Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Distinguishing insulators from enhancers by their ability to block long-range contacts using HiCAggR

Naomi Schickele
Université de Toulouse
Olivier Fosseprez
Université de Toulouse
Robel Tesfaye
Université de Toulouse
Charlène Perrois
Université de Toulouse
Nicolas Chanard
Université de Toulouse
Anne-Laure Finoux
Université de Toulouse
Eldon Emberly
Simon Fraser University, British Columbia
Pascal GP Martin
Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement
Stéphane Schaak
Université de Toulouse
Olivier Cuvier
olivier.cuvier@inserm.fr

Université de Toulouse

Research Article

Keywords: Insulators, Higher-order chromatin organization, Long-range contacts/interactions, 3D Chromatin folding, Gene expression variability, Insulator permissiveness

Posted Date: October 23rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10895521/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

Topologically Associating Domains (TADs) define domains of high levels of long-range contacts that are often bracketed by insulators. Previous work showed the ability of insulators to block an enhancer from activating a distant promoter, if interposed. The model has been questioned as few insulators were reported to regulate gene expression by preventing promiscuous enhancer-promoter (E-P) contacts over TAD borders. Analyses probing all E-P contacts depending on insulators have been challenged by the sparsity of Hi-C data in detecting promiscuous long-range contacts. How insulators influence long-range contacts and gene expression remains unclear.

Results

Here, we developed a Bioconductor package named "Hi-C oriented Aggregation Peak Analysis" (HiCAggR), which uses aggregation to quantify with great sensitivity, at high resolution, and in an oriented fashion, the long-range contacts occurring both -at and -surrounding the genomic features of interest, as assessed by Chromosome Conformation Capture techniques (3C/Hi-C). The gains in detection allow to distinguish insulators from other regulatory sequences, from their ability to block promiscuous enhancer-promoter contacts. Accordingly, insulator protein depletion results in “leakages”, defined as long-range contacts (LRCs) formed across insulators, as shown by quantifying the signal at the single locus level. Importantly, leakages include novel E-P long-range contacts that are associated with the deregulation of hundreds of neighboring genes. Single-cell multiomics confirm these deregulations and highlight an increased cell-to-cell variability in gene expression associated with the novel E-P contacts over insulators.

Conclusions

HiCAggR enables to distinguish insulators from enhancers from their ability to block 3D long-range contacts, which otherwise affect the expression of hundreds of nearby off-target genes.

INTRODUCTION

Chromatin is organized into topologically associating domains (TADs) with sizes ranging from approximately 100 kb to 2 Mbp containing multiple sub-TADs of approximately 5 to 100 kb [1–12]. TADs are often delineated by insulator proteins that were proposed to participate in gene expression homeostasis, by shielding genes from promiscuous activation by enhancers outside the same TAD [13–15]. Actually, insulator proteins were originally shown to be able to block enhancer-promoter (E-P) long-range contacts across the insulator sites, i.e. when interposed [16]. Insulator proteins include notably the conserved CCCTC-binding Factor (CTCF), and additional insulator proteins like Suppressor of Hairy-Wing

(Su(Hw)), Zeste-white 5 (Zw5) or Boundary Element Associated factor of 32 kDa (BEAF-32) in Drosophila that also binds to TAD borders [10, 17–22]. How systematically insulator proteins may favor long-range interactions (LRIs) within TADs while limiting LRIs across TAD borders, and how frequent this might influence gene expression, has been questioned. Such a model has been only partially supported by depleting insulator proteins, or their binding sites, which mildly affected the TAD structures [23–30] and provided only a few examples of neighbor genes being affected, as illustrated by polydactyl phenotypes [31]. Moreover, removal of an insulator protein site from a TAD border could alter the expression of the nearby genes, a proof of principle supporting their function [32]. At genome-wide levels, the structural and functional roles of insulator proteins have however been questioned in Drosophila, given the absence of a clear impact on TADs for mutants of the drosophila insulator protein dCTCF or its co-factors [33]. Improving the detection of the LRIs showed that dCTCF could influence 3D contacts across TADs' borders yet this was accompanied with rare changes in gene expression [34, 35]. Also, few cases of gene deregulation have been attributed to defects in insulator proteins at TAD borders and large-scale genomic rearrangements that disrupt TADs can conserve gene expression profiles [36, 37].

Of interest, the SMC complex named cohesin can entrap two separate DNA sites [38]. Also, in eukaryotes, the loop extrusion model proposes that the progressive extrusion of DNA occurs through a cohesin ring, which may account for the overall distribution of (progressing) LRIs within TADs, as detected by ensemble capture of 3D contacts using genome-wide chromosome conformation capture 3C/Hi-C [39–41]. Remarkably, the cohesin-mediating loop extrusion model was supported by studies highlighting nano-compartments representing TADs, visualized by high-resolution microscopy in mammals [42] unlike in Drosophila [43]. Loop extrusion is then blocked or paused when it reaches clusters of CTCF sites in mammals [8, 24, 39, 44–46], which may regulate E-P long-range interactions [47]. 3D clusters or hubs involving RNAPII may also regulate E-P contacts independently of cohesin [48]. Accordingly, gene expression can be regulated by numerous types of 3D contacts, including more transient or weaker long-range contacts (LRCs) e.g. due to random collisions between regulatory elements [49, 50]. This may involve additional factors including non-coding RNAs, DNA methylation, or transcription factors [51–54]. A first paradigm is that the cooperative action of enhancers can lead to persistent transcription in the absence of cohesin-mediated E-P contacts [55]. Another paradigm is found in Drosophila where insulator pairing may actually not require cohesin or at least loop extrusion [17, 56]. E-P long-range interactions may set transcriptional bursts, with insulators being able to modulate the interaction frequencies of such bursts [57], depending on the cellular context [58]. Nevertheless, E-P interactions are not the only parameter influencing transcription [47], and evaluating how insulators perturb expression depending on E-P interactions is not a straightforward matter.

Here, we developed the R/Bioconductor package HiCAggR, which provides a framework for the quantification and statistical analysis of Hi-C contact signals between and surrounding genomic features of interest, thereby enabling integration with complementary 1D genomic data. In this study, we focused on regulatory elements, including insulators and enhancers. By inspecting patterns of LRCs surrounding them, we show that HiCAggR discriminates between insulators and enhancers based on their ability to block long-range contacts. Depleting one of the major insulator proteins in Drosophila, BEAF-32, actually

leads to novel LRCs across insulator protein sites, defining “insulator permissiveness” or “leakage”. HiCAggR detects leakage over individual loci, correlating with insulator protein binding. The leakages allow novel E-P long-range contacts to form across insulators with off-target genes becoming up-regulated upon insulator protein depletion, as validated using a single-cell multiomics approach. Novel LRCs are associated with increased cell-to-cell variability in gene expression, thereby illustrating how insulators may contribute to the 3D rewiring of enhancer–gene interactions which may not strictly depend on the structural genome organization of chromosomes into TADs.

RESULTS

BEAF-32 does not affect gene expression through TAD organization

We sought to assess the possible role of insulators on gene expression through TAD borders (Fig. 1A). To this end, we efficiently depleted BEAF-32 from S2 cells (see Materials and Methods; Supplem Fig S1A-B), one of the major insulator proteins associating with TAD borders in Drosophila [59, 60]. We then performed Hi-C and gene expression analyses after dsRNA-mediated depletion of BEAF-32 (hereafter referred to as "Beaf-KD" for BEAF-32 knockdown) compared with depletion of the exogenous *luciferase* gene used as a control (hereafter referred to as "Luc-KD" or "control"; see Methods). As a result, long-range interactions (LRIs) were not drastically changed upon depletion of BEAF-32 albeit a slight increase for interactions over short distances (Fig. 1B). This result indicates that although BEAF-32 often binds to TAD borders, its depletion is not sufficient to alter the global TAD organization, as confirmed by inspecting TADs on 2D plots (Fig. 1C). Such an observation was repeated at various scales of TAD detection (Supplem. Fig S1C; see Materials and Methods), similar to what has been observed upon depletion of CTCF [61]. The positions of the detected TADs in control conditions occasionally showed modest changes upon BEAF-32 depletion (Fig. 1C; see asterisks), as detected for ~ 13% of the TADs (Fig. 1D). The altered TADs showed a significant change in insulation score upon Beaf-KD compared to the conserved TADs, which was not found for the control Luc-KD condition (Supplem. Fig S1D; p-value of 0.0051 and 0.341, respectively; Wilcoxon test). Of note, however, no systematic defect in insulation score upon Beaf-KD was found for BEAF-32 sites at TAD borders (Fig. 1E; p-value = 0.167, Wilcoxon test; see Materials and Methods), which may reflect the redundancy of BEAF-32 with additional insulator proteins, e.g. dCTCF, often present at TAD borders [14, 18]. Accordingly, the number of BEAF-32 sites was not specifically higher for the altered compared to conserved TADs (Fig. 1F; p-value = 0.355, Wilcoxon test; see Materials and Methods), highlighting that there was no specific change in insulation score depending on the presence of BEAF-32 at TAD borders. Altogether, our work is consistent with a limited structural role of BEAF-32 in organizing TADs, in agreement with a recent study [33].

Furthermore, gene expression analysis by RNA-seq showed that Beaf-KD led to the up- or down-regulation of a few hundred of genes (Fig. 1G; see Materials and Methods). However, no correlation was observed between gene deregulation and localization within a TAD altered or not upon Beaf-KD as

compared to Luc-KD control (Fig. 1H; p-value = 0.855 and 0.428, respectively; Wilcoxon test). Moreover, no significant association could be found between these deregulated genes and the borders of TADs being altered or not (Supplem. Fig. S1E; p-value = 0.979 and 0.42). Therefore, we concluded that the influence of BEAF-32 on gene expression is unlikely to occur at the scale of TADs.

BEAF-32 regulates long-range contacts across its binding sites independently of TADs

Chromatin insulators can block enhancers from activating promoters, when positioned between enhancers and promoters (Fig. 2A) [16, 20, 62, 63]. As thousands of additional BEAF-32 sites localize away from a TAD border [10, 64], we sought to test whether such sites may block long-range contacts albeit not coinciding with a TAD border. Supporting this idea, the insulation score was impaired by Beaf-KD compared to control Luc-KD, for the BEAF-32 sites not overlapping (> 2kb away) with a TAD border (Supplem. Fig. S1F, p-value = 5×10^{-8}). However, enhancer-promoter contacts can involve rare or at least transient LRCs that might suffice to influence gene expression, without changing the overall TAD organization [65, 66]. To explore this in greater detail, we sought to probe LRCs with a more sensitive method named “Aggregation” or “Aggregate peak analysis” (APA), as developed by us and others [8, 18], which improves the detection of sparse contacts by aggregating them over multiple pairs of loci. As such, LRCs were quantified across individual sites, including insulators, enhancers or control sites, by measuring the differential Hi-C signal in Beaf-KD as compared to control Luc-KD cells (Fig. 2B). To this end, we developed the Bioconductor package HiCAggR, which provides a framework for the integration of 1D genomic features with Hi-C contact maps, enabling the quantification, aggregation, visualization, and statistical analysis of the Hi-C contact signals between the genomic regions of interest and surrounding chromatin environments ([67] https://www.bioconductor.org/packages/release/bioc/html/HicAggR.html). HiCAggR notably allows to analyze and represent 3D contacts in an orientation dependent manner (e.g. E-P without mixing E-P with P-E), with proper internal normalization of reads for each locus (or submatrix) by quantilization of the individual submatrices (see Materials and Methods). BEAF-32 depletion led to a significant increase in 3D contacts in the upper-left region of the APA, corresponding to interactions across the potential insulator sites (Fig. 2B, upper and lower APA, respectively). The increase in 3D contacts, or leakage, was ranked into 20 quantiles, from the top 5% of submatrices exhibiting the strongest increase upon Beaf-KD to the submatrices showing the lowest levels of leakages (Fig. 2B-C; p-value of 5×10^{-48} and 0.586, respectively; as quantified in the region from > 2 kb to 20 kb from insulator sites; see black rectangle (see Materials and Methods); Supplem. Fig. S2A-D). Strikingly, the ranking of individual APA submatrices according to the net leakage in 3D contacts across sites showed that BEAF-32 depletion actually led to leakage most specifically over the BEAF-32 sites displaying the highest ChIP-seq signals (Fig. 2D, left heatmap), as expected if its influence was direct. On the contrary, the ranking by leakage did not properly sort out the control H3K4me1 mark signal (Fig. 2D, right heatmap), as expected. This result - quantifying leakages across individual sites - suggested how in spite of the sparsity of Hi-C data, 3D contacts occurring across insulators could reflect specific effects of insulator protein depletion.

