

基因组学、转录组学 基因的结构分析-3

张百芳



利用核酸-蛋白质互作方法研究启动子



启动子是一段能被蛋白质识别和结合的DNA序列，故检测核酸-蛋白质相互作用的研究方法都可以用于启动子的研究中。

主要方法：

- 足迹法（酶足迹法，化学足迹法）
- 电泳迁移率变动实验（EMSA）
- 染色体免疫沉淀（ChIP）



1. 足迹法 (*Footprinting*) : 利用 *DNA* 电泳条带连续性中断的图谱特点判断与蛋白质结合的 *DNA* 区域

基本流程:

DNA与蛋白质相互作用



酶切割DNA



DNA变性凝胶电泳



分析电泳图谱



(1) **酶足迹法 (Enzymatic footprinting)**: 利用能切割DNA的酶处理DNA-蛋白质混合物, 然后通过电泳进行分析

- **DNase I足迹法 (DNase I footprinting)**

利用DNase I 随机切割双链DNA, 从而确定DNA结合蛋白在DNA上结合位点。

- **核酸外切酶III足迹法 (Exonuclease III footprinting)**

利用核酸外切酶III (Exo III) 的3'→5'外切酶活性从3'末端切割双链DNA的特性, 确定蛋白质在DNA上的结合位点。



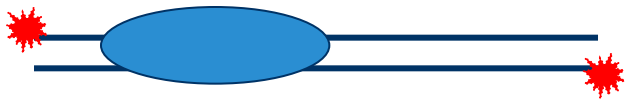
dsDNA



单链末端标记

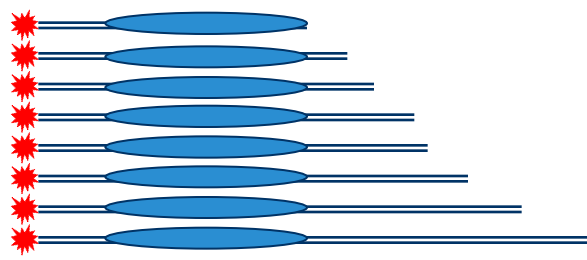


DNA结合蛋白



酶切消化
(控制反应时间)

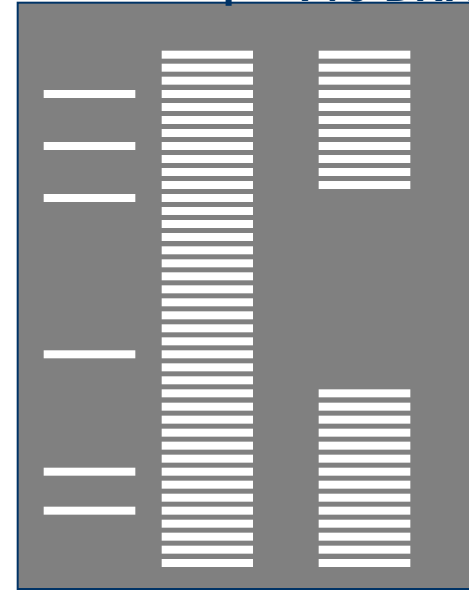
DNase I



产生长短不同的片段；
但蛋白质结合区被保护

DNase I 足迹法

M No-pro Pro-DNA



变性凝胶电泳



对凝胶上空白区域的DNA进行克隆测序,即可确定结合蛋白质的DNA序列



(2) **化学足迹法 (Chemical footprinting)**：利用能切断DNA骨架的化学试剂处理DNA-蛋白质复合物，从而通过化学试剂无法接近结合蛋白质的DNA区域而确定DNA的蛋白质结合位点

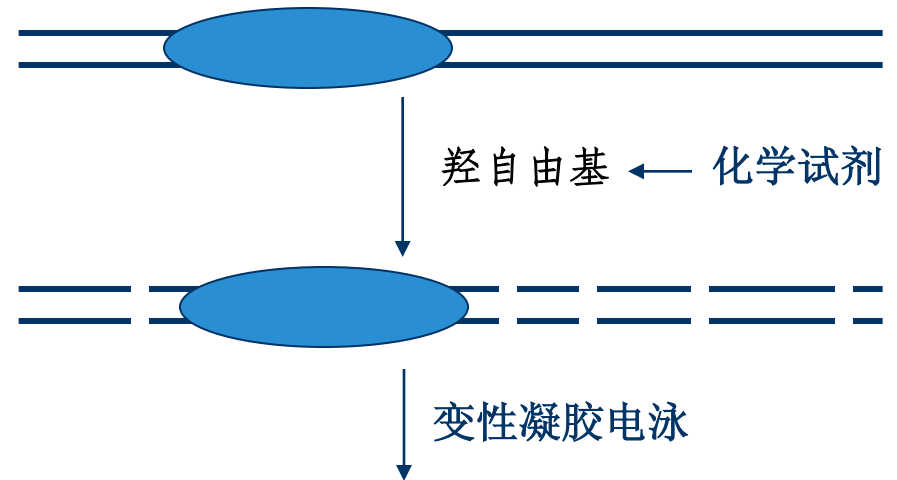
主要方法：

- 羟自由基足迹法
- 体内足迹法



羟自由基足迹法 (Hydroxyl radical footprinting)

- 利用化学试剂产生的羟自由基攻击DNA分子表面脱氧核糖骨架使DNA断裂
- 当DNA结合蛋白将脱氧核糖遮盖时，自由羟基无法攻击而使这个区域的DNA受到保护



电泳图谱上出现空白区的地方
就是结合蛋白质的DNA



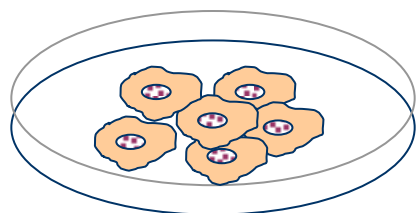
体内足迹法 (In vivo footprinting)

用化学试剂对活细胞进行体内处理，使DNA在细胞内受到化学修饰，然后裂解细胞，用化学法或酶法进行足迹实验。

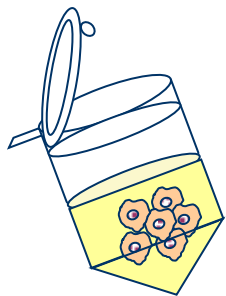
- 甲基化干扰实验 (Methylation interference assay): 利用化学试剂如硫酸二甲酯 (Dimethyl sulfate, DMS) 对活细胞DNA进行 (G) 甲基化修饰，从而干扰蛋白质与DNA的结合。
- 乙基化干扰实验 (Ethylation interference assay): 利用化学试剂对活细胞DNA进行乙基化修饰，从而干扰蛋白质与DNA的结合。



