Active enhancers strengthen TAD insulation by bRNA mediated CTCF enrichment at the TAD boundaries

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

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. The University of Trans-Disciplinary Health Sciences and Technology, IVRI road, Bangalore, Karnataka, 560064. India

4. Department of Biology, Emory University, Atlanta, Georgia, 30322. USA

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:

The genome is partitioned into Topologically Associating Domains (TADs). About half of the boundaries of these TADs exhibit transcriptional activity and are correlated with better TAD insulation. However, the role of these transcripts *per se* in TAD insulation, enhancer:promoter interactions and transcription remain unknown. Here we investigate the functional roles of these bRNAs (boundary-RNA) in boundary insulation and consequent effects on enhancer-promoter interactions and TAD transcription genome-wide and on disease relevant INK4a/ARF TAD. Using series of CTCF sites deletion and bRNA knockdown approaches at this TAD boundary, we show a direct association of CTCF with bidirectional bRNAs where the loss of bRNA triggers the concomitant loss of: CTCF clustering at TAD boundary, its insulation, enhancer:promoter interactions and gene transcription within the targeted TAD. In search of what regulates bRNA expression itself, we used another series of enhancer deletions and CRISPRi on promoters within INK4a/ARF TAD and observed that indeed, enhancers interact with boundaries and positively regulate the bRNA transcription at TAD boundaries. In return, the bRNAs recruit/stabilize the CTCF even on weaker motifs within these boundaries and supports CTCF binding in clusters, therefore enhancing TAD insulation which favors the intra-TAD enhancer:promoter interactions and robust gene transcription. Functionally, eRNAs within the boundaries are repurposed as more stable bRNAs and their knockdown exactly mimics the boundary loss. Furthermore, transcribing boundaries exhibit high TAD transcription in TCGA tumor datasets. Together, these results show that active enhancers directly mediate better insulation of TADs by activating the transcription at TAD boundaries. These transcripts trigger CTCF clustering at the boundary resulting in better insulation which favours robust intra-TAD enhancer:promoter interactions to activate the gene transcription.

78 **Introduction:**

79

80 Genomes are organized into functional and physical units known as TADs (Topologically
81 associating domains) (Dixon et al., 2012). Highly self-interacting domains are identified as contact
82 domains and the domains that exhibit corner dots (at the apex of loop) are known as loop domains.
83 Most TADs exist as contact domains and is used as an umbrella term for TADs while some TADs
84 are identified as both loop domains and contact domains (Beagan and Phillips-Cremins, 2020).
85 Loop and contact domains are both proposed to be formed by loop extrusion (Dekker and Mirny,
86 2016; Dixon et al., 2012; Fudenberg et al., 2016; Nuebler et al., 2018; Rao et al., 2014; Szabo et
87 al., 2020), and are dependent on CTCF and cohesin (Nora et al., 2017; Rao et al., 2017; Schwarzer
88 et al., 2017; Wutz et al., 2017). Length of the extruded loop is determined by the presence and
89 orientation of CTCF bound regions which contains interactions within the loop (Rao et al., 2014).
90 Thus, by not allowing the enhancers and promoters to interact outside of the loops, CTCF strength
91 and number of binding sites at boundaries provides greater insulation of the TAD (Krivega and
92 Dean, 2017; Nora et al., 2017; Rowley et al., 2017). However, CTCF consensus motif or their
93 orientation predict TADs boundaries less accurately suggesting, additional factors must influence
94 the CTCF binding and therefore TAD formation/insulation.

95

96 The perturbations in CTCF sites at the boundaries have shown none to severe effects on gene
97 transcription in the same and neighbouring TADs (Despang et al., 2019; Flavahan et al., 2016;
98 Hanssen et al., 2017; Kloetgen et al., 2020; Narendra et al., 2015; Sima et al., 2019; Williamson
99 et al., 2019; Yang et al., 2019). Within these observations, perturbation of all of the CTCF sites
100 within the boundary region or acute depletion of CTCF is essential to display effects on
101 transcription (Khoury et al., 2020; Williamson et al., 2019) suggesting, a redundancy among the
102 multiple CTCF sites at the TAD boundary. However, how multiple CTCF sites within the
103 boundary coordinate to achieve better insulation requires interrogations.

104

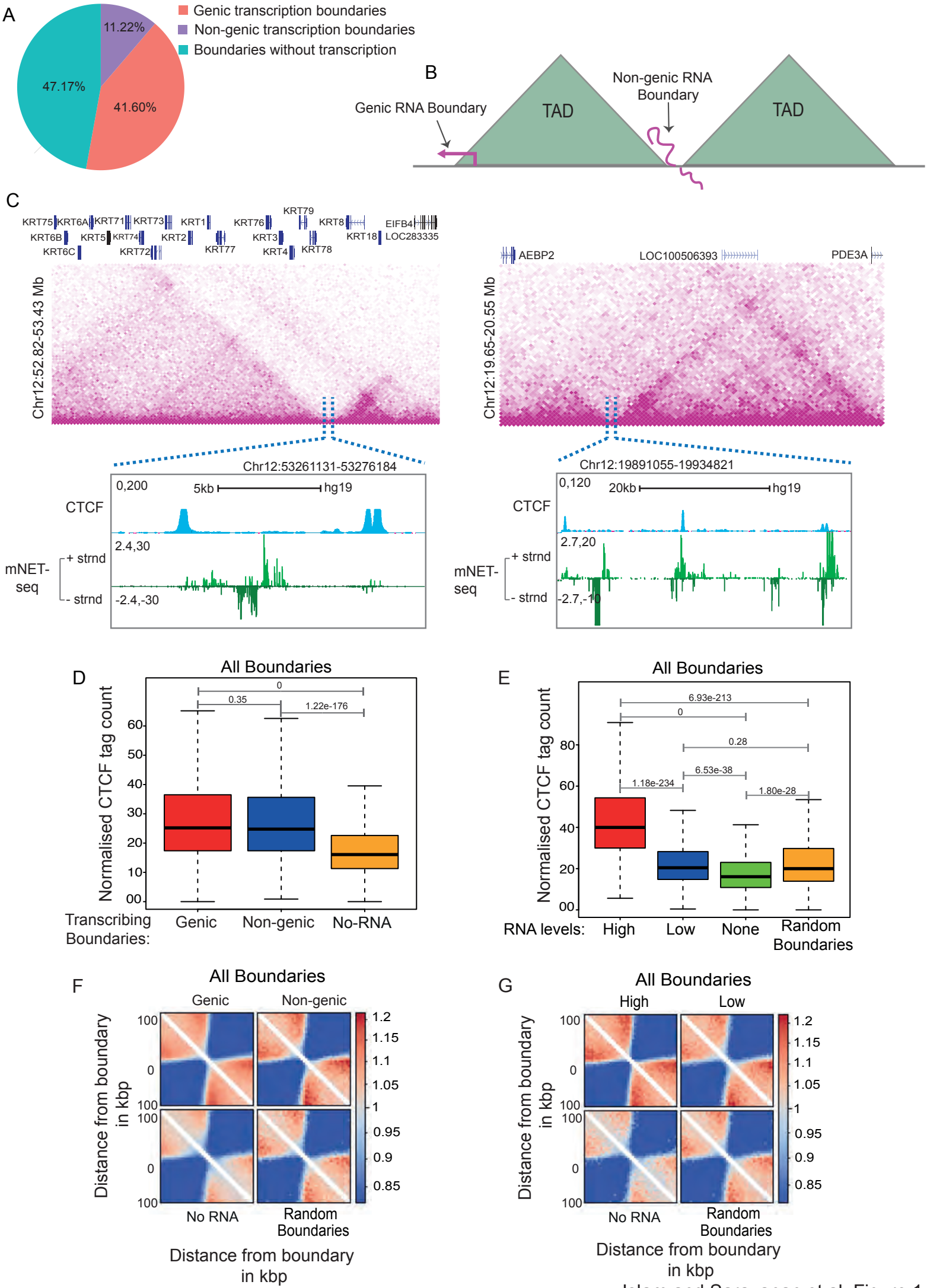
105 Further, transcription at the boundaries is correlated with better insulation and CTCF occupancy
106 (Bonev et al., 2017; Hsieh et al., 2020; Khoury et al., 2020) explaining, why transcription inhibition
107 or RNaseA treatments affect genome architecture (Barutcu et al., 2019, Thakur et al., 2019).
108 Furthermore, CTCF dimerization, important for loop extrusion, is RNA dependent and RNA

109 binding of CTCF is required for genome organization (Hansen et al., 2019; Saldaña-Meyer et al.,
110 2019), but how CTCF:RNA interactions maintain the TAD organization is not clear. Further,
111 which RNAs and their sites of transcription that are involved in such organizer functions is not
112 known.
113
114 Furthermore, if active TADs utilise these RNA mediated mechanisms to strengthen the TAD
115 boundaries remains completely unknown. For example, if these observed effects on chromatin
116 architecture are the result of transcription or transcripts from, intra TAD regions or from the
117 boundaries remains an interesting aspect of this conundrum. Further, if RNA
118 transcription/transcripts at the TAD boundaries influence enhancer:promoter interactions and gene
119 transcription remains unknown.
120
121 Interestingly, the TADs with super enhancers exhibit strong boundaries (Gong et al., 2018; D.
122 Zhang et al., 2020). Enhancers and boundaries are also known to interact with each other in some
123 instances (NandyMazumdar et al., 2020). However, if these enhancers play an active role in
124 boundary insulation or if there is a regulatory relationship between these two type of elements
125 remains to be tested.
126
127 Here we show that boundaries of about half of the TADs (with and without corner dots) exhibit
128 ongoing transcriptional activity either from a gene-body that is overlapping with boundary region
129 or from non-genic region that transcribes *de novo* non-coding RNAs within boundary. The
130 transcribing boundaries (non-genic and genic) show more CTCF enrichment and number of CTCF
131 peaks resulting in better TAD insulation (with and without corner dots). Further, we observed
132 genome-wide high corroboration between higher transcription at boundaries with the higher
133 transcriptional state of the genes within the TAD. Furthermore, the non-coding RNAs within non-
134 genic boundaries are transcribed from enhancers like regions however; these RNAs are more
135 stable, have relatively higher presence of 5' splicing sites and exhibit less abundance of
136 transcription termination signal (TTS) than eRNAs. Similarly, the genic transcripts from the
137 boundaries also exhibit more stability as compared to other mRNAs. Due to these distinct features
138 of RNA (coding and non-coding) transcribing from boundaries, we name them bRNAs (Boundary-
139 RNAs). One such non-genic boundary shows *de novo* bRNA transcribing from two CTCF sites at

140 the INK4a/ARF locus TAD. These bRNAs and CTCF physically interact with one another and
141 knockdown of these sense and anti-sense bRNAs triggers the loss of CTCF at the boundary and
142 also results in down regulation of genes within INK4a/ARF TAD which is similar to the deletion
143 of these two CTCF sites. This TAD also harbours a super enhancer and the deletions of multiple
144 enhancers within SE reveals that indeed active enhancers regulate the *de novo* bRNA transcription
145 by direct physical interactions between these enhancers and boundary elements. Transcriptional
146 activity of promoters, however, has no effect on *de novo* bRNA expression. Both, CTCF sites
147 deletions and bRNA knockdown results in the loss of enhancer:promoter interactions but
148 strengthened interactions of the enhancers with regions outside the TAD. These observations
149 suggest that active enhancers directly assist in TAD insulation by triggering the transcription at
150 the boundaries. The bRNA transcripts then recruit CTCF at the boundary and thus enhancers are
151 enforced to interact more with promoters in the same TAD resulting in robust intra-TAD
152 transcription.
153
154 **Results:**
155
156 **RNA transcribing boundaries exhibit better insulation**
157
158 Transcriptional inhibition by either DRB or α-amanitin alters the chromatin architecture (Hsieh et
159 al., 2020, Barutcu et al., 2019). This suggests that the transcriptional activity or the RNA *per se*
160 play a role in genome organization. Further, the active transcription at the TAD boundaries has
161 been speculated to play a role in TAD strength (Giorgetti et al., 2016; D. Zhang et al., 2020; Zhang
162 et al., 2019). Similarly, severe depletion of CTCF, an enriched factor at TAD boundaries also
163 affects the chromatin architecture (Khoury et al., 2020; Nakahashi et al., 2013; Nuebler et al.,
164 2018; Pugacheva et al., 2020). Thus, we asked if there is any association between the
165 transcriptional activity at the TAD boundaries and CTCF occupancy. Towards this, we first
166 interrogated the proportion of genome-wide TAD boundaries that transcribe RNA using GRO-seq
167 (Andersson et al., 2014). In order to identify boundaries, we employed HOMER that uses a high
168 degree of intra-domain interactions to their surrounding regions for TAD calling. As a result, it
169 identifies contact domain TADs, with corner dot (loop domains) and without corner-dot. Both,
170 TADs and loops are proposed to be formed by loop extrusion and are dependent on CTCF and

171 cohesin (Nora et al., 2017; Rao et al., 2017; Schwarzer et al., 2017; Wutz et al., 2017). We
172 identified 15397 TADs boundaries (5 kb resolution) using HiC from HeLa (Rao et al., 2014).
173 Within these, 8134 TADs boundaries (52.8%) showed presence of nascent transcripts in GRO-seq.
174 Several TAD boundaries have overlapping genes within them. Thus, we tested the extent of these
175 boundaries transcribing RNA as a result of genic transcription overlapping with boundary region.
176 We stratified the TAD boundaries into two categories; the boundaries with presence of transcripts
177 that arise from overlapping genes (gene body) with boundary (5 kb region), and the boundaries
178 where the transcription originates from within the boundaries (5 kb) but without the presence of
179 any gene (*de novo* non-coding transcription) (Fig. 1A-B). 41.56% (6406) out of 15397 had genic
180 transcripts. Whereas, 11.22% (1728) boundaries showed *de novo* non-coding transcription (Fig.
181 1A, C). These data suggest that a large majority of TAD boundaries show ongoing transcriptional
182 activity. Further, several boundaries exhibit *de novo* transcription originating as non-coding RNA.
183 We label both, genic and non-genic boundaries as RNA transcribing boundaries. Note, most RNA
184 transcribing boundaries (78.75%) are genic and remaining (21.25%) are non-genic boundaries.
185
186 TAD boundaries are enriched for CTCF binding and it also plays an important role in loop
187 extrusion. Thus, to test the correlation between strength of CTCF occupancy and presence of
188 transcription, we assessed the levels of CTCF on the genic, non-genic and non-transcribing
189 boundaries. To rule out the biased enrichment of CTCF due to chemical fixation used in standard
190 ChIP protocol, we used Cut & Run data. Total CTCF signal was calculated at these boundaries
191 using CTCF Cut & Run data (Skene and Henikoff, 2017) and we observed that genic and non-
192 genic RNA transcribing boundaries had significantly higher CTCF enrichment as compared to
193 non-transcribing boundaries (Fig. 1D). This correlation in (Fig. 1D) was also evident on all the
194 boundaries segregated based on RNA expression levels (Fig. 1E) (Top 1500 (10%) boundaries out
195 of the 15397, with respect to the GRO-seq signal is classified as high, 1500 non-zero low GRO-
196 seq signal as Low, random 1500 zero GRO-seq signal as None and 1500 random boundaries from
197 15397 classified as random boundaries). Further, within non-genic transcribing boundaries, we
198 observed that CTCF enrichment was correlated with the levels of RNA transcription (600
199 boundaries in all categories (10% of all non-genic boundaries)) (Fig. S1A-B). Interestingly, even
200 the non-genic boundaries with lowest transcriptional activity exhibited higher CTCF enrichments
201 as compared to non-transcribing boundaries (Fig. S1A). The same trend of CTCF correlation with

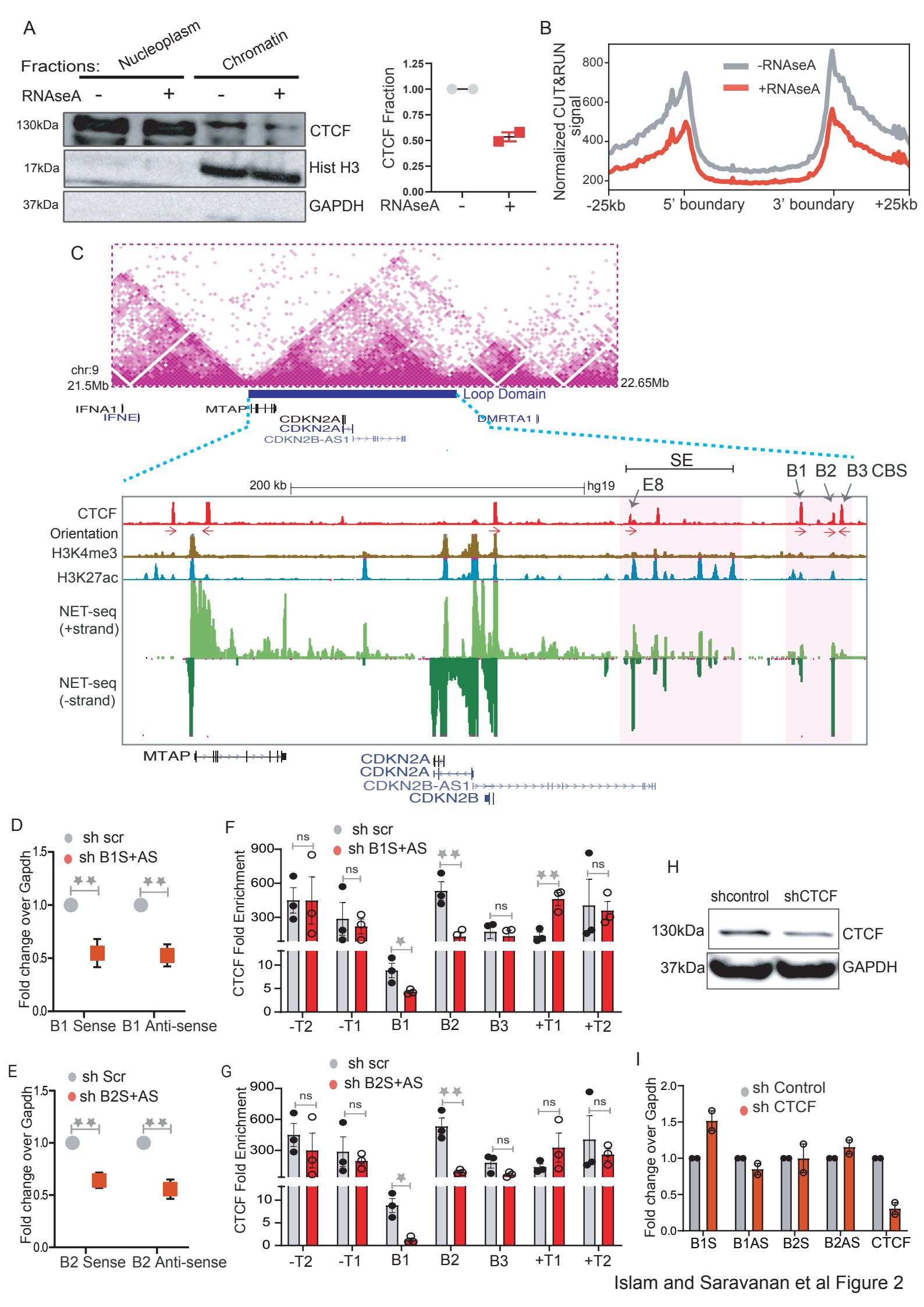
202 RNA presence on all, genic and non-genic boundaries was noted with CTCF ChIP-seq (with
203 formaldehyde fixed cells (Zhang et al., 2020)) (Fig. S1C-E). Taken together, the transcribing
204 boundaries exhibit higher CTCF binding as compared to non-transcribing and random boundaries
205 in the genome (Fig. 1D-E, S1A-E).
206
207 Boundaries insulate TADs by favoring intra-TAD interactions and their strength is measured by
208 insulation scores using HiC data. We asked if the various classes of TAD boundaries that are also
209 transcribing and show high CTCF binding indeed have differential insulation strength.
210 Interestingly, we noted a higher insulation of genic and non-genic transcribing boundaries as
211 compared to non-transcribing boundaries (Fig. 1F). Further, the insulation score was positively
212 corelated with the RNA levels in all transcribing boundaries (Fig. 1G) and non-genic transcribing
213 boundaries (Fig. S1F). Even the TADs with least transcribing non-genic boundaries were better
214 insulated as compared to non-transcribing boundaries (Fig. S1F) (Here, higher insulation score
215 refers to better insulation). To rule out resolution related biases, we extended the 5kb windows of
216 TAD boundary to 20 kb and these observations were found to be replicated even then (Fig. S2).
217 For example, total 69.2% (10658) out of 15397 TADs boundaries were found to be transcribing
218 (Fig. S2A). The CTCF enrichment on All, genic and non-genic boundaries even at 20 kb distance
219 was true as for 5 kb (Fig. S2B-E). Insulation score also followed the same trend (Fig. S2F-H).
220
221 Further, to test if the positive correlation between transcribing boundaries and CTCF enrichment
222 is also true on domains called by yet another method, we used HICCUPS (Algorithm of the Juicer
223 Software suite) that identifies loop domains. In total 3525 loop anchors were identified with 5kb
224 boundary region. Out of these, 43.09% (1519) were genic and 19.19% (675) were non-genic
225 boundaries (loop anchors) (Fig. S3A). We found similar positive correlation between transcribing
226 boundaries and extent of transcription on all transcribing boundaries with CTCF enrichment using
227 CTCF cut & run (Fig. S3B-C) and ChIP-seq data (Fig. S3D-E).
228
229 These observations of higher CTCF and insulation with extent of transcription at transcribing
230 boundaries were also confirmed in IMR90 (Rao et al., 2014) (Fig. S4A-D) and lymphoblastoid
231 cells K562 (Zirkel et al., 2018) (Fig. S4E-H) suggesting these correlations are relevant in other
232 cancer-type and normal cells. These data reveal the high CTCF occupancy and concurring higher



Islam and Saravanan et al. Figure 1

Figure 1. RNA transcribing boundaries exhibit better insulation: A. Pie diagram showing the percentage of boundaries that show genic transcription, non-genic de novo transcription and no transcription. **B.** The schematic depicting genic and non-genic transcribing boundaries. **C.** Browser shots displaying TAD structure from HiC data on two regions. The zoomed in box show *de novo* non-genic transcribing boundary region that is overlaid with CTCF and mNET-seq signal. **D.** Box plots showing the CTCF enrichment on genic, non-genic and Non-transcribing boundaries. **E.** Box plots showing the CTCF enrichment on All boundaries exhibiting varying levels of RNAs vs. non-transcribing and random boundaries. **F.** Pile up plot on transcribing genic, non-genic, non-transcribing boundaries and random boundaries. **G.** Pile up plot on high transcribing, low transcribing, non-transcribing and random boundaries. The p-values in boxplots were calculated by Mann Whitney test (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.

