

ATAC/CUT&Tag基础培训

生信帮

云南农业大学

云南省生物大数据重点实验室

2023年8月5日

表观遗传学



雲南農業大學
Yunnan Agricultural University



生信帮

云南省生物大数据重点实验室

表观遗传学研究的是基因组中**不涉及DNA序列变化**的遗传信息传递和表达。与传统的遗传学关注基因座上的碱基序列不同，表观遗传学研究着眼于基因组中的**化学修饰、染色体结构**和**非编码RNA**等因素对基因表达的调控。

表观遗传学研究的重点包括DNA甲基化、组蛋白修饰、染色质结构、非编码RNA等各种表观遗传标记以及它们与基因调控网络之间的相互作用。这些表观遗传标记可以在细胞和组织之间稳定传递，并对基因表达模式和细胞功能产生影响。

表观遗传研究的意义在于揭示了**基因表达的调控机制**，并对理解生物发育、疾病发生机制以及环境对基因表达的影响提供了重要线索。通过表观遗传研究，我们可以深入了解基因组的复杂性，以及环境和遗传因素在表达型多样性和疾病易感性中的相互作用。

总之，表观遗传学研究在揭示基因调控机制、理解生物发育和疾病发生机制以及发展个性化医疗方面具有重要意义。它为我们深入了解基因组的复杂性提供了新的视角，并推动了基因组医学的发展。





ChIP-seq（染色质免疫沉淀测序）是一种广泛应用于生物学研究中的高通量测序技术，用于研究染色质上与蛋白质和DNA互作。以下是ChIP-seq测序的基本原理：

染色质免疫沉淀（ChIP）：首先，将染色质中与特定蛋白质结合的区域选择性地免疫沉淀下来。这可以通过交联蛋白质和DNA，然后使用适当的抗体对目标蛋白质进行免疫沉淀来实现。

DNA提取：随后，通过酶解和其他化学处理方法将沉淀下来的染色质与蛋白质交联解离，并将DNA分离出来。

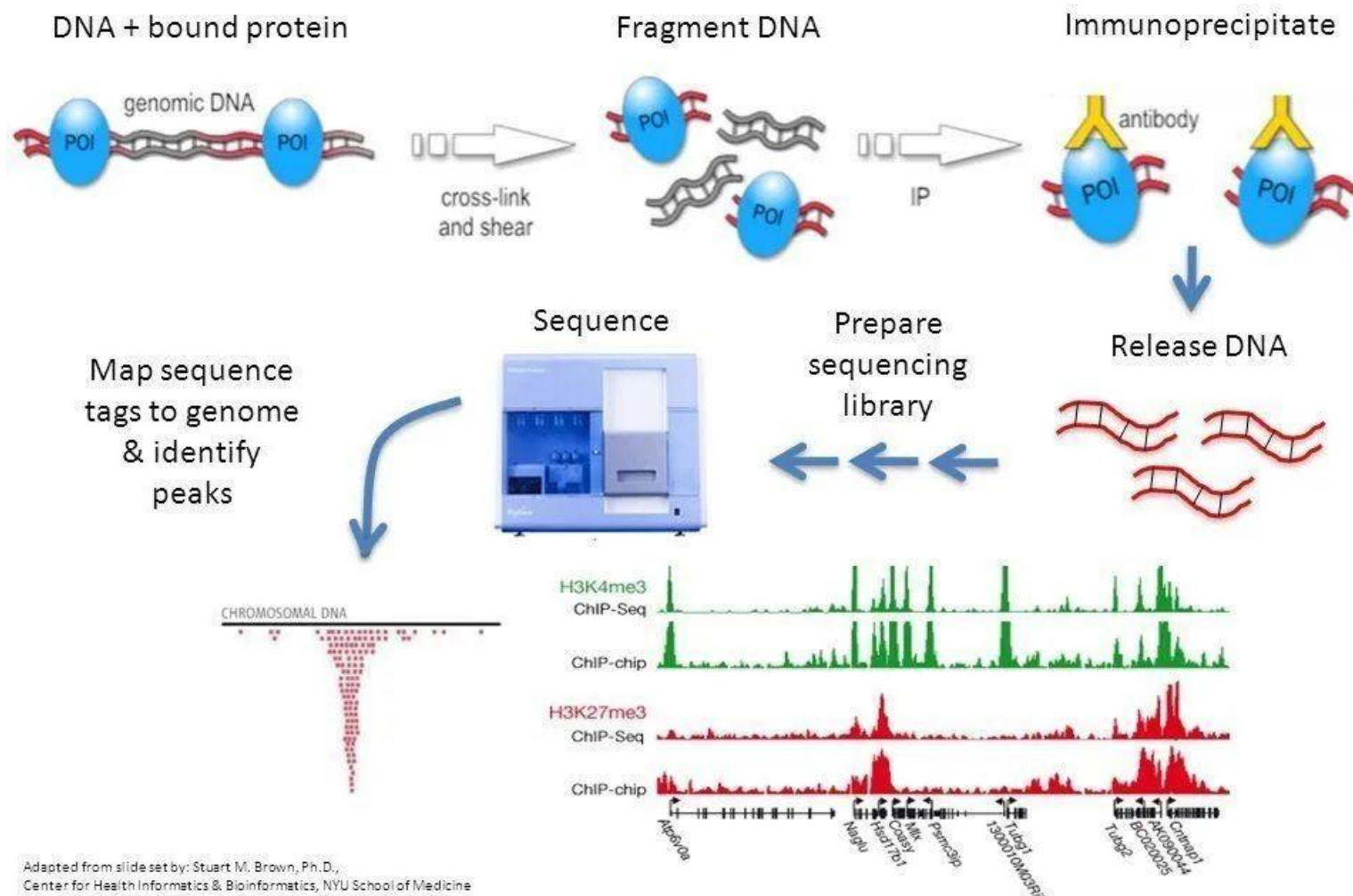
DNA片段化：接下来，将提取的DNA进行机械或酶切割，将其分为较短的片段。这些片段通常具有一定的长度范围，例如150到300碱基对。

测序库构建：将片段化的DNA进行末端修复、连接特定的DNA连接器（adapters），然后进行PCR扩增，生成测序所需的文库。此文库包含了染色质中与目标蛋白质结合的DNA片段。

