Swiss Institute of Bioinformatics

BIOLOGY-INFORMED MULTIOMICS DATA INTEGRATION AND VISUALIZATION

Multi-omics data

Deepak Tanwar

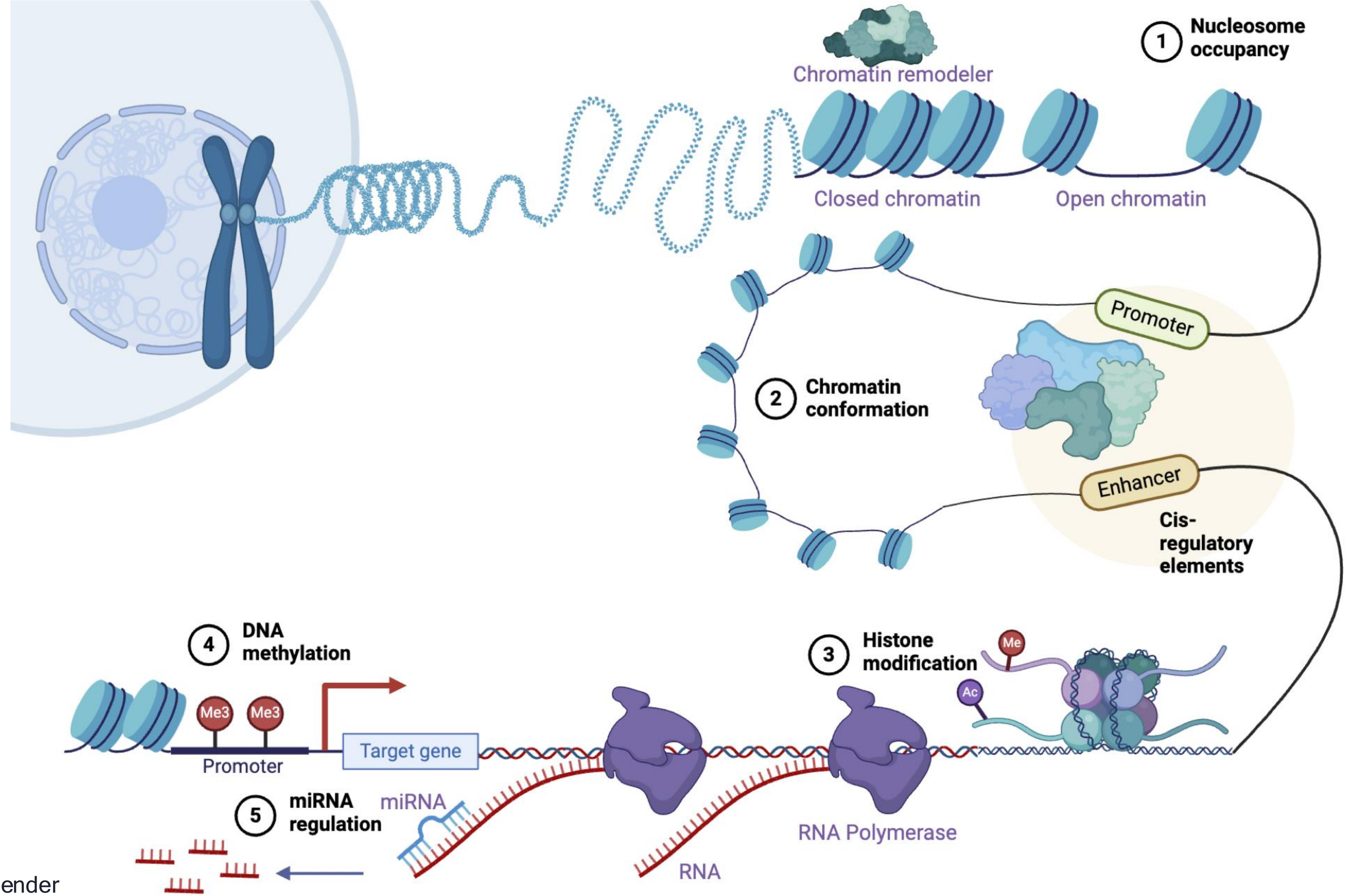
June 16-17, 2025

elixir SWITZERLAND

Learning outcomes

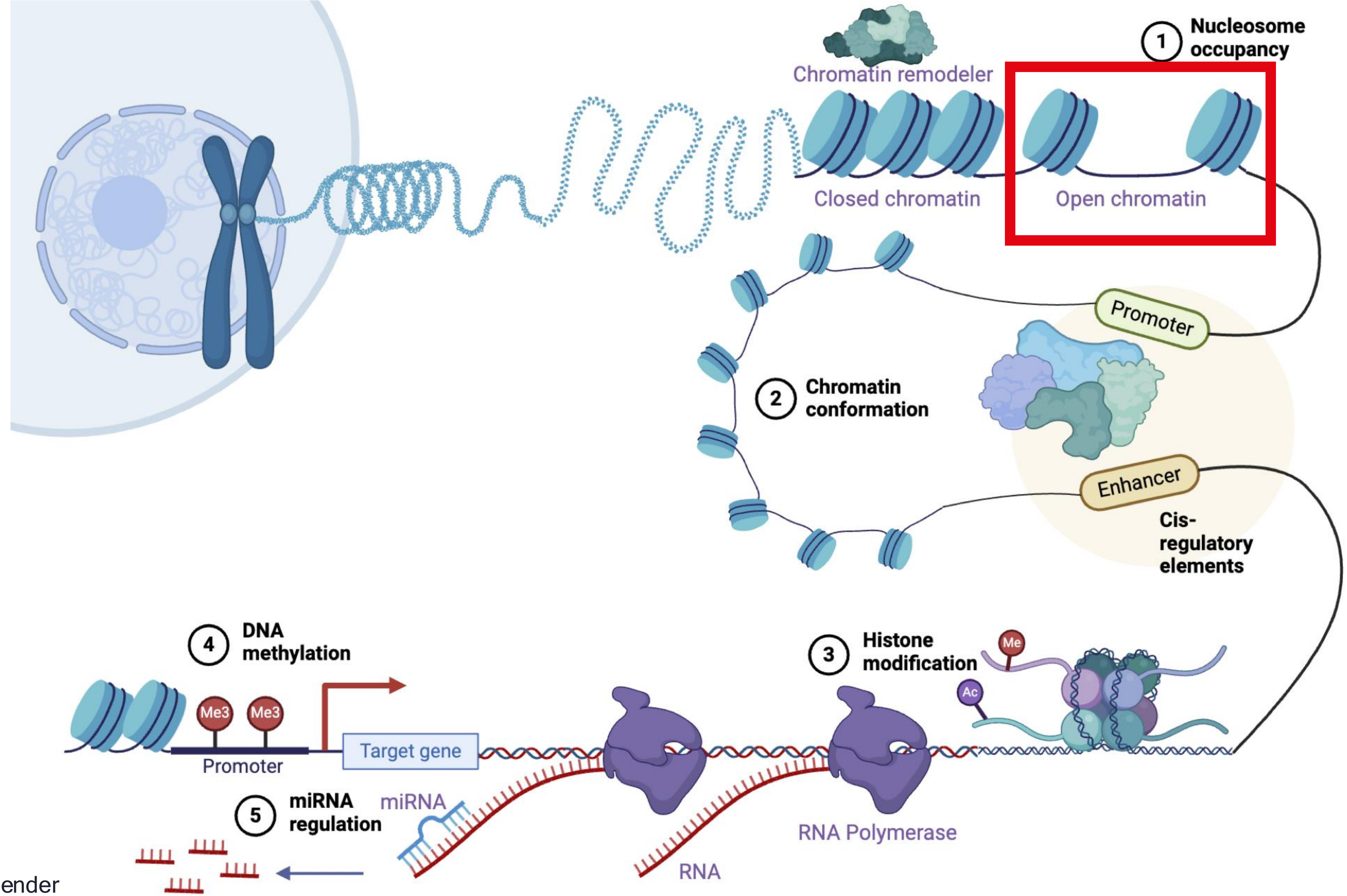
- Different layers of genome activity
- Basic pipeline for data generation
- Basic pipeline for the data processing

