Sequencing technologies for epigenomic profiling

November 6, 2023

Ben Sunkel, PhD
benjamin.sunkel@nationwidechildrens.org

Nationwide Children's
When your child needs a hospital, everything matters.℠

The Ohio State University
College of Medicine

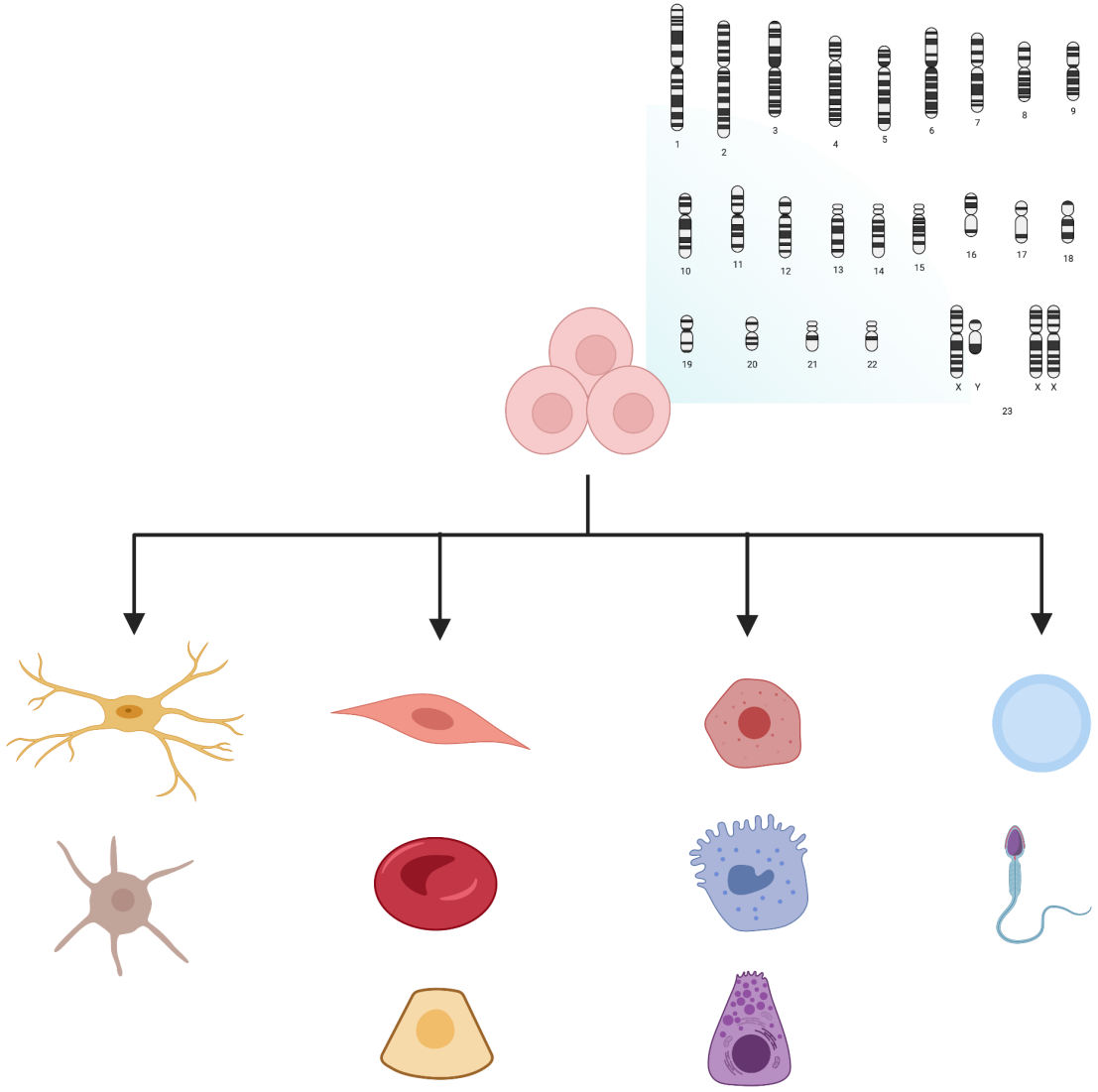
Outline

- One person's meaning of "epigenetics"
- Introduction to a few features of the epigenome
    - •DNA methylation
    - •Histones and histone post-translational modifications
    - •Chromatin remodeling
    - •Transcription factors
    - •Chromatin conformation
- An overview of three epigenomic techniques and some analysis tools
    - •ChIP-seq
    - •ATAC-seq
    - •HiC

Epigenetics (Epigenomics)

How do multicellular organisms create many distinct cells from one genome?

- Instruction manual for interpreting the genome
- Spatial and temporal control
- Heritable or at least durable

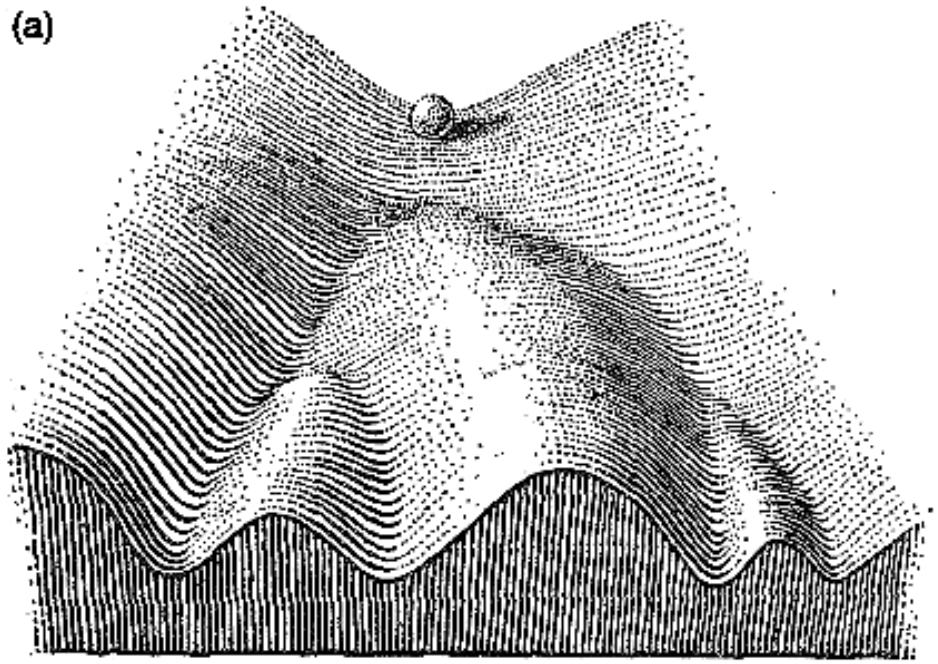


Epigenetics

From the NIH Roadmap Epigenomics Project

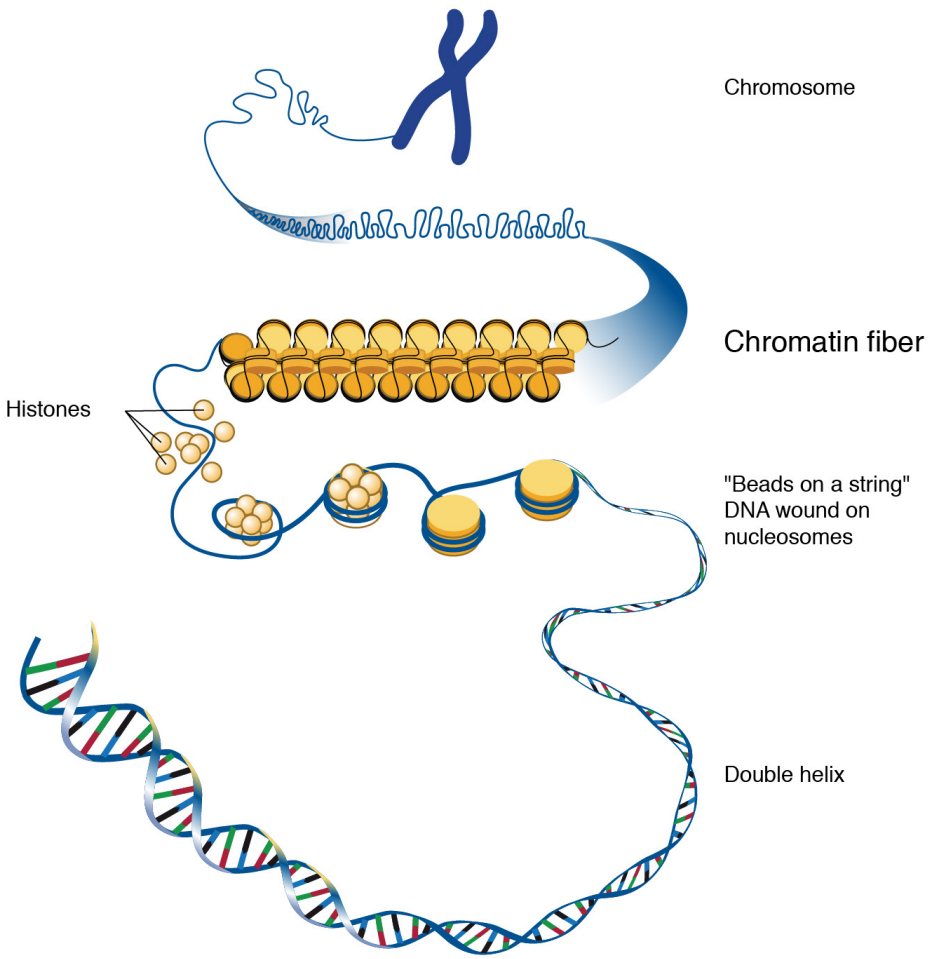
> "… the study of changes in the regulation of gene activity and expression that are not dependent on gene sequence."

> "…refers to both heritable changes in gene activity and expression and also stable, long-term alterations in the transcriptional ***potential*** of a cell that are not necessarily heritable."



Conrad Waddington's Epigenetic Landscape

The Many Features of Epigenetics



Multiple systems dedicated to the instructive packaging and appropriate decoding of our genome

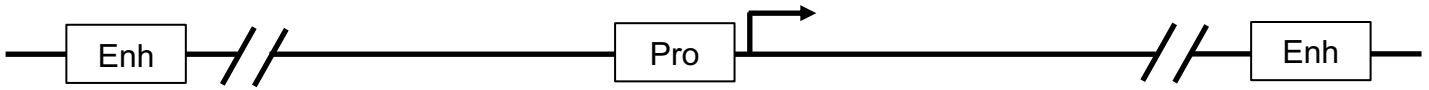
Their goal:

1. Keep the correct genes accessible and active
2. Compact and repress the rest
3. Protect and repair the genome
4. Maintain responsiveness to stimuli

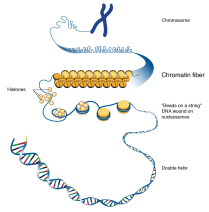
Our goal: Annotate gene regulatory elements

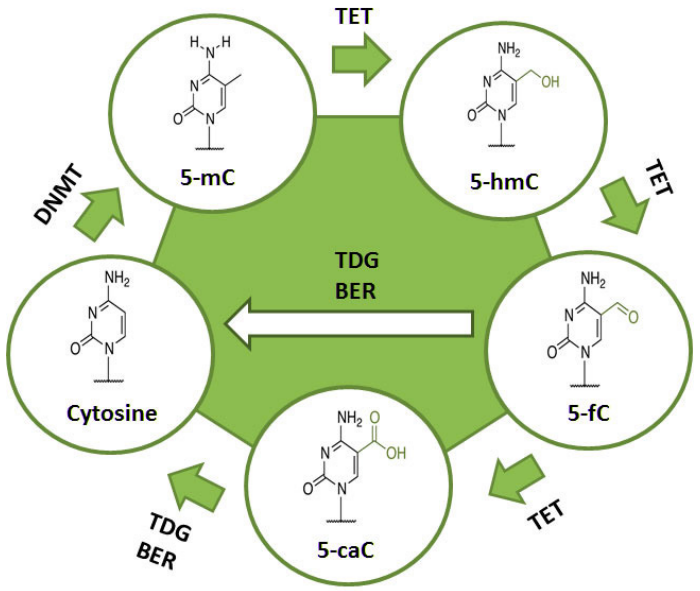
Promoters – Adjacent to TSS, assemble general transcription machinery

Enhancers – Distance/position-independent, often tissue-specific expression modifiers



The Many Features of Epigenetics



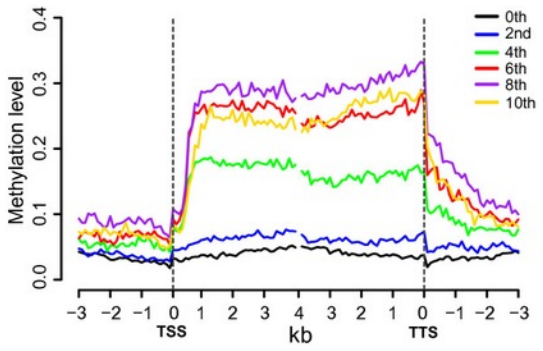


DNA Methylation (CG context)

1. "Written" by DNA Methyl Transferases (DNMTs, maintenance and *de novo*)
   - S-adenosyl methionine (SAM)
2. Progressively "removed" by ten-eleven translocation methylcytosine dioxygenases (TETs) and DNA repair machinery
   - Alpha-ketoglutarate + O2 → Succinate + CO2

