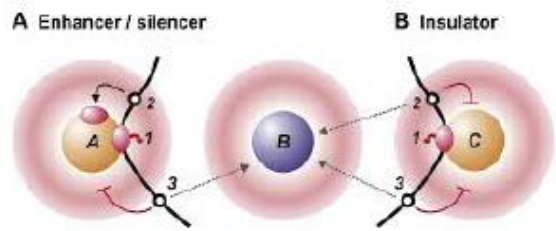
Genómica conformacional

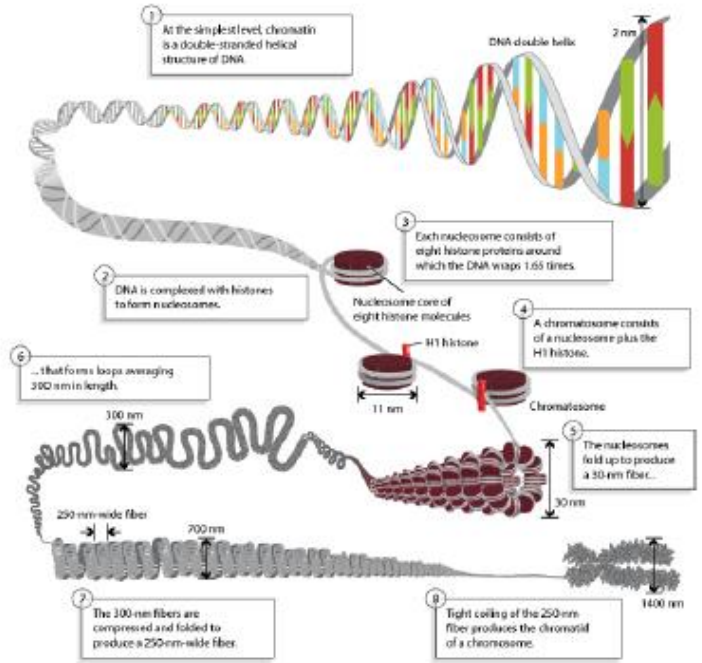
José Sotelo

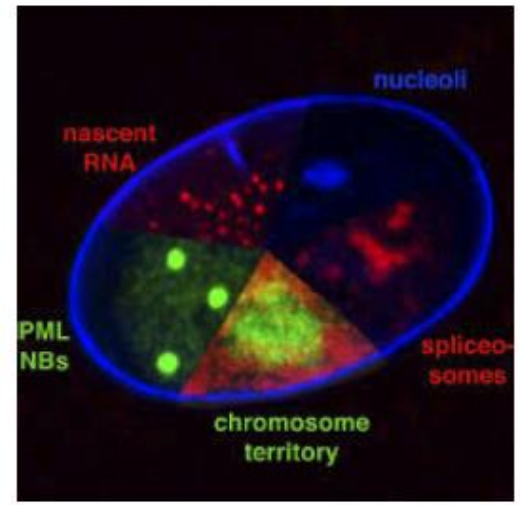
Genómica 2024

Antecedentes

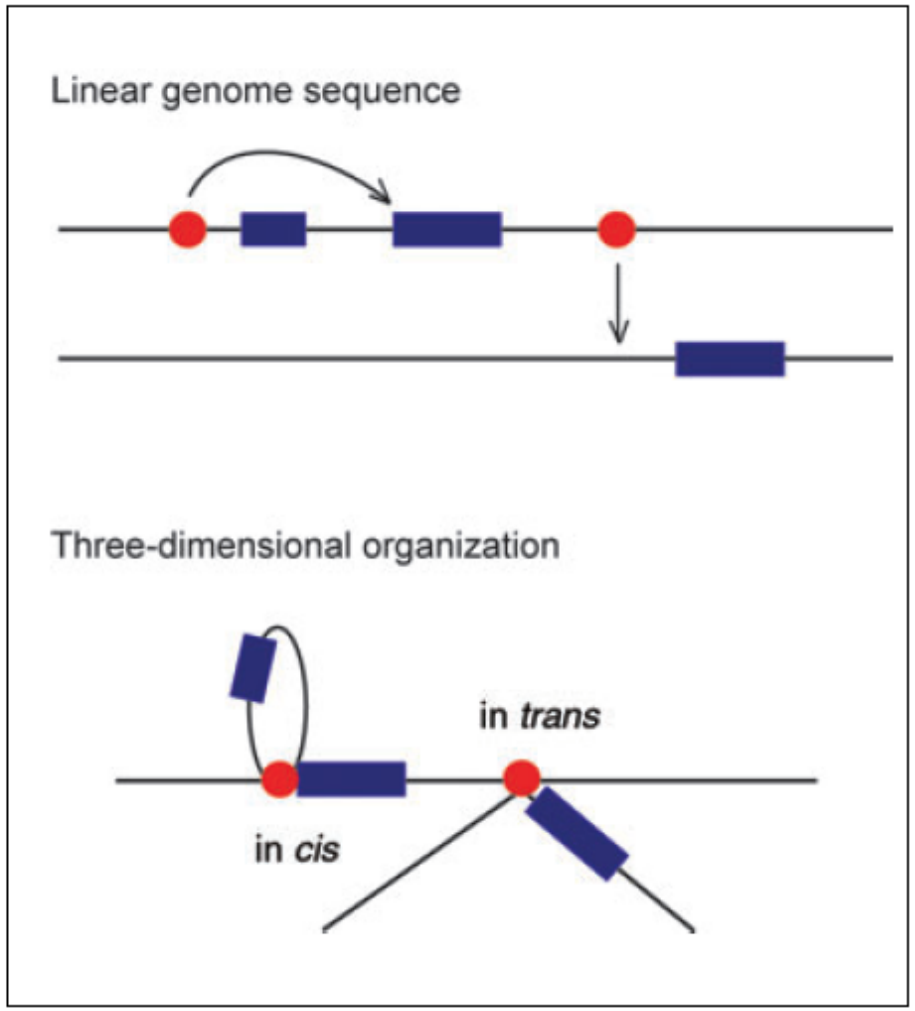
- Existencia de elementos en *cis* como promotores, potenciadores (*enhancers*), aislantes (*insulators*)
- Plegamiento y empaquetado de la cromatina.
- Regiones nucleares especializadas







Interacciones cromosómicas a distancia.



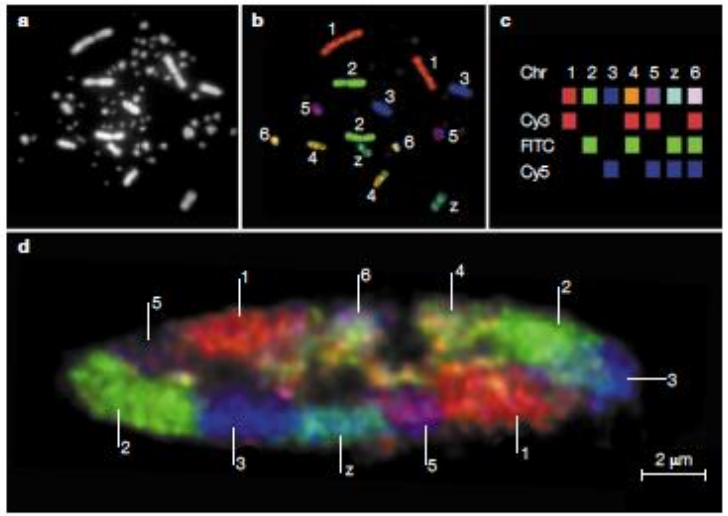
El núcleo eucariota está subdividido en compartimentos **funcionales**.

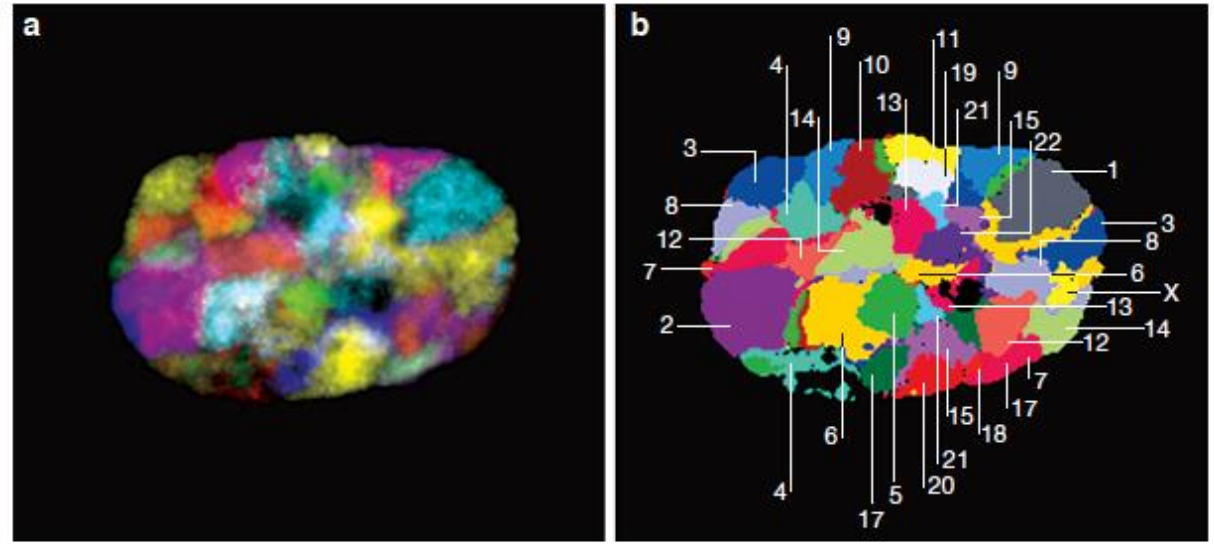
- Fábricas de transcripción
- Fábricas de replicación
- Speckles de procesamiento (splicing)
- Cuerpos de Cajal
- Cuerpos PML
- Cuerpos polycomb

Componentes **estructurales** de los compartimentos:

- Laminas (filamentos intermedios asociados a envoltura nuclear)
- Cromosomas mismos como componente estructural: heterocromatina, eucromatina

Aproximaciones por microscopía: Hibridación de fluorescencia *in situ* 3D

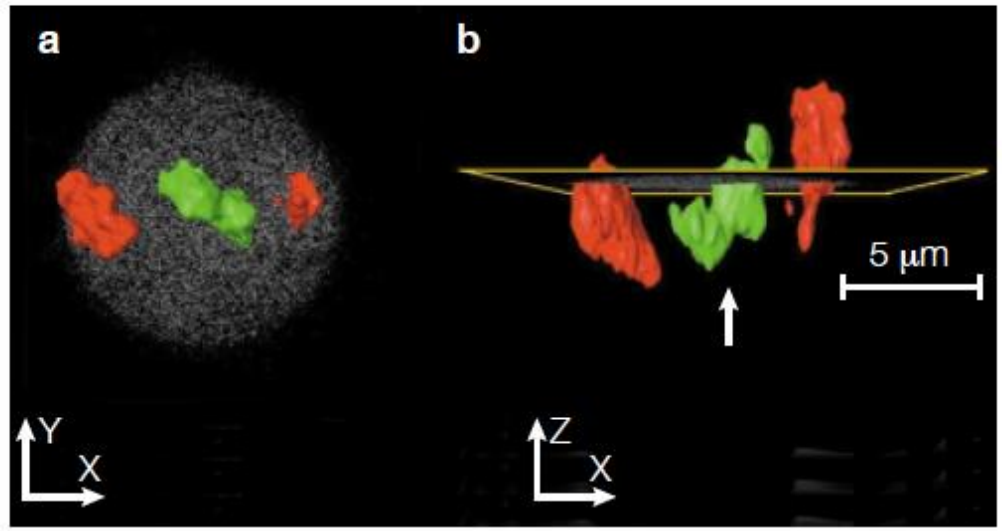




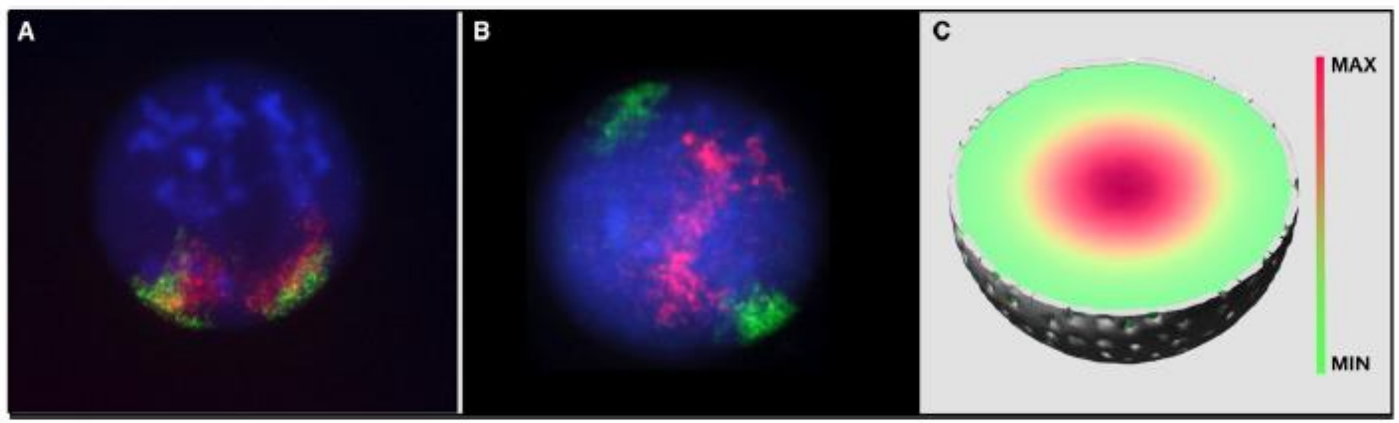
Cremer, Cremer *Nature* 2001

Speicher, Carter *Nature* 2005

Territorios y contenido de genes

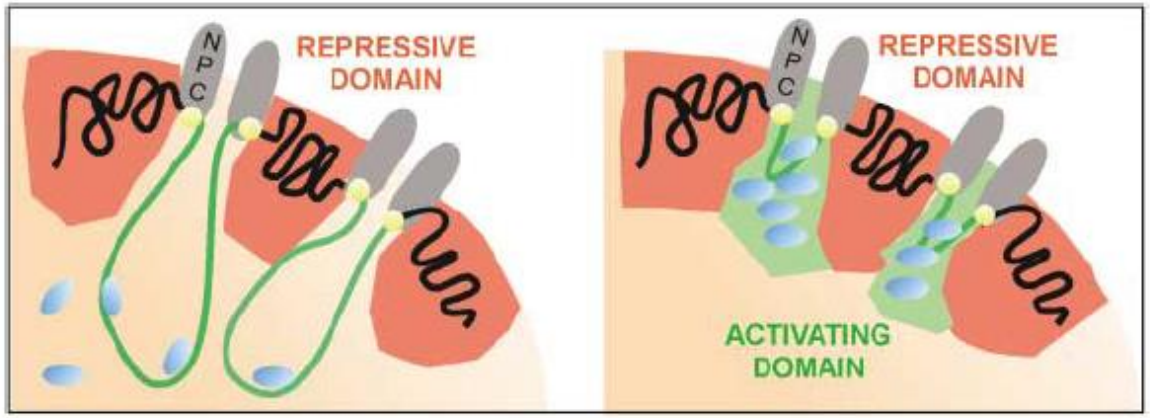


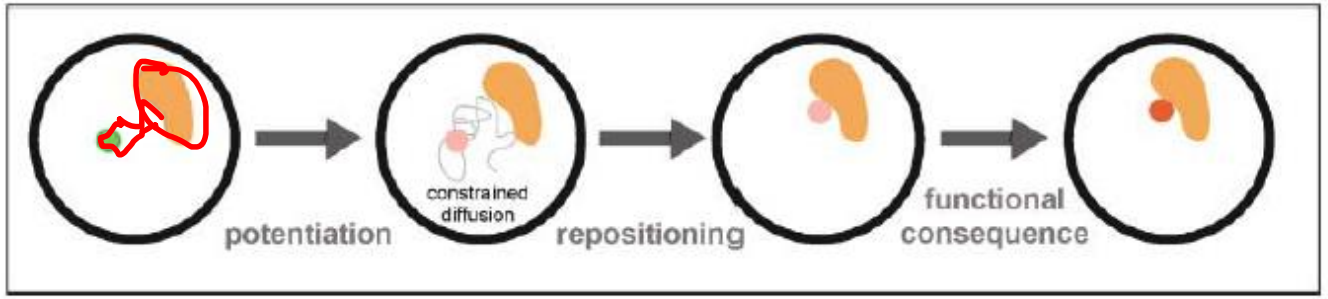
Cremer, Cremer *Nature* 2001



Bickmore, *Cell* 2013

Funciones: Rol de la posición espacial en la actividad y la estabilidad genómica

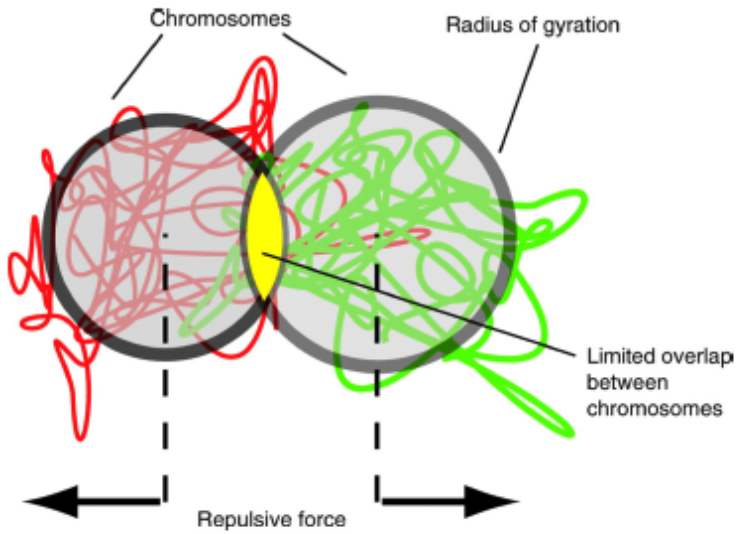


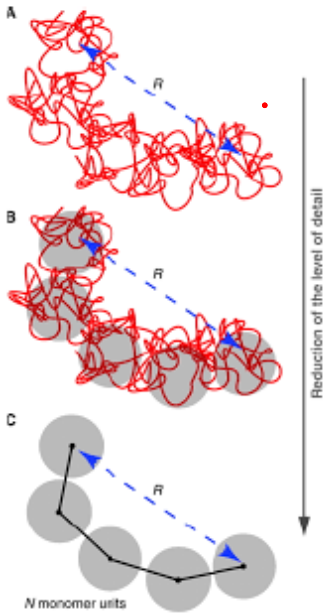


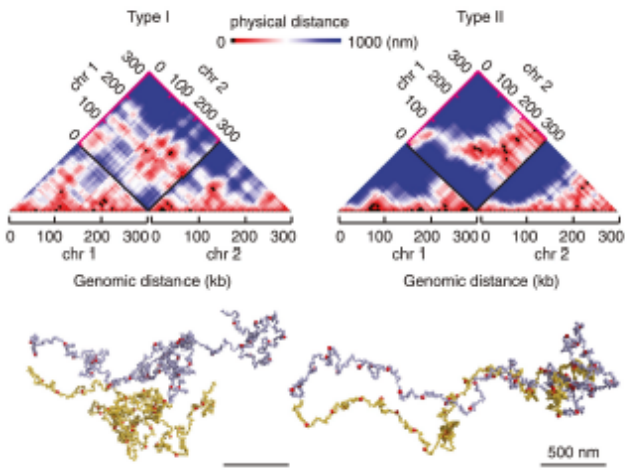
Mistelli, *Cell* 2004

Aproximaciones para el estudio de cromosomas: 3C y modelado

- 3 aproximaciones:
  - Identificación de loci que interaccionan más frecuentemente de lo esperado
  - Modelado basado en restricciones
  - Modelado basado en tratamiento de cromatina como polímero

