Role of 3D Chromatin Conformation in Gene Regulation During Early Drosophila Embryogenesis and in Differentiated Mouse Tissues.

Olivier Messina

THÈSE POUR OBTENIR LE GRADE DE DOCTEUR DE L'UNIVERSITÉ DE MONTPELLIER

En Biologie Santé

École doctorale Sciences Chimiques et Biologiques pour la Santé, CBS2

Unité de recherche Centre de Biochimie Structurale (CBS), INSERM U1054

Étude du rôle de l'organisation tridimensionnelle du génome sur la régulation transcriptionnelle des gènes.

Présentée par Olivier MESSINA
Le 12 Décembre 2023

Sous la direction de Marcelo NOLLMANN et Marie SCHAEFFER

Devant le jury composé de

Daniel JOST, Directeur de Recherche, ENS, Lyon, France.	Président du Jury, Examinateur
Antoine COULON, Chargé de recherche, Institut Curie, Paris, France.	Rapporteur
Darío Jesús LUPIANEZ GARCIA, Directeur de Recherche, CABD, Seville, Espagne.	Rapporteur
Marcelo NOLLMANN, Directeur de Recherche, CBS, Montpellier, France.	Directeur de thèse
Marie SCHAEFFER, Chargée de recherche, CBS, Montpellier, France.	Directrice de thèse

UNIVERSITÉ DE MONTPELLIER

Remerciements

Quatre ans déjà... Ce manuscrit est la concrétisation de quatre années d'un travail aussi captivant qu'épanouissant, rendu possible grâce aux financements de Fondation de la Recherche Médicale et de la Ligue contre le Cancer. Je suis heureux d'avoir mené cette aventure à son terme, et je suis conscient que ce projet n'aurait pas eu la même saveur sans le soutien inestimable de toutes les personnes qui m'ont accompagné tout au long de ces années.

Je souhaite remercier les membres de mon jury de thèse qui ont accepté de me conseiller, d'évaluer ce travail et d'être présents à son dénouement. Leur expertise, leur temps et leur investissement ont grandement contribué à l'aboutissement de ma thèse.

C'est avec une immense gratitude que je souhaite adresser mes sincères remerciements à tous les enseignants universitaires qui ont contribué de manière significative à façonner mon parcours universitaire à l'IUT Génie Biologique et à l'Université des Sciences de Montpellier. Grâce à votre mentorat bienveillant, j'ai pu développer des compétences essentielles et acquérir de solides bases de connaissances. Votre dévouement et votre expertise ont été une source d'inspiration constante tout au long de mon parcours universitaire. Je serai toujours reconnaissant de l'opportunité d'avoir été guidé par des personnes aussi inspirantes et compétentes.

Je tiens à exprimer ma profonde gratitude envers les chercheurs ainsi qu'envers les personnes qui m'ont encadré tout au long de mon parcours universitaire, et qui ont grandement contribué à ma formation académique. Je suis particulièrement reconnaissant envers Simonetta Piatti, Reini Fernandez de Luco et Mounia Lagha pour m'avoir offert cette précieuse opportunité de travailler à leurs côtés, d'apprendre de leur expertise et de participer à des projets de recherche stimulants. Leur soutien indéfectible et leur influence ont eu un impact significatif sur ma trajectoire académique. Je tiens à les remercier sincèrement pour leur contribution à mon développement académique.

Je tiens à remercier toutes les personnes avec qui j'ai pu travailler tout au long de ma thèse dans le laboratoire du Dr. Marcelo Nollmann : Antoine Le Gall, Diego Cattoni, David Lleres, Andrés Cardozo-Gizzi, Franziska Barho, Markus Goetz, Sergio Espinola, Xavier Devos, Anna Mas, Maxime Andrieu, Alexandre Mesnager et Gautham Ganesh. Je tiens particulièrement à remercier Christophe Houbron, pour son aide infaillible dans mon apprentissage des protocoles expérimentaux et pour les moments précieux que nous avons partagés à la cantine du CNRS. Marion Bardou, pour sa bienveillance et pour nos discussions interminables dans les couloirs du labo. Julian Gurgo, pour sa gentillesse et pour tous les moments précieux que nous avons partagés autour d'un maté. Christel Elkhoury Youhanna, pour sa bonne humeur constante et son aide pendant les expériences. Jean-Bernard Fiche, "the master of the topics", pour sa patience, son aide précieuse et son dévouement constant. Je tiens à le remercier également pour sa précieuse contribution au bon fonctionnement

de l'épicerie via son alimentation constante en figues séchées.

Je tiens à remercier toutes les personnes qui ont pris le temps de relire ce manuscrit, et de me faire part de leurs commentaires constructifs.

Je souhaite remercier toutes les personnes avec lesquelles j'ai eu le privilège de travailler au cours de nos collaborations : Flavien Raynal et Vera Pancaldi. Par ailleurs, je tiens à exprimer ma gratitude envers les membres de mon comité de thèse, Daniel Jost, Giacomo Cavalli et Marcelo Rubinstein, pour leurs précieux conseils durant ces quatre années.

Je tiens à remercier ma directrice de thèse Marie Schaeffer pour son aide inestimable et ses précieux conseils qui m'ont permis de me surpasser tout au long de ma thèse. Son expertise en physiologie du pancréas a été d'une grande richesse dans mes travaux de recherche. Son soutien constant et la confiance qu'elle m'a accordée ont été des éléments essentiels pour surmonter les épreuves rencontrées. Je n'oublierai jamais l'odeur des pâtes carbonara du vendredi midi dans notre bureau partagé (la grotte).

Je tiens à remercier tout particulièrement mon directeur de thèse Marcelo Nollmann pour son soutien et son engagement. Je suis également reconnaissant de la confiance qu'il m'a accordé en m'impliquant dans divers projets et collaborations. Sa vaste expertise dans des domaines aussi variés que la biologie, la biophysique, la bioinformatique, l'informatique, la modélisation, l'imagerie, l'analyse d'images, et bien d'autres, a constamment nourri ma soif d'apprentissage et m'a permis de me surpasser pour donner le meilleur de moi-même. Cela a été un véritable privilège d'avoir pu réaliser cette thèse avec toi.

Enfin, je tiens à remercier tous mes proches qui m'ont soutenu pendant ces quatre années d'aventure : Brume, Ruby et Luma, mes amis, mes cousines, mes tantes et oncles, les blaireaux (Florian et Anaïs), ainsi qu'à mes beaux-parents, mais surtout mes parents et mes grands-parents. Pour finir, je tenais à remercier ma chérie, Fanny, qui a été d'un soutien infaillible tout au long de ce projet sur "les mouches et les souris". Sa force et sa résilience ont été une source d'inspiration infinie.

Une pensée très forte à toutes ces étoiles qui brillent intensément

Prenez soin de vos rêves. [...]

Résumé en français

Dans le noyau des cellules eucaryotes, la molécule d'ADN, qui porte l'information génétique, s'associe avec des protéines telles que les histones pour former la chromatine. Au cours de la dernière décennie, les techniques de "capture de la conformation des chromosomes" ont révélé que la chromatine était organisée de manière hiérarchique en domaines physiques appelés TADs (*Topologically Associating Domains*, en anglais). Les TADs jouent un rôle cléf dans la régulation génique en facilitant les interactions physiques entre les gènes et leurs éléments régulateurs tels que les enhancers, les promoteurs et les insulateurs. Cependant, les mécanismes moléculaires et les facteurs impliqués dans la formation de ces structures pendant le développement embryonnaire, ainsi que leur influence sur l'activation et la répression des gènes dans les tissus différenciés, demeurent mal compris.

L'objet de ces travaux est d'étudier le rôle de certains facteurs dans l'établissement des interactions entre les gènes et leurs éléments régulateurs afin de mieux comprendre leur impact sur la régulation de la transcription chez la drosophile. Pour ce faire, nous avons utilisé des approches de bioinformatique et une technique de microscopie avancée (Hi-M). Grâce à ces outils novateurs, nous avons pu mettre en évidence l'importance des protéines insulatrices, dite de "Classe I", dans la mise en place de ces interactions au cours du développement embryonnaire chez la drosophile. Nous avons montré que les interactions entre les régions liées par les protéines insulatrices de Classe I étaient peu fréquentes et principalement observées en paire, et que leur formation précédait l'émergence des TADs au cycle nucléaire 14. Par la suite, nos recherches se sont penchées sur le rôle de ces interactions entre les gènes et leurs éléments régulateurs au sein des tissus différenciés chez les mammifères. Pour cela, nous avons généré des données Hi-M et nous avons développé un nouvel outil de machine learning permettant la décomposition de la structure de la chromatine en unités fondamentales, que nous avons nommées "Motifs de Pliage de la Chromatine" (*Chromatin Folding Motifs*, en anglais). Ceci nous a permis de montrer qu'un ensemble de motifs CFMs était essentiel pour expliquer la structure d'un locus spécifique. En conséquence, nous avons pu démontrer que l'architecture tridimensionnelle varie non seulement d'un tissu à l'autre, mais également d'un sous-type cellulaire à l'autre en ajustant la proportion d'un même ensemble de CFMs. Finalement, nous avons également montré que l'architecture 3D était affectée lors de la genèse de maladies telles que le diabète de type II. Dans l'ensemble, ces études ont pour but d'ouvrir de nouvelles perspectives pour appréhender les mécanismes à l'œuvre dans l'organisation du génome en 3D.

English abstract

In the nucleus of eukaryotic cells, the DNA molecule, that carries genetic information, associates with proteins such as histones to form chromatin. Over the past decade, techniques known as "chromosome conformation capture" have revealed that chromatin is hierarchically organised into physical domains called Topologically Associating Domains (TADs). TADs play a critical role in gene regulation by facilitating physical interactions between genes and their regulatory elements such as enhancers, promoters, and insulators. However, the molecular mechanisms and factors involved in the formation of these structures during embryonic development, as well as their influence on gene activation and repression in differentiated tissues, remain poorly understood.

In this study, we investigated the role of certain factors in establishing interactions between genes and their regulatory elements to better understand their impact on transcriptional regulation in *Drosophila*. To do this, we used bioinformatic approaches and an advanced microscopy technique (Hi-M). These innovative tools allowed us to highlight the importance of Class I insulator proteins in mediating these interactions during embryonic development in *Drosophila*. We were able to show that interactions between regions bound by Class I insulator proteins are rare, mainly observed in pairs, and that their formation precedes the emergence of TADs during the nuclear cycle 14. Subsequently, our research focused on the role of interactions between genes and their regulatory elements in differentiated tissues in mammals. To this end, we produced Hi-M data and developed a novel machine-learning tool to decompose chromatin structure into fundamental units, which we termed "Chromatin Folding Motifs" (CFMs). This allowed us to show that a set of CFMs is essential for explaining the structure of a given locus. As a result, we were able to show that the three-dimensional architecture varies not only between tissues, but also between different cell-types by adjusting the proportion of the CFMs. Finally, we showed that 3D architecture is affected during the onset of diseases such as type II diabetes. Collectively, these studies aim to open new perspectives for understanding the mechanisms involved in 3D genome organisation.

Long résumé en Français

Mise en contexte

L'ADN est le support de l'information génétique de nos cellules. Une mutation dans l'ADN peut conduire à l'apparition de maladies génétiques complexes. Ces dernières années, il est devenu évident que non seulement la séquence d'ADN mais aussi l'épigénome (c'est-à-dire la manière dont le génome est interprété) sont critiques dans l'apparition de nombreuses maladies, telles que certains cancers ou des malformations embryonnaires. Il est maintenant bien établi que la structure tridimensionnelle du génome est intimement liée à l'épigénétique. Dans le noyau des cellules eucaryotes, l'organisation tridimensionnelle du génome est un processus hautement régulé. Ces dernières années, le développement de techniques dite de "capture de la conformation des chromosomes" telles que la technique Hi-C [1] ont permis de révéler l'importance de la conformation 3D de l'ADN dans la régulation de nombreux processus biologiques comme la réplication, la différenciation ou encore la transcription. Ainsi, des défauts dans l'organisation 3D du génome sont retrouvés dans de nombreux types de cancers. Néanmoins, les mécanismes impliqués dans la régulation de la conformation tridimensionnelle du génome sont encore mal caractérisés. De ce fait, comprendre comment les interactions physiques entre les éléments *cis*-régulateurs du génome sont mises en place durant le développement embryonnaire et comment elles sont modelées pendant le processus de différenciation pour assurer différents programmes de transcription reste une question majeure en génétique.

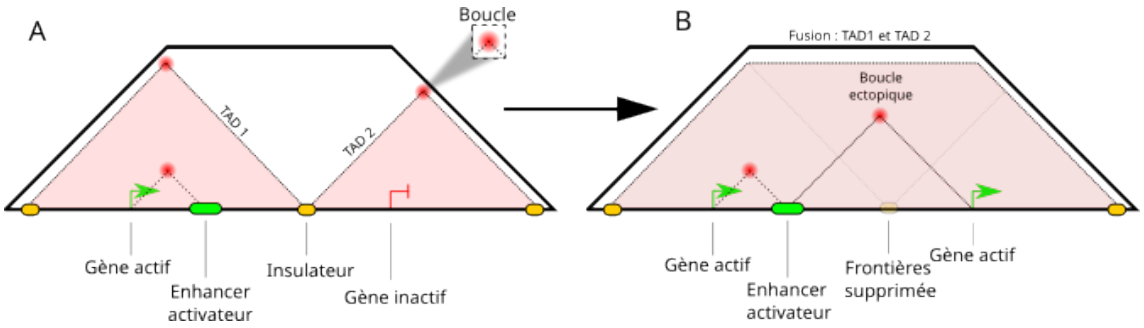


Figure I: Représentation schématique des TADs et des interactions entre les éléments régulateurs. - Sur chacune des deux représentations (A-B), le génome est représenté de manière horizontale. Les TADs apparaissent sous forme de triangles rouges. Les frontières de TADs sont représentées sous la forme de rectangle jaunes et caractérisées par la présence de protéines insulatrices (insulateur). (A) Schéma représentant la régulation d'un gène par une région activatrice (enhancer activateur) par l'intermédiaire d'interaction physique (boucle) dans le TAD 1. Le gène dans le TAD 2 est réprimé. (B) Schéma représentant la suppression de la frontière entre le TAD 1 et le TAD 2. Cette suppression entraîne la fusion des deux TADs adjacents ce qui conduit à l'activation ectopique du gène de droite suite à la formation de boucle ectopique avec l'élément activateur (enhancer) de gauche.

La technique de Hi-C a permis de montrer que le génome est localement organisé en domaines physiques appelés TADs ("Topologically Associating Domains" en anglais) [2, 3, 4]. L'organisation en TADs semble être un trait important de la fonction du génome car elle apparaît être conservée au cours de l'évolution de nombreux organismes eucaryotes tels que la mouche, la souris ou encore l'humain [5]. Dans la majorité des cas, les TADs englobent les gènes et leurs éléments *cis*-régulateurs (enhancers, promoteurs, et insulateurs) afin de favoriser leurs interactions physiques pour réguler l'activation et la répression transcriptionnelle des gènes. Les frontières des TADs sont souvent enrichies en protéines insulatrices (CTCF chez les mammifères) qui permettent de restreindre les intéractions entre les éléments *cis*-régulateurs des TADs voisins. Ainsi, la perturbation de l'organisation des TADs via la modification de leurs frontières peut entraîner des défauts d'intéractions physiques entre les éléments *cis*-régulateurs qu'ils contiennent (**Figure I** : Boucle ectopique.). Ces défauts peuvent conduire à des dérégulations transcriptomiques pouvant causer des anomalies développementales (e.g. polydactylie) [6] ou l'activation ectopique d'oncogènes à l'origine de nombreux cancers (e.g. leucémies) [7]. Récemment, il a été montré que la structure en TADs du génome émerge rapidement pendant le développement embryonnaire et coïncide avec l'établissement du programme transcriptionnel du zygote (**Figure II**) [6, 8, 9]. Néanmoins, les mécanismes moléculaires et les facteurs impliqués dans l'établissement de ces structures pendant le développement embryonnaire ainsi que leurs rôles sur l'activation et la répression des gènes dans les tissus différenciés font à ce jour l'objet d'études approfondies [10, 11, 11, 12, 13, 14]. De ce fait, la compréhension des mécanismes de formation des TADs ainsi que leurs rôles dans la régulation de l'expression des gènes pourrait aider à mieux appréhender comment des dérégulations de cette architecture 3D du génome peuvent provoquer l'apparition de maladies.

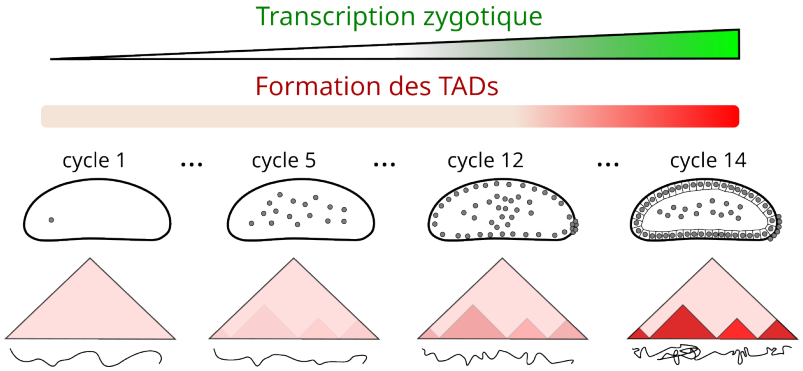


Figure II: Représentation schématique de l'émergence des TADs pendant le développement embryonnaire. - De gauche à droite sont représentés de manière schématique des embryons de drosophile à différents cycles de développement (cycles nucléaires 1,5,12 et 14) avec leur carte schématique Hi-C en bas. Les gradients vert et rouge représentent respectivement le niveau croissant de transcription zygotique et de formation de TADs au cours du développement embryonnaire.

Le but de mon projet de thèse est de mieux comprendre les mécanismes impliqués dans la formation des intéractions physiques entre les éléments *cis*-régulateurs et de disséquer leurs rôles dans la régulation de l'expression des gènes. Pour cela, j'ai

décomposé mon projet en deux objectifs majeurs qui vont être décrits sous forme de chapitres dans ce résumé en français :

- **Chapitre 1 :** Étudier le rôle des insulateurs dans le partitionnement du génome de la drosophile pendant le développement embryonnaire.
- **Chapitre 2 :** Étudier comment les réseaux d'interaction entre les éléments *cis*-régulateurs sont modulés pour réguler la transcription dans les tissus différenciés de la souris en conditions physiologiques et pathologiques.

Chapitre 1 :

Objectif : Étudier le rôle des insulateurs dans le partitionnement du génome de la drosophile pendant le développement embryonnaire.

Ce premier objectif de thèse a été réalisé dans le laboratoire du Dr. Marcelo Nollmann en collaboration avec le laboratoire du Dr. Vera Pancaldi.

Introduction

Chez les métazoaires, les premières étapes du développement embryonnaire sont contrôlées par des ARN et des protéines d'origine maternelle. Ce n'est qu'après quelques heures/jours suivant la fécondation que l'embryon devient transcriptionnellement actif et prend le contrôle de son propre développement, ce processus étant connu sous le nom de transition maternelle zygotique (MZT : "*Maternal to Zygotic Transition*" en anglais). La MZT se caractérise par deux événements importants : la dégradation des ARN maternels et l'activation du génome zygotique (ZGA : "*Zygotic Genome Activation*" en anglais). Chez la drosophile, l'embryogenèse commence par 14 cycles nucléaires rapides qui se déroulent au sein d'un syncytium. Environ deux heures après la fécondation, le génome zygotique s'active et la cellularisation débute. L'activation transcriptionnelle du génome zygotique se caractérise par l'acquisition de l'architecture tridimensionnelle du génome et l'apparition des TADs. Chez les mammifères, le couple de protéine cohesine/CTCF joue un rôle central dans la coordination du processus d'extrusion qui conduit à la formation des TADs [15]. Cependant, contrairement aux mammifères, l'organisation des domaines chez la drosophile ne repose généralement pas sur des boucles médiées par CTCF [16]. À la place, la drosophile possède un ensemble distinct de protéines de liaison aux insulateurs (IBPs) qui ne sont pas conservées au cours de l'évolution [17]. Ces facteurs incluent Su(Hw), BEAF-32, mod(mdg4), CP190, dCTCF, GAF, Zw5, Pita, ZIPIC, Z4, Chriz/Chromator, Fs(1)h [18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28]. Les insulateurs sont des facteurs capables de partitionner le génome, notamment en empêchant certains enhancers distaux d'agir sur le promoteur de gènes. Bien que le rôle des insulateurs semble crucial pour réguler la communication entre enhancers et promoteurs, leurs rôles dans la modulation de l'architecture 3D du génome et dans l'établissement des réseaux d'interactions physiques restent à ce jour mal compris.

Dans cette étude, nous avons étudié le rôle de chacun de ces facteurs liants les insulateurs dans la régulation de l'organisation tridimensionnelle du génome chez la drosophile.

Approches méthodologiques

Pour étudier le rôle des protéines de liaison aux insulateurs (IBPs) de la drosophile sur la mise en place de contact entre les éléments *cis*-régulateurs, nous avons utilisé une combinaison d'analyses bioinformatiques ainsi qu'une technique d'imagerie multiplexée de la chromatine: Hi-M.

Analyses bioinformatiques

Pour étudier le rôle des facteurs liants les insulateurs dans la formation des contacts entre les éléments *cis*-régulateurs du génome, nous avons d'abord développé un outil bioinformatique (**Figure III**) que nous avons utilisé pour mesurer le niveau d'interaction entre les régions occupées par les facteurs liants les insulateurs lors de différents cycles développementaux chez la drosophila (cycles nucléaires 12-13-14 et 3-4hpf). Cet outil intègre les informations de liaison de facteurs à la chromatine (ChIP-seq) [29] avec des données Hi-C acquises au cours du développement embryonnaire chez la drosophile [6] (**Figure III**).

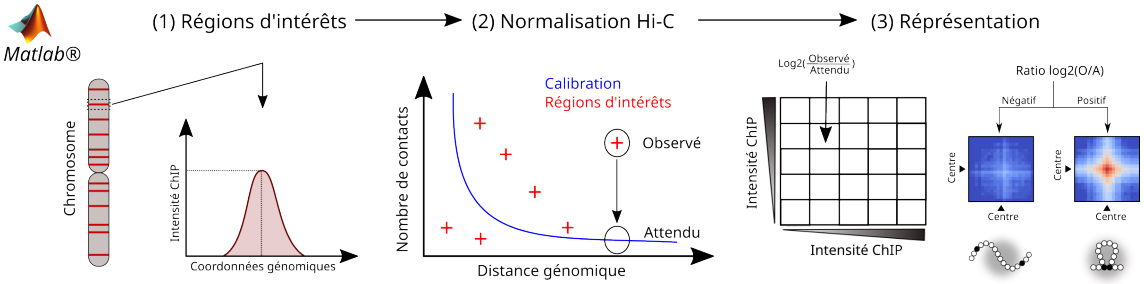


Figure III: Représentation schématique du programme bioinformatique développé pour étudier le rôle des facteurs liants la chromatine dans la formation des contacts physiques entre les éléments régulateurs. - De gauche à droite sont représentées les différentes étapes de l'analyse. (1) La sélection des régions d'intérêt basée sur les données ChIP-seq, (2) la normalisation des données Hi-C par la distance (3) et la représentation des résultats. Si le ratio log2(observé/attendu) est positif alors le facteur testé est impliqué dans la formation de contact physique. Sinon le facteur n'est pas impliqué dans l'établissement de contact.

Imagerie multiplexée de la chromatine : Hi-M

Jusqu'à présent, la plupart des méthodologies utilisées pour étudier le repliement tridimensionnel du génome reposent sur le séquençage, ce qui présente plusieurs limitations. L'une des principales limitations est l'absence d'informations spatiales sur l'échantillon étudié, rendant impossible l'analyse de la relation entre l'architecture

tridimensionnelle du génome, la transcription et la localisation des cellules au sein des tissus. Afin de surmonter ces limitations et d'accéder à l'information spatiale manquante, nous [9] et d'autres [30, 31, 32] avons développé de nouvelles technologies basées sur l'imagerie pour étudier l'architecture de la chromatine. La visualisation de plusieurs loci génomiques dans une seule expérience à l'aide de la technique ADN-FISH est essentielle pour révéler la structure tridimensionnelle d'un locus donné. Dans cette méthodologie innovante, au lieu d'imager simultanément différentes sondes fluorescentes, comme dans une méthode d'imagerie ADN-FISH conventionnelle, plusieurs loci génomiques sont imagés de manière séquentielle avec le même fluorophore grâce à des cycles d'imagerie et de photoblanchiment. Cela est rendu possible grâce à un dispositif microfluidique sophistiqué, programmé pour délivrer différentes solutions de lavage et d'hybridation (**Figure IV**).

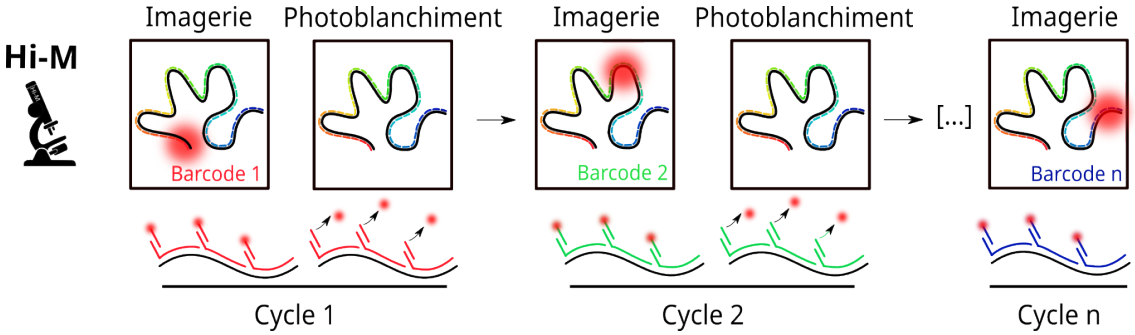


Figure IV: Représentation schématique de la stratégie d'imagerie Hi-M. - De gauche à droite sont représentés différents cycles d'imagerie et de photoblanchiment de la stratégie d'imagerie Hi-M.

Résultats

Grâce à l'outil bioinformatique développé (**Figure III**) et à des analyses complémentaires d'assortativité de réseau d'interaction réalisées par Flavien Raynal dans le groupe du Dr. Vera Pancaldi [33], nous avons pu démontrer que certains facteurs liant les insulateurs (Classe I) participent à la formation de contacts entre les éléments *cis*-régulateurs du génome chez la drosophile **(Figure V**). Parmi les insulateurs de Classe I figurent : BEAF-32, CHRO, DREF, ZIPIC, Zw5 et Z4. Les insulateurs de Classe II incluent L(3)MBT, Pita, CP190, Fs(1)h, CBP, Mod(mdg4), GAF et SU(HW).

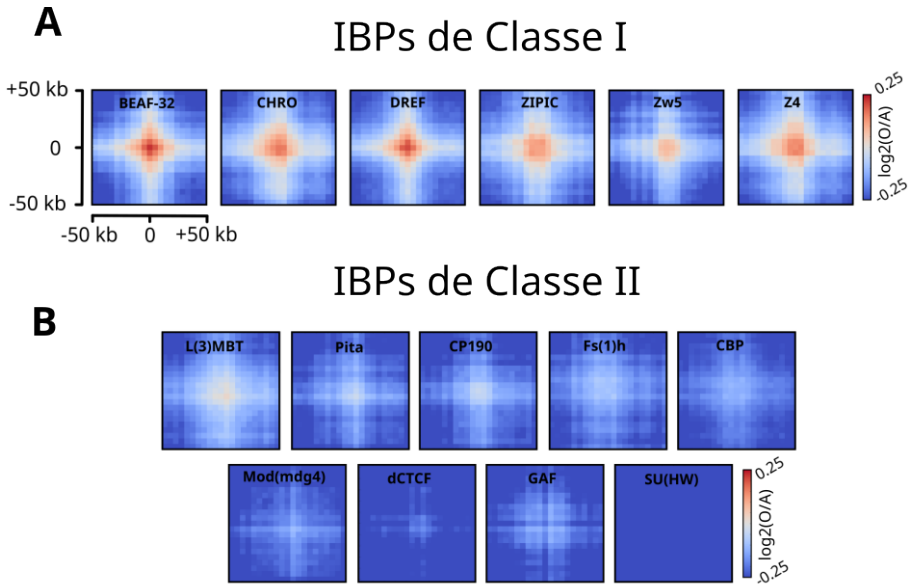


Figure V: Les insulateurs de Classe I favorisent la formation d'intéractions physiques entre les éléments *cis*-régulateurs. - (A) Matrice d'aggregation de la moyenne du niveau d'interaction centrée sur la région liée par le facteur étudié (Insulateur de Classe I). La présence d'une croix rouge au centre indique que ce groupe d'insulateur favorise la formation de contact entre les régions qu'ils lient. (B) Matrice d'aggregation de la moyenne du niveau d'interaction pour les insulateurs de Classe II. L'absence de croix rouge au centre indique que ce groupe d'insulateur ne favorise pas la formation de contact entre les régions qu'ils lient.

En utilisant des données Hi-C [6] acquises dans différents stades de développement embryonnaire chez la drosophile, nous avons pu démontrer que les intéractions entre les éléments *cis*-régulateurs liés par les insulateurs de Classe I apparaissent précocement lors du développement embryonnaire, avant l'activation transcriptionnelle du zygote (NC12), et qu'une faible proportion de ces sites (environ 14%) étaient liés par Zelda. Ces résultats suggèrent que les insulateurs de Classe I sont impliqués dans la formation de contacts indépendamment de la liaison du facteur pionnier Zelda (**Figure VI**).

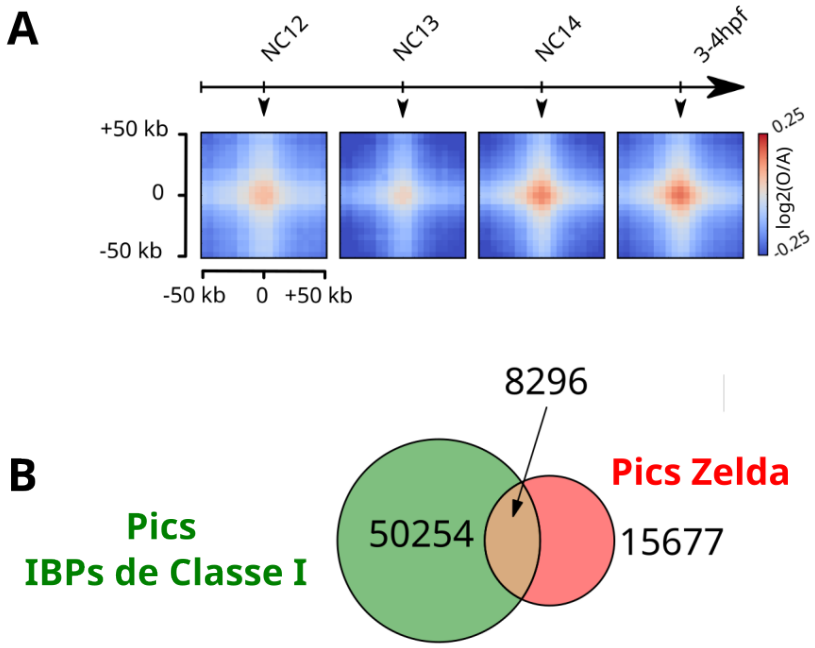


Figure VI: Les insulateurs de Classe I favorisent la formation d'interactions physiques entre les éléments *cis*-régulateurs avant l'activation transcriptionnelle et indépendamment de Zelda. - (A) Matrice d'aggregatio de la moyenne du niveau d'interaction des Insulateurs de Classe I dans 4 cycles développementaux (NC12, NC13, NC14 et 3-4hpf).(B) Diagramme de Venn représentant l'union entre les régions liées par les Insulateurs de Classe I et les régions liées par Zelda.

Par la suite, nous avons utilisé Hi-M pour quantifier la probabilité absolue de contact entre les régions liées par les insulateurs de Classe I le long d'une région chromosomique d'intérêt. Nous avons pu montrer que la fréquence de contact entre les régions liées par les insulateurs de Classe I était très légèrement supérieure à la fréquence de contact d'un set de régions contrôle non liées par les insulateurs de Classe I. Cette quantification montre que les régions liées par les insulateurs de Classe I interagissent dans environ 12% des cellules imagées (**Figure VII**).

Bien que les analyses bioinformatiques à l'échelle du génome révèlent que les insulateurs de Classe I interagissent plus fréquemment que la moyenne, les quantifications de la fréquence d'interaction par la technique Hi-M sont en contradiction

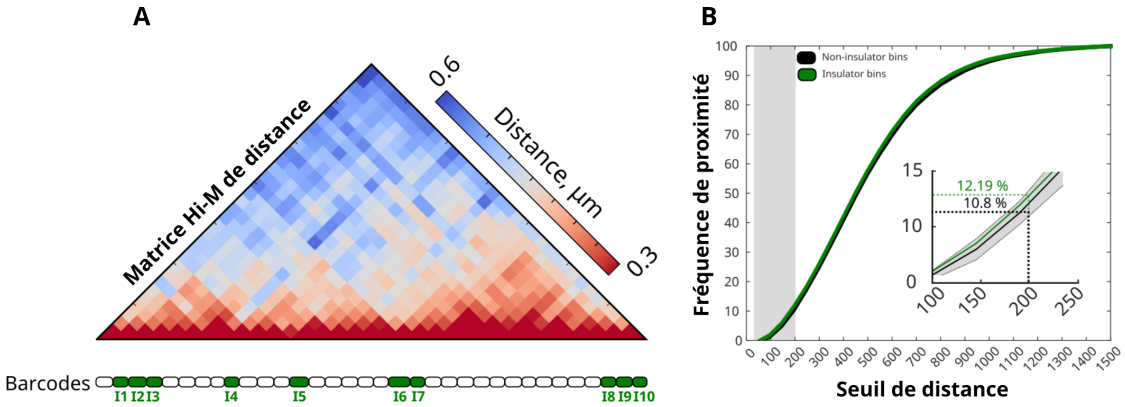


Figure VII: Quantification de la fréquence de contacts entre les régions liées par les insulateurs de Classe I sur une région chromosomique d'intérêt par Hi-M. - (A) Matrice Hi-M de distance sur une région chromosomique d'intérêt. Les barcodes Hi-M liés par les insulateurs de Classe I sont colorés en vert en dessous de la matrice. (B) Quantification de la fréquence de proximité entre paires de barcodes liés par les insulateurs de Classe I pour différents seuils de distances (courbe verte) et pour un set de régions contrôles.

avec un modèle de formation de boucles stables entre les éléments cis-régulateurs du génome chez la drosophile. Ces résultats sont présentés dans la thèse sous la forme d'un article accepté dans la revue "Nature Communications" (voir 3.1).

Chapitre 2 :

Objectif : Étudier comment les réseaux d'interaction entre les éléments cis-régulateurs sont modulés pour réguler la transcription dans les tissus différenciés de la souris en conditions physiologiques et pathologiques.

Ce second objectif de thèse a été réalisé dans le laboratoire du Dr. Marcelo Nollmann.

Introduction

L'organisation tridimensionnelle de la structure de la chromatine est considérée comme cruciale pour réguler les transitions dynamiques entre différents programmes transcriptionnels lors de la différenciation tissulaire. Cependant, de récentes études, y compris celles menées au sein du laboratoire d'accueil [12, 13], ont révélé que certaines interactions entre les enhancers et les promoteurs de gènes étaient identiques entre différents tissus pré-différenciés au début du développement embryonnaire chez la drosophile. Ces observations remettent en question l'hypothèse selon laquelle les interactions entre les éléments *cis*-régulateurs diffèrent d'un tissu complètement différencié à un autre pour réguler la transcription chez les mammifères.

Dans cette étude, notre objectif était d'explorer le rôle de l'architecture tridimensionnelle du génome dans la régulation des programmes transcriptionnels spécifiques aux tissus à l'échelle de la cellule unique chez la souris.

Approches méthodologiques

Pour explorer la modulation des réseaux d'interactions entre les éléments *cis*-régulateurs au cours de la différenciation chez les mammifères, nous avons d'abord adapté la technique Hi-M (**Figure IV**) aux cryosections de tissus différenciés de souris. Ensuite, nous avons développé un algorithme de machine learning basé sur la modélisation de topics, nommé 3DTopic, dans le but de décomposer la conformation de la chromatine à la cellule unique d'un locus particulie en blocs fondamentaux (CFMs, abréviation de *"Chromatin Folding Motifs"* en anglais).

Decomposition de la structure en blocs fondamentaux : 3DTopic

3DTopic repose sur une approche statistique non supervisée innovante basée sur la modélisation des topics. En d'autres termes, 3DTopic exploite les données Hi-M à la cellule unique relatives à la conformation de la chromatine d'un locus spécifique, puis les analyse afin d'identifier un ensemble de blocs fondamentaux (CFMs) permettant d'expliquer la conformation de ce locus particulier. Ensuite, chaque cellule peut être décomposée en fonction des différents CFMs, où chaque

CFM est associé à un poids qui détermine la manière dont la cellule présente ce repliement spécifique (**Figure VIIIA**).

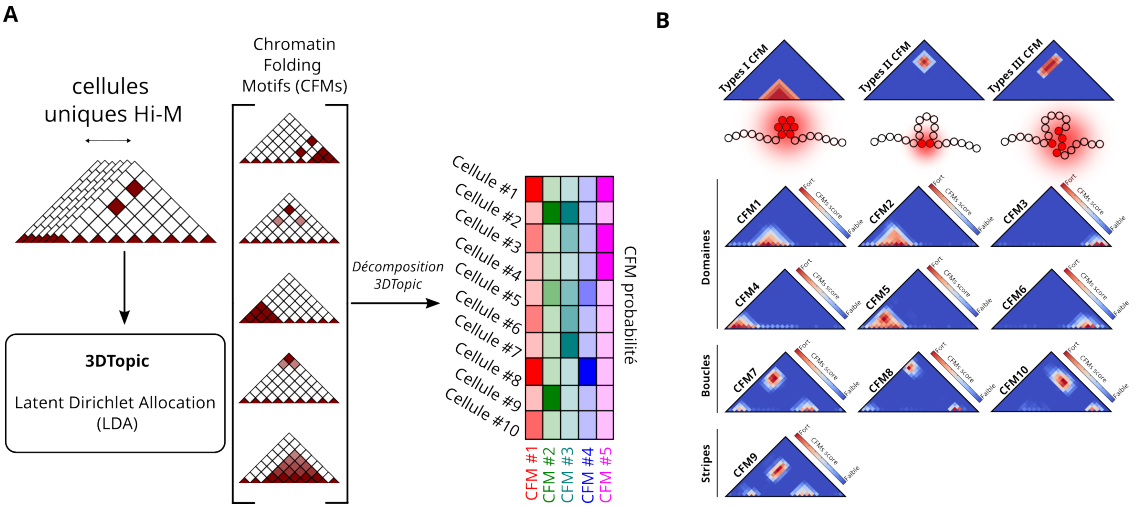


Figure VIII: Analyse par 3DTopic de la conformation de la cellule unique. - (A) Représentation schématique de la stratégie 3DTopic de décomposition de la structure en blocs fondamentaux (CFMs). (B) Représentation des différents blocs fondamentaux obtenus à partir de données Hi-M sur des cellules pancréatiques de souris. Les blocs sont catégorisés en trois catégories en fonction de la structure chromatinienne décrite dans les CFMs, à savoir les domaines, les boucles et les stripes.

Résultats

Grâce à l'adaptation de la technique Hi-M sur des cryosections de tissus différenciés de souris, nous avons réussi à reconstruire des cartes d'interactions physiques au niveau d'un locus d'intérêt dans le pancréas, ainsi que dans six autres tissus distincts, à savoir le cerveau, le poumon, le rein, le foie, le ganglion et le thymus. Les cartes d'interactions obtenues pour ces différents tissus révèlent que l'architecture 3D semble varier d'un tissu à l'autre (**Figure IXA**). Pour mieux comprendre ces variations, nous avons utilisé l'approche 3DTopic afin d'identifier les CFMs spécifiques au pancréas. Cet ensemble de CFMs englobe des domaines, des boucles ainsi que des interactions en lignes (stripes), pouvant être modélisés par la physique des polymères. Enfin nous avons appliqué cette décomposition à l'ensemble des cellules provenant de tous les tissus. De manière remarquable, nous avons observé que le même ensemble de CFMs, utilisé à des proportions différentes, permettait de décrire l'architecture 3D d'un locus d'intérêt dans les différents organes avec une grande précision (**Figure IXB**). Les changements observés dans la proportion d'utilisation des CFMs dans différents tissus pourraient signifier que les mécanismes de repliements opèrent différemment dans divers tissus, peut-être en raison de répertoires distincts de protéines de liaison à la chromatine dans le locus étudié.

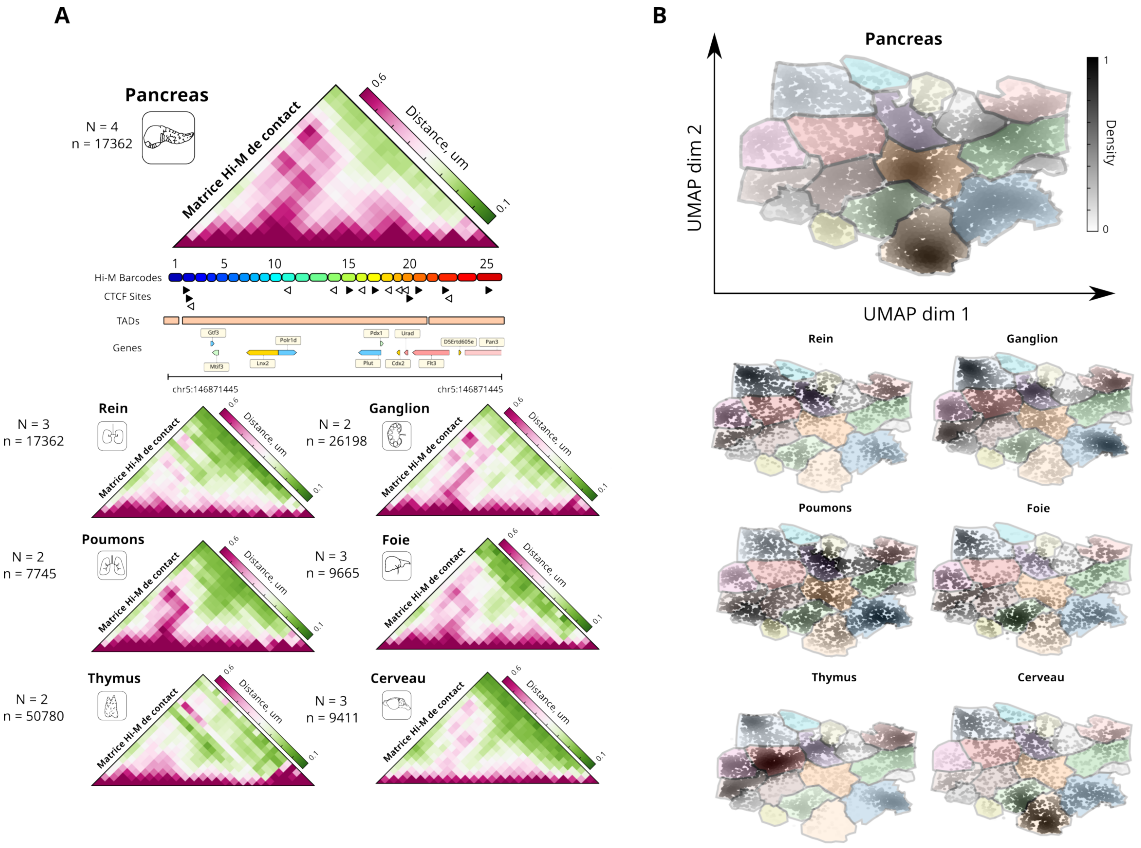


Figure IX: Cartes Hi-M d'interaction dans différents tissus différenciés chez la souris. - Matrices Hi-M de contact représentant une région chromosomique d'intérêt pour différents tissus de la souris : le pancréas, le rein, le poumon, le thymus, Le ganglion, Le foie et le cerveau. Les barcodes Hi-M sont illustrés par des rectangles de couleurs situés en-dessous de la matrice, avec l'annotation des sites CTCF, des TADs et des gènes le long du locus.

Pour étudier si les changements observés dans les réseaux d'interactions à l'échelle de différents tissus existent également au niveau des sous-types cellulaires, nous avons appliqué ces développements technologiques à l'étude des îlots de Langerhans du pancréas, qui sont composés de différents sous-types cellulaires, notamment les cellules β (exprimant l'insuline : Ins1), les cellules α (exprimant le glucagon : Gcg), les cellules γ (exprimant le polypeptide pancréatique : Ppy) et les cellules δ (exprimant la somatostatine : Sst). Pour distinguer ces différents sous-types cellulaires, nous avons conçu et validé une bibliothèque de sondes fluorescentes [34, 35] permettant de détecter plusieurs espèces d'ARN par la méthode de seqARN-FISH (**Figure XA-B**). Ensuite, nous avons couplé cette expérience avec la méthode d'imagerie Hi-M. L'analyse des cartes d'interactions reconstruites dans les différents sous-types cellulaires des îlots de Langerhans du pancréas a révélé des changements dans les réseaux d'interaction au sein de la région étudiée (**Figure XC**).

De plus, nous avons étudié comment ces réseaux sont affectés lors de l'apparition du diabète de type 2 (T2D), qui s'accompagne d'une réduction significative de l'expression de CTCF dans les îlots et d'une perte d'accessibilité de la chroma-

tine [36]. Nos données Hi-M chez les souris atteintes de T2D mettent en évidence d'importantes altérations de la structure de la chromatine au sein des sous-types cellulaires pancréatiques, notamment au niveau des boucles, des stripes et des domaines (**Figure XC**). Ces changements architecturaux sont également en accord avec la dérégulation de l'expression du gène Pdx1, présent dans le locus d'intérêt, après l'induction du T2D [37].

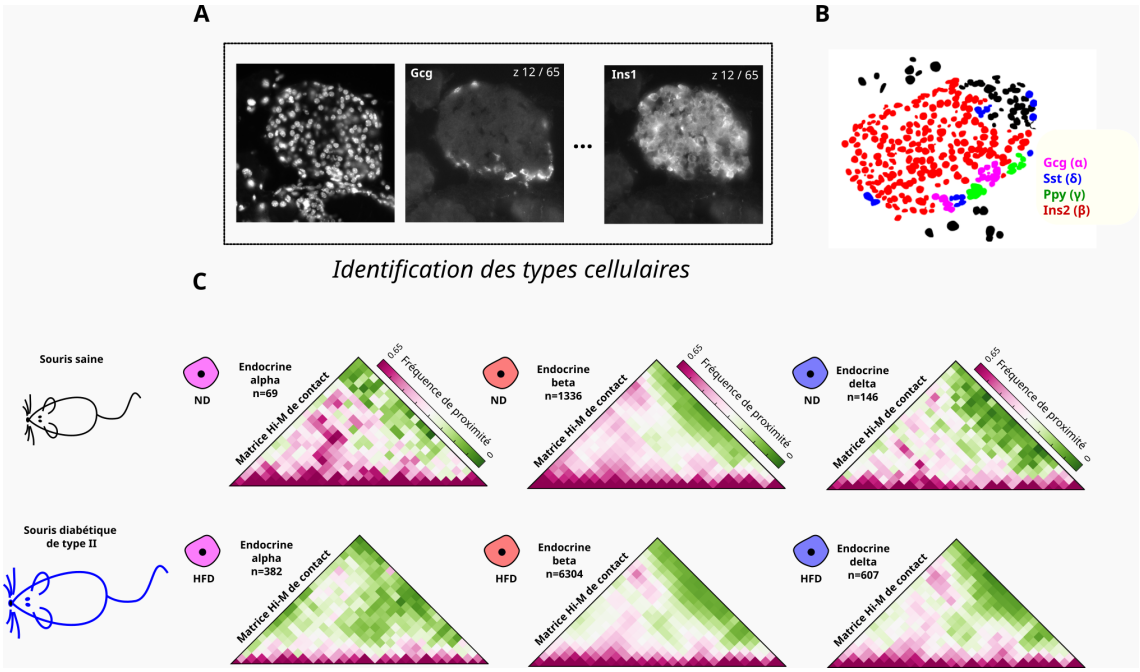


Figure X: Cartes Hi-M d'interaction dans les différents types cellulaires des îlots de Langerhans du pancréas. - (A) Images en microscopie à fluorescence d'un marquage DAPI (gauche) et d'ARN FISH séquentiel (seqARN-FISH) (milieu et droite) contre : Gcg et Ins2 au niveau d'un îlot de Langerhans. (B) Image reconstituée des sous-types cellulaires identifiées dans l'îlot imagé grâce au marquage seqARN-FISH. (C) Matrices Hi-M de contact dans les différents types cellulaires des îlots de Langerhans du pancréas. Gauche (α), Milieu (β), Droite (δ) en condition physiologique (haut) et lors de l'induction du diabète de type 2 (bas).

Ces résultats sont présentés dans la thèse sous la forme d'un article en préparation (voir 3.2).

Table of contents

List of acronyms

3C : Chromosome Conformation Capture
4C : Circular Chromosome Conformation Capture
5C : Carbone Copy Chromosome Conformation Capture
ADLD : adult-onset demyelinating leukodystrophy
A : Adenine
BX-C : Bithorac Complex
CDK : cyclin-dependent kinas
CFM : Chromatin Folding Motif
CHD : Chromodomain, Helicase, DNA binding
ChAs : Chromatin assortativity (ChAs)
ChIP : Chromatin immunoprecipitation
C : Cytosine
CIDs : Chromosomal Interacting Domains
CREs : Cis-Regulatory Elements
CRMs : Cis-Regulatory Modules
CRISPR : Clustered Regularly Interspaced Short Palindromic Repeats
CT : Chromosome Territories
DNA : Deoxyribonucleic Acid
DBD : DNA-Binding Domain (DBD)
FISH : Fluorescence in situ hybridization
G : Guanine
GAM : Genome Architecture Mapping
GWAS : Genome-Wide Association Studies
GTFs : General Transcription Factors (GTFs)
HFD : High Fat Diet
HMG : high-mobility group box
hpf : hours post-fertilization
IBPs : Insulator-binding proteins
IDRs : Intrinsically disordered regions (IDRs)
INO80 : INOsitol requiring 80
LCR : Locus control region (LCR)
LEFs : Loop extruding factors (LEFs)
ME : Mesendoderm
MERFISH : Multiplexed error-robust fluorescence in situ hybridization
mESC : Mouse embryonic stem cells.
MSC : Mesenchymal Stem Cells
NC : Nuclear cycle
ND : Normal Diet
NDR : Nucleosome Depleted Region
NGS : Next Generation Sequencing
NFR : Nucleosome Free Region
NPC : Neural Progenitor Cells
ORCA : Optical reconstruction of chromatin architecture
O/E : Observed versus expected counts (O/E)

PCR : Polymerase chain reaction
PIC : Pre-initiation complex
PEV : Position-effect variegation
PRC1/2 : Polycomb Repressive Complex 1/2
PTMs : PostTranslational Modifications
RNAPII : RNA Polymerase II
RNA : RiboNucleic Acid
SCR : Sox2 Control Region
SHH : Sonic hedgehog
SMC : Structural maintenance of chromosomes
STFs : Specific Transcription Factors
SWI/SNF : SWItching defective/Sucrose NonFermenting
T : Thymine
TAD : Topologically Associating Domain
TBP : TATA box binding protein
TFs : Transcription factors
TSS : Transcription Start Site
U : Uracile
ZGA : Zygotic genome activation (ZGA)

List of Figures

Prelude

The discovery of DNA

In biology, *abiogenesis* refers to the natural process by which life has arisen from non-living matter. For centuries, the definition of life has been a challenge for scientists and philosophers. More than 100 definitions of life have been compiled [38]. Most dictionaries define life as the quality that distinguishes the living from the dead. However, there is currently no consensus on the definition of life. Despite the impressive diversity among living organisms, they all share common structural and functional features. From a genetic point of view, life is underpinned by a genetic blueprint, a recipe book, that controls the production of essential molecules called proteins. This blueprint, the molecule of life, known as DeoxyriboNucleic Acid (DNA), was first discovered in the 1860s by Joahannes Friedich Miescher (reviewed in [39]). In his experiment, Dr. Friedich Miescher precipitated the nucleus of white blood cells and found that this substance was resistant to protease digestion, indicating that it was not a protein. He then characterized the composition of this substance as consisting of a five-carbon sugar attached to one or more phosphate groups and a nitrogenous base. As he isolated this substance from cell nuclei, he named it *nuclein* in a publication in 1874 (reviewed in [39]). This early discovery was complemented in 1881 by Albrecht Kossel, who identified *nuclein* as what is now called nucleic acid. He then went on to characterize the nitrogen-containing bases: the purine bases (Adenine : A and Guanine : G), the pyrimidine bases (Thymine : T and Cytosine : C) and confirmed that these basic building blocks are restricted to the nucleus (**Figure 1A**). In 1879, Walther Flemming pioneered the use of aniline dyes to visualize the behavior of the nucleus of salamander embryos during cell division. He observed that the dye was strongly absorbed by a structure in the nucleus that he named *chromatin*, from the Greek word for color (khrôma, "color"). This staining strategy allowed Flemming to first observe how DNA is distributed between the mother and daughter cells during the process of cell division to produce two identical halves. He beautifully summarized his observations in a series of elegant drawings between 1878 and 1888 and named this process of cell division: mitosis, from the Greek word for thread (mito, "thread") (**Figure 2**).

In the 1900s, Walter Sutton and Theodor Boveri independently proposed that the chromosome could be the carrier of genetic information to which the Mendelian laws of inheritance could be applied. It wasn't until 1944 that Oswald Avery, Colin MacLeod and Maclyn McCarty confirmed Walter Sutton and Theodor Boveri's hypothesis by demonstrating in bacteria that DNA, and not the proteins themselves, carries the genetic information of cells. However, the mechanisms by which proteins are encoded by the information carried by DNA were still a mystery. In 1949 and 1950, Erwin Chargaff and co-workers found that the molar ratios of $\frac{A}{T}$ and $\frac{C}{G}$ on the DNA molecule were very close to 1 (known as "Chargaff's rule"). In contrast, the molar ratios of $\frac{A}{C}$, $\frac{A}{G}$, $\frac{T}{C}$, and $\frac{T}{G}$ could differ significantly from 1. This observation was crucial to the later work of Watson and Crick, but Chargaff could not conceive of an explanation for these relationships. Advances in X-ray analysis by Rosalind Franklin

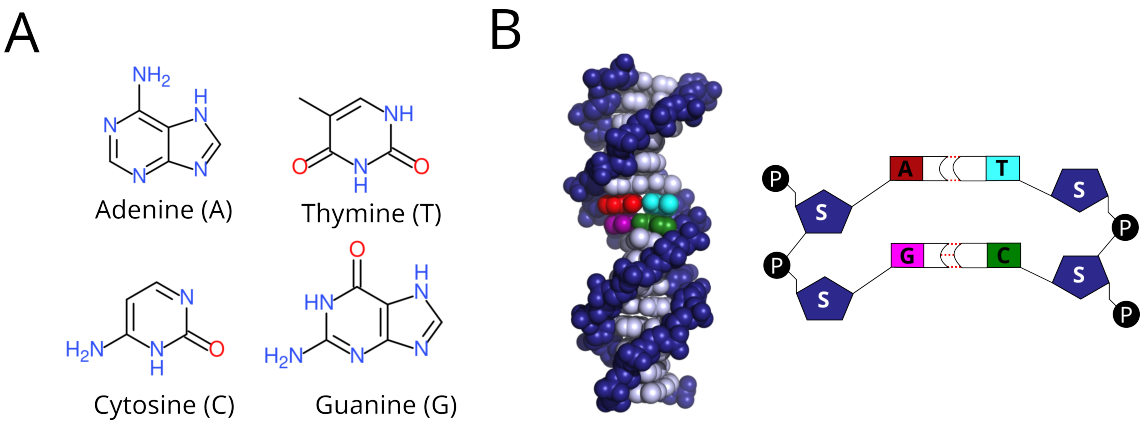


Figure 1: The structure of helical DNA - (A) The primary building blocks of DNA : the nitrogenous bases. (B) The polynucleotide chain of deoxyribonucleic acid with the sugar-phosphate DNA backbone in blue and the base pairs in gray. Zoom in on two specific pairs of nucleobases in the center of the structure.

and Maurice Wilking contributed significantly to Watson and Crick's double helix model for the structure of DNA [40]. With the help of Jerry Donohue, Francis Crick and James Watson resolved the configuration of each of the elements of DNA (the carbon, nitrogen, hydrogen and oxygen rings and the nitrogen-containing bases) and found that complementary bases fit together (i.e. A bases are always paired with T and C with G), in agreement with Chargaff's rule (**Figure 1B**).

The regulation of genetic flow

Once the structure of DNA was solved, scientists tried to make the connection between DNA and proteins. Francis Crick proposed that an unstable molecule, known as RiboNucleic Acid (RNA), could serve as a temporary copy of the genetic material (DNA) to make proteins in a unidirectional and irreversible manner. This led to the formulation of the "central dogma" of molecular biology to explain how genetic information is transferred between all 3 major classes of biopolymers (DNA, RNA and proteins). In this dogma, there are $3 \times 3 = 9$ conceivable links between the polymers, but for Francis Crick, only four of them were thought to exist in nature (*i* to *iv* in **Figure 3**). However, the discovery of an enzyme called RNA-dependent DNA polymerase, also known as reverse transcriptase, by Howard Temin and David Baltimore in 1978 challenged the long-held idea of a unidirectional flow of genetic information as proposed by the central dogma. This enzyme, found in some viruses, allowed the conversion of RNA into DNA, leading to the suggestion that the central dogma needed to be revised (see *v* arrow in **Figure 3**). So far, this very complex view of the "central dogma" has served as a scaffolding to better understand how the regulation of genetic information governs life processes.

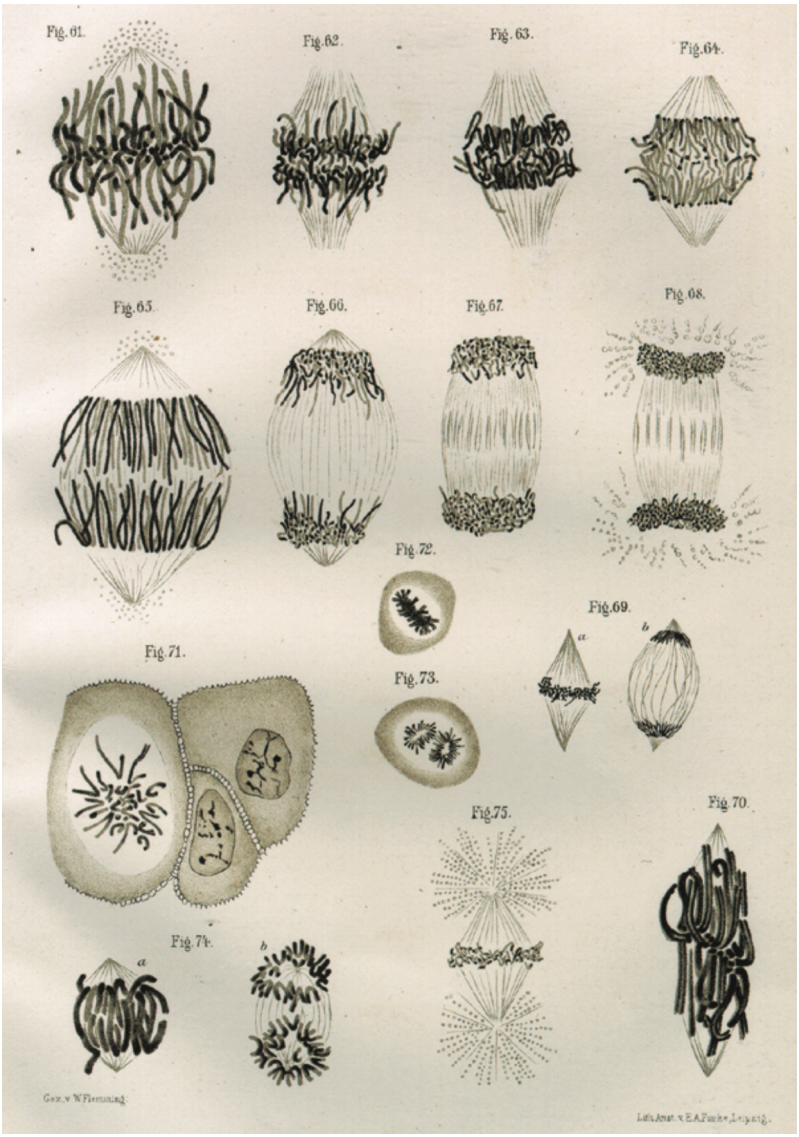


Figure 2: Flemming's drawings of mitosis - Drawings illustrating the major steps of the cell division. Image reproduced from [41]

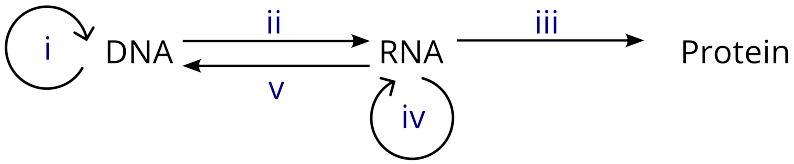


Figure 3: Central dogma of molecular biology - Schematic representation of the central dogma of molecular biology, showing all the possible conversions between DNA, RNA and protein : (*i*) DNA replication, (*ii*) Transcription, (*iii*) Translation, (*iv*) RNA replication, and (*v*) Reverse transcription.

The paradox of the DNA sequence

The genome, defined as the DNA instructions found in a cell, contains all the information needed for an individual to develop, function and reproduce. In less

than 50 years, from 1953 to 2003, the field of genetics has evolved significantly, from the solving of the structure of DNA by Watson and Crick in 1953 to the publication of the complete sequence of the human genome in 2003 (Human Genome Project). One of the major unexpected discoveries of recent years has been that the cells of an adult organism do not undergo irreversible changes when they begin to differentiate, contrary to what was proposed by Briggs and King in the early 1950s. Sir John Gurdon played a pivotal role in this discovery by demonstrating that the nuclei of intestinal epithelial cells from feeding tadpoles, when transferred into enucleated eggs, could develop into adult frogs. This groundbreaking experiment proved that every cell in a multicellular organism has identical DNA instructions [42]. Despite the fact that the cells have the same DNA instructions, they can be different in both form and function because they synthesize different sets of RNA and protein molecules through different gene expression programs. The vision of differential cell decision making was pioneered by Conrad Waddington in 1957 in his famous Waddington's landscape (**Figure 4A**), which conceptualized how cells can specialize during embryogenesis by shaping their landscape through differential gene expression (**Figure 4B**).

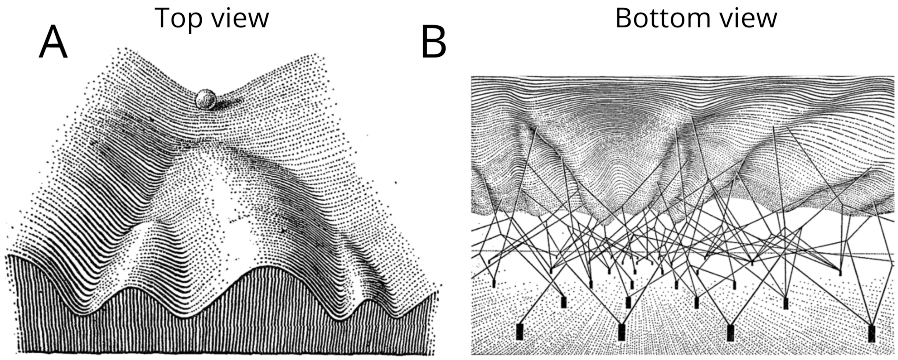


Figure 4: The Waddington landscape - (A) Top view of the landscape. The sphere represents a cell crossing the metaphorical landscape of cell differentiation. (B) Bottom view of the landscape. Pillars (representing genes) and ropes (representing the functional output of the genes) that modulate the landscape. Figure adapted from [43].

The regulation of gene expression

Gene regulation is a fundamental process that controls the expression of genes in living organisms. It plays a critical role in determining the specific protein repertoire that must be present in a given cell at a given time point. This regulation can occur at different levels, such as transcriptional, post-transcriptional, translational, and post-translational levels, and is influenced by a range of internal and external factors, such as environmental cues, cellular signaling pathways, and epigenetic modifications. The concept of gene regulation was first introduced in 1961 by François Jacob and Jacques Monod who observed that the production of β-galactosidase in *Escherichia coli* is controlled by the presence or the absence of "inducer" (lactose) in

the media : the PaJaMo experiment (**Figure 5**). Following this observation, Jacob and Monod presented a model for the regulation of gene expression : the operon model (**Figure 6**). The lac operon consists of three structural genes (lacZ, lacY and lacA) and one regulatory gene (lacI) encoding a repressor protein. In the absence of an "inducer" (lactose) in the medium. The repressor protein LacI binds to the "operator" to block the transcription of the polygenic mRNA (lacZYA) (**Figure 6A**), preventing the unnecessary production of enzymes. Conversely, in the presence of an "inducer" in the medium (lactose), the lactose binds to the repressor protein LacI and prevents it from inhibiting the expression of the lac operon genes. This allows the structural genes to be transcribed and the proteins they encode to be produced, allowing the bacteria to use lactose as an energy source.

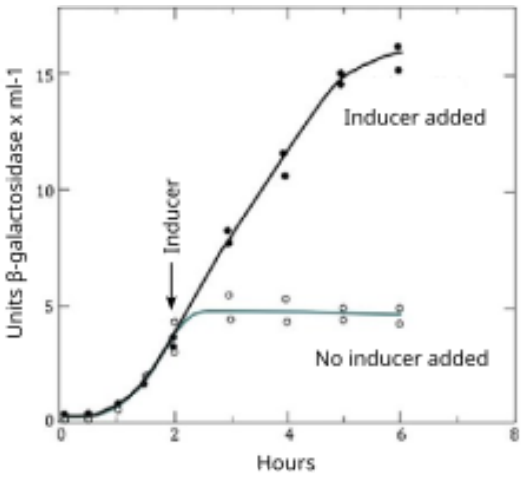


Figure 5: The PaJaMo experiment - Graph representing the evolution of units of β-galactosidase x ml^{-1} over time (measured in hours) in the presence and absence of an inducer. Figure adapted from [44]

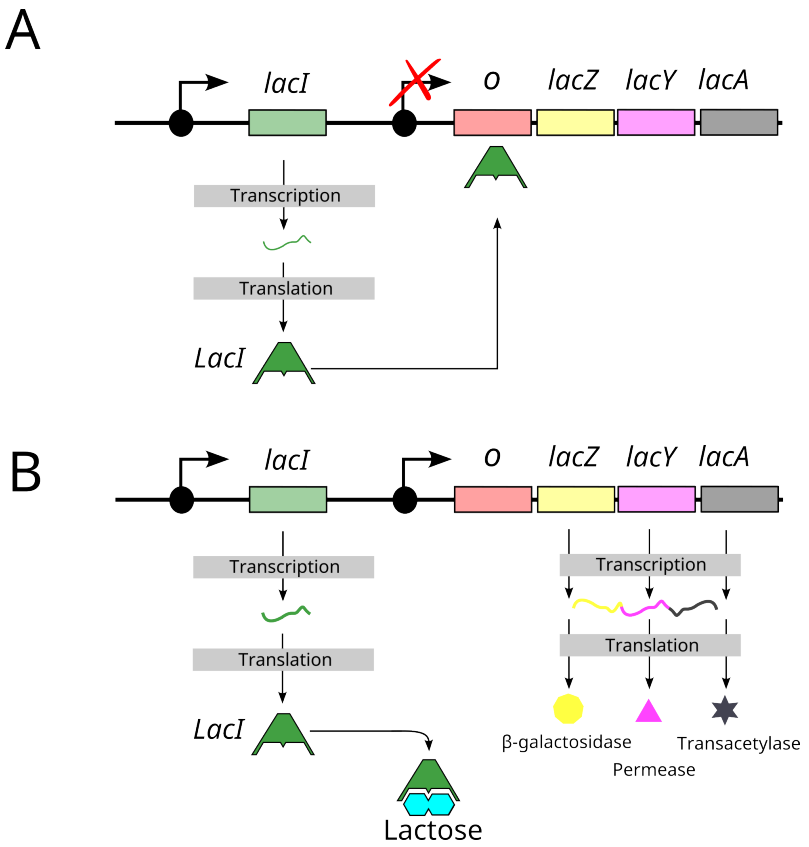


Figure 6: The lac operon - Schematic representation of the lactose operon in repressed (A) and induced state (B). The operon consists of three structural genes (lacZ, lacY and lacA) under the control of a single promoter (P) and operator (O). In the absence of inducer (lactose), the LacI repressor binds to the operator to inhibit transcription. When lactose is present in the media, it binds to the LacI repressor protein, causing a conformational change that releases it from the operator.

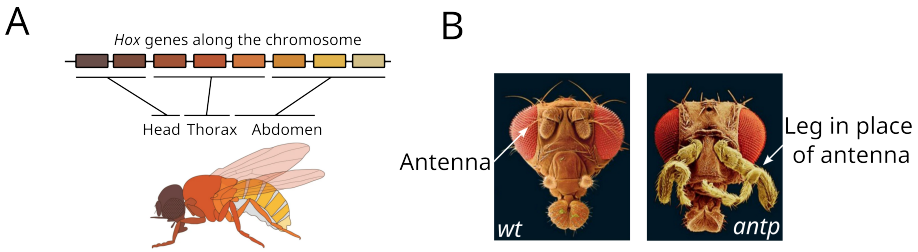


Figure 7: The homeotic genes in *Drosophila* - (A)Top: *Hox* genes distribution along the fruit fly chromosome. Bottom : The expression pattern of Hox genes along the anterior-posterior axis in the adult. (B) Right : A wild-type fruit fly. Left an *Antennapedia* mutant fruit fly.

Spatial and temporal regulation of gene expression is extremely important in the development of multicellular organisms. A striking example is the regulation of the *Hox* genes along the anterior-posterior (AP) axis in both invertebrates and vertebrates. In 1978, work by E.B. Lewis showed that the *Hox* genes play a critical role in assigning distinct spatial identities to cells in different regions along the AP axis in the fruit fly. The specific combination of homeotic genes expressed in a given segment determines whether a cell in that segment belongs to the head, thorax or abdomen of the fly (**Figure 7A**). Misexpression of a *Hox* gene called *antennapedia* in the head precursor cells of the *Drosophila* results in the ectopic formation of legs in place of antennae (Schneuwly et al. 1987) (**Figure 7B**).

Thus, understanding the mechanisms of gene regulation at the single cell level is crucial not only for basic biological processes but also for understanding how the dysregulation of gene expression can lead to the development of various diseases.

Chapter 1: Introduction

1 Hierarchical organization of the genome : From Structure to Function

Across the whole tree of life, the range of genome sizes is quite broad, with some genomes differing by as much as 8 orders of magnitude in size from <2 kb for hepatitis D virus to up to 100 Gbp for *Marbled lungfish*. Prokaryotes generally have a very compact genome, with less than 10% being non-coding DNA. On the other hand, eukaryotes, such as humans, have a much larger genome, with only about 2% being protein coding. The other 98% of the genome, originally called "junk DNA", was proposed to have "little specificity and conveys little or no selective advantage to the organism." In a Nature paper published in the 1980 by Leslie Orgel and Francis Crick. Today, we know that large portions of the genome contain important regulatory regions that control gene expression in space and time.

1.1 From 1D view to the 3D view of the genome

1.1.1 The structure of the genome in 1D

In this section, I will explore of the composition of the 1D-genome by describing each component and its role in gene regulation.

1.1.1.1 Cis-regulatory elements

Cis-Regulatory Elements (CREs) are regions of non-coding DNA consisting of promoters, enhancers, silencers, and insulators, as shown in **Figure 8**. These elements are essential for the control of spatiotemporal gene expression. Each CRE is discussed in this section.

1.1.1.1.1 Promoters

The positioning of RNA polymerase at the correct initiation site is a critical process to ensure effective transcription initiation [45]. This process is regulated by the Transcription Start Site (TSS), a minimal stretch of DNA contained within a region of the genomic DNA known as the core promoter. The length of promoters can vary, but is generally between 100 and 1000bp. The sequence of a promoter actually depends on the gene being transcribed, the type of RNA polymerase (I, II or III) and the organism. In general the core promoter contains a sequence called the TATA box [46], located about 30bp upstream of the TSS. The TATA motif is recognized by the TATA box binding protein (TBP, subunit of TFIID, which is a key component of the transcription initiation complex that recruits the RNA polymerase to the promoter. The chromatin architecture of eukaryotic gene promoters is commonly characterized by the presence of a Nucleosome Free Region or Nucleosome

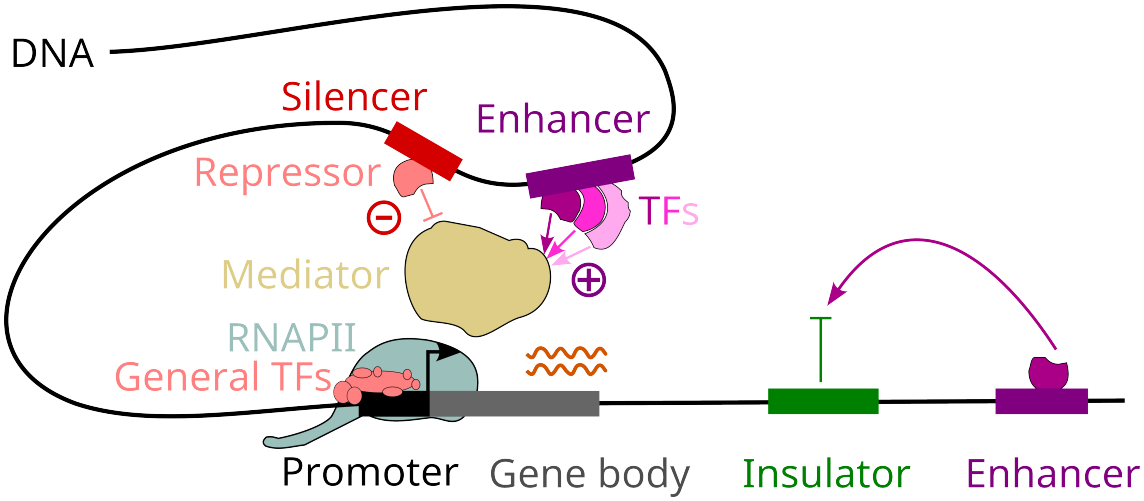


Figure 8: The different classes of cis and trans-regulatory elements - The mediator "complex" integrates information from both repressors and transcription factors (TFs) that bind to silencers and enhancers (indicated by red and purple arrows, respectively) to control the processivity of RNA polymerase II (RNAPII). Insulators act as barriers to prevent activation of the downstream enhancer.

