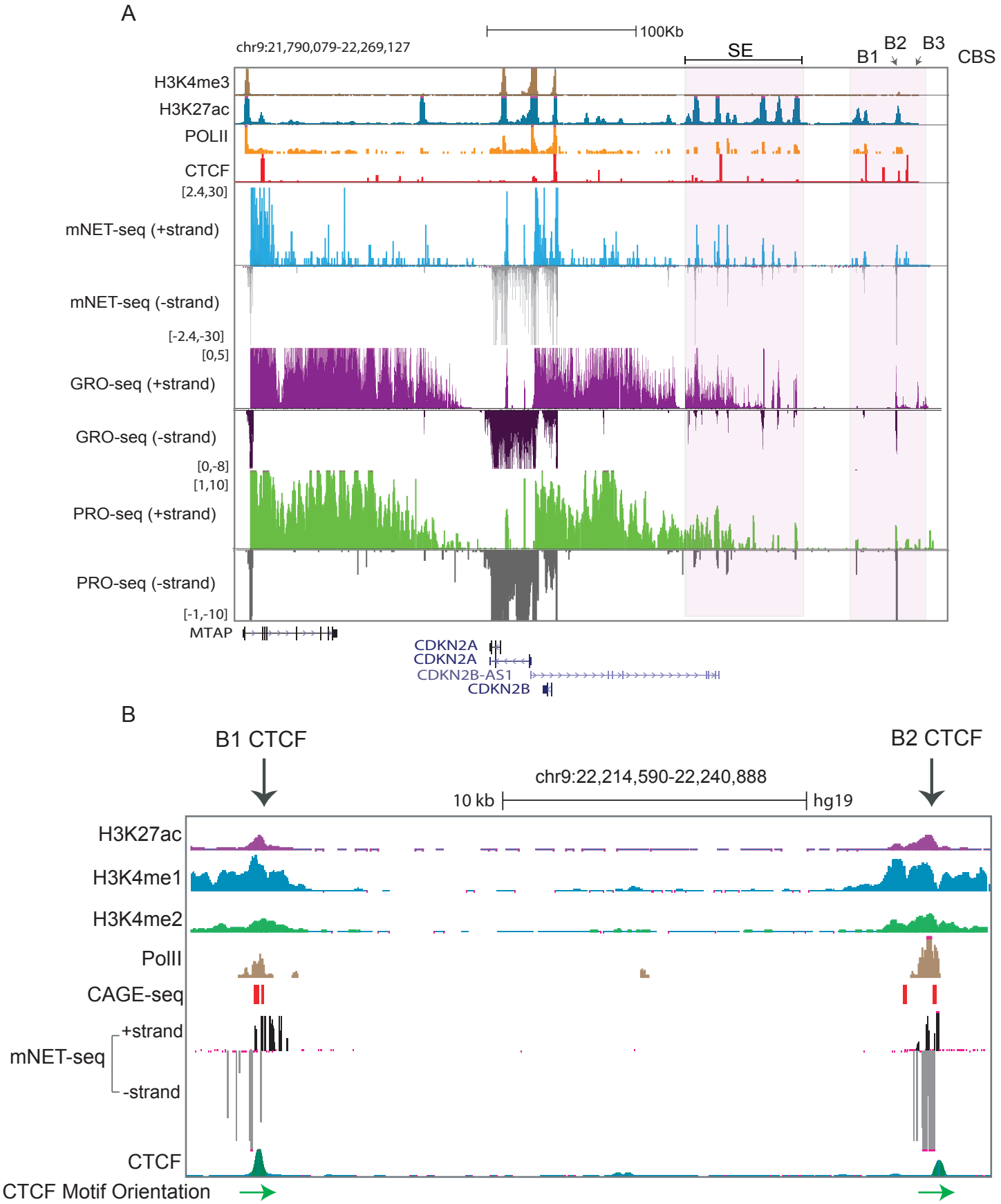
Supp Figure 3. Transcribed Loop Anchors exhibit more CTCF enrichment: A. Pie chart showing the percentage of loop anchors that overlap with regions of genic transcription and non-genic de novo transcription and regions without transcription. **B.** CTCF enrichment (CTCF CUT&RUN) at transcribed genic, non-genic and non-transcribed loop anchors. **C.** CTCF enrichment on all transcribed loop anchors with high and low RNA versus non-transcribed and random loop anchors. **D.** CTCF enrichment (CTCF ChIP-seq) at transcribed genic, non-genic and non-transcribed loop anchors. **E.** CTCF enrichment at all transcribed loop anchors with high and low RNA versus non-transcribed and random loop anchors. p-values in boxplots were calculated by Wilcoxon rank-sum test. The boxplots depict the minimum (Q1-1.5*IQR), first quartile, median, third quartile and maximum (Q3+1.5*IQR) without outliers.





