



武汉大学

Wuhan University

# 5、表观基因组研究方法

一、全基因组甲基化测序

二、甲基化DNA免疫共沉淀测序

三、染色质免疫共沉淀测序





# 一、全基因组甲基化测序

## 研究目的与内容：

- 1、构建全基因组甲基化图谱（DNA methylation profiling），全面而精确了解全基因组DNA甲基化状态及其高低甲基化位点在基因组不同区域的分布情况。
- 2、多样品间差异性甲基化区域分析

## 全基因组Bisulfite测序（Whole Genome Bisulfite Sequencing, WGBS；

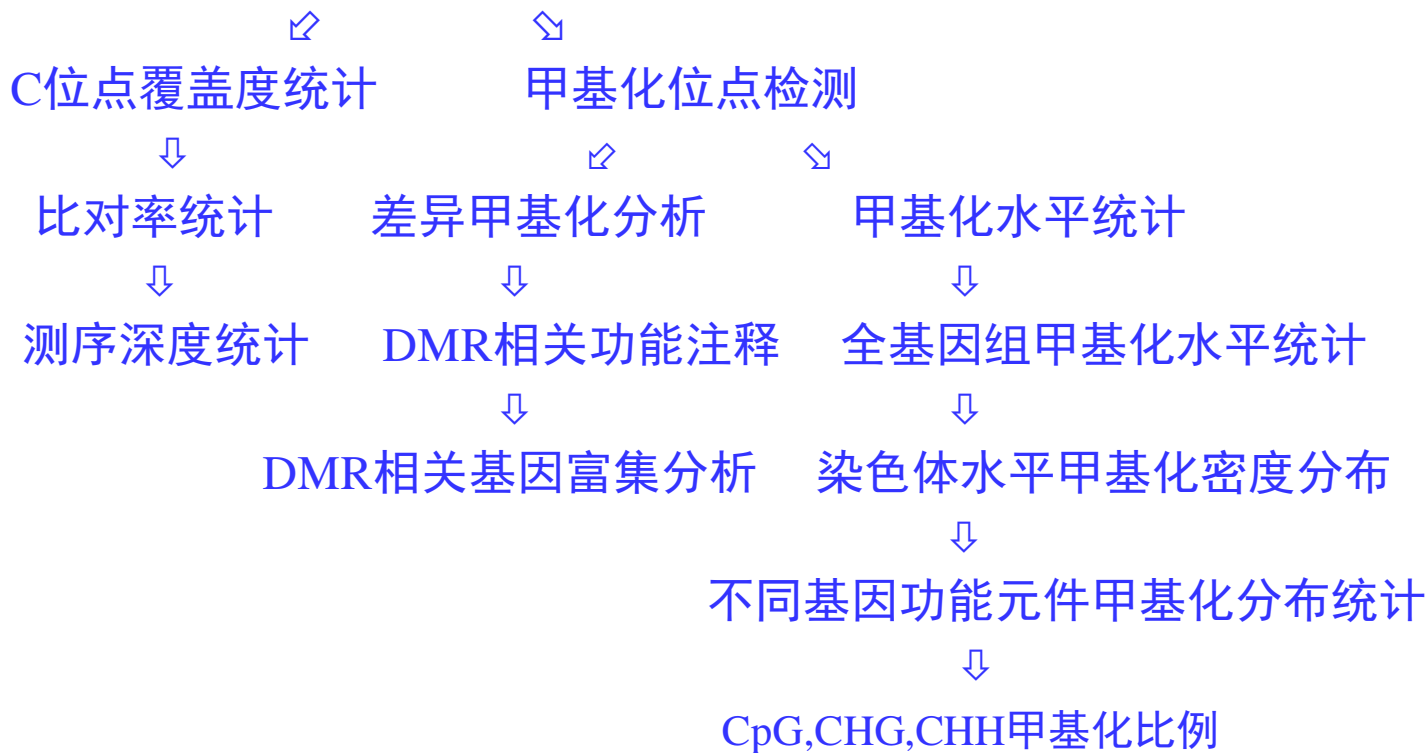
Bisulfite-Seq）： Bisulfite处理能够将基因组中未甲基化的C碱基与甲基化的C碱基区分开来，是表观遗传学研究的经典方法。将重亚硫酸盐处理方法和高通量测序（例如 Illumina HiSeq）技术结合，能够绘制出单碱基分辨率的DNA甲基化图谱，可用于研究物种特定DNA区域甲基化与特定表型的关联；也是研究表观遗传调控的机制的基础





# 全基因组甲基化测序实验技术路线

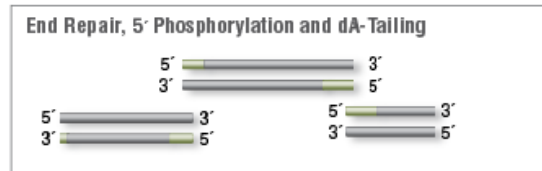
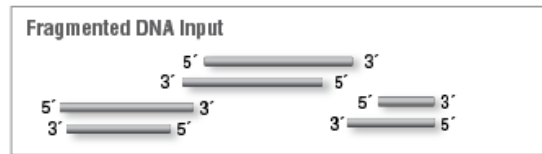
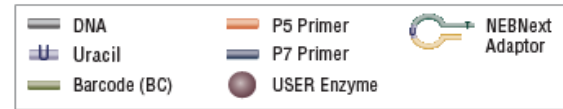
基因组DNA ⇒ 全基因组DNA甲基化文库 (DNA片段化⇒末端修复、加A⇒连接接头) ⇒ Bisulfite处理⇒PCR扩增 ⇒ Illumina HiSeq测序⇒数据整体质量评估 ⇒ 参考序列比对





