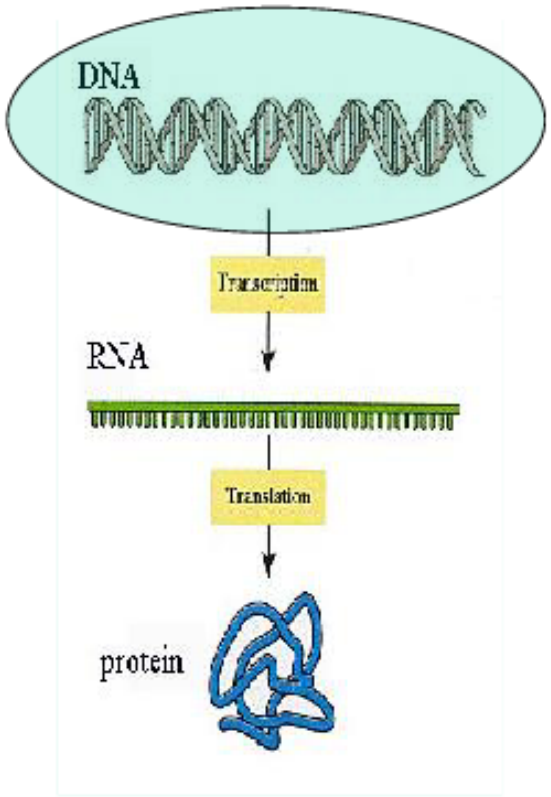
生物信息学

# 表观遗传学

**Yu Hou (侯 宇)** Zhejiang University 2025

# 基因表达调控




**转录前调控**
- 染色质的活化
- 基因扩增
- 基因重排
- 基因丢失

**转录调控**
- 顺式作用元件
- 反式作用因子

**转录后调控**
- mRNA加工成熟
- mRNA的转运

**翻译调控**
- 蛋白质合成过程

**翻译后调控**
- 蛋白质的加工成熟

# Sequencing techniques

**Genomic sequencing: （DNA）**

WGS, WES, Single-cell Genomic Sequencing

**Transcriptomic sequencing: （RNA）**

Total-RNA-seq, mRNA-seq, small RNA-seq, Ribo-seq, circRNA-seq, single-cell RNA-seq

**Epigenomic sequencing: （Modifications）**

DNA-methylation (WGBS, RRBS, MeDIP-seq)

Histone-modification (ChIP-seq, CUT&RUN, CUT&Tag)

Chromatin accessibility (ATAC-Seq, DNase-seq, NOMe-seq)

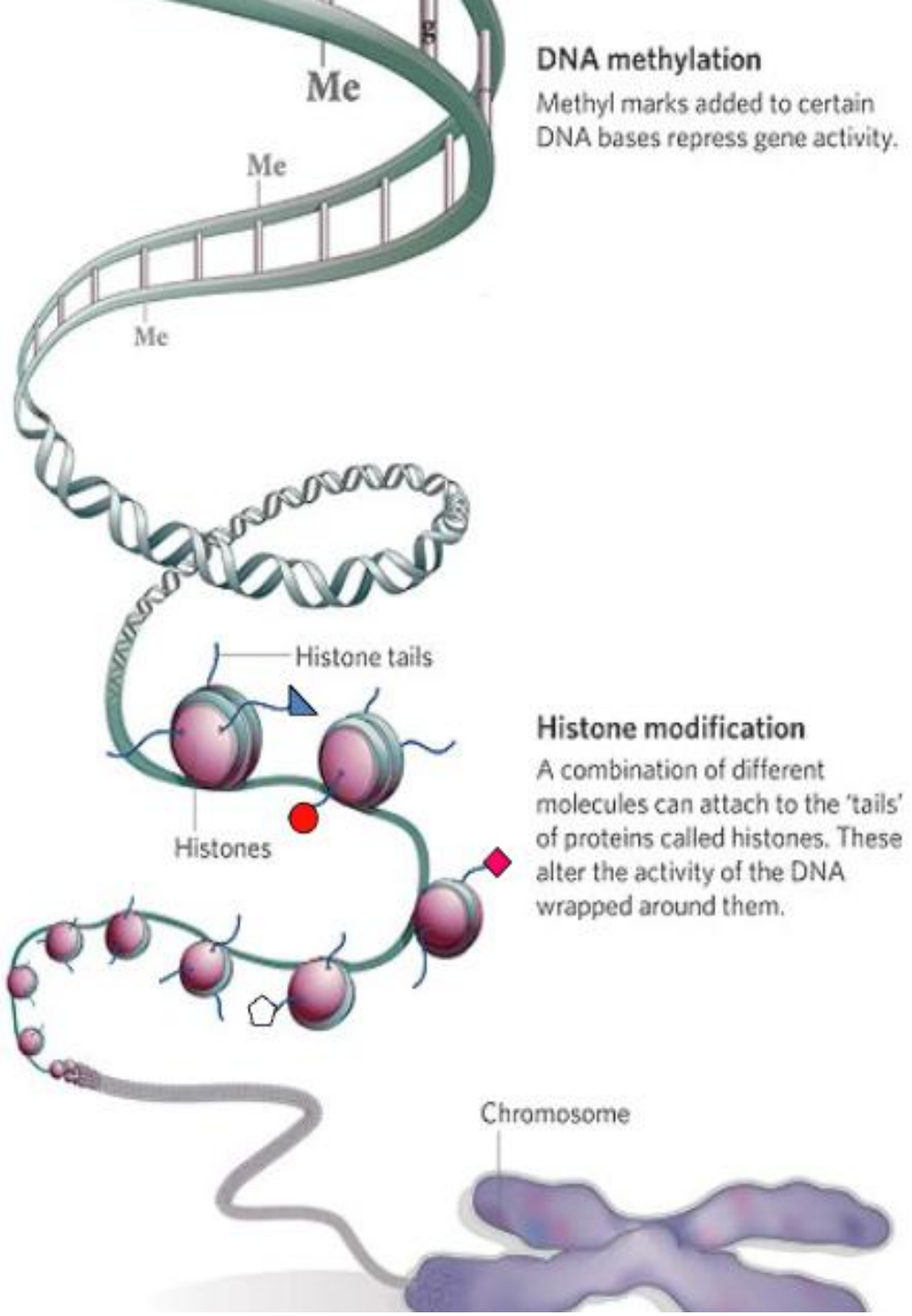
Chromatin Structure (3C-seq, Hi-C-seq)

# Epigenomics

## What is Epigenomics?

- The study of the complete set of epigenetic modifications on the genetic material of a cell.

- DNA methylation and histone modification are two primary examples of processes that result in epigenetic modifications to the genome. The changes linked to these processes can accumulate through person life and can be passed from one generation to the next.



# Epigenetic Sequencing

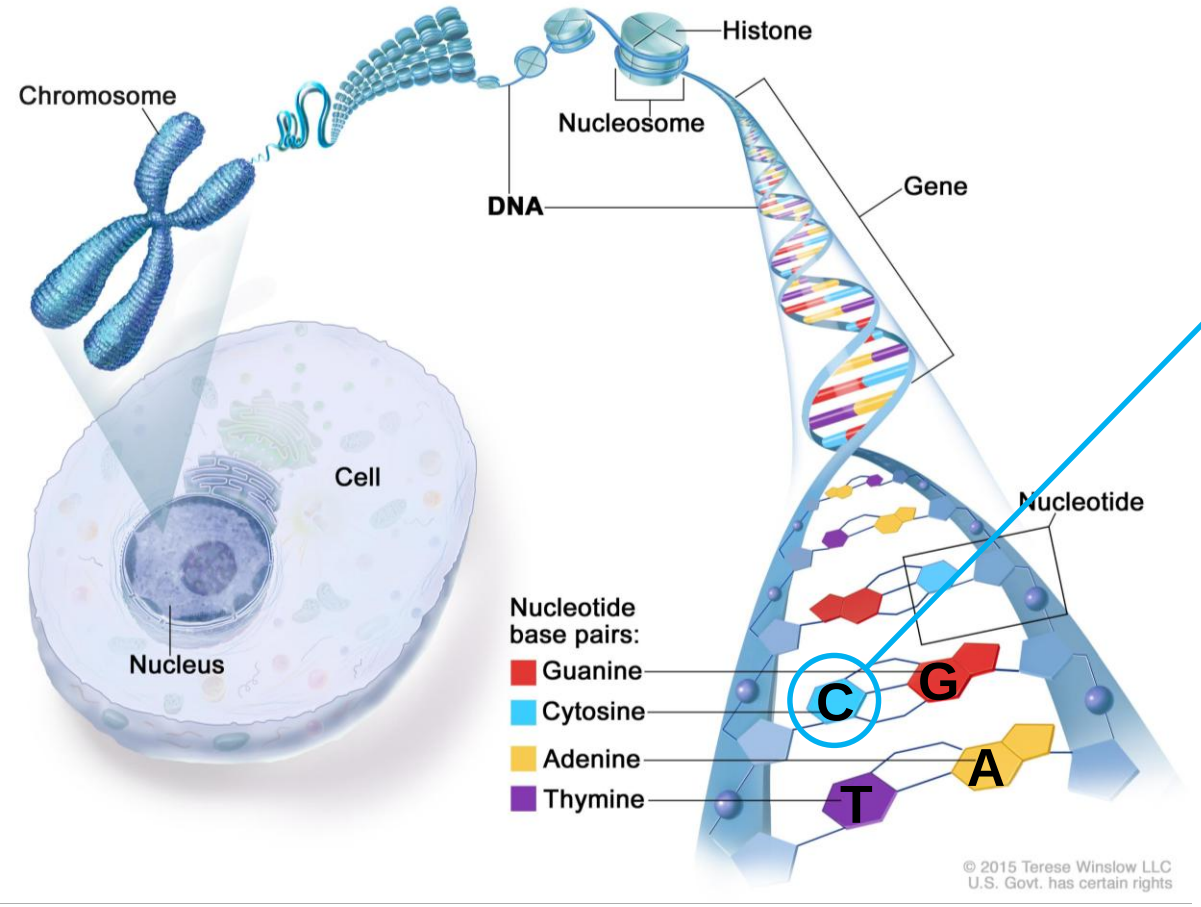
## Epigenetic Sequencing

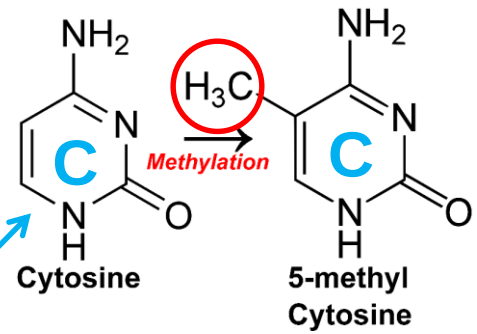
1. DNA-methylation  (WGBS, RRBS, MeDIP-seq, single-cell WGBS)
2. Histone-modification

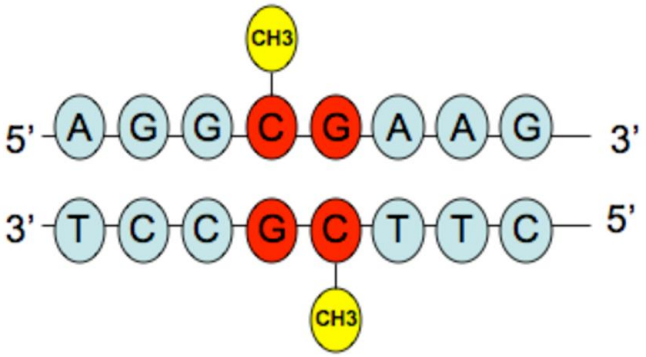
   (ChIP-seq, CUT&RUN, CUT&Tag, single-cell CUT&Tag)
3. Chromatin accessibility (ATAC-Seq, DNase-seq, NOMe-seq, single-cell ATAC-seq)
4. Chromatin Structure (3C-seq, Hi-C-seq, single-cell Hi-C)
5. Multi-omics analyses and applications

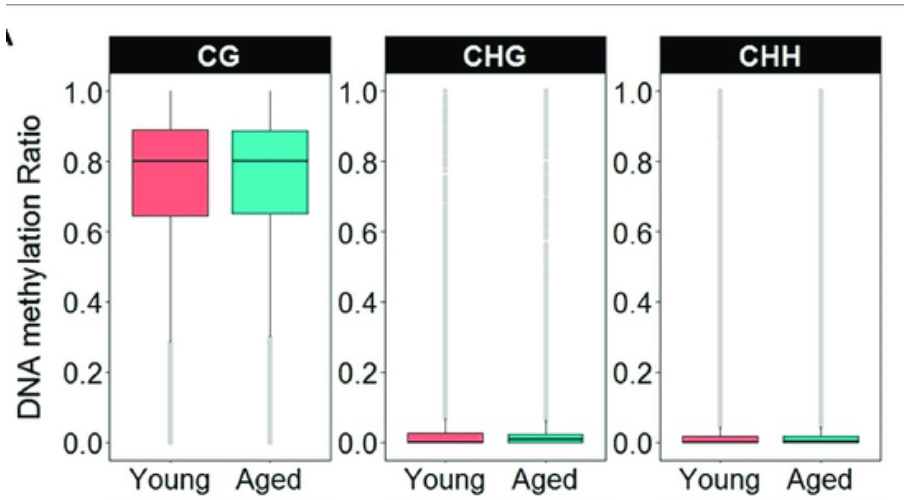
# 01 DNA Methylation

## DNA methylation (the fifth nucleotide: methyl cytosine)



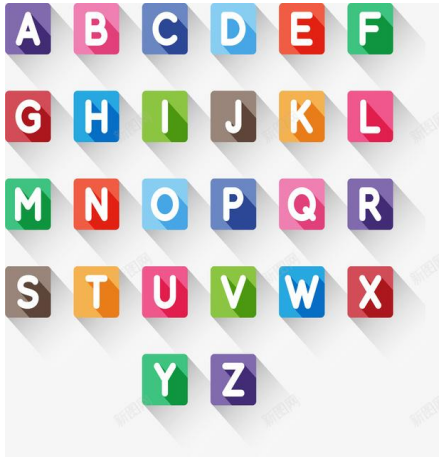



# 01 DNA Methylation

List of symbols

| Symbol[2] | Meaning/derivation | Possible bases | | | | | Complement |
|---|---|---|---|---|---|---|---|
| A | Adenine | A | | | | 1 | T (or U) |
| C | Cytosine | | C | | | | G |
| G | Guanine | | | G | | | C |
| T | Thymine | | | | T | | A |
| U | Uracil | | | | U | | A |
| W | Weak | A | | | T | 2 | S |
| S | Strong | | C | G | | | W |
| M | aMino 胺 | A | C | | | | K |
| K | Keto 酮 | | | G | T | | M |
| R | puRine 嘌呤 | A | | G | | | Y |
| Y | pYrimidine 嘧啶 | | C | | T | | R |
| B | not A (B comes after A) | | C | G | T | 3 | V |
| D | not C (D comes after C) | A | | G | T | | H |
| H | not G (H comes after G) | A | C | | T | | D |
| V | not T (V comes after T and U) | A | C | G | | | B |
| N | any Nucleotide (not a gap) | A | C | G | T | 4 | N |
| Z | Zero | | | | | 0 | Z |

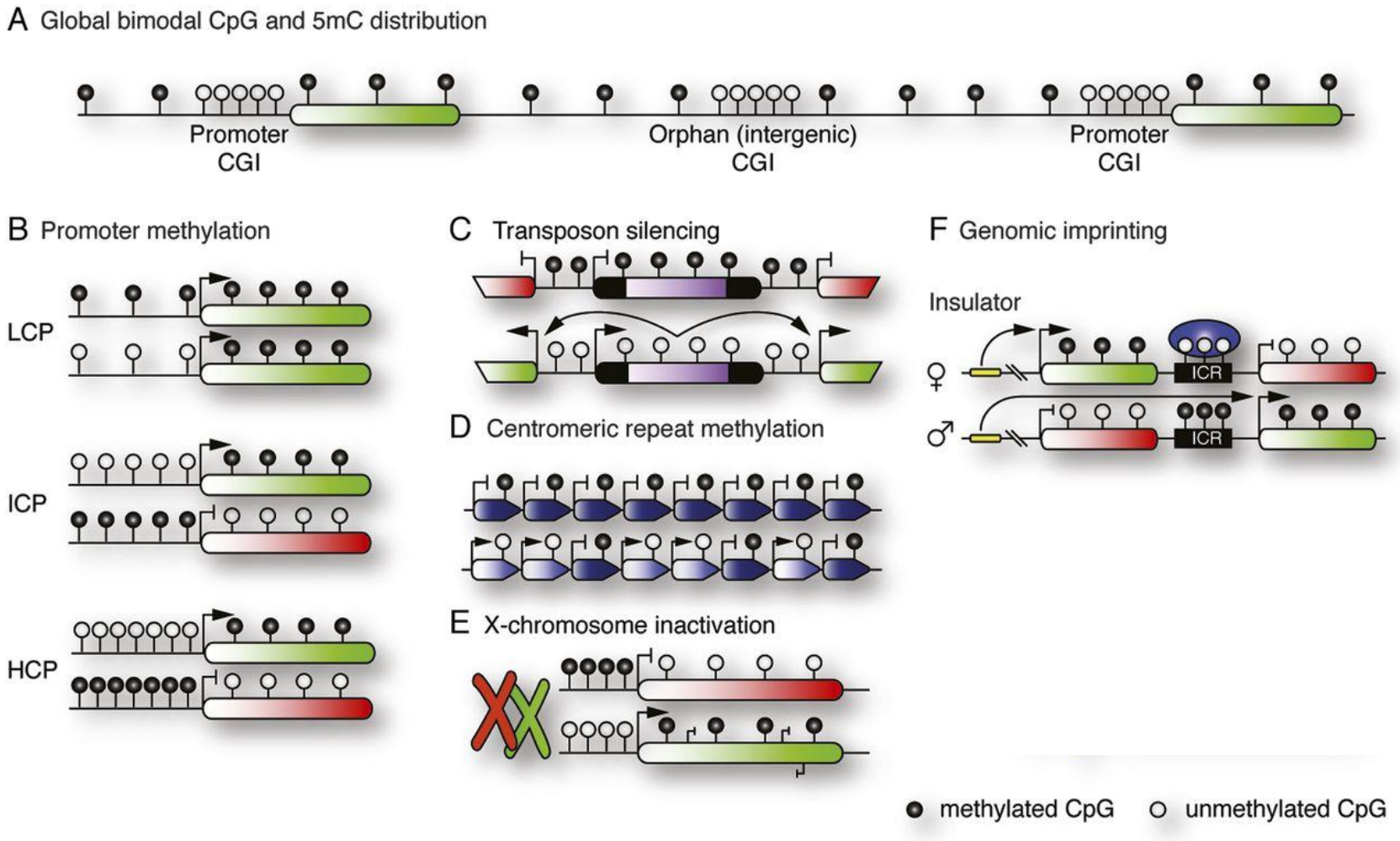



S = G or C
•S represents bases that form strong hydrogen bonds, specifically G (Guanine) and C (Cytosine). G and C pair together with three hydrogen bonds, which is stronger compared to the two hydrogen bonds between A and T. Hence, S is used to denote this stronger pairing.

W = A or T
•W represents bases that form weaker hydrogen bonds, specifically A (Adenine) and T (Thymine). Since A and T form only two hydrogen bonds, which are considered weaker compared to the three hydrogen bonds between G and C, W is used to represent these bases.

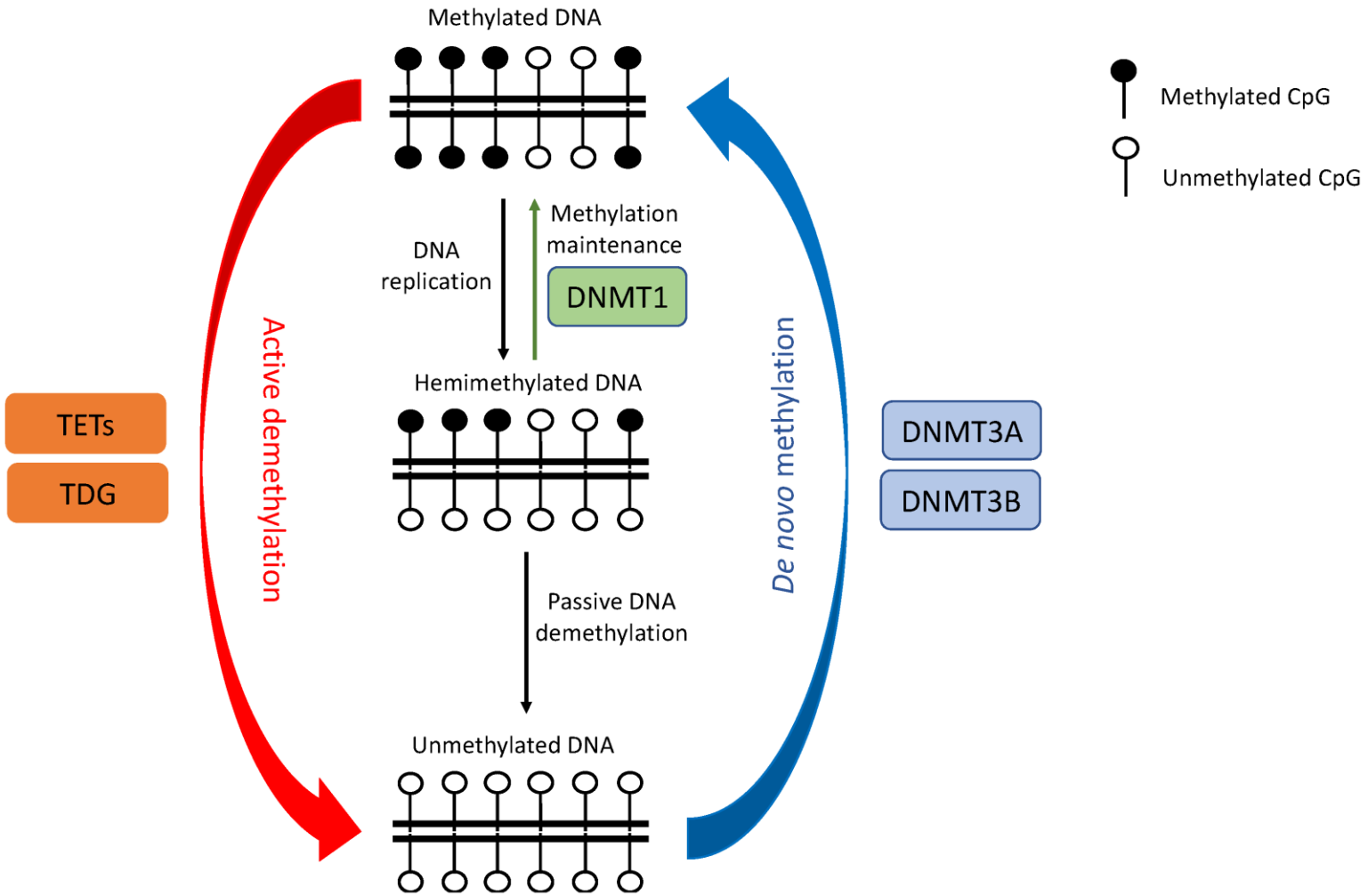
# 01 DNA Methylation

## DNA methylation



# 01 DNA Methylation

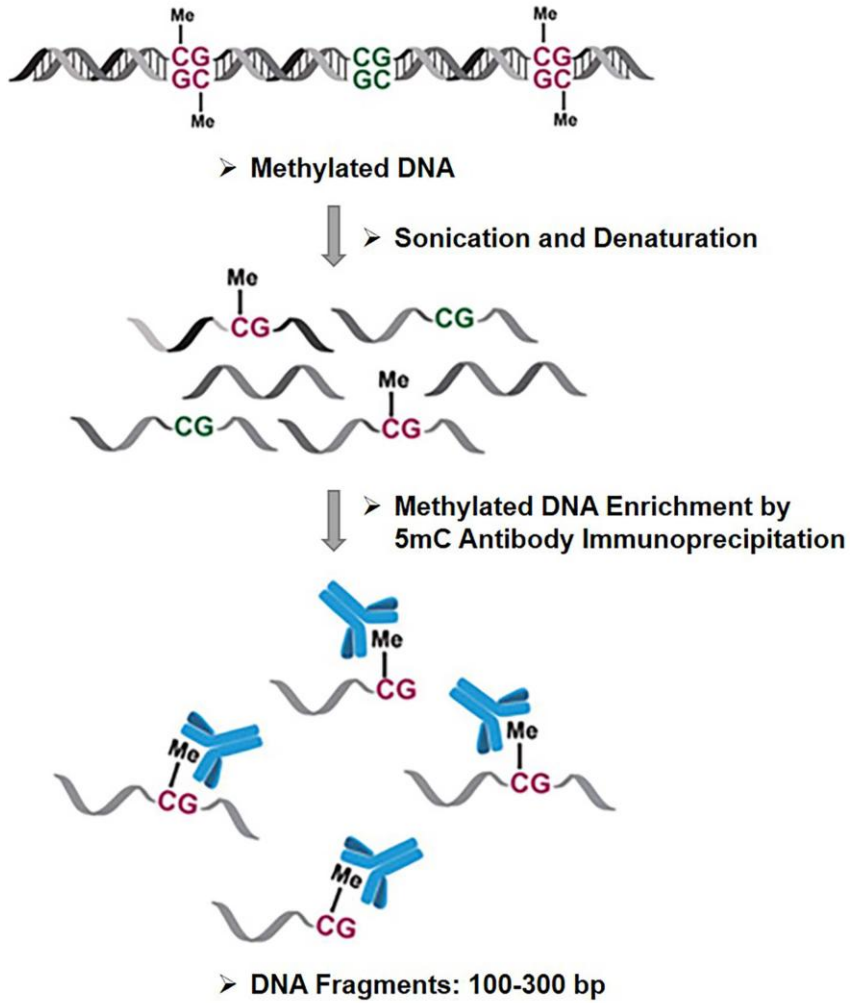
## DNA methylation



# 01 DNA Methylation

## 1.1. MeDIP-seq

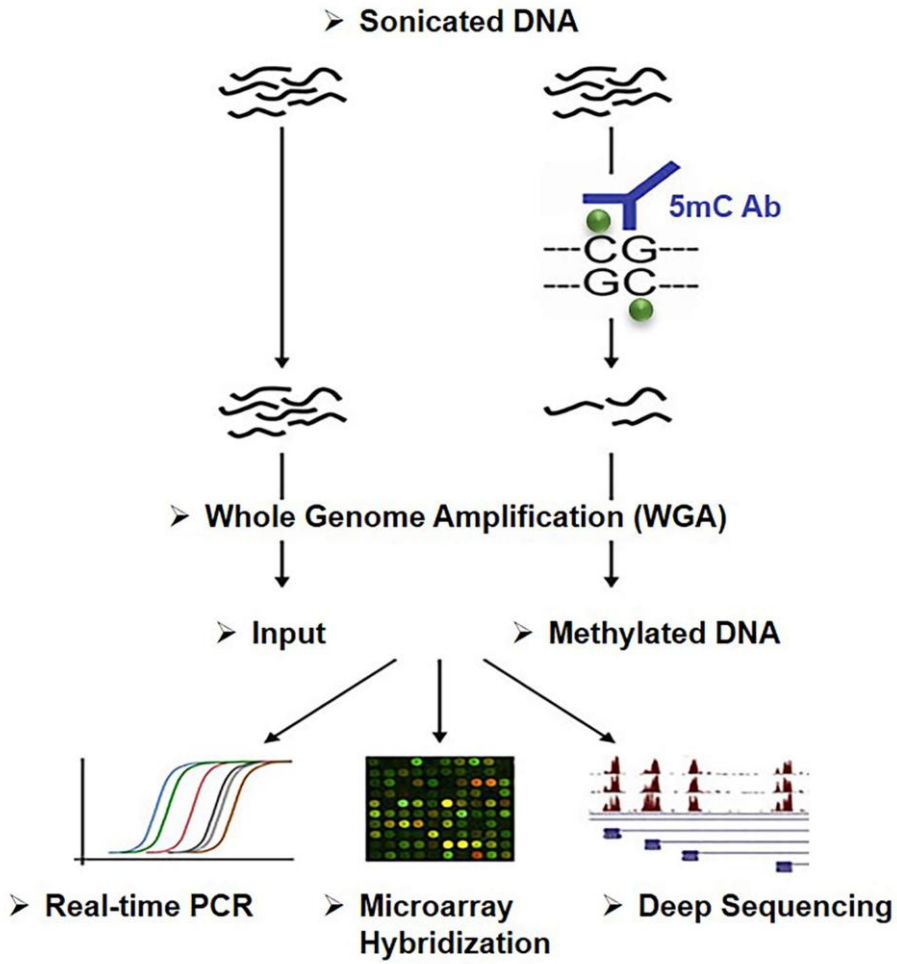
**Methylated DNA Immunoprecipitation Sequencing (MeDIP-seq)** is a high-throughput, antibody-based method used to study DNA methylation at a genome-wide scale. Unlike bisulfite sequencing methods (e.g., WGBS, RRBS), which detect single-base resolution methylation, MeDIP-seq identifies methylated regions (methylomes) based on enrichment rather than base conversion.



