Structure determination of genomes and genomic domains by satisfaction of spatial restraints

Marc A. Marti-Renom
CNAG-CRG · ICREA

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

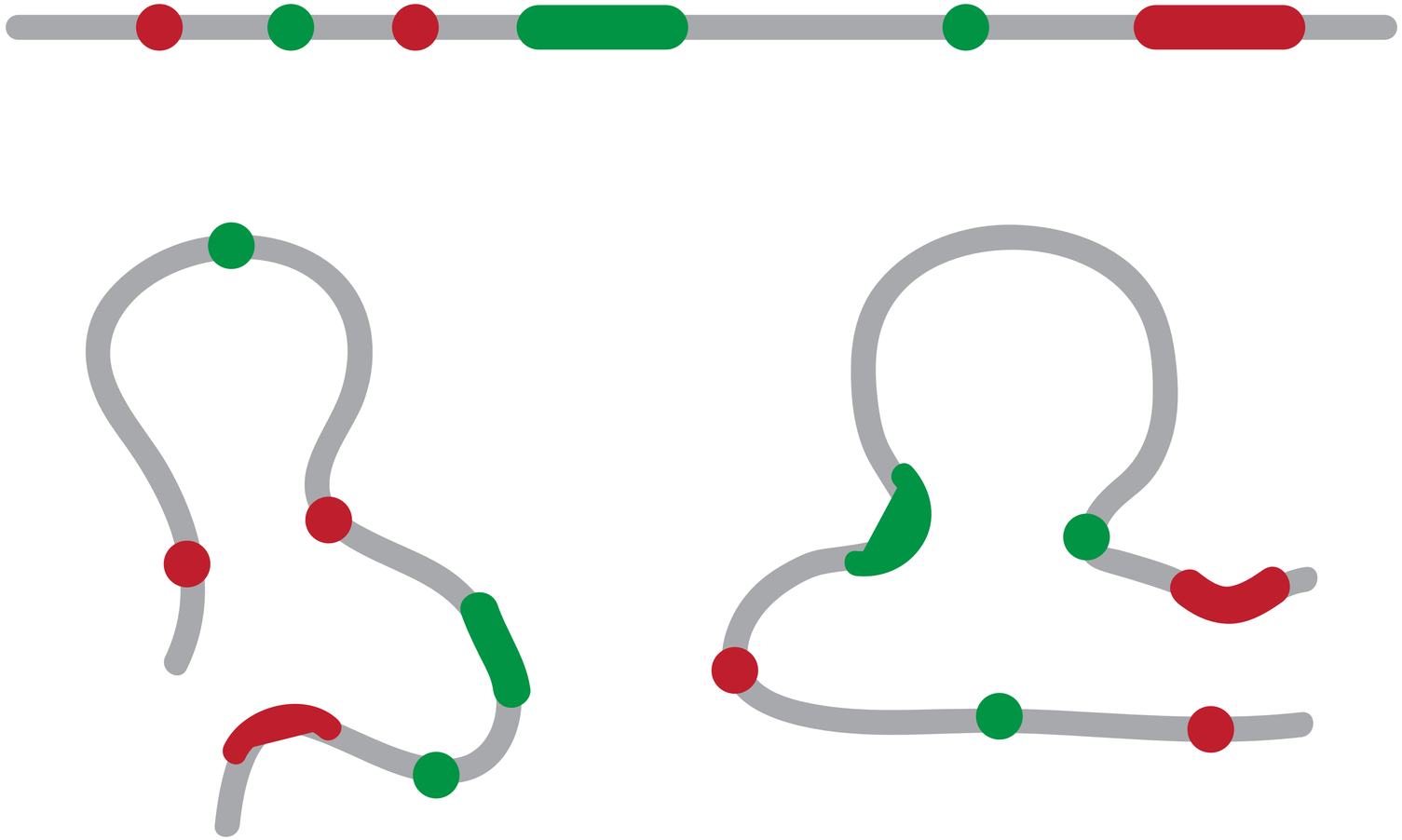
cnag CRG ICREA

All you will see in the screen will be stored here:

http://sgt.cnag.eu/www/presentations/

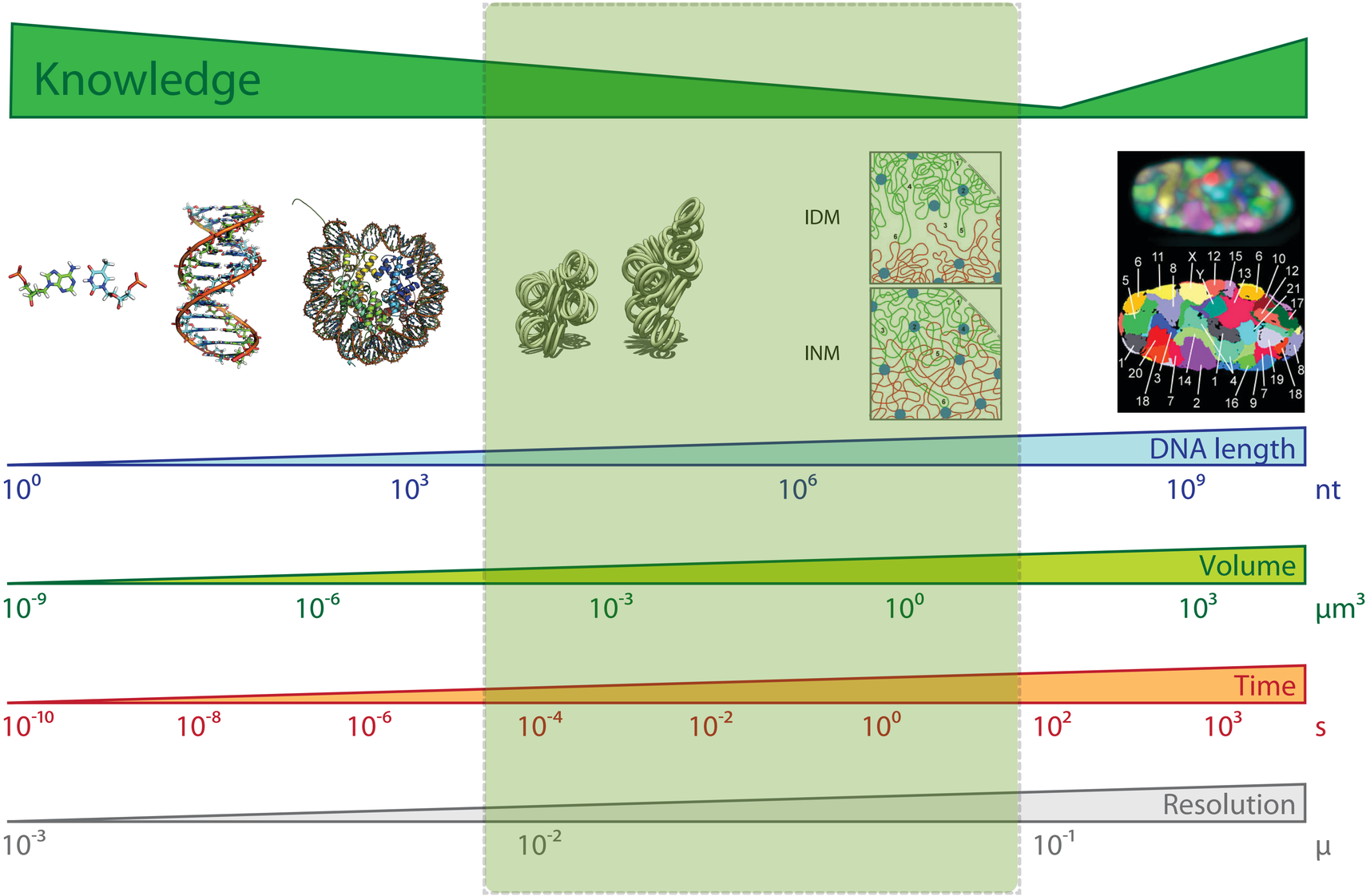
I encourage you to:

You can ask for question any time



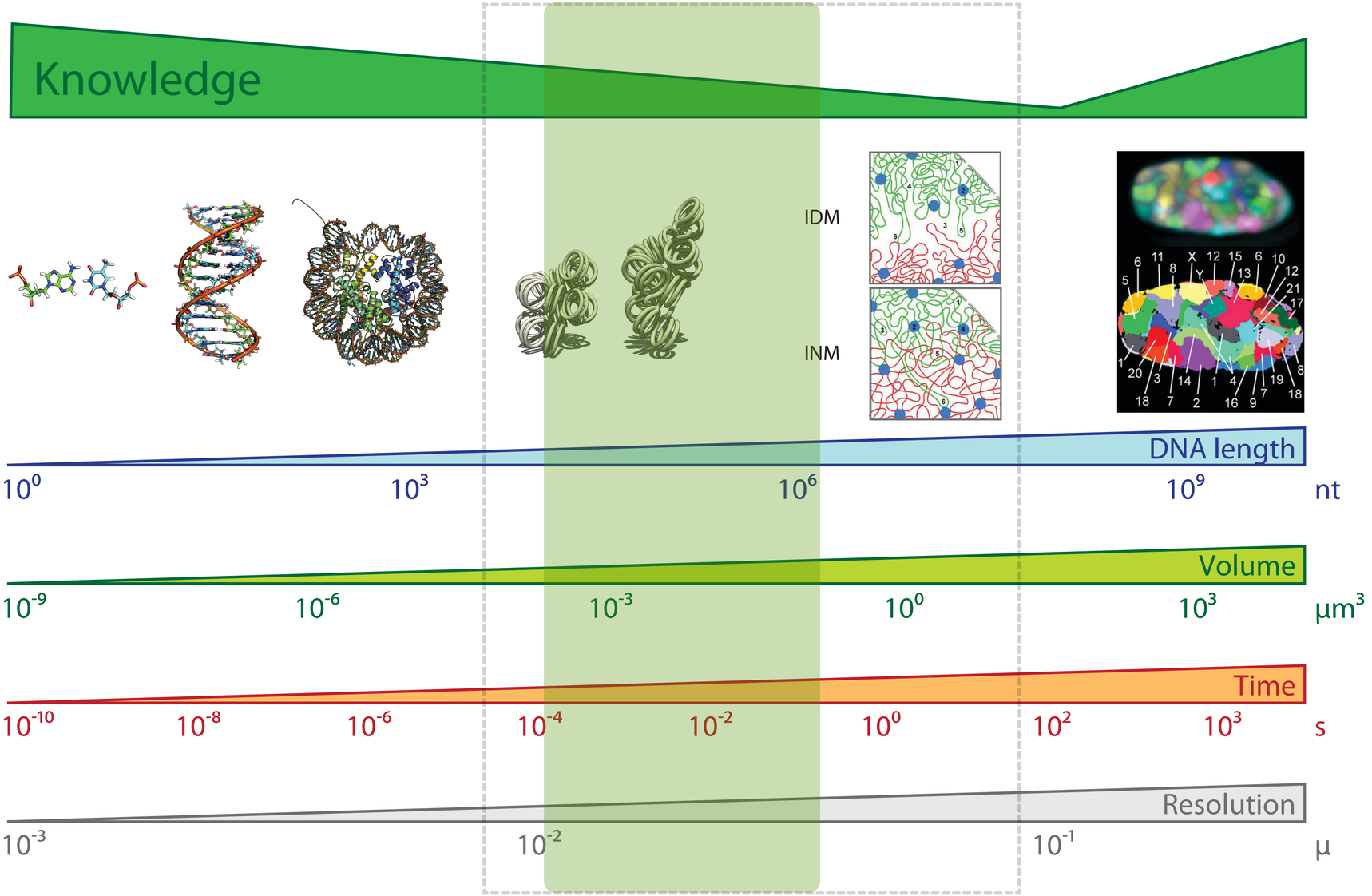
Resolution Gap

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



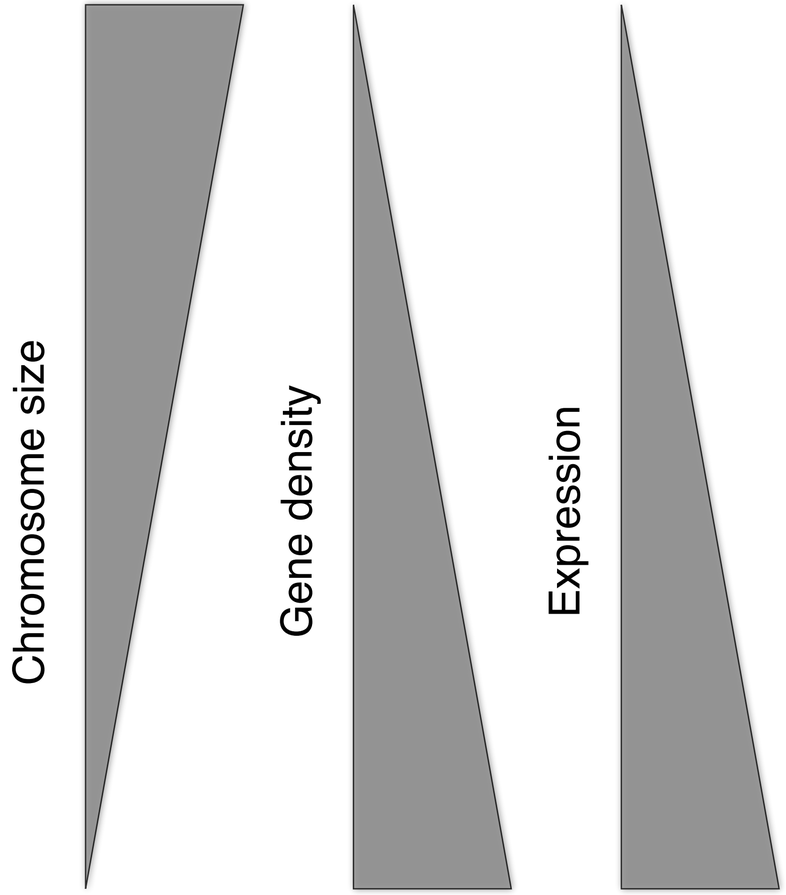
Resolution Gap

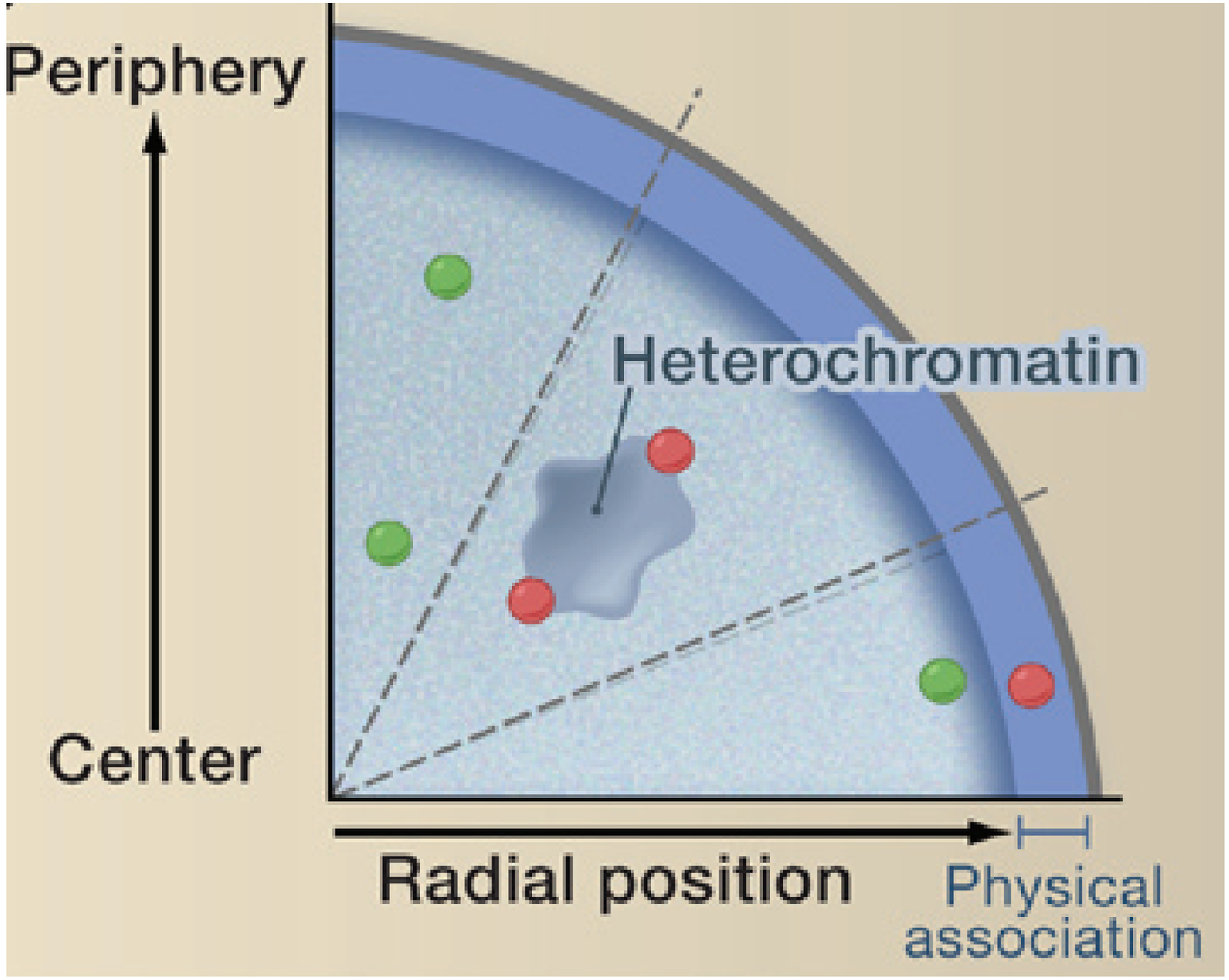
Marti-Renom, M. A. & Mirny, L. A. PLoS Comput Biol *7*, e1002125 (2011)



Level I: Radial genome organization

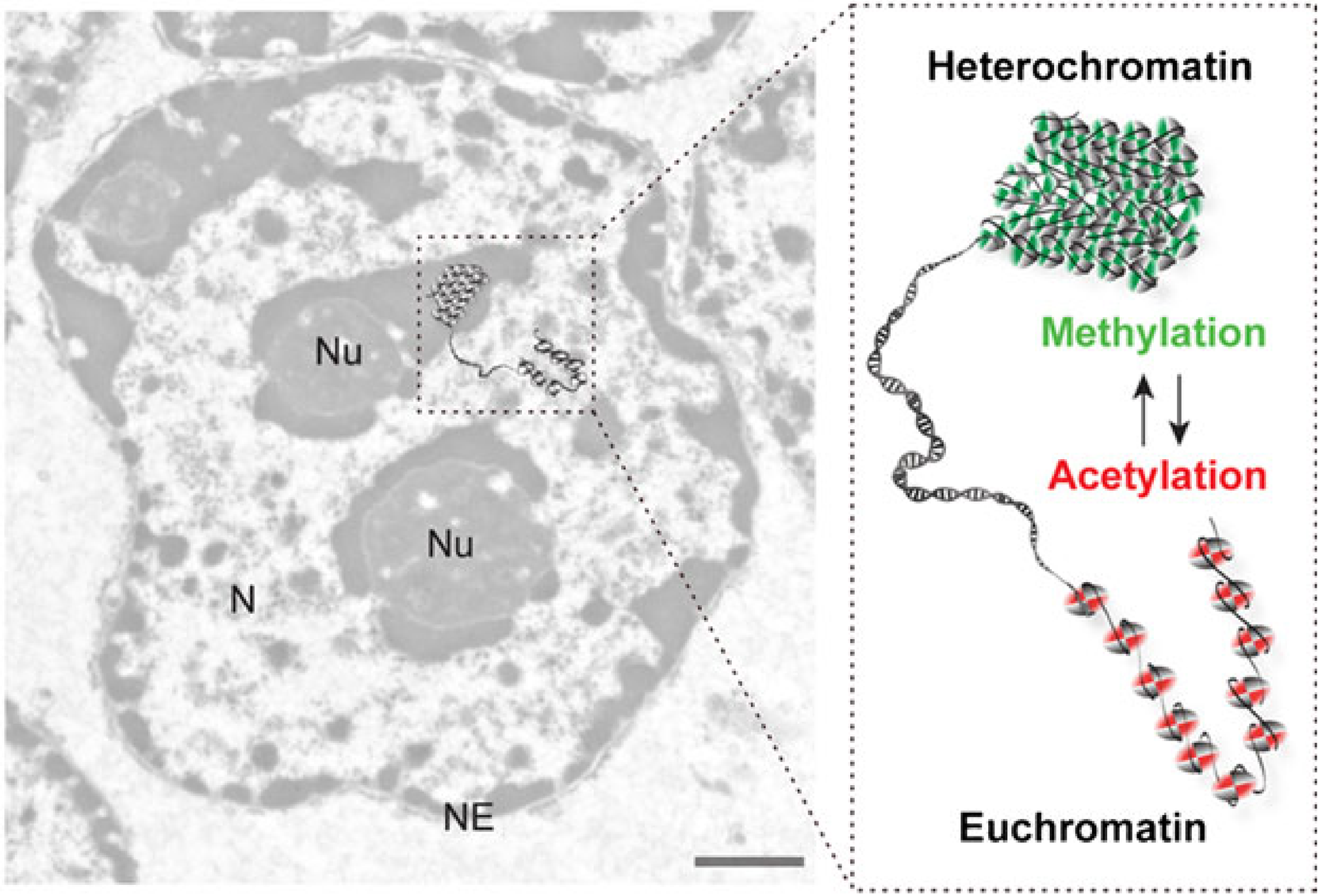
Takizawa, T., Meaburn, K. J. & Misteli, T. The meaning of gene positioning. Cell 135, 9–13 (2008).



