



GHGA Webinar

• A Beginner's Guide to Epigenetics

Dr. Kathrin Kattler
Department of Epigenetics
Saarland University

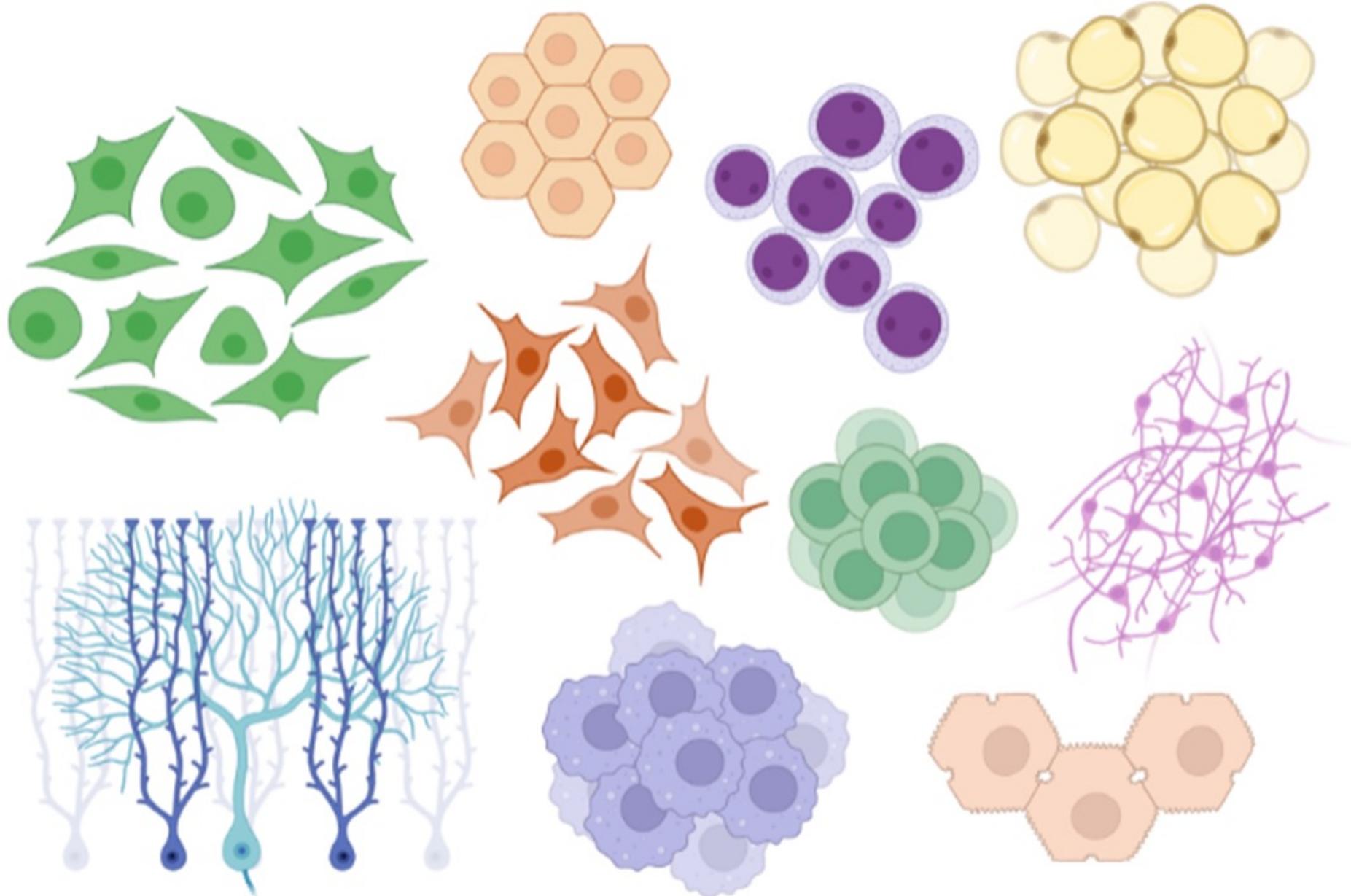


Part I

Exploring the Epigenetic Landscape

→ Epigenetic modifications and mechanisms

GHGA

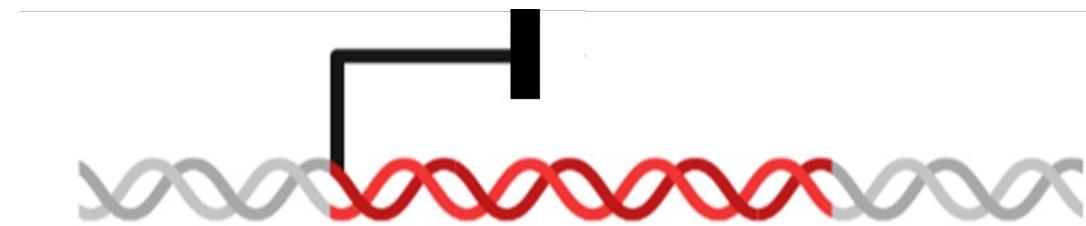
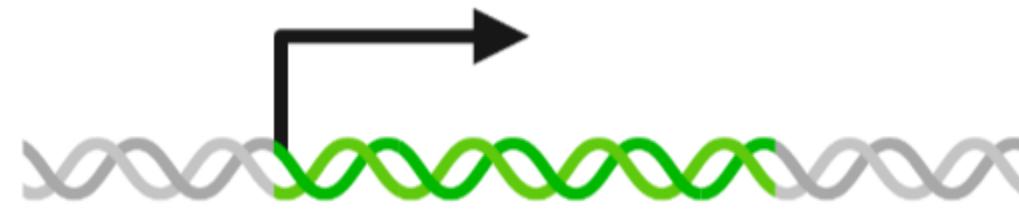




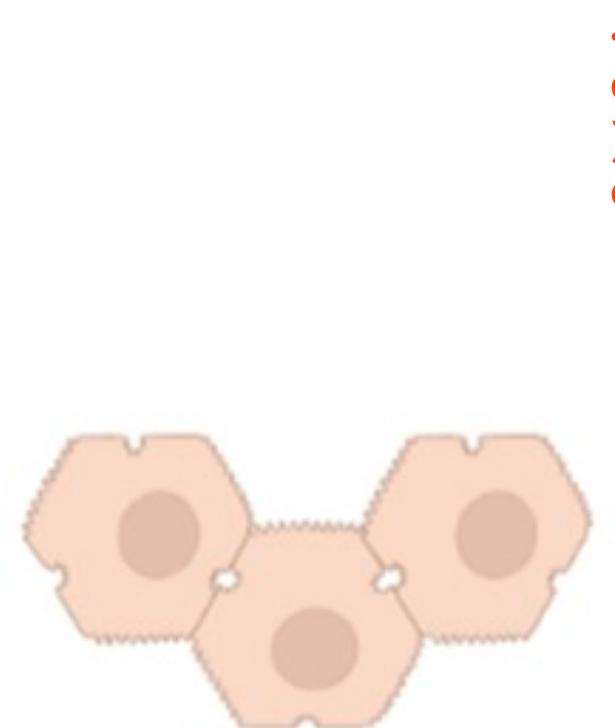
DNA

GHGA

Gene regulation



→ Epigenetics



GHGAA

Epigenetics

“the branch of biology which studies the causal interactions between genes and their products, which bring the phenotype into being”

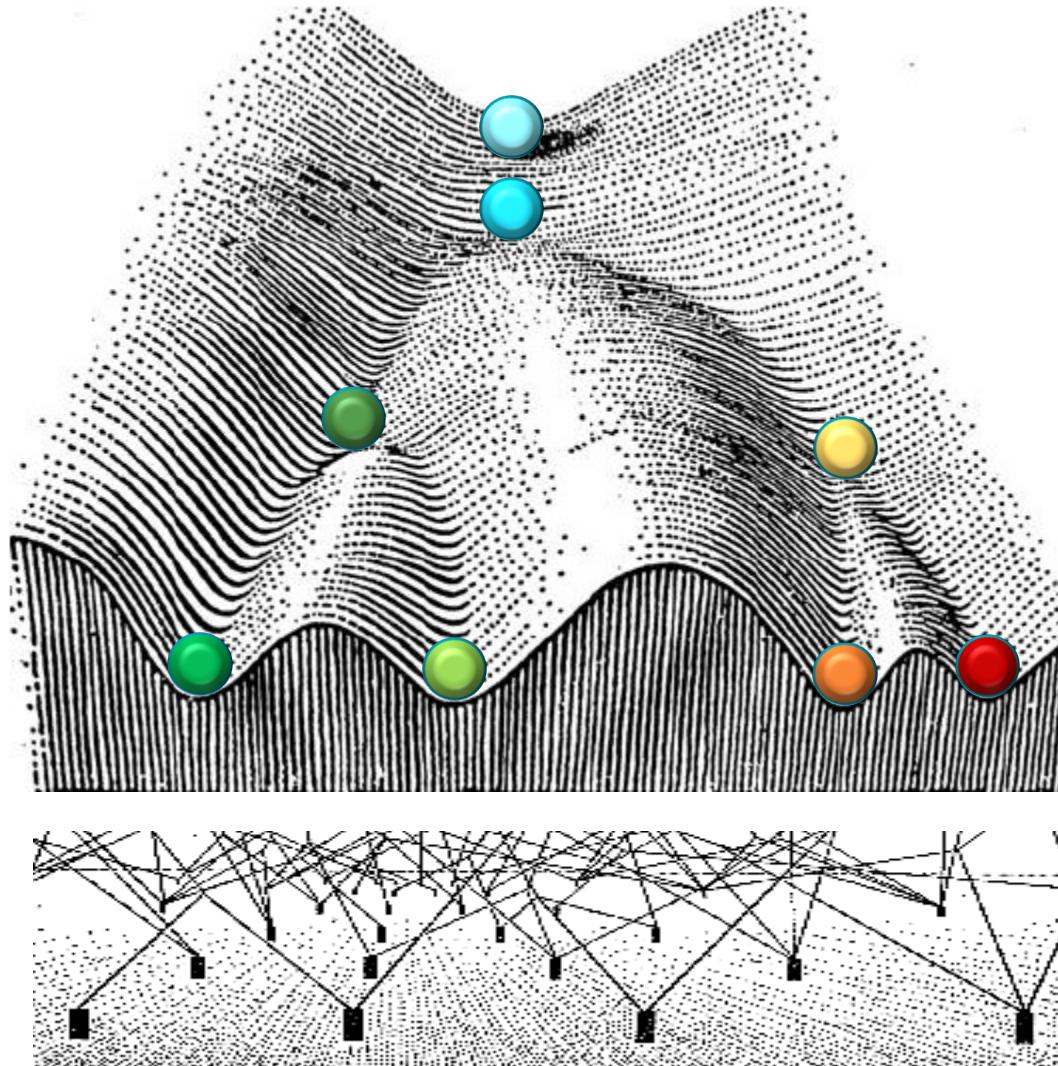
Conrad H. Waddington (1942)

GHGA

→ Study of heritable phenotypic changes

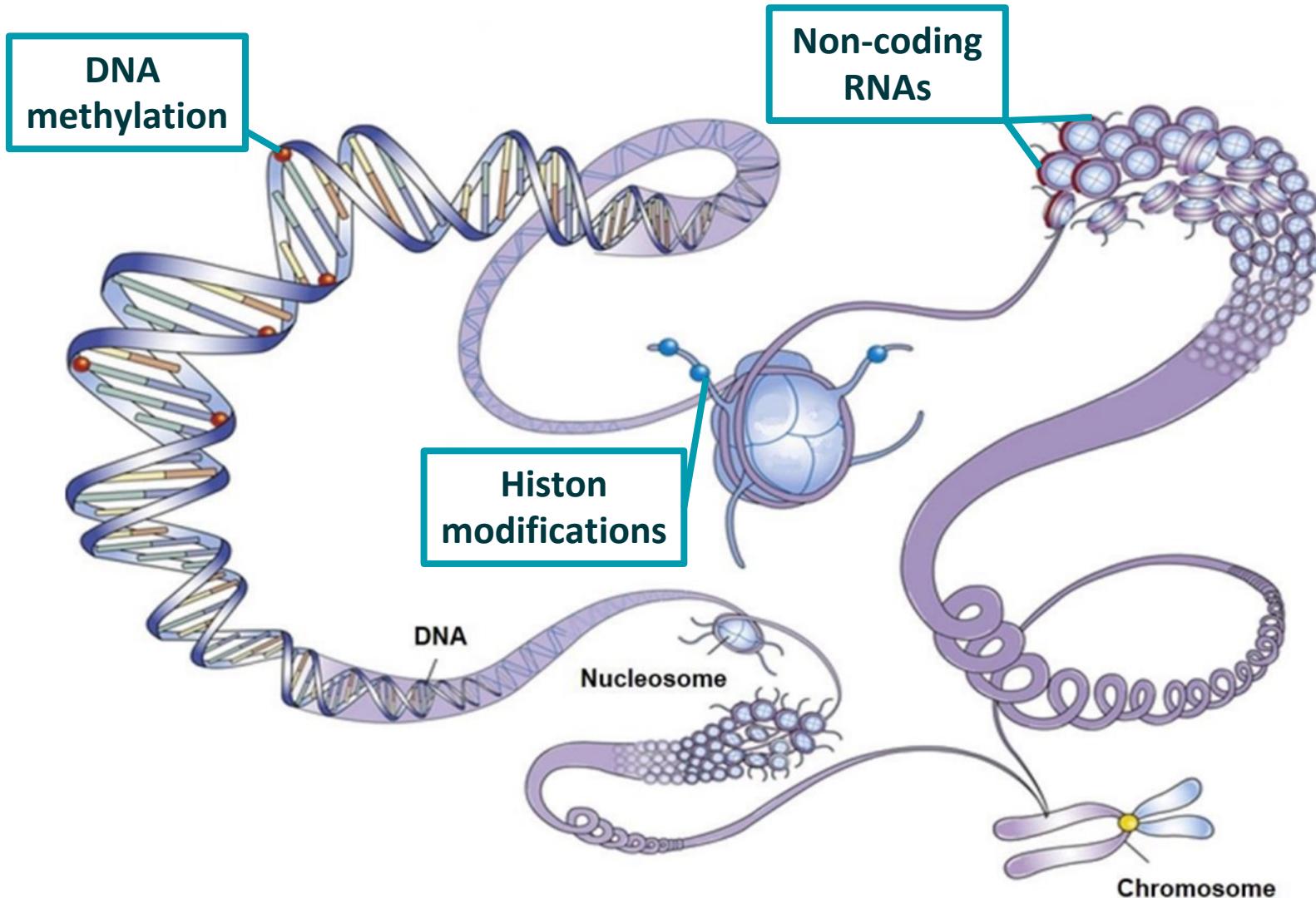
- no changes in DNA sequence
- reversible modifications of DNA or chromatin → altered gene activity