# 01 DNA Methylation

## 1.1. MeDIP-seq

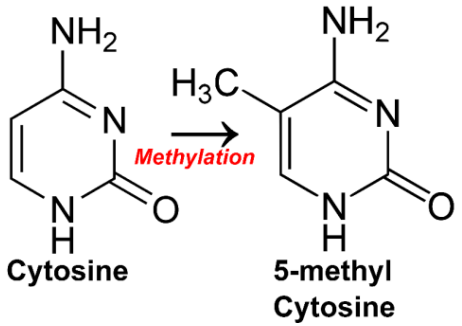
**Methylated DNA Immunoprecipitation Sequencing (MeDIP-seq)** is a high-throughput, antibody-based method used to study DNA methylation at a genome-wide scale. Unlike bisulfite sequencing methods (e.g., WGBS, RRBS), which detect single-base resolution methylation, MeDIP-seq identifies methylated regions (methylomes) based on enrichment rather than base conversion.

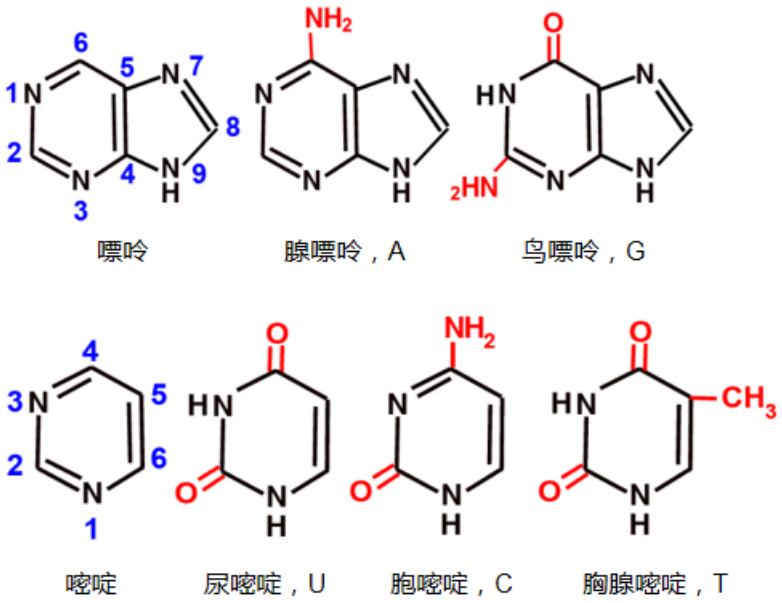


# 01 DNA Methylation

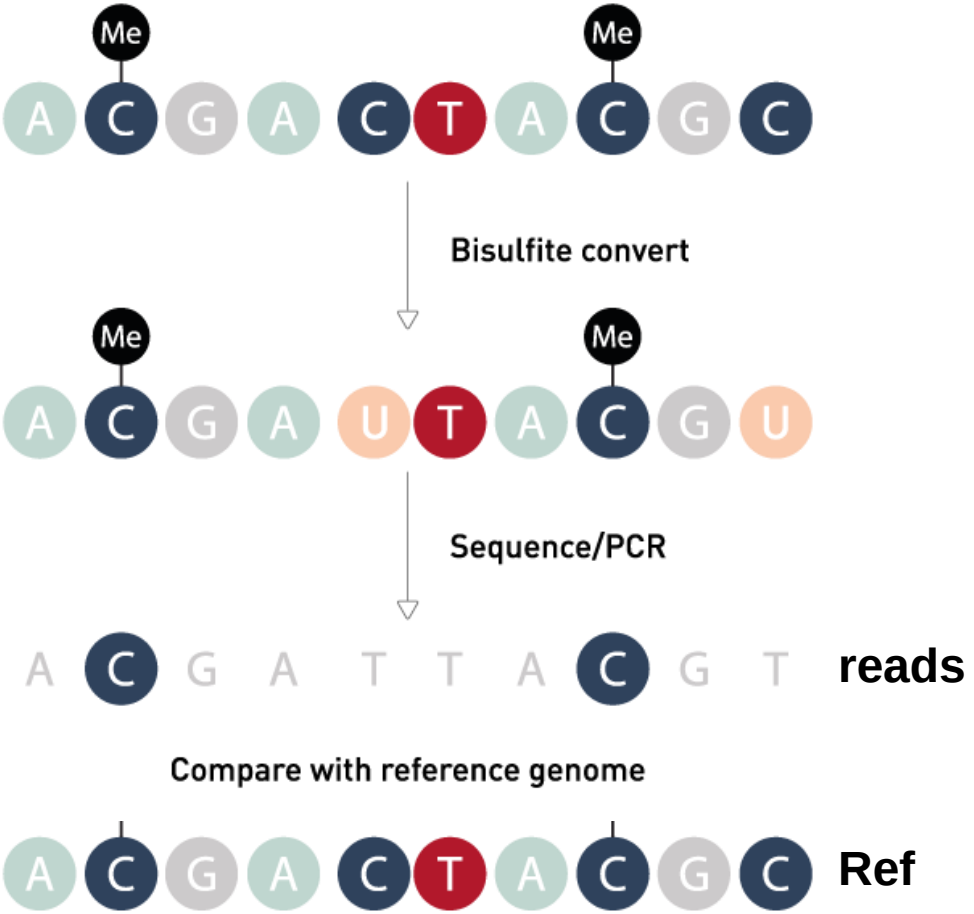
## 1.2. WGBS

**Whole Genome Bisulfite Sequencing (WGBS)** is a gold-standard method for analyzing 5-methylcytosine (5mC) modifications at single-base resolution across the entire genome.




PCR can not maintain 5mC on the amplicons



# 01 DNA Methylation

## 1.2. WGBS

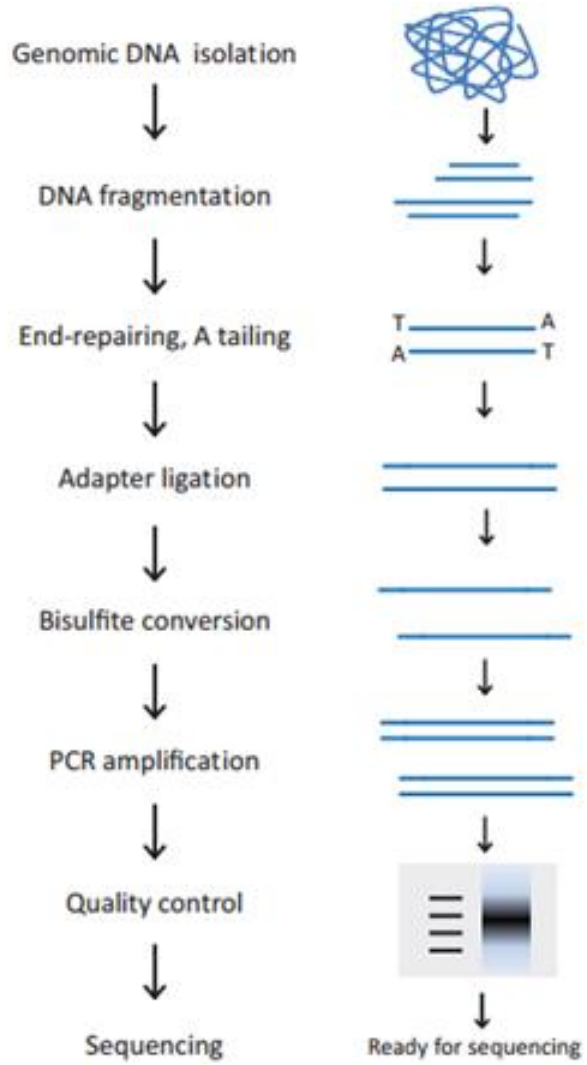
**Key Steps**

**1.Genomic DNA extraction**

**2.Library preparation** (Fragmentation, adapter ligation, Bisulfite conversion, PCR amplification)

**3.High-throughput sequencing** (Usually Illumina platform)

**4.Data analysis** (Alignment to reference genome, methylation calling, and visualization)

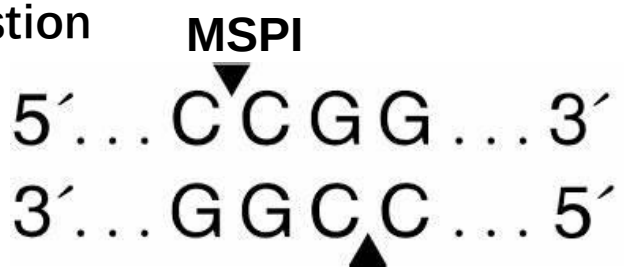


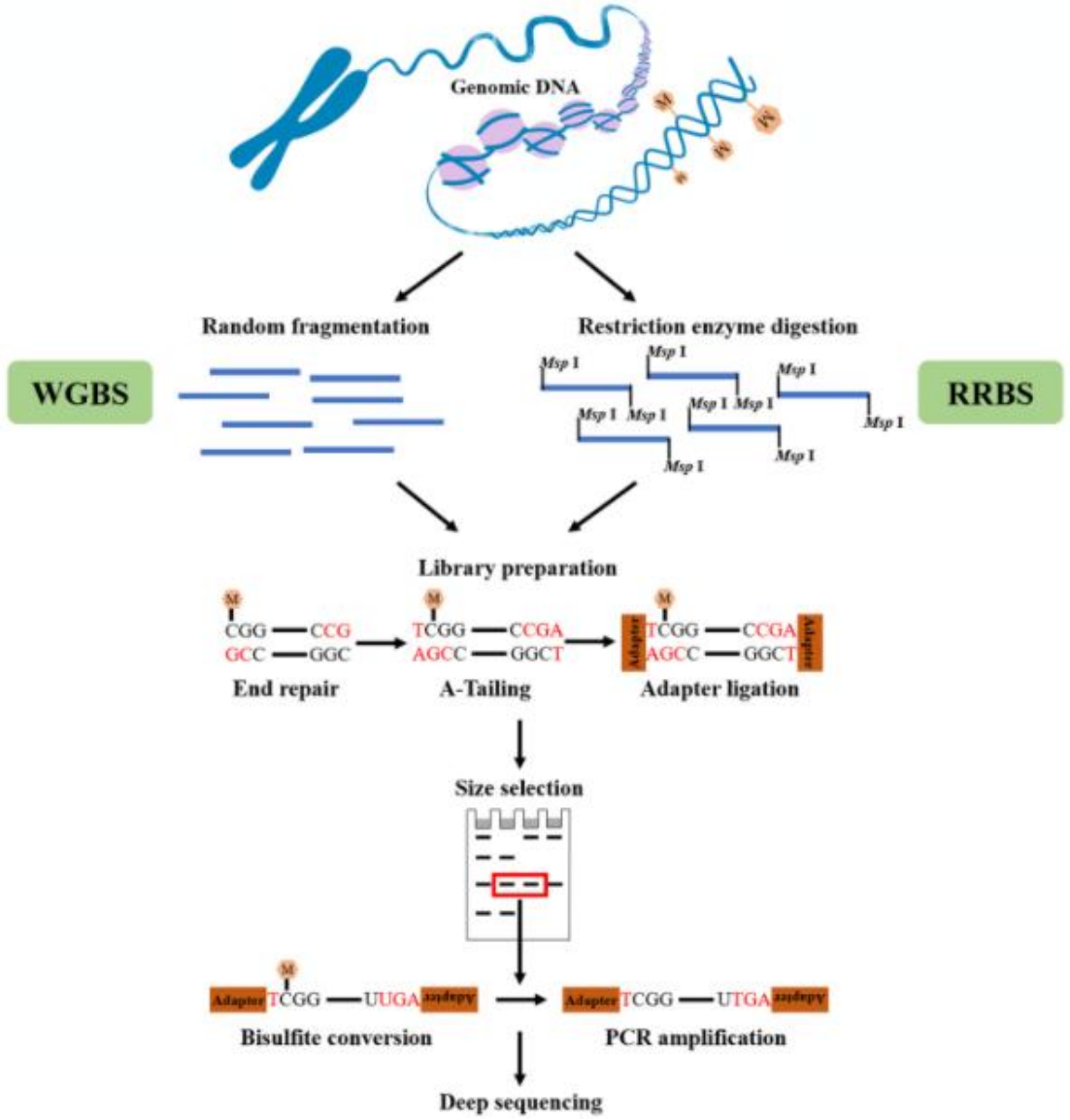
# 01 DNA Methylation

## 1.3. RRBS

**Reduced Representation Bisulfite Sequencing (RRBS)** is a cost-effective targeted bisulfite sequencing method used to study DNA methylation at a single-nucleotide resolution. It selectively enriches CpG-rich regions, such as promoters, enhancers, and CpG islands, making it an efficient alternative to Whole Genome Bisulfite Sequencing (WGBS).

**restriction enzyme digestion**



# 01 DNA Methylation

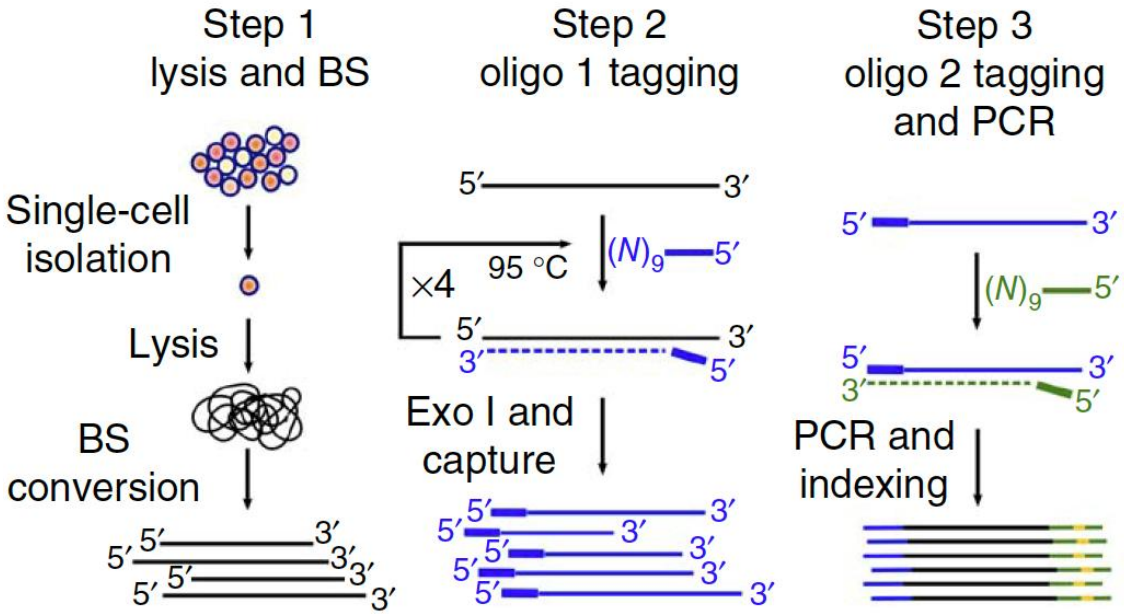
## 1.3. RRBS

| Feature | MeDIP-seq | WGBS | RRBS |
|---|---|---|---|
| **Resolution** | Region-level (~100-500 bp) | Single-base | Single-base |
| **Genome Coverage** | Whole genome (but biased toward methyl-rich regions) | Whole genome | CpG islands (~1-10% of genome) |
| **CpG & Non-CpG Detection** | ☑ Yes (CpG & CHG/CHH methylation) | ☑ Yes (CpG & non-CpG) | ✗ No (CpG-only) |
| **Sequencing Depth** | Low (~5×) | High (~30×) | Medium (~10×) |
| **Input DNA** | Low (~100 ng) | High (~500 ng) | Low (~10-100 ng) |
| **Best For** | Global methylation profiling, cost-effective | Comprehensive methylome analysis | Targeted CpG methylation |
| **Cost** | ☑ Low | ✗ High | ☑ Medium |

# 01 DNA Methylation

## 1.4. scWGBS

**Single-cell Whole Genome Bisulfite Sequencing (scWGBS)** is a high-resolution single-cell epigenomics method used to profile DNA methylation at single-base and single-cell levels. It combines whole-genome bisulfite sequencing (WGBS) with single-cell isolation techniques to study cell-to-cell heterogeneity in DNA methylation, which is critical for understanding development, cancer, and epigenetic regulation.

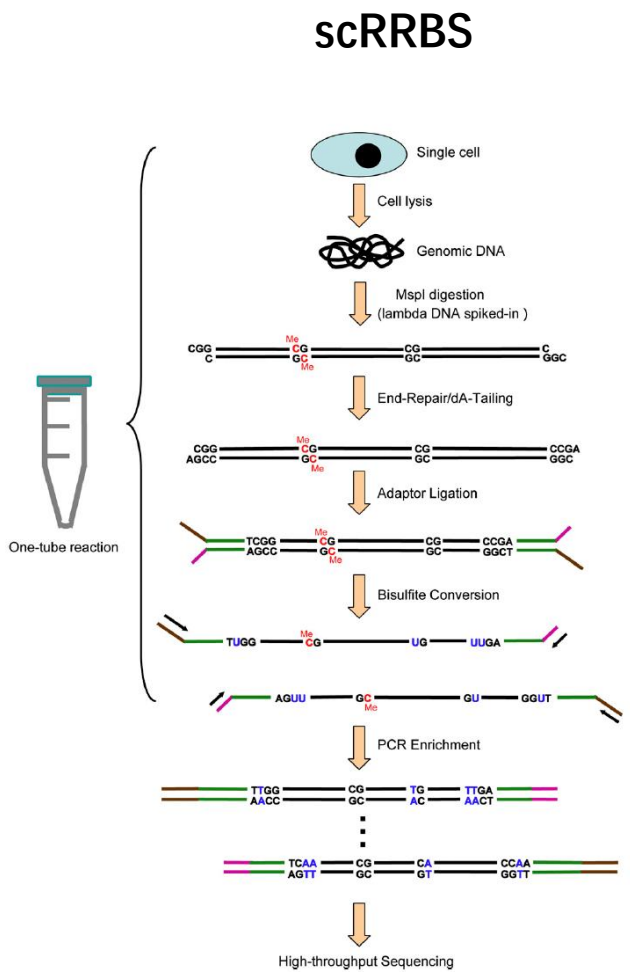



Smallwood *et al., Nature Methods, 2014*

# 01 DNA Methylation

## 1.4. scRRBS

**Single-cell Reduced Representation Bisulfite Sequencing (scRRBS)** is a targeted, single-cell epigenomics technique that profiles DNA methylation at CpG-rich regions, such as CpG islands, promoters, and enhancers. It is an adaptation of RRBS for single-cell applications, making it a cost-effective and efficient method to study epigenetic heterogeneity across individual cells.



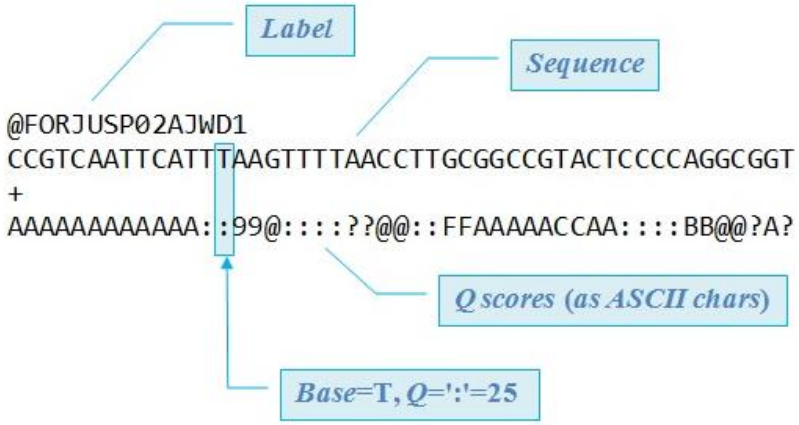

Hongshan Guo et al., *Genome Research,* 2013

# 01 DNA Methylation

## 1.5. Data Analyses

Pipeline

FASTQ → 质控 (FastQC, Trim Galore)
↓
比对到参考基因组 (Bismark)
↓
提取甲基化水平 (Bismark extractor / MethylDackel)
↓
差异甲基化分析 (methylKit / DSS)
↓
功能注释与整合分析 (annotatr / clusterProfiler)




**质量控制**

- **目的：** 去除低质量碱基与接头序列，防止比对错误
- **常用工具：**
  - FastQC
  - Trim Galore!
  - fastp

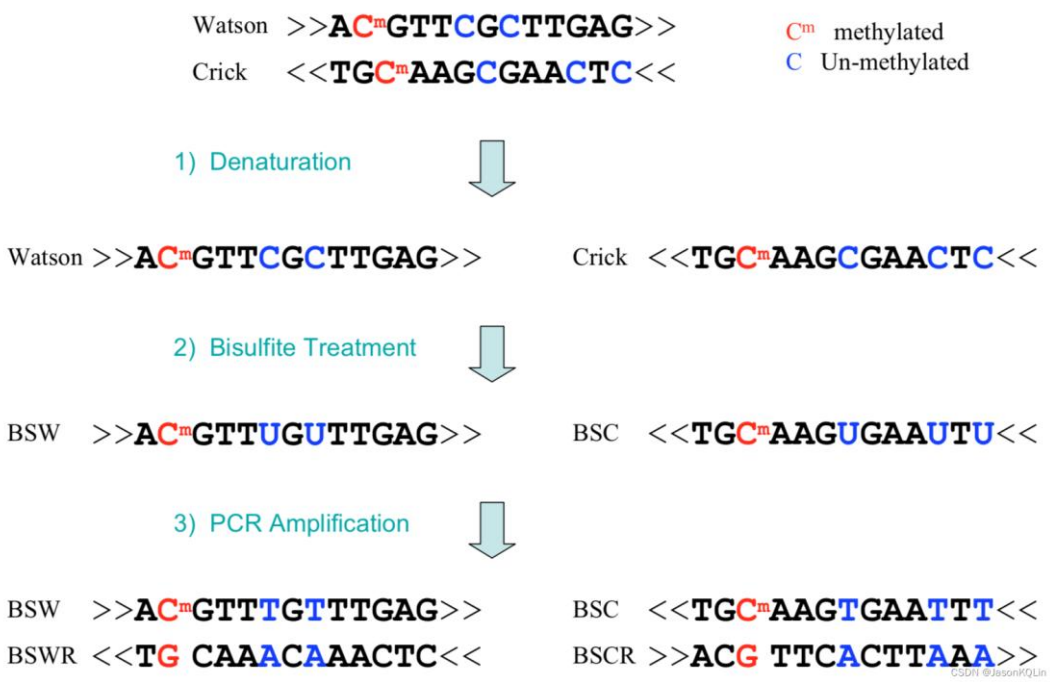
# 01 DNA Methylation

## 1.5. Data Analyses

Pipeline

FASTQ → 质控 (FastQC, Trim Galore)
↓
比对到参考基因组 (Bismark)
↓
提取甲基化水平 (Bismark extractor / MethylDackel)
↓
差异甲基化分析 (methylKit / DSS)
↓
功能注释与整合分析 (annotatr / clusterProfiler)

Bisulfite conversion




•**核心挑战**：亚硫酸氢盐处理使~90%的C变为T，导致读长与参考基因组序列严重不匹配，标准比对器（如BWA、STAR）无法直接使用。

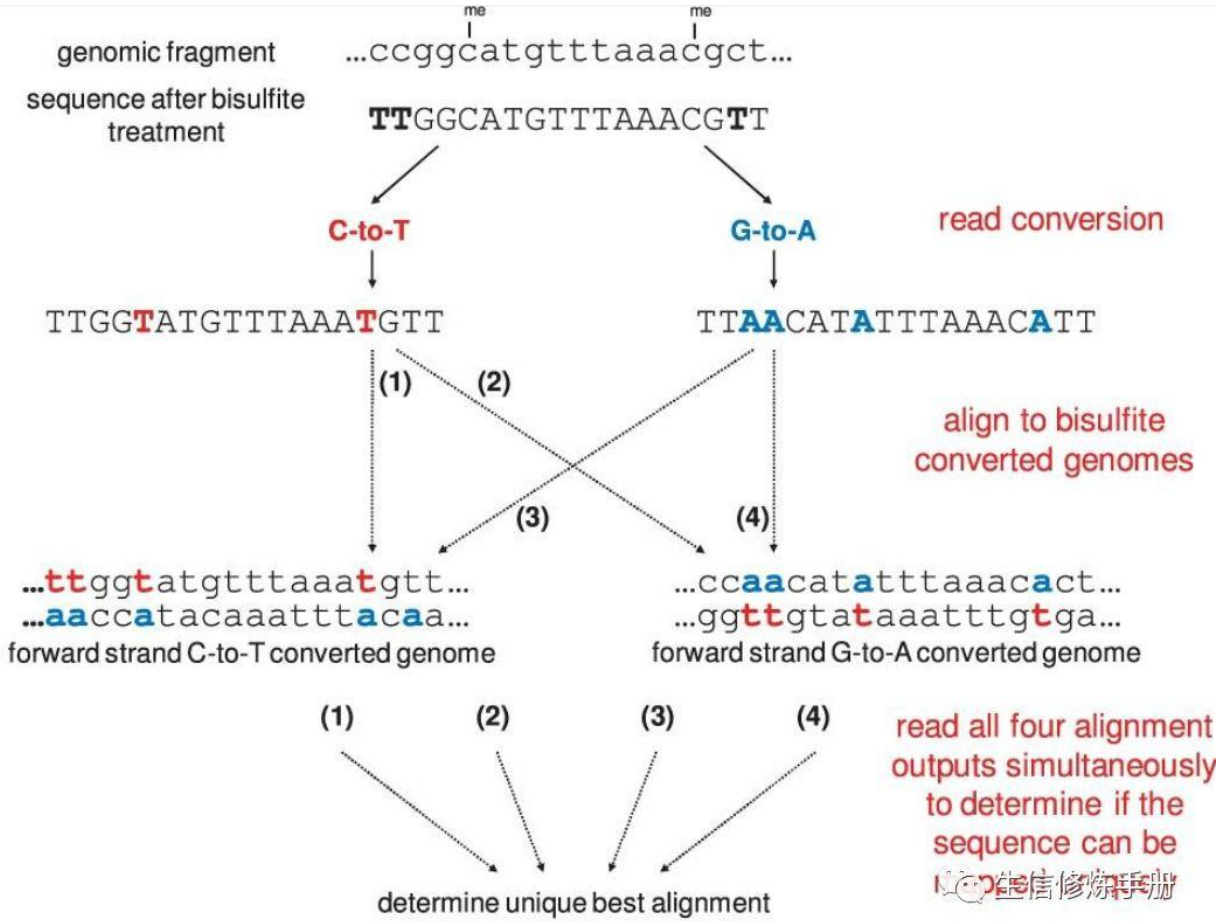
# 01 DNA Methylation

## 1.5. Data Analyses

Pipeline

FASTQ → 质控 (FastQC, Trim Galore)
↓
比对到参考基因组 (Bismark)
↓
提取甲基化水平 (Bismark extractor / MethylDackel)
↓
差异甲基化分析 (methylKit / DSS)
↓
功能注释与整合分析 (annotatr / clusterProfiler)