Permissiveness to long-range contacts distinguishes insulators from enhancers

Our analysis identified hundreds of genomic regions exhibiting increasing LRCs upon Beaf-KD, as illustrated for some of the top 5% permissive regions (Fig. 3A; Supplem. Fig. S3A). We first sought to further assess whether this actually reflected a blocking activity, i.e. specific to insulator sites. To this end, we tested the ability of insulators versus control enhancer sites to block LRCs, using receiver operating characteristic (ROC) curves [17]. Individual sites were ranked according to their leakage scores in the Beaf-KD compared to control condition. Insulator sites were efficiently sorted out over control enhancer sites in regions exhibiting the highest leakage scores, with an area under the curve (AUC) of 99.26% (Fig. 3B; left ROC). Accordingly, the top 5% of regions with the highest leakage confirmed BEAF-32 sites as preferential leakage sites over control sites, with a false-positive rate of approximately 1%. In stark contrast, the distinction between BEAF-32 and control sites was not found for regions with the lowest leakage scores, as expected (Fig. 3B; lower left ROC; AUC = 52.19%). Moreover, the ability to distinguish BEAF-32 from control enhancer sites was partially impaired by Beaf-KD including for the top 5% of loci with highest leakages, as BEAF-32 sites were not as efficiently sorted out compared to control sites, with an AUC reduced to 81.81% (Fig. 3B; upper-right ROC). The persistent score might suggest a partial maintenance of the insulating activity, maybe in part due to additional insulator proteins including dCTCF that often co-localizes with BEAF-32 [18, 19]. Of note, the insulator sites associated with leakage were not correlated with the borders of altered TADs as compared to the control TADs (Supplem. Fig. S3B), illustrating how HiCAggR could capture changes in LRCs that were not detected at the TAD scale, but may rather involve local regulation of LRCs.

Given these results, we further tested whether the HiCAggR package tools could readily distinguish insulators from enhancers, solely based on their ability to block LRCs independently of insulator protein depletion, i.e. in the control condition. To this end, the relative levels of LRCs were classified across all insulator or enhancer sites, and then compared. ROC analysis showed that all tested Drosophila insulator-associated factors, notably dCTCF, CP190, Su(Hw), and BEAF-32 (see Materials and Methods), were efficiently distinguished based on their ability to block LRCs, from enhancers (Fig. 3C). Strikingly, the AUC was 99.2%, with a false-positive rate of < 1%, thereby indicating that the ability of insulators to block LRCs provides a highly accurate means for distinguishing them from enhancers. Accordingly, the quantification of the APA signal in Beaf-KD compared to Luc-KD cells showed that BEAF-32 sites had a higher blocking activity as compared to enhancers (Fig. 3D; p-value of 3×10^{-5}; using a Wilcoxon test). The AUC value was also significant for the regions exhibiting the highest leakage upon Beaf-KD (Supplem. Fig. S3C; AUC of 99.5%). Of note, however, regions showing a lower leakage also led to sort out insulators over enhancers (AUC of 94.76%) as expected given that additional insulator proteins, e.g. dCTCF, exist in Drosophila, so that assessing leakage upon Beaf-KD cannot account for all LRCs (see Discussion).

Leakage is associated with novel enhancer-promoter contacts

Next, we sought to test whether the leakage detected upon insulator protein depletion, may be associated with novel enhancer-promoter contacts across insulators (Fig. 4A). The HiCAggR package enables detection of specific enhancer-promoter (E-P) long-range contacts, by normalizing the Hi-C reads of these E-P contacts to the interactions within the chromatin environment, in line with previous normalization procedures (see Materials and Methods) [8, 17]. E-P couples were then searched systematically to identify the most significant E-P contacts using the accompanying package HiCAggR, for the Beaf-KD and Luc-KD conditions. Aggregation plots were then performed, confirming the formation of novel E-P long-range contacts. Such E-P actually showed a strong signal in the center of the APA, which was specific to Beaf-KD cells compared to control cells (Fig. 4B, right and left APA, respectively). We observed that such LRCs were also more specific and focused for regions exhibiting the highest leakage as compared to more diffuse LRCs for regions with the lowest leakage (Fig. 4B, compare upper with lower APAs in Beaf-KD). One hundred significant E-P contacts were detected in the Beaf-KD condition corresponding to the regions exhibiting highest increases in leakage (Fig. 4C-D; see Supplementary Table S1; Materials and Methods). In contrast, only 32 new LRCs involving E-P were detected in Luc-KD condition for the same (top 5%) regions (Supplem. Fig. S4A), i.e. more than three times fewer. Moreover, only 11 new E-P contacts were detected in Beaf-KD for regions with the lowest leakage (Fig. 4C), similar to the control condition (i.e. 8 in total; Supplem. Fig. S4B). Of note, the ranking according to leakage scores did not show any bias with respect to the number of enhancer-promoter contacts (Supplem. Fig. S4C), excluding that this may account for the increase in enhancer-promoter LRCs in leakage regions. Moreover, enhancer-promoter distances were similar across all leakage-score quantiles, ruling out the possibility that the higher number of short-range contacts observed upon BEAF-32 depletion (Fig. 1B) could have hampered the detection of novel LRCs (Supplem. Fig. S4D). Taken together, these data support the conclusion that leakages across insulator sites enable the emergence of novel enhancer-promoter contacts.

The finding that insulator protein depletion leads to novel enhancer-promoter contacts led us to test whether leakages may be associated with the up-regulation of genes localized in the same genomic regions. We sought to score changes in expression that might be related to changes in LRCs. In this instance, genes whose promoters were closely (< 3kb) associated with a BEAF-32 site were filtered out as these sites are likely to directly influence gene expression, as shown previously [67–70]. Among all other genes, we observed a significant enrichment in up-regulated genes for genes pertaining to loci pairs exhibiting the highest leakages (Fig. 4E; p-value = 9.59×10^{-3}; Fisher's exact test), with a global odds-ratio of 1.67-fold to 1.86-fold as compared to control, for the loci corresponding to the range of top 5–20% leakage levels (Q1-Q4). This effect was specific to the up-regulated genes, as it was not observed for down-regulated genes (Supplem. Fig. S4E). Moreover, further statistical test using bootstrap confirmed that the association between leakage regions and up-regulated genes was significantly stronger than that for low leakage regions (Fig. 4F; left bar plots; a chi2 test confirmed that the two distributions were significantly different: p-value < 2×10^{-16}), unlike for down-regulated genes (Fig. 4F, right bar plot; p-value of 1). Furthermore, up-regulated genes were not associated with altered TAD borders upon Beaf-KD as compared to unchanged TAD borders (Supplem. Fig. S4F). This result suggests

that changes in LRCs associated with insulator permissiveness could impede gene expression, yet without changing the demarcation of TADs (see Discussion).

Insulator depletion leads to leakages and gene expression variability

Our data highlighted that leakages and gains in enhancer-promoter contacts may impede gene expression, notably leading to up-regulate numerous genes. To further characterize the changes in expression, we performed a single-cell multiomic approach aiming at integrating transcriptomic and chromatin accessibility data, using scRNA-seq and scATAC-seq (Fig. 5; see Materials and Methods). UMAP projections could not identify any cluster that was specifically associated with the control Luc-KD or the Beaf-KD condition (Fig. 5A), while the expression of *beaf-32* was efficiently suppressed in the knockdown condition (Supplem. Fig. S5A, compare right and left plots of the corresponding UMAP1/2). However, the depletion of insulator protein may not be sufficient to induce a stable reprogramming detected as a cluster. Nevertheless, scRNA-seq scored hundreds of up-regulated genes upon Beaf-KD compared to control, and importantly, up-regulated genes were enriched in genomic regions exhibiting higher leakages and LRCs, consistent with previous analyses (Fig. 5B, left matrix). This enrichment was confirmed by bootstrap analysis showing a distinct distribution of up-regulated genes with regions exhibiting the highest leakages compared to those with the lowest leakages (Fig. 5C, p-value < 2.2×10^{-16}). Moreover, although still significant, this association was far less pronounced for down-regulated genes (Fig. 5C, p-value < 1.9×10^{-2}), thus supporting a specific association of up-regulated genes with permissive insulators letting novel E-P LRCs emerge upon Beaf-KD, thereby potentially contributing to increased gene expression.

To determine whether gene up-regulation upon Beaf-KD results from the activation of new enhancers rather than from enhancer interactions released across permissive insulators, we mapped active enhancers by scATAC-seq in Beaf-KD and control conditions and assessed whether enhancer activation was associated with insulator permissiveness. In stark contrast with up-regulated genes, the associated scATAC-seq signal in Beaf-KD was not correlated with leakage (Fig. 5B, right matrix; Supplem. Fig. S5B). These results rule out the possibility that the new LRCs detected across permissive insulators arise from the activation of novel enhancers in the Beaf-KD condition.

The absence of defined clusters of cells in the Beaf-KD condition (Fig. 5A) suggested that the nature of the expression defects induced by such insulator protein depletion might be stochastic with respect to cell reprogramming scales, albeit predictable with respect to the localization of up-regulated genes, e.g. close to a permissive border. To further investigate the relationship between insulator permissiveness and transcriptional variability, we tested whether highly variable genes identified by scRNA-seq were preferentially associated with permissive insulator regions test this, we inspected the linkage between insulator permissiveness and gene expression variability as scored in scRNA-seq (see Materials and Methods). Strikingly, Beaf-KD led the genes subjected to insulator permissiveness to have a significantly higher variability in expression compared to genes in control regions with no permissiveness (Fig. 5D,

right plot; p-value < 2.2 x 10^{-16}). Moreover, such association was not found for the control Luc-KD condition (Fig. 5D, left plot; p-value = 0.592), thus supporting the view that upon Beaf-KD, gene expression changes may have a stochastic nature rather than corresponding to a coordinated transcriptional response. The increase in variability was specific to genes pertaining to loci exposed to insulator permissiveness, as genes not exposed to novel LRCs showed a significantly lower variability. Moreover, hierarchical clustering analysis showed that the genes being associated with the most permissive insulators form clusters, with cluster 3 being significantly enriched in up-regulated genes (Fig. 5E-F; p-value = 9 x $10^{-13,}$ Spearman correlation test), supporting an association between insulator permissiveness and up-regulation of genes within genomic regions with higher leakage. However, the variability of expression was not significantly higher in Beaf-KD than in Luc-KD (Supplem. Fig. S5C; p-value = 0.949; Wilcoxon test), indicating that the increased transcriptional variability is specific to genes located in loci exhibiting insulator permissiveness. Accordingly, genes present in such genomic regions define clusters enriched in up-regulated genes (Supplem. Fig. S5D-E), which are associated with Gene Ontology terms distinct from those previously identified for direct targets of dCTCF or BEAF-32 insulator proteins (Supplem. Fig. S5F) [68]. By blocking LRCs between enhancers and genes, insulators may thus contribute to the regulation of nearby genes and to the maintenance of gene expression homeostasis at adjacent loci by restricting promiscuous or stochastic enhancer-driven activation. Accordingly, diffuse and stochastic LRCs may enable enhancer activity to influence neighboring genes despite the relative stability of higher-order genome organization defined by TADs and compartments [34, 71, 72].

