Boundary stacking interactions enable cross-TAD enhancer-promoter communication during limb development

Authors: Tzu-Chiao Hung[1], David M. Kingsley[1,2], Alistair Boettiger[1,*]

[1] Department of Developmental Biology, Stanford University School of Medicine, Stanford CA

[2] Howard Hughes Medical Institute, Stanford University School of Medicine, Stanford CA

*correspondence: boettiger@stanford.edu

Abstract

While long-range enhancers and their target promoters are frequently contained within a TAD, many developmentally important genes have their promoter and enhancers within different TADs. Hypotheses about molecular mechanisms enabling such cross-TAD interactions remain to be assessed. To test these hypotheses, we use Optical Reconstruction of Chromatin Architecture (ORCA) to characterize the conformations of the *Pitx1* locus on thousands of single chromosomes in developing mouse limbs. Our data supports a model in which neighboring boundaries are stacked with each other as a result of loop-extrusion, bringing boundary-proximal *cis*-elements into contact. This stacking interaction also explains the appearance of architectural stripes in the population average maps (e.g. Hi-C data). Through molecular dynamics simulations, we further propose that increasing boundary strengths facilitates the formation of the stacked boundary conformation, counter-intuitively facilitating border bypass. This work provides a revised view of the TAD borders' function, both facilitating as well as preventing *cis*-regulatory interactions, and introduces a framework to distinguish border-crossing from border-respecting enhancer-promoter pairs.

25 **Introduction**

26

27 The mammalian genome is partitioned in its 3D organization into topologically associating

28 domains (TADs), regions in which intra-domain interactions are substantially more common than

29 interdomain interactions [1–4]. Known long range *cis*-regulatory interactions, most notably between

30 enhancers and their target promoters, frequently lie within the same TAD, and in a multitude of

31 cases mutations which disrupt this TAD organization result in corresponding disruptions to normal

32 enhancer promoter contact and changes in gene expression [5–9].

33

34 Nonetheless, several developmentally important enhancers are separated by TAD borders and

35 CTCF bound insulators from the genes they control. These include genes with key roles in axial

36 patterning [10], hematopoiesis [11] and limb patterning [12,13]. As candidate enhancer-promoter (E-P)

37 interactions are frequently hypothesized and tested based on linear proximity to a gene, guided

38 by previously mapped TAD boundaries, and studied in cell culture, it is likely that the genome-

39 wide frequency of TAD border-bypassing E-P pairs is underestimated. The known examples, just

40 mentioned, are among some of the longer-range acting enhancers and have preferentially been

41 identified by low-throughput phenotype-driven studies in developmental contexts where cell

42 culture models are often lacking. Being able to distinguish these border-crossing from border-

43 respecting enhancers is an essential feature of understanding *cis*-regulation. A notable case of

44 this border bypass occurs in the regulation of the gene *Pitx1,* where one of its major enhancers,

45 *Pen*, is located no less than three TADs away (**Fig. 1a, b**).

46

47 Several properties make *Pitx1* ideal for studying border bypass. Whereas enhancer

48 redundancy/shadow enhancers can often mask the regulatory contributions of individual

49 enhancers and functional linking between enhancers and promoters [14,15], the *Pitx1* distal

50 enhancer, called *Pen*, has obvious effects on both gene expression and morphological

51   phenotypes. Indeed, studies in human patients first implicated the region around *Pen* in regulation

52   of limb specific expression of *Pitx1* [16,17]. Subsequent work with mouse genetic models found

53   deletion of *Pen* caused a 45% reduction in *Pitx1* hindlimb expression, and increased incidence of

54   clubfoot, consistent with reduced *Pitx1* function [18]. In addition to the multi-TAD crossing *Pen*, two

55   other hind-limb specific enhancers of *Pitx1* have also been characterized, which provide intriguing

56   comparison cases. 5' of *Pitx1* lies the RA3/PDE enhancer, which crosses no TAD borders in

57   interacting with *Pitx1* [19]. Meanwhile 3' of *Pitx1*, in the neighboring TAD, lies the *PelB* enhancer in

58   the intron of a neighboring unrelated gene [20]. This enhancer is conserved through fish, but must

59   navigate one border to interact with *Pitx1*.

60

61   Another attractive feature of *Pitx1-Pen* border bypass is its tissue-specificity. Though *Pitx1* is

62   primarily expressed in hindlimb and not forelimb, the isolated *Pen* enhancer drives expression in

63   both tissues when cloned and tested in a small transgenic reporter construct [18]. Furthermore, the

64   *Pen* enhancer in the native *Pitx1* locus carries the H3K27ac mark associated with active

65   enhancers in both hindlimbs and forelimbs (**Fig. 1b** [21]). However, recent capture-Hi-C (cHi-C)

66   experiments from mouse limb revealed quantifiably higher interaction frequency between *Pen*

67   and *Pitx1* in the hindlimb compared to forelimb tissues [18]. Notably, this work also showed that this

68   difference was largely reduced when genetic aberrations removed a TAD boundary between *Pen*

69   and *Pitx1* and brought them closer in linear distance, either through deletion or inversion, which

70   led to increased *Pen-Pitx1* interaction, ectopic *Pitx1* expression in the developing forelimb, and

71   corresponding forelimb to hindlimb developmental transformations [18]. Thus, the hindlimb-specific

72   interaction topology appears to account for the normal cell-type specific gene expression, but how

73   this contact is achieved across multiple TAD boundaries remains a mystery. To understand the

74   requirements for border bypass, here we compare how the same *Pitx1* chromosomal domain

75   differs in its 3D conformations between forelimb and hindlimb, where border bypass is prohibited

76   in the former but enabled in the latter.

77

78   We employ Optical Reconstruction of Chromatin Architecture [22,23] to visualize the 3D organization

79   of the *Pitx1* chromosomal domain at single-chromatin resolution in developing limbs. This

80   microscopy based approach, similar to other recent studies [24–28], allows the 3D chromatin

81   trajectory to be visualized from individual cells, and thus has the potential to reveal what folding

82   behavior enables hindlimb cells to achieve cross-TAD border enhancer-promoter interaction. By

83   comparing 3D conformation of the *Pitx1* domain in the otherwise similar tissues of hindlimb and

84   forelimb, we aimed to uncover the molecular mechanisms which enable functional bypass of TAD

85   borders.

86

87   **Results**

88   To visualize the *Pitx1* chromosomal domain, we designed Oligopaint probes tiling a 750 kb region

89   around *Pitx1*, encompassing the downstream *PelB* enhancer, upstream enhancers RA3/PDE and

90   *Pen*, and an unrelated gene *Neurog1* in the distal polycomb-silenced region (**Fig. 1a**). These

91   probes span the 2 CTCF marked TAD borders, which we term B1 and B2, respectively, between

92   *Pitx1* and *Pen* (**Fig. 1a, b**). We hybridized these probes to developing forelimbs and hindlimbs

93   from E11.5 - E12.5 C57BL/6J mouse embryos, and visualized nearly 100,000 chromosome traces

94   at 10 kb resolution, with each trace containing 75 3D coordinates, across 5 individual experiments

95   (**Supplementary Table 1**).

96

97   The pairwise contact frequency (where two positions were found in proximity i.e., within 200 nm

98   of one another) agreed closely with the corresponding frequency at which ligated reads were

99   detected in previous work using cHi-C (**Fig. 1d**), and revealed a similar pattern of TADs, loops

100   and stripes (**Fig. 1a** and **Extended Data Fig. 1a**). Comparing the contact frequency maps from

101   hindlimb and forelimb, we see that the associations of *Pitx1* with *PelB*, *RA3*, *RA4*, and *Pen* are all

102   more frequent in the hindlimb, where these enhancers contribute to *Pitx1* expression [18–20] (**Fig.**

103   **1c,f**), consistent with previous reports [18]. These differences are also apparent when we examine

104   the absolute distances instead of contact frequencies (**Extended Data Fig. 1d-f**). In contrast to

105   these enhancer-promoter interactions, the interactions between *Pitx1* and the gene *Neurog1* are

106   more frequent in forelimb, also consistent with previous cHi-C measurements [18]. Although yet to

107   be tested, this gene-gene interaction is likely mediated by Polycomb dependent long range

108   looping [29,30], as both genes are enriched in Polycomb-associated H3K27me3 in forelimb, whereas

109   in hindlimb, only *Neurog1* is H3K27me3 rich (**Fig. 1b**). The agreement between the two technically

110   distinct assays testifies to the accuracy of the data. Additionally, the single-molecular nature of

111   ORCA data allows us to robustly test the statistical significance of the difference across thousands

112   of measurements (**Fig. 1f** and **Extended Data Fig. 1e**).

113

114   Surprisingly, while several cross-TAD interactions are more frequent in hindlimb, the TAD borders

115   are also stronger in hindlimb compared to forelimb. For example, while *RA4* shows a higher

116   frequency of proximity to *Pitx1* in hindlimb cells compared to forelimb cells, the TAD boundary B1

117   (next to RA3) is stronger, and the rest of the TAD containing RA4 (though not RA4 itself) is on

118   average farther away (**Extended Data Fig. 1d**) and experiences fewer contacts with the TAD

119   containing *Pitx1* (**Fig. 1c**, upper dashed box and middle circle). Similarly, the TAD containing the

120   enhancer, *Pen*, is also on average farther away from and better insulated from the TAD containing

121   *Pitx1*, even though *Pen* and *Pitx1* are on average closer (**Extended Data Fig. 1d**) and in more

122   frequent contact (**Fig. 1c**, lower dashed box and circle). Since the measurements are made

123   without the need for matrix balancing to address amplification biases, are imaged on the same

124   slide, and replicated across independent experiments (**Extended Data Fig. 2**), we interpret these

125   data to reflect true changes in the absolute interaction and not artifacts of normalization. We thus

126   have a paradox: stronger TAD borders, yet more frequent enhancer-promoter contacts across

127   borders.

128

129   We hypothesized this could occur through three plausible models (**Fig. 2a**). In the first model,

130   "Merge", the TADs merge in a subset of chromatins, for example due to the stochastic release of

131   boundary factors. This would allow contact between enhancer and promoter in this subset, even

132   though at the population level the boundaries would still be detected due to the contribution of

133   other chromatins. In the second model, "Stack", the TADs generally remain intact, but their

134   boundaries stack around a central hub at which the boundary-proximal enhancers and promoters

135   meet. Finally, it is possible that enhancer and promoter interact without the TADs merging or

136   stacking, which we call "Loop out". For example, the promoter and enhancer could be decompact

137   enough to leave their respective TADs and meet outside the domain. These three models, as

138   they stipulate hypotheses about how one structural aspect influences another in the very same

139   molecules, cannot be resolved by Hi-C or other bulk methods that are restricted to measuring

140   average pairwise interactions. ORCA, however, with its single-chromatin resolution, is well-suited

141   for analyzing these models.

142

143   To compare the three models against the many individual molecules visualized with ORCA, we

144   define criteria to determine if each chromatin trace fits Merge, Stack, or Loop Out configuration

145   (**Fig. 2b** and **Extended Data Fig. 3**). We then analyze whether the chromatin traces fitting a

146   particular model are more likely to have *Pitx1-Pen* contact, and whether such traces are more

147   abundant in hindlimb versus forelimb cells. We define "Merge" chromatin traces as those having

148   an insulation score of less than 1.1 (<10% difference in intra- vs. inter-domain median pairwise

149   distance, after correction for linear distance) for both TAD boundaries between *Pitx1* and the *Pen*

150   enhancer (B1 and B2 (**Fig. 1a**)), "Stack" traces as those with *Pitx1* contacting either B1 or B2,

151   and "Loop out" traces as those not belonging to either "Merge" or "Stack". For each model, there

152   are numerous chromatin molecules consistent with its configuration in both forelimb and hindlimb

153   (**Fig. 2c**). Notably, "Stack" is the only configuration that is more abundant in hindlimb than in

154   forelimb. Moreover, among molecules with *Pitx1* contacting *Pen*, about 55% can be accounted

155   for by "Stack", highest among the three models (**Fig. 2d**). In addition, the difference in fraction of

156   *Pitx1-Pen* contacting molecules between forelimb and hindlimb is most pronounced among "Stack"

157   molecules. When assessing the relative risk, an indicator of the probability of finding *Pitx1-Pen*

158   contact given one of the three configurations, "Stack" again is the only one with a value > 1,

159   indicating that chromatin traces in the "Stack" configuration are (in this case, > 2x) more likely to

160   have *Pitx1* contacting *Pen* compared to molecules not in "Stack" configuration (**Fig. 2e**). These

161   analyses show that "Stack" is the most plausible mechanism for hindlimb-enriched interaction

162   between *Pitx1* and *Pen*. Noticeably, despite nearly all hindlimb cells expressing *Pitx1*, only ~10%

163   of them exhibit a "Stack" configuration where *Pen* and *Pitx1* are in contact (**Fig. 2d**), suggesting

164   that the hub organization in "Stack", similar to the formation of TADs, is a dynamic conformation

165   instead of a static structure.

