

第六章

分子生物学基本研究法（下）

基因功能研究技术

基因表达研究技术

1. 转录组学研究
2. 分子杂交技术
3. cDNA sequencing
4. RNAi技术
5. 基因定点突变技术
6. 基因敲除技术

转录组学研究

转录组 (transcriptome), 广义上指在某一特定生理条件或环境下, 一个细胞、组织或者生物体中所有RNA的总和, 包括信使RNA (mRNA)、核糖体RNA (rRNA)、转运RNA (tRNA) 及非编码RNA (non-coding RNA或siRNA)。

现常指细胞中转录出来的所有mRNA的总和。

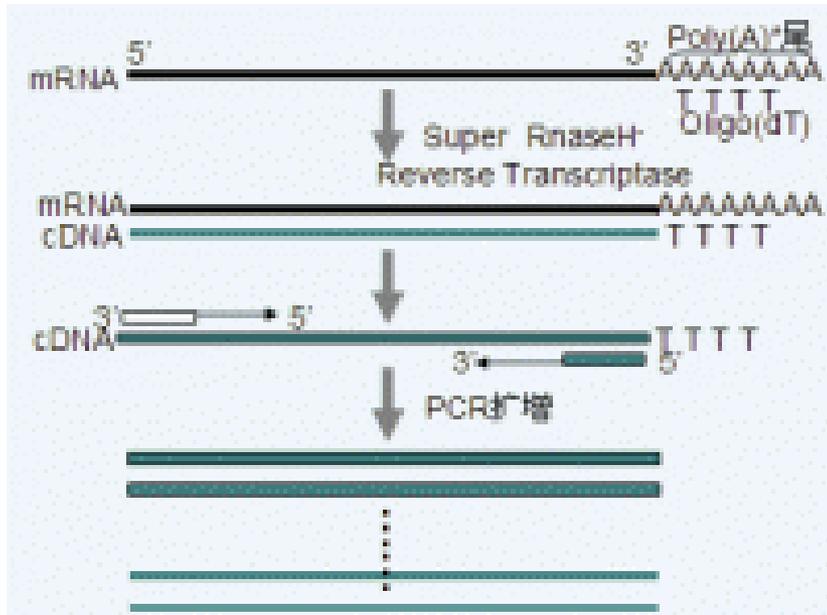
基因组 — 转录组 — 蛋白质组（**genome — transcriptome — proteome**）是中心法则在组学框架下的主要表现形式。

通过特定生理条件下细胞内的**mRNA**丰度来描述基因表达水平并外推到最终蛋白质产物的丰度是目前基因表达研究的基本思路。

转录组研究方法

- **RT-PCR**
- **Hybridization**方法
 - (1) Northern blots,
 - (2) 基因芯片技术
- **cDNA sequencing**
 - (1) 标签测表达序列序技术
(Expressed Sequence Tag, EST)
 - (2) 基因表达系列分析技术
(serial analysis of expression, SAGE)
 - (3) 高通量测序技术