DISCUSSION

Our work strengthens the current view that, while Drosophila insulators have little or no influence on the structural organization of chromosomes into TADs, they still regulate long-range contacts (LRCs) across their binding sites, which impede gene expression. Notably, the HiCAggR package could quantify LRCs surrounding regulatory elements, thereby distinguishing insulators from enhancers. Upon insulator protein depletion, hundreds of novel LRCs are formed across insulator sites, including novel E-P contacts that are associated with gene up-regulation. These data support a model in which insulators can readily limit gene expression by restricting specific 3D interactions, without necessarily altering TAD structure.

How the 3D organization of chromosomes contributes to transcriptional regulation has been studied for decades, though often at limited resolution. Generally, Hi-C approaches have revealed the basic organization as a polymer that can support 3D folding [41], which in turn favors LRIs and more transient long-range contacts (LRCs) between two distal sites. Moreover, TADs bordered by insulators may serve as three-dimensional interaction hubs that organize regulatory landscapes [17] and can influence 3D contacts [32, 73–75]. As such, removal of insulator sites from the TAD border was shown to increase its permissiveness. At genome-wide scales, the quantification of insulation and its influence on genes requires sensitive, high-resolution approaches, or alternatively single cell and temporal analyses. A Hi-M approach actually revealed that TADs could serve to insulate genes, a phenomenon that was not directly influenced by transcription [65, 76]. Our results also support a genome-wide function of insulator

proteins that contributes to maintaining gene expression homeostasis, consistent with the functions of CTCF and additional insulator proteins in Drosophila and humans [35, 77, 78]. We notably show that depleting the insulator protein BEAF-32 leads to insulator permissiveness or leakage. Of note, this often results in the stochastic activation and increased expression variability of the genes that lie in the corresponding regions exposed to gains of LRCs, with novel enhancer-promoter contacts forming across the permissive insulators. Notably, integrated scRNA-seq and scATAC-seq analyses reveal that newly activated genes or genes exhibiting increased cell-to-cell transcriptional variability are not associated with the activation of novel enhancers, but rather with their novel LRCs enabling to reach out promoters across permissive insulators, i.e. from pre-existing enhancers. This regulatory rewiring may be constrained at BEAF-32 sites located at TAD boundaries, where they overlap with dCTCF sites. TAD borders tend to harbor sites of multiple insulator proteins, which may act redundantly [18, 19]. This would account for the persistence of TADs upon Beaf-KD e.g. if overlapping dCTCF sites are present. Of interest, similar conclusions were proposed from direct measurements by microscopy [76, 77]. Accordingly, the nature of the signal, leading to gene expression variability, may therefore not be linked to TAD or structural loops but rather to stochastic or transient 3D contacts. Beyond TADs, our data support models in which insulators regulate the frequencies of E-P contacts depending on the clustering into 3D hubs of multiple regulatory elements. Such 3D hubs connecting enhancers and promoters may already assemble prior to, and hence independently of, TAD formation [78].

Beyond structural organization, transient long-range contacts are essential to activate a gene through E-P interactions. Accordingly, LRCs lasting a few seconds were shown to suffice for an enhancer to transiently engage a promoter in the 3D space for gene activation [79, 80]. For genome-wide analyses, HiCAggR appears to be particularly suited for detecting rare and possibly transient long-range contacts, at the scale of individual loci. We actually find that leakage events are detected in the absence of insulator proteins, at BEAF-32 sites, and preferentially away from a TAD border, where additional insulator proteins like CTCF were absent. Remarkably, this confirms the specificity of the leakages related to local permissiveness of insulator sites. Altogether, our results support a model in which insulators contribute to gene expression homeostasis through the regulation of promiscuous long-range contacts, which likely participate to gene expression programming through development without altering the overall TAD structures.

MATERIALS AND METHODS

All experiments were performed with independent biological replicates per condition.

Cell culture, RNAi

Drosophila melanogaster Schneider 2 (S2) cells were cultured at 25°C in Schneider's Drosophila medium (21720-024, Thermo Fisher Scientific), supplemented with 10% heat-inactivated FBS (FB-1285/500, BioSera) and 1% penicillin/streptomycin (15140-122, Thermo Fisher Scientific). Double-stranded RNAs

(dsRNAs) were generated as previously described [18, 67] from PCR templates amplified using PrimeSTAR Max DNA Polymerase (TAKR045A, Takara), purified after agarose gel electrophoresis by NucleoSpin Clean-up Kit (740609.50, Macherey Nagel), and transcribed *in vitro* using the TranscriptAid T7 High Yield Transcription Kit (K0441, Thermo Fisher Scientific), according to the manufacturer's instructions. ssRNAs were purified with phenol-chloroform-isoamyl alcohol (24/24/1 volume), precipitated with isopropanol, and paired. dsRNA targeting the exogenous *luciferase* gene was used as control. For RNAi treatment, cells were seeded at 4 x 10^6 cells/mL 24 h before exposure to dsRNA (10 µg / 10^6 cells) in serum-free medium for 2 h at 25°C. Complete medium was then added and cells were harvested 3 days post-treatment.

Depletions were verified by RT-qPCR analysis as follows: total RNA was extracted using the RNeasy Plus Mini kit (QIAGEN, Cat#74134). Reverse transcription was performed on 1 µg of total RNA using the Maxima H Minus First Strand cDNA Synthesis kit using a mix of oligo(dT) and random hexamer primers. quantitative PCR (qPCR) was performed on a ViiA 7 Real-Time PCR System (Applied Biosystems). Samples were diluted (1/20) and standard curves were generated using pooled samples. Gene expression levels were analyzed using ViiA 7 Software (Applied Biosystems) software and normalized to *act5C* reference gene using the comparative Ct (ΔΔCt) method. Depletions were also verified by Western blot analysis as previously described (see [81]).

ChIP-seq

ChIP was performed on 5 × 10^6 cells per immunoprecipitation (IP). Cells were fixed with 0.8% formaldehyde (F1635, Sigma) for 10 min at room temperature (RT) and quenched with 130mM glycine for 5 min at RT. After washing with 1x PBS, cells were permeabilized for 20 min at RT with 500 µL of 1X PBS, containing 0,2% Triton X-100 and 10 mM sodium butyrate. After centrifugation (10 min at 3,000 rpm 4°C), cell pellets were washed with 1 mL of Lysis buffer (LB; 140 mM NaCl, 15 mM HEPES pH7.6, 1m M EDTA, 0,5 mM EGTA, 1% Triton X-100, 0.1% sodium deoxycholate, 0.5 mM DTT, 10 mM sodium butyrate, 1x protease inhibitor) and resuspended in 100 µL LB supplemented with 1% SDS and 0.5% N-lauroylsarcosyl for 30 minat RT. Chromatin was sonicated for 5 cycles of 30 seconds ON and 30 seconds OFF at 4°C with the Bioruptor Pico (Diagenode) to generate DNA fragments of approximately 200–500 bp. Protein A/G beads were prepared by washing them twice with LB containing 0.1% SDS and 0.5% N-lauroylsarcosyl (LB/SDS) followed by incubation with 0.1mg/mL BSA (A1595, Sigma) for 2 hat 4°C. After 3x washes with LB/SDS, the beads were incubated with 5 µg of antibodies overnight at 4°C. Prior to immunoprecipitation, chromatin samples were pre-cleared overnight at 4°C using BSA-blocked beads lacking antibody. Antibody-coupled beads were then washed four times with LB/SDS and 10% of each chromatin sample was reserved as input. Pre-cleared chromatin samples were incubated with 10 µL of antibody-coupled beads for 4 hat 4°C. The beads were washed four times with LB/SDS and twice with TE buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA). DNA was eluted in TE buffer supplemented with 0.1% SDS at 70°C with shaking at 1,000 rpm for 10 min. Cross-links were reversed overnight at 65°C. Samples were then sequentially treated with RNase A (0.01 mg/mL, for 1 hr at 37°C) and proteinase K (10 mg/mL, for 1 hat 55°C), and DNA was purified by phenol–chloroform extraction and ethanol precipitation. After

centrifugation at 16,000 x g for 30 min at 4°C, DNA pellets were washed with 70% ethanol, centrifuged at 16,000 x g for 10 min at 4°C, dried and resuspended in TE buffer. For sequencing, libraries were prepared using the Collibri PS DNA Library Prep Kit (Thermo Fisher, A38614196) and sequenced as paired-end 100-bp reads, generating approximately 15 million reads per sample. Adapter sequences were trimmed using cutadapt (v1.8.3 [82]) and sequencing quality was assessed using FastQC (v0.11.7 [83]) within the Trim Galore pipeline (v0.4.0 [84]). ChIP-seq reads (CTCF, H3K4me1 and H3K27ac) were then aligned to the Drosophila reference genome (dm6) using BWA (v0.7.15 [85]) and filtered using SAMtools (v1.3.1 [86]). Coverage files of aligned reads were generated using the bamCoverage tool from deepTools (v2.5.3 [87]) with a bin size of 10 bp and normalized using RPGC. Peak calling was performed using MACS2 ([88]; q-value < 0.05), using narrow peak options for H3K4me1 and H3K27ac. Additional H3K27ac, H3K4me1 and H3K4me3 datasets were retrieved from GEO (GSE85191 [89]). H3K27me3, H3K36me3 and BEAF-32 datasets were obtained from the Cuvier laboratory [18, 67].

Hi-C

Hi-C was performed on 2 × 10^6 cells (in control luciferase, Luc-KD, and BEAF-32, Beaf-KD, depletion conditions) cross-linked as described for ChIP using the Arima Genomics Hi-C kit, following the manufacturer's instructions. Purified DNA was sonicated (Bioruptor Pico, Diagenode; 7–8 cycles of 15 seconds ON / 90 seconds OFF). Libraries were prepared using the KAPA HyperPrep kit (Roche, KK8500) and sequenced as paired-end, 100-bp reads, generating approximately 100 million reads per sample (BGI Genomics). Data were processed using the Juicer pipeline (v1.6 [90]), BWA (v0.7.17-r1188), Juicer-tools (v1.19.02), openjdk (v1.8.0_312) and the HiCAggR package (v1.8.0) associated with this manuscript:

(https://www.bioconductor.org/packages/release/bioc/html/HicAggR.html).

Contact matrices were generated at 1, 5 and 10 kb resolution. Hi-C data (.hic files) were first processed using the HiCExplorer suite (v3.7.2 [59]). Contact matrices were converted to HDF5 (.h5) format and analyzed at multiple resolutions (1, 5, 10, 25 and 50 kb). TADs were identified using hicFindTADs with default parameters. For the insulation score at 5 kb resolution, thresholds and FDR correction for multiple testing cutoffs were applied using a minimum threshold of the difference between the TAD-separation score of 0.01 and a q-value threshold of 0.008. TAD visualization and comparison between conditions were performed using hicPlotTADs, hicMergeDomains, and hicDifferentialTAD (default settings). Further Hi-C data analysis were performed using the Bioconductor HiCAggR package (v1.8.0). After importing raw contact count matrices, normalization and observed/expected correction were performed using, respectively, BalanceHiC (ICE method) and OverExpectedHiC included functions.

