Washington University in St. Louis
School of Medicine

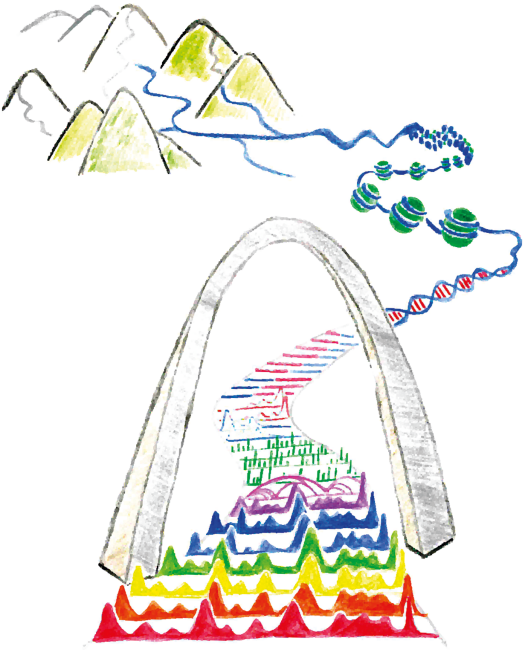
Department of Genetics

系统生物学与生物信息学

海外学者短期讲学系列课程

Current Topics in Epigenomics

表观基因组学前沿

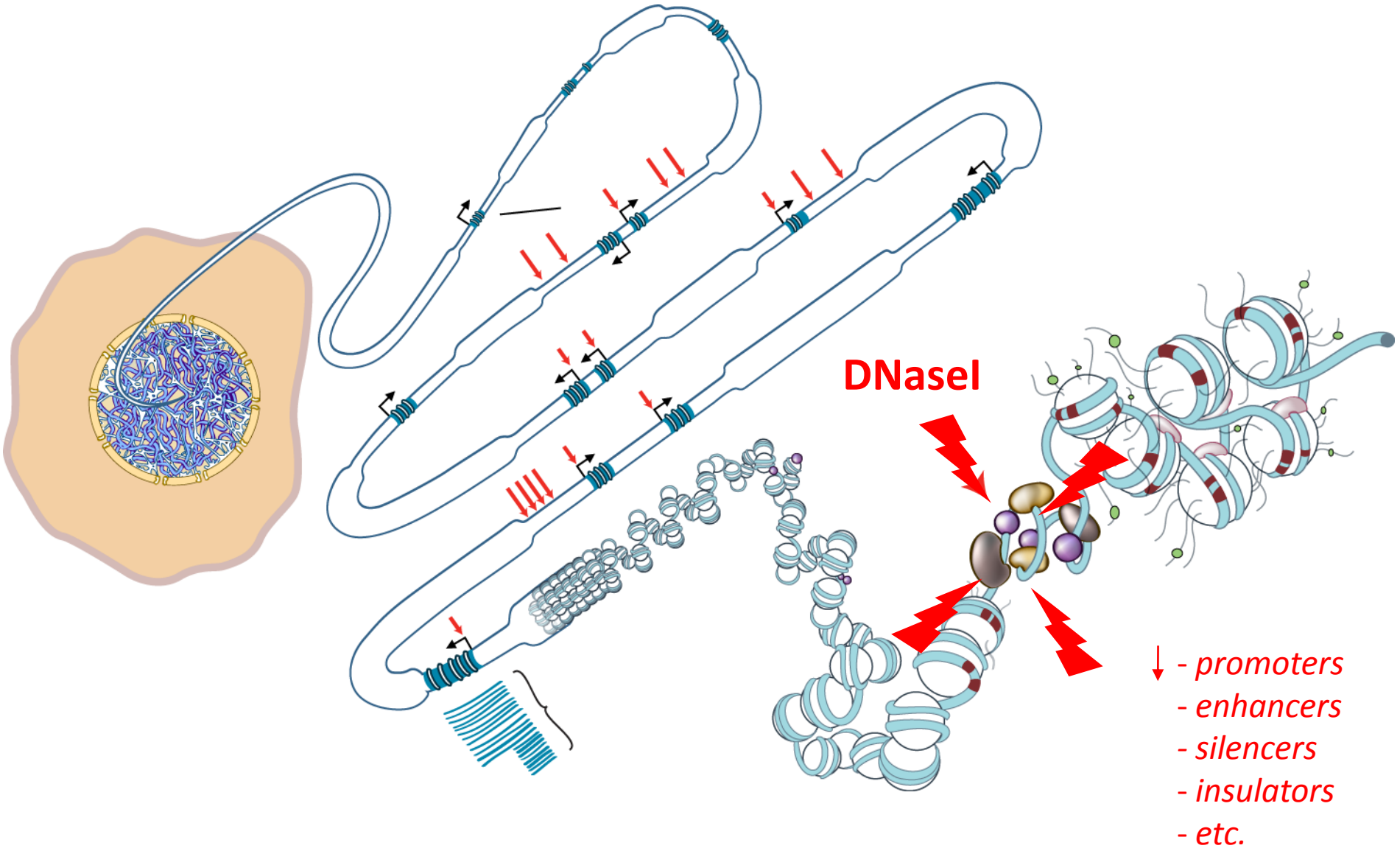


Ting Wang
Department of Genetics
Center for Genome Sciences and Systems Biology
Washington University School of Medicine

Tsinghua University
April 15-27

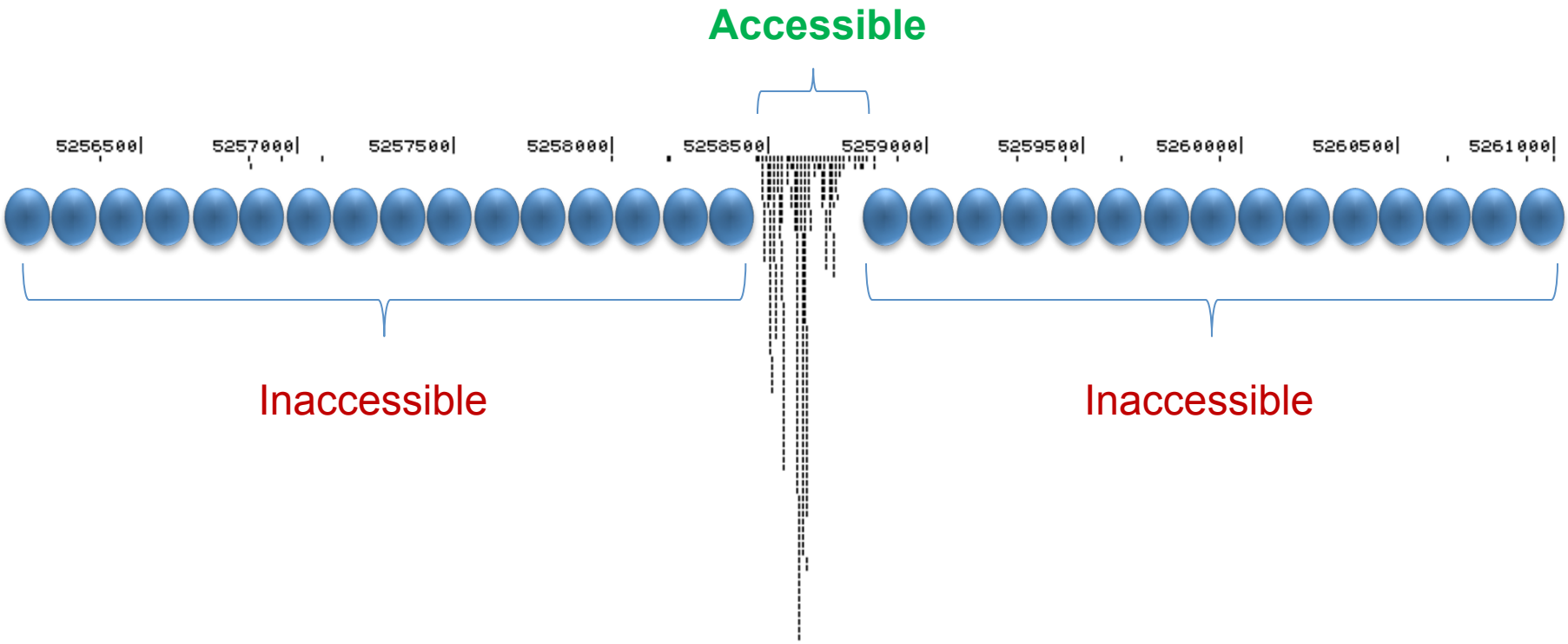
Epigenomic technology

DNase I hypersensitivity ~ Regulatory DNA

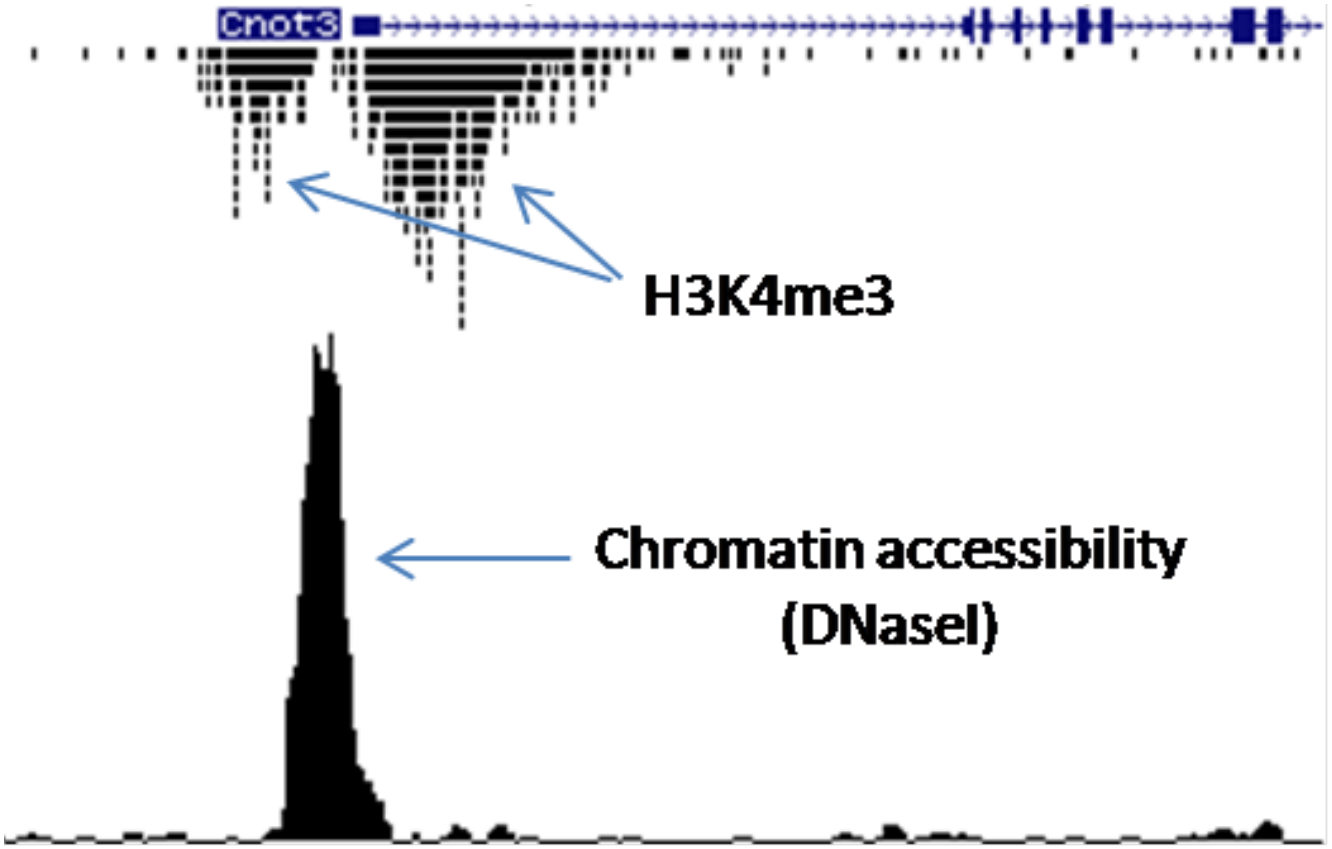


Digital DNaseI profiling

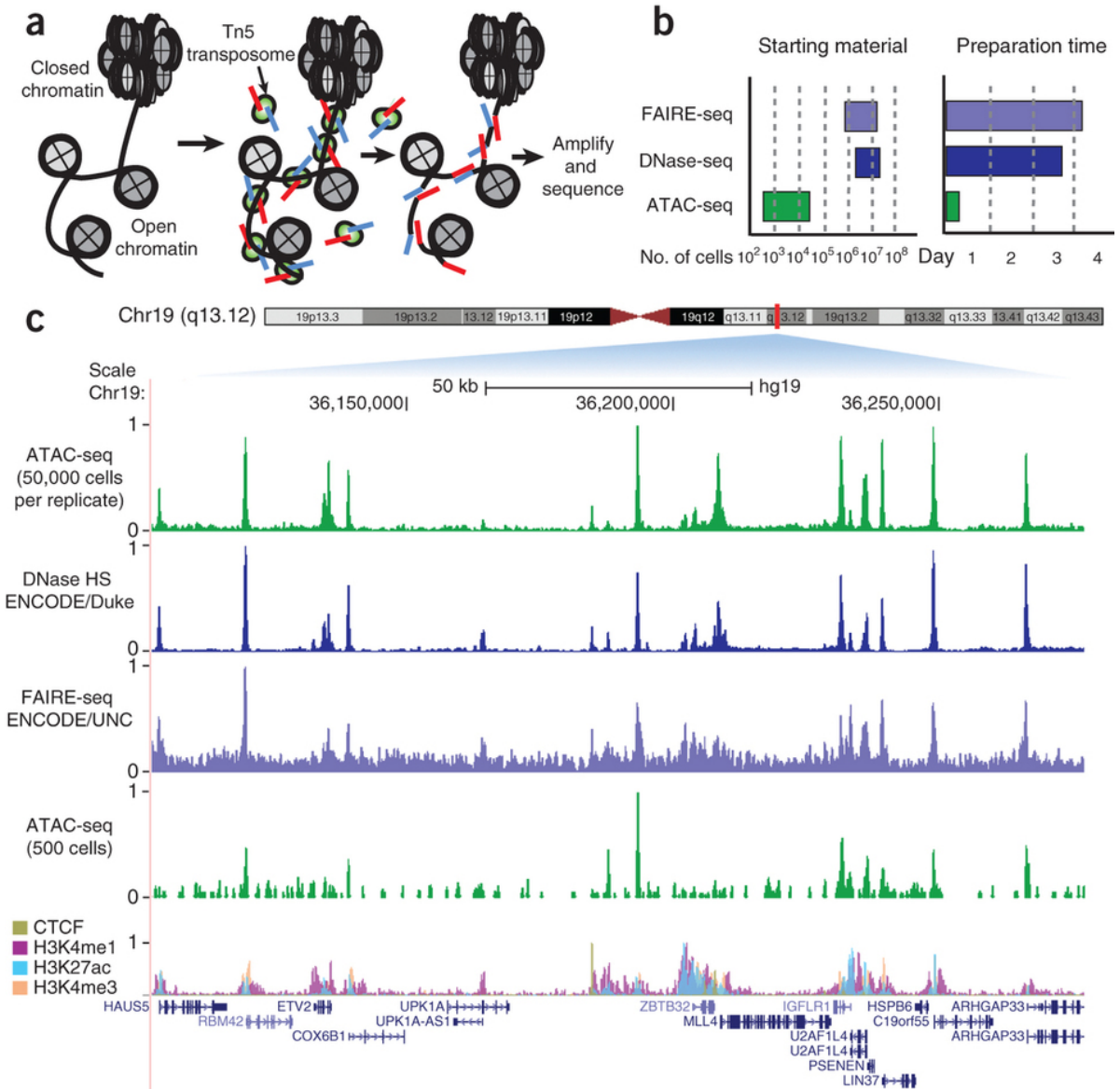
Precise delineation of the accessible regulatory DNA compartment



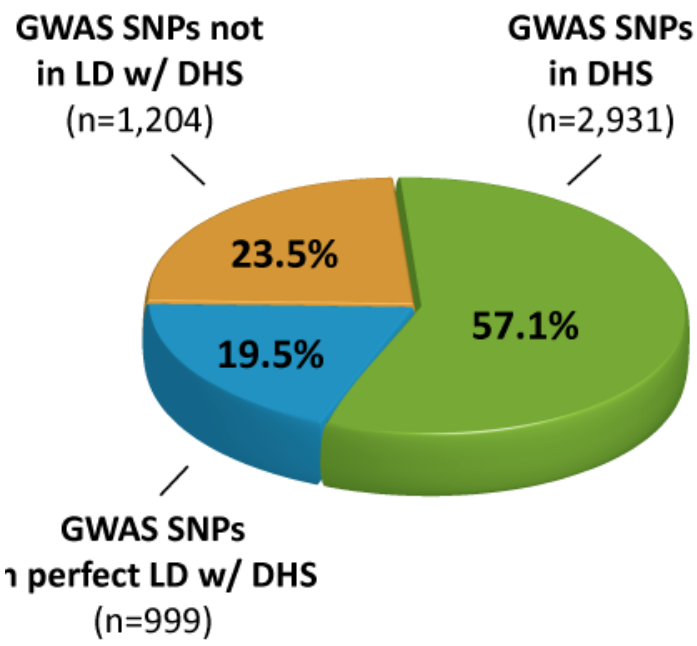
Digital DNaseI profiling: direct access to regulatory sequences

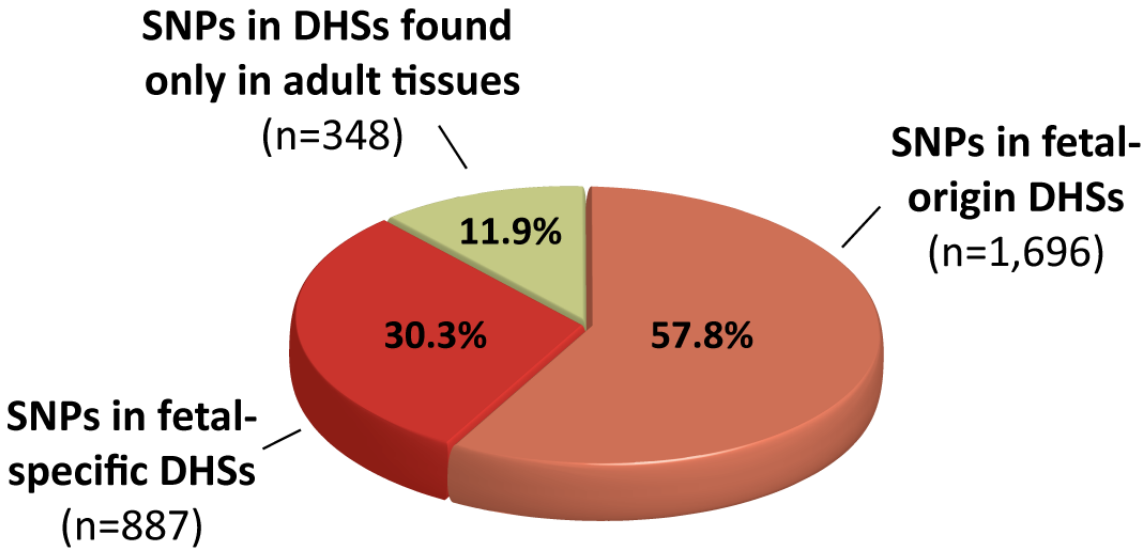


ATAC-seq
(assay for transposase-accessible chromatin)