166

167   Thus, the data indicate that while all three mechanisms, "Loop out", "Merge", and "Stack",

168   contribute to the overall contact frequency observed in hindlimb, "Stack" is the primary

169   conformation the chromosomal domain takes for *Pitx1* to bypass multiple TAD borders and make

170   preferential contact with its distal enhancer, *Pen*. Such preferential contact was shown to be

171   required for hindlimb specific up-regulation of *Pitx1* and avoidance of ectopic forelimb expression

172   [18]. Transient boundary loss or looping out across the domain occurs, but they do not enhance the

173   probability of E-P contact, as their relative risks < 1, nor show much difference in frequency

174   between hindlimb and forelimb.

175

176   The "Stack" model also helps explain another feature commonly observed in both Hi-C and ORCA

177   data: long stripes in contact frequency maps. A stripe indicates that a genomic position has higher

178   contact frequencies with almost all positions in the neighboring genomic region than other

179   positions in that region have with each other. In **Fig. 1a**, for example, one of the stripes emanates

180   from *Pitx1* and extends through almost the entire probed chromosomal domain. Hi-C analyses

181   have found similar stripe patterns throughout the genome, and in many cases the stripes also

182   connect enhancers and promoters, prompting new efforts to understand their molecular origins

183   and new models of enhancer - promoter communication [31–34]. Based on Hi-C analyses and

184   polymer modeling, it has been proposed that stripes arise from the dynamic activity of cohesin

185   molecules that load at an enhancer or promoter which is adjacent to a CTCF-site and reel-in the

186   cognate enhancer/promoter through loop extrusion in the direction not blocked by CTCF [31]. By

187   sliding the loading site across the intervening DNA in this fashion, a stripe arises in the population

188   map (**Fig 3a**). The "Stack" model provides an alternative explanation for the stripes. With elements

189   next to CTCF sites stacked together at a central hub, these elements are simultaneously closer

190   to all other elements in the domain (out in the intervening loops) than those other elements are to

191   one another (**Extended Data Fig. 4**). The centrally located boundary elements are thus more

192   likely to contact any element out in the loops through stochastic fluctuations of the polymer

193   (**Extended Data Fig. 4c**).

194

195   While both scenarios result in stripes at the populational level, they predict different behaviors at

196   the single-molecular level. The "reeling-in" model predicts that for most molecules, *Pitx1* contacts

197   no more than 1 position along the stripe: the stripe arises from capturing different chromatins in

198   each step along the scan. Therefore, when examining chromatin traces where *Pitx1* contacts, for

199   example, B1, *Pitx1* should not also be contacting the region between B2 and *Pen*, and the stripe

200   should not be observed in each individual molecule. However, as seen in **Fig. 3c**, our single-

201   molecule data shows that stripe is not only present, but stronger for molecules where *Pitx1* is

202   interacting with either B1 or B2. The "Stack" model explains this phenomenon well: By

203   conditioning on *Pitx1*-intermediate boundary interaction, we enrich for molecules with hub

204   formation, a configuration where the boundaries are closer to all intervening regions. It is worth

205   clarifying that we are not claiming "reeling-in" does not occur. To the contrary, it is likely through

206   "reeling-in" that multiple domain boundaries stack together to form the "Stack" organization.

207   However, examining individual molecules suggests that stripes at the populational level are not

208   merely an aggregate of single loops. Instead, "Stack" hub organization is seen in many

209   contributing molecules, suggesting that it is indispensable for interpreting the underlying dynamics

210   of these stripes. We observed similar aggregates of loop hubs in individual traces producing

211   stripes in the population maps for the *SOX9* locus in human cranial neural crest cells, though in

212   that case the stripes did not extend beyond TAD boundaries [35].

213

214   We next asked, what molecular mechanisms facilitate the formation of the "Stack" conformation,

215   and what molecular conditions differ between hindlimb and forelimb to bring this about in a tissue-

216   specific manner? To answer these mechanistic questions, we turned to physical polymer

217   modeling. Using the polymer modeling framework from open2c [36], derived from prior work

218   modeling loop-extrusion dynamics in chromatin [37,38], we simulate the dynamics of a chromosomal

219   domain with 4 extrusion blockers, representing *Pitx1* (promoter), B1, B2 and *Pen* (enhancer),

220   respectively, each of which correspond to TAD-borders marked by CTCF (**Fig. 1a**). We use these

221   simulations to explore interactions among sequentially arrayed border elements under distinct

222   regimes of loop extrusion.

223

224   In our simulation, increasing CTCF occupancy by modeling not only moderately strengthened the

225   TAD boundaries as expected, it also strengthened the cross-TAD contact frequency between the

226   *Pen* enhancer and the *Pitx1* promoter (**Fig. 4e**). Examination of polymer structures with E-P

227   contact showed frequent involvement of the intervening boundaries B1 and B2, consistent with

228   the "Stack" organization seen for the *Pitx1* domain. Plotting the positions of the cohesin molecules

229   reveals the presence of multi-cohesin bridges connecting the promoter to B1 to B2 to the

230   enhancer (**Fig. 4a**). Although not all E-P contacts involved the intermediate boundary interactions,

231   indicating that sometimes "Merge" or "Loop out" rather than "Stack" allows contact, "Stack"

232   configuration accounts for most of the increase in contact upon increased CTCF occupancy (**Fig.**

233   **4f,g**). Similarly, the relative risk for observing E-P contact given interaction with intermediate

234   boundaries is significantly greater than 1, whereas the relative risk given either "Merge" or "Loop

235   out" configurations is not (**Fig. 4h**).

236

237   To explore this effect more generally, we do not insist on matching the specific organization of

238   the *Pitx1* region with respect to the distances between border elements, their relative orientation

239   or strengths. However, additional simulations, in which we manually tuned these parameters to

240   give a more *Pitx1*-like structure to that observed experimentally (Pearson's R=0.93), recapitulated

241   the effect of boundary strengthening in increasing boundary bypass (**Extended Data Fig. 5**).

242

243   As a test of this hypothesis, we examined the binding of CTCF across the *Pitx1* domain in hindlimb

244   vs. forelimb in published ChIP-seq data [21]. Interestingly, CTCF peak heights were higher at *Pitx1*,

245   B1, B2 and B3 in hindlimb than forelimb, even though background binding levels between the

246   peaks were similar (**Fig. 4i**). This suggests that differences in CTCF binding to peak regions could

247   explain the distinct structural organization of the *Pitx1* domain in hindlimbs, which allows sufficient

248   interaction for activation by *Pen* specifically in hindlimb.

249

250   To see if the Stack model applies more broadly outside of the *Pitx1* domain, we explore published

251   datasets to examine the relationship between enhancer-promoter (E-P) pairs and TAD

252   boundaries. If the Stack organization is a major mechanism for border-crossing interactions, we

253   would expect that cross-TAD interacting *cis*-elements are more likely to be closer to boundaries

254   than intra-TAD interacting *cis*-elements, whose adjacency to boundary matter less since there are

255   no borders to cross. For E-P pairs, we take advantage of the Transcribed Enhancer Atlas from

256   the FANTOM 5 project, where enhancer-promoter pairs were mapped using the correlation of

257   mRNA and eRNA transcription in capped analysis of gene expression (CAGE) results across

258   diverse human cell types [39]. For boundary calling, we use the Hi-C data obtained from GM12878

259 cell line [40] (**Fig. 5a**). We deliberately choose these datasets despite the lack of recency. Rao *et*

260 *al.* remain one of the highest-resolved human Hi-C maps and demonstrated a high degree of

261 conservation of boundaries across cell types. The FANTOM 5 E-P pairs were determined in a

262 TAD-agnostic way and a substantial number of them were > 100 kb apart. Most more recent E-P

263 prediction datasets took TAD boundaries into consideration and/or uncovered relatively few E-P

264 pairs >100 kb apart, possibly due to limitations of cell-type choice and assay sensitivity [41,42],

265 rendering them inappropriate for our purpose.

266

267 We focus on long-range E-P interactions (>100 kb, ~60% of the 1e5 E-P pairs in the dataset),

268 and intersect these with TAD border calls from Hi-C. We find many of these long-range

269 interactions cross a TAD border. To see if border-stacking could play a role in a substantial

270 fraction of these cross-border interactions, we compute the largest distance to the nearest border

271 for both E and P in boundary-crossing pairs and for non-boundary-crossing pairs. We find

272 crossing pairs are significantly more likely to be near a border (*p*=2e-308, Mann-Whitney U and

273 KS-test) than non-crossing pairs, with a median distance of 40 kb from the nearest border for both

274 E and P, rather than over 100 kb (**Fig. 5b**). This statistical trend is consistent with border stacking

275 facilitating border bypass for a substantial number of E-P interactions across the human genome.

276

277

278 **Discussion**

279

280 Here we set out to understand the physical and molecular mechanisms which allow some

281 enhancers to regulate genes across TAD boundaries. Combining imaging and molecular dynamic

282 modeling, we arrived at a model in which strengthening the loop-extrusion-blocking behavior of

283 border elements increases the probability that they will be collected together at a central hub by

284 arresting consecutive cohesin-mediated loops. Enhancers and promoters proximal to these

285   borders thus also gather at the hub, where they interact with elevated frequency, whereas

286   enhancers and promoters more distal from the border are neither intermeshed by extrusion nor

287   collected at hubs. This model provides a mechanistic explanation of why some enhancers can

288   act across TAD borders more easily than others, and a predictive basis on which to identify

289   border-crossing enhancer candidates.

290

291   The picture of the Stack organization (e.g. **Fig. 5c**) has a superficial resemblance to earlier models

292   of multi-enhancer-gene hubs, sometimes described as rosettes [43]. Notable examples include

293   interactions between the protocadherin alpha genes [44], the *Tcr* locus in thymocytes [45] and

294   enhancer "archipelagos" near the *Hoxd* genes [46]. However, they differ from the organization

295   proposed here both in the molecular evidence supporting the model and in the way the structure

296   was interpreted for gene expression. Rosette structures were proposed on the basis of 3C (and

297   in some cases later 4C experiments), which showed preferential ligations among the genes and

298   enhancers. As these methods measure only pairwise interaction, the proposed multi-way hub

299   organization is by nature speculative. Entirely distinct conclusions were drawn from the inferred

300   structure as well. The cell-type specific nature of enhancer-promoter interaction was emphasized,

301   but no question of border and border bypass existed for these loci, nor was any speculation of

302   boundary effects a component of the models. As Hi-C increasingly displaced 3C approaches and

303   provided more ways to correct for the accessibility bias at enhancers and promoters [47], the

304   enhancer rosette model at these loci has given way to a TAD interpretation [48–50]. In investigating

305   how a key developmental enhancer might bypass these TAD borders, we have presented direct

306   data for specific multi-way interactions. Rather than functioning as enhancer hubs, these multiway

307   interactions allow a 3D geometry that preserves boundary function for border-distal elements

308   while enabling boundary-bypass for border-proximal elements.

309

310   In our model, the cell-type specific activation of *Pitx1* gene by the distant *Pen*-enhancer can be

311   explained by the strengthening of extrusion blocking borders in hindlimb relative to forelimb.

312   Experimental data indicates that one of the most notable increases in CTCF border binding occurs

313   at the *Pitx1* promoter itself, with a corresponding increase in Rad21 detection (**Fig. 1b & Fig. 4i**).

314   The strengthening of CTCF binding, especially near the *Pitx1* promoter, increases the frequency

315   with which *Pitx1* joins enhancers at the hub. Such differential CTCF binding could be achieved

316   through several different mechanisms. A simple hypothesis that developing hindlimb cells express

317   more CTCF finds no support in published data surveyed so far, including results from mice [21], bat

318   [51] and chicken [52]. Other more probable hypotheses include tissue-specific expression of cofactors

319   that either recruit CTCF to the *Pitx1* locus or open up binding sites for CTCF. Interference of

320   CTCF binding through DNA methylation of CTCF binding sites could also play an important

321   regulatory role [53,54]. Although we are not aware of any methylomic dataset regarding forelimbs

322   and hindlimbs of tetrapods, a previous study showed differential methylation of a transgene along

323   the anterior-posterior axis in developing mouse embryos [55]. Specifically, methylation level is

324   higher in the forelimbs compared to hindlimbs, resulting in lower expression of the transgene in

325   the forelimbs. If the methylation difference is in the same direction at the *Pitx1* locus, it would help

326   explain the forelimb-hindlimb differences in CTCF ChIP-seq binding, chromatin conformation, and

327   *Pen-Pitx1* contact frequency.