233 insulation score of transcribing boundaries. For the remaining work, we focused on TADs called
234 by HOMER with 5 kb boundary regions.
235
236 **RNA augments the recruitment of CTCF at the TAD boundary**
237
238 Since, genic and non-genic transcribing boundaries exhibit high levels of CTCF, we asked if there
239 is a causal relationship between the presence of RNAs (genic RNA or *de novo* non-coding RNA)
240 and CTCF at these sites. Towards this, we first tested the dependency of CTCF on RNA for its
241 chromatin occupancy by chromatin and nucleoplasm fractionations upon RNaseA treatments to
242 HeLa cells. We noticed a significant loss of CTCF in the chromatin fraction upon removal of RNA
243 (Fig. 2A). This data is in concordance with the reported dependence of CTCF on RNA for
244 chromatin binding (Hansen et al., 2019; Saldaña-Meyer et al., 2019, Thakur et al., 2019). To test
245 the relative loss of CTCF on specific sites in the genome upon RNAseA, we performed Cut & Run
246 for CTCF upon RNAseA treatment and interrogated the altered CTCF binding in a genome-wide
247 manner. As expected, integrated CTCF tag counts along the TADs showed highest enrichment of
248 CTCF at the TAD borders (Fig. 2B), and this enrichment at the boundaries was relatively more
249 susceptible to RNAseA treatment as compared to Intra-TAD regions (Fig. 2B). Thus, we asked
250 how RNA might regulate the binding of CTCF at the TAD borders.
251
252 We tested if such loss of CTCF at the boundaries is directly mediated via RNA *per se* or the loss
253 of CTCF is an indirect effect of general RNA degradation as seen in RNaseA Cut & Run data (Fig.
254 2B). Towards this, we performed a series of experiments on a disease relevant *INK4a/ARF* TAD
255 (approx. 500 Kb long) as a model (Fig. 2C). This TAD harbors actively transcribing cell cycle
256 regulatory genes; *CDKN2A*, *CDKN2B* and *MTAP*. The locus also harbors a super-enhancer
257 associated with several GWAS variants (Fig. 2C, high-lighted region) (Farooq et al., 2021;
258 Harismendy et al., 2011). Several enhancers within the super enhancer are active as seen by
259 H3K27ac marks and eRNA expression (Fig. 2C, mNET-seq panel) (Farooq et al., 2021). The 5'
260 boundary of the TAD is within the transcribed gene *MTAP*, whereas the 3' boundary region is non-
261 genic and contains three significant CTCF ChIP-seq peaks (B1, B2 and B3) (Fig. 2C, high-lighted
262 region). Out of these, B1 and B2 CTCF sites exhibited strong nascent transcription as observed by
263 publicly available mNET-seq, GRO-seq and PRO-seq data (Bouvy-Liivrand et al., 2017; Nilson

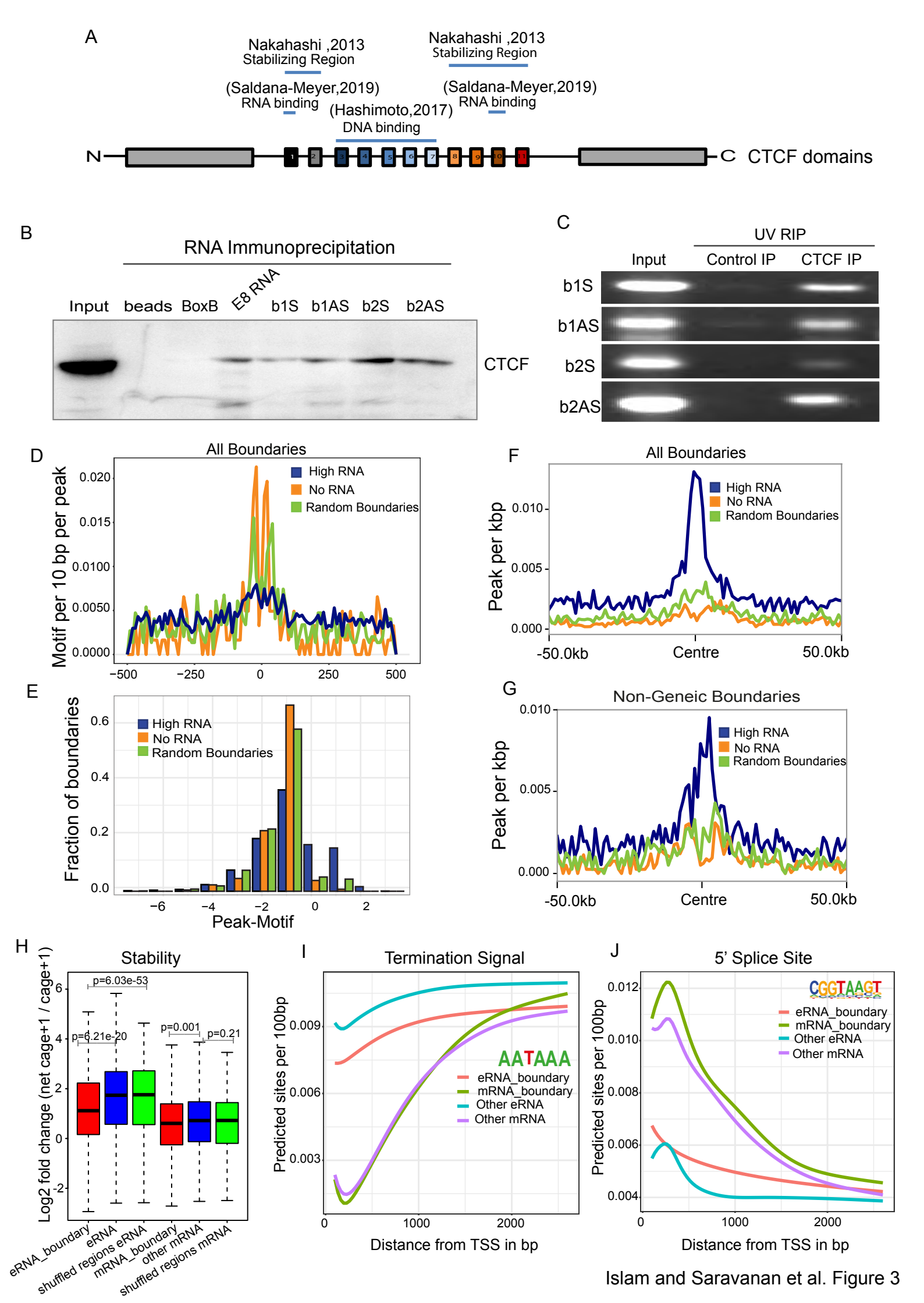


Islam and Saravanan et al Figure 2

Figure 2. RNA augments the recruitment of CTCF at the TAD boundary: A. Immunoblotting with CTCF, Histone H3 and GAPDH 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 and run signal in control and RNAse A treated conditions plotted at equally scaled TADs of 50kb length each and 25kb flanks**. C.** 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 3' boundary and super enhancer with position of enhancer E8 marked by an arrow. **D.** qRT-PCRs showing the levels of sense and anti-sense non-coding RNA at B1 CTCF site upon B1 RNA knockdown **E.** qRT-PCRs showing the levels of sense and anti-sense RNA at B2 CTCF site upon B2 RNA knockdown. **F.** CTCF ChIP enrichment before and after shRNA mediated knockdown of sense and anti-sense RNA at B1 CTCF sites. The bars show the CTCF on three CTCF sites (B1, B2 and B3) at 3' boundary, on -T1 (5' boundary) of INK4a/ARF TAD, -T2 on adjacent TAD boundary upstream, +T1 (on adjacent TAD boundary downstream) and +T2 the boundary of following TAD downstream. **G.** CTCF ChIP enrichment before and after shRNA mediated knockdown of sense and anti-sense RNA at B2 CTCF sites. The bars show the CTCF on three CTCF sites (B1, B2 and B3) at 3' boundary, on -T1 (5' boundary) of INK4a/ARF TAD, -T2 on adjacent TAD boundary upstream, +T1 (on adjacent TAD boundary downstream) and +T2 the boundary of following TAD downstream. Error bars denote SEM from three biological replicates. *p*-values were calculated by Student's two-tailed unpaired *t*-test in **D, E, F** and **G .** $^{**}p < 0.01$, $^{*}p < 0.05$, $^{ns}p > 0.05$ **H.** Immunoblot showing the drop in CTCF levels upon CTCF knockdown by the pool of shRNAs. **I.** qRT-PCRs of sense and antisense RNAs from B1 and B2 CTCF sites upon CTCF knockdown. The drop in CTCF levels are shown in the last two bars.

264 et al., 2017; Schlackow et al., 2017), whereas B3 was weakly transcribed (Fig. 2C, mNET-seq
265 panel and Fig. S5A mNET-seq, GRO-seq and PRO-seq on this TAD) and there was no annotated
266 gene or had reads from the nearest gene suggesting it is a *de novo* non-coding RNA boundary and
267 the CTCF motif within B1 and B2 site is in forward orientation (Fig. 2C, Fig. S5A-B). Since, the
268 removal of RNA by RNaseA treatment showed the loss of CTCF at the boundary regions on this
269 TAD as seen by CTCF Cut & Run upon RNaseA (Fig. S6A), we asked if these non-coding RNAs
270 recruit CTCF at this boundary (B1 and B2 CTCF sites). Towards that end, we depleted non-coding
271 RNAs at B1 and B2 CTCF sites using specific shRNAs against sense and anti-sense RNAs (Fig.
272 2D and E). We also confirmed the acute knockdown of RNAs at the B1 and B2 CTCF sites in
273 chromatin fraction by qRT-PCRs (Fig. S6B). The CTCF at B1 and B2 sites exhibited loss of CTCF
274 upon such knockdowns of sense and anti-sense RNAs. However, CTCF remained unaffected on
275 B3, the 5' boundary of the same TAD (-T1 site) and on the boundaries of its flanking TADs (-T2
276 and +T1, +T2) (Fig. 2F-G and Fig. S6C). Interestingly, loss of RNAs at B1 site perturbed the
277 CTCF occupancy at not only B1 but also at B2 site and vice versa (Fig. 2F-G). These results
278 suggest that CTCF on transcribing boundaries is recruited partly via the RNAs and different
279 transcribing CTCF sites within a boundary are potentially interdependent. Notably, shRNA
280 mediated knockdown or over-expression of CTCF had none to marginal effects on non-coding
281 RNA expression (Fig. 2H-I and Fig. S6D), which suggests that CTCF enrichment at TAD
282 boundaries is regulated by RNA but RNA levels are independent of CTCF levels. However,
283 persistent CTCF binding at B1 and B2 sites even after its acute depletion cannot be ruled out and
284 hence, can bias this result.
285
286 **Non-coding RNAs and CTCF physically interact at 9p21 TAD boundary**
287
288 Since, CTCF occupancy at B1 and B2 sites was reduced upon RNaseA treatment and RNA
289 knockdowns (Fig. 2), we asked if RNA and CTCF interact as shown on total RNA (Hansen et al.,
290 2019; Saldaña-Meyer et al., 2019) (Fig. 3A). Towards this, we first performed RNA pulldowns on
291 sense and anti-sense RNAs transcribed from B1 and B2 CTCF sites (Fig. S5B). We observed an
292 interaction between these RNAs and CTCF (Fig. 3B). Interestingly, an eRNA from E8 enhancer
293 within the super enhancer was also tested and exhibited interaction with CTCF (Fig. 3B). E8 also
294 shows CTCF occupancy on its genomic region (Fig. 2C). However, CTCF failed to interact with

295 BoxB RNA (viral RNA recognized by anti-terminator protein λN) (Fig. 3B). These observations
296 were confirmed by another orthogonal technique UV-RIP which captures the binding of RNA and
297 proteins *in vivo* (Fig. 3C).
298
299 Though, the CTCF occupancy was high on transcribing boundaries (Fig. 1D-E) but surprisingly,
300 the CTCF motif was weakly present in these boundaries as compared to non-transcribing and
301 random boundaries (Fig. 3D). This suggests that RNA may allow high affinity CTCF binding to
302 less favourable CTCF motifs. This RNA aided CTCF occupancy was also evident from more
303 number of CTCF peaks on transcribing boundaries (Fig. 3E) and more signal per peak (Fig. S6F).
304 Similarly, higher peak per bp occupancy of CTCF on transcribing boundaries despite of them
305 having weaker motif was seen on genic (Fig. 3F) and non-genic boundaries (Fig. 3G) with high
306 RNA. These data suggest that presence of RNA can overcome the limitation of a weak CTCF
307 motif and can assist in more numbers of CTCF binding events and with high affinity on genic and
308 non-genic transcribing boundaries.
309
310 Next, we compared the stability, polyadenylation and splicing features of RNAs at TAD boundary
311 with other RNAs. We compared genic transcripts arising from genic boundaries with other genic
312 RNA that are not within the boundaries. Similarly, transcribing non-genic boundaries show
313 presence of enhancer like features (Gong et al., 2018; Sun et al., 2019) thus, we compared these
314 non-coding RNAs with eRNAs. For turnover of RNA, we compared the ratio of nascent
315 transcription with steady levels of these RNAs within the same experimental setting. In general,
316 eRNAs exhibit fast turnover as seen by higher ratio of nascent: mature RNA as compared to genic
317 RNA (Fig. 3H) (Andersson et al. 2014). Interestingly, the genic RNA within boundaries were
318 slightly more stable compared to other genic RNAs (Fig. 3H). Similarly, non-coding RNAs on
319 non-genic boundaries exhibited lower ratio of nascent RNA: mature RNA as compared with
320 eRNAs (Fig. 3H) suggesting, non-coding RNAs on non-genic boundaries are more stable as
321 compared to eRNAs emerging from enhancers elsewhere. Perhaps, the stable nature of these RNAs
322 is required for stable binding of CTCF at transcribing boundaries.
323
324 Next, we tested if the genic and non-genic RNAs are different from eRNAs and mRNAs in terms
325 of transcriptional termination and 5' splicing sites. As expected, the TTS was least present around

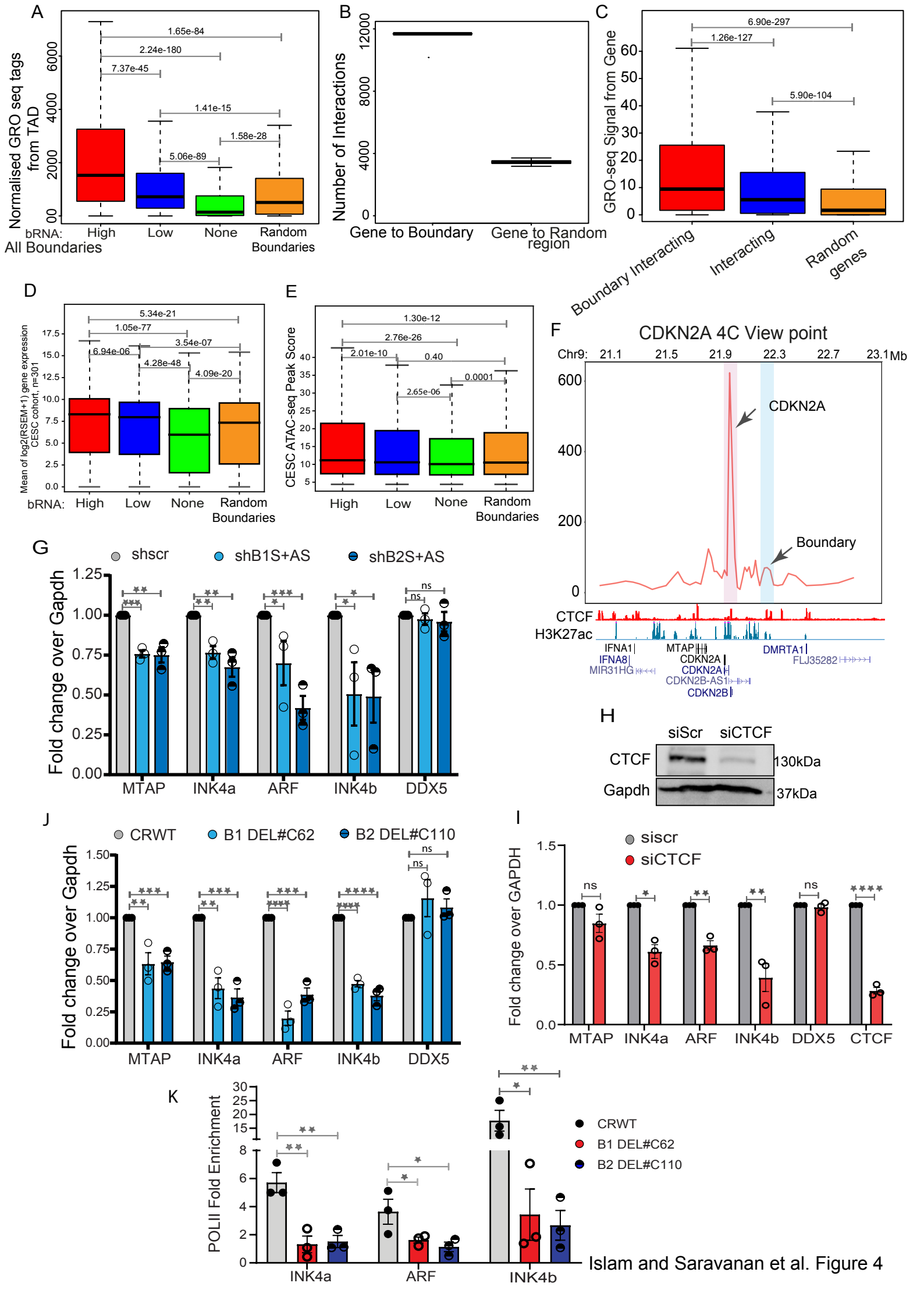


Islam and Saravanan et al. Figure 3

Figure 3. Non-coding RNAs and CTCF physically interact at 9p21 TAD boundary: A. Schematic showing CTCF domains and 11 zinc fingers. The blue lines indicate the reported regions important in its stabilization and RNA binding. **B.** Immunoblot for CTCF on RNA pulldowns performed on various RNAs (n=2). **C.** Agarose gel image showing the UV-RIP RT-PCRs for sense and anti-sense RNA arising from B1 and B2 CTCF sites (n=2). **D.** Line plots show the CTCF motif per 10 base pair per peak on various classes of boundaries. **E.** Fraction of boundaries with differential occupancy of CTCF peaks over the presence of the CTCF motif**.** **F.** CTCF peaks per kbp on all transcribing boundaries with high RNA, no RNAs and random boundaries (50 kb flanks on both side). **G.** CTCF peaks per kbp on non-genic transcribing boundaries with high bRNAs, no bRNAs and random boundaries. **H.** Log2FC of NET-CAGE+1/ CAGE+1, indicating stability of RNA at TSS identified from NET-CAGE at boundaries and TSS not at boundaries. b eRNA denotes the bRNA from non-genic boundaries, b mRNA denotes the mRNA from genic boundaries **I.** Average predicted TTS sites from the TSS identified in different classes of RNA by NET-CAGE. **J.** Average predicted 5'SS sites from the TSS identified by NET-CAGE. The p-values in boxplots were calculated by Mann Whitney test (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.