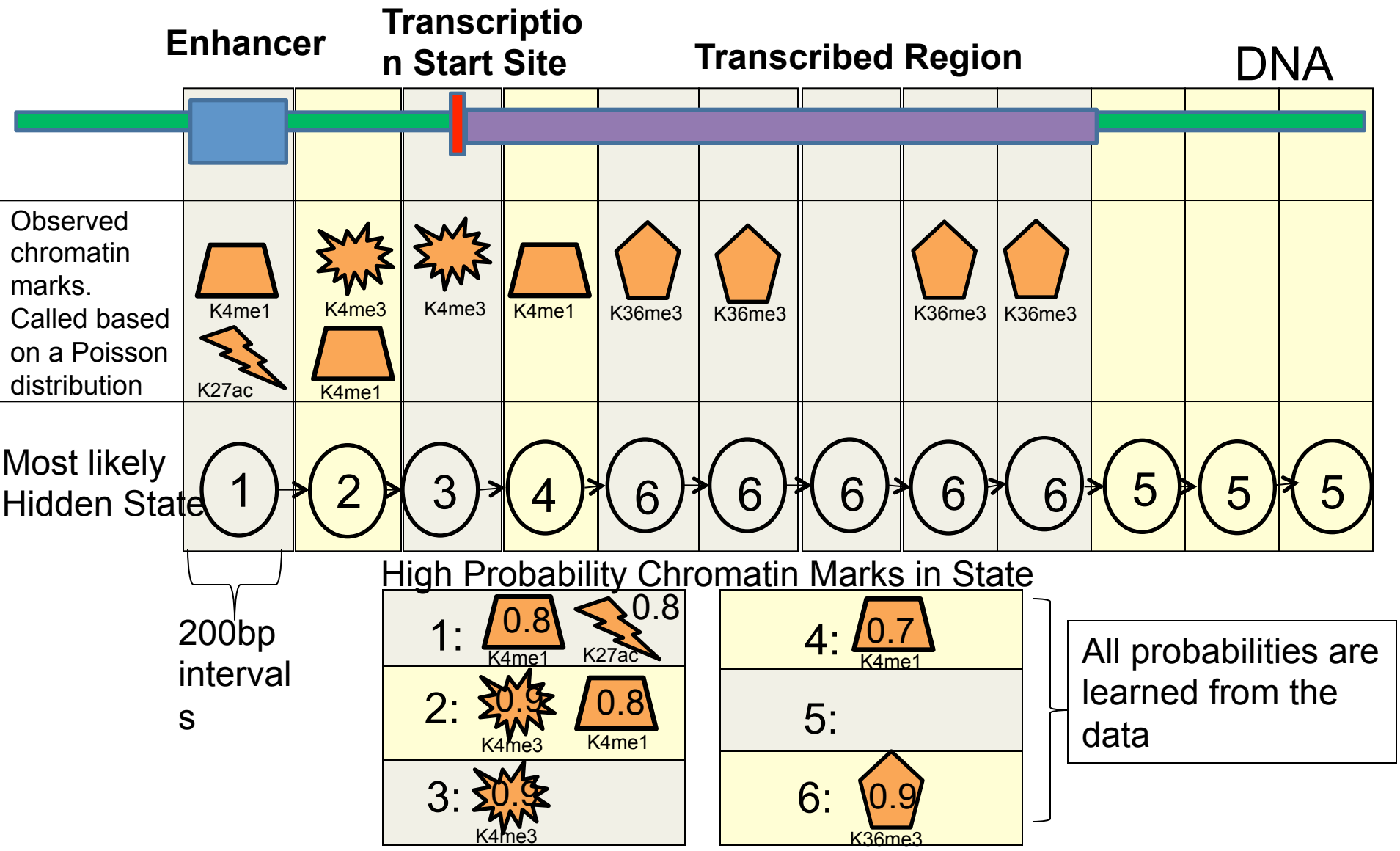


Epigenetic annotation of GWAS hits

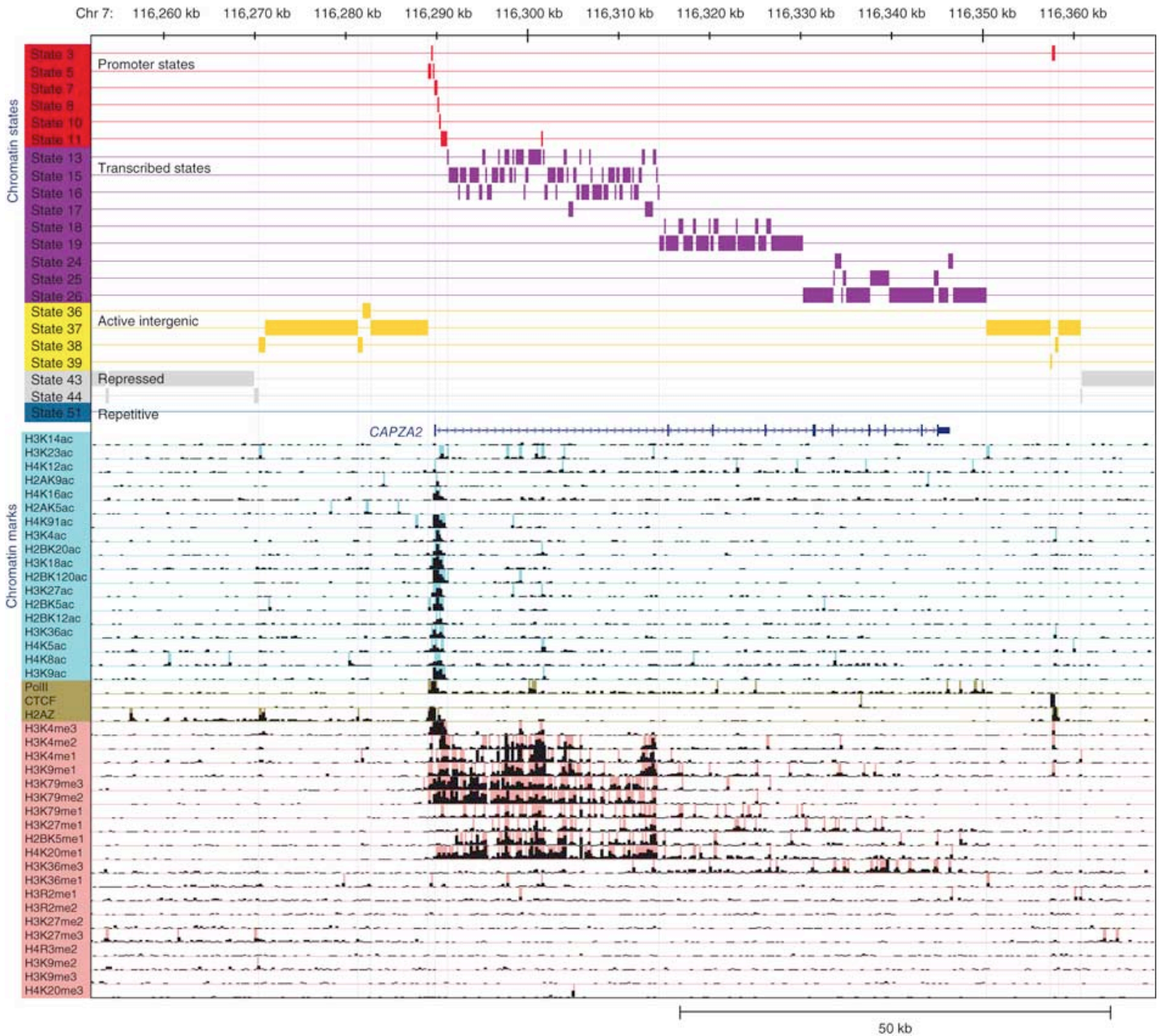




ChromHMM


Enhancer
Transcriptio n Start Site
Transcribed Region
DNA
Observed chromatin marks. Called based on a Poisson distribution
K4me1
K27ac
K4me3
K4me1
K4me3
K4me1
K36me3
K36me3
K36me3
K36me3
Most likely Hidden State
1
2
3
4
6
6
6
6
6
5
5
5
200bp interval s
High Probability Chromatin Marks in State
1: 0.8 K4me1 0.8 K27ac
2: 0.9 K4me3 0.8 K4me1
3: 0.9 K4me3
4: 0.7 K4me1
5:
6: 0.9 K36me3
All probabilities are learned from the data

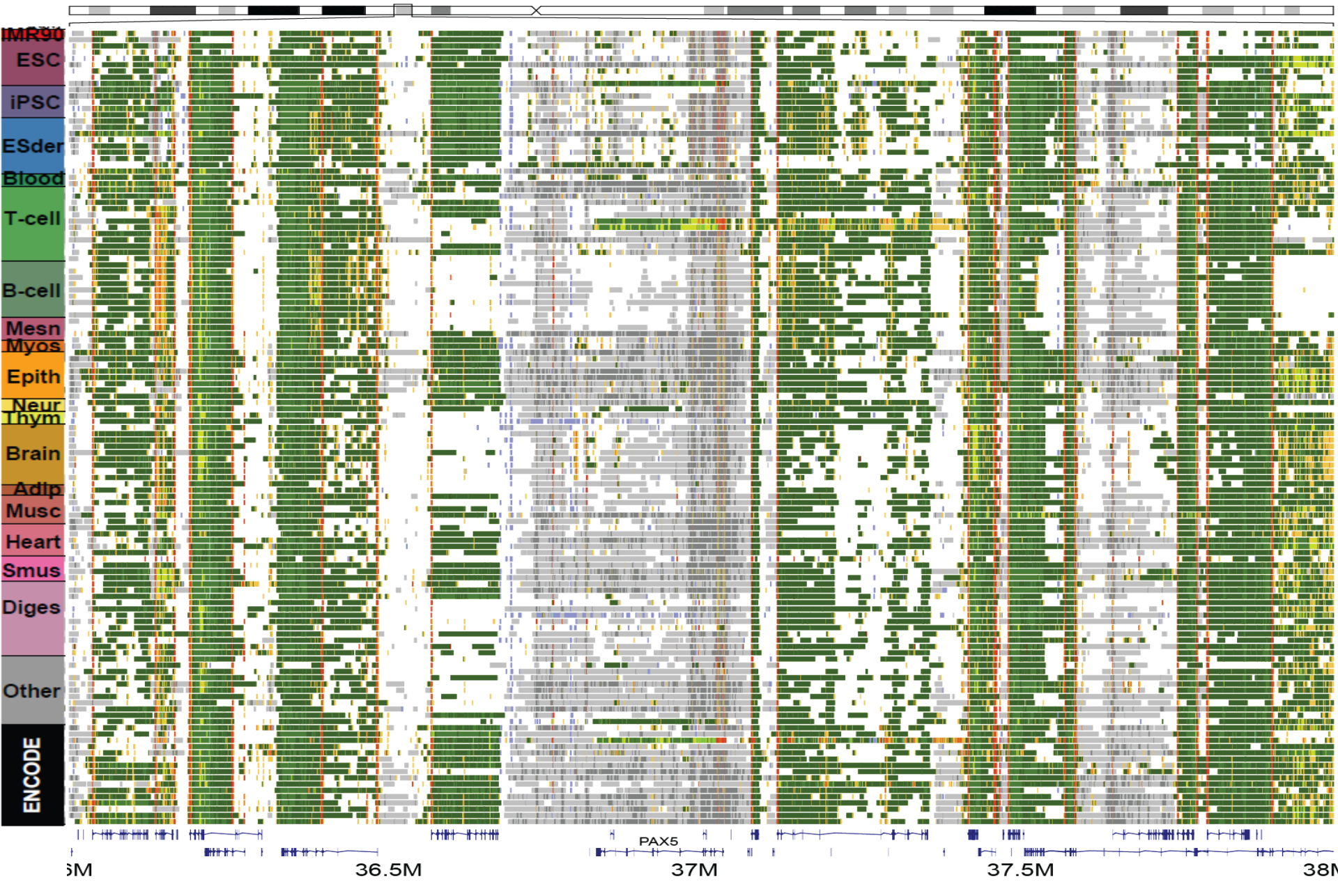
ChromHMM


Chr 7: 116,260 kb 116,270 kb 116,280 kb 116,290 kb 116,300 kb 116,310 kb 116,320 kb 116,330 kb 116,340 kb 116,350 kb 116,360 kb
Chromatin states
State 3
State 5
State 7
State 8
State 10
State 11
Promoter states
State 13
State 15
State 16
State 17
State 18
State 19
State 24
State 25
State 26
Transcribed states
State 36
State 37
State 38
State 39
Active intergenic
State 43
State 44
Repressed
State 51
Repetitive
CAPZA2
Chromatin marks
H3K14ac
H3K23ac
H4K12ac
H2AK9ac
H4K16ac
H2AK5ac
H4K91ac
H3K4ac
H2BK20ac
H3K18ac
H2BK120ac
H3K27ac
H2BK5ac
H2BK12ac
H3K36ac
H4K5ac
H4K8ac
H3K9ac
PolII
CTCF
H2AZ
H3K4me3
H3K4me2
H3K4me1
H3K9me1
H3K79me3
H3K79me2
H3K79me1
H3K27me1
H2BK5me1
H4K20me1
H3K36me3
H3K36me1
H3R2me1
H3R2me2
H3K27me2
H3K27me3
H4R3me2
H3K9me2
H3K9me3
H4K20me3
50 kb

Application of ChromHMM to 41 chromatin marks in CD4+ T-cells (Barski'07, Wang'08)

Chromatin Marks from (Barski et al, Cell 2007; Wang et al Nature Genetics, 2008); DNAseI hypersensitivity from (Boyle et al, Cell 2008); Expression Data from (Su et al, PNAS 2004); Lamina data from (Guelen et al; Naature 2008)

Chromatin state annotations across 127 epigenomes



Genome 4D

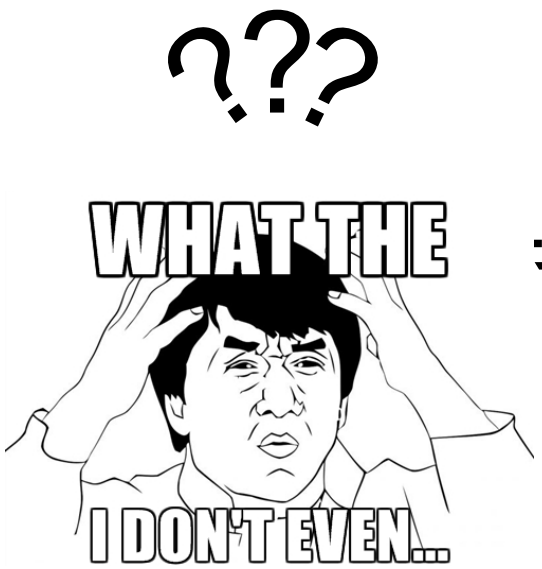
Chromatin structure

"We finished the genome map, now we can't figure out how to fold it."

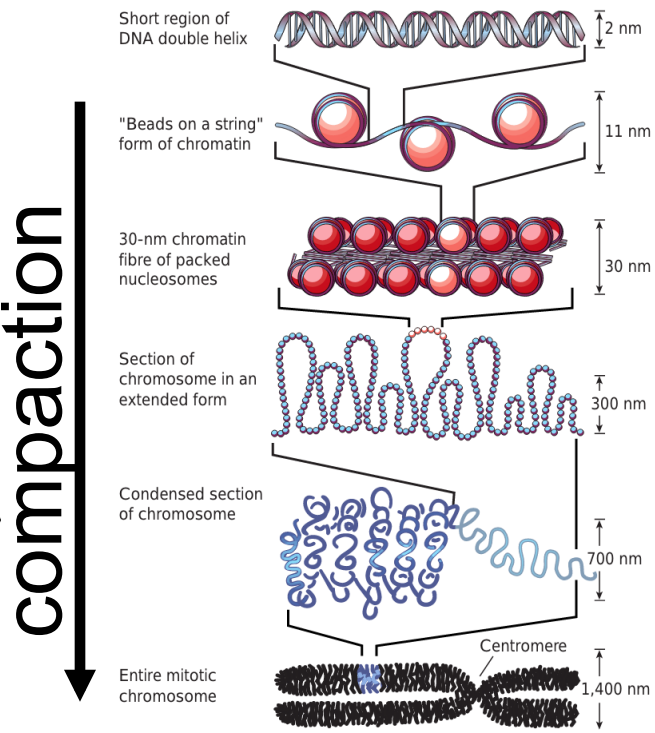
Adapted with permission from cartoonist John Chase http://www.chasetoons.com

DNA packaging problem

Total extended length of DNA in a human cell is ~1.8 m, which is longer than the average height of humans!



CAN'T BE NUCLEAR PHYSICS!?!
…or is it?