Depleted Region (NFR and NDR respectively) in the core promoter region flanked by at least one of the histone H2A variants H2A.Z [47]]. Another important feature of nucleosome-flanking promoters is the presence of some post-translational modifications of the histone tails, such as the trimethylation of histone H3 lysine 4 (H3K4me3), the acetylation of histone H3 lysine 27 or the lysine 9 (H3K27ac and H3K9ac respectively), which are generally associated with active promoters, whereas the presence of trimethylation of histone H3 lysine 27 (H3K27me3) is generally associated with repressed promoters [48, 49]. Some promoters, called bivalent promoters are decorated by both H3K4me3 and H3K27me3 [50]. These so-called "bivalent domains" are thought to poise developmental gene expression, allowing for timely activation while maintaining repression in the absence of differentiation signals. In mammals, such as mice or humans, about half of the promoters have normal G+C content. The other half of the promoters are associated with CpG islands, which are regions devoid of methylation and have a G+C content higher than the genome average. These promoters are often associated with housekeeping genes and many tissue-specific genes [51]. The core promoter alone is sufficient to assemble the RNAPII machinery and drive a basal level of accurate transcription initiation. However, when regulatory sequences such as enhancers are present the resulting transcription is enhanced as reviewed in [52].

1.1.1.1.2 Enhancers

The term of enhancer was first introduced in 1981 to describe the positive effect on transcription of a 72bp repeat sequence located 200bp upstream of the TSS of early genes in simian virus 40 (SV40) [53]. Banerji and colleagues [54] showed that

this sequence, when linked or placed at a very large distance from a rabbit β-globin gene, can increase its expression by more than 200-fold regardless of its orientation. Next, in 1983, the identification of an enhancer associated with immunoglobulin (Ig) heavy chain genes represented a major advance in our understanding of gene regulation in eukaryotes [55]. Since these initial studies, enhancers have been defined as short stretches of DNA (between 50 and 1000bp in length) that contain multiple binding sites for transcription factors (TFs) (see 1.1.1.2.1) [56, 57] and that can activate transcription over large genomic distances regardless of their orientation. With the advent of Genome-Wide Association Studies (GWAS), which allow the identification of genetic variants associated with complex traits and diseases, it has become increasingly clear that the mutations within the non-coding genome are considered to be the major genetic cause of enhancer-associated human diseases (*i.e.* enhanceropathies). Understanding the function of CREs is a critical goal, and it requires addressing several fundamental questions about enhancers, such as how to identify them ?, how to test their functionally ?, and how to identify their target gene ?

How to identify enhancers ?

The definition of enhancers evolved with the development of DNA sequencing techniques that allow the identification of unique chromatin signatures associated with them. Specifically, enhancers can be found upstream or downstream of genes but also within introns of a gene. They are characterized by an accessible chromatin region (addressed by Micrococcal Nuclease (MNase) digestion [58, 59] or Assay of Transposase Accessible Chromatin sequencing (ATAC-seq) [60]), bound by proteins such as p300/CBP, Mediator or TFs and to carrying specific modifications to the adjacent nucleosomes, such as histone 3 hyperacetylation (e.g. H3K9Ac and H3K27Ac), H3K4 mono- and dimethylation (H3K4me1 and H3K4me2) **Figure 9**. Surprisingly, enhancers have been shown to be bound by RNAPII [61, 62] and to be actively transcribed. The product of this transcription is called "enhancer RNA" or "eRNA". Its production has been shown to be highly tissue specific [63, 64, 65, 66] and can differ between different activation states of cells [67]. More recently, a new class of enhancers called "super-enhancers" has been described. Super-enhancers are a distinct type of enhancers that are larger in size than classical enhancers (approximately 8kb in mESC). They are characterized by the presence of high binding of mediator protein (Med1) and transcription factors important for mESC biology, including Klf4, Esrrb and Prdm14. Super-enhancers are often found in close proximity to genes associated with pluripotency, such as Oct4, Sox2, and Nanog, specifically in mESC [68].

These characteristic chromatin features have enabled the identification of a large number of enhancers and super-enhancers in different cell types and tissues [52] (between 10,000 and 150,000 enhancers per cell type). However, these results must be interpreted with caution because few of these putative enhancers have been experimentally characterized *in vitro* and *in vivo*. Thus, it remains unclear how many enhancers are functional in a given cell-type.

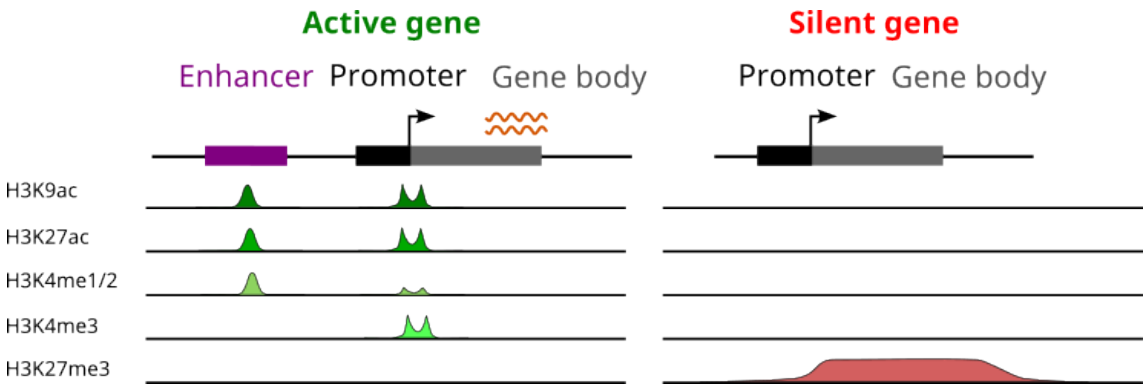


Figure 9: Distribution of histone modifications across transcriptional regulatory elements - Schematic representation of the chromatin signature across regulatory elements of an active (left) and a silent (right) gene.

How to test their activity ?

To assess the activity of enhancer elements, several *in vitro* and *in vitro* methods have been developed (FACS-seq [69], CRE-seq [70], MPRA [71], MPFD [72] and STARR-seq [73]). For example, STARR-seq (Self-Transcribing Active Regulatory Region Sequencing), is a technique that can be applied to both *Drosophila* and human cells *in vitro* to assess and quantify the activity of millions of DNA regions on transcriptional activation. In the STARR-seq method, the entire DNA genome is randomly segmented into small pieces, and adaptor sequences are ligated to either side of the size-selected fragments (see **Figure 10**). These sequences are PCR amplified and placed downstream of a reporter gene in screening vectors. Active regulatory regions are transcribed, and the strength of the enhancer is reflected by the abundance of cellular RNAs. For instance, Arnold and colleagues [73] employed this approach to test 96% of the non-repetitive *Drosophila* genome, resulting in the identification of 5499 regions with significant activity in *Drosophila melanogaster*. While this method allows the high-throughput testing of enhancers genome-wide, its current limitation is that the assay is specific to certain cell types because it relies on cultured cells and does not take into account the chromatin environment signatures of these elements.

How to identify their target gene ?

Despite advances in the field, identifying the gene(s) targeted by a given enhancer element remains complex. In some cases, enhancers remotely regulate their associated genes [75] and sometimes skip unaffected intermediate genes. It has been shown that about 60% of developmental enhancers in mice skip their immediate neighbor gene to regulate a more distal gene instead [76]. Another layer of complexity is added by the fact that enhancers can also regulate promoters located on different chromosomes [77, 78] (regulation in *trans*).

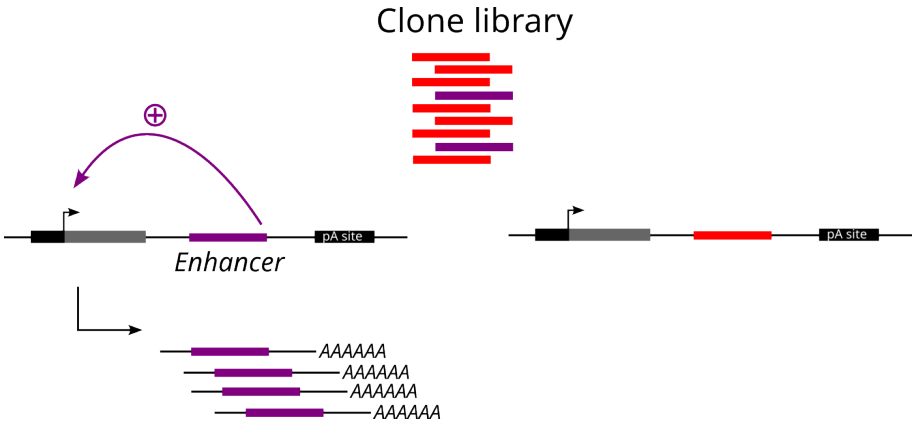


Figure 10: Schematic representation of the STARR-seq method - Schematic representation of the STARR-seq pipeline. A clone library is generated by fragmenting and cloning a source of DNA fragments. The library is transfected into cultured cells, and reporter transcripts are isolated from the total RNA and sequenced. (Left) DNA fragment mediating enhancer activity. (Right) DNA fragment not mediating enhancer activity. Figure adapted from [74].

Currently four models have been proposed to describe gene regulation by enhancers [79]. (i) the tracking model, (ii) the linking model, (iii) the relocation model and finally (iv) the looping model. According to the tracking model, a protein is bound to a distal enhancer and moves along the chromatin to reach the regulated promoter and stimulate transcription. The linking model is characterized by the formation of a protein bridge between the enhancer and the target promoter. In the relocation model, a gene is relocated to a nuclear subcompartment in the nucleus where transcription is favored. The looping model describes a direct chromatin interaction between the enhancer and the regulated promoter via the formation of a specific three-dimensional genome architecture (loop). This last model has been widely accepted by the scientific community, as it is the only one that can explain the *trans* effect of enhancers. This model is also supported by the development of Chromatin Conformation Capture and microscopy techniques (discussed in section 2) that allow for direct visualization of chromatin interactions between enhancers and promoters, providing strong evidence for the looping model.

1.1.1.1.3 Silencers

In addition to positive regulation by enhancers, transcription can also be negatively regulated by other CREs known as silencers. Numerous genomic and computational studies have focused on the prediction and characterization of enhancers. In contrast to enhancers, silencers are much less understood. Silencers are defined as CREs capable of repressing promoter activity in an orientation- and position-independent manner relative to the target promoter [80].

In prokaryotes, the cis-regulatory region lacO, shown in **Figure 6**, contains the

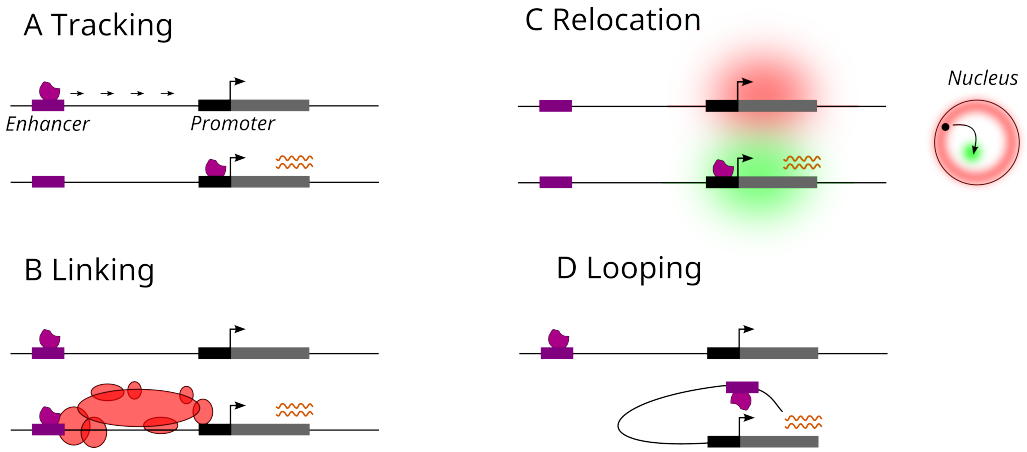


Figure 11: Models for the function of enhancers - (A) The tracking model, where a TF is loaded on the enhancer and tracks to the promoter. (B) The linking model, where a TFs loads onto the enhancer and a link between enhancer and promoter is made by protein bridge. (C) The translocation model, where a gene is translocated to a nuclear subcompartment where transcription is favored. (D) The looping model, where the enhancer comes into close proximity with the promoter it regulates. Figure adapted from [79].

binding site for the repressor protein (LacI) and has a function similar to that of a silencer region. Historically, the first silencer element, called HMRE, was discovered in yeast at the mating-type loci and was shown to be able to silence two non-mating-type promoters [81]. Similar regulatory elements were discovered in mammals upstream of the rat insulin 1 gene [82]. Several years later, in 1996 and 2004, pivotal research studies, identified a silencer located in the intron of both the mouse and the human CD4 genes. This silencer element was found to be involved in the regulation of lineage specificity and cell fate determination. Specifically, in the case of CD8+ T cells, this silencer has been shown to exert negative regulation by repressing the expression of CD4. While enhancers have a well-defined chromatin signature, the chromatin signature of silencers is poorly understood and remains an active area of research. Broadly speaking, silencers generally consist in regions of accessible chromatin that contain binding sites for regulatory factors known as transcriptional repressors. A recent study [83] found that silencers were significantly enriched for H4K20me1. However, this criteria may not be applicable to all silencers across different eukaryotes since the first silencer identified in *S. cerevisiae* lacks the PR-SET7 homolog, the protein known to deposit H4K20me1. Silencers are often associated with Polycomb repressive complexes (PRC1 and PRC2) and thus with the characteristic H3K27me3 histone mark deposited by the PRC2 complex. The study of chromatin accessible regions in H3K27me3 domains led to the identification of regions correlated with gene repression [84]. However, silencers appear to be a broad class of elements, and Polycomb-associated silencers may represent a subclass of silencing elements. Similar to enhancer studies, high-throughput reporter assays [85] were developed to attempt to identify tissue-specific silencer elements in the *Drosophila* genome (silencer-FACS-Seq). A subset of silencers was found to

be enriched for the presence of the transcriptional repressor Snail, a well-studied repressor of non-mesodermal genes in the developing mesoderm. The role of Snail was confirmed by genetic modification of the Snail motif. Such mutation was shown to reduce silencer activity [85]. More generally, the majority of silencers that were identified in *Drosophila* have also been identified as active enhancers in other cellular contexts. Thus, the traditional classification of *cis*-regulatory elements as "silencers" or "enhancers" appears to be an oversimplification.

1.1.1.1.4 Insulators

Regulatory elements such as enhancers and silencers are widely distributed throughout the genome and can affect gene expression over a long distances. To protect genes from being regulated by inappropriate enhancers or silencers, the process of chromatin insulation involves the formation of independent chromatin domains.

Early studies of the polytene chromosomes of *Drosophila* salivary glands reveal a stereotyped organization of the chromosome into independent domains, with compacted, DNA-rich "bands" alternating with extended, DNA-poor "interband" regions (reviewed in [86]) shown in **Figure 12**. This band/interband pattern of the polytene chromosome appears to also reflect the general features of chromatin organization shared by non-polytene chromosomes. The discovery of insulators provided a critical insight into the understanding of this specific patterned structure.

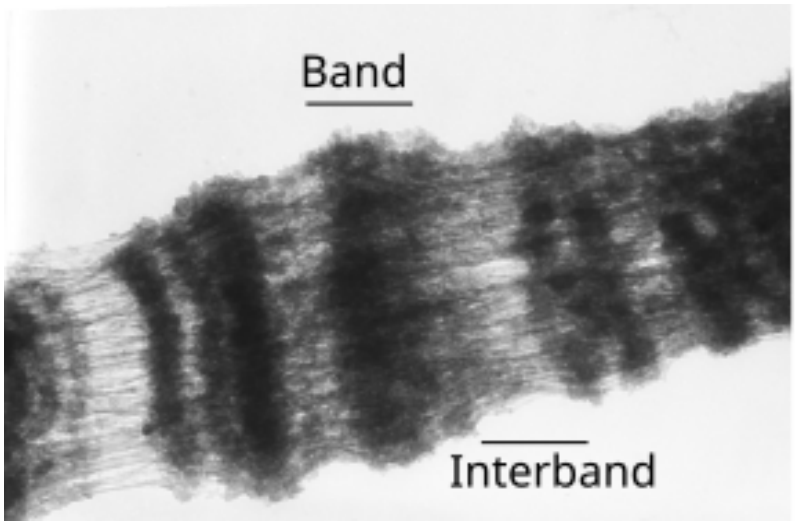


Figure 12: Polytene chromosome - High-voltage transmission electron microscope image showing the series of bands and interbands pattern of a polytene chromosome of *Drosophila melanogaster*. Figure adapted from [87].

By definition an insulator is a class of regulatory elements that establishes independent domains of transcriptional activity within the eukaryotic genome. Insulators are characterized by two distinct properties. First, insulators can protect gene

expression from the spread of active or repressive chromatin marks from neighboring regions of the genome [88]. This function is termed "barrier activity". Second, insulators can prevent enhancers and silencers from communicating with a promoter in a position-dependent manner [89]. This function is called the "enhancer blocker activity" (**Figure 13**). To carry out their functions, insulators are thought to be associated with the establishment of higher order chromatin structures [90].

Proteins associated with insulators (Insulators binding proteins (IBPs)) were first discovered in *Drosophila* using genetic screens. These factors include Su(Hw), BEAF-32, mod(mdg4), CP190, dCTCF, GAF, Zw5, Pita, ZIPIC, Z4, Chriz/Chromator, Fs(1)h [18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28]. In contrast to CTCF, which has been found in a wide range of eukaryotic organisms, from yeast to humans, all other *Drosophila melanogaster* insulator proteins are not evolutionarily conserved and are restricted to arthropods [17]. Using immunostaining on polytene chromosomes, insulator proteins were found to localize to both polytene interbands and interband boundaries. [20, 91]. Furthermore, the structural organization of polytene chromosomes was found to be disrupted when certain insulator proteins were mutated [92]. In addition to insulator-mediated structural patterning of *Drosophila* chromosomes, genome-wide analyses of the chromosomal distribution of co-expressed genes in *Drosophila* [93] and humans [94] have revealed that approximately 20 to 30% of co-regulated genes are clustered on chromosomes. This partitioning of the genome into small regulatory regions has also been proposed to be regulated by insulators.

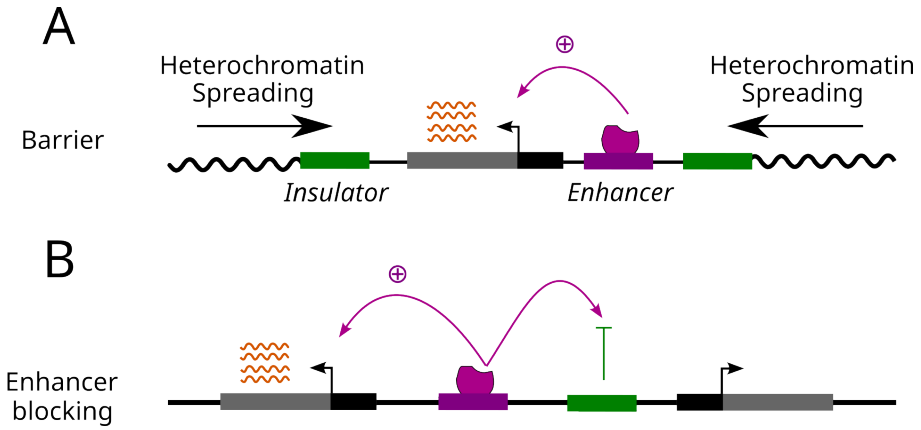


Figure 13: Insulator properties - Insulators possess two main properties. (A) They can prevent heterochromatin from spreading (barrier). (B) They can block enhancer-promoter communication (enhancer blocker).

Barrier activity

The best characterized barrier elements are located at the silent mating type loci in the yeast *Saccharomyces cerevisiae*. Haploid budding yeasts are characterized by the presence of two silent copies of the mating-type specific genes MATa (a1 and a2) and MATα (α1 α2) at the HMR and HML loci, respectively. The MATa and

MATα genes are flanked by silencer elements named *E* and *I*. The repression of these genes is ensured by the recruitment of the Sir 2/3/4 protein complex, which leads to the heterochromatization of the domain and the silencing of the HM loci between the *E* and *I* silencing elements. Depletion of the HM flanking regions (*E* and *I*) leads to the expansion of the silent domain [95]. This suggests that the heterochromatin barrier elements *E* and *I* are required to block the further spreading of silencer proteins to neighboring regions. Barrier activity has also been identified at the telomeres in budding yeast. Telomeric DNA is flanked by two subtelomeric repetitive sequences, the X and Y' elements. This sequence known as STAR for SubTelomeric Anti-Silencing Region [96]) consists of multiple binding sites for the Tbf1 protein. Tethering of Tbf1 via a GAL4 DNA-binding domain was shown to be sufficient to confer barrier activity [97]. Analogous to the *E*, *I*, and STAR elements studied in yeast, mammals also have barrier-like elements that function in a similar manner. One example is a 16 kb region of silent chromatin : the HS4 insulator sequence in chickens which separates the folate receptor gene from the β-globin genes. This sequence has been shown to act as a barrier to prevent the repressive chromatin originating from the folate receptor gene from spreading and silencing the β-globin genes [98]. These results provide evidence that barrier-like properties of insulators are common among eukaryotes.

In addition to acting as a "barriers" that prevent inappropriate spreading of active or repressive histone marks to undesired locations, insulators also regulate proper communication between enhancers, silencers, and promoters through their "enhancer blocker" activity.

Enhancer blocker activity

Experiments performed in *Drosophila* lead to the identification of three famous insulator elements that act as enhancer blockers. (i) The *scs*, (ii) the *scs'* [99] and the *gypsy* insulator [100]. Cai and Levine [101] conducted an analysis of the effect of *scs* positioning on the communication between the promoter and two enhancer elements of the even skipped (*eve*) gene. During the development of *Drosophila* embryos, the spatial expression of the *eve* gene is regulated by these two enhancers, which activate and position two expression stripes (stripes 2 and 3) along the anterior-posterior axis. The study showed that when the stripe 2 enhancer was placed upstream of the 900bp fragment of the *scs* insulator, it effectively blocked stripe 2 expression, while stripe 3 expression remained unaffected. Conversely, when the order of the stripe 2 and stripe 3 enhancers was reversed, opposite results were observed. Similar observations were made with *gypsy* in the same assay. The enhancer-blocking effect of *scs* and the gypsy insulator is not specific to a single enhancer, but was shown to exert its enhancer-blocking activity on more than 20 enhancers *in vivo*. This observation provided evidence that insulator elements can block the communication between enhancers and promoters.

The properties of insulators depend on the binding of Architectural Proteins (APs) (or insulator-binding proteins). For example, *scs* and *scs'* are bound by ZW5

[23] and BEAF32 [20], respectively, while *gypsy* is bound by Su(Hw), which recruits the Mod(mdg4) protein. Although insulators are highly efficient enhancer blockers, they do not form inacessible barrier. Under certain conditions, enhancers can bypass a *gypsy* insulator to activate transcription [102]. This effect is called "insulator bypass". This insulator bypass effect has been observed when two *gypsy* insulators are placed between an enhancer and a promoter [103]. It has been proposed that neutralization of the *gypsy* insulator function results in the formation of a chromatin loop between the two adjacent *gypsy* insulators suggesting a potential role for insulators in the organization of the three-dimensional genome (see 1.1.2).

1.1.1.2 Trans-regulatory elements

In contrast to CREs, Trans-Regulatory Elements (TREs) are essentially DNA-binding proteins that can bind to CREs by recognizing specific DNA-binding motifs.

1.1.1.2.1 The transcription factors

Transcription Factors (TFs) are the key molecules for transcriptional regulation. Since the discovery of enhancers about 30 years ago, scientists have been intensely studying how TFs can read the genome to accurately identify and bind their appropriate sequences at specific developmental time points to regulate correct spatiotemporal gene expression. TFs are a group of proteins that have the ability to bind to DNA thanks to their DNA-Binding Domain (DBD) [104]. TFs use various types of DNA-binding domains to recognize their target DNA motif, including homeodomain (HD), helix-turn-helix (HTH), and high-mobility group box (HMG)(reviewed [105]). Once bound to specific cis-regulatory sequences such as promoters, enhancers or silencers [106], TFs can exert transcriptional control by activating or repressing RNA polymerase II (Pol II) through their transactivation domain [107]. Transcription factors can be broadly divided into two main categories: General Transcription Factors (GTFs) and Specific Transcription Factors.

The general transcription factors

In vitro studies have shown that in eukaryotes, the recruitment of the RNAPII to the core promoter requires the assembly of the pre-initiation complex(PIC). The formation of the PIC is a critical step in transcriptional regulation, and it involves the assembly of a group of proteins known as the GTFs [108, 109]. The assembly process begins with the binding of TFIID to its TBP subunit, to the TATA box of the promoter. TFIIA helps in facilitating the recruitment of TFIID to the promoter, and once TFIID and TFIIA are bound to the core promoter, TFIIB is recruited to TFIID. This step allows for the recruitment of RNAPII and TFIIF to the complex [110]. Finally, the assembly of the PIC is completed with the addition of TFIIE and TFIIH, which stabilize the complex [111] (**Figure 14**). Historically, the PIC has been associated with GTFs and the RNAPII, but recent studies have shown that an

another protein complex called Mediator is a key component of the PIC (review in [112]) discussed in section 1.1.1.2.2).

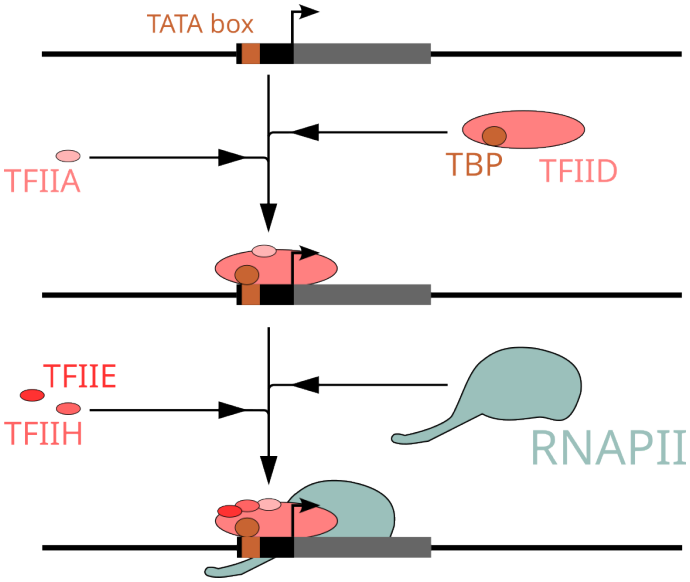


Figure 14: Assembly of the preinitiation complex - Illustration of the assembly of the preinitiation complex by the general transcription factors.

The specific transcription factors

In contrast to GTF which bind to gene promoters, Specific Transcription Factors (STFs) are factors that bind to either enhancers or silencers and regulate the transcriptional process. It is estimated that the human genome contains approximately 1557 TFs [113] and that a typical human protein-coding gene is regulated by multiple enhancers, each of which being bound by a specific combination of TFs. However, studies have shown that the majority of TFs are unable to bind to highly compacted chromatin [114]. Therefore, the action of specific factors, such as remodeling factors, is required to unwrap the DNA around nucleosomes and make it accessible for other transcription factors. These specialized transcription factors (which represent only a small subset of transcription factors) are known as pioneer factors. This specific group of TFs is described in the section 1.1.2.1).

1.1.1.2.2 The Mediator complex

The mediator, or mediator of RNA polymerase II transcription, is a transcriptional coactivator found in all eukaryotes. It is an evolutionarily conserved multisubunit complex essential for transcriptional regulation. The mediator complex consists in 25 subunits in *Saccharomyces cerevisiae* and can consist in up to 30 subunits in humans. These subunits are organized into four distinct modules: the

head, the middle, the tail, and the kinase module. The mediator does not appear to bind directly to DNA, but instead, it interacts with proteins that bind to the CREs such as TFs, RNAPII [115], or transcription elongation factors [116]. The mediator is thought to regulate gene expression through the "bridge model" allowing the conversion of biological input (binding of TFs to enhancers / silencers) into physiological response (gene expression) by regulating RNAPII initiation and elongation (**Figure 8**). According to this model, the tail module of the Mediator interacts with the activation domains of TFs bound at enhancers and recruit the complex to DNA. The head and the middle modules interact with the RNA II and GTFs bound at promoters. Finally the kinase module, which contains cyclin-dependent kinase 8 (CDK8), positively and negatively regulates transcription through RNAPII [117]. A large number of studies have reported that individual deletion of 10 of the 25 Mediator subunits in yeast leads to a global dysregulation of all protein-coding genes [118], resulting in lethality. In mammals, various knockouts of Mediator subunits result in embryonic lethality consistent with Mediator's extensive role in transcriptional activation [119].

1.1.2 The structure of the genome in 3D

In the previous section I discussed DNA as a linear sequence of nucleotides containing regulatory regions that control gene expression. However, it is becoming increasingly clear that chromosomes adopt specific conformations that allow precise regulation of gene expression. In this section I will talk about the fascinating world of chromatin folding, exploring how DNA is packed into the nucleus and how this folding is critical for gene regulation.

1.1.2.1 The chromatin fiber

In eukaryotic cells, DNA is associated with biomolecules such as proteins and RNA. One of the most important protein components is histones, which make up more than half of the total proteins associated with DNA. Histones are a group of positively charged proteins consisting of H2A, H2B, H3 and H4 [120]. These proteins assemble in pairs to form histone octamers, which then interact with the negatively charged phosphate backbone of DNA to form compact structures called the nucleosomes [121]. The nucleosome is the basic unit of the chromatin fiber in which approximately 145-147bp of DNA are wrapped around the histone octamer in 1.65 superhelical turns allowing for 7-fold compaction of the genomic DNA [122, 123]. The chromatin fiber consists of a multitude of nucleosomes, each connected to its neighboring nucleosomes by a linker DNA, creating the 10 nm fiber that can be observed as "beads-on-a-string" under an electron microscope [124]. The next level of compaction is provided by the addition of H1 to the nucleosome, allowing an additional 20 base pairs to be wrapped, resulting in two complete turns around the histone octamer to form the 30 nm chromatin fiber. Nucleosomes within the fiber can engage in close interactions with their neighbors, forming nucleosome clusters of varying sizes called "clutches" [125]. Subsequent levels of organization involve the addition of scaffolding proteins that wind the 30nm fiber to form higher order structures (**Figure 15**).

Biological processes such as transcription, replication, and DNA repair require access to the DNA template by polymerases. However, the highly compact nature of the chromatin fiber presents a significant barrier to these enzymes. To overcome this barrier, the regulation of chromatin structure is necessary. There are three major reversible mechanisms to overcome this compaction : (i) Pioneer transcription factors. (ii) ATP-dependent remodeling complexes. (iii) Covalent histone-modifying complexes.

Pioneer transcription factors

Pioneer factors are a specific class of transcription factors that have the ability to bind to their target sequences even when the DNA is highly compacted. In doing so, they increase the local accessibility and promote the subsequent binding of additional transcription factors. The concept of pioneer factor was first introduced

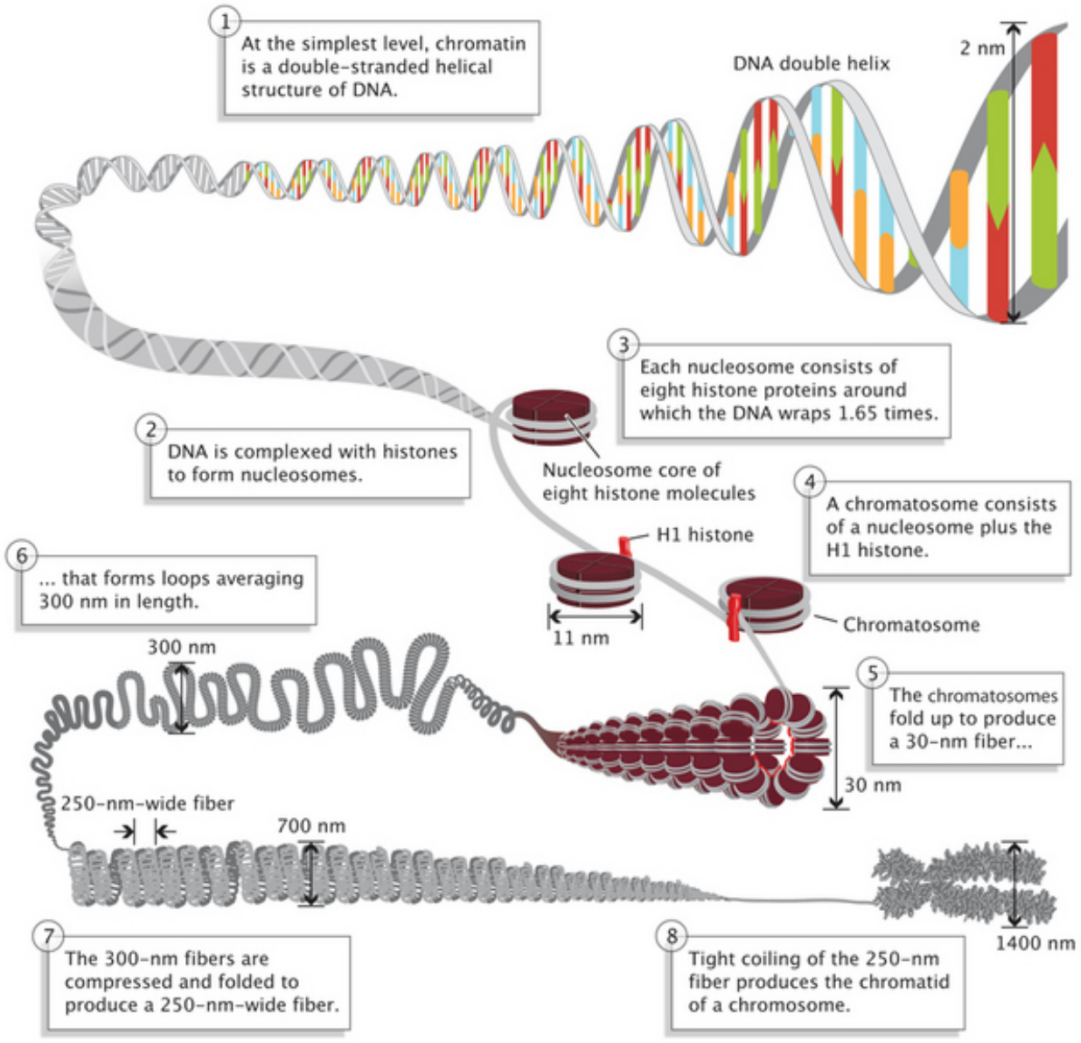


Figure 15: Hierarchical folding of the genome - The DNA molecule is coiled into a chromatin fiber with different levels of organization. First, DNA is wrapped around histone octamers to form nucleosomes. Next, the chromatin fiber is coiled into a 30 nm structure, which is looped and coiled to form a 250 nm structure, which is finally condensed to form the chromosome. Figure adapted from [126].

in 2002 when it was shown that FOXA but not others (NF1 and C/EBP), can bind the nucleosome of compacted chromatin and locally open the domain *in vitro* [58]. Thus, the nucleosome binding capacity of TFs represents a new functional class of TFs called pioneer factors. However, the binding of a pioneer TF by itself does not imply an immediate function (activation of transcription), but rather the accessibility of chromatin is required to allow the recruitment of other TFs to modulate enhancer activity. This was demonstrated by a footprinting study showing that the pioneer activity of FOXA1 is also required for the binding of GATA-4 to the albumin enhancer in mouse embryonic endoderm [127]. Zelda, a pioneer transcription factor in *Drosophila*, provides another example of how pioneer activity can facilitate the

binding of other transcription factors. Studies have shown that Zelda can locally deplete nucleosomes, allowing other transcription factors, such as Dorsal, to bind to its DNA motif [128].

ATP-dependent remodeling complexes

ATP-dependent chromatin remodeling plays a key role in regulating the dynamics of chromatin structure. By utilizing the energy derived from ATP hydrolysis, remodeling complexes are able to either disrupt the contacts between histones and DNA or induce conformational changes within nucleosomes [129], thereby regulating DNA unpacking to facilitate the access of nuclear machineries. These remodeling factors fall into four categories. First, the SWI/SNF (SWItching defective/Sucrose NonFermenting) family consists of 8 to 14 subunits that play a critical role in nucleosome rearrangement, which can involve either ejection or sliding of the nucleosome [130]. Second, the ISWI (Imitation SWItch) family which is composed of 2 to 4 subunits, plays a role in optimizing nucleosome spacing to facilitate chromatin assembly and transcriptional repression [131]. Third, the CHD (Chromodomain, Helicase, DNA binding) family is composed of 1 to 10 subunits and promotes nucleosome sliding or removal to regulate transcriptional activation [132]. Finally, INO80 (INOsitol requiring 80) contains more than 10 subunits involved in transcriptional regulation and DNA repair. SWR1, a member of the INO80 family, has the unique ability to remove the canonical H2A-H2B dimers and replace them with H2A.Z-H2B dimers during double-strand break repair [133].

Covalent histone-modifying complexes

The core histone consists of two distinct domains: a structured domain and an unstructured N- and C-terminal domain of 25-40 residues called "tails". These tails protrude into the space surrounding the nucleosomes and provide sites for a variety of Post-Translational Modifications (PTMs), including acetylation, phosphorylation and methylation, which are catalyzed by "writers" enzymes (**Figure 16**). It is becoming increasingly clear that such modifications on the histone tails can affect nucleosome-nucleosome interactions or impair the affinity between the histone octamer and nucleosomal DNA [134]. In addition, certain PTM sites can act as platforms to recruit "reader" proteins that can in turn change chromatin structure [135, 136]. The large number of possible histone tail modifications as well as the numerous protein readers that can interact with them provides scope for the tight control of chromatin structure. Epigenomic studies have shown that the distribution of histone PTMs can serve as a signature for the chromatin state within a given genomic region. Eukaryotic genomes can be broadly classified into two distinct chromatin environments : Euchromatin and Heterochromatin. Euchromatin typically associated with actively transcribed genes, is characterized by high levels of histone acetylation such as H3K9ac and H4K16ac at promoter regions, while enhancers exhibit enrichment in H3K27ac and other modifications such as H3K4me1. Di- and tri-methylations of histone H3 at K4 (H3K4me2-3) are commonly found in close proximity to active promoters, while H3K36me3 reaches its highest levels at the 3'

end of active genes. In contrast, heterochromatin is defined by the presence of compact, tightly packed chromatin and is classified into two categories: constitutive and facultative heterochromatin. On the other hand, heterochromatin is characterized by dense, tightly packed chromatin and appears as darkly stained regions. There are two types of heterochromatin : the constitutive and facultative heterochromatin. Constitutive heterochromatin is characterized by the presence of H3K9me3, which is responsible for the formation of stable heterochromatin at pericentromeric regions and at telomeres. Facultative heterochromatin, in contrast, is characterized by the presence of Polycomb repressive marks such as H3K27me3 deposited by the Polycomb repressive complex 2 (PRC2) and H3K119Ub by PRC1.

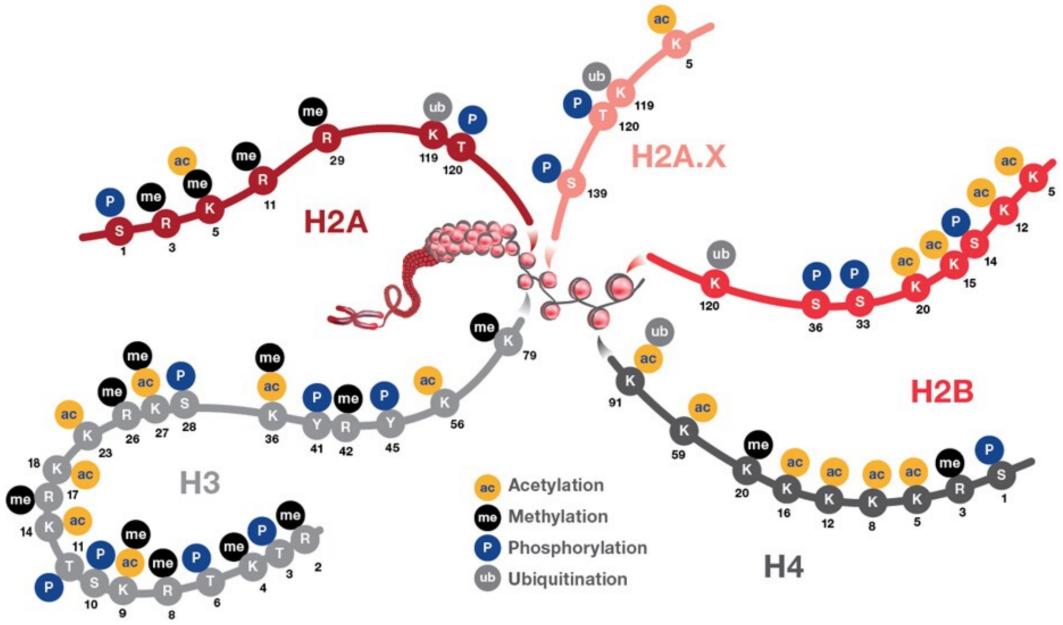


Figure 16: Sites of histone post-translational modifications - Sites of the most common histone post-translational modifications including methylation, acetylation, phosphorylation, and ubiquitination. Figure adapted from Thermofisher.

1.1.2.2 Chromosome territories

In the early stages of chromosome research, Carl Rabl and Theodor Boveri observed that interphase chromosomes were located in distinct regions within the nucleus later called "Chromosome Territories" (CTs) (**Figure 17**). The confirmation of the existence of CTs came to light through the development of the fluorescence in situ hybridization (FISH) technique (discussed in section 2), which allows for the direct visualization of a given chromosome in the intact nucleus [137]. The distinct nuclear localization of each chromosome raises a fundamental question : Are CTs randomly arranged in the nucleus ?. Advances in the flied clearly demonstrate that

the chromosomes are arranged in a non-random way. For instance, chromosome 12, 14, and 15 have been shown to form a triplet cluster in mouse lymphocytes [138]. Furthermore, CTs are semi-conserved during cell division, and their locations in the daughter cell closely resemble those in the parent cell. The development of chromosome conformation capture techniques (discussed in section 2) have revealed that the frequency of contacts within chromosomes (intra-chromosome) is significantly higher than contacts between different chromosomes (inter-chromosome). This finding, reported by Lieberman Aiden et al. in 2009 [1], provides evidence that chromosomes form distinct subcompartments within the nucleus. The radial positioning of CTs in the nuclear volume of human lymphocytes suggests a relationship between the gene density of a chromosomal region and its spatial localization within the nucleus [139]. Specifically, chromosomes with high gene density are more likely to be located closer to the center of the nucleus, while chromosomes with lower gene density are more likely to be located closer to the nuclear lamina. This supports the idea that the spatial arrangement of chromosomes is influenced by a range of nuclear processes, including transcription, DNA replication, and repair.

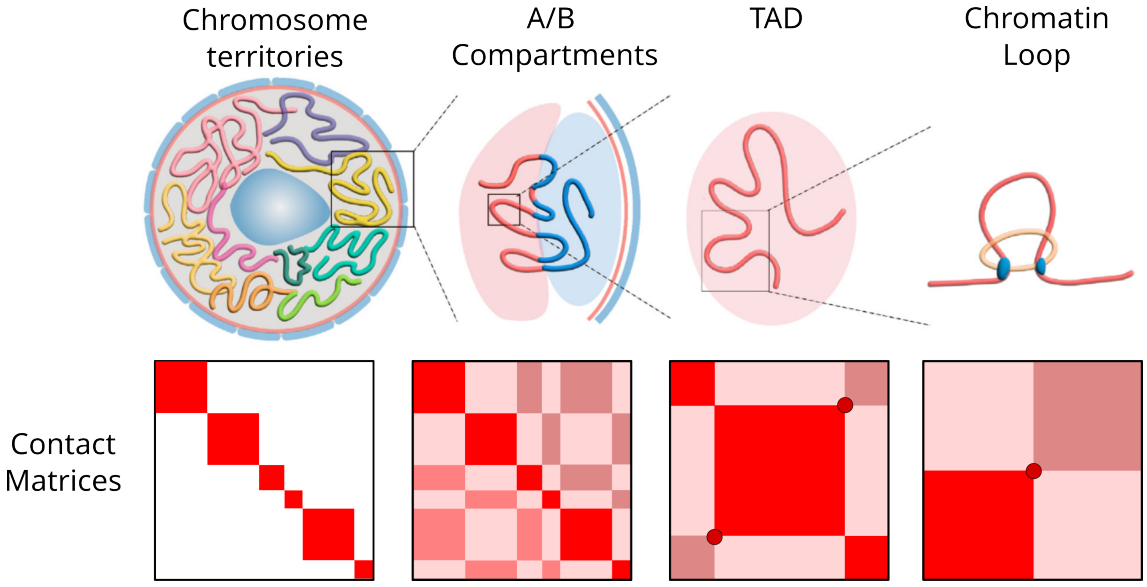


Figure 17: Different hierarchies of the 3D genome organization - (Top) Schematic illustrating the different scales of genome folding. (Bottom) Different interaction patterns in the contact matrix. Figure adapted from [140].

1.1.2.3 Chromosome compartments

Within each CT, chromosomes are partitioned into two distinct compartments (**Figure 17**). These two compartment were originally defined by Emil Heitz in 1928, who demonstrated through cytological staining that certain regions of the chromosome appear more densely stained than others. These regions were later identified as heterochromatin and euchromatin, respectively. In conventional nuclei, as opposed to inverted nuclei in murine rod cells [141], it have been observed

that the euchromatin occupies preferentially the inner part of the nucleus and the heterochromatin occupies the nuclear periphery [141, 142, 143]. This observation on chromosomal compartments was later supported by early Hi-C studies, which showed that the entire genome can be divided into two distinct spatial compartments, called A and B. These compartments were found to have preferential interactions with the same compartment-associated regions (A with A and B with B). Further analysis showed that A/B compartimentalization correlated with either accessible and transcriptionally active regions (A) or inaccessible and transcriptionally inactive regions (B). Computational approaches have further strengthened the notion of a relationship between the 1D epigenomic domains (epigenetic states) and their partitioning into discrete chromatin compartments [144, 145]. Thus histones of euchromatin are marked by H3K4me3, H3K27ac, H4K8ac, and H4K16ac, whereas those of heterochromatin are marked by H3K9me3, H3K27me3 or H4K20me3 [146]. High-resolution microscopy techniques (dSTORM) have revealed the presence of nanometer-sized epigenetic domains at the single-cell level, including both active (H3K4me3) and repressive (H3K27me3) compartments [147]. Furthermore, Hi-C data at higher resolution suggested that these two primary compartments (A and B) can be subdivided into five distinct subcompartments, each with specific patterns of histone modifications and replication characteristics [148]. The A compartment can be segregated into two distinct regions, A1 and A2, which differ in their replication timing. A1 completes replication at the beginning of S phase, while A2 continues to replicate until the middle of S phase. The B compartment can be further subdivided into three distinct subcompartments, namely B1, B2, and B3. B1 is marked by the H3K27me3 modification but lacks H3K36me3, and its replication occurs in mid-S phase. Conversely, B2 and B3 lack both H3K27me3 and H3K36me3 modifications, and their replication does not occur until the end of the S phase. In addition, chromosome compartments were shown to be a highly dynamic structure that undergoes significant A/B compartment switching during the process of differentiation. In fact, in at least one of the four hESC-derived lineages (Mesendoderm (ME), Mesenchymal Stem Cells (MSC), Neural Progenitor Cells (NPC), and Trophoblast-Like Cells (TB)) used in the study [149], it was reported that, more than a third (36%) of the genome switches compartments. Interestingly, the genes that switch from A to B show lower expression levels, while those that switch from B to A show an increase in expression level. Finally, artificially tethering of a loci to the nuclear periphery (lamina) has been shown to favor gene silencing in yeast and in mammalian cells [150, 151, 152]. Moreover unusual location of gene from a euchromatic region to a position near a heterochromatin region can lead to drastic downregulation through a process called Position-effect variegation (PEV) [153, 154]. This highlights the the crucial role of the nuclear microenvironment or compartment in regulating gene activity.

1.1.2.4 Topologically Associating Domains

Advances in the resolution of 3C methods have revolutionized the field of genomics, enabling the detection of a previously unobservable chromatin structure. One of the most exciting discoveries over the last decade has been the characteri-

zation of Topologically Associating Domains (TADs) [2, 3, 4] (**Figure 18**). TADs are large genomic regions that exhibit a unique 3D organization, with a twofold increase in contact probability with regions situated inside the TAD compared to regions located outside of it. TADs appear on a chromatin contact map as large squares of enhanced contact frequency along the diagonal. They have been shown to be present in several species, including the Drosophila [4], mouse [3], zebrafish [155], plant [156] and human [2] genomes. TADs-like domains have also been identified in bacteria, they are called Chromosomal Interacting Domains (CIDs) [157, 158].

An important feature of TADs is their insulating boundaries, which play a crucial role in maintaining the self-associating property of TADs. These boundaries are characterized by the binding of a number of proteins, including CCCTC-binding factor (CTCF), proteins in the structural maintenance of chromosomes (SMC) complex (cohesin, condensin), RNA polymerase II, and the presence of housekeeping genes. Loss of CTCF, cohesin, or the cohesin loading factor Nipbl results in the disruption of TAD. Conversely, the loss of cohesin release factor, Wapl, reinforces the boundaries of TADs. [159, 160]. The median size of TADs varies between species. For instance, in mice, the median size of TADs is approximately 880kb [161], whereas in Drosophila, TADs have a smaller median size of approximately 70kb [162]. However, it's important to note that the definition of TADs depends on both the data resolution and the computational algorithms used to identify them [163]. As a result, the size of TADs and their annotations must be considered with caution. It's believed that TADs represent the functional unit of chromosomal organization. TADs play a critical role in limiting the scope of interactions between CREs and their target genes, which are typically located within the same TAD [10]. Thus, the disruption of TAD structures by altering their boundaries can lead to ectopic contacts between CREs and gene promoters, ultimately causing gene silencing as observed in certain developmental defects and cancers [164, 165]. Surprisingly, the majority of TAD boundaries appear to be conserved during the differentiation, with more than 80% of boundaries being present in both mESCs and cortex [2]. In addition, a significant fraction of these boundaries are conserved across evolution, with 53.8% of human boundaries found in mouse, and 75.9% of mouse boundaries found in human [2]. While TADs exhibit a high degree of conservation across different cell types and species, the organization of genomic material at a sub-Mb scale within TADs may play a critical role in regulating cellular states. This hypothesis was supported by the discovery of sub-megabase-scale chromatin domains, known as sub-TADs, that are found to be nested within TADs in mammalian Hi-C maps. Initially, only a limited number of nested sub-TADs were detected in low-resolution Hi-C data. However, with the development of advanced technologies that allow the generation of ultra-high-resolution (1-4 kb) architectural maps, the detection of sub-TADs across the genome has become more feasible across the genome. Sub-TADs are similar to TADs in that they are delineated by internal boundaries. However, the insulation strength of these boundaries is weaker than that of TADs, and they are more likely to exhibit cell type-specific folding properties. In addition, DNA-FISH method using super-resolution microscopy (3D-SIM) reveals that TADs are organized into discrete chromatin nanodomains (CNDs) of varying size and number

that are present in single cells [166].

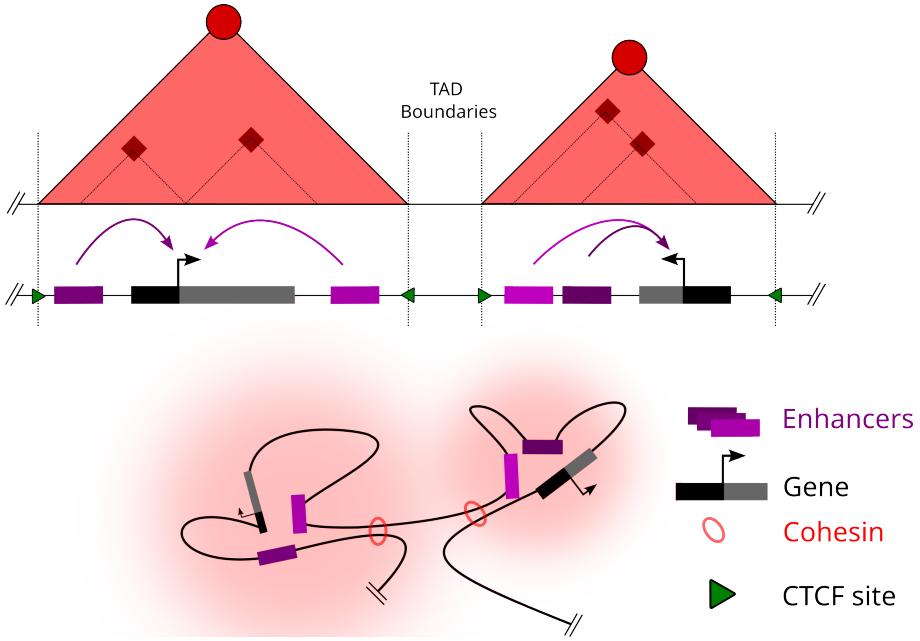


Figure 18: Topological associating domains (TADs) as functional units - TADs are visualized as triangular blocks along the diagonal of a contact matrix. TADs play a critical role in restricting interactions between *cis*-regulatory elements and genes within the same TAD. The boundaries of TADs are marked by convergent CTCF and cohesin binding sites that separate one TAD from another.

1.1.2.5 Chromatin loops

Chromatin loops represent a more intricate level of genome organization, characterized by radially symmetric peaks in contact matrices resulting from the physical convergence of distant genomic loci within the same chromosome (**Figure 17**. The size of the chromatin loops varies widely, ranging from a few kilobases to several megabases in both the human and mouse genomes [167, 168]. At an intermediate scale, numerous TAD boundaries bounded by CTCF and cohesin in mammals exhibit a "corner peak" pattern in contact matrices [2]. This particular loop between adjacent boundaries is thought to constrain the formation of loops between regulatory elements found in single TAD. These include enhancer-promoter loops, which represent a subset of chromatin interactions thought to regulate transcription. A striking illustration of the role of enhancer-promoter interaction in transcriptional regulation has been demonstrated at the β-globin locus. Erythroid cells (where theβ-globin gene is active) exhibit extensive chromatin looping between the locus control region (LCR) and the β-globin cluster compared to cells from other lineages [169]. Complementary studies [170] show that the recruitment of Ldb1 to the β-globin promoter in GATA1 null cells completely restores chromatin looping between the β-globin and LCR regions, leading to the rescue of transcription initiation and partial

rescue of transcription elongation. These findings highlight a potentially informative role for enhancer-promoter interaction in the regulation of gene expression. While the looping model seems best suited to explain the regulation of promoter activity by distal enhancers, the relationship between enhancer-promoter proximity and transcriptional activation may not be straightforward [171]. To date, three theories have been proposed regarding enhancer-promoter loops: (i) these loops are cell-type specific, (ii) they remain stable across tissues, and (iii) they are disrupted upon gene activation. (discussed in section 1.2.2.2).

Besides chromatin loops, chromatin architectural "stripes" (also termed "lines") were characterized and refer to the observation that one anchor region interacts with entire domains at high frequency [15, 172, 173, 174]. Recent discoveries in chromosome organization have also revealed the existence of "fountains" that form during genome activation at active enhancers. These fountains appear as a perpendicular line to the main diagonal and their formation has been shown to depend on the pioneer transcription factors Pou5f3, Sox19b and Nanog [175, 76].

1.2 Understanding the 3D Genome

The three-dimensional folding of chromatin in the genome is critical for numerous biological processes, including transcription, replication, and DNA repair (reviewed in [176]). However, the mechanisms underlying the establishment and regulation of this architecture to maintain proper cell identity is an area of extensive study. In this section, I will review recent breakthroughs in the field of chromatin folding, with a particular focus on the different mechanisms involved. I will explore how chromatin is folded during early stages of embryonic development, how this scaffold is adapted in differentiated tissues to regulate changes in the expression program, and how disruption of chromatin folding can impair normal cell function and lead to diseases such as cancer.

1.2.1 Mechanisms of the 3D genome folding

In the field of chromatin, a key question is how the architectural proteins bring linearly distant loci together to form a chromatin loop ?

1.2.1.1 The Role of CTCF/cohesin in Mammalian

The complex three-dimensional architecture of the genome is intricately regulated by a network of chromatin-associated proteins. Among these, CTCF has emerged as a key player in mediating long-range chromatin interactions as the majority of TAD boundaries and chromatin loop anchors in mammals are bound by CTCF. The importance of CTCF in maintaining the 3D genome structure has been

highlighted by experiments using the degron system to deplete CTCF in mouse embryonic stem cells (mESCs). Interestingly, when CTCF is depleted the 3D genome structure is drastically affected resulting in a loss of TAD insulation and a disruption of loops normally anchored by CTCF [177]. However, CTCF alone is not sufficient to promote the formation of TADs and long-range chromatin loops suggesting that other complexes may also play a role in the regulation of chromatin folding. Another protein complex that plays a significant role in loop formation is the cohesin complex. This is supported by the observation that a majority of cohesin binding sites in the human genome co-localize with CTCF sites [178]. In addition, cohesin depletion was shown to disrupt CTCF-mediated looping and reduce interactions within TADs. These findings highlight the dual role of the CTCF/cohesin complex in the formation of both TADs and long-range chromatin loops. However, the exact molecular details by which CTCF and cohesin regulate 3D genome architecture remains to be elucidated. The loop extrusion model, which proposes that the motor function of cohesin drives chromatin extrusion while CTCF acts as a barrier has emerged as a plausible explanation for loop formation [173].

Initially, the loop extrusion hypothesis was formulated to explain the mechanism by which condensin compacts mitotic chromosomes, by translocating and progressively growing DNA loops [179]. Adapted to the CTCF/cohesin complex, this model provides a robust explanation for the organization of chromosomes in interphase cells of higher eukaryotes [173]. The model is based on the idea that protein motors, in this case cohesin, can extrude DNA loops. The molecular motor loads onto chromatin in a sequence-independent manner and begins extruding loops bidirectionally until it encounters a properly oriented CTCF site bound by CTCF (*The CTCF convergence rule*, reviewed here [180]). At these sites, CTCF would stop the extrusion process. This simple mechanism provides an explanation for the accumulation of cohesin at CTCF sites and the co-localization of both proteins at loop anchors and TAD boundaries. Although the loop extrusion model provides an attractive explanation for the 3D organization of chromosomes, it is important to note that the model was developed through simulations and without direct observations of the SMC complex motor activity and loop extrusion function. The first direct observation of SMC motor activity came from real-time imaging of condensin translocation along tethered lambda DNA that was stretched under flow conditions *in vitro* [181]. Condensin is an SMC protein complex that regulates chromosome condensation in mitotic chromosomes. In this particular experiment, the rate of extrusion by condensin was estimated to be 1.5 kb/s while hydrolyzing 1 ATP/s. These measurements demonstrate the plausibility of rapid mitotic chromosome compaction by condensin via loop extrusion. Perhaps more importantly [182], cohesin-dependent loop extrusion process was observed using a similar real-time imaging approach. Similar to extrusion by condensin, cohesin-dependent loop extrusion proceeded at 2 kb/s. However unlike the majority of condensin-mediated one-sided extrusions, interphase loop extrusion by cohesin is a symetric, two-sided process. This series of *in vitro* experiments provides a clear evidence for the existence of the loop extrusion process. However, DNA in the nucleus is wrapped into chromatin and bound by proteins that can act as an obstacle that might slow down loop extrusion *in vivo*.

Matching between Hi-C experiments and simulations has led to the conclusion that the estimated rate of extrusion *in vivo* is about 10 times slower than that for naked DNA, with suggested rates on the order of 0.1 kb/s (as reviewed here, [183]). According to a recent study [184], human condensins and cohesins were observed to be able to bypass single nucleosomes without displacing them. This finding is consistent with the molecular sizes of each complex, as a single nucleosome measures 11 nm and the SMC complex ring is approximately 50 nm in size. However, it is less obvious how larger structures such as nucleosome clutches or DNA/RNA polymerases could be compatible with chromatin loop extrusion. Given the molecular size of CTCF, which is approximately 3-5 nm, it is not immediately obvious how this protein could effectively halt or pause the process of cohesin extrusion.

Several mechanisms have been proposed to explain how CTCF can prevent the process of loop extrusion (**Figure 19**). However, it has only recently been observed that CTCF is sufficient to block the diffusion of cohesin *in vitro* [185]. This phenomenon appears to be controlled by the N-terminus domain of CTCF [186], consistent with the fact that in 90% of CTCF loops, CTCF motifs point toward each other (*the CTCF convergence rule*). Since both CTCF and cohesin have a dynamic binding behavior to chromatin, 1-2 min vs 22 min of residence time respectively [187] as quantified by single-molecule imaging, there is a considerable probability that one of them will dissociate from the DNA before they meet each other [187]. However, if cohesin encounters two converging CTCF binding sites that are already occupied, a CTCF loop can be stabilized. Recent live-cell measurements of CTCF loop have provided compelling evidence that cohesin-mediated loops between CTCF sites separated by 500kb last 10-30 minutes [188, 189]. The dynamics that govern cohesin-mediated chromatin looping are regulated by two mutually exclusive factors : the cohesin loading factor NIPBL and the cohesin release factor WAPL. Depletion of WAPL significantly increases the amount of cohesin on DNA. This leads to a global compaction of chromatin [190] during interphase, as well as a dramatic redistribution of cohesin into axial chromosomal regions known as "vermicelli" and extended chromatin loops. Conversely, depletion of NIPBL [191] results in a decrease in cohesin occupancy on chromatin, leading to a widespread loss of TADs and loops.

There is increasing evidence that TADs and chromatin loops arise from the interplay between SMC complexes and barrier proteins such as CTCF through the loop extrusion process [173]. Consequently, loss of CTCF in mammals results in a significant disruption of TAD insulation and CTCF loops [177]. This finding is supported by the fact that TAD structures are completely absent in the *C. elegans* autosomes [192] and that CTCF appears to have been lost during evolution [193].

However, a CTCF-independent genome folding process may exist. This is evidenced by the fact that some loops in mammals are maintained after degradation of the loop extrusion factor [194], TAD-like structures are present on the dosage-compensated X chromosomes in *C.elegans* [192] and that the *drosophila* homologue of CTCF (dCTCF) is preferentially enriched within TADs and is only mildly enriched at TAD borders [4, 195] with no preferential convergent orientation as seen

in mammals [16].

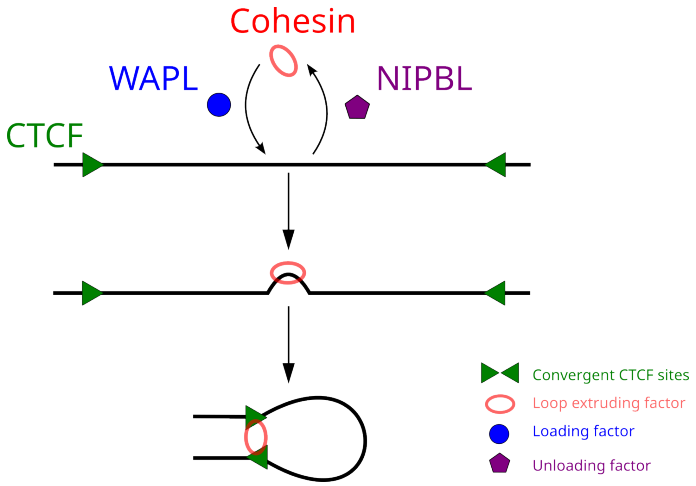


Figure 19: Loop extrusion - Simplified version of the cohesin-mediated loop extrusion model.

1.2.1.2 Alternative Mechanisms

Recent studies show that upon depletion of CTCF, cohesin or WAPL in mESC, approximately 20% of loops are significantly decreased, but the remaining 80% of loops are largely unaltered. Anchors of unaffected loops are mostly enriched in promoters and enhancers [194]. Remarkably, depletion of CTCF and cohesin has little effect on E-P/P-P loops, as approximately 80% of E-P contacts and 90% of P-P contacts remain unchanged. In addition, other recent studies have shown that tissue-specific EP interactions are weakly enriched in CTCF at both anchors [76]. Taken together, these observations suggest that "regulatory loops" involving enhancers and promoters are formed independently of CTCF/cohesin. One possibility is that these chromatin interactions between regulatory elements are mediated by transcription factors or other complexes such as Mediator.

TFs, exhibit relatively short residence times, ranging from less than a second to tens of seconds [196]. This suggests that the dynamics and probability of chromatin loops regulated by TFs may significantly differ from those regulated by CTCF, which has a moderate residence time of 1-2 minutes [197].

Through direct TFs oligomerization

One of the simplest ways to illustrate how enhancers and promoters can interact in the absence of CTCF is through the direct oligomerization of transcription factors (TFs). TFs are enriched at both enhancers and promoters. If TFs are able to interact with each other through protein-protein interactions, they may serve as

a bridge between enhancer and promoter regions. This hypothesis is supported by early observations that certain eukaryotic TFs can self-associate, thereby bringing together distant DNA segments *in vitro* [198]. Another example, is the transcription factor Yin Yang 1 (YY1), which binds to both active enhancers and promoters, forming dimers that facilitate the interaction between these DNA elements [199].

Through intermediate factors/complexes

Another possible way for enhancers and promoters to interact is through the presence of a cofactor that bridges the gap between transcription factors bound at both sites (enhancer and promoter). For example, in the mouse erythroid cells, the cofactor Lmo2 bridges the interaction of Ldb1 with the TFs Gata1, Tal1, and E2A [200]. More interestingly, the recruitment of Ldb1 (by a designed zinc finger protein) to the promoter of β-globin is sufficient to form a chromatin loop between the β-globin and LCR regions [170] (**Figure 20**). Another example of a complex that bridges enhancers and promoters is Mediator (see 1.1.1.2.2), which binds to both sites via intermediates. Depletion of Mediator results in a redistribution of the cohesin complex on chromatin and reduces enhancer-promoter interactions, highlighting the critical role of Mediator in facilitating physical communication between enhancers and promoters [201].

Through the formation of condensates

In addition to strong protein-protein interactions between transcription factors, coactivators or other complexes, weaker multivalent interactions can also occur between TFs. Most of eukaryotic transcription factors, nearly 80%, have extended intrinsically disordered regions (IDRs) [202]. IDRs are polypeptide segments that are characterized by a lack of hydrophobic amino acids. Conversely, they are highly enriched in polar or charged amino acids [203]. These regions are highly flexible and can adopt multiple conformations, allowing them to interact with different binding partners through weak, multivalent interactions. Weak multivalent interactions between IDRs have been proposed to lead to the formation of condensates or hubs that exhibit liquid-liquid phase separation properties [204]. Recent studies have shown that several TFs and cofactors form condensates or hubs *in vitro* and *in vivo* through their IDRs [205, 206, 207]. For example, Mediator and RNA Pol II have been shown to form transient condensate in living mESC that is sensitive to transcriptional inhibitors [207]. However, few experiments to date have established a causal relationship between these condensates and the shaping of the 3D genome. As an example, the interprobe distance between two estradiol-induced enhancers (NRIP1 and TFF1) measured by FISH was increased by 1,6-hexanediol treatment (a widely used drug to disrupt phase-separated condensates [208].) compared to control [209].

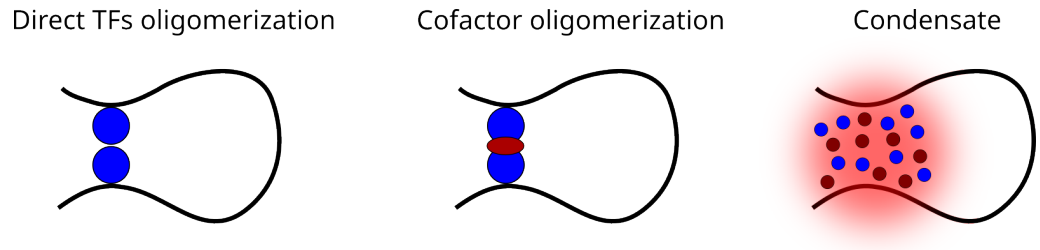


Figure 20: Modes of TF action on 3D Genome Organization - Illustration of the different modes of action of TFs in shaping the 3D genome. Left : Direct TFs oligomerization. Middle : Cofactor oligomerization. Right : Condensate formation. Figure adapted from [210].

.

1.2.2 Acquisition, Modulation, and Deregulation of 3D genome

Over the past decade, technological advances have provided a clear understanding of how chromatin is folded in the nucleus. It is becoming increasingly evident that mapping the 3D genome architecture could help to explain how regulatory elements and genes are connected in 3D to ensure precise spatiotemporal regulation of genes. In this section I will discuss how this particular 3D genome scaffold is established and modulated to support different transcriptional programs. In addition, I will explore how any dysregulation of this architecture can ultimately lead to the development of various diseases.

1.2.2.1 Acquisition of 3D genome

Interphase chromatin is organized into loops, TADs, and compartments (as discussed in section 1.1.2). Since chromatin architecture is thought to have a significant impact on transcriptional regulation, it is crucial to determine whether this particular scaffold is established *de novo* during early embryogenesis or inherited from gametes.

The genome of mouse germinal vesicle stage oocytes (immature oocytes) is also organized into loops, TADs, and compartments [211]. However, during oocyte maturation, the intensity of TADs and loops decreases significantly [211]. At a later stage of maturation, during meiosis II, mature oocytes were found to be completely devoid of TADs and loops, as demonstrated in several studies [212, 213]. Interestingly, mouse sperm chromatin, which is packaged around protamines rather than nucleosomes [214], has been found to contain TADs [213], in contrast to human sperm, which appears to lack the typical 3D folding. This difference in 3D genome architecture between mouse and human sperm is thought to be related to differences in CTCF abundance in gametes between the two species [215]. During fertilization, the two gametes, the sperm and the oocyte, fuse to form the zygote. In all animals,

the early stages of embryonic development are controlled by maternally deposited elements, including mRNAs and proteins, in the absence of zygotic transcription. These maternally deposited elements play critical roles in guiding early development, determining the body plan and initiating zygotic gene expression. Zygotic genome activation (ZGA) is characterized by a widespread recruitment of the RNA polII to promoters [216] and an increase in DNA accessibility [217]. This change in the epigenetic state of chromatin is accompanied by a substantial remodeling of the 3D chromatin structure.

In *Drosophila*, mouse and human, loops, TADs, and compartments are completely absent at the first stage of embryonic development before zygotic genome activation (ZGA), which occurs around nuclear cycle 14 in *Drosophila* [6] (**Figure 21**), at the late two-cell stage in mouse [213, 212] and at the late eight-cell stage in human [215]. During ZGA, the 3D chromatin structure gradually changes, as evidenced by the appearance of TADs, loops, and the formation of compartments. In contrast to the gradual appearance of the 3D chromatin architecture during ZGA observed in *Drosophila*, mouse, and human, zebrafish exhibits a unique pattern of successive gain and loss of the 3D scaffolding. Compartments and TADs are present prior to ZGA, at 2.25 hours post-fertilization (hpf), but then disappear at 4 hpf before recovering in the following hours of development. By 24 hpf, the 3D chromatin structure has returned to pre-ZGA intensities again.

The establishment of the 3D chromatin architecture during ZGA raises an important question regarding the potential causal relationship between genome architecture and it's function. In other words, is 3D chromatin structure necessary for the activation of transcription, or is it simply a consequence of transcriptional activation?. Studies in mice and *Drosophila* [6, 212, 213] have used inhibitors of transcription, such as triptolide or alpha-amanitin, to investigate the role of transcriptional activation and TAD formation during ZGA. Surprisingly, global transcriptional inhibition does not prevent TAD formation in mice and *Drosophila* in contrast to the human embryo, where ZGA is required for TAD formation [215]. The difference observed between human embryos and other organisms may indicate a different mechanism of genome folding during ZGA. While CTCF protein is present in mouse embryos before ZGA, it is absent in human embryos until the 8-cell stage, suggesting that zygotic expression of CTCF is critical for TAD formation in humans. This finding does not necessarily imply that transcription is required for TAD and loop formation, but rather suggests that the extrusion machinery required for proper genome folding is not yet produced until after ZGA in human embryos. In summary, although there may be some differences between species, the basic process of gradual acquisition of 3D genome structure during ZGA appears to be largely conserved.

1.2.2.2 Modulation of the 3D genome

Since chromatin structure is thought to play a critical role in gene regulation, it is important to investigate how this scaffold is modulated to support changes in transcription, particularly during mitosis, when global transcriptional shutdown

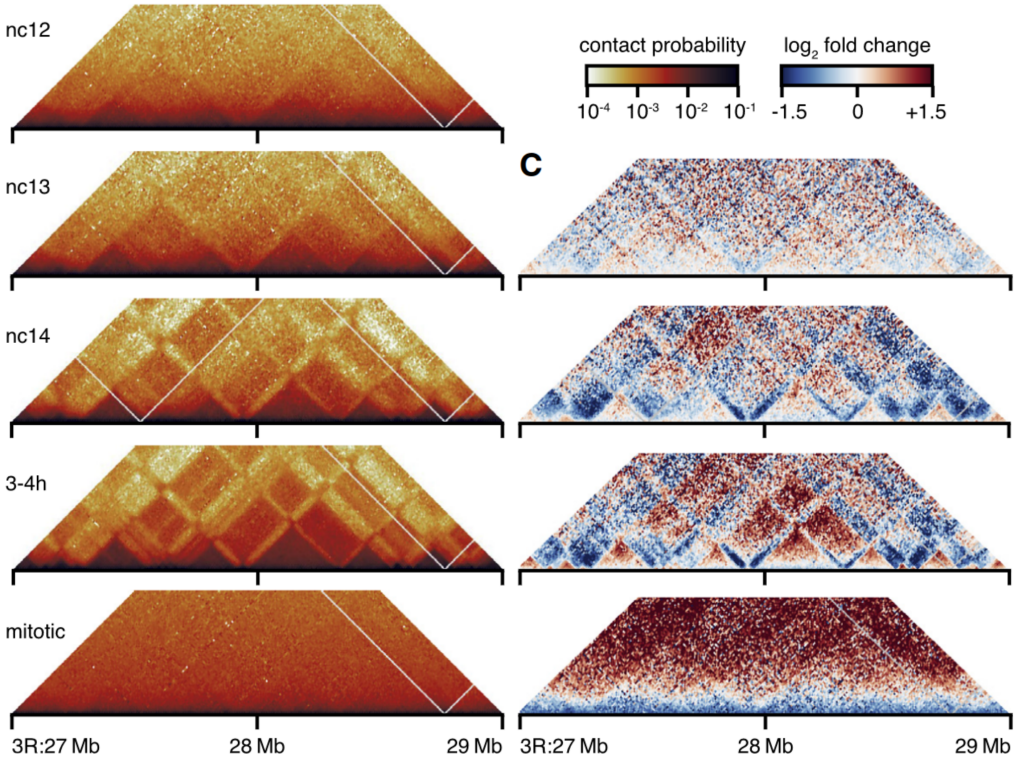


Figure 21: Acquisition of 3D genome - Left : Chromatin interaction probability maps from of *Drosophila melanogaster* at different developmental stages. Right : Differential interaction probabilities for each stage relative to nc12. Figure adapted from [6].

occurs [218, 219], and during differentiation, when cellular identity is established through the activation or repression of specific genes.

