

Chromatin organization: Meta-analysis for the identification and classification of structural patterns

Silvia Galan Martínez

TESI DOCTORAL UPF / any: 2020

Directors de la tesi:

Marc A. Marti-Renom i François Serra

STRUCTURAL GENOMICS,
GENE REGULATION, STEM CELLS AND CANCER
CENTRE NACIONAL D'ANÀLISI GENÒMICA
CENTRE FOR GENOMIC REGULATION



Universitat
Pompeu Fabra
Barcelona



cnag

centre nacional d'anàlisi genòmica
centro nacional de análisis genómico

Acknowledgements

Text dels agraïments [12 punts]

Abstract

High-throughput Chromosome Conformation Capture (3C) experiments, provide detailed three-dimensional (3D) information about genome organization. Specially, Hi-C, a 3C derivative, has become the standard technique to investigate the 3D chromatin structure, and its functional implication into cell fate determination. However, the correct bioinformatic analysis and interpretation of this data is still an active field of development.

In this thesis, we explore the ability of CTCF to form chromatin loops and their epigenetic signature, by developing metawaffle, an artificial neural network to classify structural patterns without any prior information. This classification, was used to generate a convolutional neural network to *de novo* detect chromatin loops from Hi-C contact matrices, called LOOPbit.

We also present CHESS, a bioinformatic tool for the comparison of chromatin contact maps and differential 3D feature extraction, such as Topologically Associating Domains, stripes or loops.

Resum

L'avenç de mètodes experimentals basats en la captura de la conformació genòmica (3C), estant aportant informació valuosa sobre l'estructura tri-dimensional (3D) del genoma. En especial, Hi-C, un derivat del 3C, s'ha convertit en la tècnica estàndard per estudiar l'estructura 3D de la cromatina, i la seva implicació funcional en la determinació de la identitat cel·lular. De fet, el anàlisi i la interpretació correcta d'aquesta informació és encara un camp de desenvolupament bioinformàtic actiu .

En aquesta tesi, explorem la capacitat del CTCF de formar bucles de cromatina i la seva signatura epigenètica, desenvolupant *metawaffle*, una xarxa neuronal artificial per la classificació de patrons estructurals sense informació prèvia. Aquesta classificació, permet la generació d'una xarxa neuronal convolutiva per la detecció *de novo* de bucles de cromatina en matrius de contacte Hi-C, anomenada *LOOPbit*.

També presentem *CHESS*, una eina bioinformàtica per la comparació de mapes de contactes i l'extracció d'estructures diferencials, tals com dominis (anomenats TADs), ratlles o bucles.

Preface

All living organisms are made of cells, the smallest unit of life. Interestingly, human cells contain a subcellular compartment of few micrometres in size, the nuclei. This compartment contains the genomic information, the DNA, which folding is not arbitrary.

Microscopy and Chromosome Conformation Capture (3C) technologies, helped to unveil the complex hierarchical organization of chromatin. The chromatin fiber has to be sufficiently accessible for DNA-binding proteins, which will be crucial for cell maintenance and fate, such as transcription factors, polymerases and chromatin modifiers, but maximizing its compaction. Recent advances in 3C-based technologies allowed the inspection and acquisition of increasing evidences indicating how the genome architecture influence the regulation of gene transcription. The interplay of genome architecture and function is a paramount to understand multiple biological scenarios, such as cell identity, development and disease.

This thesis consists of multiple chapters. First of all, in the Introduction I review each chromatin organization layer and its impact on the transcription regulation. Moreover, I provide information of the tools available to identify structural patterns, which can be key to regulate cell expression in different scenario.

The results obtained in the two main publications of the candidate, are presented in the core chapters I and II. In chapter I, first I present metawaffle, an algorithm to deconvolve and classify the structural pattern of DNA-binding proteins, specifically CTCF, the master regulator of chromatin loops. Then, I present LOOPbit, a convolutional neural network, trained with CTCF loops, which is able

to retrieve chromatin loops probabilities genome-wide. In chapter II, I introduce CHESS, a new bioinformatics tool to systematically compare and identify differential structural features between contact matrices. Finally, the conclusions for both chapters are added to highlight the main contributions of this thesis.

Objectives

The broad goal of this thesis was to provide an in-depth analysis of how mammalian genomes are organized in 3D. Specifically, we studied the role of the insulator protein CCCTC-Binding Factor (CTCF) in the formation of chromatin loops and its biological implications. To achieve this broad goal, were carried out two specific objectives:

1. Specific Aim 1 includes three tasks: (1) to develop a computational tool for deconvoluting the mean interaction signal between pairs of CTCF in an unsupervised manner. (2) To obtain structural subpopulations and their epigenetic signature. And (3) to develop a chromatin loop detection method.
2. Specific Aim 1 includes two tasks: (1) the design and development of an algorithm for the assessment of structural similarity between genomic regions. And (2) the identification and classification of the genome differential structures.

Table of contents

INTRODUCTION	3
THE NUCLEUS	3
THE GENETIC MATERIAL	4
NUCLEAR ORGANIZATION	7
<i>Chromosome territories</i>	8
<i>Chromosome compartments</i>	10
<i>Subdomains</i>	11
Topological Associated Domains (TADs).....	12
Lamina-Associated Domains (LADs).....	15
Nucleolus-Associated Domains (NADs).....	17
<i>DNA-looping</i>	18
Loop extrusion model	19
Chromatin loop detection methods.....	28
TECHNOLOGY TO ASSESS NUCLEAR ORGANIZATION	32
<i>Microscopy</i>	33
<i>Chromosome conformation capture technology</i>	37
FUNCTION-STRUCTURE DOGMA	41
<i>Differential chromatin interaction identification software tools</i>	43
CHAPTER I.....	46
CHAPTER II.....	47
DISCUSSION.....	48
IDENTIFICATION OF CHROMATIN LOOPS FROM HI-C INTERACTION MATRICES BY CTCF-CTCF TOPOLOGY CLASSIFICATION.....	48
QUANTITATIVE COMPARISON AND AUTOMATIC FEATURE EXTRACTION FOR CHROMATIN CONTACT DATA	54
CONCLUSIONS	59
REFERENCES	61

Table of figures

Figure 1. Illustration of the chromatin organization hierarchy in an interphase nucleus	7
Figure 2. Schematic view of the chromatin organization inside the nucleus.....	9
Figure 3. Subdomains organization	18
Figure 4. Illustration of CTCF regulation.....	21
Figure 5. Loop extrusion illustration and supporting in vitro experiments	24
Figure 6. Summary of chromatin organization effects by cohesin and CTCF mutants.	26
Figure 7. Deletion of TAD borders leading to TAD fusion and enhancer crosstalk	35
Figure 8. Visualization of TAD-like globular domains in single cells	36

Introduction

“Identity is not an object; it is a process with addresses for all the different directions and dimensions in which it moves, and so it cannot so easily be fixed with a single number.”

Lynn Margulis, 1991

The term *cell* (from Latin *cella*, meaning “small room”) was coined by Robert Hooke (1665), which has been described to be the smallest unit of life. In 1839, Theodor Schwann and Matthias Jakob Schleiden proposed the cell theory, which says that all living organisms are made of cells, whose size, number and type, will ultimately define the structure and functions of the organism.

Currently, living organisms on Earth are classified in three large kingdoms: archaea, bacteria and eukaryotes. Bacteria and archaea are also called prokaryotes and are simpler, single-celled organisms. While eukaryotes cells contain compartments, called organelles surrounded by membranes, including the mitochondria, the chloroplasts, the smooth and rough endoplasmic reticula and the nucleus. Lynn Margulis proposed the endosymbiotic theory for the origin of eukaryotic cells (Sagan, 1967), however, it is still unclear and have been proposed more than 20 different versions (Martin, Garg, & Zimorski, 2015). This subcellular organization in eukaryotes gives the opportunity to separate metabolic processes, leaving the cell nucleus a unique structure.

The nucleus

“If you know you are on the right track, if you have this inner knowledge, then nobody can turn you off... no matter what they say...”

Barbara McClintock.

The nucleus represents a milestone in evolution transition. It is enclosed by the nuclear envelope, which separates transcription from translation. This separation still needs the bidirectional trafficking allowed by the nuclear pore complex. The nuclear envelope is also essential for the anchoring of lamins for mechanical support and chromosomal positioning and segregation (Devos, Graf, & Field, 2014). Many nuclear functions such as these complex interactions governing chromosome positioning with respect to the nuclear envelope and their dynamics, are conserved across eukaryotes.

The nucleus was the first organelle discovered by Antoine van Leeuwenhoek (1632-1723), who observed a lumen, the nucleus, in salmon red blood cells. The nucleus is the largest and most easily discernable organelle in eukaryotic cells. The nucleus is a fascinating structure to study, as it regulates in space and time, genomic and metabolic functions such as transcription and genome stability.

The genetic material

"The results suggest a helical structure (which must be very closely packed) containing 2, 3 or 4 co-axial nucleic acid chains per helical unit, and having the phosphate groups near the outside."

Rosalind Franklin, 1951.

It was not until 1869 that Friedrich Miescher identified what he called "nuclein" inside the nuclei of human white blood cells (Dahm, 2005). By the twentieth century, Miescher's term fell into oblivion, and

nowadays it is known as deoxyribonucleic acid or DNA. In 1909, Phoebus Levene discovered the deoxyribose (carbohydrate element of DNA) and in 1929 the ribose (carbohydrate element of ribonucleic acid or RNA). Moreover, in 1919, Levene proposed the ‘polynucleotide model’, which claims that nucleic acids were composed by a series of nucleotides, and each nucleotide was composed of just one of the four nitrogen-containing bases, a glucose molecule and a phosphate group (Levene, 1919). There are two main nitrogen-containing bases classes: purines (adenine (A) and guanine (G); with two fused rings each) and pyrimidines (cytosine (C), thymine (T) and uracil (U); each with a single ring). It is also known that RNA contains A, G, C and U, while DNA contains A, G, C, and T. In 1944, Erwin Chargaff studied the differences on DNA composition between species (Chargaff et al., 1950) and proposed the “Chargaff’s rule”, which claims that the amount of A was similar to T and the amount of G was similar to C. He shared his studies with James Watson and Francis Crick, who were benefited by it and a DNA X-ray image generated by Rosalind Franklin (Franklin & Gosling, 1953), to claim that the DNA is based by two polynucleotide chains twisted around each other to form a double helix (Watson & Crick, 1953).

The chromatin (or similar caption)

Each human diploid cell contains a 2 meters length of DNA fiber, which needs to be efficiently accessible to DNA-binding proteins (DBPs) and at the same time strengthen and compact. This is possible thanks to the wrapping of DNA fiber around octamers of histone proteins. Together this structure, the nucleosome, is composed by two

of each of the four histones (H2A, H2B, H3 and H4), discovered in 1884 by Albrecht Kossel, with \cong 147 bp of DNA wrapped around. In the 1970s was identified the histone H1 to be responsible to link adjacent nucleosomes (with 20-80 bp distance). Altogether, the nucleosomes and the linker DNA connecting them, adopt an open beads-on-a-string-like structure (Van Holde, 1989)(Figure 1). Each histone has highly disordered N-terminal and C-terminal tails that will be susceptible to post-translational modifications (PTMs), which are relevant for the maintenance of the proper cell identity. PTMs are reversible and can regulate the binding of DBPs. There are described a large number of PTMs including methylation, acetylation or phosphorylation. The interaction between DNA and histone tails is proposed to drive the liquid-liquid phase separation, which is also facilitated by the linker histone H1 (Gibson et al., 2019). Using live-cell super-resolution imaging, was observed that chromatin domains behave as “liquid drops”, suggesting that are phase-separated through self-assembly providing plasticity to the chromatin to conduct many functions such as, DNA repair, gene expression and cell-cycle (Nozaki et al., 2017). The macromolecular complex composed by nucleosomes, linker DNA and other DBPs is called chromatin. Under physiological salt conditions this structure is condensed *in vitro* into a fiber of around 30 nm of diameter, referred to as 30 nm chromatin fiber (Horowitz-Scherer & Woodcock, 2006)(Figure 1). However, this structure is not still clear *in vivo*.

The chromatin fiber can be highly packed in a more condensed structure, the chromosomes, observed in 1842 by Karl Wilhelm von Nägeli. It was not until 1902, thanks to the work of Walter Sutton and

Theodor Boveri, when he described the chromosome theory of inheritance, identified chromosomes as the carriers of genetic information. Considering all the mentioned discoveries, chromatin organization was started to be considered of fundamental importance for basic biological processes, such as gene expression.

In order to study chromatin structure is key to obtain the genomic sequence, being the first human genome sequence published almost twenty years ago (Venter et al., 2001). Its discovery contributed into the understanding of the human evolution, disease, and the interplay between environment and heredity.

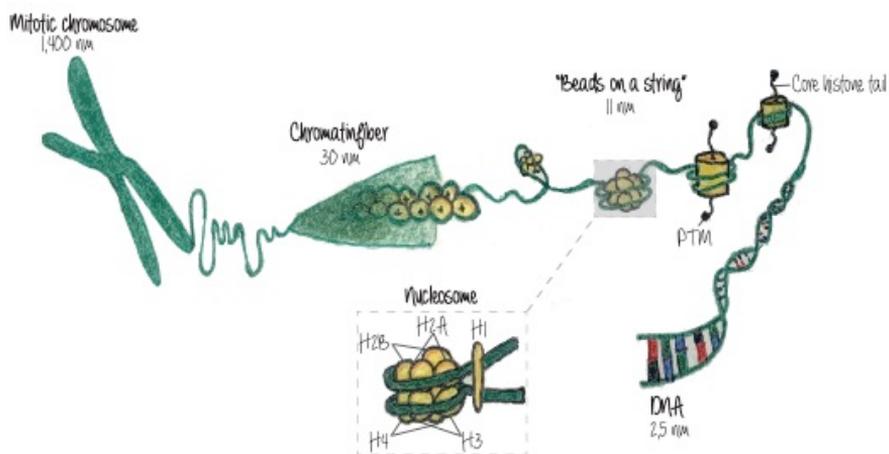


Figure 1. Illustration of the chromatin organization hierarchy in an interphase nucleus. First the double helix DNA is wrapped around the nucleosome, which is shown in detail in the dashed square. The histone tails can present PTMs, enabling chromatin plasticity. The nucleosomes and the linker DNA, form the “beads-on-a-string” like structure. This structure, can be packed to form the chromatin fiber. The chromatin will ultimately be packed conforming the chromosome.

Nuclear organization

The biggest obstacle for women to remain our best in science, I think it is really the combination of career and science. You could say this is not unique to science except that science is really a time eater.

The spatial organization of the human genome in the nucleus is known to play a key role in transcriptional regulation. In order to shed light into the assessment of the three-dimensional (3D) architecture of the nucleus, various methods are used, i.e. microscopy and 3C-based technology. Thanks to them, currently is known that the genome is organized in multiple layers, and each layer has its own regulatory system. Here are discussed the main features in each layer, from larger to finest structures.



Chromosome territories

~100 Mb

In 1885 Carl Rabl suggested a territorial organization of chromosomes, but was in 1909, when Theodor Boveri introduced the term chromosome territories (CTs). He described the non-random distribution of chromosomes during interphase and the maintained structure in the daughter nuclei (Boveri, 1909). From 1970 to 1980 the scientific community believed that chromatin fibers were almost randomly intermingled, picturing the still today very spread image of the spaghetti dish. In 1988 was possible to directly observe CTs in a microscope using chromosome painting (Lichter, Cremer, Borden, Manuelidis, & Ward, 1988). Since then, the details about the organization of chromosomes have been emerging. It has been described that larger chromosomes, that contain higher number of heterochromatic regions, are more present on the nuclear periphery. Whereas smaller chromosomes with higher euchromatic regions are

localized at the center of the nucleus (Tanabe et al., 2002). The boundaries between these CTs have also been precisely described using laser UV microbeam, with the discovery of “intermingling” or “kissing” between chromosomes (Cremer, Cremer, Baumann, et al., 1982). Later, fluorescence in situ hybridization (FISH) techniques indicated a probability of interchromosomal associations to occur (Bolzer et al., 2005). This event changed the way to understand how genes can be coregulated considering the third dimension. For example, the olfactory receptor genes, which are around 1,400 genes located across 18 different chromosomes, are gathered at the time of expression into the same interchromatin space, called “olfatosome” (Horta, Monahan, Bashkirova, & Lomvardas, 2019).