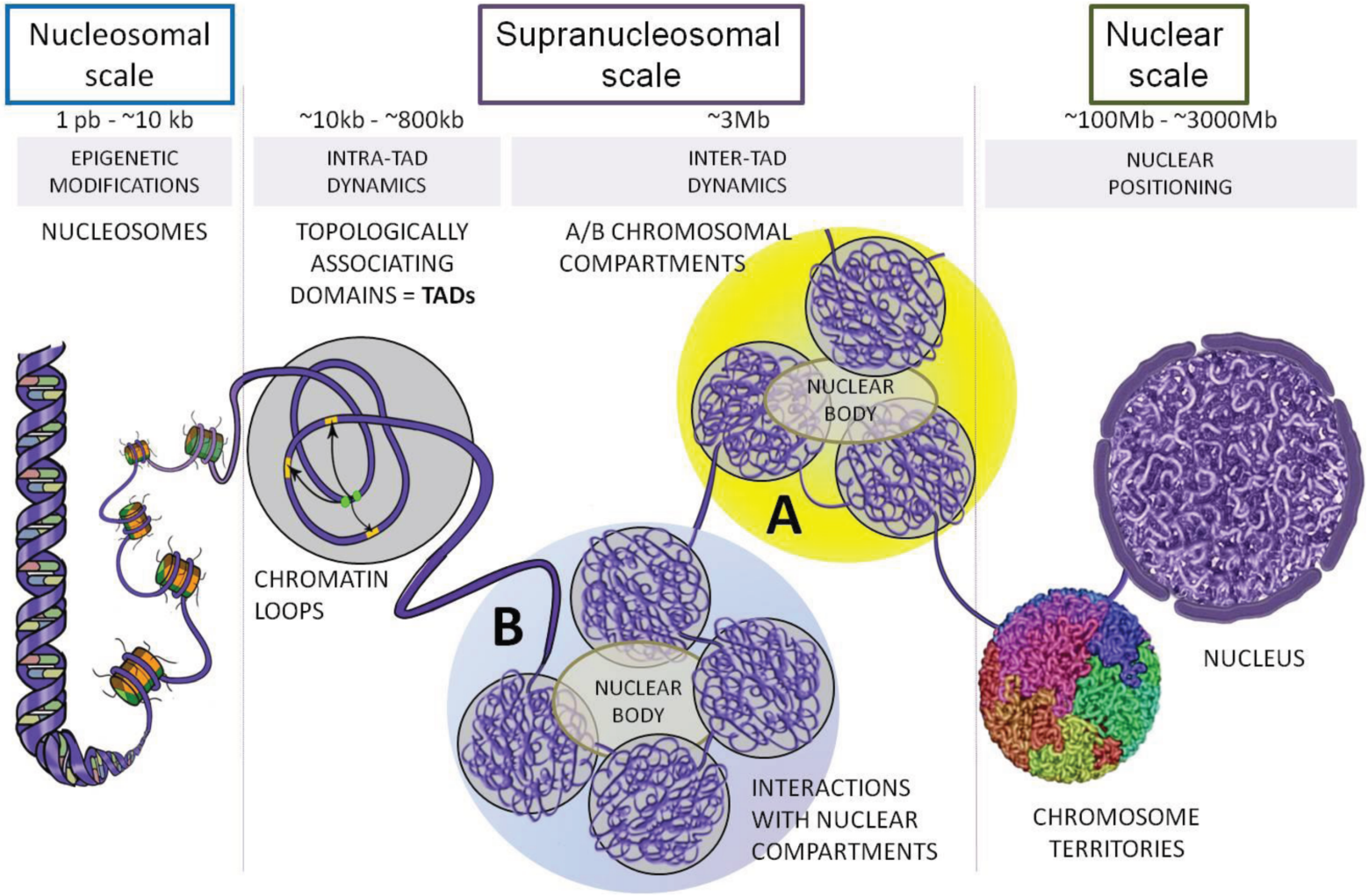


Structural hierarchy of human DNA that is fit into a nucleus with a diameter of <10 μm

---

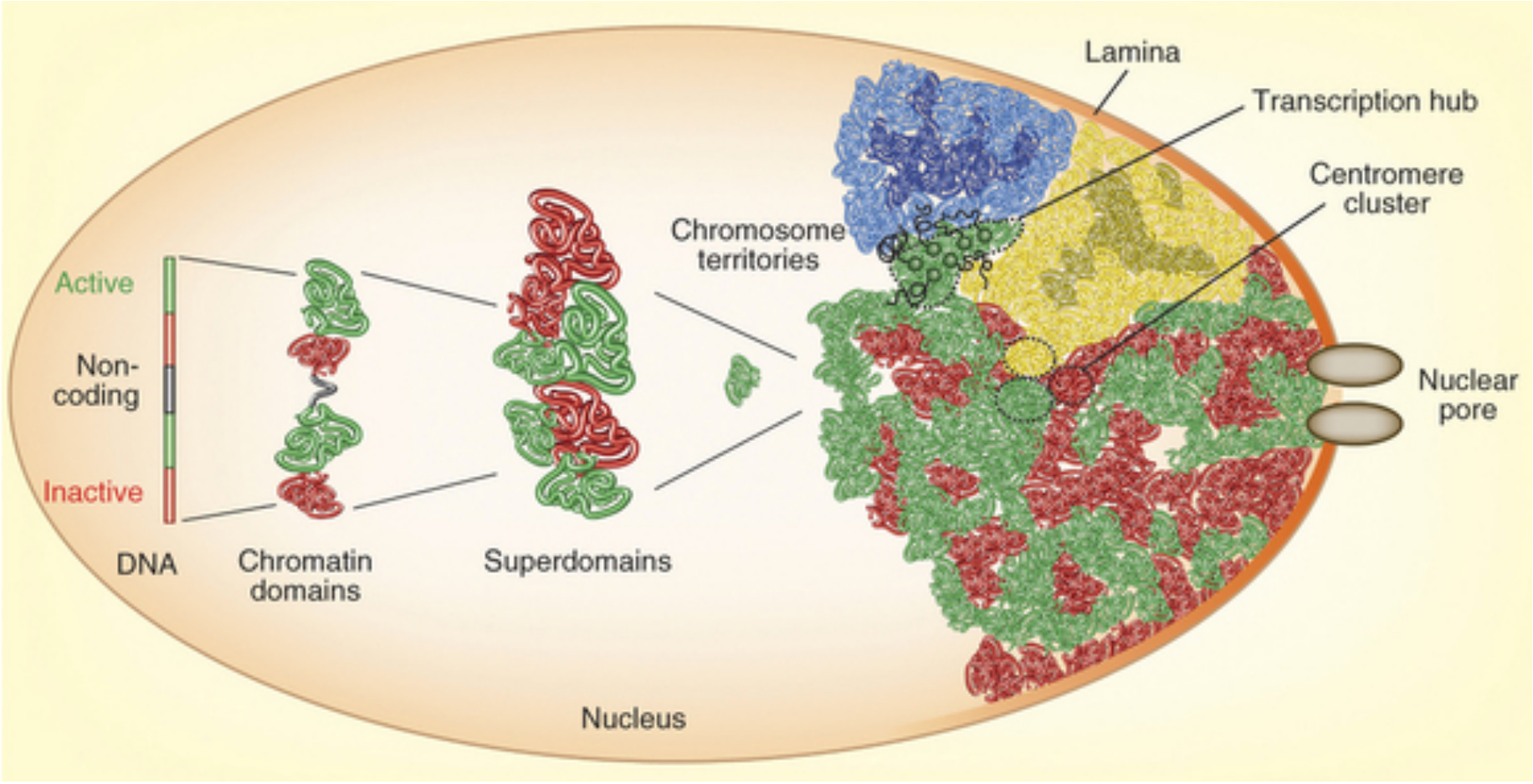
Adapted from
http://thetechjournal.com/science/8-2-percent-dna-functional.xhtml/attachment/man-climbing-dna
Controlling the double helix. ***Felsenfeld G & Groudine M.*** *Nature (2003)*

Genome organization in mammals



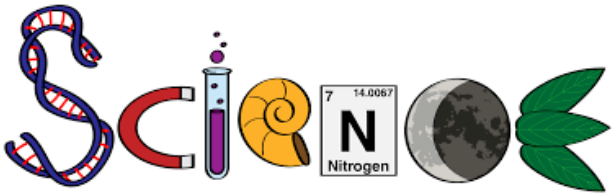
Contribution of Topological Domains and Loop Formation to 3D Chromatin Organization. ***Ea*** *et al. Genes (2015)*

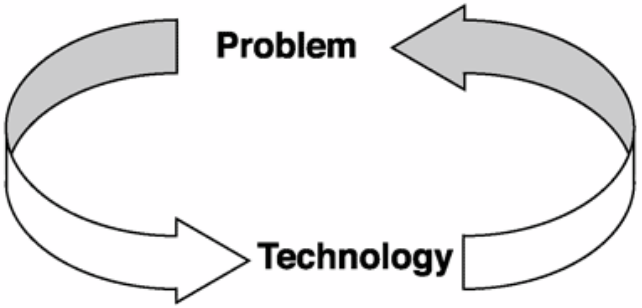
Global *view* of the nucleus



Adapted from Functional implications of genome topology. ***Cavalli G & Misteli T****. Nature Structural and Molecular Biology. (2013)*

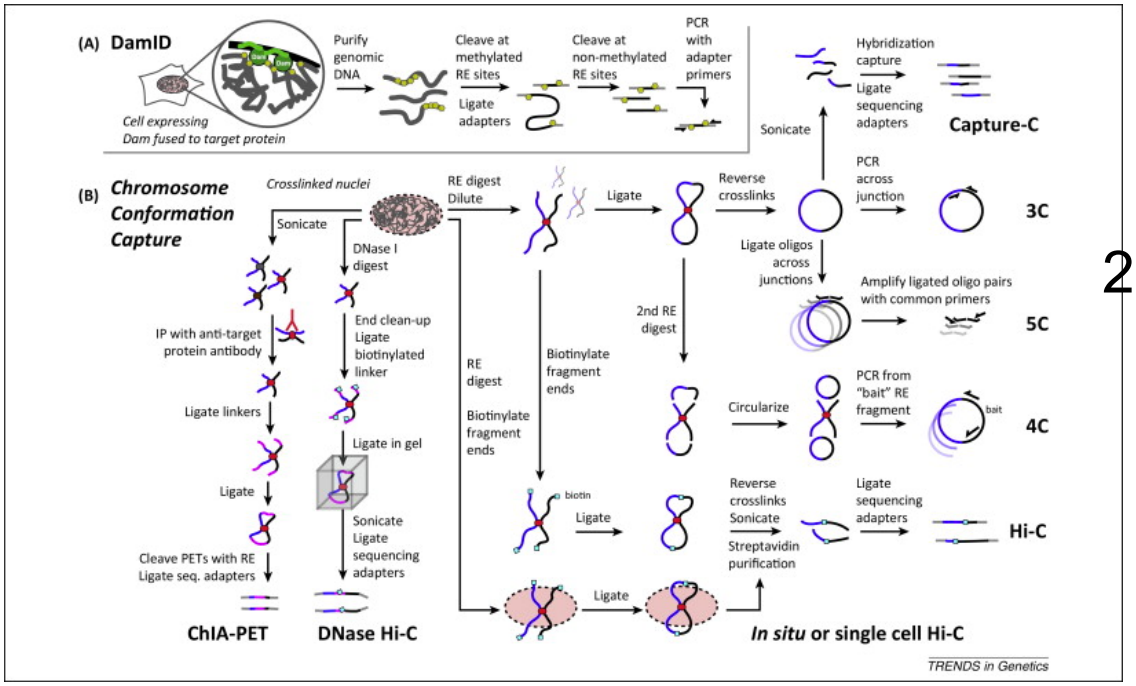
But how do we *"view"* it?





Adapted with changes from Savage Chickens by Doug Savage, and Google Images.

Technologies used (and developed) to study genome folding



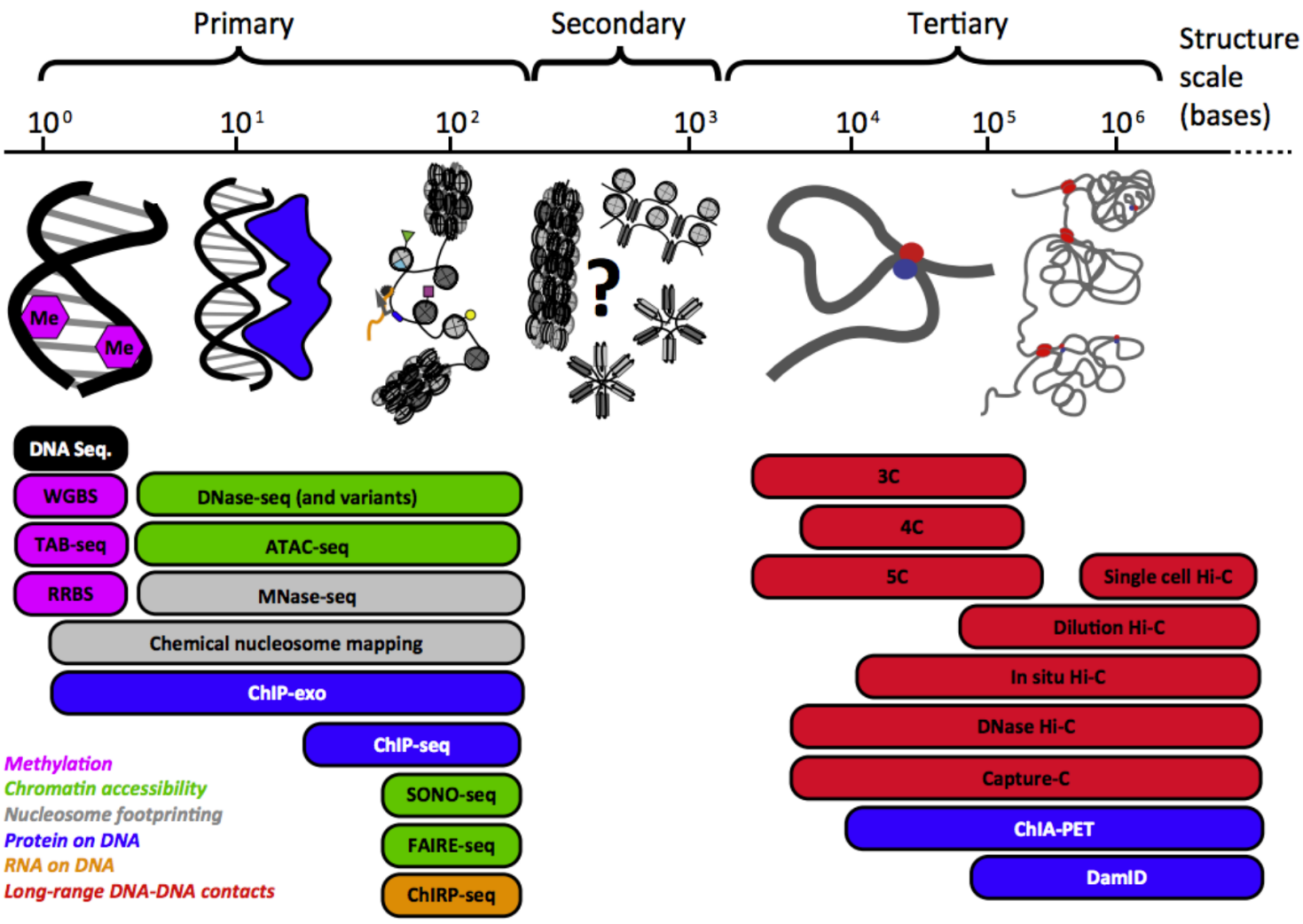
Can be divided into two broad categories:

1. Imaging
   1. Bright-field
   2. Fluorescence
   3. EM
   4. Fluorescence *in-situ* hybridization (FISH), etc.
2. Genome-wide
   1. DamID
   2. ChIA-PET
   3. Chromosome conformation capture-derived, etc.

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

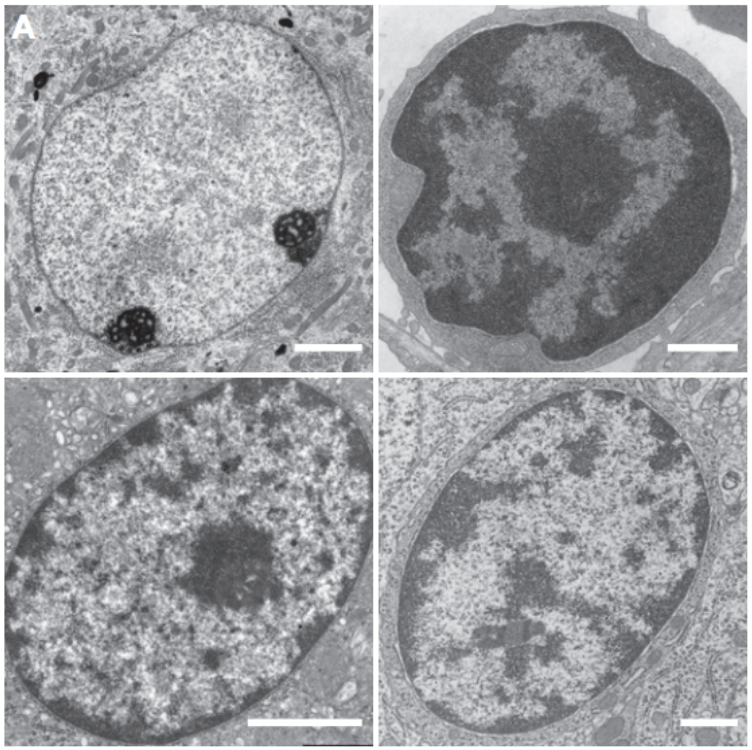
Unraveling the 3D genome: genomics tools for multiscale exploration. ***Viviana I. Risca and William J. Greenleaf.*** *Trends in Genetics (2015)*

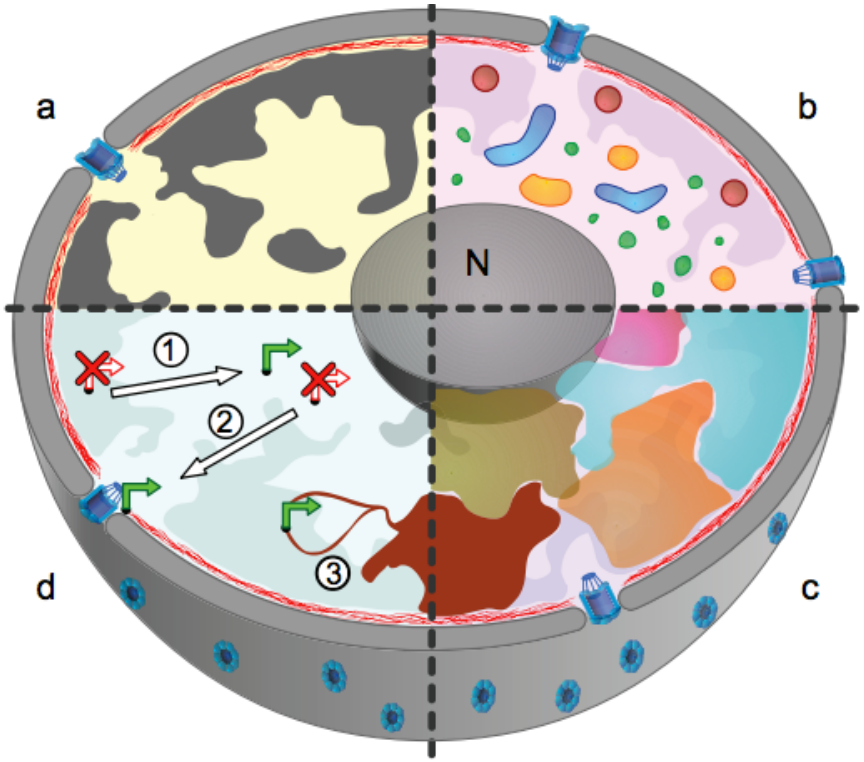
Overview of chromatin structure and assays at three scales