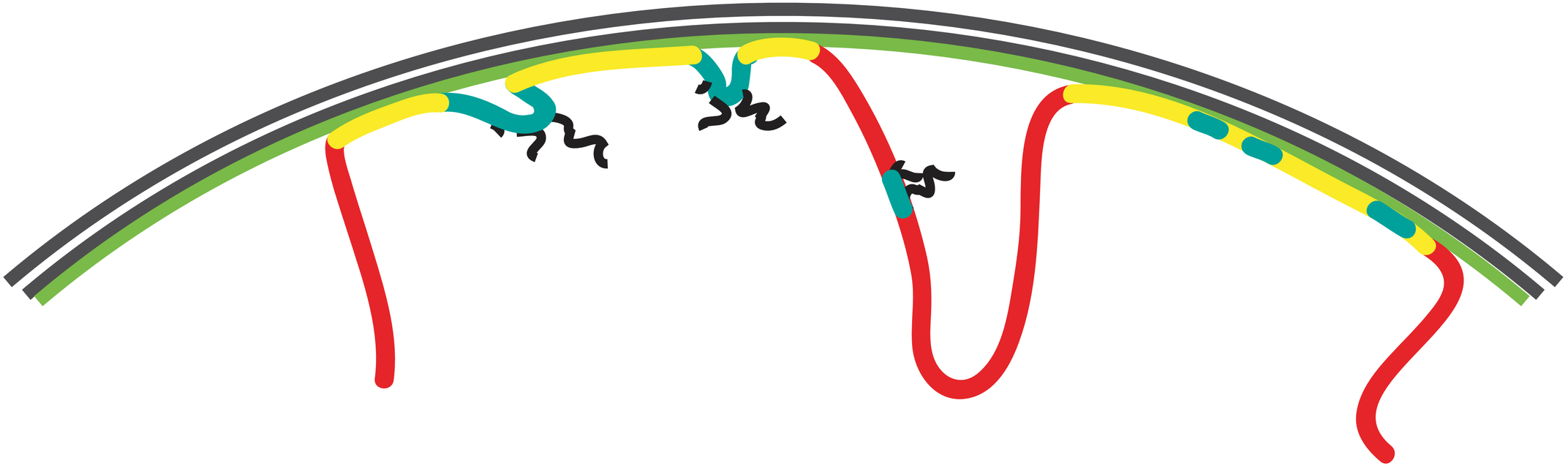


Level II: Euchromatin vs heterochromatin

Electron microscopy



Level III: Lamina-genome interactions



nuclear membrane
nuclear lamina

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

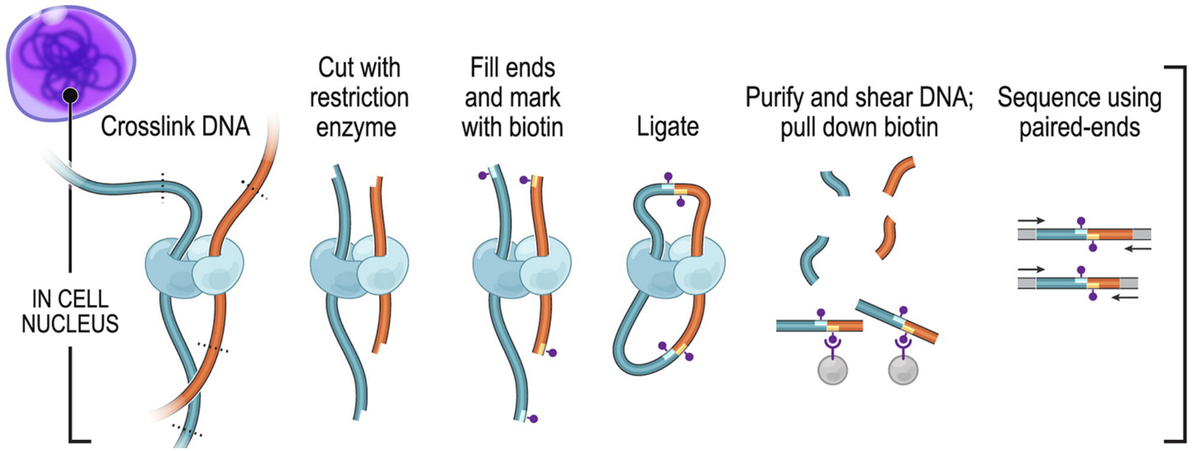
Genes

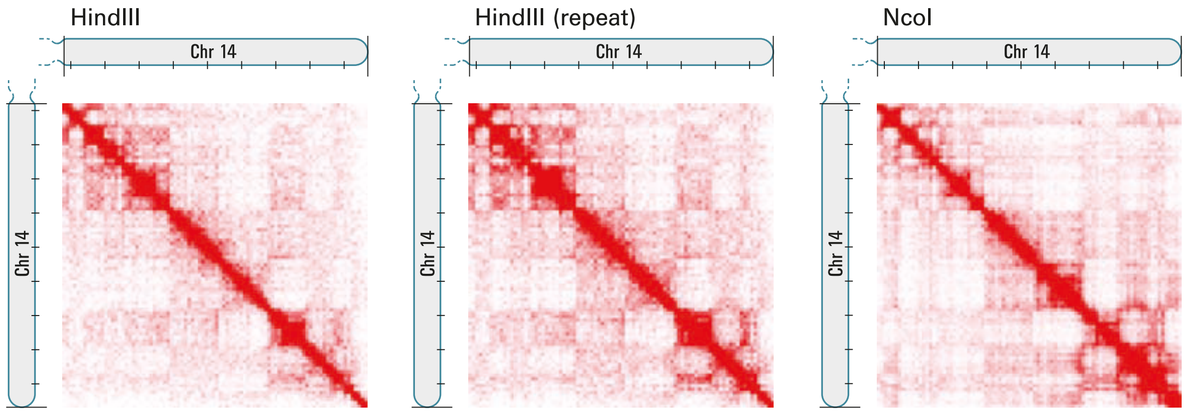
mRNA

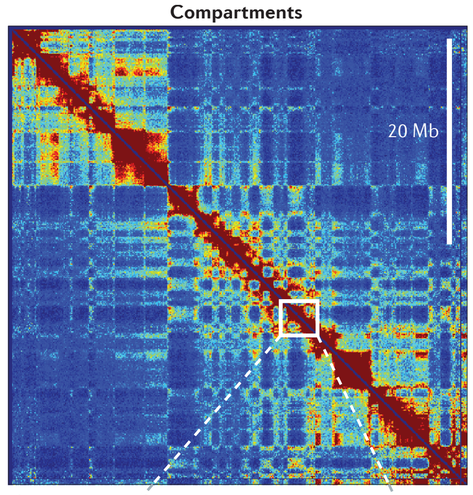
Adapted from Molecular Cell 38, 603-613, 2010

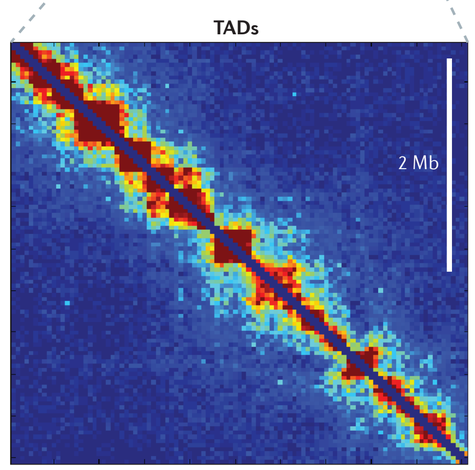
Level IV: Higher-order organization

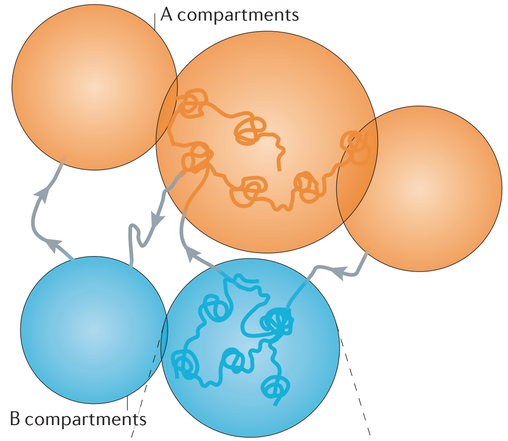
Dekker, J., Marti-Renom, M. A. & Mirny, L. A. Nat Rev Genet 14, 390–403 (2013).

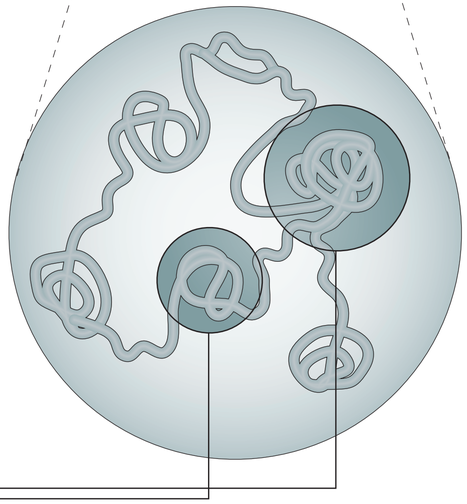




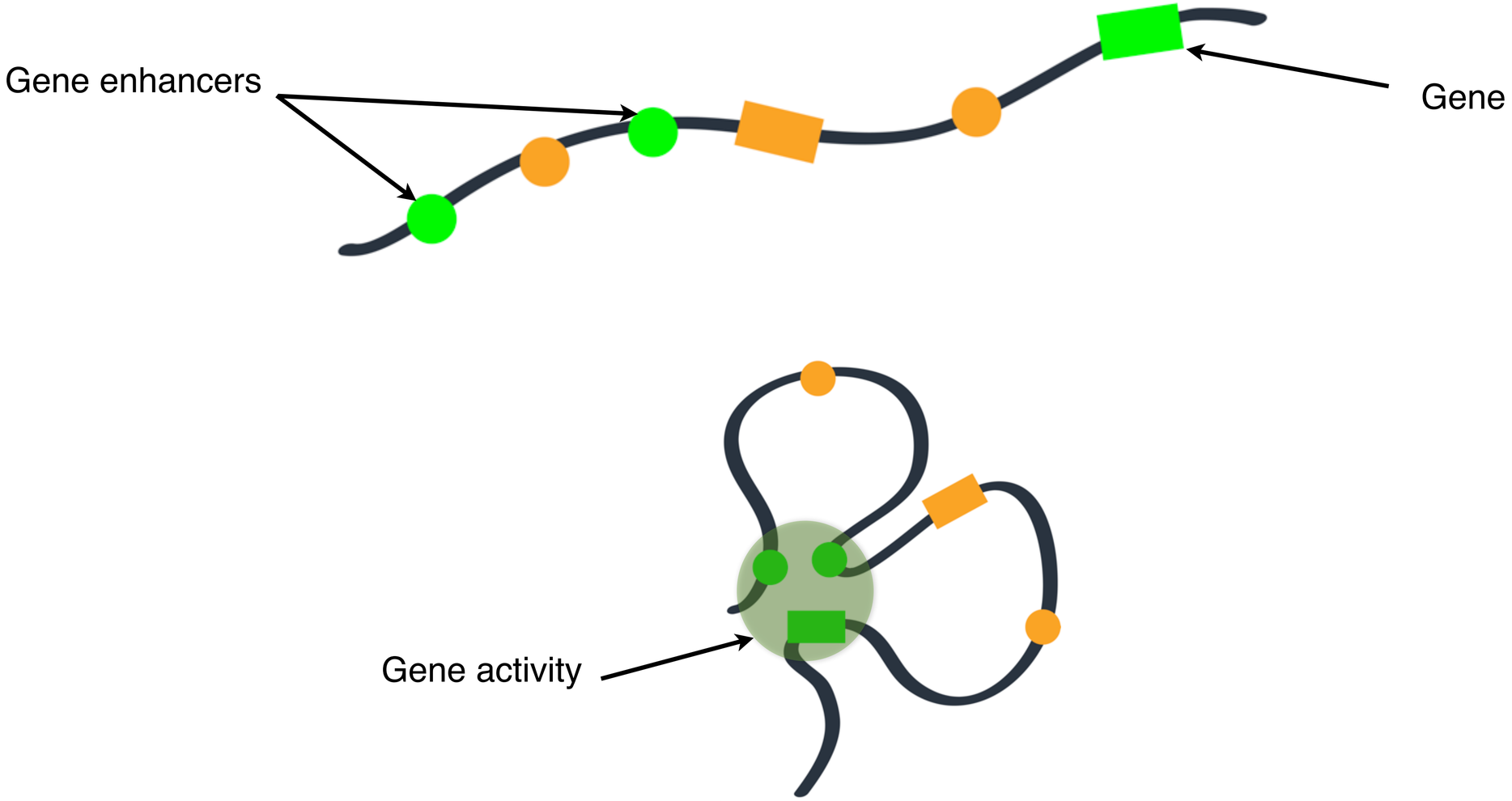






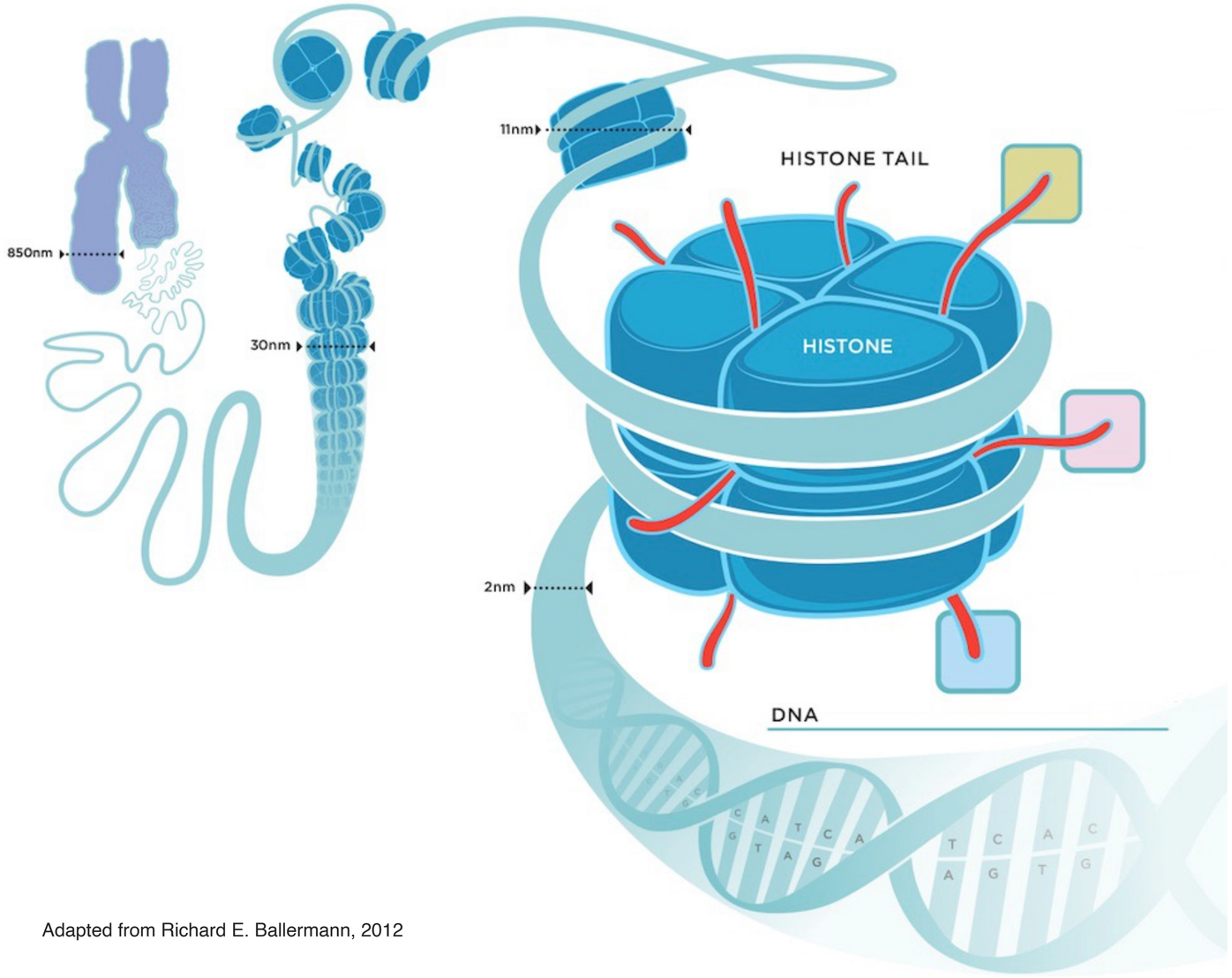


Level V: Chromatin loops



Level VI: Nucleosome

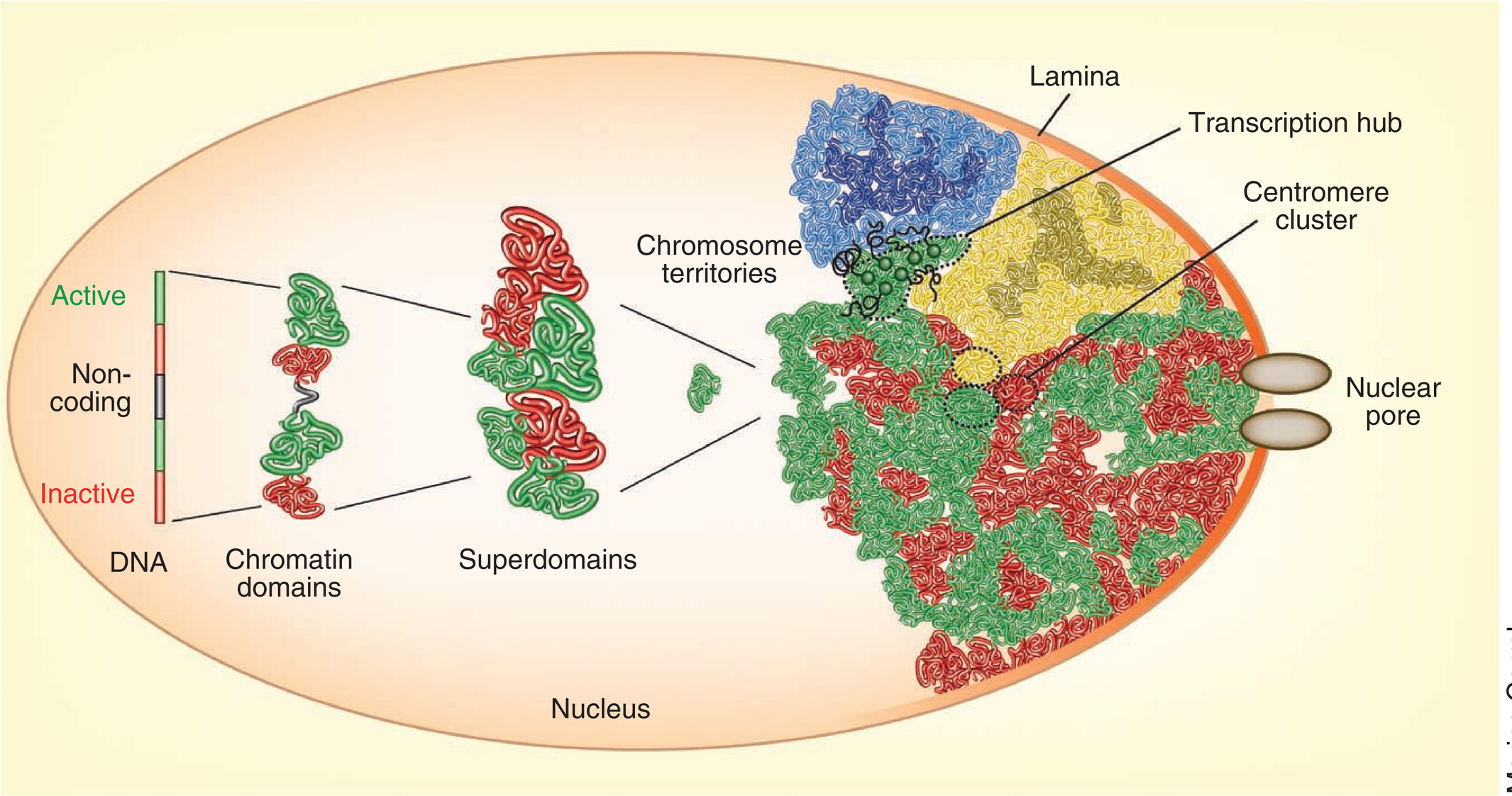
Chromosome **Chromatin fibre** **Nucleosome**