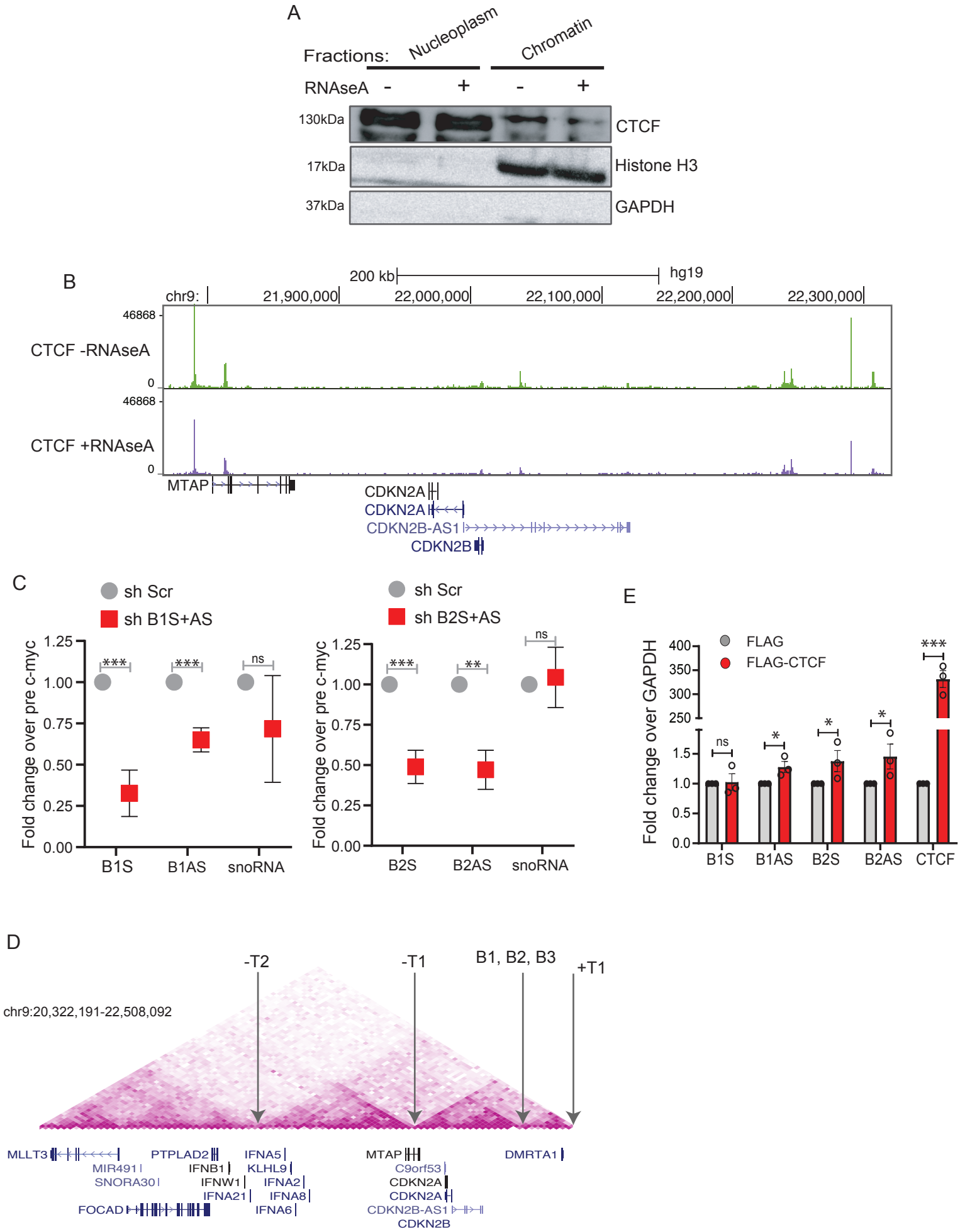
Islam and Saravanan et al Sup Fig. 4

Supp Figure 4. Transcribed boundaries in IMR90 and K562 exhibit better insulation: A. CTCF enrichment at transcribed genic, non-genic and non-transcribed boundaries in IMR90 cells. **B.** CTCF enrichment at non-genic transcribed boundaries with high and low non-coding RNA versus non-transcribed and random boundaries in IMR90 cells. **C.** Density plots showing insulation scores for transcribed genic, non-genic and non-transcribed boundaries in IMR90 cells**. D.** Density plots showing insulation scores for non-genic transcribed boundaries exhibiting varying levels of non-coding RNAs versus non-transcribed and random boundaries in IMR90. **E-F**. Same as A-B in K562 cells. **G-H.** Same as C-D in K562 cells. p-values in boxplots were calculated by Wilcoxon rank-sum test. The boxplots depict the minimum (Q1-1.5*IQR), first quartile, median, third quartile and maximum (Q3+1.5*IQR) without outliers.



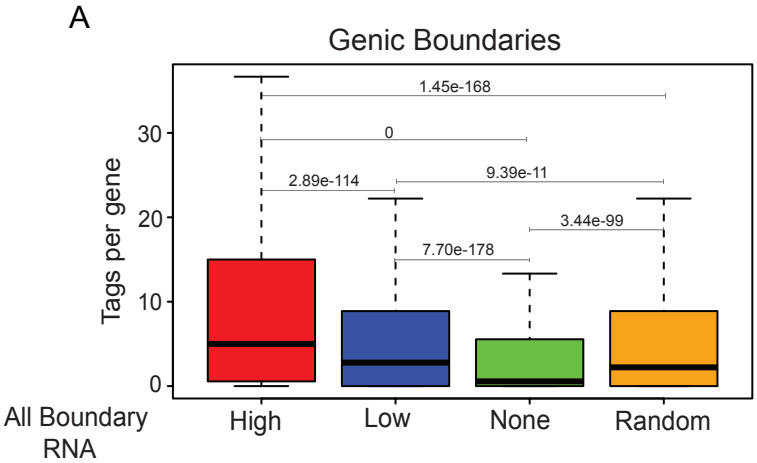
Islam and Saravanan et al. Supp Figure 5