The Epigenetic Landscape



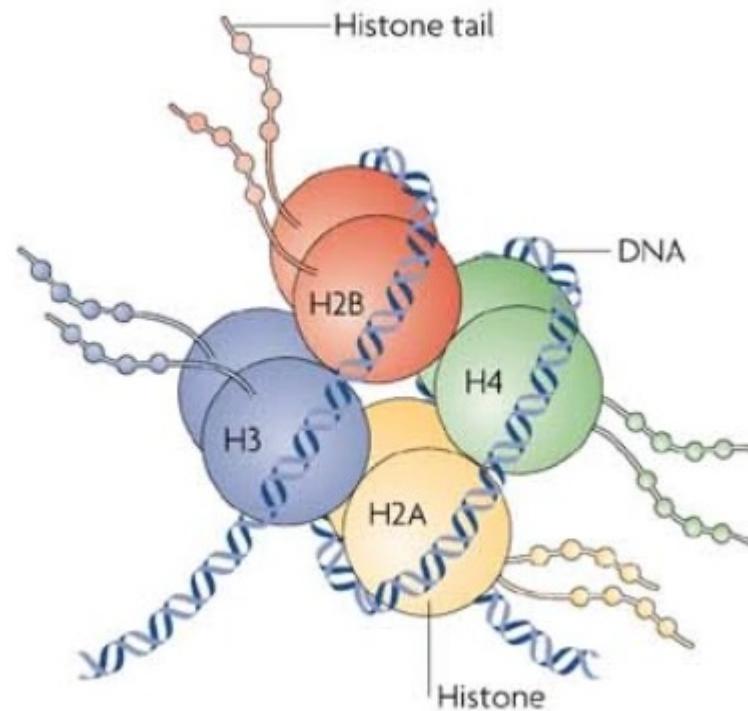
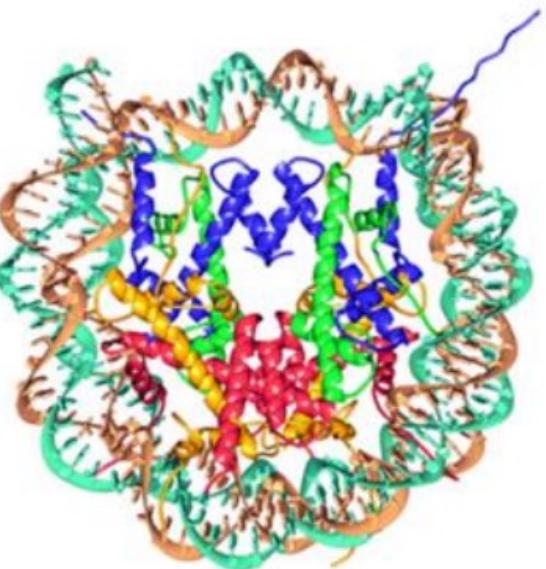
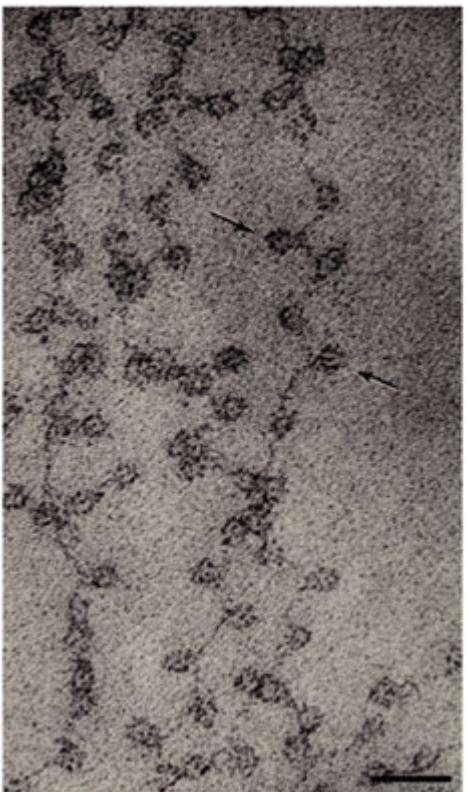
GHGA

Epigenetic modifications



GHGA

How is the DNA packaged into the nucleus?

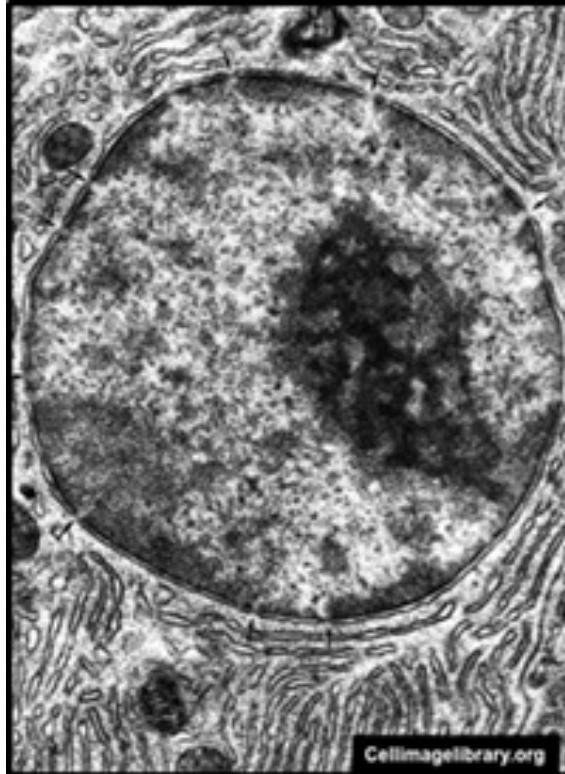


GHGAA

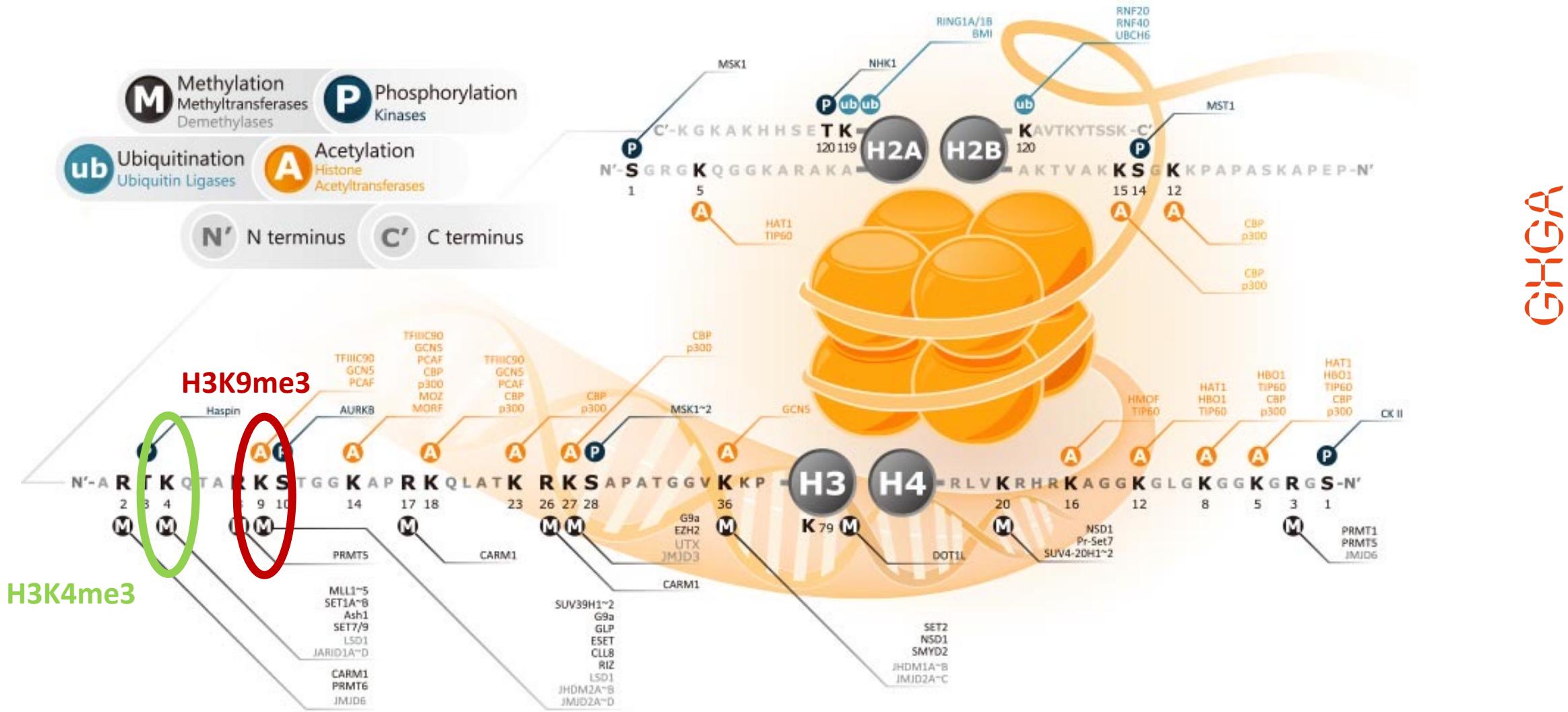
→ chromatin

Chromatin: foundation of epigenetic control

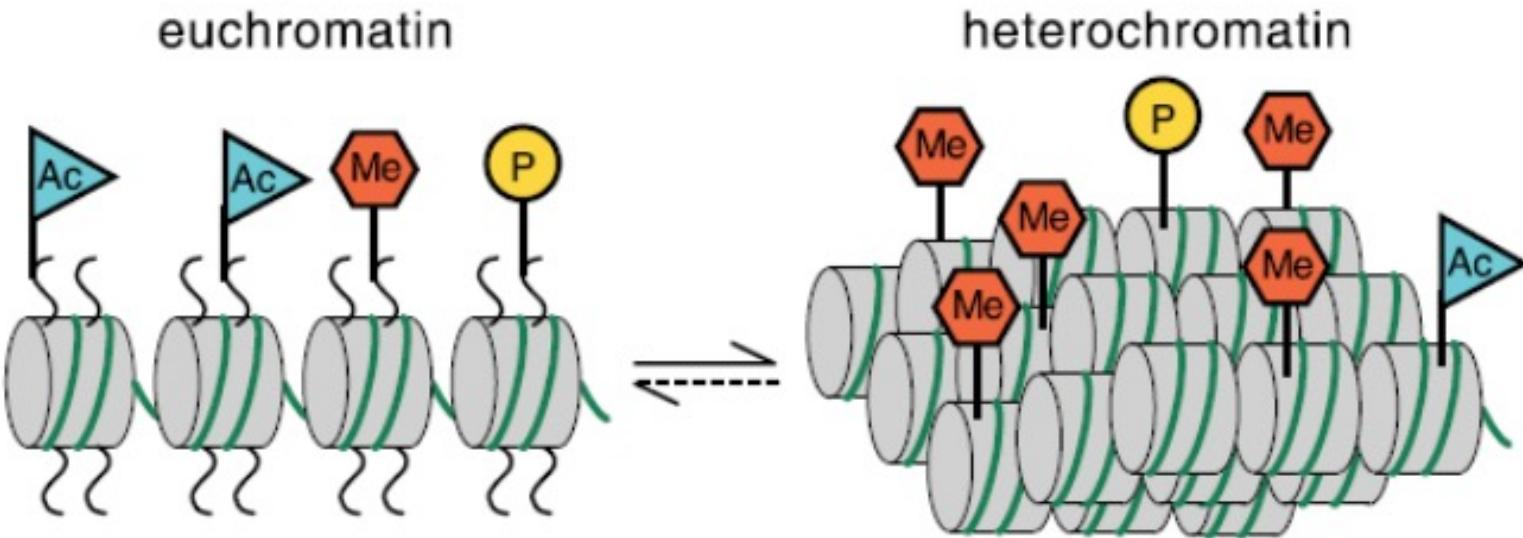
Pancreatic cell
→ highly active euchromatin



Histone modifications



The histone code

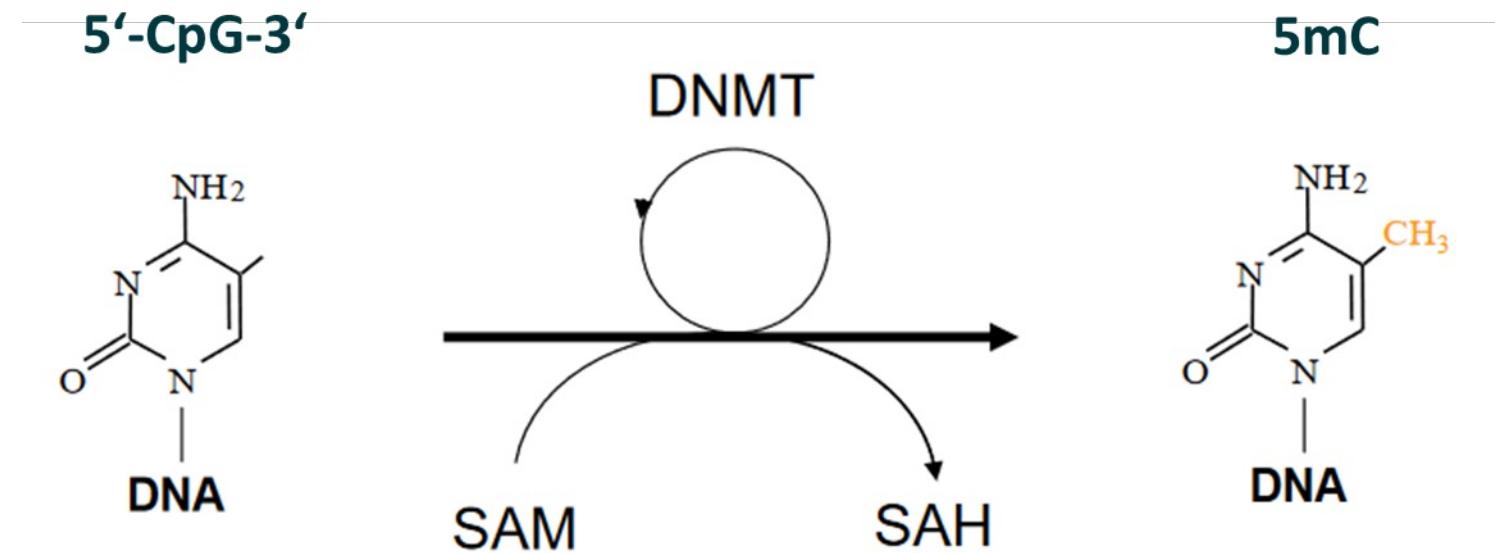


- H3K4me3 (trimethylation of lysine 4 on histone H3)
- H3K9ac (acetylation of lysine 9 on histone H3)
- H3K27ac (acetylation of lysine 27 on histone H3)
- H3K36me3 (trimethylation of lysine 36 on histone H3)
- H4K16ac (acetylation of lysine 16 on histone H4)
- H2BK5ac (acetylation of lysine 5 on histone H2B)
- H2BK120ac (acetylation of lysine 120 on histone H2B)
- H3S10ph (phosphorylation of serine 10 on histone H3)
- H3K79me2 (dimethylation of lysine 79 on histone H3)
- ...