Changes during the cell cycle

During the cell cycle, chromatin switches between two different folding states. In interphase (G1-, S-, and G2-phases), chromatin is largely decondensed. However, in metaphase of mitosis, chromatin progressively undergoes a gradual conversion into a highly condensed set of rod-shaped structures. However, most studies based on chromosome conformation (C-based) tend to focus on chromatin organization in non-synchronous cells, where almost 95% of the cells are in interphase, potentially masking heterogeneity in 3D organization during the cell cycle [220]. To gain a better understanding of chromosome architecture during the cell cycle, researchers have explored chromosome conformation during early G1-, mid-G1-, S-, and M-phase cells. Their findings suggest that chromatin organization during mitosis significantly differs from that observed in each of the different phases of interphase [221]. In mitotic cells, the previously described levels of organization in compartment, TADs and loops tend to disappear. The ability to erase and re-establish chromatin structure during the cell cycle appears to depend on the interplay between cohesin and con-

densin. This loss of the 3D structure is associated with the dissociation of cohesin from chromosome arms, and the re-folding of chromosomes into helical arrays of nested loops by condensin. As cells exit mitosis, the transition from condensin to cohesin results in the reassembly of the 3D genome in early G1 [221, 222, 223].

Changes during the differentiation

The proper organization of chromatin structure is thought to be critical for regulating the dynamic switches between transcriptional programs during lineage specification and cell differentiation. For example, during the differentiation of human ESCs into four derived lineages, Mesendoderm (ME), Mesenchymal Stem Cells (MSC), Neural Progenitor Cells (NPC), 36% of the genome undergoes compartment switches in at least one of the lineages [149], although these chromosome compartment switches are correlated with changes in gene expression (**Figure 22**. In contrast, TAD boundaries are relatively more stable than compartments during differentiation [10, 224]. However, changes in chromatin structure mostly occur within TADs during cell differentiation. For example, promoter-capture HiC (pcHiC) study in 17 human primary hematopoietic cell types showed that CRE interactions are highly cell type specific [10]. Similarly, Hi-C studies on lineage committed cells from ESCs to cortical neurons reported that CREs interaction are correlated with up-regulation of cortical neurons genes. Genome Architecture Mapping (GAM), a ligation-free sequencing method (as introduced in section 2), was used to generate high-resolution chromatin interaction map for three different neural cell types (oligodendrocytes, pyramidal glutamatergic neurons and dopaminergic neurons), in mice *in vivo*. This study [14] revealed that the majority of the significant differential chromatin contacts between pyramidal glutamatergic neurons and dopaminergic neurons involve regions containing binding sites for different sets of transcription factors that are differentially expressed in each neuronal cell type. These results suggest that differential binding of TFs may fine-tune cell-types-specific chromatin interactions.

However, other studies have shown that specific interactions between CREs are established before the target genes are activated [12, 13] during early embryonic development. For example, in *Drosophila*, the three promoters of the *dorsocross* genes, namely *doc1*, *doc2*, and *doc3*, contained in the doc TADs, show significant interactions with four putative enhancers, CRMa, CRMb, CRMc, and CRMd, in both pre-differentiated tissues where the doc genes are expressed, (dorsal ectoderm), and in tissues where *doc* is repressed (mesoderm and neuroectoderm). This study suggests that pre-established interactions between CREs may be bound by different sets of transcription factors, either activators or repressors, to regulate different transcriptional programs of the target genes in specific tissues. Similar studies in the human epidermal differentiation system found that a set of CRE contacts referred to as "stable" are found in both undifferentiated progenitor and differentiated cells [225].

The two contrasting views of enhancer-promoter interactions in transcriptional activation, namely "stable" and "dynamic," may seem surprising. However, recent

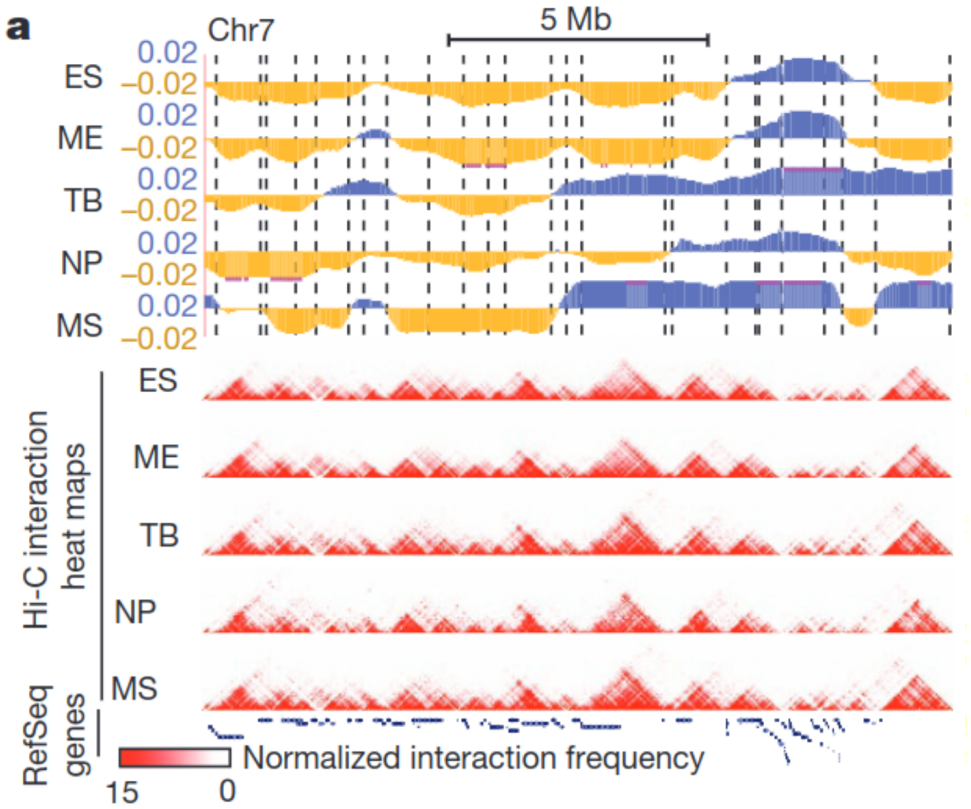


Figure 22: Dynamic of chromatin structure during differentiation of human ES cells - Hi-C contact matrices in H1 ES cells and H1-derived lineages, demonstrating that chromatin compartments and intra-TAD interactions are dynamic. Embryonic Stem cell (ES), MEsendoderm cell (ME), TrophoBlast-like cell (TB), Neural Progenitor cell (NP), Mesenchymal Stem cell (MS). Figure adapted from [149].

studies have attempted to reconcile these models by quantifying the behavior of enhancer-promoter interactions across the genome in ten different mouse embryonic tissues with distinct transcriptional programs. This study shows that a small fraction (13.6%) of E-P chromatin interactions are "stable" across the ten embryonic tissues examined. Stable E-P contacts have been proposed to be associated with proximal CTCF binding at both ends and to be uncorrelated with enhancer activity [76].

It is worth noting that the conventional model of promoter–enhancer looping has been challenged at some loci in recent years. For example, a recent live imaging approach in mESC fails to provide evidence for enhancer-promoter spatial proximity driving transcription between the promoter of *Sox2* and its essential enhancer Sox2 Control Region (SCR) [226]. More surprisingly, another study shows that the activation of the morphogenesis gene sonic hedgehog (*Shh*) during the differentiation of mESC to neural progenitors is accompanied by an increase in spatial proximity between the promoter and its enhancers [227]. Finally, it has been observed that

the mobility of enhancers and promoters undergoing differentiation-associated activity increases in living mESC [228]. The results suggest a "collision" model that allows for the dynamic communication between promoters and enhancers within a decompacted chromatin domain. However, while this model has been demonstrated for only a few specific regions, it is unclear how broadly it applies to other genomic regions.

1.2.2.3 Deregulation of the 3D genome

At the genetic level, diseases are often associated with the acquisition of genetic variants (such as substitutions, inversions, insertions or deletions) located within the coding or the noncoding genome. Given the importance of chromatin folding in gene regulation, there is an increasing interest in investigating the relationship between genetic variation, alterations in chromatin structure, and the emergence of diseases. Initially, attention has been paid to the study of coding variants because they are easier to characterize. For example, mutations in the genes encoding CTCF and cohesin, or their regulators such as WAPL or NIPBL, have been frequently associated with human diseases including cohesinopathies [229, 230] and developmental abnormalities. In mice, the cardiac-specific depletion of CTCF is sufficient to induce heart failure, to alter the strength of TAD borders, to impair the A/B compartmentalization and lead to the misregulation of disease-causing genes. Nervertheless, a significant number of genetic variants are located in the non-coding regions of the genome, particularly in distal regulatory elements such as enhancers or insulators, which have been documented to influence gene expression in cancer [231] as well as developmental disorders. Alteration of local chromatin domains such as boundary elements can lead to ectopic enhancer-promoter interaction also known as "enhancer hijacking" resulting in transcriptional deregulation.

For example, human limb syndromes, including brachydactyly, F syndrome and polysyndactyly, are caused by deletions, inversions or duplications near the boundaries of a TAD-spanning WNT6/IHH/EPHA4/PAX3 locus [232]. Using CRISPR–Cas9 genome editing, researchers generated mouse models with the same genetic alterations as those seen in humans. Disruption of the boundary element led to aberant interactions between promoters and non-coding DNA which are associated with ectopic limb expression of another gene in the locus including *Pax3*, *Wnt6*, and *Ihh* (**Figure 23**). In another case, the disruption of the boundary element at the lamin B1 locus leads to ectopic interaction between the three forebrain-directed enhancers and the *Lmnb1* promoter. This ectopic interaction results in the overexpression of lamin B1 and myelin degeneration that is observed in adult-onset demyelinating leukodystrophy (ADLD) [233].

In addition to developmental disorders, some genetic variants can also lead to tumorigenesis caused by enhancer hijacking by oncogenes. Disruption of such boundaries in non-malignant cells has been shown to activate proto-oncogenes, leading to malignancy [7]. For example, the deletion of specific TAD boundaries has been associated with dysregulation of IRS4 in sarcoma and squamous cancers [165]. In

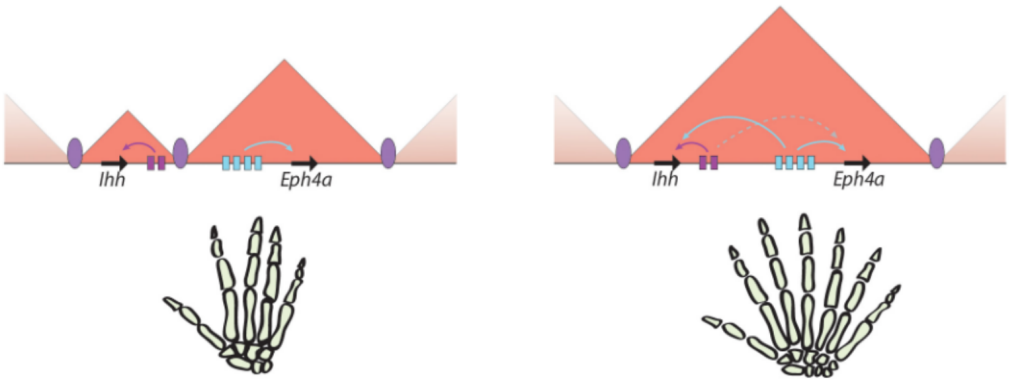


Figure 23: Deregulation of the 3D genome structure - Illustration of the phenotypic output of deletion of the TADs boundary element at the Ihh/Eph4a locus in mice. Figure adapted from [234].

addition, genomic duplications have been observed to create new chromatin domains, resulting in the overexpression of IGF2 in colorectal cancer [165].

Thus, these findings demonstrate the importance of TAD boundary sequences for *in vivo* genome function and reinforce the critical need to carefully consider the potential pathogenicity of noncoding deletions affecting TAD boundaries during clinical genetic screening [235].

2 Methodologies to study the folding of the 3D genome

The study of chromatin organization has made significant advancements in recent years, primarily due to advances in molecular biology, high-throughput sequencing, and microscopy. In this section, I will review some of the most important techniques, describe their advantages and limitations to study 3D genome architecture.

2.1 Sequencing based methods

2.1.1 3C (one-vs-one)

Chromosome Conformation Capture or 3C is a molecular biology method that aims at measuring the frequency of interaction between pairs of loci "one versus one" in order to determine how DNA sequences are physically connected to one another in the nucleus [236] (**Figure 24**). To that end, nuclei are cross-linked using a chemical crosslinking agent such as formaldehyde. Then, the cross-linked DNA are digested into small fragments using a restriction enzyme, such as *BglII, DpnII, EcoRI, HindIII*, to generate pairs of cross-linked fragments that are physically close in space but may be distant in the linear genome. These fragments are then re-ligated using a ligase to form chimeric molecules, which are further amplified by Polymerase Chain Reaction (PCR) using a specific set of primers and sequenced. Thus, genomic regions that are in close proximity are statistically more likely to be detected in a population of cells than those that do not interact frequently. The first application of the 3C technique was in haploid NKY2997 cells from *Saccharomyces cerevisiae* along the 320-kb of the chromosome III [236]. Dekker and colleagues designed 13 primers covering the entire chromosome and calculated the interaction frequency for all of the 78 combinations of primers 1 to 13. They found that the two telomere sequences of chromosome III interact more frequently than expected based on their linear genomic distance, suggesting that the chromosome is not circular but rather appears as a contorted ring. Subsequently, the technique has been widely adopted and applied to various biological samples, including those from mammals, on several loci such as the β-globin locus [169, 237, 238], the T-helper type 2 cytokine locus [239], the immunoglobulin k locus [240], and the Igf2 imprinted locus [241]. This extensive use of the method has allowed scientists to study the relationship between 3D genome architecture and transcriptional regulation. One example is the study of the conformation of the β-globin locus both in the embryonic mouse liver, where β-globin genes are transcribed, and in the brain, where they are repressed. This experiment shows that changes in the interaction between the Locus Control Regions (LCRs [242]) and the promoter of the βmaj gene are linked with gene activity [237]. Although the 3C method provides an important insights into our understanding of chromatin architecture, this method is restricted to a limited number of pairwise interactions between known DNA sequences. To overcome this

limitation, several 3C-based methods have been developed.

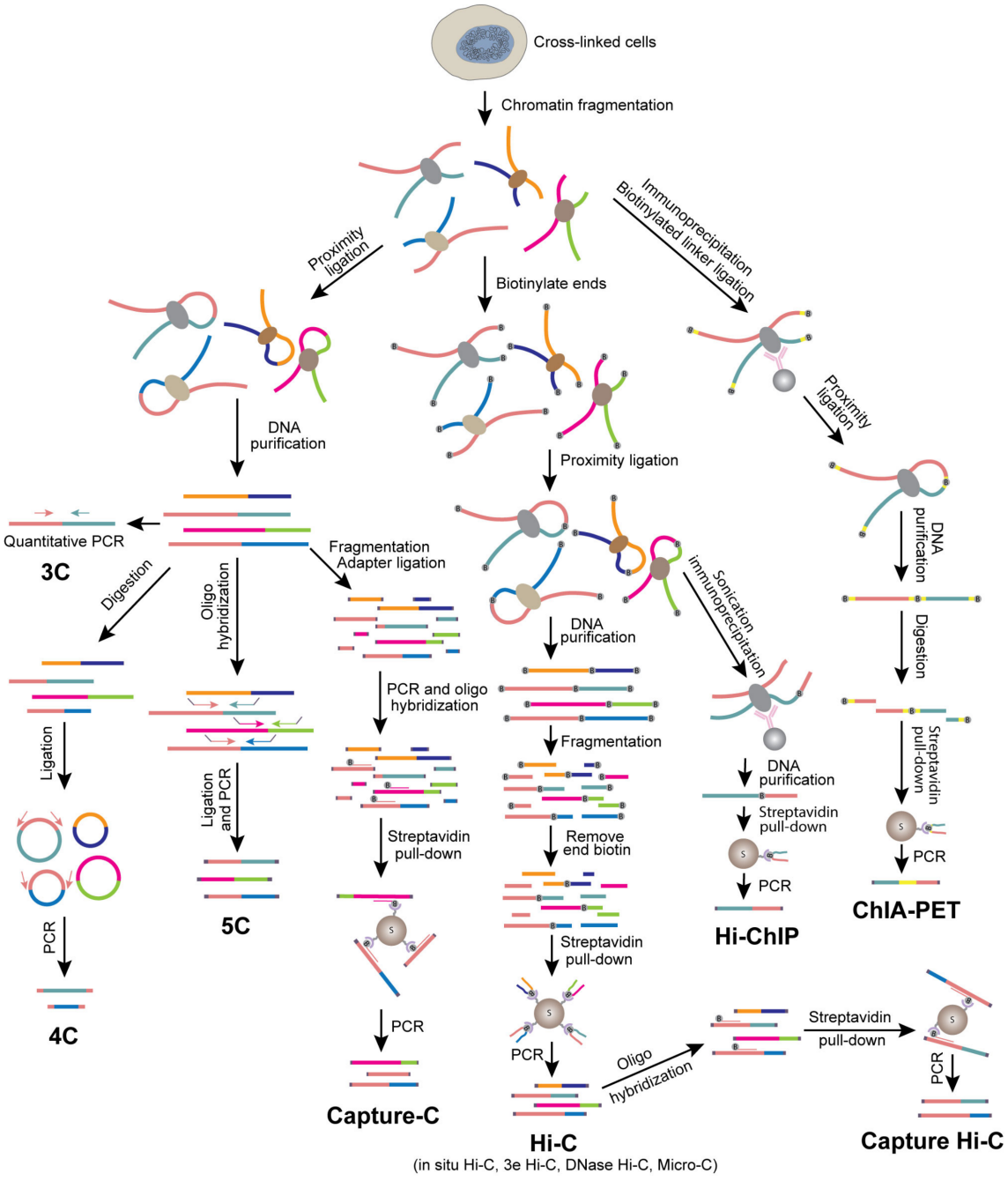


Figure 24: Overview of 3C-based methods - Chromosome conformation techniques use cross-linking of DNA fragments to study their spatial organization. The fragments are linked together and then analyzed using different detection methods. The detection principle varies from technique to technique. Figure adapted from [243].

2.1.2 4C (one-vs-all)

While the original 3C experiment typically involves the study of a limited short-range pairwise interaction (< 1 Mb), the Circular Chromosome Conformation Capture (4C) experiment allows the detection of chromatin interaction between a single restricted fragment (the viewpoint : enhancers, promoters ...) with all genomic regions sometimes located several megabases away "one versus all". A fundamental advantage of this method over the 3C method is that only the nature of the viewpoint is needed to design the primers and to capture the interactions with unknown interacting sequences. To achieve this goal, the chimeric molecules formed after the first ligation process are self-circularized by an additional ligation step. These circularized molecules are then used as templates for inverse-PCR (iPCR) with a pair of primers that recognize both ends of the viewpoint sequence [4, 244, 245]. In this way, the known viewpoint sequence is used to amplify the unknown interacting sequence. Next the amplified regions can be revealed with microarrays or Next Generation Sequencing (NGS) approaches (**Figure 24**). As a result, 4C data are represented in a 2D graphic showing the number of reads over a given genomic region centered on the viewpoint. Typically, the 4C signal starting from the viewpoint is high and decreases as the genomic distance increases, reaching a certain level of noise at larger distances. Alternatively, 4C can be represented by an arc connecting a locus i and j when a contact Cij exceeds a certain threshold value. 4C experiments at the β-globin locus in both transcriptionally active and inactive tissues show that the β-globin genes can contact a different set of distal genes that have the same transcriptional status in the two tissues [244]. Other studies show that the promoter of odorant genes can form 3D interactions with a regulatory non-coding region, the H enhancer [78] in mouse. Thus, changes in 3D architecture can facilitate communication between genes and regulatory elements to modulate their activity. 4C can therefore be used to identify putative regulatory elements of a given gene or genes regulated by a given enhancer. The throughput of 4C can be easily increased to allow the study of multiple viewpoints in a single experiment. However, this cannot be applied to the study of all possible interactions within a given genomic region because it would require the design of multiple pairs of primers.

2.1.3 5C (many-vs-many)

To adress this need, the Carbone Copy Chromosome Conformation Capture (5C) was proposed [246]. Basically, 5C uses the same principle as 3C, but it allows the simultaneous interrogation of multiple regions spanning several megabases at once "many-vs-many". To do this, chimeric molecules from the 3C library are incubated with a complex bioinformatically designed mix of primers (a forward primer and reverse 5' phosphorylated primers). Primers that are designed to face each other anneal to the ligation junction of the chimeric fragments. The annealed primers are then ligated with Taq ligase prior to amplification of the 5C library using known

sequences located at each extremities of the forward and reverse primers. The amplified 5C library is then analyzed using high-throughput sequencing or with microarrays (**Figure 24**). Typically, 5C results are represented as a square heat map with the x and y axes representing the genomic regions and the color corresponding to the number of contacts. Alternatively the genome can be plotted as a circle with contacts represented by edges between pairs of loci, or less commonly, a graph with nodes as loci and edges as contact is used [33] (see 2.4.2). Since the β-globin locus has been extensively studied with 3C and 4C approaches, it also serves as validation for the 5C approach. 5C was applied to a 400-kb region containing the human β-globin locus and successfully identified previously observed looping interactions. However, because 5C is able to capture all combinations of interactions between regions spanning the locus, it also identified a new looping interaction between the LCR and the γ–β-globin intergenic region. 5C has also been used to study the organization of the human chromosome 21 through the cell cycle [221] and the 3D architecture of the Hox clusters in human and mouse [247, 248]. In practice, the size of the genomic region that can be analyzed in a 5C experiment is limited to a few megabases. Even though web-based programs such as my5C [249] have been developed to facilitate the design of 5C experiment, the need to design hundreds of primers is a significant limitation of the method. To overcome these limitations, new methods have been developed to capture all types of interactions, in an unbiased manner.

2.1.4 Hi-C (all-vs-all)

Hi-C is a high-throughput chromosome conformation capture technique developed by Lieberman-Aiden et al. [1] that allows the capture of all genome-wide interactions "all-vs-all" without *a priori*. Again, the protocol of Hi-C is very similar to other 3C-based assays with the generation of 3C template by a digestion and a slightly modified ligation step. After restriction digestion of the cross-linked DNA fragments, the DNA overhangs are filled with biotinylated nucleotides. The blunt ends are then ligated together. The biotinylated chimeric fragments are then pulled down with steptavidin beads. Finally, adapters are ligated to the 5' and 3' ends of the chimeric molecules prior to PCR amplification and sequencing using the paired-end sequencing method (**Figure 24**). In the original Hi-C study, researchers reconstructed for the first time the spatial proximity maps of the human genome with a resolution of 1Mb. This revealed that the entire human genome is hierarchically organized at a multiscale level (see 1.1.2). According to Rao and colleagues, [148] the definition of the resolution of Hi-C experiment is the "minimum size window which, when used to calculate the genome coverage, leads to 80% of the windows covered by at least 1000 reads". Thus Hi-C resolution is limited by (i) the sequencing depth (ii) and (ii) the restriction enzymes used in the assay.

Motivated by the biological significance of the 3D genome architecture, significant efforts have been made to push the Hi-C resolution to its limits. One of the most striking efforts to improve the resolution of Hi-C data has been to increase the

sequencing depth. In the original Hi-C paper, Lieberman-Aiden et al. generated about 30 million read pairs to create a genome-wide contact matrix, dividing the genome into 1Mb blocks. Five years later, Rao et al. [148] went up to 6.5 billion paired-end reads to reach the 1kb resolution barrier in the human genome with a Hi-C map binned at 950bp. More recently, Bonev et al. 2017 [10] achieved even higher coverage in the mouse genome with up to 7.3 billion read pairs to reach a resolution of 850bp. It is important to note that when Hi-C is applied to smaller genomes (e.g., *Drosophila*), a smaller number of reads is needed to reach high coverage. For example, Wang and colleagues [250] achieved a resolution of 500bp in *Drosophila* with only 1.5 billion read pairs. Although sequencing costs have decreased, the cost of generating a Hi-C matrix grows quadratically with the number of bins in the matrix. Goel et al. [251] estimated that the cost of sequencing alone would be approximately $1.6 billion to statistically generate one read per nucleosome across the human genome. To address this cost limitation, one approach is to reduce the size of the region studied using capture-C approaches, as detailed in the next section (see 2.1.6). In addition to the sequencing depth, it is also important to note that the resolution of Hi-C can also be limited by the restriction enzyme used in the experiment. In fact, the length of the site recognized by an enzyme determines the average length of the fragmented regions and therefore the resolution. For example, the 6bp cutter *HindIII* will generate an average fragment length of approximately 4kb ($4^6 = 4096$bp), while the 4bp cutter *DpnII* should theoretically result in a mean digested regions of 256bp ($4^4 = 256$bp). Over the last decade the use of *DpnII* or *MboI* as a 4bp cutter has become the protocol of choice for mapping chromatin interactions at high resolution. However, the use of restriction enzymes limited the resolution of Hi-C to 500bp to 1kb. To overcome these limitations, researchers have alternatively turned to DNAse I or MNase as sequence-agnostic nucleases [252, 253].

2.1.5 Micro-C

Together with DNAse Hi-C [252], Micro-C [253] are methods that offer the advantage of being independent of restriction enzymes (**Figure 24**). In the case of Micro-C the fragmentation of the chromatin is performed at the nucleosome level by micrococcal nuclease (MNase), resulting in an increase in fragment density. This increase in fragment density leads to a reduction of the sequencing depth required to build a 1 kb resolution matrix. For example, Lee and colleagues [254] estimated that approximately 2 billion reads are required to achieve a resolution of 1 kb with Micro-C, while Hi-C requires approximately 6.5 billion reads to achieve the same resolution [148]. Initially developed in budding yeast [253] and then adapted to the study of both human and mouse cells [253], Micro-C has enabled the study of the fine-scale chromatin structure of nucleosomal interactions. Surprisingly, they found some evidence that the 30nm fiber is organized into "clutches" supporting the zig-zag tetra-nucleosome folding model [255]. Side-by-side comparison of Micro-C data [253] with the deepest Hi-C data [10] shows that Micro-C is able to detect previously reported 3D chromatin structures such as compartments, TADs and loops with high reproducibility. In addition to structures larger than 20kb, Micro-C has

shown a remarkable ability to assess local chromatin folding on the scale of 100bp to 20kb. These short-range interactions correspond mainly to contacts between *cis*-regulatory modules (CRMs) such as promoter-promoter, enhancer-promoter or enhancer-enhancer. Micro-C has dramatically improved our view of the local chromatin architecture by efficiently detecting interactions between CRMs. However, it is not as efficient as Hi-C at mapping long-range interactions and interchromosomal contacts [256]. Interestingly, this apparent discrepancy can be compensated by increasing the sequencing depth. In fact, Micro-C data with 3 billion paired reads systematically reported more long-range loops than Hi-C data with 1 billion paired reads [254]. With the development of restriction enzyme-free assays, the resolution of 3C approaches is no longer dependent on fragment size but is determined by sequencing depth. To better resolve chromosomal structure while reducing the sequencing costs associated with genome-wide high-resolution Hi-C or Micro-C experiments, scientists are developing Capture-C approaches.

2.1.6 Capture-C

Capture-C approaches are a group of techniques that allow for "many-vs-many" analysis within a region of interest. Several methods have been developed such as Chromosome Conformation Capture Carbon Copy (5C) (see 2.1.3) (**Figure 24**), Targeted Chromatin Capture (T2C) [257], Capture Hi-C (cHi-C) [258], Region Capture Micro-C (RCMC) [259], Micro-Capture-C (MCC) [260] and Tiled-Micro-Capture-C (Tiled-MCC) [261]. The Tiled-MCC approach allows for the generation of regional contact matrices with extremely high resolution (up to 20bp) and for relatively low sequencing costs. Cells are permeabilized with digitonin instead of classical detergents. Next, the chromatin is digested with MNase, similar to what is done in Micro-C assays. A highly complex biotinylated oligonucleotide library is then designed to densely cover the region of interest. These improvements result in a high quality dataset with an higher fraction of usable reads with approximately 80% on-target reads for the generation of high resolution local contact matrices. A comparison of Tiled-MCC and Micro-C data shows that Tiled-MCC is able to capture contacts between *cis*-regulatory elements, such as enhancer-promoter interactions and long-range CTCF interactions that are not detected by conventional Micro-C approach.

This type of capture-C approach can also be extended to study thousands of sequences like promoter regions. Promoter capture Hi-C (PcHiC) [258] can generate a genome-scale map of interactions between gene promoters and their regulatory elements. This allow the precise identification of potential *cis*-regulatory regions of a given gene. Applied to human pancreatic islets [11] PcHiC allowed the identification of 1300 *cis*-regulatory regions of glucose-dependent genes.

2.1.7 Other sequencing-based methods

3C-based methods have been developed to generate highly detailed genome-wide interaction profiles for various biological samples. Increased resolution now allows the detection of short- and long-range interactions between *cis*-regulatory elements of the genome. By combining the one-dimensional profile of protein binding obtained through ChIP-seq with the contact matrix over a region of interest, it is possible to indirectly associate contacts with the presence of specific factor on the chromatin. The ChIA-PET [262] or HiChiP [263] techniques allow the combination of 3C and chromatin immunoprecipitation sequencing. Thus, these techniques can be used to directly map short- and long-range interactions associated with a protein of interest. By investigating the role of cohesin and CTCF binding in loop formation with HiChiP, it was shown that the vast majority (80%) of genome-wide loops are bound by CTCF and cohesin, which is in close agreement with Hi-C data [263]. All of the 3C methods described above rely on digestion and ligation to capture interacting DNA segments. More recently, a novel genome-wide method called GAM [264] for Genome Architecture Mapping has been developed to capture three-dimensional proximities between genomic loci without ligation. GAM is a method that uses ultracryosectioning, laser microdissection, and DNA sequencing to study genome organization. Cells are thinly cryosectioned and individual nuclear profiles (NPs) are isolated by laser microdissection. The DNA from each NP is extracted and sequenced. Loci that are found within the same NP are detected and identified as interacting partners. The collection of large NPs sliced at random orientations allows the reconstruction of contact matrices similar to those generated in Hi-C. A significant advantage of GAM over conventional Hi-C is its ability to identify clustering of multiple gene loci, also known as multi-way interactions.

Overall, these methods for studying 3D genome folding based on sequencing have several limitations. First, except for the GAM [264] and Tri-C [265] methods, 3C-based approaches can only measure pairwise interactions, making it impossible to study clustering of multiple loci, such as hubs formation. Second, most 3C studies with the exception of some single cell Hi-C experiments [211] are performed in bulk. Thus, they provide only average information about genome organization. Individual cells can have different 3D structures. Third, 3C methods are not suitable for studying physical properties such as distances between loci and compartments, or nuclear volume. Fourth, sequencing-based techniques do not provide spatial information, making it impossible to study the link between 3D genome architecture, transcription, and the positioning of cells within a tissue. To overcome these limitations and to access the missing information, many imaging based technologies have been developed.

2.2 Microscopy based methods

2.2.1 Conventional DNA-FISH

Over the past decade, *in situ* hybridization (ISH) techniques have become increasingly popular because they allow direct observation of specific DNA fractions in the nucleus. The technique consists of hybridizing DNA probes complementary to a specific genomic region of interest. The first ISH procedure was used to stain *Xenopus* DNA [266] with a radioactive DNA probes in solution. Soon after this pioneering experiment, fluorescent labels replaced the radioactive ones [267, 268, 269], being less hazardous and more stable. Historically, DNA fluorescence *in situ* hybridization probes were prepared by cloning chromosomal of the targeted regions into bacterial artificial chromosomes (BACs) and subsequently labeling them with fluorescent dyes [270]. This time-consuming and logistically challenging limitation of conventional FISH labeling approaches was uplifted by Oligopaint [271], which enables flexible, rapid, and efficient design and synthesis of FISH probesets. Since its implementation, DNA FISH techniques have become increasingly popular because they allow the direct observation of the 3D genome architecture in a manner that is complementary to 3C methods. In recent years, several studies have used DNA-FISH combined with super-resolution microscopy to visualize chromosome territories [271, 272, 273], chromatin folding in different epigenetic states [147, 274], and TADs in single cells [147, 275]. However, a major limitation of FISH, is the small number of spectrally distinct fluorophores that can be combined in a single experiment, limiting the number of loci that can be detected simultaneously. Several methods have been developed to overcome this limitation with "combinatorial labeling". For example, the multicolor DOPE-FISH or MiL-FISH approaches use double or quadruple-labeled oligonucleotide probes with different dyes to visualize up to six targets [276, 277]. More recently, confocal laser scanning microscopy combined with white-light laser technology methods has allowed the unambiguous detection of eight fluorophores with distinct spectral properties [278]. A major challenge in FISH has been to increase the number of loci that can be analyzed simultaneously. Innovative oligonucleotide designs combined with microfluidics, have made it possible to image multiple genomic loci simultaneously in a single experiment.

2.2.2 Sequential DNA-FISH

Visualization of multiple genomic loci in a single experiment using the DNA-FISH technique is key to revealing the structure of a given locus in 3D. These innovative methods take a different approach. Instead of simultaneously imaging different probes with different dyes, multiple genomic loci are sequentially imaged with the same fluorophore within imaging/bleaching cycles. Technically, DNA-FISH is performed with a series of small oligonucleotides. Each oligonucleotide within the library contains three regions: (1) A region of homology complementary to the target genomic regions. (2) A tail consisting of locus-specific readout sequence, which is further recognized by fluorescent dyes. (3) A forward and reverse sequence primer

positioned at the 5' and 3' ends for PCR amplification of the library. All probes contained in the library are first amplified by emulsion PCR (emPCR) and *in vitro* transcription (see 3.2.2). Subsequently, the library is hybridized to the genomic DNA. The sample is then placed in a microfluidic device programmed to deliver different wash and hybridization solutions. An imaging oligo capable of recognizing the locus-specific readout sequence is hybridized to the sample prior to imaging. Afterward, the imaging oligo is washed out, and the cycle is repeated with another imaging oligo hybridizing to a different locus. Repeating this cycle n times allows the imaging of all the individual n loci encoded in the library. This sophisticated design combined with microfluidic devices, allows the probing and recording of the three-dimensional positions of tens to hundreds of loci within single cells or population-based manner at 100kb resolution [31]. Further improvements of this method led to the development of a series of chromatin tracing methods to visualize TAD structure in cell culture at 25kb resolution [32]. Combined with the detection of specific RNA species, these techniques provide a unique tool to study the relationship between chromosome architecture and gene expression. The Nollmann lab developped Hi-M (**Figure 25**), a method that allows for the detection of chromatin conformation and transcription of a gene in the intact *Drosophila* embryo at very high resolution (3kb) [9, 279]. In parallel with these advances, other groups have also developed a comparable technique on cryosectioning of *Drosophila* embryos (ORCA : optical reconstruction of chromatin architecture) [30]. More recently, sequential encoding schemes, specifically DNA-MERFISH, have been used to detect the conformation of whole chromosomes, encompassing over 1,000 genomic loci, along with the transcriptional activity of more than 1,000 genes within the same cells [280]. SeqFISH+ techniques achieve genome-wide coverage in chromatin imaging at resolutions of 0.5 to 1Mb along with the detection of various RNA species and nuclear landmarks [281]. The integration of cutting-edge chromatin tracing techniques with RNA-FISH, immunostaining, and imaging of chromatin marks has brought forth a new perspective to our understanding of gene regulation. These spatial multi-omics datasets provide valuable insights into the intricate dynamics of gene expression and chromatin organization.

At the chromosomal scale, the mapping of the spatial organization of the central 100kb regions of all 34 TADs on chromosome 21 in IMR90 cells reveals that TADs are largely organized into two compartments that are spatially arranged in a polarized manner [31]. This spatial segregation of TADs along the chromosome is consistent with the A/B compartmentalization of active and repressed chromatin (see 1.1.2.3). At higher resolution, tracking chromatin organization in successive 30-kb segments has revealed the formation of spatially distinct globular structures within single cells, termed TADs-like structures. The existence of TADs in single cells is consistent with previous results obtained using multicolor 3D-FISH combined with super-resolution microscopy [275]. However, it is important to note that these studies were performed using cultured cells with asynchronous cell cycle progression and partially distinct transcriptional programs. This inherent variability makes it challenging to establish a direct correlation between chromatin structure and function.

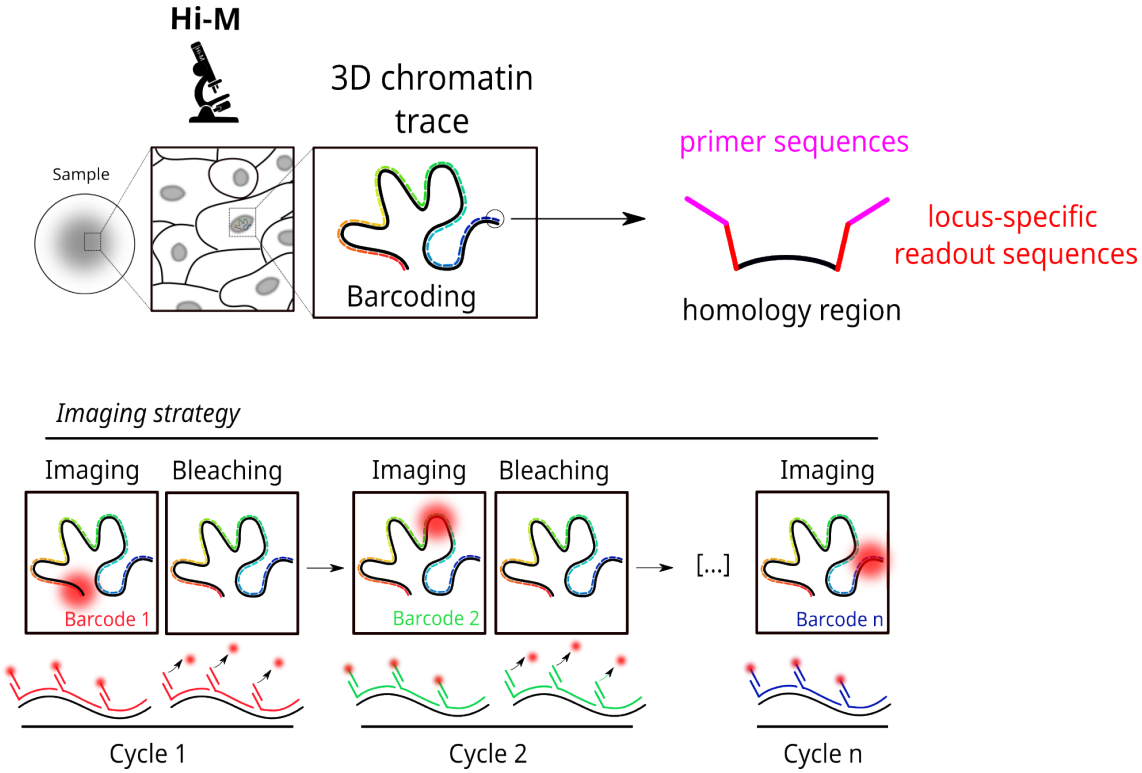


Figure 25: Schema of Hi-M experiment - Overview of the strategy used for the sequential DNA-FISH experiment. The Hi-M probes consist of a homologous region, a locus-specific readout sequence, and primer sequences. The imaging and bleaching cycles are shown below.

To further study the role of chromatin architecture and gene regulation, the Nollmann lab has been applying the chromatin tracing method Hi-M on staged *Drosophila* embryos at very high resolution (3kb). The ability of Hi-M to trace chromosome architecture and RNA expression in the whole *Drosophila* embryo is fundamental to classify cells with similar patterns of gene expression in a population-based manner. Hi-M revealed significant changes in the internal organization of *snail* TADs in the presumptive mesoderm as compared to cells in the dorsal ectoderm. The high-resolution capabilities of Hi-M, along with its ability to simultaneously probe chromatin conformation and gene transcription have enabled the specific detection of TADs structure in the presumptive mesoderm, endoderm, and ectoderm of *Drosophila* embryos [12, 282]. Hi-M enabled the detection of enhancer-promoter interactions at the *dorsocross* locus that contain 3 genes (*doc1*, *doc2* and *doc3*). These studies demonstrate that chromatin conformation exhibits significant variability at the single cell level [282]. Furthermore, they show that the physical proximity of *cis*-regulatory modules (CRMs), such as enhancer-promoter, enhancer-enhancer, and promoter-promoter, does not necessarily dictate transcriptional states [12]. Moreover, these loops between CRMs are established in early development, before zygotic genome activation (ZGA) and the emergence of TADs. These observations indicate

that such contacts can be observed in both transcribing and non-transcribing cells, challenging the notion that proximity alone directly influences gene expression (see 4.6). Invariant chromatin topologies were also been observed by ensemble sequencing methods at specific loci [283, 13]. The ability of Hi-M to trace chromosome conformation at single cell level make it suitable for studying multiway interactions between CRMs. Analysis of multiway interactions by Hi-M reveals that CRMs at the *doc* locus spatially coalesce to form hubs in single cells.

Complementary studies using ORCA revealed changes in intra-TAD organization, changes between body segments, and correlated with the spatial patterns of *Hox* gene expression of the Bithorac Complex (BX-C) locus in cryosectioned *Drosophila* embryos 10-12h post-fertilization [30]. More recently, a chromatin tracing technique called multiplexed imaging of nucleome architecture (MINA) was deployed to visualize chromatin architecture in E14.5 mouse fetal liver along with the imaging of hundred of RNA species. MINA revealed changes in A/B compartments between cell-types and increased in proximity between enhancers-promoters in hepatocyte versus non-hepatocyte cells [284]. In addition, SeqFISH+ revealed cell-type specific association and scaffolding of DNA loci around nuclear bodies that correlated with differential expression levels [285].

2.3 Simulations of polymers

The emergence of sequencing and imaging technologies has allowed to comprehensively map 3D genome architecture in different tissues under different physiological conditions. However, a full comprehension of the mechanisms behind DNA folding is still far from being understood. Polymer physics-based theoretical and analytical approaches have been used to gain insight into the intricate nature of chromatin architecture data and to elucidate the fundamental mechanisms that contribute to shaping its structure.

One of the great benefits of simulation in biology is the ability to test the underlying mechanisms of 3D genome architecture that would explain experimental observations. In polymer physics, chromatin is described as a polymer of beads on a string, where each bead represents a portion of the chromatin fiber, which may be one or more nucleosomes. In recent years, several models based on polymer physics have emerged that provide a valuable understanding of the minimal requirements for the formation of higher order chromatin structures.

2.3.1 Homopolymer model

Prior to the formulation of more sophisticated models, chromatin was considered to be a simple homogeneous polymer described by a chain of connected identical

monomers whose dynamics is controlled by thermal forces, polymer stiffness, and non-specific interactions between monomers. In such a simple model [286], the chromatin chain is represented by a series of N monomers, each having a specific size and occupying a defined position in 3D. To simulate a 3D conformation, the polymer is subjected to general forces based on chain connectivity and the non-specific short-range interactions between monomers. To achieve this, the polymer is subjected to general forces that take into account the connectivity of the chain and the non-specific short-range interactions between adjacent monomers. These forces collectively influence the folding and organization of the chromatin chain, resulting in the formation of higher order structures. However, it is important to note that while this simple model provided initial insights into chromatin behavior, more sophisticated models have been developed to model the nature of the interactions between monomers. These interactions may include DNA bridging and packing by specific structural proteins such as cohesin [287] and CTCF [288], long-range interactions between enhancers and promoters, epigenetic modifications, or interactions between the chromatin fiber and the nuclear lamina [289], which collectively contribute to the dynamic and complex nature of chromatin in the cell. Modeling such specific interactions between monomers requires to consider the mosaic properties of the chromatin. The most common way to incorporate these different properties is to assign attractive interactions between blocks with the same property.

2.3.2 Block Copolymer model

Copolymers, also known as heteropolymers, are a class of polymers in which the monomers have different physical properties (**Figure 26**). For example, nucleosomes with different histone marks or chromatin regions with different transcriptional activity can be modeled as monomers with different stiffness or interaction forces. Thus, the interaction of beads is dictated by their epigenetic properties, which induce non-mixing between monomers of different states. The formulation of this model is motivated by many experimental evidence showing that loci with different chromatin states preferentially interact with each other. For example, Polycomb group [290] or HP1 [291] have been shown to act as bridges between heterochromatin regions. Jost el al. 2014 [145] used a block copolymer model that incorporates local epigenetic information to model 3D genome interaction of *Drosophila melanogater* genome. The model considers only two types of interactions : (i) non-specific interaction between each pair of monomers and (ii) specific attraction between monomers with the same epigenetic state. The model assumes that the attractive forces is identical between monomers of the same types whatever the spanning distance and can be described by this simple equation (1). Where the V_{chain} represents the bonding potential between adjacent beads and the V_{inter} describes interactions between non-bonding monomers [292].

$$V = V_{\text{chain}} + V_{\text{inter}} \tag{1}$$

This two-parameter model was used to predict the chromatin conformation of dif-

ferent chromatin regions alternating between four types of blocks : Polycomb bound region (blue chromatin), HP1/H3K9me2-3 heterochromatin (green chromatin), active regions (red chromatin), null heterochromatin (black chromatin). Despite their simplicity, such model convincingly recapitulate large scale Hi-C contact such as A/B compartments which are observed experimentally [145, 292]. Ghosh and Jost [293] simulated the folding of the 3D organization of chromosome 3R (20 Mbp) in *Drosophila* with 5 different blocks representing different epigenetic states. This model successfully reproduces the formation of TADs and compartments observed experimentally. Similar copolymer models were used to simulate the spatial arrangement of the chromatin with three types of blocks representing euchromatin, heterochromatin and constitutive heterochromatin [294]. This model successfully recapitulated the spatial arrangement of chromatin observed by imaging techniques and suggested that heterochromatin plays a critical role in the spatial organization of chromatin. Later, [295] built a copolymer model to incorporate interactions between heterochromatin and the nuclear lamina, which successfully reproduced heterochromatin segregation and LAD formation in mammalian nuclei. In summary, the block copolymer model successfully reproduces the intricacies of chromatin folding at multiple scales, encompassing TADs, LADs, and compartments.

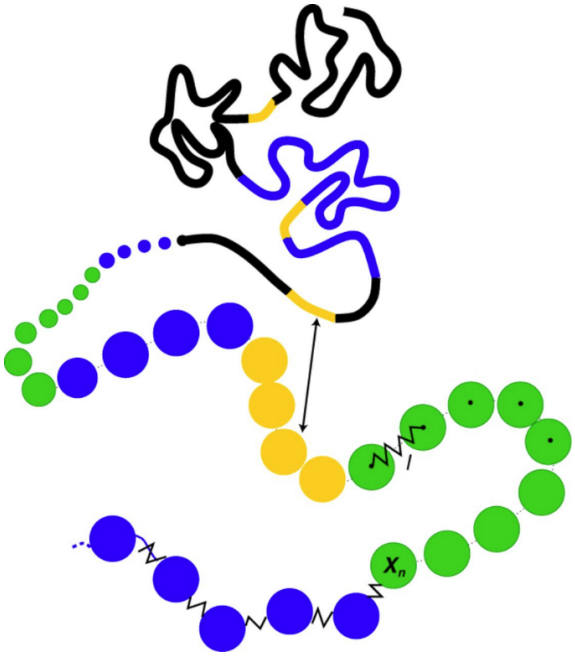


Figure 26: Block Copolymer models - Chromatin is modeled as a chain of self-avoiding beads, with each bead color-coded according to its epigenetic state. Figure adapted from [145].

2.3.3 Strings and binders switch model

A variation of the block copolymer model called to as Strings and Binders Switch modeling (SBS) allows interaction between chromatin regions occupied by specific factors (binders) [296, 297] (**Figure 27**). The model is composed of a polymer chain with binding sites to simulate the chromatin and diffuse beads called binders that could bind to binding sites on the monomers. This model can be described by the following equation (2). Where V_{backbone} is similar to the V_{chain} in (1), V_{EV} describes the repulsive interactions between pairs of beads and $V_{\text{B-S}}$ describes attractive interactions between binders and binding sites.

$$V = V_{\text{backbone}} + V_{\text{EV}} + V_{\text{B-S}} \tag{2}$$

The SBS model offers flexibility by allowing the adjustment of various parameters, such as the type and number of binding sites in the monomers, the number of binders and their relative concentrations, and the affinity between them. As a result, numerous studies have revealed different facets of chromatin structure. C. A. Brackley *et al.* [298] use the locations of DNase1 hypersensitive sites (DHSs) as a proxy for binding sites of a generic type of protein bridge, CTCF binding sites and H3K4me1 marks. With this model, which considers only 3 binders they were able to reproduce the 3D genome architecture of the α globin locus and the β globin locus. M. Barbieri *et al.* [299] established an SBS model that takes into account active genes and polycomb-repressed genes. Their model effectively reproduced DNA-FISH observations, supporting the idea that genes with similar states tend to be located in close proximity.

In summary, the SBS model combined with epigenetic data such as ChIP-seq or ATAC-seq can provide useful information on the different folding patterns of the heterochromatin and euchromatin. The SBS model has primarily been used to model fine-scale structures and has not yet been applied to larger-scale structures such as compartments.

2.3.4 Loop-extrusion models

Since the loop formation is an important features of chromatin organization and function, the effect of loop formation have been extensively studied and several models have been elaborated to explain this phenomenon. As discussed in section 1.2.1.1), loop extrusion has been proposed as a mechanism underlying the formation of loops and TADs. In this model, loop extruding factors (LEFs) such as cohesins bind to the chromatin and slide along the chromatin fiber until they encounter extrusion barriers such as properly oriented and occupied CTCF binding sites or when two LEFs meet together. The model is essentially defined by a set of four parameters : the lifetime, the velocity, the separation between LEFs and the permeability of the barriers [173]. To simulate features of chromatin folding at the scale of TADs, Fudenberg *et al.* [173] use monomers representing multiple nucleosomes to simulate a region spanning 10 to 50 Mb of chromatin. Using an optimized set of parameters, they were able to model the formation of a corner peak between

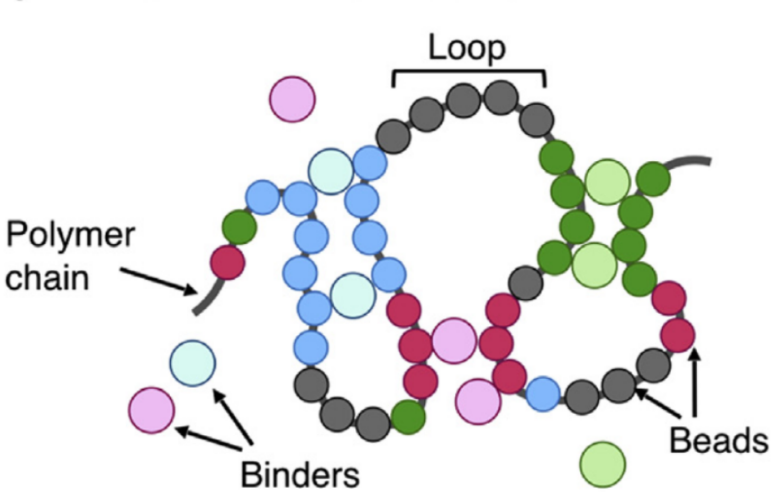


Figure 27: Strings and Binders Switch model - Chromatin is modeled as a chain of self-avoiding beads. Homotypic interactions between beads and cognate binders drive the folding of the polymer chain. Figure adapted from [300].

convergent CTCF binding sites as well as an increase of the contact probability within a TAD as observed in typical contact matrices. The model also recapitulated stripes emerging from one side of a barrier and the power-law decay of the contact probability with genomic distance. The model was also used to make experimental predictions for biological perturbations such as depletion of CTCF or cohesins. Consistent with experimental data, the deletion of site-specific barriers, such as CTCF binding sites, led to the loss of TADs and associated loops. Intriguingly, the depletion of loop extrusion factors (LEFs) not only resulted in the loss of TADs and loops but also induced a widespread decompaction of the chromatin structure. However, these models currently focus on naked DNA rather than the chromatin context. Therefore, it is still not fully understood how SMC complexes (50nm) can efficiently translocate along chromatin fibers composed of nucleosomes (10nm). The presence of nucleosomes poses significant challenge to SMC translocation, as they can potentially act as obstacles in the process.

2.4 Downstream analysis methods

2.4.1 Pattern detection

Downstream analysis of Chromosome Conformation Capture data consists mainly in extracting interaction patterns from contact maps. For example, checkerboard or plaid patterns indicate the presence of compartments, triangular patterns along the diagonal indicate the formation of TADs, while focal contacts reveal the presence of loops (**Figure 17**). However, the detection of such information often necessitates the application of normalization techniques and specific algorithms. For example,

intra-chromosomal contact frequency matrices follow power-law decay of the contact probability with genomic distance. This tendency can be easily seen on the contact maps as a decrease in contact probability when moving away from the diagonal. Consequently, regions that are close on the linear genome exhibit a higher probability of interaction than regions that are further apart. This distance-dependent decay is often removed in downstream analyses to study the long-range interactions or compartments that are often invisible in unnormalized matrices. One example is normalization using the ratio of observed versus expected counts (O/E) [1]. This approach effectively normalizes the interactions between two loci that are separated by a given distance by considering the average interaction among all loci separated by the same distance. In other words, the O/E matrix is computed by taking each $M_{i,j}$ and dividing the number of individual counts by the average number of counts for a given genomic distance $I_{i,j}$ [1].

Since chromatin loops represent the smaller functional level of 3D genome organization, it is becoming increasingly important to detect them in the contact matrix. In recent years, many algorithms have been developed including HiCExplorer [301], Mustache [302], HiCCUPS [148], SIP [303], Homer [304] and Chromosight [305]. These methods rely on efficient modeling of the background signal to assess whether two distant genomic loci have a higher frequency of interaction than expected based on the background model. The result is a list of pairs of loci that interact more frequently than expected.

However, such analyses require very deep Hi-C libraries, on the order of more than a billion Hi-C contacts [148]. Consequently, the majority of Hi-C datasets cannot be used to identify loops. However, these datasets are still valuable because they can be used to quantify the average loop strength. This involves assessing the enrichment of contacts within these loops compared to their local background [306]. To this end, Rao *et al.* developed Aggregate Peak Analysis (APA) or pile-up analysis, a method that facilitates the statistical analysis of loop interactions [148]. These methods are based on calculating the average of all areas of the distance normalized (O/E) Hi-C maps containing loops.

In addition to quantifying the strength of known features (*i.e.* loops), APA analysis can be used *de novo* to investigate whether certain regions, defined for example by the binding of a particular factor (ChIP-seq peaks), tend to interact more frequently with each other on average than would be expected by chance. This method has been successfully used to study the interaction between pluripotency factor binding sites in mESC [307]. Such a method can aid in the discovery of novel drivers of 3D genome interaction.

2.4.2 Networks

More recently, network-based analysis has gained in popularity for interpreting 3C chromatin capture experiments (**Figure 28**). In such a representation, chro-

matin can be symbolized as a network, where nodes represent chromatin regions and edges connecting two nodes indicate preferential interaction (or looping) between the corresponding regions [308]. The nodes within the chromatin network can be effectively characterized by feature vectors, facilitating the projection of various epigenetic information derived from the ChIP-seq dataset onto the corresponding node. Using such network analysis Thibodeau *et al.* found that enhancers interact more frequently with each other compared to their typical counterparts [309]. Next Pancaldi *et al.* introduced the concept of chromatin assortativity [308] to determine whether chromatin fragments (nodes) enriched in specific epigenetic features tend to interact preferentially in 3D. In network analysis, chromatin assortativity (ChAs) refers to a measure that quantifies the tendency of nodes in a chromatin network to interact preferentially with other nodes that have similar attributes. This measure indicates whether nodes with similar or dissimilar attributes are more likely to be connected. ChAs applied to a collection of over 80 epigenetic features projected onto the mESC promoter-capture Hi-C dataset highlight the pivotal role of Polycomb group proteins in facilitating promoter-mediated interactions within the chromatin network [33]. Thus, network representations are useful and complementary approaches to study 3D genome architecture.

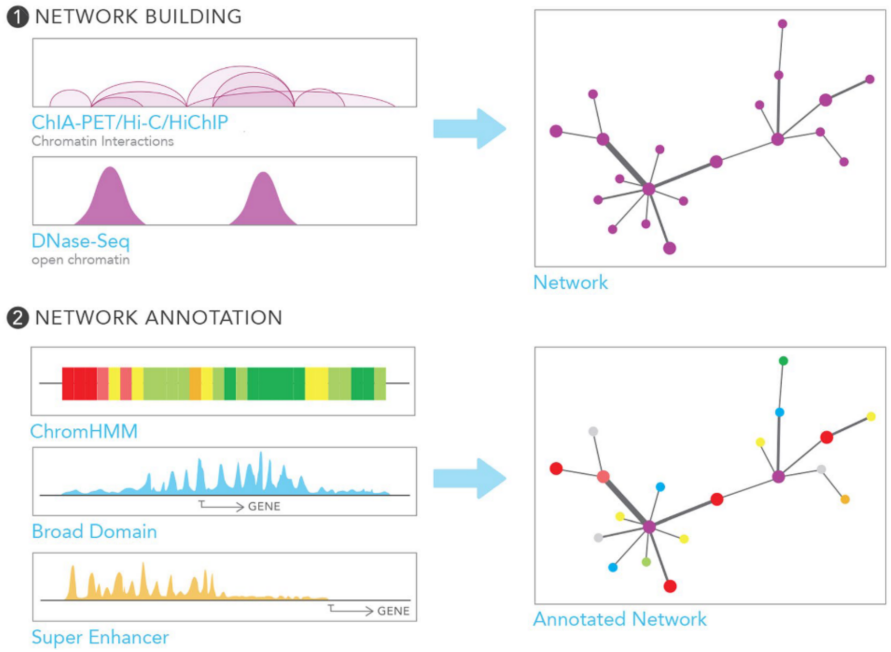


Figure 28: Chromatin network analysis - Illustration of network analysis based on chromatin interactions. Nodes in the network represent chromatin regions and edges represent interactions detected by 3C methods. The network can be annotated with different chromatin signatures to identify regulatory elements. Figure adapted from [309].

3 Thesis project

The last decade has been marked by the development of chromosome conformation analysis techniques (e.g. Hi-C). This has revealed that the eukaryotic genome is organized into physical domains called topologically associating domains (TADs). The establishment of these TAD structures occurs rapidly in the zygote during embryonic development and coincides with the establishment of its transcriptional program. Disruption of the TAD organization of the genome by altering their boundaries can lead to defects in the physical interactions between the regulatory elements they encompass, such as enhancers (E), promoters (P), and insulators. These architectural defects can lead to transcriptomic dysregulation, resulting in developmental abnormalities (e.g., polydactyly) or ectopic activation of oncogenes that contribute to many cancers (e.g., leukemia). The goal of my thesis is to gain a better understanding of the mechanisms involved in the formation of physical interactions between regulatory elements of the genome and how these interactions are altered during the onset of a disease.

To that aim my thesis was conducted with two main objectives :

First, I combined bioinformatics analysis and chromatin tracing method (Hi-M) to investigate the role of chromatin insulators in the folding of the chromosome during early embryogenesis in *Drosophila melanogaster* [see 3.1].

Second, I implemented the chromosome tracing method (Hi-M) to investigate the role of higher order chromatin structure within single cells in fully differentiated mouse tissues, both under physiological conditions and during the onset of a disease [see 3.2].

Chapter 2: Results

Since the groundbreaking publication of the first Hi-C paper [1], the use of this technique to explore the architecture of the 3D genome has revolutionized our understanding of genome organization. More importantly, recent advances in chromosome tracing techniques have further improved our ability to directly visualise the spatial arrangement of chromosomes within individual cells, together with their transcriptional status. During my PhD, my work focused mainly on studying 3D structural organisation in single cells at TAD and sub-TAD scales. My first aim was to investigate how insulator proteins contribute to the folding of the *Drosophila* genome. This aim is materialised by a research article currently accepted for publication in the journal Nature Communications. Next, in my second aim, I used mouse tissues, to shed light into the role of chromatin architecture in the regulation of tissue-specific transcriptional programmes in single cells. The results of this ongoing work are presented here as a manuscript in preparation.

3.1 Research article

3.1.1 Rationale of the work

In mammals, the CTCF protein plays a central role in coordinating loop extrusion processes, leading to the formation of well-defined topologically associating domains (TADs) within the chromatin structure. However, in contrast to mammals, domain organisation in *Drosophila* does not typically rely on CTCF-mediated looping. Instead, *Drosophila* possesses a distinct set of insulator-binding proteins (IBPs) that are not evolutionarily conserved. In this study, we aimed to investigate the role of IBPs in regulating the three-dimensional genome organisation in *Drosophila*. To this end, we used a combination of bioinformatic analysis and multiplexed chromatin imaging techniques.

3.1.2 3D chromatin interactions involving *Drosophila* insulators are infrequent but preferential and arise before TADs and transcription

3D chromatin interactions involving *Drosophila* insulators are infrequent but preferential and arise before TADs and transcription

Olivier Messina[1], Flavien Raynal[2], Julian Gurgo[1], Jean-Bernard Fiche[1], Vera Pancaldi[2,3*], Marcelo Nollmann[1,*]

Affiliations

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

[2] *CRCT, Université de Toulouse, Inserm, CNRS, Université Toulouse III-Paul Sabatier, Centre de Recherches en Cancérologie de Toulouse, Toulouse, France .*

[3] *Barcelona Supercomputing Center, Barcelona, Spain.*

Keywords : Insulators, TADs, Drosophila, Embryogenesis.

[*] Corresponding authors: vera.pancaldi@inserm.fr, marcelo.nollmann@cbs.cnrs.fr

Abstract

In mammals, insulators contribute to the regulation of loop extrusion to organize chromatin into topologically associating domains. In *Drosophila* the role of insulators in 3D genome organization is, however, under current debate. Here, we addressed this question by combining bioinformatics analysis and multiplexed chromatin imaging. We describe a class of *Drosophila* insulators enriched at regions forming preferential chromatin interactions genome-wide. Notably, most of these 3D interactions do not involve TAD borders. Multiplexed imaging shows that these interactions occur infrequently, and only rarely involve multiple genomic regions coalescing together in space in single cells. Finally, we show that non-border preferential 3D interactions enriched in this class of insulators are present before TADs and transcription during *Drosophila* development. Our results are inconsistent with insulators forming stable hubs in single cells, and instead suggest that they fine-tune existing 3D chromatin interactions, providing an additional regulatory layer for transcriptional regulation.

Introduction

Eukaryotic chromosomes are organized in a multi-layered structure comprising chromosome territories, compartments, topologically-associating domains (TADs) and nano-domains [1,2]. Notably, this multi-scale organization of the genome is conserved from *Drosophila* to mammals [3–6]. However, the mechanisms responsible for the acquisition and maintenance of these structures seem to differ between species.

In vertebrates, TADs are often separated from each other by convergent CCCTC-binding factor (CTCF) sites localized at TAD boundaries. TAD borders bound by CTCF/cohesin form "focal chromatin loops" in contact matrices [7]. These specific looping interactions may facilitate the communication between genes and their cis-regulatory elements (CREs, e.g. enhancers and promoters) most often localized within TADs [8,9]. In mammals, the formation of TADs is thought to involve loop extrusion, a mechanism by which Structure Maintenance of Chromosome (SMC) proteins (e.g. cohesin) bind chromatin and reel it in until they encounter properly-oriented CTCF sites [10]. In contrast, the *Drosophila* homologue of CTCF (dCTCF) binds preferentially within TADs and is only mildly enriched at TAD borders [4,6] with no preferential convergent orientation as in mammals [11]. These results suggest that other mechanisms may instead be responsible for the establishment of TADs in *Drosophila*.

While CTCF is the main chromatin insulator in mammals, *Drosophila* contains tens of non-evolutionary conserved insulator binding proteins (hereafter IBPs) [12]. Since before the genomic era, *Drosophila* insulators were shown to be involved in the regulation of long-range chromatin interactions: either by blocking enhancer-promoter interactions or by establishing barriers between chromatin states [13–15]. Early genome-wide studies showed that insulators preferentially bind to genomic regions containing housekeeping genes and highly transcribed regions [16]. In addition, IBPs frequently bind to TADs borders [4,6,17–19] that can often interact in 3D [20]. Taken together, these data suggest that insulators may be involved in the organization of *Drosophila* TADs.

Recent studies suggest different modes of action. On one hand, insulators may promote TAD border interactions by forming contacts between insulator factors [21–23]. On the other hand, insulators may not form CTCF-like focal chromatin loops, but rather restrict interactions between domains [24].

Here, we investigated the role of insulators in the 3D organization of the *Drosophila* genome by combining advanced bioinformatics analysis and Hi-M, an imaging-based method we recently developed to detect the 3D positions of multiple genomic loci in single cells [25]. First, we show that genomic regions occupied by insulators display preferential interactions genome-wide. These preferred interactions occur inside TADs and can also span TAD borders. Second, we show that chromatin regions displaying the most prominent 3D interactions are preferentially bound by insulators. We detect TAD border preferential interactions, but these represent the minority of the interactions detected. Interestingly, non-border interactions quantitatively increased with the occupancy of IBPs. By visualizing 3D chromatin structure at the *dpp* locus, we observed, however, that spatial colocalization between insulators is infrequent and similar to neighboring regions not bound by insulators. Finally, by mapping preferential interactions during development, we found that non-border regions harboring insulators display a tendency to preferentially interact before the emergence of TADs and transcription.

Results

Genomic regions displaying preferential interactions are predominantly bound by chromatin insulators.

To shed light onto the roles of *Drosophila* insulators in 3D genome organization during early embryogenesis, we deployed Chromatin Assortativity analysis (ChAs) [26,27]. Assortativity measures the preference for the nodes of a network to interact with other nodes that have the same characteristics. In ChAs analysis, a chromatin interaction network is built from a genome-wide contact map [28]. This network represents the genomic loci (nodes) displaying high frequency interactions amongst each other (**Fig. 1a**, see *chromatin assortativity* in Methods). Chromatin assortativity for a given factor is calculated by estimating whether nodes bound by this factor interact with other nodes with the same factor more than expected at random. Thus, a factor with positive assortativity is enriched in chromatin loci that preferentially interact.

We applied ChAs analysis to study chromatin organization of *Drosophila* embryos at nuclear cycle 14 (nc14) [20], a developmental stage coinciding with the zygotic genome activation (ZGA) and with the emergence of TADs [20]. For this, we obtained chromatin interaction networks by mapping preferentially interacting chromatin regions using Chromosight [29] on Hi-C data (**Figs. S1a-d**). Remarkably, the constructed network exhibits high overlap with previously annotated loops in the *Drosophila* embryo (**Fig. S1a**) [30]. Chromosight detects preferential chromatin interactions by segmenting the genomic regions displaying local maxima in the observed/expected Hi-C map. In mammals, loops often appear as clear focal peaks [7], however most of the Chromosight-annotated interactions from nc14 HiC data do not appear as focal peaks in the observed HiC map (**Fig. S1d**). This is consistent with many preferential contacts in *Drosophila* representing low-frequency interactions. Next, we annotated these chromatin networks with the binding patterns of publicly available ChIP-seq datasets (features, **Fig. 1a**) and calculated chromatin assortativities for a wide panel of chromatin binding factors, including insulator and insulator-associated proteins (BEAF-32, CBP, CHRO, CP190, dCTCF, DREF, FS(1)h, GAF, L(3)MBT, Pita, Mod(mdg4), Su(HW), Z4, ZIPIC and Zw5), pioneering factors (Zelda), RNA polymerase II (RNAPII CTD phospho-Ser5 : S5P), Polycomb group proteins (Pc, Ph) and the cohesin subunit (Rad21).

Chromatin assortativity Z-scores (hereafter ChAs Z-scores) are calculated to estimate if ChAs for a feature is higher than expected for regions separated by similar genomic distances, indicating the importance of 3D interactions for establishing preferential contacts. Regions enriched in Zelda, Polycomb group proteins (Pc and Ph), and RNAPII CTD phospho-Ser5 (S5P) displayed positive ChAs Z-scores (**Fig. S1e**), consistent with previous findings [31–33]. In contrast, ChAs Z-scores were highly variable between IBPs (**Fig. 1b**), indicating that different insulators may contribute unequally to the formation of preferential contacts. A sub-group of IBPs displayed high assortativities (ChAs Z-score > 2), including the insulator and insulator-associated proteins: BEAF-32, CHRO, DREF, L(3)MBT, Pita, Z4, ZIPIC and Zw5 (**Fig. 1b**). Notably, cohesin (Rad21), dCTCF, and a second sub-group of IBPs including CBP, CP190, Fs(1)h, GAF, Mod(mdg4) and SU(HW) displayed low assortativity and low Z-scores (ChAs Z-score < 2, **Figs. 1b, S1e**). To validate the robustness of these results, we performed similar analysis for different sets of Chromosight parameters (see *chromatin*

assortativity in Methods) generating larger networks that include lower-frequency interactions. ChAs Z-scores were highly correlated between networks, and the insulator factors exhibiting the highest ChAs Z-scores were the same independently of the network size or loop size distribution (**Figs. S1f-h**). For some insulators the ChAs Z-score increase was larger than proportional in the networks including longer-range contacts (e.g. GAF), while for others the ChAs Z-score increased less than proportionally (e.g. Fs1h, CTCF). This is consistent with these factors being slightly more/less assortative depending on the network loop size distribution. We note, however, that these factors still displayed the lowest assortativities in all networks.

Low assortativity scores can arise when the presence of a factor is not associated with a preferential interaction (**Fig. 1a**), or if the factor is present either in a very small or in a very large proportion of them. For instance, GAF is often bound to the anchors of focal loops clearly visible in Hi-C and micro-C datasets [30,31,33]. These focal loops, however, represent a small proportion of preferential interactions in our network (~11%, **Fig. S1a**), consistent with the low ChAs Z-scores we observed. We note that GAF binds to thousands of sites genome-wide (3842), however only a small fraction of these sites correspond to focal loop anchors (<620) [30]. Taken together, these results are consistent with only a small number of GAF binding peaks being involved in focal loops and in regulating transcriptional activation and repression [30,31,33].

Insulator binding increases the strength of preferential chromatin contacts

Next, we complemented ChAs with aggregation peak analysis (APA) [7]. This method relies on the calculation of pairwise, intra-arm autosomal contact frequencies between genomic regions bound by a given factor (i.e. peak) (**Fig. S1i**). The statistical relevance of these contacts is estimated by calculating the average of the log2(Observed/Expected) distribution of the Hi-C signal at all peak regions (*see Log2(O/E) and Hi-C aggregate plot analysis* in Methods). Thus, the log2(O/E) ratio is positive when contacts occur at frequencies higher than expected and is negative when contact frequencies are lower than expected for regions separated by the same genomic distance.

Notably, the positive correlation between ChAs and log2(O/E) (**Fig. 1c,** Supplementary Table 4) indicates that factors displaying high assortativities are bound to chromatin regions that exhibit the most preferential interactions. Remarkably, most of the insulator factors displaying positive ChAs Z-Scores also exhibited positive log2(O/E) (BEAF-32, CHRO, DREF, Z4, ZIPIC and Zw5) (hereafter referred to as Class I insulators) (**Figs. 1b, S1j**). The peaks observed for negative log2(O/E) values (referred to as *peak 2* in **Fig. S1j**) are related to longer-range contacts. Consequently, it can be inferred that Class I insulator sites exhibit a higher tendency to interact with each other at shorter distances (<250kb, **Figs. S1k-l**). Thus, Class I insulators occupy genomic regions displaying the most preferential interactions, and conversely, the genomic regions they occupy tend to preferentially interact in 3D in nc14 embryos.

Next, we investigated the specificity of preferential chromatin interactions by using Hi-C aggregate plot analysis [20] (**Fig. 1d**, see *Log2(O/E) and Hi-C aggregate plot analysis* in Methods). Class I IBPs displayed a well-defined center spot, indicating that presence of IBPs at both loop anchors reinforces preferential 3D interactions (**Fig. 1e**). Similar results were observed for Zelda and RNAPII (**Fig. S1m**). In contrast, factors with low assortativity and/or

negative log2(O/E) did not exhibit centered spots (**Figs. S1m-n**), likely due to positive and negative log2(O/E) values for different regions averaging out.

Preferential interactions captured by Chromosight are highly variable and often do not appear as focal peaks (**Fig. S1d**). We further analyzed the impact of this variability in our analysis by focusing on BEAF-32 –the insulator displaying the highest log2(O/E) ratio and ChAs Z-score– and investigated how the interaction preference depended on the number of peaks aggregated. For this, we first calculated the distribution of log2(O/E) values for different numbers of BEAF-32 peaks averaged using bootstrapping (**Fig. S1o, left panel,** see *Log2(O/E) and Hi-C aggregate plot analysis* in Methods). On average, most of the 2- and 5-peak aggregations displayed low or no preference. Nonetheless, most aggregations exhibited positive log(O/E) values when 25 or more BEAF-32-bound regions were averaged. Overall, these results indicate that interactions between different BEAF-32 anchors are highly variable and often display low or no preference. In support of these conclusions, well-centered peaks in Hi-C aggregate analysis were observed only after a sufficient number of BEAF-32-bound regions were aggregated (**Fig. S1o**, right panel). All in all, these analyses agree with our previous observations (**Fig. S1d**), and suggest that interactions between insulator-bound genomic regions are on average preferential, but highly variable and often weak.

To investigate whether IBPs act together to promote preferential chromatin interactions, we employed Cross-ChAs and AND-ChAs [27]. Cross-ChAs measures assortativity of two different proteins, giving information about frequency of interactions joining fragments with one protein on either side. Instead, AND-ChAs measures assortativity of two different proteins considering that connected nodes are bound by a pair of factors, and therefore provides information about interaction frequencies of co-occupied regions. We computed Cross-ChAs and AND-ChAs Z-Scores for each pair of factors investigated previously (**Figs. S1p-q**). Cross-ChAs shows that class I insulators (BEAF-32, Chromator, Z4, PolII, Zelda, L3(MBT), DREF) tend to display high cross-assortativities, suggesting that anchors bound by either of these factors tend to preferentially interact. AND-ChAs shows that DNA fragments containing colocalized class I insulators (BEAF-32, Chromator, Z4, PolII, Zelda, L3(MBT), ZIPIC) interact preferentially with each other. Thus, pairs of class I IBPs can be found at each anchor of strong loops. These results are consistent with Class I IBPs often interacting together to promote formation of preferential chromatin contacts in nc14 embryos.

