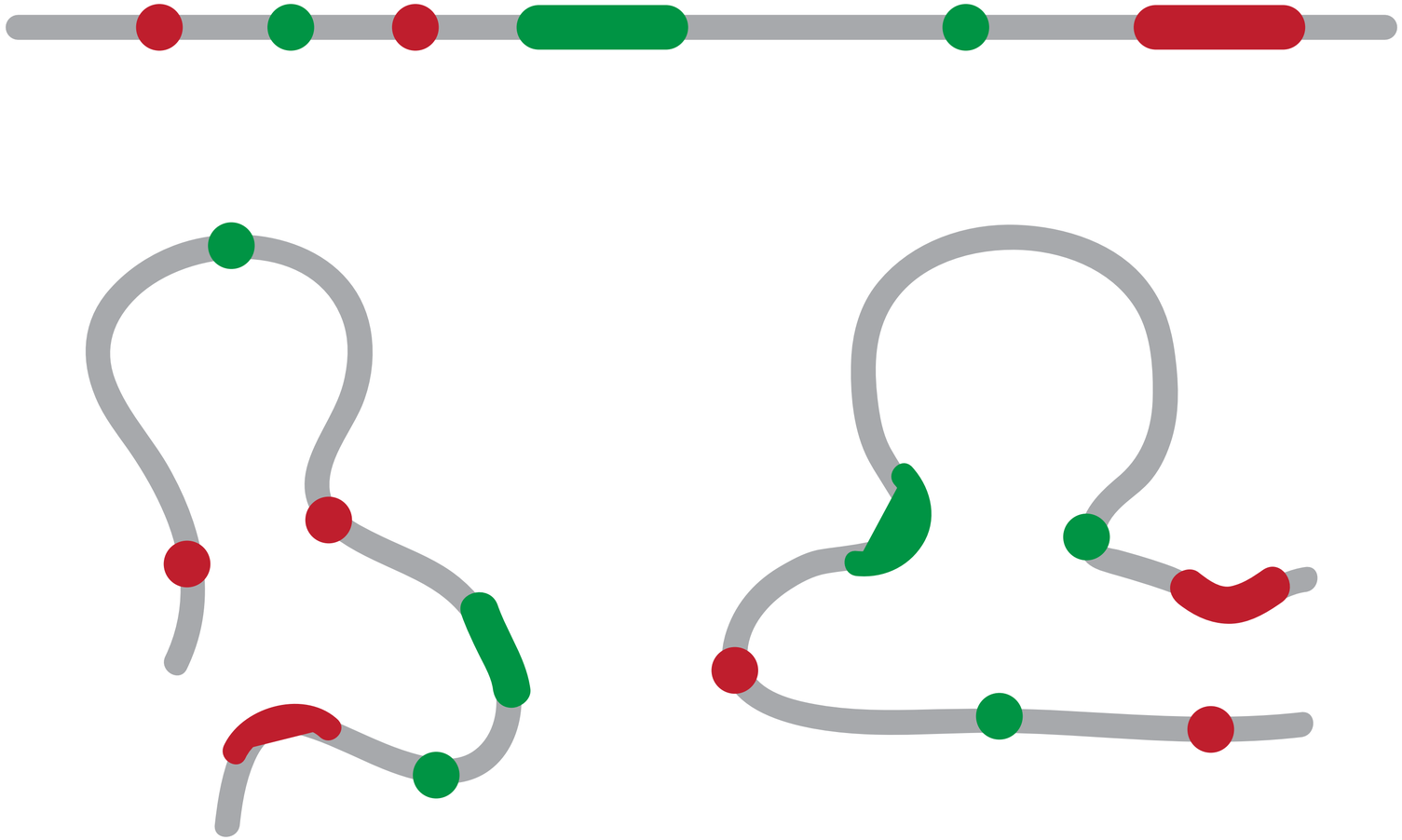
3D Genomics

Marc A. Marti-Renom

CNAG-CRG · ICREA

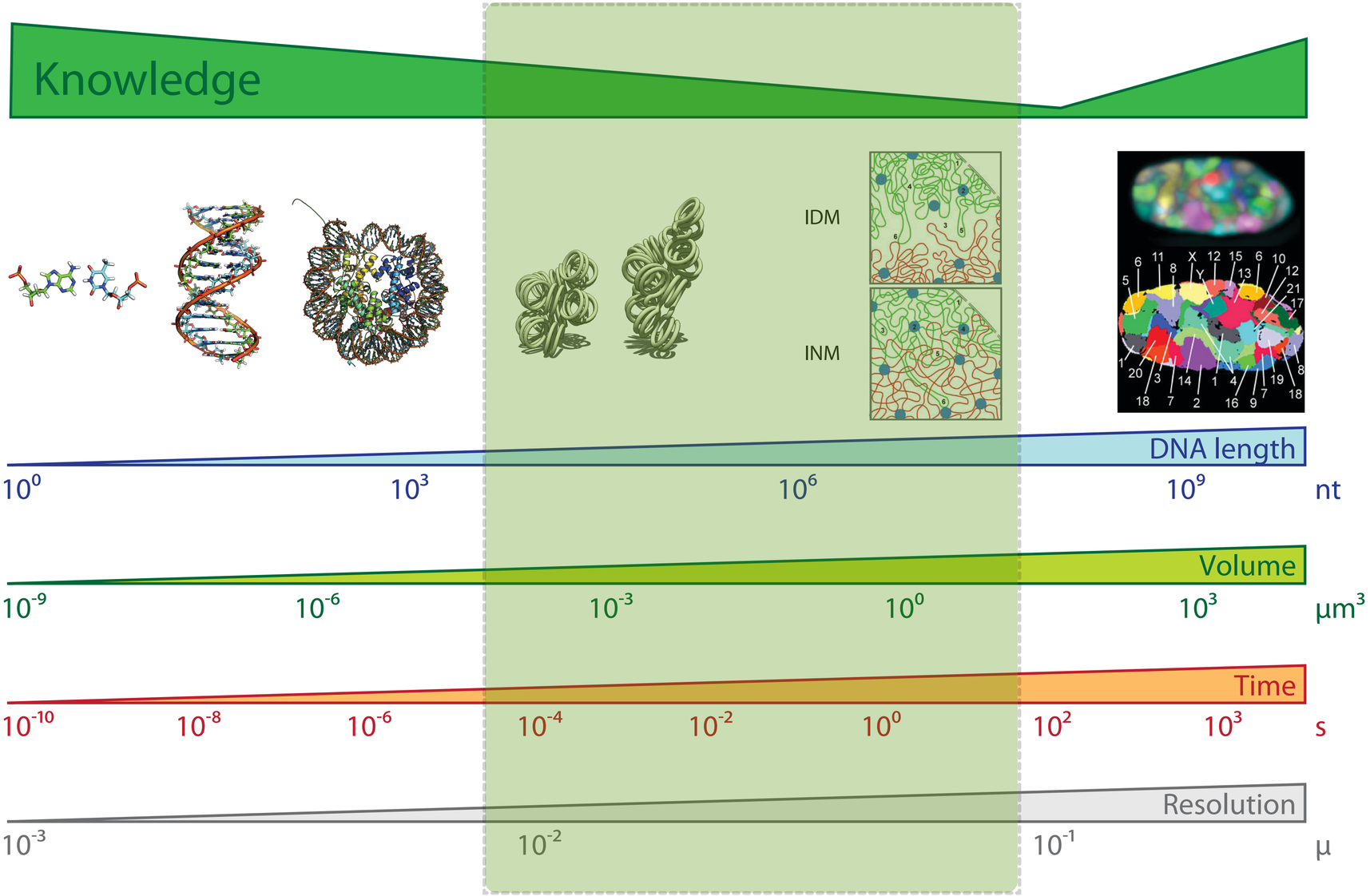
http://marciuslab.org
http://3DGenomes.org
http://cnag.crg.eu

cnag CRG ICREA



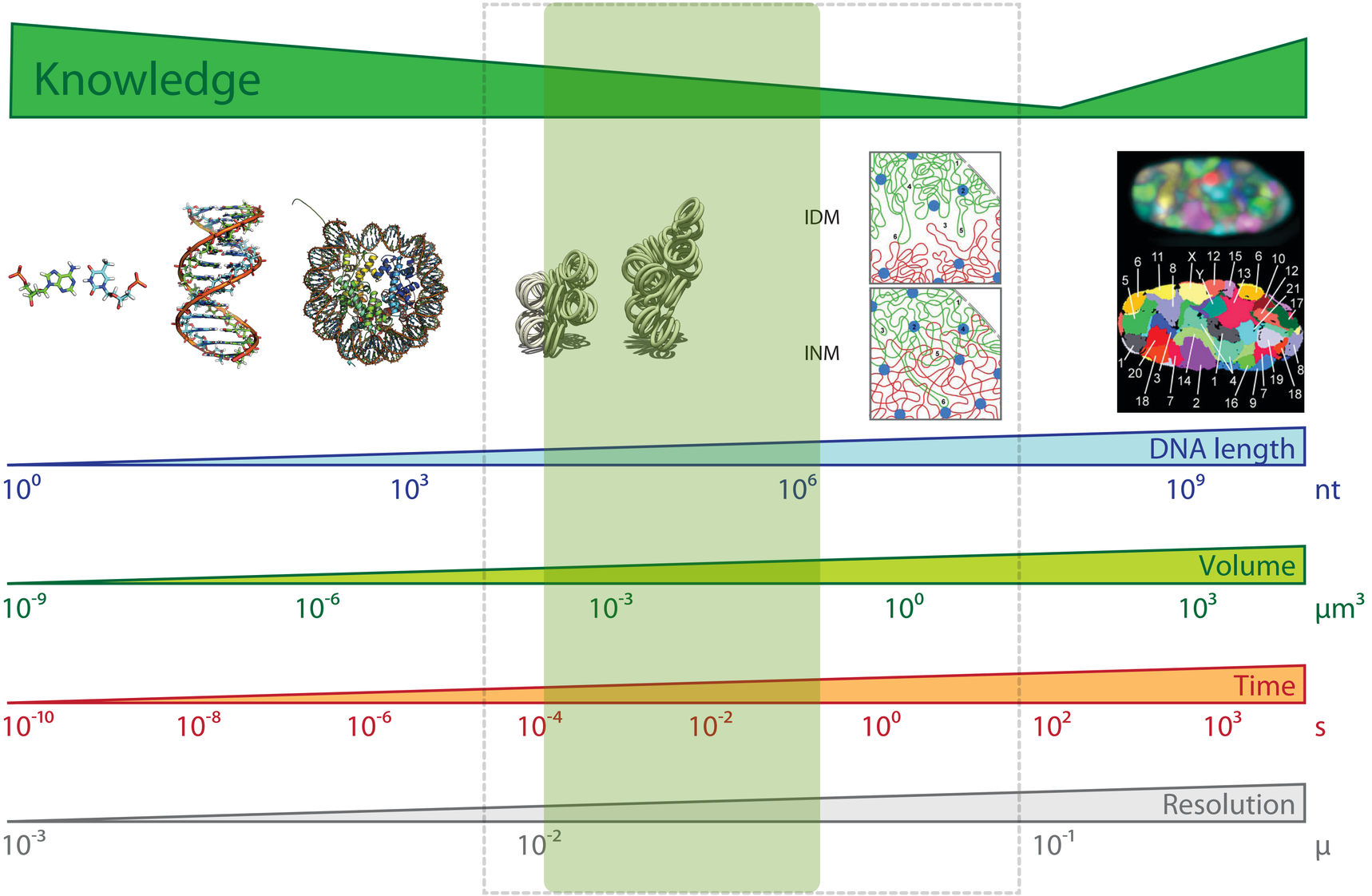
Resolution Gap

Marti-Renom, M. A. & Mirny, L. A. PLoS Comput Biol *7*, e1002125 (2011)