RT-PCR

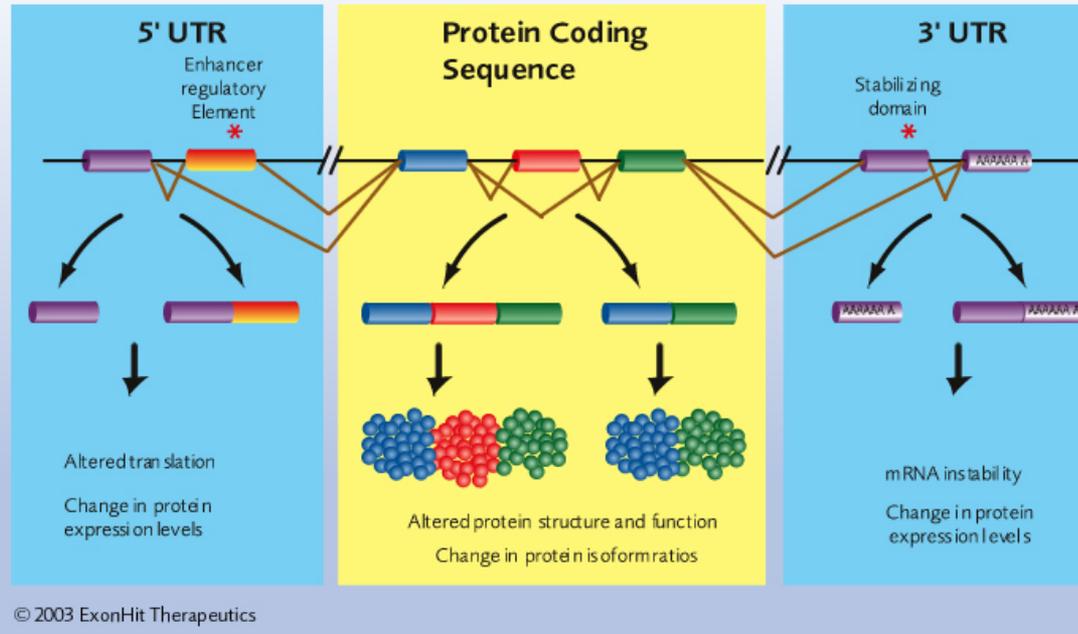


- 在mRNA反转录之后进行的PCR。将以RNA为模板的cDNA合成同PCR结合在一起，提供了一种分析基因表达的快速灵敏的方法。
- RT-PCR用于对表达信息进行检测或定量。另外，这项技术还可以用来检测基因表达差异。

RT-PCR的模板可以为总RNA或poly(A)+选择性RNA。逆转录反应可以使用逆转录酶，以随机引物、oligo(dT)或基因特异性的引物（GSP）起始。RT-PCR可以一步法或两步法的形式进行。

