Epigenomics for everyone!

Amelia Weber Hall, PhD

Ellinor Lab

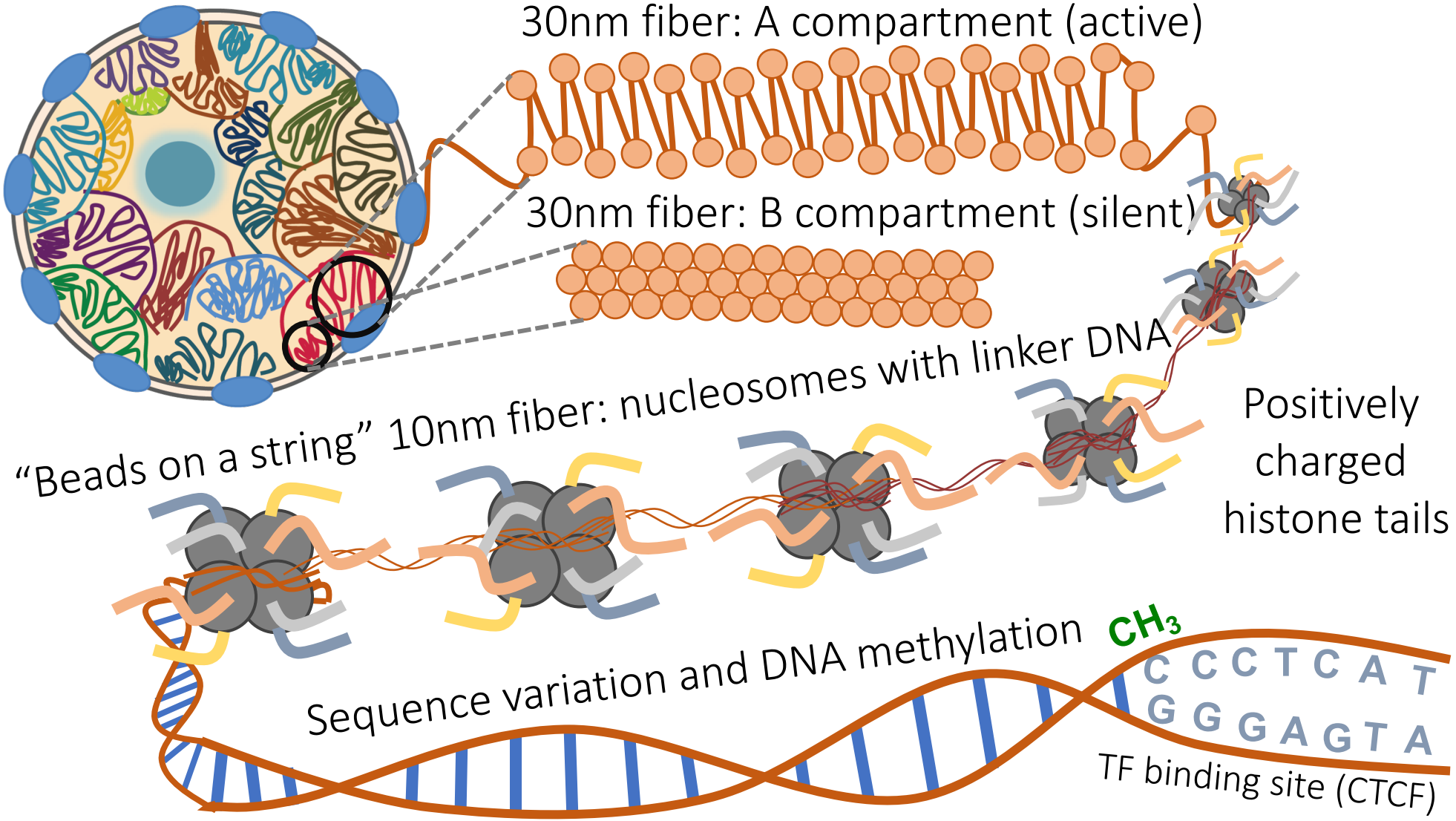
AHA SFRN Webinar

April 2020

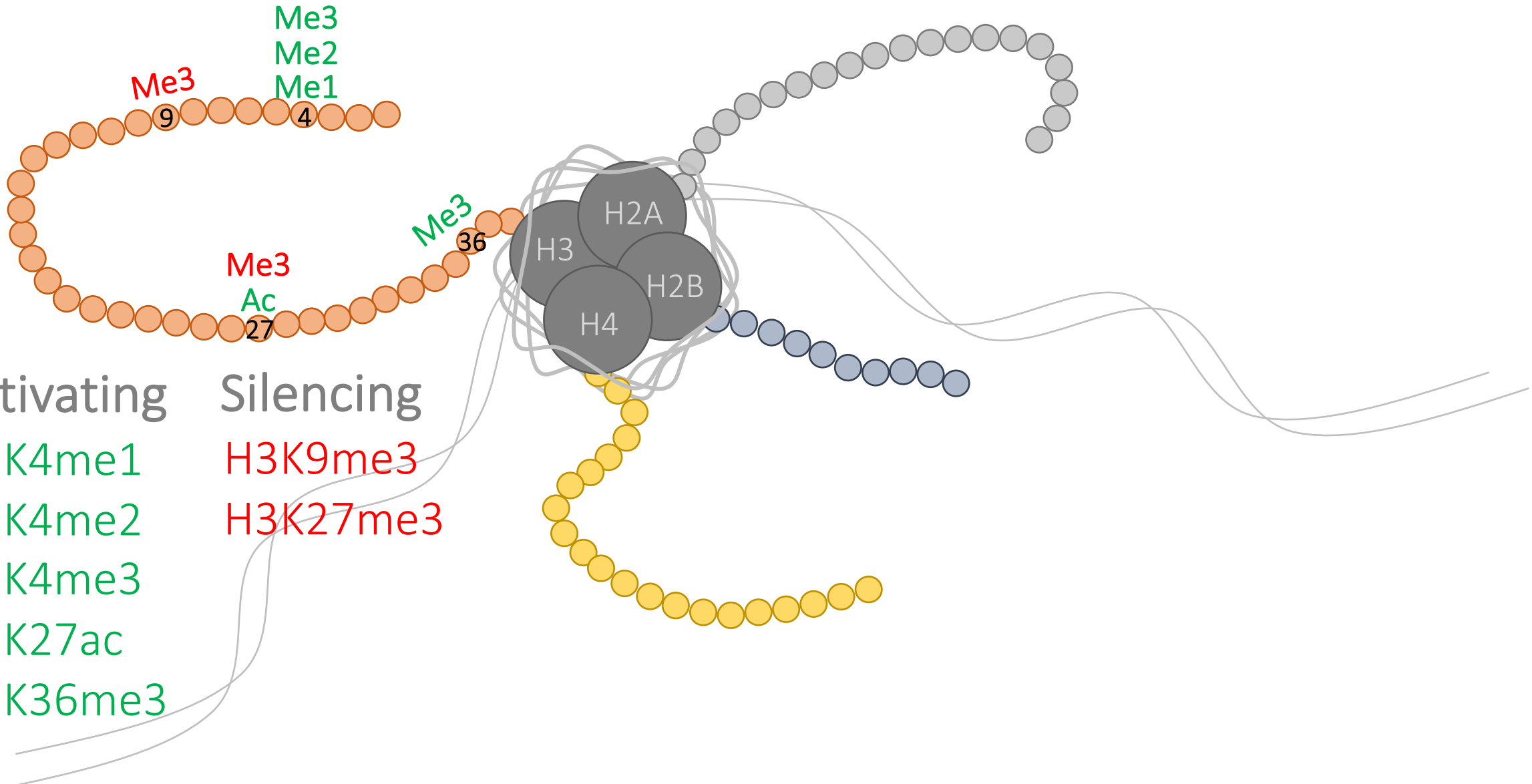
Outline

1. Architecture of DNA in the nucleus
2. History of epigenomics
3. Techniques for studying the epigenome
4. Analysis methods for the epigenome
5. Epigenomics and clinical treatments/trials

There are 2 meters of DNA in every nucleated cell in the human body



Histones provide structure for DNA in the nucleus



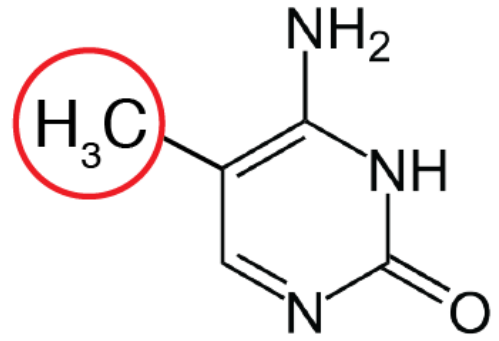
Activating	Silencing
H3K4me1	H3K9me3
H3K4me2	H3K27me3
H3K4me3	
H3K27ac	
H3K36me3	