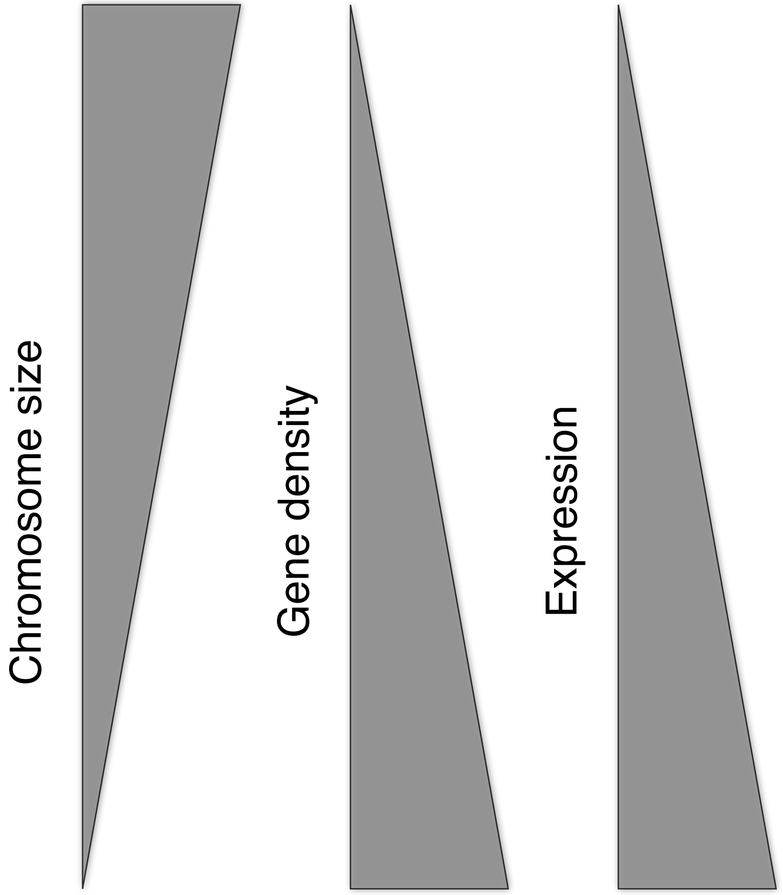
Resolution Gap

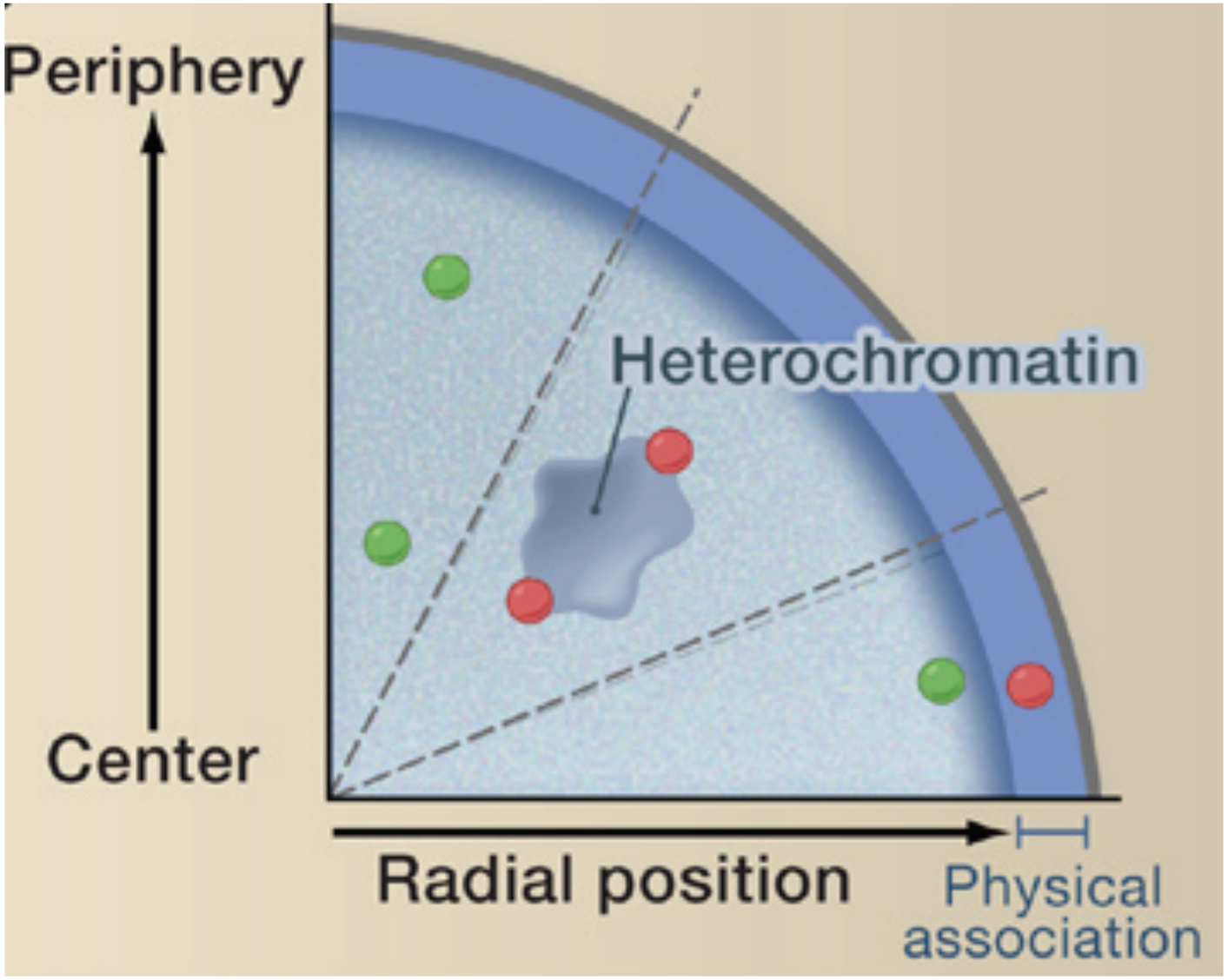
Marti-Renom, M. A. & Mirny, L. A. PLoS Comput Biol *7*, e1002125 (2011)



Level I: Radial genome organization

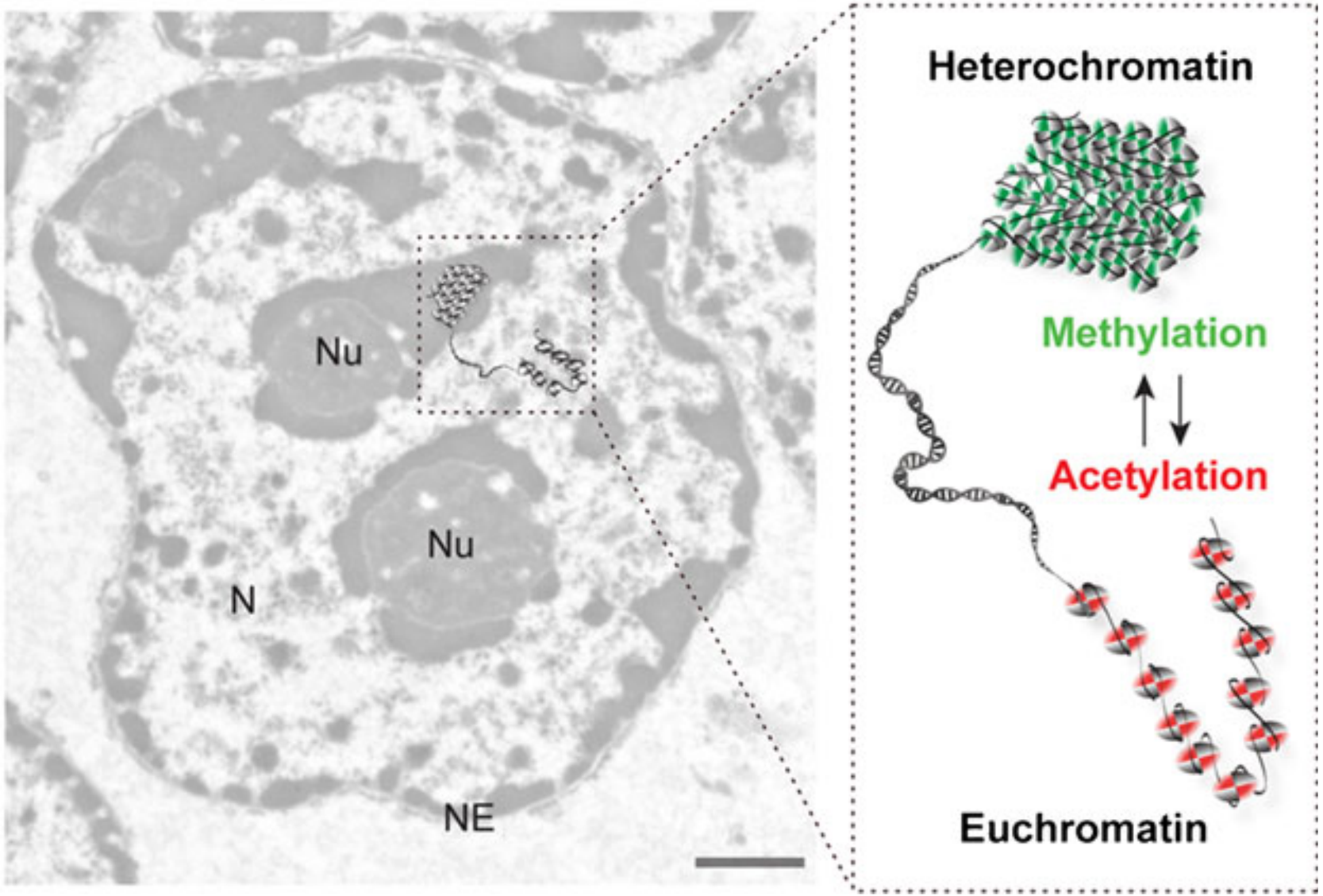
Takizawa, T., Meaburn, K. J. & Misteli, T. The meaning of gene positioning. Cell 135, 9–13 (2008).



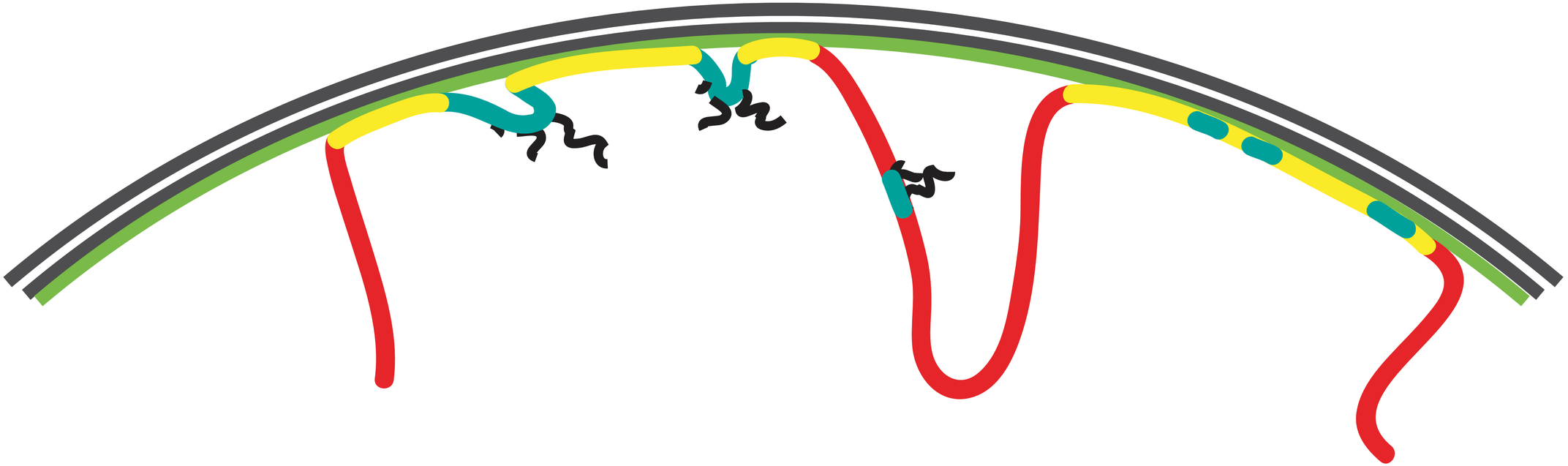


Level II: Euchromatin vs heterochromatin

Electron microscopy



Level III: Lamina-genome interactions



nuclear membrane
nuclear lamina

internal chromatin (mostly active)
lamina-associated domains (repressed)

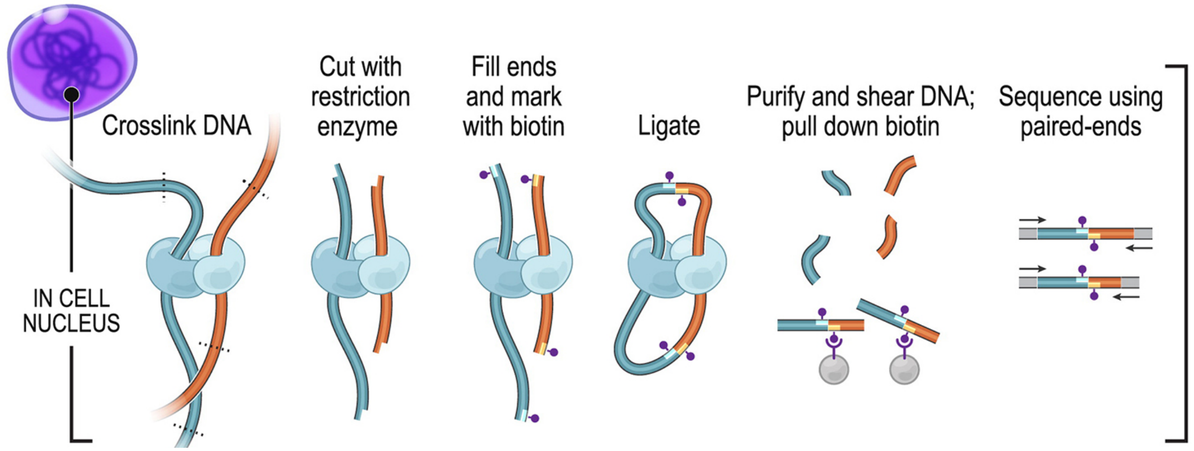
Genes

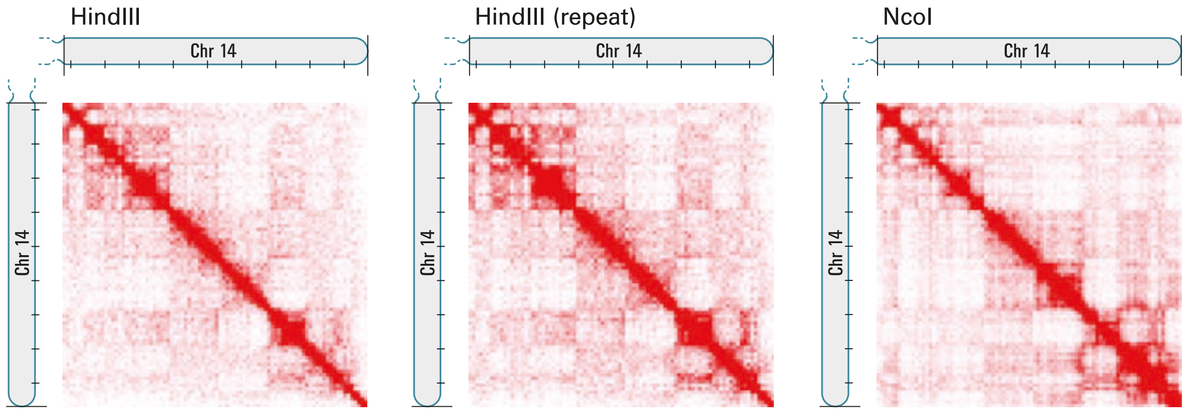
mRNA

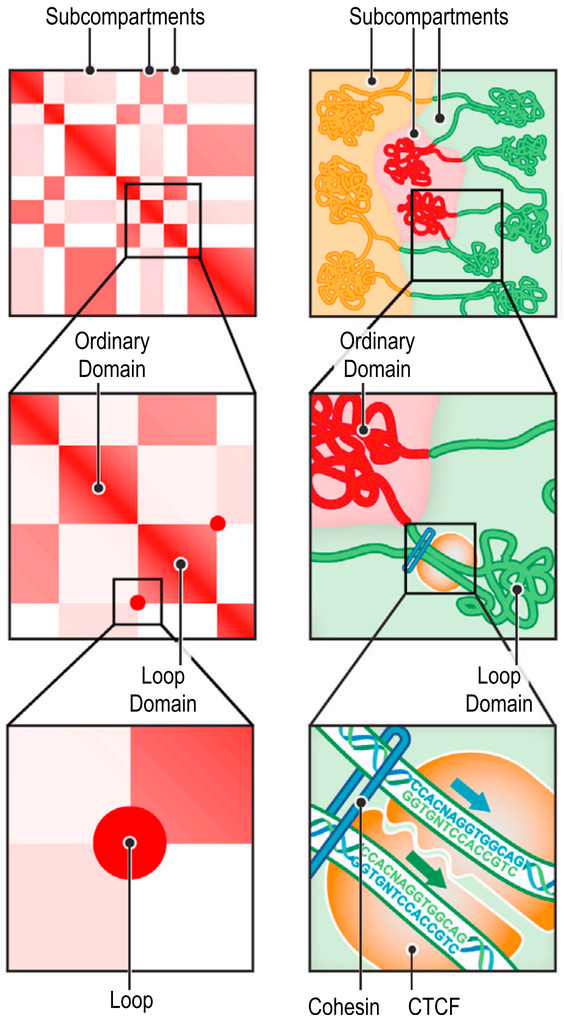
Adapted from Molecular Cell 38, 603-613, 2010

