

Gene looping and 3D chromatin architecture

Bio 5488

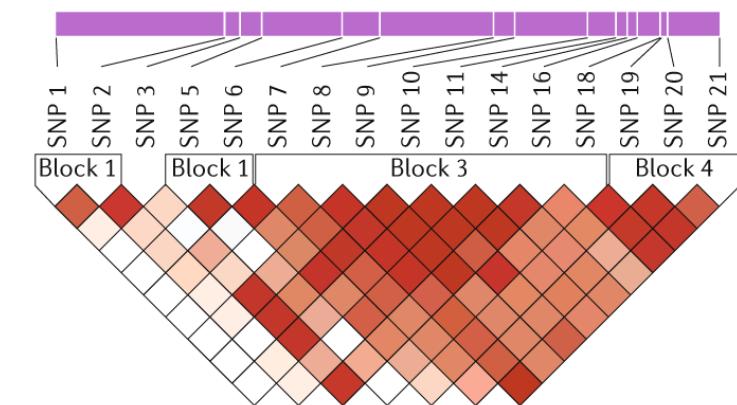
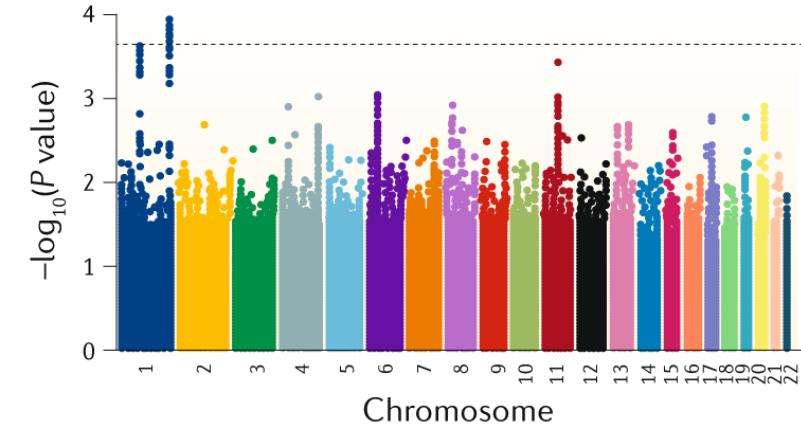
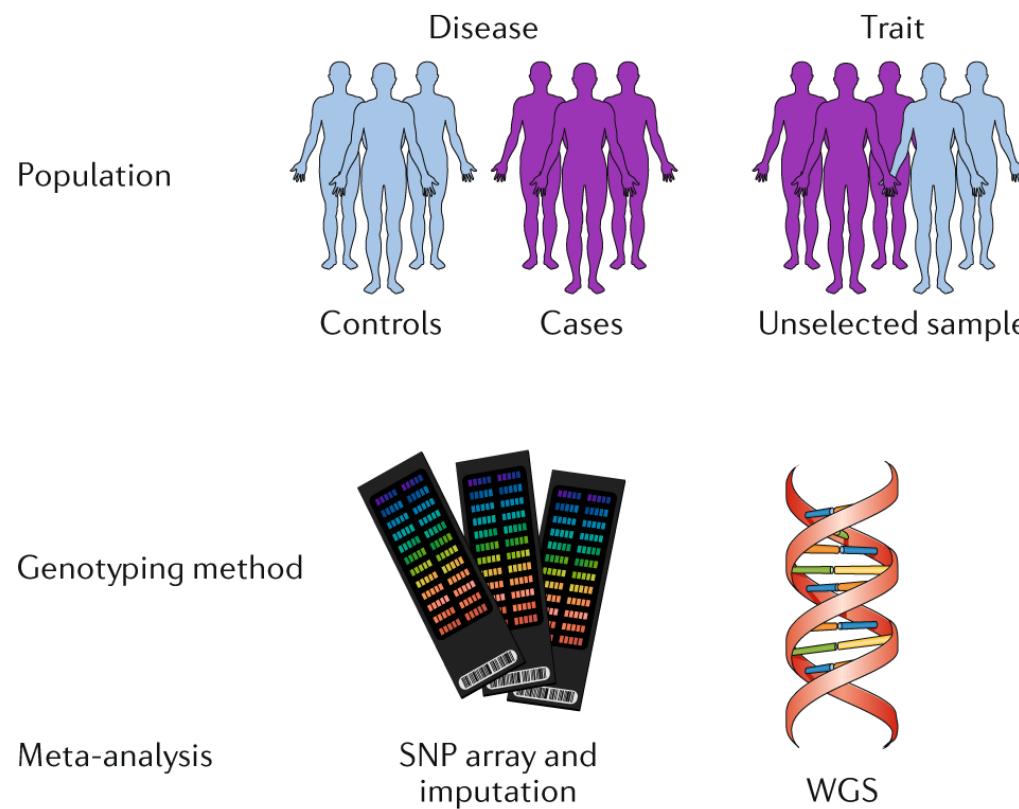
Michael Meers

2/11/2026

Outline

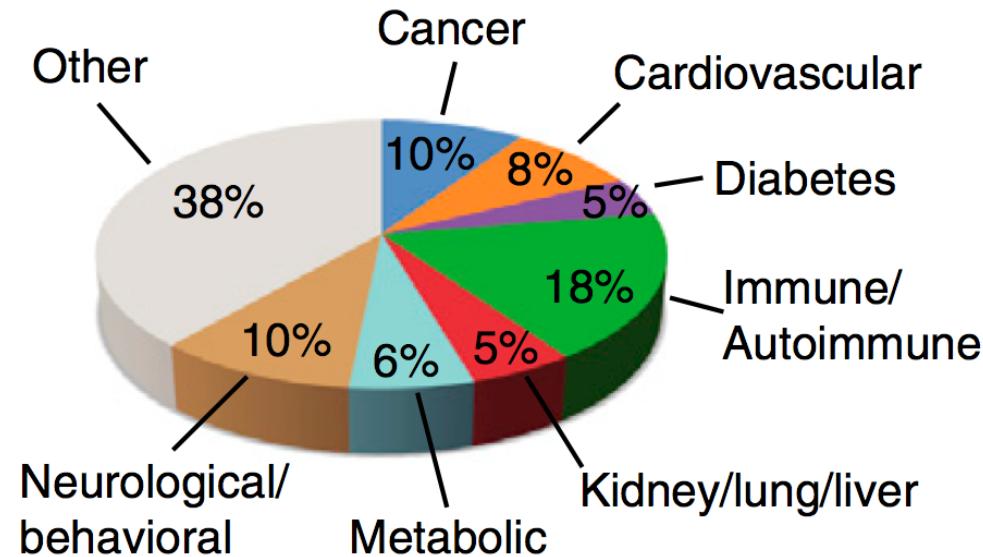
- Motivation for studying 3D genome
- Method to investigate 3D genome organization
- 3D genome organization and human diseases
- 4DN project (<https://www.4dnucleome.org>)

Identify disease or trait-associated variants by genome-wide association studies (GWAS)

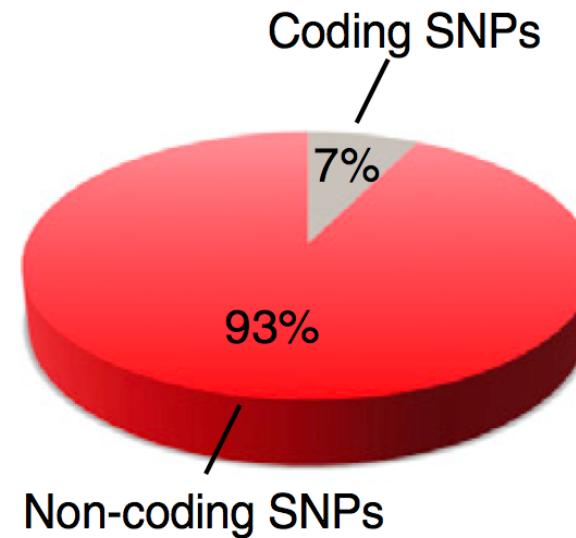


Majority of the GWAS hits are located in the non-coding regions

5,303 SNPs from 1,675 GWAS studies

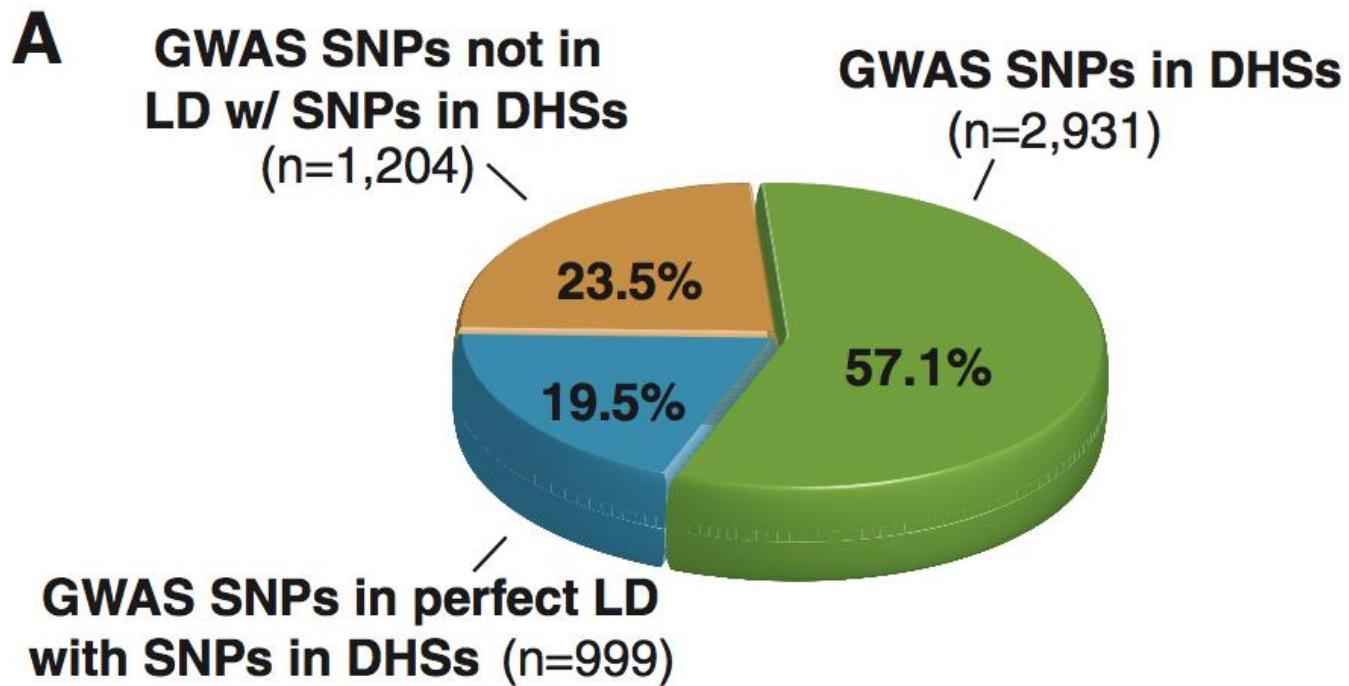


Coding vs. non-coding SNPs



Hnisz et al, Cell, 2013

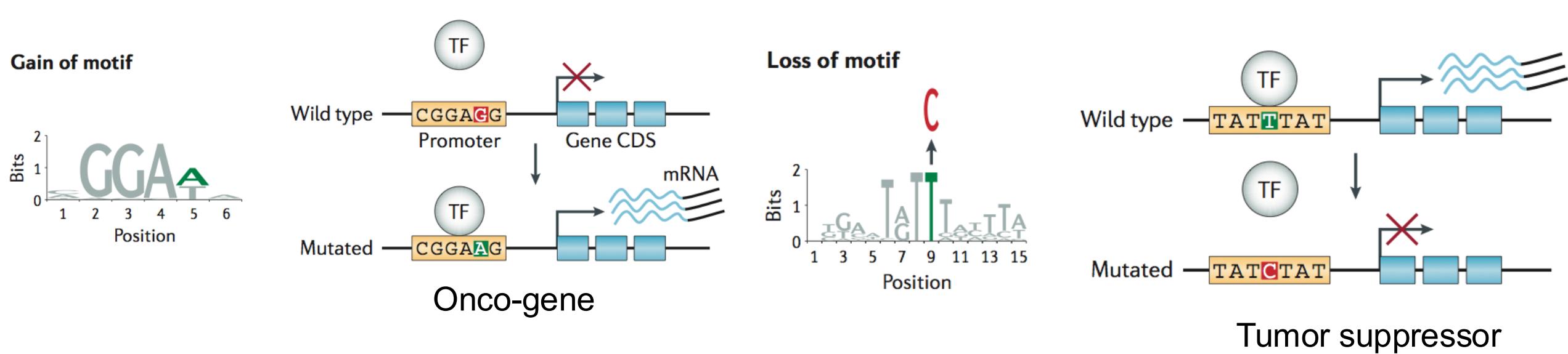
Majority of the GWAS hits are located in the cis-regulatory elements



Maurano et al Science 2012

How non-coding mutations contribute to disease

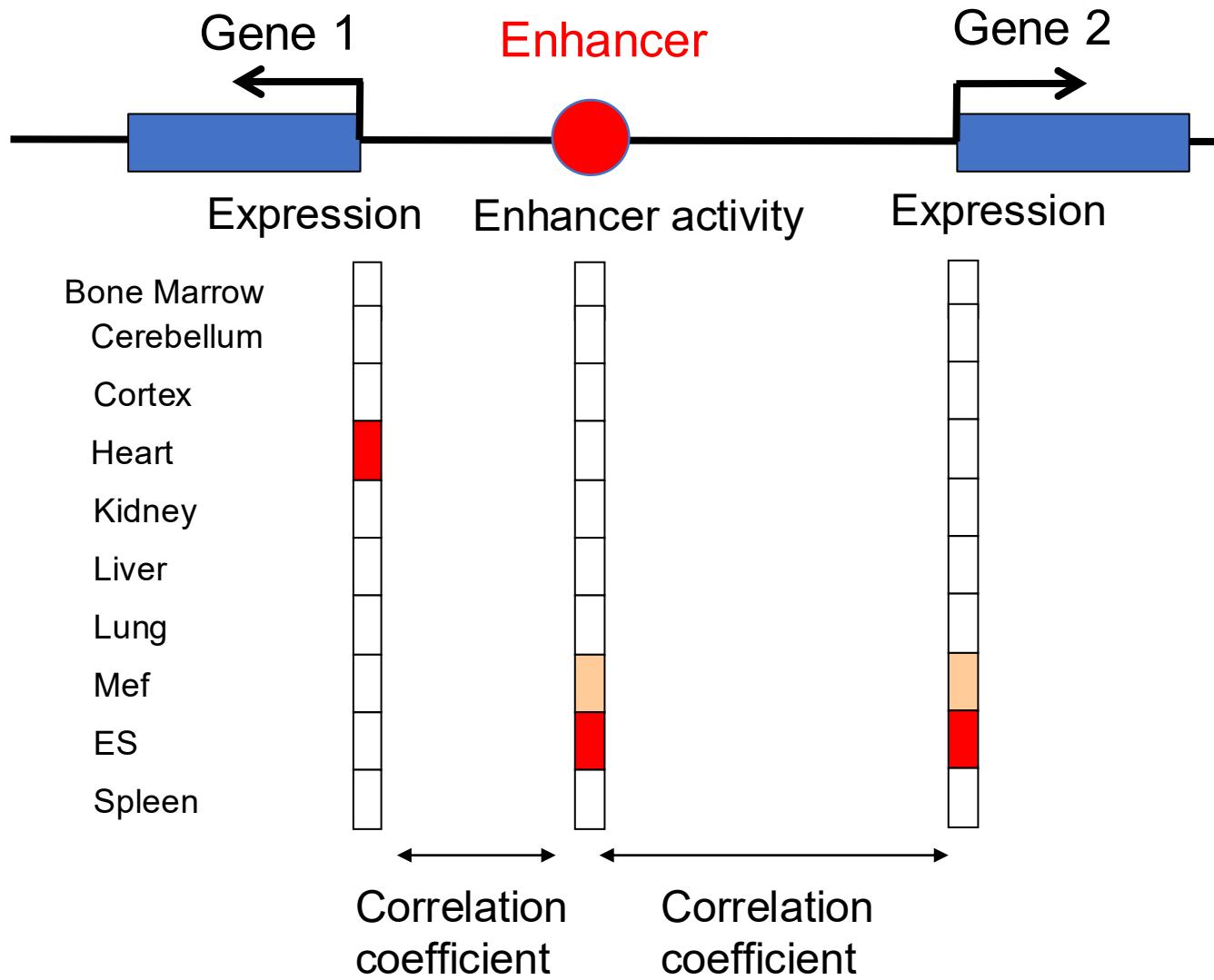
- Gain of TF binding sites;
- Loss of TF binding sites;
- Enhancer hijacking



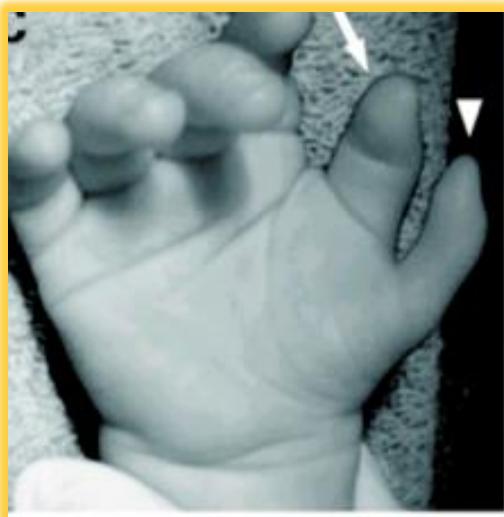
How do we find the target genes for distal enhancers?

- Nearest genes by genomic distance;
- Correlation based on gene expression and enhancer activities;
- 3D space proximity;
- CRISPR/Cas9 to delete enhancers and observe their effect on gene expression;

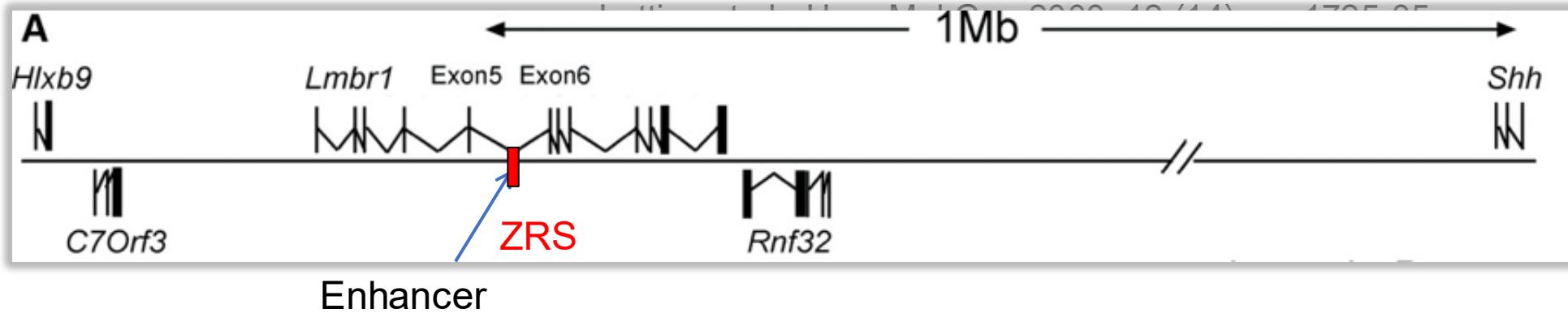
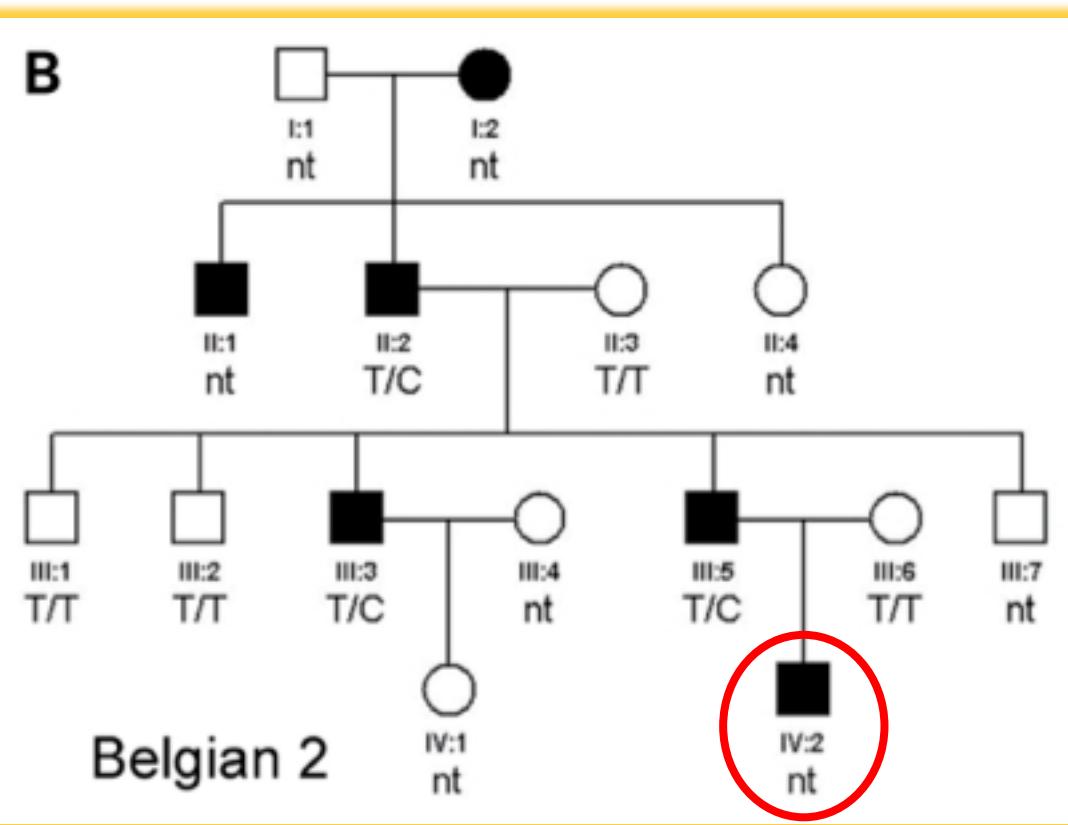
Linking enhancers to target promoters



