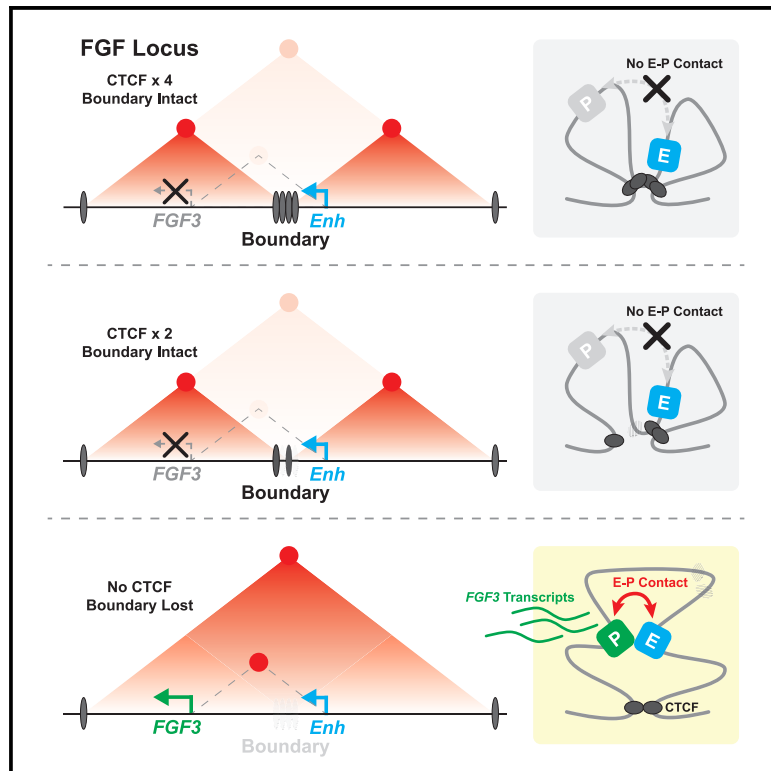


## Dissection of a CTCF topological boundary uncovers principles of enhancer-oncogene regulation

### Graphical abstract



### Authors

Kyung Lock Kim, Gilbert J. Rahme, Viraat Y. Goel, Chadi A. El Farran, Anders S. Hansen, Bradley E. Bernstein

### Correspondence

bradley\_bernstein@dfci.harvard.edu

### In brief

Enhancer-gene communication is dependent on chromosome topology, which helps enhancers find their correct gene targets. Kim et al. characterize topological boundaries that insulate the oncogenes *FGF3* and *FGF4* from a potent enhancer and show that combinatorial disruption of CTCF sites in the boundary results in enhancer-promoter contact and gene induction.

### Highlights

- FGF oncogenes insulated from a potent enhancer by multiple redundant CTCF sites
- RCMC maps delineate CTCF contacts and boundaries that separate genes from enhancer
- Combinatorial disruption of CTCF sites results in aberrant enhancer-promoter contact
- Contact between enhancer and promoter initiation sites correlates with FGF3 induction

Short article

# Dissection of a CTCF topological boundary uncovers principles of enhancer-oncogene regulation

Kyung Lock Kim,<sup>1,2,3</sup> Gilbert J. Rahme,<sup>1,2,3</sup> Viraat Y. Goel,<sup>2,4,5</sup> Chadi A. El Farran,<sup>1,2,3</sup> Anders S. Hansen,<sup>2,4,5</sup> and Bradley E. Bernstein<sup>1,2,3,6,\*</sup>

<sup>1</sup>Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, MA 02215, USA

<sup>2</sup>Gene Regulation Observatory, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA

<sup>3</sup>Departments of Cell Biology and Pathology, Harvard Medical School, Boston, MA 02215, USA

<sup>4</sup>Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

<sup>5</sup>Koch Institute for Integrative Cancer Research, Cambridge, MA 02139, USA

<sup>6</sup>Lead contact

\*Correspondence: [bradley\\_bernstein@dfci.harvard.edu](mailto:bradley_bernstein@dfci.harvard.edu)

<https://doi.org/10.1016/j.molcel.2024.02.007>

## SUMMARY

Enhancer-gene communication is dependent on topologically associating domains (TADs) and boundaries enforced by the CCCTC-binding factor (CTCF) insulator, but the underlying structures and mechanisms remain controversial. Here, we investigate a boundary that typically insulates fibroblast growth factor (*FGF*) oncogenes but is disrupted by DNA hypermethylation in gastrointestinal stromal tumors (GISTs). The boundary contains an array of CTCF sites that enforce adjacent TADs, one containing *FGF* genes and the other containing *ANO1* and its putative enhancers, which are specifically active in GIST and its likely cell of origin. We show that coordinate disruption of four CTCF motifs in the boundary fuses the adjacent TADs, allows the *ANO1* enhancer to contact *FGF3*, and causes its robust induction. High-resolution micro-C maps reveal specific contact between transcription initiation sites in the *ANO1* enhancer and *FGF3* promoter that quantitatively scales with *FGF3* induction such that modest changes in contact frequency result in strong changes in expression, consistent with a causal relationship.

## INTRODUCTION

Gene regulation is a major driver of cell identity, allowing cells that share the same DNA to differentiate into the diverse lineages that underlie tissue and organismal biology.<sup>1,2</sup> By recent estimates, the human genome contains more than a million enhancer elements thought to dictate the precise spatiotemporal patterns of gene expression crucial for this specification and tissue function.<sup>3–5</sup> Their locations and cell type specificities can be predicted from genomic maps of chromatin accessibility or histone modifications.<sup>6</sup> However, a fundamental unanswered question remains: how are these enhancers, which are distributed across vast swaths of chromosomal DNA, guided to their appropriate gene targets. The ability of some enhancers to activate genes across distances of tens or hundreds of kilobases (kb) suggests that genome conformation plays an important role while also alluding to the complexity of enhancer targeting mechanisms.

The role of chromosome conformation (“topology”) and physical contact between enhancers and promoters in gene regulation and transcriptional control is an area of active investigation. Topology can be mapped by chromosome conformation capture-based methods that measure contact frequency between

genomic loci by restricting and ligating DNA in its native chromatin context.<sup>7</sup> The Hi-C method measures such interactions in an all-by-all manner by deep sequencing of ligation junctions.<sup>8</sup> Although Hi-C produces genome-wide maps, its effective resolution is limited by restriction site availability and high sequencing requirement. This has prompted the development of alternative approaches that increase resolution by digesting with micrococcal nuclease (micro-C),<sup>9</sup> enrich loci of interest by hybrid capture or amplification (e.g., capture-C),<sup>10</sup> or measure one-to-all interactions from a prespecified viewpoint (e.g., 4C).<sup>11</sup>

Application of these conformational assays has revealed that the genome is partitioned into discrete regions of self-interacting chromatin, termed topologically associating domains (TADs). TADs correspond to loops of chromatin extruded through cohesin rings<sup>12,13</sup> and are frequently bounded by binding sites for the CCCTC-binding factor (CTCF) insulator, which halts extrusion.<sup>5,14</sup> A compelling interpretation is that enhancers and promoters are constrained by these topological structures, such that regulatory interactions primarily occur between elements in the same TAD.<sup>5,14–16</sup> However, this straightforward interpretation of TADs as “insulated neighborhoods” has been challenged by microscopy studies that find TADs to represent relatively short-lived structures.<sup>17,18</sup> Furthermore, enhancer-promoter

interactions are not clearly evident in Hi-C maps<sup>14</sup> and are difficult to appreciate by microscopy,<sup>19–22</sup> leaving open the question of whether and how proximity translates to transcriptional activation. Although mainstream models assume that physical enhancer-promoter contacts direct transcription, alternate models invoke nonlinear<sup>23–25</sup> or linear<sup>26,27</sup> relationships to quantitatively relate physical contact to transcriptional activity. Clarifying these relationships is critical for understanding enhancer functions and the role of TADs in regulating their activities. The lack of clarity has also prompted alternate proposals in which contact-free effects are mediated through condensates or diffusion of transcription factors (TFs).<sup>24,28</sup>

Human genetics and disease studies have provided important insights into the functions of TADs and boundaries. Germline mutations that alter chromosome topology in the *EphA4*, *Shh*, and *Sox9* loci disrupt enhancer-promoter interactions crucial for limb development, leading to polydactyly.<sup>29–31</sup> Somatic mutations that disrupt TAD boundaries and/or CTCF binding sites have been associated with aberrant gene activation in several cancer types.<sup>32–35</sup> The DNA hypermethylation associated with various cancer types can also disrupt CTCF binding and thereby weaken TAD boundaries.<sup>36–38</sup> We previously documented widespread loss of CTCF insulators in hypermethylated isocitrate dehydrogenase (IDH)-mutant gliomas and identified a specific CTCF insulator whose disruption caused an enhancer to aberrantly contact and activate the canonical brain tumor oncogene *PDGFRA*.<sup>36,39</sup> We identified a similar phenomenon in gastrointestinal stromal tumors (GISTs) wherein insulator disruption activates fibroblast growth factor (FGF) oncogenes.<sup>37</sup> In contrast to these striking individual examples, perturbation studies that used genome editing to disrupt multiple CTCF insulators suggest that a relatively small proportion of boundaries functionally impact gene expression in any given cellular context.<sup>40–43</sup> A critical goal is now to gain a mechanistic understanding of how these specific insulator aberrations can have such profound effects on gene activity.

GISTs are a relatively common form of sarcoma typically driven by gain-of-function mutations affecting the *KIT* or *PDGFRA* oncogenes.<sup>44</sup> However, ~10% of the GISTs are driven by loss of the succinate dehydrogenase (SDH) complex. This subtype does not respond to standard-of-care therapy, highlighting an urgent need for new understanding and therapeutic strategies. Succinate overload in SDH-deficient tumors inhibits TET family DNA demethylases and results in profound DNA hypermethylation.<sup>45</sup> Hundreds of CTCF-binding sites are disrupted by the hypermethylation, including sites within an insulator that protects the developmental regulators and proto-oncogenes *FGF3* and *FGF4*. Disruption of this insulator by methylation is a recurrent event in SDH-deficient GIST associated with strong induction of oncogenic FGF ligands.<sup>37</sup> Moreover, the insulator may also be disrupted by genetic rearrangement, as documented in a very small cohort of GISTs.<sup>46</sup> Beyond its clinical implications, this profoundly responsive locus has potential to provide key insights into how TADs and their boundaries impact gene regulation. However, we neither yet understand which of the many CTCF sites in the locus direct boundary formation nor how their disruption impacts locus topology, enhancer-promoter communication, or transcriptional activity.

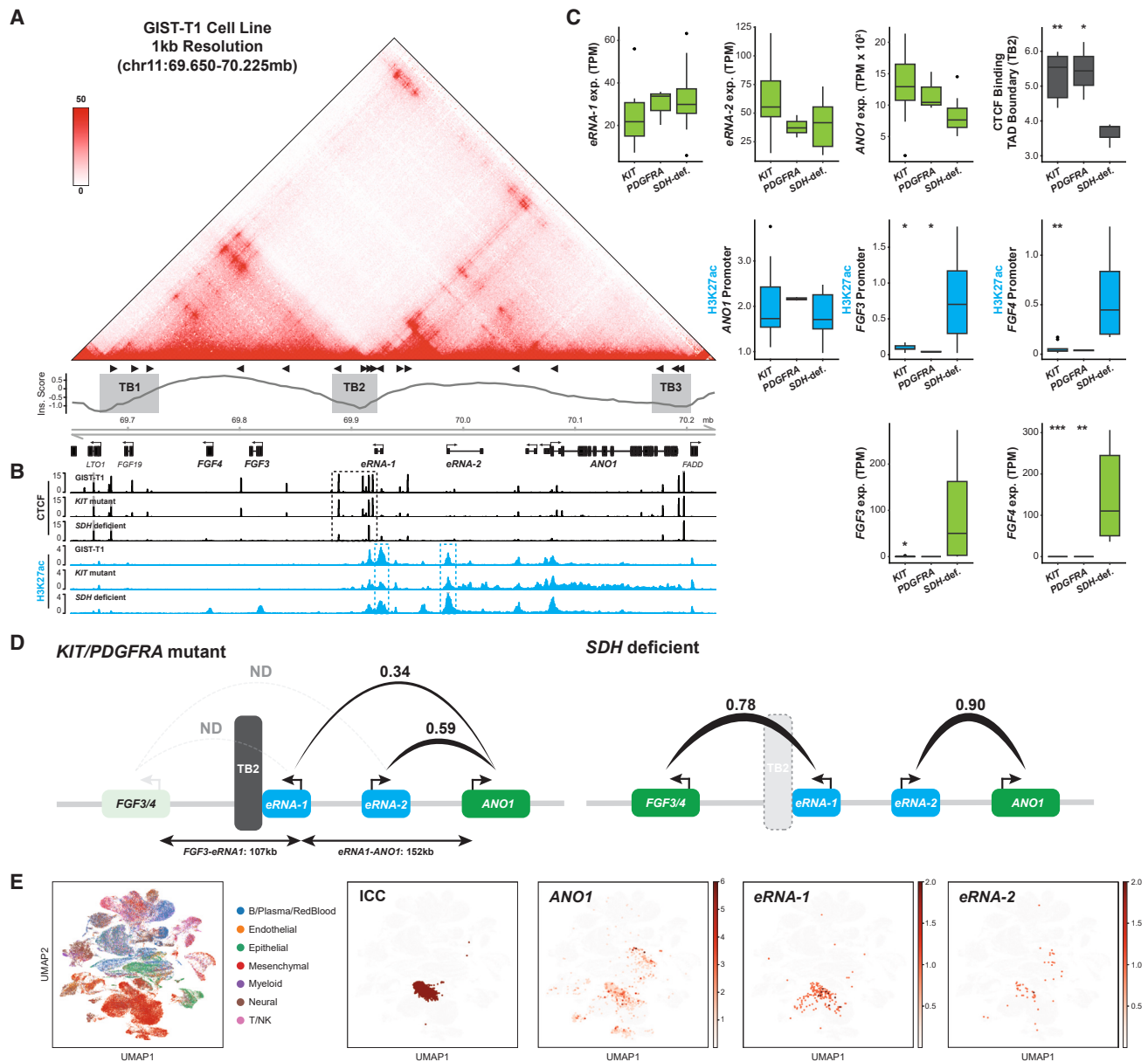
Here, we characterized the structure and function of the *FGF3/4* locus and the boundary element that insulates the region from an enhancer in the adjacent TAD. By integrating chromosome conformation mapping with combinatorial CRISPR-Cas9 perturbations in a GIST cell line, we identify specific CTCF binding sites within the insulator that enforce the boundaries of two adjacent TADs, one containing *FGF3* and *FGF4* and the other containing a potent enhancer active in GIST and its presumed cell of origin. We used region-capture micro-C (RCMC)<sup>9</sup> to quantify locus topology across GIST cell line derivatives with different CTCF disruptions and relate enhancer-promoter contact frequency to transcription. We find that a minimum of 4 CTCF sites must be disrupted to disable the insulator function, thus fusing the adjacent TADs to induce *FGF3*. Remarkably, micro-C measurements indicated that contact frequency between the enhancer and *FGF3* promoter scales linearly with *FGF3* transcriptional induction. In contrast, *FGF4* did not contact the enhancer and remained inactive. In conclusion, we deeply characterized a topological boundary that insulates potent oncogenes from a nearby enhancer, documented an array of redundant CTCF insulators as critical for its function, and provided direct evidence that enhancer-promoter contacts govern transcriptional activity of a key developmental regulator.

