MULTI-SCALE STUDY OF THE GENOME ARCHITECTURE AND ITS DYNAMICAL FACETS

Paula Soler Vila

TESI DOCTORAL UPF / 2019

Director de la tesi

Dr. Marc A. Marti-Renom

STRUCTURAL GENOMICS

GENE REGULATION, STEM CELLS AND CANCER

NATIONAL CENTER FOR GENOMIC ANALYSIS

CENTRE FOR GENOMIC REGULATION

A la meua família d'Ontinyent, de Torrent, de València, de Castelló, d'Alacant, de Toledo, de Barcelona, de Gènova, de Roma, de Madrid, de Galícia, de Boston, de Grècia, de Mallorca...

Per a què mai deixe de créixer.

ABSTRACT (150 WORDS aprox)

High-throughput Chromosome Conformation Capture (3C) has provided a comprehensive overview of the genome architecture. Hi-C, a derivative of 3C, has become a reference technique to study the 3D chromatin structure and its relationship with gene activity and the functional state of the cell. However, several aspects of the analysis and the interpretation of Hi-C data remain a challenge and may hide a potential yet to be unveiled.

In this thesis, we explore the structural landscape of multiple chromatin features. We developed an integrative approach combining *in situ* ***Hi-C data with nine additional omic layers and revealed a new dynamic and transitional state of genome enriched in poised and polycomb-repressed chromatin. This novel*** *intermediate* ***compartment plays an important role in the modulation of the genome during B cells fate decision and upon neoplastic transformation, specifically in chronic lymphocytic leukemia (CLL) or mantle cell lymphoma (MCL) patients.***

We also developed TADpole, a computational tool designed to characterize the entire hierarchy of topologically-associated domains (TADs) using Hi-C interaction matrices. We demonstrated its technical and biological robustness, and its capacity to reveal topological differences in high-resolution capture Hi-C experiments.

RESUMEN

En años recientes, el desarrollo de métodos experimentales basados en** Chromosome Conformation Capture **(3C) nos han permitido tener una visión más detallada y global de como el genoma se pliega en el núcleo celular. En particular, los experimentos Hi-C, derivados de 3C, se han convertido en el método estándar de analizar la arquitectura genómica, así como su relación con la actividad funcional de la célula. La generación de nuevos datos Hi-C ha derivado en una serie de retos de como analizar y interpretar los resultados que nos permitan extraer todo el potencial de los experimentos.

En esta tesis, hemos explorado como se pliega el genoma analizando experimentos Hi-C conjuntamente con datos múltiples de cromatina. Hemos desarrollado un análisis integrativo combinando datos de** in situ **Hi-C con nueva capas epigenéticas de las mismas muestras celulares. Nuestros análisis han relevado la existencia de un nuevo compartimiento genómico caracterizado por su dinámica y capacidad de transición entre estados. Este nuevo compartimiento** intermedio**, enriquecido en cromatina reprimida por Polycomb, juega un papel importante en la modulación del genoma durante la diferenciación en líneas de paciente de células B derivadas en neoplasias, en particular de leucemia linfocítica crónica (CLL) o de linfomas indolentes y de células del manto (MCL).

En esta tesis, además, hemos desarrollado un nuevo método de detección de dominios de genoma (TADs). El método, llamado TADpole, toma como entrada mapas de interacciones de Hi-C. El nuevo método se ha demostrado muy robusto a tanto los datos como en las replicas de experimentos además de ser útil en el estudio de diferencias topológicas usando experimentos de alta resolución de Capture Hi-C.

Keyword list related to the PhD thesis in Catalan or Spanish AND English

Genome architecture, Hi-C, chromatin compartments, topologically-associated domains, bioinformatics tools

PREFACE

The cell is the fundamental unit of an organism. For instance, approximately 3.72×10^{13} cells conform the average human being (100 times more than the stars counted in the Milky Way) clustering into more than 200 different cell types (1). Each of our cells contains around 3 billion DNA base pairs (bp) organized in 23 pairs of chromosomes (22 autosomes and 2 sexual chromosomes). Each base pair is about 0.34 nanometer long, therefore each diploid cell contains approximately 2 meters of DNA folded up and packaged around specific proteins, forming a complex fiber called chromatin, in a nucleus of few micrometers in size. This high compaction of the DNA fiber is folded up to higher-order structures, allowing to maximize the DNA compaction, ensuring an accurate segregation during DNA replication and cell division, while remaining sufficiently accessible for multiples DNA-binding proteins, such as transcription factors, polymerases, nucleases or histones marks that have been reported to play a fundamental role in the genome maintenance and gene regulation.

Thanks to the complementary efforts of the microscopy and molecular biology techniques (especially chromosome conformation capture (3C) technologies), it has been possible to unravel that the DNA folding, as well as promoting a dimensionality-reduction of the fiber, plays a prominent role in the cell function. Increasing evidence indicates that genome architecture regulates gene transcription with implications on cell-fate decisions, development, and disease occurrences such as congenital abnormalities and neoplastic transformations.

This thesis is composed of multiple chapters. In the introduction, we review how the genome is folded in the nucleus following a hierarchical organization and how this folding has a direct impact on the transcription regulation

across genomic scales. The core of the thesis, in chapters 1 and 2, presents the results obtained in the two main publications of the candidate. In chapter 1, we present an integrative multi-omics approach that allowed us to study the dynamics of genome architecture and chromatin function during human B cell differentiation and neoplastic transformation. In chapter 2, a new bioinformatics tool, called TADpole, is proposed to annotate the underlying hierarchical organization of chromatin. Finally, a conclusion is added to highlight the main contributions of this thesis.

OBJECTIVES

The global objective of this thesis is to explore in detail how the chromatin is organized inside the nucleus, specifically at the level of compartments and topologically associated domains (TADs), and assess the biological relevance of this chromatin organization during cell differentiation and upon neoplastic transformation. To achieve this main goal, two main projects were accomplished:

1. ***An exhaustive study of the modulation of chromatin structure during B cell differentiation and upon neoplastic transformation applying an integrative multi-omics approach.***

2. ***Development of a computational tool designed to identify and analyze the entire hierarchy of topologically associated domains (TADs) in intra-chromosomal interaction matrices.***

TABLE OF CONTENTS

TABLE OF FIGURES

INTRODUCTION

The molecular structure of the DNA

From the University of Cambridge to the King College London, five scientists; Maurice Wilkins, Rosalind Franklin with Raymond Gosling (2), James Watson and Francis Crick (3), laid the foundations to determine the structure of the human DNA double-helix.

One by one, 3 billion of recurring structural blocks, known as nucleotides, are precisely hooked and stabilized to write the DNA book, the human genetic instructions. Each nucleotide is composed by a single phosphate group, a pentose sugar and a nitrogenous base. The phosphate group and the pentose are the same for all nucleotides and form the sugar-phosphate backbone of the DNA molecule. Additionally, there are two basic types of nitrogenous bases, purines (adenine(A) and guanine(G)) and pyrimidines (cytosine (C) and thymine (T)). The base pairing (adenine always pairs with thymine, and cytosine with guanine) forms each "rung of the DNA ladder" maintained by hydrogen bonds (Figure 1***). Each side of the ladder is known as a strand, and two sister strands, normally called positive and negative, are twisted around a common axis to form a double-helical structure oriented in an antiparallel sense. As a consequence of the base pair geometry, two grooves, called minor of 12Å and major of 22Å, arise with unequal size in the extension of the polymer. These grooves are potentially binding sites to accommodate DNA binding proteins involved, among many other cellular processes, in the replication and transcription.***

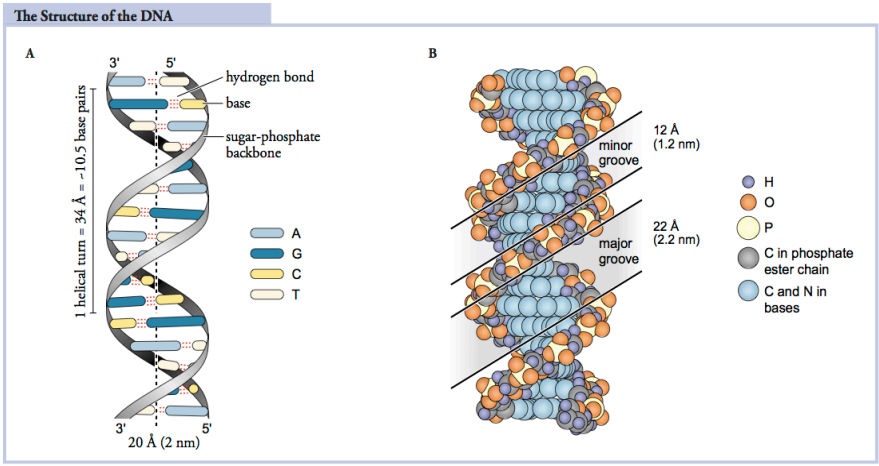


Figure 1: The structure of the DNA. (A) ***Schematic representation of the DNA double helix. The four covalent linked building blocks (A, T, G, C) form the polynucleotide DNA strand. The DNA molecule is formed by two antiparallel strands stabilized by hydrogen bonds established by base-pairing. One turn of the helix spans 3.4 nm (10.5 bp).*** **(B)** ***The alternative distribution of the sugar and phosphates residues together with the bases projected inward the DNA core promotes the creation of the minor (12Å) and the major (22Å) grooves. The latter constitutes an accessible chemical environment to accommodate multiples binding molecules. Adapted from (4).***

In 2003, after 13 years of work involving a great international effort, the first reference sequence of the human genome was published (5-7). At this point and with the subsequent improvements of the reference, the challenge has been to endow functionality to the sequences embedded in the genome. Interestingly, around 99% of the genome does not code for proteins, and for many years this huge portion of the genome was considered as *DNA junk* ***or*** *garbage DNA*. ***Over the years, numerous studies determined how non-coding regions in the human genome can harbor many functionally significant elements, and, as a consequence, play an important role in the regulation and maintenance of the genome (8).***

The first level of DNA folding: Chromatin

Each diploid human cell contains about 2 meters of DNA distributed over 23 pairs of chromosomes that have to fit into a ~10 micrometers cellular nucleus. As if the DNA was a thread, it has to be wrapped around specific protein complexes called nucleosomes to acquire a certain degree of compaction that led it to pack inside the nucleus. The X-ray structure of the nucleosome determined by Luger in 1997 (9) allowed seeing at near-atomic resolution (2.8 Å) how the core components of the nucleosome are assembled and how the 145-147 base pairs (bp) of a linear polymer of eukaryotic DNA are wrapped around it.

The structural organization of the nucleosome was determined by DNA digestion with specific enzymes called deoxyribonucleases (DNases). Particularly, micrococcal nuclease (one type of DNase) was used to break down the DNA by cutting between nucleosomes, making possible to determine that each nucleosome mainly consists of a structured core and an unstructured tail domain. The structured nucleosome core is made up of eight positively charged proteins called histones, which are known as a histone octamer. Each histone octamer is composed of two copies of each histone protein H2A, H2B, H3, and H4 stabilized by the union of the linker histone H1 (*Figure 3*). The latest binds to the DNA entry/exit sites on the surface of the nucleosomal core and stabilize the chromatin into a higher-order structure known as chromatosome (10, 11). The terminal portion of the nucleosome (composed of 15-30 basic residues) are disordered tails that extend outwards from the core, becoming an exposed surface for potential interactions and post-translational modifications (PTM). These type of epigenetic modifications are chemical alterations of the DNA and histones that act as switches implicated in the regulation of gene expression that do not produce changes in the DNA sequence (12).

The classical assumption that nucleosomes are static units is being changed toward dynamics and instructive participants in all chromosomal processes as transcription, replication, DNA repair, etc… (13). Composition alteration, covalent modifications, and translational reposition are the three main dynamic properties that categorize the nucleosomes, conforming an epigenetic diversity known as the "nucleosome landscape" (14). Firstly, the composition alteration promotes changes in the canonical nucleosomes' configuration forming a nucleosome variant. Examples of histone variants as H2A.Z and H3.3 have been demonstrated to play an important role in chromatin structure and gene regulation, associated to a repressive or active state of gene transcription, respectively (15, 16). Secondly, the nucleosome tails, and also the histone core, undergo PTMs that change their interaction with DNA and convert them in potential targets for nuclear proteins. Methylation, acetylation, phosphorylation, and ubiquitination are some of the large repertoires of PTMs that can modulate chromatin structure and transcriptional activity (17). In general, these histone modifications are catalyzed by different enzymes including histone methyltransferases, histone acetyltransferases, histone deacetylases, and kinases that are responsible to add or remove specific covalent modifications leading to activation and repression of transcription, depending on the nature and the position of the PTM (18, 19). All these dynamic modifications conform a collection of combinatorial or sequential signals constituting the so-called "histone code", and their effects on gene transcription can be broadly categorized into active and repressive marks (Figure 2). *For example, histone H3 trimethylated on lysine 4 (H3K4me3) has been associated to active promoters near transcription start sites (TSS) while histone H3 trimethylated on lysine 9 (H3K9me3) and lysine 27 (H3K27me3) usually are associated with silent and poised promoters, respectively, with a general downregulation of nearby genes (20). Describing the histone code is crucial to understanding how the functional associations of the covalent histone modifications affect gene*

expression and their impact in chromatin folding (21). Finally, apart from the histone proteins, other non-histone proteins, such as high mobility group nucleosome-binding proteins, can establish a direct nucleosome-protein interaction that regulate chromatin structure either locally or globally (22).

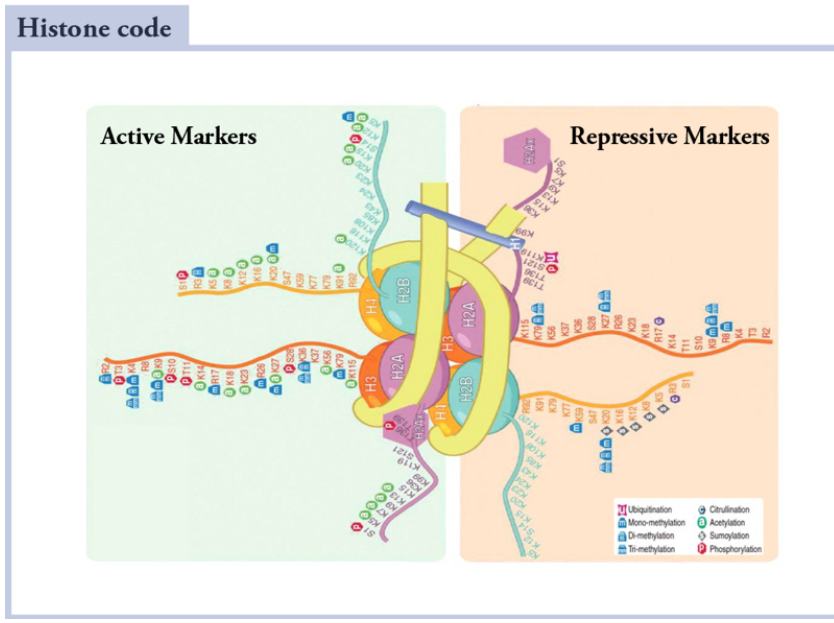


Figure 2: The histone code divided into active and repressive markers. ***DNA is wrapped around the histone cores, whose tails are composed by different amino acids susceptible to be covalently modified. Active marks are represented on the left part of the figure while the repressive mars are located on the right. The common nomenclature of the modifications is the name of the histone, the type and position of the amino-acid and the type and number of modification(s). Lysine (K), arginine (R), serine (S), and threonine (T). Adapted from (23)***

Most genomic DNA is occupied by nucleosomes in a not randomly distributed manner. In fact, many functional regions (promoters, enhancers, and others) are depleted in nucleosomes and some regions are largely nucleosome-free (24). The dynamic nature of nucleosome position has a direct influence in the gene regulation. The ATP-dependent nucleosome remodeling complexes such as SWI/SNF or ISWI regulate the access to DNA sequences promoting the mobilization or rejection of nucleosomes (25).

*Nucleosomes, considered as the "genome's guardian" provide not just the structural support and the requirements to compact the DNA, but also play a crucial role in the control of cell fate and maintaining the integrity of the genome (26). Nucleosome wrapping shortens the chromatin fiber about sevenfold from the naked DNA becoming the first determinant of the DNA accessibility. All these nucleosomal arrays constitute the primary beads-on-a-string 10 nm chromatin fiber. The consequence folding of this chromatin fiber establishes new high-order chromatin structures to culminate with the mitotic chromosome (***Figure 3***)*

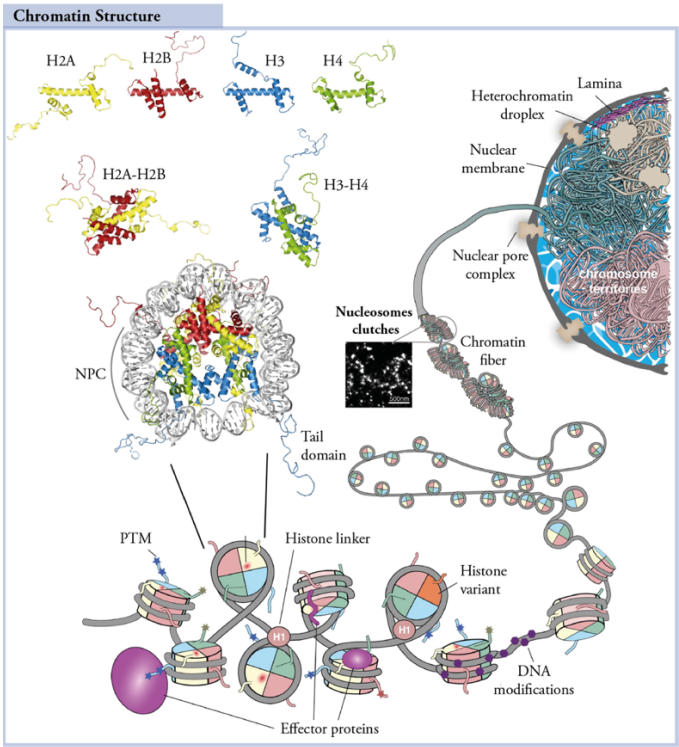


Figure 3: Hierarchical overview of the chromatin structure. ***Two copies of the four histone proteins (H2A-H2B, H3-H4) conform the nucleosome core particle (NPC) around which 145-147 bp of DNA are wrapped. Nucleosomes display a dynamic nature in terms of composition and conformation. Post-translational modifications (PTMs) and histone-variant composition alter the structure of the nucleosome and, consequently, its interaction properties. A linker histone H1 interacts with NPC and the linker DNA to facilitate the folding in non-periodic, irregular high-order structures (such as nucleosomes clutches) to further culminate with the largest structural assembly within the nucleus: the chromosomes territories (CTs). Adapted from (26).***

30-nm chromatin fiber *in vitro* vs *in vivo*

When the chromatin structure is studied isolated in an** in vitro **system, outside the nuclear envelope, it apparently folds around a central axis with a uniform fiber-like conformation of 30 nm diameter. This structural model proposed by Finch and Flug in 1976 (27), assumed that the nucleosomes are distributed consecutively next to each other in a solenoidal one-start-helix. At this point, the fiber of 30 nm was categorized as the basic structural unit of the chromatin. Over time, modifications of this first model emerged, such as zigzag model, which proposed a zigzag arrangement of nucleosomes along the chromatin fiber but always considering a stable and periodic structure of chromatin at 30 nm (28).

However, thanks to the advances in the imaging field, it seems that the chromatin tends to be organized in a non-uniform manner with less regularity folded structures** in vivo **systems (29). Cryogenic electron microscopy (Cryo-EM) and cryo-electron tomography (Cryo-ET) studies applied in multiples species have the potential to observe the cell close-to-the native structure (30, 31). These studies suggested that the nucleosome fiber does not undergo 30 nm folding, highlighting the existence of a disordered and interdigitated state of compactness. Recently, using stochastic optical reconstruction microscopy (STORM), it was possible to visualize at high resolution (~20 nm) that the nucleosomes are distributed in discrete heterogeneous domains, called "nucleosomes clutches" in the interphase nuclei of mammalian cells (32). Interestingly, large clutches with high nucleosome compaction were associated to heterochromatin regions with an increase of H1, whereas the small clutches with a low-density of nucleosomes were associated to active regions (32). Moreover, other techniques such as ChromEMT (a combination of EM tomography and targeted labeling) determined that the chromatin is a disordered granular 5 to 24 nm diameter curvilinear chains packed with many nucleosome rearrangements and structural conformations (33).

The previous results indicate that the 10 nm fiber follows a non-periodicity and irregular folding with less physical constraints that increase the dynamism and accessibility of the DNA. Moreover, the level of chromatin compaction has a direct impact on the degree of DNA exposure to damage-inducing factors and repair pathways (34) and could determine how the cellular machinery accesses genes and consequently its transcription.

How chromosomes are organized in the nucleus? Chromosome Territories.

Since the 19th century, thanks to the microscopy techniques, many relevant features of the chromatin organization are known. Carl Rabl (35) and later Boveri (36), between 1902 and 1904, proposed that the DNA, from the animal interphase chromosomes, is organized in a defined volume, forming discrete entities called chromosome territories (CTs) inside the nucleus. Boveri, in particular, suggested that chromosomes retain its individuality during the interphase but with a certain possibility to overlap with its neighbors' regions.

During the 1950s to 1970s, the first electron microscopy images started to describe a different organization of the chromosomes in the interphase nucleus. This model, which resembles a bowl of spaghetti, considered that the DNA fiber (10-30 nm in diameter) is randomly entangled in the nucleus with a high degree of intermingling. In light of this situation, many scientists tried to demonstrate one of the two proposed models. Among them, two brothers, Thomas and Christoph Cremer did the first indirect evidence of the existence of CTs using laser-UV-micro irradiation experiments (37). The hypothesis of this experiment was based on the fact that DNA-damage distribution and the affected area depend on how chromosomes are arranged in the nucleus with two possible scenarios based on the hypothetical organization of the

*chromosomes (*Figure 4A*). The Cremer brothers experiment clearly showed that micro irradiation in a specific part of the nucleus only damaged segments of the affected CT and its neighboring chromosome territories, without massive damage expansion throughout the rest of the genome (*Figure 4B*).*

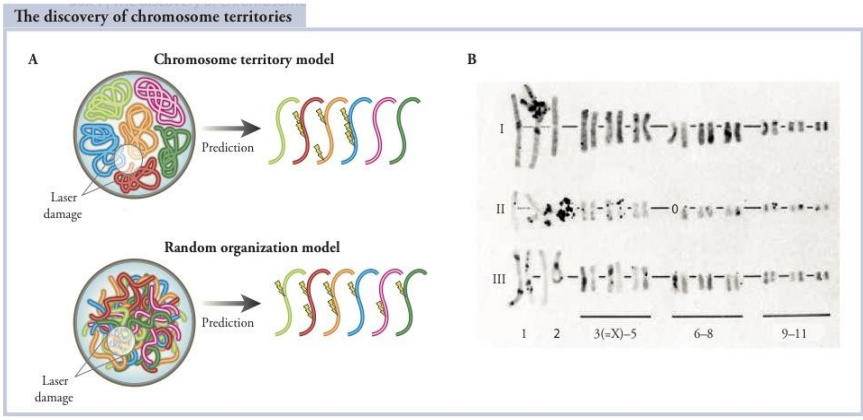


Figure 4: Chromosome territory models (A) *The outline of the two possible models that would explain the distribution of the chromatin fiber in the interphase nuclei. It also indicates what the predicted result would be after a micro-irradiation process.* **(B)** *Three subsets of hamster chromosomes after irradiation. The damage affected mainly chromosome 1 and 2 without a significant expansion through the rest of the chromosomes. Adapted from (38).*

Few years later, using DNA fluorescence in situ hybridization (FISH) and its subsequent improvements as 3D FISH (39) or cryo-FISH (40) it was possible to directly visualize CTs in interphase cells (fixed or lived) of many higher eukaryotes, demonstrating that chromosomes preferably occupy specific non-random areas in the nucleus (40).

However, the precise compaction nature of the chromosomes remained a mystery due to the limited resolution and throughput of these approaches. The discovery of nuclear ligation assay (41) inspired the generation of chromosome conformation capture (3C) technology that was fundamental to overcome the imaging limitations and thus disentangle the genome topology at high resolution. These techniques were the first molecular methods that

*could detect the physical interaction of DNA segments that are close in the nuclear 3D space. The common experimental steps of these methodologies are the following: (i) a cell population is cross-linked by treatment with formaldehyde to promote the covalent bonds between DNA fragments, (ii) isolation and digestion of the chromatin using specific restriction enzyme (it will affect the size of the final fragments and thus the maximal resolution of the experiment) leaving 5' overhangs, (iii) proximity ligation of the restriction fragments, (iv) reverse crosslink and DNA purification and (v) interrogation of the proximity ligation fragments by PCR or sequencing technologies which reflect the interaction frequency between pairs of genomic loci that are close in the 3D space (42, 43) (***Figure 5A***).*

The original 3C method, which allows to determine interactions between one pair of loci(that is, "one-to-one"), evolved in the development of multiple 3C-based techniques: circularized chromosome conformation capture (4C, "one-to-all") (44), chromosome conformation capture carbon copy (5C, "many-to-many") (45), and high-throughput 3C (Hi-C, "all-to-all") (46) among others **(Figure 5B-C).** *Specific to Hi-C, additional experimental steps are added compared to the rest of the 3C-based methods. After DNA digestion, the resulting overhangs are filled with biotinylated-nucleotides. Next, the DNA is shearing and then purified using a biotin pull-down experiment that uses streptavidin beads to ensure capture only the biotinylated DNA junctions for high-through sequencing and subsequent computational analysis. The basic data analysis of a Hi-C experiment involves mainly 5 aspects: (i) read mapping, paired-end reads have to be aligned to the reference genome, (ii) read filtering, non-informative/error fragments (such as un-ligated, self-ligated, PCR artefacts,… etc.) have to be removed to keep only valid pairs, (iii) building the contact matrix, the genome is divided into non-overlapping bins where each bin contains the number of the read pairs that have interactions (the bin size of the Hi-C map is also referred to as "resolution"), (iv) bin filtered with low interactions counts and (v) matrix*

normalization, two main strategies exist to normalize the Hi-C data: explicit methods, that assume that all the biases that affect the data are known such as GC content, mappability, frequency restriction sites and implicit or balancing methods, that assume equal visibility for all bins. Once normalization is completed, a contact heatmap can be generated and inferred proximity information of the entire genome (43, 47).

Hi-C opened the way to interrogate all the genomic interaction pairs in an unbiased genome-wide fashion, which corroborated the existence of CTs. Indeed, a clear preference was detected for the interactions that occur between pairs of loci that come from the same chromosomes (*cis* or intra-chromosome interactions) compared to those established between different chromosomes (*trans* or inter-chromosome interactions) (46). The *cis* interactions can be between pairs of loci apart several kilobases (between promoters and terminators) up to tens of kilobases or even megabases away to allow the interactions between promoters and enhancers (48).

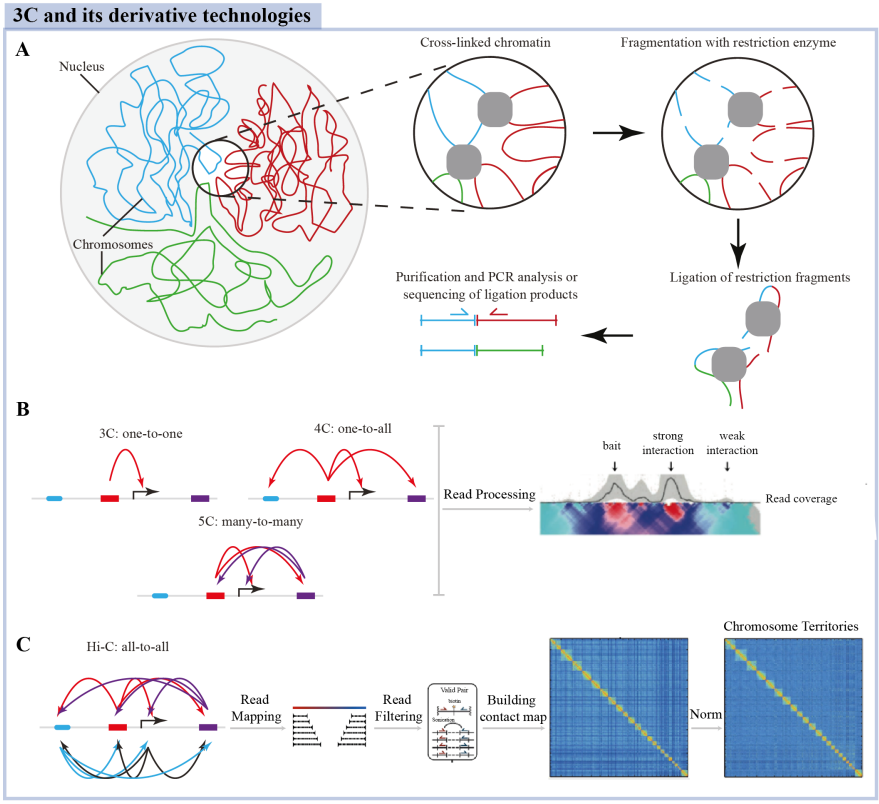


Figure 5: Experimental common basis of the 3C-based methods. (A) ***These methods involve cross-linking cells with formaldehyde to promote covalent bonds between DNA fragments that are physically close in the nuclear 3D space, followed by DNA digestion with a specific restriction enzyme and subsequent proximity ligation of the fragments. These chromatin complexes are purified and analyzed (using PCR or high through sequencing).*** **(B)** ***Variations of 3C technique. 4C (one locus with the rest of the genome) and 5C (all-vs-all in specific genomic regions). After the read preprocessing, a domainogram can be created depicting interaction intensities across the viewpoint(s) to the rest.*** **(C)** ***Unbiased interaction across the entire genome (Hi-C). After read preprocessing and correction bias (read mappability, GC content, number of restriction sites), a normalized matrix*** (M) ***is created containing interaction frequency values in each cell*** M_{ij}***. From it, multiples structural chromatin features can be annotated. Adapted from (49, 50)***

The chromosome territories appear like subnuclear environments that extend into neighboring domains creating intermingling areas. These "areas of contact" between chromosome territories provide a chance to establish potentially functional interactions between different chromosomes, a

phenomenon that has been termed "chromosome kissing", "chromosome intermingling" and more recently "non-homologous chromosomal contacts" (NHCCs) (40, 51-53). Indeed, around 46% of each chromosome intermingles with other chromosomes, highlighting two important features of the chromatin, mobility and plasticity **(Figure 6A)**. *At low resolution, one of the most well-known and largest phenomena of NHCCs become visible, the formation of the nucleolus (53). In human nuclei, the inter-chromosomal and nucleolar associated of five acrocentric chromosomes (13, 14, 15, 21 and 22) which bear the nuclear organizing regions (NOR) behaves the formation of this large sub-nuclear conserved domain (54). At higher resolution, NHCCs have been determined between specific enhancers and target genes. For example, a well-known functional significance of an inter-chromosomal association was characterized between a promoter region of the IFN-gamma gene on chromosome 10 and the regulatory regions of the Th2 cytokine locus on chromosome 11. This association favors the creation of a "poised chromatin hub" that enhances the expression of both Th1 and Th2 cytokines at naive CD4(+) T cells (55).*

The CT localization follows a non-random radial distribution in the nucleus, which is determined by the position of a target chromosome or gene relative to the center of the nucleus and appears to be evolutionarily conserved (56). Interestingly, several factors have been proposed to participate in the organization of the CTs in the nuclear space, such as the chromosome size, replication timing, gene density and transcriptional activity (57, 58). The radial position of the chromosomes correlated with the length of their sequence; the longest chromosomes are preferably located in the peripheral part of the nucleus while the shortest tend to be localized more internally. CTs near the nuclear envelope and perinucleolar space are mainly associated with a decrease of gene expression and often these transcriptionally repressed genes are attached to the nuclear lamina (59). In contrast, gene-rich chromosomes (such as human chromosomes 16, 17, 19 and 22) tend to be

*concentrated in a central position in the nucleus (60) (***Figure 6B***). One of the clearest examples, between the implication of the peripheral location with the loss of expression, is the inactivation and maintenance of chromosome X silencing in female mammals (61). Additionally, several examples have been reflected how the relative position of a gene in the nucleus have a link with their functional state: for example, IgH and IgK loci are preferentially located at the nuclear periphery in hematopoietic progenitor cells, however in pro-B-nuclei appear in the central part of the nucleus, suggesting an association between nuclear positioning and transcriptional regulation during lymphocyte development (62).*

*However, the CT location follows a probabilistic pattern, despite to exist a preferred average position of the chromosomes inside the nucleus, the location in individual cells and tissues show a great variability (38). For example, it has been determined that the chromosome 5 is located preferentially toward the nucleus center in liver cells and lymphocytes compared to the more peripheral position that acquires in lung cells (63) (***Figure 6C***). A recent study, combining direct imaging and transcription information at single-cell level (seqFISH), revealed that nascent site of RNA synthesis tends to be localized at the surface of the CT with a high variable distribution among individual cells promoting inter-chromosomal contacts. These active regions are not dynamically positioned according to the immediate transcriptional activity of the single cell (64).*

There are several well-documented examples where gene nuclear positioning correlate with its expression, however the actual position of a gene is not essential for its normal function. Rather, this positioning can be a result of a clustering of co-regulated genes with similar expression patterns that contribute to their proper expression and regulation (38).

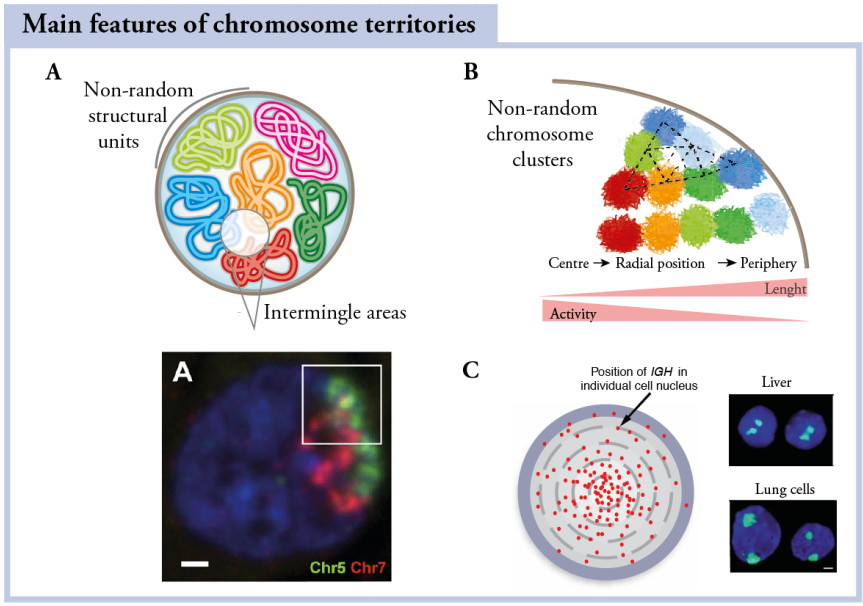


Figure 6: Main features of chromosome territories. (A) *Top. Schematic representation of the individual chromosomes territories that highlight their potential to establish contact regions. Bottom.* ***Intermingling area between chromosome 5 and 7 in human lymphocytes detected by fluorescence microscopy. Adapted from (40).*** **(B)** ***Radial distribution of the CT respect to the nucleus. The chromosome size and the gene expression are factors that inversely correlated with the position of CT inside the nucleus.*** **(C)** *Left.* ***The probabilistic pattern of CT distribution. Every dot is the location of the IgH locus in different cells. Adapted from (38).*** *Right.* ***FISH of chromosome 5 (green) in the liver (located at the nuclear core) and lung cell (located at the nuclear periphery) nuclei. DNA counterstaining with DAPI is in blue. Adapted from (63).***

Chromosomal arrangements can also change during states of differentiation (65), spermatogenesis (66), DNA damage response (67) and in response to changes in cellular homeostasis (as, for instance, serum-starved cells or softer extracellular matrices) (68). Specifically, during these unstable situations, CTs can restore their original positions within minutes to hours depending on the cell type and the context-specific response (68). The altered positions of the CTs have also been described in diseases and neoplasms transformation. The aberrant nuclear position of CT (chromosome translocations or chromosome content imbalance) has the potential to either promote the CT displacement where the translocation has occurred (69) or,

as in the case of aneuploidies, that the chromosome affected alters or re-locate another CTs of its usual nuclear position (70).

Nuclear Neighborhoods

As the precise position of a gene in the nucleus is not enough to determine its activity, alternative mechanisms that have a direct or indirect effect on gene regulation have been studied. Among them, the existence of two kinds of nuclear neighborhoods have proposed: (i) those that are associated with transcriptional repression, such as internal nuclear membrane/nuclear lamina, and (ii) those are associated with transcriptional activation such as the surroundings of nuclear pore complex (NPC), and many nuclear bodies (as nuclear speckles, Cajal bodies or promyelocytic leukemia bodies among others) (71).

The eukaryotic nucleus is a confined cellular organelle surrounded by a lipid bilayer membrane known as a nuclear membrane or nuclear envelope (NE). The NE consists of two parts: the outer nuclear membrane (ONM) and inner nuclear membrane (INM) that are populated by nuclear envelope transmembrane proteins, which associate with the lamin-binding proteins on the INM face to form the nuclear lamina (NL) (72). The NL, composed by a fibrous multi-protein network, can bind many proteins, including chromatin components such as heterochromatin protein 1 (HP1) and histones (72). Using DNA Adenine Methyltransferase Identification (DamID) technology, it has been possible to discover that chromatin establishes a molecular contact with the NL through lamina-associated domains (LADs) (73). These LADs, which vary in size from 100 kilobases (Kb) to 10 megabases (Mb), bear several similarities with the heterochromatin. It (i) harbors silent or low expression genes, (ii) overlaps with late replication timing regions, (iii) has a low density of genes and a large one of gene desserts, (iv) is depleted in RNA polymerase II (PolII) and

*H3K4me2, and (v) enriched in H3K9me2 and H3K27me3 (74). In fact, H3K27me3 is enriched at the outer membrane (ONM), possibly to prevent the spreading of the active chromatin into same LADs (75). However, recent findings suggested that LADs are not necessarily restricted to the nuclear periphery, having the possibility to harbor either heterochromatin and euchromatin domains which contain active genes and regulatory elements (76, 77) (*Figure 7A*). In this line, the visualization of the LAD dynamics in single-cells during the cell cycle demonstrated a step-wise organization with a clear modulation of its aggregation state and its localization in the nucleus. Interestingly, during mitosis, many interactions are measured between LADs and non-LADs regions, and gradually, these inter-regional interactions are reduced during early G1 while the intra-LADs interactions increased (77) (*Figure 7B*).*

The borders of LADs are often enriched in active promoters, CpG islands and CTCF proteins, demarcating the structural limits between the repressed LADs domains and the neighborhoods active regions (74, 77). Some LADs are cell-types specific, while others appear largely conserved in size and position (constitutive LADs). However, a highly orchestrated reorganization of NL interaction has been detected upon progressive cell differentiation (78). In fact, some studies have shown an increased expression of targeted genes when they move away from the lamina. However, the expression of many other genes is not reduced upon similar experimental manipulations, bringing out that nuclear periphery is compatible with transcription (72, 79).