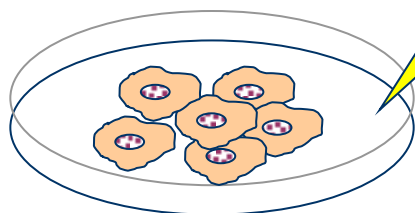
正常对照



提取DNA



化学修饰



化学试剂

提取DNA



切割DNA

DNase I
或化学试剂

变性凝胶电泳分析

体内足迹法

- 化学修饰对蛋白质与DNA的结合有干扰，因此，体内足迹实验也叫干扰实验
- 电泳图谱需与未修饰的DNA样品进行比较，在未修饰样品中出现空白区的位置是体内发生化学修饰的DNA区域



2. 电泳迁移率变动实验研究启动子

(Electrophoretic mobility shift assay, EMSA)

利用结合蛋白质的**DNA**片段在凝胶中迁移滞后的特点，通过电泳分离研究核酸-蛋白质互作的方法

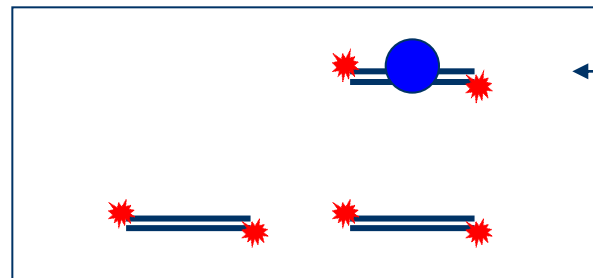
- 又称为凝胶阻滞实验 (**Gel retardation assay**)
- 只能说明**DNA**序列中含有核蛋白结合位点，但不能确定具体序列。



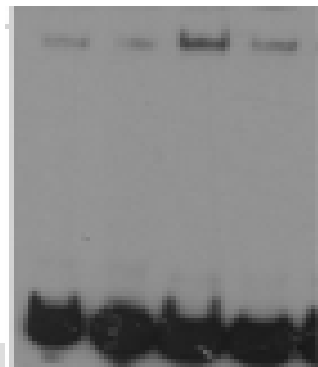
电泳迁移率变动实验 (EMSA)



↓ 凝胶电泳



↓ 显影





利用核酸-蛋白质互作方法研究启动子



3. 染色体免疫沉淀法鉴定启动子

(Chromatin immunoprecipitation, ChIP)

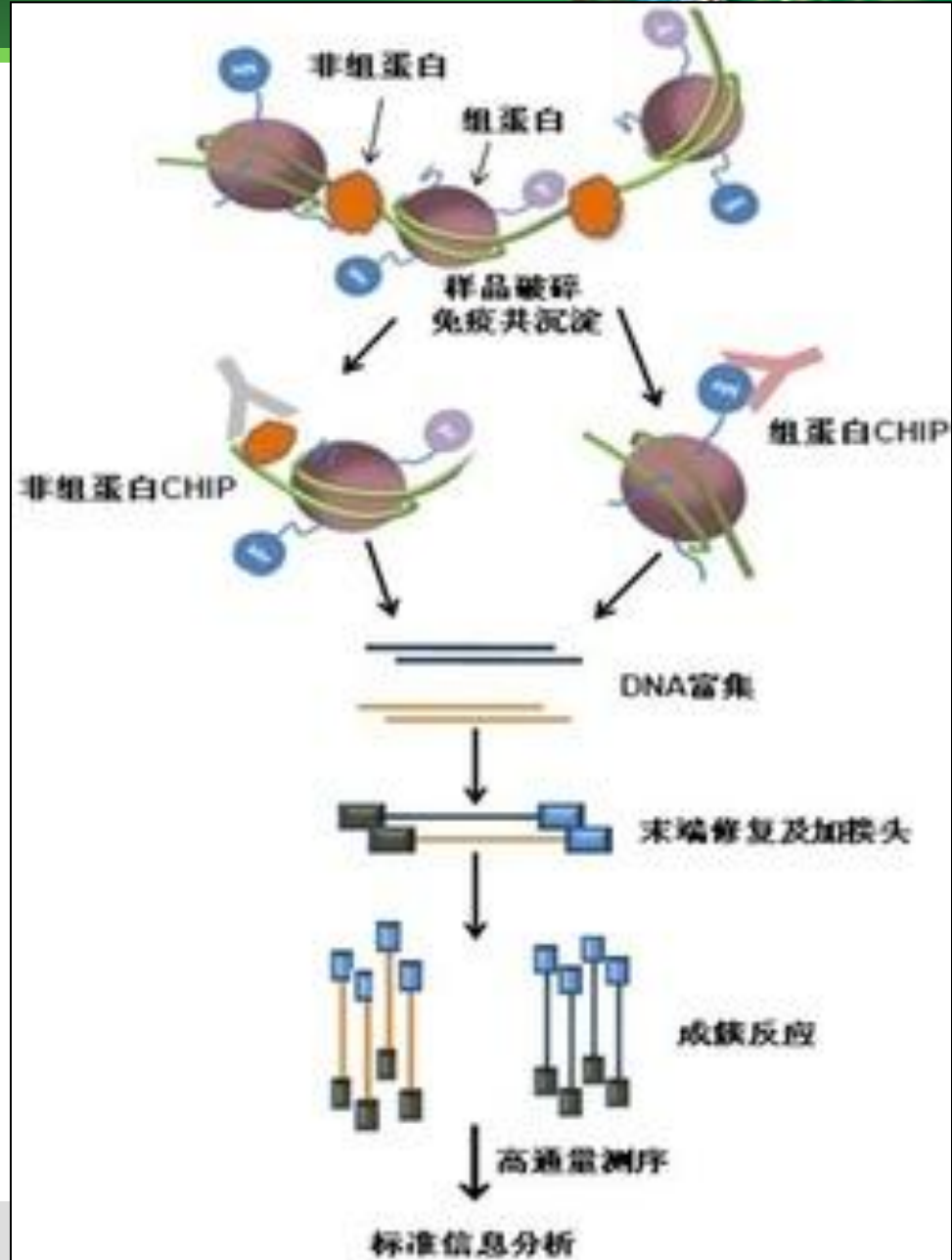
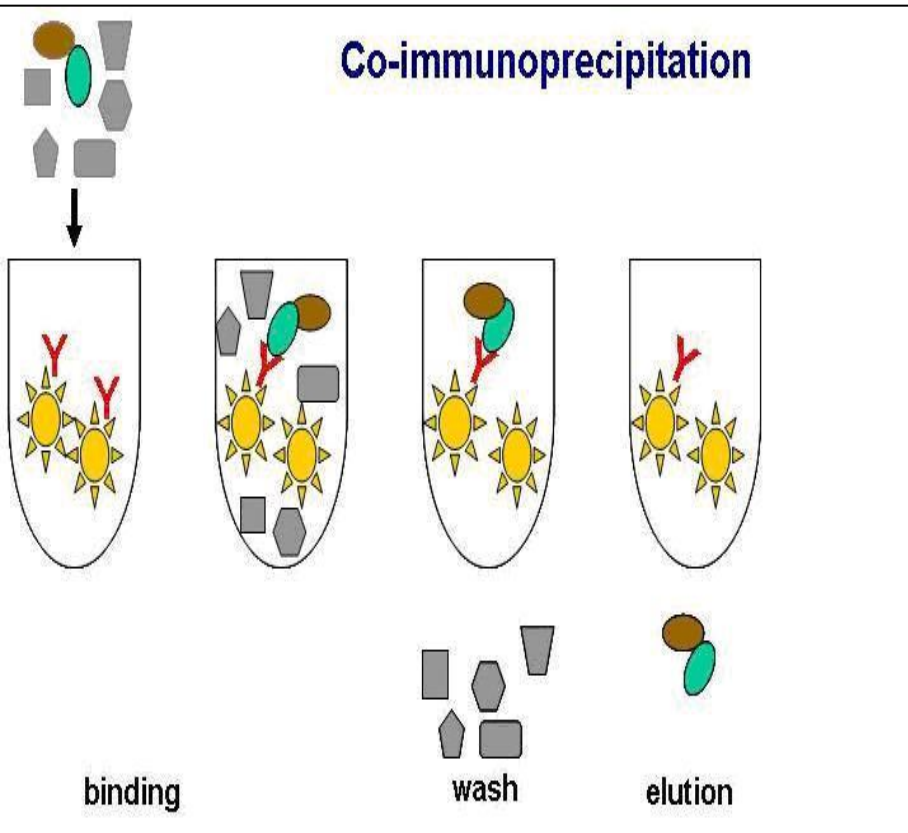
是在保持蛋白质与染色体DNA结合的同时，将染色体切割成小片段并沉淀下来。