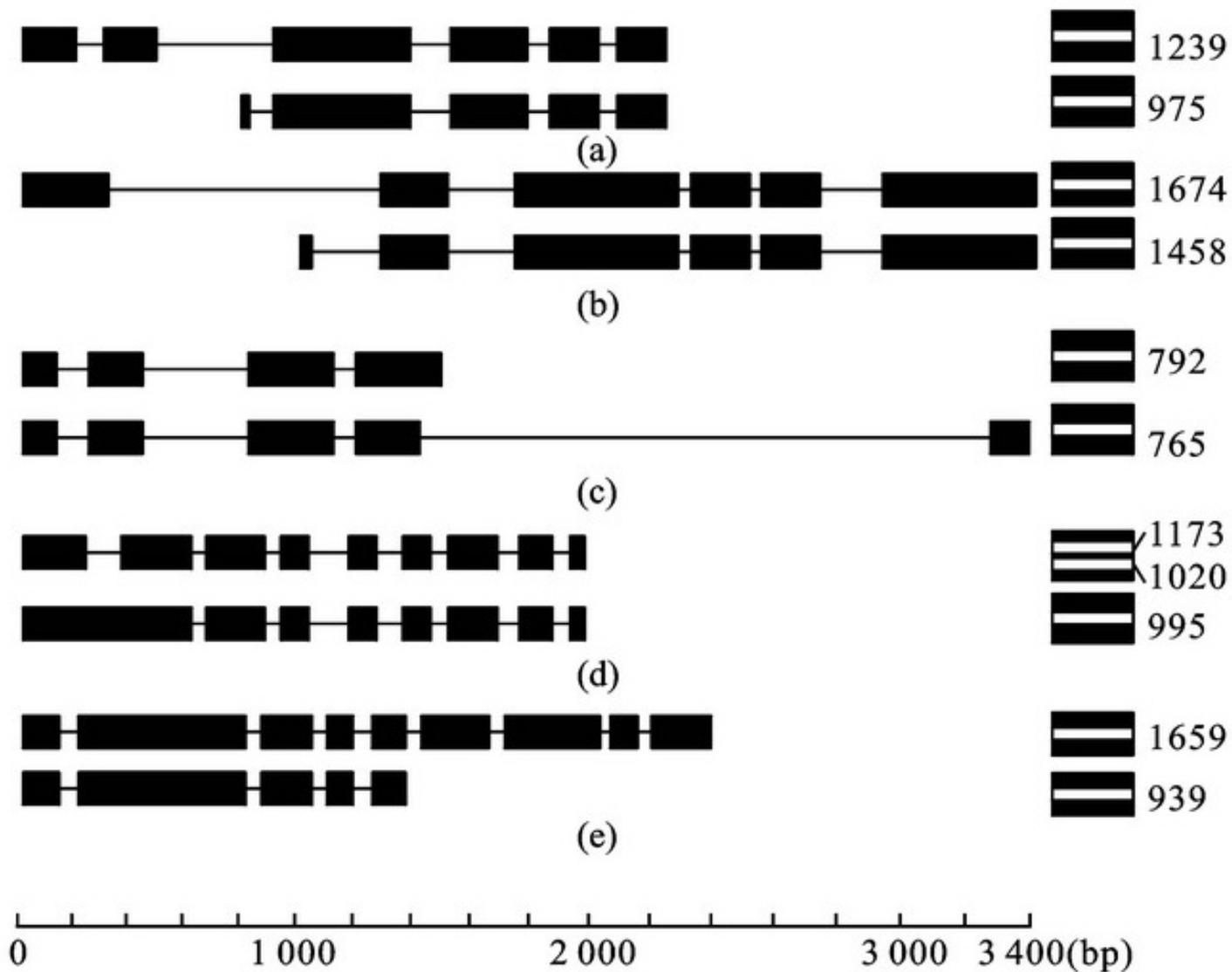
RNA的选择性剪接技术

The impact of alternative RNA splicing

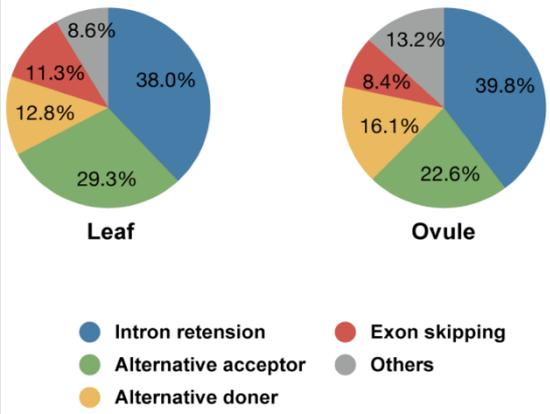


- **RNA**的选择性剪接是指用不同的剪接方式从一个**mRNA**前体产生不同的**mRNA**剪接异构体的过程。
- 包括：平衡剪接、**5'**选择性剪接、**3'**选择性剪接、外显子遗漏型剪接及相互排斥性剪接。
- 通过**RT-PCR**可以检测基因的选择性剪接。

RT-PCR检测5个拟南芥转录调控因子基因的选择性剪切

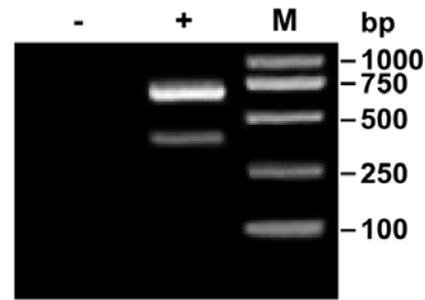
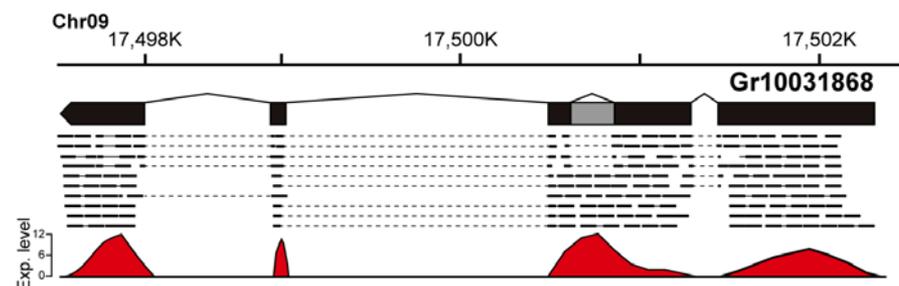


