



Chromatin folding motifs define the basic grammar of TAD architecture in single alleles

Olivier Messina, Jean-Bernard Fiche, Christel Elkhoury Youhanna, Gautham Ganesh, Yasmine Kemkem, Alexandre Mesnager, Ildem Akerman, David J Hodson, Marie Schaeffer, Marcelo Nollmann

► To cite this version:

Olivier Messina, Jean-Bernard Fiche, Christel Elkhoury Youhanna, Gautham Ganesh, Yasmine Kemkem, et al.. Chromatin folding motifs define the basic grammar of TAD architecture in single alleles. 2025. <hal-05358689>

HAL Id: hal-05358689

<https://hal.science/hal-05358689v1>

Preprint submitted on 11 Nov 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Chromatin folding motifs define the basic grammar of TAD architecture in single alleles

Olivier Messina¹, Jean-Bernard Fiche¹, Christel Elkhoury Youhanna¹, Gautham Ganesh¹, Yasmine Kemkem², Alexandre Mesnager¹, Ildem Akerman^{3,4}, David J. Hodson⁵, Marie Schaeffer^{1*}, Marcelo Nollmann^{1,6,*}

Affiliations

¹ *Centre de Biologie Structurale, Univ Montpellier, CNRS UMR 5048, INSERM U1054, 34090 Montpellier, France.*

² *Centre for Craniofacial and Regenerative Biology, King's College London, London, UK.*

³ *Institute of Metabolism and Systems Research (IMSR), University of Birmingham, Birmingham, UK.*

⁴ *Centre for Endocrinology, Diabetes and Metabolism, Birmingham Health Partners, Birmingham, UK.*

⁵ *Oxford Centre for Diabetes, Endocrinology and Metabolism (OCDEM), NIHR Oxford Biomedical Research Centre, Churchill Hospital, Radcliffe Department of Medicine, University of Oxford, Oxford, UK.*

Keywords: Single-Cell Imaging, Mammalian Tissues, Chromatin Architecture, Chromatin Folding Motifs, Spatial Genomics.

*Corresponding authors: marie.schaeffer@inserm.fr, marcelo.nollmann@cbs.cnrs.fr

⁶ Lead author: marcelo.nollmann@cbs.cnrs.fr

Abstract

Chromatin structure varies dramatically among single cells, yet the specific motifs underpinning this organization remain poorly understood. Using imaging-based spatial genomics, we visualized 3D chromatin organization across seven mouse tissues. We demonstrate that single chromatin traces can be decomposed into loop and domain Chromatin Folding Motifs (CFMs). Polymer simulations revealed that loop-extrusion reproduces loop CFMs, while cis-regulatory interactions are primarily responsible for domain CFMs. Notably, a unique set of CFMs describes single-cell chromatin architecture across multiple cell types, underscoring the universality of CFMs in organizing chromatin. In the pancreas, cells with different transcriptional programs exhibited distinct CFM usage, highlighting the functional relevance of CFMs. Perturbation of transcription and CTCF levels in a type 2 diabetes model altered CFM usage, demonstrating that disease induction affects both architectural and cis-regulatory interactions. Our results uncover how chromatin is organized in single cells and reveal the alterations that occur during disease.

Introduction

Eukaryotic chromosomes are organized at multiple scales, with topologically associating domains (TADs) (Dixon et al. 2012; Nora et al. 2012; Sexton et al. 2012) encapsulating genes and most of their *cis*-regulatory elements (Schwarzer and Spitz 2014; de Laat and Duboule 2013). In mammals, TADs are formed by the combined action of cohesin-mediated loop extrusion (Nasmyth 2001; Fudenberg et al. 2016; Davidson and Peters 2021) and interactions between *cis*-regulatory elements (CRE), such as enhancers (E) and promoters (P) (Furlong and Levine 2018; Schoenfelder and Fraser 2019; Vermunt, Zhang, and Blobel 2019; Hafner and Boettiger 2023). Loop extrusion itself can also bring enhancers and promoters into close proximity to regulate transcription (Karpinska and Oudelaar 2023). However, how loop extrusion and *cis*-regulatory interactions impact TAD architecture and transcriptional regulation in single cells remains an open question (Schaeffer and Nollmann 2023).

Initially, TADs were viewed as stable structural units essential for maintaining proper gene regulation (Dixon et al. 2015). Subsequent single-cell studies have shown that TAD architecture displays considerable structural variability (Giorgetti et al. 2014; Cattoni et al. 2017; Flyamer et al. 2017; Bintu et al. 2018; Finn et al. 2019), possibly due to the internal reconfiguration of nano-compartments (Ricci et al. 2015; Szabo et al. 2020; Aljahani et al. 2022). Modeling studies demonstrated that loop extrusion is a highly dynamic process, giving rise to highly heterogeneous TAD architectures (Fudenberg et al. 2016; Conte et al. 2022), with more recent live imaging evidence showing that CCCTC factor (CTCF)-mediated loops are infrequent and short-lived (Gabriele et al. 2022; Mach et al. 2022). Additionally, multiple studies have revealed that enhancer-promoter proximity occurs at low frequencies and with low specificity (Mateo et al. 2019; Espinola et al. 2021; Xiao, Hafner, and Boettiger 2021) and is highly dynamic (Zuin et al. 2022; H. Chen et al. 2018; Alexander et al. 2019; Li et al. 2020). Thus, the two main mechanisms sculpting TAD structure are intrinsically dynamic and give rise to highly heterogeneous TAD architectures in single snapshots of time. Yet, it is unclear if common structural patterns in single cells exist, if and how they change between cell types, and how they contribute to cell type-specific transcriptional regulation.

In this study, we addressed this question using multiplexed DNA-FISH imaging and machine learning tools. We reveal that 3D chromatin organization in single cells can be decomposed by a discrete set of chromatin folding motifs. These motifs can robustly describe the 3D chromatin structure across multiple cell types, and can be directly linked to loop extrusion and *cis*-regulatory interactions. Furthermore, we found that different cell types utilize chromatin folding motifs distinctly, influencing chromatin organization and transcription. Importantly, we showed that perturbation of cell-type specific transcription and CTCF levels significantly alter chromatin folding motif usage.

Results

Single-allele 3D TAD structures can be decomposed into discrete structural motifs

To explore heterogeneity in 3D chromatin organization in adult mammalian tissues, we used Hi-M on cryosectioned mouse pancreas (**Fig. 1a**, Methods) (Cardozo Gizzi et al. 2019). Specifically, we focused on a genomic region spanning 625 kb around the pancreatic and duodenal homeobox 1 gene (*Pdx1*), a transcription factor essential for pancreas specification/development and beta cell function (Ahlgren et al. 1998; Oliver-Krasinski et al. 2009; Offield et al. 1996). We covered the *Pdx1* locus using an oligopaint library of 25 Hi-M primary barcodes, each spanning on average 25 kb (**Fig. 1b**, Supplementary Data 1). Image acquisition, analysis and validation were performed as previously described (Cardozo Gizzi et al. 2020; Espinola et al. 2021) (**Figs. S1a-f**, and Methods). The proximity frequency and ensemble pairwise distance (PWD) maps of the *Pdx1* locus display a single TAD with multiple overlapping loops and sub-TADs (**Figs. 1b, S1g**). Notably, these structural features are also clearly distinguishable in ensemble Hi-C maps for the same locus (**Fig. S1h**), which correlate well with Hi-M proximity frequencies (**Fig. S1i**).

To discover and characterize patterns of chromatin organization within individual traces, we developed 3DTopic –an unsupervised Bayesian framework based on topic modeling using latent Dirichlet allocation (Methods). 3DTopic decomposes single 3D chromatin traces into a discrete set of orthogonal 3D Chromatin Folding Motifs (**CFMs**), where ‘motif’ refers to a recurring 3D chromatin folding pattern, rather than a DNA sequence (**Figs. 1c-d**). Therefore, CFMs capture common, unique, and non-overlapping components of the chromatin folding landscape explored by experimental 3D chromatin structures (**Figs. 1e, S2a**). Notably, a modest number of CFMs (30) suffices to obtain robust decompositions (**Figs. S2b-d**). The main CFMs for the *Pdx1* locus represent either domains (10/30) or loops (17/30) (**Figs. 1e, S2a**). Domain CFMs exhibit a typical size of 75 ± 25 kb, and represent short-range interactions between neighboring monomers (Fig. S2e). In contrast, loop CFMs depict motifs where genomically distant monomers (200 ± 25 kb for the *Pdx1* locus, **Fig. S2e**) come into close spatial proximity. In addition, single 3D chromatin traces from *Isl1*, a cell-specific transcription factor (TF) (Habener and Stoffers 1998; Bohuslavova et al. 2023), and from *Ins2*, one of the two genes in mice encoding insulin, can be effectively decomposed into loop and domain CFMs (**Figs. 1f and S3-S4**). Finally, 3DTopic also finds loops and domain CFMs as the main motifs representing chromatin traces from human cells (**Figs. 1f, S5**). Thus, individual 3D chromatin structures can be decomposed into orthogonal sets of motifs representing mainly short-range interaction domains and distal loops.

Next, we examined whether CFMs accurately represent the chromatin conformations of single traces. For this, we mapped the CFM decompositions of individual chromatin structures using Uniform Manifold Approximation and Projection (UMAP). We color-coded each single trace according to the CFM with the highest component (hereafter major CFM, **Fig. 1g**), and superimposed them with pure CFMs, which represent decompositions with a single CFM component (stars on the UMAP). This visualization shows that single traces with similar CFM usage cluster together in space. Consistently, the average proximity matrix of traces sharing the same major component closely resembled that of the corresponding pure CFM (**Fig. 1g**, insets). We note, however, that very few cells displayed the structure of a pure CFM, indicating that multiple CFMs are necessary to describe the organization of single traces (**Fig. 1g**). Therefore, while single CFMs often dominate single trace decompositions, they rarely represent single chromatin structures.

Having established that CFMs are essential but not singular representations of chromatin structures, we further explored the frequency of major CFMs among single traces. We quantified this by measuring the frequency at which loop and domain CFMs appear with at least 10% weight in pancreatic cells (**Figs. 1h and S6a-d**). Domain CFMs consistently

showed the highest frequencies (10-25%) in the *Pdx1*, *Isl1*, *Ins2* loci and the locus of Bintu *et al.* (**Figs. 1h, S6a-d**, black lines). In contrast, loop CFMs exhibited significantly lower frequencies (5-15%), comparable to the looping frequencies between distant CTCF sites in live cultured cells (3-30%) (Gabriele *et al.* 2022; Mach *et al.* 2022). As expected, these frequencies vary with changes in the minimum CFMs weight considered in single traces (**Figs. S6a-d**, blue, green and yellow lines). These findings show that short-range interactions leading to TAD sub-domains are more prevalent than long-range loops within TADs.

Evidence from ensemble methods suggested that TADs often comprise multiple overlapping patterns involving CTCF loops and *cis*-regulatory contacts (Phillips-Cremins *et al.* 2013; Rao *et al.* 2014). We tested this hypothesis by quantifying whether specific combinations of CFMs appear together in single traces (**Fig. 1i**). Notably, while single traces are composed of multiple CFMs, specific combinations of CFMs were infrequent. Consistent with these results, color-coding of the UMAP based on the second CFM component revealed a quasi-random spatial distribution (**Fig. S6e**). We note that the most dominant correlations were amongst domain and loop CFMs, indicating the simultaneous contribution from both CFM types to defining single allele chromatin structures (**Fig. 1i**). Overall, these experiments and analyses show that single chromatin traces can be decomposed into a discrete number of 3D structural motifs representing short-range domains and long-range loops.

Experimental CFMs can be reconstructed by loop extrusion and block copolymer interactions

To shed light into the mechanisms involved in CFM formation, we performed molecular dynamics simulations (Frenkel and Smit 2023). First, to simulate domain CFMs, we used a block copolymer model where a subset of neighboring monomers was allowed to interact when found at close spatial proximity (**Figs. 2a, S7a**) (see *Methods*) (Jost *et al.* 2014). Using this model, we generated 7500 independent single polymer structures and decomposed them using 3DTopic. Polymer structures could be decomposed by combinations of domain-like CFMs encompassing the interacting monomers (**Fig. 2a**, right). Second, we performed molecular dynamics simulations of loop extrusion with two inverted CTCF sites (**Figs. 2b, S7b**) (see *Methods*). The main CFM in the decomposition of 7500 single polymer conformations was a loop between CTCF sites (**Fig. 2b**, right). This motif represents conformations where cohesins/loop extruding factors (LEFs) are stalled at the CTCF sites, while the remaining motifs represent incompletely extruded polymers. Thus, domain and loop CFMs can be obtained from block copolymer and loop extrusion simulations, respectively.