比对：Bismark



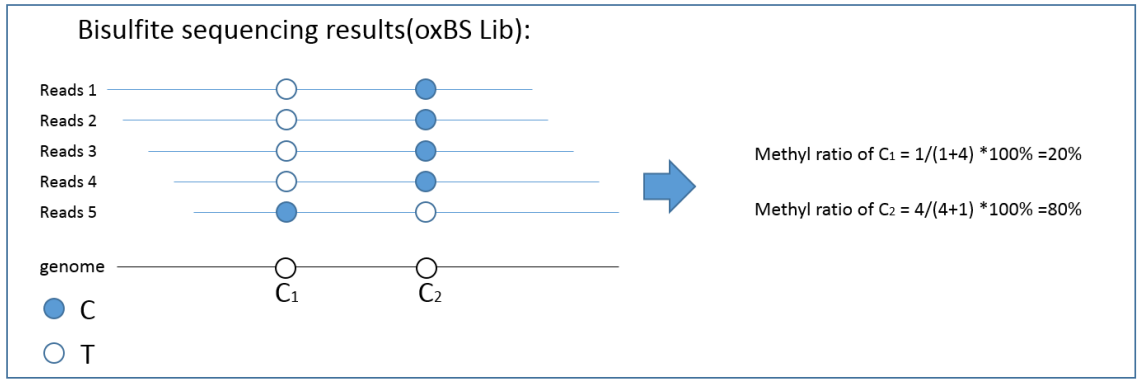
# 01 DNA Methylation

## 1.5. Data Analyses

Pipeline

FASTQ → 质控 (FastQC, Trim Galore)
↓
比对到参考基因组 (Bismark)
↓
提取甲基化水平 (Bismark extractor / MethylDackel)
↓
差异甲基化分析 (methylKit / DSS)
↓
功能注释与整合分析 (annotatr / clusterProfiler)

**甲基化位点提取**
**Bismark methylation_extractor**




- 检查测序读长中对应位置的碱基：
  - 如果是 **C**，说明在转化中它**未被转化**，即**甲基化**。
  - 如果是 **T**，说明在转化中它**被转化了**，即**未甲基化**。

**输出文件**：
•**CpG_context.txt**：包含所有CpG位点的甲基化统计。

| 染色体 | 位置 | 链 | 上下文 | 甲基化状态 | 总覆盖深度 | 甲基化读数 | 未甲基化读数 |
|---|---|---|---|---|---|---|---|
| chr1 | 3000821 | + | CG | CG | 70 | 32 | 38 |

**如果是单细胞甲基化数据呢？**

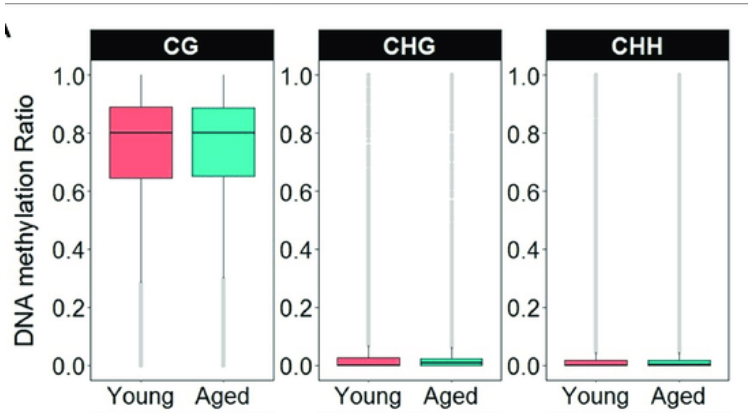
# 01 DNA Methylation

## 1.5. Data Analyses

**Bisulfite 转化效率**

WGBS 实验 BS 处理 C->U (T) 转化效率不是 100% 的，转化失败的部分 C 会成为假阳性信号，所以 BS 转化率也是很重要的指标。

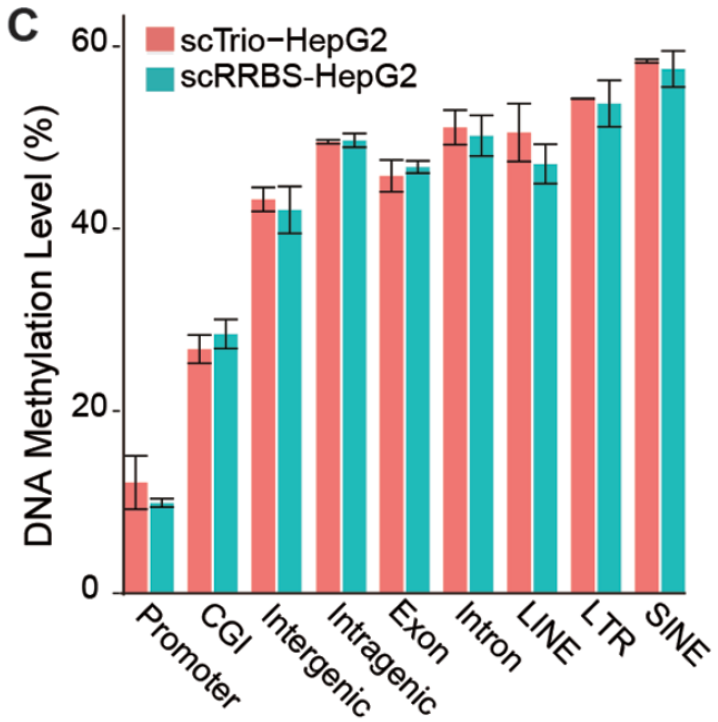
1. Lambda DNA: 计算转化率需要建库实验参入已知不含甲基化的 DNA ,
2. CHH, CHG

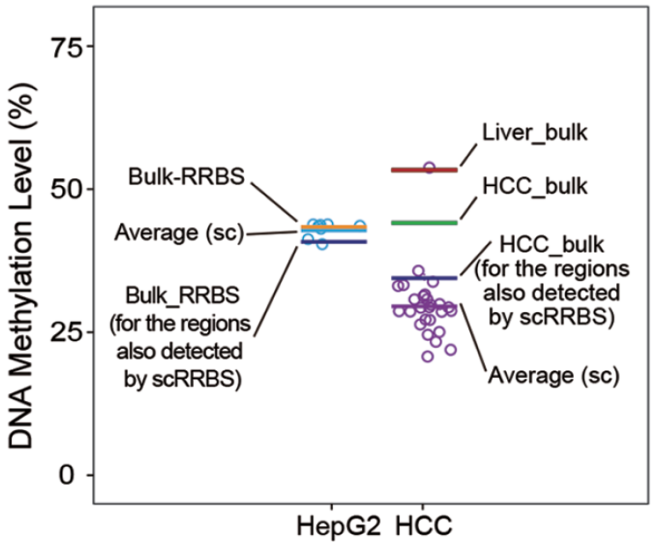


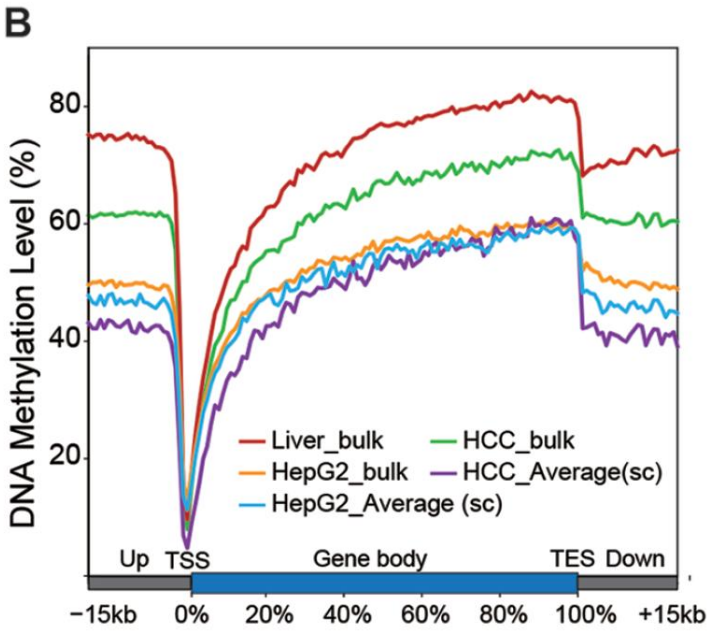
# 01 DNA Methylation

## 1.5. Data Analyses

甲基化程度计算

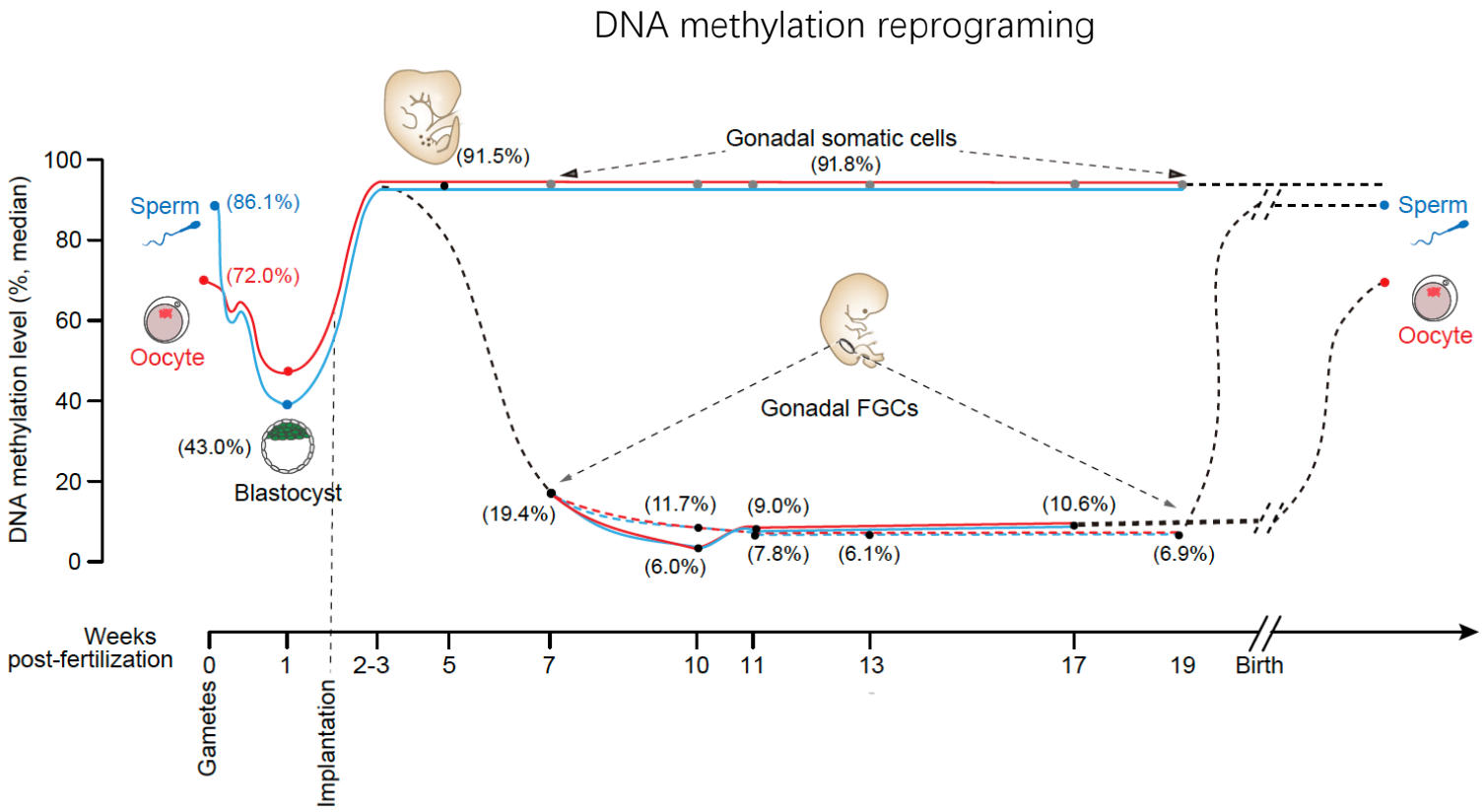


# 01 DNA Methylation

## 1.5. Data Analyses



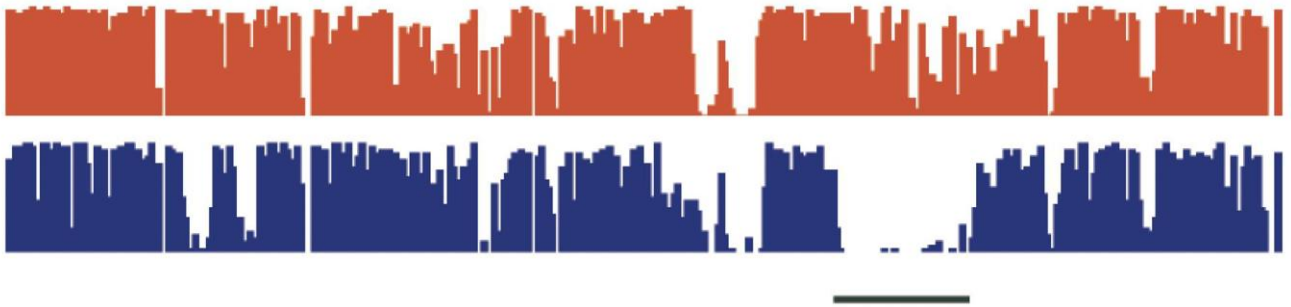

FGC: fetal germ cell

Lu Wen *et al.*, *Molecular Cell,* 2019

# 01 DNA Methylation

## 1.5. Data Analyses

甲基化程度可视化 IGV



# 01 DNA Methylation

## 1.5. Data Analyses

聚类分析

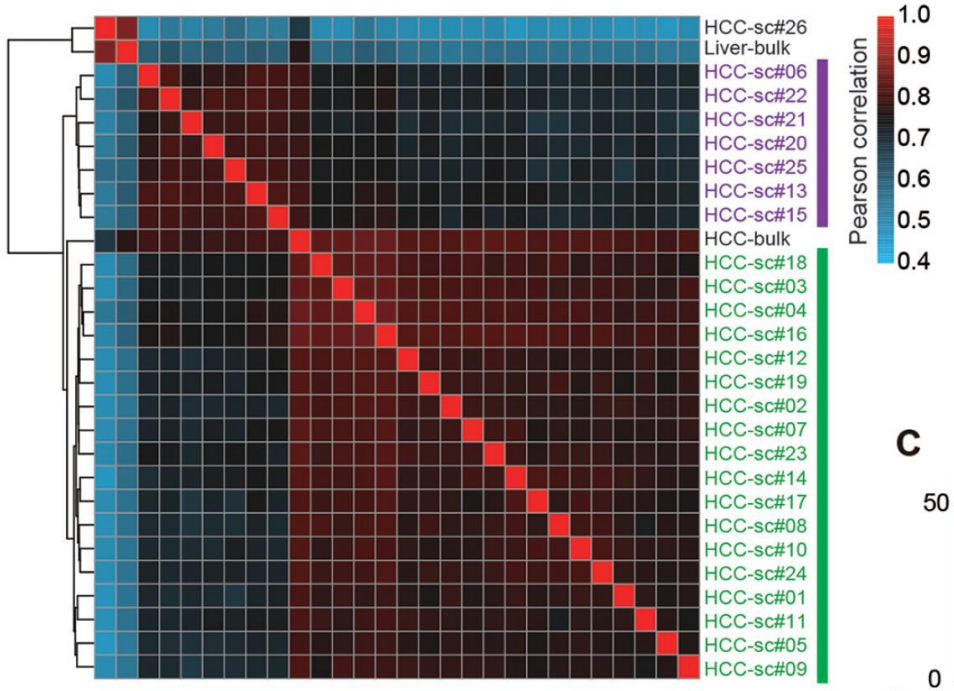
Clustering by the methylation level of :

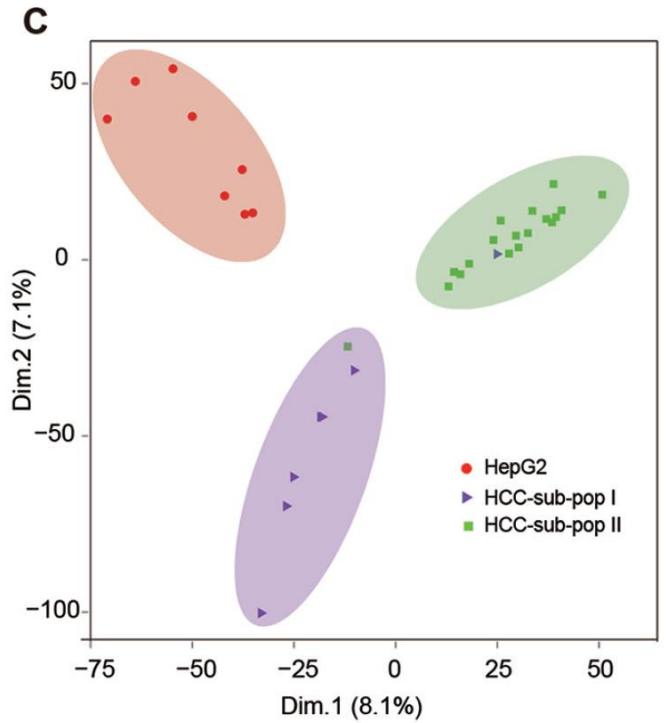
Single CpG

Genomic window (100bp, 1kb, 1M)

Promoter region

CGI

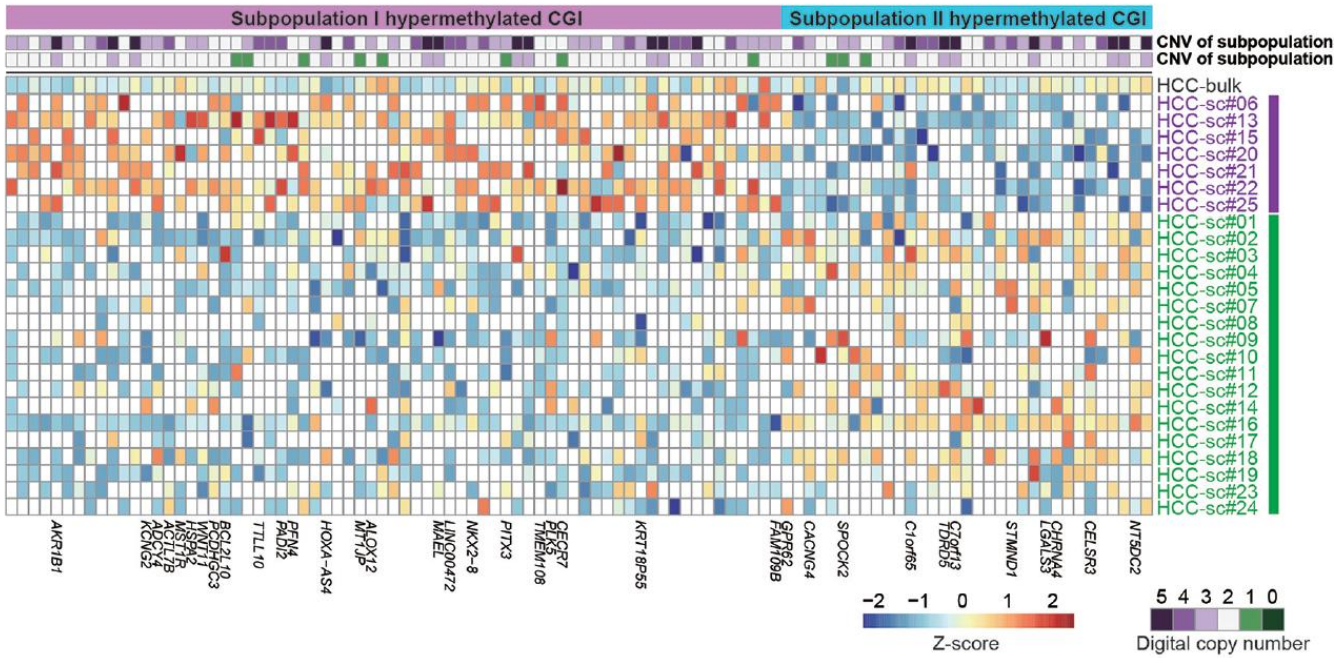


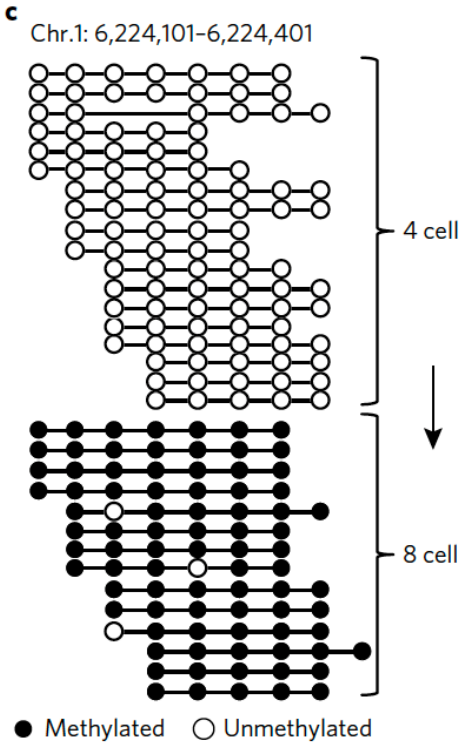



# 01 DNA Methylation

## 1.5. Data Analyses

差异甲基化区域



# Epigenetic Sequencing

## Epigenetic Sequencing

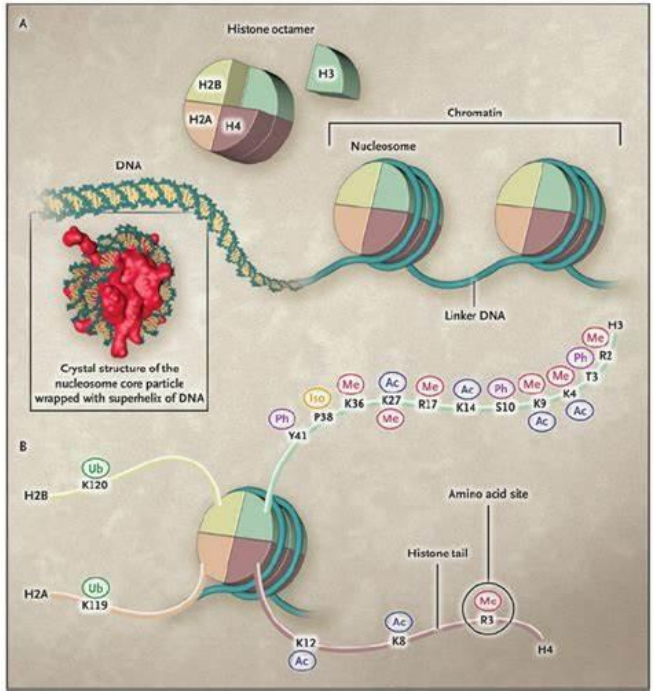
1. DNA-methylation  (WGBS, RRBS, MeDIP-seq, single-cell WGBS)
2. Histone-modification

   (ChIP-seq, CUT&RUN, CUT&Tag, single-cell CUT&Tag)
3. Chromatin accessibility (ATAC-Seq, DNase-seq, NOMe-seq, single-cell ATAC-seq)
4. Chromatin Structure (3C-seq, Hi-C-seq, single-cell Hi-C)

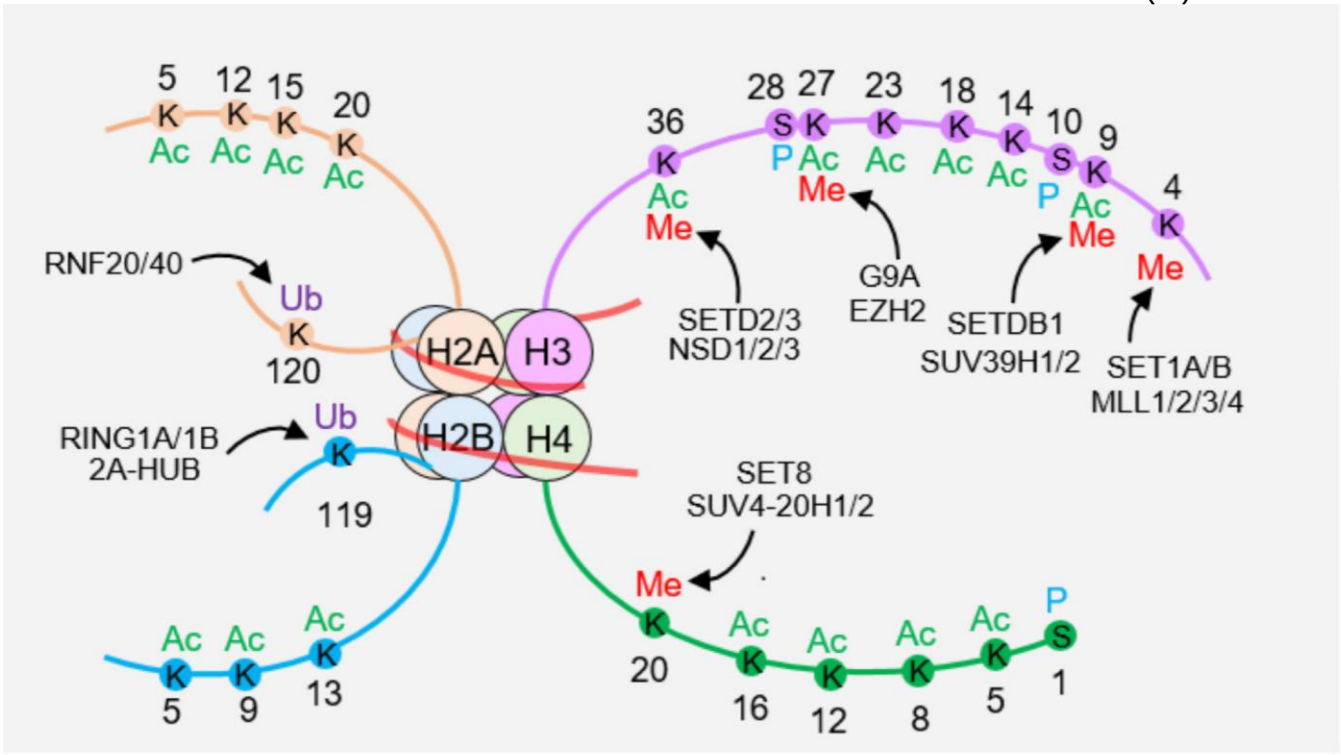
# 02 Histone Modification

## Histone Modifications

Histone modifications, such as methylation, acetylation, phosphorylation, and ubiquitination, regulate gene expression and chromatin structure.




Lysine (K)
Serine (S)



# 02 Histone Modification

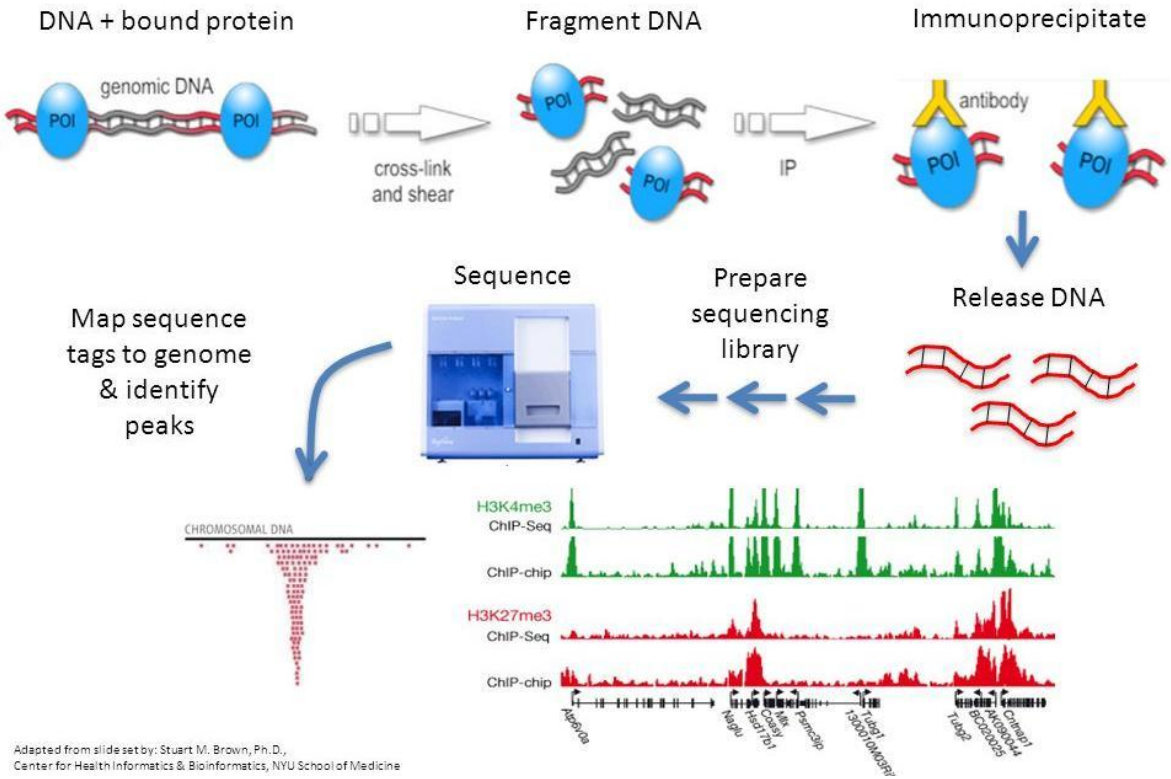
## Histone Modifications

1. **ChIP-seq** (Chromatin Immunoprecipitation Sequencing)
2. **CUT&RUN** (Cleavage Under Targets and Release Using Nuclease)
3. **CUT&TAG** (Cleavage Under Targets and Tagmentation)

# 02 Histone Modification

## 2.1. ChIP-seq

**Chromatin Immunoprecipitation Sequencing (ChIP-seq)** is a powerful technique used to study protein-DNA interactions on a genome-wide scale. It combines chromatin immunoprecipitation (ChIP) with next-generation sequencing (NGS) to identify binding sites of DNA-associated proteins, such as transcription factors, histones, or other chromatin-modifying enzymes.