Simplified illustration of transcriptional regulation



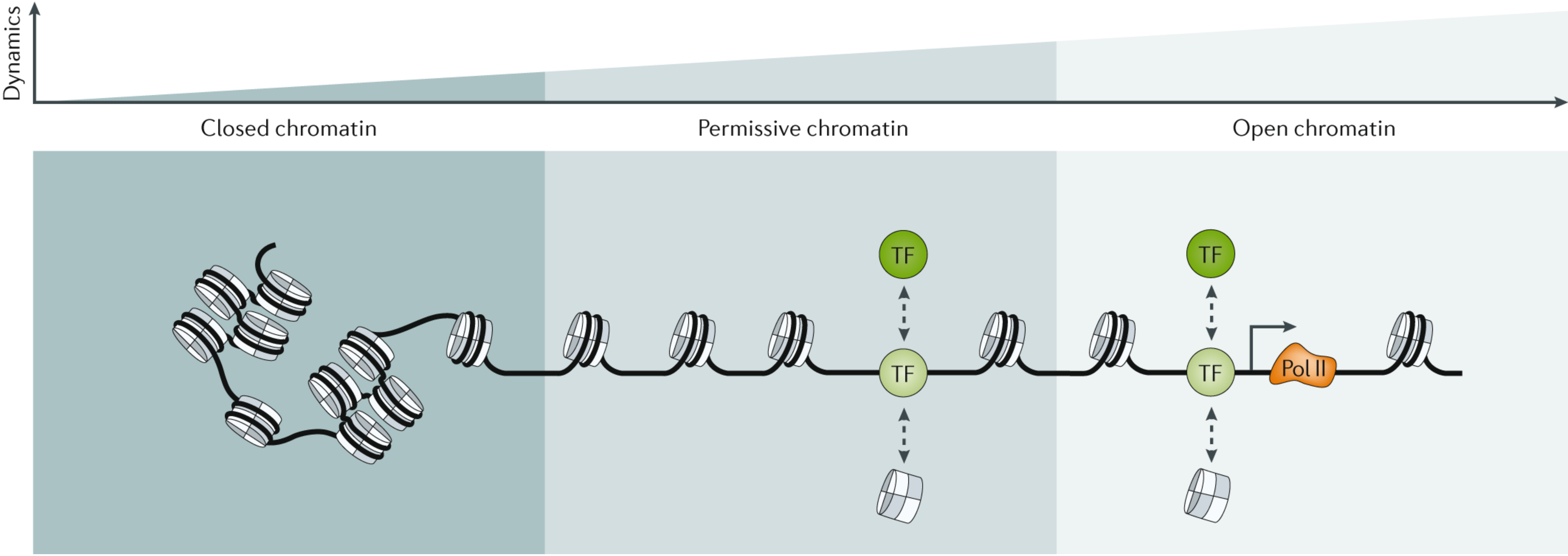
Generated using Biorender

Simplified illustration of transcriptional regulation



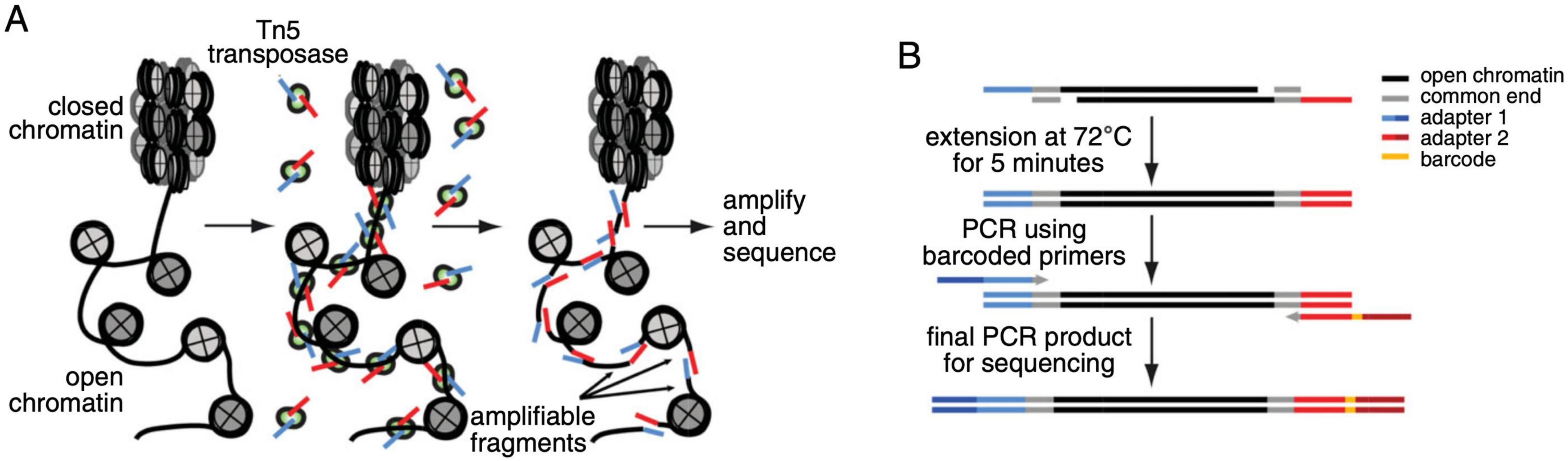
Generated using Biorender

Chromatin accessibility



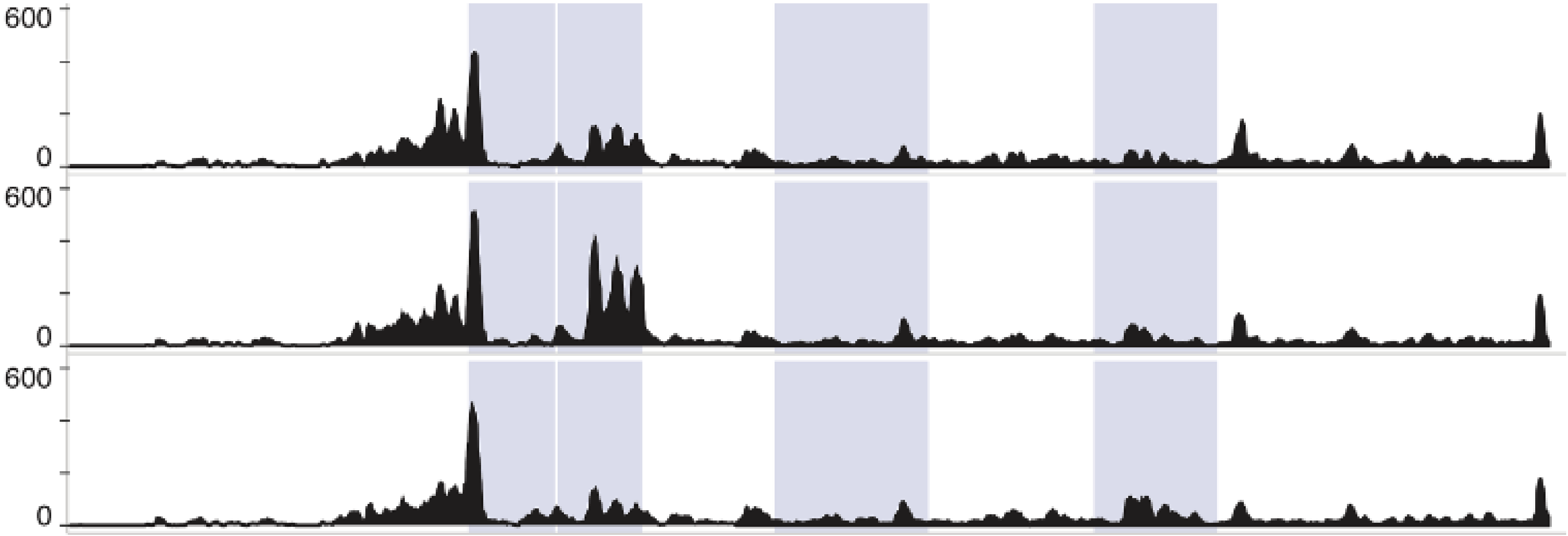
doi.org/10.1038/s41576-018-0089-8

Library preparation for ATAC-seq

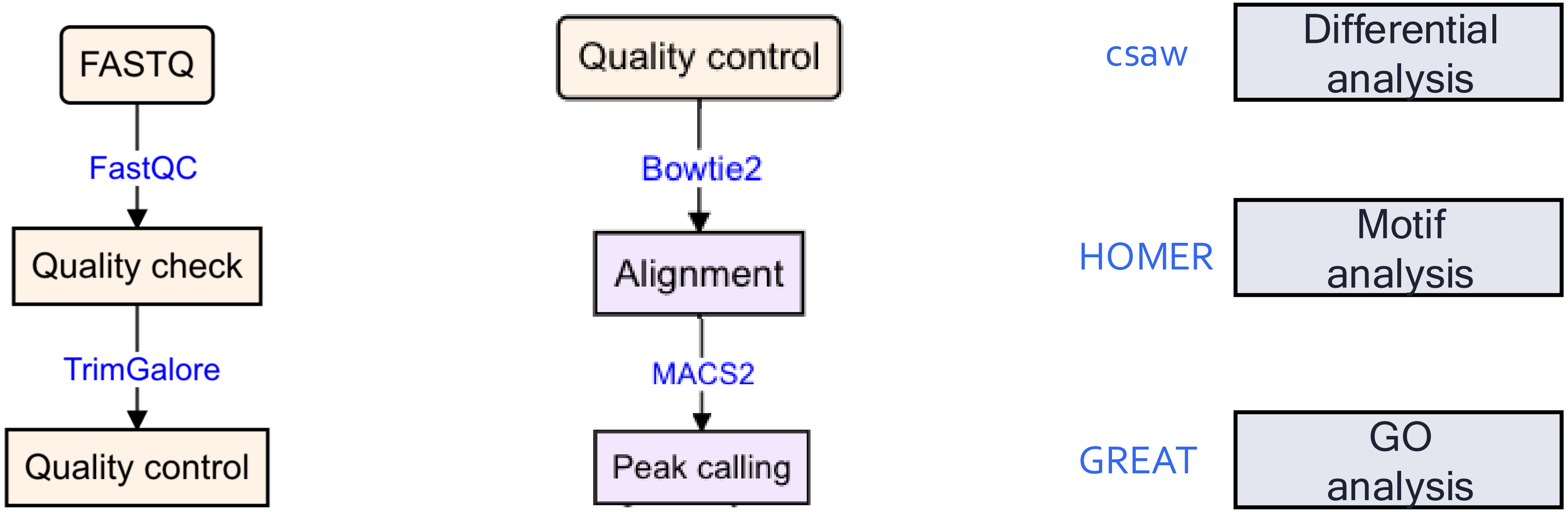


doi.org/10.1038/s41576-018-0089-8

Example snapshot of chromatin accessibility



ATAC-seq data analysis pipeline



FASTQ

FastQC

Quality check

TrimGalore

Quality control

Quality control

Bowtie2

Alignment

MACS2

Peak calling

csaw

Differential analysis

HOMER

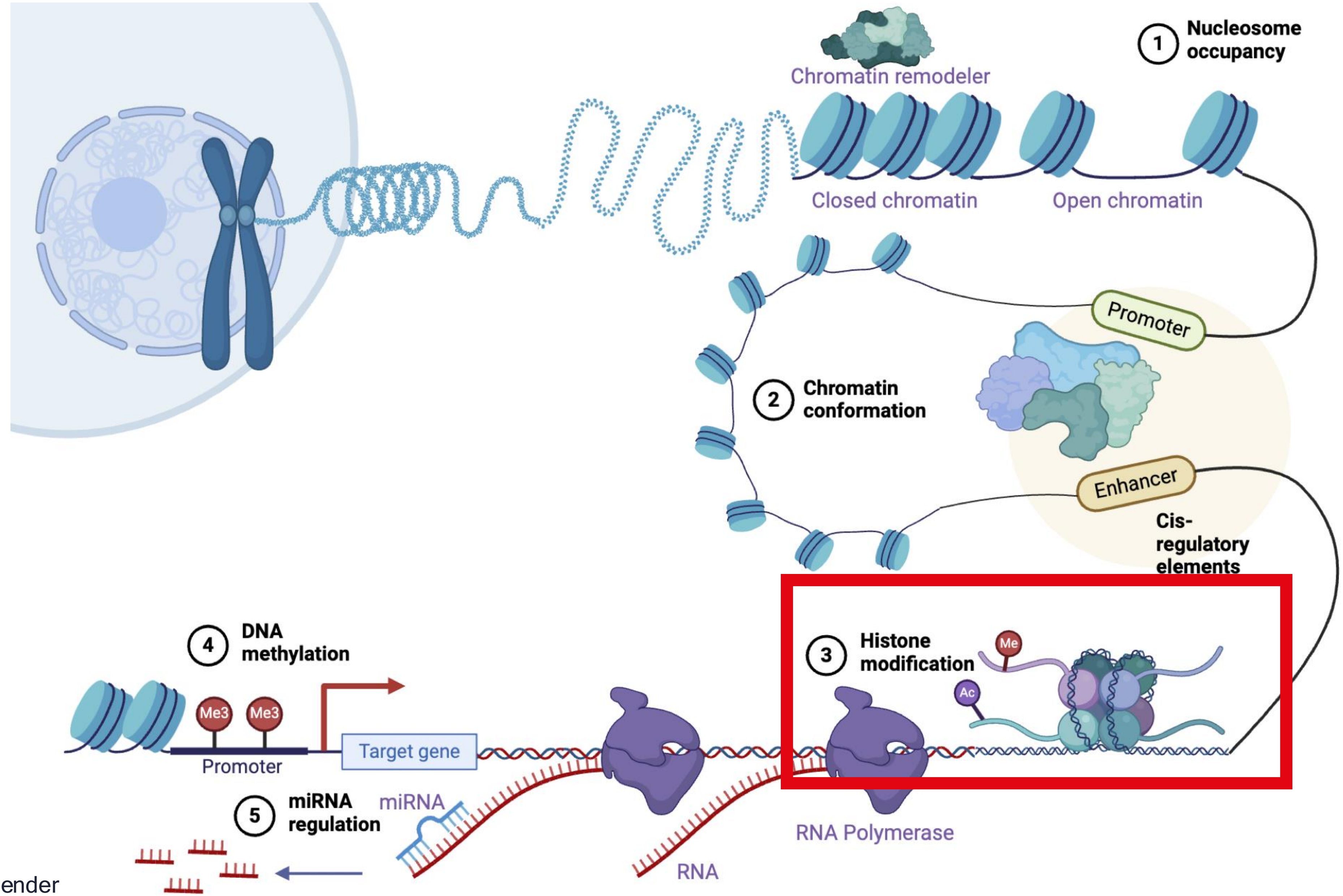