高通量测序：最后，将测序文库进行高通量测序，通常使用Illumina测序技术。测序得到的短读长片段将被映射到参考基因组上，从而确定DNA片段是来自哪些区域。

通过分析ChIP-seq测序结果，可以确定与目标蛋白质结合的DNA序列在基因组上的定位和丰度。这对于研究转录调控、染色质结构和功能等方面非常有用。

ChIP-seq overview



CUT&Tag技术



雲南農業大學
Yunnan Agricultural University

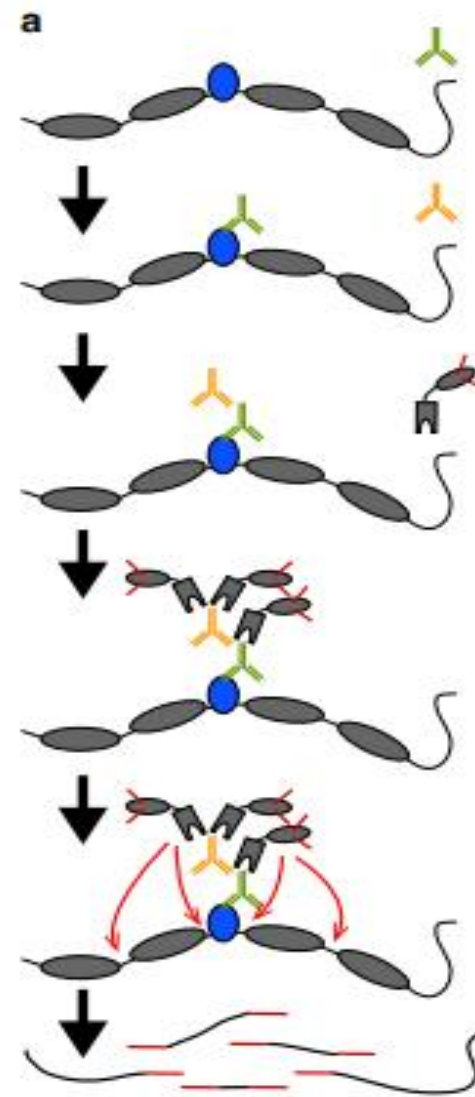
云南省生物大数据重点实验室



生信帮

CUT&Tag技术是一种高效的染色质分析方法，其原理如下：

1. 细胞固定：首先，将待研究的活细胞与特异性抗体（一抗）混合，使一抗能够结合到目标蛋白上。
 2. 抗体结合：二级抗体（二抗）与一抗结合，并且二抗上携带有特定的分子标记（例如酶、荧光染料等）。这样做的目的是通过放大信号，增强抗体结合的检测敏感性。
 3. 转座酶结合：使用pA/G-Tn5融合蛋白，它能够与一抗和二抗结合。这个融合蛋白携带有转座酶（Tn5）和DNA结合域（pA/G），通过与一抗和二抗结合，将转座体间接固定在靶蛋白上。
 4. 靶标位点切割：添加Mg²⁺离子来激活Tn5转座酶，使其识别并切割靶标区域。切割的同时，在片段化的DNA两端直接添加测序接头，这是因为Tn5转座体本身携带有测序接头。
 5. 文库构建：在切割后，使用PCR扩增来扩增带有测序接头的片段化DNA，并得到完整的DNA文库。这些文库可以用于后续的高通量测序，以研究特定蛋白质与DNA之间的相互作用。
- 总的来说，CUT&Tag技术通过利用抗体和转座酶的组合，实现了对靶蛋白结合位点的准确鉴定和高效分析，为研究蛋白质-DNA相互作用提供了一种强大的工具。



ChIP-seq Vs CUT&Tag



雲南農業大學
Yunnan Agricultural University

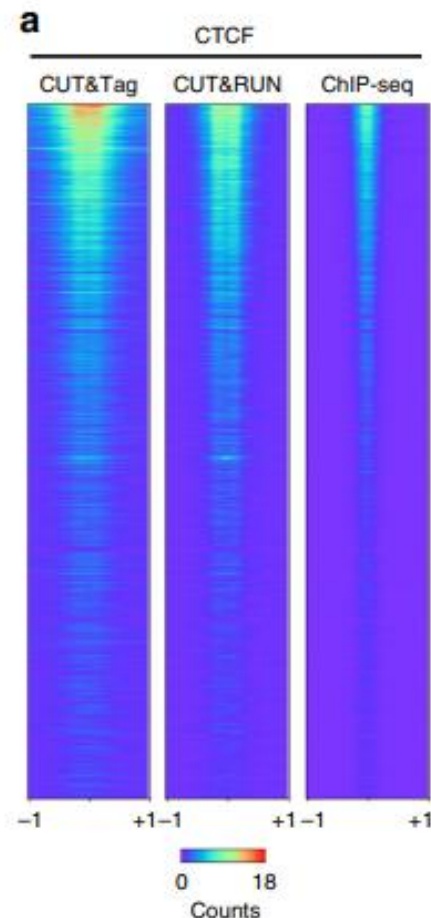
云南省生物大数据重点实验室



生信帮

快速和高效； 低背景信号， 无需要input;更少的DNA需求;更简化的实验流程

区别点	ChIP-Seq	CUT&Tag
是否需要固定	需要	不需要
染色质片段化方法	超声或限制性内切酶	基于Tn5的片段化
所需细胞数量	20-100万	可应用于单细胞
所需测序深度*	20-50M reads	2M reads**
是否整合建库过程	否， 需要另外的建库步骤	是， 通过片段化加Tn5接头
适用的靶点	广泛适用于组蛋白修饰、转录因子和辅因子	主要适用于组蛋白修饰， 少许转录因子和辅因子
所需时间	36-48小时	< 12小时