Chromatin regulatory motifs and factors

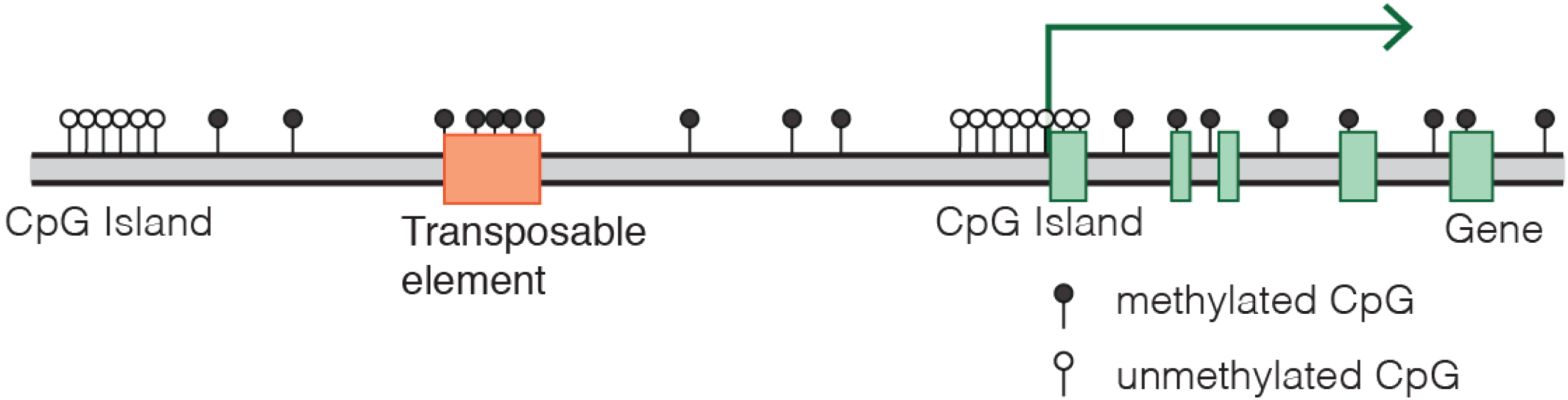
Histone PTMs identify chromatin regulatory motifs		Histone PTM co-localization
H3K4me3	Active promoters	K4me3, K27ac, K9ac, (CTCF)
H3K4me2	Cell type specific TF loci	
H3K4me1	Active enhancers (with H3K27ac co-localization)	H3k4me1 alone is inactive
H3K36me3	Active transcription (gene bodies)	With H3K79Me2
H3K27ac	Active promoters/enhancers	Methylation of H3K4 is differential
H3K9me3	Heterochromatin/inactive	Large multi-Mb tracts, centromeres
H3K27me3	Inactive/polycomb silenced promoters	Smaller tracts, bivalency (with K4me3)
ATAC-seq	Chromatin accessibility/Active state	All accessible chromatin
CTCF	Insulator binding/transcriptional activation	Alone, alongside strong promoters

DNA methylation: background

- Methyl residues can be added to cytosines directly at the 5 carbon position
- These methylated cytosines are located within “CpG islands” which are often (not always) located near gene promoters or regulatory elements
  - Methylation generally implies silencing when it is highly present over a region

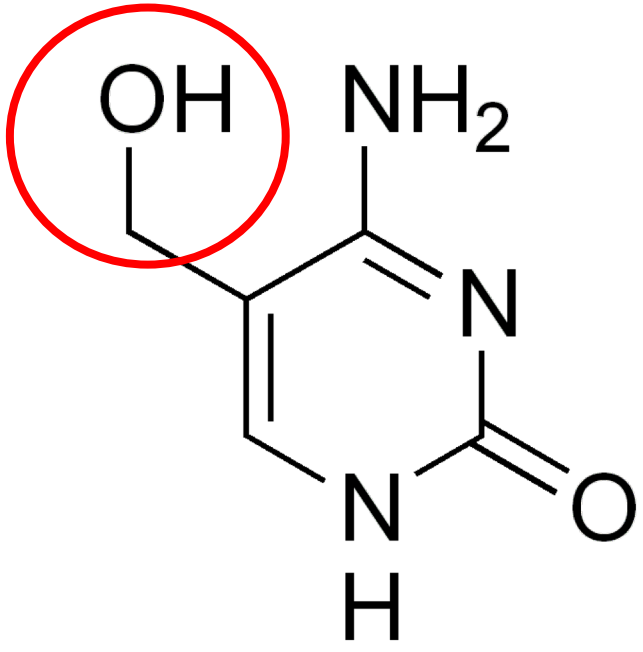


methylation



DNA methylation: 5-hydroxymethyl cytosine

- A form of methylation first detected in the 70s, validated in 2009 by the Heintz lab
- Produced by the enzyme Tet acting on a methylated cytosine
  - Less well studied, present in detectable amounts in mammalian brain, heart, other tissues: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3711458/
- Thought to be involved in activation of cell type specific elements
  - https://pubmed.ncbi.nlm.nih.gov/22829908/



Outline

1. Architecture of DNA in the nucleus
2. History of epigenomics
3. Techniques for studying the epigenome
4. Analysis methods for the epigenome
5. Epigenomics and clinical treatments/trials

History of Epigenomics: DNA methylation

- First discovered in 1948 (Hotchkiss); the original modification of DNA!
  - Identified a modified cytosine in calf thymus
- DNA methylation involved in gene expression and regulation
  - Holliday & Pugh, 1975; Compere & Palmiter, 1981
- Methylated DNA in de-enriched in CpG islands
  - Bird et al., 1985
- Most gene promoters reside within CpG islands
  - Saxonov et al., 2006
- 5-hydroxymethyl cytosine discovered, the "6th base"
  - https://pubmed.ncbi.nlm.nih.gov/19372393/

History of Epigenomics: histone PTMs I

- Histones were described in 1884 by Albrecht Kossel
- In 1950, histones were proposed to act as transcriptional modulators
  - Stedman & Stedman, 1950
- Histone acetylation correlated with gene activation
  - Allfrey & Mirsky, 1964
- ”Open” chromatin without histones bound is rare in eukaryotes
  - Clark & Felsenfeld, 1971; Cedar & Felsenfeld, 1973
- DNA packaged into nucleosomes
  - Kornberg & Thomas, 1974
- In 1984, John Lis develops the Chromatin Immunoprecipitation technique
  - Using antibodies to selectively isolate different proteins bound to DNA

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3941222/

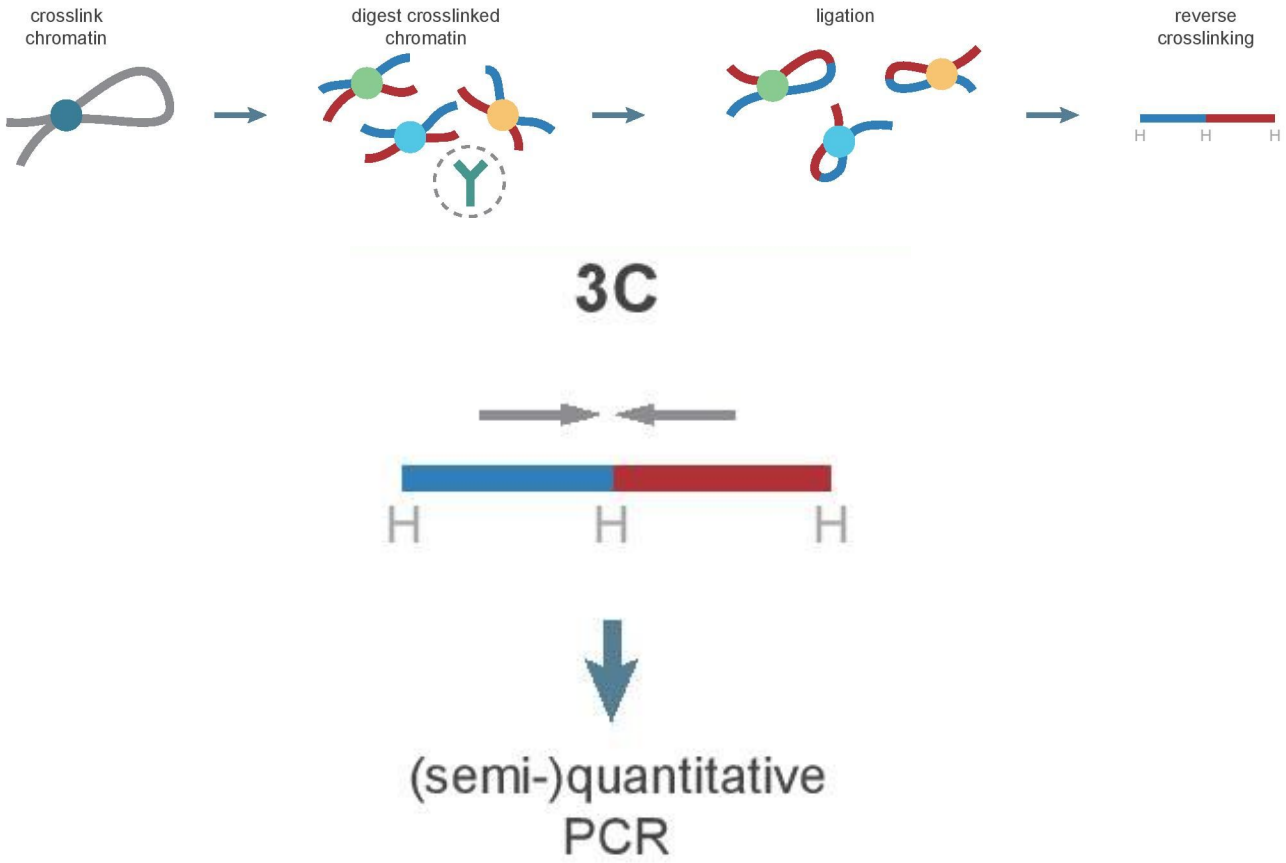
History of Epigenomics: histone PTMs II

- Histone tails necessary for activation/silencing control
  - Wallis et al., 1980; Durrin et al., 1991
- Understanding homology across species with regard to histone modifying enzymes
  - Brownell et al., 1996, yeast/Tetrahymena; histone acetylase
  - Taunton et al., 1996, yeast/human; histone deacetylase
- Nucleosome remodeling complexes
  - Peterson & Herskowitz 1992; Tsukiyama & Wu 1995
- Chromatin IP and sequencing: entry to the NGS era
  - Robertson et al., 2007