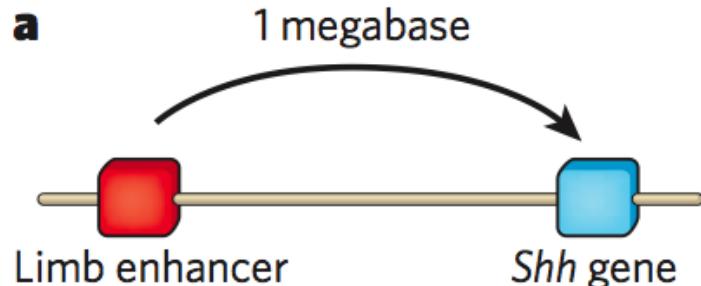
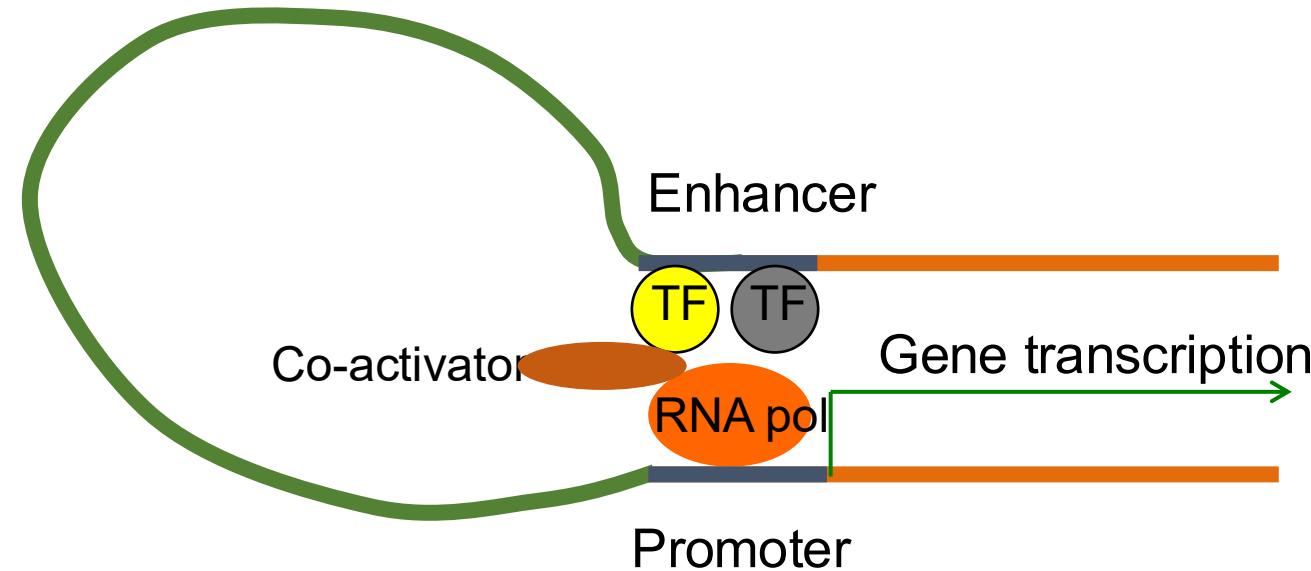
Regulation at a distance



Preaxial polydactyly



3D genome and transcriptional regulation



2005

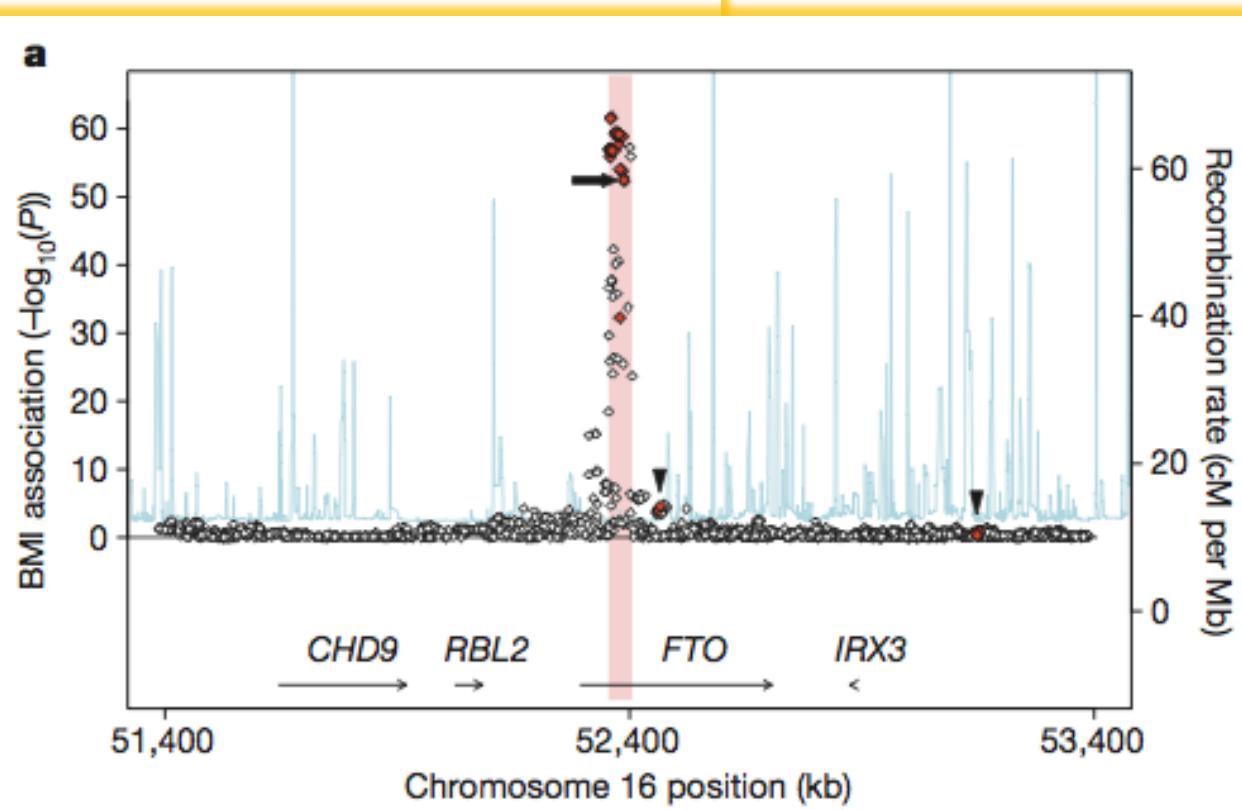
Regulation at a distance

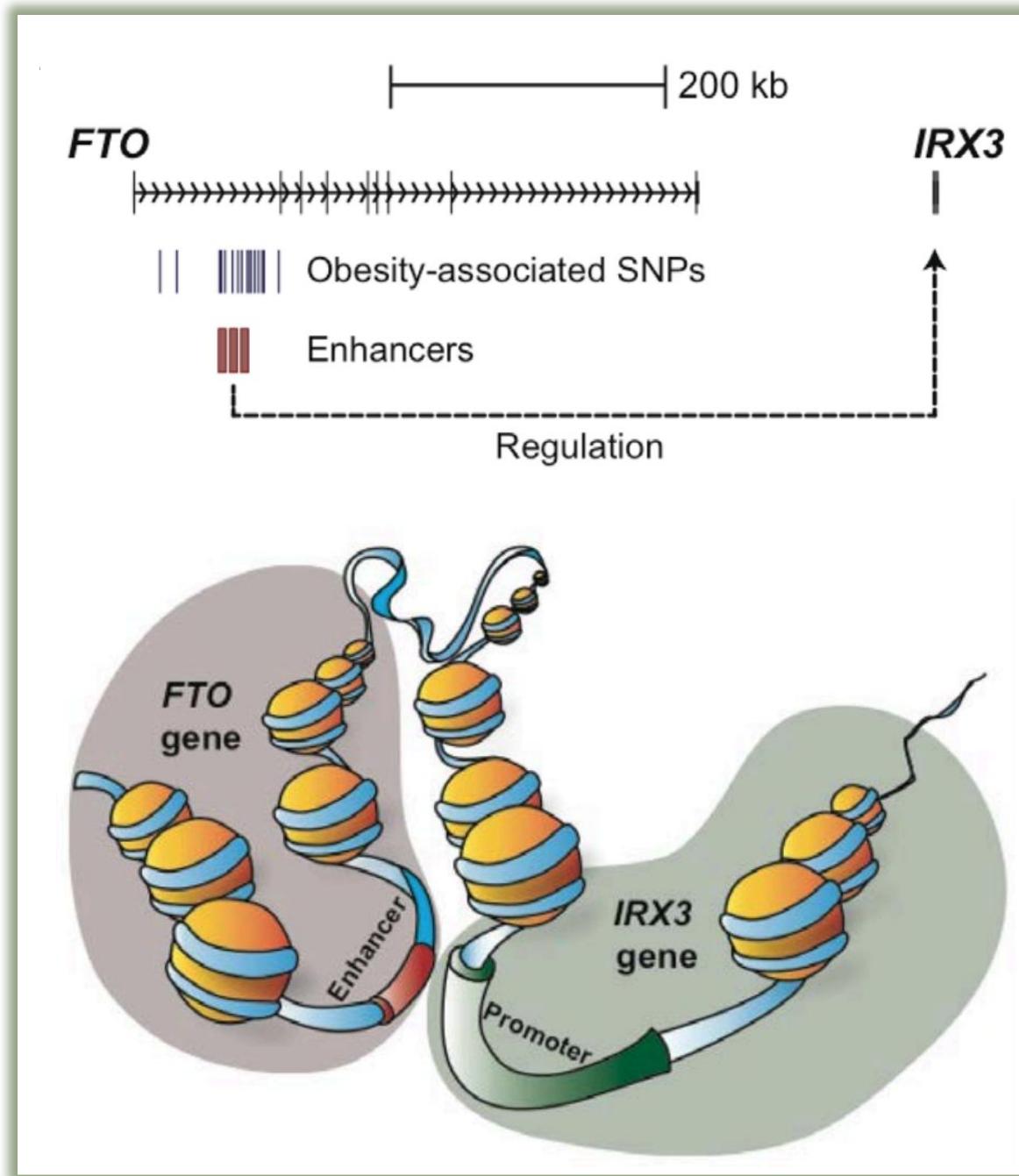
LETTER

doi:10.1038/nature13138

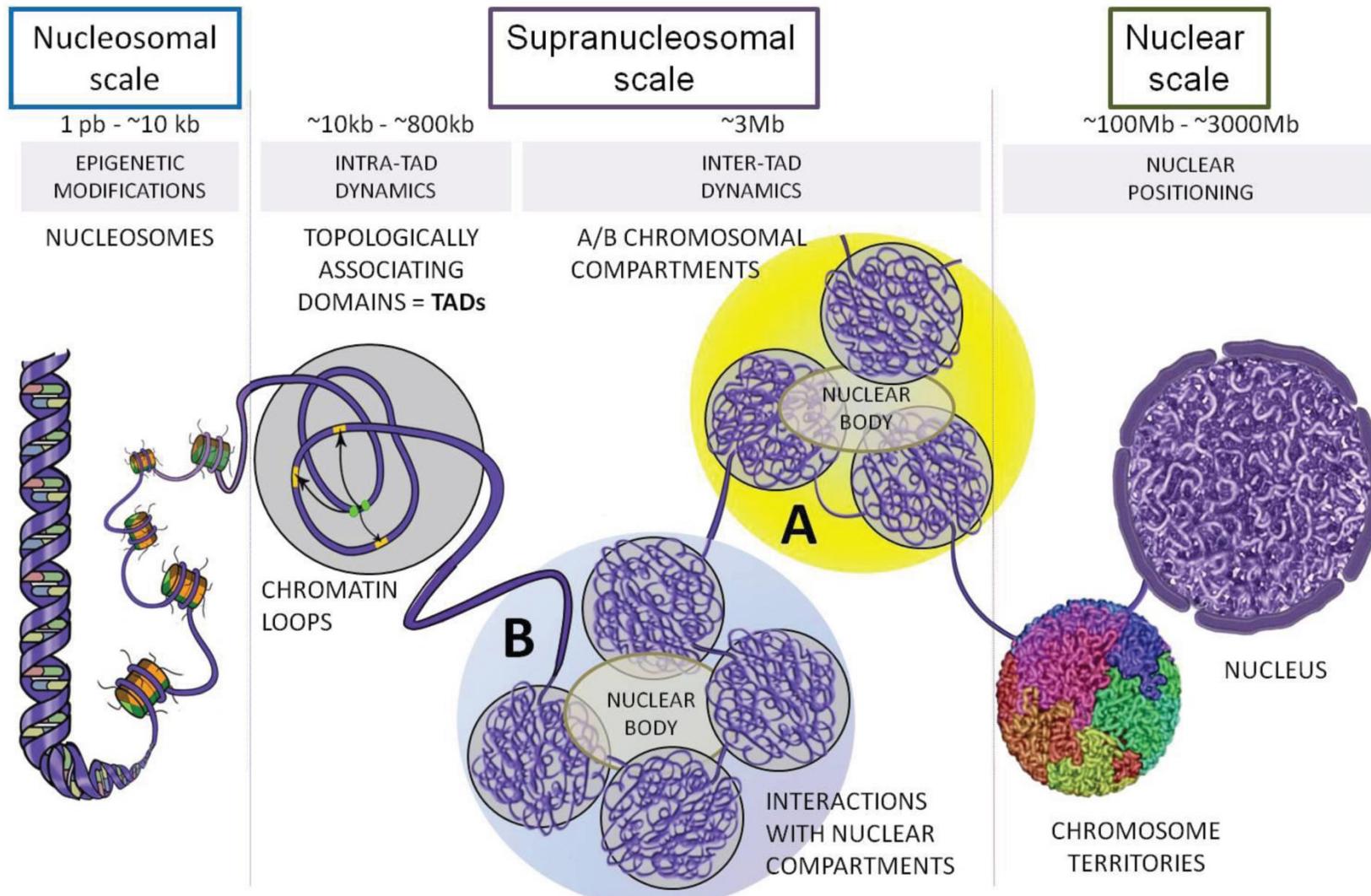
Obesity-associated variants within *FTO* form long-range functional connections with *IRX3*

Scott Smemo^{1*}, Juan J. Tena^{2*}, Kyoung-Han Kim^{3*}, Eric R. Gamazon⁴, Flavia L. Credidio¹, Débora R. Sobreira¹, Nora F. Wasserman¹, Ju Hee Lee¹, Joe Eun Son⁵, Niki Alizadeh Vakili³, Hoon-Ki Sung⁵, Silvia Naranjo², Nancy J. Cox^{1,4}, Chi-Chung Hui³, Jose Luis Gomez-Skarmeta² & Marcelo

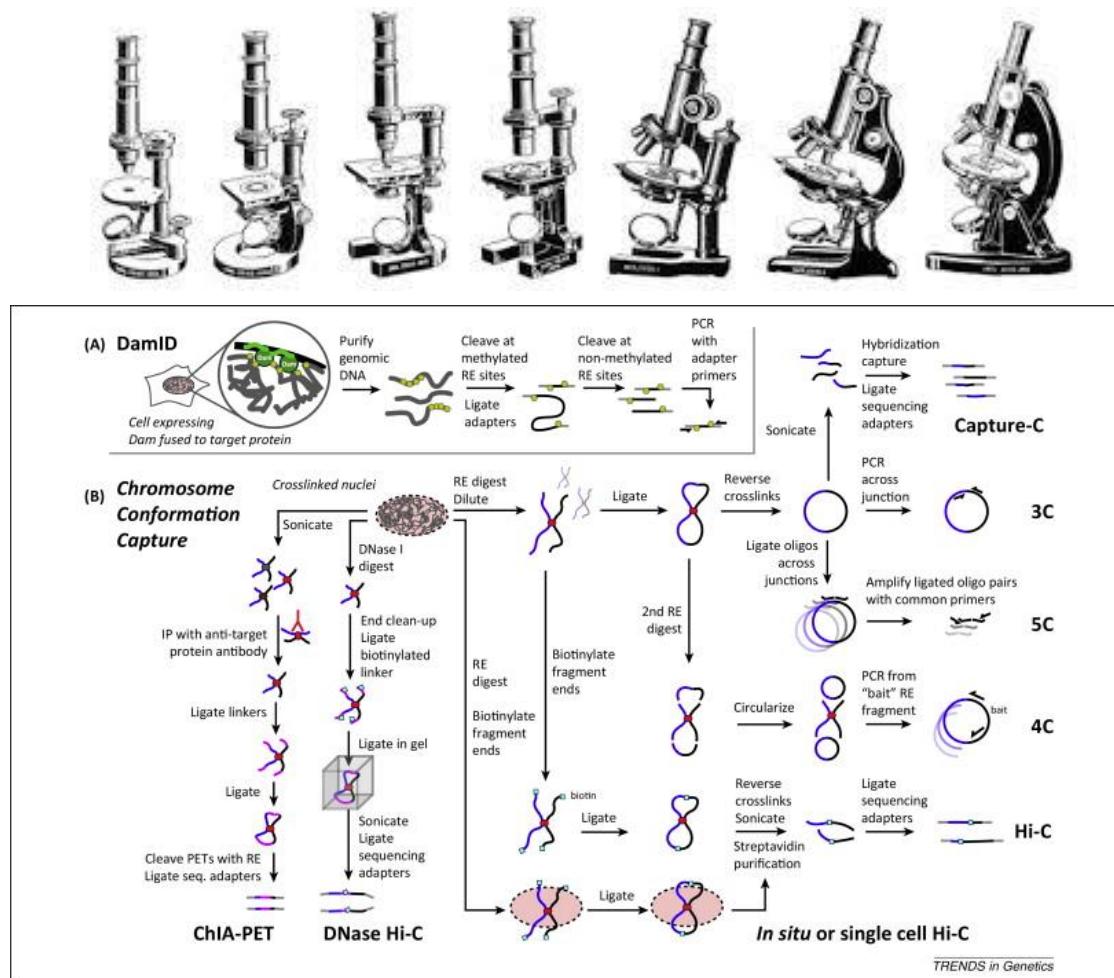




Genome organization in mammals



Technologies used (and developed) to study genome folding



Can be divided into two broad categories:

1. Imaging

1. Bright-field
2. Fluorescence
3. EM
4. Fluorescence in-situ hybridization (FISH), etc.

2. Genomics

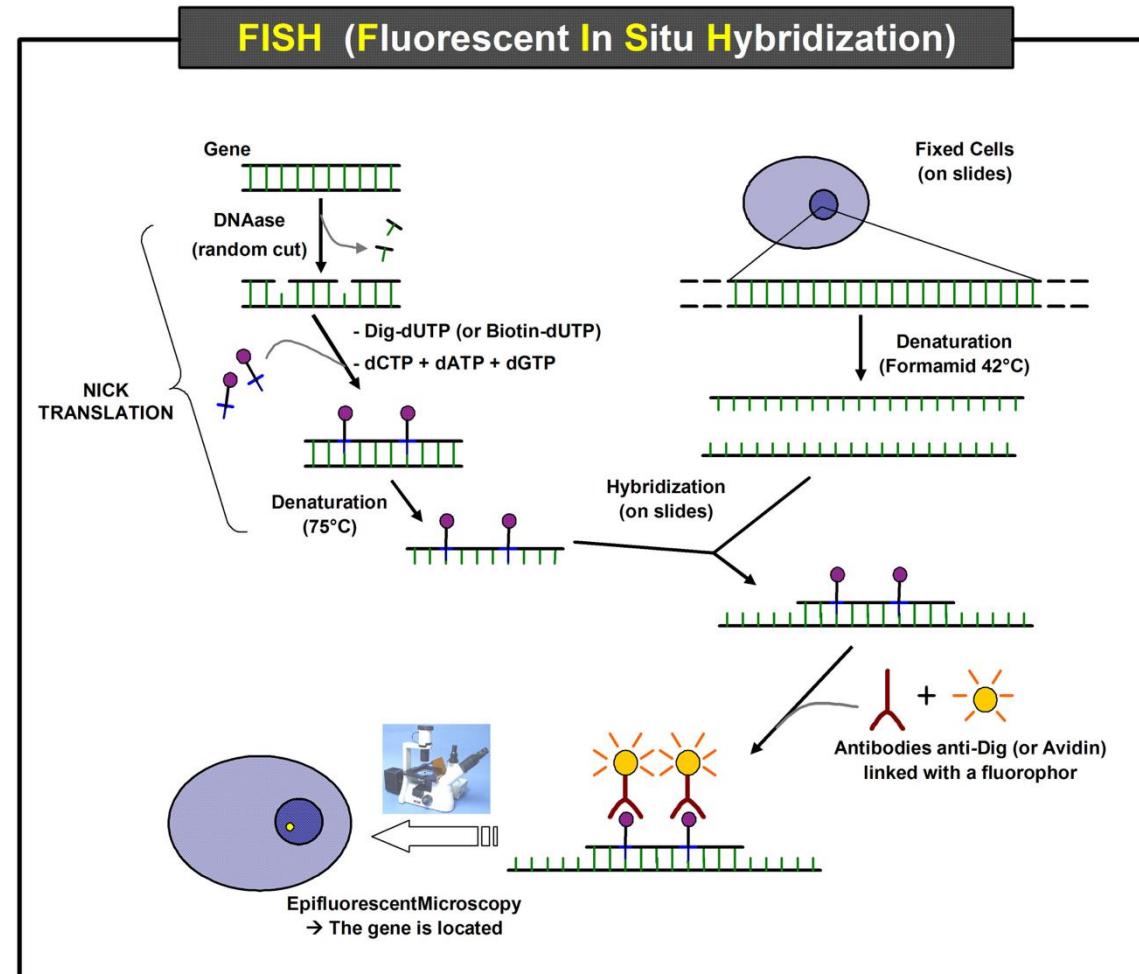
1. DamID
2. ChIA-PET
3. GAM
4. Chromosome conformation capture-derived, etc.