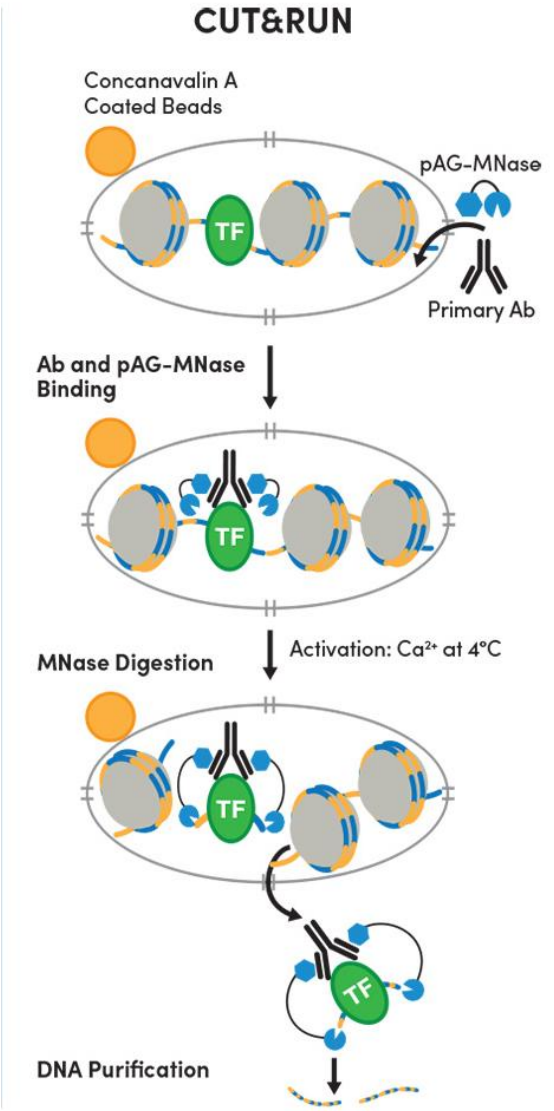

**Workflow:**

1.Chromatin crosslinking & shearing (via sonication or MNase digestion).
2.Immunoprecipitation (IP) with modification-specific antibodies.
3.Library preparation & sequencing (typically Illumina short reads).

# 02 Histone Modification

## 2.2. CUT&RUN

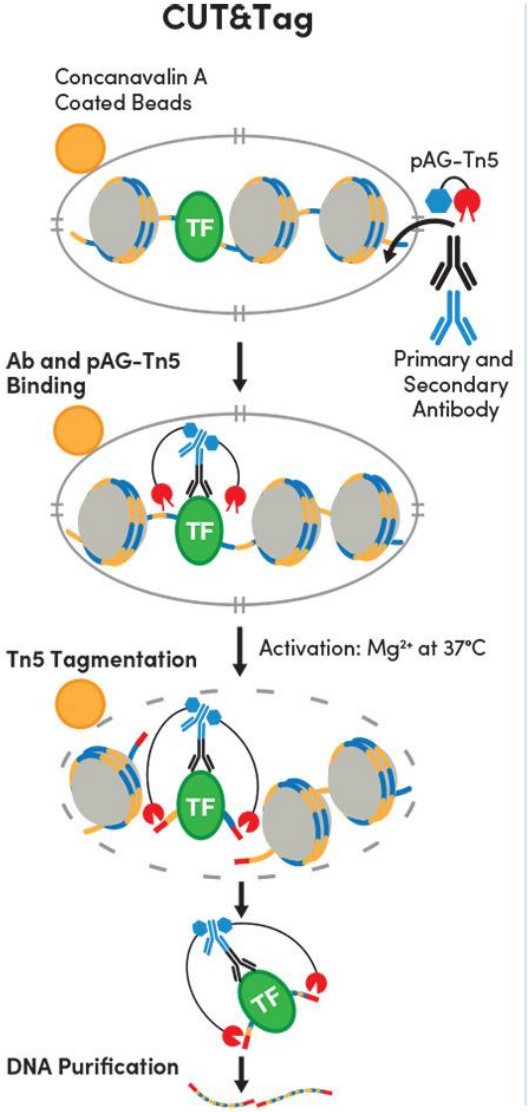
**CUT&RUN (Cleavage Under Targets & Release Using Nuclease) :** is an innovative, high-resolution method for mapping protein-DNA interactions in the genome. It is more efficient and requires fewer cells than traditional **ChIP-seq**, making it ideal for studying rare cell populations or low-abundance proteins.



# 02 Histone Modification

## 2.3. CUT&Tag

**CUT&Tag (Cleavage Under Targets & Tagmentation)** is an advanced epigenomic profiling technique that builds upon CUT&RUN, offering even higher sensitivity and scalability for mapping protein-DNA interactions. Instead of using Micrococcal Nuclease (MNase) like CUT&RUN, CUT&TAG employs a fusion protein of Protein A-Tn5 transposase (pA-Tn5) to simultaneously cleave and tag DNA at protein-binding sites.



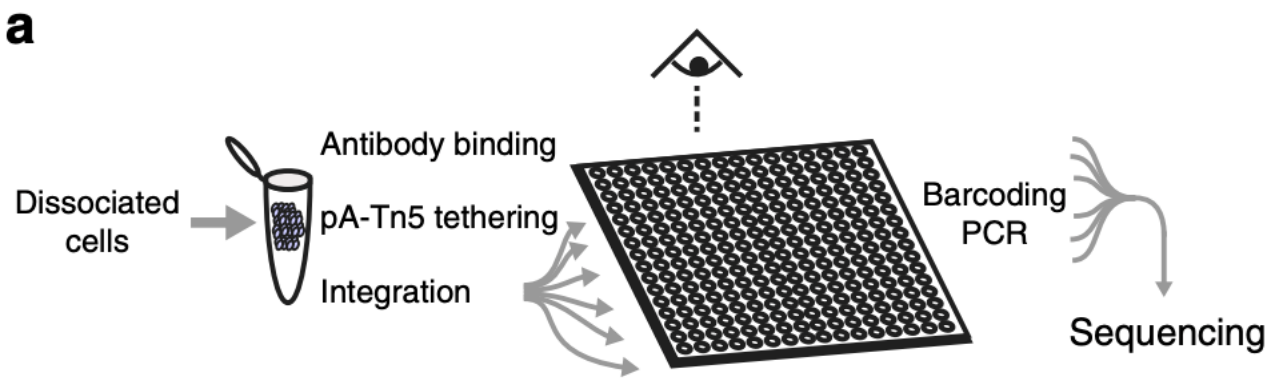
# 02 Histone Modification

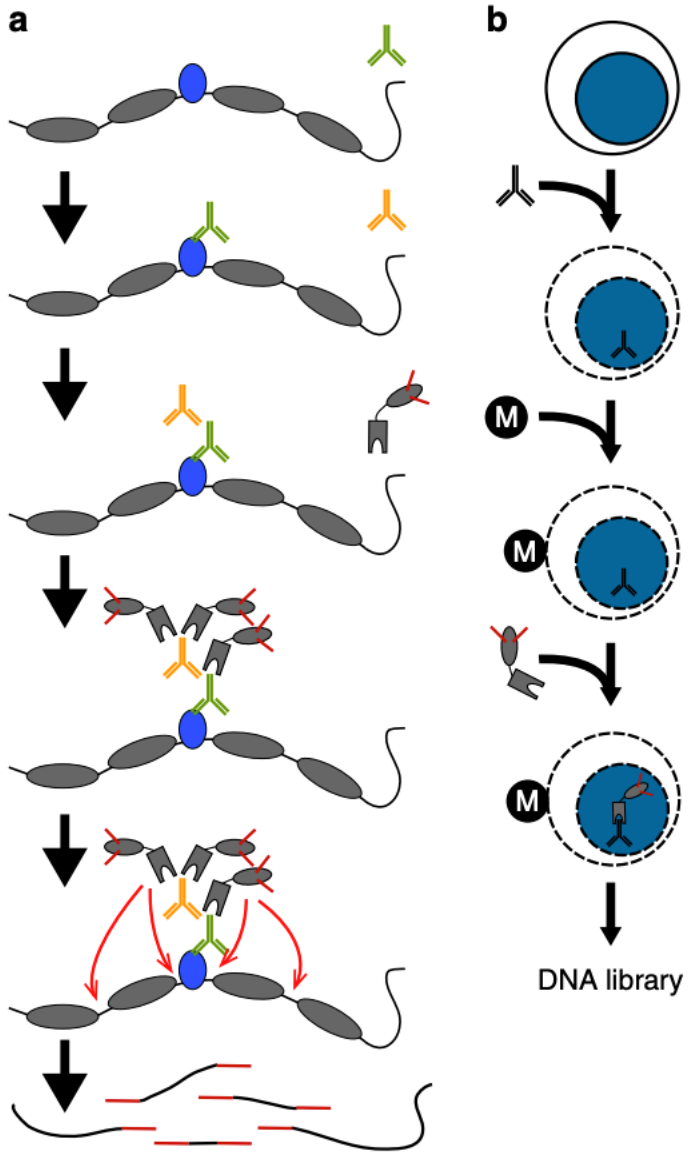
## 2.3. CUT&Tag

| Method | Input (Cells/DNA) | Resolution | Advantages | Disadvantages |
|---|---|---|---|---|
| **ChIP-seq** | High (~1M cells) | ~200 bp | Well-established, widely used | High background noise, requires crosslinking |
| **CUT&RUN** | Low (~100-1000 cells) | ~300 bp | High signal-to-noise, no crosslinking | |
| **CUT&TAG** | Ultra-low (~10-100 cells) | ~100 bp | Direct tagmentation, high resolution | Tn5 bias |

# 02 Histone Modification

## 2.4. single cell CUT&Tag



# 02 Histone Modification

## 2.5. Data analyses

FASTQ → 质控 (FastQC, Trim Galore)
↓
比对到参考基因组 (bowtie2)
↓
去除PCR duplicate （samtools markdup)
↓
提取修饰peak (MACS2)

质控

```
fastqc sample_R1.fastq.gz sample_R2.fastq.gz -o qc/
multiqc qc/ -o qc_report/
```

测序质量、接头污染、序列重复水平、GC 偏差、过低质量的 bases。

# 02 Histone Modification

## 2.5. Data analyses

FASTQ → 质控 (FastQC, Trim Galore)
↓
比对到参考基因组 (bowtie2)
↓
去除PCR duplicate （samtools markdup)
↓
提取修饰peak (MACS2)

比对到参考基因组

```
bowtie2 -x hg38_index -1 trimmed_R1.fq.gz -2
trimmed_R2.fq.gz -S sample.sam --threads 8

samtools view -bS sample.sam > sample.bam
samtools sort -o sample.sorted.bam sample.bam
samtools index sample.sorted.bam
```

# 02 Histone Modification

## 2.5. Data analyses

FASTQ → 质控 (FastQC, Trim Galore)
↓
比对到参考基因组 (bowtie2)
↓
去除PCR duplicate （samtools markdup)
↓
提取修饰peak (MACS2)

去除PCR duplicate

```
picard MarkDuplicates I=sample.sorted.mapq30.bam
O=sample.dedup.bam M=dup_metrics.txt REMOVE_DUPLICATES=true
# 去除 chrM
samtools idxstats sample.dedup.bam | cut -f1 | grep -v chrM | xargs
samtools view -b sample.dedup.bam > sample.filtered.bam
# 排除 blacklist
bedtools intersect -v -abam sample.filtered.bam -b hg38_blacklist.bed >
sample.clean.bam
samtools index sample.clean.bam
```

# 02 Histone Modification

## 2.5. Data analyses

FASTQ → 质控 (FastQC, Trim Galore)
↓
比对到参考基因组 (bowtie2)
↓
去除PCR duplicate （samtools markdup)
↓
提取修饰peak (MACS2)

Peak calling

窄峰（H3K4me3、H3K27ac 等）：MACS2 callpeak --nomodel --extsize/--shift 或默认模式。

宽峰（H3K27me3、H3K36me3）：MACS2 --broad、SICER、epic2

```
macs2 callpeak -t ChIP.bam -c Input.bam -f BAM -g hs -n sample_H3K27ac -q 0.01 --outdir macs2_out
```

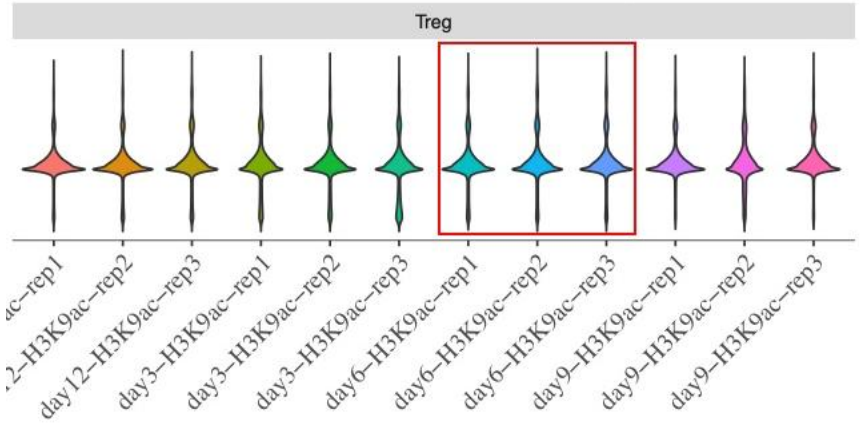
```
macs2 callpeak -t ChIP.bam -c Input.bam -f BAM -g hs --broad -n sample_H3K27me3 --broad-cutoff 0.1
```

# 02 Histone Modification

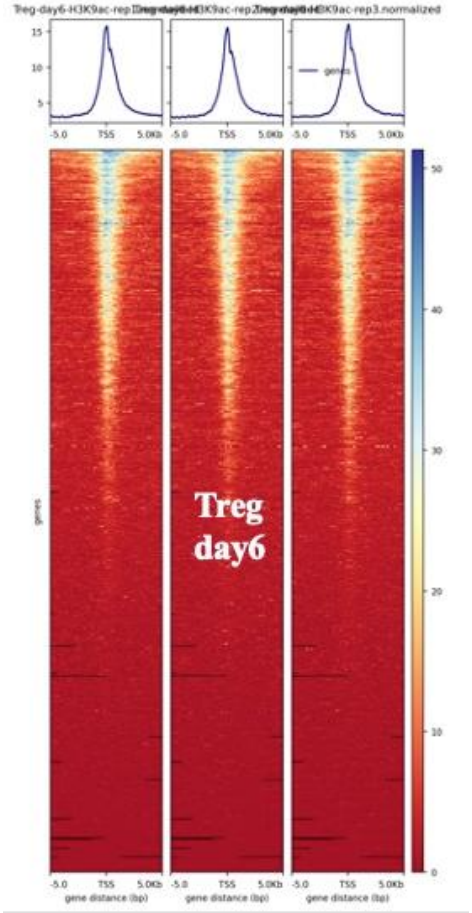
## 2.5. Data analyses

质控

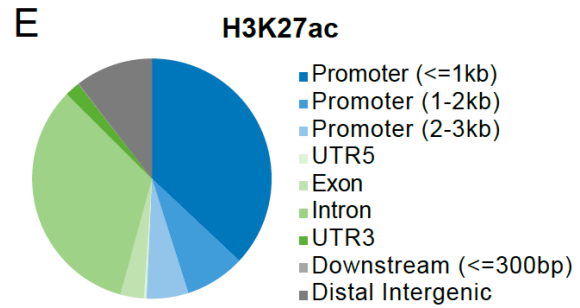
Peak 长度




Peak 分布




Peak 分布

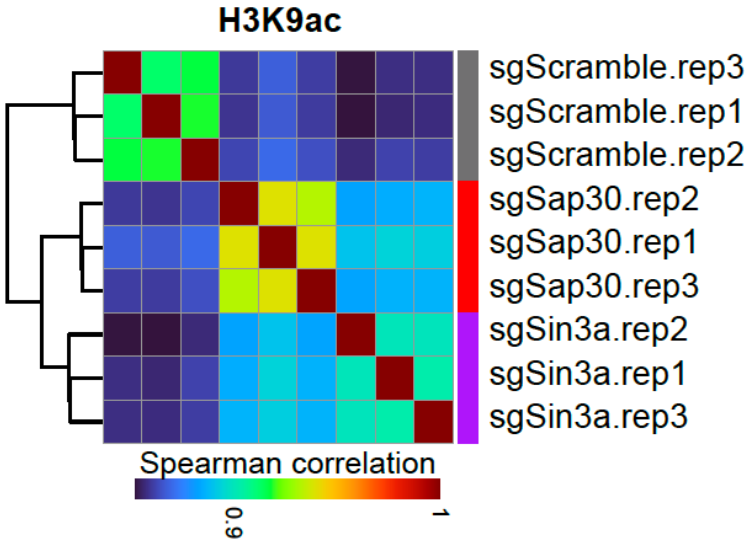


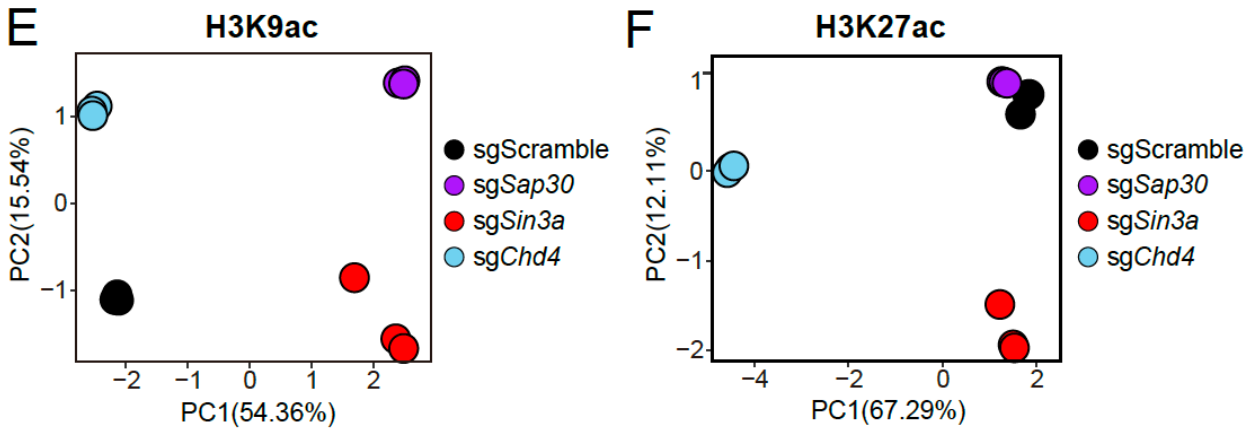
# 02 Histone Modification

## 2.5. Data analyses

样品聚类

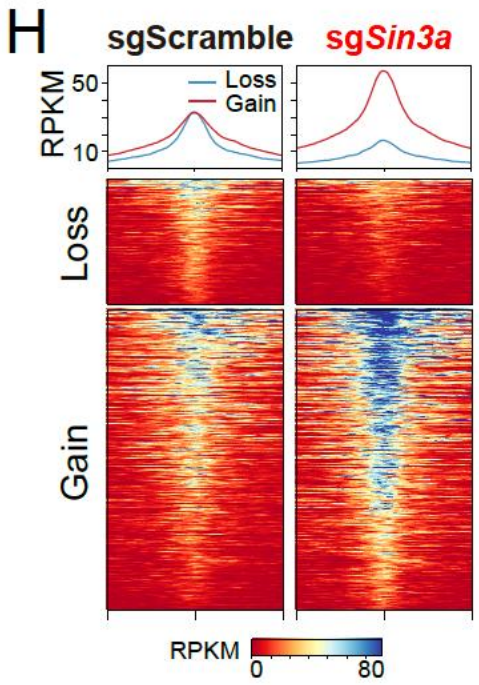
按照peak 或 promoter 修饰程度均可聚类



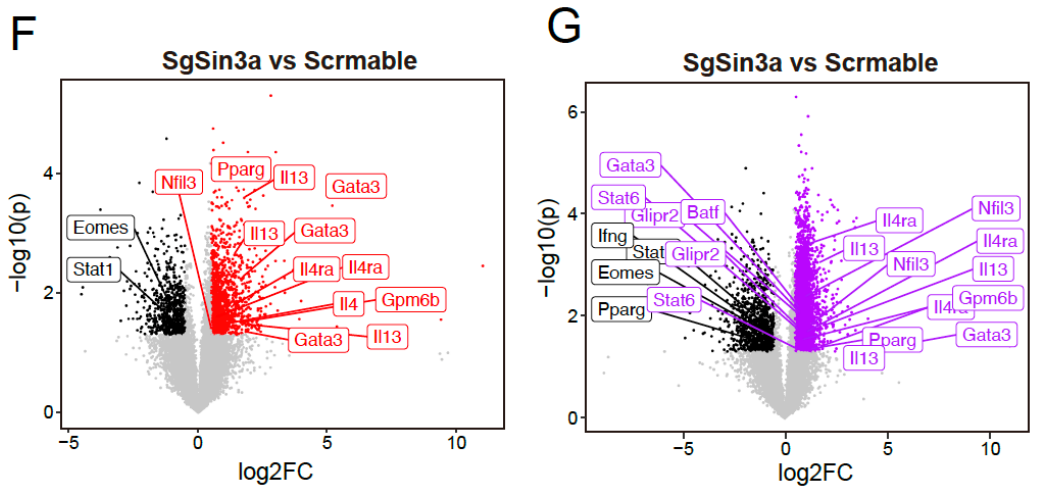



## 2.5. Data analyses

差异peak



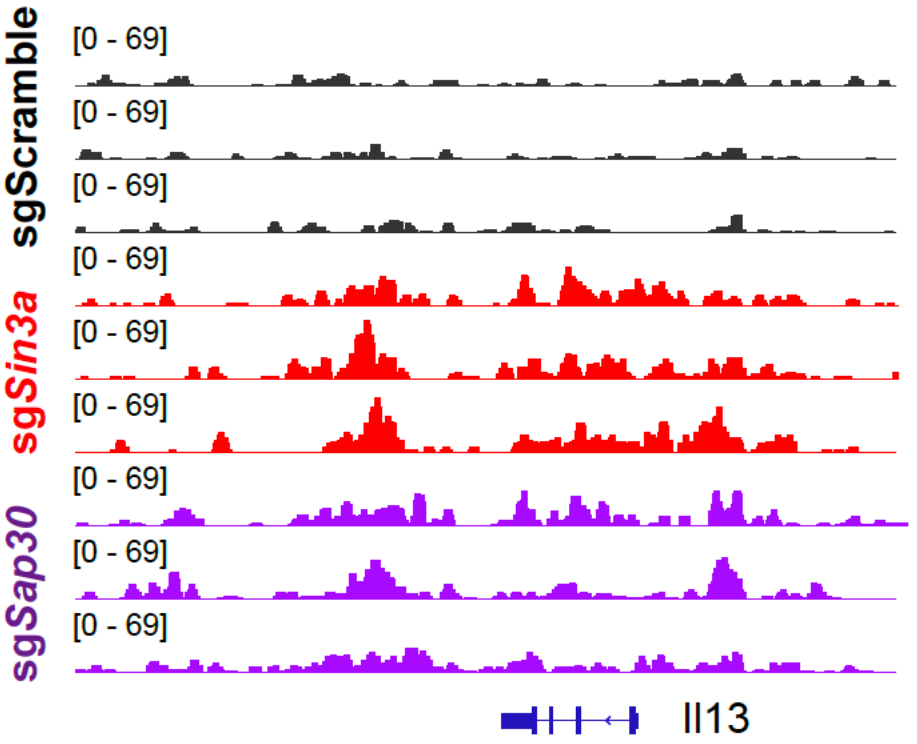

差异peak注释



# 02 Histone Modification

## 2.5. Data analyses

IGV 可视化



# Epigenetic Sequencing

## Epigenetic Sequencing

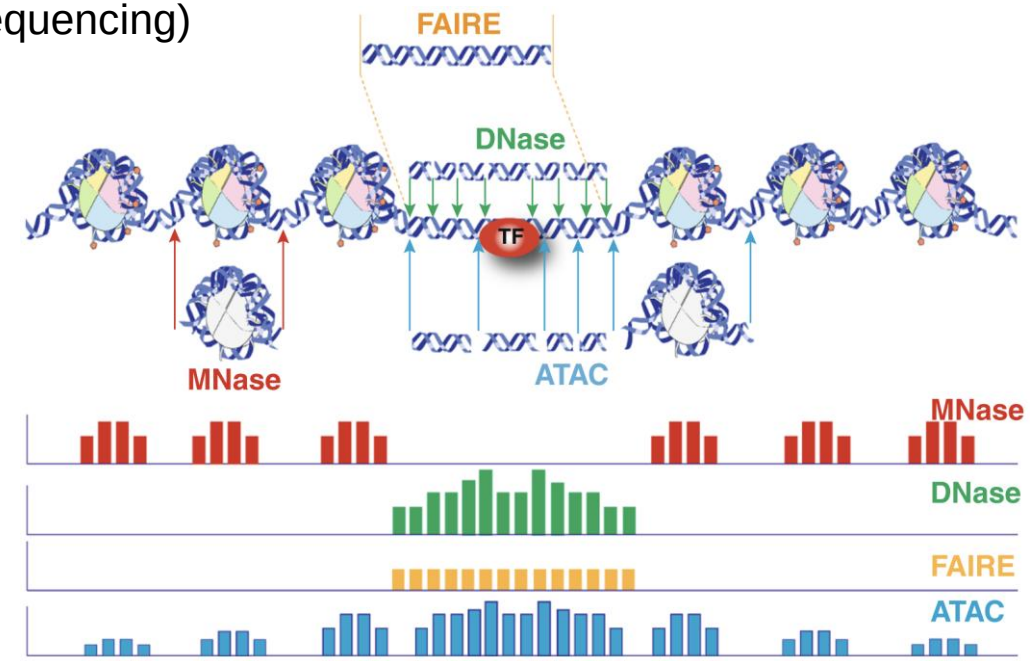
1. DNA-methylation  (WGBS, RRBS, MeDIP-seq, single-cell WGBS)
2. Histone-modification

   (ChIP-seq, CUT&RUN, CUT&Tag, single-cell CUT&Tag)
3. Chromatin accessibility (ATAC-Seq, DNase-seq, MNase-seq, single-cell ATAC-seq)
4. Chromatin Structure (3C-seq, Hi-C-seq, single-cell Hi-C)

# 03 Chromatin accessibility

## Chromatin accessibility

1. **ATAC-seq** (Assay for Transposase-Accessible Chromatin with Sequencing)
2. **MNase-seq** (Micrococcal Nuclease Sequencing)
3. **DNase-seq** (DNase I Hypersensitive Sites Sequencing)
4. **scATAC-seq** (Single-Cell ATAC-seq)



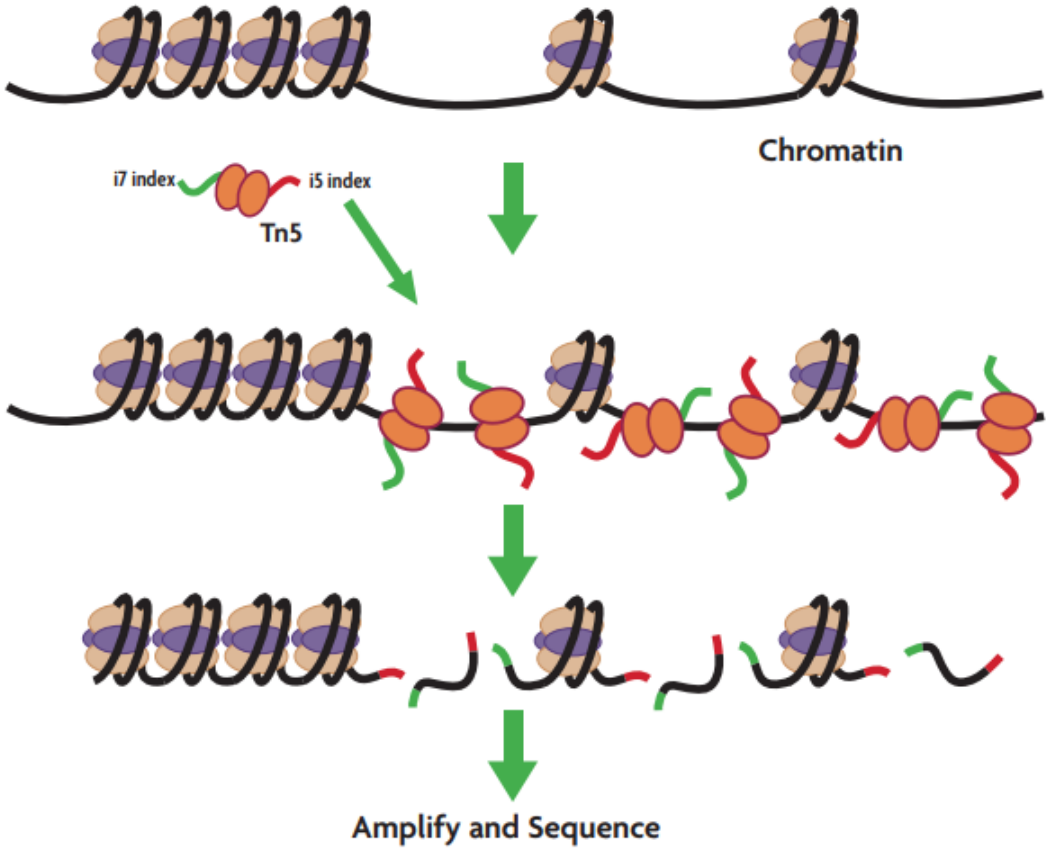

**Open (accessible) chromatin** → easier for proteins to bind → **active gene expression**

**Closed (inaccessible) chromatin** → tightly packed around histones → **gene repression**

# 03 Chromatin accessibility

## 3.1 ATAC-seq

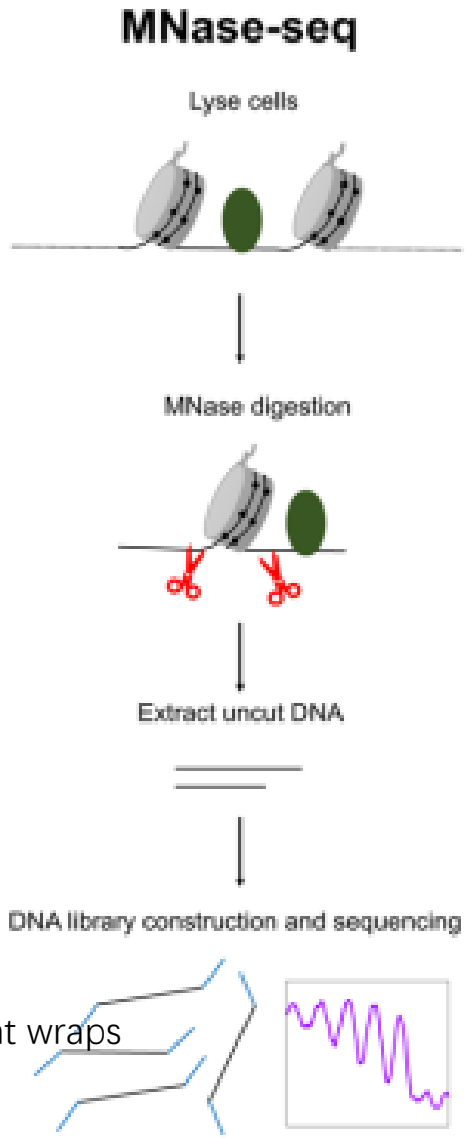
**ATAC-seq** (Assay for Transposase-Accessible Chromatin using Sequencing) is a powerful technique to profile genome-wide chromatin accessibility, revealing regions of open chromatin associated with active regulatory elements (e.g., enhancers, promoters). It uses a hyperactive Tn5 transposase to simultaneously cut and tag accessible DNA, followed by sequencing.