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3941222/

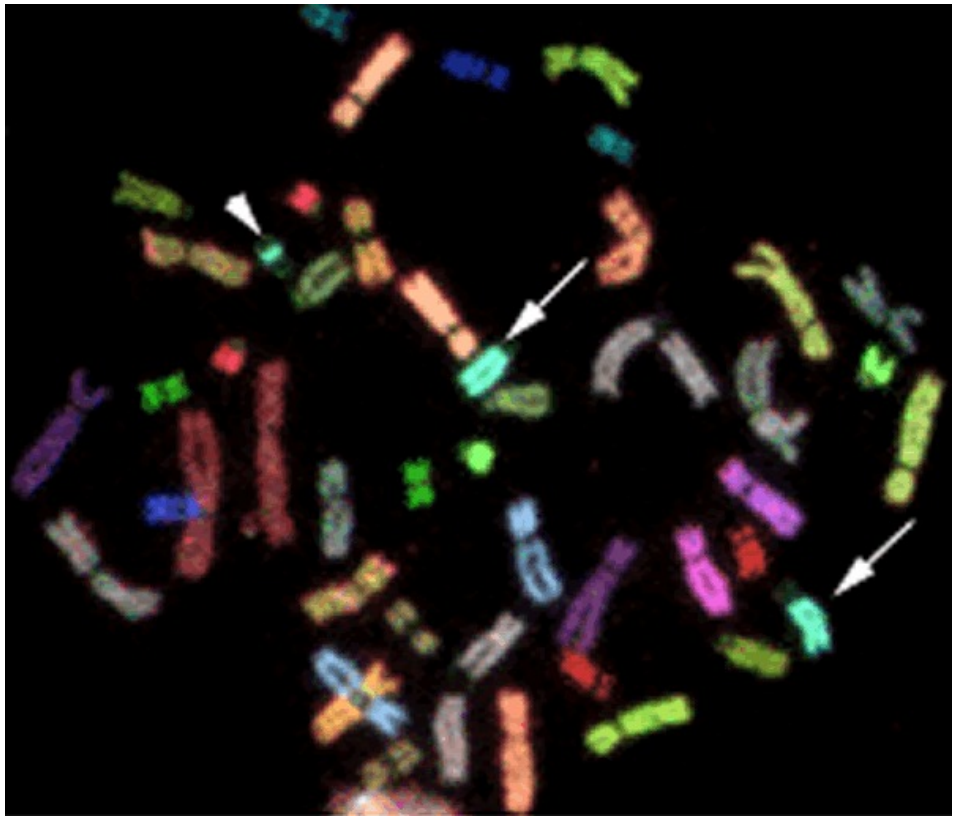
History of Epigenomics: chromatin architecture

Chromosome conformation capture
Job Dekker, 2002



https://bmcgenomics.biomedcentral.com/articles/10.1186/1471-2164-15-S12-S11

Chromosome painting (FISH-based, late 90s)
S. Uhrig et al, 1999



https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1377944/

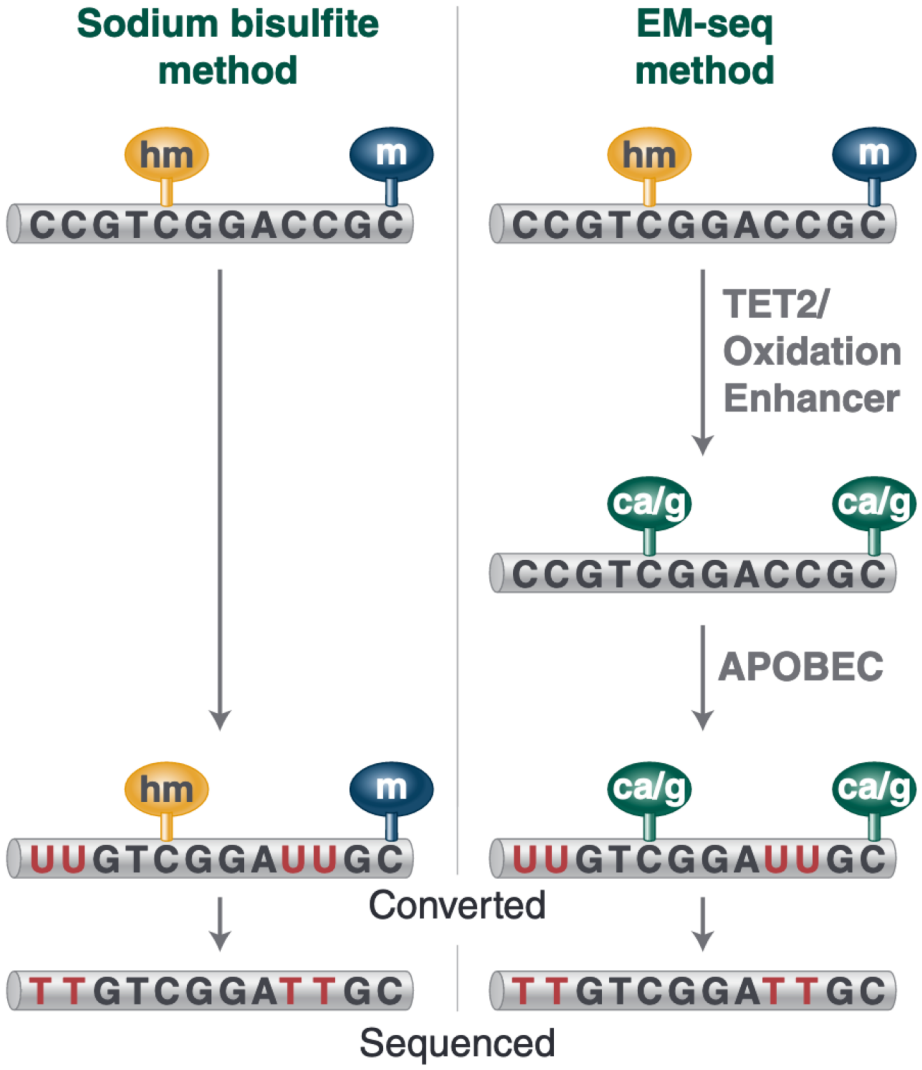
Outline

1. Architecture of DNA in the nucleus
2. History of epigenomics
3. Techniques for studying the epigenome
4. Analysis methods for the epigenome
5. Epigenomics and clinical treatments/trials

Techniques for studying the epigenome

- Direct DNA methylation:
  - Whole Genome Bisulfite Sequencing (WGBS)
  - EM-seq (same as WGBS, but enzymatic)
  - Methyl chips or arrays
- DNA-protein interactions
  - ChIP-seq
  - CUT&RUN
- Chromatin architecture/topology
  - “C” based methods
  - FISH (Fluorescence in-situ hybridization)
    - Used for validation