Motif analysis

GREAT

GO analysis

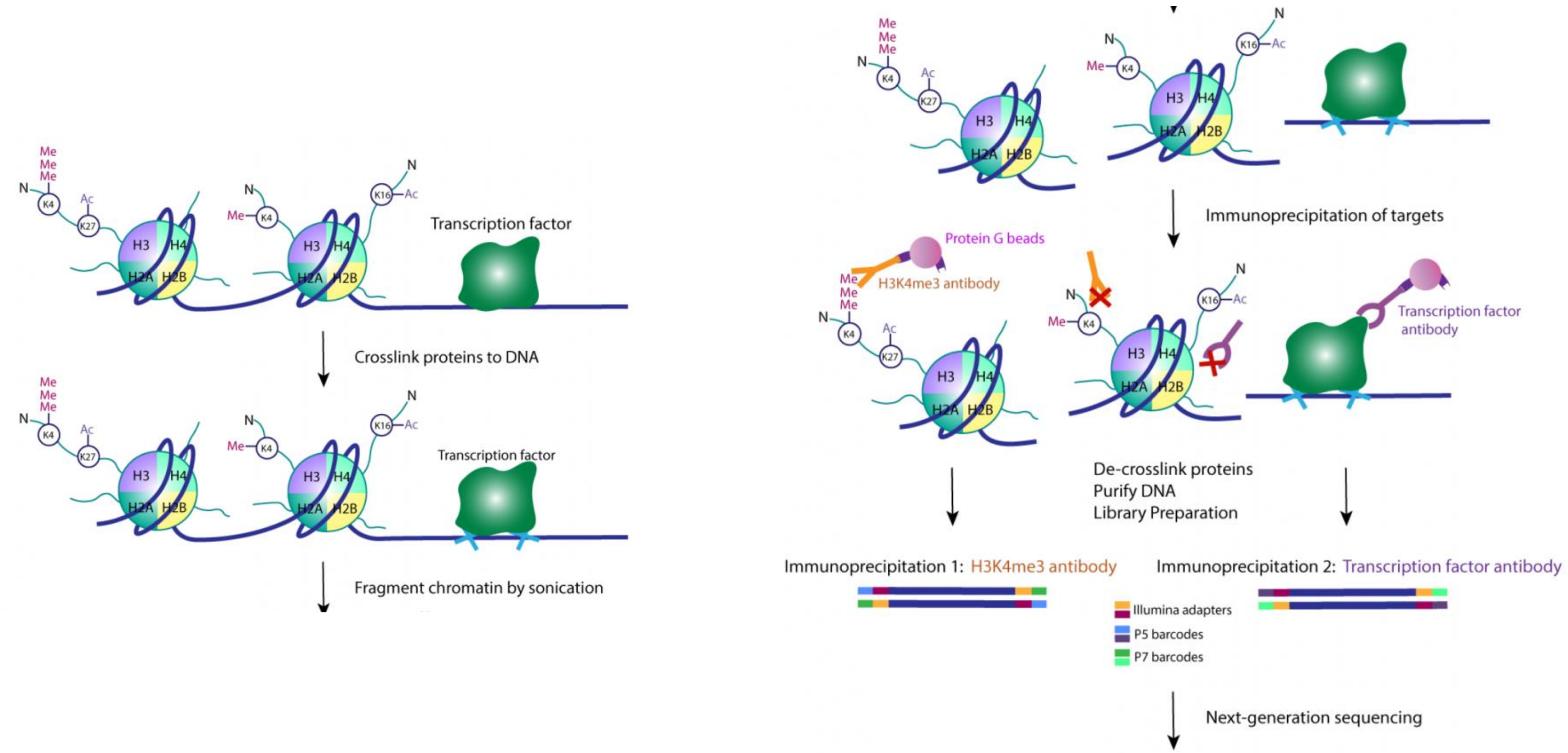
https://doi.org/10.3929/ethz-b-000537329

Simplified illustration of transcriptional regulation



Generated using Biorender

Chromatin Immunoprecipitation sequencing



Crosslink proteins to DNA

Fragment chromatin by sonication

Immunoprecipitation of targets

De-crosslink proteins
Purify DNA
Library Preparation

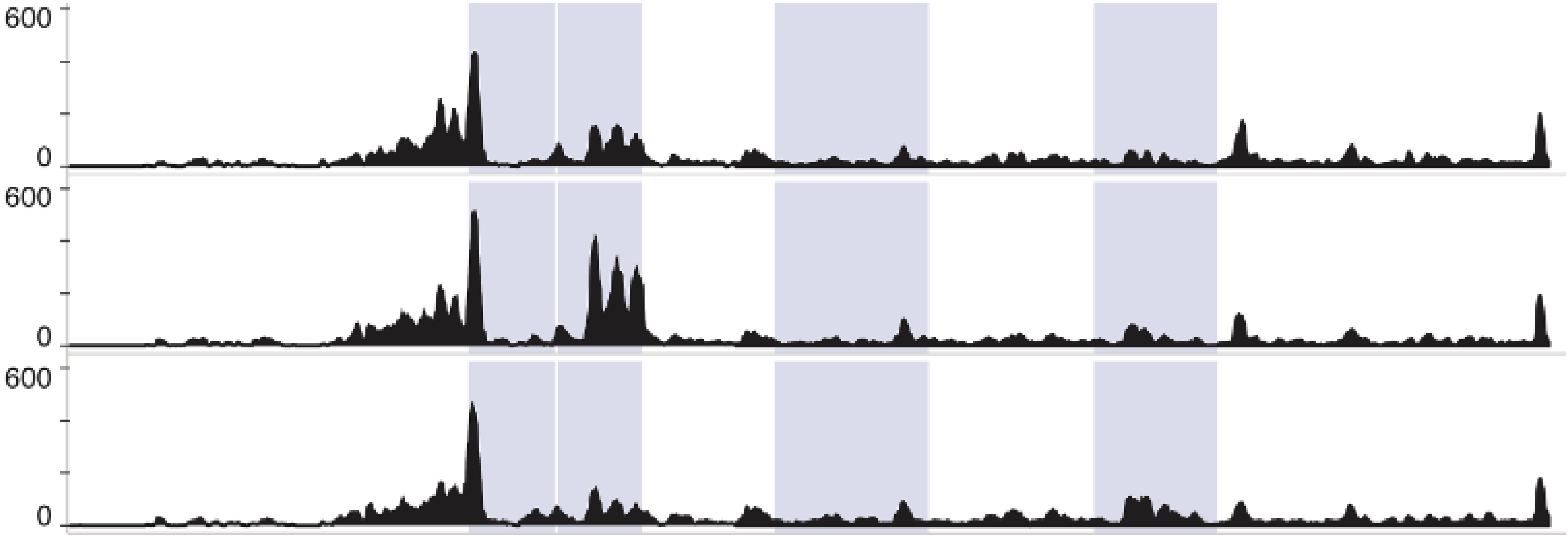
Immunoprecipitation 1: H3K4me3 antibody

Immunoprecipitation 2: Transcription factor antibody

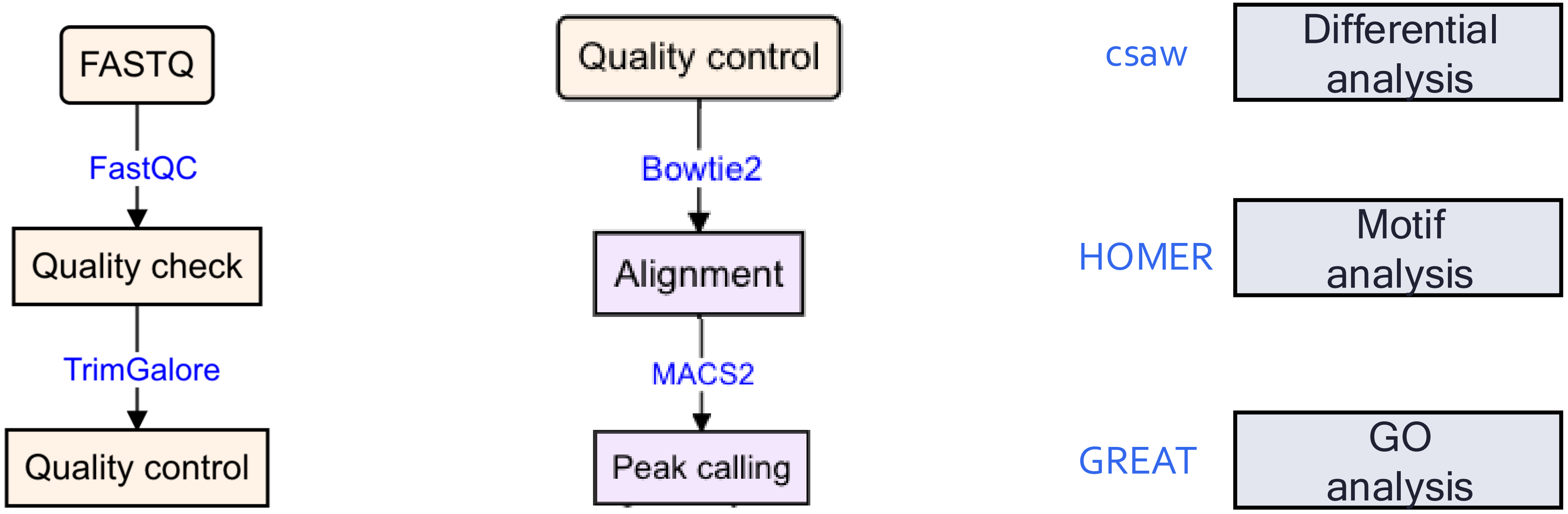
Next-generation sequencing

https://medresearch.umich.edu/office-research/about-office-research/biomedical-research-core-facilities/epigenomics-core/histone-modifications