Most insulator-bound preferential contacts involve non-border chromatin regions

TAD borders in *Drosophila* are mostly occupied by insulators [4,6], with only 4% of borders lacking insulator sequences [19]. In *Drosophila*, ensemble analysis showed that adjacent TAD borders tend to preferentially interact [20], however contact maps do not display focal chromatin peaks as those observed in mammals [7,34,35]. To determine whether our unbiased network analysis was able to recover preferential interactions between TAD borders, we calculated whether TAD borders were assortative in the network of chromatin interactions generated by Chromosight for nc14 embryos. This analysis shows that TAD borders appear highly connected to each other in the interaction network (**Fig. 2a, S2a**). This connectivity results in considerably higher ChAs values for TADs borders (**Fig. 2b**, blue dashed line) as compared to randomized networks (**Fig. 2b**, black distribution). As expected, and consistent with

previous analysis [20], TAD borders exhibited a well-centered interaction spot in Hi-C aggregate plots (**Fig. S2b**).

The strength of a TAD border, as assessed by its insulation score (IS), is positively correlated to the binding level of insulator proteins [19,36]. Given this correlation, we tested if the presence of IBPs at TAD borders is also associated with their interaction preference by stratifying TAD borders into five equally-sized categories according to their IS and by computing Hi-C aggregate plots for each category. Notably, the level of preferential interactions between TAD borders increased with insulation strength (**Fig. 2c**), providing indirect evidence for a role of IBPs in contributing to TAD border interactions.

Next, we wondered whether preferential chromatin contacts may be detected in locations other than TAD boundaries and what their determinants may be. To this aim, we divided the interactions in our chromatin network into three categories: border/border (blue), border/non-border (red), and non-border/non-border (black) (**Fig. 2d**). Then, we quantified the occurrence of each type of interaction by quantifying the presence of a border on each loop anchor. Notably, preferential interactions involved a border in one or both of their anchors in a minority of cases (<1% for border/border and ~6% for border/non-border), with the overwhelming majority of preferential interactions involving non-borders (>93%) (**Fig. 2e**).

To better understand the role of insulators in each of these interaction categories (i.e. border/border, border/non-border and non-border/non-border), for each category we calculated the proportion of interactions displaying binding of Class I IBPs in two, one or none of the anchors. The vast majority of the anchors for all categories were bound by at least one Class I IBPs (>92%, **Fig. 2e**, right pie charts). Anchors in border-border interactions are most often bound by two Class I IBPs (~90%), and in a smaller proportion by a single class I IBP. This trend was similar for the other categories, further supporting a role of class I IBPs in the mediation of chromatin loops that in most cases do not involve TAD borders.

The number of loop anchors corresponding to TAD borders is considerably larger than the number of non-borders. Thus, we estimated the probability with which a border may take part in a loop by calculating the proportion of borders participating in loops (either in one or both anchors). We found that ~38% of borders take part in loops in our Chromosight network (Fig. S2c), with the majority of them participating as a single anchor (~36.8%) (Fig. S2d). Next, we calculated similar statistics for Class I IBPs non overlapping with borders. Notably, we found that the propensity of non-border IBP peaks to form loops was always lower than that of TAD borders (Figs. S2c-d). Overall, these results are consistent with Class I IBPs binding at loci displaying preferential looping, at both border and non-border regions.

To further support this conclusion, we performed aggregation Hi-C analysis on non-border/non-border regions occupied by Class I IBPs. Notably, this analysis displays a clear peak (**Fig. 2f**), suggesting preferential interactions between anchors containing Class I IBP sites. Consistently, interactions mediated by Class I IBPs at non-border regions increased with ChIP intensity (binding strength) (**Figs. 2g, S2e**) of both anchors (**Figs. 2h, S2f**). All in all, these analyses suggest that class I insulators participate in mediating preferential interactions between border and non-border chromatin regions. These analyses, however, do not inform us on how frequently these preferential interactions occur in single cells, or whether they involve two or multiple anchors.

Insulator-bound chromatin regions only infrequently co-localize in 3D

Sequencing-based 3C methods only provide relative interaction frequencies, thus we turned to DNA-FISH, a technique that can quantify absolute physical proximity frequencies [37,38]. As conventional DNA-FISH can only measure proximity between a limited number of genomic targets, we used Hi-M, a multiplexed imaging method that enables the detection of tens of genomic loci at once [25,32,39]. Specifically, we imaged the 3D chromatin organization of the *dpp* locus (*chr2L: 2343645-2758688* dm6) in intact nc14 *Drosophila* embryos at ~12 kb resolution (**Figs. 3a**). The *dpp* locus contains three TADs, multiple preferential loops (**Fig. S3a**), and several regions displaying high levels of class I insulator binding, named barcode I1 to I10 (**Fig. 3b**). To cover this locus, we designed 34 equally-spaced barcodes that label insulator-bound and insulator-free genomic regions (**Fig. 3c**). Nuclei and barcodes were registered, segmented and localized as in previous studies [25,32] (**Fig. S3b**, see *Image processing* in Methods), with similar barcode detection efficiencies (**Figs. S3c-d**). Ensemble pairwise distance maps were built by calculating the median of the full pairwise distance (PWD) distributions (**Fig. S3e**). Proximity maps were constructed by calculating the frequency of co-localization for each pair of barcodes from chromatin traces (**Fig. S3f**) using a pre-established distance threshold that maximizes the correlation between Hi-M and Hi-C datasets (d = 200 nm, **Figs. S3g-h**), and that was previously used for similar studies [32]. The number of traces acquired was sufficient to ensure a statistically representative ensemble map (**Fig. S3i,** see *Image processing* in Methods).

The proximity and PWD distance maps revealed multiple regions displaying preferential 3D spatial proximity (**Fig. 3c**). These mostly corresponded to the TADs called from Hi-C data (Figs. 3b-c, blue arrows) and from Hi-M proximity frequency maps (Fig. 3c, insulation score, and domainogram). We note that TAD3 is more insulated than the other two TADs in this region, and that it is flanked by multiple IBP peaks. This is consistent with the role of IBPs in TAD insulation. To quantify the frequency at which insulator-bound regions spatially co-localized in a population of single cells, we calculated the cumulative average proximity frequencies between insulator-bound regions and control regions for different cutoff distances (**Fig. S3j**). At the cutoff distance used to calculate proximity maps (200 nm), insulator barcodes co-localized on average only in a small fraction of cells (~12.19%, **Fig. S3j**, green curve and inset). As expected, the proximity frequency monotonously increased with cutoff distance, but remained low for cutoff distance thresholds used in this and other studies (<200 nm) [25,32,40]. Thus, we conclude that the average colocalization between insulator-bound regions within and between TADs is rather infrequent, consistent with colocalization of insulator barcodes occurring only in a small proportion of cells (i.e. large cell-to-cell heterogeneity) or/and with colocalization being highly dynamic.

Next, we investigated the specificity of insulator barcode co-localizations by calculating the proximity frequency versus cutoff distance curve for non-insulator (control) barcodes located at similar genomic distances (**Fig. S3j**, black curve). For this, we averaged 10 sets of control barcodes. At a cutoff distance of 200 nm, control barcodes co-localized at similar frequencies than insulator barcodes (10.8% and 12.19%, respectively). Next, we calculated how proximity frequency depended on genomic distance for both insulator and control barcodes, using a fixed cutoff distance of 200 nm (Fig. 3d). This analysis revealed that, at least at the *dpp* locus, barcodes co-localize at similar frequencies irrespective of whether they contain insulators. Proximity frequencies dropped with genomic distance, as expected, but the difference between insulator and non-insulator barcodes remained small for all genomic

distances. We note that use of larger cutoff distances increases the proximity frequency, but this would happen for both insulator and non-insulator barcodes. Overall, these results show that insulators coalesce in space infrequently, and only at slightly higher frequencies than non-insulator regions.

Insulator barcodes most frequently co-localize in pairs

The existence of multiple focal peaks in the Hi-M matrix can be explained by two different models. On one hand, a low fraction of single cells can form rosette-like structures where multiple insulator-bound regions come together in space at once, as suggested by previous models [41]. On the other hand, different combinations of insulator barcodes may co-localize at low-frequencies in a pairwise manner in single cells. In this case, the multiplicity of peaks in the Hi-M matrix would arise from ensemble averaging. To discern between these two models, we calculated how often insulator barcodes were proximal (i.e. at a distance ≤ 200 nm) to any other insulator barcode in single cells. This frequency was comparable for all the insulator barcodes investigated, and on average lower than 12% (**Fig. 3e**). Thus, in single cells, insulator barcodes interact with any other (genomically close) insulator barcode at low frequency.

Finally, to explore if these rare spatial encounters involved multiple insulator-bound regions, we calculated the proportion of clusters containing two (i.e. pairwise cluster) or multiple insulator barcodes (multiway cluster). Clusters containing only two insulator targets were the most common in all cases (>65%) (**Fig. 3f**). Next, we calculated the frequency of multiway clusters as a function of the number of barcodes in a cluster for all barcodes combined (**Fig. 3f**) or for each barcode independently at a distance ≤ 200 nm (**Fig. 3g**) and for different distance thresholds (**Fig. S3k**). We note that at larger cutoff distances (e.g. 400 nm) multiple barcodes can frequently coalesce in space, but we don't consider these to represent multiway clusters because of the large distances involved. The frequency of multiway clusters rapidly decreased with the number of co-localizing targets but was still slightly higher than what would be expected by chance (**Fig. 3f,** see *Multiway proximity frequency analysis* Methods). All in all, these results indicate that insulator-bound regions rarely form clusters with more than two insulators, and when they do, they contain only a very limited number of insulator-bound regions.

Preferential interactions between class I insulators arise before TADs and transcription

Previous studies showed that preferential spatial proximity between *cis*-regulatory elements (e.g. enhancers and promoters) can occur before nc14 [32], the nuclear cycle at which most zygotic genes get activated and when TADs first emerge [20]. We reasoned that insulators may display similar features. To test this hypothesis, we first analyzed the changes in accessibility of class I IBP sites at different time points within nc12-13 [42]. Surprisingly, we found that many of these sites are accessible as early as nc12, with a progressive acquisition of accessibility with time within this restricted time window (**Figs. 4a, S4a**).

To investigate whether these accessible insulator sites preferentially interacted before nc14, we performed APA analysis for nc12/nc13 (pre-ZGA), nc14 (ZGA) and 3-4 hours post fertilization (hpf) (post-ZGA). As expected, preferential interactions between TAD borders first appear at nc14 and are sustained thereafter (**Fig. 4b**), consistent with previous analyses [20].

Thus, interactions between insulator-bound regions occupying TAD borders arise at the same time as TADs.

Next, we quantified the timing at which preferential interactions between non-border, insulator-bound regions emerged. For this, we performed APA analysis for non-borders for different developmental timings. Notably, we found that preferential interactions between non-border regions bound by insulators were already present in nc12 embryos for most Class I IBPs (**Figs. 4c, 4d**). We note, however, that further studies will be required to fully establish whether these sites are actually bound by Class I IBPs at these early stages of development.

To determine if interactions between Class I IBPs occurred at similar or reduced frequencies before nc14, we performed Hi-M imaging at nc12. The overall structure of the *dpp* locus displayed relatively minor changes between these two nuclear cycles (**Figs. 4e-f**), which agree with those expected from the emergence of TADs at nc14 (**Figs. 4f and S4b-d**). To better dissect how the proximity between insulator barcodes changed between nc12 and nc14, we calculated the proximity frequency versus cutoff distance curves for insulator and control regions (**Fig. 4g**). This analysis reveals that Class I IBPs co-localize with each other with similarly low frequencies in nc12 and nc14 embryos (12.83% vs 12.19% respectively). Thus, preferential interactions between Class I IBPs can be detected before the ZGA, but they occur at low frequencies.

To further investigate the origin of these weak interactions, we performed APA analysis from nc14 embryos treated with triptolide and alpha-amanitin, two small-molecule inhibitors of RNA Pol II activity [20]. Notably, preferential interactions between non-borders increased under these chemical perturbations (**Figs. 4i, S4e**). In contrast, interactions between TAD borders were relatively undisturbed (**Fig. 4h**). The increase in interactions between non-border insulator-bound regions is consistent with enhanced inter-TAD interactions [20] (**Fig. S4f**). As RNA Pol II activity in these embryos is inhibited before they are transcriptionally active, our result indicates that preferential interactions between non-border, insulator-bound regions do not require active transcription.

Finally, to shed light onto the mechanism of preferential interactions between non-border IBP sites, we performed APA analysis on embryos depleted in Zelda, a pioneering factor involved in establishing early accessibility of *cis*-regulatory elements [43]. Surprisingly, preferential interactions between non-border IBP sites were overall unaffected in Zelda-depleted embryos (**Figs. 4i, S4e**), suggesting that binding of class I insulators to non-border regions may not require chromatin opening by Zelda. To test this hypothesis, we first calculated the fraction of class I IBP binding sites overlapping with Zelda sites. This analysis revealed that only ~14% of the class I IBP sites corresponded to Zelda sites (**Fig. 4j**). Next, we calculated the accessibility of class I IBP sites at nc14 for all sites and for two subclasses: sites not bound by Zelda, and sites also bound by Zelda (**Fig. 4k-l)**. Sites displaying both Class I IBPs and Zelda binding exhibited high accessibility, as expected. Notably, accessibility of Class I IBP sites not overlapping with Zelda represented the majority of sites and displayed significant accessibility. Overall, these results explain why preferential contacts between Class I insulators are not affected by Zelda depletion, and suggest that this class of insulators rely on other means to access chromatin during early embryogenesis.

Discussion

In this study, we applied bioinformatic analysis to investigate the role of *Drosophila* insulator binding proteins in the folding of the zygotic genome during early embryogenesis, and combined it with novel imaging-based chromosome conformation capture approaches to quantify the absolute frequency and specificity of pairwise and multiway chromatin interactions involving insulators.

Drosophila insulator proteins are highly enriched at TADs borders [4,6,19,21] and contribute to the insulation of TADs [44–46]. Our bioinformatics analysis reveals that most preferential chromatin interactions genome-wide involve regions bound by class I insulators that do not involve TAD borders (>90%). This finding suggests that class I insulators are likely also involved in modulating interactions within TADs and across TAD boundaries. Members of the class I insulator group (e.g. BEAF-32) tend to co-localize with promoter regions [47,48] and tend to demarcate differentially-expressed genes [49], suggesting that class I insulators may play a role in modulating contacts between *cis*-regulatory modules within and between TADs. Direct promoter regulation and reduction in TAD insulation can only account for a minority (20%) of the genes downregulated upon depletion of BEAF-32 [45]. Non-border chromatin interactions by Class I IBPs appear before TADs and the onset of zygotic transition, suggesting that they may contribute to defining pre-established topologies to demarcate *cis*-regulatory networks. *Drosophila* homologous chromosomes are often paired, and several factors, including insulators, play a role in this process [50], therefore contacts between insulators bound to different homologous chromosomes could also contribute to *cis*-regulation [51].

Zelda plays a central role in rendering the zygotic genome accessible [52–55]. However, we found that interactions between Class I IBPs at non-border regions are not affected by the depletion of Zelda. This surprising result may be explained by our finding that a significant portion of class I IBPs peaks (~90%) are open at early developmental cycles (e.g. nc12) but do not colocalize with Zelda, suggesting that other unidentified pioneering factors may be required to provide access to most Class I IBPs.

Despite the genome-wide enrichment of IBPs at regions displaying 3D preferential interactions, the quantification of absolute proximity frequencies using Hi-M shows that insulator-bound regions (borders and non-borders) physically co-localize in space infrequently (~12%), and marginally more frequently than neighboring genomic regions (10.8%). This observation is consistent with low proximity frequencies between TAD borders measured in S2 cells (~10%) [37]. The low proximity frequencies between insulator-enriched regions are consistent with a recent study showing that depletion of insulators only partially weakens the strength of TAD borders [45], and with the overall absence of “focal loops" involving class I insulators in Hi-C contact maps [4,6,20,31,56]. Finally, our genome-wide analysis shows that interactions between insulator-bound regions are on average preferential, but highly variable and often weak.

The early discovery of insulator bodies led to the proposal that insulators mediate the formation of stable, rosette-like hubs involving multiple insulator-bound genomic regions [57–60]. More recently, it was shown that CP190 and Su(HW) insulator bodies formed in cultured-cells under stress conditions exhibit liquid-liquid phase separation properties[61]. This model predicts that genomically-close insulators should interact in space often, nucleating interactions between multiple partners. In contrast, we observed low-frequencies of pairwise proximities that rapidly decrease with the number of interacting partners (<5 % for 3-way interactions and

<1 % for 4-way interaction). Therefore, these results do not provide support for a widespread role of stable insulator hubs or LLPS-mediated insulator bodies in the 3D organization of the *Drosophila* genome, at least in normal physiological conditions at the *dpp* locus.

Previous studies proposed a role for *Drosophila* IBPs in mediating distant interactions [13–15,62]. Our genome-wide analysis and imaging data are inconsistent with stable interactions between class I IBPs, and suggest that these insulators may play a role at stabilizing 3D distant chromatin conformations arising from other processes, including polymer dynamics [63,64]. It is well established that binding peaks from multiple insulators often cluster together [16,49]. In this scenario, combinatorial binding of multiple insulator binding sites at single genomic locations [19,45] would provide a means to modulate the strength of the stabilization, to regulate its specificity, and to enable a locus to time-share 3D interactions with multiple genomic locations in an asynchronous manner. Consistent with this concept, analyzing binding of RNAPII and polycomb members in mouse embryonic stem cell promoter-centered chromatin interactions using network measures such as bridgeness and betweenness centrality, it was suggested that RNAPII-bound chromatin fragments would belong to multiple communities at once, whereas polycomb bound fragments appeared to participate in multiple interactions at once [26].

Direct measurements of residence times have, unfortunately, not been reported for class I *Drosophila* insulators. However, recent studies showed that GAF and mammalian CTCF can remain bound to their cognate chromatin sites for minutes [65,66], and that CTCF loops are dynamic [67,68]. These data are consistent with a model whereby insulators help modulate the dynamics of specific interactions between distant cis-regulatory regions, but do not form stable scaffolds. These transient structures, however, may be more stable than the typical residence time of transcription factors (~10 seconds) [69]. In this picture, insulators could help promote transcription by stabilizing transient *cis*-regulatory interactions to allow for the rapid binding and unbinding of transcription factors, or rather contribute to transcriptional repression by promoting 3D conformations that prevent functional interactions. This said, the lack of clear focal peaks, the high variability in interaction strength genome-wide, and the low proximity frequencies between class I insulator-bound regions, argue for the involvement of additional molecular actors in the 3D regulation of transcription.

Finally, the methods used in this manuscript to show that *Drosophila* insulators only moderately increase the frequency of border and non-border chromatin interactions may be used to investigate insulator mechanisms in other organisms.

Methods

Drosophila stocks and embryo collection.

The *yw* fly stocks were maintained either in a 21°C room or in a 25°C incubator with a natural light-dark circadian cycle. Following a pre-laying period of 16-18 h in cages with yeasted apple juice agar plates, flies were allowed to lay eggs during 1.5 h on new plates. Layed embryos were then incubated at 25°C for an extra 2.5 h to reach the desired developmental stage. Embryos were collected and fixed as previously described [39]. Briefly, embryos were dechorionated with 2.6% freshly opened bleach for 5 min and thoroughly rinsed with water. Then, embryos were fixed in 10mL of a 1:1 mixture of fixation buffer (4% methanol-free formaldehyde in PBS and heptane). They were then agitating for 25 min at RT. The bottom formaldehyde layer was replaced by 5mL of methanol and embryos were vortexed for at least

30 s. Embryos that sank to the bottom of the tube, devitellinized, were rinsed three times with methanol. Embryos were then stored in methanol at -20°C until further use.

Hi-M libraries.

Oligopaint libraries were constructed as in previous studies [25,32,39]. Briefly, each oligo had an homology region of 35-41 nt followed by a flap encoding a sequence complementary to the readout probes. We selected 138 genomic regions of interest (barcodes) in the The *dpp* locus (2L:2343645..2758688 BDGP *Release 6* + ISO1 MT/*dm6*). For each barcode we used ~50 probes, covering ~3 kb. The coordinates of the targeted genomic regions are listed in Supplementary Data 2. Each oligonucleotide in the pool (CustomArray) consisted of 5 regions: (i) a 21-mer forward primer region; (ii) two 20-mers separated by an A sequence for the barcoding; (iii) a 35/45-mer genome homology region; (iv) an extra 20-mer readout region for barcoding; and (v) a 21-mer reverse priming region. The designed oligonucleotide pools were ordered from CustomArray. The procedure to amplify a given library from the pool was previously described [39]. Briefly, the seven-step strategy consist of (i) emulsion PCR (emPCR) to extract the desired library from the pool using specific couple of primer; (ii) limited-cycle PCR from the emPCR product to determine the optimal amplification cycle; (iii) large-scale PCR with T7 promoter on the reverse primer; (iv) in-vitro transcription using T7 RNA polymerase; (v) reverse transcription; (vi) alkaline hydrolysis; and (vii) purification and concentration of the ssDNA. The sequences of the primers used for amplification of the library are listed in Supplementary Table 3.

For imaging, we used a combination of 4 barcodes to cover ~12kb, the list of positions of the barcodes are listed in Supplementary Table 2. Each adapter consists of a 20-mer region complementary to the readout sequence that can recognize the barcode bind to a unique Alexa Fluor-647-labeled oligonucleotide (containing a disulfide linkage). Between each cycle, the fluorophore attached via a disulfide linkage can be cleavable by the mild reducing agent tris(2-carboxyethyl)phosphine (TCEP), as previously described here [39]. For fiducial, we used an adapter complementary to the reverse primer that can be bound by an unique Atto 550 labeled oligonucleotide. The sequences of the adapters and labeled barcodes purchased from Integrated DNA Technology (IDT) are listed in Supplementary Data 1.

Hybridization of Hi-M primary library.

The ssDNA library is hybridized to the DNA as previously described [39]. Briefly, embryos were rehydrated and permeabilized by sequential dilution of methanol with 0.1% Tween-20 PBS (PBT) : 90%MeOH; 70%MeOH; 50%MeOH; 30%MeOH; 100%PBT (5min each). Embryos were RNase A treated during 2h, permeabilized 1h with 0.5% Triton in PBS and rinsed with increased concentration of Triton/pHM buffer. pHM (pHM = 2X SSC, NaH_2PO_4 0.1M pH = 7, 0.1% Tween-20, 50% formamide (v/v)) : 20%pHM; 50%pHM; 80%pHM ; 100%pHM (20 min each). Then, 225 pmols of ssDNA were diluted in 25µL of Fish Hybridization Buffer (FHB = 50% Formamide, 10% dextran sulfate, 2X SSC, Salmon Sperm DNA 0.5 mg/mL). The ssDNA and embryos were preheated at 80°C during 15 minutes in separated tubes. The supernatant of the embryo's tube (pHM) is removed and the 25µL of FHB containing the ssDNA is added. Next the mixture is transferred in a PCR-tube and deposited in the thermomixer set at 80°C. Immediately, the thermomixer is set to decrease to 0.1°C/min until it reaches 37°C for an overnight incubation. The next day, the embryos were transferred to a new 1.5 mL eppendorf tube and washed two times at 37°C during 20 min with 50%

formamide, 2X SSC. Next, embryos were sequentially washed at 37°C for 20 min with serial dilutions of formamide/PBT: 50% formamide / 2xSSC; 40% formamide / 2xSSC; 30% formamide / 70% PBT; 20% formamide / 80% PBT; 20% formamide / 80% PBT; 10% formamide / 90% PBT; 100% PBT. An additional crosslink step with PFA 4% was performed and labelled embryos were washed, resuspended in PBS and stored at -20°C for months until further use.

Imaging system.

Experiments were performed on a home-made imaging setup built on a RAMM modular microscope system (Applied Scientific Instrumentation) coupled to an improved microfluidic device, as the one described previously [39]. Software-controlled microscope components, including camera, stages, lasers, needles, pump and valves, were run using Qudi-HiM, an homemade software developed in python [70] (RRID, record ID: SCR_022114). Embryos were imaged using an ×60 Plan-Achromat water-immersion objective (numerical aperture=1.2; Nikon) mounted on a closed-loop piezoelectric stage (Nano-F100, Mad City Labs Inc.). The Illumination was provided by three lasers (OBIS-405nm nm and Sapphire-LP-561nm from Coherent and VFL-0-1000-642-OEM1 from MPB communications Inc.) and the images were acquired using an sCMOS camera (ORCA Flash 4.0V3, Hamamatsu, Japan). A homemade autofocus system was used to correct for axial drift in real time using a 785nm laser (OBIS-785nm from Coherent).

Acquisition of Hi-M datasets.

Embryos were aligned on a 2% agar:PBS pad, attached to a 1:10 poly(L-lysine):water coated coverslip and mounted into a FCS2® flow chamber (Bioptechs, USA). ~20-30 embryos were selected and imaged using two regions of interest (ROI 200x200µm²). . Then, a mixture containing the fiducial adapter (25nM Atto-550 imager probe, 25nM of adapter to the reverse primer, 2× SSC, 40% v:v formamide) was injected in the chamber and let incubate for 15 min to allow complete hybridization on the primary FISH library. Embryos were washed for 10 min with a washing buffer solution (2× SSC, 40% v:v formamide) and for 5 min with 2×SSC before injecting 0.5 µg.ml−1 of DAPI in PBS to stain nuclei. Prior to imaging, the imaging buffer (1x PBS, 5% w:v glucose, 0.5mg/ml of glucose oxidase and 0.05mg/ml of catalase) was injected to reduce photobleaching of the fiducial barcode. A stack of images was acquired for DAPI and the fiducial tagged with Atto550 (z-step size of 200 nm and a total range of 20 µm) using 405 nm and 561 nm sequential illumination. Next, the sample was sequentially hybridized as follows. A solution containing the barcode and the imager oligo was injected (25nM Alexa-SS-647 probe, 25nM barcode, 2× SSC, 40% v:v formamide) and incubated for 15 min. Then, the embryos were washed with 1.5mL of washing buffer and with 1.5 mL of 2x SSC before injecting the imaging buffer. In each cycle, fiducials and readout probes were sequentially imaged with 561 nm and 647 nm excitation lasers. After imaging, the fluorescent tag of the readout probes was cleaved and discarded using 1 mL of chemical bleaching buffer (2× SCC, 50mM TCEP hydrochloride). Finally, samples were washed with 1 mL of 2× SSC for 5 min before a new hybridization cycle started. Further details can be found on our previously published protocol [39].

Image processing.

DCIMG files were converted to TIFF using proprietary software from Hamamatsu. TIFF images were then deconvolved using Huygens Professional 21.04 (Scientific Volume

Imaging, https://svi.nl). The analysis was performed using our pyHiM analysis pipeline (https://pyhim.readthedocs.io/en/latest/). Briefly, images were first z-projected using either sum (DAPI channel) or maximum intensity projections (barcodes, fiducials). Fiducial images from each hybridization cycle were used to register barcode images using global and local registration methods. Next, barcode images were segmented in 3D using stardist [71] and the positions of the centers of barcodes were detected with subpixel resolution using Big-FISH (https://github.com/fish-quant/big-fish) [72]. DAPI images were segmented in 3D using stardist. Barcodes were then attributed to each single nucleus mask by using their XY coordinates. Finally, pairwise distance matrices were calculated for each single nucleus. From the list of pairwise distance maps, we calculated the proximity frequencies as the number of nuclei in which pairwise distances were within 200 nm normalized by the number of nuclei containing both barcodes. Hi-M maps of nc14 embryos were generated from a total of 23531 traces from 22 embryos from 2 separate experiments. The maps for nc12 embryos are constructed from 1792 traces from 4 embryos from 2 separate experiments.

Insulation score derived Hi-M dataset

Insulation scores derived from the Hi-M dataset were computed by moving an n-by-n square window along the diagonal of the median pairwise distance and summing the distances within this square. Domainogram were calculated by smoothing a matrix obtained by computing the IS with an increased window size (from 1-by-1 to 6-by-6) over the Hi-M matrix.

Multiway proximity frequency analysis.

The proportion of multiway contacts is calculated from single nucleus proximity frequency nc14 matrices [73]. Briefly, we counted the number of multiway contacts where the selected anchor barcode was interacting with other partners within a 200 nm radius. These values were normalized by the number of pairwise interactions for each anchor. The expected proximity frequency is derived by considering all events as independent. For this, we computed the mean of the product of all possible barcode combinations for various numbers of interacting partners.

Chip-Seq data processing

Insulator proteins ChIP-Seq fastq files were downloaded from Gene Expression Omnibus (GEO) with GSE62904 and GSE54337 primary accession numbers. The quality of the reads was estimated with FastQC [74] (0.11.7). Sequencing reads were aligned to the reference *Drosophila melanogaster* genome assembly (dm6) using Burrows-Wheeler Aligner [75] (0.7.17-r1188) with default parameters. Finally, peak calling was performed using MACS2 [76] (2.2.7.1) with default parameters. BEAF-32, ZELDA, Zw5, PolIISer5 and Pc / Ph raw data were downloaded from GEO series accession code GSE62904, GSE30757, GSE76997, GSE62925, GSE60428 respectively, and processed as previously described. ChIP-on-Chip Insulator proteins data from GSE26905 GEO serie have been downloaded as bed files and peak coordinates have been converted from dm3 to dm6 by using FlyBase's sequence coordinates converter (FB2021_04, released August 17, 2021). The accession number of the data used in this study are listed in Supplementary Table 1.

ATAC-Seq data processing

ATAC-seq data were downloaded from GSE83851 [42]. *Wig* files were converted to *BigWig* using *wigToBigWig* from UCSC. Heatmaps of ATAC-seq profiles were then plotted over +/- 1kb window centered on Class I IBPs sites using *computeMatrix* followed by *plotProfiles* from *deepTools* [77]. Average ATAC-seq profiles derived from the heatmaps for individual IBPs were constructed using a custom Matlab script. Venn diagrams between Class I IBPs and Zelda peaks were generated from bed files using *Intervene* [78] and plotted using a custom python script. Different bed files coming from each group were generated using *intersect* and *subtractBed* from bedtools v.2.3.

Boundary calling

We used the previously annotated list of TAD boundaries from Hug *et al*. [20]. Briefly, boundaries were called using the insulation score metric defined by Crane *et al.* [79] using a 5kb balanced contact matrix with a window size of 8 bins.

Chromatin Assortativity

In order to build networks needed for Chromatin Assortativity, Hi-C contact matrices were used with a 5 kb resolution. Chromosight [29] (1.3.3) was used with different sets of parameters to create different networks. Network 1 was built with the following parameter set: --pearson 0.3, --min-dist 20kb, --max-dist 2Mb, --min-sep 5kb, --max_perc_0 10. Network 2 was built using: --pearson 0.3, --min-dist 10kb, --max-dist 200Mb, --min-sep 5kb, --max_perc_0 50. Network 3 was built using: --pearson 0.2, --min-dist 10kb, --max-dist 200Mb, --min-sep 5kb, --max_perc_0 50. In all cases, Chromosight was used with the "--norm" parameter set to "auto" to instruct Chromosight to use matrices normalized using the Knight-Ruiz balancing algorithm [80]. As a pre-processing step, Chromosight normalizes HiC matrices by genomic distance using observed/expected values. Significant chromatin loops are called on these normalized matrices. A specific genome scale chromatin network was built where nodes are genomic fragments and edges are significant interactions between two fragments. Then, chromatin networks were loaded on R and ChIP-seq peaks were used as features assigned to nodes using the ChAseR R package [81] (0.0.0.9) to calculate chromatin assortativity. For each feature, 1000 randomized networks preserving genomic distances and corresponding chromatin assortativity values were computed.

The chromatin assortativity Z-Scores calculation is given by the following formula :

$$z = \frac{X - \mu}{\sigma}$$

where X is the feature chromatin assortativity value, μ is the randomizations ChAs average and σ is the standard deviation from the randomization distribution.

The calculation of the Z-score allows estimating the significance of the assortativity values with respect to the assortativity expected based purely on correlation of feature values along the linear genome. For example, domains of a feature that span multiple bins in the Hi-C matrix are more likely to produce high ChAs values, but this does not imply the importance of the 3D contacts.

Cytoscape (3.8.0) was used for chromatin network visualization. Methods for Cross-ChAs and AND-ChAs are provided elsewhere [27].

Log2(O/E) and Hi-C aggregate plot analysis

ChIP-seq peaks were used to extract a list of regions bound by a set of putative factors. Next, for each autosomal arm, we computed the log2(Observed/Expected) Hi-C contact value by calculating the average contact frequency for all the combinations of regions separated by a certain genomic distance on a 5kb-Hi-C dataset [20] (E-MTAB-4918). We then selected the interaction between the regions bound by the set of factors on the distance normalized Hi-C dataset. The distribution of log2(O/E) between all pairwise combinations of regions bound by the investigated factors was then displayed in a violin plot or divided into equal-size groups depending either on protein occupancy or insulation score and displayed in Hi-C aggregate plots.

Hi-C aggregate plots were performed using a homemade analysis pipeline developed in MATLAB Release R2019b (The MathWorks, Inc., Natick, United States). The distance-normalized sub-matrices over a window of 100kb surrounding the intersection between two anchored peaks were extracted. Finally, the aggregate plots were then created by averaging all of the sub-matrices together. For the bootstrapping method (Fig. S1o), we performed a series of iterations by randomly selecting *N* BEAF-32 anchors from the full list of anchors (*N* values were chosen as 2, 5, 10, 25, 50, 100, 250, and 300). We then calculated the mean log2(O/E) for this set of *N* anchors and repeated this process 10000 times for each value of *N*.

Data availability

The single nucleus pairwise distance matrices as well as XYZ coordinates of chromatin traces generated in this study have been deposited at our Open Science Framework project (https://osf.io/aqtxj/) with DOI: 10.17605/OSF.IO/AQTXJ. Source data are provided with this paper. The list of previously published datasets used in this study is provided in Supplementary Table 1.

Code availability

The code used for aggregation plot analysis and for post-processing Hi-M matrices are accessible at https://github.com/NollmannLab/messina_2022. For a permanent link, see DOI: 10.17605/OSF.IO/AQTXJ. Hi-M data were acquired using qudi-HiM [70]. The current version of qudi-HiM is found at https://github.com/NollmannLab/qudi-HiM, and an archived version at https://zenodo.org/record/6379944 (DOI: 10.5281/zenodo.6379944). Hi-M data were analyzed using pyHiM release 0.6, available at https://github.com/marcnol/pyHiM.

References

1. Szabo, Q., Bantignies, F. & Cavalli, G. Principles of genome folding into topologically associating domains. *Sci Adv* **5,** eaaw1668 (2019).
2. Rowley, M. J. & Corces, V. G. The three-dimensional genome: principles and roles of long-distance interactions. *Curr. Opin. Cell Biol.* **40,** 8–14 (2016).

3. Dixon, J. R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J. S. & Ren, B. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485,** 376–380 (2012).
4. Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F. d. R., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A. & Cavalli, G. Three-dimensional folding and functional organization principles of the Drosophila genome. *Cell* **148,** 458–472 (2012).
5. Nora, E. P., Lajoie, B. R., Schulz, E. G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N. L., Meisig, J., Sedat, J., Gribnau, J., Barillot, E., Blüthgen, N., Dekker, J. & Heard, E. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485,** 381–385 (2012).
6. Hou, C., Li, L., Qin, Z. & Corces, V. Gene density, transcription, and insulators contribute to the partition of the Drosophila genome into physical domains. *Mol. Cell* **48,** 471–484 (2012).
7. Rao, S. S. P., Huntley, M. H., Durand, N. C., Stamenova, E. K., Bochkov, I. D., Robinson, J. T., Sanborn, A. L., Machol, I., Omer, A. D., Lander, E. S. & Aiden, E. L. A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell* **162,** 687–688 (2015).
8. Symmons, O., Uslu, V. V., Tsujimura, T., Ruf, S., Nassari, S., Schwarzer, W., Ettwiller, L. & Spitz, F. Functional and topological characteristics of mammalian regulatory domains. *Genome Res.* **24,** 390–400 (2014).
9. Robson, M. I., Ringel, A. R. & Mundlos, S. Regulatory Landscaping: How Enhancer-Promoter Communication Is Sculpted in 3D. *Mol. Cell* **74,** 1110–1122 (2019).
10. Banigan, E. J. & Mirny, L. A. Loop extrusion: theory meets single-molecule experiments. *Curr. Opin. Cell Biol.* **64,** 124–138 (2020).
11. Matthews, N. E. & White, R. Chromatin Architecture in the Fly: Living without CTCF/Cohesin Loop Extrusion?: Alternating Chromatin States Provide a Basis for Domain Architecture in Drosophila. *Bioessays* **41,** e1900048 (2019).
12. Schoborg, T. A. & Labrador, M. The phylogenetic distribution of non-CTCF insulator

proteins is limited to insects and reveals that BEAF-32 is Drosophila lineage specific. *J. Mol. Evol.* **70,** 74–84 (2010).

13. Vogelmann, J., Valeri, A., Guillou, E., Cuvier, O. & Nollmann, M. Roles of chromatin insulator proteins in higher-order chromatin organization and transcription regulation. *Nucleus* **2,** 358–369 (2011).
14. Kyrchanova, O. & Georgiev, P. Chromatin insulators and long-distance interactions in Drosophila. *FEBS Lett.* **588,** 8–14 (2014).
15. Schoborg, T. & Labrador, M. Expanding the roles of chromatin insulators in nuclear architecture, chromatin organization and genome function. *Cell. Mol. Life Sci.* **71,** 4089–4113 (2014).
16. Bushey, A. M., Ramos, E. & Corces, V. G. Three subclasses of a Drosophila insulator show distinct and cell type-specific genomic distributions. *Genes Dev.* **23,** 1338–1350 (2009).
17. Ulianov, S. V., Khrameeva, E. E., Gavrilov, A. A., Flyamer, I. M., Kos, P., Mikhaleva, E. A., Penin, A. A., Logacheva, M. D., Imakaev, M. V., Chertovich, A., Gelfand, M. S., Shevelyov, Y. Y. & Razin, S. V. Active chromatin and transcription play a key role in chromosome partitioning into topologically associating domains. *Genome Res.* **26,** 70–84 (2016).
18. Luzhin, A. V., Flyamer, I. M., Khrameeva, E. E., Ulianov, S. V., Razin, S. V. & Gavrilov, A. A. Quantitative differences in TAD border strength underly the TAD hierarchy in Drosophila chromosomes. *J. Cell. Biochem.* **120,** 4494–4503 (2019).
19. Van Bortle, K., Nichols, M. H., Li, L., Ong, C.-T., Takenaka, N., Qin, Z. S. & Corces, V. G. Insulator function and topological domain border strength scale with architectural protein occupancy. *Genome Biol.* **15,** R82 (2014).
20. Hug, C. B., Grimaldi, A. G., Kruse, K. & Vaquerizas, J. M. Chromatin Architecture Emerges during Zygotic Genome Activation Independent of Transcription. *Cell* **169,** 216–228.e19 (2017).
21. Wang, Q., Sun, Q., Czajkowsky, D. M. & Shao, Z. Sub-kb Hi-C in D. melanogaster

reveals conserved characteristics of TADs between insect and mammalian cells. *Nat. Commun.* **9,** 188 (2018).

22. Vogelmann, J., Le Gall, A., Dejardin, S., Allemand, F., Gamot, A., Labesse, G., Cuvier, O., Nègre, N., Cohen-Gonsaud, M., Margeat, E. & Nollmann, M. Chromatin insulator factors involved in long-range DNA interactions and their role in the folding of the Drosophila genome. *PLoS Genet.* **10,** e1004544 (2014).
23. Liang, J., Lacroix, L., Gamot, A., Cuddapah, S., Queille, S., Lhoumaud, P., Lepetit, P., Martin, P. G. P., Vogelmann, J., Court, F., Hennion, M., Micas, G., Urbach, S., Bouchez, O., Nollmann, M., Zhao, K., Emberly, E. & Cuvier, O. Chromatin immunoprecipitation indirect peaks highlight long-range interactions of insulator proteins and Pol II pausing. *Mol. Cell* **53,** 672–681 (2014).
24. Rowley, M. J., Nichols, M. H., Lyu, X., Ando-Kuri, M., Rivera, I. S. M., Hermetz, K., Wang, P., Ruan, Y. & Corces, V. G. Evolutionarily Conserved Principles Predict 3D Chromatin Organization. *Mol. Cell* **67,** 837–852.e7 (2017).
25. Cardozo Gizzi, A. M., Cattoni, D. I., Fiche, J.-B., Espinola, S. M., Gurgo, J., Messina, O., Houbron, C., Ogiyama, Y., Papadopoulos, G. L., Cavalli, G., Lagha, M. & Nollmann, M. Microscopy-Based Chromosome Conformation Capture Enables Simultaneous Visualization of Genome Organization and Transcription in Intact Organisms. *Mol. Cell* (2019). doi:10.1016/j.molcel.2019.01.011
26. Pancaldi, V., Carrillo-de-Santa-Pau, E., Javierre, B. M., Juan, D., Fraser, P., Spivakov, M., Valencia, A. & Rico, D. Integrating epigenomic data and 3D genomic structure with a new measure of chromatin assortativity. *Genome Biol.* **17,** 152 (2016).
27. Madrid-Mencía, M., Raineri, E., Cao, T. B. N. & Pancaldi, V. Using GARDEN-NET and ChAseR to explore human haematopoietic 3D chromatin interaction networks. *Nucleic Acids Res.* **48,** 4066–4080 (2020).
28. Pancaldi, V. Network models of chromatin structure. *Curr. Opin. Genet. Dev.* **80,** 102051 (2023).
29. Matthey-Doret, C., Baudry, L., Breuer, A., Montagne, R., Guiglielmoni, N., Scolari, V.,

Jean, E., Campeas, A., Chanut, P. H., Oriol, E., Méot, A., Politis, L., Vigouroux, A., Moreau, P., Koszul, R. & Cournac, A. Computer vision for pattern detection in chromosome contact maps. *Nat. Commun.* **11,** 5795 (2020).

30. Batut, P. J., Bing, X. Y., Sisco, Z., Raimundo, J., Levo, M. & Levine, M. S. Genome organization controls transcriptional dynamics during development. *Science* **375,** 566–570 (2022).
31. Ogiyama, Y., Schuettengruber, B., Papadopoulos, G. L., Chang, J.-M. & Cavalli, G. Polycomb-Dependent Chromatin Looping Contributes to Gene Silencing during Drosophila Development. *Mol. Cell* **71,** 73–88.e5 (2018).
32. Espinola, S. M., Götz, M., Bellec, M., Messina, O., Fiche, J.-B., Houbron, C., Dejean, M., Reim, I., Cardozo Gizzi, A. M., Lagha, M. & Nollmann, M. Cis-regulatory chromatin loops arise before TADs and gene activation, and are independent of cell fate during early Drosophila development. *Nat. Genet.* **53,** 477–486 (2021).
33. Loubiere, V., Papadopoulos, G. L., Szabo, Q., Martinez, A.-M. & Cavalli, G. Widespread activation of developmental gene expression characterized by PRC1-dependent chromatin looping. *Sci Adv* **6,** eaax4001 (2020).
34. Rao, S. S. P., Huang, S.-C., Glenn St Hilaire, B., Engreitz, J. M., Perez, E. M., Kieffer-Kwon, K.-R., Sanborn, A. L., Johnstone, S. E., Bascom, G. D., Bochkov, I. D., Huang, X., Shamim, M. S., Shin, J., Turner, D., Ye, Z., Omer, A. D., Robinson, J. T., Schlick, T., Bernstein, B. E., Casellas, R., Lander, E. S. & Aiden, E. L. Cohesin Loss Eliminates All Loop Domains. *Cell* **171,** 305–320.e24 (2017).
35. Nora, E. P., Goloborodko, A., Valton, A.-L., Gibcus, J. H., Uebersohn, A., Abdennur, N., Dekker, J., Mirny, L. A. & Bruneau, B. G. Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell* **169,** 930–944.e22 (2017).
36. Gong, Y., Lazaris, C., Sakellaropoulos, T., Lozano, A., Kambadur, P., Ntziachristos, P., Aifantis, I. & Tsirigos, A. Stratification of TAD boundaries reveals preferential insulation of super-enhancers by strong boundaries. *Nat. Commun.* **9,** 542 (2018).

37. Cattoni, D. I., Cardozo Gizzi, A. M., Georgieva, M., Di Stefano, M., Valeri, A., Chamousset, D., Houbron, C., Déjardin, S., Fiche, J.-B., González, I., Chang, J.-M., Sexton, T., Marti-Renom, M. A., Bantignies, F., Cavalli, G. & Nollmann, M. Single-cell absolute contact probability detection reveals chromosomes are organized by multiple low-frequency yet specific interactions. *Nat. Commun.* **8,** 1753 (2017).
38. Finn, E. H., Pegoraro, G., Brandão, H. B., Valton, A.-L., Oomen, M. E., Dekker, J., Mirny, L. & Misteli, T. Extensive Heterogeneity and Intrinsic Variation in Spatial Genome Organization. *Cell* **176,** 1502–1515.e10 (2019).
39. Cardozo Gizzi, A. M., Espinola, S. M., Gurgo, J., Houbron, C., Fiche, J.-B., Cattoni, D. I. & Nollmann, M. Direct and simultaneous observation of transcription and chromosome architecture in single cells with Hi-M. *Nat. Protoc.* **15,** 840–876 (2020).
40. Mateo, L. J., Murphy, S. E., Hafner, A., Cinquini, I. S., Walker, C. A. & Boettiger, A. N. Visualizing DNA folding and RNA in embryos at single-cell resolution. *Nature* **568,** 49–54 (2019).
41. Cubeñas-Potts, C. & Corces, V. G. Architectural proteins, transcription, and the three-dimensional organization of the genome. *FEBS Lett.* **589,** 2923–2930 (2015).
42. Blythe, S. A. & Wieschaus, E. F. Establishment and maintenance of heritable chromatin structure during early embryogenesis. *Elife* **5,** (2016).
43. Schulz, K. N. & Harrison, M. M. Mechanisms regulating zygotic genome activation. *Nat. Rev. Genet.* **20,** 221–234 (2019).
44. Kaushal, A., Dorier, J., Wang, B., Mohana, G., Taschner, M., Cousin, P., Waridel, P., Iseli, C., Semenova, A., Restrepo, S., Guex, N., Aiden, E. L. & Gambetta, M. C. Essential role of Cp190 in physical and regulatory boundary formation. *Sci Adv* **8,** eabl8834 (2022).
45. Cavalheiro, G. R., Girardot, C., Viales, R. R., Feng, S., Pollex, T., Ngoc Cao, T. B., Lacour, P., Rabinowitz, A. & Furlong, E. E. M. CTCF, BEAF-32 and CP190 are not required for the initial establishment of TADs in early Drosophila embryos, but have locus specific roles. *bioRxiv* 2022.07.27.501678 (2022). doi:10.1101/2022.07.27.501678
46. Chathoth, K. T., Mikheeva, L. A., Crevel, G., Wolfe, J. C., Hunter, I., Beckett-Doyle, S.,

Cotterill, S., Dai, H., Harrison, A. & Zabet, N. R. The role of insulators and transcription in 3D chromatin organization of flies. *Genome Res.* **32,** 682–698 (2022).

47. Jiang, N., Emberly, E., Cuvier, O. & Hart, C. M. Genome-wide mapping of boundary element-associated factor (BEAF) binding sites in Drosophila melanogaster links BEAF to transcription. *Mol. Cell. Biol.* **29,** 3556–3568 (2009).
48. Cubeñas-Potts, C., Rowley, M. J., Lyu, X., Li, G., Lei, E. P. & Corces, V. G. Different enhancer classes in Drosophila bind distinct architectural proteins and mediate unique chromatin interactions and 3D architecture. *Nucleic Acids Res.* (2016). doi:10.1093/nar/gkw1114
49. Negre, N., Brown, C. D., Shah, P. K., Kheradpour, P., Morrison, C. A., Henikoff, J. G., Feng, X., Ahmad, K., Russell, S., White, R. A., Stein, L., Henikoff, S., Kellis, M. & White, K. P. A comprehensive map of insulator elements for the Drosophila genome. *PLoS Genet.* **6,** e1000814 (2010).
50. Rowley, M. J. & Corces, V. G. Organizational principles of 3D genome architecture. *Nat. Rev. Genet.* **19,** 789–800 (2018).
51. Galouzis, C. C. & Prud'homme, B. Transvection regulates the sex-biased expression of a fly X-linked gene. *Science* **371,** 396–400 (2021).
52. Harrison, M. M., Li, X.-Y., Kaplan, T., Botchan, M. R. & Eisen, M. B. Zelda binding in the early Drosophila melanogaster embryo marks regions subsequently activated at the maternal-to-zygotic transition. *PLoS Genet.* **7,** e1002266 (2011).
53. Nien, C.-Y., Liang, H.-L., Butcher, S., Sun, Y., Fu, S., Gocha, T., Kirov, N., Manak, J. R. & Rushlow, C. Temporal coordination of gene networks by Zelda in the early Drosophila embryo. *PLoS Genet.* **7,** e1002339 (2011).
54. Sun, Y., Nien, C.-Y., Chen, K., Liu, H.-Y., Johnston, J., Zeitlinger, J. & Rushlow, C. Zelda overcomes the high intrinsic nucleosome barrier at enhancers during Drosophila zygotic genome activation. *Genome Res.* **25,** 1703–1714 (2015).
55. Schulz, K. N., Bondra, E. R., Moshe, A., Villalta, J. E., Lieb, J. D., Kaplan, T., McKay, D. J. & Harrison, M. M. Zelda is differentially required for chromatin accessibility,

transcription factor binding, and gene expression in the early Drosophila embryo. *Genome Res.* **25,** 1715–1726 (2015).

56. Eagen, K. P., Aiden, E. L. & Kornberg, R. D. Polycomb-mediated chromatin loops revealed by a subkilobase-resolution chromatin interaction map. *Proc. Natl. Acad. Sci. U. S. A.* (2017). doi:10.1073/pnas.1701291114
57. Gerasimova, T. I., Byrd, K. & Corces, V. G. A chromatin insulator determines the nuclear localization of DNA. *Mol. Cell* **6,** 1025–1035 (2000).
58. Labrador, M. & Corces, V. G. Setting the boundaries of chromatin domains and nuclear organization. *Cell* **111,** 151–154 (2002).
59. Byrd, K. & Corces, V. G. Visualization of chromatin domains created by the gypsy insulator of Drosophila. *J. Cell Biol.* **162,** 565–574 (2003).
60. Melnikova, L. S., Georgiev, P. G. & Golovnin, A. K. The Functions and Mechanisms of Action of Insulators in the Genomes of Higher Eukaryotes. *Acta Naturae* **12,** 15–33 (2020).
61. Amankwaa, B., Schoborg, T. & Labrador, M. insulator proteins exhibit in vivo liquid-liquid phase separation properties. *Life Sci Alliance* **5,** (2022).
62. Kyrchanova, O. V., Bylino, O. V. & Georgiev, P. G. Mechanisms of enhancer-promoter communication and chromosomal architecture in mammals and Drosophila. *Front. Genet.* **13,** 1081088 (2022).
63. Tortora, M. M., Salari, H. & Jost, D. Chromosome dynamics during interphase: a biophysical perspective. *Curr. Opin. Genet. Dev.* **61,** 37–43 (2020).
64. Chen, H., Levo, M., Barinov, L., Fujioka, M., Jaynes, J. B. & Gregor, T. Dynamic interplay between enhancer-promoter topology and gene activity. *Nat. Genet.* **50,** 1296–1303 (2018).
65. Hansen, A. S., Pustova, I., Cattoglio, C., Tjian, R. & Darzacq, X. CTCF and cohesin regulate chromatin loop stability with distinct dynamics. *Elife* **6,** (2017).
66. Tang, X., Li, T., Liu, S., Wisniewski, J., Zheng, Q., Rong, Y., Lavis, L. D. & Wu, C. Kinetic principles underlying pioneer function of GAGA transcription factor in live cells. *Nat.*

Struct. Mol. Biol. **29,** 665–676 (2022).

67. Hansen, A. S., Cattoglio, C., Darzacq, X. & Tjian, R. Recent evidence that TADs and chromatin loops are dynamic structures. *Nucleus* **9,** 20–32 (2018).
68. Gabriele, M., Brandão, H. B., Grosse-Holz, S., Jha, A., Dailey, G. M., Cattoglio, C., Hsieh, T.-H. S., Mirny, L., Zechner, C. & Hansen, A. S. Dynamics of CTCF- and cohesin-mediated chromatin looping revealed by live-cell imaging. *Science* **376,** 496–501 (2022).
69. Lu, F. & Lionnet, T. Transcription Factor Dynamics. *Cold Spring Harb. Perspect. Biol.* **13,** (2021).
70. Barho, F., Fiche, J.-B., Bardou, M., Messina, O., Martiniere, A., Houbron, C. & Nollmann, M. Qudi-HiM: an open-source acquisition software package for highly multiplexed sequential and combinatorial optical imaging. *Open Research Europe* **2,** 46 Preprint at https://doi.org/10.12688/openreseurope.14641.2 (2022)
71. Schmidt, U., Weigert, M., Broaddus, C. & Myers, G. Cell Detection with Star-Convex Polygons. *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018* 265–273 Preprint at https://doi.org/10.1007/978-3-030-00934-2_30 (2018)
72. Imbert, A., Ouyang, W., Safieddine, A., Coleno, E., Zimmer, C., Bertrand, E., Walter, T. & Mueller, F. FISH-quant v2: a scalable and modular tool for smFISH image analysis. *RNA* **28,** 786–795 (2022).
73. Gurgo, J., Walter, J.-C., Fiche, J.-B., Houbron, C., Schaeffer, M., Cavalli, G., Bantignies, F. & Nollmann, M. Multiplexed chromatin imaging reveals predominantly pairwise long-range coordination between Drosophila Polycomb genes. *bioRxiv* 2022.05.16.492046 (2022). doi:10.1101/2022.05.16.492046
74. Wingett, S. W. & Andrews, S. FastQ Screen: A tool for multi-genome mapping and quality control. *F1000Res.* **7,** 1338 (2018).
75. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26,** 589–595 (2010).
76. Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E.,

Nusbaum, C., Myers, R. M., Brown, M., Li, W. & Liu, X. S. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9,** R137 (2008).

77. Ramírez, F., Ryan, D. P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A. S., Heyne, S., Dündar, F. & Manke, T. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44,** W160–5 (2016).
78. Khan, A. & Mathelier, A. Intervene: a tool for intersection and visualization of multiple gene or genomic region sets. *BMC Bioinformatics* **18,** 287 (2017).
79. Crane, E., Bian, Q., McCord, R. P., Lajoie, B. R., Wheeler, B. S., Ralston, E. J., Uzawa, S., Dekker, J. & Meyer, B. J. Condensin-driven remodelling of X chromosome topology during dosage compensation. *Nature* **523,** 240–244 (2015).
80. Knight, P. A. & Ruiz, D. A fast algorithm for matrix balancing. *IMA J. Numer. Anal.* **33,** 1029–1047 (2012).
81. Madrid-Mencía, M., Raineri, E. & Pancaldi, V. GARDEN-NET and ChAseR: a suite of tools for the analysis of chromatin networks. *bioRxiv* 717298 (2019). doi:10.1101/717298

Acknowledgements

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', and the Drosophila facility (BioCampus Montpellier, CNRS, INSERM, Univ Montpellier, Montpellier, France). The CBS is a member of the France-BioImaging, 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. F. R. and V.P. were supported by Fondation Toulouse Cancer Santé and Pierre Fabre Research Institute as part of the Chair of Bioinformatics in Oncology of the CRCT.

Author Contributions

O.M., V.P. and M.N. conceived the study and the design. O.M. acquired the data. O.M., F.R., and J.G. analyzed the data. O.M., J.-B.F., and F.R. wrote the software. J-B.F. built the microscope. O.M., F.R., V.P. and M.N. interpreted the data. M.N., O.M., and V.P. wrote the manuscript. M.N. and V.P. supervised the study and acquired funds.

Competing interests statement

The authors declare no competing interests.

Figure legends

Figure 1. Genomic regions displaying preferential interactions are predominantly bound by chromatin insulators.

a Cartoon illustrating Chromatin Assortativity (ChAs) of chromatin binding factors in a network of chromatin contacts. 5-kb genomic bins are represented by nodes in the chromatin network. Nodes are connected to each other if they form loops in Hi-C data, called by Chromosight [29]. Nodes are color-coded according to the presence or absence of a given chromatin binding factor (features). The assortativity is then calculated for each different factor (see *Methods*). **b** Bar plot illustrating ChAs-Z-Scores for 15 IBPs in the nc14 chromatin network classified by alphabetical order. The horizontal red dashed line represents the ChAs-Z-Score=2 threshold considered in this study. **c** Pearson's correlation between ChAs-Z-Scores from ChAs (y) and mean of log2(O/E) from APA (x) for the 15 IBPs tested. Class I and II are delineated by a vertical red dashed line centered at log2(O/E) = 0. **d** Cartoon illustrating the Hi-C aggregate map procedure around pairs of specific chromatin binding factors (peak). The first line illustrates the log2(O/E) aggregate map expected for a given factor involved in preferential contact formation and the second for a factor not involved in preferential contact formation (see *Methods*). **e** Aggregate Hi-C plots of Class I IBPs regions in nc14 embryos [20]. Maps show the log2(O/E) in a 50 kb window around the crossing point of two Class I IBPs regions : BEAF-32, CHRO, DREF, ZIPIC, Zw5, Z4. Source data are provided as a Source Data file.

Figure 2. Border-Border and non-border interactions are favored by an increase in insulation-score and an increase in IBPs binding respectively.

a Chromosight chromatin subnetwork from Hi-C data in nc14 embryos [20]. Each node of the network is a chromatin fragment, blue nodes represent nodes where a TAD boundary is found, and edges represent significant 3D interactions. **b** ChAs-Z-score for TAD borders (blue) versus distribution of ChAs scores for randomized networks (black). The ChAs-Z-Score is calculated for TAD borders based on comparing ChAs values with the distribution of the ChAs in randomized networks (see *Methods*). **c** Aggregation Hi-C plots for TADs borders in nc14 embryos stratified into five equal size category groups (I, II, III, IV and V) with an increasing level of insulation score. **d** Cartoon illustrating the different types of interaction observed in Hi-C dataset. Genomic bins are represented by color-coded nodes. Arcs represent interactions between pairs of genomic bins. Border / Border interactions are shown in blue, Border / Non-Border in red and Non-Border / Non-Border in black. **e** Donut chart representing the loop distribution called by Chromosight into the different types of interactions (left panel). Donut charts illustrating the quantification of Class I IBPs bound on each side of the anchored interaction (right panel). **f** Aggregation Hi-C plots for non-border regions bound by Class I IBPs in nc14 embryos. **g** Aggregation Hi-C plots for Class I IBPs in nc14 stratified into five equal size category groups (I, II, III, IV and V) with an increasing ChIP signal. **h** Log2(O/E) average interaction frequencies between five categories of Class I IBPs regions ranked by increasing ChIP signal in nc14 embryos. Source data are provided as a Source Data file.

Figure 3. Hi-M reveals visible interactions between insulator-bound chromatin regions.

a Cartoon illustrating the imaging based strategy used to study chromosome conformation at

the single cell level in intact *Drosophila melanogaster* embryos (Hi-M). **b** Top: nc14 Hi-C matrix along the *dpp* locus (2L:2343645-2758688) in *Drosophila melanogaster* (dm6). Purple and green represent high and low contact probabilities, respectively. Identified TADs borders from nc14 embryos [20] are represented by blue triangles. TADs are highlighted on the matrix with black dashed lines. Barcodes used for Hi-M sequential imaging are represented as boxes, with barcodes bound by Class I IBPs displayed in green. Bottom : ChIP-seq profiles for Class I IBPs (BEAF-32, Chromator, DREF, Z4, ZIPIC, Zw5) aligned with genomic coordinates and gene locations. **c** Top: Hi-M pairwise distance (PWD) matrix for nc14 embryos constructed from 23531 traces from 22 embryos. Red and blue represent low and high distances, respectively. Middle: insulation score derived from Hi-M data with different window sizes (1, 2 and 3 bins), and domainogram (see *Methods*). Bottom: proximity frequency matrix from nc14 embryos (cutoff distance: 200 nm). Pink and green represent high and low proximity frequencies, respectively. **d** Scatter plot illustrating the dependence of proximity frequency with genomic distance (cutoff distance: 200 nm) for Class I IBPs barcodes (green) and for non-Class I IBPs barcodes (black). Dashed black and green lines represent polynomial fits for each distribution, along with 95% confidence intervals. **e** Violin plot distributions representing the frequency with which each insulator barcode interacts with each other insulator barcode in our oligopaint library for nc14 embryos (see *Methods*). The dashed red line represents the mean. **f** Observed and expected proximity frequency versus number of interacting partners for Class I IBPs barcodes for nc14 embryos (green and red, respectively). **g** Histograms of Class I IBPs preferential interaction as a function of the number of interacting partners normalized by the number of pairwise interactions for nc14 embryos. Insulator barcodes are indicated on the left (I1 to I10). The color-scale represents the normalized frequency in log2-scale. Source data are provided as a Source Data file.

Figure 4. Border-border interactions are formed at nc14, while non-border interactions are gradually formed during development.

a Series of heat maps showing the ATACseq signal for Class I IBPs regions across two developmental stages (nc12 and nc13), within a window of +/- 1kb. **b** Aggregation Hi-C plots for TADs borders at different developmental stages (nc12, nc13, nc14, 3-4 hpf). **c** Aggregation Hi-C plots for non-border Class I IBPs group at different developmental stages (nc12, nc13, nc14, 3-4 hpf). **d** Aggregation Hi-C plots for each individual protein of the non-border Class I IBPs group at different developmental stages (nc12, nc13, nc14, 3-4hpf). **e** Hi-M pairwise distance (PWD) matrices for nc12 constructed with 1792 traces from 4 embryos and nc14 embryos constructed with 23531 traces from 22 embryos are shown at the top and at the bottom, respectively. **f** Differential pairwise distance matrix between nc14 and nc12. Red and blue respectively represent closer and farther distances in nc14 as compared to nc12. **g** Cumulative proximity frequency versus different cutoff distances curve for class I IBPs barcodes (green) and for 10 sets of control barcodes (black) for nc12 embryos. **h** Aggregation Hi-C plots for TADs borders for different biological conditions and treatments (nc14 triptolide-treated, nc14 alpha-amanitin-treated and nc14 knockdown of Zelda). **i** Aggregation Hi-C plots for non-border Class I IBPs group for different biological conditions and treatments (nc14 triptolide-treated, nc14 alpha-amanitin-treated and nc14 knockdown of Zelda). **j** Venn diagram representing the overlap between Class I IBPs peaks and Zelda peaks. **k** Series of heat maps showing the ATACseq signal for the different groups of peaks shown in 4j. **l** Metagene profiles of the ATACseq signal for Class I IBPs peaks bound by Zelda (orange) and not bound by Zelda (green). Source data are provided as a Source Data file.

Figure 1

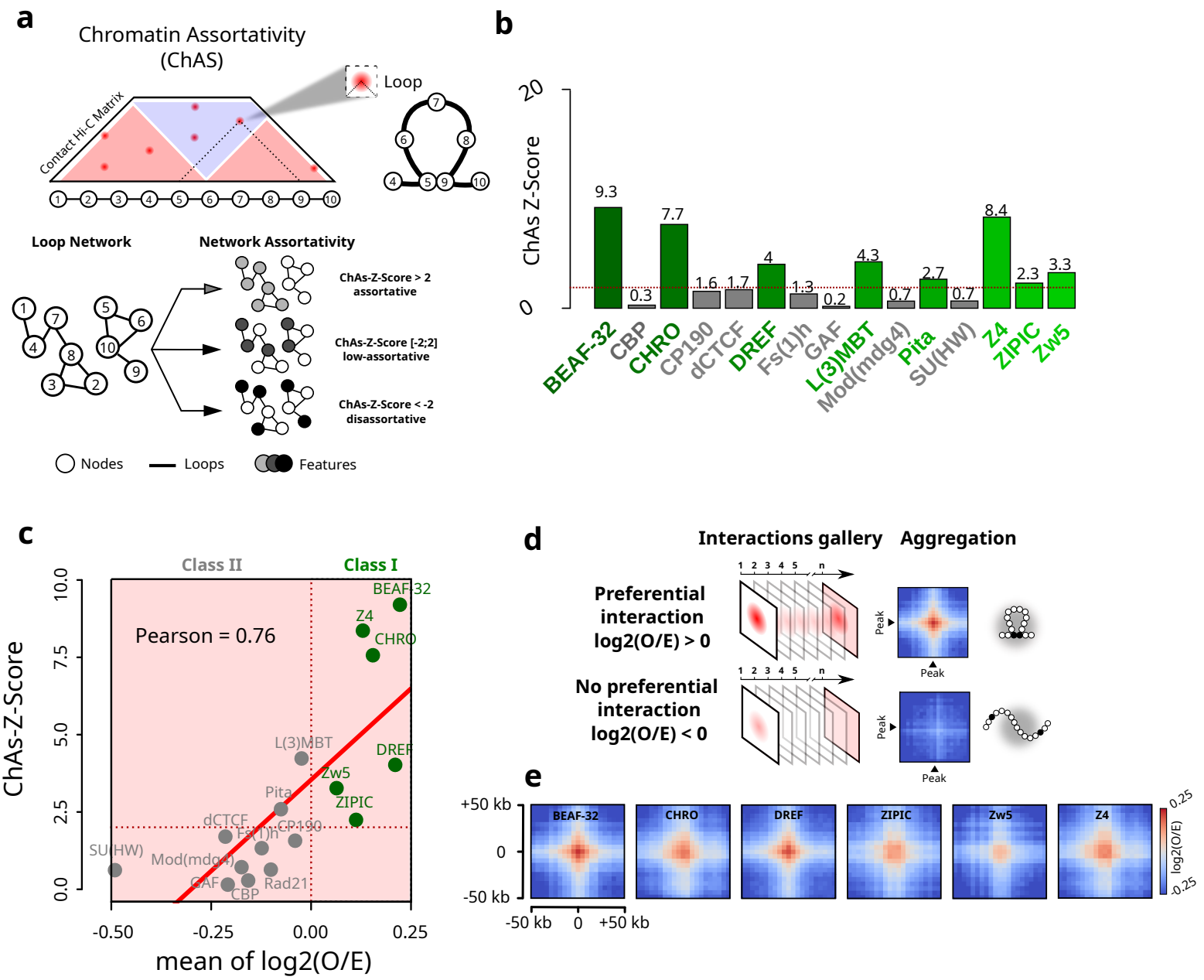

a
Chromatin Assortativity
(ChAS)
Contact Hi-C Matrix
Loop
Loop Network
Network Assortativity
ChAs-Z-Score > 2
assortative
ChAs-Z-Score [-2;2]
low-assortative
ChAs-Z-Score < -2
disassortative
Nodes
Loops
Features
b
ChAs Z-Score
9.3
0.3
7.7
1.6
1.7
4
1.3
0.2
4.3
0.7
2.7
0.7
8.4
2.3
3.3
BEAF-32
CBP
CHRO
CP190
dCTCF
DREF
Fs(1)h
GAF
L(3)MBT
Mod(mdg4)
Pita
SU(HW)
Z4
ZIPIC
Zw5
c
Class II
Class I
Pearson = 0.76
ChAs-Z-Score
mean of log2(O/E)
d
Interactions gallery
Aggregation
Preferential
interaction
log2(O/E) > 0
No preferential
interaction
log2(O/E) < 0
Peak
e
+50 kb
0
-50 kb
BEAF-32
CHRO
DREF
ZIPIC
Zw5
Z4
log2(O/E)

Figure 2

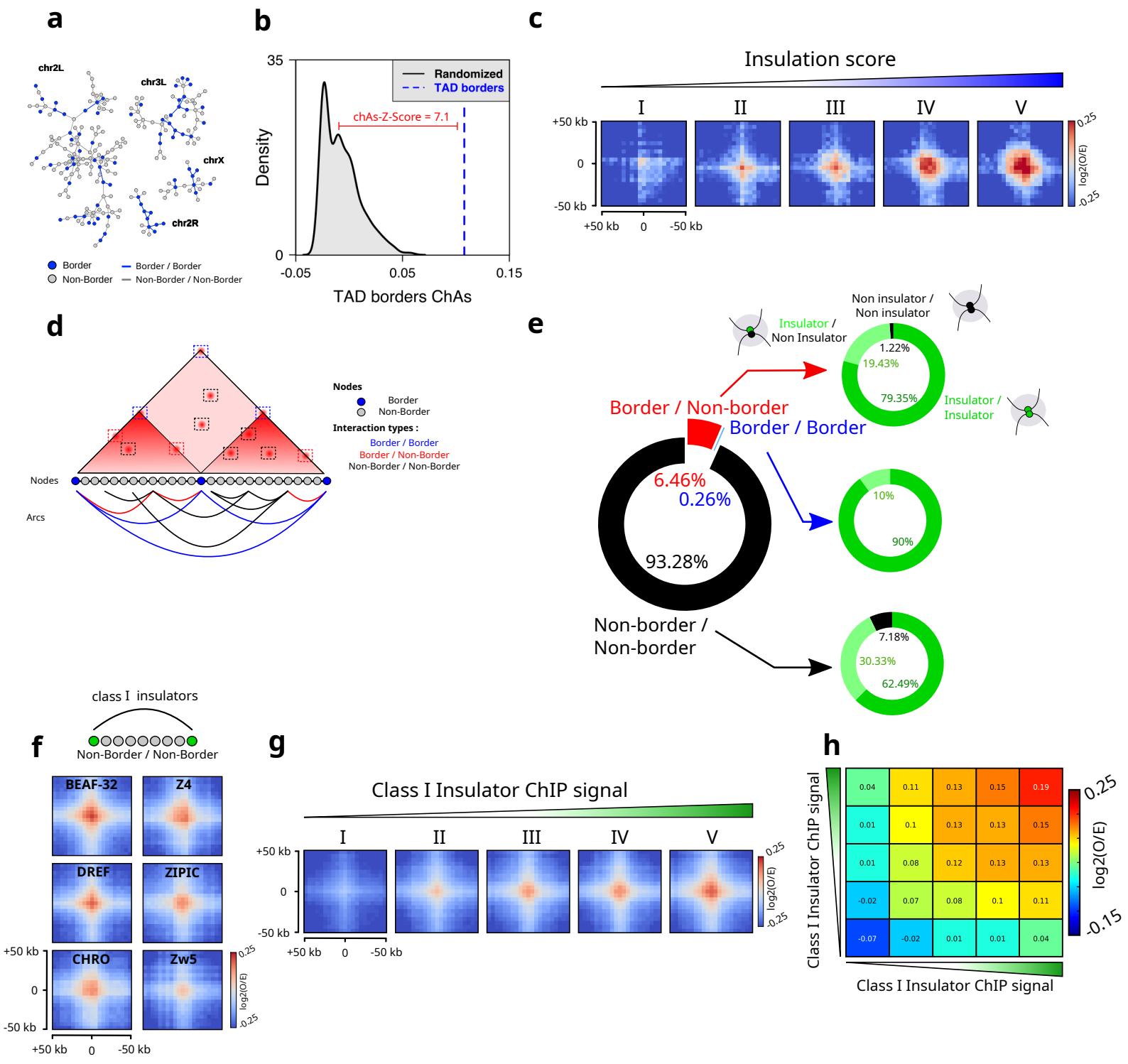

a
chr2L
chr3L
chrX
chr2R
Border
Non-Border
Border / Border
Non-Border / Non-Border
b
35
0
Density
Randomized
TAD borders
chAs-Z-Score = 7.1
-0.05
0.05
0.15
TAD borders ChAs
c
Insulation score
I
II
III
IV
V
+50 kb
0
-50 kb
log2(O/E)
0.25
-0.25
d
Nodes
Arcs
Nodes
Border
Non-Border
Interaction types :
Border / Border
Border / Non-Border
Non-Border / Non-Border
e
Non insulator /
Non insulator
Insulator /
Non Insulator
1.22%
19.43%
79.35%
Insulator /
Insulator
Border / Non-border
Border / Border
6.46%
0.26%
93.28%
10%
90%
Non-border /
Non-border
7.18%
30.33%
62.49%
f
class I insulators
Non-Border / Non-Border
BEAF-32
Z4
DREF
ZIPIC
CHRO
Zw5
+50 kb
0
-50 kb
log2(O/E)
0.25
-0.25
g
Class I Insulator ChIP signal
I
II
III
IV
V
+50 kb
0
-50 kb
log2(O/E)
0.25
-0.25
h
Class I Insulator ChIP signal
0.04 0.11 0.13 0.15 0.19
0.01 0.1 0.13 0.13 0.15
0.01 0.08 0.12 0.13 0.13
-0.02 0.07 0.08 0.1 0.11
-0.07 -0.02 0.01 0.01 0.04
Class I Insulator ChIP signal
log2(O/E)
0.25
-0.15