## RESULTS

### Topological organization and regulatory landscape of the *FGF3-FGF4* locus

To investigate how locus topology impacts transcription of the *FGF3* and *FGF4* proto-oncogenes, we used RCMC to generate a 1 kb-resolution contact map of the locus in GIST-T1 cells (Figure 1A). We acquired 12.2 million reads to map a 1.0 Mb region that contains genes encoding FGF ligands, as well as *ANO1*, which encodes a chloride channel expressed in all GIST subtypes. For perspective, the resolution of our RCMC map is equivalent to a genome-wide contact map with ~40 billion reads. Visual inspection of the contact map revealed that the FGF genes are contained within an ~200 kb TAD marked by high levels of self-interacting chromatin (Figure 1A, red heat in the contact map). An adjacent TAD contains *ANO1* and two enhancers that harbor histone 3 lysine 27 acetylation (H3K27ac) and transcribe enhancer-associated RNAs (eRNAs) in GIST tumors and the GIST-T1 cell line. The respective TADs are separated by a robust boundary (TB2) that is evident as a break between the regions of high self-interaction and confirmed quantitatively by insulation scores<sup>47</sup> derived from the contact map (Figure 1A). The boundary and overall topology of the locus are highly conserved across Hi-C and micro-C maps for diverse tissue and cell types (Figures S1A and S1B).

The central TB2 boundary contains four strongly bound CTCF binding sites whose motif orientations are informative<sup>48</sup> (Figure 1A). CTCF is understood to establish TAD boundaries by halting the extrusion of chromatin through cohesin. The extruded chromatin joins the region of high self-interaction, whereas the remainder is excluded or “insulated” from the TAD. Importantly, this effect is orientation dependent, in that a CTCF site must be oriented toward the TAD to halt extrusion and create a boundary.<sup>49–51</sup> Of the four main CTCF sites in TB2, one is oriented



**Figure 1. FGF locus topology and regulatory landscape**

(A) RCMC heatmap depicts contact frequency between genomic positions across a 575 kb interval that includes genes encoding FGF ligands and the GIST biomarker ANO1. The region contains two large TADs (red triangles) flanked by boundaries (TB1–3) that are evident as dips in the insulation score metric. (B) CTCF (black) and H3K27ac (blue) profiles are shown for GIST-T1 cells and for clinical samples corresponding to different GIST subtypes (ChIP-seq data are reads per million [RPM] normalized). Multiple CTCF sites in the TB2 boundary (black-dashed box) that separates the FGF genes from the H3K27ac-marked enhancers that encode eRNAs (blue dashed box). (C) Boxplots depict eRNA and ANO1 expression (top, green), CTCF binding (top, gray), promoter acetylation (middle), and FGF expression (bottom) in the respective GIST subtypes (data from GEO: GSE107447<sup>37</sup>). Two-sided t test *p* values, \**p* < 0.05, \*\**p* < 0.01, \*\*\**p* < 0.001. TPM, transcripts per million. (D) Schematic indicates correlations between eRNA expression, ANO1 expression, and averaged FGF3 and FGF4 expression across multiple GISTs (13 SDH-proficient, 6 SDH-deficient). The strong correlation between eRNA-1 and the FGF genes in SDH-deficient GISTs implicates the enhancer in the activation of these oncogenes. ND, not determined. (E) Uniform manifold approximation and projection (UMAP) visualization of single-cell RNA-seq data for gastrointestinal tissue highlights a cluster of ICC that highly express ANO1 and eRNA-1. See also Figure S1.

toward the FGF TAD, whereas the other three are oriented toward the TAD containing the enhancers and *ANO1*. This suggests that the single TB2 element actually encodes two boundaries—one for the FGF TAD and one for the *ANO1* TAD—through multiple, potentially redundant CTCF sites.

Comparison of 6 SDH-deficient and 13 SDH-proficient GIST samples indicates that CTCF binding is reduced in SDH-deficient GISTs (Figures 1B and 1C), consistent with our prior study.<sup>37</sup> Overall CTCF occupancy across TB2 is ~34% lower in the hypermethylated SDH-deficient tumors. We also examined the two enhancers in the *ANO1* TAD. We found that both elements transcribe eRNAs in all three GIST subtypes, a feature that is suggestive of strong activity.<sup>52</sup> Using these eRNAs as surrogates for enhancer activity, we were able to correlate enhancer activity with *FGF3* and *FGF4* transcription across GIST samples using RNA sequencing (RNA-seq) data. Neither enhancer correlated with these genes in *KIT*-mutant or *PDGFRA*-mutant GIST, which retain TB2. However, the more proximal enhancer (*eRNA-1*) strongly correlated with *FGF3/4* expression in the SDH-deficient tumors (Figures 1D and S1C). By contrast, transcription from the distal enhancer (*eRNA-2*) was lower in tumors with high *FGF3/4* expression. We speculate that this reflects a switch from a baseline state in which *eRNA-1* interacts with *eRNA-2* and *ANO1* to an alternate state in which *eRNA-1* preferentially interacts with *FGF3/4*, with the latter state primarily occurring in cells with a disrupted insulator. This suggests that the proximal enhancer may communicate with and potentially direct the robust induction of *FGF3* and *FGF4* in hypermethylated GIST tumors that have lost TB2.

The nomination of an enhancer element and *eRNA-1* as correlates and potential drivers of oncogene activation during tumorigenesis could provide insight into the cell of origin in SDH-deficient GIST. We therefore evaluated the expression of this eRNA systematically across gastrointestinal tissues using published single-cell RNA-seq data.<sup>53</sup> Clustering of these expression data distinguished a spectrum of gastrointestinal cell types, a small subset of which expressed *eRNA-1* transcript at levels ranging from 2 to 2.5 normalized gene counts in the single cells (Figure 1E). The most prominent cluster of *eRNA-1* expressing cells corresponded to the interstitial cell of Cajal (ICC), a specialized cell type with pacemaker functions in healthy gastrointestinal tissue that is a proposed cell of origin for GIST.<sup>44</sup> *eRNA-1* and the nearby *ANO1* gene were both selectively expressed in the cluster of cells annotated as ICC. Hence, disruption of TB2 could represent a truncal driver event that causes the proximal enhancer to activate FGF oncogenes in ICC or a closely related pre-malignant cell type.

### Experimental modeling of combinatorial CTCF site disruptions

We next sought to validate the insulatory function of TB2 and pinpoint the specific CTCF sites whose disruption underlies the induction of FGF genes in SDH-deficient GIST. We used CRISPR-Cas9 editing to disrupt different combinations of CTCF sites in the locus, prioritizing the TB2 sites most frequently lost in tumors (Figures 2A and S2A). We engineered 8 derivatives of GIST-T1, a human cell line derived from a *KIT*-mutant GIST with intact SDH expression. We verified that the parental GIST-

T1 cells retain a GIST-like enhancer landscape (Figure 1B) and that the FGF boundary remains intact and hypomethylated in TB2, with robust CTCF binding and physiologic locus topology (Figures S2A and S2B). We then confirmed efficient and selective editing of the targeted CTCF motifs in the respective GIST-T1 derivatives. Quantitation of transcript levels by real-time PCR revealed robust induction of *FGF3* in five of the eight derivatives ( $\Delta$ InsD/E/F/G/H) relative to non-targeting (NT) control (Figure 2B). A common thread across these five derivatives was the disruption of all four CTCF binding sites in TB2. *FGF4* was not strongly induced in any of the cell line derivatives, which suggests that the transcription of this neighboring gene is affected by additional conformational or epigenetic factors (Figure S2C).

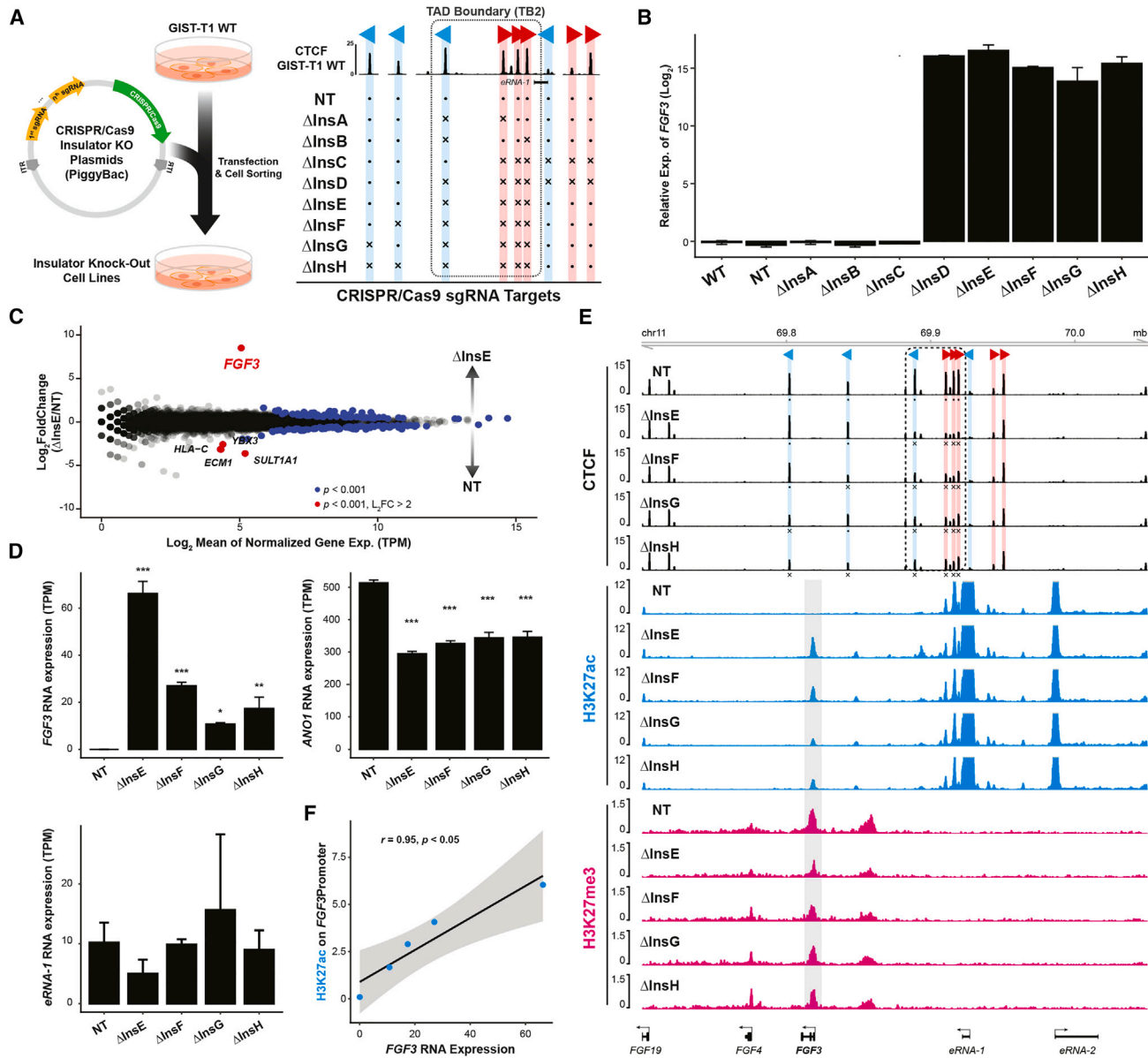
To evaluate further the specificity of induction, we used RNA-seq to query gene expression changes in four responsive derivatives,  $\Delta$ InsE/F/G/H, relative to control. Strikingly, despite the robust induction of *FGF3*, we did not identify any other genes that were significantly regulated by TB2 disruption (Figures 2C, S2D, and S2E). Within the locus, *ANO1* and *eRNA-2* were weakly downregulated (20%–30%) in these derivatives, whereas *eRNA-1* and other transcripts remained stable (Figure 2D).