Example snapshot of enrichment



ChIP-seq data analysis pipeline



FASTQ

FastQC

Quality check

TrimGalore

Quality control

Quality control

Bowtie2

Alignment

MACS2

Peak calling

csaw

Differential analysis

HOMER

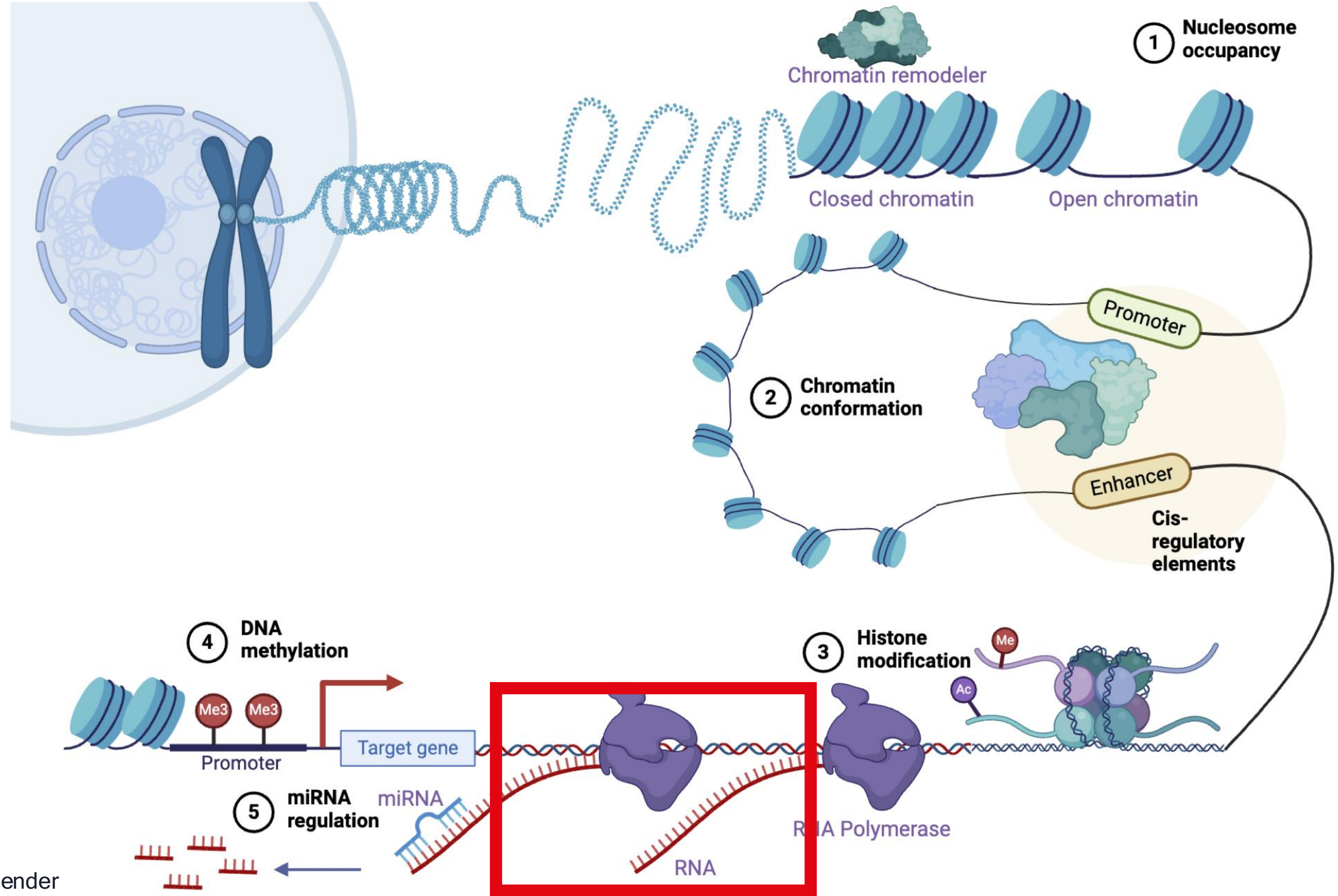
Motif analysis

GREAT

GO analysis

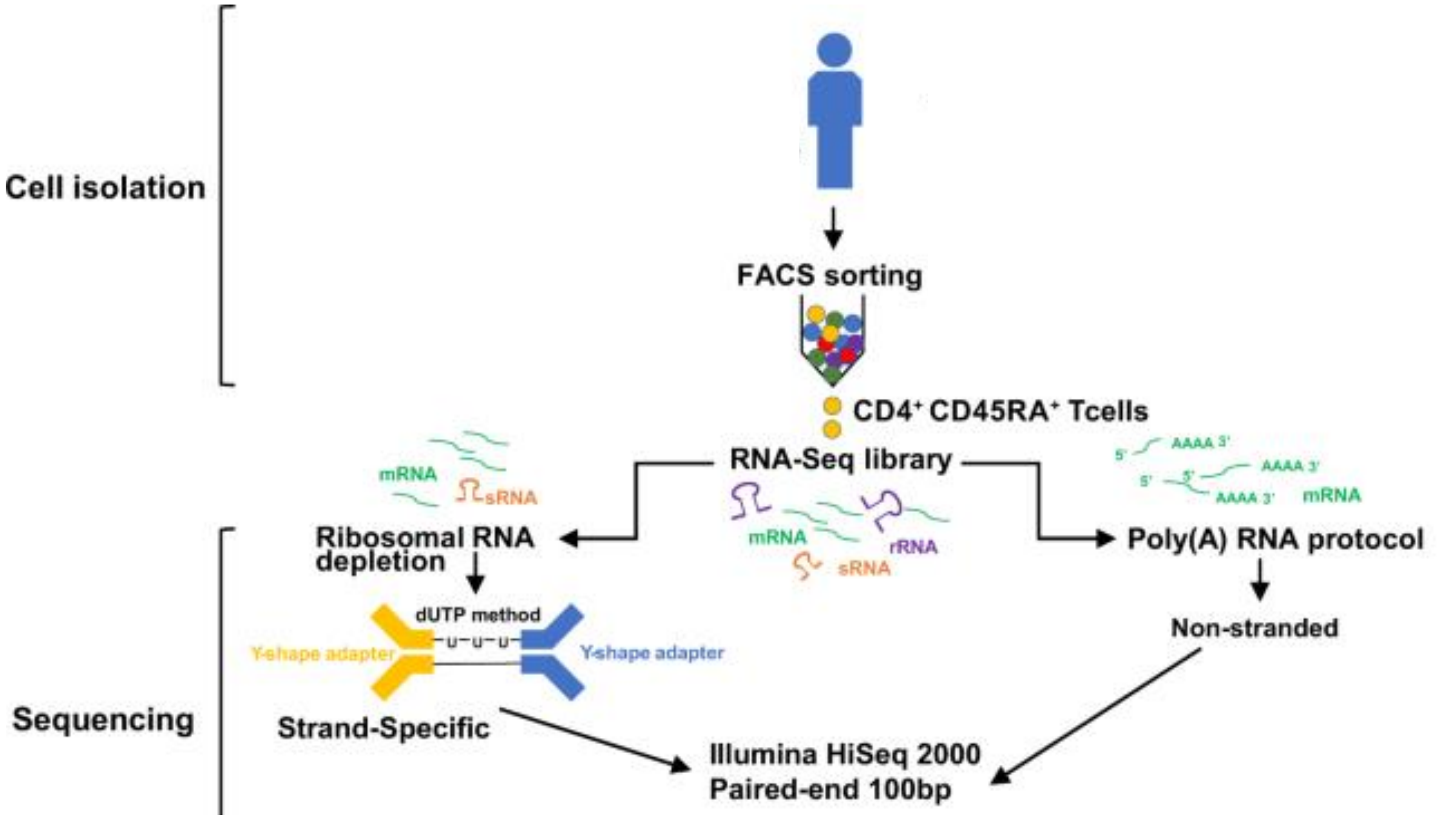
https://doi.org/10.3929/ethz-b-000537329