Adapted from Richard E. Ballermann, 2012

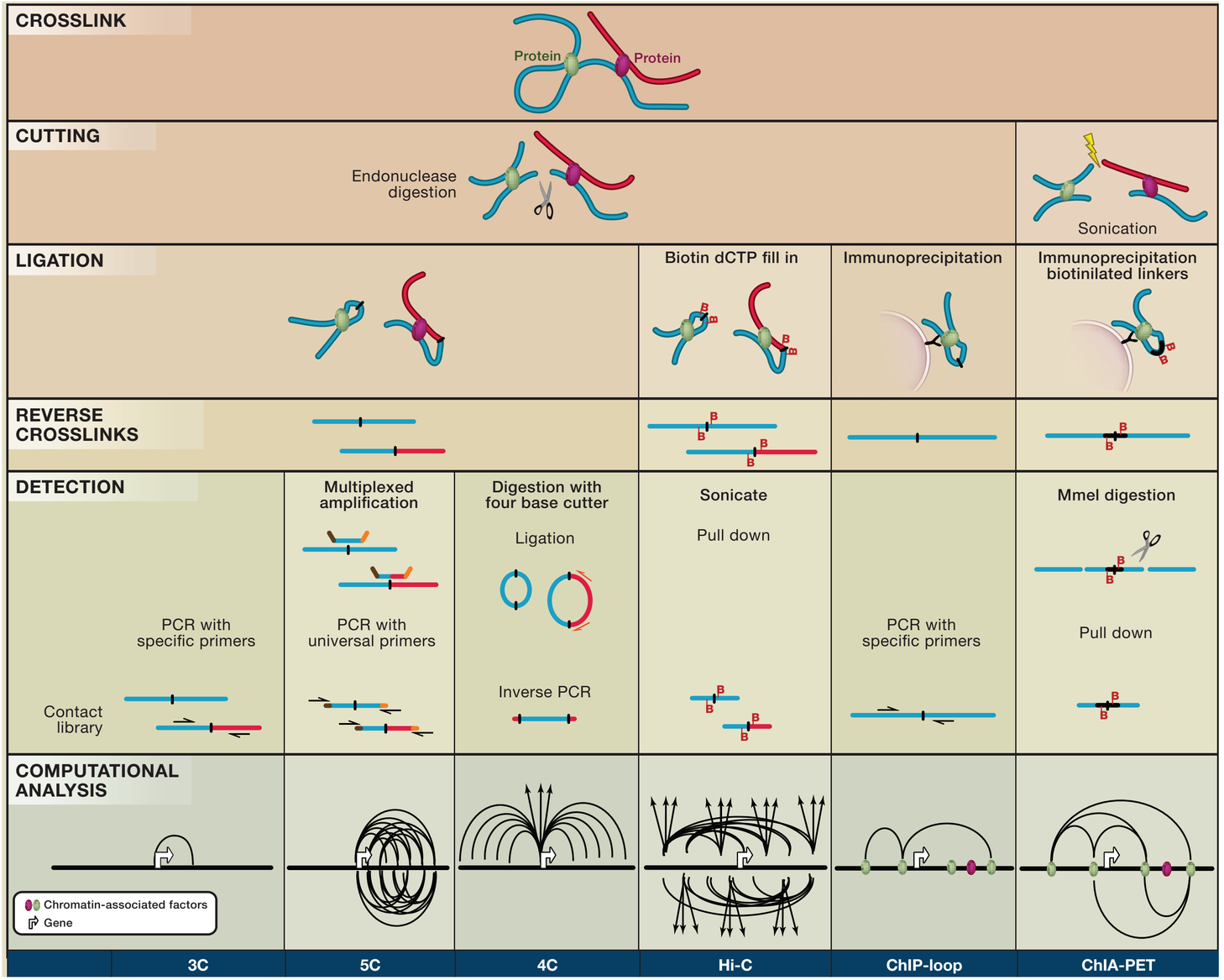
Complex genome organization

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



Marina Corral

Chromosome Conformation Capture



Hakim, O., & Misteli, T. (2012). SnapShot: Chromosome Confirmation Capture. Cell, 148(5), 1068–1068.e2.

ARTICLE

doi:10.1038/nature12593

Single-cell Hi-C reveals cell-to-cell variability in chromosome structure

Takashi Nagano[1]*, Yaniv Lubling[2]*, Tim J. Stevens[3]*, Stefan Schoenfelder[1], Eitan Yaffe[2], Wendy Dean[4], Ernest D. Laue[3], Amos Tanay[2] & Peter Fraser[1]

LETTER

doi:10.1038/nature20158

Capturing pairwise and multi-way chromosomal conformations using chromosomal walks

Pedro Olivares-Chauvet[1], Zohar Mukamel[1], Aviezer Lifshitz[1], Omer Schwartzman[1], Noa Oded Elkayam[1], Yaniv Lubling[1], Gintaras Deikus[2], Robert P. Sebra[2] & Amos Tanay[1]

nature genetics
ARTICLES
https://doi.org/10.1038/s41588-018-0161-5

Enhancer hubs and loop collisions identified from single-allele topologies

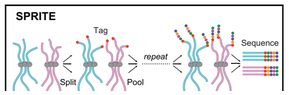
Amin Allahyar[1,2,7], Carlo Vermeulen[3,7], Britta A. M. Bouwman[3], Peter H. L. Krijger[3], Marjon J. A. M. Verstegen[3], Geert Geeven[3], Melissa van Kranenburg[3], Mark Pieterse[3], Roy Straver[1], Judith H. I. Haarhuis[4], Kees Jalink[5], Hans Teunissen[6], Ivo J. Renkens[1], Wigard P. Kloosterman[1], Benjamin D. Rowland[4], Elzo de Wit[4], Jeroen de Ridder[1]* and Wouter de Laat[3]*

Resource

Cell

Higher-Order Inter-chromosomal Hubs Shape 3D Genome Organization in the Nucleus

Graphical Abstract



Authors

Sofia A. Quinodoz, Noah Ollikainen, Barbara Tabak, ..., Patrick McDonel, Manuel Garber, Mitchell Guttman

Correspondence

mguttman@caltech.edu

nature COMMUNICATIONS

ARTICLE

DOI: 10.1038/s41467-018-06961-0 OPEN

Chromatin conformation analysis of primary patient tissue using a low input Hi-C method

Noelia Díaz[1], Kai Kruse[1], Tabea Erdmann[2], Annette M. Staiger[3,4,5], German Ott[3], Georg Lenz[2] & Juan M. Vaquerizas[1]

Article | Published: 11 February 2021

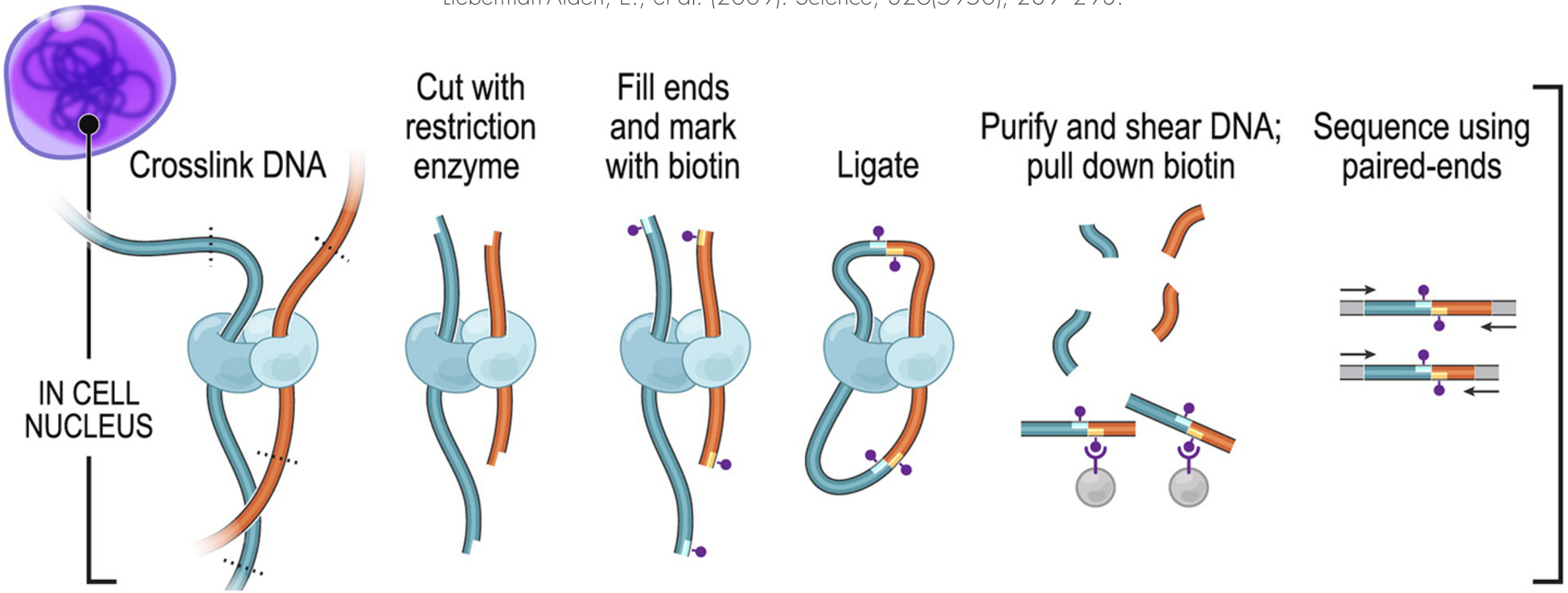
Liquid chromatin Hi-C characterizes compartment-dependent chromatin interaction dynamics

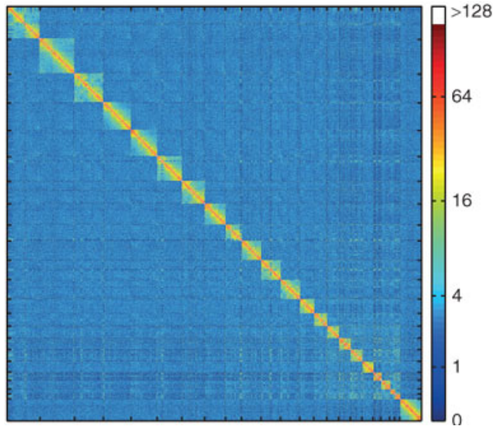
Houda Belaghzal, Tyler Borrman, Andrew D. Stephens, Denis L. Lafontaine, Sergey V. Venev, Zhiping Weng, John F. Marko & Job Dekker

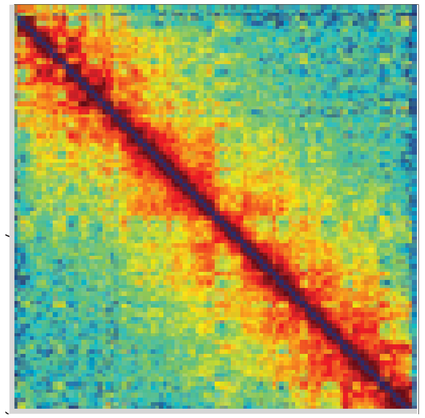
Nature Genetics **53**, 367–378 (2021) | Cite this article

Chromosome Conformation Capture

Dekker, J., Rippe, K., Dekker, M., & Kleckner, N. (2002). Science, 295(5558), 1306–1311.
Lieberman-Aiden, E., et al. (2009). Science, 326(5950), 289–293.







Hi-C 3.0

Akgol Oksuz, et al. Nature Methods 2021

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

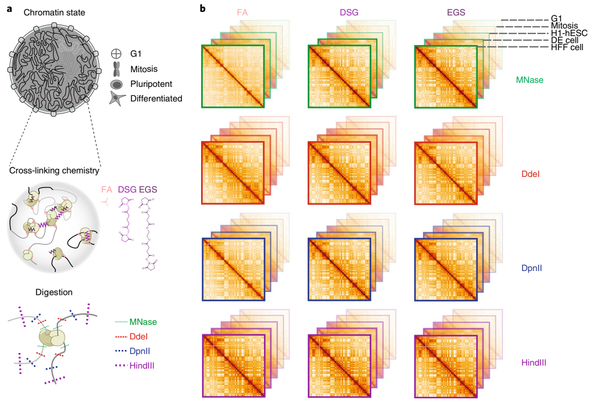
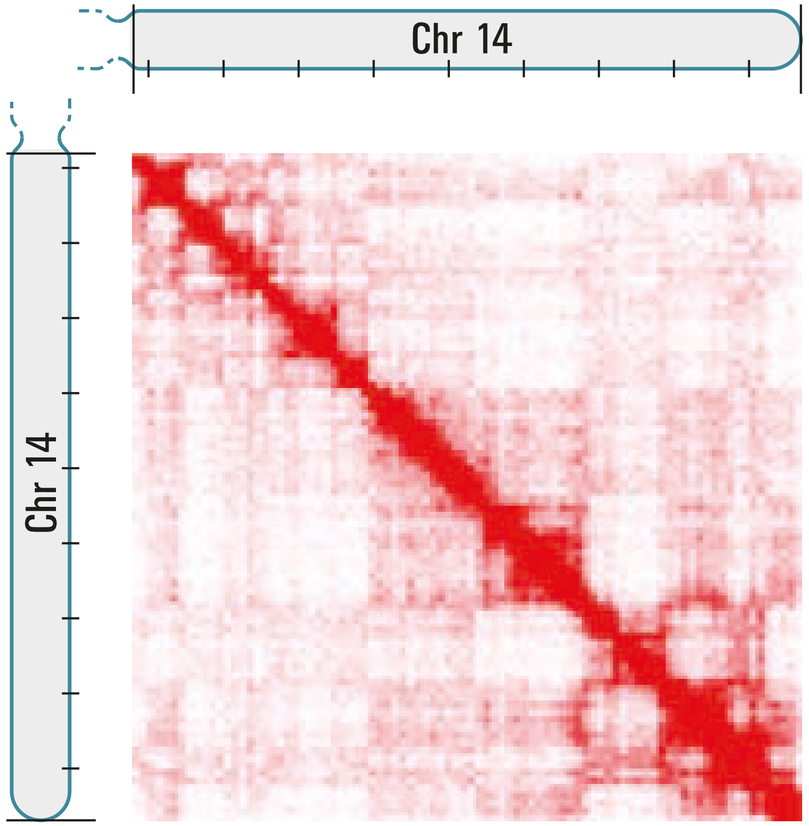
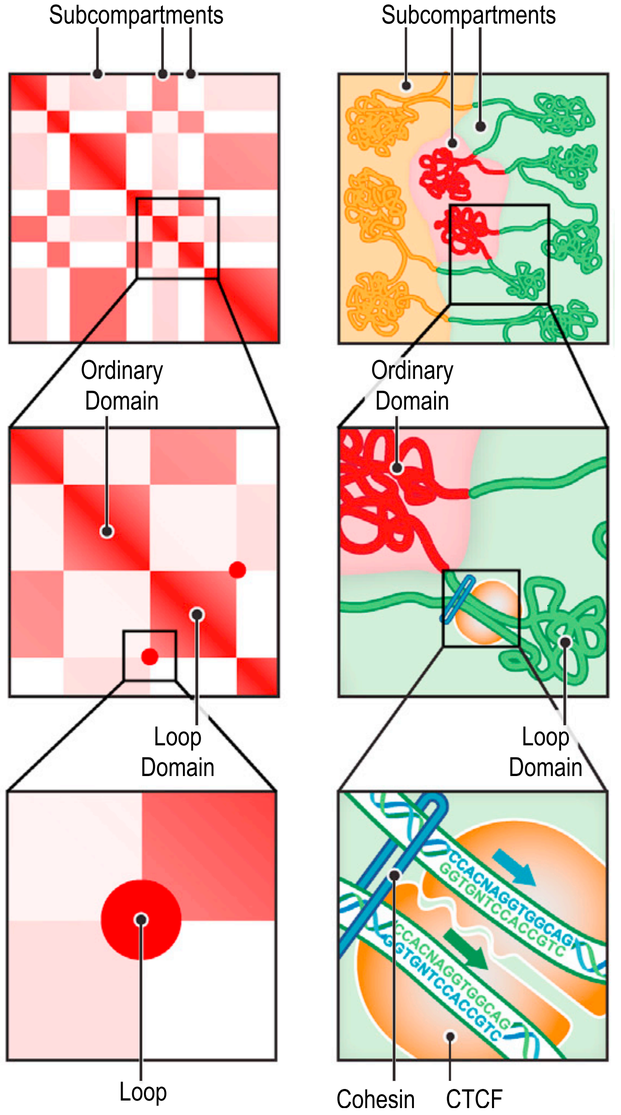


Fig. 1 | Outline of the experimental design. a, Experimental design for conformation capture for various cells, cross-linkers and enzymes. **b**, Representation of interaction maps from experiments in **a**.

Hierarchical genome organisation

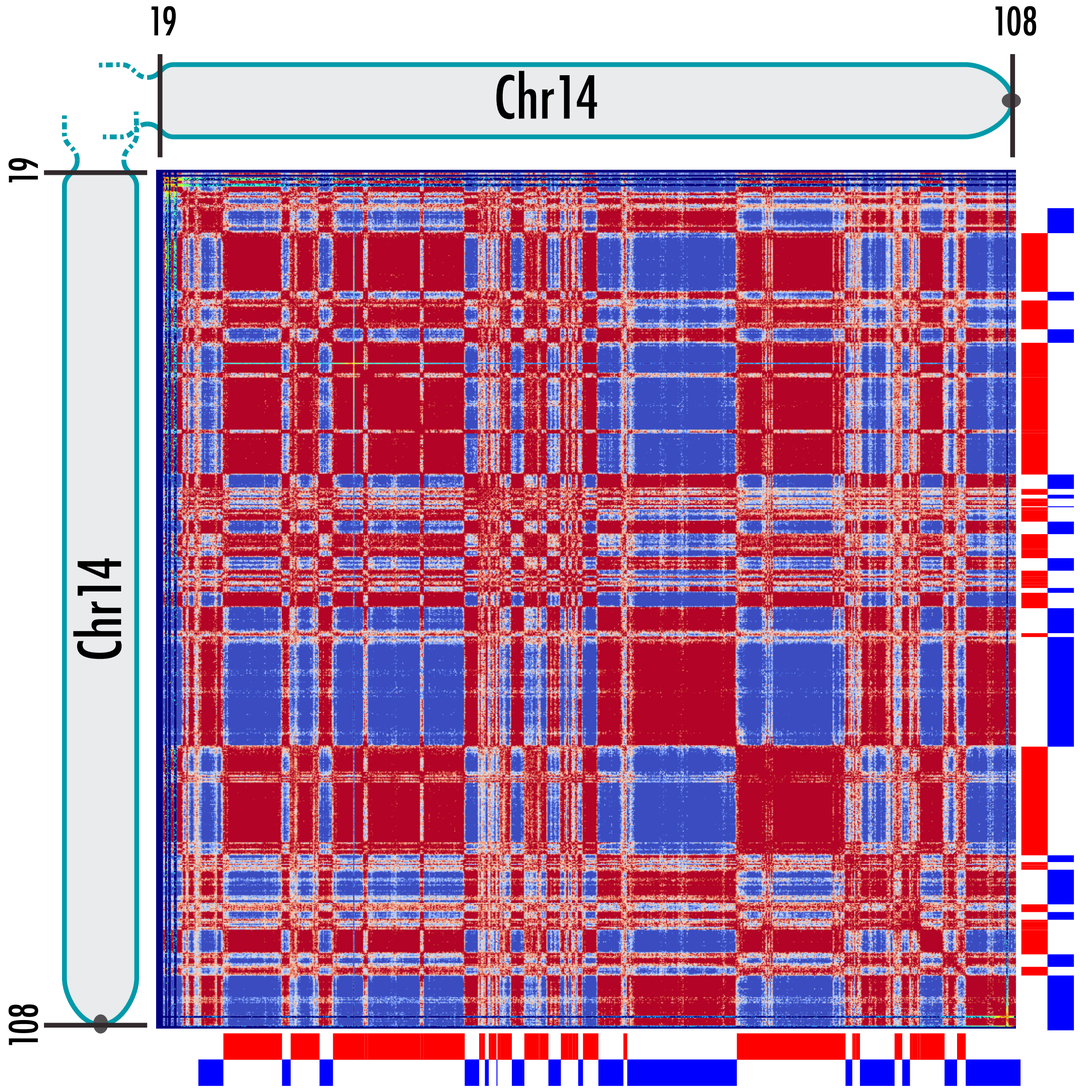
Lieberman-Aiden, E., et al. (2009). Science, 326(5950), 289–293.
Rao, S. S. P., et al. (2014). Cell, 1–29.