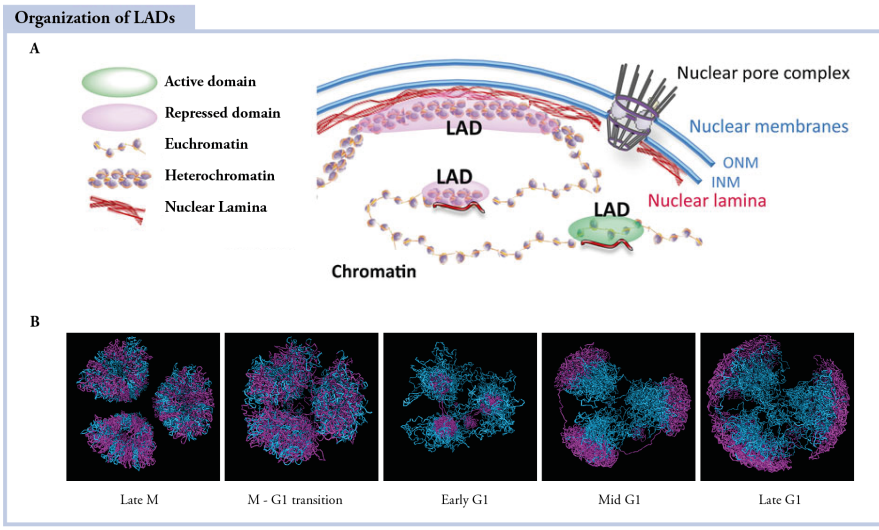


Figure 7: Organization of LADs. (A) ***Schematic view of the possible locations (from peripheral to nuclear interior) of the LADs inside the nucleus. Adapted from (76)*** **(B)** ***Structural molecular dynamics through mitosis (M) to G1. Three chromosome cartoons show the modulation in the location and the aggregation of LADs (colored by magenta) and non-LAD (colored by cyan) during the cell cycle. Adapted from (77).***

Whereas there is a clear association of NL to heterochromatin, the nuclear pore complex (NPC) have been linked with active genes and euchromatin (80). The NPC (that break the continuity of the NE) that has been extensively characterized as nucleo-cytoplasmic molecule exchange, also play important transport-independent roles in the cell, including gene expression, chromatin organization or maintenance of genome integrity. (81). The role of NPCs in transcriptional regulation has been continuously rebound between positive (active NPC-associated compartment) or negative (repressive lamina-associated) modulation of gene expression (82). Another important observation suggested dynamic functional clusters of active RNA polymerase II forming nuclear transcription factories within the nuclear space in the living and fixed cells (83). These active genes tend to localize on the edge of their corresponding CTs (84). Many of co-regulated genes show relocation in a single transcription factory ("gene kissing") (85). For

example,* Hbb *and* Hba *globin genes in mouse erythroid cells, preferentially associate with hundreds of other transcriptional partners in transcription factories to coordinate and increase its transcription in a regulatory landscape (86).

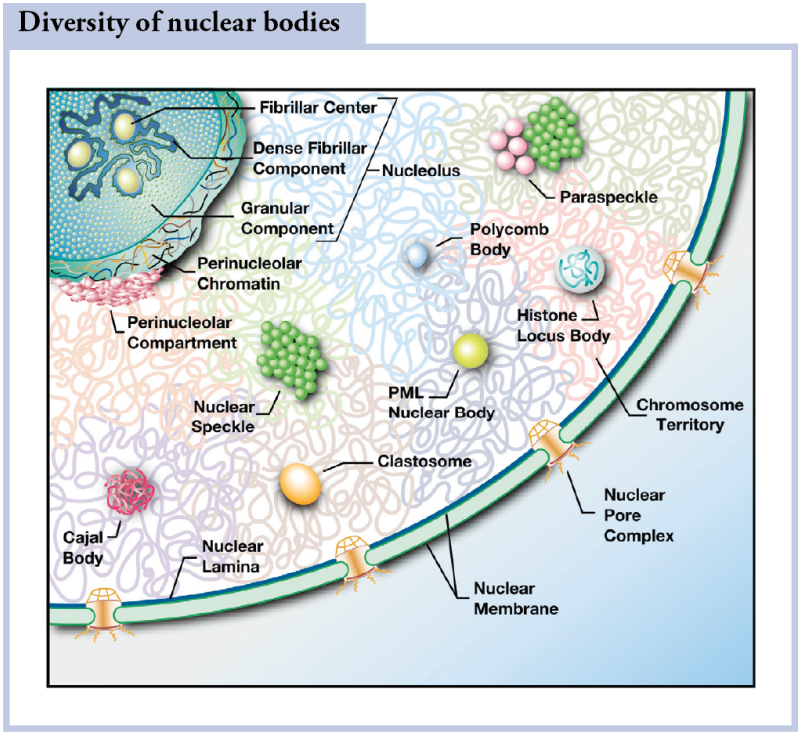


Figure 8: Nuclear neighborhood. ***The cartoon represents the mammalian nuclear landscape in interphase. The nucleus is physically separated from the cytoplasm thanks to the nuclear membrane (NL). The NL is interrupted by the nuclear pore complex that, besides to control nuclear transport, it has been linked to the regulation of gene activity. Under the inner nuclear membrane (INM), the nuclear lamina (NL) appears as a mechanical support and contributes to chromatin organization. Within the nucleus, individual chromosomes (represented as colored threads) occupied limited and non-random regions known as CT. Besides them, the nuclear landscape harbors a wide variety of dynamic nuclear bodies including nucleolus, Cajal bodies, nuclear speckles, paraspeckles, Polycomb bodies, etc…, which have an important role in modulation of numerous nuclear processes. Adapted from (87).***

The non-random distribution of the CT together with its nuclear neighborhood, besides contributing to the structural organization of the genome, help in coordinated gene regulation, encouraging the functional compartmentalization of the genome (38). In the next section, we discuss further levels of segregation of the genome within the CT with functional reverberations.

Compartmentalization at Mb-based scale of the chromatin.

About ~8,4 million paired reads were enough to build the first human genome-wide interaction map using Hi-C technique (46). At low resolution (1 Mb), two main important aspects were retrieved. First, the enrichment of intra-chromosome interactions compared to inter-chromosome interactions (reflecting the chromosome territories) and, second, the power-low decrease of the intra-chromosomal interaction frequency as a function of the genomic distance that points to a set of possible polymer models describing the large-scale chromatin organization (88-90)

By focusing on individual chromosomes, it is possible to see that distinct sets of chromosomal regions, known as compartments, tend to interact preferentially with each other more than expected for a random polymer conformation. To elucidate this level of organization, sequential mathematical transformations of the Hi-C contact matrix (M) are applied (46): (i) a normalization strategy (91), to remove the inherent biases of the experiment producing the normalized matrix (M_{norm}), (ii) a Pearson correlation, computed between the rows and the columns of the M_{norm}, and (iii) a principal component analysis (PCA), that clearly and visually highlights the transitions between these compartments (46). The signature of the compartments is a checkerboard pattern that reflects the preference to keep close loci that present the same interaction profile, epigenomic status and genomic content while separating them from those that have opposite

features; manifesting a functional segregation of the genome into, at least, two compartments (92).

*Normally the first eigenvector (EV_1) of the PCA describes the division of the chromosome into different compartments that show similar interaction behaviors. Regions that present a similar pattern of active chromatin marks (such as H3K36me3 or H3K4me3), DNAseI hypersensitivity, transcription activity, enrichment of RNA polymerase II, early replication domains, high GC content, and high gene density have similar EV_1 values and are categorized as A compartments. By contrast, regions enriched in inactive chromatin marks (such as H3K9me2 or H3K9me3), lamina-associated domains, late replication domains, and present low gene density tend to have the opposite sign of EV_1 values and are annotated as B compartments (46, 93) (*Figure 10B*). On this basis, the A and B spatial segregated compartments were associated with the euchromatin ("active" or "open" chromatin) and heterochromatin ("repressed" or "closed" chromatin), respectively.*

However, the signal of H3K9me3 and H3K27me3 does not perfectly delineate with the bimodal compartmentalization of the genome, suggesting a different gene regulation mechanism of these two histone marks (93). In fact, analyzing the local A/B compartment composition in Arabidopsis *genome, high levels of H3K27me3 were determined in both A and B compartments. This suggests that H3K27me3, a hallmark of the Polycomb Group (PcG) proteins, essential during cell development due to its ability to modulate the chromatin repressing targeted genes, can also be involved in the local chromatin organization (94). Indeed, 3C-based experiments have demonstrated the capacity of PcG to form discrete self-interaction domains with uniform close-range interactions that, thanks to their permissive regulatory topology, can influence in the condition and maintenance of the silenced state of the cell (95, 96). In fact, the folding and the chromatin*

properties of Polycomb-bound chromatin appear to be unique compared with the other compartments with a certain degree of variability between cells. Using 3D-STORM in 46 different epigenetic domains allowed to define three major epigenetic states in Drosophila *cells: (i) active state (enriched in H3K4me2 or H3K79me3), (ii) inactive state (depleted of PcG proteins and transcriptional activators), and (iii) Polycomb-repressed state (enriched in H3K27me3 or PcG proteins) (97). The latter presents an individual compact structure with a high degree of intermingling within itself and a clear tendency to spatially exclude neighboring domains (97). In this line, the capacity of PcG complex to polymerase and establish multivalent interactions suggests its potential ability to form permissive micro-compartments (98, 99) (***Figure 9***).*

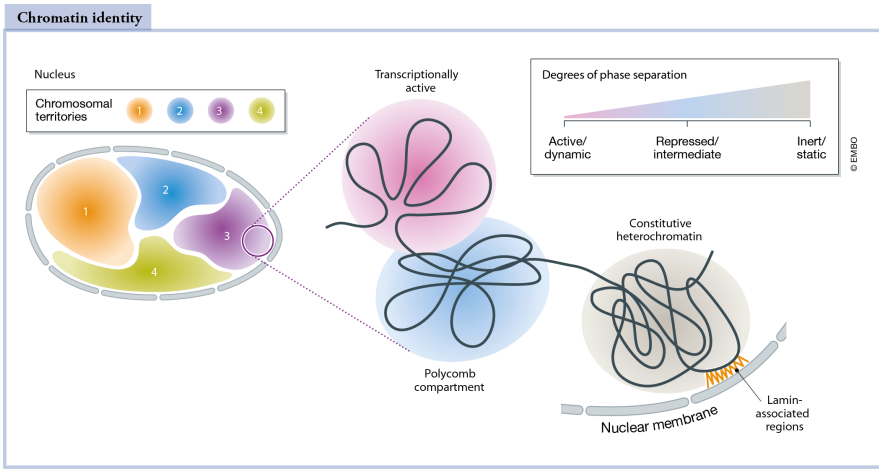


Figure 9: Chromatin identities. ***Chromosomes occupy specific non-random territories in the nucleus and inside them, other degree of compartmentalization can be distinguished. Transcriptionally and open chromatin compartments are spatially segregated from other two more repressive compartments that are formed by constitutive heterochromatin, a transcriptionally inert compartment, and Polycomb compartment, categorized by repressed/intermediate dynamic state. Adapted from (98).***

High-resolution Hi-C studies in humans and mice, also revealed that there are not just two opposite chromatin segregation compartments, but there is a

continuous state of several compartments that can capture better the complexity of the genome interaction landscape (100). Other study revealed a finer cluster of chromatin structure in three specific compartments: one GC-rich and transcriptionally active loci and the other two characterized by low genomic activity and low GC content, distinguished by its relative distance from the centromere, *(i)* ***centromere-proximal and*** *(ii)* ***centromere-distal domains (49). At high resolution (25kb), other divisions of the chromatin compartmentalization were proposed including six (sub)compartments (A1, A2, B1, B2, B3, B4) (101). (Sub)compartments A1 and A2 were related to highly gene dense regions decorating with active histone marks, B1 and B2/B3 were associated with facultative and constitutive heterochromatin respectively and B4 was categorized as a special manually annotated (sub)compartment only present on human chromosome 19, that is enriched in the KRAB-ZNF superfamily genes (101) (*Figure 10A*). Recently, A1 and A2 (sub)compartments were correlated with transcription hot zones, where A1 was associated with the periphery of nuclear speckles and A2 was located to intermediate distance from nuclear speckles. Additionally, B1 was correlated with intermingled regions enriched in polycomb-silenced regions while B2 and B3 were linked to LADs (102).***

Analyzing these alternative chromatin subdivisions, it seems that as we increase the resolution of the Hi-C data, the number of the compartments categories increases, identifying even more distinct patterns at finer resolutions (103, 104). For instance, in the *Drosophila* ***genome, finer (sub)compartments were annotated at 10kb of resolution (104). If these new subdivisions are related to the presence of new biological features or if just the ability to define new boundaries at high resolutions remains to be determined (103).***

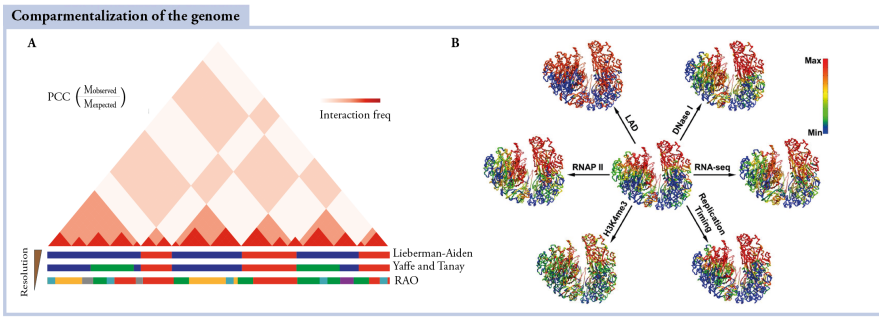


Figure 10: Compartmentalization of the genome. (A) ***Cartoon of a checkerboard pattern in a Hi-C map. At the bottom, examples of alternative (sub)compartmentalization of the genome are shown:*** *(i)* ***classical A (red) and B (blue) segmentation from Lieberman-Aiden,*** *(ii)* ***high-activity cluster (red), low activity centromere-proximal clusters (green) and low-activity centromere-distal clusters (blue) proposed by Yaffe and Tanay and*** *(iii)* ***A1 (green), A2 (aquamarine), B1 (red), B2 (yellow), B3 (gray), B4 (violet) determined by Rao.*** **(B)** ***Mapping the spatial segregation of genome features in 3D chromatin model from IMR90 human cell line. The degree of compartmentalization is shown in the core of the figure. The red color represents enrichment in the case of DNase I hypersensitive sites (DNase I), RNA polymerase II (RNAP II), active gene expression (RNA-seq) and H3K4me3, and deficiency of lamina-associated domain (LAD) and early replication timing. A degree of compartmentalization was computed in IMR90 cells showing how the positive values were located in the interior of compartment A whereas low negative values were located in the inner part of the B compartments with a clear separation by compartments boundaries with intermediate values between them. Adapted from (93).***

The existence of the chromatin compartmentalization was questioned as an experimental Hi-C effect of the average population cell or as if its existence were a transient or simple consequence of shared genomic features (72). However, imaging of numerous genomic regions suggested a spatial arrangement in a polarized manner of the genome along individual chromosomes (105). Interestingly, these compartments are shown as physical structures that in single-cells appear as individual, distinct entities or as entangled structures with a certain variability from cell to cell (106).

In general, the high correlation between the compartment assignment of the genome with the previous biochemical features, highlighted the role of chromatin structure as the emerging regulator of gene expression and the

possible diffusor of the transcription factors to a certain part of the chromatin (93).

How dynamic are the genome compartments?

The organization of genomic compartments is highly dynamic. Numerous studies have shown a great modulation of the chromatin compartments in terms of number, type, and size during cell fate decision. For example, about 36% of the active or inactive chromosomal compartments switched during human embryonic stem cell differentiation correlating in several cases with changes in the gene expression (107). A minor percentage has been reported during human cardiogenesis, were about 19% of the genome changes either from A to B compartments or vice versa. The transition toward more active compartments is coincident with an increase in gene expression and DNA accessibility whereas the transition toward inactive compartments does not necessarily associate with the genomic functional state of the cells (108). During reprogramming of somatic cells into pluripotent cells, changes in subnuclear compartmentalization follow a similar trajectory of the transcriptome, suggesting that changes in the nuclear topology frequently precedes the transcription changes (109).

The compartmentalization of the genome is often perturbed upon variation in the homeostasis cell environment (such as hormone-induced or hyperosmotic stress) (110, 111) and also during neoplastic transformation and diseases. Interestingly, around 12% of all the compartments determined in mammary epithelial cells (at 250kb of resolution) presented a clear transition to the opposite compartment in the breast cancer cell with a higher increment of open compartmentalization (112). Another study determined 32% of compartmental changes between normal cultured B cells (GM12878) with multiple myeloma cell lines (RPMI-8226 and U266) associated with and up- or down-regulation of gene expression (113).

Overall, the nuclear architecture of the genome is highly dynamic and can respond upon environmental perturbations generating a rapid response, while retaining its capacity to restore its initial state (110, 111). Their global and local rearrangements during cell differentiation and upon neoplastic transformation and diseases generally produce a change in gene expression determining a close correlation between structure and function.

Topological-associated domains.

At the tens of kilobase resolution, Hi-C and 5-C experiments revealed the existence of sub-megabase blocks of dense chromatin interactions. These chromatin domains were termed as topologically associating domains (TADs) and were first described by two main features: *(i)* ***a high preference of physical intra-TAD interactions in comparison with inter-TAD interactions and*** *(ii)* ***a non-continuous contact frequency with an abrupt transition between topological domains characterized by a significant reduction of interactions and by enrichment of barrier elements (114, 115) (*****Figure 11*****). On this basis, TADs bring linear distal cis-regulatory elements, such as promoters and enhancers, into a 3D proximity in the nucleus, whereas barrier elements act as insulators to avoid interactions (116).***

Structurally, they appear as globular units formed by looping structures, isolated from the rest in the 3D space even if there are adjacent along the genome (72). Functionally, TADs have been considered as "genomic regulons" to allow spatially proximity of the genes that work in a coordinated fashion (117). TADs have been extensively described in many species as *Drosophila melanogaster, Mus musculus,* ***and*** *Homo sapiens,* ***multiple cell types and even at individual cells, suggesting evolutionary conserved domains and an inherent principle of chromatin folding (115, 118, 119).***

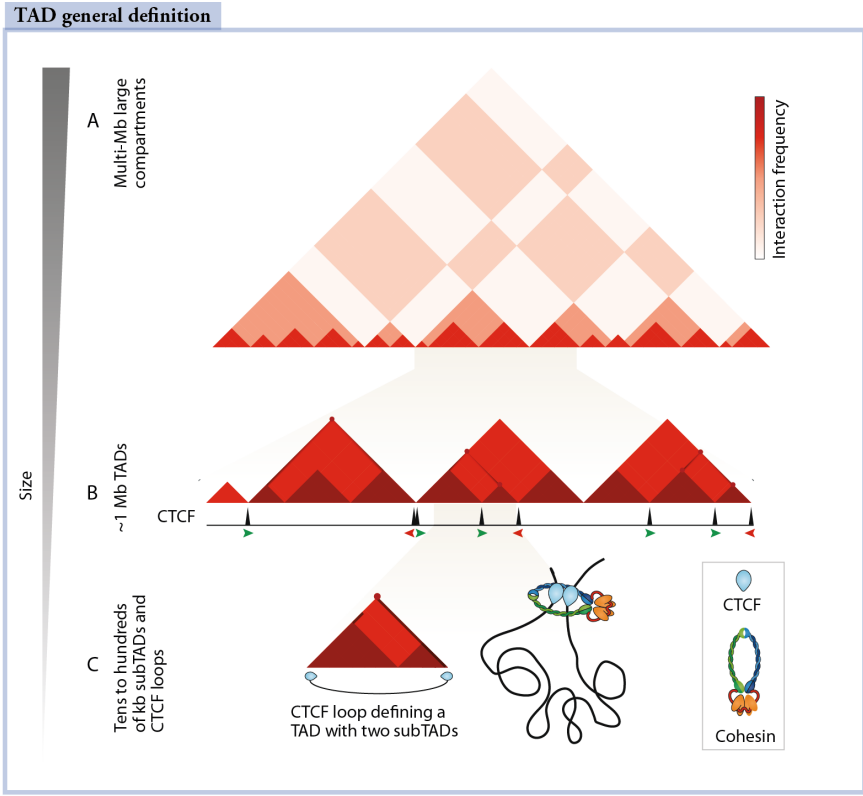


Figure 11: Hierarchical organization of the chromatin at different levels of resolution. (A) ***Low-resolution Hi-C map showing the checkerboard pattern segmenting the genome into A (red) and B (blue) mega-sized compartment.*** **(B)** ***Definition of topologically associating domains (TADs) at 40/50kb of resolution. TADs contain smaller sub-TADs characterized by an increase of interaction frequencies. Some of them are confined by a specific architectural proteins called CTCF.*** **(C)** ***CTCF loop definition, characterized by a strong dark peak on the Hi-C map. Its formation is due to the result of a looping structure establishes between DNA sequences that recruit two convergent CTCF motives and their partner cohesin complex. Adapted from (120).***

TAD boundaries are largely invariant over many cell divisions, across cell-types and evolution (115). However, since the TAD detection is sensitive to the resolution of the Hi-C matrix and the method used to annotate it, their biological importance and even their existence have been extensively debated (98). The ability to detect them as individually privileged self-interaction structures and the duplication/deletion of TAD boundaries associated with

*gene misexpression and developments defects are some arguments that encourage the existence of TADs (98). Alternative methods such as i3C (intrinsic 3C-based method without crosslink)(121), genome architecture mapping (GAM) a ligation-free method using ultrathin cryosections of the nucleus (122) and multiplexed super-resolution imaging methods (123) have reported the existence of TAD-like structures from the population-based experiments to single-cell data (*Figure 12*).*

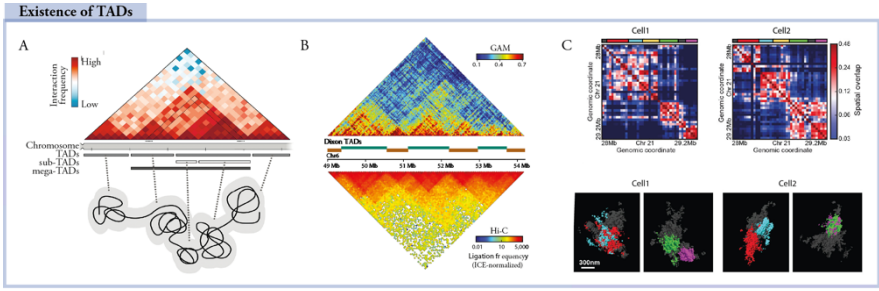


Figure 12: Alternatives methods that support the existence of the TADs at different size-scales. (A) ***Hypothetical Hi-C interaction map highlighting the presence of several hierarchical levels of domains separated by sharp boundaries. Adapted from (124).*** **(B)** ***Genome architecture mapping (GAM) identifies TADs previously detected using Hi-C data. Adapted from (122)*** **(C)** *Top. Spatial distance matrices from two individual IMR90 cells of the chr21. The genomic regions mark in multiples colors represent sub-TADs observed in a population-based experiment. Bottom. 3D STORM images corresponding to the top cells highlighting the different (sub)domains. Adapted from (123).*

Despite the debate on the existence of TADs, dozens of computational methods have been developed to computationally annotated them (125). Four main computational categories can be distinguished based on their mathematical approach: *(i)* ***linear scores associated with each bin,*** *(ii)* ***statistical models based on the interaction distributions,*** *(iii)* ***clustering approaches applied to the contact matrices and*** *(iv)* ***graph theory building dense TADs (sub)networks (125). From them, two main organizations can be retrieved either disjointed and unrelated TADs or overlaid/nested TADs with shared content. Indeed, thanks to the increased resolution of Hi-C***

maps, numerous experimental and theoretical studies proposed that TADs, far from being simple plain triangles close to the diagonal, are organized in nested hierarchical levels, where smaller TADs are part of larger ones, showing a wide range of sizes. In mammalian cells, concepts such as "metaTADs" (126) or "sub-TADs" (127) have been used to define the different sizes and scales that can be used to annotate TADs. "Meta-TAD" is used to define a superior hierarchy of domains-within-domains that are modulated during cell differentiation (126) while "sub-TAD" is used to emphasize how and where the cis-regulatory elements establish physical interactions that contribute to gene regulation (127).

TADs are separated by boundaries enriched in multiples factors that could contribute to their formation. Such factors associate with active promoters, gene bodies, housekeeping genes, transcriptional start sites (TSS), Alu SINE elements and specific chromatin architectural proteins, such cohesion and CCCTC-binding factor (CTCF) factor in vertebrates. Specifically, CTCF and cohesin are considered master regulators of chromatin architecture (128). CTCF, an 11-zinc-finger DNA-binding domain, with the ring shape multi-subunit complex of cohesin, composed of SMC1, SMC3, Rad21, and SA1/2, have been considered as key components of the TAD for its implication of boundary formation and its maintenance. Depletion of CTCF promotes a massive change in TAD topology that loses its structure but does not seem to affect the segregation of the genome at the level of compartments (129). Super-resolution experiments have been shown that after cohesin depletion, the single-cell TAD-like structures persist, suggesting that cohesin is not essential for the maintenance and initial establishment of these structures. However, after this depletion, the preferential positioning of the boundaries to CTCF is abolished, highlighting the dependence on cohesin-CTCF interaction (123).

Changes in the extent of TAD insulation have been reported from a modest effect on gene expression to dramatic consequences in gene regulation, which can contribute to developmental defects or even cancer as the oncogene activation in IDH mutant gliomas (130, 131). From this clear discrepancy, a recent experiment leading by Despang in 2019 (116) suggested that TADs are formed by a redundant system of CTCF sites, the insulation between TADs is not required for developmental gene regulation and the inversion/insertion of boundary elements (redirection of TADs substructures) can induce gene misexpression and diseases.

Chromatin Loops.

Improvements in the original Hi-C protocol as well as deeper sequencing have allowed to increase the number of informative reads up to 5 billion, which has resulted in interaction matrices at a few kilobases of resolution. In this type of matrices, small contact domains (smaller than previously described TADs) were found interacting preferentially over the rest of domain creating strong spots in the contact map, called "peaks" or "chromatin loops" (101). These loops have a great variability, in terms of length and duration time (can be divided into temporal loops created dynamically and strong loops more conserved during cell cycle) (132). Around 10,000 peaks or loops have been detected in human genome-wide matrix at 5-kb of resolution, and they present very interesting properties: *(i)* ***most loops are short-range (<2Mb),*** *(ii)* ***often conserved across cell types and between human and mouse,*** *(iii)* ***many of them promote the association between promoters and enhancers with a clear influence in the gene activity,*** *(iv)* ***often demarcate TADs domain,*** *(v)* ***not present overlapping degree, and*** *(vi)* ***86% of the anchor loops are closely associated with CTCF (90% of the cases in convergent orientation) and cohesin (101).***

How dynamic are TADs and loops?

Given that the Hi-C technique provides a single static snapshot of the chromatin organization by averaging the conformation of millions of cells, many authors highlight its main limitation as its inability to study its dynamics and cellular heterogeneity. However, thanks to the decrease in sequencing costs, more refined 3C-based time-resolved experiments have been done to monitor chromatin dynamics over time. TADs and loop modulation have been extensively described during cell reprogramming (109), cell differentiation (133), after environments perturbations (117) or during the cell cycle (134). All these data have the potential to be integrated, using restraint-based modeling and molecular dynamics approaches, to study at high resolution how the chromatin conformation has been adapted during cell fate conversion (135).

Nowadays, high-throughput DNA sequencing technology allows us to reliably measure many genomic features at the level of single-cell, including RNA-seq, ATAC-seq, and Hi-C for 3D genome architecture (136, 137). Single-cell Hi-C together with DNA-FISH experiments have revealed an extensive cell-to-cell variability at multiple structural levels of chromatin, which provides the opportunity to explore the complex underlying functional aspects that are occurring in the cell. Despite this structural stochasticity, it has been determined how the general chromatin structure is probabilistically linked with genome activity patterns. In fact, biochemical and single-molecule imaging studies (involving CTCF and cohesin complex) suggest that TADs and chromatin loops are dynamic structures that continuously form and fall apart through the cell cycle (138).

Loop extrusion model and phase separation.

The characterization of the different topological levels of chromatin entails a question about how they originate, and which are the molecular and biophysical mechanisms that allow their formation. One of the main

advances in this topic was the determination and assessment of how TADs and loops are formed. The model that currently is supported by the scientist community proposes that TADs and loops are formed via loop extrusion model carried out by the cohesin complex. This model operates in four main steps, (i) the extrusion cohesin complex is formed by two subunits that attach to the intact DNA forming a small loop (ii) the two subunits slide along DNA in opposite directions making the loop larger (iii) the extrusion cohesin complex halts when it finds a specific motif, CTCF motif, that acts as extrusion barriers and (iv) two pairs of CTCF in convergence orientation promote the stop of growing looping and the formation of loop domain (90) **(Figure 13).** *The loops generated by this mechanism are unknotted, promoting fast and easy access to the genetic information just stretched them out and the correct segregation of the chromosomes avoiding possibly detrimental entanglements between them.*

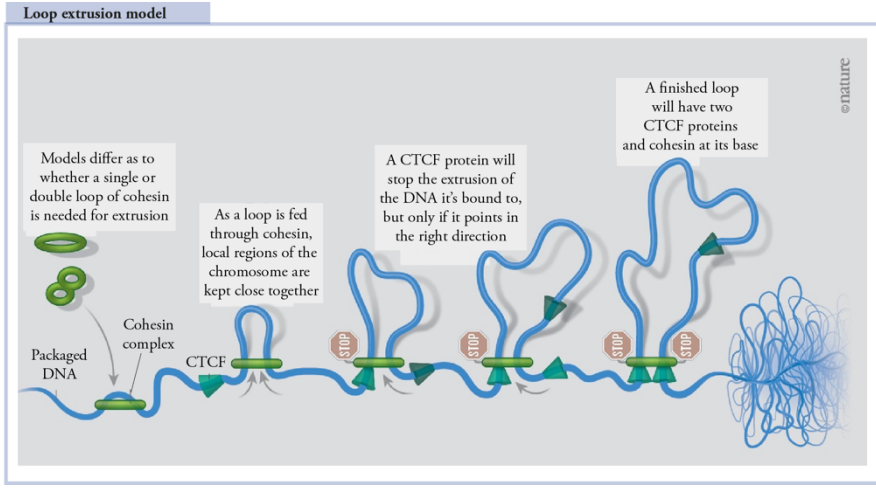


Figure 13: The loop extrusion model. ***The model proposes that TADs are formed when cis-acting loop-extruding factors (such as cohesin) bind to DNA and form progressively large loops until boundary elements (such as CTCF) are attached in a convergent orientation and stop the growth of the loop. Adapted from Nik Spencer illustration for Nature (139).***

Nevertheless, much less is known about the mechanism that promotes the segregation of chromatin into compartments. A phase-separation model has been proposed to explain how interactions between compartments of the same type (A to A or B to B) can generate attractive forces between them while establishing a repulsion force with the compartments of the opposite type, contributing to the physical segregation of the genome (140). However, these attractions can also be the result of the association of the domains with sub-nuclear bodies to form liquid-liquid phase separation. Recently a new variant of Hi-C, called liquid chromatin Hi-C, suggest that the compartmentalization (that can occur when a particular domain present at least 10kb in length) is mainly promoted by the stable heterochromatin interactions while associations between open regions (close to the nuclear speckles) and polycomb-bound regions present highly dynamism (141). Compartmentalization, therefore, appears to be the default mechanism of 3D genome folding, whereas loop extrusion establishes insulated genomic regions that are resistant to further compartmentalization (142) **(Figure 14).**

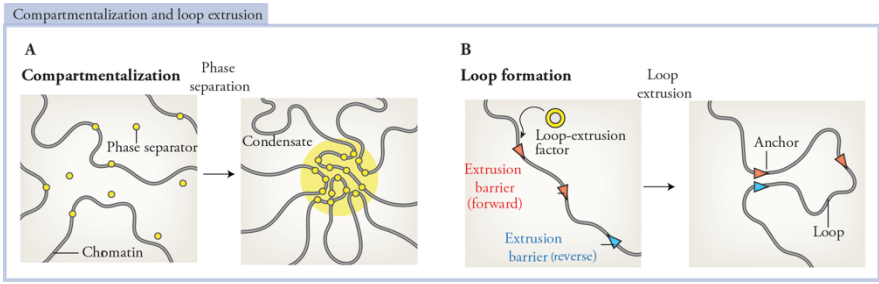


Figure 14: Loop extrusion model and phase separation. (A) ***Compartmentalization model that induces phase-separation as a consequence of the attractive forces between regulatory elements creating 3D hubs in the nucleus.*** **(B)** ***Chromatin loop formation resulted from the loop-extrusion model where cohesin engages the chromatin to start the extrusion until it stops at an extrusion barrier (CTCF). Adapted from (142)***

*In conclusion, all these data suggest that chromatin follows a hierarchical organization that implies a large degree of chromatin structure organization with functional regulation implications. Chromosomes are organized non-randomly in chromosomes territories inside the interphase nucleus with a high frequency of contact inside them in comparison to contacts between them (but with a certain degree of intermingling). At intrachromosomal level, segregation of the genome can be compartmentalized in relation to their functional state. Inside compartments, functional, structural and evolutionary conserved units, called TADs, promote the physical interactions between regulatory elements via loops that necessary for the proper cell function (*Figure 15*).*

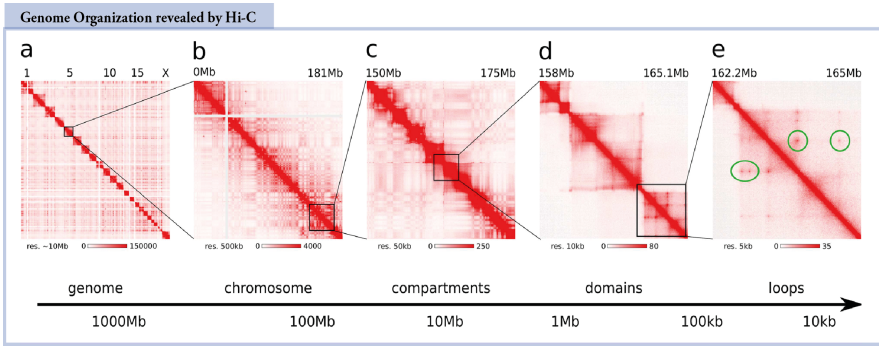


Figure 15: Hierarchical nature of the chromatin architecture determined by Hi-C experiments. (A) *Chromosome territories.* **(B)** *Intrachromosomal interactions.* **(C)** *Segregation of the chromatin in (sub)compartments* **(D)** *Topological associated domains* **(E)** *Looping structures. Adapted from (132)*

CHAPTER 1

Dynamics of genome architecture and chromatin function during human B cell differentiation and neoplastic transformation

Roser Vilarrasa-Blasi, Paula Soler-Vila….

Dynamics of genome architecture and chromatin function during human B cell differentiation and neoplastic transformation

doi:

Dynamics of genome architecture and chromatin function during human B cell differentiation and neoplastic transformation

Roser Vilarrasa-Blasi[1,2,14], Paula Soler-Vila[3,14], Núria Verdaguer-Dot[1], Núria Russiñol[1], Marco Di Stefano[3], Vicente Chapaprieta[1], Guillem Clot[1,5], Irene Farabella[3], Pol Cuscó[4], Xabier Agirre[5,6], Felipe Prosper[5,6,7], Renée Beekman[1,5], Silvia Beà[1,5], Dolors Colomer[1,5,8], Hendrik G. Stunnenberg[9], Ivo Gut[3,10], Elias Campo[1,2,5,11], Marc A. Marti-Renom[3,10,12,13,15], José Ignacio Martin-Subero[1,2,5,12,15]

1. Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Barcelona, Spain.

2. Departament de Fonaments Clínics, Facultat de Medicina, Universitat de Barcelona, Barcelona, Spain.

3. CNAG-CRG, Centre for Genomic Regulation (CRG), Barcelona Institute of Science and Technology (BIST), Barcelona, Spain.

4. Gastrointestinal and Endocrine Tumors Group, Vall d'Hebron Institute of Oncology (VHIO), Barcelona, Spain.

5. Centro de Investigación Biomédica en Red de Cáncer (CIBERONC), Madrid, Spain.

6. Área de Oncología, Centro de Investigación Médica Aplicada (CIMA), Instituto de Investigación Sanitaria de Navarra (IdiSNA), Universidad de Navarra, Pamplona, Spain.

7. Departamento de Hematología, Clínica Universidad de Navarra, Universidad de Navarra, Pamplona, Spain.

8. Hematopathology Section, Hospital Clinic of Barcelona, Barcelona, Spain.

9. Molecular Biology, NCMLS, FNWI, Radboud University, Nijmegen, The Netherlands.

10. Universitat Pompeu Fabra (UPF), Barcelona, Spain.

11. Fundació Clínic per a la Recerca Biomèdica, Barcelona, Spain.

12. ICREA, Barcelona, Spain.

13. Centre for Genomic Regulation (CRG), Barcelona Institute of Science and Technology (BIST), Barcelona, Spain.

14. These authors have contributed equally

15. These authors have jointly supervised this work

Abstract

***We integrate** in situ **Hi-C and nine additional omic layers to define and biologically characterize the three-dimensional (3D) genome architecture across normal B cell differentiation and in patient samples from two distinct B cell tumors. Beyond conventional active (A) and inactive (B) compartments, we identify a highly-dynamic intermediate compartment enriched in poised and polycomb-repressed chromatin. During B cell development, 28% of the compartments change and mostly involve the intermediate compartment. The transition from naive to germinal center B cells is associated with widespread chromatin activation, which reverts into the naive state upon further maturation into memory B cells. The analysis of neoplastic B cells points both to entity-specific alterations in chromosome organization, which entails large chromatin blocks containing key disease-specific genes. This study indicates that 3D genome interactions are extensively modulated during normal B cell differentiation and that the genome of B cell neoplasms acquires a tumor-specific 3D genome architecture.*

Introduction

Over the last decades, our understanding of higher-order chromosome organization in the eukaryotic interphase nucleus and its regulation of cell

state, function, specification and fate has profoundly increased (Rowley and Corces, 2018; Szalaj and Plewczynski, 2018).