Simplified illustration of transcriptional regulation



Generated using Biorender

Gene expression



Cell isolation

FACS sorting

CD4⁺ CD45RA⁺ Tcells

RNA-Seq library

Ribosomal RNA depletion

Poly(A) RNA protocol

dUTP method

Y-shape adapter

Y-shape adapter

Strand-Specific

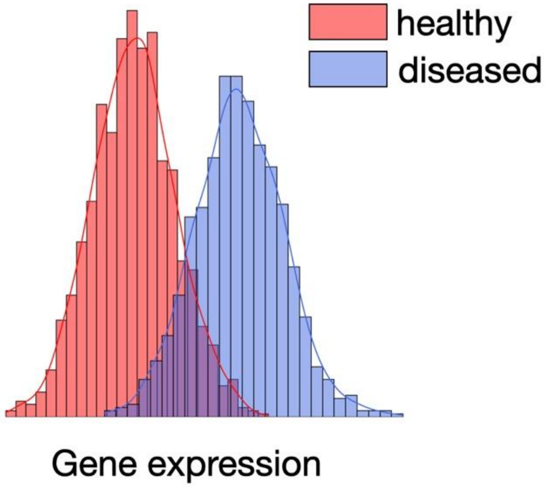
Non-stranded

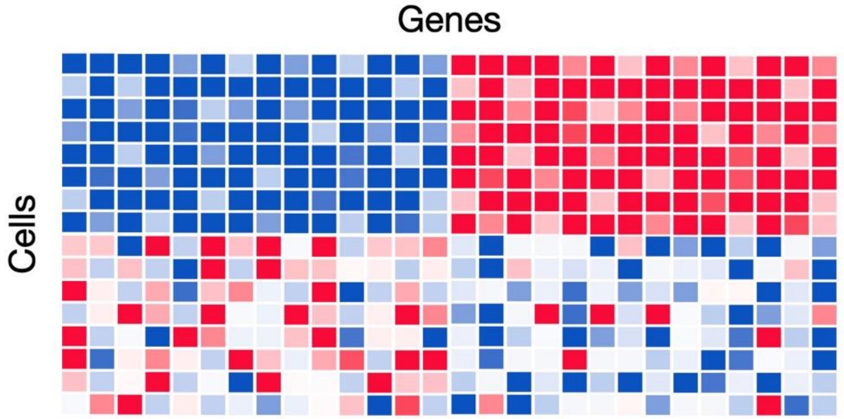
Illumina HiSeq 2000
Paired-end 100bp

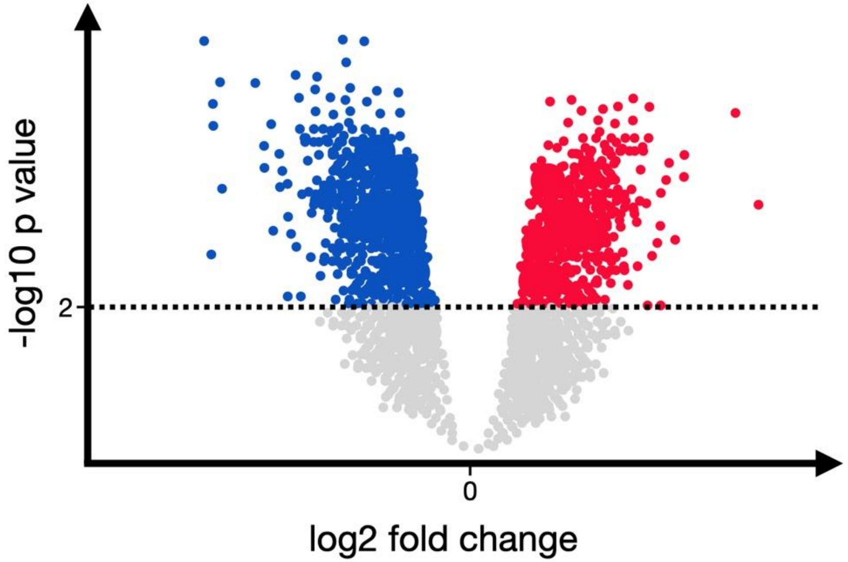
Sequencing

http://doi.org/10.1038/s41597-020-00719-4

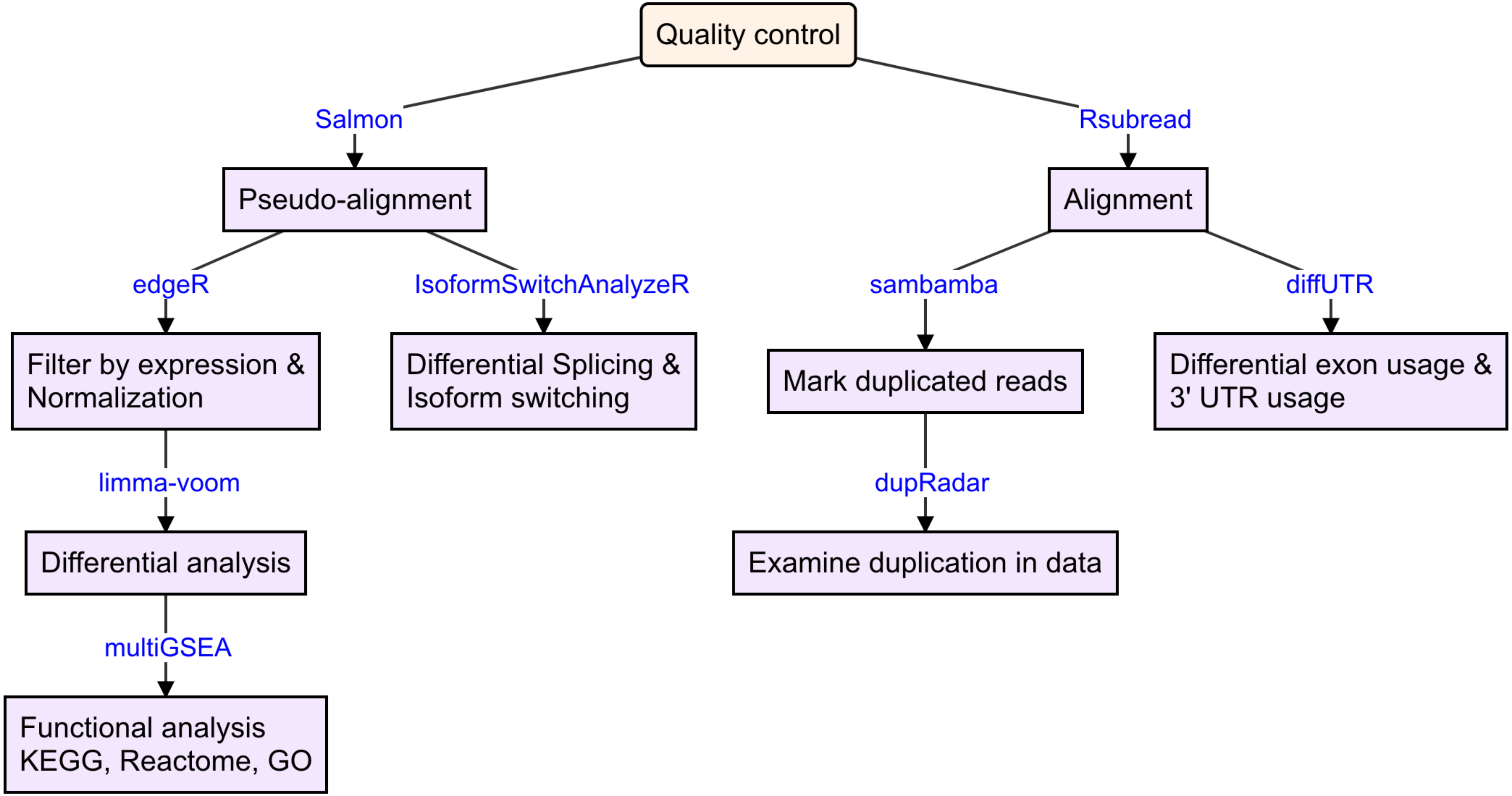
Differences in gene expression





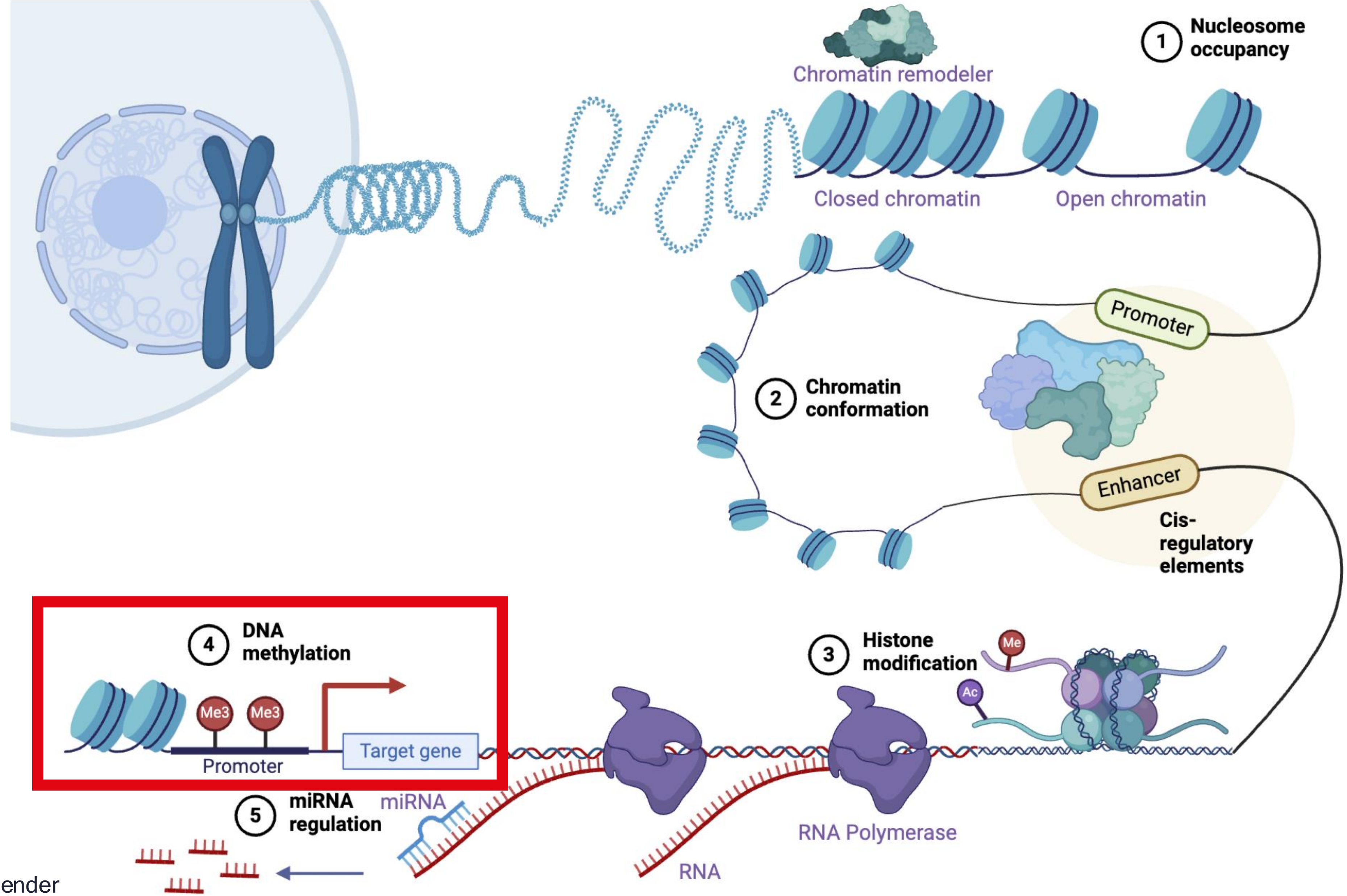


RNA-seq data analysis pipeline



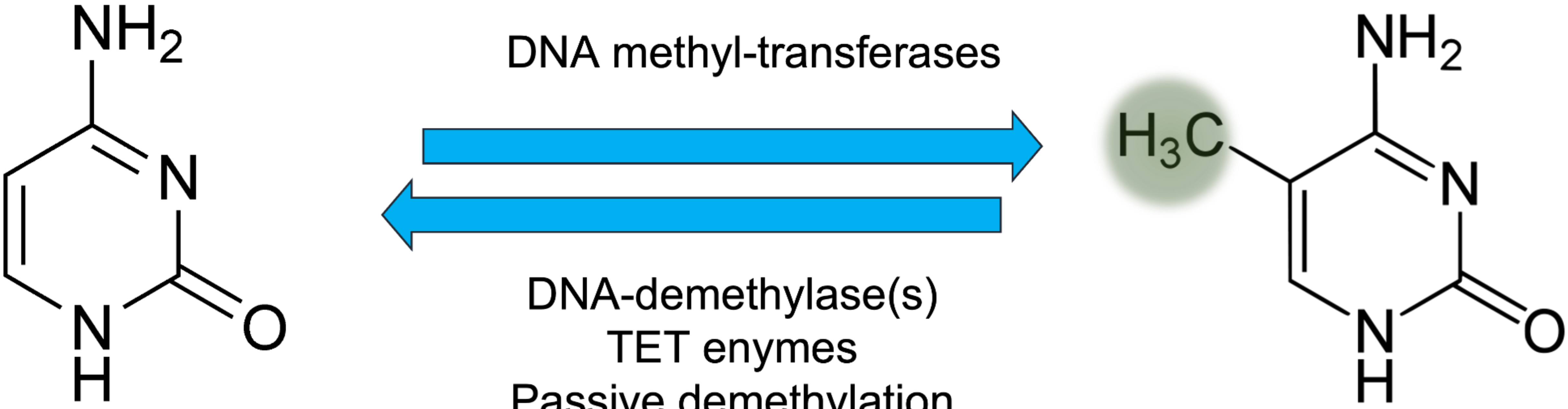
https://doi.org/10.3929/ethz-b-000537329

Simplified illustration of transcriptional regulation



Generated using Biorender

DNA cytosine methylation is a biological process by which methyl groups are added to the DNA molecule