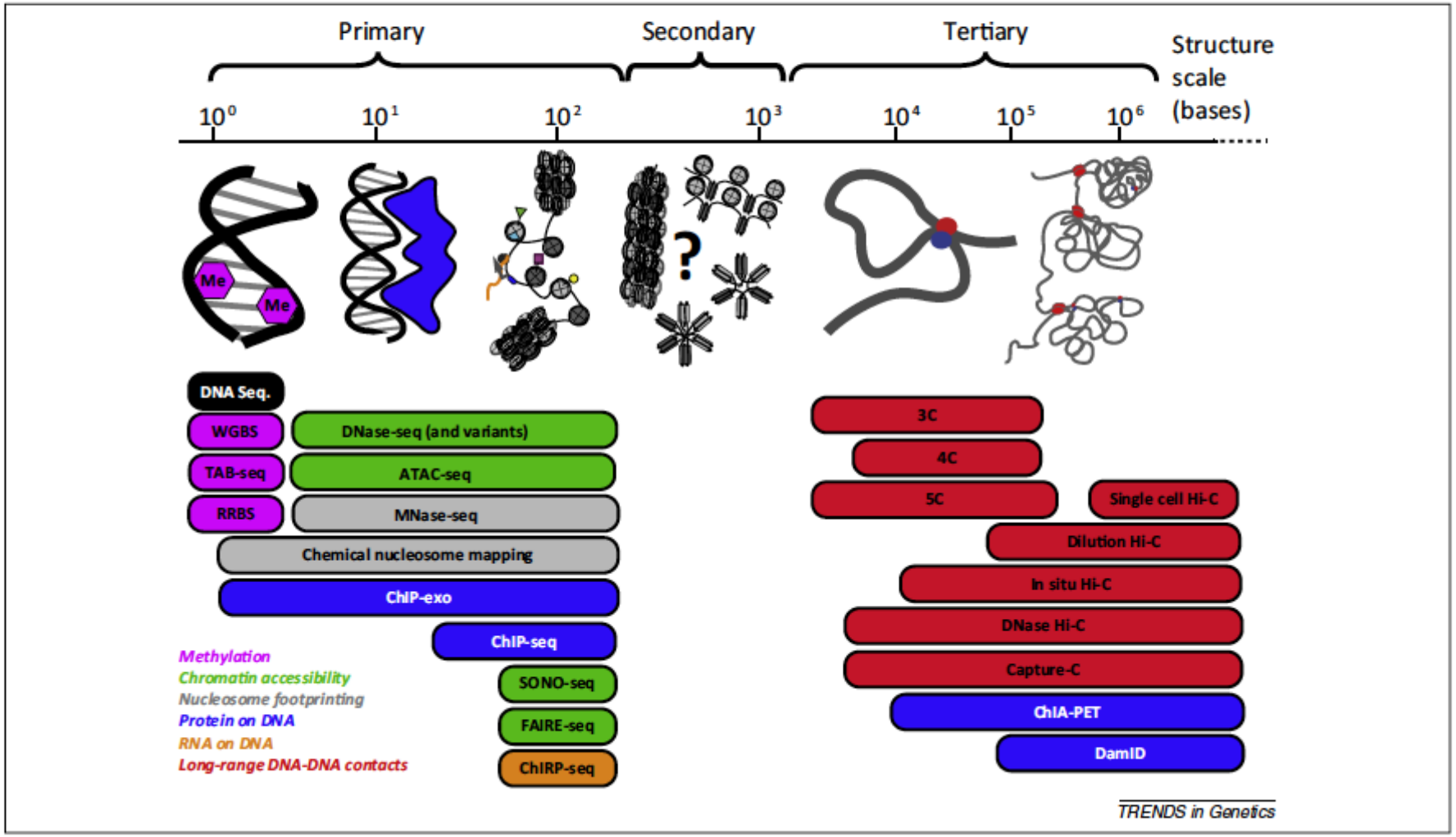






	Type I	Type II	p value
Interchromosome contacts	2.8 ± 1.9	7.2 ± 4.9	< 0.0001

Diferentes escalas estructurales y diferentes métodos de análisis



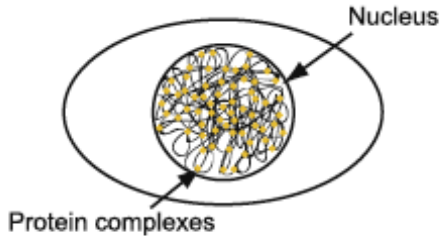
Principios del método 3C: chromosome conformation capture

- Esta metodología y sus derivados miden la frecuencia promediada por población a la cual dos loci cromosomales se asocian fisicamente en el espacio tridimensional.
- Esta medición se basa en la tendencia de estos dos loci a quedar unidos químicamente durante la adición de formaldehído.
- La cromatina intercruzada se solubiliza y fragmenta, usando una enzima de restricción.
- Los fragmentos interactuantes se ligan y purifican, creando una librería genómica de moléculas de ADN quimérico.
- La abundancia relativa de quimeras específicas está relacionada a la probabilidad de que los fragmentos que componen la quimera estén asociados en el espacio tridimensional en la población celular estudiada.
- Una biblioteca 3C incluye una gran cantidad (10^{11}) de productos de ligación entre fragmentos de aprox 4 kb y se puede analizar de muchas maneras, según los objetivos del estudio.
- 3C: “uno contra algunos”, 4C: “uno contra todos”, 5C: “muchos contra muchos”, Hi-C: “todos contra todos”, ChIA-PET (“algunos contra todos”)

C3 method

1. Crosslinking
2. Digestion
3. Ligation
4. Reverse cl
5. PCR

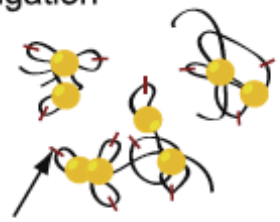
A) 1. Crosslinking



2. Digestion

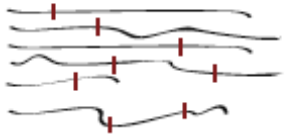
DNA - Protein clumps

3. Ligation



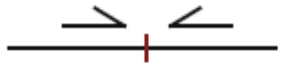
Ligation junction

4. Reverse crosslinks



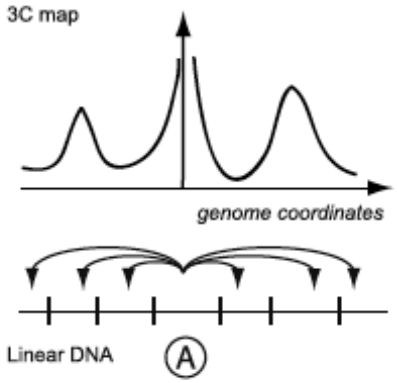
3C Library

5. PCR detection



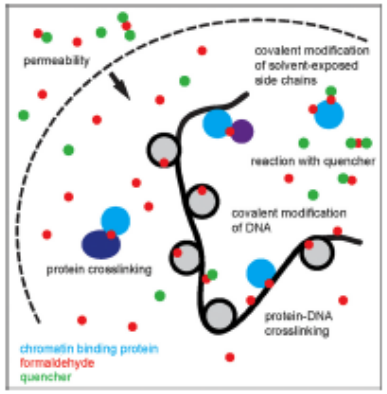
6. Gel Quantification

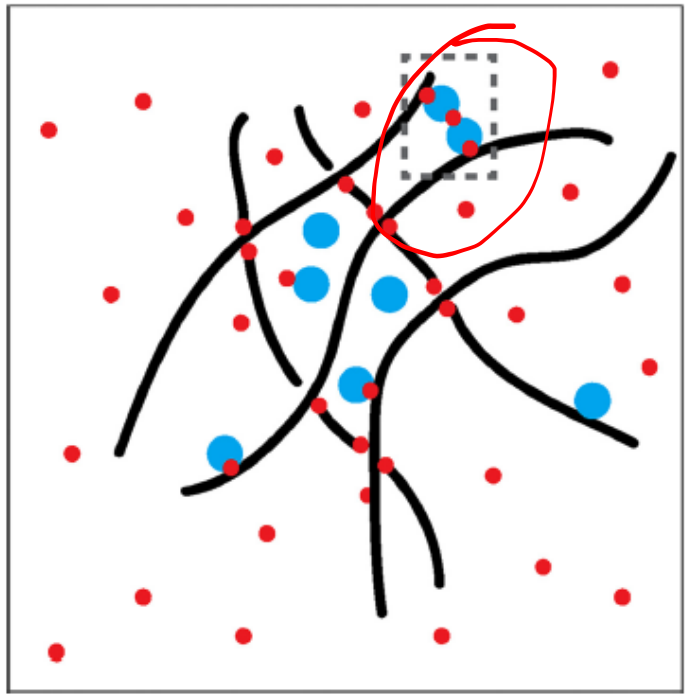
7. Plot 3C profile



Naumova *et al* 2012 Nat meth

Entrecruzamiento (*crosslinking*) de cromatina mediante formaldehído





Step 1

Protein + Formaldehyde ⇌ Methylol $\xrightarrow{-H_2O}$ Schiff Base

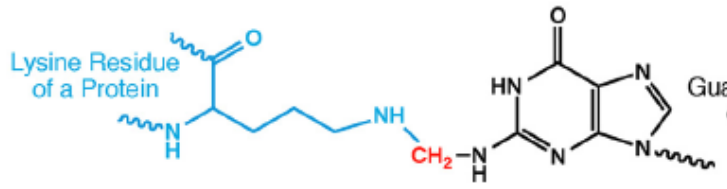
R_1-NH_2 + $H_2C=O$ ⇌ $R_1-NH-CH_2-OH$ $\xrightarrow{-H_2O}$ $R_1-N=CH_2$

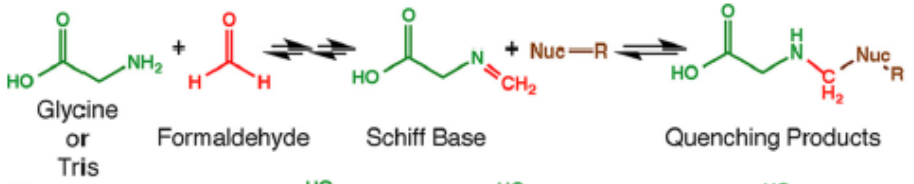
Step 2

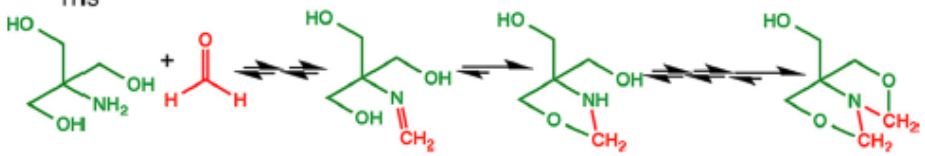
Schiff Base + DNA or Protein ⇌ Crosslinked Product