基本原理：用甲醛处理活细胞，使蛋白质与DNA在细胞内发生交联，然后分离染色体并随机打断成染色质小片段（200-2000bp），然后以目标蛋白的特异性抗体免疫沉淀核蛋白-DNA复合物，再通过PCR扩增DNA片段并进行鉴定。

ChIP是目前既可真实、完整地反映结合基因组DNA上的调控蛋白因子，又能确定与特定蛋白质结合的基因组区域（包括启动子等顺式作用元件）的最好方法。

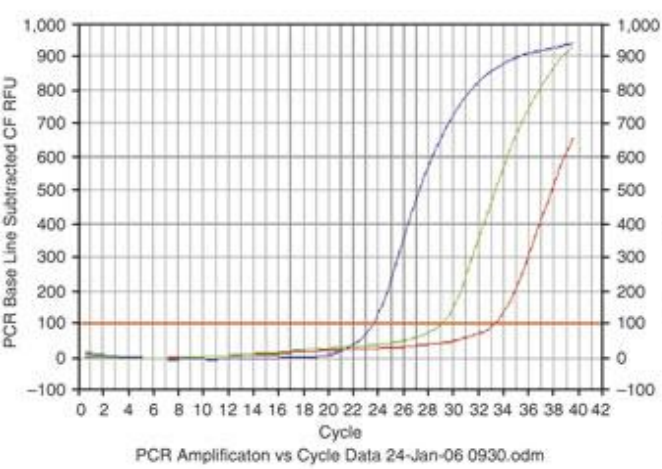
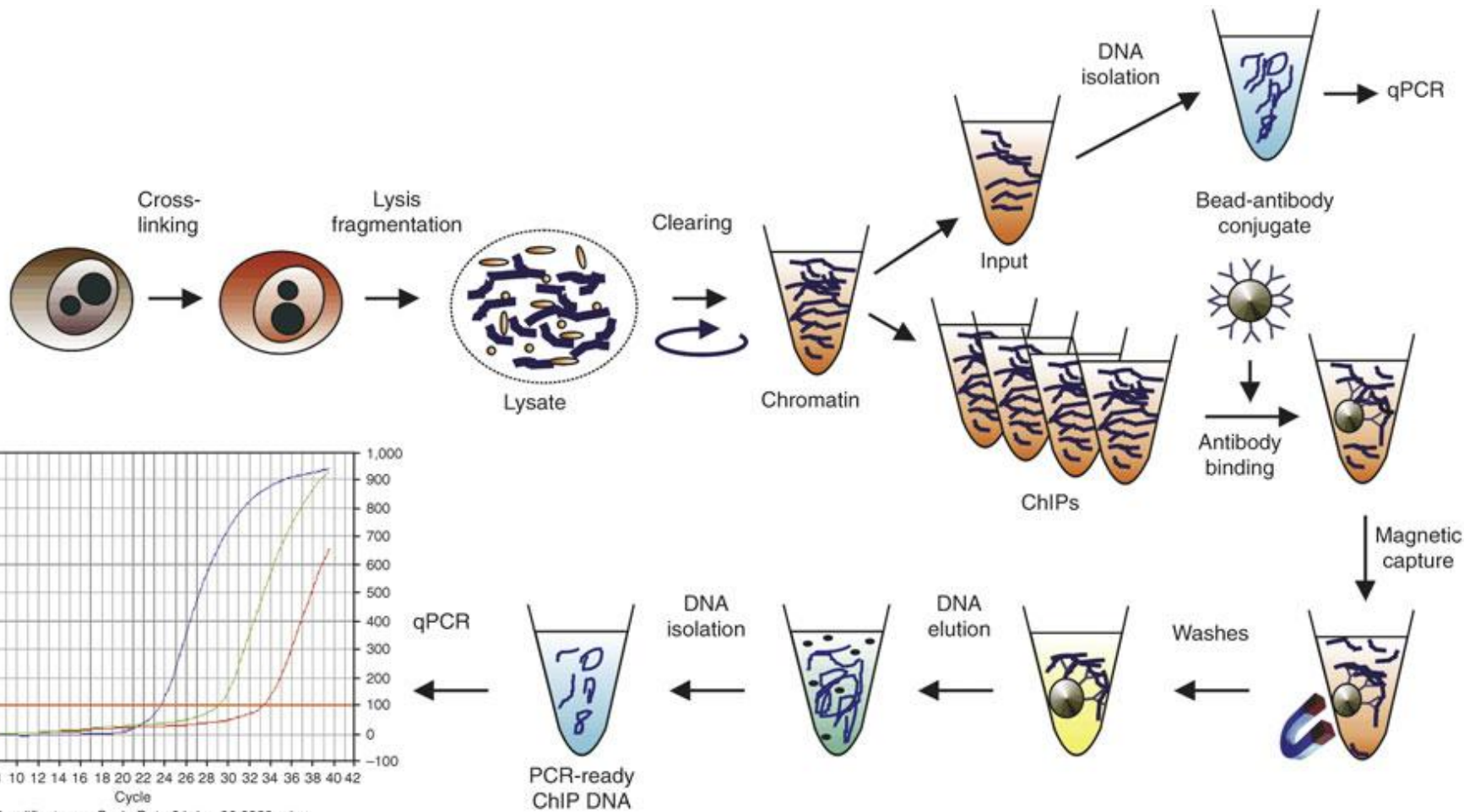