# 03 Chromatin accessibility

## 3.2 MNase-seq

**MNase-seq （Micrococcal Nuclease Sequencing)** is a technique used to study nucleosome positioning and chromatin structure by digesting linker DNA between nucleosomes, leaving protected DNA wrapped around histones for sequencing. Unlike ATAC-seq (which profiles open chromatin), MNase-seq provides detailed information about nucleosome occupancy, spacing, and histone-DNA interactions.

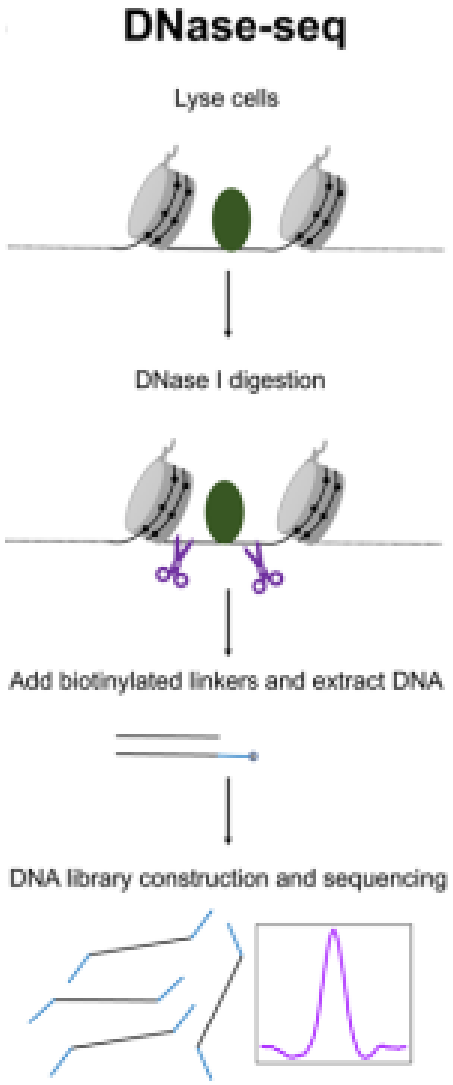



**147 base pairs (bp)** of DNA that wraps around this histone core

# 03 Chromatin accessibility

## 3.3 DNase-seq

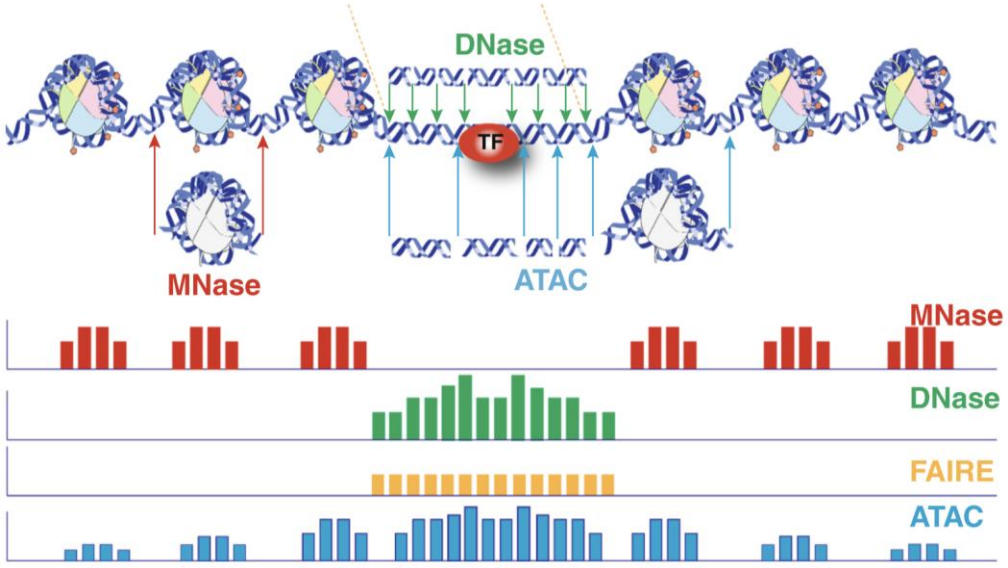
**DNase-seq** (DNase I Hypersensitive Sites Sequencing) is a high-resolution sequencing method used to identify open chromatin regions in the genome, also known as DNase I hypersensitive sites (DHSs). These sites are associated with active regulatory elements like promoters, enhancers, insulators, and other cis-regulatory regions.



# 03 Chromatin accessibility

## 3.3 DNase-seq

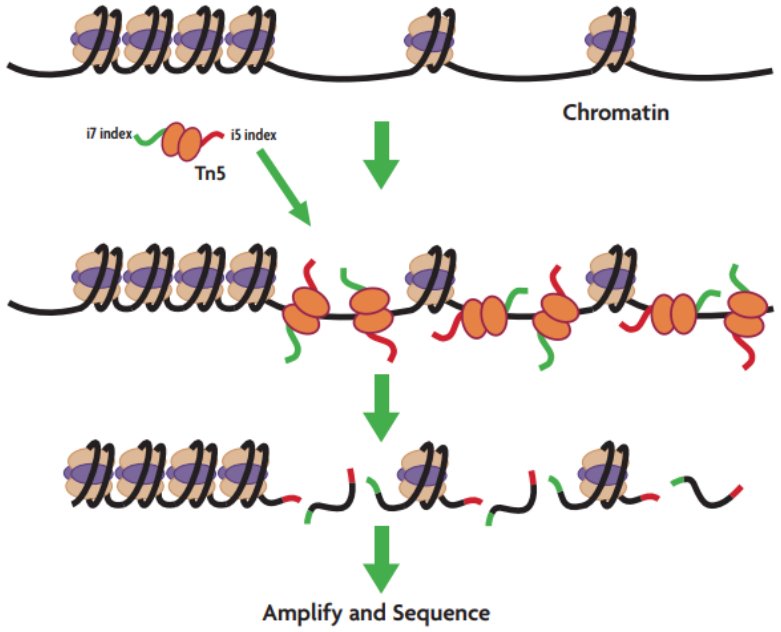
| Method | Main Purpose | Enzyme Used | Resolution | Input Required | Key Application |
|---|---|---|---|---|---|
| **ATAC-seq** | Chromatin accessibility profiling | Tn5 transposase | High (~100 bp) | Low (~50K cells) | Identifying open chromatin regions, TF binding sites |
| **MNase-seq** | Nucleosome positioning & chromatin compaction | Micrococcal nuclease (MNase) | High (~150 bp) | Medium (~1M cells) | Mapping nucleosome positioning |
| **DNase-seq** | Chromatin accessibility & TF footprinting | DNase I | High (~200 bp) | High (~10M cells) | Identifying DNase hypersensitive sites (DHSs) |



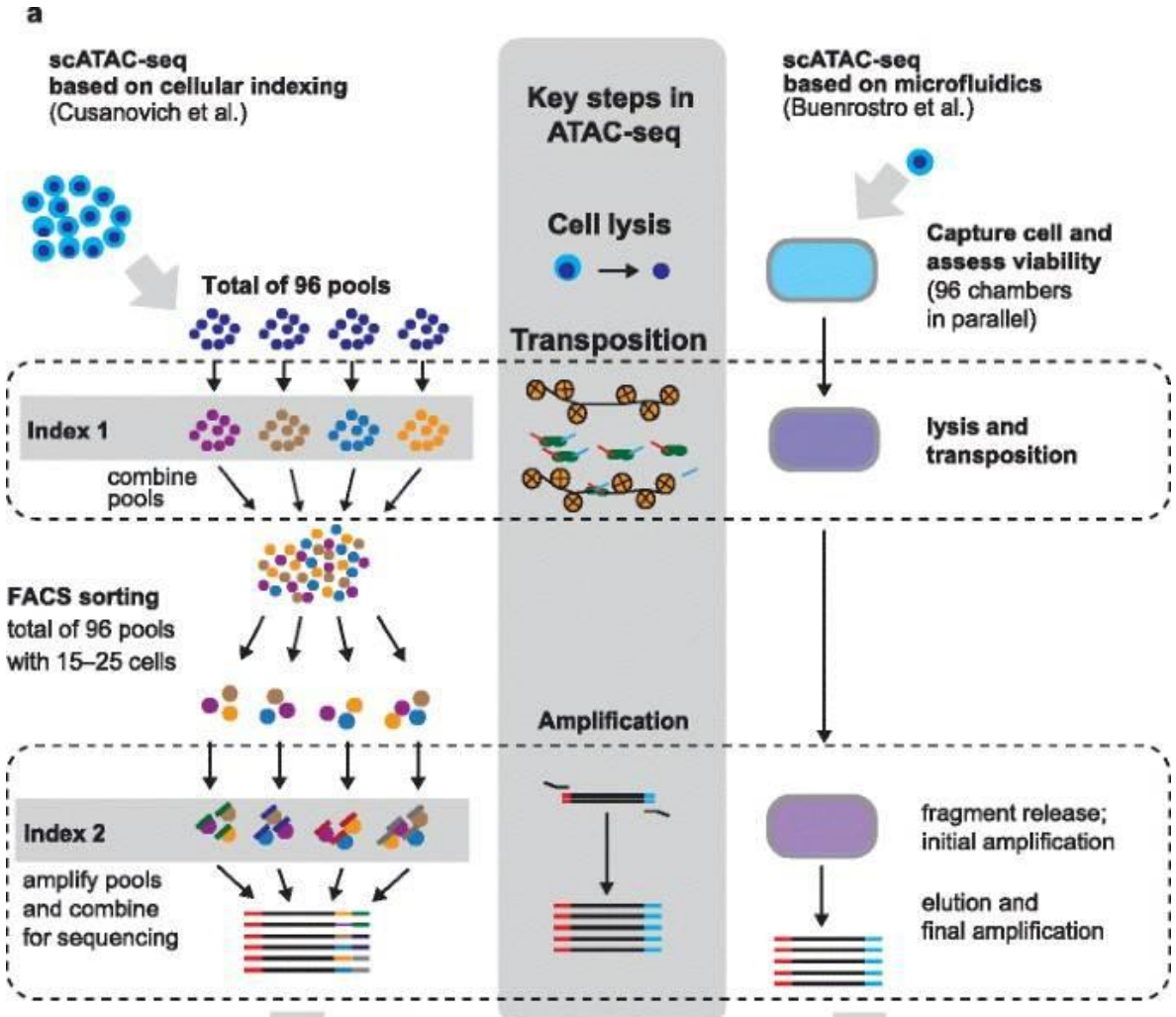
# 03 Chromatin accessibility

## 3.4 scATAC-seq

**Single-cell ATAC-seq (scATAC-seq)** is a powerful method for profiling chromatin accessibility at the single-cell resolution.

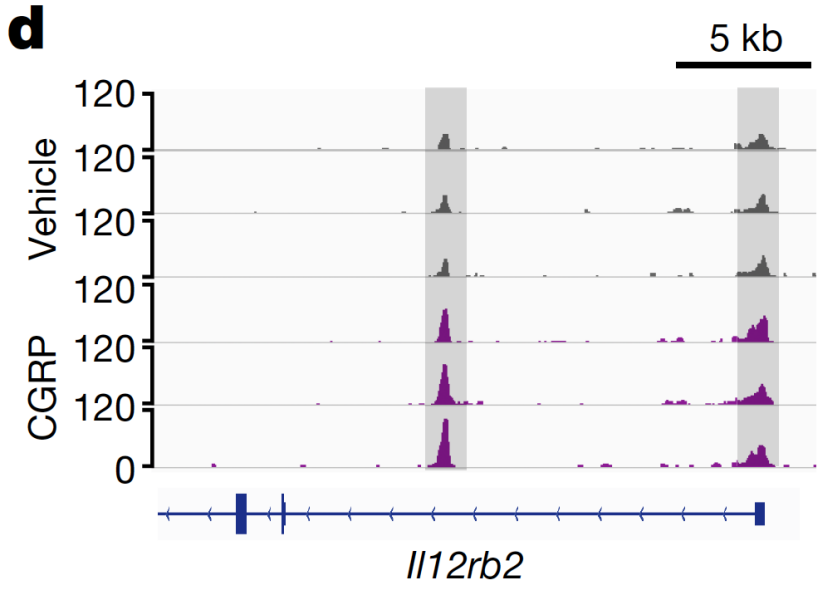


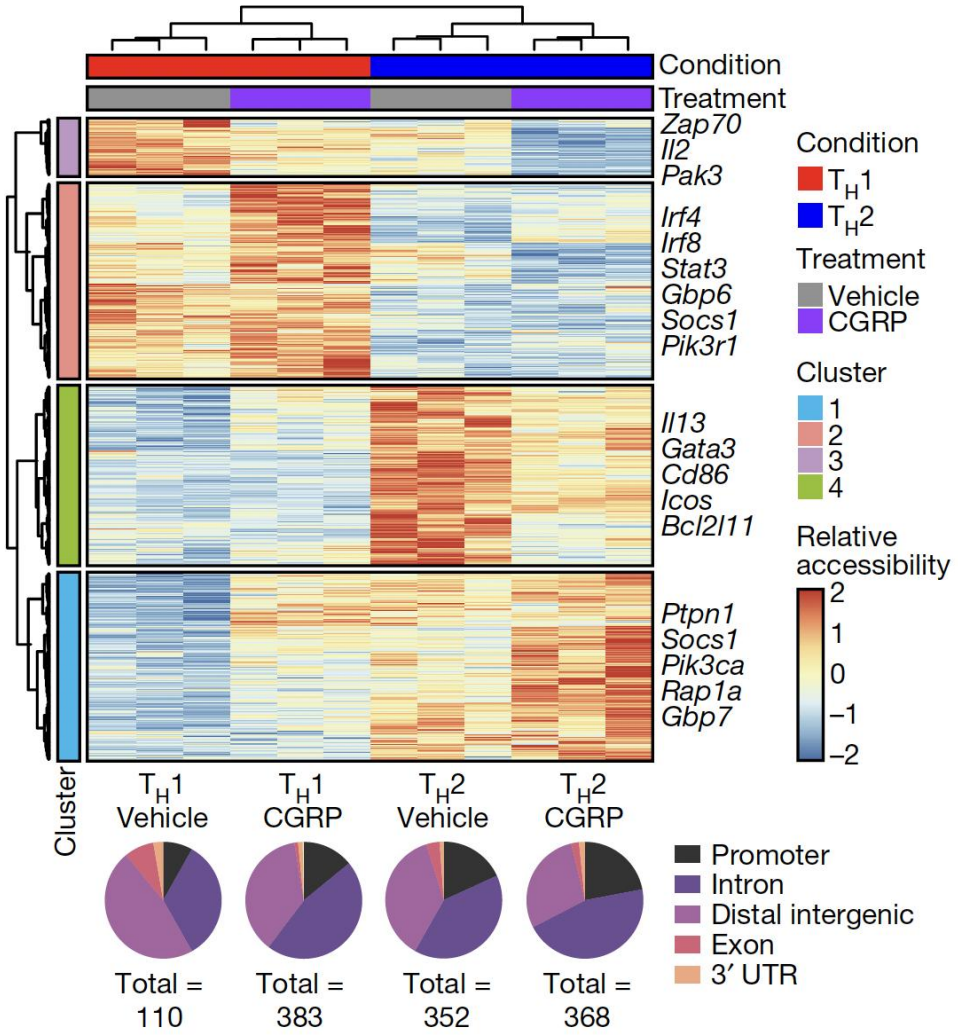

单细胞unique的标签组合



# 03 Chromatin accessibility

## 3.5 Data Analyses (ATAC-seq)



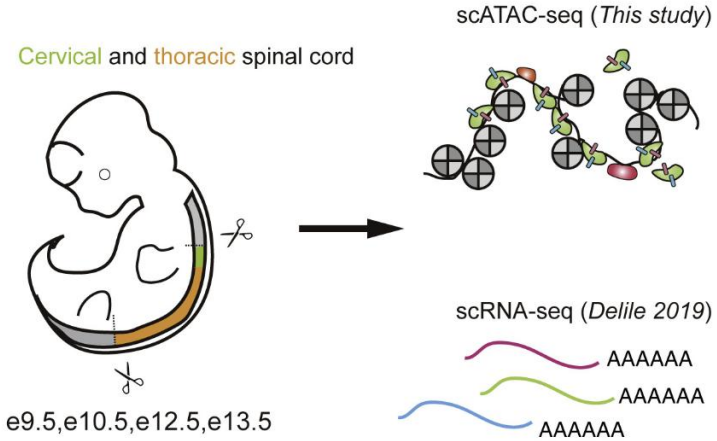



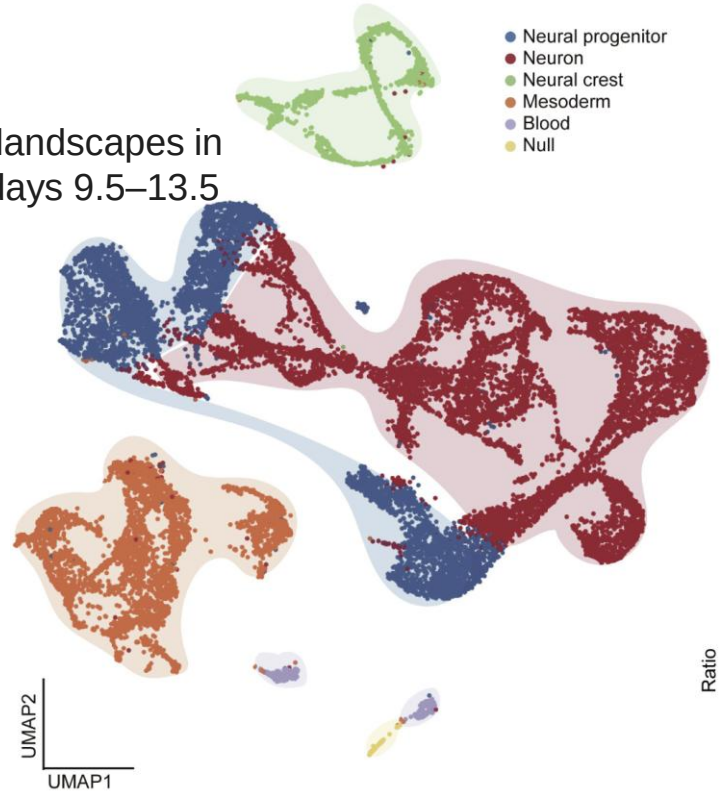

Yu Hou et al., *Nature*, 2024

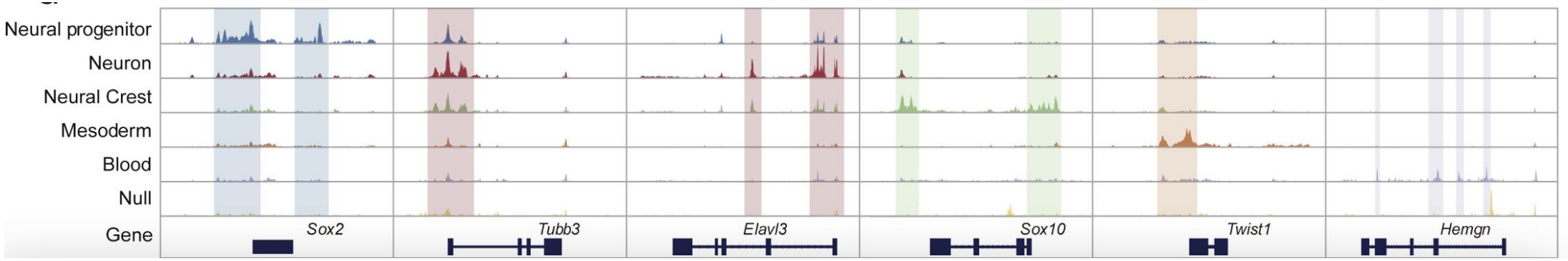
# 03 Chromatin accessibility

## 3.5 Data Analyses (single cell ATAC-seq)

profiled single-cell chromatin accessibility landscapes in mouse neural tubes spanning embryonic days 9.5–13.5




Muya Shu et al., *Developmental Cell*, 2022

# Epigenetic Sequencing

## Epigenetic Sequencing

1. DNA-methylation  (WGBS, RRBS, MeDIP-seq, single-cell WGBS)
2. Histone-modification

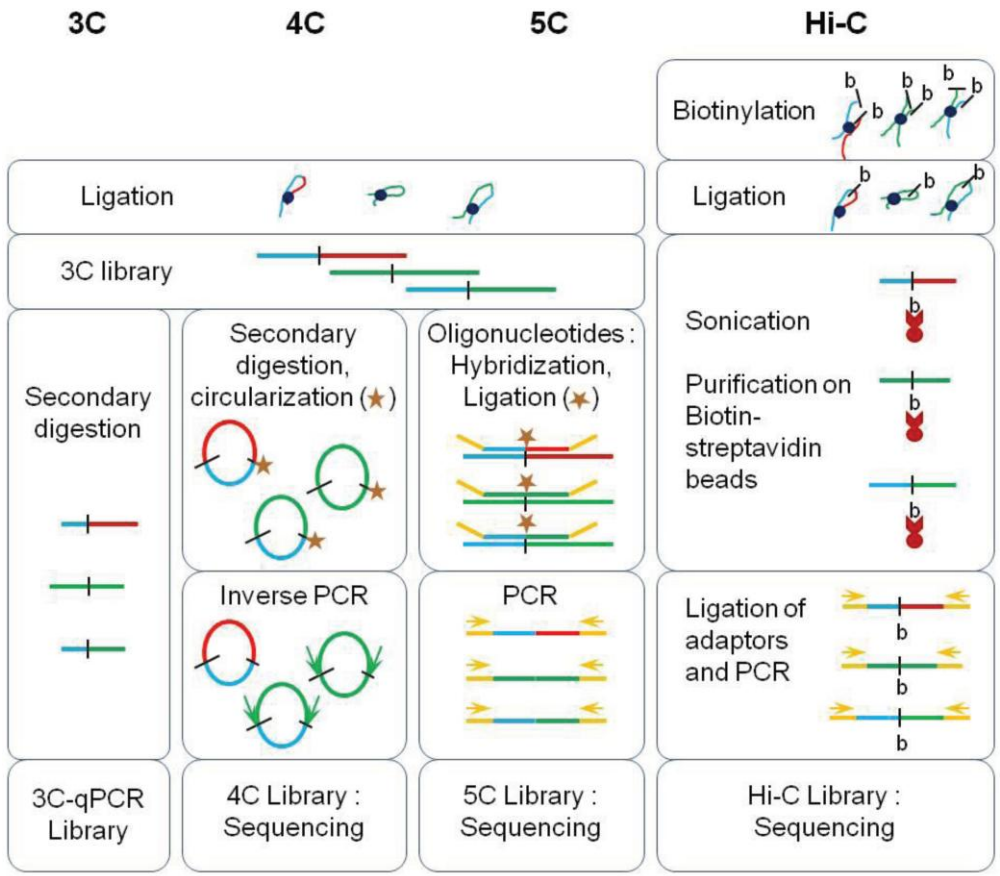
   (ChIP-seq, CUT&RUN, CUT&Tag, single-cell CUT&Tag)
3. Chromatin accessibility (ATAC-Seq, DNase-seq, MNase-seq, single-cell ATAC-seq)
4. Chromatin Structure (3C, 4C, 5C, Hi-C, single-cell Hi-C)

# 04 Chromatin Structure

## Chromatin Structure

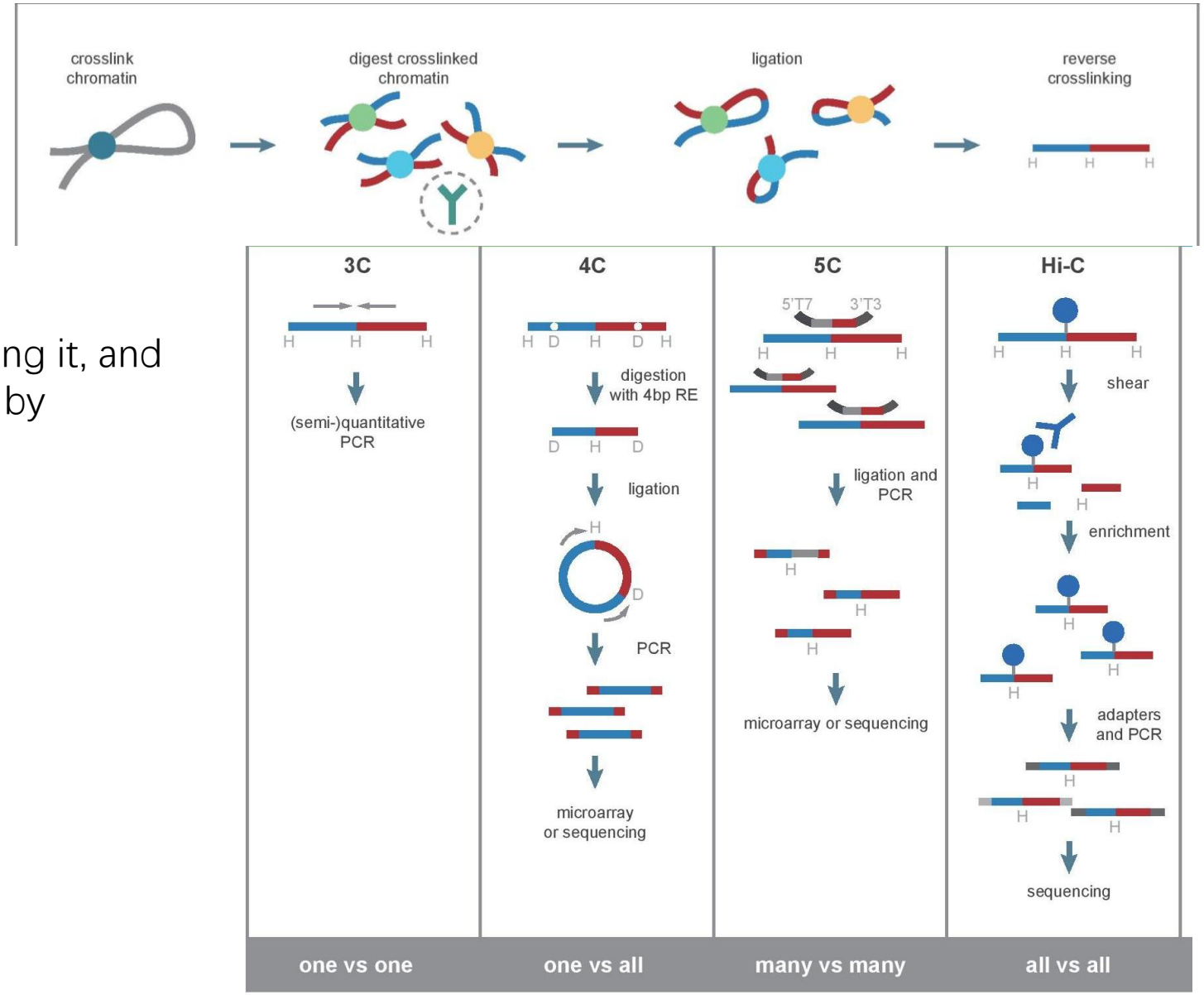
1. **3C** (Chromosome Conformation Capture)
2. **4C** (Circular Chromosome Conformation Capture)
3. **5C** (Chromosome Conformation Capture Carbon Copy)
4. **Hi-C** (High-Throughput Chromosome Conformation Capture)

Chromatin structure sequencing techniques are used to study the 3D organization of the genome, which plays a crucial role in gene regulation, enhancer-promoter interactions, and overall chromatin dynamics.