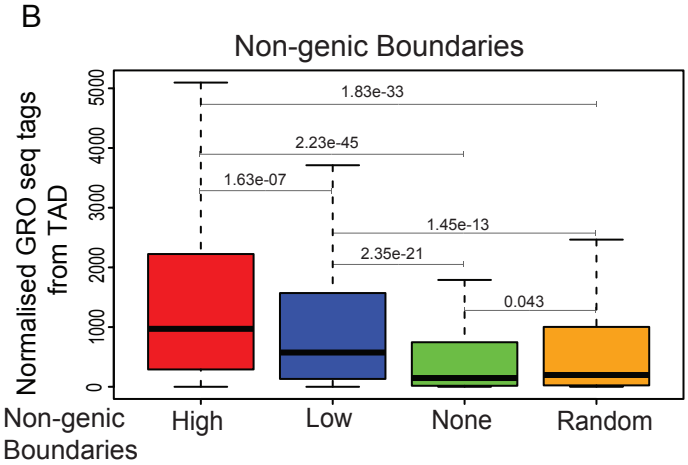
Supp Figure 5. Browser shots of the *INK4a/ARF* loop domain: A. A browser shot showing the *INK4a/ARF* locus in the 9p21 region. Below are H3K4me3, H3K27ac (enhancers), PolII, CTCF, mNET-seq tracks, GRO-seq tracks, PRO-seq tracks and gene annotations. The highlighted region shows the B1, B2 and B3 CTCF sites at the 3' boundary and super-enhancer. **B**. Browser shot showing the zoomed-in region of the 3' boundary (B1 and B2 CTCF sites). Below are H3K27ac, H3K4me1, H3K4me2, PolII, CAGE seq tracks, mNET-seq tracks and CTCF peaks. Arrowheads at the bottom show CTCF motif orientation at the B1 and B2 CTCF sites.



Islam and Saravanan et al. Supp Figure 6

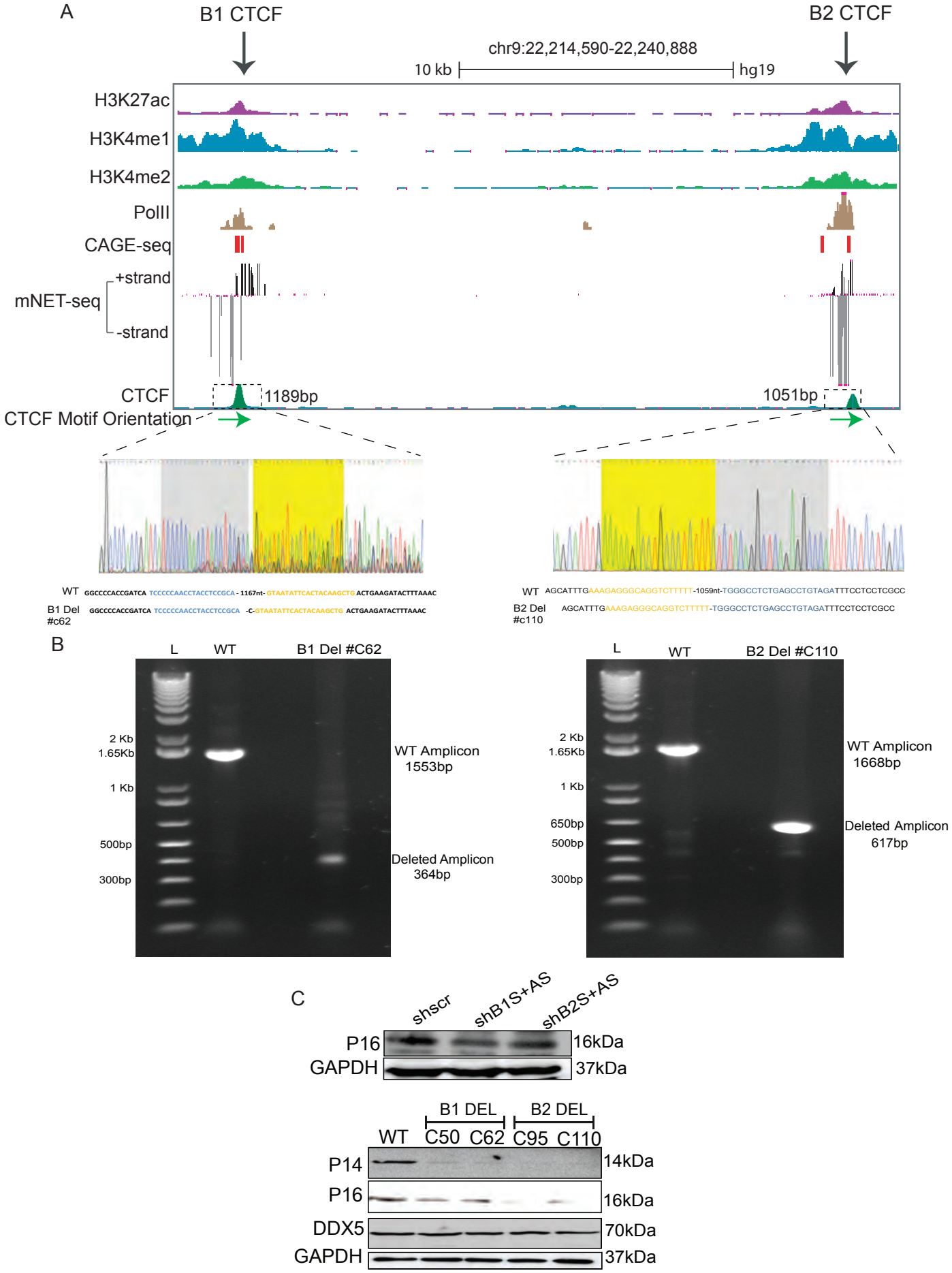
Supp Figure 6. RNA increases CTCF occupancy at TAD boundary: A. Immunoblotting with CTCF, Histone H3 and GAPDH on soluble nucleoplasm and chromatin-bound fractions of nuclei treated with or without RNaseA. **B.** Browser shot showing CTCF CUT&RUN peaks at the *INK4a/ARF* TAD with and without RNaseA treatment. **C.** qRT-PCRs showing the levels of sense and anti-sense non-coding RNA at B1and B2 CTCF sites and of snoRNA upon B1 and B2 RNA knockdown in the chromatin associated RNA fraction. **D.** TADs around the *INK4a/ARF* TAD. Positions of arrows show the CTCF sites at upstream and downstream TAD boundaries where CTCF occupancy was interrogated. **E.** qRT-PCRs showing the fold-change in sense and anti-sense RNAs from B1 and B2 CTCF sites upon Flag-CTCF overexpression. The last two bars show the increase in CTCF levels. Error bars denote SEM from three biological replicates. p-values were calculated by Student's two-tailed unpaired t test in **C** and **E**. ****$p < 0.0001$, ***$p < 0.001$, **$p < 0.01$, *$p < 0.05$, $^{ns}p > 0.05$.