326 the TSS of genic transcripts as compared to non-genic RNAs and eRNAs (Fig. 3I) but, non-genic
327 RNAs on boundaries showed lesser enrichment of TTS as compared to eRNAs again pointing
328 towards their stable nature (Fig. 3I). 5' splicing sites were under-represented in eRNAs and RNAs
329 from non-genic boundaries compared to genic transcripts (Fig. 3J). Interestingly, RNAs from non-
330 genic boundaries exhibited slight enrichment of splice sites compared to eRNAs (Fig. 3J).
331
332 These data suggest that non-coding RNAs from non-genic boundaries are more stable and exhibit
333 less termination signal and more 5' splicing sites as compared to eRNAs. Further, even the genic
334 transcripts from genic boundaries exhibited more stability and 5' splice sites. Furthermore, these
335 RNAs are transcribed at the CTCF-rich regions of boundaries. Thus, overall the RNAs from
336 boundaries (coding or non-coding) have different features than other RNAs in the genome thus,
337 we named them as bRNAs (Boundary-RNA).
338
339 **TAD transcription is regulated by bRNAs**
340
341 Next, we compared the transcriptional potential of the TADs with high, low, random non-
342 transcribing boundaries and random boundaries and found that the transcriptional output of the
343 entire TAD (Fig. 4A) or of the individual genes (Fig. S7A) was correlated with the levels of bRNAs
344 at boundaries. This observation was found to be true on non-genic boundaries as well (Fig. S7B).
345 This data shows that the presence of not only the genic RNA but also the non-genic RNAs at the
346 boundaries is positively correlated with high transcriptional activity of genes with in the TAD. We
347 asked if this positive correlation is also exhibited as a result of high interactions between promoters
348 and these boundaries. Towards this, we analysed the publicly available promoter capture-seq data
349 in HeLa cells (Thiecke et al., 2020). The promoters exhibited higher interactions with boundaries
350 of respective TADs as compared to random regions, suggesting that promoters tend to interact
351 more with boundaries than expected (Fig. 4B). Further, we found that the genes that interacted
352 with boundaries were relatively more active than genes that interacted with any other genomic
353 regions (Fig. 4C). In order to test the generality of these observation, we asked if the transcribing
354 boundaries in HeLa show open chromatin features (a proxy for active transcriptional potential) in
355 cervical tumors in TCGA datasets. We observed a positive correlation between the level of bRNAs
356 in HeLa and ATAC-seq signal on the same boundaries in cervical tumors (Fig. S7C). Further, the

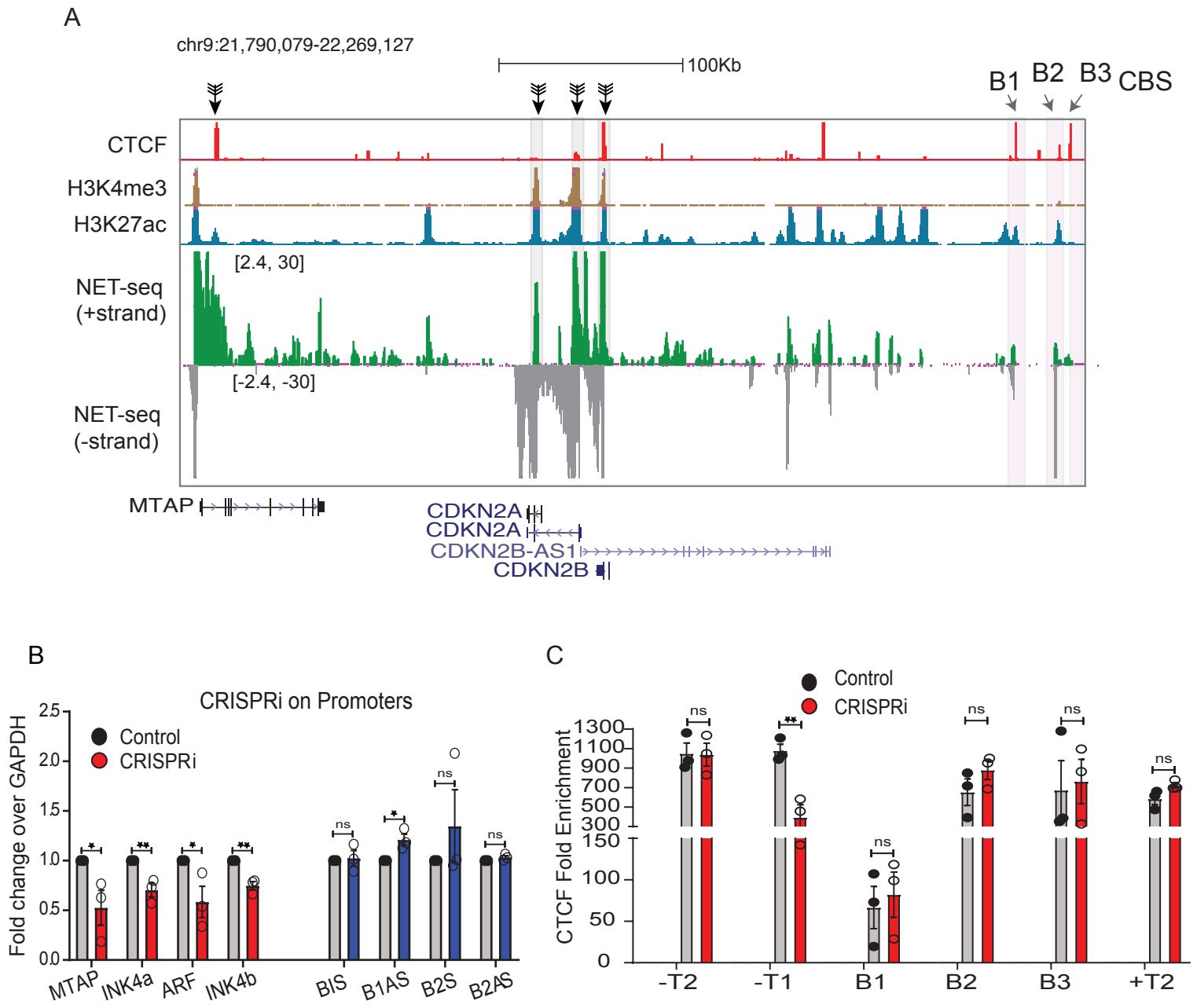
357 transcriptional status and overall ATAC-seq signal in these TADs was also proportional to bRNAs
358 levels. (Fig. 4D-E). Together, these data suggest that transcribing boundaries tend to interact with
359 promoters and are positively associated with the higher rate of TAD transcription.
360
361 Next, we confirmed the direct physical interaction between *CDKN2A* promoter and boundary
362 regions by 4C-seq at *CDKN2A* promoter (Fig. 4F). We then asked if bRNAs play any direct roles
363 in transcriptional activity of genes within this TAD. Interestingly, we observed downregulation of
364 INK4a, ARF (Transcribed from *CDKN2A*), INK4b (Transcribed from *CDKN2B*) and MTAP
365 expression upon shRNA mediated knockdown of both sense and anti-sense bRNAs at the B1 and
366 B2 CTCF sites, whereas the transcription of DDX5 (non-INK4a/ARF TAD gene) remained
367 unchanged upon these perturbations (Fig. 4G). These data suggest that *de novo* bRNAs positively
368 regulate the transcription of the genes within the *INK4a/ARF* TAD where promoter physically
369 interacts with TAD boundary.
370
371 However, the above observation cannot distinguish between the role of CTCF or bRNA *per se*,
372 since the loss of bRNA also causes the loss of CTCF binding at the TAD boundary. Thus, we
373 asked if loss of CTCF levels would also cause the gene repression. We observed downregulation
374 of INK4a/ARF genes upon depletion of CTCF levels using pool of specific siRNAs (Fig. 4H-I)
375 suggesting, bRNA and CTCF loss individually causes INK4a/ARF gene down regulation.
376 However, global CTCF knockdown affect all CTCF bound regions within the TAD. Hence, the
377 effects of other CTCF sites apart from B1 and B2 CTCF sites at boundary cannot be ruled out. To
378 test the effects of loss of CTCF at the boundary only, we deleted the B1 and B2 CTCF sites
379 individually in HeLa cells (Fig. S8A-B) and interrogated the expression of INK4a/ARF genes
380 within the TAD. Similar to bRNA and CTCF knockdown, we observed the down regulation of
381 these genes in boundary deleted cells (Fig. 4J). The downregulation of mRNA was also noticed at
382 p16 protein levels upon both, bRNA knockdown and boundary deletion (Fig. S8C). Further, the
383 loss of PolII at these promoters confirms the transcriptional inhibition of INK4a/ARF promoters
384 (Fig. 4K). These data suggest that the bRNA mediated loading of CTCF at the boundary
385 functionally regulates the expression of genes within the *INK4a/ARF* TAD.
386



Islam and Saravanan et al. Figure 4

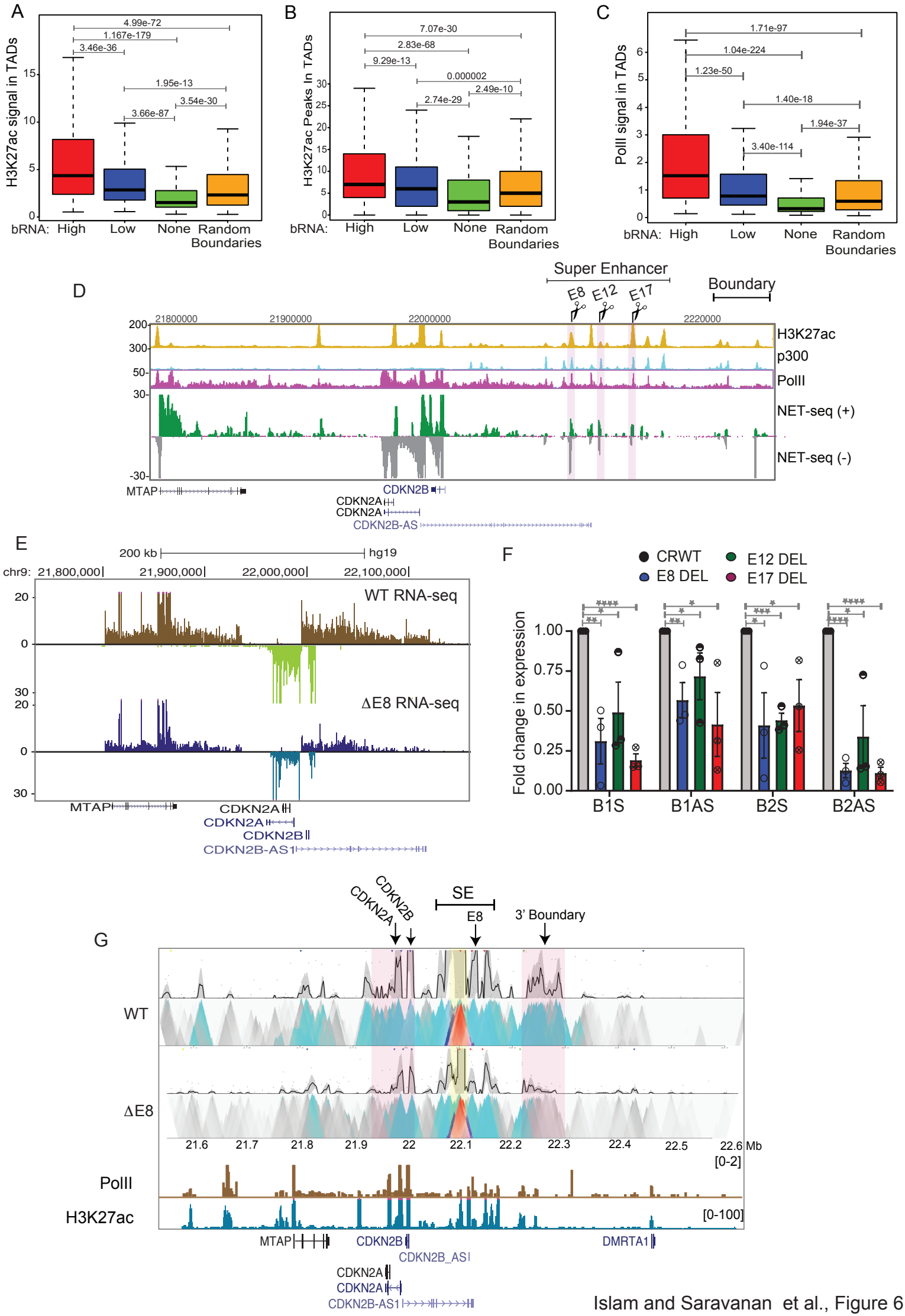
Figure 4. TAD transcription is regulated by bRNAs: A. Box plots showing GRO-seq tags from the entire TAD with high, low, non-transcribing and random boundaries. **B.** Graph showing the number of interactions between gene and; boundaries, and 100 randomizations of interactions between boundaries and random regions. **C.** Box plots showing the GRO-seq tag counts from genes when they interact with boundary or any region vs. random interactions. **D.** Box plots showing the RNA-seq expression of TADs with boundaries exhibiting different levels of bRNAs (HeLa) in Cervical Squamous Cell Carcinoma tumors from TCGA. **E.** Box plots showing the ATAC-seq signal of TADs with boundaries exhibiting different levels of bRNAs in Cervical Squamous Cell Carcinoma tumors from TCGA. **F.** 4C plot at CDKN2A promoter viewpoint, the interaction with boundary is highlighted in blue. The plot is overlaid with H3K27ac, CTCF and gene annotations in HeLa. **G.** qRT-PCR plot shows the change in gene expression upon shRNA mediated knockdowns of bRNAs arising from B1 and B2 CTCF sites at INK4a/ARF TAD boundary. **H.** Immunoblot showing the drop in CTCF protein levels upon its siRNA mediated knockdown. **I**. qRT-PCRs showing the down regulation of INK4A/ARF genes and levels of DDX5 and CTCF upon CTCF knockdown. **J.** qRT-PCR showing the changes in expression of different genes in INK4a/ARF TAD and DDX5 upon cas9 mediated independent deletion of B1 and B2 CTCF sites. **K.** 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 **G, I, J** and **K.** ****$p < 0.0001$,***$p < 0.001$,**$p < 0.01$,*$p < 0.05$, $^{ns}p > 0.05$. p-values in boxplots were calculated by Mann Whitney test (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.

387 **Transcriptional activity of promoters does not regulate non-genic transcription at**
388 **boundaries**
389
390 bRNAs regulate *INK4a/ARF* TAD transcription but the question remains as to what causes the
391 transcription of *de novo* bRNA at this boundary. We asked if the transcriptional activity of the
392 *INK4a/ARF* promoters indeed causes the transcriptional activation at the boundaries of the same
393 TAD. Towards this, we blocked promoters (*CDKN2A*, *CDKN2B* and *MTAP*) via specific gRNAs
394 and dCas9-KRAB (Fig. 5A) and tested the levels of bRNAs at B1 and B2 CTCF sites. We first
395 confirmed the successful CRISPRi by observing the transcriptional down-regulation of these genes
396 (Fig. 5B, red bars). Interestingly, the levels of bRNAs were unaffected upon such transcriptional
397 inhibition of the promoters (Fig. 5B, blue bars). In concordance with unperturbed bRNA
398 expression, we did not observe any change in CTCF occupancy at B1 and B2 CTCF sites but we
399 observed the loss of CTCF on 5' boundary that overlaps with MTAP gene potentially due to down-
400 regulation of *MTAP* gene (Fig. 5C). The boundary of flanking TADs also did not exhibit any
401 perturbations in CTCF occupancy (Fig. 5C). These data suggest that bRNAs regulate the
402 transcription of promoters in *INK4a/ARF* TAD but this transcriptional activity is not the cause of
403 bRNA expression at the boundary.
404
405 **bRNAs are positively regulated by enhancers within the TAD**
406
407 Since, the transcriptional potential of TADs with transcribing boundaries (genic RNA and non-
408 genic *de novo* RNA) was higher, we asked if the number of active enhancers in these TADs were
409 also high to account for the active gene transcription. In agreement with high transcriptional levels
410 of genes, these TADs were also enriched in total H3K27ac levels and number of H3K27ac marked
411 enhancers (Fig. 6A and B). Notably, PolII enrichment was also high in accordance with the
412 presence of high gene transcription and active enhancers in these TADs as compared to the TADs
413 with non-transcribing boundaries and random boundaries (Fig. 6C). These observations were also
414 true for non-genic boundaries (Fig. S9A-C). To test if there is a regulatory axis between enhancers
415 and bRNA expression at *INK4a/ARF* TAD, we used series of lines with active enhancer deletions
416 within the super enhancer (E8, 12 and 17) (Fig. 6D, Farooq et al., 2021). The enhancer cluster
417 contains 21 enhancers and down-regulation of *INK4a/ARF* genes was observed by RNA-seq in



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 antisense bRNA 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.** $^{**}p < 0.01$, $^{*}p < 0.05$, $^{ns}p > 0.05$.

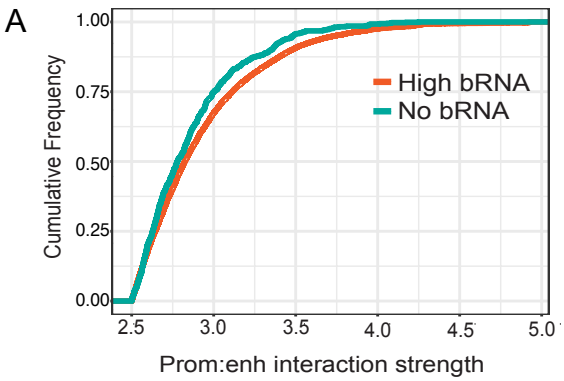


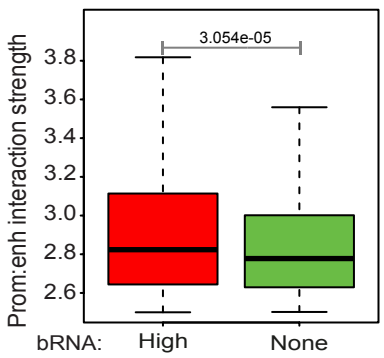
Islam and Saravanan et al., Figure 6

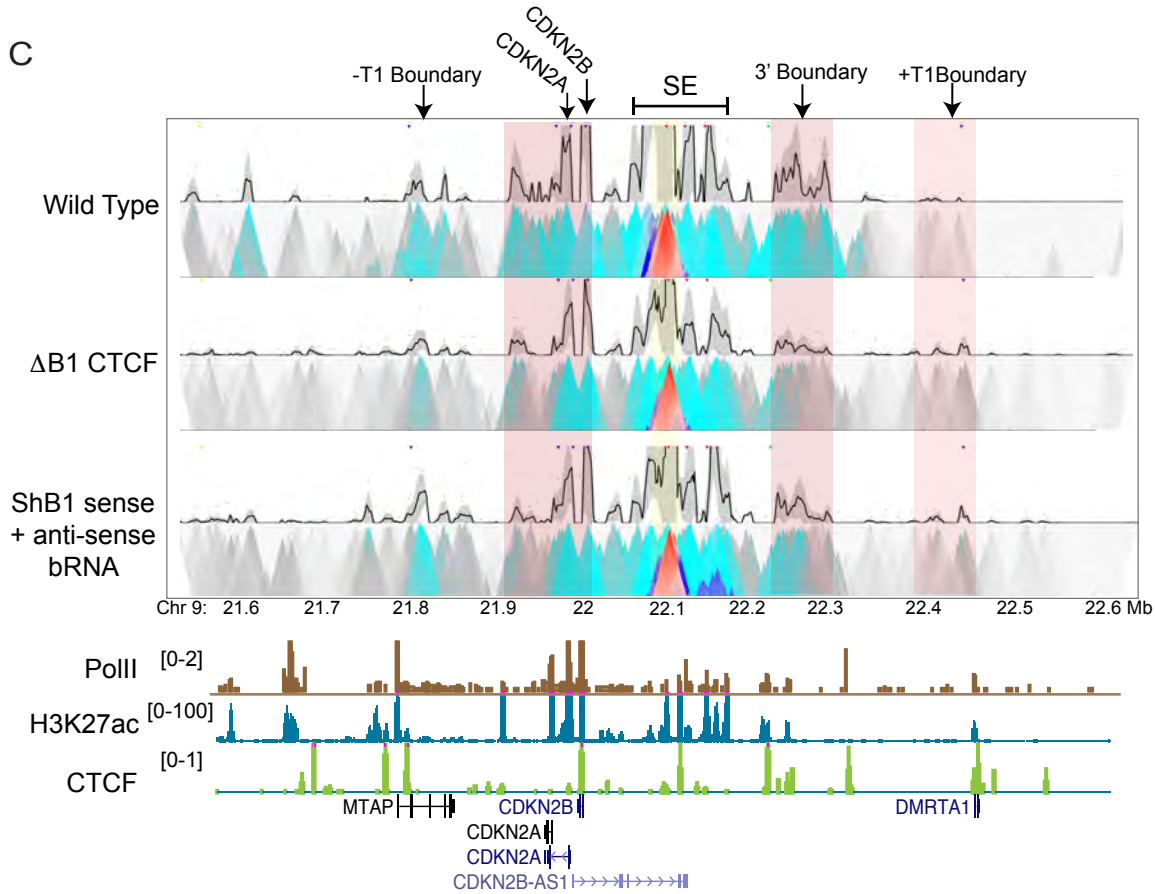
Figure 6. bRNAs are positively regulated by enhancers within the TAD: A. Boxplots displaying enrichment of H3K27ac in TADs with high, low, non-transcribing and random boundaries. **B.** Boxplot showing number of H3K27ac peaks in TADs with various boundaries. **C.** Levels of PolII in TADs with different boundaries. **D.** UCSC browser shot showing H3K27ac, p300, PolII ChIP-seq and mNET-seq on 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 line. **F.** qRT-PCR shows levels of sense and antisense bRNA arising from B1 and B2 CTCF sites upon E8, E12 and E17 enhancer deletions. **G.** 4C-seq plot in WT and in E8 deletion line on an enhancer beside E8 enhancer as a viewpoint (Yellow highlight) exhibiting interactions with promoters and 3' boundary region (Pink highlight). The interaction heatmap is overlaid by 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 Mann Whitney test (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.