Further support for the specificity of *FGF3* induction emerged from an analysis of histone modifications across the locus by chromatin immunoprecipitation sequencing (ChIP-seq). In unedited GIST-T1 cells and *KIT* mutant GIST tumors, the FGF TAD is devoid of activating H3K27ac chromatin (Figure 1B). Rather, it is diffusely enriched for histone 3 lysine 27 trimethylation (H3K27me3), a Polycomb-associated modification that silences key developmental loci in lineages where they are not required (Figure S2F). The *ANO1* TAD showed the opposite pattern with strong H3K27ac signals over the two eRNA-producing enhancers and essentially no H3K27me3. These chromatin patterns are entirely consistent with the baseline transcriptional activity of the respective TADs in the GIST subtypes and the cell line, which retain TB2.

However, these patterns were markedly changed in the GIST-T1 derivatives with combinatorial CTCF deletions and *FGF3* expression. We specifically observed a striking increase in H3K27ac over the *FGF3* promoter, consistent with the transcriptional activation and potentially indicative of enhancer-promoter contact (Figure 2E). H3K27me3 levels were decreased over the induced *FGF3* promoter. Comparing across the different GIST-T1 derivatives, we found that the amount of H3K27ac increase over the *FGF3* promoter closely correlated with the degree of *FGF3* induction ( $r \sim 0.95$ ; Figure 2F). Notably, the acetylation patterns in the GIST-T1 derivatives mirror those in SDH-deficient GIST specimens in which the boundary is disrupted by methylation (Figures 1B and 1C). These experimental models and analyses pinpoint specific combinations of CTCF sites as critical for the insulatory function of TB2 and document transcriptional and acetylation changes suggestive of aberrant long-range enhancer-promoter interactions in insulator-disrupted cells.

### High-resolution contact maps reveal interconnected, CTCF-dependent topological loops

To understand the mechanisms by which these combinatorial CTCF disruptions activate *FGF3* while sparing *FGF4*, we turned to the high-resolution RCMC contact map for the GIST-T1 line



**Figure 2. Redundant CTCF sites in the TAD boundary mediate robust insulation**

(A) Schematic depicts combinatorial disruption of CTCF sites in the TAD boundary (TB2) region by CRISPR-Cas9 editing. Table lists engineered GIST-T1 derivatives in which different combinations of CTCF sites are disrupted (targeted sites indicated by "x") and the NT control.

(B) Relative *FGF3* mRNA expression (real-time qPCR) in the derivative lines. Two biologically independent replicates (p values [one-way ANOVA] < 0.001 for  $\Delta$ InsD/E/F/G/H relative to control).

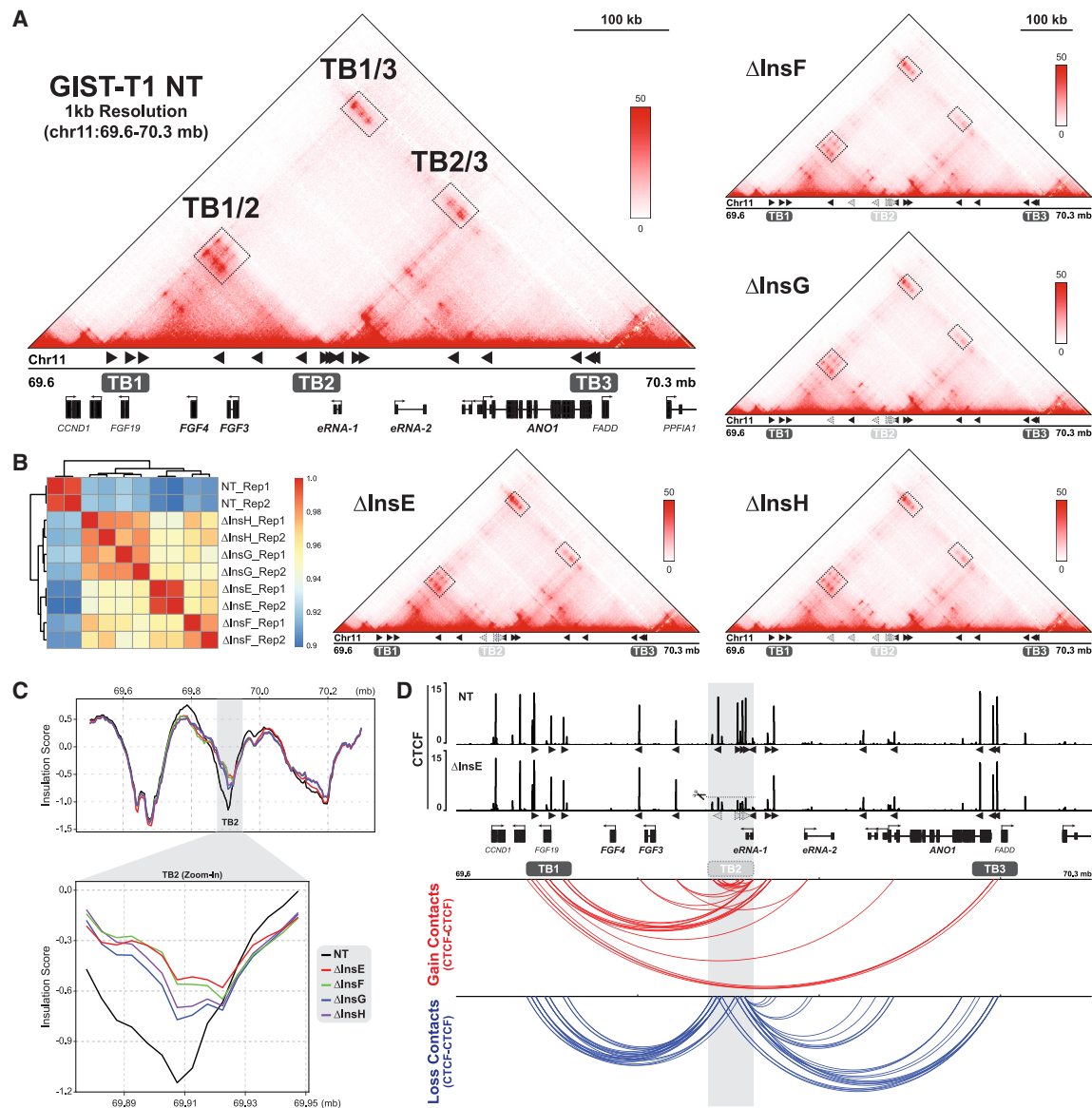
(C) Scatter plot depicts log<sub>2</sub> normalized RNA-seq expression (TPM) versus log<sub>2</sub> fold change in  $\Delta$ InsE (TPM + 1) compared with NT control (TPM + 1). It reveals specific upregulation of *FGF3* in the insulator-disrupted derivative.

(D) Barplots show *FGF3*, *eRNA-1*, and *ANO1* RNA-seq expression in NT and InsKO lines. Data represent two independent biological replicates (one-way ANOVA for  $\Delta$ InsD/E/F/G/H relative to control: \**p* < 0.05, \*\**p* < 0.01, \*\*\**p* < 0.001).

(E) CTCF (black), H3K27ac (blue), and H3K27me3 (magenta) profiles are shown for NT and InsKO lines over the genomic interval containing the FGF genes and eRNAs. Triangles indicate the CTCF motif orientation (red:sense; blue:antisense).

(F) Correlation of *FGF3* expression (RNA-seq) and H3K27ac at the *FGF3* promoter. RNA expression values in (C), (D), and (F) are TPM normalized. Error bars in (B) and (D) represent standard deviations. *r* indicates Pearson correlation coefficient. Shaded area represents confidence interval.

See also Figure S2.



**Figure 3. High-resolution contact maps reveal complex topological changes associated with combinatorial CTCF disruptions**

(A) RCMC contact maps shown for the FGF locus in control GIST-T1 (NT; as in Figure 1A) and derivatives (InsKO). Loop domain contacts between the boundaries flanking the two main TADs are indicated by dashed boxes.

(B) Heatmap depicts Pearson correlations between replicate RCMC profiles for the control and derivative lines.

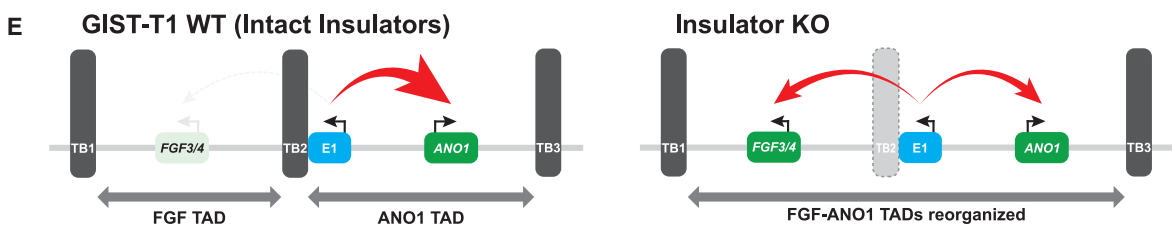
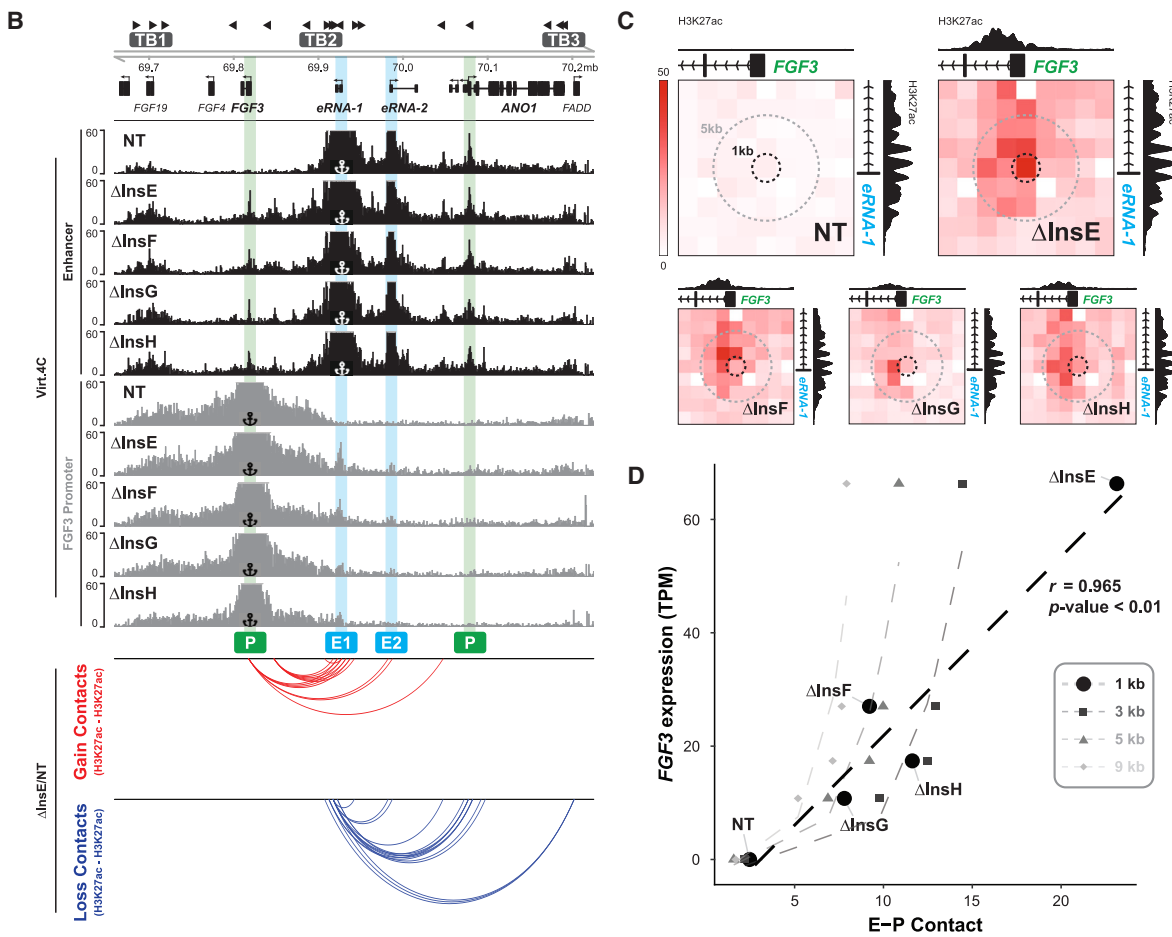
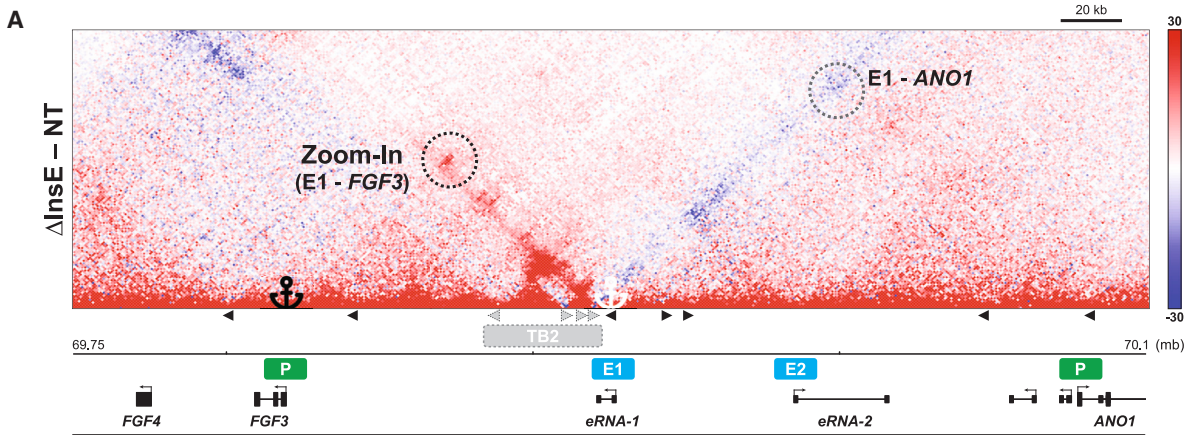
(C) Insulation scores are plotted for the control and derivative lines across a portion of the FGF locus including the TB2 boundary (gray bar), which is expanded below.