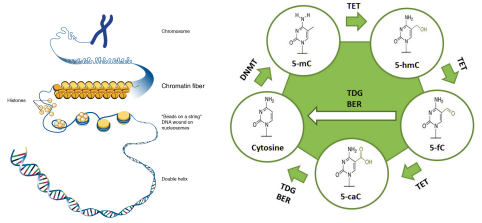
Genome-wide Patterns

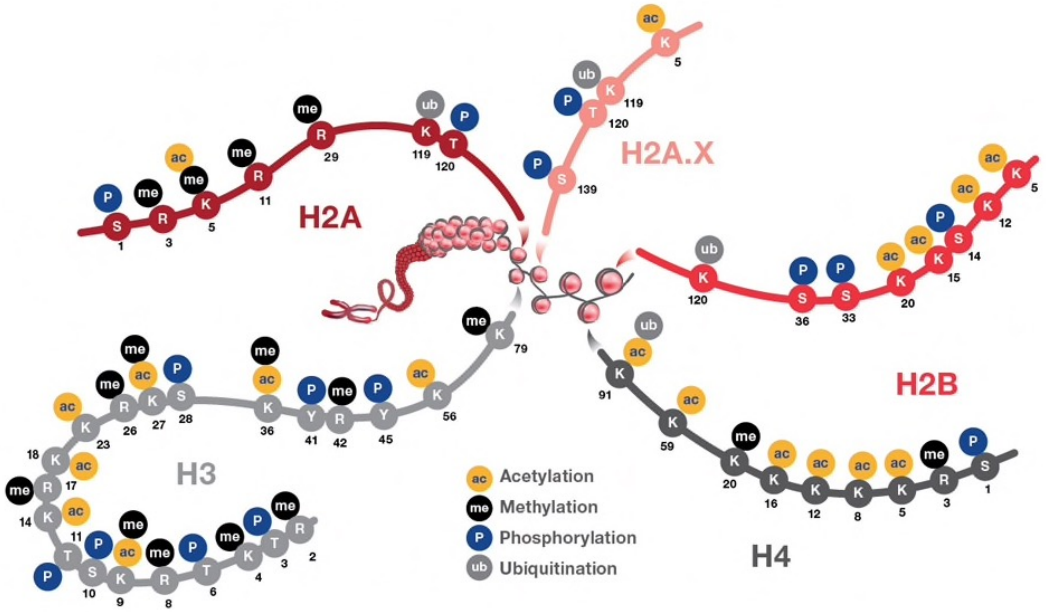
1. Majority of CG dinucleotides are methylated
2. Transposable elements are silenced by cytosine methylation
3. "CG islands" in active gene promoters tend to be hypomethylated
4. Gene expression is positively correlated with gene body methylation
5. DNA methylation is correlated with other epigenetic features



Wang et al., BMC Genomics 2014

The Many Features of Epigenetics





Histone post-translational modifications

1. Too many enzymes to list – acetyltransferases/deacetylases, kinases/phosphatases, ubiquitin ligases, etc.
2. Can result in charge neutralization, disrupting histone-DNA interaction
3. Creates substrates for myriad protein recognition motifs: bromodomains, chromodomains, etc.

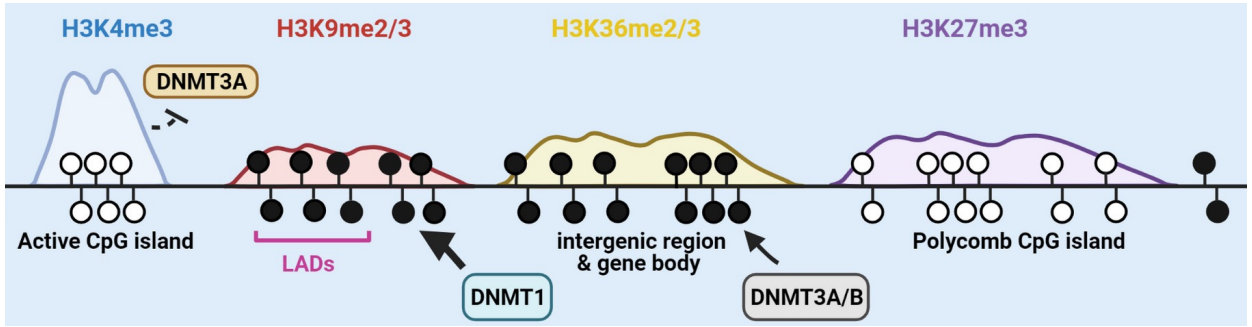
Shorthand Nomenclature

Histone H3, lysine 4, trimethylation = H3K4me3

Histone H2A, lysine 119, ubiquitination = H2AK119ub

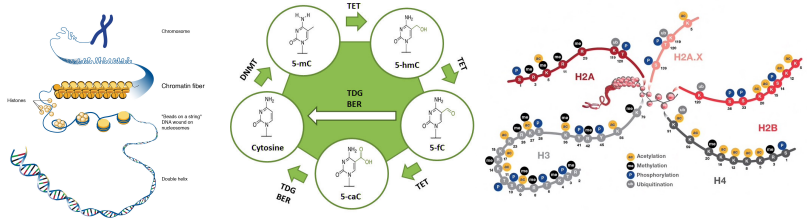
Genome-wide Patterns

1. H3K4me3 in promoters, H3K4me1 in enhancers
2. H3K27ac in active regulatory elements, H3K27me3 in repressed regions
3. H3K36me3 in elongating genes
4. H3K9me3 in repressed heterochromatin
5. γH2AX at double-stranded breaks

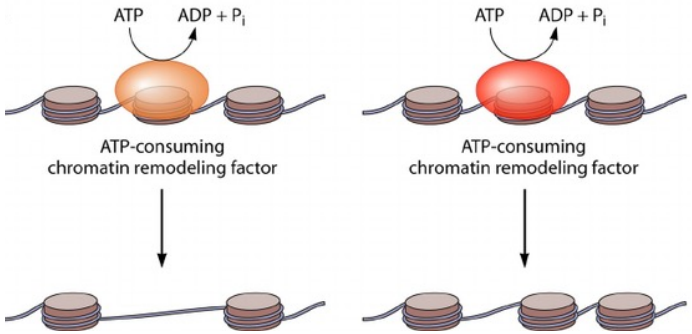


Li et al., EMBO Reports 2021

The Many Features of Epigenetics

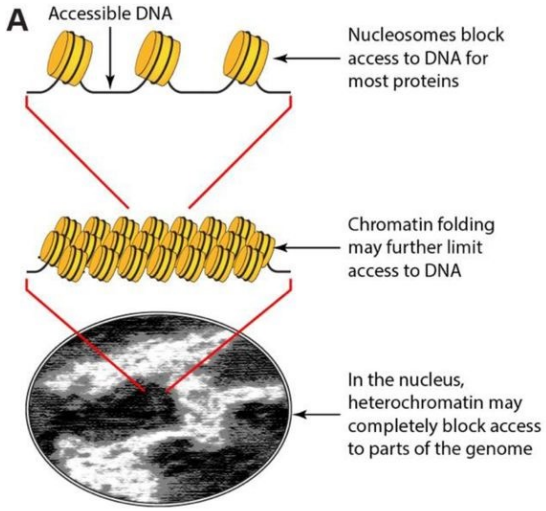


Chromatin Remodeling



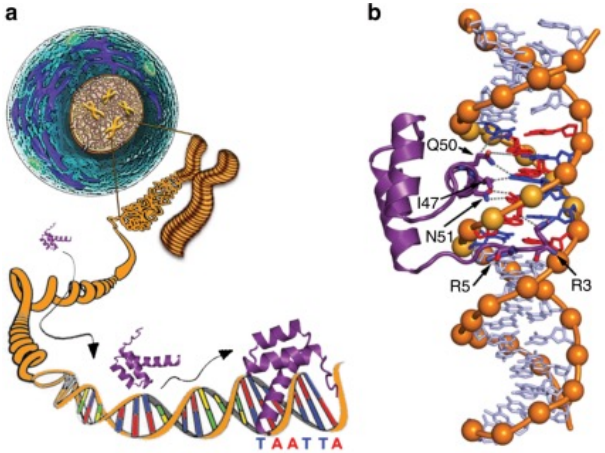
- Multi-subunit protein complexes
- Establish proper nucleosome positioning/spacing in regulatory elements
- Example: SWI/SNF

Chromatin Accessibility



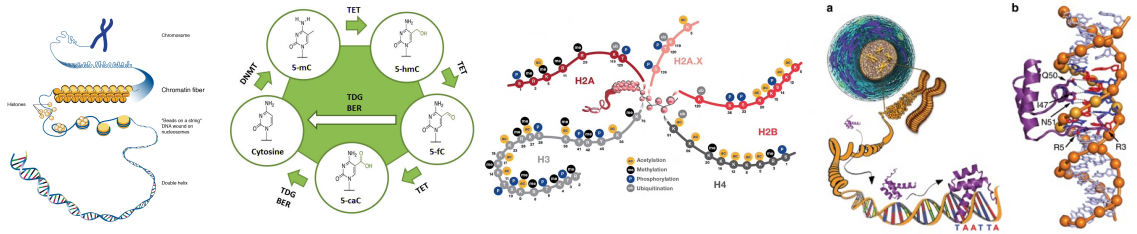
- Varies throughout the genome
- Often a signature of a functional regulatory element
- Required for binding of many proteins to the genome

Transcription Factor Binding

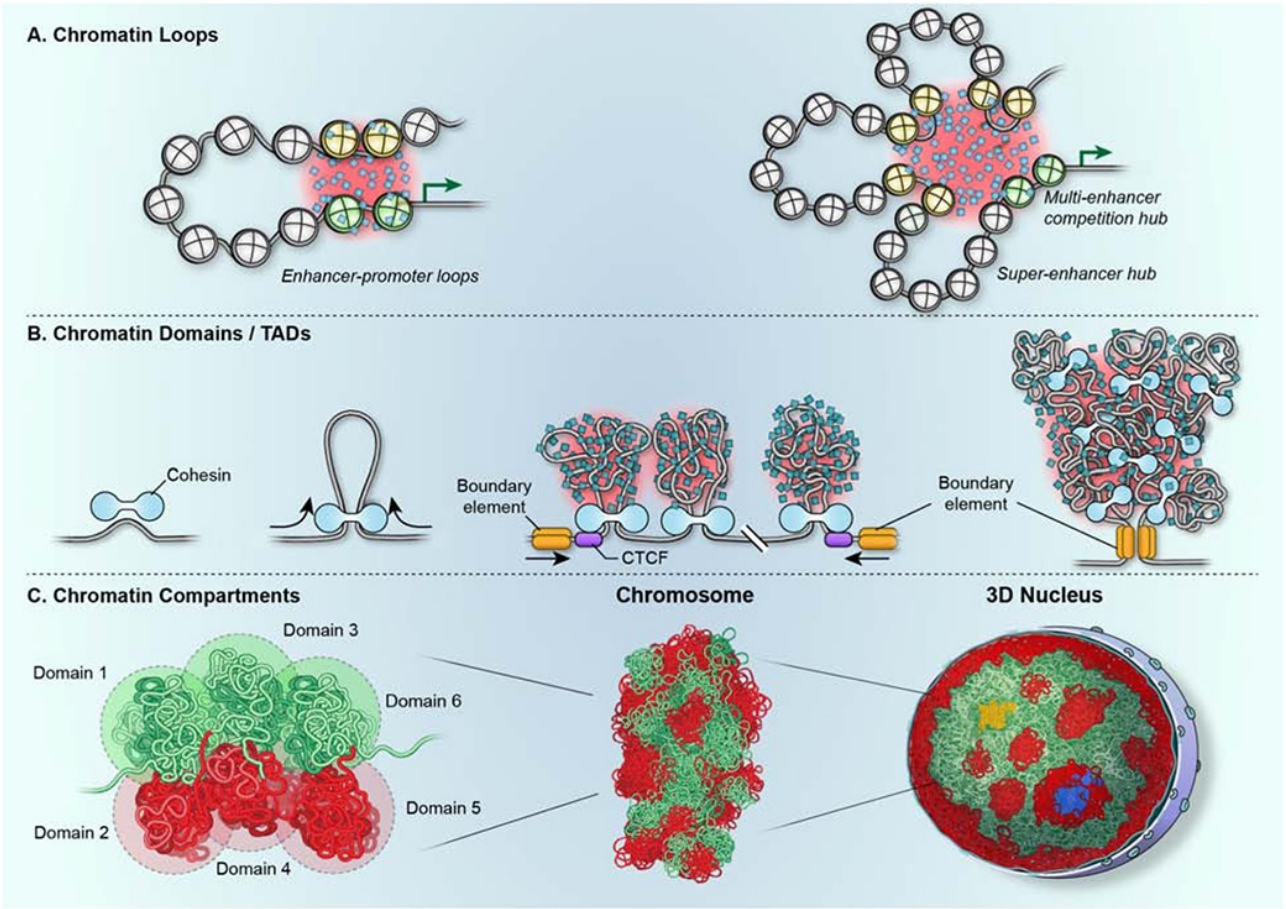


- Large families of tissue-specific proteins
- Bind to specific DNA sequence motifs, mostly in accessible chromatin
- Mediate recruitment of additional gene-regulatory factors