328

329   Our model provides an explanation for some other long-standing puzzles in developmental *cis*-

330   regulation. From flies to mammals, genes in the Hox family are controlled by distal enhancers

331   which act across CTCF-marked TAD boundaries [10,22,56–62]. The *Hoxd genes* are regulated in the

332   vertebrate limb by a series of enhancer "islands" scattered throughout the 1 Mb upstream region

333   [46,50]. Many of these islands, as well as the genes they regulate, are closely linked to CTCF borders

334   [13]. Similarly in human hematopoietic stem cells, enhancers of anterior *Hoxa* genes, located 1.3

335   Mb apart from their promoters and separated by TAD borders, are themselves juxtaposed to a

336   strong TAD border [11,63]. In Drosophila, the promoter of the Hox gene *Abd-B* sits near a boundary

337   and communicates across intervening borders to its distal enhancers *iab-7* and *iab-5* [10], as does

338   the Hox gene *Scr*, whose enhancers also bypass the intervening gene *ftz* [62]. Several of these

339   boundaries (a.k.a. insulators) not only coincide with TAD borders but have been shown to

340   functionally block enhancer-promoter communication in a transgene context and prevent ectopic

341   activation at native loci. Leading geneticists termed this cross-border regulation puzzle "the

342   boundary paradox" [10]. Our model predicts that these borders, rather than inhibit or attenuate,

343   actually facilitate action of the distal enhancers through the preferential stacking of the borders

344   themselves. These examples of border-crossing, boundary-associated enhancers of a few major

345   developmental control genes, along with our genome-wide analysis on the relationship between

346   tens of thousands of inferred E-P pairs and their distances to TAD boundaries, suggest that the

347   border stacking model proposed from *Pitx1* could be informative more generally. Such generality

348   provides a new perspective on the contexts in which insulators will be most and least effective.

349   Adding an insulator between an enhancer and a promoter already anchored by CTCF boundaries

350   may have more subtle effects on contact interactions than when either the enhancer, promoter or

351   both are not closely associated with TAD borders. This may contribute to the surprising recent

352   findings that a variety of previously characterized insulators provide little attenuation between

353   *Sox2* and its distal *SCR* enhancer [64,65].

354

355   Domain boundaries and the process of loop extrusion have long been two of the fundamental

356   concepts by which we contemplate the dynamics of chromatin organization. Despite their billing

357   as features of "3D genome organization", TADs and their boundaries are often thought-of primarily

358   in a quasi-1D manner. In the past, boundaries were regarded as barriers to genomic cross talk:

359   elements only interact with other elements in the same TAD, and each TAD functioned

360   independently from each other (**Fig. 5c**, top panel). In this view the case of *Pen* and other border

361   crossing elements appear to be unexplained exceptions. Our 3D image data adds an additional

362   dimension to this quasi-1D model. While multiple borders cluster together in the Stack

363   organization, border-proximal enhancers and promoters are brought into proximity, even if

364   separated by multiple TADs. Meanwhile, border distal regions of the TAD extrude radially from

365   this hub and are not intermixed, allowing the border function to persist (**Fig. 5c**). Moreover, this

366   border stacking organization is consistent with the loop-extrusion model, as long as cohesin is

367   sufficiently dense on chromatin for multiple cohesins to arrest at opposite sides of an intermediate

368   border, facilitating the formation of a stack. Our model adds a new framework to help us reconcile

369   loop extrusion and TAD borders with previously paradoxical examples of long-range *cis*-regulation,

370   including border bypass and gene skipping behavior, bringing us a step closer to a predictive

371   understanding of genome function.

372

373   **Acknowledgements**

377

378 **Material and Methods**

379

380 **Animal procedures**

381 All animal procedures followed guidelines approved by Stanford University's Administrative

382 Panel on Laboratory Animal Care (APLAC).

383

384 **Mouse limb bud section preparation**

385 C57Bl/6 embryos were harvested at E11.5 or E12.5. Pregnancy was timed by checking vaginal

386 plugs and the developmental stage is confirmed by embryo morphology. The pregnant mouse

387 was euthanized with CO_2 inhalation for 8 minutes followed by cervical dislocation. Limb buds were

388 dissected in PBS and fixed in 4% PFA in PBS at 4°C for 2 hours. (Fixation reduces background

389 signals.) Afterwards limb buds were washed in PBS at 4°C for 5 minutes 3 times and then

390 individually oriented in OCT blocks, frozen in liquid nitrogen and preserved in -80°C. For

391 sectioning, forelimb-bud and hindlimb-bud of the same embryo from the same side (right or left)

392 were sectioned together. The OCT blocks were placed at -20°C for 30 minutes to equilibrate.

393 Limb buds were sectioned from the proximal end along the proximal - distal axis and parallel to

394 the anterior - posterior axis with 8 μm thickness. Sections of forelimb-bud and hindlimb-bud were

395 placed on the same slide, which was coated with chromium–gelatin.

396 For additional details, please refer to "Embryo cryosectioning" (steps 69 - 75) of [23]. Note that for

397 all our samples, we proceeded to probe hybridizing immediately without storing the slides in the

398 fridge or freezer.

399

400 **ORCA experiment**

401 We closely followed the method comprehensively described in [23].

402 **Probe Design**

403   We tiled the mouse *Pitx1* chromosomal domain, mm10 chr13:55,615,000-56,475,00, with 86 10-

404   kb barcodes, each consisting of ~250 individual probes. Each probe is 80 bp long, with 20 bp

405   complementary to fiducial probes, followed by 20 bp of barcode, and ended with 40 bp of unique

406   target sequence. The probes were designed by the software described previously [23]. Note that

407   we only report results for the first 75 barcodes, as not all experiments used all 86 barcodes, and

408   regions beyond barcode 75 were not crucial for the main conclusions in this work. The complete

409   sequences of probes and their genomic targets can be found on the 4DN data portal. The probe

410   design software can be found on https://github.com/BoettigerLab/ORCA-public.The probe library

411   was ordered from CustomArray (now operated by Genscript) as an oligo pool.

412

413   **ORCA hybridization**

414   We followed the procedure from (Tracing DNA paths and RNA profiles in cultured cells and tissues

415   with ORCA | Nature Protocols). See "Hybridization with DNA primary probes", steps 147 - 166.

416   Following are details the protocol did not specify or modifications for our experiment. For most

417   steps, the slide was placed inside a plastic petri dish with a diameter of 6 cm, with buffers added

418   by pouring from a 50 mL tube and removed by suction. During the 37°C incubation with RNaseA

419   (step 152), the petri dish containing the slide was placed inside a humidifying chamber (a tip box

420   with water inside). We used a 22 x 22 mm coverslip at steps 154 and 157. At step 156, our

421   hybridization mixture was composed of 26 µL of hybridization no. 2 plus 3 uL of 500 ng/uL primary

422   probe (29 uL in total).

423

424   **Imaging system setup**

425   For the microscope setup, see [22]. (Also the "Microscope setup" section in Introduction of [23].) All

426   experiments were imaged on the system with the Ti2 body, which used a 1,536 x 1,536 (px) field

427   of view. Briefly, it is composed of an IR-laser based autofocus system, a high performance 3D

428   stage with 500 nm range piezo based z-stage (Ludl), and a custom fluidics system. Illumination

429   is provided by a 561-nm solid-state laser (MPB 2RU-VFL-P-2000-560-B1R, for fiducial probes)

430   and a 647-nm solid-state laser (MPB 2RU-VFL-P-2000-647-B1R, for readout probes). The

431   microscope   setup   is   available   on   the   Micro-Meta:

432   https://data.4dnucleome.org/microscope-configurations/33ea326f-a557-4457-becf-

433   55bc860c4bdf/).

434

435   For the automated fluidics system setup, see the "Fluidics setup" section in Introduction of [23].

436   Briefly, a custom built robotic system with a 3-axis CNC router engraver, buffer reservoirs and

437   hybridization wells (96-well deep well plate) on the 3-axis stage, ETFE tubing, imaging chamber

438   (FCS2, Bioptechs), a needle, and peristaltic pump (Gilson F155006). The needle moved between

439   buffers of hybridization wells and flowed the liquid across the sample in the imaging chamber

440   through the tubings with the pump.

441

442   **Imaging process**

443   For the imaging process, see the "ORCA imaging" section in [22] and steps 167 - 197 of [23]. Briefly,

444   we used open-source software to control and coordinate the fluidic system and the microscope.

445   First, 2x SSC buffer was flown into the sample chamber, and fields-of-view (FOV) were selected

446   to cover either various parts or the whole limb section. Each FOV was then photo-bleached with

447   both 561-nm and 647-nm at maximum intensity for 1 - 3 minutes.

448

449   For each readout barcode, we first flowed in 200–500 µL probe solution (25% ethylene carbonate

450   and 2x SSC containing the readout probe and oligos complementary to the previous readout

451   probe to peel off remaining probes) and incubated for 15 mins to hybridize. Then we flowed 1 mL

452   of wash buffer (30% formamide in 2x SSC buffer) for 2 mins, followed by washing with 1 mL of 2x

453   SSC buffer for 2 mins. At the imaging stage, we filled the chamber with imaging buffer (0.5 mg/ml

454   glucose oxidase, 40 µg/ml catalase and 10% w/v glucose in 2x SSC). Each FOV was then

455   sequentially imaged with 561-nm and 647-nm in z-stacks of 100-nm steps for a total of 100 steps.

456   After all FOVs were imaged, 2x SSC was flown and each FOV was photobleached for 3 seconds

457   by the 647-nm laser. This process was repeated until all barcodes was imaged. Depending on

458   the number of FOVs, each round of imaging took ~1 - 2 hours.

459

460   **Analysis**

461   Image processing was performed following steps 198 onward of [23]. For each experiment, the

462   resultant data are the 3D coordinates of each barcode of each detected chromatin. For each

463   detected chromatin, its 3D coordinates can be converted into a pairwise distance matrix for ease

464   of analysis.

465

466   **Masking of bad hybridizations**

467   When presenting population-average level data, either contact frequencies or median pairwise

468   distances, we mask the bad hybridizations and interpolate with adjacent values. Bad

469   hybridizations, namely 16, 21, 26, 29, 30, 32, 33, 36, 39, 40, 44, 49, 53, 63, 65 (refer to **Fig. 1a**

470   for their genomic positions), are determined by visual examination of the raw imaging data, where

471   either weak signal or off-target bright dots are present. Many of these bad hybridizations are a

472   result of some intrinsic properties of the readout probes. The same probes have yielded poor

473   hybridization in ORCA experiments targeting completely different regions in different organisms.

474

475   **Normalization**

476   The average 3D physical sizes of the probed region may vary among experiments due to

477   variations in time a particular slide spent in different reagents, as well as how individual limbs

478   were snipped and subjected to fixation. To address this batch effect when comparing between

479   experiments, we normalized the median of all pairwise distances between probes 1 - 68 (common

480   to all experiments) measured in a given limb in a given experiment.

481

Insulation score (categorizing "Merge" molecules)

We defined the insulation score of a TAD boundary for a single molecule as the median cross-TAD pairwise distances divided by the median intra-TAD pairwise distances.

485

Simulation

Polymer simulations were performed with the open source polychrom software [36] from open2c. Codes for performing and analyzing results, including the parameters used in each condition, are available on our Github repository. Briefly, the software constructs Langevin dynamic simulations of flexible polymers moving under thermal noise with user defined energy potentials to describe the dynamics of the polymer and molecular interactions among monomers, including the links produced by loop extrusion factors. The simulations use the freely distributed openMM framework [66] to provide GPU accelerated computation. For the generalized simulations in **Fig. 4**, the CTCF loading rates for each CTCF binding site are 0.18 and 0.7 for the low and high CTCF conditions, respectively. For the fine-tuned simulations in **Supp. Fig. 5**, the strength of each CTCF binding site is different, with the low CTCF condition having a 60% CTCF loading rate compared to that of the high CTCF condition.