Figure 3

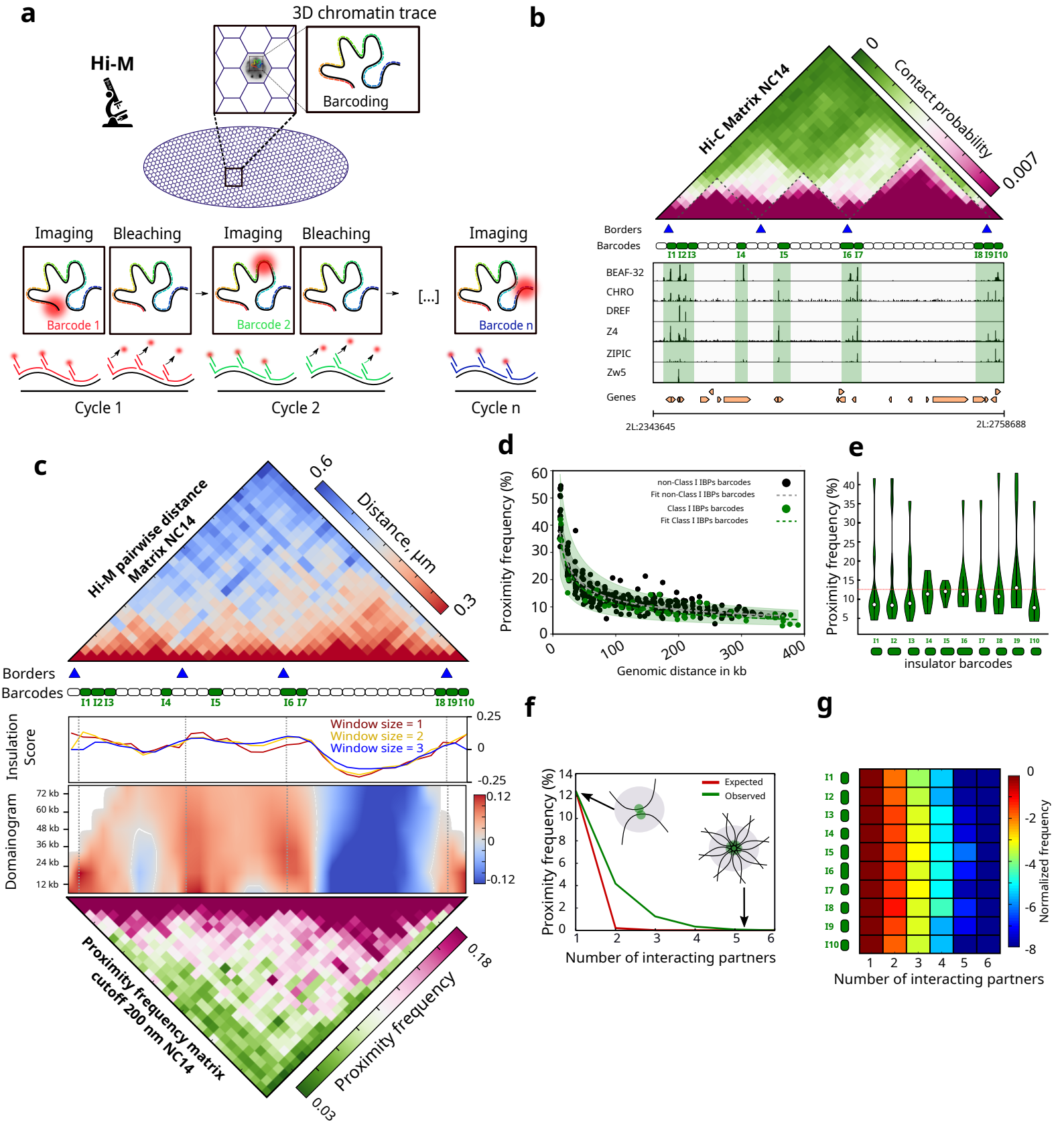
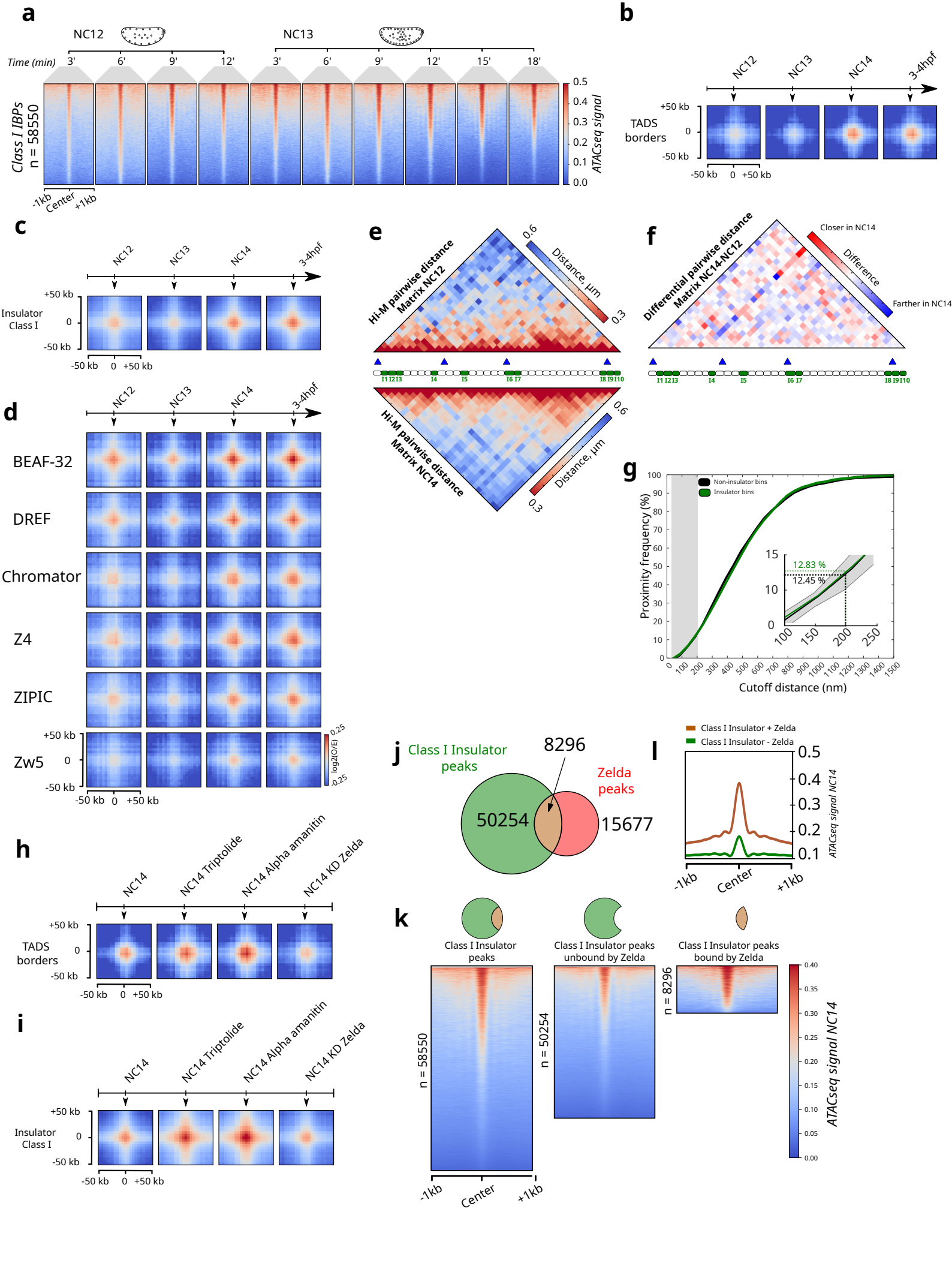

a
Hi-M
3D chromatin trace
Barcoding
Imaging
Bleaching
Imaging
Bleaching
Imaging
[...]
Barcode 1
Barcode 2
Barcode n
Cycle 1
Cycle 2
Cycle n
b
Hi-C Matrix NC14
0
Contact probability
0.007
Borders
Barcodes
I1 I2 I3
I4
I5
I6 I7
I8 I9 I10
BEAF-32
CHRO
DREF
Z4
ZIPIC
Zw5
Genes
2L:2343645
2L:2758688
c
Hi-M pairwise distance Matrix NC14
0.6
Distance, µm
0.3
Borders
Barcodes
I1 I2 I3
I4
I5
I6 I7
I8 I9 I10
Insulation Score
Window size = 1
Window size = 2
Window size = 3
0.25
0
-0.25
Domainogram
72 kb
60 kb
48 kb
36 kb
24 kb
12 kb
0.12
0
-0.12
Proximity frequency matrix cutoff 200 nm NC14
0.18
Proximity frequency
0.03
d
Proximity frequency (%)
non-Class I IBPs barcodes
Fit non-Class I IBPs barcodes
Class I IBPs barcodes
Fit Class I IBPs barcodes
Genomic distance in kb
e
Proximity frequency (%)
insulator barcodes
f
Proximity frequency (%)
Expected
Observed
Number of interacting partners
g
Normalized frequency
Number of interacting partners

Figure 4


a
NC12
NC13
Time (min)
3'
6'
9'
12'
15'
18'
Class I IBPs
n = 58550
ATACseq signal
-1kb
Center
+1kb
b
NC12
NC13
NC14
3-4hpf
TADS
borders
+50 kb
0
-50 kb
c
Insulator
Class I
d
BEAF-32
DREF
Chromator
Z4
ZIPIC
Zw5
log2(O/E)
e
Hi-M pairwise distance
Matrix NC12
Hi-M pairwise distance
Matrix NC14
Distance, µm
f
Differential pairwise distance
Matrix NC14-NC12
Closer in NC14
Difference
Farther in NC14
g
Non-insulator bins
Insulator bins
Proximity frequency (%)
12.83 %
12.45 %
Cutoff distance (nm)
h
NC14
NC14 Triptolide
NC14 Alpha amanitin
NC14 KD Zelda
i
j
Class I Insulator
peaks
8296
Zelda
peaks
50254
15677
k
Class I Insulator
peaks
Class I Insulator peaks
unbound by Zelda
Class I Insulator peaks
bound by Zelda
n = 58550
n = 50254
n = 8296
ATACseq signal NC14
l
Class I Insulator + Zelda
Class I Insulator - Zelda
ATACseq signal NC14

SUPPLEMENTARY INFORMATION RELATED TO

3D chromatin interactions involving *Drosophila* insulators are infrequent but preferential and arise before TADs and transcription

Olivier Messina[1], Flavien Raynal[2], Julian Gurgo[1], Jean-Bernard Fiche[1], Vera Pancaldi[2,3*], Marcelo Nollmann[1,*]

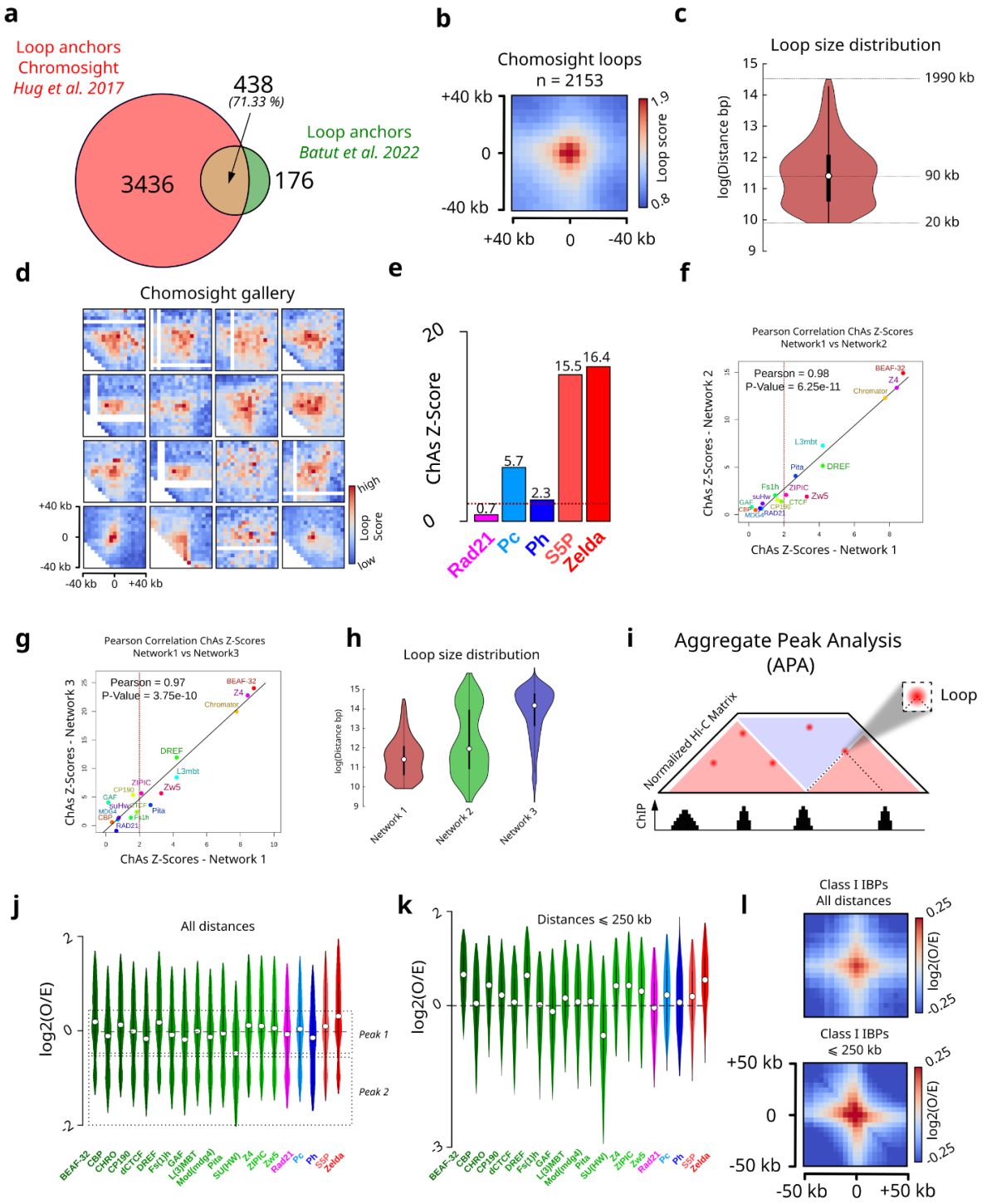
[1] *Centre de Biologie Structurale, Univ Montpellier, CNRS UMR 5048, INSERM U1054, Montpellier, France.*

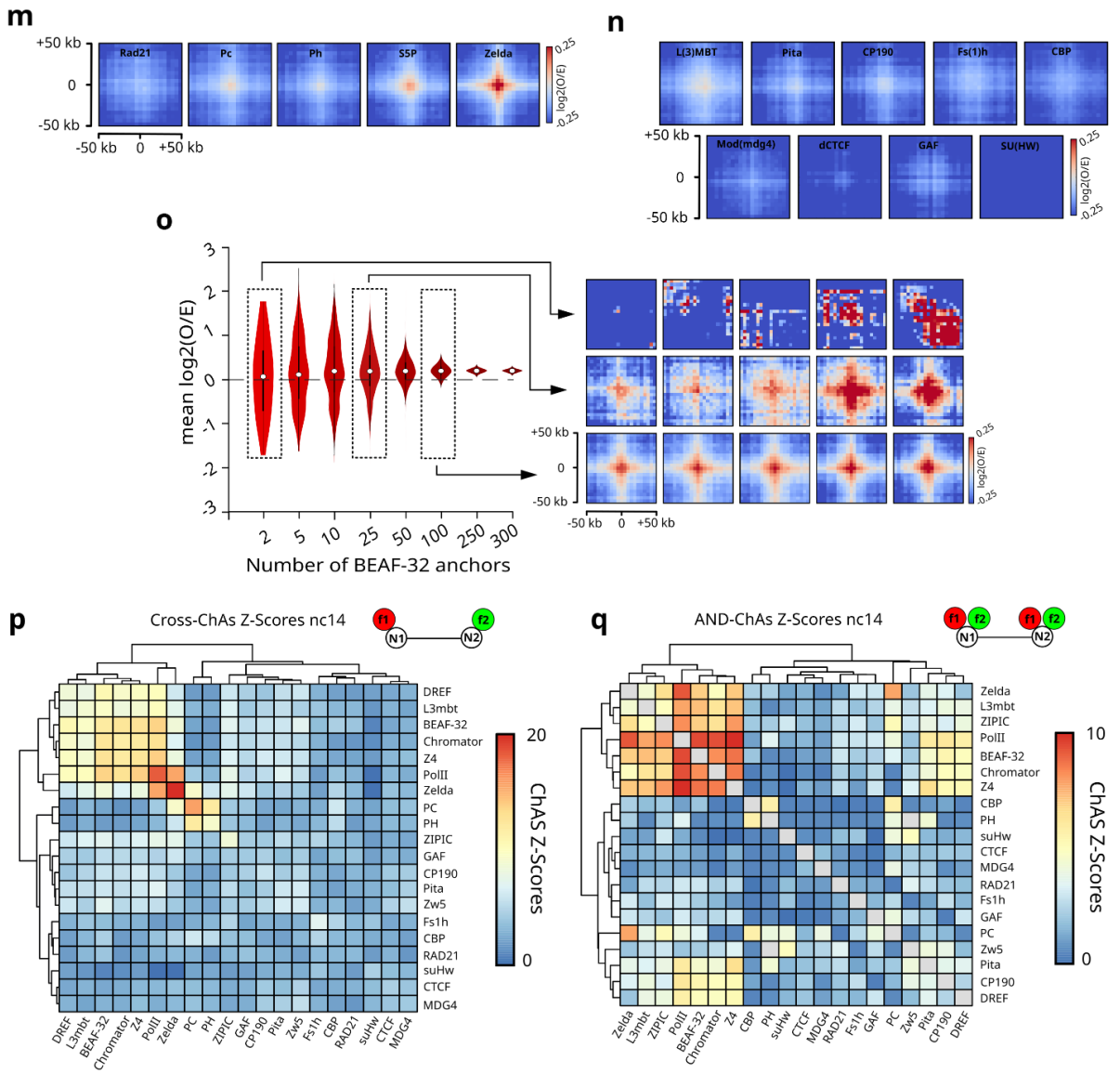
[2] *Université de Toulouse, Inserm, CNRS, Université Toulouse III-Paul Sabatier, Centre de Recherches en Cancérologie de Toulouse, Toulouse, France.*

[3] *Barcelona Supercomputing Center, Barcelona, Spain.*

[*] Corresponding authors: vera.pancaldi@inserm.fr, marcelo.nollmann@cbs.cnrs.fr

Supplementary Figure S1. Supplementary Data complementary to Figure 1.

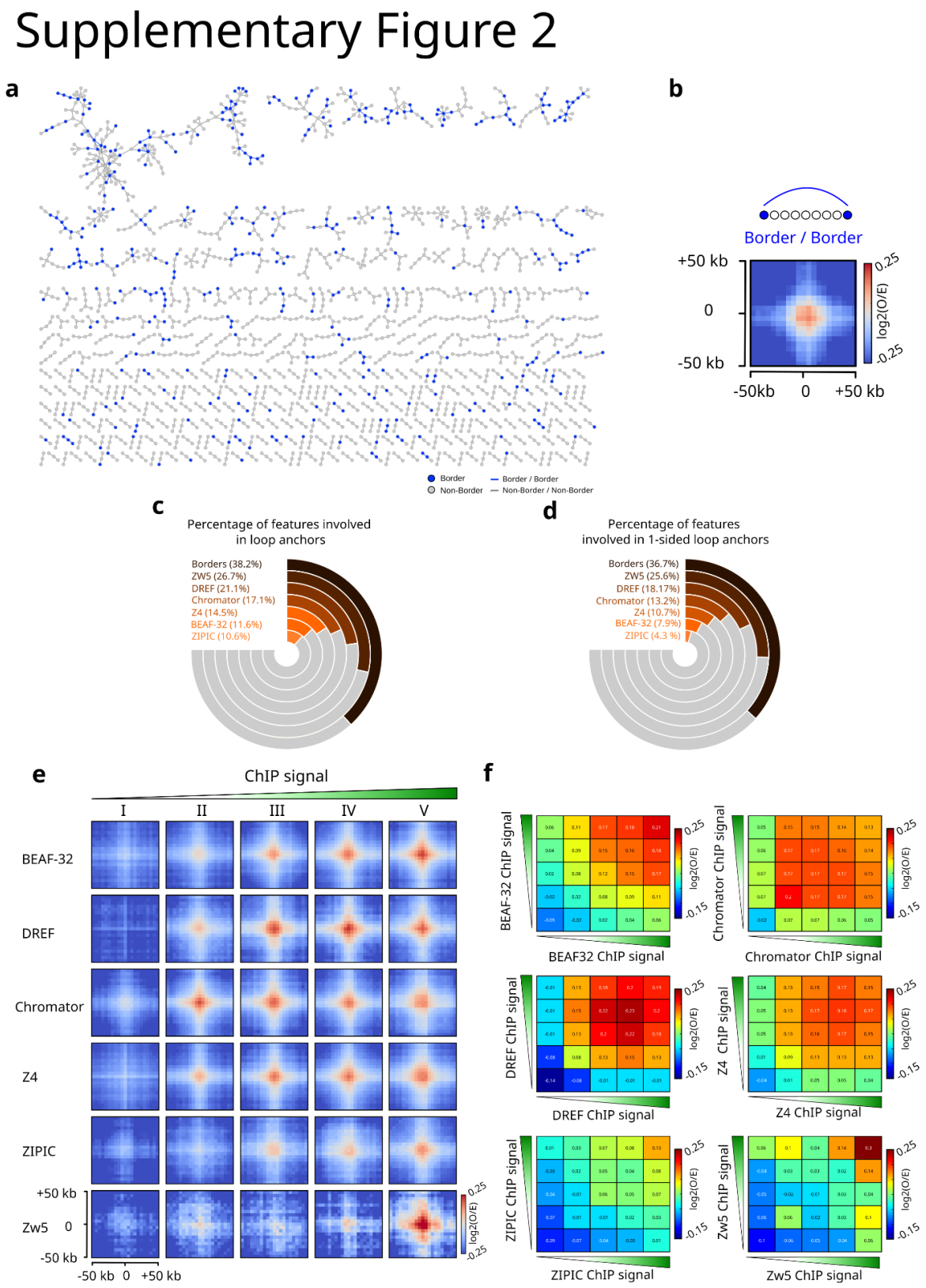




a. Venn diagram representing the overlap, at +/- 10kb, between loop anchors obtained from chromosight on data from Hug *et al.* [1] using the nc14 dataset, and loop anchors detected in *Batut et al.* 2022. [2] **b**. Pileup plot of all detected loops (2153) from Chromosight in nc14 embryos, centered on loop anchors, within a window of +/- 40kb. Red and blue represent high and low loop score values, respectively. **c**. Violin plot illustrating the loop size distribution from our Chromosight network 1. **d**. Examples of Hi-C maps showing regions with loops detected by Chromosight. Red and blue represent high and low loop scores, respectively. Matrices are distance normalized. **e**. Barplot illustrating the ChAs-Z-Scores for the cohesin subunit (Rad21), Polycomb group proteins (Pc, Ph), RNA polymerase II (RNAPII CTD phospho-Ser5) and the pioneering factors (Zelda) in the nc14 chromatin network classified by alphabetical order. **f**. Pearson correlation between ChAs Z-Scores of Network 1 and Network 2. (see *Methods*) **g.** Pearson correlation between ChAs Z-Scores of Network 1 and Network 3 (see *Methods*). **h.** Violin plots illustrating the loop size distribution from Chromosight network 1, 2 and 3. **i**. Cartoon illustrating Aggregate Peak Analysis (APA) of chromatin binding factors in distance normalized Hi-C matrix. **j**. Violin plots illustrating the distribution of the log2(O/E) for 15 IBPs, green, the cohesin subunit (Rad21), pink, Polycomb group proteins (Pc, Ph), blue, RNA polymerase II (RNAPII CTD phospho-Ser5) and the pioneering factors (Zelda), red, in the nc14 chromatin network for all genomic distances classified by

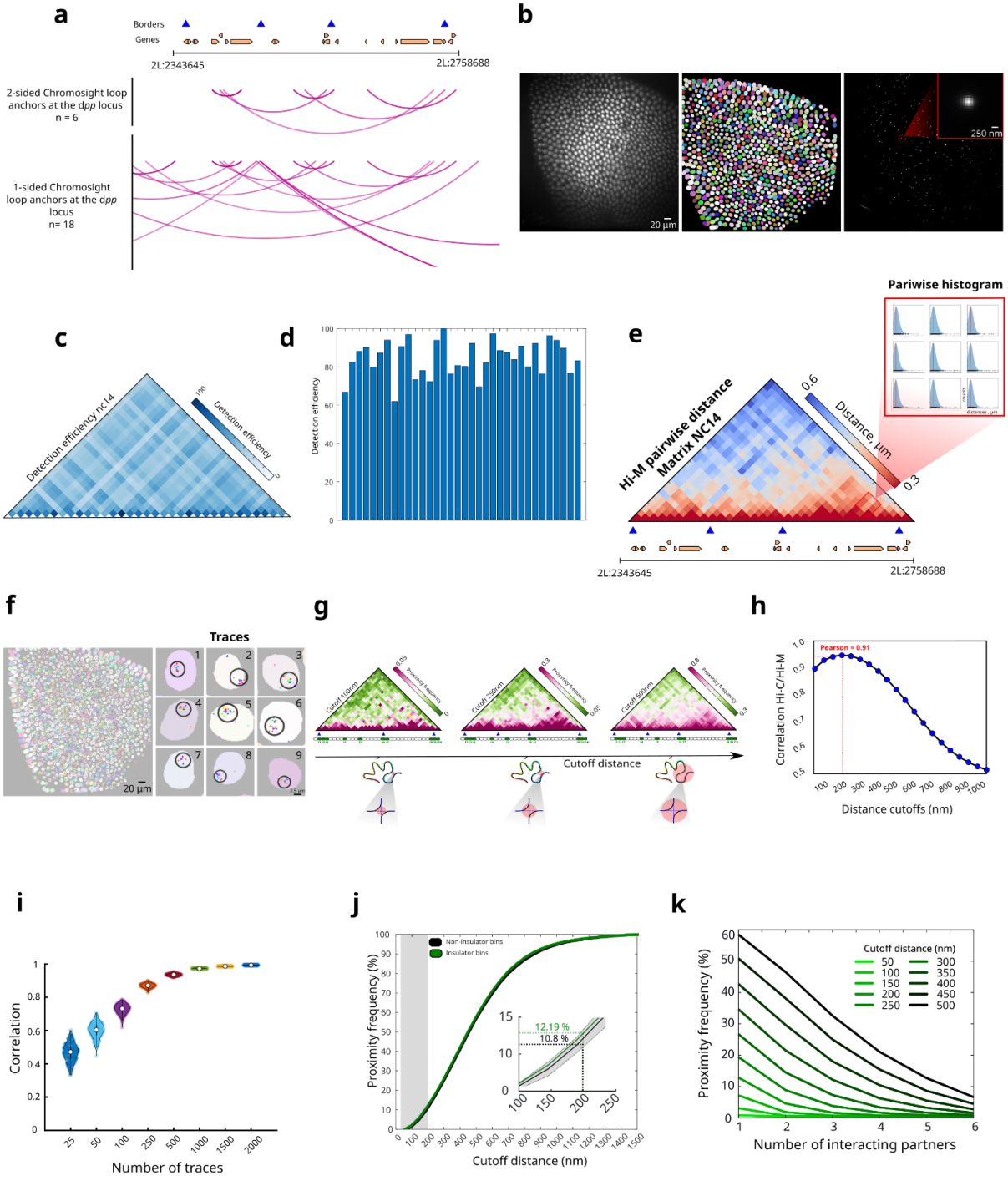
alphabetical order. Dashed rectangles pinpoint the two different peaks observed in the distribution. **k**. For genomic distances shorter than 250 kb. **l.** Aggregation Hi-C plots for nc14 embryos for Class I IBPs at all genomic distances (top) and when a maximum distance threshold of 250 kb is applied (bottom). **m**. Aggregation Hi-C plots for nc14 embryos for the cohesin subunit (Rad21), Polycomb group proteins (Pc, Ph), RNA polymerase II (RNAPII CTD phospho-Ser5) and the pioneering factors (Zelda), for all genomic distances. **n.** Aggregation Hi-C plots for nc14 embryos for Class II IBPs that do not display a positive log2(O/E) : L(3MBT), Pita, CP190, Fs(1)h, CBP, Mod(mdg4), dCTCF, GAF, SU(HW), for all genomic distances. **o**. Violin plots illustrating the distribution of the mean log2(O/E) for different sets of BEAF-32 anchors over 10000 iterations (left). Example of aggregation Hi-C plots for nc14 embryos for different numbers of BEAF-32 anchors (right) (see *Methods*). **p**. Heat map representing the ChAS Z-Scores from Cross-ChAs analysis on the nc14 chromatin network (see *Methods*). **q**. Heat map representing the ChAS Z-Scores from AND-ChAs analysis on the nc14 chromatin network (see *Methods*).

Supplementary Figure S2. Supplementary Data complementary to Figure 2.



a. Full chromosight chromatin network from Hi-C data at nc14 embryos [1]. Subnetworks containing less than 3 nodes are not represented. **b**. Aggregation Hi-C plots based on Hi-C data for TADs borders called in nc14. **c**. Radial bar chart representing the percentage of features (i.e borders or insulator peaks) involved in loop anchors. Only non-border interactions are considered. **d**. Radial bar chart representing the percentage of features (i.e borders or insulator peaks) involved in 1-sided loop anchors. Only non-border interactions are considered. The majority of features are involved in 1-sided loops. **e**. Aggregation Hi-C plots for each individual protein of the Class I IBP group in nc14 embryos classified by increasing ChIP signal. **f**. log2(O/E) for each individual protein of the Class I IBP group in nc14 ranked by increasing insulator enrichment in nc14 embryos.

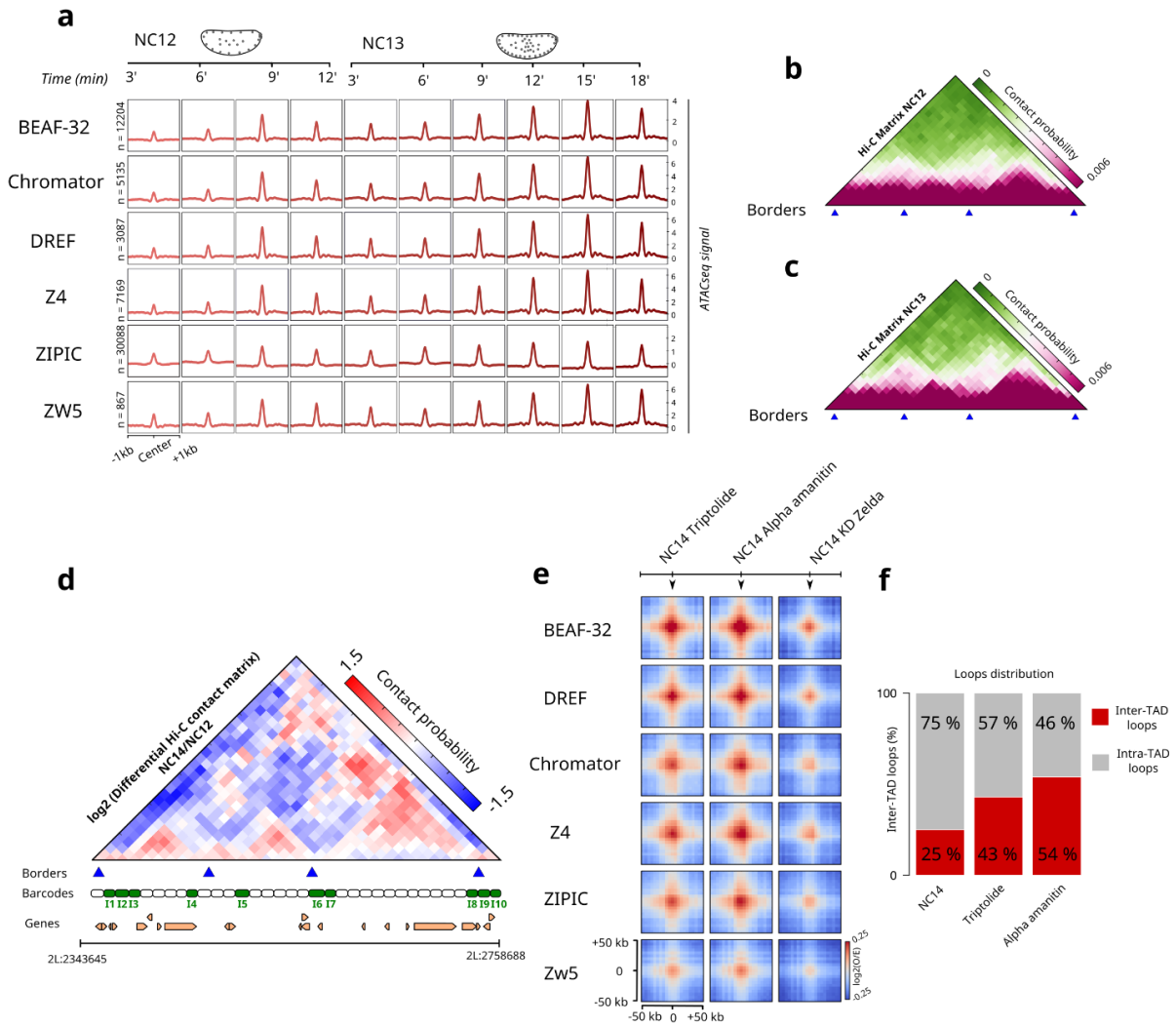
Supplementary Figure S3. Supplementary Data complementary to Figure 3.



a. Interaction arc diagrams showing called loops from Chromosight at the *Dpp* locus. Top: loops where both anchors are present at the *Dpp* locus are shown. Bottom: loops where at least 1 anchor is found at the *Dpp* locus. **b**. Left: Gray-scale image of DAPI-stained nuclei from nc14 Drosophila embryos. Middle: Segmented DAPI masks after nuclei segmentation. Right: Maximum intensity projection of the fluorescence signal from a single barcode in the same field of view. **c**. Detection efficiency matrices normalized by the most detected barcode pair. Dark blue and white represent high and low detection efficiencies respectively. **d**. Barplot representing the detection efficiency for all barcodes normalized by the most detected barcode. **e**. Hi-M pairwise distance (PWD) matrix for nc14 embryos. The zoom shows the histogram of the pairwise distribution of distances from a subset of a barcode pair. **f**. Example

of segmented barcodes in traces in the same field. The color code indicates different barcodes found in a trace. **g**. Hi-M proximity frequency matrices generated with different cutoff distances for nc14 embryos. From left to right the cutoffs considered are 100, 250, 500 nm. **h**. Pearson correlation coefficient between interpolated Hi-C contact map and Hi-M proximity frequency maps generated with different cutoff distances. The higher correlation is reached with a cutoff of 200 nm. **i**. Violin plots representing the Pearson correlation between and the Hi-M ensemble matrix and matrices generated by sampling subsets of traces by bootstrapping. The Hi-M ensemble matrix was obtained by considering all the traces available. For each condition, 250 bootstrapping cycles were used. **j**. Cumulative proximity frequency versus different cutoff distances for Class I IBP barcodes (green) and for 10 sets of control barcodes (black) for nc14 embryos. For the control, the solid black line represents the mean and the gray shade represents two standard deviations calculated from the variability of controls. **k**. Proximity frequency plotted against the number of interacting partners for Class I IBP barcodes in nc14 embryos, considering different cutoff distances.

Supplementary Figure S4. Supplementary Data complementary to Figure 4.



a. Piled ATACseq profiles for each individual Class I IBP across multiple developmental stages (+/- 1kb window). **b**. Nc12 Hi-C matrix from Hug. *et al* [1]. **c**. Nc13 Hi-C matrix from Hug. *et al* [1]. **d**. Log2 differential pairwise Hi-C contact matrix comparing nc14 and nc12. Red and blue indicate higher and lower contacts in nc14 relative to nc12, respectively. **e**. Aggregation Hi-C plots for each individual protein of the non-border Class I IBPs group for different biological conditions and treatments (nc14 triptolide-treated, nc14 alpha-amanitin-treated and nc14 knockdown of Zelda). **h**. Barplot showing the distribution of the percentage of inter-TAD and intra-TAD loops in nc14 wt, nc14 triptolide-treated, nc14 alpha-amanitin-treated.

3.2 Manuscript in preparation

3.2.1 Rationale of the work

The importance of 3D genome organisation in the regulation of gene expression has become increasingly clear in recent years. Research has revealed remarkable changes in chromatin structure, including both intra- and inter-TAD chromatin interactions, particularly in the context of cellular differentiation. In this study, we aimed to investigate how changes in chromatin structure occur at the single cell level in different mouse tissues and cell types under physiological conditions as well as during the onset of disease. To this end, we implemented the chromosome tracing method (Hi-M) and an innovative machine learning based tool to decompose the 3D structure of single cells (3DTopic).

3.2.2 Discrete chromatin folding motifs define single-cell 3D structures in mammalian tissue

Discrete chromatin folding motifs define single-cell 3D structures in mammalian tissues

Olivier Messina[1], Jean-Bernard Fiche[1], Christel Elkhoury Youhanna[1], Gautham Ganesh[1], Alexandre Mesnager[1], Ildem Akerman[2,3], David J. Hodson[4], Marie Schaeffer[1], Marcelo Nollmann[1]

Affiliations

[1] Centre de Biologie Structurale, Univ Montpellier, CNRS UMR 5048, INSERM U1054, 34090 Montpellier, France.
[2] Institute of Metabolism and Systems Research (IMSR), University of Birmingham, Birmingham, UK.
[3] Centre for Endocrinology, Diabetes and Metabolism, Birmingham Health Partners, Birmingham, UK.
[4] Oxford Centre for Diabetes, Endocrinology and Metabolism (OCDEM), NIHR Oxford Biomedical Research Centre, Churchill Hospital, Radcliffe Department of Medicine, University of Oxford, Oxford OX3 7LE, UK.

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

Abstract

The structure of chromatin is highly variable among single cells, yet whether and how this variability is related to cell-type specific programs has been so far unexplored. Here, we used imaging-based spatial genomics to visualize 3D chromatin organization across multiple cell types and tissues. Single-cell chromatin traces can be decomposed into a discrete number of chromatin folding motifs (CFMs) that comprise loops, stripes and domains. We used bioinformatics analysis and polymer simulations to show that loop extrusion is sufficient to produce loop and stripe CFMs, but that *cis*-regulatory interactions contribute to the formation of domain CFMs. By comparing CFM utilization across multiple cell types in seven different mouse tissues, we show that the single cell 3D chromatin architecture can be articulated through a unique set of CFMs, each contributing with varying proportions. In the pancreas, we observe chromatin structural ensembles and motif utilization that vary between cell types with different transcriptional programs. Notably, induction of obesity and type 2 diabetes lead to changes in 3D chromatin structure and utilization of CFMs in the different pancreatic cell types, indicating that metabolic deregulation leads to changes in both architectural and *cis*-regulatory interactions. Our approach provides valuable insights into the organization of chromatin in single cells within complex tissues, and how it changes in a disease state.

Introduction

Eukaryotic chromosomes exhibit intricate folding patterns over a range of length scales that include chromosome territories, A/B compartments and topologically associating domains (TADs) (Dixon et al., 2012; Nora et al., 2012; Sexton et al., 2012). In mammals, most TADs are delimited by convergent CCCTC binding sites (CTCF) and their formation relies primarily on loop extrusion by the cohesin complex (Fudenberg et al., 2016; Rao et al., 2015). In addition, cis-regulatory communication involving enhancers (E) and promoters (P) can lead to the formation of regulatory chromatin loops that are cell type specific (Furlong & Levine, 2018; Kragesteen et al., 2018; Peslak et al., 2023; Schoenfelder & Fraser, 2019). Loop extrusion is considered as one of the major mechanisms bringing enhancers and promoters into close proximity to regulate transcription (Karpinska & Oudelaar, 2023). Disruption of TADs can lead to ectopic contacts between enhancers and promoters, ultimately leading to gene misexpression and to developmental defects (Lupiáñez et al., 2015; Tena & Santos-Pereira, 2021). However, how changes in TAD architecture and E-P communication contribute to cell type-specific transcriptional regulation is currently under intense debate in the field.

Several experimental studies established that TAD are highly structurally variable between single cells (Bintu et al., 2018; Cattoni et al., 2017; Finn et al., 2019; Flyamer et al., 2017; Giorgetti et al., 2014; Stevens et al., 2017), challenging the traditional notion that TADs are stable domains favoring cis-regulatory interactions. More recently, measurements of single-cell chromatin dynamics showed that loop extrusion is a highly dynamic process and that CTCF-mediated loops are infrequent and short-lived (Gabriele et al., 2022; Mach et al., 2022), underlying the highly dynamic structure of chromatin in single cells and providing a rationale for the large variations in chromatin conformations observed in single cells by static methods. Consistent with this picture, multiplexed imaging showed that E-P proximity occurs at low frequencies and with low specificity (Espinola et al., 2021; Mateo et al., 2019). Despite this, enhancers can regulate transcriptional dynamics robustly by means of non-linear responses (Xiao et al., 2021; Zuin et al., 2022). Overall, these findings suggest that TAD architecture results from multiple dynamic mechanisms acting simultaneously. So far, it is unclear whether these mechanisms are modulated between cell types executing different transcriptional programs.

Here, we addressed this issue by reconstructing the single cell chromatin structure of multiple cell types from seven different mouse tissues. We developed 3DTopic, an unsupervised Bayesian framework based on topic modeling to analyze variability in single-cell chromatin organization. We found that 3D chromatin structures can be decomposed in a small set of discrete chromatin folding motifs (hereafter CFMs) presenting just three basic structural elements: loops, domains and stripes. Remarkably, this single set of CFMs can robustly describe the 3D chromatin structures of cell types from multiple mouse tissues. The folding of CFMs in loops, domains and stripes can be effectively modeled using loop-extrusion polymer models; however *cis*-regulatory interactions are also required to recover the experimentally-determined CFMs. Interestingly, we observe the existence of cell type-specific chromatin structures within different pancreatic cell types that can be described by varying CFMs utilization. Finally, we show that induction of obesity and type 2 diabetes (T2D) results in changes in how pancreatic cell types use loop, domain and stripe CFMs. Our discovery and quantification of chromatin folding motifs usage across cell types contributes to our understanding of the role of 3D genome organization in cell type specific transcription and to the mechanisms involved.

Results

The 3D architecture of the Pdx1 locus can be described by a set of discrete Chromatin Folding Motifs.

To explore single-cell heterogeneity in 3D chromatin organization in adult mammalian tissues, we adapted Hi-M, a microscopy-based chromosome conformation capture technique to trace chromatin folding in cryosectioned mouse pancreas (Fig. 1a) (Cardozo Gizzi et al., 2019). Specifically, we imaged the 3D chromatin organization of a 625-kb region surrounding the pancreatic and duodenal homeobox 1 gene (Pdx1) gene at ~25-kb resolution. Pdx1 encodes for a transcription factor that plays a central role in pancreatic β-cell function. To image this locus in single cells, we hybridized a library with 25 Hi-M primary barcodes, each spanning approximately 25 kb. Nuclei and barcodes were registered, segmented and localized as in previous studies (Figs. S1a-d). Ensemble pairwise distances (PWD) maps were built by calculating the median of the pairwise distance distribution for all combinations of captured barcode pairs. Proximity maps were constructed by calculating the frequency of co-localization for each pair of barcodes in a chromatin trace, using a distance threshold of 200 nanometers (nm), a value that optimizes the correlation between Hi-M and Micro-C datasets and that was used in other studies (d= 200 nm) (Figs. S1e,g). We acquired Hi-M traces for the Pdx1 locus for 4 individual experiments and generated sufficient data to ensure a statistically representative ensemble map (Figs. 1b, S1f). The region of Pdx1 contains a single TAD demarcated by a prominent apical loop. We note also the presence of sub-TADs nested within the TADs delineated by a moderate increase in insulation score (Fig. 1b).

To study how the Pdx1 locus is organized in single cells, we developed 3DTopic, an innovative unsupervised Bayesian framework based on topic modeling. 3DTopic uses Latent Dirichlet Allocation (LDA) to decompose single-cell 3D structures into a discrete number of folding topics that we named Chromatin Folding Motifs (CFMs) (Fig. 1c, *Methods*). CFMs represent an orthogonal base of 3D motifs that are sufficient to uniquely decompose any single 3D structure (Fig. 1d). The set-decomposed cells can in turn be used to reconstruct the experimental ensemble proximity frequency matrix with high accuracy (93% Pearson correlation), underscoring the accuracy and power of our approach to describe the complex landscape of chromatin organization at the single cell level (Fig. S1h). Interestingly, a relatively small number of 3DTopics (36) was necessary to robustly decompose single chromatin structures at the Pdx1 locus (Fig. 1d).

The Pdx1 locus can be decomposed into CFMs that can be classified as three types: domains (7/36), loops (10/36) or stripes (12/36, Fig. 1e). We applied 3DTopic to decompose single 3D chromatin structures from other genomic regions in mice (Fig. S1i-q) as well as from human cells (Fig. S1r-t). In all cases, we observe that a small set of CFMs representing domains, loops and stripes is sufficient to decompose single chromatin structures at the sub-Mb scale. Notably, the proportions between these three classes of CFMs are similar between datasets (Fig. S1i-t).

Next, we calculated the frequency with which CFMs appeared in the population of single chromatin traces. Notably, single CFMs are represented in a maximum of 25% of the traces at a minimal weight of 10% (Fig. S1j). This proportion remained low even for other minimal weights (Fig. S1j). Therefore, single CFMs are not predominant in a population of traces, and consequently there is no single trace whose structure looks like that of any single CFM. In the pancreas, the apical TAD loop was the most prominent CFM (15% of traces), followed closely by domain and stripe CFMs (~8-15% of traces at most) (Figs. 1d, S1j).

We then turned to Uniform Manifold Approximation and Projection (UMAP), an unsupervised nonlinear dimension reduction approach, followed by Leiden clustering to dissect the similarities in 3D organization of different chromatin traces (Fig. 1f). Remarkably, single traces segregate into distinct Leiden clusters. Single traces within a single Leiden cluster are characterized, in most cases, by one major CFM (Fig. 1e). Consistently, the average

structure within each Leiden cluster closely resembles that of the major CFM (Fig. 1f, insets). However, we note that multiple CFMs are in all cases required to decompose single traces (Figs. 1g-S1k).

Notably, CFMs can be mutually exclusive. For instance, the Leiden clusters where CFM 2-5 are most prominent (e.g. 0, 6, 12, 18, 8) do not contain CFM0, a motif representing a prominent CTCF loop (Fig. 1e). Conversely, this loop only occurs in a subset of Leiden clusters (2, 14, 15) representing a small proportion of pancreatic cells (~25%, Figs. 1f-g, S1k). Other loop CFMs are even less frequent (Figs. S1j-k) and are not the major CFM in any Leiden cluster (Fig. 1g). The apical TAD loop (CFM13) is present in most Leiden clusters but in a low fraction of cells (~10%, Figs. 1g, S1k). Stripe and domain CFMs were more common in single traces (Fig. S1k) and more often represent the major CFMs in a cluster (Fig. 1g). All in all, our data show that 3D structures of single chromatin traces can be understood in terms of a small number of structural building blocks comprising loops, stripes and domains, with motifs representing multiple contacts (i.e. domains and stripes) appearing more often in single traces.

CFMs rely on loop extrusion and *cis*-regulatory interactions

To better understand the possible mechanisms at the origin of CFMs, we examined the relative contributions of CTCF and *cis*-regulatory interactions. For this, we first generated template maps encoding the interaction patterns expected from each of these mechanisms, derived from pancreas-specific chromatin accessibility and CTCF peaks at the Pdx1 locus, respectively (Figs. 2a-b, *Methods*). These templates were correlated to each CFM to obtain scores estimating the contributions from CTCF and *cis*-regulatory peaks (Fig. 2c). Interactions within most CFMs contained both CTCF and *cis*-regulatory contributions, as reflected by their intermediate scores (Fig. 2c). Remarkably, a subset of CFMs displayed high CTCF and low *cis*-regulatory scores, and corresponded to loop or stripe-like patterns (Fig. 2c). Conversely, a complementary subset of CFMs exhibited high *cis*-regulatory and low CTCF scores, and corresponded to domain-like structures. The size of domain CFMs were in all cases smaller than the TADs identified by Hi-C (Figs. 1e, S2i). Notably, we observed similar results for other genomic regions and in human cells (Fig. S1r-t). This analysis suggests the presence of two distinct mechanisms contributing to the formation of CFMs.

To test this hypothesis, we performed molecular dynamics simulations of loop extrusion (see *Methods*). For this, we used fine-tuned extrusion parameters and a defined set of CTCF sites from existing pancreas ChIP-seq data at the Pdx1 locus (see *Methods*). The ensemble PWD map displayed many of the loops observed in the experimental matrix (Fig. 2e). Remarkably, single traces from loop extrusion simulations could also be decomposed using 36 CFMs, with most of these representing either loops (16/36) or stripes (14/36) (Fig. S2c). We also note the presence of a small number of domain CFMs (3/36) arising from short-range interactions between CTCF sites. Interestingly, less than half of the domain CFMs observed experimentally (3/7) can be explained by a loop extrusion model (Fig. S2e). Overall, these analyses suggest that loop extrusion is sufficient to form loop and stripe CFMs, and that other mechanisms such as *cis*-regulatory interactions contribute to the formation of domain CFMs.

A single set of CFMs can describe the 3D single-cell structures of multiple cell types

Numerous lines of evidence suggest that 3D genome folding is important for regulating different transcriptional programs. Thus, we naturally wondered whether the 3D genome architecture of the Pdx1 locus changes between cell types from multiple mouse tissues. To address this question, we imaged the organization of this locus with Hi-M in six other adult mouse tissues including 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 (Figs. 3a, S3a). Consistently, other genomic loci also displayed 3D changes in

chromatin architecture amongst multiple cell types (Fig. S2b-c).

Next, we wondered whether CFMs derived uniquely from a collection of pancreatic cells could be used to decompose single chromatin traces from cell types in other tissues. To test this hypothesis, we embedded the single traces from the six tissues displayed above using a dictionary of pancreatic CFMs (see Figs. 1e, S1i). Remarkably, this dictionary contained enough conformational diversity to represent the 3D chromatin structures explored by differentiated cell types that diverged in fate during embryonic development (Fig. 3b). We observed similar results in other genomic loci (Fig. S3d).

The use of a common dictionary of motifs, however, does not imply that different cell types use it in the same manner. To shed light into this question, we embedded CFM decompositions from different tissues into a common UMAP and overlaid them with the Leiden clusters from pancreas. For the pancreas, Leiden clusters showed similar densities of single traces (Figs. 3c). However, densities between Leiden clusters are highly inhomogeneous between tissues (Figs. 3d, S3f) and often different from the densities observed in pancreas. We reach similar conclusions for the *Igf2-Ins2* locus (Fig. S3e, S3g, S3h). Overall, these results indicate that single-cells from different tissues explore the same conformational space, but the likelihood with which a conformation occurs depends highly on cell type.

To study if domain, loop and stripe CFMs are differentially impacted, we calculated the relative risk as the ratio of the probability of a CFM in pancreas to the probability of the same CFM in another tissue (Fig. 3e). Notably, all CFM types display changes between tissues, with domain CFMs exhibiting higher risk in the pancreas than in most other tissues (up arrows, Fig. 1e). In contrast, a subset of loops (e.g. CFM1, CFM6) and stripes (e.g. CFM19, CFM20) showed higher consistent risk in the pancreas, while others displayed the opposite trend (e.g. down arrows: CFM2, CFM3, CFM7, CFM13, CFM14, CFM17, CFM23). We observed similar results for other genomic loci (Figs. S3i). All in all, these analyses indicate that single-cells adapt their use of domain, loop and stripe chromatin folding motifs between cell types.

Cell-type specific use of chromatin folding motifs

To further understand why certain CFMs may be more highly represented in certain cell types, we performed sequential RNA-FISH imaging in the pancreas followed by Hi-M (Fig. 4a). This pipeline enabled the reconstruction of 3D PWD maps for the main pancreatic cell types (Fig. 4b). For this, we relied on the detection of insulin (ins1), glucagon (gcg), and somatostatin (sst) transcripts to detect beta, alpha and delta cells (endocrine), while all other cells were classified as exocrine cells. Reconstruction of ensemble PWD maps from each of these species show clear differences in 3D chromatin organization between cell types at the Pdx1 locus (Figs. 4c, S4a-e). Pdx1 is specifically expressed in beta cells, thus we subtracted the PWD maps of beta cells and exocrine cells to detect changes occurring specifically in the cell type where Pdx1 is expressed (Fig. 4d). We observe multiple changes in 3D chromatin organization that occur at short and long-ranges from the Pdx1 promoter. We quantified these changes by performing volcano plot analysis (Fig. 4e), where the fold change in 3D distance is correlated with a p-value drawn from single-cell analysis (see *Methods*). This analysis shows a number of bins displaying statistically-relevant, large changes in 3D distances. Further analysis is needed to identify putative beta cell specific enhancers to relate the observed changes in 3D organization to *cis*-regulatory mechanisms.

Next, we trained a 3DTopic model to predict CFMs based on the single trace structures of exocrine cells, and used this model to decompose beta, alpha and delta cells (Figs. 4f). As before, we obtained domain, loop and stripe CFMs, with similar proportions between the three types (Fig. S4f). To study differential CFM utilization between the different cell types we calculated the relative risk that a CFM appeared in beta, alpha and delta cells with respect to exocrine cells (Fig. 4g). We observed that the patterns in the relative risk ratio map were

different for each cell type, thus each cell type uses CFMs in a different manner. In addition, all CFM types displayed changes, therefore cell type specific changes in 3D chromatin organization may implicate both loop extrusion and *cis*-regulatory interactions. Interestingly, the smallest overall relative risk differences seem to appear between beta cells and exocrine cells. However, further analysis is required to assign cell-type specific enhancers in order to correlate differential use of specific CFMs to *cis*-regulatory elements differentially occupied between cell types.

Disease onset affects the use of domain, loop and stripe motifs

In the previous section, we established that specific sets of CFMs preferentially change between pancreatic cell types, particularly between exo- and endocrine cells. Recently, it was shown that diet-induced metabolic deregulation leads to changes in endocrine-specific CTCF levels and accessible chromatin regions (Wang et al., 2022) in a well-established mouse model of obesity and type 2 diabetes (T2D). To investigate whether these deregulations lead to differential changes in CFM utilization between pancreatic cell types, we performed RNA-FISH followed by Hi-M experiments in 14-weeks old mice fed with a normal diet (ND) or with a high-fat diet (HFD, Figs. 5a, S5a-f). As expected, islet size and number of beta cells per islet increased considerably in HFD with respect to ND mice (Fig. 5b) (Kemkem et al., 2020).

The overall 3D organization at the Pdx1 locus displayed similar patterns of interactions in HFD and ND, and both conditions displayed noticeable differences between cell types (Fig. 5c). Surprisingly, metabolic deregulation led to changes in 3D chromatin organization in both endocrine and exocrine cells (Fig. 5d). Exocrine cells were the least affected, and mainly displayed higher distances within the Pdx1 TAD and shorter distances with the neighboring TAD. Delta cells displayed distributed distance changes, whereas beta and alpha cells showed a marked increase in intra and inter-TAD distances (Figs. 5d-e). Notably, this translates into a general loss of proximity between the Pdx1 promoter with any putative *cis*-regulatory regions within the Pdx1 TAD in obese/diabetic mice (Figs. 5d-e). Thus, diet-induced metabolic perturbation leads to cell-type specific changes in 3D chromatin organization at the Pdx1 locus.

To shed light into the contributions of loop extrusion and *cis*-regulatory interactions in these cell-type specific changes, we decomposed single traces from pancreatic cell types in HFD and ND mice using the base of CFMs from our previous training (Fig. S4f, S5l). We observe that both loop/stripe and domain CFMs are perturbed according to relative risk analysis (Fig. 5f). This supports the notion that both loop extrusion and *cis*-regulatory interactions are perturbed during metabolic deregulation, consistent with previous observations indicating changes in CTCF expression and in accessibility peaks (Wang et al., 2022). Overall, these results indicate that multiple chromatin architectural elements may be perturbed in single cells within diseased tissues.

Discussion

Here, we used chromatin tracing to study chromatin organization across cell types and tissues. Our main finding is that the structure of chromatin in single cells can be described by discrete 3D folding motifs. These basic units of chromatin organization comprise short-range domains, long-range loops, and stripes. Loop CFMs typically involve long-range proximity between single distal elements. In contrast, domains and stripes are formed by the coalescence of multiple genomic loci in single cells. The former implicate short-range (25-100kb) interactions, while the latter requires distal long-range contacts.

A second important finding is that multiple cell types from different organs explore the same CFM conformational space. In other words, a single dictionary of folding motifs can be used to represent single cell conformations of very different cell types. This suggests, provocatively, that the same mechanisms of chromatin folding are available to different cell types. Critically, the frequency with which each motif is used varies between cell types, consistent with cells regulating how these folding mechanisms act in a cell type specific manner.

The detection of multiple motif types describing the structures of single cells offers the unique opportunity to quantify the frequency at which these motifs appear in single cells within tissues and to dissect how they vary between cell types. Our quantifications show that long-range loops occur in a small number of cells at any given time (~10-20%), comparable to measurements performed using time-lapse imaging (Gabriele et al., 2022; Mach et al., 2022). Our results show that CTCF loops between TAD borders or between CTCF sites within TADs are equally infrequent in tissues. These frequencies depend on cell type, and therefore may represent a mechanism by which cells tune architectural features to favor cell-type specific *cis*-regulatory interactions (Karpinska & Oudelaar, 2023).

An important advantage of our approach, in contrast to time-lapse imaging, is that it visualizes spatial colocalization between multiple genomic targets at once. This allowed us to detect and quantify domain and stripe CFMs, both of which involve the short- or long-range co-localization of multiple targets. These motifs were surprisingly as common as loops in single cells, and represent a fundamental building block of chromatin organization.

Our observation of stripe CFMs is consistent with recent studies reporting the visualization of stripes in single cells (Hafner et al., 2023; Hung et al., 2023), inconsistent with the reel-in model, and rather supporting a model whereby multiple CTCF loops stack in 3D. 3DTopic allowed us to automatically detect and quantify stripe patterns from single chromatin traces, and show that these patterns are widespread: they represent a large proportion of the folding motifs we observed, they are present in all cell types and organs we investigated, and their utilization varies between cell types. Overall, these observations support the idea that multiple loops between CTCF sites are common and participate in cell type specific transcriptional programs.

Domain CFMs involve multiple genomic loci interacting together at short-ranges (25-100kb). Thus, domain CFMs are considerably smaller than typical mammalian TADs, which typically range from hundreds of kilobases (kb) to a few megabases (Mb) in size. These results imply that TADs are not formed often enough in single cells to elicit TAD-scale CFMs, consistent with previous results (Bintu et al., 2018; Cattoni et al., 2017; Finn & Misteli, 2019; Luppino et al., 2019). In contrast, domain CFMs represent shorter-scale clusters of loci occurring in single cells, and may be related to nucleosome clusters (clutches) (Ricci et al., 2015), or chromatin nanodomains (Szabo et al., 2020). Our simulations show that loop extrusion alone can generate domain CFMs by clustering of proximal CTCF sites. However, experimentally we observe many more domain CFMs than predicted by loop extrusion alone, suggesting that additional mechanisms leading to short-range loops may be involved, such as *cis*-regulatory interactions. Further support for this hypothesis comes from our observed changes in domain CFM utilization between cell types and tissues. This hypothesis is also consistent with recent studies showing that enhancers and promoters contact frequently in micro-compartments (Espinola et al., 2021; Goel et al., 2023; Oudelaar et al., 2018).

Importantly, loop extrusion is sufficient to generate many of the loop and stripe CFMs observed experimentally, suggesting that this is the main mechanism to generate this motif type. Previous studies in classical loci (e.g. beta-globin) showed that CTCF loops can be cell-type specific (Hou et al., 2010; Phillips & Corces, 2009). Our finding that loop and stripe CFMs change dramatically between cell types and tissues generalizes this observation to multiple cell types from different tissues. It is unclear, however, what may be the mechanism underpinning these changes. CTCF, cohesin, WAPL and NIPBL are general factors present in all cell types, and affect chromatin organization genome-wide. Thus, locus-specific modulation of loop extrusion would require cell-type and locus-specific modulation of cohesin loading, unloading or blocking by affecting the binding of NIPBL, WAPL or CTCF. This would be best achieved by transcription factors regulating chromatin accessibility and binding of loop extrusion components. Alternatively, recent studies proposed that changes in CTCF loops arise from cell-type specific enhancer-promoter interactions rather than to changes in CTCF binding (Hua et al., 2021; Kragesteen et al., 2018). In this model, activation of promoters and enhancers would result in increased loop extrusion activity through increased loading of cohesin (Hua et al., 2021). Clearly, more work will be necessary to investigate the mechanisms underlying cell-type specific utilization of loops and stripe CFMs, which may be locus- or model-specific.

Metabolic stress induced by a high-fat diet leads to obese and diabetic mice, and to morphological and functional changes in specific pancreatic cell types (e.g. beta cells). Surprisingly, these are accompanied by changes in CTCF levels and chromatin accessibility in endocrine cells that are thought to contribute to beta cell dysfunction (Wang et al., 2022). Our results additionally show that HFD influences chromatin organization in multiple pancreatic cell types. Particularly, beta cells display a general loss of interactions of the Pdx1 promoter with putative enhancers, suggesting a mechanism that would deregulate Pdx1 expression and therefore beta cell identity and function (Gao et al., 2014; Guo et al., 2013; Sachdeva et al., 2009). These changes involve multiple CFM types, suggesting that both loop extrusion and *cis*-regulatory interactions may be affected. These would be consistent with the observed down-regulation of CTCF levels specifically in beta cells. More work is required to understand the specific consequences of these structural changes to the regulation of PDX1 transcription. Unexpectedly, we also observed smaller but noticeable changes in 3D chromatin organization in exocrine cells, which are not involved in blood glucose homeostasis. We note, however, that HFD-induced obesity and diabetes can lead to other pancreatic malfunctions such as exocrine pancreatic insufficiency (Alkaade & Vareedayah, 2017), which could in principle also arise from changes in exocrine-specific transcriptional programs.

The last decade has seen a renewed interest in mapping epigenetic, accessibility and transcriptional changes across cell types and tissues to ultimately understand how the same DNA code can be interpreted to execute different transcriptional programs. Here, 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. 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 handling

Animal studies were conducted according to the European guidelines for animal welfare (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 h light/12 h dark cycle and were given chow and water ad libitum. C57BL/6J mice were purchased from Janvier-SAS (Le Genest-St-Isle, France). Mice were fed with normal diet (ND) until age 6 weeks and then fed with either ND or HFD (63% energy from fat) (Safe Diets, France) for 14 weeks. No data were excluded unless animals died during experimentation. Intraperitoneal glucose tolerance test (IPGTT) was 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 (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 h 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 (Bioptechs) 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 min at room temperature. Slides were left in water for 5-10 min and rinsed 2x10 min in a water bath with agitation. A solution containing 0.5% glutaraldehyde (Sigma) in PBS was added for 30 min and rinsed with water. The slides were then coated with 0.1 mg/mL poly-D-lysine (Sigma) in PBS for 1 h and incubated O/N with 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.

Imaging system

Experiments were performed on a home-made imaging setup built on a RAMM modular microscope system (Applied Scientific Instrumentation) coupled to a microfluidic device, as the one described previously (Barho et al., 2022; Cardozo Gizzi et al., 2020).

RNA-FISH

Sequential RNA-FISH Libraries were constructed following the procedure described in (Chen et al., 2015) with the library design script available here 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 30bp allowing 20bp overlap between adjacent probes to maximize the number of probes for a gene. Each probe is 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.

Slides were taken out of the freezer and kept on the bench for 1h to dry. Tissues were post-fixed 10 min with PFA 4% and washed with PBS. Slides were incubated with cold 70% EtOh. The day after, slides were rehydrated in 2xSSC for 5 min at RT and incubated for 3f in 30% formamide wash buffer at 37 °C. Slides were placed upside down in glass petridishes in contact with 2uL of 5 to 10 ug/uL ssDNA library, in 20 uL of HB (HB = 30% Formamide, 10% dextran sulfate, 2X SSC, 500uL tRNA stock (20 mg/ml), 100 µl of RVC stock (200 mM)) and incubated in a humidity-controlled 37 °C incubator for 36 h. After staining slides were washed two times for 30 min each with a 30% formamide wash buffer at 45°C and stored in 2xSSC at

4°C until imaging.

DNA-FISH

Oligopaint libraries from a public database (http://genetics.med.harvard.edu/oligopaints), consisting of unique 35/45-mer sequences with genome homology. Hi-M probe sets were designed for each locus and amplified following the methodology described previously (Cardozo Gizzi et al., 2020). In total, we designed three Hi-M libraries. One encompassing the Pdx1 gene containing 25 barcodes of approximately 25kb each. We designed two more Hi-M libraries to investigate other genes of interest. One encompassing the Ins2/Igf2 genes and containing 25 barcodes targeting regions of approximately 6kb each. The other library encompassed the Isl1 gene and consisted of 24 barcodes labeling regions of approximately 26kb each. The fiducial library used for drift correction comprises three sequences specifically designed to target minor and major satellites in the mouse (Guenatri et al., 2004; Jagannathan et al., 2018). Barcode coordinates and sequences are available in Supplementary Table Sx.

Slides were treated with RNAse A at 200 µg/ml in PBS for 1h at RT and incubated in Sodium Citrate 10mM 5 min at RT. Next, slides were incubated with Sodium Citrate 10mM 25 min at 80°C in a water bath and left on the bench for 1h. Slides were washed with 2xSSC 5min and incubated in 50% Formamide wash buffer 2h at RT. Slides were placed upside down in glass petri dishes in contact with 2uL of 5 to 10 ug\uL library, 1uL of 100uM the fiducial library, in 20uL of FHB (FHB = 50% Formamide, 10% dextran sulfate, 2X SSC, Salmon Sperm DNA 0.5 mg/mL). Slides were incubated 3h at 45°C in the water bath and a heat shock was performed at 85°C 5min in the heat block. Slides were incubated in a humidity-controlled 37°C incubator o/n. The next day, tissues were washed under agitation at 80 rpm, 2 times 40 min in 50 % Formamide wash buffer, 1 time 20 min in 40% Formamide wash buffer, 1 time 20 min in 30% Formamide wash buffer, 1 time 20 min in 20% Formamide wash buffer, 1 time 20 min 10% Formamide wash buffer and 1 time 20 min in 2x SSC. Slides were post fixed with 4% PFA 10 min at RT and stored in 2xSSC at 4°C until imaging.

Acquisition of Hi-M datasets

Slides were mounted in an FCS2® flow chamber (Bioptechs, USA). ~20-30 regions were selected and imaged using two regions of interest (ROI 200x200µm²). A mixture containing an adapter complementary to the reverse primer sequence was injected for “mask0” imaging (25nM of adapter to the reverse primer, 2× SSC, 50% v:v formamide). Tissues were washed with 2.8mL of wash buffer solution (2× SSC, 50% v:v formamide) and incubated for 15 min with a readout hybridization mixture (25nM of Atto-488 imager probe complementary to the fiducial library, 25nM Alexa-SS-647 probe complementary to the “mask0” adapter, 2× SSC, 50% v:v formamide). Tissues were then washed with 2.8mL of wash buffer solution (2× SSC, 50% v:v formamide) and flushed with 1.5mL of 0.5 µg.ml−1 of DAPI solution in 2× SSC to stain the nuclei. A stack of images (z-step size of 250 nm and a total area of 20 µm) was acquired for DAPI (405 nm), the fiducial library tagged with Atto488 (488nm) and the “mask0” tagged with Alexa647 (640 nm). After imaging, the chambre was flushed with 1mL of chemical bleach buffer (2× SCC, 50mM TCEP hydrochloride) to bleach the “mask0” imager. The sample was then sequentially hybridized as follows. A solution containing the imager oligo (25nM Alexa-SS-647 probe, 2× SSC, 50% v:v formamide) was injected and incubated for 15 min. Tissues were then washed with 2mL of wash buffer and with 1 mL of 2x SSC before injecting the imaging buffer (IB, 1xPBS, 5% w/v glucose, 0.5 mg/mL glucose oxidase and 0.05 mg/mL catalase). In each cycle, fiducials and readout probes were sequentially imaged with 488 nm and 647 nm excitation lasers respectively. After imaging, the fluorescent tag of the readout probes was cleaved with 1mL the reducing agent tris(2-carboxyethyl)phosphine (TCEP), as previously described. Finally, the samples were washed with 1 mL of 2× SSC before a new hybridization cycle started.

Image processing

DCIMG files were converted to TIFF using proprietary software from Hamamatsu. TIFF images were then deconvolved using Huygens Professional 21.04 (Scientific Volume Imaging, https://svi.nl). The analysis was performed using our pyHiM analysis pipeline (https://pyhim.readthedocs.io/en/latest/) with release 0.7.

Chip-seq, ATAC-seq, CUT&Tag and RNA-seq data processing

ATAC-seq and CTCF data were reanalyzed from raw data (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 CUT&Tag data were downloaded from https://data.mendeley.com/datasets/mwgxv7m927/2 (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.

Topic Modeling

3DTopic uses Latent Dirichlet Allocation (LDA) to decompose single cell 3D structures into a discrete number of folding topics called Chromatin Folding Motifs (CFMs). To implement 3D Topic on the scHi-M data, we used the LDA implementation of the scikit-learn Python package (Pedregosa et al. 2011). As described in the cisTopic approach (Bravo González-Blas et al. 2019), a scHi-M contact matrix is computed for each individual cell and is considered as a single "document". This document consists of a set of "words" representing the binary pairwise interactions between pairs of loci. Since LDA does not handle 2D data, contact matrices are linearized into a 1D vector (only the first half of the map is used, without the diagonal). For example, in our specific context, a 25x25 contact map is transformed into a 1D vector composed of 300-bit 'words'. A word is equal to 1 if a contact was detected, 0 if no contact was detected or the data was missing. Therefore, LDA decomposition is used to group contacts that are most likely to be detected together and to generate CFMs that are characteristic of the cells under study.

In this work, LDA is applied to scHiM contact data with a minimum barcode detection of 65% and decomposed into a set of 36 CFMs. This set of 36 CFMs is in turn used to decompose the scHi-M data from a 300-bit 1D vector describing contact values into 36 values describing the weight associated with each CFM.

Polymer simulation

The loop extrusion simulations were performed using the OpenMM-based polychrom package (Imakaev et al., 2019) developed by the Open Chromosome Collective (Open2C) using an approach similar to that used in Fudenberg et. al. (Fudenberg et al., 2016). First, the dynamics of cohesins/loop extruding factors (LEFs) were simulated on a one-dimensional lattice with CTCF sites and their LEF capture and release rates assigned based on ChIP-seq data (Table 2). LEF positions were recorded over time and were used to rapidly construct simple contact frequency maps based on the bridging action of LEFs, which then facilitated manual optimization of the loop extrusion parameter values given in Table 1. Next, the recorded LEF positions were used to mimic loop extrusion on a flexible polymer whose non-bonded interactions were dictated by a simple polynomial repulsive potential during molecular dynamics simulations (parameters provided in Table 3). Polymer conformations from the last time steps of the simulation after thermalization were saved and used to

construct pairwise distance maps.

Data availability

Single nucleus pairwise distance matrices as well as XYZ coordinates of chromatin traces generated in this study were deposited at our Open Science Framework project with DOI: XXX. The list of previously published datasets used in this study is provided in Supplementary Table SX.

Code availability

The code used for post-processing Hi-M matrices is accessible at X. For a permanent link, see DOI:XXX. 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, and an archived version at https://zenodo.org/record/6379944 (DOI: 10.5281/zenodo.6379944). Hi-M data were analyzed using pyHiM release 0.6, available at https://github.com/marcnol/pyHiM.

References

Alkaade, S., & Vareedayah, A. A. (2017). A primer on exocrine pancreatic insufficiency, fat malabsorption, and fatty acid abnormalities. *The American Journal of Managed Care*, *23*(12 Suppl), S203–S209.

Barho, F., Fiche, J.-B., Bardou, M., Messina, O., Martiniere, A., Houbron, C., & Nollmann, M. (2022). Qudi-HiM: an open-source acquisition software package for highly multiplexed sequential and combinatorial optical imaging. In *Open Research Europe* (Vol. 2, p. 46). https://doi.org/10.12688/openreseurope.14641.2

Bintu, B., Mateo, L. J., Su, J.-H., Sinnott-Armstrong, N. A., Parker, M., Kinrot, S., Yamaya, K., Boettiger, A. N., & Zhuang, X. (2018). Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science*, *362*(6413). https://doi.org/10.1126/science.aau1783

Bravo González-Blas C, Minnoye L, Papasokrati D, Aibar S, Hulselmans G, Christiaens V, Davie K, Wouters J, Aerts S. (2019). cisTopic: cis-regulatory topic modeling on single-cell ATAC-seq data. *Nat Methods*. 2019 http://doi: 10.1038/s41592-019-0367-1

Cardozo Gizzi, A. M., Cattoni, D. I., Fiche, J.-B., Espinola, S. M., Gurgo, J., Messina, O., Houbron, C., Ogiyama, Y., Papadopoulos, G. L., Cavalli, G., Lagha, M., & Nollmann, M. (2019). Microscopy-Based Chromosome Conformation Capture Enables Simultaneous Visualization of Genome Organization and Transcription in Intact Organisms. *Molecular*

Cell. https://doi.org/10.1016/j.molcel.2019.01.011

Cardozo Gizzi, A. M., Espinola, S. M., Gurgo, J., Houbron, C., Fiche, J.-B., Cattoni, D. I., & Nollmann, M. (2020). Direct and simultaneous observation of transcription and chromosome architecture in single cells with Hi-M. *Nature Protocols*, *15*(3), 840–876.

Cattoni, D. I., Cardozo Gizzi, A. M., Georgieva, M., Di Stefano, M., Valeri, A., Chamousset, D., Houbron, C., Déjardin, S., Fiche, J.-B., González, I., Chang, J.-M., Sexton, T., Marti-Renom, M. A., Bantignies, F., Cavalli, G., & Nollmann, M. (2017). Single-cell absolute contact probability detection reveals chromosomes are organized by multiple low-frequency yet specific interactions. *Nature Communications*, *8*(1), 1753.

Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S., & Zhuang, X. (2015). RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science*, *348*(6233), aaa6090.

Dixon, J. R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J. S., & Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, *485*(7398), 376–380.

Espinola, S. M., Götz, M., Bellec, M., Messina, O., Fiche, J.-B., Houbron, C., Dejean, M., Reim, I., Cardozo Gizzi, A. M., Lagha, M., & Nollmann, M. (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), 477–486.

Finn, E. H., & Misteli, T. (2019). Molecular basis and biological function of variability in spatial genome organization. *Science*, *365*(6457). https://doi.org/10.1126/science.aaw9498

Finn, E. H., Pegoraro, G., Brandão, H. B., Valton, A.-L., Oomen, M. E., Dekker, J., Mirny, L., & Misteli, T. (2019). Extensive Heterogeneity and Intrinsic Variation in Spatial Genome Organization. *Cell*, *176*(6), 1502–1515.e10.

Flyamer, I. M., Gassler, J., Imakaev, M., Brandão, H. B., Ulianov, S. V., Abdennur, N., Razin, S. V., Mirny, L. A., & Tachibana-Konwalski, K. (2017). Single-nucleus Hi-C reveals unique chromatin reorganization at oocyte-to-zygote transition. *Nature Publishing Group*. https://doi.org/10.1038/nature21711

Fudenberg, G., Imakaev, M., Lu, C., Goloborodko, A., Abdennur, N., & Mirny, L. A. (2016). Formation of Chromosomal Domains by Loop Extrusion. *Cell Reports*, *15*(9), 2038–2049.

Furlong, E. E. M., & Levine, M. (2018). Developmental enhancers and chromosome topology. *Science*, *361*(6409), 1341–1345.