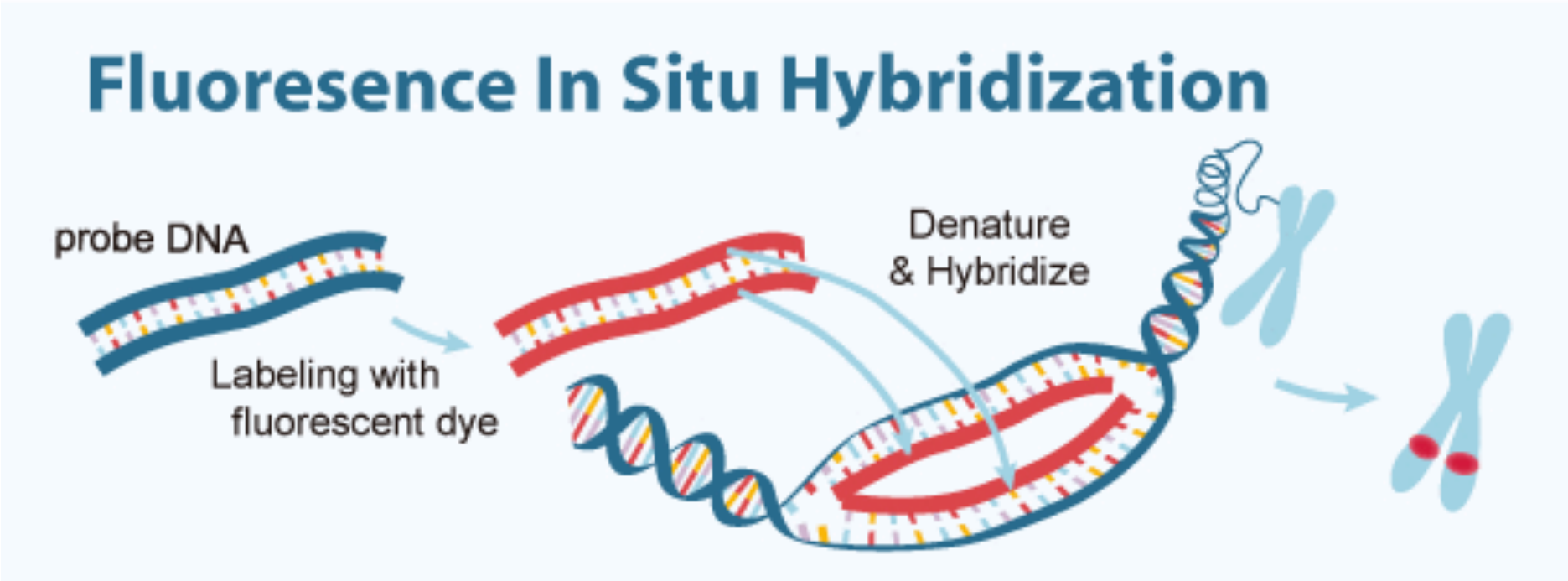
Unraveling the 3D genome: genomics tools for multiscale exploration. ***Viviana I. Risca and William J. Greenleaf.*** *Trends in Genetics (2015)*

Principles of nuclear organization revealed using microscopic techniques





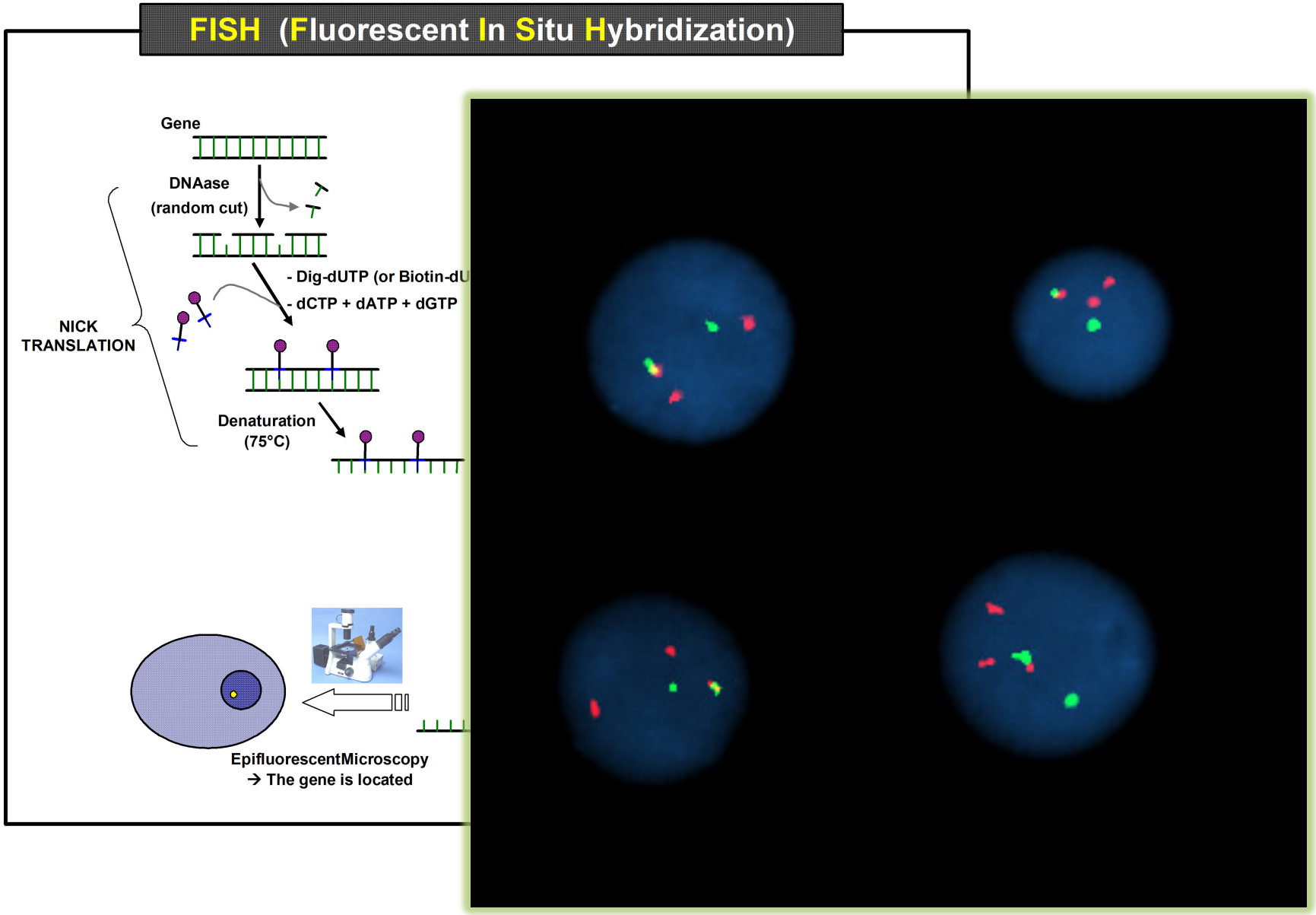
From single genes to entire genomes: the search for a function of nuclear organization. ***Pueschel*** *et al. Development (2016)*



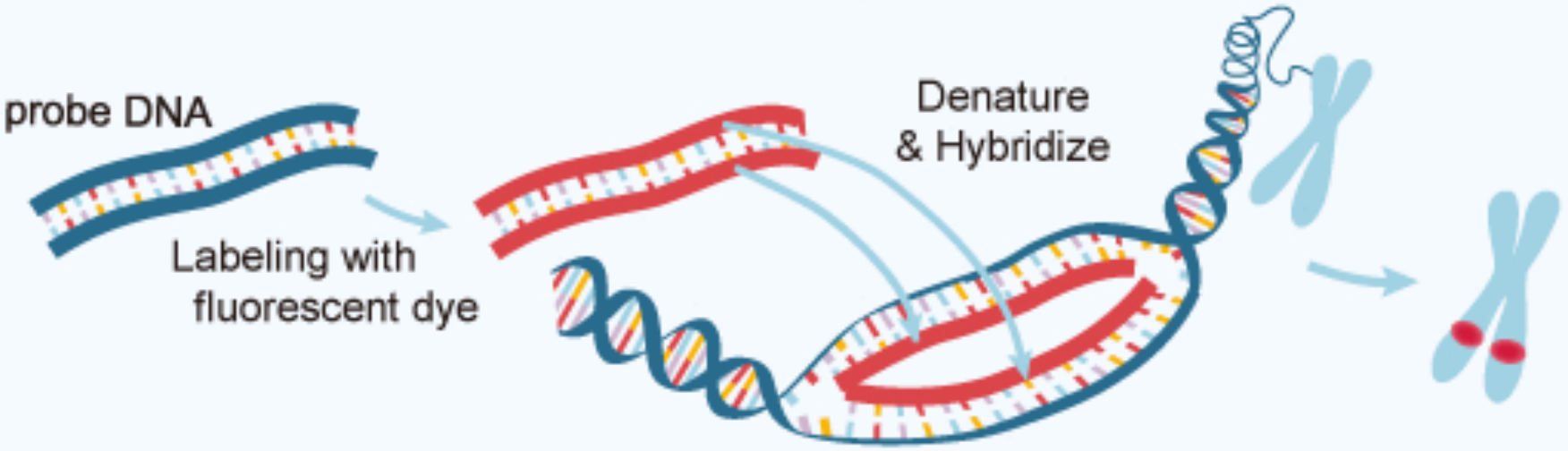
- Cytogenetic technique
- Uses fluorescent molecules to “paint” (regions of interest on) chromosomes in cells often in Metaphase or Interphase
- Aids in analysis of chromosome structure, structural aberrations, ploidy determination, etc.

Adapted from http://www.abnova.com/images/content/support/Banner_fish.gif

FISH analysis of genomic loci



Fluoresence In Situ Hybridization



ADVANTAGES

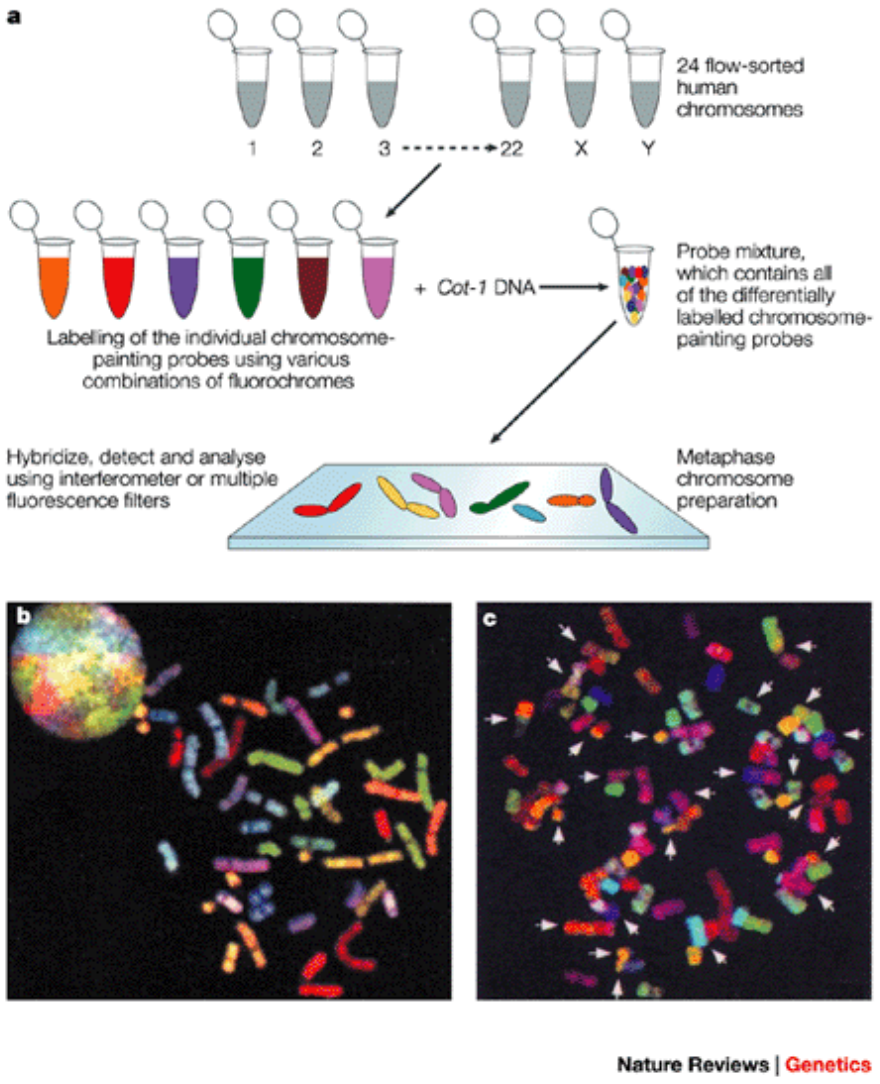
- Rapid and sensitive
- Lots of cells can be analyzed
- No cell culture needed

DISADVANTAGES

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

Adapted from http://www.abnova.com/images/content/support/Banner_fish.gif

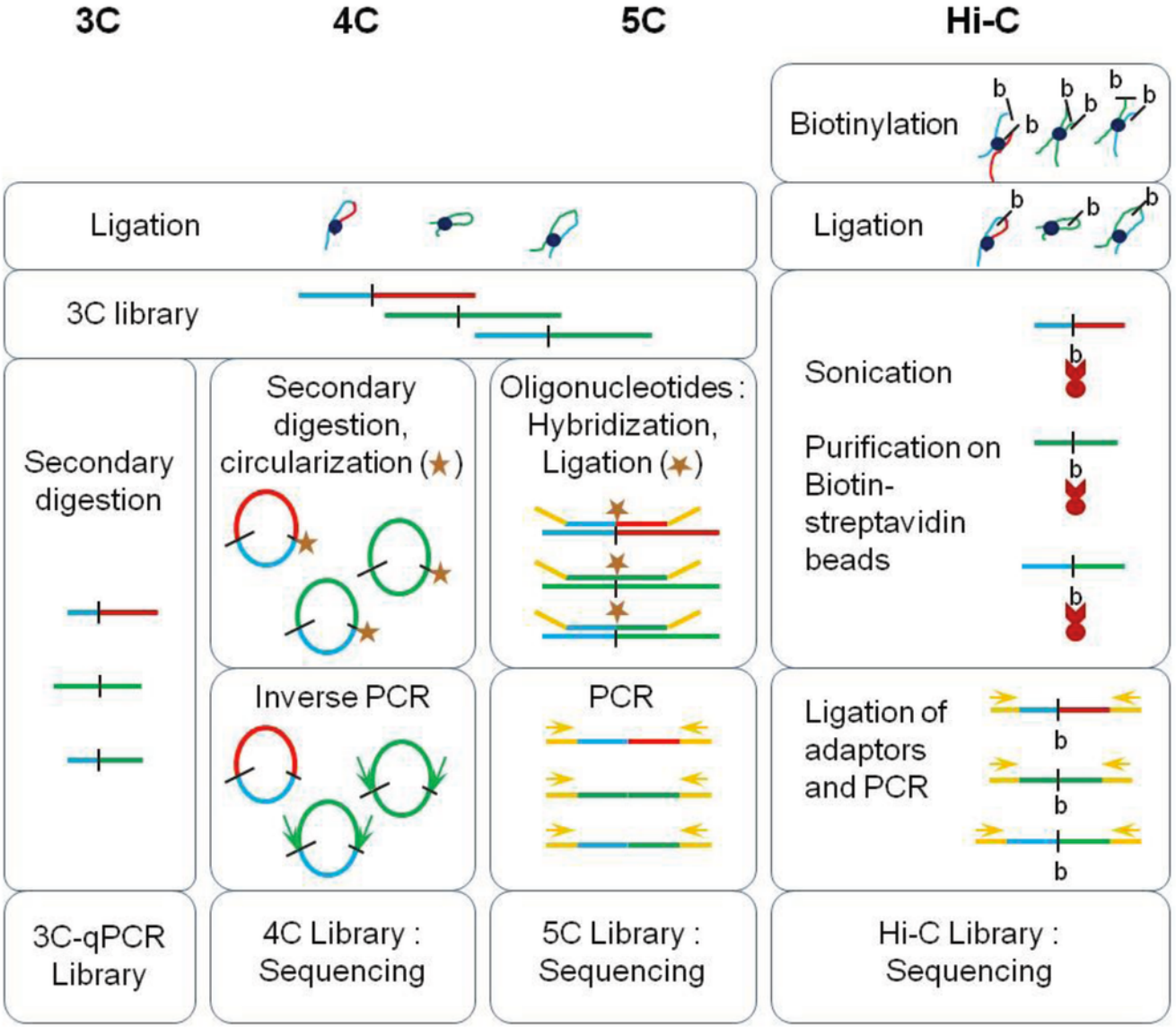
Chromosome painting



Chromosome painting analysis of interphase nuclei reveals chromosome territories

Bolzer et al., PLoS Biol. 2005

Methodology of “C”-technologies

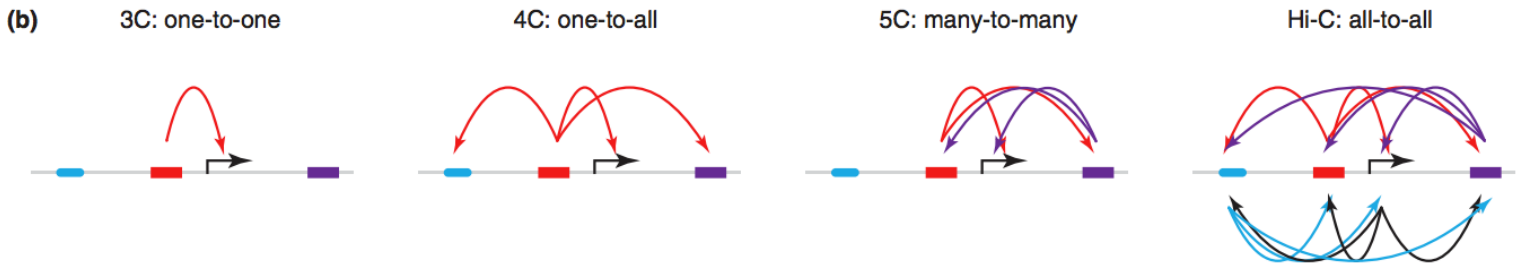


Contribution of Topological Domains and Loop Formation to 3D Chromatin Organization. ***Ea*** *et al. Genes (2015)*

Advantages and drawbacks of "C"-technologies

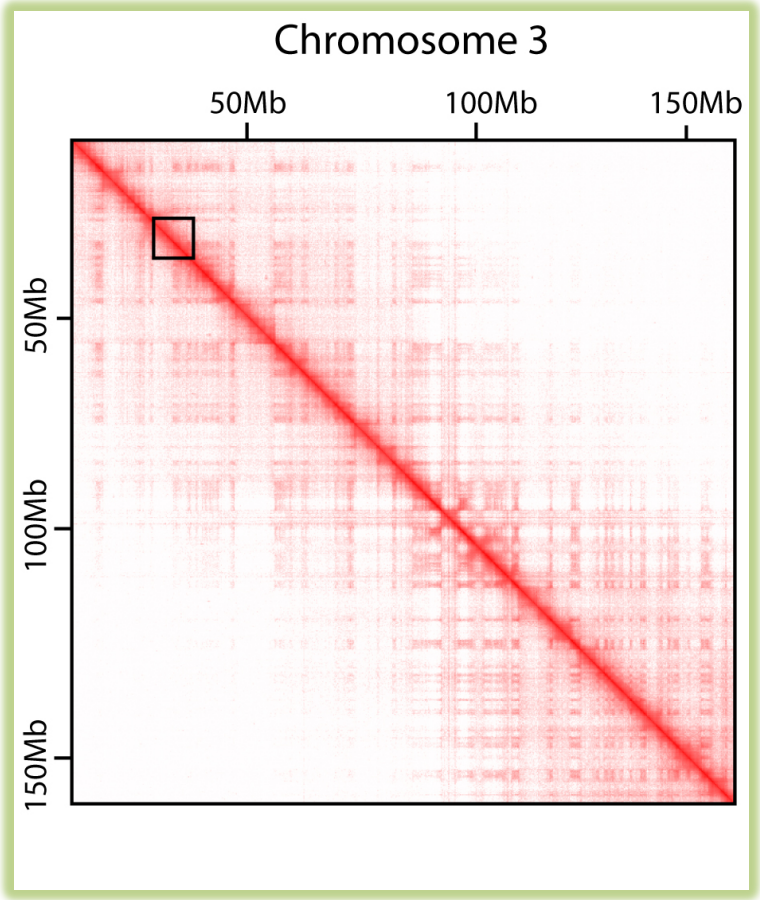
Table 1. Advantages and limits of 3C-derived methods.