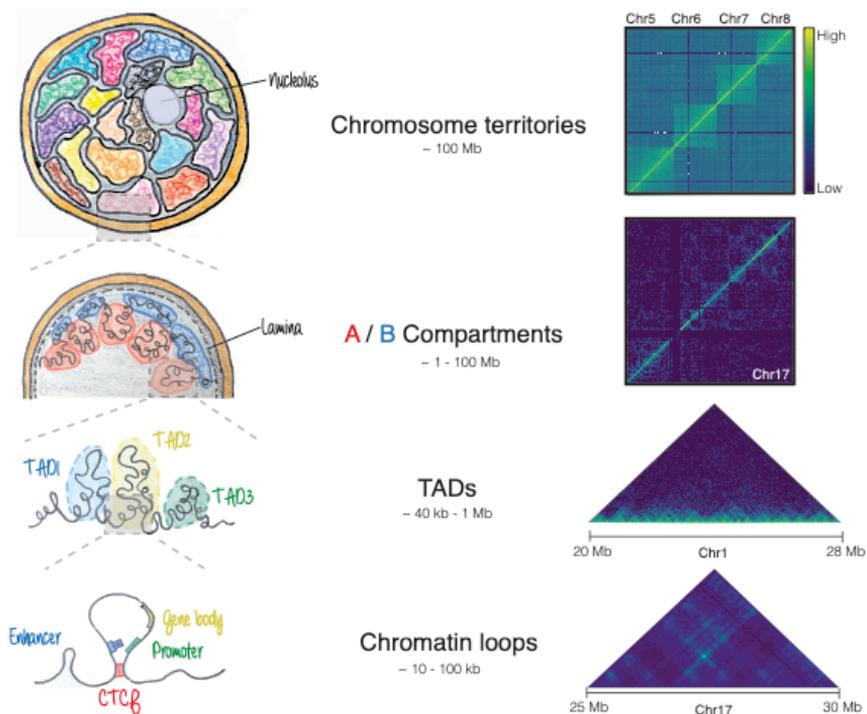


Figure 2. Schematic view of the chromatin organization inside the nucleus. In the left column are illustrated the different chromatin organization layers from chromosome territories to chromatin loops. In the right column a representation of Hi-C maps from GM12878 cells (S. S. Rao et al., 2014) at different genomic scales,

reflect the different layers of chromatin organization. First, in the illustration, chromosomes are represented by different colors, representing the chromosome territories. This organization is also visible in the Hi-C map, showing higher interaction within chromosomes. Then, chromatin compartments are represented in the illustration using red for the A compartment, and blue for the B compartment, this organization is reflected in the Hi-C map by the checkboard pattern segmentation. At 40 kb – 1 Mb are represented the TAD structures, observed in the rotated Hi-C map as triangles, with high interaction frequencies. Finally, are illustrated chromatin loops, which are demarcated by CTCF and help to bring in close proximity genes and their regulatory units. This structure is found in the Hi-C map by a strong dark peak.



Chromosome compartments

~ 1 Mb

It was not until the emergence of a new molecular technique, the high-throughput conformation capture (Hi-C), to be able to discern finer structures inside CTs, which can be as large as several Mb (Lieberman-Aiden et al., 2009). Two major chromatin compartments were elucidated: the A compartment containing more active and open chromatin, and the B compartment containing more inactive and closed chromatin. These two major compartments match with euchromatin and heterochromatin regarding compaction, replication timing, repetitive elements distribution, gene location and expression (Croft et al., 1999). This organization has been maintained in eukaryotes over more than 500 million years, with only few exemptions. Chromosome compartments are cell-type specific, defining the identity of the cell. The 3D organization of the genome is partly driven by the high affinity of active regions for other active regions, which can explain the infrequent euchromatin-heterochromatin interactions. Euchromatin regions are enriched in housekeeping genes, and replicate early in S-phase, while heterochromatin is gene-poor, with tissue-specific genes and with late S-phase replication. Moreover, euchromatin contains most of the short

interspersed repetitive sequences (SINEs), whereas retrotransposon-related long elements (LINEs) and long terminal repeats (LTRs) are located preferentially in heterochromatin. Mitosis disrupts this euchromatin-heterochromatin segregation, and then gradually this separation is restored in two phases during interphase. Which are the mechanisms that maintain this organization is not fully known. Various studies have proposed that sequences exhibit high affinity to each other, causing the separation of the two compartments. This can be caused by the attraction of homotypic chromatin marked by the same repetitive sequences and enforced by the binding of architectural and epigenetic factors (Gibcus & Dekker, 2013). Moreover, the level of segregation has been seen to correlate with the cell differentiation state. For instance, embryonic stem cells (ESCs) are devoid of compact heterochromatin domains, but possess hyperdynamic organization, are transcriptionally promiscuous and have an open chromatin configuration. This flexibility is thought to be required for the maintenance of the pluripotent state and for its progression. With cell differentiation there is a loss in potency and the genome is partitioned into larger euchromatin and heterochromatin domains with replication synchrony. For instance, in human ESCs around the 40 % of the genome switches towards B compartment during cell differentiation (Dixon et al., 2015).



Subdomains

~ 50 kb

The chromatin is organized in a subscale into multiple subdomains, which are important for the proper nuclear homeostasis. According to their nuclear localization, and their contact regions, have been mainly

described the topologically associating domains (TADs), the nucleolus associated domains (NAR) and the lamin associated domains (LADs).

Topologically Associating Domains (TADs)

Hi-C techniques revealed that within compartments, chromatin is organized in sub-domains called TADs. TADs are characterized to interact more within themselves, than between adjacent domains. TADs might serve as a structural platforms for dynamic *cis*-interactions between regulatory elements. In fact, TADs have helped improving our understanding of the relation between enhancers and their target genes. Nowadays, has been observed that actively transcribed genes can form mini-domains that interact more frequently with other active genes. Then, clusters of active genes can form multi-gene domains, with all the belonging genes with a similar transcriptional activity. Within the TAD, chromatin presents a common epigenetic signature and replication timing. Moreover, TADs are highly conserved across cell types, substantiating their importance in cell homeostasis regulation (Dixon et al., 2012; Nora et al., 2012). The high conservation of TAD features in mammals (Dixon et al., 2012; Vietri Rudan et al., 2015), suggests that TADs might represent a basic and ancient structure of the chromatin organization in eukaryotes. The differential TAD sizes between organisms can be explained by the relative sizes of active and inactive chromatin segments, for example, in *Drosophila*, TADs have a mean size of 60 kb, while in mammals is around 800 kb (Dekker & Heard, 2015). As TADs can differ in size, chromatin features and formation mechanisms, they can be classified into different classes or subtypes, each with a specific structural and functional properties. It is relevant to

notice that the identification of TADs is highly dependent on the experiment resolution and the software used to annotate them. Recently, high sequencing depths and resolution have revealed a finer TAD patterning (Forcato et al., 2017; S. S. Rao et al., 2014; Rowley et al., 2017; Zufferey, Tavernari, Oricchio, & Ciriello, 2018). TADs are demarcated by boundaries, which are enriched in multiple genomic features, such as transcription start sites (TSS) and binding sites of CCCTC-binding factor (CTCF). Multiple studies showed the importance of TAD borders to regulate gene expression, for instance their deletion can lead to TAD-fusion events and gene deregulation (Nora et al., 2012). Another studies showed that the disruption or relocation of TAD borders, lead to ectopic contacts between cis-regulatory elements, and finally contributing to developmental disorders or cancer (Flavahan et al., 2016; Franke et al., 2016; Hnisz et al., 2016; Kraft et al., 2019; Lupianez et al., 2015).

Due to the hierarchical fashion of chromatin organization, the compartmentalization of the genome (see Chromosome compartments section) is responsible for both long-range (as for genomic compartments) and local domains (as for TADs). However, their regulatory cross-talk is not fully described. A study of CTCF in loop and TAD formation in mammals, showed that the loss of CTCF using an auxin-mediated system, was translated into a loss of TAD domains, while compartments were maintained (E. P. Nora et al., 2017). Furthermore, the transcriptional activity was slightly affected, suggesting TAD compartmentalization as a fundamental support for the regulation of transcription. In order to shed light into the interplay between transcription and TAD formation, has been recently suggested that can take place even after the inhibition of transcription in

Drosophila embryos. However, it is not totally clear as under the triptolide treatment used, the RNA-Pol II (RNAPII) remains bound to promoter genes (Hug, Grimaldi, Kruse, & Vaquerizas, 2017). Nevertheless, differential gene expression in multiple cell types, can result in the formation of distinct compartmental domains. Supporting the idea that TADs, which are formed by compartments, may be different between cell types with different transcriptional patterns. The regulation of the chromatin organization by gene expression, can be explained in the organisms, like in *Drosophila*, in which CTCF is not found in TADs. However, as mentioned before, in mammals CTCF is an essential architectural protein, and super-resolution microscopy experiments suggested that CTCF together with cohesin are required for TAD boundaries position rather than for TAD formation. FISH labelling, proved the presence of TAD-like structural units in single cells, in wild type and cohesin-depleted conditions. While the position of the TAD boundaries in the wild-type lies more often at CTCF sites, it is random in cohesin-depleted samples (Bintu et al., 2018). Some studies showed that there is ~ 20 % of TAD boundaries independent of CTCF, suggesting their resilience after the loss of CTCF (E. P. Nora et al., 2017). These boundaries might be associated to transcription, as proposed before for *Drosophila*, (Bonev et al., 2017; Dixon et al., 2012) or can correspond to frontier between active and inactive chromatin regions, such as A and B compartment types (S. S. Rao et al., 2014; Rowley et al., 2017). Nevertheless, recent studies showed using CRISPR-dCas9-mediated transcriptional activation its inability to create new TAD borders (Bonev et al., 2017). This result suggests that transcription is not sufficient to demarcate CTCF-independent TAD boundaries. TADs have been observed to appear gradually in early

mouse embryogenesis, and they can still be observed after inhibiting transcription with α -amanitin. However, the inhibition of DNA-replication with aphidicolin blocked the TAD establishment, indicating the potential of replication for the primary establishment of TADs (Du et al., 2017; Ke et al., 2017). Finally, exists a correlation between the conserved TADs and the conservation of the CTCF binding sites and their motif orientation at their boundaries. Interestingly, it has been observed that changes within TADs are correlated with changes in the binding and orientation of CTCF. Thus, TADs are maintained as core structures during evolution, being each a platform imprinted with a specific functional regulatory scenario. Indeed genomic sequences at TAD boundaries are hotspots for genomic rearrangements as they appear to be locally open chromatin with higher frequency of double strand breaks (Guelen et al., 2008).

Lamina-Associated Domains (LADs)

The nuclear lamina is a meshwork of intermediate filament proteins called lamins, which is subjacent to the internal side of the nuclear membrane. These lamina consist of lamins A and C (also referred as lamins A/C or lamin A; both splice variants of the *LMNA* gene) and lamins B1 and B2 (products of *LMNB1* and *LMNB2* genes) (de Leeuw, Gruenbaum, & Medalia, 2018). Specifically, Lamin B-receptor (LBR) and Lamin A/C have been described as major LAD tethers (Solovei et al., 2013). LADs are enriched in LINEs, and in heterochromatin compartment and have a size between 100 kb and 10 Mb (Guelen et al., 2008). Various DNA motifs and proteins has been described to play a role driving LADs to the nuclear periphery. The use of the DNA

adenine methyltransferase identification (DamID) method (van Steensel & Henikoff, 2000) was key to infer the LAD composition and function, disentangling the basic principles of genome organization. Lamin B1 is observed at the nuclear periphery, while lamin A is also present in the nuclear interior (Kind et al., 2013; E. Lund et al., 2013). While peripheral LADs are gene-poor, transcriptionally silent and heterochromatic, intranuclear LADs tend to be more gene-rich, transcriptionally active and euchromatic (Gesson et al., 2016; E. G. Lund, Duband-Goulet, Oldenburg, Buendia, & Collas, 2015). This spatial distribution of lamin A, may explain its impact on the radial positioning of chromatin and its dynamics (Solovei et al., 2013; Vivante, Brozgol, Bronshtein, Levi, & Garini, 2019). However, it is relevant to know that lamin A is not sufficient to anchor heterochromatin into the nuclear periphery, suggesting the need of lamin-associated protein complexes containing integral proteins of the inner nuclear membrane (Buchwalter, Kaneshiro, & Hetzer, 2019). Lamin B1, despite its exclusively peripheric location, also seems to be associated to a relatively complex regulation mechanism as it can be found in euchromatin, presenting a dynamic role in the execution of the Epithelial-Mesenchymal Transition (EMT) transcriptional program (Pascual-Reguant et al., 2018). Recent experiments using FISH, have shown a radial repositioning of TADs dependent on the presence of lamin B1 LADs (Forsberg, Brunet, Ali, & Collas, 2019). The interplay between TADs and LADs to orchestrate spatial genome topology has been studied in a differentiation system showing its fundamental role to shape the 4-dimensional genome during differentiation (Paulsen et al., 2019).

Nucleolus-Associated Domains (NADs)

The largest substructure in the nucleus is the nucleolus, which is the responsible of the ribosome biogenesis. This process is initiated by the transcription of ribosomal RNA (rRNA) genes, the basic component of the nucleolus. The ribosome biogenesis is vital for the assembly of the ribosome to regulate the translational state of the cell. The chromatin regions in contact with the nucleolus are referred as NADs. These subdomains were firstly observed through a very specific sonication of the nuclei that allowed to obtain unbroken nucleoli (Sullivan et al., 2001). Some years later, NADs were mapped in HeLa, IMR90 and HT1080 human cell lines (Nemeth et al., 2010; van Koningsbruggen et al., 2010). NADs are composed by regions with low gene density and transcriptional levels and enriched in repressive histone modifications (H3K27me3, H3K9me3 and H4K20me3). Microscopy studies identified centromeric and pericentromeric satellite repetitive regions and subtelomeric regions as NADs. Moreover, NADs have been observed to cover around 40 % of the genome, and revealed a considerable overlap with LADs (van Koningsbruggen et al., 2010). The unveiling of the role of the nucleolus in genome architecture has been impeded by its membrane-less substructure, which makes difficult the precise mapping of the chromatin domains that contacting it. This factor limits the application of the DamID technology as done for the genome-wide mapping of the LADs. However, it have been hypothesized that alterations at rRNA repeats would alter the nucleolus in its structure and protein composition, promoting the required concentration of rRNA regulatory factors to establish repressive states (Bersaglieri & Santoro, 2019). Finally, both NADs and LADs are

suggested to represent the hub for organizing the inactive or heterochromatin genome regions.

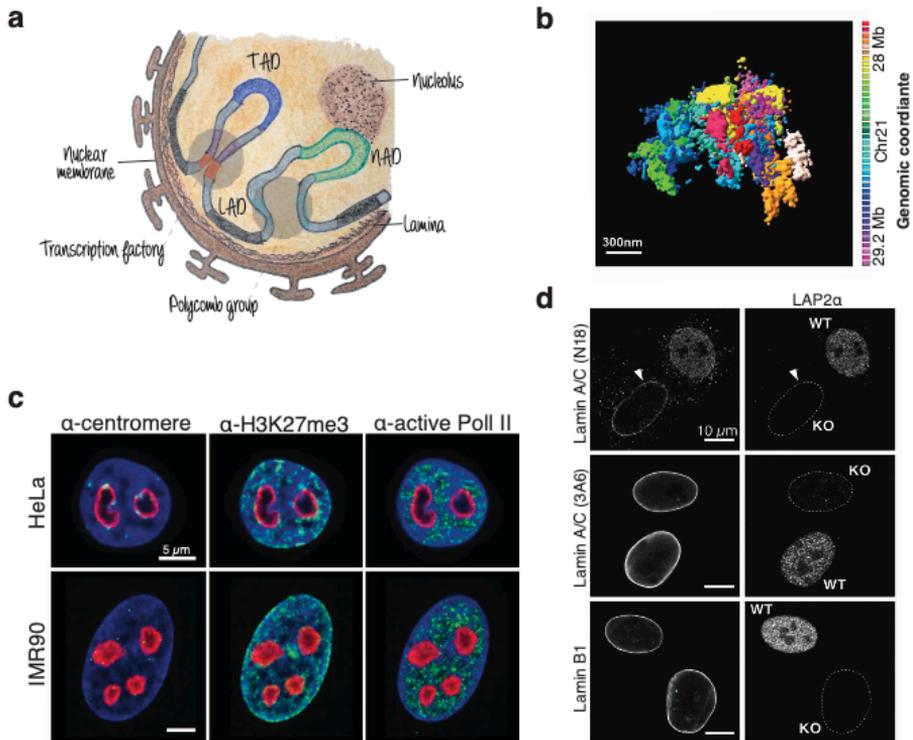


Figure 3. Subdomains organization. **a**, Illustration of a cell nuclei with the three subdomains explained in the Subdomains section. The transcription factory is highlighted in a grey circle next to the TAD represented as a blue rectangle, helping to gather regulatory units highlighted in colored boxes, to enhance gene expression. LADs are represented as black rectangles, being in contact with the lamina. Polycomb group proteins are highlighted with a grey circle, being relevant to regulate cell identity. Finally, NADs are represented using green rectangles, being in contact with the nucleolus. **b**, 3D STORM images of 41 consecutive 30 kb chromatin segments from a 1.2 Mb of IMR90 chromosome 21 (Adapted from (Bintu et al., 2018)). **c**, Immunostaining in HeLa and IMR90 cells of α -centromere, α -H3K27me3 and α -active Pol II signals shown in green, nucleolar staining in red, and DAPI stain in blue (Adapted from (Nemeth et al., 2010)). **d**, Confocal immunofluorescence microscopic images of wild-type and knock-out mouse dermal fibroblast double-stained for LAP2 α and lamins. The unstained nuclei are highlighted using a white dashed line (Adapted from: (Gesson et al., 2016)).



DNA-looping
 $\sim 10 - 100$ kb

Genetic studies have demonstrated that translocations or deletions can affect the correct transcription regulation of genes located at distal regions (tens to hundreds of kilobases). It indicates the ability of regulatory elements to exert their function over large genomic distances. Enhancers have been described as key gene-regulatory elements that can control gene expression in a cell-specific and spatiotemporal manner through long-range chromosomal interactions. In mammal genomes, such as mouse and human, hundreds of thousands of putative enhancers have been mapped. The elucidation of enhancer-promoter pairs solely based on the linear distance leads to high number of false positive assignments. Thanks to major technological advances that allowed the genome-wide mapping of enhancer-promoter contacts at high resolution, elucidated the principles of enhancer function and enhancer-promoter communication. These type of events are known as DNA-looping or loops, which permit the regulation of genes by distant regulatory elements.