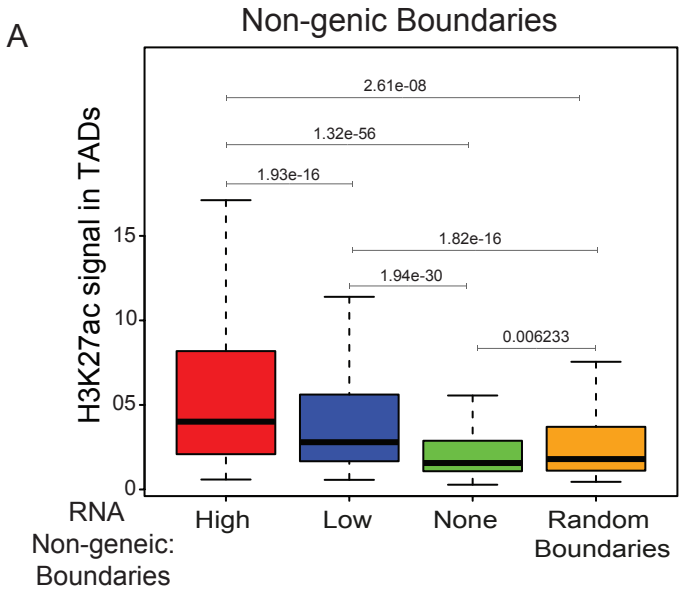
Islam and Saravanan et al. Supp Figure 7

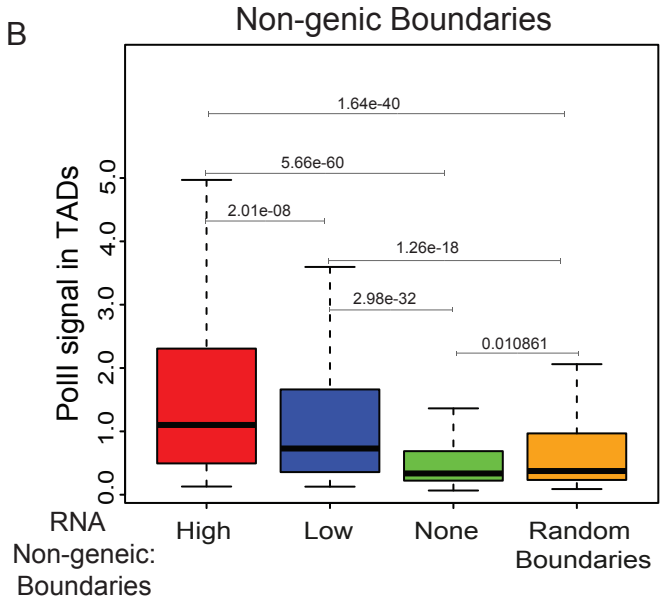
Supp Figure 7. TAD transcription is correlated with transcribed boundaries: A. Box plots showing the normalized GRO-seq reads per gene within TADs of high, low, non-transcribed and random boundaries. **B.** Box plots showing the normalized GRO-seq tags within the TADs of transcribed non-genic boundaries with high and low boundary RNAs, non-transcribed and random boundaries.