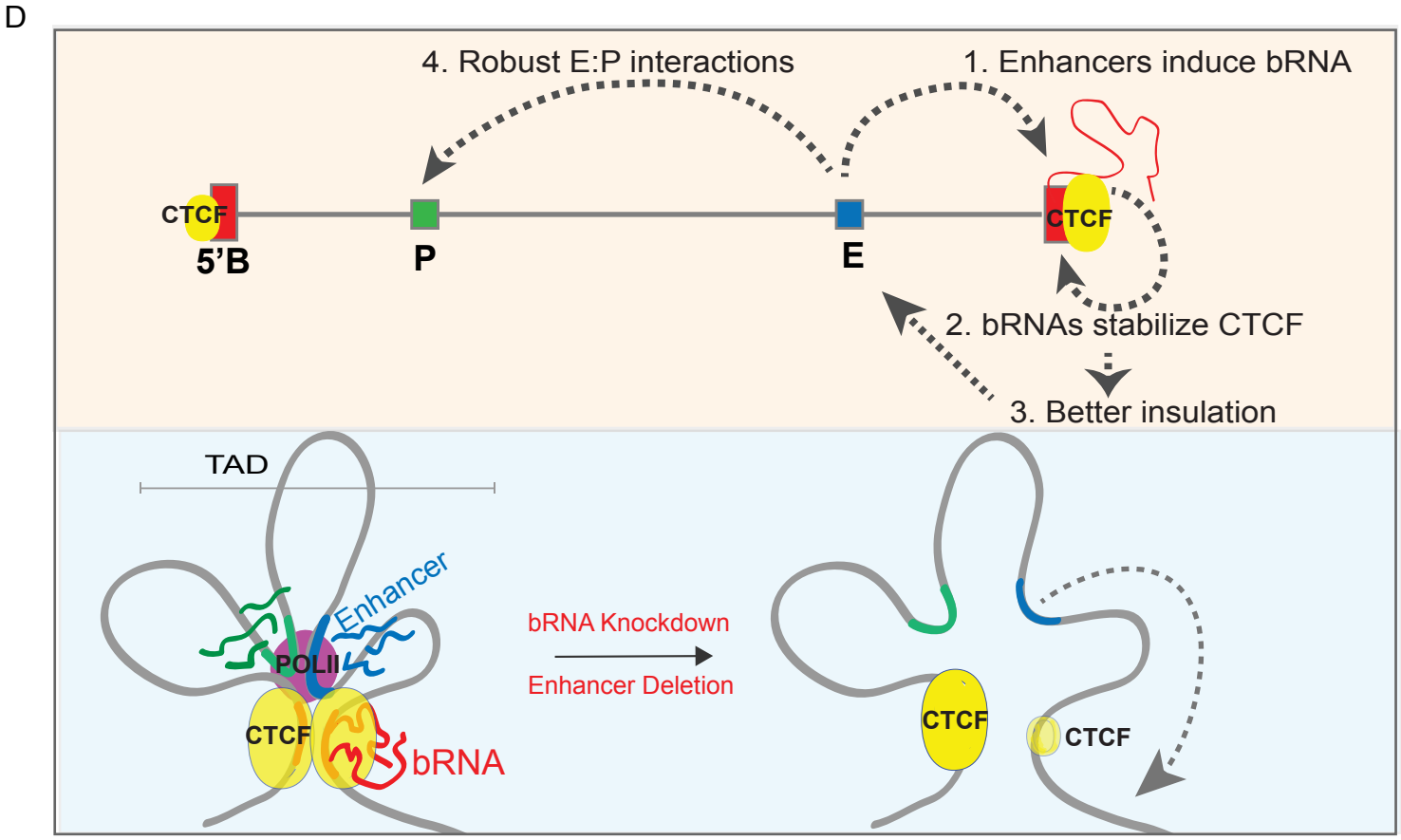
418 these deleted lines (Fig. 6E) (Farooq et al., 2021), confirming that these enhancers are functional
419 enhancers. We then asked if these enhancers indeed regulate the boundaries as well. Interestingly,
420 we noted the down-regulation of sense and anti-sense bRNAs at the B1 and B2 CTCF sites upon
421 E8, E12 or E17 deletions (Fig. 6F) suggesting, enhancers indeed transcriptionally activate *de novo*
422 bRNAs at the boundary region. Further, we performed 4C on an enhancer beside E8 enhancer as
423 a viewpoint. As expected, we observed strong interaction of this enhancer with other enhancers in
424 the super enhancer cluster including E8. We also noted its interactions with *CDKN2A* and
425 *CDKN2B* promoters and with the boundary region (Fig. 6G, right highlighted pink). Intriguingly,
426 we observed the loss of SE interaction with the boundary and the promoters upon deletion of the
427 E8 enhancer (Fig. 6G). Such interactions between boundaries and enhancers are reported in
428 literature (Sun et al., 2019) but it is unknown if they have any regulatory relation. These data
429 suggest that apart from activating their target promoters, functional enhancers positively regulate
430 the transcription at the boundaries by direct interactions. To test if high levels of bRNAs and
431 enhancer activity insulate the boundary better *in vivo* we further interrogated the INK4a/ARF TAD
432 upon senescence where TAD gains more activation transcriptionally due to the cell cycle
433 regulatory roles of this locus. We analyzed publicly available data on proliferative and senescent
434 fibroblasts (replicative senescence). Data suggests that upon senescence, the bidirectional eRNA
435 expression increases by many folds on the super enhancer cluster (Fig. S10A, blue highlighted
436 region). The INK4a/ARF gene expression and the bRNA expression also increases under this
437 scenario (Fig. S10A, pink highlighted region). We validated the gain of gene transcription, bRNA
438 expression and CTCF enrichment at the boundary upon X-ray induced senescence (non-replicative
439 senescence) (Fig. S10B-D) and the elevated p16 and p21 protein levels (Fig. S10E). Using HiC
440 data, we then observed the gain of intra-TAD interactions at this TAD in senescent cells as
441 compared to proliferative cells (Fig. S10F). Together these data suggest that as the enhancers gain
442 more activity, they interact with boundaries to induce bRNA expression. The induced bRNA in
443 turn recruits more CTCF to increase the insulation.
444
445 **bRNAs provide better insulation to favour robust enhancer:promoter interactions**
446
447 Enrichment of CTCF at boundary regions is an indicator of a better insulated boundary and due to
448 the better insulation, enhancers are contained within respective TADs. Since, the levels of CTCF

449 and insulation score was high on transcribing boundaries thus, we asked if the enhancer and
450 promoter interactions were stronger within the TADs of transcribing boundaries vs. non-
451 transcribing boundaries. Interestingly, we observed relatively higher strength of
452 enhancer:promoter interactions in these TADs as compared to the TADs with non-transcribing
453 boundaries (Fig. 7A-B). Further, since bRNA knockdown reduced the CTCF occupancy at the
454 boundary and also the transcription of genes was downregulated in the *INK4a/ARF* TAD (Fig. 2,
455 4), we asked if the insulation of this TAD and SE:promoter interactions were also affected upon
456 bRNA depletion. Towards this, we performed 4C assays at enhancer viewpoint upon knockdown
457 of sense and anti-sense bRNAs at the B1 CTCF site. Not surprisingly and in concordance with
458 CTCF levels loss, we observed the loss of boundary: enhancer interactions (Fig. 7C). Importantly,
459 we observed the weakening of enhancer: INK4a/ARF promoter interactions as well. Further, the
460 enhancers gained interactions with the +1 TAD boundary indicating that the enhancer was now
461 able to interact with the boundary of the next TAD (Fig. 7C). Intriguingly, we also observed the
462 gain of CTCF at +T1 boundary upon bRNA depletion (Fig. 2F). Then to test if these inter-TAD
463 interactions were indeed indicative of loss of boundary functions, we extended the 4C on the same
464 enhancer viewpoint in cells with B1 CTCF site deletion. We noted a similar loss of enhancer:
465 promoter and enhancer: boundary interactions and gain of interactions with +T1 boundary region
466 (Fig. 7C). This data confirms that bRNA loss weakens the insulation of the TAD which is similar
467 to boundary loss (Fig. 7D).
468
469 **Discussion**
470
471 RNaseA treatment followed by HiC shows the perturbations in 3D chromatin architecture
472 suggesting, the RNA is an integral player in chromatin organization (Barutcu et al., 2019). Further,
473 genome organizer CTCF interacts with RNA and mutations in its RNA binding ZF1 and ZF11
474 exhibit weakening of TADs favoring more inter-TAD interactions (Hansen et al., 2019; Saldaña-
475 Meyer et al., 2019). However, these experiments have not provided the identity of RNA and its
476 transcribing sites that are important in such chromatin organization roles. Towards this,
477 transcription at TAD boundaries is speculated to regulate TAD functions (D. Zhang et al., 2020;
478 Zhang et al., 2019). Here we show that bRNAs emerging at TAD boundaries (from genes
479 overlapping with boundary or *de novo* non-coding transcripts) have a direct role in CTCF









Islam and Saravanan et al. Figure 7

Figure 7. bRNAs provide better insulation to allow robust enhancer:promoter interactions: A. Cumulative frequency of the interaction strength between enhancer and promoters interacting with boundaries showing high and no transcription. **B.** Enhancer:promoter interaction strength within TADs with high and non-transcribing boundaries. **C.** 4C-seq plot on region beside E8 enhancer as a viewpoint (Yellow highlight) in WT and upon B1 CTCF site deletion (ΔB1) and knockdown of B1 sense and anti-sense bRNAs. The anchored region exhibiting interactions with promoters and 3' boundary region and +T1 boundary (Pink highlights). The interaction heatmap is overlaid by PolII, H3K27ac and CTCF ChIP-seq tracks and gene annotations. **D.** Model depicting that: 1. Active enhancer physically interacts with boundary and activates the bRNA transcription at the boundary. 2. The bRNA in return stabilizes the CTCF at these boundaries thereby, 3. The insulation of these TADs is strengthened. 4. This favors intra-TAD enhancer: promoter interactions to facilitate robust gene transcription. The loss of bRNA/enhancer deletion reduces the bRNA levels which triggers the loss of CTCF and insulation of TADs. Enhancer:promoter interactions weaken in these scenario which causes concomitant loss of gene transcription. Kolmogorov–Smirnov two-sided test was used to calculate significance in (A). p-values in boxplots were calculated by Mann Whitney test (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.

480 recruitment at these boundaries and therefore in TAD strength and gene regulation. Our data,
481 provide the direct evidence that indeed bRNAs at boundary create insulated neighbourhood by the
482 virtue of binding with CTCF that favours enhancer:promoter interactions for robust gene activation
483 thus, bRNAs at boundaries make genome topology less "hardwired" and more regulated.
484
485 Using a series of enhancer and CTCF sites deletions at INK4a/ARF TAD, we observed that
486 enhancers interact and positively regulate the expression of bRNA (gene transcripts or non-coding
487 RNAs) at TAD boundaries. bRNAs in turn stabilizes the binding of CTCF at the boundary, in
488 particular on weak CTCF motifs (Fig. 3D). These boundaries show more number of CTCF peaks
489 (Fig. 3E-F) and hence, the TAD boundary is better insulated. With these CTCF features, perhaps,
490 bRNA transcribing boundaries do not let cohesin ring slip to the next TAD (Hansen, 2020;
491 Kentepozidou et al., 2020) thereby insulation is enhanced. Any RNA perhaps would add in CTCF
492 recruitment if CTCF motif is present however, the transcription sites due to high levels of overall
493 RNA in surroundings would do this efficiently. Furthermore, the stable nature of RNA at
494 transcribing boundaries would be crucial for stable loop formation. Towards this, perhaps eRNAs
495 within the boundaries are repurposed as bRNAs for these added functions.
496
497 During development or signaling cues when enhancers are activated, the weaker CTCF sites if
498 divergent may act as a stronger boundary by the virtue of bRNA and may thereby stabilize the
499 dynamic chromatin changes that are the result of enhancer activation in these TADs giving rise to
500 differential gene regulation. For example, we observed higher insulation of INK4a/ARF TAD
501 upon onset of senescence where the TAD transcription is induced and coincides with the gain of
502 enhancer functions and bRNA expression at the TAD boundary (Fig. S10A, F).
503
504 The bRNA loss reduces the enrichment of CTCF at the boundary and shows concomitant loss of
505 enhancers and promoter interactions at INK4a locus (Fig. 2F, 7C). Whereas, interactions between
506 these enhancers and the boundary (+T1) of neighbouring TAD strengthen, this gained looping also
507 enhances the CTCF enrichment at +T1 (Fig. 2F, 7C) suggesting, the insulation of INK4a/ARF
508 boundary upon bRNA perturbations is reduced. These data indicate that bRNAs stabilize CTCF at
509 the boundary thus, interactions between the genes within these TADs and their cognate enhancers
510 is favored resulting in robust transcription of genes due to better insulation of these TAD

511 boundaries. These results contrast the observations that deletion of boundary had no effect on E-P
512 interaction (Yokoshi et al., 2020) but agree with other literature (Arzate-Mejía et al., 2020).
513
514 Transcriptional inhibition also affects the chromatin interactions. This could arise from RNA *per*
515 *se* or from the PolII and its associating machineries that occupy the enhancers and promoters.
516 Towards this, we used CRIPSRi that not only inhibits the transcription thereby the transcripts but
517 also inhibits the binding of PolII machinery. Interestingly, CRISPRi mediated active silencing of
518 TAD did not alter the expression of bRNA and thereby CTCF occupancy at the boundary (Fig. 5
519 B-C) suggesting that neither transcriptional activity of promoters or the alterations in the PolII
520 machinery, directly change the bRNA mediated insulation of boundary. Thus, the effects on
521 genome organization upon transcription inhibition perhaps do not arise from the promoters. Our
522 data suggests that observed effects on TAD topology upon transcription inhibition could arise from
523 inhibition of enhancer transcription thereby its activity which can alter the boundary functions
524 leading to compromised TAD insulation as seen by the loss of bRNA and super enhancer-boundary
525 interactions upon deletion of E8 enhancer in INK4a/ARF TAD (Fig. 6F-G).
526
527 These observation are in agreement with the fact that TADs with super enhancers exhibit stronger
528 boundaries (Gong et al., 2018; Sun et al., 2019). This is also evident by the presence of more
529 number of enhancers in TADs with bRNA transcribing boundaries (Fig. 6A-B and S9A-B). The
530 data indicates that as the TAD acquires enhancer activity, the enhancers loop with boundary to
531 activate the bRNA expression. Then, the boundary insulates the TAD better thereby enhancers
532 stably engage with the cognate promoters. The transcription assisted TAD boundary strengthening
533 also corroborates with the observed zygotic activation dependent boundary formation (Hug et al.,
534 2017) and in X chromosome inactivation (Giorgetti et al., 2016). Further, a combination between
535 CTCF site at the boundary and active enhancers are required to induce bRNA-dependent insulated
536 boundary. This suggests that mere activation of transcription at the boundary is not sufficient to
537 create the TAD boundary (Bonev et al., 2017; Hug et al., 2017). Together, these results show that
538 active enhancers directly mediate better insulation of TADs by activating the transcription at TAD
539 boundaries which favours robust intra-TAD enhancer:promoter interactions to activate the gene
540 transcription.
541

542 The boundaries per se also harbor H3K27ac and PolII marked regions suggesting the presence of
543 active regulatory elements at these boundaries (Gong et al., 2018; Sun et al., 2019; Hsieh et al.,
544 2020; Matthew et al., 2018). The *de novo* RNA (bRNA) that emerges from these regions tend to
545 be bidirectional resembling the bi-directional nature of eRNAs emerging from functional
546 enhancers (Li et al., 2016, 2013). The strong CTCF binding along with DNA structural features at
547 these active regions in boundaries may give these elements a structural and functional ability that
548 is different from CTCF-bound enhancers therefore, these boundaries functionally behave
549 differently. Nonetheless, CTCF binding, insulation at the boundaries and enhancer:promoter
550 interactions in TADs corresponds to the bRNA levels (Fig. 1D-E, F and Fig. S1) suggesting the
551 RNA (from genes or from regulatory elements within boundaries) potentially determines the
552 strength of these boundaries.
553
554 **Limitations of the study:**
555
556 While our study provides bRNA as the additional layer that determines the boundary insulation,
557 the other players such as transcription factors and DNA sequences that augments the insulation
558 function of TAD boundaries are to be discovered. The role of the dynamics of RNA with cohesin
559 has not been probed in this manuscript but may add to the above mentioned mechanism. Further,
560 apart from binding with CTCF, if the stable nature of bRNAs provide any added function over
561 eRNAs remains to be studied.
562
563 **Acknowledgements:** We thank Tapas K. Kundu for the discussions. We are grateful to Henrietta
564 Lacks and to her family members for their contributions to biomedical research. We acknowledge
565 support of the Department of Atomic Energy, Government of India, under project no.12-R&D-
566 TFR-5.04-0800 and intramural funds from NCBS-TIFR (to DN and RS). DN is an EMBO Global
567 Investigator. We also acknowledge the funding support from Welcome-IA: IA/1/14/2/501539 (to
568 DN). UF is supported by CSIR (Govt. of India) senior research fellowship. KW and RM are
569 supported by the TIFR-NCBS graduate program. KW is also supported by the CSIR-SPM (Govt.
570 of India) fellowship.
571

572 **Contribution:** We thank the members of DN lab for comments and discussion on the manuscript.
573 The study was conceived and designed by ZI, BS and DN. BS performed NGS data analysis. Most
574 experiments were performed by ZI with help from UF and KW. AKS and RS executed tumor
575 dataset analysis. JT performed CTCF Cut & Run. RNaseA experiment in SH lab. AP performed
576 NGS library preparation and sequencing. DN, ZI and BS wrote the manuscript with inputs from
577 all the authors.
578
579 **Materials and Methods:**
580
581
582 **Cell Culture:**
583
584 HeLa cells were obtained from ATCC and were cultured in (DMEM) Dulbecco's Modified Eagle's
585 medium supplemented with 10% Fetal Bovine Serum and 1% Penicillin/Streptomycin. The cells
586 were maintained at 37°C in presence of 5% CO_2 in a humidified incubator
587
588 **Chromatin Immunoprecipitation:**
589
590 Cells were crosslinked with 1% formaldehyde at room temperature for 10 mins with constant
591 shaking. Glycine was added to a final concentration of 125 mM to quench the formaldehyde. Cells
592 were washed three times with 1X ice cold PBS then were scraped and pelleted down 1X PBS at 4
593 ^{o}C. Cells were gently resuspended in nuclear lysis buffer (50 mM Tris-HCl pH 7.4, 1% SDS, 10
594 mM EDTA pH 8.0) supplemented with 1X PIC and incubated on ice for 10 mins. Lysate was
595 subjected to fragmentation using Diagenode Bioruptor Pico for 20 cycles (30 s ON and 30 s OFF)
596 to generate fragments of around ~500 bp. Cell lysate was cleared at 12K rpm for 12 mins. 100 ug
597 of sheared chromatin was taken for each IP. The lysates were diluted by adding Dilution Buffer
598 (DB) (20 mM Tris-HCl pH 7.4, 100 mM NaCl, 2 mM EDTA pH 8.0, 0.5% Triton X-100)
599 supplemented with 1X PIC in 1:1.5 ratio (1 volume of sheared chromatin and 1.5 volumes of
600 Dilution Buffer). 1 ug of antibody was added to immunoprecipitate the DNA and incubated on a
601 rocking platform overnight at 4°C. 15ul of Pre-blocked (with 1% BSA) Protein G dynabeads
602 (10004D, Invitrogen) were added to the tubes and rotated for 4h at 4 °C. Beads were collected,
603 flow through was discarded and the samples were washed with Wash Buffer I (20 mM Tris-HCl
604 pH 7.4, 150 mM NaCl, 0.1% SDS, 2 mM EDTA pH 8.0, 1% Triton X-100) at 4 ^{o}C on a rocking

605 platform. Washings were sequentially repeated with Wash Buffer II (20 mM Tris-HCl pH 7.4, 500
606 mM NaCl, 2 mM EDTA pH 8.0, 1% Triton X-100), Wash Buffer III (10 mM Tris-HCl pH 7.4,
607 250 mM LiCl, 1% NP-40, 1% Sodium Deoxycholate, 1 mM EDTA pH 8.0 and 1X TE (10 mM
608 Tris pH 8.0, 1mM EDTA pH 8.0). Chromatin was eluted with 200 ul of Elution Buffer (100 mM
609 NaHCO3, 1% SDS) for 30 mins at 37ºC in a thermomixer. The supernatant was collected in
610 separate tubes and 14 µl NaCl (5 M) was added to eluted samples and kept at 65 °C overnight for
611 de-crosslinking. Immunoprecipitated DNA was purified by Phenol:chloroform:isoamyl alcohol,
612 followed by ethanol precipitation. The final air dried DNA pellet was dissolved in 100 ul of 1X
613 TE. These samples were then used for qRT-PCRs using SYBR Green. Mean values for all the
614 regions analyzed in different conditions were expressed as fold-enrichment compared over beads.
615 Statistically significant differences between different conditions were computed using a ***T*-test** (*p*-
616 value < 0.05), ChIP primers are listed in Supplementary Table. 1.
617
618 **RNA Isolation and cDNA synthesis:**
619
620 Cells were lysed in 1 ml of Trizol (Thermo Fisher Inc.). 200 ul chloroform was added to the
621 sample, briefly vortexed and centrifuged at 12K rpm for 12 mins. The aqueous phase was carefully
622 collected and transferred to the fresh tube. 1 volume of isopropanol was added to the sample and
623 incubated at room temperature for 10 mins to precipitate the RNA. The samples were centrifuged
624 at 12K rpm for 12 mins, supernatant was discarded without disturbing the pellet. The pellet
625 obtained was washed with 75% ethanol. Pellet was air dried and dissolved in RNase free water.
626 RNA obtained was treated with ezDNase (Invitrogen) to remove the traces of contaminating DNA.
627 1 µg of RNA was used for each cDNA synthesis reaction by Superscript IV (Invitrogen) and
628 random hexamers as per manufacturer's recommendation. The CFX96 touch (Biorad) real time
629 PCR was used for qRT-PCRs. qRT-PCRs were performed using three biological replicates for
630 each sample. The fold changes were calculated by the ΔΔCt method and individual expression
631 data was normalized to GAPDH mRNA. The *p*-values were calculated by Student's unpaired two-
632 tailed *t*-test. qRT-PCR primers are listed in Supplementary Table 1.
633
634 **CRISPR Cas9 mediated deletion:**
635

636 gRNAs were designed with crispr.mit.edu tool. gRNAs were selected based on the highest score
637 and a least number of off targets. All the gRNAs were cloned in pgRNA humanized vector
638 (Addgene #44248) (a gift from Stanley Qi Lab) between BstX1 and Xho1 restriction sites. gRNAs
639 were co-transduced with lenti-Cas9 vector. Cells were put under selection using puromycin
640 (3ug/ml) for 48 hours. Single cells were seeded in a 96 well plate. Wells with single cells were
641 marked and allowed to grow till they formed colonies. The cells were trypsinzed and shifted to 48
642 well plates and allowed to grow to confluence and then half of the cells were taken for the surveyor
643 assay and the other half was plated again. Surveyor assay was carried out by PCR based method
644 using primers listed in the Supplementary Table 1.

645 **CRISPRi: gRNA design and cloning:**

646 We used gRNA design tool (https://portals.broadinstitute.org/gpp/public/analysis-tools/sgrna-
647 design) to design the gRNAs against TSS of *MTAP*, *CDKN2A* and *CDKN2B* in the INK4a/ARF
648 TAD. For each gene, two gRNAs were designed. The gRNAs were cloned in a customized PX459
649 vector (pSpCas9(BB)-2A-Puro V2.0, Addgene # 62988) was a gift from Feng Zhang lab. The Cas9
650 enzyme cassette was removed from PX459 by digesting it with Xba1 and Not1 followed by blunt
651 end ligation. The cloning of the gRNAs in the customized PX459 plasmid was performed as per
652 the Zhang Lab's general cloning protocol. dcas9 plasmid along with gRNA specific to genes were
653 transfected in 70% confluent cells. Transfections were done using Lipofectamin 2000 (Invitrogen,
654 cat no# 11668027), CRISPRi cells were harvested for ChIP and gene expression analysis.

655 **Lentiviral transduction:**

656

657 HEK293FT cells were grown in poly-D lysine coated culture dishes. These cells were co-
658 transfected with lentiviral packaging plasmids like VSVG a gift from Bob Weinberg (addgene
659 #8454); PAX2 a gift from Didier Trono (addgene #12260) along with the plasmid of interest using
660 Lipofectamine 2000 Invitrogen (cat# 11668027). Media was changed after 6 hours. Viral soup was
661 collected after 48 h and 72 hours, pooled together, filtered with 0.44µm syringe filter and then
662 finally added to cells along with 8ug/ml of polybrene. Transduction was stopped after 16 hours.
663

664 **RNA Immunoprecipitation:**

665

666 For RNA, primers were designed to amplify desired genomic regions that correspond to peak of
667 bRNA (GRO-seq) at B1 and B2 CTCF sites of TAD boundary. The amplified products were cloned
668 in pcDNA3 BoxB a gift from Howard Chang (Addgene #29729) plasmid. All the clones were
669 confirmed by sanger sequencing, and subsequently used for RNA synthesis using T7 RNA
670 polymerase (Promega Inc. P207e) and biotin RNA labeling mix (Roche11685597910). Primer
671 sequences are mentioned in the Supplementary Table 1.