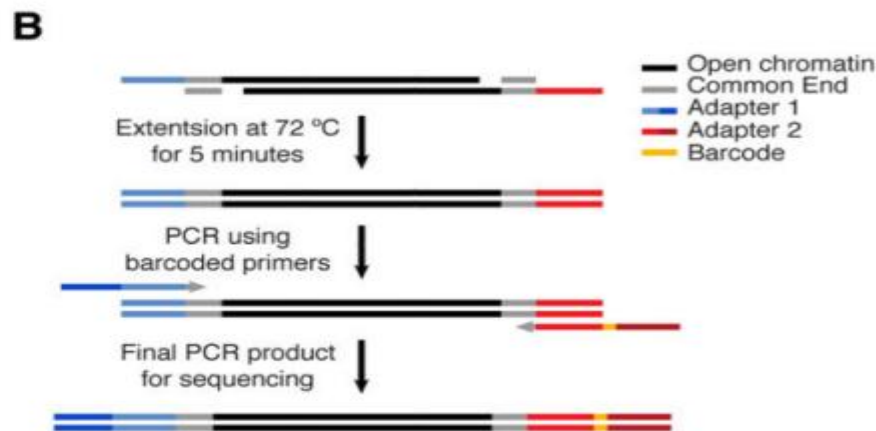
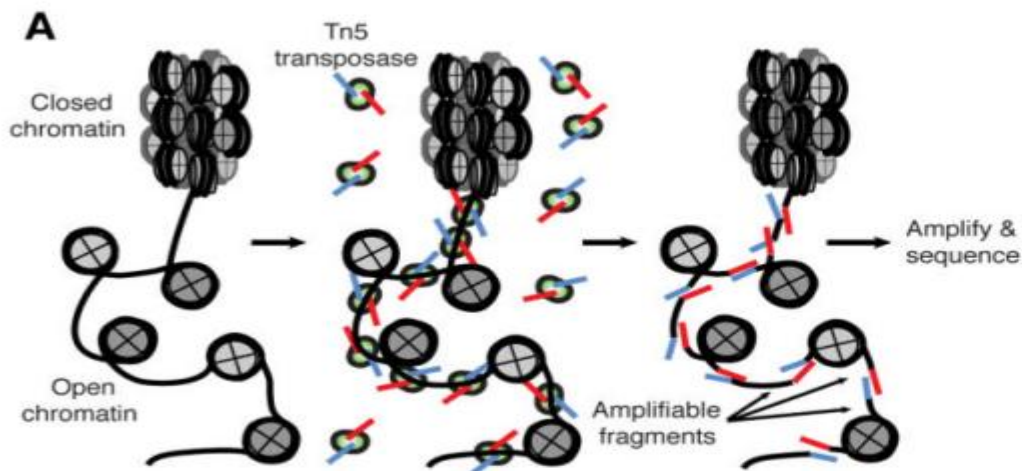




ATAC (Assay for Transposase-Accessible Chromatin) 技术是一种用于分析**染色质可及性**的高通量测序方法，它可以帮助我们确定**基因组中开放染色质的位置**。这些区域代表着细胞中的活动基因组区域，包括启动子、增强子等。这种方法简单而高效，能够帮助我们全面了解基因组中的开放染色质和基因调控元件，以及它们与表观遗传学和基因表达之间的关系。

ATAC技术的原理如下：

- 细胞固定和渗透：首先，将待研究的活细胞进行固定，并使用非离子性洗涤剂（如Triton X-100）使得细胞膜渗透。这一步骤有助于使转座酶能够进入细胞内。
- 转座反应：使用一种活化的转座酶（例如Tn5转座酶），将转座酶与DNA结合并切割开放染色质的特定区域。转座酶在切割的同时会直接嵌入测序适配体。
- 洗脱和扩增：将测序适配体与转座酶结合的DNA片段从细胞中洗脱出来，并进行PCR扩增。PCR扩增的目的是为了增加片段数量以便后续测序的需要。
- 文库构建和测序：通过对PCR扩增产物进行文库构建，将其转化为测序所需的DNA文库。然后，使用高通量测序技术（例如Illumina测序）对文库进行测序。
- 数据分析：通过分析测序数据，可以确定开放染色质区域的位置和特征。