Next, we wondered if this simple *in silico* approach would be able to reconstruct the experimental CFMs. For this, we first performed block copolymer simulations where mouse pancreas accessible regions along the *Pdx1* locus (C. Liu *et al.* 2019) were used as proxies for *cis*-regulatory elements, and allowed to interact with each other when in close spatial proximity (**Fig. 2c**). Consistently, 3DTopic decompositions produced only domain CFMs (**Figs. 2c, S7c**). In addition, we performed loop extrusion simulations where CTCF sites were positioned by using the CTCF binding sites from mouse pancreas at the *Pdx1* locus (R. R. Wang *et al.* 2022) (see *Methods*). In this case, most reconstructed motifs represented loops (**Figs. 2d, S7b, S7d**), albeit domain motifs were also occasionally observed when CTCF sites were located in close genomic proximity (**Fig. S7e**). Therefore, block copolymer and loop extrusion simulations of the *Pdx1* locus reproduce multiple domain and loop CFMs reflecting the underlying accessibility and CTCF binding patterns.

To test whether these simulations can restore experimental motifs, we calculated the Pearson correlation between simulated and experimental CFMs. Notably, all the experimental domain CFMs can be matched by at least one CFM simulated using a block copolymer model (**Fig. 2e**, top-left quadrant). We remark that block copolymer simulations are unable to restore experimental loop CFMs with our simulation parameters assuming weak monomer

interactions (**Fig. 2e**, bottom-left quadrant). Instead, these CFMs can be properly restored by loop extrusion simulations (**Figs. 2e-f**, bottom-right quadrant). In addition, we note that loop extrusion can also reproduce a subset of the experimental domain CFMs that arise from loops between genomically-close CTCF sites, however with lower scores than CFMs derived from block copolymer simulations (**Fig. 2e**, top-right quadrant). Thus, experimental CFMs can be fully reproduced by simulations that take into account loop extrusion and interactions between *cis*-regulatory elements and that are simply based on the available CTCF binding and accessibility profiles for mouse pancreas.

To further validate these findings, we conducted additional analyses that quantify the contributions of loop extrusion and *cis*-regulatory interactions to the formation of experimental CFMs. For each experimental CFM, we derived loop extrusion and block copolymer scores by identifying the highest correlation with CFMs generated from loop extrusion and block copolymer models (**Fig. 2g**, see *Methods*). These scores were used to estimate the contributions of both mechanisms to the folding of experimental CFMs. As expected, the majority of domain CFMs (CFM1-4, CFM7-8) display high block copolymer scores (>0.5) and low loop-extrusion scores (<0.5), indicating that interactions between ATAC-seq open regions are involved in their formation. Conversely, all loop CFMs except for 12, 15 and 18 exhibit high loop extrusion scores and low block copolymer scores, consistent with loop extrusion being necessary to reproduce most experimental loop CFMs. In line with previous observations (**Fig. S7e**), we observe that a subset of domain CFMs (CFM5, CFM6) exhibit higher loop extrusion than block copolymer scores, consistently with these domains being formed by either or both *cis*-regulatory interactions and looping between neighboring CTCF sites. Overall, these molecular dynamics simulations strongly support the idea that both loop extrusion and *cis*-regulatory interactions are needed to restore experimentally observed CFMs.

Single-allele TAD structures from multiple cell types are described by a single CFM dictionary

Numerous lines of evidence suggest that 3D genome folding is critical for the regulation of cell-type-specific transcription (Schaeffer and Nollmann 2023). The *Pdx1* locus contains multiple genes displaying tissue-specific expression, for instance *Pdx1* is primarily expressed in the pancreas, *Gsx1* in the brain, and *Urad* in the liver, while other genes are expressed in multiple tissues (e.g. *Polr1d*, *Pan3*, *Gtf3*, or *Mtif3*) (**Figs. S8a-b**). This locus displays multiple chromatin regions accessible in all tissues, and a subset of regions with tissue specificity (notably around *Pdx1*, **Fig. S8a**). Thus, we wondered whether the set of CFMs derived from pancreas could decompose the single allele conformations of the *Pdx1* locus from other mouse tissues despite these differences in gene expression and accessibility. To test this, we first acquired Hi-M data from six additional tissues: brain, kidney, liver, lung, lymph node and thymus. The reconstructed proximity and PWD distance Hi-M maps reveal remarkable structural variations within the *Pdx1* TAD among the different tissues (**Figs. 3a, S8c**). Consistently, other genomic loci containing genes with differential gene expression patterns also displayed 3D changes in chromatin architecture amongst multiple cell types (**Figs. S8d-e**).

As multiple ATAC-seq and CTCF peaks in the *Pdx1* locus are present across all tissues examined (**Fig. S8a**), we wondered whether single chromatin structures from multiple cell types shared a similar set of motifs. To test this, we decomposed these structures using the dictionary of CFMs from the mouse pancreas (**Figs. 1e, S2a**). Remarkably, this dictionary contained enough conformational diversity to accurately represent the 3D chromatin structures across all seven tissues analyzed (**Figs. 3b, S9a-b**). Similar results were observed when deriving a CFM dictionary from a dataset of chromatin traces from all tissues (data not shown). However, the use of a common dictionary of motifs does not necessarily imply that different cell types use CFMs with the same frequency. To explore this, we embedded single

trace CFM decompositions from various tissues into UMAPs. Remarkably, the UMAP density distributions varied dramatically between tissues (**Figs. 3c-d**). Similar observations were made at the *Ins2* locus (**Figs. S9c-d**). These findings indicate that while a single CFM dictionary can effectively decompose chromatin structures from different tissues, the specific usage patterns of CFMs vary considerably across cell types.

Next, we evaluated the statistical differences in CFM utilization by calculating the odds that a CFM would be differentially used in a given tissue compared to the pancreas. An Odds Ratio (OR) of one indicates equal usage of the CFM in both the tissue and the pancreas, an OR greater than one means the CFM is used more in the pancreas, and an OR less than one means it is used less in the pancreas. To ensure that our results were robust and not influenced by the specific CFM decomposition used, we performed 500 independent CFM decompositions (as in Fig. 1d-e) and calculated the distributions of OR for each CFM across different tissues (**Fig. 3e**). Notably, different cell types can utilize the same CFMs at varying frequencies, independent of the specific CFM decomposition. Overall, these results demonstrate that CFM utilization varies significantly across cell types, illustrating how a single dictionary of motifs can produce distinct chromatin organization patterns in different tissues (**Figs. S9a-b**).

The *Pdx1* locus contains multiple genes with tissue-specific expression, likely driven by *cis*-regulatory elements, leading us to anticipate considerable changes in domain CFM utilization across tissues. Indeed, we observed significant variations in domain CFM usage, with only about $9.8 \pm 6.5\%$ of domains remaining unchanged (**Fig. 3f**). CFMs associated with ATAC-seq peaks that varied between tissues (variable, **Fig. 3f**) showed substantial changes, further supporting our modeling results that link domain CFMs to *cis*-regulatory interactions. Interestingly, even CFMs with stable ATAC-seq peaks across tissues exhibited notable, albeit more limited, changes (stable, **Fig. 3f**). This finding suggests that 3D chromatin interactions between open chromatin regions can change despite stable ATAC-seq signatures.

Our modeling work strongly suggests that loop CFMs are primarily formed by loop extrusion. To further test this model, we examined the variation in loop CFM usage across tissues and found that approximately $87.8 \pm 2.8\%$ of loops exhibited either gain or loss compared to the pancreas (**Fig. 3g**). Similar results were observed at other genomic loci (**Figs. S9e-f**). Notably, the *Pdx1* locus contains both stable and variable CTCF sites (**Fig. S8a**). Loop CFMs with variable CTCF anchors showed significant changes (variable, **Fig. 3g**), likely due to the cell-type specific binding of CTCF. This finding further supports a role of loop extrusion in the formation of loop CFMs. Interestingly, loop CFMs with stable anchors also exhibited substantial changes between tissues (stable, **Fig. 3g**). This could arise from multiple processes, including cohesin loading at tissue-specific enhancers (Vos et al. 2021; Karpinska and Oudelaar 2023). This analysis indicates that loop extrusion processes vary significantly between cell types, resulting in distinct 3D structural ensembles that may contribute to cell type-specific transcriptional regulation. Overall, these results demonstrate that while a single CFM dictionary can decompose single trace structures from multiple cell types, the relative frequencies of domain and loop CFM utilization are highly tissue-specific.

Cell-type specific transcription affects the use of domain and loop CFMs in single cells

Given these widespread changes in 3D chromatin structure, we aimed to further characterize differential CFM utilization in specific cell types within a tissue to understand its impact on transcriptional regulation. To achieve this, we performed sequential RNA-FISH imaging in the pancreas followed by Hi-M on the *Pdx1* locus (**Figs. 4a-b**). We classified pancreatic cells into beta cells, which express insulin (*Ins1/Ins2*), and exocrine cells as cells that do not express any of the major endocrine cell markers, i.e. *Ins1/Ins2*, glucagon (*Gcg*), and somatostatin (*Sst*) (**Fig. 4b**, top). Next, we reconstructed ensemble Hi-M maps for beta and exocrine cells (**Fig. 4b**, bottom) and decomposed single allele conformations using the dictionary of CFMs

from pancreas (**Figs. 1e, S2a**), and embedded these decompositions using UMAPs (**Fig. 4c**). The differences between the UMAPs of these well-defined cell types were considerably smaller than those previously observed between cell types from different tissues (**Figs. 3c-d**), which is consistent with both beta and exocrine cells originating from a common multipotent pancreatic progenitor. Nonetheless, inspection of differential UMAPs (**Fig. 4d**) reveals multiple differences, suggesting CFMs may be differentially used between beta and exocrine cells.

Indeed, OR analysis revealed multiple significant differences in the utilization of domains and loops between the two pancreatic cell types (**Figs. 4e-f**). To investigate whether these differences are associated with *Pdx1* transcription, we examined changes in CFMs involving the *Pdx1-Plut* locus between beta cells, where *Pdx1-Plut* are highly transcribed, and exocrine cells, where it is expressed at low levels (**Fig. S10a**). We detected significant alterations affecting both domains and loops (**Fig. 4e-f**). Surprisingly, CFMs bringing *Pdx1-Plut* into proximity of close (domains) or distant (loops) regulatory elements were in most cases more prevalent in exocrine than in beta cells (stars in **Figs. 4e-f**).

To rationalize this unexpected finding, we leveraged the ability of Hi-M to visualize multiple genomic regions concurrently to assess the spatial distances between *Pdx1-Plut* and other genomic regions across cell types. Differential maps and single-cell analysis indicate that *Pdx1-Plut* becomes locally decompacted and spatially isolated from other genomic loci, including close and distant CREs. This occurs specifically in beta cells, where *Pdx1-Plut* are expressed (**Figs. 4g-h**). To further validate this result, we performed Hi-M using a high-resolution oligopaint library that zooms-in on the *Pdx1-Plut* region. Remarkably, these high-resolution chromatin tracing experiments revealed that local decompaction specifically encompasses the *Pdx1-Plut* gene unit (**Figs. 4i-j, S10b-c**). Finally, we observed similar local gene decompaction for *Isl1*, an endocrine cell-specific TF (**Figs. 4k-l, S10d-e**) (Habener and Stoffers 1998; Bohuslavova et al. 2023). In brief, these results indicate that gene transcription leads to local chromatin decompaction, resulting in a reduction of contacts with both proximal and distal regulatory and non-regulatory elements.

To shed light into the mechanisms involved in this loss of proximity upon transcriptional activation, we analyzed the relative changes and variability in CFM utilization between chromatin traces from beta and exocrine cells. For this, we plotted the OR versus the p-value for each CFM (**Fig. 4m, Methods**). The three least used CFMs in beta cells, where *Pdx1* is more highly expressed, encode interactions involving *Pdx1-Plut* with proximal and distal regions (CFM2, CFM9, CFM22, **Fig. 4e-f, 4m**). Remarkably, these CFMs are also significantly more prevalent in tissues where *Pdx1* is not expressed (**Figs. 3f-g**). CFM2 represent contacts between *Pdx1-Plut* and their proximal CREs, CFM22 anchors *Pdx1-Plut* to a cluster of distal CREs, and CFM9 to the distant TAD border (**Fig. 4m**). These changes agree well with increased pairwise distances between *Pdx1-Plut* and their enhancers and TAD border (**Fig. 4h**). Reconstruction of these CFMs requires either block-copolymer interactions between neighboring open chromatin regions or loop extrusion (**Figs. 2e-g**), thus these findings suggest that both mechanisms shape the observed changes in CFM utilization. In contrast, CFMs more frequently occurring in chromatin traces from beta cells did not involve *Pdx1-Plut* (**Fig. 4m**), and may represent compensatory structural changes induced by *Pdx1-Plut* expression which decompacts chromatin at the TAD scale. Similar conclusions can be reached from high-resolution tracing data of the *Pdx1-Plut* locus and from the *Isl1* locus (**Fig. S10f-k**). Thus, these data show that cell-type specific gene expression is accompanied by multiple changes in the use of both domain and loop CFMs, leading to structural TAD rearrangements involving both *cis*-regulatory interactions and long-range CTCF loops.