Level IV: Higher-order organization

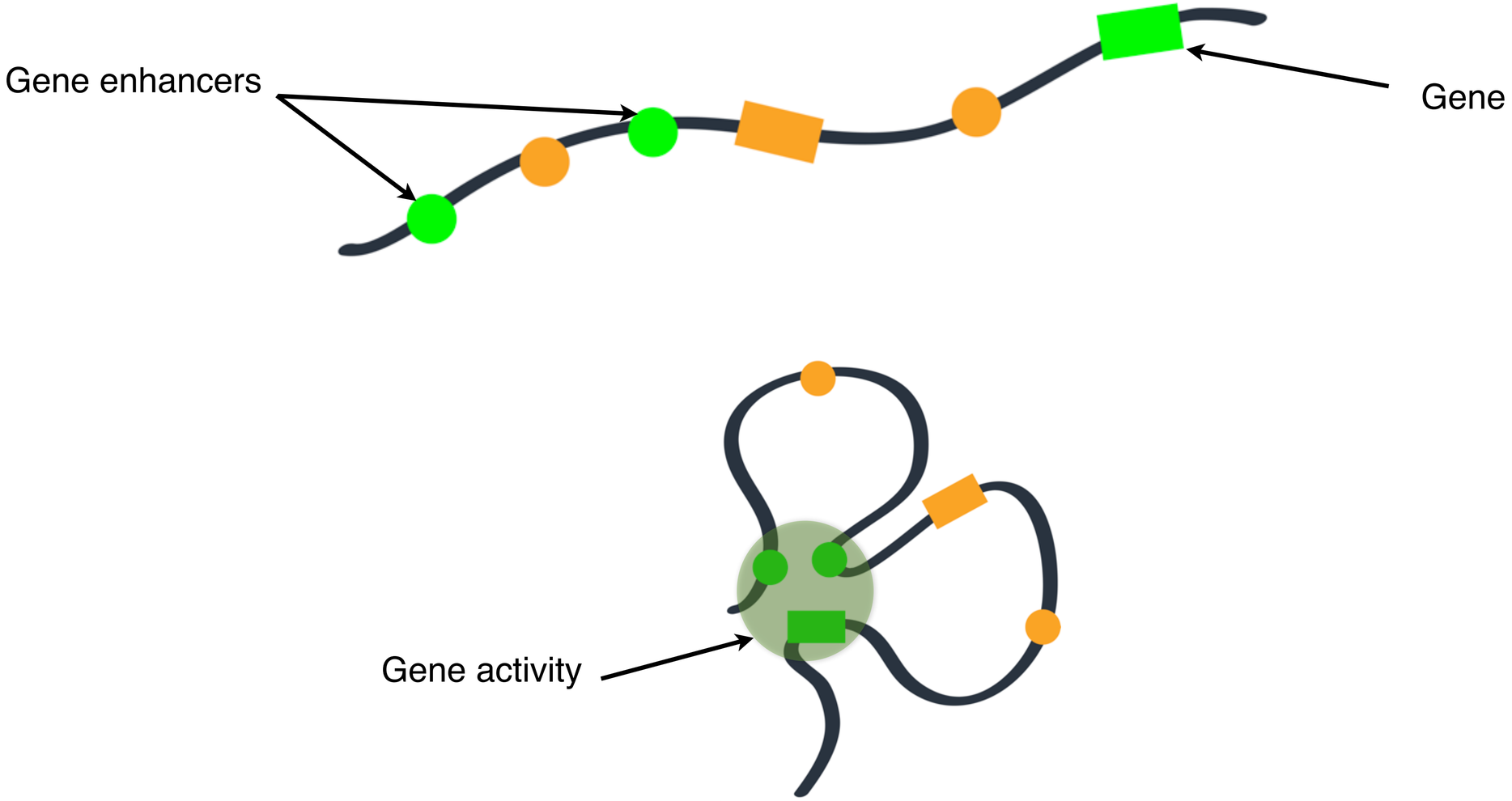
Lieberman-Aiden, E., et al. (2009). Science, 326(5950), 289–293.
Dekker, J., Marti-Renom, M. A. & Mirny, L. A. Nat Rev Genet 14, 390–403 (2013).
Rao, S. S. P., et al. (2014). Cell, 1–29.





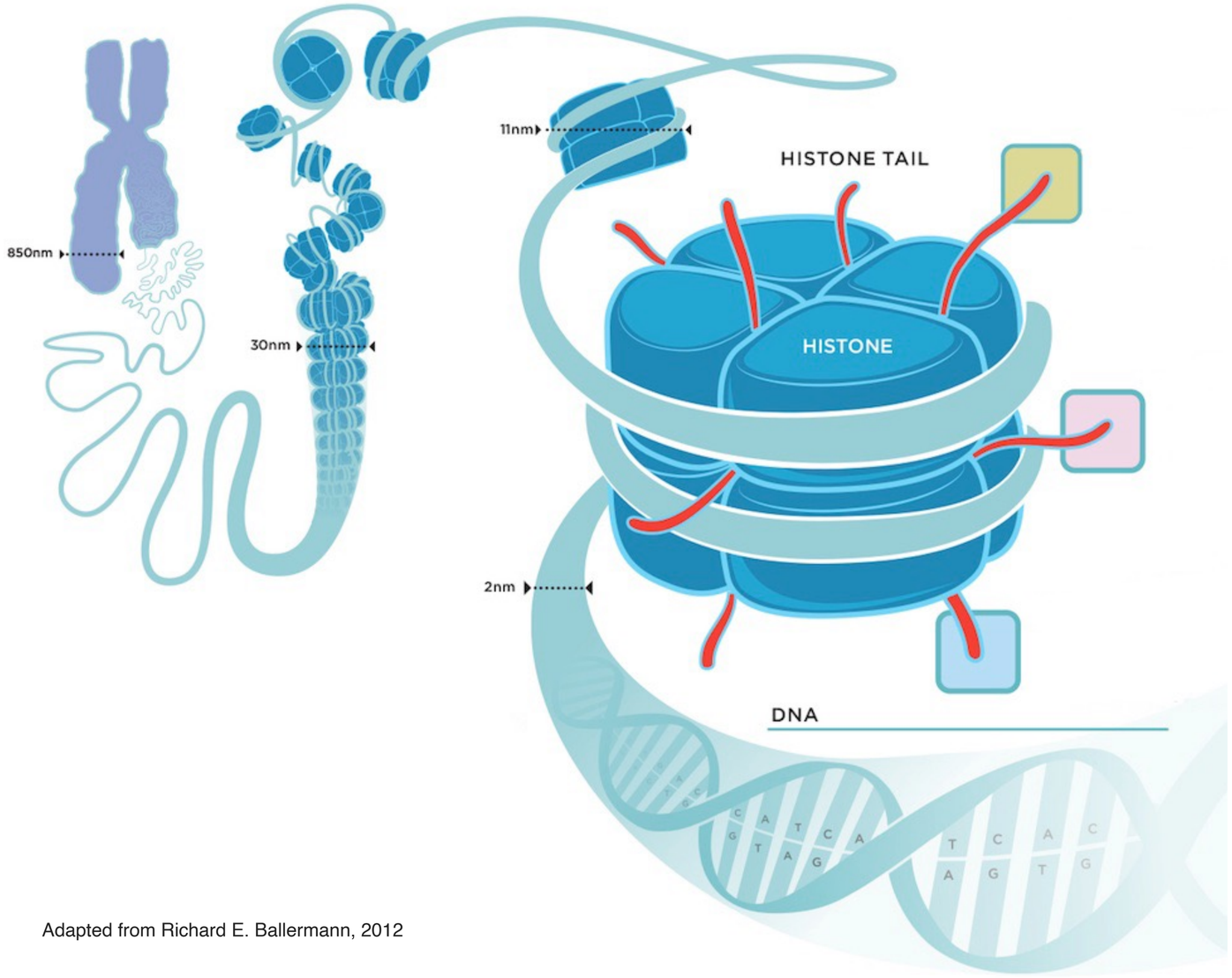


Level V: Chromatin loops



Level VI: Nucleosome

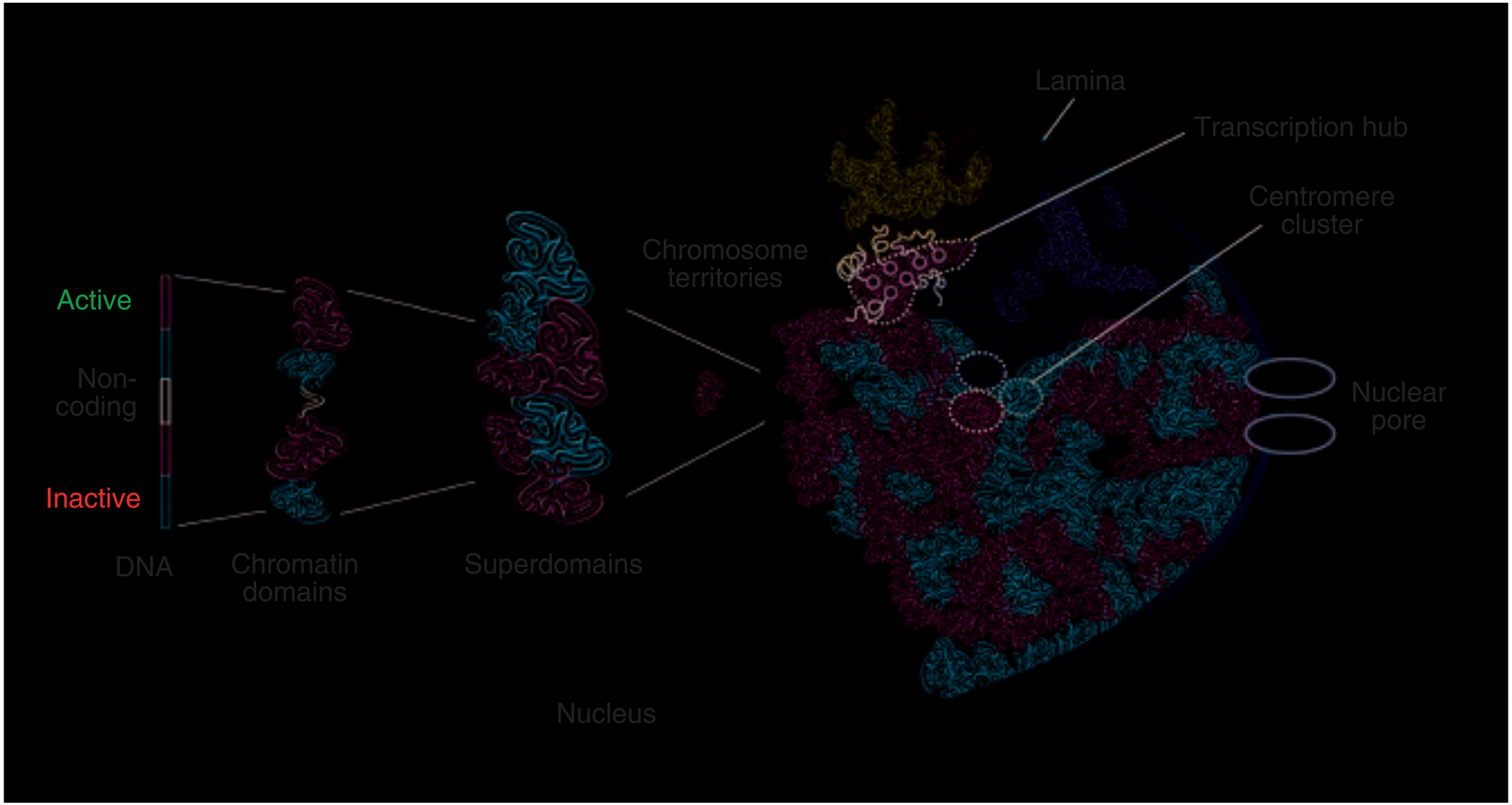
Chromosome **Chromatin fibre** **Nucleosome**