Chromatin conformation capture techniques have been used to elucidate the genome compartmentalization (Dekker et al., 2002; Denker and de Laat, 2016). It is widely accepted that the genome is segregated into two large compartments, named A-type and B-type (Lieberman-Aiden et al., 2009), which undergo widespread remodeling during cell differentiation (Andrey and Mundlos, 2017; Dixon et al., 2015; Peric-Hupkes et al., 2010; Stadhouders et al., 2018; Szalaj and Plewczynski, 2018). These compartments have been associated with different GC content, DNAseI hypersensitivity, gene density, gene expression, replication time, and chromatin marks (Lieberman-Aiden et al., 2009; Ryba et al., 2010). Alternative subdivisions of genome compartmentalization have been proposed, including three compartments (Yaffe and Tanay, 2011) or even five compartment subtypes with distinct genomic and epigenomic features (Rao et al., 2014). All of these studies highlight the role of genome three-dimensional (3D) organization in the regulatory decisions associated with cell fate. However, the majority of these studies have been performed using cell lines, animal models or cultured human cells (Dixon et al., 2015; Hu et al., 2018; Johanson et al., 2018; Schmitt et al., 2016; Stadhouders et al., 2018), and although few analyze sorted cells from healthy human individuals (Bunting et al., 2016; Javierre et al., 2016), there is limited information regarding 3D genome dynamics across the differentiation program of a single human cell lineage (Bunting et al., 2016).

Normal human B cell differentiation is an ideal model to study the dynamic 3D chromatin conformation during cell maturation, as these cells show different transcriptional features and biological behaviors, and can be accurately isolated due to their distinct surface phenotypes (Kurosaki et al., 2010; Matthias and Rolink, 2005). Moreover, how the 3D genome is linked to cancer development using primary samples from patients is also widely

unknown (Li et al., 2018). In this context, several types of neoplasms can originate from B cells at distinct differentiation stages (Swerdlow et al., 2017). Out of them, chronic lymphocytic leukemia (CLL) and mantle cell lymphoma (MCL) are derived from mature B cells and show a broad spectrum of partially overlapping biological features and clinical behaviors (Puente et al., 2018). Both diseases can be categorized according to the mutational status of the immunoglobulin heavy chain variable region (IGHV), a feature that seems to be related to the maturation stage of the cellular origin (Chiorazzi and Ferrarini, 2011). CLL cases lacking IGHV somatic hypermutation are derived from germinal center-independent B cells whereas CLL with mutated IGHV derive from germinal center-experienced B cells (Kipps et al., 2017). In CLL, this variable is strongly associated with the clinical features of the patients, with mutated IGHV (mCLL) cases correlating with good prognosis and those lacking IGHV mutation (uCLL) with poorer clinical outcome (Kipps et al., 2017). In MCL, although two groups based on the IGHV mutational status can be recognized and partially correlate with clinical behavior, other markers such as expression of the *SOX11* ***oncogene are used to classify cases into clinically-aggressive conventional MCL (cMCL) and clinically-indolent non-nodal leukemic MCL (nnMCL) (Jares et al., 2012; Navarro et al., 2012; Puente et al., 2018; Royo et al., 2012).***

From an epigenomic perspective, previous reports have identified that B cell maturation and neoplastic transformation to CLL or MCL entails extensive modulation of the DNA methylome and histone modifications (Beekman et al., 2018a; Kulis et al., 2012, 2015; Oakes and Martin-Subero, 2018; Oakes et al., 2016; Queirós et al., 2016). However, whether such epigenetic changes are also linked to modulation of the higher-order chromosome organization is yet unknown (Johanson et al., 2019).

Here, to decipher the 3D genome architecture of normal and neoplastic B cells, we generated *in situ* ***high-throughput chromosome conformation capture***

(Hi-C) maps of cell subpopulations spanning the B cell maturation program as well as of neoplastic cells from MCL and CLL patients. Next, we mined the data together with whole-genome maps of six different histone modifications, chromatin accessibility, DNA methylation, and gene expression obtained from the same human cell subpopulations and patient samples. This multi-omics approach allowed us to identify a widespread modulation of the chromosome organization during human B cell maturation and neoplastic transformation, including the presence of recurrent aberrations in the chromosome organization of regions containing deregulated disease-specific genes.

Results

Multi-omics analysis during human B cell differentiation

We used in situ **Hi-C** *to generate genome-wide chromosome conformation maps of normal human B cells across their maturation program. These included three biological replicates each of naive B cells (NBC), germinal center B cells (GCBC), memory B cells (MBC), and plasma cells (PC) (***Figure 1A-1B** *and* **Table S1***). From the same B cell subpopulations, we analyzed nine additional omics layers generated as part of the* **BLUEPRINT** *consortium (Adams et al., 2012; Beekman et al., 2018a). Specifically, we obtained data for chromatin immunoprecipitation with massively parallel sequencing (ChIP-seq) of six histone modifications with non-overlapping functions (H3K4me3, H3K4me1, H3K27ac, H3K36me3, H3K9me3, H3K27me3), transposase-accessible chromatin with high-throughput sequencing (ATAC-seq), whole genome bisulfite sequencing (WGBS), and gene expression (RNA-seq).*

We initially explored the intra- and inter-subpopulation variability and observed that the Hi-C replicas were concordant, as quantified measuring

*and clustering the reproducibility score (RS) (Yan et al., 2017) (*Figure 1C *and* Figure S1A*). Furthermore, the comparison of samples suggests that the overall genome architecture of NBC is more similar to MBC, and clearly different from GCBC and PC, which belong to a different cluster (*Figure 1C*). This finding was also reflected in the first component of the principal component analysis (PCA) of histone modifications, chromatin accessibility and gene expression (*Figure 1D*). In contrast to other omics marks, the first component of DNA methylation data resulted in a division of GCBC, MBC and PC separated from the NBC. These analyses suggest fundamental differences between chromatin-based epigenetic marks, including chromosome conformation data, and DNA methylation. In fact, changes in DNA methylation linearly accumulate throughout B cell maturation (Kulis et al., 2015; Oakes et al., 2016), which explains the clear differences between NBC and MBC in spite of their converging transcriptomes.*

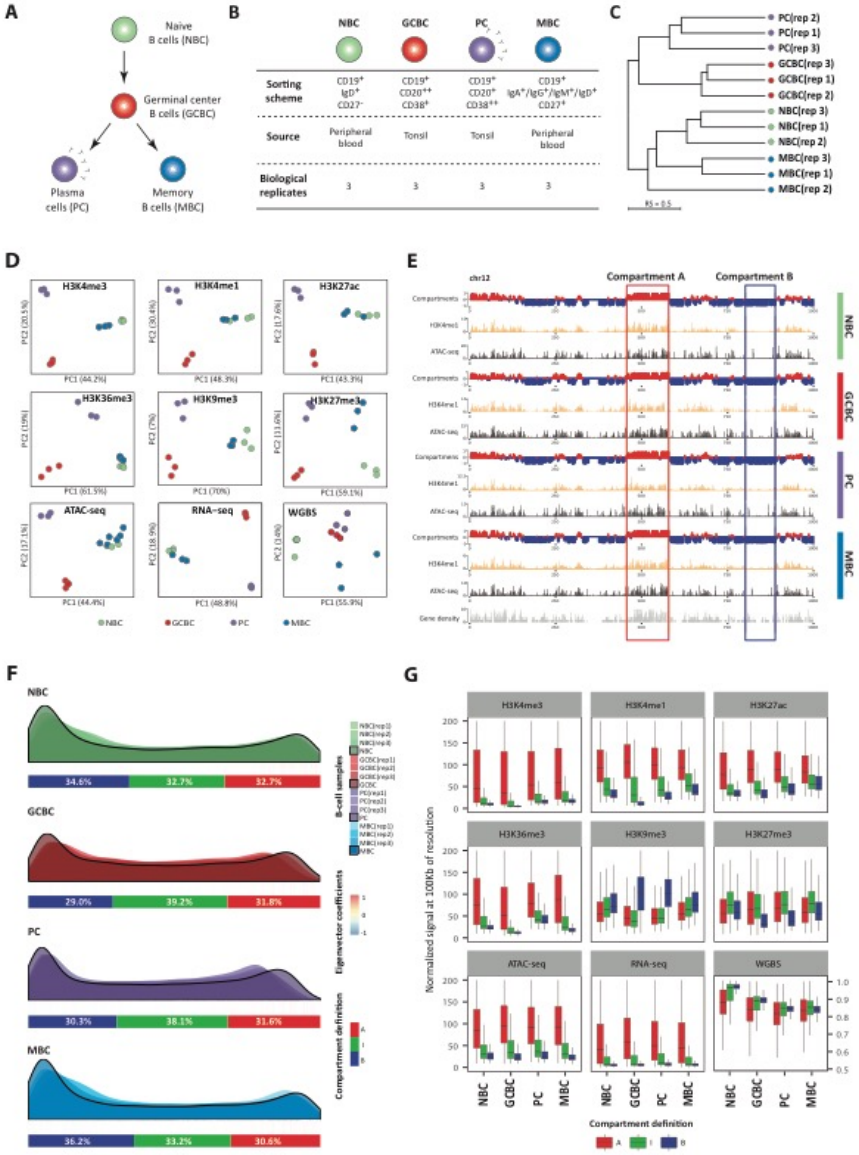


Figure 1. Multi-omics view of B cell differentiation and identification of an intermediate compartment. A - ***Schematic overview of mature B cell differentiation showing the four B cell subpopulations considered in this study.*** **B -** ***Sample description and*** *in situ* ***Hi-C sequencing experimental design for normal B cell differentiation subpopulations. NBC, naive B cells; GCBC, germinal center B cells; MBC, memory B cells and PC, plasma cells.*** **C -** ***Dendrogram of the reproducibility score of B cell subpopulation replicates for normalized Hi-C contact maps at 100Kb resolution.*** **D -** ***Unsupervised principal component analysis (PCA) for nine omics layers: chromatin immunoprecipitation followed by sequencing (ChIP-seq) of six histone marks (H3K4me3 n=46,184 genomic regions, H3K4me1 n=44,201 genomic***

regions, H3K27ac n=72,222 genomic regions, H3K36me3 n=25,945 genomic regions, H3K9me3 n=40,704 genomic regions, and H3K27me3 n=20,994 genomic regions), chromatin accessibility measured by ATAC-seq (n=99,327 genomic regions), DNA methylation measured by whole-genome bisulfite sequencing (WGBS, n=15,089,887 CpGs) and gene expression measured by RNA-seq (n=57,376 transcripts). Three independent biological replicates of NBC, GCBC, PC, and MBC were studied for all omic layers, with the exception of ATAC-seq for which six biological replicates of MBC were used. **E** *- Example on chromosome 12 (chr12) comparing the profile of three-dimensional (3D) data (in situ Hi-C), H3K4me1 ChIP-seq signal, chromatin accessibility (ATAC-seq) and gene density. The red and blue rectangles highlight the features of A and B compartments, respectively.* **F** *- Distribution of the first eigenvector of each B cell subpopulation (three replicates and merge). The relative abundance of A-type, B-type and intermediate (I)-type compartments per merged B cell subpopulations are indicated below each distribution. Compartment definition based on eigenvalue thresholds: A-type, 1 to 0.43; I-type, 0.43 to -0.63; B-type, -0.63 to -1.* **G** *– Boxplots showing the association between the three compartments (A-type, I-type and B-type) and each of the nine additional omics layers under study.*

Polycomb-associated chromatin defines an intermediate and moldable 3D genome compartment

*To study the compartmentalization of the genome during B cell differentiation, we next merged all biological replicates per B cell subpopulation resulting in interaction Hi-C maps with around 300 million valid reads each. These Hi-C interaction maps were further segmented into positive and negative eigenvalues based on the eigenvector decomposition (Imakaev et al., 2012; Lieberman-Aiden et al., 2009), and regions were assigned to the A-type (active) and B-type (inactive) compartments using the association with histone modifications (*Figure 1E *and* Figure S1B*). A pairwise correlation of the first eigenvector of each B cell subpopulation showed that NBC and MBC on the one hand, and GCBC and PC on the other hand, have similar compartmentalization (*Figure S1C*), confirming previous results using the RS (*Figure 1C*). Unexpectedly, the H3K27me3 histone mark, which is deposited by the polycomb repressive complex (Margueron and Reinberg, 2011), was neither correlated with positive nor with negative eigenvector coefficients (*Figure S1B*). We then speculated that, as H3K27me3 was not related with standard A or B compartments, this*

*histone mark may be linked to a different type of chromatin compartmentalization. In this context, a visual inspection of the first eigenvector distribution revealed a positive extreme, a negative extreme and a long intermediate valley (*Figure 1F*). Indeed, applying the Bayesian Information Criterion, we observed that a classification into three compartments was the best compromise between distribution fitting accuracy and minimum number of compartments (*Figure S1D*). Subsequently, we modelled the eigenvector distribution to establish the thresholds segmenting the data into an A-type, B-type and intermediate (I)-type compartments (*Figure S1E-F*). Analyzing these three compartments together with other omics layers revealed the expected association of A-type compartment with active chromatin, B-type compartment with H3K9me3, and a remarkably association between the I-type compartment and the presence of H3K27me3 (*Figure 1G*). Indeed, a chromHMM-based chromatin state model specific for B cells (Beekman et al., 2018a; Ernst and Kellis, 2017) revealed that the regions associated with the I-type compartment were enriched for poised-promoter and polycomb-repressed chromatin states (*Figure 2A *and* Figure S2A*).*

*We next quantified the compartment interactions by computing the compartment score (C-score) as the ratio of intra-compartment interactions over the total chromosomal interactions per compartment (*Figure S2B*). Interestingly, the I-type compartment was associated with lower C-score than the A-type and B-type compartments (*Figure S2C*). We further explored this phenomenon by dividing the I-type compartment into two blocks differentiating positive (IA) and negative (IB) eigenvector components (*Figure S2D*). The analysis showed that the I-type compartment, regardless being IA or IB, was consistently having lower C-score than the A or B-type compartments. This finding further supports the existence of the I-type compartment as an independent chromatin structure different from A and B-type compartments. Additionally, it suggests that the I-type compartment*

tends to interact not only with itself but also with A and B-type compartments, and as such it may represent an interconnected space between the fully active and inactive compartments.

*To study the potential role of the I-type compartment during B cell differentiation, we selected poised promoters or polycomb repressed regions within this compartment in NBC and studied how they change in both compartment and chromatin state upon differentiation into GCBC (*Figure 2B*). The majority of compartment transitions (69.1% of poised promoter and 73.0% of polycomb repressed) change into A-type compartment, a consistent fraction (21.9% and 21.1%) into B-type, and only a small fraction (9% and 5.9%) maintain their intermediate definition. This finding indicates that the regions with a most prominent I-type compartment character undergo a widespread structural modulation during NBC to GCBC differentiation step. Interestingly, transitions from I-type to A-type compartment (activation events) were paired with a reduction of poised promoters (56.7% loss) and polycomb repressed states (70.2% loss). These reductions were associated with an increase of A-related chromatin states (1.31- or 1.33-fold change coming from poised promoter or polycomb-repressed, respectively) such as promoter, enhancer and transcription (*Figure 2B*). Conversely, poised promoters and polycomb-repressed regions associated with I-type compartments in NBC that changed into B compartments in GCBC (inactivation events) were related to an increase of B-related chromatin states (3.81 or 1.4-fold change coming from poised promoter or polycomb-repressed, respectively) such as heterochromatin characterized by H3K9me3 (*Figure 2B*).*

Altogether, these results point to the existence of an intermediate transitional compartment with biological significance, enriched in poised and polycomb-repressed chromatin states, interconnected with A and B -type compartments, and amenable to rewire the pattern of interactions leading to active or inactive chromatin state transitions upon cell differentiation.

Changes in genome compartmentalization are reversible during B cell differentiation

*Mapping A, I and B-type compartments in NBC, GCBC, MBC and PC Hi-C maps revealed that 28.1% of the genome dynamically changes compartment during B cell differentiation (*Figure 2A *and* Figure S2A*). B cell differentiation is not a linear process, NBC differentiate into GCBC, which then branch into long-lived MBC or antibody-producing PC. Thus, we studied the 3D genome compartment dynamics along these two main differentiation paths (NBC-GCBC-PC and NBC-GCBC-MBC). At each differentiation step, we classified the genome into three different dynamics: (i) compartments undergoing activation events (B-type to A-type, B-type to I-type, or I-type to A-type), (ii) compartments undergoing inactivation events (A-type to B-type, A-type to I-type, or I-type to B-type), and (iii) stable compartments (*Figure 2C-D*). The NBC-GCBC-MBC differentiation path suggests that the extensive remodeling taking place from NBC to GCBC is followed by an overall reversion of the compartmentalization in MBC, achieving a profile similar to NBC (*Figure 2C*). To assess the capacity of the genome to revert to a past 3D configuration, we analyzed the compartments in NBC as compared to those in PC and MBC. Indeed, we globally observed that 72.7% of the regions in MBC re-acquire the same compartment type as in NBC. This phenomenon was mostly related to compartments undergoing activation in GCBC, as 82.9% of them reverted to inactivation upon differentiation into MBC. This finding is in line with solid evidence showing that NBC and MBC, in spite of representing markedly different maturation B cell stages, are phenotypically similar (Agirre et al., 2019; Klein et al., 2003) (*Figure 1D*). In the case of PC, the compartment reversibility accounted only for 30.8% of the genome (*Figure 2D*). To determine whether this compartment reversibility was also accompanied by a functional change, we analyzed the chromatin state dynamics within the compartments becoming uniquely active in GCBC as compared to NBC,*

*MBC and PC (n=937) (*Table S2*). We observed that the transient compartment activation from NBC to GCBC is related to an increase of A-related chromatin states (1.36-fold change). Conversely, the subsequent 3D genome inactivation upon differentiation into MBC and PC was related to an increase in B-related chromatin states (1.21- and 1.15-fold change, respectively) (*Figure 2E *left). Furthermore, those regions had a significant increase in chromatin accessibility and gene expression in GCBC as compared to NBC and MBC, but not in PC (*Figure 2E *right). These findings suggest that structural 3D reversibility in MBC is accompanied by a functional reversibility whereas PC partially maintains gene expression levels and chromatin accessibility similar to GCBC in spite of the compartment changes. Interestingly, in contrast to chromatin-based marks, DNA methylation was overall unrelated to compartment or chromatin state dynamics of the B cell differentiation (*Figure 2E *right).*

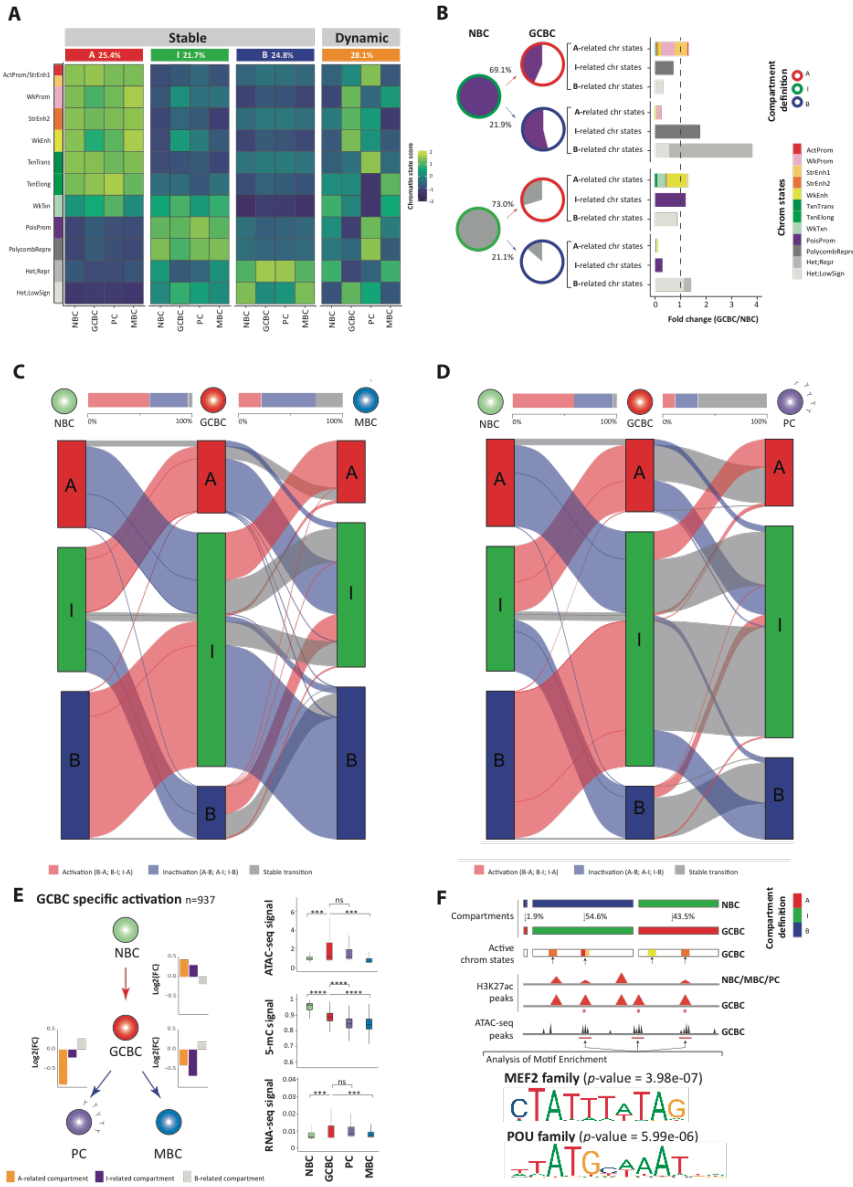


Figure 2. Chromatin dynamics across B cell differentiation. A - ***Functional association of the conserved and dynamic compartments during B cell maturation using eleven chromatin states (normalized by sample and chromatin state). Conserved compartments were segmented into A-type, I-type and B-type compartments. The percentage of each conserved or dynamic compartment is indicated for all B cell subpopulations. ActProm-StrEnh1, Active Promoter-Strong Enhancer 1; WkProm, Weak Promoter; StrEnh2, Strong Enhancer 2; WkEnh, Weak Enhancer; TxnTrans, Transcription Transition; TxnElong, Transcription Elongation; WkTxn, Weak Transcription; PoisProm, Poised promoter; PolycombRepr, Polycomb repressed; Het;Repr, Heterochromatin;Repressed;***

Het;LowSign, Heterochromatin;Low Signal. **B** *- Intermediate compartment dynamics. Pie charts represent poised promoters (top, violet color) or polycomb-repressed (bottom, light gray color) within the I-type compartment in NBC which shifts to A-type and B-type compartments in GCBC. The pie charts under GCBC represent the fraction that maintains the previous chromatin state (colored as previously defined) or changed chromatin states (not colored). Bar graphs represent the fold change between GCBC and NBC of each three groups of chromatin states (arranged by their relationship to the A-type, I-type and B-type compartments). Active Promoter, Weak Promoter, Strong Enhancer 1, Strong Enhancer 2, Weak Enhancer, Transcription Transition, Transcription Elongation, Weak Transcription were A-type compartment-related states. Heterochromatin/Repressed and Heterochromatin/Low signal were B-type compartment-related states. Poised Promoter or Polycomb repressed chromatin states were I-type compartment-related states.* **C/D** *- Alluvial diagrams showing the compartment dynamics in the two branches of mature B cell differentiation: NBC-GCBC-MBC (***C***) and NBC-GCBC-PC (***D***). Activation, in red, represents changes from compartment B-type to A-type, B-type to I-type and I-type to A-type. Inactivation, in blue, represents changes from A-type to B-type, A-type to I-type and I-type to B-type compartments. The non-changed compartments are represented in gray. On the top, the bar plots between B cell subpopulations represent the total percentage of regions changing to active or inactive, and regions that conserve its previous compartment definition.* **E** *– Multi-omics characterization of the 937 regions (of 100Kb resolution) gaining activity exclusively in GCBC.* **Left:** *Scheme of B cell differentiation and chromatin state dynamics, in which the barplots indicate the log2 fold change of active, intermediate or inactive -related chromatin state groups.* **Right:** *Boxplots of chromatin accessibility (ATAC-seq signal), DNA methylation (5-mC signal) and gene expression (RNA-seq signal) per B cell subpopulations compared using the Wilcoxon's test. *p-value<0.05, **p-value<0.001, ***p-value<0.0001, ****p-value<0.00001.* **F** *- Enrichment analysis of transcription factor binding motifs.* **Top***: Schematic representation of the analytic strategy.* **Bottom***: Binding motifs of MEF2 and POU TF families are highly enriched in active and accessible loci in the GCBC specific regions gaining activity (n=171 independent genomic loci) versus the background (n=268 independent genomic loci). p-values were calculated using the AME-MEME suite. Out of the list of all enriched transcription factor binding motifs, we considered only those expressed in the three GCBC replicates.*

The 3D genome of GCBC undergoes extensive compartment activation

*Our analyses revealed that the NBC and GCBC transition was associated to a large structural reconfiguration of compartments involving 96.0% of all dynamic compartments (***Figure 2A***). Interestingly, 61.5% of the changes between NBC and GCBC involved compartment activation (***Figure 2C-D***). As the germinal center reaction is known to be mediated by specific transcription factors (TFs) (De Silva and Klein, 2015; Song and Matthias,*

*2018) and those may be involved in shaping the spatial organization of the genome (Bunting et al., 2016; Johanson et al., 2018; Stadhouders et al., 2018), we further explored the presence of TF binding motifs in the newly activated compartments. We identified significantly enriched motifs for MEF2 and POU families (*Figure 2F *and* Table S3*), which are essential TFs involved in germinal center formation (Brescia et al., 2018; Schubart et al., 2001; Wilker et al., 2008; Ying et al., 2013). Furthermore, the newly activated compartments hosted about 100 genes significantly upregulated in GCBC as compared to the rest of B cell subpopulations (FDR<0.05) (*Table S4*). Remarkably, among them was the Activation Induced Cytidine Deaminase (AICDA) gene, which is essential for class-switch recombination and somatic hypermutation in GCBC and is specifically expressed in GCBC (de Yébenes and Ramiro, 2006). Indeed, the AICDA locus was globally remodeled from an inactive state in NBC to a global chromatin activation in GCBC, which included an increase in the ratio of GCBC/NBC 3D interactions as well as increased levels of active chromatin states (that is, active promoter and enhancers as well as transcriptional elongation), open chromatin, and gene expression (*Figure 3A-B*). This analysis also revealed the presence of possible upstream and downstream AICDA-specific enhances that gain interactions with the gene promoter in GCBC (*Figure 3B*). Interestingly, this multilayer chromatin activation at the AICDA locus was reverted to the inactive ground state once GCBC differentiate into MBC or PC.*

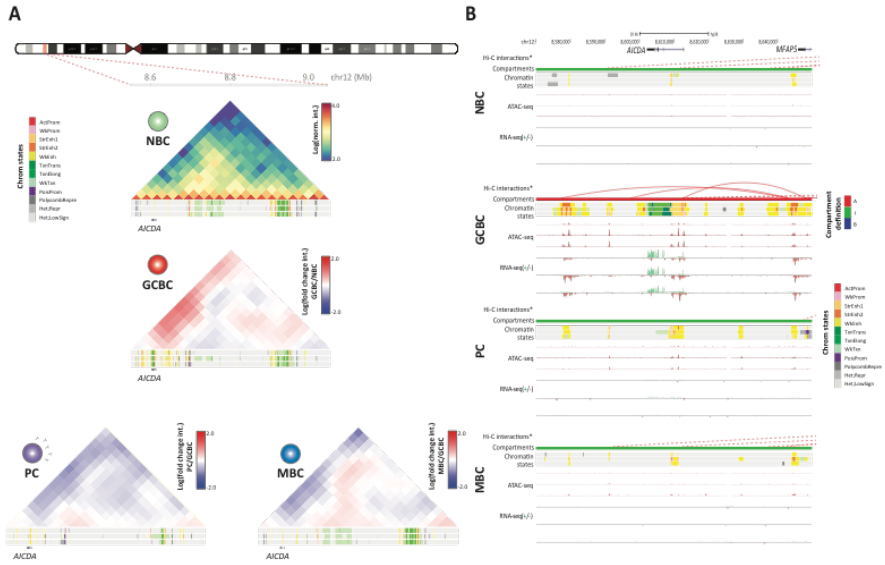


Figure 3. Chromatin organization at the *AICDA* locus. A - ***Normalized Hi-C contact map of the domain structure surrounding the AICDA gene in NBC. The log fold change interaction ratio between GCBC, MBC or PC as compared to NBC was computed. Below each interaction map, chromatin state tracks of three biological replicates per B cell subpopulation are shown. The coordinates of the represented region are chr12:8,550,000-9,050,000, GRCh38.*** **B –** ***Multi-layer epigenomic characterization of AICDA gene region (chr12:8,598,290-8,615,591, GRCh38) in four B cell subpopulations. Arc diagrams indicate the Hi-C significant interactions (continuous red lines involve the region of interest, while dashed red lines involve other regions of chromosome 12). Below them, we show compartment definition (red, compartment A-type: green, compartment I-type), chromatin states, chromatin accessibility (ATAC-seq, y-axis signal from 0 to 105) and gene expression (RNA-seq, y-axis signal from 0 to 4 for the positive strand and from 0 to -0.1 for the negative strand). Tracks of Hi-C interactions and compartment definition are based on merged replicates whereas chromatin states, chromatin accessibility and gene expression tracks of each replicate is shown separately. The coordinates of the represented region are chr12:8,570,000-8,670,000, GRCh38.***

B cell neoplasms undergo disease-specific 3D genome reorganization

Next, we analyzed whether the observed 3D genome organization during normal B cell differentiation is further altered upon neoplastic transformation. To address this, we performed in situ Hi-C in fully characterized tumor cells from patients with chronic lymphocytic leukemia (CLL, n=7) or mantle cell lymphoma (MCL, n=5). Within each neoplasm,

*we included cases of two subtypes, IGVH mutated (m, n=5) and unmutated (u, n=2) CLL as well as conventional (c, n=2) and non-nodal leukemic (nn, n=3) MCL (*Figure 4A *and* Table S5*). An initial unsupervised clustering of the RS from the entire Hi-C dataset indicated that CLL and MCL, similarly to the PCA from other omic layers generated from the same patient samples, clustered separately from each other and within a major cluster that included NBC and MBC (*Figure 4B-C *and* Figure S3A*). Interestingly, NBC and MBC have been described as potential cells of origin of these neoplasms (Puente et al., 2018). Furthermore, pairwise eigenvector correlation analysis of the cancer samples suggested that the 3D genome configuration of the two clinico-biological subtypes of CLL was rather homogeneous (*Figure S3B-C*). This was not the case for the two MCL subtypes, which were more heterogeneous (*Figure S3D-E*).*

*The differential clustering of CLL and MCL samples hint into disease-specific changes of their 3D genome organization (*Figure 4B*). To further detect those changes, we took the fraction of the genome with stable compartments during normal B cell differentiation and compared them to each lymphoid neoplasm. Qualitatively, we observed that roughly one quarter of the genome changes compartments in at least one CLL (23.8%) and at least one MCL sample (27.3%) as compared to normal B cells (*Figure 4D-E *left). Using a more stringent quantitative approach, we aimed at detecting changes associated with CLL or MCL as whole, which revealed a total of 348 and 82 significant compartment changes (absolute difference in the eigenvalue>0.4 and FDR<0.05) in CLL and MCL, respectively. The larger number of regions changing compartments in CLL correlates with the results of the Hi-C based clustering (*Figure 4B*), which indicates that MCL is more similar to NBC/MBC than CLL. Moreover, the observed compartment changes tended towards inactivation in CLL (57.5%) (*Figure 4D *middle) and towards activation in MCL (57.0%) (*Figure 4E *middle) compared to the normal B cells. These 3D genome organization changes were*

*associated with the expected changes in chromatin function. Inactivation at the 3D genome level in CLL was linked to a shift to poised promoter and polycomb-repressed chromatin states, and a significant loss of chromatin accessibility and gene expression (*Figure 4D *right). Activation at the 3D genome level in MCL was accompanied with an enrichment of active chromatin states and a significantly increase in chromatin accessibility and gene expression (*Figure 4E *right). Overall, these results point to the presence of recurrent and specific changes in the 3D genome organization in CLL and MCL, being the former more extensively altered than the latter.*

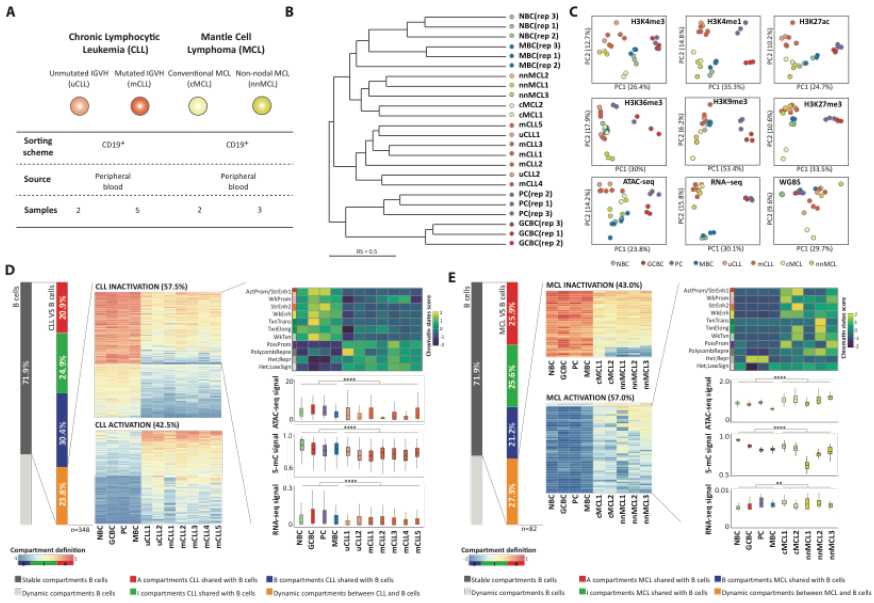


Figure 4. Characterization of the chromatin architecture of human B cell neoplasms. A - ***Sample description and*** *in situ* ***Hi-C experimental design in CLL and MCL cases.* B -** ***Dendrogram of the reproducibility score for normalized Hi-C contact maps at 100Kb for B cell subpopulations replicates and samples from B cell neoplasia patients. IGVH unmutated (u)CLL; IGVH mutated (m)CLL; conventional (c)MCL and non-nodal (nn)MCL.* C -** ***Unsupervised principal component analysis (PCA) for nine omic layers generated in the same patient samples as Hi-C: chromatin immunoprecipitation followed by sequencing (ChIP-seq) of six histone marks (H3K4me3 n=53,241 genomic regions, H3K4me1 n=54,653 genomic regions, H3K27Ac n=106,457 genomic regions, H3K36me3 n=50,530 genomic regions, H3K9me3 n=137,933 genomic regions, and H3K27me3 n=117,560 genomic***

regions), chromatin accessibility measured by ATAC-seq (n=140,187 genomic regions), DNA methylation measured by whole-genome bisulfite sequencing (WGBS, n=14,088,025 CpGs) and gene expression measured by RNA-seq (n=57,376 transcripts). In addition to the normal B cell subpopulations explained in **figure 1D***, we studied 7 CLL patient samples (2 uCLL and 5 mCLL) and 5 MCL patient samples (2 cMCL and 3 nnMCL).* **D –** *Compartment changes upon CLL transformation.* **Left***: First bar graph represents the percentage of conserved and dynamic compartments during normal B cell differentiation. Second bar graph shows the percentage of compartments stable and differential in CLL as compared to normal B cells. A total of 23.8% of the compartments change in at least one CLL sample.* **Middle***: Heatmaps showing eigenvector coefficients of the 348 compartments significantly losing (n=200) or gaining activation (n=148) between all CLL samples and normal B cells.* **Right***: Multi-omics characterization of the 200 regions losing activity in CLL. We show chromatin states, chromatin accessibility (ATAC-seq signal), DNA methylation (5-mC signal) and gene expression (RNA-seq signal) in CLL and normal B cells. Comparisons were performed using the Wilcoxon's test. ****p-value<0.00001.* **E –** *Compartment changes upon MCL transformation.* **Left:** *First bar graph represents the percentage of conserved and dynamic compartments in B cells. Second bar graph shows the percentage of conserved compartments between B cells and MCL, being 27.29% non-conserved compartment in MCL.* **Middle:** *Heatmaps showing eigenvector coefficients of significant dynamic compartments (n=82) between MCL and B cells. Regions were split in two groups (MCL activation, n=35 or inactivation, n=47) according to the structural modulation of the MCL compared to B cells.* **Right:** *Example of the MCL activation subset (mostly those B-type compartments in B cells which significantly increase eigenvector coefficients in MCL) showing the chromatin states pattern, chromatin accessibility (ATAC-seq signal), DNA methylation (5-mC signal) and gene expression (RNA-seq signal). Comparisons were performed using the Wilcoxon's test. *p-value<0.05, **p-value<0.001, ***p-value<0.0001, ****p-value<0.00001.*

EBF1 downregulation in CLL is linked to extensive 3D genome reorganization

To further characterize the compartmentalization of neoplastic B cells, we classified the changing compartments as common (between CLL and MCL) or entity-specific (either in CLL or MCL). We detected 31 compartments commonly altered in both malignancies, revealing the existence of a core of regions that distinguish normal and neoplastic B cells (*Figure 5A-B*). A targeted analysis of CLL and MCL revealed 89 CLL-specific (41 and 48 inactivated and activated, respectively) and only 3 MCL-specific compartment changes (*Figure 5C*,* Figure 6A *and* Figure S4A*). Interestingly, the set of 41 compartments inactivated in CLL were

significantly enriched (p-value=0.0060) in downregulated genes (n=11) as compared to normal B cells and MCL samples, being the Early B cell Factor 1 (EBF1) a remarkable example (*Figure 5C-D *and* Table S6*). EBF1 downregulation has been described to be a diagnostic marker in CLL (Navarro et al., 2017), and its low expression may lead to reduced levels of numerous B cell signaling factors contributing to the anergic signature of CLL cells (Mockridge et al., 2013; Muzio et al., 2008) and low susceptibility to host immunorecognition (Schultze et al., 1996; Seifert et al., 2012). To obtain insights into the mechanisms underlying EBF1 silencing in CLL, we analyzed in detail a 2Mb region hosting the gene, which also contains two nearby protein coding genes, RNF145 and UBLCP1, and a lncRNA, LINC02202. We observed that a large fraction of 3D interactions involving the EBF1 region in normal B cells were lost in CLL resulting in a change from A-type to I-type compartment and a sharp inactivation of the gene, as shown by the analysis of chromatin states (*Figure 5E*). Remarkably, in spite of the global reduction of 3D interactions, the two adjacent genes (RNF145 and UBLCP1) were located in the only region (spanning 200Kb) that remained as A-type compartment in the entire 2Mb region, maintaining thus an active state. To obtain further insights into the EBF1 genome structure, we modeled its spatial organization in NBC and CLL by using the restraint-based modeling approach implemented in TADbit (Baù and Marti-Renom, 2012; Serra et al., 2017) (*Figure 5F *and* Figure S4B-C*). The EBF1 domain in CLL resulted in larger structural variability as compared with the models in NBC due to the depletion of interactions in neoplastic cells (*Figure S4B*). The 3D models revealed that the EBF1 gene is located in a topological domain, isolated from the rest of the region in NBC, hosting active enhancer elements (*Figure 5F*). Remarkably, the active enhancer elements together with the interactions are lost in CLL (*Figure 5F*), resulting in more collapsed conformations (*Figure 5G*). Overall, these analyses suggest that EBF1

silencing in CLL is linked to a compartment shift of a large genomic region leading to the abrogation of interactions and regulatory elements.