(D) CTCF profiles are shown for the FGF locus in control GIST-T1 and the ΔInsE derivative. Arc plots show differential contacts ( $p < 0.05$ ) between pairs of CTCF sites that are increased (red) or decreased (blue) in the ΔInsE derivative.

See also Figure S3.

(Figure 3A). In addition to delineating the neighboring FGF and ANO1 TADs, the contact map reveals long-range contacts between CTCF sites in the respective boundaries (TB1, TB2, and TB3). These long-range contacts reflect loop domain structures that organize the corresponding TADs. In particular, the contact map highlights multiple points of contact between CTCF sites in TB1 and TB2, which bound the FGF TAD, and between CTCF

sites in TB2 and TB3, which bound the ANO1 TAD. The contacts reflect states in which the respective boundaries are brought into close proximity by paused cohesin loops.<sup>17</sup> Consistent with a dynamic topology, long-range interactions spanning the entire FGF-ANO1 locus can also be appreciated in the heatmap (TB1-TB3). However, with an exception of this distant contact, physical interactions between the FGF and ANO1 TADs were



(legend on next page)



conspicuously absent in unedited GIST-T1 cells (parental or NT control).

We next used RCMC to generate high-resolution contact maps for GIST-T1 derivatives in which combinatorial disruption of CTCF sites in TB2 induced *FGF3* (Figure 3A). Although the overall topology of the locus remained largely intact ( $r > 0.9$ ), hierarchical clustering of the contact maps revealed consistent differences between the respective derivatives (Figure 3B). We evaluated the nature of these changes in several ways. First, we computed insulation scores across the locus by summing contacts spanning each position (Figures 3C and S3A). This showed that the insulation strength of TB2 was reduced in all four derivatives. Second, we compared pairwise contact frequencies across the locus to identify interactions that increased or decreased in the GIST-T1 derivative “ $\Delta$ InsE,” which showed the strongest induction of *FGF3* (Figure 3D). This analysis revealed a loss of interactions involving sites in the TB2 element, consistent with the loss of CTCF binding (Figures 3D, S3B, and S3C). Conversely, we observed an increase in “cross-TAD” interactions between sites in the FGF TAD and sites in the ANO1 TAD, consistent with the loss of TB2 boundary function (Figures 3D, S3B, and S3C). Notably, although cross-TAD interactions were increased in all four of the GIST-T1 derivatives, the magnitudes and patterns of change were highly variable (Figure S3D). This raised the possibility that variations in locus topology underlie the differences in *FGF3* transcription levels across these GIST-T1 derivatives.

### Contact frequency between *FGF3* and *eRNA-1* initiation sites predicts *FGF3* induction

To investigate specific topological changes that could mediate *FGF3* induction, we computed differential contact maps between control GIST-T1 cells and the derivatives (Figures 4A and S4A). This highlighted key changes in interactions involving the proximal enhancer E1 in the ANO1 TAD. The enhancer gained interactions with multiple sites across the FGF TAD in the InsE derivative that could be visualized as a stripe of increased contact frequency in the differential map (red). Importantly, this stripe extended only as far as the *FGF3* promoter and did not involve *FGF4* (Figure S4B). Conversely, the enhancer lost interactions with multiple sites across the ANO1 TAD in the InsE derivative that could be visualized as a stripe of reduced contact frequency in the differential map (blue). Importantly, the interac-

tions with the FGF TAD were highly specific to the derivative and essentially undetectable in control GIST-T1 cells. Hence, disruption of TB2 redirects the enhancer encoding *eRNA-1* toward the FGF TAD and the *FGF3* promoter.

To evaluate the interactions between enhancer and *FGF3* promoter more precisely, we imputed “4C” profiles from the RCMC data (Figure 4B). For each derivative and control GIST-T1 line, we imputed one profile for interactions made by the enhancer across the locus, and a second profile for interactions made by the *FGF3* promoter. In the control cells, the E1 enhancer contacted the *ANO1* gene and the second enhancer (E2) in the ANO1 TAD. In the derivatives, the E1 enhancer also contacted sites in the FGF3 TAD, with a sharp peak directly over the *FGF3* promoter. Consistent with this result, the 4C profiles imputed for *FGF3* revealed a contact peak over the E1 enhancer but did not reveal interactions with other sites in the ANO1 TAD (Figure S4C). Closer examination of the 4C profiles suggested two additional important points. First, the profiles indicated that the peak interaction occurs between the transcription start sites of *eRNA-1* and *FGF3*. Second, they indicated that contact frequency between these sites varied across the GIST-T1 derivatives.

Based on these findings, we hypothesized that physical interaction between these specific positions in the E1 enhancer and the *FGF3* promoter drives transcriptional induction of *FGF3*. As such, their contact frequency should predict differential *FGF3* levels across the derivatives. To test this, we computed contact frequency between positions centered over the *eRNA-1* transcription start site in the E1 enhancer and the *FGF3* transcription start site (Figure 4C). We initially quantified contact over 5 kb windows, based on prior convention,<sup>23</sup> and correlated the values against *FGF3* RNA expression across control and derivative lines (Figures 4D and S4D). Although this yielded a positive correlation, the relationship was nonlinear, as contact frequency appeared to reach a plateau in the highest *FGF3* expressors. We therefore leveraged the high-resolution RCMC data to quantify contacts for 1 kb windows directly centered on the *eRNA-1* and *FGF3* transcription start sites. Remarkably, these more precise interaction measurements correlated linearly with *FGF3* RNA expression ( $r = 0.97$ ,  $p < 0.01$ ; Figures 4D, S4E, and S4F). This striking linear correlation suggests that *FGF3* expression is precisely controlled by interactions with the unleashed E1 enhancer or, under more physiologic conditions, may scale

### Figure 4. Enhancer-promoter contact frequency scales linearly with *FGF3* transcriptional activation

(A) Heatmap depicts differential RCMC contact frequency between  $\Delta$ InsE and control GIST-T1 cells. Dashed circles indicate interactions between the *eRNA-1* transcriptional start site and the *FGF3* promoter (left) or the *ANO1* promoter (right). Gene tracks, promoter and enhancer regions, and CTCF (triangles) are indicated below. Anchors indicate the transcription start sites of *FGF3* and *eRNA-1*. Upon boundary disruption, *eRNA-1* gains interaction with *FGF3* (red heat in E1-*FGF3* dashed circle), but loses interaction with sites across the ANO1 TAD (blue stripe extending to the E1-*ANO1* dashed circle).  
(B) Virtual 4C tracks (inferred from RCMC) depict contact frequency between *eRNA-1* (top, black) or the *FGF3* promoter (bottom, gray) and all genomic positions across the FGF locus in control GIST-T1 (NT) and InsKO lines. Arc plots below show differential contacts (two-sided t test  $p$ -value  $< 0.05$ ) between H3K27ac sites that are increased (red) or decreased (blue) in the  $\Delta$ InsE derivative.  
(C) Heatmaps depict RCMC contact frequency between 10 kb windows centered over the transcription start sites of *FGF3* and *eRNA-1* in control GIST-T1 and the respective derivatives. Horizontal and vertical tracks flanking the heatmaps depict H3K27ac signal over the *FGF3* promoter and *eRNA-1*, respectively.  
(D) *FGF3* expression (RNA-seq) is plotted against contact frequency between the transcription start sites of *FGF3* and *eRNA-1* for control GIST-T1 and derivative lines. Contact frequencies and correlations were computed for different window sizes (symbols and lines). Transcription scales linearly with enhancer-promoter contact when calculated between start sites at 1 kb resolution.  $r$  indicates Pearson correlation coefficient.  
(E) Schematic summarizes the impact of TB2 disruption on TAD organization and re-targeting or “hijacking” of the key enhancer to activate FGF oncogenes. See also Figure S4.

with cumulative contact with the lineage-specific enhancers that regulate its expression during development (Figure 4E).

## DISCUSSION

Understanding the molecular mechanisms by which chromatin structure and genome topology regulate gene expression is a critical goal relevant to human biology and disease. Here, we drilled down on a recurrent epigenetic lesion that disrupts a boundary element to drive oncogene expression in GIST. We find that redundant CTCF sites in the boundary enforce adjacent TADs that robustly insulate the FGF proto-oncogenes, and we chart topological alterations incurred upon boundary disruption that allow aberrant enhancer-promoter interactions and transcriptional induction. Our study exemplifies how disease-associated epigenetic or genetic alterations can provide fundamental insights into genome organization, boundary elements, and their impact on gene regulation.

First, we find that the FGF boundary element achieves robustness through redundancy. Multiple CTCF sites within the element enforce the boundaries of two adjacent TADs, which, respectively, ensconce the FGF genes and the *ANO1* enhancer. Combinatorial editing of the cognate-binding motifs revealed that any one of the four central CTCF sites is sufficient to maintain insulation of the proto-oncogenes. The sufficiency of even a single CTCF site to prevent enhancer hijacking is consistent with a model in which the coordinate extrusion of enhancer and promoter through cohesin incurs the long-range physical interactions required for transcription.<sup>12,13</sup> The presence of even a single intervening CTCF site would then suffice to exclude either enhancer or promoter from the extruding loop and thereby prevent contact and transcriptional induction. In the absence of extrusion-driven co-localization, diffusion-limited enhancer-promoter contacts are presumably too rare to activate transcription due to the significant distance between *FGF3* and the enhancer. Our results are consistent with the linearity of the activity-by-contact (ABC) model<sup>26,27</sup> and help explain the remarkable robustness of the boundary.

Second, we gained key insights into how boundary destabilization translates to transcriptional activation by analyzing high-resolution contact maps for derivative GIST lines lacking different combinations of CTCF sites. Although the FGF boundary is clearly compromised in all of the derivatives with transcriptional induction of *FGF3*, the respective TADs remain partially intact. Importantly, we found that the different combinatorial CTCF disruptions incur subtly different conformational changes. We leveraged this variability to evaluate the relationship between conformational change and transcriptional activity. We found that contact frequency between enhancer and *FGF3* promoter correlated with *FGF3* induction. Importantly, the correlation was strongest and remarkably linear when we focused precisely on contact between 1 kb intervals corresponding to the *eRNA-1* transcriptional start site and the *FGF3* transcriptional start site. This analysis was only made possible by the high-resolution nature of the RCMC contact maps.<sup>9</sup> The linear association was lost when we measured interactions between 5 kb windows, which has been the standard in prior analyses.<sup>23</sup> This supports a model in which contact between transcriptional complexes in enhancer

and promoter, which may be stabilized by biophysical interactions,<sup>54,55</sup> translates directly to activation of this oncogene. It also indicates that a relatively small change in TAD organization can lead to remarkably strong gene induction, provided it increases physical enhancer-promoter communication.