IP与ChIP的比较





染色体免疫沉淀法鉴定启动子



qPCR
PCR-ready
ChIP DNA



利用报告基因研究启动子活性



- 启动子通常是基因上游参与基因转录调控的**DNA**序列。由于启动子中的顺式作用元件在基因的特异性表达中发挥重要作用，因此，可以通过连接**报告基因**研究启动子的功能。

1. 报告基因 (Reporter gene)

- 在细胞培养条件下或动植物体内作为筛选标志的易检测信号，通过基因操作将发光蛋白或酶的编码基因附加到一个感兴趣基因上或插入基因调控序列下游，从而监测感兴趣基因的表达或分析基因调控序列的活性。



•常用的报告基因

•荧光蛋白编码基因:

绿色荧光蛋白 (GFP)

•在蓝色光源照射下发绿光

红色荧光蛋白 (dsRed)

•蛋白酶:

荧光素酶 (luciferase)

•能催化荧光素 (luciferin) 发生氧化反应发光

β -半乳糖苷酶

•能使细菌在X-gal存在条件下变成蓝色



报告基因的应用

- 监测基因的转染效率

报告基因与目的基因分别插入各自启动子下游，实现报告基因的组成性表达模式

- 监控目的基因的表达

报告基因与目的基因融合共同受控于一个启动子，报告基因的表达即代表目的基因的表达

- 研究启动子的活性

报告基因插入被研究启动子下游，通过观察报告基因的表达情况推测启动子活性



启动子捕获技术 (promoter trapping)

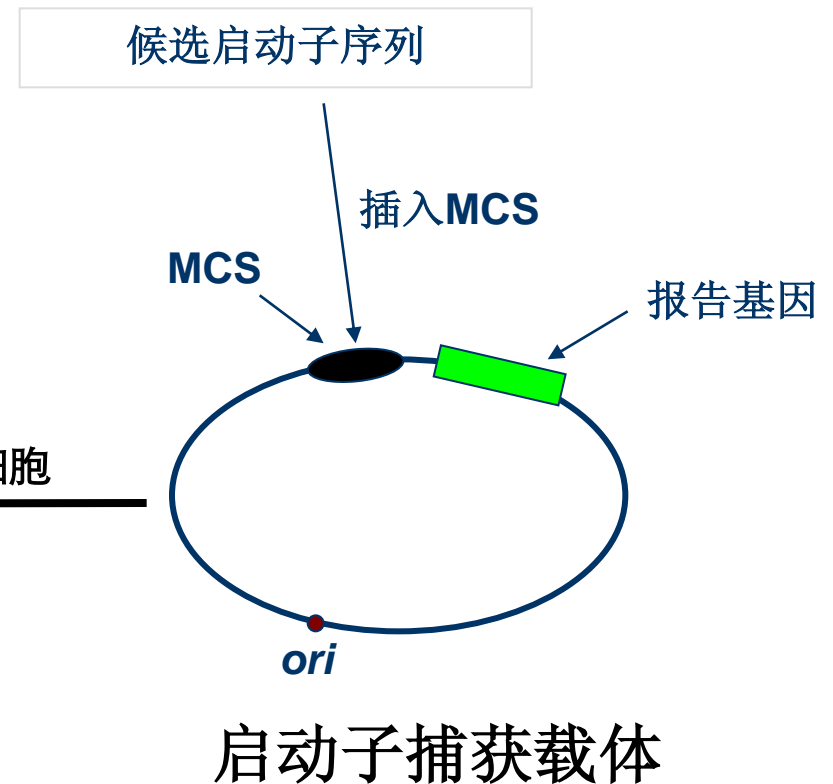
是一种研究启动子活性的筛选方法

• 基本流程：

① 构建启动子捕获载体

② 观察报告基因的表达

观察报告基因的表达 ← 转染细胞





三) 基因转录起始点的序列分析方法

转录起始点 (TSS)