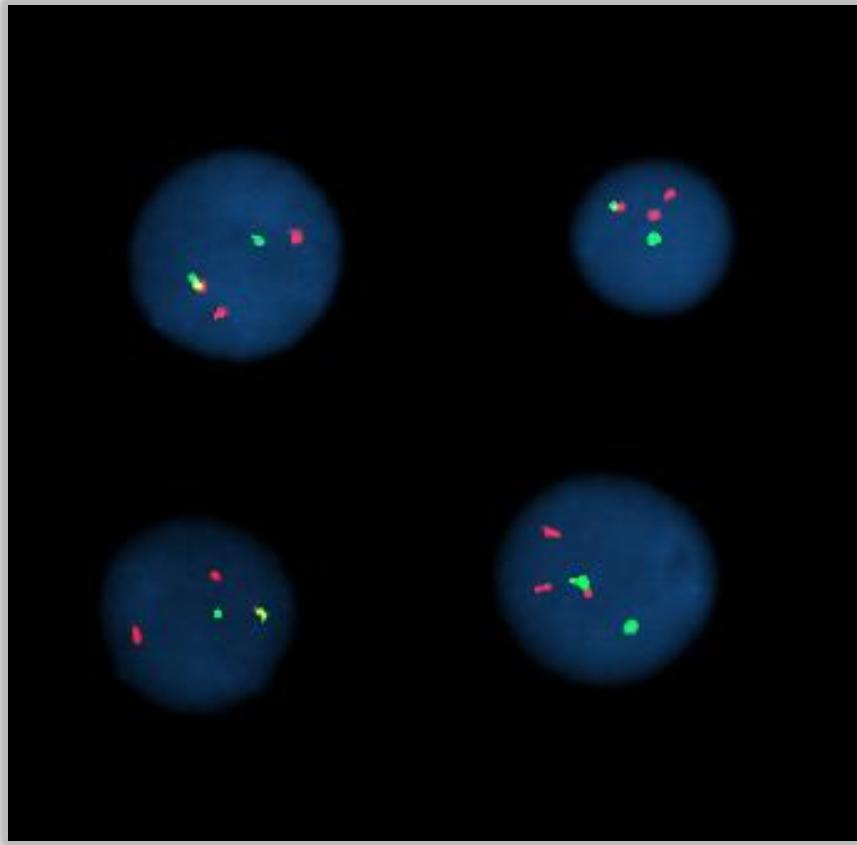
Adapted from <http://web.uvic.ca/ail/equipment.html>

Unraveling the 3D genome: genomics tools for multiscale exploration. Viviana I. Risca and William J. Greenleaf. *Trends in Genetics* (2015)

Fluorescence in situ hybridization

- Cytogenetic technique
- Uses fluorescent molecules to “paint” (regions of interest on) chromosomes in cells often in Metaphase or Interphase
- Aids in analysis of chromosome structure, structural aberrations, ploidy determination, etc.





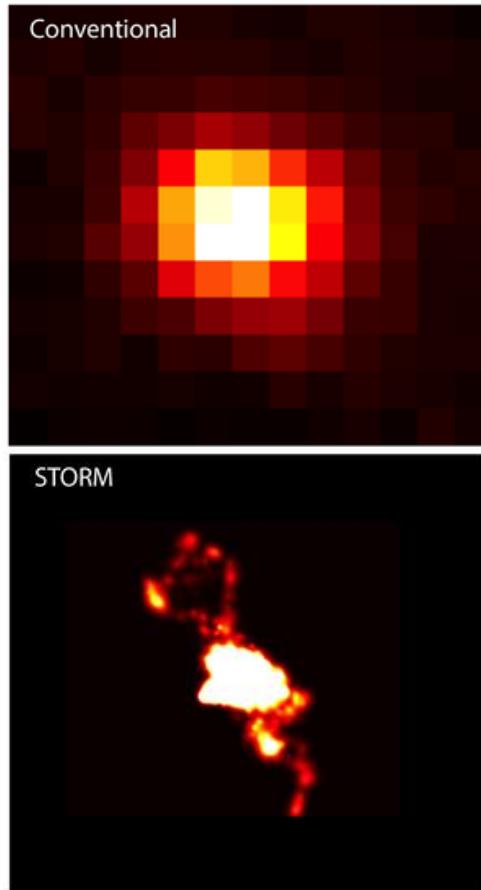
ADVANTAGES

- Rapid and sensitive
- Lots of cells can be analyzed
- No cell culture needed

DISADVANTAGES

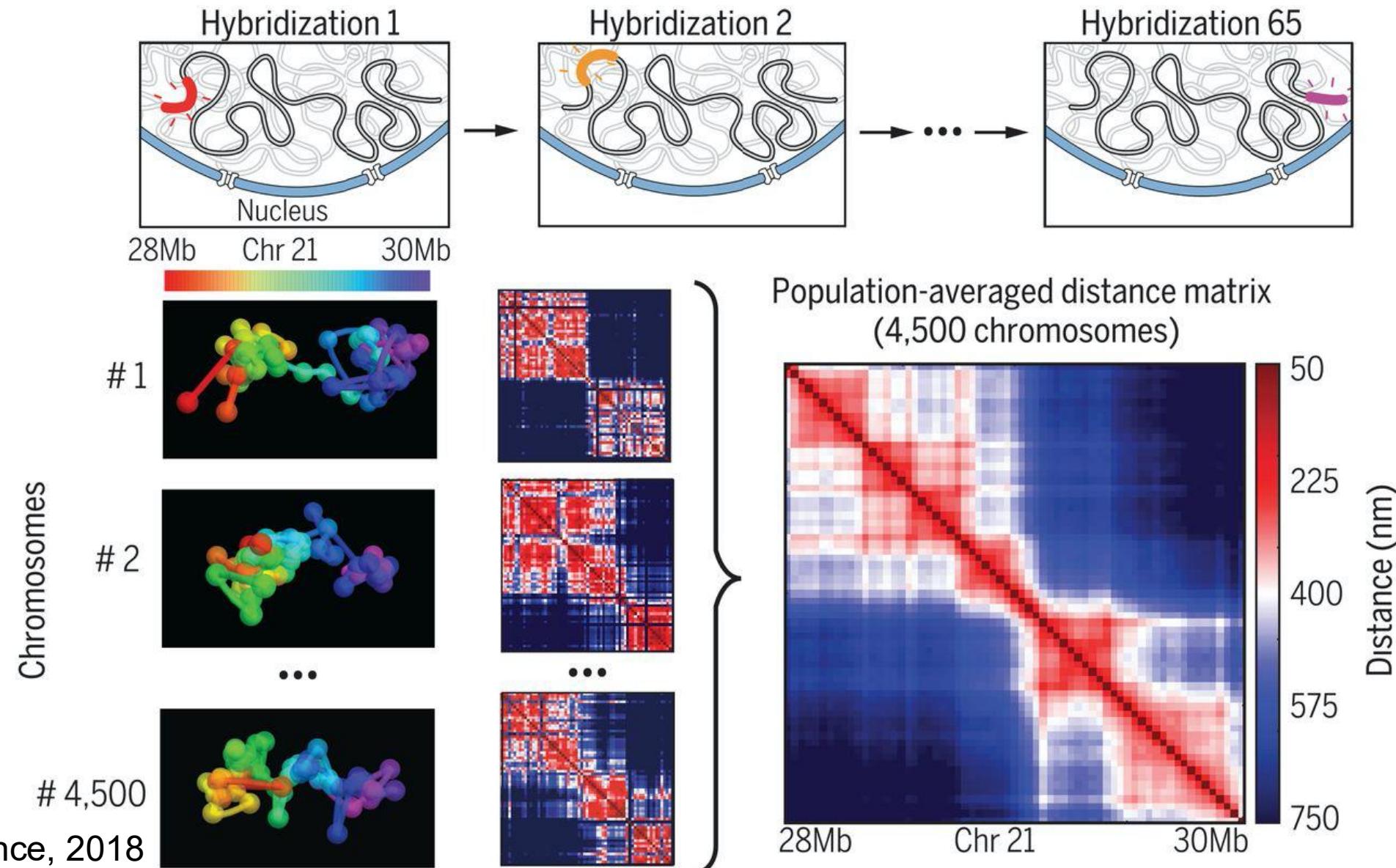
- Low-throughput
- Limited number of commercial probes available
- Needs specialized camera and image capture system

Multiplexed super-resolution FISH and stochastic optical reconstruction microscopy (STORM)

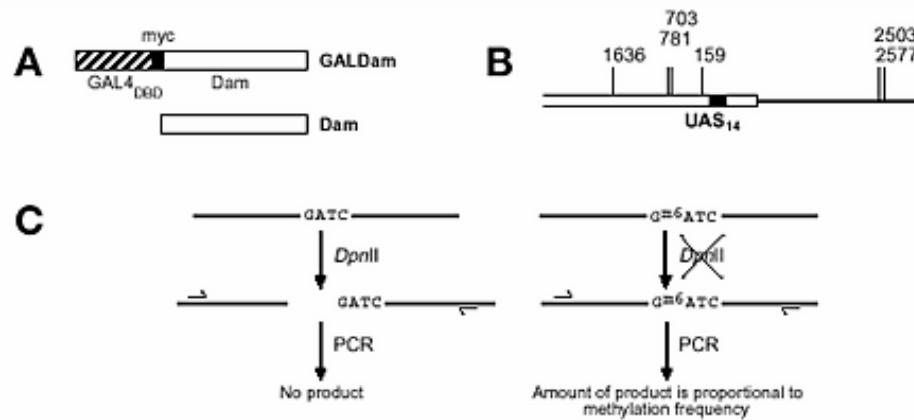


Beliveau et al. Nat Comm., 2015

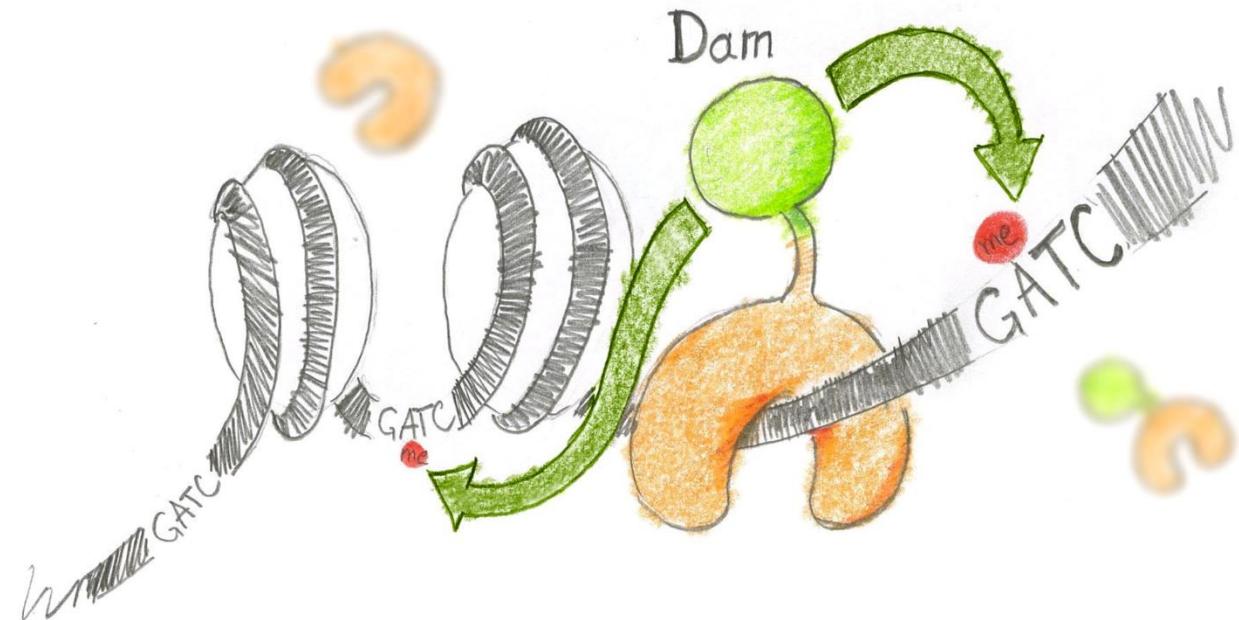
Bintu et al. Science, 2018



Dam-ID: Adenine methylation tagging to map nuclear interactions

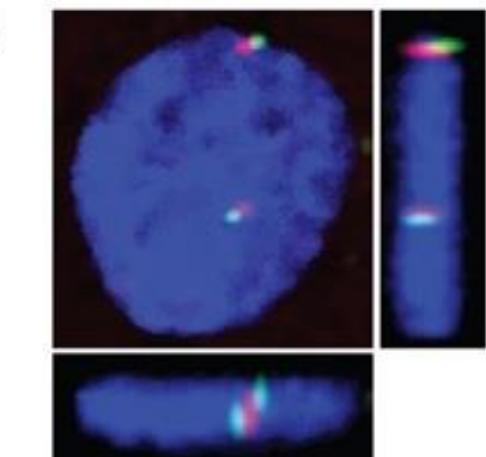
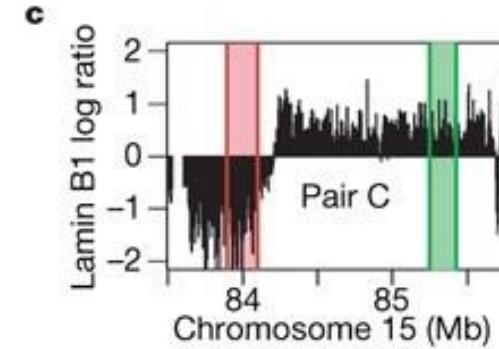
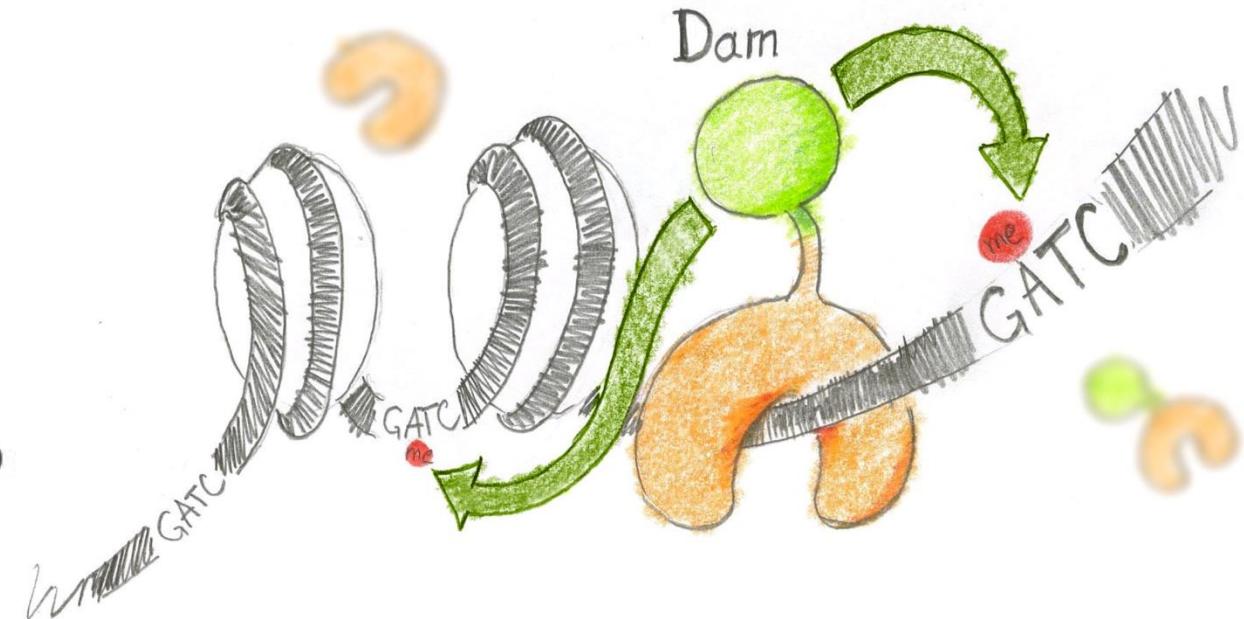
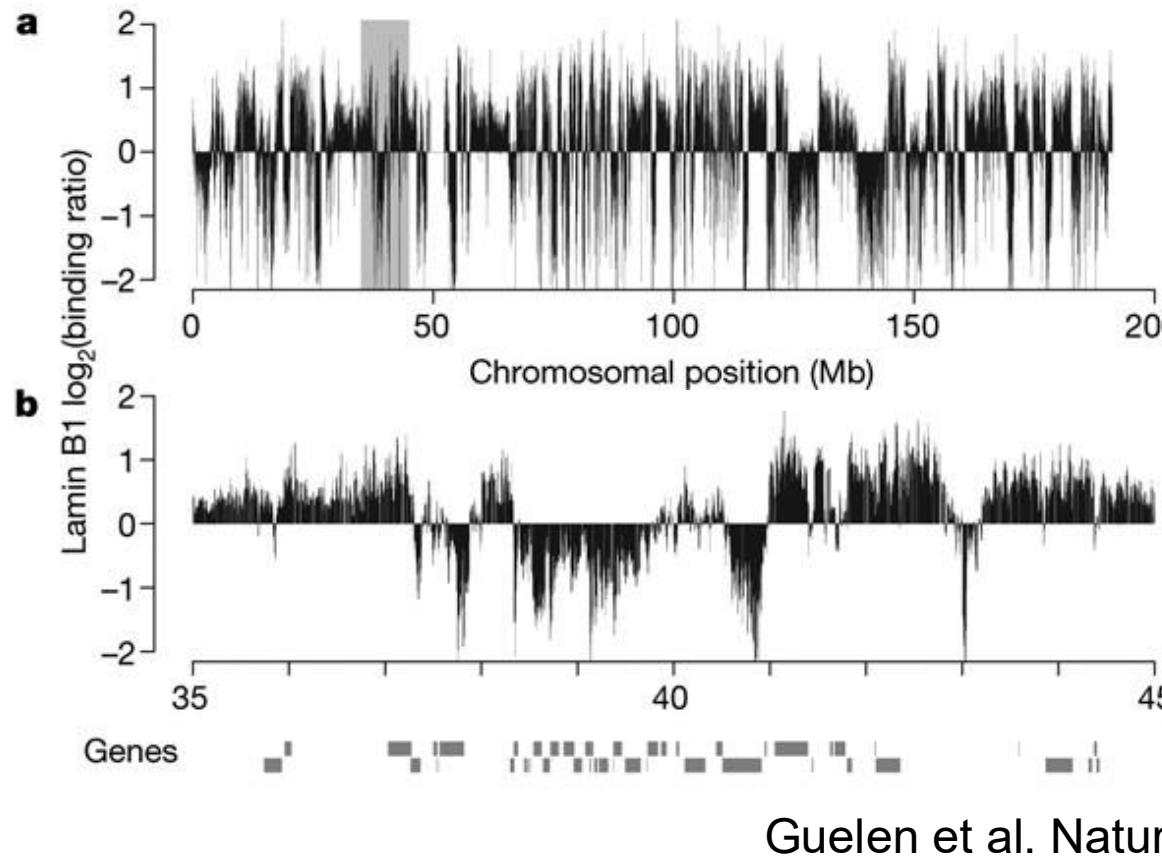