Loop extrusion model

Over the years the study of how genes can be regulated from a distant elements has been of high interest, being the “loop model” the more prevalent (Ptashne, 1986). As mentioned in previous sections (see Topologically Associating Domains), CTCF has been observed to play a major role in chromatin organization. CTCF was first described as a repressor of the *C-MYC* oncogene in chicken, mouse and human (Filippova et al., 1996). Classically, CTCF has been observed as an insulator, blocking the communication between gene promoters and

distal enhancers. The first study proving this CTCF capacity was in transgenic assays in the chicken β -globin locus, where the enhancer and the reporter gene are separated by 1.2 kb DNA segment. This study revealed that the insulator ability of CTCF was dependent on its relative position, as was only observed when placed between the enhancers and the promoter (Bell & Felsenfeld, 1999). CTCF is a zinc-finger protein highly conserved across bilaterians, and absent in yeast, derived nematodes as *Caenorhabditis elegans*, fungi and plants (Heger, Marin, Bartkuhn, Schierenberg, & Wiehe, 2012). The central region of CTCF consist of 11 zinc fingers, with around 20 bp well-conserved and non-palindromic as a core region, also referred as M1 motif (Schmidt et al., 2012)(Figure 4). What makes CTCF binding motif unique, in comparison to the majority of the known transcription factors (TFs), is that it is long and information-enriched, meaning that can possess pleiotropic functions by diverse combination of its 11 zinc fingers (Nakahashi et al., 2013). Moreover, in mammals, have been described shorter motifs of around 10 bp, up- and downstream to M1 which can help to stabilize or destabilize the CTCF binding, respectively. Despite some CTCF sites lack any sequence motif, they present a markedly lower affinity than those with the core motif (Nakahashi et al., 2013)(Figure 4).

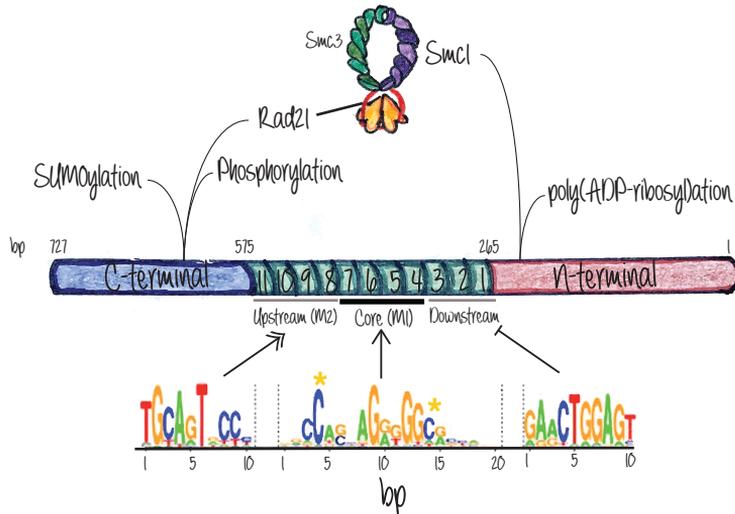


Figure 4. Illustration of CTCF regulation. CTCF is depicted in a reverse orientation to reflect its binding. CTCF can be regulated by diverse post-translation modifications (PTMs), such as SUMOylation, phosphorylation (by casein kinase 2) or Poly(ADP-ribosylation) by PARP1. The C-terminal of CTCF can interact with the cohesin subunit, Rad21. CTCF can regulate its binding to chromatin by the central 11 zing finger domains. The core motif (M1) consist of the 4 central zinc fingers; whereas the up and downstream motifs can stabilize or destabilize its binding, respectively. The methylation of two cytosines located in the core motif can impede the CTCF binding.

The binding of CTCF into chromatin correlates to gene density, especially in intergenic regions, gene bodies and near TSSs (Holwerda & de Laat, 2013). Moreover, its binding capacity can be regulated at multiple levels. First, methylation of cytosines in the M1 motif will impair the CTCF binding (Wang et al., 2012). Second, the binding has been observed normally to happen in nucleosome-free regions (Fu, Sinha, Peterson, & Weng, 2008). Finally, post-transcriptional modifications in the CTCF protein such as SUMOylation, poly(ADP-ribosylation) and phosphorylation by casein-kinase 2 (CK2), can interfere CTCF function without affecting its chromatin binding (El-Kady & Klenova, 2005; Kitchen & Schoenherr, 2010; Pavlaki et al., 2018)(Figure 4).

Mammalian genomes contain around 50,000 CTCF binding sites, with around 10-20 % located at TAD boundaries and 60-70 % in intra-domain regions (Cuddapah et al., 2009; S. S. Rao et al., 2014; Tang et al., 2015). CTCF has been described to be an architectural protein, as well as, cohesin and Mediator, all three are involved in chromatin-looping (Parelho et al., 2008; Rubio et al., 2008), but not fully required for the maintenance of TADs. Cohesin is a highly conserved protein complex assembled in a ring-like structure by two proteins from the “Structural Maintenance of Chromosomes” family, Smc1 and Smc3, and a kleisin family subunit, Rad21. The cohesin complex is mainly known to be responsible for tethering sister chromatids during mitosis by encircling the DNA fiber. However, cohesin has been found in non-dividing cells, and mutations on its subunits can lead to developmental impairments in humans, suggesting its role in gene regulation (Brooker & Berkowitz, 2014). In vertebrates, it was observed that around 50 to 80 % of CTCF sites were co-occupied by cohesin, which can interact directly with the C-terminal of CTCF through the Rad21 subunit (Figure 4)(Xiao, Wallace, & Felsenfeld, 2011). While cohesin is not needed for the binding and positioning of CTCF, CTCF is necessary for the positioning of cohesin. As expected and according to all these observations, CTCF and cohesin have been found to participate in the chromatin loop formation, having a direct effect on gene activation. Recent studies have showed how the CTCF N-terminus interacts with Smc1 subunits of human cohesin, probing that this interaction is required for the positioning of cohesin in CTCF-anchored loops (Li et al., 2020)(Figure 4). Nevertheless, CTCF is able to form loops via homodimerization, or by interacting with other proteins: such as, nucleophosmin (NPM) (Yusufzai, Tagami, Nakatani, & Felsenfeld,

2004), thyroid hormone receptor (TR) (Lutz et al., 2003), chromodomain helicase 8 (CHD8) (Ishihara, Oshimura, & Nakao, 2006) and transcription factor Yin Yang 1 (YY1)(C. Guo et al., 2011). Despite all the possible interactors of CTCF, none has been observed to co-localize genome-wide as extensively as with cohesin. This suggest that CTCF might have different interactors depending on the epigenetic scenario, participating in a wide range of protein complex modulations to adapt its function to specific loci in a given conditions. For example, various loops can appear during differentiation, without changes in CTCF binding, which have been observed to be enriched in other TFs binding motifs (Bonev et al., 2017; Phanstiel et al., 2017). Hi-C experiments showed that CTCF chromatin loops occur preferentially between motifs in convergent orientation (S. S. Rao et al., 2014). This orientation preference was tested by CRISPR-mediated inversions of different CTCF motifs, which disrupted the loops and created new enhancer-promoter interactions (Y. Guo et al., 2015). Moreover, it is possible to accurately predict the changes in CTCF loops after inversions or deletions of CTCF motifs (Sanborn et al., 2015). This preference on motifs orientation suggests that CTCF pairs are formed in a dimension-restricted space, via an extrusion process mediated by the cohesin complex (Fudenberg et al., 2016; Nichols & Corces, 2015; Sanborn et al., 2015). Loop formation can be explained by the loop extrusion model, which suggests that SMC cohesin subunits progressively extrude chromatin until halted by convergent-oriented CTCFs (Fudenberg et al., 2016)(Figure 5a). The capacity of CTCF to stall the extrusion may be caused by its ability to induce chromatin conformational changes, such as nucleosome repositioning (Clarkson et al., 2019; Fu et al., 2008). Atomic force microscopy experiments showed

that DNA wraps around bound CTCF forming structures of around 67- 80 nm in diameter (Mawhinney et al., 2018), larger than proteins able to block cohesin sliding in vitro (Davidson et al., 2016; Stigler, Camdere, Koshland, & Greene, 2016)(Figure 5c).

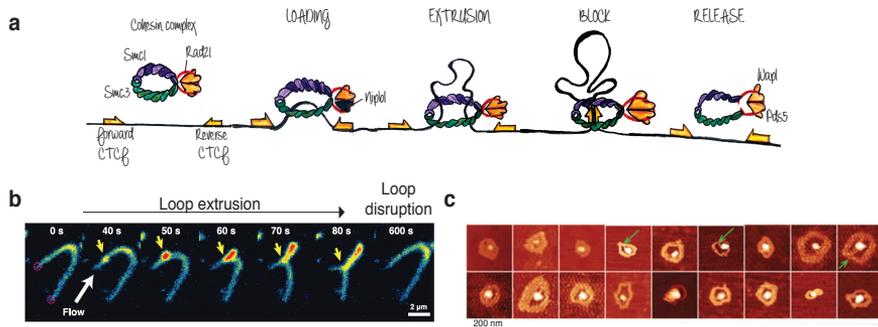


Figure 5. Loop extrusion illustration and supporting in vitro experiments. **a**, Schematic view of the loop extrusion model, from the loading of cohesin into chromatin until its release. Here we can observe CTCF pairs in a convergent oriented manner, as yellow arrows, which will halt the extrusion. Are also depicted Nipbl, the cohesin loader, and Wapl the release factor (Adapted from (Rowley & Corces, 2018)). **b**, Snapshots showing DNA loop extrusion by condensin on a SxO-stained double-tethered γ -DNA. The constant flow (white arrow) maintains the DNA in the imaging plane and extrudes the loop. The position of the loop base is pointed by a yellow arrow. At 40 s starts to appear a small loop that will grow over time until 80 s. The loop is disrupted after 600 s (Adapted from (Ganji et al., 2018)). **c**, Atomic force images of DNA in the presence of the 11 zinc fingers domains of CTCF. Green arrows indicate multiple DNA strands (Adapted from (Mawhinney et al., 2018)).

Several *in vitro* studies proved the diffusing capacity of cohesin along the DNA, until being blocked by CTCF (Davidson et al., 2016; Stigler et al., 2016). Importantly, it has been observed the condensin mediating loop extrusion *in vitro* (Ganji et al., 2018; Terakawa et al., 2017)(Figure 5b). Interestingly, a single cohesin ring can capture two DNA fragments, but only if one is single stranded (Murayama, Samora, Kurokawa, Iwasaki, & Uhlmann, 2018), suggesting that a single ring is able to lead this process. Polymer physics simulations, CHIP-exo and CHIP-nexus experiments have shed light on understanding how the cohesin ring is randomly loaded into chromatin, until blocked by CTCF when reaching

the 3' end of the motif (i.e. within the loop) (Fudenberg et al., 2016; Nagy et al., 2016; Tang et al., 2015). Moreover, while cohesin depletion leads to CTCF loops loss (S. S. P. Rao et al., 2017; Schwarzer et al., 2017; Wutz et al., 2017), the depletion of WAPL, a cohesin release factor, results in the formation of longer loops due to an extended period of cohesin bound into chromatin, and a decreased long-range interactions between compartmental domains (Gassler et al., 2017; Haarhuis et al., 2017). The deletion of Nipbl, the cohesin loader, and Rad21 subunit, did not affected the CTCF binding, but resulted in a general loss of CTCF loops and a stronger compartmental segregation (S. S. P. Rao et al., 2017; Schwarzer et al., 2017; Wutz et al., 2017)(Figure 6). Auxin-mediated degradation of Rad21 showed that 6 h was enough to lose CTCF loops, suggesting the need of being constantly extruded by cohesin (Fudenberg et al., 2016). Then, after 40 min of restoring cohesin, CTCF loops as large as 900 kb were established (S. S. P. Rao et al., 2017). The condensin extrusion rate in bacteria was assessed by real-time imaging *in vitro*, being around 600 bp/s (Ganji et al., 2018), however, it is still unmeasured in mammals. These studies suggest that the size of the loop might depend on the time that cohesin is able to extrude chromatin until stopped by CTCF, and that cohesin mediates other interactions apart from those involving CTCF. Moreover, the loop extrusion process must coexist with mechanisms responsible for interactions between compartments of the same type.

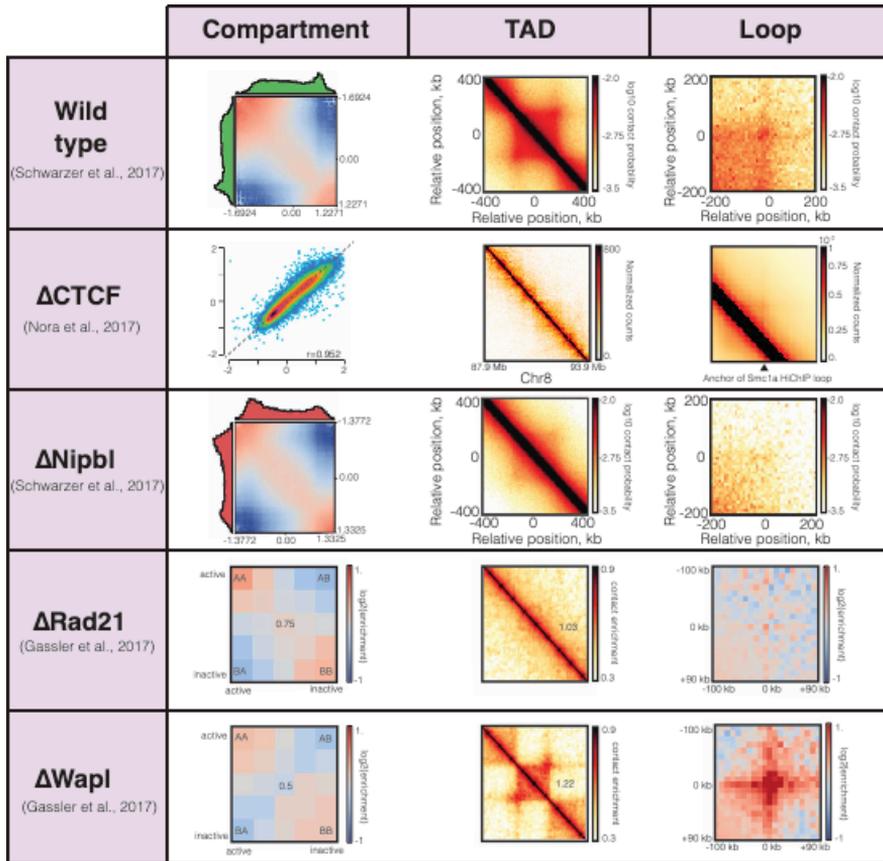


Figure 6. Summary of chromatin organization effects by cohesin and CTCF mutants. Each row represents a condition, and each column represent the chromatin organization layer being analyzed. **a**, In the first row the results from a wild-type mouse are shown. The compartmentalization is represented by saddle plots in which the average interaction frequencies between pairs of loci (100 kb bins) arranged by their compartment signal (eigenvector). The histograms on the axis show the distributions of eigenvector values. TADs are shown as an average Hi-C map of all the TADs called in the dataset. The loop information is shown as the average Hi-C map around 102 peaks. **b**, Results of degrading CTCF in mESCs. The compartments are not affected, as can be observed by the correlation between the mutant and the wild-type, of the first eigenvector values. However, TADs disappeared, as shown in a 6 Mb segment of Hi-C data at 20 kb resolution. Loops were also affected, visualized by aggregating the Hi-C signal from Smc1 Hi-ChIP loops separated by 280-380 kb. **c**, Results from a mouse mutant, with a deletion of Nipbl. Here the compartment information is retrieved like in the wild-type. Notice an enrichment of AA and a depletion of AB interactions. Can also be observed a TAD and loop lose. **d**, Deletion of Rad21 results in mouse zygotes. The average loops, TADs and compartments are done by pooling together the maternal and paternal data. TADs and loops were entirely absent. However, the compartmentalization of active and inactive chromatin was increased. **e**, Study of deleting WAPL in mouse zygotes. Here the same analysis

as in Rad21 knock-out is applied. This mutant showed stronger signal of TADs and loops, while the compartments became weaker.

How cohesin can move along the DNA *in vivo* is not known, and multiple models have been proposed, such as diffusion, motor activity or involvement of other macromolecules (Brackley et al., 2018; Terakawa et al., 2017; Vian et al., 2018). For example, it has been proposed that cohesin can be relocated by transcription, directly pushed by RNAPII (Ocampo-Hafalla, Munoz, Samora, & Uhlmann, 2016) or indirectly by chromatin supercoiling (Racko, Benedetti, Dorier, & Stasiak, 2018). A study showed that transcription elongation resulted in a decrease in CTCF looping. Strikingly this decrease was weaker than upon ATP depletion, showing the emergence of hundreds of CTCF-independent loops (Vian et al., 2018). Moreover, loop extrusion via RNAPII fails to explain the formation of inactive domains and present a slower elongation rate compared to the estimated loop extrusion speed, $\sim 9\text{-}90$ bps/s versus $\sim 374\text{-}850$ bp/s (Jonkers & Lis, 2015; S. S. P. Rao et al., 2017). However, it can suggest that RNAPII will interfere cohesin extrusion over transcriptionally active regions. Overall, the proposed models are supported by some evidence, nevertheless each presents its limitations, suggesting that a combination between some may underlie the extrusion process.

Finally, CTCF can mediate cell-to-cell gene expression variability by regulating enhancer-promoter interactions (Ren et al., 2017). It has been shown that around the 41 % differential CTCF binding through 19 human cell types is due to methylation, as disruption of CTCF binding is associated with increased methylation at promoter sites (Wang et al., 2012). GWAS studies have revealed various mutations in CTCF binding sites, which can affect TAD organization or enhancer-promoter

interactions, leading to an increased variability of gene expression (Hnisz et al., 2016; Katainen et al., 2015). Thus, CTCF can contribute to cellular heterogeneity in mammals by mediating transcriptional pausing (Paredes, Melgar, & Sethupathy, 2013) and alternative mRNA splicing (Shukla et al., 2011). In the future, data from single-cell analysis of chromatin organization and epi-transcriptomics, will be needed to provide more evidences of the contribution of CTCF to cellular heterogeneity and its relevance in diverse biological scenario.