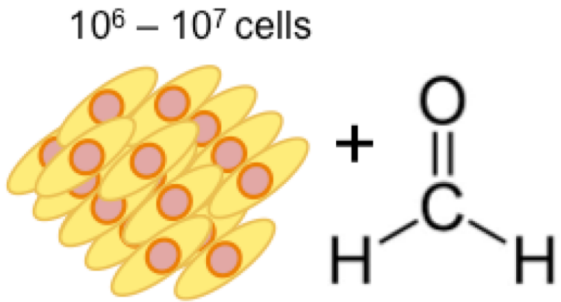
Methyl-seq: WGBS vs EM-seq



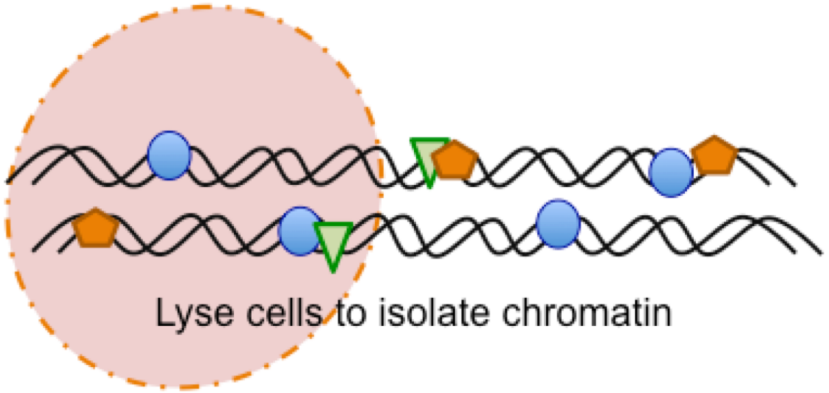
Genomic DNA Shearing	Input is 10–200 ng of genomic DNA, sheared to 300 bp	
End Repair/ dA-Tailing	DNA is end-repaired and dA-tailed	NEBNext® Ultra™ II reagents
EM-seq Adaptor Ligation	DNA is ligated to the EM-seq adaptors	
Oxidation of 5mC and 5hmC	TET2 and Oxidation Enhancer protect 5mC/5hmC from deamination	
Deamination of C to U	APOBEC deaminates cytosines to uracils; oxidized forms of 5mC/ 5hmC are not deaminated	
PCR Amplification	Library amplification using NEBNext Q5U Master Mix and NEBNext index primers	
Sequencing	Sequencing on the Illumina® platform	

https://www.neb.com/products/e7120-nebnext-enzymatic-methyl-seq-kit

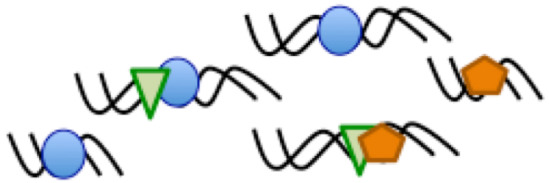
ChIP-seq allows for whole genome profiling of DNA-protein interactions



Cross-link cells with formaldehyde

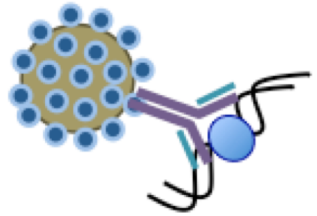


Lyse cells to isolate chromatin



fragment chromatin (sonication)