Cell-type specific transcriptional perturbation affects use of domain and loop CFMs

Our previous analyses showed that loop and domain CFMs arise primarily due to the action of loop extrusion and to cell-type-specific *cis*-regulatory interactions. We directly tested this hypothesis by performing a perturbation that downregulates the expression of beta cell-specific transcription factors, and the levels and binding pattern of CTCF in beta cells. This was achieved by feeding mice with a high-fat diet (R. R. Wang et al. 2022), a well-established preclinical model of obesity and type 2 diabetes (T2D) (C.-Y. Wang and Liao 2012). We performed sequential RNA-FISH followed by Hi-M on pancreas cryosections of mice fed for 14 weeks with either normal diet (ND) or high-fat diet (HFD) (**Fig. 5a**). As expected, islet size and number of beta cells per islet increased considerably in HFD-fed with respect to ND-fed mice (**Fig. 5b**) (Kemkem et al. 2020) as well as body weight and blood glucose levels (**Figs. S11a-d**). While the *Pdx1* locus exhibited broadly similar proximity patterns in HFD and ND conditions (**Figs. 5c, S11e**), significant differences emerged specifically in beta cells between the two conditions (**Fig. 5d**). Specifically, distances between *Pdx1-Plut* and their CREs were higher in beta cells from ND mice (**Figs. 5e**). *Pdx1-Plut* expression is down-regulated during obesity and T2D in humans and mice (Akerman et al. 2017; Kaneto et al. 1999), thus these results are consistent with our previous findings that *Pdx1-Plut* expression leads to increased distances to regulatory elements (**Fig. 4g-h**). We note, however, that multiple additional changes in 3D chromatin structure occur in HFD mice (compare **Figs. 5d** and **4g**), unsurprising given the large deregulation in beta cell-specific TF expression and CTCF levels in obese/T2D mice. In contrast, distances between *Is11* and its CRE in beta cells were lower in ND than in HFD mice (**Fig. S12a-c**). In short, these results indicate that deregulation of cell-type specific TFs and CTCF in beta cells leads to significant changes in the organization of TADs and in the distances between genes and their regulatory elements.

To determine whether these 3D structural changes affected the relative utilization of CFMs in HFD mice, we decomposed beta cell chromatin traces from ND and HFD mice using our standard dictionary of CFMs from pancreas (**Fig. 1g**) and visualized the results on a UMAP (**Figs. 5f-g**). We observed significant changes in the use of both domain and loop CFMs (**Fig. 5h-i**). Notably, loop CFMs that bring *Pdx1-Plut* close to their distant regulatory elements or to the TAD border (CFM22, CFM25 and CFM9) were overrepresented in obese/T2D mice (**Figs. 5i, 5j**). Conversely, the domain CFMs that promote contacts between *Pdx1-Plut* and its close enhancers (CFM2) were underrepresented in these mice (**Fig. 5h**). Additionally, several other CFMs not involving *Pdx1-Plut* (e.g., CFM4, CFM17, CFM12) were also perturbed in HFD mice (**Fig. 5h-j**). Similar results were observed at the *Is11* locus (**Figs. S12d-f**). These findings indicate that metabolic perturbation specifically affecting beta-cell transcription factors and CTCF levels leads to significant changes in both domain and loop CFMs. These data suggest a direct link between transcription factor deregulation and the reshaping of 3D chromatin structure through altered *cis*-regulatory interactions and loop extrusion in single cells.

Discussion

Here, we combined chromatin tracing with machine learning to study how chromatin is organized within TADs in single alleles and across multiple cell types. Our main finding is that chromatin structure within TADs can be described by combinations of unique 3D chromatin folding motifs. This discrete set of motifs represents basic units of chromatin organization within TADs, and comprise short-range interactions between multiple neighboring genomic regions (domain CFMs) and long-range interactions between distal elements (loop CFMs). Notably, a single dictionary of CFMs can be used to represent single allele conformations from very different cell types, suggesting that the same building blocks are used to build the TAD structure of cells from different origins and with different transcriptional states. In addition, the frequency with which each motif is used depends on the cell type, and therefore

may represent a mechanism by which cells fine tune the internal organization of TADs to favor cell type-specific *cis*-regulatory interactions.

Ensemble studies have shown that TADs are internally organized into a nested, partially-overlapping mixture of subdomains (or sub-TADs) (Phillips-Cremins et al. 2013; Rao et al. 2014), which can link CTCF sites, enhancers and promoters, and which change during cell differentiation (Phillips-Cremins et al. 2013; Beagan and Phillips-Cremins 2020). While considerable efforts have been recently devoted to identifying the nature of sub-TADs (Weinreb and Raphael 2016; Haddad, Vaillant, and Jost 2017; Cao et al. 2023; Xu et al. 2024), it is so far unclear whether sub-TADs exist in single cells, at which frequencies, if they are mutually exclusive, or whether they change upon transcriptional activation. Our ability to unbiasedly decode and quantify patterns of chromatin interactions within TADs allowed us to decompose the internal structure of TADs into defined motifs, to measure the absolute frequencies with which these patterns occur in single cells, and how they vary between cell types, critical to understand how chromatin organization within TADs may be linked to transcription at the single allele level.

We find that domain CFMs consistently exhibit the highest frequencies (~15-25%), while loop CFMs are less common (5-15%). These observations align well with recent live-imaging measurements of looping frequencies between single pairs of CTCF sites in cultured cells (3-30%) (Mach et al. 2022; Gabriele et al. 2022), and additionally show that CTCF sites also interact rarely in differentiated mouse tissues even for TADs with multiple CTCF sites. Our analyses showed that co-occurrence of CFMs is rare in single traces, consistent with the transient binding of transcription factors (Lu and Lionnet 2021) and the un-coordinated translocation of LEFs (Ganji et al. 2018; Davidson et al. 2023). The rare co-occurrence between loop CFMs in single traces suggests that typically only one distal CTCF loop is present at any given time. Conversely, dominant domain CFMs frequently correlate with loop CFMs, indicating that short-range domain interactions often co-occur with various distal CTCF loops. Finally, 3Dtopic captures the most prominent structural features, thus other motifs such as intermediate partially-extruded conformations (Fudenberg et al. 2016; Beagan and Phillips-Cremins 2020) are likely present but are far less common due to their structurally heterogeneous nature. Our findings show that despite the sparse and dynamic nature of chromatin, there are common patterns of chromatin organization within TADs linked to LEFs and CRE interactions that, together, shape chromatin organization in single cells. In addition, we provide critical quantitative constraints to guide future physical models combining loop extrusion and *cis*-regulatory interactions to understand how these processes operate *in vivo*.

Enhancers and promoters were proposed to coalesce in 3D to form micro-environments where transcriptional regulation takes place (J. H. Yang and Hansen 2024; Lim and Levine 2021; Wurmser and Basu 2022). This model makes several predictions: the detection of CFMs involving interactions between multiple distant regulatory sequences, the existence of correlations between CFM linking together enhancers and promoters, and the overrepresentation of CFMs bringing together genes and their enhancers in expressing cells. Our data indicates that formation of regulatory micro-environments may be very uncommon in our model system.

Multiple studies established that TADs are structurally highly heterogeneous (Fudenberg et al. 2016; Cattoni et al. 2017; Bintu et al. 2018; Finn et al. 2019; Götz et al. 2022; Boettiger et al. 2016). TADs were proposed to be composed of higher-order chromatin structures termed chromatin nanodomains with a typical size of ~100 kb (Szabo et al. 2020). These sub-TAD structures are disrupted by hyperacetylation, though their molecular origin remains unclear. Our ability to identify and decode the genomic composition of CFMs provides new insights into the mechanisms behind these higher-order structures. Specifically, experimental domain CFMs can be reconstructed through short-range interactions between *cis*-regulatory elements or loop extrusion between proximal CTCF sites, implicating these processes in the

formation of nano-domains. This is further supported by our observation that domain CFM utilization varies significantly between cell types with different gene expression patterns, and is disturbed by perturbations affecting *cis*-acting factors and CTCF levels. These findings suggest that both short-range *cis*-regulatory interactions and loop extrusion shape chromatin architecture at the sub-TAD level, underscoring their importance in cell type-specific transcriptional regulation.

Recent single-cell transcriptomic studies have revealed that many diseases stem from transcriptional dysregulation in single cells or small groups of cells (Rajewsky et al. 2020). However, it remains unclear whether these transcriptional changes are accompanied by alterations in 3D chromatin structure at the single-cell level. Our results demonstrate that obesity and T2D in mice significantly disrupts the organization of TADs associated with genes involved in disease pathogenesis. Notably, both architectural loops and *cis*-regulatory domains are affected, leading to a widespread perturbation of interactions between promoters and putative enhancers. This disruption is consistent with a mechanism that could deregulate the expression of key beta cell markers, such as *Pdx1* and *Isl1* (Guo et al. 2013; Sachdeva et al. 2009; Gao et al. 2014). These findings underscore the critical role of 3D chromatin structure in maintaining proper gene regulation and suggest that changes in CFMs could be early indicators of transcriptional dysregulation during disease progression. By identifying specific CFM alterations, we may predict transcriptional changes that arise at later disease stages, providing a molecular basis for understanding disease mechanisms. Furthermore, CFM changes could serve as proxies for identifying diseased states, offering potential biomarkers for early diagnosis and targeted intervention.

In summary, we provide a blueprint for using imaging-based spatial genomics to chart 3D chromosome organization in single cells across multiple cell types, and apply machine learning to unveil the building blocks of chromatin organization in single cells in different healthy tissues and during disease. These technologies will be critical to further dissect the role of 3D chromatin structure in the regulation of cell type specific transcriptional programs, and the molecular mechanisms involved.

Methods

Mice

Animal studies were conducted according to the European Animal Welfare Guidelines (2010/63/EU). Protocols were approved by the Institutional Animal Care and Use Committee (CEEA-LR-1434) and the French Ministry of Agriculture (APAFIS#13044). Mice were housed in a conventional facility on a 12 hour light/12 hour dark cycle and were given chow and water *ad libitum*. C57BL/6J mice were purchased from Janvier-SAS (RRID: IMSR_JAX:000664). Mice were fed with normal diet (ND) until 6 weeks of age and then fed with either ND or high-fat diet (HFD) (63% energy from fat) (Safe Diets, France) for 14 weeks. No data were excluded unless animals died during experimentation. The intraperitoneal glucose tolerance test (IPGTT) was performed as described (Michau et al. 2016). We chose a glucose dose of 3 g/kg body weight to ensure full beta cell challenge and generation of an insulin peak (J.-K. Yang et al. 2018).

Sample preparation

Mice were anesthetized with ketamine/xylazine and perfused with cold PBS through the intracardiac route. The collected tissues were then fixed in 4% paraformaldehyde in PBS (v/v) (Sigma) at room temperature for 4 hours under agitation, followed by overnight incubation at 4°C. Subsequently, the tissues were transferred to a PBS solution containing 30% sucrose until they sank to the bottom of the tube. Tissues were then embedded in OCT (Sigma) and stored at -80°C until cryosectioning. Prior to cryosectioning, 40 mm round coverslips (Bioprotech) were washed with 70% ethanol in water (v/v) and activated with air plasma for 30 seconds. Slides were then covered with 100 µL of pure 3-aminopropyltrimethoxysilane (Sigma) for 5 minutes at room temperature. Slides were left in water for 5-10 minutes and rinsed 2x10 minutes in a water bath with agitation. A solution containing 0.5% glutaraldehyde (Sigma) in PBS was added for 30 minutes and rinsed with water. The slides were then coated with 0.1 mg/mL poly-D-lysine (Sigma) in water for 1 hour and incubated O/N in water. Ten µm tissue sections were cut with a cryostat and added to the coated slides, dried at room temperature for 1-2 hours for immediate use, or frozen at -20°C for later use.

RNA-FISH libraries

Sequential RNA-FISH Libraries were constructed following the procedure described in (K. H. Chen et al. 2015) with the library design script available at https://github.com/ZhuangLab/MERFISH_analysis. Briefly, a maximum of 90 probes were designed to hybridize to cDNA with a target sequence region of 30 bp allowing 20 bp overlap between adjacent probes to maximize the number of probes for a gene. Each probe was customized to add a tail composed of gene-specific readout sequence and a forward and reverse primer sequence positioned at the 5' and 3' ends for PCR amplification of the library following the method described in Cardozo Gizzi et al. (2020). The full list of barcode and probe sequences is available in Supplementary Data 2 and 3. RNA-FISH labeling followed standard protocols (Cardozo Gizzi et al. 2020).