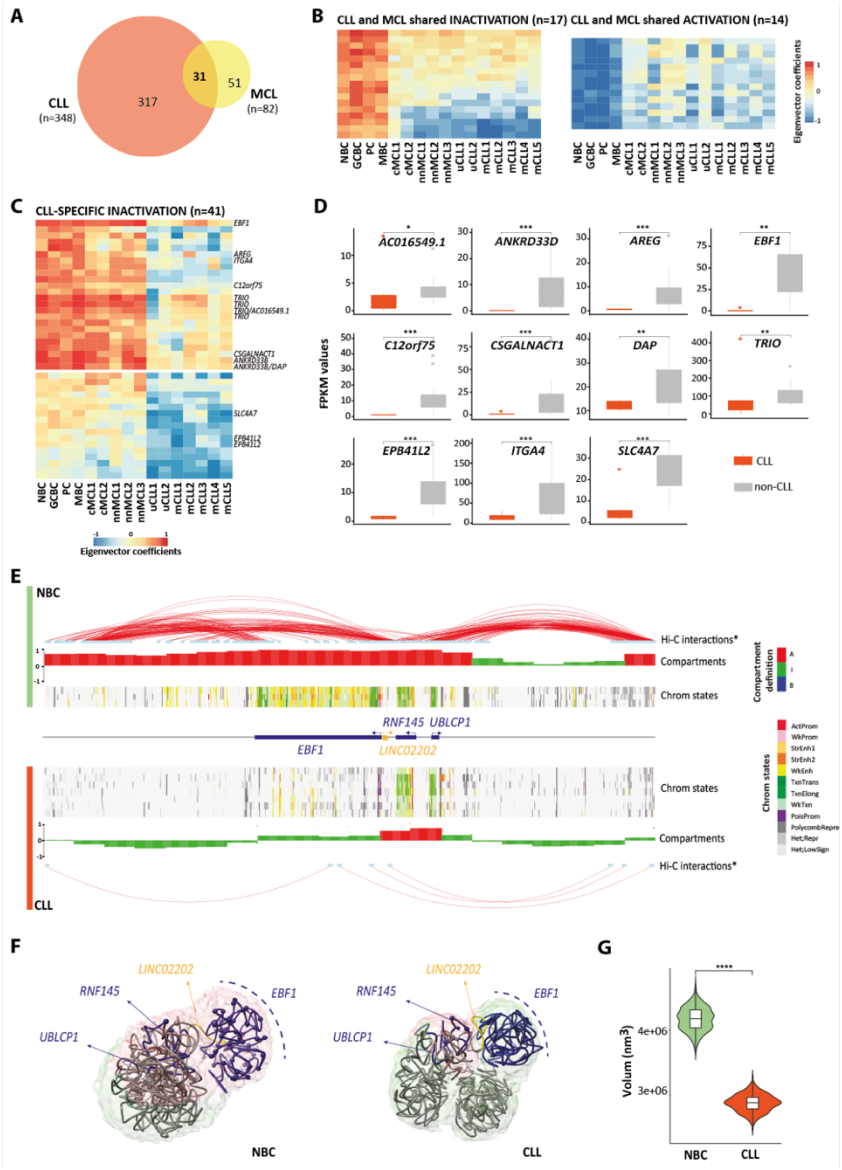


Figure 5. *EBF1* silencing in CLL is accompanied by structural changes affecting a 2Mb region. A - ***Venn diagram showing the significant number of dynamic compartments in CLL and MCL as compared to normal B cell differentiation and the regions shared between both B cell neoplasms (n=31).*** **B -**

Heatmaps showing eigenvector coefficients of compartments significantly losing or gaining activation between B cell neoplasms (MCL and CLL together) and B cells. **C** ***- Heatmap showing the eigenvector coefficients of the compartments losing activation specifically in CLL (n=41). Significantly downregulated genes (FDR<0.05) associated to each compartment are shown on the right of the heatmap (p-value=0.0038, calculated from the total number of genes picked on 48 random compartments per 10,000 times).*** **D** ***- FPKM values of all the CLL-specific significantly downregulated genes within compartments losing activation. *adjusted p-value<0.05, **adjusted p-value<0.005, ***adjusted p-value<0.0005.*** **E** ***- Map of the*** EBF1 ***regulatory landscape. Significant Hi-C interactions (p-value=0.001) and compartment type from merged NBC and a representative CLL sample, followed by chromatin state tracks from each NBC (n=3) and CLL (n=7). The coordinates of the represented region are chr5:158,000,000-160,000,000, GRCh38.*** **F** ***- Restraint-based model at 5Kb resolution of the 2Mb region containing*** EBF1 ***(total 400 particles,*** EBF1 ***locus localized from 139 to 220 particle). Data from merged NBC (***top***) and CLL (***bottom***) was used. Surface represents the ensemble of 1,000 models and is color-coded based on the compartment definition (A-type, B-type and I-type in red, blue and green, respectively). The top-scoring model is shown as trace, where protein-coding genes are colored in blue and long non-coding RNAs in yellow. Spheres represents enhancer regions.*** **G** ***- Violinplot of the convex hull volume involving the 81 particles from the*** EBF1 ***region. Comparison was performed using Wilcoxon's test. ****p-value=0.00001.***

Our analysis also detected 48 regions that changed towards more active compartment exclusively in CLL (Figure 6A***). As expected, these regions were significantly enriched in upregulated genes (p-value=0.0038) and harbored 9 genes with increased expression (***Figure 6B *and* Table S7***). As previously shown for regions gaining activity in GCBC (***Figure 2E***), we evaluated whether particular TFs were related to the CLL-specific increase in 3D interactions. Indeed, we found an enrichment in TF binding motifs of the TCF (p-value=0.00004) and NFAT (p-value=0.00647) families, which have been described to be relevant for CLL pathogenesis (Beekman et al., 2018a; Gutierrez et al., 2010; Le Roy et al., 2012) (***Figure 6C *and* Table S8***). One of the nine upregulated genes in CLL-specific active compartments was*** KSR2, ***a gene whose upregulation has a strong diagnostic value in CLL (Navarro et al., 2017). Importantly, this gene contained several motifs for the TCF4 transcription factor (***Figure 6D***), which itself is overexpressed in CLL as compared to normal B cells (Beekman et al., 2018a), suggesting in this***

particular example that TCF4 overexpression may lead to aberrant binding to KSR2 regulatory elements and a global remodeling of its 3D interactions.

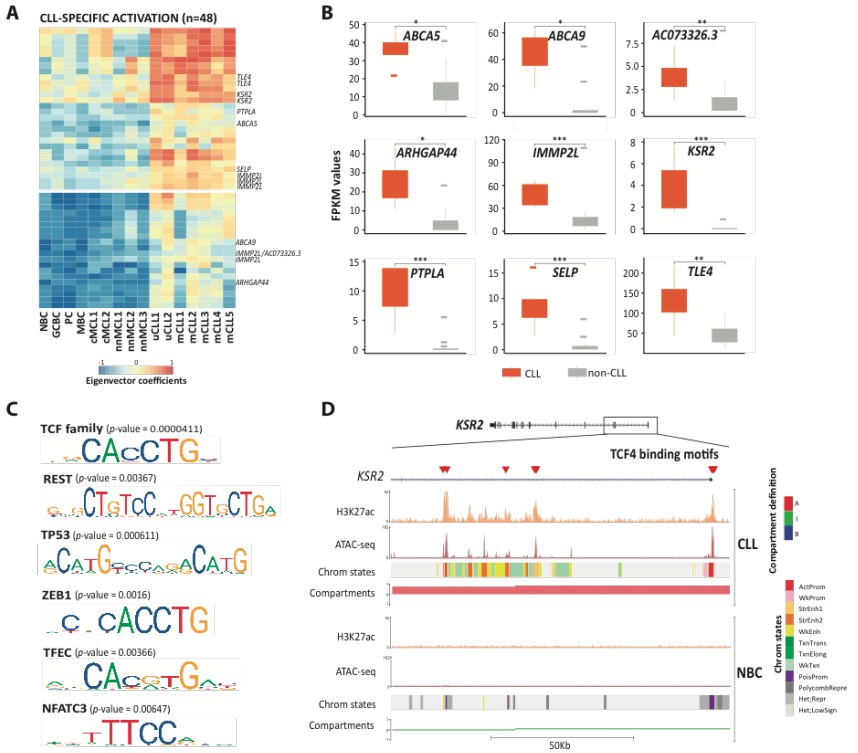


Figure 6. Transcription factors associated to CLL-specific activated compartments. A - *Heatmap showing the first eigenvector coefficients of the compartments gaining activation specifically in CLL (n=48). Significantly upregulated genes (FDR=0.05) associated to each compartment are shown on the right of the heatmap (p-value=0.006).* **B** - *FPKM values of all the CLL-specific significantly upregulated genes within compartments gaining activation. *adjusted p-value<0.05, **adjusted p-value<0.005, ***adjusted p-value<0.0005.* **C** - *Enrichment analysis of transcription factor binding motifs. We show the most significant TF binding motifs enriched in active and accessible loci within the CLL-specific regions gaining activity (n=25 independent genomic loci) versus the background (n=28 independent genomic loci). p-values were calculated using the AME-MEME suite. Out of the list of all enriched transcription factor binding motifs, we considered only those expressed in all CLL samples (n=7).* **D** - *Example of TCF4 binding motifs at the KSR2 promoter region in CLL and NBC. We show the following tracks: H3K27ac, chromatin accessibility (ATAC-seq) and chromatin states of a representative NBC replicate and CLL sample. The coordinates of the represented region are chr12:117,856,977-117,975,164, GRCh38.*

Increased 3D interactions across a 6.1Mb region including the *SOX11* oncogene in aggressive MCL

***In addition to entity-specific 3D genome changes, our initial analyses also suggested that different clinico-biological subtypes may have a different 3D genome organization, especially in MCL (*Figure 4B*). To identify subtype differences within each B cell neoplasia, we selected regions with homogeneous compartments within each disease subtype and classified them as distinct if the difference between the Hi-C matrices cross-correlation eigenvalues was greater than 0.4. Applying this criterion, we defined 47 compartment changes between uCLL and mCLL, and 673 compartment changes between nnMCL and cMCL (*Figure 7A*). This finding confirmed the previous analyses (*Figure S3B-E*), and indicated that the two MCL subtypes have a markedly different 3D genome organization. Two thirds of the compartments changing in the MCL subtypes (n=435, 64.6%) gained activity in the clinically aggressive cMCL, and one third gained activity in nnMCL. We then characterized the chromosomal distribution of these compartment shifts, which, surprisingly, was significantly biased towards specific chromosomes (*Figure 7B*). In particular, those regions gaining 3D interactions in aggressive cMCL were highly enriched in chromosome 2, being 22.3% (n=97) of all 100Kb compartments located in that chromosome (*Figure 7B*). We next analyzed chromosome 2 of cMCL in detail and we observed a* de novo *gain of A-type and I-type compartments accumulated at band 2p25 as compared to both normal B cells and nnMCL (*Figure 7C*). The entire region of about 6.1Mb had a dramatic increase of interactions and active chromatin states in cMCL as compared to nnMCL (*Figure 7D *and* Figure S5A*). Most interestingly, this region contains* SOX11*, whose overexpression in cMCL represents the main molecular marker to differentiate these two MCL subtypes (Fernandez et al., 2010), and has been shown to play multiple oncogenic functions in cMCL pathogenesis (Balsas et al., 2017; Palomero et al., 2016; Vegliante et al., 2013). However, as* SOX11

*is embedded into a large block of 6.1Mb gaining activation in cMCL, we wondered whether additional genes could also become upregulated as a consequence of the large-scale spatial organization of chromosomal band 2p25. Indeed, mining the expression data from the 5 MCL cases studied herein as well as two additional published cohorts (Navarro et al., 2017; Scott et al., 2017), we observed that 13 (43%) of the 30 expressed genes within the 6.1Mb region were over-expressed in cMCL as compared to nnMCL in at least one cohort (***Figure 7D** *and* **Figure S5B-C**), which may also contribute to cMCL pathogenesis and clinical aggressiveness.*

.

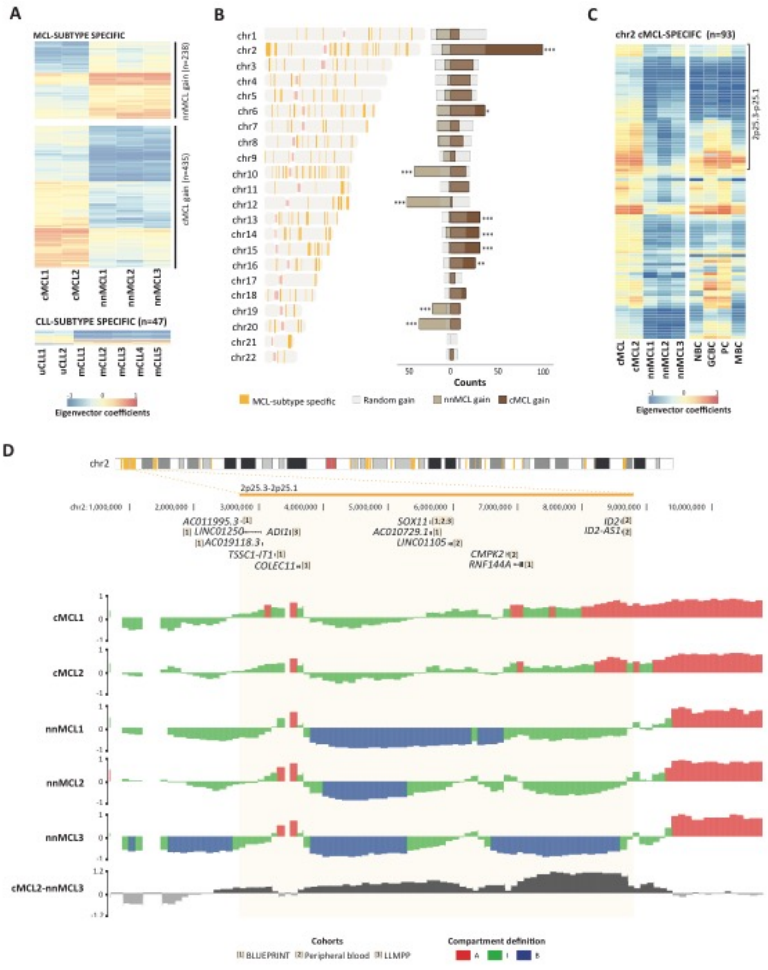


Figure 7. Long-range chromatin remodeling of a 6.1Mb involving *SOX11* in cMCL A - ***Heatmaps showing eigenvector coefficients of compartments significantly changing in cMCL versus nnMCL (n=673) and in uCLL versus mCLL (n=47).*** **B - Left:** ***Genome-wide distribution of compartments changing in MCL subtypes. The vertical orange lines point to the chromosome location of the regions.*** **Right:** ***Relative abundance of the compartments significantly gaining activity in cMCL or nnMCL as compared with a random probability. A gain in compartment activation was defined as an increase of eigenvector coefficient of at least 0.4. *p-value<0.05, **p-value<0.005, ***p-value<0.0005.*** **C -** ***Heatmap showing eigenvector coefficients of the chromosome 2 compartments specifically gaining activation in cMCL (n=93). On the top of the heatmap, we show the 6.1Mb genomic block gaining activation in 2p25.*** **D - Top:** ***Differentially expressed genes between cMCL and nnMCL in each of the three cohorts of transcriptional data of MCL patients.*** **Bottom:** ***Compartment type tracks on all the MCL samples under study. Eigenvalue subtraction between representative cMCL and nnMCL samples highlighting the 6.1Mb region gaining activity in the former.***

Discussion

We present a comprehensive analysis of the dynamic genome architecture reorganization during normal human B cell differentiation and upon neoplastic transformation into CLL and MCL. The integration of 3D genome data with nine additional omics layers including DNA methylation, chromatin accessibility, six histone modifications and gene expression, has allowed us to obtain new insights into 3D genome functional compartmentalization, cellular transitions across B cell differentiation and 3D genome aberrations in neoplastic B cells. We initially explored the distribution of Hi-C eigenvector coefficients and identified that a categorization into three components seemed to be more appropriate than the well-established dichotomous separation of the genome into A and B compartments (Lieberman-Aiden et al., 2009). Between the active (A) and repressed (B) compartments, we revealed the presence of an intermediate (I) component which contained more inter-compartment interactions than fully active or inactive chromatin, and is enriched in H3K27me3 located within poised promoters and polycomb-repressive chromatin states. Thus, this I-type compartment may represent a labile state of the high-order chromatin organization that may evolve either into active or inactive chromatin compartments. The existence of an intermediate compartment may be supported by several lines of published evidence. For example, during T cell commitment, a correlation between intermediate compartment scores with intermediate levels of gene expression was observed (Hu et al., 2018). Recently, using super-resolution imaging, it was found that some compartments could belong to active or inactive states depending on the observed cell (Nir et al., 2018), which could resemble an intermediate compartment in a population-based analysis such as Hi-C. Finally, these evidences are also in line with the observation that the polycomb repressive complex forms discrete subnuclear chromatin domains (Boettiger et al., 2016;

Kundu et al., 2017; Wani et al., 2016) that can be dynamically modulated during cell differentiation (Mas et al., 2018; Rada-Iglesias et al., 2018).

The three compartments had extensive modulation during human B cell differentiation, a process whose 3D genome architecture has been previously studied in cell lines and primary mouse cells (Johanson et al., 2018; Kieffer-Kwon et al., 2013; Lin et al., 2012; Martin et al., 2015; Mumbach et al., 2017; Stadhouders et al., 2018) or during the human germinal center reaction (Bunting et al., 2016). We observed that 28.1% of the genome is dynamically altered in particular B cell maturation transitions, a magnitude that is in line with compartment transitions observed during the differentiation of human embryonic stem cells into four cell lineages (Dixon et al., 2015) or the reprogramming of mouse somatic cells into induced pluripotent stem cells (Krijger et al., 2016; Stadhouders et al., 2018), but lower than an analysis of compartment transitions across 21 human cells and tissues, which reached 60% of the genome (Schmitt et al., 2016). The compartment modulation linked to B cell maturation was mainly related to two phenomena, a large-scale activation from NBC to GCBC and a reversion of the 3D genome organization of MBC back to the one observed in less mature NBC. As the number of mid-range 3D interactions upon activation has been suggested to decrease (Le Dily et al., 2014), our result on the GCBC structural activation supports a previous study in which the chromatin structure of GCBC undergoes global de-compaction (Bunting et al., 2016). In this context, TFs have been described to act as the architects instructing structural changes in the genome (Natoli, 2010) and a recent report has described that TFs are able to drive topological genome reorganizations even before detectable changes in gene expression (Stadhouders et al., 2018). A detailed analysis of regions that become exclusively active in GCBC as compared to any other B cell subpopulation under study revealed an enrichment in TF binding motifs of MEF2 and POU families, which have been described to play a key role in the germinal center formation (Song and Matthias, 2018). In line with this

important role of TFs in activating chromatin in GCBC, we also identified that NFAT and TCF binding motifs are enriched in those compartments specifically activated in CLL, and these TFs have also been previously linked to** de novo **active regulatory elements in CLL and its pathobiology (Beekman et al., 2018a). All these results are concordant with studies in which lineage-restricted transcription factors have been proposed to establish and maintain genome architecture of specific lineages (Heinz et al., 2010; Johanson et al., 2018; Montefiori et al., 2016; Natoli, 2010). The outcome of the germinal center reaction are PC and MBC, which are phenotypically and functionally distinct subpopulations. GCBC and PC show an overall high level of conservation of their 3D genome organization, but the differentiation into MBC is related to extensive changes. Remarkably, we observed roughly three quarters of the changes in MBC compartments reverted back to the compartment profile observed in NBC. This reversibility of the higher-order chromatin structure is very much in line with the previously observed similarity of histone modifications, chromatin accessibility and gene expression profiles in NBC and MBC. In sharp contrast to this congruent behavior of chromatin-based traits, DNA methylation is rather different between NBC and MBC, as this mark follows an accumulative pattern during cell differentiation (Kulis et al., 2015; Shearstone et al., 2011) and can be used to faithfully track the lineage trajectory of the cells (Gaiti et al., 2019).

We describe that B cell neoplasms show tumor-specific changes in the 3D genome organization that can span over large DNA stretches and contain genes linked to their pathogenesis. Of particular interest was the observation of the structural activation of 6.1Mb affecting the entire chromosome band 2p25.2 in aggressive cMCL, which contains the** SOX11 **oncogene, a biomarker whose expression defines this MCL subtype (Fernandez et al., 2010) and plays key functional roles in its pathogenesis (Beekman et al., 2018b). Although the** SOX11 **oncogene expression is related to the presence

of active histone modifications in the promoter region (Vegliante et al., 2011) and the establishment of novel 3D loops with a distant enhancer element (Queirós et al., 2016), our finding indicates that such looping is embedded into long-range alterations in the 3D genome structure. This change is not only linked to SOX11 overexpression, but seems to be related to the simultaneous overexpression of multiple genes within the target region. This phenomenon of long-range epigenetic changes has been observed at the DNA methylation level, as the hypermethylation over one chromosomal band of 4Mb that has been linked to silencing of several genes in colorectal cancer (Frigola et al., 2006). Additionally, in prostate cancer, long-range chromatin activation or inactivation analyzed by histone modifications has been shown to target oncogenes, microRNAs and cancer biomarker genes (Bert et al., 2013). The presence of large-range epigenetic remodeling in cancer (Achinger-Kawecka et al., 2016; Bert et al., 2013; Dallosso et al., 2009; Frigola et al., 2006; Hitchins et al., 2007; Novak et al., 2008; Rafique et al., 2015; Seng et al., 2008; Stransky et al., 2006; Taberlay et al., 2016) shall support a more generalized use of genome-wide chromosome conformation capture techniques as part of the global characterization of primary human tumors. Beyond the identification of a concerted deregulation of multiple contiguous genes with a potential role in cancer biology, targeting long-range aberrations in the 3D genome structure may itself represent a therapeutic target.

In conclusion, we provide an integrative and functional view of the 3D genome topology during human B cell differentiation and neoplastic transformation. Beyond revealing the presence of a novel compartment related to the polycomb-repressive complex, our analysis points to a highly dynamic 3D genome organization in normal B cells, including extensive activation from NBC to GCBC and a reversibility in MBC. In neoplastic cells from CLL and MCL, we identify disease and subtype-specific change in the 3D genome organization, which include large chromatin blocks

containing genes playing key roles in their pathogenesis and clinical behavior.

.

Acknowledgments

This research was funded by the European Union's Seventh Framework Programme through the Blueprint Consortium (grant agreement 282510), the World Wide Cancer Research Foundation Grant No. 16-1285 (to J.I.M-S.), the ERC (grant agreement 609989 to M.A.M-R.), European Union's Horizon 2020 research and innovation programme (grant agreement 676556 to M.A.M-R.). We also knowledge the support of Spanish Ministerio de Ciencia, Innovación y Universidades through SAF2012-31138 and SAF2017-86126-R to J.I.M-S., SAF2015-64885-R to E.C., BFU2017-85926-P to M.A.M-R. and PMP15/00007 to E.C. which is part of Plan Nacional de I+D+I and co-financed by the ISCIII-Sub-Directorate General for Evaluation and the European Regional Development Fund (FEDER-"Una manera de Hacer Europa") (to E.C.), the International Cancer Genome Consortium (Chronic Lymphocytic Leukemia Genome consortium to E.C.), La Caixa Foundation (CLLEvolution-HE17-00221, to E.C.). Furthermore, the authors would like to thank the support of the Generalitat de Catalunya Suport Grups de Recerca AGAUR 2017-SGR-736 (to J.I.M.-S.), 2017-SGR-1142 (to E.C.) and 2017-SGR-468 (to E.C.), the Accelerator award CRUK/AIRC/AECC joint funder-partnership, the CERCA Programme/Generalitat de Catalunya and CIBERONC (CB16/12/00225, CB16/12/00334 and CB16/12/00489). R.V-B. (BES-2013-064328) and P.S-V. (BES-2014-070327) were supported by a predoctoral FPI Fellowship from the Spanish Government and N.R. by the Acció instrumental d'incorporació de científics i tecnòlegs PERIS 2016 from the Generalitat de Catalunya. The authors thank the Barcelona Supercomputing Center for access to computational resources. This work was partially developed at the

Centro Esther Koplowitz (CEK, Barcelona, Spain). CRG acknowledges support from 'Centro de Excelencia Severo Ochoa 2013-2017', SEV-2012-0208 and the CERCA Programme/Generalitat de Catalunya. We thank Marta Kulis for critical reading of this manuscript.

Author Contributions

X.A., F.P., S.B., D.C., E.C., contributed to sample collection as well as to their biological and clinical annotation; R.V-B., N.V-D., N.R., R.B., H.G.S., I.G., E.C., and J.I.M.-S. performed, coordinated and/or supported in situ Hi-C, histone mark, ATAC-seq, 4C-seq, methylome and transcriptome data generation; R.V-B., P.S-V., M.D-S., V.C., G.C., I.F., P.C., R.B., M.A.M-R., and J.I.M.-S. performed, coordinated and/or supported computational data analysis; R.V-B., P.S-V. M.D-S., I.F., P.C., E.C., M.A.M-R., J.I.M-S. participated in the study design and/or data interpretation. M.A.M-R. and J.I.M-S. directed the research and wrote the manuscript together with R.V-B. and P.S-V.

Declaration of Interests

The authors declare no competing interests.

Data availability

In situ Hi-C data generated within this study is in the process of being deposited in a public data repository. The remaining raw data analyzed in this study has been deposited at the European Genome-Phenome Archive (EGA, http://www.ebi.ac.uk/ega/), which is hosted at the European Bioinformatics Institute (EBI). We have not yet created a specific dataset unifying all the raw data of our study, but all epigenomic data normal B cells, CLL and MCL can be found under accession numbers EGAS00001000326 (ChIPseq), EGAS00001001596 (ATACseq), EGAS00001000418 (WGBS)

and EGAS00001000327 (RNAseq). It is important to note here that the consent agreements signed by the participants in the BLUEPRINT project do not allow anonymous access to the raw data. If reviewers wish to access raw data, they will need to disclose their identity. To help reviewers to visualize the generated EBF1 models using TADkit and the multi omics data in a browser session we have created a website accompanying the manuscript: http://resources.idibaps.org/paper/dynamics-of-genome-architecture-and-chromatin-function-during-human-b-cell-differentiation-and-neoplastic-transformation, which can be anonymously accessed.

Methods

Isolation of B cell subpopulations for *in situ* Hi-C experiment

Four B cell subpopulations spanning mature normal B cell differentiation were sorted for in situ Hi-C as previously described (Kulis et al., 2015). Briefly, peripheral blood B cell subpopulations i.e. naive B cells (NBC) and memory B cells (MBC) were obtained from buffy coats for healthy adult male donors from 56 to 61 years of age, obtained from Banc de Sang i Teixits (Catalunya, Spain). Germinal center B cells (GCBC) and plasma cells (PC) were isolated from tonsils of male children undergoing tonsillectomy (from 2 to 12 years of age), obtained from the Clínica Universidad de Navarra (Pamplona, Spain). Samples were cross-linked before FACS sorting, to separate each of the B cell subpopulations, and afterwards were snap frozen and kept at -80°C. Three replicates per B cell subpopulation were processed and each replicate was derived from individual donors with the exception of plasma cells, for which two of the three replicates proceeded from the pool of four different donors. The use of the samples analyzed in the present study was approved by the ethics committee of the Hospital Clínic de Barcelona and Clínica Universidad de Navarra.

Patient Samples

The samples from CLL (n=7) (Beekman et al., 2018a) and MCL (n=5) patients were obtained from cryopreserved mononuclear cells from the Hematopathology collection registered at the Biobank (Hospital Clínic-IDIBAPS; R121004-094). All samples were >85% tumor content. Clinical and biological characteristics of the patients are shown in Table S5.

The enrolled patients gave informed consent for scientific study following the ICGC guidelines and the ICGC Ethics and Policy committee (Consortium, 2010). This study was approved by the clinical research ethics committee of the Hospital Clínic of Barcelona.

In situ Hi-C

In situ ***Hi-C was performed based on the previously described protocol (Rao et al., 2014). Two million of cross-linked cells per sample were used as starting material. Chromatin was digested adding 100U DpnII (New England BioLabs) on overnight incubation. After the fill-in with bio-dCTP (Life-Technologies, 19518-018), nuclei were centrifuged 5 minutes, 3000rpm at 4°C and ligation was performed for 4 hours at 16°C adding 2µl of 2000U/µl T4 DNA ligase on total 1.2mL of ligation mix (120µl of 10X T4 DNA ligase buffer; 100µl of 10% Triton X-100; 12µl of 10mg/ml BSA; 966µl of H_20). Following ligation, nuclei were pelleted and resuspended with 400µl 1X NEBuffer2 (New England BioLabs). Then, 10µl of RNAseA (10mg/ml) was added to the nuclei and incubated during 15 minutes at 37°C while shaking (300rpm), and after that 20µl of proteinakse K (10mg/mL) was added and incubated overnight at 65°C while shaking (600rpm). After reversion of the cross-link, DNA was purified by phenol/chloroform/isoamyl alcohol and DNA was precipitated by adding to the upper aqueous phase: 0.1X of 3M sodium acetate pH 5.2, 2.5X of pure ethanol and 50µg/ml glycogen. Samples were mixed and incubated***

overnight at -80°C. Next, samples were centrifuged 30 minutes at 13,000rpm at 4°C and pellet was washed with 1mL of EtOH 70% followed by a 15 minutes centrifugation at 13,000rpm at 4°C. The supernatant was discarded and the pellet air-dried for 5 minutes and resuspended in 130µl of 1X Tris buffer (10 mM TrisHCl, pH 8.0), which to be fully dissolved was incubated at 37°C for 15 minutes. Purified DNA was sonicated using Covaris S220, and then the final volume was adjusted to 300µl with 1X Tris buffer. Sonicated DNA was mixed with washed magnetic streptavidin T1 beads (total of 100µl 10mg/ml beads), split in two tubes (150µl each), and incubated for 30 minutes at room temperature (RT) under rotation. Subsequently, beads were separated on the magnet, the supernatant discarded and the DNA was washed with 400µl of BB 1X, twice. Sonicated DNA conjugated with beads was washed with 100µl of 1X T4 DNA ligase buffer, pooling the two tubes per condition. After that, beads were reclaimed in end-repair mix. Once incubated during 30 minutes at RT the beads were washed twice with 400µl of BB 1X. Then, beads were washed with 100µl of NEBuffer2 and reclaimed in A-tailing mix, incubated during 30 minutes at 37°C and washed twice with 400µl of BB 1X, followed by a wash in 100µl of 1X T4 DNA ligase buffer. Afterwards, the beads were resuspended in 50µl of 1X Quick ligation buffer, 2.5µl of Illumina adaptors and 4,000U of T4 DNA ligase and incubated during 15 minutes at RT. Then, beads were washed twice with 400µl BB 1X and resuspended in 30µl of 1X Tris buffer. In the end, libraries were amplified by eight cycle of PCR using 8.3µl of beads and pooling a total of 4 PCRs per sample. The PCR products were mixed by pipetting with an equal volume of AMPure XP beads and incubated at RT for 5 minutes. Beads were washed with 700µl of EtOH 70%, without mixing, twice, and left the EtOH evaporate at RT without over-drying the beads (aprox. 4 minutes). Finally, the beads were resuspended with 30µl 1X Tris buffer, incubated during 5 minutes and supernatant containing the purified library was transferred in a new tube and stored at -20°C. DNA was

quantified by Qubit dsDNA High Sensitivity Assay, the library profile was evaluated on the Bioanalayzer 2100 and the ligation was assessed. Libraries were sequenced on HiSeq 2500. Table S1 summarizes the number of reads sequenced and quality metrics for each B cell subpopulation replicate and B cell neoplasm.

Hi-C data pre-processing, normalization and interaction calling

The sequencing reads of Hi-C experiments were processed with TADbit (Serra et al., 2017). Briefly, sequencing reads were aligned to the reference genome (GRCh38) applying a fragment-based strategy; dependent on GEM mapper (Marco-Sola et al., 2012). The mapped reads were filtered to remove those resulting from unspecified ligations, errors or experimental artefacts. Specifically, we applied seven different filters using the default parameters in TADbit: self-circles, dangling ends, errors, extra dangling-ends, over-represented, duplicated and random breaks (Serra et al., 2017). Hi-C data were normalized using the OneD correction (Vidal et al., 2018) at 100Kb of resolution to remove known experimental biases. The significant Hi-C interactions were called with the analyzeHiC *function of the HOMER software suite (Heinz et al., 2010), binned at 10Kb of resolution and with the default* p*-value threshold of 0.001.*

Reproducibility of Hi-C replicas

The agreement between Hi-C replicates was assessed using the reproducibility score (Yan et al., 2017). The RS is a measure of matrix similarity ranging between 0 (totally different matrices) and 1 (identical matrices). A genome-wide RS was defined for each experiment as the average RS between pairs of corresponding normalized chromosome matrix (Figure S1A, Figure S3B and S3D). Then, the matrix representing all the genome-

wide RSs was analyzed using a hierarchical clustering algorithm with the Ward's agglomeration method using hclust function from R stats package.

ChIP-seq and ATAC-seq data generation and processing

ChIP-seq of the six different histone marks and ATAC-seq data were generated as described in (http://www.blueprint-epigenome.eu/index.cfm?p=7BF8A4B6-F4FE-861A-2AD57A08D63D0B58) (Beekman et al., 2018a). Briefly, fastq files of ChIP-seq data were aligned to the GRCh38 reference genome using bwa 0.7.7 (Li and Durbin, 2009), PICARD (http://broadinstitute.github.io/picard/) and SAMTOOLS (Li et al., 2009), and wiggle plots were generated (using PhantomPeakQualTools R package) as described (http://dcc.blueprint-epigenome.eu/#/md/methods). Peaks of the histone marks were called as described in http://dcc.blueprint-epigenome.eu/#/md/methods using MACS2 (version 2.0.10.20131216) (Zhang et al., 2008) with input control. ATAC-seq fastq files were aligned to genome build GRCh38 using bwa 0.7.7 (parameters: -q 5 –P -a 480) (Li and Durbin, 2009) and SAMTOOLS v1.3.1 (default settings) (Li et al., 2009). BAM files were sorted and duplicates were masked using PICARD tools v2.8.1 with default settings (http://broadinstitute.github.io/picard/). Finally, low quality and duplicate reads were removed using SAMTOOLS v1.3.1 (parameters: -b -F 4 -q 5,-b, -F 1024) (Li et al., 2009). ATAC-seq peaks were determined using MACS2 (version 2.1.1.20160309, parameters: -g hs q 0.05 -f BAM –nomodel - shift -96 extsize 200 - keep -dup all) without input (Zhang et al., 2008).
For each mark a set of consensus peaks (chr1-22) present in the normal B cells (n=12 biologically independent samples for histone marks and n=15 biologically independent samples for ATAC-seq) was generated by merging the locations of the separate peaks per individual sample. Also, a second set of consensus peaks was generated taking into account normal B cells, CLL

(n=7 biologically independent samples) and MCL (n=5 biologically independent samples). For the histone marks, the number of reads per sample per consensus peak was calculated using the genomecov function of bedtools suite (Quinlan and Hall, 2010). For ATAC-seq, the number of insertions of the TN5 transposase per sample per consensus peaks was calculated determining the estimated insertion sites (shifting the start of the first mate 4bp downstream), followed by the genomecov function of bedtools suite (Quinlan and Hall, 2010). The number of consensus peaks for normal B cell samples were 46,184 (H3K4me3), 44,201 (H3K4me1), 72,222 (H3K27ac), 25,945 (H3K36me3), 40,704 (H3K9me3), 20,994 (H3K27me3), 99,327 (ATAC-seq), while the number of consensus peaks for normal B cells, CLL and MCL samples were 53,241 (H3K4me3), 54,653 (H3K4me1), 106,457 (H3K27ac), 50,530 (H3K36me3), 137,933 (H3K9me3), 117,560 (H3K27me3), 140,187 (ATAC-seq). Using DESeq2 R package (Love et al., 2014), counts for all consensus peaks were transformed by means of the variance stabilizing transformation (VST) with blind dispersion estimation. Principal component analysis (PCAs) were generated with the prcomp function from the stats package in R using the VST values.