Bind antibody to protein-DNA complex

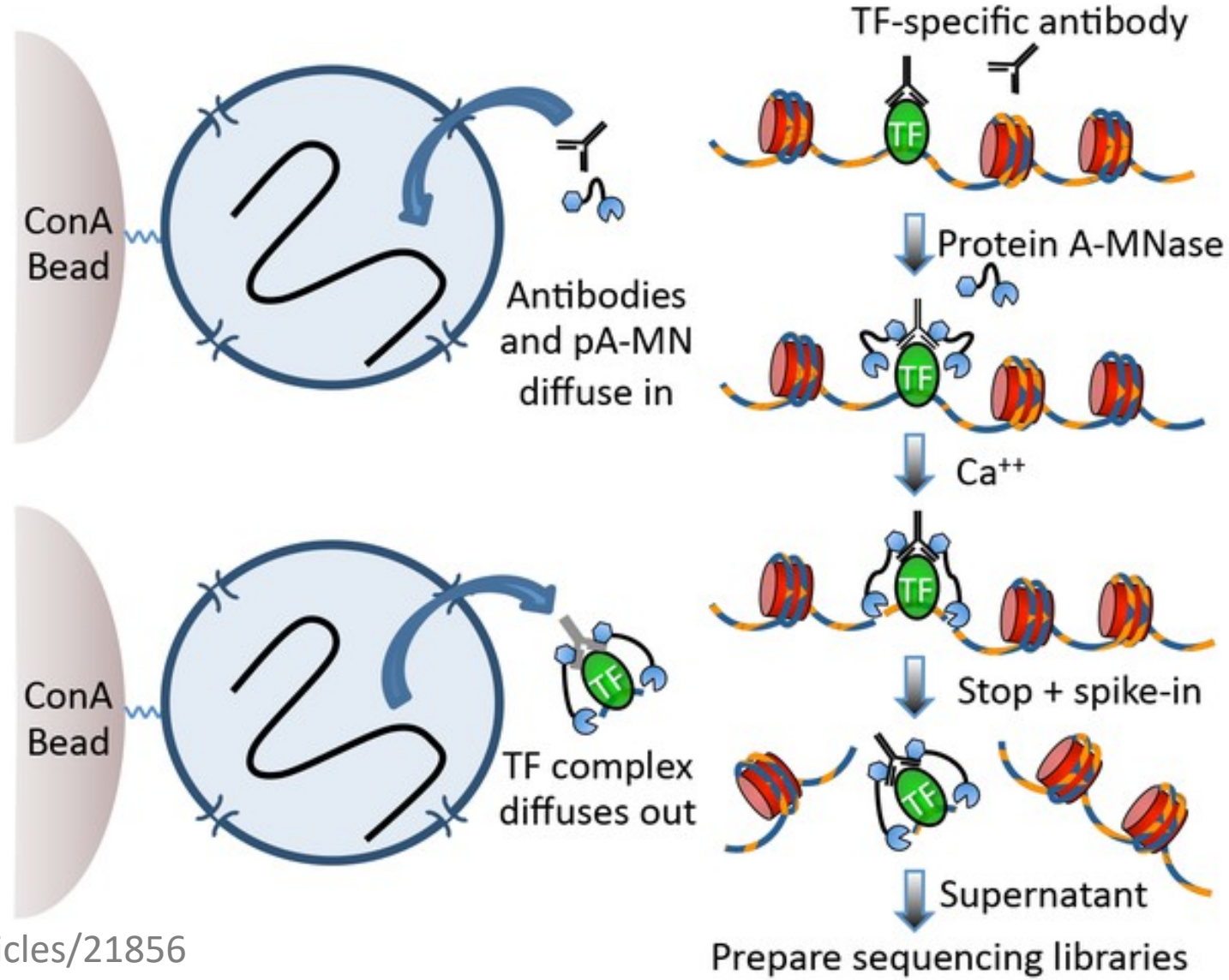


Capture antibody on protein A/G beads

Digest protein, purify DNA

Prepare library, sequence

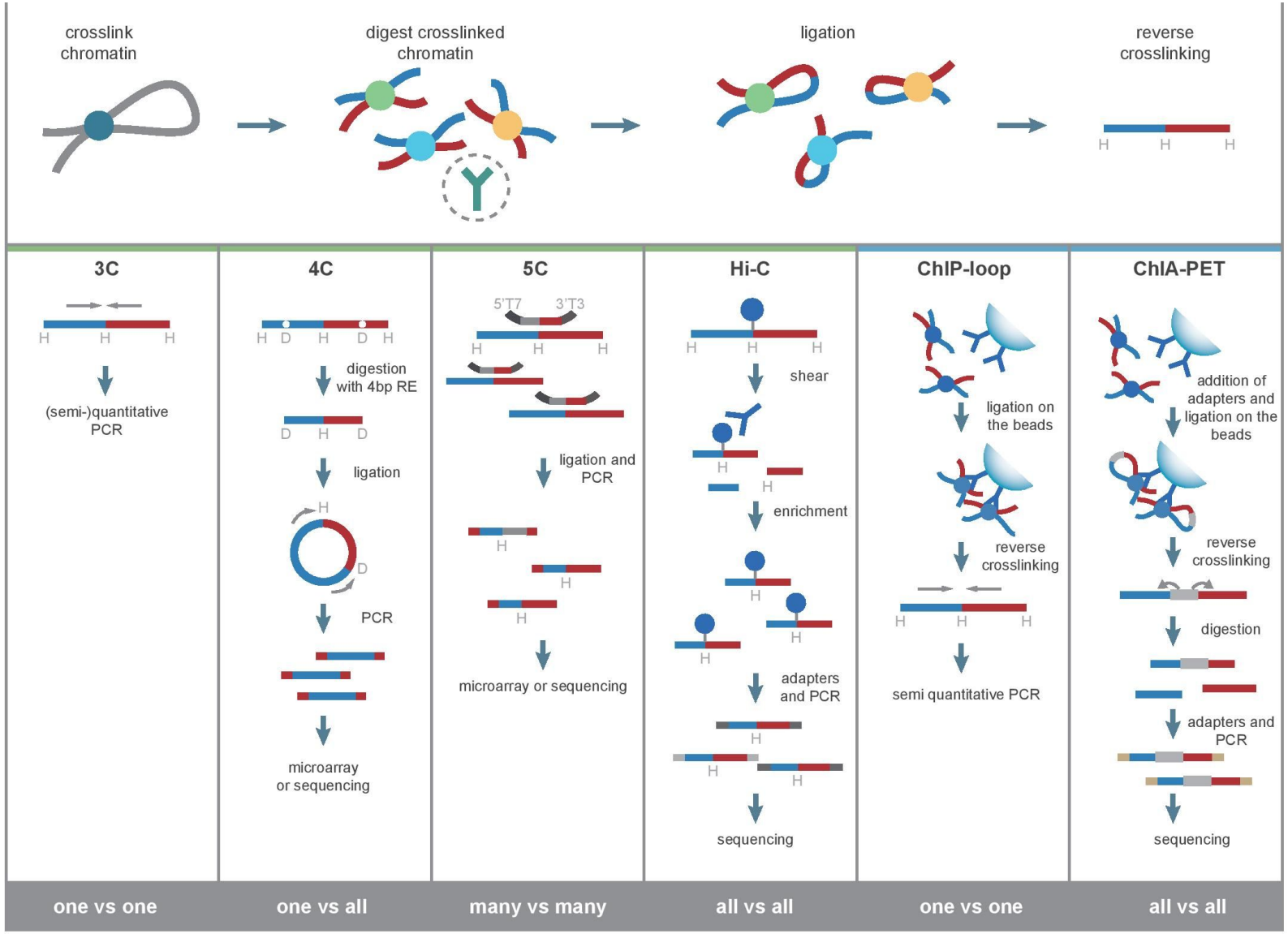
Low input ChIP *in situ*: CUT&RUN



https://elifesciences.org/articles/21856

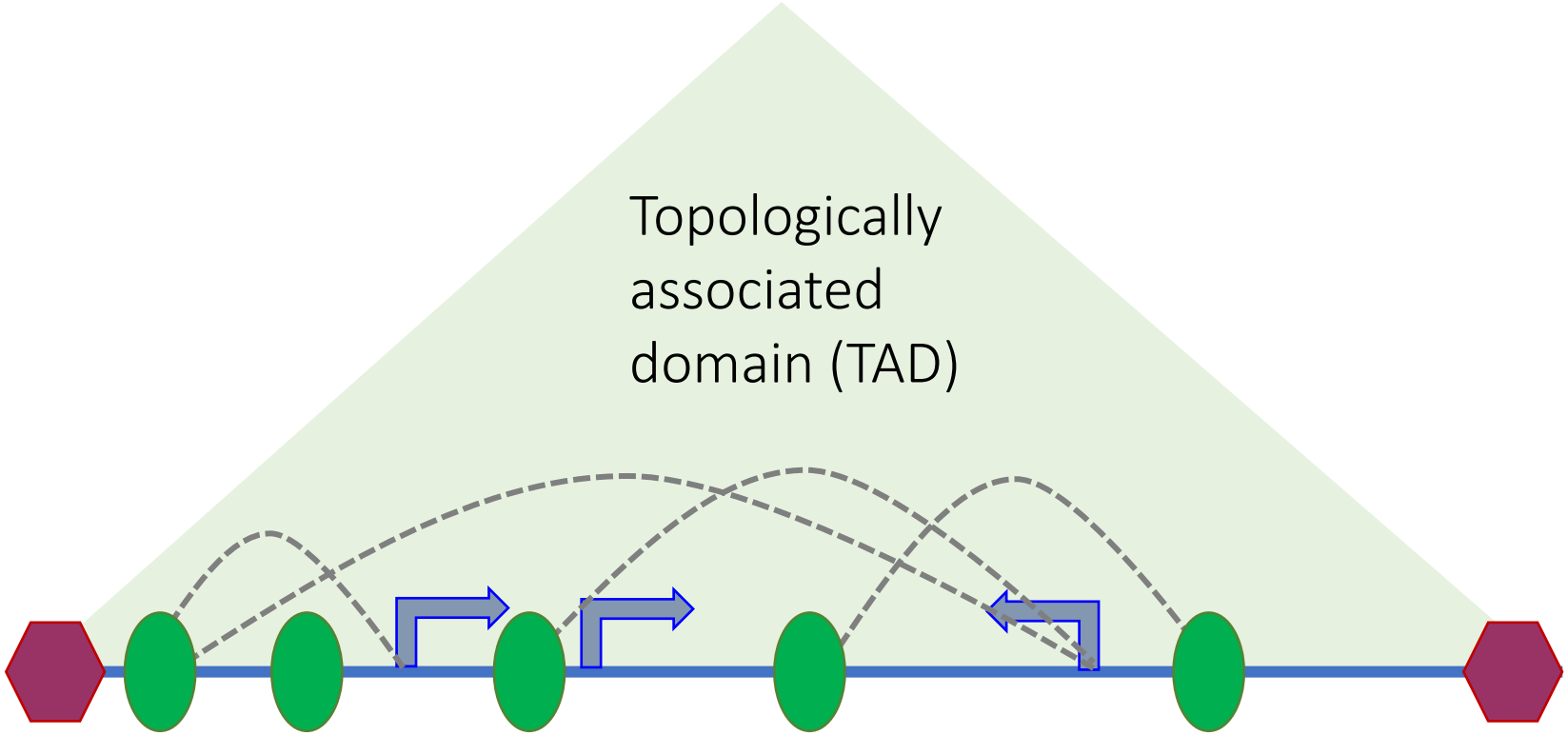
https://elifesciences.org/articles/46314

Overview of “C” based methods

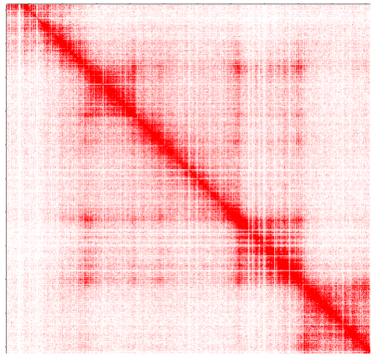


https://bmcgenomics.biomedcentral.com/articles/10.1186/1471-2164-15-S12-S11

Chromatin Conformation topology (HiC)



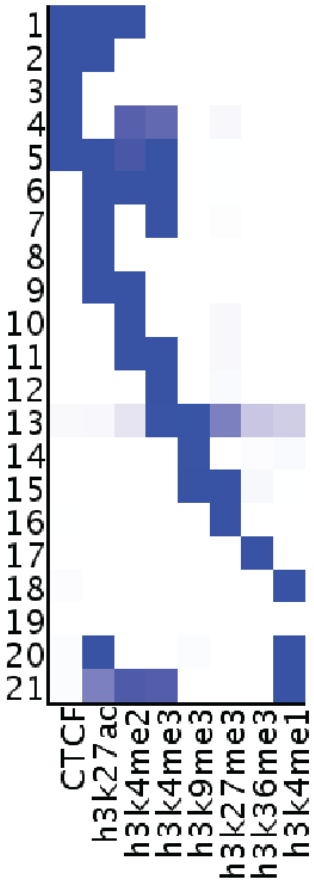
Chromatin
topology (HiC)



Outline

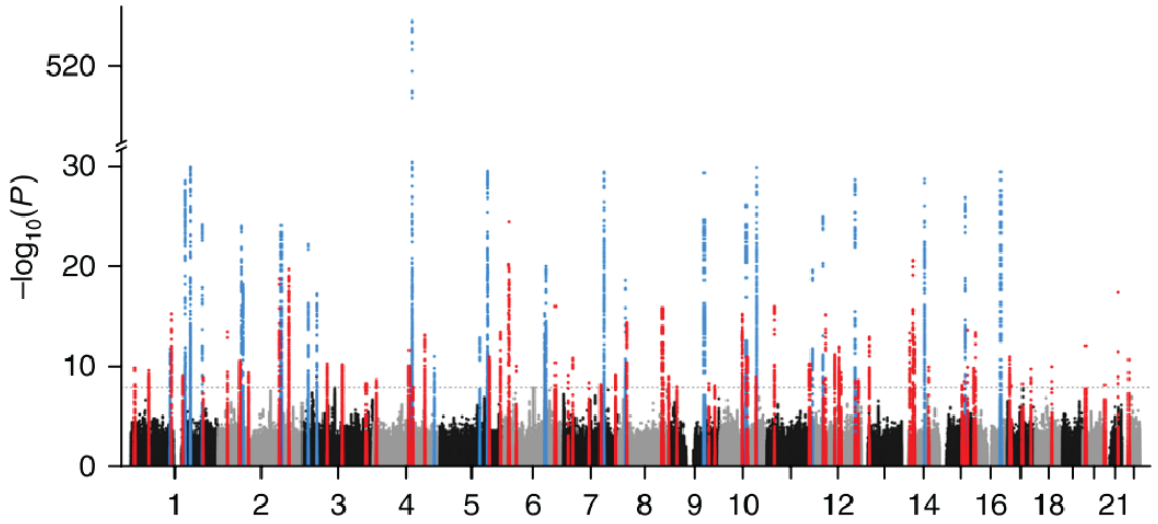
1. Architecture of DNA in the nucleus
2. History of epigenomics
3. Techniques for studying the epigenome
4. Analysis methods for the epigenome
5. Epigenomics and clinical treatments/trials

Can we use epigenetic data to inform the function of GWAS loci?