分析流程

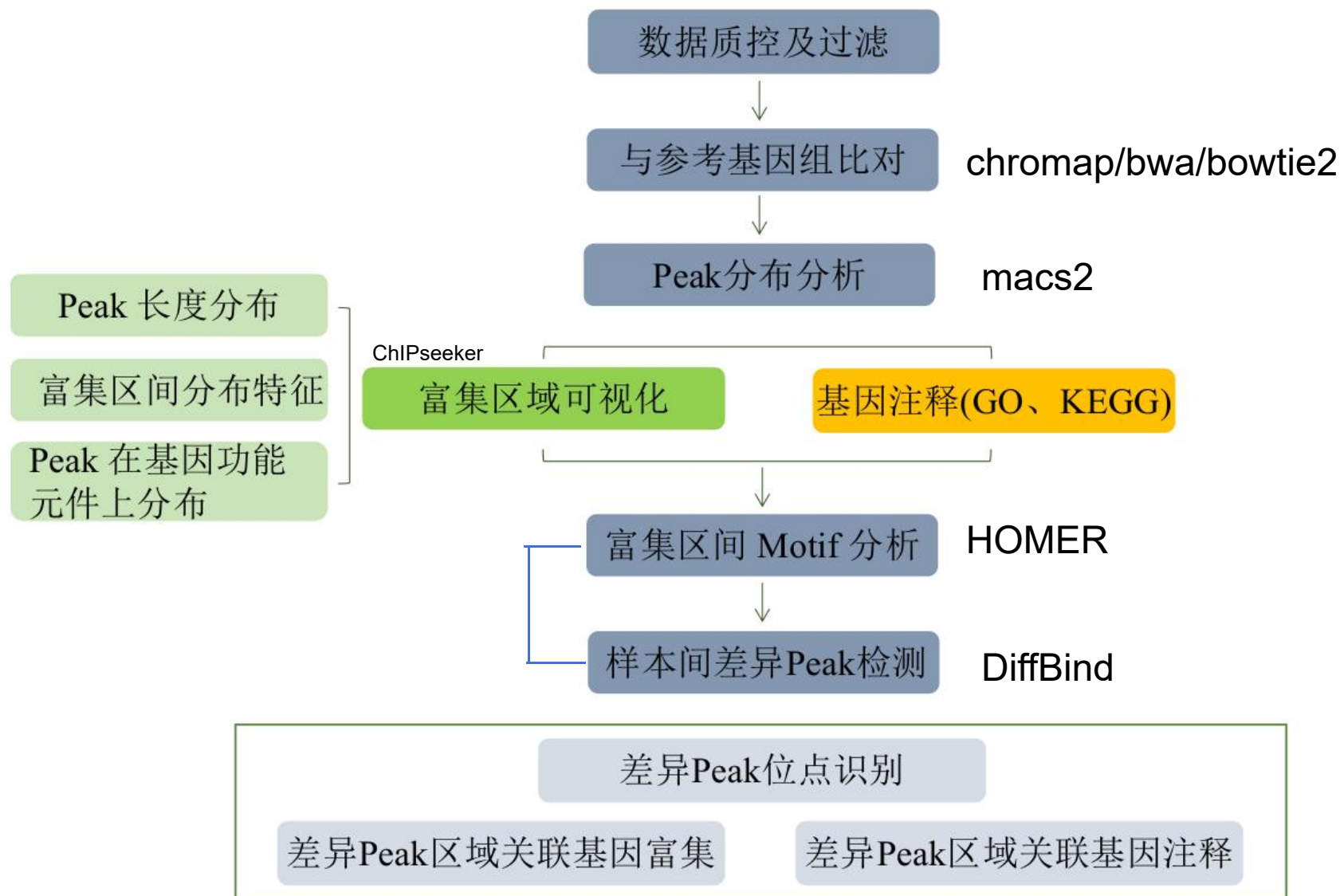


雲南農業大學
Yunnan Agricultural University

云南省生物大数据重点实验室



生信帮



ATAC文库评估

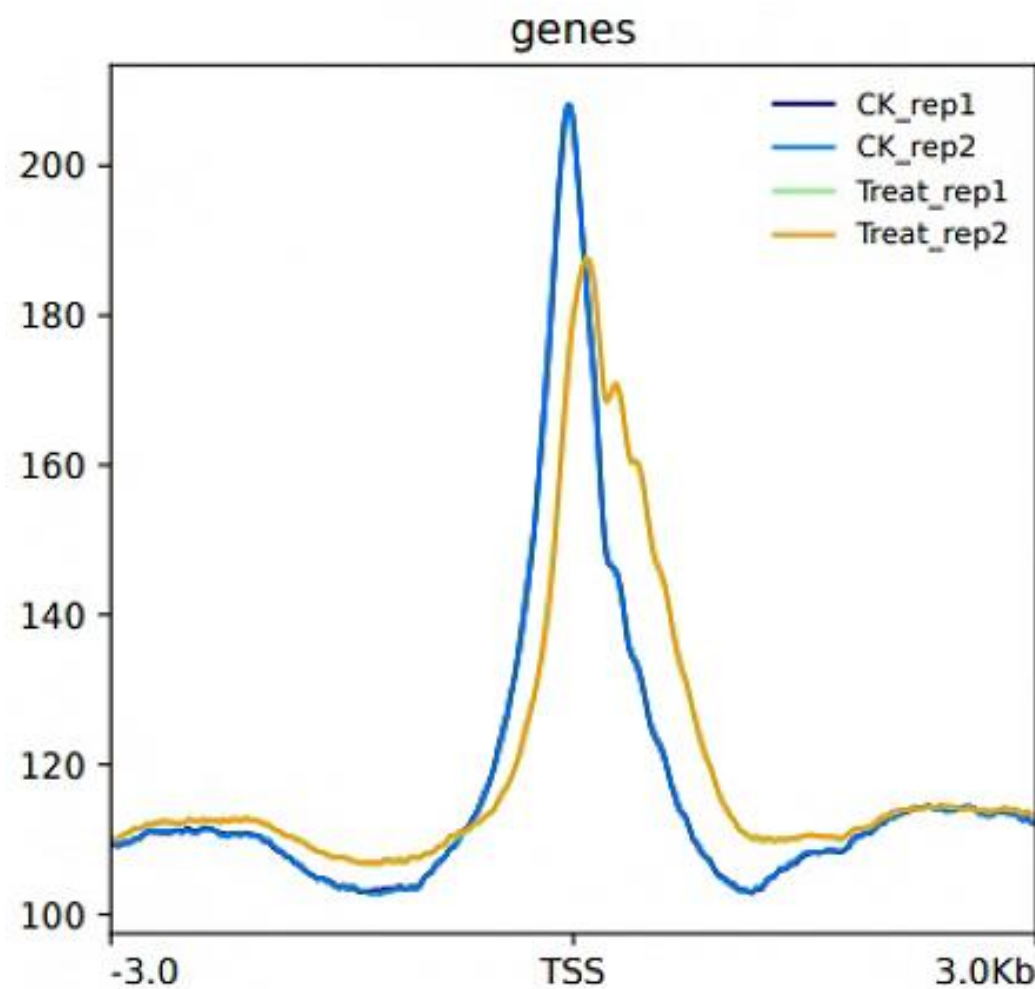
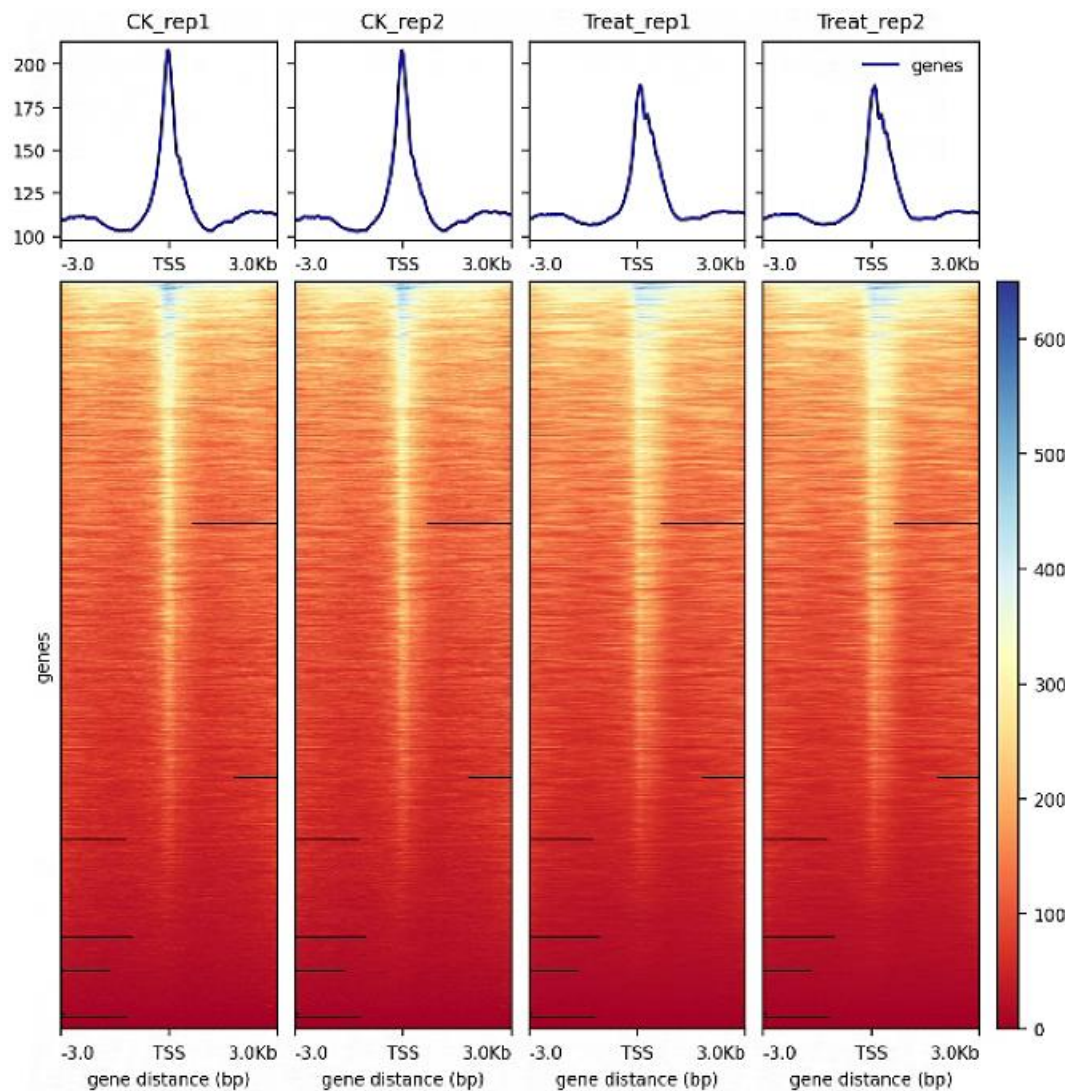