Chromatin loop detection methods

“Impostor syndrome is the frequent feeling of not deserving one’s success, and of being of a failure despite a sustained record of achievements. Highly successful people often experience it throughout their careers, especially when they are members of a group that is underrepresented in their profession—such as female scientists or engineers”.

Maria Klawe, 2014

The advent of genome-wide chromatin architecture (mainly thanks to Hi-C) has made necessary the development of loop detection methods. During the last two decades, a lot effort has been put to develop a loop detection software. There is a plethora of different software (Table 1). Among these, a selection has been benchmarked in detail in (Forcato et al., 2017). Here is a brief description of their methodology:

1. *Hi-C Computational Unbiased Peak Search (HiCCUPS) (Durand et al., 2016; S. S. Rao et al., 2014)*

HiCCUPS is implemented as a part of Juicer suite of tools for the analysis and visualization of Hi-C experiments. This software by default

needs as input the normalized contact matrix with Knight-Ruiz matrix balancing (Knight & Ruiz, 2013). Its algorithm looks for clusters of contact matrix entries in which the contact frequency is enriched compared to the local background. It scans each pixel and compares its number of contacts with four neighboring areas, giving the possibility to cluster the significant peaks. Implemented using GPUs, as it has to analyze trillions of pixels in kilobase-resolution experiments. It is programmed in JAVA language.

2. *Fit-Hi-C* (Ay, Bailey, & Noble, 2014) and *Fit-Hi-C 2* (Kaul, Bhattacharyya, & Ay, 2020)

This software is designed to identify mid-range intra-chromosomal contacts, and inter-chromosomal with the new version. The algorithm relies in a model with two splines. The first spline models the observed counts according to the genomic distance between all the possible pairs. The second spline is used to fit to calculate a refined null model. The biases are computed using the Iterative Correction and Eigenvector decomposition normalization (ICE) (Imakaev et al., 2012) and incorporated in the expected contact probability. Using a binomial distribution are calculated the p-values and corrected for multiple testing. Fit-Hi-C is programmed in Python language.

3. *GOTHic* (Mifsud et al., 2017)

This software takes as input the aligned reads, it pairs them, and assigns the pairs to enzyme-specific restriction fragments and finally discards the ones separated by less than 10 kb. To normalize the counts applies a similar Vanilla Coverage (Lieberman-Aiden et al., 2009) and a binomial test to capture the significant interactions and a Benjamini-

Hochberg multiple testing correction (Benjamini & Hochberg, 1995). The output contains both *cis* and *trans* interactions with the \log_2 of observed over expected counts, p-value, FDR and the number of read pairs of the interaction. GOTHiC is programmed in R.

4. HOMER (Heinz *et al.*, 2010)

HOMER is a command-line based software. As input takes the aligned reads, which are paired and filtered. In order to identify the significant interaction bins, HOMER generates a background model that normalizes the genomic interactions for linear distance and coverage at a specific bin size. This model permits to estimate the expected read count and to apply a binomial test to get the significant *cis* and *trans* interactions. Finally, HOMER outputs a p-value, the false discovery rate (FDR), the number of read pairs supporting the interaction, and the interaction distance. HOMER is programmed in Perl and R languages.

5. High-throughput Identification Pipeline for Promoter Interacting Enhancer elements (HIPPIE) (Hwang *et al.*, 2015)

HIPPIE identifies the significant interactions at a restriction fragment-level. To be run requires a computing cluster with Open Grid Scheduler or other Sun Grid Engine (SGE) compatible with job schedulers. HIPPIE consists in five steps: mapping, quality control, Hi-C interactions identification, enhancer-target gene interactions and its prediction analysis. It classifies the read pairs as specific or non-specific if the sum of the distance of the reads from the closest restriction enzyme site is smaller or bigger than a given size selection parameter (Yaffe & Tanay, 2011). The biases are computed as in (Jin *et al.*, 2013) which estimates the expected random contact frequencies considering

mappability, GC content, fragment length and fragment distance (for intra-chromosomal read pairs). The significant interactions are identified by fitting a negative binomial distribution. HIPPIE retrieves the *cis* and *trans* restriction fragment-based interactions with the associated p-value. The software is programmed in Python, Perl and R languages.

6. *DiffHiC* (Lun & Smyth, 2015)

DiffHiC is intended to identify significant differential interactions between Hi-C experiments. However, it contains a method to call the interactions on individual samples based on the signal enrichment over a local background. The software consists on read mapping, filtering and genome partitioning to obtain the counts between the genomic bin pairs. It applies a “local enrichment” algorithm to identify the bin pairs with more reads than their neighbors. It computes the \log_2 fold change between the number of read pairs of the target-bin and the neighbor region with greatest abundance. DiffHiC does not applies statistical test, meaning that there is no significance value associated to the interaction. It is programmed in R and Python languages.

<i>Software</i>	<i>Input data</i>	<i>Algorithm</i>
HMRFBayesHiC (Xu, Zhang, Jin, et al., 2016)	O/E Hi-C matrices	HMRF model
FastHiC (Xu, Zhang, Wu, Li, & Hu, 2016)	O/E Hi-C matrices	HMRF based on simulated field approximation
Binless (Spill, Castillo, Vidal, & Marti-Renom, 2019)	Hi-C mapped reads (TSV)	Negative binomial likelihood

r3C-seq (Thongjuea, Stadhouders, Grosveld, Soler, & Lenhard, 2013)	3C mapped reads (BAM)	Background scaling method (Z-scores)
ChiaSig (Paulsen, Rodland, Holden, Holden, & Hovig, 2014)	ChIA-PET interaction file (BEDPE)	NCHG distribution
CHiCAGO (Cairns et al., 2016)	Capture Hi-C data mapped reads (BAM)	Convolution background model
Mustache (Ardakany, Gezer, Lonardi, & Ay, 2020)	Raw contact map (TSV)	2D Gaussian

Table 1. Other available software for the identification of significant chromatin interactions. For each software the input file, and the used algorithm, are specified. All the software of the table run statistical test to retrieve the significant chromatin interactions. Bayesian Additive Regression Trees (BART); Hidden Markov random field model (HMRF); Non-central hypergeometric (NCHG).

Technology to assess nuclear organization

“Although we don't know what is outside our universe, astronomers still wonder. Several pictures of what there might be have been dreamed up. An interesting one, called multiverse, has lots of universes. Picture it as a foam of bubbles. Our universe would be one bubble, and we'd be surrounded by lots of other bubbles.”

Jocelyn Bell, 2013

Understanding the three-dimensional spatial genome organization is paramount to fully characterize its function. Therefore, it is fundamental to the use and develop microscopy technologies, both conventional and super-resolution, as well as chromosome conformation capture (3C) techniques. Each of these two types of technologies allow a given resolution and the inspection of specific chromatin topology levels (François Le Dily, Serra, & Marti-Renom, 2017).

Microscopy

As mentioned in the “Nuclear organization” section, many of the basic principles of genome architecture were discovered by microscopy techniques. For instance light and electron microscopy proved the existence of nuclear bodies, such as the nucleolus, nuclear speckles Cajal bodies and polycomb bodies (Denker & de Laat, 2016). Using this technology was possible to observe the nuclear localization of active euchromatin which is more densely stained than heterochromatin (Heitz, 1928). Moreover, it has been shown that chromosomes territories occupy a specific position with a slight intermingling (Cremer, Cremer, Schneider, et al., 1982). It was also observed that some genes occupy specific nuclear positions according to their transcriptional status (Brown et al., 1997). Thanks to recent advances in super-resolution microscopy techniques, it is now possible to visualize from whole chromosomes to few kilobases interactions among nearby cis-regulatory elements. Advances in robotics and microfluidics, as well as accelerated super-resolution, it is also possible to inspect thousands to hundreds of thousands of individual cells, thereby increasing robustly the statistical power of imaging. The breakthrough of the Oligopaints technology has transformed the sequence-specific imaging of genome organization. Stochastic optical reconstruction microscopy (STORM) technology, captured images of active, inactive and Polycomb-repressed (PC-repressed) epigenetic states in *Drosophila*. This study revealed that each state obeys a different power-law scaling of domain size as function of length. Meaning that these three states have different packaging density, levels of self-interaction and levels of interaction

between them. For instance, in *Drosophila* and mouse, PC domains are compacted globular structures, contrasting with active state structures, which are more extended. Due to the epigenetic role of PC in maintaining transcriptionally silent throughout cell divisions, it has been hypothesized that its characteristic densely packed structure and high degree of spatial exclusion of neighboring active domains, contributes to this ability (Boettiger et al., 2016; Kundu et al., 2018). Recently, the optical reconstruction of chromatin architecture (ORCA) method was able to visualize very small TADs (< 20 kb) in several cells. This method allowed to study a strong cell-type specific physical partition of the posterior Hox complex in *Drosophila*, in which two transcriptionally active genes are separated. Surprisingly, the boundary of these two genes did not corresponded to a divergent epigenetic state, as both *Ubx* and *abd-A* gene regions were in an active epigenetic state, and a deletion of ~ 4 kb between these two domains disrupted their correct expression (Mateo et al., 2019). This proved that the epigenetic state is not the unique mechanism determining the chromatin organization, and showed the contribution of the boundary elements *in vivo* (Figure 7). Since the discovery of TADs thanks to 3C-based methods, which are mainly population-based run in millions of cells, their existence in single-cells has been very controversial. Microscopy techniques, showed that while chromosome territories correspond to a physical structure, TADs only exist as statistical features in population of cells. However, only around 100 cells are needed to provide a proper delineation of TADs, subTADs and loops.

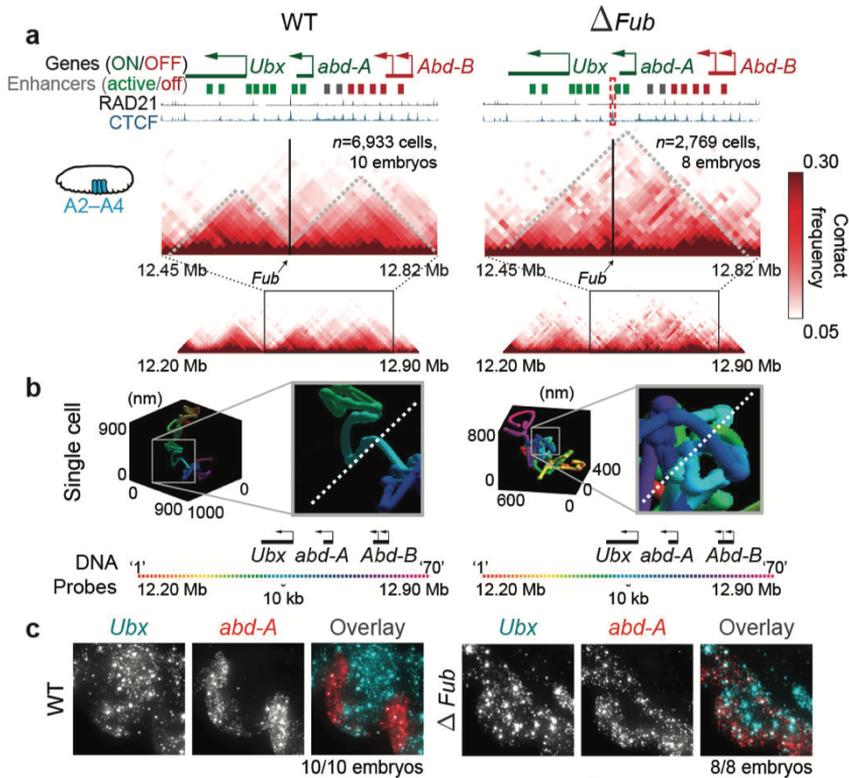


Figure 7. Deletion of TAD borders leading to TAD fusion and enhancer crosstalk. **a**, Contact frequency computed from ORCA for A2-A4 in wild-type (WT) and *Fub* mutants, with a deletion of a TAD boundary (mark with red dotted lines), at 10 kb resolution. Dashed grey lines mark the TADs. Below are marked the bithorax-complex (BX-C) genes, and the ChIP-seq of CTCF and Rad21 from WT embryos. **b**, ORCA snapshots from one experiment of a 700 kb domain containing the BX-C genes. Barcode position is marked by color. The dashed line in the zoom-in images, indicate the 3D separation of up and downstream regions of the *Fub* locus. **c**, Comparison of the expression of two genes: *abd-A* and *Ubx* in WT and *Fub* mutant embryos in individual regions (Adapted from (Mateo et al., 2019)).

The globular structures or TAD-like features are present in single cells, and persist even in the absence of cohesin, whereas TADs in identical compartment types (sharing an epigenetic state) disappear (Gassler et al., 2017; Haarhuis et al., 2017; S. S. P. Rao et al., 2017). Microscopy methods explained the emerging of this pattern as an uniform statistical position of the boundaries instead of biased to CTCF binding sites (Bintu et al., 2018)(Figure 8). This new data has promoted multiple

models to explain the competing mechanistic of chromatin folding (Barbieri et al., 2012; Brackley et al., 2018; Fudenberg et al., 2016; Giorgetti et al., 2014; Jost, Carrivain, Cavalli, & Vaillant, 2014; Sanborn et al., 2015).

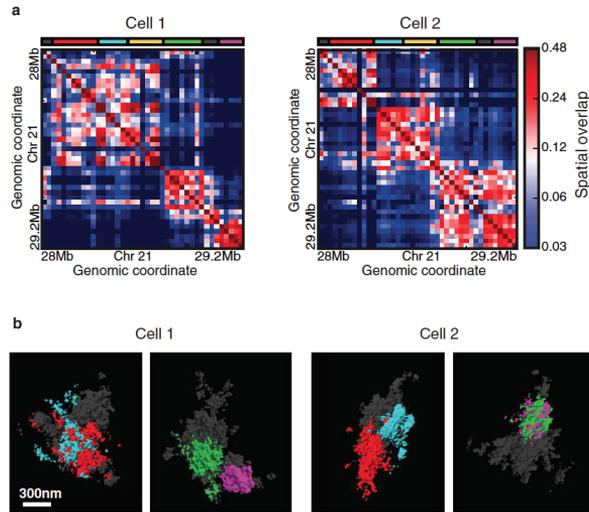


Figure 8. Visualization of TAD-like globular domains in single cells. **a**, Matrices from the spatial-overlap from a 1.2 Mb region (chr21: 28-29.2 Mb) from two individual IMR90 cells. The color barcode below the contact matrix indicate the five sub-TADs identified at population level. **b**, Images from multiplexed 3D STORM corresponding to the two cells shown before. Per cell are shown two images, in each highlighting two sub-TADs and with a specific rotation to ease their visualization (Adapted from (Bintu et al., 2018)).

Finally, recent advances in live imaging techniques validated the dynamic and transient nature of chromatin interactions (Bintu et al., 2018; Finn et al., 2019; Mateo et al., 2019). Being possible to study the transcriptional dynamics, memory, and gene co-regulation (Alexander, Guan, Huang, Lomvardas, & Weiner, 2018; Chen et al., 2018; Ferraro et al., 2016). For the future, it will be highly important to combine the flexibility and ease use of microscopy methods, to enhance the resolution, image DNA and protein at the same time, and increase its throughput and the robustness.

Chromosome conformation capture technology

In the last decade, the development of 3C-based technology (3C) and thereby their modelling and analysis, has revolutionized the understanding of 3D chromatin architecture. First, 3C techniques were applied to yeast chromosomes (Dekker, Rippe, Dekker, & Kleckner, 2002), then, as explained before, to observe the long distance chromatin loops between enhancer and the target genes in the β -globin locus (Tolhuis, Palstra, Splinter, Grosveld, & de Laat, 2002). Finally, this technique was adapted to its high-throughput variants (Table 2). All these techniques are based on formaldehyde crosslinking of chromatin, which allows the capture of a snapshot of the interactions among any pair of genomic loci in the three-dimensional nuclear space. Then chromatin is fragmented by digestion, and re-ligated in order to convert the interacting loci into unique DNA ligation products to be finally detected by different methods. Originally, PCR with locus-specific primers was used to detect ligation products one at a time. The development of deep-sequencing techniques allowed the high-throughput detection of ligation products and permitted the interrogation of the chromatin architecture genome-wide, the most popular implementation of this method is called Hi-C (Lieberman-Aiden et al., 2009).

<i>Technology</i>	<i>Type of approach</i>	<i>Advantage</i>	<i>Limitation</i>
3C (Dekker et al., 2002)	One-to-one	Cheap and simple	Amplification efficiency
4C (Simonis et al., 2006)	One-to-all	No need a priori knowledge	Amplification efficiency

5C (Dostie et al., 2006)	Many-to-many	Unbiased readout and great coverage	Need a priori knowledge
ChIA-PET (Fullwood et al., 2009)	Many-to-many	Enrichment in rare interactions	Difficult quantification
Capture-C (Hughes et al., 2014)	Many-to-all	Detection of SNPs	Sequence enrichment efficiency
Micro-C (T. H. Hsieh et al., 2015)	All-to-all	High resolution	Increase noise in long distance
HiChIP (Mumbach et al., 2016)	Many-to-all	Low input requirement	RE site proximity / Antibody efficiency

Table 2. Existing 3C-derived methods. In the table are listed the main advantages and drawbacks of the most used technologies, as well as the information retrieved.