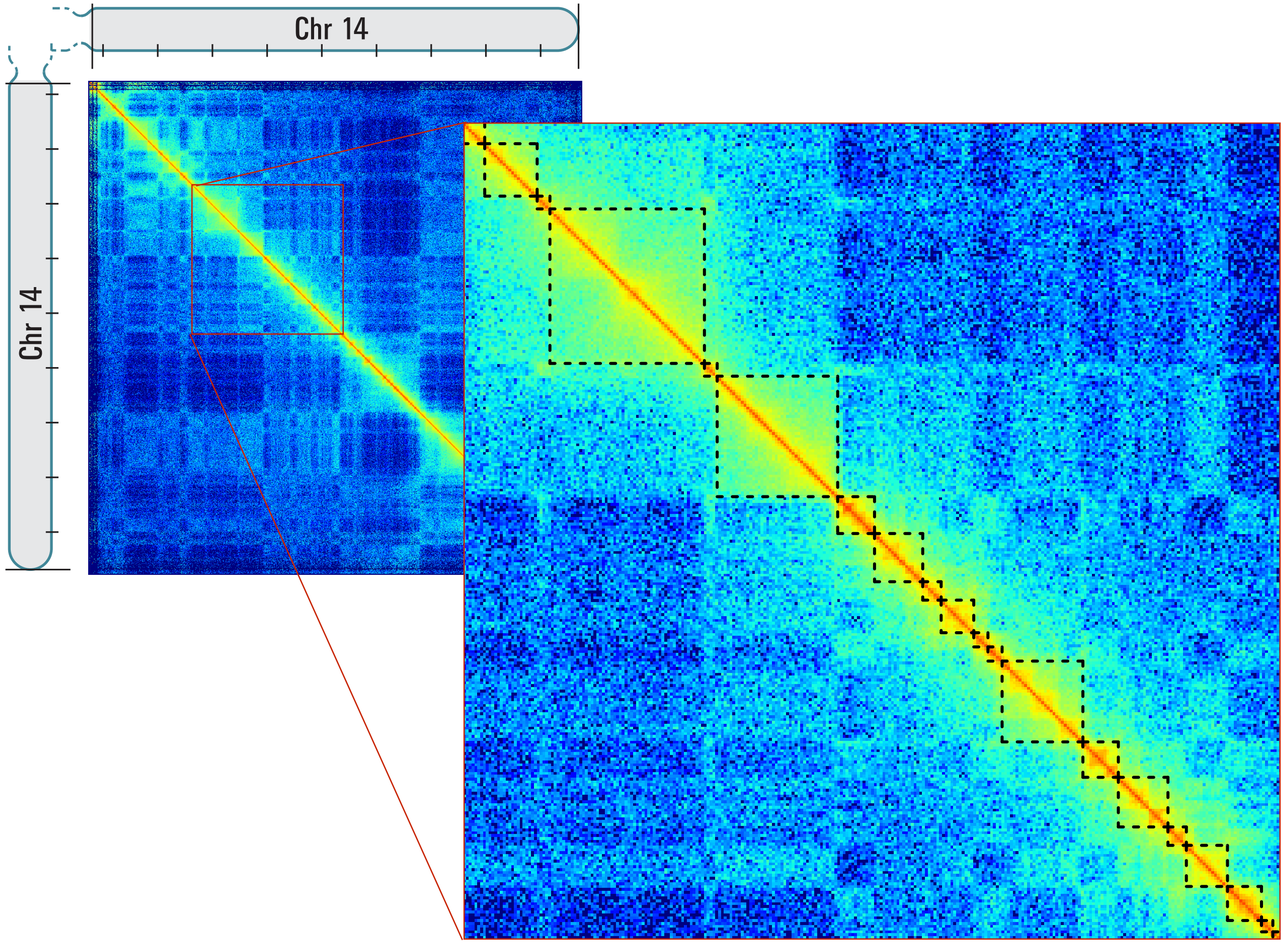
A/B Compartment

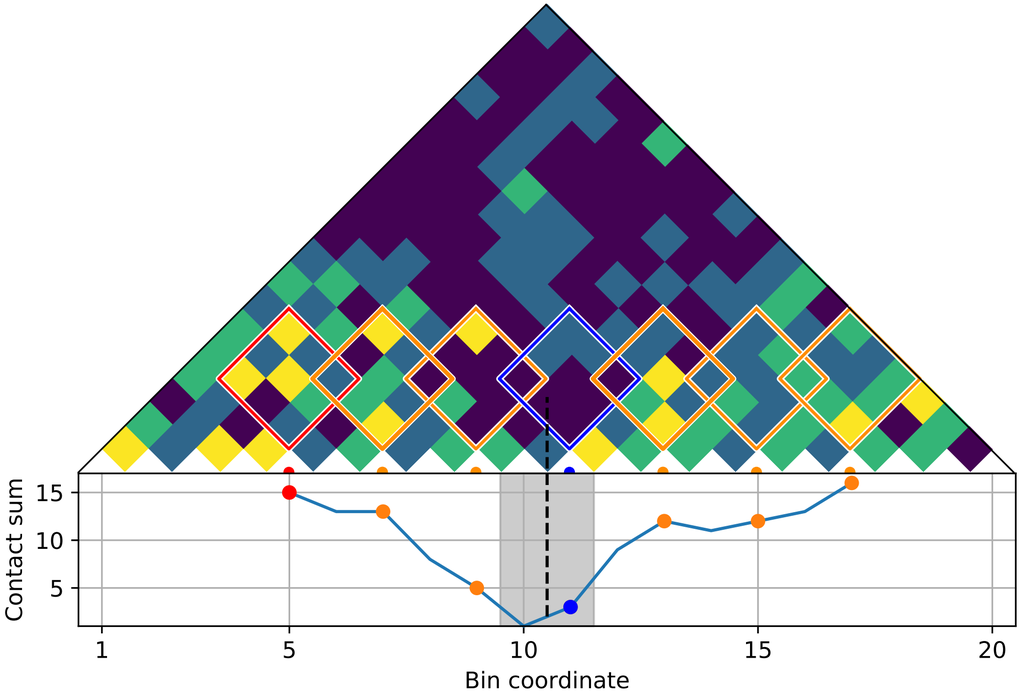
Chromosome 14



TADs

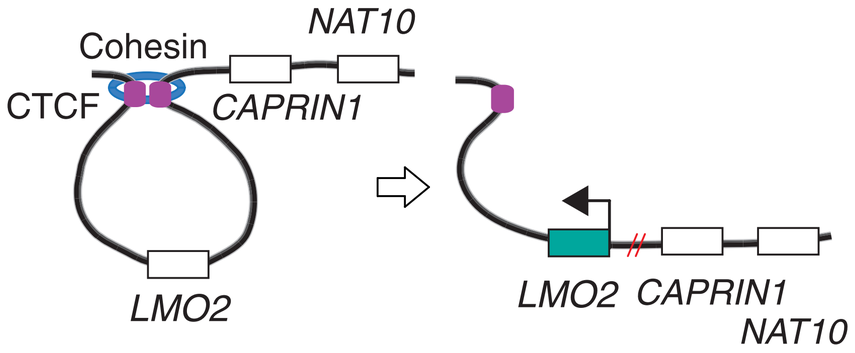
Chromosome 14

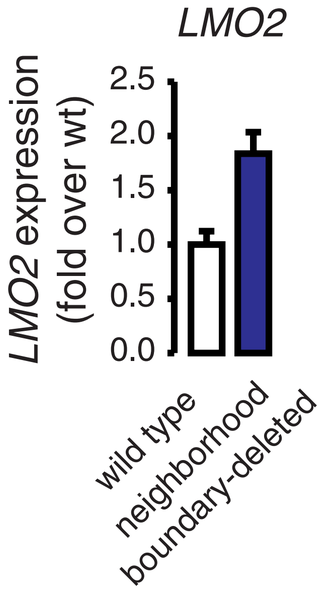


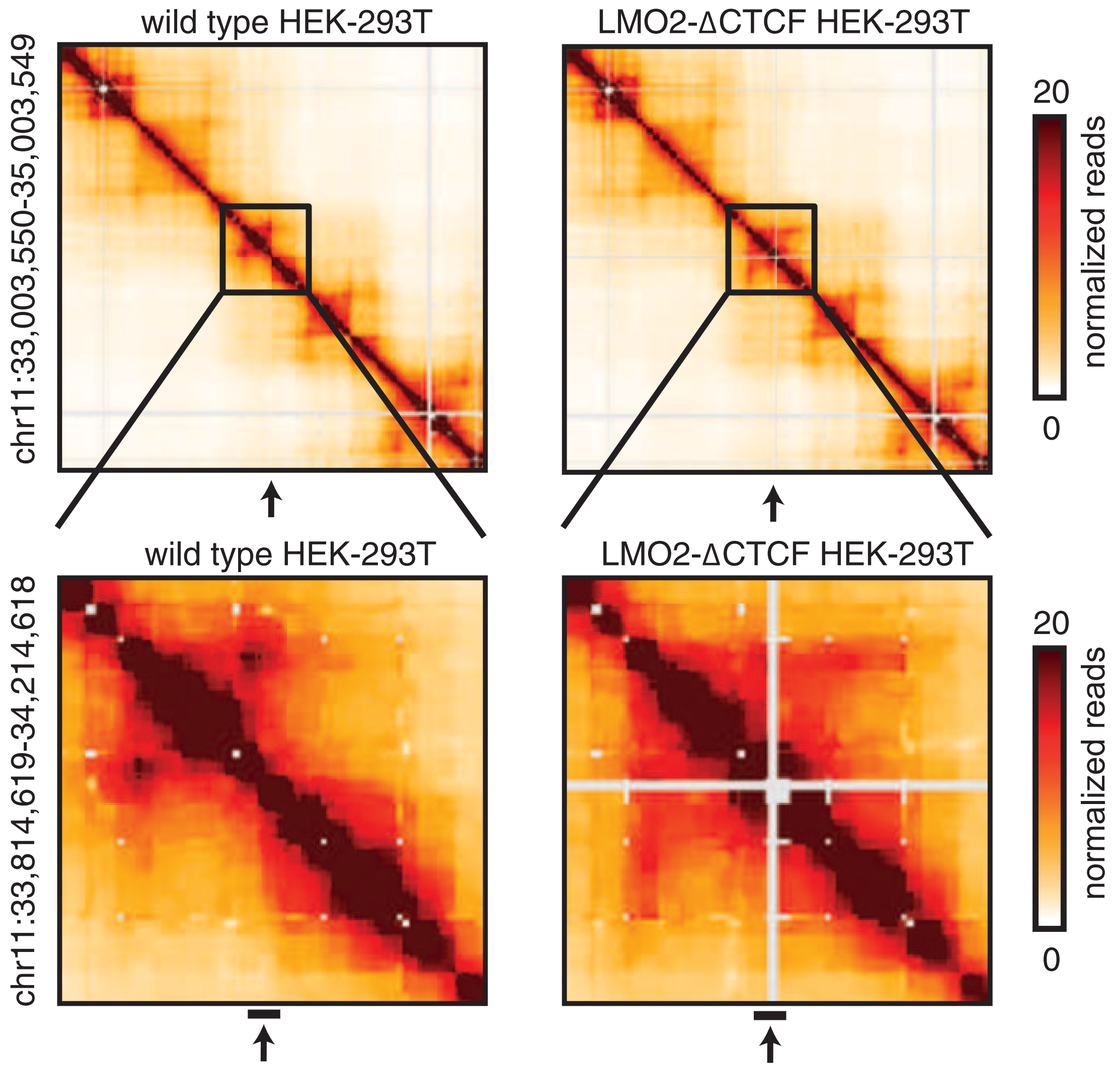


TADs are functional units

Hnisz, D., et al. (2016). Science

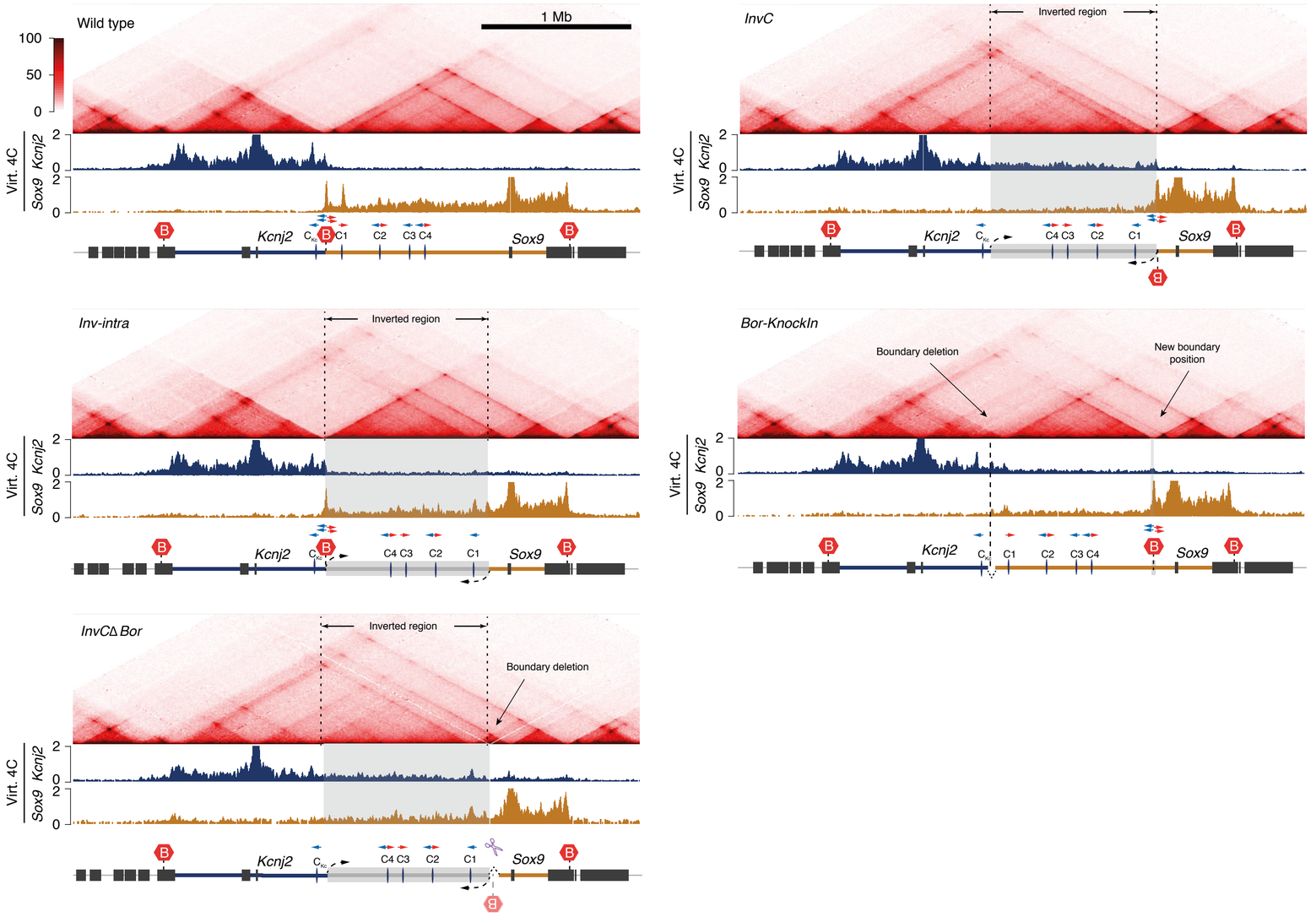






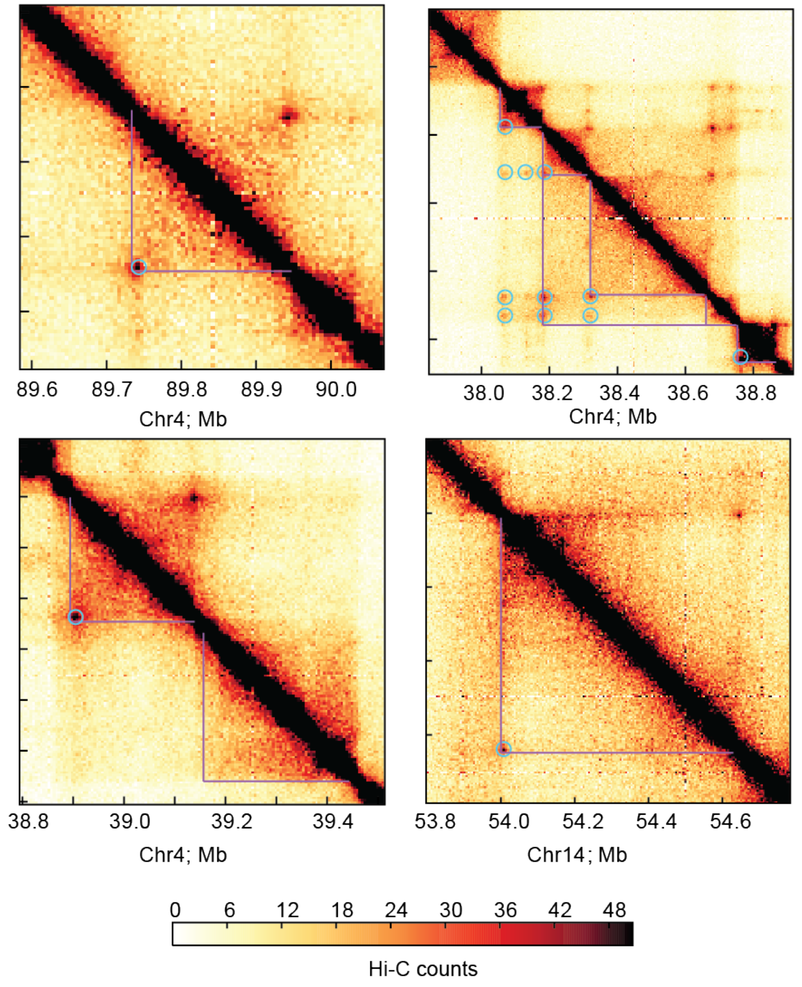
TADs are functional units

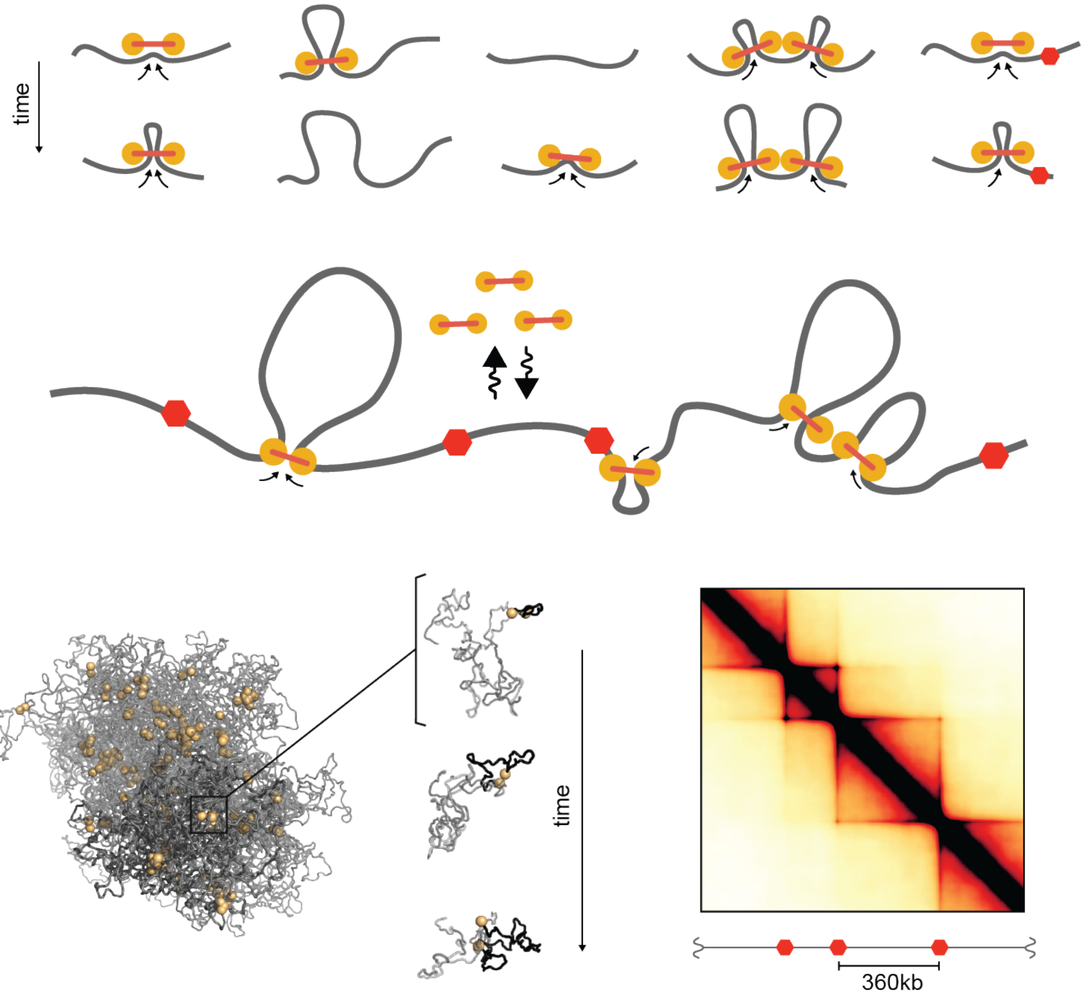
Despang, et al. (2019). Nature Genetics 51,1263–1271 (2019)