DNA methyl-transferases

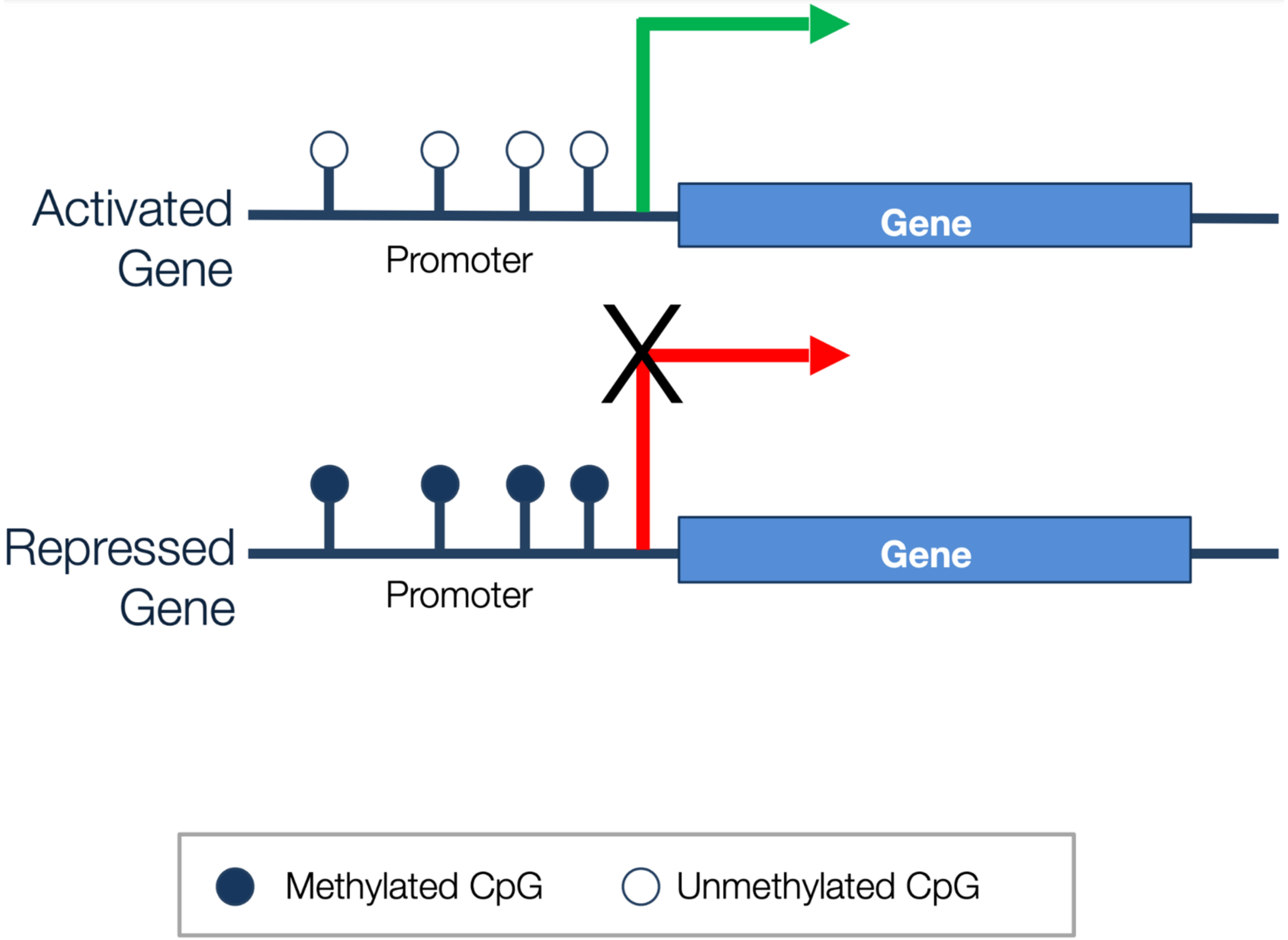
DNA-demethylase(s)
TET enymes
Passive demethylation

Cytosine

5-methyl Cytosine

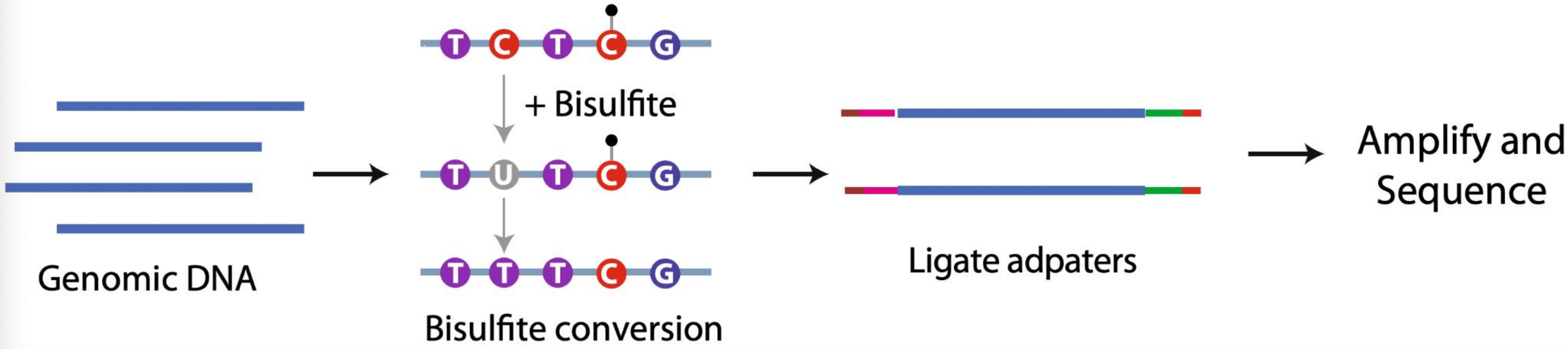
https://en.wikipedia.org/wiki/DNA_methylation

Role of DNA methylation in transcriptional regulation



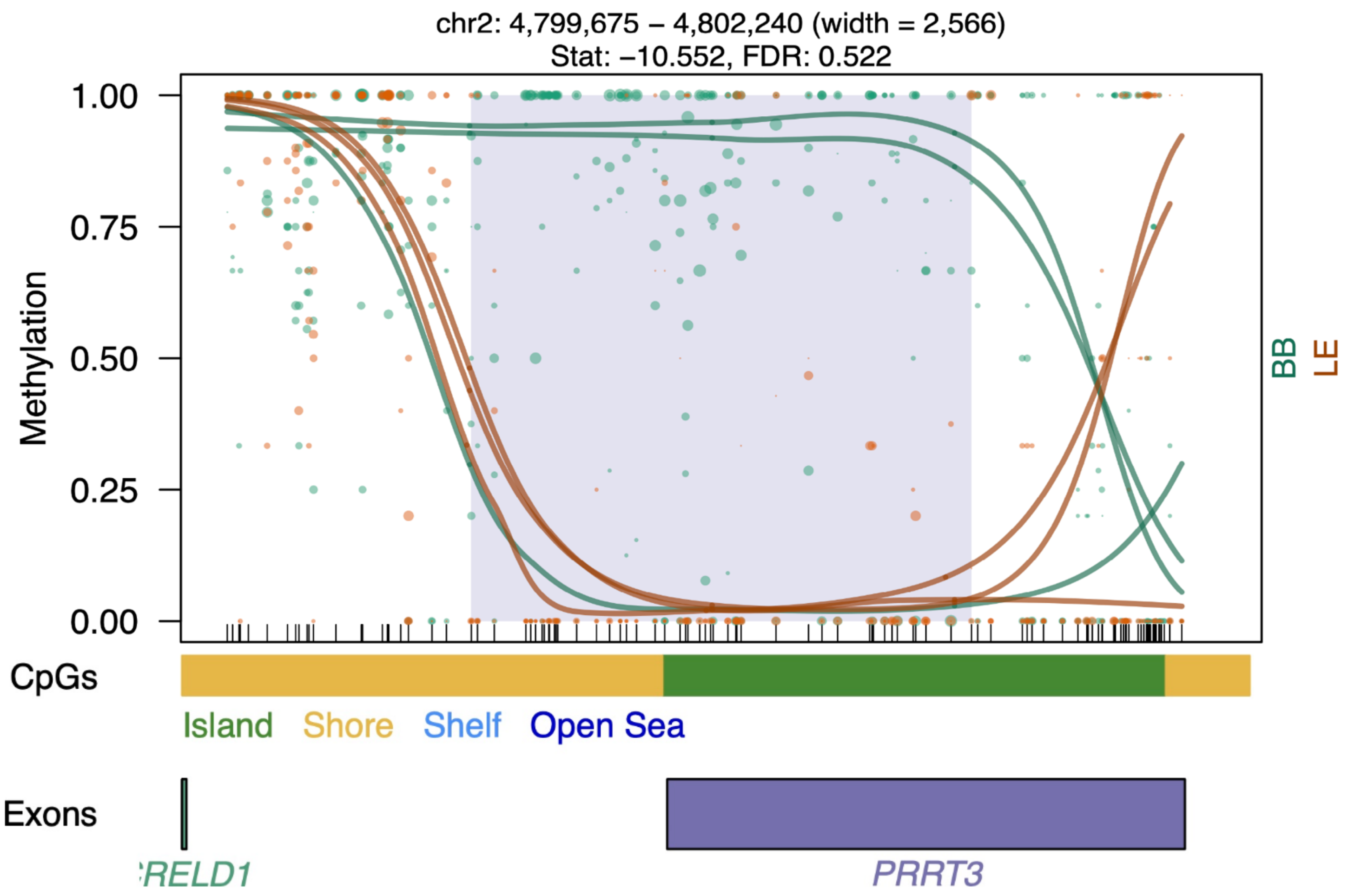
https://en.wikipedia.org/wiki/DNA_methylation