Chromatin state model

+



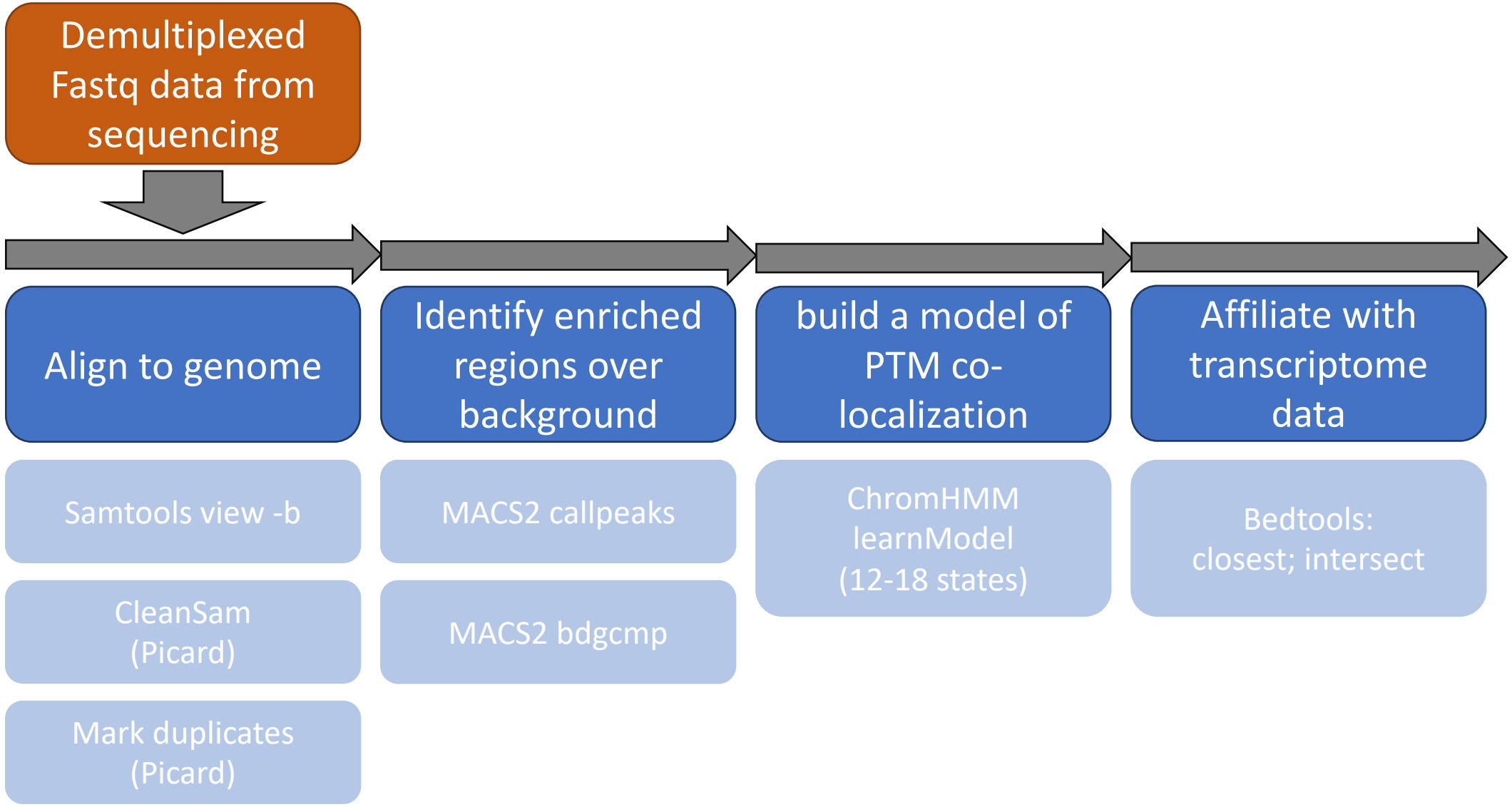
CVD GWAS loci & proxies $R^2 > 0.8$

Roselli et al., Nature Genetics, 2018

Analysis methods for the epigenome

- Data processing (after alignment):
  - Signal/noise detection (peak calling, for ChIP and CUT&RUN data)
  - “variant” detection – for methyl-seq, to identify methylated cytosines
  - Loop detection/validation
- Downstream analysis of data:
  - Peak annotation – proximity to transcripts
  - Pathway analysis of proximal genes

Pipeline: processing ChIP-seq data to build a model



Useful resources and utilities I

- Gene annotation: Gencode
  - https://www.gencodegenes.org/
- Gene expression: GTEx
  - https://gtexportal.org/home/
- Pre-existing chromatin annotation datasets:
  - Roadmap Epigenomics: http://www.roadmapepigenomics.org/data/
  - ENCODE project data: https://www.encodeproject.org/
  - VISTA enhancer collection: https://enhancer.lbl.gov/
- UCSC genome browser: http://genome.ucsc.edu
  - Annotations: genes, DNA repeats, conservation, ENCODE data (not all), multiple species available

Useful resources and utilities II

- R: Bioconductor https://www.bioconductor.org/
- Python: Biopython https://biopython.org/
- Genome arithmetic: bedtools https://bedtools.readthedocs.io/en/latest/
- Signal/noise discrimination: MACS2 https://github.com/taoliu/MACS
- Motif identification and analysis: MEME-Suite http://meme-suite.org/
- Online course on NGS (fastq and alignment tutorial): https://wikis.utexas.edu/display/CoreNGSTools/Core+NGS+Tools+Home
- HUGO Genenames: https://www.genenames.org/
  - For validating gene names across genome assemblies

Useful resources and utilities III

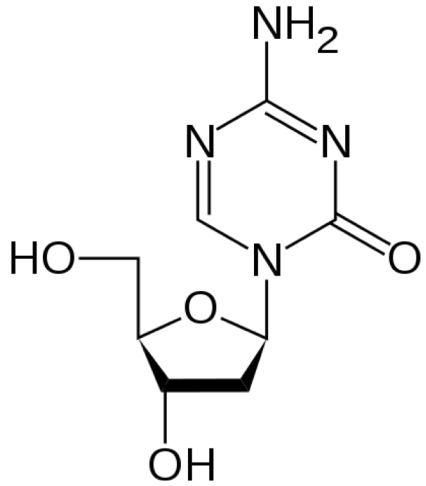
- Pathway analysis:
  - Enrichr: https://amp.pharm.mssm.edu/Enrichr/
  - DAVID: https://david.ncifcrf.gov/
  - GeneOntology/Panther: http://geneontology.org/
- Non model organism genomes:
  - Ensembl: https://useast.ensembl.org/index.html
  - NCBI genebank: https://www.ncbi.nlm.nih.gov/genbank/
- GWAS trait lookup:
  - EBI-GWAS catalog: https://www.ebi.ac.uk/gwas/
- Protein information:
  - Uniprot: https://www.uniprot.org/

Outline

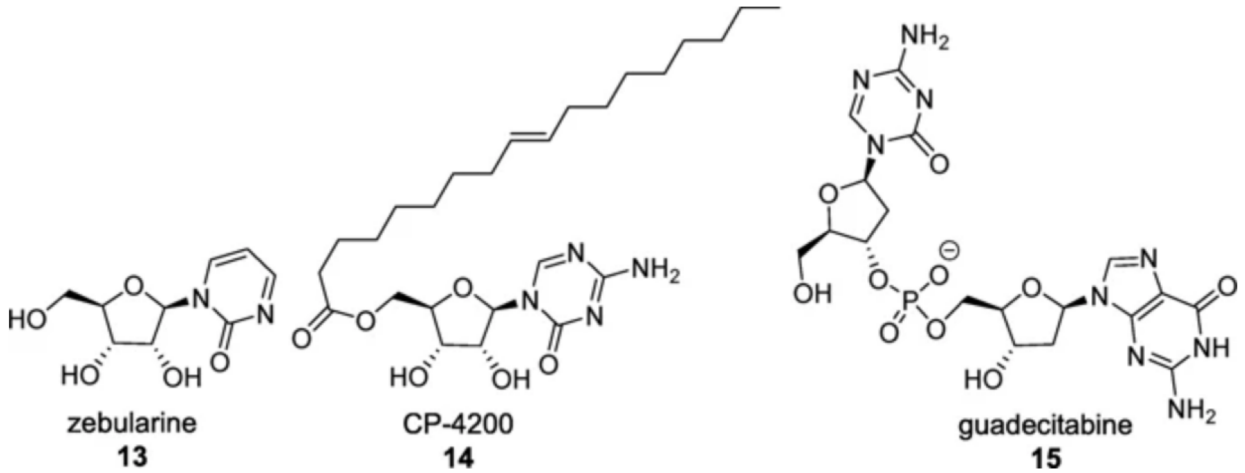
1. Architecture of DNA in the nucleus
2. History of epigenomics
3. Techniques for studying the epigenome
4. Analysis methods for the epigenome
5. Epigenomics and clinical treatments/trials