Loop-extrusion as a TAD forming mechanism

Fudenberg, G., Imakaev, M., Lu, C., Goloborodko, A., Abdennur, N., & Mirny, L. A. (2018).
Cold Spring Harb Symp Quant Biol 2017. 82: 45-55





Spatio-temporal regulatory landscape of sex-determination

or a case for:

Structure-Based Genome Editing (SBGE) guided by genome spatial auto-correlation analysis

Marc A. Marti-Renom
CNAG · CRG · ICREA

BioRxiv

http://marciuslab.org
http://3DGenomes.org

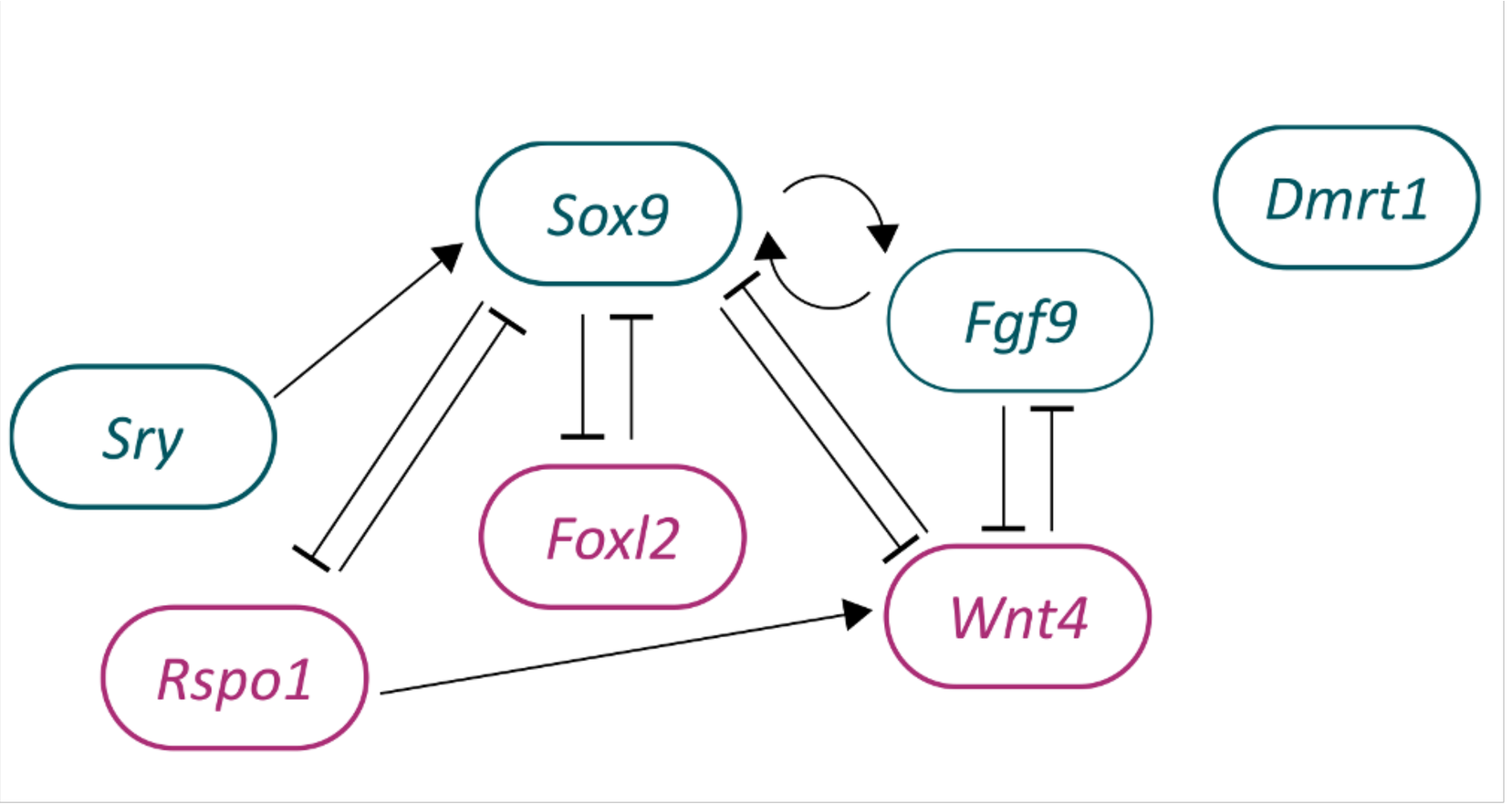
cnag CRG ICREA

Juan A. Rodríguez
CNAG-CRG
now @GLobe I. Denmark

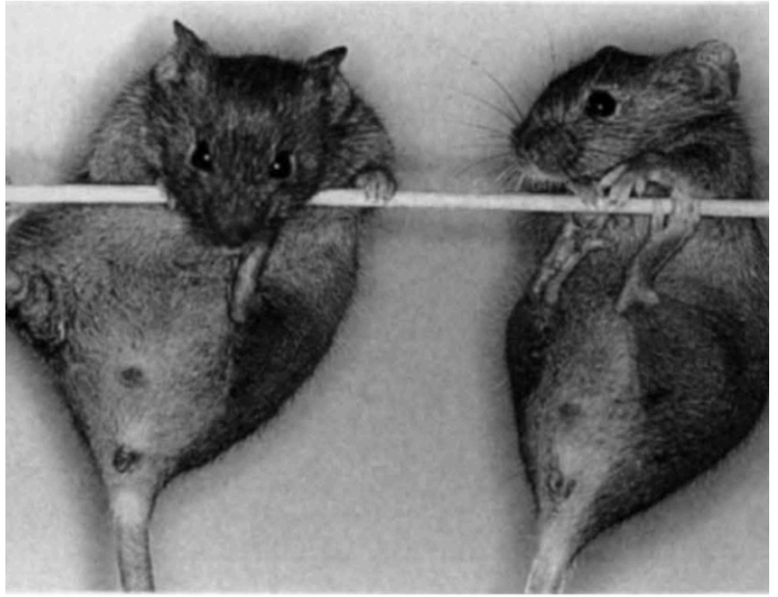
Irene Mota
MDC Berlin

Dario Lupiañez
MDC Berlin
Now @CABD

Sex determination: a 3,000 year-old enigma



Genetics

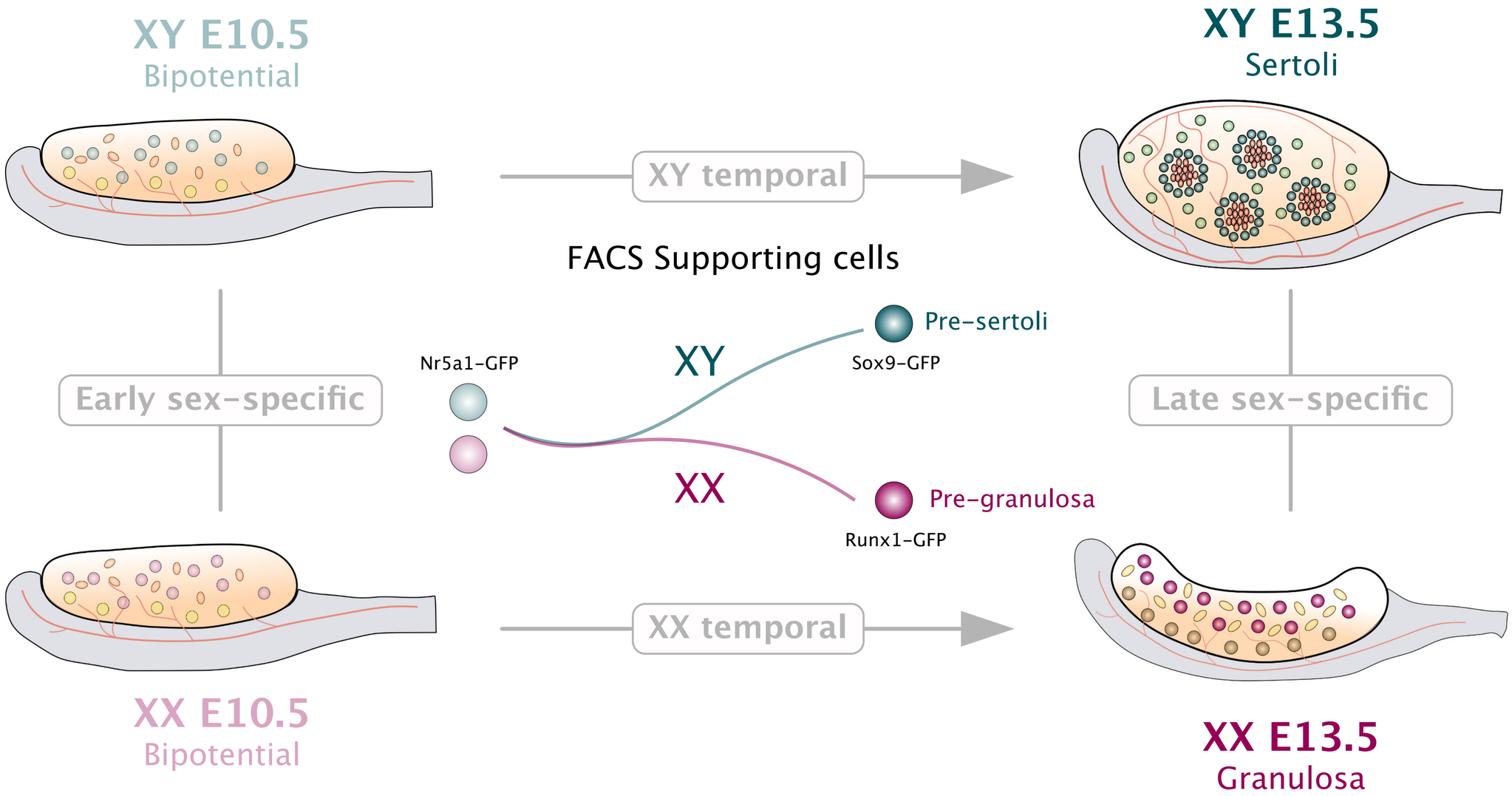


Discovery of *Sry* gene

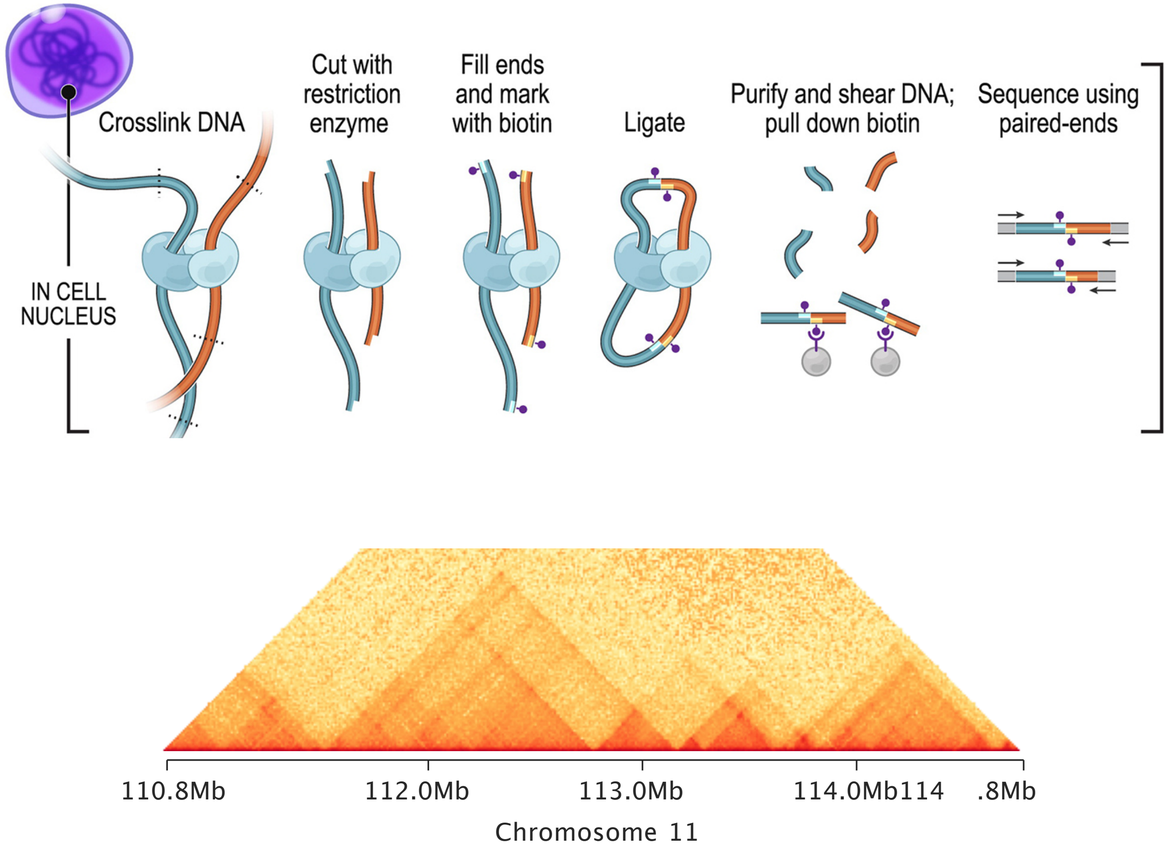
Koopman et al., Nature, 1991

(Goodfellow & Lovell-Badge labs)

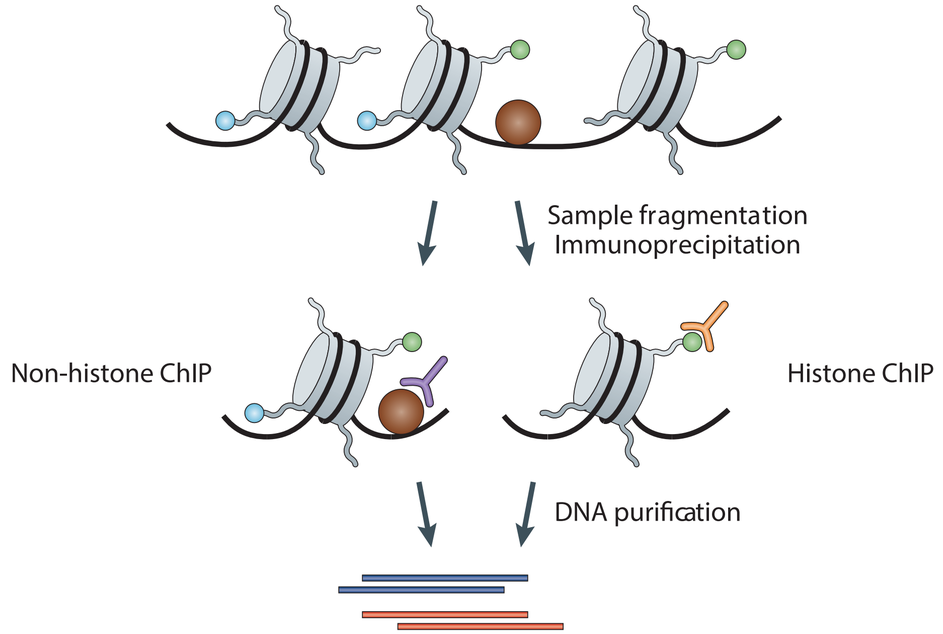
Sex-determination as a model for "bipotential" commitment

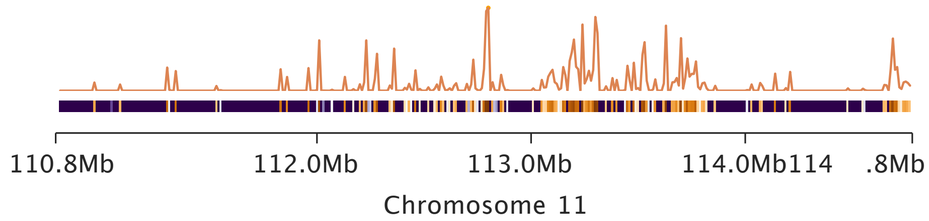


Hi-C & ChIP-seq



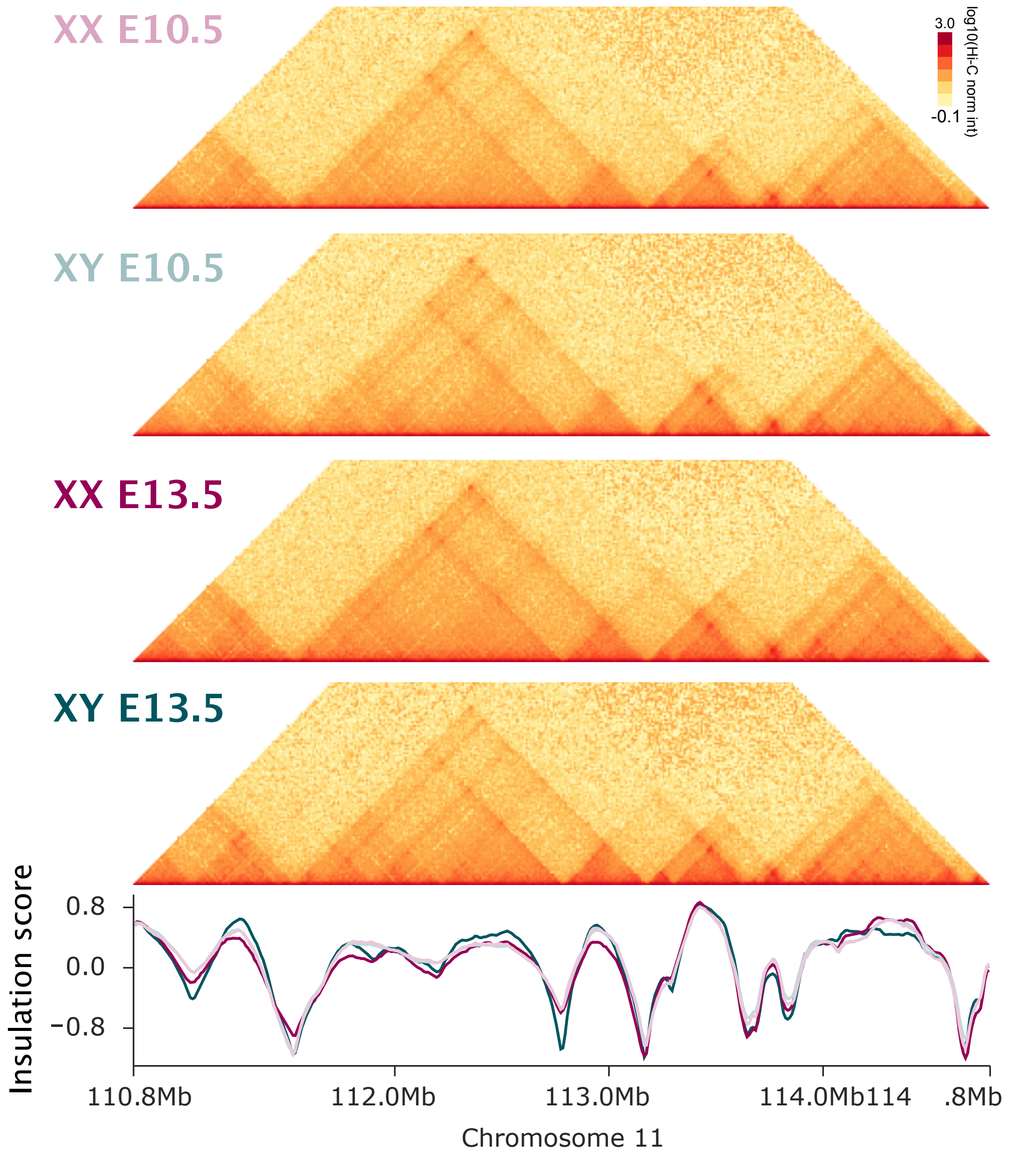
Dekker, J., Rippe, K., Dekker, M., & Kleckner, N. (2002). Science, 295(5558), 1306–1311.
Lieberman-Aiden, E., et al. (2009). Science, 326(5950), 289–293.