Hi-M libraries

Oligopaint libraries were obtained from a public database (<http://genetics.med.harvard.edu/oligopaints>) containing unique 35/45mer sequences homologous to the *Mouse musculus* genome (mm10). Hi-M probe sets were designed for each locus and amplified according to the method described by Cardozo Gizzi et al. (2020). Four Hi-M libraries were designed. The first two libraries encompassed the *Pdx1* gene: one containing 25 barcodes of approximately 25 kb each, named *Pdx1* low-resolution library, and

one containing 18 barcodes of approximately 8 kb each, named *Pdx1* high-resolution library. The third library was designed along the *Ins2* gene and contained 25 barcodes, each targeting regions of approximately 6 kb. The final library was designed along the *Isl1* gene and consisted of 24 barcodes, each targeting regions of approximately 26 kb. The full list of barcode and probes sequence is available in Supplementary Data 1 and 2. DNA-FISH labeling followed previously published protocols (Cardozo Gizzi et al. 2020).

Imaging system

Experiments were performed on two similar acquisition setups. A home-made imaging setup built on a RAMM modular microscope system (Applied Scientific Instrumentation), as described previously (Cardozo Gizzi et al. 2020; Barho et al. 2022) and a modified Zeiss microscope (Axio Observer) with a similar design. Both setups were equipped with a water immersion objective (Nikon Plan-Apo 60x NA=1.2 and Zeiss W DICII 63x NA=1.2 respectively), a sCMOS camera (Flash4-v3, Hamamatsu) and a home-made fluidic system.

Combined sequential RNA-FISH and Hi-M

After seq RNA-FISH experiment, the slide was removed from the FCS2® flow chamber and DNA-FISH labeling protocol was performed. The slide was then mounted again in the FCS2® flow chamber and realigned with micron precision using triangulation marks on both the coverslip and the flow chamber glass. Pixel accuracy was achieved by correcting the TIFF images by cross-correlating the DAPI staining acquired during RNA and Hi-M acquisitions using a custom Python script.

Image analysis

TIFF images were then deconvolved using Huygens Professional 21.04 (Scientific Volume Imaging, <https://svi.nl>). The analysis was performed using the pyHiM analysis pipeline we developed (<https://pyhim.readthedocs.io/en/latest/>) with release 0.7 (Devos et al. 2024). To obtain proximity frequency maps from pairwise distance matrices, we used a distance threshold of 100 nm, a value that optimizes the correlation between Hi-M and Hi-C datasets (Figs. S1h-i).

Cell type assignment to Hi-M traces

Cell type assignment was accomplished using RNA-FISH signals. For the segmentation of alpha, beta, and delta cells, specific RNA-FISH signals for each cell marker, *Gcg*, *Ins2* and *Sst*, were used. These images were used to train random forest models in Ilastik. The trained models were then applied, in conjunction with a custom Python script for image registration, to assign traces to specific cell types within the endocrine tissues. Exocrine cells were identified as cells lacking these markers.

Trace filtering and normalization

Variability in 3D distance distributions can sometimes be observed between experiments and is likely to be due to small variations in Hi-M labeling and acquisition conditions. To minimize these variations, the following normalization procedure was used: First, the physical vs. genomic distance curves are calculated for each experiment. By pooling all data, an average curve is also calculated. A correction factor is then derived for each experiment by minimizing the deviation from the average curve.

Missing barcodes in Hi-M traces were imputed as follows: If a barcode was missing in a Hi-M trace, but the two adjacent barcodes were present, the position of the missing barcode was imputed by averaging the positions of the two adjacent barcodes.

Domainogram

Insulation Scores (IS) derived from the Hi-M dataset were calculated as previously described (Messina et al. 2023). Briefly, an n-by-n square window was moved along the diagonal of the median PWD matrix and the distances within the square were summed. Next, a domainogram was generated by smoothing a matrix obtained by calculating the IS with increasing window size (ranging from 1 x 1 to 6 x 6) over the median Hi-M PWD matrix.

Volcano plots

Volcano plots from Hi-M matrices were computed for two conditions. First, the differential median PWD matrix was calculated by determining the ratio of each pair of bins between the two conditions. Then, p-values were derived using the Wilcoxon rank-sum test, comparing the single-cell distance distributions for each pair of bins. Finally, the log2 values of the differential matrix were plotted against the -log10(p-values) from the Wilcoxon test.

CTCF sites prediction

Genomic sequences corresponding to specific loci were retrieved from Ensembl (<https://www.ensembl.org>) and subsequently analyzed to predict CTCF binding using the CTCFBSDB 2.0 database (<http://insulatordb.uthsc.edu/>) (Ziebarth, Bhattacharya, and Cui 2013).

Hi-C, ChIP-seq, ATAC-seq, CUT&Tag, RNA-seq and scRNA-seq data processing

ATAC-seq data were reanalyzed from raw data (C. Liu et al. 2019; Shen et al. 2012). Sequencing reads were aligned to the reference mus musculus genome assembly (mm10) using Burrows-Wheeler Aligner (0.7.17-r1188) with default parameters. Finally, peak calling was performed using MACS2 (2.2.7.1) with default parameters. CTCF ChIP-seq were reanalyzed from raw data (Shen et al. 2012) with default parameters. CTCF CUT&Tag data were downloaded from <https://data.mendeley.com/datasets/mwgxv7m927/2> (R. R. Wang et al. 2022) in .bw and converted to .bedgraph using bigWigToBedGraphNext. Peaks were then called using *macs2 bdgpeakcall* with -c option set to 3. Hi-C data acquired in different mouse tissues were downloaded in .mcool format from accession number GSE150704 (Leidescher et al. 2022).

Annotated single-cell RNA sequencing (scRNA-seq) data were downloaded from the Tabula Muris GitHub repository (<https://github.com/czbiohub-sf/tabula-muris>) (Tabula Muris Consortium et al. 2018). These data were re-plotted using a custom Python script. Quickly, the annotated data were filtered to select only the tissues used in this study, specifically Brain, Kidney, Lung, Liver, Pancreas, and Thymus. Within the brain tissue, both Brain_Non-Myeloid and Brain_Myeloid categories were combined and referred to as Brain. Next, tSNE1 and tSNE2 coordinates for cells from these tissues were used to represent the data in a 2D space. Expression count data were downloaded from Figshare (<https://ndownloader.figshare.com/articles/5829687/versions/4>) and used to represent gene expression levels in single cells within the tSNE space.

The relative expression of genes in exocrine and beta cells was determined by calculating the ratio between the count of the target gene, and the count of a housekeeping gene : Actb, in each single cell. Cells were classified as exocrine or beta cells based on the “cell_ontology_class” in the annotated scRNA-seq data. Type B pancreatic cells are referred to as beta cells, while pancreatic acinar cells are referred to as exocrine cells.

3DTopic modeling

3DTopic uses LDA to decompose single-cell pairwise contact maps into discrete topics. Each CFM represents contacts that are likely to occur together in the data. Only traces with at least

65% of the barcodes were selected. Pairwise distance maps from single alleles were pre-processed and used for training. The number of topics was chosen based on the maximum coherence (Röder et al. 2015) calculated with different numbers of topics, ranging from 5 to 50. Multiple LDA runs were performed to assess reproducibility, and the final decomposition was selected based on consensus and uniqueness of the topics.

3Dtopic reconstructions

Each single-cell decomposed matrix was reconstructed by multiplying the weight of each CFM by its corresponding CFM matrix. The reconstructed single-cell matrix was then generated by summing all weighted CFM matrices for that cell. Finally, the mean 3DTopic reconstructed map was obtained by averaging the reconstructed single-cell matrices.

Odds ratio

The odd ratio calculation is given by the following formula :

$$OR = \frac{DN}{HN} / \frac{DE}{HE}$$

where DN and DE are the number of cells for which a given CFMs weight exceeds a minimum threshold (0.2) in test and reference tissue respectively. HN and HE are the number of cells for which a given CFMs weight does not exceed the minimum threshold (0.2) in test and reference tissue respectively.

To generate the volcano plot of odds ratios (OR), we performed over 100 distinct 3D Topic decompositions. We then clustered similar CFMs from these decompositions that had a Pearson correlation greater than 0.8 with the reference decomposition. For each CFM, we calculated the $-\log_{10}$ of the p-values from the Wilcoxon test applied to the OR distribution (y-axis) and the $\log_{10}$ of the median OR value (x-axis).

Consensual CFMs

To test the robustness and reproducibility of the 3D topic decomposition, consensus CFMs were computed. Multiple independent decompositions ($n > 10$) are pooled and similar CFMs (with a Pearson correlation > 0.62) are clustered together. If a cluster contains at least $n/2$ topics, i.e. these topics are found in at least half of the decompositions, a consensus CFMs is derived by calculating the mean of the individual CFMs grouped within the selected cluster.

Domain and loop CFM classification

CFMs were classified into two categories: domains and loops, using a custom Python script. Briefly, a binary mask is created based on thresholding. Connected components within the mask are analyzed for their shape and size. CFMs are categorized based on the presence and characteristics of these components: domains are identified by diagonal components, and loops by compact non-diagonal components. When necessary, visual examinations are applied to correct for misclassifications.

Sankey diagrams

Sankey diagrams were generated to evaluate the percentages of domains and loops that are gained, lost, or unaffected across different tissues as compared to a reference tissue. The

odds ratio for loops and domains in CFMs from a single decomposition was calculated for each tissue relative to a reference tissue. CFMs were then classified into three categories (gained, lost, unaffected) based on an arbitrary log10 threshold of -0.1 and 0.1. This process was repeated over 100 different decompositions.

Polymer simulations

All polymer simulations were performed using the OpenMM-based polychrom package (Imakaev, Goloborodko, and hbbrandao 2019; Goloborodko et al. 2016; Fudenberg et al. 2016) developed by the Open Chromosome Collective (open2c, <https://doi.org/10.5281/zenodo.3579473>). As in Fudenberg et al. (Fudenberg et al. 2016), the chromatin segment corresponding to the locus was modeled as a bead-spring chain with some stiffness and excluded volume, where each bead (or monomer) was considered to be 1 kb.

Chromatin accessibility-based heteropolymers consisted of two different types of monomers; those corresponding to accessible sites from ATAC-seq data were assigned *type 1*, and all the others, *type 0*. We then used a smoothed square well potential (equations further below, parameters in **Table 1**) to introduce a distance-dependent attraction between all non-bonded *type 1* monomers (i.e., attraction between all accessible sites).

For the loop extrusion simulations, the dynamics of cohesins/loop extruding factors (LEFs) were first simulated on a one-dimensional lattice with CTCF sites and their LEF capture rates (**Table 2, Fig. S7b**) assigned based on the peak sizes of CTCF from CUT&Tag data (R. Wang et al. 2022). LEF positions were recorded over time and were used to rapidly construct simple contact frequency maps based on loops created during the trajectory, which then facilitated manual optimization of loop extrusion parameters (**Table 3**). Next, the recorded LEF positions were used to mimic loop extrusion on the homopolymer bead-spring chain described above under Langevin dynamics (Zwanzig 2007) using molecular dynamics simulations (parameters in **Table 4**).

Equilibrated polymer conformations from these simulations were saved, coarse-grained to match the experimental resolution (**Table 5**), and used to construct pairwise distance matrices.

Potentials

All the potentials used in our simulations describe the attractive or repulsive energy between non-bonded monomers based on the distance from the center of the monomer, denoted by r and given in monomer-monomer bond length units.

Homopolymer polynomial repulsive potential (Fig. S7a) (Imakaev, Goloborodko, and hbbrandao 2019; Goloborodko et al. 2016; Fudenberg et al. 2016),

The repulsive potential used for excluded volume is given by

$$E_{rep}(r) = E_0 \left(1 + \left(\sqrt{\frac{6}{7}} \frac{r}{r_{max}} \right)^{12} \left(\left(\sqrt{\frac{6}{7}} \frac{r}{r_{max}} \right)^2 - 1 \right) \left(\frac{823543}{46656} \right) \right)$$

where $r_{max} = 1.05$ and $E_0 = 3 kT$ are the width and the height of the potential respectively

Heteropolymer smoothed-square well potential (Fig. S7a) (Imakaev, Goloborodko, and hbbrandao 2019; Goloborodko et al. 2016; Fudenberg et al. 2016),

The repulsive part of the heteropolymer potential is given by

$$E_{rep}(r) = E_{rep}^{tot} \left(1 + \left(\sqrt{\frac{6}{7}} \frac{r}{r_{rep}} \right)^{12} \left(\left(\sqrt{\frac{6}{7}} \frac{r}{r_{rep}} \right)^2 - 1 \right) \left(\frac{823543}{46656} \right) \right)$$

where the width of the potential $r_{rep} = 1.05$, and the total repulsive energy is

$$E_{rep}^{tot} = E_{rep}^{common} + E_{rep}^{select},$$

Here, $E_{rep}^{common} = 3 kT$ and $E_{rep}^{select} = 0 kT$ are the repulsive energies between all and selective non-bonded monomers respectively. The former accounts for excluded volume.