672

673 RNA immunoprecipitation was adopted from (Jayani et al., 2017), with some modifications. HeLa
674 cells were grown to 90–95% confluency and were harvested by scraping and washed twice with
675 ice cold 1X PBS to gently resuspend in 2 ml of PBS. 2 ml of nuclear isolation buffer (NIB)(40
676 mM Tris–HCl (pH 7.5), 20 mM MgCl2, 1.28 M sucrose, 4 % Triton X-100, 1 mM PMSF, protease
677 inhibitors, and 20 U/ml SUPERase inhibitor (Thermo Fisher Scientific, catalog # AM2694) was
678 added and the pellet was gently resuspend. Then, 6 ml of distilled water was added and kept on ice
679 for 20 min with intermittent gentle shaking. Nuclei were Pelleted by centrifugation at 2500×*g* for
680 5 min at 4ºC. The pellet was resuspended in 1 ml of RIP buffer (25 mM Tri–HCl (pH 7.4), 150
681 mM KCl, 0.5 mM DTT, 0.5 % NP40, 1 mM PMSF, protease inhibitors, and 20 U/ml SUPERase
682 inhibitor) and incubated on ice for 5 min. Nuclei were sheared by 10 cycles of sonication (30s on
683 and 30 s off, bioruptor) followed by centrifugation at 12K rpm for 10 min at 4ºC. 1 µg of
684 biotinylated RNA was incubated with 20 µl of RNA structure buffer (RSB) (10 mM Tris–HCl (pH
685 7.0), 100 mM KCl, 10 mM $MgCl_2$, 1 mM PMSF, protease inhibitors, and 20 U/ml SUPERase
686 inhibitor) at room temperature for 5 min. Folded RNA was mixed with 1 mg of nuclear extract in
687 500ul of RIP buffer and rotated at 4^0C for 1h. 15 ul of DynabeadsMyOne Streptavidin beads were
688 added and rotation continued for one more hour. Samples were washed 3 times with RIP buffer
689 and beads with proteins were boiled in 2XSDS for 10 min. Immunoblotting for CTCF was
690 performed using the CTCF antibody (cell signaling, 3418).

691

692 **UV RIP:**

693

694 Ultraviolet-RNA Immunoprecipitation was performed as in (Rahnamoun et al., 2018) with some
695 modifications. Hela cells at confluency of 80–90% were crosslinked by UV irradiation in

696 Stratalinker UV cross-linker. Cells were washed three times with cold 1XPBS and scraped in PBS
697 before subjecting to centrifugation at 2500 rpm. Pellet was lysed in RIP lysis buffer (25 mM
698 HEPES-KOH at pH 7.5, 150 mM KCl, 0.5% NP40, 1.5 mM $MgCl_2$, 10% glycerol, 1 mM EDTA,
699 0.4 U RNase inhibitor (Invitrogen, 18091050), and 1X protease inhibitor cocktail (PIC) on ice for
700 30 min. Cleared cell lysates were subjected to IP with CTCF antibody bound Protein G Dynabeads
701 (Invitrogen, 10004D) overnight. Samples were subsequently washed three times with RIP lysis
702 buffer and RNA samples were eluted using TRIzol reagent. Prior to cDNA synthesis RNA was
703 treated with RNase-Free DNase Qiagen (cat no.79256) to remove the traces of contaminating
704 DNA. cDNA was prepared using random hexamers by Superscript IV (Invitrogen 18091050) as
705 per manufacturer's recommendation and analyzed by PCR primers listed in Supplementary Table
706 1 and were run on 1% Agarose gel.
707
708 **Cell Fractionation:**
709
710 Cell fractionation and RNAseA treatment of nuclei was adapted from (Gagnon et al., 2014).
711 Briefly, cells were washed three times with cold 1XPBS and scraped in 1ml of PBS, pelleted down
712 at 2500 rpm for 5 min at 4^0C. Pellet was then resuspended in Hypotonic Lysis Buffer (HLB)(10mM
713 Tris HCl pH 7.5, 10mM NaCl, 3mM MgCl2, 0.3% NP-40, 10% glycerol, 1XPIC), the mix was
714 incubated on ice for 10min and then centrifuged at 800g for 8min at 4^0C. The supernatant was
715 transferred to a new tube and marked as Cytoplasmic fraction. The nuclei pellet after centrifugation
716 was washed twice with HLB and centrifuged at 200g for 2 min at 4^0C. The pellet was then
717 resuspended in 700 ul of ice cold Modified Wuarin-Schiebler buffer (MWS)(10 mM Tris-HCl (pH
718 7.5), 300 mM NaCl, 4 mM EDTA, 1 M urea, 1% NP-40, 1% glycerol,1XPIC) with or without
719 0.1ul of 100mg/ml RNaseA (Qaigen cat no 19101) for 15min on ice briefly tapped few times
720 during incubation time and then centrifuged at 1000g for 5min at 4^0C. The supernatant was
721 collected as a nucleoplasmic fraction. The chromatin pellet was washed twice with MWS buffer
722 and centrifuged at 500g for 3min at 4^0C. The pellet was then dissolved in 500ul of Nuclear Lysis
723 Buffer (NLB) (20mM Tris HCl ph 7.5, 150mM KCl, 3mM MgCl2, 0.3% NP-40, 10%
724 glycerol,1XPIC) and sonicated for 10 cycles with (30 s ON, 30 s OFF) in Bioruptor (Diagenode).
725 SDS loading dye was added and the samples were boiled before immunoblotting.
726

727 **Chromatin associated RNA isolation:**

728

729 Chromatin associated RNA isolation was adapted from (Conrad T et al., 2017). Trypsinized cells
730 were collected in 1.5ml tubes and washed twice with 1xPBS and pellet down at 200xg for 2
731 min. To each pellet was added 400ul of Cell lysis buffer(10mM Tris HCl ph7.4, 150mM NaCl,
732 0.15% Igepal), gently pipetted 3-5times and then incubated for 5min on ice. To the Lo-bind
733 1.5ml tube was added 1ml(2.5 volume) of the cold sucrose buffer(10mM Tris HCl ph7.4,
734 150mM NaCl, 24% Sucrose) and gently overlaid with the cell lysate. Centrifuged at 3500xg
735 for 5min,to collect the nuclear pellet. Isolated nuclei were rinsed with 1ml of ice cold 1xPBS-
736 EDTA solution followed by short spin at 3500xg. The nuclei were resuspended in 250ul of
737 glycerol buffer(20mM Tris HCl ph7.4, 75mM NaCl, 0.5mM EDTA ph8.0, 50% Glycerol and
738 then immediately was added 250ul of Urea buffer(10mM Tris HCl ph7.4, 1M Urea, 0.3M
739 NaCl, 7.5mM MgCl2, 0.2mM EDTA,1%Igepal). Mixed with vortexing for 4s and then
740 incubated on ice for 2min. Centrifuged at 13000xg for 2min to precipitate the Chromatin-RNA
741 complex, the chromatin-RNA pellete was washed with 1xPBS-EDTA once and RNA was
742 isolated using Trizol Reagent. RNA obtained was treated with ezDNase (Ambion) to remove the
743 traces of contaminating DNA. 1 μg of RNA was used for each cDNA synthesis reaction by
744 Superscript IV (Invitrogen) and random hexamers as per manufacturer's recommendation. The
745 CFX96 touch (Biorad) real time PCR was used for qRT-PCRs. qRT-PCRs were performed using
746 three biological replicates for each sample. The fold changes were calculated by the ΔΔCt method
747 and individual expression data was normalized to pre c-myc. The *p*-values were calculated by
748 Student's unpaired two-tailed *t*-test. qRT-PCR primers are listed in Supplementary Table 1.

749 **X-ray induced Senescence:**

750 BJ fibroblast were cultured in (DMEM) Dulbecco's Modified Eagle's medium supplemented
751 with 10% Fetal Bovine Serum and 1% Penicillin/Streptomycin at 37 °C in a humidified
752 incubator containing 5% CO2. Cells were first immortalized by expressing hTERT. Cells were
753 then seeded at around 40% confluency. The day after seeding, a single dose of 10 Gy was
754 delivered using CIX3 cabinet x-ray irradiator (xstrahl Life Sciences) with (195 kV, 10 mA).The
755 irradiated cells were not passaged, but the medium was changed on Days 4, 8,12 and 16 and cells
756 were either harvested on day 10 or day 20 and used for RNA isolation and Chip experiments

757 respectively. Senescence was confirmed by SA-β-gal staining (Cell Signaling technology #9860)
758 and p21 expression (p21 Santacruz Biotechnology sc-397).

759 **siRNA transfection:**

760

761 siRNAs SMARTpools for scrambled (cat no D-001810-10-05) and CTCF siRNA (L-020165-00-
762 0005) were purchased from GE Dharmacon. Transfections were done using Lipofectamine 2000
763 (Invitrogen cat no 11668027).

764

765 **Antibodies:**

766

767 Antibodies used were, CTCF (cell signaling, 3418), , DDX5 (abcam, ab-21696), Cdkn2A p16
768 (abcam, ab-108349), ARF p14 (Santacruz Biotechnology, sc-53639), PolII (Santacruz
769 Biotechnology, sc-55492), GAPDH (Santacruz Biotechnology, sc-32233), Histone H3 (Sigma-
770 Aldrich H0164).

771

772 **TCGA tumor data analysis:**
773
774
775 Cancer specific ATAC-seq peaks for CESC from TCGA cohort were obtained from
776 https://gdc.cancer.gov/. Gene expression for CESC belonging to the TCGA cohort were also
777 obtained from https://gdc.cancer.gov/. TADs and boundaries from HeLa were interrogated in
778 tumor datasets.

779
780
781 **ChIP-seq and Cut and Run analysis:**

782

783 The sequenced reads were aligned to hg19 assembly using default Bowtie2 options. Tag directories
784 were made from the aligned reads using HOMER (http://homer.ucsd.edu/homer/)
785 makeTagDirectory command. A 200bp sliding window was used to identify narrow peaks of
786 transcription factors' ChIP-seq and CTCF Cut and Run which are characteristic of transcription
787 factor peaks. The artifacts from clonal amplification were neglected as only one tag from each
788 unique genomic position was considered (-tbp 1). The threshold was set at a false discovery rate

789 (FDR) of 0.001 determined by peak finding using randomized tag positions in a genome with an
790 effective size of 2 × 10^9 bp. For histone marks, seed regions were initially found using a peak
791 size of 500bp (FDR<0.001) to identify enriched loci. Enriched regions separated by <1kb were
792 merged and considered as blocks of variable lengths. HOMER makeUCSCfile command was used
793 to generate bedgraph files of the read densities across the genome and this track was uploaded to
794 UCSC genome browser. The tag directories generated were used to quantify signal at various
795 regions of interest using annotatePeaks.pl. The datasets used in the study are mentioned in
796 Supplementary Table 3.
797
798 **GRO-seq analysis:**
799
800 The sequenced reads were aligned to hg19 assembly using default Bowtie2 options. Tag directories
801 were made from the aligned reads using HOMER makeTagDirectory command. Tag directories
802 were used to quantify signal and quantify gene expression. HOMER makeUCSCfile command
803 was used to generate strand specific bedgraph files of the read densities across the genome and
804 this track was uploaded to UCSC genome browser.
805
806 **TADs and loop domain calling:**
807
808 The raw paired reads were trimmed and individually mapped to hg19 assembly using bowtie2. The
809 HOMER program makeTagDirectory was first used to create tag directories with tbp 1 option.
810 Data was further processed by HOMER in order to remove small fragment and self-ligations using
811 makeTagDirectory with the following options: -removePEbg -removeSpikes 10000 5. Next,
812 findTADsAndLoops.pl was used to obtain overlapping TADs (TADs and sub-TADs), produced
813 at a 5kb resolution with 10kb windows. This program was also used to generate two bedGraph
814 files that describe the directionality index and the insulation score. tagDir2hicFile.pl was used to
815 generate .hic files. WashU epigenome browser was used to visualize the HiC data. Boundaries of
816 the TADs were defined as 5kb or 20kb regions centred at the end of the TADs called. HICCUPS
817 was used to obtain loops anchors from the earlier obtained .hic file at a 5kb resolution. HOMER
818 annotatePeaks.pl was used to quantify signal from tagDirectories, count peaks, quantify insulation
819 and annotate features at the TAD boundaries. Non-genic TAD boundaries were identified as those

820 that did NOT intersect with an annotated gene body. GRO-seq signal was used to rank the
821 boundaries to then select 10% unique boundaries in each category (high transcription, low
822 transcription, no transcription and random). The .hic file was converted to .cool. The obtained file
823 was then used to plot the averaged insulation score valleys at 5 kb resolution and 100 kb around
824 the boundary using coolpup.py.
825
826 **NET-CAGE and CAGE Analysis:**
827
828 NET-CAGE and CAGE sequenced reads were aligned to hg19 using bowtie2. The HOMER
829 program makeTagDirectory was first used to create tag directories. findPeaks with "–style tss"
830 was used to obtain NET-CAGE TSS. TSS identified within 1kb of each other on the same strand
831 were then merged. The NET-CAGE identified TSS intersecting the refseq annotated TSS was
832 called mRNA TSS and the rest were defined as eRNA TSS. Those TSS which fall in boundaries
833 were identified as bmRNA or beRNAs. 2kb around the identified TSS was used to quantify NET-
834 CAGE and CAGE signal to calculate the log2 fold change between the two.
835
836 **5' Splice site and Transcription Termination Site Sequence Analysis:**
837
838 To identify the 5' SS motif, HOMER denovo motif analysis was performed on the 500bp region
839 after the Refseq annotated TSS with other H3K27ac 500bp regions as background. The top hit was
840 observed to be the splice acceptor site. The "AATAAA" motif was used as the terminator polyA
841 sequence. HOMER scanMotifGenomeWide was used to identify "predicted" splice sites and
842 polyA terminator sites using the earlier identified motifs. The occurrence of the predicted sites was
843 plotted at the earlier identified TSS.
844
845 **Circular Chromatin Conformation Capture-seq:**
846
847 4C was performed as per the protocol described in van de Werken et al., 2012 with minor
848 variations. HeLa cells were fixed with fresh formaldehyde (1.5%) and quenched with glycine
849 (125mM) followed by washes with ice-cold PBS (2x) and scraped, pelleted and stored at -80^{0}C.
850 Lysis buffer [Tris-Cl pH 8.0 (10mM), NaCl (10mM), NP-40 (0.2%), PIC (1x)] was added to the

851 pellets and homogenized by Dounce homogenizer (20 stroked with pestle A followed by pestle B).
852 The 3C digestion was performed with HindIII (400units, NEB) and ligation was performed by the
853 T4 DNA ligase and ligation mix [Triton X-100 (1%), Ligation buffer (1x), BSA (0.105mg/ml),
854 ATP (1.05mM)]. 10x Ligation buffer- Tris-Cl pH 7.5 (500mM), MgCl2 (100mM), DTT (100mM).
855 The ligated samples were then purified by PCI purification and subjected to ethanol precipitation
856 and the pellet was eluted in TE (pH 8.0) to obtain the 3C library. The 4C digestion was performed
857 by DpnII (50units, NEB) and the samples were ligated, purified and precipitated similar to the 3C
858 library to obtain the 4C library. The 4C library was subjected to RNaseA treatment and purified
859 by the QIAquick PCR purification kit. The concentration of the library was then measured by
860 Nanodrop and subjected to PCRs using the oligos for the enhancer viewpoint. The samples were
861 next PCR purified using the same kit and subjected to next-generation sequencing with Illumina
862 HiSeq2500 using 50bp single-end reads. The 4C data is mentioned in Supplementary Table 2.
863
864 **4C method and analysis:**
865
866 The sequenced reads were aligned to hg19 assembly as described in R.4Cker
867 (https://github.com/rr1859/R.4Cker) with Bowtie2. Data analysis was further performed using
868 4C-ker with default parameters. 4Cseqpipe
869 (http://compgenomics.weizmann.ac.il/tanay/?page_id=367) was also used to process the
870 sequenced data. 4C-seq images were generated using truncated mean at a 10kb resolution.
871
872 **Motif analysis:**
873
874 annotatePeaks.pl was used with –hist option with the CTCF motif probability matrix to plot the
875 motif occurrence in the peaks.
876
877 **Promoter capture analysis:**
878
879 Processed CHiCAGO data from Thiecke et al. (https://osf.io/brzuc/) was obtained. BEDTools suite
880 was used to obtain interactions of interest and perform further analysis. The promoter interacting
881 region of the Bedpe files of pcHiC interactions were then intersected with boundaries to obtain

882 promoters that interact with boundaries. The strength of the interactions obtained from the
883 processed CHiCAGO data was plotted for interactions from boundary interacting promoters to
884 enhancers (H3K27ac regions).
885
886 **RNA-seq analysis:**
887
888 The raw reads were mapped to hg19 assembly using hisat2 in a strand specific manner. Tag
889 directories were made from the aligned reads using HOMER makeTagDirectory command.
890 HOMER makeUCSCfile command was used to generate strand specific bedgraph files of the read
891 densities across the genome and this track was uploaded to UCSC genome browser.
892
893
894
895
896
897
898
899
900
901
902
903

904 **References:**

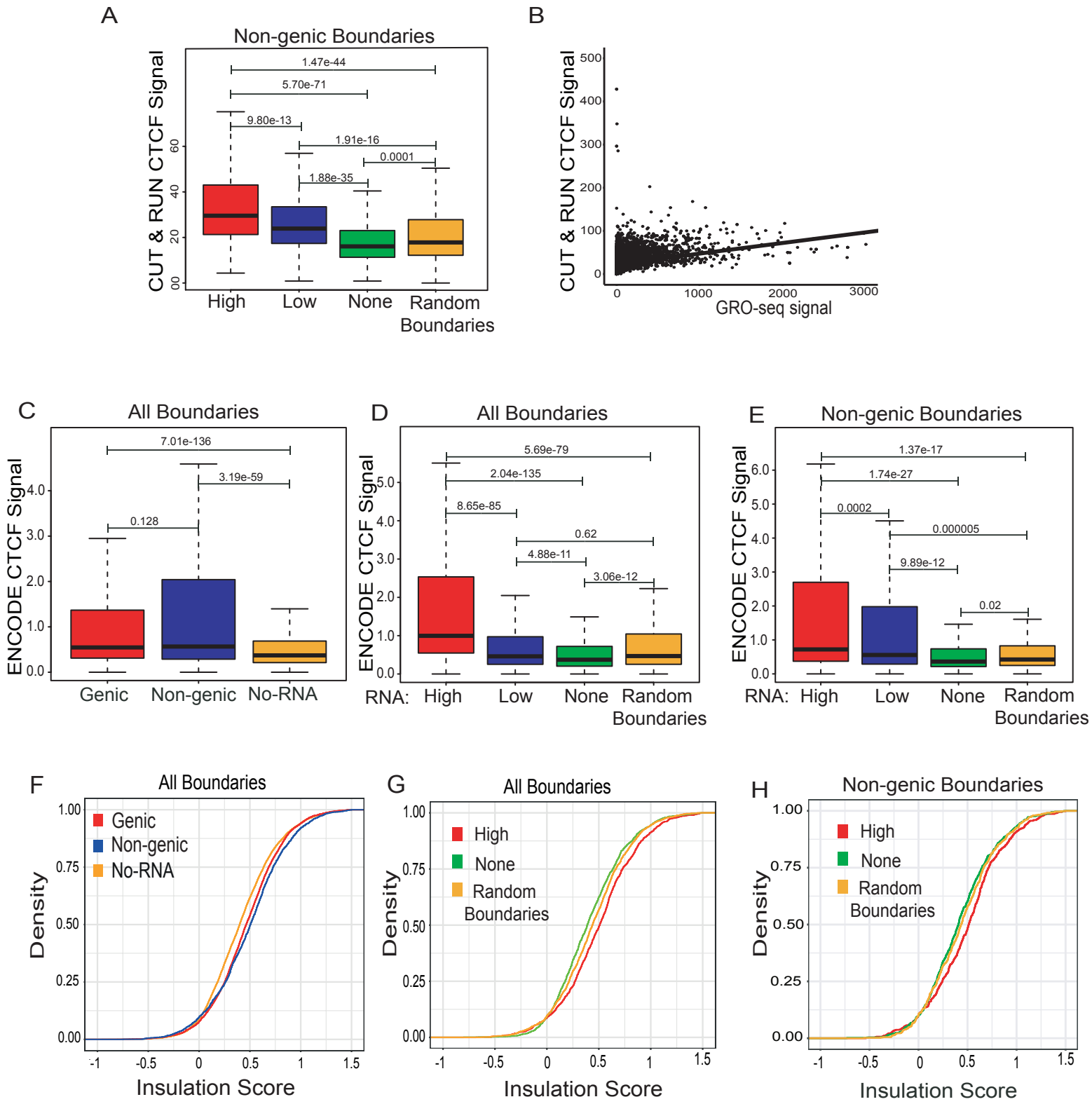
905

906 Andersson, R., Gebhard, C., Miguel-Escalada, I. et al. An atlas of active enhancers across
907 human cell types and tissues. Nature 507, 455–461 (2014).
908 https://doi.org/10.1038/nature12787
909 Arzate-Mejía RG, Josué Cerecedo-Castillo A, Guerrero G, Furlan-Magaril M, Recillas-Targa F.
910 2020. In situ dissection of domain boundaries affect genome topology and gene
911 transcription in Drosophila. *Nat Commun*. doi:10.1038/s41467-020-14651-z
912 Barutcu AR, Blencowe BJ, Rinn JL. 2019. Differential contribution of steady-state RNA and
913 active transcription in chromatin organization . *EMBO Rep*. doi:10.15252/embr.201948068
914 Bonev B, Mendelson Cohen N, Szabo Q, Fritsch L, Papadopoulos GL, Lubling Y, Xu X, Lv X,
915 Hugnot JP, Tanay A, Cavalli G. 2017. Multiscale 3D Genome Rewiring during Mouse
916 Neural Development. *Cell*. doi:10.1016/j.cell.2017.09.043
917 Bouvy-Liivrand M, De Sande AH, Pölönen P, Mehtonen J, Vuorenmaa T, Niskanen H,
918 Sinkkonen L, Kaikkonen MU, Heinäniemi M. 2017. Analysis of primary microRNA loci
919 from nascent transcriptomes reveals regulatory domains governed by chromatin
920 architecture. *Nucleic Acids Res*. doi:10.1093/nar/gkx680
921 Dekker J, Mirny L. 2016. The 3D Genome as Moderator of Chromosomal Communication. *Cell*.
922 doi:10.1016/j.cell.2016.02.007
923 Despang A, Schöpflin R, Franke M, Ali S, Jerković I, Paliou C, Chan WL, Timmermann B,
924 Wittler L, Vingron M, Mundlos S, Ibrahim DM. 2019. Functional dissection of the Sox9–
925 Kcnj2 locus identifies nonessential and instructive roles of TAD architecture. *Nat Genet*.
926 doi:10.1038/s41588-019-0466-z
927 Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B. 2012. Topological
928 domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*.
929 doi:10.1038/nature11082
930 Farooq U, Saravanan B, Islam Z, Walavalkar K, Singh AK, Jayani RS, Meel S, Swaminathan S,
931 Notani D. An interdependent network of functional enhancers regulates transcription and
932 EZH2 loading at the INK4a/ARF locus. *Cell Rep*. doi: 10.1016/j.celrep.2021.108898.
933 Flavahan WA, Drier Y, Liau BB, Gillespie SM, Venteicher AS, Stemmer-Rachamimov AO,
934 Suvà ML, Bernstein BE. 2016. Insulator dysfunction and oncogene activation in IDH
935 mutant gliomas. *Nature*. doi:10.1038/nature16490
936 Fudenberg G, Imakaev M, Lu C, Goloborodko A, Abdennur N, Mirny LA. 2016. Formation of
937 Chromosomal Domains by Loop Extrusion. *Cell Rep*. doi:10.1016/j.celrep.2016.04.085
938 Gagnon KT, Li L, Janowski BA, Corey DR. 2014. Analysis of nuclear RNA interference in
939 human cells by subcellular fractionation and Argonaute loading. *Nat Protoc*.
940 doi:10.1038/nprot.2014.135
941 Giorgetti L, Lajoie BR, Carter AC, Attia M, Zhan Y, Xu J, Chen CJ, Kaplan N, Chang HY,
942 Heard E, Dekker J. 2016. Structural organization of the inactive X chromosome in the
943 mouse. *Nature*. doi:10.1038/nature18589
944 Gong Y, Lazaris C, Sakellaropoulos T, Lozano A, Kambadur P, Ntziachristos P, Aifantis I,
945 Tsirigos A. 2018. Stratification of TAD boundaries reveals preferential insulation of super-
946 enhancers by strong boundaries. *Nat Commun*. doi:10.1038/s41467-018-03017-1
947 Hansen AS. 2020. CTCF as a boundary factor for cohesin-mediated loop extrusion: evidence for
948 a multi-step mechanism. *Nucleus*. doi:10.1080/19491034.2020.1782024
949 Hansen AS, Hsieh THS, Cattoglio C, Pustova I, Saldaña-Meyer R, Reinberg D, Darzacq X, Tjian