498

Data availability

The chromosome trace data has been converted to the NIH 4DN data standard, FOF-CT (FISH-omic Format, Chromosome Tracing) and can be accessed at 4DN data portal (https://data.4dnucleome.org/) with accession 4DNES4TC13IL. Simulation data are available on Zenodo with the following DOIs: 10.5281/zenodo.7566077, 10.5281/zenodo.7566087, 10.5281/zenodo.7566087, 10.5281/zenodo.7572059.
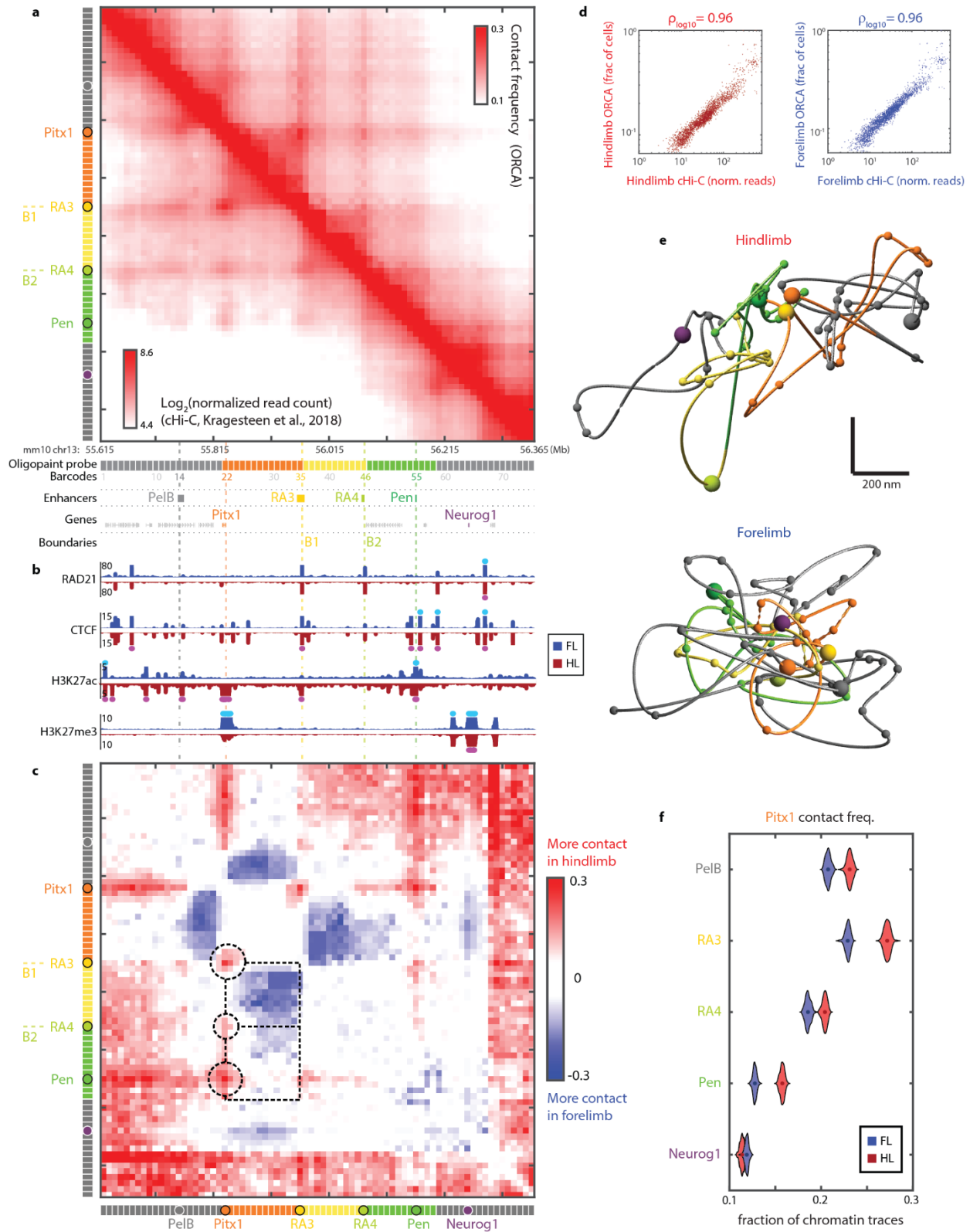
505

Code availability

507 Scripts for analyses and figure generation are available in our github repository:

508 https://github.com/BoettigerLab/ORCA-Pitx1-2022. Polymer simulations also require the

509 simulation toolkit adapted from the open2c project, which is available here:

510 https://github.com/BoettigerLab/polychrom. Probe design and image analysis software is

511 available at https://github.com/BoettigerLab/ORCA-public.

512

513

514 **Figures**

515

516 **Fig. 1 | Chromatin conformation at *Pitx1* domain examined by ORCA agrees with published**

517 **cHi-C data**

518 **a,** Chromatin conformation at *Pitx1* domain in hindlimb cells as detected by ORCA (upper-right half, from 44,966

519 individual chromatin traces) and cHi-C [18] (lower-left half). **b,** ChIP-seq data from E10.5 limb buds (adapted from [21]).

520 Blue: Forelimb data; Red: Hindlimb data. The tracks are thickened to increase visibility. Signals higher than the y-axis

521 limit are marked with bright turquoise and magenta for FL and HL, respectively. **c,** Difference in contact frequency,

522 shown as the log2 ratio of hindlimb to forelimb. See text for discussion on the paradox highlighted by the dashed boxes

523 and circles. **d,** Correlation of cHi-C and ORCA from hindlimb and forelimb. **e,** Representative reconstructed chromatin

524 traces of the *Pitx1* chromosomal domain from hindlimb and forelimb visualized using ORCA. Spheres represent

525 important genes and enhancers. See colorbar in **a** or **c** for reference. **f,** Violin plots quantifying the difference in contact

526 between Pitx1 promoter and the indicated elements. All comparisons are statistically significant. (p = 2.6e-34, p = 2.6e-

527 34, p = 2.8e-13, p = 2.6e-34, p = 2.6e-34, Wilcoxon test.)

528
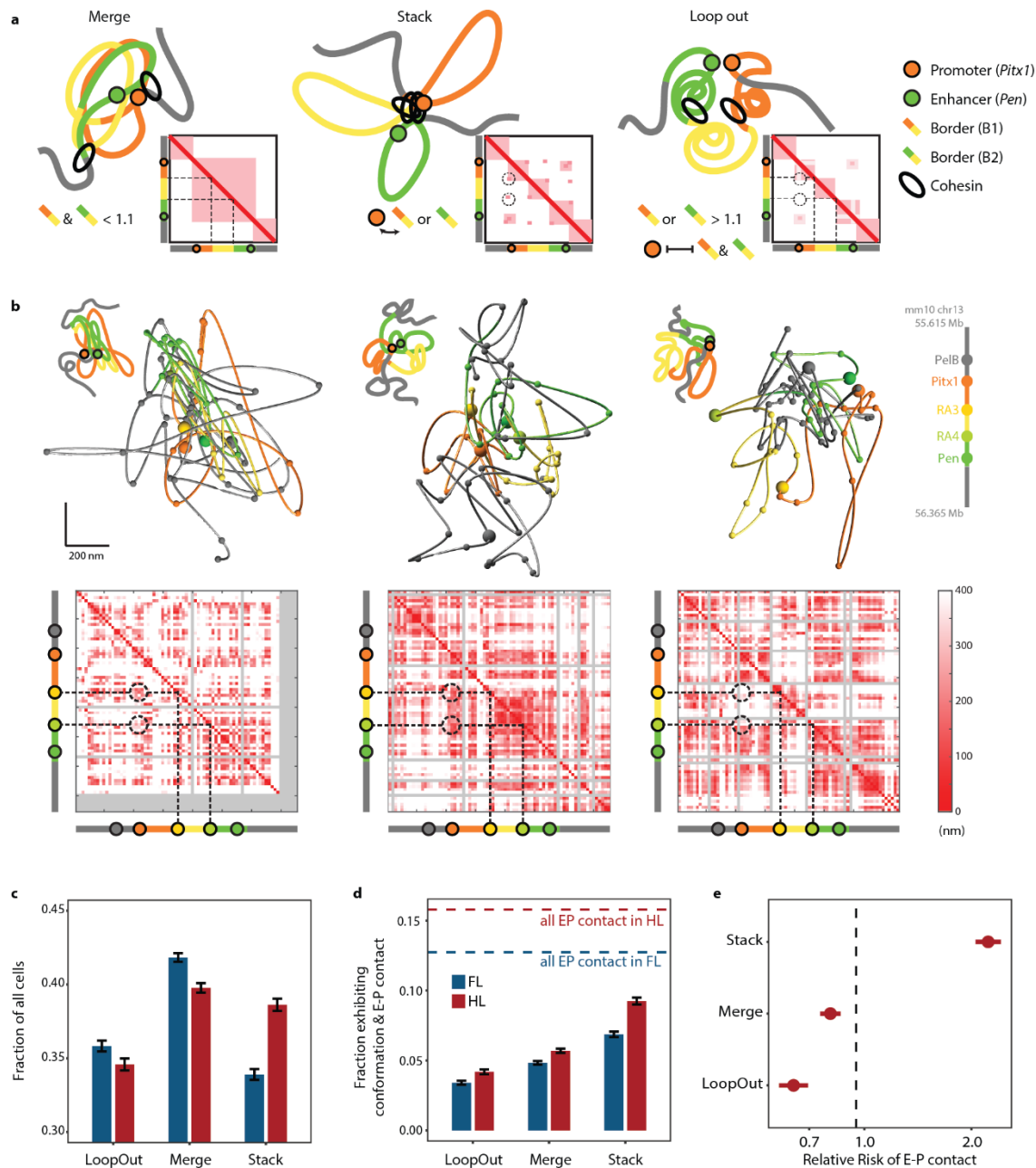
Fig. 2 | Testing three hypotheses regarding cross boundary enhancer promoter interaction

a, Schematic representation of the three models, each showing a cartoon chromatin trace, an idealized distance matrix at lower right, and the defining criteria at lower left. "Merge" is characterized by the weakening of boundaries B1 (orange-yellow) and B2 (yellow-green) between the enhancer and the promoter, emphasized by the dotted lines. "Stack" is characterized by promoter contacting intermediate boundaries B1 or B2, emphasized by the dotted circles. "Loop out" chromatins are those that neither merge nor stack (clear boundaries along the dotted lines and no contact in the dotted circles). **b,** Representative chromatin traces for each model. The top row shows reconstructed chromatin traces

536 along with simplified cartoons that highlights the positions of enhancer, promoter and boundaries. The bottom row

537 shows single-trace distance matrices. Gray rows and columns indicate missing data. The dotted lines mark the

538 boundaries B1 and B2. The dotted circles mark the contacts between *Pitx1* and B1 and B2. Note how the contacts are

539 present in "Stack" but not "LoopOut". **c,** Fraction of all molecules exhibiting configurations matching each model in

540 forelimb (blue) and hindlimb (red). Error bars represent standard deviations. **d,** Fraction of molecules where *Pitx1-Pen*

541 interact exhibiting configurations matching each hypothesis in forelimb (blue) and hindlimb (red). Error bars represent

542 standard deviations. Dashed lines indicate the fraction of all molecules exhibiting *Pitx1-Pen* interaction. **e,** Relative risk

543 of having *Pitx1-Pen* contact for hindlimb molecule configurations matching each model. Bar indicates 95% confidence