Adapted from Richard E. Ballermann, 2012

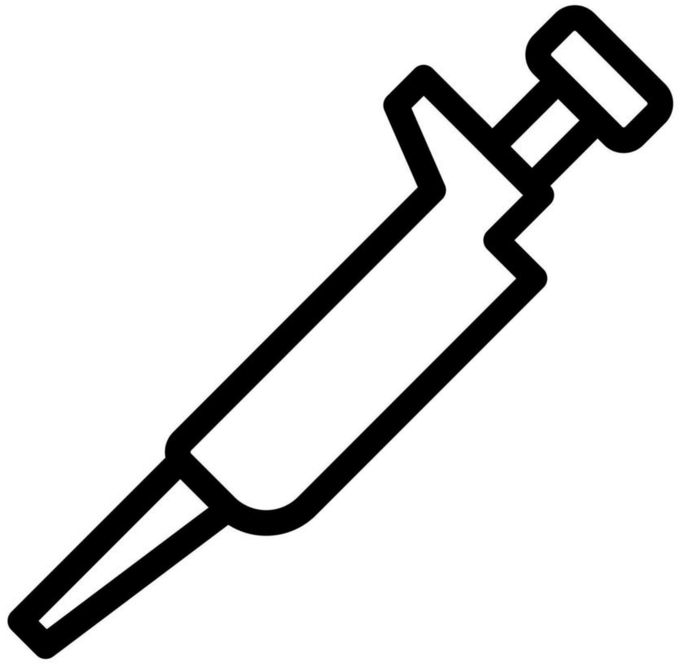
Complex genome organization

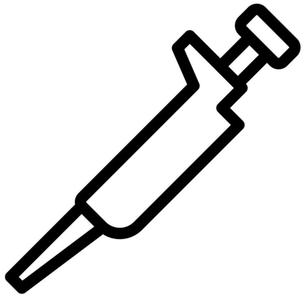
Cavalli, G. & Misteli, T. Functional implications of genome topology. Nat Struct Mol Biol 20, 290–299 (2013).



Marina Corral

Technologies





Hi-C 3.0

Akgol Oksuz, et al. Nature Methods 2021 & keep an eye on a possible soon paper for 4DNucleome

ANALYSIS
https://doi.org/10.1038/s41592-021-01248-7

nature methods

Check for updates

OPEN

Systematic evaluation of chromosome conformation capture assays

Betul Akgol Oksuz[1,10], Liyan Yang[1,10], Sameer Abraham[2], Sergey V. Venev[1], Nils Krietenstein[3], Krishna Mohan Parsi[4,5], Hakan Ozadam[1,6], Marlies E. Oomen[1], Ankita Nand[1], Hui Mao[4,5], Ryan M. J. Genga[4,5], Rene Maehr[4,5], Oliver J. Rando[3], Leonid A. Mirny[2,7,8], Johan H. Gibcus[1]✉ and Job Dekker[1,9]✉

Chromosome conformation capture (3C) assays are used to map chromatin interactions genome-wide. Chromatin interaction maps provide insights into the spatial organization of chromosomes and the mechanisms by which they fold. Hi-C and Micro-C are widely used 3C protocols that differ in key experimental parameters including cross-linking chemistry and chromatin fragmentation strategy. To understand how the choice of experimental protocol determines the ability to detect and quantify aspects of chromosome folding we have performed a systematic evaluation of 3C experimental parameters. We identified optimal protocol variants for either loop or compartment detection, optimizing fragment size and cross-linking chemistry. We used this knowledge to develop a greatly improved Hi-C protocol (Hi-C 3.0) that can detect both loops and compartments relatively effectively. In addition to providing benchmarked protocols, this work produced ultra-deep chromatin interaction maps using Micro-C, conventional Hi-C and Hi-C 3.0 for key cell lines used by the 4D Nucleome project.

Chromosome conformation capture (3C)-based assays[1] have become widely used to generate genome-wide chromatin interaction maps[2]. Analysis of chromatin interaction maps has led to detection of several features of the folded genome. Such features include precise looping interactions (at the 0.1–1 Mb scale) between pairs of specific sites that appear as local dots in interaction maps. Many of such dots represent loops formed by cohesin-mediated loop extrusion that is stalled at convergent CCCTC-binding factor (CTCF) sites[3–5]. Loop extrusion also produces other features in interaction maps such as stripe-like patterns anchored at specific sites that block loop extrusion. The effective depletion of interactions across such blocking sites leads to domain boundaries (insulation). At the megabase scale, interaction maps of many organisms including mammals display checkerboard patterns that represent the spatial compartmentalization of two main types of chromatin: active and open A-type chromatin domains, and inactive and more closed B-type chromatin domains[6].

The Hi-C protocol has evolved over the years. While initial protocols used restriction enzymes such as HindIII that produces relatively large fragments of several kilobases[6], over the last 5 years Hi-C using DpnII or MboI digestion has become the protocol of choice for mapping chromatin interactions at kilobase resolution[3]. More recently, Micro-C, which uses MNase instead of restriction enzymes as well as a different cross-linking protocol, was shown to allow generation of nucleosome-level interaction maps[7–9]. It is critical to ascertain how key parameters of these 3C-based methods, including cross-linking and chromatin fragmentation, quantitatively influence the detection of chromatin interaction frequencies and the detection of different chromosome folding features that range from local looping between small intra-chromosomal (cis) elements to global compartmentalization of megabase-sized domains. Here, we systematically assessed how different cross-linking and fragmentation methods yield quantitatively different chromatin interaction maps.

Results

We explored how two key parameters of 3C-based protocols, cross-linking and chromatin fragmentation, determine the ability to quantitatively detect chromatin compartment domains and loops. We selected three cross-linkers widely used for chromatin: 1% formaldehyde (FA), conventional for most 3C-based protocols; 1% FA followed by incubation with 3 mM disuccinimidyl glutarate (the FA + DSG protocol); and 1% FA followed by incubation with 3 mM ethylene glycol bis(succinimidylsuccinate) (the FA + EGS protocol) (Fig. 1a). We selected four different nucleases for chromatin fragmentation: MNase, DdeI, DpnII and HindIII, which fragment chromatin in sizes ranging from single nucleosomes to multiple kilobases. Combined, the three cross-linking and four fragmentation strategies yield a matrix of 12 distinct protocols (Fig. 1b). To determine how performance of these protocols varies for different states of chromatin we applied this matrix of protocols to multiple cell types and cell cycle stages. We analyzed four different cell types: pluripotent H1 human embryonic stem cells (H1-hESCs), differentiated endoderm (DE) cells derived from H1-hESCs, fully differentiated human foreskin fibroblast (HFF) cells (12 protocols for each), and HeLa-S3 cells (9 protocols). We analyzed two cell cycle stages: G1 and mitosis, in HeLa-S3 cells (9 protocols for each; Fig. 1). Each interaction library was then sequenced on a single lane of a HiSeq4000 instrument, producing ~150–200 million uniquely mapping read pairs (Supplementary Table 1). We used the Distiller pipeline to align the sequencing reads, and pairtools and cooler[10] packages to process mapped reads and create multi-resolution contact maps (Methods). Given that the density of restriction sites for DdeI, DpnII and HindIII fluctuates along chromosomes, we observed different read coverages in raw interaction maps obtained from datasets using these enzymes (Extended Data Fig. 1h). These differences were removed after matrix balancing[11].

We first assessed the size range of the chromatin fragments produced after digestion by the 12 protocols for HFF cells (Methods). Digestion with HindIII resulted in 5–20-kb DNA fragments; DpnII and DdeI produced fragments of 0.5–5 kb; and MNase protocols included a size selection step to ensure that the ligation product involved two mononucleosome-sized fragments (~150 bp) (Extended Data Fig. 1). Different cross-linkers did not affect the size ranges produced by the different nucleases, although DSG cross-linking lowered digestion efficiency slightly (Extended Data Fig. 1b).

All 3C-based protocols can differentiate between cell states. We first assessed the similarity between the 63 datasets by global and pairwise correlations using HiCRep and hierarchical clustering (Extended Data Fig. 1c)[12,13]. We found that the datasets are highly correlated and cluster primarily by cell type and state and then by cell type similarity, for example H1-hESCs and H1-hESC-derived DE cells cluster together; and the most distinct cluster is formed by mitotic HeLa cells. MNase protocols show slightly lower correlations with Hi-C experiments.

Extra cross-linking yields more intra-chromosomal contacts. Given that chromosomes occupy individual territories, intra-chromosomal (cis) interactions are more frequent than inter-chromosomal (trans) interactions[14]. The cis:trans ratio is commonly used as an indicator of Hi-C library quality given that inter-chromosomal interactions are a mixture of true chromatin interactions and interactions that are the result of random ligations[14,15]. For all enzymes and cell types, we found that the addition of DSG or EGS to FA cross-linking decreased the percentage of trans interactions (Fig. 2a for HFF and Extended Data Fig. 2a for H1-hESC, DE, HeLa-S3).

Regarding intra-chromosomal interactions, we noticed two distinct patterns. First, digestion into smaller fragments increased short-range interactions. MNase digestion generated more interactions between loci separated by less than 10 kb, whereas digestion with either DdeI, DpnII or HindIII resulted in a relatively larger number of interactions between loci separated by more than 10 kb (Fig. 2a,b for HFF and Extended Data Fig. 2a,b for DE, H1-hESC, HeLa-S3). Second, $P(s)$ plots showed that the addition of either DSG or EGS resulted in a steeper decay in interaction frequency as a function of genomic distance for all fragmentation protocols. Moreover, for a given chromatin fragmentation level, additional cross-linking with DSG or EGS reduced trans interactions, as shown for HFF cells and all other cell types and cell stages studied (Fig. 2c,d and Extended Data Fig. 2c). The addition of DSG or EGS could have reduced fragment mobility and the formation of spurious ligations, resulting in a steeper slope of the $P(s)$. We note a difference in slopes for data obtained with different cell types and cell cycle stages, which could reflect state-dependent differences in chromatin compaction.

Random ligation events between un-cross-linked, freely diffusing fragments lead to noise that is mostly seen in trans and long-range cis interactions. Experiments that use DpnII and

[1]Program in Systems Biology, Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, MA, USA. [2]Department of Physics, Massachusetts Institute of Technology, Cambridge, MA, USA. [3]Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, MA, USA. [4]Program in Molecular Medicine, University of Massachusetts Medical School, Worcester, MA, USA. [5]Program in Molecular Medicine, Diabetes Center of Excellence, University of Massachusetts Medical School, Worcester, MA, USA. [6]Department of Molecular Biosciences, University of Texas at Austin, Austin, TX, USA. [7]Institute for Medical Engineering and Science, Massachusetts Institute of Technology, Cambridge, MA, USA. [8]Graduate Program in Biophysics, Harvard University, Cambridge, MA, USA. [9]Howard Hughes Medical Institute, Chevy Chase, MD, USA. [10]These authors contributed equally: Betul Akgol Oksuz, Liyan Yang. ✉e-mail: Johan.Gibcus@umassmed.edu; Job.Dekker@umassmed.edu

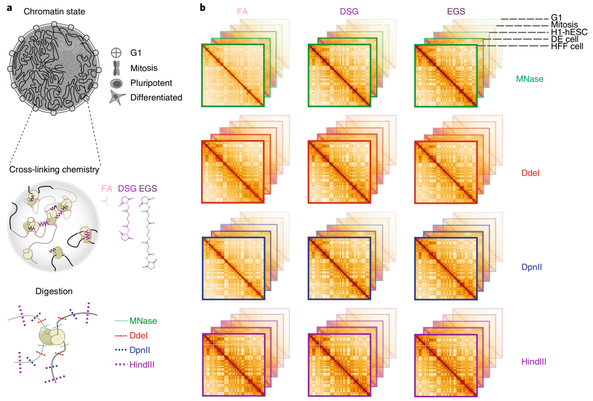
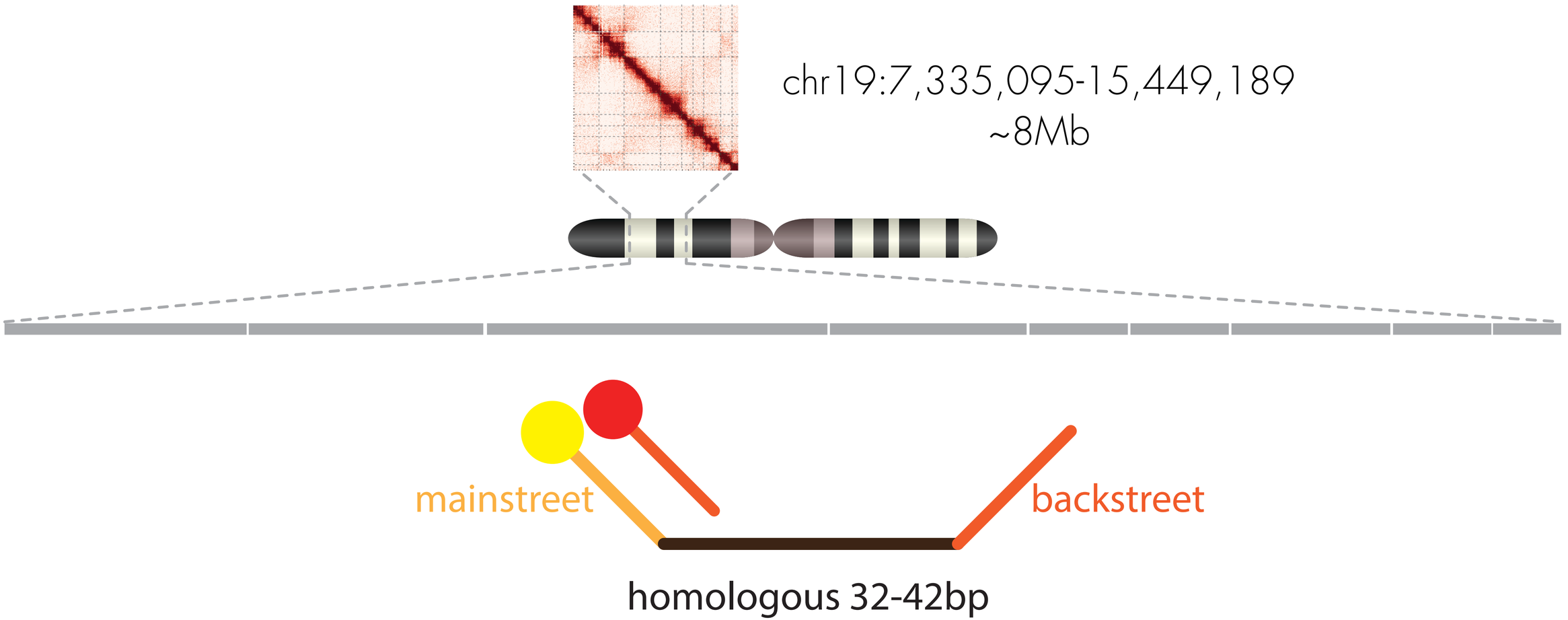


Fig. 1 | Outline of the experimental design. a, Experimental design for conformation capture for various cells, cross-linkers and enzymes. **b**, Representation of interaction maps from experiments in **a**.