G. raimondii 基因组中四种主要的选择性剪切事件

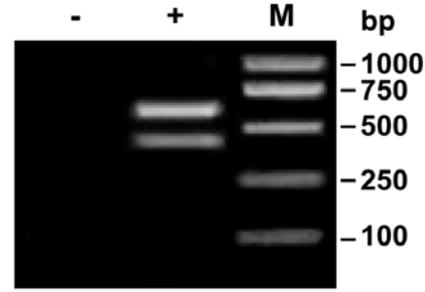
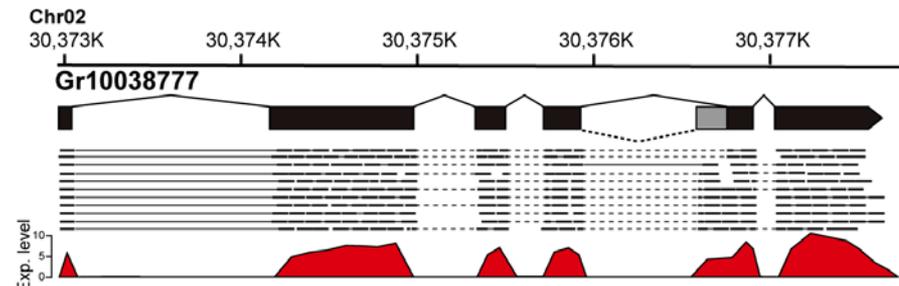


IR is the most frequent AS event in *G. raimondii*.

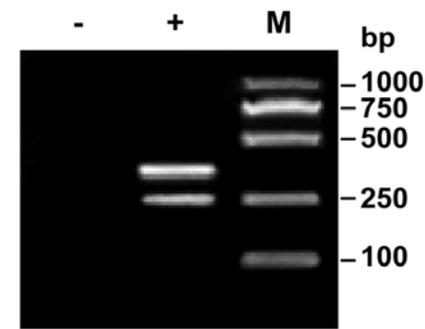
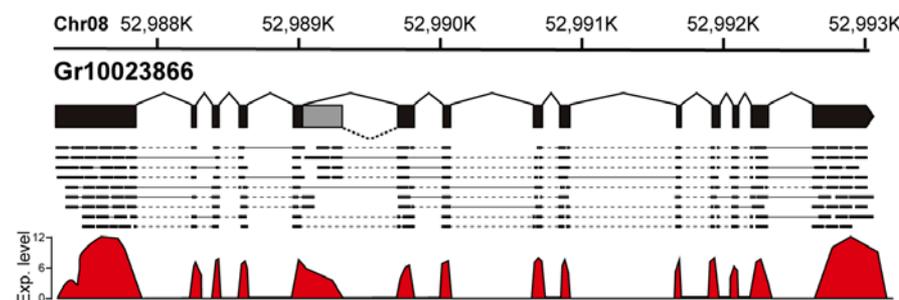
A Intron retention



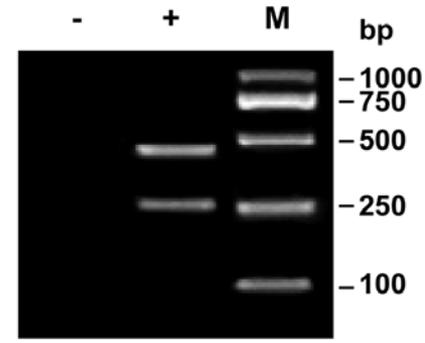
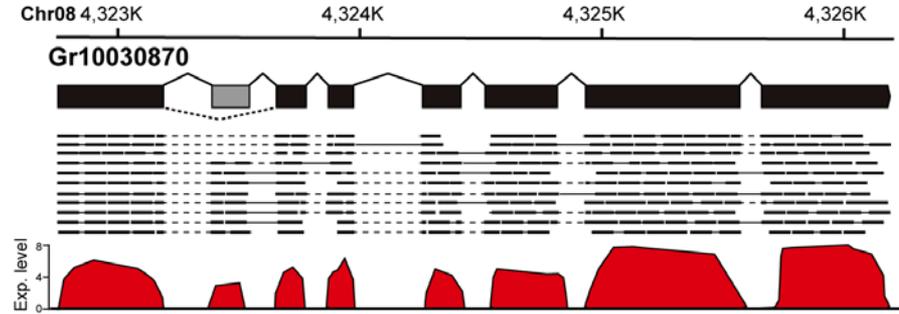
B Alternative acceptor