Van Steensel and Henikoff Nature Biotech., 2000



Dam-ID: Adenine methylation tagging to map nuclear interactions

Lamin-Dam fusion to tag DNA interactions with the nuclear periphery

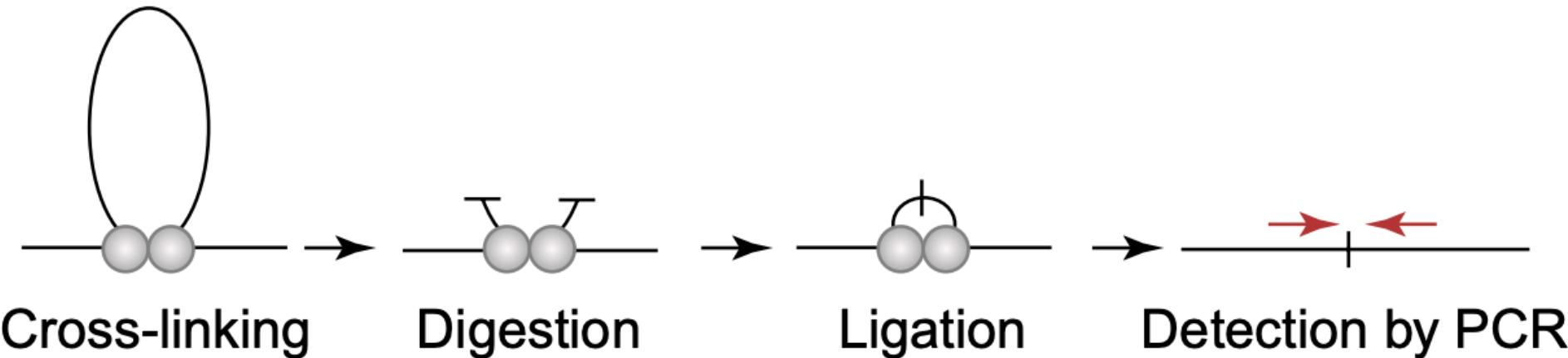


Capturing Chromosome Conformation

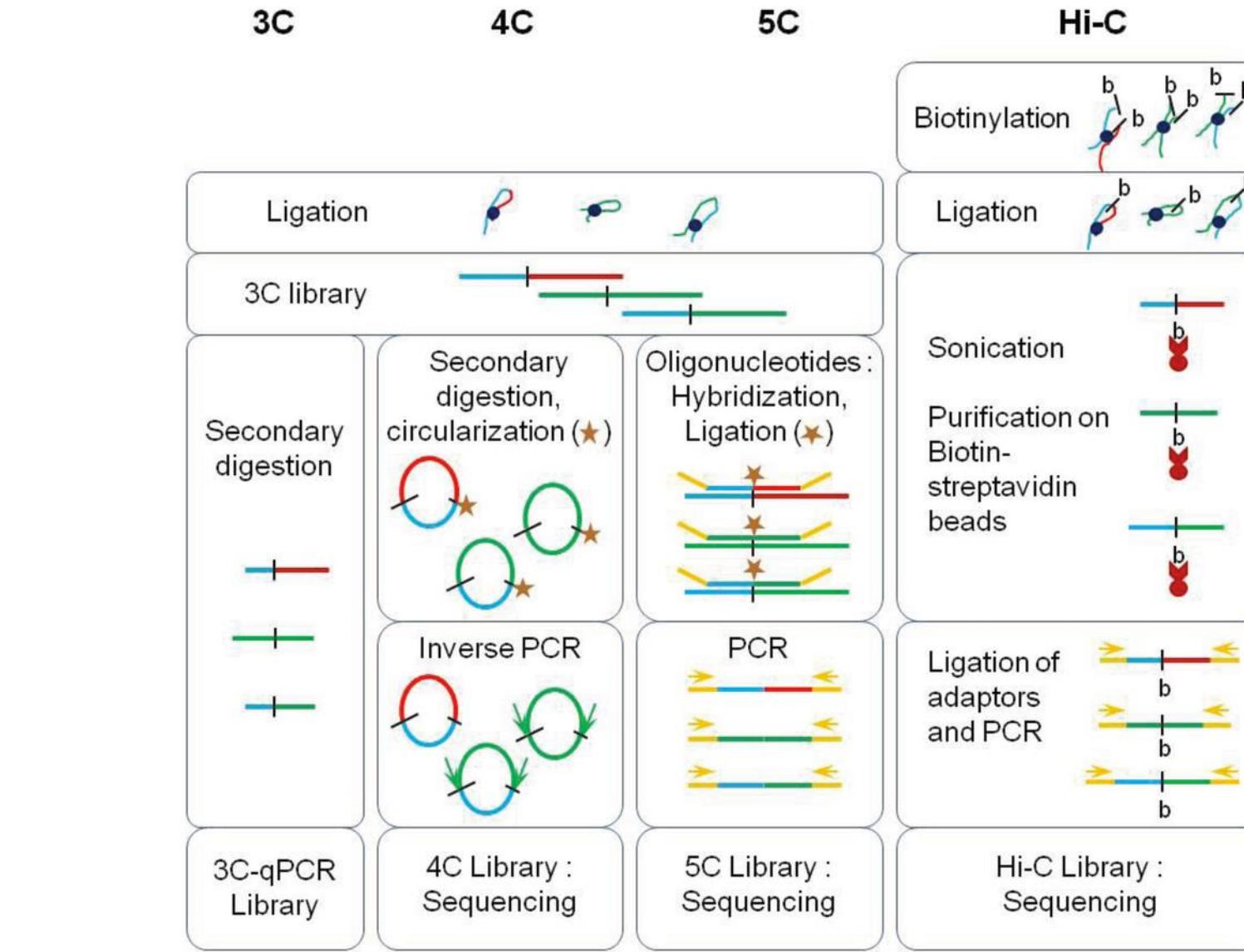
Job Dekker,^{1*} Karsten Rippe,² Martijn Dekker,³ Nancy Kleckner¹

15 FEBRUARY 2002 VOL 295 SCIENCE www.sciencemag.org

(a) 3C



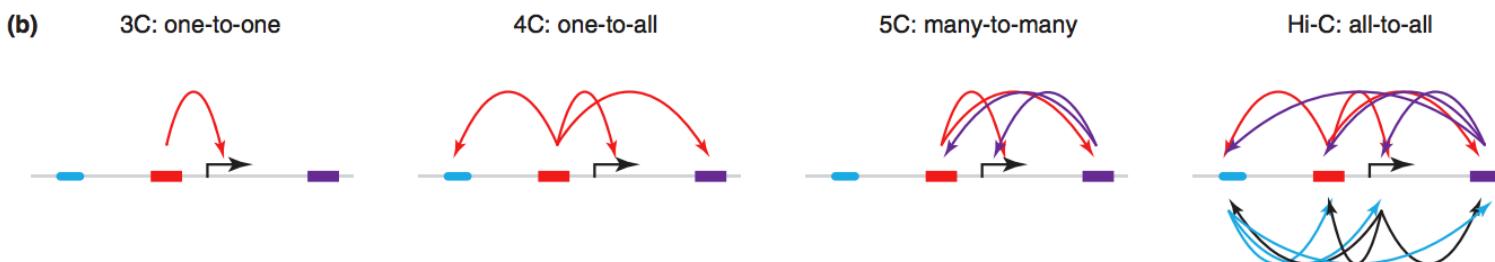
Methodology of “C”-technologies



Advantages and drawbacks of “C”-technologies

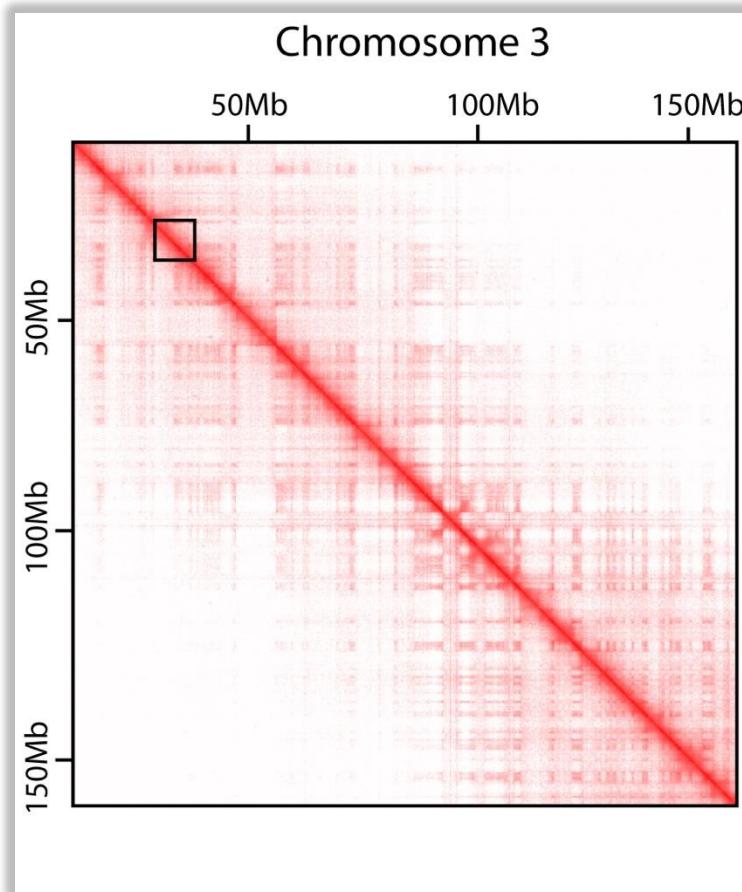
Table 1. Advantages and limits of 3C-derived methods.

Method	Genomic Scale Investigated	Advantages	Limits
3C-qPCR	~250 kilobases	Very high dynamic range (highly quantitative), easy data analysis	Very low throughput: limited to few viewpoints in a selected region
4C	Complete genome	Good sensitivity at large separation distances	Genome-wide contact map limited to a unique viewpoint (few viewpoints if multiplex sequencing is used)
5C	Few megabases	Good dynamic range, complete contact map (all possible viewpoints) of a specific locus	The contact map obtained is limited to a selected region
Hi-C	Complete genome	Very high throughput (complete contact map)	Poor dynamic range, complex data processing

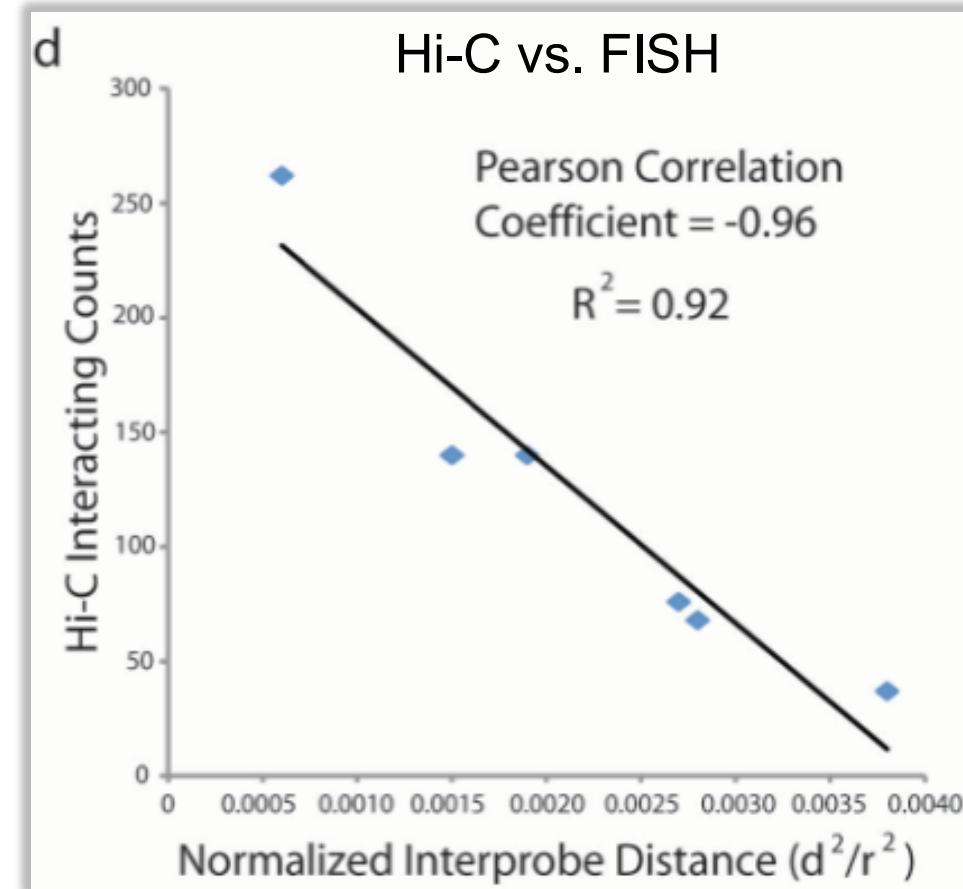


Contribution of Topological Domains and Loop Formation to 3D Chromatin Organization. Ea et al. Genes (2015)
Genome organization influences partner selection for chromosomal rearrangements. Patrick J. Wijchers & Wouter de Laat. Trends in Genetics (2011)

Hi-C for genome-wide analysis of higher order chromatin structure



Mouse ES cells
(from 433 Million Reads)



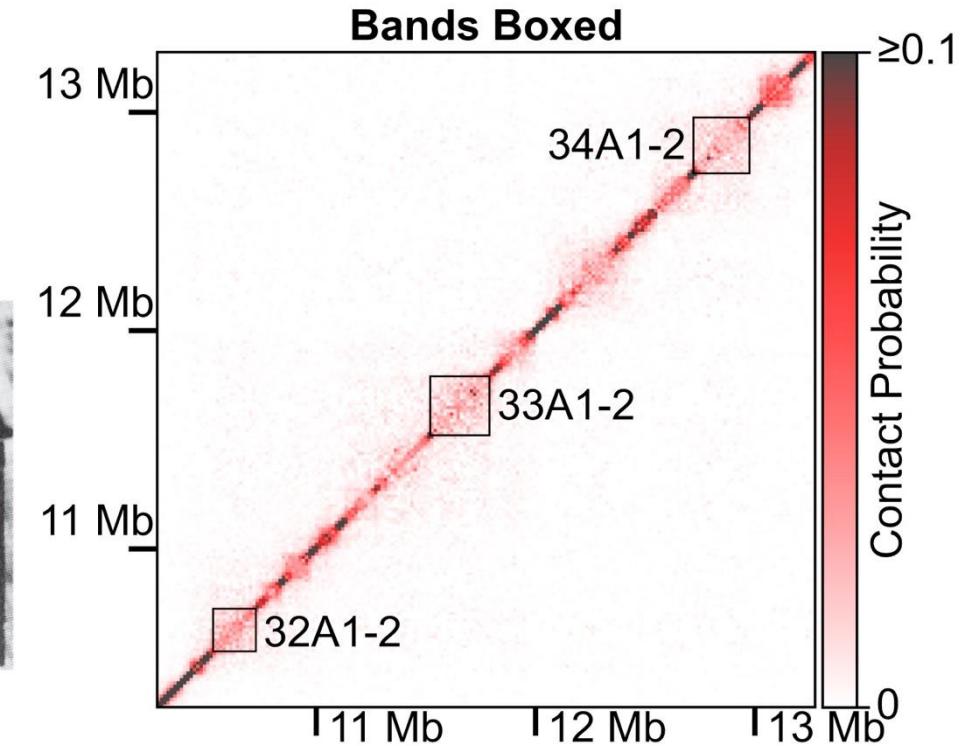
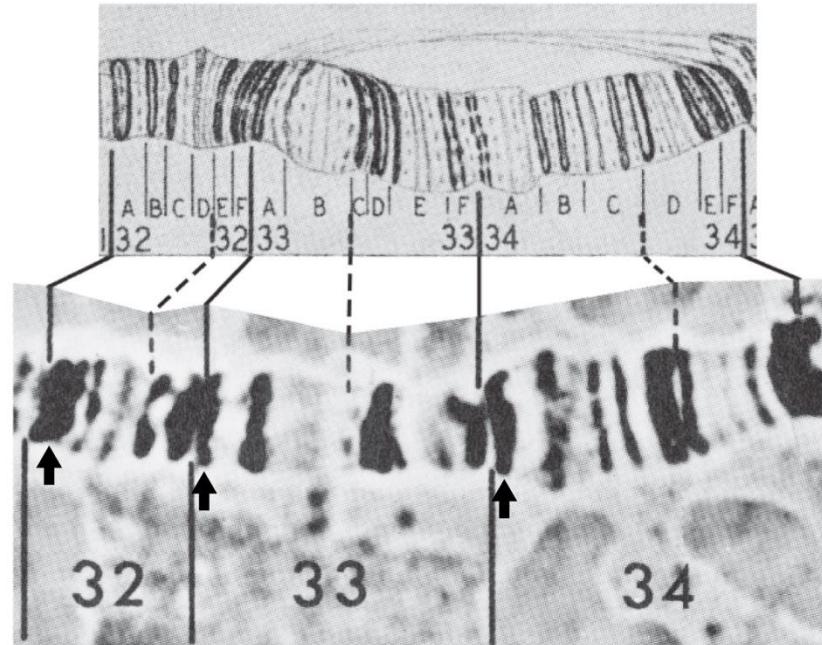
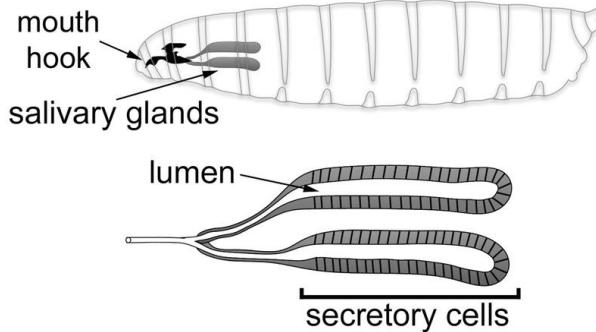
Dixon et al., Nature, 2012

Hi-C recapitulates *Drosophila* polytene chromosome banding patterns observed 80 years prior



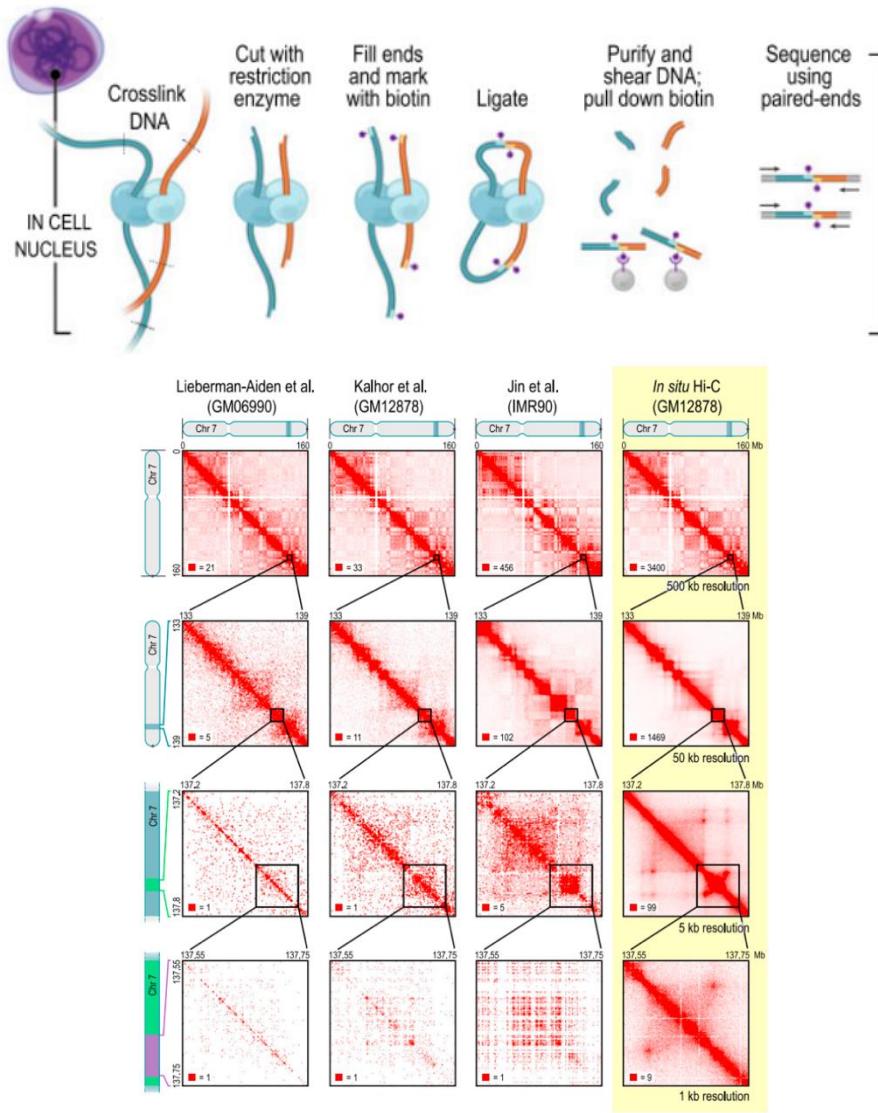
Calvin Blackman Bridges, 1927.
Photo courtesy of Cold Spring Harbor Laboratory Archives.