High-resolution imaging

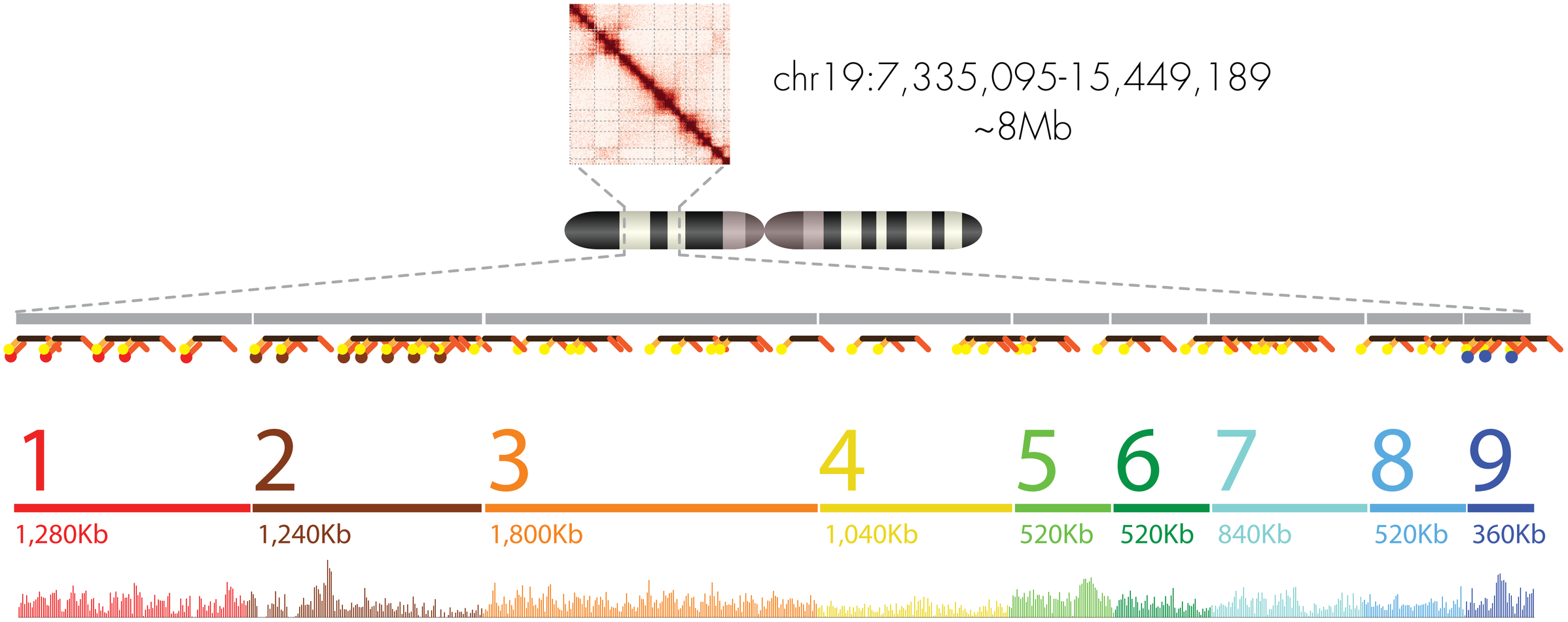
Tracing chromosomes with OligoSTORM & fluidics cycles in PGP1 cells



Beliveau et al. Nat. Comm. 2015

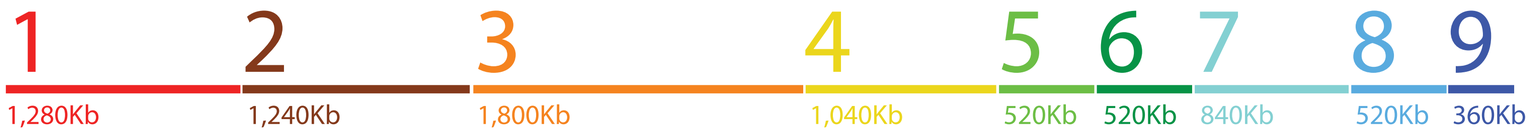
High-resolution imaging

Tracing chromosomes with OligoSTORM & fluidics cycles in PGP1 cells



High-resolution imaging

Tracing chr19:7,335,095-15,449,189 ~8Mb



Picture from the book: Castells i Castellers. Una voluntat col·lectiva.

Fossilized chromosomes from woolly mammoth

Marc A. Marti-Renom

CNAG-CRG · ICREA

Cell. In press.

cnag

CRG Centre for Genomic Regulation

Juan Antonio Rodríguez

Marcela Sandoval Velasco

Tom Gilbert

Baylor College of Medicine

Olga Dudchenko

Cynthia Perez Estrada

Erez Lieberman Aiden

CPG Centre for Palaeogenetics

Love Dalén

Jordan Rowley

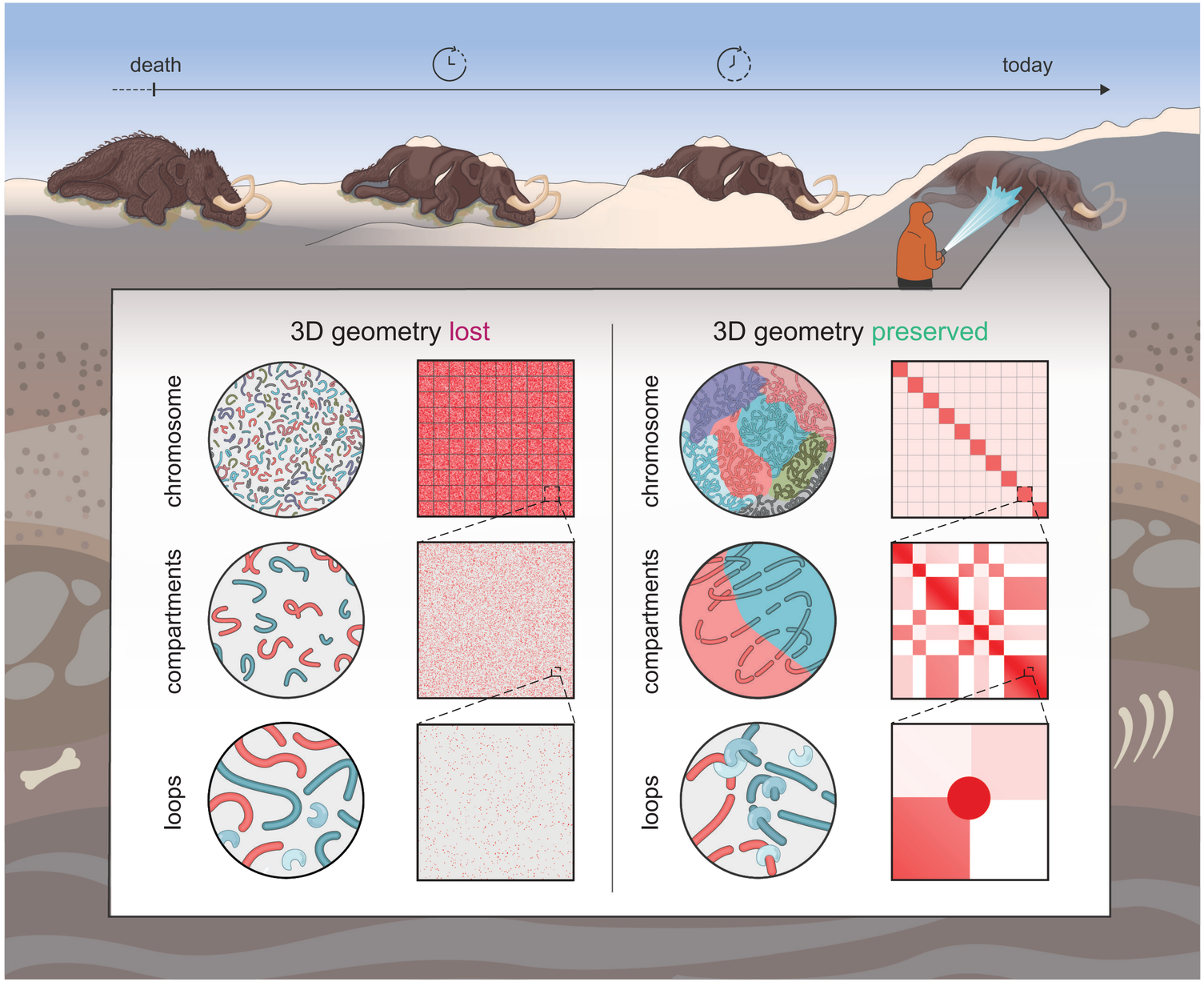
UAB

Aurora Ruiz-Herrera

Kerstin Lidblad-Toh,
Federica Di Palma et al.

The DNAZoo

What happens to the nucleus in 10s of thousands of years?



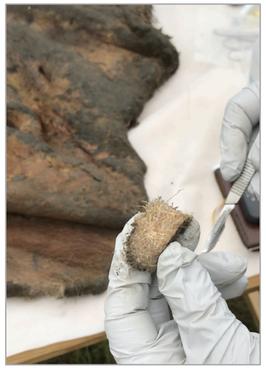
A “whoolly” phenomenal sample

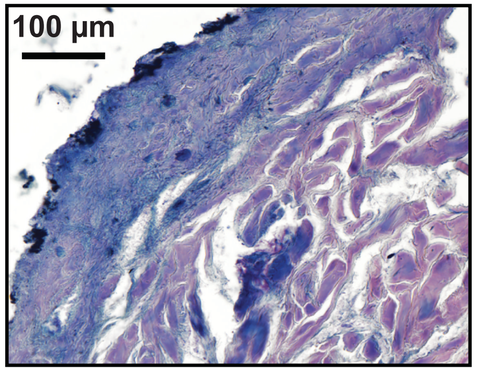
Photo credit: Chris Waddle

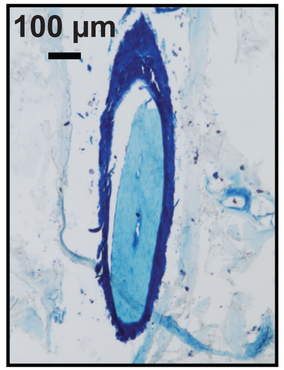
Dan Fisher
UMich, Museum of Paleontology

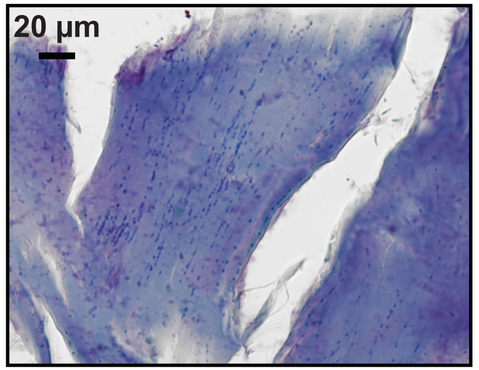
Valeri Plotnikov
Sakha Academy of Sciences

- Found in permafrost in the summer of 2018
- Belaya Gora in Yakutia, Russia
- Date beyond the range of radiocarbon dating but older than >45,000 years









Paleo-HiC complements ancient DNA-seq

Limitations of (a)DNA-Seq

What is in the genome?

Need chrom-length de novo assemblies!
aDNA-Seq relies on modern references

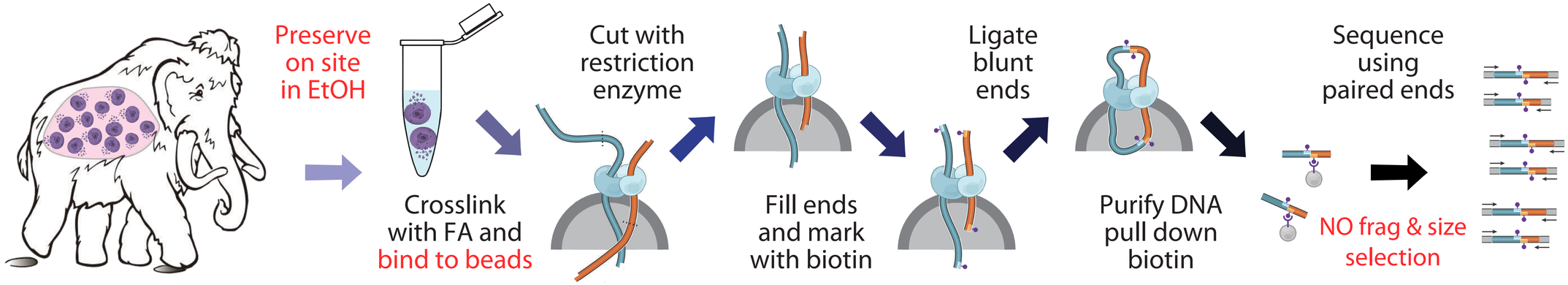
What is expressed in individual tissues?