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

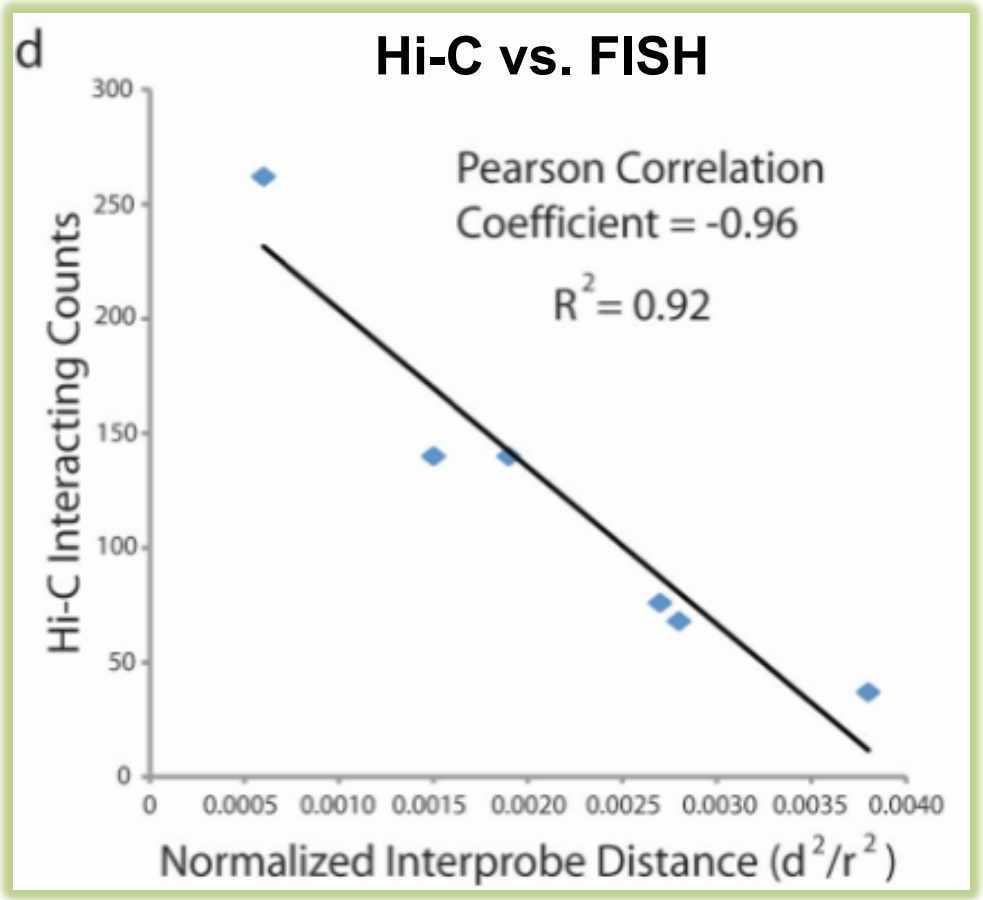


Contribution of Topological Domains and Loop Formation to 3D Chromatin Organization. ***Ea*** *et al. Genes (2015)*
Genome organization influences partner selection for chromosomal rearrangements. ***Patrick J. Wijchers & Wouter de Laat****. Trends in Genetics (2011)*

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

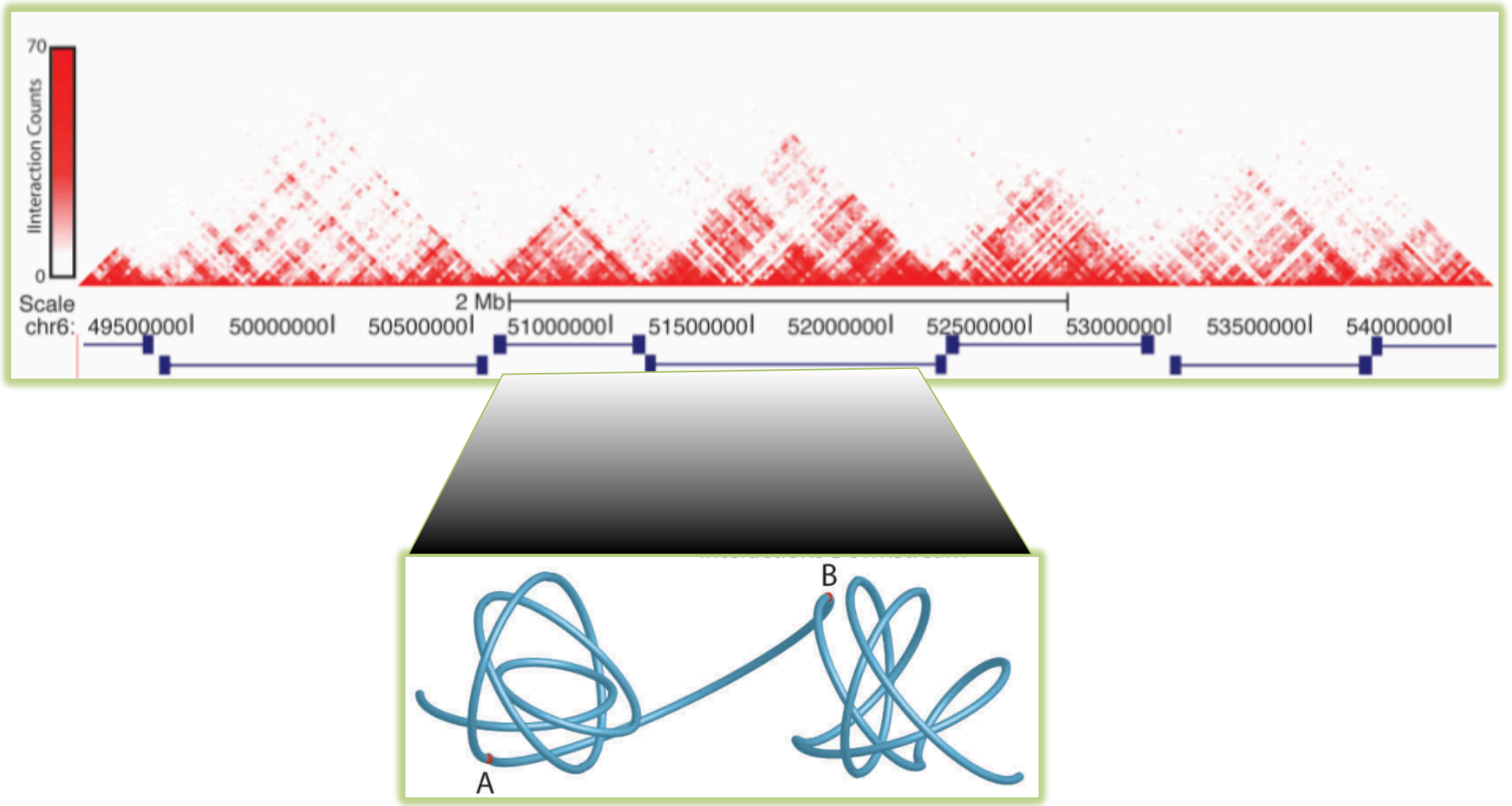


Mouse ES cells
(from 433 Million Reads)

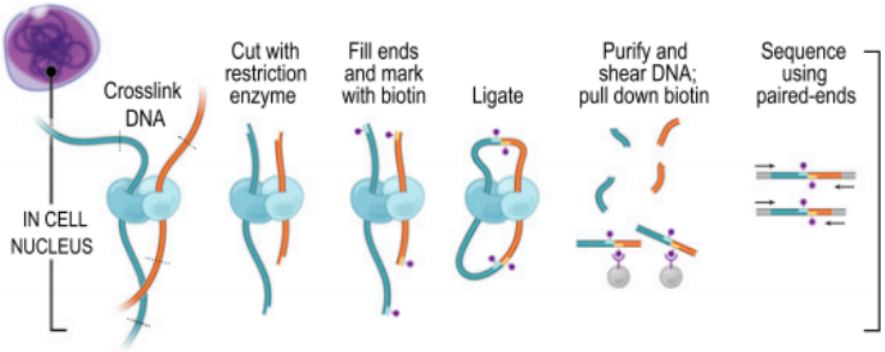


Dixon et al., Nature, 2012

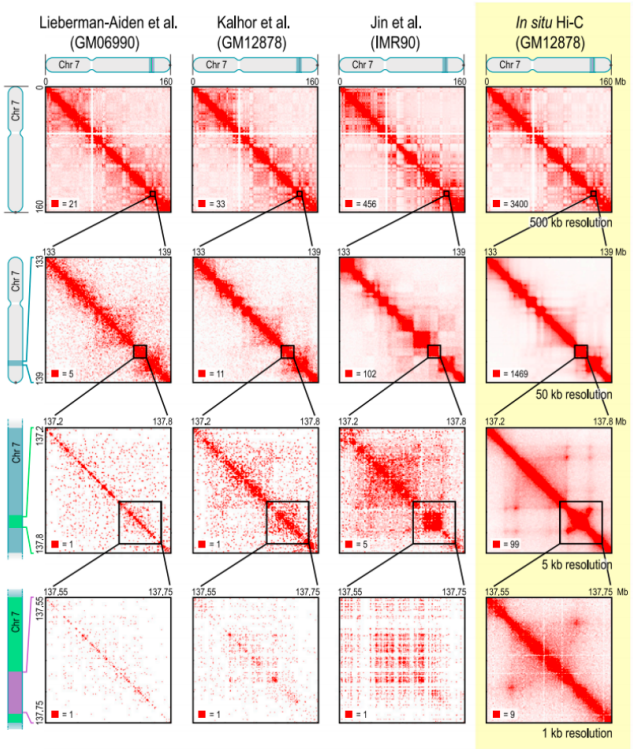
The topological domains or TADs



in situ Hi-C



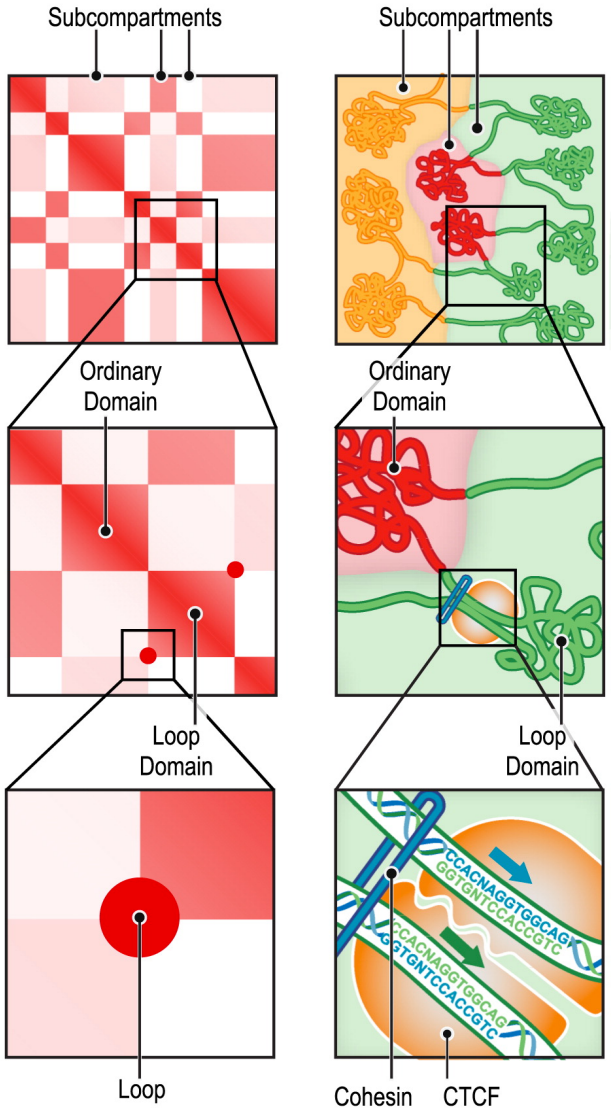
In situ Hi-C maps DNA–DNA contacts occurring in intact nuclei, by proximity ligation.



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.

A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. ***Rao SS*** *et al. Cell (2014)*

Overview of features revealed by Hi-C maps



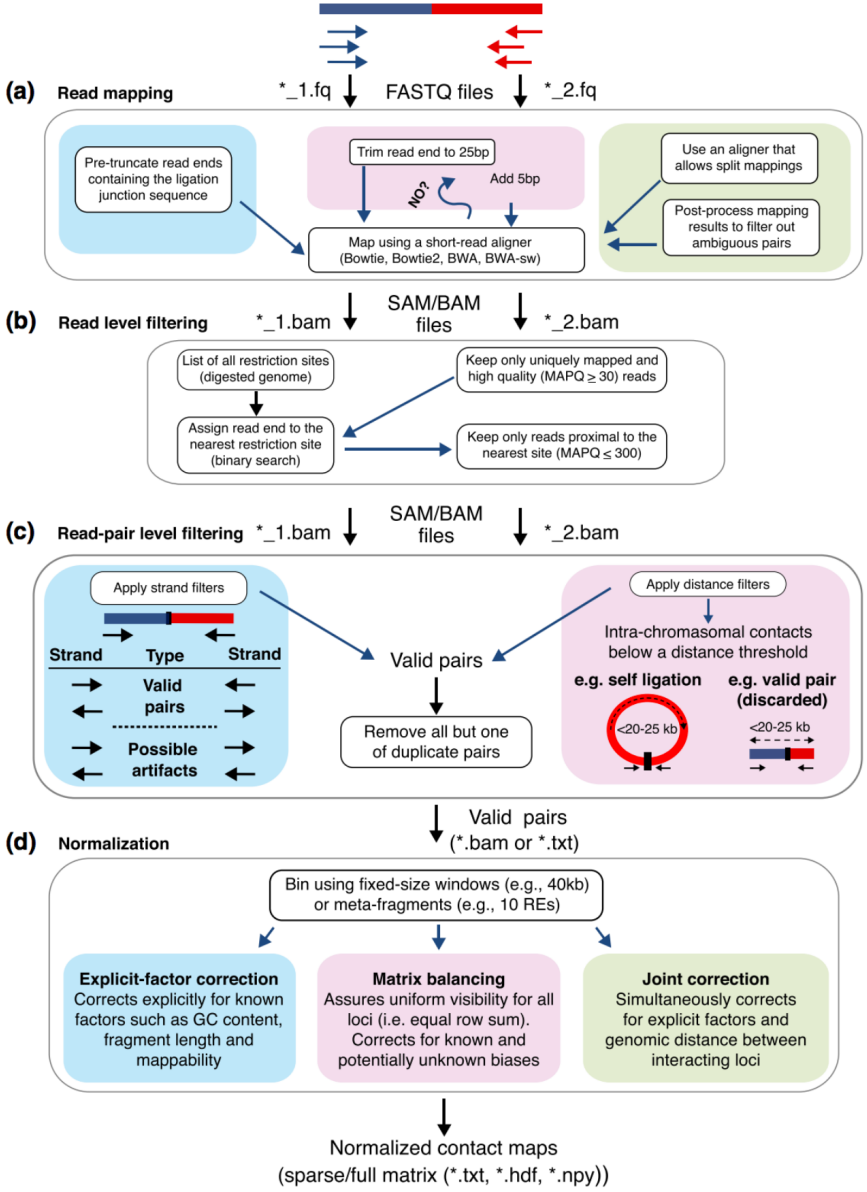
Top: the long-range contact pattern of a locus (left) indicates its nuclear neighborhood (right).

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 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. ***Rao SS*** *et al. Cell (2014)*

Overview of Hi-C analysis pipelines

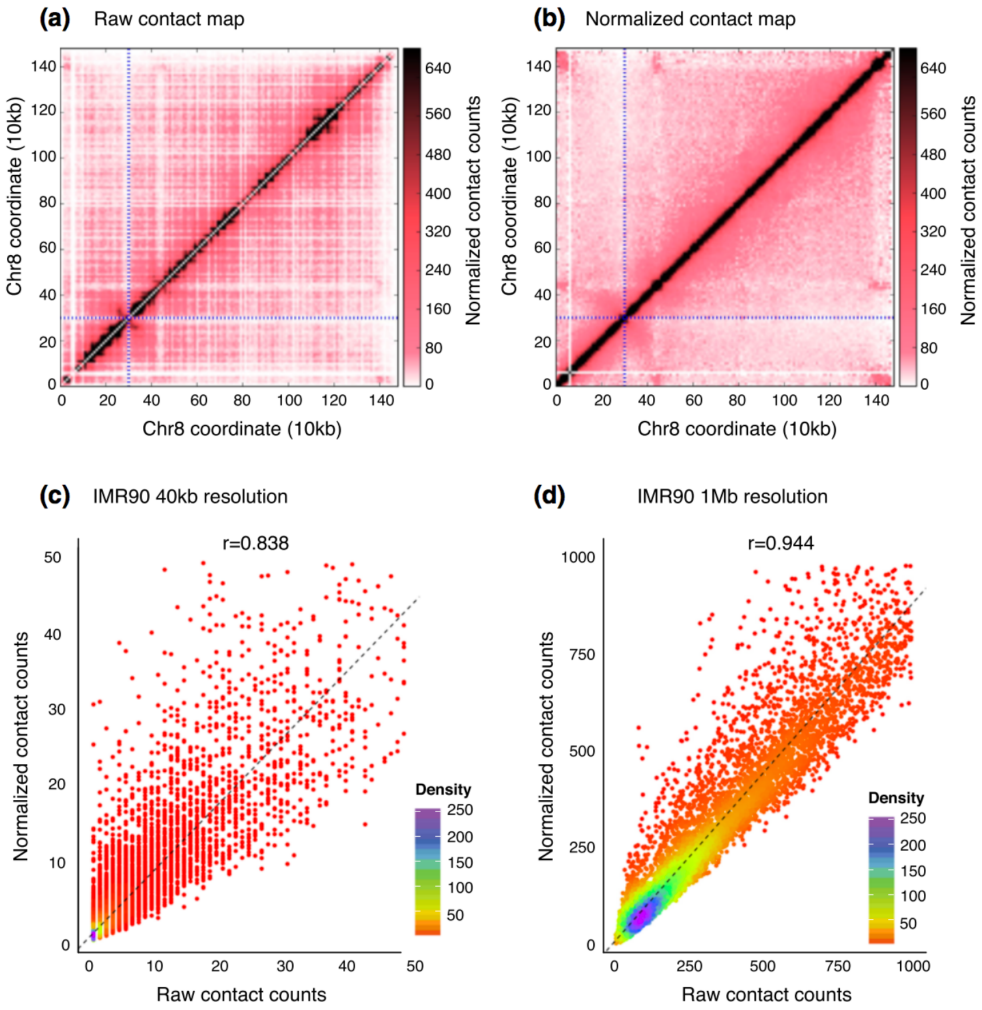


These pipelines start from raw reads and produce raw and normalized contact maps for further interpretation. Colored boxes represent alternative ways to accomplish a given step in the pipeline. RE, restriction enzyme. At each step, commonly used file formats (‘.fq’, ‘.bam’, and ‘.txt’) are indicated.