$R_1-N=CH_2$ + $Nuc-R_2$ ⇌ $R_1-NH-CH_2-Nuc-R_2$

Example Protein-DNA Crosslink



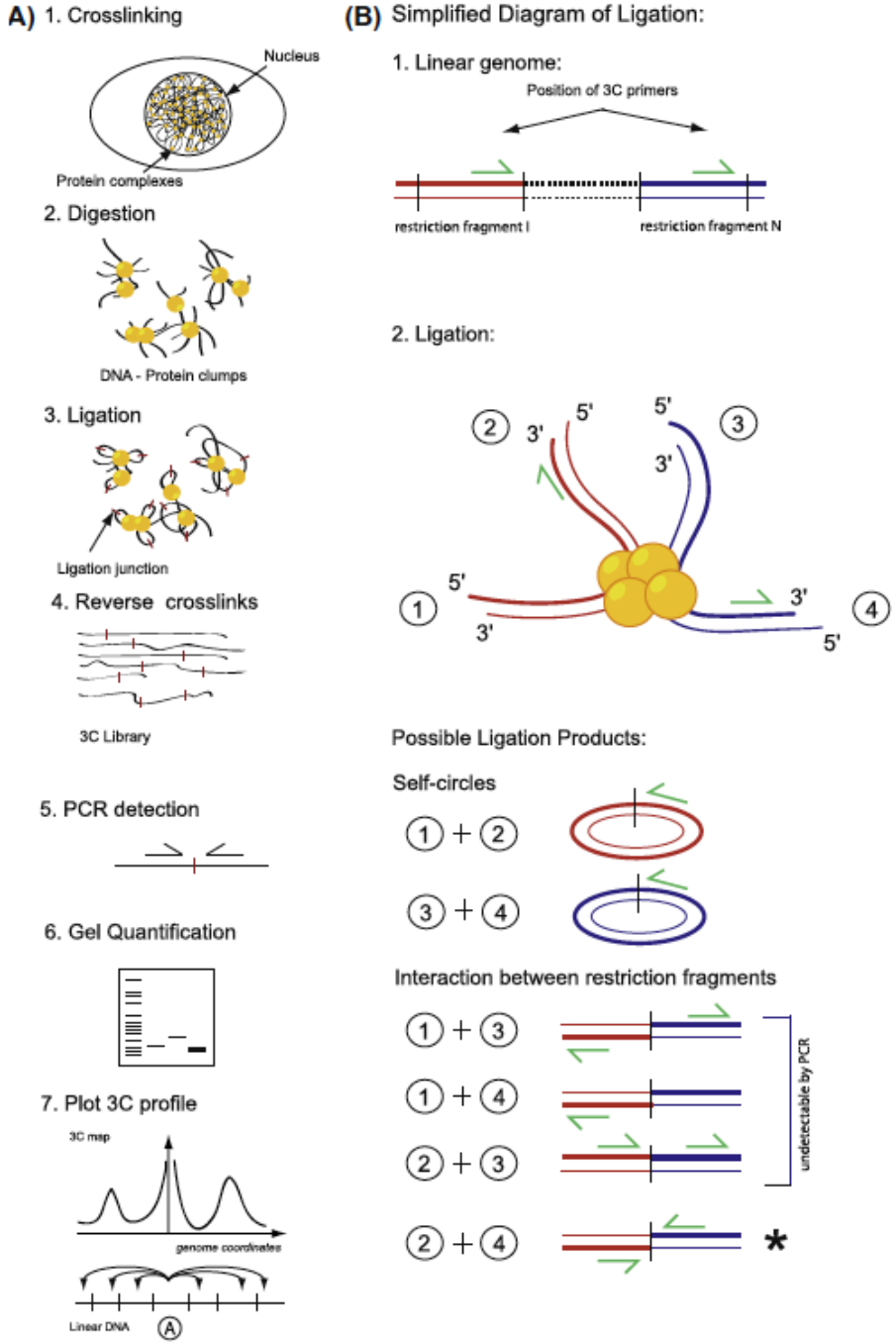




quenching

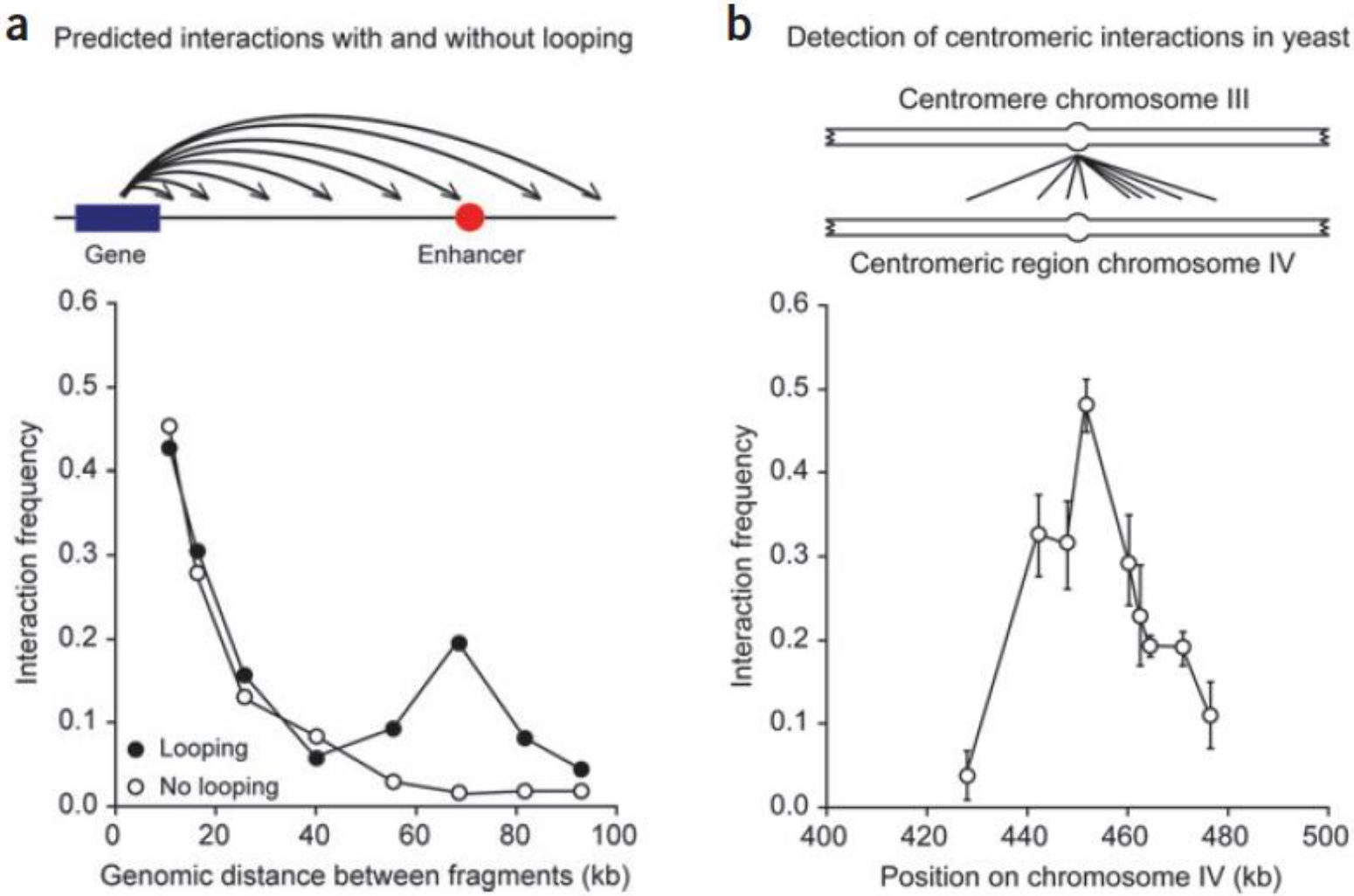
C3 method

1. Crosslinking
2. Digestion
3. Ligation
4. Reverse cl
5. PCR



Naumova *et al* 2012 Nat meth

Ejemplos de análisis 3C



Dekker, 2006

Chromosome Conformation Capture (3C) Technologies

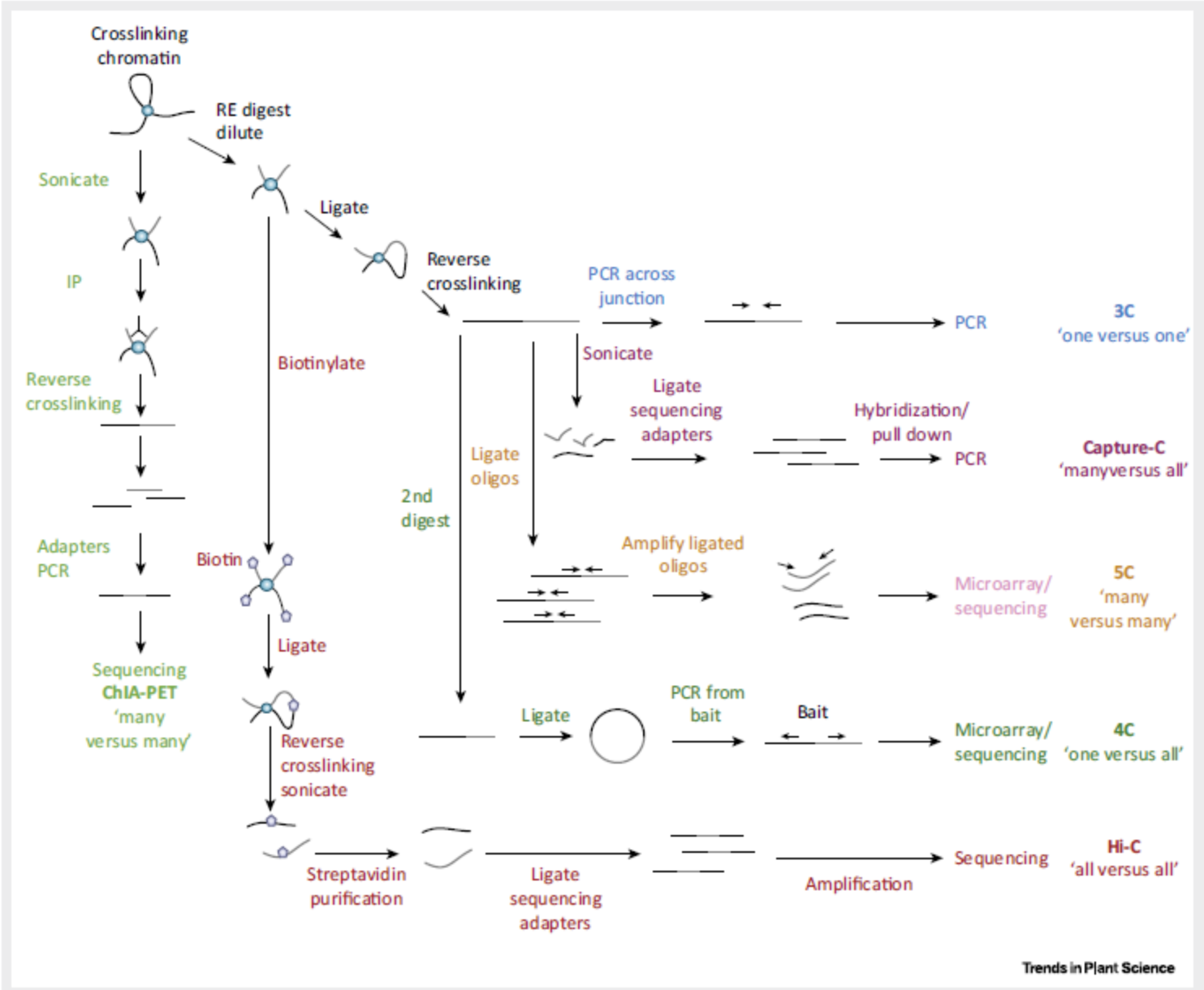
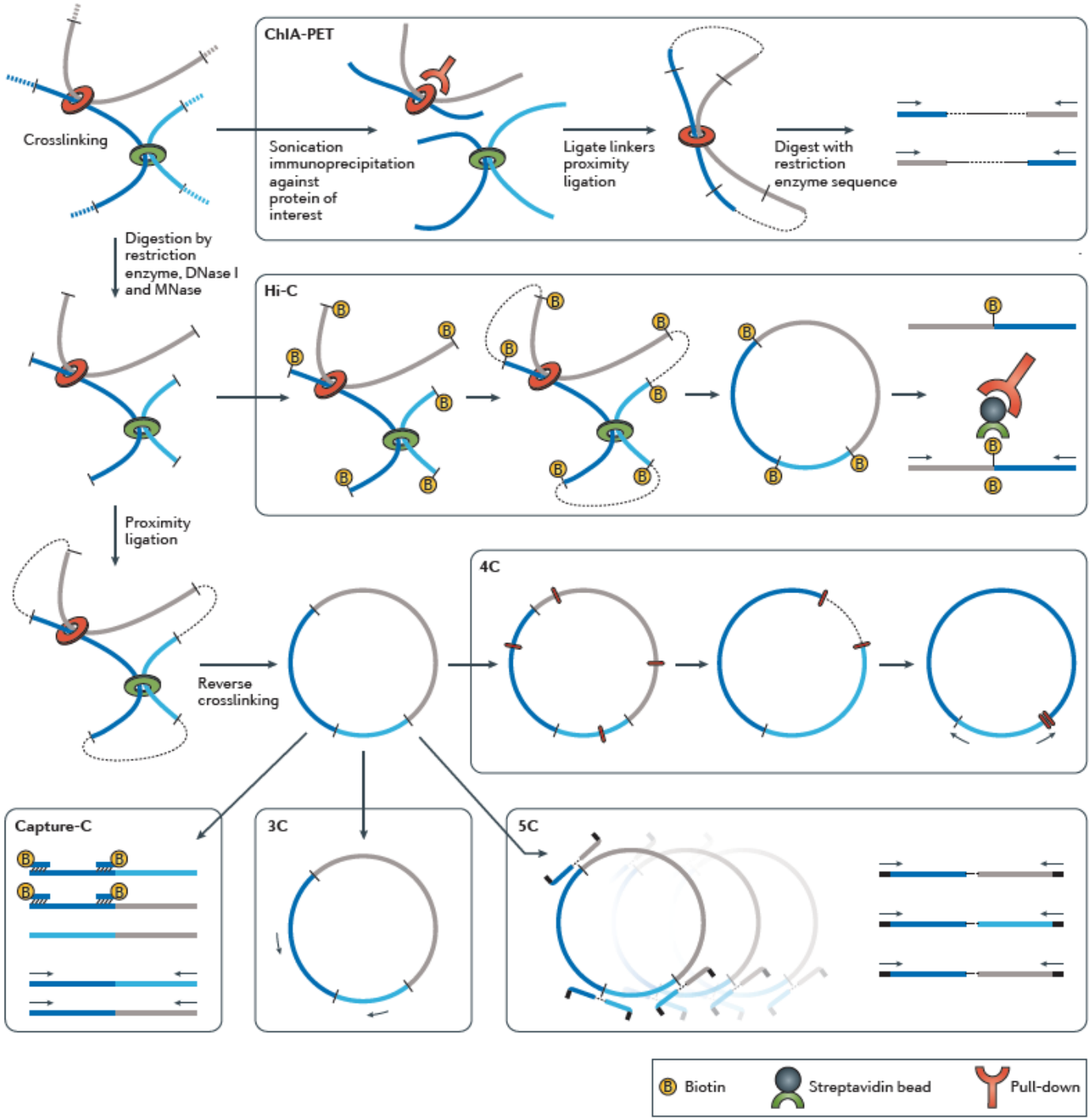


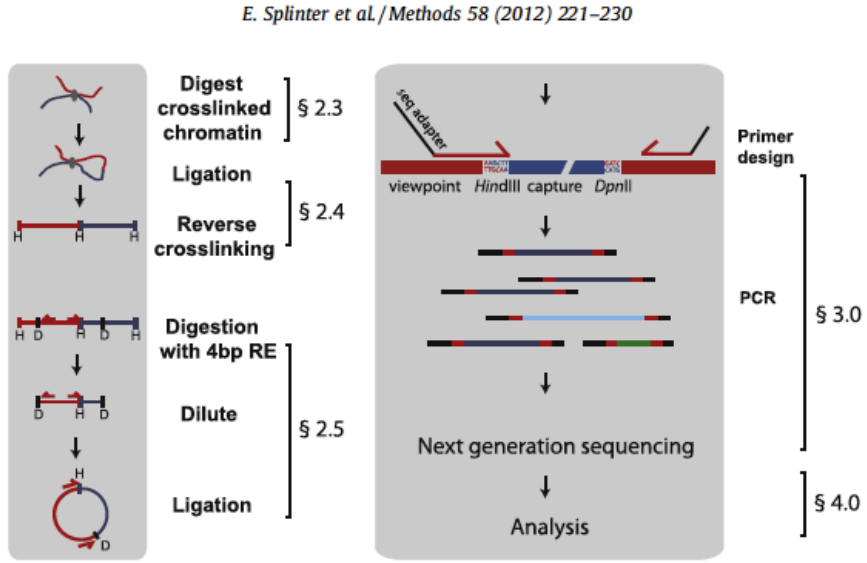
Figure I. Schematic Representation of Chromosome Conformation Capture (3C) and 3C-Derived Methods. These methods help to elucidate nuclear

Sotelo-Silveira et al., 2018

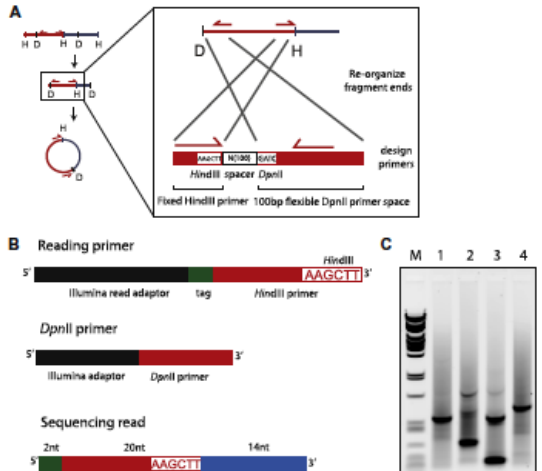

Crosslinking
ChIA-PET
Sonication immunoprecipitation against protein of interest
Ligate linkers proximity ligation
Digest with restriction enzyme sequence
Digestion by restriction enzyme, DNase I and MNase
Hi-C
Proximity ligation
Reverse crosslinking
4C
Capture-C
3C
5C
Biotin
Streptavidin bead
Pull-down