# 武汉大学

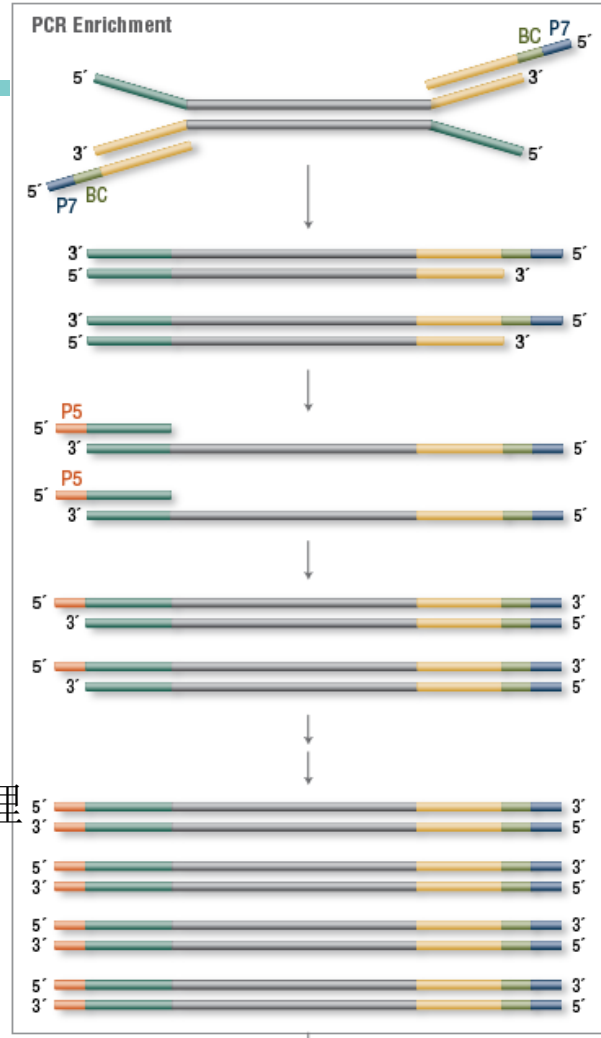
## Wuhan University



甲基化修饰接头



亚硫酸盐处理





## 测序样品要求：

**样品纯度：**OD 260/280值 在1.8 ~ 2.0之间；RNA应去除干净

**样品浓度：**最低浓度不低于50ng/ul

**样品总量：**每个样品总量不少于15ng

**样品溶剂：**应溶解在H<sub>2</sub>O或TE (pH8.0) 中

**样品运输：**DNA低温运输 (-20°C) ；

用parafilm将管口密封好，以防出现污染

**测序数据质量控制：**包括reads数、总base数，Q20比例、

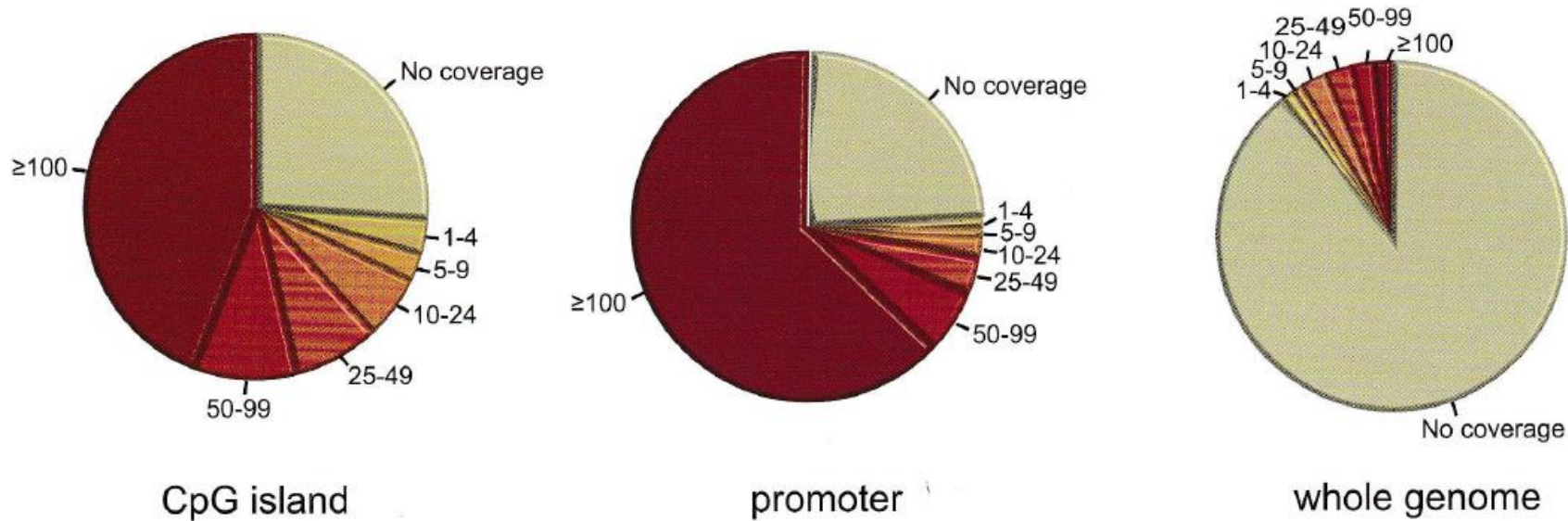
Q20位点分布图、碱基分布图

**测序数据预处理：**包括去除总体质量偏低的reads，将质量大于20碱基所占比例小于50%的reads去除；去除3'端质量Q低于20的碱基；去除reads中所有的接头序列；去除长度小于20的测序片段 (reads)

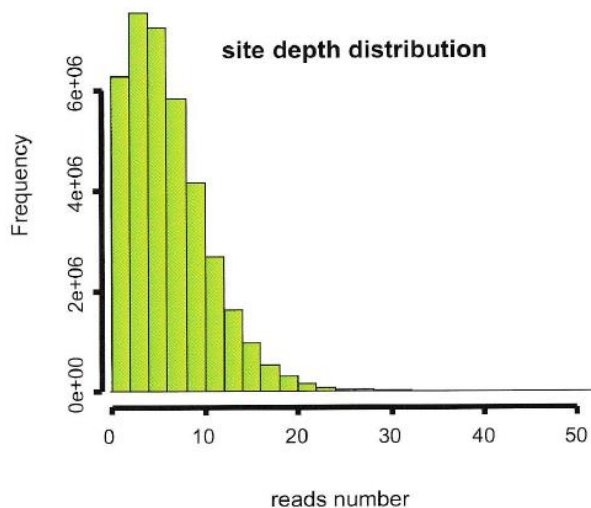