Gabriele, M., Brandão, H. B., Grosse-Holz, S., Jha, A., Dailey, G. M., Cattoglio, C., Hsieh, T.-H. S., Mirny, L., Zechner, C., & Hansen, A. S. (2022). Dynamics of CTCF- and cohesin-mediated chromatin looping revealed by live-cell imaging. *Science*, *376*(6592), 496–501.

Gao, T., McKenna, B., Li, C., Reichert, M., Nguyen, J., Singh, T., Yang, C., Pannikar, A., Doliba, N., Zhang, T., Stoffers, D. A., Edlund, H., Matschinsky, F., Stein, R., & Stanger, B. Z. (2014). Pdx1 maintains β cell identity and function by repressing an α cell program. *Cell Metabolism*, *19*(2), 259–271.

Giorgetti, L., Galupa, R., Nora, E. P., Piolot, T., Lam, F., Dekker, J., Tiana, G., & Heard, E. (2014). Predictive polymer modeling reveals coupled fluctuations in chromosome conformation and transcription. *Cell*, *157*(4), 950–963.

Goel, V. Y., Huseyin, M. K., & Hansen, A. S. (2023). Region Capture Micro-C reveals coalescence of enhancers and promoters into nested microcompartments. *Nature Genetics*, *55*(6), 1048–1056.

Guenatri, M., Bailly, D., Maison, C., & Almouzni, G. (2004). Mouse centric and pericentric satellite repeats form distinct functional heterochromatin. *The Journal of Cell Biology*, *166*(4), 493–505.

Guo, S., Dai, C., Guo, M., Taylor, B., Harmon, J. S., Sander, M., Robertson, R. P., Powers, A. C., & Stein, R. (2013). Inactivation of specific β cell transcription factors in type 2 diabetes. *The Journal of Clinical Investigation*, *123*(8), 3305–3316.

Hafner, A., Park, M., Berger, S. E., Murphy, S. E., Nora, E. P., & Boettiger, A. N. (2023). Loop stacking organizes genome folding from TADs to chromosomes. *Molecular Cell*, *83*(9), 1377–1392.e6.

Hou, C., Dale, R., & Dean, A. (2010). Cell type specificity of chromatin organization mediated by CTCF and cohesin. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(8), 3651–3656.

Hua, P., Badat, M., Hanssen, L. L. P., Hentges, L. D., Crump, N., Downes, D. J., Jeziorska, D. M., Oudelaar, A. M., Schwessinger, R., Taylor, S., Milne, T. A., Hughes, J. R., Higgs, D. R., & Davies, J. O. J. (2021). Defining genome architecture at base-pair resolution. *Nature*, *595*(7865), 125–129.

Hung, T.-C., Kingsley, D. M., & Boettiger, A. (2023). Boundary stacking interactions enable cross-TAD enhancer-promoter communication during limb development. In *bioRxiv*. https://doi.org/10.1101/2023.02.06.527380

Imakaev, M., Goloborodko, A., & hbbrandao. (2019). *mirnylab/polychrom: v0.1.0*. Zenodo. https://doi.org/10.5281/ZENODO.3579473

Jagannathan, M., Cummings, R., & Yamashita, Y. M. (2018). A conserved function for pericentromeric satellite DNA. *eLife*, *7*. https://doi.org/10.7554/eLife.34122

Karpinska, M. A., & Oudelaar, A. M. (2023). The role of loop extrusion in enhancer-mediated gene activation. *Current Opinion in Genetics & Development*, *79*, 102022.

Kemkem, Y., Nasteska, D., de Bray, A., Bargi-Souza, P., Peliciari-Garcia, R. A., Guillou, A., Mollard, P., Hodson, D. J., & Schaeffer, M. (2020). Maternal hypothyroidism in mice influences glucose metabolism in adult offspring. *Diabetologia*, *63*(9), 1822–1835.

Kragesteen, B. K., Spielmann, M., Paliou, C., Heinrich, V., Schöpflin, R., Esposito, A., Annunziatella, C., Bianco, S., Chiariello, A. M., Jerković, I., Harabula, I., Guckelberger, P., Pechstein, M., Wittler, L., Chan, W.-L., Franke, M., Lupiáñez, D. G., Kraft, K., Timmermann, B., … Andrey, G. (2018). Dynamic 3D chromatin architecture contributes to enhancer specificity and limb morphogenesis. *Nature Genetics*, *50*(10), 1463–1473.

Liu, C., Wang, M., Wei, X., Wu, L., Xu, J., Dai, X., Xia, J., Cheng, M., Yuan, Y., Zhang, P., Li, J., Feng, T., Chen, A., Zhang, W., Chen, F., Shang, Z., Zhang, X., Peters, B. A., & Liu, L. (2019). An ATAC-seq atlas of chromatin accessibility in mouse tissues. *Scientific Data*, *6*(1), 65.

Lupiáñez, D. G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., Horn, D., Kayserili, H., Opitz, J. M., Laxova, R., Santos-Simarro, F., Gilbert-Dussardier, B., Wittler, L., Borschiwer, M., Haas, S. A., Osterwalder, M., Franke, M., Timmermann, B., Hecht, J., … Mundlos, S. (2015). Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell*, *161*(5), 1012–1025.

Luppino, J. M., Park, D. S., Nguyen, S. C., Lan, Y., Xu, Z., & Joyce, E. F. (2019). Cohesin promotes stochastic domain intermingling to ensure proper regulation of boundary-proximal genes. In *bioRxiv* (p. 649335). https://doi.org/10.1101/649335

Mach, P., Kos, P. I., Zhan, Y., Cramard, J., Gaudin, S., Tünnermann, J., Marchi, E., Eglinger, J., Zuin, J., Kryzhanovska, M., Smallwood, S., Gelman, L., Roth, G., Nora, E. P., Tiana, G., & Giorgetti, L. (2022). Cohesin and CTCF control the dynamics of chromosome folding. *Nature Genetics*, *54*(12), 1907–1918.

Mateo, L. J., Murphy, S. E., Hafner, A., Cinquini, I. S., Walker, C. A., & Boettiger, A. N. (2019). Visualizing DNA folding and RNA in embryos at single-cell resolution. *Nature*, *568*(7750), 49–54.

Michau, A., Hodson, D. J., Fontanaud, P., Guillou, A., Espinosa-Carrasco, G., Molino, F., Peters, C. J., Robinson, I. C., Le Tissier, P., Mollard, P., & Schaeffer, M. (2016). Metabolism Regulates Exposure of Pancreatic Islets to Circulating Molecules In Vivo. *Diabetes*, *65*(2), 463–475.

Nora, E. P., Lajoie, B. R., Schulz, E. G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N. L., Meisig, J., Sedat, J., Gribnau, J., Barillot, E., Blüthgen, N., Dekker, J., & Heard, E. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature*, *485*(7398), 381–385.

Oudelaar, A. M., Davies, J. O. J., Hanssen, L. L. P., Telenius, J. M., Schwessinger, R., Liu, Y., Brown, J. M., Downes, D. J., Chiariello, A. M., Bianco, S., Nicodemi, M., Buckle, V. J., Dekker, J., Higgs, D. R., & Hughes, J. R. (2018). Single-allele chromatin interactions identify regulatory hubs in dynamic compartmentalized domains. *Nature Genetics*, *50*(12), 1744–1751.

Pedregosa, F., Varoquaux, G.,Gramfort, A. and Michel, V., Thirion, B. and Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research,* 12, 2825-2830.

Peslak, S. A., Demirci, S., Chandra, V., Ryu, B., Bhardwaj, S. K., Jiang, J., Rupon, J. W., Throm, R. E., Uchida, N., Leonard, A., Essawi, K., Bonifacino, A. C., Krouse, A. E., Linde, N. S., Donahue, R. E., Ferrara, F., Wielgosz, M., Abdulmalik, O., Hamagami, N., … Blobel, G. A. (2023). Forced enhancer-promoter rewiring to alter gene expression in animal models. *Molecular Therapy. Nucleic Acids*, *31*, 452–465.

Phillips, J. E., & Corces, V. G. (2009). CTCF: master weaver of the genome. *Cell*, *137*(7), 1194–1211.

Rao, S. S. P., Huntley, M. H., Durand, N. C., Stamenova, E. K., Bochkov, I. D., Robinson, J. T., Sanborn, A. L., Machol, I., Omer, A. D., Lander, E. S., & Aiden, E. L. (2015). A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell*, *162*(3), 687–688.

Ricci, M. A., Manzo, C., García-Parajo, M. F., Lakadamyali, M., & Cosma, M. P. (2015). Chromatin fibers are formed by heterogeneous groups of nucleosomes in vivo. *Cell*, *160*(6), 1145–1158.

Sachdeva, M. M., Claiborn, K. C., Khoo, C., Yang, J., Groff, D. N., Mirmira, R. G., & Stoffers, D. A. (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), 19090–19095.

Schoenfelder, S., & Fraser, P. (2019). Long-range enhancer-promoter contacts in gene expression control. *Nature Reviews. Genetics*, *20*(8), 437–455.

Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F. d. R., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A., & Cavalli, G. (2012). Three-dimensional folding and functional organization principles of the Drosophila genome. *Cell*, *148*(3), 458–472.

Shen, Y., Yue, F., McCleary, D. F., Ye, Z., Edsall, L., Kuan, S., Wagner, U., Dixon, J., Lee, L.,

Lobanenkov, V. V., & Ren, B. (2012). A map of the cis-regulatory sequences in the mouse genome. *Nature*, *488*(7409), 116–120.

Stevens, T. J., Lando, D., Basu, S., Atkinson, L. P., Cao, Y., Lee, S. F., Leeb, M., Wohlfahrt, K. J., Boucher, W., O'Shaughnessy-Kirwan, A., Cramard, J., Faure, A. J., Ralser, M., Blanco, E., Morey, L., Sansó, M., Palayret, M. G. S., Lehner, B., Di Croce, L., … Laue, E. D. (2017). 3D structures of individual mammalian genomes studied by single-cell Hi-C. *Nature*, *544*(7648), 59–64.

Szabo, Q., Donjon, A., Jerković, I., Papadopoulos, G. L., Cheutin, T., Bonev, B., Nora, E. P., Bruneau, B. G., Bantignies, F., & Cavalli, G. (2020). Regulation of single-cell genome organization into TADs and chromatin nanodomains. *Nature Genetics*, *52*(11), 1151–1157.

Tena, J. J., & Santos-Pereira, J. M. (2021). Topologically Associating Domains and Regulatory Landscapes in Development, Evolution and Disease. *Frontiers in Cell and Developmental Biology*, *9*, 702787.

Wang, R. R., Qiu, X., Pan, R., Fu, H., Zhang, Z., Wang, Q., Chen, H., Wu, Q. Q., Pan, X., Zhou, Y., Shan, P., Wang, S., Guo, G., Zheng, M., Zhu, L., & Meng, Z. X. (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

Xiao, J. Y., Hafner, A., & Boettiger, A. N. (2021). How subtle changes in 3D structure can create large changes in transcription. *eLife*, *10*. https://doi.org/10.7554/eLife.64320

Yang, J.-K., Lu, J., Yuan, S.-S., Asan, Cao, X., Qiu, H.-Y., Shi, T.-T., Yang, F.-Y., Li, Q., Liu, C.-P., Wu, Q., Wang, Y.-H., Huang, H.-X., Kayoumu, A., Feng, J.-P., Xie, R.-R., Zhu, X.-R., Liu, C., Yang, G.-R., … Xu, A. (2018). From Hyper- to Hypoinsulinemia and Diabetes: Effect of KCNH6 on Insulin Secretion. *Cell Reports*, *25*(13), 3800–3810.e6.

Zuin, J., Roth, G., Zhan, Y., Cramard, J., Redolfi, J., Piskadlo, E., Mach, P., Kryzhanovska, M., Tihanyi, G., Kohler, H., Eder, M., Leemans, C., van Steensel, B., Meister, P., Smallwood, S., & Giorgetti, L. (2022). Nonlinear control of transcription through

enhancer-promoter interactions. *Nature*, *604*(7906), 571–577.

Acknowledgements

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-BioImaging, 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.

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. prepared and characterized mice. I.A. and D.H. participated in experimental design. A.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

The authors declare no competing interests.

Figure legends

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

a Schematic illustrating the imaging-based strategy used to study chromosome conformation at single cell level in cryosectioned mouse tissues (Hi-M). **b** Top: Hi-M pairwise distance (PWD) matrix along the Pdx1 locus (chr5:146871445-147499625) in mouse pancreas (mm10) constructed from 17362 traces (n=4 experiments, 2 different mice). Middle : barcodes used for Hi-M sequential imaging are represented as color-coded boxes. CTCF sites are shown as black and white triangles, representing forward and reverse orientations, respectively. TADs are represented with rectangles. Genes locations are displayed along the locus. Bottom: insulation score derived from Hi-M data with different window sizes (1, 2 and 3 bins), and domainogram (see *Methods*). *N* represents the number of biological replicates, and *n* the number of traces. **c** Schematic illustrating 3DTopic workflow and application for Hi-M single-cell data. Single cell 3D structures are used to determine a set of CFMs and each cell is iteratively assigned to a probability distribution for each of the CFM. **d** Probability distribution showing the decomposition of pancreatic cells in 36 CFMs. Red and blue represent high and low probability, respectively. **e** Subsets of CFMs, selected from the complete gallery of 36 CFMs of one decomposition of the Pdx1 locus on pancreatic cells (see *Methods*). CFMs are categorized into three distinct groups according to the structure they describe (Domain, Loop and Stripe). **f** 3DTopic UMAP, derived from CFM decomposition of pancreatic cells, color-coded by Leiden clusters. On the right, two examples of matrices originating from an individual cluster are shown. The top matrix shows 3DTopic reconstruction, taking into account the contribution of CFMs within this cluster, while the bottom matrix illustrates the major CFM from this cluster. **g** Matrix representing the normalized frequency of CFMs occurring in more than 10% in cell decompositions within each Leiden cluster. A selection of CFMs are shown below, ‘S’: stripes, ‘D’: domains, ‘L’: loops.

Figure 2. CTCF and accessible regions are required to fold CFMs.

a Co-accessibility matrix derived from ATAC profile for the Pdx1 locus in pancreas. Red and blue represent high and low Co-accessibility, respectively. ATAC-seq profile from pancreatic tissue is displayed below with gene locations and genomic coordinates. **b** Co-binding matrix derived from CTCF profile for the Pdx1 locus in pancreas. Red and blue represent high and low Co-binding, respectively. ChiP-seq profile for CTCF from pancreatic tissue is displayed below with gene locations and genomic coordinates. **c** Association scores between features (ATAC and CTCF) and CFMs for the Pdx1 locus. Black and green curves indicate the association score for ATAC and CTCF, respectively. CFMs are sorted based on decreasing CTCF scores, with Domain CFMs numbers indicated in bold. **d** Examples of CFM matrices associated with high CTCF score (left), high ATAC score and medium CTCF score (middle), and high ATAC score and low CTCF score (right). **e** Simulated Hi-M matrix along the Pdx1 locus using the loop extrusion model (see *Methods*). **f** Subset of CFMs, selected from the complete gallery of 36 CFMs of one decomposition of the Pdx1 locus from simulation.

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

a Median Hi-M PWD matrices along the Pdx1 locus (chr5:146871445-147499625) in 7 different mouse tissues. *N* represents the number of biological replicates, and *n* the number of traces. **b** Probability distribution showing the decomposition of 7 different mouse tissues in 36 CFMs. Red and blue represent high and low probability, respectively. **c** Density distribution of pancreatic cells in the UMAP landscape. Dark black and light black represent high and low density, respectively. Leiden clusters are overlaid on the UMAP. **d** Density distribution in the UMAP landscape for six other mouse tissues overlaid with the Leiden clusters. **e** Relative

risks for all tissues and for all CFMs for the Pdx1 library, sorted by categories (domains, loops and stripes) in comparison to CFM probabilities in the pancreas. Orange and green show higher and lower probabilities relative to the pancreas, respectively. **f** Examples of CFMs for each category that are either more or less likely to occur in the pancreas compared to other tissues.

Figure 4. Cell-type specific 3D chromatin organization can be described by CFMs.

a Schematic illustrating the seqRNA-FISh strategy used to identify cell types followed by Hi-M. **b** 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 α, β, δ, and γ cells. **c** Left column : Grayscale maximum intensity projection of a DAPI-stained nuclei image, centered on a pancreatic islet. Grayscale maximum intensity projection of Ins1, Gcg and Sst RNA-FISH markers, corresponding to α, β, and δ-cells, respectively. Right column : Corresponding median PWD Hi-M matrix for each cell type along the Pdx1 locus. *n* represents the number of traces. Experiments are from two biological replicates. **d** Differential PWD Hi-M matrices between β-cells and exocrine tissues. Blue and red represent shorter and larger distances in β-cells compared to exocrine tissue, respectively. **e** Volcano plot showing the log2 (exocrine/β-cells) PWD distance (x) and -log10 (P-value). Horizontal dashed red line represents P-value < 0.05, T-test. Vertical dashed gray line is centered at x = 0. Left and right parts of the graph indicate larger and shorter distances in β-cells compared to exocrine cells. **f** Probability distribution showing the decomposition of exocrine, α, β, and δ-cells in 36 CFMs. **g** Relative risks for α, β, δ-cells and for all CFMs for the Pdx1 library, sorted by categories (domains, loops and stripes) in comparison to CFM probabilities in the exocrine tissue. Orange and green show higher and lower probabilities relative to the exocrine tissue, respectively.

Figure 5. Perturbation of loop extrusion and CREs changes cell distribution into folding motifs.

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 β-cells 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** Median PWD distance matrices for exocrine, α, β, and δ-cells from HFD-fed mice. *N* represents the number of biological replicates, and *n* the number of traces. **d** Differential PWD Hi-M Matrices of the Pdx1 locus from ND-fed mice versus the same cell type in HFD-fed mice. Red represents larger distances in HFD-fed mice compared to ND-fed mice. **e** Volcano plot showing the log2 (cell-types in HFD/cell-types in HFD). Color code represents differential pairwise distances (red: larger distances in ND, blue: larger distances in HFD). **f** Left : Relative risks for β-ND, β-HFD and exocrine-HFD for all CFMs for the Pdx1 library, sorted by categories (domains, loops and stripes) in comparison to CFM probabilities in the exocrine-ND. Right : Relative risks for exocrine-ND, exocrine-HFD and β-HFD for all CFMs for the Pdx1 library, sorted by categories (domains, loops and stripes) in comparison to CFM probabilities in the β-ND.

Table 1: Loop extrusion parameters

Parameter	Value
Average lifetime of a LEF	480
Average separation between LEFs	120
Steepness of the logistic function used to compute LEF capture probability by CTCF	0.012
Midpoint of the logistic function	390.2307692307692
LEF release probability by CTCF	0.005

Table 2: CTCF positions and probabilities

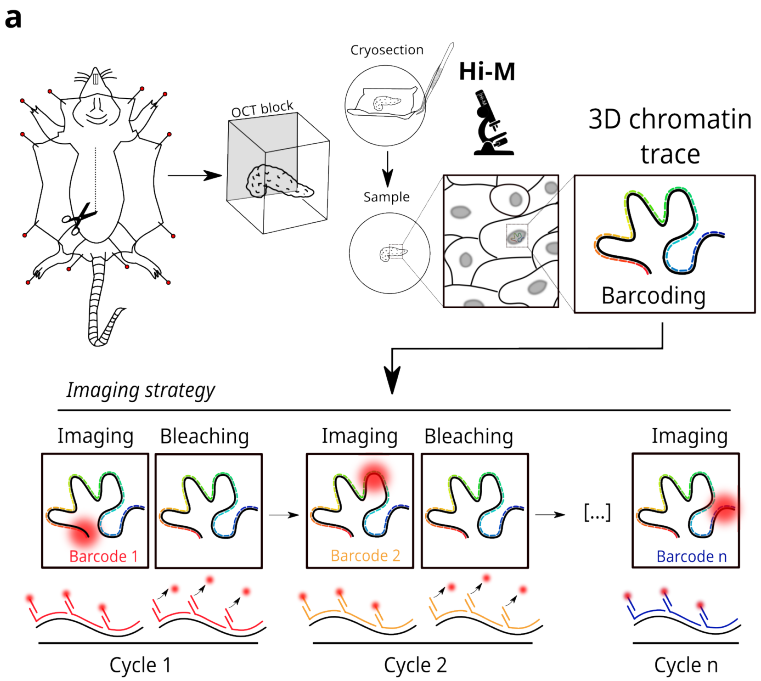
Position	LEF capture probability
33	0.81716109
34	0.25298232
68	0.05804334
111	0.02980649
183	0.03271153
205	0.25754473
356	0.03891178
396	0.0161943
398	0.0219935
421	0.58254526
427	0.3759512
448	0.99999798
452	0.66133089

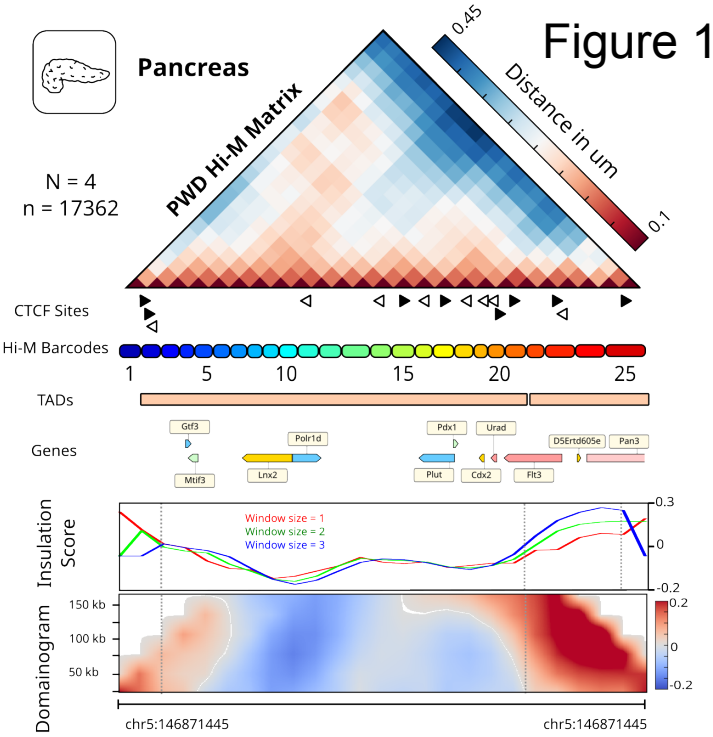
Table 3: Molecular dynamics simulation parameters

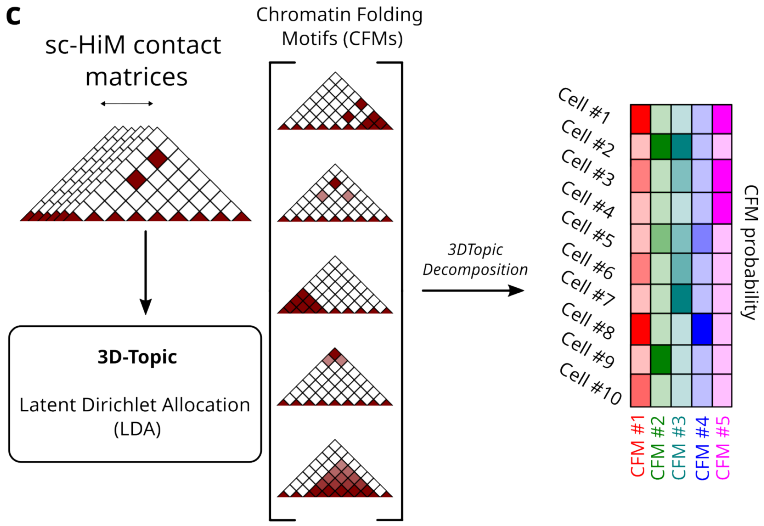
Parameter	Value
Density	0.1
Monomer-monomer harmonic bond length	1.0
Monomer-monomer harmonic bond wiggle distance	0.1

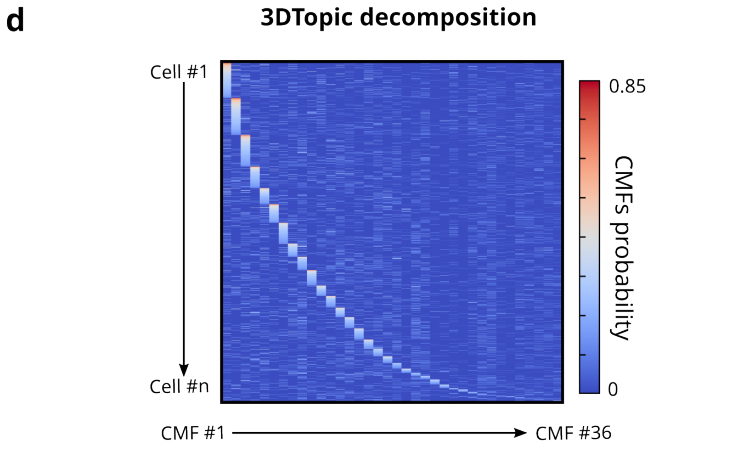
Chain stiffness	1.5
Repulsive potential value at the center of the monomer	1.5
Radius at which the repulsive force becomes null	1.05
LEF-mediated harmonic bond length	0.5
LEF-mediated harmonic bond wiggle distance	0.2

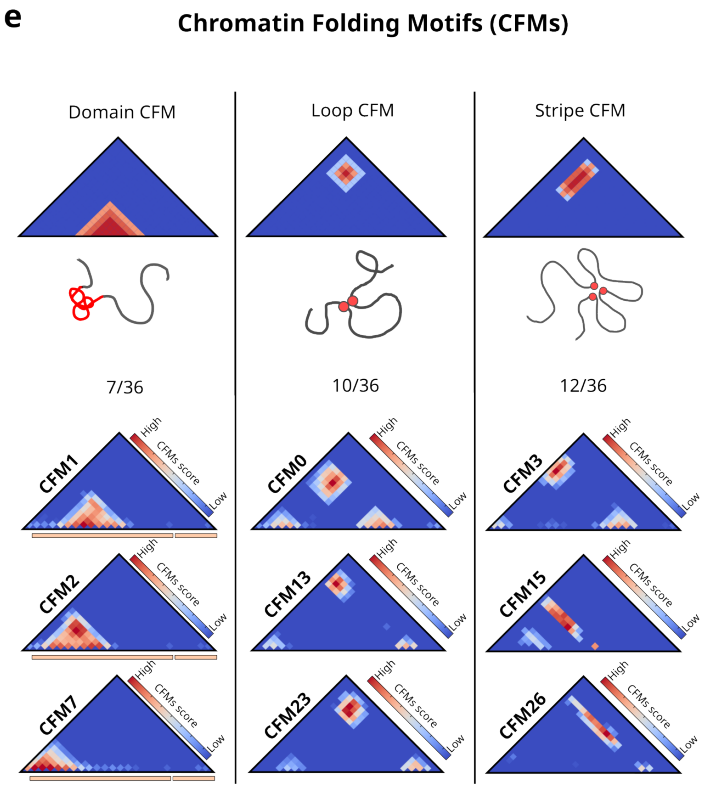
Figure 1

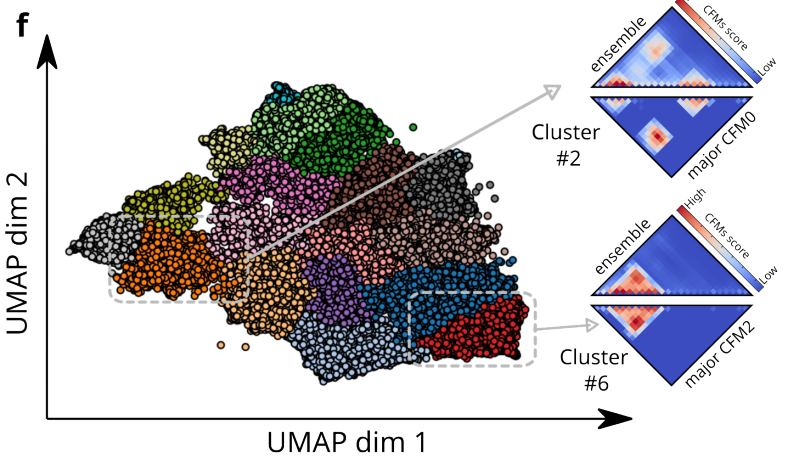

a
Cryosection
Hi-M
OCT block
Sample
3D chromatin trace
Barcoding
Imaging strategy
Imaging
Bleaching
Imaging
Bleaching
Imaging
[...]
Barcode 1
Barcode 2
Barcode n
Cycle 1
Cycle 2
Cycle n


b
Pancreas
PWD Hi-M Matrix
N = 4
n = 17362
0.45
Distance in um
0.1
CTCF Sites
Hi-M Barcodes
1
5
10
15
20
25
TADs
Genes
Gtf3
Mtif3
Polr1d
Lnx2
Pdx1
Plut
Urad
Cdx2
D5Ertd605e
Flt3
Pan3
Insulation Score
Window size = 1
Window size = 2
Window size = 3
0.3
0
-0.2
Domainogram
150 kb
100 kb
50 kb
0.2
0
-0.2
chr5:146871445
chr5:146871445


c
sc-HiM contact matrices
Chromatin Folding Motifs (CFMs)
3D-Topic
Latent Dirichlet Allocation (LDA)
3DTopic Decomposition
Cell #1
Cell #2
Cell #3
Cell #4
Cell #5
Cell #6
Cell #7
Cell #8
Cell #9
Cell #10
CFM #1
CFM #2
CFM #3
CFM #4
CFM #5
CFM probability


d
3DTopic decomposition
Cell #1
Cell #n
0.85
0
CMFs probability
CMF #1
CMF #36


e
Chromatin Folding Motifs (CFMs)
Domain CFM
Loop CFM
Stripe CFM
7/36
10/36
12/36
CFM1
CFM2
CFM7
CFM0
CFM13
CFM23
CFM3
CFM15
CFM26
High
CFMs score
Low


f
UMAP dim 2
UMAP dim 1
Cluster #2
Cluster #6
ensemble
major CFM0
major CFM2
High
CFMs score
Low

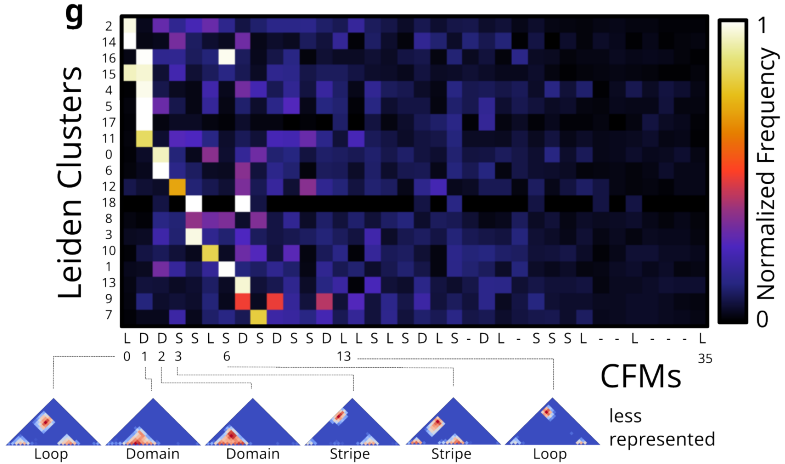

g
Leiden Clusters
Normalized Frequency
1
0
L D D S S L S D S D S S D L L S L S D L S - D L - S S S L - - L - - - L
0 1 2 3 6 13 35
CFMs
more represented
less represented
Loop
Domain
Domain
Stripe
Stripe
Loop

Figure 2

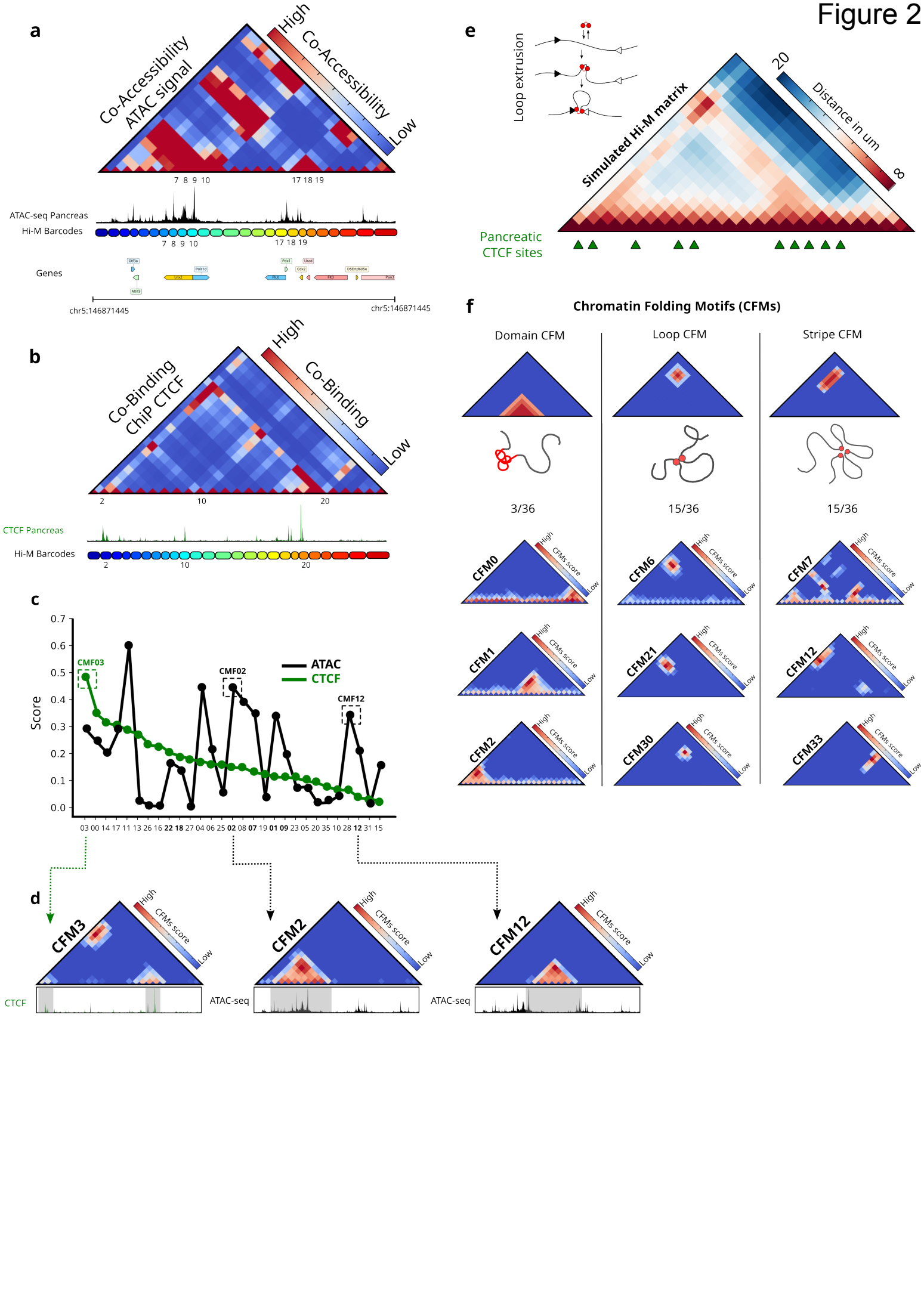

a
Co-Accessibility ATAC signal
High
Co-Accessibility
Low
7 8 9 10
17 18 19
ATAC-seq Pancreas
Hi-M Barcodes
Genes
chr5:146871445
chr5:146871445
b
Co-Binding ChiP CTCF
High
Co-Binding
Low
2
10
20
CTCF Pancreas
Hi-M Barcodes
c
Score
CMF03
CMF02
CMF12
ATAC
CTCF
d
CFM3
CFM2
CFM12
High
CFMs score
Low
CTCF
ATAC-seq
ATAC-seq
e
Loop extrusion
Simulated Hi-M matrix
20
Distance in um
8
Pancreatic CTCF sites
f
Chromatin Folding Motifs (CFMs)
Domain CFM
Loop CFM
Stripe CFM
3/36
15/36
15/36
CFM0
CFM6
CFM7
CFM1
CFM21
CFM12
CFM2
CFM30
CFM33

Figure 3

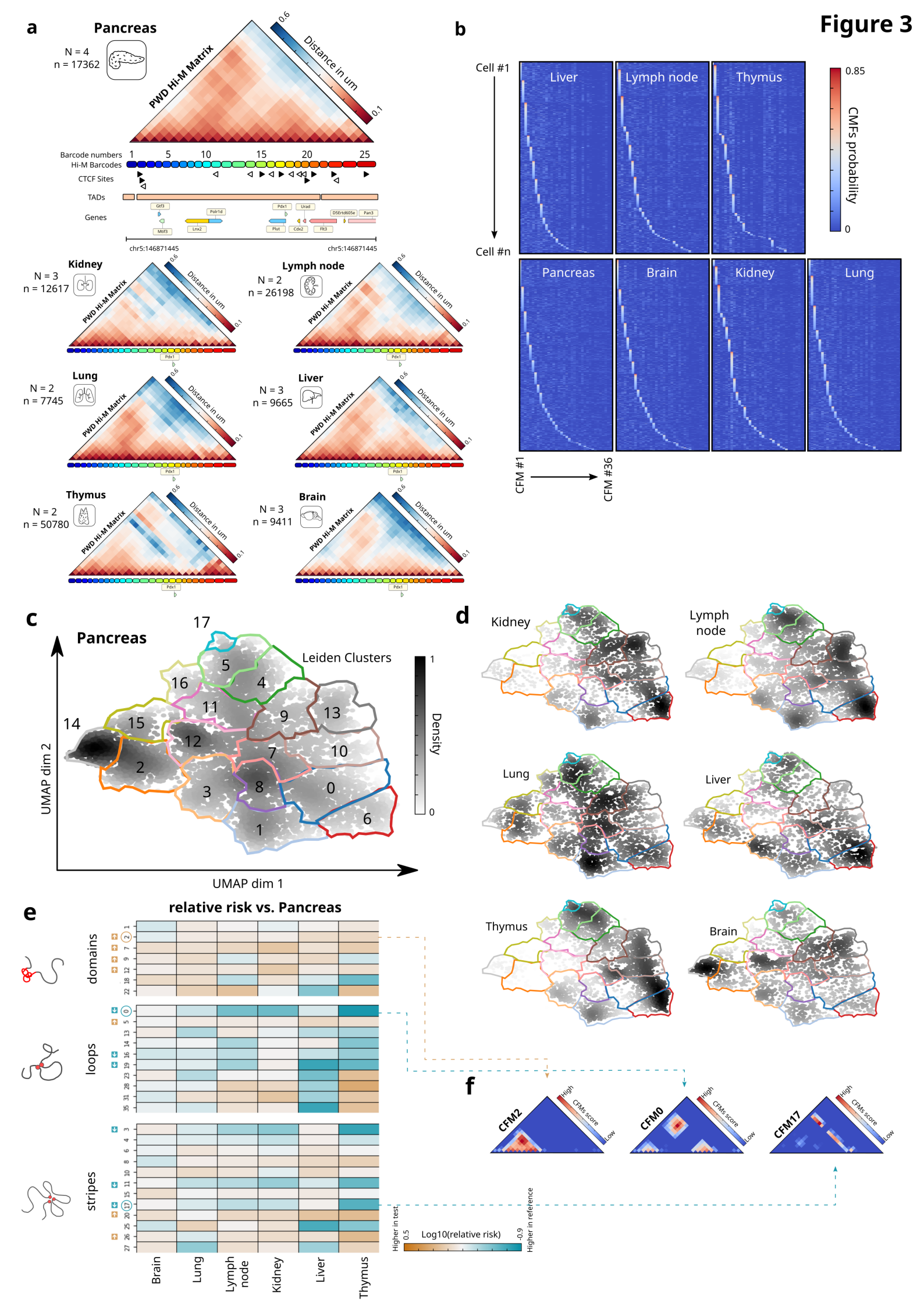

a
Pancreas
N = 4
n = 17362
PWD Hi-M Matrix
Distance in um
Barcode numbers
Hi-M Barcodes
CTCF Sites
TADs
Genes
chr5:146871445
chr5:146871445
Kidney
N = 3
n = 12617
Lymph node
N = 2
n = 26198
Lung
N = 2
n = 7745
Liver
N = 3
n = 9665
Thymus
N = 2
n = 50780
Brain
N = 3
n = 9411
b
Cell #1
Cell #n
Liver
Lymph node
Thymus
Pancreas
Brain
Kidney
Lung
CMFs probability
CFM #1
CFM #36
c
Pancreas
Leiden Clusters
Density
UMAP dim 1
UMAP dim 2
d
Kidney
Lymph node
Lung
Liver
Thymus
Brain
e
relative risk vs. Pancreas
domains
loops
stripes
Brain
Lung
Lymph node
Kidney
Liver
Thymus
Higher in test
Log10(relative risk)
Higher in reference
f
CFM2
CFM0
CFM17
High
CFMs score
Low

Figure 4

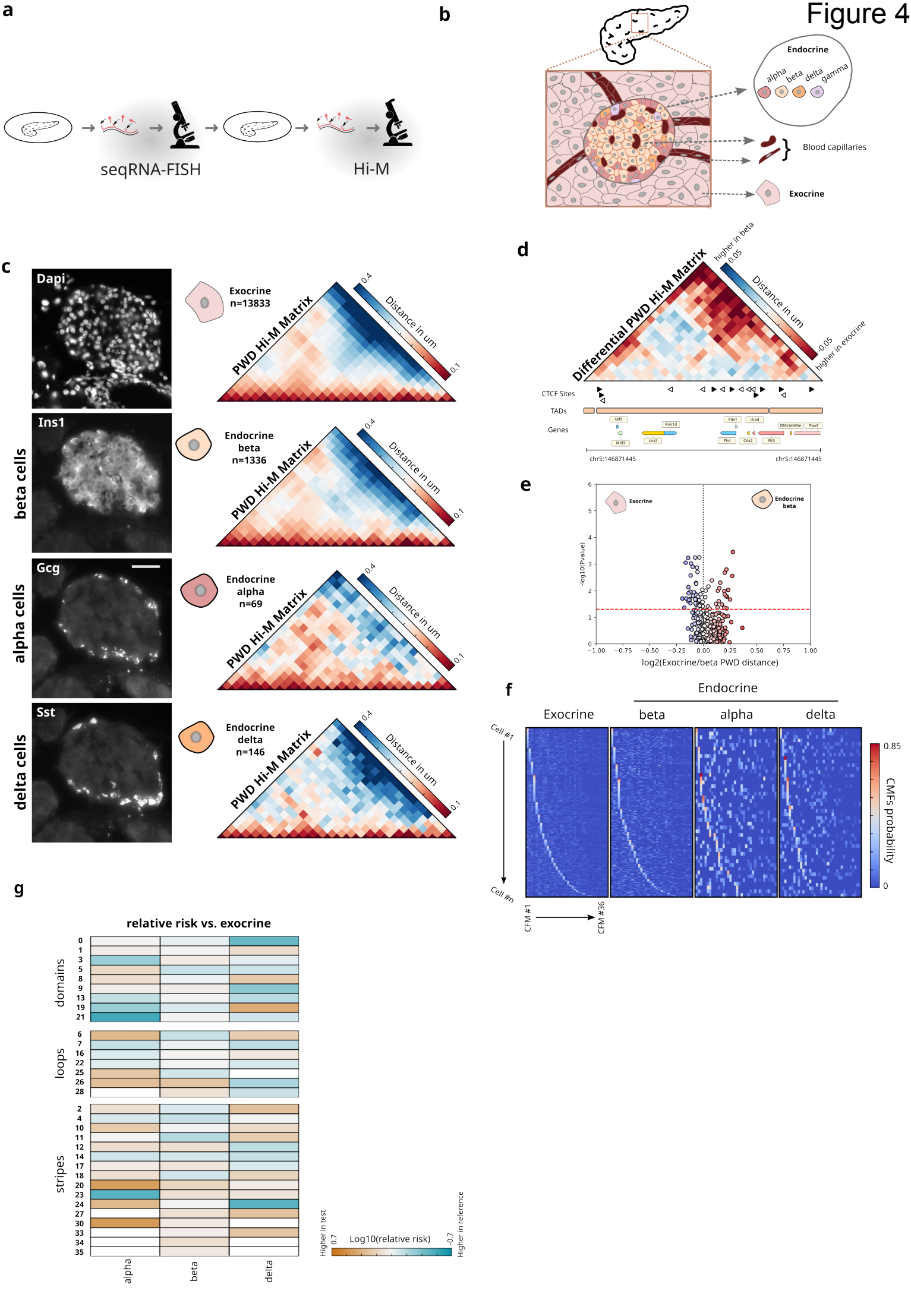
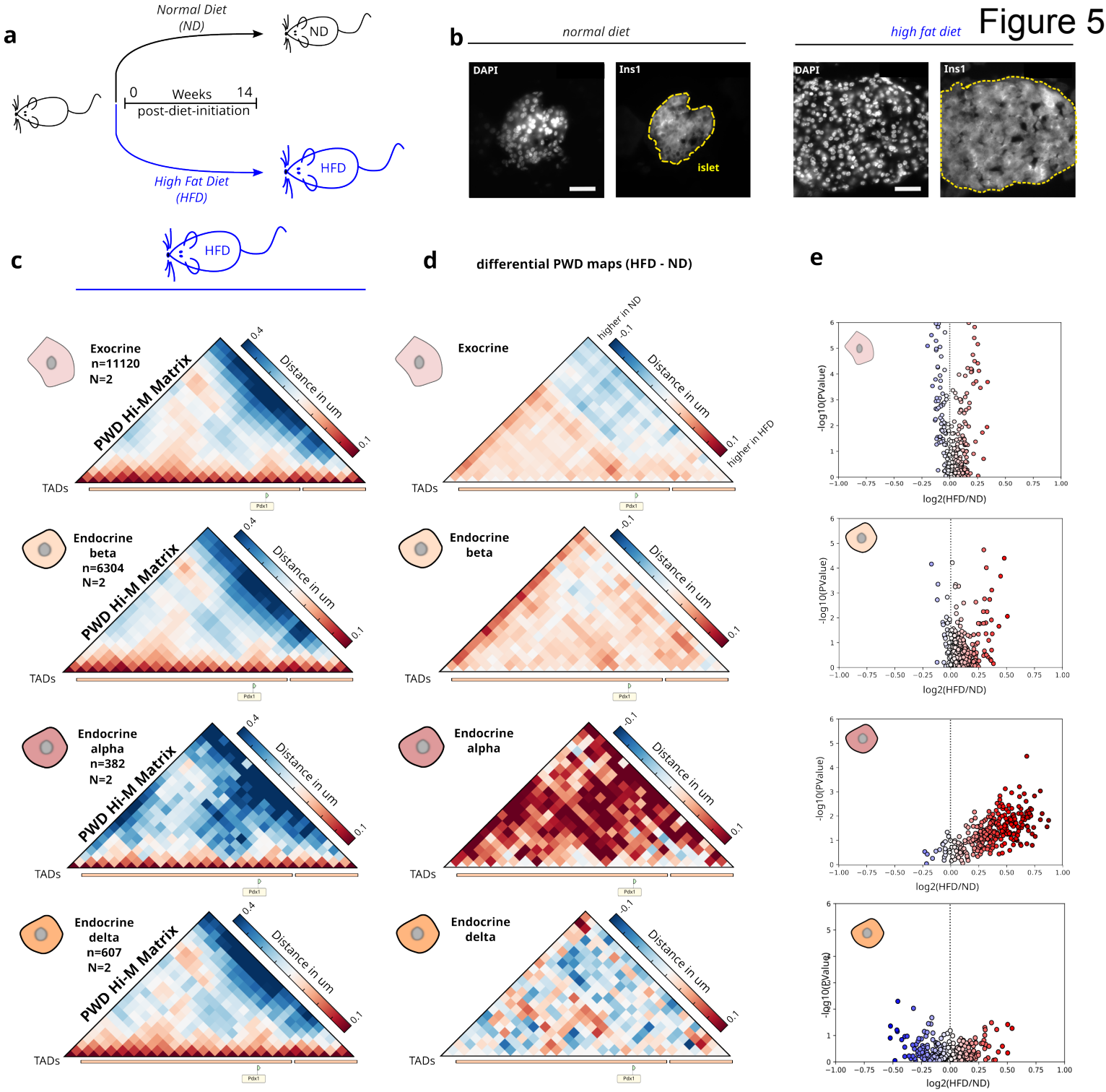
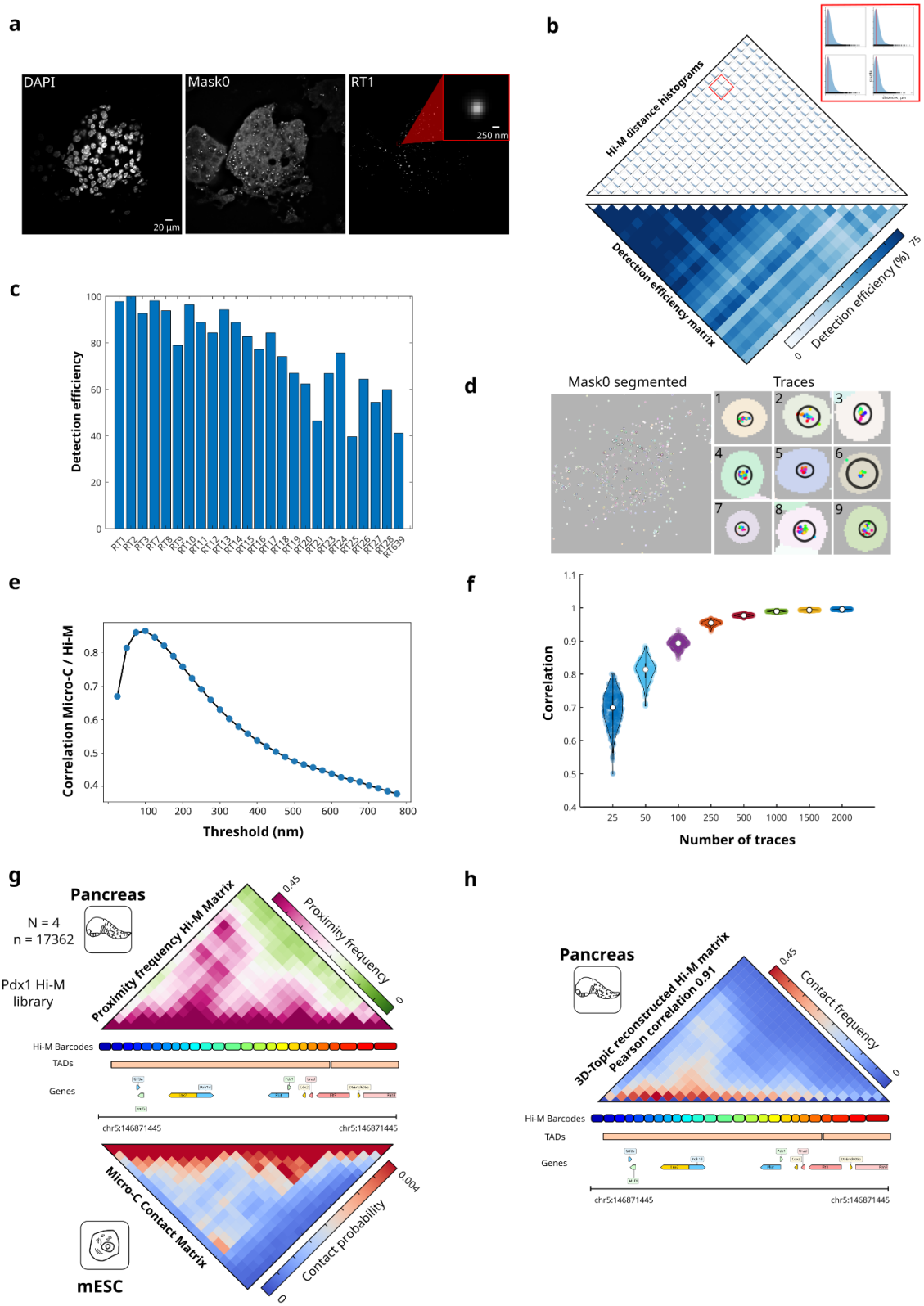

a
seqRNA-FISH
Hi-M
b
Endocrine
alpha
beta
delta
gamma
Blood capillaries
Exocrine
c
Dapi
Ins1
Gcg
Sst
beta cells
alpha cells
delta cells
Exocrine n=13833
Endocrine beta n=1336
Endocrine alpha n=69
Endocrine delta n=146
PWD Hi-M Matrix
Distance in um
0.4
0.1
d
Differential PWD Hi-M Matrix
higher in beta
0.05
-0.05
higher in exocrine
CTCF Sites
TADs
Genes
chr5:146871445
chr5:146871445
e
Exocrine
Endocrine beta
-log10(Pvalue)
log2(Exocrine/beta PWD distance)
f
Endocrine
Exocrine
beta
alpha
delta
Cell #1
Cell #n
CFM #1
CFM #36
0.85
CMFs probability
0
g
relative risk vs. exocrine
domains
loops
stripes
alpha
beta
delta
Higher in test
Log10(relative risk)
Higher in reference
0.7
-0.7

Figure 5


a
Normal Diet (ND)
ND
0
Weeks
14
post-diet-initiation
High Fat Diet (HFD)
HFD
b
normal diet
high fat diet
DAPI
Ins1
islet
c
HFD
d
differential PWD maps (HFD - ND)
e
Exocrine n=11120 N=2
Endocrine beta n=6304 N=2
Endocrine alpha n=382 N=2
Endocrine delta n=607 N=2
PWD Hi-M Matrix
Distance in um
TADs
higher in ND
higher in HFD
-log10(PValue)
log2(HFD/ND)
f
relative risk vs. exocrine ND
relative risk vs. beta ND
domains
loops
stripes
beta ND
beta HFD
exo HFD
exocrine ND
exocrine HFD
beta HFD
Higher in test
Log10(relative risk)
Higher in reference

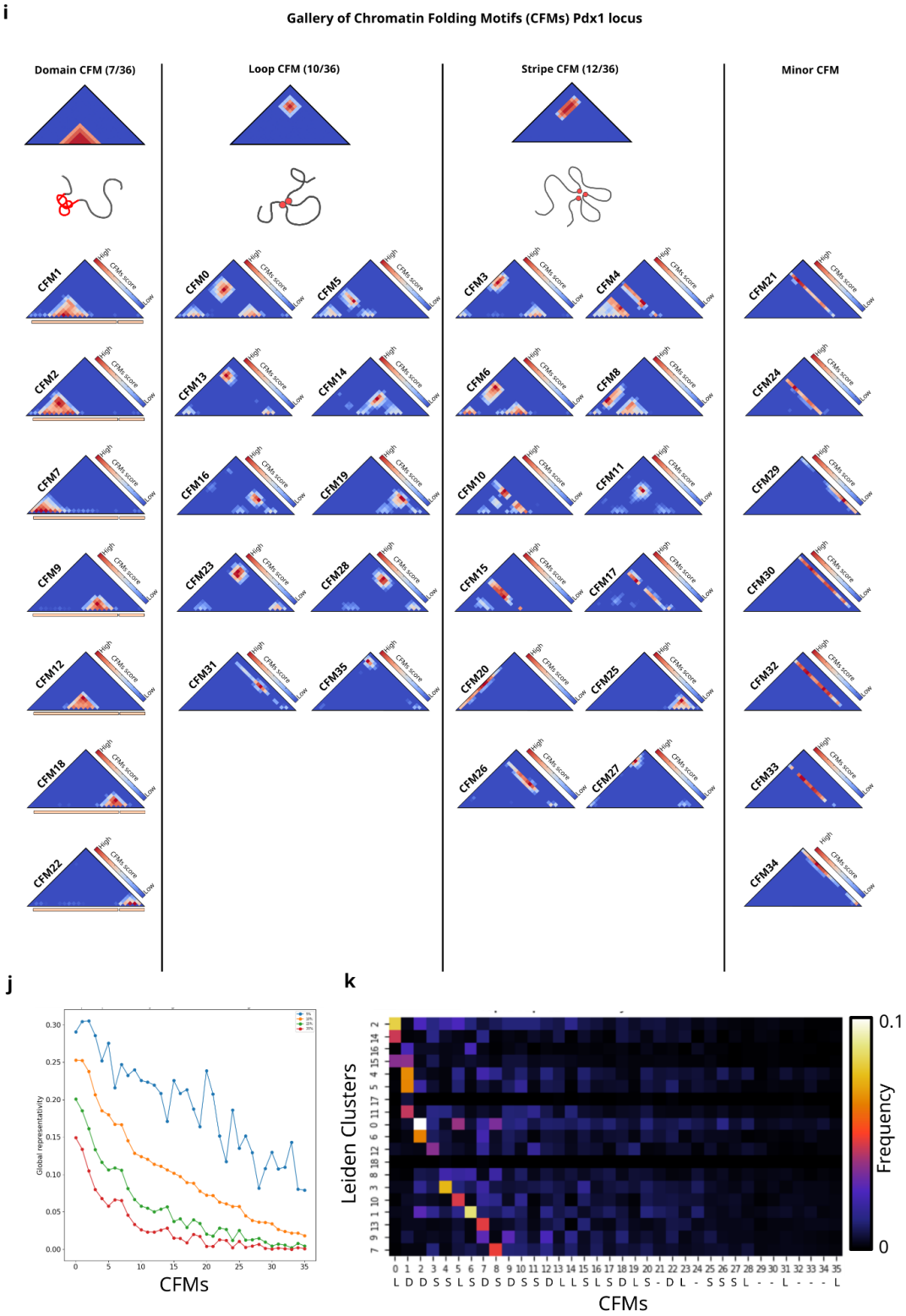
Supplementary Figure S1.



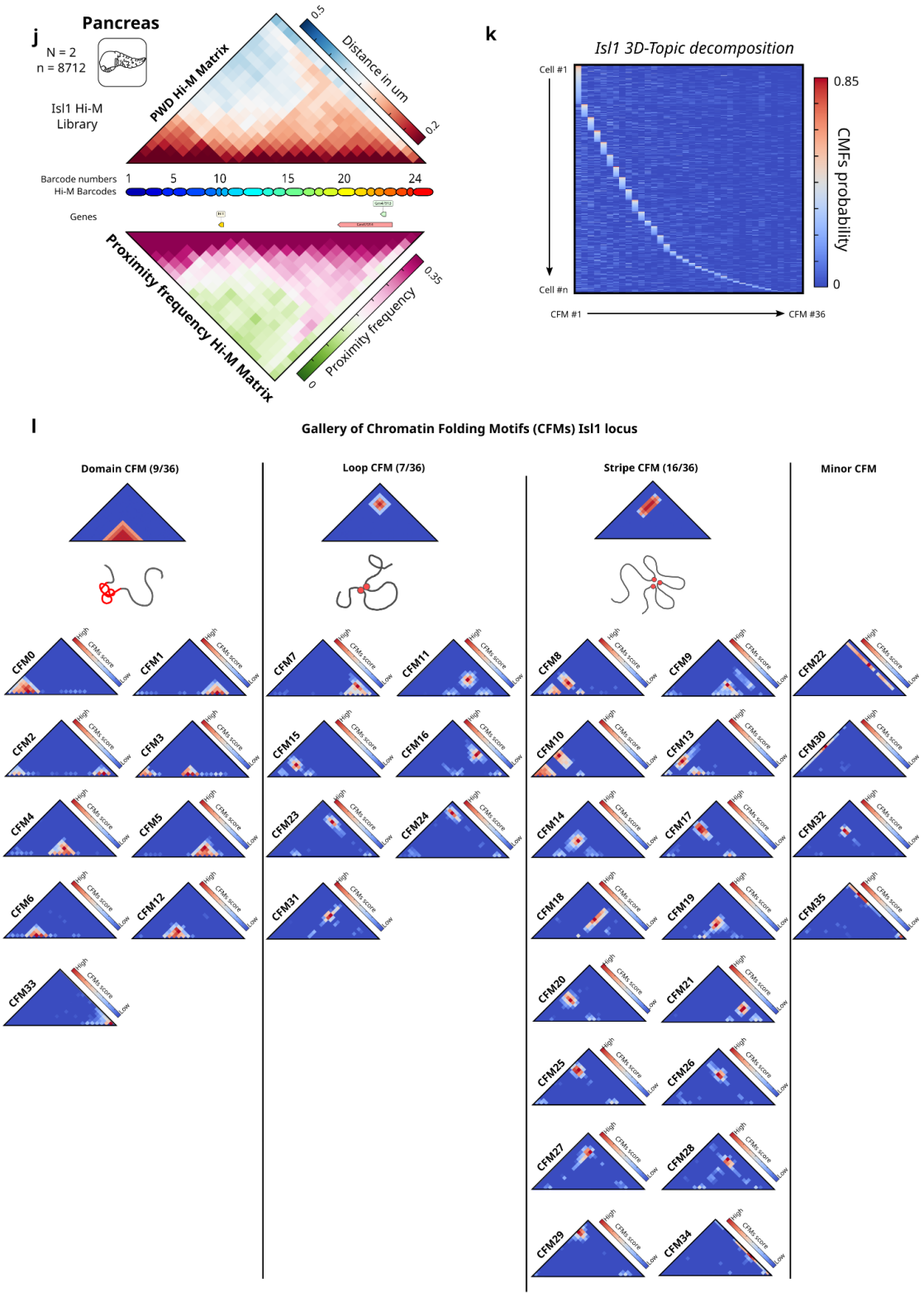
Supplementary Data complementary to Figure 1.

a. Left: Grayscale DAPI-stained nuclei image of a mouse pancreas, centered on a

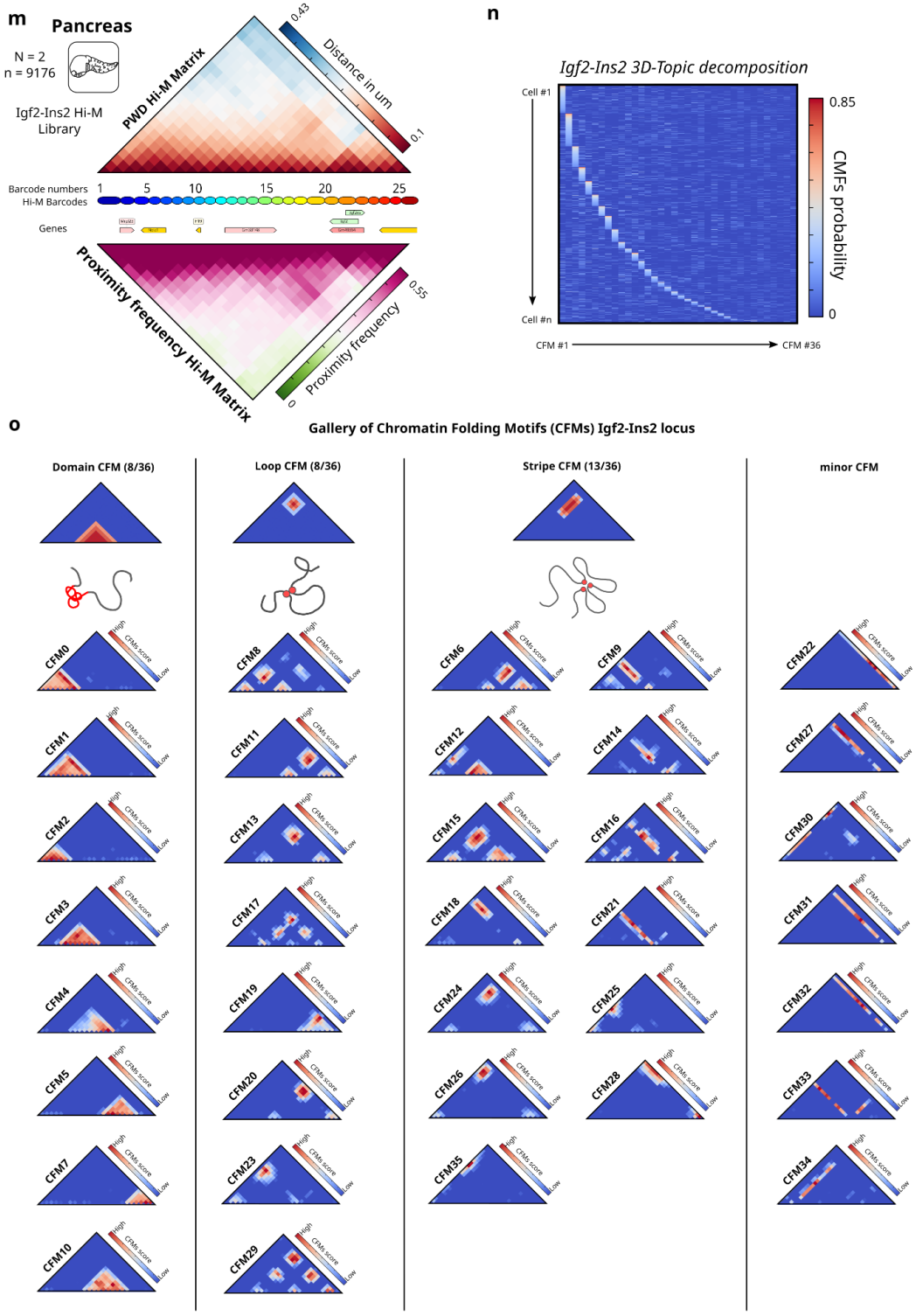
pancreatic islet. Middle : Maximum intensity projection of the fluorescence signal from the entire Pdx1 Hi-M library (Mask0). Right : Maximum intensity projection of the fluorescence signal from a single barcode (RT1) in the same field of view. **b.** Top : Matrix displaying the histogram of pairwise distance distributions within all barcode pairs acquired on mouse pancreatic tissue. Bottom : Detection efficiency matrix normalized by the most detected barcode pair. Dark blue and white represent high and low detection efficiencies, respectively. **c.** Bar-plot representing the detection efficiency for all barcodes normalized by the most detected barcode. **d.** Example of segmented barcodes in traces in the same field. The color code indicates different barcodes found in a trace. **e.** Pearson correlation coefficient between interpolated Micro-C contact map and Hi-M proximity frequency maps generated with different cutoff distances. **f.** Violin plots representing the Pearson correlation between the Hi-M ensemble matrix and matrices generated by sampling subsets of traces by bootstrapping. For each condition, 250 bootstrapping cycles were used. **g.** Top: Proximity Hi-M proximity frequency matrix of the Pdx1 locus (chr5:146871445-146871445) from mouse pancreas (cutoff distance: 200 nm). Pink and green represent high and low proximity frequencies, respectively. The map has been constructed with 17362 traces from 4 experiments. Barcodes used for Hi-M sequential imaging are represented as boxes with different colors aligned with TADs annotation, gene locations. Bottom : Micro-C contact probability matrix along the Pdx1 locus in mESC. Red and blue represent high and low contact probabilities, respectively. **h.** 3DTopic reconstructed Hi-M matrix along the Pdx1 locus.



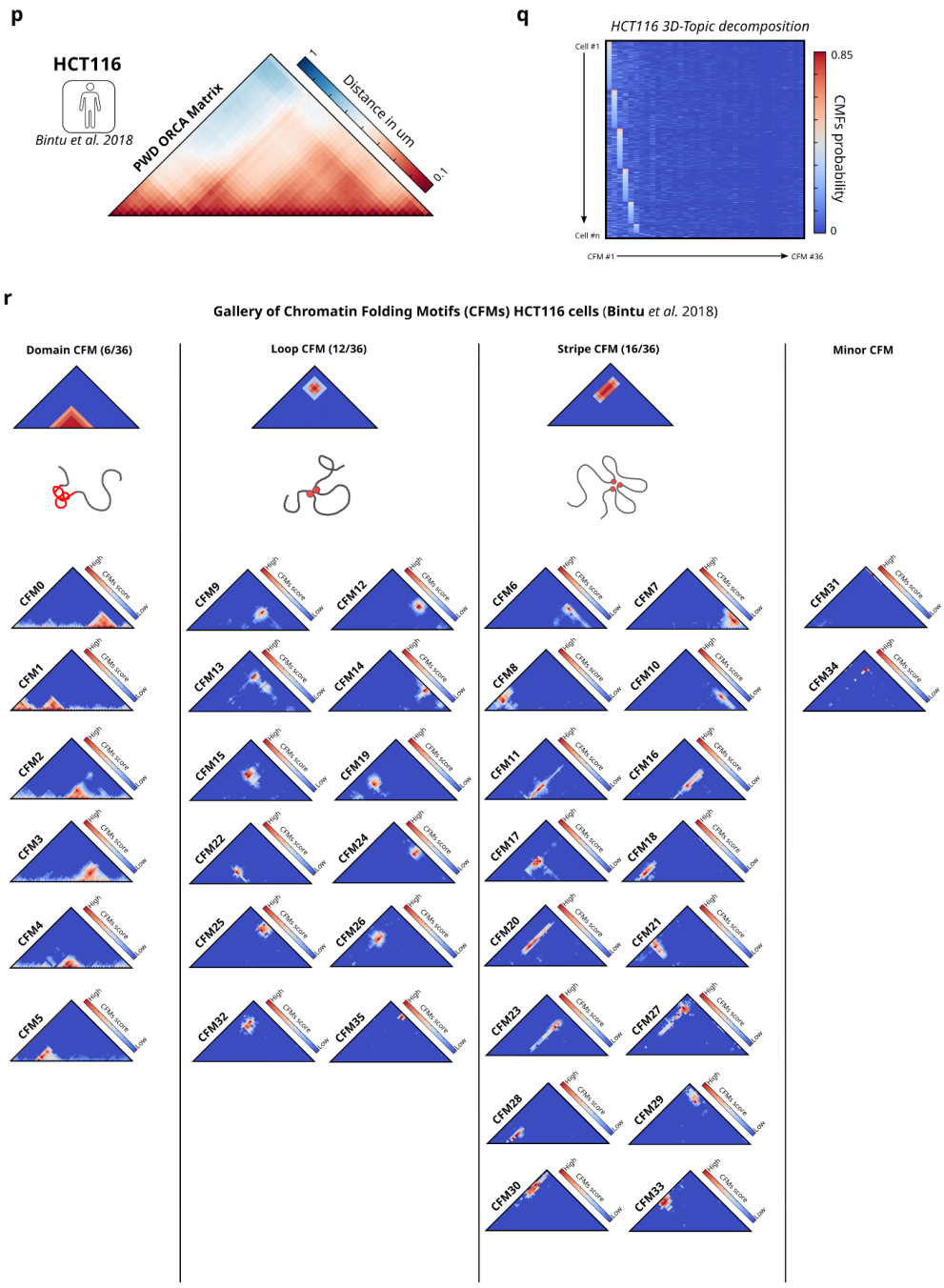
i. Gallery of chromatin folding motifs obtained from one decomposition of single cell Hi-M data along the Pdx1 Locus. CFMs are categorized into three distinct groups according to the structure they describe (Domain, Loop and Stripe). Minor CFMs refer to CFMs that are infrequently used in the decomposition of a single cell. **j.** Graph representing the frequency of CFMs occurring in more than X % in cell decompositions. (X = 5,10,15 and 20 %) **k.** Matrix representing the frequency of CFMs occurring more than 10 % in cell decompositions.



l. Top: Hi-M PWD matrix along the Isl1 locus (chr5:116102031-116759941) in mouse pancreas (mm10) constructed from 8712 traces derived from 2 experiments. Bottom : Proximity Hi-M proximity frequency of the Isl1 locus (cutoff distance: 200 nm). Genes locations are displayed along the locus. **m.** Probability distribution showing the decomposition of pancreatic cells for the Isl1 locus in 36 CFMs. **n.** Gallery of chromatin folding motifs obtained from one decomposition of single cell Hi-M data along the Isl1 Locus.

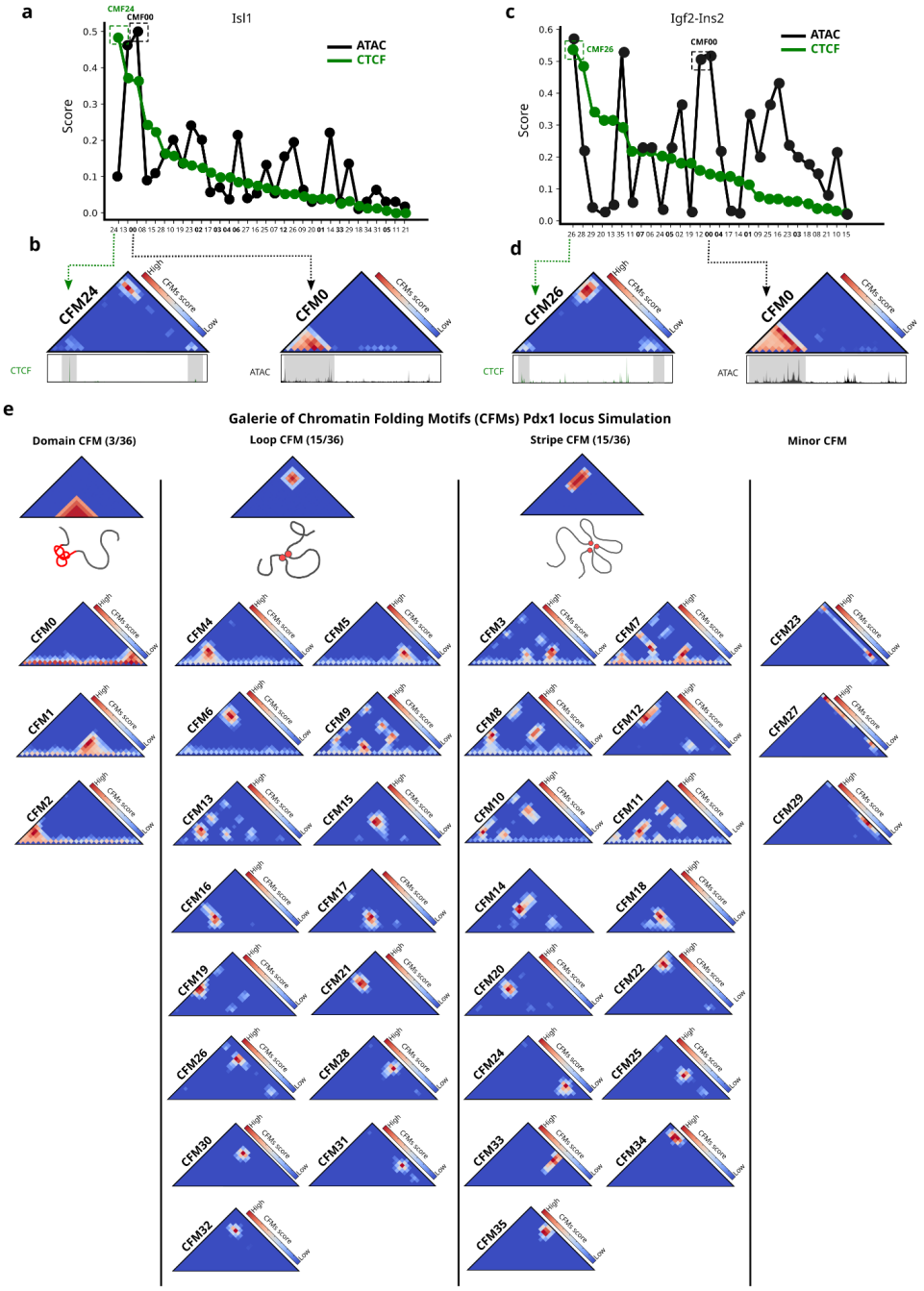


o. Top: Hi-M PWD matrix along the Igf2/Ins2 locus (chr7:142520024-142699856) in mouse pancreas (mm10) constructed from 8712 traces derived from 2 experiments. Bottom : Proximity Hi-M proximity frequency matrix of the Igf2/Ins2 locus (cutoff distance: 200 nm). **p.** Probability distribution showing the decomposition of pancreatic cells for the Igf2/Ins2 locus in 36 CFMs. **q.** Gallery of chromatin folding motifs obtained from one decomposition of single cell Hi-M data along the Igf2/Ins2 Locus.