Third, our study provides insights into disease mechanisms. Having pinpointed a specific enhancer as a driver of aberrant FGF expression in SDH-deficient GIST, we used the associated eRNA as a surrogate to track its activity across gastrointestinal cell types using single-cell data. The eRNA was strongly and specifically expressed in the ICC, a candidate cell of origin for GIST.<sup>44</sup> However, the enhancer should not impact the FGF genes under physiologic conditions as the boundary is constitutive across tissues and lineages. The boundary should also be resilient to point mutations or short deletions, given its multiple redundant CTCF sites. However, it is vulnerable to DNA hypermethylation, which spreads across the element and displaces multiple CTCF sites in SDH-deficient GIST. Genetic rearrangements that effectively displace the entire boundary have also been described<sup>29</sup> but are extremely rare. Hence, epigenetic or genetic disruption of this otherwise constitutive boundary in a specific cellular context causes FGF induction and a proliferative advantage that drives tumorigenesis. We acknowledge that this boundary element and the impact of its loss may not be representative of TAD boundaries. It was identified because of its potent effect on FGF expression and tumorigenesis. Moreover, its regulatory impact is highly context dependent and might only be appreciated in studies of GIST, the rarified ICC, or potentially specific neural cell types that also express *ANO1*. Indeed, we note that studies of other TAD boundaries have tended to observe more subtle effects on neighboring gene expression.

In conclusion, we have deeply characterized the determinants and impact of a uniquely robust boundary element that insulates FGF ligand genes, which represent key developmental regulators and proto-oncogenes. Our results explain why DNA hypermethylation is uniquely suited for the coordinate disruption of the redundant CTCF sites that enforce the boundary and document activity of the relevant enhancer in a likely cell of origin for GIST. Finally, our high-resolution contact maps of derivative lines with compromised insulation provide critical support for a model in which long-range transcriptional activation is driven by direct physical interaction between initiation sites in enhancer and promoter.

## Limitations of the study

We note several limitations of our study. Although SDH-deficient GISTs can express *FGF3* and *FGF4*, our CTCF perturbations specifically activate *FGF3*. The pattern of activation is consistent with the observation that *FGF4* does not physically interact with the enhancer or any other position in the *ANO1* TAD. Nonetheless, we cannot explain why the aberrant contacts unleashed upon FGF boundary disruption failed to extend to *FGF4*, even when an additional CTCF site positioned between *FGF3* and *FGF4* was disrupted. One possibility is that partial methylation of the *FGF3* promoter, which is evident in SDH-deficient GISTs (Figure S4G), may redirect the enhancer toward *FGF4* on some alleles in these tumors. A further limitation is that although

RCMC enables contact mapping with remarkable resolution and precision, it nonetheless provides static, averaged measurements of a highly dynamic process. We can neither track loop extrusion or stabilization over time nor can we evaluate the temporal relationship between enhancer-promoter contacts and transcription. We speculate that a cascade of conformational changes incurred upon boundary disruption alters boundary structure and TAD organization, thereby affecting enhancer-promoter contact and *FGF3* transcription. Finally, although our perturbations and high-resolution contact maps are supportive of a causal role for physical enhancer-promoter contact, the possibility that contact frequency is affected by transcription or does not linearly control transcription cannot be ruled out. Future studies and new technologies are needed to address critical issues regarding the kinetics of loop extrusion, TAD function, and the causality of long-range enhancer contact.

### STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Cell cultures
- METHOD DETAILS
  - Insulator knockout cell line
  - Quantitative real-time PCR
  - RNA-seq
  - Chromatin immunoprecipitation
  - Locus bisulfite sequencing
  - Region Capture Micro-C
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - CTCF motif analysis
  - RNA-seq analysis
  - ChIP-seq analysis
  - Mapping and normalizing RCMC
  - Chromatin contact analysis
  - Analysis of ENCODE, 4DN, Gut Cell Atlas data
  - Analysis of scRNA-seq data
  - Group comparisons

### SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.molcel.2024.02.007>.

### ACKNOWLEDGMENTS

We thank members of the Bernstein and Hansen labs and the Gene Regulation Observatory for thoughtful discussions and feedback. This work was supported by funds from the NCI/NIH Director's Fund (DP1CA216873 to B.E.B.), the NIH New Innovator Award (DP2GM140938 to A.S.H.), the Gene Regulation Observatory at the Broad Institute, and the Bertarelli Foundation at Harvard Medical School.

### AUTHOR CONTRIBUTIONS

K.L.K., A.S.H., and B.E.B. developed the concepts and designed the study. K.L.K. and G.J.R. performed the experiments. K.L.K., V.Y.G., and C.A.E.F. performed bioinformatic analysis. K.L.K., G.J.R., and B.E.B. wrote the manuscript with feedback from all authors. B.E.B. supervised the project.

### DECLARATION OF INTERESTS

B.E.B. declares outside interests in Fulcrum Therapeutics, HiFiBio, Arsenal Biosciences, Chroma Medicine, Cell Signaling Technologies, and Design Pharmaceuticals. V.Y.G. and A.S.H. have filed a provisional patent on RCMC.

Received: October 3, 2023

Revised: January 3, 2024

Accepted: February 8, 2024

Published: March 6, 2024

### REFERENCES

1. Furlong, E.E.M., and Levine, M. (2018). Developmental enhancers and chromosome topology. *Science* 361, 1341–1345.
2. Cramer, P. (2019). Organization and regulation of gene transcription. *Nature* 573, 45–54.
3. Heinz, S., Romanoski, C.E., Benner, C., and Glass, C.K. (2015). The selection and function of cell type-specific enhancers. *Nat. Rev. Mol. Cell Biol.* 16, 144–154.
4. Rozowsky, J., Gao, J., Borsari, B., Yang, Y.T., Galeev, T., Gürsoy, G., Epstein, C.B., Xiong, K., Xu, J., Li, T., et al. (2023). The EN-TE resource of multi-tissue personal epigenomes & variant-impact models. *Cell* 186, 1493–1511.e40.
5. Schoenfelder, S., and Fraser, P. (2019). Long-range enhancer-promoter contacts in gene expression control. *Nat. Rev. Genet.* 20, 437–455.
6. Preissl, S., Gaulton, K.J., and Ren, B. (2023). Characterizing cis-regulatory elements using single-cell epigenomics. *Nat. Rev. Genet.* 24, 21–43.
7. Hildebrand, E.M., and Dekker, J. (2020). Mechanisms and functions of chromosome compartmentalization. *Trends Biochem. Sci.* 45, 385–396.
8. Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326, 289–293.
9. Goel, V.Y., Huseyin, M.K., and Hansen, A.S. (2023). Region Capture Micro-C reveals coalescence of enhancers and promoters into nested microcompartments. *Nat. Genet.* 55, 1048–1056.
10. Hughes, J.R., Roberts, N., McGowan, S., Hay, D., Giannoulidou, E., Lynch, M., De Gobbi, M., Taylor, S., Gibbons, R., and Higgs, D.R. (2014). Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment. *Nat. Genet.* 46, 205–212.
11. Kempfer, R., and Pombo, A. (2020). Methods for mapping 3D chromosome architecture. *Nat. Rev. Genet.* 21, 207–226.
12. Ganji, M., Shaltiel, I.A., Bish, S., Kim, E., Kalichava, A., Haering, C.H., and Dekker, C. (2018). Real-time imaging of DNA loop extrusion by condensin. *Science* 360, 102–105.
13. Davidson, I.F., and Peters, J.-M. (2021). Genome folding through loop extrusion by SMC complexes. *Nat. Rev. Mol. Cell Biol.* 22, 445–464.
14. Hafner, A., and Boettiger, A. (2023). The spatial organization of transcriptional control. *Nat. Rev. Genet.* 24, 53–68.
15. Chakraborty, S., Kopitchinski, N., Zuo, Z., Eraso, A., Awasthi, P., Chari, R., Mitra, A., Tobias, I.C., Moorthy, S.D., Dale, R.K., et al. (2023). Enhancer-promoter interactions can bypass CTCF-mediated boundaries and contribute to phenotypic robustness. *Nat. Genet.* 55, 280–290.
16. Ren, G., Jin, W., Cui, K., Rodriguez, J., Hu, G., Zhang, Z., Larson, D.R., and Zhao, K. (2017). CTCF-mediated enhancer-promoter interaction is a

- critical regulator of cell-to-cell variation of gene expression. *Mol. Cell* 67, 1049–1058.e6.
17. Gabriele, M., Brandão, H.B., Grosse-Holz, S., Jha, A., Dailey, G.M., Cattoglio, C., Hsieh, T.-H.S., Mirny, L., Zechner, C., and Hansen, A.S. (2022). Dynamics of CTCF- and cohesin-mediated chromatin looping revealed by live-cell imaging. *Science* 376, 496–501.
  18. Mach, P., Kos, P.I., Zhan, Y., Cramard, J., Gaudin, S., Tünnermann, J., Marchi, E., Eglinger, J., Zuin, J., Kryzhanovska, M., et al. (2022). Cohesin and CTCF control the dynamics of chromosome folding. *Nat. Genet.* 54, 1907–1918.
  19. Alexander, J.M., Guan, J., Li, B., Maliskova, L., Song, M., Shen, Y., Huang, B., Lomvardas, S., and Weiner, O.D. (2019). Live-cell imaging reveals enhancer-dependent transcription in the absence of enhancer proximity. *eLife* 8, e41769.
  20. Heist, T., Fukaya, T., and Levine, M. (2019). Large distances separate coregulated genes in living embryos. *Proc. Natl. Acad. Sci. USA* 116, 15062–15067.
  21. Mateo, L.J., Murphy, S.E., Hafner, A., Cinquini, I.S., Walker, C.A., and Boettiger, A.N. (2019). Visualizing DNA folding and RNA in embryos at single-cell resolution. *Nature* 568, 49–54.
  22. Benabdallah, N.S., Williamson, I., Illingworth, R.S., Kane, L., Boyle, S., SenGupta, D., Grimes, G.R., Therizols, P., and Bickmore, W.A. (2019). Decreased enhancer-promoter proximity accompanying enhancer activation. *Mol. Cell* 76, 473–484.e7.
  23. Zuin, J., Roth, G., Zhan, Y., Cramard, J., Redolfi, J., Piskadlo, E., Mach, P., Kryzhanovska, M., Tihanyi, G., Kohler, H., et al. (2022). Nonlinear control of transcription through enhancer-promoter interactions. *Nature* 604, 571–577.
  24. Xiao, J.Y., Hafner, A., and Boettiger, A.N. (2021). How subtle changes in 3D structure can create large changes in transcription. *eLife* 10, e64320.
  25. Tsujimura, T., Takase, O., Yoshikawa, M., Sano, E., Hayashi, M., Hoshi, K., Takato, T., Toyoda, A., Okano, H., and Hishikawa, K. (2020). Controlling gene activation by enhancers through a drug-inducible topological insulator. *eLife* 9, e47980.
  26. Fulco, C.P., Nasser, J., Jones, T.R., Munson, G., Bergman, D.T., Subramanian, V., Grossman, S.R., Anyoha, R., Doughty, B.R., Patwardhan, T.A., et al. (2019). Activity-by-contact model of enhancer-promoter regulation from thousands of CRISPR perturbations. *Nat. Genet.* 51, 1664–1669.
  27. Nasser, J., Bergman, D.T., Fulco, C.P., Guckelberger, P., Doughty, B.R., Patwardhan, T.A., Jones, T.R., Nguyen, T.H., Ulirsch, J.C., Lekschas, F., et al. (2021). Genome-wide enhancer maps link risk variants to disease genes. *Nature* 593, 238–243.
  28. Karr, J.P., Ferrie, J.J., Tjian, R., and Darzacq, X. (2022). The transcription factor activity gradient (TAG) model: contemplating a contact-independent mechanism for enhancer-promoter communication. *Genes Dev.* 36, 7–16.
  29. Lupiáñez, D.G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., Horn, D., Kayserili, H., Opitz, J.M., Laxova, R., et al. (2015). Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* 161, 1012–1025.
  30. Despang, A., Schöpflin, R., Franke, M., Ali, S., Jerković, I., Paliou, C., Chan, W.L., Timmermann, B., Wittler, L., Vingron, M., et al. (2019). Functional dissection of the *Sox9-Kcnj2* locus identifies nonessential and instructive roles of TAD architecture. *Nat. Genet.* 51, 1263–1271.
  31. Symmons, O., Pan, L., Remeseiro, S., Aktas, T., Klein, F., Huber, W., and Spitz, F. (2016). The *shh* topological domain facilitates the action of remote enhancers by reducing the effects of genomic distances. *Dev. Cell* 39, 529–543.
  32. Hnisz, D., Weintraub, A.S., Day, D.S., Valton, A.-L., Bak, R.O., Li, C.H., Goldmann, J., Lajoie, B.R., Fan, Z.P., Sigova, A.A., et al. (2016). Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science* 351, 1454–1458.
  33. Guo, Y.A., Chang, M.M., Huang, W., Ooi, W.F., Xing, M., Tan, P., and Skanderup, A.J. (2018). Mutation hotspots at CTCF binding sites coupled to chromosomal instability in gastrointestinal cancers. *Nat. Commun.* 9, 1520.
  34. Liu, E.M., Martinez-Fundichely, A., Diaz, B.J., Aronson, B., Cuykendall, T., MacKay, M., Dhingra, P., Wong, E.W.P., Chi, P., Apostolou, E., et al. (2019). Identification of cancer drivers at CTCF insulators in 1,962 whole genomes. *Cell Syst.* 8, 446–455.e8.
  35. Debaugny, R.E., and Skok, J.A. (2020). CTCF and CTCFL in cancer. *Curr. Opin. Genet. Dev.* 61, 44–52.
  36. Flavahan, W.A., Drier, Y., Liao, B.B., Gillespie, S.M., Venteicher, A.S., Stemmer-Rachamimov, A.O., Suvà, M.L., and Bernstein, B.E. (2016). Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature* 529, 110–114.
  37. Flavahan, W.A., Drier, Y., Johnstone, S.E., Hemming, M.L., Tarjan, D.R., Hegazi, E., Shareef, S.J., Javed, N.M., Raut, C.P., Eschle, B.K., et al. (2019). Altered chromosomal topology drives oncogenic programs in SDH-deficient GISTs. *Nature* 575, 229–233.
  38. Johnstone, S.E., Reyes, A., Qi, Y., Adriaens, C., Hegazi, E., Pelka, K., Chen, J.H., Zou, L.S., Drier, Y., Hecht, V., et al. (2020). Large-scale topological changes restrain malignant progression in colorectal cancer. *Cell* 182, 1474–1489.e23.
  39. Rahme, G.J., Javed, N.M., Puorro, K.L., Xin, S., Hovestadt, V., Johnstone, S.E., and Bernstein, B.E. (2023). Modeling epigenetic lesions that cause gliomas. *Cell* 186, 3674–3685.e14.
  40. Tarjan, D.R., Flavahan, W.A., and Bernstein, B.E. (2019). Epigenome editing strategies for the functional annotation of CTCF insulators. *Nat. Commun.* 10, 4258.
  41. Williamson, I., Kane, L., Devenney, P.S., Flyamer, I.M., Anderson, E., Kilanowski, F., Hill, R.E., Bickmore, W.A., and Lettice, L.A. (2019). Developmentally regulated expression is robust to TAD perturbations. *Development* 146, dev179523.
  42. Amândio, A.R., Beccari, L., Lopez-Delisle, L., Mascrez, B., Zakany, J., Gitto, S., and Duboule, D. (2021). Sequential in mutagenesis in vivo reveals various functions for CTCF sites at the mouse cluster. *Genes Dev.* 35, 1490–1509.
  43. Ghavi-Helm, Y., Jankowski, A., Meiers, S., Viales, R.R., Korbel, J.O., and Furlong, E.E.M. (2019). Highly rearranged chromosomes reveal uncoupling between genome topology and gene expression. *Nat. Genet.* 51, 1272–1282.
  44. Blay, J.-Y., Kang, Y.-K., Nishida, T., and von Mehren, M. (2021). Gastrointestinal stromal tumours. *Nat. Rev. Dis. Primers* 7, 22.
  45. Killian, J.K., Kim, S.Y., Miettinen, M., Smith, C., Merino, M., Tsokos, M., Quezado, M., Smith, W.I., Jr., Jahromi, M.S., Xekouki, P., et al. (2013). Succinate dehydrogenase mutation underlies global epigenomic divergence in gastrointestinal stromal tumor. *Cancer Discov.* 3, 648–657.
  46. Urbini, M., Astolfi, A., Indio, V., Nannini, M., Schipani, A., Bacalini, M.G., Angelini, S., Ravegnini, G., Calice, G., Del Gaudio, M., et al. (2020). Gene duplication, rather than epigenetic changes, drives FGF4 overexpression in KIT/PDGFR $\alpha$ /SDH/RAS-P WT GIST. *Sci. Rep.* 10, 19829.
  47. Crane, E., Bian, Q., McCord, R.P., Lajoie, B.R., Wheeler, B.S., Ralston, E.J., Uzawa, S., Dekker, J., and Meyer, B.J. (2015). Condensin-driven remodelling of X chromosome topology during dosage compensation. *Nature* 523, 240–244.
  48. Rao, S.S.P., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., et al. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 159, 1665–1680.
  49. Pugacheva, E.M., Kubo, N., Loukinov, D., Tajmul, M., Kang, S., Kovalchuk, A.L., Strunnikov, A.V., Zentner, G.E., Ren, B., and Lobanov, V.V. (2020). CTCF mediates chromatin looping via N-terminal domain-dependent cohesin retention. *Proc. Natl. Acad. Sci. USA* 117, 2020–2031.