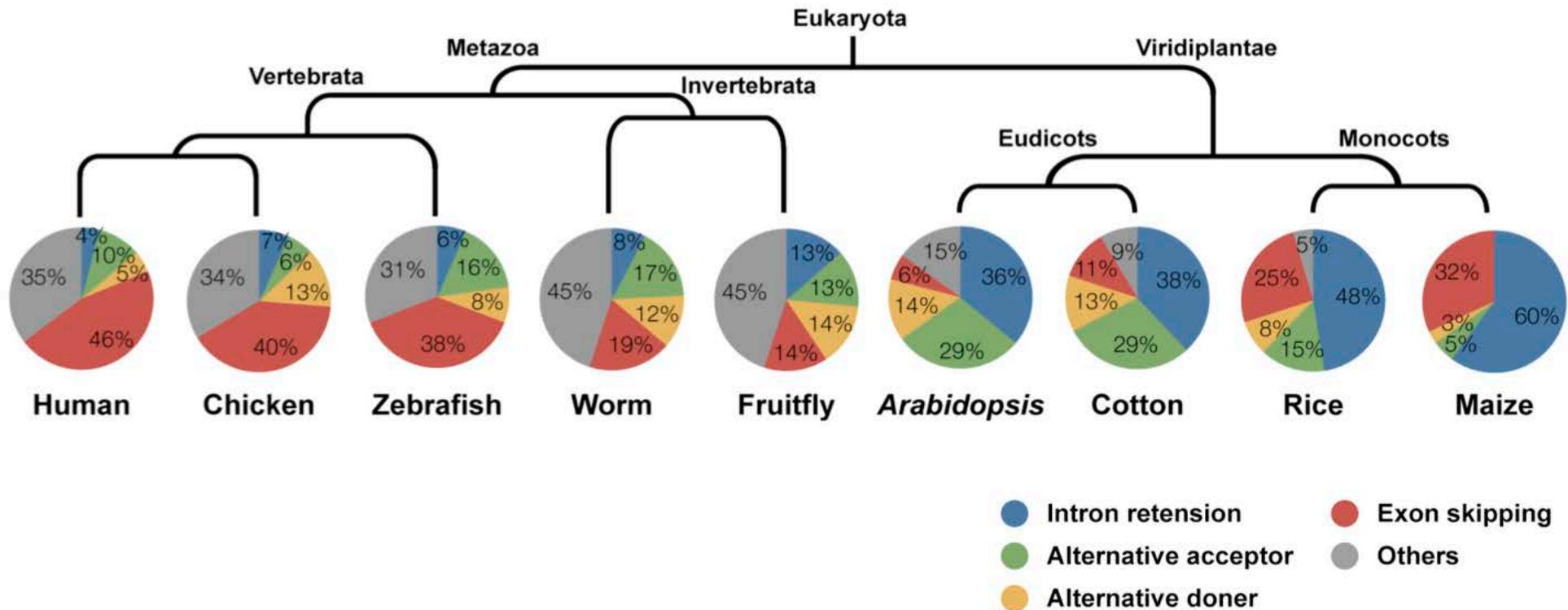
C Alternative donor



D Exon skipping



AS events profiling across nine species



核酸的分子杂交

放射性同位素技术

同位素：原子序数相同而质量不同的元素。

- **放射性同位素** 同位素可分为稳定同位素和不稳定同位素，后者又称为放射性同位素。不稳定同位素原子核结构不稳定，易自发产生衰变，产生 α -射线、 β -射线和 γ -射线等，同时从不稳定同位素变成稳定同位素。
- 在分子生物学研究中，得到应用的是放射性同位素。

放射性同位素

➤ 放射性的衰变类型

α -射线：带正电荷的高速粒子流

β -射线：带负电荷的粒子流

γ -射线：光子流

➤ 分子生物学研究中常用的放射性同位素及其性质

元素名称	^3H	^{14}C	^{32}P	^{35}S	^{60}Co
半衰期	12.1年	5700年	14.3天	87.1天	5.8年

➤ 放射性强度的探测

(1) 盖革计数器 (Geiger counter tuber)，闪烁计数器。

(2) 放射自显影：X-光乳胶片覆盖于样品，在黑暗中， -70°C ，曝光。

➤ 放射性分子的制备：核物理，化学，生化反应，生化分离技术

分子杂交类型

根据鉴定对象不同，可将分子杂交法分为3种类型：

类型	Southern杂交 Southern印迹	Northern杂交 Northern印迹	Western杂交 Western印迹
检测对象	DNA	RNA	蛋白质
探针	DNA或RNA	DNA或RNA	蛋白质