RNA-seq data generation and processing

Single-stranded RNA-seq data were generated as previously described (Ecker et al., 2017). Briefly, RNA was extracted using TRIZOL (Life Technologies) and libraries were prepared using TruSeq Stranded Total RNA kit with Ribo-Zero Gold (Illumina). Adapter-ligated libraries were amplified and sequenced using 100bp single-end reads. RNA-seq data of the 24 samples, some (n=19) mined from a previous study (Beekman et al., 2018a), were aligned to the reference human genome build GRCh38 (Table S5). Signal files were produced and gene quantifications (gencode 22, 60,483 genes) were calculated as described (http://dcc.blueprint-

epigenome.eu/#/md/methods) using the GRAPE2 pipeline with STAR-RSEM profile (adapted from the ENCODE Long RNA-Seq pipeline). The expected counts and fragments per kilobase million (FPKM) estimates were used for downstream analysis. The PCA of the RNA-seq data was generated with the prcomp *function from the stats package in R in the 12 analyzed normal B cell samples or 24 analyzed normal and neoplastic B cell samples.*

WGBS data generation and processing

WGBS was generated as previously described (Kulis et al., 2015). Mapping and determination of methylation estimates were performed as described (http://dcc.blueprint-epigenome.eu/#/md/methods) using GEM3.0. Per sample, only methylation estimates of CpGs with ten or more reads were used for downstream analysis. The principal component analysis (PCA) of the DNA methylation data was generated with the prcomp *function from the stats package in R using methylation estimates of 15,089,887 CpGs (chr1-22) with available methylation estimates in all 12 analyzed normal B cell samples or 14,088,025 CpGs (chr1-22) in all 24 analyzed normal and neoplastic B cell samples.*

Definition of sub-nuclear genome compartmentalization

The segmentation of the genome into compartments was determined as previously described (Lieberman-Aiden et al., 2009). In short, normalized chromosome-wide interaction matrices at 100Kb resolution were transformed into Pearson correlation matrices. These correlation matrices were then used to perform PCA for which the first eigenvector (EV) normally delineates genome segregation. All EVs were visually inspected to ensure that the EV selected corresponded to genomic compartments (Lieberman-Aiden et al., 2009). Since the sign of the EV is arbitrary, a rotation factor based on the

histone mark H3K4me1 signal and ATAC-seq signal were applied to correctly call the identity of the compartments. A Pearson correlation coefficient was computed between the EVs for each pair of merged B cell subpopulation (Figure S1C). Each merged sample was also correlated with its replica (Figure S1C). The multi-modal distribution of the EV coefficients from the B cells dataset was modelled as a Gaussian mixture with three components (k=3). To estimate the mixture distribution parameters, an Expectation Maximization algorithm using the normalmixEM *function from the* mixtools **R package was applied (Benaglia et al., 2009).**

A Bayesian Information Criterion (BIC) was computed for the specified mixture models of clusters (from 1 to 10) using mclustBIC **function from** mclust **package in R (Scrucca et al., 2016) (Figure S1D). Three underlying structures were defined; an alternative compartmentalization into A-type (with the most positive EV values), B-type (with the most negative EV values) and I-type (an intermediate-valued region with a distinct distribution) compartments. Two intersection values (IV1, IV2) were defined at the intersection points between two components. The mean IV1 and IV2 values across all the B cell replicas (n=12) were then used as standard thresholds to categorize the data into the three different components (that is, A-type compartment was defined for EV values between +1.00 and +0.63, I-type compartment as of "Intermediate" was defined for EV values between +0.63 and -0.43, and B-type compartment was defined for EV values between -0.43 and -1.00) (Figure S1E).**

Characterizing compartment types in B cells by integrating nine omics layers

***Given a set of peaks as previous defined by Beekman** et al., **(Beekman et al., 2018a) from nine different omic layers including six histone marks (H3K4me3, H3K4me1, H3K27ac, H3K36me3, H3K9me3, H3K27me3), gene accessibility (ATAC-seq), gene expression (RNA-seq) and DNA*

methylation (WGBS), a** bedmap **function from BEDOPS software (Neph et al., 2012) was applied to get the mean scoring peak over the 100Kb intervals genome-wide. Next, Pearson correlation coefficients were computed between the EV coefficients and the mean scoring value of each epigenetic mark at 100Kb intervals (Figure S1B). Finally, the mean scoring values were normalized by the total sum of the values for each mark and grouped by the three defined genomic compartments (A, I, B-type; Figure 1G). A Wilcoxon test was used to compute the significance between all the possible pairwise comparisons of the signal distribution.

Compartment Interaction Score (C-Score)

The compartment score is defined as the ratio of contacts between regions within the same compartment (intra-compartment contacts) over the total chromosomal contacts per compartment (intra-compartment + inter-compartment). To compute the compartment score, all the compartments that shared the same genomic segmentation were merged.

Chromatin states enrichment by genomic compartments

The genome was segmented into 12 different chromatin states at 200bp interval as previously described (Beekman et al., 2018a). The active promoter and strong enhancer1 were merged as a unique state, giving a total of 11 chromatin states. The genome compartmentalization was next split into 4 groups; 3 conserved groups, in which the B cells samples shared A-type compartment (n=6,409), B-type compartment (n=6,267) or I-type compartment (n=5,467) and a dynamic group (n=7,099) of non-conserved compartmentalization among B cells subpopulation. Each group was correlated with the defined 11 chromatin states using** foverlaps **function from** data.table **R package. The frequency of each chromatin state (corrected by the

total frequency in the genome) was computed per each genomic compartment. The chromatin state score is thus the median frequency of the three replicas scaled by the columns and the rows using scale *function from* base **R package.**

Description of chromatin states in the intermediate (I)-type compartment

200bp-windows containing poised promoter (n=547) or polycomb repressed (n=11,665) chromatin states were extracted from the NBC intermediate compartments (n=1,885). From those regions, two main sub-groups were distinguished according to the chromatin state shown in the next state of differentiation (GCBC): (1) those regions that maintained their chromatin state (poised promoter or polycomb repressed), and (2) those regions that changed their chromatin state; which were further classified into three categories: (i) I-related chromatin states (poised promoter or polycomb repressed), (ii) B-related chromatin states (repressive heterochromatin and low signal heterochromatin), (iii) A-related chromatin states (active promoter/strong enhancer1, weak promoter, strong enhancer2, transcription transition, transcription elongation and weak transcription). Finally, the fold-change of related chromatin states between GCBC and NBC was computed.

Analysis of chromatin state dynamics upon B cell differentiation

B cell differentiation axis was divided into two main branches: (i) NBC-GCBC-PC, (ii) NBC-GCBC-MBC. Both branches presented a common step from NBC to GCBC and then a divergence step in PC or MBC. The 5,445 common compartments from both branches were considered for the analysis.

The general modulation of chromatin structure was drawn using the** alluvial **function from** alluvial **R package.

Transcription factor analyses

From GCBC-specific 937 active compartments (B to A-type, n =18; B to I-type, n=512 and I to A-type, n=407) were narrowed down to 171 peaks due to the following filtering steps: (i) only the 200bp-windows contain active promoter, strong enhancer1 and strong enhancer2 chromatin states were retained (n=1,907 regions). (ii) Regions where H3K27ac peaks were differentially enriched in GCBC replicates compared to the rest of normal B cell subpopulations (FDR<0.05) computed using DEseq2 R package (Love et al., 2014) were retained. (iii) Regions with a presence of ATAC-seq peaks in at least two GCBC replicates were retained (n=171 peaks). The background considered was the rest of the ATAC-seq peaks (n=268) presented at the 1,907 regions in at least two GCBC replicates.

From CLL-specific 48 active compartments (in normal B cells defined as I-type: n=28 and B-type: n=20), were narrowed down to 25 peaks due to the following filtering steps: (i) Regions where H3K27ac peaks were differentially enriched (FDR<0.05) comparing CLL from all normal B cells and MCL using DEseq2 package (Love et al., 2014), (ii) Regions where ATAC-seq peaks were presented in at least five CLL (n=25). The background considered was all the resting ATAC-seq peaks (n=28) on the 48 compartments presented in at least five CLL.

On both analysis, FASTA sequences of targeted regions (GCBC-specific regions and CLL-specific regions) were extracted using** getfasta **function from** bedtools **suite (Quinlan and Hall, 2010) using GRCh38 as reference assembly. An analysis of motif enrichment was done by the** AME-MEME **suite

(McLeay and Bailey, 2010) using non-redundant transcription factor (TF) binding profiles of Homo sapiens *Jaspar 2018 database (Khan et al., 2018) as a reference motif database. The database contained a set of 537 DNA motifs. Maximum odd scores were used as a scoring method and one-tailed Wilcoxon rank-sum as motif enrichment test. Only TF genes that were expressed (FPKM median values>1) were included.*

TCF4 binding motif example from *KSR2* gene

A FASTA sequences of 25 ATAC-seq peaks detected in CLL-specific active compartments were extracted using GRCh38 as reference assembly. A search of individual motif occurrences analysis was done using AME-FIMO *suite (Grant et al., 2011) library(BSgenome.Hsapiens.UCSC.hg38,masked) with a custom random model (letter frequencies: A, 0.262: C, 0.238: G, 0.238 and T, 0.262). A* p-value<0.0001 *was established as a threshold to determine 23 significant motif occurrences where TCF4 binding motif (MA0830.1) was one of the top candidates.*

Log-ratio of normalized interactions in the AICDA *regulatory landscape*
Normalized Hi-C maps were analyzed at 50Kb of resolution at the specific genomic region, chr12:8,550,000-9,050,000 (GRCh38), from the four B cell subpopulations. A logarithmic ratio of the contact maps was computed between NBC and GCBC and GCBC with PC and MBC. The result array was convolved with a 1-dimensional Gaussian filter of standard deviation (sigma) of 1.0 using and interpolated with a nearest-neighbor approach using scipyndimage *Python package.*

Statistical testing for detecting significant changed compartment regions

Briefly, 100Kb regions that had at least one missing value among the compared samples were removed from the analysis. Then, two different groups were defined, case and control, according to the case-control pair analyzed. A T-test was computed to compare each case-control pair, and the resulting p-values were adjusted using the false discovery rate (FDR) (Benjamini and Hochberg, 1995). The regions with significantly different means and fold changes were selected based on two specific thresholds: a p-adjustment value less than 0.05 and a fold change greater than 0.4. The results were then generated for a total of 4 different case-control pairs.

- ***(I) control: all regions conserved across all B cell samples without missing values in CLL (A-type, n=3,967, I-type, n=4,301 and B-type, n=5,226), case: all CLL regions non-conserved in B cell samples (n=3,217). The analysis resulted in 348 B cell_CLL significantly changed regions.***
- ***(II) control: all regions conserved across all B cell samples without missing values in MCL (A-type n=6,167, I-type n=5,299, B-type n=5,812), case: all MCL regions non-conserved in B cell samples (n=4,716). The analysis resulted in 82 B cell_MCL significantly changed regions.***
- ***(III) control: B cell-CLL significantly changed regions (n=348) - MCL-CLL overlapping (n=31) = B cell-CLL specific regions (n=317), case: MCL regions (A-type n=97, I-type n=154, B-type n=61; total n=312). The analysis resulted in 89 B cell_CLL-specific regions.***
- ***(IV) control: B cell-MCL significantly changed regions (n=82) - MCL-CLL overlapping (n=31) = B cell-MCL specific regions (n=51), case: CLL regions (n=41). The analysis resulted in 3 B cell_MCL-specific regions.***

Integrative 3D modelling of EBF1 and structural analysis

Hi-C interactions matrices from the merging of three replicas of NBC and the seven cases of CLL were used to model chr5:158,000,000:160,000,000 (GRCh38) at 5Kb of resolution. For NBC and CLL merged Hi-C interaction maps, a MMP score was calculated to assess the modeling potential of the region, resulting in 0.79 for NBC and 0.84 for CLL indicative of good quality Hi-C contact maps for accurate 3D reconstruction (Trussart et al., 2015). Next, this region was modelled using a restraint-based modelling approach as implemented in TADbit (Serra et al., 2017), where the experimental frequencies of interaction are transformed into a set of spatial restraints (Baù and Marti-Renom, 2012). Briefly, each 5Kb bin of the interaction Hi-C map was represented as a spherical particle in the model, which resulted in 400 particles each of radius equal to 25nm. All the particles in the models were restrained in the space based on the frequency of the Hi-C contacts, the chain connectivity and the excluded volume. The TADbit optimal parameters (maxdist=-1.0; lowfreq=1.0; upfreq=200; and dcutoff=150) resulted in the best Spearman correlations of 0.61 (NBC) and 0.63 (CLL) between the Hi-C interaction map and the models contact map. Next, a total of 5,000 models per cell type were generated, and the top 1,000 models that best satisfied the imposed restraints were retained for the analysis. To assess the structural similarities among the 3D models, the distance root-mean-square deviations (dRMSD) value was computed for all the possible pairs of top models (1,000 in NBC and 1,000 in CLL) and a hierarchical clustering algorithm was applied on the resulting dRMSD matrix using** *ward.D* **method from stats package in R (Figure S4C). The convex hull volume spanned by the 81 particles of the** *EBF1* **gene (chr5:158,695,000-159,000,000, GRCh38) was computed in each model using the** *convexhull* **function from the** *scipy.spatial* **Python package (Figure 5G).

Differential Gene expression analyses

Differentially expressed genes were defined using the DEseq2 R package (Love et al., 2014), *nbinomWaldTest*, ***on all the genes. Then, the genes present on the compartments of interest were selected and Benjamini y Hochberg (BH) test (FDR<0.05) was applied. In detail, expected counts were used on the following considered comparisons: (i) for GCBC-specific activate compartments, GCBC samples (n=3) versus the rest of normal B cells samples (NBC, PC, MBC; n=9); (ii) for CLL-specific active compartments, CLL samples (n=7) versus the rest of the samples (normal B cells and MCL, n=17); (iii) for CLL-specific inactive compartments, all normal B cells and MCL samples (total n=17) versus CLL samples (n=7) and (iv) for cMCL, cMCL (n=2) versus nnMCL (n=3) samples were studied. Then, the expression of the genes differentially expressed on each comparison of interest was assessed. Only genes that were expressed (FPKM median values>1) were included.***

The *findOverlaps* ***function from*** *GenomicRanges* ***R package (Lawrence et al., 2013) was used to annotated genes that overlapped with these defined regions. One tailed Monte-Carlo method was applied to evaluate the significant number of differentially expressed genes in CLL-specific compartments (this process was randomly repeated 10,000 times).***

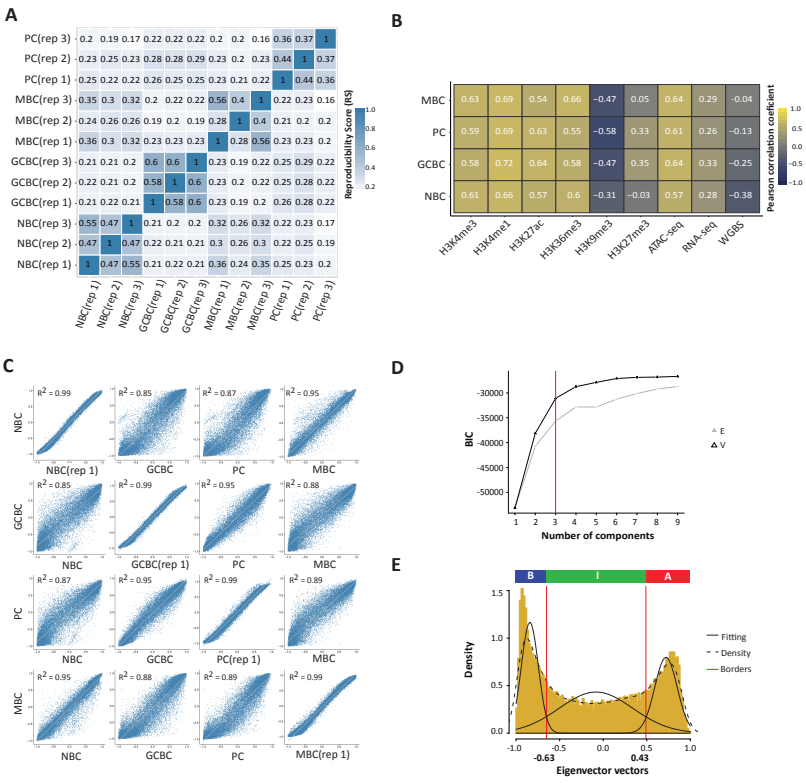
Defining de novo (in)active regions in sub-type specific neoplastic group

MCL and CLL patient samples were grouped according to their biological and clinical characteristics. This classification resulted in two conventional (c) and three leukemic non-nodal (nn) MCL cases and two IGVH-unmutated (u) and five IGVH-mutated (m) CLL cases.

First, the non-assigned neoplasia compartments were removed from the analysis. A sample homogenization was applied to reduce the intra-subtype variance; the samples that presented a difference of EV smaller than 0.4 were retained (91.29% in MCL, 87.1% CLL). Next, to study the inter-subtype variance, the mean of the EV from each subtype of B cell malignancy was computed. Significant regions were determined if the difference between the two subtypes (cMCL vs nnMCL and uCLL vs mCLL) was equal or higher than 0.4, which resulted in 673 regions in MCL and 47 in CLL. MCL-subtype specific regions where split into two groups according to the value of its EV coefficient (n=435 region called cMCL gain, n=238 regions called nnMCL gain). The distribution and the frequency of the significantly changed regions were studied per chromosome and compared with the probability of finding them by chance in each chromosome. N-subsamples of 100Kb size were selected from the GRCh38 genome and their frequency was calculated per chromosome (this process was randomly repeated 10,000 times). One tailed Monte-Carlo method was applied to compute *p* ***-values. The*** *findOverlaps* ***function from*** *GenomicRanges* ***R package (Lawrence et al. 2013) was next used to annotate protein coding genes that overlapped with these defined regions. Differentially expressed genes among cMCL and nnMCL on chr2:2,700,000-8,800,000 (GRCh38) was compute using*** *Deseq2* ***(Love et al. 2014) (using a FDR<0.05). The expression analysis was validated on two independent published cohorts, i.e.: a series with 30 conventional and 24 leukemic non-nodal mantle cell lymphoma (GEO GSE79196) from peripheral blood (Navarro et al., 2017) and a second series from the lymphoma/leukemia molecular profiling project (LLMPP) (GEO GSE93291) (Scott et al., 2017). The microarrays were normalized using the R*** *frma* ***(McCall et al., 2010) method and*** *limma* ***R package (Smyth, 2004) was used to identify differentially expressed genes with adjusted*** *p* ***-value<0.05. Standardized expression matrices were used to do the heatmaps using*** *pheatmap* ***R package. Gene differentially expressed on the identified cohort: [1] RNAseq from***

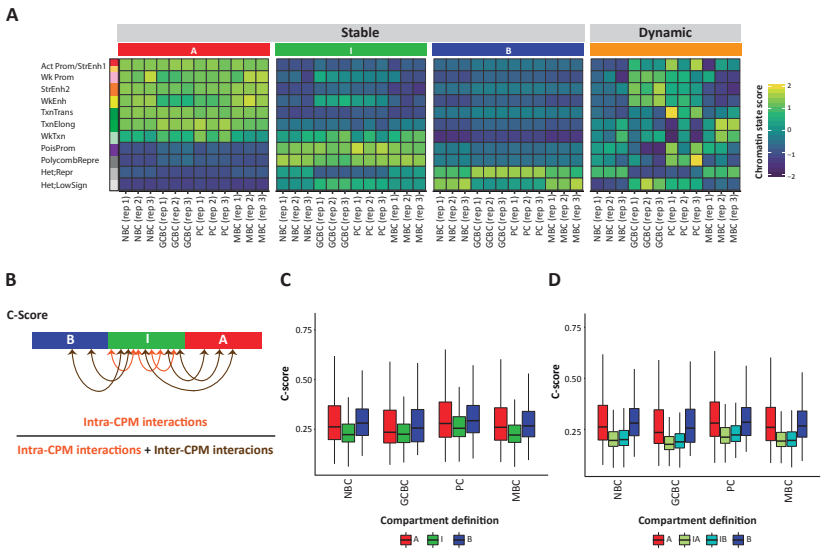
BLUEPRINT data, [2] peripheral blood and [3] LLMPP. The magnitude of the compartmentalization change was calculated subtracting the EV of cMCL1 and nnMCL2. The karyotype and the chromosome 2 were designed using the karyoploteR library (Gel and Serra, 2017).

Extended Data Figure 1



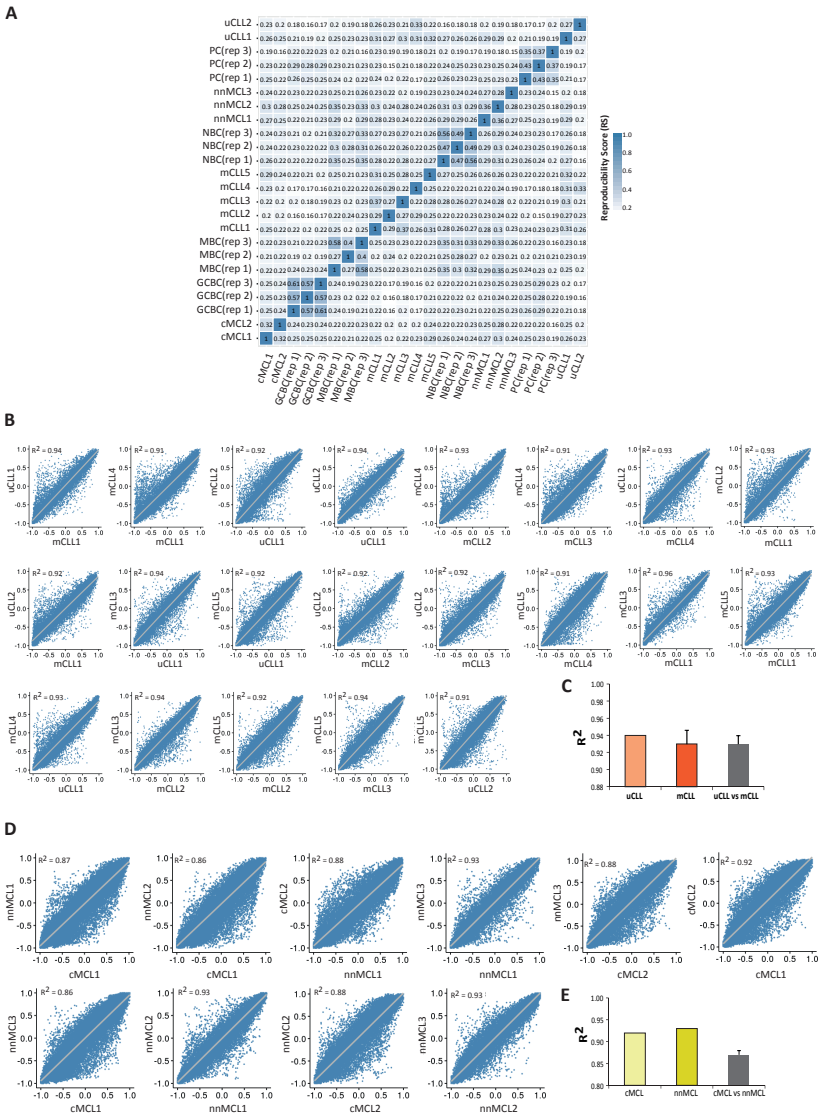
Extended Data Figure 1. A. ***Average genome-wide reproducibility score matrix of each B cell subpopulation replicates at 100Kb resolution. The reproducibility score ranging between 0 (totally different matrices) and 1 (identical matrices).*** **B.** ***Pearson correlation between the eigenvector coefficients, which defines 3D compartments per B cell subpopulation, with six histone marks, chromatin accessibility (ATAC-seq), gene expression (RNA-seq) and DNA methylation (WGBS). Positive values of the eigenvector show higher correlation with H3K4me1 (enhancer mark) and chromatin accessibility.*** **C.** ***Genome-wide scatterplots of coefficients from the first eigenvector showing the correlation between pairs of B cell subpopulations at 100Kb resolution. The squared correlation coefficient (R^2) is indicated.*** **D.** ***Bayesian Information Criterion (BIC) plot for the equal (E) and unequal (V) variance model parameterization ranged from 1 to 10 clusters.*** **E.** ***Compartment definition model. The x-axis shows the distribution of the eigenvector coefficients and the y-axis indicates the density. The fitting model proposed is highlighted using solid black line. The red lines mark the intersection points (EV1 = -0.63 and EV2 = 0.43) used to distinguish the three different compartments (A-type, I-type, B-type).***

Extended Data Figure 2



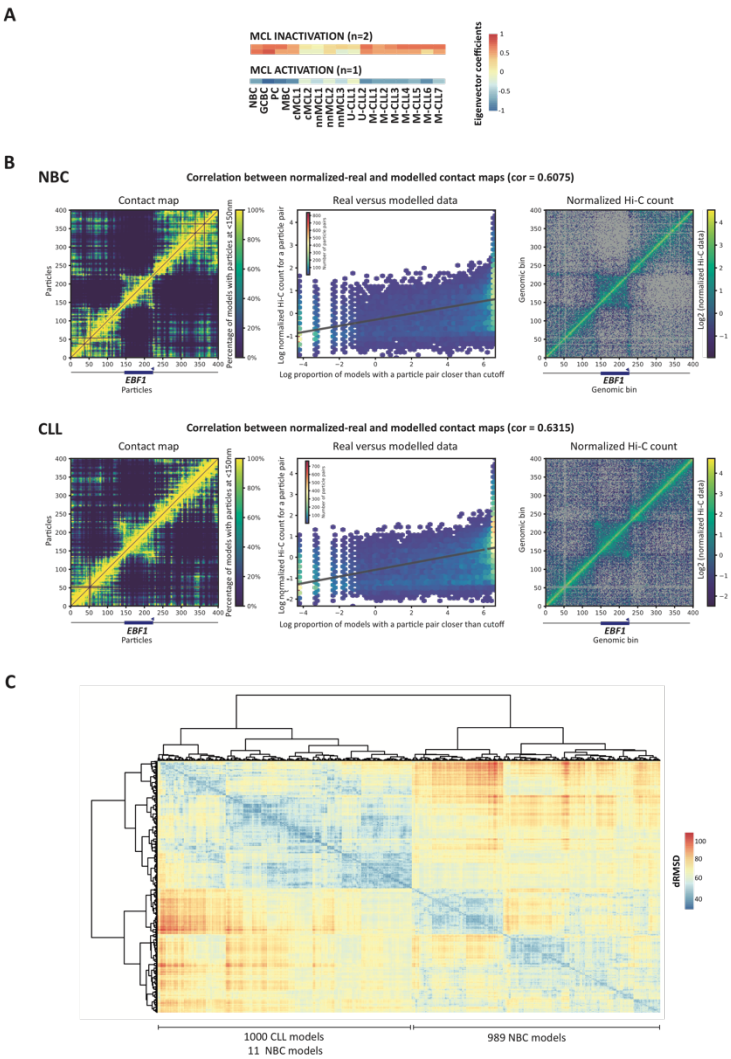
Extended Data Figure 2. A. ***Functional validation of the conserved (A-type, I-type and B-type) and dynamic compartments in all B cell subpopulations replicates using eleven different chromatin states. The chromatin state score is normalized by sample and chromatin state.*** **B.** ***C-score. Method defined by the ratio of contacts between regions within the same compartment (intra-compartment contacts) over the total chromosomal contacts per compartments (intra- and inter-chromosomal interactions).*** **C.** ***C-score distributions on the three defined compartments A-type, I-type and B-type.*** **D.** ***C-score distributions segmenting the I-type compartment onto positive (IA) or negative (IB) eigenvector coefficients.***

Extended Data Figure 4



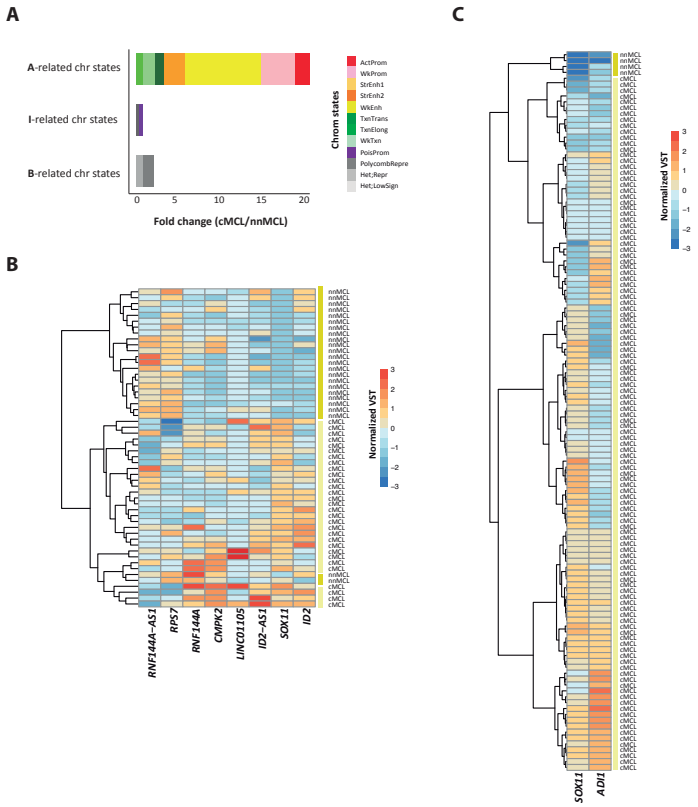
Extended Data Figure 4. A. ***Average genome-wide reproducibility score matrix of each B cell subpopulation replicates and B cell neoplasia at 100Kb. The reproducibility score ranging between 0 (totally different matrices) and 1 (identical matrices).*** **B/D.** ***Genome-wide scatterplots of the first eigenvector showing the correlation between pairs of each B cell malignancy samples at 100Kb resolution. CLL (B). MCL (D). The squared correlation coefficient (R^2) is indicated.*** **C/E.** ***Mean and standard deviation of the squared correlation coefficients calculated intra- or inter- each neoplasia subtype. CLL (C). MCL (E).***

Extended Data Figure 5



Extended Data Figure 5. A. ***Heatmap showing the eigenvector coefficients of the compartments losing (***top***) or gaining (***bottom***) activation specifically in MCL.*** **B.** ***Correlation between normalized Hi-C and modeled contact maps in*** *EBF1* ***regulatory landscape.*** **Left:** ***Contact map computed from the restrained-based model.*** **Middle:** ***Scatterplot of Hi-C normalized map versus modeled contact data with linear regression.*** **Right:** ***Normalized Hi-C data.*** **Top:** ***NBC.*** **Bottom:** ***CLL. The position of*** *EBF1* ***is indicated in blue at the bottom of the matrix plots.*** **C.** ***Heatmap of the hierarchical clustering of the dRMSD values computed for all the possible pairs of generated models (1,000 in NBC and 1,000 in CLL).***

Extended Data Figure 7



Extended Data Figure 7. A. ***Bar graphs represent the fold change between cMCL and nnMCL of each three groups of chromatin states (arranged by their relationship to the A-type, I-type and B-type compartments). Active Promoter, Weak Promoter, Strong Enhancer 1, Strong Enhancer 2, Weak Enhancer, Transcription Transition, Transcription Elongation, Weak Transcription were A-type compartment-related states. Heterochromatin; Repressed and Heterochromatin; Low signal were B-type compartment-related states. Poised Promoter or Polycomb repressed chromatin states were I-type compartment-related states.*** **B/C.** ***Heatmaps of the differentially expressed gens between MCL samples classified as cMCL (light yellow) and nnMCL (dark yellow) subtypes. Peripheral blood (*** **B***) and LLMPP (*** **C***) cohorts. The VST values were normalized by genes.***

Supplementary Tables

Supplementary Table 1. *In situ* ***Hi-C experimental quality metrics.***

Supplementary Table 2. ***GCBC specific 3D active compartments on a three-column bed file format (chromosome, start position and end position).***

Supplementary Table 3. ***List of the identified enriched binding motifs expressed on GCBC.***

Supplementary Table 4. ***Genes differentially upregulated (FDR<0.05) in GCBC specific regions. The coordinates of the compartment or compartments each gene belongs to is indicated***

Supplementary Table 5. ***Patient characteristics and general overview of the omic layers analyzed.***

Supplementary Table 6. ***Genes differentially expressed (FDR<0.05) at CLL-specific inactive compartments. The coordinates of the compartment or compartments each gene belongs to is indicated***

Supplementary Table 7. ***Genes differentially expressed (FDR<0.05) at CLL-specific active compartments. The coordinates of the compartment or compartments each gene belongs to is indicated***

Supplementary Table 8. ***List of the identified enriched binding motifs expressed on CLL-specific active compartments.***

References (main text and methods)

Achinger-Kawecka, J., Taberlay, P.C., and Clark, S.J. (2016). Alterations in Three-Dimensional Organization of the Cancer Genome and Epigenome. Cold Spring Harb. Symp. Quant. Biol. 81, 41–51.

Adams, D., Altucci, L., Antonarakis, S.E., Ballesteros, J., Beck, S., Bird, A., Bock, C., Boehm, B., Campo, E., Caricasole, A., et al. (2012). BLUEPRINT to decode the epigenetic signature written in blood. Nat. Biotechnol. 30, 224–226.

Agirre, X., Meydan, C., Jiang, Y., Garate, L., Doane, A.S., Li, Z., Verma, A., Paiva, B., Martín-Subero, J.I., Elemento, O., et al. (2019). Long non-coding RNAs discriminate the stages and gene regulatory states of human humoral immune response. Nat. Commun. 10, 821.

Andrey, G., and Mundlos, S. (2017). The three-dimensional genome: regulating gene expression during pluripotency and development. Development 144, 3646–3658.

Balsas, P., Palomero, J., Eguileor, Á., Rodríguez, M.L., Vegliante, M.C., Planas-Rigol, E., Sureda-Gómez, M., Cid, M.C., Campo, E., and Amador, V. (2017). SOX11 promotes tumor protective microenvironment interactions through CXCR4 and FAK regulation in mantle cell lymphoma. Blood 130, 501–513.

Baù, D., and Marti-Renom, M.A. (2012). Genome structure determination via 3C-based data integration by the Integrative Modeling Platform. Methods 58, 300–306.

Beekman, R., Chapaprieta, V., Russiñol, N., Vilarrasa-Blasi, R., Verdaguer-Dot, N., Martens, J.H.A., Duran-Ferrer, M., Kulis, M., Serra, F., Javierre, B.M., et al. (2018a). The reference epigenome and regulatory chromatin landscape of chronic lymphocytic leukemia. Nat. Med. 24, 868–880.

Beekman, R., Amador, V., and Campo, E. (2018b). SOX11, a key oncogenic factor in mantle cell lymphoma. Curr. Opin. Hematol. 25, 299–306.

Benaglia, T., Chauveau, D., Hunter, D.R., and Young, D. (2009). mixtools : An R Package for Analyzing Finite Mixture Models. J. Stat. Softw. 32.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J. R. Stat. Soc. Ser. B 57, 289–300.

Bert, S.A., Robinson, M.D., Strbenac, D., Statham, A.L., Song, J.Z., Hulf, T., Sutherland, R.L., Coolen, M.W., Stirzaker, C., and Clark, S.J. (2013). Regional Activation of the Cancer Genome by Long-Range Epigenetic Remodeling. Cancer Cell 23, 9–22.

Boettiger, A.N., Bintu, B., Moffitt, J.R., Wang, S., Beliveau, B.J., Fudenberg, G., Imakaev, M., Mirny, L.A., Wu, C., and Zhuang, X. (2016). Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. Nature 529, 418–422.

Brescia, P., Schneider, C., Holmes, A.B., Shen, Q., Hussein, S., Pasqualucci, L., Basso, K., and Dalla-Favera, R. (2018). MEF2B Instructs Germinal Center Development and Acts as an Oncogene in B Cell Lymphomagenesis. Cancer Cell 34, 453-465.e9.

Bunting, K.L., Soong, T.D., Singh, R., Jiang, Y., Béguelin, W., Poloway, D.W., Swed, B.L., Hatzi, K., Reisacher, W., Teater, M., et al. (2016). Multi-tiered Reorganization of the Genome during B Cell Affinity Maturation Anchored by a Germinal Center-Specific Locus Control Region. Immunity 45, 497–512.

Chiorazzi, N., and Ferrarini, M. (2011). Cellular origin(s) of chronic lymphocytic leukemia: cautionary notes and additional considerations and possibilities. Blood 117, 1781–1791.

Consortium, T.I.C.G. (2010). International network of cancer genome projects. Nature 464, 993–998.

Dallosso, A.R., Hancock, A.L., Szemes, M., Moorwood, K., Chilukamarri, L., Tsai, H.-H., Sarkar, A., Barasch, J., Vuononvirta, R., Jones, C., et al. (2009). Frequent Long-Range Epigenetic Silencing of Protocadherin Gene Clusters on Chromosome 5q31 in Wilms' Tumor. PLoS Genet. 5, e1000745.

Dekker, J., Rippe, K., Dekker, M., and Kleckner, N. (2002). Capturing chromosome conformation. Science (80-.). 295, 1306–1311.

Denker, A., and de Laat, W. (2016). The second decade of 3C technologies: detailed insights into nuclear organization. Genes Dev. 30, 1357–1382.

Le Dily, F.L., Baù, D., Pohl, A., Vicent, G.P., Serra, F., Soronellas, D., Castellano, G., Wright, R.H.G., Ballare, C., Filion, G., et al. (2014). Distinct structural transitions of chromatin topological domains correlate with coordinated hormone-induced gene regulation. Genes Dev. 28, 2151–2162.

Dixon, J.R., Jung, I., Selvaraj, S., Shen, Y., Antosiewicz-Bourget, J.E., Lee,

A.Y., Ye, Z., Kim, A., Rajagopal, N., Xie, W., et al. (2015). Chromatin architecture reorganization during stem cell differentiation. Nature 518, 331–336.

Ecker, S., Chen, L., Pancaldi, V., Bagger, F.O., Fernández, J.M., Carrillo de Santa Pau, E., Juan, D., Mann, A.L., Watt, S., Casale, F.P., et al. (2017). Genome-wide analysis of differential transcriptional and epigenetic variability across human immune cell types. Genome Biol. 18, 18.

Ernst, J., and Kellis, M. (2017). Chromatin-state discovery and genome annotation with ChromHMM. Nat. Protoc. 12, 2478–2492.

Fernandez, V., Salamero, O., Espinet, B., Sole, F., Royo, C., Navarro, A., Camacho, F., Bea, S., Hartmann, E., Amador, V., et al. (2010). Genomic and Gene Expression Profiling Defines Indolent Forms of Mantle Cell Lymphoma. Cancer Res. 70, 1408–1418.

Frigola, J., Song, J., Stirzaker, C., Hinshelwood, R.A., Peinado, M.A., and Clark, S.J. (2006). Epigenetic remodeling in colorectal cancer results in coordinate gene suppression across an entire chromosome band. Nat. Genet. 38, 540–549.

Gaiti, F., Chaligne, R., Gu, H., Brand, R.M., Kothen-Hill, S., Schulman, R.C., Grigorev, K., Risso, D., Kim, K.-T., Pastore, A., et al. (2019). Epigenetic evolution and lineage histories of chronic lymphocytic leukaemia. Nature 569, 576–580.

Gel, B., and Serra, E. (2017). karyoploteR: an R/Bioconductor package to plot customizable genomes displaying arbitrary data. Bioinformatics 33, 3088–3090.

Grant, C.E., Bailey, T.L., and Noble, W.S. (2011). FIMO: scanning for occurrences of a given motif. Bioinformatics 27, 1017–1018.

Gutierrez, A., Tschumper, R.C., Wu, X., Shanafelt, T.D., Eckel-Passow, J., Huddleston, P.M., Slager, S.L., Kay, N.E., and Jelinek, D.F. (2010). LEF-1 is a prosurvival factor in chronic lymphocytic leukemia and is expressed in the preleukemic state of monoclonal B-cell lymphocytosis. Blood 116, 2975–2983.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and 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, 576–589.