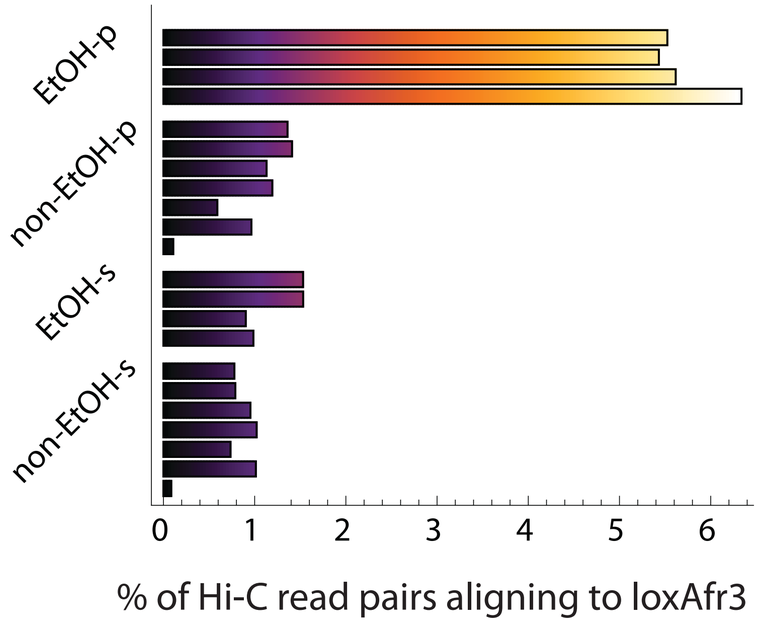
Need to probe transcriptional activity!

How expression patterns arise?

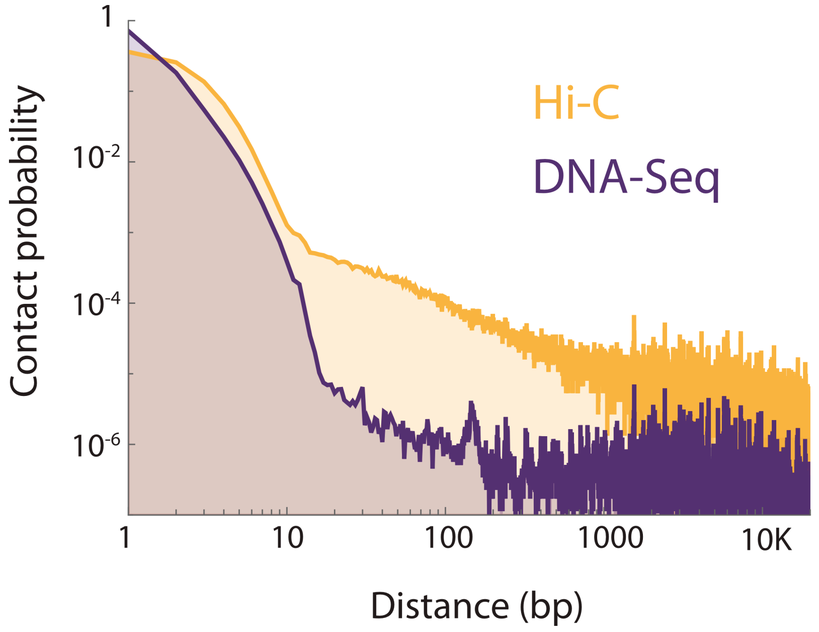
Need to probe genetic regulation!

Paleo-HiC improves endogenous long-range contact recovery



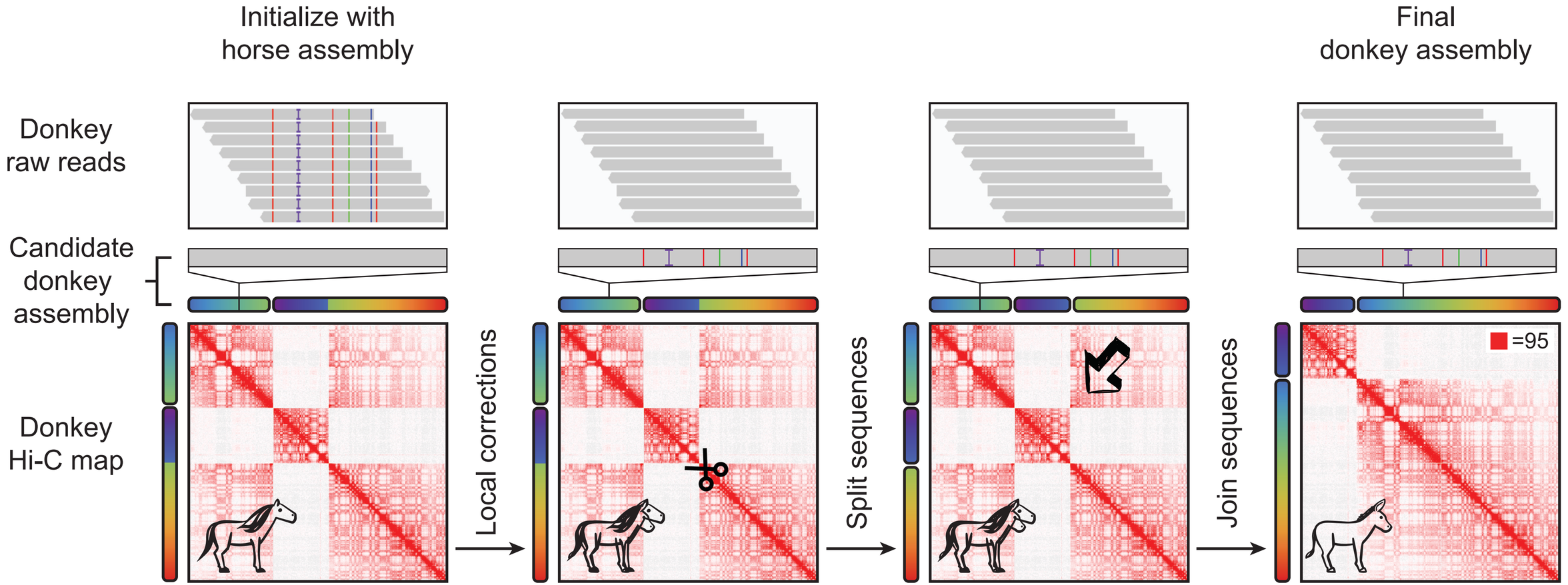


Total read count	4,444,894,354
Unique paired alignments (loxAfr3)	24,415,411
Unique paired (%)	0.55%
Long-range (20kb)	1,763,225
Long-range (%)	0.04%



Hi-C assisted assembly

Dubchenko et al. Science. 2017 Apr 7;356(6333):92-95

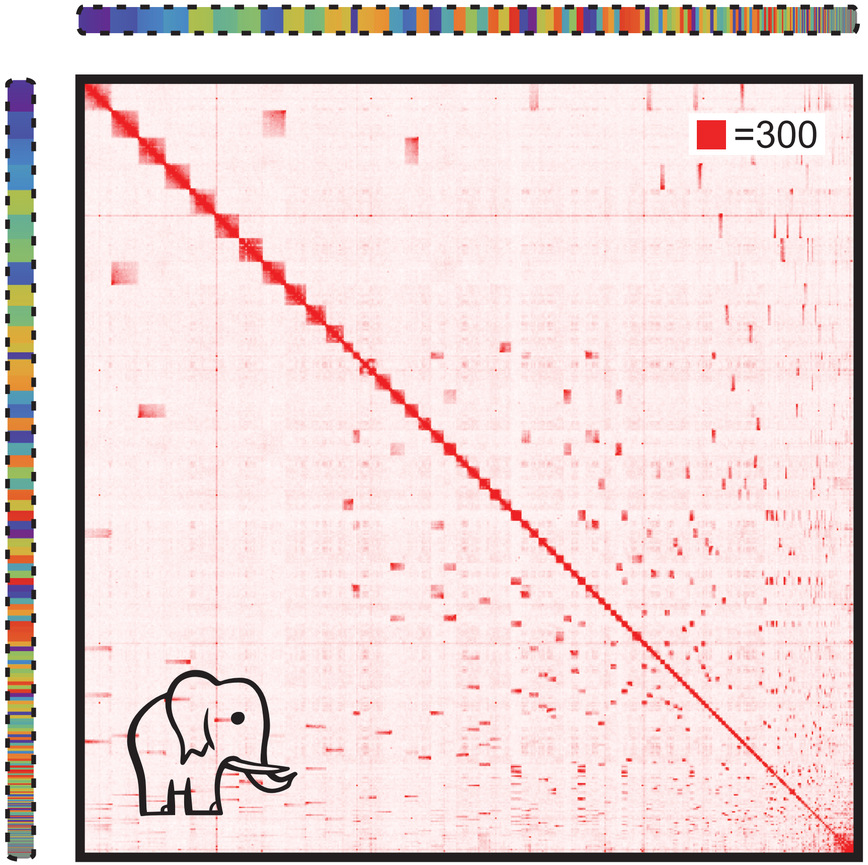


correct · split · orient · order

This is a Hi-C from mammoth

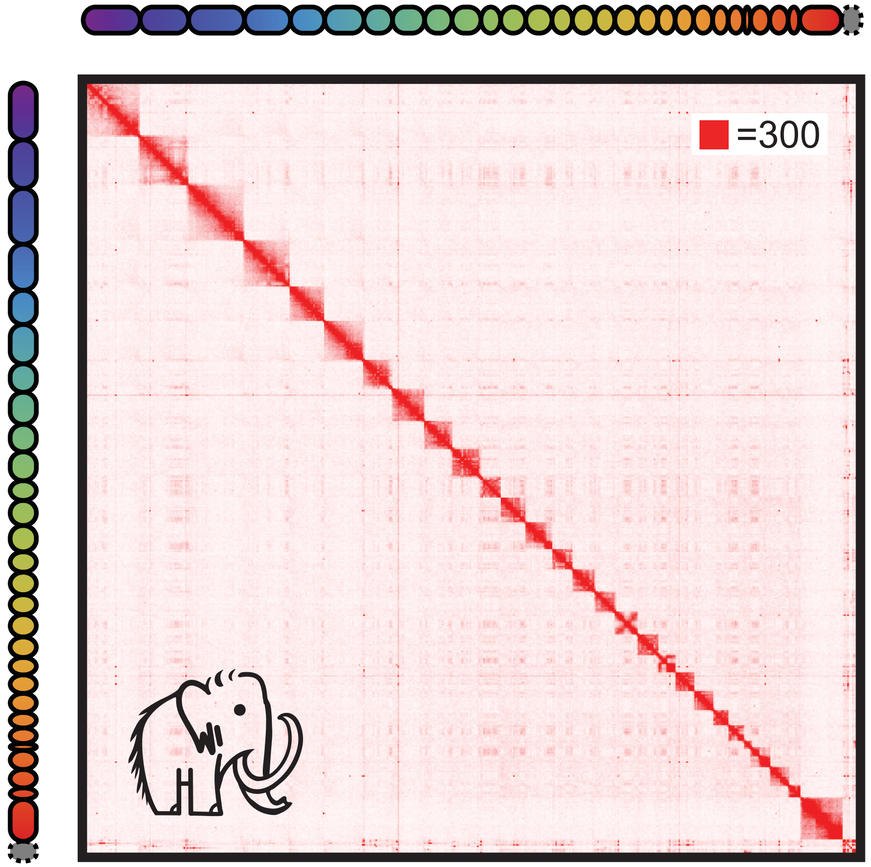
based on Loxafr3.0

PaleoHi-C vs Loxafr3.0,
fragmentary African elephant assembly



3D assisted
assembly

PaleoHi-C vs MamPri_Loxafr3.0_assisted_HiC,
chromosome-length mammoth assembly



Paleo-HiC complements ancient DNA-seq

Limitations of (a)DNA-Seq

What is in the genome?

Need chrom-length de novo assemblies!
aDNA-Seq relies on modern references

What is expressed in individual tissues?

Need to probe transcriptional activity!

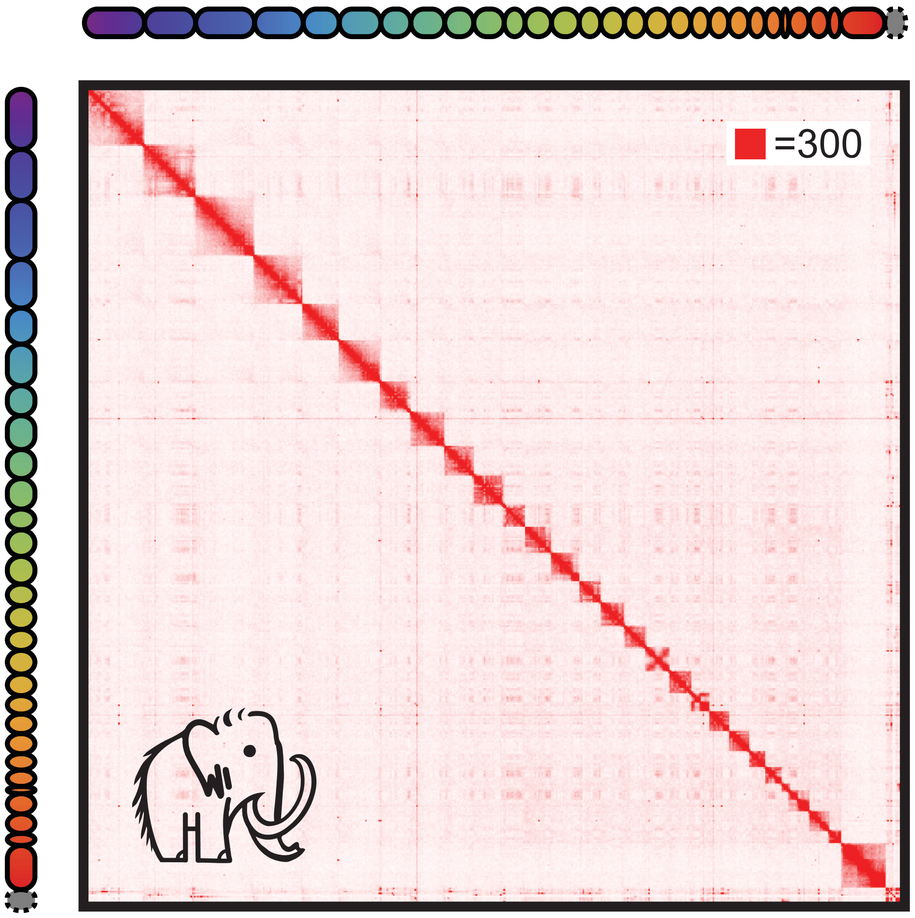
How expression patterns arise?