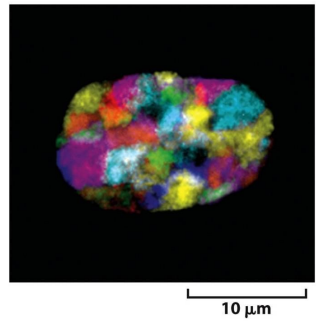
The Many Features of Epigenetics

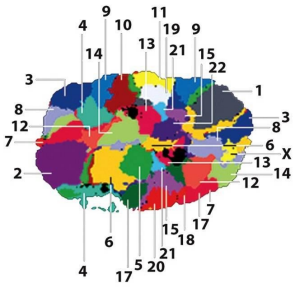


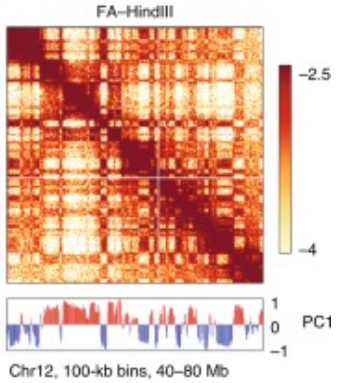
Chromatin Conformation



Early imaging techniques teach us that our genome organizes into chromosome territories

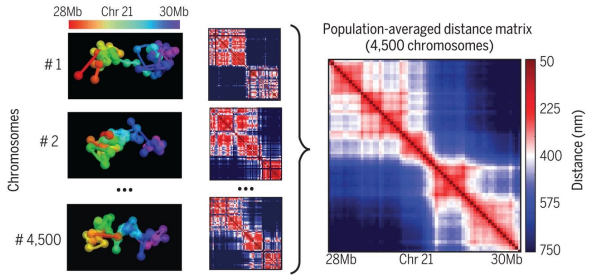






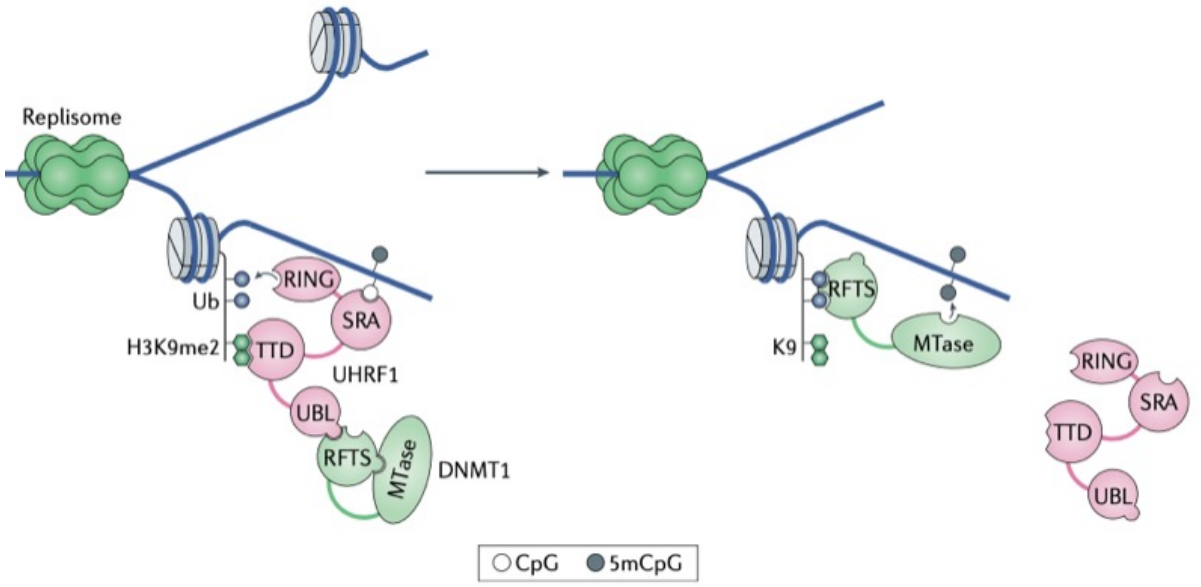
Proximity ligation techniques coupled with NGS reveal principles of genome folding at high resolution (< 1 kb)

New imaging and single-cell techniques reveal stochasticity of chromatin interaction events

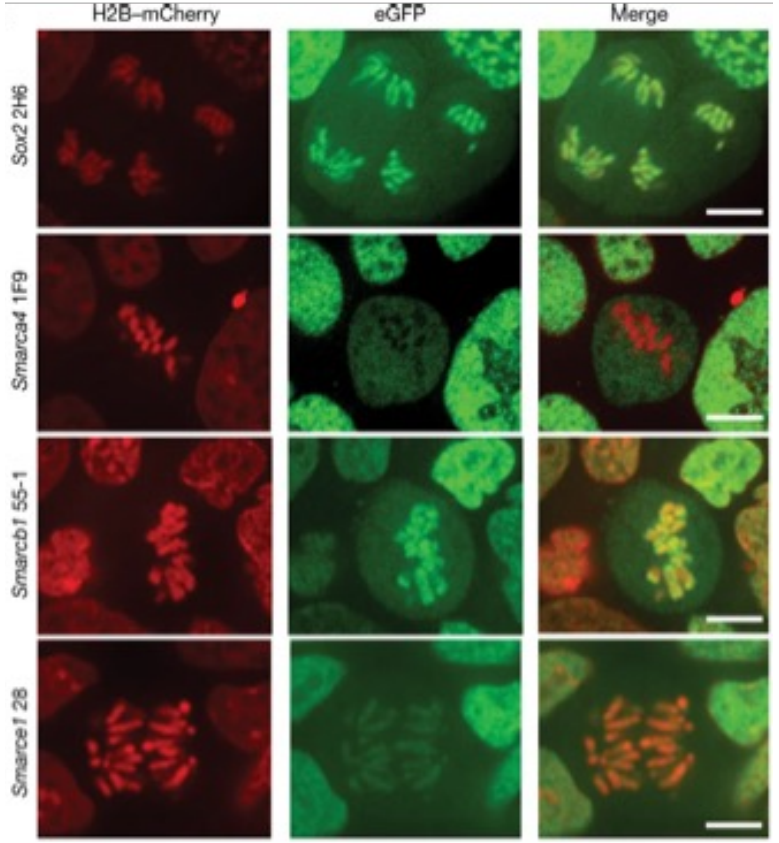


Misteli, Cell 2020

Mitotic transmission of epigenetic states



Greenberg and Bourc'his, *Nat Rev Mol Cell Biol* 2019

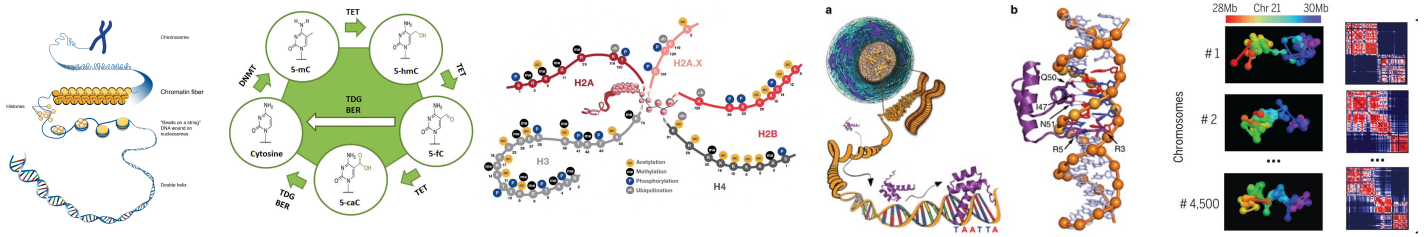


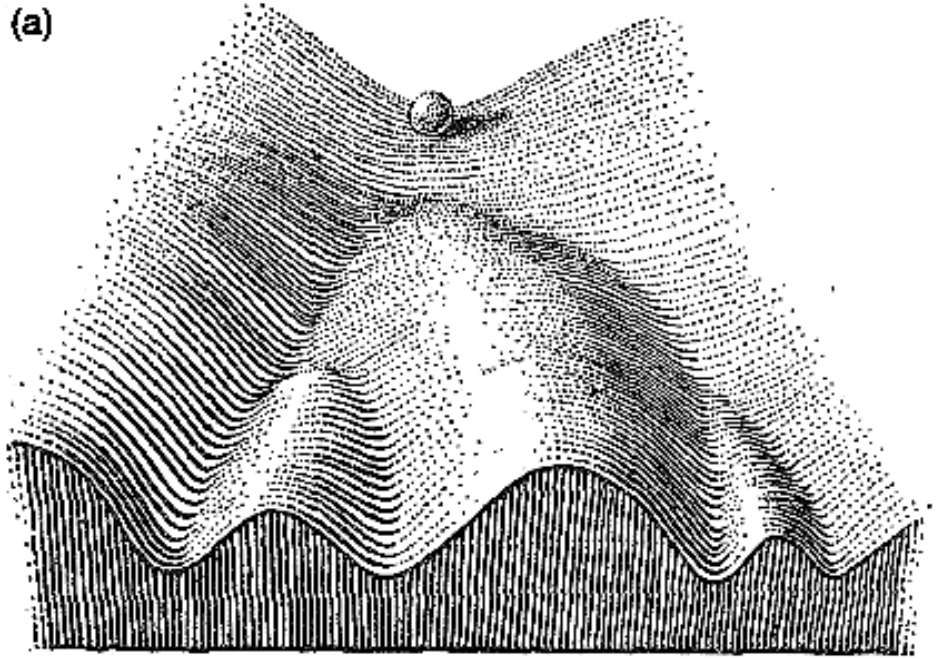
Zhu et al., *Nature* 2023

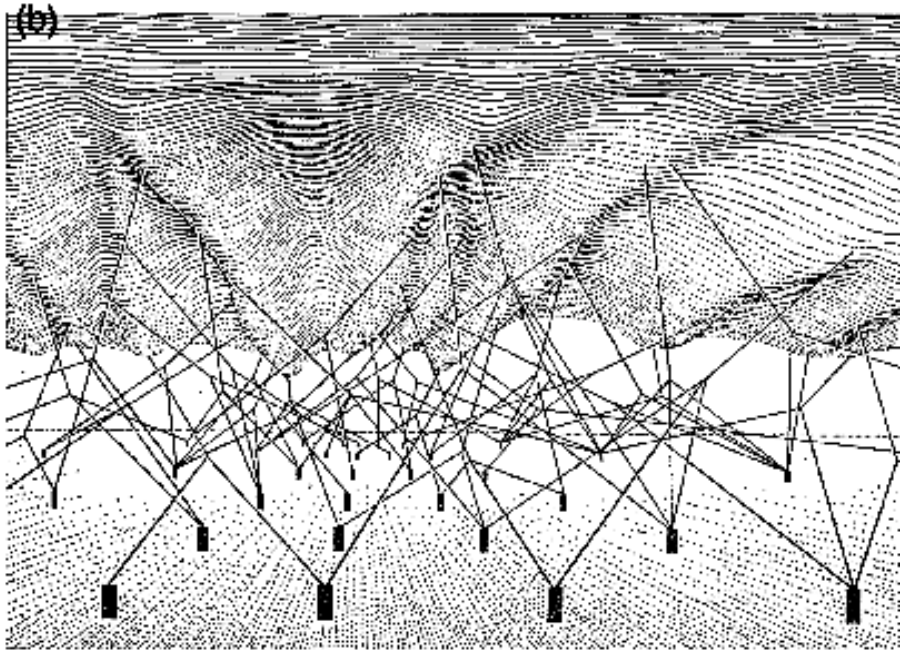
Example of mechanisms facilitating epigenetic memory

1) Maintenance DNMT1 modifying hemi-methylated CpGs during DNA replication
2) Mitotic retention of lineage defining transcription factors in addition to core subunits of chromatin remodeling complexes

The Many Features of Epigenetics

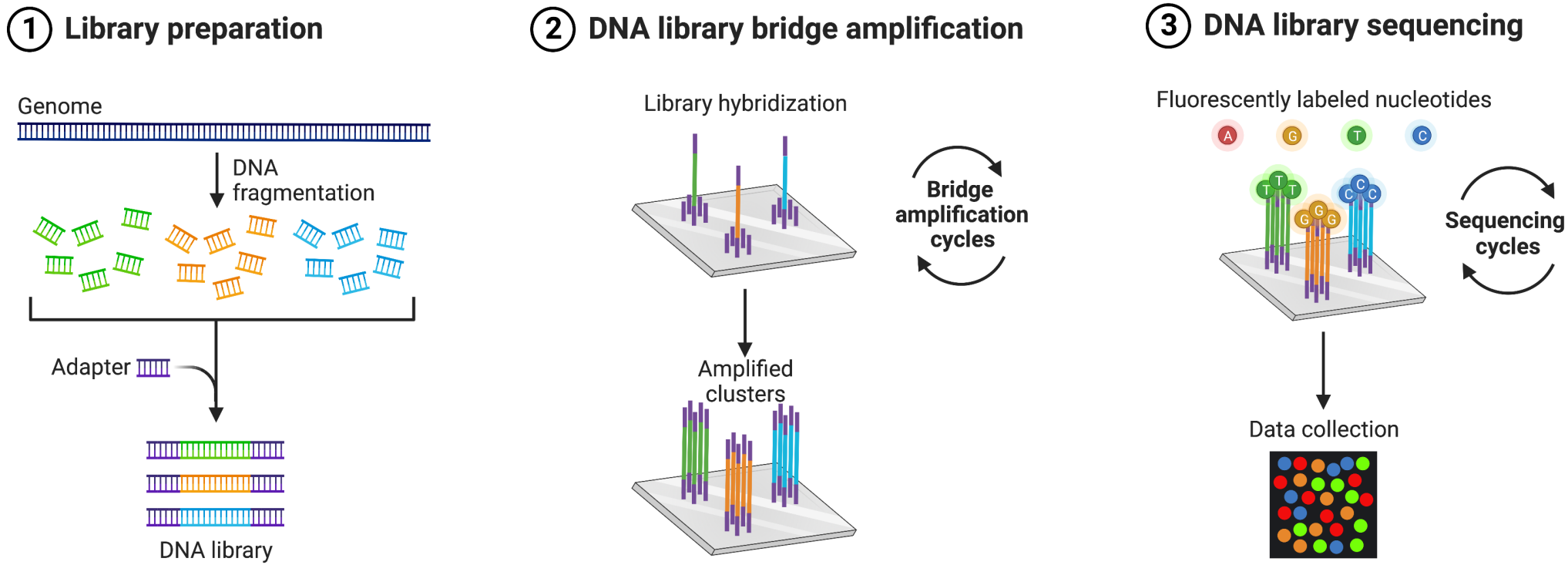






DNA Methylation, Histone Modification, Nucleosome Positioning, Transcription Factors, Chromatin Conformation