Hitchins, M.P., Lin, V.A., Buckle, A., Cheong, K., Halani, N., Ku, S., Kwok, C.T., Packham, D., Suter, C.M., Meagher, A., et al. (2007). Epigenetic inactivation of a cluster of genes flanking MLH1 in microsatellite-unstable colorectal cancer. Cancer Res. 67, 9107–9116.

Hu, G., Cui, K., Fang, D., Hirose, S., Wang, X., Wangsa, D., Jin, W., Ried, T., Liu, P., Zhu, J., et al. (2018). Transformation of Accessible Chromatin and 3D Nucleome Underlies Lineage Commitment of Early T Cells. Immunity 48, 227-242.e8.

Imakaev, M., Fudenberg, G., McCord, R.P., Naumova, N., Goloborodko, A., Lajoie, B.R., Dekker, J., and Mirny, L.A. (2012). Iterative correction of Hi-C data reveals hallmarks of chromosome organization. Nat. Methods 9, 999–1003.

Jares, P., Colomer, D., Campo, E., Jares, P., Colomer, D., and Campo, E. (2012). Molecular pathogenesis of mantle cell lymphoma Find the latest version : Review series Molecular pathogenesis of mantle cell lymphoma. 122, 3416–3423.

Javierre, B.M., Burren, O.S., Wilder, S.P., Kreuzhuber, R., Hill, S.M., Sewitz, S., Cairns, J., Wingett, S.W., Várnai, C., Thiecke, M.J., et al. (2016). Lineage-Specific Genome Architecture Links Enhancers and Non-coding Disease Variants to Target Gene Promoters. Cell 167, 1369-1384.e19.

Johanson, T.M., Lun, A.T.L., Coughlan, H.D., Tan, T., Smyth, G.K., Nutt, S.L., and Allan, R.S. (2018). Transcription-factor-mediated supervision of global genome architecture maintains B cell identity. Nat. Immunol. 19, 1257–1264.

Johanson, T.M., Chan, W.F., Keenan, C.R., and Allan, R.S. (2019). Genome organization in immune cells: unique challenges. Nat. Rev. Immunol. 19, 448–456.

Khan, A., Fornes, O., Stigliani, A., Gheorghe, M., Castro-Mondragon, J.A., van der Lee, R., Bessy, A., Chèneby, J., Kulkarni, S.R., Tan, G., et al. (2018). JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. Nucleic Acids Res. 46, D260–D266.

Kieffer-Kwon, K.R., Tang, Z., Mathe, E., Qian, J., Sung, M.H., Li, G., Resch, W., Baek, S., Pruett, N., Grøntved, L., et al. (2013). Interactome maps of mouse gene regulatory domains reveal basic principles of transcriptional regulation. Cell 155, 1507–1520.

Kipps, T.J., Stevenson, F.K., Wu, C.J., Croce, C.M., Packham, G., Wierda,

W.G., O'Brien, S., Gribben, J., and Rai, K. (2017). Chronic lymphocytic leukaemia. Nat. Rev. Dis. Prim. 3, 16096.

Klein, U., Tu, Y., Stolovitzky, G.A., Keller, J.L., Haddad, J., Miljkovic, V., Cattoretti, G., Califano, A., and Dalla-Favera, R. (2003). Transcriptional analysis of the B cell germinal center reaction. Proc. Natl. Acad. Sci. 100, 2639–2644.

Krijger, P.H.L., Di Stefano, B., De Wit, E., Limone, F., Van Oevelen, C., De Laat, W., and Graf, T. (2016). Cell-of-origin-specific 3D genome structure acquired during somatic cell reprogramming. Cell Stem Cell 18, 597–610.

Kulis, M., Heath, S., Bibikova, M., Queirós, A.C., Navarro, A., Clot, G., Martínez-Trillos, A., Castellano, G., Brun-Heath, I., Pinyol, M., et al. (2012). Epigenomic analysis detects widespread gene-body DNA hypomethylation in chronic lymphocytic leukemia. Nat. Genet. 44, 1236–1242.

Kulis, M., Merkel, A., Heath, S., Queirós, A.C., Schuyler, R.P., Castellano, G., Beekman, R., Raineri, E., Esteve, A., Clot, G., et al. (2015). Whole-genome fingerprint of the DNA methylome during human B cell differentiation. Nat. Genet. 47, 746–756.

Kundu, S., Ji, F., Sunwoo, H., Jain, G., Lee, J.T., Sadreyev, R.I., Dekker, J., and Kingston, R.E. (2017). Polycomb Repressive Complex 1 Generates Discrete Compacted Domains that Change during Differentiation. Mol. Cell 65, 432-446.e5.

Kurosaki, T., Shinohara, H., and Baba, Y. (2010). B Cell Signaling and Fate Decision. Annu. Rev. Immunol. 28, 21–55.

Lawrence, M., Huber, W., Pagès, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M.T., and Carey, V.J. (2013). Software for Computing and Annotating Genomic Ranges. PLoS Comput. Biol. 9, e1003118.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079.

Li, R., Liu, Y., Hou, Y., Gan, J., Wu, P., and Li, C. (2018). 3D genome and its disorganization in diseases. Cell Biol. Toxicol. 34, 351–365.

Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M.,

Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., et al. (2009). Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. Science (80-.). 326, 289–293.

Lin, Y.C., Benner, C., Mansson, R., Heinz, S., Miyazaki, K., Miyazaki, M., Chandra, V., Bossen, C., Glass, C.K., and Murre, C. (2012). Global changes in the nuclear positioning of genes and intra-and interdomain genomic interactions that orchestrate B cell fate. Nat. Immunol. 13, 1196–1204.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550.

Marco-Sola, S., Sammeth, M., Guigó, R., and Ribeca, P. (2012). The GEM mapper: fast, accurate and versatile alignment by filtration. Nat. Methods 9, 1185–1188.

Margueron, R., and Reinberg, D. (2011). The Polycomb complex PRC2 and its mark in life. Nature 469, 343–349.

Martin, P., McGovern, A., Orozco, G., Duffus, K., Yarwood, A., Schoenfelder, S., Cooper, N.J., Barton, A., Wallace, C., Fraser, P., et al. (2015). Capture Hi-C reveals novel candidate genes and complex long-range interactions with related autoimmune risk loci. Nat. Commun. 6, 1–7.

Mas, G., Blanco, E., Ballaré, C., Sansó, M., Spill, Y.G., Hu, D., Aoi, Y., Le Dily, F., Shilatifard, A., Marti-Renom, M.A., et al. (2018). Promoter bivalency favors an open chromatin architecture in embryonic stem cells. Nat. Genet. 50, 1452–1462.

Matthias, P., and Rolink, A.G. (2005). Transcriptional networks in developing and mature B cells. Nat. Rev. Immunol. 5, 497–508.

McCall, M.N., Bolstad, B.M., and Irizarry, R.A. (2010). Frozen robust multiarray analysis (fRMA). Biostatistics 11, 242–253.

McLeay, R.C., and Bailey, T.L. (2010). Motif Enrichment Analysis: a unified framework and an evaluation on ChIP data. BMC Bioinformatics 11, 165.

Mockridge, C.I., Potter, K.N., Wheatley, I., Neville, L.A., Packham, G., Stevenson, K., Dc, W., and Stevenson, F.K. (2013). Reversible anergy of sIgM-mediated signaling in the two subsets of CLL defined by V H -gene mutational status Reversible anergy of sIgM-mediated signaling in the two subsets of CLL defined by V H -gene mutational status. 109, 4424–4431.

Montefiori, L., Wuerffel, R., Roqueiro, D., Lajoie, B., Guo, C., Gerasimova, T., De, S., Wood, W., Becker, K.G., Dekker, J., et al. (2016). Extremely Long-Range Chromatin Loops Link Topological Domains to Facilitate a Diverse Antibody Repertoire. Cell Rep. 14, 896–906.

Mumbach, M.R., Satpathy, A.T., Boyle, E.A., Dai, C., Gowen, B.G., Cho, S.W., Nguyen, M.L., Rubin, A.J., Granja, J.M., Kazane, K.R., et al. (2017). Enhancer connectome in primary human cells identifies target genes of disease-associated DNA elements. Nat. Genet. 49, 1602–1612.

Muzio, M., Apollonio, B., Scielzo, C., Frenquelli, M., Vandoni, I., Boussiotis, V., Caligaris-Cappio, F., and Ghia, P. (2008). Constitutive activation of distinct BCR-signaling pathways in a subset of CLL patients: a molecular signature of anergy. Blood 112, 188–195.

Natoli, G. (2010). Maintaining cell identity through global control of genomic organization. Immunity 33, 12–24.

Navarro, A., Clot, G., Royo, C., Jares, P., Hadzidimitriou, A., Agathangelidis, A., Bikos, V., Darzentas, N., Papadaki, T., Salaverria, I., et al. (2012). Molecular Subsets of Mantle Cell Lymphoma Defined by the IGHV Mutational Status and SOX11 Expression Have Distinct Biologic and Clinical Features. Cancer Res. 72, 5307–5316.

Navarro, A., Clot, G., Martínez-Trillos, A., Pinyol, M., Jares, P., González-Farré, B., Martínez, D., Trim, N., Fernández, V., Villamor, N., et al. (2017). Improved classification of leukemic B-cell lymphoproliferative disorders using a transcriptional and genetic classifier. Haematologica 102, e360–e363.

Neph, S., Kuehn, M.S., Reynolds, A.P., Haugen, E., Thurman, R.E., Johnson, A.K., Rynes, E., Maurano, M.T., Vierstra, J., Thomas, S., et al. (2012). BEDOPS: high-performance genomic feature operations. Bioinformatics 28, 1919–1920.

Nir, G., Farabella, I., Pérez Estrada, C., Ebeling, C.G., Beliveau, B.J., Sasaki, H.M., Lee, S.D., Nguyen, S.C., McCole, R.B., Chattoraj, S., et al. (2018). Walking along chromosomes with super-resolution imaging, contact maps, and integrative modeling. PLOS Genet. 14, e1007872.

Novak, P., Jensen, T., Oshiro, M.M., Watts, G.S., Kim, C.J., and Futscher, B.W. (2008). Agglomerative epigenetic aberrations are a common event in human breast cancer. Cancer Res. 68, 8616–8625.

Oakes, C.C., and Martin-Subero, J.I. (2018). Insight into origins, mechanisms, and utility of DNA methylation in B-cell malignancies. Blood

132, ***999–1006.***

Oakes, C.C., Seifert, M., Assenov, Y., Gu, L., Przekopowitz, M., Ruppert, A.S., Wang, Q., Imbusch, C.D., Serva, A., Brocks, D., et al. (2016). DNA methylation dynamics during B cell maturation underlie a continuum of disease phenotypes in chronic lymphocytic leukemia. Nat. Genet. *48,* ***253–264.***

Palomero, J., Vegliante, M.C., Eguileor, A., Rodríguez, M.L., Balsas, P., Martínez, D., Campo, E., and Amador, V. (2016). SOX11 defines two different subtypes of mantle cell lymphoma through transcriptional regulation of BCL6. Leukemia *30,* ***1596–1599.***

Peric-Hupkes, D., Meuleman, W., Pagie, L., Bruggeman, S.W.M., Solovei, I., Brugman, W., Gräf, S., Flicek, P., Kerkhoven, R.M., van Lohuizen, M., et al. (2010). Molecular Maps of the Reorganization of Genome-Nuclear Lamina Interactions during Differentiation. Mol. Cell *38,* ***603–613.***

Puente, X.S., Jares, P., and Campo, E. (2018). Chronic lymphocytic leukemia and mantle cell lymphoma: Crossroads of genetic and microenvironment interactions. Blood *131,* ***2283–2296.***

Queirós, A.C., Beekman, R., Vilarrasa-Blasi, R., Duran-Ferrer, M., Clot, G., Merkel, A., Raineri, E., Russiñol, N., Castellano, G., Beà, S., et al. (2016). Decoding the DNA Methylome of Mantle Cell Lymphoma in the Light of the Entire B Cell Lineage. Cancer Cell *30,* ***806–821.***

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26,* ***841–842.***

Rada-Iglesias, A., Grosveld, F.G., and Papantonis, A. (2018). Forces driving the three-dimensional folding of eukaryotic genomes. Mol. Syst. Biol. *14,* ***e8214.***

Rafique, S., Thomas, J.S., Sproul, D., and Bickmore, W.A. (2015). Estrogen-induced chromatin decondensation and nuclear re-organization linked to regional epigenetic regulation in breast cancer. Genome Biol. *16,* ***1–19.***

Rao, S.S.P., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., et al. (2014). A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. Cell *159,* ***1665–1680.***

Rowley, M.J., and Corces, V.G. (2018). Organizational principles of 3D genome architecture. Nat. Rev. Genet. *19,* ***789–800.***

Le Roy, C., Deglesne, P.-A., Chevallier, N., Beitar, T., Eclache, V., Quettier, M., Boubaya, M., Letestu, R., Levy, V., Ajchenbaum-Cymbalista, F., et al. (2012). The degree of BCR and NFAT activation predicts clinical outcomes in chronic lymphocytic leukemia. Blood 120, 356–365.

Royo, C., Navarro, A., Clot, G., Salaverria, I., Giné, E., Jares, P., Colomer, D., Wiestner, A., Wilson, W.H., Vegliante, M.C., et al. (2012). Non-nodal type of mantle cell lymphoma is a specific biological and clinical subgroup of the disease. Leukemia 26, 1895–1898.

Ryba, T., Hiratani, I., Lu, J., Itoh, M., Kulik, M., Zhang, J., Schulz, T.C., Robins, A.J., Dalton, S., and Gilbert, D.M. (2010). Evolutionarily conserved replication timing profiles predict long-range chromatin interactions and distinguish closely related cell types. Genome Res. 20, 761–770.

Schmitt, A.D., Hu, M., Jung, I., Xu, Z., Qiu, Y., Tan, C.L., Li, Y., Lin, S., Lin, Y., Barr, C.L., et al. (2016). A Compendium of Chromatin Contact Maps Reveals Spatially Active Regions in the Human Genome. Cell Rep. 17, 2042–2059.

Schubart, K., Massa, S., Schubart, D., Corcoran, L.M., Rolink, A.G., and Matthias, P. (2001). B cell development and immunoglobulin gene transcription in the absence of Oct-2 and OBF-1. Nat. Immunol. 2, 69–74.

Schultze, J., Nadler, L.M., and Gribben, J.G. (1996). B7-mediated costimulation and the immune response. Blood Rev. 10, 111–127.

Scott, D.W., Abrisqueta, P., Wright, G.W., Slack, G.W., Mottok, A., Villa, D., Jares, P., Rauert-Wunderlich, H., Royo, C., Clot, G., et al. (2017). New Molecular Assay for the Proliferation Signature in Mantle Cell Lymphoma Applicable to Formalin-Fixed Paraffin-Embedded Biopsies. J. Clin. Oncol. 35, 1668–1677.

Scrucca, L., Fop, M., Murphy, T.B., and Raftery, A.E. (2016). mclust 5: Clustering, Classification and Density Estimation Using Gaussian Finite Mixture Models. R J. 8, 289–317.

Seifert, M., Sellmann, L., Bloehdorn, J., Wein, F., Stilgenbauer, S., Dürig, J., and Küppers, R. (2012). Cellular origin and pathophysiology of chronic lymphocytic leukemia. 209, 2183–2198.

Seng, T.J., Currey, N., Cooper, W.A., Lee, C.S., Chan, C., Horvath, L., Sutherland, R.L., Kennedy, C., McCaughan, B., and Kohonen-Corish, M.R.J. (2008). DLEC1 and MLH1 promoter methylation are associated with poor prognosis in non-small cell lung carcinoma. Br. J. Cancer 99, 375–382.

Serra, F., Baù, D., Goodstadt, M., Castillo, D., Filion, G., and Marti-Renom, M.A. (2017). Automatic analysis and 3D-modelling of Hi-C data using TADbit reveals structural features of the fly chromatin colors. PLoS Comput. Biol. 13, 1–17.

Shearstone, J.R., Pop, R., Bock, C., Boyle, P., Meissner, A., and Socolovsky, M. (2011). Global DNA Demethylation During Mouse Erythropoiesis in Vivo. Science (80-.). 334, 799–802.

De Silva, N.S., and Klein, U. (2015). Dynamics of B cells in germinal centres. Nat. Rev. Immunol. 15, 137–148.

Smyth, G.K. (2004). Linear Models and Empirical Bayes Methods for Assessing Differential Expression in Microarray Experiments. Stat. Appl. Genet. Mol. Biol. 3, 1–25.

Song, S., and Matthias, P.D. (2018). The Transcriptional Regulation of Germinal Center Formation. Front. Immunol. 9, 2026.

Stadhouders, R., Vidal, E., Serra, F., Di Stefano, B., Le Dily, F., Quilez, J., Gomez, A., Collombet, S., Berenguer, C., Cuartero, Y., et al. (2018). Transcription factors orchestrate dynamic interplay between genome topology and gene regulation during cell reprogramming. Nat. Genet. 50, 238–249.

Stransky, N., Vallot, C., Reyal, F., Bernard-Pierrot, I., De Medina, S.G.D., Segraves, R., De Rycke, Y., Elvin, P., Cassidy, A., Spraggon, C., et al. (2006). Regional copy number-independent deregulation of transcription in cancer. Nat. Genet. 38, 1386–1396.

Swerdlow, S. H., Campo, E., Harris, N. L., Jaffe, E.S., Pileri, S.A., Stein, H., Thiele, J., Arber, D.A., Hasserjian, R.P., Le Beau, M.M., Orazi, A., Siebert, R. (2017). WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues. International Agency for Research on Cancer, Lyon.

Szalaj, P., and Plewczynski, D. (2018). Three-dimensional organization and dynamics of the genome. Cell Biol. Toxicol. 34, 381–404.

Taberlay, P.C., Achinger-Kawecka, J., Lun, A.T.L., Buske, F.A., Sabir, K., Gould, C.M., Zotenko, E., Bert, S.A., Giles, K.A., Bauer, D.C., et al. (2016). Three-dimensional disorganization of the cancer genome occurs coincident with long-range genetic and epigenetic alterations. Genome Res. 26, 719–731.

Trussart, M., Serra, F., Bau, D., Junier, I., Serrano, L., Marti-Renom, M. a., Trussart, M., and Ba, D. (2015). Assessing the limits of restraint-based 3D

modeling of genomes and genomic domains. Nucleic Acids Res. 43, 3465–3477.

Vegliante, M.C., Royo, C., Palomero, J., Salaverria, I., Balint, B., Martín-Guerrero, I., Agirre, X., Lujambio, A., Richter, J., Xargay-Torrent, S., et al. (2011). Epigenetic Activation of SOX11 in Lymphoid Neoplasms by Histone Modifications. PLoS One 6, e21382.

Vegliante, M.C., Palomero, J., Perez-Galan, P., Roue, G., Castellano, G., Navarro, A., Clot, G., Moros, A., Suarez-Cisneros, H., Bea, S., et al. (2013). SOX11 regulates PAX5 expression and blocks terminal B-cell differentiation in aggressive mantle cell lymphoma. Blood 121, 2175–2185.

Vidal, E., le Dily, F., Quilez, J., Stadhouders, R., Cuartero, Y., Graf, T., Marti-Renom, M.A., Beato, M., and Filion, G.J. (2018). OneD: increasing reproducibility of Hi-C samples with abnormal karyotypes. Nucleic Acids Res. 46, e49–e49.

Wani, A.H., Boettiger, A.N., Schorderet, P., Ergun, A., Münger, C., Sadreyev, R.I., Zhuang, X., Kingston, R.E., and Francis, N.J. (2016). Chromatin topology is coupled to Polycomb group protein subnuclear organization. Nat. Commun. 7, 10291.

Wilker, P.R., Kohyama, M., Sandau, M.M., Albring, J.C., Nakagawa, O., Schwarz, J.J., and Murphy, K.M. (2008). Transcription factor Mef2c is required for B cell proliferation and survival after antigen receptor stimulation. Nat. Immunol. 9, 603–612.

Yaffe, E., and Tanay, A. (2011). Probabilistic modeling of Hi-C contact maps eliminates systematic biases to characterize global chromosomal architecture. Nat. Genet. 43, 1059–1065.

Yan, K.-K., Yardımcı, G.G., Yan, C., Noble, W.S., and Gerstein, M. (2017). HiC-spector: a matrix library for spectral and reproducibility analysis of Hi-C contact maps. Bioinformatics 33, 2199–2201.

de Yébenes, V.G., and Ramiro, A.R. (2006). Activation-induced deaminase: light and dark sides. Trends Mol. Med. 12, 432–439.

Ying, C.Y., Dominguez-Sola, D., Fabi, M., Lorenz, I.C., Hussein, S., Bansal, M., Califano, A., Pasqualucci, L., Basso, K., and Dalla-Favera, R. (2013). MEF2B mutations lead to deregulated expression of the oncogene BCL6 in diffuse large B cell lymphoma. Nat. Immunol. 14, 1084–1092.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein,

B.E., Nussbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based Analysis of ChIP-Seq (MACS). Genome Biol. 9, R137.

CHAPTER 2

Hierarchical chromatin organization detected by TADpole

Paula Soler-Vila, Pol Cuscó Pons, Irene Farabella, Marco Di Stefano, Marc A. Marti-Renom.

Hierarchical chromatin organization detected by TADpole

doi: ***https://doi.org/10.1101/698720***

Hierarchical levels of chromatin organization detected by TADpole.

*Paula Soler-Vila1,†, Pol Cuscó Pons2,†, Irene Farabella1, Marco Di Stefano1 and Marc A. Marti-Renom1,3,4,5,**

1 CNAG-CRG, Centre for Genomic Regulation (CRG), Barcelona Institute of Science and Technology (BIST), Barcelona, Spain.
2 Gastrointestinal and Endocrine Tumors Group, Vall d'Hebron Institute of Oncology (VHIO), Barcelona, Spain.
3 Centre for Genomic Regulation (CRG), Barcelona Institute of Science and Technology (BIST), Barcelona, Spain.
4 Universitat Pompeu Fabra (UPF), Barcelona, Spain.
5 ICREA, Barcelona, Spain.

† Joint first authors

** To whom correspondence should be addressed. Emails: martirenom@cnag.crg.eu & marco.distefano@cnag.crg.eu*

Abstract

The rapid development of chromosome conformation capture (3C-based) techniques as well as super-resolution imaging together with bioinformatics analyses has been fundamental for unveiling that chromosomes are organized into the so-called topologically associating domains or TADs. While these TADs appear as nested patterns in the 3C-based interaction matrices, the vast majority of available computational methods are based on the hypothesis that TADs are individual and unrelated chromatin structures. Here we introduce TADpole, a computational tool designed to identify and analyze the entire hierarchy of TADs in intrachromosomal interaction

matrices. TADpole combines principal component analysis and constrained hierarchical clustering to provide an unsupervised set of significant partitions in a genomic region of interest. TADpole identification of domains is robust to the data resolution, normalization strategy, and sequencing depth. TADpole domain borders are enriched in CTCF and cohesin binding proteins, while the domains are enriched in either H3K36me3 or H3k27me3 histone marks. We show TADpole usefulness by applying it to capture Hi-C experiments in wild-type and mutant mouse strains to pinpoint statistically significant differences in their topological structure.

Introduction

The organization of the genome in the cell nucleus has been shown to play a prominent role in the function of the cell. Increasing evidence indicates that genome architecture regulates gene transcription (1,2) with implications on cell-fate decisions (3-5), development (6), and disease occurrences such as developmental abnormalities (7,8) and neoplastic transformations (9-11).

The genome organization is characterized by complex and hierarchical layers (1). For example, fluorescence *in-situ* ***hybridization revealed that chromosomes are positioned in preferential areas of the nucleus called chromosome territories (12). This large-scale feature has been confirmed by high-throughput Chromosome Conformation Capture (Hi-C) experiments (13), that provided a genome-wide picture in which inter-chromosomal interactions are depleted relative to intra-chromosomal. Analysis of Hi-C data also revealed the segregation of the genome in multi-megabase compartments characterized by different GC-content, gene density, and chromatin marks (13-15). Microscopy approaches, in spite of considerable variability, have corroborated the spatial segregation of such compartments at the single cell level (16). At the sub-megabase level, Hi-C experiments also***

revealed the presence, validated by microscopy approaches (17-19), of self-interacting regions termed Topologically Associated Domains (TADs) (20,21). TADs are composed by dense chromatin interactions, which promote 3D spatial proximity between genomic *loci* ***that are distal in the linear genome sequence. Since many of these interacting*** *loci* ***are cis-regulatory elements, TADs are usually considered as the structural functional units of the genome that define the regulatory landscape (22,23), and are conserved across cell types and species (20,24). Moreover, TADs boundaries are often demarcated by housekeeping genes, transcriptional start sites and specific chromatin insulators proteins, such as CTCF factor and cohesin complex (20,25). TADs appear to be further organized in a hierarchical fashion. For example, in mammalian cells, concepts such as "metaTADs" (26) or "sub-TADs" (27) have been introduced. The former is used to define a superior hierarchy of domains within domains that are modulated during cell differentiation (26) while the latter to emphasize how and where the cis-regulatory elements establish physical interactions that contribute to gene regulation (27).***

Several computational methods to identify and characterize TADs from 3C-based interaction data have been reported (28,29). Based on different *a priori* ***assumptions on the TADs subdivision, these methods can be mainly classified as disjointed or overlapping. The former considers TADs as individual and unrelated structures with no possible mutual intersections (e.g. directionality index (DI) (20), insulation score (IS) (30), ClusterTAD (31), ICFinder (32)).***

The latter assume that TADs are overlapping and related structures with a shared content (e.g. Arrowhead (13,15), armatus (33), TADtree (34), 3DNetMod (35)). However, only few algorithms (CaTCH (36), GMAP (37), matryoshka (38), and PSYCHIC (39)) can identify nested domains

where each domain contain other sub-domain profiling a hierarchical chromatin architecture.

Here, we present TADpole, a bioinformatics tool to disentangle the full structural chromatin hierarchy that automatically determines an optimal division level. Notably, TADpole is robust both at technical and biological benchmarks based on a published study (29) and does not rely on mandatory parameters. We prove the effectiveness of TADpole to investigate the chromatin hierarchy in capture Hi-C data (cHi-C) (40) where the chromosome topology is altered with local genomic inversions that drive gene misexpression associated to congenital malformations in mouse (41).

Material and Methods

The TADpole pipeline

TADpole consists in three main steps (*Figure 1A*): (i) pre-processing of the input Hi-C dataset, (ii) constrained hierarchical clustering optimization, and (iii) genome segmentation. TADpole has been implemented as an R package available at https://github.com/3DGenomes/TADpole.

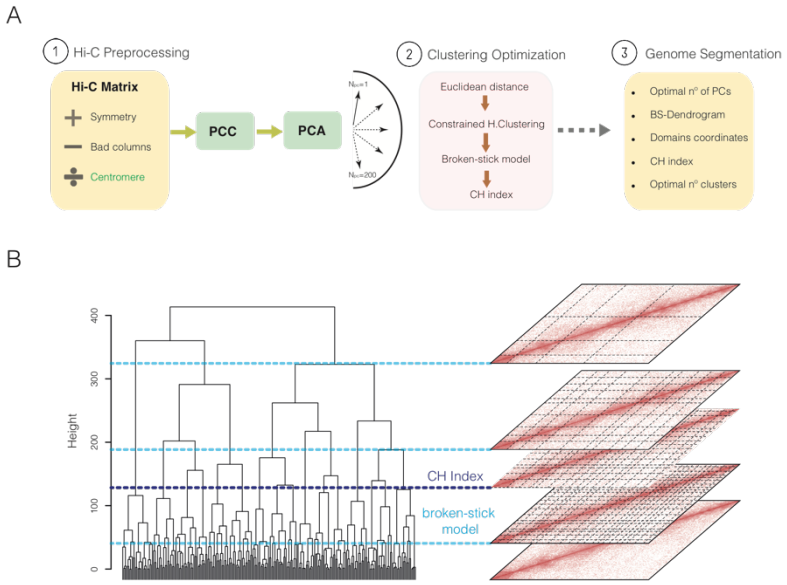
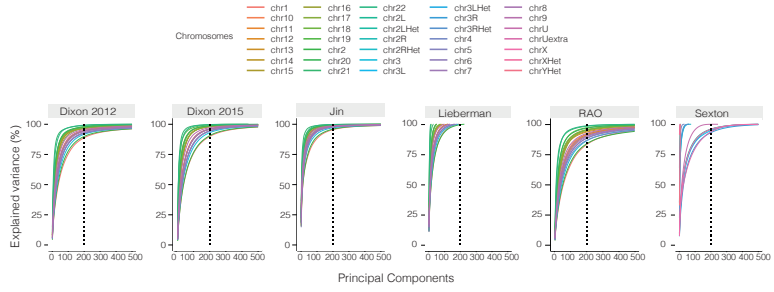


Figure 1. General overview of TADpole tool. (A) ***Schematic view of TADpole algorithm. (1) TADpole input is an all-vs-all interaction matrix. The matrix is checked for symmetry, and low-quality columns (bad columns) are removed. Large matrices of entire chromosomes are optionally split at the centromere to create two smaller sub-matrices corresponding to the chromosomal arms. Next, denoising and dimensionality reduction steps take place by computing the corresponding Pearson's correlation coefficient (PCC) matrix, and its principal component analysis (PCA). (2) Per each number of first PCs retained (from 2 to 200), the PC matrix is transformed into the corresponding Euclidean distance, that serves as input to perform the constrained hierarchical clustering. The range of possible clustering levels in the hierarchy is given by an upper bound according to a broken-stick model, and then the Calinski-Harabasz (CH) index is used to select the optimal partition. (3) As output, TADpole returns the number of first PCs retained to obtain the optimal set of TADs, the dendrogram with the significant hierarchical levels, the coordinates of the chromatin domains for each partition with the associated CH index, and the optimal number of clusters.*** **(B)** ***Example of TADpole tool applied to a 10Mb-region from a Hi-C matrix of human chr18 at 40kb resolution. The complete dendrogram (left) is cut using the broken-stick model to prune nonsignificant partitions. The first (4 clusters) and last (21 clusters) significant subdivisions are shown as light blue dashed lines. A selection of the corresponding partitions (right) is mapped on the analyses Hi-C matrix as light grey dashed lines. The optimal division in 16 clusters, identified by the highest Calinski-Harabasz (CH) index, is shown as a dark blue dashed line.***

(i) Preprocessing of the input dataset. ***TADpole is designed to process all-vs-all intrachromosomal interactions matrices representing an entire chromosome, or a continuous chromosome region. Input data are formatted as matrices***

containing the interaction values in each cell** ij**. An optional filtering step can be applied to exclude columns with a low number of interactions, which are typically local biases (42). Specifically, the rows (and columns) that contain an empty cell at the main diagonal, or those whose cumulative interactions are below the first percentile (default) are excluded from the analysis. To enhance the signal-to-noise ratio, the interaction matrix is transformed into its Pearson correlation coefficient (PCC) matrix as in (13), and a principal component analysis (PCA) is performed using the* prcomp *function from the* stats **R package (R Core Team, 2013). Only the first NPC (by default, 200) principal components are retained, which are enough to extract more than 85% of the variance in the test datasets (**Supplementary Figure 1**). To reduce memory usage and processing time, TADpole has the option to divide the interaction matrix by the centromere (considered to be the longest contiguous stretch of columns with no interactions in the Hi-C matrix) and process each chromosomal arm separately. This is particularly recommended when working with matrices of more than 15,000 bins.

A



B

Dataset	Cell type	Enzyme	ID	Filtered reads	Resolution
Lieberman-Aiden	GM06990	HindIII	SRR027956	3,653,331	1Mb
Sexton	Fly Embryo	DpnII	SRR389762, SRR389763,SRR389764,SRR389765,SRR389766,SRR389767,SRR389768	47,321,181	40kb
Dixon_2012	H1-hESC	HindIII	SRR400260,SRR400261,SRR400262,SRR400263	22,912,612	40kb
Jin	IMR90	HindIII	SRR639030,SRR639031,SRR639032,SRR639033	167,135,412	40kb
Rao	GM12879	MboI	SRR1658602	64,941,983	40kb
Dixon 2015	H1-hESC	HindIII	SRR1030718,SRR1030719,SRR1030720,SRR1030721	221,757,193	40kb

Supplementary Figure 1. Percentage of explained variance as a function of the number of retained principal components of for various datasets. (A) ***Each continuous line represents a different chromosome, and the vertical dashed lines mark the default maximum value of (200) first PCs retained in TADpole.*** **(B)** ***The Hi-C experimental datasets used characterized with five descriptors: cell type, restriction enzyme, the NCBI accession numbers, number of the valid reads retrieved***

after filtering using an in-house pipeline based on TADbit (50), and binning. The datasets from different NCBI entries are merged and the resulting matrices after filtering are binned using an equal bin-width of 40kb, but for Lieberman-Aiden dataset (13) in which bin-width is 1Mb.

(ii) Constrained hierarchical clustering optimization. ***Per each value of N_{PC}, the dimensionally-reduced matrix is transformed into a Euclidean distance matrix. This distance matrix is then partitioned into topological domains using a constrained hierarchical clustering procedure as implemented in the Constrained Incremental Sums of Squares clustering method (*coniss*) of the *rioja* R package (Juggins et al., 2017). This analysis explicitly assumes the following two priors: first, the genome is organized in a hierarchical manner, with higher-order structures containing lower-order ones, and second, every pair of contiguous genomic loci must either belong to the same self-interacting domain or to the immediately contiguous one. The constrained hierarchical clustering results in a tree-like description of the organization of the genome. Next, using the broken-stick model as implemented in the *bstick* function from the *rioja* R package (Juggins et al., 2017), the dendrogram is cut at the sensible maximum number of statistically significant clusters ($max(N_D)$). Next, the Calinski-Harabasz (CH) index is computed per each of the obtained significant partitions using the *calinhara* function from the *fpc* R package (Henning C, 2018). The maximum CH is associated to the optimal chromatin subdivision, while all the other significant hierarchical levels correspond the ones with the optimal number of first N_{PC}*** **(Figure 1B).**

(iii) Genome Segmentation. ***TADpole generates four main descriptors that recapitulate the entire sets of results, namely: *(i)* the optimal number of principal components; *(ii)* the dendrogram cut at the maximum significant level; *(iii)* the start and end coordinates of the domains and the CH index per each significant level; *(iv)* and the optimal number of domains. All the TADpole output is organized in a comprehensive R object.***

TADpole benchmark analysis

Benchmark Hi-C dataset and scripts. ***A pre-existing benchmark dataset, that comprises Hi-C interaction matrices of the entire chromosome 6 in the human cell line GM12787, was used for the analysis (29). A total of 24 different conditions were tested:*** *(i)* ***twelve matrices given by the combination of four different resolutions (10kb, 50kb, 100kb and 250kb) and three normalization strategies (raw, Iterative Correction and Eigenvector decomposition (ICE) (14) and parametric model of Local Genomic Feature (LGF) (43), and*** *(ii)* ***twelve matrices obtained by down-sampling the ICE interaction matrix at 50kb resolution*** **(Figure 2A).** ***The scripts for benchmarking were downloaded and used as released in the repository https://github.com/CSOgroup/TAD-benchmarking-scripts (29)*** **(Data availability).** ***The processed Hi-C dataset was shared by Zufferey and colleagues, this eliminated from the analysis possible biases associated with the use of different pipelines for Hi-C interaction data reconstruction (28). To compare on equal footing with the other 22 TAD callers analyzed using the same benchmark, only levels of division that comprise at least 10 chromatin domains were taken into consideration for the analysis. Within these levels, the optimal partition was identified using the CH index as described before.***

The technical benchmark. ***TADpole optimal topological partitions were compared over different resolutions, normalization strategies, and sequencing depths as previously described (29). To compare the conservation of the TAD borders between two partitions, two different metrics were applied:***

1. ***The overlap score (29) was used to compare partitions across different resolutions. This is the percentage of overlapping borders, with one bin of tolerance. The statistical significance of each overlap***

score was estimated by drawing 10,000 random partitions at the finer resolution (preserving the number of optimal clusters of the real case) and computing their overlap with the subdivision at the coarser resolution. The p-value of the real-case overlap was computed as the fraction of randomized partitions with larger overlap.

2. ***The Measure of Concordance (MoC) (29), was used to compare partitions across different resolutions and normalization strategies. MoC ranges from 1 for a perfect match to 0 for poorly scoring comparisons.***

The biological benchmark. ***To test the biological relevance of the TADs identified by TADpole, the enrichment of main architectural proteins determined at TAD borders (CTCF, SMC3, and RAD21) and within TADs (H3K27me3, and H3K36me3) were studied. CTCF and cohesin sub-units, such as SMC3, and RAD21 have been shown, in fact, to be enriched at TAD borders (15,44) while H3K27me3 and H3K36me3 marks have been related to acting as a differentiator of TADs because topological domains are enriched in either one or the other, but not both (13,15,20,21). The ChIP-seq profiles were downloaded from ENCODE (45) (https://www.encodeproject.org/) (*Supplementary Table 1*). For each protein, a consensus profile was determined as the intersection of the peaks identified in each experiment using the* *multiIntersectBed* ***function from the BEDTools suite (Quinlan and Hall, 2010). Similar to Zufferey et al. (29), the fold change enrichments of CTCF, RAD21 and SMC3 at TAD borders, and the H3k27me3/H3k36me3 log10-ratio for a given partition were computed.***

Chromatin Marks	Encode ID
CTCF	ENCSR000DRZ, ENCSR000DKV, ENCSR000DZN, ENCSR000AKB
SMC3	ENCSR000DZP
RAD21	ENCSR000BMY, ENCSR000EAC
H3K36me3	ENCSR000DRW
H3K27me3	ENCSR000DRX

Supplementary Table 1. ***Encode IDs of the Chip-seq experiments used in the biological benchmark analysis.***

Difference score between topological partitions (DiffT)

The TADpole tool was next applied to two Capture Hi-C (cHi-C) datasets in embryonic day E11.5 mouse limb buds (41). Specifically, two homozygous strains were considered comprising the wild type (WT) and the so-called inversion1 (Inv1). The cHi-C interaction maps were downloaded from GEO (46) at the GSM3261968 and GSM3261969 entries for WT and Inv1, respectively. The region chr1:73.92-75.86 Mb was extracted and used for further analysis.