The attractive part of the potential is described by

$$E_{att}(r) = E_{att}^{tot} \left(\left(\sqrt{\frac{6}{7}} \frac{(r - r_{rep} - \delta)}{\delta} \right)^{12} \left(\left(\sqrt{\frac{6}{7}} \frac{(r - r_{rep} - \delta)}{\delta} \right)^2 - 1 \right) \left(\frac{823543}{46656} \right) - 1 \right)$$

where $\delta = \frac{1}{2} \cdot (r_{att} - r_{rep})$, the width of the potential measured from $r = 0$ is $r_{att} = 1.8$, and the total attractive energy is

$$E_{att}^{tot} = E_{att}^{common} + E_{att}^{select},$$

with $E_{att}^{common} = 0 kT$ and selective attraction between *type 1* monomers $E_{att}^{select} = 0.2 kT$

These two parts are then stitched together using a Heaviside step function

$$E_{tot}(r) = H(r_{rep} - r) \times E_{rep}(r) + H(r - r_{rep}) \times E_{att}(r)$$

where $H(x) = 0$ if $x < 0$ and $H(x) = 1$ if $x \geq 0$

Code availability

Hi-M data were acquired using qudi-HiM (Barho et al. 2022). The current version of qudi-HiM is found at <https://github.com/NollmannLab/qudi-HiM>. Hi-M data were analyzed using pyHiM release 0.7, available at <https://github.com/marcnol/pyHiM> (Devos et al. 2024).

References

- Ahlgren, U., J. Jonsson, L. Jonsson, K. Simu, and H. Edlund. 1998. "Beta-Cell-Specific Inactivation of the Mouse *Ip1/Pdx1* Gene Results in Loss of the Beta-Cell Phenotype and Maturity Onset Diabetes." *Genes & Development* 12 (12): 1763–68.
- Akerman, Ildem, Zhidong Tu, Anthony Beucher, Delphine M. Y. Rolando, Claire Sauty-Colace, Marion Benazra, Nikolina Nakic, et al. 2017. "Human Pancreatic β Cell lncRNAs Control Cell-Specific Regulatory Networks." *Cell Metabolism* 25 (2): 400–411.
- Alexander, Jeffrey M., Juan Guan, Bingkun Li, Lenka Maliskova, Michael Song, Yin Shen, Bo Huang, Stavros Lomvardas, and Orion D. Weiner. 2019. "Live-Cell Imaging Reveals Enhancer-Dependent Sox2 Transcription in the Absence of Enhancer Proximity." *eLife* 8

- (May). <https://doi.org/10.7554/eLife.41769>.
- Aljahani, Abrar, Peng Hua, Magdalena A. Karpinska, Kimberly Quililan, James O. J. Davies, and A. Marieke Oudelaar. 2022. "Analysis of Sub-Kilobase Chromatin Topology Reveals Nano-Scale Regulatory Interactions with Variable Dependence on Cohesin and CTCF." *Nature Communications* 13 (1): 2139.
- Barho, Franziska, Jean-Bernard Fiche, Marion Bardou, Olivier Messina, Alexandre Martiniere, Christophe Houbon, and Marcelo Nollmann. 2022. "Qudi-HiM: An Open-Source Acquisition Software Package for Highly Multiplexed Sequential and Combinatorial Optical Imaging." *Open Research Europe*. <https://doi.org/10.12688/openreseurope.14641.2>.
- Beagan, Jonathan A., and Jennifer E. Phillips-Cremins. 2020. "On the Existence and Functionality of Topologically Associating Domains." *Nature Genetics* 52 (1): 8–16.
- Bintu, Bogdan, Leslie J. Mateo, Jun-Han Su, Nicholas A. Sinnott-Armstrong, Mirae Parker, Seon Kinrot, Kei Yamaya, Alistair N. Boettiger, and Xiaowei Zhuang. 2018. "Super-Resolution Chromatin Tracing Reveals Domains and Cooperative Interactions in Single Cells." *Science* 362 (6413). <https://doi.org/10.1126/science.aau1783>.
- Boettiger, Alistair N., Bogdan Bintu, Jeffrey R. Moffitt, Siyuan Wang, Brian J. Beliveau, Geoffrey Fudenberg, Maxim Imakaev, Leonid A. Mirny, Chao-Ting Wu, and Xiaowei Zhuang. 2016. "Super-Resolution Imaging Reveals Distinct Chromatin Folding for Different Epigenetic States." *Nature*, January. <https://doi.org/10.1038/nature16496>.
- Bohuslavova, Romana, Valeria Fabriciova, Laura Lebrón-Mora, Jessica Malfatti, Ondrej Smolik, Lukas Valihrach, Sarka Benesova, et al. 2023. "ISL1 Controls Pancreatic Alpha Cell Fate and Beta Cell Maturation." *Cell & Bioscience* 13 (1): 53.
- Bravo González-Blas, Carmen, Liesbeth Minnoye, Dafni Papasokrati, Sara Aibar, Gert Hulselmans, Valerie Christiaens, Kristofer Davie, Jasper Wouters, and Stein Aerts. 2019. "cisTopic: Cis-Regulatory Topic Modeling on Single-Cell ATAC-Seq Data." *Nature Methods* 16 (5): 397–400.
- Cao, Yaqiang, Shuai Liu, Kairong Cui, Qingsong Tang, and Keji Zhao. 2023. "Hi-TrAC Detects Active Sub-TADs and Reveals Internal Organizations of Super-Enhancers." *Nucleic Acids Research* 51 (12): 6172–89.
- Cardozo Gizzi, Andrés M., Diego I. Cattoni, Jean-Bernard Fiche, Sergio M. Espinola, Julian Gurgo, Olivier Messina, Christophe Houbon, et al. 2019. "Microscopy-Based Chromosome Conformation Capture Enables Simultaneous Visualization of Genome Organization and Transcription in Intact Organisms." *Molecular Cell*, February. <https://doi.org/10.1016/j.molcel.2019.01.011>.
- Cardozo Gizzi, Andrés M., Sergio M. Espinola, Julian Gurgo, Christophe Houbon, Jean-Bernard Fiche, Diego I. Cattoni, and Marcelo Nollmann. 2020. "Direct and Simultaneous Observation of Transcription and Chromosome Architecture in Single Cells with Hi-M." *Nature Protocols* 15 (3): 840–76.
- Cattoni, Diego I., Andrés M. Cardozo Gizzi, Mariya Georgieva, Marco Di Stefano, Alessandro Valeri, Delphine Chamoussset, Christophe Houbon, et al. 2017. "Single-Cell Absolute Contact Probability Detection Reveals Chromosomes Are Organized by Multiple Low-Frequency yet Specific Interactions." *Nature Communications* 8 (1): 1753.
- Chen, Hongtao, Michal Levo, Lev Barinov, Miki Fujioka, James B. Jaynes, and Thomas Gregor. 2018. "Dynamic Interplay between Enhancer-Promoter Topology and Gene Activity." *Nature Genetics* 50 (9): 1296–1303.
- Chen, Kok Hao, Alistair N. Boettiger, Jeffrey R. Moffitt, Siyuan Wang, and Xiaowei Zhuang. 2015. "RNA Imaging. Spatially Resolved, Highly Multiplexed RNA Profiling in Single Cells." *Science* 348 (6233): aaa6090.
- Conte, Mattia, Ehsan Irani, Andrea M. Chiariello, Alex Abraham, Simona Bianco, Andrea Esposito, and Mario Nicodemi. 2022. "Loop-Extrusion and Polymer Phase-Separation Can Co-Exist at the Single-Molecule Level to Shape Chromatin Folding." *Nature Communications* 13 (1): 4070.
- Davidson, Iain F., Roman Barth, Maciej Zaczek, Jaco van der Torre, Wen Tang, Kota Nagasaka, Richard Janissen, et al. 2023. "CTCF Is a DNA-Tension-Dependent Barrier to

- Cohesin-Mediated Loop Extrusion." *Nature* 616 (7958): 822–27.
- Davidson, Iain F., and Jan-Michael Peters. 2021. "Genome Folding through Loop Extrusion by SMC Complexes." *Nature Reviews. Molecular Cell Biology* 22 (7): 445–64.
- Devos, Xavier, Jean-Bernard Fiche, Marion Bardou, Olivier Messina, Christophe Houbbron, Julian Gurgo, Marie Schaeffer, et al. 2024. "pyHiM: A New Open-Source, Multi-Platform Software Package for Spatial Genomics Based on Multiplexed DNA-FISH Imaging." *Genome Biology* 25 (1): 47.
- Dixon, Jesse R., Inkyung Jung, Siddarth Selvaraj, Yin Shen, Jessica E. Antosiewicz-Bourget, Ah Young Lee, Zhen Ye, et al. 2015. "Chromatin Architecture Reorganization during Stem Cell Differentiation." *Nature* 518 (7539): 331–36.
- Dixon, Jesse R., Siddarth Selvaraj, Feng Yue, Audrey Kim, Yan Li, Yin Shen, Ming Hu, Jun S. Liu, and Bing Ren. 2012. "Topological domains in mammalian genomes identified by analysis of chromatin interactions." *Nature* 485 (7398): 376–80.
- Espinola, Sergio Martin, Markus Götz, Maelle Bellec, Olivier Messina, Jean-Bernard Fiche, Christophe Houbbron, Matthieu Dejean, et al. 2021. "Cis-Regulatory Chromatin Loops Arise before TADs and Gene Activation, and Are Independent of Cell Fate during Early Drosophila Development." *Nature Genetics* 53 (4): 477–86.
- Finn, Elizabeth H., Gianluca Pegoraro, Hugo B. Brandão, Anne-Laure Valton, Marlies E. Oomen, Job Dekker, Leonid Mirny, and Tom Misteli. 2019. "Extensive Heterogeneity and Intrinsic Variation in Spatial Genome Organization." *Cell* 176 (6): 1502–15.e10.
- Flyamer, Ilya M., Johanna Gassler, Maxim Imakaev, Hugo B. Brandão, Sergey V. Ulianov, Nezar Abdennur, Sergey V. Razin, Leonid A. Mirny, and Kikuë Tachibana-Konwalski. 2017. "Single-Nucleus Hi-C Reveals Unique Chromatin Reorganization at Oocyte-to-Zygote Transition." *Nature Publishing Group*, March. <https://doi.org/10.1038/nature21711>.
- Frenkel, Daan, and Berend Smit. 2023. *Understanding Molecular Simulation: From Algorithms to Applications*. Elsevier.
- Fudenberg, Geoffrey, Maxim Imakaev, Carolyn Lu, Anton Goloborodko, Nezar Abdennur, and Leonid A. Mirny. 2016. "Formation of Chromosomal Domains by Loop Extrusion." *Cell Reports* 15 (9): 2038–49.
- Furlong, Eileen E. M., and Michael Levine. 2018. "Developmental Enhancers and Chromosome Topology." *Science* 361 (6409): 1341–45.
- Gabriele, Michele, Hugo B. Brandão, Simon Grosse-Holz, Asmita Jha, Gina M. Dailey, Claudia Cattoglio, Tsung-Han S. Hsieh, Leonid Mirny, Christoph Zechner, and Anders S. Hansen. 2022. "Dynamics of CTCF- and Cohesin-Mediated Chromatin Looping Revealed by Live-Cell Imaging." *Science* 376 (6592): 496–501.
- Ganji, Mahipal, Indra A. Shaltiel, Shveta Bisht, Eugene Kim, Ana Kalichava, Christian H. Haering, and Cees Dekker. 2018. "Real-Time Imaging of DNA Loop Extrusion by Condensin." *Science* 360 (6384): 102–5.
- Gao, Tao, Brian McKenna, Changhong Li, Maximilian Reichert, James Nguyen, Tarjinder Singh, Chenghua Yang, et al. 2014. "Pdx1 Maintains β Cell Identity and Function by Repressing an α Cell Program." *Cell Metabolism* 19 (2): 259–71.
- Giorgetti, Luca, Rafael Galupa, Elphège P. Nora, Tristan Piolot, France Lam, Job Dekker, Guido Tiana, and Edith Heard. 2014. "Predictive polymer modeling reveals coupled fluctuations in chromosome conformation and transcription." *Cell* 157 (4): 950–63.
- Goloborodko, Anton, Maxim V. Imakaev, John F. Marko, and Leonid Mirny. 2016. "Compaction and Segregation of Sister Chromatids via Active Loop Extrusion." *eLife* 5 (May): e14864.
- Götz, Markus, Olivier Messina, Sergio Espinola, Jean-Bernard Fiche, and Marcelo Nollmann. 2022. "Multiple Parameters Shape the 3D Chromatin Structure of Single Nuclei at the Doc Locus in Drosophila." *Nature Communications* 13 (1): 5375.
- Guo, Shuangli, Chunhua Dai, Min Guo, Brandon Taylor, Jamie S. Harmon, Maike Sander, R. Paul Robertson, Alvin C. Powers, and Roland Stein. 2013. "Inactivation of Specific β Cell Transcription Factors in Type 2 Diabetes." *The Journal of Clinical Investigation* 123 (8): 3305–16.