RNA-seq and Single-Cell Multiome (scRNA-seq + scATAC-seq)

For RNA-seq, total RNA was extracted from 2 × 10^6 cells using TRIzol (Thermo Fisher) and sent to BGI genomics for RT and library preparation, using rRNA depletion, strand-specific library preparation and sequencing paired-end 100-bp reads, generating approx. 20 million reads/sample. Adapter sequences

were trimmed using cutadapt (v1.8.3 [82]) and sequencing quality was assessed using FastQC (v0.11.2 [83]) within the Trim Galore pipeline (v0.4.0 [84]). Reads were aligned to the Drosophila melanogaster genome dm6 using STAR (v2.5.2). Only uniquely mapped reads were retained and processed with SAMtools (v1.3.1 [86]). All analyses were performed using R (v4.1.2). Genomic data (.bed) were imported using rtracklayer and handled as GRanges objects with the GenomicRanges R/Bioconductor package. Active genes were defined from RNA-seq data as regions with normalized read counts > 50pb in both control (Luc-KD) and BEAF-32 knockdown (Beaf-KD) conditions (n = 8643 genes). Differential expression analysis was performed using DESeq2 (v1.34.0; [91]) with the Wald test and log2 fold change shrinkage (lfcShrink). Genes with p-value < 0.05 were considered differentially expressed; log2FC > 0 defined up-regulated genes (n = 437) and log2FC < 0 down-regulated genes (n = 312).

Single-Cell multiomics was performed at GenScan facility (CBI Toulouse) using the Chromium Single Cell Multiome ATAC + Gene Expression kit (10x Genomics), with two replicates per condition. Libraries were sequenced by BGI Genomics and processed with the Cell Ranger ARC pipeline. After initial quality filtering with Cell Ranger ARC, 38,373 cells were retained. Sequencing depth averaged approximately 26,000 ATAC read pairs and approx. 19,000 RNA reads per cell. Single-cell data were processed using Seurat (v5.4.0) and Signac (v1.16.0). Data were normalized using SCTransform, and dimensionality reduction was performed using PCA and LSI. Multimodal integration was performed using the weighted nearest neighbors (WNN) framework, followed by Louvain graph-based clustering (resolution = 0.5). Data were visualized using UMAP. ScRNA-seq and scATAC-seq data were imported using Read10X function and processed using standard Seurat and Signac workflows. scATAC-seq data were represented as ChromatinAssay using CreateChromatinAssay function with dm6 (BDGP6.46) annotations. Cells were retained based on the following thresholds (5%-95%); 1,149 – 21,448 unique molecular identifiers (UMIs), 395-2,817 detected genes, 896 – 31,544 fragments, and 340-7,762 peaks. Additional filtering criteria included mitochondrial content < 0.545, transcription start site (TSS) enrichment > 1.49, nucleosome signal < 1.28. An apoptosis-related gene expression score was computed using Seurat's AddModuleScore function based on a curated list of apoptosis-associated genes (Apoptosis score < 0). Cells outside these thresholds or identified as doublets using Scrublet (6%) were excluded. After applying the additional QC criteria, 24,917 cells were retained, 24,917 cells were retained. scRNA-seq data were normalized using SCTransform with regression of mitochondrial content and sequencing depth, followed by PCA. scATAC-seq data were processed using TF-IDF normalization and latent semantic indexing (LSI), excluding the first component. Data were embedded using UMAP on the first 30 RNA principal components and LSI dimensions. Multimodal integration was performed using Seurat's weighted nearest neighbor (WNN) framework, followed by Louvain graph-based clustering (resolution = 0.5). Gene expression and chromatin accessibility were visualized using FeaturePlot. Highly variable genes were selected based on gene expression variance computed across both Luc-KD and Beaf-KD cells. Correlations between genes at the most permissive sites (top 5%) across all cells were assessed using Spearman's rank correlation test. Hierarchical clustering was performed on the Spearman correlation matrix of the top 5% genes associated with leakage regions to identify gene clusters (k = 4).

Enrichment of up-regulated genes within clusters was evaluated using Fisher's exact test, with statistical significance defined as p-value < 0.05.

Aggregation of Hi-C data

APA was performed using HiCAggR (v1.8.0). Genomic features (TSS, BEAF-32 peaks, and enhancers defined by H3K4me1 and H3K27ac marks) were indexed at 1 kb resolution. Interactions between anchor and bait regions were identified within a genomic distance of 40–200 kb. Generation of Hi-C submatrices consists of extracting Hi-C contacts surrounding the bin corresponding to couples of anchors. As anchor1 at position *i*, anchor2 at position *j*, a squared matrix, called submatrix in this paper, of 41x41 bins around positions *i* and *j* in the full chromosome matrix were extracted and stored with metadata such as protein or features associated to the corresponding anchors. For a 1kb resolution Hi-C contact map, submatrix corresponded to a square centered on putative interaction between bin anchor1 and bin anchor2, and spanning on +/- 20kb horizontally and vertically (20bins x 1kb upstream, 1kb anchored bin, 20bins x 1kb downstream). Each matrix was then converted to an integer between 1 and 500, i.e. vectorized between 1 and 500 and assigned to submatrix values according to their rank (after filtering non-attributed values) relative to the contacts in the regions surrounding anchors. To align all submatrices, with first and second anchors always kept in that order, the submatrices with inversed anchors were reoriented (consisting of a flip of the matrix along the APA diagonal), before aggregation using median quantilized signal. Contact pairs were filtered to keep intrachromosomal contacts with the distances between the two anchors ranging from 40kb to 200kb. Single locus quantification was done by quantifying contacts for each submatrix after KR normalization, and before aggregation. Insulator permissiveness or leakage signal was quantified on the upper-left 18x18 square of the APA. Quantifications were statistically analyzed using ggplot2 and statistical functions in R (v4.3.1) using the HiCAggR GetQuantif function. Significant novel enhancer-promoter LRCs were selected with the CompareToBackground function, using a z-test (p-value < 0.05). Leakage scores were computed from Hi-C interaction matrices for BEAF-32 promoter pairs defined under control (Luc-KD) and BEAF-32 depletion (Beaf-KD) conditions. A normalized leakage score was defined as the ratio of interaction intensity relative to background, in Beaf-KD relative to Luc-KD. Pairs were ranked by decreasing leakage score and ranked into 20 quantiles (5% each), with group 1 corresponding to highest leakage and group 20 to the lowest leakage. Leakage scores were computed genome-wide and associated with genomic features, including BEAF-32 binding sites and enhancers. Receiver operating characteristic (ROC) analysis [92] was performed in R (v4.3.1) using a custom function available from the Cuvier lab GitHub repository to evaluate the ability of leakage scores to discriminate BEAF-32 binding sites from enhancer regions from APA matrices BEAF-32-active TSS and Enhancer-active TSS respectively. Performance was assessed using area under the curve (AUC) values derived from true and false positive rates across thresholds.

Bootstrap analysis of gene enrichment in leakage regions

Bootstrap resampling (1,000 iterations) was used to assess the distribution of gene sets across leakage regions. Up-regulated genes (bulk RNA-seq, n = 437; single-cell RNA-seq, n = 229) and highly variable genes (Luc-KD : n = 2,551, Beaf-KD : n = 1,264) were analyzed separately. At each iteration, 100 genes

from the up-regulated gene set or 500 genes from the highly variable gene set, were randomly sampled from the corresponding gene set, and genes associated with leakage regions (group top 5%, Q1 or group lowest 5%, Q20) was computed. The resulting distributions were used to evaluate whether these gene sets were preferentially associated with leakage regions.

Data availability

All sequencing data, including replicates of Hi-C, of scATAC-seq and scRNA-seq and RNA-seq, in Beaf-KD and Luc-KD, have been deposited in GEO under accession GSE338858.

All computational code is available at the CuvierLab GitHub: https://github.com/CuvierLab/ under the repositories 'https://github.com/CuvierLab/Insulators_regulating_3D_long-range_contacts'

Abbreviations

Genome-wide 3C/Hi-C
Chromosome Conformation Capture and high throughput sequencing
APA
Aggregate peak analysis
TADs
Topologically associating domains
CTCF
CCCTC-binding protein
BEAF-32
Boundary-Element Associated Factor of 32 KDa
Enhancer-promoter (long-range) contacts
LRI
long-range interaction

Declarations

Running title

Hi-C aggregation with the HiCAggR package to quantify long-range contacts

DISCLOSURE DECLARATION

The authors declare no conflict of interest regarding the publication of this work.

ETHICS, CONSENT

Ethics, Consent to Participate, and Consent to Publish declarations: not applicable.

Author Contribution

N.S. performed all bioinformatics analyses and data integration with the help of N.C, E.E, P.M. and R.T. for starting APA analysis and S.S. for aligning and checking Hi-C data. O.F. did the Hi-C analyses, and C.P. performed the sc-multiomics with the help of A-L.F. O.C. designed the analysis with N.S., supervised the work and wrote the manuscript.

Acknowledgement

We thank the Genscan platform of the CBI and in particular Thomas Clouaire for help with the processing of multiomic data, and Marion Aguirrebengoa for her kind advice to N.S. We also thank Fernando Muzoppapa and Fabian Erdel for critical reading of the manuscript. The laboratory of O. Cuvier was supported by ANR grant (HELICO), the MSD-Avenir program (Hit Hidden HBV), a fellowship of the Ministry of Research Science and Technology (MRST; O.F.), the INSERM (S.S and OC) and the CNRS (OC's lab).

Data Availability

All sequencing data, including replicates of Hi-C, of scATAC-seq and scRNA-seq and RNA-seq, in Beaf-KD and Luc-KD, have been deposited in GEO under accession GSE338858.All computational code is available at the CuvierLab GitHub: https://github.com/CuvierLab/ under the repositories 'https://github.com/CuvierLab/Insulators_regulating_3D_long-range_contacts'

References

1. Beagan JA, Phillips-Cremins JE. On the existence and functionality of topologically associating domains. Nat Genet. 2020;52:8–16. https://doi.org/10.1038/s41588-019-0561-1.
2. Cremer M, Schmid VJ, Kraus F, Markaki Y, Hellmann I, Maiser A, et al. Initial high-resolution microscopic mapping of active and inactive regulatory sequences proves non-random 3D arrangements in chromatin domain clusters. Epigenetics Chromatin. 2017;10:39. https://doi.org/10.1186/s13072-017-0146-0.
3. Dekker J, Misteli T. Long-Range Chromatin Interactions. Cold Spring Harb Perspect Biol. 2015;7:a019356. https://doi.org/10.1101/cshperspect.a019356.
4. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature. 2012;485:376–80. https://doi.org/10.1038/nature11082.