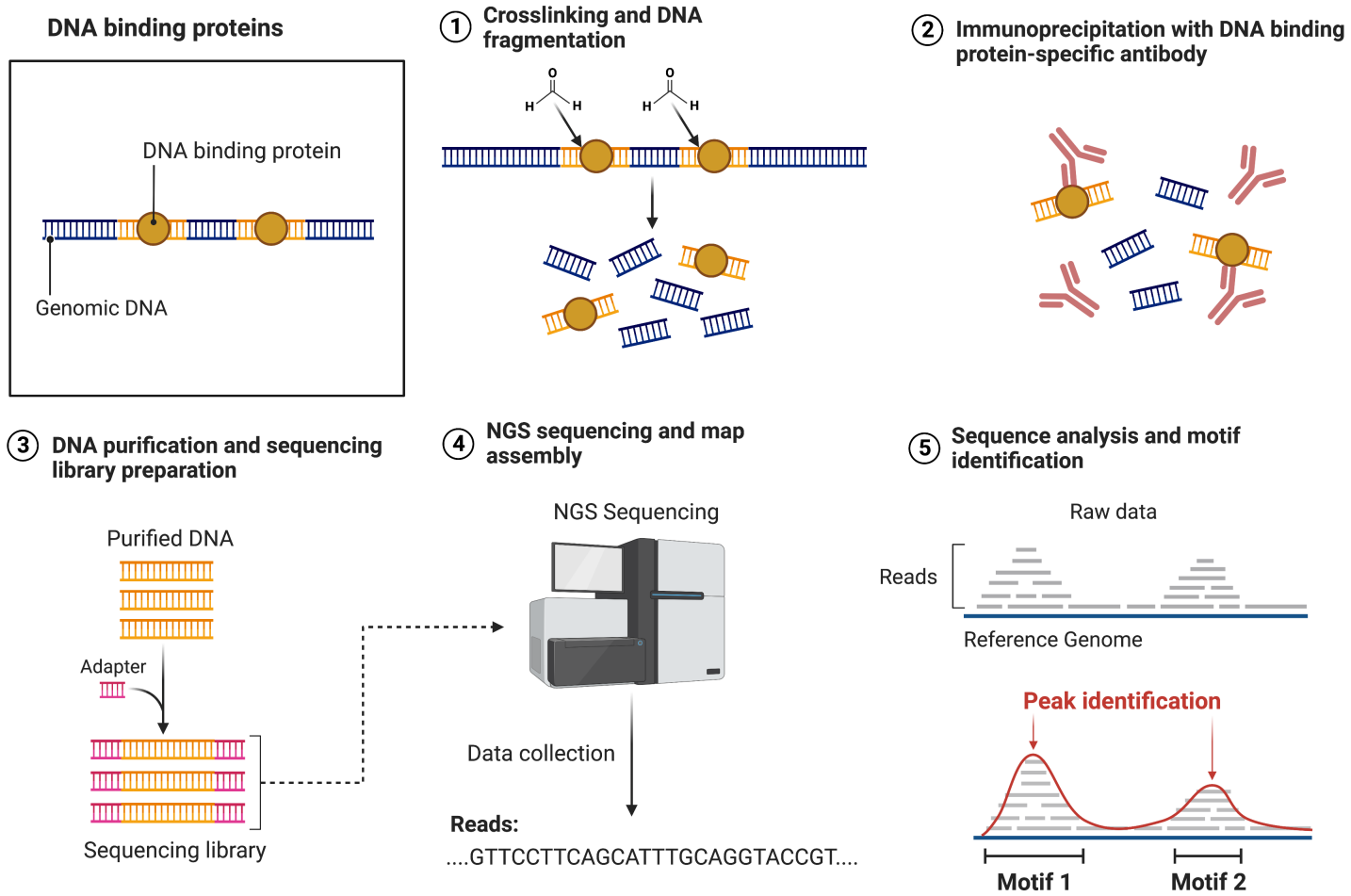
DNA sequencing methods for profiling epigenomic features - Illumina



Other players - Element, Singular, Ultima, **PacBio**, **Oxford Nanopore**

Chromatin Immunoprecipitation – ChIP-seq

Goal: Identify regions of the genome occupied by a protein of interest

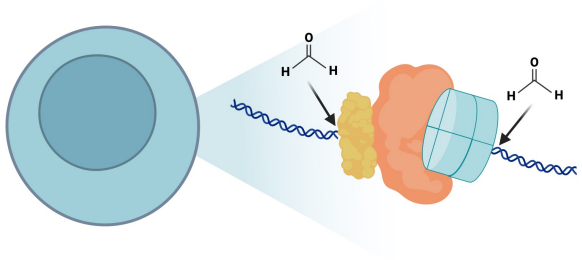


Targets:

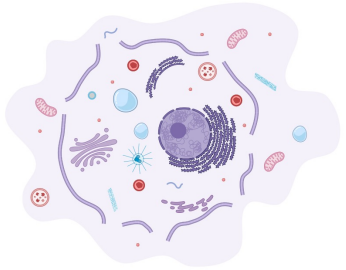
- Histone isoforms
- Histone PTMs
- Transcription factors
- Chromatin modifiers
- Architectural proteins

ChIP-seq Methodology: Chromatin Preparation and Immunoprecipitation

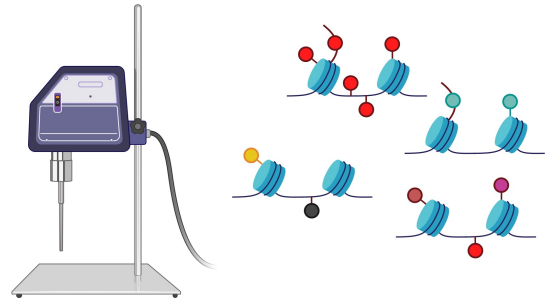
1) Cells are minimally crosslinked with 1% formaldehyde, linking chromatin-associated proteins to DNA



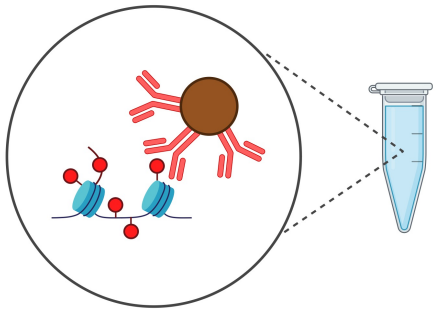
2) Optionally, nuclei are isolated before incubating in SDS-containing lysis buffer for sonication



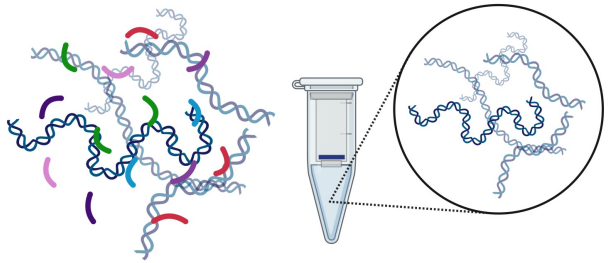
3) Chromatin is solubilized using high-frequency ultrasonication, producing fragments of a desired length (set aside ***Input***)



4) Residual cellular debris is removed and chromatin is pre-cleared before incubation with primary antibody and ProteinA/G beads.

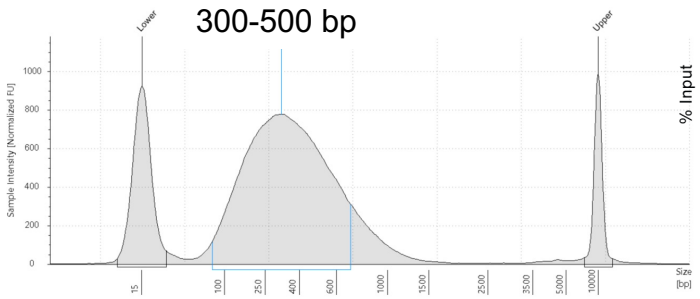


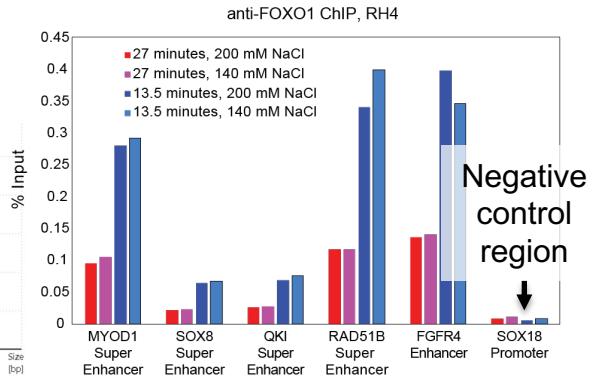
5) Following extensive washing, enriched chromatin is decrosslinked and DNA is recovered for downstream analysis



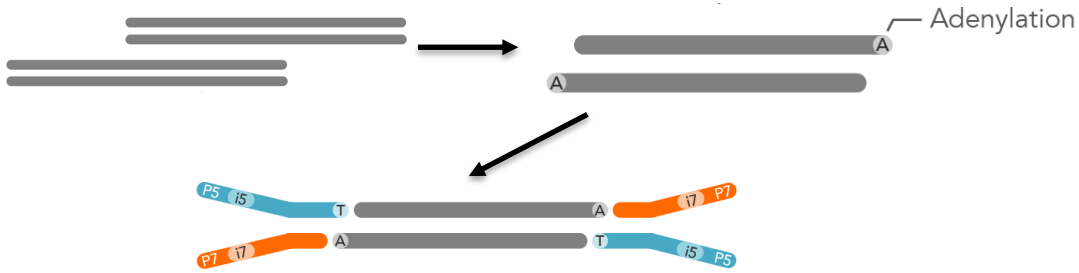
ChIP-seq Methodology: Library Preparation and Sequencing

1) Ideally, QC is performed to assess fragmentation and ChIP efficiency prior to library preparation





2) Sonicated DNA ends are repaired/blunted before A-tailing and adapter ligation. No further fragmentation is required

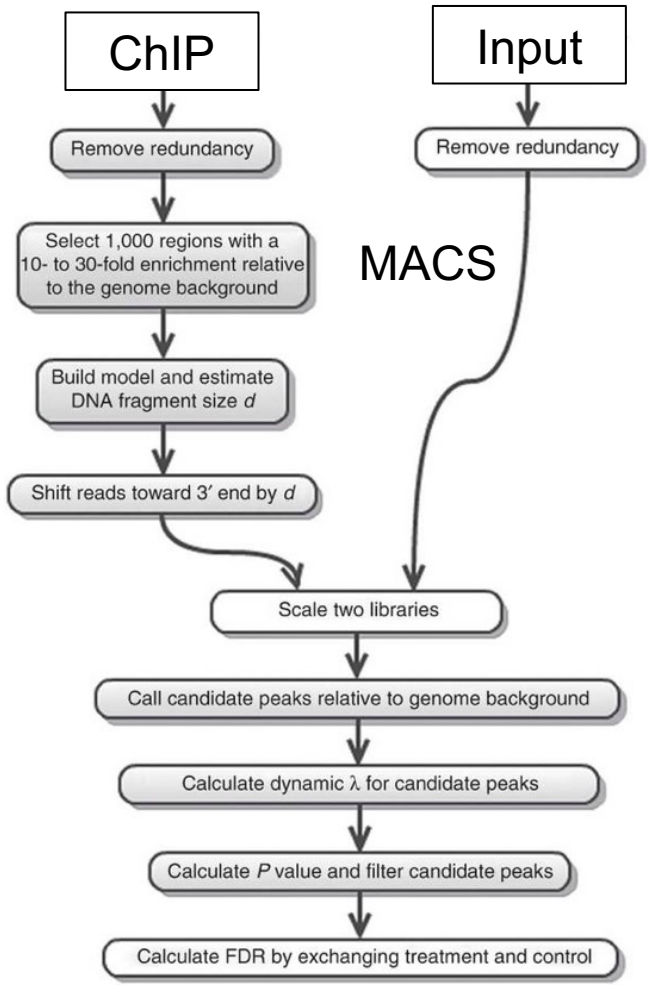


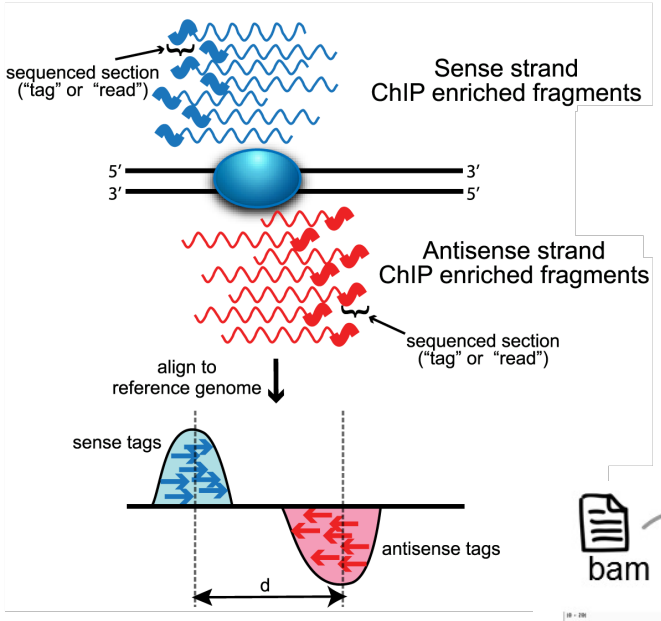
3) Minimal PCR amplification and size-selection are performed to enhance downstream data quality