3rd instar larva (L3)



Eagen et al., Cell, 2015

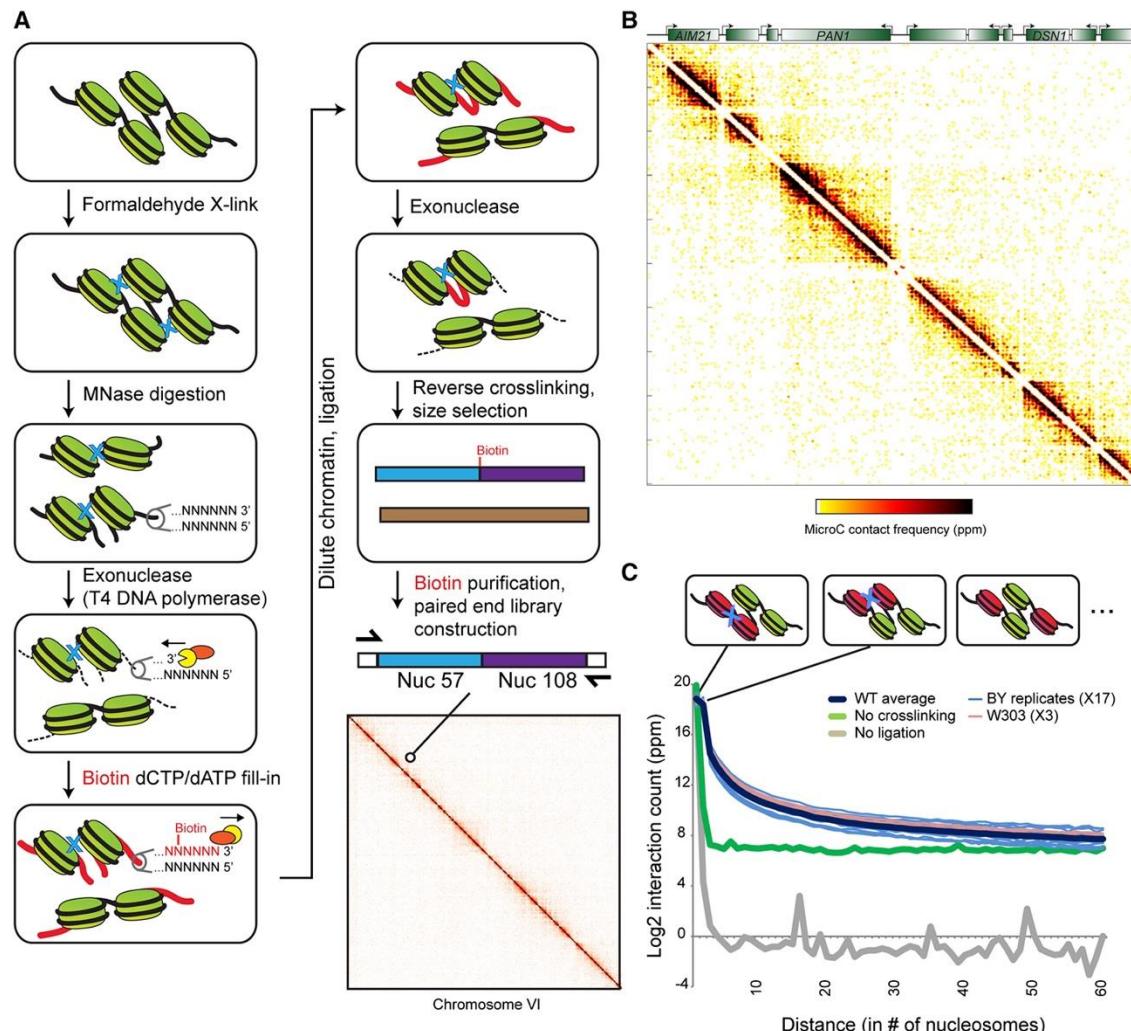
in situ Hi-C



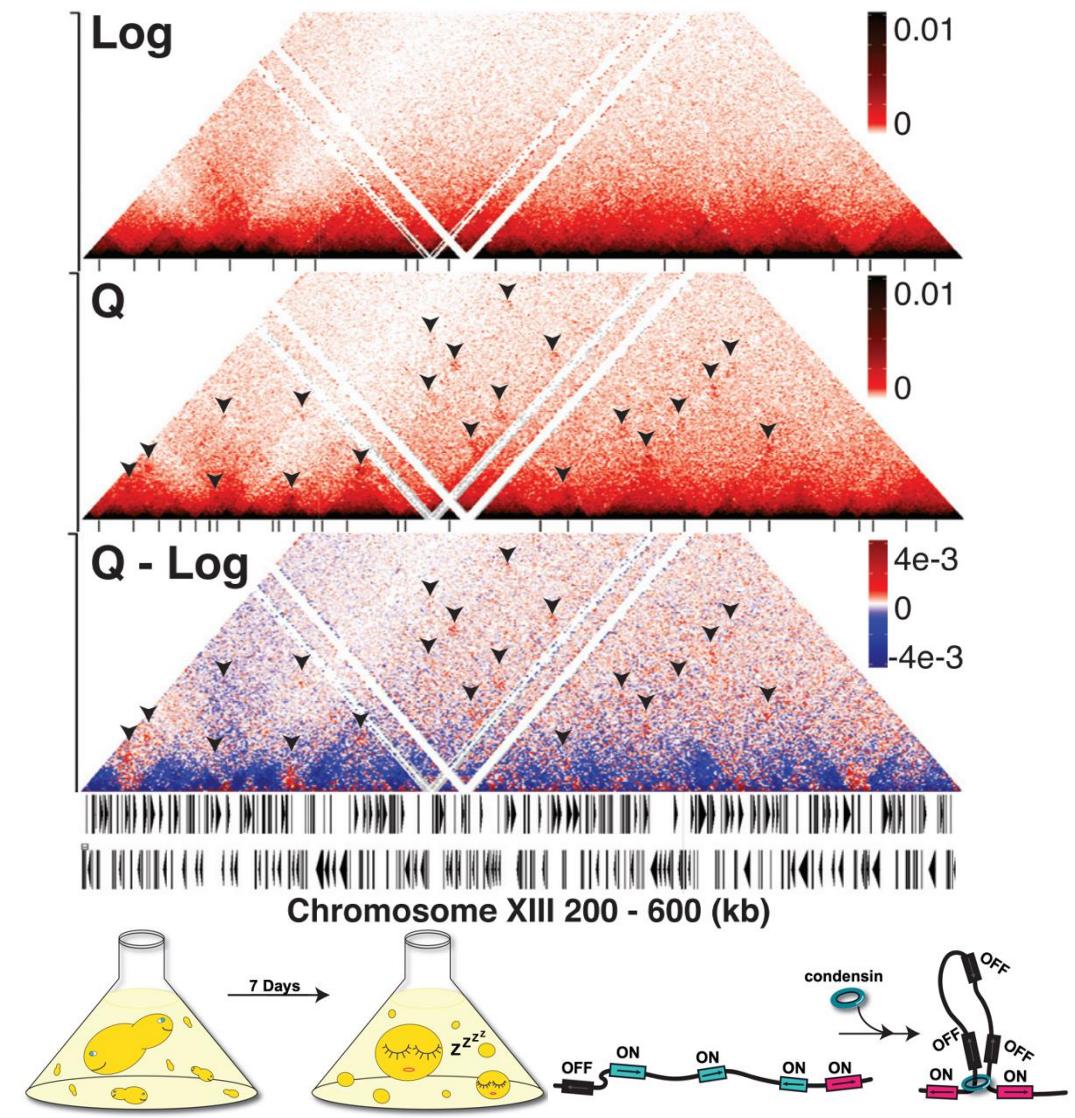
In situ Hi-C maps DNA–DNA contacts occurring in intact nuclei, by proximity ligation.

While initial studies achieved only megabase resolution, the latest study with 15 billion contact reads, reaches kilobase resolution—which is a function of both the size of the restriction enzyme recognition sequence and the sequencing depth of the library.

Micro-C: Nucleosome resolution Hi-C

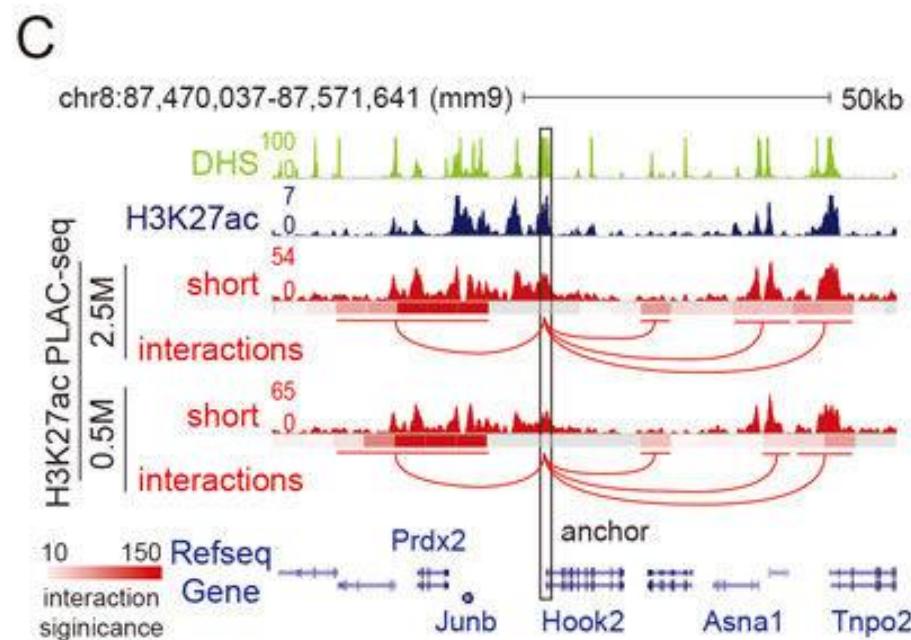
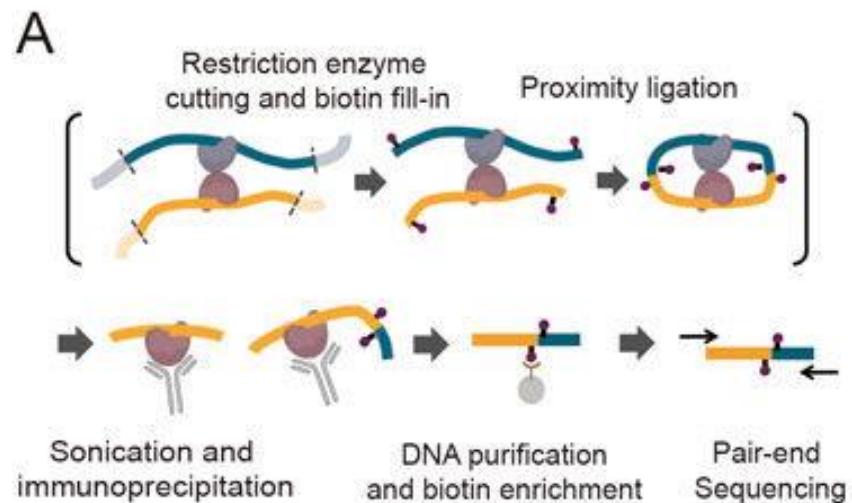


Hsieh et al. Cell, 2015



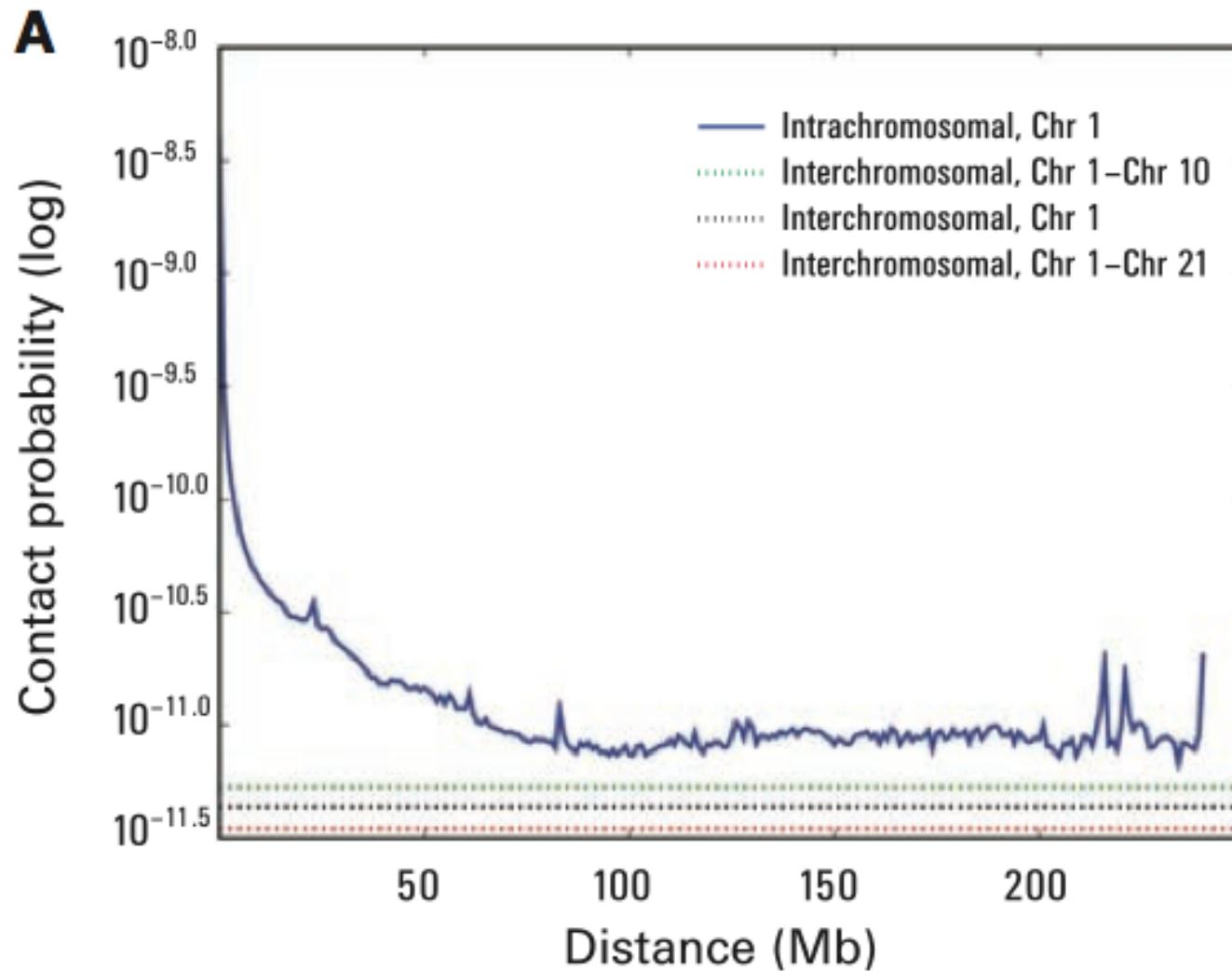
Swygert et al. Mol Cell, 2019

HiChIP/PLAC-seq

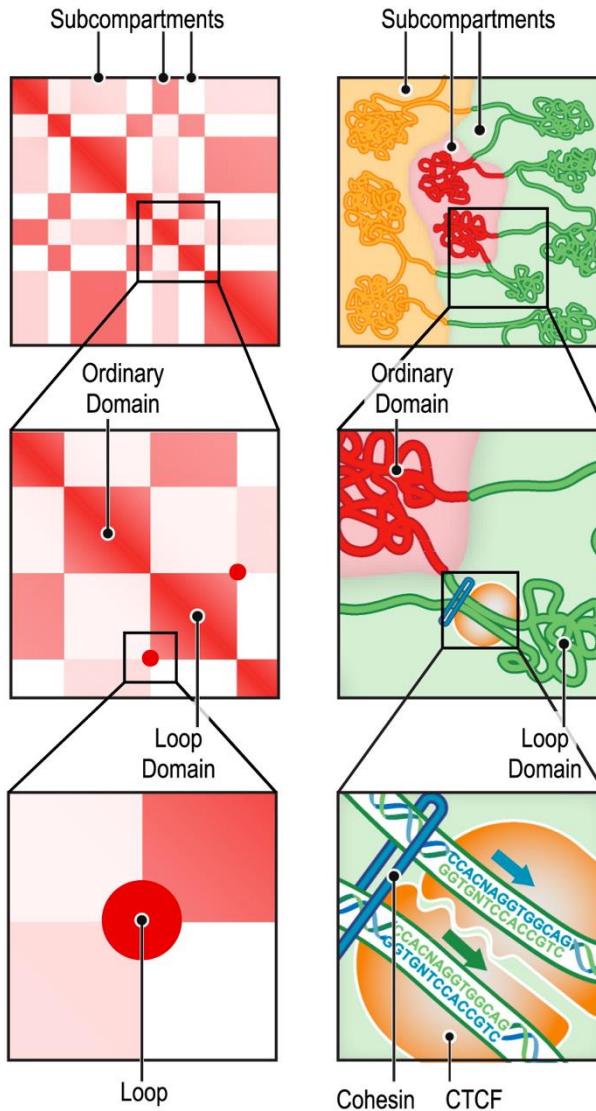


Fang et al., 2016
Mumbach et al., 2016

Distribution of contact frequencies



Overview of features revealed by Hi-C maps

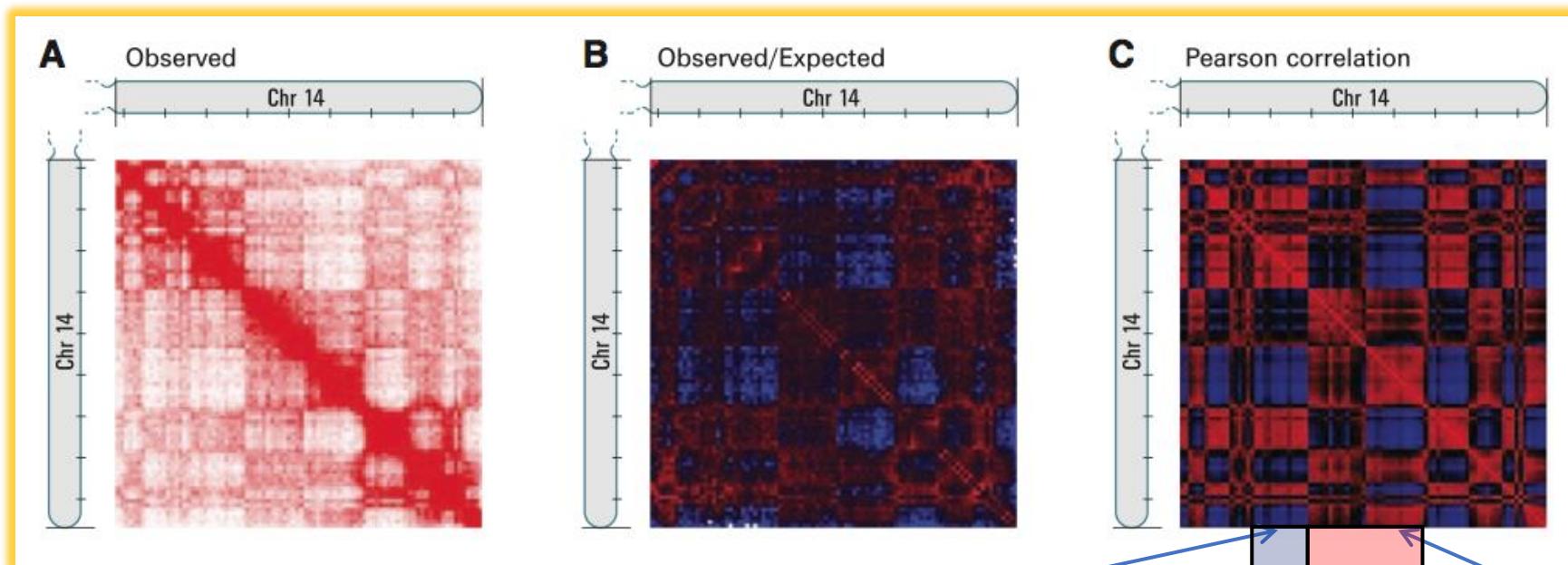


Top: the long-range contact pattern of a locus (left) indicates its nuclear neighborhood (right).

Middle: squares of enhanced contact frequency along the diagonal (left) indicate the presence of small domains of condensed chromatin.

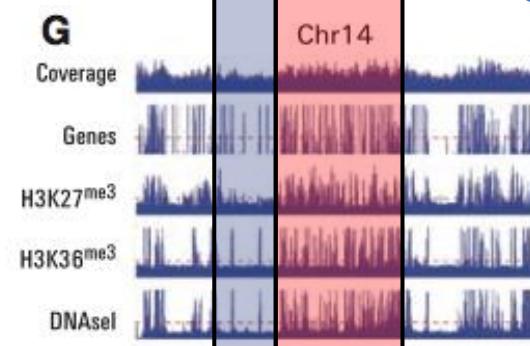
Bottom: peaks in the contact map (left) indicate the presence of loops (right). These loops tend to lie at domain boundaries and bind CTCF in a convergent orientation.

A and B Compartments



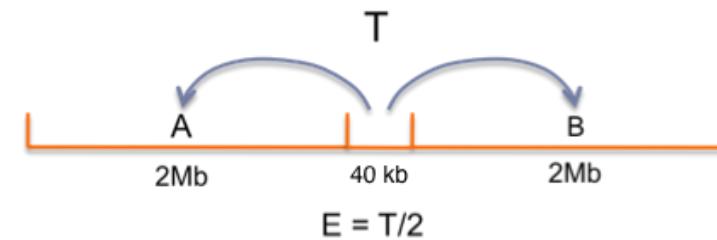
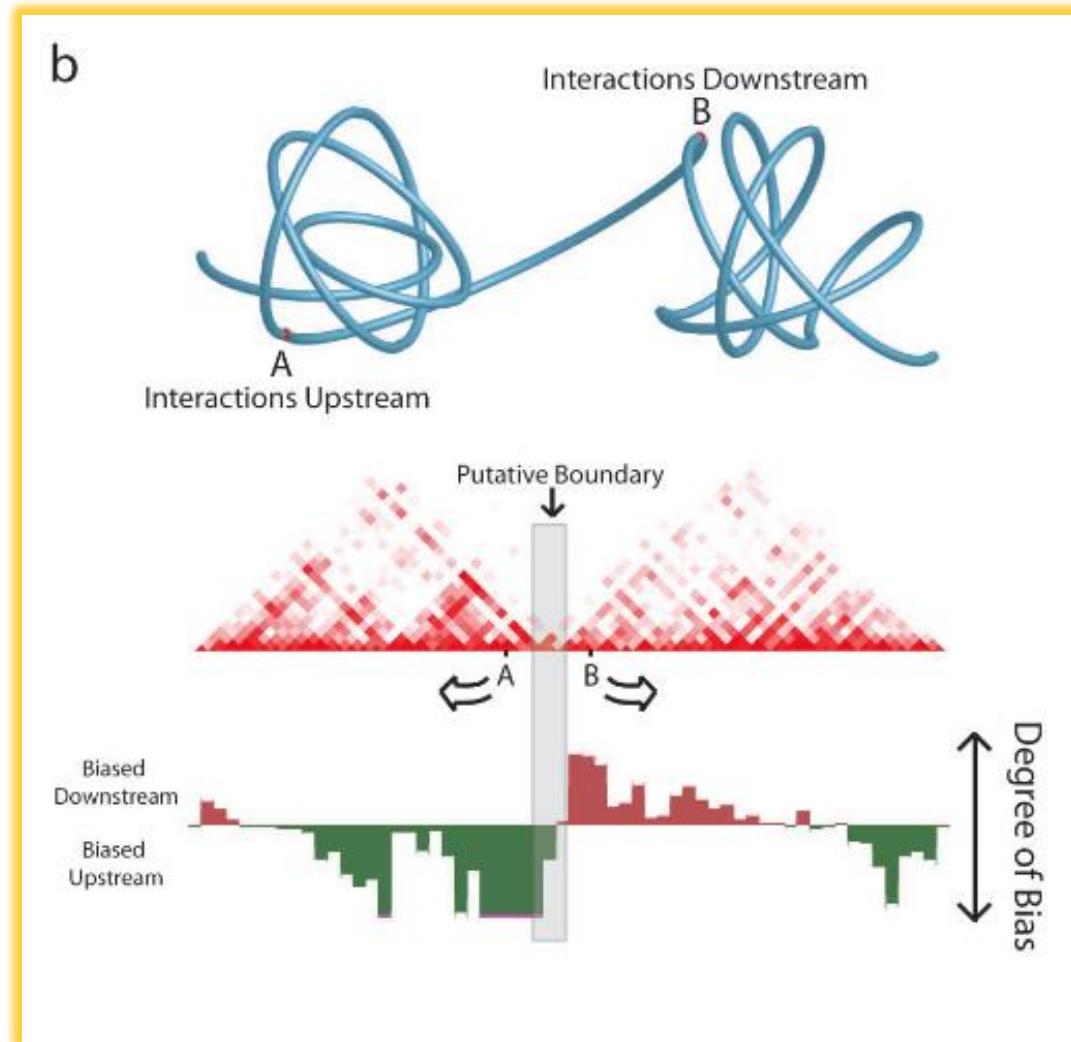
Compartment B