High-throughput conformation capture (Hi-C)

This technique allows the interrogation of all loci at once. The experiment requires around 5 million of cells as an input material in order to have enough library complexity for sequencing. Although this limitation has been addressed in recent versions of the Hi-C protocol, it is still normally conducted on a population of cells and the resulting 3D genome organization represents the ensemble of single cells of the population. The development of the Hi-C assay represents a breakthrough in the understanding of genome organization as it is unbiased and unsupervised. As in the others 3C-based methods, the chromatin is crosslinked with formaldehyde and digested by a restriction enzyme (RE), which leaves a 5' overhang to be filled by free nucleotides some of which biotinylated. Then the repaired blunt ends are ligated and the DNA is sheared and purified using a biotin pull-down with streptavidin beads. The final product of the experiment before sequencing is a library of ligation junctions that will be sequenced

using a paired-end approach. On the bioinformatics side, both reads will be mapped to a reference genome, obtaining the *cis* and *trans* contacts (respectively inter- and intra-chromosomal contacts) to generate the contact matrices. It has to be considered that the resolution of the experiment is highly dependent on the RE sites frequency. One solution to increase x-fold the resolution is to sequence x^2 more pairs, keeping in mind that the resolution could never go deeper than the RE fragment sizes. The first version of Hi-C corroborated the existence of A and B chromatin compartments. Since the first Hi-C assay it has been exponentially improved retrieving kilobase contact matrices allowing the detection of sub-compartmentalization of the chromatin (Imakaev et al., 2012) (S. S. Rao et al., 2014).

The bioinformatics analysis of Hi-C experiments consists mainly on 5 steps:

- I. Hi-C quality check and mapping to reference genome.
- II. Read filtering: removal of PCR artifacts as well as products from non-standard ligation. Only the valid pairs will be kept.
- III. Building of contact matrices: the genome is chunked into non-overlapping bins of fixed size.
- IV. Bin filtering: removal of matrix columns with low counts. If the observed number of reads in a bin is much lower than average, it is expected that they belong only from mapping artifacts.
- V. Matrix normalization: remove the inherent biases of the experiment. There are two types of normalization approaches: explicit methods, that suppose that all the biases are known, like GC content, mappability or number of RE sites per bin; such as OneD normalization (Vidal et al., 2018). And implicit or balancing methods, which assume an equal experimental

visibility of each bin, including ICE (Imakaev et al., 2012), square root vanilla coverage, vanilla coverage normalization (S. S. Rao et al., 2014) and Knight-Ruiz Matrix Balancing (KR)(Knight & Ruiz, 2013).

However, Hi-C protocols present caveats, including an overrepresentation of inter-chromosomal contacts (~ 60 %) due to random ligations between unrelated DNA fragments. This overrepresentation is typically expected to be due to random ligation between unrelated DNA fragments (i.e., uncross-linked). This high percentage will not affect the intra-chromosomal contacts measured from short and medium-range distance (< 2 Mb), but the long-distance (> 10 Mb) between and within chromosomes. The inter-chromosomal contacts were reduced to < 20 % by placing the ligation *in situ* inside the nuclei instead of *in solution* (Nagano et al., 2015). Another critical step and potential source of biases is the use of formaldehyde to fix DNA fragments. Formaldehyde links proteins or protein and DNA however the fixation could differ depending on the proteins and potentially chromatin state, being a drawback to capture protein-mediated loops. Moreover, highly dynamics and fluctuating interactions might not be captured as formaldehyde takes at least 5 seconds to crosslink. Another important issue is to consider the meaning of the retrieved quantitative contact by 3C experiments, as these methods assess the ligation frequencies between cross-linked and fragmented DNA segments. For a DNA segments to participate in the 3C contact profiles it needs to i) be cross-linkable, ii) have DNA ends available to ligate, and iii) in the cross-linked DNA-protein aggregates compete with other fragments to ligate. These three requisites depends on size, chromatin composition,

duration and strictness of fixation (Dekker, 2006). The ligation efficiency can work as a proxy for contact frequencies, which must be validated by microscopy or genetics (i.e., DNA deletions).

Despite its limitations, Hi-C is the chosen method to obtain the genome's interactome. During the past 10 years, in the majority of the labs the Hi-C technology became ordinary, thus there are Hi-C contact maps available for the most used cell types from human and mouse, human tissues and organoids. The 3D genome organization together with epigenomics and transcriptomics data would help to interpret the cell homeostasis and the genetic variations associated to disease. All the recent advances in single-cell Hi-C, will help to reveal the heterogeneity of specific chromatin contacts within cell populations and cell-to-cell during multiple biologic processes (i.e., development).

Function-Structure dogma

“I don't want to say epigenetics isn't exciting ... [but] there's a gap between the fact and the fantasy. Now the facts are having to catch up.”

Edith Heard, 2013

The relationship between structure and function is a paramount in structural biology. First, was reported that changes in structure were directly related to sequence modifications in proteins (Chothia & Lesk, 1986). In the case of 3D genome organization, as described in previous sections (see Nuclear Organization), TADs may represent regulatory domains due to their high conservation in mammals and the co-regulation of their contained genes (Dixon et al., 2015; F. Le Dily et al., 2014). Moreover, the functional relevance of TADs has been validated by studies in diseases caused by structural variations (SVs), which the

disruption of TADs caused congenital limb malformations, gene misexpression and cancer (Dixon et al., 2018; Franke et al., 2016; Lupianez et al., 2015; Spielmann, Lupianez, & Mundlos, 2018). However, neither CTCF nor cohesin depletion resulted in large repercussions on gene expression (Elphege P Nora et al., 2017; S. S. P. Rao et al., 2017; Wutz et al., 2017). These studies challenge the crucial implication of 3D genome organization for gene regulation, suggesting a more complex and multi-layer effect of chromatin architecture on gene regulation. Moreover, high resolution experiments showed that not all the regions of the genome follow the same rules. For example, gene dense regions do not present a canonical 3D structure, as they can form multiple smaller interacting domains, suggesting an alternative regulation as large TAD and loops. Recently, thanks to higher resolution and more detailed analysis, 3D architectural stripes have been described, which are asymmetrical patterns of contacts that can span several 100 kb, reflecting a unidirectional loop extrusion process. Architectural stripes are associated with strong and active enhancers, which are scanning the genome for a target gene, in close proximity to a CTCF boundary, defined as stripe anchor. (Barrington et al., 2019; Kraft et al., 2019; Vian et al., 2018). Similarly, the development of the micro-C method, which provides 3D chromatin structure at nucleosomal resolution, revealed the existence of micro-TADs, with a median size from 5.4 to 40 kb, reflecting individual transcriptional units (T. S. Hsieh et al., 2020). It makes essential to understand the relationship between 3D chromatin structure and gene regulation to decipher the possible regulatory scenarios.

Differential chromatin interaction identification software tools

Humans are allergic to change. They love to say, “We’ve always done it this way”. I try to fight that. That’s why I have a clock on my wall that runs counter-clockwise.

Grace Murray Hopper, 1987

Nowadays there are a large number of 3C experiments coming from different cell types, tissues and/or conditions or diseases. This has made necessary the development of software tools to compare them to shed light on the biological implications of genome structural changes.

The vast majority of available software to identify differential interactions are bin or pixel based, limiting the identification of chromatin loops (Table 2). Moreover, some of the tools presented in chromatin loop detection methods section present a version to assess differential interactions between experiments. Here are summarized the most used:

1. *HOMER* (Heinz *et al.*, 2010)

First, HOMER identifies the interactions in the first experiment as described in the gene dense regions does not present a canonical 3D structure. Then, quantifies the number of reads per interaction in the second experiment. In order to retrieve significant interactions independent statistics are computed for the second experiment on its background model and compared to the first experiment. The output contains the Z-score, logP and number of reads supporting the interaction between the experiment and the background.

2. *DiffHiC* (Lun & Smyth, 2015)

As previously explained, the main purpose of DiffHiC is to capture the differential significant interactions between experiments. In this case applies a QL F-test that yields a p-value per bin. It also corrects for multiple testing to control the FDR with the Benjamini-Hochberg method (Benjamini & Hochberg, 1995). These two values can be used to threshold the differential interactions.

	<i>Datatype</i>	<i>Normalization</i>	<i>Algorithm</i>
HOMER	Hi-C	✓	Background model
DiffHiC	Hi-C, Capture Hi-C*	✓	QLF-test
HiBrowse (Paulsen, Sandve, et al., 2014)	Hi-C, ChIA- PET	✓	Monte Carlo
HiCdat (Schmid, Grob, & Grossniklaus, 2015)	Hi-C, ChIP- Seq, RNA-seq, BS-seq, genome annotation	✓	Signed difference matrices
FIND (Djekidel, Chen, & Zhang, 2018)	Hi-C	✓	Spatial Poisson
Selfish (Ardakany, Ay, & Lonardi, 2019)	Hi-C		Self-similarity metric
Chicdiff (Cairns, Ochard, Malysheva, & Spivakov, 2019)	Promoter Capture Hi-C	✓	IHW
HiCcompare (Stansfield, Cresswell, Vladimirov, & Dozmorov, 2018)	Hi-C	✓	Z-score from loess
MultiHiCcompare (Stansfield, Cresswell, & Dozmorov, 2019)	Hi-C	✓	GLM
ACCOST (Cook, Hristov, Le Roch, Vert, & Noble, 2020)	Hi-C		Extended model used by DEseq**

Table 2. Comparison between features of some currently available software for differential chromatin contact data analysis. In this table are included the methods explained before as well as other available methods. All the software result in statistical information of the structural changes. differentia * might be implemented in future versions. ** (Anders & Huber, 2010) Independent Hypothesis Weighting (IHW); locally weighted linear regression (loess); general linear model (GLM).

Chapter I

Identification of chromatin loops from Hi-C interaction matrices
by CTCF-CTCF topology classification

Galan S, Serra F, Marti-Renom MA.
**Identification of chromatin loops from Hi-C
interactions matrices by CTCF-CTCF topology
classification.**

Chapter II

Quantitative comparison and automatic feature extraction for chromatin contact data

Galan S., Machnik N., Kruse K., Díaz N., Marti-Renom MA., Vaquerizas JM. Quantitative comparison and automatic feature extraction for chromatin contact data.

Discussion

Identification of chromatin loops from Hi-C interaction matrices by CTCF-CTCF topology classification

During the last years an outstanding number of studies proved the high relevance of the 3D organization of chromatin and its tight relationship with functional and regulatory processes. Specifically, chromatin loops are able to bring close into space regulatory units, forming transcription factories or hubs. The vast majority of these loops are flanked by two CTCF in a convergent oriented manner. However, not all the chromatin loops present a canonical CTCF binding and the same functional signature. What is more, not all convergent CTCFs are observed to form chromatin loops and same-oriented CTCFs can also be part of 3D structural rearrangements. Until now, the studies to unveil the role of features, which may play a role in chromatin organization, have been based on their average interaction frequency. The main limitation of piling up pairs of genomic coordinates, is the possibility to lose small clusters with a similar, and non-average, structure and function.

In Chapter I, we present an algorithm to deconvolve the structural signal of Hi-C experiments in the context of colocalizing DNA-binding proteins, called *metawaffle*. Specifically, we deconvolved the genomic average CTCF-CTCF interaction pattern within 45 kb and 1.5 Mb, due to its relevant role in formation and maintenance of chromatin loops. A total of 10 CTCFs subpopulations were identified after applying *metawaffle* and clustering. Thanks to this classification, we were able to check the distribution of various features, such as binding motif orientation, compartment type and genomic distance. First, we

observed that the genomic distance between pairs of CTCF is relevant to form a chromatin loop. It can be explained by the chromatin loop extrusion speed and the exchange dynamics of the SMC complex, and its requirement of ATP to translocate the DNA (Ganji et al., 2018; Terakawa et al., 2017). Moreover, polymer simulation modelling allowed the study of TAD borders, probing the interval of distance to insulate neighbouring TADs, comparable to our results (Fudenberg et al., 2016). As expected, the convergent oriented CTCFs, highly correlated with the loop structure pattern, whereas divergent oriented CTCFs presented the opposite trend. It has been observed that the directionality imposed by the DNA bending-initiated loop extrusion model produces a higher interaction frequency with the DNA on one side of it, which agrees with the low interacting frequency between divergent CTCF sites (Y. Guo et al., 2015; S. S. Rao et al., 2014). As for same-oriented CTCF pairs, which can be encountered during loop extrusion, nonetheless their anti-parallel orientation would be unfavourable for dimerization, and the extrusion will continue until finding a convergent site. Interestingly, the A/B compartment types were segregated thanks to the signal deconvolution, with an enrichment of A compartment in the loop structure clusters. We wanted to study in detail the chromatin states distribution through the CTCF-CTCF structural clusters. The promoter-enhancer state correlated with the A compartment type, being enriched in the canonical loop structure clusters, in line with the loop function, which is to bring together regulatory units for the proper gene expression. The heterochromatin state was found in between the two cluster extremes, which had the expected structural pattern according to the genomic distance between CTCF pairs. It may be caused by the role of active and bivalent chromatin into 3D chromatin

organization, consequently organizing the heterochromatin domains. The two clusters with depleted interactions were enriched by promoter-polycomb state, which may represent cell-specific “forbidden” loops. Surprisingly, the polycomb-polycomb state was enriched in those clusters with more interactions than expected. We hypothesize that they can contribute to gene silencing of the already described polycomb-dependent loops (Eagen, Aiden, & Kornberg, 2017), or simply appear by the overlapping of polycomb-polycomb and CTCF-CTCF driven interactions. Single-cell Hi-C will shed light into the dynamics of chromatin looping, allowing the comparison between pre-established versus de novo loops, and how are the loops maintained in non-expressing cells. Moreover, the study of other candidates, apart from CTCF and cohesin, and their interplay is essential to have a complete overview. Here has been inspected the interval of distance from 45 kb to 1.5 Mb in which CTCF it is known to play a relevant role. However, it can be applied to other distances, which would shed light on the complex regulation of the hierarchical structural layers.

The signal deconvolution of CTCFs pairs, also allowed us to obtain those specifically forming chromatin loops, which were used to train a CNN as a loop caller, called here *LOOPbit*. The lack of ground-truth-positive and ground-truth-negative controls, hinders the robust quantification of the specificity and the sensitivity of its performance. To overcome this limitation, *LOOPbit* was compared to a previously published benchmark in which large set of experiments by 6 different loop callers were analysed (Forcato et al., 2017). *LOOPbit* was applied to 33 Hi-C experiments at 5 kb resolution and to 4 experiments at 40 kb resolution. It presented the same trend to identify more *cis* interactions when having greater number of filtered reads. However, this trend was

more pronounced in *LOOPbit*, as the algorithm relies on the immediate surrounding of the loop for its identification. Nevertheless, this higher tendency was not observed in the Hi-C datasets at 40 kb resolution. The loops identified by *LOOPbit* presented a similar average distance between their anchors when compared to other methods. All the loop callers presented poor reproducibility between biological replicates, which normally are pooled together before the analysis to generate a unique sample with higher number of reads. Then, it was not surprising to obtain a similar reproducibility for our method. However, *LOOPbit* was more affected by low coverage experiments, as expected by the higher tendency of *cis*-interaction identification by greater number of filtered reads. Nonetheless, it is expected to perform better with higher coverage samples, as observed in the analysis of Hi-C experiments with targeted degraded CTCF (E. P. Nora et al., 2017), which presented a high reproducibility between untreated and recovered CTCF samples, and low reproducibility of these two samples with the CTCF degraded replicates. This overall poor reproducibility between replicates, may be explained by the fact that Hi-C experiments are an ensemble of cells in different cell states and cell cycle phases, thus not having identical chromatin contacts. Interestingly, anchors of the detected chromatin loops by *LOOPbit*, were highly enriched in promoter-enhancer state, and depleted in heterochromatin and a biologically less plausible state, being consistent with their biological description (Y. Guo et al., 2015; S. S. Rao et al., 2014). This can be a result of the training set used to build *LOOPbit*, which consisted in CTCF-CTCF interactions that were classified by *metawaffle* according to their structural pattern. Moreover, the chromatin loop anchors identified in (E. P. Nora et al., 2017), were enriched of CTCF ChIP-Seq peaks in the untreated and CTCF-

recovered samples, whereas it was not observed in the degraded CTCF experiments. Our results indicate that *LOOPbit* is more specific than other tools available as it captured higher proportion of functional loops.

Altogether, *LOOPbit* is comparable to other previously published methods for loop detection in terms of reproducibility and sensitivity, but results in a better performance on detecting functional loops, as it detected twice as many promoter-enhancer loops than most of the other callers. We can conclude, that there is no gold standard algorithm to identify chromatin loops. However, it is important to consider the usability, interoperability, stability of the implementation, and the computing resources required for each algorithm. Considering the fast pace on data production, it is key to provide computational tools as *LOOPbit* and *metawaffle* able to deal with high resolution datasets with reasonable amounts of computational resources with easily exchangeable data formats, following the FAIR principles (findable, accessible, interoperable and re-usable) (Marti-Renom et al., 2018).

Quantitative comparison and automatic feature extraction for chromatin contact data

The development of C-based experiments and their technical improvements, allowed their implementation as a lab routine technique, highly increasing the number of publicly available datasets. It opened the need of new algorithms able to compare chromatin structure between conditions and/or species. Nowadays, there are no available methods able to automatically compare and classify the identified 3D structural changes between contact matrices. To overcome this limitation, we developed *CHESS*, an automated, systematic and feature extraction tool, which correspond to the visual perception of structural differences. It is important to notice that *CHESS* is not limited to compare specific regions within a dataset, but genome-wide comparisons between samples, cell types, developmental stages and species. To note that other available methods for Hi-C comparison are bin-based, meaning that they rely on the comparison between individual bins from the contact matrices, or that will bias towards detecting local structural patterns such as loops or TADs. What makes *CHESS* unique is that it is structure-free and region-based. It will scan full genomic regions, and will compute their degree of similarity, from which the differential structural features would be extracted and classified. Furthermore, *CHESS* performance was faster and highly efficient with a small memory footprint, compared to *diffHiC*, *HOMER* and *ACCOST*. *CHESS* achieved 4, 7 and 320 times speed up, respectively, at 3 times lower memory consumption than the three other methods, making *CHESS* more usable without the need of an advanced computational infrastructure. Moreover, *CHESS* can be parallelizable,

allowing to do a countless number of comparisons if wanted, to address more complex biological questions.