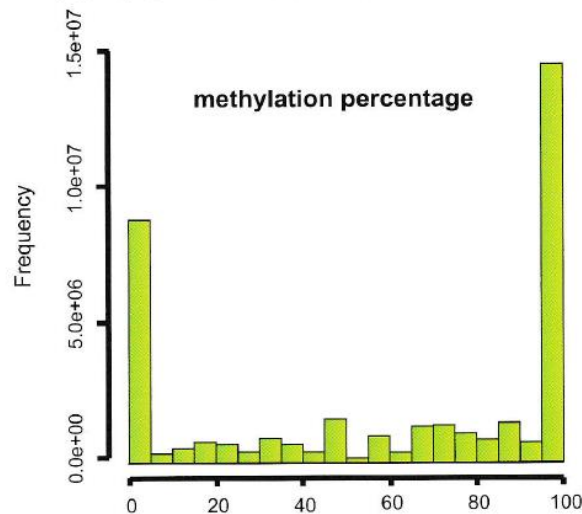
## CpG岛区域、启动子区域及整个基因组的覆盖及不同深度关系



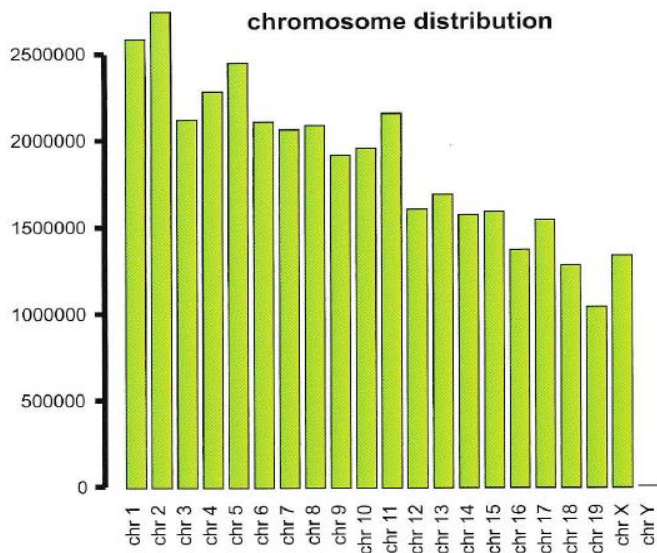
【1】 CpG/CHG/CHH位点测序深度



【2】 CpG/CHG/CHH位点甲基化比例

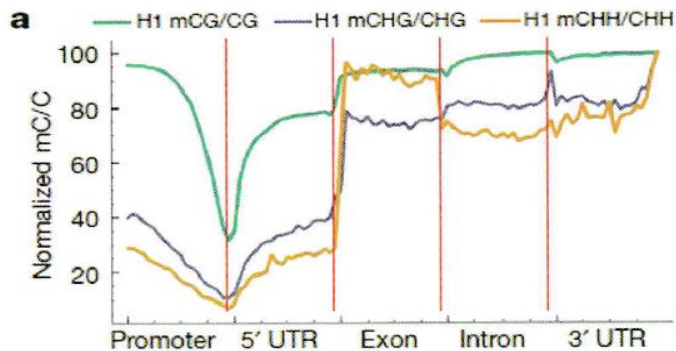


【3】 CpG/CHG/CHH位点染色体分布



【4】 高甲基化位点 (甲基化比例 > 70%) 和低甲基化位点 (甲基化比例 < 30%) 的基因功能区分布

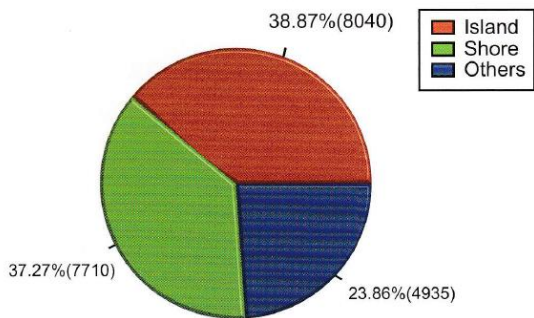
【5】 甲基化程度 (%) 对于基因的不同功能区域的分布





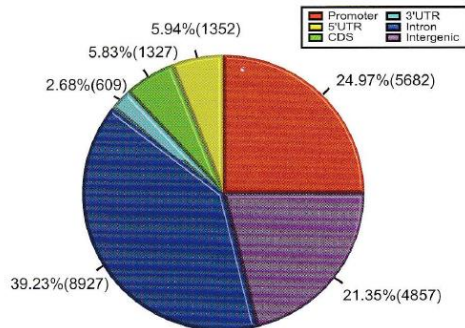
【1】 差异甲基化位点或区段CpG岛分布