50. Li, Y., Haarhuis, J.H.I., Sedeño Cacciatore, Á., Oldenkamp, R., van Ruiten, M.S., Willems, L., Teunissen, H., Muir, K.W., de Wit, E., Rowland, B.D., et al. (2020). The structural basis for cohesin-CTCF-anchored loops. *Nature* 578, 472–476.
51. Nora, E.P., Caccianini, L., Fudenberg, G., So, K., Kameswaran, V., Nagle, A., Uebersohn, A., Hajj, B., Saux, A.L., Coulon, A., et al. (2020). Molecular basis of CTCF binding polarity in genome folding. *Nat. Commun.* 11, 5612.
52. Sartorelli, V., and Laubert, S.M. (2020). Enhancer RNAs are an important regulatory layer of the epigenome. *Nat. Struct. Mol. Biol.* 27, 521–528.
53. Elmentaite, R., Kumasaka, N., Roberts, K., Fleming, A., Dann, E., King, H.W., Kleshchevnikov, V., Dabrowska, M., Pritchard, S., Bolt, L., et al. (2021). Cells of the human intestinal tract mapped across space and time. *Nature* 597, 250–255.
54. Ohishi, H., Shinkai, S., Owada, H., Fujii, T., Hosoda, K., Onami, S., Yamamoto, T., Ohkawa, Y., and Ochiai, H. (2023). Transcription-coupled changes in higher-order genomic structure and transcription hub viscosity prolong enhancer-promoter connectivity. Preprint at bioRxiv. <https://doi.org/10.1101/2023.11.27.568629>.
55. Barshad, G., Lewis, J.J., Chivu, A.G., Abuhashem, A., Krietenstein, N., Rice, E.J., Ma, Y., Wang, Z., Rando, O.J., Hadjantonakis, A.-K., et al. (2023). RNA polymerase II dynamics shape enhancer-promoter interactions. *Nat. Genet.* 55, 1370–1380.
56. Taguchi, T., Sonobe, H., Toyonaga, S.-I., Yamasaki, I., Shuin, T., Takano, A., Araki, K., Akimaru, K., and Yuri, K. (2002). Conventional and molecular cytogenetic characterization of a new human cell line, GIST-T1, established from gastrointestinal stromal tumor. *Lab. Invest.* 82, 663–665.
57. Kramer, N.E., Davis, E.S., Wenger, C.D., Deoudes, E.M., Parker, S.M., Love, M.I., and Phanstiel, D.H. (2022). Plotgardener: cultivating precise multi-panel figures in R. *Bioinformatics* 38, 2042–2045.
58. van der Weide, R.H., van den Brand, T., Haarhuis, J.H.I., Teunissen, H., Rowland, B.D., and de Wit, E. (2021). Hi-C analyses with Genova: a case study with cohesin variants. *NAR Genom. Bioinform.* 3, lqab040.
59. Frankish, A., Diekhans, M., Jungreis, I., Lagarde, J., Loveland, J.E., Mudge, J.M., Sisu, C., Wright, J.C., Armstrong, J., Barnes, I., et al. (2021). GENCODE 2021. *Nucleic Acids Res.* 49, D916–D923.
60. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21.
61. Wolf, F.A., Angerer, P., and Theis, F.J. (2018). SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* 19, 15.

**STAR★METHODS**

**KEY RESOURCES TABLE**

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Antibodies</b>		
Rabbit anti CTCF (clone D31H2)	Cell Signaling Technologies	Cat# 3418, RRID:AB_2086791
Rabbit anti H3K27ac	Active motif	Cat# 39133, RRID:AB_2561016
Rabbit anti H3K27me3 (clone C36B11)	Cell Signaling Technologies	Cat# 9733, RRID:AB_2616029
<b>Bacterial and virus strains</b>		
NEB stable competent E. coli	New England Biolabs	C3040I
<b>Chemicals, peptides, and recombinant proteins</b>		
Agencourt AMPure XP	Beckman Coulter	A63882
LipoD293	Signagen	SL100668
Fetal Bovine Serum	GIBCO	26140079
DMEM, high glucose, GlutaMAX Supplement, pyruvate	GIBCO	10569010
Penicillin-Streptomycin	GIBCO	15140122
Formaldehyde	ThermoFisher	28906
DSG (disuccinimidyl glutarate)	ThermoFisher	20593
1 M Tris-HCl Buffer, pH 7.5	Invitrogen	15567027
NP-40 Substitute	Abcam	ab142227
Protease Inhibitor Cocktail (EDTA-free)	Sigma-Aldrich	5056489001
MNase (micrococcal nuclease)	New England BioLabs	M0247S
EGTA	BioWorld	40520008
BSA (bovine serum albumin)	Sigma-Aldrich	B8667
BSA, Molecular Biology Grade	New England BioLabs	B9000S
T4 PNK (polynucleotide kinase)	New England BioLabs	M0201
ATP	New England BioLabs	P0756
DTT (dithiothreitol)	Sigma-Aldrich	10197777001
DNA Polymerase I Klenow Fragment	New England BioLabs	M0210
dNTP set	Invitrogen	10297018
Biotin-14-dATP	Jena Bioscience	NU-835-BIO14
Biotin-11-dCTP	Jena Bioscience	NU-809-BIOX
EDTA	Invitrogen	15575020
T4 DNA Ligase	New England BioLabs	M0202
Exonuclease III	New England BioLabs	M0206
Proteinase K	New England BioLabs	P8107S
RNase A, DNase- and protease-free	ThermoFisher	EN0531
SDS (sodium dodecyl sulfate)	Invitrogen	AM9822
Agarose	VWR	97062
Dynabeads MyOne Streptavidin T1	Invitrogen	65601
Tween-20	Sigma-Aldrich	P8074
Multiplex Oligos for Illumina Primer Set 1	New England BioLabs	E7335
Q5 High-Fidelity DNA Polymerase	New England BioLabs	M0491
Capture Custom Panel	Twist Bioscience	101001
Standard Hybridization Mix	Twist Bioscience	104178
Universal Blockers	Twist Bioscience	100578
Streptavidin Binding Beads	Twist Bioscience	100983
Equinox Library Amplification Mix	Twist Bioscience	104178

(Continued on next page)

**Continued**

REAGENT or RESOURCE	SOURCE	IDENTIFIER
DNA Purification Beads	Twist Bioscience	100983
SYBR™ Green PCR Master Mix	Applied Biosystems	4309155
KAPA HiFi Uracil+ HotStart ReadyMix	KAPA	KK2800

**Critical commercial assays**

NextSeq 500/550 High Output Kit v2.5 (75 Cycles)	Illumina	20024906
Qubit dsDNA HS Assay kit	Invitrogen	Q32854
Qiagen Maxiprep plasmid kit	Qiagen	12163
Bioanalyzer D1000 screentape	Agilent	5067-5582
DNA Clean & Concentrator kit	Zymo Research	D4034
Gel Purification kit	Zymo Research	D4008
Quick Blunting Kit	New England BioLabs	E1201
NEBNext Ultra II DNA Library Prep Kit for Illumina	New England BioLabs	E7645
NEBNext Ultra II RNA Library Prep Kit for Illumina	New England BioLabs	E7770
NEBNext Poly(A) mRNA Magnetic Isolation Module	New England BioLabs	E7490
RNeasy Plus Kit	Qiagen	74134
ProtoScript II First Strand cDNA Synthesis Kit	New England BioLabs	E6560
Monarch Genomic DNA Purification Kit	New England BioLabs	T3010
EZ DNA Methylation-Lightning Kit	Zymo Research	D5030

**Deposited data**

Processed data	GEO	GSE241927
Raw data (ChIP-seq)	GEO	GSE241924
Raw data (Micro-C)	GEO	GSE241925
Raw data (RNA-seq)	GEO	GSE241926
GIST Patients <sup>37</sup> (ChIP-seq, RNA-seq, DNA methylation)	GEO	GSE107447
scRNA-seq <sup>53</sup> (Gut Cell Atlas)	ArrayExpress	E-MTAB-9543, E-MTAB-9536, E-MTAB-9532, E-MTAB-9533

**Experimental Models: Cell Lines**

GIST-T1 cell line	Cosmo Biosciences	PMC-GIST01C
-------------------	-------------------	-------------