- TSS位于基因编码序列的5'端
- 多肽链是以mRNA为模板经翻译合成的

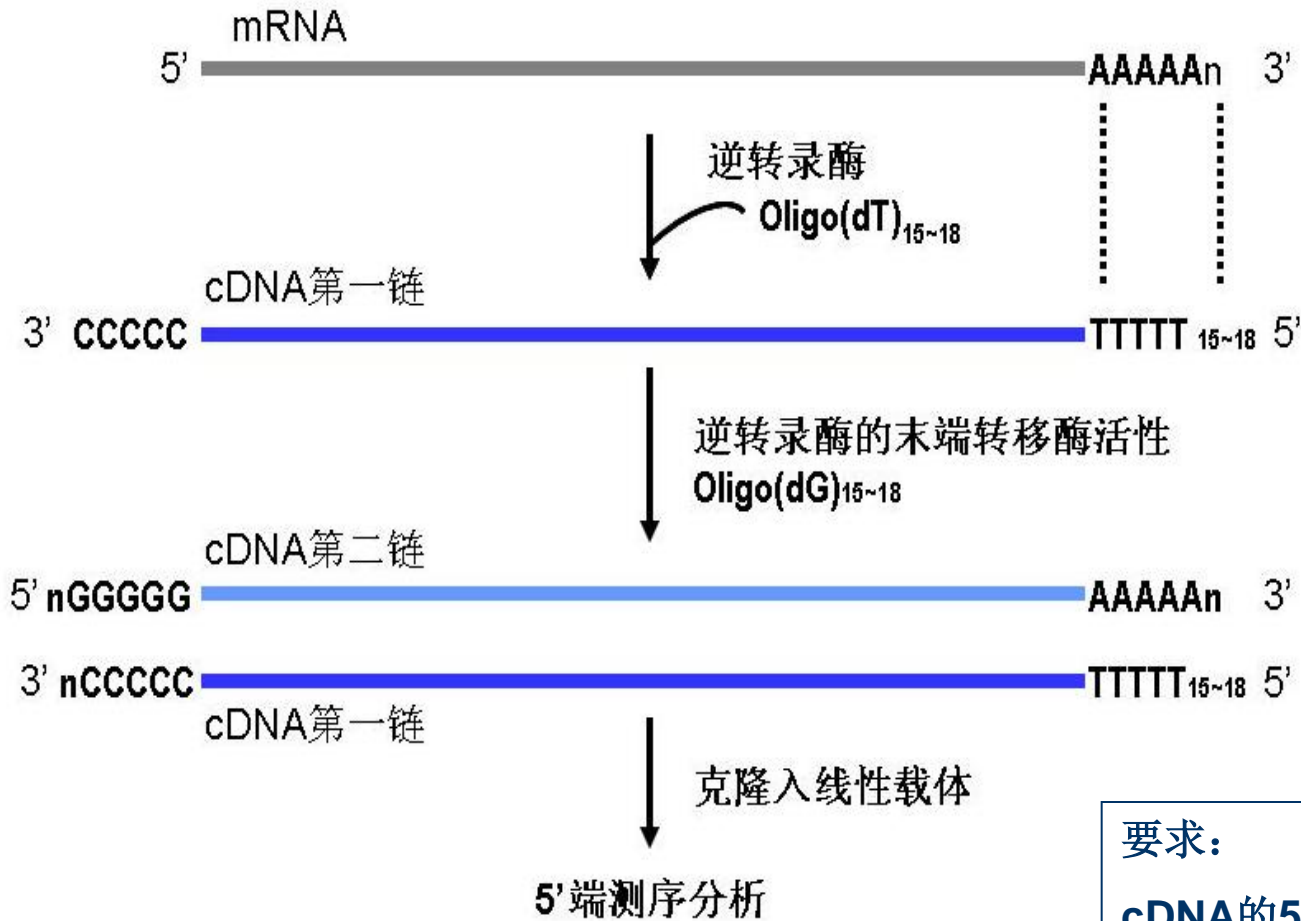
所以：分析鉴定TSS的方法都是以cDNA为切入点

分析TSS方法

1. cDNA克隆测序
2. cDNA末端快速扩增技术(RACE)
3. 连续分析基因TSS
4. 利用数据库搜索TSS



cDNA克隆测序



要求:
cDNA的5'端完整无缺

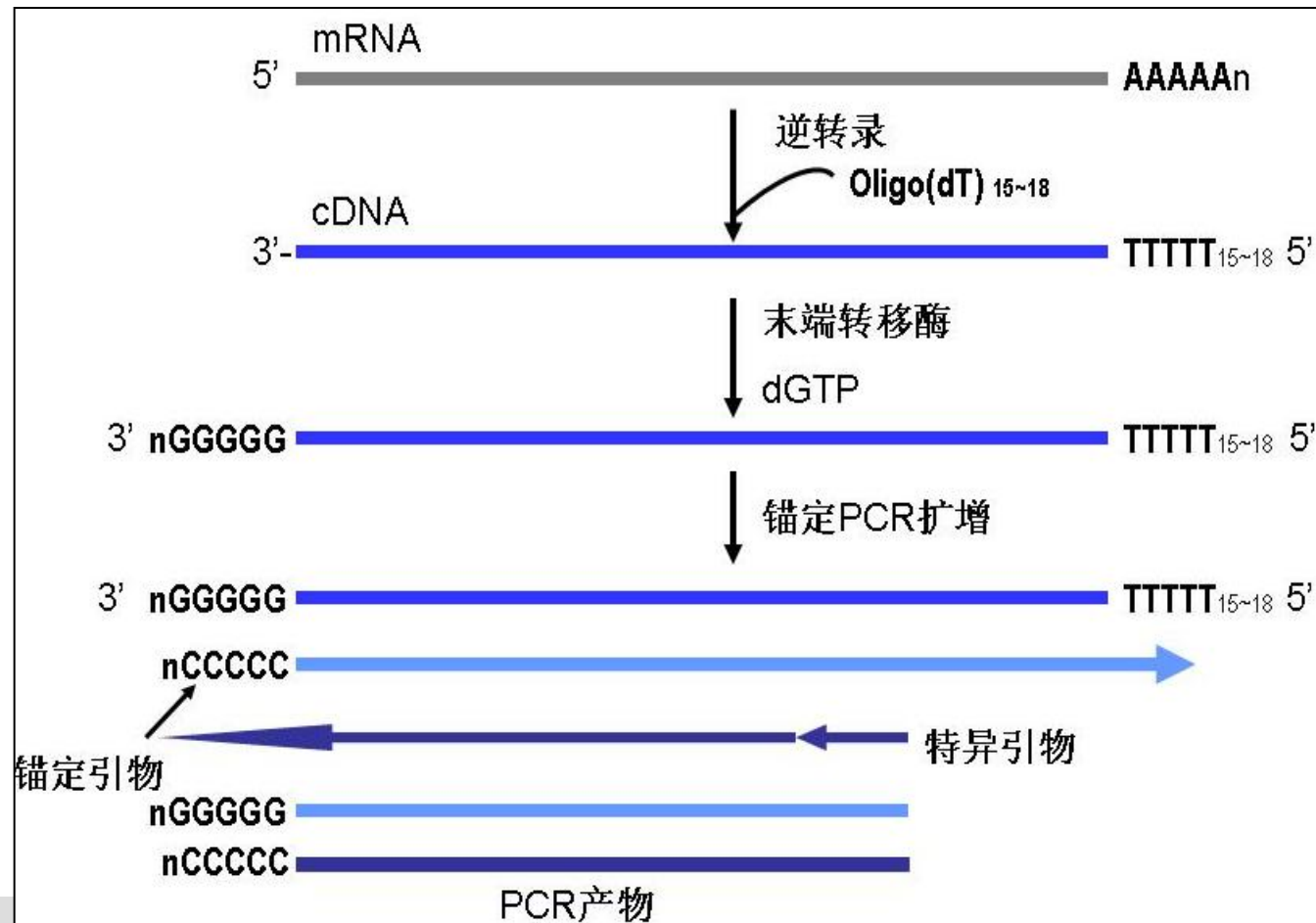


cDNA末端快速扩增技术(RACE)