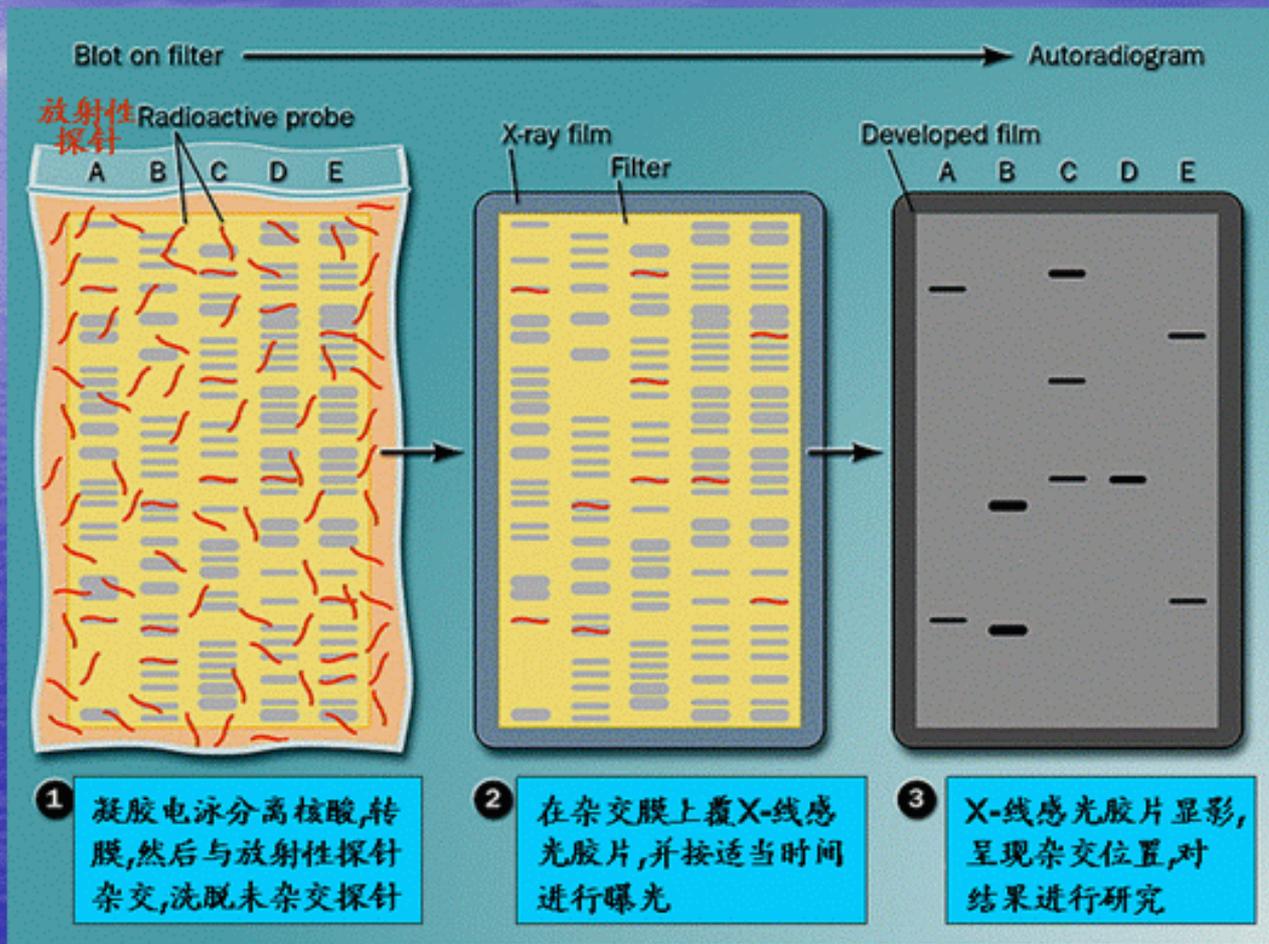
Southern杂交：将DNA分子从电泳凝胶转移至杂交滤膜上，然后与核酸探针进行杂交。由E. M. Southern于1975年创立。