雲南農業大學
Yunnan Agricultural University

云南省生物大数据重点实验室



生信帮



比对到参考基因组序列的reads会在DNA开放区域(或者蛋白质结合区域) 位点及其附近成富集趋势，通过识别 reads比对结果，可以获取富集峰（Peak），即可获得潜在的染色质开放区域。由于测序得到的read在基因组上的分布是随机的，因此某个碱基上覆盖的read的数目应该服从二项分布。当总read数很大时，二项分布可以近似用泊松分布代替。根据分布情况，可以算出在某个置信概率（如0.05）下（随机情况下），某个碱基上可以覆盖的read的数目的最小值。当实际观察到的read数目超过这个值（单侧检验）时，则认为该碱基是一个染色质开放区域热点。

我们保留Unique mapped Reads ($\text{MapQ} \geq 30$)，并去掉PCR扩增导致的冗余的比对结果，最后使用MACS v2.1.1软件进行峰提取。MACS2包含四步:去除冗余的Reads,调整Read位置,计算峰富集 (peak enrichment)和估计经验错误发现率(FDR)。

差异Peak



云南农业大学
Yunnan Agricultural University

云南省生物大数据重点实验室



生信帮

1) 全基因组范围转录因子的调控差异

比较**不同类型样本**中的ATAC染色质可及性的差异，推导出对应的motif的差异，进一步获得不同样本转录因子表达的差异性，因而可以结合RNA-seq的结果，从源头推测出转录差异的原因是否是由转录起始的某些调控转录因子引起的。