Compartment A



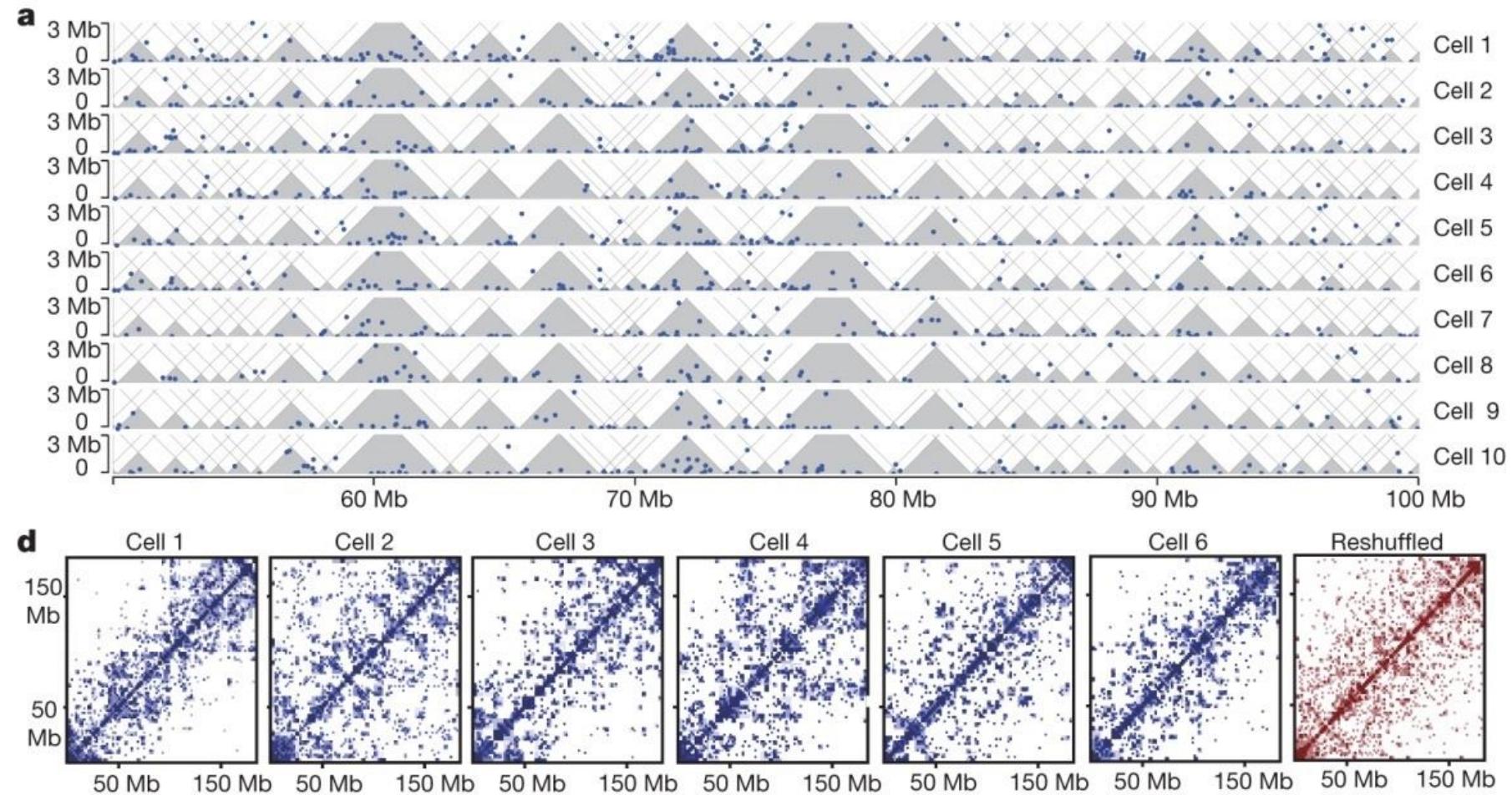
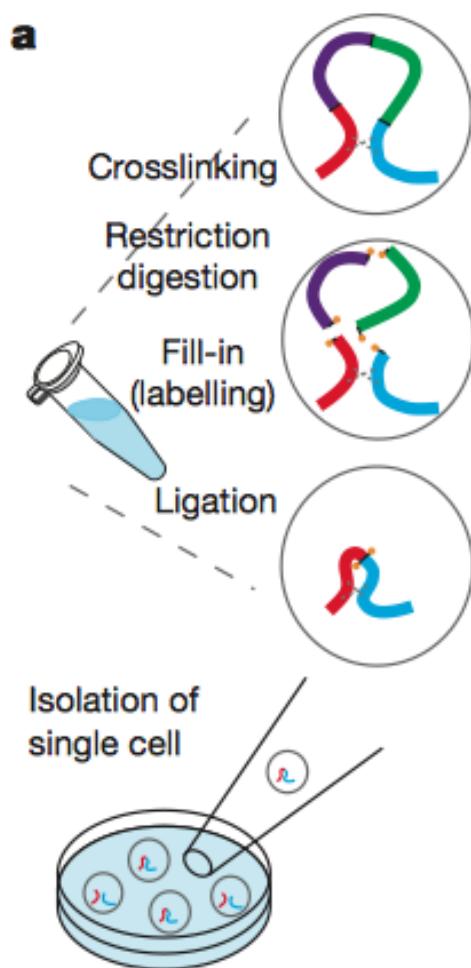
Lieberman-Aiden et al., 2009

Hi-C data reveals strong local chromatin interaction domains

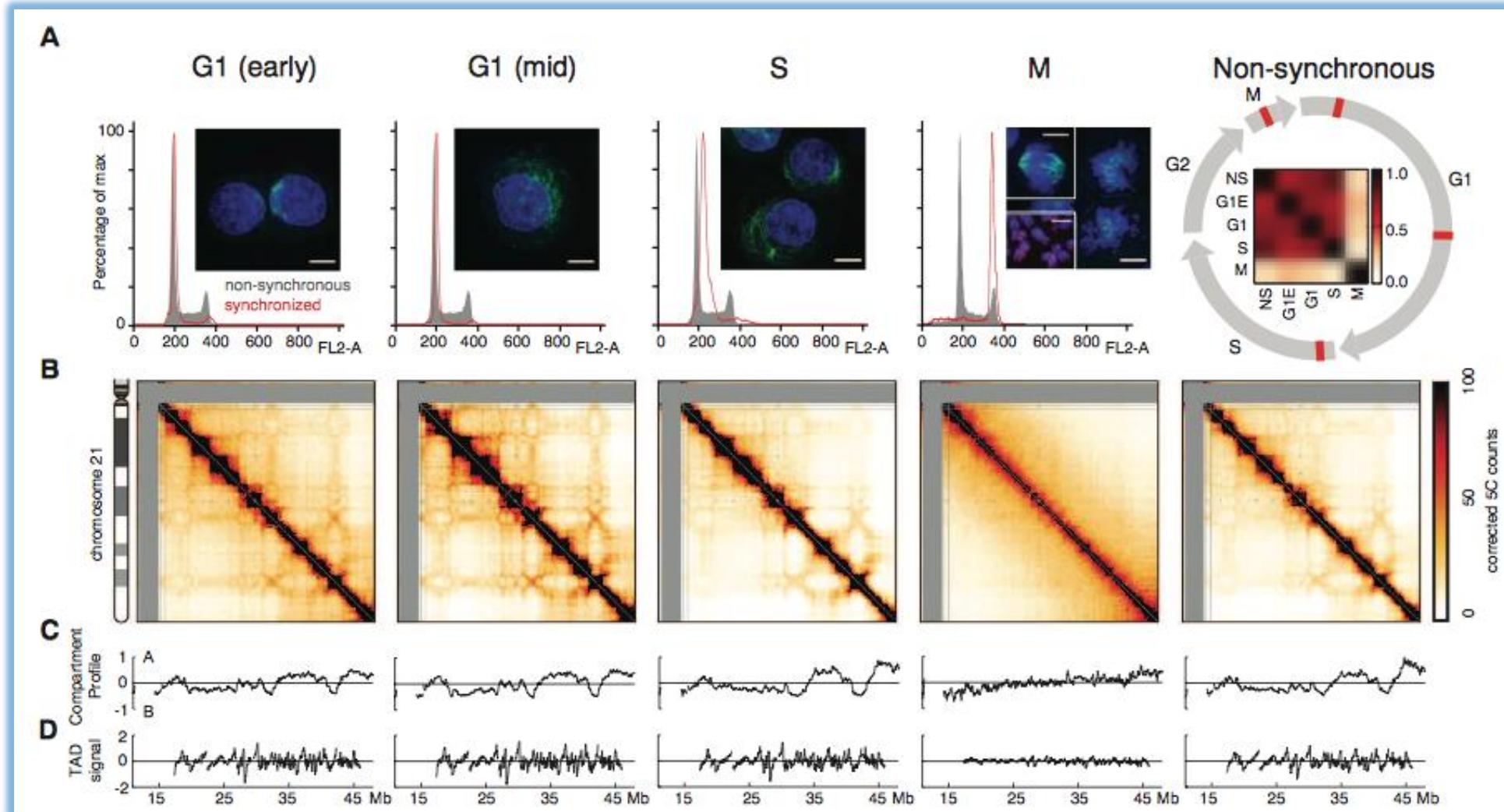


$$DI(x) = \frac{(B - A)}{\text{abs}(B - A)} \left(\frac{(A - E)^2}{E} + \frac{(B - E)^2}{E} \right)$$

Distribution of contact frequencies – cell to cell variability



Dynamic chromatin organization during cell cycle



Visualization of Hi-C data

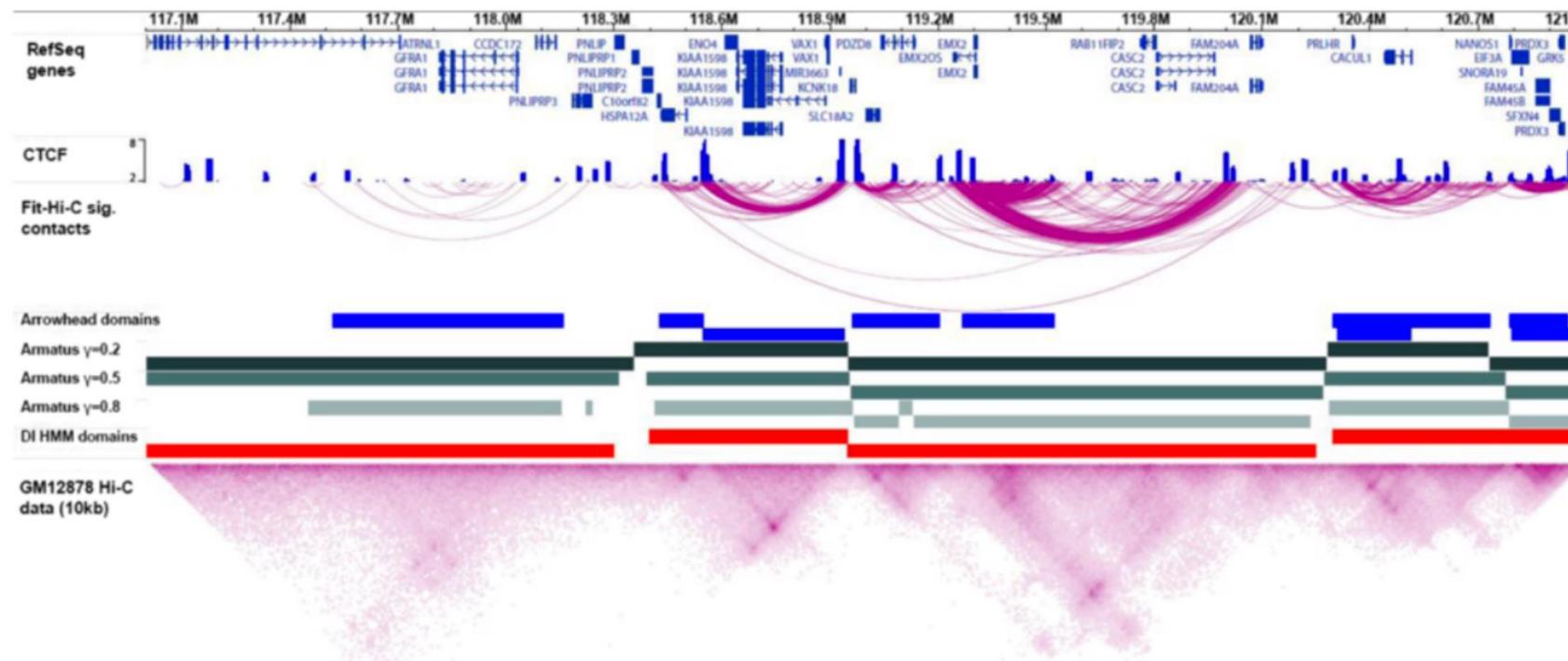
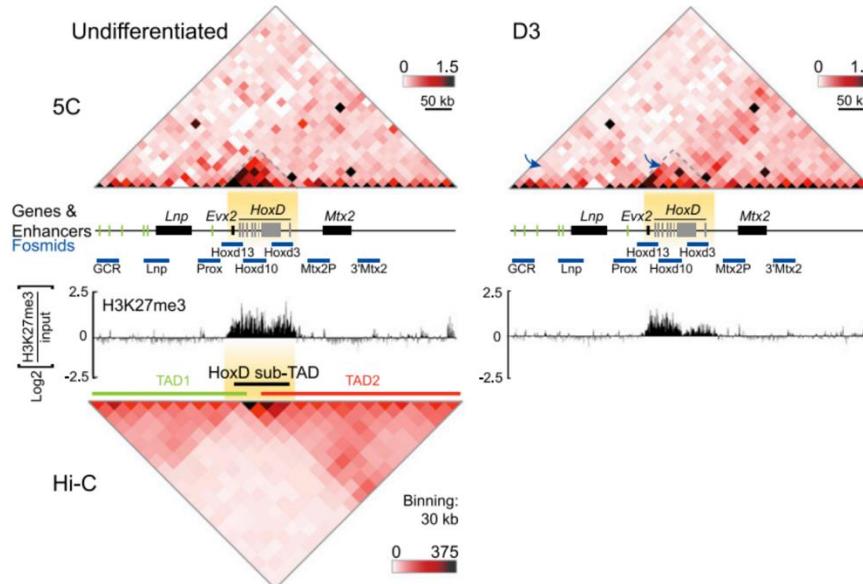


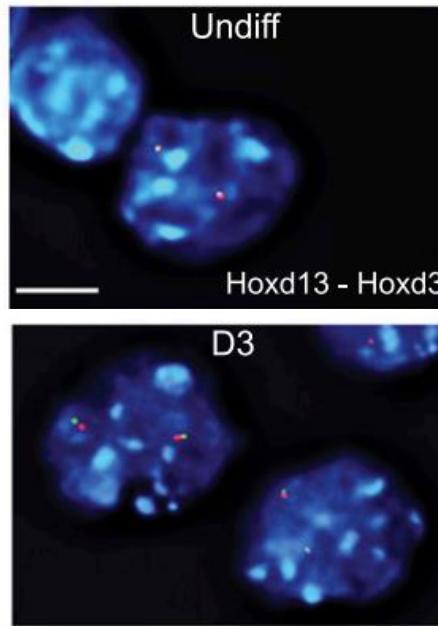
Fig. 3 Visualization of Hi-C data. An Epigenome Browser snapshot of a 4 Mb region of human chromosome 10. Top track shows Refseq genes. All other tracks display data from the human lymphoblastoid cell line GM12878. From top to bottom these tracks are: smoothed CTCF signal from ENCODE [130]; significant contact calls by Fit-Hi-C using 1 kb resolution Hi-C data (only the contacts >50 kb distance and $-\log(p\text{-value}) \leq 25$ are shown) [20]; arrowhead domain calls at 5 kb resolution [18]; Armatus multiscale domain calls for three different values of the domain-length scaling factor γ [87]; DI HMM TAD calls at 50 kb resolution [15]; and the heatmap of 10 kb resolution normalized contact counts for GM12878 Hi-C data [18]. The color scale of the heatmap is truncated to the range 20 to 400, with higher contact counts corresponding to a darker color

What you “C” might not be what you see



“C” technologies

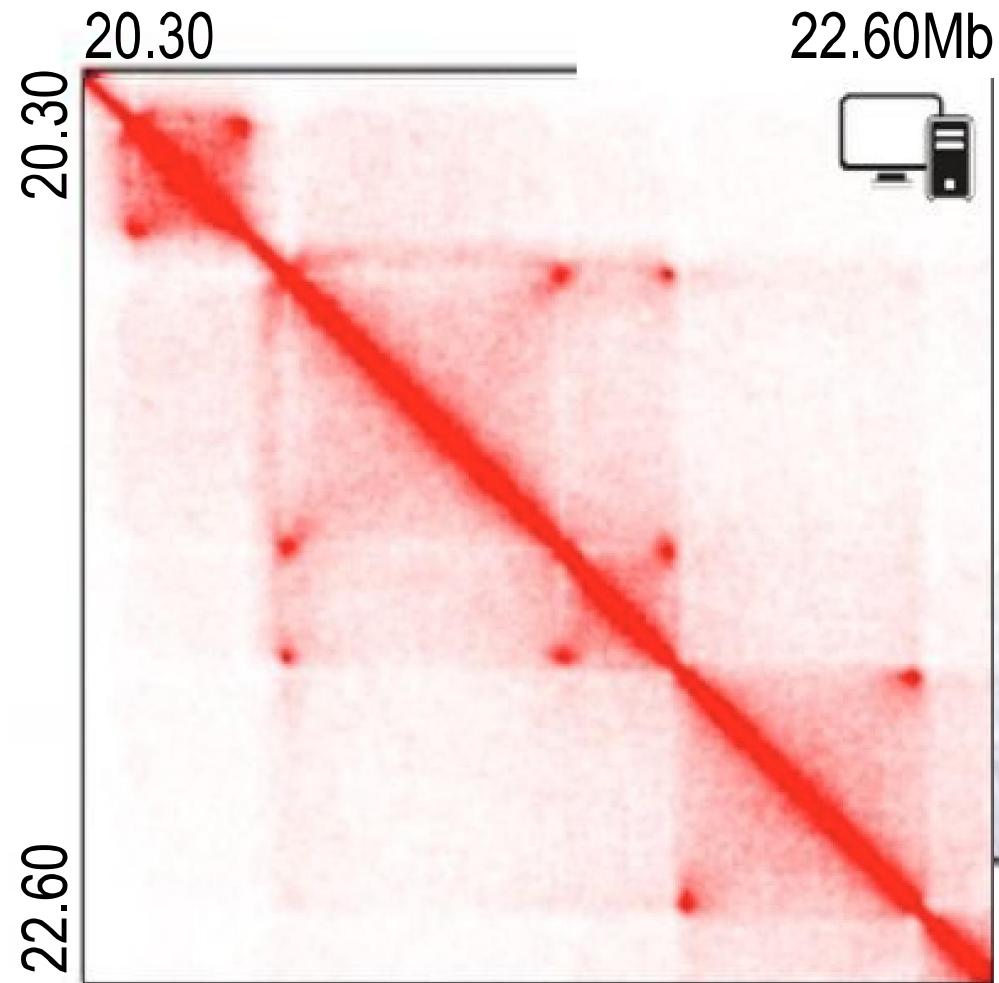
- ✓ high throughput
- ✓ high resolution
- ✗ low ligation (detection) efficiency
- ✗ indirect cross-linking via nuclear structure
- ✗ not readily applicable to low cell number



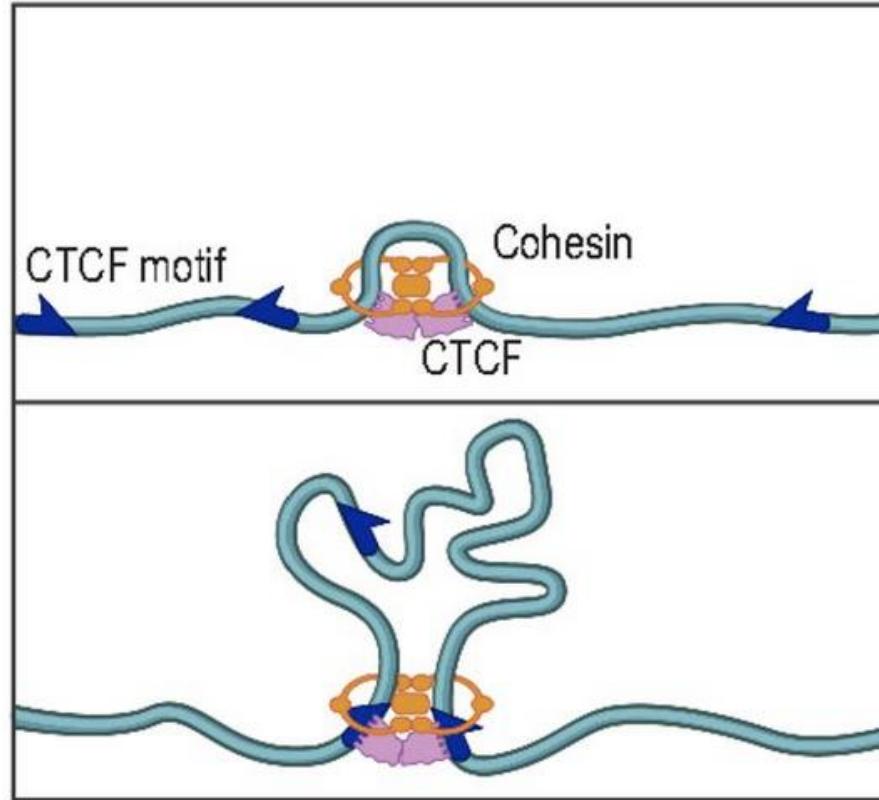
Imaging

- ✗ low throughput
- ✗ limited resolution
- ✓ high detection efficiency
- ✓ direct visualization of proximity
- ✓ readily applicable to single cells