To compare the WT and Inv1 partitions identified by TADpole at a fixed level of the hierarchy, we defined a difference topology score (DiffT). Specifically, the partitioned matrices were transformed into binary forms W for WT, and analogously V for Inv1, in which each entry w_{ij} (v_{ij}) is equal to 1 if the bins i and j are in the same TAD and 0 otherwise. Then, DiffT is computed as the normalized (from 0 to 1) difference between the binarized matrices as a function of the bin index l as:

$$S(l) = \frac{\sum_{i=1}^{l}\sum_{j=1}^{N}|w_{ij} - v_{ij}|}{\sum_{i=1}^{N}\sum_{j=1}^{N}|w_{ij} - v_{ij}|}$$

where N is the size of the matrix.

To test whether the identified partition in Inv1 is different from WT, at each level of the chromatin hierarchy, a statistical analysis was introduced. This analysis assesses the significance of DiffT at each bin of the matrix. A total of 10,000 random partitions of the region were simulated excluding the bad columns of the Inv1 matrix. The DiffT score was computed between simulated and WT partitions ($DiffT_{simulated\text{-}wt}$). At each bin, the fraction of $DiffT_{simulated\text{-}wt}$ lower or equal to the $DiffT_{Inv1\text{-}wt}$ score estimates the p-value. A p-value < 0.05 means that a significant difference is located from the bin

under consideration onwards. Hence, the bin with the minimum p-value marks the starting point of the genomic region where the most significant fraction of the DiffT score is located.

Results

TADpole benchmark analysis

To quantitatively compare TADpole with other 22 TAD callers, we applied the multiple conditions test proposed in (29) on the same reference benchmark dataset **(Figure 2A and Material and Methods).**

Technical benchmarking. ***We assessed various technical aspects of TADpole as well as the robustness of TADpole identified domains with respect to different resolutions, normalization strategies, and sequencing depths of the input matrix*** **(Figure 2A).** ***Firstly, we examined the number and the size (in kilobases and in bins) of the optimal domain partition of the ICE normalized maps at different resolutions*** **(Figure 2B).** ***We found that, as the resolution of the Hi-C interaction map decreased, both the numbers of TADs and the mean TAD size in bins decreased with a 4-fold reduction. TADpole followed a similar grow tendency (positive when the TADs are measure in kilobases and negative with the TADs are measured in bins) as the majority of the other TAD callers independently on the normalized strategy applied*** **(Supplementary Table 2).** ***We also inspected if TADpole identified robust boundaries that were conserved at different resolutions. To measure this conservation, we tested if a border detected in the ICE normalized Hi-C matrices at a certain resolution was conserved in the resolution immediately finer*** **(Figure 2C).** ***At the coarser resolutions, that is 250kb vs. 100kb, we found a high agreement (67%), that decreased only slightly to (59%) at intermediate ones (100*** *vs.* ***50kb). Interestingly, we found that even at the finer resolutions (50kb*** *vs.* ***10kb), where the 48% of the borders were***

conserved, this analysis was consistent with a statistically significant overlap (p-value<0.05).

Resolution	Raw N° TADs	Raw Size (kb)	Raw Size(bins)	ICE N° TADs	ICE Size (kb)	ICE Size (bins)	LGF N° TADs	LGF Size (kb)	LGF Size (bins)
250kb	118	1425.85	5.7	116	1465.52	5.86	120	1402.08	5.61
100kb	193	868.91	8.68	193	879.79	8.79	193	884.46	88.45
50kb	217	772.35	15.45	205	814.39	16.29	208	805.77	16.12
10kb	535	313.01	31.3	494	338.7	33.87	510	328.29	32.83

Supplementary Table 2. ***The total number of TADs and the corresponding average size detected in raw and normalized (by ICE and LGF) Hi-C matrices across different resolutions.***

Next, we used the Measure of Concordance (MoC) **(Material and Methods)** ***to estimate if the number and the position of the borders of chromatin domains identified by TADpole were affected by the matrix resolutions and by different normalization strategies. Interestingly, we found that the MoC over different matrix resolutions had values in the [0.45:0.82] range with an average MoC of 0.63, and ranked first when compared with the other 22 TAD callers previously benchmarked (29). TADpole was also robust over different normalization strategies with an average MoC of 0.74, ranking 9th over the 22 TAD callers. Comparing the average of resolutions vs. normalizations MoC values of TADpole with the rest of TADcallers*** **(Figure 2D),** ***we found that TADpole appeared in the top-right corner of the plot demonstrating its overall high robustness and confidence to identify optimal chromatin domains independently of the resolution or the normalization of the input Hi-C matrix. We also tested the TADpole propensity to identify consistent optimal chromatin domains independently of the sequencing depth*** **(Figure 2E).** ***We compared the partitions obtained by doing 12 different sub-sampling of the ICE-normalized interaction matrix at 50kb with the full interaction matrix using the MoC. We found that TADpole partitions were clearly robust to down-sampling with a MoC score of 0.79 with just 0.1% of the total data. This feature classified TADpole as the top TAD caller with respect to the other 22 tools.***

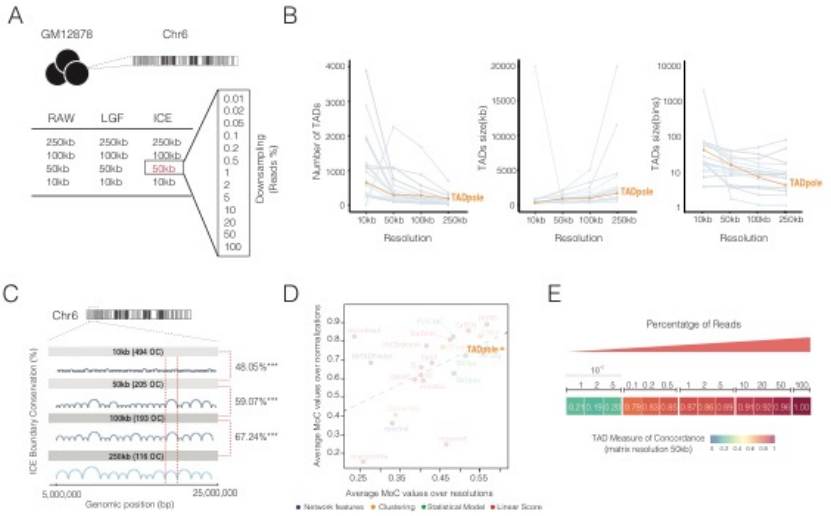


Figure 2. Technical benchmark of TADpole tool. (A) ***Descriptions of the dataset used for the TADpole benchmark analysis of Zufferey*** *et al.* ***(29). The dataset includes the Hi-C interactions matrices for chromosome 6 in GM12878 cells organized in 24 different forms*** **(Material and Methods). (B-E)** ***Technical benchmark of TADpole optimal partition and comparison with 22 other TAD callers considered in (29).*** **(B)** ***Analysis of the number of TADs at different matrix binning and the TAD size in terms of kb and number of bins.*** **(C)** ***Fraction of conserved TADs boundaries at different resolutions in the entire chromosome 6. The scheme illustrates the analysis focusing on the region from 5 to 25Mb.*** **(D)** ***Average Measure of Concordance (MoC) values across normalizations vs the average MoC value across resolutions. The color scheme reflects the specific approach used in each TAD caller.*** **(E)** ***The MoC values over different down-sampling levels of the ICE-normalized interaction matrix at 50kb. Panel B and D have been adapted from Figure 2C of Zufferey*** *et al.* ***(29) to include TADpole in the comparison of TAD callers.***

Biological benchmarking. ***With the lack of a gold standard to define TADs in Hi-C interaction maps (28,29), we investigated the biological relevance of the domains identified by TADpole in terms of their association with biological features that have been shown to have an important role in the formation and maintenance of TADs. We found that the intensity profiles of the CTCF, RAD21, and SMC3 signals were peaked at TADpole chromatin borders*** **(Figure 3A).** ***To compare these results with the set of other TAD callers, we computed the fold change enrichments at the peak with respect to the flanking regions*** **(Figure 3B).** ***TADpole resulted in a fold change enrichment***

around 1 for each of the three architectural proteins (1.18 in CTCF, 1.06 in RAD21 and 0.97 in SMC3, respectively), that was consistent with a significantly high peak at the border compared with the background (p-value<10^{-5}). In this analysis, TADpole ranked as the 6th TAD caller. Additionally, more than 40% of the tagged boundaries are enriched in one or more of these three architectural proteins being CTCF (42%) and SMC3 (42%) the most abundant ones. Considering the enrichment at TADs boundaries, TADpole ranked 3rd within the set of 22 TAD examined callers **(Figure 3C).** *To further study the enrichment of these biological features at domain borders, we performed an analysis of the fold change of CTCF, RAD21 and SMC3 in each of the chromosome arms. In all the identified levels, there was a positive fold-change with certain variability with the level of nested data, being CTCF in all the cases, the most enriched architectural protein in the border regions* **(Figure 3D).**

TADs are usually expected to be transcriptionally either active or inactive (15) with the TAD body enriched in active or inactive histone mark. To assess if the interior of TADpole detected partitions was indeed enriched in either active or inactive chromatin, we considered the signals of two marks H3K36me3 for transcriptional activity and H3K27me3 for repression, and measured the fraction of TADs where the Log10 of their ratio(H3K27me3/H3K36me3) was significantly high or low (FDR<0.1). Notably, we found that the majority (57%) of the TADpole identified TADs have a defined active or inactive state, locating TADpole within the top four TAD callers based on this criterion **(Figure 3E).**

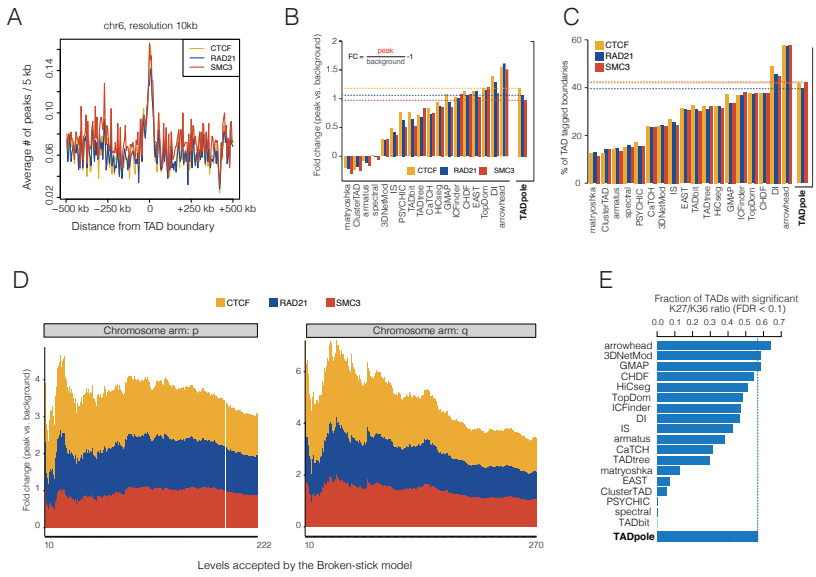


Figure 3. Biological benchmark of TADpole tool. ***TADpole borders for the optimal level of the hierarchical partition and chromatin-organizer proteins.*** **(A)** ***The mean ChIP-seq signal profiles (5-kb intervals) in 1Mb region around the TADpole domains borders are shown for CTCF, RAD21 and SMC3.*** **(B)** ***The fold-change enrichment of CTCF, RAD21 and SMC3 at domain borders vs. background and*** **(C)** ***the percentage identified TADs boundaries tagged with CTCF, RAD21 and SMC3 are shown as bar plots.*** **(D)** ***Cumulative fold-change of the enrichment in CTCF, RAD21 and SMC3 at domain borders vs. background for all the significant levels (minimum 10 partitions as in (29)) retrieved by the broken-stick model in the two chromosomes arms (p, q) of chromosome 6.*** **(E)** ***The fractions of TADs with significant log10 ratio between H3K36me3 and H3K27me3*** **(Material and Methods)** ***in TADpole and the 22 TAD callers are represented as bar plots. Panels B, C and E have been adapted from Figure 5D, 5E, 5I of Ref. (29) reporting on the results of the other TAD callers.***

Applications to capture Hi-C datasets

To show the effectiveness of TADpole, we applied our caller to cHi-C experiment of embryonic day E11.5 mouse cells (41). Kraft et al. investigated the pathogenic consequences of balanced chromosomal rearrangements in embryonic mouse limb buds, focusing on a 1.9Mb region in chr1 (chr1:73.92-75.86 Mb) where a cluster of genomic regulators of *Epha4 locus* ***is located. The authors generated, together with the wild-type (WT), a total of 4 mutant strains, each inducing different inversions. We compared the WT strain with***

the sole inversion producing a homozygous strain (here called Inv1), that is located between the telomeric site of *Epha4* ***enhancer cluster and the promoters of*** *Resp18* ***(breakpoint at chr1:75,275,966-75,898,706*** **Figure 4A*****). Analysis of the entire TADpole dendrogram revealed the existence of 19 and 17 significant partition levels in WT and Inv1, respectively. The optimal ones were 11 for WT and 2 for Inv1. At a visual inspection, the maps in*** **Figure 4A** ***show that the difference between the chromatin partitions increases with the partition level, and accumulates in the region of the inversion.***

To statistically quantify and localize the significant topological differences between the WT and Inv1 matrices, we computed their DiffT score profiles (**Material and Methods** ***and*** **Figure 4B** ***for the partition in 9 domains), at each level of the hierarchical partition. We found that the DiffT profiles sharply increased close to the point of the inversion*** **(Figure 4C).** ***Based on the p-value profiles*** **(Figure 4D),** ***we identified two regions where the minimum pvalues (one per partition level) accumulated. Notably, 70% of minimum p-values were located within a region, spanning 50kb, from the point where inversion was induced, suggesting that the significant topological changes between WT and Inv1 accumulated in the inverted region.***

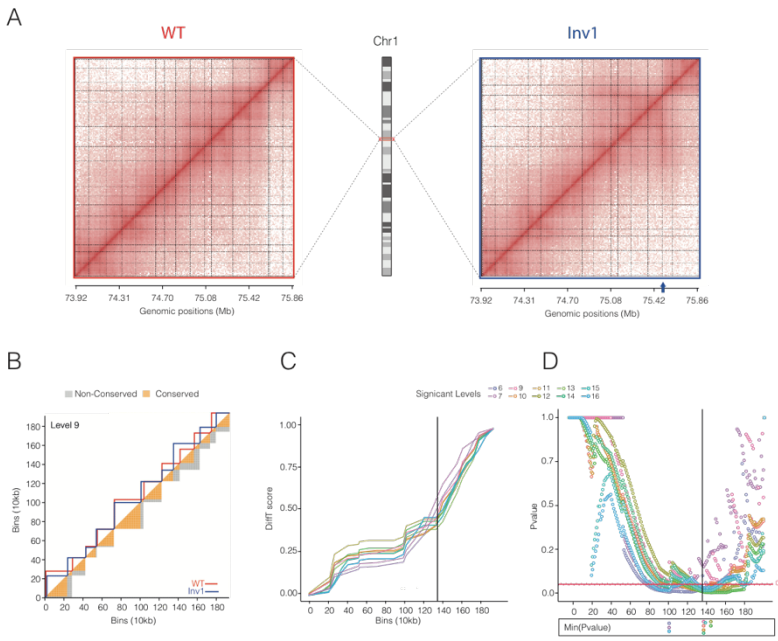


Figure 4. Characterization of topological difference in capture Hi-C datasets. **(A)** ***Capture Hi-C maps of chr1:73.92-75.86Mb region in wild-type (WT) and inversion 1 (Inv1) strains (41). In both matrices, TADpole significant partitions are shown as light gray dashed lines. The blue arrow indicates the centromeric breakpoint located at the promoter of Resp18 gene in Inv1.*** **(B)** ***DiffT score scheme for level 9. The upper triangle of the matrix shows the TADs borders identified by TADpole in WT and Inv1 matrices as red and blue continuous lines, respectively. The lower triangle of the matrix shows the conserved and non-conserved areas of the TADs in orange and gray, respectively. In the panels C and D, the Inv1 breakpoint is highlighted with a solid black line and only the significant levels (with at least a p-value < 0.05) are shown.*** **(C)** ***DiffT score profiles as a function of the matrix bins*** **(Material and Methods). (D)** ***P-value profiles per bin for automated detection of significant differences. In the lower panel, the bins associated with minimum p-values per level are marked with empty dots.***

Discussion and conclusion

In this work we introduced TADpole, a tool to identify hierarchical topological domains from all-vs-all intra-chromosomal interaction matrices. In line with previously introduced concepts such as metaTADs (26) and sub-TADs (27), we propose that there is not a single meaningful subdivision of chromatin domains, but rather a set of hierarchical levels associated with

different genomic features. Notably, TADpole characterizes the entire hierarchy of TADs while assessing the significance at various levels of partition, paving the way for deepening our understanding of the nested topology and its biological role. Indeed, the principles behind this nested structure are not yet fully understood. Different levels of the hierarchy can be involved in the dynamical modulations of TADs upon perturbation, responding to or causing changes in gene regulation (3,22,47). Alternatively, these nested organization can rise as effect of the variability in the topological conformation observed in individual cells (48). However, we cannot exclude the possibility that this nested structure exists in single cells as multi-site interactions conformation acting together to establish robust gene regulation networks. All these evidences highlight the importance to have a tool like TADpole, that systematically characterizes the entire TAD architecture from all-vs-all intrachromosomal interaction matrices.

*Here, we compared TADpole's performance with a set of other 22 TAD callers following the benchmark analysis performed by Zufferey et al. (29). TADpole identifies a number of TADs over different resolutions that is in agreement with other TAD callers (***Figure 2B***). The identified domains have an average size of 855kb, in agreement with the reported average TADs size in mammalian cells (~900-1000kb) (15). TADpole shows one of the largest consistencies over different normalization strategies (including also non-normalized data), resolutions and sequencing depths (***Figure 2D** *and* **Supplementary Table 2***). These observations make TADpole potentially suitable for analyzing sparse datasets. TADpole has been shown to be technically robust. Indeed, the optimal TADpole chromatin partition borders present a high enrichment of architectural proteins such as CTCF, SMC3, and RAD21, and this enrichment is maintained, for certain partitions, over all the significant hierarchical levels (***Figure 3D***). The identified TADs are enriched in either active (H3K36me3) or inactive (H3K27me3) marks*

*(*Figure 3E*), suggesting a strong consistency between the structural definition of TADs and their biological characterization. A handful of algorithms for the detection of chromatin domains as nested TADs have been implemented (CaTCH (36), GMAP (37), matryoshka (38), and PSYCHIC (39)). However, the uniqueness of TADpole is its ability to provide multiple significant partitions and define the optimal one in an unsupervised manner by using the Broken-Stick model and the Calinski-Harabasz index criteria. Notably, in the benchmark analysis presented here (Figures 2 and 3), TADpole performs generally better than all the other nested TAD callers when considering the technical and biological benchmarking performed here.*

A possible advantage of TADpole over existing TAD callers is the pre-processing data step. Indeed, the PCC transformation and the PCA application regularize the input matrix so that the specific normalization applied on the input and the sparsity of the data have little effect on identifying TADs. Previously, other architectural features of the chromatin have been already studied using PCA. The first principal component is widely used to identify the chromatin segregation into compartments (13). The second and the third PCs have been associated instead to intra-arm features mainly centromere-centromere and telomere-telomere interactions enrichment (14). Moreover, the first PCs have been used to assess the similarity between two interaction maps (14) as well as to quantify their reproducibility (49). Here we have demonstrated that there exists an optimal set of PCs capable of identifying the hierarchical structure of chromatin, extending the current application of PCA to characterize genome topology.

*We provide a proof of TADpole's usability on a topologically complex region analyzing cHi-C data in both a wild-type strain and a mutant one carrying a genomic inversion (41) (*Figure 4B*). The use of TADpole in combination with the DiffT score is able to identify the inverted region as the one with the*

highest difference in topological partitions, proving that this strategy can isolate bin-dependent and statistically significant topological dissimilarities. Overall, we prove that the DiffT score allows to evaluate a priori the location where the most significant topological differences between two hierarchical subdivisions are accumulated.

In summary, TADpole combines straightforward bioinformatic analyses such as PCA and hierarchical clustering to study continuous nested hierarchical segmentation of an all-vs-all intra-chromosomal interactions matrix. Additionally, we demonstrated the technical and biological robustness of TADpole, and its usability in identifying topological difference in high-resolution capture Hi-C experiments. TADpole is released as a publicly-available, open-source and numerically-efficient R tool. As such, TADpole represents a comprehensive tool that fulfils the needs of the scientific community for an accurate TAD caller able to comprehensively study the interplay between the hierarchical chromatin topology and genomic function.

Data availability

The TADpole is freely available for download as an R package at https://github.com/3DGenomes/TADpole. The scripts for the technical and biological benchmarks were obtained from the repository https://github.com/CSOgroup/TADbenchmarking-scripts (28). Specifically, the script *fig2_fig3_fig4_fig5_moc_calc.*R ***was used for panels*** **Figure 2B** ***to E, the script*** *StructProt_EnrichBoundaries_script.*R ***for panels*** **Figure 3A** ***to*** **D**, ***and the script*** *HistMod_script.sh* ***for panel in*** **Figure 3E**. ***Default parameters were applied.***

Acknowledgement

We thank Dr. M. Zufferey and Dr. G. Ciriello for providing us with the dataset used in Ref.(29), that make possible an easy and quick comparison of TADpole with 22 other TAD callers. We thank also Dr. K. Kraft and Dr. S. Mundlos for helping in the interpretation of the Capture Hi-C datasets in Ref. (41). We acknowledge the ENCODE consortium and the ENCODE production laboratories that generated the datasets used in the manuscript.

Funding

This work was partially supported by the European Research Council under the 7th Framework Program FP7/2007-2013 (ERC grant agreement 609989), the European Union's Horizon 2020 research and innovation programme (grant agreement 676556) and the Spanish Ministerio de Ciencia, Innovación y Universidades (BFU2013-47736-P and BFU2017-85926-P to M.A.M-R. and BES-2014-070327 to P.S-V.). We also knowledge support from 'Centro de Excelencia Severo Ochoa 2013-2017', SEV-2012-0208 and the CERCA Programme/Generalitat de Catalunya to the CRG.

Conflict of interest

No conflict of interests declared.

REFERENCES

1. *Sexton, T. and Cavalli, G. (2015) The role of chromosome domains in shaping the functional genome. Cell, 160, 1049-1059.*
2. *Dekker, J. and Mirny, L. (2016) The 3D Genome as Moderator of Chromosomal Communication. Cell, 164, 1110-1121.*
3. *Stadhouders, R., Vidal, E., Serra, F., Di Stefano, B., Le Dily, F., Quilez, J., Gomez, A., Collombet, S., Berenguer, C., Cuartero, Y. et al. (2018) Transcription factors orchestrate dynamic interplay between genome topology and gene regulation during cell reprogramming. Nat Genet, 50, 238-249.*
4. *Paulsen, J., Liyakat Ali, T.M., Nekrasov, M., Delbarre, E., Baudement, M.O., Kurscheid, S., Tremethick, D. and Collas, P. (2019) Long-range interactions between topologically associating domains shape the four-dimensional genome during differentiation. Nat Genet, 51, 835-843.*
5. *Bonev, B., Mendelson Cohen, N., Szabo, Q., Fritsch, L., Papadopoulos, G.L., Lubling, Y., Xu, X., Lv, X., Hugnot, J.P., Tanay, A. et al. (2017) Multiscale 3D Genome Rewiring during Mouse Neural Development. Cell, 171, 557-572 e524.*
6. *Zheng, H. and Xie, W. (2019) The role of 3D genome organization in development and cell differentiation. Nat Rev Mol Cell Biol.*
7. *Lupianez, D.G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., Horn, D., Kayserili, H., Opitz, J.M., Laxova, R. et al. (2015) Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. Cell, 161, 1012-1025.*
8. *Franke, M., Ibrahim, D.M., Andrey, G., Schwarzer, W., Heinrich, V., Schopflin, R., Kraft, K., Kempfer, R., Jerkovic, I., Chan, W.L. et al. (2016) Formation of new chromatin domains determines pathogenicity of genomic duplications. Nature, 538, 265-269.*

9. *Groschel, S., Sanders, M.A., Hoogenboezem, R., de Wit, E., Bouwman, B.A.M., Erpelinck, C., van der Velden, V.H.J., Havermans, M., Avellino, R., van Lom, K. et al. (2014) A single oncogenic enhancer rearrangement causes concomitant EVI1 and GATA2 deregulation in leukemia. Cell, 157, 369-381.*
10. *Flavahan, W.A., Drier, Y., Liau, B.B., Gillespie, S.M., Venteicher, A.S., Stemmer- Rachamimov, A.O., Suva, M.L. and Bernstein, B.E. (2016) Insulator dysfunction and oncogene activation in IDH mutant gliomas. Nature, 529, 110-114.*
11. *Krijger, P.H. and de Laat, W. (2016) Regulation of disease-associated gene expression in the 3D genome. Nat Rev Mol Cell Biol, 17, 771-782.*
12. *Cremer, T. and Cremer, C. (2001) Chromosome territories, nuclear architecture and gene regulation in mammalian cells. Nat Rev Genet, 2, 292-301.*
13. *Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O. et al. (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science, 326, 289-293.*
14. *Imakaev, M., Fudenberg, G., McCord, R.P., Naumova, N., Goloborodko, A., Lajoie, B.R., Dekker, J. and Mirny, L.A. (2012) Iterative correction of Hi-C data reveals hallmarks of chromosome organization. Nat Methods, 9, 999-1003.*
15. *Rao, S.S., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S. et al. (2014) A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping.Cell, 159, 1665-1680.*
16. *Nir, G., Farabella, I., Perez Estrada, C., Ebeling, C.G., Beliveau, B.J., Sasaki, H.M.,Lee, S.D., Nguyen, S.C., McCole, R.B., Chattoraj, S. et*

al. (2018) Walking along chromosomes with super-resolution imaging, contact maps, and integrative modeling. PLoS Genet, 14, e1007872.

17. *Boettiger, A.N., Bintu, B., Moffitt, J.R., Wang, S., Beliveau, B.J., Fudenberg, G., Imakaev, M., Mirny, L.A., Wu, C.T. and Zhuang, X. (2016) Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. Nature, 529, 418-422.*
18. *Bintu, B., Mateo, L.J., Su, J.H., Sinnott-Armstrong, N.A., Parker, M., Kinrot, S.,Yamaya, K., Boettiger, A.N. and Zhuang, X. (2018) Super-resolution chromatintracing reveals domains and cooperative interactions in single cells. Science, 362.*
19. *Szabo, Q., Jost, D., Chang, J.M., Cattoni, D.I., Papadopoulos, G.L., Bonev, B.,Sexton, T., Gurgo, J., Jacquier, C., Nollmann, M. et al. (2018) TADs are 3D structural units of higher-order chromosome organization in Drosophila. Sci Adv, 4, eaar8082.*
20. *Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S. and Ren, B. (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature, 485, 376-380.*
21. *Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N.L., Meisig, J., Sedat, J. et al. (2012) Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature, 485, 381-385.*
22. *Le Dily, F., Baù, D., Pohl, A., Vicent, G.P., Serra, F., Soronellas, D., Castellano, G., Wright, R.H., Ballare, C., Filion, G. et al. (2014) Distinct structural transitions of chromatin topological domains correlate with coordinated hormone-induced gene regulation. Genes Dev, 28, 2151-2162.*
23. *Le Dily, F., Vidal, E., Cuartero, Y., Quilez, J., Nacht, A.S., Vicent, G.P., Carbonell-Caballero, J., Sharma, P., Villanueva-Canas, J.L.,*

Ferrari, R. et al. (2019) Hormonecontrol regions mediate steroid receptor-dependent genome organization. Genome Res, 29, 29-39.

24. *Dixon, J.R., Jung, I., Selvaraj, S., Shen, Y., Antosiewicz-Bourget, J.E., Lee, A.Y., Ye, Z., Kim, A., Rajagopal, N., Xie, W. et al. (2015) Chromatin architecture reorganization during stem cell differentiation. Nature, 518, 331-336.*
25. *Bonev, B. and Cavalli, G. (2016) Organization and function of the 3D genome. Nat Rev Genet, 17, 661-678.*
26. *Fraser, J., Ferrai, C., Chiariello, A.M., Schueler, M., Rito, T., Laudanno, G., Barbieri, M., Moore, B.L., Kraemer, D.C., Aitken, S. et al. (2015) Hierarchical folding and reorganization of chromosomes are linked to transcriptional changes in celular differentiation. Mol Syst Biol, 11, 852.*
27. *Berlivet, S., Paquette, D., Dumouchel, A., Langlais, D., Dostie, J. and Kmita, M. (2013) Clustering of tissue-specific sub-TADs accompanies the regulation of HoxA genes in developing limbs. PLoS Genet, 9, e1004018.*
28. *Forcato, M., Nicoletti, C., Pal, K., Livi, C.M., Ferrari, F. and Bicciato, S. (2017)Comparison of computational methods for Hi-C data analysis. Nat Methods, 14, 679-685.*
29. *Zufferey, M., Tavernari, D., Oricchio, E. and Ciriello, G. (2018) Comparison of computational methods for the identification of topologically associating domains. Genome Biol, 19, 217.*
30. *Crane, E., Bian, Q., McCord, R.P., Lajoie, B.R., Wheeler, B.S., Ralston, E.J., Uzawa, S., Dekker, J. and Meyer, B.J. (2015) Condensin-driven remodelling of X chromosome topology during dosage compensation. Nature, 523, 240-244.*
31. *Oluwadare, O. and Cheng, J. (2017) ClusterTAD: an unsupervised machine learning approach to detecting topologically associated domains of chromosomes from Hi-C data. BMC Bioinformatics, 18,*

480.

32. *Haddad, N., Vaillant, C. and Jost, D. (2017) IC-Finder: inferring robustly thehierarchical organization of chromatin folding. Nucleic Acids Res, 45, e81.*
33. *Filippova, D., Patro, R., Duggal, G. and Kingsford, C. (2014) Identification of alternative topological domains in chromatin. Algorithms Mol Biol, 9, 14.*
34. *Weinreb, C. and Raphael, B.J. (2016) Identification of hierarchical chromatin domains. Bioinformatics, 32, 1601-1609.*
35. *Norton, H.K., Emerson, D.J., Huang, H., Kim, J., Titus, K.R., Gu, S., Bassett, D.S. and Phillips-Cremins, J.E. (2018) Detecting hierarchical genome folding with network modularity. Nat Methods, 15, 119-122.*
36. *Zhan, Y., Mariani, L., Barozzi, I., Schulz, E.G., Bluthgen, N., Stadler, M., Tiana, G.and Giorgetti, L. (2017) Reciprocal insulation analysis of Hi-C data shows that TADs represent a functionally but not structurally privileged scale in the hierarchical folding of chromosomes. Genome Res, 27, 479-490.*
37. *Yu, W., He, B. and Tan, K. (2017) Identifying topologically associating domains and subdomains by Gaussian Mixture model And Proportion test. Nat Commun, 8, 535.*
38. *Malik, L. and Patro, R. (2018) Rich Chromatin Structure Prediction from Hi-C Data. IEEE/ACM Trans Comput Biol Bioinform.*
39. *Ron, G., Globerson, Y., Moran, D. and Kaplan, T. (2017) Promoter-enhancer interactions identified from Hi-C data using probabilistic models and hierarchical topological domains. Nat Commun, 8, 2237.*
40. *Dryden, N.H., Broome, L.R., Dudbridge, F., Johnson, N., Orr, N., Schoenfelder, S., Nagano, T., Andrews, S., Wingett, S., Kozarewa, I. et al. (2014) Unbiased analysis of potential targets of breast cancer susceptibility loci by Capture Hi-C. Genome Res, 24, 1854-1868.*

41. *Kraft, K., Magg, A., Heinrich, V., Riemenschneider, C., Schopflin, R., Markowski, J., Ibrahim, D.M., Acuna-Hidalgo, R., Despang, A., Andrey, G. et al. (2019) Serial genomic inversions induce tissue-specific architectural stripes, gene misexpression and congenital malformations. Nat Cell Biol, 21, 305-310.*
42. *Vidal, E., le Dily, F., Quilez, J., Stadhouders, R., Cuartero, Y., Graf, T., Marti-Renom, M.A., Beato, M. and Filion, G.J. (2018) OneD: increasing reproducibility of Hi-C samples with abnormal karyotypes. Nucleic Acids Res, 46, e49.*
43. *Hu, M., Deng, K., Selvaraj, S., Qin, Z., Ren, B. and Liu, J.S. (2012) HiCNorm: removing biases in Hi-C data via Poisson regression. Bioinformatics, 28, 3131-3133.*
44. *Kojic, A., Cuadrado, A., De Koninck, M., Gimenez-Llorente, D., Rodriguez-Corsino, M., Gomez-Lopez, G., Le Dily, F., Marti-Renom, M.A. and Losada, A. (2018) Distinct roles of cohesin-SA1 and cohesin-SA2 in 3D chromosome organization. Nat Struct Mol Biol, 25, 496-504.*
45. *Davis, C.A., Hitz, B.C., Sloan, C.A., Chan, E.T., Davidson, J.M., Gabdank, I., Hilton, J.A., Jain, K., Baymuradov, U.K., Narayanan, A.K. et al. (2018) The Encyclopedia of DNA elements (ENCODE): data portal update. Nucleic Acids Res, 46, D794-D801.*
46. *Barrett, T., Wilhite, S.E., Ledoux, P., Evangelista, C., Kim, I.F., Tomashevsky, M., Marshall, K.A., Phillippy, K.H., Sherman, P.M., Holko, M. et al. (2013) NCBI GEO: archive for functional genomics data sets--update. Nucleic Acids Res, 41, D991-995.*
47. *Narendra, V., Bulajic, M., Dekker, J., Mazzoni, E.O. and Reinberg, D. (2016) CTCF mediated topological boundaries during development foster appropriate gene regulation. Genes Dev, 30, 2657-2662.*
48. *Nagano, T., Lubling, Y., Stevens, T.J., Schoenfelder, S., Yaffe, E.,*

Dean, W., Laue, E.D., Tanay, A. and Fraser, P. (2013) Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. Nature, 502, 59-64.

49. *Yan, K.K., Yardimci, G.G., Yan, C., Noble, W.S. and Gerstein, M. (2017) HiCspector: a matrix library for spectral and reproducibility analysis of Hi-C contact maps. Bioinformatics, 33, 2199-2201.*
50. *Serra, F., Bau, D., Goodstadt, M., Castillo, D., Filion, G.J. and Marti-Renom, M.A. (2017) Automatic analysis and 3D-modelling of Hi-C data using TADbit reveals structural features of the fly chromatin colors. PLoS Comput Biol, 13, e1005665.*

DISCUSSION

Dynamics of genome architecture and chromatin function during human B cell differentiation and neoplastic transformation

The three-dimensional eukaryotic genome is organized in a hierarchical manner which is tightly linked with the functional and regulatory DNA processes. Hi-C, and its derivative techniques have opened the door to uncover the 3D genome organization within the nucleus at multiple scales of resolution. However, the high complexity of biological systems makes it necessary to integrate multiple layers of information, coming from high-throughput technologies and imaging methods, to draw a comprehensive biological view of the interrelationship between the spatial organization of the chromatin and its role in the genomic functions.

B cells are central in the humoral immune system and abnormal gene regulation in these cells can be associated with cancer development and is accompanied by a critical chromatin remodeling. In Chapter 1 of this Thesis, we present an integrative multi-omics approach combining *in situ* ***Hi-C data with nine additional layers of information (six histone modifications, chromatin accessibility, gene expression, and DNA methylation) to explore how the 3D genome is modulated during normal B cell differentiation and upon neoplastic transformation and its link with the functional state of the cell.***

Our Hi-C study on chromatin segregation into more transcriptionally active or repressed compartments suggested a general positive correlation between changes in the type of compartmentalization with gene expression, DNA accessibility and histone modification patterns. In fact, the active histone marks H3K4me3, H3K4me1, H3K27ac and H3K36me3 had a higher Pearson correlation with the first principal component (that usually defines genome segmentation) than the repressive marks H3K9me3 and

H3K27me3. In fact, it has been already suggested to use only active histone marks to annotate compartments(93), which reveals a differential relationship between active and repressive histone marks with the genome compartmentalization. Xie and colleagues {Xie, 2017, 0RyX} even propose a different regulatory mechanism for the repressive histone marks, especially for H3K27me3.

Upon a closer examination of Hi-C data and specifically of the first eigenvector coefficients, we determined a multimodal distribution of continuous data with two extreme modes connected by a valley of intermediate values. The application of a global clustering using a Gaussian mixture model categorized the data into three compartments: A-type, B-type and a novel compartment, that we annotated as I-type, identified as an intermediate and transitional compartment with significant enrichment in two main chromatin states, poised promoter and polycomb-repressive chromatin. This new compartment definition better captures the complexity of the genome topology, yielding a better description of how the genome is modulated during cell fate decision. According to recent studies, polycomb-bound chromatin segregated into central nuclear (sub)compartments that have a different 3D folding, compared to active or constitutive inactive chromatin (98). 4C-seq experiments during mouse embryonic stem cells differentiation showed how the polycomb-bound chromatin is switched between A and B compartments accompanied by a massive loss of intra and inter-interactions between polycomb-bond loci, which suggests transitional and dynamic features of this facultative heterochromatin (133). In fact, we show that this dynamic compartment can evolve either into active or inactive compartments, with high correlation with the gain of A-related chromatin states or B-related chromatin states respectively, highlighting the association between the modulation of the 3D genome together with the epigenomic properties of the chromatin. Our data confirm that the majority of the compartment conversions during normal B cell differentiation involve

transitioning to or from I-type compartments, and that abrupt transitions between extreme compartments, such as A-to-B or B-to-A) are minimal. In fact, during T cell differentiation, permissive chromatin microenvironments (with intermediate compartment scores) were detected inside repressed B compartments decorated with active histone marks, where they can propagate into neighboring regions and turn into active compartments (143). Interestingly, 1,218 genomic regions (determined at 20kb resolution) exhibited a compartment switch. In our study, around 28% of the entire genome presented a change in the compartmentalization (determined at 100kb resolution) in at least one of the 4 analyzed B cell subpopulations. Large spatial plasticity of the genome (36% of switched compartments) was annotated by studying the human embryonic stem cell differentiation (107) (compartments determined at 40kb resolution). While another, much more extensive study (a compendium of 21 human cell and tissue types) determined that ~60% of the genome was dynamically compartmentalized between the samples studied (144) (compartments were determined at 1Mb resolution).