A. The blue, pink and green boxes correspond to pre-truncation, iterative mapping and allowing split alignments, respectively.
B. Several filters are applied to individual reads
C. The blue and pink boxes correspond to strand filters and distance filters, respectively.
D. Three alternative methods for normalization

Analysis methods for studying the 3D architecture of the genome. ***Ferhat Ay and William S. Noble****.*
Genome Biology (2015)

Impact of normalization on Hi-C contact maps



(a, b): Hi-C contact maps of chr8 from the schizont stage of the parasite Plasmodium falciparum at 10 kb resolution before and after normalization. Blue dashed lines represent the centromere location.

(c, d): Density scatter plots of counts before (x-axis) and after (y-axis) normalization of Hi-C data from the human cell line IMR90 at two different resolutions. Correlation values are computed using all intra-chromosomal contacts within human chr8. Only a subset of points are shown for visualization purposes

Analysis methods for studying the 3D architecture of the genome. ***Ferhat Ay and William S. Noble****.*
Genome Biology (2015)

Visualization of Hi-C data

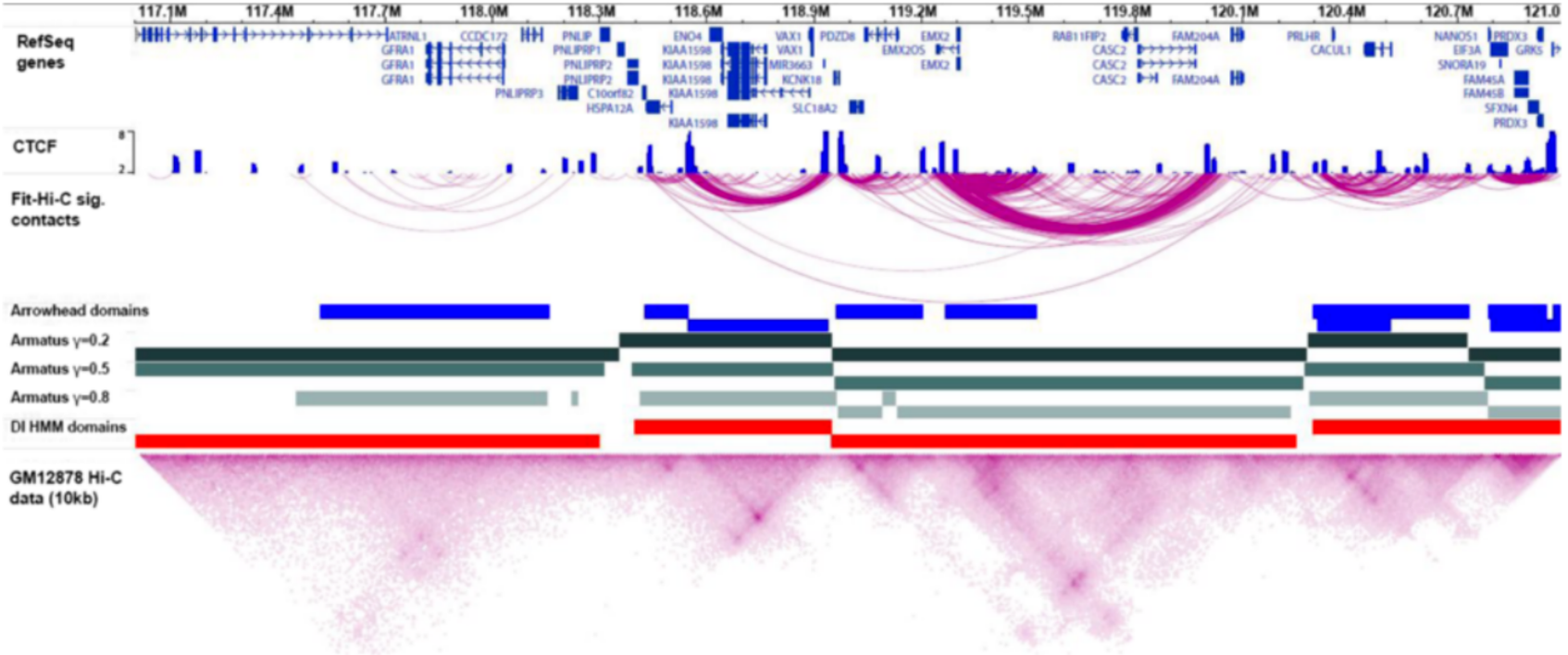
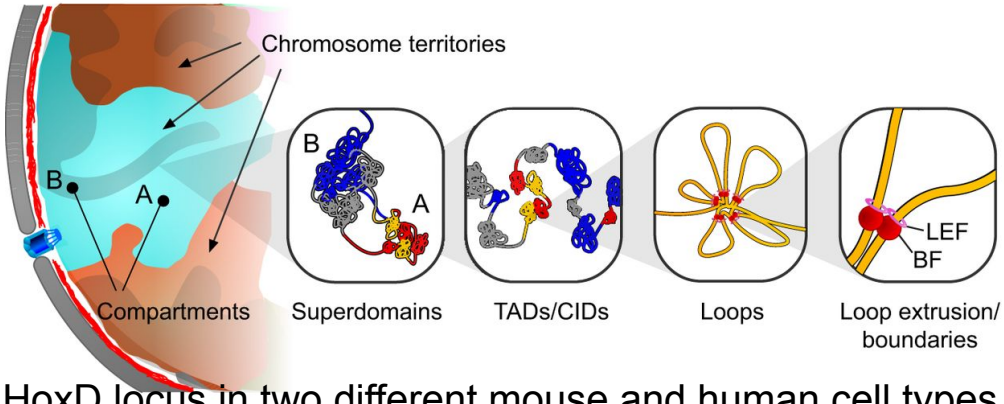


Fig. 3 Visualization of Hi-C data. An Epigenome Browser snapshot of a 4 Mb region of human chromosome 10. Top track shows Refseq genes. All other tracks display data from the human lymphoblastoid cell line GM12878. From top to bottom these tracks are: smoothed CTCF signal from ENCODE [130]; significant contact calls by Fit-Hi-C using 1 kb resolution Hi-C data (only the contacts >50 kb distance and $-\log(\text{p-value}) \leq 25$ 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

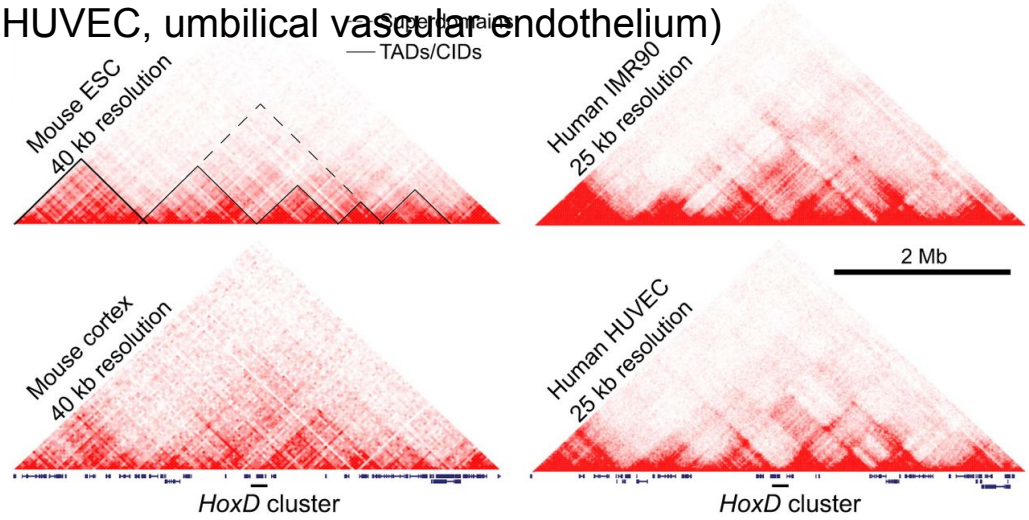
---

Analysis methods for studying the 3D architecture of the genome. ***Ferhat Ay and William S. Noble****.*
Genome Biology (2015)

Chromatin domain folding occurs at different scales



HoxD locus in two different mouse and human cell types (ESC, embryonic stem cell; IMR90, fetal lung fibroblast; HUVEC, umbilical vascular endothelium)

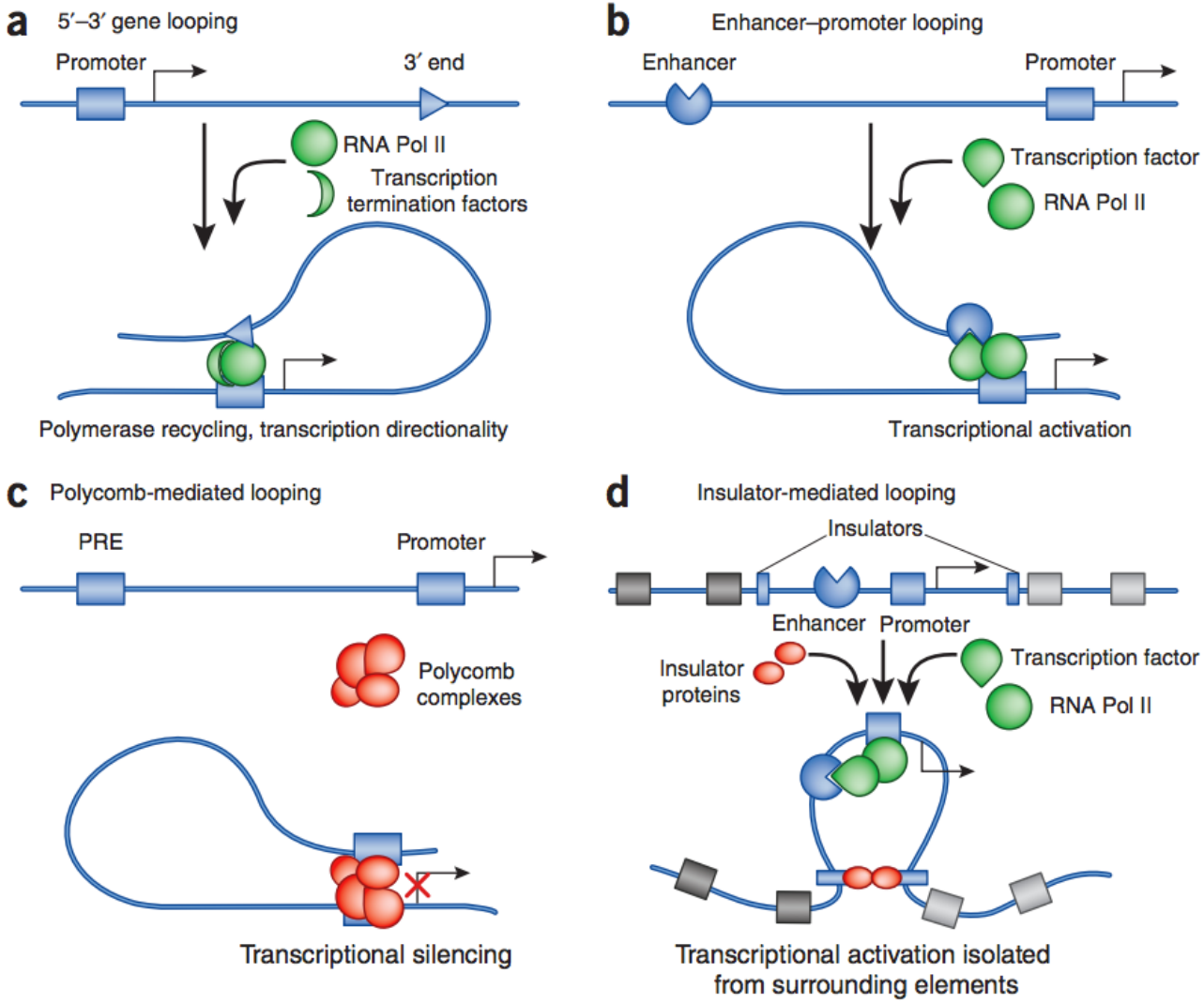


Hierarchical chromatin folding inside the nucleus, as uncovered by chromosome conformation capture.

TADs/CIDs are conserved between cell types and paralogous regions between species. The TAD structure is outlined (solid lines), as well as one superdomain (dashed lines). Overall patterns of contacts are very similar between cell types of the same species as well as between species

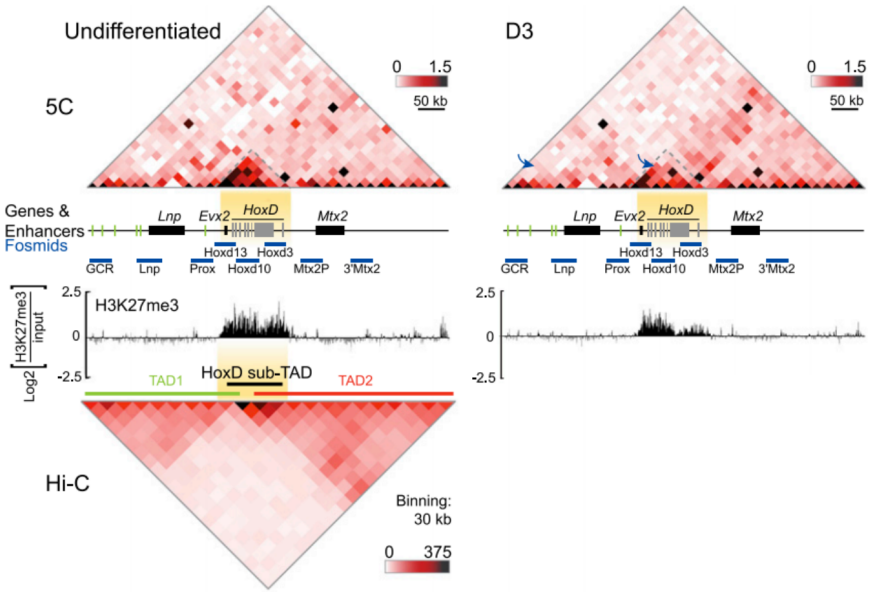
From single genes to entire genomes: the search for a function of nuclear organization. ***Pueschel*** *et al. Development (2016)*

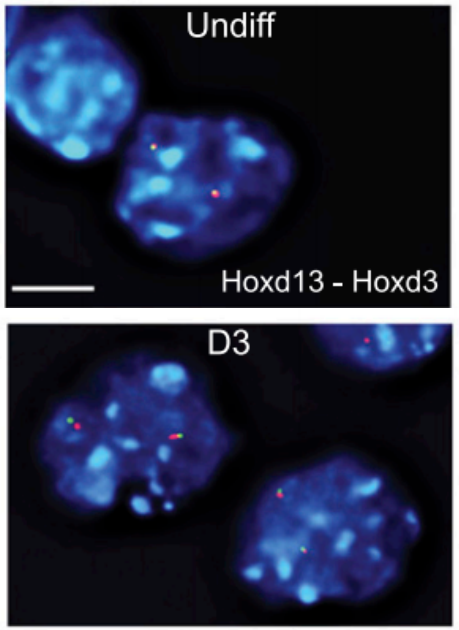
Four types of transcription regulatory chromatin loops



Adapted from Functional implications of genome topology. ***Cavalli G & Misteli T****. Nature Structural and Molecular Biology. (2013)*

What you *"C"* might not be what you see





"C" technologies

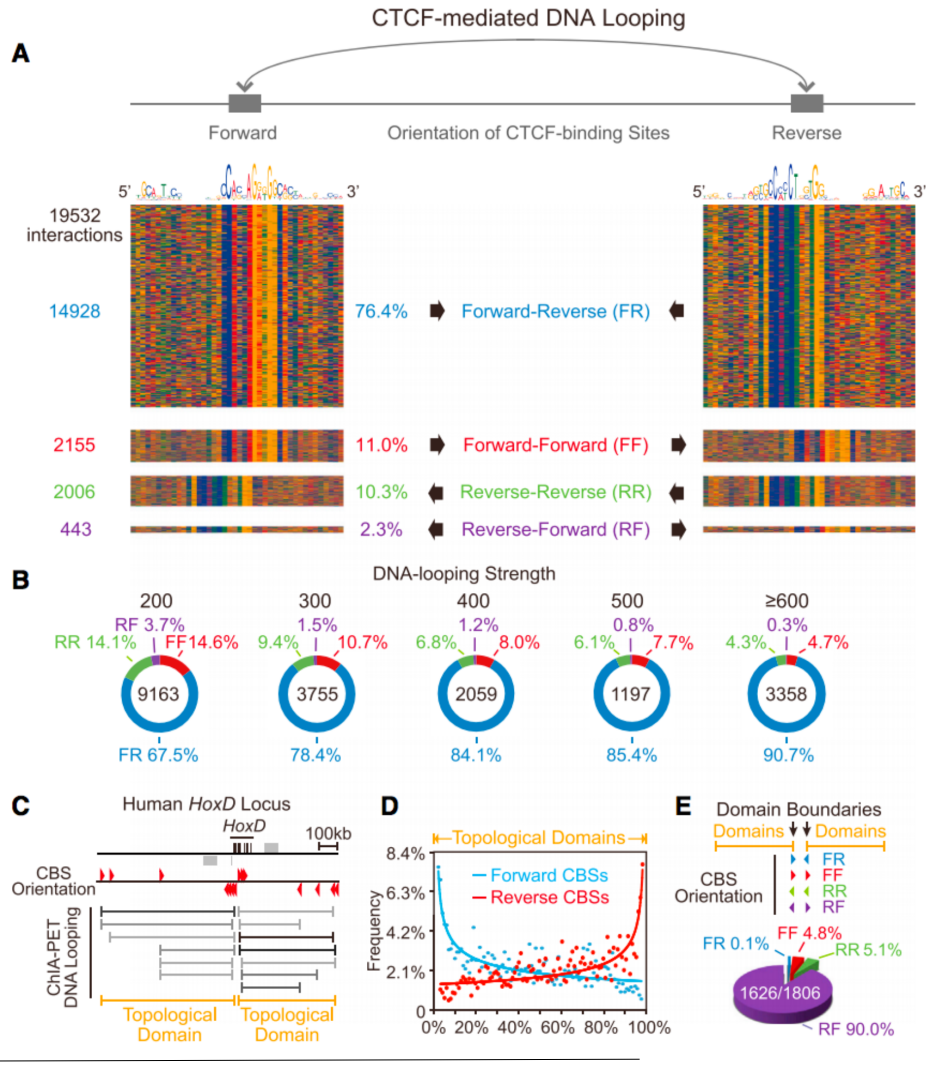
✓ high throughput
✓ high resolution
✗ low ligation (detection) efficiency
✗ indirect cross-linking via nuclear structure
✗ not readily applicable to low cell number

Imaging

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

Spatial genome organization: contrasting views from chromosome conformation capture and fluorescence in situ hybridization. ***Williamson*** *et al. Genes and Dev (2014)*