CpG content and neighborhood context



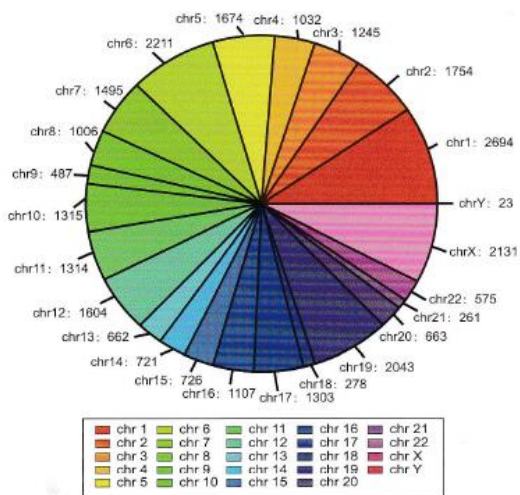
【2】 差异甲基化位点或区段不同基因功能元件分布

Pie Chart of generic features



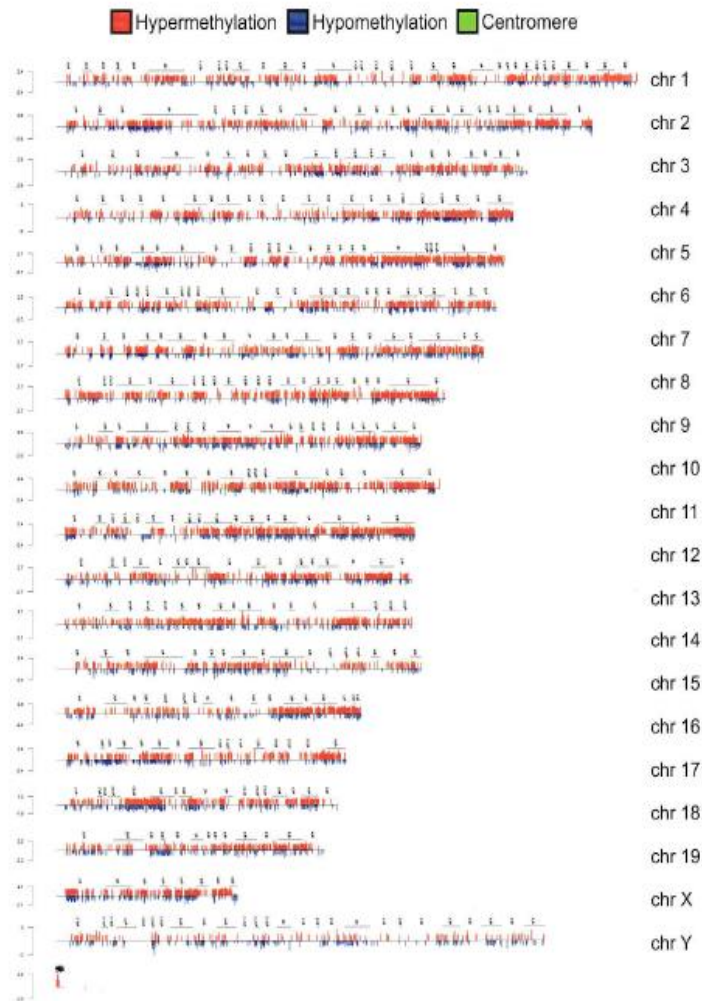
【3】 差异甲基化位点或区段染色体分布

Chromosome location



全基因组甲基化图谱

Chromosome Graph



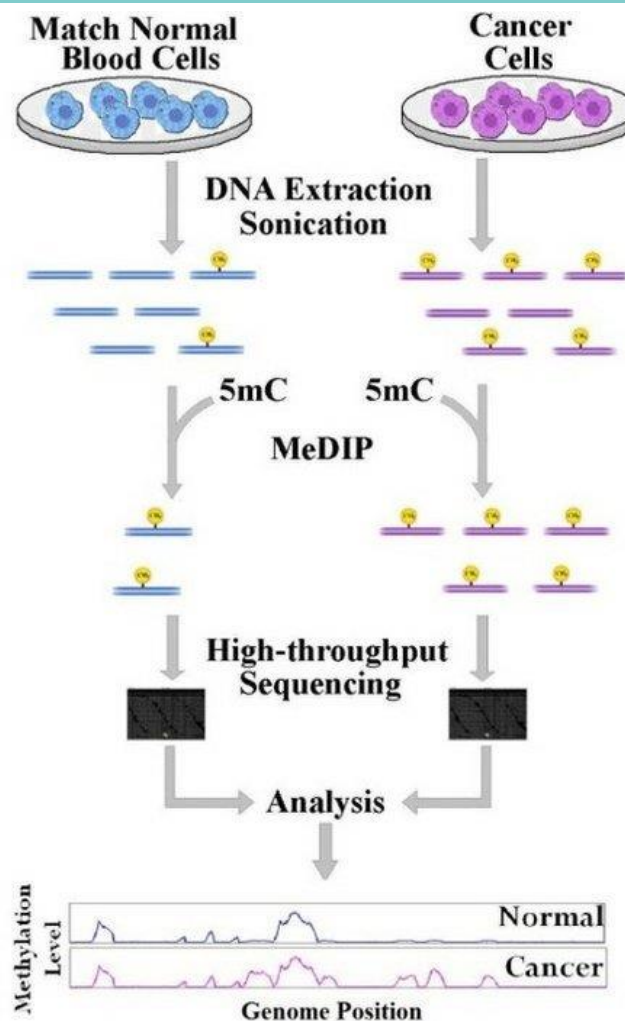




## 二、甲基化DNA免疫共沉淀测序

**MeDIP-Seq** (Methylated DNA Immunoprecipitation Sequencing) 测序是基于抗体富集原理进行测序的全基因组甲基化检测技术，采用甲基化DNA免疫共沉淀技术，通过5'-甲基胞嘧啶抗体特异性富集基因组上发生甲基化的DNA片段，然后通过高通量测序可以在全基因组水平上进行高精度的CpG密集的高甲基化区域研究。