Northern杂交：将RNA分子从电泳凝胶转移至杂交滤膜上，然后与核酸探针进行杂交。

Western杂交：将蛋白质从电泳凝胶中转移到杂交滤膜上，然后同放射性同位素¹²⁵I标记的特定蛋白质进行反应。

核酸的变性、复性与杂交

核酸的杂交

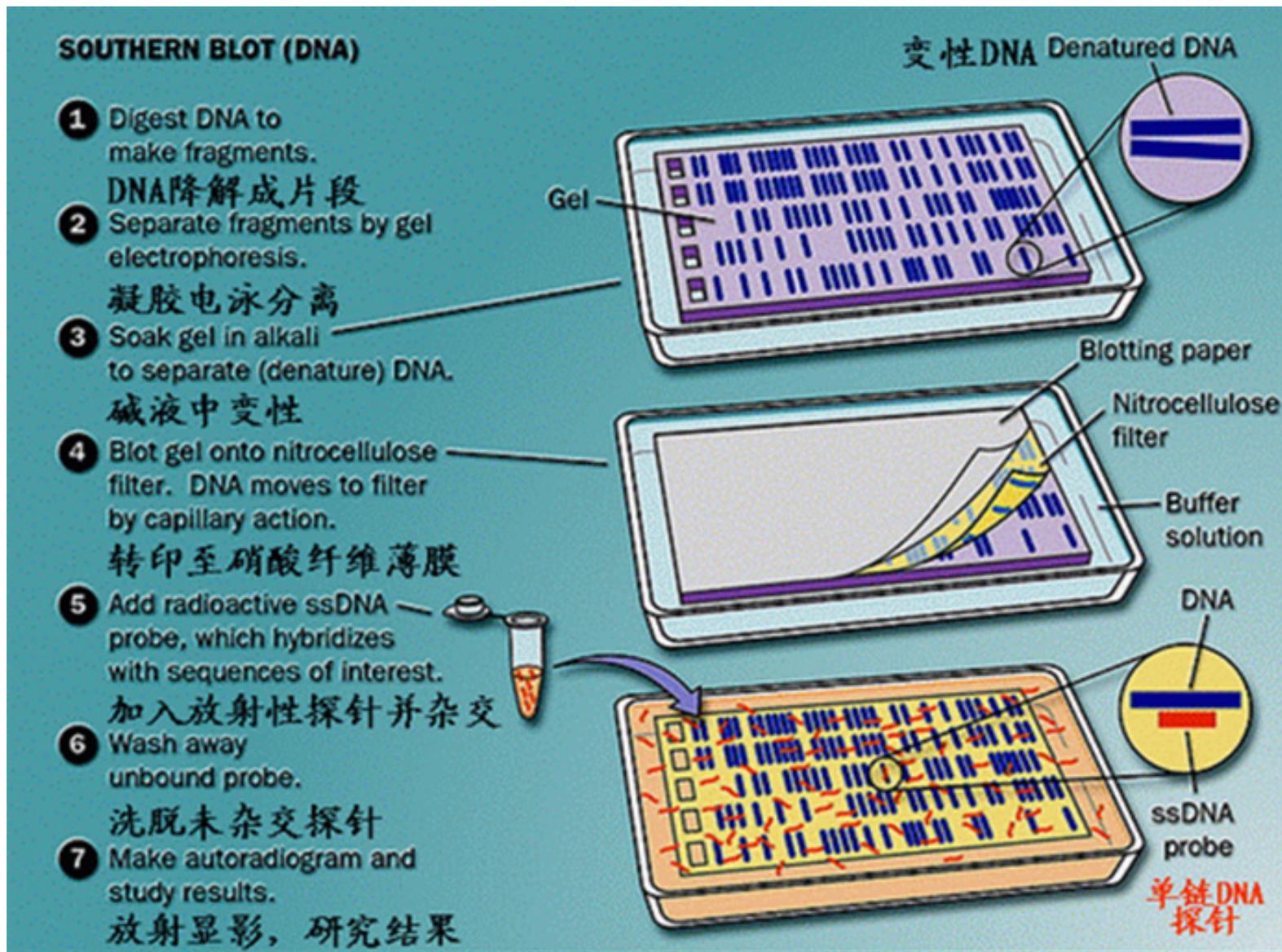


复性（退火）：在一定条件下变性的两条单链重新结合为双链的过程。

杂交：来源不同的两条单链结合为双链分子的过程。

核酸探针：指序列或功能特性已知的标记分子，为双链或单链，长度通常为几十至几百bp，包括DNA探针、RNA探针和cDNA探针。

Southern杂交 (Southern印迹)