CHESS also proved to be highly robust to experimental noise and usable even on shallow sequenced datasets. This makes *CHESS* widely applicable to a large amount of additional types of chromatin conformation capture datasets. For instance, *CHESS* was applied to mouse high-resolution Capture-C data (Despang et al., 2019) from the *Sox9-Kcnj2* locus, where they studied how genome-editing affected TAD function and gene expression. *CHESS* proved to identify the same structural rearrangements as in the experimental study with manually curated detection of structural differences. Nonetheless, it was able to spot subtle changes, such as the gain of a stripe in the *Inv-Intra* mutant, which had the *Kcnj2* TAD inverted without affecting the border between the two genes. This mutant showed a structural rearrangement without drastic effects on gene expression. Loop extrusion models have suggested that stripes/flames at CTCF sites are formed when an extruder finds a single barrier in one orientation but can continue tracking along the chromatin fiber in the other direction. Stripes have been identified in population-based methods, which accumulate snapshots of this dynamic process, giving rise to an increase of contact frequency coming out from the barrier site (Mirny, Imakaev, & Abdennur, 2019).

As stated before, *CHESS* is a structure-free method, which is very promising to *de novo* discovery of structurally similar regions between two experiments. For instance, *CHESS* was applied to mice and human “structurally syntenic” region pairs providing fundamental insights on the evolution of nuclear architecture and its 3D constraints. *CHESS* also identified different degrees of structural conservation across

mammalian evolution, indicating variable evolutionary 3D genomic rearrangements paces. The results, thus, suggests that evolutionary conserved regions will require a specific regulatory requirements, which will need the formation of large structured domains, while less conserved regions will not have the same demands and will depend less on such structures.

Additionally, several studies identified the association between structural variation and 3D rearrangements, being translated to gene expression mis-regulation. The identification of structural changes will help to shed light to evaluate the contribution of chromatin organization and disease-generating processes. To demonstrate how *CHESS* can contribute, it was applied to detect chromatin conformation alterations genome-wide in human B-cells from a diffuse large B-cell lymphoma (DLBCL) patient in comparison with normal B-cells, without the need of previous knowledge of the aberrations. Interestingly, the identified differential regions contained long non-coding RNAs (lncRNAs) and pseudogenes instead of protein coding genes. For instance, lncRNAs have been observed to participate in normal B-cell differentiation, and its deregulation can lead to B-cell malignancies such as DLBCL, being their discovery of prognostic value (Dahl, Kristensen, & Gronbaek, 2018). Moreover, from those differential regions, the gain of structural features, classified as TADs and loops, were systematically identified.

As mentioned before, the vast majority of available algorithms to detect structural differences between experiments are structure-biased, such as TAD and loop callers. It limits the study between structure and function relationship as they are focused on the role of mammalian TADs in

spatially regulating *cis*-regulatory elements, which is not clear to apply to all scenario and species. Nonetheless, as *CHESS* is a feature-free algorithm, it will not be biased and will provide all the set of structural changes regarding their structure. An integrative analysis of other patient-matched genome-wide datasets, such as RNA-seq, ATAC-seq, will be highly useful to determine the cause and consequence of the *CHESS* detected 3D structural rearrangements and its relation to disease. Furthermore, apart from the mentioned scenario in which *CHESS* can be applied, it might also be used to assess the reproducibility between biological replicates. However, this utility has not been inspected in detail in Chapter II.

In conclusion, *CHESS* is an unsupervised algorithm that can automatically retrieve a quantification of the structural similarity and the differential structural feature classification between two genomic regions from chromosome conformation data. Moreover, *CHESS* proved to be highly robust and tolerant to library size and noise level of datasets. Finally, *CHESS* can be applied to a wide range of biological scenarios, including the ranking of known region pairs by structural similarity, such as syntenic regions, and the discovery of structural changes between conditions, such as disease-associated structural variations for clinical applications.

Conclusions

From **Chapter I**, we can specifically infer:

- I. We developed *metawaffle*, an algorithm to deconvolve the structural patterns of DNA binding proteins within specific intervals of distances by combining Hi-C and ChIP-seq data.
- II. We revealed the presence of 10 CTCF-CTCF subpopulations within 45 kb and 1.5 Mb interval of distance, with different structural patterns and functional signatures.
- III. Each CTCF subpopulation presented an epigenetic enrichment, for instance the canonical loop was enriched in enhancer-promoter and enhancer mark, whereas polycomb-promoter mark was found in those regions with less interactions than expected.
- IV. The classification of *metawaffle* of the CTCF loops, was used to train a CNN, *LOOPbit*, which can be used to scan genome-wide and identify *cis*-interactions.
- V. *LOOPbit* proved useful to capture functional chromatin loops, which were enriched in enhancer-promoter and enhancer marks.
- VI. *Metawaffle* and *LOOPbit* are publicly-available and open-source python packages.

From **Chapter II**, we can specifically infer:

- I. We developed *CHESS*, an algorithm to systematically compare and classify differential structural features between contact matrices.
- II. *CHESS* is highly robust and consistent over different levels of experimental noise, resolutions and sequencing depths.
- III. *CHESS* is highly efficient and fast, with a very small memory footprint in comparison to other similar methods.
- IV. *CHESS* can be used to study multiple biological scenarios, by comparing samples, cell types, developmental stages and even species.
- V. *CHESS* can be used to analyze chromatin conformation capture datasets, as Capture-C, by extracting and classifying differential structural features, such as TADs and loops.
- VI. *CHESS* is a publicly-available and open-source python package.

VII. References

- Alexander, J. M., Guan, J., Huang, B., Lomvardas, S., & Weiner, O. D. (2018). Live-Cell Imaging Reveals Enhancer-dependent Sox2 Transcription in the Absence of Enhancer Proximity. *bioRxiv*, 409672. doi:10.1101/409672
- Anders, S., & Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biol*, 11(10), R106. doi:10.1186/gb-2010-11-10-r106
- Ardakany, A. R., Ay, F., & Lonardi, S. (2019). Selfish: discovery of differential chromatin interactions via a self-similarity measure. *Bioinformatics*, 35(14), I145-I153. doi:<https://doi.org/10.1093/bioinformatics/btz362>
- Ardakany, A. R., Gezer, H. T., Lonardi, S., & Ay, F. (2020). Mustache: Multi-scale Detection of Chromatin Loops from Hi-C and Micro-C Maps using Scale-Space Representation. *bioRxiv*. doi:<https://doi.org/10.1101/2020.02.24.963579>
- Ay, F., Bailey, T. L., & Noble, W. S. (2014). Statistical confidence estimation for Hi-C data reveals regulatory chromatin contacts. *Genome Res*, 24(6), 999-1011. doi:10.1101/gr.160374.113
- Barbieri, M., Chotalia, M., Fraser, J., Lavitas, L. M., Dostie, J., Pombo, A., & Nicodemi, M. (2012). Complexity of chromatin folding is captured by the strings and binders switch model. *Proc Natl Acad Sci U S A*, 109(40), 16173-16178. doi:10.1073/pnas.1204799109
- Barrington, C., Georgopoulou, D., Pezic, D., Varsally, W., Herrero, J., & Hadjur, S. (2019). Enhancer accessibility and CTCF occupancy underlie asymmetric TAD architecture and cell type specific genome topology. *Nat Commun*, 10(1), 2908. doi:10.1038/s41467-019-10725-9
- Bell, A. C., & Felsenfeld, G. (1999). Stopped at the border: boundaries and insulators. *Curr Opin Genet Dev*, 9(2), 191-198. doi:S0959-437X(99)80029-X [pii] 10.1016/S0959-437X(99)80029-X
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Royal Stat Soc B*, 57, 289-300.
- Bersaglieri, C., & Santoro, R. (2019). Genome Organization in and around the Nucleolus. *Cells*, 8(6). doi:10.3390/cells8060579
- Bintu, B., Mateo, L. J., Su, J. H., Sinnott-Armstrong, N. A., Parker, M., Kinrot, S., . . . Zhuang, X. (2018). Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science*, 362(6413). doi:10.1126/science.aau1783

- Boettiger, A. N., Bintu, B., Moffitt, J. R., Wang, S., Believeau, B. J., Fudenberg, G., . . . Zhuang, X. (2016). Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. *Nature*, *529*(7586), 418-422. doi:10.1038/nature16496
- Bolzer, A., Kreth, G., Solovei, I., Koehler, D., Saracoglu, K., Fauth, C., . . . Cremer, T. (2005). Three-dimensional maps of all chromosomes in human male fibroblast nuclei and prometaphase rosettes. *PLoS Biol*, *3*(5), e157. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=15839726
- Bonev, B., Mendelson Cohen, N., Szabo, Q., Fritsch, L., Papadopoulos, G. L., Lubling, Y., . . . Cavalli, G. (2017). Multiscale 3D Genome Rewiring during Mouse Neural Development. *Cell*, *171*(3), 557-572 e524. doi:10.1016/j.cell.2017.09.043
- Bovery, T. (1909). Die Blastomerenkerne von *Ascaris megalocephala* und die Theorie der Chromosomenindividualität. *Arch Zellforsch*, *3*, 181-268.
- Brackley, C. A., Johnson, J., Michieletto, D., Morozov, A. N., Nicodemi, M., Cook, P. R., & Marenduzzo, D. (2018). Extrusion without a motor: a new take on the loop extrusion model of genome organization. *Nucleus*, *9*(1), 95-103. doi:10.1080/19491034.2017.1421825
- Brooker, A. S., & Berkowitz, K. M. (2014). The roles of cohesins in mitosis, meiosis, and human health and disease. *Methods Mol Biol*, *1170*, 229-266. doi:10.1007/978-1-4939-0888-2_11
- Brown, K. E., Guest, S. S., Smale, S. T., Hahm, K., Merckenschlager, M., & Fisher, A. G. (1997). Association of transcriptionally silent genes with Ikaros complexes at centromeric heterochromatin. *Cell*, *91*(6), 845-854. doi:10.1016/s0092-8674(00)80472-9
- Buchwalter, A., Kaneshiro, J. M., & Hetzer, M. W. (2019). Coaching from the sidelines: the nuclear periphery in genome regulation. *Nat Rev Genet*, *20*(1), 39-50. doi:10.1038/s41576-018-0063-5
- Cairns, J., Freire-Pritchett, P., Wingett, S. W., Varnai, C., Dimond, A., Plagnol, V., . . . Spivakov, M. (2016). CHiCAGO: robust detection of DNA looping interactions in Capture Hi-C data. *Genome Biol*, *17*(1), 127. doi:10.1186/s13059-016-0992-2
- Cairns, J., Orchard, W. R., Malysheva, V., & Spivakov, M. (2019). Chicdiff: a computational pipeline for detecting differential chromosomal interactions in Capture Hi-C data. *Bioinformatics*, *35*(22), 4764-4766. doi:<https://doi.org/10.1093/bioinformatics/btz450>
- Chargaff, E., Magasanik, B., Vischer, E., Green, C., Doniger, R., & Elson, D. (1950). Nucleotide composition of pentose nucleic acids from yeast and mammalian tissues. *J Biol Chem*, *186*(1), 51-67. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/14778803>
- Chen, H., Levo, M., Barinov, L., Fujioka, M., Jaynes, J. B., & Gregor, T. (2018). Dynamic interplay between enhancer-promoter topology and

- gene activity. *Nat Genet*, 50(9), 1296-1303. doi:10.1038/s41588-018-0175-z
- Chothia, C., & Lesk, A. M. (1986). The relation between the divergence of sequence and structure in proteins. *EMBO J*, 5(4), 823-826. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/3709526>
- Clarkson, C. T., Deeks, E. A., Samarista, R., Mamayusupova, H., Zhurkin, V. B., & Teif, V. B. (2019). CTCF-dependent chromatin boundaries formed by asymmetric nucleosome arrays with decreased linker length. *Nucleic Acids Res*, 47(21), 11181-11196. doi:10.1093/nar/gkz908
- Cook, K. B., Hristov, B. H., Le Roch, K. G., Vert, J. P., & Noble, W. S. (2020). Measuring significant changes in chromatin conformation with ACCOST. *Nucleic Acids Res*, 48(5), 2303-2311. doi:10.1093/nar/gkaa069
- Cremer, T., Cremer, C., Baumann, H., Luedtke, E. K., Sperling, K., Teuber, V., & Zorn, C. (1982). Rabl's model of the interphase chromosome arrangement tested in Chinese hamster cells by premature chromosome condensation and laser-UV-microbeam experiments. *Hum Genet*, 60(1), 46-56. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/7076247>
- Cremer, T., Cremer, C., Schneider, T., Baumann, H., Hens, L., & Kirsch-Volders, M. (1982). Analysis of chromosome positions in the interphase nucleus of Chinese hamster cells by laser-UV-microirradiation experiments. *Hum Genet*, 62(3), 201-209. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/7169211>
- Croft, J. A., Bridger, J. M., Boyle, S., Perry, P., Teague, P., & Bickmore, W. A. (1999). Differences in the localization and morphology of chromosomes in the human nucleus. *J Cell Biol*, 145(6), 1119-1131. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/10366586>
- Cuddapah, S., Jothi, R., Schones, D. E., Roh, T. Y., Cui, K., & Zhao, K. (2009). Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains. *Genome Res*, 19(1), 24-32. doi:gr.082800.108 [pii] 10.1101/gr.082800.108
- Dahl, M., Kristensen, L. S., & Gronbaek, K. (2018). Long Non-Coding RNAs Guide the Fine-Tuning of Gene Regulation in B-Cell Development and Malignancy. *Int J Mol Sci*, 19(9). doi:10.3390/ijms19092475
- Dahm, R. (2005). Friedrich Miescher and the discovery of DNA. *Dev Biol*, 278(2), 274-288. doi:10.1016/j.ydbio.2004.11.028
- Davidson, I. F., Goetz, D., Zaczek, M. P., Molodtsov, M. I., Huis In 't Veld, P. J., Weissmann, F., . . . Peters, J. M. (2016). Rapid movement and transcriptional re-localization of human cohesin on DNA. *EMBO J*, 35(24), 2671-2685. doi:10.15252/emboj.201695402

- de Leeuw, R., Gruenbaum, Y., & Medalia, O. (2018). Nuclear Lamins: Thin Filaments with Major Functions. *Trends Cell Biol*, 28(1), 34-45. doi:10.1016/j.tcb.2017.08.004
- Dekker, J. (2006). The three 'C' s of chromosome conformation capture: controls, controls, controls. *Nat Methods*, 3(1), 17-21. doi:nmeth823 [pii] 10.1038/nmeth823
- Dekker, J., & Heard, E. (2015). Structural and functional diversity of Topologically Associating Domains. *FEBS Lett*, 589(20 Pt A), 2877-2884. doi:10.1016/j.febslet.2015.08.044
- Dekker, J., Rippe, K., Dekker, M., & Kleckner, N. (2002). Capturing chromosome conformation. *Science*, 295(5558), 1306-1311. doi:10.1126/science.1067799 295/5558/1306 [pii]
- Denker, A., & de Laat, W. (2016). The second decade of 3C technologies: detailed insights into nuclear organization. *Genes Dev*, 30(12), 1357-1382. doi:10.1101/gad.281964.116
- Despang, A., Schopflin, R., Franke, M., Ali, S., Jerkovic, I., Paliou, C., . . . Ibrahim, D. M. (2019). Functional dissection of the Sox9-Kcnj2 locus identifies nonessential and instructive roles of TAD architecture. *Nat Genet*, 51(8), 1263-1271. doi:10.1038/s41588-019-0466-z
- Devos, D. P., Graf, R., & Field, M. C. (2014). Evolution of the nucleus. *Curr Opin Cell Biol*, 28, 8-15. doi:10.1016/jceb.2014.01.004
- Dixon, J. R., Jung, I., Selvaraj, S., Shen, Y., Antosiewicz-Bourget, J. E., Lee, A. Y., . . . Ren, B. (2015). Chromatin architecture reorganization during stem cell differentiation. *Nature*, 518(7539), 331-336. doi:10.1038/nature14222
- Dixon, J. R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., . . . Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, 485(7398), 376-380. doi:10.1038/nature11082 nature11082 [pii]
- Dixon, J. R., Xu, J., Dileep, V., Zhan, Y., Song, F., Le, V. T., . . . Yue, F. (2018). Integrative detection and analysis of structural variation in cancer genomes. *Nat Genet*, 50(10), 1388-1398. doi:10.1038/s41588-018-0195-8
- Djekidel, M. N., Chen, Y., & Zhang, M. Q. (2018). FIND: differential chromatin INteractions Detection using a spatial Poisson process. *Genome Res*, 28(3), 412-422. doi:10.1101/gr.212241.116
- Dostie, J., Richmond, T. A., Arnaout, R. A., Selzer, R. R., Lee, W. L., Honan, T. A., . . . Dekker, J. (2006). Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements. *Genome Res*, 16(10), 1299-1309. doi:gr.5571506 [pii]