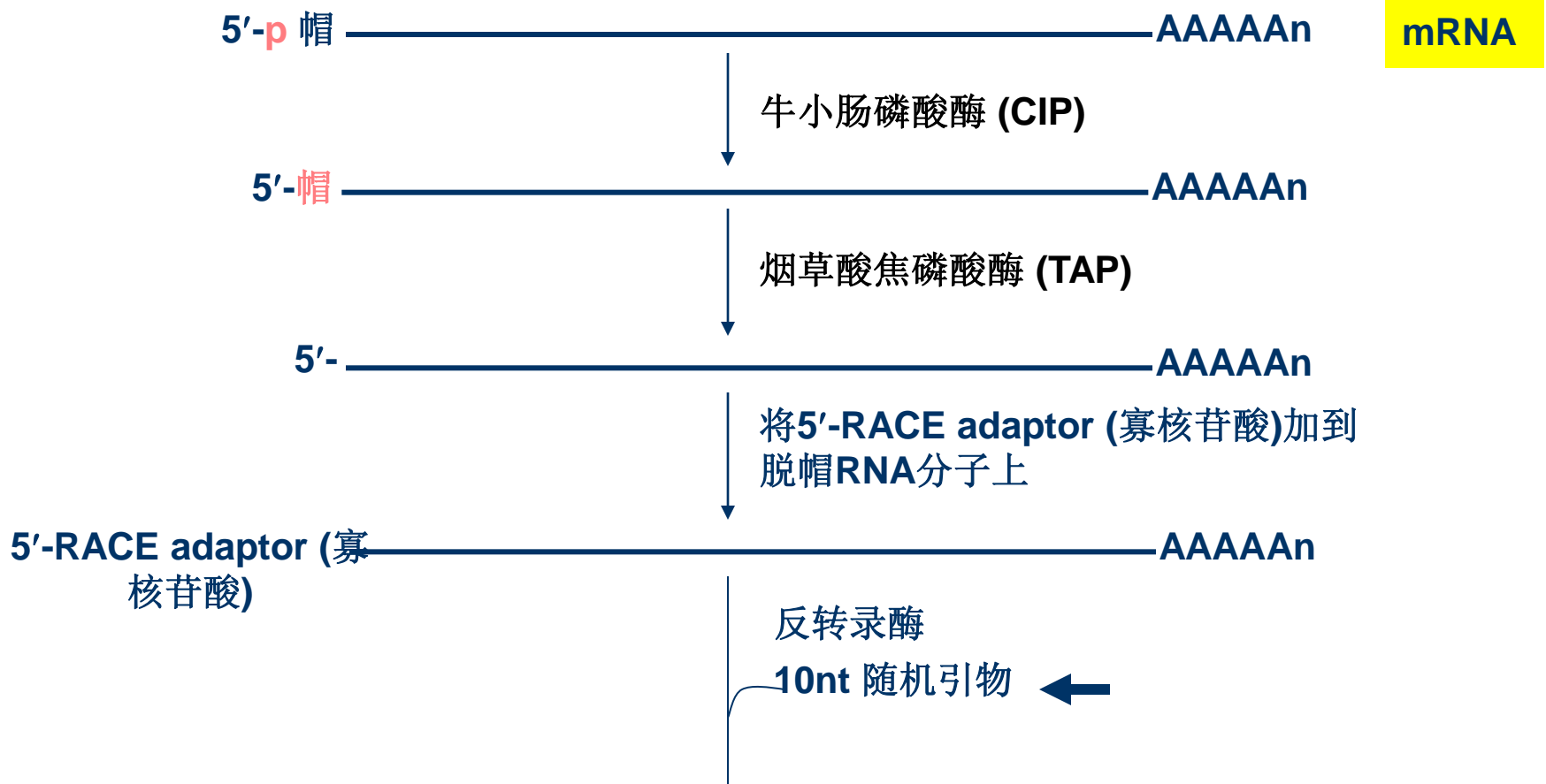
rapid amplification of cDNA ends

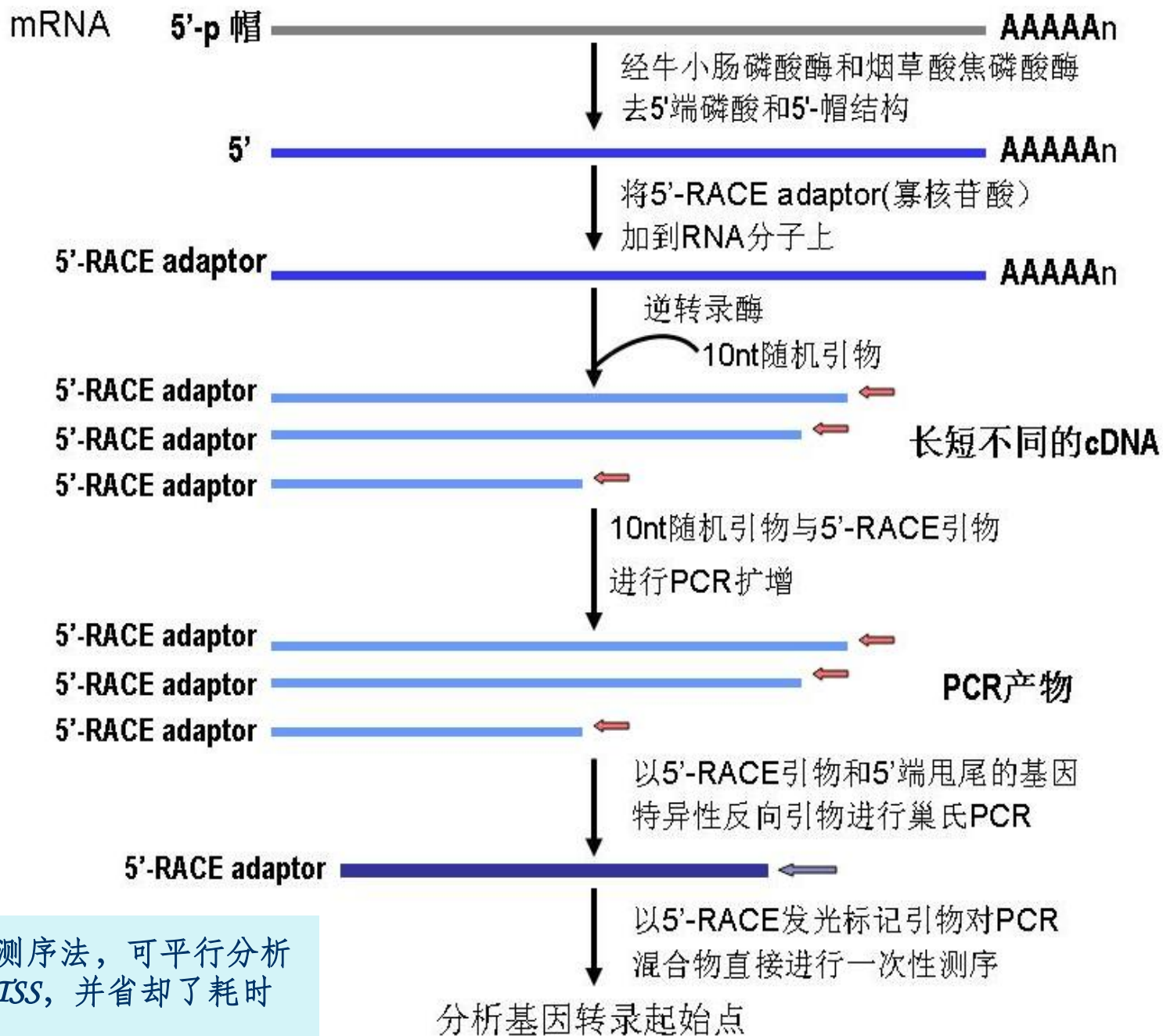
• 传统的RACE





Deep-RACE: 用寡核苷酸替代mRNA的5'端帽结构, 采用发光标记引物进行巢式PCR, 实现高通量鉴定转录起始点





借助于最新的发光测序法，可平行分析多达几百个基因的TSS，并省却了耗时的克隆步骤。

分析基因转录起始点



连续分析基因转录起始点



- 在**RACE**基础上，通过在转录本5'端引入一个特殊的限制性核酸内切酶位点，实现了基因5'端短片段串联连接产物一次测序分析多个**TSS**的目的
- 主要有两种方法：
 - **5'端连续分析基因表达**（5'-end serial analysis of gene expression, **5' SAGE**）
 - **帽分析基因表达**（cap analysis gene expression, **CAGE**）