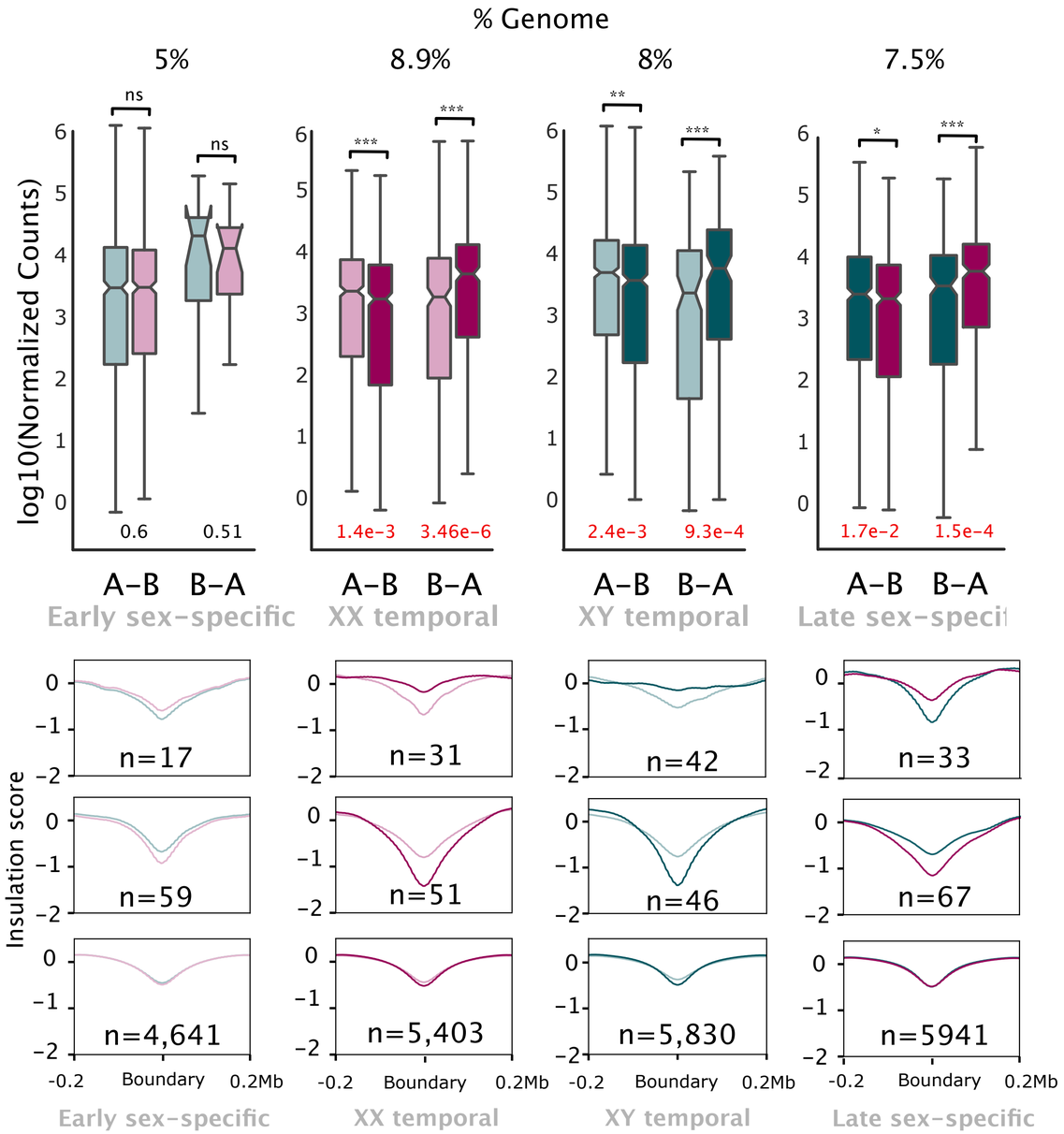


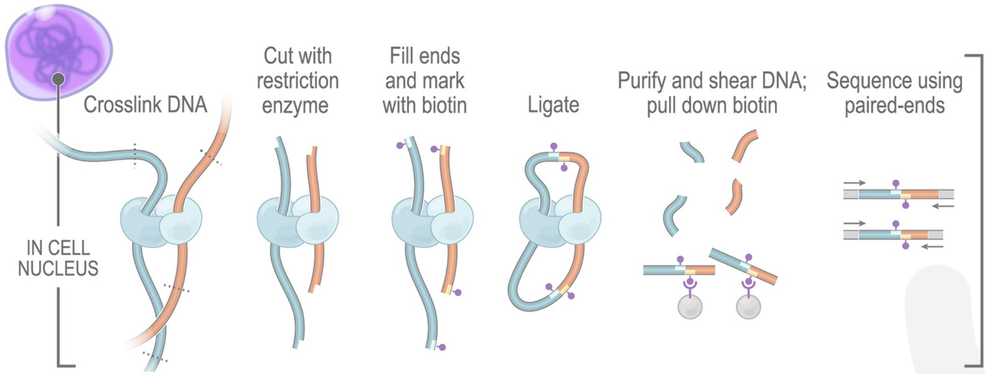


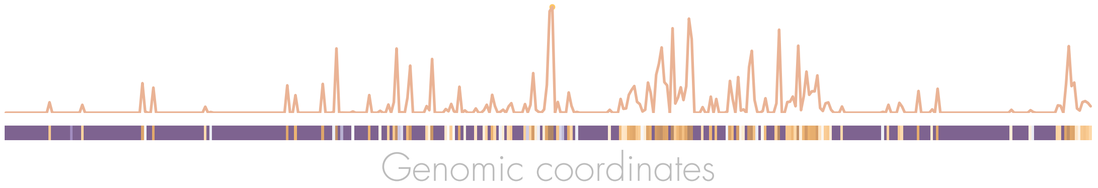
Solomon, M. J., Larsen, P. L. & Varshavsky, A. (1988) Cell 53, 937–947.
Park, P.J. (2009) Nature Reviews Genetics 10, 669–680.

No major structural (apparent) differences





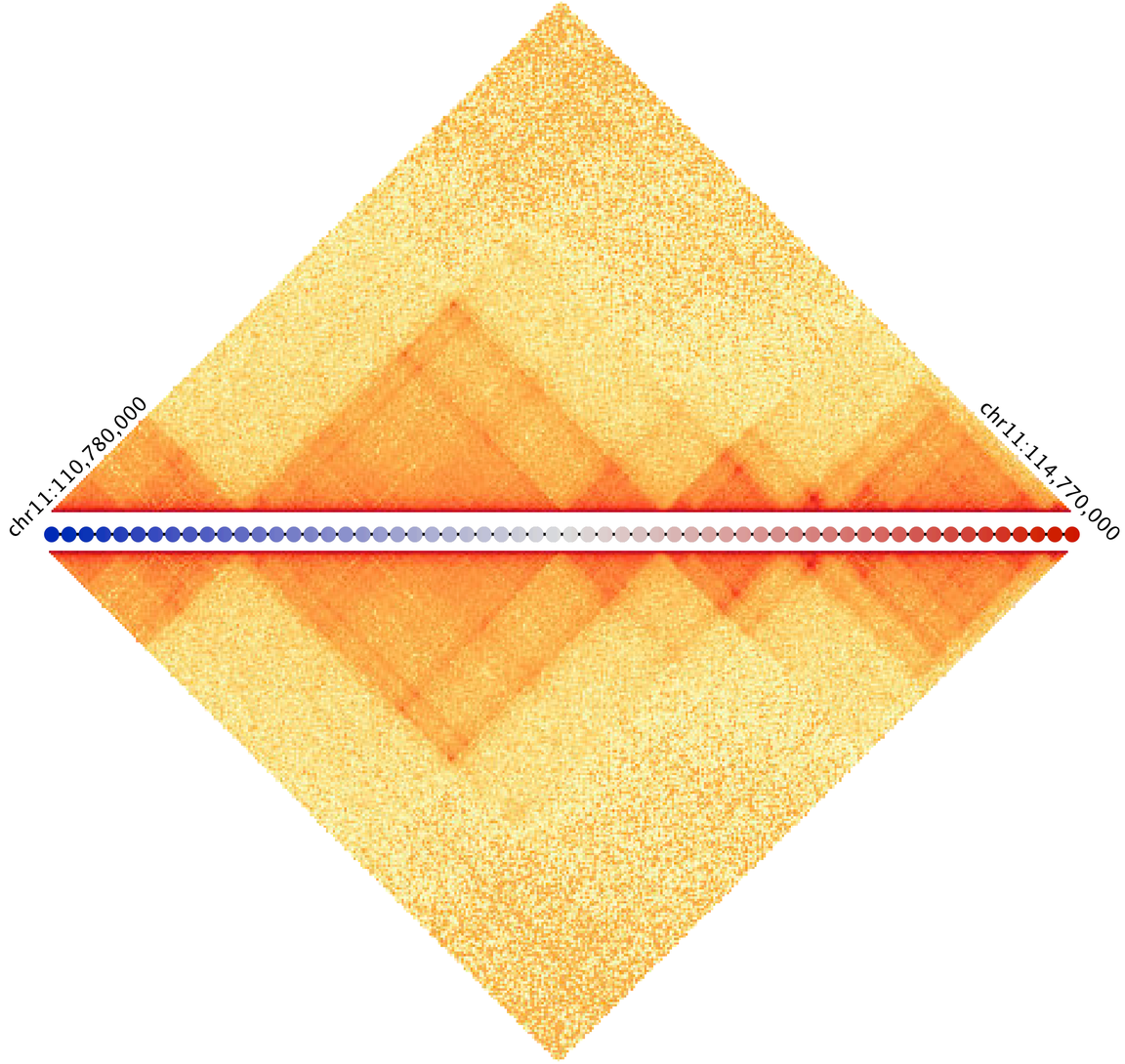

Crosslink DNA
IN CELL NUCLEUS
Cut with restriction enzyme
Fill ends and mark with biotin
Ligate
Purify and shear DNA; pull down biotin
Sequence using paired-ends