544 interval. X-axis is in log scale.
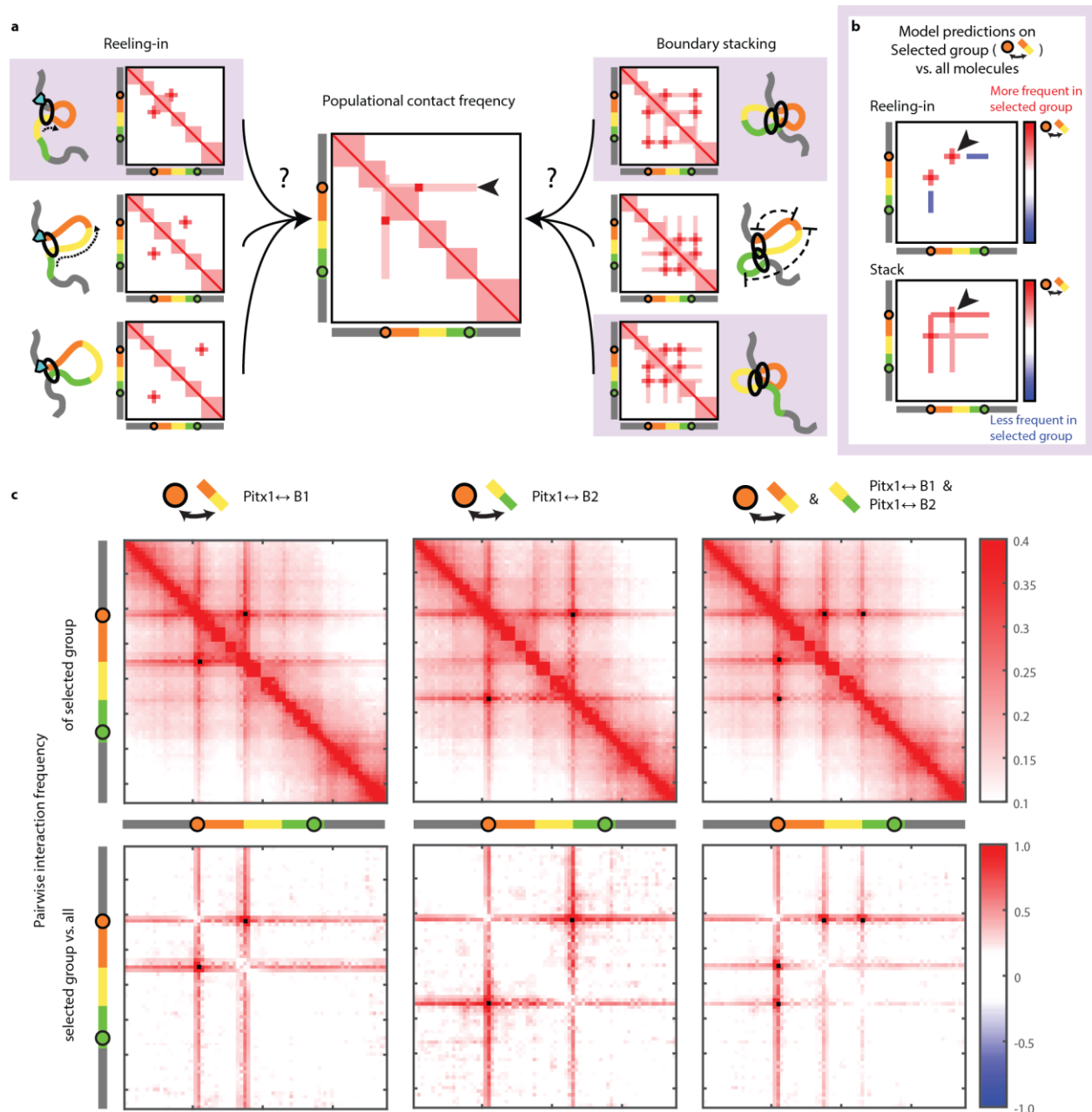
26

Fig. 3 | "Stripes" in single molecule data

a, Two models for the origin of "stripes". (Center) Population-average pairwise contact frequency of the *Pitx1* locus in hindlimbs shows stripe patterns (arrowhead, refer to **Fig. 1a** for actual data.) (Left) Schematic representation of how stripes result through a "reeling-in" mechanism. The anchor at the CTCF site (cyan wedge) on the top, due to directional loop extrusion, interacts more frequently with the colored region compared to any other positions within the same region. Individual chromatin traces are each responsible for a dot along the stripe. (Right) Schematic representation of how stripes result from a "Stack" hub organization. Strong CTCF sites are collected in the central hub in 3D conformation.

553   Due to its central location, the average distance from the hub to any position is shorter compared to pairwise distances

554   between non-hub positions, as illustrated in the center-right panel. **b,** Testing the two models with single-molecule data.

555   When comparing with all molecules, a selected group whose *Pitx1* and B1 are interacting (eg. shaded molecules in **a**)

556   will by definition have a stronger "loop" between the said two positions, regardless of the model (arrowheads). However,

557   "reeling-in" model predicts that other positions along the stripe would have less interaction compared to the population

558   (blue lines), whereas "Stack" model predicts the opposite. **c,** (Top) Pairwise interaction of selected molecules on which

559   *Pitx1* is contacting B1 (left), B2 (center) or both (right). (Bottom) Comparison of selected groups with the whole

560   population, shown as their log2 ratio. The black pixels indicate the interactions used to select the group.
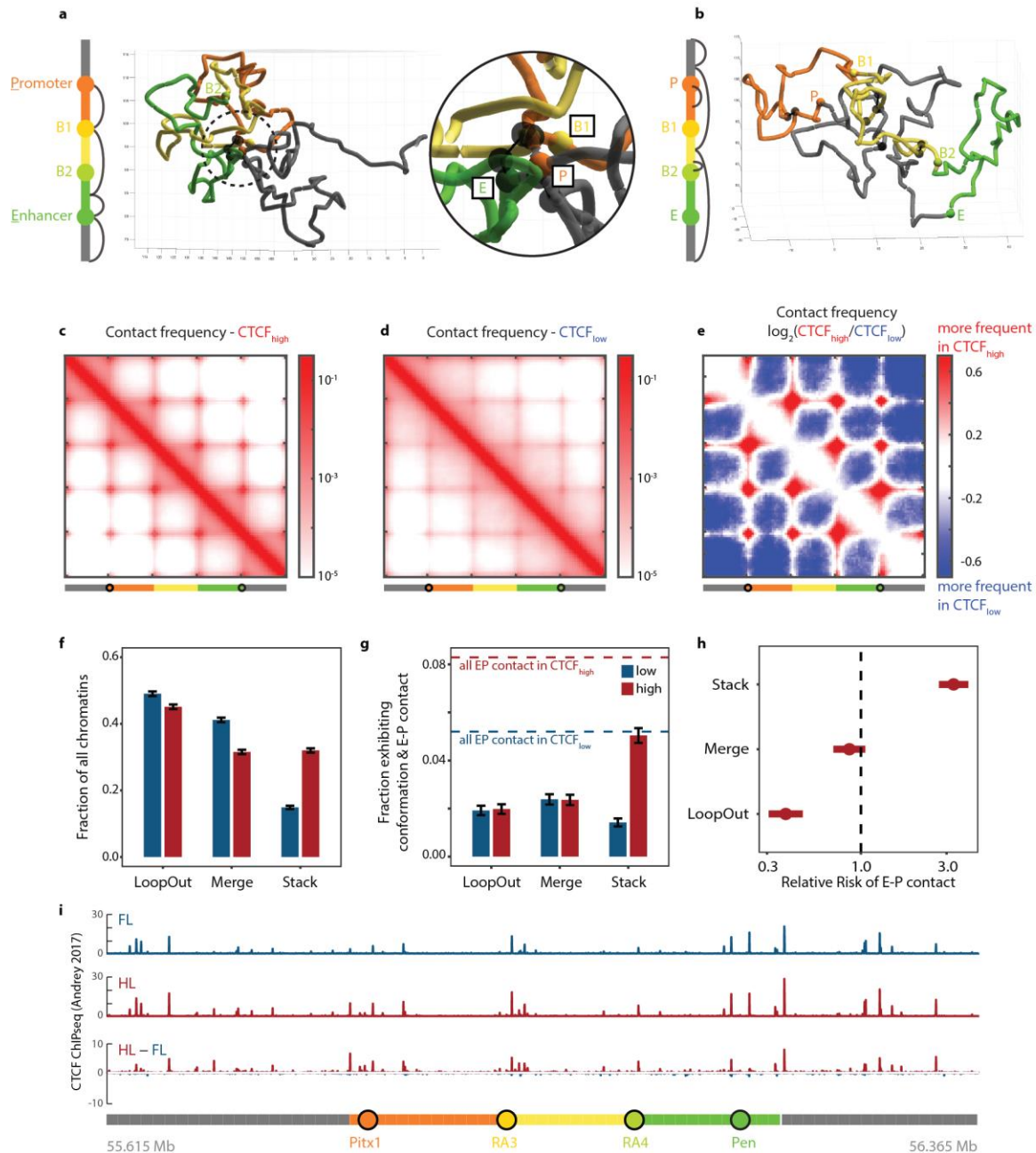
Fig. 4 | Modeling effects of increasing CTCF activity.

a, b, Snapshots of individual simulations. The bars on the left depict the polymer used in simulation straightened, with a color scheme parallel to that used for the actual *Pitx1* chromosomal domain. The colored circles are CTCF sites. Black arcs represent positions brought together by cohesin molecules in that snapshot. **a,** A polymer with high CTCF binding rate and thus more cohesin (transparent black dumbbells) arrested at each CTCF site. The inset shows a zoomed-in view of the circled area, where 3 CTCF sites, the promoter (P), enhancer (E) and intermediate boundary B1, are brought together by arrested cohesins stalled between them in a bridge. **b,** Simulation with low CTCF concentration.

569 **c,** Contact frequency at high CTCF. **d,** Contact frequency at low CTCF. **e,** log2 ratio of data in **c** and **d**. **f,** Fraction of all

570 simulated polymers exhibiting configurations matching each model in low CTCF (blue) and high CTCF (red) conditions.

571 Error bars represent standard deviations. **g,** Fraction of all simulated polymers that have E-P interaction exhibiting

572 configurations matching each model in low CTCF (blue) and high CTCF (red) conditions. Error bars represent standard

573 deviations. Dotted lines represent the fraction of all polymers exhibiting E-P interaction. **h,** Relative risk of having E-P

574 contact for polymers under high CTCF condition while exhibiting configurations matching a model. Bar indicates 95%

575 confidence interval. **i,** Comparison of CTCF ChIP-seq peak heights [21] in hindlimb (red) vs. forelimb (blue) at E10.5.
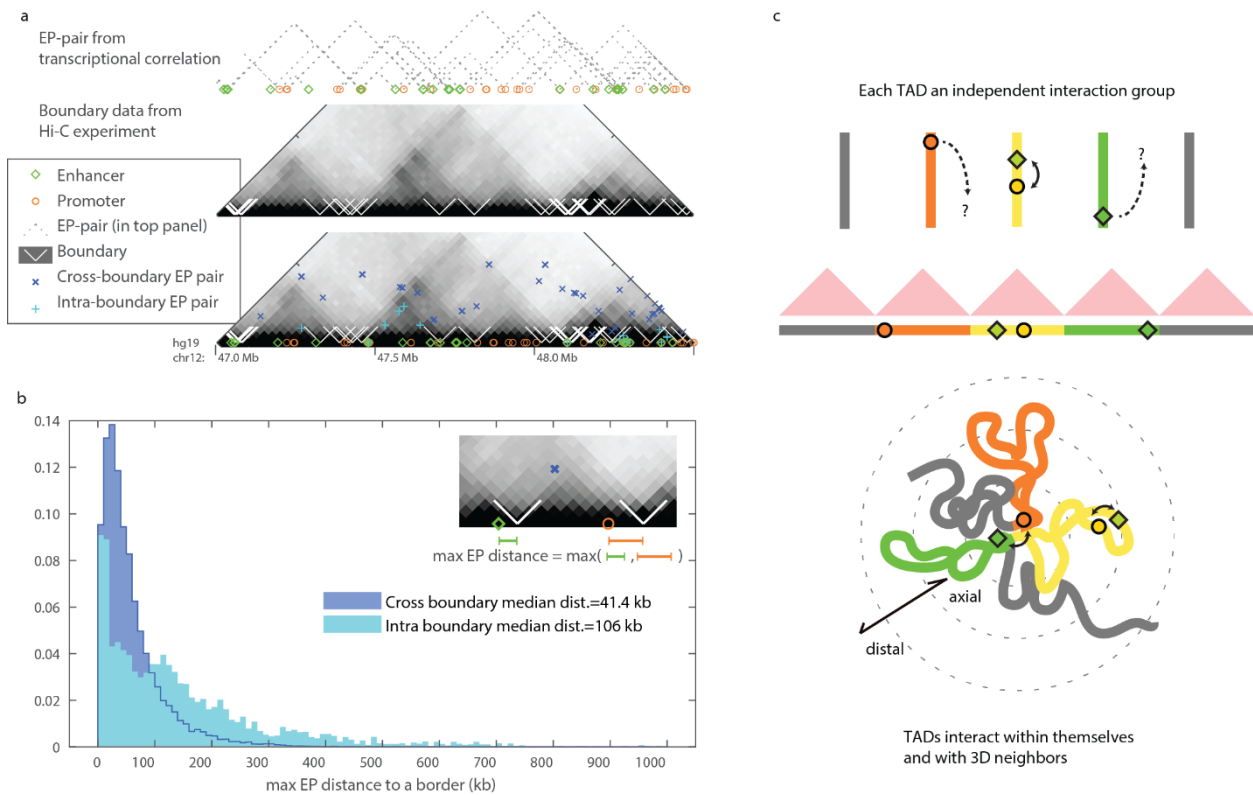

576

577 **Fig. 5 | Genome-wide analysis of cross-TAD enhancer-promoter interactions.**

578 **a, b, Boundary-proximity is enriched for E-P pairs that interact across boundaries. a,** Combining information of

579 E-P pairs across the genome [39] and boundary information from Hi-C data [40], we categorized each E-P pair as cross-

580 boundary or intra-boundary. **b,** Histogram of maximum distance to a border for E-P pairs that are >100 kb apart. **c,**

581 **New perspective on the role of chromatin boundaries.** (Top) Conventional models consider each TAD as largely

582 independent and do not account for cross-TAD interactions. (Bottom) Stack model adds in the 3D relationship between