- Habener, J. F., and D. A. Stoffers. 1998. "A Newly Discovered Role of Transcription Factors Involved in Pancreas Development and the Pathogenesis of Diabetes Mellitus." *Proceedings of the Association of American Physicians* 110 (1): 12–21.
- Haddad, Noelle, Cédric Vaillant, and Daniel Jost. 2017. "IC-Finder: Inferring Robustly the Hierarchical Organization of Chromatin Folding." *Nucleic Acids Research* 45 (10): e81.
- Hafner, Antonina, and Alistair Boettiger. 2023. "The Spatial Organization of Transcriptional Control." *Nature Reviews. Genetics* 24 (1): 53–68.
- Imakaev, Maksim, Anton Goloborodko, and hbrandao. 2019. *Mirnylab/polychrom: v0.1.0*. Zenodo. <https://doi.org/10.5281/ZENODO.3579473>.
- Jost, Daniel, Pascal Carrivain, Giacomo Cavalli, and Cédric Vaillant. 2014. "Modeling Epigenome Folding: Formation and Dynamics of Topologically Associated Chromatin Domains." *Nucleic Acids Research* 42 (15): 9553–61.
- Kaneto, H., Y. Kajimoto, J. Miyagawa, T. Matsuoka, Y. Fujitani, Y. Umayahara, T. Hanafusa, Y. Matsuzawa, Y. Yamasaki, and M. Hori. 1999. "Beneficial Effects of Antioxidants in Diabetes: Possible Protection of Pancreatic Beta-Cells against Glucose Toxicity." *Diabetes* 48 (12): 2398–2406.
- Karpinska, Magdalena A., and Aukje Marieke Oudelaar. 2023. "The Role of Loop Extrusion in Enhancer-Mediated Gene Activation." *Current Opinion in Genetics & Development* 79 (April): 102022.
- Kemkem, Yasmine, Daniela Nasteska, Anne de Bray, Paula Bargi-Souza, Rodrigo A. Peliciari-Garcia, Anne Guillou, Patrice Mollard, David J. Hodson, and Marie Schaeffer. 2020. "Maternal Hypothyroidism in Mice Influences Glucose Metabolism in Adult Offspring." *Diabetologia* 63 (9): 1822–35.
- Laat, Wouter de, and Denis Duboule. 2013. "Topology of mammalian developmental enhancers and their regulatory landscapes." *Nature* 502 (7472): 499–506.
- Leidescher, Susanne, Johannes Ribisel, Simon Ullrich, Yana Feodorova, Erica Hildebrand, Alexandra Galitsyna, Sebastian Bultmann, et al. 2022. "Spatial Organization of Transcribed Eukaryotic Genes." *Nature Cell Biology* 24 (3): 327–39.
- Li, Jieru, Angela Hsu, Yujing Hua, Guanshi Wang, Lingling Cheng, Hiroshi Ochiai, Takashi Yamamoto, and Alexandros Pertsinidis. 2020. "Single-Genome Imaging Links Genome Topology, Promoter-Enhancer Communication and Transcription Control." *Nature Structural & Molecular Biology* 27 (11): 1032–40.
- Lim, Bomyi, and Michael S. Levine. 2021. "Enhancer-Promoter Communication: Hubs or Loops?" *Current Opinion in Genetics & Development* 67 (April): 5–9.
- Liu, Chuanyu, Mingyue Wang, Xiaoyu Wei, Liang Wu, Jiangshan Xu, Xi Dai, Jun Xia, et al. 2019. "An ATAC-Seq Atlas of Chromatin Accessibility in Mouse Tissues." *Scientific Data* 6 (1): 65.
- Liu, Lin, Lin Tang, Wen Dong, Shaowen Yao, and Wei Zhou. 2016. "An Overview of Topic Modeling and Its Current Applications in Bioinformatics." *SpringerPlus* 5 (1): 1608.
- Lu, Feiyue, and Timothée Lionnet. 2021. "Transcription Factor Dynamics." *Cold Spring Harbor Perspectives in Biology* 13 (11). <https://doi.org/10.1101/cshperspect.a040949>.
- Mach, Pia, Pavel I. Kos, Yinxiu Zhan, Julie Cramard, Simon Gaudin, Jana Tünnermann, Edoardo Marchi, et al. 2022. "Cohesin and CTCF Control the Dynamics of Chromosome Folding." *Nature Genetics* 54 (12): 1907–18.
- Mateo, Leslie J., Sedona E. Murphy, Antonina Hafner, Isaac S. Cinquini, Carly A. Walker, and Alistair N. Boettiger. 2019. "Visualizing DNA Folding and RNA in Embryos at Single-Cell Resolution." *Nature* 568 (7750): 49–54.
- Messina, Olivier, Flavien Raynal, Julian Gurgo, Jean-Bernard Fiche, Vera Pancaldi, and Marcelo Nollmann. 2023. "3D Chromatin Interactions Involving Drosophila Insulators Are Infrequent but Preferential and Arise before TADs and Transcription." *Nature Communications* 14 (1): 6678.
- Michau, Aurélien, David J. Hodson, Pierre Fontanaud, Anne Guillou, Gabriel Espinosa-Carrasco, François Molino, Catherine J. Peters, et al. 2016. "Metabolism Regulates Exposure of Pancreatic Islets to Circulating Molecules In Vivo." *Diabetes* 65 (2): 463–75.

- Nasmyth, Kim. 2001. "Disseminating the Genome: Joining, Resolving, and Separating Sister Chromatids During Mitosis and Meiosis." *Annual Review of Genetics* 35 (Volume 35, 2001): 673–745.
- Nora, Elphège P., Bryan R. Lajoie, Edda G. Schulz, Luca Giorgetti, Ikuhiro Okamoto, Nicolas Servant, Tristan Piolot, et al. 2012. "Spatial partitioning of the regulatory landscape of the X-inactivation centre." *Nature* 485 (7398): 381–85.
- Offield, M. F., T. L. Jetton, P. A. Labosky, M. Ray, R. W. Stein, M. A. Magnuson, B. L. Hogan, and C. V. Wright. 1996. "PDX-1 Is Required for Pancreatic Outgrowth and Differentiation of the Rostral Duodenum." *Development* 122 (3): 983–95.
- Oliver-Krasinski, Jennifer M., Margaret T. Kasner, Juxiang Yang, Michael F. Crutchlow, Anil K. Rustgi, Klaus H. Kaestner, and Doris A. Stoffers. 2009. "The Diabetes Gene Pdx1 Regulates the Transcriptional Network of Pancreatic Endocrine Progenitor Cells in Mice." *The Journal of Clinical Investigation* 119 (7): 1888–98.
- Phillips-Cremins, Jennifer E., Michael E. G. Sauria, Amartya Sanyal, Tatiana I. Gerasimova, Bryan R. Lajoie, Joshua S. K. Bell, Chin-Tong Ong, et al. 2013. "Architectural protein subclasses shape 3D organization of genomes during lineage commitment." *Cell* 153 (6): 1281–95.
- Rajewsky, Nikolaus, Geneviève Almouzni, Stanislaw A. Gorski, Stein Aerts, Ido Amit, Michela G. Bertero, Christoph Bock, et al. 2020. "LifeTime and Improving European Healthcare through Cell-Based Interceptive Medicine." *Nature* 587 (7834): 377–86.
- Rao, Suhas S. P., Miriam H. Huntley, Neva C. Durand, Elena K. Stamenova, Ivan D. Bochkov, James T. Robinson, Adrian L. Sanborn, et al. 2014. "A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping." *Cell* 159 (7): 1665–80.
- Řehůřek, R., and P. Sojka. 2010. "Software Framework for Topic Modelling Withlarge Corpora." In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, 45–50.
- Ricci, Maria Aurelia, Carlo Manzo, María Filomena García-Parajo, Melike Lakadamyali, and Maria Pia Cosma. 2015. "Chromatin Fibers Are Formed by Heterogeneous Groups of Nucleosomes in Vivo." *Cell* 160 (6): 1145–58.
- Sachdeva, Mira M., Kathryn C. Claiborn, Cynthia Khoo, Juxiang Yang, David N. Groff, Raghavendra G. Mirmira, and Doris A. Stoffers. 2009. "Pdx1 (MODY4) Regulates Pancreatic Beta Cell Susceptibility to ER Stress." *Proceedings of the National Academy of Sciences of the United States of America* 106 (45): 19090–95.
- Schaeffer, Marie, and Marcelo Nollmann. 2023. "Contributions of 3D Chromatin Structure to Cell-Type-Specific Gene Regulation." *Current Opinion in Genetics & Development* 79 (April):102032.
- Schoenfelder, Stefan, and Peter Fraser. 2019. "Long-Range Enhancer-Promoter Contacts in Gene Expression Control." *Nature Reviews. Genetics* 20 (8): 437–55.
- Schwarzer, Wibke, and François Spitz. 2014. "The Architecture of Gene Expression: Integrating Dispersed Cis-Regulatory Modules into Coherent Regulatory Domains." *Current Opinion in Genetics & Development* 27 (August):74–82.
- Sexton, Tom, Eitan Yaffe, Ephraim Kenigsberg, Fr d. Ric Bantignies, Benjamin Leblanc, Michael Hoichman, Hugues Parrinello, Amos Tanay, and Giacomo Cavalli. 2012. "Three-dimensional folding and functional organization principles of the Drosophila genome." *Cell* 148 (3): 458–72.
- Shen, Yin, Feng Yue, David F. McCleary, Zhen Ye, Lee Edsall, Samantha Kuan, Ulrich Wagner, et al. 2012. "A Map of the Cis-Regulatory Sequences in the Mouse Genome." *Nature* 488 (7409): 116–20.
- Szabo, Quentin, Axelle Donjon, Ivana Jerković, Giorgio L. Papadopoulos, Thierry Cheutin, Boyan Bonev, Elphège P. Nora, Benoit G. Bruneau, Frédéric Bantignies, and Giacomo Cavalli. 2020. "Regulation of Single-Cell Genome Organization into TADs and Chromatin Nanodomains." *Nature Genetics* 52 (11): 1151–57.
- Tabula Muris Consortium, Overall coordination, Logistical coordination, Organ collection and processing, Library preparation and sequencing, Computational data analysis, Cell type

- annotation, Writing group, Supplemental text writing group, and Principal investigators. 2018. "Single-Cell Transcriptomics of 20 Mouse Organs Creates a Tabula Muris." *Nature* 562 (7727): 367–72.
- Vermunt, Marit W., Di Zhang, and Gerd A. Blobel. 2019. "The Interdependence of Gene-Regulatory Elements and the 3D Genome." *The Journal of Cell Biology* 218 (1): 12–26.
- Vos, Erica S. M., Christian Valdes-Quezada, Yike Huang, Amin Allahyar, Marjon J. A. M. Verstegen, Anna-Karina Felder, Floor van der Vegt, Esther C. H. Uijttewaal, Peter H. L. Krijger, and Wouter de Laat. 2021. "Interplay between CTCF Boundaries and a Super Enhancer Controls Cohesin Extrusion Trajectories and Gene Expression." *Molecular Cell* 81 (15): 3082–95.e6.
- Wang, Chao-Yung, and James K. Liao. 2012. "A Mouse Model of Diet-Induced Obesity and Insulin Resistance." *Methods in Molecular Biology* 821:421–33.
- Wang, R. R., X. Qiu, R. Pan, H. Fu, Z. Zhang, Q. Wang, H. Chen, et al. 2022. "Dietary Intervention Preserves β Cell Function in Mice through CTCF-Mediated Transcriptional Reprogramming." *The Journal of Experimental Medicine* 219 (7). <https://doi.org/10.1084/jem.20211779>.
- Weinreb, Caleb, and Benjamin J. Raphael. 2016. "Identification of Hierarchical Chromatin Domains." *Bioinformatics* 32 (11): 1601–9.
- Wurmser, Annabelle, and Srinjan Basu. 2022. "Enhancer-Promoter Communication: It's Not Just About Contact." *Frontiers in Molecular Biosciences* 9 (April):867303.
- Xiao, Jordan Yupeng, Antonina Hafner, and Alistair N. Boettiger. 2021. "How Subtle Changes in 3D Structure Can Create Large Changes in Transcription." *eLife* 10 (July). <https://doi.org/10.7554/eLife.64320>.
- Xu, Jingxuan, Xiang Xu, Dandan Huang, Yawen Luo, Lin Lin, Xuemei Bai, Yang Zheng, et al. 2024. "A Comprehensive Benchmarking with Interpretation and Operational Guidance for the Hierarchy of Topologically Associating Domains." *Nature Communications* 15 (1): 4376.
- Yang, Jin H., and Anders S. Hansen. 2024. "Enhancer Selectivity in Space and Time: From Enhancer-Promoter Interactions to Promoter Activation." *Nature Reviews. Molecular Cell Biology* 25 (7): 574–91.
- Yang, Jin-Kui, Jing Lu, Sha-Sha Yuan, Asan, Xi Cao, Hai-Yan Qiu, Ting-Ting Shi, et al. 2018. "From Hyper- to Hypoinsulinemia and Diabetes: Effect of KCNH6 on Insulin Secretion." *Cell Reports* 25 (13): 3800–3810.e6.
- Ziebarth, Jesse D., Anindya Bhattacharya, and Yan Cui. 2013. "CTCFBSDB 2.0: A Database for CTCF-Binding Sites and Genome Organization." *Nucleic Acids Research* 41 (Database issue): D188–94.
- Zuin, Jessica, Gregory Roth, Yinxu Zhan, Julie Cramard, Josef Redolfi, Ewa Piskadlo, Pia Mach, et al. 2022. "Nonlinear Control of Transcription through Enhancer-Promoter Interactions." *Nature* 604 (7906): 571–77.
- Zwanzig, Robert. 2007. "Langevin Theory of Polymer Dynamics in Dilute Solution." In *Advances in Chemical Physics*, 325–31. John Wiley & Sons, Ltd.