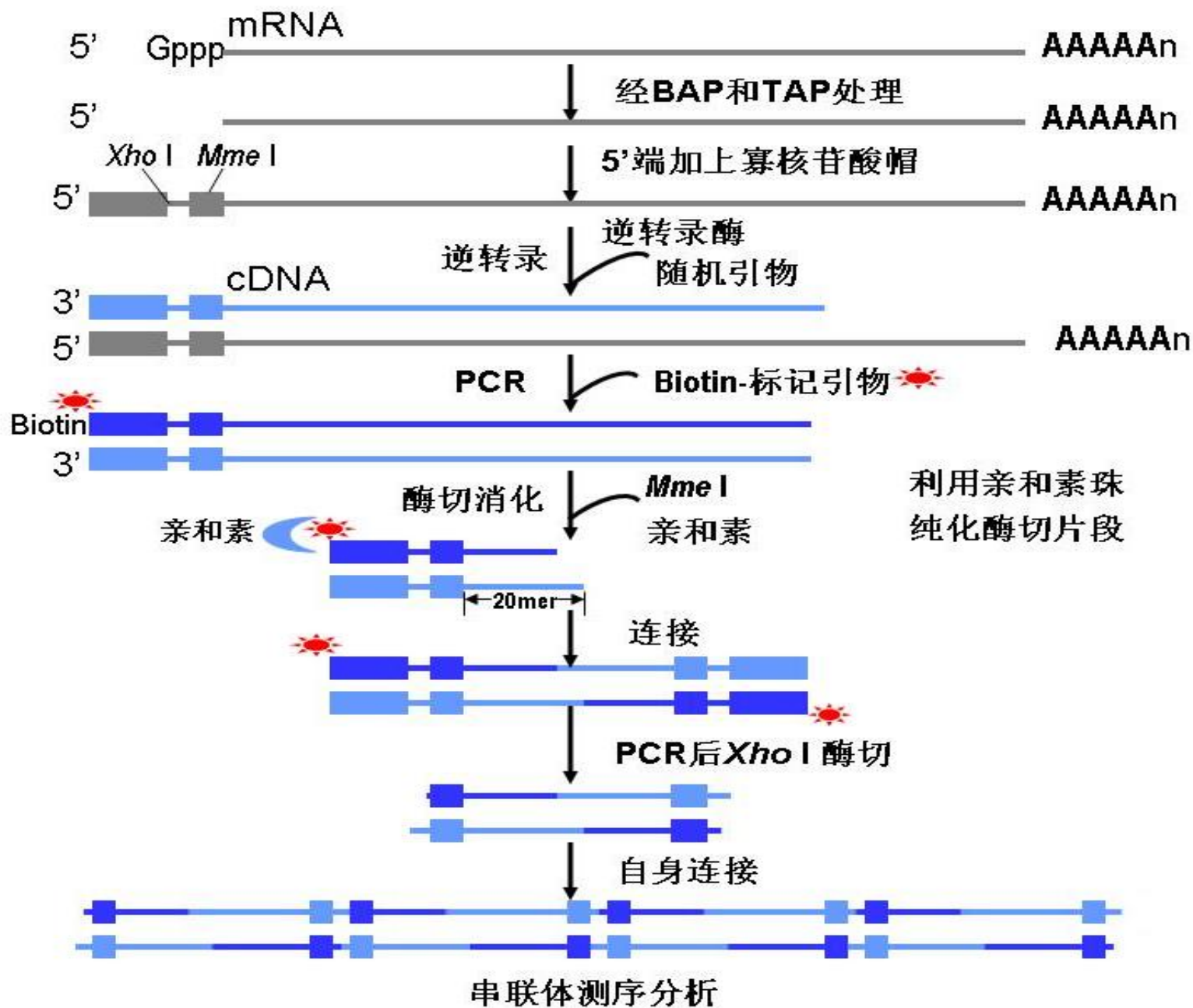
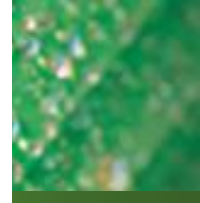


1. 5' SAGE (5'端连续分析基因表达)

- 即在PCR过程中将*MmeI*酶切位点引入cDNA的5'端，通过酶切和连接获得不同短片段重复序列，并对重复序列进行测序获得大量片段序列信息
- 不同序列的短片段代表不同基因的转录起始点 (TSS)

MmeI:

- 是一种特殊的限制性核酸内切酶
- 识别的序列不是回文结构，而是不对称的DNA序列5' - TCCRAC-3' (R代表G或A)
- 在识别位点下游18~20碱基处切开双链DNA