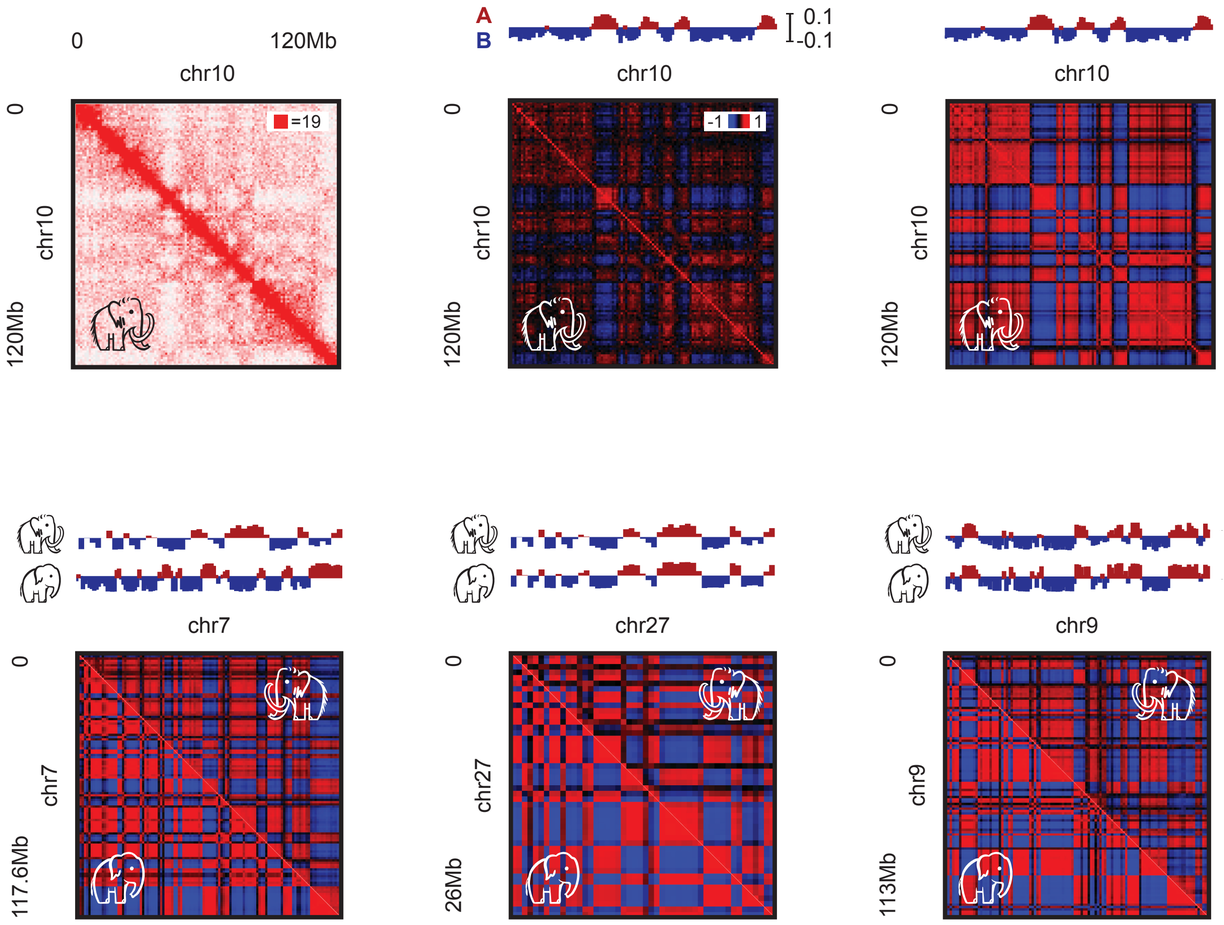
Need to probe genetic regulation!

Hallmarks of a successful Hi-C experiment

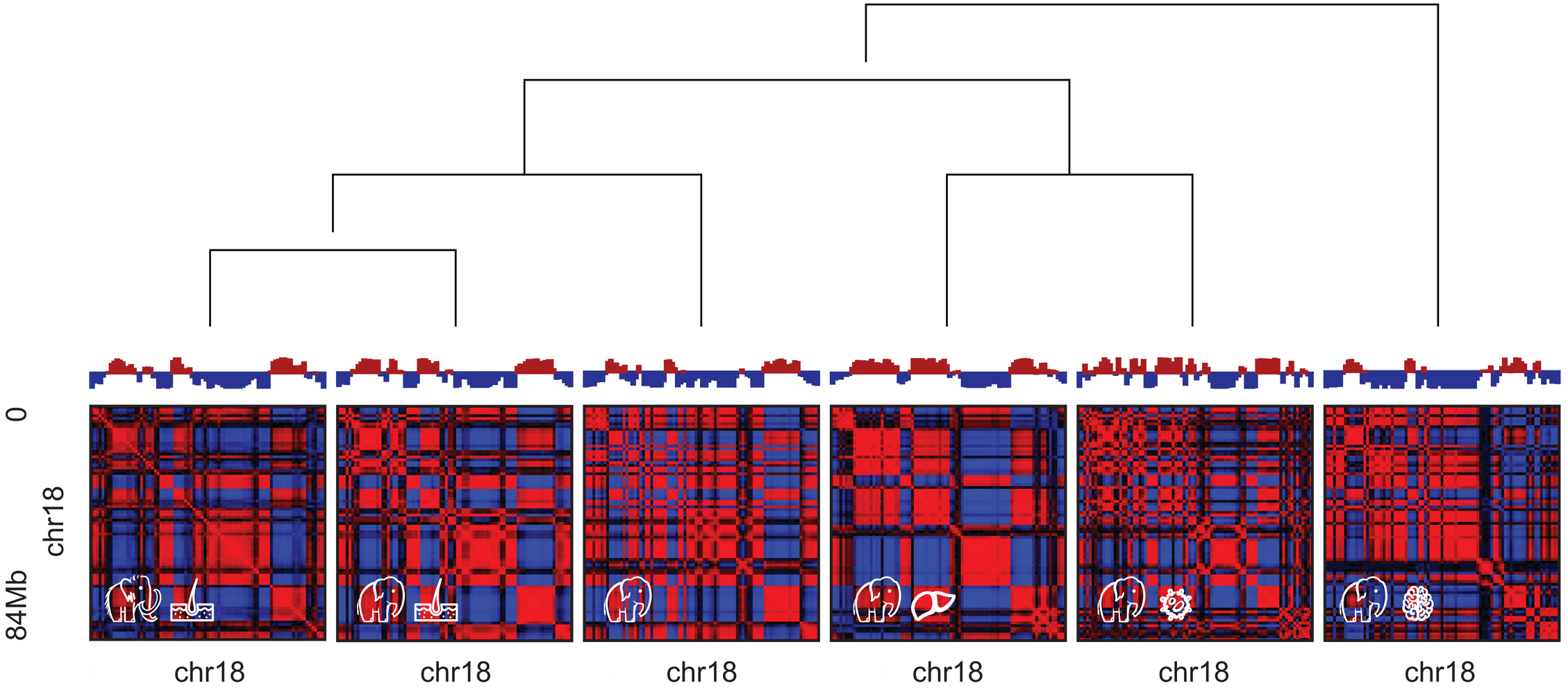
- Chromosome territories
  Facilitates **de novo assembly of whole chromosomes**

Compartments preserved in a 47K years old sample

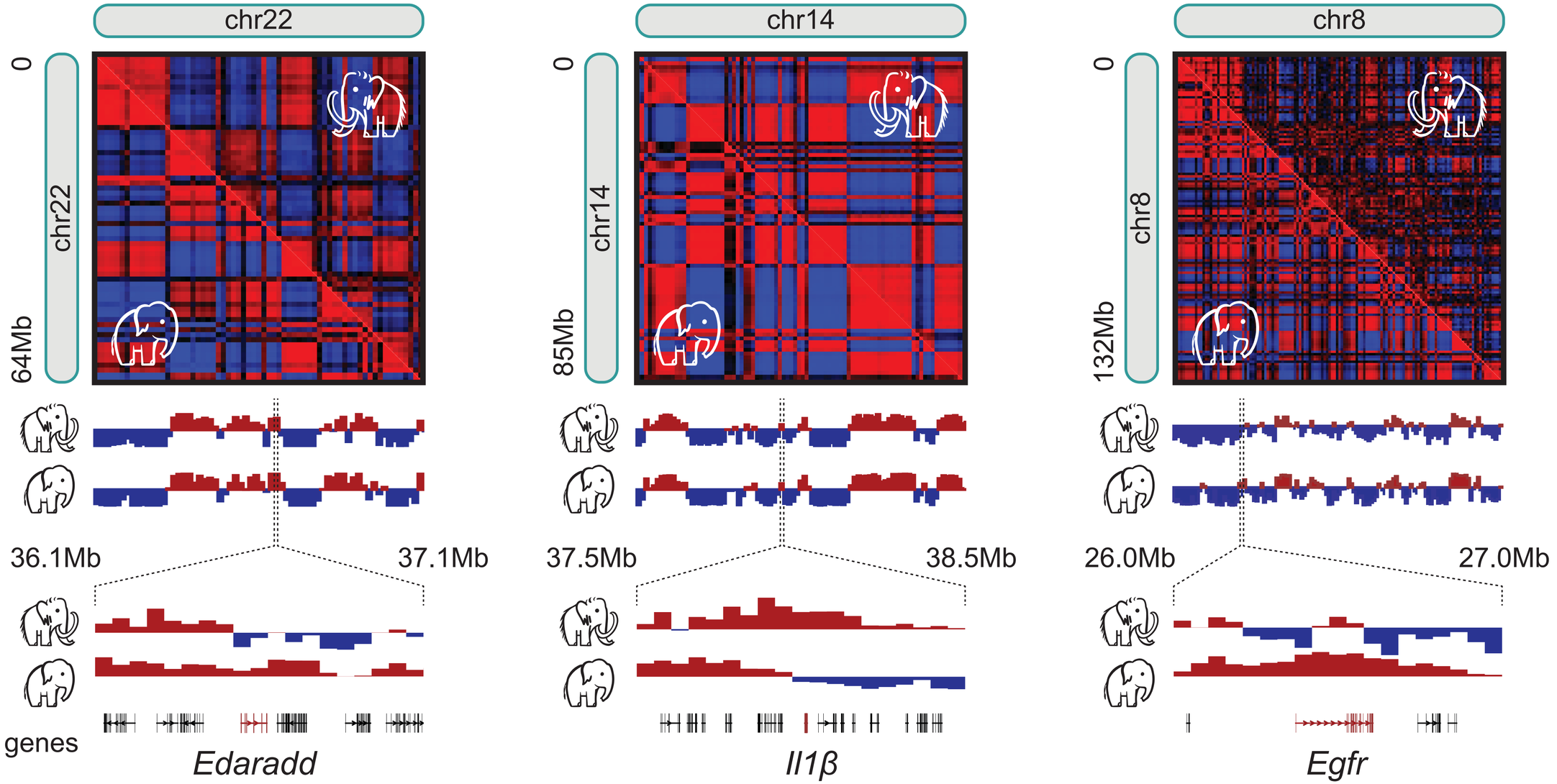




Tissue specific compartmentalization



52 Mammoth Altered Regulation Sequences (MARS)



Paleo-HiC complements ancient DNA-seq

Limitations of (a)DNA-Seq

What is in the genome?

Need chrom-length de novo assemblies!
aDNA-Seq relies on modern references

What is expressed in individual tissues?

Need to probe transcriptional activity!

How expression patterns arise?

Need to probe genetic regulation!