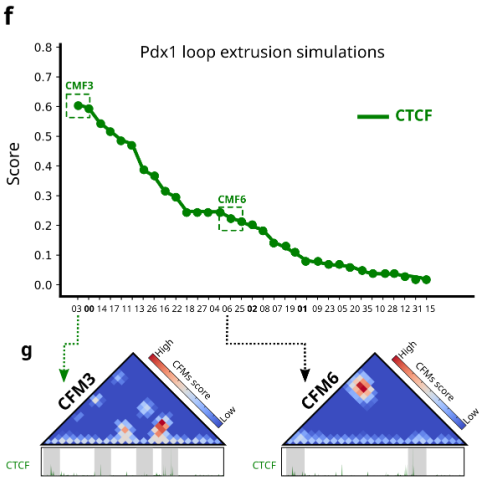
r. Median PWD matrix from Bintu *et al*. 2018 generated on *HCT 116* cell line (Bintu et al., 2018). **s.** Probability distribution showing the decomposition of *HCT 116* cells along the studied locus in 36 CFMs. **t.** Gallery of chromatin folding motifs obtained from one decomposition of single cell chromatin tracing data.

Supplementary Figure S2.



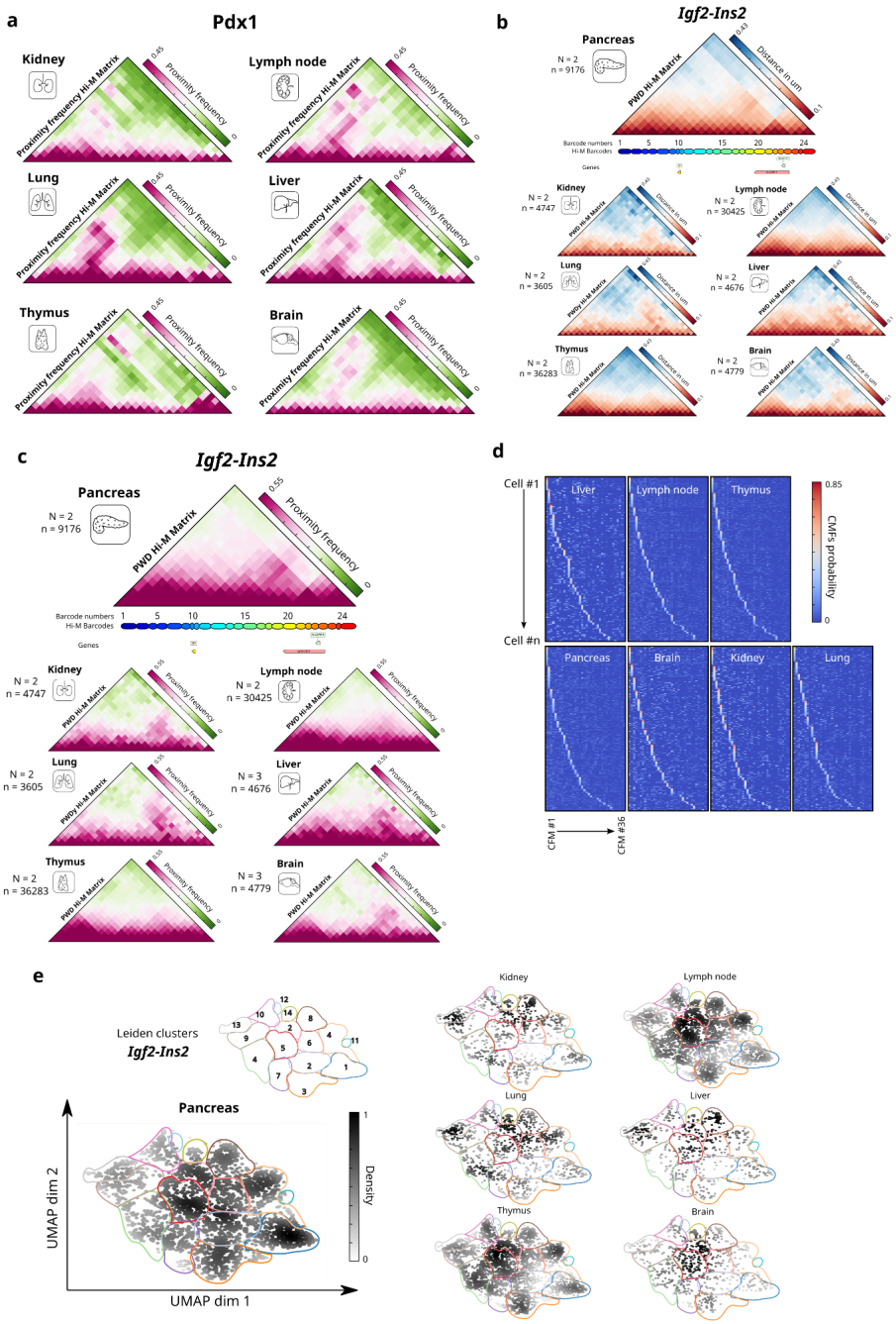
Supplementary Data complementary to Figure 2.

a Association scores between features (ATAC and CTCF) and CFMs for the Isl1 locus. Black and green curves indicate the association score for ATAC and CTCF, respectively. CFMs are sorted by descending order based on decreasing CTCF scores, with Domain CFMs numbers indicated in bold. **b** Examples of CFM matrices associated with high CTCF score (left) and high ATAC score (right) along the Isl1 locus. **c** Association scores between features (ATAC and CTCF) and CFMs for the Isl1 locus. **d** Examples of CFM matrices associated with high CTCF score (left) and high ATAC score (right) along the Igf2-/ns2 locus. **e** Gallery of chromatin folding motifs obtained from one decomposition of simulated Hi-M data using the loop-extrusion model along the Pdx1 Locus (see *Methods*).



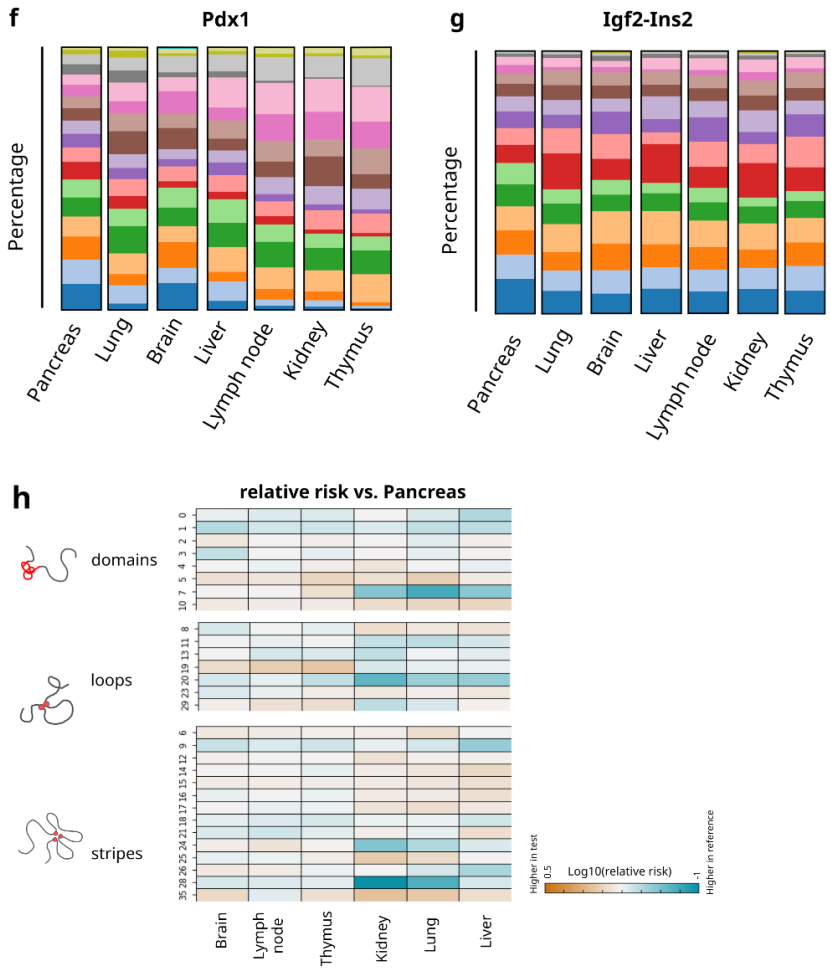
f. Association scores between CTCF and CFMs obtained from loop extrusion simulations along the Pdx1 Locus (see Methods and Tables below). **g.** Examples of CFM matrices.

Supplementary Figure S3.



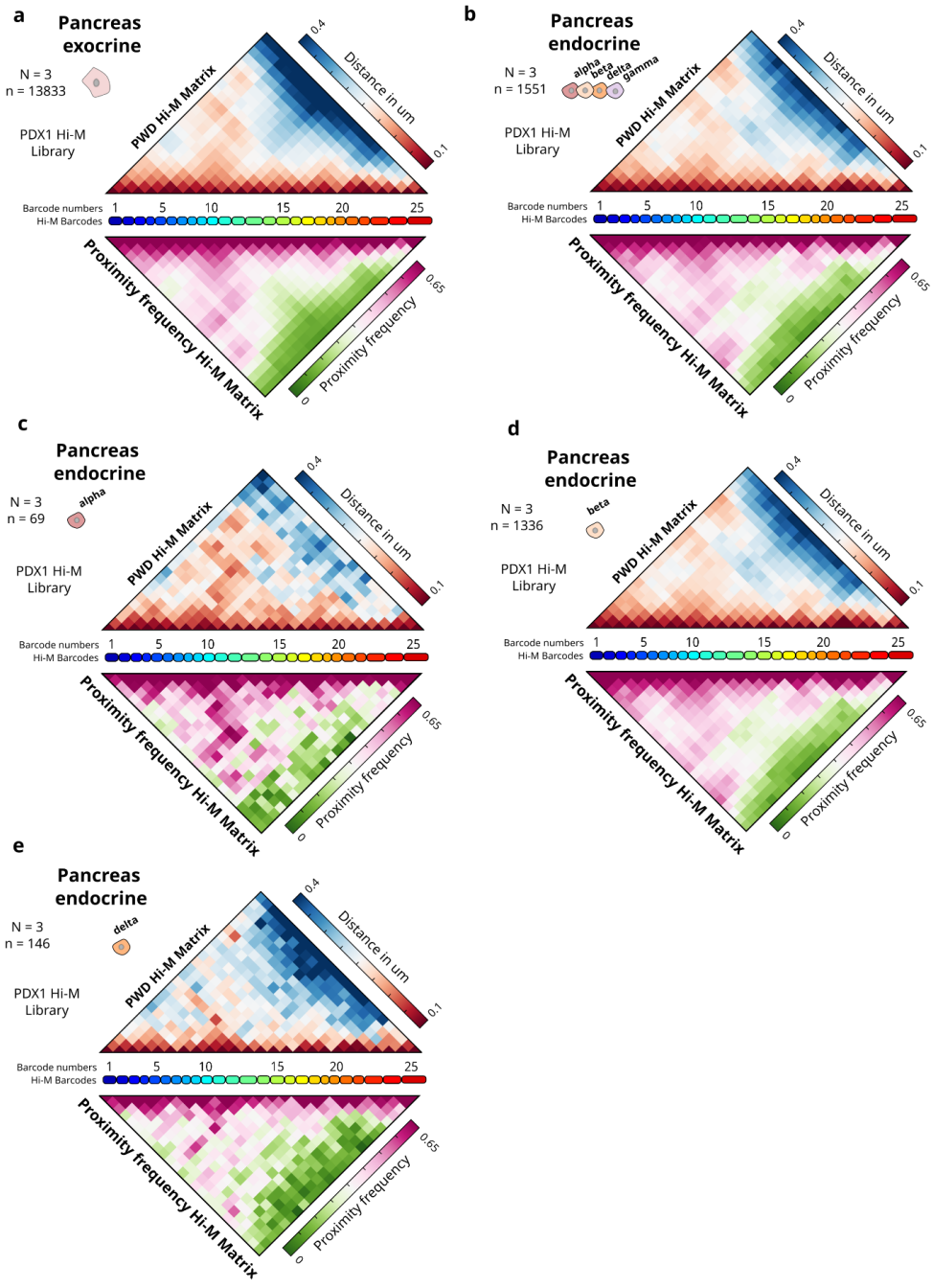
Supplementary Data complementary to Figure 3.

a Proximity Hi-M proximity frequency matrices of the Pdx1 locus for six other mouse tissues including Kidney, Lymph node, Lung, Liver, Thymus and Brain. (cutoff distance: 200 nm). **b** Median PWD Hi-M matrices for the mouse pancreas and six other mouse tissues including Kidney, Lymph node, Lung, Liver, Thymus and Brain along the Igf2/Ins2 locus. **c** Derived proximity frequency matrices (cutoff distance: 200 nm). **d** Probability distribution showing the decomposition of 7 tissues for the Igf2/Ins2 locus in 36 CFMs. **e** Top left: Leiden clusters within the UMAP landscape of the Igf2/Ins2 Hi-M library. Bottom left: Density distribution of pancreatic cells in the UMAP landscape, overlaid with Leiden clusters. Dark black and light black indicate high and low density, respectively. Right: Density distribution of cells from other tissues within the same UMAP landscape.



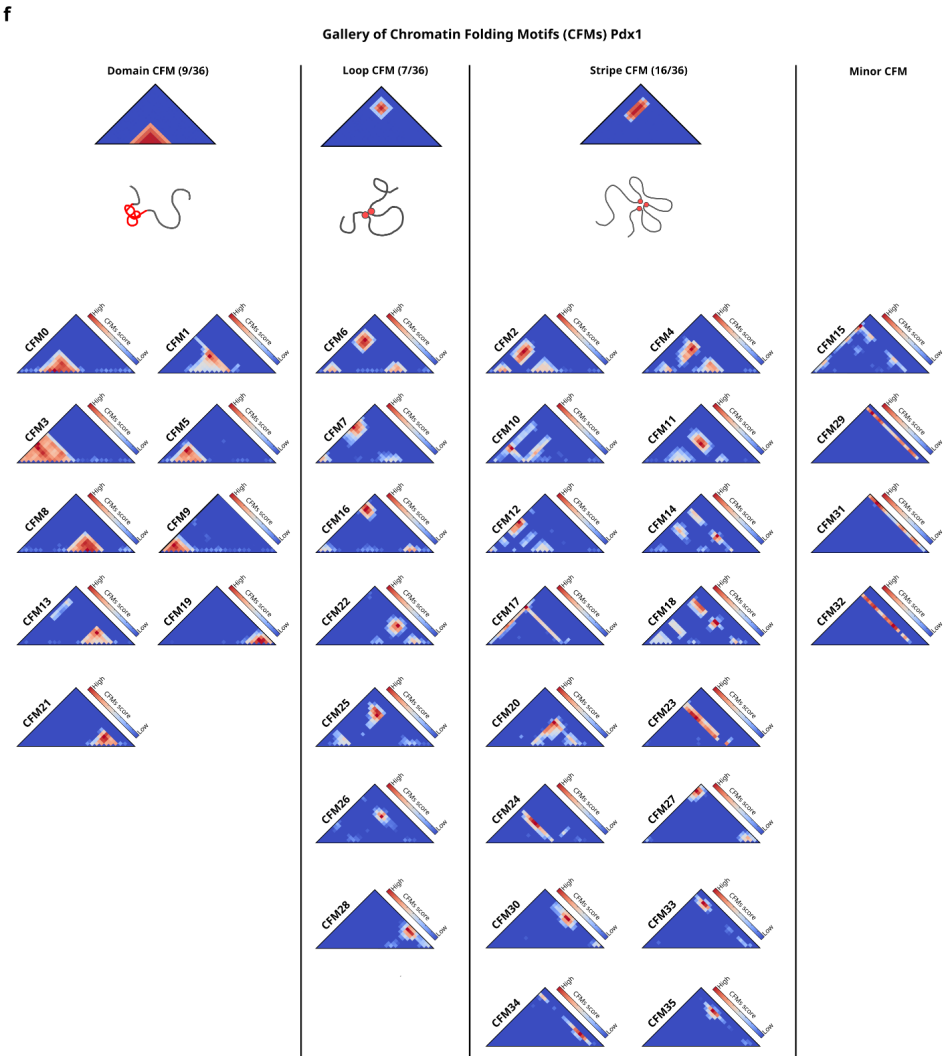
f Stacked bar plots representing the percentage of cells within each Leiden cluster for the Pdx1 Hi-M library for all tissues studied. Stacked bar plots for the Igf2/Ins2 Hi-M library for all tissues studied. **h** Relative risks for all tissues and for all CFMs for the Igf2/Ins2 library, sorted by categories (domains, loops and stripes) in comparison to CFM probabilities in the pancreas. Orange and green show higher and lower probabilities relative to the pancreas, respectively.

Supplementary Figure S4.



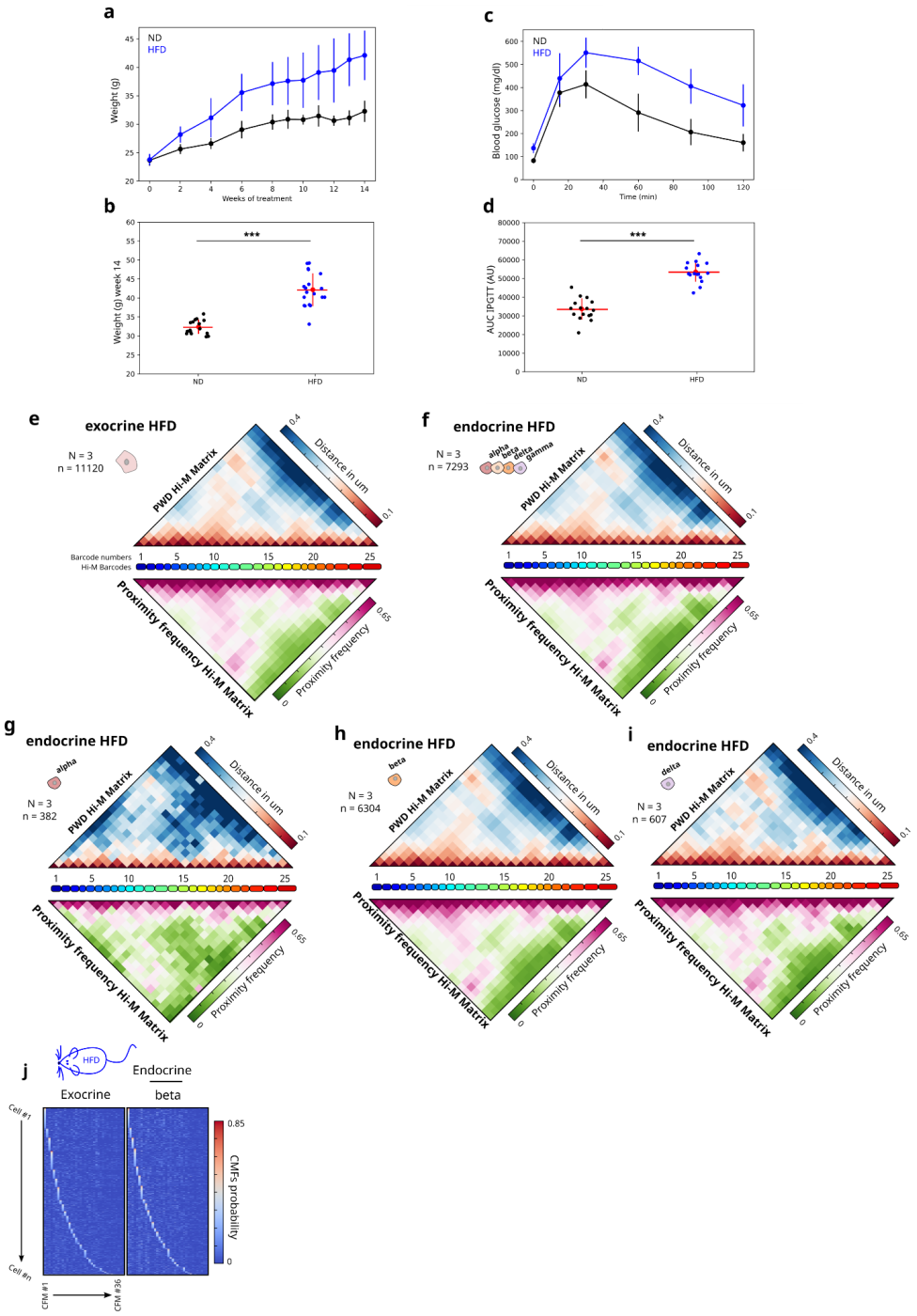
Supplementary Data complementary to Figure 4.

a. Top: Hi-M median PWD matrix for exocrine cells along the Pdx1 locus constructed from 13833 traces derived from 3 experiments. Bottom : Proximity Hi-M proximity frequency matrix (cutoff distance: 200 nm). **b.** Top: Hi-M median PWD matrix for endocrine cells constructed from 1551 traces derived from 3 experiments. Bottom : Proximity Hi-M proximity frequency matrix (cutoff distance: 200 nm). **c.** Top: Hi-M median PWD matrix for alpha cells constructed from 69 traces derived from 3 experiments. Bottom : Proximity Hi-M proximity frequency matrix (cutoff distance: 200 nm). **d.** Top: Hi-M median PWD matrix for beta cells constructed from 1336 traces derived from 3 experiments. Bottom : Proximity Hi-M proximity frequency matrix (cutoff distance: 200 nm). **e.** Top: Hi-M median PWD matrix for delta cells constructed from 146 traces derived from 3 experiments. Bottom : Proximity Hi-M proximity frequency matrix (cutoff distance: 200 nm).



f. Gallery of chromatin folding motifs obtained from one decomposition of simulated Hi-M data using the Pdx1 Locus on Normal Diet mice (see *Methods*).

Supplementary Figure S5.



Supplementary Data complementary to Figure 5.

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.** Top: Hi-M median PWD matrix for exocrine cells along the Pdx1 locus constructed from 11120 traces derived from 3 experiments on HFD-fed mice. Bottom : Proximity Hi-M proximity frequency matrix (cutoff distance: 200 nm). **f.** Top: Hi-M median PWD matrix for endocrine cells constructed from 7293 traces derived from 3 experiments on HFD mice. Bottom : Proximity Hi-M proximity frequency matrix (cutoff distance: 200 nm). **g.** Top: Hi-M

median PWD matrix for alpha cells constructed from 382 traces derived from 3 experiments on HFD mice. Bottom : Proximity Hi-M proximity frequency matrix (cutoff distance: 200 nm). **h.** Top: Hi-M median PWD matrix for beta cells constructed from 6304 traces derived from 3 experiments on HFD mice. Bottom : Proximity Hi-M proximity frequency matrix (cutoff distance: 200 nm). **i.** Top: Hi-M median PWD matrix for delta cells constructed from 607 traces derived from 3 experiments on HFD mice. Bottom : Proximity Hi-M proximity frequency matrix (cutoff distance: 200 nm). **j.** Probability distribution showing the decomposition of exocrine tissue and beta cells for the Pdx1 locus in 36 CFMs.

Chapter 3: Discussion

Over the past decade, studies in chromosome conformation have unveiled a hierarchical landscape of physical interactions along chromosomes. The discovery of the structure of TADs is one of the most exciting discoveries of recent years, as they are thought to play an important role in transcriptional gene regulation. Within TADs, researchers have identified sub-TADs, which are intricately nested structures. From a structural point of view, the distinction between TADs and sub-TADs remains challenging, as the identification of such structures is strongly influenced by the resolution of the dataset and because of ensemble averaging. This raises the question of whether TADs and sub-TADs are truly distinct structures, or whether they simply represent the same features of genome organisation repeated at a smaller scale. To date, the main observation suggesting that TADs and sub-TADs may represent distinct structures of genome organisation is that TADs have been characterised as a invariant across cell types, whereas changes at the sub-TAD level have been observed during the process of cell differentiation [149].

During my PhD, my work focused mainly on studying 3D structural organisation in single cells at TAD and sub-TADs scales. My first aim was to investigate how insulator proteins contribute to the folding of the *Drosophila* genome (see 3.1) and how such structure can affect transcriptional gene regulation (see 4.6). Next I used mouse tissues, to shed light into the role of chromatin architecture in the regulation of tissue-specific transcriptional programmes in single cells (see 3.2). I reasoned that using different model organisms would allow us to compare the conservation and specificity of mechanisms of genome folding and function across species. This discussion is divided into two sections, each focusing on a specific project: one on *Drosophila* and the other on mouse.

Work on *Drosophila*

Despite the similarities in domain organisation between flies and mammals, studies in flies have revealed distinct features in the signature of their TADs. In general, *Drosophila* TADs lack the characteristic interaction hotspots indicative of highly localised CTCF anchor elements observed in mammalian TADs. Although a small number of *Drosophila* domains with apical hotspots have been identified, their formation does not depend on dCTCF binding. In addition to dCTCF, *Drosophila* possesses a distinct set of insulator-binding proteins (IBPs) that are not evolutionarily conserved [17] and that are highly enriched at TAD borders [310, 311]. This intrinsic observation suggests that insulator proteins may contribute to TAD folding in *Drosophila*.

Our bioinformatic analysis using aggregate peaks analysis (APA) [148] and Chromatin Assortativity (ChAs) [33] shows that the majority of preferential chromatin interactions observed in *Drosophila* embryos at nc14 are associated with a specific Class of insulators, which we termed Class I IBPs. This Class includes BEAF-32,

CHRO, DREF, L(3)MBT, Pita, Z4, ZIPIC and Zw5. More interestingly, a significant proportion of these interactions do not appear to involve TAD boundaries. This suggests that these interactions may correspond to regulatory interaction between CREs as recently observed between tethering elements in *Drosophila* embryos [312]. However, Class I IBPs and tethering elements, which mediate preferential interactions between CREs, represent distinct classes of interactions because they do not involve the same set of factors bound to the loop anchors. Additionally, the depletion of Zelda, which is a hallmark of tethering elements, as little effect on Class I IBPs interactions.

Previous studies have demonstrated that TADs emerge during the initiation of transcriptional activation, particularly during *Drosophila* embryogenesis [6, 9, 8] (in nc14). In contrast, our results indicate that preferential interactions between regions of Class I IBPs occur prior to transcriptional activation (e.g. nc12). This is consistent with our previous study showing that interactions between CREs at the *doc* locus are formed as early as nc12 [12]. Depletion of BEAF-32, Chro, and Dref leads to to changes in gene expression [313]. Consequently, we hypothesized that the majority of Class I IBP interactions may represent "regulatory interactions" involving enhancers and promoters rather than "structural" ones involving TADs borders.

To better understand if these regulatory interactions between Class I IBPs are stable or stochastic, we turned to Hi-M to quantify the absolute proximity frequencies of ten Class I IBPs regions located within a specific locus of interest. We observed that Class I IBPs regions physically co-localize in space infrequently (12%), and marginally more frequently than neighboring genomic regions (10.8%). The infrequent preferential interactions observed between Class I IBP regions may arise from a high degree of structural variability from cell to cell, as observed in single-nucleus Hi-C studies [211]. This observation is consistent with the low proximity frequencies observed between "structural" interactions among TAD borders in S2 cells (approximately 10%) [147].

Previous studies have shown the formation of insulator bodies in cultured cells that exhibit properties of liquid-liquid phase separation (LLPS) [90]. This LLPS model suggests that genomically close insulators should frequently interact in close proximity and may involve multiple interacting partners to form rosette-like structures. Our observations are inconsistent with such a model, as pairwise proximity frequencies decrease rapidly with increasing number of interacting partners ($<$5% for 3-way interactions and $<$1% for 4-way interactions).

All in all, our observation support a model in which Class I insulators help to regulate the dynamics of specific interactions between distant *cis*-regulatory regions by stabilising the 3D structure. This model may be consistent with the concept of loop extrusion, but may involve different molecular machinery. For example, certain Class I IBPs identified in *Drosophila* or other proteins could potentially act as loop extrusion boundaries, serving as alternatives to CTCF. However, it's important to

note that concrete evidence for the presence of extruding mechanism in *Drosophila* is currently lacking.

In continuing this work, we anticipate that further research will require the use of a combination of experimental and computational approaches to better understand how preferential interactions between Class I IBP regions are formed.

Firstly, it will be crucial to elucidate the presence of a potential extrusion factor in *Drosophila*. For this, we anticipate that the use of state-of-the-art single-molecule assays for the visualization of putative extrusion factor, as detailed in [181], will play a pivotal role in understanding the mechanisms of genome organization in *Drosophila*. If such a factor exists, it would be interesting to investigate the mechanisms by which Class I IBPs effectively block extrusion *in vitro*, similar to recent *in vitro* work on mammalian CTCF and cohesin [185]. Secondly, while polymer-based simulations have proved to be valuable tools for modelling 3D genome organisation in mammals [15] and elucidating epigenomic domains formation in *Drosophila* [145], further development or adaptation of existing polymer simulation methods will be essential to accurately model genome folding at the scale of TADs and sub-TADs in *Drosophila*.

While the mechanisms underlying TAD folding appear to diverge between *Drosophila* and mammals, the conservation of TADs and sub-TADs structures in various species (reviewed in [5]) has raised significant questions regarding their role in regulating transcription. To date, several studies have focused on understanding the role of TADs in transcriptional gene regulation using sequencing-based techniques and/or imaging. However, these investigations have produced conflicting findings (discussed in 17). In the field, A persistent challenge is the inherent nature of the techniques used, which rely on correlation between chromosome conformation capture data (e.g. Hi-C) and gene expression data (e.g. RNA-seq), often obtained in separate experiments. This indirect correlation between conformation and transcription makes it difficult to establish a direct link between chromatin structure and function. Recent advances in microscopy techniques have revolutionised our ability to simultaneously image both chromosome architecture and transcription at the single cell level (see *Cardozo et al. Mol Cell 2019 contribution* in 4.6 and discussed in 2). These advances allow researchers to directly study the impact of genome architecture on the regulation of gene expression. A particularly noteworthy approach involves sequential DNA-FISH experiments, such as Hi-M, coupled with intronic labelling of specific RNA species (discussed in 2). Such an experiment allows the high-resolution study of chromatin conformation alongside the actual transcriptional status within each individual cell (see *Espinola, Götz et al. Nat Gen 2021 contribution* in 4.6). As an example of the power of this approach, experiments in *Drosophila* embryos have visualized significant enhancer-promoter interactions in both transcriptionally active and inactive cells [12]. These results are consistent with observations using sequencing methods [13]. Such observations suggest a model in which the chromatin architecture, in particular the interactions between enhancers and promoters, acts as a scaffold. Transcriptional output is then determined by the differential binding

of transcriptional activators or repressors to both enhancers and the promoter region.

To test the applicability of this model of enhancer-promoter action in other organisms and to gain deeper insights into the fundamental principles of genome folding and its role in cell-type specific transcriptional regulation, we shifted our focus to mice in the second aim of my thesis.

Work on Mouse

Previous studies have demonstrated the crucial role of 3D genome architecture in the regulation of gene transcription. For example, research has revealed remarkable changes in chromatin structure during the process of cell differentiation [14, 149]. These changes include loops formed between convergent CTCF sites as well as interactions between regulatory elements (such as enhancers and promoters) [10]. Remarkably, depletion of CTCF and cohesin has little effect on E-P/P-P loops, as approximately 80% of E-P contacts and 90% of P-P contacts remain unchanged [194]. These observations suggest that "structural loops" between CTCF sites and "regulatory loops" involving enhancers and promoters are distinct entities. This distinction is further supported by the differing dynamics in their formation during embryogenesis [6, 12] (also described here 3.1).

The distinction between the existence of these two types of loops raises several questions about their nature: Are these loop types present in individual cells? Are they stable or do they exhibit stochastic behaviour as observed in *Drosophila*? Are they mutually exclusive or do they depend on each other to form?.

To address theses questions we imaged the 3D genome architecture of the Pdx1 locus in cryosectioned mouse pancreatic tissue at a resolution of approximately 25 kb. This level of resolution allows the detection of typical features of mammalian genome organisation, including TADs, sub-TADs and apical CTCF interactions. To investigate the organization of this locus and to examine the behavior of these structures in individual cells we deployed a machine learning tool that we called 3DTopic. 3DTopic enable the decomposition of a specific locus in a set of discrete Chromatin Folding Motifs (called CFMs) that serve as fundamental building blocks of chromatin architecture. The CFMs are in turn used to decompose each individual cell by assigning a weight to all CFMs that better fit the experimental observation. Interestingly, our observations revealed the presence of three distinct types of Chromatin CFMs: "loop", "domain", and "stripe". These findings suggest that the organization of a locus can be characterized by these three discrete architectural elements.

Interestingly, the "Loop" and "Stripes" categories encompasses previously described boundary elements interactions, specifically CTCF-CTCF interactions which can effectively be modelled by a loop-extrusion [173] and tend to be classified as "structural loops". In contrast, the "Domain" category is more closely associated

with short-range interactions between chromatin-accessible regions and is more likely to encompass "regulatory loops".

As the single-cell structure of the Pdx1 locus is decomposed among all CFMs categories, it suggests that both "structural loops" and "regulatory loops" may co-operate within single cells. However, a more comprehensive analysis is needed to determine whether certain CFMs are more likely to co-occur with others. For example, it is conceivable that CFMs involving regulatory interactions between CREs may be more prevalent in cells where CFMs involving interactions between TAD boundaries are also present. This would be consistent with the proposed role of TADs in facilitating intra-TAD interactions.

In line with previous observations, we identified a restructuring of the intra-TAD structure at the Pdx1 locus in various mouse tissues. More interestingly, our results shows that the same set of CFMs is employed to accurately model tissue-specific structures, albeit with varying proportions.

Consistently with previous observations, we observed restructuration of the intra-TAD structure at the Pdx1 locus in different mouse tissues and three different pancreatic cell-types. More interestingly, our results show that the same set of CFMs is used to accurately describe the tissue-specific 3D structure, albeit with different proportions. These changes observed in the proportion of CFMs used in different tissues may imply different folding events, possibly due to distinct repertoires of chromatin-binding proteins bound to CREs.

Previous research has highlighted the dysregulation of the epigenetic landscape in pancreatic chromatin associated with the onset of type 2 diabetes (T2D) [36]. The studies have shown that the development of T2D coincides with a reduction in islet-specific gene expression, a decrease in chromatin accessibility as well as a decrease in CTCF expression. This observation suggests that the 3D genome organization and specifically the usage of CFMs may vary between healthy and T2D pancreatic cells. Our Hi-M data are consistent with such model as T2D mice show extensive alterations in chromatin structure within pancreatic cells impacting loop, stripe, and domain CFMs.

All in all, our observations support a model in which the structural scaffold is regulated by the chromatin signature (e.g. accessibility and repertoires of chromatin-binding proteins) and in turn influence regulation of gene expression.

While sequential imaging techniques such as Hi-M have contributed to our understanding by simultaneously visualising chromatin conformation and transcriptional states in a single cell, it is important to consider that their inherent limitations is that they can only study a limited number of loci at a time an cannot access chromatin dynamics. Therefore, further advances will require the use of novel methods capable of analysing enhancer-promoter interaction networks and the transcriptional status of genes on a genome-wide scale. One promising approach will be the cleaver

design of oligopaint libraries using combinatorial coding schemes, similar to those successfully used in MERFISH [34]. Preliminary efforts have already introduced combinatorial labelling for sequential imaging of DNA structure [281]. However, the current resolution of such approaches does not allow to effectively resolve inter-TAD structure thus changes in enhancer-promoter contact may not be detected.

In continuing this work, we anticipate that the use of larger scale imaging techniques such as those discussed above will greatly enhance our understanding of genome conformation. While the 3DTopic approach has proven effective in understanding the fundamental building blocks of 3D genome organisation at the TAD scale, extending this methodology to broader genomic contexts will require careful consideration. A key consideration will be the number of CFMs required to effectively model chromosome-scale structure. This challenge arises from the possible interconnected nature of the genome, where regulatory elements from distal regions can co-occur and influence each other's behaviour. Consequently, if 36 CFMs were sufficient to effectively model 625 kb regions, how many will be needed to model hundreds of megabases? Will this number increase proportionally to the size of the regions being studied, or will it be exponential?. Understanding whether different CFMs from distant genomic regions can coexist harmoniously or whether they exhibit mutual exclusivity is therefore crucial to understanding chromatin behaviour. Such knowledge is essential for effectively adapting existing polymer models to accurately represent the intricacies of chromatin architecture.

Chapter 4: Materials and Methods

4.1 Library design

4.1.1 Sequential RNA-FISH Library design

Sequential RNA-FISH libraries were constructed following the procedure described in [314] with slightly modified version of the library design script available here `https://github.com/ZhuangLab/MERFISH_analysis`. Briefly, a maximum of 90 probes were designed to hybridize to cDNA with a target sequence region of 30bp. Notably, we allowed for a deliberate overlap of 20 bp between adjacent probes, to maximize the number of probes for a gene. Each probe is customized to add a tail consisting of a gene-specific readout sequence and a forward and reverse primer sequence positioned at the 5' and 3' ends for PCR amplification of the library. In summary, each oligo in the library consists of 134 nucleotides. From 5' to 3' the library is composed of :

- A 21-nt universal forward primer region for library amplification.
- A 20-nt gene-specific readout sequence.
- A 1-nt spacer.
- A 20-nt gene-specific readout sequence.
- A 30-nt region of homology to mRNA.
- A 1-nt spacer.
- A 20-nt locus-specific readout sequence.
- A 21-nt universal reverse primer region for library amplification.

4.1.2 Hi-M Library design

Hi-M libraries were constructed following the procedure described in [9, 279]. Briefly, oligopaints probes with 'Balance' settings from Drosophila melanogaster (dm6) or Mouse (mm10), generated with Oligominer [35] were downloaded from `https://oligopaints.hms.harvard.edu/genome-files`. The files consist of a list of homology regions of 35-41 nt that can be used for DNA-FISH probe design. These Oligominer files were used to extract probes using a custom bash script based on the start and the end coordinates of the target region. Next, each oligo is customized to add a tail consisting of a locus-specific readout sequence and a forward and reverse primer sequence positioned at the 5' and 3' ends for PCR amplification of the library. In summary, each oligo in the library consists of approximately 150 nucleotides. From 5' to 3' the library is composed of :

- A 21-nt universal forward primer region for library amplification.
- A 20-nt locus-specific readout sequence.
- A 1-nt spacer.
- A 20-nt locus-specific readout sequence.
- A 35 to 41-nt region of homology to chromosomal DNA.
- A 20-nt locus-specific readout sequence.
- A 20-nt locus-specific readout sequence.
- A 21-nt universal reverse primer region for library amplification.

The resolution of the Hi-M experiment is determined by the number of oligos with the same locus-specific readout sequence. In the case of Drosophila, approximately one probe can be designed for every 60-bp. Thus, in a typical experiment, around 50 probes are used to image a 3-kb genomic region. Conversely, in mouse samples, one probe can be designed for every 100-bp. Thus, in a typical experiment, approximately 230 probes are used to image a 25-kb genomic region.

4.2 Library amplification

The libraries were ordered from GenScript in the 12K oligonucleotide pool format. The procedure for amplifying a specific library from the pool has been described previously [9, 279]. Briefly, the seven-step strategy consists of (i) emulsion PCR (emPCR) to extract the desired library from the pool using a specific pair of primers; (ii) limited-cycle PCR from the emPCR product to determine the optimal amplification cycle; (iii) large-scale PCR with T7 promoter on the reverse primer ;(iv) *in-vitro* transcription using T7 RNA polymerase; (v) reverse transcription; (vi) alkaline hydrolysis; and (vii) purification and concentration of the ssDNA.

4.3 Sample preparation

4.3.1 Drosophila embryo

For this work, the *yw* strain was used as the wild type. The full protocol for embryo collection, fixation and hybridization of the Hi-M library can be found in the lab's Nature Protocols paper [279] and will soon appear in a chapter of Methods in Molecular Biology see *Fiche, Schaeffer, Houbron et al. Methods in Molecular Biology contribution* in 4.6.

4.3.2 Mouse tissues

In this study, Marie Schaeffer prepared the tissues from the wild-type *C57BL/6* mouse strain. To collect the required tissues, the mice were anesthetized with ketamine/xylazine and perfused with cold PBS through the intracardiac route. The collected tissues were then fixed in 4% PFA for 4h at room temperature with gentle agitation, followed by overnight incubation at 4°C. Subsequently, the tissues were then transferred to a PBS solution containing 30% sucrose until they sank to the bottom of the tube. Tissues were then embedded in OCT and stored at -80°C.

4.3.2.1 Attachment of tissues to coverslip

The protocol for tissue attachment was adapted from [315] by Marie Schaeffer. 40mm round coverslips (Bioptechs) were washed with EtOH 70% and activated with air plasma for 30 seconds. The slides were then covered with 100μL of pure 3-aminopropyltrimethoxysilane for 5 min at RT. Slides were left in water for 5-10 min and rinsed 2x10 min in a ddH2O with agitation. A solution containing 0.5% glutaraldehyde in PBS was added for 30 min and rinsed with water. The slides were then coated with 0.1 mg/mL poly-D-lysine in PBS for 1h and incubated O/N with water. 10 μm tissue sections were cut with a cryostat and added to the coated slides, dried at RT for 1-2h, and frozen at -20°C for later use.

4.3.2.2 Sequential RNA-FISH labelling

For RNA-FISH labelling, the following reagents are required :

Product	**Catalog #**
Formamide	Sigma-Aldrich F9037
Paraformaldehyde	Sigma-Aldrich P6148
Dextran sulfate	Sigma-Aldrich D8906
Saline–sodium citrate buffer	Thermo Fisher Scientific J61815.AK
tRNA	Thermo Fisher Scientific AM7119
RVC	Biolabs
BSA	Ambion AM2616

The following solution need to be prepare beforehand :

- **2XSSC :** Mix 100 ml of 20X SSC and 900 ml of distilled water.
- **PFA 4% :** Mix 12.5 mL of PFA 32%, 10mL of 20X SSC and 77.5 mL of $dd\mathrm{H_2O}$.
- **Hybridization buffer (HB) :** Mix 1 mL of 20XSCC, 1g of Dextran sulfate, 3mL of formamide, 500μL of tRNA (20 mg/mL),100μL of RVC (200mM), 40μL BSA and 5.3mL $dd\mathrm{H_2O}$.

- **30 % Formamide wash buffer :** Mix 1.2 mL of Formamide and 2.8mL of PBS.

Protocol :

1. Slides were taken out of the freezer and left on the bench for 1h to dry.

2. Slides were fixed 10 min in PFA 4% and washed in PBS.

3. Slides were incubated with cold 70% EtOh (v/v) in RNAse-free water o/n at 4°C.

4. Slides were rehydrated in 2XSSC for 5 min at RT.

5. Slides were incubated in 30% formamide wash buffer at 37 °C for at least 3h.

6. Slides were placed upside down in petridishes in contact with 2μL of 5 to 10 μg/μL library, in 20 μL of HB and incubated in a humidity-controlled 37°C incubator for 36h.

7. After staining, the slides were washed twice for 30 min each with 30% formamide wash buffer at 45 °C.

8. Slides were whashed twice in 2XSSC and stored in 2XSSC at 4°C until imaging.

4.3.2.3 Hybridization of Hi-M library

For DNA-FISH labelling, the following reagents are required :

Product	**Catalog #**
tri-Sodium citrate dihydrate	Sigma-Aldrich S1804
Salmon sperm	Ambion AM9680
DPBS 1X	Gibco 14190094
RNAse A	Thermo Fisher Scientific EN0531
Tween 20	Sigma-Aldrich P2287

The following solution must be prepared beforehand :

- **2XSSC :** Mix 100 ml of 20X SSC and 900 ml of distilled water.
- **PFA 4% :** Mix 12.5 mL of PFA 32%, 10mL of 20X SSC and 77.5 mL of $dd\mathrm{H_2O}$.
- **FISH hybridization buffer (FHB) :** Mix 5mL of formamide, 2mL of 50% dextran sulfate, 1mL of 20xSSC, 500μL of salmon sperm and 1.5mL of $dd\mathrm{H_2O}$.

- **50 % Formamide wash buffer :** Mix 4mL of Formamide, 0.8mL of 20xSSC and 3.2mL of $dd\mathrm{H_2O}$.
- **40 % Formamide wash buffer :** Mix 1.6mL of Formamide, 0.4mL of 20xSSC and 2mL of $dd\mathrm{H_2O}$.
- **30 % Formamide wash buffer :** Mix 1.2mL of Formamide, 2.8mL of PBS.
- **20 % Formamide wash buffer :** Mix 0.8mL of Formamide, 3.2mL of PBS.
- **10 % Formamide wash buffer :** Mix 0.4mL of Formamide, 3.6mL of PBS.
- **Sodium citrate 10mM, 0.05% Tween, pH 6.0 :** Mix 2.94g of Sodium citrate, 500μL of Tween in 1000 mL.

Protocol :

1. Slides were taken out of the freezer and left on the bench for 1h to dry.

2. Slides were washed with PBS.

3. Slides were incubated with Sodium Citrate 10mM 5 min at RT

4. Slides were incubated with Sodium Citrate 10mM 25 min at 80°C in water bath.

5. Slides were left on the bench for 1h.

6. Slides were washed with 2XSSC 5min.

7. Slides were incubated in 50 % Formamide wash buffer 2h at RT.

8. Slides were placed upside down in petri dishes in contact with 2μL of 5 to 10 μg/μL library,1μL of 100uM the fiducial library , in 20 μL of FHB.

9. Slides were incubated 3h at 45°C in the water bath.

10. Slides were heat-shocked at 85°C 5min in the heat block.

11. Slides were incubated in a humidity-controlled incubator at 37°C O/N.

12. Slides were washed under agitation at 80 rpm, 2 times 40 min in 50 % Formamide wash buffer, 1 time 20 min in 40 % Formamide wash buffer, 1 time 20 min in 30 % Formamide wash buffer, 1 time 20 min in 20 % Formamide wash buffer, 1 time 20min 10 % Formamide wash buffer and 1 time 20 min in 2XSSC.

13. Slides were post fixed with 4% PFA 10 min at RT.

14. Slides were washed 3 times with 2XSSC and stored in 2XSSC at 4°C until imaging.

4.4 Microscope setup

Experiments were performed on a home-made imaging setup built on a RAMM modular microscope system (Applied Scientific Instrumentation) coupled to an improved microfluidic device, as the one described previously [9, 279]. Software-controlled microscope components, including camera, stages, lasers, needles, pump and valves, were run using Qudi-HiM, an homemade software developed in python.

4.5 Image acquisition

For Image acquisition, the following reagents are required :

Product	**Catalog #**
Glucose oxidase	Sigma-Aldrich G2133
DAPI	Sigma-Aldrich D9542
Catalase	Sigma-Aldrich C9322
Alexa 647 readout probes	Integrated DNA Technologies
Alexa 488 readout probes	Integrated DNA Technologies
Primer adapter	Integrated DNA Technologies
Tris(2-carboxyethyl)phosphine hydrochlorid	Sigma-Aldrich 646547

The following solution must be prepared beforehand :

- **2XSSC :** Mix 100 ml of 20XSSC and 900 ml of ddH2O.
- **50 % Formamide wash buffer :** Mix 4mL of Formamide, 0.8mL of 20XSSC and 3.2mL of $dd\mathrm{H_2O}$.
- **30 % Formamide wash buffer :** Mix 1.2mL of Formamide, 2.8mL of PBS.
- **Gloxy :** Mix 50 mg of glucose oxydase, 100μL of catalase and 900μL of 11mM Tris-HCL / 55 mM NaCl.
- **Imaging Buffer :** Mix 110μL pf Gloxy, 1mL of Glucose 50% and 8.89 mL of PBS.
- **Chemical bleaching :** Mix 1mL ampoule of TCEP with 9 mL of 2XSSC.
- **Imaging oligo mix :** Mix 0.6μL of imaging oligo (100uM) in 2mL 50 % Formamide wash buffer.
- **Fiducial and mask0 oligo mix :** Mix 0.6μL of imaging oligo for fiducial (100uM), 0.6μL of imaging oligo for mask0 (100uM) in 2mL 50 % Formamide wash buffer .

- **Primer adapter :** Mix 0.6μL of primer adapter (100uM) in 2mL 50 % Formamide wash buffer.
- **DAPI :** Prepare 0.5 μg/ml of DAPI 2mL PBS.

4.5.1 Sequential RNA-FISH acquisition

Slides were mounted in FCS2 chamber prior to RNA-FISH acquisition.

Protocol :

1. Several 200μm x 200μm regions of interest (ROI) were selected on the tissue. These regions were bleached at the maximum laser power to reduce tissue autofluorescence. 488 nm, 561nm and 647nm lasers were used simultaneously.

2. 1.4 mL of the DAPI solution was injected and incubated for 600 seconds.

3. Acquisitions were performed with DAPI (405nm) on all ROIs in Z-stacks of 15 - 20 μm.

4. RNA-FISH experiment was launched as follows.

Image acquisition		
Buffer	Volume (μL)	Flow rate (μ/min)
Imaging oligo	1400	150
Incubation	600 sec	
30 % Formamide wash buffer	2000	150
2XSSC	1000	150
Imaging Buffer	900	150
Bleaching		
TCEP	750	150
2XSSC	750	150

4.5.2 Hi-M acquisition

Slides were mounted in FCS2 chamber pior to Hi-M acquisition.

Protocol :

1. Several 200μm x 200μm regions of interest (ROI) were selected on the tissue. These regions were bleached at the maximum laser power to reduce tissue autofluorescence.

2. 1.4 mL of the primer adapter solution was injected and incubated for 600 seconds.

3. Samples were washed with 2.8 mL of the 50% Formamide wash buffer.

4. 1.4 mL of the fiducial and mask0 oligo mix was injected and incubated for 600 seconds to stain the mask of the library (mask0) and the fiducial library.

5. Samples were washed with 2.8 mL of the 50% Formamide wash buffer.

6. Samples were washed with 1.4 mL of 2XSCC.

7. 1.4 mL of the DAPI solution was injected and incubated for 600 seconds.

8. 900 μL of Imaging Buffer (IB) was injected.

9. Acquisitions were performed with DAPI (405nm), fiducial (488nm), and mask0 (647nm) on all ROIs in Z-stacks of 15 - 20 μm.

10. The fluorescent signal from the mask0 was bleached with 1 mL of TCEP solution

11. Samples were washed with 1 mL of 2XSCC and the Hi-M experiment was launched as follows.

Image acquisition		
Buffer	Volume (μL)	Flow rate (μ/min)
Imaging oligo	1400	150
Incubation	600 sec	
50 % Formamide wash buffer	2000	150
2xSSC	1000	150
Imaging Buffer	900	150
Bleaching		
TCEP	750	150
2xSSC	750	150

4.6 Data analysis

TIFF images were then deconvolved using Huygens Professional 21.04 (Scientific Volume Imaging, `https://svi.nl`). The analysis was performed using our pyHiM analysis pipeline `https://pyhim.readthedocs.io/en/latest/`.

Chapter 5: Appendix

Cardozo et al. Mol Cell 2019 contribution

In this work published in **Molecular Cell** in 2019, we presented our recently developed imaging tracing technique, Hi-M, which allows for the detection of chromatin conformation and transcription of a gene in intact *Drosophila* embryo at very high resolution. We applied this technique to study the relation between chromosome conformation and transcription at the *sna* and *esg* locus in *Drosophila*. We show that TADs emerge at the onset of zygotic genome activation (ZGA), in agreement with Hi-C data [6, 8]. We observed a partial similarity in the TAD structure between paired and unpaired chromosomes. We show that TADs in cells where the genes are active have a different architecture than those in inactive cells. This suggests that transcriptional activity plays an important role in shaping the TADs at the *sna* and *esg* locus.

In this research project, I contributed to the development of the Hi-M technique, working closely with Julian Gurgo and Christophe Houbron on various aspects of the development, including fly handling, embryo collection, fixation, and optimization of the labeling protocol. In addition, I actively participated with Sergio Espinola and Jean-Bernard Fiche in the acquisition of the Hi-M dataset using the microscope. This collaborative effort involved the collective work of Andres Cardozo-Gizzi, Sergio Espinola, Julian Gurgo, Christophe Houbron, Jean-Bernard Fiche, and was conducted under the supervision of Marcelo Nollmann. My work was supervied by Diego Cattoni and Marcelo Nollmann.

Title : Microscopy-Based Chromosome Conformation Capture Enables Simultaneous Visualization of Genome Organization and Transcription in Intact Organisms.
Authors : Andrés M Cardozo Gizzi, Diego I Cattoni, Jean-Bernard Fiche, Sergio M Espinola, Julian Gurgo, **Olivier Messina**, Christophe Houbron, Yuki Ogiyama, Giorgio L Papadopoulos, Giacomo Cavalli, Mounia Lagha, Marcelo Nollmann.
Journal : Molecular Cell.
Statut : Published.
Year : 2019.
URL : `https://pubmed.ncbi.nlm.nih.gov/30795893/`

Espinola, Götz et al. Nat Gen 2021 contribution

In this work published in **Nature Genetics** in 2021, we used chromosome tracing approach Hi-M to study the looping interactions within topologically associating domains (TADs) at the *doc* locus. We show that chromatin loops within TADs of the *doc* locus correspond to CREs, including E–P (enhancer-promoter), P–P (promoter-promoter), and E–E (enhancer-enhancer). Surprisingly, we show that these contacts are established even before the formation of TADs, as early as nc11. In particular, we found that CREs can act as hubs, facilitating intricate multi-way interactions among themselves. Strikingly, these CRE-mediated interactions are not restricted to *doc* transcribing cells, such as the dorsal ectoderm, but are also present in non-transcribing cells, including the mesoderm and neuroectoderm. Finally we found that the pioneer factor Zelda is required for the establishment of CREs interaction at the *doc* locus.

In this research project, I developed a bioinformatics algorithm based on published Hi-C datasets [6] to investigate the impact of the pioneer factor Zelda on the formation of physical contacts between CREs genome-wide. I observed that regions bound by Zelda exhibit a higher frequency of interactions than what would be expected by chance. Importantly, I found that these interactions are significantly affected by Zelda depletion, but do not necessarily require active transcription to form. My work was supervised by Mounia Lagha and Marcelo Nollmann.

Title : Cis-regulatory chromatin loops arise before TADs and gene activation, and are independent of cell fate during early Drosophila development.
Authors : Sergio Martin Espinola, Markus Götz, Maelle Bellec, **Olivier Messina**, Jean-Bernard Fiche, Christophe Houbron, Matthieu Dejean, Ingolf Reim, Andrés M Cardozo Gizzi, Mounia Lagha, Marcelo Nollmann.
Journal : Nature Genetics.
Statut : Published.
Year : 2021.
URL : `https://pubmed.ncbi.nlm.nih.gov/33795867/`

Götz et al. Nat Com 2022 contribution

In this work published in **Nature Communications** in 2022, we used chromosome tracing approach Hi-M to study the looping interactions between CREs at the *doc* locus at the single cell level in the *Drosophila* embryo. We show that the emergence of TADs partially segregates the conformational space explored by single nuclei during the early development of *Drosophila* embryos. More surprisingly, we show that active transcription within a TAD leads to minor changes to the local inter- and intra-TAD chromatin conformation in single nuclei and only weakly affects insulation to the neighboring TAD. Collectively, our results provide insight into the role of TAD insulation, TAD condensation and transcription in shaping chromatin structure at the single cell level.

In this research project, I actively participated with Sergio Espinola both in the acquisition of the Hi-M dataset on the microscope and in the subsequent data analyses. I was responsible for the analysis of the ChIP-seq and ATAC-seq data used in the study. Additionally, I have worked on data visualization and integration using Jupyter Notebook along with the CoolBox package (`https://gangcaolab.github.io/CoolBox/installation.html`). This allowed us to effectively interpolate the Hi-M pairwise distance matrix and integrate ATAC-seq and ChIP-seq tracks. My work was supervised by Marcelo Nollmann.

Title : Multiple parameters shape the 3D chromatin structure of single nuclei at the doc locus in Drosophila.
Authors : Markus Götz, **Olivier Messina**, Sergio Espinola, Jean-Bernard Fiche, Marcelo Nollmann.
Journal : Nature Communications.
Statut : Published.
Year : 2022.
URL : `https://pubmed.ncbi.nlm.nih.gov/36104317/`

Barho et al. Open Research Europe 2022 contribution

In this work published in **Open Research Europe** in 2022, we developed Qudi-HiM, a software package written in Python 3 that allows the acquisition of thousands of three-dimensional multicolor microscopy images, the manipulation of microfluidic devices, and the remote monitoring of ongoing acquisitions and real-time analysis.

In this research project, I actively participated in the validation of the acquisition software by identifying and reporting bugs. Qudi-HiM is now used as a routine software in the lab for Hi-M acquisition and is available in open access for the microscopy community. My work was supervised by Marcelo Nollmann.

Title : Qudi-HiM: an open-source acquisition software package for highly multiplexed sequential and combinatorial optical imaging .
Authors : Franziska Barho, Jean-Bernard Fiche, Marion Bardou, **Olivier Messina**, Alexandre Martiniere, Christophe Houbron, Marcelo Nollmann.
Journal : Open Research Europe.
Statut : Published.
Year : 2022.
URL : `https://pubmed.ncbi.nlm.nih.gov/37645324/`

Devos, Fiche et al. Genome Biology contribution

In this work under-review in **Genome Biology**, we developed pyHiM, a software package written in Python 3 that allows the analysis of multiplexed DNA-FISH data.

In this research project, I actively participated in the validation of the acquisition software by identifying and reporting bugs. pyHiM is now used as a routine software in the lab for Hi-M data analysis and is available in open access for the microscopy community `https://pyhim.readthedocs.io/en/latest/`. My work was supervised by Marcelo Nollmann.

Title : pyHiM, a new open-source, multi-platform software package for spatial genomics based on multiplexed DNA-FISH imaging.
Authors : Devos Xavier, Fiche Jean-Bernard, Bardou Marion, **Messina Olivier**, Houbron Christophe, Gurgo Julian, Schaeffer Marie, Götz Markus, Walter Thomas, Mueller Florian, Nollmann Marcelo.
Journal : Genome Biology.
Statut : Under-review.
Year : 2023.
URL :
`https://www.biorxiv.org/content/10.1101/2023.09.19.558412v1.full.pdf`

Fiche, Schaeffer, Houbron et al. Methods in Molecular Biology contribution

In this submission to **Methods in Molecular Biology**, we had the honor of contributing to the volume dedicated to Fluorescent In Situ Hybridization. Our contribution provides a comprehensive presentation of the Hi-M protocol and Qudi-HiM for the broader scientific community.

During this research project, I played an active role in optimizing the Hi-M protocol for use in *Drosophila* embryos. Additionally, I contributed to the validation of the acquisition software Qudi-HiM by identifying and reporting bugs. My work was conducted under the supervision of Marcelo Nollmann.

Title : Hi-M: a Multiplex oligopaint FISH method to capture chromatin conformations in situ and accompanying open-source acquisition software.
Authors : Jean-Bernard Fiche, Marie Schaeffer, Christophe Houbron, Christel Elkhoury Youhanna, **Olivier Messina**, Franziska Barho, Marcelo Nollmann.
Journal : Methods in Molecular Biology.
Statut : Under-review.
Year : 2023.

References

[1] Erez Lieberman-Aiden, Nynke L van Berkum, Louise Williams, Maxim Imakaev, Tobias Ragoczy, Agnes Telling, Ido Amit, Bryan R Lajoie, Peter J Sabo, Michael O Dorschner, Richard Sandstrom, Bradley Bernstein, M A Bender, Mark Groudine, Andreas Gnirke, John Stamatoyannopoulos, Leonid A Mirny, Eric S Lander, and Job Dekker. "Comprehensive mapping of long-range interactions reveals folding principles of the human genome." In: *Science (New York, N.Y.)* (2009).

[2] Jesse R Dixon, Siddarth Selvaraj, Feng Yue, Audrey Kim, Yan Li, Yin Shen, Ming Hu, Jun S Liu, and Bing Ren. "Topological domains in mammalian genomes identified by analysis of chromatin interactions." In: *Nature* (2012).

[3] Elphège P Nora, Bryan R Lajoie, Edda G Schulz, Luca Giorgetti, Ikuhiro Okamoto, Nicolas Servant, Tristan Piolot, Nynke L van Berkum, Johannes Meisig, John Sedat, Joost Gribnau, Emmanuel Barillot, Nils Blüthgen, Job Dekker, and Edith Heard. "Spatial partitioning of the regulatory landscape of the X-inactivation centre." In: *Nature* (2012).

[4] Tom Sexton, Eitan Yaffe, Ephraim Kenigsberg, Frédéric Bantignies, Benjamin Leblanc, Michael Hoichman, Hugues Parrinello, Amos Tanay, and Giacomo Cavalli. "Three-dimensional folding and functional organization principles of the Drosophila genome." In: *Cell* (2012).

[5] Quentin Szabo, Frédéric Bantignies, and Giacomo Cavalli. "Principles of genome folding into topologically associating domains." In: *Science advances* (2019).

[6] Clemens B Hug, Alexis G Grimaldi, Kai Kruse, and Juan M Vaquerizas. "Chromatin Architecture Emerges during Zygotic Genome Activation Independent of Transcription." In: *Cell* (2017).

[7] Denes Hnisz, Abraham S Weintraub, Daniel S Day, Anne-Laure Valton, Rasmus O Bak, Charles H Li, Johanna Goldmann, Bryan R Lajoie, Zi Peng Fan, Alla A Sigova, Jessica Reddy, Diego Borges-Rivera, Tong Ihn Lee, Rudolf Jaenisch, Matthew H Porteus, Job Dekker, and Richard A Young. "Activation of proto-oncogenes by disruption of chromosome neighborhoods." In: *Science (New York, N.Y.)* (2016).

[8] Yuki Ogiyama, Bernd Schuettengruber, Giorgio L Papadopoulos, Jia-Ming Chang, and Giacomo Cavalli. "Polycomb-Dependent Chromatin Looping Contributes to Gene Silencing during Drosophila Development." In: *Molecular cell* (2018).

[9] Andrés M Cardozo Gizzi, Diego I Cattoni, Jean-Bernard Fiche, Sergio M Espinola, Julian Gurgo, Olivier Messina, Christophe Houbron, Yuki Ogiyama, Giorgio L Papadopoulos, Giacomo Cavalli, Mounia Lagha, and Marcelo Nollmann. "Microscopy-Based Chromosome Conformation Capture Enables Simultaneous Visualization of Genome Organization and Transcription in Intact Organisms." In: *Molecular cell* (2019).

[10] Boyan Bonev, Netta Mendelson Cohen, Quentin Szabo, Lauriane Fritsch, Giorgio L Papadopoulos, Yaniv Lubling, Xiaole Xu, Xiaodan Lv, Jean-Philippe Hugnot, Amos Tanay, and Giacomo Cavalli. "Multiscale 3D Genome Rewiring during Mouse Neural Development." In: *Cell* (2017).

[11] Irene Miguel-Escalada, Silvia Bonàs-Guarch, Inês Cebola, Joan Ponsa-Cobas, Julen Mendieta-Esteban, Goutham Atla, Biola M Javierre, Delphine M Y Rolando, Irene Farabella, Claire C Morgan, Javier García-Hurtado, Anthony Beucher, Ignasi Morán, Lorenzo Pasquali, Mireia Ramos-Rodríguez, Emil V R Appel, Allan Linneberg, Anette P Gjesing, Daniel R Witte, Oluf Pedersen, Niels Grarup, Philippe Ravassard, David Torrents, Josep M Mercader, Lorenzo Piemonti, Thierry Berney, Eelco J P de Koning, Julie Kerr-Conte, François Pattou, Iryna O Fedko, Leif Groop, Inga Prokopenko, Torben Hansen, Marc A Marti-Renom, Peter Fraser, and Jorge Ferrer. "Human pancreatic islet three-dimensional chromatin architecture provides insights into the genetics of type 2 diabetes." In: *Nature genetics* (2019).

[12] Sergio Martin Espinola, Markus Götz, Maelle Bellec, Olivier Messina, Jean-Bernard Fiche, Christophe Houbron, Matthieu Dejean, Ingolf Reim, Andrés M Cardozo Gizzi, Mounia Lagha, and Marcelo Nollmann. "Cis-regulatory chromatin loops arise before TADs and gene activation, and are independent of cell fate during early Drosophila development." In: *Nature genetics* (2021).

[13] Elizabeth Ing-Simmons, Roshan Vaid, Xin Yang Bing, Michael Levine, Mattias Mannervik, and Juan M Vaquerizas. "Independence of chromatin conformation and gene regulation during Drosophila dorsoventral patterning." In: *Nature genetics* (2021).

[14] Warren Winick-Ng, Alexander Kukalev, Izabela Harabula, Luna Zea-Redondo, Dominik Szabó, Mandy Meijer, Leonid Serebreni, Yingnan Zhang, Simona Bianco, Andrea M Chiariello, Ibai Irastorza-Azcarate, Christoph J Thieme, Thomas M Sparks, Sílvia Carvalho, Luca Fiorillo, Francesco Musella, Ehsan Irani, Elena Torlai Triglia, Aleksandra A Kolodziejczyk, Andreas Abentung, Galina Apostolova, Eleanor J Paul, Vedran Franke, Rieke Kempfer, Altuna Akalin, Sarah A Teichmann, Georg Dechant, Mark A Ungless, Mario Nicodemi, Lonnie Welch, Gonçalo Castelo-Branco, and Ana Pombo. "Cell-type specialization is encoded by specific chromatin topologies." In: *Nature* (2021).

[15] Geoffrey Fudenberg, Maxim Imakaev, Carolyn Lu, Anton Goloborodko, Nezar Abdennur, and Leonid A Mirny. "Formation of Chromosomal Domains by Loop Extrusion." In: *Cell reports* (2016).

[16] Nicholas E Matthews and Rob White. "Chromatin Architecture in the Fly: Living without CTCF/Cohesin Loop Extrusion?: Alternating Chromatin States Provide a Basis for Domain Architecture in Drosophila." In: *BioEssays : news and reviews in molecular, cellular and developmental biology* (2019).

[17] Peter Heger, Rebecca George, and Thomas Wiehe. "Successive gain of insulator proteins in arthropod evolution." In: *Evolution; international journal of organic evolution* (2013).

[18] S M Parkhurst and V G Corces. "Interactions among the gypsy transposable element and the yellow and the suppressor of hairy-wing loci in Drosophila melanogaster." In: *Molecular and cellular biology* (1986).

[19] C Spana, D A Harrison, and V G Corces. "The Drosophila melanogaster suppressor of Hairy-wing protein binds to specific sequences of the gypsy retrotransposon." In: *Genes development* (1988).

[20] K Zhao, C M Hart, and U K Laemmli. "Visualization of chromosomal domains with boundary element-associated factor BEAF-32." In: *Cell* (1995).

[21] T I Gerasimova, D A Gdula, D V Gerasimov, O Simonova, and V G Corces. "A Drosophila protein that imparts directionality on a chromatin insulator is an enhancer of position-effect variegation." In: *Cell* (1995).

[22] A C Bell, A G West, and G Felsenfeld. "The protein CTCF is required for the enhancer blocking activity of vertebrate insulators." In: *Cell* (1999).

[23] M Gaszner, J Vazquez, and P Schedl. "The Zw5 protein, a component of the scs chromatin domain boundary, is able to block enhancer-promoter interaction." In: *Genes development* (1999).

[24] K C Scott, A D Taubman, and P K Geyer. "Enhancer blocking by the Drosophila gypsy insulator depends upon insulator anatomy and enhancer strength." In: *Genetics* (1999).

[25] K Büchner, P Roth, G Schotta, V Krauss, H Saumweber, G Reuter, and R Dorn. "Genetic and molecular complexity of the position effect variegation modifier mod(mdg4) in Drosophila." In: *Genetics* (2000).

[26] Chi-Yun Pai, Elissa P Lei, Dolanchanpa Ghosh, and Victor G Corces. "The centrosomal protein CP190 is a component of the gypsy chromatin insulator." In: *Molecular cell* (2004).

[27] Hanlim Moon, Galina Filippova, Dmitry Loukinov, Elena Pugacheva, Qi Chen, Sheryl T Smith, Adam Munhall, Britta Grewe, Marek Bartkuhn, Rüdiger Arnold, Les J Burke, Renate Renkawitz-Pohl, Rolf Ohlsson, Jumin Zhou, Rainer Renkawitz, and Victor Lobanenkov. "CTCF is conserved from Drosophila to humans and confers enhancer blocking of the Fab-8 insulator." In: *EMBO reports* (2005).

[28] Nikolay Zolotarev, Anna Fedotova, Olga Kyrchanova, Artem Bonchuk, Aleksey A Penin, Andrey S Lando, Irina A Eliseeva, Ivan V Kulakovskiy, Oksana Maksimenko, and Pavel Georgiev. "Architectural proteins Pita, Zw5,and ZIPIC contain homodimerization domain and support specific long-range interactions in Drosophila." In: *Nucleic acids research* (2016).

[29] Nicolas Nègre, Christopher D Brown, Parantu K Shah, Pouya Kheradpour, Carolyn A Morrison, Jorja G Henikoff, Xin Feng, Kami Ahmad, Steven Russell, Robert A H White, Lincoln Stein, Steven Henikoff, Manolis Kellis, and Kevin P White. "A comprehensive map of insulator elements for the Drosophila genome." In: *PLoS genetics* (2010).

[30] Leslie J Mateo, Sedona E Murphy, Antonina Hafner, Isaac S Cinquini, Carly A Walker, and Alistair N Boettiger. "Visualizing DNA folding and RNA in embryos at single-cell resolution." In: *Nature* (2019).

[31] Siyuan Wang, Jun-Han Su, Brian J Beliveau, Bogdan Bintu, Jeffrey R Moffitt, Chao-ting Wu, and Xiaowei Zhuang. "Spatial organization of chromatin domains and compartments in single chromosomes." In: *Science (New York, N.Y.)* (2016).

[32] Bogdan Bintu, Leslie J Mateo, Jun-Han Su, Nicholas A Sinnott-Armstrong, Mirae Parker, Seon Kinrot, Kei Yamaya, Alistair N Boettiger, and Xiaowei Zhuang. "Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells." In: *Science (New York, N.Y.)* (2018).

[33] Vera Pancaldi, Enrique Carrillo-de-Santa-Pau, Biola Maria Javierre, David Juan, Peter Fraser, Mikhail Spivakov, Alfonso Valencia, and Daniel Rico. "Integrating epigenomic data and 3D genomic structure with a new measure of chromatin assortativity." In: *Genome biology* (2016).

[34] Jeffrey R Moffitt, Junjie Hao, Guiping Wang, Kok Hao Chen, Hazen P Babcock, and Xiaowei Zhuang. "High-throughput single-cell gene-expression profiling with multiplexed error-robust fluorescence in situ hybridization." In: *Proceedings of the National Academy of Sciences of the United States of America* (2016).

[35] Brian J Beliveau, Jocelyn Y Kishi, Guy Nir, Hiroshi M Sasaki, Sinem K Saka, Son C Nguyen, Chao-Ting Wu, and Peng Yin. "OligoMiner provides a rapid, flexible environment for the design of genome-scale oligonucleotide in situ hybridization probes." In: *Proceedings of the National Academy of Sciences of the United States of America* (2018).

[36] Ruo-Ran Wang, Xinyuan Qiu, Ran Pan, Hongxing Fu, Ziyin Zhang, Qintao Wang, Haide Chen, Qing-Qian Wu, Xiaowen Pan, Yanping Zhou, Pengfei Shan, Shusen Wang, Guoji Guo, Min Zheng, Lingyun Zhu, and Zhuo-Xian Meng. "Dietary intervention preserves cell function in mice through CTCF-mediated transcriptional reprogramming." In: *The Journal of experimental medicine* (2022).

[37] Ildem Akerman, Zhidong Tu, Anthony Beucher, Delphine M Y Rolando, Claire Sauty-Colace, Marion Benazra, Nikolina Nakic, Jialiang Yang, Huan Wang, Lorenzo Pasquali, Ignasi Moran, Javier Garcia-Hurtado, Natalia Castro, Roser Gonzalez-Franco, Andrew F Stewart, Caroline Bonner, Lorenzo Piemonti, Thierry Berney, Leif Groop, Julie Kerr-Conte, Francois Pattou, Carmen Argmann, Eric Schadt, Philippe Ravassard, and Jorge Ferrer. "Human Pancreatic Cell lncRNAs Control Cell-Specific Regulatory Networks." In: *Cell metabolism* (2017).

[38] Edward N Trifonov. "Vocabulary of definitions of life suggests a definition." In: *Journal of biomolecular structure dynamics* (2011).

[39] Ralf Dahm. "Discovering DNA: Friedrich Miescher and the early years of nucleic acid research." In: *Human genetics* (2008).

[40] J D WATSON and F H CRICK. “Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid.” In: *Nature* (1953).

[41] Flemming W. “Zellsubstanz, Kern und Zelltheilung”. In: *Book* (1882).

[42] J B GURDON. “The developmental capacity of nuclei taken from intestinal epithelium cells of feeding tadpoles.” In: *Journal of embryology and experimental morphology* (1962).

[43] Waddington Conrad Hal. “The Strategy of Genes”. In: *Book* (1957).

[44] Monod J. Pardee AB Jacob F. “The genetic control and cytoplasmic expression of“inducibility” in the synthesis of -galactosidase in Escherichia coli”. In: *J Mol Biol* (1959).

[45] Stephen T Smale and James T Kadonaga. “The RNA polymerase II core promoter.” In: *Annual review of biochemistry* (2003).

[46] R P Lifton, M L Goldberg, R W Karp, and D S Hogness. “The organization of the histone genes in Drosophila melanogaster: functional and evolutionary implications.” In: *Cold Spring Harbor symposia on quantitative biology* (1978).

[47] Istvan Albert, Travis N Mavrich, Lynn P Tomsho, Ji Qi, Sara J Zanton, Stephan C Schuster, and B Franklin Pugh. “Translational and rotational settings of H2A.Z nucleosomes across the Saccharomyces cerevisiae genome.” In: *Nature* (2007).