5' SAGE的基本流程

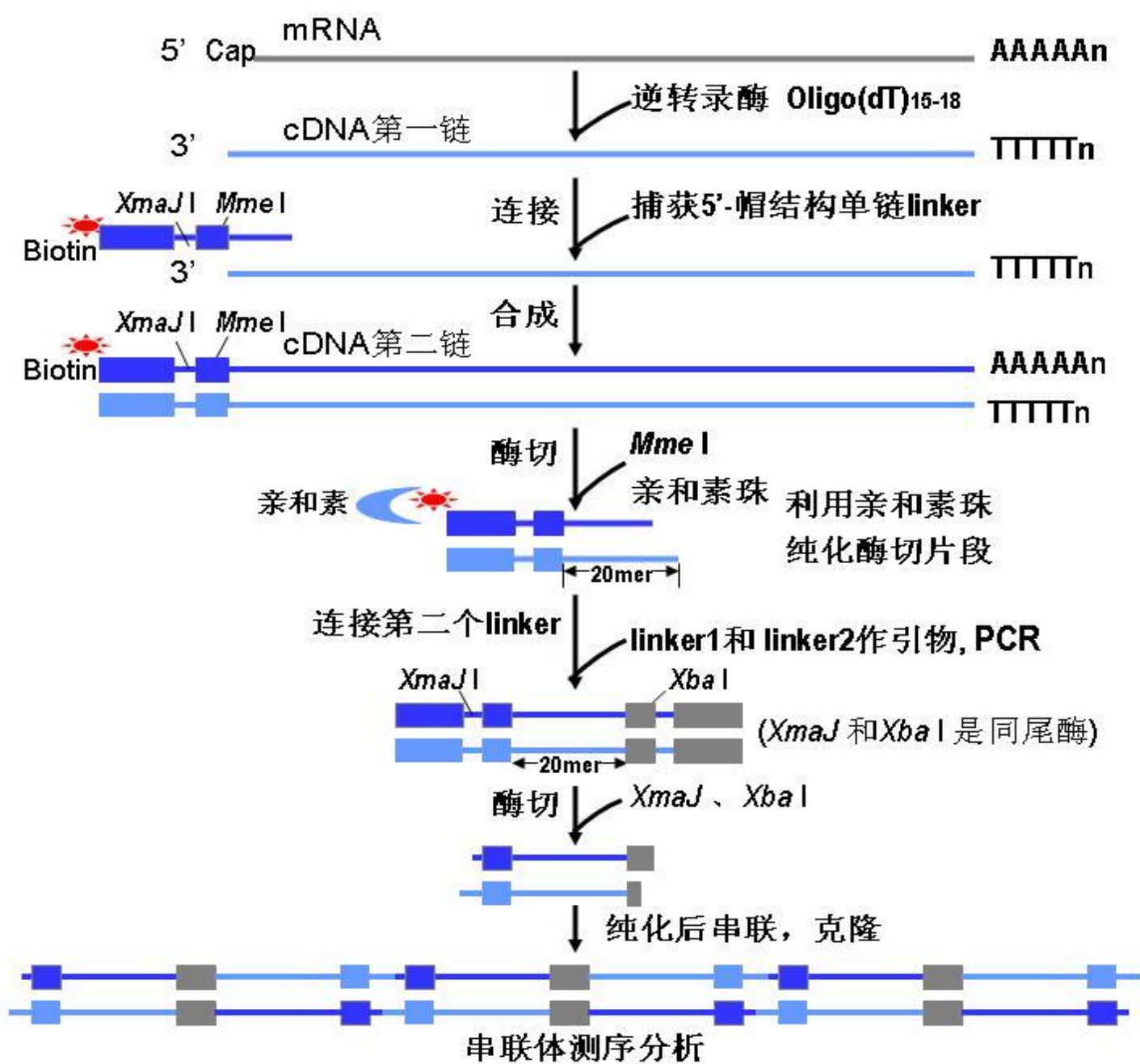


2. CAGE (帽分析基因表达)

CAGE流程相似于5'SAGE

不同之处:

- CAGE不需要在RNA上加接头，而是用oligo(dT)引物先进行第一链cDNA的合成
- 然后通过捕获帽结构，将含有*MmeI*和另一内切酶位点如*XmaI*的linker加到单链全长cDNA的3'末端



CAGE的基本流程



三、基因编码区结构分析方法



一) 基因编码序列的结构特征



编码序列 (coding sequence): 指能体现在蛋白质氨基酸序列中的基因信息。cDNA文库为基础的基因结构研究。

基因编码序列的特征性序列:

- 蛋白质翻译的开放阅读框架 (open reading frame, **ORF**)、位于起始密码子和终止密码子之间
- 真核基因的外显子和内含子之间有mRNA选择性剪接特殊序列 (—**GU**-----**AG**—)
- 基因外显子包括能被翻译成蛋白质的编码区, 5'-非翻译区 (5'UTR), 3'-非翻译区 (3'UTR)



分析一段DNA序列中是否存在ORF:

可对dsDNA序列的6种ORF进行分析，每一条链分析三种ORF

例如:

1) **5'-UCU AAA AUG GGU GAC-3'**

(其中AUG是起始密码子)

2) **5'-U CUA AAA UGG GUG AC-3'**

3) **5'-UC UAA AAU GGG UGA C-3'**

(其中UAA是终止密码子)

只有真正的ORF可以不遇到终止密码子



mRNA的选择性剪接 (alternative splicing):

- 基因外显子转录产物RNA进行切割再连接的加工过程
- 经剪接所产生的mRNA可以翻译成不同的蛋白质，从而导致一个基因可以编码一个以上蛋白质

真核基因的内含子在与外显子交界区域有共通序列 (consensus sequences):

- 内含子的5'端有GU序列，3'端有AG序列



基因外显子可以被分成三部分

- 能够被翻译成蛋白质的编码区

- 5'-非翻译区 (5'UTR)

- 3'-非翻译区 (3'UTR)

•有作为蛋白质翻译起始重要元件的Kozak序列：
由起始密码子AUG及其周围序列组成

•3'UTR位于终止密码子下游，含有poly A尾的加尾信号AATAAA序列



二) 基因编码序列的结构分析



与mRNA互补的cDNA成为研究编码序列的主要切入点

主要方法：

- cDNA文库的编码序列筛选
 - RNA剪接分析编码序列
 - 基因的拷贝数分析
- 对各种方法所获得的cDNA片段的序列在基因数据库中进行同源性比对，通过染色体定位分析、内含子/外显子分析、ORF分析及表达谱分析等



RNA剪接分析基因编码区的选择性剪接序列和位点

- **DNA芯片分析**

常用外显子微阵列或外显子/外显子交接的阵列，以cDNA为探针筛选RNA剪接体，以确定基因的编码序列。

- **交联免疫沉淀 (CLIP)**

先用紫外线将蛋白质与RNA交联，然后用蛋白特异性抗体将蛋白质-RNA复合物沉淀分离，再分析蛋白质结合的RNA序列，确定RNA的剪接位点，以推导基因编码序列和内含子交界区序列。

- **体外报告基因测定法**



基因的拷贝数分析

1. Southern blot分析基因的拷贝数变化

2. 差异显示PCR技术也可分析基因的拷贝数

利用随机引物对不同组织来源的DNA进行扩增，电泳图谱可能出现具有差异的片段，从而推断基因的拷贝数。

内参照相对定量PCR: 利用管家基因作为PCR体系中内参照。

3. 代表性寡核苷酸阵列分析(ROMA)是分析基因拷贝数的新方法 (representational oligonucleotide microarray analysis)

设计能与代表性片段杂交的70 mer的Oligo探针（覆盖整个基因组的重复序列），将探针制成芯片，与代表性差异分析筛选出来的PCR产物进行杂交，计算机分析结果。



小 结



- 依据基因的结构特征，基因结构的分析可从实验检测和数据库搜索比对两方面入手。
- 基因结构特点成为基因组范围内高通量扫描基因的重要靶标，基因的转录起始点、启动子以及编码序列是基因的重要结构特征
- 基因结构分析的切入点已经从一个基因的克隆测序，发展到如今在基因组范围的高通量筛选，因此，研究策略也发生了变化，基因数据库占据了重要地位。

See You !