583 TADs and introduces the dimension of axial - distal position for each element relative to the whole molecule.
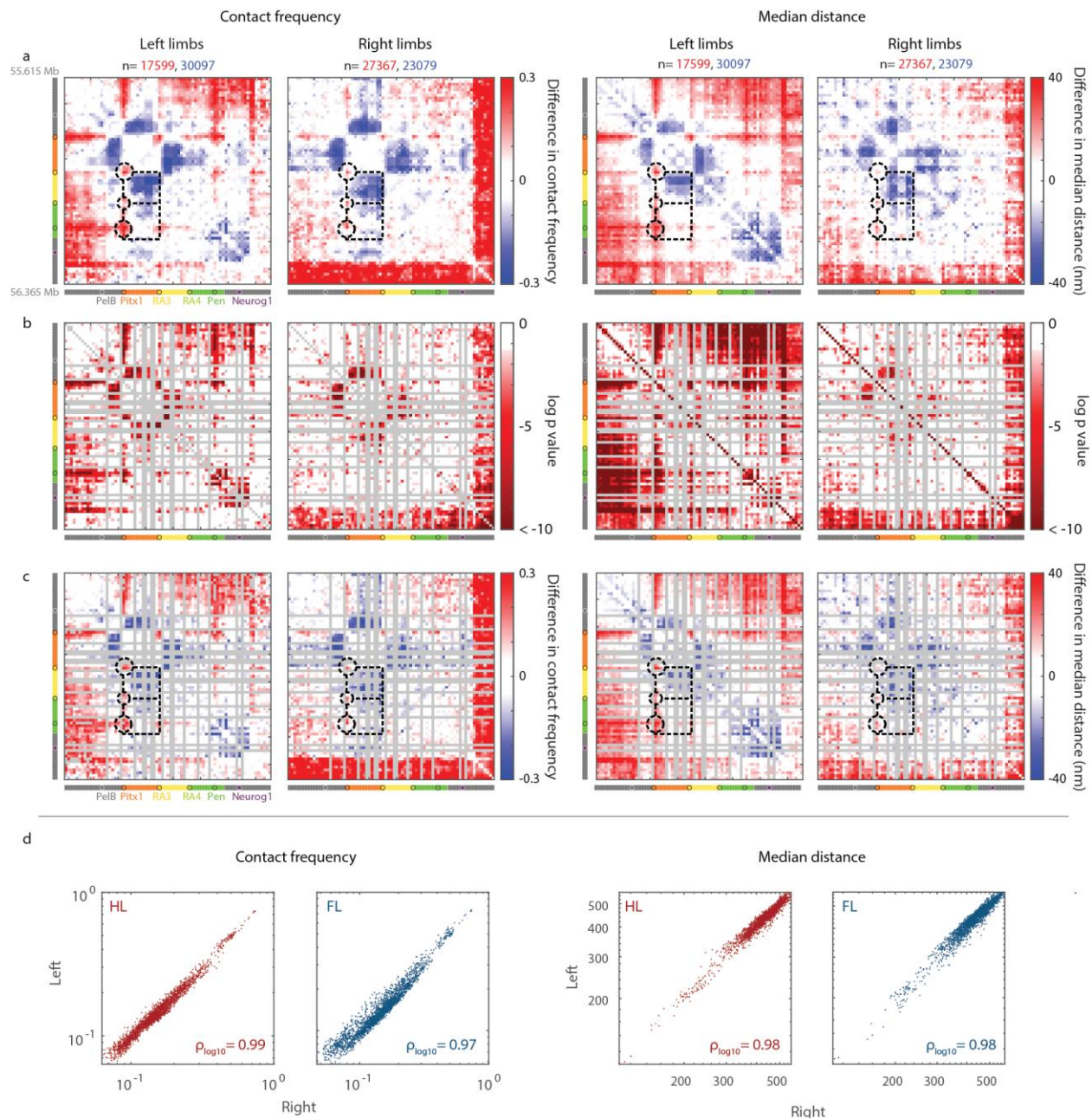
584   **Extended Data Figures**
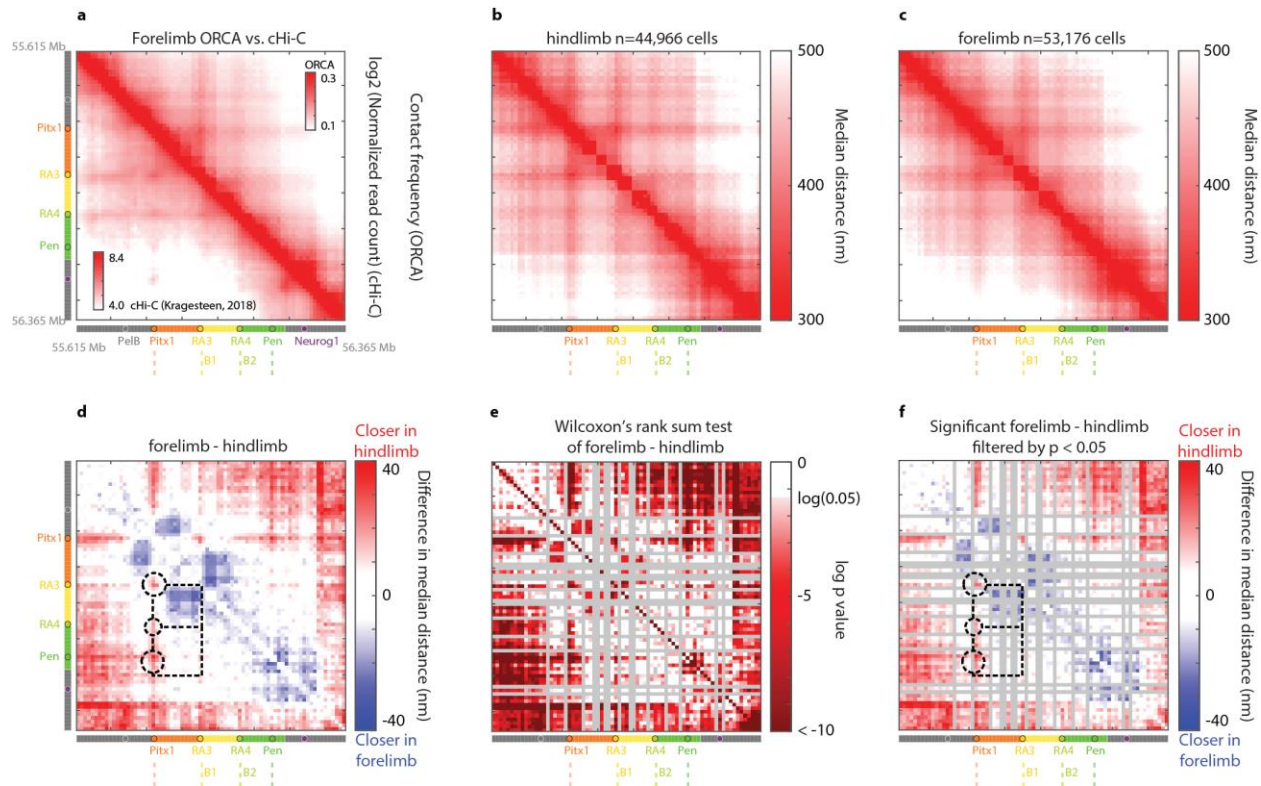
585

586   **Extended Data Fig. 1 | Forelimb contact frequency and median pairwise distance data**

587   **a,** Chromatin conformation at *Pitx1* domain in forelimb cells as detected by ORCA (upper-right half, from 53,176

588   individual chromatin traces) and cHi-C [18] (lower-left half). **b, c,** ORCA median pairwise distance of the *Pitx1*

589   chromosomal domain in hindlimb and forelimb cells, respectively. **d,** Difference in distance between forelimb and

590   hindlimb cells. Red indicates that distance is smaller (closer) in hindlimb. **e,** Matrix of p-values for all distance differences

591   shown in **d**, using Wilcoxon rank-sum test for each pairwise distance. The colorbar turns from white to pink at p = 0.05.

592   Bad hybridizations (see Material and Methods) are masked gray. **f,** Significant (raw p-value < 0.05) distance differences.

593   Non significant differences are masked white.

Extended Data Fig. 2 | Concordance of results between independent ORCA experiments.

The left two columns are the maps for forelimb-hindlimb differences in pairwise contact frequency from experiments using either left limbs or right limbs and independent hybridizations. The right two columns are the maps for forelimb-hindlimb differences in pairwise distances. The two numbers of n denote the number of chromatin traces observed for hindlimb and forelimb, respectively. **a,** The original maps. **b,** Matrix of p-values for all differences shown in **a**, using a chi-square test for each pairwise contact frequency and Wilcoxon rank-sum test for each pairwise distance. The colorbar turns from white to pink at p = 0.05. Bad hybridizations are masked gray. **c,** Significant (raw p-value < 0.05)

602   differences. Non significant differences are masked white. For **a** and **c**, red indicates more contact or closer distance

603   in hindlimb. Difference in contact frequency is shown as the log2 ratio of hindlimb to forelimb. Key features discussed

604   in this paper, highlighted by the dashed boxes and circles, are consistent between experiments from left or right limbs,

605   such as the hindlimb-enriched contacts between *Pitx1* promoter and its enhancers and forelimb-enriched contacts

606   between orange TAD and yellow-green TAD. The "red borders" seen in the final hybridizations (68 - 75) of the right

607   limb is likely an experimental artifact from the end of a run. It is only seen in experiment #1, and the other right limb

608   experiment (#5, see **Supplementary Table 1**) does not have data for these hybridizations. **d,** Correlation between left

609   and right limb experiments for pairwise contact frequencies (left panels) and median distances (right panels).

610
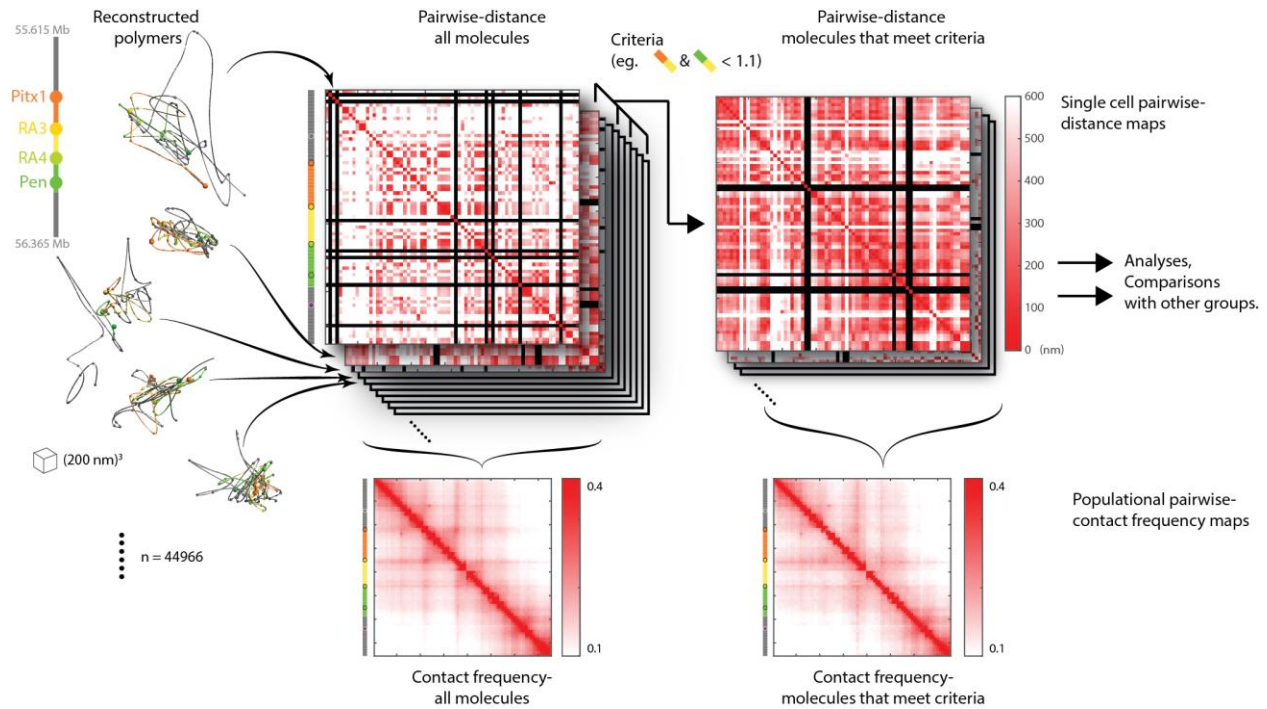
Extended Data Fig. 3 | Selecting groups of molecules in ORCA data based on model criteria