Epigenomics and clinical treatments/trials

- Some existing drugs are epigenetic modulators!
  - 5-azacytidine (5-azaC or azacytidine), 5-aza-2′-deoxycytidine (5-aza-dC or decitabine)
    - **DNA methyltransferase (DNMT)** inhibitors
  - Used in treatment for myelodysplastic syndrome (MDS), a bone marrow disorder with a high risk of progressing to AML that occurs primarily in elderly patients and is characterized by the production of abnormal blood cells



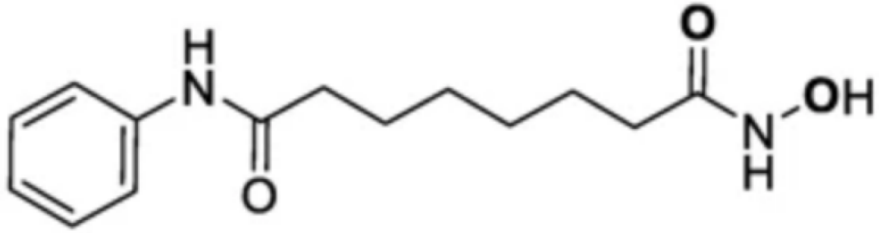
decitabine



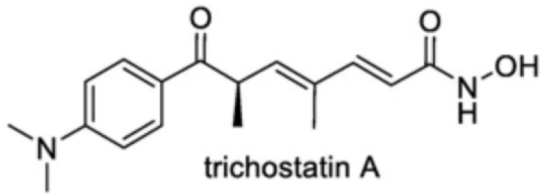
https://clinicalepigeneticsjournal.biomedcentral.com/articles/10.1186/s13148-019-0776-0

HDAC inhibitors in current usage I

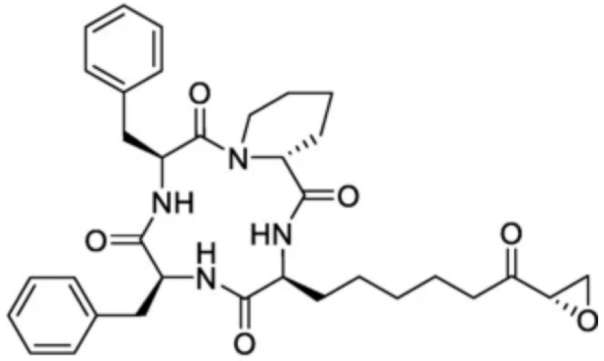
- HDAC: histone de-acetylase
  - So a HDAC inhibitor would prevent removal of acetyl (activating) groups by HDAC enzymes
- Vorinostat & Romidepsin
  - T cell lymphoma
- Trichostatin A and Trapoxin A
  - antifungals



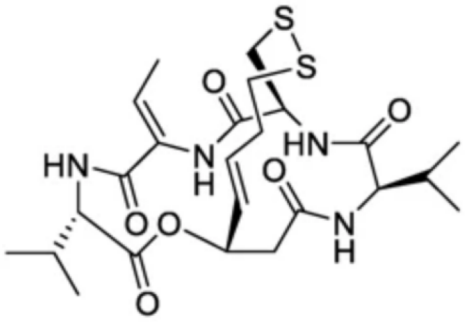
suberoylanilide hydroxamic acid, **10**
(SAHA, vorinostat, Zolinza)
2 μM



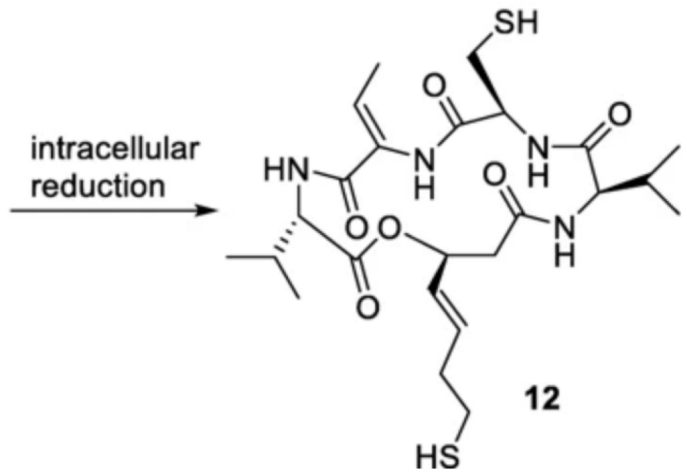
trichostatin A



trapoxin A



FK228 or romidepsin
11

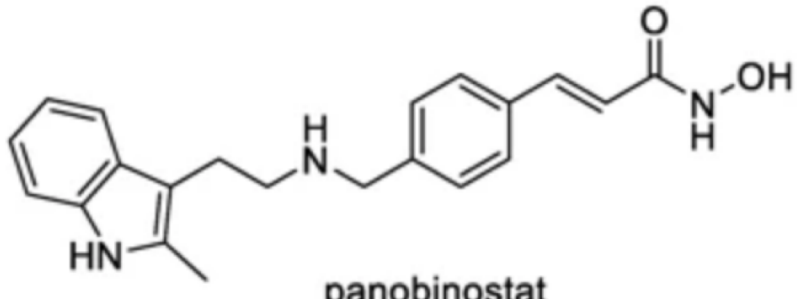


12

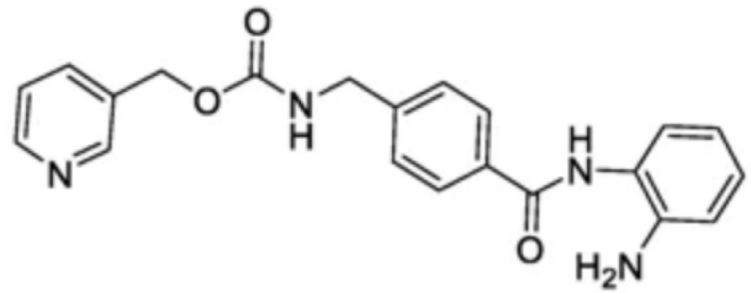
https://clinicalepigeneticsjournal.biomedcentral.com/articles/10.1186/s13148-019-0776-0

HDAC inhibitors in current usage II

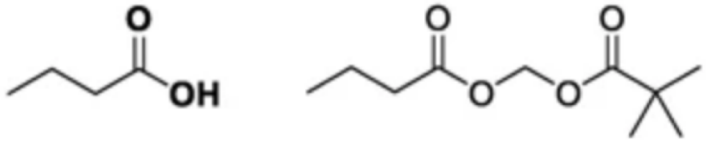
- Second generation versions of Vorinostat
  - Hydroxamic acid based compounds
- Benzamide HDACi
  - Breast cancer clinical trials
- Carboxylic acid HDACi
  - Butyric acid reported to be a HDACi as early as 1978
  - Treatments for Spinal Muscular Dystrophy, Epilepsy



panobinostat
18

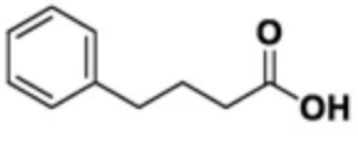


entinostat
23

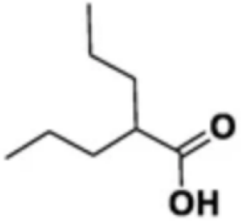


butyric acid
26

Pivanex
27



phenylbutyric acid
28

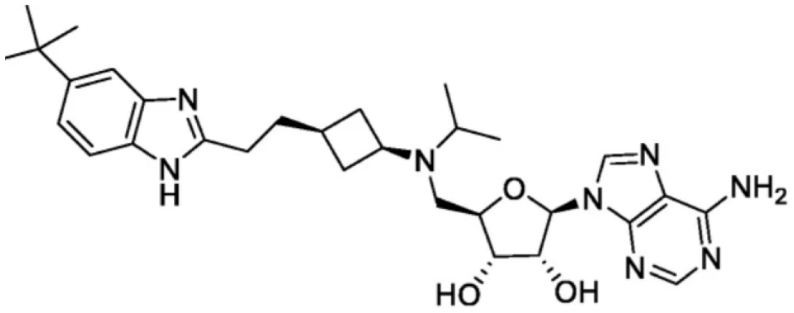


valproic acid
29

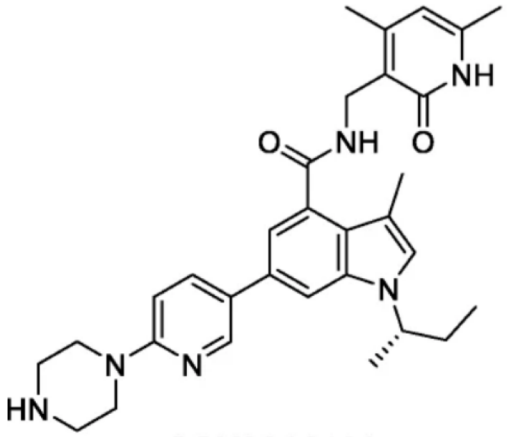
https://clinicalepigeneticsjournal.biomedcentral.com/articles/10.1186/s13148-019-0776-0