4C: circular 3C (“uno contra todos”) un cebo, muchas interacciones desconocidas

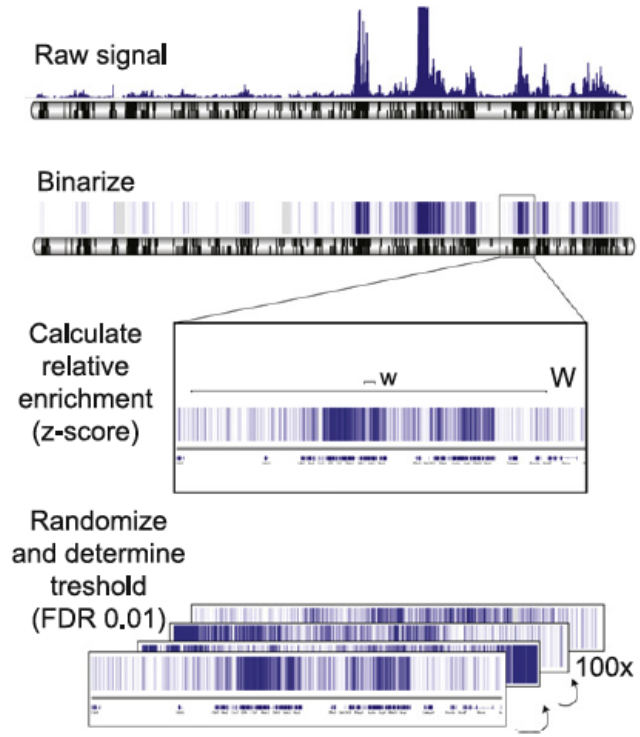
Esquema general



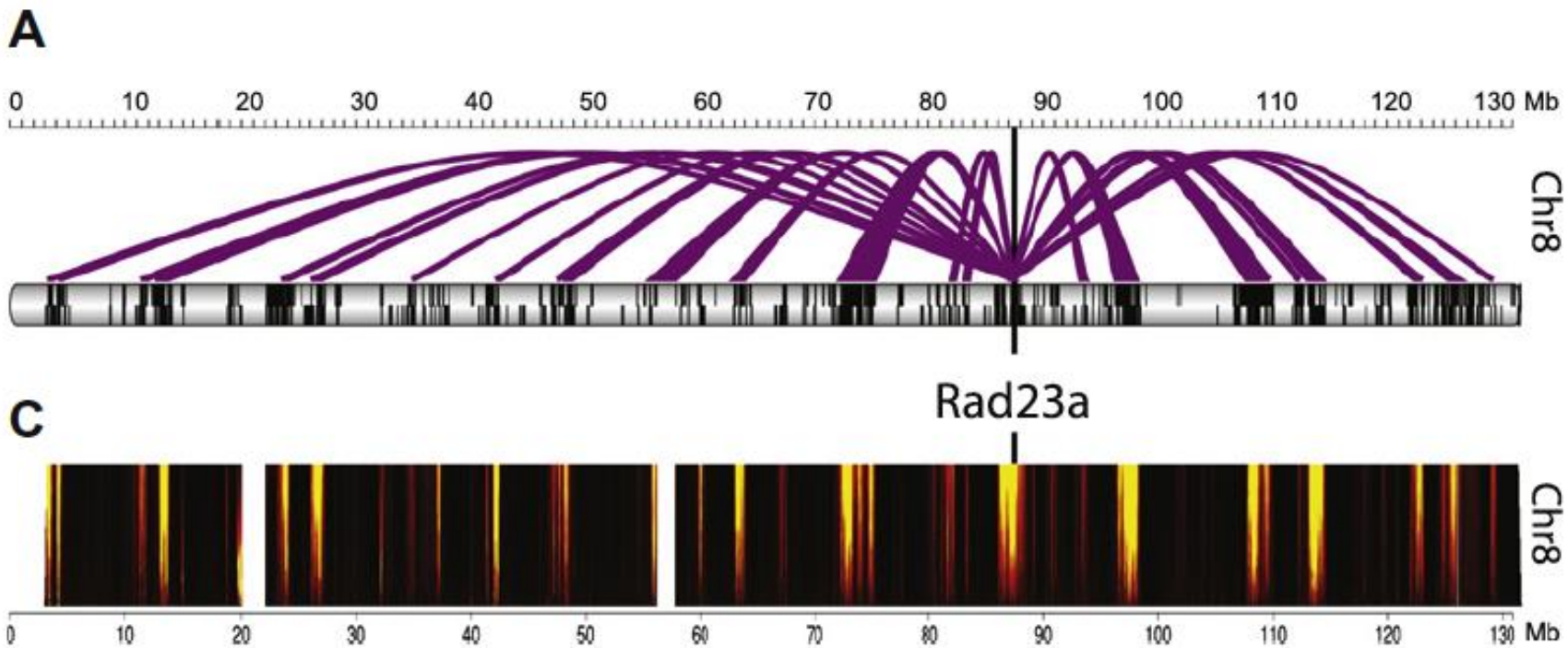
Diseño de primers

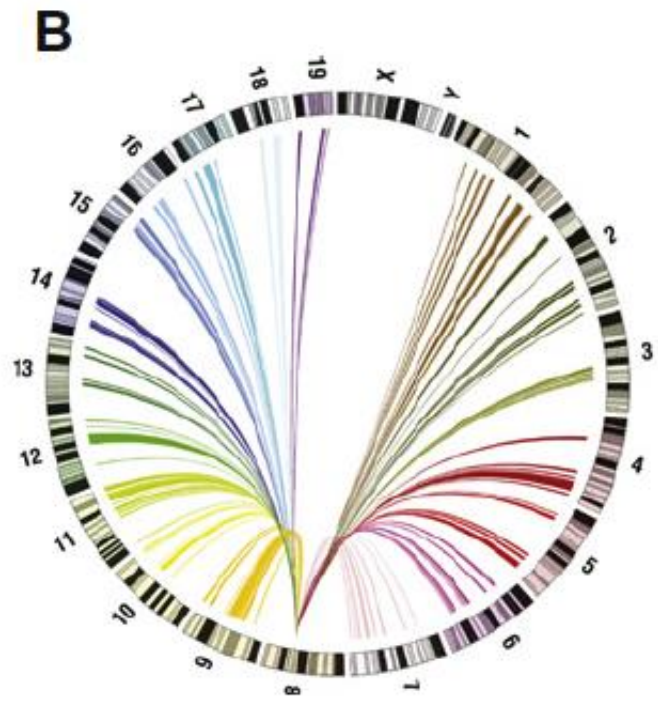


Tratamiento de datos

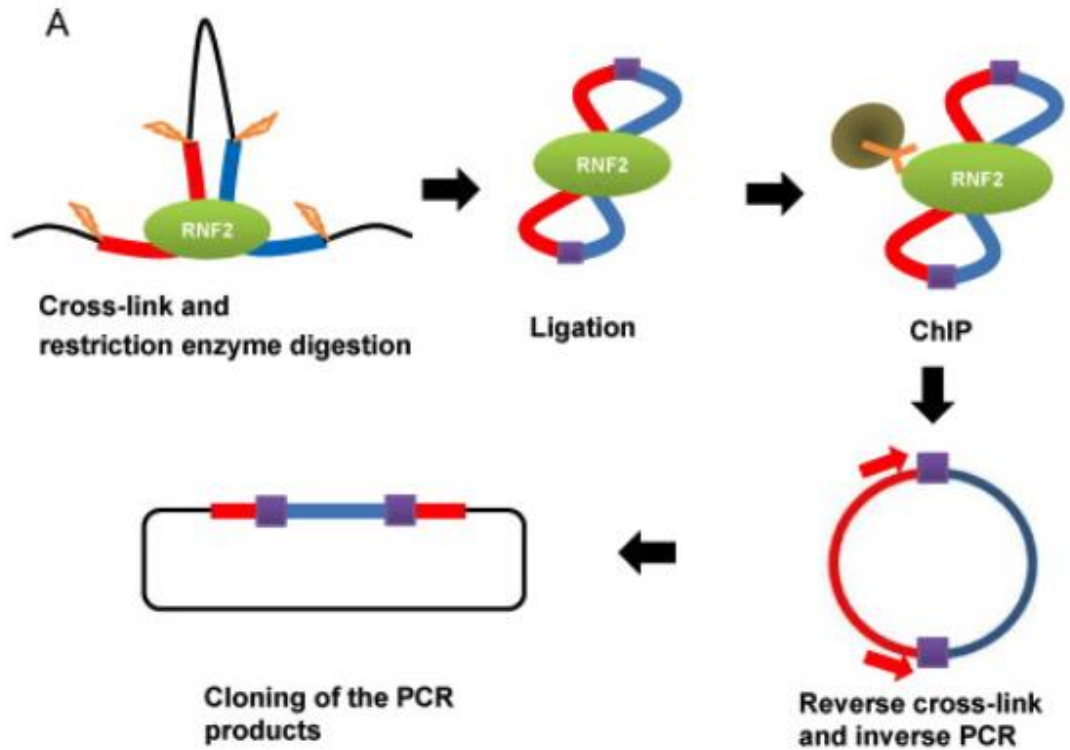


Representación de resultados 4C



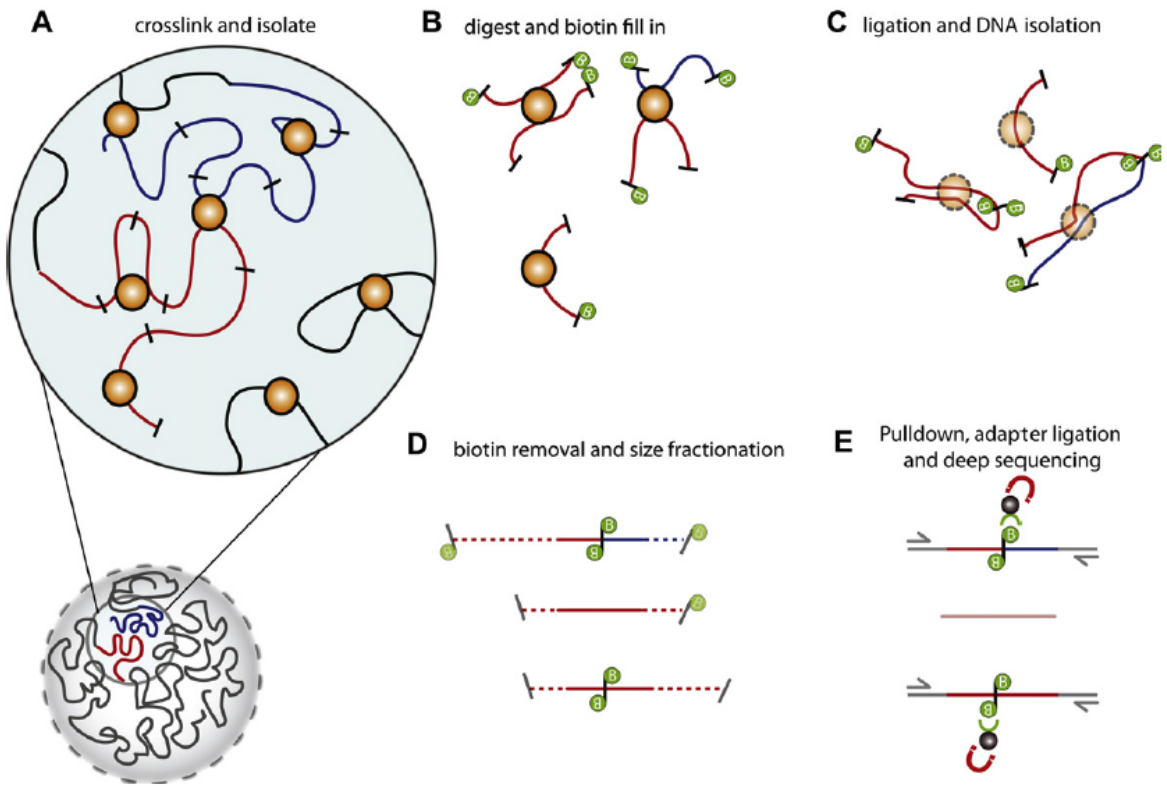


Variación de 4C: coIP-4C

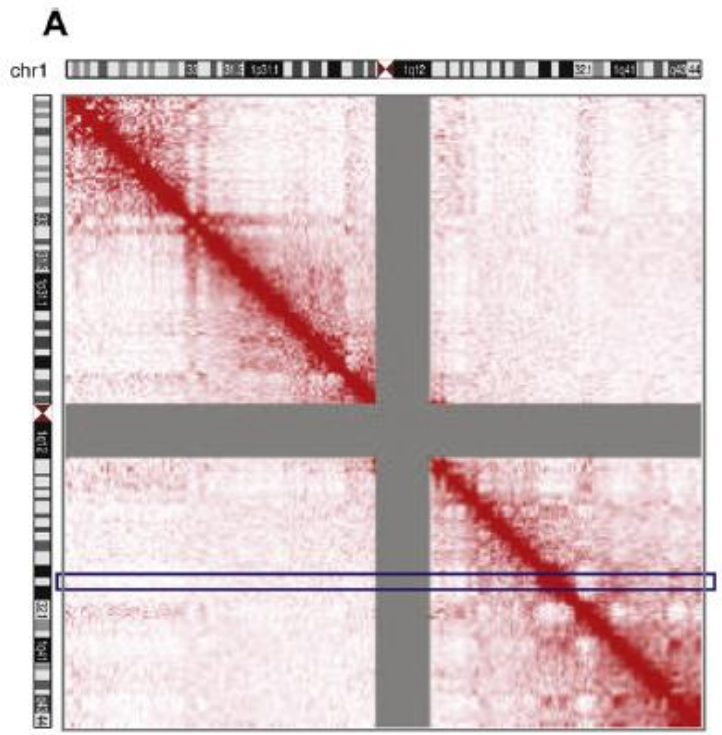


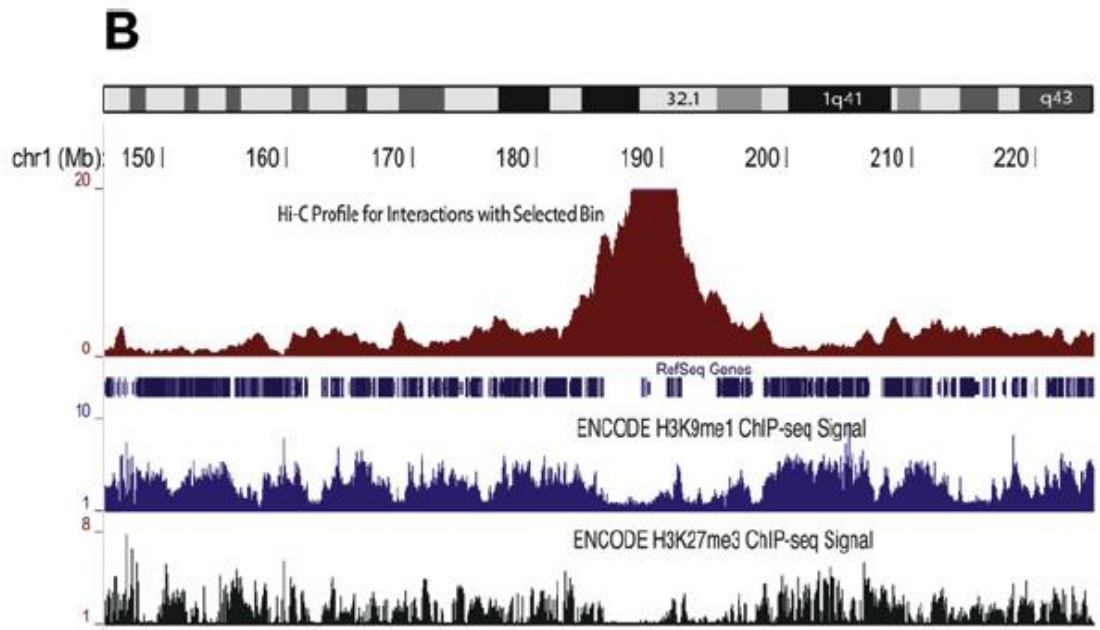
Hi-C: “todos contra todos”

Poder de la técnica: los puntos de unión se marcan con biotina y pueden purificarse y secuenciarse

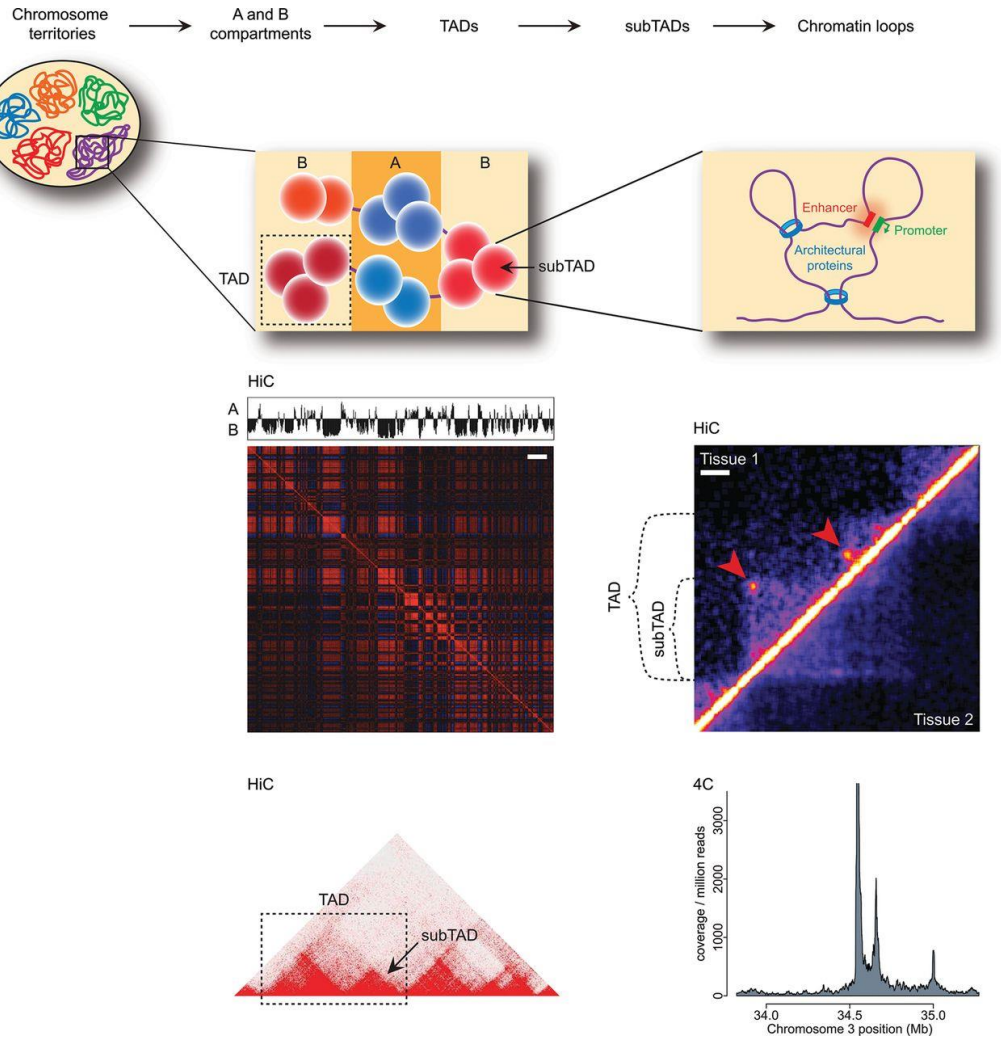


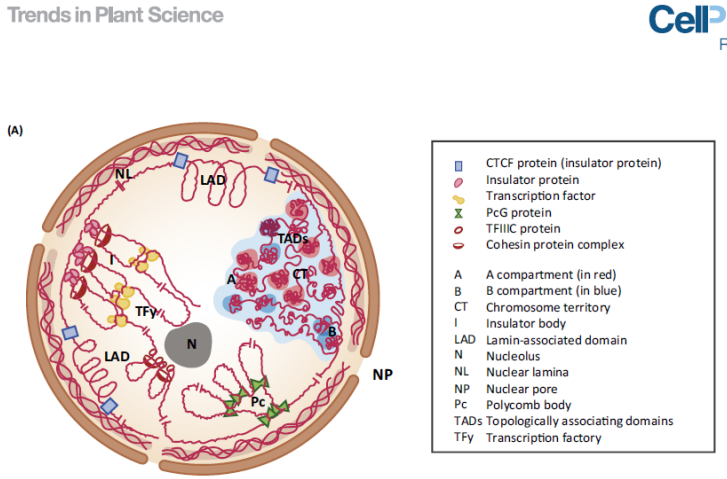
Visualización de datos Hi-C

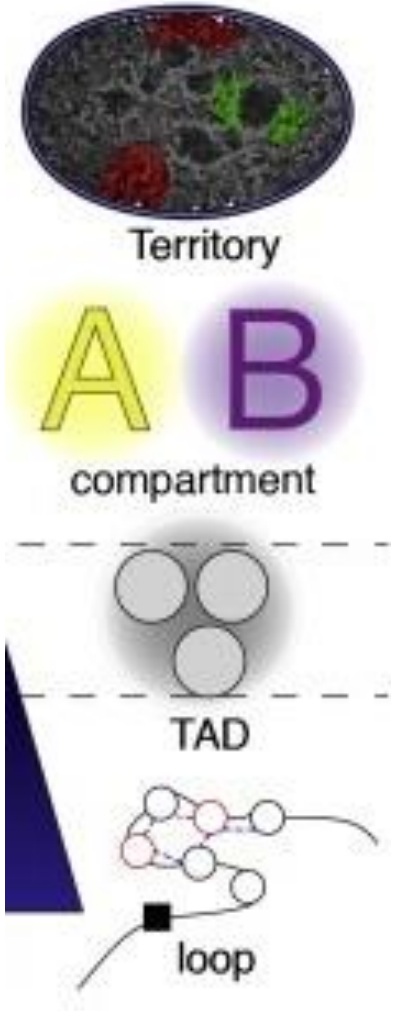




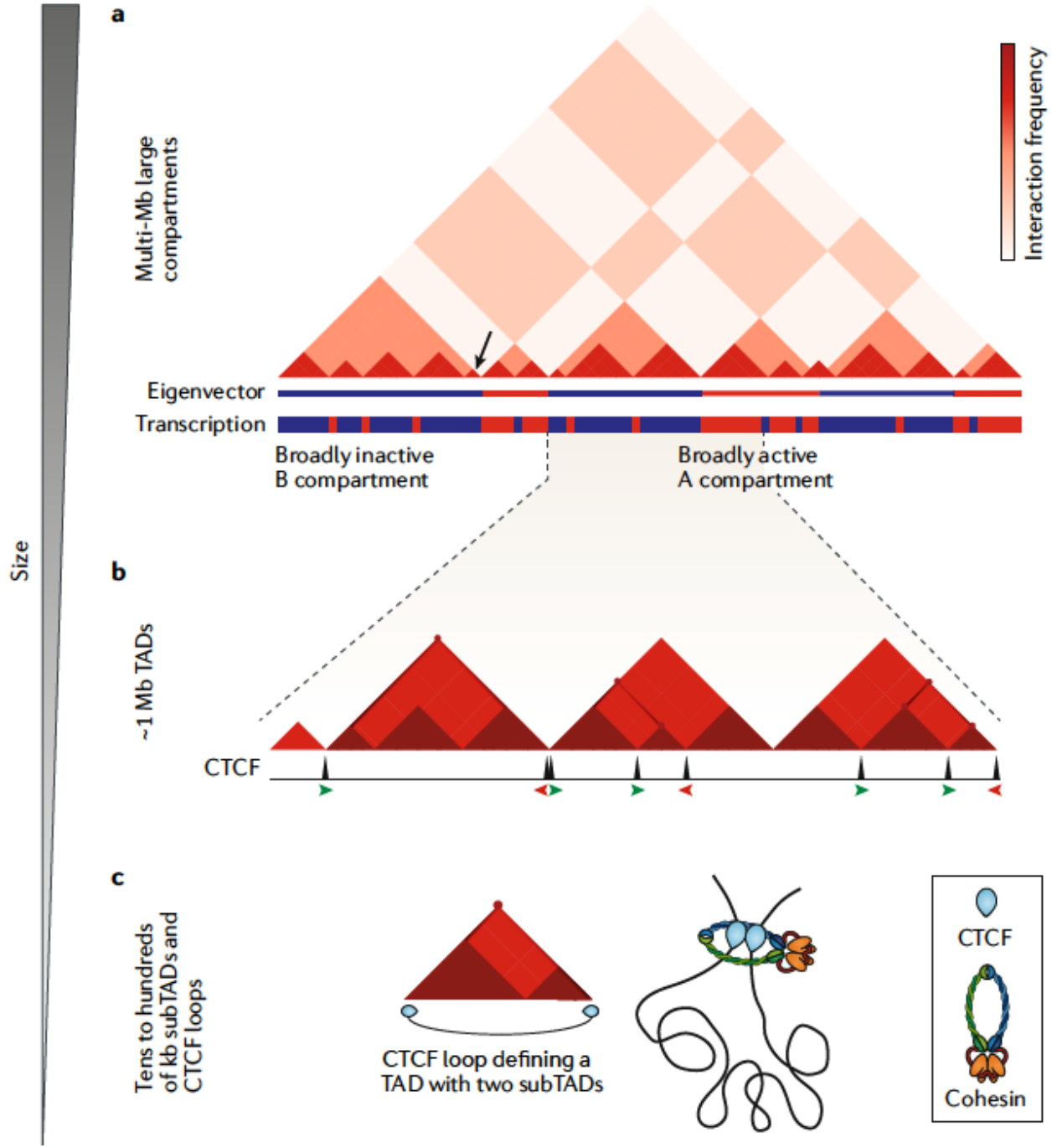
Hierarchical genome organization.