# 04 Chromatin Structure

## 4.1 3C vs 4C vs 5C




These methods rely on crosslinking DNA, digesting it, and re-ligating interacting DNA fragments, followed by sequencing to infer chromatin interactions.

# 04 Chromatin Structure

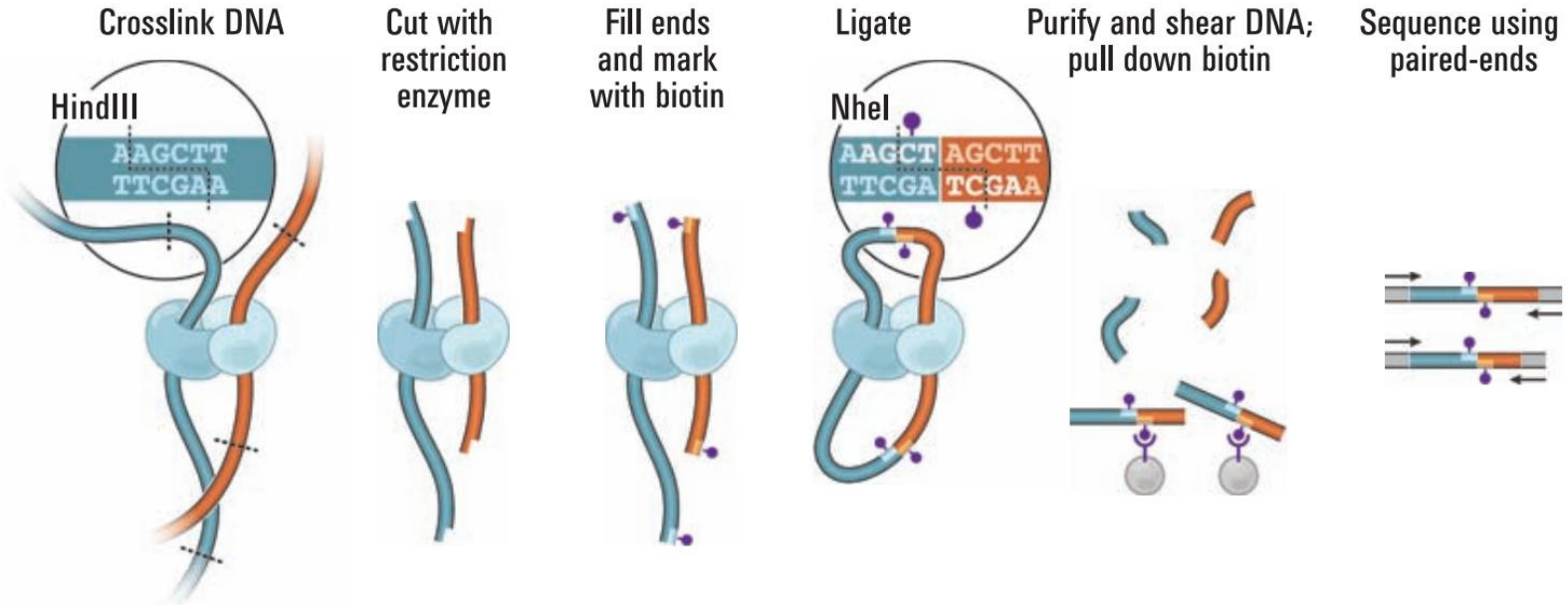
## 4.1 3C vs 4C vs 5C

| Technique | 3C (Chromosome Conformation Capture) | 4C (Circular Chromosome Conformation Capture) | 5C (Chromosome Conformation Capture Carbon Copy) |
|---|---|---|---|
| **Main Purpose** | Measures physical interactions between two genomic loci | Identifies all interactions of a specific locus | Captures interactions between multiple loci |
| **Interaction Type** | One vs. one | One vs. all | Many vs. many |
| **Genome Coverage** | Targeted (single locus pair) | Genome-wide (from a single viewpoint) | Targeted (specific genomic region) |
| **Best Used For** | Studying specific interactions (e.g., enhancer-promoter loops) | Finding all possible contacts of a specific locus (e.g., enhancer hubs) | Mapping chromatin architecture in specific regions (e.g., TADs) |

# 04 Chromatin Structure

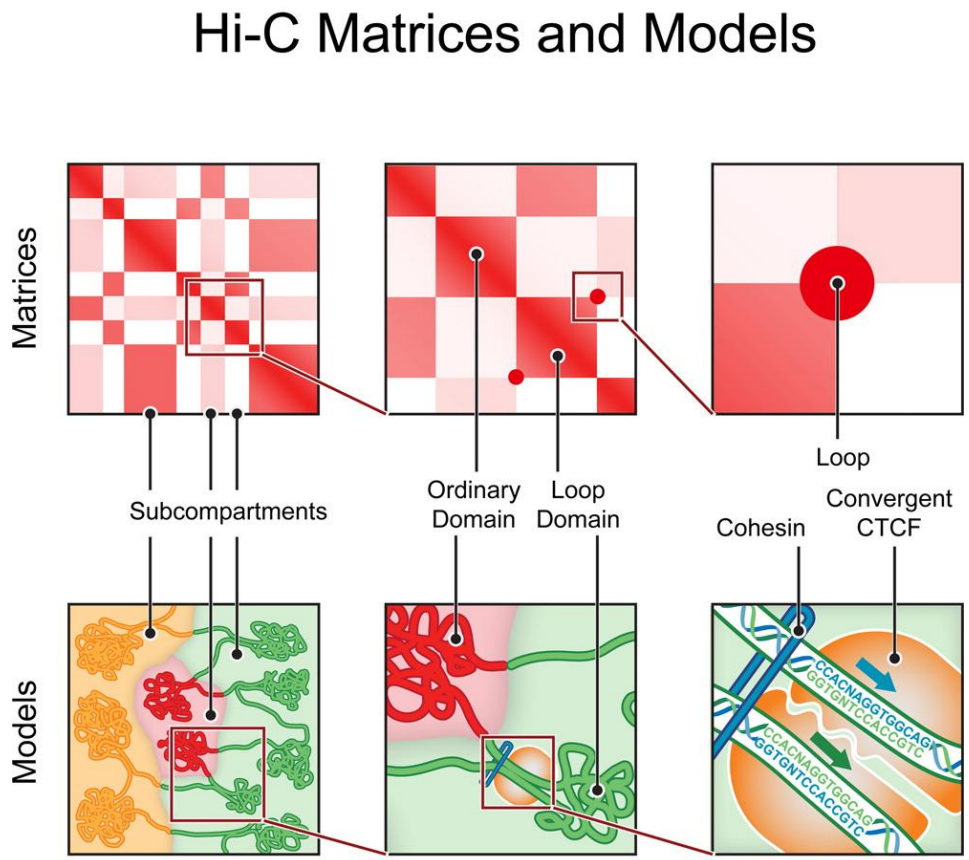
## 4.2 Hi-C

Hi-C is a genome-wide chromosome conformation capture (3C) technique that identifies chromatin interactions across the entire genome. It provides insights into 3D genome organization, revealing features such as topologically associating domains (TADs), chromatin loops, and A/B compartments.



# 04 Chromatin Structure

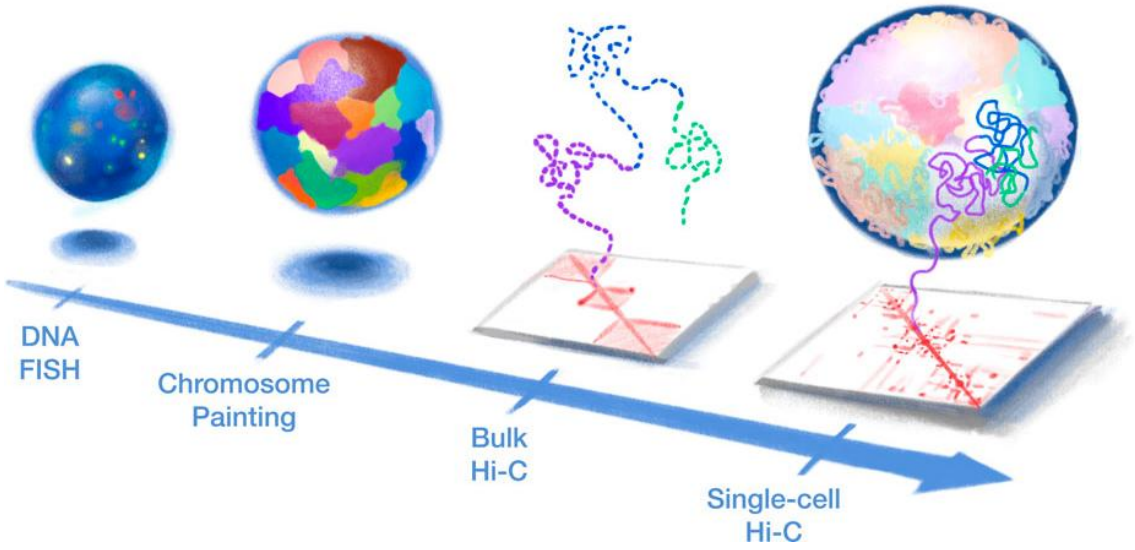
## 4.2 Hi-C

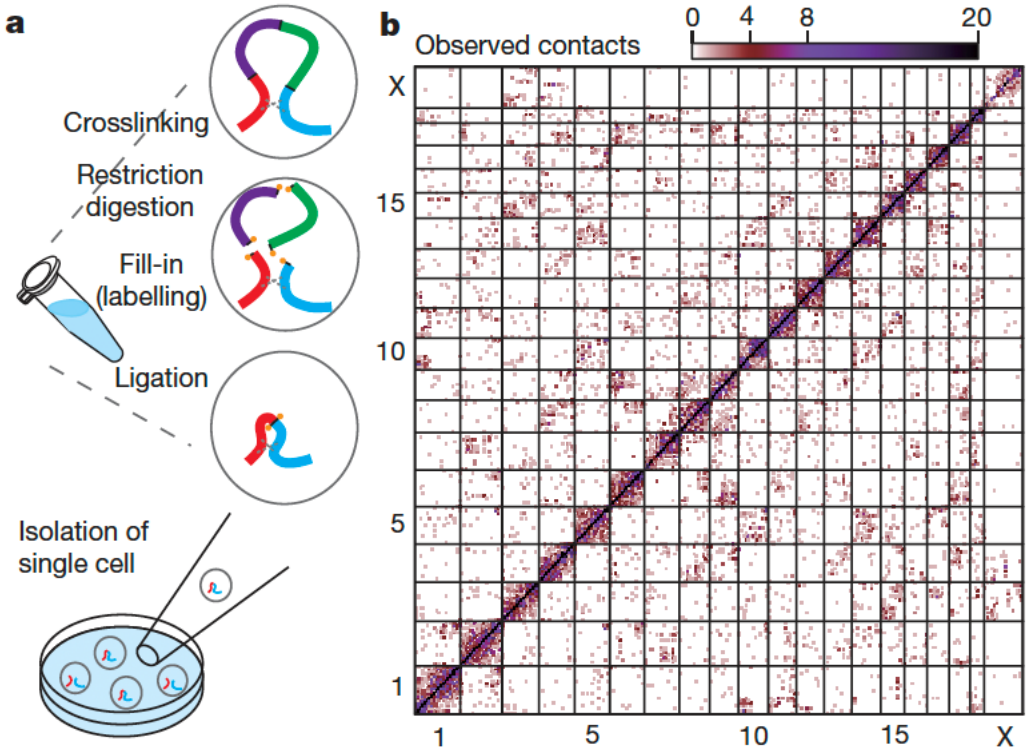



Compartment A: 包含的基因数目较多，基因表达较高，区域开放，转录激活的区域
Compartment B:包含的基因数目较少，基因表达较低，转录沉默的区域

# 04 Chromatin Structure

## 4.3 single cell Hi-C



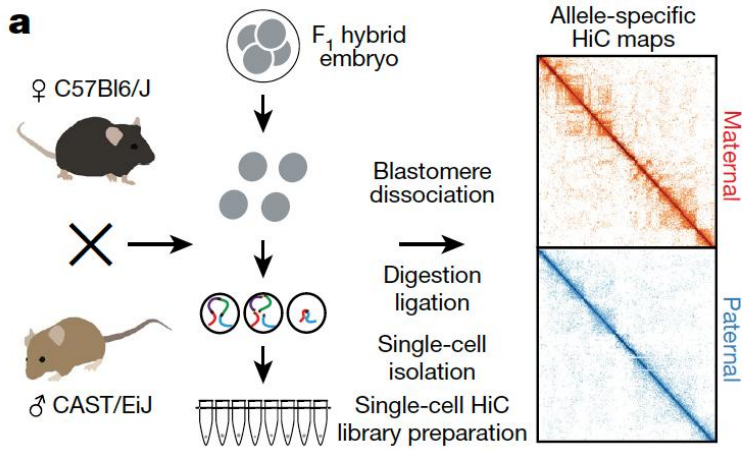



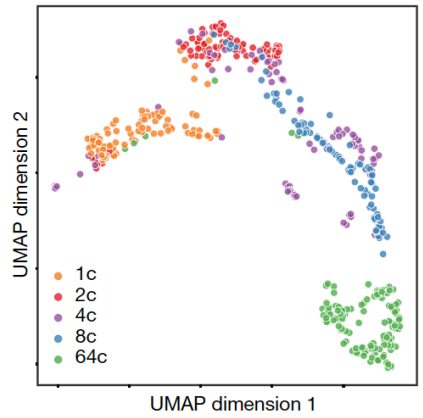

**Single-cell Hi-C** (scHi-C) is an extension of traditional Hi-C that enables the study of **chromatin interactions at the single-cell level**, offering insights into **chromatin conformation heterogeneity** within populations of cells.

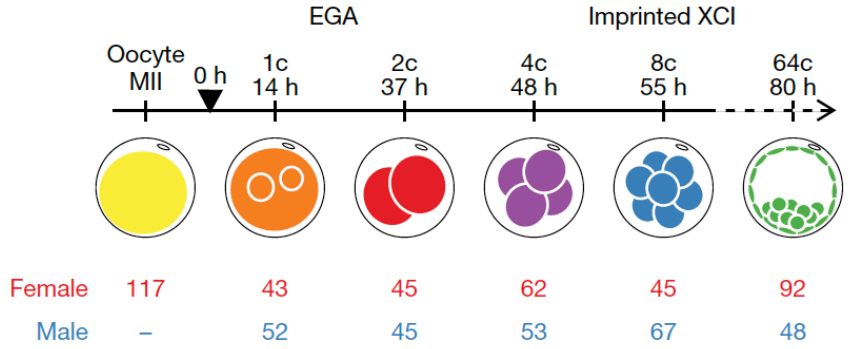
*Nagano et al.*, *Nature*, 2013

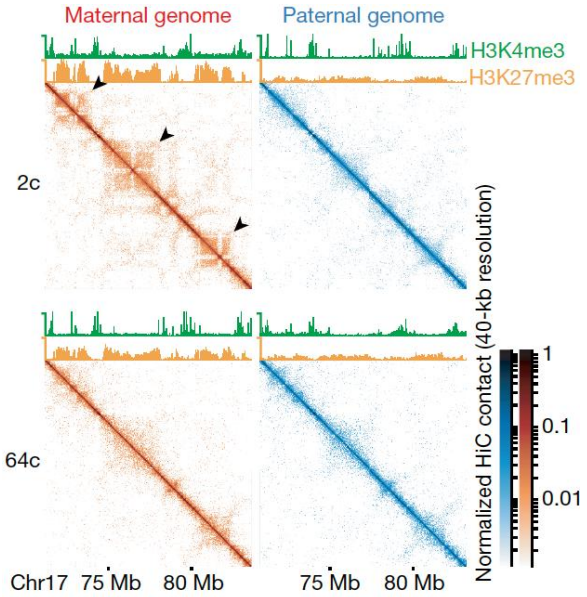
# 04 Chromatin Structure

## 4.3 single cell Hi-C



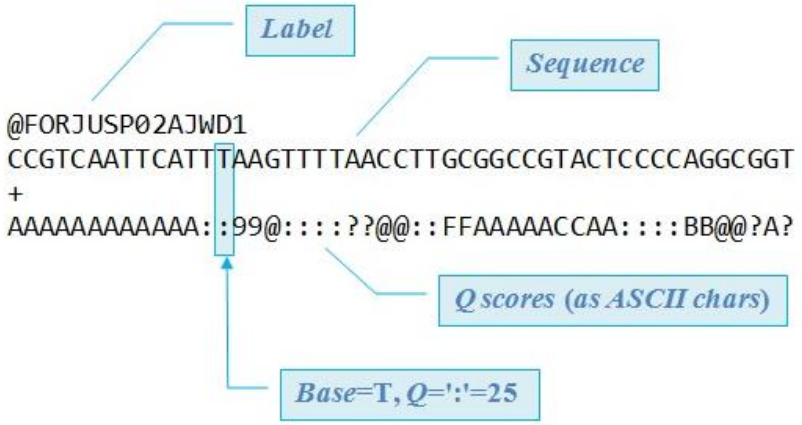




Collombet et al., *Nature*, 2020

# 04 Chromatin Structure

## 4.4 Data analyses

FASTQ → 质控 (Trimmomatic, Trim Galore)
↓
比对到参考基因组 (bwa / bowtie2)
↓
鉴定Valid Pairs (pairtools)
↓
Bin and count junctions (cooler)
↓
鉴定compartment / TAD / loop / stripe (cooltools / hicFindTADs / FitHiC2 / StripeCaller)




**质量控制**

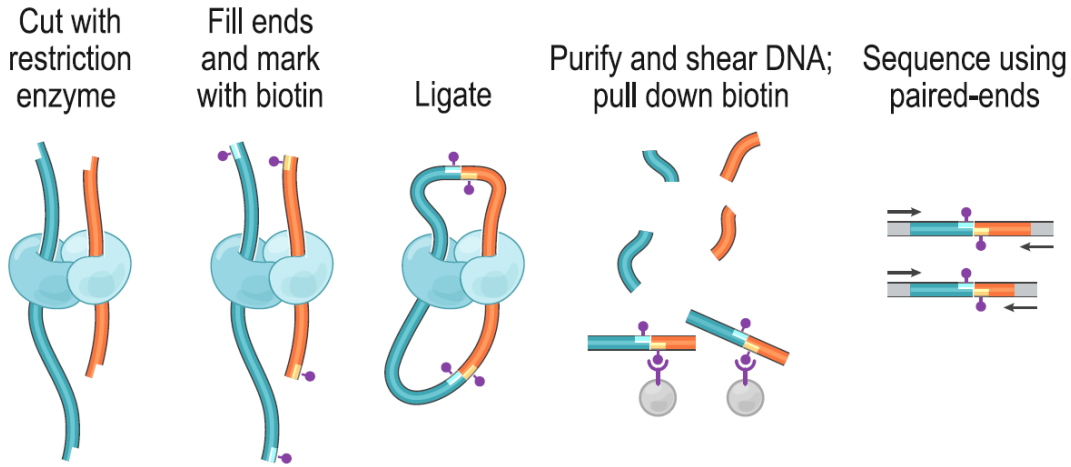
- **目的：** 去除低质量碱基与接头序列，防止比对错误
- **常用工具：**
  - Trimmomatic
  - Trim Galore

# 04 Chromatin Structure

## 4.4 Data analyses

FASTQ → 质控 (Trimmomatic, Trim Galore)
↓
比对到参考基因组 (bwa / bowtie2)
↓
鉴定Valid Pairs (pairtools)
↓
Bin and count junctions (cooler)
↓
鉴定compartment / TAD / loop / stripe (cooltools / hicFindTADs / FitHiC2 / StripeCaller)

鉴定Valid Pairs



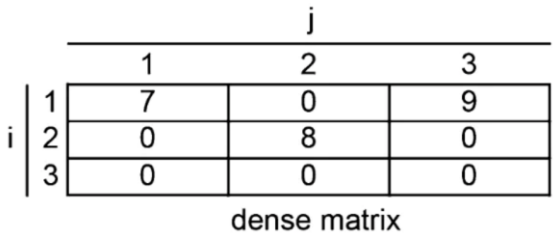

•**核心**：HiC产生的reads主要来源于基因组上的两个非连续位置的连接，因此每一条reads代表了空间上的一个DNA互作

## 4.4 Data analyses

FASTQ → 质控 (Trimmomatic, Trim Galore)
↓
比对到参考基因组 (bwa / bowtie2)
↓
鉴定Valid Pairs (pairtools)
↓
Bin and count junctions (cooler)
↓
鉴定compartment / TAD / loop / stripe (cooltools / hicFindTADs / FitHiC2 / StripeCaller)

Bin and count junctions

| Bin1坐标 | Bin2坐标 | Reads count | 互作频率 |
|---|---|---|---|
| 0 | 0 | 2435 | 0.270307 |
| 0 | 1 | 937 | 0.134423 |
| 0 | 2 | 271 | 0.0248747 |
| 0 | 3 | 129 | 0.011422 |
| 0 | 4 | 53 | 0.00667344 |
| 0 | 5 | 51 | |
| 0 | 6 | 66 | 0.00487654 |
| 0 | 7 | 51 | 0.00392599 |
| 0 | 8 | 49 | 0.00325238 |
| 0 | 9 | 24 | 0.00209715 |
| 0 | 10 | 31 | 0.002481 |
| 0 | 11 | 24 | 0.00157917 |
| 0 | 12 | 24 | 0.00147929 |
| 0 | 13 | 28 | 0.00196714 |
| 0 | 14 | 20 | 0.00158235 |



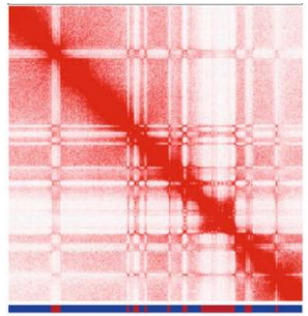

这一步决定最终contact矩阵的分辨率
Bin越小，位于其区间的reads越少
conatct矩阵需要校正（内切酶的偏好性、基因组本身质量、基因组序列特异性）

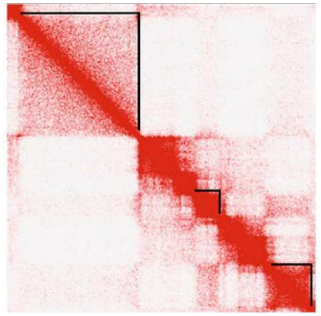
# 04 Chromatin Structure

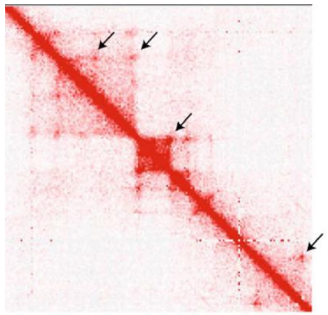
## 4.4 Data analyses

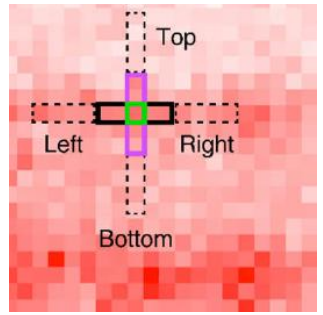
FASTQ → 质控 (Trimmomatic, Trim Galore)
↓
比对到参考基因组 (bwa / bowtie2)
↓
鉴定Valid Pairs (pairtools)
↓
Bin and count junctions (cooler)
↓
鉴定compartment / TAD / loop / stripe (cooltools / hicFindTADs / FitHiC2 / StripeCaller)

compartment/TAD/loop/stripe


compartment


TAD


loop



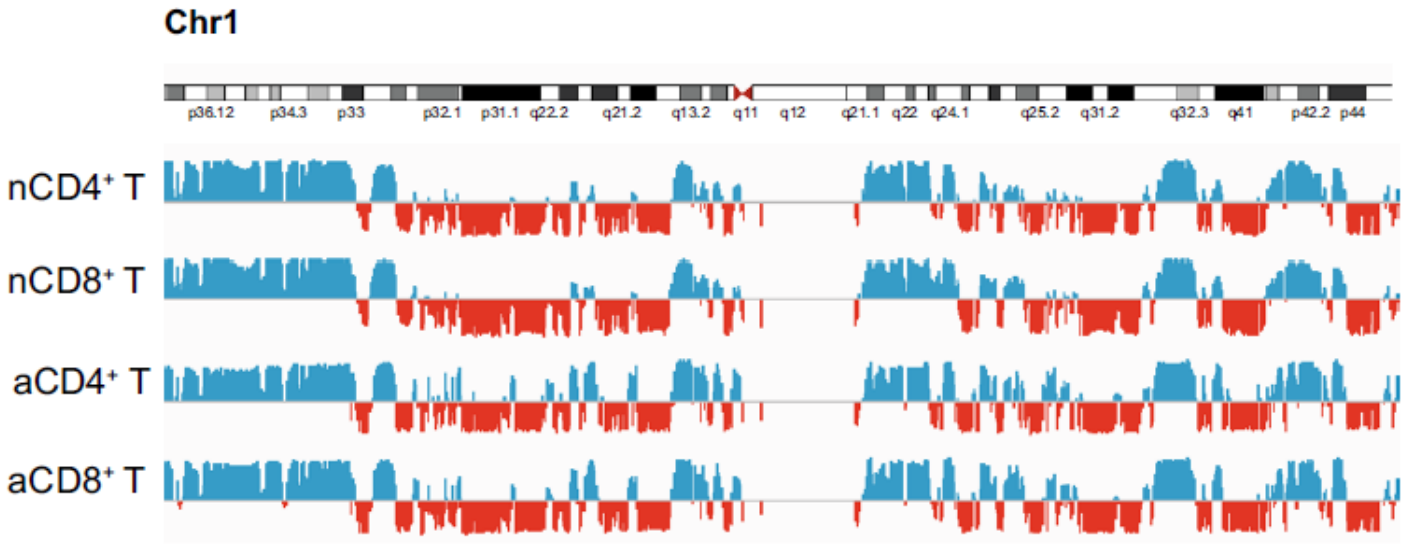
stripe

分辨率确定：
•80%以上的bin里都有1000个contact以上
•10kb分辨率需要的有效测序量：
(3G / 10k) * 0.8 * 1000 = 240M reads (PE150测序量：72G)