- H3K9me2 (dimethylation of lysine 9 on histone H3)
- **H3K9me3** (trimethylation of lysine 9 on histone H3)
- H3K27me3 (trimethylation of lysine 27 on histone H3)
- H4K20me3 (trimethylation of lysine 20 on histone H4)
- H3K9me1 (monomethylation of lysine 9 on histone H3)
- H3K27me2 (dimethylation of lysine 27 on histone H3)
- H3K79me3 (trimethylation of lysine 79 on histone H3)
- ...

GHGAA

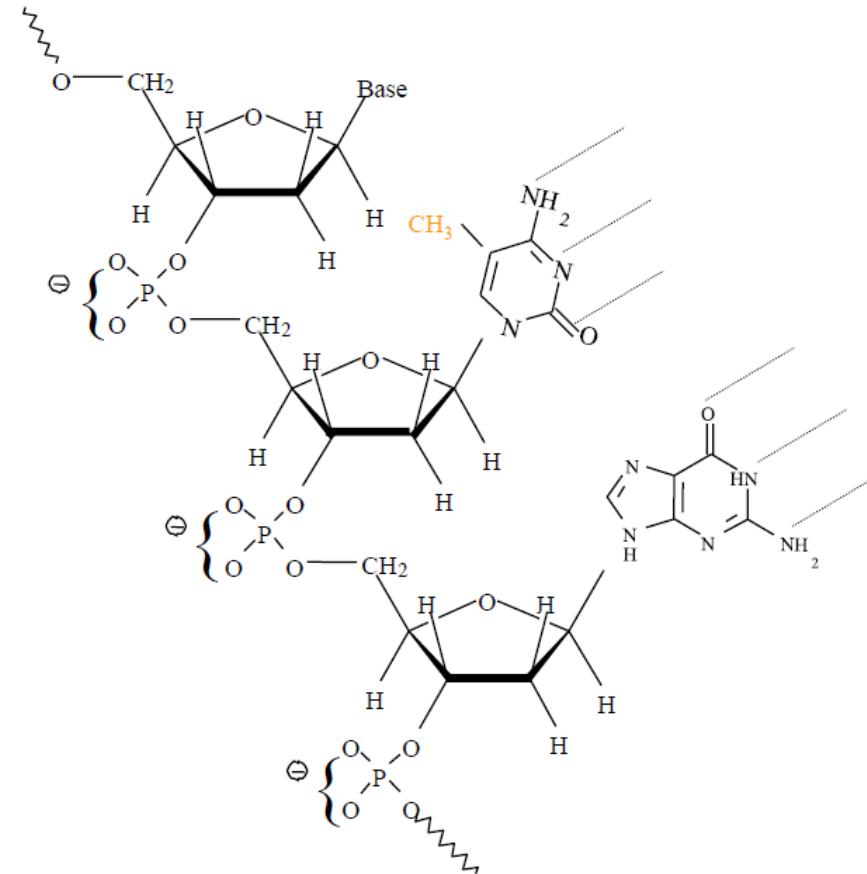
DNA methylation



GHGA

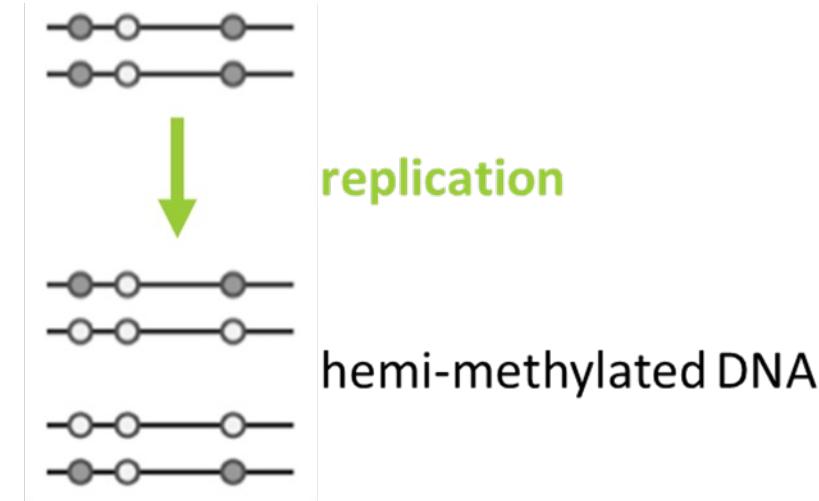
DNA methylation

- 5mC does not change the DNA sequence
- Palindromic sequence 
5'...CG...3'
3'...GC...5' 
- 5mC is more hydrophobic than unmethylated C → effect on DNA structure and stability
- 5mC is reversible