Islam and Saravanan et al Supp Fig. 8

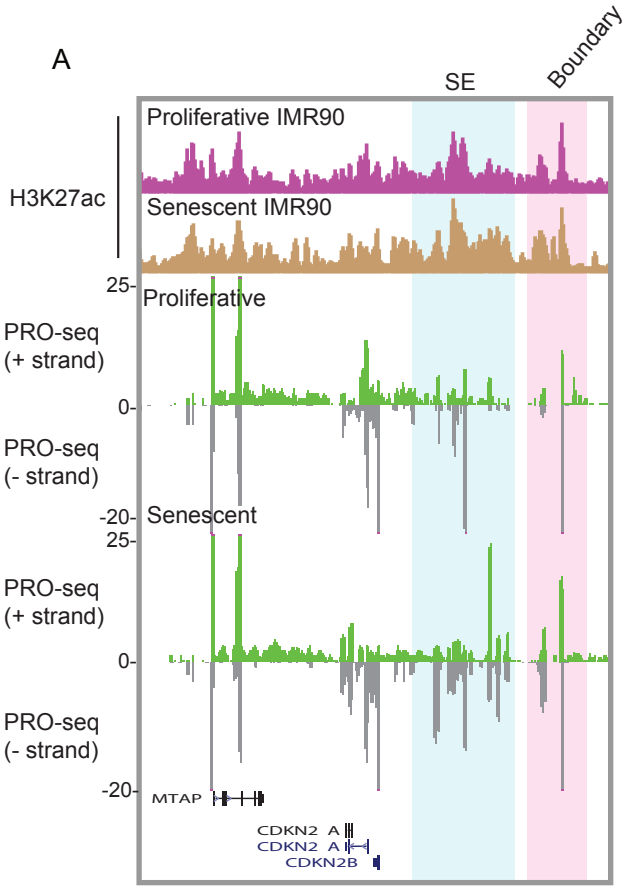
Supp Figure 8. Deletion of B1 and B2 CTCF sites in the *INK4a/ARF* TAD boundary: A. UCSC genome browser track depicting the deleted B1 and B2 CTCF sites (boxed region at the bottom). Below are CAGE, NET-seq signal, CTCF peaks and CTCF orientation (Green arrows). The tracks zoom in to the chromatograms from Sanger sequencing on deleted amplicons to verify the deletion. **B.** Agarose gel images showing the amplicons from a Surveyor assay of B1 and B2 deletions. **C.** Immunoblot showing the protein levels of p16 after Scr shRNA knockdown against sense and anti-sense RNA at B1 and B2 CTCF sites (Upper panel). Immunoblot showing the protein levels of p14, p16, DDX5 and GAPDH in WT and in cells with B1 and B2 CTCF site deletions (Lower panel).