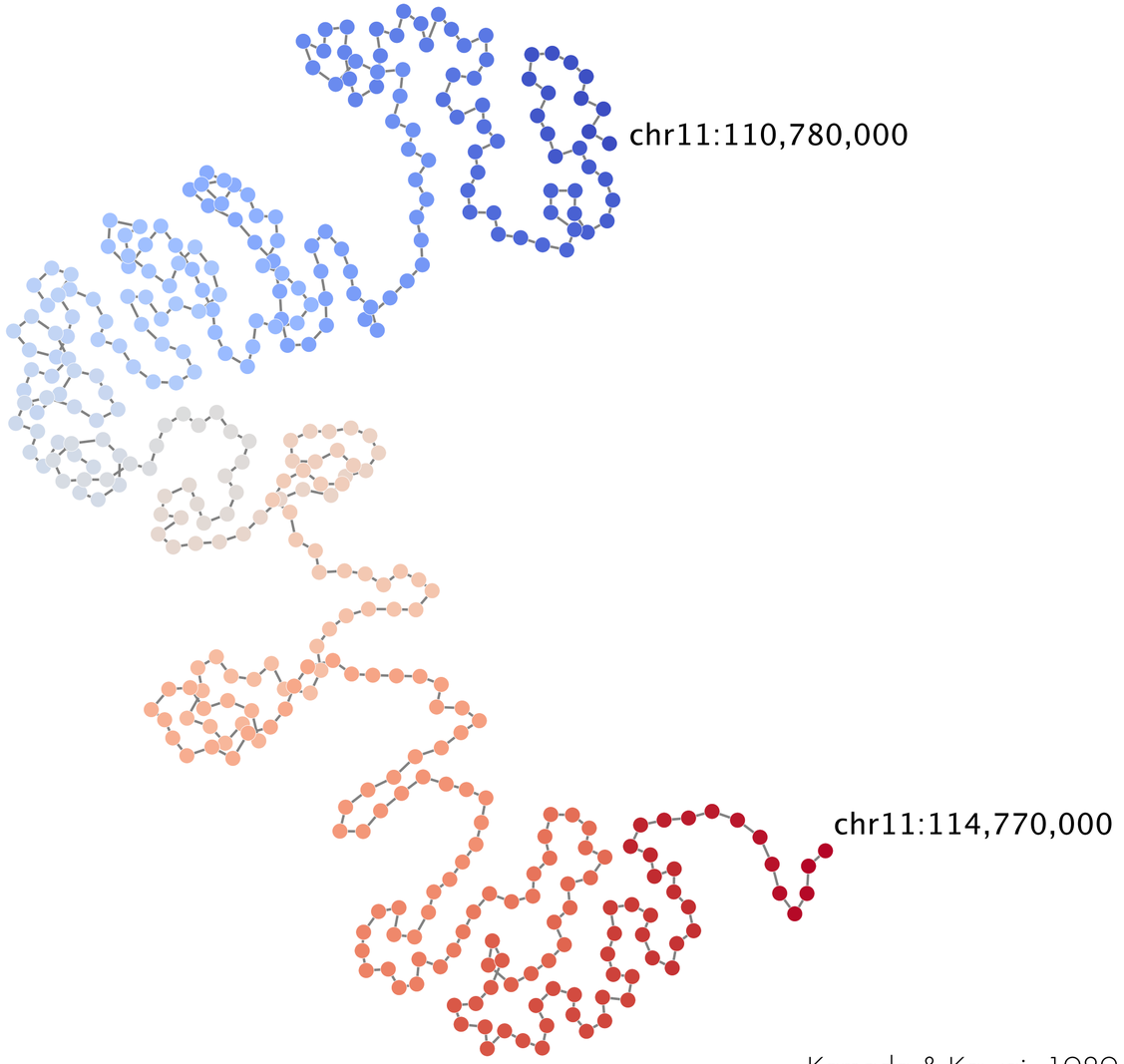

Genomic coordinates

METALoci

Spatial lay-out of significant interactions

chr11:110780000-114770000 @10kb

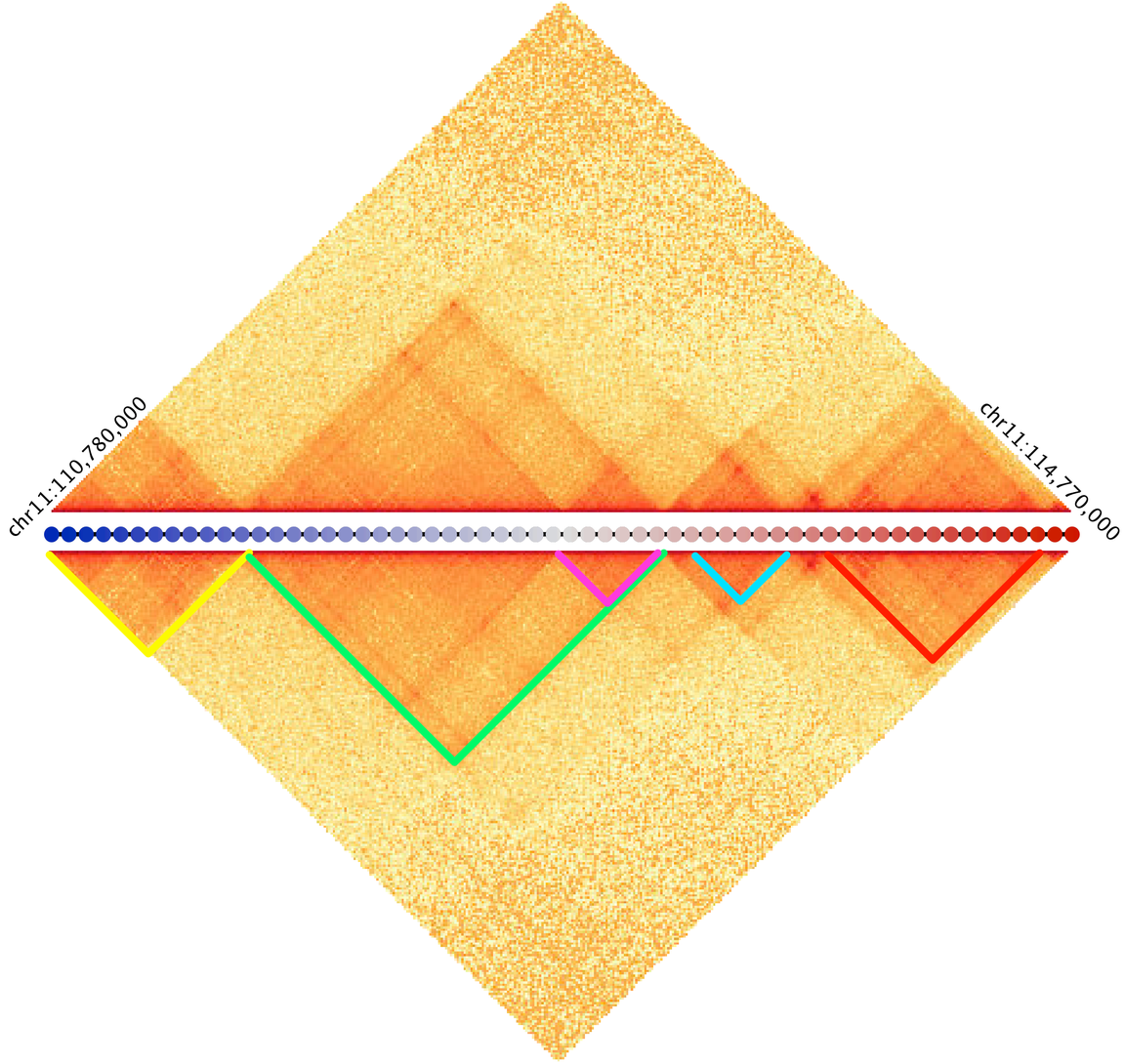


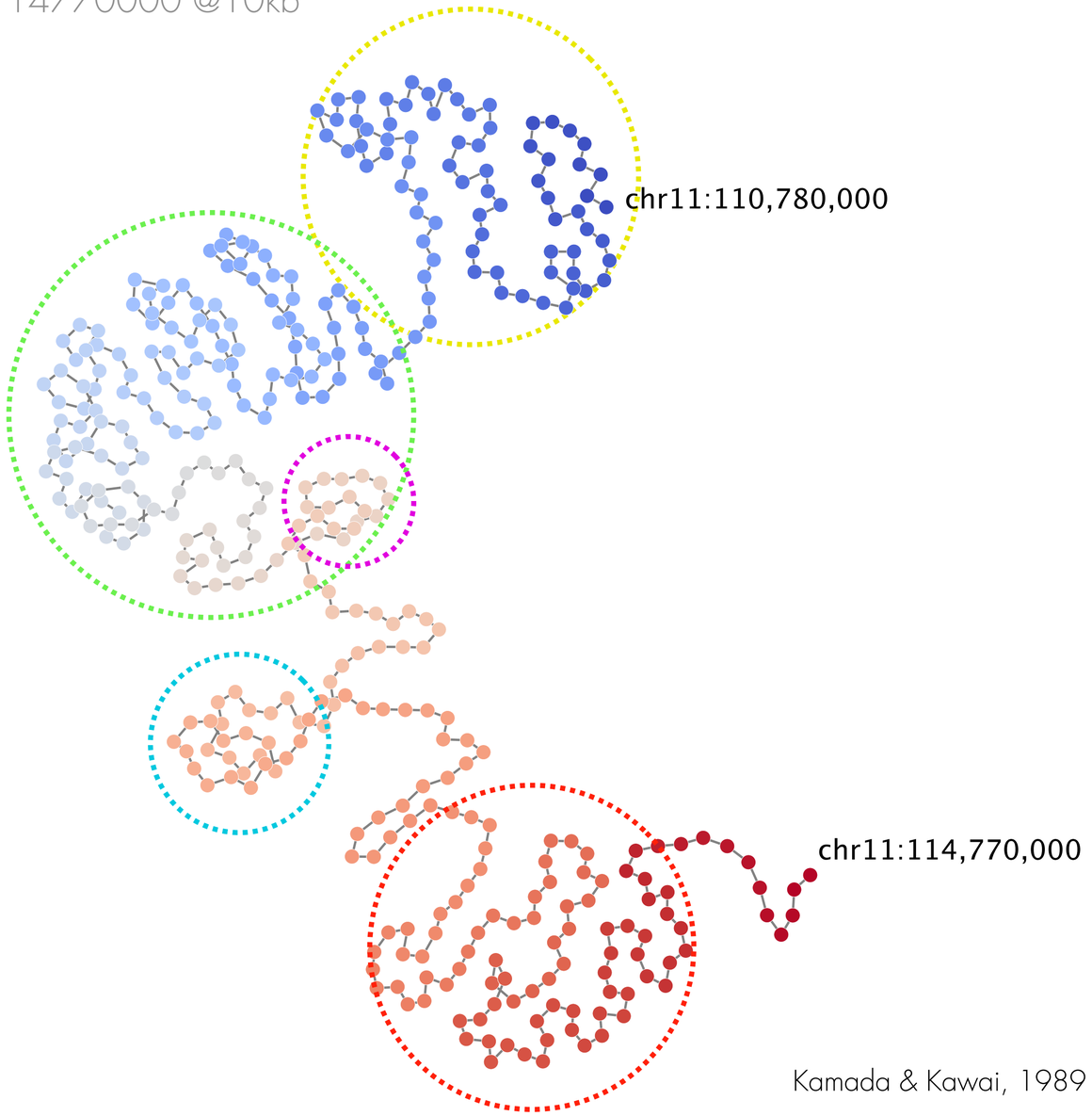


Kamada & Kawai, 1989

Spatial lay-out of significant interactions

chr11:110780000-114770000 @10kb

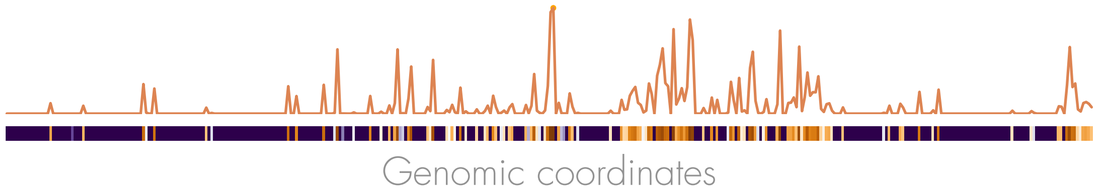


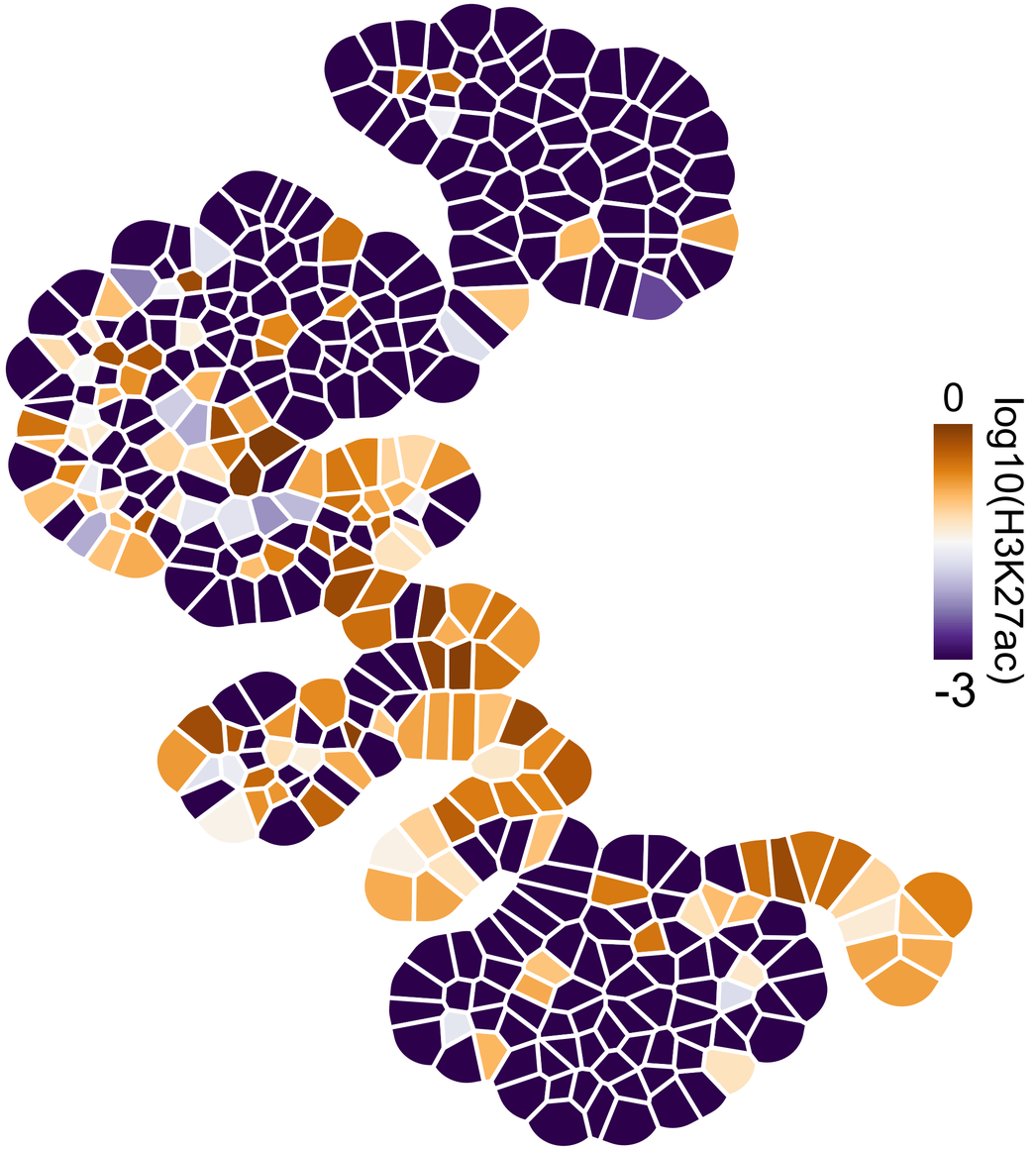


Kamada & Kawai, 1989

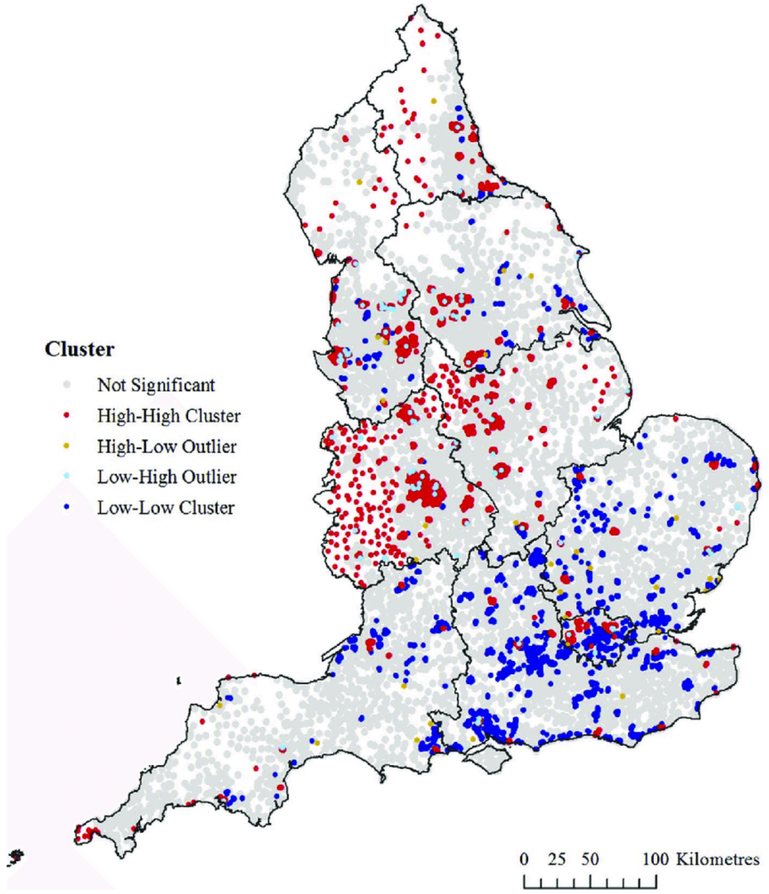
Marker (H3K27ac) into 2D mapping

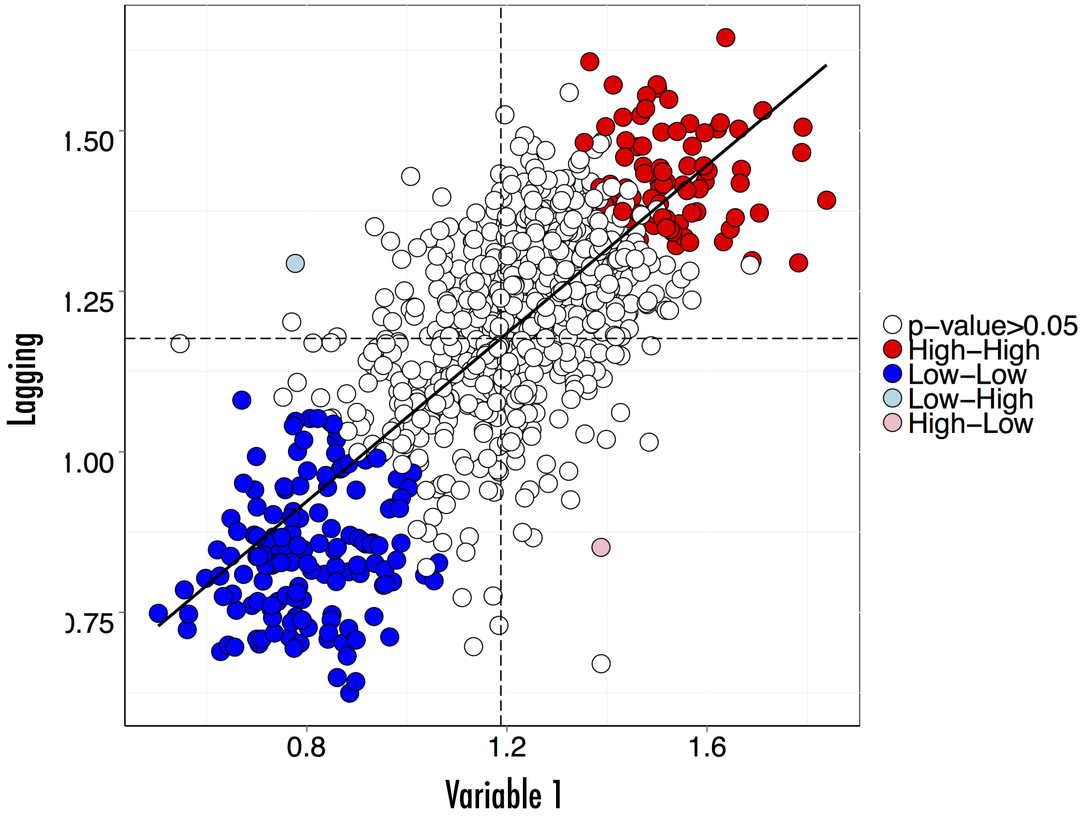
chr11:110780000-114770000



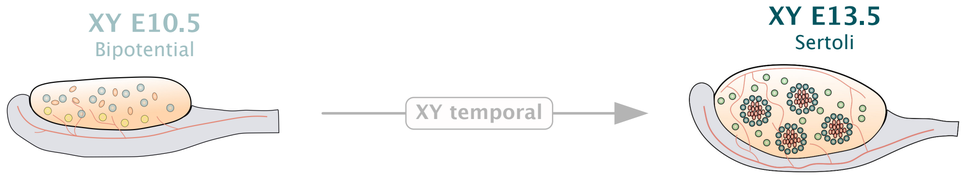


Local Moran Index

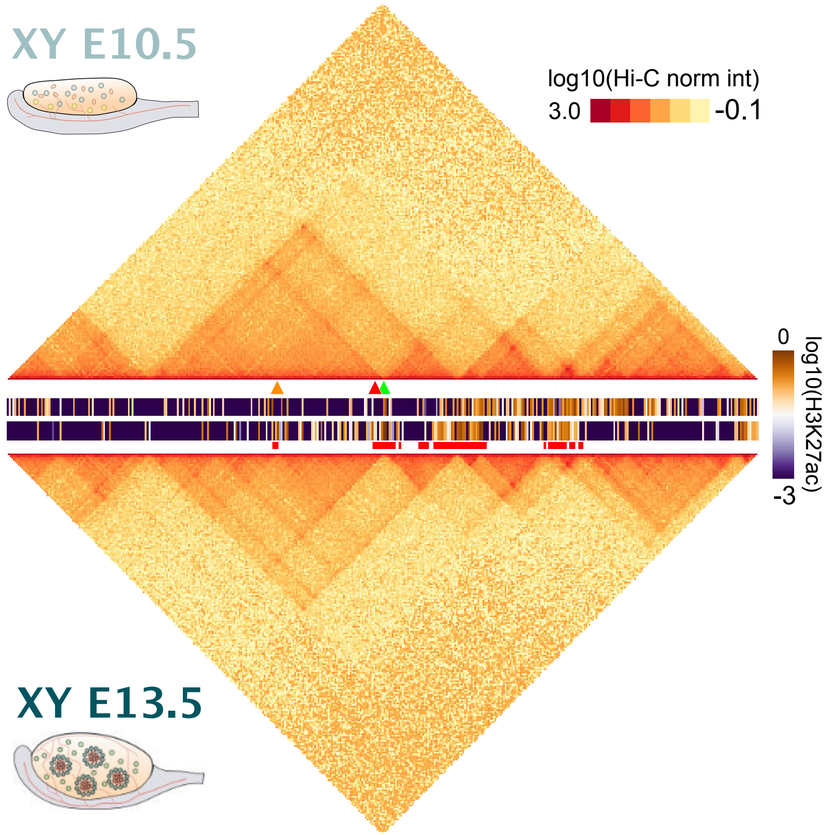


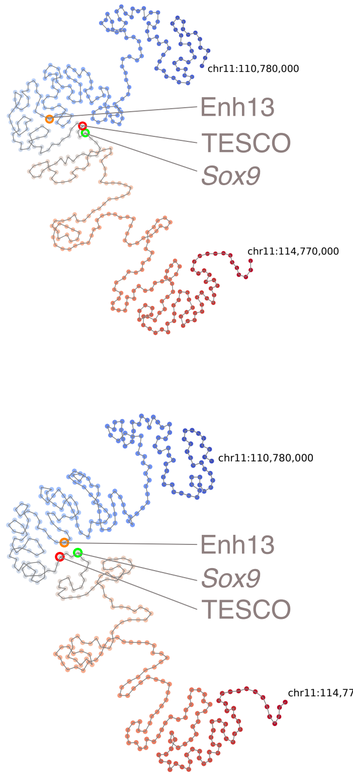


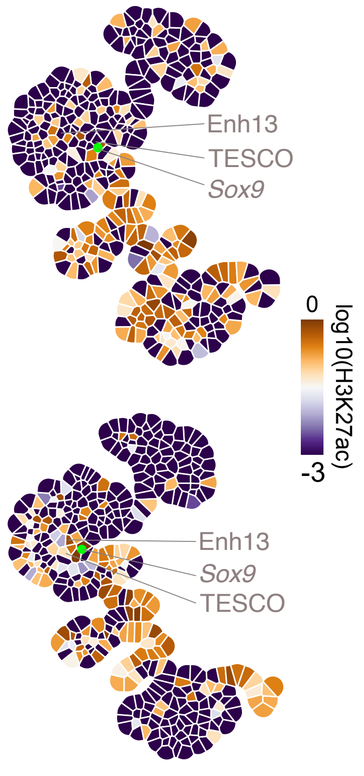
Quantifying regulatory environments bin by bin

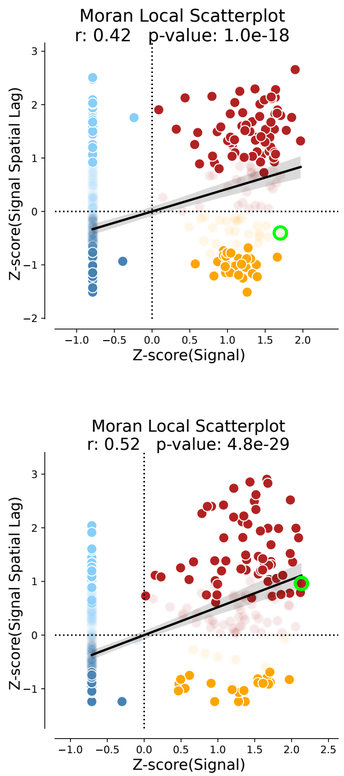


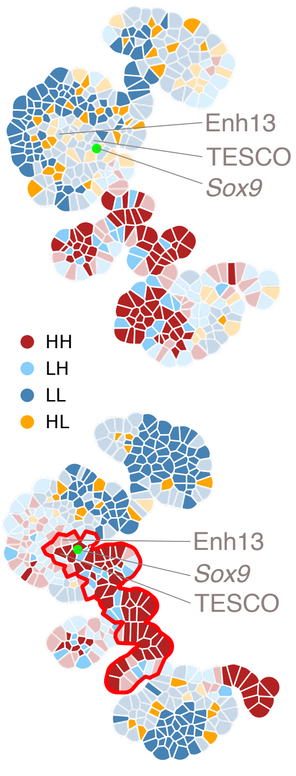
Sox9 locus chr11:110,780,000-114,770,000