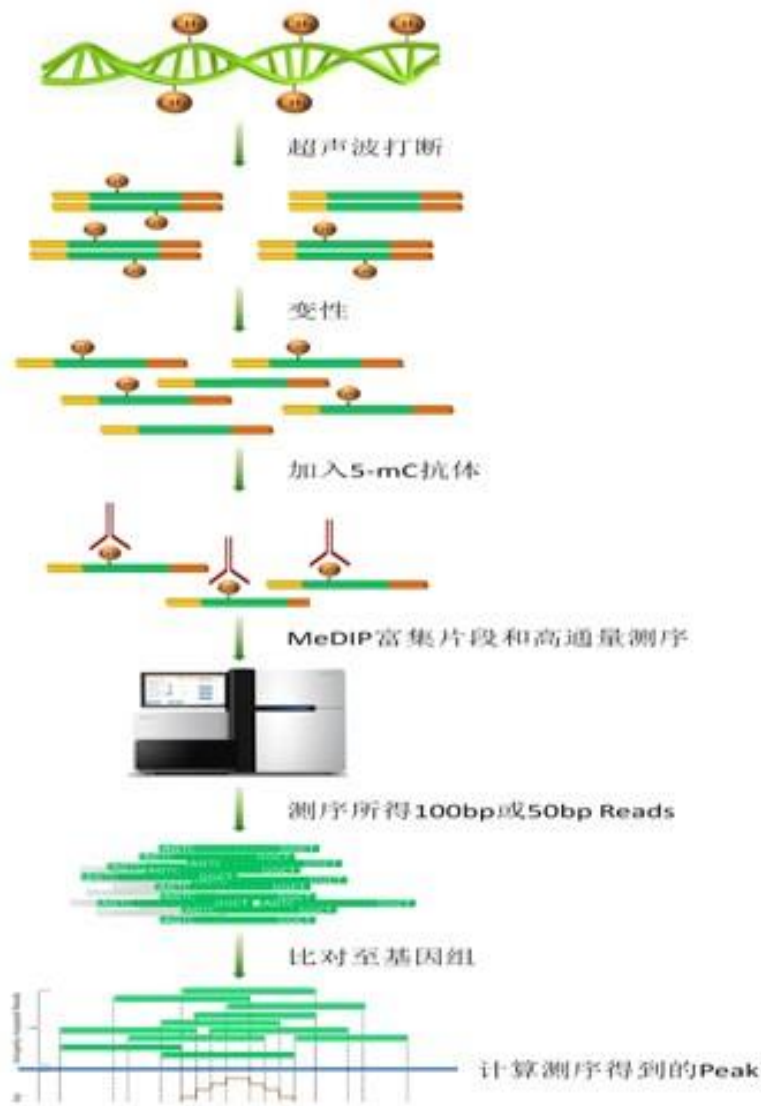
利用MeDIP-Seq技术可以快速有效地寻找基因组上的甲基化区域，从而比较不同细胞、组织或疾病样本间的DNA甲基化修饰模式的差异。