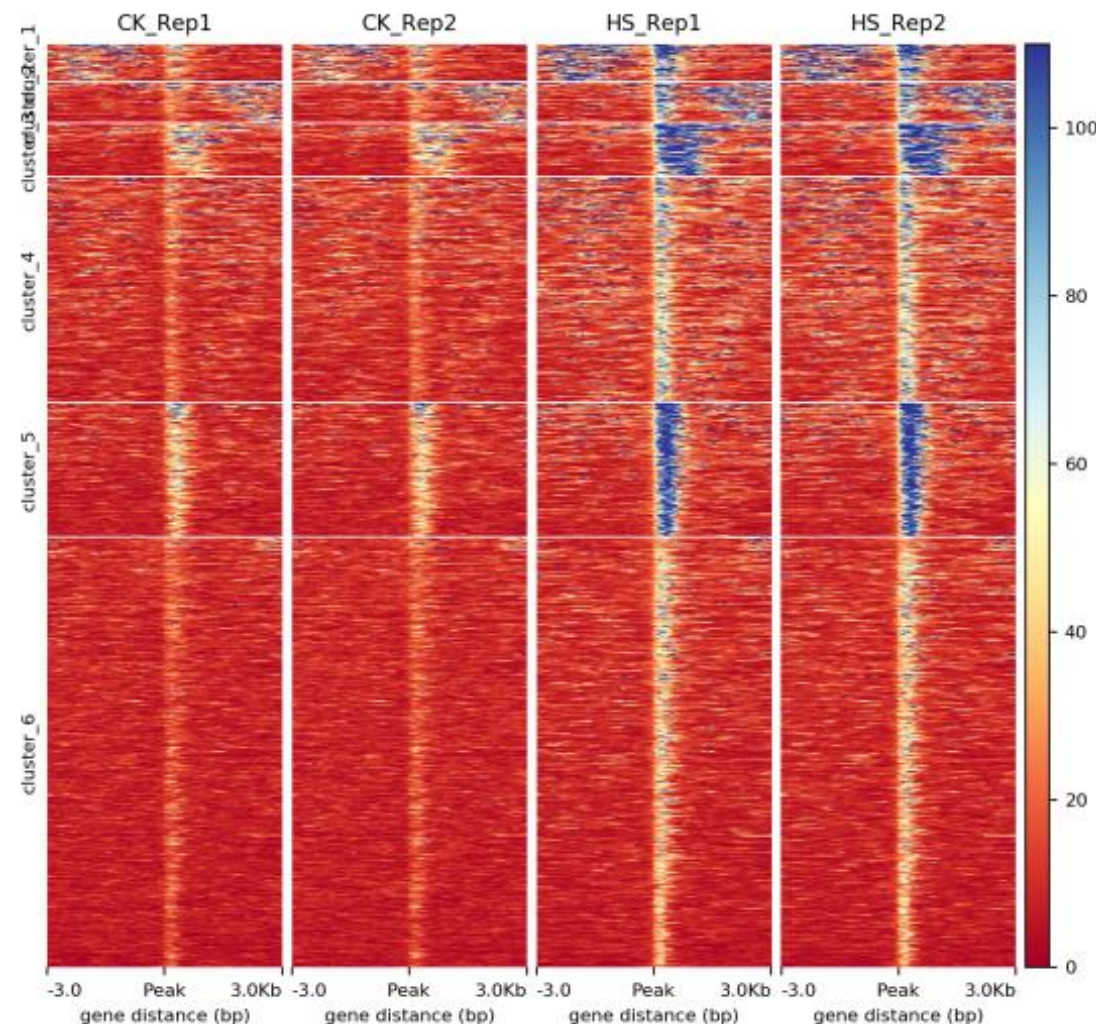
2) 特定转录因子的检索

对于无法获得特定转录因子特异性抗体，可以采用ATAC-seq进行更为全面的研究。利用鉴定的Motif找到关注的转录因子，观察两个不同样本的差异。

3) 差异Peak鉴定

采用DiffBind包进行差异Peak分析。具体为计算共有Peak在每个样品中支持的read count数，并基于count数来计算affinity得分(即标准化后的read)。affinity得分作为DESeq2 v1.24.0软件输入，用于对每组样品进行差异筛选，筛选标准： $|\log_2(\text{fold change})| \geq 1$ ， $\text{FDR} \leq 0.01$ 。

使用deepTools 工具，采用kmeans方法进行聚类，绘制差异Peak在每组样品中的测序Read的富集热图。



富集分析



雲南農業大學
Yunnan Agricultural University



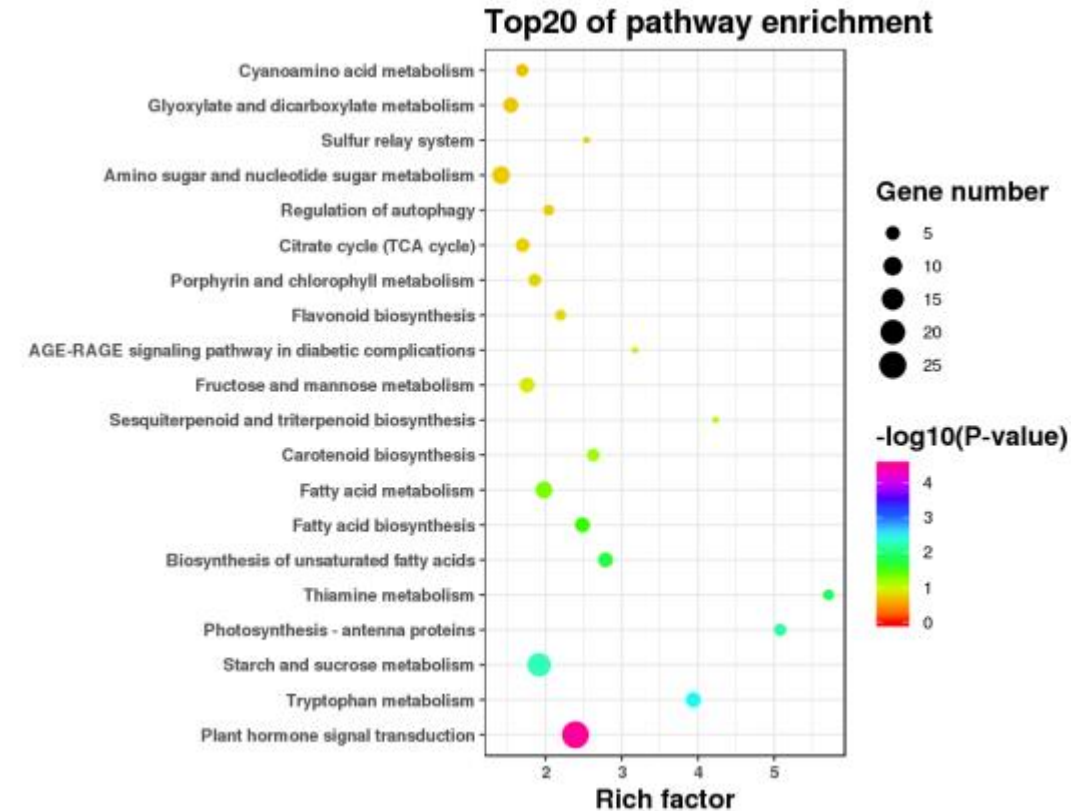
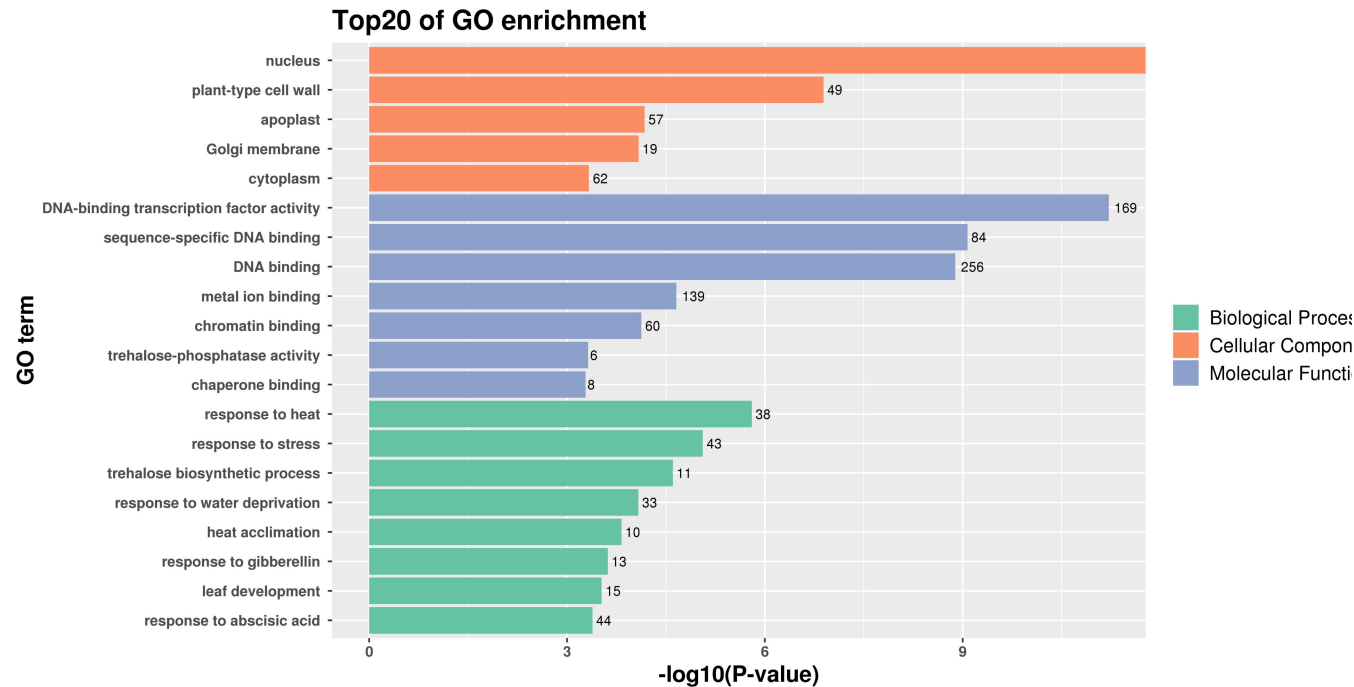
生信帮