**Oligonucleotides**

FGF4 real-time PCR primers: CCAACAACACTACAACGCCTACGA (Forward) CCCTTCTTGGTCTTCCCATTCT (Reverse)	IDT	N/A
FGF3 real-time PCR primers: ATGCTTCGGAGCACTACAGC (Forward) CACGTACCACAGTCTCTCGG (Reverse)	IDT	N/A
Ribosomal protein, large, P0 (RPLP0) real-time PCR primers: TCCCACCTTGCTGAAAAGGTCA (Forward) CCGACTCTTCCTTGGCTTCA (Reverse)	IDT	N/A
Locus-specific Bisulfite-sequencing (CTCF#1): AAATTTAAGAAATATTAAGGTGGGAAAG (Forward) ACCTAAAAATTACAATTAACCAACCC (Reverse)	IDT	N/A
Locus-specific Bisulfite-sequencing (CTCF#2): TGGTGGTTTAGTTTGTGTTTGAATTAAGA (Forward) ACCCACCTTAAAATAAAATTAACCAACCA (Reverse)	IDT	N/A
Locus-specific Bisulfite-sequencing (CTCF#3): ATTGTGATTGGTTGTGTTTATATGGTGTA (Forward) AATCCTAAATCCAAACCAATCC (Reverse)	IDT	N/A
Locus-specific Bisulfite-sequencing (CTCF#4): ATGAAATGTAGTAATGTTTTTGTATATGG (Forward) ACTCTATCCTTTAAAAACAACCC (Reverse)	IDT	N/A

**Recombinant DNA**

U6-sgRNA-CMV-Cas9-T2A-eGFP (piggybac construct)	This study	N/A
---	------------	-----

(Continued on next page)

**Continued**

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and algorithms		
R version 4.1.3	R Core Team	<a href="https://www.rproject.org">https://www.rproject.org</a>
Benchling (for sgRNA design)	Benchling	<a href="https://www.benchling.com">https://www.benchling.com</a>
Juicer tools	Github	<a href="https://github.com/aidenlab/juicer">https://github.com/aidenlab/juicer</a>
BWA aln version 0.7.4	Github	<a href="https://github.com/lh3/bwa">https://github.com/lh3/bwa</a>
Picard toolkit 2.9.2	Broad Institute	<a href="https://broadinstitute.github.io/picard/">https://broadinstitute.github.io/picard/</a>
MEME suite (version 4.11.3-1)	MEME suite	<a href="https://memesuite.org/meme/doc/download.html">https://memesuite.org/meme/doc/download.html</a>
RSEM version 1.3.2	Github	<a href="https://github.com/deweylab/RSEM">https://github.com/deweylab/RSEM</a>
DESeq2 version 1.40.2	Bioconductor	<a href="https://bioconductor.org/packages/release/bioc/html/DESeq2.html">https://bioconductor.org/packages/release/bioc/html/DESeq2.html</a>
plotgardener version 1.6.1	Bioconductor	<a href="https://bioconductor.org/packages/release/bioc/html/plotgardener.html">https://bioconductor.org/packages/release/bioc/html/plotgardener.html</a>
strawr version 0.0.91	CRAN	<a href="https://cran.r-project.org/web/packages/strawr/index.html">https://cran.r-project.org/web/packages/strawr/index.html</a>
GENOVA version 1.0.1	Github	<a href="https://github.com/robinweide/GENOVA">https://github.com/robinweide/GENOVA</a>
Pheatmap version 1.0.12	CRAN	<a href="https://cran.r-project.org/web/packages/pheatmap/index.html">https://cran.r-project.org/web/packages/pheatmap/index.html</a>
SCANPY	Github	<a href="https://github.com/scverse/scanpy">https://github.com/scverse/scanpy</a>

**RESOURCE AVAILABILITY**

**Lead contact**

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Bradley E. Bernstein ([bradley\\_bernstein@dfci.harvard.edu](mailto:bradley_bernstein@dfci.harvard.edu)).

**Materials availability**

Plasmids generated in this study are available upon request.

**Data and code availability**

- Accession information of published data sets analyzed in the manuscript are included in the [key resources table](#) along with information for original datasets. Original data generated in this study have been deposited at GEO (GSE241927) and are publicly available as of the date of publication.
- This paper does not report any original code.
- Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request.

**EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS**

**Cell cultures**

The GIST-T1 cell line, which was derived from a KIT-mutant GIST and maintains a GIST-like enhancer landscape and the FGF boundary,<sup>56</sup> was obtained from Cosmo Biosciences (no. #PMC-GIST01C). This cell line was cultured using DMEM with 10% fetal bovine serum, 1× penicillin-streptomycin and 1× Glutamax (Life Technologies) at 37°C in a humidified CO<sub>2</sub>-controlled (5%) incubator.

**METHOD DETAILS**

**Insulator knockout cell line**

For genome editing, we designed a U6-sgRNA-CMV-Cas9-T2A-eGFP piggyBac plasmid with the following sgRNAs: human non-targeting: 5'-ACGGAGGCTAAGCGTCGCAA-3'; CTCF site 1: 5'-GTCCCACTGCCACCACAAGA-3'; CTCF site 2: 5'-ATGACATGATGGC CAGCAGA-3'; CTCF site 3: 5'-GAAAAGCAACCGCCTCTAGG-3'; CTCF site 4: 5'-GGGCCAGGCCCGCCGCCAGG-3'; CTCF site 5: 5'-ATTTGGGAATCCTGGCTGCA-3'; CTCF site 6: 5'-AGAGAAGGGAAGGCCACTAG-3'; CTCF site 7: 5'-TTGTTTCTGTTGCCACC ACA-3'; CTCF site 8: 5'-CTGCCCGACTCTCCAGCAGA-3'; CTCF site 9: 5'-TACTGAGGCCTACCAGCAGA-3'; sgRNAs were designed to disrupt the NGG portions of the CTCF motif and to have high specificity using the Benchling software. Constructs harboring



two sequential sgRNAs cassettes were synthesized by TwistBioscience and cloned using standard methods. We verified insulator disruptions at target loci using CTCF ChIP-seq assay.

GIST-T1 WT cells were transfected with 1:1 ratio of the genome editing construct along with the piggyBac transposase construct using LipoD293 per manufacturer's guidelines. Cells were sorted twice for GFP positivity on a Sony SH800 flow cytometer, once 48 hours after transfection, and another time after the cells reached confluence from the initial sort. Crosslinked cells were harvested for ChIP analysis to verify loss of CTCF binding.

### Quantitative real-time PCR

Total RNA was isolated from Cell samples using the RNeasy Plus Kit (Qiagen no. 74134) and used to synthesize cDNA with the ProtoScript II First Strand cDNA Synthesis Kit (NEB no. E6560). cDNA was analyzed using the SYBR mastermix (Applied Biosystems) on a 7500 Fast Real Time system (Applied Biosystems). Gene expression primers were as follows: *FGF4* set 1 forward 5'-CCAA CAACTACAACGCCTACGA-3'; *FGF4* set 1 reverse 5'-CCCTTCTGGTCTTCCCATTCT-3'; *FGF4* set 2 forward 5'-GCAGCAAGGG CAAGCTCTAT-3'; *FGF4* set 2 reverse 5'-CGGTTCCCCTTCTGGTCTT-3'; *FGF3* forward 5'-ATGCTTCGGAGCACTACAGC-3'; *FGF3* reverse 5'-CACGTACCACAGTCTCTCGG-3'. All gene expression results were normalized to primers for ribosomal protein, large, P0 (*RPLP0*) as follows: forward 5'-TCCCACCTTGCTGAAAAGGTCA-3'; reverse 5'-CCGACTCTTCTGGCTTCA-3'.

### RNA-seq

Total RNA was isolated from cell samples using the RNeasy Plus Kit (Qiagen) and quality was determined via TapeStation (Agilent). Libraries were prepared using the NEBNext Poly(A) mRNA Magnetic Isolation Module (NEB no. E7490) and the NEBNext Ultra II RNA Library Prep Kit (NEB no. E7770), and equimolar multiplexed libraries were sequenced with 38 base paired end reads on an Illumina NextSeq 500.

### Chromatin immunoprecipitation

ChIP-seq was performed as described previously.<sup>37,39</sup> Briefly, cultured cells were crosslinked in 1% formaldehyde for 12 minutes, flash frozen in liquid nitrogen, and then stored at -80 °C. Chromatin was fragmented using a Branson Sonifier calibrated to shear DNA to between 200-600 bp fragment length. CTCF was precipitated using a monoclonal rabbit antibody (Cell signaling clone D31H2, no. 3418). H3K27ac was precipitated using a polyclonal rabbit antibody (Active Motif no. 39133). H3K27me3 was precipitated using a polyclonal rabbit antibody (Cell signaling clone C36B11, no. 9733). Eluted ChIP DNA was used to generate sequencing libraries by NEBNext Ultra II DNA Library Prep Kit (NEB no. E7645). Barcoded fragments were amplified by PCR using NEBNext Ultra II Q5 Master Mix for 16 cycles and double-size selected using AMPure XP beads (Beckman Coulter no. A63880) for fragments between 300-500 bp.

### Locus bisulfite sequencing

Briefly, genomic DNA was extracted (Monarch Genomic DNA Purification Kit, NEB no. T3010) and subjected to bisulfite conversion using EZ DNA Methylation-Lightning Kit (Zymo Research no. D5030). The conversion was split to four independent PCR reactions and CTCF sites were then PCR amplified using KAPA HiFi HotStart Uracil+ReadyMix PCR Kit (KAPA no. KK2801) and the following primers: CTCF 1 forward 5'-AAATTTAAGAATATTAAGGTGGGAAAG-3', reverse 5'-ACCTAAAAATTACAATTAACCTCAACCC-3'; CTCF 2 forward 5'-TGGTGGTTTAGTTTGTGGTGAATTAAGA-3', reverse 5'-ACCCACCTTAAAATAAAAATTAACCAA-3'; CTCF 3 forward 5'-ATTGTGATTGGTTGTGTTTATATGGTGA-3', reverse 5'-AATCCTAAATCCAAACCCAAATCC-3'; CTCF 4 forward 5'-ATGAAATGTAGTAATGTTTTTGTATATGG-3', reverse 5'-ACTCTATCCTTTAAAAACAACCC-3'. Sequencing reads (Illumina) were aligned to the bisulfite converted locus and the frequency of methylated to unmethylated Cytosines was calculated.

### Region Capture Micro-C

#### Crosslinking

RCMC was performed as described previously.<sup>9</sup> Briefly, trypsinized cells were doubly crosslinked to fix protein-protein and protein-DNA interactions using 3 mM disuccinimidyl glutarate (ThermoFisher no. 20593) and 1% methanol-free formaldehyde (ThermoFisher no. 28906) in 1 × PBS (ThermoFisher no. 10010023), respectively. The crosslinking reaction was gently mixed at room temperature for 35 min, after which formaldehyde was added to a final concentration of 1%. The double crosslinking reaction was mixed at room temperature for an additional 10 min before quenching with Tris buffer pH 7.5 (ThermoFisher no. 15567027) at a final concentration of 0.375 M. Crosslinked cells were washed twice with 1 × PBS containing 100 μg/mL BSA (NEB no. B9000), recounted to quantify any sample loss during fixation and then partitioned into 5 M cell aliquots that were pelleted and snap-frozen in liquid nitrogen for storage at -80 °C.

#### MNase digestion

Cell membranes were solubilized to extract intact nuclei by resuspending crosslinked 5 M cell pellets in Micro-C Buffer #1 (MB#1; 50 mM NaCl, 10 mM Tris-HCl pH 7.5, 5 mM MgCl<sub>2</sub>, 1 M CaCl<sub>2</sub>, 0.2% NP-40 Substitute (Abcam no. ab142227), 1 × Protease Inhibitor Cocktail (Sigma-Aldrich no. 5056489001)) at 1 M cells per 100 μL for 20 min on ice. Following an MB#1 wash, samples were resuspended in 500 μL MB#1 and 1-2 μL MNase (NEB no. M0247) was added. This digestion reaction was mixed at 37 °C for 20 min on a thermomixer before being quenched with 20 mM EGTA (bioWORLD no. 40520008) and heat inactivated at

65 °C for 10 min. Digested nuclei were washed twice with ice-cold Micro-C Buffer #2 (50 mM NaCl, 10 mM Tris-HCl pH 7.5, 10 mM MgCl<sub>2</sub> and 100 μg/mL BSA).

#### **End repair and labeling**

To generate blunt ends on digested DNA fragments before proximity ligation and add biotinylated nucleotides, a series of enzymatic processing steps were performed. First, to catalyze the addition of 5'-phosphate groups and the removal of 3'-phosphate groups, digested samples generated from 5 M cell inputs were incubated in end-repair reactions (50 U T4 Polynucleotide Kinase (NEB no. M0201), 50 mM NaCl, 10 mM Tris-HCl pH 7.5, 10 mM MgCl<sub>2</sub>, 100 μg/mL BSA, 2 mM ATP (NEB no. P0756) and 5 mM DTT (Sigma-Aldrich no. 10197777001), in water) at 37 °C for 15 min while mixing. To create 5' fragment overhangs for end-blunting and labeling, 50 U DNA Polymerase I Klenow Fragment (NEB no. M0210) was added to the reaction and incubated at 37 °C for 15 min while mixing. Next, a mixture of dNTPs in end-labeling buffer (66 μM each of dTTP (Invitrogen no. 10297018), dGTP (Invitrogen no. 10297018), biotin-dATP (Jena Bioscience no. NU-835-BIO14), biotin-dCTP (Jena Bioscience no. NU-809-BIOX) and 100 μg/mL BSA in 1 × T4 DNA Ligase Buffer) was added to the reaction. This reaction was incubated at room temperature for 45 min with interval mixing before being quenched by 30 mM EDTA (Invitrogen no. 15575020) and heat inactivated at 65 °C for 20 min. Finally, end-blunted and biotin-labeled nuclei were washed once with Micro-C Buffer #3 (50 mM Tris-HCl pH 7.5, 10 mM MgCl<sub>2</sub> and 100 μg/mL BSA).