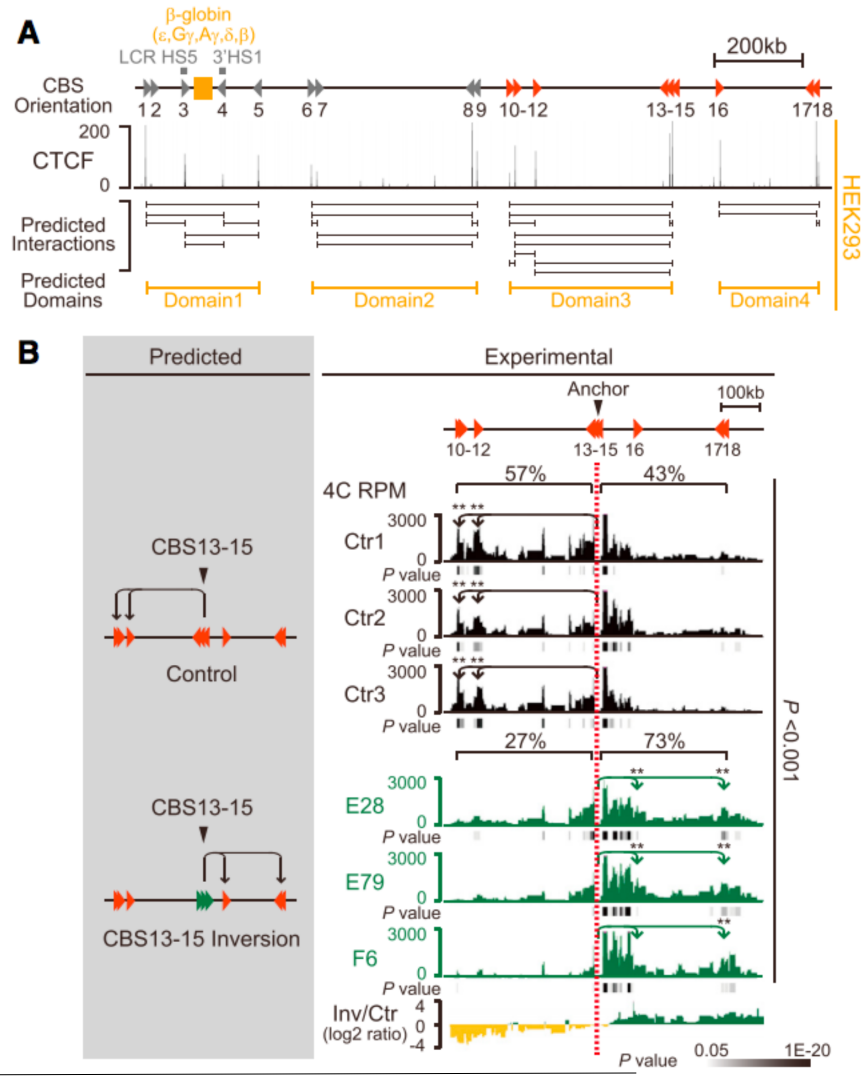
Role of CBS location and orientation in CTCF-mediated genome-wide DNA looping



CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. ***Guo*** *et al. Cell (2015)*

- Most CTCF-mediated long-range chromatin-looping interactions occur between CBS pairs in forward-reverse orientation
- The percentage of CBS pairs in the forward reverse orientation increases as chromatin-looping strength is enhanced
- Schematic of two TADs in HoxD locus with CBS orientations
- Cumulative patterns of CBS orientations of TADs in humans
- Distribution of genome-wide orientation configurations of CBS pairs (vast majority in RF) located in boundaries between neighboring TADs in human K562 genome

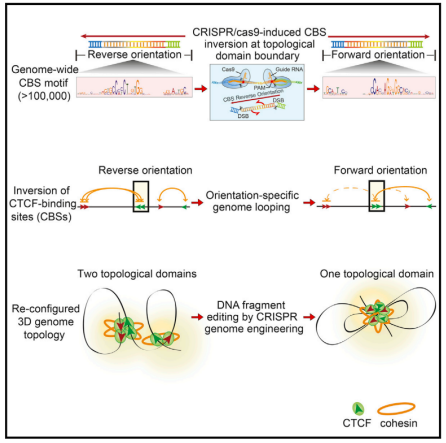
CRISPR inversion confirms CTCF/Cohesin-mediated directional looping mechanism

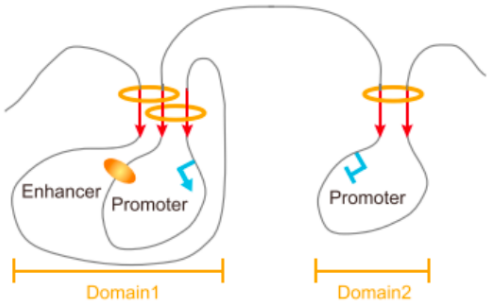


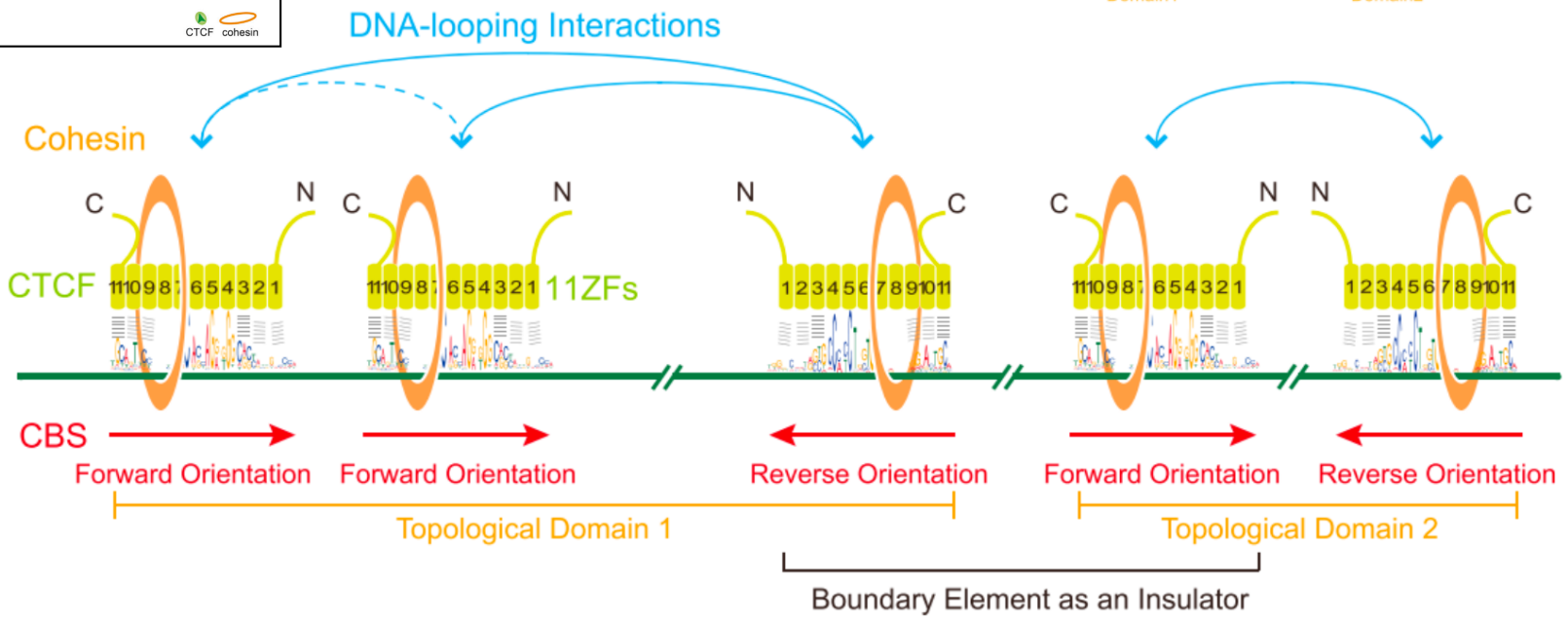
- Prediction of looping interactions and TADs based on CTCF occupancy in HEK293 cells in the Human b-globin region
- Predicted interactions (left) and the altered looping directions (right) in three subcloned CRISPR cell lines with inversion of the CBS13-15 confirmed by 4C with CBS13-15 as an anchor. Average log2 ratios of interactions between inversions and controls are also indicated. **p < 0.01

CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. ***Guo*** *et al. Cell (2015)*

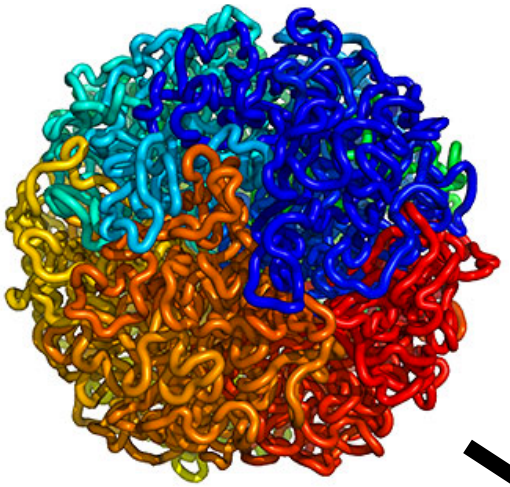
Model of CTCF/Cohesin-Mediated Topological 3D Genome Folding and Gene Regulation

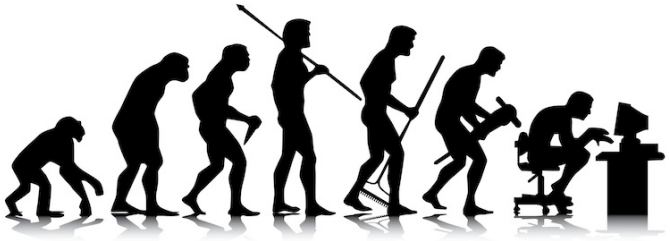






CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. ***Guo*** *et al. Cell (2015)*





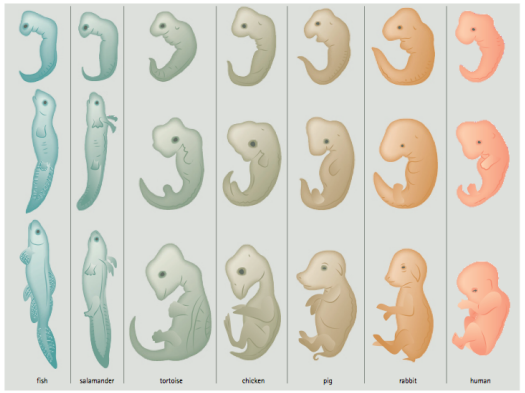
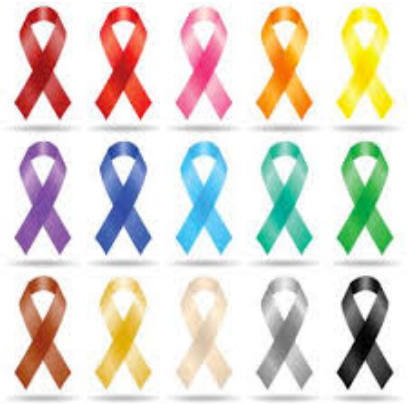


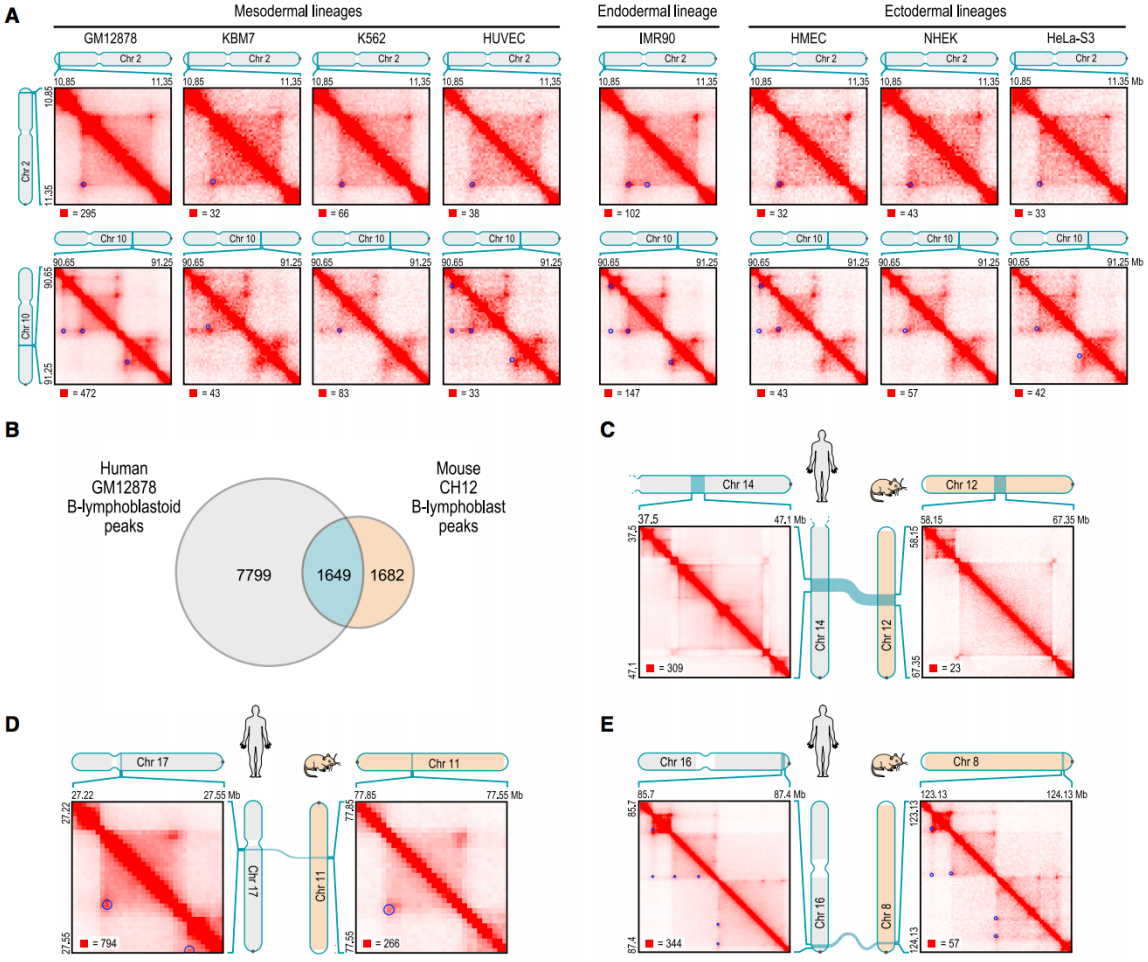
Figure 6.2.6 Stages in the embryonic development of vertebrates



Evolution Development Disease

Adapted from Google Images

Loops are often preserved across cell types and from human to mouse

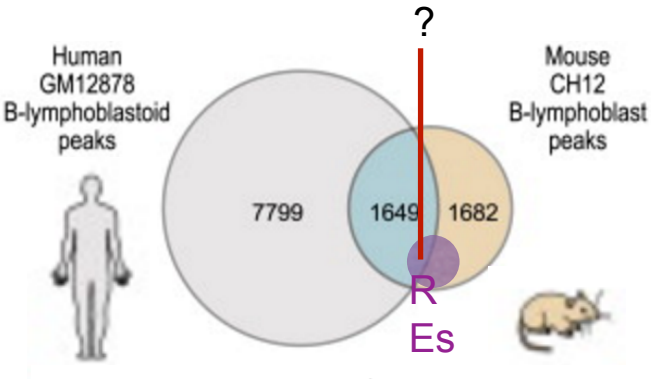


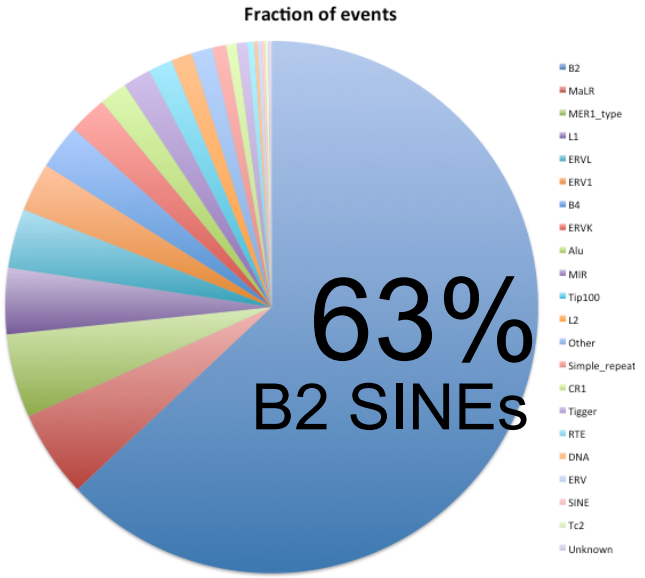
(A) Examples of peak and domain preservation across cell types. Annotated peaks are circled in blue. All annotations are completely independent.
(B) Of the 3,331 loops we annotate in mouse CH12-LX, 1,649 (50%) are orthologous to loops in human GM12878.
(C–E) Conservation of 3D structure in synteny blocks. The contact matrices in (C) are shown at 25 kb resolution. (D) and (E) are shown at 10 kb resolution

A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. ***Rao SS*** *et al. (Cell 2014)*

Hi-C analysis gives insights into anchoring site contribution by repetitive elements

- Despite 75 million years of evolution, TAD boundaries are strikingly conserved (~50%) between human and mouse[1,2]
- Rao SS et al. report 3331 looping events in Mouse CH12-LX (B-lymphoblasts)
- Of these 6662 total looping sites, 3832 are anchored at a pair of convergent CTCF/RAD21/ SMC3 binding sites that contribute to the loop formation.
- 473 such binding sites intersect with repetitive elements
  - 298 such events take place solely in mouse-specific family of B2 SINEs
    - 144 such events take place solely in mouse-specific subfamily of B3 elements