950 R. 2019. Distinct Classes of Chromatin Loops Revealed by Deletion of an RNA-Binding
951 Region in CTCF. *Mol Cell*. doi:10.1016/j.molcel.2019.07.039
952 Hanssen LLP, Kassouf MT, Oudelaar AM, Biggs D, Preece C, Downes DJ, Gosden M, Sharpe
953 JA, Sloane-Stanley JA, Hughes JR, Davies B, Higgs DR. 2017. Tissue-specific CTCF-
954 cohesin-mediated chromatin architecture delimits enhancer interactions and function in
955 vivo. *Nat Cell Biol*. doi:10.1038/ncb3573
956 Harismendy O, Notani D, Song X, Rahim NG, Tanasa B, Heintzman N, Ren B, Fu XD, Topol
957 EJ, Rosenfeld MG, Frazer KA. 2011. 9p21 DNA variants associated with coronary artery
958 disease impair interferon-γ 3 signalling response. *Nature*. doi:10.1038/nature09753
959 Hsieh THS, Cattoglio C, Slobodyanyuk E, Hansen AS, Rando OJ, Tjian R, Darzacq X. 2020.
960 Resolving the 3D Landscape of Transcription-Linked Mammalian Chromatin Folding. *Mol*
961 *Cell*. doi:10.1016/j.molcel.2020.03.002
962 Hug CB, Grimaldi AG, Kruse K, Vaquerizas JM. 2017. Chromatin Architecture Emerges during
963 Zygotic Genome Activation Independent of Transcription. *Cell*.
964 doi:10.1016/j.cell.2017.03.024
965 Jayani RS, Singh A, Notani D. 2017. Isolation of nuclear RNA-associated protein
966 complexesMethods in Molecular Biology. doi:10.1007/978-1-4939-6716-2_9
967 Kentepozidou E, Aitken SJ, Feig C, Stefflova K, Ibarra-Soria X, Odom DT, Roller M, Flicek P.
968 2020. Clustered CTCF binding is an evolutionary mechanism to maintain topologically
969 associating domains. *Genome Biol*. doi:10.1186/s13059-019-1894-x
970 Khoury A, Achinger-Kawecka J, Bert SA, Smith GC, French HJ, Luu PL, Peters TJ, Du Q, Parry
971 AJ, Valdes-Mora F, Taberlay PC, Stirzaker C, Statham AL, Clark SJ. 2020. Constitutively
972 bound CTCF sites maintain 3D chromatin architecture and long-range epigenetically
973 regulated domains. *Nat Commun*. doi:10.1038/s41467-019-13753-7
974 Kloetgen A, Thandapani P, Ntziachristos P, Ghebrechristos Y, Nomikou S, Lazaris C, Chen X,
975 Hu H, Bakogianni S, Wang J, Fu Y, Boccalatte F, Zhong H, Paietta E, Trimarchi T, Zhu Y,
976 Van Vlierberghe P, Inghirami GG, Lionnet T, Aifantis I, Tsirigos A. 2020. Three-
977 dimensional chromatin landscapes in T cell acute lymphoblastic leukemia. *Nat Genet*.
978 doi:10.1038/s41588-020-0602-9
979 Krivega I, Dean A. 2017. CTCF fences make good neighbours. *Nat Cell Biol*.
980 doi:10.1038/ncb3584
981 Li W, Notani D, Ma Q, Tanasa B, Nunez E, Chen AY, Merkurjev D, Zhang J, Ohgi K, Song X,
982 Oh S, Kim HS, Glass CK, Rosenfeld MG. 2013. Functional roles of enhancer RNAs for
983 oestrogen-dependent transcriptional activation. *Nature*. doi:10.1038/nature12210
984 Li W, Notani D, Rosenfeld MG. 2016. Enhancers as non-coding RNA transcription units: Recent
985 insights and future perspectives. *Nat Rev Genet*. doi:10.1038/nrg.2016.4
986 Matthews BJ, Waxman DJ. Computational prediction of CTCF/cohesin-based intra-TAD loops
987 that insulate chromatin contacts and gene expression in mouse liver. Elife. 2018 May
988 14;7:e34077. doi: 10.7554/eLife.34077
989 Nakahashi H, Kwon KRK, Resch W, Vian L, Dose M, Stavreva D, Hakim O, Pruett N, Nelson
990 S, Yamane A, Qian J, Dubois W, Welsh S, Phair RD, Pugh BF, Lobanenkov V, Hager GL,
991 Casellas R. 2013. A Genome-wide Map of CTCF Multivalency Redefines the CTCF Code.
992 *Cell Rep*. doi:10.1016/j.celrep.2013.04.024
993 NandyMazumdar M, Yin S, Paranjapye A, Kerschner JL, Swahn H, Ge A, Leir SH, Harris A.
994 2020. Looping of upstream cis-regulatory elements is required for CFTR expression in
995 human airway epithelial cells. *Nucleic Acids Res*. doi:10.1093/nar/gkaa089

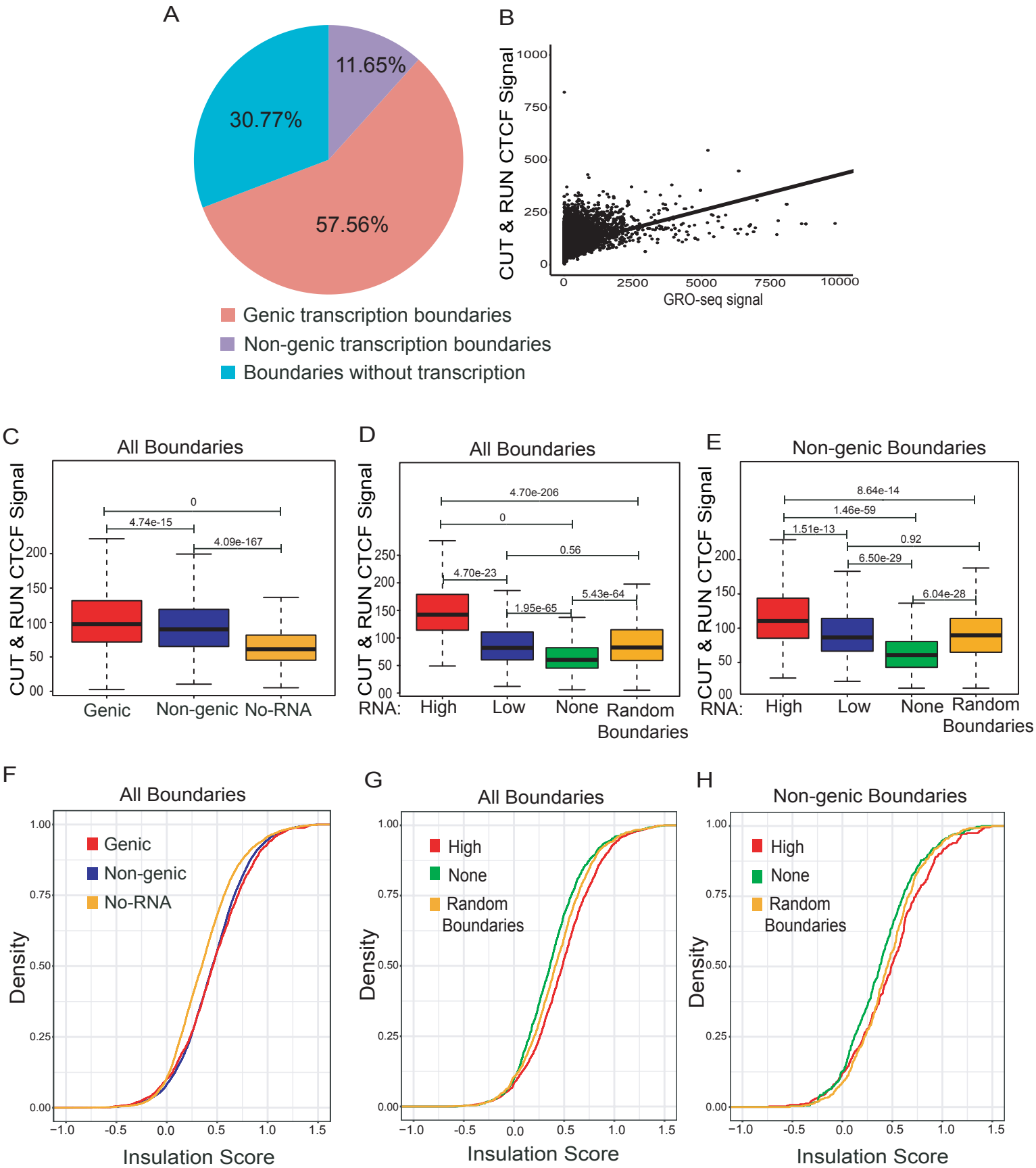
996 Narendra V, Rocha PP, An D, Raviram R, Skok JA, Mazzoni EO, Reinberg D. 2015. CTCF
997 establishes discrete functional chromatin domains at the Hox clusters during differentiation.
998 *Science (80-)*. doi:10.1126/science.1262088
999 Nilson KA, Lawson CK, Mullen NJ, Ball CB, Spector BM, Meier JL, Price DH. 2017. Oxidative
1000 stress rapidly stabilizes promoter-proximal paused Pol II across the human genome. *Nucleic*
1001 *Acids Res*. doi:10.1093/nar/gkx724
1002 Nora EP, Goloborodko A, Valton AL, Gibcus JH, Uebersohn A, Abdennur N, Dekker J, Mirny
1003 LA, Bruneau BG. 2017. Targeted Degradation of CTCF Decouples Local Insulation of
1004 Chromosome Domains from Genomic Compartmentalization. *Cell*.
1005 doi:10.1016/j.cell.2017.05.004
1006 Nuebler J, Fudenberg G, Imakaev M, Abdennur N, Mirny LA. 2018. Chromatin organization by
1007 an interplay of loop extrusion and compartmental segregation. *Proc Natl Acad Sci U S A*.
1008 doi:10.1073/pnas.1717730115
1009 Pugacheva EM, Kubo N, Loukinov D, Tajmul M, Kang S, Kovalchuk AL, Strunnikov A V.,
1010 Zentner GE, Ren B, Lobanenkov V V. 2020. CTCF mediates chromatin looping via N-
1011 terminal domain-dependent cohesin retention. *Proc Natl Acad Sci U S A*.
1012 doi:10.1073/pnas.1911708117
1013 Rahnamoun H, Lee J, Sun Z, Lu H, Ramsey KM, Komives EA, Lauberth SM. 2018. RNAs
1014 interact with BRD4 to promote enhanced chromatin engagement and transcription
1015 activation. *Nat Struct Mol Biol*. doi:10.1038/s41594-018-0102-0
1016 Rao SSP, Huang SC, Glenn St Hilaire B, et al. Cohesin loss eliminates all loop domains. *Cell*.
1017 2017;171(2):305–320.e24.
1018 Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL,
1019 Machol I, Omer AD, Lander ES, Aiden EL. 2014. A 3D map of the human genome at
1020 kilobase resolution reveals principles of chromatin looping. *Cell*.
1021 doi:10.1016/j.cell.2014.11.021
1022 Rowley MJ, Nichols MH, Lyu X, Ando-Kuri M, Rivera ISM, Hermetz K, Wang P, Ruan Y,
1023 Corces VG. 2017. Evolutionarily Conserved Principles Predict 3D Chromatin Organization.
1024 *Mol Cell*. doi:10.1016/j.molcel.2017.07.022
1025 Saldaña-Meyer R, Rodriguez-Hernaez J, Escobar T, Nishana M, Jácome-López K, Nora EP,
1026 Bruneau BG, Tsirigos A, Furlan-Magaril M, Skok J, Reinberg D. 2019. RNA Interactions
1027 Are Essential for CTCF-Mediated Genome Organization. *Mol Cell*.
1028 doi:10.1016/j.molcel.2019.08.015
1029 Schlackow M, Nojima T, Gomes T, Dhir A, Carmo-Fonseca M, Proudfoot NJ. 2017. Distinctive
1030 Patterns of Transcription and RNA Processing for Human lincRNAs. *Mol Cell*.
1031 doi:10.1016/j.molcel.2016.11.029
1032 Schwarzer W, Abdennur N, Goloborodko A, et al. Two independent modes of chromatin
1033 organization revealed by cohesin removal. *Nature*. 2017;551(7678):51–56
1034 Skene PJ, Henikoff S. An efficient targeted nuclease strategy for high-resolution mapping of
1035 DNA binding sites. Elife. 2017 Jan 16;6:e21856. doi: 10.7554/eLife.21856
1036 Sima J, Chakraborty A, Dileep V, Michalski M, Klein KN, Holcomb NP, Turner JL, Paulsen
1037 MT, Rivera-Mulia JC, Trevilla-Garcia C, Bartlett DA, Zhao PA, Washburn BK, Nora EP,
1038 Kraft K, Mundlos S, Bruneau BG, Ljungman M, Fraser P, Ay F, Gilbert DM. 2019.
1039 Identifying cis Elements for Spatiotemporal Control of Mammalian DNA Replication. *Cell*.
1040 doi:10.1016/j.cell.2018.11.036
1041 Sun F, Chronis C, Kronenberg M, Chen XF, Su T, Lay FD, Plath K, Kurdistani SK, Carey MF.

1042 2019. Promoter-Enhancer Communication Occurs Primarily within Insulated
1043 Neighborhoods. *Mol Cell*. doi:10.1016/j.molcel.2018.10.039
1044 Szabo Q, Donjon A, Jerković I, Papadopoulos GL, Cheutin T, Bonev B, Nora EP, Bruneau BG,
1045 Bantignies F, Cavalli G. 2020. Regulation of single-cell genome organization into TADs
1046 and chromatin nanodomains. *Nat Genet*. doi:10.1038/s41588-020-00716-8
1047 Thakur J, Fang H, Llagas T, Disteche CM, Henikoff S. Architectural RNA is required for
1048 heterochromatin organization. doi: https://doi.org/10.1101/784835
1049 Thiecke MJ, Wutz G, Muhar M, Tang W, Bevan S, Malysheva V, Stocsits R, Neumann T, Zuber
1050 J, Fraser P, Schoenfelder S, Peters JM, Spivakov M. 2020. Cohesin-Dependent and -
1051 Independent Mechanisms Mediate Chromosomal Contacts between Promoters and
1052 Enhancers. *Cell Rep*. doi:10.1016/j.celrep.2020.107929
1053 Williamson I, Kane L, Devenney PS, Flyamer IM, Anderson E, Kilanowski F, Hill RE,
1054 Bickmore WA, Lettice LA. 2019. Developmentally regulated Shh expression is robust to
1055 TAD perturbations. *Dev*. doi:10.1242/dev.179523
1056 Wutz G, Várnai C, Nagasaka K, et al. Topologically associating domains and chromatin loops
1057 depend on cohesin and are regulated by CTCF, WAPL, and PDS5 proteins. *Embo J*.
1058 2017;36(24):3573–3599
1059 Yang M, Vesterlund M, Siavelis I, Moura-Castro LH, Castor A, Fioretos T, Jafari R, Lilljebjörn
1060 H, Odom DT, Olsson L, Ravi N, Woodward EL, Harewood L, Lehtiö J, Paulsson K. 2019.
1061 Proteogenomics and Hi-C reveal transcriptional dysregulation in high hyperdiploid
1062 childhood acute lymphoblastic leukemia. *Nat Commun*. doi:10.1038/s41467-019-09469-3
1063 Yokoshi M, Segawa K, Fukaya T. 2020. Visualizing the Role of Boundary Elements in
1064 Enhancer-Promoter Communication. *Mol Cell*. doi:10.1016/j.molcel.2020.02.007
1065 Zhang D, Huang P, Sharma M, Keller CA, Giardine B, Zhang H, Gilgenast TG, Phillips-Cremins
1066 JE, Hardison RC, Blobel GA. 2020. Alteration of genome folding via contact domain
1067 boundary insertion. *Nat Genet*. doi:10.1038/s41588-020-0680-8
1068 Zhang J, Lee D, Dhiman V, Jiang P, Xu Jie, McGillivray P, Yang H, Liu J, Meyerson W, Clarke
1069 D, Gu M, Li S, Lou S, Xu Jinrui, Lochovsky L, Ung M, Ma L, Yu S, Cao Q, Harmanci A,
1070 Yan KK, Sethi A, Gürsoy G, Schoenberg MR, Rozowsky J, Warrell J, Emani P, Yang YT,
1071 Galeev T, Kong X, Liu S, Li X, Krishnan J, Feng Y, Rivera-Mulia JC, Adrian J, Broach JR,
1072 Bolt M, Moran J, Fitzgerald D, Dileep V, Liu T, Mei S, Sasaki T, Trevilla-Garcia C, Wang
1073 S, Wang Y, Zang C, Wang D, Klein RJ, Snyder M, Gilbert DM, Yip K, Cheng C, Yue F,
1074 Liu XS, White KP, Gerstein M. 2020. An integrative ENCODE resource for cancer
1075 genomics. *Nat Commun*. doi:10.1038/s41467-020-14743-w
1076 Zhang Y, Li T, Preissl S, Amaral ML, Grinstein JD, Farah EN, Destici E, Qiu Y, Hu R, Lee AY,
1077 Chee S, Ma K, Ye Z, Zhu Q, Huang H, Fang R, Yu L, Izpisua Belmonte JC, Wu J, Evans
1078 SM, Chi NC, Ren B. 2019. Transcriptionally active HERV-H retrotransposons demarcate
1079 topologically associating domains in human pluripotent stem cells. *Nat Genet*.
1080 doi:10.1038/s41588-019-0479-7
1081 Zirkel A, Nikolic M, Sofiadis K, Mallm JP, Brackley CA, Gothe H, Drechsel O, Becker C,
1082 Altmüller J, Josipovic N, Georgomanolis T, Brant L, Franzen J, Koker M, Gusmao EG,
1083 Costa IG, Ullrich RT, Wagner W, Roukos V, Nürnberg P, Marenduzzo D, Rippe K,
1084 Papantonis A. 2018. HMGB2 Loss upon Senescence Entry Disrupts Genomic Organization
1085 and Induces CTCF Clustering across Cell Types. *Mol Cell*.
1086 doi:10.1016/j.molcel.2018.03.030
1087