#### **Proximity ligation and removal of unligated biotin**

Proximity ligation was performed by incubating labeled chromatin in a ligation reaction (10,000 U T4 DNA Ligase (NEB no. M0202), 1 × T4 DNA Ligase Buffer, 100 μg/mL BSA) at 25 °C for overnight with gentle mixing. To remove biotinylated dNTPs from all unligated fragment ends, samples were digested by 1,000 U Exonuclease III (NEB no. M0206) in reaction buffer (1 × NEBuffer #1 and 100 μg/mL BSA in water) at 37 °C for 15 min with interval mixing.

#### **DNA purification and size selection**

To prepare ligated DNA for library generation, DNA was reverse crosslinked and proteins and RNA were digested by adding 200 μL ProK Buffer (20 mM Tris-HCl pH7.5, 100 mM NaCl, 1 mM EDTA, 1 mM EGTA, 0.5 % Tx-100, 0.2 % SDS (Invitrogen no. AM9822), 24 U Proteinase K (NEB no. P8107), and 100 μg/mL RNaseA (Invitrogen no. EN0531), in water) to the samples and incubating at 65 °C overnight. DNA was extracted using the Zymo DNA Clean & Concentrator kit (Zymo Research no. D4034) according to the kit manual.

Dinucleosome-sized DNA fragments (250–350 bp) were isolated by extraction from a 1% agarose gel (VWR no. 97062). Gel extracts were purified using the Zymo Gel Purification kit (Zymo Research no. D4008), and samples were quantified by Qubit 1 × dsDNA High Sensitivity Assay (Invitrogen no. Q33231). Sample ends were polished and blunted again using the Quick Blunting Kit (NEB no. E1201) at 25 °C for 45 min, followed by reaction inactivation at 65 °C for 10 min.

Ligated DNA contact fragments were isolated by pulling down biotin-bound fragments using Dynabeads MyOne Streptavidin T1 (Invitrogen no. 65601). DNA samples were bound to beads in a Binding and Wash Buffer (1 M NaCl, 5 mM Tris-HCl pH 7.5, 500 μM EDTA, 0.1% Tween-20 (Sigma-Aldrich no. P8074) in water) at room temperature for at least 30 min with mixing. After two washes with the Binding and Wash Buffer, the bead-bound samples were washed once with 10 mM Tris-HCl pH 7.5 before library preparation.

#### **Library preparation**

Illumina library preparation was performed using the NEBNext Ultra II kit (NEB no. E7645) and NEB Multiplex Oligos for Illumina Primer Set 1 (NEB no. E7335) to end-repair, A-tail, and adaptor ligate the bead-bound samples. All steps were performed as directed by the manual, except that incubations included interval shaking (1 min on, 3 min off) at 1,000 rpm. Sample washes were performed using Binding and Wash Buffer and Tris Buffer (10 mM Tris-HCl pH 7.5 in water). After adaptor ligation, samples were amplified by PCR for 7–8 cycles and then purified using AmPure XP beads.

#### **Capture of target loci and sequencing**

80-mer probes were designed to tile end-to-end without overlap across the capture loci through Twist Bioscience. Probes with high predicted likelihoods of off-target pull-down (for example, such as those in high-repeat regions) were masked and removed from the probe tiling, and probe coverage was double-checked to ensure the inclusion of key genomic features (for example, all promoters and CTCF/K27ac peaks in the locus) before finalization. Probe panels were synthesized and purchased as Custom Target Enrichment Panels from Twist Bioscience.

Capture enrichment was performed in accordance with Twist Bioscience's Target Enrichment Standard Hybridization v1 Protocol. Briefly, pooled sample libraries were dried and mixed with Hybridization Mix (Twist Bioscience no. 104178), Custom Panels (Twist Bioscience no. 101001) and Universal Blockers (Twist Bioscience no. 100578). The library pool was hybridized to the biotinylated probe panel overnight, after which streptavidin beads (Twist Bioscience no. 100983) were used to pull down probes with hybridized ligated fragments and then washed (Twist Bioscience no. 104178) to remove unbound fragments. Bead-bound libraries were amplified by PCR using the Equinox Library Amplification Mix (Twist Bioscience no. 104178) for 10 cycles and purified using AmPure XP beads. Pooled libraries were sequenced by paired-end 2 × 50 cycle sequencing kits with Illumina NovaSeq S1 flow cells on a NovaSeq 6000 system by the Broad Institute of MIT and Harvard's Walk-Up Sequencing services.

## **QUANTIFICATION AND STATISTICAL ANALYSIS**

#### **CTCF motif analysis**

CTCF peaks of GIST-T1 cells were called by MACS3 (version 3.0.0) and merged by bedtools (version 2.31.0). Peaks were then centered around the CTCF motif where found by FIMO (MEME 4.7) at a 100 bp window around the peak center, based on the

“JASPAR\_CORE\_2021\_vertbrates” database (MA0139.1). If multiple motifs were detected, we kept the one with the highest score.

### RNA-seq analysis

Libraries were sequenced as paired 38 base end reads on an Illumina NextSeq 500. Reads were aligned using STAR 2.5.345 to the human reference (hg38). RNA-seq data for clinical GISTs were previously published (GEO: GSE107447).<sup>37</sup> Gene expression was estimated by RSEM (version 1.3.2). Data were visualized using R version 4.3.1.

### ChIP-seq analysis

Libraries were sequenced as paired 38 base end reads on an Illumina NextSeq 500. Reads were then aligned to the hg38 reference genomes using BWA aln version 0.7.4, removing reads with MAPQ score lower than 10. PCR duplicates were removed by Picard toolkit 2.9.2. Peaks were called with MACS3, correcting against input controls. Differential analysis of CTCF peaks and quantification of reads per peak were previously done. CTCF/H3K27ac ChIP-seq on GIST patients was downloaded from GEO: GSE107447,<sup>37</sup> normalized by DESeq2 and analyzed. BigWig files normalized for reads per million (RPM) were visualized using the plotgardner package (version 1.6.1)<sup>57</sup> on R version 4.3.1. Processed (RPM normalized BigWig), and raw data have been deposited on GEO (accession numbers: GSE241924).

### Mapping and normalizing RCMC

RCMC paired-end reads generated by the Illumina NovaSeq sequencers were downloaded as.fastq files for each sample, pair mate, and flow cell lane. Read quality was verified using FastQC (v0.11.9). Paired end reads were aligned to the UCSC hg38 genome using Juicer (default setting, version 1.6). Deduplicated paired reads were filtered by mapping quality score (MAPQ  $\geq$  30) and filtered to retain those reads where both read mates lay within the capture region of interest (chr11:69.4-70.4Mb). These filtered reads were converted to contact matrices (.hic format file) using Juicerbox\_Tools with Pre at base pair delimited resolutions of 1 Mb, 500 kb, 100 kb, 25 kb, 10 kb, 5 kb, 1 kb and 500 bp. All contact matrices used for further analysis were KR-normalized with Juicer. RCMC contact maps were visualized alongside genomic annotations and ChIP-seq datasets using the ‘plotgardner’ package (version 1.6.1) on R version 4.3.1.

### Chromatin contact analysis

#### Insulation score/boundary

Insulation scores and TAD-boundaries have been calculated using the ‘GENOVA’ package (version 1.0.1)<sup>58</sup> on R version 4.3.1. For the TAD calling performed on RCMC maps (.hic files), we used 1 kb contact matrices, with a window size of 100 kb and a min\_strength of -0.2 to find TAD-boundaries. Insulation scores and TAD-boundaries were plotted using R.

#### Correlation matrix

Contact map data KR-normalized at 1 kb resolution were extracted from.hic files (MAPQ  $\geq$  30) using the ‘strawr’ package (version 0.0.91) on R version 4.3.1. The extracted data were filtered to bins comprising CTCF/H3K27ac peaks that were called from CTCF/H3K27ac ChIP-seq data by MACS3 with a minimum distance of 20 kb. From these filtered matrices, the pairwise correlation matrix between RCMC contact maps has been visualized using the ‘Pheatmap’ package (version 1.0.12) on R.

#### Differential contacts

Contact map data KR-normalized at 5 kb resolution were extracted from.hic files (MAPQ  $\geq$  30) using the ‘strawr’ package in R and filtered with a minimum distance of 20 kb. From this data, all differential contacts with adjusted p value  $<$  0.05 and absolute log2fold-change  $>$  0.5 were identified using the ‘DESeq2’ package (version 1.40.2) in R version 4.3.1. To extract CTCF-CTCF or H3K27ac-H3K27ac differential contacts, all differential contacts are filtered to contacts between the bins comprising CTCF or H3K27ac peaks that were called from CTCF or H3K27ac ChIP-seq data by MACS3, respectively. Differential contacts were visualized using the ‘plotgardner’ package (version 1.6.1) on R.

#### Virtual 4C

Virtual 4C tracks were generated by using Juicebox. Horizontal and vertical 1D tracks of the pixel at 1 kb resolution for the interaction between *FGF3* promoter (chr11: 69,819,000-69,820,000) and either *eRNA-1* (chr11: 69,926,000-69,927,000), *eRNA-2* (chr11: 69,986,000-69,987,000) or *ANO1* promoter (chr11: 70,078,000-70,079,000) were generated with Juicebox ‘generate 1D track’ function. Wig files were visualized using the ‘plotgardner’ package (version 1.6.1) on R version 4.3.1.

#### IntraTAD and CrossTAD

Contact frequencies at 5 kb resolution were extracted from.hic files (MAPQ  $\geq$  30) using the ‘strawr’ package in R and filtered with a maximum distance of 10 kb. The normalization factors were calculated from the filtered frequencies using the ‘DESeq2’ package (version 1.40.2) in R version 4.3.1. Contact map data KR-normalized at 5 kb resolution were extracted from.hic files (MAPQ  $\geq$  30) using the ‘strawr’ package in R and filtered with a minimum distance of 20 kb and then applied by the normalization factors. Contacts in chr11:69.675-70.195 mb are counted for total contacts in the FGF-ANO1 locus. Contacts in chr11:69.675-69.905 mb and chr11:69.910-70.195 mb are counted for IntraTAD contacts in the FGF and ANO1 TADs, respectively. Contacts between chr11:69.675-69.905 mb and chr11:69.910-70.195 mb are counted for CrossTAD contacts between the FGF and ANO1 TADs. The ratio of CrossTAD/IntraTAD and the number of total contacts were plotted using R version 4.3.1.

### Analysis of ENCODE, 4DN, Gut Cell Atlas data

Normalized ChIP-seq data and insulation scores with boundary calls from various human tissues and cell lines were downloaded from ENCODE and 4DN, respectively. Genome locus figures were plotted using the ‘plotgardner’ package (version 1.6.1) on R version 4.3.1. Expression of *ANO1*, *eRNA-1* and *eRNA-2* in scRNA-seq from human gut cells was plotted from Gut Cell Atlas, dataset ArrayExpress: E-MTAB-9543, E-MTAB-9536, E-MTAB-9532, E-MTAB-9533.

### Analysis of scRNA-seq data

The published 10X 5’ libraries<sup>53</sup> were mapped to Human Gencode 41 assembly<sup>59</sup> using STARsolo.<sup>60</sup> We have set barcode and umi parameters in STARsolo as follows “–soloType CB\_UMI\_Simple –soloCBstart 1 –soloCBlen 16 –soloUMIstart 17 –soloUMIlen 10 –soloCBmatchWLtype 1MM\_multi\_Nbase\_pseudocounts”. We have adjusted the 5p clipping to 39 “–clip5pNbases 39 0”. We have also set the following parameters to ensure that the gene expression matrices generated are compatible with the downstream analysis steps “–soloUMIdedup 1MM\_CR –outFilterScoreMin 30 –outSAMtype BAM SortedByCoordinate –outSAMattributes CR UR CY UY CB UB –soloUMIfiltering MultiGeneUMI\_CR –soloBarcodeMate 1”

The raw matrix outputs of STARsolo were then processed using SCANPY.<sup>61</sup> The matrices were normalized, log1p transformed and then scaled. Technical variabilities due to total counts and mitochondrial content were regressed out using the “sc.pp.regress\_out” function. PCA, clustering and UMAP generation were performed using SCANPY’s inbuilt functions.

### Group comparisons

Pearson r values were computed for all correlations. Two-sided t test P values were computed when comparing two groups. One-way ANOVA P values were computed when comparing more than 2 groups.