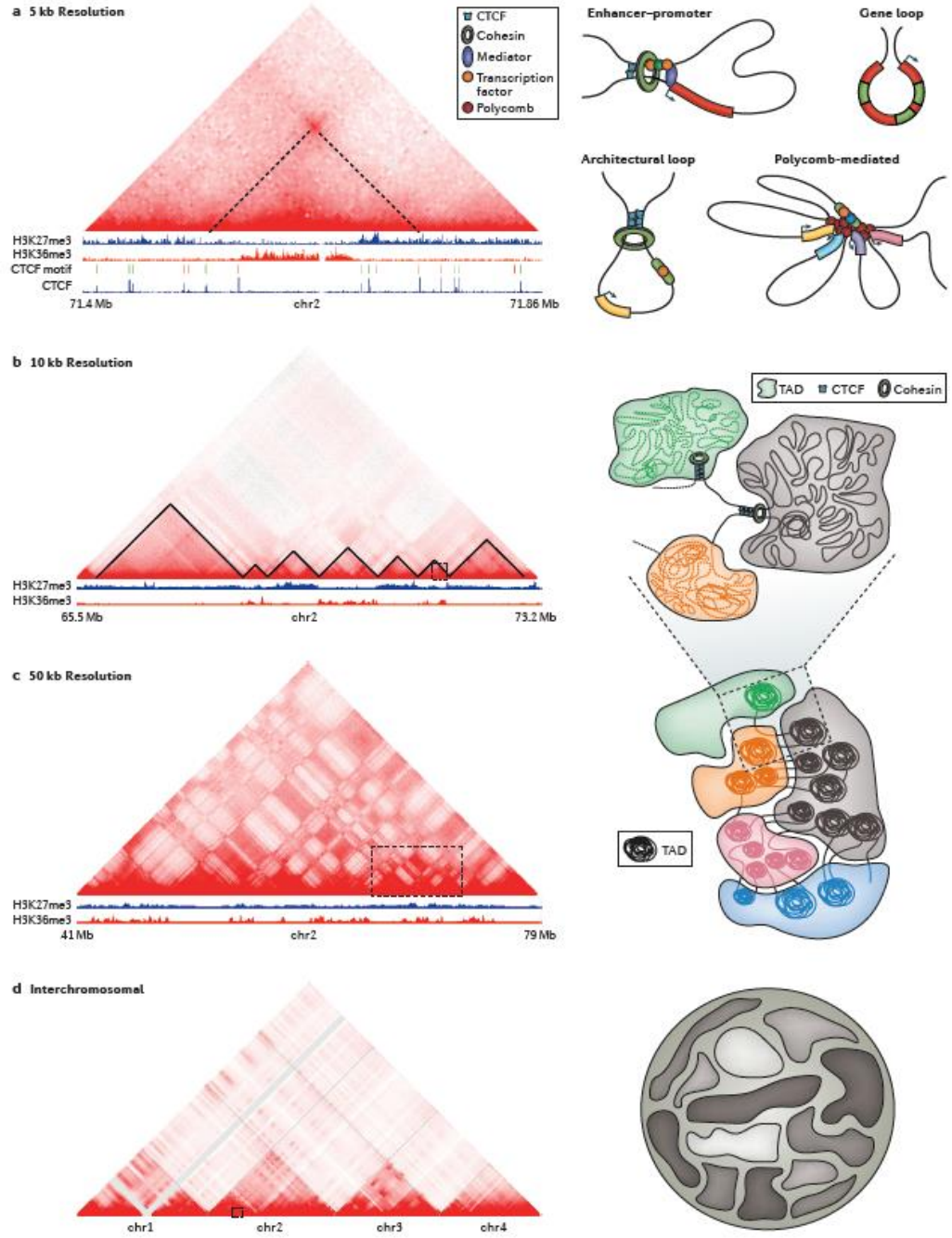


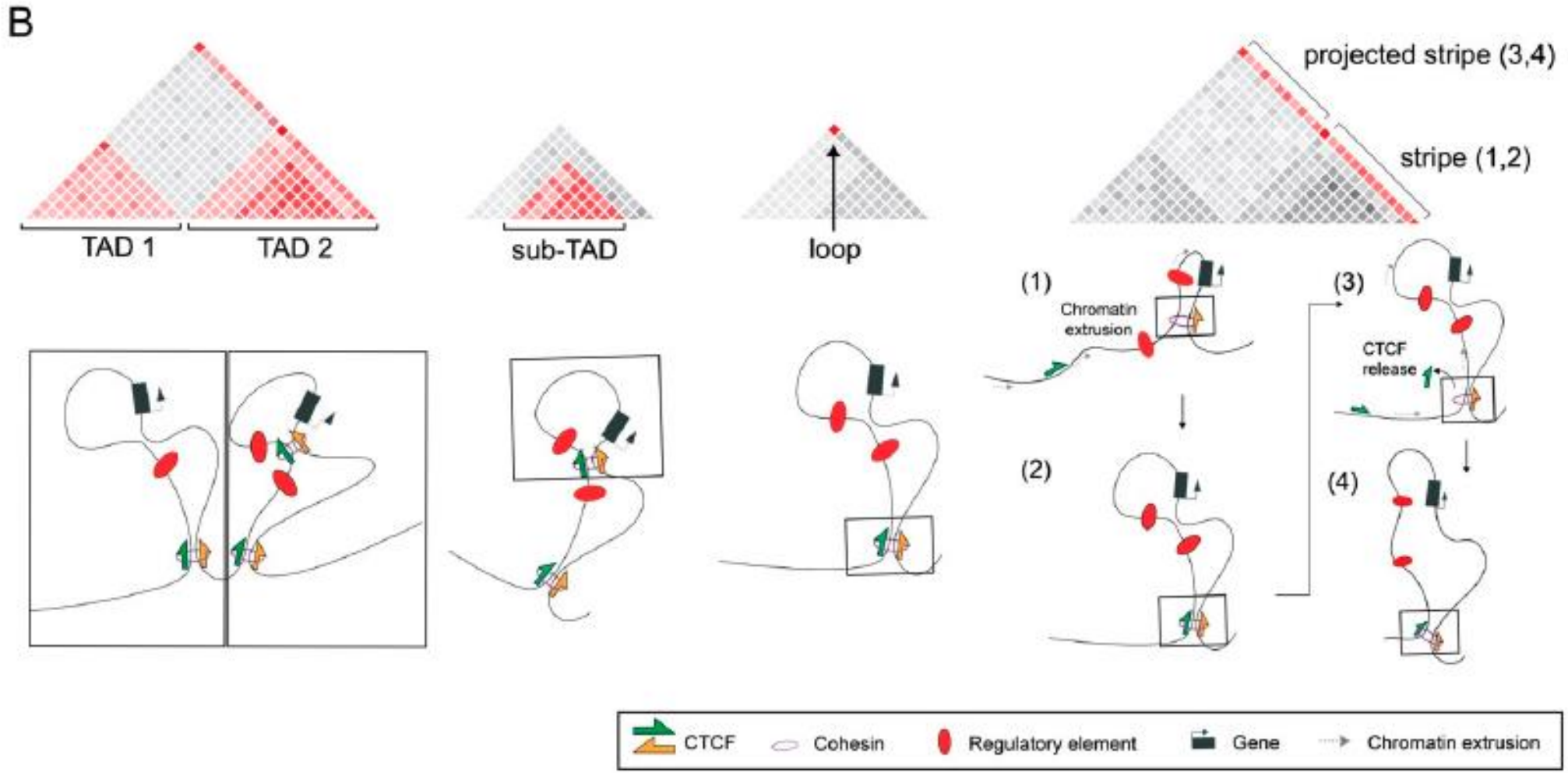




Annette Denker, and Wouter de Laat Genes Dev. 2016;30:1357-1382


Size
a
Multi-Mb large compartments
Interaction frequency
Eigenvector
Transcription
Broadly inactive B compartment
Broadly active A compartment
b
~1 Mb TADs
CTCF
c
Tens to hundreds of kb subTADs and CTCF loops
CTCF loop defining a TAD with two subTADs
CTCF
Cohesin




B
TAD 1
TAD 2
sub-TAD
loop
projected stripe (3,4)
stripe (1,2)
(1)
Chromatin extrusion
(2)
(3)
CTCF release
(4)
CTCF
Cohesin
Regulatory element
Gene
Chromatin extrusion

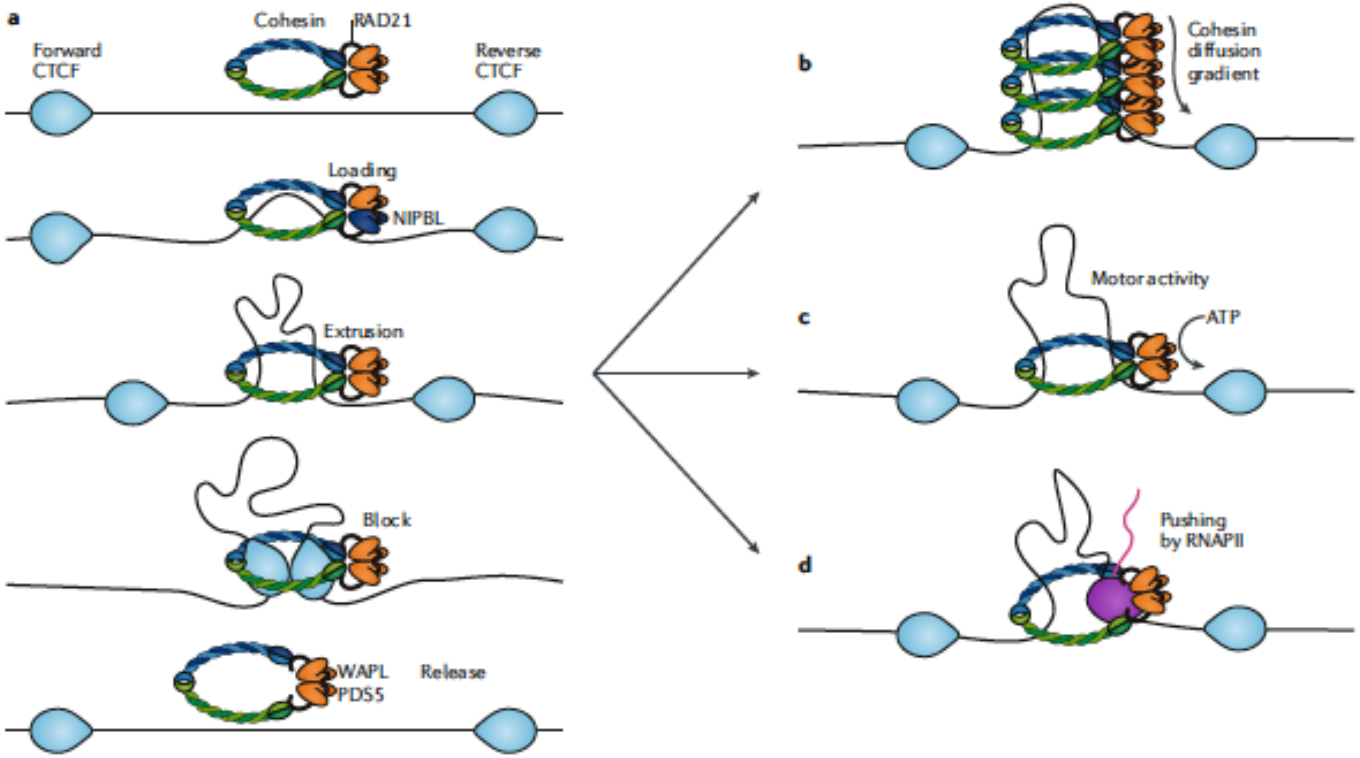


Fig. 3 | **Mechanisms of loop extrusion. a** | General model of loop extrusion. The extrusion process involves cohesin composed of structural maintenance of chromosomes (SMC) proteins SMC1 and SMC3 and RAD21; cohesin is loaded onto chromatin via NIPBL[94]. Extrusion is blocked at CTCF sites arranged in a convergent head-to-head orientation[45–49]. Some proportion of cohesin is released throughout this process by the activity of WAPL and PDS5 (REF.[94]). **b** | Extrusion via cohesin diffusion. Extrusion may occur by constant loading of cohesin resulting in a diffusion gradient[70]. **c** | Extrusion via cohesin motor activity. An alternative explanation for extrusion is that the process is driven by the motor activity of cohesin via ATP hydrolysis[54,72]. **d** | Extrusion via pushing of cohesin by RNA polymerase II (RNAPII). Other factors able to move along chromatin, such as RNAPII (purple), may help cohesin to extrude DNA[50,51,73–76].

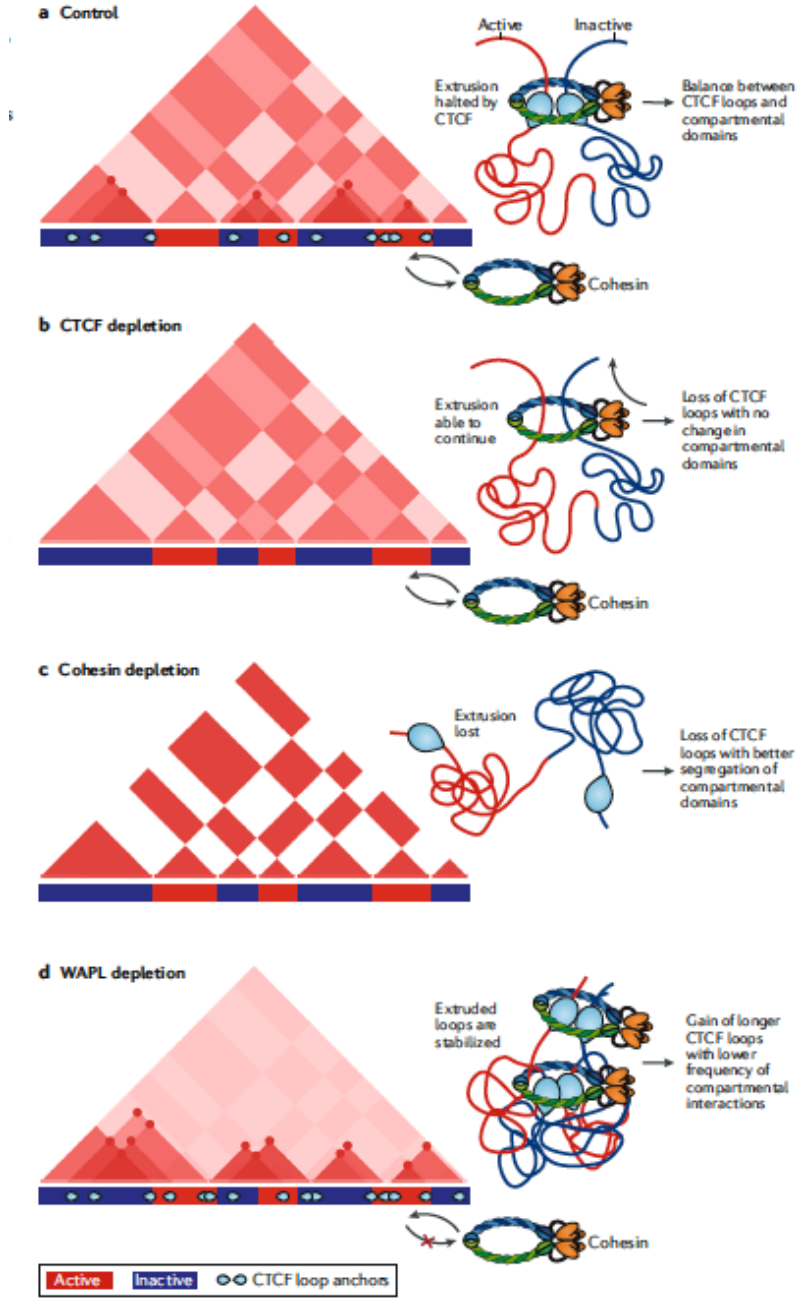
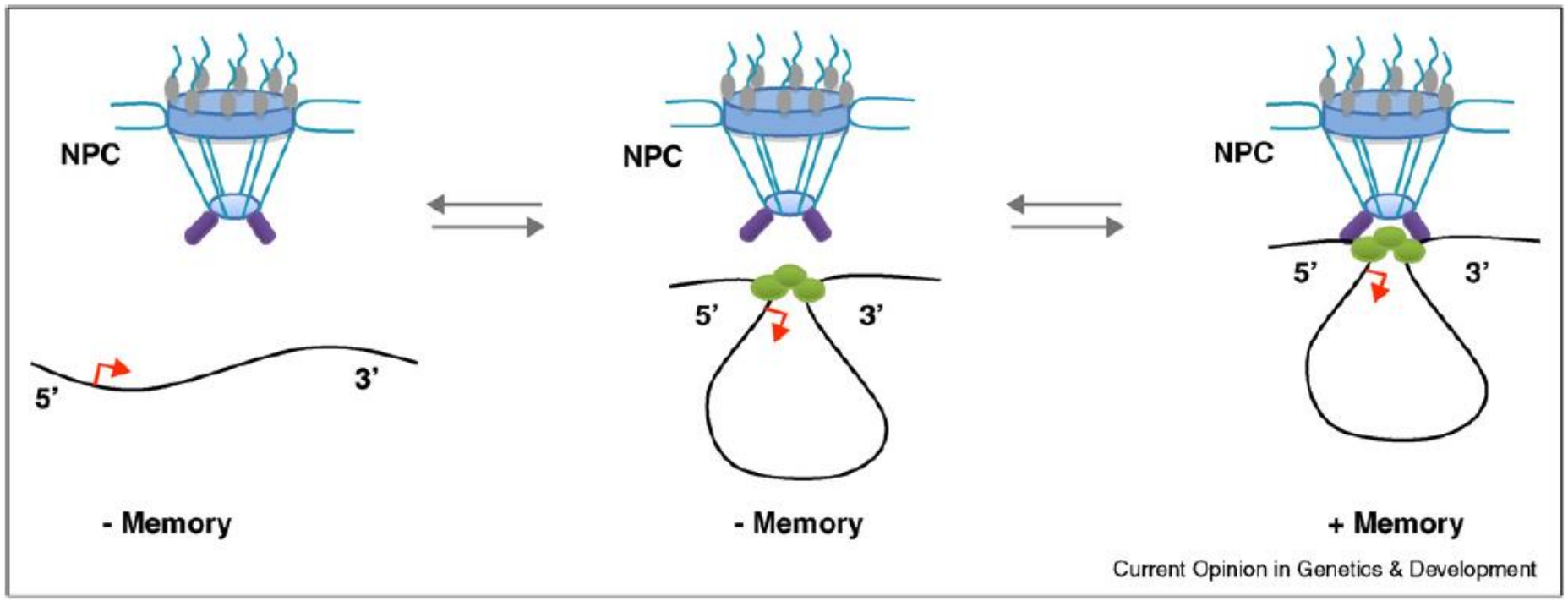
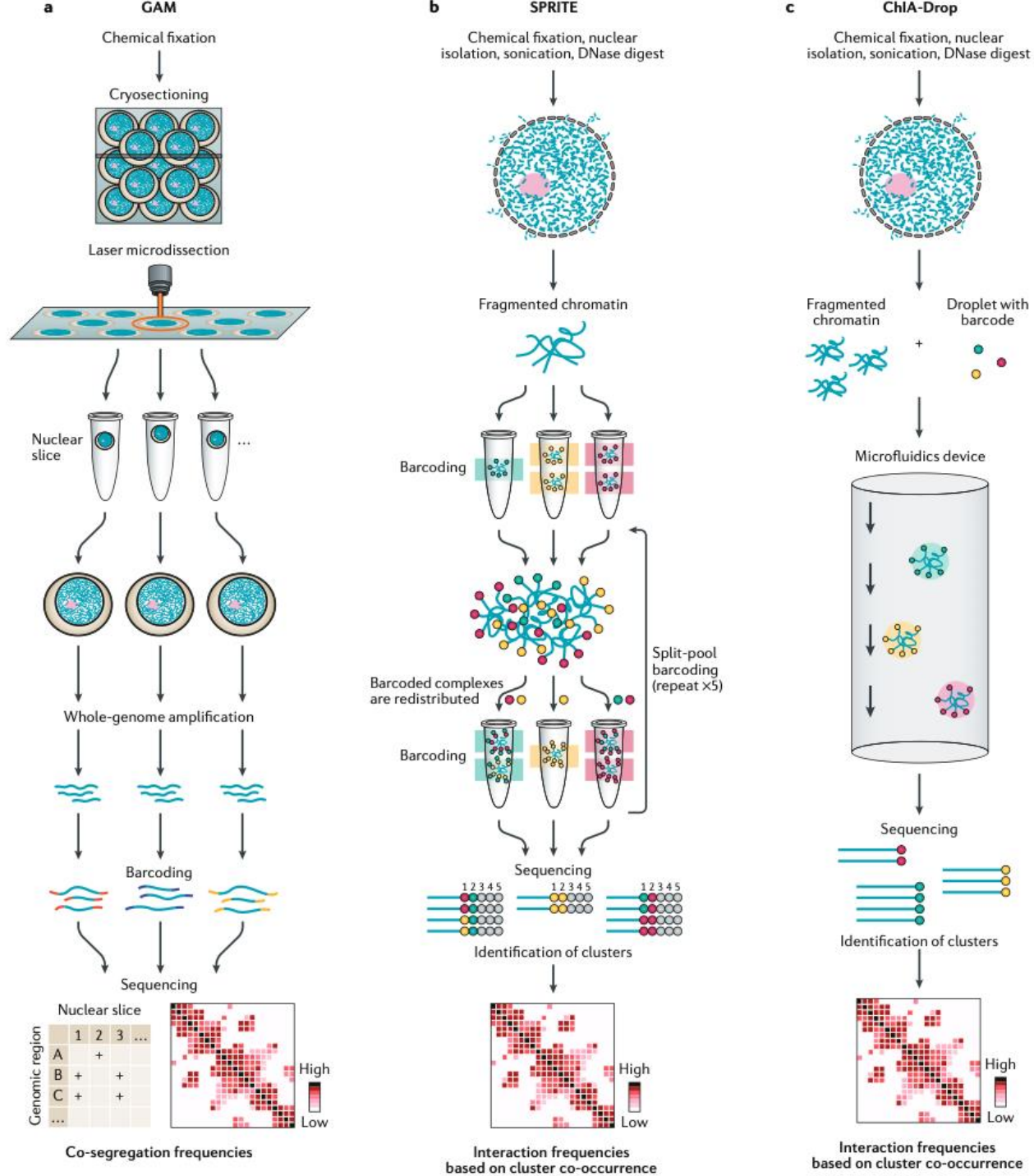


Fig. 4 | **Effects of CTCF, cohesin or WAPL depletion on 3D chromatin organization.** **a** | Chromatin is organized in the 3D nuclear space by CTCF loops and compartmental domains.