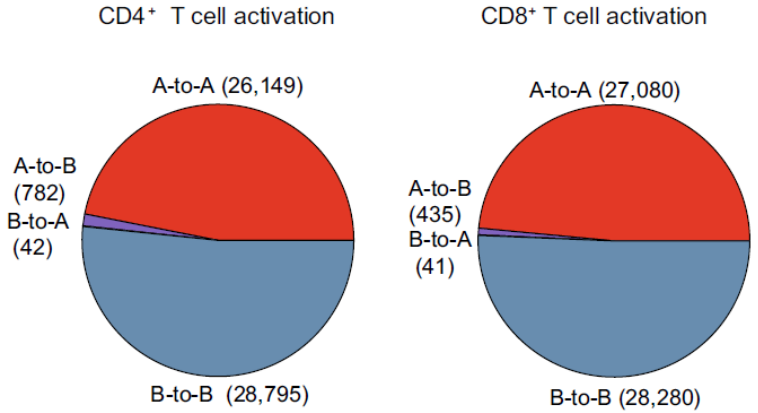
•Compartment→TAD→Loop、Stripe
•分辨率逐渐增大

# 04 Chromatin Structure

## 4.4 Data analyses




A/B compartments of resting and activated T cells along chromosome 1

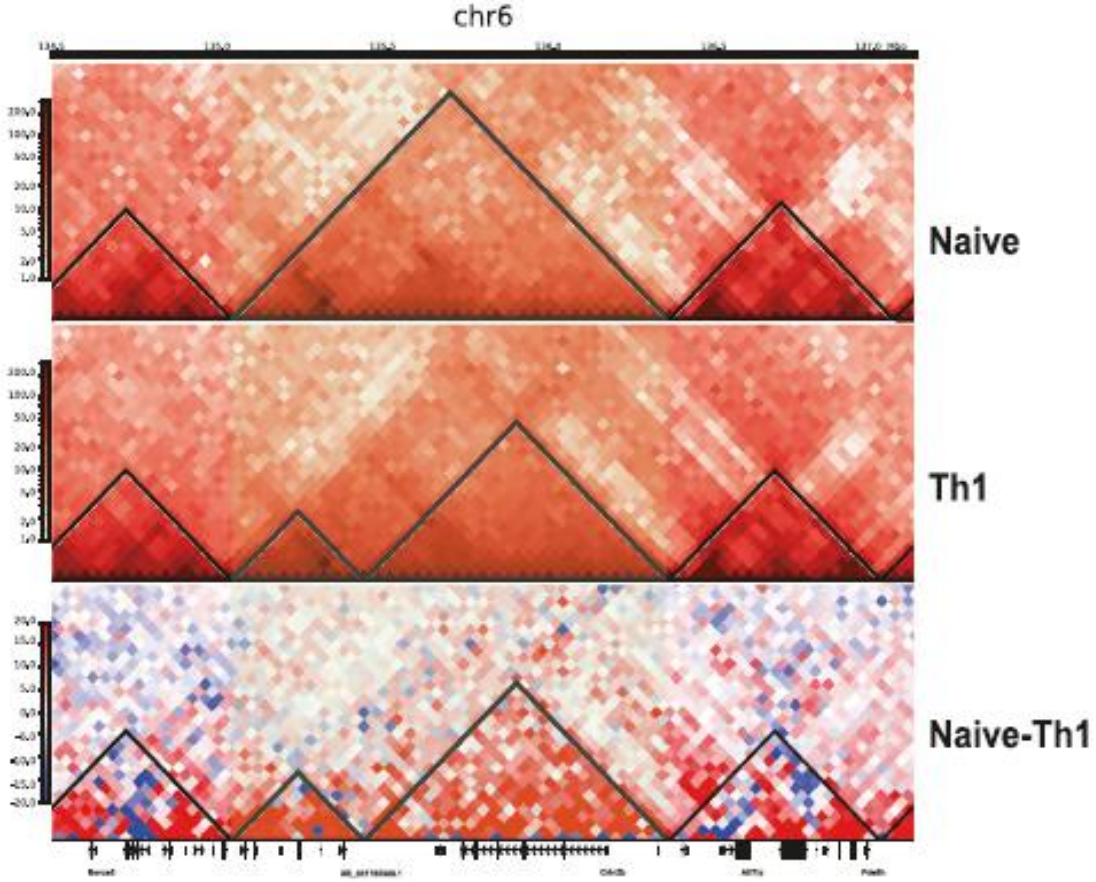



compartment changes: stable (A to A, B to B) or flipping (A to B and B to A)

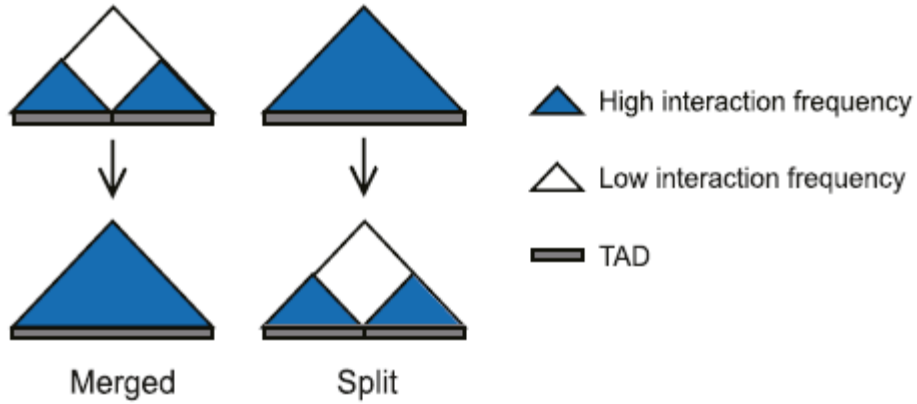
Bediaga, N.G. *et al*., *Sci Rep,* 2021

# 04 Chromatin Structure

## 4.4 Data analyses

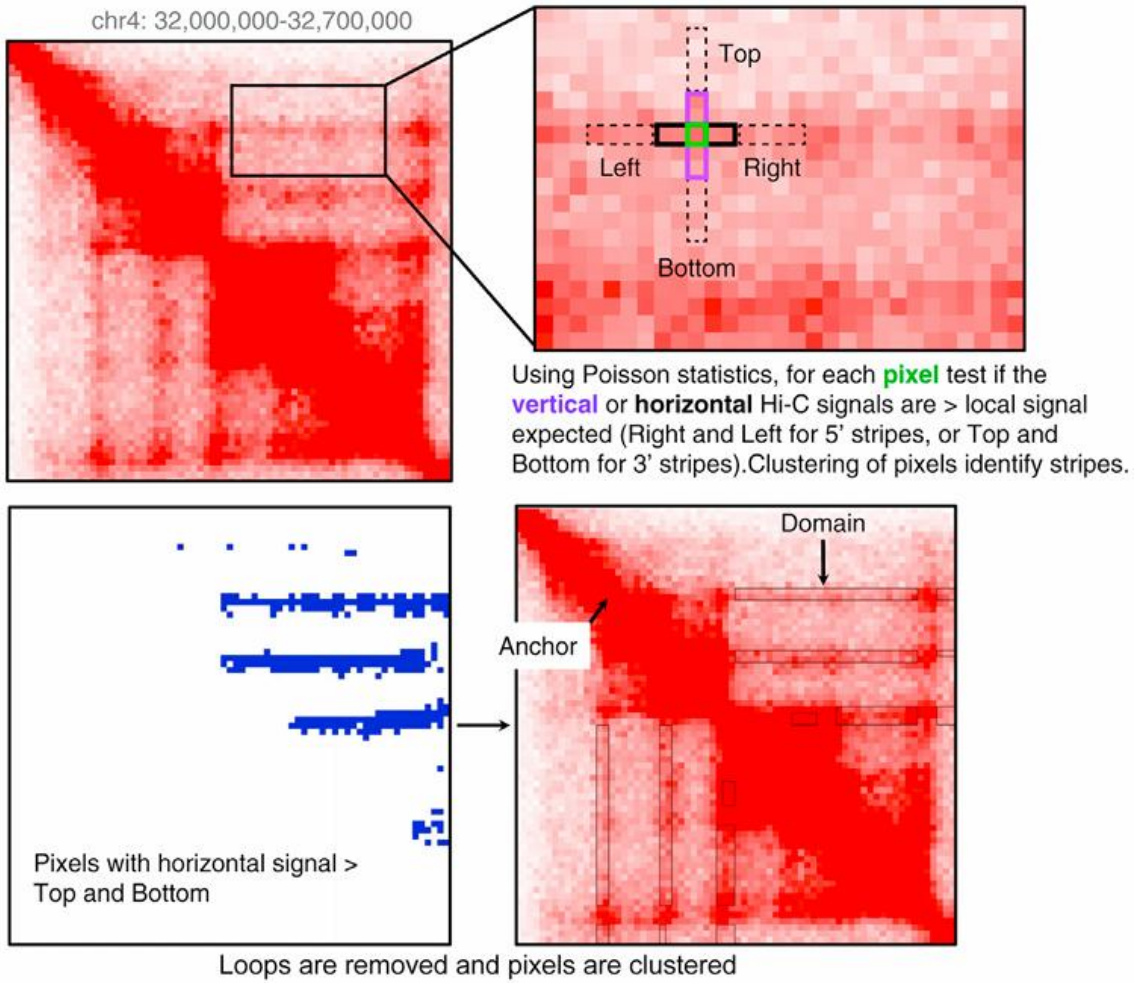


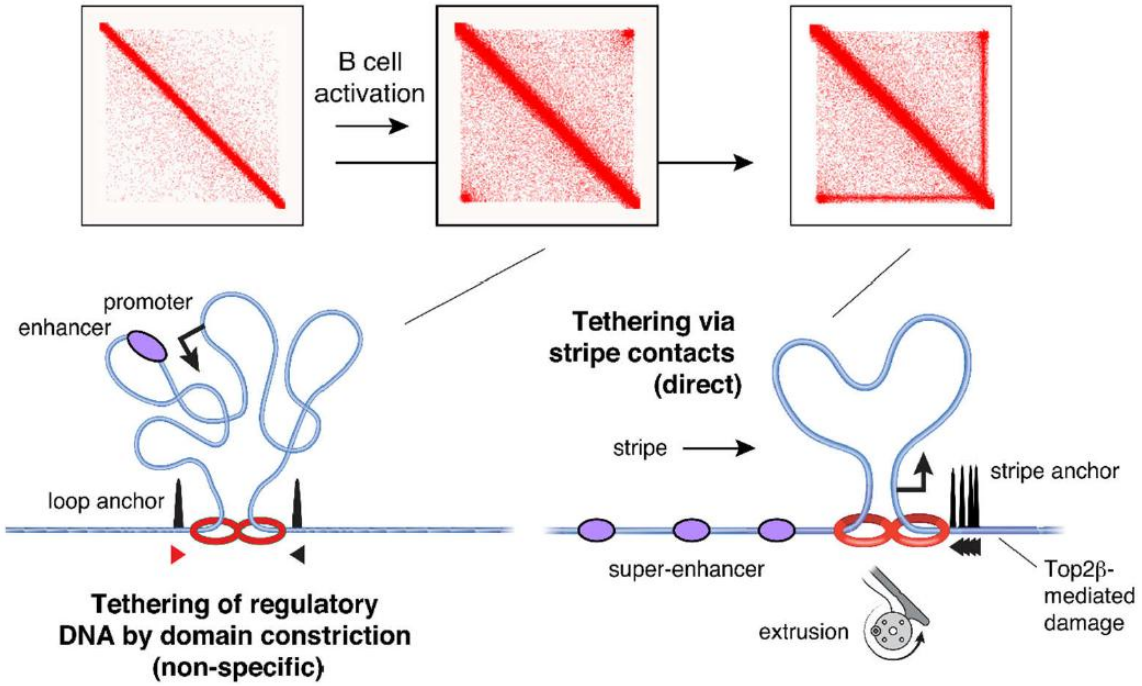

T cell differentiation results in TAD rearrangement




Zhang, G.. *et al*., *Commun Biol,* 2023

## 4.4 Data analyses



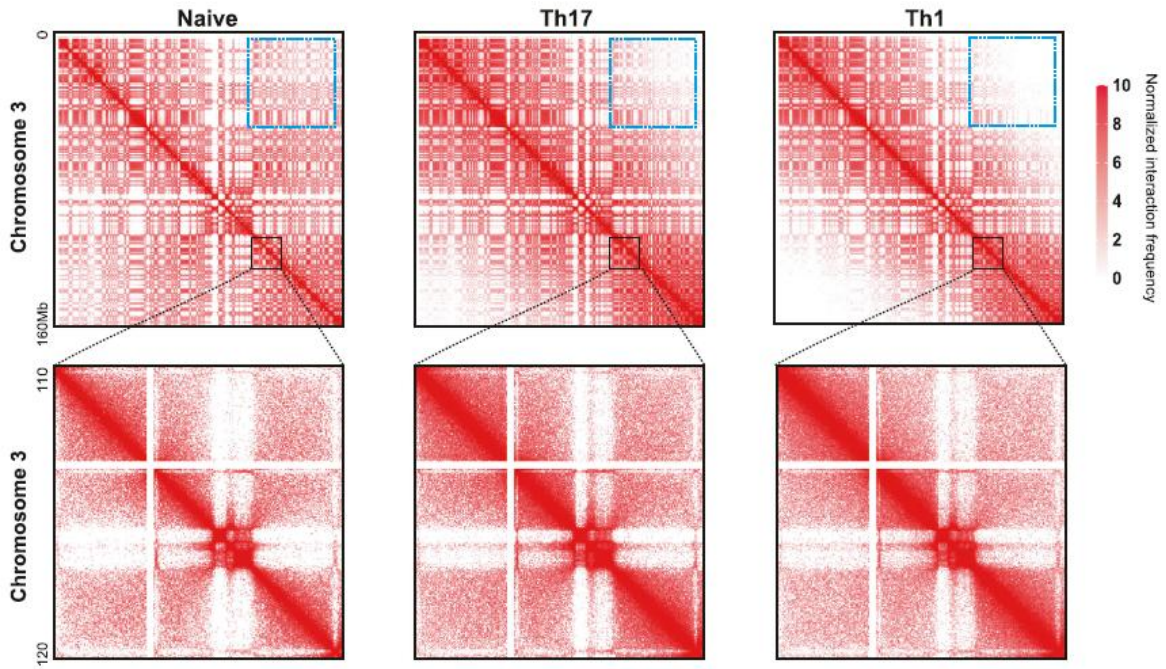



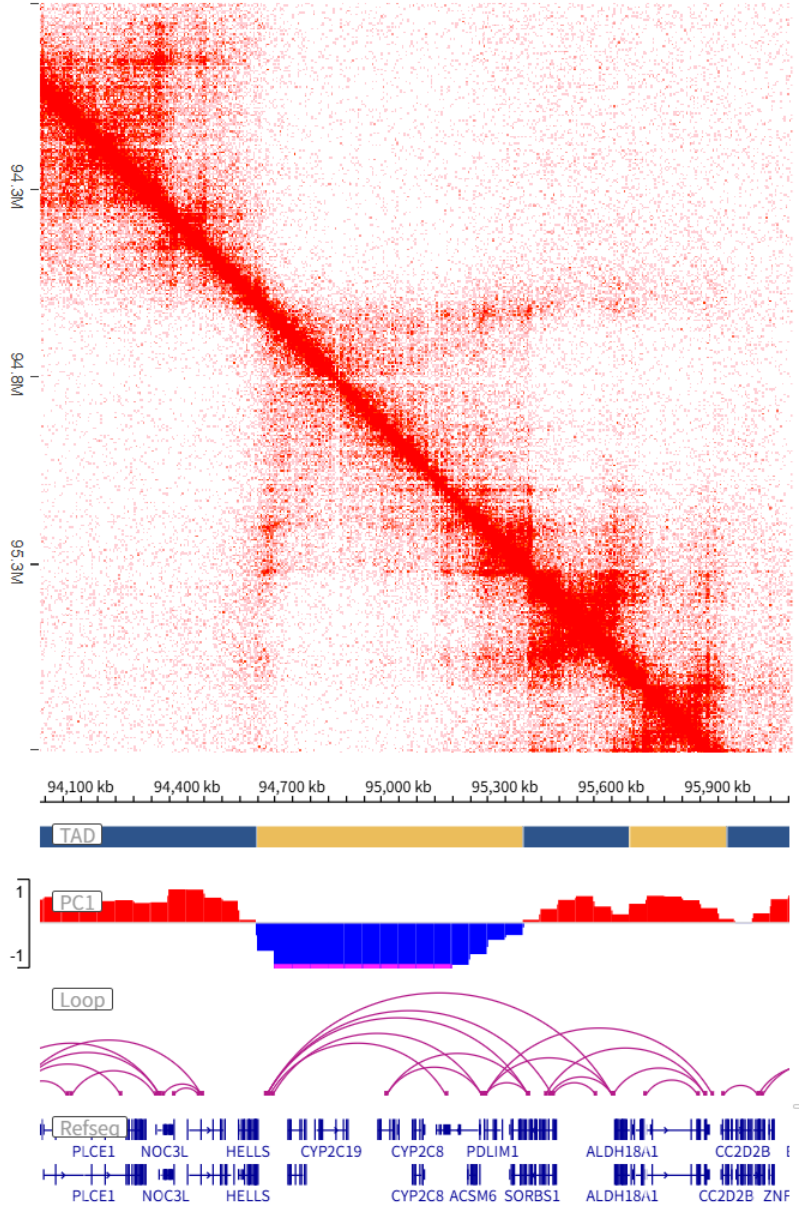

Vian, Laura. *et al*., *Cell,* 2018

# 04 Chromatin Structure

## 4.4 Data analyses

### HiC可视化 Juicerbox / HiCExplorer



# Epigenetic Sequencing

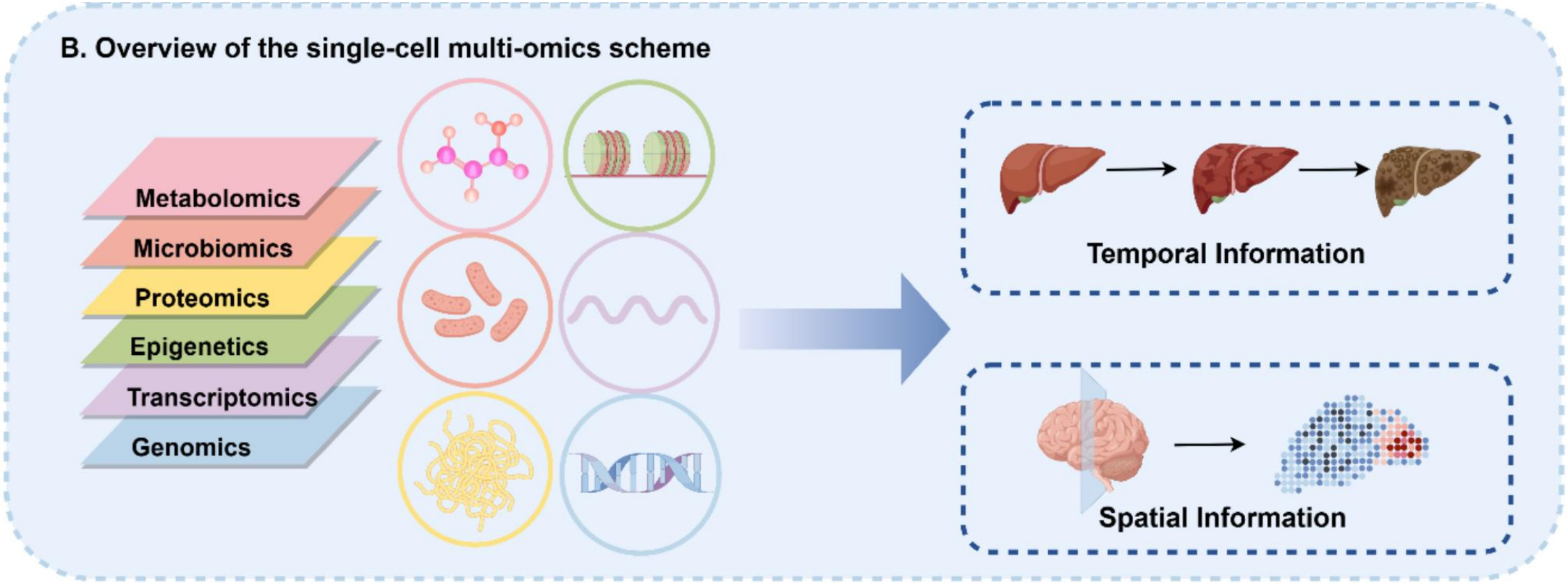
## Epigenetic Sequencing

1. DNA-methylation  (WGBS, RRBS, MeDIP-seq, single-cell WGBS)
2. Histone-modification

   (ChIP-seq, CUT&RUN, CUT&Tag, single-cell CUT&Tag)
3. Chromatin accessibility (ATAC-Seq, DNase-seq, NOMe-seq, single-cell ATAC-seq)
4. Chromatin Structure (3C-seq, Hi-C-seq, single-cell Hi-C)
5. Multi-omics analyses and applications

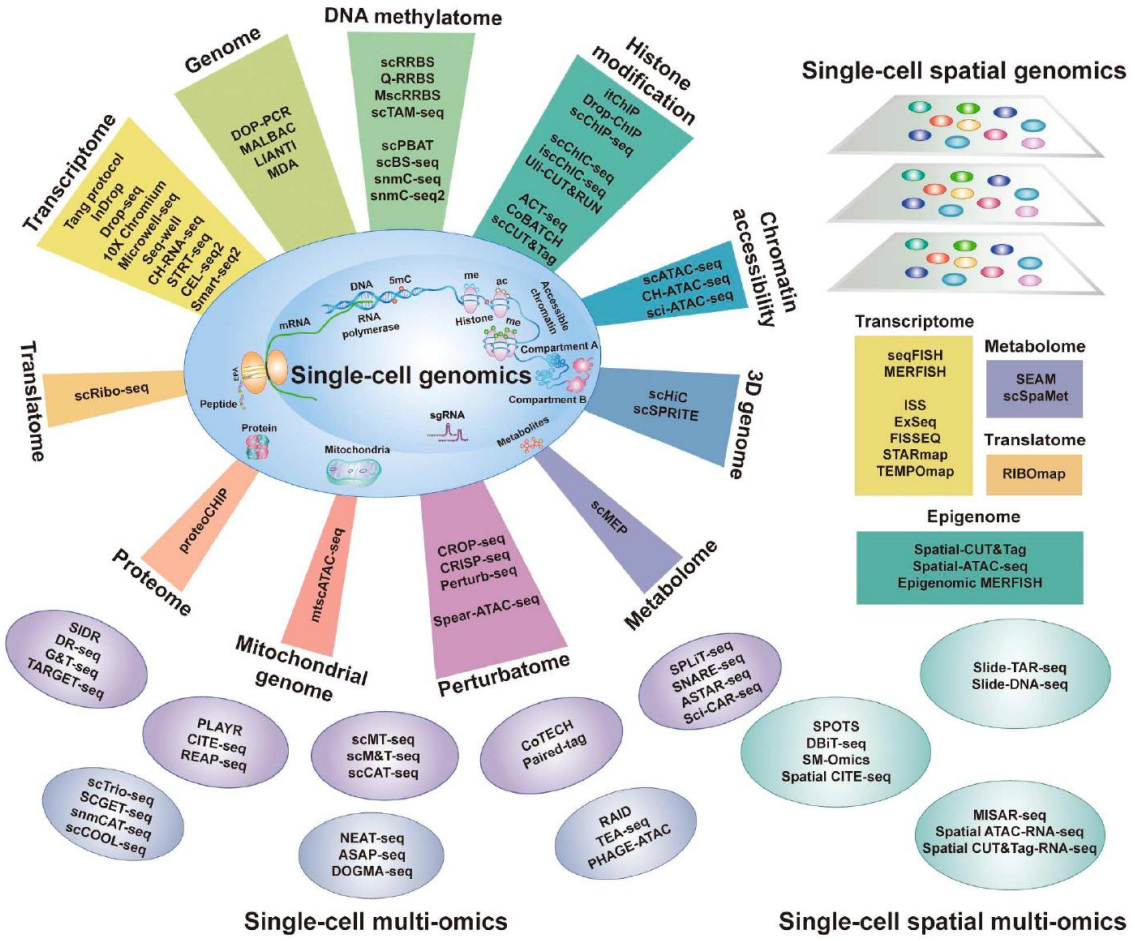
# Multi-omics

## 5.1 single-cell Multi-omics sequencing



# Multi-omics

## 5.1 single-cell Multi-omics sequencing



# 05 Multi-omics

## 5.1 single-cell Multi-omics sequencing

### Dual-omics :

1. RNA + DNA : **G&T-seq**, Target-seq, DR-seq
2. RNA + ATAC : SHARE-seq, ISSAAC-seq, Paired-seq, SNARE-seq, ASTAR-seq
3. RNA + Histone : **Paired-tag**, CoTECH,
4. 5mC + ATAC : scNOME-seq
5. Chromatin structure + 5mC : scMethyl-HiC, sn-m3C-seq
6. RNA + Protein : **CITE-seq**
7. RNA + Perturbatome : Perturb-seq, **CROP-seq**
8. RNA + spatial: seqFISH, MERFISH, DBiT-seq, Slide-sesq, Stereo-seq

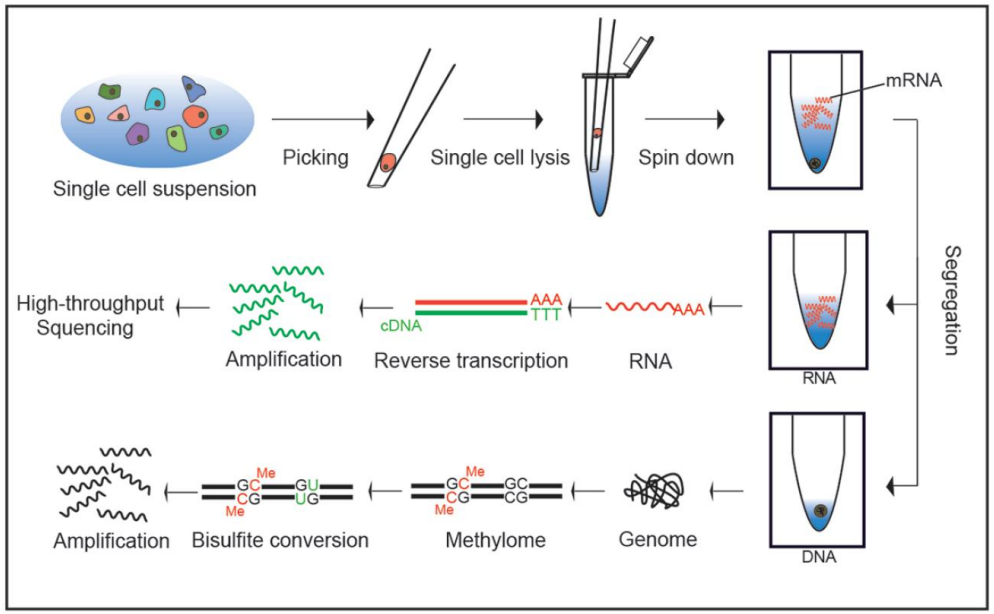
### Triple-omics :