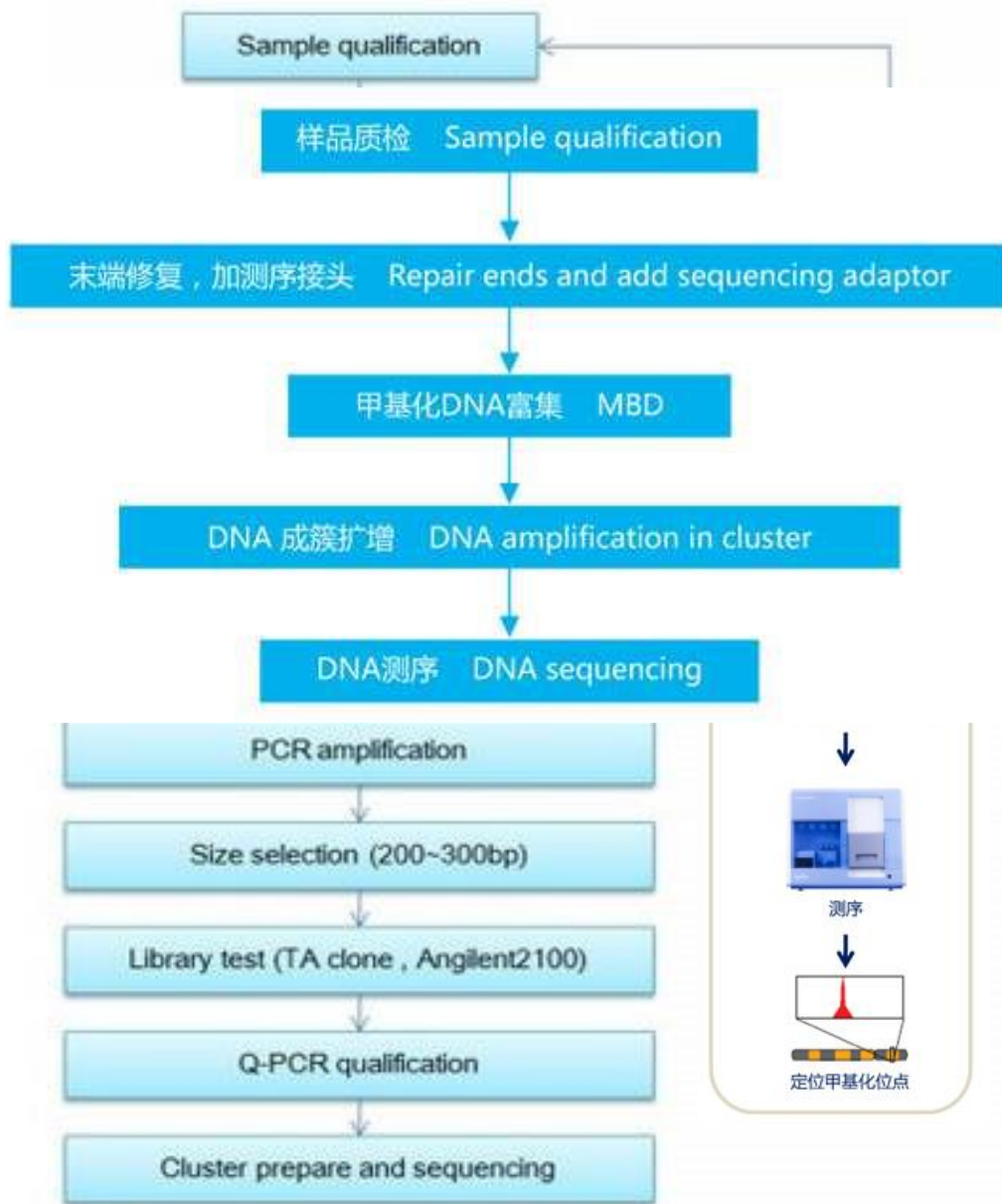


# 武汉大学

Wuhan University



## 技术路线:





## 生物信息分析：

- 1) 基本数据分析：获得原始序列数据（Base calling）、过滤接头。
- 2) 序列数据的数据的质量评估（QC）：QC图（碱基分布、每个循环的质量）。
- 3) 将reads比对到基因组：将预处理reads与reference genome进行mapping，最后得到mapping的sam结果文件，给出mapping结果统计。
- 4) Peak 查找：应用sam文件进行peak富集区查找；组间差异Peak筛选；MeDIP-seq数据和表达谱数据结合分析。
- 5) Peak长度和染色体分布。
- 6) 基因和CpG岛关联：根据找到的peaks的位置信息，确定其关联的基因和CpG岛
- 7) 关联基因GO富集分析：对peaks关联的基因进行GO功能富集分析。





## 样品采集和运输要求

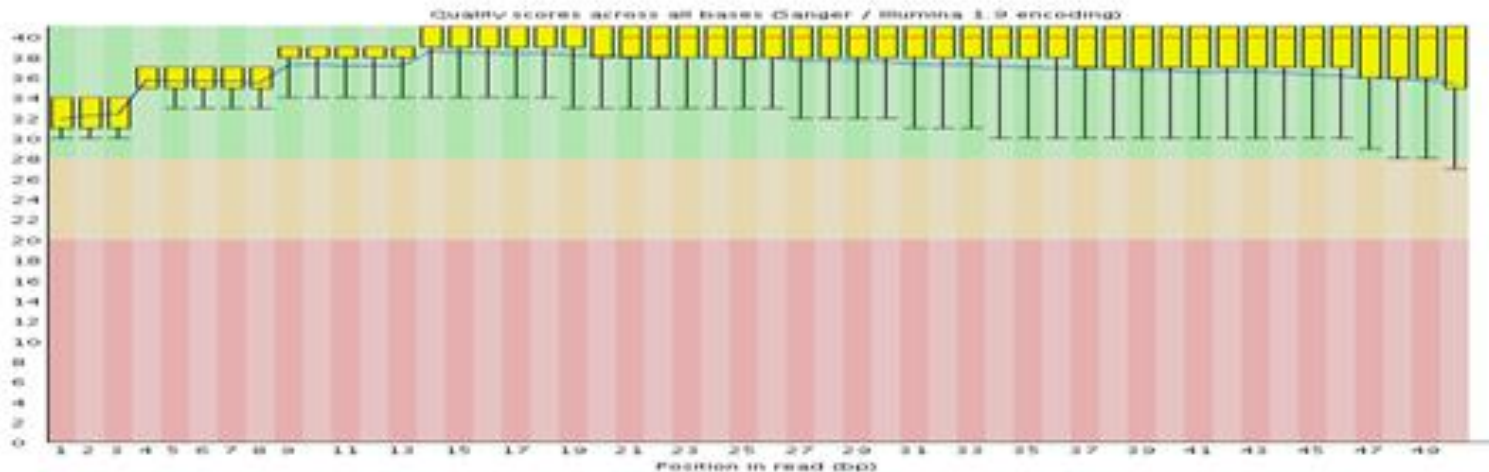
1. 样品纯度：OD 260/280值应在1.8~2.0 之间，OD 260/230 值应大于等于1.5，无降解及RNA污染。
2. 样品浓度：最低浓度不低于50ng/ $\mu$ l。
3. 样品总量：每个样品总量不少于8 $\mu$ g。
4. 样品溶剂：要溶解在H<sub>2</sub>O或TE (pH 8.0)中。
5. 样品运输：DNA低温运输（-20 $^{\circ}$ C）；且在运输过程中请用parafilm将管口密封好，以防出现污染。





## 1. 数据量和数据质量

质量盒图统计 (盒图即是质量分位图, 长方形最下面的边为所有质量值中占25%的质量数值, 依此往上分别为占50%, 占75%对应的质量值, 上下的黑线代表占10%和90%对应的质量值), 蓝色的线表示质量数值的平均值、绿色背景部分代表高质量数值部分、橘色背景部分代表合理质量数值部分、红色背景部分代表低质量数值部分, 横坐标为测序Reads 1~100的位置, 纵坐标为所有Reads在1~100位置的质量(Q:0~40)分布。