云南省生物大数据重点实验室

对差异Peak关联的基因进行GO富集分析,, 同时利用Fisher进行显著性检验 ($p \leq 0.01$)。下图中展示了前20个最为富集的GO term。

对关联的基因进行KEGG富集分析, 并用Fisher进行显著性检验 ($p \leq 0.05$) , 下图中展示了前20个最为富集的pathway条目。

软件: clusterprofile



Motif分析



雲南農業大學
Yunnan Agricultural University

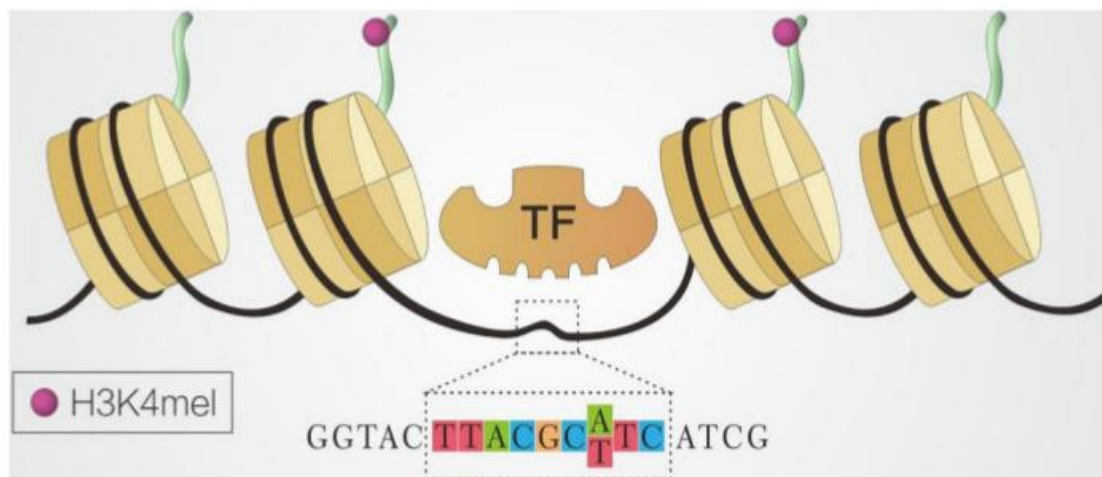


生信帮

云南省生物大数据重点实验室

Motif 是具有某种生物特性,在一组相关序列中重复出现,与某些蛋白质高亲和度的特定碱基序列模式,是研究基因表达调控机制的重要信息。ATAC-seq/CUT&Tag获得的染色质peak区域的motif,可推测其结合的转录因子等调控蛋白。

差异Peak的Motif分析, 本处我们主要通过HOMER 中的findMotifsGenome.pl脚本对差异的peak进行motif的鉴定, 主要的参数为 -size given -len 6,8,10,12。最终获得两个结果一个是从头预测的motif结果 (homerResults.html), 另外一个是根据数据库中已知的motif进行序列匹配获取的结果 (knownResults.html), 同时最后我们获取鉴定的motif在peak和整个基因组上的位置。



Motifs in accessible chromatin (KIT ⁺ vs PS)		
	E value	TF-binding sites in the motifs
KIT ⁺ -unique	 4.8x10 ⁻⁵⁷	SOX10 HOMEZ NFATC1
	 1.5x10 ⁻³²	FOS JUND JDP2
	 2.7x10 ⁻¹⁹	NR5A2 ESRRA NR2F6
PS-unique	 3.2x10 ⁻²²⁰	LHX2 NFYA NFYB
	 3.6x10 ⁻¹³⁶	POU2F2 POU3F1 POU5F1
	 4.3x10 ⁻¹⁷⁴	POU6F1 POU2F1 RARA
		SOX10 HOMEZ NFATC1