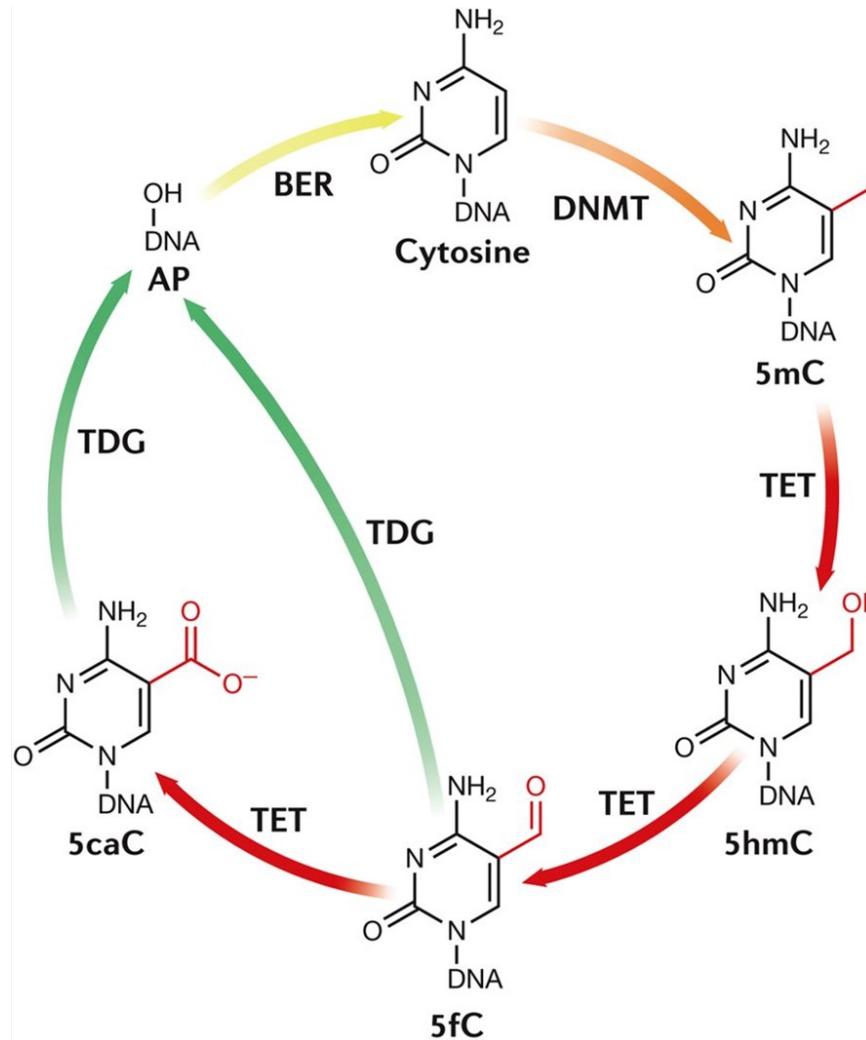
GHGA

Passive DNA Demethylation



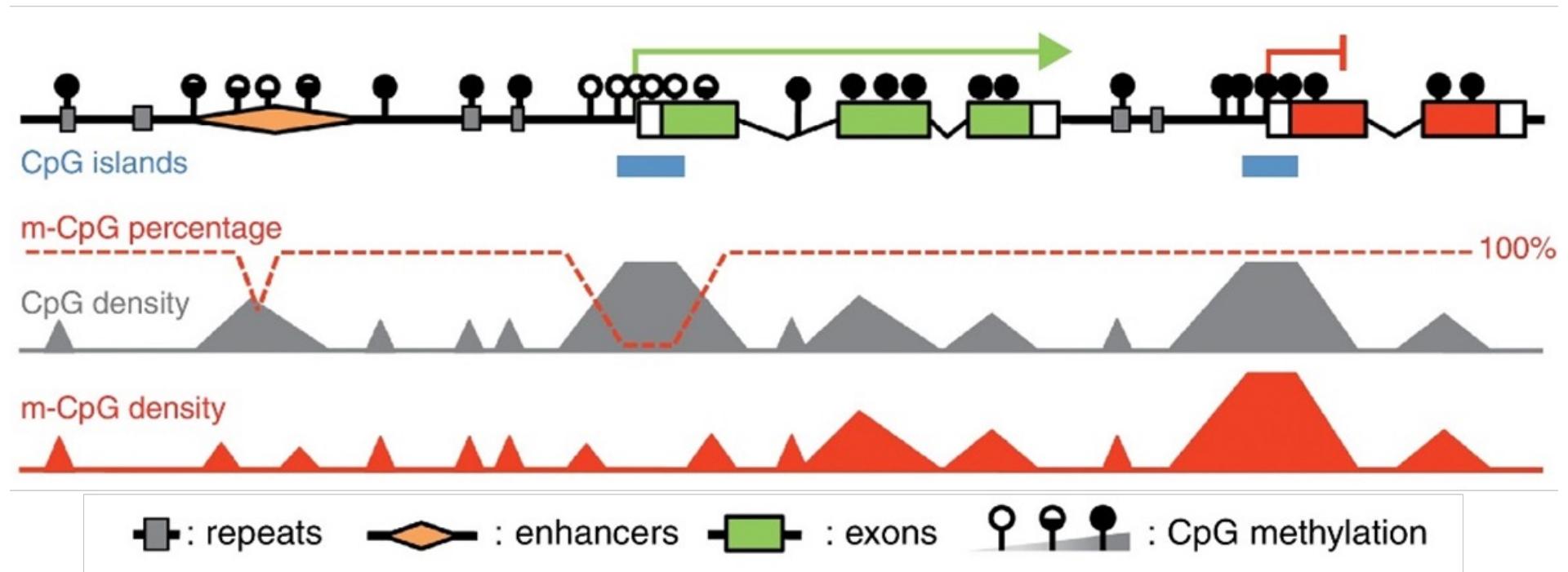
GHGA

Active DNA Demethylation

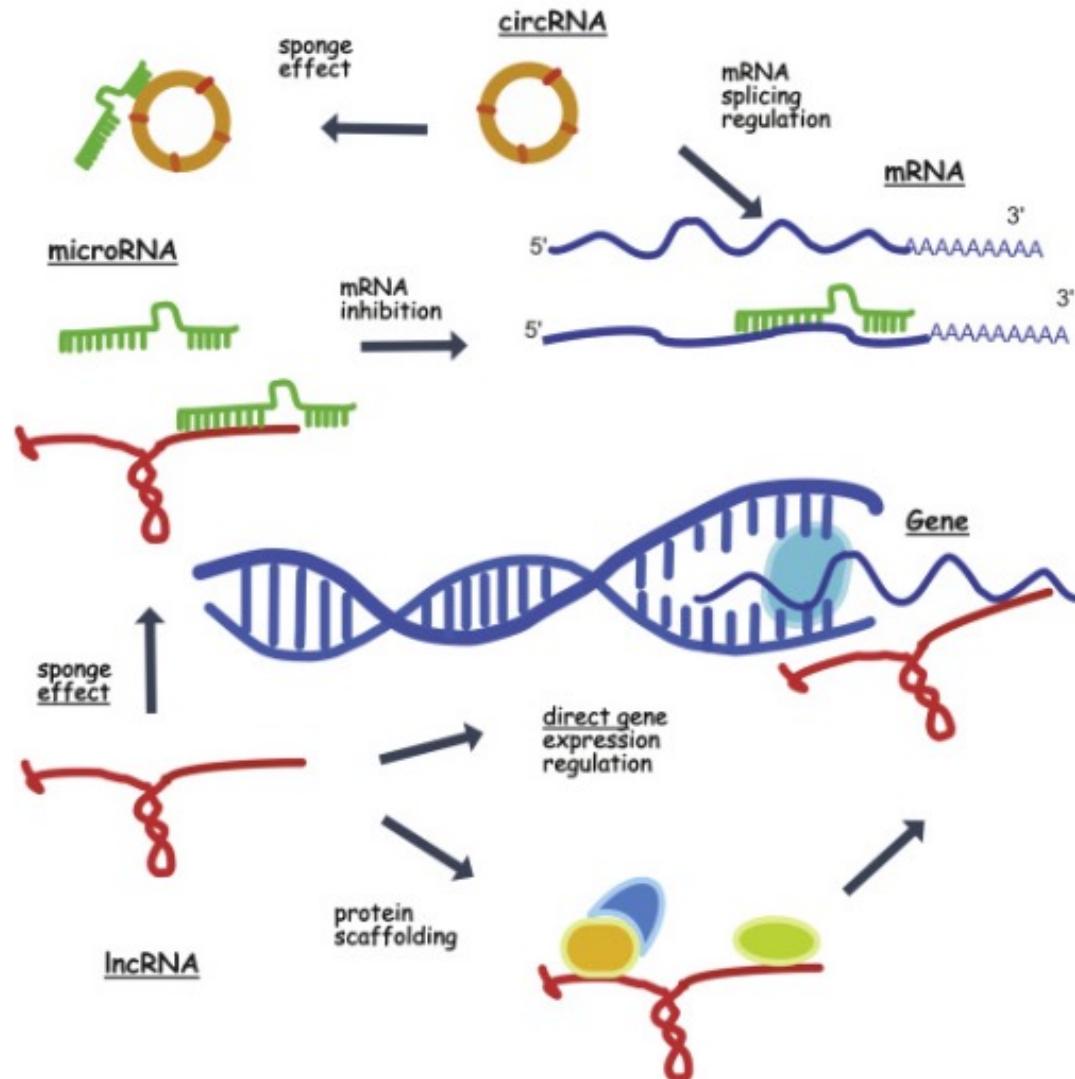


GHGA

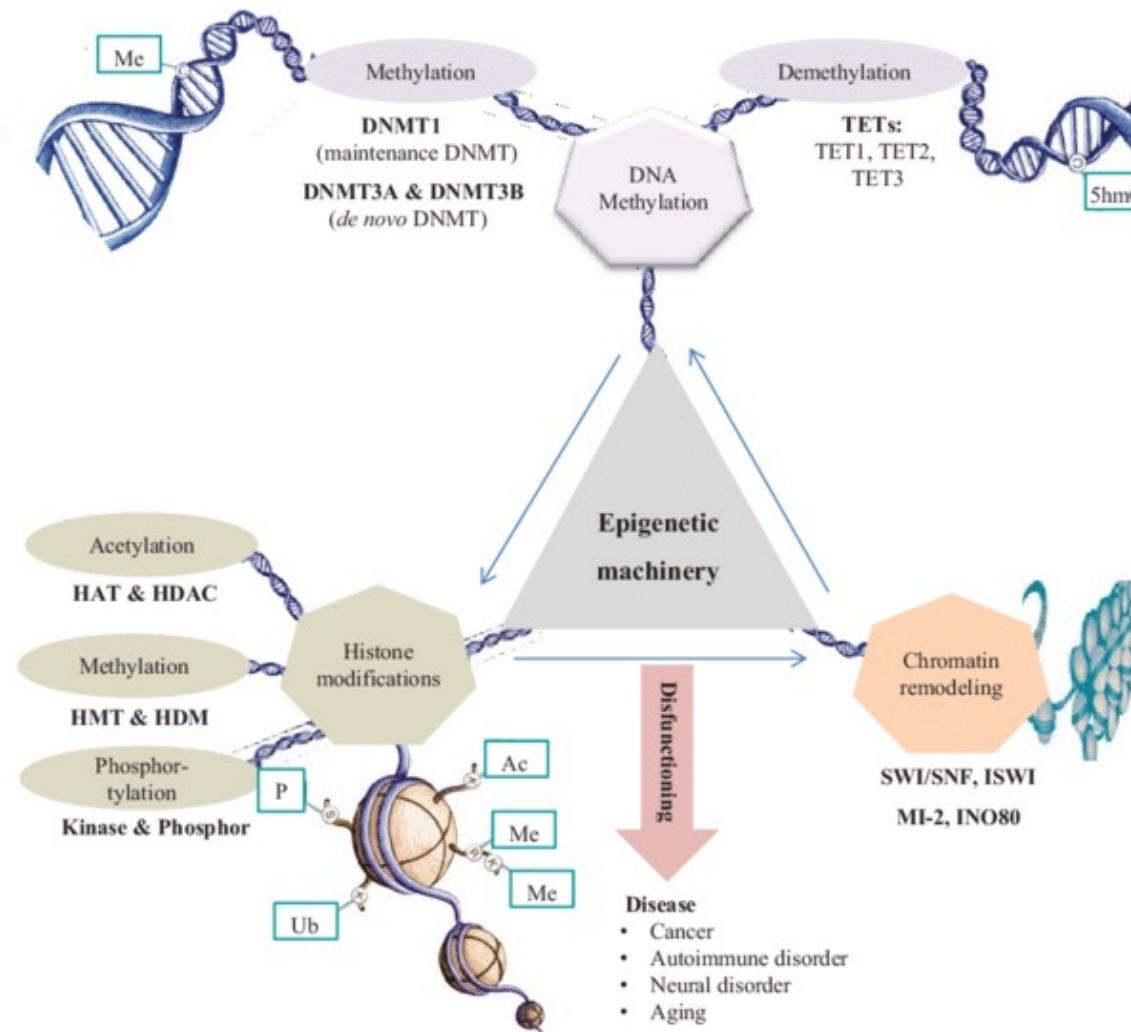
DNA methylation in regulatory elements



Non-coding regulatory RNAs

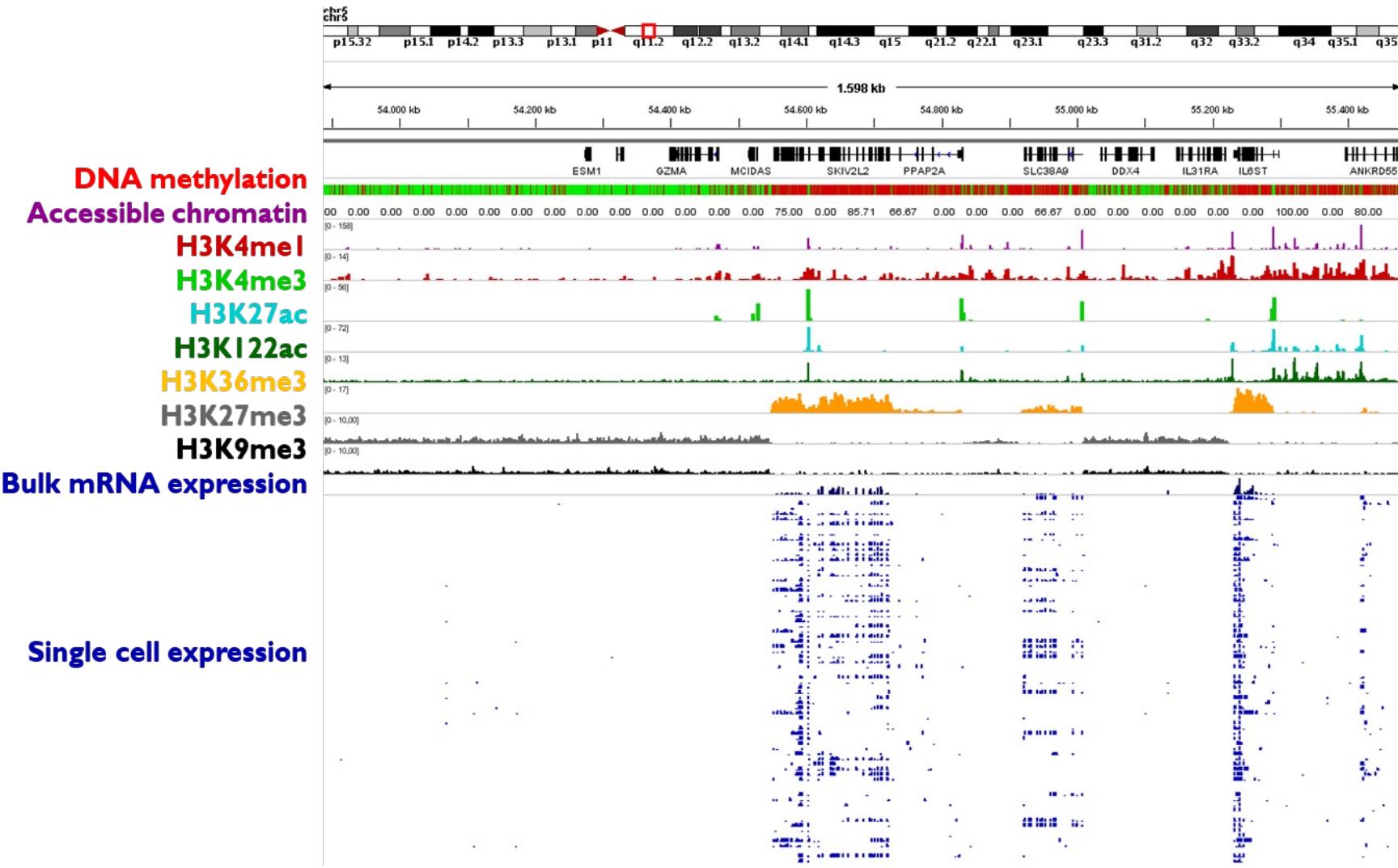


Interplay of epigenetic modifications



GHGΑ

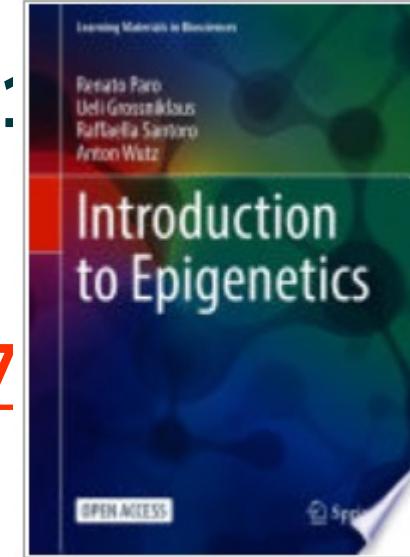
Interplay of epigenetic modifications



Resources – Part I

- Paro, Grossniklaus, Santoro, Wutz (2021)
Introduction to Epigenetics
(Springer) [DOI:10.1007/978-3-030-6867](https://doi.org/10.1007/978-3-030-6867)

[open access]



- <https://epigenie.com/epigenie-learning-center/>



QUESTIONS?

Part I

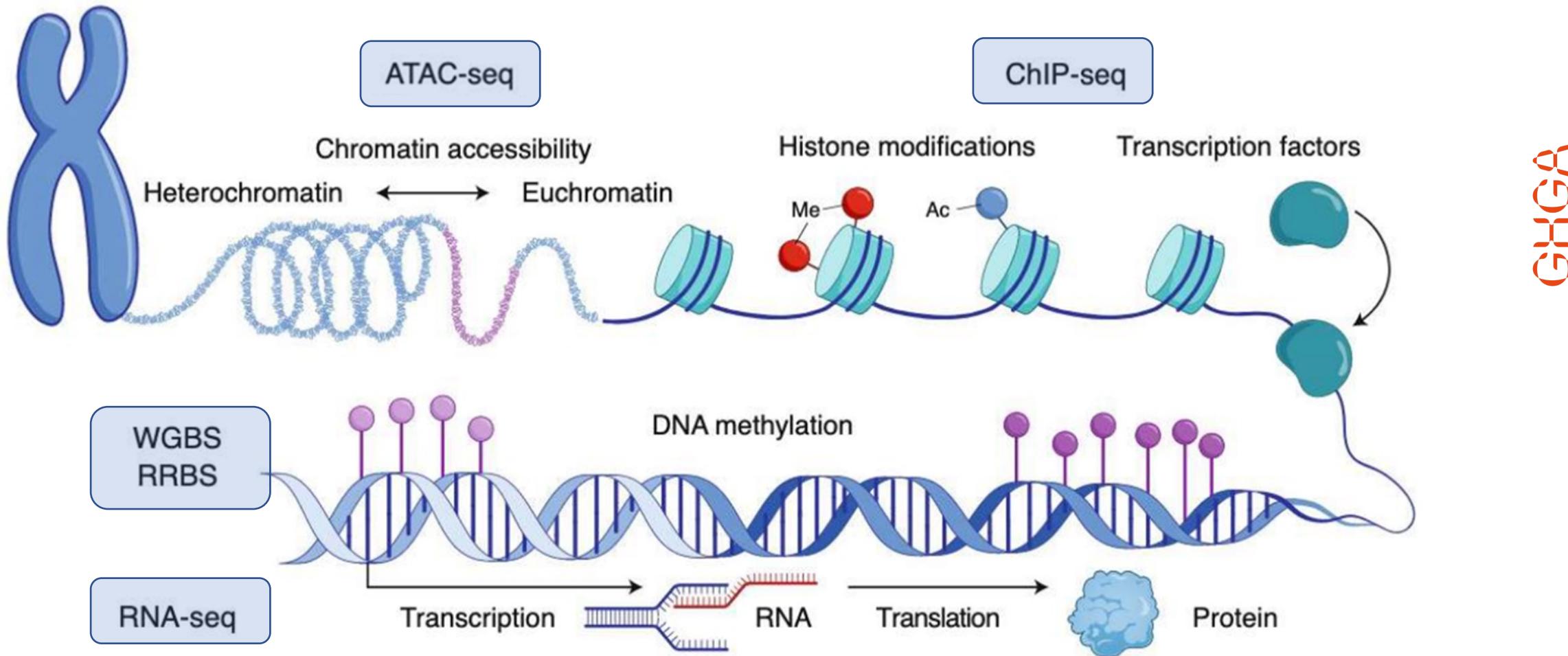
Part II

Epigenetic Toolbox: Introduction into Epigenomic Data Analysis

- Methods for genome-wide analysis of epigenetic modifications
- Data analysis tools and strategies

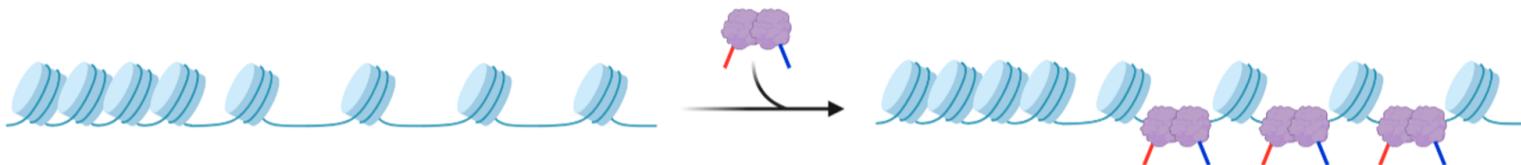


Key methods to analyse epigenomic layers on a genome-wide scale

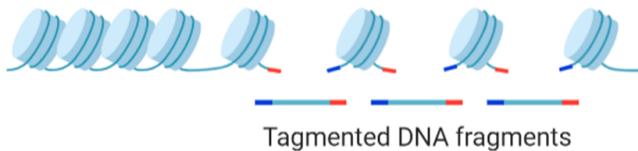


Chromatin accessibility analysis: ATAC-seq

- 1 Binding of Tn5 transposases to open chromatin



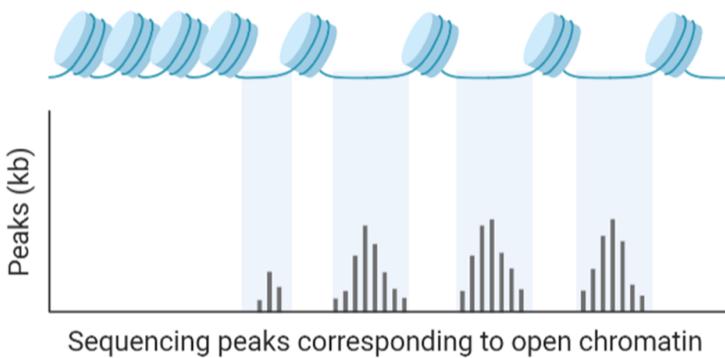
- 2 Fragmentation and tagging (tagmentation) by Tn5 transposase