5. Dostie J, Bickmore WA. Chromosome organization in the nucleus - charting new territory across the Hi-Cs. Curr Opin Genet Dev. 2012;22:125–31. https://doi.org/10.1016/j.gde.2011.12.006.
6. Hsieh T-HS, Cattoglio C, Slobodyanyuk E, Hansen AS, Rando OJ, Tjian R, et al. Resolving the 3D Landscape of Transcription-Linked Mammalian Chromatin Folding. Mol Cell. 2020;78:539–e5538. https://doi.org/10.1016/j.molcel.2020.03.002.
7. Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature. 2012;485:381–5. https://doi.org/10.1038/nature11049.
8. Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell. 2014;159:1665–80. https://doi.org/10.1016/j.cell.2014.11.021.
9. Seitan VC, Faure AJ, Zhan Y, McCord RP, Lajoie BR, Ing-Simmons E, et al. Cohesin-based chromatin interactions enable regulated gene expression within preexisting architectural compartments. Genome Res. 2013;23:2066–77. https://doi.org/10.1101/gr.161620.113.
10. Sexton T, Yaffe E, Kenigsberg E, Bantignies F, Leblanc B, Hoichman M, et al. Three-dimensional folding and functional organization principles of the Drosophila genome. Cell. 2012;148:458–72. https://doi.org/10.1016/j.cell.2012.01.010.
11. Szabo Q, Bantignies F, Cavalli G. Principles of genome folding into topologically associating domains. Sci Adv. 2019;5:eaaw1668. https://doi.org/10.1126/sciadv.aaw1668.
12. Vara C, Paytuví-Gallart A, Cuartero Y, Dily FL, Garcia F, Salvà-Castro J, et al. Three-Dimensional Genomic Structure and Cohesin Occupancy Correlate with Transcriptional Activity during Spermatogenesis. Cell Rep Elsevier. 2019;28:352–e3679. https://doi.org/10.1016/j.celrep.2019.06.037.
13. Vogelmann J, Le Gall A, Dejardin S, Allemand F, Gamot A, Labesse G, et al. Chromatin insulator factors involved in long-range DNA interactions and their role in the folding of the Drosophila genome. PLoS Genet. 2014;10:e1004544. https://doi.org/10.1371/journal.pgen.1004544.
14. Van Bortle K, Nichols MH, Li L, Ong C-T, Takenaka N, Qin ZS, et al. Insulator function and topological domain border strength scale with architectural protein occupancy. Genome Biol. 2014;15:R82. https://doi.org/10.1186/gb-2014-15-5-r82.
15. Bhattacharya M, Lyda SF, Lei EP. Chromatin insulator mechanisms ensure accurate gene expression by controlling overall 3D genome organization. Curr Opin Genet Dev. 2024;87:102208. https://doi.org/10.1016/j.gde.2024.102208.
16. Cai H, Levine M. Modulation of enhancer-promoter interactions by insulators in the Drosophila embryo. Nature. 1995;376:533–6. https://doi.org/10.1038/376533a0.
17. Heurteau A, Perrois C, Depierre D, Fosseprez O, Humbert J, Schaak S, et al. Insulator-based loops mediate the spreading of H3K27me3 over distant micro-domains repressing euchromatin genes. Genome Biol. 2020;21:193. https://doi.org/10.1186/s13059-020-02106-z.

18. Liang J, Lacroix L, Gamot A, Cuddapah S, Queille S, Lhoumaud P, et al. Chromatin immunoprecipitation indirect peaks highlight long-range interactions of insulator proteins and Pol II pausing. Mol Cell. 2014;53:672–81. https://doi.org/10.1016/j.molcel.2013.12.029.
19. Nègre N, Brown CD, Shah PK, Kheradpour P, Morrison CA, Henikoff JG, et al. A comprehensive map of insulator elements for the Drosophila genome. PLoS Genet. 2010;6:e1000814. https://doi.org/10.1371/journal.pgen.1000814.
20. Phillips JE, Corces VG. CTCF: Master Weaver of the Genome. Cell. 2009;137:1194–211. https://doi.org/10.1016/j.cell.2009.06.001.
21. Van Bortle K, Ramos E, Takenaka N, Yang J, Wahi JE, Corces VG. Drosophila CTCF tandemly aligns with other insulator proteins at the borders of H3K27me3 domains. Genome Res. 2012;22:2176–87. https://doi.org/10.1101/gr.136788.111.
22. Gaszner M, Vazquez J, Schedl P. The Zw5 protein, a component of the scs chromatin domain boundary, is able to block enhancer–promoter interaction. Genes Dev Cold Spring Harbor Lab. 1999;13:2098–107.
23. Sanborn AL, Rao SSP, Huang S-C, Durand NC, Huntley MH, Jewett AI, et al. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. Proc Natl Acad Sci U S A. 2015;112:E6456–65. https://doi.org/10.1073/pnas.1518552112.
24. Fudenberg G, Imakaev M, Lu C, Goloborodko A, Abdennur N, Mirny LA. Formation of Chromosomal Domains by Loop Extrusion. Cell Rep. 2016;15:2038–49. https://doi.org/10.1016/j.celrep.2016.04.085.
25. Merkenschlager M, Nora EP. CTCF and Cohesin in Genome Folding and Transcriptional Gene Regulation. Annu Rev Genom Hum Genet. 2016;17:17–43. https://doi.org/10.1146/annurev-genom-083115-022339.
26. Wutz G, Várnai C, Nagasaka K, Cisneros DA, Stocsits RR, Tang W, et al. Topologically associating domains and chromatin loops depend on cohesin and are regulated by CTCF, WAPL, and PDS5 proteins. EMBO J. 2017;36:3573–99. https://doi.org/10.15252/embj.201798004.
27. Rao SSP, Huang S-C, St Hilaire BG, Engreitz JM, Perez EM, Kieffer-Kwon K-R, et al. Cohesin loss eliminates all loop domains. Cell. 2017;171:305–e32024. https://doi.org/10.1016/j.cell.2017.09.026.
28. Nora EP, Goloborodko A, Valton A-L, Gibcus JH, Uebersohn A, Abdennur N, et al. Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. Cell. 2017;169:930–e94422. https://doi.org/10.1016/j.cell.2017.05.004.
29. Schwarzer W, Abdennur N, Goloborodko A, Pekowska A, Fudenberg G, Loe-Mie Y, et al. Two independent modes of chromatin organization revealed by cohesin removal. Nature. 2017;551:51–6. https://doi.org/10.1038/nature24281.
30. Chathoth KT, Mikheeva LA, Crevel G, Wolfe JC, Hunter I, Beckett-Doyle S, et al. The role of insulators and transcription in 3D chromatin organization of flies. Genome Res. 2022;32:682–98. https://doi.org/10.1101/gr.275809.121.

31. Lupiáñez DG, Kraft K, Heinrich V, Krawitz P, Brancati F, Klopocki E, et al. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. Cell. 2015;161:1012–25. https://doi.org/10.1016/j.cell.2015.04.004.
32. Arzate-Mejía RG, Josué Cerecedo-Castillo A, Guerrero G, Furlan-Magaril M, Recillas-Targa F. In situ dissection of domain boundaries affect genome topology and gene transcription in Drosophila. Nat Commun Nat Publishing Group. 2020;11:894. https://doi.org/10.1038/s41467-020-14651-z.
33. Cavalheiro GR, Girardot C, Viales RR, Pollex T, Cao TBN, Lacour P et al. CTCF, BEAF-32, and CP190 are not required for the establishment of TADs in early Drosophila embryos but have locus-specific roles. Sci Adv. 2023;9:eade1085. https://doi.org/10.1126/sciadv.ade1085
34. Bonev B, Cavalli G. Organization and function of the 3D genome. Nat Rev Genet. 2016;17:661–78. https://doi.org/10.1038/nrg.2016.112.
35. Kahn TG, Savitsky M, Kuong C, Jacquier C, Cavalli G, Chang J-M, et al. Topological screen identifies hundreds of Cp190- and CTCF-dependent Drosophila chromatin insulator elements. Sci Adv. 2023;9:eade0090. https://doi.org/10.1126/sciadv.ade0090.
36. Despang A, Schöpflin R, Franke M, Ali S, Jerković I, Paliou C, et al. Functional dissection of the Sox9-Kcnj2 locus identifies nonessential and instructive roles of TAD architecture. Nat Genet. 2019;51:1263–71. https://doi.org/10.1038/s41588-019-0466-z.
37. Ghavi-Helm Y, Jankowski A, Meiers S, Viales RR, Korbel JO, Furlong EEM. Highly rearranged chromosomes reveal uncoupling between genome topology and gene expression. Nat Genet. 2019;51:1272–82. https://doi.org/10.1038/s41588-019-0462-3.
38. Losada A, Hirano T. Dynamic molecular linkers of the genome: the first decade of SMC proteins. Genes Dev. 2005;19:1269–87. https://doi.org/10.1101/gad.1320505.
39. Banigan EJ, Mirny LA. The interplay between asymmetric and symmetric DNA loop extrusion. eLife 9:e63528. https://doi.org/10.7554/eLife.63528
40. Dekker J, Rippe K, Dekker M, Kleckner N. Capturing chromosome conformation. Volume 295. New York, N.Y.: Science; 2002. pp. 1306–11. https://doi.org/10.1126/science.1067799.
41. Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science. 2009;326:289–93. https://doi.org/10.1126/science.1181369.
42. Szabo Q, Donjon A, Jerković I, Papadopoulos GL, Cheutin T, Bonev B, et al. Regulation of single-cell genome organization into TADs and chromatin nanodomains. Nat Genet. 2020;52:1151–7. https://doi.org/10.1038/s41588-020-00716-8.
43. Szabo Q, Jost D, Chang J-M, Cattoni DI, Papadopoulos GL, Bonev B, et al. TADs are 3D structural units of higher-order chromosome organization in Drosophila. Sci Adv. 2018;4:eaar8082. https://doi.org/10.1126/sciadv.aar8082.
44. Barth R, Janissen R, Muras L, van der Torre J, Litos G, van der Sluis E, et al. Two CTCF motifs impede cohesin-mediated DNA loop extrusion. Mol Cell. 2025;85:4304–e43179. https://doi.org/10.1016/j.molcel.2025.11.001.

45. Chang L-H, Ghosh S, Papale A, Luppino JM, Miranda M, Piras V, et al. Multi-feature clustering of CTCF binding creates robustness for loop extrusion blocking and Topologically Associating Domain boundaries. Nat Commun. 2023;14:5615. https://doi.org/10.1038/s41467-023-41265-y.
46. Liu Y, Ding B, Zheng L, Xu P, Liu Z, Chen Z, et al. Regulatory elements can be essential for maintaining broad chromatin organization and cell viability. Nucleic Acids Res. 2022;50:4340–54. https://doi.org/10.1093/nar/gkac197.
47. Galouzis CC, Furlong EEM. Regulating specificity in enhancer–promoter communication. Curr Opin Cell Biol. 2022;75:102065. https://doi.org/10.1016/j.ceb.2022.01.010.
48. Zhang S, Übelmesser N, Barbieri M, Papantonis A. Enhancer-promoter contact formation requires RNAPII and antagonizes loop extrusion. Nat Genet. 2023;55:832–40. https://doi.org/10.1038/s41588-023-01364-4.
49. Alexander JM, Guan J, Li B, Maliskova L, Song M, Shen Y, et al. Live-cell imaging reveals enhancer-dependent Sox2 transcription in the absence of enhancer proximity. Elife. 2019;8:e41769. https://doi.org/10.7554/eLife.41769.
50. Hansen AS, Cattoglio C, Darzacq X, Tjian R. Recent evidence that TADs and chromatin loops are dynamic structures. Nucleus. 2017;9:20–32. https://doi.org/10.1080/19491034.2017.1389365.
51. Fraser J, Williamson I, Bickmore WA, Dostie J. An Overview of Genome Organization and How We Got There: from FISH to Hi-C. Microbiol Mol Biol Rev. 2015;79:347–72. https://doi.org/10.1128/MMBR.00006-15.
52. Hnisz D, Shrinivas K, Young RA, Chakraborty AK, Sharp PA. A phase separation model predicts key features of transcriptional control. Cell. 2017;169:13–23. https://doi.org/10.1016/j.cell.2017.02.007.
53. Sabari BR, Dall'Agnese A, Boija A, Klein IA, Coffey EL, Shrinivas K, et al. Coactivator condensation at super-enhancers links phase separation and gene control. Science. 2018;361:eaar3958. https://doi.org/10.1126/science.aar3958.
54. Rowley MJ, Lyu X, Rana V, Ando-Kuri M, Karns R, Bosco G, et al. Condensin II Counteracts Cohesin and RNA Polymerase II in the Establishment of 3D Chromatin Organization. Cell Rep. 2019;26:2890–e29033. https://doi.org/10.1016/j.celrep.2019.01.116.
55. Hansen KL, Adachi AS, Braccioli L, Kadvani S, Boileau RM, Martinovic M, et al. Synergy between regulatory elements can render cohesin dispensable for distal enhancer function. Volume 391. New York, N: Science; 2026. p. eadt4221. https://doi.org/10.1126/science.adt4221.
56. Bing X, Ke W, Fujioka M, Kurbidaeva A, Levitt S, Levine M, et al. In: Levine MT, Desplan C, editors. Chromosome structure in Drosophila is determined by boundary pairing not loop extrusion. Volume 13. eLife. eLife Sciences Publications, Ltd; 2024. p. RP94070. https://doi.org/10.7554/eLife.94070.
57. Fukaya T, Lim B, Levine M. Enhancer control of transcriptional bursting. Cell. 2016;166:358–68. https://doi.org/10.1016/j.cell.2016.05.025.
58. Chakraborty S, Kopitchinski N, Zuo Z, Eraso A, Awasthi P, Chari R, et al. Enhancer–promoter interactions can bypass CTCF-mediated boundaries and contribute to phenotypic robustness. Nat Genet. 2023;55:280–90. https://doi.org/10.1038/s41588-022-01295-6.