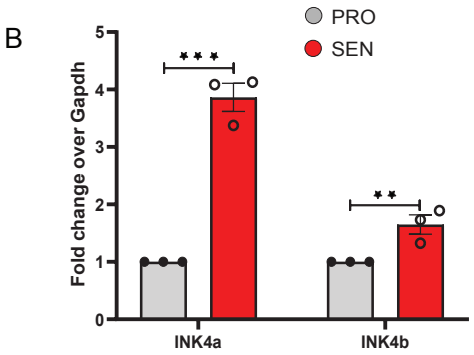


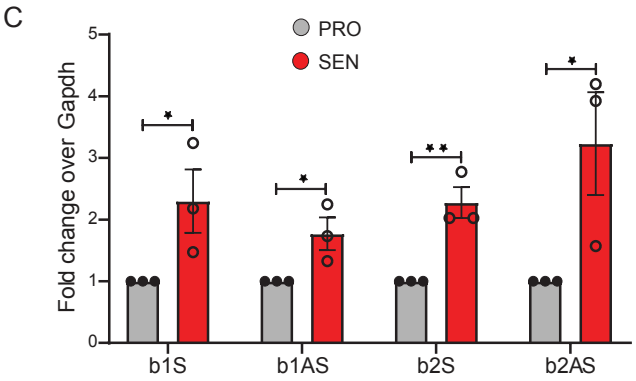


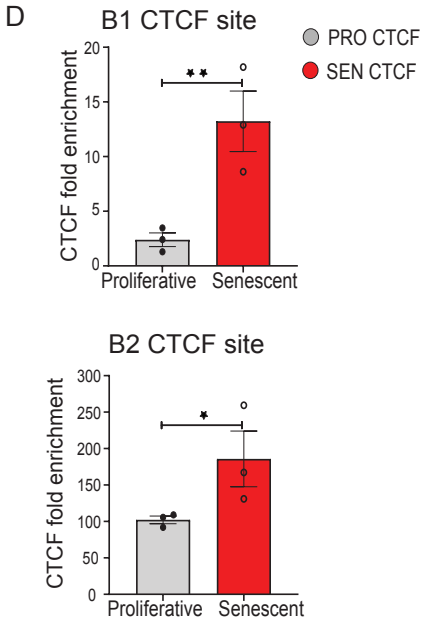
Islam and Saravanan Supp Fig. 9

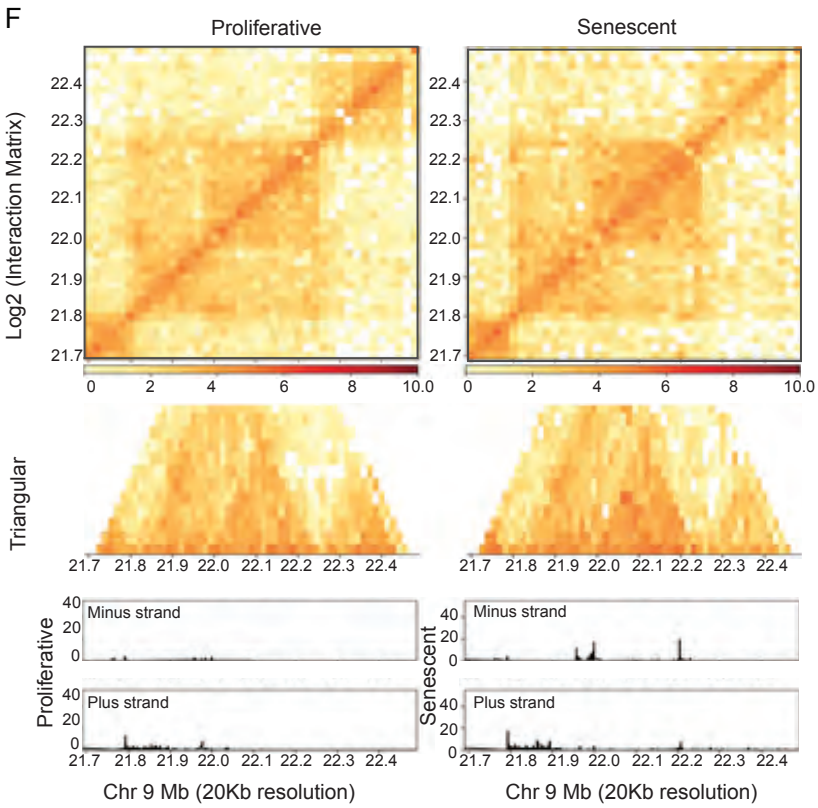
Supp Figure 9. TADs with transcribed boundaries harbor more active enhancers: A. Boxplots displaying enrichment of H3K27ac in TADs with high and low transcribed non-genic HeLa boundaries versus non-transcribed and random boundaries. **B.** Levels of PolII in TADs with different non-genic transcribed boundaries vs. non-transcribed and random boundaries.

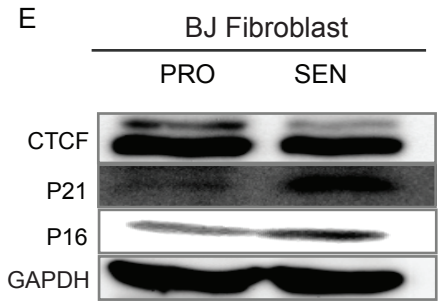












Islam and Saravanan et al. Supp Figure 10

Supp Figure 10. Transcriptional state of *INK4a/ARF* TAD in senescence : A. Browser images showing H3K27ac, PRO-seq reads on *INK4a/ARF* TAD from young and replicative senescent IMR90 fibroblasts. Highlighted regions depict the super-enhancer and the boundary. **B**. qRT-PCR plots show the fold change in expression of *INK4a* and *INK4b* in young and DNA damage-induced senescent BJ fibroblasts. **C.** qRT-PCR plots show the fold change in expression of boundary RNA from B1 and B2 CTCF sites in young and DNA damage-induced senescent BJ fibroblasts. **D.** ChIP qRT-PCRs showing the changes in fold enrichment of CTCF at B1 and B2 CTCF sites in young and DNA damage-induced senescent BJ fibroblasts. Error bars denote SEM from three biological replicates. *p*-values were calculated by Student's two-tailed unpaired *t*-test in **B**, **C**, and **D .** ***$p < 0.001$, **$p < 0.01$, *$p < 0.05$, $^{ns}p > 0.05$. **E.** Immunoblot showing the changes in protein levels of CTCF, p21, p16 and GAPDH in young and DNA damage-induced senescent BJ fibroblasts. **F.** *INK4a/ARF* TAD structure in proliferating and senescent cells. Lower boxes represent the transcriptional levels.

Table 1: Oligonucleotide sequences used in the study

Primers used for ChIP-qPCRs

CTCF sites	Forward Primer (5'-3')	Reverse Primer (5'-3')
-T2	TAAATCTCCTGCCAGGAGG	GGCAGGTTCCATTTTACTTTGAG
-T1	GGCAGTACTGAGTTGACAC	CCCAGGAAACTGGACCTCTAA
B1	AGCTCTCTGCACCACATTGCCACTGCC	CCAACCTACCTCCGCAAGGACACC
B2	TTGAGGACATCTAGCGGCG	GCTGCAGATTCTGGAATGAGC
B3	GTCAGGACTAGGCCCATCTGAC	TAGGTGTCCAAGTGAAGATTAAAGAG
+T1	CAACTCCTTCCGAGTGGAAA	ACACTGTGACCGCTCCATTC
+T2	GCATAGATGTTTGGAGGGCAAC	TTACAGAATCTGGCTACTGCCC
INK4a	TCGCCAGGAGGAGGTCTGTGATTAC	CAGGTGGGTAGAGGGTCTGCAGC
ARF	ACCCAGGATATTCGGGACTCACTGAC	CGTCTCTAGCCCAGGCTAGGAGG
INK4b	CCGTCGTCCTTCTGCGGCTTG	AGTGAGGACTCCGCGACGCGT