Acknowledgements

We thank the laboratory members for the helpful discussion and feedback. We are grateful to Marcelo Rubinstein, Maria Cristina Gambetta, and David Lleres for their critical reading of the manuscript. This project was funded by the European Union's Horizon 2020 Research and Innovation Program (Grant ID 724429) (M.N.). We acknowledge the Bettencourt-Schueller Foundation for their prize 'Coup d'élan pour la recherche Française'. The CBS is a member of the France-Biolmaging, a national infrastructure supported by the French National Research Agency (ANR-10-INBS-04-01). O.M. was supported by an FRM and Ligue Contre la Cancer PhD fellowships. D.J.H. was supported by MRC (MR/S025618/1), Diabetes UK (17/0005681 and 22/0006389) and UKRI ERC Frontier Research Guarantee (EP/X026833/1)

Grants. This work was supported on behalf of the “Steve Morgan Foundation Type 1 Diabetes Grand Challenge” by Diabetes UK and SMF (grant number 23/0006627 to D.J.H. and I.A.).

Author Contributions

O.M., M.S. and M.N. conceived the study and the design. O.M., M.S., C.E.Y. acquired the data. O.M., J-B.F., M.S., C.E.Y, G.G., M.N. analyzed the data. O.M., J-B.F., G.G. and M.N. wrote analysis software. G.G and J-B.F performed loop extrusion simulations. J-B.F. built the microscope. O.M., M.S., J-B.F, G.G. and M.N. interpreted the data. M.S. and Y.K. prepared and characterized mice. I.A. and D.H. participated in experimental design. A.M and O.M. designed and tested RNA-FISH libraries. M.N., O.M., M.S. and J-B.F. wrote the manuscript. M.N. and M.S. supervised the study and acquired funds.

Competing interests statement

D.J.H. has filed patents on type 2 diabetes therapy.

Figure legends

Figure 1. Chromatin is organized in discrete Chromatin Folding Motifs (CFMs).

a Schematic illustrating the imaging-based strategy used to trace chromatin architecture at single-cell level in cryosectioned mouse tissues (Hi-M). **b** Top: Hi-M proximity frequency matrix of the *Pdx1* locus (chr5:146871445-147499625, mm10) from mouse pancreas generated with a cutoff distance of 100 nm, constructed from $n = 82754$ traces ($N = 5$ experiments, 2 different mice). Loops are indicated with arrows, while sub-TADs are marked with dashed lines. Middle: TADs are represented with orange rectangles. Barcodes used for Hi-M sequential imaging are represented as color-coded boxes. Gene locations are displayed along the locus. CTCF sites predicted from sequence (see *Methods*) are shown as triangles pointing to the right or to the left, depending on sequence orientation. Upright triangles represent peaks that are not predicted from sequence but that are bound by CTCF in pancreas (see blue plot below). Triangles in gray represent predicted sites that do not display CTCF signals in pancreas. Bottom: domainogram showing the insulation score derived from Hi-M data with different window sizes (see *Methods*) along with genomic coordinates. **c** Schematic illustrating 3DTopic workflow and application for single cell Hi-M data. Single cell 3D structures are used to determine a set of Chromatin Folding Motifs (CFMs) and each cell is iteratively assigned to a probability distribution for each CFM. **d** Probability distribution showing the decomposition of pancreatic cells into 30 CFMs. Red and blue represent high and low CFMs probability, respectively. **e** Subsets of CFMs, selected from the complete gallery of 30 CFMs of one decomposition of the *Pdx1* locus on pancreatic cells (see *Methods*). CFMs are categorized into two distinct categories according to the structure they describe (Domain and Loop). **f** Top: Hi-M proximity frequency matrices for the *Isl1* (chr13:116102031-116759941, mm10) and the *Ins2* locus (chr7:142520024-142699856, mm10) in mouse pancreas, and for IMR90 fibroblast cells (Bintu et al. 2018). Middle: Hi-M barcodes, genes, and predicted CTCF sites. Bottom: subsets of domain and loop CFMs, selected from the complete gallery of 30 CFMs from a single decomposition (see Figs. S3c, S4c, S5c for complete list). **g** Low-dimensional representation (UMAP) of single traces decomposed using the set of CFMs displayed in Figs. 1e, S2a. Single traces are color-coded by the CFM with the highest weight in the decomposition. Stars are used to display virtual decompositions containing a single CFM component (pure CFM). On the right, we provide two examples where the major CFM proximity matrix (top) is compared to the proximity matrix obtained by averaging the maps of all traces with the same major CFM. Training sample size: 6174 traces. **h** Graph representing the frequency at which loop and domain CFMs appear with at least 10% weight in pancreatic cells (representativity). Red circles represent domain CFMs, blue circles represent loop CFMs, and gray circles represent CFMs that are not attributed to either loops or domains (N/A). Other thresholds are shown in Fig. S6. **i** Matrix representing the co-representativity (conditional probability) of CFMs occurring in more than 10% in pancreatic cells. The matrix is asymmetric because of the higher frequency of domain over loop CFMs impacting the conditional probabilities. See Methods section for more details.

Figure 2. Loop extrusion and block copolymer models reproduce experimental CFMs.

a Gallery of 3 CFMs derived from the decomposition of a simulated chromatin domain, formed around adjacent chromatin accessible regions. The simulation was conducted using a block copolymer model (see *Methods*). **b** Gallery of 3 CFMs derived from the decomposition of a simulated chromatin loop, formed between distant convergent CTCF sites. The simulation was conducted using a polymer model based on loop extrusion (see *Methods*). **c** Subsets of CFMs, selected from the complete gallery of consensual CFMs (see *Methods*) of simulated Hi-M matrix along the *Pdx1* locus using a block copolymer model based on chromatin accessible regions determined by ATAC-seq data (SRR8119841) (C. Liu et al. 2019). **d** Subsets of CFMs, selected from the complete gallery of consensual CFMs (see *Methods*) of one decomposition of simulated Hi-M matrix along the *Pdx1* locus using the loop

extrusion model. CTCF sites are determined from CUT&Tag data (R. R. Wang et al. 2022). **e** Cross-correlation matrix comparing the experimental CFMs of one decomposition, including domains and loops, with consensual simulated CFMs generated using either the block copolymer model or the loop extrusion model (see *Methods*). **f** Example of one experimental CFM domain and one experimental CFM loop that are highly correlated and can be reproduced using polymer physics. **g** Graph representing the block-copolymer score (y) and the loop extrusion score (x) for each individual experimental CFM at the *Pdx1* locus. Domain CFMs are represented by red squares, while loop CFMs are depicted by blue circles.

Figure 3. Single cell structures from seven mouse tissues can be decomposed with the same set of CFMs.

a Hi-M proximity matrices for the *Pdx1* locus in seven different mouse tissues. *N* represents the number of biological replicates, and *n* the number of traces. **b** Probability distribution showing the decomposition of seven different mouse tissues in 30 CFMs. Red and blue represent high and low probability, respectively. **c** Density distribution of pancreatic cells in the UMAP landscape. Red and blue represent high and low density, respectively. Subsets of pure CFMs are displayed on the UMAP. **d** Density distribution in the UMAP landscape for six other mouse tissues. Colorbar is as in panel c. **e** Distribution of the odds ratio across 500 different decompositions for three distinct CFMs (0, 12, 17) and for all tissues compared to the pancreas (see *Methods*) with their associated CFM matrix on the right. The dashed red line indicates $\log_{10}(\text{odds ratio}) = 0.1$ and -0.1 . **f** Top: Matrix showing the odds ratio of a single decomposition for all domains CFMs in the different tissues compared to the pancreas along the *Pdx1* locus. Green indicates lower probabilities, and orange indicates higher probabilities in a given tissue as compared to the pancreas. 'ATAC status' indicates whether the barcodes involved in a CFM display a stable ATAC-seq signal (similar between tissues) or a variable ATAC-seq signal (changing between tissues). Bottom: Sankey diagrams showing the proportion of domains that are gained (orange), lost (green), or unaffected in the six different tissues compared to the pancreas across 100 different decompositions. **g** Same as f, but for loops CFMs. 'CTCF status' indicates whether the anchors of a CFM are stable or variable across tissues. We consider a CTCF status as stable only if both anchors are stable, otherwise the CFM is classified as variable.

Figure 4. Cell type specific 3D chromatin organization can be described by differential usage of CFMs.

a Schematic illustrating the different cell types of the pancreas. The pancreas is divided into exocrine and endocrine tissues, with the islet representing the basic unit of the endocrine tissue. Islets are constituted of various cell types, including alpha, beta, delta, and gamma cells. **b** Top: Grayscale maximum intensity projection of a DAPI-stained nuclei image, centered on a pancreatic islet. Exocrine cells (outside the contour in dashed lines) are defined as cells not displaying any of the endocrine markers (*Ins1/Ins2*, *Gcg*, *Sst*). Grayscale maximum intensity projection of *Ins1*, *Gcg* and *Sst* RNA-FISH markers, corresponding to alpha, beta, and delta-cells, respectively. Beta cells are defined as *Ins1*+ cells (inside solid line contour). Bottom: Corresponding proximity frequency matrices for exocrine and beta cells along the *Pdx1* locus. *n* represents the number of traces. CTCF sites are shown as black and white triangles, representing forward and reverse orientations, respectively. Pancreatic enhancers determined by enhancer atlas (<http://enhanceratlas.org/>) are represented by pink rectangles. **c** Left: Density distribution of exocrine tissue in the UMAP landscape. Right: Distribution of beta cells. Red and blue represent high and low density, respectively. **d** Differential UMAP density plot showing the relative density of exocrine and beta cells in the UMAP space. Blue indicates lower density, and red indicates higher density in beta cells compared to exocrine tissue. **e** Matrix showing the odds ratio of a single decomposition for all domains CFMs in beta cells compared to exocrine tissue. Green indicates lower probabilities, and orange indicates higher probabilities in beta cells relative to exocrine tissue. **f** Same as e, but for loops CFMs. **g** Differential PWD Hi-M matrices between beta cells and exocrine cells.

Blue and red represent higher and lower distances in beta cells compared to exocrine cells, respectively. **h** Volcano plot showing the $\log_2(\text{beta cells/exocrine cells})$ PWD distance (x-axis) and $-\log_{10}(\text{P-value})$. Gray circles indicate data points where the P-Value (y-axis) is below the 0.05. **i, j** Same as **g** and **h**, but for the higher resolution *Pdx1* Hi-M library. **k, l** Same as **g** and **h**, but for the *Isl1* locus. **m** Volcano plot showing the $\log_{10}$ odd ratio (x-axis) between exocrine cells and beta cells, and $-\log_{10}(\text{P-value})$ for each individual CFMs across 500 different decompositions. Domain CFMs are represented by squares, while loop CFMs are depicted by circles. Stars indicate CFMs related to the *Pdx1* gene.

Figure 5. Perturbation of transcription factors and CTCF levels changes CFM utilization.

a Schematic illustrating the perturbation model to induce type II diabetes (T2D). **b** Left: Grayscale DAPI-stained nuclei image of a mouse pancreas fed with normal diet (ND), centered on a pancreatic islet. Maximum intensity projection of the fluorescence signal from a beta cell marker (*Ins1*). Right: Microscopy images of mouse pancreas fed with a high fat diet (HFD). In both cases, islets are circled with yellow dashed lines. **c** Proximity frequency matrices along the *Pdx1* locus in beta cells under ND (left) and HFD (right) conditions. **d** Differential PWD Hi-M matrices between beta cells ND and beta cells HFD. Blue and red represent higher and lower distances in beta cells ND compared to beta cells HFD, respectively. **e** Volcano plot showing the $\log_2(\text{beta cells ND/ beta cells HFD})$ PWD distance (x-axis) and $-\log_{10}(\text{P-value})$. Gray circles indicate data points where the P-Value (y-axis) is below the 0.05. **f** UMAP density distribution of CFM decomposition of beta cell chromatin traces for ND (left) and HFD (right) conditions. **g** Differential UMAP density plot showing the relative density of beta cells ND and beta cells HFD in the UMAP space. Blue indicates higher density, and red indicates lower density in beta cells ND mice compared to beta cells in HFD mice. **h** Matrix showing the odds ratio of a single decomposition for all domains CFMs in beta cells HFD compared to beta cells HFD. Green indicates lower probabilities, and orange indicates higher probabilities in beta cells HFD relative to beta cells HFD. **i** Same as **g**, but for loops CFMs. **j** Volcano plot showing the $\log_{10}$ odd ratio (x-axis) between beta cells ND and beta cells HFD, and $-\log_{10}(\text{p-value})$ for each individual CFMs across 500 different decompositions. Domain CFMs are represented by squares, while loop CFMs are depicted by circles. Stars indicate CFMs related to the *Pdx1* gene.