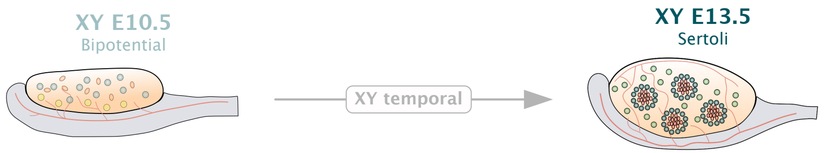


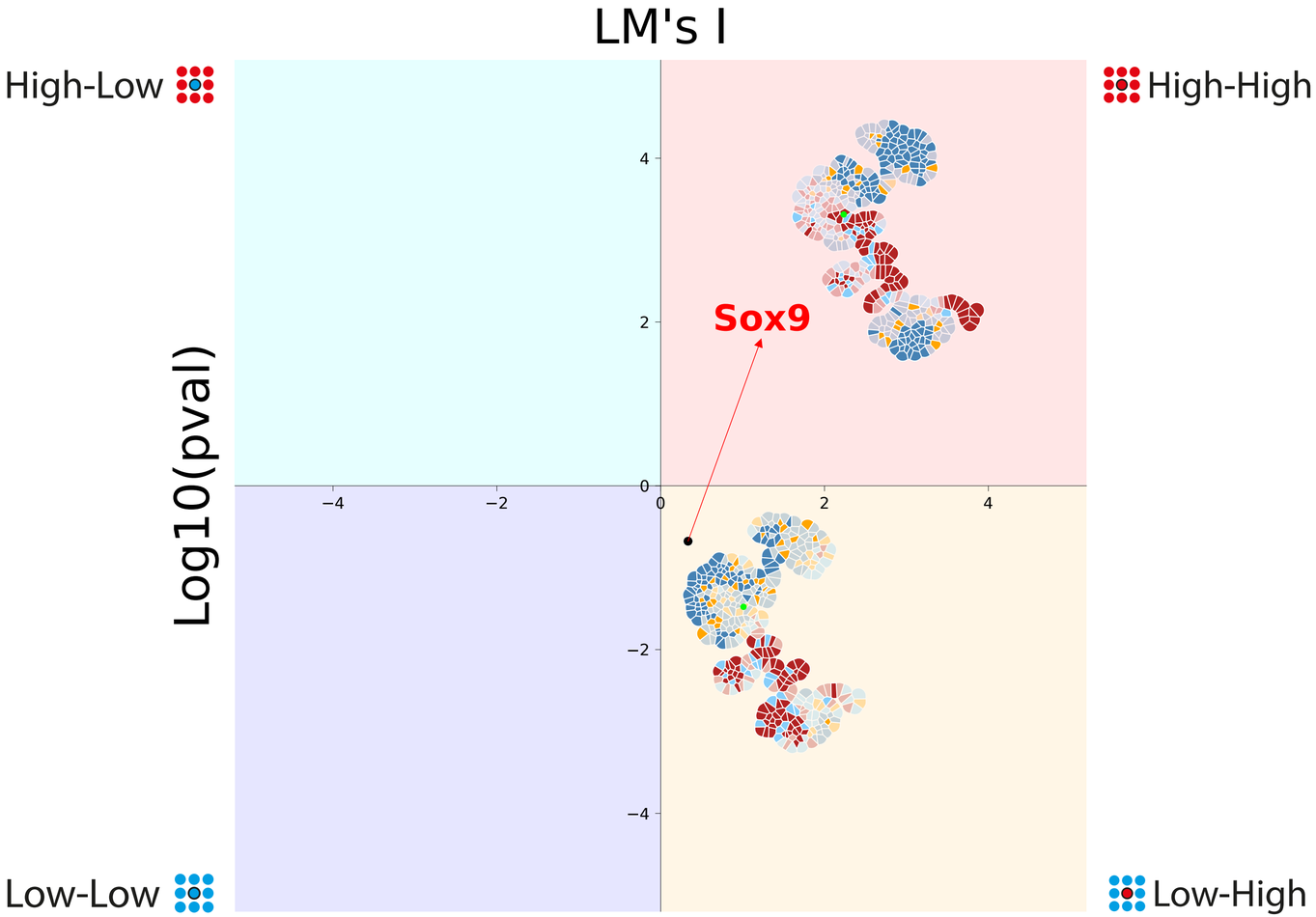




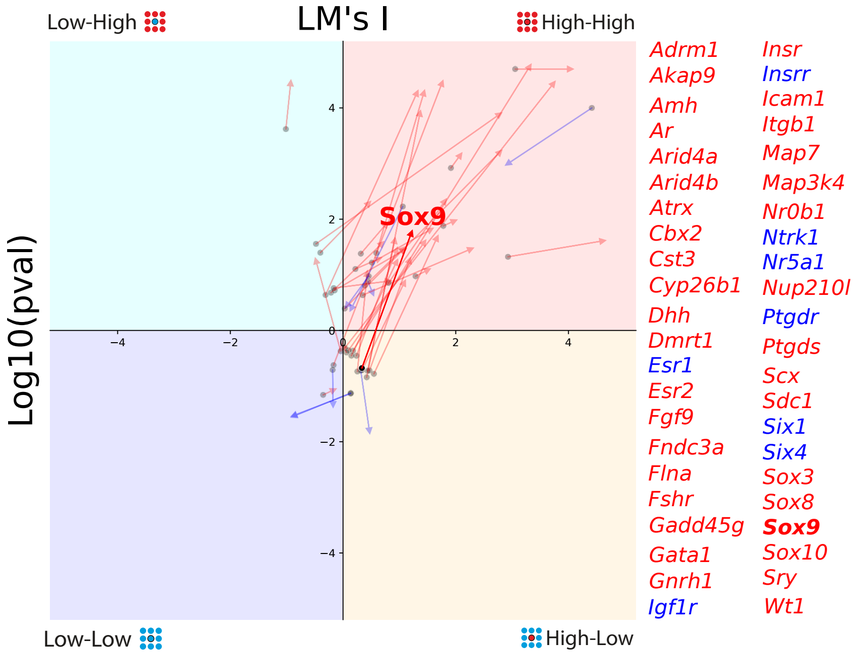


LMI Trip for Sox9 gene

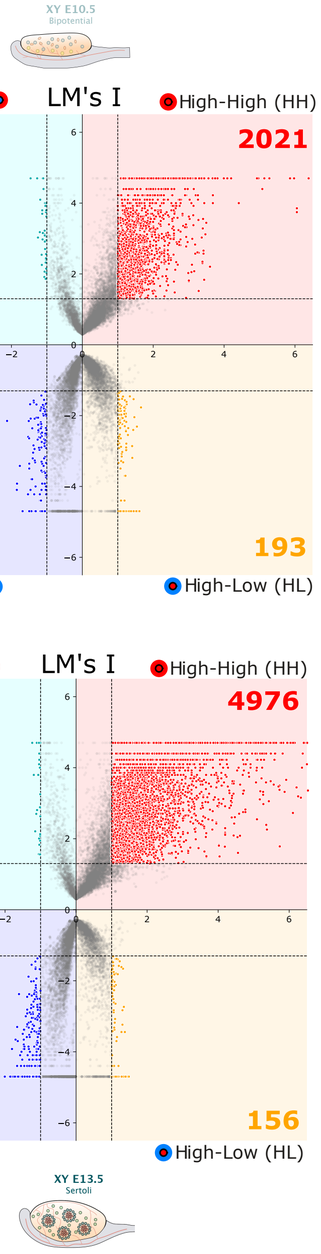


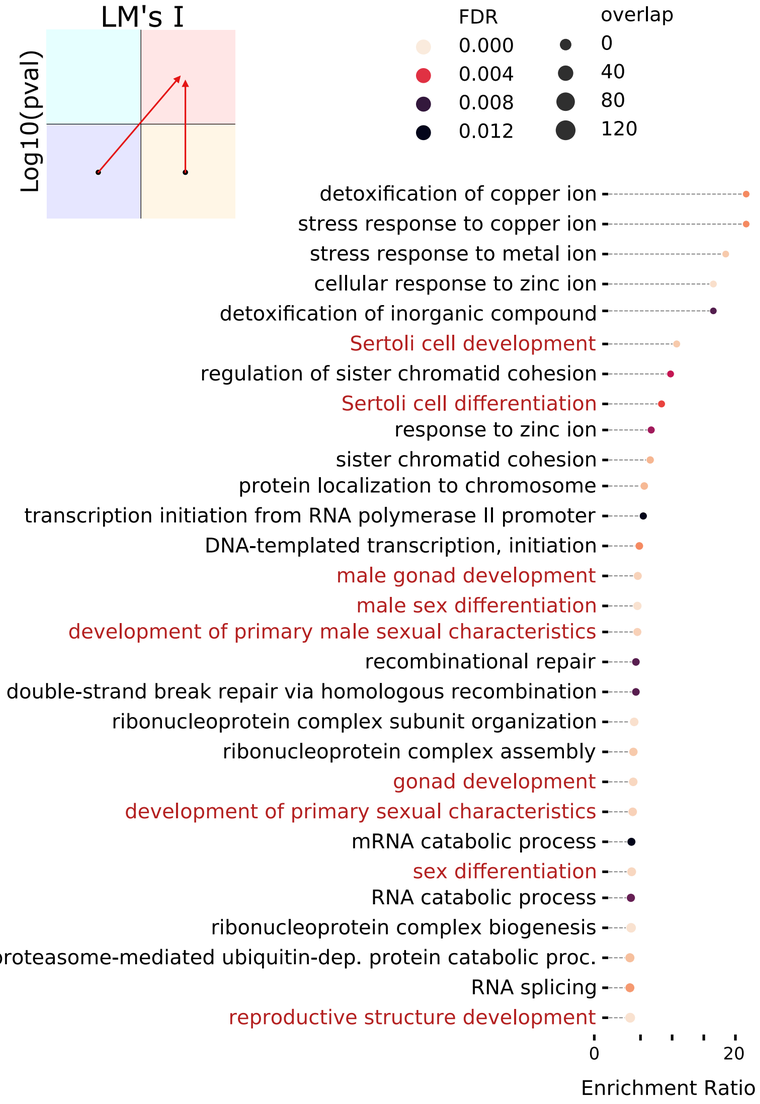


All genes LMI Trip



Adrm1
Akap9
Amh
Ar
Arid4a
Arid4b
Atrx
Cbx2
Cst3
Cyp26b1
Dhh
Dmrt1
Esr1
Esr2
Fgf9
Fndc3a
Flna
Fshr
Gadd45g
Gata1
Gnrh1
Igf1r
Insr
Insrr
Icam1
Itgb1
Map7
Map3k4
Nr0b1
Ntrk1
Nr5a1
Nup210l
Ptgdr
Ptgds
Scx
Sdc1
Six1
Six4
Sox3
Sox8
Sox9
Sox10
Sry
Wt1



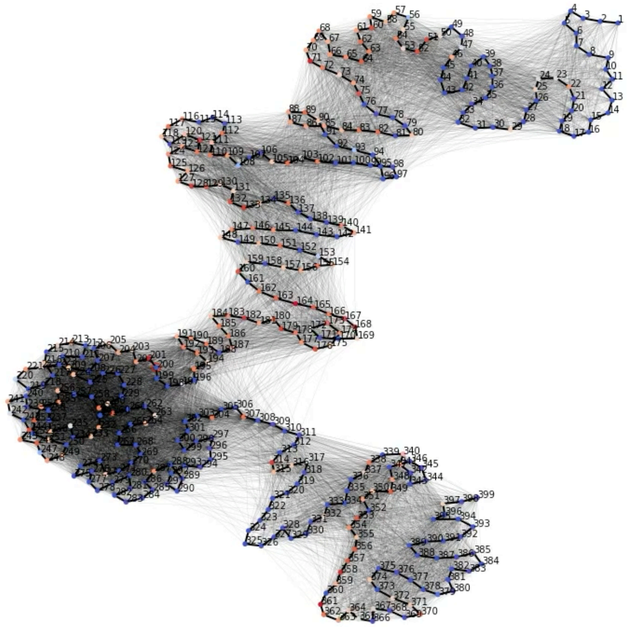


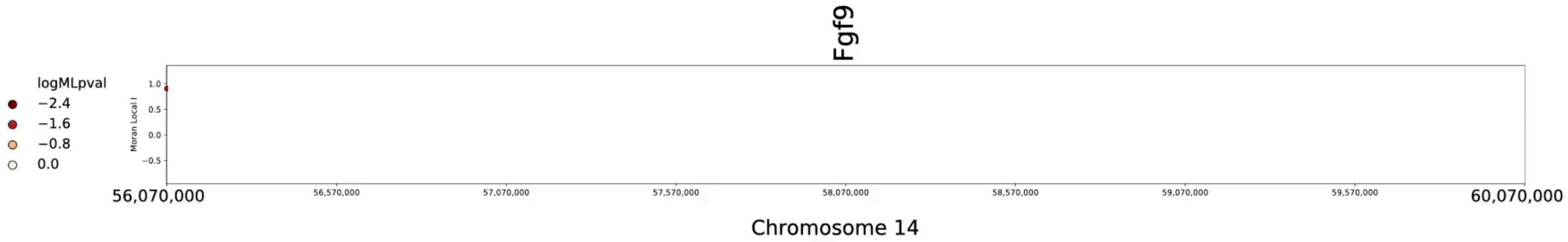
Now that we know the genes….

Can we identify regulatory elements using 

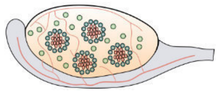
METALoci predictive mode

Fgf9 locus chr14:56,070,000-60,070,000



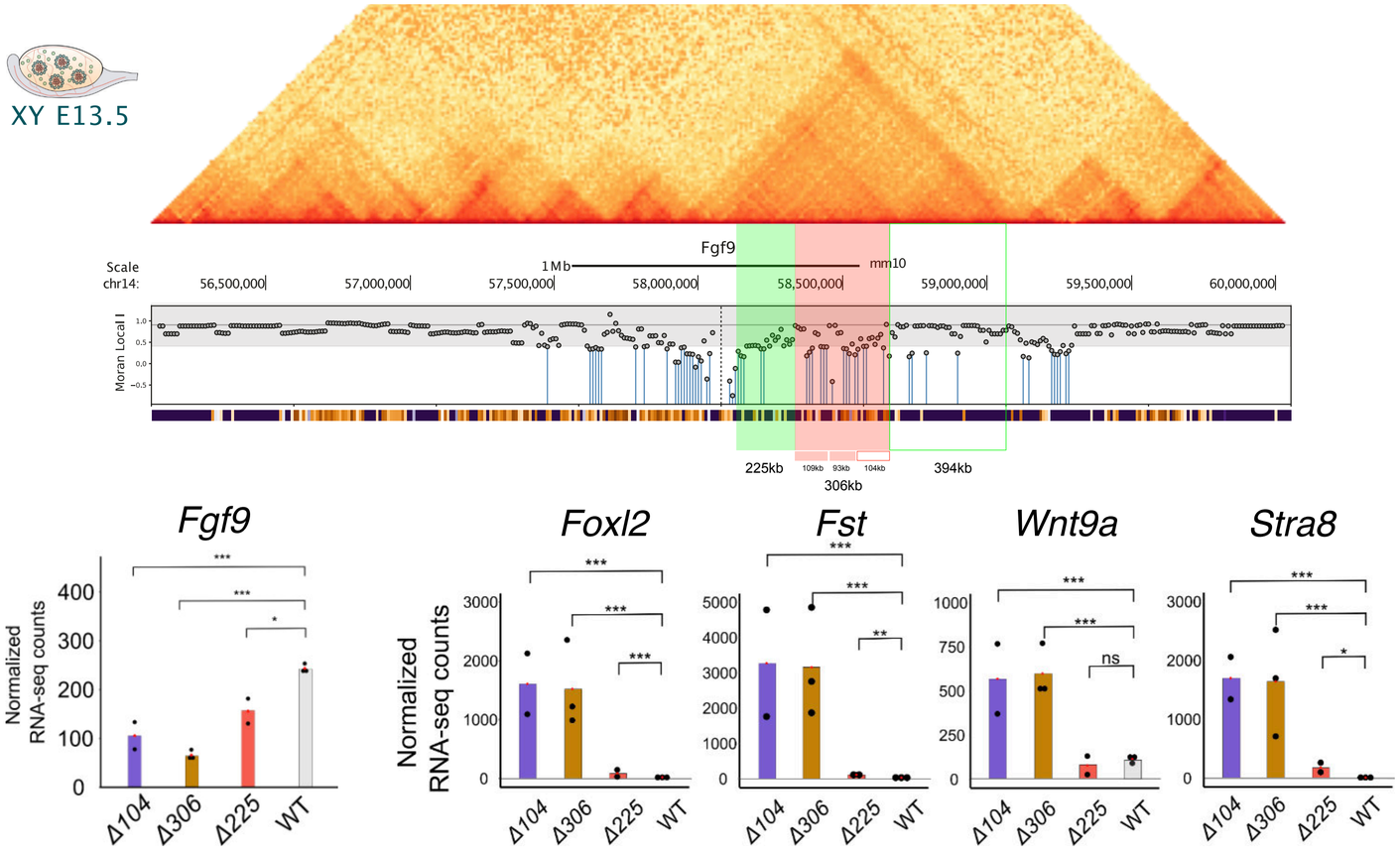


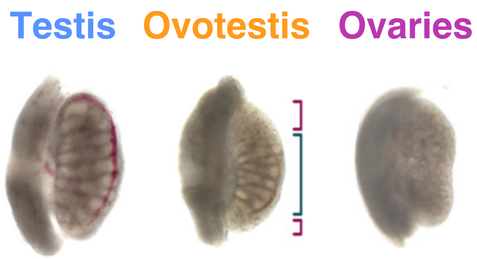
In silico prediction of regulatory regions at the Fgf9 locus

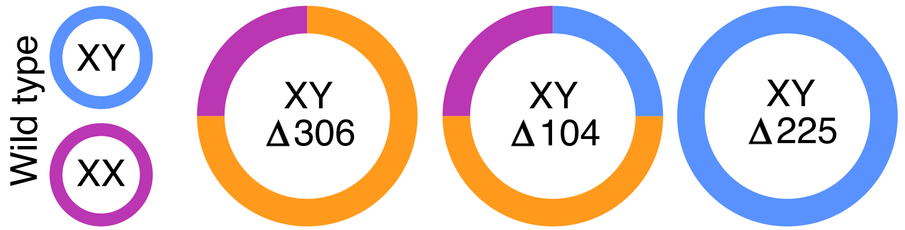


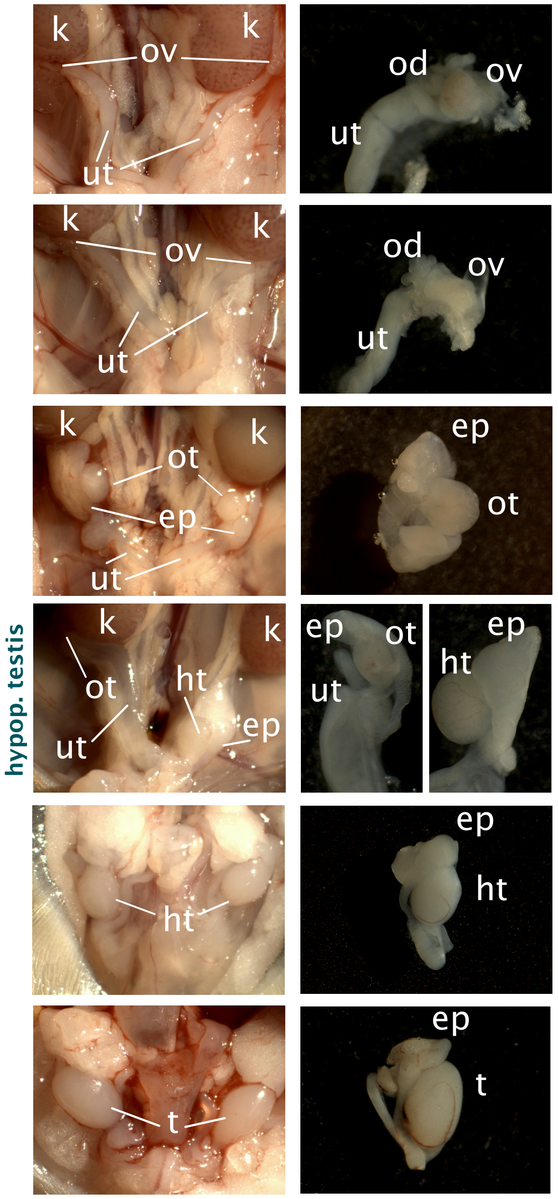
XY E13.5

Phenotype confirms METALoci predictions









Take home messages:

- First characterization of the 3D regulatory landscape of sex determination
- METALoci is an unbiased approach to quantify gene regulatory activity
- METALoci is a predictive tool to identify critical regulatory loci
- Discovery of a novel non-coding region controlling sex determination

@marciuslab
@mamartirenom

http://marciuslab.org

Alexander Barclay
Nikolai Bykov
Iana Kim
Peter Hoboth
Anne Lee
Iago Maceda
John Markham
Maria Marti-Marimon
Ana Nikolovska
Mireia Novell
Meritxell Novillo
Maria Roy
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