Primers used for qRT-PCRs

mRNA/bRNA	Forward Primer (5'-3')	Reverse Primer (5'-3')
MTAP	TTGACAGGACCACTATGAGACCTC	AGGTCCCTCGATTGTGACCATT
INK4a	ATGGAGCCTTCGGCTGACT	CACCAGCGTGTCCAGGAAG
ARF	AACATGGTGCGCAGGTTCT	CACCAGCGTGTCCAGGAAG
INK4b	CGTTAAGTTTACGGCCAACG	CCATCATCATGACCTGGATCG
DDX5	CCTACCTTGTCCTTGATGAAGC	CACTCAGTTCAAGTGCACCA
CTCF	TACAAACACACCCACGAGAAG	GGGTAAACCGAGCATGACAA
b1S	CCACAGCAGAATAGAGCACC	AAGTCCAGCCAGTCTCAGCA
b1AS	TGAGGCTTGTAGGAACTCAAGTTC	AGTGGCAATGTGGTGCAGAG
b2S	ACAGAGGTCGGATGTTCGTG	TAAACCGCCGCTAGATGTCC
b2AS	CTCCTGAGCTCTAAGTGAGATAGAGG	GGTTATTCACTCACAGTGCTCTTCAG
sno RNA	CGTGTAGAGCACCGAAAACC	CACTCAGACCGCGTTCTCTC
pre c-MYC	CTTCTCTCCGTCCTCGGATTC	CCTTCCTAATAAGAGTGGCCCG
GAPDH	CGCTCTCTGCTCCTCCTGTT	CCATGGTGTCTGAGCGATGT

shRNAs used for knockdown

shRNA	Primer sequence (5'-3')
shRNA1 Boundary 1 S F	CCGGGAGTTGTGAGACCCTCATTTCCTCGAGGAAATGAGGGTCTCACAACTCTTTTTG
shRNA1 Boundary 1 S R	AATTCAAAAAGAGTTGTGAGACCCTCATTTCCTCGAGGAAATGAGGGTCTCACAACTC
shRNA2 Boundary 1 F	CCGGGGACCCTGAGCTGTGAATAACCTCGAGGTTATTCACAGCTCAGGGTCCTTTTTG
shRNA2 Boundary 1 R	AATTCAAAAAGGACCCTGAGCTGTGAATAACCTCGAGGTTATTCACAGCTCAGGGTCC
shRNA3 Boundary 1 S F	CCGGGAGTCCTAGCCAGGTAATATTCTCGAGAATATTACCTGGCTAGGACTCTTTTTG
shRNA3 Boundary 1 S R	AATTCAAAAAGAGTCCTAGCCAGGTAATATTCTCGAGAATATTACCTGGCTAGGACTC
shRNA1 Boundary 1 AS	CCGGCTGGCTCCGGCTGGTTTAAATCTCGAGATTTAAACCAGCCGGAGCCAGTTTTTG

F	
shRNA1 Boundary 1 AS R	AATTCAAAAACTGGCTCCGGCTGGTTTAAATCTCGAGATTTAAACCAGCCGGAGCCAG
shRNA2 Boundary 1 AS F	CCGGCGGCTCTGCTTTCAGCTATAACTCGAGTTATAGCTGAAAGCAGAGCCGTTTTTG
shRNA2 Boundary 1 AS R	AATTCAAAAACGGCTCTGCTTTCAGCTATAACTCGAGTTATAGCTGAAAGCAGAGCCG
shRNA3 Boundary 1 AS F	CCGGAGGTAAGATGTAAGCATTAATCTCGAGATTAATGCTTACATCTTACCTTTTTTG
shRNA3 Boundary 1 AS R	AATTCAAAAAAGGTAAGATGTAAGCATTAATCTCGAGATTAATGCTTACATCTTACCT
shRNA1 Boundary 2 S F	CCGGCGCAATCCAACTATTCAATTTCTCGAGAAATTGAATAGTTGGATTGCGTTTTTG
shRNA1 Boundary 2 S R	AATTCAAAAACGCAATCCAACTATTCAATTTCTCGAGAAATTGAATAGTTGGATTGCG
shRNA2 Boundary 2 S F	CCGGGCTTTCCTGAAGGAGAGTAATCTCGAGATTACTCTCCTTCAGGAAAGCTTTTTG
shRNA2 Boundary 2 S R	AATTCAAAAAGCTTTCCTGAAGGAGAGTAATCTCGAGATTACTCTCCTTCAGGAAAGC
shRNA3 Boundary 2 S F	CCGGTAGCAATGGCAGAACCTATTTCTCGAGAAATAGGTTCTGCCATTGCTATTTTTG
shRNA3 Boundary 2 S R	AATTCAAAAATAGCAATGGCAGAACCTATTTCTCGAGAAATAGGTTCTGCCATTGCTA
shRNA1 Boundary 2 AS F	CCGGAGGAACTACATTGACTTATTTCTCGAGAAATAAGTCAATGTAGTTCCTTTTTTG
shRNA1 Boundary 2 AS R	AATTCAAAAAAGGAACTACATTGACTTATTTCTCGAGAAATAAGTCAATGTAGTTCCT
shRNA2 Boundary 2 AS F	CCGGCGACCTCTGTAAATGAAATAGCTCGAGCTATTTCATTTACAGAGGTCGTTTTTG
shRNA2 Boundary 2 AS R	AATTCAAAAACGACCTCTGTAAATGAAATAGCTCGAGCTATTTCATTTACAGAGGTCG
shRNA3 Boundary 2 AS F	CCGGTGGATTGCGATTTCCTGAAATCTCGAGATTTCAGGAAATCGCAATCCATTTTTG
shRNA3 Boundary 2 AS R	AATTCAAAAATGGATTGCGATTTCCTGAAATCTCGAGATTTCAGGAAATCGCAATCCA