- 3 DNA fragment purification and PCR amplification

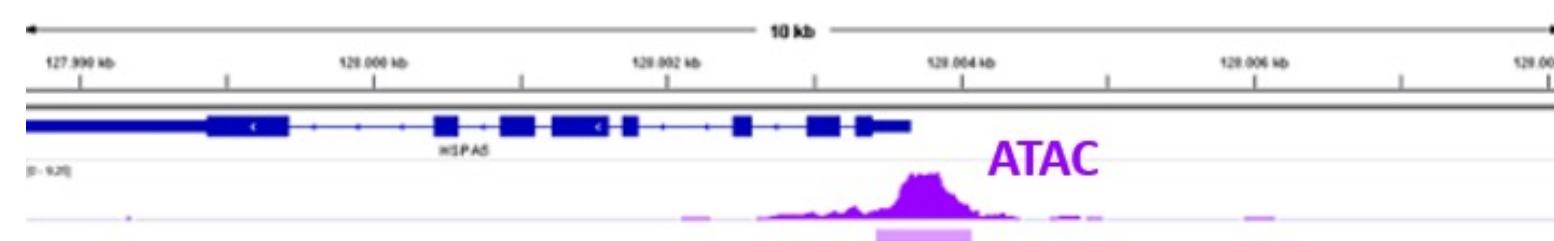
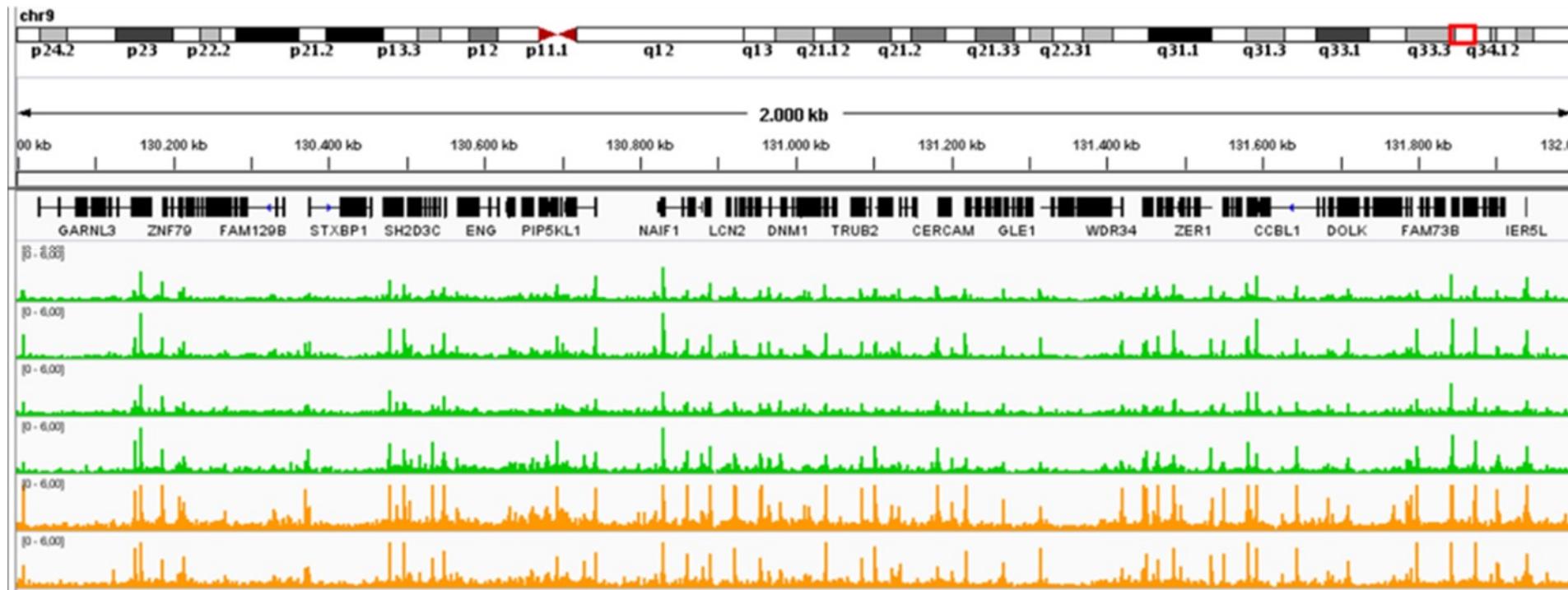


- 4 Next generation sequencing of PCR product and data analysis



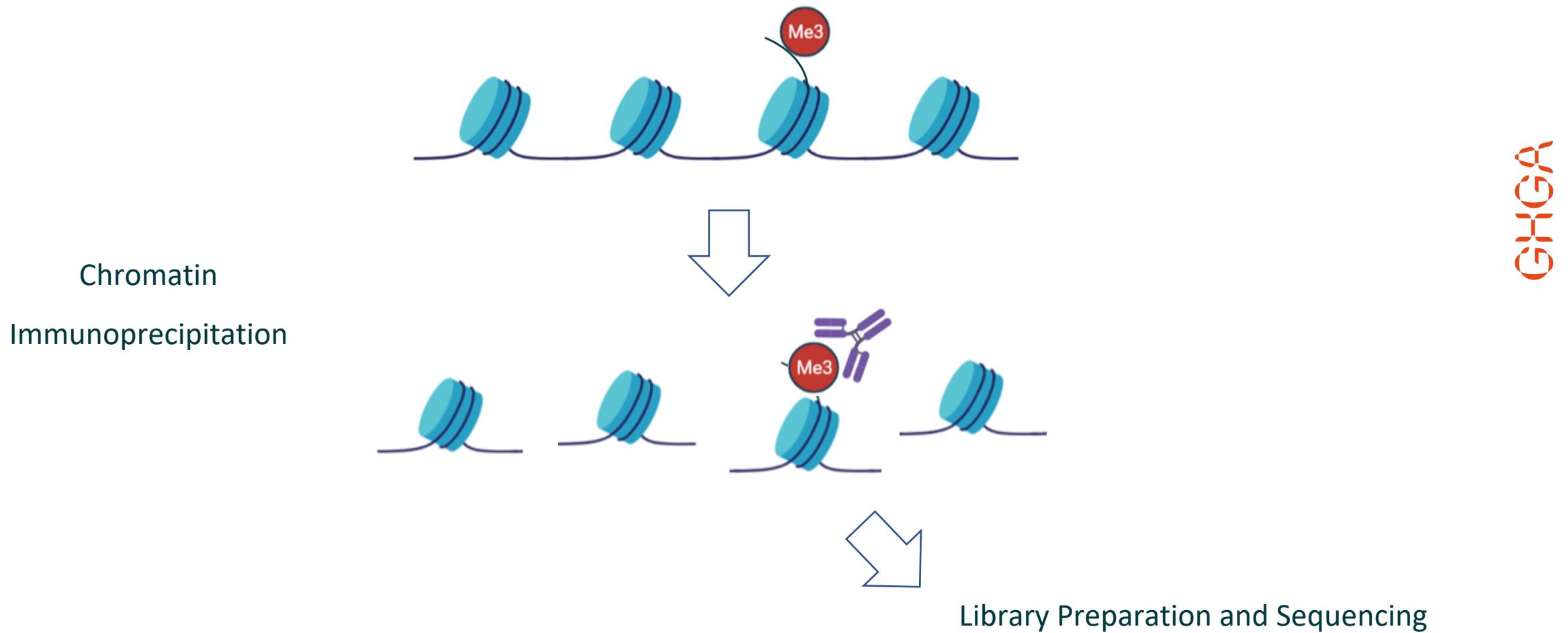
GHGA

ATAC-seq profiles

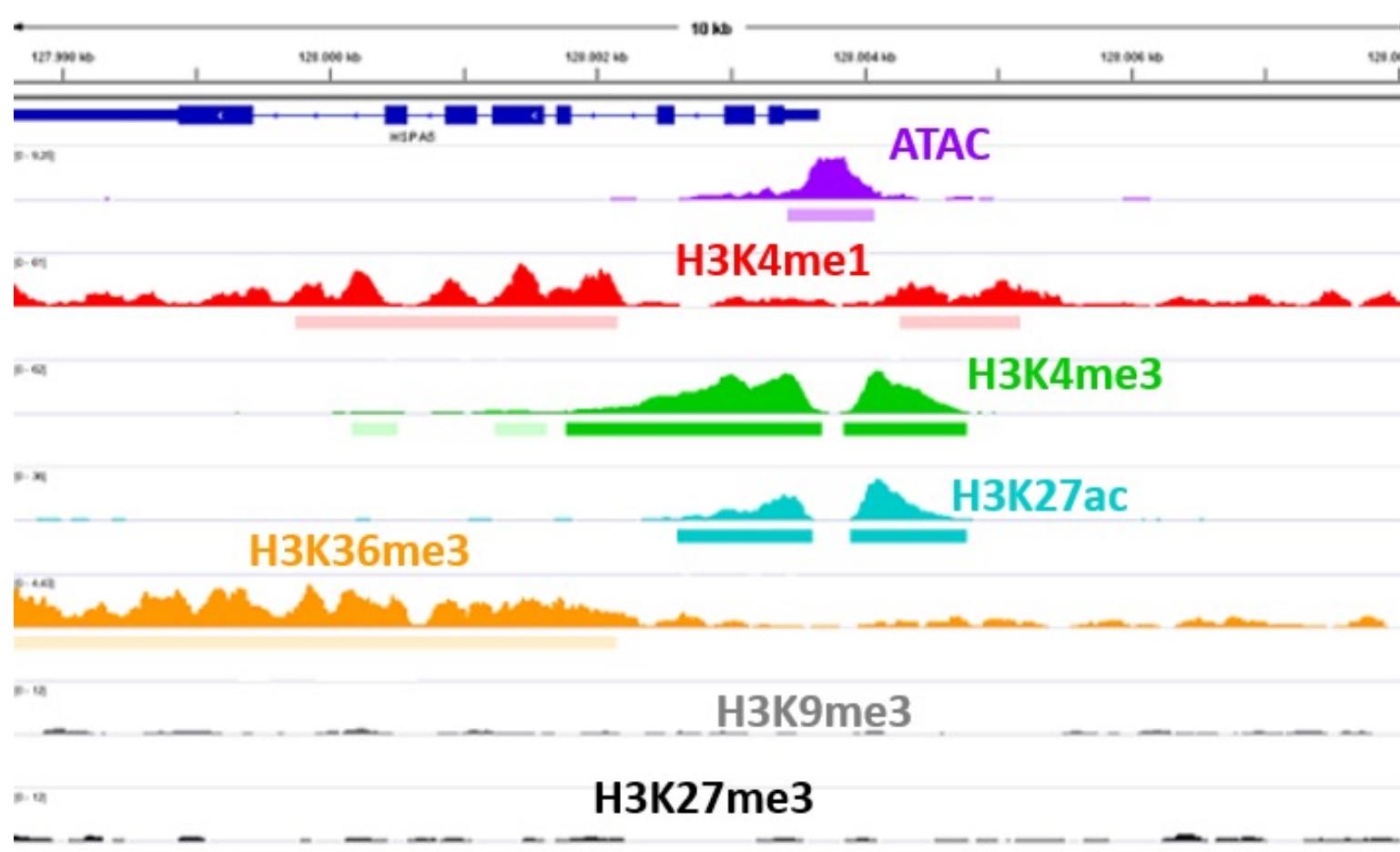


GHGA

Histone modification / TF binding analysis: ChIP-seq

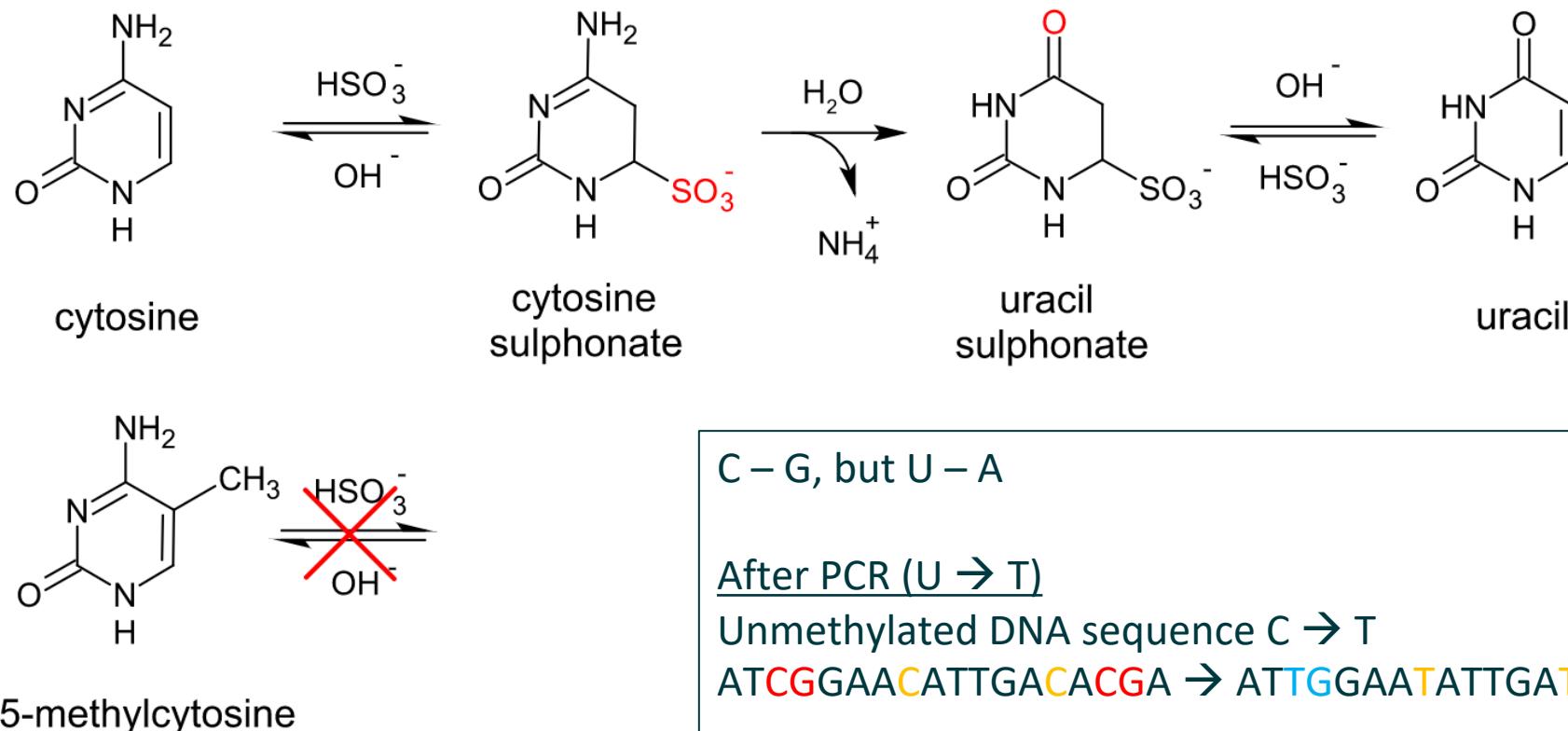


Typical ChIP-seq profile at active promoters



GHGA

DNA methylation analysis: Bisulfite conversion



C – G, but U – A

After PCR ($U \rightarrow T$)