10.1101/gr.5571506

- Du, Z., Zheng, H., Huang, B., Ma, R., Wu, J., Zhang, X., . . . Xie, W. (2017). Allelic reprogramming of 3D chromatin architecture during early mammalian development. *Nature*, *547*(7662), 232-235. doi:10.1038/nature23263
- Durand, N. C., Shamim, M. S., Machol, I., Rao, S. S., Huntley, M. H., Lander, E. S., & Aiden, E. L. (2016). Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Syst*, *3*(1), 95-98. doi:10.1016/j.cels.2016.07.002
- Eagen, K. P., Aiden, E. L., & Kornberg, R. D. (2017). Polycomb-mediated chromatin loops revealed by a subkilobase-resolution chromatin interaction map. *Proc Natl Acad Sci U S A*, *114*(33), 8764-8769. doi:10.1073/pnas.1701291114
- El-Kady, A., & Klenova, E. (2005). Regulation of the transcription factor, CTCF, by phosphorylation with protein kinase CK2. *FEBS Lett*, *579*(6), 1424-1434. doi:10.1016/j.febslet.2005.01.044
- Ferraro, T., Esposito, E., Mancini, L., Ng, S., Lucas, T., Coppey, M., . . . Lagha, M. (2016). Transcriptional Memory in the Drosophila Embryo. *Curr Biol*, *26*(2), 212-218. doi:10.1016/j.cub.2015.11.058
- Filippova, G. N., Fagerlie, S., Klenova, E. M., Myers, C., Dehner, Y., Goodwin, G., . . . Lobanenkoy, V. V. (1996). An exceptionally conserved transcriptional repressor, CTCF, employs different combinations of zinc fingers to bind diverged promoter sequences of avian and mammalian c-myc oncogenes. *Mol Cell Biol*, *16*(6), 2802-2813. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=8649389
- Finn, E. H., Pegoraro, G., Brandao, H. B., Valton, A. L., Oomen, M. E., Dekker, J., . . . Misteli, T. (2019). Extensive Heterogeneity and Intrinsic Variation in Spatial Genome Organization. *Cell*, *176*(6), 1502-1515 e1510. doi:10.1016/j.cell.2019.01.020
- Flavahan, W. A., Drier, Y., Liao, B. B., Gillespie, S. M., Venteicher, A. S., Stemmer-Rachamimov, A. O., . . . Bernstein, B. E. (2016). Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature*, *529*(7584), 110-114. doi:10.1038/nature16490
- Forcato, M., Nicoletti, C., Pal, K., Livi, C. M., Ferrari, F., & Bicciato, S. (2017). Comparison of computational methods for Hi-C data analysis. *Nat Methods*, *14*(7), 679-685. doi:10.1038/nmeth.4325
- Forsberg, F., Brunet, A., Ali, T. M. L., & Collas, P. (2019). Interplay of lamin A and lamin B LADs on the radial positioning of chromatin. *Nucleus*, *10*(1), 7-20. doi:10.1080/19491034.2019.1570810
- Franke, M., Ibrahim, D. M., Andrey, G., Schwarzer, W., Heinrich, V., Schopflin, R., . . . Mundlos, S. (2016). Formation of new chromatin domains determines pathogenicity of genomic duplications. *Nature*, *538*(7624), 265-269. doi:10.1038/nature19800

- Franklin, R. E., & Gosling, R. G. (1953). Evidence for 2-chain helix in crystalline structure of sodium deoxyribonucleate. *Nature*, *172*(4369), 156-157. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=13072614
- Fu, Y., Sinha, M., Peterson, C. L., & Weng, Z. (2008). The insulator binding protein CTCF positions 20 nucleosomes around its binding sites across the human genome. *PLoS Genet*, *4*(7), e1000138. doi:10.1371/journal.pgen.1000138
- Fudenberg, G., Imakaev, M., Lu, C., Goloborodko, A., Abdennur, N., & Mirny, L. A. (2016). Formation of Chromosomal Domains by Loop Extrusion. *Cell Rep*, *15*(9), 2038-2049. doi:10.1016/j.celrep.2016.04.085
- Fullwood, M. J., Liu, M. H., Pan, Y. F., Liu, J., Xu, H., Mohamed, Y. B., . . . Ruan, Y. (2009). An oestrogen-receptor-alpha-bound human chromatin interactome. *Nature*, *462*(7269), 58-64. doi:nature08497 [pii] 10.1038/nature08497
- Ganji, M., Shaltiel, I. A., Bisht, S., Kim, E., Kalichava, A., Haering, C. H., & Dekker, C. (2018). Real-time imaging of DNA loop extrusion by condensin. *Science*, *360*(6384), 102-105. doi:10.1126/science.aar7831
- Gassler, J., Brandao, H. B., Imakaev, M., Flyamer, I. M., Ladstätter, S., Bickmore, W. A., . . . Tachibana, K. (2017). A mechanism of cohesin-dependent loop extrusion organizes zygotic genome architecture. *EMBO J*, *36*(24), 3600-3618. doi:10.15252/embj.201798083
- Gesson, K., Rescheneder, P., Skoruppa, M. P., von Haeseler, A., Dechat, T., & Foisner, R. (2016). A-type lamins bind both hetero- and euchromatin, the latter being regulated by lamina-associated polypeptide 2 alpha. *Genome Res*, *26*(4), 462-473. doi:10.1101/gr.196220.115
- Gibcus, J. H., & Dekker, J. (2013). The hierarchy of the 3D genome. *Mol Cell*, *49*(5), 773-782. doi:10.1016/j.molcel.2013.02.011
- Gibson, B. A., Doolittle, L. K., Schneider, M. W. G., Jensen, L. E., Gamarra, N., Henry, L., . . . Rosen, M. K. (2019). Organization of Chromatin by Intrinsic and Regulated Phase Separation. *Cell*, *179*(2), 470-484 e421. doi:10.1016/j.cell.2019.08.037
- Giorgetti, L., Galupa, R., Nora, E. P., Piolot, T., Lam, F., Dekker, J., . . . Heard, E. (2014). Predictive polymer modeling reveals coupled fluctuations in chromosome conformation and transcription. *Cell*, *157*(4), 950-963. doi:10.1016/j.cell.2014.03.025
- Guelen, L., Pagie, L., Brasset, E., Meuleman, W., Faza, M. B., Talhout, W., . . . van Steensel, B. (2008). Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature*, *453*(7197), 948-951. doi:nature06947 [pii]

10.1038/nature06947

- Guo, C., Yoon, H. S., Franklin, A., Jain, S., Ebert, A., Cheng, H. L., . . . Alt, F. W. (2011). CTCF-binding elements mediate control of V(D)J recombination. *Nature*, *477*(7365), 424-430.
doi:10.1038/nature10495
- Guo, Y., Xu, Q., Canzio, D., Shou, J., Li, J., Gorkin, D. U., . . . Wu, Q. (2015). CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. *Cell*, *162*(4), 900-910.
doi:10.1016/j.cell.2015.07.038
- Haarhuis, J. H. I., van der Weide, R. H., Blomen, V. A., Yanez-Cuna, J. O., Amendola, M., van Ruiten, M. S., . . . Rowland, B. D. (2017). The Cohesin Release Factor WAPL Restricts Chromatin Loop Extension. *Cell*, *169*(4), 693-707 e614. doi:10.1016/j.cell.2017.04.013
- Heger, P., Marin, B., Bartkuhn, M., Schierenberg, E., & Wiehe, T. (2012). The chromatin insulator CTCF and the emergence of metazoan diversity. *Proc Natl Acad Sci U S A*, *109*(43), 17507-17512.
doi:10.1073/pnas.1111941109
- Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y. C., Laslo, P., . . . Glass, C. K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell*, *38*(4), 576-589.
doi:10.1016/j.molcel.2010.05.004
- Heitz, E. (1928). Das Heterochromatin der Moose. *Jahrb Wiss Botanik*, *69*, 762-818.
- Hnisz, D., Weintraub, A. S., Day, D. S., Valton, A. L., Bak, R. O., Li, C. H., . . . Young, R. A. (2016). Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science*, *351*(6280), 1454-1458.
doi:10.1126/science.aad9024
- Holwerda, S. J., & de Laat, W. (2013). CTCF: the protein, the binding partners, the binding sites and their chromatin loops. *Philos Trans R Soc Lond B Biol Sci*, *368*(1620), 20120369. doi:10.1098/rstb.2012.0369
rstb.2012.0369 [pii]
- Horowitz-Scherer, R. A., & Woodcock, C. L. (2006). Organization of interphase chromatin. *Chromosoma*, *115*(1), 1-14.
doi:10.1007/s00412-005-0035-3
- Horta, A., Monahan, K., Bashkirova, E., & Lomvardas, S. (2019). Cell type-specific interchromosomal interactions as a mechanism for transcriptional diversity. *bioRxiv*.
- Hsieh, T. H., Weiner, A., Lajoie, B., Dekker, J., Friedman, N., & Rando, O. J. (2015). Mapping nucleosome resolution chromosome folding in yeast by Micro-C. *Cell*, *162*, 108-119.
- Hsieh, T. S., Cattoglio, C., Slobodyanyuk, E., Hansen, A. S., Rando, O. J., Tjian, R., & Darzacq, X. (2020). Resolving the 3D Landscape of Transcription-Linked Mammalian Chromatin Folding. *Mol Cell*.
doi:10.1016/j.molcel.2020.03.002

- Hug, C. B., Grimaldi, A. G., Kruse, K., & Vaquerizas, J. M. (2017). Chromatin Architecture Emerges during Zygotic Genome Activation Independent of Transcription. *Cell*, *169*(2), 216-228 e219. doi:10.1016/j.cell.2017.03.024
- Hughes, J. R., Roberts, N., McGowan, S., Hay, D., Giannoulatou, E., Lynch, M., . . . Higgs, D. R. (2014). Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment. *Nat Genet*, *46*(2), 205-212. doi:10.1038/ng.2871
- Hwang, Y. C., Lin, C. F., Valladares, O., Malamon, J., Kuksa, P. P., Zheng, Q., . . . Wang, L. S. (2015). HIPPIE: a high-throughput identification pipeline for promoter interacting enhancer elements. *Bioinformatics*, *31*(8), 1290-1292. doi:10.1093/bioinformatics/btu801
- Imakaev, M., Fudenberg, G., McCord, R. P., Naumova, N., Goloborodko, A., Lajoie, B. R., . . . Mirny, L. A. (2012). Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nat Methods*, *9*(10), 999-1003. doi:10.1038/nmeth.2148
- Ishihara, K., Oshimura, M., & Nakao, M. (2006). CTCF-dependent chromatin insulator is linked to epigenetic remodeling. *Mol Cell*, *23*(5), 733-742. doi:10.1016/j.molcel.2006.08.008
- Jin, F., Li, Y., Dixon, J. R., Selvaraj, S., Ye, Z., Lee, A. Y., . . . Ren, B. (2013). A high-resolution map of the three-dimensional chromatin interactome in human cells. *Nature*, *503*(7475), 290-294. doi:10.1038/nature12644
- Jonkers, I., & Lis, J. T. (2015). Getting up to speed with transcription elongation by RNA polymerase II. *Nat Rev Mol Cell Biol*, *16*(3), 167-177. doi:10.1038/nrm3953
- Jost, D., Carrivain, P., Cavalli, G., & Vaillant, C. (2014). Modeling epigenome folding: formation and dynamics of topologically associated chromatin domains. *Nucleic Acids Res*, *42*(15), 9553-9561. doi:10.1093/nar/gku698
- Katainen, R., Dave, K., Pitkanen, E., Palin, K., Kivioja, T., Valimaki, N., . . . Aaltonen, L. A. (2015). CTCF/cohesin-binding sites are frequently mutated in cancer. *Nat Genet*, *47*(7), 818-821. doi:10.1038/ng.3335
- Kaul, A., Bhattacharyya, S., & Ay, F. (2020). Identifying statistically significant chromatin contacts from Hi-C data with FitHiC2. *Nat Protoc*. doi:10.1038/s41596-019-0273-0
- Ke, Y., Xu, Y., Chen, X., Feng, S., Liu, Z., Sun, Y., . . . Liu, J. (2017). 3D Chromatin Structures of Mature Gametes and Structural Reprogramming during Mammalian Embryogenesis. *Cell*, *170*(2), 367-381 e320. doi:10.1016/j.cell.2017.06.029
- Kind, J., Pagie, L., Ortobozkoyun, H., Boyle, S., de Vries, S. S., Janssen, H., . . . van Steensel, B. (2013). Single-cell dynamics of genome-nuclear lamina interactions. *Cell*, *153*(1), 178-192. doi:10.1016/j.cell.2013.02.028
- S0092-8674(13)00217-1 [pii]

- Kitchen, N. S., & Schoenherr, C. J. (2010). Sumoylation modulates a domain in CTCF that activates transcription and decondenses chromatin. *J Cell Biochem*, 111(3), 665-675. doi:10.1002/jcb.22751
- Knight, P. A., & Ruiz, D. (2013). A fast algorithm for matrix balancing. *IMA Journal of Numerical Analysis*, 33(3), 1029-1047.
- Kraft, K., Magg, A., Heinrich, V., Riemenschneider, C., Schopflin, R., Markowski, J., . . . Mundlos, S. (2019). Serial genomic inversions induce tissue-specific architectural stripes, gene misexpression and congenital malformations. *Nat Cell Biol*, 21(3), 305-310. doi:10.1038/s41556-019-0273-x
- Kundu, S., Ji, F., Sunwoo, H., Jain, G., Lee, J. T., Sadreyev, R. I., . . . Kingston, R. E. (2018). Polycomb Repressive Complex 1 Generates Discrete Compacted Domains that Change during Differentiation. *Mol Cell*, 71(1), 191. doi:10.1016/j.molcel.2018.06.022
- Le Dily, F., Baù, D., Pohl, A., Vicent, G. P., Serra, F., Soronellas, D., . . . Beato, M. (2014). Distinct structural transitions of chromatin topological domains correlate with coordinated hormone-induced gene regulation. *Genes Dev*, 28(19), 2151-2162. doi:10.1101/gad.241422.114
- Le Dily, F., Serra, F., & Marti-Renom, M. A. (2017). 3D modeling of chromatin structure: is there a way to integrate and reconcile single cell and population experimental data? *Wiley Interdisciplinary Reviews: Computational Molecular Science*, e1308-n/a. doi:10.1002/wcms.1308
- Levene, P. A. (1919). The structure of yeast nucleic acid: IV. Ammonia hydrolysis. *J. Biol. Chem.*, 40, 415-424.
- Li, Y., Haarhuis, J. H. I., Sedenò Cacciatore, A., Oldenkamp, R., van Ruiten, M. S., Willems, L., . . . Panne, D. (2020). The structural basis for cohesin-CTCF-anchored loops. *Nature*. doi:10.1038/s41586-019-1910-z
- Lichter, P., Cremer, T., Borden, J., Manuelidis, L., & Ward, D. C. (1988). Delineation of individual human chromosomes in metaphase and interphase cells by in situ suppression hybridization using recombinant DNA libraries. *Hum Genet*, 80(3), 224-234. doi:10.1007/bf01790090
- Lieberman-Aiden, E., van Berkum, N. L., Williams, L., Imakaev, M., Ragozy, T., Telling, A., . . . Dekker, J. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, 326(5950), 289-293. doi:10.1126/science.1181369
- Lun, A. T., & Smyth, G. K. (2015). diffHic: a Bioconductor package to detect differential genomic interactions in Hi-C data. *BMC Bioinformatics*, 16, 258. doi:10.1186/s12859-015-0683-0
- Lund, E., Oldenburg, A. R., Delbarre, E., Freberg, C. T., Duband-Goulet, I., Eskeland, R., . . . Collas, P. (2013). Lamin A/C-promoter interactions specify chromatin state-dependent transcription

- outcomes. *Genome Res*, 23(10), 1580-1589.
doi:10.1101/gr.159400.113
- Lund, E. G., Duband-Goulet, I., Oldenburg, A., Buendia, B., & Collas, P. (2015). Distinct features of lamin A-interacting chromatin domains mapped by ChIP-sequencing from sonicated or micrococcal nuclease-digested chromatin. *Nucleus*, 6(1), 30-39.
doi:10.4161/19491034.2014.990855
- Lupianez, D. G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., . . . Mundlos, S. (2015). Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell*, 161(5), 1012-1025. doi:10.1016/j.cell.2015.04.004
- Lutz, M., Burke, L. J., LeFevre, P., Myers, F. A., Thorne, A. W., Crane-Robinson, C., . . . Renkawitz, R. (2003). Thyroid hormone-regulated enhancer blocking: cooperation of CTCF and thyroid hormone receptor. *EMBO J*, 22(7), 1579-1587. doi:10.1093/emboj/cdg147
- Marti-Renom, M. A., Almouzni, G., Bickmore, W. A., Bystricky, K., Cavalli, G., Fraser, P., . . . Torres-Padilla, M. E. (2018). Challenges and guidelines toward 4D nucleome data and model standards. *Nat Genet*, 50(10), 1352-1358. doi:10.1038/s41588-018-0236-3
- Martin, W. F., Garg, S., & Zimorski, V. (2015). Endosymbiotic theories for eukaryote origin. *Philos Trans R Soc Lond B Biol Sci*, 370(1678), 20140330. doi:10.1098/rstb.2014.0330
- 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*. doi:10.1038/s41586-019-1035-4
- Mawhinney, M. T., Liu, R., Lu, F., Maksimoska, J., Damico, K., Marmorstein, R., . . . Urbanc, B. (2018). CTCF-Induced Circular DNA Complexes Observed by Atomic Force Microscopy. *J Mol Biol*, 430(6), 759-776. doi:10.1016/j.jmb.2018.01.012
- Mifsud, B., Martincorena, I., Darbo, E., Sugar, R., Schoenfelder, S., Fraser, P., & Luscombe, N. M. (2017). GOTHiC, a probabilistic model to resolve complex biases and to identify real interactions in Hi-C data. *PLoS ONE*, 12(4), e0174744. doi:10.1371/journal.pone.0174744
- Mirny, L. A., Imakaev, M., & Abdennur, N. (2019). Two major mechanisms of chromosome organization. *Curr Opin Cell Biol*, 58, 142-152. doi:10.1016/j.ceb.2019.05.001
- Mumbach, M. R., Rubin, A. J., Flynn, R. A., Dai, C., Khavari, P. A., Greenleaf, W. J., & Chang, H. Y. (2016). HiChIP: efficient and sensitive analysis of protein-directed genome architecture. *Nat Methods*, 13(11), 919-922. doi:10.1038/nmeth.3999
- Murayama, Y., Samora, C. P., Kurokawa, Y., Iwasaki, H., & Uhlmann, F. (2018). Establishment of DNA-DNA Interactions by the Cohesin Ring. *Cell*, 172(3), 465-477 e415. doi:10.1016/j.cell.2017.12.021