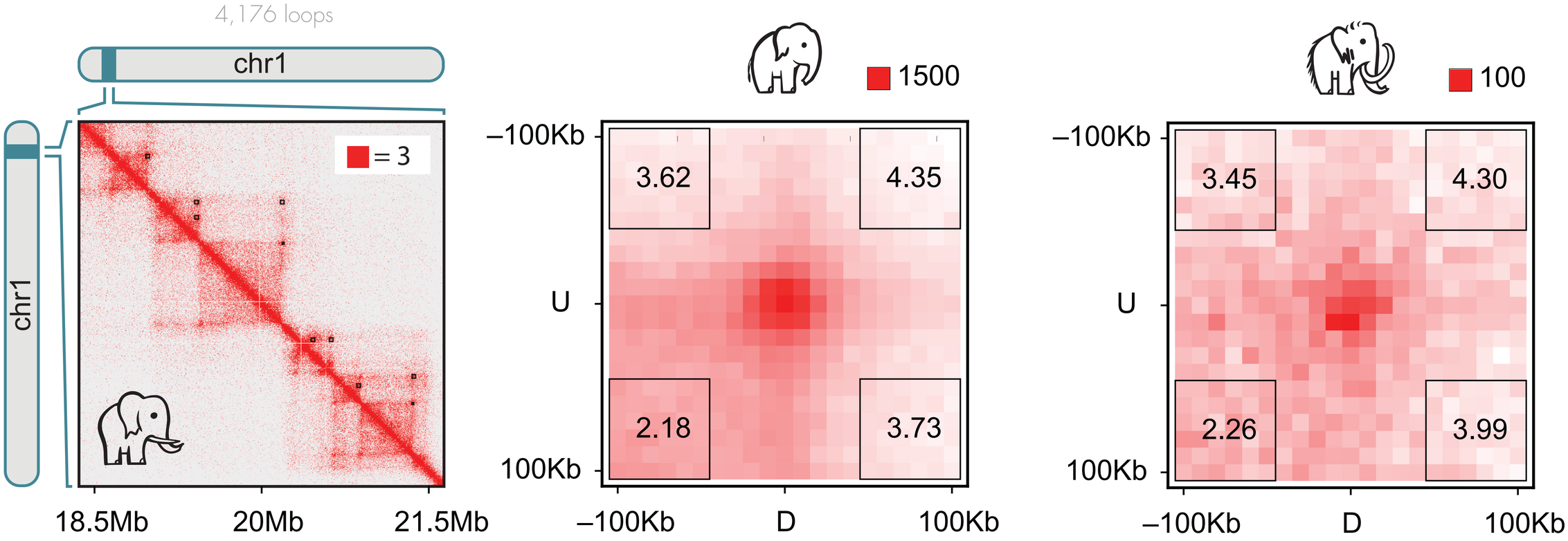
Hallmarks of a successful Hi-C experiment

- Chromosome territories
  Facilitates **de novo assembly of whole chromosomes**

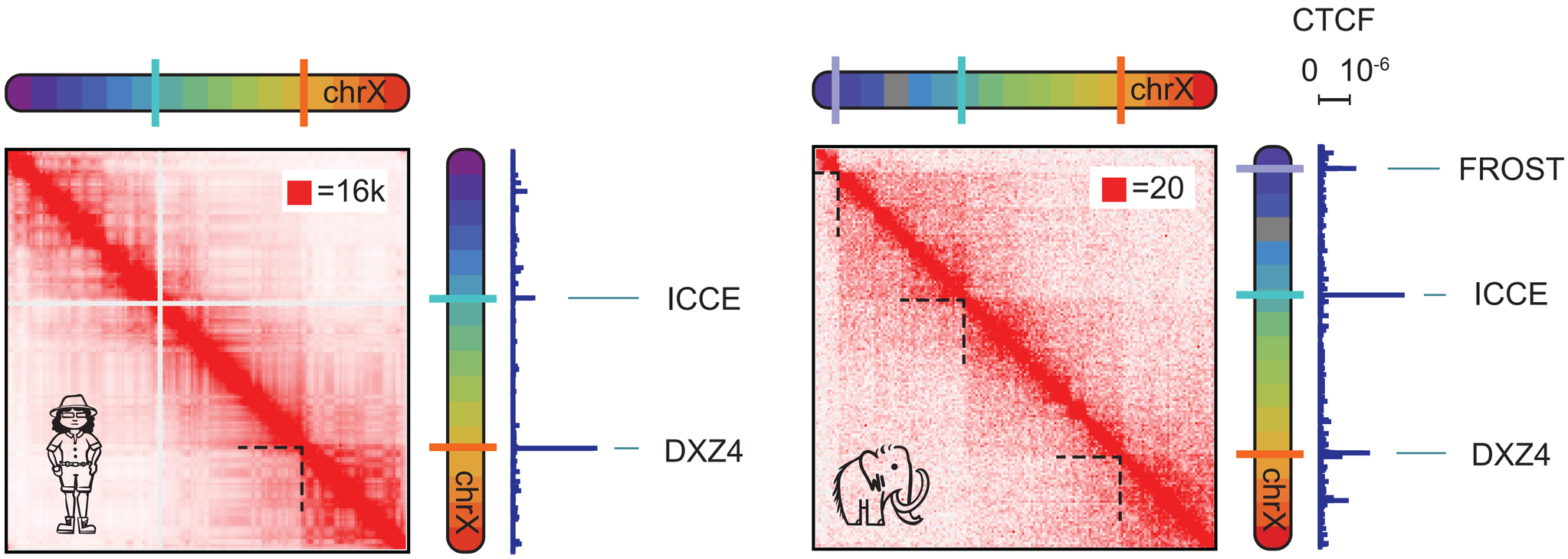
- Active and inactive chromatin compartments
  Probes **Transcriptional activity**

Paleo-hic recovers loop signatures!

Rao, Huntley et al., Cell 2014



Inactive chromosome X segregates



Paleo-HiC complements ancient DNA-seq

Limitations of (a)DNA-Seq

What is in the genome?

Need chrom-length de novo assemblies!
aDNA-Seq relies on modern references

What is expressed in individual tissues?

Need to probe transcriptional activity!

How expression patterns arise?

Need to probe genetic regulation!

Hallmarks of a successful Hi-C experiment

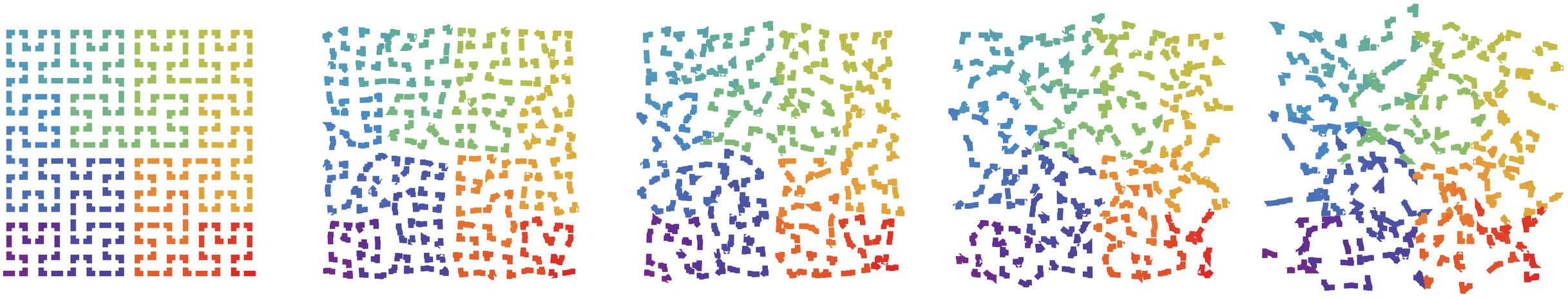
- Chromosome territories
  Facilitates **de novo assembly of whole chromosomes**
- Active and inactive chromatin compartments
  Probes **Transcriptional activity**
- Chromatin Loops
  Reveals **regulation of individual genes**
- Barr body of the inactive X
  Reflects **chromosome-scale dosage compensation**

How is this possible?

The "chromoglass" hypothesis

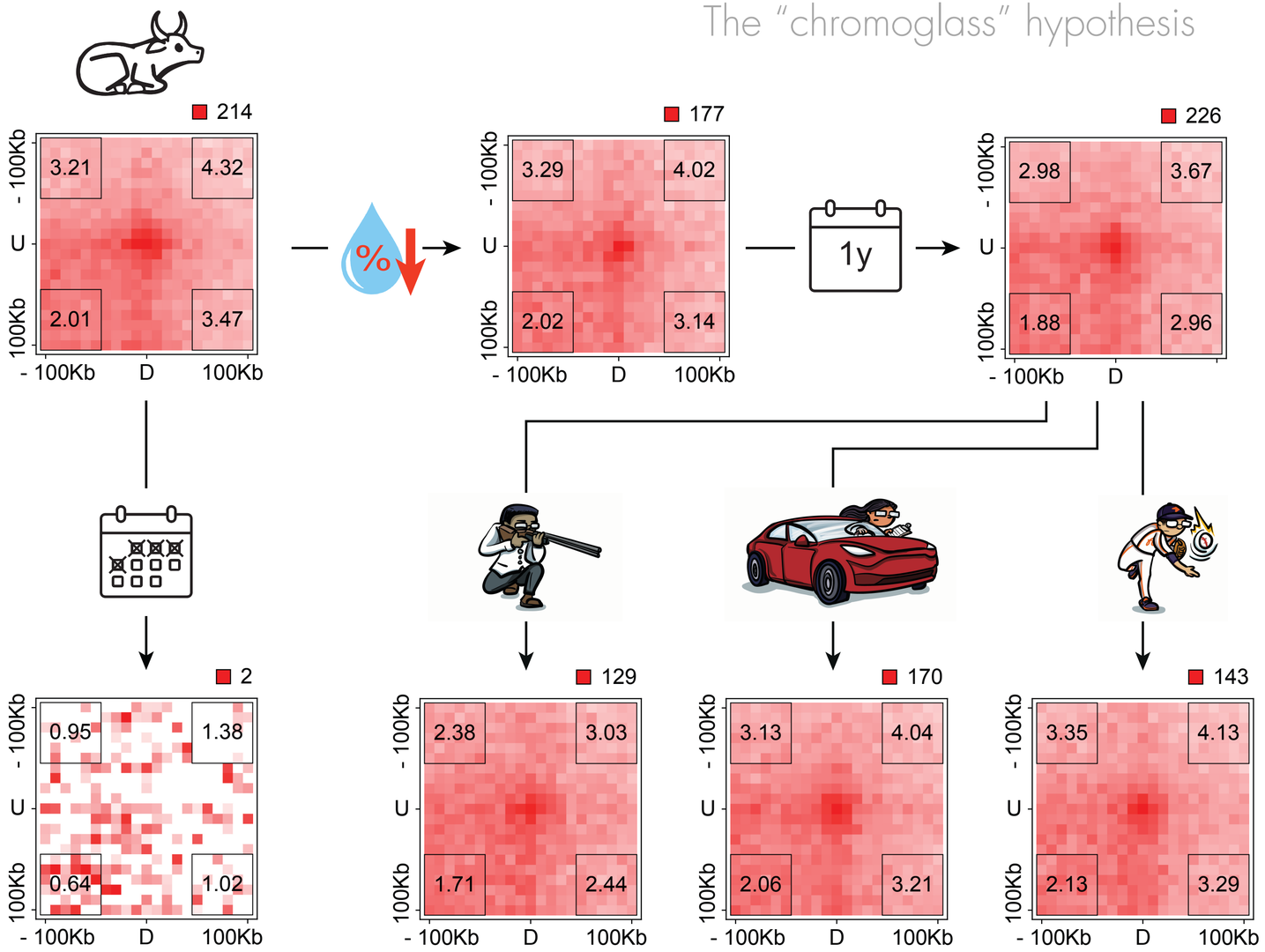
Initial structure

Diffusion



How is this possible? (q.k.a. reviewer #3)

The "chromoglass" hypothesis



THREE-DIMENSIONAL GENOME ARCHITECTURE PERSISTS IN A 52,000-YEAR-OLD WOOLLY MAMMOTH SKIN SAMPLE

Marcela Sandoval-Velasco#, Olga Dudchenko#,†, Juan Antonio Rodríguez#, Cynthia Pérez Estrada#, Marianne Dehasque, Claudia Fontsere, Sarah S.T. Mak, Ruqayya Khan, Vinícius G. Contessoto, Antonio B. Oliveira Junior, Achyuth Kalluchi, Bernardo J. Zubillaga Herrera, Jiyun Jeong, Renata P. Roy, Ishawnia Christopher, David Weisz, Arina D. Omer, Sanjit S. Batra, Muhammad S. Shamim, Neva C. Durand, Brendan O'Connell, Alfred L. Roca, Maksim V. Plikus, Mariya A. Kusliy, Svetlana A. Romanenko, Natalya A. Lemskaya, Natalya A. Serdyukova, Svetlana A. Modina, Polina L. Perelman, Elena A. Kizilova, Sergei I. Baiborodin, Nikolai B. Rubtsov, Gur Machol, Krisha Rath, Ragini Mahajan, Parwinder Kaur, Andreas Gnirke, Isabel Garcia-Treviño, Rob Coke, Joseph P. Flanagan, Kelcie Pletch, Aurora Ruiz-Herrera, Valerii Plotnikov, Innokentiy S. Pavlov, Naryya I. Pavlova, Albert V. Protopopov, Michele Di Pierro, Alexander S. Graphodatsky, Eric S. Lander, M. Jordan Rowley, Peter G. Wolynes, José N. Onuchic, Love Dalén, Marc A. Marti-Renom†, M. Thomas P. Gilbert†, Erez Lieberman Aiden†

Cell 2024

Take home messages:

Mammoth foot
Photo credit: Love Dalén

- Hi-C was done in a 52,000-year-old well conserved sample.
- Chromosome fossils also enable to assemble the entire genome of extinct species.
- Chromosome fossils help to interpret how the genomes of those species were organized in space as well as its functional activity.
- Key mammoth genes associated with hair follicle development were active in mammoth compared to modern elephants.
- Specific loop interactions in the genome regulating gene expression were also visible and conserved in the mammoth sample.
- Chromoglass (a glass-like-state of the chromosomes) allowed the genome structure to be physically conserved over such long period of time.

https://tinyurl.com/MammothPaper

@mamartirenom

http://marciuslab.org

Alexander Barclay
Nikolai Bykov
Iana Kim
Peter Hoboth
Zoé Horisberger
Anne Lee
John Markham
Maria Marti-Marimon
Anastasiia Nikitina
Ana Nikolovska
Mireia Novell
Meritxell Novillo
Aleksandra Sparavier
Leo Zuber

.: Our current sponsors :.

European Commission

MINISTERIO DE CIENCIA E INNOVACIÓN

AGENCIA ESTATAL DE INVESTIGACIÓN

Generalitat de Catalunya

NIH National Human Genome Research Institute