Indeed, there is a large variability on the percentage of compartment changes in the genome upon cell fate decision and cell types. Taking into account the biological and technical biases associated to each experiment, and the statistical analyses applied, the resolution (the bin size of the Hi-C interaction matrix) used in each study could be a parameter affecting the results of downstream analyses. The resolution of the Hi-C matrix dictates the scale of the 3D organization observable from the data. There is not a standard method to determine which is the best resolution for annotating the compartments. The annotation of this sub-compartmentalization could be very useful to provide a finer 3D genome perspective concerning its spatial position in the nucleus and its influence on gene regulation. From this point of view, we have contributed to detect genome compartment differences that take place during cell differentiation and correlate them with the functional

events that are occurring in the cell, all thanks to the detailed annotation of A-type, I-type and B-type compartment.

The main structural features that we detected during B cell differentiation were: *(i)* ***the great modulation to the active state of the NBC to the GCBC and (ii) the reversion of the compartment organization that suffers MBC to the organization profiled in a naive state, the NBC. The dramatic structural change of NBC to GCBC is consistent with the findings of another study (145) where the authors suggested a global decompaction of the germinal center related to a significant loss of inter-arm chromosomal interactions and*** *de novo* ***coordination of specific transcription factories (145). Analogously, we observed a global activation of the germinal center together with an increase of chromatin accessibility, all decorated in a I-type compartment environment. A transcription factor motif analysis in germinal-specific active regions showed an enrichment of two main transcription factor families, MEF2 and POU, that are functionally connected with developmental processes and linked with the formation to the germinal center (146). The germinal center response gives rise to two different cell types, MBC and PC, with divergent immune functions. Surprisingly, about 75% of the total detected compartment changes in MBC reverts into a naive B cells state. This change, or this structural memory, is also reflected in the content of the other analyzed multi-omics layers, where histone modifications, DNA accessibility, and gene expression follow the same general pattern either in NBC and in MBC.***

Extensive disorganization of the genome has been described during neoplastic transformation: genomic structural alterations, epigenomic remodeling, atypical gene expression, etc. There is also a high degree of variability between different cancer types (101, 147). In this context, using Hi-C *in situ* ***data, we have corroborated important structural and functional aspects that affect the neoplastic cells:*** *(i)* ***an overall spatial conservation of chromatin in normal B cells with CLL and MCL neoplastic samples (~70%***

of the total genome), (ii) even with this high degree of conservation, a significant number of changed compartments were detected between cancer samples (MBC, CLL) with normal B-cell subpopulations but also between cancer subtypes, showing tumor-specific changes in the 3D genome, and (iii) the link between the changes in cancer genome and epigenome was correlated with a 3D chromatin reorganization of the cells (115, 146). Interestingly, the significant regions detected in CLL mainly progress towards an inactivation state, this observation is consistent with a recent study in which an enhanced heterochromatin organization in the neoplastic sample was suggested compared to normal NBC cells (146).

Chromosome conformation capture techniques can be used, not only to characterize the global chromatin architecture, but also as a diagnostic tool to distinguish normal and cancerous cell types. This idea is supported by the presence of blocks of compartment changes between normal and neoplastic samples, such as the 2Mb region on chromosome 5 associated with the silencing of EBF1 in CLL, and the entire 2p25.2 chromosome band that embedded SOX11. In fact, low levels of Early B-cell Factor 1 (EBF1), a key B-cell transcription factor, can result in reduced levels of B-cell signaling that may contribute to an anergic phenotype of CLL cells, granting it a potential diagnostic value (148). High levels of SOX11, an oncogene specific of clinical-aggressive MCL, have been demonstrated to be highly sensitive molecular marker to classify the different entities of this kind of lymphoma (149).

Hierarchical chromatin organization detected by TADpole

While a standard mathematical procedure exists to segment the genome in (sub)compartments (46), the different ways to algorithmically detect the topologically associated domains are hugely diverse. Linear scores, graph theory models and clustering methods are some of the approaches that have been applied to determine TADs; each one with its weaknesses and strengths. The variable results from these multiple approaches point to the fact that

there is not a standard definition of what a TAD is, and this may potentially affect the conclusions of the scientific studies that are based on them (151). Basically, two main trends were observed among these different algorithms: TADs are defined as individual and disjoint domains, or TADs are considered a hierarchy of block-like structures. From the last observation, concepts such as "metaTADs" or "sub-TADs" emerged from the need to characterize the chromatin organization at different scales (126, 127). Instead, we propose that these subdivisions represent different levels of the spectrum of chromatin scales that bring us the possibility to extract more details about how the chromatin is organized at scales of a few kilobases.

A recent review article, which compares a total of 22 TAD callers, stated that the limited concordance between the results was evidenced when the number and size of TADs were analyzed on multiple scales of resolution. This fact reflects an underlying hierarchical chromatin organization that can only be partially captured by a few hierarchical methods that generally have four main limitations: *(i)* ***low TAD consistency across replicate experiments,*** *(ii)* ***need for high-quality and high-resolution data,*** *(iii)* ***high computational time and*** *(iv)* ***retrieving of only a few levels of the complete hierarchy (125). These limitations require an improvement of the hierarchical methods so they can be widespread used. TADpole addresses these specific needs characterizing the entire hierarchical structure of TADs, providing a significance assessment of each level with great consistency in terms of resolution and sequencing depth, that has been validated by multiple biological, TAD-associated features.***

TADpole adopts a different strategy than the ones proposed by other TAD caller algorithms. Firstly, the vast majority of them use a local insulation score or the total global of interactions from the normalized Hi-C map to annotate TADs. Rather, we enhance the signal-to-noise ratio of the matrix by computing its Pearson correlation coefficients. To reduce the computation

time and increase the precision of the detection, a principal component analysis (PCA). In fact, we have demonstrated that there exists an optimal set of PCs capable of identifying the hierarchical structure of chromatin extending the current application of PCA to characterize genome topology. Secondly, while other methods can retrieve a hierarchical structure of the chromatin map, TADpole provides a sensible number of significant partitions and define one level as the optimal one in an unsupervised manner, using the broken-stick model and the Calinski-Harabasz index. Thirdly, the majority of tools have multiple parameters that have to be set or can be modified by the users, which can affect the results in ways that can be hard to anticipate. Instead, TADpole does not require mandatory parameters. Fourth, TADpole uses a standard tab-separated matrix format of the Hi-C data, where each cell contains the interaction value, avoiding potential incompatibilities with special input formats that other tools require. Finally, the computational requirements and facilities to install each program are factors that the user considers to use or not use a determined software. TADpole R package is an open-source tool with minimum dependencies that can be installed with ease across platforms and operating systems.

From the technical benchmarking perspective, TADpole is robust to variation over normalization strategies (even with non-normalized data), resolution (tested between 250kb and 10kb) and sequencing depths. This suggests that TADpole is potentially suitable for sparse data, that is, for Hi-C datasets at high resolution. From the biological benchmarking perspective, the TAD boundaries detected by TADpole show significant enrichment in the main architectural proteins (CTCF, SMC3, and RAD21) over all the significant hierarchical levels detected highlighting their role to act as chromatin barriers. The H3K36me3/H3K27me3 ratio established that TADpole can annotate separately (sub)domains that present a different chromatin state as a structural unit. However, it would be interesting to

expand the biological benchmark to other types of omic layers to study in detail the relationship of each level of the chromatin hierarchy with gene expression, DNA accessibility, DNA methylation, and also increase the pool of proteins and histone modifications previously analyzed.

TADpole can also be used to interrogate a specific targeted Hi-C experiment (such as capture Hi-C) (152). This dataset comes from a mouse model comprising a series of inversions that induce gene misexpression and congenital malformations. We focus our study in a specific 1.9Mb region on chromosome 1 (chr1:73.92-75.86 Mb) where a cluster of genomic features of the *Epha4* ***locus is located. We selected this study, and specifically this region because the annotation of TADs proved to be a challenge. The authors pointed out that this gene-dense region did not show a clear structure as it lacked defined chromatin boundaries. To tackle this problem, we used TADpole and, notably, found that*** *(i)* ***the region of interest indeed had a structure and we could determine it,*** *(ii)* ***there existed clear topological dissimilarities between the control and the case experiments and*** *(iii)* ***we were able to determine the location of the highest accumulation of topological differences between different hierarchical chromatin levels using a novel score called the DiffT. Overall, we proved that the combination of TADpole and the DiffT score can contribute to study chromatin organization in a hierarchical fashion and also assess the locations of the most significant topological differences between two specific hierarchical levels.***

Based on the above findings, we can generally conclude that an integrative approach is essential to shed light on the details of how the chromatin is hierarchically organized inside the nucleus and its potential link with the functional state of the cell. Combining high throughput technologies and imaging methods, together with robust bioinformatic tools, we could establish the basis to study the intrinsic and complex regulatory cellular network. Furthermore, the recent single-cell techniques have the potential to help us

characterize the cell-to-cell variability and detect the specific genetic and epigenetic status of the cell during cell fate decision, development, aging, disease, and neoplastic transformations.

CONCLUSIONS

From* chapter 2*, we can specifically conclude:

1. ***We developed a multi-omics approach combining Hi-C** in situ **data with nine additional layers (six histone modifications, DNA accessibility, gene expression, and DNA methylation) using B cell subpopulation samples and two types of neoplastic samples patients (MCL and CLL).***
2. ***We revealed the presence of a novel intermediate and dynamic compartment enriched in poised and polycomb-repressed chromatin, which is prone to change both in normal B cell differentiation and neoplastic transformation.***
3. ***One-third of the entire genome undergoes compartment changes during B cell differentiation. These changes are mostly related to two phenomena: a widespread chromatin activation from naive to germinal center B cells and the structural reversion of the memory B cells subpopulation into a state similar to that of naive B cells.***
4. ***Even with a high degree of conservation between normal and neoplastic cells, significant compartment switches were detected between them and also between cancer subtypes, with tumor-specific changes of the 3D genome linked to epigenetic changes.***
5. ***We identified large blocks of changed compartments harboring essential neoplasia-specific genes. These included:** (i) **the silencing of** EBF1 **gene in CLL, located in a 2Mb region on chromosome 5 and** (ii) **the overexpression of** SOX11 **oncogene in clinically-aggressive MCL in a 6.1Mb region on chromosome 2. Both genes present a highly sensitive potential to act as a molecular marker.***

From chapter 3*, we can specifically conclude:*

1. ***We developed TADpole that combines straightforward bioinformatics analyses such as PCA and constrained hierarchical clustering to study continuous nested hierarchical segmentation of an all-vs-all intra-chromosomal interactions matrix.***
2. ***TADpole results in one of the largest consistencies and robustness over different Hi-C normalization strategies, resolutions and sequencing depths.***
3. ***Detected TAD boundaries are significantly enriched in the main architectural proteins associated to TADs (CTCF, SMC3, RAD21).***
4. ***We developed a DiffT score to significantly detect where the main topological differences between two hierarchical levels are accumulated.***
5. ***TADpole is a publicly-available, open-source and numerically-efficient R tool.***

REFERENCES

1. Bianconi E, Piovesan A, Facchin F, Beraudi A, Casadei R, Frabetti F, et al. An estimation of the number of cells in the human body. Annals of Human Biology. 2013;40(6):463-71.

2. Franklin RE, Gosling RG. Molecular configuration in sodium thymonucleate. Nature. 1953;171(4356):740-1.

3. Crick F, Watson JD. The Molecular Structure of Nucleic Acids: The Classic Papers from Nature, 25 April 19531953 1953. 5 p.

4. Watson JD, Baker TA, Bell SP. Molecular Biology of the Gene: Benjamin-Cummings Publishing Company; 2014 2014. 872 p.

5. Consortium IHGS, International Human Genome Sequencing C. Initial sequencing and analysis of the human genome. Nature. 2001;409(6822):860-921.

6. International Human Genome Sequencing C. Finishing the euchromatic sequence of the human genome. Nature. 2004;431(7011):931-45.

7. Hood L, Galas D. The digital code of DNA. Nature. 2003;421(6921):444-8.

8. Kellis M, Wold B, Snyder MP, Bernstein BE, Kundaje A, Marinov GK, et al. Defining functional DNA elements in the human genome. Proc Natl Acad Sci U S A. 2014;111(17):6131-8.

9. Luger K, Mäder AW, Richmond RK, Sargent DF, Richmond TJ. Crystal structure of the nucleosome core particle at 2.8 Å resolution. Nature. 1997;389(6648):251-60.

10. Fyodorov DV, Zhou B-R, Skoultchi AI, Bai Y. Emerging roles of linker histones in regulating chromatin structure and function. Nat Rev Mol Cell Biol. 2018;19(3):192-206.

11. Hergeth SP, Schneider R. The H1 linker histones: multifunctional proteins beyond the nucleosomal core particle. EMBO Rep. 2015;16(11):1439-53.

12. Mariño-Ramírez L, Kann MG, Shoemaker BA, Landsman D. Histone structure and nucleosome stability. Expert Rev Proteomics. 2005;2(5):719-29.

13. Kornberg RD, Lorch Y. Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. Cell. 1999;98(3):285-94.

14. Saha A, Wittmeyer J, Cairns BR. Chromatin remodelling: the industrial revolution of DNA around histones. Nat Rev Mol Cell Biol. 2006;7(6):437-47.

15. Subramanian V, Fields PA, Boyer LA. H2A.Z: a molecular rheostat for transcriptional control. F1000Prime Reports. 2015;7.

16. Ahmad K, Henikoff S. The histone variant H3.3 marks active chromatin by replication-independent nucleosome assembly. Mol Cell. 2002;9(6):1191-200.

17. Gansen A, Tóth K, Schwarz N, Langowski J. Opposing roles of H3- and H4-acetylation in the regulation of nucleosome structure—a FRET study. Nucleic Acids Res. 2015;43(3):1433-43.
18. Sadakierska-Chudy A, Filip M. A comprehensive view of the epigenetic landscape. Part II: Histone post-translational modification, nucleosome level, and chromatin regulation by ncRNAs. Neurotox Res. 2015;27(2):172-97.
19. Alaskhar Alhamwe B, Khalaila R, Wolf J, von Bülow V, Harb H, Alhamdan F, et al. Histone modifications and their role in epigenetics of atopy and allergic diseases. Allergy Asthma Clin Immunol. 2018;14:39.
20. Prakash K, Fournier D. Evidence for the implication of the histone code in building the genome structure. Biosystems. 2018;164:49-59.
21. Prakash K, Fournier D. Histone Code and Higher-Order Chromatin Folding: A Hypothesis. Genom Comput Biol. 2017;3(2).
22. Gerlitz G. HMGNs, DNA repair and cancer. Biochim Biophys Acta. 2010;1799(1-2):80-5.
23. Kim YZ. Altered histone modifications in gliomas. Brain Tumor Res Treat. 2014;2(1):7-21.
24. Struhl K, Segal E. Determinants of nucleosome positioning. Nat Struct Mol Biol. 2013;20(3):267-73.
25. Gangaraju VK, Bartholomew B. Mechanisms of ATP dependent chromatin remodeling. Mutat Res. 2007;618(1-2):3-17.
26. van Emmerik CL, van Ingen H. Unspinning chromatin: Revealing the dynamic nucleosome landscape by NMR. Prog Nucl Magn Reson Spectrosc. 2019;110:1-19.
27. Finch JT, Klug A. Solenoidal model for superstructure in chromatin. Proc Natl Acad Sci U S A. 1976;73(6):1897-901.
28. Woodcock CL, Frado LL, Rattner JB. The higher-order structure of chromatin: evidence for a helical ribbon arrangement. J Cell Biol. 1984;99(1 Pt 1):42-52.
29. Gross DS, Chowdhary S, Anandhakumar J, Kainth AS. Chromatin. Curr Biol. 2015;25(24):R1158-63.
30. Eltsov M, Maclellan KM, Maeshima K, Frangakis AS, Dubochet J. Analysis of cryo-electron microscopy images does not support the existence of 30-nm chromatin fibers in mitotic chromosomes in situ. Proc Natl Acad Sci U S A. 2008;105(50):19732-7.
31. Cai S, Chen C, Tan ZY, Huang Y, Shi J, Gan L. Cryo-ET reveals the macromolecular reorganization of mitotic chromosomes in vivo. Proc Natl Acad Sci U S A. 2018;115(43):10977-82.
32. Ricci MA, Manzo C, García-Parajo MF, Lakadamyali M, Cosma MP. Chromatin fibers are formed by heterogeneous groups of nucleosomes in vivo. Cell. 2015;160(6):1145-58.
33. Ou HD, Phan S, Deerinck TJ, Thor A, Ellisman MH, O'Shea CC. ChromEMT: Visualizing 3D chromatin structure and compaction in interphase and mitotic cells. Science. 2017;357(6349).

34. Stadler J, Richly H. Regulation of DNA Repair Mechanisms: How the Chromatin Environment Regulates the DNA Damage Response. Int J Mol Sci. 2017;18(8).
35. Rabl C. Uber Zelltheilung1885 1885. 330 p.
36. Cremer T, Cremer C. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. Nature Reviews Genetics. 2001;2(4):292-301.
37. Cremer T, Cremer C, Schneider T, Baumann H, Hens L, Kirsch-Volders M. Analysis of chromosome positions in the interphase nucleus of Chinese hamster cells by laser-UV-microirradiation experiments. Hum Genet. 1982;62(3):201-9.
38. Meaburn KJ, Misteli T. Cell biology: chromosome territories. Nature. 2007;445(7126):379-781.
39. Cremer M, Grasser F, Lanctôt C, Müller S, Neusser M, Zinner R, et al. Multicolor 3D fluorescence in situ hybridization for imaging interphase chromosomes. Methods Mol Biol. 2008;463:205-39.
40. Branco MR, Pombo A. Intermingling of chromosome territories in interphase suggests role in translocations and transcription-dependent associations. PLoS Biol. 2006;4(5):e138.
41. Cullen KE, Kladde MP, Seyfred MA. Interaction between transcription regulatory regions of prolactin chromatin. Science. 1993;261(5118):203-6.
42. Dekker J, Rippe K, Dekker M, Kleckner N. Capturing chromosome conformation. Science. 2002;295(5558):1306-11.
43. Han J, Zhang Z, Wang K. 3C and 3C-based techniques: the powerful tools for spatial genome organization deciphering. Mol Cytogenet. 2018;11:21.
44. Simonis M, Klous P, Splinter E, Moshkin Y, Willemsen R, de Wit E, et al. Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). Nat Genet. 2006;38(11):1348-54.
45. Dostie J, Richmond TA, Arnaout RA, Selzer RR, Lee WL, Honan TA, et al. Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements. Genome Res. 2006;16(10):1299-309.
46. Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science. 2009;326(5950):289-93.
47. Lajoie BR, Dekker J, Kaplan N. The Hitchhiker's guide to Hi-C analysis: practical guidelines. Methods. 2015;72:65-75.
48. Dekker J, Misteli T. Long-Range Chromatin Interactions. Cold Spring Harb Perspect Biol. 2015;7(10):a019356.

49. Yaffe E, Tanay A. Probabilistic modeling of Hi-C contact maps eliminates systematic biases to characterize global chromosomal architecture. Nat Genet. 2011;43(11):1059-65.
50. Wijchers PJ, de Laat W. Genome organization influences partner selection for chromosomal rearrangements. Trends Genet. 2011;27(2):63-71.
51. Cavalli G. Chromosome kissing. Current Opinion in Genetics & Development. 2007;17(5):443-50.
52. Cremer T, Cremer M. Chromosome territories. Cold Spring Harb Perspect Biol. 2010;2(3):a003889.
53. Maass PG, Barutcu AR, Rinn JL. Interchromosomal interactions: A genomic love story of kissing chromosomes. J Cell Biol. 2019;218(1):27-38.
54. Sun HB, Shen J, Yokota H. Size-dependent positioning of human chromosomes in interphase nuclei. Biophys J. 2000;79(1):184-90.
55. Spilianakis CG, Lalioti MD, Town T, Lee GR, Flavell RA. Interchromosomal associations between alternatively expressed loci. Nature. 2005;435(7042):637-45.
56. Tanabe H, Muller S, Neusser M, von Hase J, Calcagno E, Cremer M, et al. Evolutionary conservation of chromosome territory arrangements in cell nuclei from higher primates. Proceedings of the National Academy of Sciences. 2002;99(7):4424-9.
57. Grasser F, Neusser M, Fiegler H, Thormeyer T, Cremer M, Carter NP, et al. Replication-timing-correlated spatial chromatin arrangements in cancer and in primate interphase nuclei. J Cell Sci. 2008;121(11):1876-86.
58. Heride C, Ricoul M, Kiêu K, von Hase J, Guillemot V, Cremer C, et al. Distance between homologous chromosomes results from chromosome positioning constraints. J Cell Sci. 2010;123(Pt 23):4063-75.
59. Schneider R, Grosschedl R. Dynamics and interplay of nuclear architecture, genome organization, and gene expression. Genes Dev. 2007;21(23):3027-43.
60. Boyle S. The spatial organization of human chromosomes within the nuclei of normal and emerin-mutant cells. Human Molecular Genetics. 2001;10(3):211-9.
61. Basu R, Zhang L-F. X chromosome inactivation: A silence that needs to be broken. Genesis. 2011;49(11):821-34.
62. Kosak ST, Skok JA, Medina KL, Riblet R, Le Beau MM, Fisher AG, et al. Subnuclear compartmentalization of immunoglobulin loci during lymphocyte development. Science. 2002;296(5565):158-62.
63. Parada LA, McQueen PG, Misteli T. 10.1186/gb-2004-5-7-r44. Genome Biol. 2004;5(7):R44.
64. Shah S, Takei Y, Zhou W, Lubeck E, Yun J, Eng C-HL, et al. Dynamics and Spatial Genomics of the Nascent Transcriptome by Intron seqFISH. Cell. 2018;174(2):363-76.e16.
65. Rozwadowska N, Kolanowski T, Wiland E, Siatkowski M, Pawlak P, Malcher A, et al. Characterisation of nuclear architectural alterations

during in vitro differentiation of human stem cells of myogenic origin. PLoS One. 2013;8(9):e73231.

66. Foster HA, Abeydeera LR, Griffin DK, Bridger JM. Non-random chromosome positioning in mammalian sperm nuclei, with migration of the sex chromosomes during late spermatogenesis. J Cell Sci. 2005;118(Pt 9):1811-20.

67. Mehta IS, Kulashreshtha M, Chakraborty S, Kolthur-Seetharam U, Rao BJ. Chromosome territories reposition during DNA damage-repair response. Genome Biol. 2013;14(12):R135.

68. Pradhan R, Ranade D, Sengupta K. Emerin modulates spatial organization of chromosome territories in cells on softer matrices. Nucleic Acids Res. 2018;46(11):5561-86.

69. Harewood L, Schutz F, Boyle S, Perry P, Delorenzi M, Bickmore WA, et al. The effect of translocation-induced nuclear reorganization on gene expression. Genome Res. 2010;20(5):554-64.

70. Kemeny S, Tatout C, Salaun G, Pebrel-Richard C, Goumy C, Ollier N, et al. Spatial organization of chromosome territories in the interphase nucleus of trisomy 21 cells. Chromosoma. 2018;127(2):247-59.

71. Zhao R, Bodnar MS, Spector DL. Nuclear neighborhoods and gene expression. Curr Opin Genet Dev. 2009;19(2):172-9.

72. Fraser J, Williamson I, Bickmore WA, Dostie J. An Overview of Genome Organization and How We Got There: from FISH to Hi-C. Microbiol Mol Biol Rev. 2015;79(3):347-72.

73. Vogel MJ, Peric-Hupkes D, van Steensel B. Detection of in vivo protein–DNA interactions using DamID in mammalian cells. Nature Protocols. 2007;2(6):1467-78.

74. Guelen L, Pagie L, Brasset E, Meuleman W, Faza MB, Talhout W, et al. Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. Nature. 2008;453(7197):948-51.

75. van Steensel B, Belmont AS. Lamina-Associated Domains: Links with Chromosome Architecture, Heterochromatin, and Gene Repression. Cell. 2017;169(5):780-91.

76. Oldenburg AR, Collas P. Mapping Nuclear Lamin-Genome Interactions by Chromatin Immunoprecipitation of Nuclear Lamins. Methods Mol Biol. 2016;1411:315-24.

77. Luperchio TR, Sauria MEG, Hoskins VE, Xianrong W, DeBoy E, Gaillard M-C, et al. The repressive genome compartment is established early in the cell cycle before forming the lamina associated domains. Genomics. 2018.

78. Peric-Hupkes D, Meuleman W, Pagie L, Bruggeman SWM, Solovei I, Brugman W, et al. Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation. Mol Cell. 2010;38(4):603-13.

79. Finlan LE, Sproul D, Thomson I, Boyle S, Kerr E, Perry P, et al. Recruitment to the nuclear periphery can alter expression of genes in human cells. PLoS Genet. 2008;4(3):e1000039.
80. Gibcus JH, Dekker J. The hierarchy of the 3D genome. Mol Cell. 2013;49(5):773-82.
81. Raices M, D'Angelo MA. Nuclear pore complexes and regulation of gene expression. Current Opinion in Cell Biology. 2017;46:26-32.
82. D'Angelo MA. Nuclear pore complexes as hubs for gene regulation. Nucleus. 2018;9(1):142-8.
83. Cisse II, Izeddin I, Causse SZ, Boudarene L, Senecal A, Muresan L, et al. Real-time dynamics of RNA polymerase II clustering in live human cells. Science. 2013;341(6146):664-7.
84. Umlauf D, Mourad R. The 3D genome: From fundamental principles to disease and cancer. Semin Cell Dev Biol. 2019;90:128-37.
85. Pueschel R, Coraggio F, Meister P. From single genes to entire genomes: the search for a function of nuclear organization. Development. 2016;143(6):910-23.
86. Schoenfelder S, Sexton T, Chakalova L, Cope NF, Horton A, Andrews S, et al. Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. Nat Genet. 2010;42(1):53-61.
87. Mao YS, Zhang B, Spector DL. Biogenesis and function of nuclear bodies. Trends Genet. 2011;27(8):295-306.
88. Grosberg A, Rabin Y, Havlin S, Neer A. Crumpled Globule Model of the Three-Dimensional Structure of DNA. Europhysics Letters (EPL). 1993;23(5):373-8.
89. Mirny LA. The fractal globule as a model of chromatin architecture in the cell. Chromosome Res. 2011;19(1):37-51.
90. Sanborn AL, Rao SSP, Huang S-C, Durand NC, Huntley MH, Jewett AI, et al. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. Proc Natl Acad Sci U S A. 2015;112(47):E6456-65.
91. Ay F, Noble WS. Analysis methods for studying the 3D architecture of the genome. Genome Biol. 2015;16:183.
92. Bonev B, Cavalli G. Organization and function of the 3D genome. Nat Rev Genet. 2016;17(11):661-78.
93. Xie WJ, Meng L, Liu S, Zhang L, Cai X, Gao YQ. Structural Modeling of Chromatin Integrates Genome Features and Reveals Chromosome Folding Principle. Sci Rep. 2017;7(1):2818.
94. Dong P, Tu X, Chu P-Y, Lü P, Zhu N, Grierson D, et al. 3D Chromatin Architecture of Large Plant Genomes Determined by Local A/B Compartments. Mol Plant. 2017;10(12):1497-509.
95. Cruz-Molina S, Respuela P, Tebartz C, Kolovos P, Nikolic M, Fueyo R, et al. PRC2 Facilitates the Regulatory Topology Required for Poised Enhancer Function during Pluripotent Stem Cell Differentiation. Cell Stem Cell. 2017;20(5):689-705.e9.

96. Kundu S, Ji F, Sunwoo H, Jain G, Lee JT, Sadreyev RI, et al. Polycomb Repressive Complex 1 Generates Discrete Compacted Domains that Change during Differentiation. Mol Cell. 2017;65(3):432-46.e5.

97. Boettiger AN, Bintu B, Moffitt JR, Wang S, Beliveau BJ, Fudenberg G, et al. Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. Nature. 2016;529(7586):418-22.

98. Rada-Iglesias A, Grosveld FG, Papantonis A. Forces driving the three-dimensional folding of eukaryotic genomes. Molecular Systems Biology. 2018;14(6):e8214.

99. Isono K, Endo TA, Ku M, Yamada D, Suzuki R, Sharif J, et al. SAM domain polymerization links subnuclear clustering of PRC1 to gene silencing. Dev Cell. 2013;26(6):565-77.

100. Imakaev M, Fudenberg G, McCord RP, Naumova N, Goloborodko A, Lajoie BR, et al. Iterative correction of Hi-C data reveals hallmarks of chromosome organization. Nat Methods. 2012;9(10):999-1003.

101. Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell. 2014;159(7):1665-80.

102. Chen Y, Zhang Y, Wang Y, Zhang L, Brinkman EK, Adam SA, et al. Mapping 3D genome organization relative to nuclear compartments using TSA-Seq as a cytological ruler. J Cell Biol. 2018;217(11):4025-48.

103. Eagen KP. Principles of Chromosome Architecture Revealed by Hi-C. Trends Biochem Sci. 2018;43(6):469-78.

104. Rowley MJ, Nichols MH, Lyu X, Ando-Kuri M, Rivera ISM, Hermetz K, et al. Evolutionarily Conserved Principles Predict 3D Chromatin Organization. Mol Cell. 2017;67(5):837-52.e7.

105. Wang S, Su J-H, Beliveau BJ, Bintu B, Moffitt JR, Wu C-T, et al. Spatial organization of chromatin domains and compartments in single chromosomes. Science. 2016;353(6299):598-602.

106. Nir G, Farabella I, Pérez Estrada C, Ebeling CG, Beliveau BJ, Sasaki HM, et al. Walking along chromosomes with super-resolution imaging, contact maps, and integrative modeling. PLoS Genet. 2018;14(12):e1007872.

107. Dixon JR, Jung I, Selvaraj S, Shen Y, Antosiewicz-Bourget JE, Lee AY, et al. Chromatin architecture reorganization during stem cell differentiation. Nature. 2015;518(7539):331-6.

108. Bertero A, Fields PA, Ramani V, Bonora G, Yardimci GG, Reinecke H, et al. Dynamics of genome reorganization during human cardiogenesis reveal an RBM20-dependent splicing factory. Nat Commun. 2019;10(1):1538.

109. Stadhouders R, Vidal E, Serra F, Di Stefano B, Le Dily F, Quilez J, et al. Transcription factors orchestrate dynamic interplay between genome topology and gene regulation during cell reprogramming. Nat Genet. 2018;50(2):238-49.

110. Zhou Y, Gerrard DL, Wang J, Li T, Yang Y, Fritz AJ, et al. Temporal dynamic reorganization of 3D chromatin architecture in hormone-induced breast cancer and endocrine resistance. Nat Commun. 2019;10(1):1522.
111. Amat R, Böttcher R, Le Dily F, Vidal E, Quilez J, Cuartero Y, et al. Rapid reversible changes in compartments and local chromatin organization revealed by hyperosmotic shock. Genome Res. 2019;29(1):18-28.
112. Barutcu AR, Lajoie BR, McCord RP, Tye CE, Hong D, Messier TL, et al. Chromatin interaction analysis reveals changes in small chromosome and telomere clustering between epithelial and breast cancer cells. Genome Biol. 2015;16:214.
113. Wu P, Li T, Li R, Jia L, Zhu P, Liu Y, et al. 3D genome of multiple myeloma reveals spatial genome disorganization associated with copy number variations. Nat Commun. 2017;8(1):1937.
114. Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature. 2012;485(7398):381-5.
115. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature. 2012;485(7398):376-80.
116. Despang A, Schöpflin R, Franke M, Ali S, Jerkovic I, Paliou C, et al. Functional dissection of TADs reveals non-essential and instructive roles in regulating gene expression. Genetics. 2019.
117. Le Dily F, Baù D, Pohl A, Vicent GP, Serra F, Soronellas D, et al. Distinct structural transitions of chromatin topological domains correlate with coordinated hormone-induced gene regulation. Genes Dev. 2014;28(19):2151-62.
118. Crane E, Bian Q, McCord RP, Lajoie BR, Wheeler BS, Ralston EJ, et al. Condensin-driven remodelling of X chromosome topology during dosage compensation. Nature. 2015;523(7559):240-4.
119. Sexton T, Yaffe E, Kenigsberg E, Bantignies F, Leblanc B, Hoichman M, et al. Three-dimensional folding and functional organization principles of the Drosophila genome. Cell. 2012;148(3):458-72.
120. Rowley MJ, Corces VG. Organizational principles of 3D genome architecture. Nat Rev Genet. 2018;19(12):789-800.
121. Brant L, Georgomanolis T, Nikolic M, Brackley CA, Kolovos P, van Ijcken W, et al. Exploiting native forces to capture chromosome conformation in mammalian cell nuclei. Mol Syst Biol. 2016;12(12):891.
122. Beagrie RA, Scialdone A, Schueler M, Kraemer DCA, Chotalia M, Xie SQ, et al. Complex multi-enhancer contacts captured by genome architecture mapping. Nature. 2017;543(7646):519-24.
123. Bintu B, Mateo LJ, Su J-H, Sinnott-Armstrong NA, Parker M, Kinrot S, et al. Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. Science. 2018;362(6413).

124. Razin SV, Gavrilov AA. Structural-Functional Domains of the Eukaryotic Genome. Biochemistry. 2018;83(4):302-12.
125. Zufferey M, Tavernari D, Oricchio E, Ciriello G. Comparison of computational methods for the identification of topologically associating domains. Genome Biol. 2018;19(1):217.
126. Fraser J, Ferrai C, Chiariello AM, Schueler M, Rito T, Laudanno G, et al. Hierarchical folding and reorganization of chromosomes are linked to transcriptional changes in cellular differentiation. Mol Syst Biol. 2015;11(12):852.
127. Berlivet S, Paquette D, Dumouchel A, Langlais D, Dostie J, Kmita M. Clustering of tissue-specific sub-TADs accompanies the regulation of HoxA genes in developing limbs. PLoS Genet. 2013;9(12):e1004018.
128. Mehra P, Kalani A. What's in the "fold"? Life Sci. 2018;211:118-25.
129. Nora EP, Goloborodko A, Valton A-L, Gibcus JH, Uebersohn A, Abdennur N, et al. Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. Cell. 2017;169(5):930-44.e22.
130. Lupiáñez DG, Kraft K, Heinrich V, Krawitz P, Brancati F, Klopocki E, et al. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. Cell. 2015;161(5):1012-25.
131. Flavahan WA, Drier Y, Liau BB, Gillespie SM, Venteicher AS, Stemmer-Rachamimov AO, et al. Insulator dysfunction and oncogene activation in IDH mutant gliomas. Nature. 2016;529(7584):110-4.
132. Szalaj P, Plewczynski D. Three-dimensional organization and dynamics of the genome. Cell Biol Toxicol. 2018;34(5):381-404.
133. Bonev B, Mendelson Cohen N, Szabo Q, Fritsch L, Papadopoulos GL, Lubling Y, et al. Multiscale 3D Genome Rewiring during Mouse Neural Development. Cell. 2017;171(3):557-72 e24.
134. Nagano T, Lubling Y, Varnai C, Dudley C, Leung W, Baran Y, et al. Cell-cycle dynamics of chromosomal organization at single-cell resolution. Nature. 2017;547(7661):61-7.
135. Stefano MD, Di Stefano M, Stadhouders R, Farabella I, Castillo D, Serra F, et al. Dynamic simulations of transcriptional control during cell reprogramming reveal spatial chromatin caging.
136. Nagano T, Lubling Y, Stevens TJ, Schoenfelder S, Yaffe E, Dean W, et al. Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. Nature. 2013;502(7469):59-64.
137. Liu J, Lin D, Yardimci GG, Noble WS. Unsupervised embedding of single-cell Hi-C data. Bioinformatics. 2018;34(13):i96-i104.
138. Hansen AS, Cattoglio C, Darzacq X, Tjian R. Recent evidence that TADs and chromatin loops are dynamic structures. Nucleus. 2018;9(1):20-32.

139. Dolgin E. DNA's secret weapon against knots and tangles. Nature. 2017;544(7650):284-6.
140. Nuebler J, Fudenberg G, Imakaev M, Abdennur N, Mirny LA. Chromatin organization by an interplay of loop extrusion and compartmental segregation. Proc Natl Acad Sci U S A. 2018;115(29):E6697-e706.
141. Belaghzal H, Borrman T, Stephens AD, Lafontaine DL, Venev SV, Marko JF, et al. Compartment-dependent chromatin interaction dynamics revealed by liquid chromatin Hi-C. Genomics. 2019.
142. Stadhouders R, Filion GJ, Graf T. Transcription factors and 3D genome conformation in cell-fate decisions. Nature. 2019;569(7756):345-54.
143. Hu G, Cui K, Fang D, Hirose S, Wang X, Wangsa D, et al. Transformation of Accessible Chromatin and 3D Nucleome Underlies Lineage Commitment of Early T Cells. Immunity. 2018;48(2):227-42.e8.
144. Schmitt AD, Hu M, Jung I, Xu Z, Qiu Y, Tan CL, et al. A Compendium of Chromatin Contact Maps Reveals Spatially Active Regions in the Human Genome. Cell Rep. 2016;17(8):2042-59.
145. Bunting KL, Soong TD, Singh R, Jiang Y, Béguelin W, Poloway DW, et al. Multi-tiered Reorganization of the Genome during B Cell Affinity Maturation Anchored by a Germinal Center-Specific Locus Control Region. Immunity. 2016;45(3):497-512.
146. Mallm Jp, Iskar M, Ishaque N, Klett LC, Kugler SJ, Muino JM, et al. Linking aberrant chromatin features in chronic lymphocytic leukemia to transcription factor networks. Molecular Systems Biology. 2019;15(5):e8339.
147. Sauerwald N, Kingsford C. Quantifying the similarity of topological domains across normal and cancer human cell types. Bioinformatics. 2018;34(13):i475-i83.
148. Meijers RWJ, Muggen AF, Leon LG, de Bie M, van Dongen JJM, Hendriks RW, et al. Responsiveness of Chronic Lymphocytic Leukemia cells to B cell receptor stimulation is associated with low expression of regulatory molecules of the Nuclear Factor-κB pathway. Haematologica. 2019.
149. Szostakowska M, Szymczyk M, Badowska K, Tudek B, Fabisiewicz A. SOX11 expression as a MRD molecular marker for MCL in comparison with t(11;14) and IGH rearrangement. Med Oncol. 2018;35(4):49.
150. Cristiano S, Leal A, Phallen J, Fiksel J, Adleff V, Bruhm DC, et al. Genome-wide cell-free DNA fragmentation in patients with cancer. Nature. 2019;570(7761):385-9.
151. Dali R, Blanchette M. A critical assessment of topologically associating domain prediction tools. Nucleic Acids Res. 2017;45(6):2994-3005.
152. Kraft K, Magg A, Heinrich V, Riemenschneider C, Schöpflin R, Markowski J, et al. Serial genomic inversions induce tissue-specific architectural stripes, gene misexpression and congenital malformations. Nat Cell Biol. 2019;21(3):305-10.