Memoria transcripcional y estructura de la cromatina.




a
GAM
Chemical fixation
Cryosectioning
Laser microdissection
Nuclear slice
...
Whole-genome amplification
Barcoding
Sequencing
Nuclear slice
Genomic region
1 2 3 ...
A + +
B + +
C + +
...
High
Low
Co-segregation frequencies
b
SPRITE
Chemical fixation, nuclear isolation, sonication, DNase digest
Fragmented chromatin
Barcoding
Barcoded complexes are redistributed
Barcoding
Split-pool barcoding (repeat ×5)
Sequencing
1 2 3 4 5
Identification of clusters
High
Low
Interaction frequencies based on cluster co-occurrence
c
ChIA-Drop
Chemical fixation, nuclear isolation, sonication, DNase digest
Fragmented chromatin
+
Droplet with barcode
Microfluidics device
Sequencing
Identification of clusters
High
Low
Interaction frequencies based on cluster co-occurrence

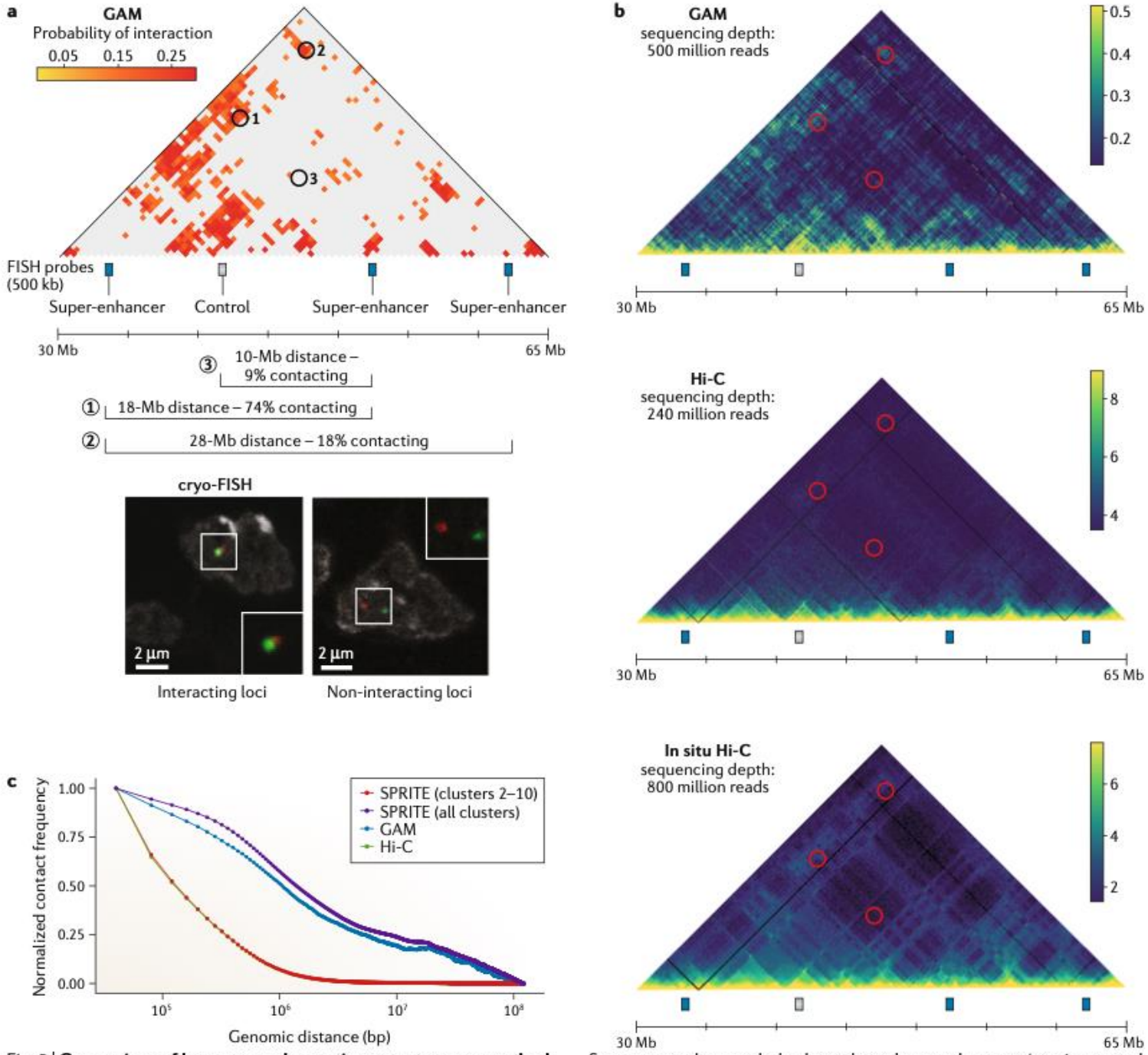
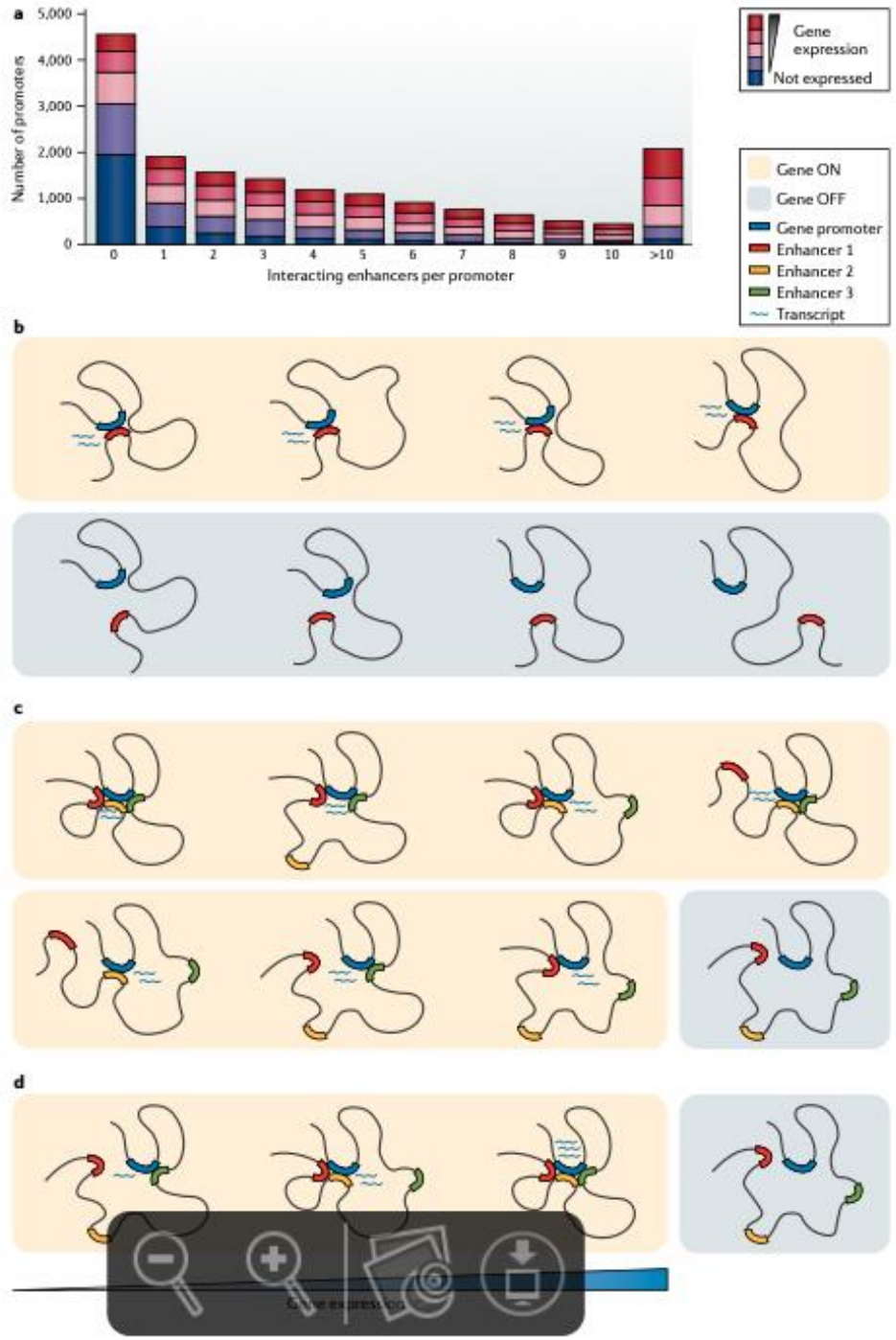


Fig. 5 | **Comparison of long-range chromatin contacts across methods.** Scores are colour-coded, where the colour-code range (maximum and



Pioneer transcription factors
A class of transcription factors that can bind to target sites in compacted ('closed') chromatin, which is covered by nucleosomes and is not DNase I hypersensitive. Pioneer transcription factors are thought to remodel the chromatin landscape during early steps of cell fate decisions to facilitate the

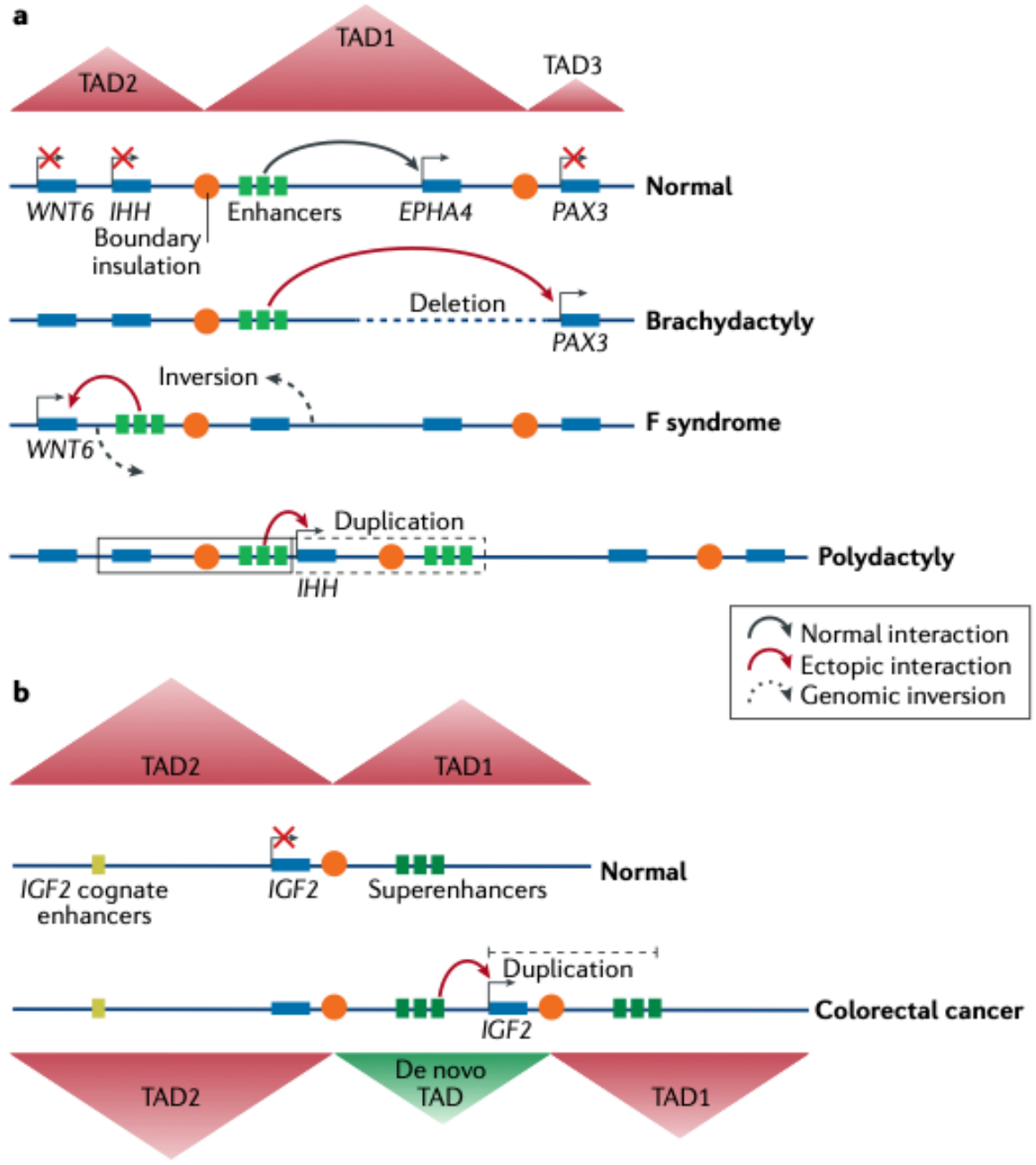
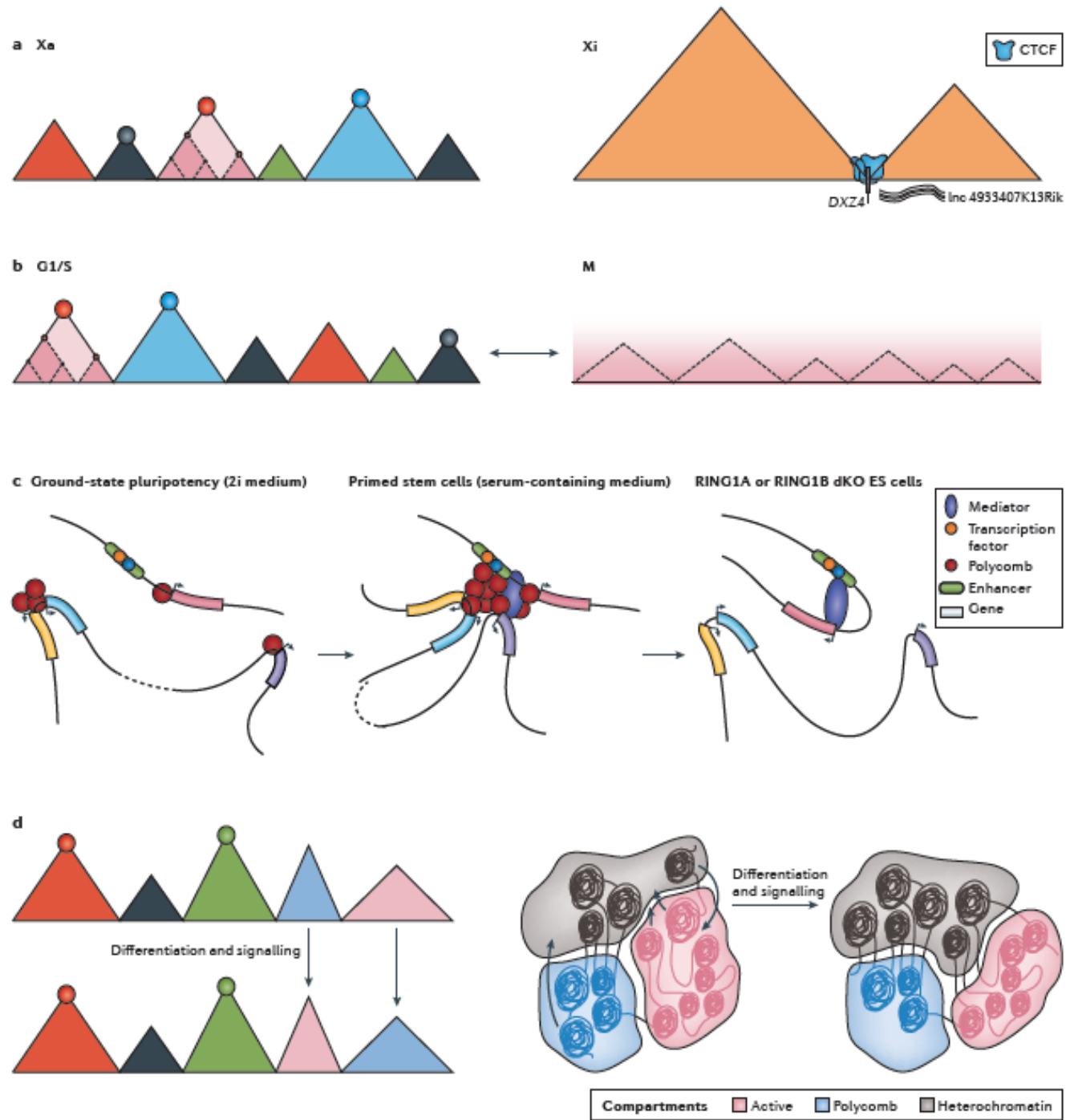


Fig. 5 | **Misregulation of the 3D genome in human diseases. a** | Three types of limb malformation that are caused by alterations in chromatin structure at the *EPHA4* locus. Normally, the interactions of enhancers in the topologically associating domain (TAD)


a Xa
Xi
CTCF
DXZ4
lnc 4933407K13Rik
b G1/S
M
c Ground-state pluripotency (2i medium)
Primed stem cells (serum-containing medium)
RING1A or RING1B dKO ES cells
Mediator
Transcription factor
Polycomb
Enhancer
Gene
d
Differentiation and signalling
Differentiation and signalling
Compartments
Active
Polycomb
Heterochromatin