Test: draw the loop structures

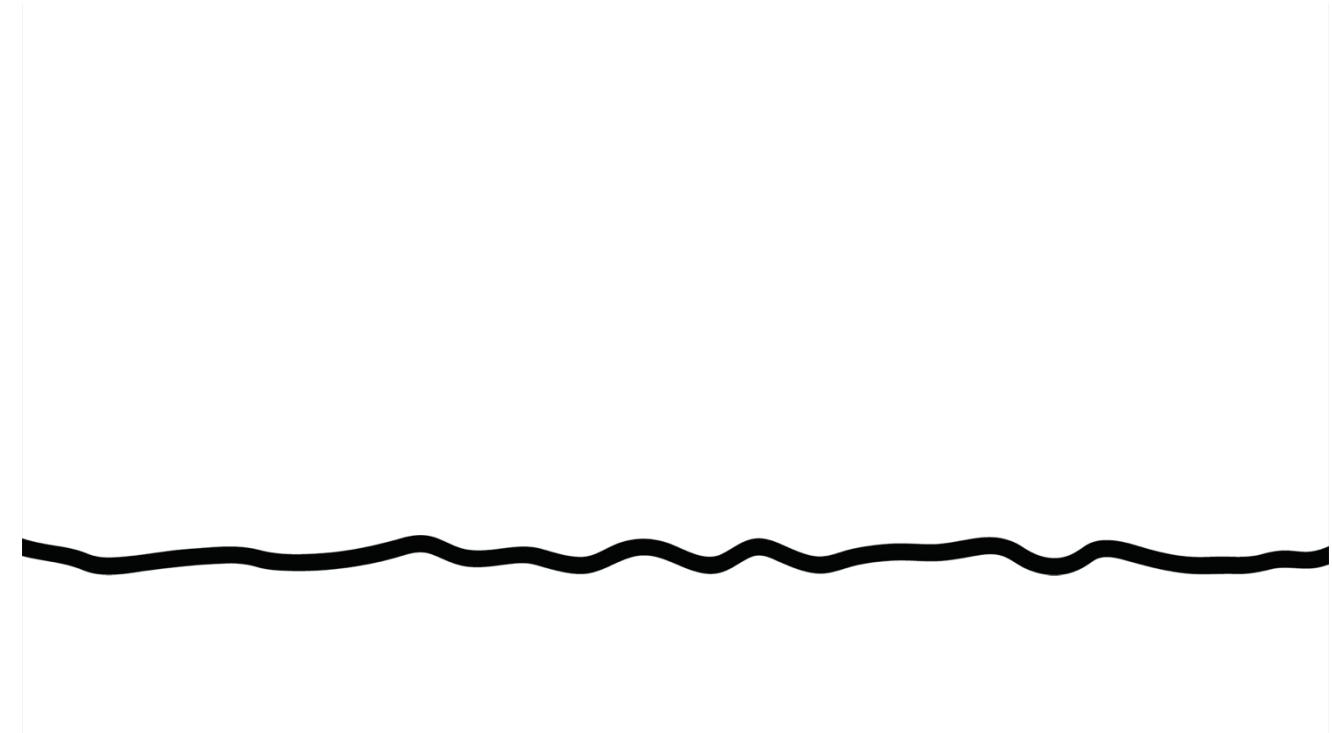
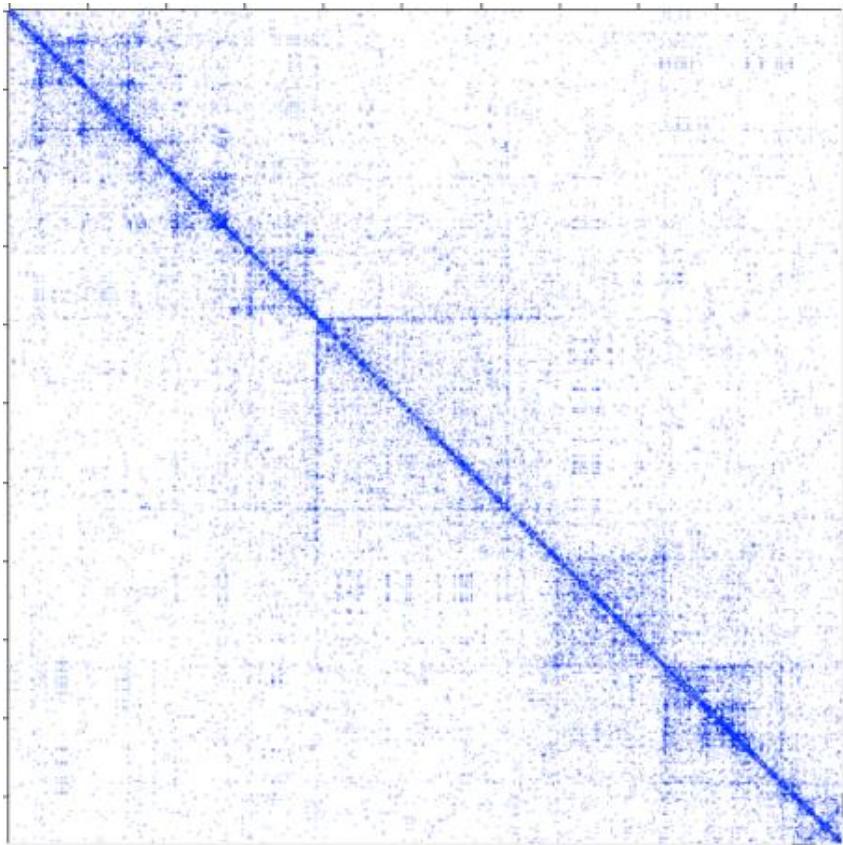


Loop extrusion model

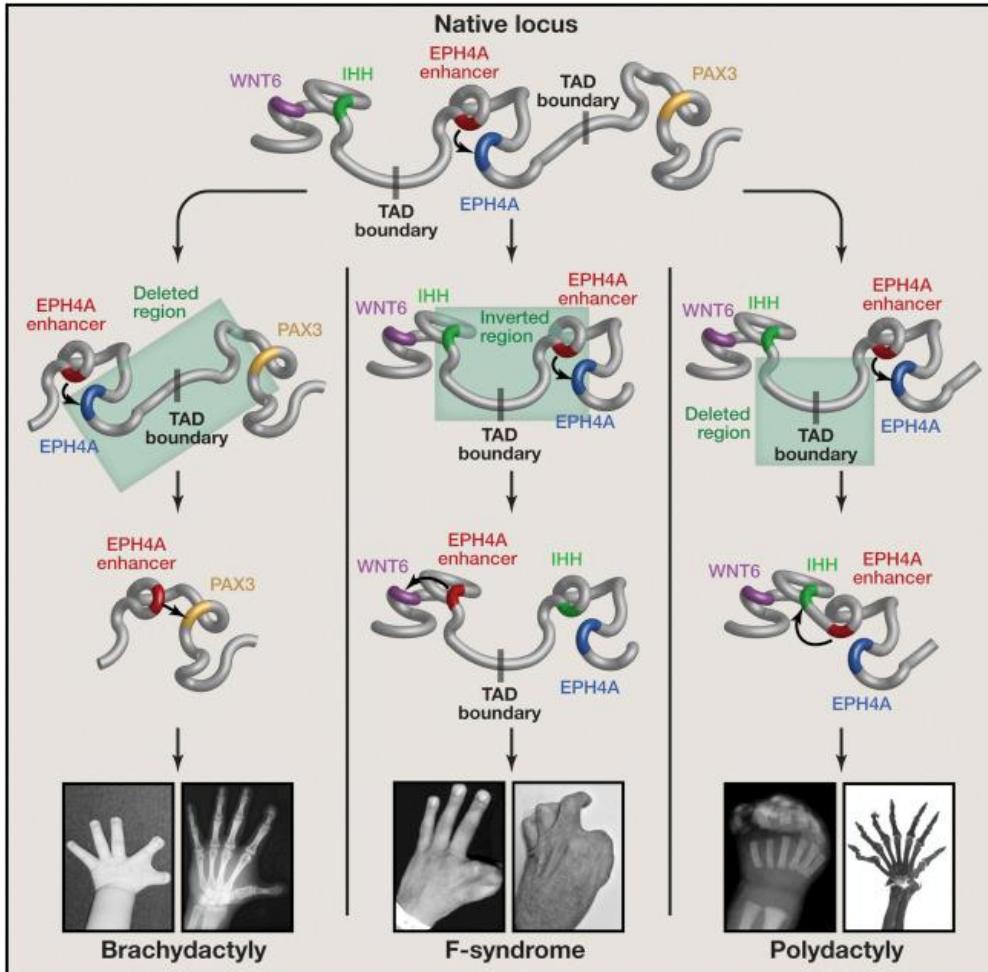


- <https://www.youtube.com/watch?v=Tn5qgEqWgW8>

One sided extrusion



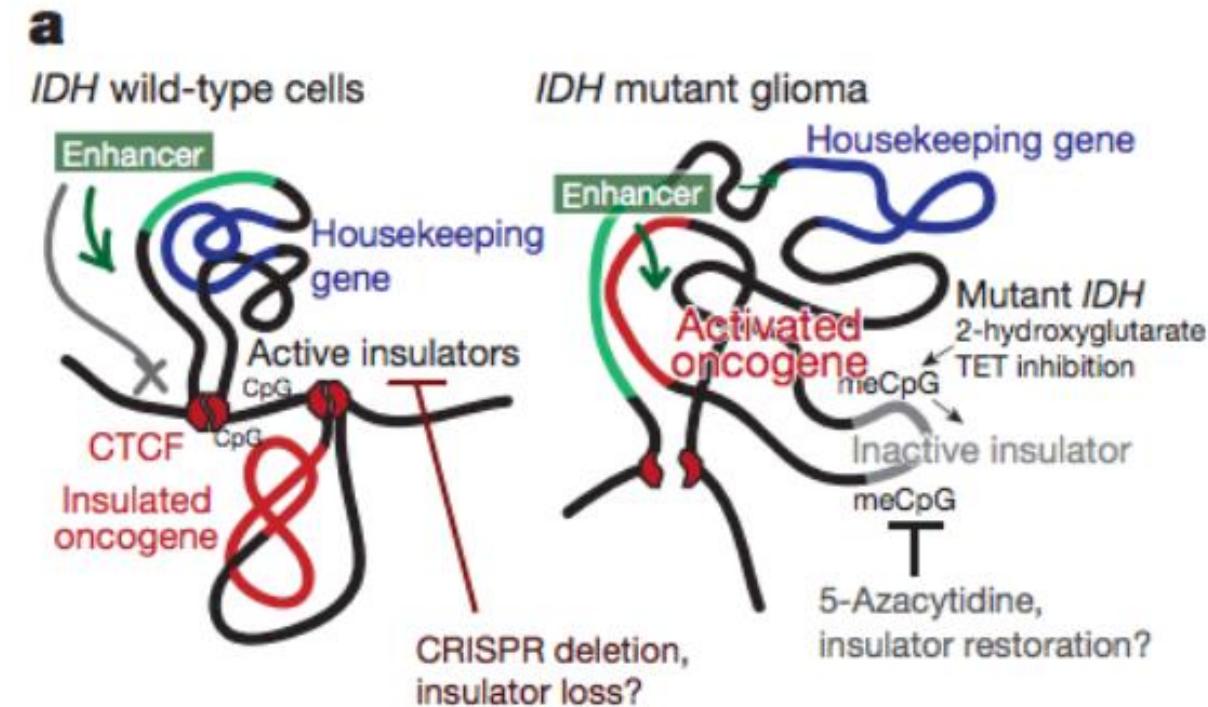
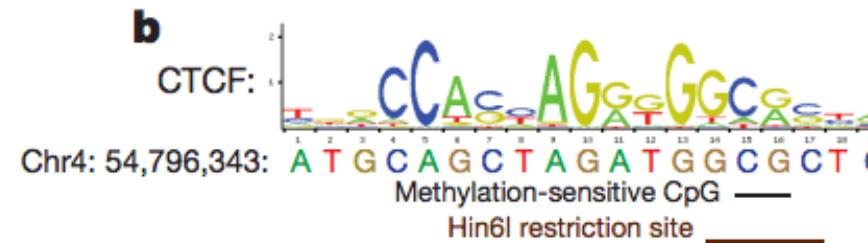
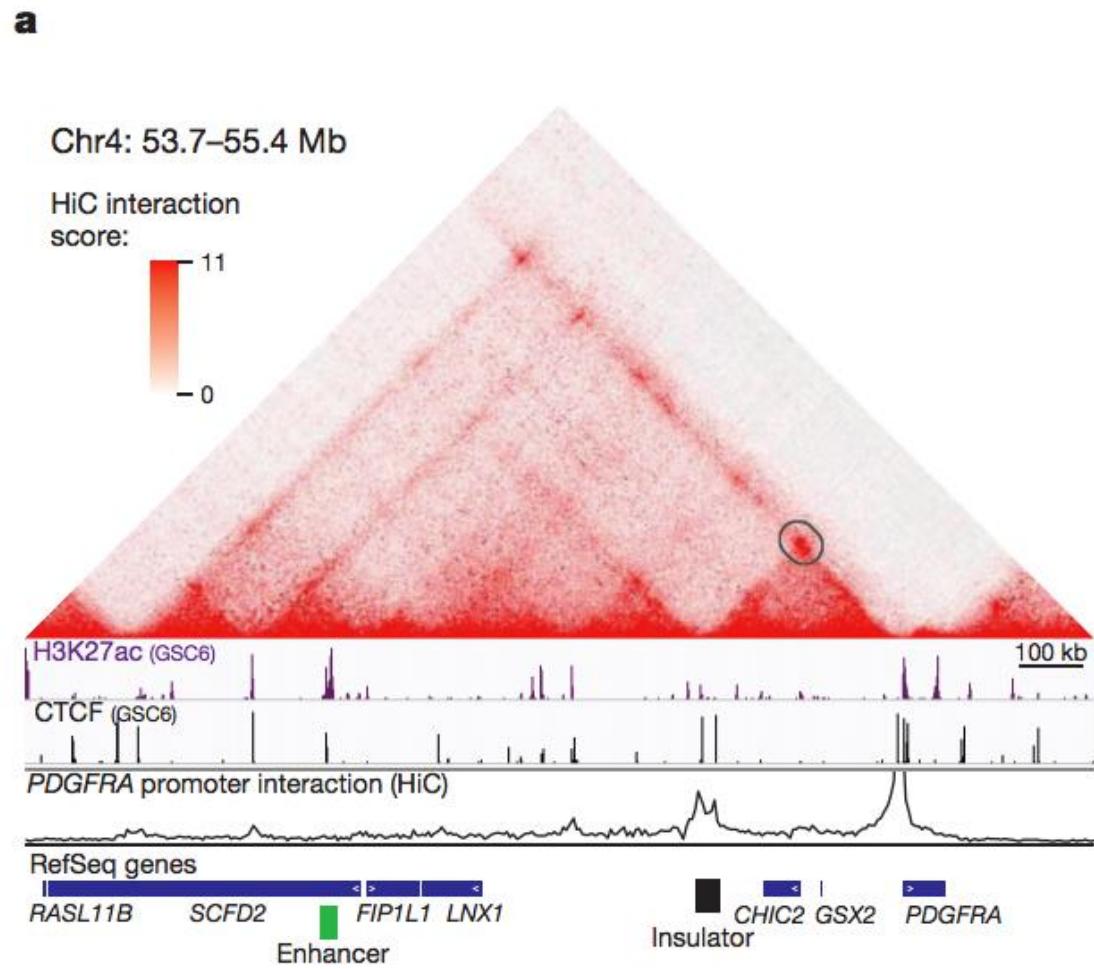
Breaking TADs: Role of chromatin topology in developmental disorders



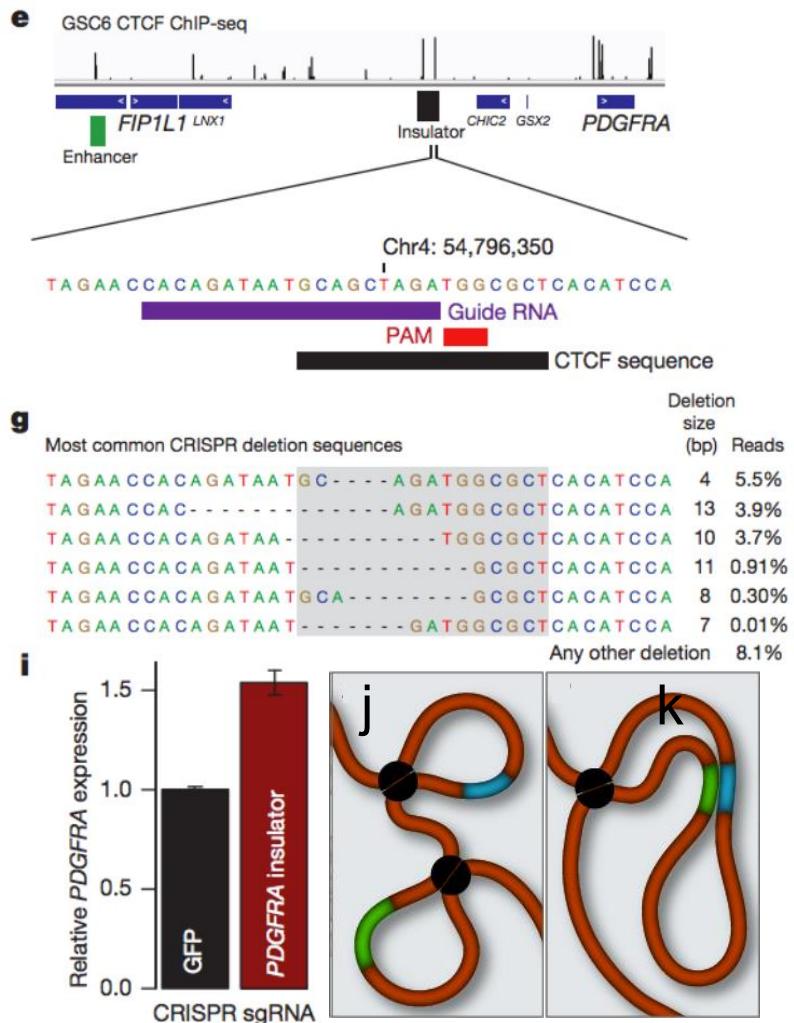
Disruption of chromatin organization by structural variation at a genetic locus containing genes and enhancers relevant to mammalian limb formation leads to pathological rewiring of genetic regulatory interactions resulting in three related human genetic disorders

Adapted from A CRISPR Connection between Chromatin Topology and Genetic Disorders. Ren B & Dixon JR. *Cell* (2015)
Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions. Lupiáñez DG et al. *Cell* (2015)

Insulator loss allows PDGFRA to interact with a constitutive enhancer in IDH gliomas



Boundary methylation and CTCF occupancy affect PDGFRA expression and proliferation



(E) CTCF binding shown for the FIP1L1/PDGFRα region. Expanded view shows CTCF motif in the insulator targeted for CRISPR-based deletion. sgRNA and PAM direct Cas9 nuclease to the motif

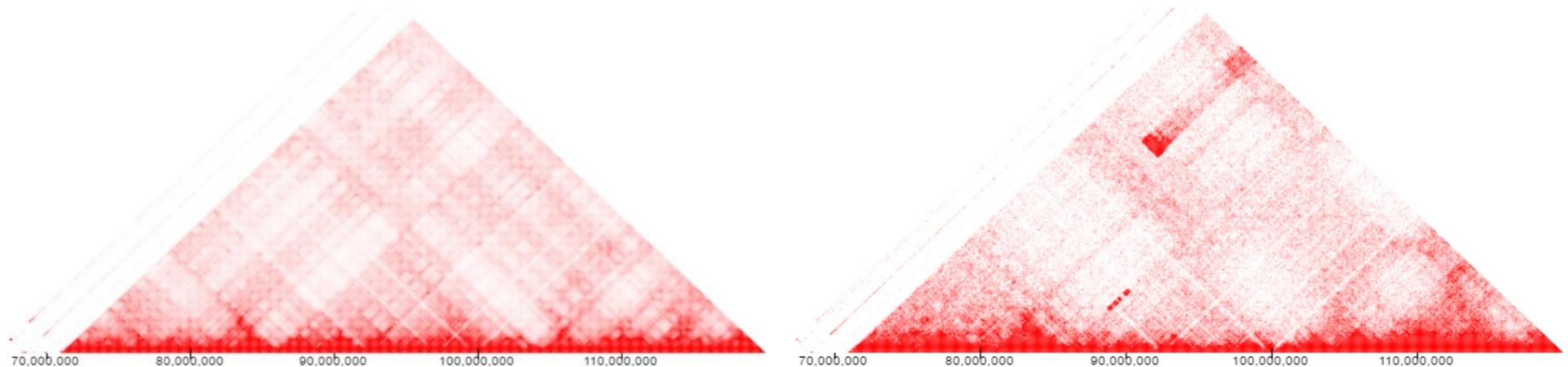
(G) Sequencing of target site reveals the indicated deletions. CTCF motif disrupted on ~25% of alleles (compare to <0.01% in control)

(I) qPCR reveals increased PDGFRA expression in insulator CRISPR cells.