质量盒图统计

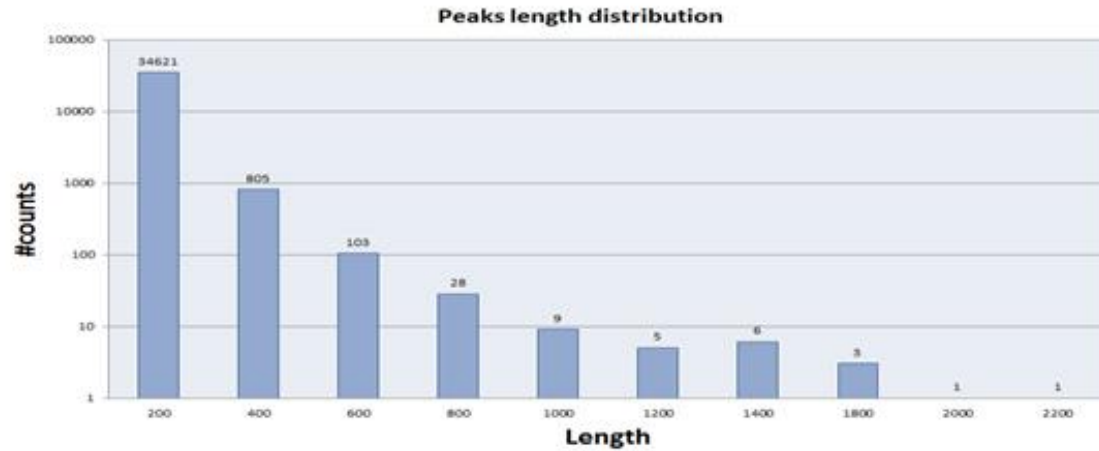




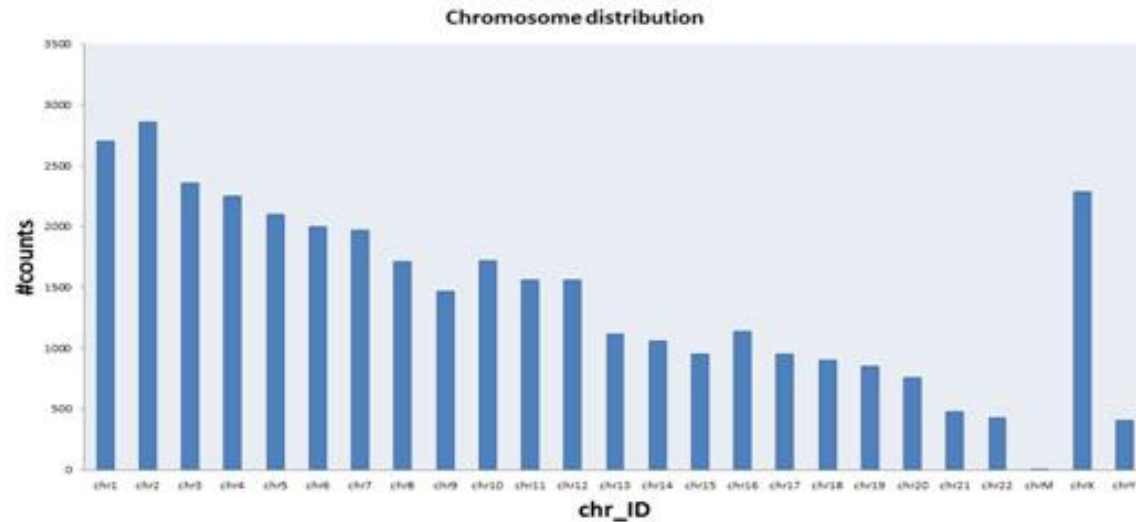


武汉大学

Wuhan University



Peak长度分布



染色体分布





## 三、染色质免疫共沉淀测序

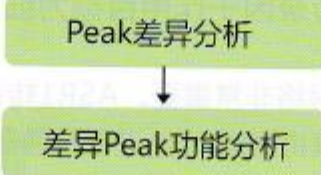
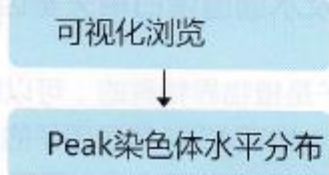
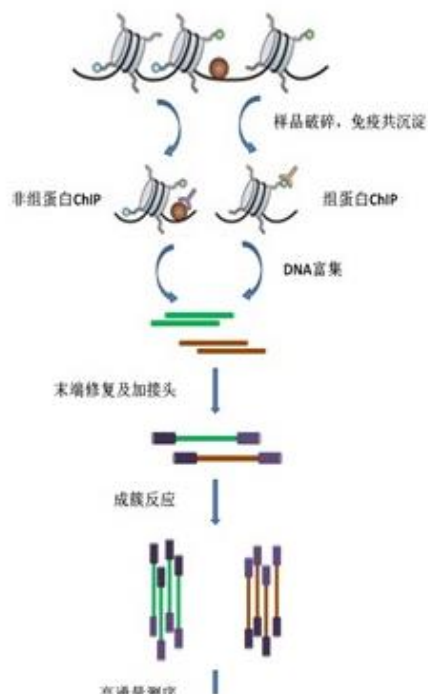
染色质免疫共沉淀(Chromatin Immunoprecipitation,ChIP)是研究蛋白质与DNA相互作用的经典实验方法,广泛的应用于组蛋白特异性修饰位点、转录因子结合位点等研究。

**染色质免疫共沉淀测序** (Chromatin Immunoprecipitation sequencing)

ChIP-Seq的原理是：首先通过染色质免疫共沉淀技术（ChIP）特异性地富集目的蛋白结合的DNA片段，并对其进行纯化与文库构建；然后对富集得到的DNA片段进行高通量测序。研究人员通过将获得的数百万条序列标签精确定位到基因组上，从而获得全基因组范围内与组蛋白、转录因子等互作的DNA区段信息。