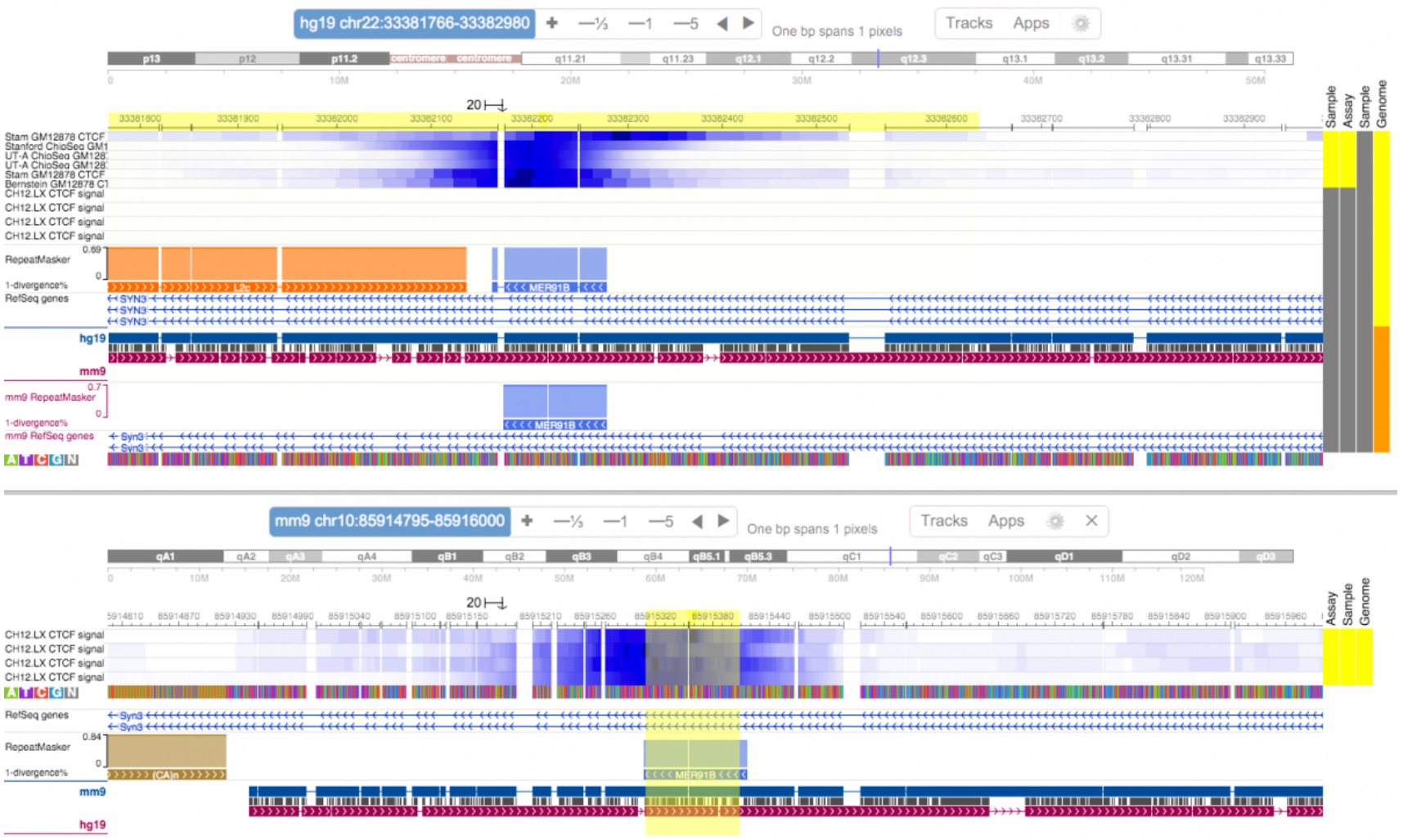


1. Topological domains in mammalian genomes identified by analysis of chromatin interactions. ***Dixon JR*** *et al. Nature (2012)*

2. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. ***Rao SS*** *et al. (Cell 2014)*

3. Choudhary (Unpublished data, March 2016)

Conserved anchoring site deposited by the same TEs

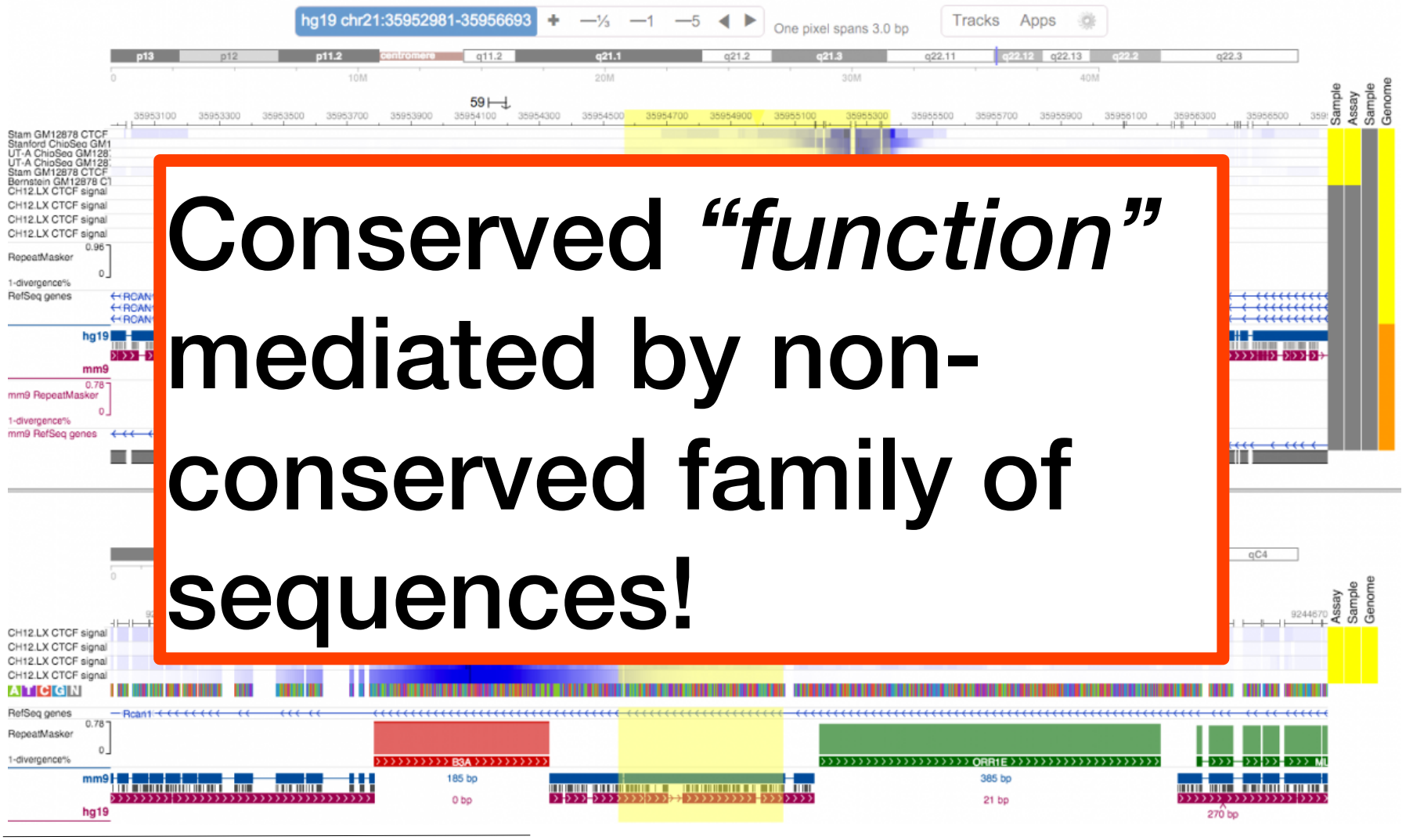


Choudhary (Unpublished observations, March 2016)

Conserved anchoring site deposited by different TEs

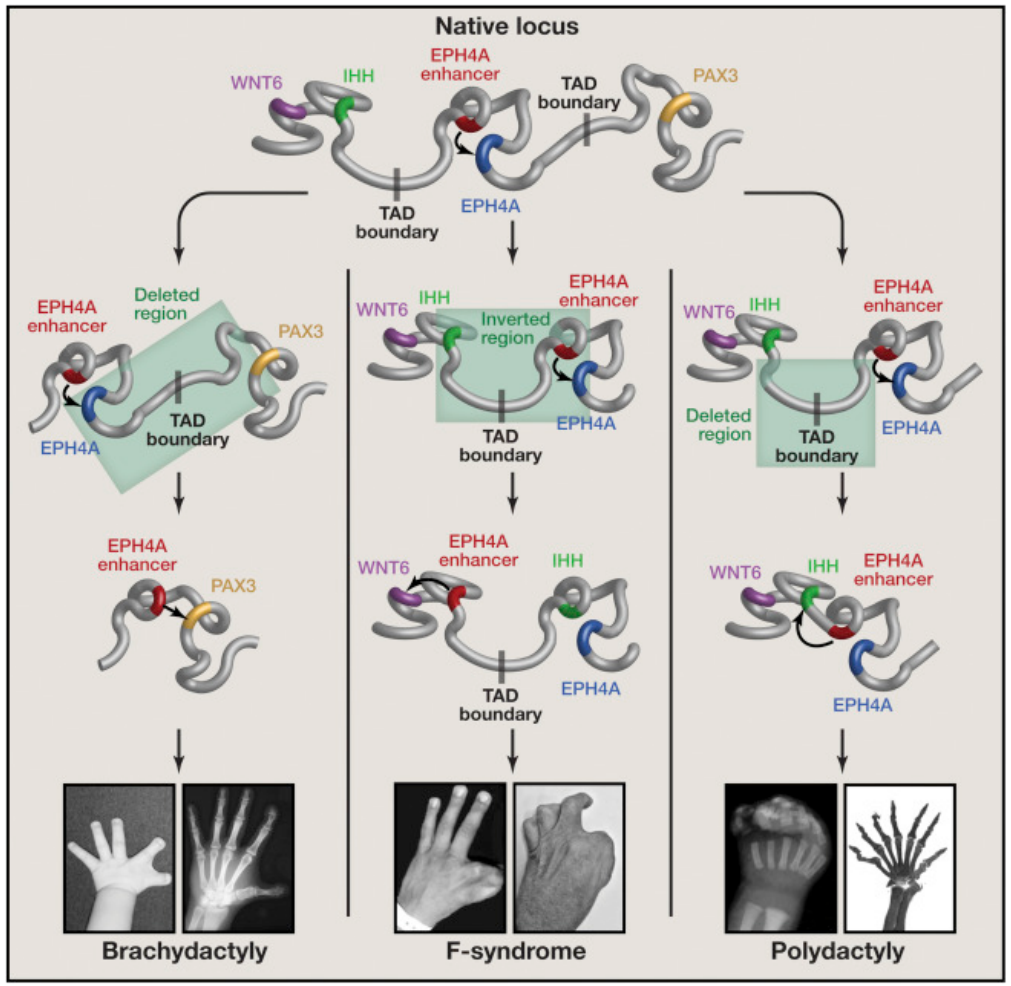
Choudhary (Unpublished observations, March 2016)

Conserved anchoring site deposited by a species-specific TEs *(turnover)*



Choudhary (Unpublished observations, March 2016)

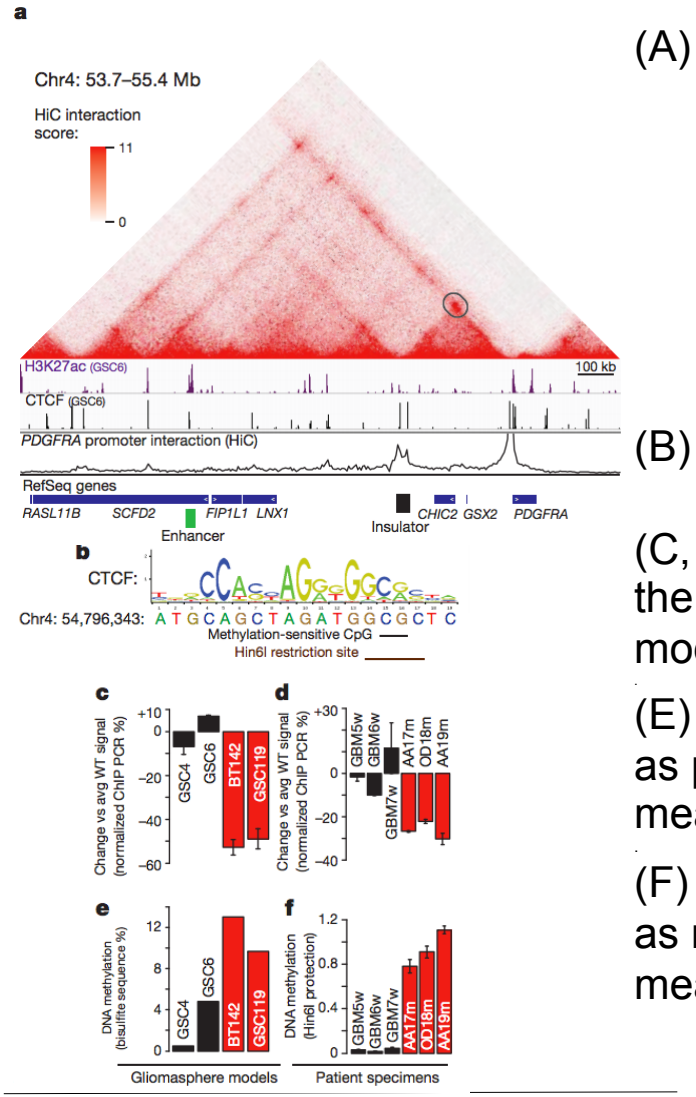
Breaking TADs: Role of chromatin topology in developmental disorders



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



(A) HiC map for IMR90 cells. Contact domain structure shown for a 1.7-Mb region containing PDGFRA. Convergent CTCF sites anchor a loop that separates PDGFRA and FIP1L1 (black circle). Interaction trace (below) depicts HiC signals between the PDGFRA promoter and all other positions in the region. Genes, FIP1L1 enhancer (per H3K27ac) and insulator (per HiC and CTCF binding) are indicated

(B) The right CTCF peak in the insulator contains a CpG methylation sensitive CTCF-motif.

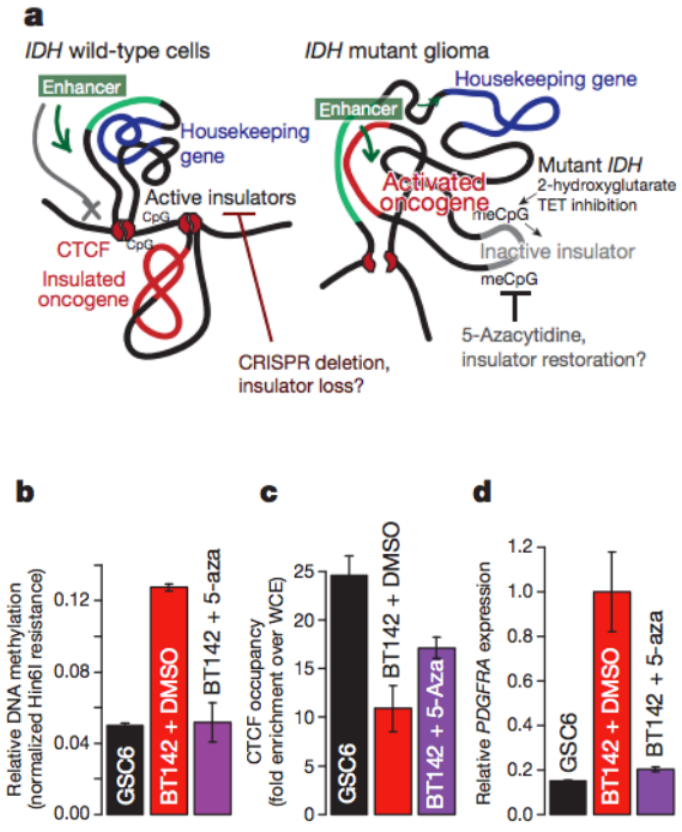
(C, D) ChIP-qPCR data shows that CTCF occupancy over the boundary is reduced in IDH mutant (red) gliomas and models, relative to wild type (black)

(E) Methylation levels of the CpG in the CTCF motif plotted as percentage of alleles protected from conversion measured in gliomaspheres by bisulfite sequencing

(F) Methylation levels of the CpG in the CTCF motif plotted as relative protection from methylation-sensitive restriction measured in gliomaspheres

Insulator dysfunction and oncogene activation in IDH mutant gliomas. ***Flavahan WA*** *et al. Nature (2016)*

Boundary methylation and CTCF occupancy affect PDGFRA expression and proliferation



(A) Schematic depicts chromatin loops and boundaries in the PDGFRA locus. In IDH wild-type cells (left), intact boundary insulates oncogene. Disruption of the boundary by removing the CTCF motif should activate the oncogene. In IDH mutant cells (right), hypermethylation blocks CTCF, compromising the boundary and allowing enhancer to activate the oncogene. Demethylation should restore CTCFmediated insulation. meCpG, methylated CpG

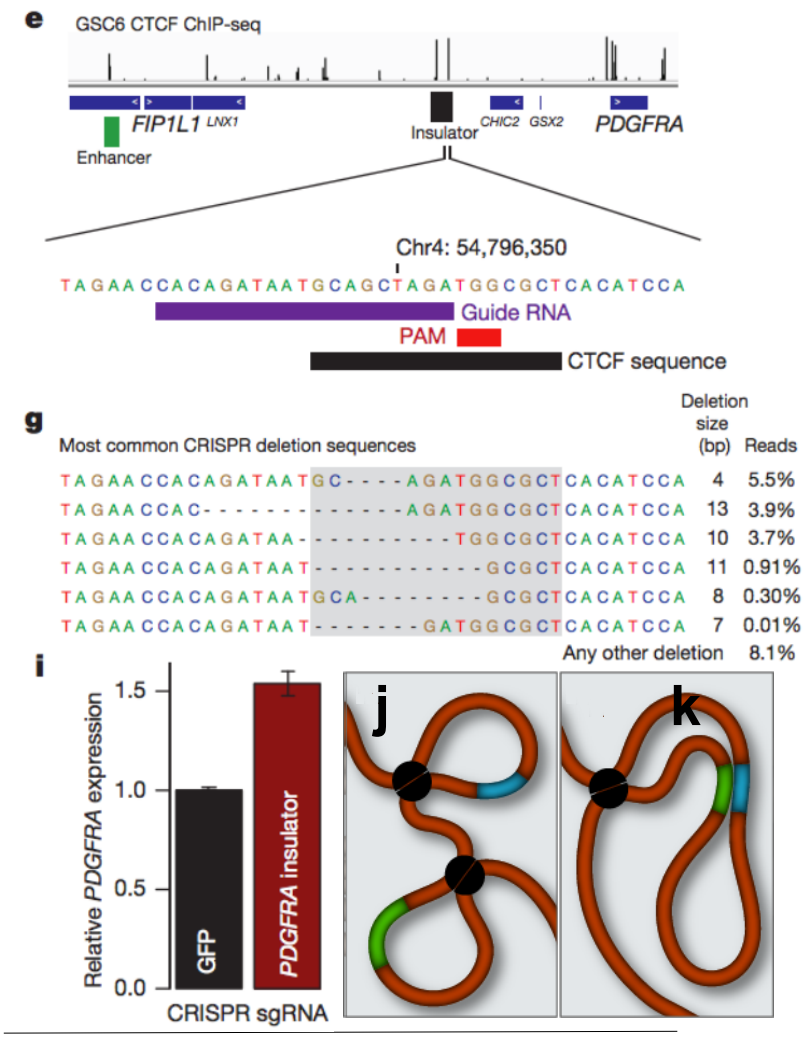
(B) Plot compares CpG methylation in the CTCF motif in IDH wild-type gliomaspheres (black), IDH1 mutant gliomaspheres (red), and IDH1 mutant gliomaspheres treated with 5μM 5-aza for 8 days (purple).

(C) Plot compares CTCF occupancy over the boundary. DMSO, dimethylsulfoxide; WCE, whole-cell extract.

(D) Plot compares PDGFRA expression. Demethylation restores PDGFRA insulation in IDH1 mutant gliomaspheres.

Insulator dysfunction and oncogene activation in IDH mutant gliomas. ***Flavahan WA*** *et al. Nature (2016)*

Boundary methylation and CTCF occupancy affect PDGFRA expression and proliferation



(E) CTCF binding shown for the FIP1L1/ PDGFRA 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

Insulator dysfunction and oncogene activation in IDH mutant gliomas. ***Flavahan WA*** *et al. Nature (2016)*
Adapted from image by Lauren Solomon, Broad Communications

Summary (1)

- 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

Summary (2)

- 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

Other interesting topics

- RNA component
  - RNAi, miRNA, X inactivation, HOTAIR, PiwiRNA
- Reprogramming
- Cloning
- Population epigenetics
- Evolution of DNA methylation
- Evolution of epigenome