4) Short-read sequencing is performed, with read depth determined by ChIP target

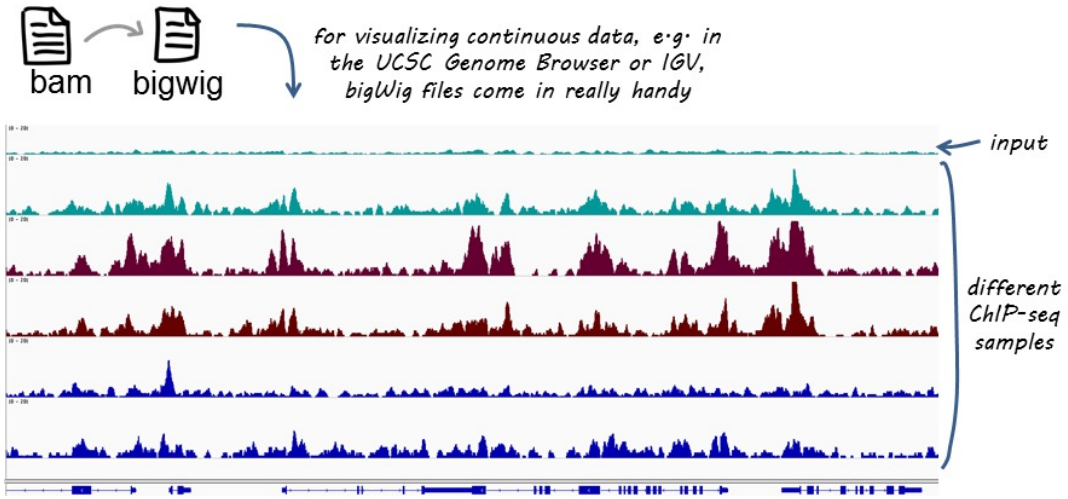
"Broad" histone – 45 million reads
"Narrow" histone – 20 million reads
Transcription factors – 20 million reads

ChIP-seq Methodology: Data Analysis (after QC, alignment, remove dups)





- Basic analysis will produce a set of significant peak locations (.bed file) and a genome-wide signal file (.bigwig file)
- Both operations ideally utilize the **input** library as background

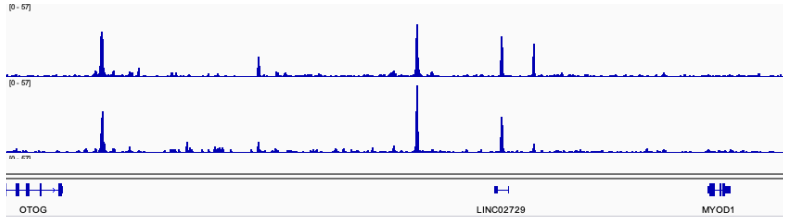


https://hbctraining.github.io/Intro-to-ChIPseq-flipped/lessons/06_peak_calling_macs.html

https://hbctraining.github.io/Intro-to-ChIPseq/lessons/10_data_visualization.html

ChIP-seq Methodology: Downstream Analysis and Interpretation

Single Locus

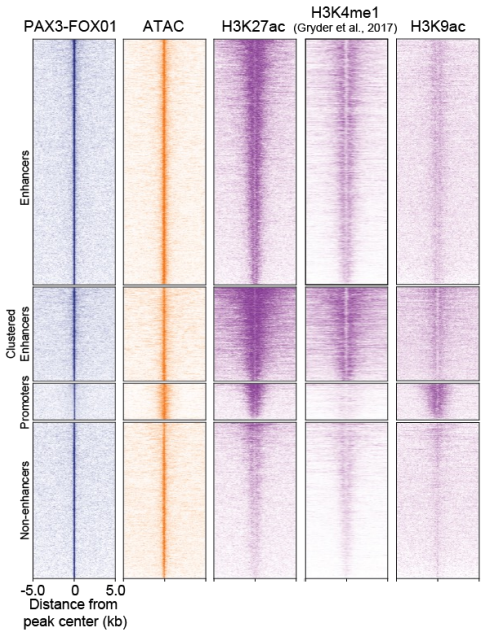


UCSC Genome Browser or
Integrative Genomic Viewer

Identify enriched DNA sequence motifs
to find potential co-binders

Rank	Motif	P-value	log P-pvalue	% of Targets	% of Background	STD(Bg STD)	Best Match/Details	Motif File
1		1e-44	-1.026e+02	41.36%	35.97%	55.1bp (60.1bp)	E2F3/MA0469.2/Jaspar(0.807) More Information | Similar Motifs Found	motif file (matrix)
2		1e-41	-9.650e+01	36.19%	31.13%	56.3bp (61.5bp)	ETV5/MA0765.1/Jaspar(0.603) More Information | Similar Motifs Found	motif file (matrix)
3		1e-36	-8.505e+01	29.74%	25.27%	55.6bp (63.2bp)	ZBED1/MA0749.1/Jaspar(0.706) More Information | Similar Motifs Found	motif file (matrix)
4		1e-34	-7.970e+01	26.25%	22.12%	55.8bp (62.5bp)	E2F4(E2F)/K562-E2F4-ChIP-Seq(GSE31477)/Homer(0.654) More Information | Similar Motifs Found	motif file (matrix)
							Sox10(HMG)/SciaticNerve-Sox3-ChIP-	motif

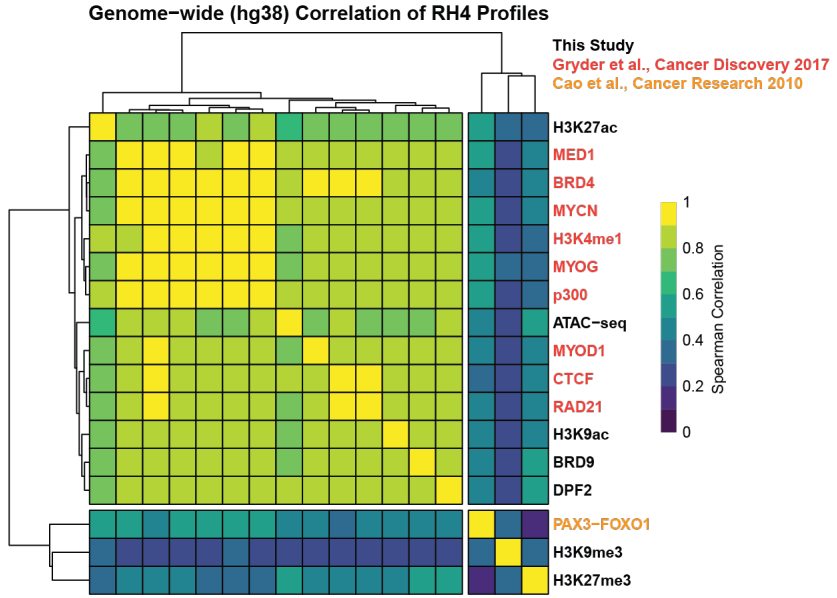
HOMER



Summary figures displaying signal over all peaks, clustered by co-occurrence of additional chromatin factors

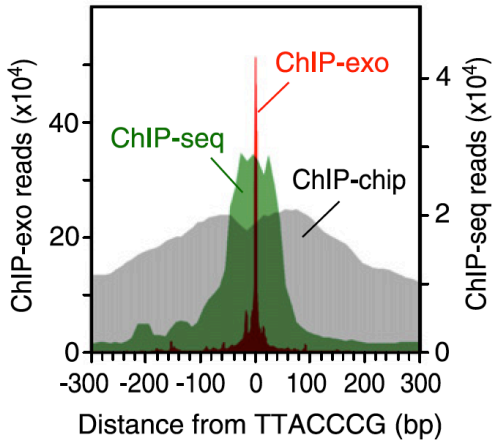
Genome-wide comparisons can point to functional similarity of multiple factors

deepTools



Advances in ChIP-based Methods

1) Improve resolution

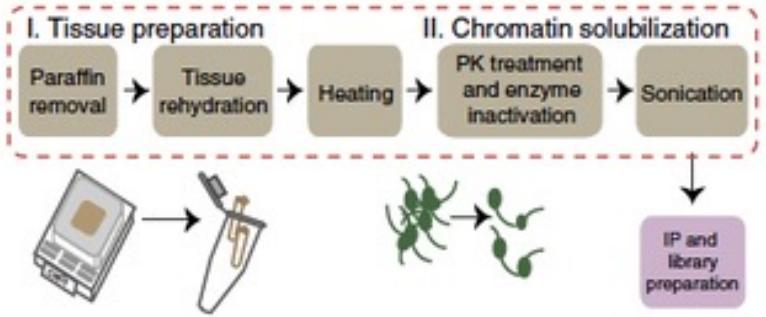


ChIP-exo can produce base pair-resolution, genome-wide footprinting information

2) Reduce cell/tissue input

CUT&RUN and CUT&Tag can reduce input and sequencing costs for profiling of certain stably bound factors

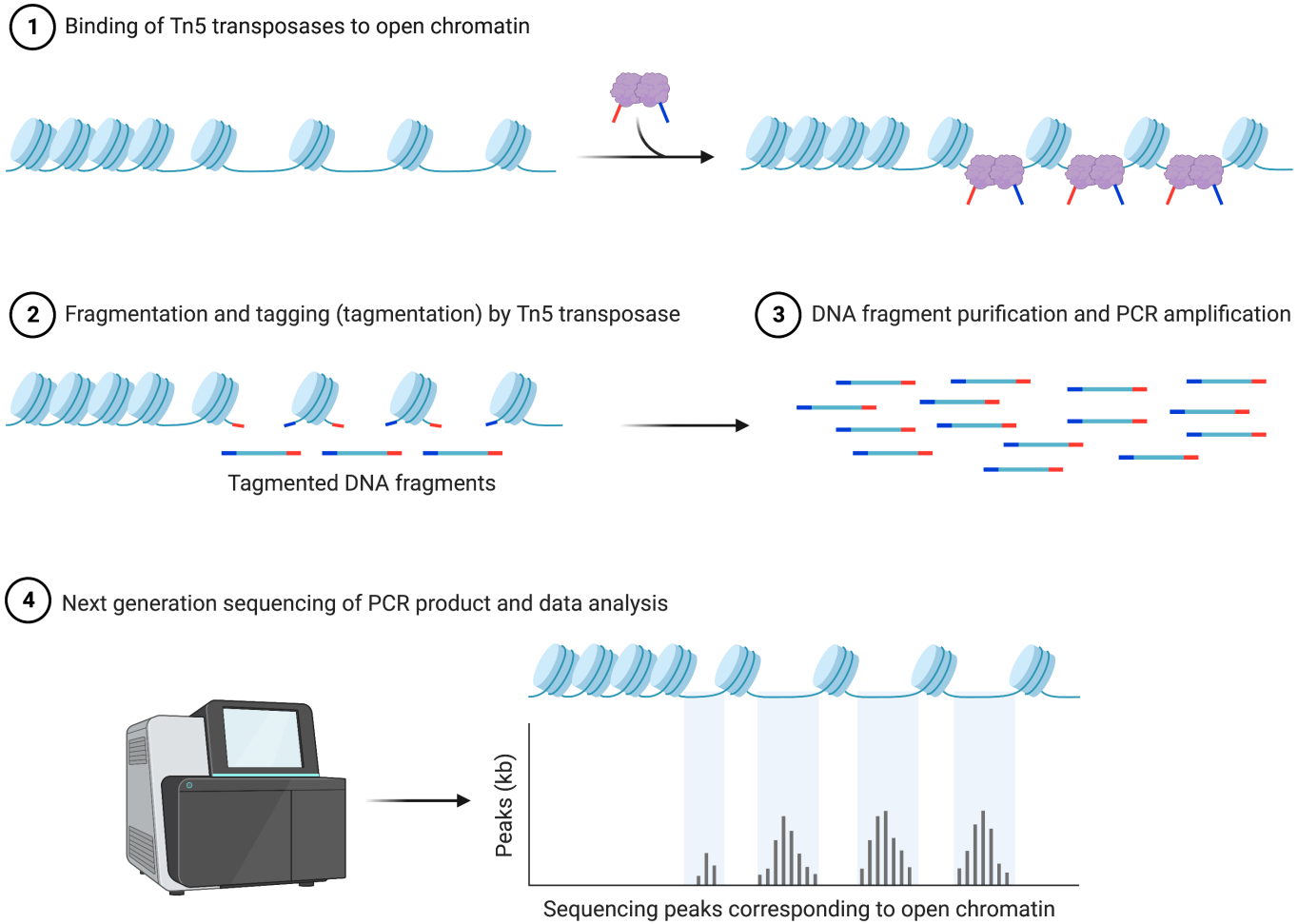
3) Utilize archival tissue



Methods like FiT-seq/FiTAc-seq can standardize the epigenomic analysis of FFPE tissues