Primers used for RNA IP

Target	**Primer sequence (5'-3')**
B1 S in pCDNA 3 Box B F	AGACTGGCGGCCGCTAGGCCTCAAGGACTGCAA
B1 S in pCDNA 3 Box B R	AGACTGTCTAGACTCCTGGGAGCCAGAGAA
B1 AS in pCDNA 3 Box B F	AGACTGGAATTCAGTAGGTCGCATGCAACCACA
B1 AS in pCDNA 3 Box B R	AGACTGGCGGCCGCTCCTACATCAGCTGTGGCACA
B2 S in pCDNA 3 Box B F	AGACTGGAATTCGCCTCTGAGCCTGTAGAGACCAA
B2 S in pCDNA 3 Box B R	AGACTGGCGGCCGCCTCTGGGTGAGGCAGTGAACAC
B2 AS in pCDNA 3 Box B F	AGACTGGAATTCATCCTCCAGGACCACAGCAGT
B2 AS in pCDNA 3 Box B R	AGACTGGCGGCCGCGCATACCCTGGAGTGAGAGGTTG
E8 in pCDNA 3 Box B F	AGACTGGAATTCAATGCCATATGGTGGCAATG
E8 in pCDNA 3 Box B R	AGACTGGCGGCCGCATCTGTTCCTTGCCTCTTCCAC

gRNAs used for deletions and CRISPRi

gRNA	Sequence
B1 DEL gRNA 1	TTAAATTGGTGTCCTTGCGG
B1 DEL gRNA 2	GTAGTGAATATTACCTGGCT
B2 DEL gRNA 1	CAGACTTCCGTAGTCCCCAT
B2 DEL gRNA 2	TTGGGTAGCATCTTCATCGG
MTAP gRNA 1	AGCAGTGCGGAGACCAGGGG
MTAP gRNA 2	CTCACTTGCCGCGCGAGGAG
INK4a gRNA 1	AGCCGTTTTACACGCAGGAG
INK4a gRNA 2	GACAGGACAGTATTTGAAGC
ARF gRNA 1	ATGCGCCCCGGACTTTTCGA
ARF gRNA 2	CTGCCCCCTTAACTGCAGAC
INK4b gRNA 1	AAAGAATTCCGTTTTCAGCT
INK4b gRNA 2	CTAGGAAGGAGAGAGTGCGC

Surveyor Oligos

Target	Forward Primer (5'-3')	Reverse Primer (5'-3')
B1 DEL	TGGAGTGGTAGTAGTGTGGGAAG	GTACGAGGGTTCTCTTTTGCTTG
B2 DEL	CACAACCCCCATCTCTCTGTAAC	TGAAGTGAGGGGAGTAGACAAGG

4C oligos for enhancer and Promoter viewpoints

View Point	Primer sequence (5'-3')
Enhancer F	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCTCTTCCGATCTGGTATTCTGACTAGAAGCTT
Enhancer R	CAAGCAGAAGACGGCATACGATAGTGGTGCTACAGTACCAC
Promoter F	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCTCTTCCGATCTGGAGGAGCTAGGGCAAGCTT
Promoter R	CAAGCAGAAGACGGCATACGACAGATGGATAAGTAGTAGCAAG

Table 3: Datasets used in the study

Experiment	Annotation	Accession Number	References
HeLa GRO-seq	SRR1596500	GSM1518913	Andersson et al., 2014. Nat Commun.
HeLa GRO-seq	SRR1596501	GSM1518914	Andersson et al., 2014. Nat Commun.
HeLa HiC	SRR1050968	-	Burton et al., 2013. *Nat Biotechnol*.
HeLa HiC	SRR1050969	-	Burton et al., 2013. *Nat Biotechnol*.
HeLa HiC	SRR3203899-SRR3203906	-	Rao et al., 2014. Cell.
HeLa PolII	wgEncodeSydhTfbsHelas3Pol2StdSig	GSM935395	ENCODE
HeLa CTCF	wgEncodeUwTfbsHelas3CtcfStdRawRep1	GSM749729	ENCODE
HeLa CTCF	wgEncodeUwTfbsHelas3CtcfStdRawRep2	GSM749739	ENCODE
HeLa H3K27ac	wgEncodeBroadHistoneHelas3H3k27acStdSig	GSM733684	ENCODE
IMR90 HiC	SRR5494761	GSM2595584	Zirkel etal., 2018. Mol Cell
IMR90 PRO seq	SRR6205688	GSM2828779	Sen et al., 2019. Mol Cell
IMR90 CTCF	wgEncodeSydhTfbsImr90CtcfbIggrabSig	GSM935404	ENCODE
K562 HiC	SRR1658693-SRR1658702	GSM1551618-GSM1551623	Rao et al., 2014. Cell.
K562 CTCF	wgEncodeUwTfbsGm12878CtcfStdAlnRep1	GSM749690	ENCODE
K562 GROseq	SRR1552484	GSM1480325	Core et al., 2014. Nat Genet.
HeLa_GRO-Seq	SRR5109939	GSM2428725	Bouvy-Liivrand et al., 2017. Nucleic Acids Res
PRO-Seq-DMSO	SRR5799635	GSM2692352	Nilson et al., 2017.Nucleic Acids Res
Total mNET-Seq	SRR4436479	GSM2357382	Schlackow et al., 2017. Mol Cell