## 技术路线

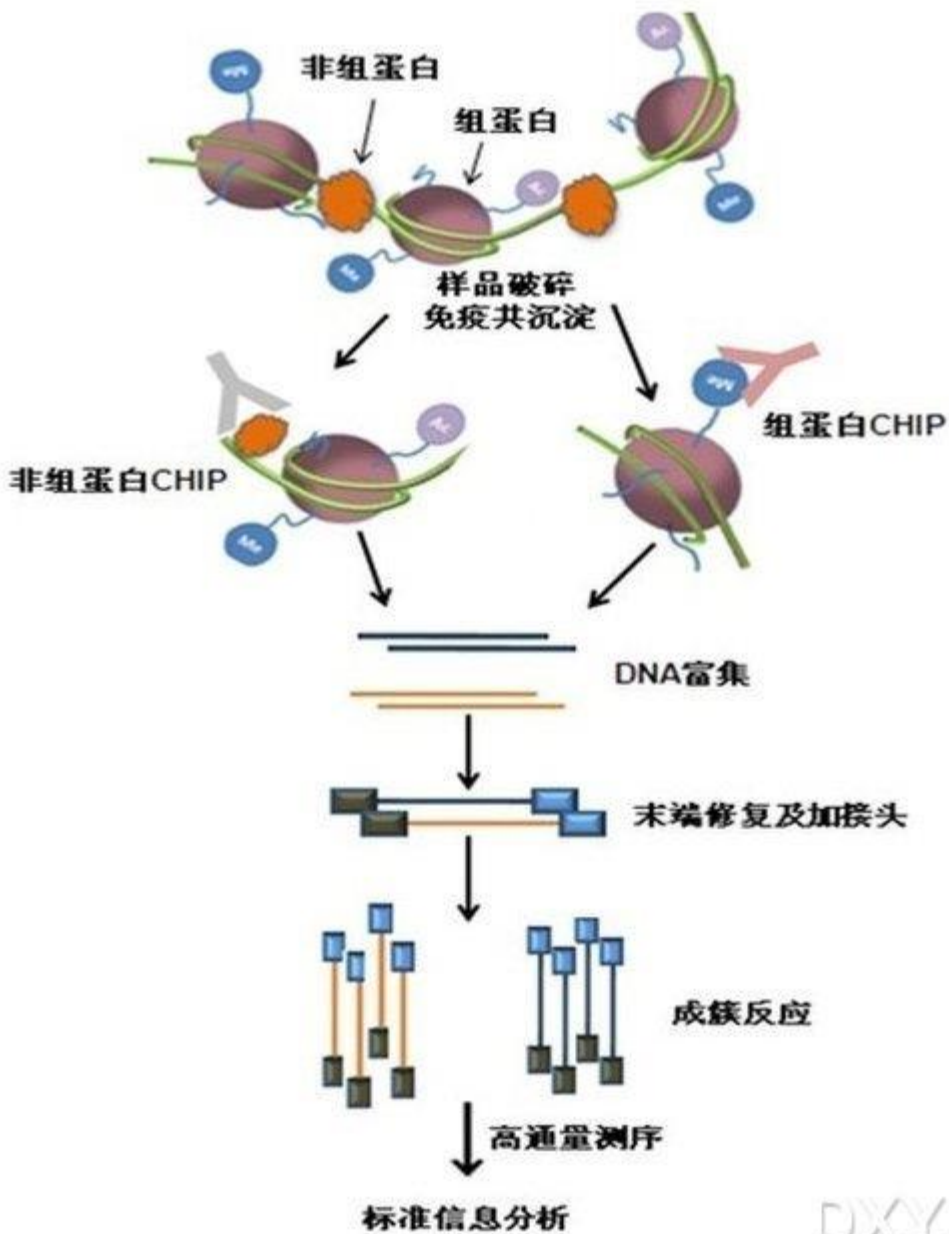


## 实验流程



图2. ChIP建库流程图







## 应用领域

由于ChIP-Seq的数据是DNA测序的结果，为研究者提供了进一步深度挖掘生物信息的资源，研究者可以在以下几方面展开研究：

- (1) 判断DNA链的某一特定位置会出现何种组蛋白修饰；
- (2) 检测RNA polymerase II及其它反式因子在基因组上结合位点的精确定位；
- (3) 研究组蛋白共价修饰与基因表达的关系；
- (4) CTCF转录因子研究。

## ChIP-Seq相比ChIP-chip有哪些优势？

第一，ChIP-Seq 能实现真正的全基因组分析。目前所能获得的芯片上固定的探针只能代表全基因组部分序列，所获得的杂交信息具有偏向性；

第二，对于结合位点分析，ChIP-Seq 通过寻找“峰”，结合分辨率可精确到10~30 bp，而芯片上探针由于长度所限，无法精确定位，即使目前最高水平的商业芯片都无法提供可与ChIP-Seq 媲美的分辨率；

第三是所需样本数量。ChIP-chip 需要多达4~5  $\mu\text{g}$  的起始样本，在杂交之前需要进行LM-PCR，但可能导致背景增高，竞争性扩增等导致假阳性。而ChIP-Seq 仅需要纳克级起始材料，如SOLiD 起始材料可低至20ng。







## CHIP-on-chip 与技术特点如下:

研究方法	CHIP-on-chip	CHIP-Seq
分辨率	30~100bp	1bp
覆盖范围	受芯片容量限制, 只能选择性地扫描特定区域, 无法覆盖全基因组	只要测定的序列 (Reads) 能够定位到基因组上, 就能获得全部基因组信息
缺陷	探针和非特异性区域杂交	测序数据会有一些 GC 含量偏向
性价比	只能研究在基因组上广泛存在的目的位点 (Broadening binding)	可以扫描全基因组; 可以研究在基因组上存在的稀有目的位点 (Sharp binding)
需要的 DNA 量	高	低 (10~50bp)
动态量程	弱信号会被遗弃; 强信号会饱和	没有局限
选择数据产出量	不可以	可以





武汉大学

Wuhan University

谢谢!



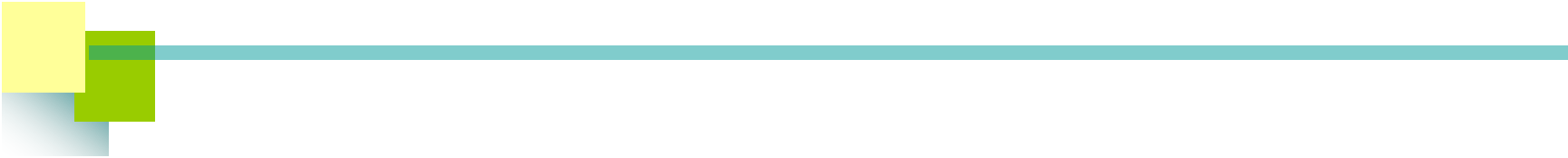
Laboratory of plant molecular cytogenetics





武汉大学

Wuhan University



Laboratory of plant molecular cytogenetics