59. Ramírez F, Bhardwaj V, Arrigoni L, Lam KC, Grüning BA, Villaveces J, et al. High-resolution TADs reveal DNA sequences underlying genome organization in flies. Nat Commun Nat Publishing Group. 2018;9:189. https://doi.org/10.1038/s41467-017-02525-w.
60. Mourad R, Cuvier O. TAD-free analysis of architectural proteins and insulators. Nucleic Acids Res. 2018;46:e27. https://doi.org/10.1093/nar/gkx1246.
61. Kaushal A, Mohana G, Dorier J, Özdemir I, Omer A, Cousin P, et al. CTCF loss has limited effects on global genome architecture in Drosophila despite critical regulatory functions. Nat Commun. 2021;12:1011. https://doi.org/10.1038/s41467-021-21366-2.
62. Kellum R, Schedl P. A position-effect assay for boundaries of higher order chromosomal domains. Cell. 1991;64:941–50. https://doi.org/10.1016/0092-8674(91)90318-s.
63. Zhao K, Hart CM, Laemmli UK. Visualization of chromosomal domains with boundary element-associated factor BEAF-32. Cell. 1995;81:879–89. https://doi.org/10.1016/0092-8674(95)90008-x.
64. Mourad R, Li L, Cuvier O. Uncovering direct and indirect molecular determinants of chromatin loops using a computational integrative approach. PLoS Comput Biol. 2017;13:e1005538. https://doi.org/10.1371/journal.pcbi.1005538.
65. Götz M, Messina O, Espinola S, Fiche J-B, Nollmann M. Multiple parameters shape the 3D chromatin structure of single nuclei at the doc locus in Drosophila. Nat Commun. 2022;13:5375. https://doi.org/10.1038/s41467-022-32973-y.
66. Schoenfelder S, Fraser P. Long-range enhancer–promoter contacts in gene expression control. Nat Rev Genet Nat Publishing Group. 2019;20:437–55. https://doi.org/10.1038/s41576-019-0128-0.
67. Lhoumaud P, Hennion M, Gamot A, Cuddapah S, Queille S, Liang J, et al. Insulators recruit histone methyltransferase dMes4 to regulate chromatin of flanking genes. EMBO J. 2014;33:1599–613. https://doi.org/10.15252/embj.201385965.
68. Emberly E, Blattes R, Schuettengruber B, Hennion M, Jiang N, Hart CM, et al. BEAF Regulates Cell-Cycle Genes through the Controlled Deposition of H3K9 Methylation Marks into Its Conserved Dual-Core Binding Sites. PLoS Biol. 2008;6:e327. https://doi.org/10.1371/journal.pbio.0060327.
69. Jiang N, Emberly E, Cuvier O, Hart CM. Genome-wide mapping of boundary element-associated factor (BEAF) binding sites in Drosophila melanogaster links BEAF to transcription. Mol Cell Biol. 2009;29:3556–68. https://doi.org/10.1128/MCB.01748-08.
70. Maksimenko O, Bartkuhn M, Stakhov V, Herold M, Zolotarev N, Jox T, et al. Two new insulator proteins, Pita and ZIPIC, target CP190 to chromatin. Genome Res. 2015;25:89–99. https://doi.org/10.1101/gr.174169.114.
71. Phillips-Cremins JE, Corces VG. Chromatin insulators: linking genome organization to cellular function. Mol Cell. 2013;50:461–74. https://doi.org/10.1016/j.molcel.2013.04.018.
72. Rowley MJ, Corces VG. Organizational principles of 3D genome architecture. Nat Rev Genet. 2018;19:789–800. https://doi.org/10.1038/s41576-018-0060-8.
73. Li L, Lyu X, Hou C, Takenaka N, Nguyen HQ, Ong C-T, et al. Widespread rearrangement of 3D chromatin organization underlies Polycomb-mediated stress-induced silencing. Mol Cell.

2015;58:216–31. https://doi.org/10.1016/j.molcel.2015.02.023.

74. Ulianov SV, Khrameeva EE, Gavrilov AA, Flyamer IM, Kos P, Mikhaleva EA, et al. Active chromatin and transcription play a key role in chromosome partitioning into topologically associating domains. Genome Res. 2016;26:70–84. https://doi.org/10.1101/gr.196006.115.
75. Cubeñas-Potts C, Rowley MJ, Lyu X, Li G, Lei EP, Corces VG. Different enhancer classes in Drosophila bind distinct architectural proteins and mediate unique chromatin interactions and 3D architecture. Nucleic Acids Res. 2017;45:1714–30. https://doi.org/10.1093/nar/gkw1114.
76. Messina O, Raynal F, Gurgo J, Fiche J-B, Pancaldi V, Nollmann M. 3D chromatin interactions involving Drosophila insulators are infrequent but preferential and arise before TADs and transcription. Nat Commun. 2023;14:6678. https://doi.org/10.1038/s41467-023-42485-y.
77. Ren G, Jin W, Cui K, Rodrigez J, Hu G, Larson DR, et al. CTCF-mediated enhancer-promoter interaction is a critical regulator of cell-to-cell variation of gene expression. Mol Cell. 2017;67:1049–e10586. https://doi.org/10.1016/j.molcel.2017.08.026.
78. Espinola SM, Götz M, Bellec M, Messina O, Fiche J-B, Houbron C, et al. Cis-regulatory chromatin loops arise before TADs and gene activation, and are independent of cell fate during early Drosophila development. Nat Genet. 2021;53:477–86. https://doi.org/10.1038/s41588-021-00816-z.
79. Dowen JM, Fan ZP, Hnisz D, Ren G, Abraham BJ, Zhang LN, et al. Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. Cell. 2014;159:374–87. https://doi.org/10.1016/j.cell.2014.09.030.
80. Chen H, Levo M, Barinov L, Fujioka M, Jaynes JB, Gregor T. Dynamic interplay between enhancer-promoter topology and gene activity. Nat Genet. 2018;50:1296–303. https://doi.org/10.1038/s41588-018-0175-z.
81. Depierre D, Perrois C, Schickele N, Lhoumaud P, Abdi-Galab M, Fosseprez O, et al. Chromatin in 3D distinguishes dMes-4/NSD and Hypb/dSet2 in protecting genes from H3K27me3 silencing. Life Sci Alliance. 2023;6:e202302038. https://doi.org/10.26508/lsa.202302038.
82. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. 2011;17:10–2. https://doi.org/10.14806/ej.17.1.200.
83. Andrew. Andrews S. (2010) FastQC A Quality Control Tool for High Throughput Sequence Data. - References - Scientific Research Publishing [Internet]. 2010 [cited 2026 May 24]. https://www.scirp.org/reference/referencespapers?referenceid=4024153. Accessed 24 May 2026.
84. Krueger. Babraham Bioinformatics - Trim Galore! [Internet]. 2015 [cited 2026 May 24]. https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/. Accessed 24 May 2026.
85. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25:1754–60. https://doi.org/10.1093/bioinformatics/btp324.
86. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25:2078–9. https://doi.org/10.1093/bioinformatics/btp352.

87. Ramírez F, Dündar F, Diehl S, Grüning BA, Manke T. deepTools: a flexible platform for exploring deep-sequencing data. Nucleic Acids Res. 2014;42:W187–191. https://doi.org/10.1093/nar/gku365.
88. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-based Analysis of ChIP-Seq (MACS). Genome Biol. 2008;9:R137. https://doi.org/10.1186/gb-2008-9-9-r137.
89. Henriques T, Scruggs BS, Inouye MO, Muse GW, Williams LH, Burkholder AB et al. Widespread transcriptional pausing and elongation control at enhancers. Genes Dev [Internet]. Cold Spring Harbor Lab; 2018 [cited 2023 May 30]; https://doi.org/10.1101/gad.309351.117
90. Durand NC, Shamim MS, Machol I, Rao SSP, Huntley MH, Lander ES, et al. Elsevier. 2016;3:95–8. https://doi.org/10.1016/j.cels.2016.07.002. Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. cels.
91. Zhu A, Ibrahim JG, Love MI. Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. Bioinformatics. 2019;35:2084–92. https://doi.org/10.1093/bioinformatics/bty895.
92. Fawcett T. An introduction to ROC analysis. Pattern Recognit Lett. 2006;27:861–74. https://doi.org/10.1016/j.patrec.2005.10.010.