List of tables

Table 1: Chromatin accessibility from ATAC-seq data and corresponding monomer in the heteropolymer chain

Table 2: CTCF positions, orientations and their LEF capture probabilities for loop extrusion simulations

Table 3: Lattice-based loop extrusion simulation parameters

Table 4: Molecular dynamics simulation parameters

Table 5: Barcode target bins for polymer coarse-graining

Supplementary Data

SUPPLEMENTARY DATA 1. List of sequences of primary Hi-M probes.

SUPPLEMENTARY DATA 2. List of sequence of imaging (io), adapter oligos and barcodes used in this study.

SUPPLEMENTARY DATA 3. List of sequences of primary RNA-FISH probes.

Supplementary Tables

SUPPLEMENTARY TABLE 1. List of publicly available data used in this study.

SUPPLEMENTARY TABLE 2. List of primers for library amplification used in this study.

Supplementary Figure legends

Supplementary Figure S1.

a Left: Grayscale DAPI-stained nuclei image of a mouse pancreas, centered on a pancreatic islet. Middle: Maximum intensity projection of the fluorescence signal from the entire *Pdx1* Hi-M library (Mask0). Right: Maximum intensity projection of the fluorescence signal from a single barcode (RT1) in the same field of view. **b** Top: Matrix displaying the histogram of pairwise distance distributions within all barcode pairs acquired on mouse pancreatic tissue. Bottom: Relative detection efficiency matrix normalized by the most detected barcode pair. Dark blue and white represent high and low relative detection efficiencies, respectively. **c** Bar-plot representing the detection efficiency for each barcode normalized by the most detected barcode. **d** Violin plots representing the Pearson correlation between the ensemble PWD matrix obtained using the full dataset and that using subsets containing different numbers of single allele data. Distributions were obtained by bootstrapping using 50 cycles per condition. **e** Median PWD matrices reconstructed for the *Pdx1* locus for each individual Hi-M experiment in mouse pancreas. *n* indicates the number of traces used for reconstructing each map. Two animals were used. **f** Matrix representing the Pearson correlation between each pair of Hi-M replicates in mouse pancreas. **g** Median Hi-M pairwise distance (PWD) matrix along the *Pdx1* locus (chr5:146871445-147499625) in mouse pancreas (mm10) constructed from *n* = 82754 traces (*N* = 5 experiments, 2 different mice). Barcodes used for Hi-M sequential imaging are represented as boxes with different colors. **h** Hi-C contact matrices from four different mouse tissues. Red and blue represent high and low contact frequencies, respectively. Data downloaded from accession number GSE150704 ((Leidescher et al. 2022)). **i** Pearson correlation coefficient between the interpolated Hi-C contact map and Hi-M proximity frequency matrix was calculated for various cutoff distances. The highest correlation was achieved at a threshold of 100 nm, as indicated.

Supplementary Figure S2.

a Complete gallery of chromatin folding motifs obtained from a decomposition along the *Pdx1* locus. The 30 CFMs are grouped into 2 distinct categories: Domains, Loops. Unidentified structures are marked as N/A. **b** 3DTopic reconstructed proximity frequency matrix following 3DTopic decomposition (see *Methods*) alongside the experimental proximity frequency map for the *Pdx1* locus in mouse pancreas. **c** Distribution of the coherence score for ten separate decompositions with varying numbers of CFMs (5, 10, 15, 20, 25, 30, 40, 50). The highest mean of coherence score was achieved with 30 CFMs, as highlighted (see *Methods*). **d** Probability distribution illustrating the decomposition of pancreatic cells into 30 CFMs for three independent decompositions along the *Pdx1* locus. Decompositions were obtained by selecting different single traces as training sets. Examples of three distinct CFMs identified in each decomposition are shown at the bottom. **e** Violin plot showing the distribution of genomic distances between each pair of bins for the two CFM categories, domains and loops.

Supplementary Figure S3.

a Median Hi-M PWD matrix and Hi-M proximity map (100 nm cutoff distance) for the *Isl1* locus (chr13:116102031-116759941) in mouse pancreas constructed from 12586 traces from 2 experiments. Barcodes, genes, and predicted CTCF sites are displayed along the locus. **b** Probability distribution illustrating the decomposition of pancreatic cells into 30 CFMs along the *Isl1* locus. **c** Complete gallery of chromatin folding motifs obtained from a decomposition along the *Isl1* locus.

Supplementary Figure S4.

a Median Hi-M PWD matrix and Hi-M proximity map (100 nm cutoff distance) for the *Ins2* locus (chr7:142520024-142699856) in mouse pancreas constructed from 9176 traces from 2 experiments. Barcodes, genes, and predicted CTCF sites are displayed along the locus. **b** Probability distribution illustrating the decomposition of pancreatic cells into 30 CFMs along the *Ins2* locus. **c** Complete gallery of chromatin folding motifs obtained from a decomposition along the *Ins2* locus.

Supplementary Figure S5.

a Median Hi-M PWD matrix and Hi-M proximity map (100 nm cutoff distance) from chromatin tracing data generated on *HCT 116* cell line (Bintu et al. 2018). **b** Probability distribution illustrating the decomposition of *HCT 116* cells into 30 CFMs along the presented locus. **c** Complete gallery of chromatin folding motifs obtained from a decomposition.

Supplementary Figure S6.

a Graph representing the global representativity of each CFM occurring in more than X% in pancreatic cells along the *Pdx1* locus, with X ranging from 5% (blue line), 10% (black line), 15% (gray line), to 20% (yellow line). Red circles represent domain CFMs, green circles represent loop CFMs, and gray circles represent CFMs that are not attributed to either loops or domains. **b** Same as a, but for the *Isl1* Hi-M library. **c** Same as a, but for the *Ins2* Hi-M library. **d** Same as a, but for the locus described in Bintu et al 2018. **e** 3DTopic UMAP, derived from CFM decomposition of pancreatic cells along the *Pdx1* locus, color-coded by the second most used CFM. Stars mark the precise locations of cells decomposed by a pure CFM, with a subset of these stars displayed in the color of their corresponding CFM.

Supplementary Figure S7.

a Potentials used to define forces between non-bonded monomers in the polymer chain during molecular dynamics simulations. Left: smoothed square-well potential for heteropolymers. Defines a common repulsive force between all types of monomers for excluded volume, and a soft attractive force only between monomers of the same type. Right: a truncated version of the above that only accounts for excluded volume in homopolymers. Distances are given in bond length units. **b** Logistic curve used to compute LEF capture probabilities by CTCF based on the latter's peak intensity from CUT&Tag data. Curve mid-point $x_0 = 390.231$, steepness $k = 0.01$. **c** Complete gallery of consensual CFMs obtained from polymer simulation using a block copolymer model (see *Methods*). **d** Same as c, but using a polymer model based on loop extrusion (see *Methods*). **e** Gallery of three CFMs derived from the decomposition of a loop extrusion simulation using two adjacent and convergent CTCF sites (see *Methods*). Due to the genomic proximity of these CTCF sites, the resulting main motif corresponds to a domain CFMs, with the second and third motif corresponding to other loop-extruded conformations extending to the right or the left of the two CTCF sites.

Supplementary Figure S8.

a RNA-seq (black), ATAC-seq (red) and CTCF ChIP-seq (Brain, Kidney, Liver, Lung, Thymus) or CUT&Tag (Pancreas) (blue) profiles (see *Methods*) along the *Pdx1* locus in six different tissues, with gene locations and CTCF binding sites (Tabula Muris Consortium et al. 2018; Shen et al. 2012; R. R. Wang et al. 2022). **b** Left: tSNE plot of cells from six different mouse tissues, color-coded by tissue type (see *Methods*). Right: tSNE plot displaying the expression levels of four different genes located along the *Pdx1* locus in single cells. **c** Median PWD matrices for the *Pdx1* locus for Pancreas, Kidney, Lymph node, Lung, Liver, Thymus and Brain. **d-e** Median PWD matrices and Hi-M proximity maps for the *Ins2* locus for the same tissues as in panel c (cutoff distance: 100 nm).

Supplementary Figure S9.

a 3DTopic reconstructed Hi-M matrix for the mouse pancreas and six other mouse tissues along the *Pdx1* locus. **b** Matrix displaying the Pearson correlation between the experimental Hi-M contact matrix and the 3DTopic reconstructed Hi-M matrix for each of the seven tissues tested. **c** Probability distribution illustrating one decomposition of cells from mouse pancreas and for six mouse tissues into 30 CFMs along the *Ins2* locus. **d** Density distribution in the UMAP landscape for the pancreas and six other mouse tissues along the *Ins2* locus. Subsets of pure CFMs are displayed on the UMAP. **e** Top: Matrix showing the odds ratio of a single decomposition for all domains CFMs in the different tissues compared to the pancreas. Green indicates lower probabilities, and orange indicates higher probabilities in a given tissue as compared to the pancreas along the *Ins2* locus. Bottom: Sankey diagrams showing the proportion of domains that are gained (orange), lost (green), or unaffected in the six different tissues compared to the pancreas across 100 different decompositions. **f** Same as f, but for loops CFMs along the *Ins2* locus.

Supplementary Figure S10.

a Violin plot showing the relative expression of *Pdx1* or *Plut* compared to *Actb* in single cells in exocrine cells and in beta cells (see *Methods*), Top and bottom, respectively. **b** Top: Hi-M proximity map along the *Pdx1* locus in mouse pancreas for the low resolution library. Bottom: Hi-M proximity map along the *Pdx1* locus in mouse pancreas for the high resolution library (chr5:147180024-147324699), (mm10) constructed from $n = 27462$ traces ($N = 2$ experiments, 2 different mice). **c** Top: Hi-M proximity map along the *Pdx1* locus for the high resolution library in exocrine pancreas. Bottom: Hi-M proximity map in beta cells. **d** Violin plot showing the relative expression of *Is1* compared to *Actb* in single cells in exocrine cells and in beta cells (see *Methods*). **e** Left: Hi-M proximity map along the *Is1* locus in exocrine

pancreas. Right : Median Hi-M proximity map in beta cells. **f** Matrix showing the odds ratio of a single decomposition for all domains CFMs in beta cells compared to exocrine tissue for the high resolution *Pdx1* Hi-M library . Green indicates lower probabilities, and orange indicates higher probabilities in beta cells relative to exocrine tissue. **g** Same as f, but for loops CFMs. **h** Volcano plot showing the log₁₀ odd ratio (x-axis) between exocrine cells and beta cells, and -log₁₀ (P-value) for each individual CFMs across 100 different decompositions. Domain CFMs are represented by squares, while loop CFMs are depicted by circles. Stars indicate CFMs related to the *Pdx1* gene. **i, j and k** Same as f, g and k, but for the *Is1* Hi-M library.

Supplementary Figure S11.

a Body weight (g) changes in mice fed with a normal diet (ND, black curve, n = 16) or a high-fat diet (HFD, blue curve, n = 19) over weeks. **b** Body weights after 14 weeks of treatment for ND-fed and HFD-fed mice. P-value < 0.01, Mann-Whitney *U* test. **c** Intraperitoneal glucose tolerance tests (IPGTT) in ND-fed and HFD-fed mice (see *Methods*). **d** Area Under the Curve (AUC) analysis of IPGTT curve in c *P* < 0.01, Mann-Whitney *U* test. **e** Left : Median Hi-M PWD matrices along the *Pdx1* locus in beta cells under ND conditions. Right : under HFD conditions.

Supplementary Figure S12.

a Left : Median Hi-M PWD matrices along the *Is1* locus in beta cells under ND conditions. Right : under HFD conditions. **b** Differential PWD Hi-M matrices between beta cells ND and beta cells HFD. Blue and red represent higher and lower distances in beta cells ND compared to beta cells HFD, respectively. **c** Volcano plot showing the log₂(beta cells ND/ beta cells HFD) PWD distance (x-axis) and -log₁₀ (P-value) for the *Is1* Hi-M library. Gray circles indicate data points where the P-Value (y-axis) is below the 0.05. **d** Matrix showing the odds ratio of a single decomposition for all domains CFMs in beta cells HFD compared to beta cells HFD. Green indicates lower probabilities, and orange indicates higher probabilities in beta cells HFD relative to beta cells HFD. **e** Same as d, but for loops CFMs. **f** Volcano plot showing the log₁₀ odd ratio (x-axis) between beta cells ND and beta cells HFD, and -log₁₀ (P-value) for each individual CFMs accross 100 different decompositions. Domain CFMs are represented by squares, while loop CFMs are depicted by circles. Stars indicate CFMs related to the *Is1* gene.