Islam and Saravanan et al Sup Fig. 1

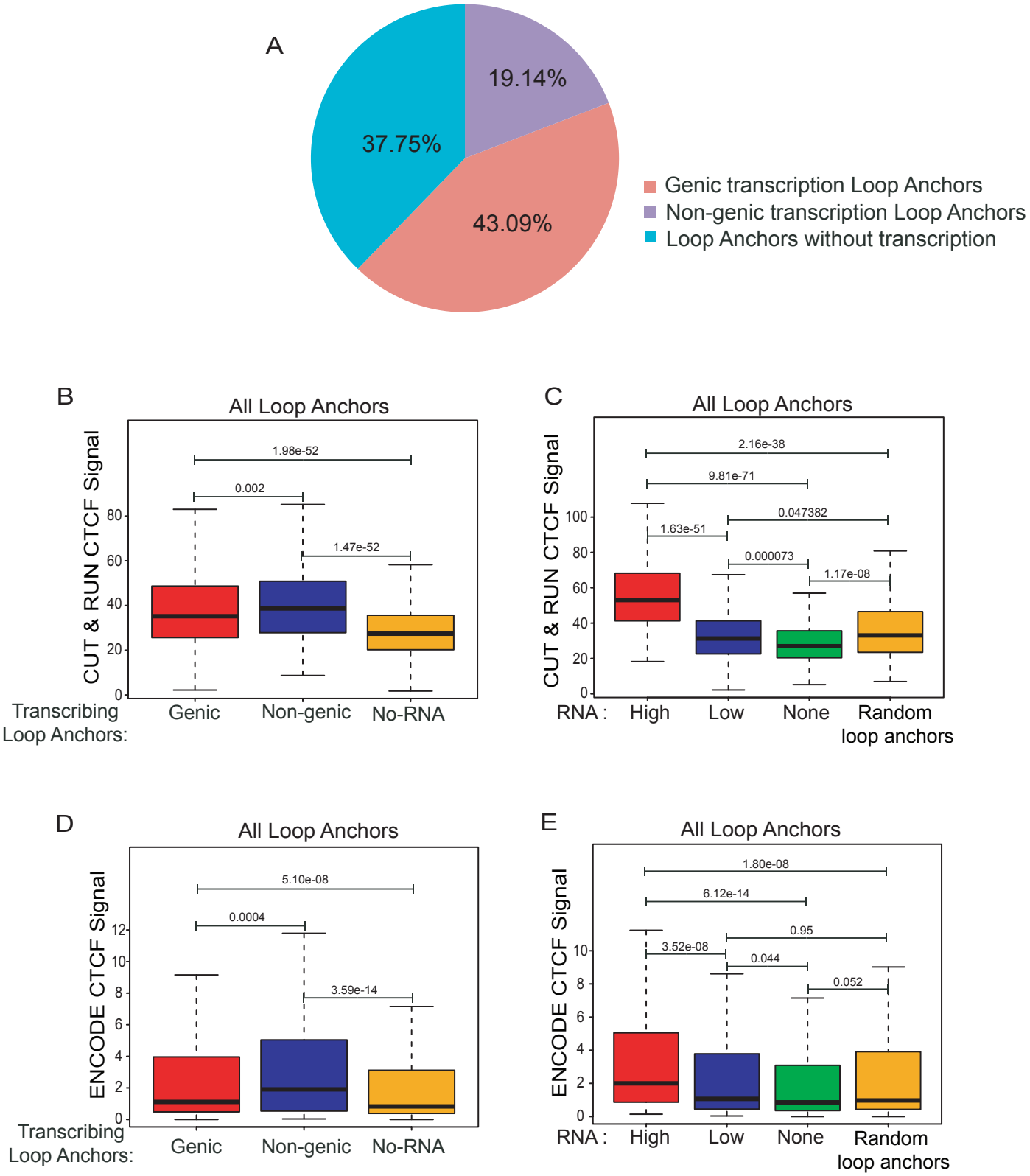
Supp Figure 1. Transcribing boundaries exhibit better insulation: A. CTCF enrichment (CTCF cut & run) on non-genic transcribing boundaries with high, low RNA vs. non-transcribing and random boundaries **B.** levels of RNA and CTCF occupancy at the TAD boundaries have direct positive correlation. **C.** CTCF enrichment (ChIP-seq) on transcribing genic, non-genic and non-transcribing boundaries. **D.** CTCF enrichment on all transcribing boundaries with high, low RNA vs. non-transcribing and random boundaries. **E.** CTCF enrichment on non-genic transcribing boundaries with high, low non-coding RNA vs. non-transcribing and random boundaries. **F.** Density plot showing insulation score of transcribing genic, non-genic and non-transcribing boundaries **G.** Density plot showing insulation score of all transcribing boundaries exhibiting varying levels of RNA vs. non-transcribing and random boundaries **H.** Density plot showing insulation score of non-genic transcribing boundaries exhibiting varying levels of RNA vs. non-transcribing and random boundaries. p-values in boxplots were calculated by Mann Whitney test (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 Sup Fig. 2

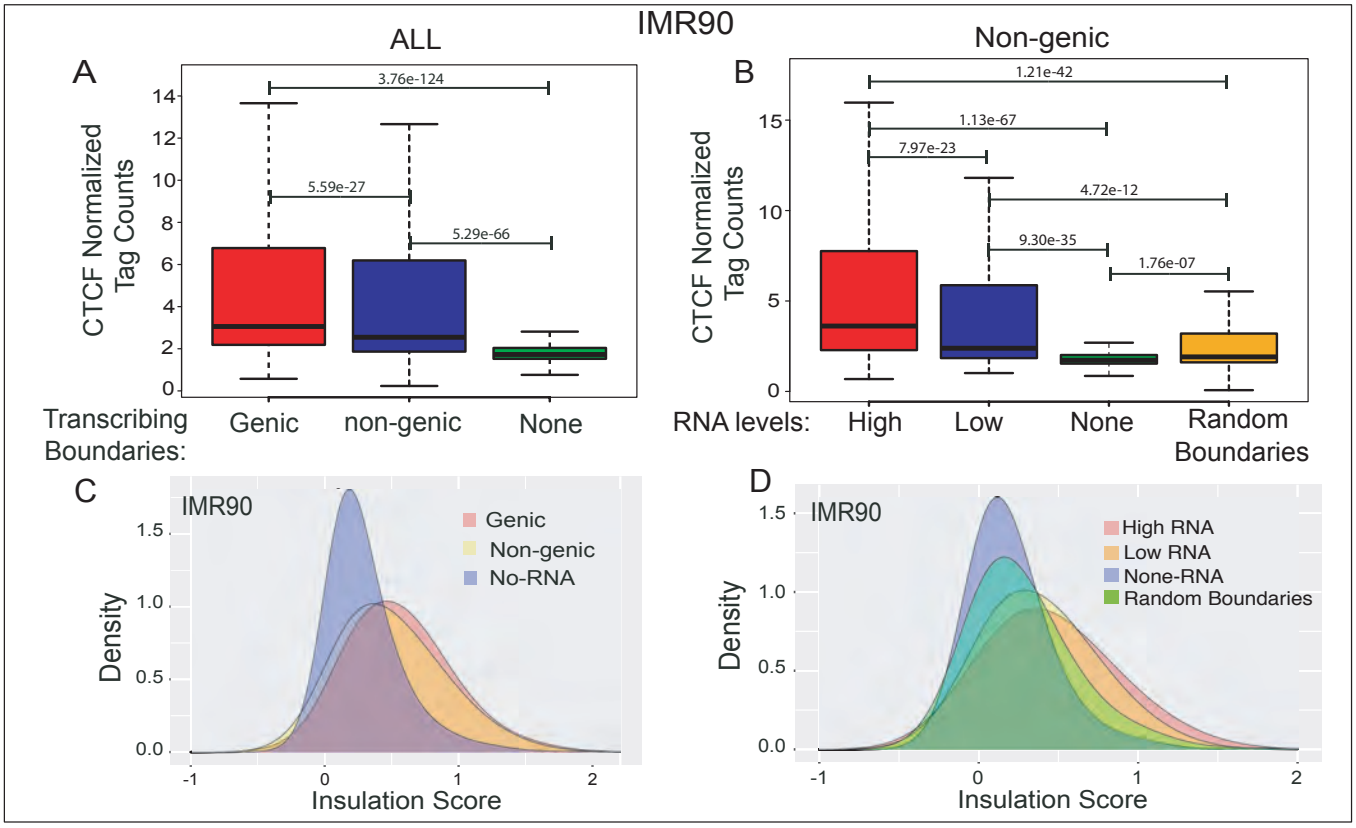
Supp Figure 2. boundaries defined by 20 kb resolution show the similar features as 5kb resolution:

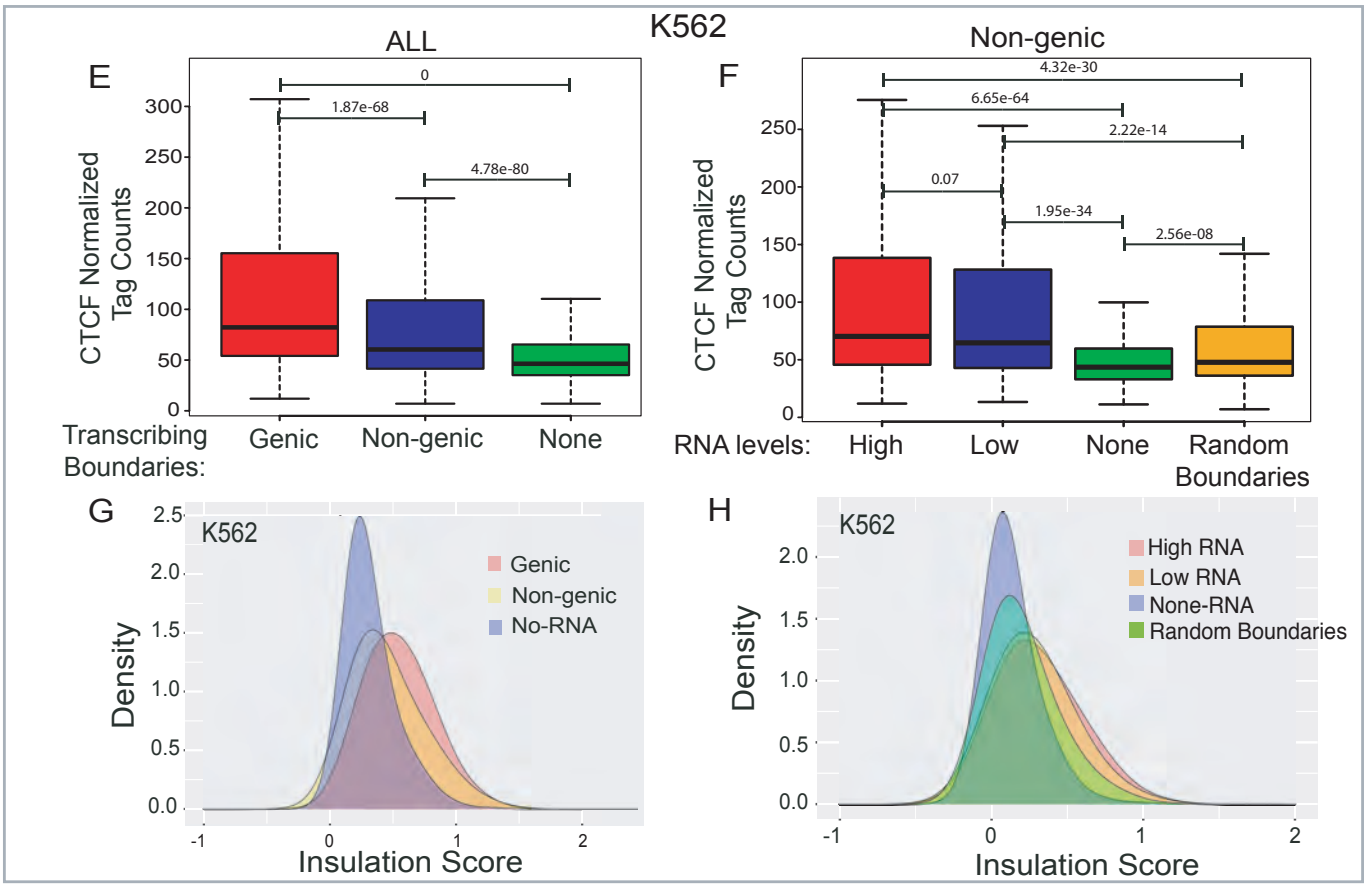
A. Pie diagram showing the percentage of boundaries that show genic, non-genic transcription and not transcribing boundaries. **B.** levels of RNA and CTCF occupancy at the TAD boundaries have direct positive correlation. **C.** CTCF enrichment on transcribing genic, non-genic and non-transcribing boundaries. **D.** CTCF enrichment on all transcribing boundaries with high, low RNA vs. non-transcribing and random boundaries. **E.** CTCF enrichment on non-genic transcribing boundaries with high, low non-coding RNA vs. non-transcribing and random boundaries. **F.** Density plot showing insulation score of transcribing genic, non-genic and non-transcribing boundaries. **G.** Density plot showing insulation score of all transcribing boundaries exhibiting varying levels of RNA vs. non-transcribing and random boundaries. **H** Density plot showing insulation score of non-genic transcribing boundaries exhibiting varying levels of RNA vs. non-transcribing and random boundaries. p-values in boxplots were calculated by Mann Whitney test (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 Sup Fig. 3

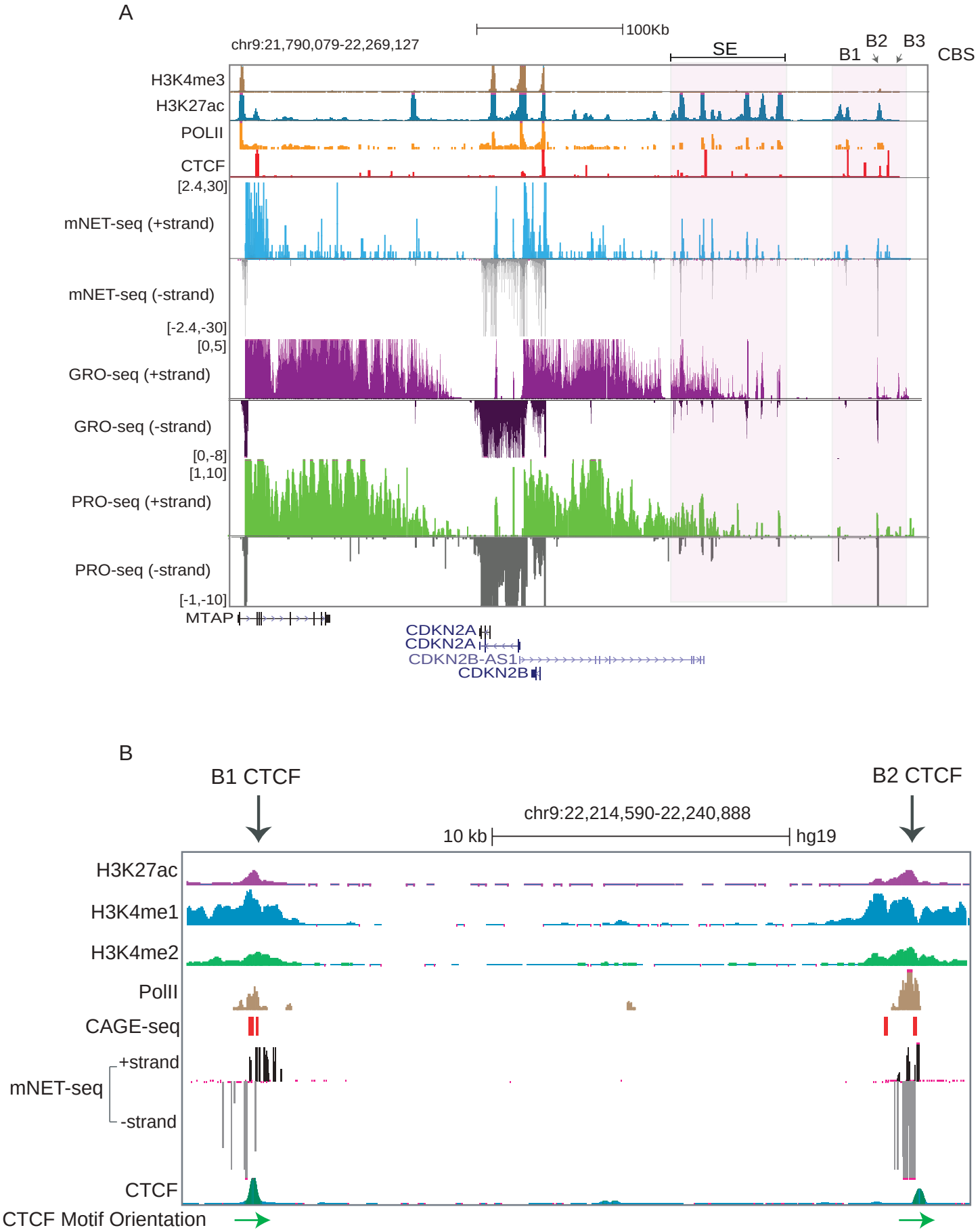
Supp Figure 3. Transcribing Loop Anchors exhibit more CTCF enrichment: A. Pie diagram showing the percentage of loop anchors that overlap with genic transcription, non-genic de novo transcription and without transcription. **B.** CTCF enrichment (CTCF cut & run) on transcribing genic, non-genic and non-transcribing loop anchors. **C.** CTCF enrichment on all transcribing loop anchors with high, low RNA vs. non-transcribing and random loop anchors. **D.** CTCF enrichment (CTCF ChIP-seq) on transcribing genic, non-genic and non-transcribing loop anchors. **E.** CTCF enrichment on all transcribing loop anchors with high, low RNA vs. non-transcribing and random loop anchors. p-values in boxplots were calculated by Mann Whitney test (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.





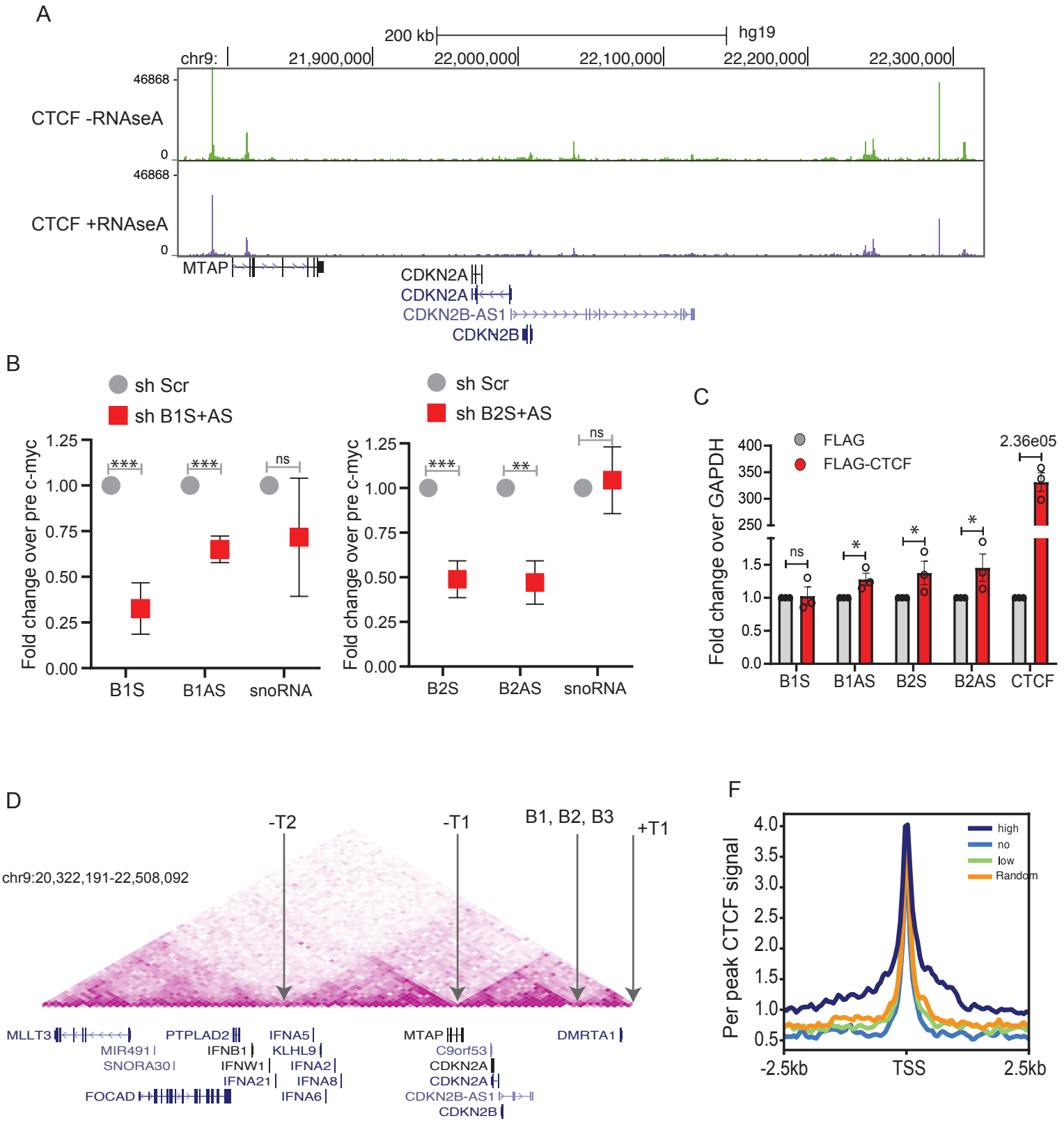
Islam and Saravanan et al Sup Fig. 4

Supp Figure 4. Transcribing boundaries in IMR90 and K562 also exhibit better insulation: A. CTCF enrichment on transcribing genic, non-genic and non-transcribing boundaries in IMR90. **B.** CTCF enrichment on non-genic transcribing boundaries with high, low non-coding RNA vs. non-transcribing and random boundaries in IMR90. **C.** Density plots showing insulation score of transcribing genic, non-genic and non-transcribing boundaries in IMR90**. D.** Density plots showing insulation score of non-genic transcribing boundaries exhibiting varying levels of non-coding RNAs vs. non-transcribing and random boundaries in IMR90. **E-F**. Plots similar to A-B in K562. **G-H.** Plots similar to C-D in K562. p-values in boxplots were calculated by Mann Whitney test (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.