The current generation of drugs in development I

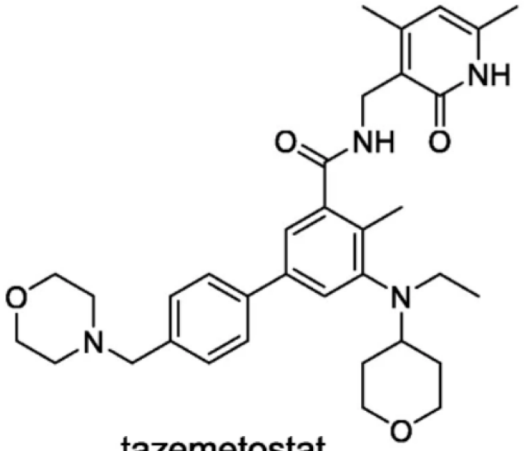
- Modulating Lysine Histone Methyl Transferases
  - These enzymes add methyl groups to lysine residues on histones
  - *DOT1L*
    - H3K79 methyl transferase (activating mark), pinometostat in trials for leukemia
  - *EZH2* inhibitors
    - EZH2 catalyzes H3K27me3 (repressive) mark, GSK and tazemetostat in trials for B-cell lymphoma



pinometostat



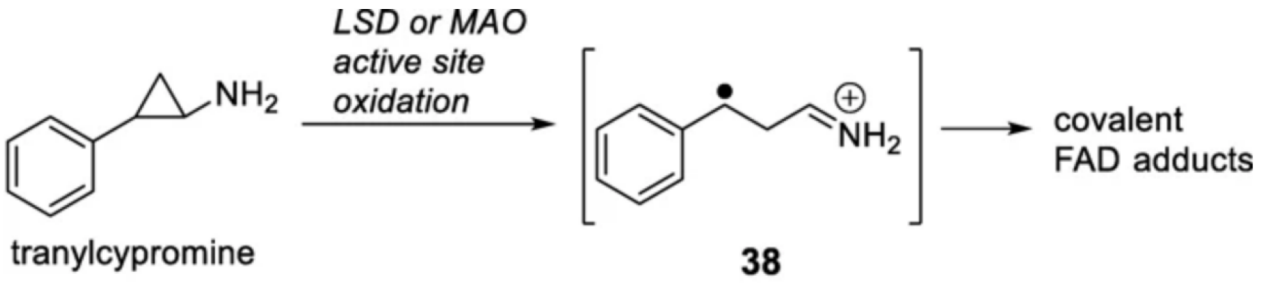
GSK2816126



tazemetostat

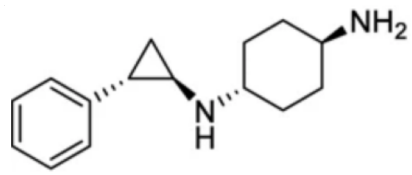
The current generation of drugs in development II

- Lysine demethylase inhibitors
  - These enzymes remove the methyl groups to lysine residues on histones
- *KDM1, LSD1/2*
  - Mono and dimethyl lysine transferases, similar to MAOIs
  - Drug repurposing! Trials for AML which is resistant to ATRA



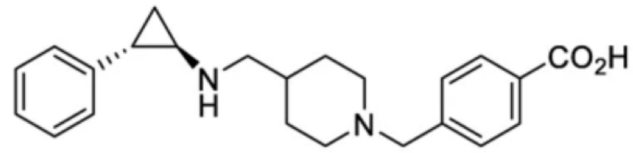
tranylcypromine

38



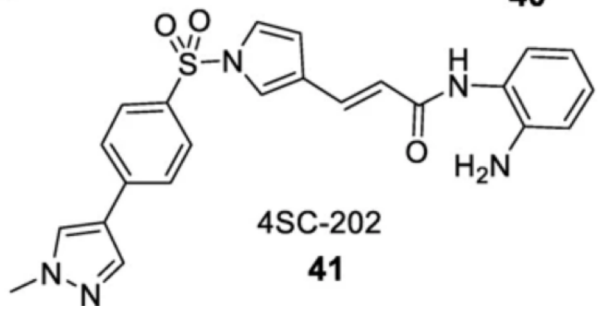
ORY-1001

39



GSK2879552

40

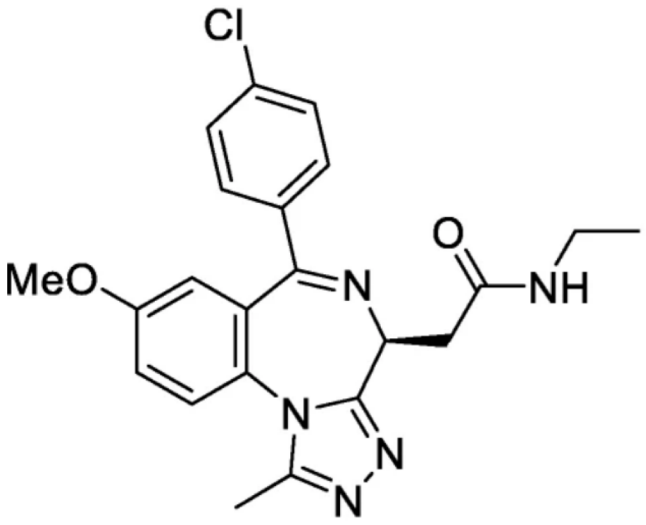


4SC-202

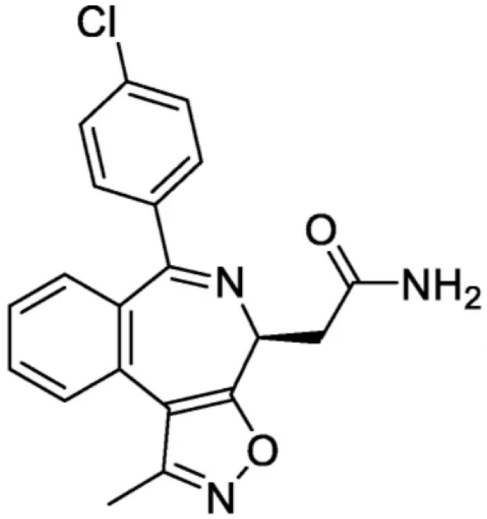
41

The current generation of drugs in development III

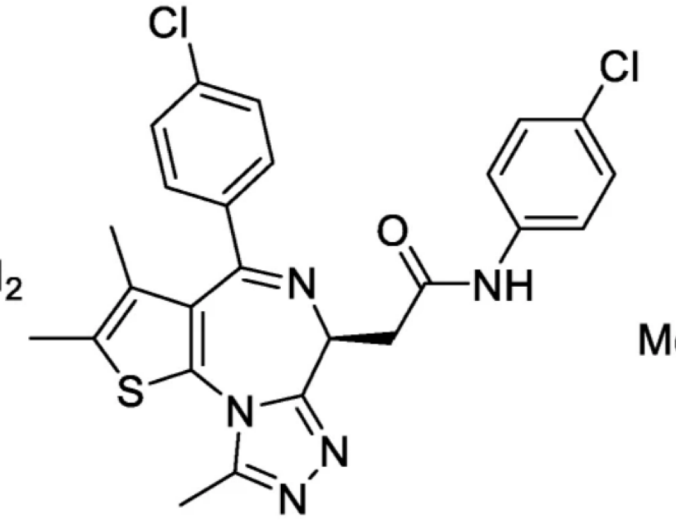
- Bromodomains
  - acetyl “readers”
- *BRD2, BRD3, BRD4*
  - Inhibition inhibits “stemness”
  - Also relevant for diabetes, HIV-1 latent reactivation



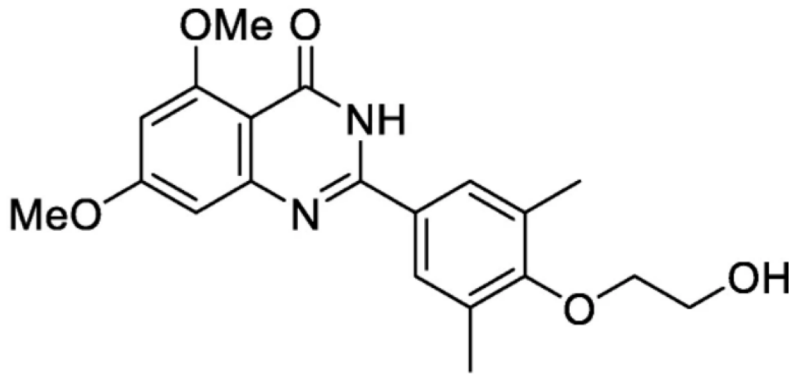
I-BET762



CPI-0610



OTX015



RVX-280

Thank you! Questions?

- Need slides, or answers to questions about epigenomic/transcriptomic analyses in detail? Email me!
  - ahall22@mgh.harvard.edu
- Working on a computational workgroup (virtual) where we will go over pathway enrichment, RNA-seq analysis
  - https://groups.google.com/a/broadinstitute.org/forum/#!forum/computational-workgroup-bio
  - Happy to provide notes from previous sessions (Jan and Feb, have been on hiatus since late March)