转录因子



雲南農業大學
Yunnan Agricultural University

云南省生物大数据重点实验室

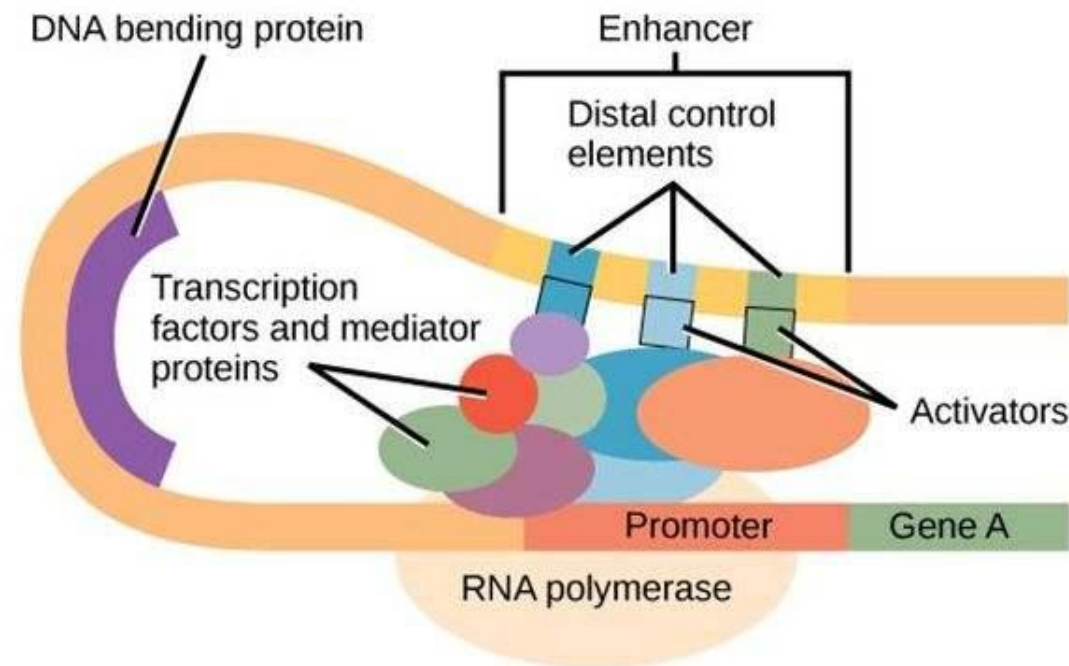


生信帮

转录因子(transcription factor)是一群能与基因5'端上游特定序列专一性结合，从而保证目的基因以特定的强度在特定的时间与空间表达的蛋白质分子。

真核生物转录起始往往需要多种蛋白因子的协助，转录因子与RNA聚合酶 II 形成转录起始复合体，共同参与转录起始的过程。

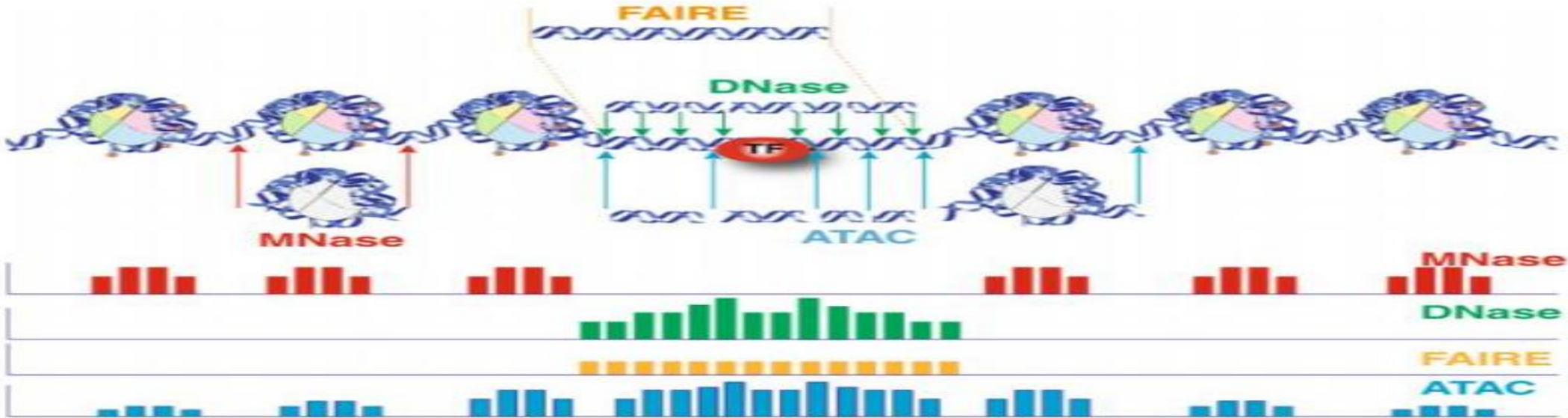
根据转录因子的作用特点可分为二类；第一类为普遍转录因子,它们与RNA聚合酶 II 共同组成转录起始复合体；第二类转录因子为**组织细胞特异性转录因子**，这些TF是在特异的组织细胞或是受到一些类固醇激素\生长因子或其它刺激后，开始表达某些特异蛋白质分子时。



其他染色质可及性技术



研究方法	优点	缺点
MNase-seq	直接定位核小体或转录因子占据的区域	需要大量细胞，酶的用量需要精准控制
DNase-seq	开放区域	需要大量细胞，样本制备过程复杂，酶的用量需要精准控制
FAIRE-seq	直接获取开放区域，灵敏度高	信噪比低，数据解读困难
ATAC-seq	直接获取开放区域，细胞需求量减少，灵敏度高，重复性好	容易引入线粒体污染



总结



雲南農業大學
Yunnan Agricultural University

云南省生物大数据重点实验室



生信帮

- ATAC-Seq与CUT&Tag的不同的是ATAC-Seq是全基因组范围内检测染色质的开放程度，可以得到全基因组范围内的蛋白质可能结合的位点信息，一般用于不知道特定的转录因子，用此方法与其他方法结合筛查感兴趣的特定调控因子：
- CUT&Tag是明确知道感兴趣的转录因子是什么，根据感兴趣的转录因子设计抗体去做ChIP实验拉DNA，验证感兴趣的转录因子是否与DNA存在相互作用。
- ATAC-Seq、CUT&Tag、Dnase-Seq、MNase-Seq、FAIRE-Seq整体的分析思路一致，找到富集区域，对富集区域进行功能分析。
- CUT&Tag是揭示特定转录因子或蛋白复合物的结合区域，实际是研究DNA和蛋白质的相互作用，利用抗体将蛋白质和DNA一起富集，并对富集到的DNA进行测序。
- DNase-Seq、ATAC-Seq、FAIRE-Seq都是用来研究开放染色质区域。DNase-Seq是用的DNase内切酶识别开放染色质区域，而ATAC-seq是用的Tn5转座酶，随后进行富集和扩增；FAIRE-Seq是先进行超声裂解，然后用酚-氯仿富集。



雲南農業大學
Yunnan Agricultural University

云南省生物大数据重点实验室

生信帮，答你所问