[48] Artem Barski, Suresh Cuddapah, Kairong Cui, Tae-Young Roh, Dustin E Schones, Zhibin Wang, Gang Wei, Iouri Chepelev, and Keji Zhao. “High-resolution profiling of histone methylations in the human genome.” In: *Cell* (2007).

[49] Boris Lenhard, Albin Sandelin, and Piero Carninci. “Metazoan promoters: emerging characteristics and insights into transcriptional regulation.” In: *Nature reviews. Genetics* (2012).

[50] Bradley E Bernstein, Tarjei S Mikkelsen, Xiaohui Xie, Michael Kamal, Dana J Huebert, James Cuff, Ben Fry, Alex Meissner, Marius Wernig, Kathrin Plath, Rudolf Jaenisch, Alexandre Wagschal, Robert Feil, Stuart L Schreiber, and Eric S Lander. “A bivalent chromatin structure marks key developmental genes in embryonic stem cells.” In: *Cell* (2006).

[51] M Cuadrado, M Sacristán, and F Antequera. “Species-specific organization of CpG island promoters at mammalian homologous genes.” In: *EMBO reports* (2001).

[52] Daria Shlyueva, Gerald Stampfel, and Alexander Stark. “Transcriptional enhancers: from properties to genome-wide predictions.” In: *Nature reviews. Genetics* (2014).

[53] C Benoist and P Chambon. “In vivo sequence requirements of the SV40 early promotor region.” In: *Nature* (1981).

[54] J Banerji, S Rusconi, and W Schaffner. “Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences.” In: *Cell* (1981).

[55] J Banerji, L Olson, and W Schaffner. "A lymphocyte-specific cellular enhancer is located downstream of the joining region in immunoglobulin heavy chain genes." In: *Cell* (1983).

[56] E M Blackwood and J T Kadonaga. "Going the distance: a current view of enhancer action." In: *Science (New York, N.Y.)* (1998).

[57] J Omar Yáñez-Cuna, Evgeny Z Kvon, and Alexander Stark. "Deciphering the transcriptional cis-regulatory code." In: *Trends in genetics : TIG* (2013).

[58] Lisa Ann Cirillo, Frank Robert Lin, Isabel Cuesta, Dara Friedman, Michal Jarnik, and Kenneth S Zaret. "Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4." In: *Molecular cell* (2002).

[59] Makiko Iwafuchi-Doi, Greg Donahue, Akshay Kakumanu, Jason A Watts, Shaun Mahony, B Franklin Pugh, Dolim Lee, Klaus H Kaestner, and Kenneth S Zaret. "The Pioneer Transcription Factor FoxA Maintains an Accessible Nucleosome Configuration at Enhancers for Tissue-Specific Gene Activation." In: *Molecular cell* (2016).

[60] Jason D Buenrostro, Paul G Giresi, Lisa C Zaba, Howard Y Chang, and William J Greenleaf. "Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position." In: *Nature methods* (2013).

[61] Francesca De Santa, Iros Barozzi, Flore Mietton, Serena Ghisletti, Sara Polletti, Betsabeh Khoramian Tusi, Heiko Muller, Jiannis Ragoussis, Chia-Lin Wei, and Gioacchino Natoli. "A large fraction of extragenic RNA pol II transcription sites overlap enhancers." In: *PLoS biology* (2010).

[62] Tae-Kyung Kim, Martin Hemberg, Jesse M Gray, Allen M Costa, Daniel M Bear, Jing Wu, David A Harmin, Mike Laptewicz, Kellie Barbara-Haley, Scott Kuersten, Eirene Markenscoff-Papadimitriou, Dietmar Kuhl, Haruhiko Bito, Paul F Worley, Gabriel Kreiman, and Michael E Greenberg. "Widespread transcription at neuronal activity-regulated enhancers." In: *Nature* (2010).

[63] Frederic Koch, Romain Fenouil, Marta Gut, Pierre Cauchy, Thomas K Albert, Joaquin Zacarias-Cabeza, Salvatore Spicuglia, Albane Lamy de la Chapelle, Martin Heidemann, Corinna Hintermair, Dirk Eick, Ivo Gut, Pierre Ferrier, and Jean-Christophe Andrau. "Transcription initiation platforms and GTF recruitment at tissue-specific enhancers and promoters." In: *Nature structural molecular biology* (2011).

[64] Sarah Djebali, Carrie A Davis, Angelika Merkel, Alex Dobin, Timo Lassmann, Ali Mortazavi, Andrea Tanzer, Julien Lagarde, Wei Lin, Felix Schlesinger, Chenghai Xue, Georgi K Marinov, Jainab Khatun, Brian A Williams, Chris Zaleski, Joel Rozowsky, Maik Röder, Felix Kokocinski, Rehab F Abdelhamid, Tyler Alioto, Igor Antoshechkin, Michael T Baer, Nadav S Bar, Philippe Batut, Kimberly Bell, Ian Bell, Sudipto Chakrabortty, Xian Chen, Jacqueline Chrast, Joao Curado, Thomas Derrien, Jorg Drenkow, Erica Dumais, Jacqueline Dumais, Radha Duttagupta, Emilie Falconnet, Meagan Fastuca, Kata

Fejes-Toth, Pedro Ferreira, Sylvain Foissac, Melissa J Fullwood, Hui Gao, David Gonzalez, Assaf Gordon, Harsha Gunawardena, Cedric Howald, Sonali Jha, Rory Johnson, Philipp Kapranov, Brandon King, Colin Kingswood, Oscar J Luo, Eddie Park, Kimberly Persaud, Jonathan B Preall, Paolo Ribeca, Brian Risk, Daniel Robyr, Michael Sammeth, Lorian Schaffer, Lei-Hoon See, Atif Shahab, Jorgen Skancke, Ana Maria Suzuki, Hazuki Takahashi, Hagen Tilgner, Diane Trout, Nathalie Walters, Huaien Wang, John Wrobel, Yanbao Yu, Xiaoan Ruan, Yoshihide Hayashizaki, Jennifer Harrow, Mark Gerstein, Tim Hubbard, Alexandre Reymond, Stylianos E Antonarakis, Gregory Hannon, Morgan C Giddings, Yijun Ruan, Barbara Wold, Piero Carninci, Roderic Guigó, and Thomas R Gingeras. "Landscape of transcription in human cells." In: *Nature* (2012).

[65] Erik Arner et al. "Transcribed enhancers lead waves of coordinated transcription in transitioning mammalian cells." In: *Science (New York, N.Y.)* (2015).

[66] Feng Liu. "Enhancer-derived RNA: A Primer." In: *Genomics, proteomics bioinformatics* (2017).

[67] Steven Blinka, Michael H Reimer, Kirthi Pulakanti, Luca Pinello, Guo-Cheng Yuan, and Sridhar Rao. "Identification of Transcribed Enhancers by Genome-Wide Chromatin Immunoprecipitation Sequencing." In: *Methods in molecular biology (Clifton, N.J.)* (2017).

[68] Sebastian Pott and Jason D Lieb. "What are super-enhancers?" In: *Nature genetics* (2015).

[69] Stephen S Gisselbrecht, Luis A Barrera, Martin Porsch, Anton Aboukhalil, Preston W Estep, Anastasia Vedenko, Alexandre Palagi, Yongsok Kim, Xianmin Zhu, Brian W Busser, Caitlin E Gamble, Antonina Iagovitina, Aditi Singhania, Alan M Michelson, and Martha L Bulyk. "Highly parallel assays of tissue-specific enhancers in whole Drosophila embryos." In: *Nature methods* (2013).

[70] Ilaria Mogno, Jamie C Kwasnieski, and Barak A Cohen. "Massively parallel synthetic promoter assays reveal the in vivo effects of binding site variants." In: *Genome research* (2013).

[71] Alexandre Melnikov, Anand Murugan, Xiaolan Zhang, Tiberiu Tesileanu, Li Wang, Peter Rogov, Soheil Feizi, Andreas Gnirke, Curtis G Callan, Justin B Kinney, Manolis Kellis, Eric S Lander, and Tarjei S Mikkelsen. "Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay." In: *Nature biotechnology* (2012).

[72] Rupali P Patwardhan, Joseph B Hiatt, Daniela M Witten, Mee J Kim, Robin P Smith, Dalit May, Choli Lee, Jennifer M Andrie, Su-In Lee, Gregory M Cooper, Nadav Ahituv, Len A Pennacchio, and Jay Shendure. "Massively parallel functional dissection of mammalian enhancers in vivo." In: *Nature biotechnology* (2012).

[73] Cosmas D Arnold, Daniel Gerlach, Christoph Stelzer, Łukasz M Boryń, Martina Rath, and Alexander Stark. "Genome-wide quantitative enhancer activity maps identified by STARR-seq." In: *Science (New York, N.Y.)* (2013).

[74] Felix Muerdter, Łukasz M Boryń, and Cosmas D Arnold. "STARR-seq - principles and applications." In: *Genomics* (2015).

[75] Tanya Vavouri, Klaudia Walter, Walter R Gilks, Ben Lehner, and Greg Elgar. "Parallel evolution of conserved non-coding elements that target a common set of developmental regulatory genes from worms to humans." In: *Genome biology* (2007).

[76] Chen Zhuoxin, Snetkova Valentina, Bower Grace, Jacinto Sandra, Clock Benjamin, Barozzi Iros, J Mannion Brandon, Alcaina-Caro Ana, Lopez-Rios Javier, Dickel Diane E, Visel Axel, Pennacchio Len, and Evgeny Z Kvon. "Widespread Increase in Enhancer—Promoter Interactions during Developmental Enhancer Activation in Mammals". In: *BioRxiv* (2022).

[77] J R Morris, P K Geyer, and C T Wu. "Core promoter elements can regulate transcription on a separate chromosome in trans." In: *Genes development* (1999).

[78] Stavros Lomvardas, Gilad Barnea, David J Pisapia, Monica Mendelsohn, Jennifer Kirkland, and Richard Axel. "Interchromosomal interactions and olfactory receptor choice." In: *Cell* (2006).

[79] Petros Kolovos, Tobias A Knoch, Frank G Grosveld, Peter R Cook, and Argyris Papantonis. "Enhancers and silencers: an integrated and simple model for their function." In: *Epigenetics chromatin* (2012).

[80] A H Brand, L Breeden, J Abraham, R Sternglanz, and K Nasmyth. "Characterization of a "silencer" in yeast: a DNA sequence with properties opposite to those of a transcriptional enhancer." In: *Cell* (1985).

[81] S Lee and D S Gross. "Conditional silencing: the HMRE mating-type silencer exerts a rapidly reversible position effect on the yeast HSP82 heat shock gene." In: *Molecular and cellular biology* (1993).

[82] L Laimins, M Holmgren-König, and G Khoury. "Transcriptional "silencer" element in rat repetitive sequences associated with the rat insulin 1 gene locus." In: *Proceedings of the National Academy of Sciences of the United States of America* (1986).

[83] Baoxu Pang and Michael P Snyder. "Systematic identification of silencers in human cells." In: *Nature genetics* (2020).

[84] Di Huang, Hanna M Petrykowska, Brendan F Miller, Laura Elnitski, and Ivan Ovcharenko. "Identification of human silencers by correlating cross-tissue epigenetic profiles and gene expression." In: *Genome research* (2019).

[85] Stephen S Gisselbrecht, Alexandre Palagi, Jesse V Kurland, Julia M Rogers, Hakan Ozadam, Ye Zhan, Job Dekker, and Martha L Bulyk. "Transcriptional Silencers in Drosophila Serve a Dual Role as Transcriptional Enhancers in Alternate Cellular Contexts." In: *Molecular cell* (2020).

[86] Tatyana Yu Zykova, Victor G Levitsky, Elena S Belyaeva, and Igor F Zhimulev. "Polytene Chromosomes - A Portrait of Functional Organization of the Drosophila Genome." In: *Current genomics* (2018).

[87] H Ris. "Stereoscopic electron microscopy of chromosomes." In: *Methods in cell biology* (1981).

[88] X Bi and J R Broach. "UASrpg can function as a heterochromatin boundary element in yeast." In: *Genes development* (1999).

[89] T I Gerasimova and V G Corces. "Chromatin insulators and boundaries: effects on transcription and nuclear organization." In: *Annual review of genetics* (2001).

[90] Mariano Labrador and Victor G Corces. "Setting the boundaries of chromatin domains and nuclear organization." In: *Cell* (2002).

[91] Keith Byrd and Victor G Corces. "Visualization of chromatin domains created by the gypsy insulator of Drosophila." In: *The Journal of cell biology* (2003).

[92] Swarnava Roy, Yian Yee Tan, and Craig M Hart. "A genetic screen supports a broad role for the Drosophila insulator proteins BEAF-32A and BEAF-32B in maintaining patterns of gene expression." In: *Molecular genetics and genomics : MGG* (2007).

[93] Alexander M Boutanaev, Alla I Kalmykova, Yuri Y Shevelyov, and Dmitry I Nurminsky. "Large clusters of co-expressed genes in the Drosophila genome." In: *Nature* (2002).

[94] Martin J Lercher, Araxi O Urrutia, and Laurence D Hurst. "Clustering of housekeeping genes provides a unified model of gene order in the human genome." In: *Nature genetics* (2002).

[95] D Donze, C R Adams, J Rine, and R T Kamakaka. "The boundaries of the silenced HMR domain in Saccharomyces cerevisiae." In: *Genes development* (1999).

[96] G Fourel, E Revardel, C E Koering, and E Gilson. "Cohabitation of insulators and silencing elements in yeast subtelomeric regions." In: *The EMBO journal* (1999).

[97] Kenzo Fukunaga, Yukinori Hirano, and Katsunori Sugimoto. "Subtelomere-binding protein Tbf1 and telomere-binding protein Rap1 collaborate to inhibit localization of the Mre11 complex to DNA ends in budding yeast." In: *Molecular biology of the cell* (2012).

[98] M N Prioleau, P Nony, M Simpson, and G Felsenfeld. "An insulator element and condensed chromatin region separate the chicken beta-globin locus from an independently regulated erythroid-specific folate receptor gene." In: *The EMBO journal* (1999).

[99] A Udvardy, E Maine, and P Schedl. "The 87A7 chromomere. Identification of novel chromatin structures flanking the heat shock locus that may define the boundaries of higher order domains." In: *Journal of molecular biology* (1985).

[100] J van der Vlag, J L den Blaauwen, R G Sewalt, R van Driel, and A P Otte. "Transcriptional repression mediated by polycomb group proteins and other chromatin-associated repressors is selectively blocked by insulators." In: *The Journal of biological chemistry* (2000).

[101] H Cai and M Levine. "Modulation of enhancer-promoter interactions by insulators in the Drosophila embryo." In: *Nature* (1995).

[102] J R Morris, J L Chen, P K Geyer, and C T Wu. "Two modes of transvection: enhancer action in trans and bypass of a chromatin insulator in cis." In: *Proceedings of the National Academy of Sciences of the United States of America* (1998).

[103] H N Cai and P Shen. "Effects of cis arrangement of chromatin insulators on enhancer-blocking activity." In: *Science (New York, N.Y.)* (2001).

[104] D S Latchman. "Transcription factors: an overview." In: *International journal of experimental pathology* (1993).

[105] K Struhl. "Helix-turn-helix, zinc-finger, and leucine-zipper motifs for eukaryotic transcriptional regulatory proteins." In: *Trends in biochemical sciences* (1989).

[106] James P Noonan and Andrew S McCallion. "Genomics of long-range regulatory elements." In: *Annual review of genomics and human genetics* (2010).

[107] S G Roberts. "Mechanisms of action of transcription activation and repression domains." In: *Cellular and molecular life sciences : CMLS* (2000).

[108] Mary C Thomas and Cheng-Ming Chiang. "The general transcription machinery and general cofactors." In: *Critical reviews in biochemistry and molecular biology* (2006).

[109] Joseph C Reese. "Basal transcription factors." In: *Current opinion in genetics development* (2003).

[110] S Buratowski, S Hahn, L Guarente, and P A Sharp. "Five intermediate complexes in transcription initiation by RNA polymerase II." In: *Cell* (1989).

[111] F C Holstege, P C van der Vliet, and H T Timmers. "Opening of an RNA polymerase II promoter occurs in two distinct steps and requires the basal transcription factors IIE and IIH." In: *The EMBO journal* (1996).

[112] Julie Soutourina. "Transcription regulation by the Mediator complex." In: *Nature reviews. Molecular cell biology* (2018).

[113] Edgar Wingender, Torsten Schoeps, Martin Haubrock, and Jürgen Dönitz. "TFClass: a classification of human transcription factors and their rodent orthologs." In: *Nucleic acids research* (2015).

[114] Karolin Luger, Mekonnen L Dechassa, and David J Tremethick. "New insights into nucleosome and chromatin structure: an ordered state or a disordered affair?" In: *Nature reviews. Molecular cell biology* (2012).

[115] Tilman Borggrefe and Xiaojing Yue. "Interactions between subunits of the Mediator complex with gene-specific transcription factors." In: *Seminars in cell developmental biology* (2011).

[116] Hidehisa Takahashi, Tari J Parmely, Shigeo Sato, Chieri Tomomori-Sato, Charles A S Banks, Stephanie E Kong, Henrietta Szutorisz, Selene K Swanson, Skylar Martin-Brown, Michael P Washburn, Laurence Florens, Chris W Seidel, Chengqi Lin, Edwin R Smith, Ali Shilatifard, Ronald C Conaway, and Joan W Conaway. "Human mediator subunit MED26 functions as a docking site for transcription elongation factors." In: *Cell* (2011).

[117] Célia Jeronimo and François Robert. "Kin28 regulates the transient association of Mediator with core promoters." In: *Nature structural molecular biology* (2014).

[118] F C Holstege, E G Jennings, J J Wyrick, T I Lee, C J Hengartner, M R Green, T R Golub, E S Lander, and R A Young. "Dissecting the regulatory circuitry of a eukaryotic genome." In: *Cell* (1998).

[119] Mitsuhiro Ito, Hirotaka J Okano, Robert B Darnell, and Robert G Roeder. "The TRAP100 component of the TRAP/Mediator complex is essential in broad transcriptional events and development." In: *The EMBO journal* (2002).

[120] J Zlatanova, S H Leuba, and K van Holde. "Chromatin fiber structure: morphology, molecular determinants, structural transitions." In: *Biophysical journal* (1998).

[121] R D Kornberg. "Chromatin structure: a repeating unit of histones and DNA." In: *Science (New York, N.Y.)* (1974).

[122] K Luger, A W Mäder, R K Richmond, D F Sargent, and T J Richmond. "Crystal structure of the nucleosome core particle at 2.8 A resolution." In: *Nature* (1997).

[123] Timothy J Richmond and Curt A Davey. "The structure of DNA in the nucleosome core." In: *Nature* (2003).

[124] A L Olins and D E Olins. "Spheroid chromatin units (v bodies)." In: *Science (New York, N.Y.)* (1974).

[125] Maria Aurelia Ricci, Carlo Manzo, María Filomena García-Parajo, Melike Lakadamyali, and Maria Pia Cosma. "Chromatin fibers are formed by heterogeneous groups of nucleosomes in vivo." In: *Cell* (2015).

[126] Huan Tao, Hao Li, Kang Xu, Hao Hong, Shuai Jiang, Guifang Du, Junting Wang, Yu Sun, Xin Huang, Yang Ding, Fei Li, Xiaofei Zheng, Hebing Chen, and Xiaochen Bo. "Computational methods for the prediction of chromatin interaction and organization using sequence and epigenomic profiles." In: *Briefings in bioinformatics* (2021).

[127] L A Cirillo and K S Zaret. "An early developmental transcription factor complex that is more stable on nucleosome core particles than on free DNA." In: *Molecular cell* (1999).

[128] Yujia Sun, Chung-Yi Nien, Kai Chen, Hsiao-Yun Liu, Jeff Johnston, Julia Zeitlinger, and Christine Rushlow. "Zelda overcomes the high intrinsic nucleosome barrier at enhancers during Drosophila zygotic genome activation." In: *Genome research* (2015).

[129] Geeta J Narlikar, Ramasubramanian Sundaramoorthy, and Tom Owen-Hughes. "Mechanisms and functions of ATP-dependent chromatin-remodeling enzymes." In: *Cell* (2013).

[130] Lisette Mohrmann and C Peter Verrijzer. "Composition and functional specificity of SWI2/SNF2 class chromatin remodeling complexes." In: *Biochimica et biophysica acta* (2005).

[131] Davide F V Corona and John W Tamkun. "Multiple roles for ISWI in transcription, chromosome organization and DNA replication." In: *Biochimica et biophysica acta* (2004).

[132] S A Denslow and P A Wade. "The human Mi-2/NuRD complex and gene regulation." In: *Oncogene* (2007).

[133] Toyoko Tsukuda, Alastair B Fleming, Jac A Nickoloff, and Mary Ann Osley. "Chromatin remodelling at a DNA double-strand break site in Saccharomyces cerevisiae." In: *Nature* (2005).

[134] Andrew T Fenley, David A Adams, and Alexey V Onufriev. "Charge state of the globular histone core controls stability of the nucleosome." In: *Biophysical journal* (2010).

[135] Tony Kouzarides. "Chromatin modifications and their function." In: *Cell* (2007).

[136] Swaminathan Venkatesh and Jerry L Workman. "Histone exchange, chromatin structure and the regulation of transcription." In: *Nature reviews. Molecular cell biology* (2015).

[137] T Cremer, C Cremer, T Schneider, H Baumann, L Hens, and M Kirsch-Volders. "Analysis of chromosome positions in the interphase nucleus of Chinese hamster cells by laser-UV-microirradiation experiments." In: *Human genetics* (1982).

[138] Luis A Parada, Philip G McQueen, and Tom Misteli. "Tissue-specific spatial organization of genomes." In: *Genome biology* (2004).

[139] M Cremer, J von Hase, T Volm, A Brero, G Kreth, J Walter, C Fischer, I Solovei, C Cremer, and T Cremer. "Non-random radial higher-order chromatin arrangements in nuclei of diploid human cells." In: *Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology* (2001).

[140] Muyu Yang and Jian Ma. "Machine Learning Methods for Exploring Sequence Determinants of 3D Genome Organization." In: *Journal of molecular biology* (2022).

[141] Irina Solovei, Moritz Kreysing, Christian Lanctôt, Süleyman Kösem, Leo Peichl, Thomas Cremer, Jochen Guck, and Boris Joffe. "Nuclear architecture of rod photoreceptor cells adapts to vision in mammalian evolution." In: *Cell* (2009).

[142] Asifa Akhtar and Susan M Gasser. "The nuclear envelope and transcriptional control." In: *Nature reviews. Genetics* (2007).

[143] Bas van Steensel and Andrew S Belmont. "Lamina-Associated Domains: Links with Chromosome Architecture, Heterochromatin, and Gene Repression." In: *Cell* (2017).

[144] Raphaël Mourad and Olivier Cuvier. "Predicting the spatial organization of chromosomes using epigenetic data." In: *Genome biology* (2015).

[145] Daniel Jost, Pascal Carrivain, Giacomo Cavalli, and Cédric Vaillant. "Modeling epigenome folding: formation and dynamics of topologically associated chromatin domains." In: *Nucleic acids research* (2014).

[146] Paul B Talbert and Steven Henikoff. "Histone variants on the move: substrates for chromatin dynamics." In: *Nature reviews. Molecular cell biology* (2017).

[147] Diego I Cattoni, Andrés M Cardozo Gizzi, Mariya Georgieva, Marco Di Stefano, Alessandro Valeri, Delphine Chamousset, Christophe Houbron, Stephanie Déjardin, Jean-Bernard Fiche, Inma González, Jia-Ming Chang, Thomas Sexton, Marc A Marti-Renom, Frédéric Bantignies, Giacomo Cavalli, and Marcelo Nollmann. "Single-cell absolute contact probability detection reveals chromosomes are organized by multiple low-frequency yet specific interactions." In: *Nature communications* (2017).

[148] Suhas S P Rao, Miriam H Huntley, Neva C Durand, Elena K Stamenova, Ivan D Bochkov, James T Robinson, Adrian L Sanborn, Ido Machol, Arina D Omer, Eric S Lander, and Erez Lieberman Aiden. "A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping." In: *Cell* (2014).

[149] Jesse R Dixon, Inkyung Jung, Siddarth Selvaraj, Yin Shen, Jessica E Antosiewicz-Bourget, Ah Young Lee, Zhen Ye, Audrey Kim, Nisha Rajagopal, Wei Xie, Yarui Diao, Jing Liang, Huimin Zhao, Victor V Lobanenkov, Joseph R Ecker, James A Thomson, and Bing Ren. "Chromatin architecture reorganization during stem cell differentiation." In: *Nature* (2015).

[150] Lee E Finlan, Duncan Sproul, Inga Thomson, Shelagh Boyle, Elizabeth Kerr, Paul Perry, Bauke Ylstra, Jonathan R Chubb, and Wendy A Bickmore. "Recruitment to the nuclear periphery can alter expression of genes in human cells." In: *PLoS genetics* (2008).

[151] K L Reddy, J M Zullo, E Bertolino, and H Singh. "Transcriptional repression mediated by repositioning of genes to the nuclear lamina." In: *Nature* (2008).

[152] Michael I Robson, Jose I de Las Heras, Rafal Czapiewski, Phú Lê Thành, Daniel G Booth, David A Kelly, Shaun Webb, Alastair R W Kerr, and Eric C Schirmer. "Tissue-Specific Gene Repositioning by Muscle Nuclear Membrane Proteins Enhances Repression of Critical Developmental Genes during Myogenesis." In: *Molecular cell* (2016).

[153] T Grigliatti. "Position-effect variegation–an assay for nonhistone chromosomal proteins and chromatin assembly and modifying factors." In: *Methods in cell biology* (1991).

[154] S Henikoff and J G Henikoff. "Position-based sequence weights." In: *Journal of molecular biology* (1994).

[155] Lucas J T Kaaij, Robin H van der Weide, René F Ketting, and Elzo de Wit. "Systemic Loss and Gain of Chromatin Architecture throughout Zebrafish Development." In: *Cell reports* (2018).

[156] Pengfei Dong, Xiaoyu Tu, Po-Yu Chu, Peitao Lü, Ning Zhu, Donald Grierson, Baijuan Du, Pinghua Li, and Silin Zhong. "3D Chromatin Architecture of Large Plant Genomes Determined by Local A/B Compartments." In: *Molecular plant* (2017).

[157] Tung B K Le, Maxim V Imakaev, Leonid A Mirny, and Michael T Laub. "High-resolution mapping of the spatial organization of a bacterial chromosome." In: *Science (New York, N.Y.)* (2013).

[158] Martial Marbouty, Antoine Le Gall, Diego I Cattoni, Axel Cournac, Alan Koh, Jean-Bernard Fiche, Julien Mozziconacci, Heath Murray, Romain Koszul, and Marcelo Nollmann. "Condensin- and Replication-Mediated Bacterial Chromosome Folding and Origin Condensation Revealed by Hi-C and Super-resolution Imaging." In: *Molecular cell* (2015).

[159] Judith H I Haarhuis, Robin H van der Weide, Vincent A Blomen, J Omar Yáñez-Cuna, Mario Amendola, Marjon S van Ruiten, Peter H L Krijger, Hans Teunissen, René H Medema, Bas van Steensel, Thijn R Brummelkamp, Elzo de Wit, and Benjamin D Rowland. "The Cohesin Release Factor WAPL Restricts Chromatin Loop Extension." In: *Cell* (2017).

[160] Gordana Wutz, Csilla Várnai, Kota Nagasaka, David A Cisneros, Roman R Stocsits, Wen Tang, Stefan Schoenfelder, Gregor Jessberger, Matthias Muhar, M Julius Hossain, Nike Walther, Birgit Koch, Moritz Kueblbeck, Jan Ellenberg, Johannes Zuber, Peter Fraser, and Jan-Michael Peters. "Topologically associating domains and chromatin loops depend on cohesin and are regulated by CTCF, WAPL, and PDS5 proteins." In: *The EMBO journal* (2017).

[161] Miao Yu and Bing Ren. "The Three-Dimensional Organization of Mammalian Genomes." In: *Annual review of cell and developmental biology* (2017).

[162] Vuthy Ea, Tom Sexton, Thierry Gostan, Laurie Herviou, Marie-Odile Baudement, Yunzhe Zhang, Soizik Berlivet, Marie-Noëlle Le Lay-Taha, Guy Cathala, Annick Lesne, Jean-Marc Victor, Yuhong Fan, Giacomo Cavalli, and Thierry Forné. "Distinct polymer physics principles govern chromatin dynamics in mouse and Drosophila topological domains." In: *BMC genomics* (2015).

[163] Darya Filippova, Rob Patro, Geet Duggal, and Carl Kingsford. "Identification of alternative topological domains in chromatin." In: *Algorithms for molecular biology : AMB* (2014).

[164] William A Flavahan, Yotam Drier, Brian B Liau, Shawn M Gillespie, Andrew S Venteicher, Anat O Stemmer-Rachamimov, Mario L Suvà, and Bradley E Bernstein. "Insulator dysfunction and oncogene activation in IDH mutant gliomas." In: *Nature* (2016).

[165] Joachim Weischenfeldt, Taronish Dubash, Alexandros P Drainas, Balca R Mardin, Yuanyuan Chen, Adrian M Stütz, Sebastian M Waszak, Graziella Bosco, Ann Rita Halvorsen, Benjamin Raeder, Theocharis Efthymiopoulos, Serap Erkek, Christine Siegl, Hermann Brenner, Odd Terje Brustugun, Sebastian M Dieter, Paul A Northcott, Iver Petersen, Stefan M Pfister, Martin Schneider, Steinar K Solberg, Erik Thunissen, Wilko Weichert, Thomas Zichner, Roman Thomas, Martin Peifer, Aslaug Helland, Claudia R Ball, Martin Jechlinger, Rocio Sotillo, Hanno Glimm, and Jan O Korbel. "Pan-cancer analysis of somatic copy-number alterations implicates IRS4 and IGF2 in enhancer hijacking." In: *Nature genetics* (2017).

[166] Quentin Szabo, Axelle Donjon, Ivana Jerković, Giorgio L Papadopoulos, Thierry Cheutin, Boyan Bonev, Elphège P Nora, Benoit G Bruneau, Frédéric Bantignies, and Giacomo Cavalli. "Regulation of single-cell genome organization into TADs and chromatin nanodomains." In: *Nature genetics* (2020).

[167] Fabian Grubert, Rohith Srivas, Damek V Spacek, Maya Kasowski, Mariana Ruiz-Velasco, Nasa Sinnott-Armstrong, Peyton Greenside, Anil Narasimha, Qing Liu, Benjamin Geller, Akshay Sanghi, Michael Kulik, Silin Sa, Marlene Rabinovitch, Anshul Kundaje, Stephen Dalton, Judith B Zaugg, and Michael Snyder. "Landscape of cohesin-mediated chromatin loops in the human genome." In: *Nature* (2020).

[168] Tsung-Han S Hsieh, Claudia Cattoglio, Elena Slobodyanyuk, Anders S Hansen, Oliver J Rando, Robert Tjian, and Xavier Darzacq. "Resolving the 3D Landscape of Transcription-Linked Mammalian Chromatin Folding." In: *Molecular cell* (2020).

[169] Robert-Jan Palstra, Bas Tolhuis, Erik Splinter, Rian Nijmeijer, Frank Grosveld, and Wouter de Laat. "The beta-globin nuclear compartment in development and erythroid differentiation." In: *Nature genetics* (2003).

[170] Wulan Deng, Jongjoo Lee, Hongxin Wang, Jeff Miller, Andreas Reik, Philip D Gregory, Ann Dean, and Gerd A Blobel. "Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor." In: *Cell* (2012).

[171] Annabelle Wurmser and Srinjan Basu. "Enhancer-Promoter Communication: It's Not Just About Contact." In: *Frontiers in molecular biosciences* (2022).

[172] Laura Vian, Aleksandra Pękowska, Suhas S P Rao, Kyong-Rim Kieffer-Kwon, Seolkyoung Jung, Laura Baranello, Su-Chen Huang, Laila El Khattabi, Marei Dose, Nathanael Pruett, Adrian L Sanborn, Andres Canela, Yaakov Maman, Anna Oksanen, Wolfgang Resch, Xingwang Li, Byoungkoo Lee, Alexander L Kovalchuk, Zhonghui Tang, Steevenson Nelson, Michele Di Pierro, Ryan R Cheng, Ido Machol, Brian Glenn St Hilaire, Neva C Durand, Muhammad S Shamim, Elena K Stamenova, José N Onuchic, Yijun Ruan, Andre Nussenzweig, David Levens, Erez Lieberman Aiden, and Rafael Casellas. "The Energetics and Physiological Impact of Cohesin Extrusion." In: *Cell* (2018).

[173] Geoffrey Fudenberg, Nezar Abdennur, Maxim Imakaev, Anton Goloborodko, and Leonid A Mirny. "Emerging Evidence of Chromosome Folding by Loop Extrusion." In: *Cold Spring Harbor symposia on quantitative biology* (2017).

[174] Liang-Fu Chen, Hannah Katherine Long, Minhee Park, Tomek Swigut, Alistair Nicol Boettiger, and Joanna Wysocka. "Structural elements promote architectural stripe formation and facilitate ultra-long-range gene regulation at a human disease locus." In: *Molecular cell* (2023).

[175] N. Isiaka Bolaji, Semple Jennifer I, Haemmerli Anja, Thapliyal Saurabh, Stojanovski Klement, Das Moushumi, Gilbert Nick, Glauser Dominique A, Towbin Benjamin, Jost Daniel, and Peter Meister. "Cohesin forms fountains at active enhancers in C. elegans". In: *BioRxiv* (2023).

[176] Boyan Bonev and Giacomo Cavalli. "Organization and function of the 3D genome." In: *Nature reviews. Genetics* (2016).

[177] Elphège P Nora, Anton Goloborodko, Anne-Laure Valton, Johan H Gibcus, Alec Uebersohn, Nezar Abdennur, Job Dekker, Leonid A Mirny, and Benoit G Bruneau. "Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization." In: *Cell* (2017).

[178] Kerstin S Wendt, Keisuke Yoshida, Takehiko Itoh, Masashige Bando, Birgit Koch, Erika Schirghuber, Shuichi Tsutsumi, Genta Nagae, Ko Ishihara, Tsuyoshi Mishiro, Kazuhide Yahata, Fumio Imamoto, Hiroyuki Aburatani, Mitsuyoshi Nakao, Naoko Imamoto, Kazuhiro Maeshima, Katsuhiko Shirahige, and Jan-Michael Peters. "Cohesin mediates transcriptional insulation by CCCTC-binding factor." In: *Nature* (2008).

[179] K Kimura, V V Rybenkov, N J Crisona, T Hirano, and N R Cozzarelli. "13S condensin actively reconfigures DNA by introducing global positive writhe: implications for chromosome condensation." In: *Cell* (1999).

[180] Iain F Davidson and Jan-Michael Peters. "Genome folding through loop extrusion by SMC complexes." In: *Nature reviews. Molecular cell biology* (2021).

[181] Mahipal Ganji, Indra A Shaltiel, Shveta Bisht, Eugene Kim, Ana Kalichava, Christian H Haering, and Cees Dekker. "Real-time imaging of DNA loop extrusion by condensin." In: *Science (New York, N.Y.)* (2018).

[182] Stefan Golfier, Thomas Quail, Hiroshi Kimura, and Jan Brugués. "Cohesin and condensin extrude DNA loops in a cell cycle-dependent manner." In: *eLife* (2020).

[183] Edward J Banigan and Leonid A Mirny. "Loop extrusion: theory meets single-molecule experiments." In: *Current opinion in cell biology* (2020).

[184] Muwen Kong, Erin E Cutts, Dongqing Pan, Fabienne Beuron, Thangavelu Kaliyappan, Chaoyou Xue, Edward P Morris, Andrea Musacchio, Alessandro Vannini, and Eric C Greene. "Human Condensin I and II Drive Extensive ATP-Dependent Compaction of Nucleosome-Bound DNA." In: *Molecular cell* (2020).

[185] Iain F Davidson, Roman Barth, Maciej Zaczek, Jaco van der Torre, Wen Tang, Kota Nagasaka, Richard Janissen, Jacob Kerssemakers, Gordana Wutz, Cees Dekker, and Jan-Michael Peters. "CTCF is a DNA-tension-dependent barrier to cohesin-mediated loop extrusion." In: *Nature* (2023).

[186] Elphège P Nora, Laura Caccianini, Geoffrey Fudenberg, Kevin So, Vasumathi Kameswaran, Abigail Nagle, Alec Uebersohn, Bassam Hajj, Agnès Le Saux, Antoine Coulon, Leonid A Mirny, Katherine S Pollard, Maxime Dahan, and Benoit G Bruneau. "Molecular basis of CTCF binding polarity in genome folding." In: *Nature communications* (2020).

[187] Anders S Hansen, Iryna Pustova, Claudia Cattoglio, Robert Tjian, and Xavier Darzacq. "CTCF and cohesin regulate chromatin loop stability with distinct dynamics." In: *eLife* (2017).

[188] Michele Gabriele, 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. "Dynamics of CTCF- and cohesin-mediated chromatin looping revealed by live-cell imaging." In: *Science (New York, N.Y.)* (2022).

[189] Pia Mach, Pavel I Kos, Yinxiu Zhan, Julie Cramard, Simon Gaudin, Jana Tünnermann, Edoardo Marchi, Jan Eglinger, Jessica Zuin, Mariya Kryzhanovska, Sebastien Smallwood, Laurent Gelman, Gregory Roth, Elphège P Nora, Guido Tiana, and Luca Giorgetti. "Cohesin and CTCF control the dynamics of chromosome folding." In: *Nature genetics* (2022).

[190] Antonio Tedeschi, Gordana Wutz, Sébastien Huet, Markus Jaritz, Annelie Wuensche, Erika Schirghuber, Iain Finley Davidson, Wen Tang, David A Cisneros, Venugopal Bhaskara, Tomoko Nishiyama, Alipasha Vaziri, Anton Wutz, Jan Ellenberg, and Jan-Michael Peters. "Wapl is an essential regulator of chromatin structure and chromosome segregation." In: *Nature* (2013).

[191] Wibke Schwarzer, Nezar Abdennur, Anton Goloborodko, Aleksandra Pekowska, Geoffrey Fudenberg, Yann Loe-Mie, Nuno A Fonseca, Wolfgang Huber, Christian H Haering, Leonid Mirny, and Francois Spitz. "Two independent modes of chromatin organization revealed by cohesin removal." In: *Nature* (2017).

[192] Emily Crane, Qian Bian, Rachel Patton McCord, Bryan R Lajoie, Bayly S Wheeler, Edward J Ralston, Satoru Uzawa, Job Dekker, and Barbara J Meyer. "Condensin-driven remodelling of X chromosome topology during dosage compensation." In: *Nature* (2015).

[193] Peter Heger, Birger Marin, Marek Bartkuhn, Einhard Schierenberg, and Thomas Wiehe. "The chromatin insulator CTCF and the emergence of metazoan diversity." In: *Proceedings of the National Academy of Sciences of the United States of America* (2012).

[194] Tsung-Han S Hsieh, Claudia Cattoglio, Elena Slobodyanyuk, Anders S Hansen, Xavier Darzacq, and Robert Tjian. "Enhancer-promoter interactions and transcription are largely maintained upon acute loss of CTCF, cohesin, WAPL or YY1." In: *Nature genetics* (2022).

[195] Chunhui Hou, Li Li, Zhaohui S Qin, and Victor G Corces. "Gene density, transcription, and insulators contribute to the partition of the Drosophila genome into physical domains." In: *Molecular cell* (2012).

[196] J Christof M Gebhardt, David M Suter, Rahul Roy, Ziqing W Zhao, Alec R Chapman, Srinjan Basu, Tom Maniatis, and X Sunney Xie. "Single-molecule imaging of transcription factor binding to DNA in live mammalian cells." In: *Nature methods* (2013).

[197] Juan José Bonfiglio, Pietro Fontana, Qi Zhang, Thomas Colby, Ian Gibbs-Seymour, Ilian Atanassov, Edward Bartlett, Roko Zaja, Ivan Ahel, and Ivan Matic. "Serine ADP-Ribosylation Depends on HPF1." In: *Molecular cell* (2017).

[198] W Su, S Jackson, R Tjian, and H Echols. "DNA looping between sites for transcriptional activation: self-association of DNA-bound Sp1." In: *Genes development* (1991).

[199] Abraham S Weintraub, Charles H Li, Alicia V Zamudio, Alla A Sigova, Nancy M Hannett, Daniel S Day, Brian J Abraham, Malkiel A Cohen, Behnam Nabet, Dennis L Buckley, Yang Eric Guo, Denes Hnisz, Rudolf Jaenisch, James E Bradner, Nathanael S Gray, and Richard A Young. "YY1 Is a Structural Regulator of Enhancer-Promoter Loops." In: *Cell* (2017).

[200] Paul E Love, Claude Warzecha, and LiQi Li. "Ldb1 complexes: the new master regulators of erythroid gene transcription." In: *Trends in genetics : TIG* (2014).

[201] Ramasamy Shyam, Aljahani Abrar, Karpinska Magdalena A, Ngoc Cao T. B., Cruz J. Neos, and A. Marieke Oudelaar. "The Mediator complex regulates enhancer-promoter interactions". In: *BioRxiv* (2022).

[202] Jiangang Liu, Narayanan B Perumal, Christopher J Oldfield, Eric W Su, Vladimir N Uversky, and A Keith Dunker. "Intrinsic disorder in transcription factors." In: *Biochemistry* (2006).

[203] V N Uversky, J R Gillespie, and A L Fink. "Why are "natively unfolded" proteins unstructured under physiologic conditions?" In: *Proteins* (2000).

[204] Salman F Banani, Hyun O Lee, Anthony A Hyman, and Michael K Rosen. "Biomolecular condensates: organizers of cellular biochemistry." In: *Nature reviews. Molecular cell biology* (2017).

[205] Shasha Chong, Claire Dugast-Darzacq, Zhe Liu, Peng Dong, Gina M Dailey, Claudia Cattoglio, Alec Heckert, Sambashiva Banala, Luke Lavis, Xavier Darzacq, and Robert Tjian. "Imaging dynamic and selective low-complexity domain interactions that control gene transcription." In: *Science (New York, N.Y.)* (2018).

[206] Benjamin R Sabari, Alessandra Dall'Agnese, Ann Boija, Isaac A Klein, Eliot L Coffey, Krishna Shrinivas, Brian J Abraham, Nancy M Hannett, Alicia V Zamudio, John C Manteiga, Charles H Li, Yang E Guo, Daniel S Day, Jurian Schuijers, Eliza Vasile, Sohail Malik, Denes Hnisz, Tong Ihn Lee, Ibrahim I Cisse, Robert G Roeder, Phillip A Sharp, Arup K Chakraborty, and Richard A Young. "Coactivator condensation at super-enhancers links phase separation and gene control." In: *Science (New York, N.Y.)* (2018).

[207] Won-Ki Cho, Jan-Hendrik Spille, Micca Hecht, Choongman Lee, Charles Li, Valentin Grube, and Ibrahim I Cisse. "Mediator and RNA polymerase II clusters associate in transcription-dependent condensates." In: *Science (New York, N.Y.)* (2018).

[208] Amy R Strom, Alexander V Emelyanov, Mustafa Mir, Dmitry V Fyodorov, Xavier Darzacq, and Gary H Karpen. "Phase separation drives heterochromatin domain formation." In: *Nature* (2017).

[209] Sreejith J Nair, Lu Yang, Dario Meluzzi, Soohwan Oh, Feng Yang, Meyer J Friedman, Susan Wang, Tom Suter, Ibraheem Alshareedah, Amir Gamliel, Qi Ma, Jie Zhang, Yiren Hu, Yuliang Tan, Kenneth A Ohgi, Ranveer Singh Jayani, Priya R Banerjee, Aneel K Aggarwal, and Michael G Rosenfeld. "Phase separation of ligand-activated enhancers licenses cooperative chromosomal enhancer assembly." In: *Nature structural molecular biology* (2019).

[210] Seungsoo Kim and Jay Shendure. "Mechanisms of Interplay between Transcription Factors and the 3D Genome." In: *Molecular cell* (2019).

[211] Ilya M Flyamer, Johanna Gassler, Maxim Imakaev, Hugo B Brandão, Sergey V Ulianov, Nezar Abdennur, Sergey V Razin, Leonid A Mirny, and Kikuë Tachibana-Konwalski. "Single-nucleus Hi-C reveals unique chromatin reorganization at oocyte-to-zygote transition." In: *Nature* (2017).

[212] Zhenhai Du, Hui Zheng, Bo Huang, Rui Ma, Jingyi Wu, Xianglin Zhang, Jing He, Yunlong Xiang, Qiujun Wang, Yuanyuan Li, Jing Ma, Xu Zhang, Ke Zhang, Yang Wang, Michael Q Zhang, Juntao Gao, Jesse R Dixon, Xiaowo Wang, Jianyang Zeng, and Wei Xie. "Allelic reprogramming of 3D chromatin architecture during early mammalian development." In: *Nature* (2017).

[213] Yuwen Ke, Yanan Xu, Xuepeng Chen, Songjie Feng, Zhenbo Liu, Yaoyu Sun, Xuelong Yao, Fangzhen Li, Wei Zhu, Lei Gao, Haojie Chen, Zhenhai Du, Wei Xie, Xiaocui Xu, Xingxu Huang, and Jiang Liu. "3D Chromatin Structures of Mature Gametes and Structural Reprogramming during Mammalian Embryogenesis." In: *Cell* (2017).

[214] Yoon Hee Jung, Michael E G Sauria, Xiaowen Lyu, Manjinder S Cheema, Juan Ausio, James Taylor, and Victor G Corces. "Chromatin States in Mouse Sperm Correlate with Embryonic and Adult Regulatory Landscapes." In: *Cell reports* (2017).

[215] Xuepeng Chen, Yuwen Ke, Keliang Wu, Han Zhao, Yaoyu Sun, Lei Gao, Zhenbo Liu, Jingye Zhang, Wenrong Tao, Zhenzhen Hou, Hui Liu, Jiang Liu, and Zi-Jiang Chen. "Key role for CTCF in establishing chromatin structure in human embryos." In: *Nature* (2019).

[216] Shelby A Blythe and Eric F Wieschaus. "Zygotic genome activation triggers the DNA replication checkpoint at the midblastula transition." In: *Cell* (2015).

[217] Shelby A Blythe and Eric F Wieschaus. "Establishment and maintenance of heritable chromatin structure during early Drosophila embryogenesis." In: *eLife* (2016).

[218] D M PRESCOTT and M A BENDER. "Synthesis of RNA and protein during mitosis in mammalian tissue culture cells." In: *Experimental cell research* (1962).

[219] C A Spencer, M J Kruhlak, H L Jenkins, X Sun, and D P Bazett-Jones. "Mitotic transcription repression in vivo in the absence of nucleosomal chromatin condensation." In: *The Journal of cell biology* (2000).

[220] Ana Rita Araujo, Lendert Gelens, Rahuman S M Sheriff, and Silvia D M Santos. "Positive Feedback Keeps Duration of Mitosis Temporally Insulated from Upstream Cell-Cycle Events." In: *Molecular cell* (2016).

[221] Natalia Naumova, Maxim Imakaev, Geoffrey Fudenberg, Ye Zhan, Bryan R Lajoie, Leonid A Mirny, and Job Dekker. "Organization of the mitotic chromosome." In: *Science (New York, N.Y.)* (2013).

[222] Haoyue Zhang, Daniel J Emerson, Thomas G Gilgenast, Katelyn R Titus, Yemin Lan, Peng Huang, Di Zhang, Hongxin Wang, Cheryl A Keller, Belinda Giardine, Ross C Hardison, Jennifer E Phillips-Cremins, and Gerd A Blobel. "Chromatin structure dynamics during the mitosis-to-G1 phase transition." In: *Nature* (2019).

[223] Kristin Abramo, Anne-Laure Valton, Sergey V Venev, Hakan Ozadam, A Nicole Fox, and Job Dekker. "A chromosome folding intermediate at the condensin-to-cohesin transition during telophase." In: *Nature cell biology* (2019).

[224] Jordan Yupeng Xiao, Antonina Hafner, and Alistair N Boettiger. "How subtle changes in 3D structure can create large changes in transcription." In: *eLife* (2021).

[225] Adam J Rubin, Brook C Barajas, Mayra Furlan-Magaril, Vanessa Lopez-Pajares, Maxwell R Mumbach, Imani Howard, Daniel S Kim, Lisa D Boxer, Jonathan Cairns, Mikhail Spivakov, Steven W Wingett, Minyi Shi, Zhixin Zhao, William J Greenleaf, Anshul Kundaje, Michael Snyder, Howard Y Chang, Peter Fraser, and Paul A Khavari. "Lineage-specific dynamic and pre-established enhancer-promoter contacts cooperate in terminal differentiation." In: *Nature genetics* (2017).

[226] Jeffrey M Alexander, Juan Guan, Bingkun Li, Lenka Maliskova, Michael Song, Yin Shen, Bo Huang, Stavros Lomvardas, and Orion D Weiner. "Live-cell imaging reveals enhancer-dependent Sox2 transcription in the absence of enhancer proximity." In: *eLife* (2019).

[227] Nezha S Benabdallah, Iain Williamson, Robert S Illingworth, Lauren Kane, Shelagh Boyle, Dipta Sengupta, Graeme R Grimes, Pierre Therizols, and Wendy A Bickmore. "Decreased Enhancer-Promoter Proximity Accompanying Enhancer Activation." In: *Molecular cell* (2019).

[228] Bo Gu, Tomek Swigut, Andrew Spencley, Matthew R Bauer, Mingyu Chung, Tobias Meyer, and Joanna Wysocka. "Transcription-coupled changes in nuclear mobility of mammalian cis-regulatory elements." In: *Science (New York, N.Y.)* (2018).

[229] Jinglan Liu and Ian D Krantz. "Cohesin and human disease." In: *Annual review of genomics and human genetics* (2008).

[230] Adrian J McNairn and Jennifer L Gerton. "The chromosome glue gets a little stickier." In: *Trends in genetics : TIG* (2008).

[231] Phillippa C Taberlay, Aaron L Statham, Theresa K Kelly, Susan J Clark, and Peter A Jones. "Reconfiguration of nucleosome-depleted regions at distal regulatory elements accompanies DNA methylation of enhancers and insulators in cancer." In: *Genome research* (2014).

[232] Darío G Lupiáñez, Katerina Kraft, Verena Heinrich, Peter Krawitz, Francesco Brancati, Eva Klopocki, Denise Horn, Hülya Kayserili, John M Opitz, Renata Laxova, Fernando Santos-Simarro, Brigitte Gilbert-Dussardier, Lars Wittler, Marina Borschiwer, Stefan A Haas, Marco Osterwalder, Martin Franke, Bernd Timmermann, Jochen Hecht, Malte Spielmann, Axel Visel, and Stefan Mundlos. "Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions." In: *Cell* (2015).

[233] Elisa Giorgio, Daniel Robyr, Malte Spielmann, Enza Ferrero, Eleonora Di Gregorio, Daniele Imperiale, Giovanna Vaula, Georgios Stamoulis, Federico Santoni, Cristiana Atzori, Laura Gasparini, Denise Ferrera, Claudio Canale, Michel Guipponi, Len A Pennacchio, Stylianos E Antonarakis, Alessandro Brussino, and Alfredo Brusco. "A large genomic deletion leads to enhancer adoption by the lamin B1 gene: a second path to autosomal dominant adult-onset demyelinating leukodystrophy (ADLD)." In: *Human molecular genetics* (2015).

[234] Juan J Tena and José M Santos-Pereira. "Topologically Associating Domains and Regulatory Landscapes in Development, Evolution and Disease." In: *Frontiers in cell and developmental biology* (2021).

[235] Sudha Rajderkar, Iros Barozzi, Yiwen Zhu, Rong Hu, Yanxiao Zhang, Bin Li, Ana Alcaina Caro, Yoko Fukuda-Yuzawa, Guy Kelman, Adyam Akeza, Matthew J Blow, Quan Pham, Anne N Harrington, Janeth Godoy, Eman M Meky, Kianna von Maydell, Riana D Hunter, Jennifer A Akiyama, Catherine S Novak, Ingrid Plajzer-Frick, Veena Afzal, Stella Tran, Javier Lopez-Rios, Michael E Talkowski, K C Kent Lloyd, Bing Ren, Diane E Dickel, Axel Visel, and Len A Pennacchio. "Topologically associating domain boundaries are required for normal genome function." In: *Communications biology* (2023).

[236] Job Dekker, Karsten Rippe, Martijn Dekker, and Nancy Kleckner. "Capturing chromosome conformation." In: *Science (New York, N.Y.)* (2002).

[237] Bas Tolhuis, Robert Jan Palstra, Erik Splinter, Frank Grosveld, and Wouter de Laat. "Looping and interaction between hypersensitive sites in the active beta-globin locus." In: *Molecular cell* (2002).

[238] Christopher R Vakoc, Danielle L Letting, Nele Gheldof, Tomoyuki Sawado, M A Bender, Mark Groudine, Mitchell J Weiss, Job Dekker, and Gerd A Blobel. "Proximity among distant regulatory elements at the beta-globin locus requires GATA-1 and FOG-1." In: *Molecular cell* (2005).

[239] Charalampos G Spilianakis and Richard A Flavell. "Long-range intrachromosomal interactions in the T helper type 2 cytokine locus." In: *Nature immunology* (2004).

[240] Zhe Liu and William T Garrard. "Long-range interactions between three transcriptional enhancers, active Vkappa gene promoters, and a 3' boundary sequence spanning 46 kilobases." In: *Molecular and cellular biology* (2005).

[241] Adele Murrell, Sarah Heeson, and Wolf Reik. "Interaction between differentially methylated regions partitions the imprinted genes Igf2 and H19 into parent-specific chromatin loops." In: *Nature genetics* (2004).

[242] F Grosveld, G B van Assendelft, D R Greaves, and G Kollias. "Position-independent, high-level expression of the human beta-globin gene in transgenic mice." In: *Cell* (1987).

[243] Keji Zhao Shuai Liu. "The Toolbox for Untangling Chromosome Architecture in Immune Cells". In: *Front Immunol* (2021).

[244] Marieke Simonis, Petra Klous, Erik Splinter, Yuri Moshkin, Rob Willemsen, Elzo de Wit, Bas van Steensel, and Wouter de Laat. "Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C)." In: *Nature genetics* (2006).

[245] Harmen J G van de Werken, Gilad Landan, Sjoerd J B Holwerda, Michael Hoichman, Petra Klous, Ran Chachik, Erik Splinter, Christian Valdes-Quezada, Yuva Oz, Britta A M Bouwman, Marjon J A M Verstegen, Elzo de Wit, Amos Tanay, and Wouter de Laat. "Robust 4C-seq data analysis to screen for regulatory DNA interactions." In: *Nature methods* (2012).

[246] Josée Dostie, Todd A Richmond, Ramy A Arnaout, Rebecca R Selzer, William L Lee, Tracey A Honan, Eric D Rubio, Anton Krumm, Justin Lamb, Chad Nusbaum, Roland D Green, and Job Dekker. "Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements." In: *Genome research* (2006).

[247] Mathieu Rousseau, Jennifer L Crutchley, Hisashi Miura, Matthew Suderman, Mathieu Blanchette, and Josée Dostie. "Hox in motion: tracking HoxA cluster conformation during differentiation." In: *Nucleic acids research* (2014).

[248] Kevin C Wang, Yul W Yang, Bo Liu, Amartya Sanyal, Ryan Corces-Zimmerman, Yong Chen, Bryan R Lajoie, Angeline Protacio, Ryan A Flynn, Rajnish A Gupta, Joanna Wysocka, Ming Lei, Job Dekker, Jill A Helms, and Howard Y Chang. "A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression." In: *Nature* (2011).

[249] Bryan R Lajoie, Nynke L van Berkum, Amartya Sanyal, and Job Dekker. "My5C: web tools for chromosome conformation capture studies." In: *Nature methods* (2009).

[250] Qi Wang, Qiu Sun, Daniel M Czajkowsky, and Zhifeng Shao. "Sub-kb Hi-C in D. melanogaster reveals conserved characteristics of TADs between insect and mammalian cells." In: *Nature communications* (2018).

[251] Viraat Y Goel and Anders S Hansen. "The macro and micro of chromosome conformation capture." In: *Wiley interdisciplinary reviews. Developmental biology* (2021).

[252] Vijay Ramani, Darren A Cusanovich, Ronald J Hause, Wenxiu Ma, Ruolan Qiu, Xinxian Deng, C Anthony Blau, Christine M Disteche, William S Noble, Jay Shendure, and Zhijun Duan. "Mapping 3D genome architecture through in situ DNase Hi-C." In: *Nature protocols* (2016).

[253] Tsung-Han S Hsieh, Assaf Weiner, Bryan Lajoie, Job Dekker, Nir Friedman, and Oliver J Rando. "Mapping Nucleosome Resolution Chromosome Folding in Yeast by Micro-C." In: *Cell* (2015).

[254] Beoung Hun Lee, Zexun Wu, and Suhn K Rhie. "Characterizing chromatin interactions of regulatory elements and nucleosome positions, using Hi-C, Micro-C, and promoter capture Micro-C." In: *Epigenetics chromatin* (2022).

[255] Feng Song, Ping Chen, Dapeng Sun, Mingzhu Wang, Liping Dong, Dan Liang, Rui-Ming Xu, Ping Zhu, and Guohong Li. "Cryo-EM study of the chromatin fiber reveals a double helix twisted by tetranucleosomal units." In: *Science (New York, N.Y.)* (2014).

[256] Betul Akgol Oksuz, Liyan Yang, Sameer Abraham, Sergey V Venev, Nils Krietenstein, Krishna Mohan Parsi, Hakan Ozadam, Marlies E Oomen, Ankita Nand, Hui Mao, Ryan M J Genga, Rene Maehr, Oliver J Rando, Leonid A Mirny, Johan H Gibcus, and Job Dekker. "Systematic evaluation of chromosome conformation capture assays." In: *Nature methods* (2021).

[257] Ilias Boltsis, Karol Nowosad, Rutger W W Brouwer, Przemko Tylzanowski, Wilfred F J van IJcken, Danny Huylebroeck, Frank Grosveld, and Petros Kolovos. "Low Input Targeted Chromatin Capture (Low-T2C)." In: *Methods in molecular biology (Clifton, N.J.)* (2021).

[258] Borbala Mifsud, Filipe Tavares-Cadete, Alice N Young, Robert Sugar, Stefan Schoenfelder, Lauren Ferreira, Steven W Wingett, Simon Andrews, William Grey, Philip A Ewels, Bram Herman, Scott Happe, Andy Higgs, Emily LeProust, George A Follows, Peter Fraser, Nicholas M Luscombe, and Cameron S Osborne. "Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C." In: *Nature genetics* (2015).

[259] Viraat Y Goel, Miles K Huseyin, and Anders S Hansen. "Region Capture Micro-C reveals coalescence of enhancers and promoters into nested microcompartments." In: *Nature genetics* (2023).

[260] Peng Hua, Mohsin Badat, Lars L P Hanssen, Lance D Hentges, Nicholas Crump, Damien J Downes, Danuta M Jeziorska, A Marieke Oudelaar, Ron Schwessinger, Stephen Taylor, Thomas A Milne, Jim R Hughes, Doug R Higgs, and James O J Davies. "Defining genome architecture at base-pair resolution." In: *Nature* (2021).

[261] Abrar Aljahani, Peng Hua, Magdalena A Karpinska, Kimberly Quililan, James O J Davies, and A Marieke Oudelaar. "Analysis of sub-kilobase chromatin topology reveals nano-scale regulatory interactions with variable dependence on cohesin and CTCF." In: *Nature communications* (2022).

[262] Guoliang Li, Liuyang Cai, Huidan Chang, Ping Hong, Qiangwei Zhou, Ekaterina V Kulakova, Nikolay A Kolchanov, and Yijun Ruan. "Chromatin Interaction Analysis with Paired-End Tag (ChIA-PET) sequencing technology and application." In: *BMC genomics* (2014).

[263] Maxwell R Mumbach, Adam J Rubin, Ryan A Flynn, Chao Dai, Paul A Khavari, William J Greenleaf, and Howard Y Chang. "HiChIP: efficient and sensitive analysis of protein-directed genome architecture." In: *Nature methods* (2016).

[264] Robert A Beagrie, Antonio Scialdone, Markus Schueler, Dorothee C A Kraemer, Mita Chotalia, Sheila Q Xie, Mariano Barbieri, Inês de Santiago, Liron-Mark Lavitas, Miguel R Branco, James Fraser, Josée Dostie, Laurence Game, Niall Dillon, Paul A W Edwards, Mario Nicodemi, and Ana Pombo. "Complex multi-enhancer contacts captured by genome architecture mapping." In: *Nature* (2017).

[265] A Marieke Oudelaar, James O J Davies, Lars L P Hanssen, Jelena M Telenius, Ron Schwessinger, Yu Liu, Jill M Brown, Damien J Downes, Andrea M Chiariello, Simona Bianco, Mario Nicodemi, Veronica J Buckle, Job Dekker, Douglas R Higgs, and Jim R Hughes. "Single-allele chromatin interactions identify regulatory hubs in dynamic compartmentalized domains." In: *Nature genetics* (2018).

[266] M L Pardue and J G Gall. "Molecular hybridization of radioactive DNA to the DNA of cytological preparations." In: *Proceedings of the National Academy of Sciences of the United States of America* (1969).

[267] G T Rudkin and B D Stollar. "High resolution detection of DNA-RNA hybrids in situ by indirect immunofluorescence." In: *Nature* (1977).

[268] J G Bauman, J Wiegant, P Borst, and P van Duijn. "A new method for fluorescence microscopical localization of specific DNA sequences by in situ hybridization of fluorochromelabelled RNA." In: *Experimental cell research* (1980).

[269] P R Langer-Safer, M Levine, and D C Ward. "Immunological method for mapping genes on Drosophila polytene chromosomes." In: *Proceedings of the National Academy of Sciences of the United States of America* (1982).

[270] J Roohi, M Cammer, C Montagna, and E Hatchwell. "An improved method for generating BAC DNA suitable for FISH." In: *Cytogenetic and genome research* (2008).

[271] Brian J Beliveau, Eric F Joyce, Nicholas Apostolopoulos, Feyza Yilmaz, Chamith Y Fonseka, Ruth B McCole, Yiming Chang, Jin Billy Li, Tharanga Niroshini Senaratne, Benjamin R Williams, Jean-Marie Rouillard, and Chaoting Wu. "Versatile design and synthesis platform for visualizing genomes with Oligopaint FISH probes." In: *Proceedings of the National Academy of Sciences of the United States of America* (2012).

[272] Magda Bienko, Nicola Crosetto, Leonid Teytelman, Sandy Klemm, Shalev Itzkovitz, and Alexander van Oudenaarden. "A versatile genome-scale PCR-based pipeline for high-definition DNA FISH." In: *Nature methods* (2013).

[273] Leah F Rosin, Son C Nguyen, and Eric F Joyce. "Condensin II drives large-scale folding and spatial partitioning of interphase chromosomes in Drosophila nuclei." In: *PLoS genetics* (2018).

[274] Alistair N Boettiger, Bogdan Bintu, Jeffrey R Moffitt, Siyuan Wang, Brian J Beliveau, Geoffrey Fudenberg, Maxim Imakaev, Leonid A Mirny, Chaoting Wu, and Xiaowei Zhuang. "Super-resolution imaging reveals distinct chromatin folding for different epigenetic states." In: *Nature* (2016).

[275] Quentin Szabo, Daniel Jost, Jia-Ming Chang, Diego I Cattoni, Giorgio L Papadopoulos, Boyan Bonev, Tom Sexton, Julian Gurgo, Caroline Jacquier, Marcelo Nollmann, Frédéric Bantignies, and Giacomo Cavalli. "TADs are 3D structural units of higher-order chromosome organization in Drosophila." In: *Science advances* (2018).

[276] Kilian Stoecker, Christiane Dorninger, Holger Daims, and Michael Wagner. "Double labeling of oligonucleotide probes for fluorescence in situ hybridization (DOPE-FISH) improves signal intensity and increases rRNA accessibility." In: *Applied and environmental microbiology* (2010).

[277] Mario P Schimak, Manuel Kleiner, Silke Wetzel, Manuel Liebeke, Nicole Dubilier, and Bernhard M Fuchs. "MiL-FISH: Multilabeled Oligonucleotides for Fluorescence In Situ Hybridization Improve Visualization of Bacterial Cells." In: *Applied and environmental microbiology* (2016).

[278] Michael Lukumbuzya, Markus Schmid, Petra Pjevac, and Holger Daims. "A Multicolor Fluorescence in situ Hybridization Approach Using an Extended Set of Fluorophores to Visualize Microorganisms." In: *Frontiers in microbiology* (2019).

[279] Andrés M Cardozo Gizzi, Sergio M Espinola, Julian Gurgo, Christophe Houbron, Jean-Bernard Fiche, Diego I Cattoni, and Marcelo Nollmann. "Direct and simultaneous observation of transcription and chromosome architecture in single cells with Hi-M." In: *Nature protocols* (2020).

[280] Jun-Han Su, Pu Zheng, Seon S Kinrot, Bogdan Bintu, and Xiaowei Zhuang. "Genome-Scale Imaging of the 3D Organization and Transcriptional Activity of Chromatin." In: *Cell* (2020).

[281] Yodai Takei, Jina Yun, Shiwei Zheng, Noah Ollikainen, Nico Pierson, Jonathan White, Sheel Shah, Julian Thomassie, Shengbao Suo, Chee-Huat Linus Eng, Mitchell Guttman, Guo-Cheng Yuan, and Long Cai. "Integrated spatial genomics reveals global architecture of single nuclei." In: *Nature* (2021).

[282] Markus Götz, Olivier Messina, Sergio Espinola, Jean-Bernard Fiche, and Marcelo Nollmann. "Multiple parameters shape the 3D chromatin structure of single nuclei at the doc locus in Drosophila." In: *Nature communications* (2022).

[283] Yad Ghavi-Helm, Felix A Klein, Tibor Pakozdi, Lucia Ciglar, Daan Noordermeer, Wolfgang Huber, and Eileen E M Furlong. "Enhancer loops appear stable during development and are associated with paused polymerase." In: *Nature* (2014).

[284] Miao Liu, Yanfang Lu, Bing Yang, Yanbo Chen, Jonathan S D Radda, Mengwei Hu, Samuel G Katz, and Siyuan Wang. "Multiplexed imaging of nucleome architectures in single cells of mammalian tissue." In: *Nature communications* (2020).

[285] Yodai Takei, Shiwei Zheng, Jina Yun, Sheel Shah, Nico Pierson, Jonathan White, Simone Schindler, Carsten H Tischbirek, Guo-Cheng Yuan, and Long Cai. "Single-cell nuclear architecture across cell types in the mouse brain." In: *Science (New York, N.Y.)* (2021).

[286] Surya K Ghosh and Daniel Jost. "Genome organization via loop extrusion, insights from polymer physics models." In: *Briefings in functional genomics* (2020).

[287] Kim Nasmyth and Christian H Haering. "Cohesin: its roles and mechanisms." In: *Annual review of genetics* (2009).

[288] Jennifer E Phillips and Victor G Corces. "CTCF: master weaver of the genome." In: *Cell* (2009).

[289] Jop Kind and Bas van Steensel. "Genome-nuclear lamina interactions and gene regulation." In: *Current opinion in cell biology* (2010).

[290] Nicole J Francis, Robert E Kingston, and Christopher L Woodcock. "Chromatin compaction by a polycomb group protein complex." In: *Science (New York, N.Y.)* (2004).

[291] Daniele Canzio, Maofu Liao, Nariman Naber, Edward Pate, Adam Larson, Shenping Wu, Diana B Marina, Jennifer F Garcia, Hiten D Madhani, Roger Cooke, Peter Schuck, Yifan Cheng, and Geeta J Narlikar. "A conformational switch in HP1 releases auto-inhibition to drive heterochromatin assembly." In: *Nature* (2013).

[292] Rui Zhou and Yi Qin Gao. "Polymer models for the mechanisms of chromatin 3D folding: review and perspective." In: *Physical chemistry chemical physics : PCCP* (2020).

[293] Surya K Ghosh and Daniel Jost. "How epigenome drives chromatin folding and dynamics, insights from efficient coarse-grained models of chromosomes." In: *PLoS computational biology* (2018).

[294] Martin Falk, Yana Feodorova, Natalia Naumova, Maxim Imakaev, Bryan R Lajoie, Heinrich Leonhardt, Boris Joffe, Job Dekker, Geoffrey Fudenberg, Irina Solovei, and Leonid A Mirny. "Heterochromatin drives compartmentalization of inverted and conventional nuclei." In: *Nature* (2019).

[295] Michael Chiang, Davide Michieletto, Chris A Brackley, Nattaphong Rattanavirotkul, Hisham Mohammed, Davide Marenduzzo, and Tamir Chandra. "Polymer Modeling Predicts Chromosome Reorganization in Senescence." In: *Cell reports* (2019).

[296] Carlo Annunziatella, Andrea M Chiariello, Andrea Esposito, Simona Bianco, Luca Fiorillo, and Mario Nicodemi. "Molecular Dynamics simulations of the Strings and Binders Switch model of chromatin." In: *Methods (San Diego, Calif.)* (2018).

[297] Mariano Barbieri, Mita Chotalia, James Fraser, Liron-Mark Lavitas, Josée Dostie, Ana Pombo, and Mario Nicodemi. "Complexity of chromatin folding is captured by the strings and binders switch model." In: *Proceedings of the National Academy of Sciences of the United States of America* (2012).

[298] Chris A Brackley, Jill M Brown, Dominic Waithe, Christian Babbs, James Davies, Jim R Hughes, Veronica J Buckle, and Davide Marenduzzo. "Predicting the three-dimensional folding of cis-regulatory regions in mammalian genomes using bioinformatic data and polymer models." In: *Genome biology* (2016).

[299] Mariano Barbieri, Sheila Q Xie, Elena Torlai Triglia, Andrea M Chiariello, Simona Bianco, Inês de Santiago, Miguel R Branco, David Rueda, Mario Nicodemi, and Ana Pombo. "Active and poised promoter states drive folding of the extended HoxB locus in mouse embryonic stem cells." In: *Nature structural molecular biology* (2017).

[300] Simona Bianco, Andrea M Chiariello, Mattia Conte, Andrea Esposito, Luca Fiorillo, Francesco Musella, and Mario Nicodemi. "Computational approaches from polymer physics to investigate chromatin folding." In: *Current opinion in cell biology* (2020).

[301] Joachim Wolff, Vivek Bhardwaj, Stephan Nothjunge, Gautier Richard, Gina Renschler, Ralf Gilsbach, Thomas Manke, Rolf Backofen, Fidel Ramírez, and Björn A Grüning. "Galaxy HiCExplorer: a web server for reproducible Hi-C data analysis, quality control and visualization." In: *Nucleic acids research* (2018).

[302] Abbas Roayaei Ardakany, Halil Tuvan Gezer, Stefano Lonardi, and Ferhat Ay. "Mustache: multi-scale detection of chromatin loops from Hi-C and Micro-C maps using scale-space representation." In: *Genome biology* (2020).

[303] M Jordan Rowley, Axel Poulet, Michael H Nichols, Brianna J Bixler, Adrian L Sanborn, Elizabeth A Brouhard, Karen Hermetz, Hannah Linsenbaum, Gyorgyi Csankovszki, Erez Lieberman Aiden, and Victor G Corces. "Analysis of Hi-C data using SIP effectively identifies loops in organisms from C. elegans to mammals." In: *Genome research* (2020).

[304] Sven Heinz, Christopher Benner, Nathanael Spann, Eric Bertolino, Yin C Lin, Peter Laslo, Jason X Cheng, Cornelis Murre, Harinder Singh, and Christopher K Glass. "Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities." In: *Molecular cell* (2010).

[305] Cyril Matthey-Doret, Lyam Baudry, Axel Breuer, Rémi Montagne, Nadège Guiglielmoni, Vittore Scolari, Etienne Jean, Arnaud Campeas, Philippe Henri Chanut, Edgar Oriol, Adrien Méot, Laurent Politis, Antoine Vigouroux, Pierrick Moreau, Romain Koszul, and Axel Cournac. "Computer vision for pattern detection in chromosome contact maps." In: *Nature communications* (2020).

[306] Ilya M Flyamer, Robert S Illingworth, and Wendy A Bickmore. "Coolpup.py: versatile pile-up analysis of Hi-C data." In: *Bioinformatics (Oxford, England)* (2020).

[307] Elzo de Wit, Britta A M Bouwman, Yun Zhu, Petra Klous, Erik Splinter, Marjon J A M Verstegen, Peter H L Krijger, Nicola Festuccia, Elphège P Nora, Maaike Welling, Edith Heard, Niels Geijsen, Raymond A Poot, Ian Chambers, and Wouter de Laat. "The pluripotent genome in three dimensions is shaped around pluripotency factors." In: *Nature* (2013).

[308] Vera Pancaldi. "Network models of chromatin structure." In: *Current opinion in genetics development* (2023).

[309] Asa Thibodeau, Eladio J Márquez, Dong-Guk Shin, Paola Vera-Licona, and Duygu Ucar. "Chromatin interaction networks revealed unique connectivity patterns of broad H3K4me3 domains and super enhancers in 3D chromatin." In: *Scientific reports* (2017).

[310] Kevin Van Bortle, Michael H Nichols, Li Li, Chin-Tong Ong, Naomi Takenaka, Zhaohui S Qin, and Victor G Corces. "Insulator function and topological domain border strength scale with architectural protein occupancy." In: *Genome biology* (2014).

[311] Fidel Ramírez, Vivek Bhardwaj, Laura Arrigoni, Kin Chung Lam, Björn A Grüning, José Villaveces, Bianca Habermann, Asifa Akhtar, and Thomas Manke. "High-resolution TADs reveal DNA sequences underlying genome organization in flies." In: *Nature communications* (2018).

[312] Philippe J Batut, Xin Yang Bing, Zachary Sisco, João Raimundo, Michal Levo, and Michael S Levine. "Genome organization controls transcriptional dynamics during development." In: *Science (New York, N.Y.)* (2022).

[313] Keerthi T Chathoth, Liudmila A Mikheeva, Gilles Crevel, Jareth C Wolfe, Ioni Hunter, Saskia Beckett-Doyle, Sue Cotterill, Hongsheng Dai, Andrew Harrison, and Nicolae Radu Zabet. "The role of insulators and transcription in 3D chromatin organization of flies." In: *Genome research* (2022).

[314] Kok Hao Chen, Alistair N Boettiger, Jeffrey R Moffitt, Siyuan Wang, and Xiaowei Zhuang. "RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells." In: *Science (New York, N.Y.)* (2015).

[315] Y L Wang and R J Pelham. "Preparation of a flexible, porous polyacrylamide substrate for mechanical studies of cultured cells." In: *Methods in enzymology* (1998).