Gold standard to study DNA methylation patterns across the genome: WGBS

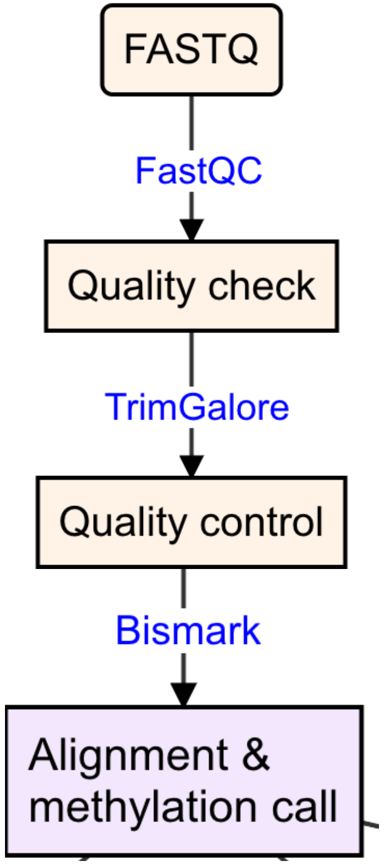


https://doi.org/10.1007%2Fs11357-018-0005-3

Example plot for DNA methylation



DNAme data analysis pipeline



FASTQ

FastQC

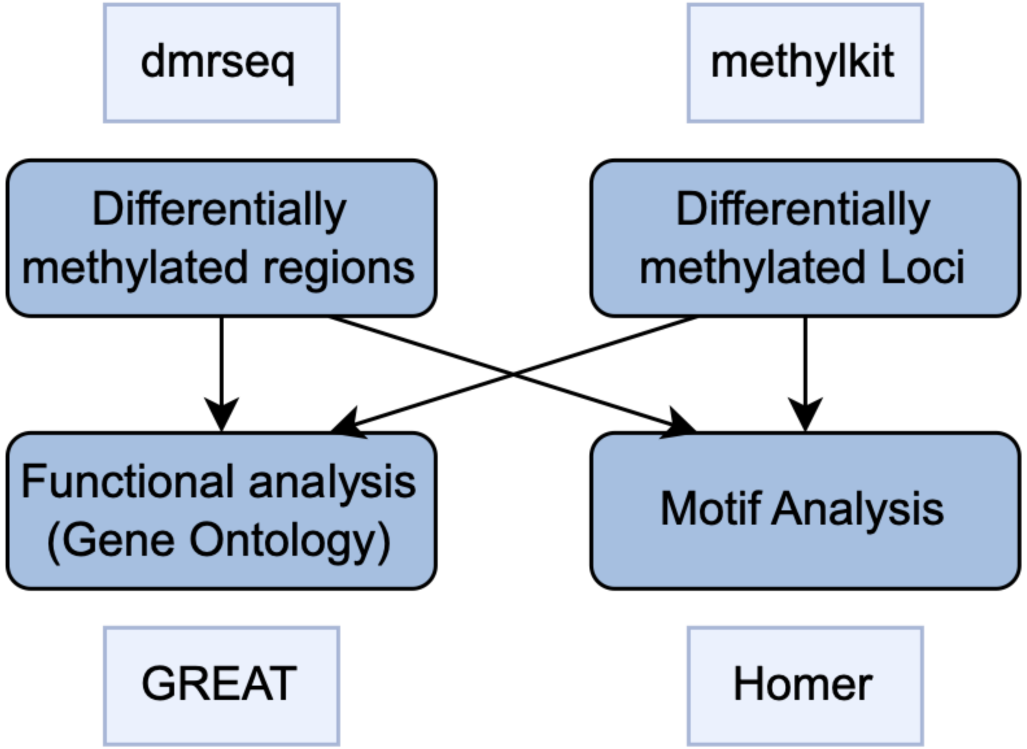
Quality check

TrimGalore

Quality control

Bismark

Alignment & methylation call



dmrseq

methylkit

Differentially methylated regions

Differentially methylated Loci

Functional analysis (Gene Ontology)