- Nagano, T., Varnai, C., Schoenfelder, S., Javierre, B. M., Wingett, S. W., & Fraser, P. (2015). Comparison of Hi-C results using in-solution versus in-nucleus ligation. *Genome Biol*, *16*, 175. doi:10.1186/s13059-015-0753-7
- Nagy, G., Czipa, E., Steiner, L., Nagy, T., Pongor, S., Nagy, L., & Barta, E. (2016). Motif oriented high-resolution analysis of ChIP-seq data reveals the topological order of CTCF and cohesin proteins on DNA. *BMC Genomics*, *17*(1), 637. doi:10.1186/s12864-016-2940-7
- Nakahashi, H., Kieffer Kwon, K. R., Resch, W., Vian, L., Dose, M., Stavreva, D., . . . Casellas, R. (2013). A genome-wide map of CTCF multivalency redefines the CTCF code. *Cell Rep*, *3*(5), 1678-1689. doi:10.1016/j.celrep.2013.04.024
- Nemeth, A., Conesa, A., Santoyo-Lopez, J., Medina, I., Montaner, D., Peterfia, B., . . . Langst, G. (2010). Initial genomics of the human nucleolus. *PLoS Genet*, *6*(3), e1000889. doi:10.1371/journal.pgen.1000889
- Nichols, M. H., & Corces, V. G. (2015). A CTCF Code for 3D Genome Architecture. *Cell*, *162*(4), 703-705. doi:10.1016/j.cell.2015.07.053
- Nora, E. P., Goloborodko, A., Valton, A.-L., Gibcus, J. H., Uebersohn, A., Abdennur, N., . . . Bruneau, B. (2017). Targeted degradation of CTCF decouples local insulation of chromosome domains from higher-order genomic compartmentalization. *bioRxiv*. doi:10.1101/095802
- Nora, E. P., Goloborodko, A., Valton, A. L., Gibcus, J. H., Uebersohn, A., Abdennur, N., . . . Bruneau, B. G. (2017). Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell*, *169*(5), 930-944 e922. doi:10.1016/j.cell.2017.05.004
- Nora, E. P., Lajoie, B. R., Schulz, E. G., Giorgetti, L., Okamoto, I., Servant, N., . . . Heard, E. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature*, *485*(7398), 381-385. doi:10.1038/nature11049
- Nozaki, T., Imai, R., Tanbo, M., Nagashima, R., Tamura, S., Tani, T., . . . Maeshima, K. (2017). Dynamic Organization of Chromatin Domains Revealed by Super-Resolution Live-Cell Imaging. *Mol Cell*, *67*(2), 282-293 e287. doi:10.1016/j.molcel.2017.06.018
- Ocampo-Hafalla, M., Munoz, S., Samora, C. P., & Uhlmann, F. (2016). Evidence for cohesin sliding along budding yeast chromosomes. *Open Biol*, *6*(6). doi:10.1098/rsob.150178
- Paredes, S. H., Melgar, M. F., & Sethupathy, P. (2013). Promoter-proximal CCCTC-factor binding is associated with an increase in the transcriptional pausing index. *Bioinformatics*, *29*(12), 1485-1487. doi:10.1093/bioinformatics/bts596
- Parelho, V., Hadjir, S., Spivakov, M., Leleu, M., Sauer, S., Gregson, H. C., . . . Merckenschlager, M. (2008). Cohesins functionally associate with

- CTCF on mammalian chromosome arms. *Cell*, 132(3), 422-433.
doi:10.1016/j.cell.2008.01.011
- Pascual-Reguant, L., Blanco, E., Galan, S., Le Dily, F., Cuartero, Y., Serra-Bardenys, G., . . . Peiro, S. (2018). Lamin B1 mapping reveals the existence of dynamic and functional euchromatin lamin B1 domains. *Nat Commun*, 9(1), 3420. doi:10.1038/s41467-018-05912-z
- Paulsen, J., Liyakat Ali, T. M., Nekrasov, M., Delbarre, E., Baudement, M. O., Kurscheid, S., . . . Collas, P. (2019). Long-range interactions between topologically associating domains shape the four-dimensional genome during differentiation. *Nat Genet*, 51(5), 835-843. doi:10.1038/s41588-019-0392-0
- Paulsen, J., Rodland, E. A., Holden, L., Holden, M., & Hovig, E. (2014). A statistical model of ChIA-PET data for accurate detection of chromatin 3D interactions. *Nucleic Acids Res*, 42(18), e143. doi:10.1093/nar/gku738
- Paulsen, J., Sandve, G. K., Gundersen, S., Lien, T. G., Trengereid, K., & Hovig, E. (2014). HiBrowse: multi-purpose statistical analysis of genome-wide chromatin 3D organization. *Bioinformatics*, 30(11), 1620-1622. doi:10.1093/bioinformatics/btu082
- Pavlaki, I., Docquier, F., Chernukhin, I., Kita, G., Gretton, S., Clarkson, C. T., . . . Klenova, E. (2018). Poly(ADP-ribosylation) associated changes in CTCF-chromatin binding and gene expression in breast cells. *Biochim Biophys Acta Gene Regul Mech*, 1861(8), 718-730. doi:10.1016/j.bbagr.2018.06.010
- Phanstiel, D. H., Van Bortle, K., Spacek, D., Hess, G. T., Shamim, M. S., Machol, I., . . . Snyder, M. P. (2017). Static and Dynamic DNA Loops form AP-1-Bound Activation Hubs during Macrophage Development. *Mol Cell*, 67(6), 1037-1048 e1036. doi:10.1016/j.molcel.2017.08.006
- Ptashne, M. (1986). Gene regulation by proteins acting nearby and at a distance. *Nature*, 322(6081), 697-701. doi:10.1038/322697a0
- Racko, D., Benedetti, F., Dorier, J., & Stasiak, A. (2018). Transcription-induced supercoiling as the driving force of chromatin loop extrusion during formation of TADs in interphase chromosomes. *Nucleic Acids Res*, 46(4), 1648-1660. doi:10.1093/nar/gkx1123
- Rao, S. S., Huntley, M. H., Durand, N. C., Stamenova, E. K., Bochkov, I. D., Robinson, J. T., . . . Aiden, E. L. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*, 159(7), 1665-1680. doi:10.1016/j.cell.2014.11.021
- Rao, S. S. P., Huang, S. C., Glenn St Hilaire, B., Engreitz, J. M., Perez, E. M., Kieffer-Kwon, K. R., . . . Aiden, E. L. (2017). Cohesin Loss Eliminates All Loop Domains. *Cell*, 171(2), 305-320 e324. doi:10.1016/j.cell.2017.09.026
- Ren, G., Jin, W., Cui, K., Rodrigez, J., Hu, G., Zhang, Z., . . . Zhao, K. (2017). CTCF-Mediated Enhancer-Promoter Interaction Is a Critical

- Regulator of Cell-to-Cell Variation of Gene Expression. *Mol Cell*, 67(6), 1049-1058 e1046. doi:10.1016/j.molcel.2017.08.026
- Rowley, M. J., & Corces, V. G. (2018). Organizational principles of 3D genome architecture. *Nat Rev Genet*, 19(12), 789-800. doi:10.1038/s41576-018-0060-8
- Rowley, M. J., Nichols, M. H., Lyu, X., Ando-Kuri, M., Rivera, I. S. M., Hermetz, K., . . . Corces, V. G. (2017). Evolutionarily Conserved Principles Predict 3D Chromatin Organization. *Mol Cell*, 67(5), 837-852 e837. doi:10.1016/j.molcel.2017.07.022
- Rubio, E. D., Reiss, D. J., Welch, P. L., Distche, C. M., Filippova, G. N., Baliga, N. S., . . . Krumm, A. (2008). CTCF physically links cohesin to chromatin. *Proc Natl Acad Sci U S A*, 105(24), 8309-8314. doi:0801273105 [pii] 10.1073/pnas.0801273105
- Sagan, L. (1967). On the origin of mitosing cells. *J Theor Biol*, 14(3), 255-274. doi:10.1016/0022-5193(67)90079-3
- Sanborn, A. L., Rao, S. S., Huang, S. C., Durand, N. C., Huntley, M. H., Jewett, A. I., . . . Aiden, E. L. (2015). Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc Natl Acad Sci U S A*, 112(47), E6456-6465. doi:10.1073/pnas.1518552112
- Schmid, M. W., Grob, S., & Grossniklaus, U. (2015). HiCdat: a fast and easy-to-use Hi-C data analysis tool. *BMC Bioinformatics*, 16, 277. doi:10.1186/s12859-015-0678-x
- Schmidt, D., Schwalie, P. C., Wilson, M. D., Ballester, B., Goncalves, A., Kutter, C., . . . Odom, D. T. (2012). Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. *Cell*, 148(1-2), 335-348. doi:10.1016/j.cell.2011.11.058
- Schwarzer, W., Abdennur, N., Goloborodko, A., Pekowska, A., Fudenberg, G., Loe-Mie, Y., . . . Spitz, F. (2017). Two independent modes of chromatin organization revealed by cohesin removal. *Nature*, 551(7678), 51-56. doi:10.1038/nature24281
- Shukla, S., Kavak, E., Gregory, M., Imashimizu, M., Shutinoski, B., Kashlev, M., . . . Oberdoerffer, S. (2011). CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature*, 479(7371), 74-79. doi:10.1038/nature10442
- Simonis, M., Klous, P., Splinter, E., Moshkin, Y., Willemsen, R., de Wit, E., . . . de Laat, W. (2006). Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nat Genet*, 38(11), 1348-1354. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=17033623
- Solovei, I., Wang, A. S., Thanisch, K., Schmidt, C. S., Krebs, S., Zwerger, M., . . . Joffe, B. (2013). LBR and lamin A/C sequentially tether

- peripheral heterochromatin and inversely regulate differentiation. *Cell*, 152(3), 584-598. doi:10.1016/j.cell.2013.01.009
- Spielmann, M., Lupianez, D. G., & Mundlos, S. (2018). Structural variation in the 3D genome. *Nat Rev Genet*, 19(7), 453-467. doi:10.1038/s41576-018-0007-0
- Spill, Y. G., Castillo, D., Vidal, E., & Marti-Renom, M. A. (2019). Binless normalization of Hi-C data provides significant interaction and difference detection independent of resolution. *Nat Commun*, 10(1), 1938. doi:10.1038/s41467-019-09907-2
- Stansfield, J. C., Cresswell, K. G., & Dozmorov, M. G. (2019). multiHiCcompare: joint normalization and comparative analysis of complex Hi-C experiments. *Bioinformatics*, 35. doi:10.1093/bioinformatics/btz048
- Stansfield, J. C., Cresswell, K. G., Vladimirov, I. V., & Dozmorov, M. G. (2018). HiCcompare: An R-package for Joint Normalization and Comparison of HI-C Datasets. *BMC Bioinformatics*, 19. doi:10.1186/s12859-018-2288-x
- Stigler, J., Camdere, G. O., Koshland, D. E., & Greene, E. C. (2016). Single-Molecule Imaging Reveals a Collapsed Conformational State for DNA-Bound Cohesin. *Cell Rep*, 15(5), 988-998. doi:10.1016/j.celrep.2016.04.003
- Sullivan, G. J., Bridger, J. M., Cuthbert, A. P., Newbold, R. F., Bickmore, W. A., & McStay, B. (2001). Human acrocentric chromosomes with transcriptionally silent nucleolar organizer regions associate with nucleoli. *EMBO J*, 20(11), 2867-2874. doi:10.1093/emboj/20.11.2867
- Tanabe, H., Muller, S., Neusser, M., von Hase, J., Calcagno, E., Cremer, M., . . . Cremer, T. (2002). Evolutionary conservation of chromosome territory arrangements in cell nuclei from higher primates. *Proc Natl Acad Sci U S A*, 99(7), 4424-4429. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=11930003
- Tang, Z., Luo, O. J., Li, X., Zheng, M., Zhu, J. J., Szalaj, P., . . . Ruan, Y. (2015). CTCF-Mediated Human 3D Genome Architecture Reveals Chromatin Topology for Transcription. *Cell*, 163(7), 1611-1627. doi:10.1016/j.cell.2015.11.024
- Terakawa, T., Bisht, S., Eeftens, J. M., Dekker, C., Haering, C. H., & Greene, E. C. (2017). The condensin complex is a mechanochemical motor that translocates along DNA. *Science*, 358(6363), 672-676. doi:10.1126/science.aan6516
- Thongjuea, S., Stadhouders, R., Grosveld, F., Soler, E., & Lenhard, B. (2013). r3Cseq: An R/Bioconductor Package for the Discovery of Long-Range Genomic Interactions From Chromosome Conformation Capture and Next-Generation Sequencing Data. *Nucleic Acids Res*, 41, e132. doi:10.1093/nar/gkt373

- Tolhuis, B., Palstra, R. J., Splinter, E., Grosveld, F., & de Laat, W. (2002). Looping and interaction between hypersensitive sites in the active beta-globin locus. *Mol Cell*, *10*(6), 1453-1465. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=12504019
- Van Holde, K. E. (1989). *Chromatin*. (C. Springer Ed.). Heidelberg.
- van Koningsbruggen, S., Gierlinski, M., Schofield, P., Martin, D., Barton, G. J., Ariyurek, Y., . . . Lamond, A. I. (2010). High-resolution whole-genome sequencing reveals that specific chromatin domains from most human chromosomes associate with nucleoli. *Mol Biol Cell*, *21*(21), 3735-3748. doi:10.1091/mbc.E10-06-0508
- van Steensel, B., & Henikoff, S. (2000). Identification of in vivo DNA targets of chromatin proteins using tethered dam methyltransferase. *Nat Biotechnol*, *18*(4), 424-428. doi:10.1038/74487
- Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., . . . Nodell, M. (2001). The sequence of the human genome. *Science*, *291*(5507), 1304-1351. Retrieved from PM:11181995
- Vian, L., Pekowska, A., Rao, S. S. P., Kieffer-Kwon, K. R., Jung, S., Baranello, L., . . . Casellas, R. (2018). The Energetics and Physiological Impact of Cohesin Extrusion. *Cell*, *173*(5), 1165-1178 e1120. doi:10.1016/j.cell.2018.03.072
- Vidal, E., le Dily, F., Quilez, J., Stadhouders, R., Cuartero, Y., Graf, T., . . . Filion, G. J. (2018). OneD: increasing reproducibility of Hi-C samples with abnormal karyotypes. *Nucleic Acids Res*, *46*(8), e49. doi:10.1093/nar/gky064
- Vietri Rudan, M., Barrington, C., Henderson, S., Ernst, C., Odom, D. T., Tanay, A., & Hadjur, S. (2015). Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture. *Cell Rep*, *10*(8), 1297-1309. doi:10.1016/j.celrep.2015.02.004
- Vivante, A., Brozgol, E., Bronshtein, I., Levi, V., & Garini, Y. (2019). Chromatin dynamics governed by a set of nuclear structural proteins. *Genes Chromosomes Cancer*, *58*(7), 437-451. doi:10.1002/gcc.22719
- Wang, H., Maurano, M. T., Qu, H., Varley, K. E., Gertz, J., Pauli, F., . . . Stamatoiyannopoulos, J. A. (2012). Widespread plasticity in CTCF occupancy linked to DNA methylation. *Genome Res*, *22*(9), 1680-1688. doi:10.1101/gr.136101.111
- Watson, J. D., & Crick, F. H. (1953). Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*, *171*(4356), 737-738. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=13054692
- Wutz, G., Varnai, C., Nagasaka, K., Cisneros, D. A., Stocsits, R. R., Tang, W., . . . Peters, J. M. (2017). Topologically associating domains and chromatin loops depend on cohesin and are regulated by CTCF,

- WAPL, and PDS5 proteins. *EMBO J*, 36(24), 3573-3599.
doi:10.15252/embj.201798004
- Xiao, T., Wallace, J., & Felsenfeld, G. (2011). Specific sites in the C terminus of CTCF interact with the SA2 subunit of the cohesin complex and are required for cohesin-dependent insulation activity. *Mol Cell Biol*, 31(11), 2174-2183. doi:10.1128/MCB.05093-11
- Xu, Z., Zhang, G., Jin, F., Chen, M., Furey, T. S., Sullivan, P. F., . . . Li, Y. (2016). A hidden Markov random field-based Bayesian method for the detection of long-range chromosomal interactions in Hi-C data. *Bioinformatics*, 32(5), 650-656. doi:10.1093/bioinformatics/btv650
- Xu, Z., Zhang, G., Wu, C., Li, Y., & Hu, M. (2016). FastHiC: a fast and accurate algorithm to detect long-range chromosomal interactions from Hi-C data. *Bioinformatics*, 32(17), 2692-2695.
doi:10.1093/bioinformatics/btw240
- Yaffe, E., & Tanay, A. (2011). Probabilistic modeling of Hi-C contact maps eliminates systematic biases to characterize global chromosomal architecture. *Nat Genet*, 43(11), 1059-1065. doi:10.1038/ng.947
- Yusufzai, T. M., Tagami, H., Nakatani, Y., & Felsenfeld, G. (2004). CTCF tethers an insulator to subnuclear sites, suggesting shared insulator mechanisms across species. *Mol Cell*, 13(2), 291-298.
doi:10.1016/s1097-2765(04)00029-2
- Zufferey, M., Tavernari, D., Oricchio, E., & Ciriello, G. (2018). Comparison of computational methods for the identification of topologically associating domains. *Genome Biol*, 19(1), 217. doi:10.1186/s13059-018-1596-9