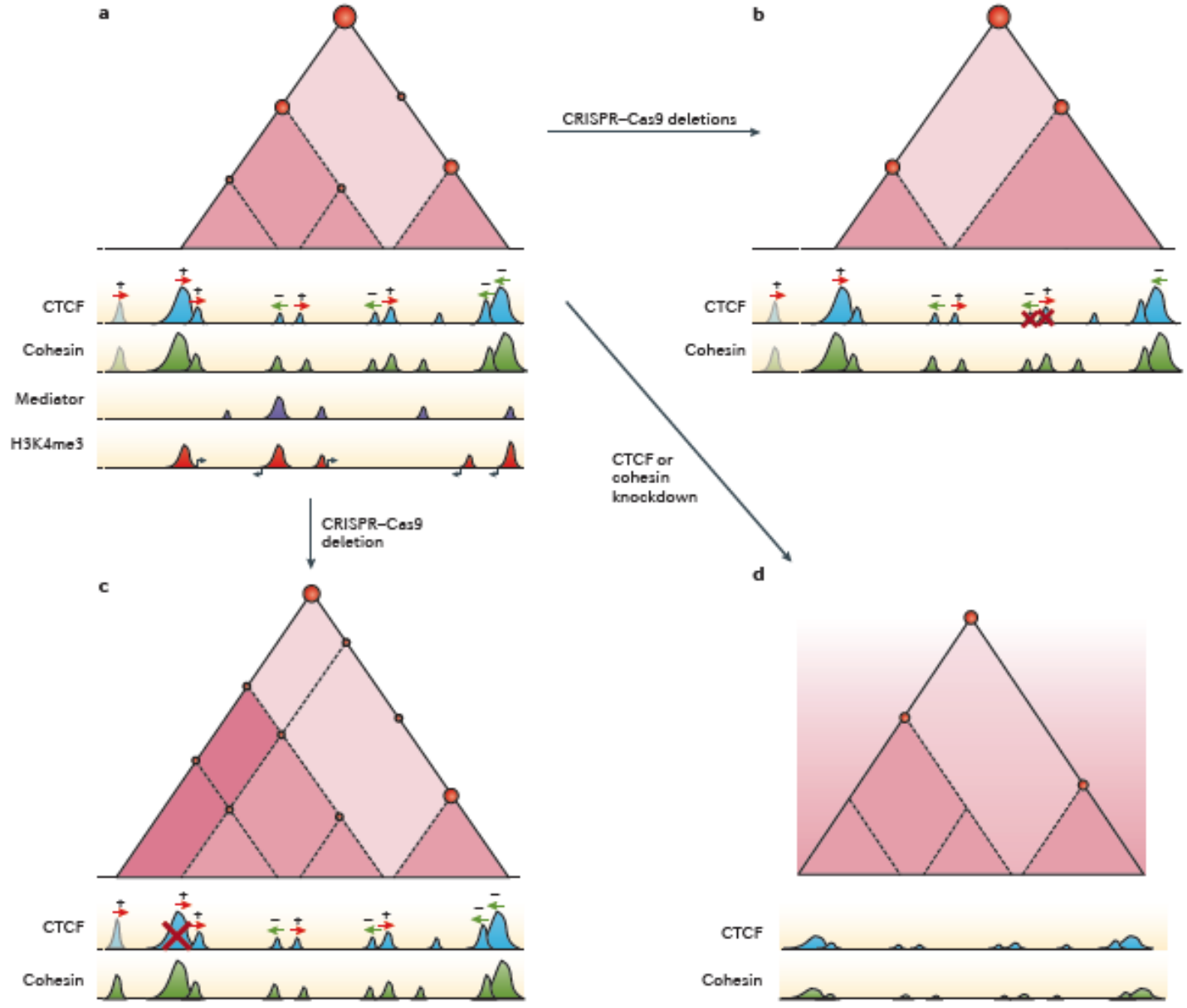


Figure 4 | **Importance of CTCF polarity on 3D chromatin organization.** **a** | Schematic representation of a typical contact domain, demarcated by a strong chromatin loop between the domain boundaries (red circle). Notice that several loops are also present within the topologically associating

contacts and the emergence of novel chromatin loops[18,72]. **c** | Deletion of a CTCF loop located at the boundary of a TAD leads to an expansion of the domain to the closest upstream CTCF-binding site with a motif in the same

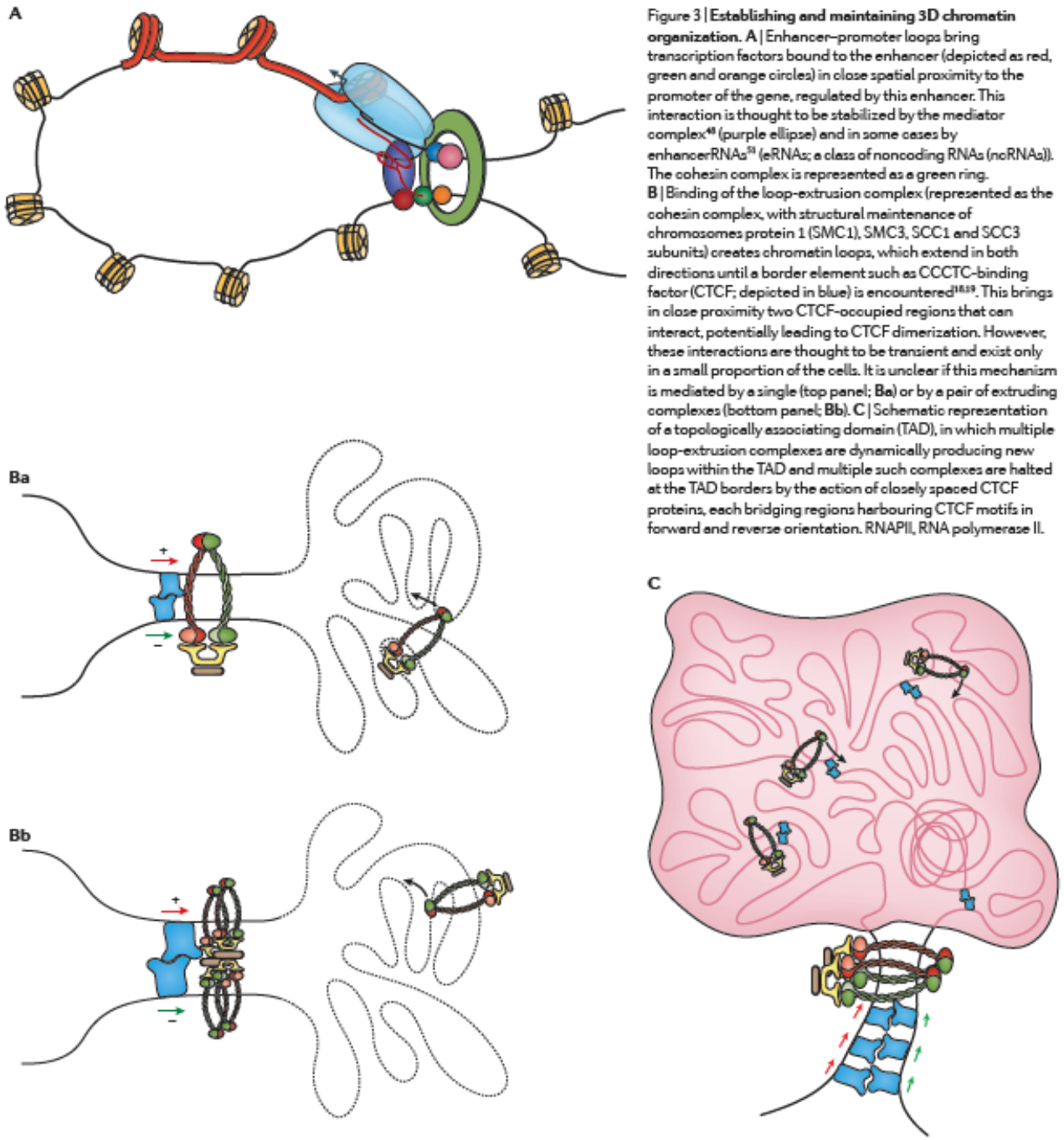


Figure 3 | **Establishing and maintaining 3D chromatin organization. A** | Enhancer–promoter loops bring transcription factors bound to the enhancer (depicted as red, green and orange circles) in close spatial proximity to the promoter of the gene, regulated by this enhancer. This interaction is thought to be stabilized by the mediator complex[48] (purple ellipse) and in some cases by enhancerRNAs[51] (eRNAs; a class of noncoding RNAs (ncRNAs)). The cohesin complex is represented as a green ring. **B** | Binding of the loop-extrusion complex (represented as the cohesin complex, with structural maintenance of chromosomes protein 1 (SMC1), SMC3, SCC1 and SCC3 subunits) creates chromatin loops, which extend in both directions until a border element such as CCCTC-binding factor (CTCF; depicted in blue) is encountered[18,19]. This brings in close proximity two CTCF-occupied regions that can interact, potentially leading to CTCF dimerization. However, these interactions are thought to be transient and exist only in a small proportion of the cells. It is unclear if this mechanism is mediated by a single (top panel; **Ba**) or by a pair of extruding complexes (bottom panel; **Bb**). **C** | Schematic representation of a topologically associating domain (TAD), in which multiple loop-extrusion complexes are dynamically producing new loops within the TAD and multiple such complexes are halted at the TAD borders by the action of closely spaced CTCF proteins, each bridging regions harbouring CTCF motifs in forward and reverse orientation. RNAPII, RNA polymerase II.

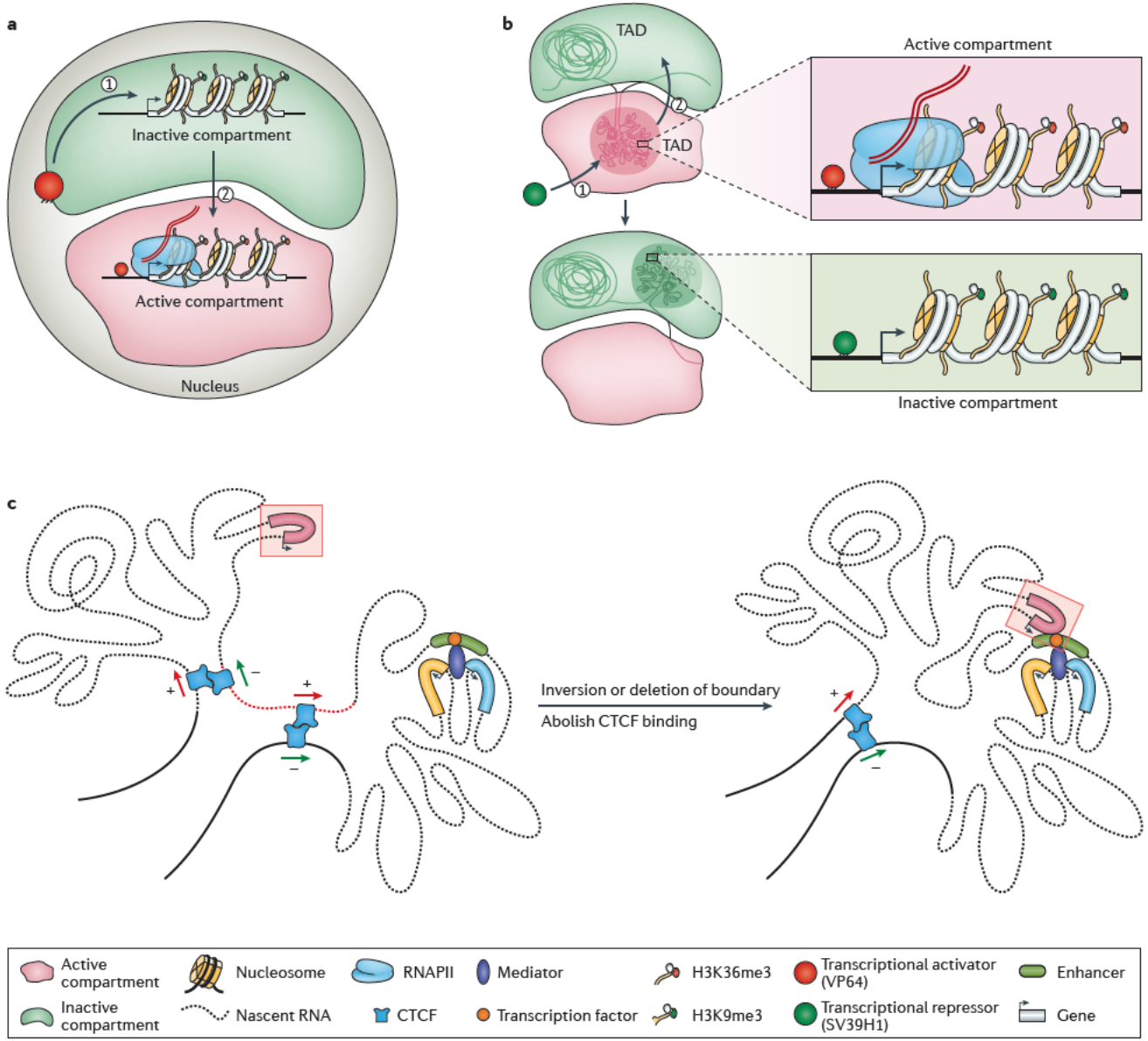


Figure 6 | **3D genome organization and gene expression. a** | Artificial recruitment of a transcriptional activator (such as VP64, depicted by the red circle) or chromatin decondensation alone is sufficient to reposition a locus located normally in

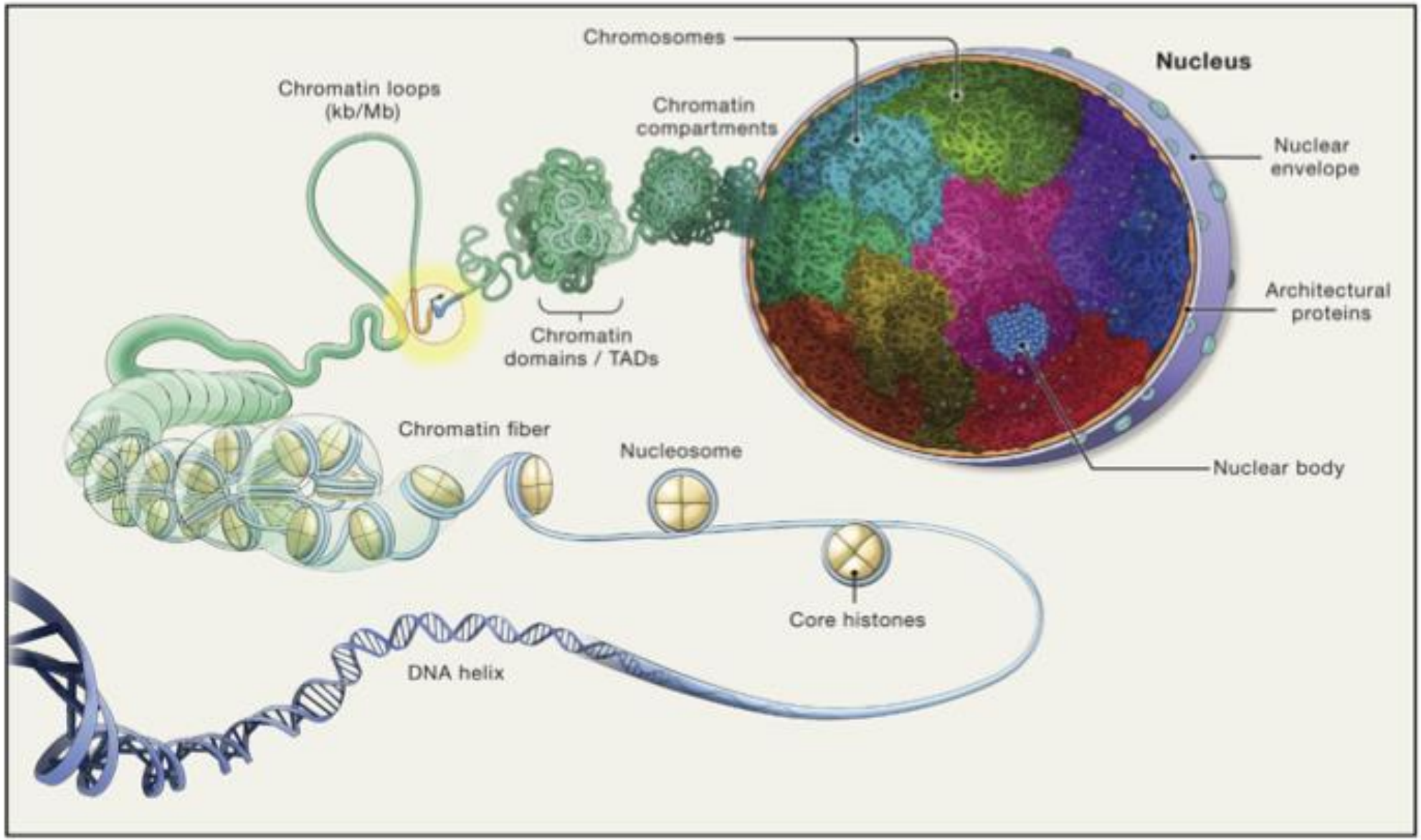
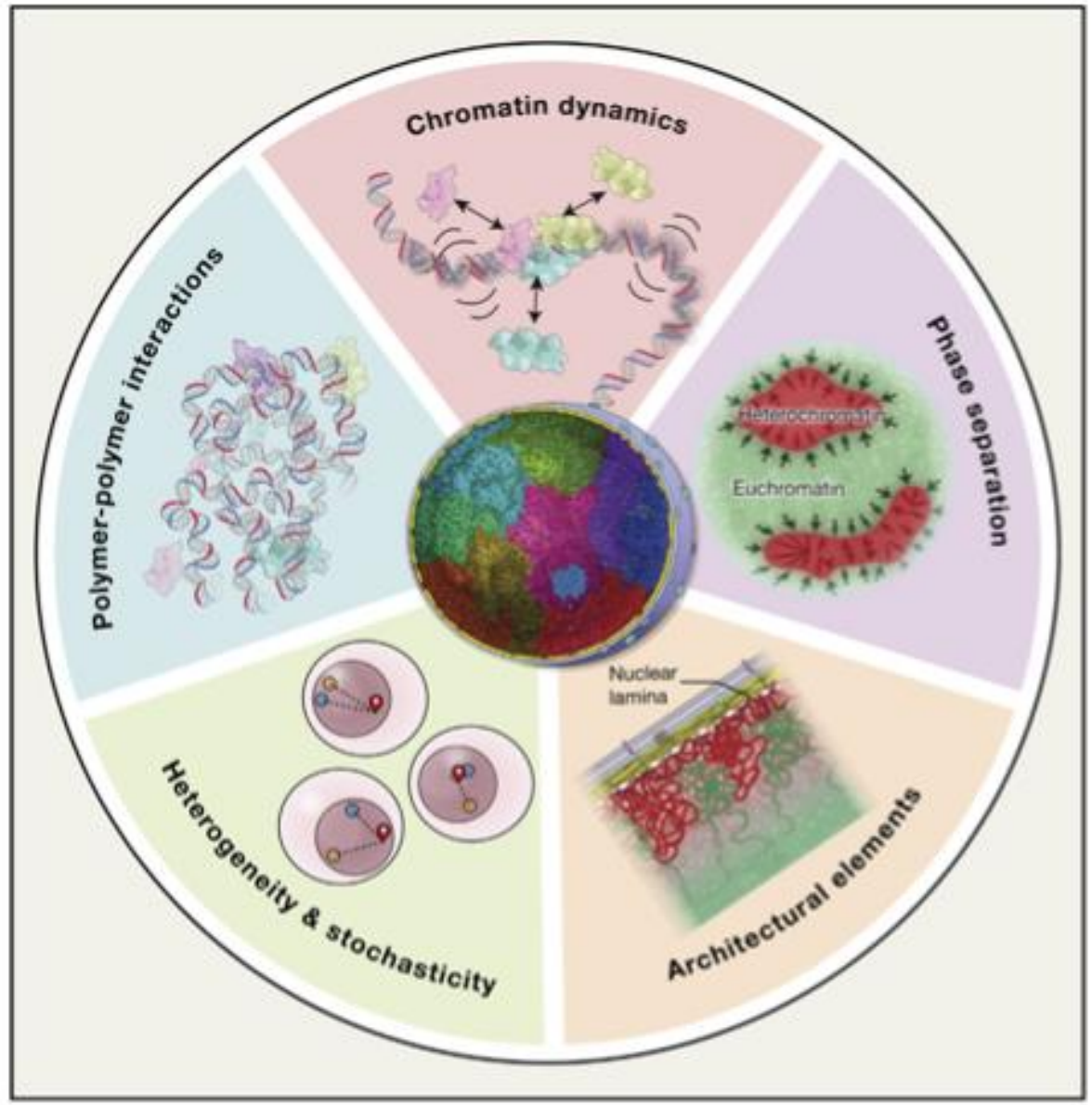
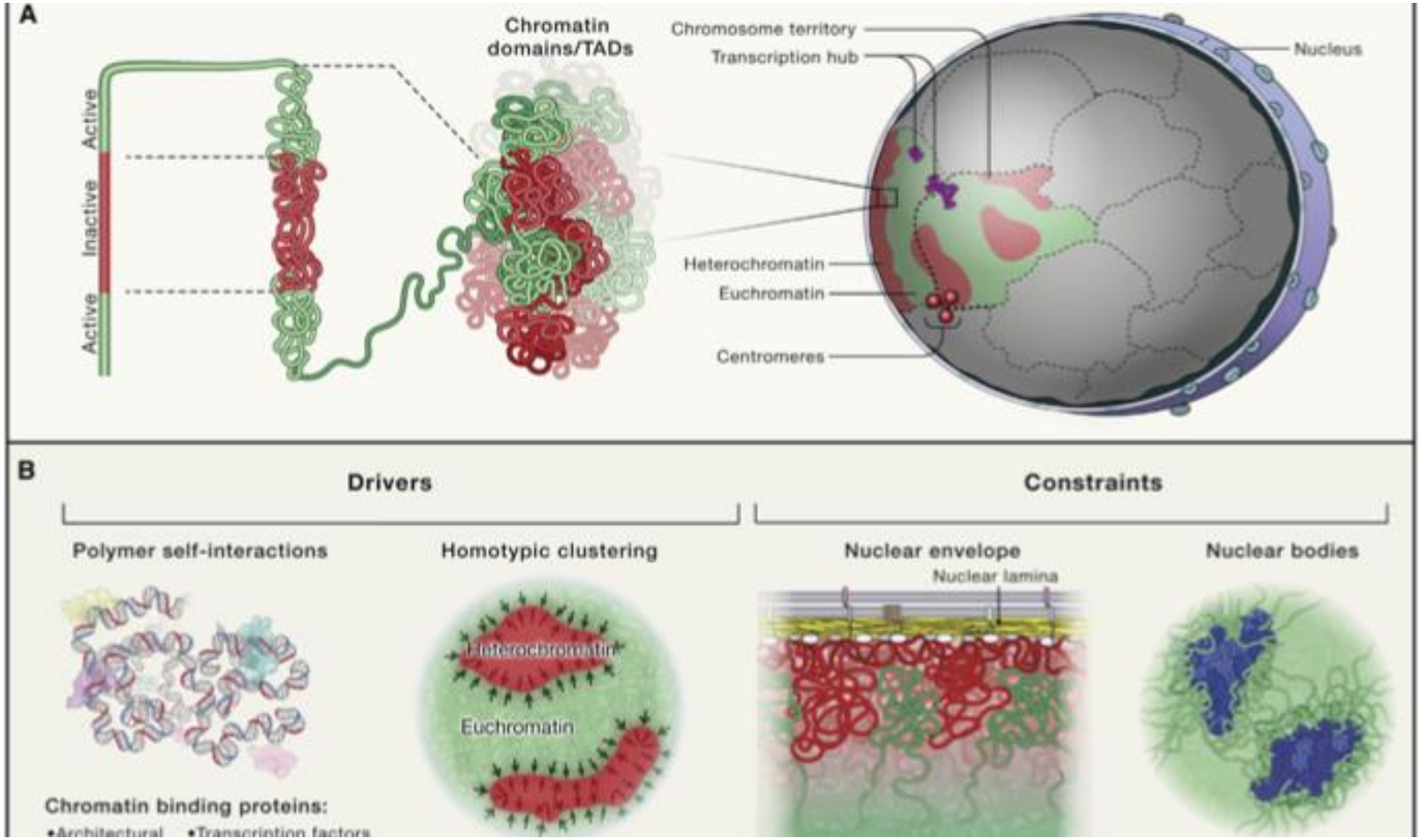