1. RNA + CNV+ 5mC : scTrio-seq; **scTrio-seq2**
2. CNV+ ATAC + 5mC : **Cool-seq**
3. ATAC + 5mC + RNA : **scNOMeRe-seq**

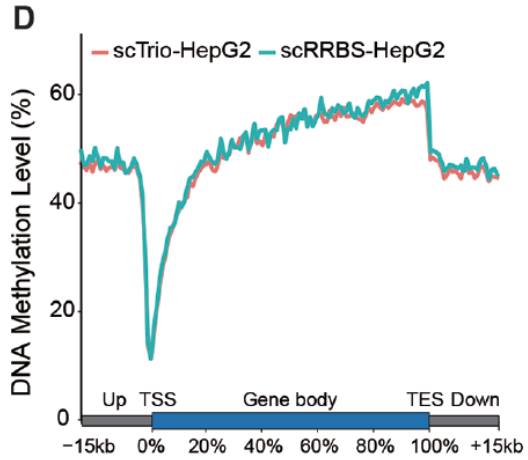
5mC : DNA methylation
CNV : DNA copy number variation

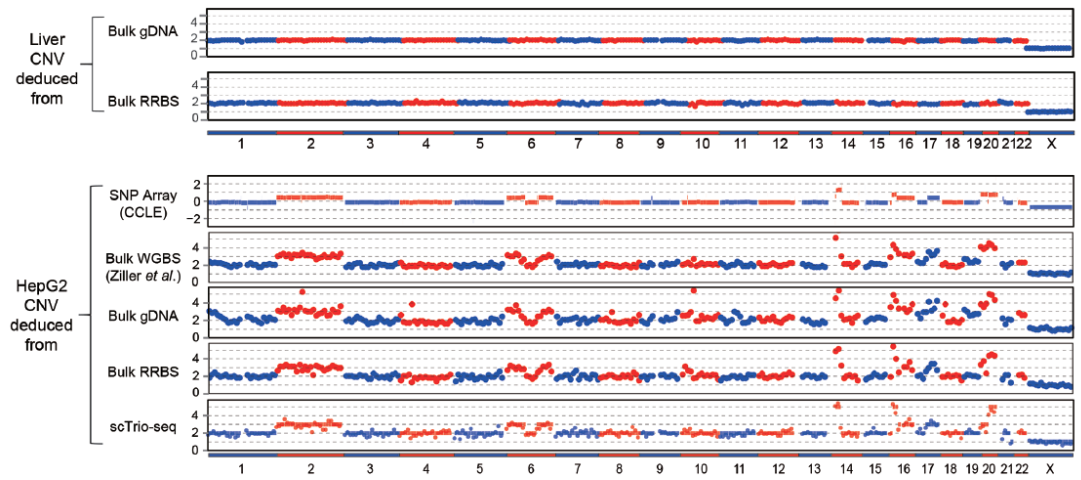
## 5.2 Multi-omics sequencing application

scTrio-seq: RNA + CNV+ 5mC




Combine scRNA-seq and PBAT




Hou *et al*., *Cell Research*, 2016

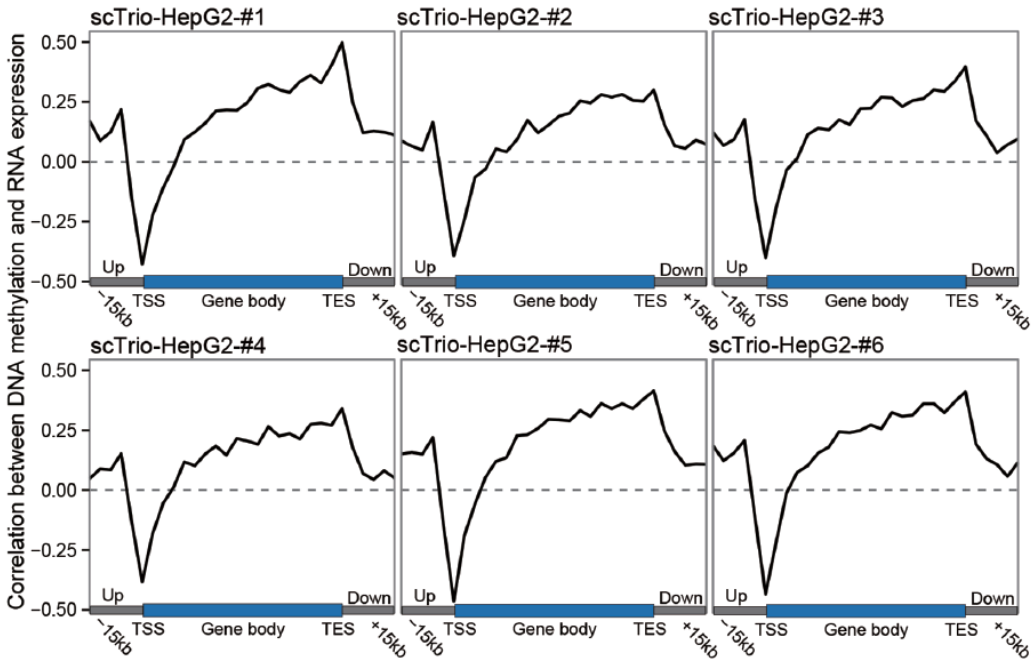
## 5.2 Multi-omics sequencing application

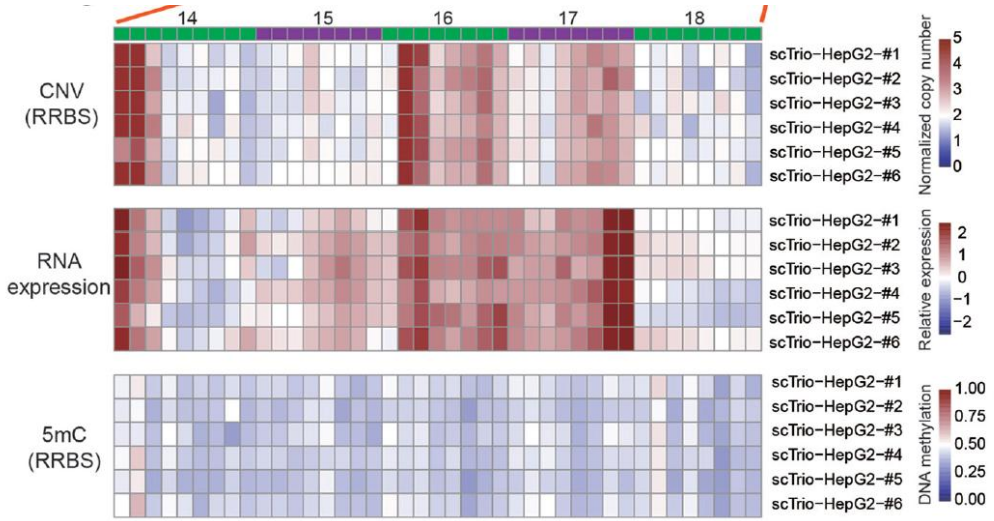
**Table 1** Number of the detected CpG sites, genes, and MspI-digested fragments in single HepG2 cells

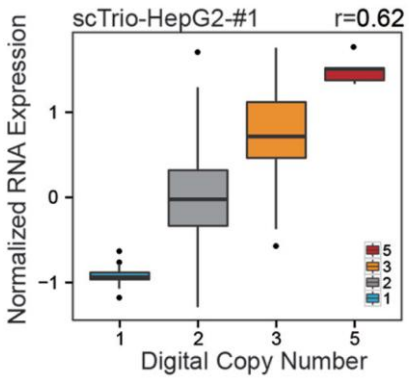
| Sample | Unique CpGs (1×) | Unique CpGs (3×) | Genes (FPKM≥0.1) | Genes (FPKM≥1) | MspI-digested fragments |
|---|---|---|---|---|---|
| scTrio-HepG2-#1 | 1 834 536 | 1 276 842 | 6 083 | 4 373 | 150 288 |
| scTrio-HepG2-#2 | 1 239 255 | 819 238 | 6 440 | 4 746 | 103 892 |
| scTrio-HepG2-#3 | 1 217 007 | 709 874 | 6 271 | 5 122 | 104 884 |
| scTrio-HepG2-#4 | 1 251 747 | 725 124 | 5 808 | 4 329 | 105 635 |
| scTrio-HepG2-#5 | 1 762 799 | 1 201 953 | 6 437 | 4 904 | 145 361 |
| scTrio-HepG2-#6 | 1 820 527 | 1 308 313 | 6 036 | 4 702 | 146 772 |
| Mean of scTrio-HepG2 | 1 520 979 | 1 006 891 | 6 179 | 4 696 | 126 139 |
| scRRBS-HepG2-#1 | 1 336 924 | 780 377 | / | / | 115 853 |
| scRRBS-HepG2-#2 | 1 199 569 | 701 340 | / | / | 105 278 |
| scRNA-HepG2-#1 | / | / | 6 099 | 4 335 | / |
| scRNA-HepG2-#2 | / | / | 6 542 | 4 987 | / |

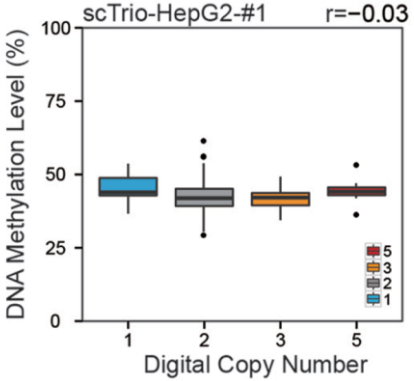
Hou *et al*., *Cell Research*, 2016

## 5.2 Multi-omics sequencing application

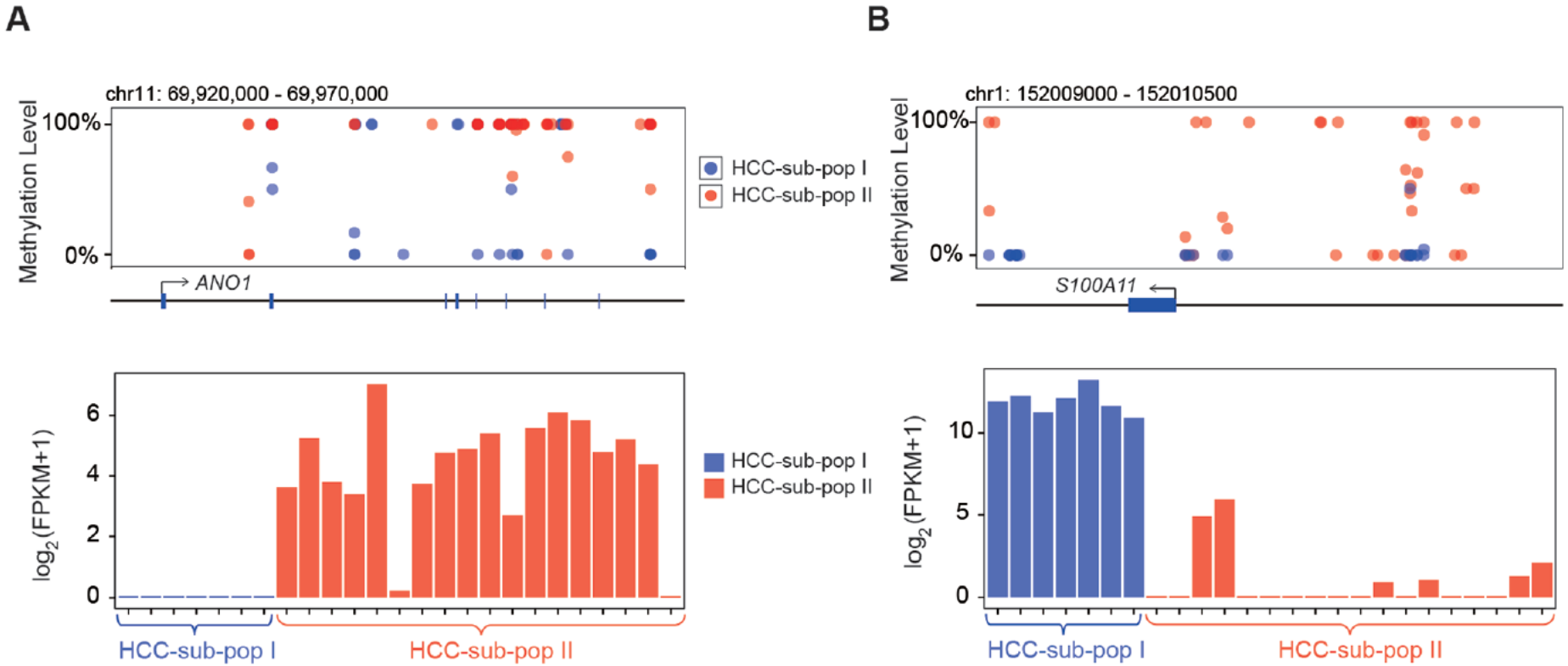


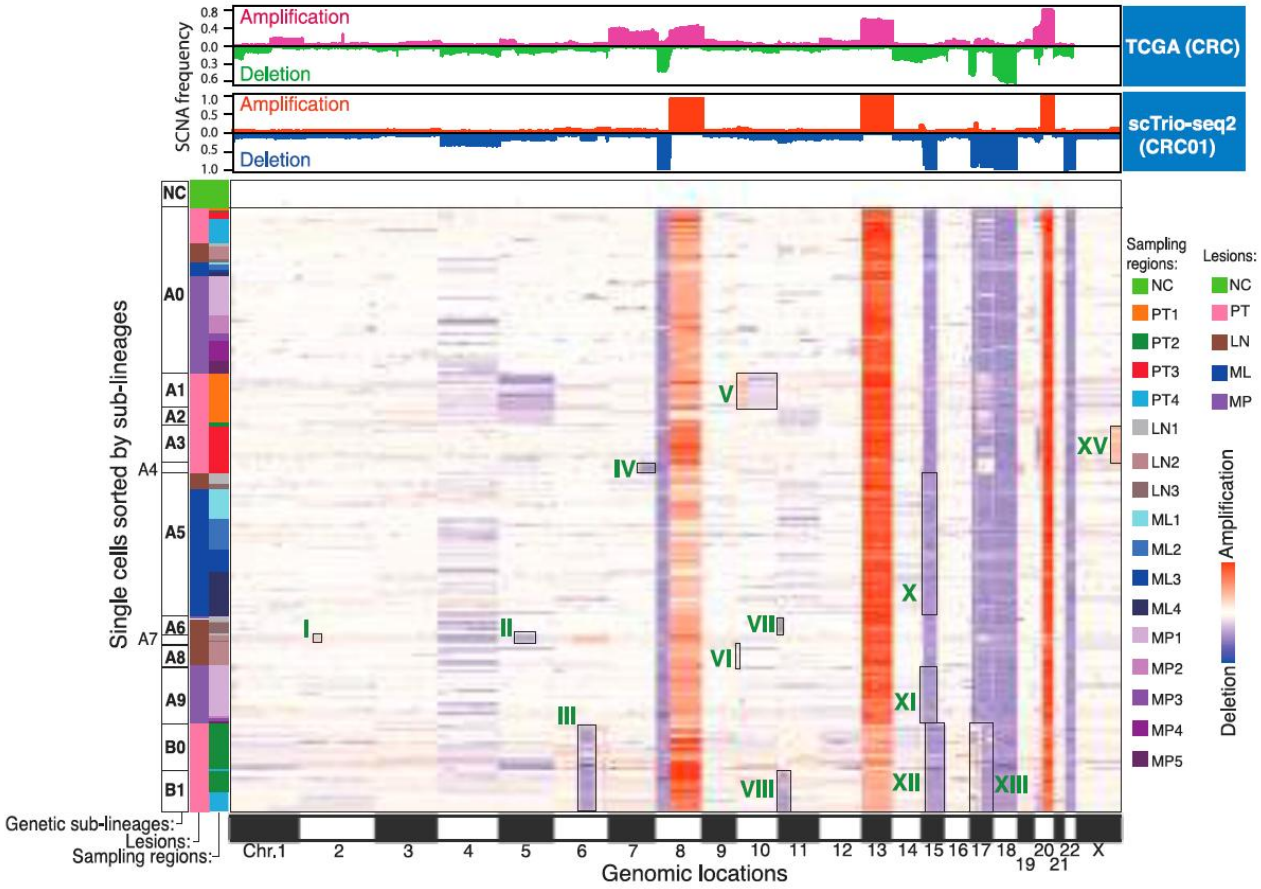



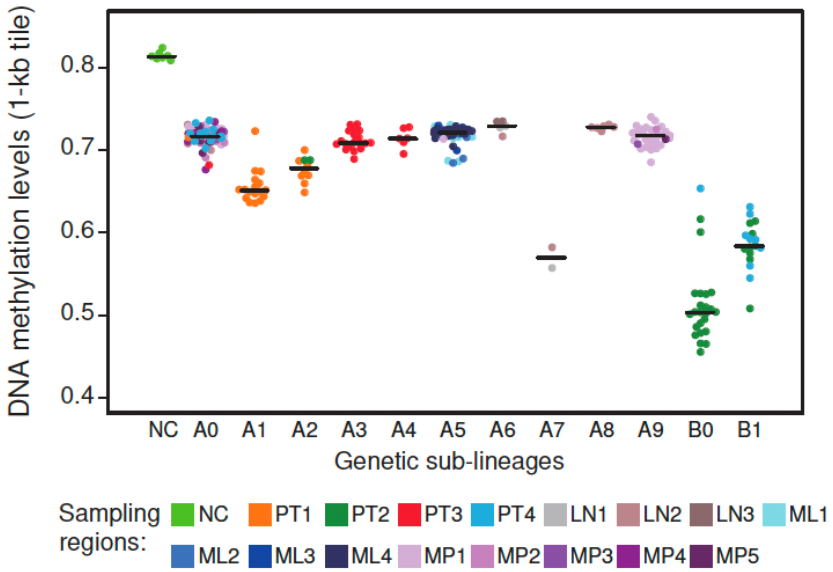



## 5.2 Multi-omics sequencing application

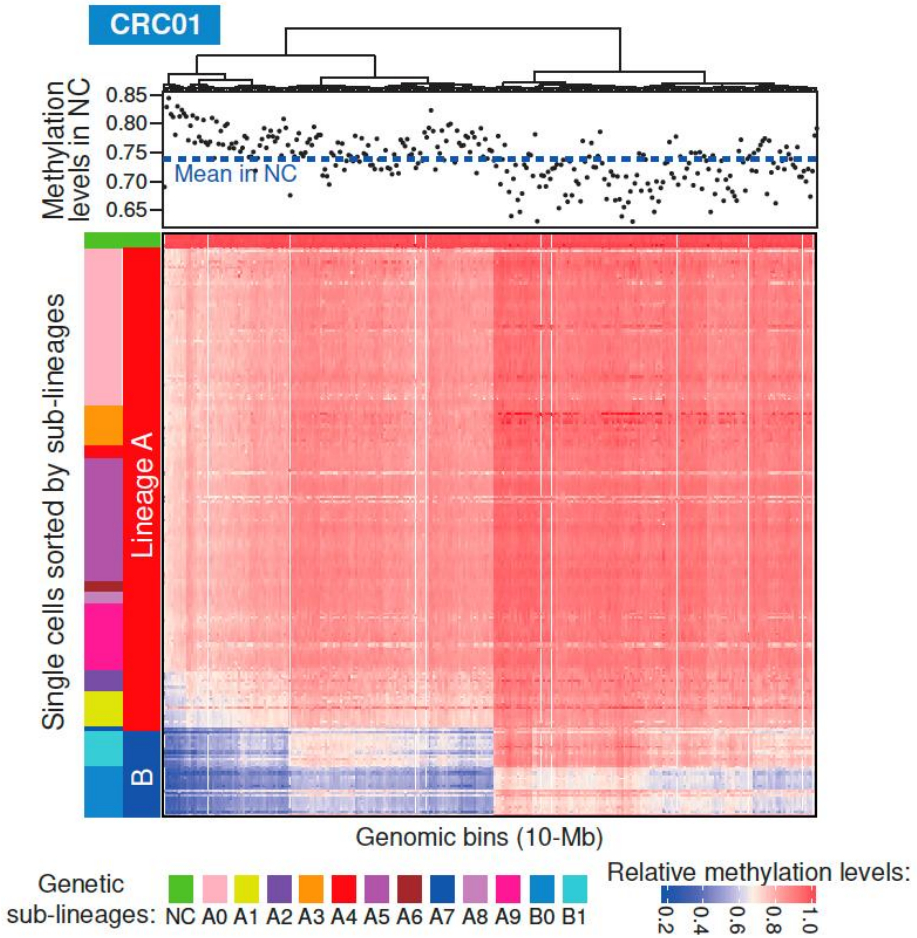


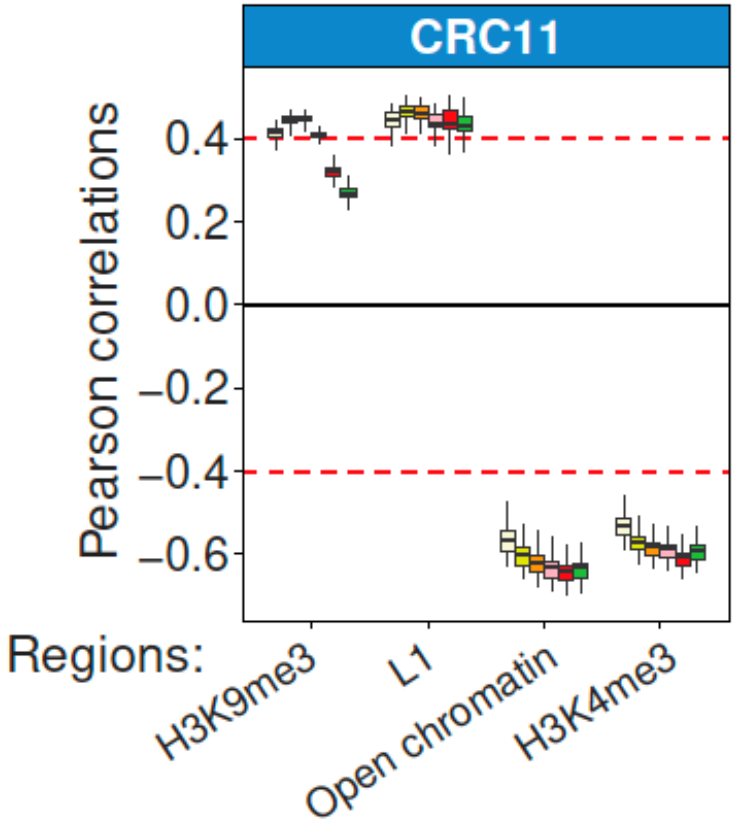
## 5.2 Multi-omics sequencing application



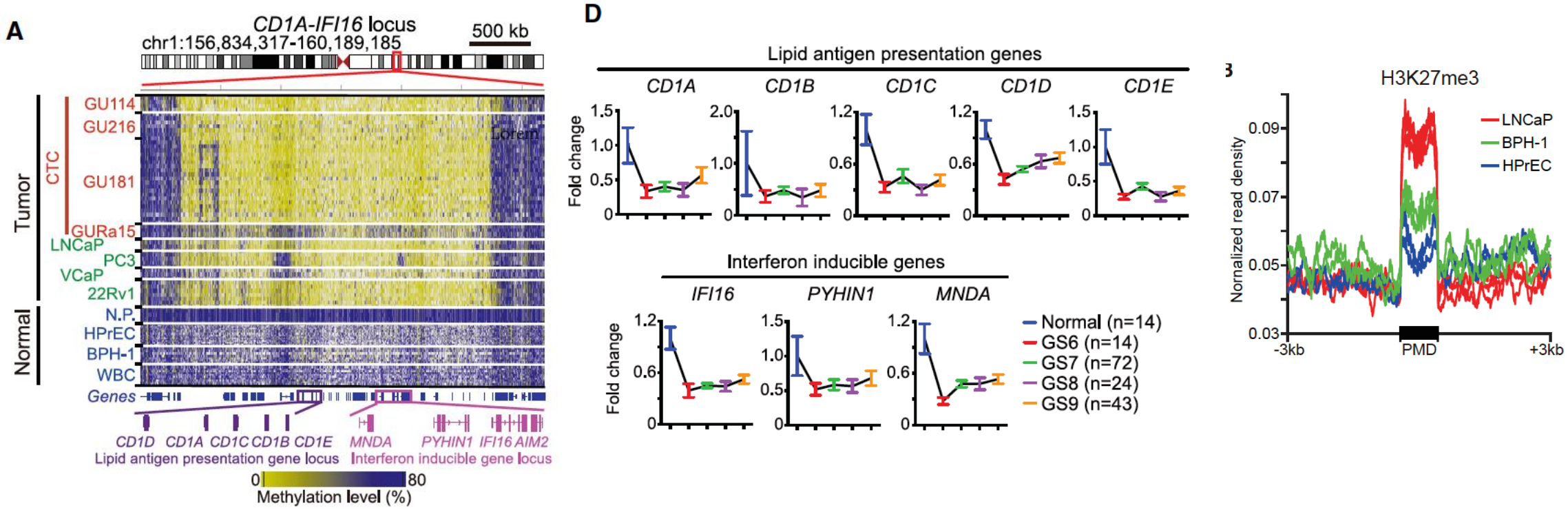



## 5.2 Multi-omics sequencing application



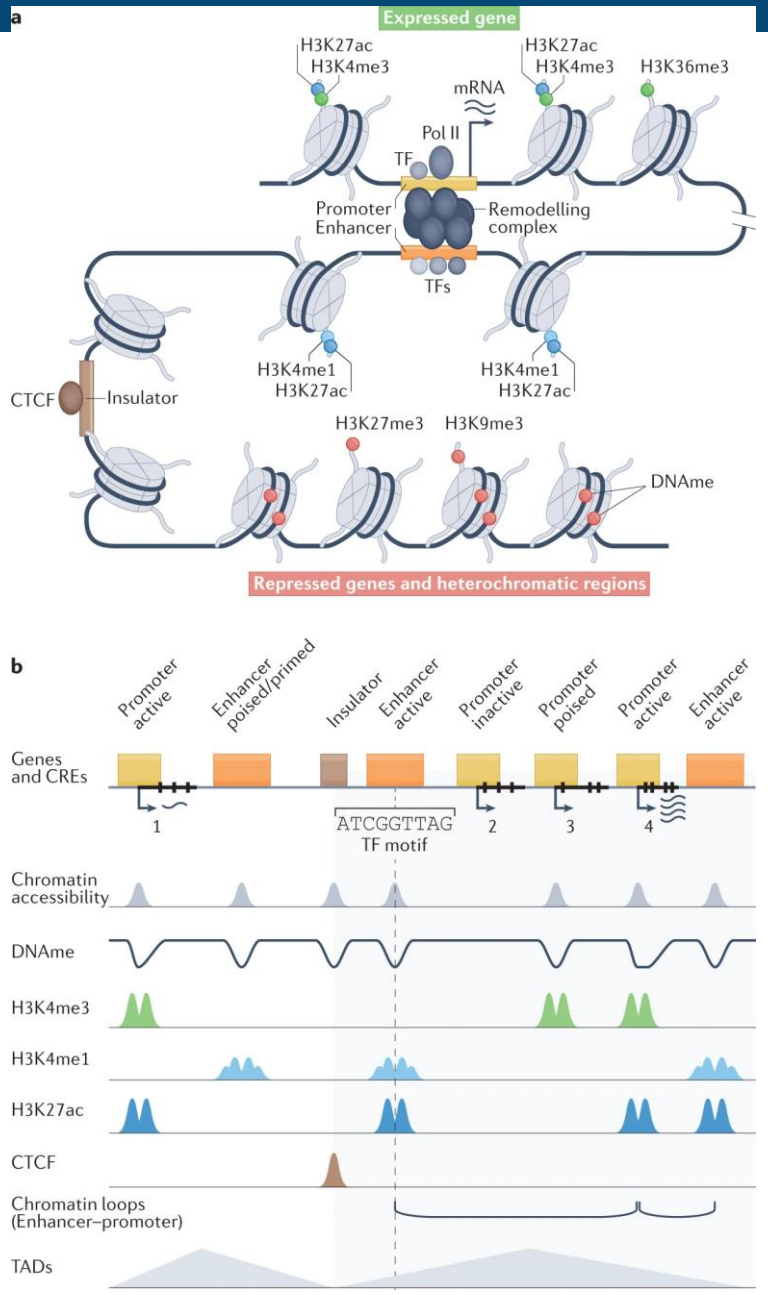
## 5.2 Multi-omics sequencing application



# Take Home Message

## Epigenetic Sequencing

1. DNA-methylation (WGBS, RRBS, MeDIP-seq, single-cell WGBS)
2. Histone-modification
   (ChIP-seq, CUT&RUN, CUT&Tag, single-cell CUT&Tag)
3. Chromatin accessibility (ATAC-Seq, DNase-seq, NOMe-seq, single-cell ATAC-seq)
4. Chromatin Structure (3C-seq, Hi-C-seq, single-cell Hi-C)
5. Multi-omics analyses and applications



# Thank you

**Yu Hou** Zhejiang University 2025