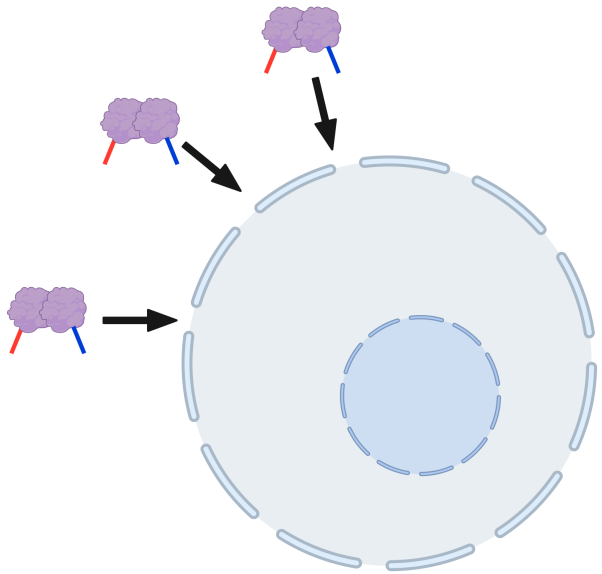
Assay for Transposase-Accessible Chromatin – ATAC-seq

Goal: Identify accessible (active) regions of the genome

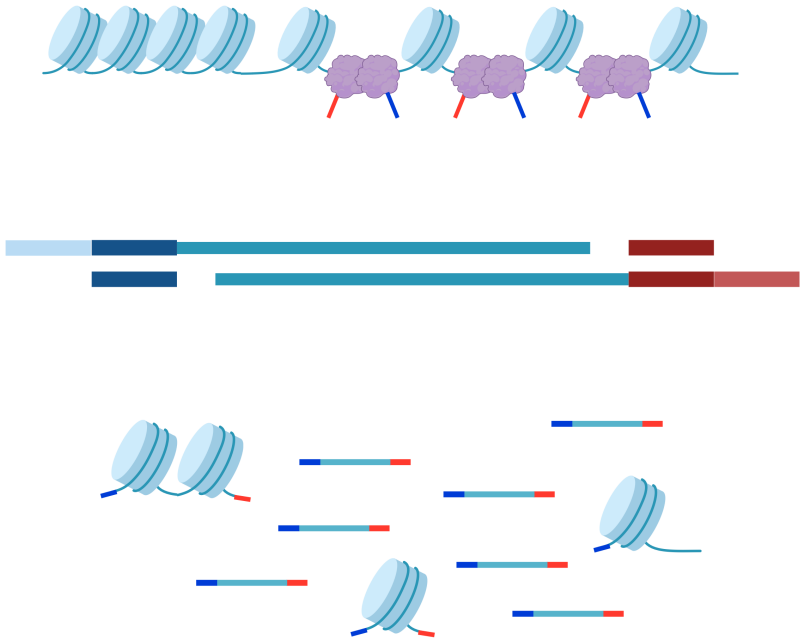


ATAC-seq Methodology: Cell Preparation and Transposition

1) 50,000 Fresh cells are gently permeabilized on ice

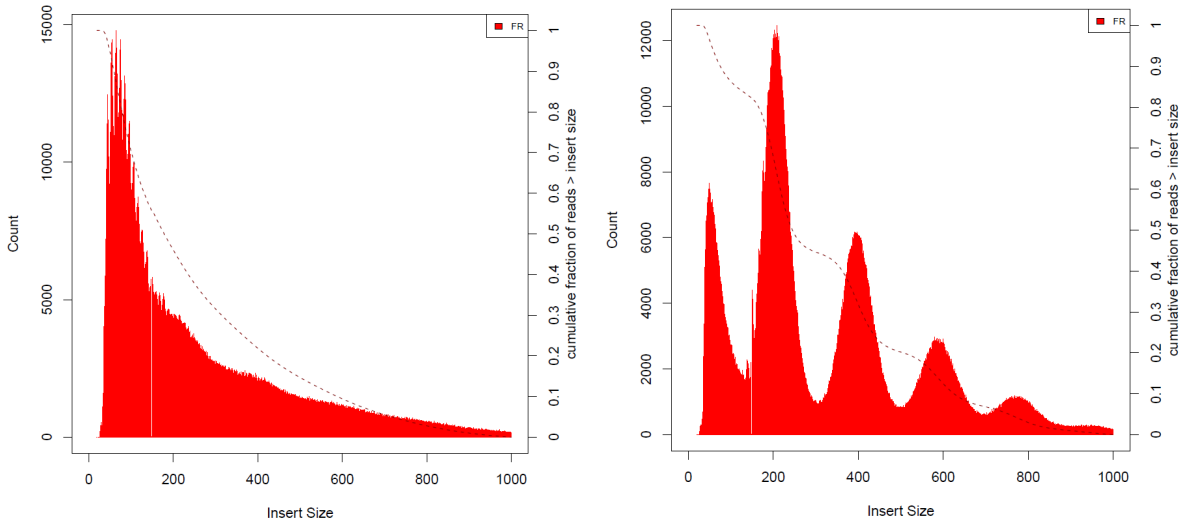


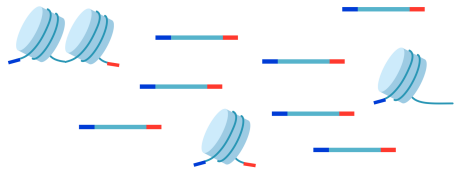
2) Add Tn5 transposase pre-loaded with adapter, and incubate at 37ºC for 30 minutes.



ATAC-seq Methodology: Library Amplification and Sequencing

1) Transposed DNA is purified and minimally amplified by PCR without further size-selection





2) Short-read sequencing to appropriate depth

25 million unique, non-mitochondrial fragments for standard analysis

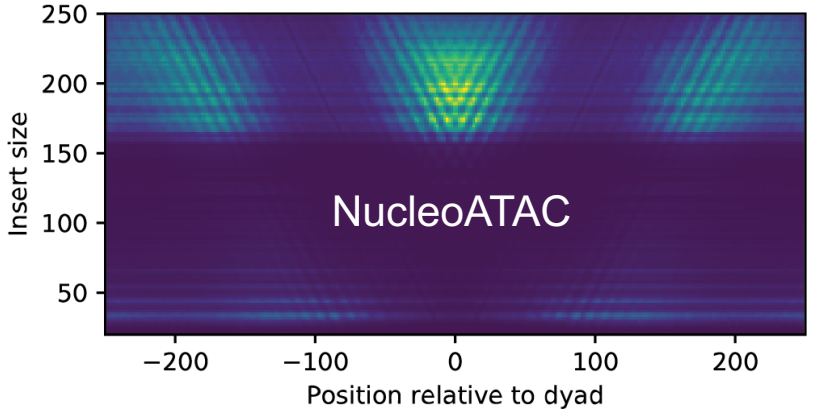
- 25 million single-end
- 50 million paired-end

More reads required for additional, advanced analysis

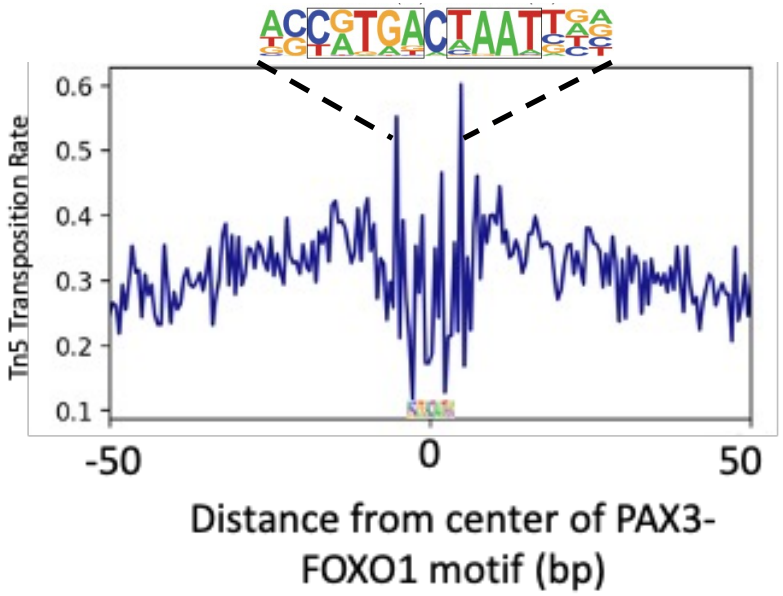
ATAC-seq Methodology: Data Analysis

Basic ATAC-seq data analysis is very similar to ChIP-seq
- Different read shifting
- No input – MACS doesn't mind
- Remove mitochondrial reads
- Additional QC (e.g., insert size distribution)