This figure illustrates how, for example, "Merge" molecules or "Stack" molecules were selected from all molecules. (Left) First, the shape of individual *Pitx1* regulatory domains are reconstructed from raw ORCA data. (Center) Each molecule can be represented as a pairwise-distance map. (Right) Maps that match a certain criteria (in this example, weak boundary strengths on both B1 and B2) are selected into a group for further analysis. (See **Fig. 2 c-e**). (Bottom) For any groups of molecules, a populational representation can be obtained using a pairwise-contact frequency map, in which the value at (x, y) represents the percentage of molecules having the distance between positions x and y within 200 nm.
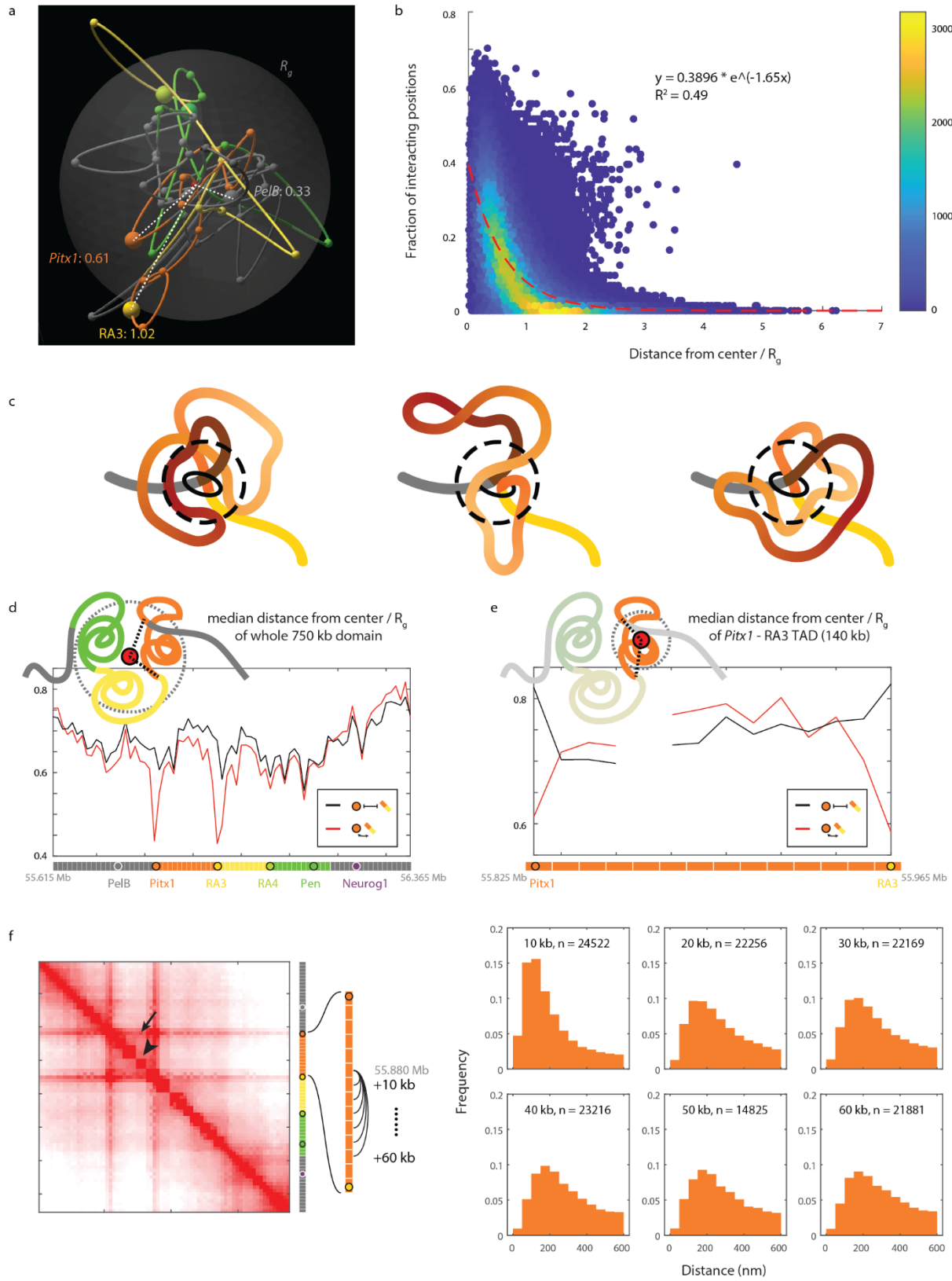
621 **Extended Data Fig. 4 | Central position within a molecule contributes significantly to**

622 **contacting other parts of the molecule.**

623 **a,** Example chromatin trace showing how centrality is measured: for a position within the probed domain (colored

624 spheres along the trace), we measure its distance from the center of the probed domain (small red sphere with radiating

625 dotted white lines), normalized by the domain's radius of gyration (R_g, translucent gray globe). The centrality of *PelB*

626 (gray), *Pitx1* (orange) and RA3 (yellow) are shown as examples. **b,** The relationship between centrality within a

627 molecule and contact frequency with other parts of the molecule, shown by data drawn from 641,687 10kb segments

628 from 10,089 high-quality molecules across 5 independent experiments. Each dot represents a 10kb segment within the

629 750 kb probed domain. The X axis denotes distance from the center. The Y axis denotes the percentage of other

630 segments within the same molecule contacting this segment (i.e. within 200 nm). Contacting multiple other segments

631 happens mostly when a segment is positioned centrally in the molecule. Red dash line: exponential regression of the

632 data, with its formula and R^2 value indicated. **c,** Schematics that convey how central positioning can facilitate contact

633 with all parts of the molecule. The dotted circles denote contact distance threshold. **d,** Distance from center of each

634 10kb segment across the *Pitx1* domain. Compared to all the molecules (black line), the subset of chromatin traces

635 where the *Pitx1* promoter is contacting the RA3 enhancer (red line) are more likely to have both positions be at the

636 center of the molecule. **e,** Even when considering the centrality only within the TAD between *Pitx1* and RA3, the two

637 are still centrally positioned when they are in contact (red), despite their terminal linear positions. As illustrated by the

638 cartoons on top of each subfigure, the differences between **d** and **e** are the choice of "center" (red) and radius of

639 gyration (dashed gray circle). (Missing data: due to technical issues, the hybridization did not go well for the 5th data

640 point in **e**. In other figures we interpolated for its position, but because doing so would result in a striking visual feature

641 here, we left this segment blank). **f,** (Left) Contact frequency for molecules where *Pitx1* and RA3 are interacting, same

642 as **Fig. 2c**. It might seem counterintuitive that being central to a region confers more contact than being linearly closer

643 along the chromosome. (Compare the pixel marked by the arrow, representing interaction between two segments 80

644 kb apart, and that marked by the arrowhead, representing interaction between two segments 20 kb apart). However,

645 what seems like a few pixel's distance on the plot is actually many kilobases apart (since are regions tiled at 10 kb

646 intervals), and within a TAD, which is not a rigid structure, the chromatin is rather flexible. (Right) Distributions of 3D

647 distances between segment #27 (mm10, chr13: 55,875,000 - 55,885,000) and increasingly distant segments show that,

648 despite being in the same TAD, distance distributions for segments over 30 kb apart do not differ much from each other,

649 indicating that increased contact frequency conferred by linear proximity is limited to very proximal elements.

650
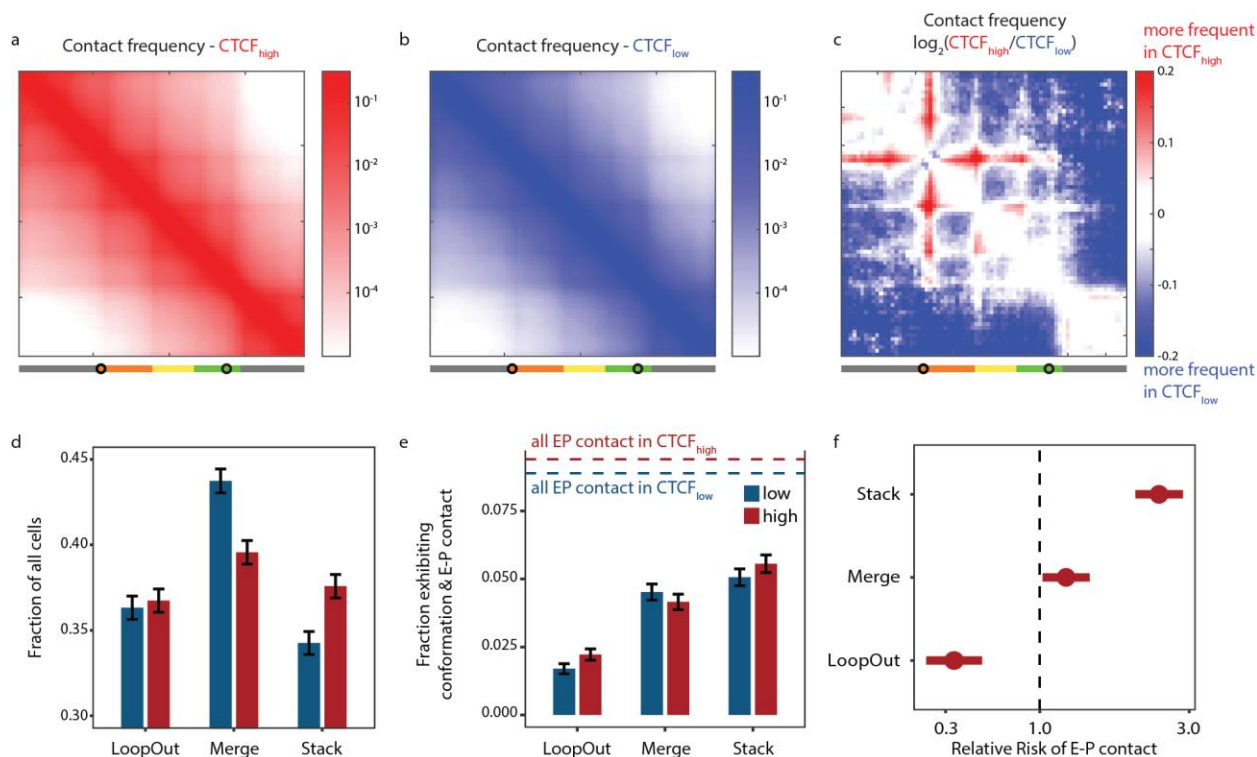
Extended Data Fig. 5 | Modeling effects of increasing CTCF activity with fine-tuned parameters.

a, Contact frequency at high CTCF. **b,** Contact frequency at low CTCF. **c,** log2 ratio of data in **a** and **b**. **d,** Fraction of all simulated polymers exhibiting configurations matching each model in low CTCF (blue) and high CTCF (red) conditions. Error bars represent standard deviations. Stack is the only configuration higher in the high CTCF condition. **e,** Fraction of all simulated polymers that have E-P interaction exhibiting configurations matching each model in low CTCF (blue) and high CTCF (red) conditions. Error bars represent standard deviations. Dotted lines represent the fraction of all polymers exhibiting E-P interaction. **f,** Relative risk of having E-P contact for polymers under high CTCF condition while exhibiting configurations matching a model. Bar indicates 95% confidence interval. Stack configuration grants the highest RR for EP contact. Although the RR for Merge is also positive, it is far more frequent in the low CTCF condition (see panel D).

664 **Supplementary Data**

665

666 **Supplementary Table. 1 | Summary of the five experiments.**

667 Experiments 2 and 3 used different sections from the same limbs. Experiments 4 and 5 used different embryos from

668 the same litter.

669

Experiment#	Stage	Left/Right	Barcodes probed	Forelimb N	Hindlimb N
1	E12.5	Right	1-75	10087	11607
2*	E11.5	Left	1-75	7156	7326
3*	E11.5	Left	1-70	14220	3814
4	E11.5	Left	1-75	8721	6459
5	E11.5	Right	1-67	12992	15760

670

671

672 **References**

673

674 1. Dixon, J. R., Gorkin, D. U. & Ren, B. Chromatin Domains: The Unit of Chromosome

675   Organization. *Mol. Cell* **62**, 668–680 (2016).

676 2. Dekker, J. & Mirny, L. A. The 3D Genome as Moderator of Chromosomal Communication.

677   *Cell* **164**, 1110–1121 (2016).

678 3. Rowley, M. J. & Corces, V. G. Organizational principles of 3D genome architecture. *Nat.*

679   *Rev. Genet.* **19**, 789–800 (2018).