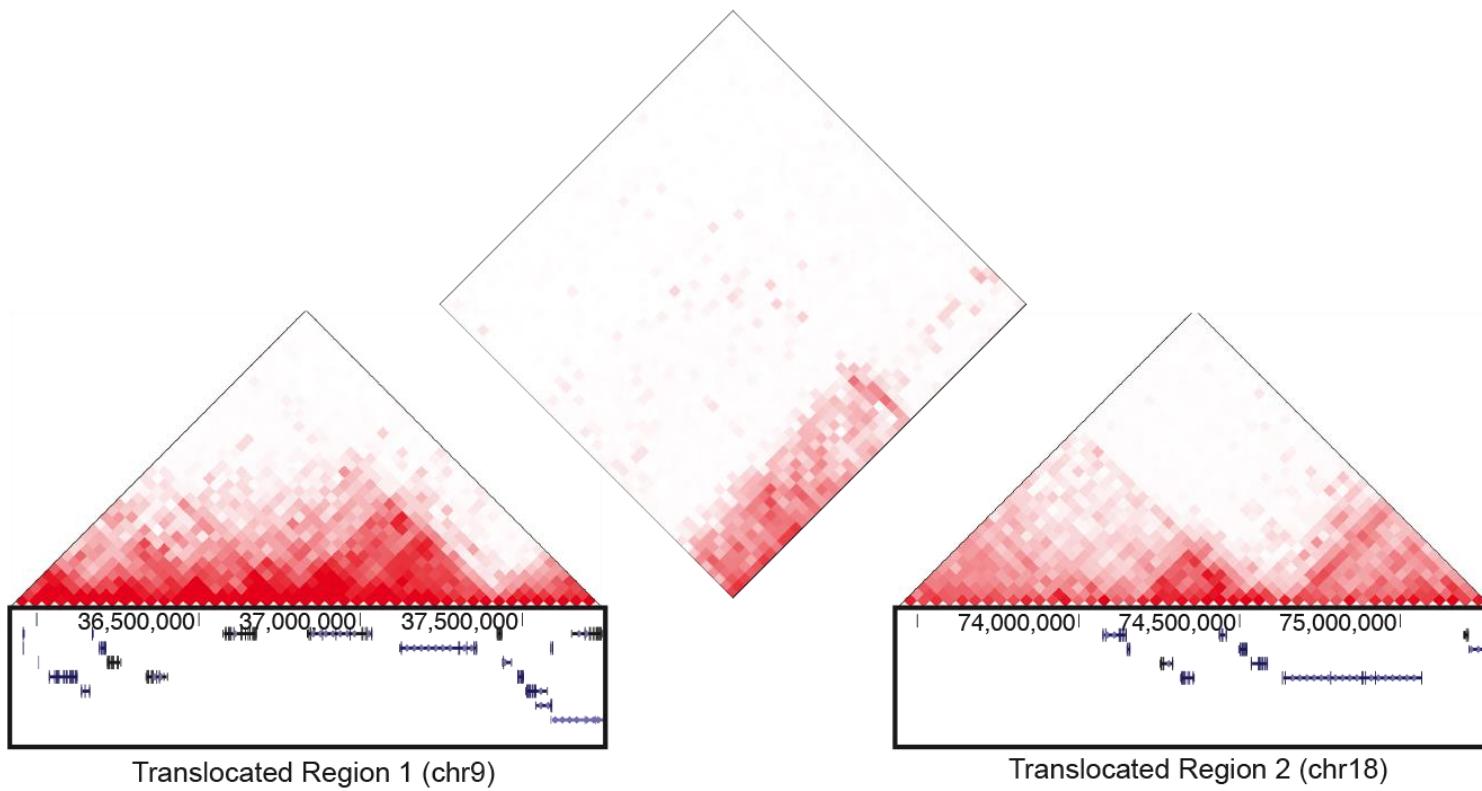
(J) PDGFRA and FIP1L1, which are normally confined to separate loop domains rarely interact

(K) But can become closely associated in IDH-mutant tumors

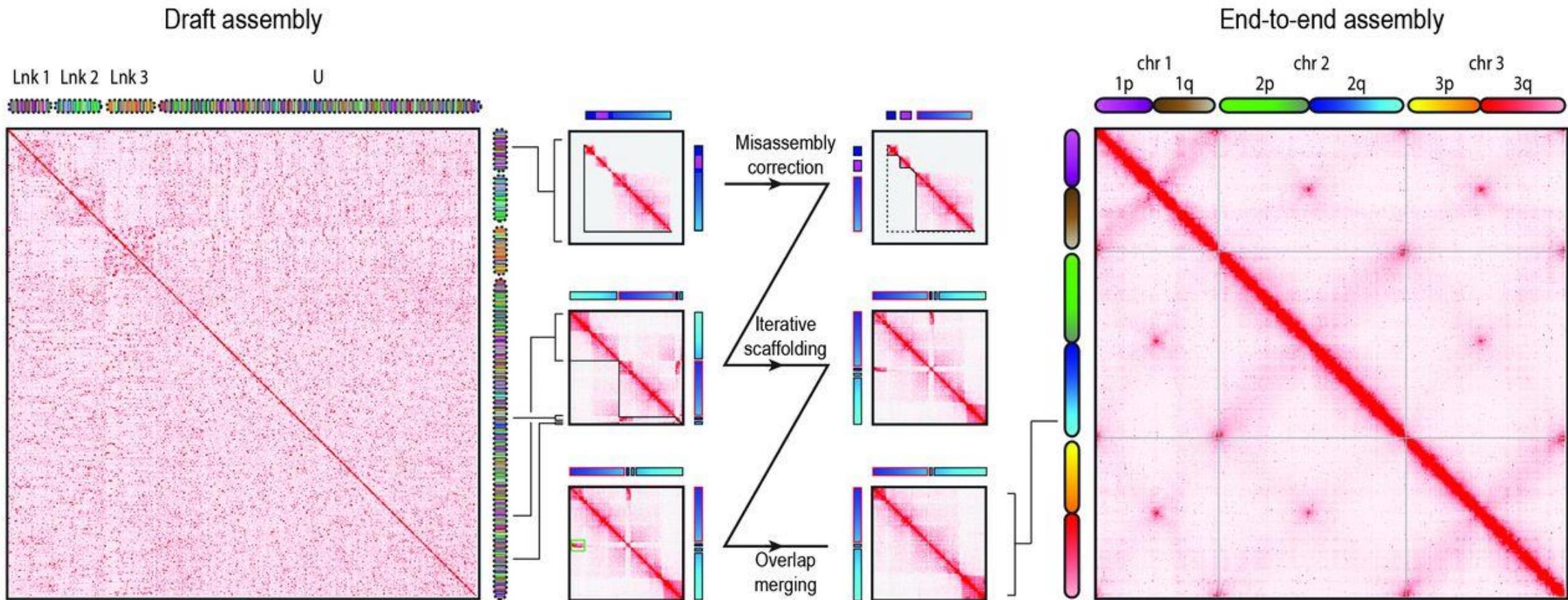
Formation of new TADs as a result of translocation



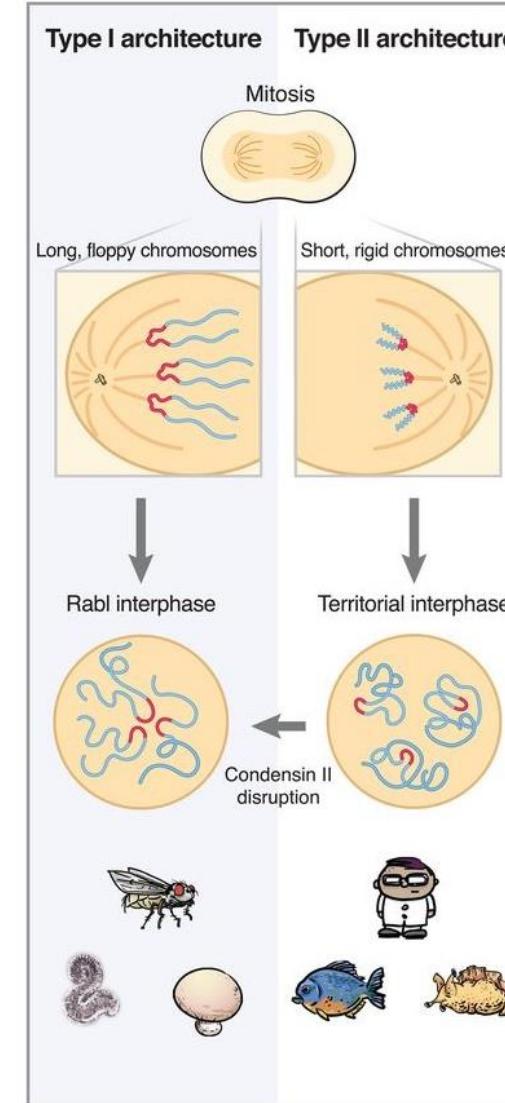
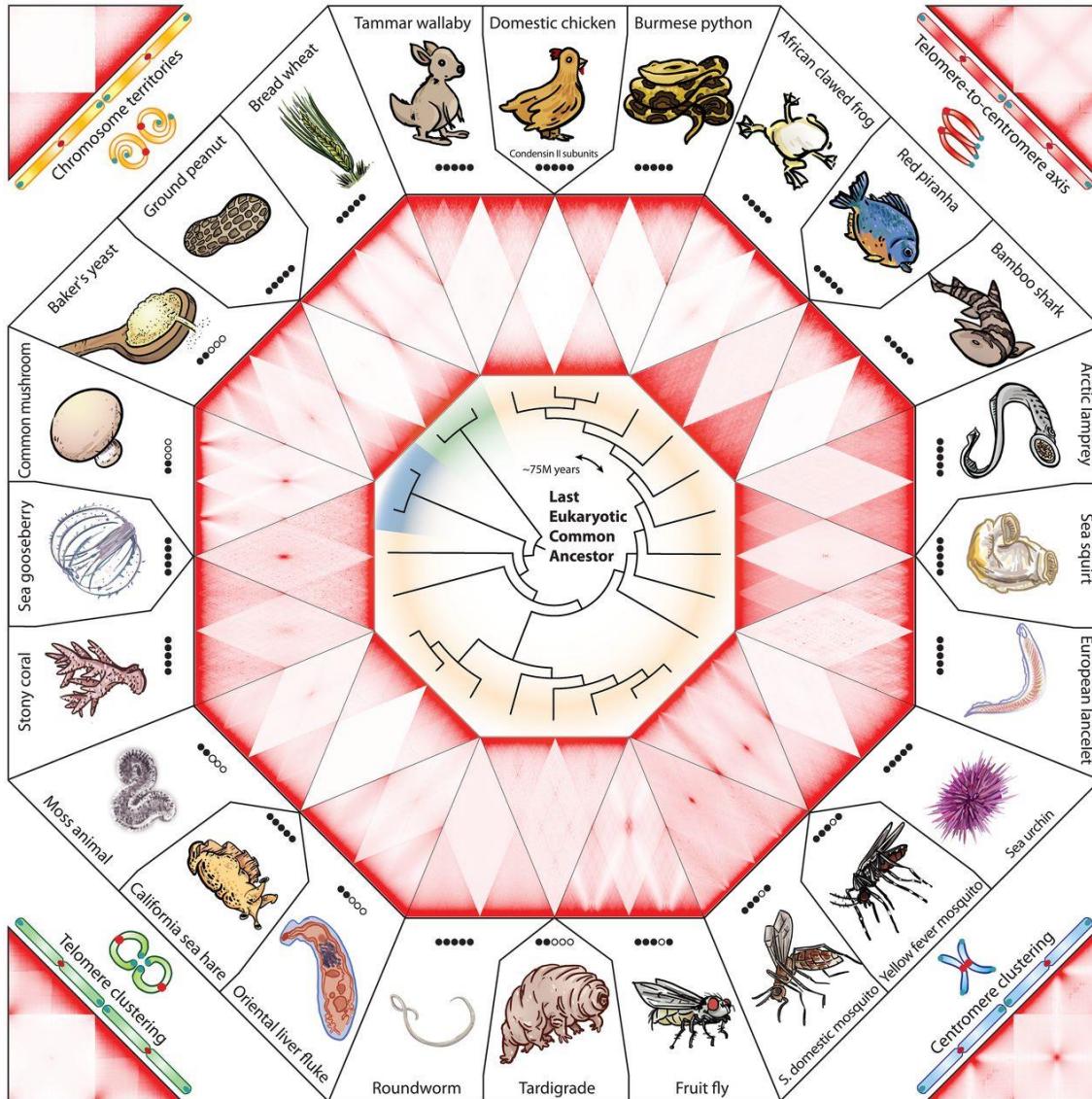
Formation of new TADs as a result of translocation



Hi-C contact maps can help with *de novo* genome assembly



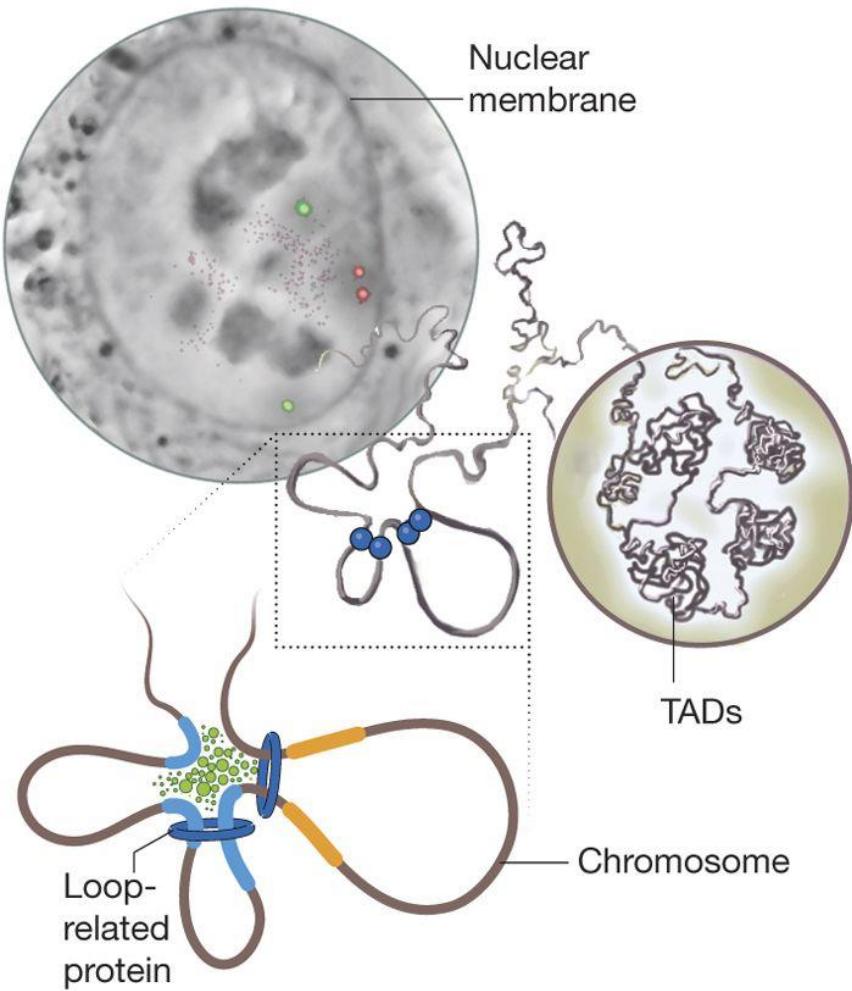
Chromatin architecture across the tree of life



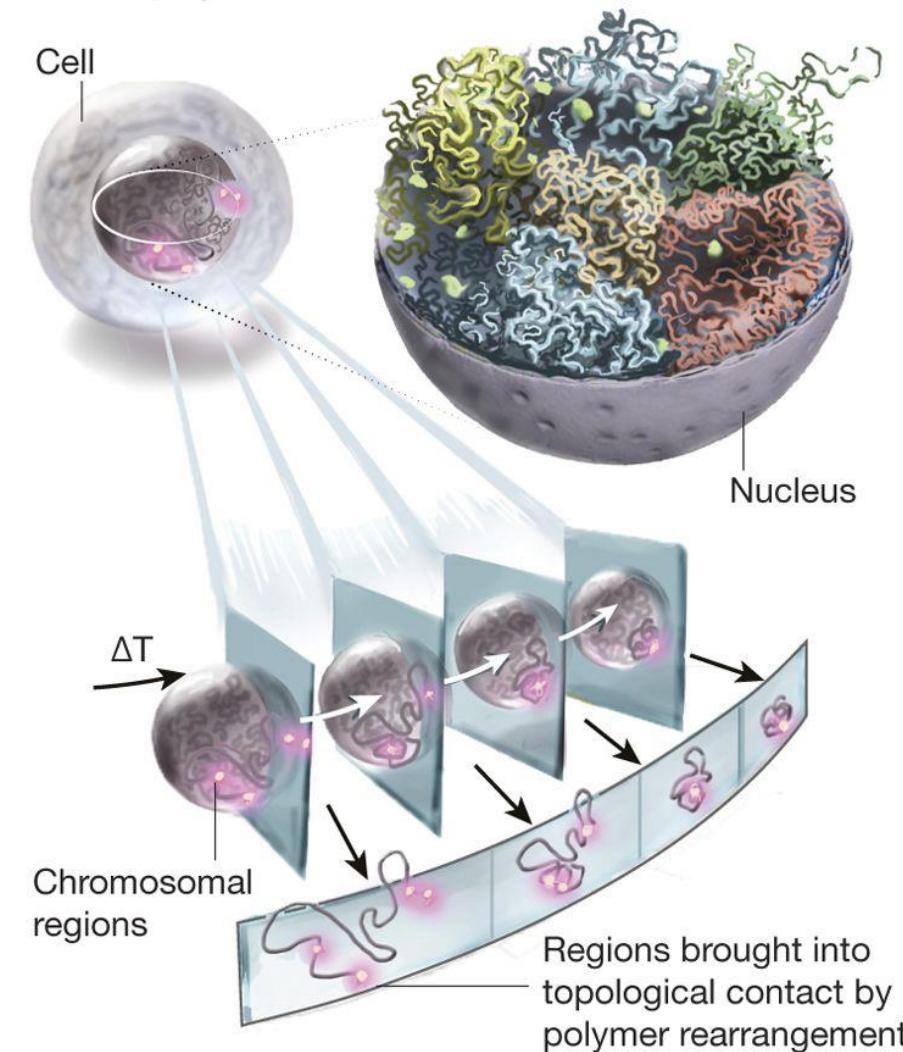
- Hi-C defines species-specific chromosome architectures: Rabl-like (T-to-T, C-to-C, C-to-T axis) or territorial
- Architecture either actively (Condensin II-defined territories) or passively (Rabl configuration) specified post-mitosis
- Condensin II knockout in territorial species results in Rabl-like configuration post-mitosis

4D Nucleome Project

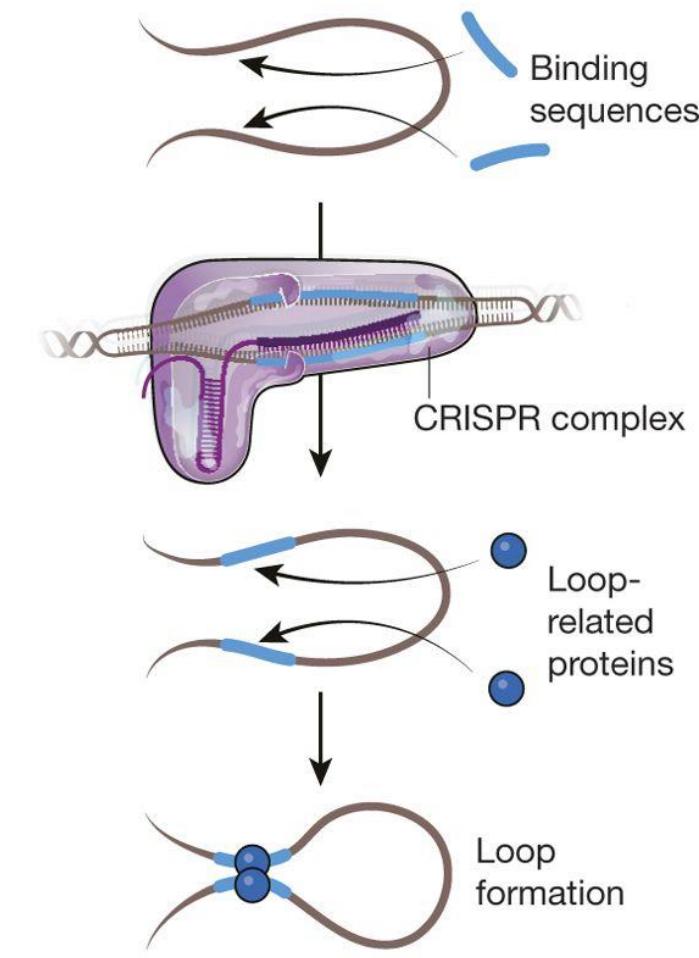
a **Mapping** – molecular genomic mapping of contacts, imaging of contacts and dynamics



b **Model building** – by constraint modelling and polymer simulation



c **Functional validation** – by genetic and biophysical perturbation experiments



Summary

- Hi-C analysis reveals that the mammalian genome is spatially compartmentalized, and consists of mega-base sized topological domains (also known as TADs).
- Topological domains have been independently observed in flies (Sexton et al. Cell 2012; Hou et al, Mol Cell 2012) and with different approaches (5C, Nora et al., Nature 2012).
- Topological domains are stable across cell types and largely preserved during evolution, suggesting that they are a basic property of the chromosome architecture.
- Partitioning of the genome into topological domains would naturally restrict the enhancers to selective promoters.
- Long-range looping interactions between enhancers and promoters correlate with higher transcriptional responsiveness of promoters.
- Cell specific enhancer/promoter interactions are formed in each cell type, some time prior to activation of the genes, and are not significantly altered by transient signaling induction.
- Pre-existing, lineage specific chromatin looping interactions between enhancers and promoters predict transcriptional responses to extracellular signaling, suggesting that chromatin conformation is another layer of transcriptional control.
- Hi-C types of assays interrogate population average!