Unmethylated DNA sequence C → T

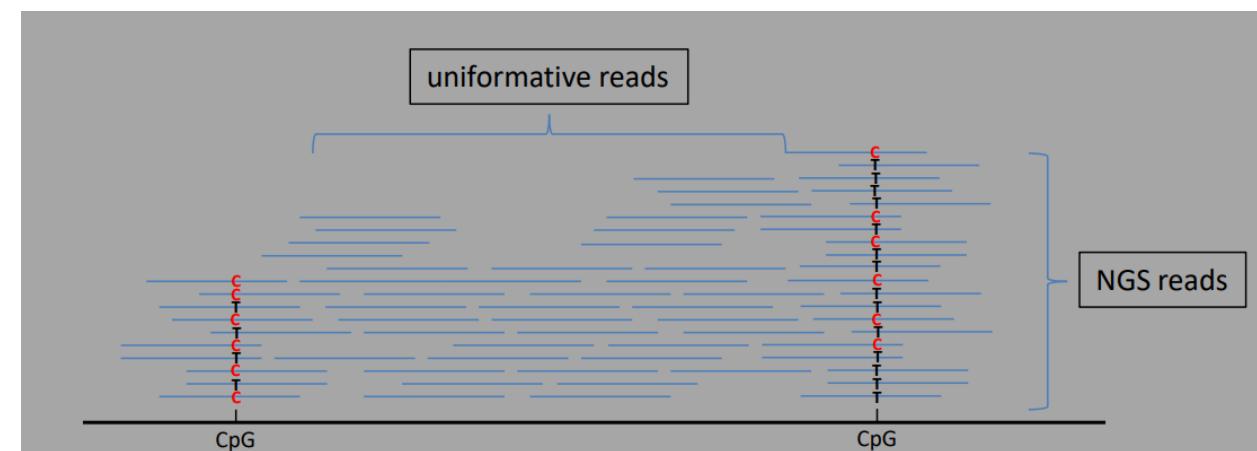
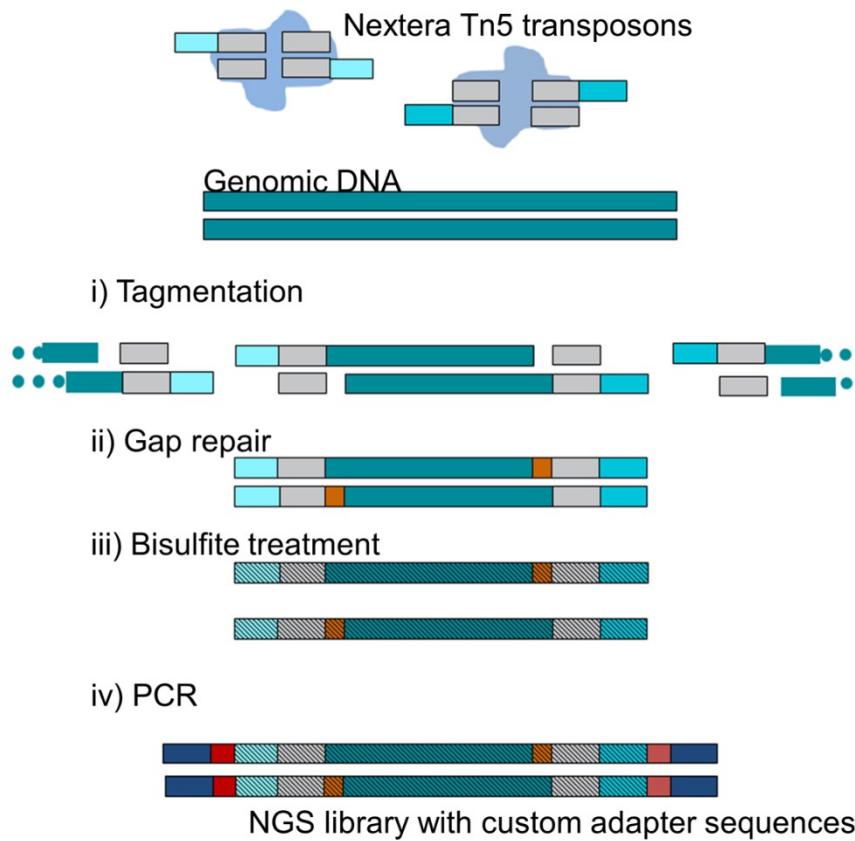
AT~~CG~~GAACATTGACAC~~CG~~A → AT~~T~~GGAATATTGATA~~T~~GA

Methylated DNA sequence → no conversion

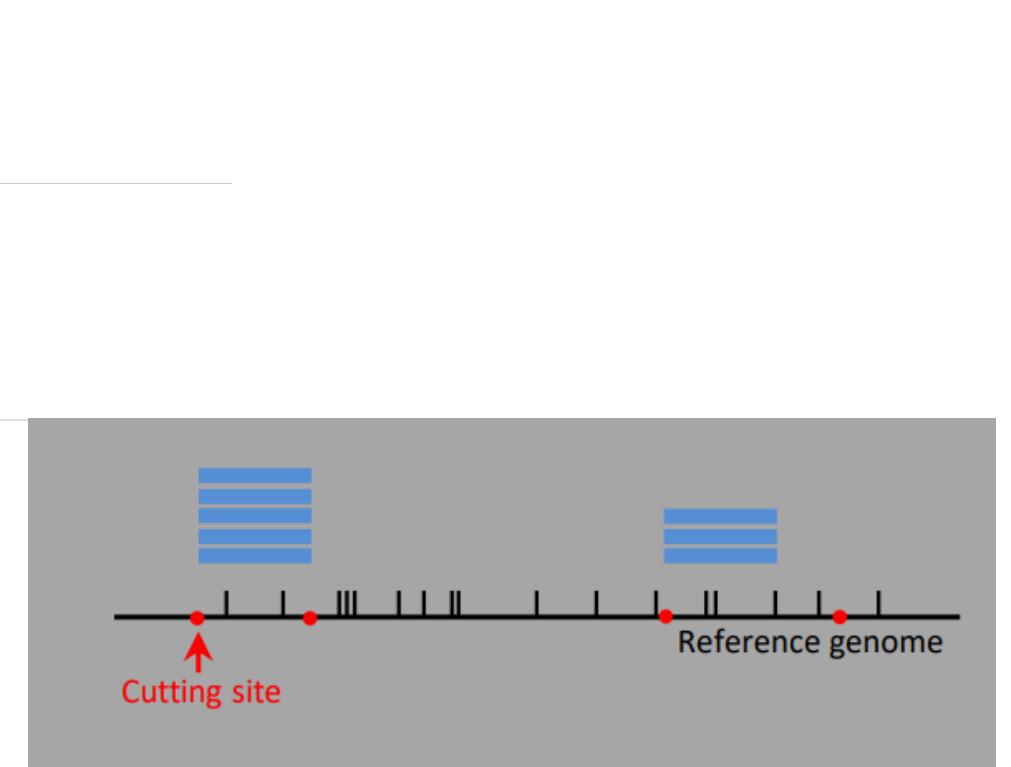
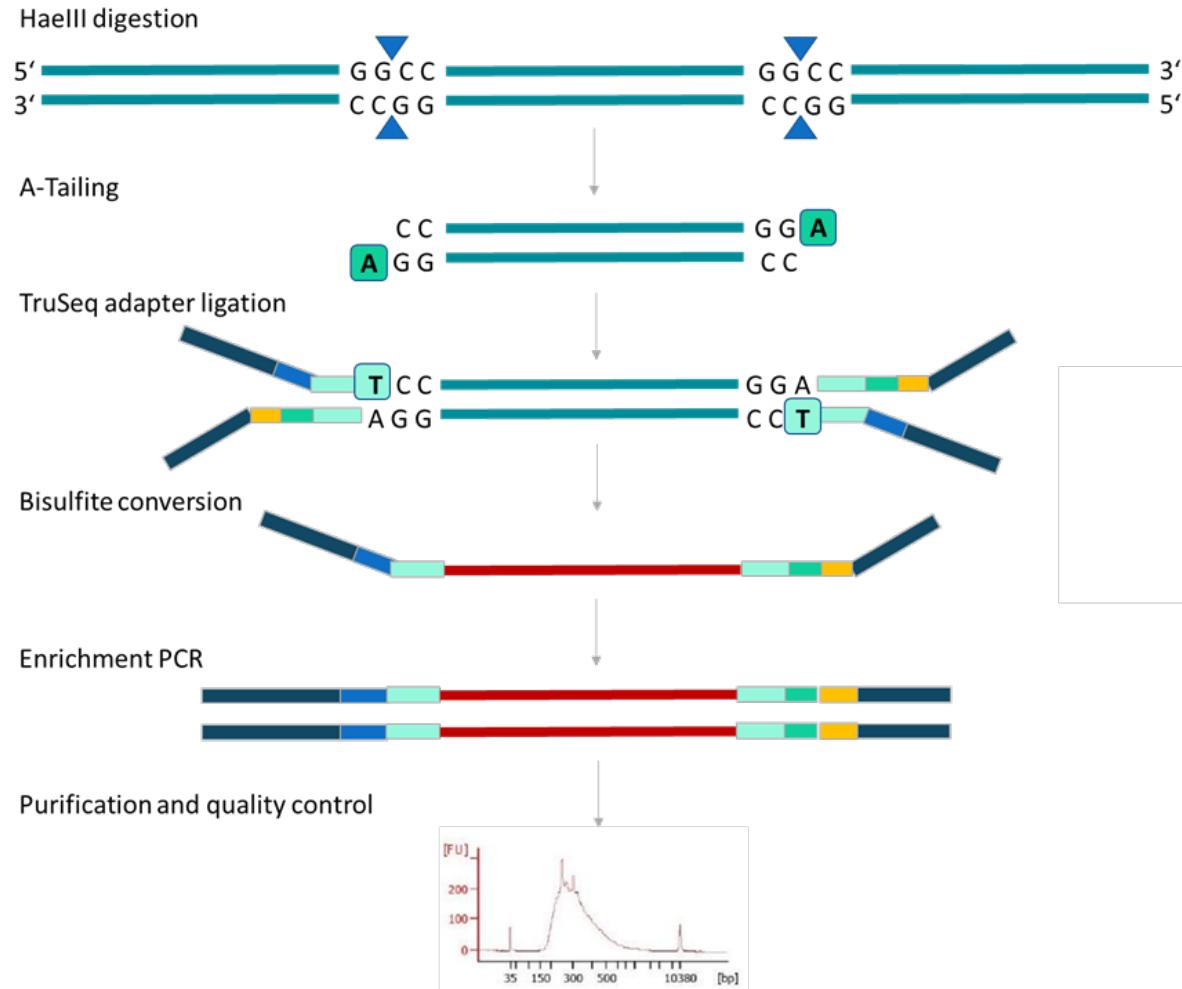
AT~~CG~~GAACATTGACAC~~CG~~A → AT~~CG~~GAATATTGATA~~CG~~A

DNA methylation analysis: Whole genome bisulfite sequencing (WGBS)

GHGA

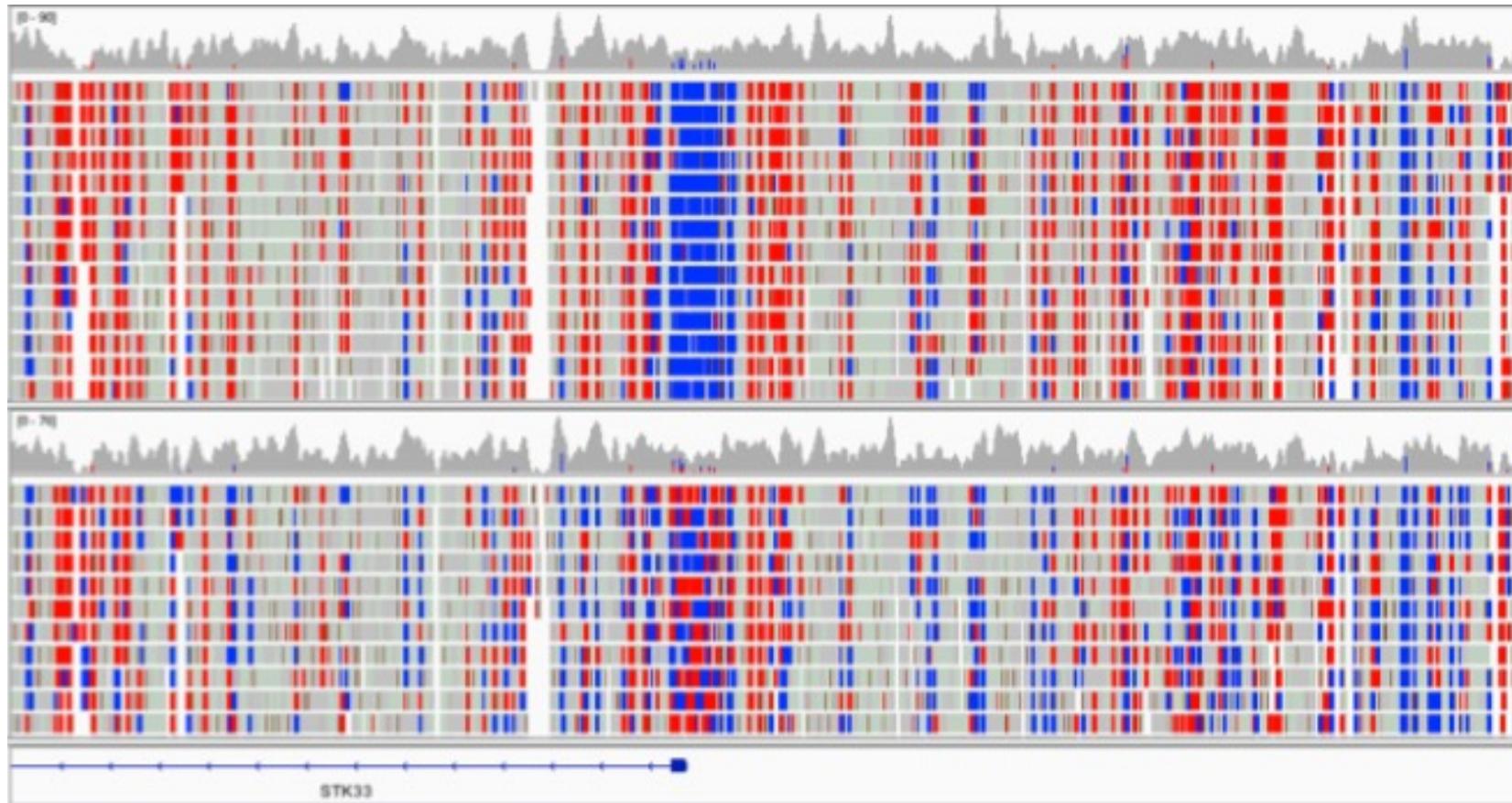


DNA methylation analysis: Reduced Representation Bisulfite Seq (RRBS)