Inferring Nucleosome Positioning

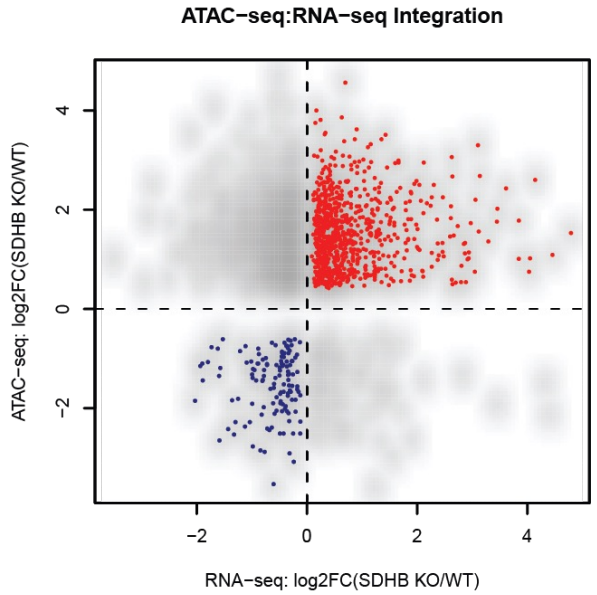


Transcription Factor Footprinting



DiffBind for differential accessibility analysis

Repurposes DESeq2, treating peaks as features akin to genes in RNA-seq



Advances in ATAC-seq

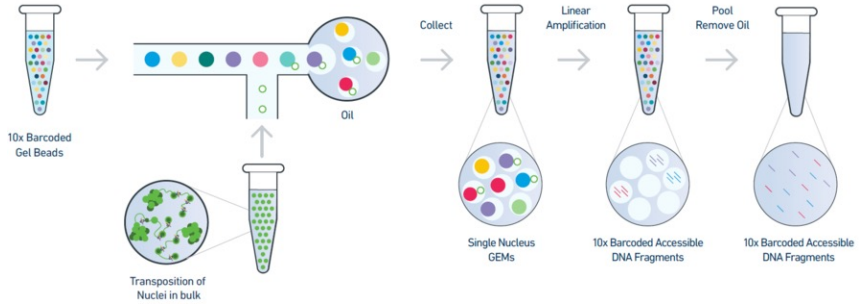
1) Improve mitochondrial contamination

2) Repurpose Tn5

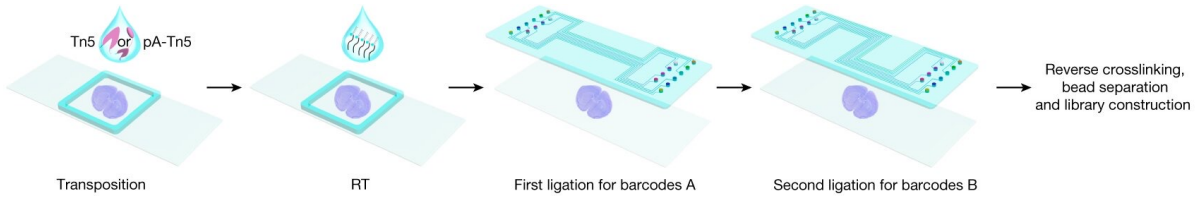
3) Improve methods for archival tissue

4) Reduce input requirements

pA-Tn5 fusion for CUT&Tag



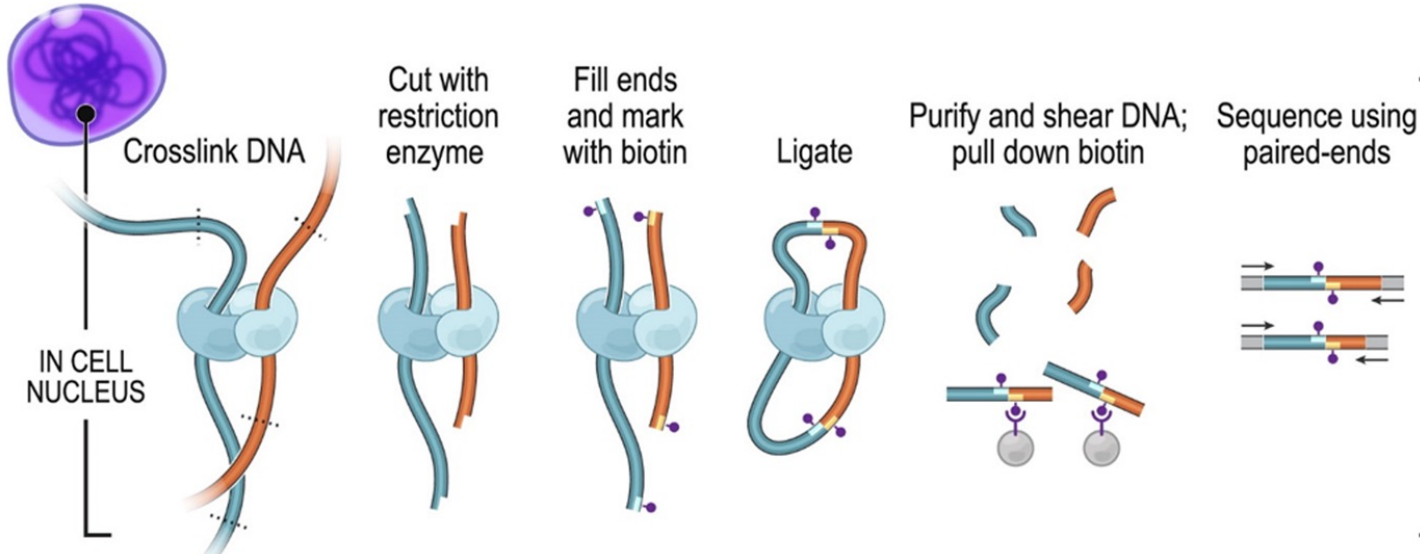
10X Genomic scATAC-seq

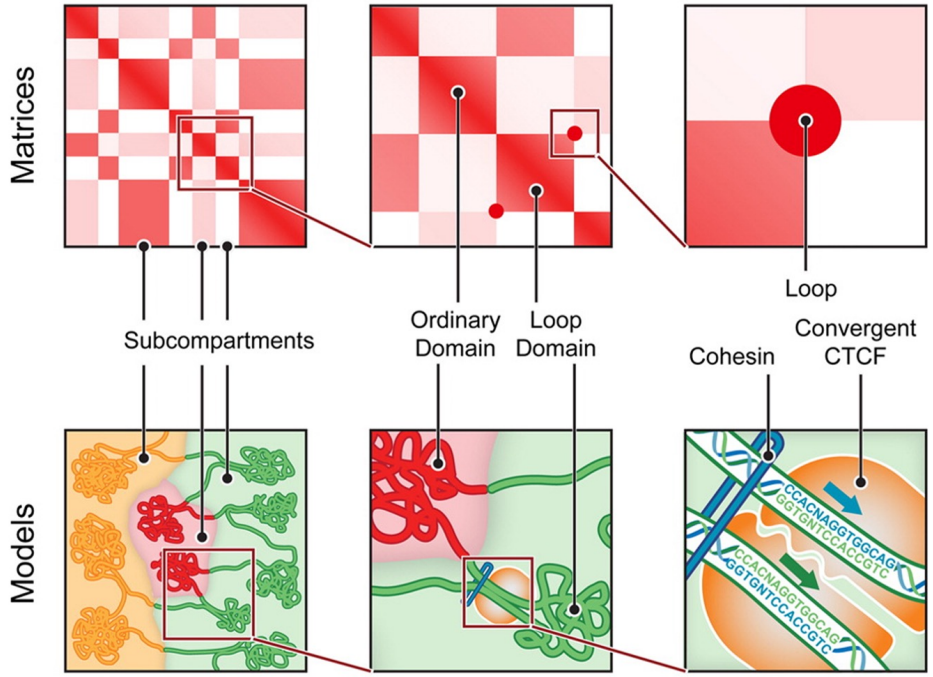


Microfluidics-enabled spatial ATAC-seq
(Zhang et al., Nature 2023)

Chromosome Conformation Capture - HiC

Goal: Identify all pair-wise interactions between genomic loci

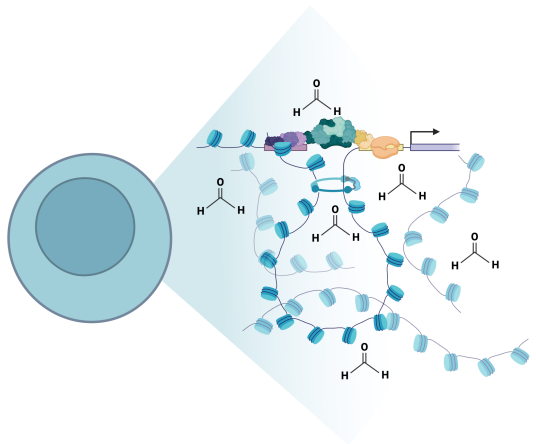




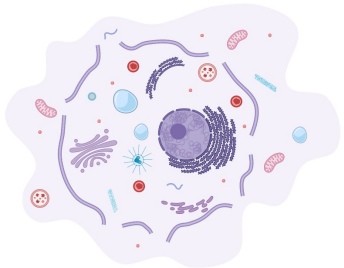
Gagnon, *Microscopy Today* 2020

HiC Methodology: Chromatin Fixation, Digestion, Ligation

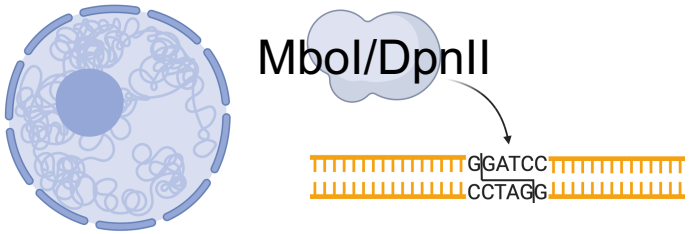
1) Cells are crosslinked with 1% formaldehyde



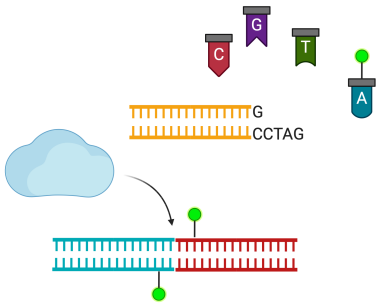
2) Nuclei are isolated from fixed cells and suspended in 0.5% SDS solution to permeabilize



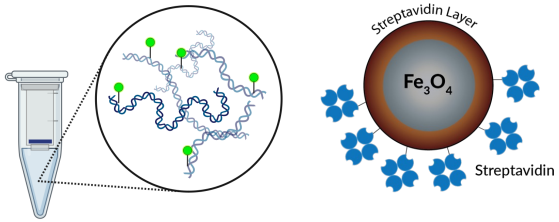
3) Permeabilized nuclei are suspended in restriction enzyme buffer for chromatin digestion

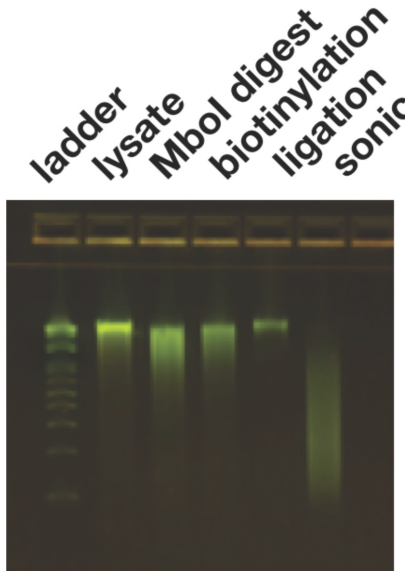


4) Resulting DNA overhangs are filled in with biotinylated dNTPs, and blunt end ligation joins DNA fragments that are close to one another in 3D space



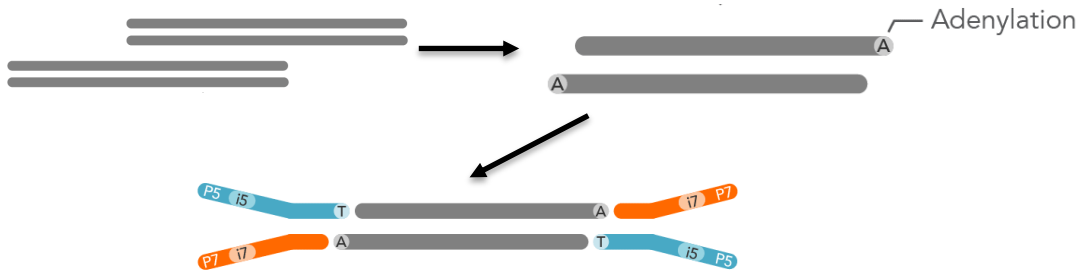
5) Chromatin is decrosslinked, DNA is fragmented by sonication, and biotinylated fragments are captured on streptavidin beads

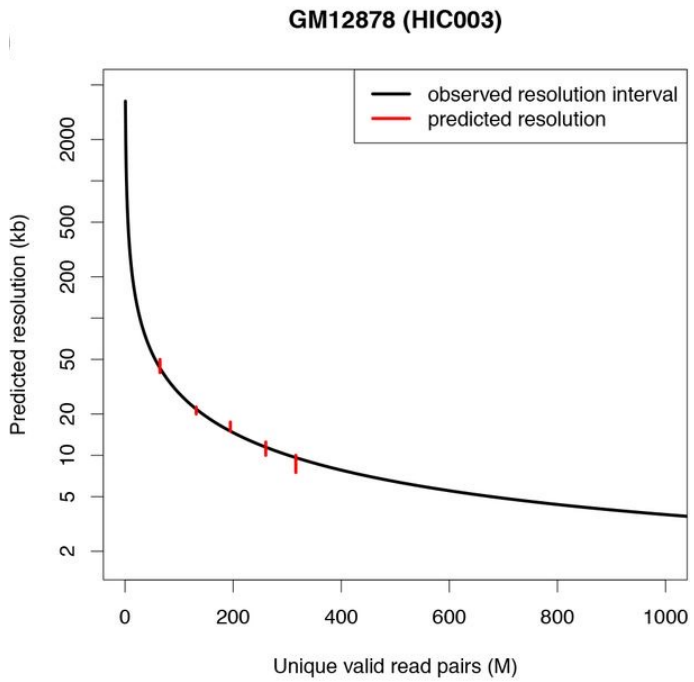




HiC Methodology: Library Preparation and Sequencing

On-bead library preparation is performed on captured, biotinylated DNA





Marchal et al., *bioRxiv* 2020

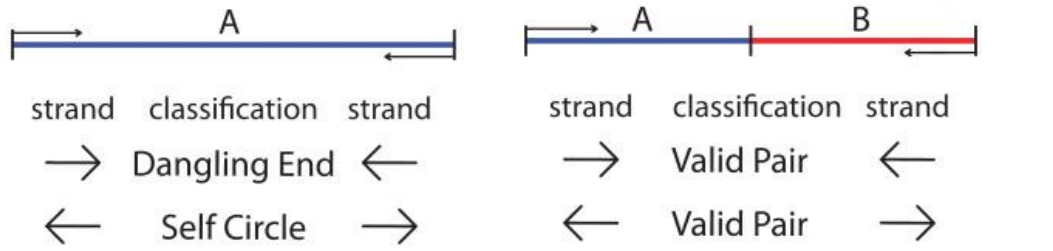
Short-read sequencing is performed to the desired depth, depending on analysis goals

Resolution can be defined as the smallest division of the genome (i.e., bin size) for which 80% of bins contain 1000 reads

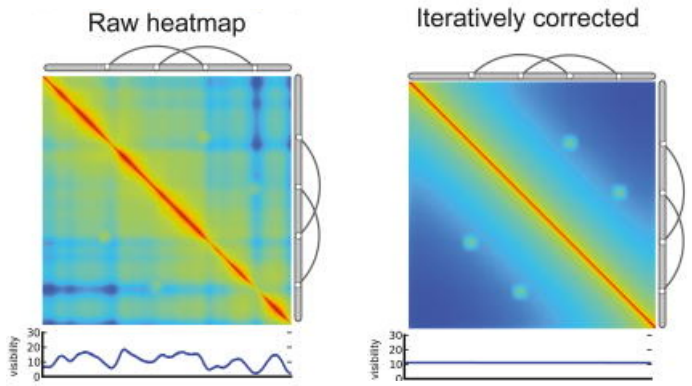
HiC Methodology: Data Analysis

1) Read mapping – performed on individual reads
2) Fragment assignment – To which restriction fragment does the read align?
3) Fragment filtering
   - PCR Duplicates
   - Same fragment
4) Binning
   - Smooths the interaction matrix
5) Balancing
   - Corrects biases to achieve equal