Figures

Figure 1

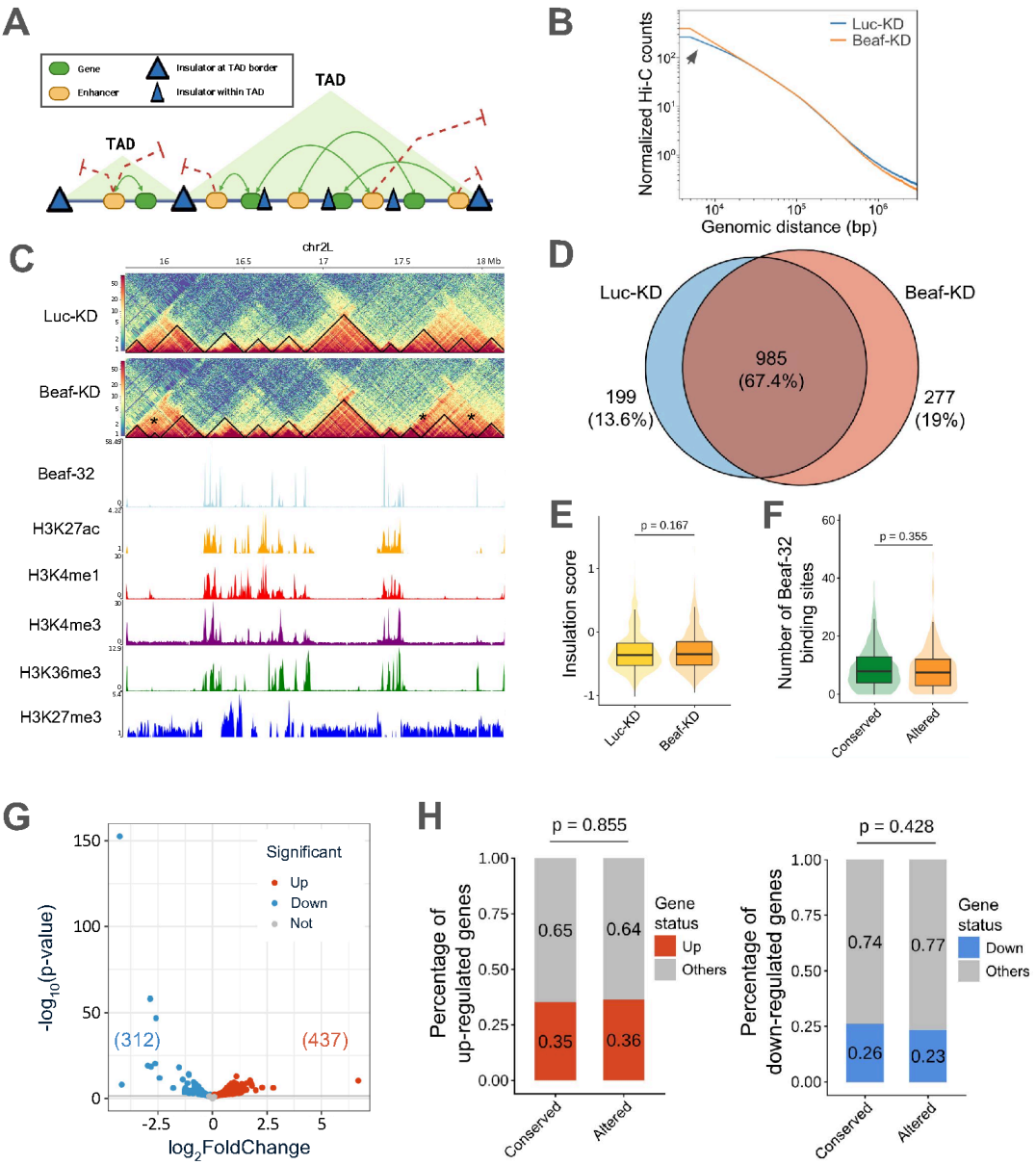

Figure 1

Insulator proteins do not influence gene expression through TADs.

A. Schematic view of functional long-range contacts between enhancers (orange ovals) and gene promoters (green ovals), depending on their co-localization within the same TAD (green arrows) or not (dotted red arrows). Insulators localizing at TAD borders (triangles in blue) may restrict the frequency of

3D contacts between enhancers in one TAD with promoters in the neighbor TADs. Note that insulators bound by CTCF or BEAF-32 may also be present within TADs (small triangles in blue) where they are thought to be non-functional (see text).

B. Plot showing the relative frequencies of interactions as a function of the distances between the interacting sites, as detected by Hi-C performed in BEAF-32 depleted cells (Beaf-KD) and in exogenous *luciferase* gene depleted control cells (Luc-KD). The arrow pinpoints relatively higher contacts in Beaf-KD for short distances.

C. 2D plots showing the landscape of long-range interactions (LRIs) as detected by Hi-C performed in luciferase control (Luc-KD; upper plot) or BEAF-32 depleted cells (Beaf-KD; lower plot) aligned with the averaged binding profiles as obtained by ChIP-seq of the insulator protein BEAF-32 or the histone post-translational modifications H3K27ac (active enhancers), H3K4me1 (enhancers), H3K4me3 (active promoters), H3K36me3 (active gene bodies) or H3K27me3 (inactive, Polycomb-associated chromatin) marks (see Materials and Methods). Note that BEAF-32 is found at TAD borders but also within TAD units. The asterisk marks an example of a region with a change of TAD between Beaf-KD and control cells.

D. Venn diagram showing the total number and percentages of TADs unique or common between Beaf-KD and control.

E. Boxplot showing the distribution of insulation score for BEAF-32 sites at TAD borders conserved between Beaf-KD and control. The non-significance of the difference was tested using a Wilcoxon test.

F. Boxplot showing the distribution of the number of BEAF-32 sites for TADs being conserved between Beaf-KD and control, or not. The non-significance of the difference was tested using a Wilcoxon test.

G. Volcano plot showing the differential gene expression levels as detected by DEseq2 (see Materials and Methods) from RNA-seq data in Beaf-KD compared to control cells.

H. Bar plot showing the proportion of genes being up-regulated (in red, left plot) or down-regulated (in blue, right plot) upon Beaf-KD compared to control cells, falling into TADs that change upon Beaf-KD compared to control. N.S., not significant enrichment of deregulated genes to be associated with changes in TADs upon Beaf-KD.

Figure 2

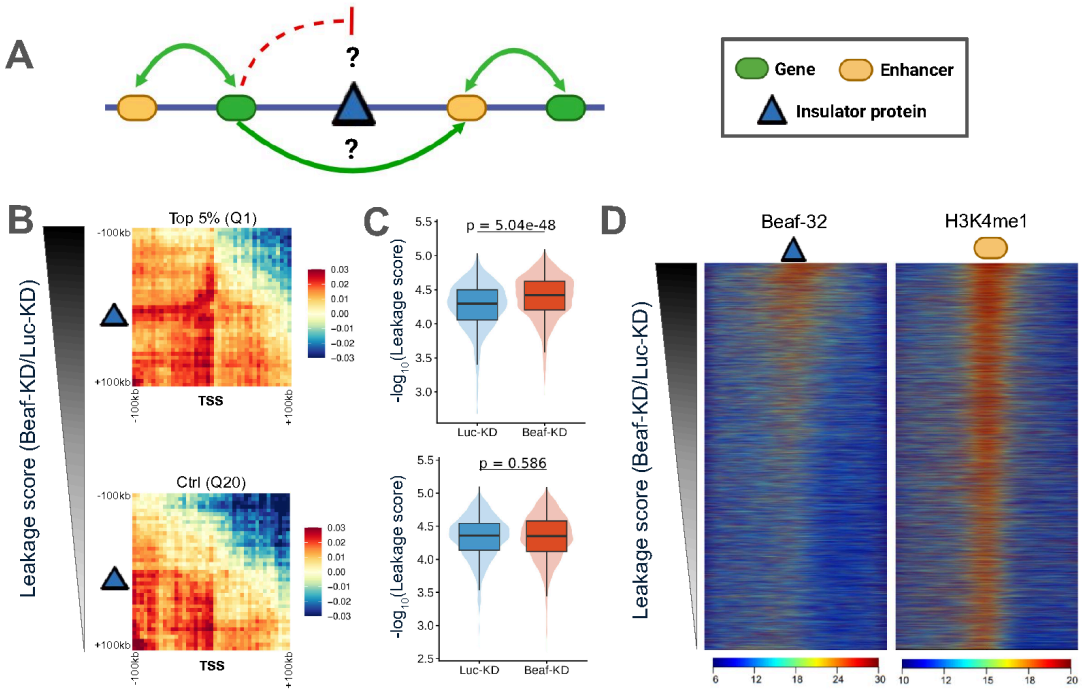

Figure 2

BEAF-32 depletion increases long-range contacts across insulator sites.

A. Scheme representing what was quantified by aggregation in panel B, namely whether long-range contacts form across insulator sites (blue triangle). Such long-range contacts may occur between any site including a distant enhancer (orange oval) and a gene promoter (green oval).

B. Quantification by quantilized Aggregation peak analysis (APA) of the long-range contacts occurring across regulatory element sites (enhancers or insulators). The aggregated Hi-C submatrices were ranked according to leakage (top left corner of the APA; see dotted black rectangle), as defined by the inscrease of long-range contacts occurring across the regulatory element site upon Beaf-KD as compared to control cells. The differential leakages were normalized to differences in the levels of interactions between Beaf-KD and control. The upper and lower APA panels show the top 5% and lowest levels of leakages, respectively (see Materials and Methods).

C. Boxplot showing the quantification of the normalized leakages in Beaf-KD as compared to control cells corresponding to the top- or lowest- 5% of all submatrices (2,717 per group). The significance was tested using a paired Wilcoxon test.

D**.** Heatmaps showing the ChIP-seq signal of BEAF-32 (left) or enhancer-associated H3K4me1 mark (right) after ranking all sites according to their differential leakage scores, as detected by Hi-C, as described in panels A and B.

Figure 3

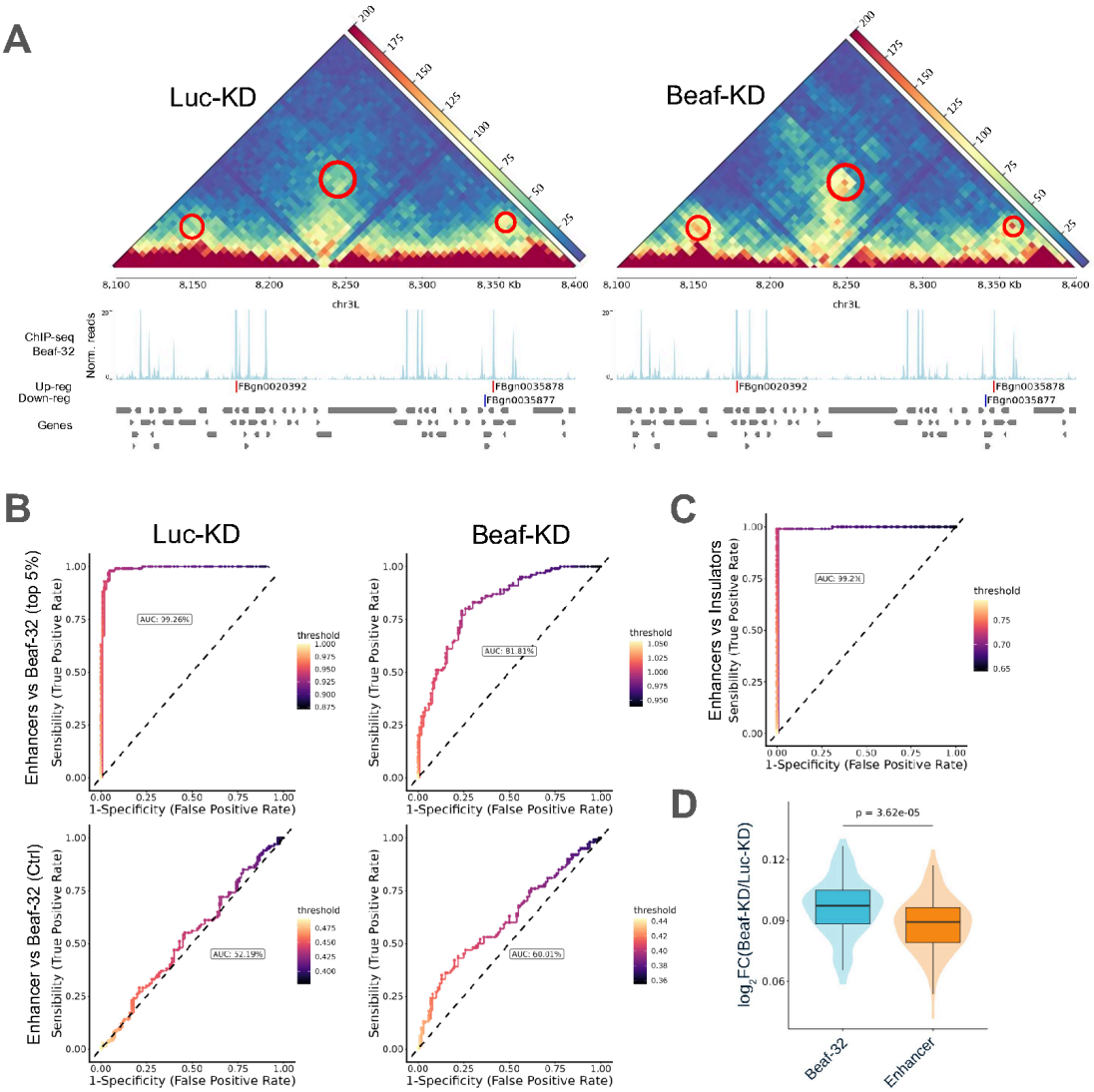

Figure 3

Detection of leakages at the level of single loci by HiCAggR distinguish insulators from enhancers.