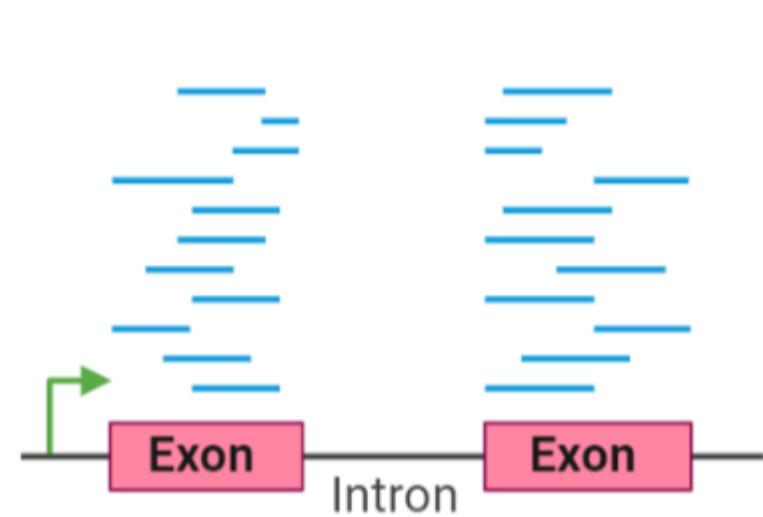
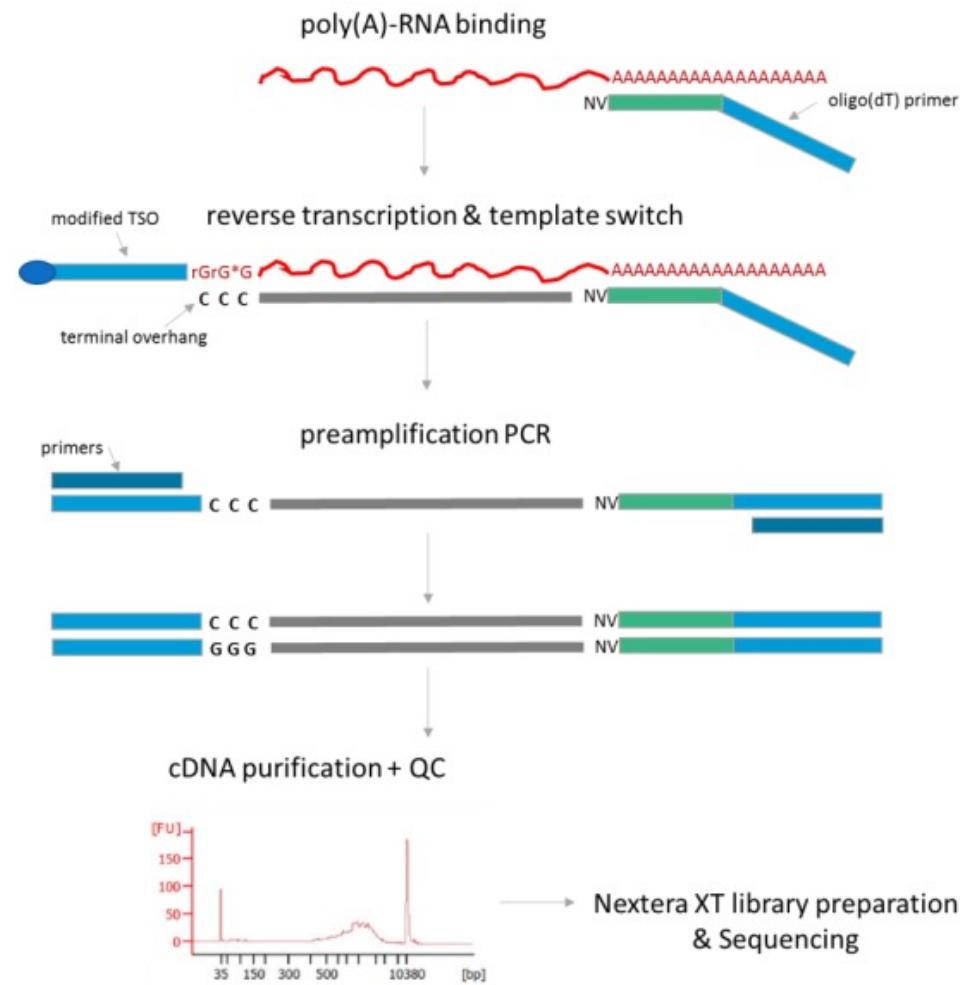
GHGA

DNA methylation profiles



GHGA

RNA-seq → functional read-out of the epigenetic landscape



GHGA

- GOfuncR:** The Bioconductor (R) package for GO analyses, which provides an R interface to FUNC. This program has options that allow you to work with pretty much any method, so you can remove all those pesky limitations that come with a basic GO run, like genomic background and gene length.
- REViGO:** Shorten your long list of Gene Ontology terms by removing redundant ones, so you can get a systems level perspective that isn't driven by the top term.
- Enrichr:** Find out about transcriptional regulation, pathways, ontologies, and much more from this neat little web tool.
- ConsensusPathwayDB:** A master tool that pulls from a large number of databases to provide ontology and pathway analysis for humans, mice, and yeast.
- GREAT:** Genomic Regions Enrichment of Annotations Tool (GREAT) gives biological context to non-coding genomic regions by analyzing the annotations of the nearby genes. It's great for analyzing genomic coordinates from your ChIP-seq and DNA methylation data.
- WGNA:** an R package for weighted correlation network analysis that can be used to find correlated gene clusters.
- Gene Set Enrichment Analysis (GSEA):** The name says it all, this pioneering program lets you compare against the Molecular Signatures Database (MSigDB).
- The Database for Annotation, Visualization and Integrated Discovery (DAVID):** The granddaddy of them all.
- STRING:** STRING is a database of known and predicted protein interactions that lets you visualize interacting networks.
- GeneMANIA:** GeneMANIA lets you visualize your gene lists and finds other related genes by using a very large set of functional association data.
- gProfiler:** Ontologies, pathways, and more from your gene list. Currently available for 200+ species.
- ToppGene:** A suite of tools to see what is enriched for in your gene list.

Motifs:

- Homer:** Discover motifs critical to the differences between your sample groups. The art and jokes are just a bonus. Also, don't forget to check out the **MARGE R package**.
- Uso:** epigenetic Landscape In Silico deletion Analysis determines the transcription factors and chromatin regulators behind your gene list.
- The MEME Suite:** Motif Based Sequence Analysis Tools.
- MotifMotif:** A cell-type specific database with transcription factor binding site motifs and accompanying DNA Methylation profiles.
- Epigramp:** An analysis pipeline that predicts histone modification and DNA methylation patterns from DNA motifs. Check out coverage of Epigramp.
- ePOSSUM:** Analyze your gene list for transcription factor binding sites from a number of species.
- Classification of Human Transcription Factors:** A large database that classifies human transcription factors.

Other Useful Tools:

- FORGE2:** Use chromatin state data to analyze your GWAS data
- GAT:** Genomic Association Tester (GAT) lets you compute the significance of the overlap between all your fancy data sets.
- GeneOverlap:** A Bioconductor (R) package to statistically test and then visualize gene overlaps between multiple gene lists.
- LOLA:** A Bioconductor (R) package that lets you test your genomic coordinates for enrichments in a large variety of internal and external datasets.
- COCOA:** A Bioconductor (R) package for understanding epigenetic variation among samples. It works with genomic coordinates, such as those from DNA methylation and chromatin accessibility data.
- nga.plot:** Visualize your results at functional genomic regions.
- ZENBU:** Japanese for all, entire, whole, altogether. This browser lets you integrate and interact with your data in a nice visual environment.
- CTCF Insulator Database:** *In silico* prediction for all your genomic insulation needs!
- EpiExplorer:** Import your very own data and compare it to ENCODE.
- Podbat:** A positioning database and analysis tool that takes data from a number of sources and to allow for the detailed dissection of a range of chromatin modifications. It can be used to analyze, visualize, store and share your data. Check out coverage on Podbat.
- BioWardrobe:** Lets you store, visualize, analyze and integrate epigenomic and transcriptomic data using a web-based graphical user interface that doesn't require programming expertise.
- Galaxy:** Provides an interface to help you with all the fancy code needed for genomic and transcriptomic analyses.
- Babulomics 5:** A user-friendly interface for a suite of tools for gene expression and genomic data.

<https://epigenie.com/epigenetic-tools-and-databases/>

Epigenetic data processing

Chromatin:

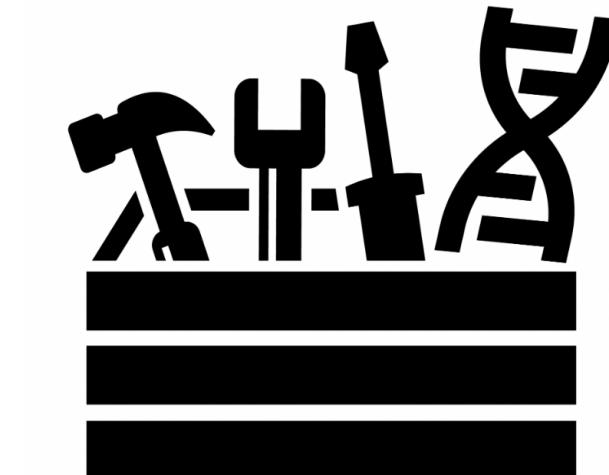
- nf-core/chipseq:** A pipeline used for Chromatin Immunoprecipitation sequencing (ChIP-seq) data built using Nextflow, a workflow tool to run tasks across multiple compute infrastructures in a very portable manner.
- BWAL:** The ChIP-seq aligner.
- MACS:** Model-based Analysis of ChIP-seq (MACS) is a go to peak-finding algorithm.
- deepTools:** A suite of python tools that tackle a lot of that complex ChIP-seq pipeline.
- DESeq2:** A Bioconductor (R) package for detecting differential peaks in your ChIP-seq data.
- EdgeR:** A Bioconductor (R) package for detecting differential peaks in your ChIP-seq data.
- PAVIS:** PAVIS (Peak Annotation and Visualization) lets you annotate and visualize your ChIP-seq and BS-seq data.
- EaSeq:** Lets you analyze and visualize your ChIP-seq data with graphical user interface that runs on a typical PC.
- ALEA:** Lets you analyze ChIP-seq or RNA-seq data to correlate allele-specific differences with epigenomic status.
- CENTDIST:** A web-application that identifies transcription factors hanging around your ChIP-seq peaks.
- ChIP-Array v2.0:** Integrate your ChIP-seq or ChIP-CHIP data with gene expression to build a regulatory network. Works for human, mouse, yeast, fly, and anabiosis data.
- CoSBI:** The histone code, dare you crack it? Learn more about CoSBI from Epigenie.
- EpigenomiX:** A Bioconductor (R) package that lets you integrate your RNA-seq or microarray data with your ChIP-seq data. It lets you preprocess and create differential gene lists for both data sets.
- HMGan:** a tool to call peaks in ChIP-seq/ATAC-seq data generated from cancer cells. It corrects for GC-content bias and DNA copy number aberrations.
- HMGan-diff:** a tool to detect differential chromatin modifications in cancer ChIP-seq data with a correction for copy number aberrations.
- LILY:** A method to call super-enhancers in cancer cells with DNA copy number aberrations.

RNA:

- nf-core/museq:** A bioinformatic analysis pipeline used for RNA sequencing data built using Nextflow, a workflow tool to run tasks across multiple compute infrastructures in a very portable manner.
- STAR:** RNA-seq aligner that performs simultaneous read mapping and counting. While you can probably get away without adapter trimming in most cases with a local aligner like STAR, it's probably still worth the effort to remove the biases. So why not add a Trim Galore run to your pipeline?
- Kallisto:** A program for quantifying abundances of transcripts from RNA-seq data. Uses pseudoalignment to skip the alignment step.
- Sleuth:** A program for analysis of RNA-Seq experiments for which transcript abundances have been quantified with kallisto.
- Salmon:** A tool for quantifying the expression of transcripts using RNA-seq data.
- DESeq2:** A Bioconductor (R) package for detecting differential expression of transcripts in your RNA-seq data.
- EdgeR:** A Bioconductor (R) package for differential expression analysis of RNA-seq data.
- RNA22V2:** Get your miRNA targeting on with an unbiased algorithm that not only considers 3'UTR binding but also 5'UTR binding.
- DIANA Tools:** A suite of tools that include target prediction algorithms and experimentally verified miRNA targets.
- TargetScan:** Predicts miRNA targets by searching for conserved sites that match the seed of each miRNA. There are different version for humans, mouse, worm, fly, and fish.
- miRWalk:** miRNA binding sites within the complete sequence of a gene and a comparison of binding sites from 12 existing miRNA-target prediction programs.
- GCRMA:** Pre-processing algorithm for affymetrix arrays.
- Chainy:** a cross-platform web tool providing systematic pipelines and steady criteria to process RT-PCR data.