Figure 1. The Organization of the Eukaryotic Genome


Chromatin dynamics
Phase separation
Heterochromatin
Euchromatin
Architectural elements
Nuclear lamina
Heterogeneity & stochasticity
Polymer-polymer interactions


A
Chromatin domains/TADs
Active
Inactive
Active
Chromosome territory
Transcription hub
Nucleus
Heterochromatin
Euchromatin
Centromeres
B
Drivers
Constraints
Polymer self-interactions
Homotypic clustering
Nuclear envelope
Nuclear bodies
Nuclear lamina
Heterochromatin
Euchromatin
Chromatin binding proteins:
•Architectural •Transcription factors

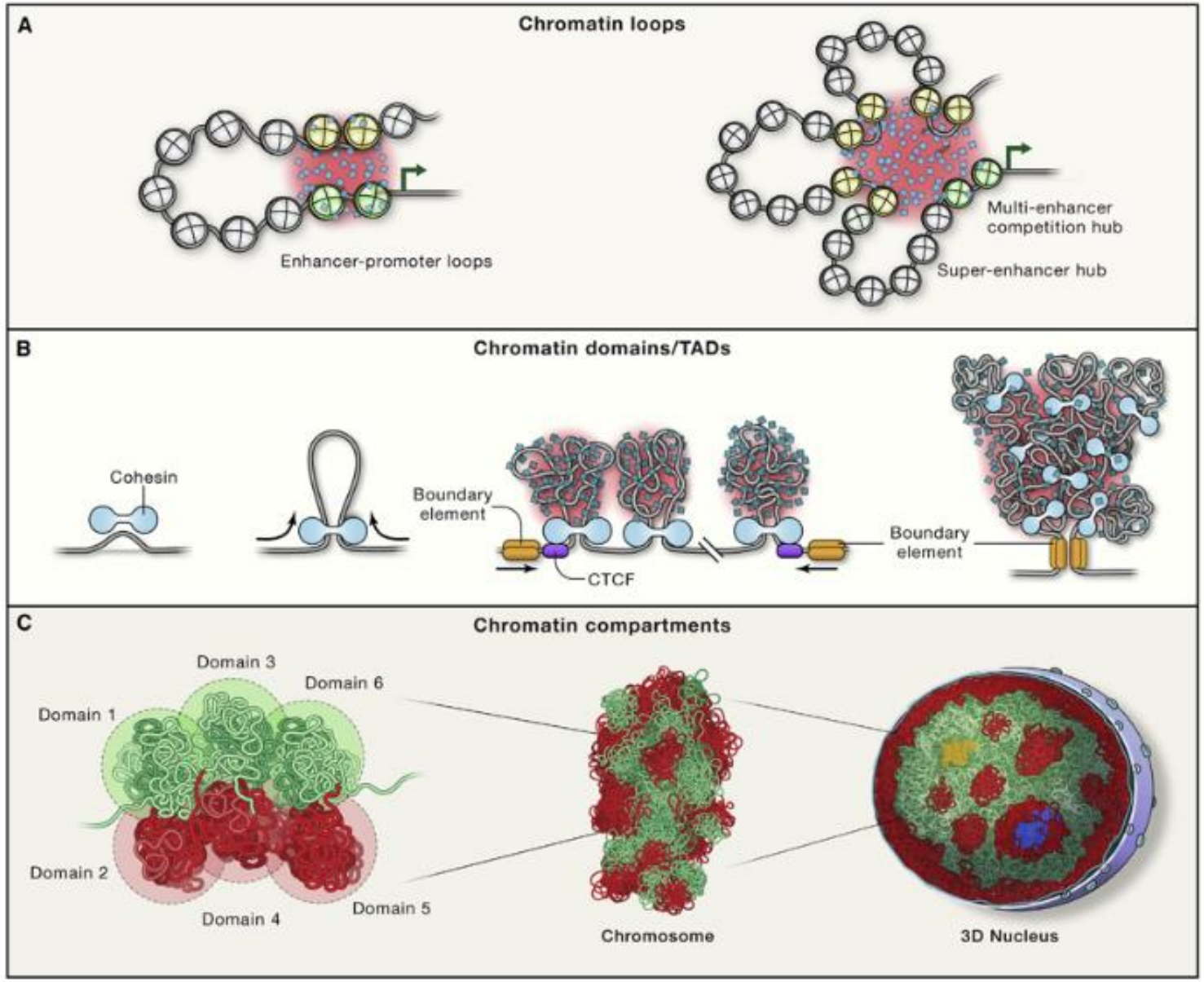


Figure 4. Formation of Chromatin Features by Self-Organization

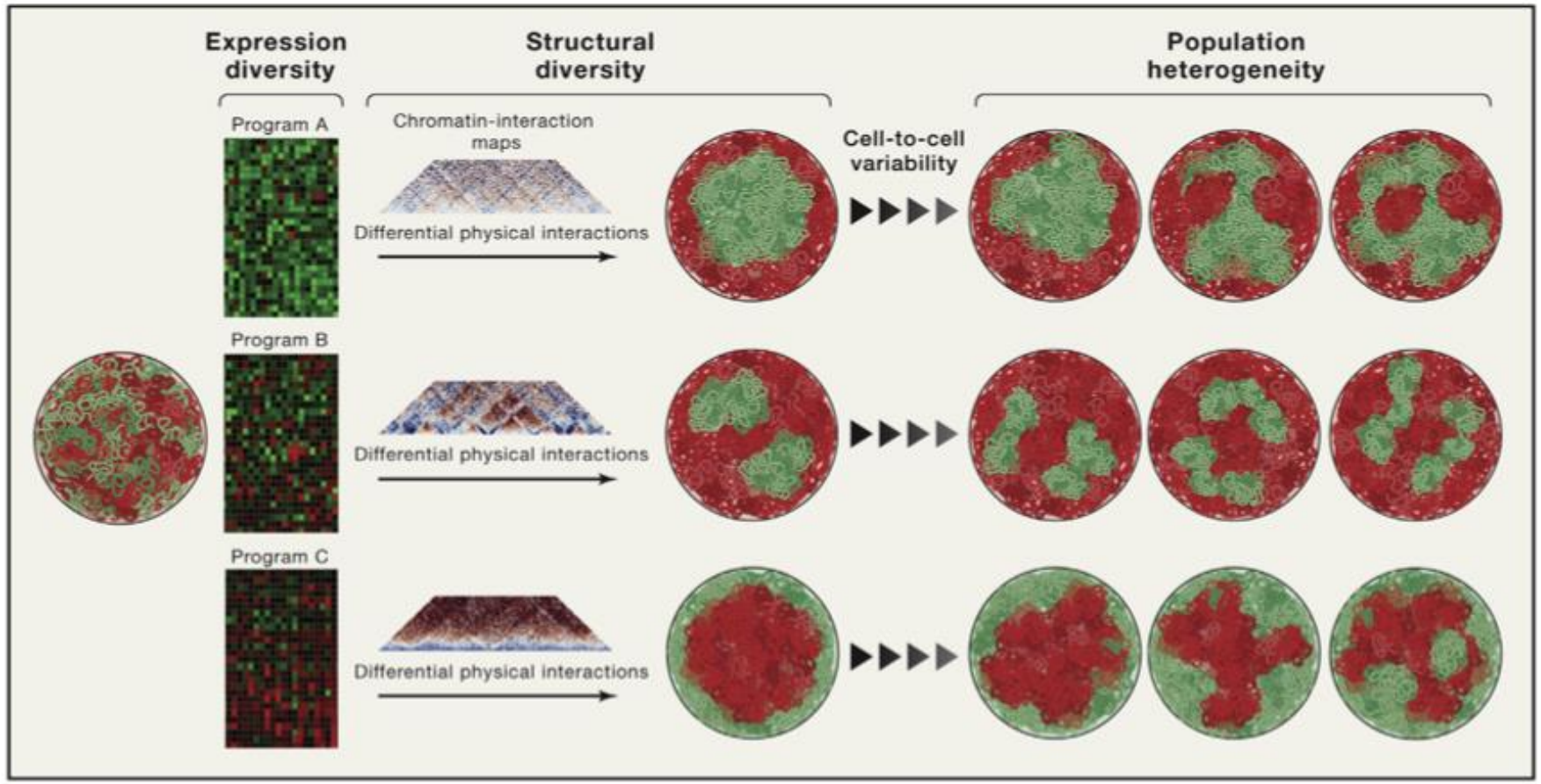


Figure 5. Genome Function Drives Structure

All cells of an organism contain genomes of identical sequence (left single nucleus), but different cell types express distinct gene expression programs (heatmaps) and assume different genome topologies as detected by chromatin interaction maps. The cell-type-specific homotypic chromatin-chromatin interactions drive higher-order genome organization and generate distinct overall genome topologies in different cell types, resulting in cell-type-specific patterns of euchromatin (green) and heterochromatin (red). (right) Within a cell type, the heterogeneity of chromatin-chromatin interactions generates cell-to-cell variability in the population.

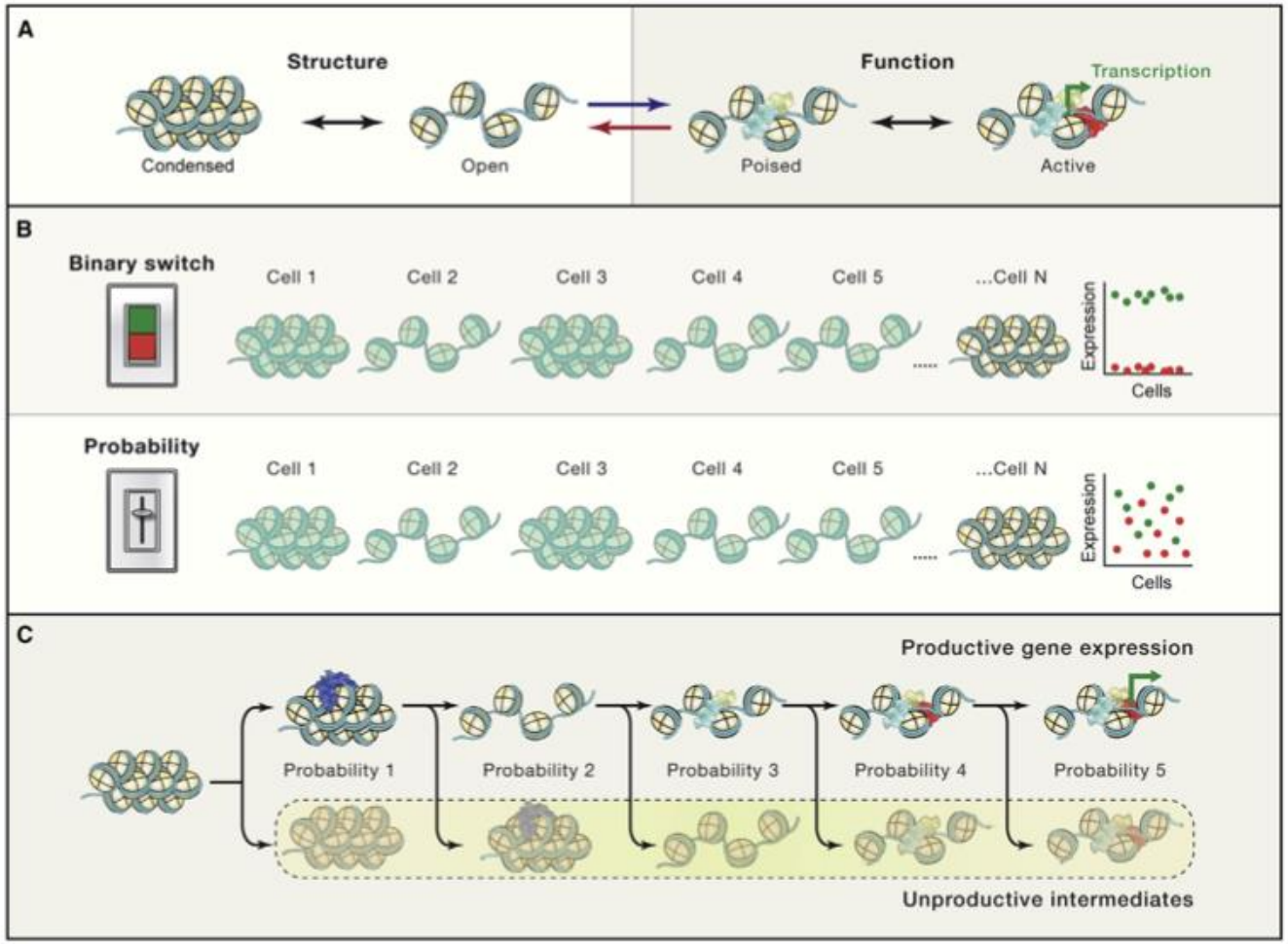


Figure 6. The Probabilistic Genome

(A) Bi-directional interplay of chromatin structure (light shaded area, left) and function (dark shaded area, right). Chromatin structure affects gene function (blue forward arrow). Chromatin structure reversibly oscillates between a condensed and open state, which facilitates association of transcription factors (blue, yellow), leading to a poised state and upon association of RNA polymerase (red) enables transcription (green bent arrow). Conversely, transcription affects structure (red reverse arrow), by maintaining an open chromatin structure.

(B) Chromatin structure does not act as a deterministic, binary switch, but rather as a probabilistic modulator of function.

In a binary switch model, the expression level of a given gene in individual cells is either fully on (green points) or fully off (red points) reflecting the open and closed configuration of the gene locus in single cells.

In a modulatory model, gene expression levels are heterogeneous (red, green points) in individual cells reflecting the probabilistic nature of gene activity in open and closed chromatin as typically observed experimentally.

(C) Gene expression is a multi-step processes and is inherently probabilistic. Each step required to activate a gene represents an equilibrium of a productive event toward gene activation versus a reverse, unproductive event with a certain probability. Probability 1: stable association of a chromatin remodeling factor (blue) versus transient interaction as they diffuse through the nucleus. Probability 2: maintenance of decondensed chromatin versus reversion to condensed state. Probability 3: association of early transcription factors (light blue, yellow) which promote the likelihood of association of additional transcription factors versus disassociation of early transcription factors. Probability 4: association of RNA polymerase (red) versus loss of activating transcription factors. Probability 5: transcriptional activation (bent green arrow) versus unproductive dissociation of RNA polymerase. As a result of the probabilistic nature of each step, the activation process as a whole is relatively inefficient, thus generating the stochastic patterns of activation observed for most genes.