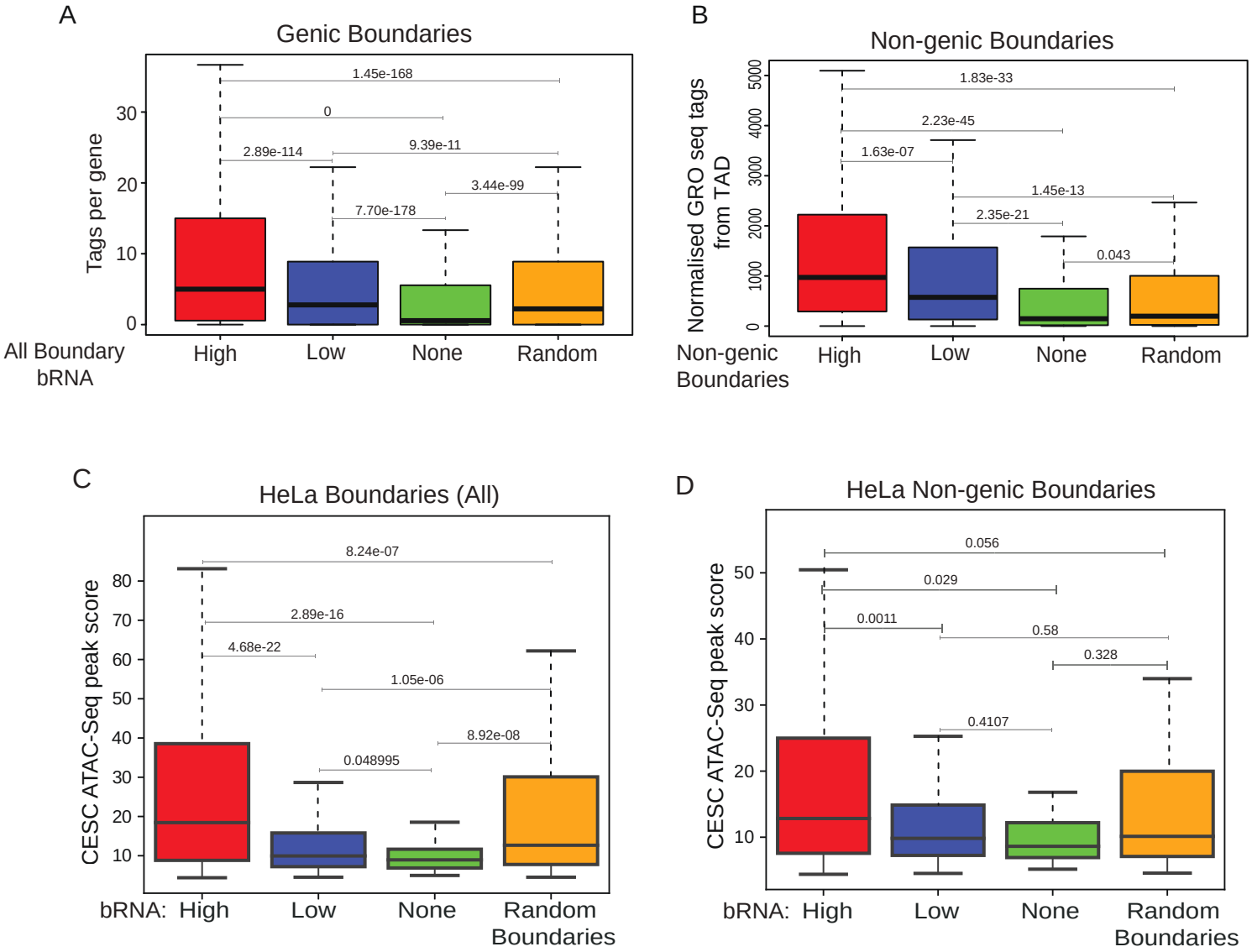
Islam and Saravanan et al. Supp Figure 5

Supp Figure 5. Browser shots of INK4a/ARF loop domain: A. The browser shot showing INK4a/ARF locus on 9p21 region. This is overlaid by the position of, 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 on 3' boundary, and super enhancer. **B**. The browser shot showing the zoomed in region of 3' boundary (B1 and B2 CTCF sites). Which is overlaid by the position of, H3K27ac, H3K4me1, H3K4me2, PolII, CAGE seq track, mNET-seq tracks and CTCF peaks. The Arrow heads at the bottom show CTCF motif orientation at the B1 and B2 CTCF sites .



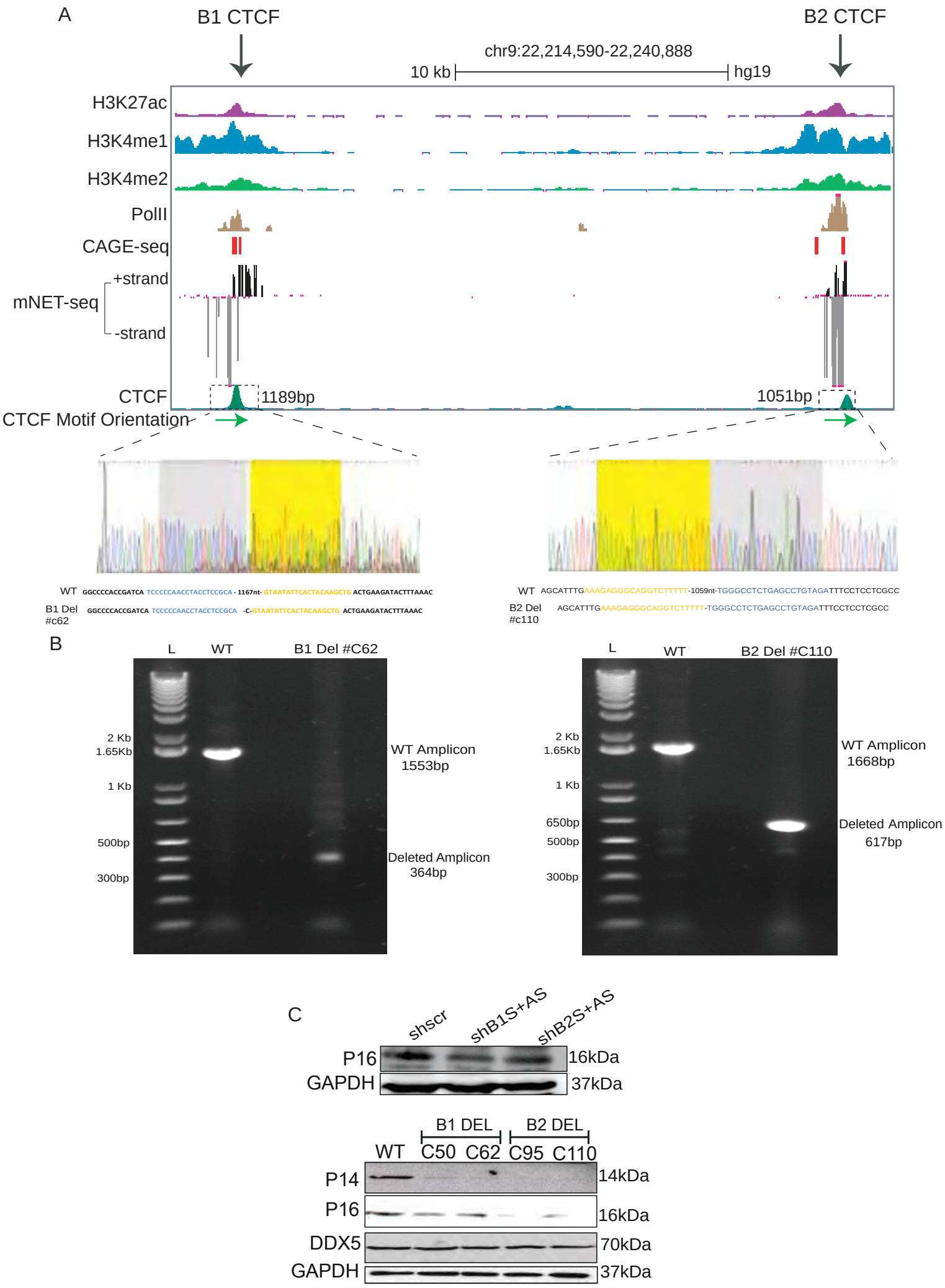
Islam and Saravanan et al. Supp Figure 6

Supp Figure 6. RNA augments the recruitment of CTCF at the TAD boundary: A. The browser shot showing the CTCF cut and run peaks at INK4a/ARF TAD with and without RNaseA treatment. **B.** 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 chromatin associated RNA fraction**. C.** qRT-PCRs showing the fold change in sense and anti-sense RNAs from B1 and B2 CTCF sites upon Flag-CTCF over expression. Last two bars show the increase in CTCF levels. **D.** TADs around INK4a/ARF TAD and position of arrows shows the CTCF sites at upstream and downstream TAD boundaries where CTCF occupancy was interrogated. **E.** Immunoblotting with CTCF and Histone H3 on soluble nucleoplasm and chromatin bound fractions of the nucleus treated with or without RNaseA. **F.** CTCF cut & Run signal at CTCF sites within different categories of boundaries. Error bars denote SEM from three biological replicates. *p*-values were calculated by Student's two-tailed unpaired *t*-test in **B** and **C.** ****$p < 0.0001$,****$p < 0.0001$,*$p < 0.05$, $^{ns}p > 0.05$.



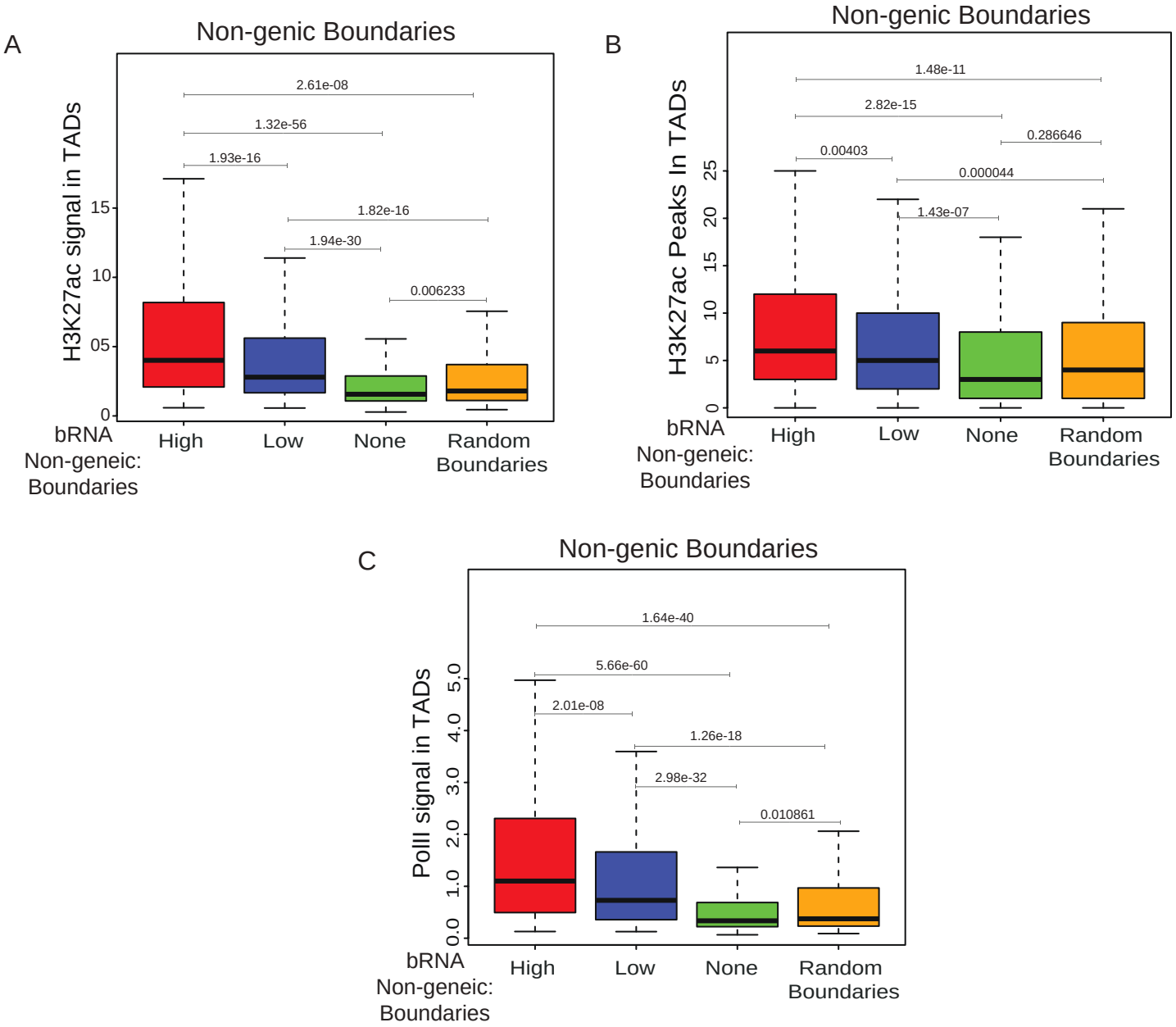
Islam and Saravanan et al. Supp Figure 7

Supp Figure 7. TAD transcription is correlated with transcribing boundaries: A. Box plots showing the normalized GRO-seq reads per gene within the TADs of high, low, non-transcribing and random boundaries. **B.** Box plots showing the normalized GRO-seq tags within the TADs of transcribing non-genic boundaries with high and low bRNAs, non-transcribing and random boundaries. **C.** Box plots showing the ATAC signal at All boundaries exhibiting different levels of bRNAs (From HeLa) vs. non-transcribing and random boundaries in Cervical Squamous Cell Carcinoma tumors from TCGA datasets. **D.** Box plots showing the ATAC signal at the non-genic boundaries exhibiting different levels of bRNAs (From HeLa) vs. non-transcribing and random boundaries in Cervical Squamous Cell Carcinoma tumors from TCGA datasets.



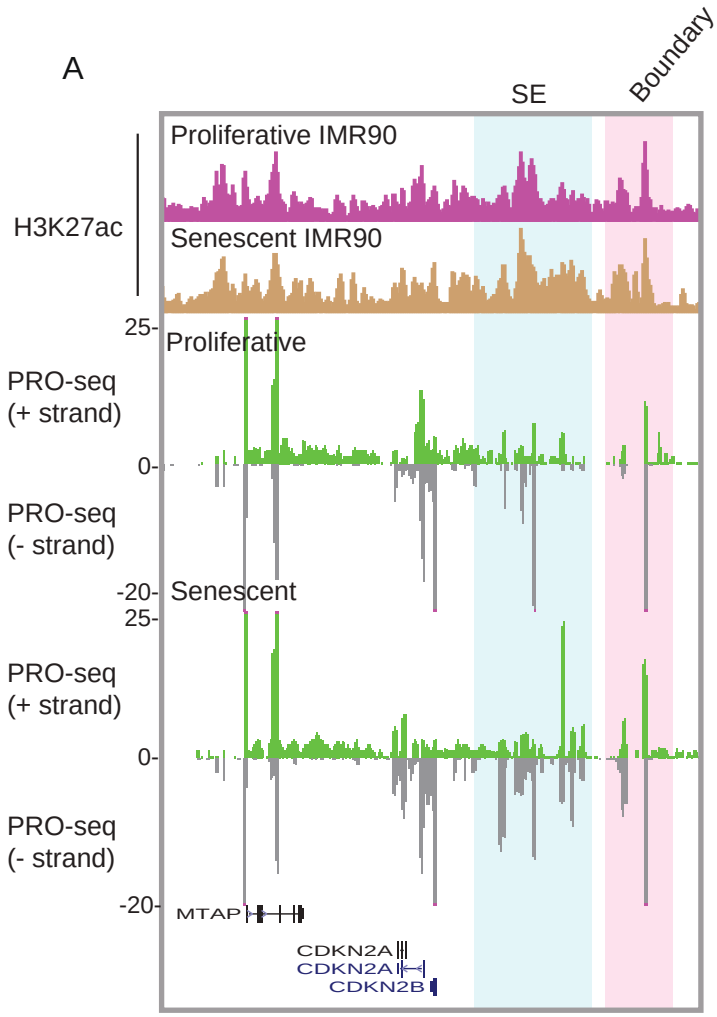
Islam and Saravanan et al Supp Fig. 8

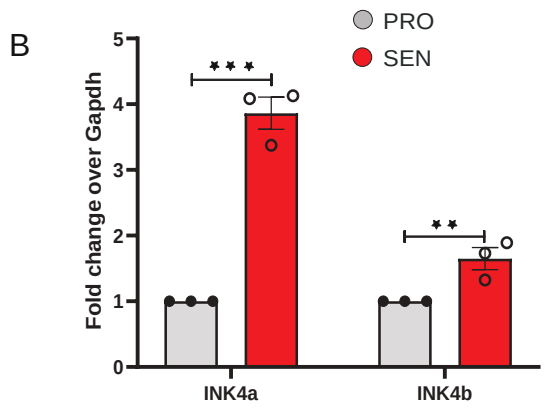
Supp Figure 8. Deletion of B1 and B2 CTCF sites in INK4a/ARF TAD boundary: A. UCSC genome bowser track depicting the deleted B1 and B2 CTCF sites (boxed region at the bottom). The sites are overlaid with various histone marks, CAGE, NET-seq signal, CTCF peaks and CTCF orientation (Green arrows). The tracks zoom in to the chromatograms from sanger sequencing on deleted amplicon to verify the deletion. **B.** Agarose gel images showing the amplicons from surveyor assay from B1 and B2 deletions. **C.** Immunoblot showing the protein levels of p16 in Scr, shRNA 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 sites deletions (Lower panel).

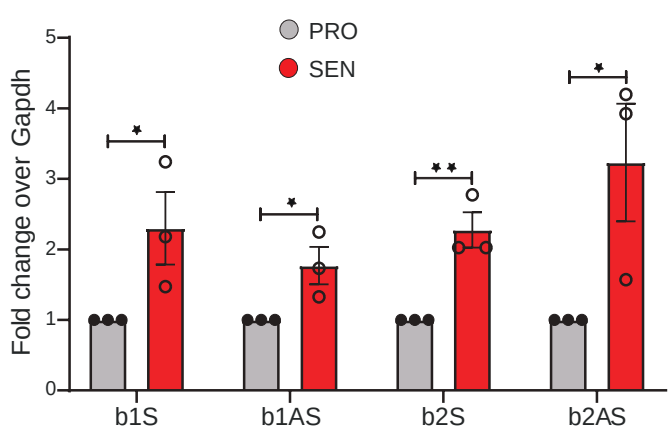


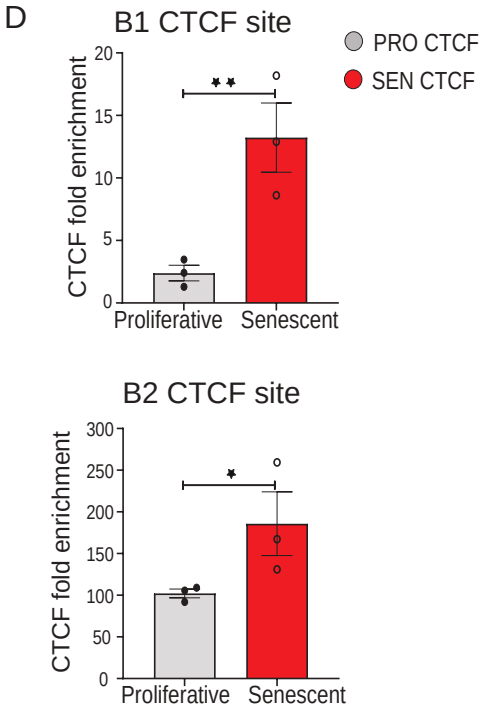
Islam and Saravanan Supp Fig. 9

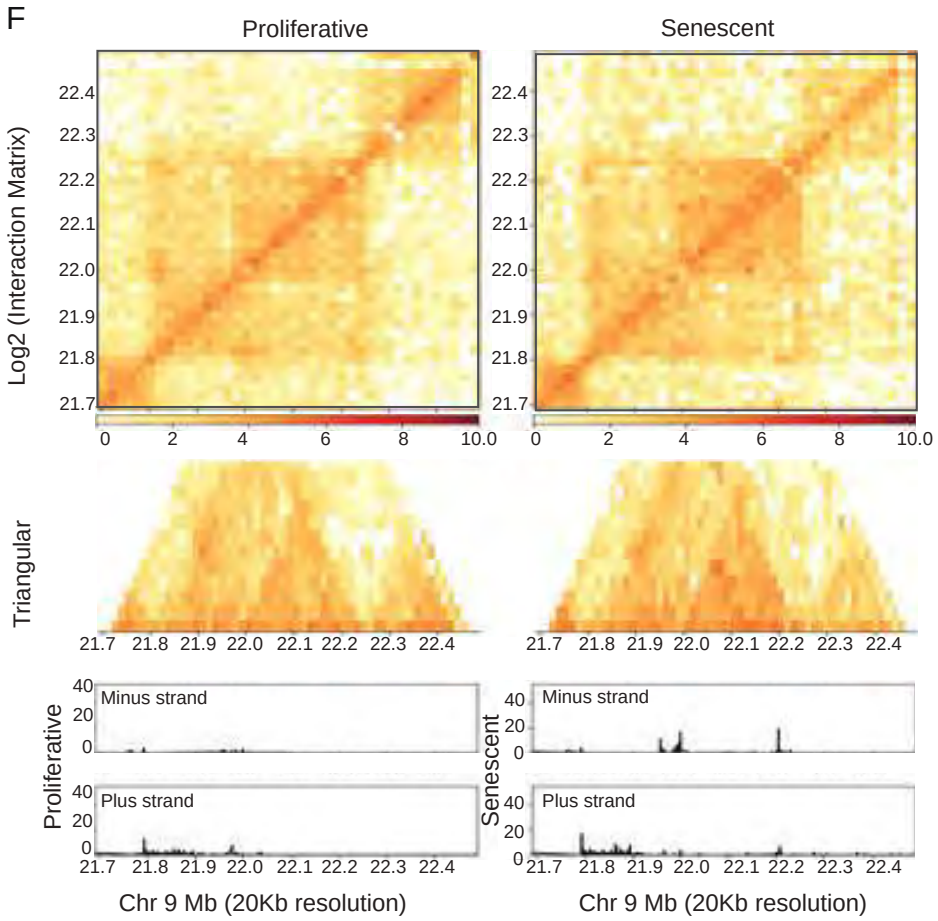
Supp Figure 9. TADs with transcribing boundaries harbour more active enhancers: A. Boxplots displaying enrichment of H3K27ac in TADs with high, low transcribing non-genic HeLA boundaries vs. non-transcribing and random boundaries. **B.** Boxplot showing number of H3K27ac peaks within the TADs with non-genic transcribing boundaries with high and low bRNA vs. non-transcribing and random boundaries. **C.** Levels of PolII in TADs with different non-genic transcribing boundaries vs. non-transcribing and random boundaries.

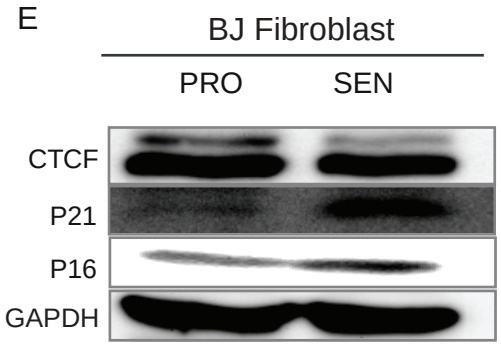












Islam and Saravanan et al. Supp Figure 10

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 super enhancer and boundary region. **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 bRNA 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