DNA Methylation:

- Pipelines and analyses can be so much work, so if you've got WGBS, RRBS, or enzymatic methyl-seq (EM-seq) data, why not check out these two repos by our very own Epigenetics Editor who made them with you mind:
 - DMRICH:** A R package and executable for the statistical analysis and visualization of differentially methylated regions (DMRs) from CpG count matrices (Bismark genome-wide cytosine reports). It primarily utilizes the `dmrseq` and `bsseq` algorithms and provides upstream pre-processing as well as downstream analyses and data visualization.
 - CpG_Me:** A whole genome bisulfite sequencing (WGBS) pipeline for the alignment and QC of DNA methylation that goes from raw reads (FastQ) to a CpG count matrix (Bismark genome-wide cytosine reports).
- methylKit:** If you love single CpG statistics, then this is the Bioconductor (R) package for you. It's focused on high-throughput bisulfite sequencing methods, such high-coverage WGBS, RRBS and its variants, target-capture methods, as well as ShmC protocols such as oxBS-Seq and TAB-Seq. All it needs are your Bismark aligned BAM files.
- RnBeads:** A Bioconductor (R) package for comprehensive analysis of DNA methylation data from Illumina Infinium arrays (450K and EPIC) and BS-seq. MeDIP-seq and MBD-seq are also supported after some external processing.
- MEIDPS:** A Bioconductor (R) package for methylated DNA immunoprecipitation (MeDIP) experiments followed by sequencing (MeDIP-seq).
- Minfi:** A Bioconductor (R) package for your Illumina Infinium arrays (450K and EPIC) that provides comprehensive analysis and takes cellular heterogeneity into account, after all variety is the spice of life.
- DMRcate:** A Bioconductor (R) package for DMR identification from the human genome using WGBS and Illumina Infinium array (450K and EPIC) data.
- eFORGE:** eFORGE is an EWAS analysis tool for your 450K and EPIC array data that identifies tissue or cell type-specific signal by analyzing for overlap with DNase I hypersensitive sites (DHSs). It provides both graphical and tabulated outputs. Its adjunct tool **eFORGE TF** lets you explore chromatin accessibility near array probes and calculate the significance of overlap with transcription factor binding sites.
- CHAMP:** A Bioconductor (R) package that offers QC/QA metrics and a number of normalization methods in order to identify DMRs and copy number variations in Illumina Infinium array (450K and EPIC) data.
- FEM:** A Bioconductor (R) package that offers integrative analysis of DNA methylation and gene expression data.
- coMET:** A Bioconductor (R) package for the visualisation of Epigenome-Wide Association Study (EWAS) from a genomic region perspective.
- ReptileR:** A Bioconductor (R) package for the analysis of enrichment-based DNA methylation data.
- ELMER:** Use DNA methylation array and gene expression data to discover the regulatory element landscape and transcription factor network.
- Wanderer:** An interactive viewer to explore DNA-methylation and gene expression data in human cancer
- Methylation plotter:** A dynamic Web tool for easy methylation data visualization



Epigenetic data processing: Nf-core pipelines

- Nextflow pipelines
- Docker / Singularity containers
- Reproducible
- Standardized
- Great documentation

GHGA



ATAC-seq data analysis → Data processing and downstream analysis



1. Raw read QC ([FastQC](#))
2. Adapter trimming ([Trim Galore!](#))
3. Choice of multiple aligners 1.([BWA](#)) 2.([Chromap](#)) 3.([Bowtie2](#)) 4.([STAR](#))
4. Mark duplicates ([picard](#))
5. Merge alignments from multiple libraries of the same sample ([picard](#))
 1. Re-mark duplicates ([picard](#))
 2. Filtering to remove: mitochondrial DNA / blacklisted regions.... ([SAMtools](#), [BEDTools](#))
 3. Alignment-level QC and estimation of library complexity ([picard](#), [Presea](#))
 4. Create normalised bigWig files scaled to 1 million mapped reads ([BEDTools](#), [bedGraphToBigWig](#))
 5. Generate gene-body meta-profile from bigWig files ([deepTools](#))
 6. Calculate genome-wide enrichment ([deepTools](#))
 7. Call broad/narrow peaks ([MACS2](#))
 8. Annotate peaks relative to gene features ([HOMER](#))
 9. Create consensus peakset across all samples and create tabular file to aid in the filtering of the data ([BEDTools](#))
 10. Count reads in consensus peaks ([featureCounts](#))
 11. Differential accessibility analysis, PCA and clustering ([R](#), [DESeq2](#))
 12. Generate ATAC-seq specific QC html report ([ataqv](#))
6. Merge filtered alignments across replicates ([picard](#)) → re-do step 5.
7. Create IGV session file containing bigWig tracks, peaks and differential sites for data visualisation ([IGV](#)).
8. Present QC for raw read, alignment, peak-calling and differential accessibility results ([ataqv](#), [MultiQC](#), [R](#))

ChIP-seq data analysis → Data processing and downstream analysis

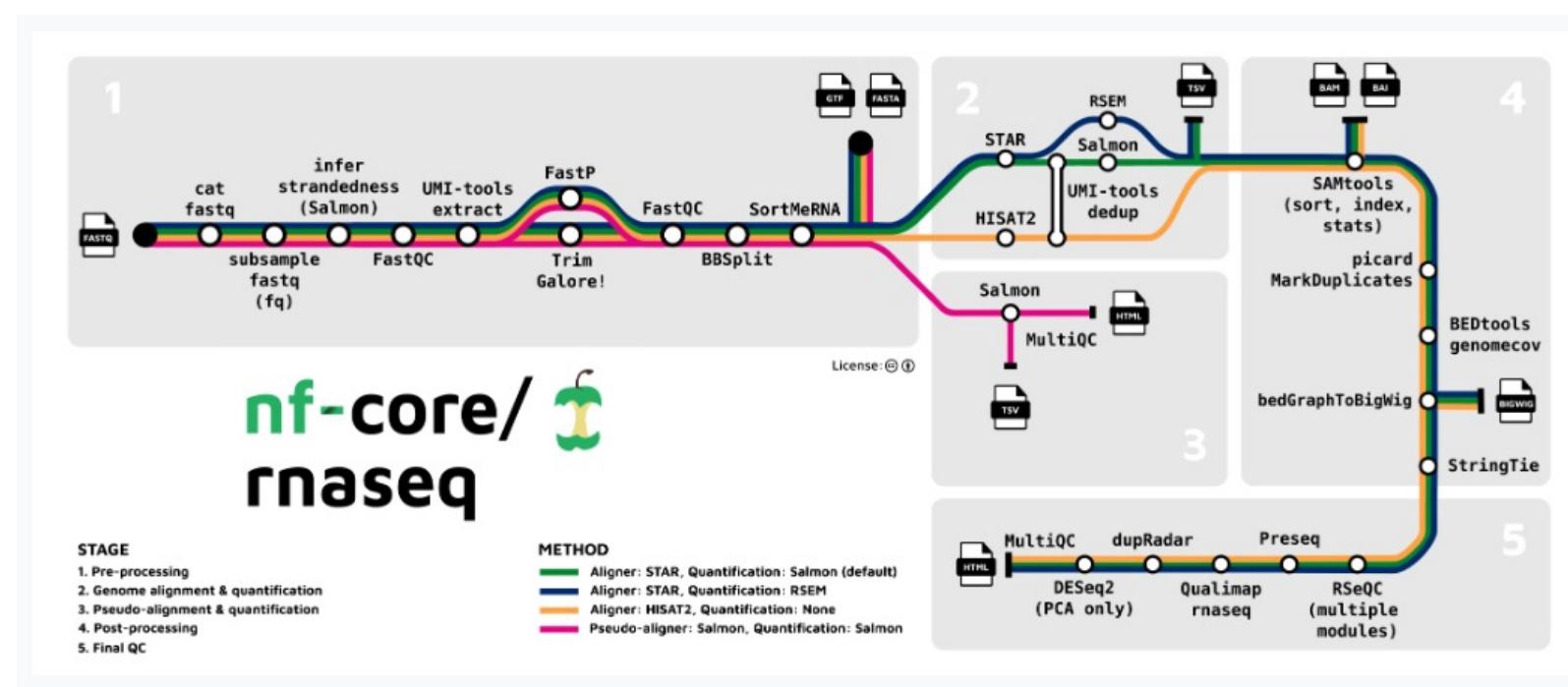


RNA-seq data analysis

→ Data processing and downstream analysis

nf-core/rnaseq

RNA sequencing analysis pipeline using STAR, RSEM, HISAT2 or Salmon with gene/isoform counts and extensive quality control.



DNA methylation data analysis → Data processing

nf-core/methylseq

Methylation (Bisulfite-Sequencing) analysis pipeline using Bismark or bwa-meth + MethylDackel

[bisulfite-sequencing](#) [dna-methylation](#) [methyl-seq](#)

Step	Bismark workflow	bwa-meth workflow
Generate Reference Genome Index (<i>optional</i>)	Bismark	bwa-meth
Merge re-sequenced FastQ files	cat	cat
Raw data QC	FastQC	FastQC
Adapter sequence trimming	Trim Galore!	Trim Galore!
Align Reads	Bismark	bwa-meth
Deduplicate Alignments	Bismark	Picard MarkDuplicates
Extract methylation calls	Bismark	MethylDackel
Sample report	Bismark	-
Summary Report	Bismark	-
Alignment QC	Qualimap	Qualimap
Sample complexity	Preseq	Preseq
Project Report	MultiQC	MultiQC

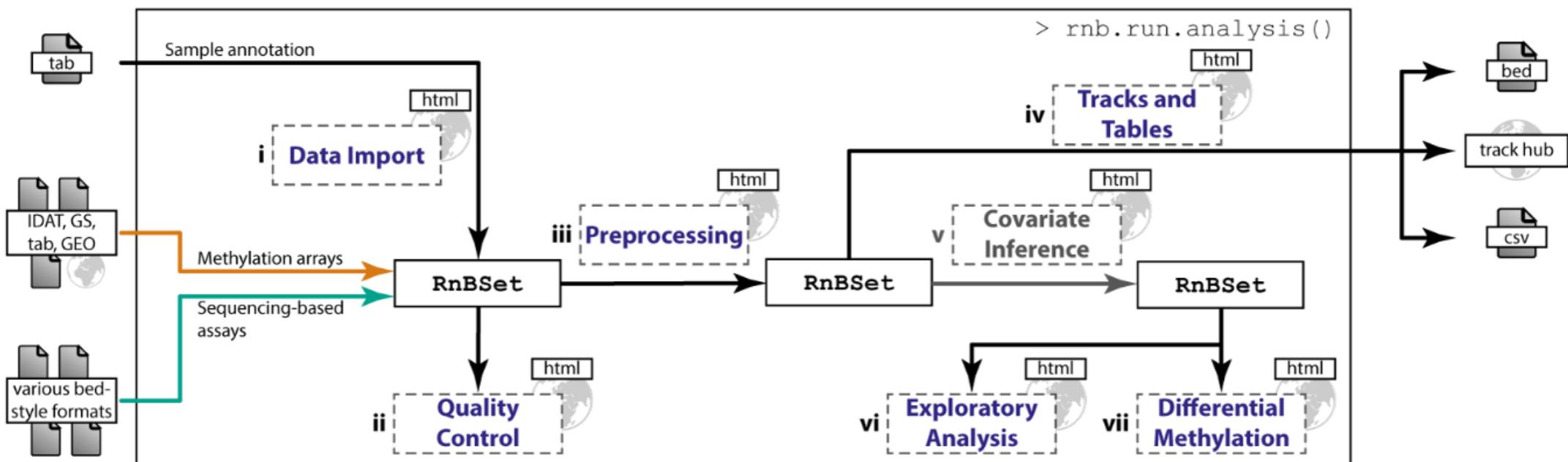
DNA methylation data analysis tools

→ Downstream analysis

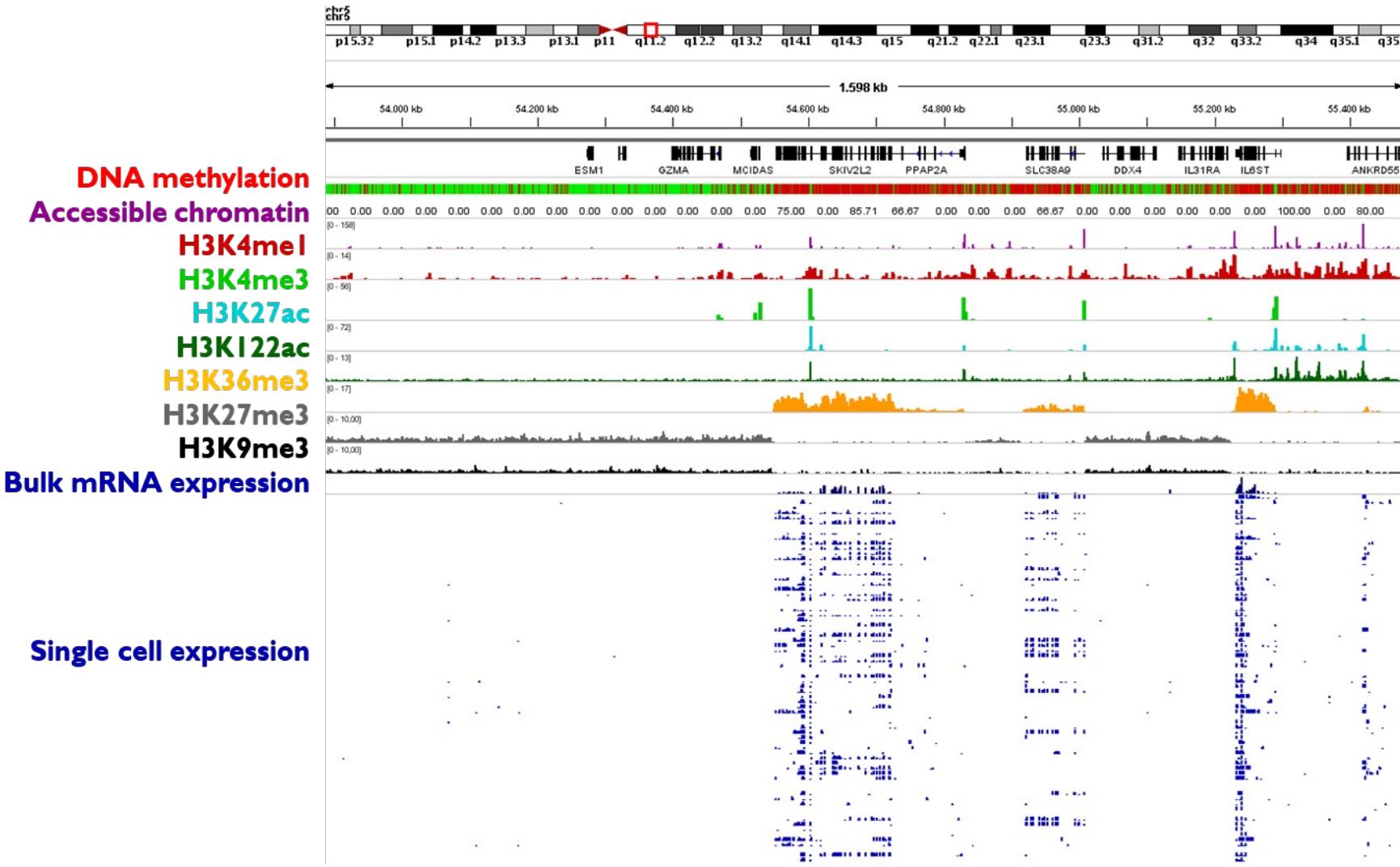


<https://rnbeads.org/>

GHGA



Interplay of epigenetic modifications



Resources – Part II

- <https://epigenie.com/epigenetic-tools-and-databases/>
- <https://nf-co.re/>
- <https://rnbeads.org/>



QUESTIONS?

Part II

A large, semi-transparent teal circle is positioned in the center-left area of the slide. It is surrounded by several small, scattered dots in various colors: red, blue, and black. These dots are arranged in a loose, circular pattern around the teal circle.

Thank you!