   "visibility" of all genomic loci

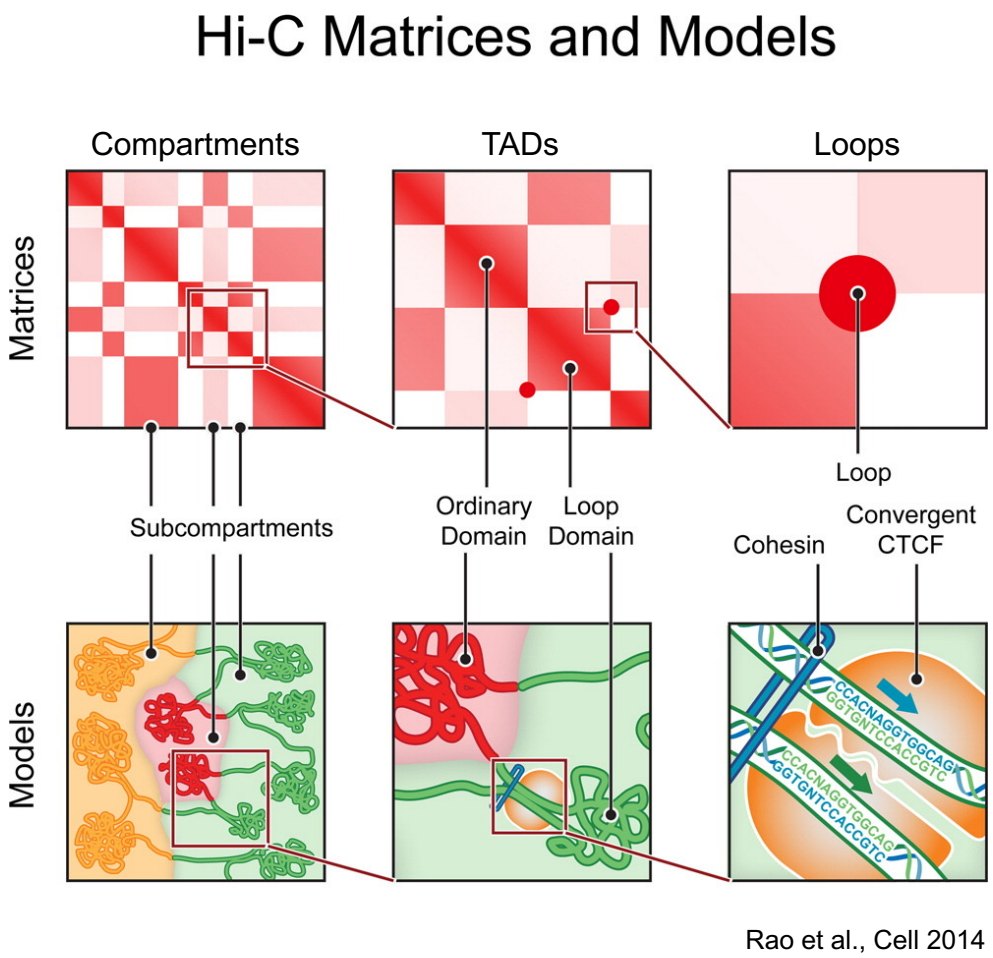


Lajoie et al., *Methods* 2015



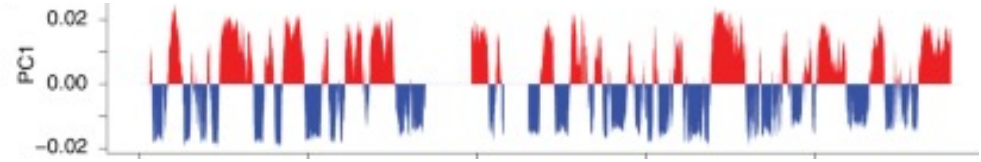
Imakaev et al., Nat *Methods* 2012

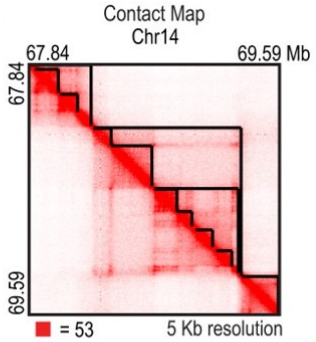
HiC Methodology: Downstream Analysis and Interpretation



Rao et al., Cell 2014

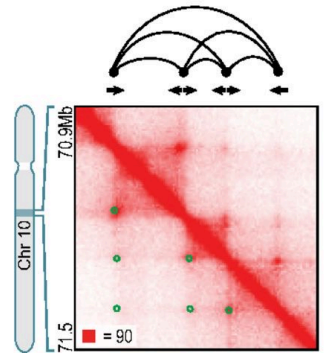
Compartments defined by first principal component of the interaction matrix



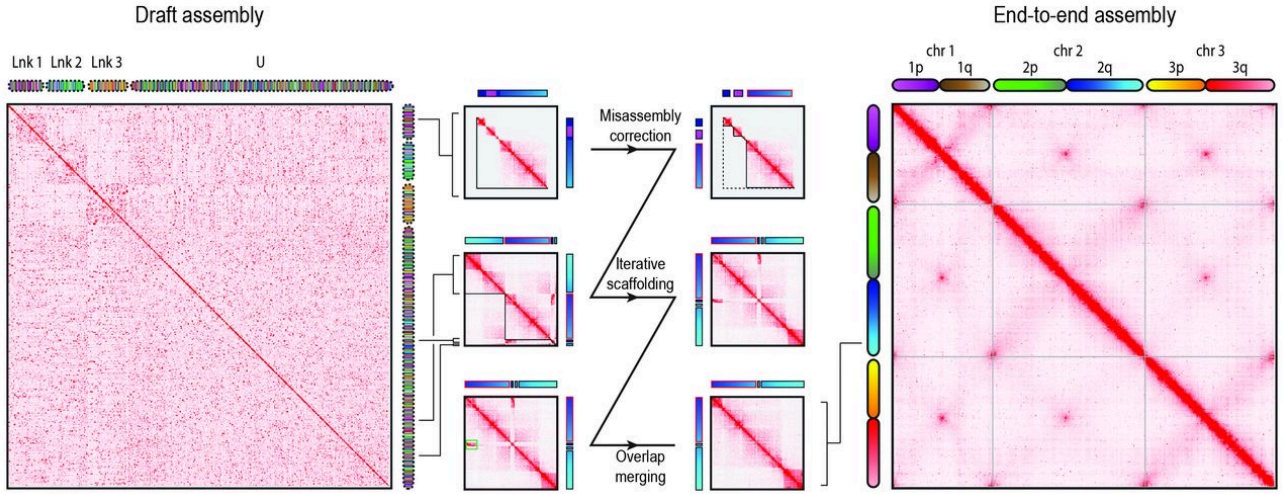


Define TADs by up-stream vs. down-stream sequence bias

Define ”loops” as higher than expected, off-diagonal signals



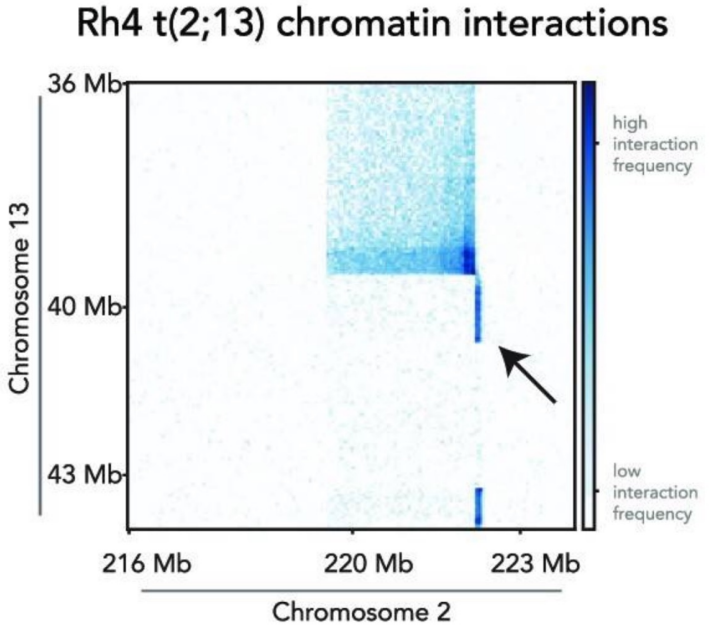
Interesting HiC Applications



Dubchenko et al., Science 2017

Correcting draft genome assemblies by identifying mis-joins

dnazoo.org

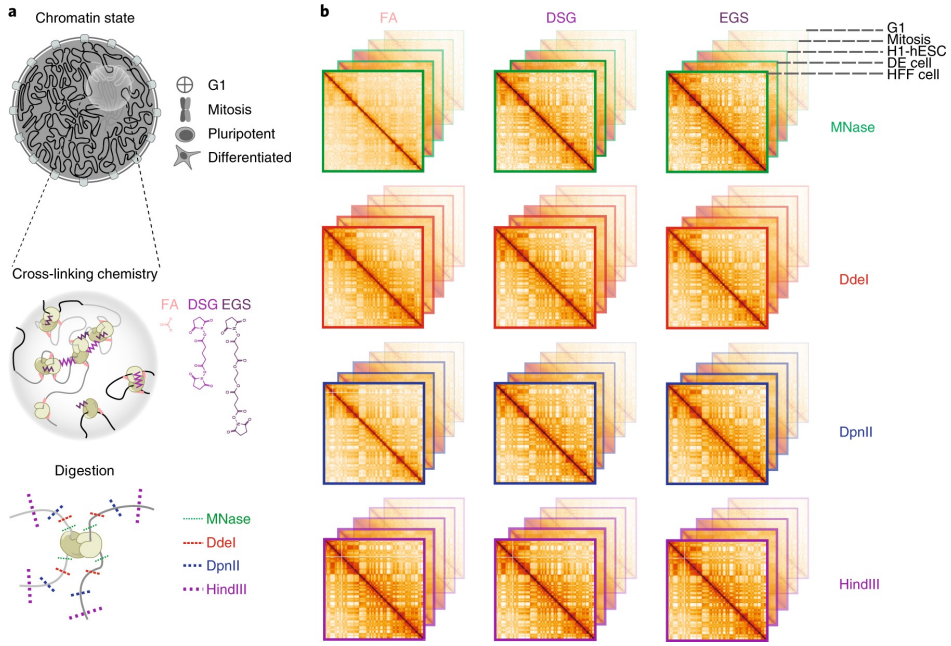


Wang et al., NAR Cancer 2023

Identifying novel translocations

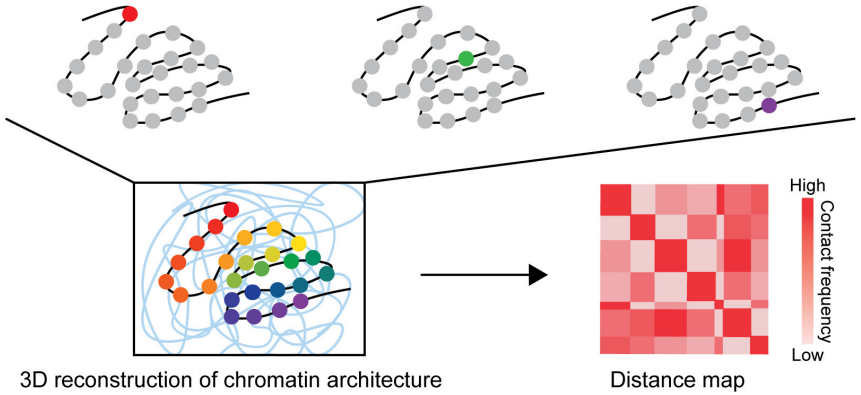
Advances in HiC Approaches

1) Enhanced resolution
2) Reducing material input
3) Tracking dynamics
4) Multi-way contacts

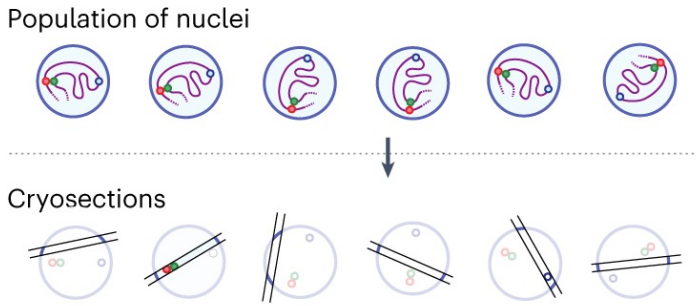


Oksuz et al., Nature Methods 2021

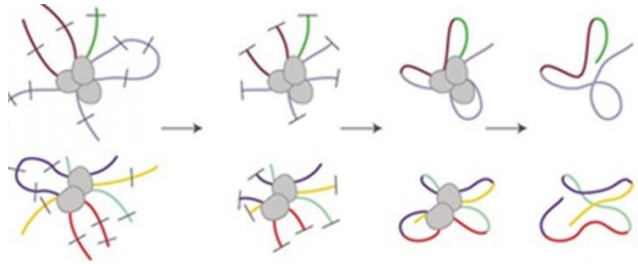
ORCA



GAM



Pore-C



Some tools to get you started

ChIP-seq

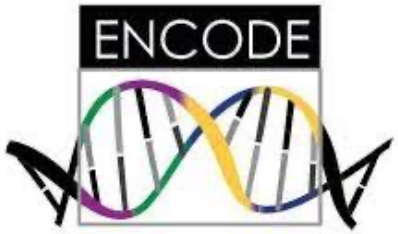
- https://github.com/ENCODE-DCC/chip-seq-pipeline2
- https://deeptools.readthedocs.io/en/develop/
- http://homer.ucsd.edu/homer/motif/

ATAC-seq

- https://github.com/ENCODE-DCC/atac-seq-pipeline
- https://github.com/GreenleafLab/NucleoATAC

HiC

- https://github.com/ENCODE-DCC/hic-pipeline
- https://github.com/aidenlab/juicer



Galaxy PROJECT

Cistrome Data Browser