A. 2D plot of one of the loci showing increased long-range contacts over an insulator site (BEAF-32 peak) upon Beaf-KD as compared to control cells.

B. Receiver operating characteristic (ROC) curves testing whether the differential levels of long-range contacts or leakage, as quantified by APA in Beaf-KD compared to control cells, as scored across all sites, allows to sort out insulators (defined by a BEAF-32 ChIP-seq peak) versus control enhancer sites. The ROC curves from top to bottom represent top 5% and lowest 5% respectively (AUC, Area under the Curve =99.26%).

C. ROC curve testing insulator-associated factors (dCTCF, CP190, Su(Hw), and BEAF-32 versus enhancers for their ability to block long-range contacts in control (Luc-KD) cells.

D. Boxplot showing the quantification of the leakages (log2FC signal) normalized to background, in Beaf-KD compared to control cells (see Materials and Methods), for the top 5% of BEAF-32 and Enhancer sites, respectively.

Figure 4

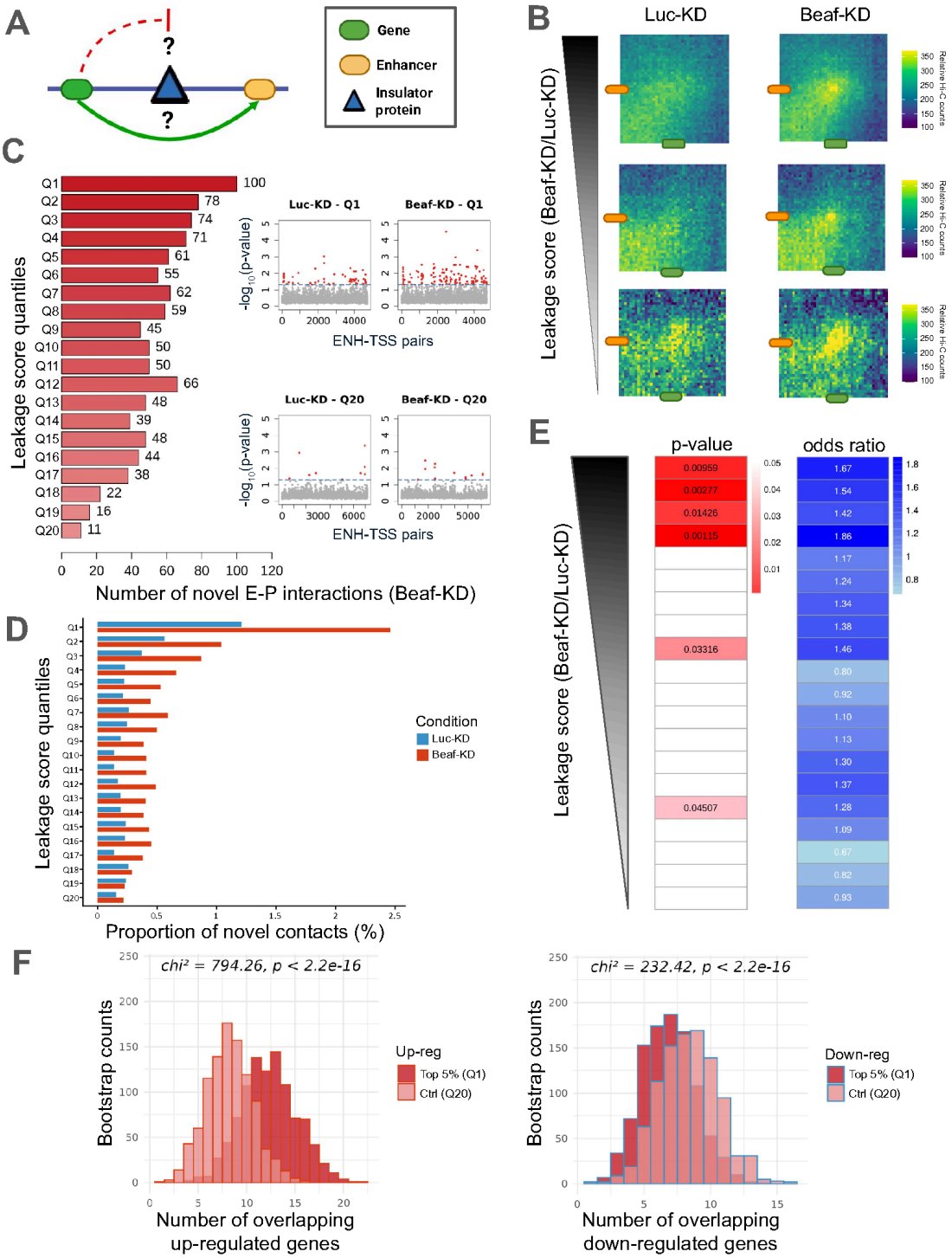

Figure 4

HiCAggR detects that Beaf-KD induces novel enhancer-promoter long-range contacts.

A. Scheme representing how Beaf depletion may lead to novel enhancer-promoter contacts across BEAF-32 sites, as compared to control.

B. Aggregation representing the novel enhancer-promoter contacts detected in Beaf-KD, as a function of high, medium or low leakage (upper, middle or lower APA, respectively) upon Luc-KD or Beaf-KD (left and right columns, respectively), across BEAF-32 sites. The novel significant enhancer-promoter contacts across BEAF-32 sites were detected systematically upon Beaf-KD or Luc-KD condition using HiCAggR (see Materials and Methods). Note that enhancer-promoter contacts are more easily detected upon Beaf-KD compared to Luc-KD, notably corresponding to regions showing highest leakage (top right APA).

C. Left: number of novel enhancer-promoter long-range contacts detected specifically in Beaf-KD, as detected in loci harboring high or low leakage (or insulator permissiveness). Right: scatter plot of the number of significant enhancer-promoter interactions in leakage regions for the top 5% (top) and lowest 5% (bottom) in control (left) and Beaf-KD (right) conditions.

D. Barplot representing the proportion of significant enhancer-promoter contacts identified using HiCAggR in Beaf-KD and control Luc-KD cells, relative to the total number of enhancer-promoter interactions.

E. Intersection matrix testing the enrichment of differentially up-regulated genes according to the extent of leakage defined as the enhanced long-range contacts across insulator sites upon Beaf-KD compared to control cells (as tested using a Fisher's exact test).

F. Distributions of up-regulated genes in regions with the highest leakage scores upon Beaf-KD as compared to regions with the lowest leakage scores, as assessed by a bootstrap-based chi-square test test (see Materials and Methods).

Figure 5

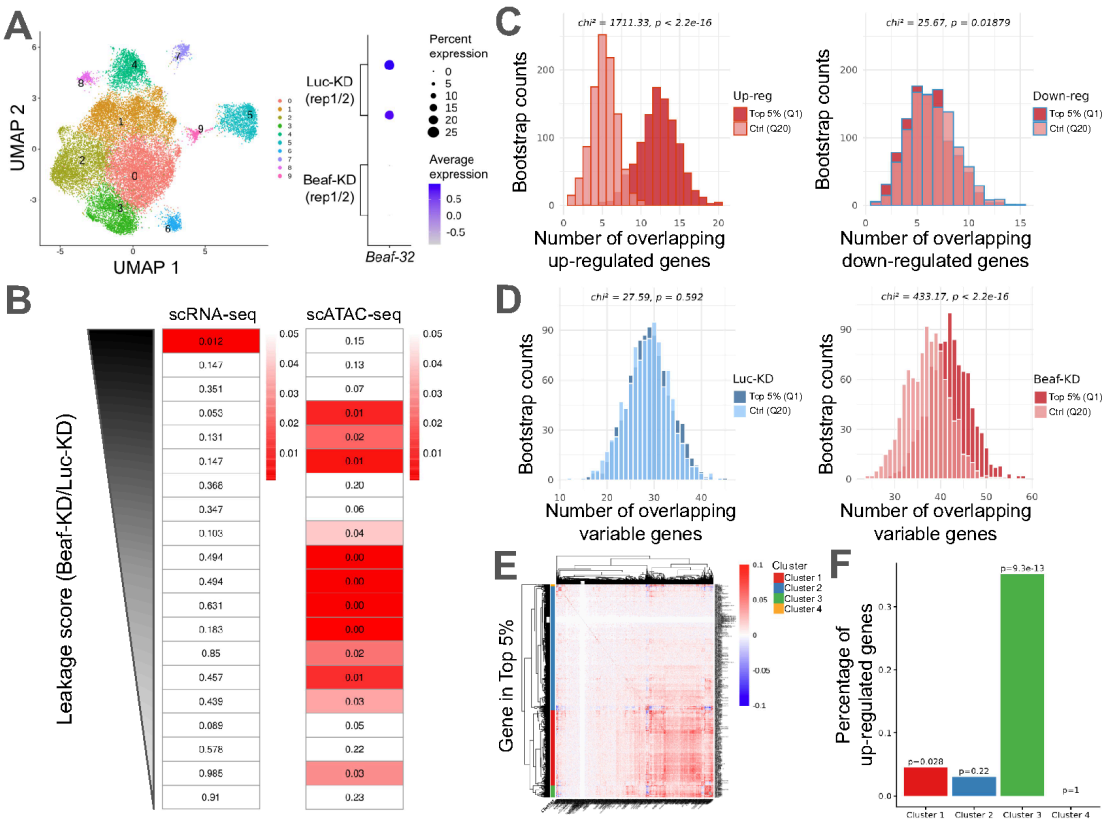

Figure 5

Single cell multiomics show that genes under the influence of leakages define clusters of up-regulated genes with higher variability in expression.

A. UMAP of single-cell RNA-seq analysis in Beaf-KD (upper scRNA-seq plot) compared to control (Luc-KD, lower scRNA-seq plot) along with the normalized expression of the *beaf-32* gene in each cell. The

expression plot on the right shows the averaged expression of *beaf-32* in each Beaf-KD (values close to 0) or control replicates (see Materials and Methods).

B. Intersection matrix testing the enrichment of differentially up-regulated genes or differentially active enhancers as obtained in single cell multiomics combining scRNA-seq and scATAC-seq in the same cells. The enrichments were tested according to the extent of leakage across insulator sites upon Beaf-KD compared to control cells.

C. Bootstrap analysis of the distribution of up- and down-regulated genes (left and right plots, respectively) within the regions exhibiting the highest leakage (top 5%) compared with regions with the lowest leakage (lowest 5%) (see Materials and Methods).

D. Bootstrap analysis of the distribution of highly variable genes within loci exhibiting the highest insulator permissiveness (top 5%) compared with loci exhibiting the lowest insulator permissiveness (lowest 5%), in Luc-KD (blue bars) and Beaf-KD (red bars) (see Materials and Methods).

E. Heatmap showing the Spearman correlation coefficients calculated for genes in top 5% of regions displaying leakages (Log Beaf-KD/control ratio) after hierarchical clustering.

F. Bar plot showing the enrichment of up-regulated genes across the 4 clusters identified in the correlation heatmap. Only cluster 1 and cluster 3 showed a significant enrichment (Fisher's exact test, p-values = 2.8×10^{-2} and 9.3×10^{-13} respectively).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- TableS1.csv
- SchickeleetalSupplem.pdf