680 4. Dekker, J. & Heard, E. Structural and functional diversity of Topologically Associating

681   Domains. *FEBS Lett.* **589**, 2877–2884 (2015).

682 5. Ibrahim, D. M. & Mundlos, S. The role of 3D chromatin domains in gene regulation: a multi-

683   facetted view on genome organization. *Curr. Opin. Genet. Dev.* **61**, 1–8 (2020).

684 6. Ibrahim, D. M. & Mundlos, S. Three-dimensional chromatin in disease: What holds us

685   together and what drives us apart? *Curr. Opin. Cell Biol.* **64**, 1–9 (2020).

686 7. Cavalheiro, G. R., Pollex, T. & Furlong, E. E. To loop or not to loop: what is the role of

687   TADs in enhancer function and gene regulation? *Curr. Opin. Genet. Dev.* **67**, 119–129

688   (2021).

689 8. McCord, R. P., Kaplan, N. & Giorgetti, L. Chromosome Conformation Capture and Beyond:

690   Toward an Integrative View of Chromosome Structure and Function. *Mol. Cell* **77**, 688–708

691   (2020).

692 9. Jerkovic´, I. & Cavalli, G. Understanding 3D genome organization by multidisciplinary

693   methods. *Nat. Rev. Mol. Cell Biol.* 1–18 (2021).

694 10. Kyrchanova, O. *et al.* The boundary paradox in the Bithorax complex. *Mech. Dev.* **138**,

695   122–132 (2015).

696 11. Wang, X. Q. D. *et al.* Three-dimensional regulation of *HOXA* cluster genes by a *cis*-element

697   in hematopoietic stem cell and leukemia. *bioRxiv* 2020.04.16.017533 (2020)

698   doi:10.1101/2020.04.16.017533.

699   12. Beccari, L. *et al. Dbx2* regulation in limbs suggests interTAD sharing of enhancers. *Dev.*

700   *Dyn.* (2021) doi:10.1002/dvdy.303.

701   13. Rodríguez-Carballo, E. *et al.* The *HoxD* cluster is a dynamic and resilient TAD boundary

702   controlling the segregation of antagonistic regulatory landscapes. *Genes Dev.* **31**, 2264–

703   2281 (2017).

704   14. Cannavò, E. *et al.* Shadow Enhancers Are Pervasive Features of Developmental

705   Regulatory Networks. *Curr. Biol.* **26**, 38–51 (2016).

706   15. Osterwalder, M. *et al.* Enhancer redundancy provides phenotypic robustness in mammalian

707   development. *Nature* **554**, 239–243 (2018).

708   16. Spielmann, M. *et al.* Homeotic arm-to-leg transformation associated with genomic

709   rearrangements at the *PITX1* locus. *Am. J. Hum. Genet.* **91**, 629–635 (2012).

710   17. Al-Qattan, M. M., Al-Thunayan, A., Alabdulkareem, I. & Al Balwi, M. Liebenberg syndrome

711   is caused by a deletion upstream to the *PITX1* gene resulting in transformation of the upper

712   limbs to reflect lower limb characteristics. *Gene* **524**, 65–71 (2013).

713   18. Kragesteen, B. K. *et al.* Dynamic 3D chromatin architecture contributes to enhancer

714   specificity and limb morphogenesis. *Nat. Genet.* **50**, 463–473 (2018).

715   19. Sarro, R. *et al.* Disrupting the three-dimensional regulatory topology of the *Pitx1* locus

716   results in overtly normal development. *Development* **145**, (2018).

717   20. Thompson, A. C. *et al.* A novel enhancer near the *Pitx1* gene influences development and

718   evolution of pelvic appendages in vertebrates. *Elife* **7**, (2018).

719   21. Andrey, G. *et al.* Characterization of hundreds of regulatory landscapes in developing limbs

720   reveals two regimes of chromatin folding. *Genome Res.* **27**, 223–233 (2017).

721   22. Mateo, L. J. *et al.* Visualizing DNA folding and RNA in embryos at single-cell resolution.

722   *Nature* **568**, 49–54 (2019).

723   23.   Mateo, L. J., Sinnott-Armstrong, N. & Boettiger, A. N. Tracing DNA paths and RNA profiles
724   in cultured cells and tissues with ORCA. *Nat. Protoc.* **16**, 1647–1713 (2021).
725   24.   Cardozo Gizzi, A. M. *et al.* Direct and simultaneous observation of transcription and
726   chromosome architecture in single cells with Hi-M. *Nat. Protoc.* **15**, 840–876 (2020).
727   25.   Cardozo Gizzi, A. M. *et al.* Microscopy-Based Chromosome Conformation Capture Enables
728   Simultaneous Visualization of Genome Organization and Transcription in Intact Organisms.
729   *Mol. Cell* **74**, 212–222.e5 (2019).
730   26.   Liu, M. *et al.* Multiplexed imaging of nucleome architectures in single cells of mammalian
731   tissue. *bioRxiv* 2019.12.20.885277 (2019) doi:10.1101/2019.12.20.885277.
732   27.   Liu, M. *et al.* Chromatin tracing and multiplexed imaging of nucleome architectures (MINA)
733   and RNAs in single mammalian cells and tissue. *Nat. Protoc.* **16**, 2667–2697 (2021).
734   28.   Takei, Y. *et al.* Integrated spatial genomics reveals global architecture of single nuclei.
735   *Nature* **590**, 344–350 (2021).
736   29.   Cheutin, T. & Cavalli, G. The multiscale effects of polycomb mechanisms on 3D chromatin
737   folding. *Crit. Rev. Biochem. Mol. Biol.* **54**, 399–417 (2019).
738   30.   Rhodes, J. D. P. *et al.* Cohesin Disrupts Polycomb-Dependent Chromosome Interactions in
739   Embryonic Stem Cells. *Cell Rep.* **30**, 820–835.e10 (2020).
740   31.   Laura Vian, A. *et al.* The Energetics and Physiological Impact of Cohesin Extrusion. *Cell*
741   **173**, 1–14 (2018).
742   32.   Kraft, K. *et al.* Serial genomic inversions induce tissue-specific architectural stripes, gene
743   misexpression and congenital malformations. *Nat. Cell Biol.* **21**, 305–310 (2019).
744   33.   Krietenstein, N. *et al.* Ultrastructural Details of Mammalian Chromosome Architecture. *Mol.*
745   *Cell* **78**, 554–565.e7 (2020).
746   34.   Hsieh, T.-H. S. *et al.* Resolving the 3D Landscape of Transcription-Linked Mammalian
747   Chromatin Folding. *Mol. Cell* **78**, 539–553.e8 (2020).
748   35.   Chen, L.-F. *et al.* Structural elements facilitate extreme long-range gene regulation at a

749   human disease locus. *bioRxiv* 2022.10.20.513057 (2022) doi:10.1101/2022.10.20.513057.

750 36. Imakaev, M., Goloborodko, A. & Brandao, H. *polychrom: Polymer simulations of*

751   *chromosomes and generating 'in silico' Hi-C maps.* (Github, 2019).

752   doi:https://zenodo.org/badge/latestdoi/178608195.

753 37. Fudenberg, G. *et al.* Formation of Chromosomal Domains by Loop Extrusion. *Cell Rep.* **15**,

754   2038–2049 (2016).

755 38. Nuebler, J., Fudenberg, G., Imakaev, M., Abdennur, N. & Mirny, L. A. Chromatin

756   organization by an interplay of loop extrusion and compartmental segregation. *Proc. Natl.*

757   *Acad. Sci. U. S. A.* **115**, E6697–E6706 (2018).

758 39. Andersson, R. *et al.* An atlas of active enhancers across human cell types and tissues.

759   *Nature* **507**, 455–461 (2014).

760 40. Rao, S. S. P. *et al.* A 3D map of the human genome at kilobase resolution reveals

761   principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).

762 41. Gasperini, M. *et al.* A Genome-wide Framework for Mapping Gene Regulation via Cellular

763   Genetic Screens. *Cell* **176**, 377–390.e19 (2019).

764 42. Nasser, J. *et al.* Genome-wide enhancer maps link risk variants to disease genes. *Nature*

765   **593**, 238–243 (2021).

766 43. Gómez-Díaz, E. & Corces, V. G. Architectural proteins: regulators of 3D genome

767   organization in cell fate. *Trends Cell Biol.* **24**, 703–711 (2014).

768 44. Guo, Y. *et al.* CTCF/cohesin-mediated DNA looping is required for protocadherin α

769   promoter choice. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 21081–21086 (2012).

770 45. Seitan, V. C. *et al.* A role for cohesin in T-cell-receptor rearrangement and thymocyte

771   differentiation. *Nature* **476**, 467–471 (2011).

772 46. Montavon, T. *et al.* A Regulatory Archipelago Controls *Hox* Genes Transcription in Digits.

773   *Cell* **147**, 1132–1145 (2011).

774 47. Hafner, A. & Boettiger, A. The spatial organization of transcriptional control. *Nat. Rev.*

775   *Genet.* **24**, 53–68 (2023).

776   48. Ba, Z. *et al.* CTCF orchestrates long-range cohesin-driven V(D)J recombinational scanning.

777   *Nature* **586**, 305–310 (2020).

778   49. Kiefer, L. *et al.* Cohesin erases genomic-proximity biases to drive stochastic Protocadherin

779   expression for proper neural wiring. *bioRxiv* 2022.03.09.483674 (2022)

780   doi:10.1101/2022.03.09.483674.

781   50. Andrey, G. *et al.* A switch between topological domains underlies *HoxD* genes collinearity

782   in mouse limbs. *Science* **340**, 1234167 (2013).

783   51. Eckalbar, W. L. *et al.* Transcriptomic and epigenomic characterization of the developing bat

784   wing. *Nat. Genet.* **48**, 528–536 (2016).

785   52. Farlie, P. G. *et al.* Co-option of the cardiac transcription factor *Nkx2.5* during development

786   of the emu wing. *Nat. Commun.* **8**, 132 (2017).

787   53. Hark, A. T. *et al.* CTCF mediates methylation-sensitive enhancer-blocking activity at the

788   *H19/Igf2* locus. *Nature* **405**, 486–489 (2000).

789   54. Wang, H. *et al.* Widespread plasticity in CTCF occupancy linked to DNA methylation.

790   *Genome Res.* **22**, 1680–1688 (2012).

791   55. Donoghue, M. J., Patton, B. L., Sanes, J. R. & Merlie, J. P. An axial gradient of transgene

792   methylation in murine skeletal muscle: genomic imprint of rostrocaudal position.

793   *Development* **116**, 1101–1112 (1992).

794   56. Berlivet, S. *et al.* Clustering of Tissue-Specific Sub-TADs Accompanies the Regulation of

795   *HoxA* Genes in Developing Limbs. *PLoS Genet.* **9**, e1004018 (2013).

796   57. Narendra, V. *et al.* CTCF establishes discrete functional chromatin domains at the *Hox*

797   clusters during differentiation. *Science* **347**, 1017–1021 (2015).

798   58. Narendra, V., Bulajić, M., Dekker, J., Mazzoni, E. O. & Reinberg, D. CTCF-mediated

799   topological boundaries during development foster appropriate gene regulation. *Genes Dev.*

800   **30**, 2657–2662 (2016).

801   59. Ishihara, K., Nakamoto, M. & Nakao, M. DNA methylation-independent removable insulator

802   controls chromatin remodeling at the *HOXA* locus via retinoic acid signaling. *Hum. Mol.*

803   *Genet.* **25**, 5383–5394 (2016).

804   60. Rodríguez-Carballo, E. *et al.* Chromatin topology and the timing of enhancer function at the

805   *HoxD* locus. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 31231–31241 (2020).

806   61. Amândio, A. R. *et al.* Sequential in *cis* mutagenesis *in vivo* reveals various functions for

807   CTCF sites at the mouse *HoxD* cluster. *Genes Dev.* **35**, 1490–1509 (2021).

808   62. Batut, P. J. *et al.* Genome organization controls transcriptional dynamics during

809   development. *Science* **375**, 566–570 (2022).

810   63. Neijts, R. *et al.* Polarized regulatory landscape and Wnt responsiveness underlie Hox

811   activation in embryos. *Genes Dev.* **30**, 1937–1942 (2016).

812   64. Huang, H. *et al.* CTCF mediates dosage- and sequence-context-dependent transcriptional

813   insulation by forming local chromatin domains. *Nat. Genet.* 1–11 (2021).

814   65. Chakraborty, S. *et al.* High affinity enhancer-promoter interactions can bypass

815   CTCF/cohesin-mediated insulation and contribute to phenotypic robustness. *bioRxiv*

816   2021.12.30.474562 (2022) doi:10.1101/2021.12.30.474562.

817   66. Eastman, P. *et al.* OpenMM 7: Rapid development of high performance algorithms for

818   molecular dynamics. *PLoS Comput. Biol.* **13**, e1005659 (2017).

819