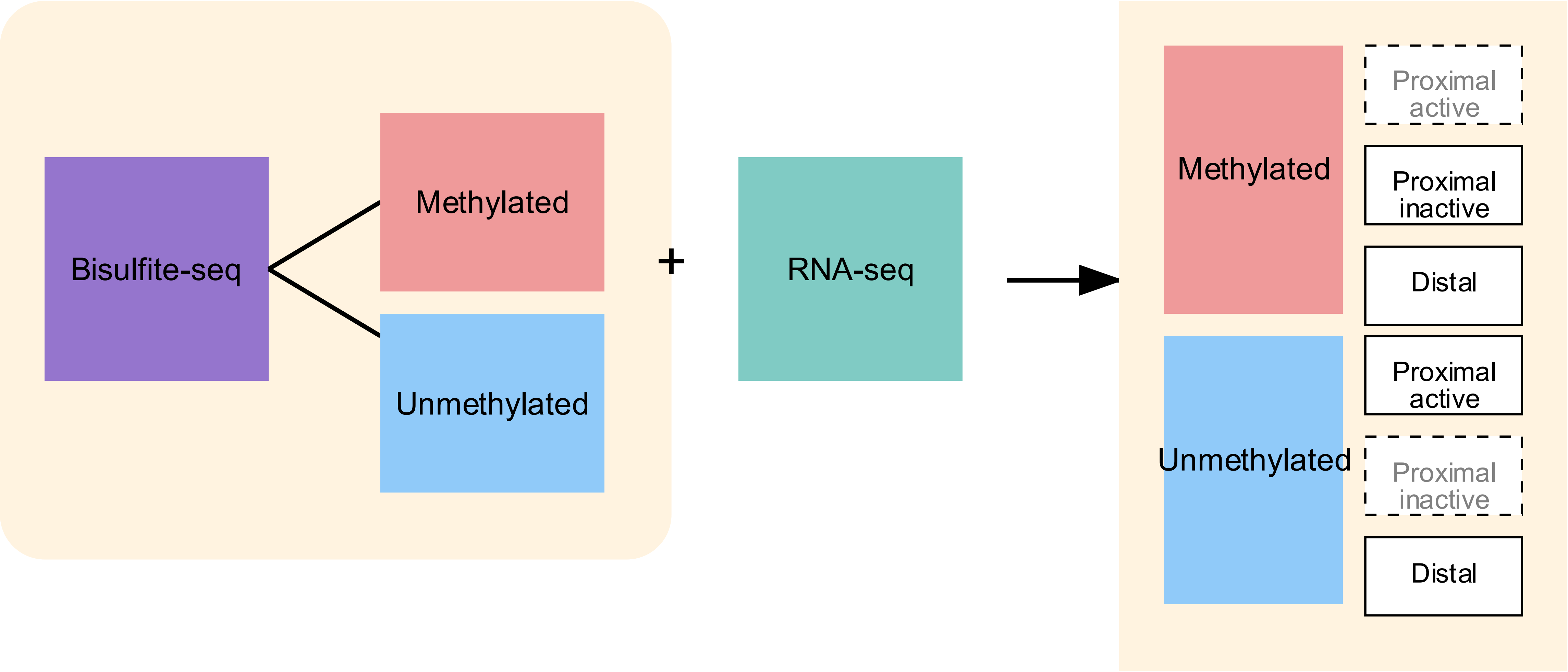
Motif Analysis

GREAT

Homer

https://doi.org/10.3929/ethz-b-000537329

Integration of RNA-seq and WGBS



Bisulfite-seq

Methylated

Unmethylated

+

RNA-seq

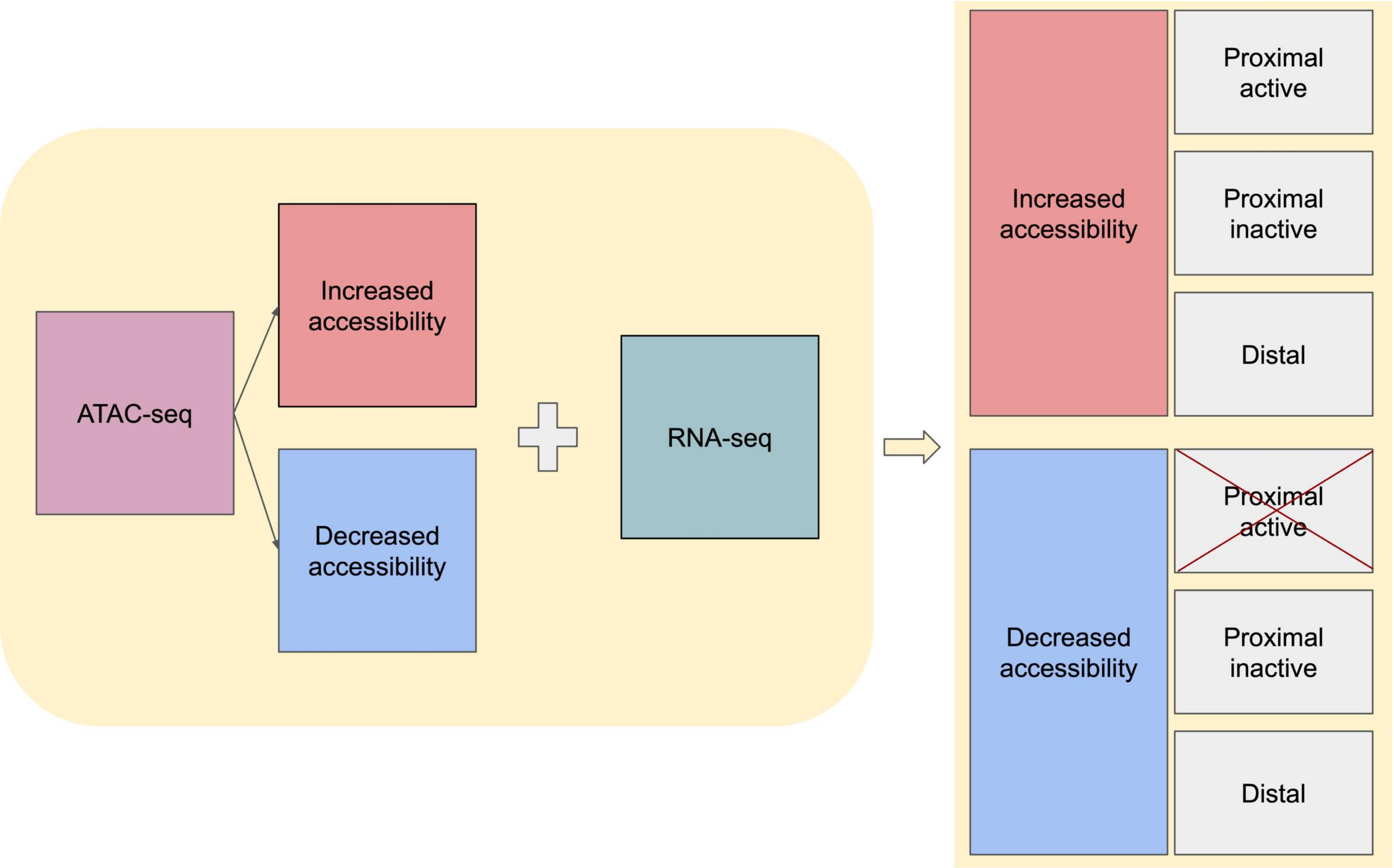
Methylated

- Proximal active
- Proximal inactive
- Distal

Unmethylated

- Proximal active
- Proximal inactive
- Distal

Integration of RNA-seq and ATAC-seq



ATAC-seq

Increased accessibility

Decreased accessibility

RNA-seq

Increased accessibility

Proximal active

Proximal inactive

Distal

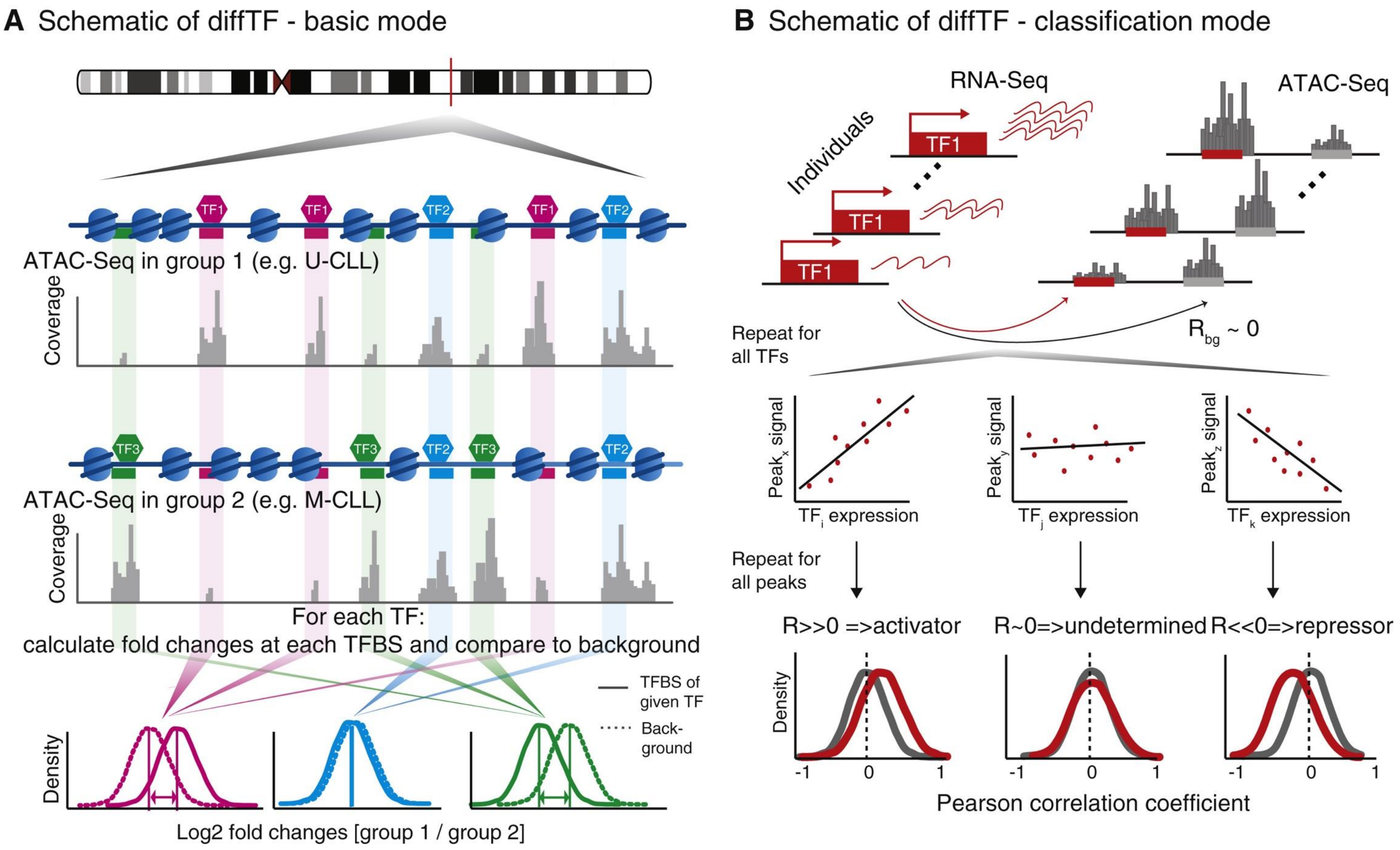
Decreased accessibility

~~Proximal active~~

Proximal inactive

Distal

Integration of RNA-seq and ATAC-seq: diffTF

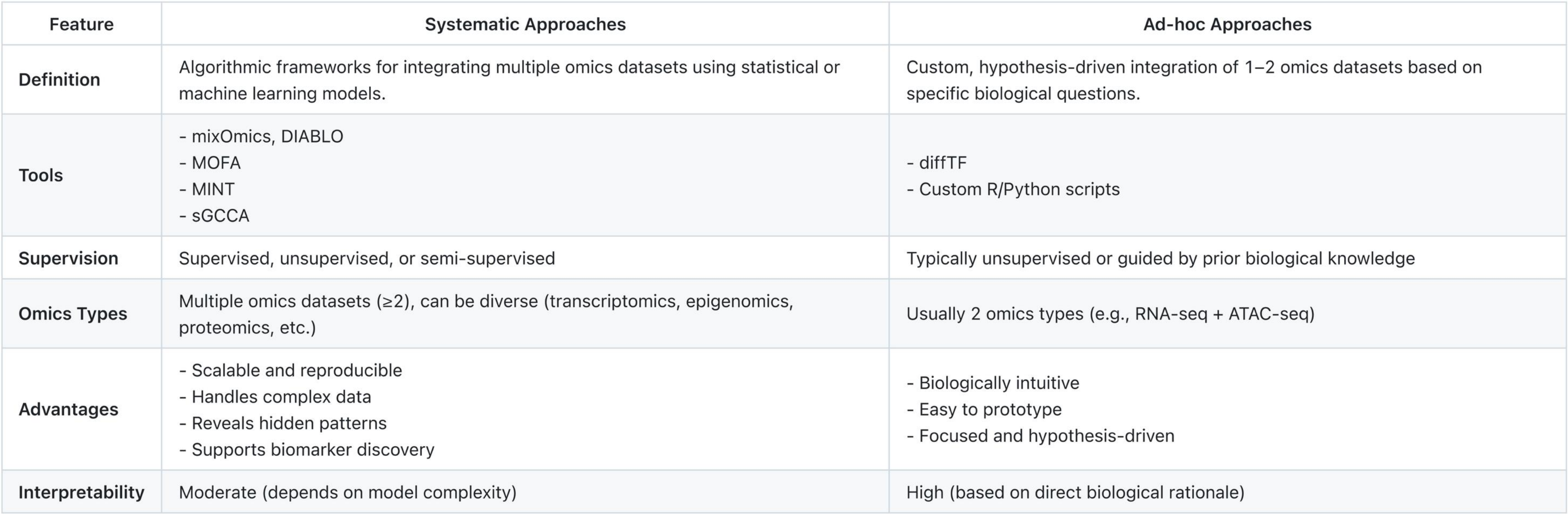


[10.1016/j.celrep.2019.10.106 External Link](https://doi.org/10.1016/j.celrep.2019.10.106)

Multi-omics integration overview

Feature	Systematic Approaches	Ad-hoc Approaches
Definition	Algorithmic frameworks for integrating multiple omics datasets using statistical or machine learning models.	Custom, hypothesis-driven integration of 1–2 omics datasets based on specific biological questions.
Tools	- mixOmics, DIABLO - MOFA - MINT - sGCCA	- diffTF - Custom R/Python scripts
Supervision	Supervised, unsupervised, or semi-supervised	Typically unsupervised or guided by prior biological knowledge
Omics Types	Multiple omics datasets (≥2), can be diverse (transcriptomics, epigenomics, proteomics, etc.)	Usually 2 omics types (e.g., RNA-seq + ATAC-seq)
Advantages	- Scalable and reproducible - Handles complex data - Reveals hidden patterns - Supports biomarker discovery	- Biologically intuitive - Easy to prototype - Focused and hypothesis-driven
Interpretability	Moderate (depends on model complexity)	High (based on direct biological rationale)

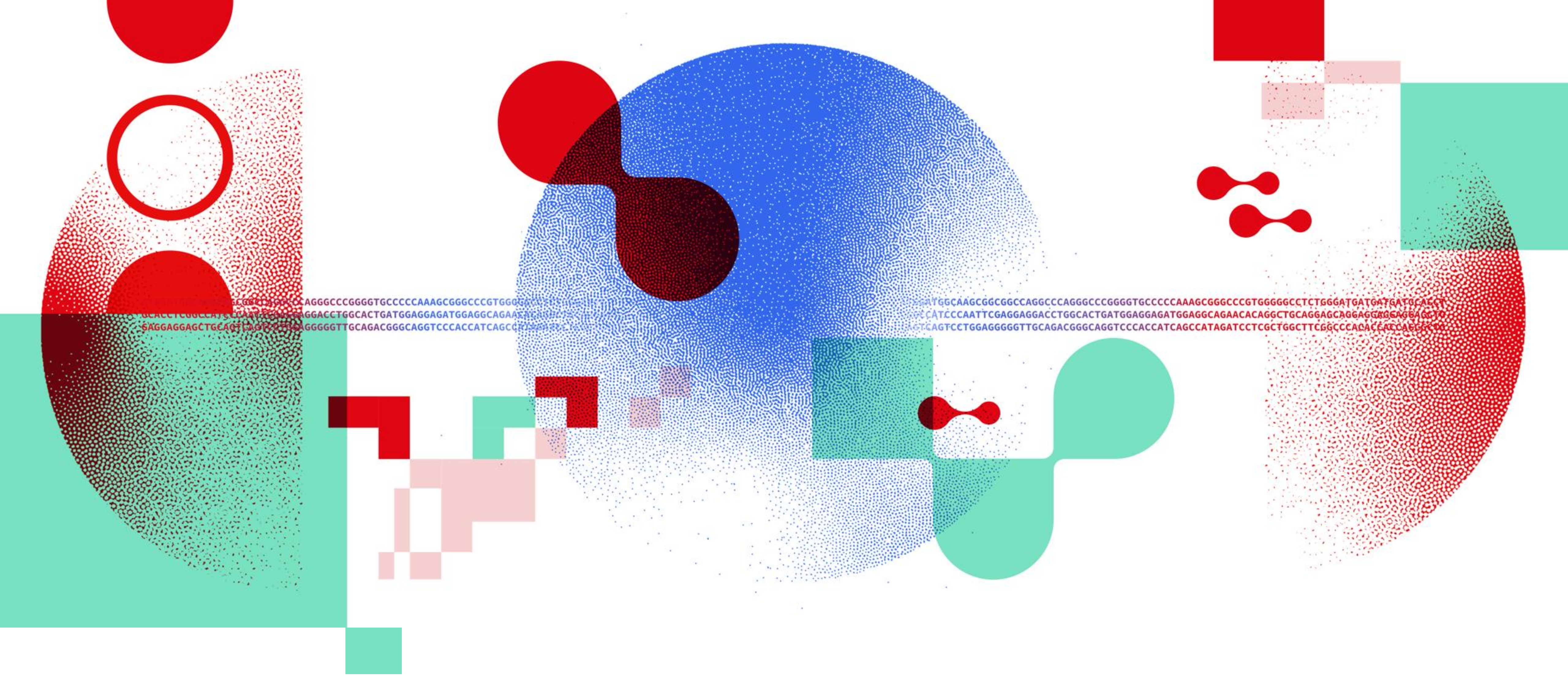
Multi-omics integration



Feature	Systematic Approaches	Ad-hoc Approaches
Definition	Algorithmic frameworks for integrating multiple omics datasets using statistical or machine learning models.	Custom, hypothesis-driven integration of 1–2 omics datasets based on specific biological questions.
Tools	- mixOmics, DIABLO - MOFA - MINT - sGCCA	- diffTF - Custom R/Python scripts
Supervision	Supervised, unsupervised, or semi-supervised	Typically unsupervised or guided by prior biological knowledge
Omics Types	Multiple omics datasets (≥2), can be diverse (transcriptomics, epigenomics, proteomics, etc.)	Usually 2 omics types (e.g., RNA-seq + ATAC-seq)
Advantages	- Scalable and reproducible - Handles complex data - Reveals hidden patterns - Supports biomarker discovery	- Biologically intuitive - Easy to prototype - Focused and hypothesis-driven
Interpretability	Moderate (depends on model complexity)	High (based on direct biological rationale)

We will learn ad-hoc approaches for multi-omics in this course

Quiz: 6-10



Thank you

DATA SCIENTISTS FOR LIFE

sib.swiss

elixir SWITZERLAND

SIB