可有效地分析基因结构或检测待测DNA样本中是否含有特定的序列。

Northern 杂交 (Northern 印迹)

NORTHERN BLOT (RNA)

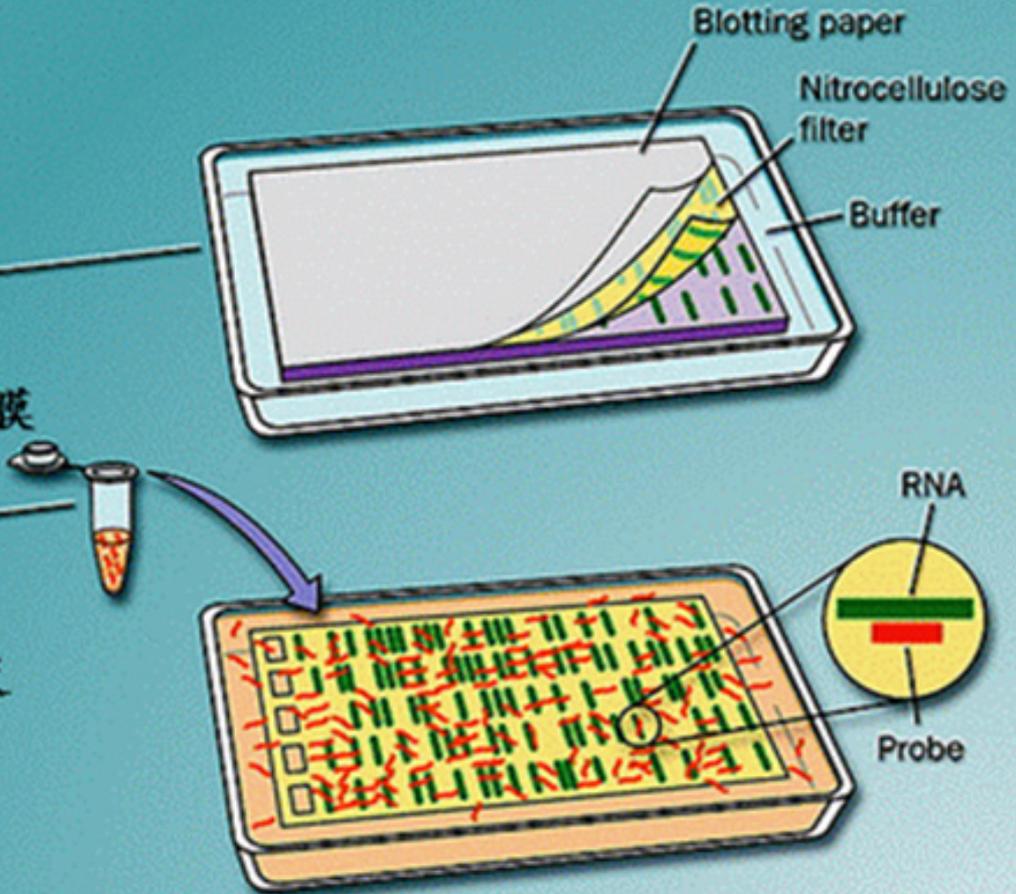
- 1 Separate RNA molecules by gel electrophoresis.
凝胶电泳分离 RNA

- 2 Blot gel onto nitrocellulose filter. RNA moves to filter by capillary action.
转印至硝酸纤维薄膜

- 3 Add radioactive RNA probe, which hybridizes with sequences of interest.
加入 RNA 探针并杂交

- 4 Wash away unbound probe.
洗脱未杂交探针

- 5 Make autoradiogram and study the results.
放射显影, 研究结果



原位杂交技术

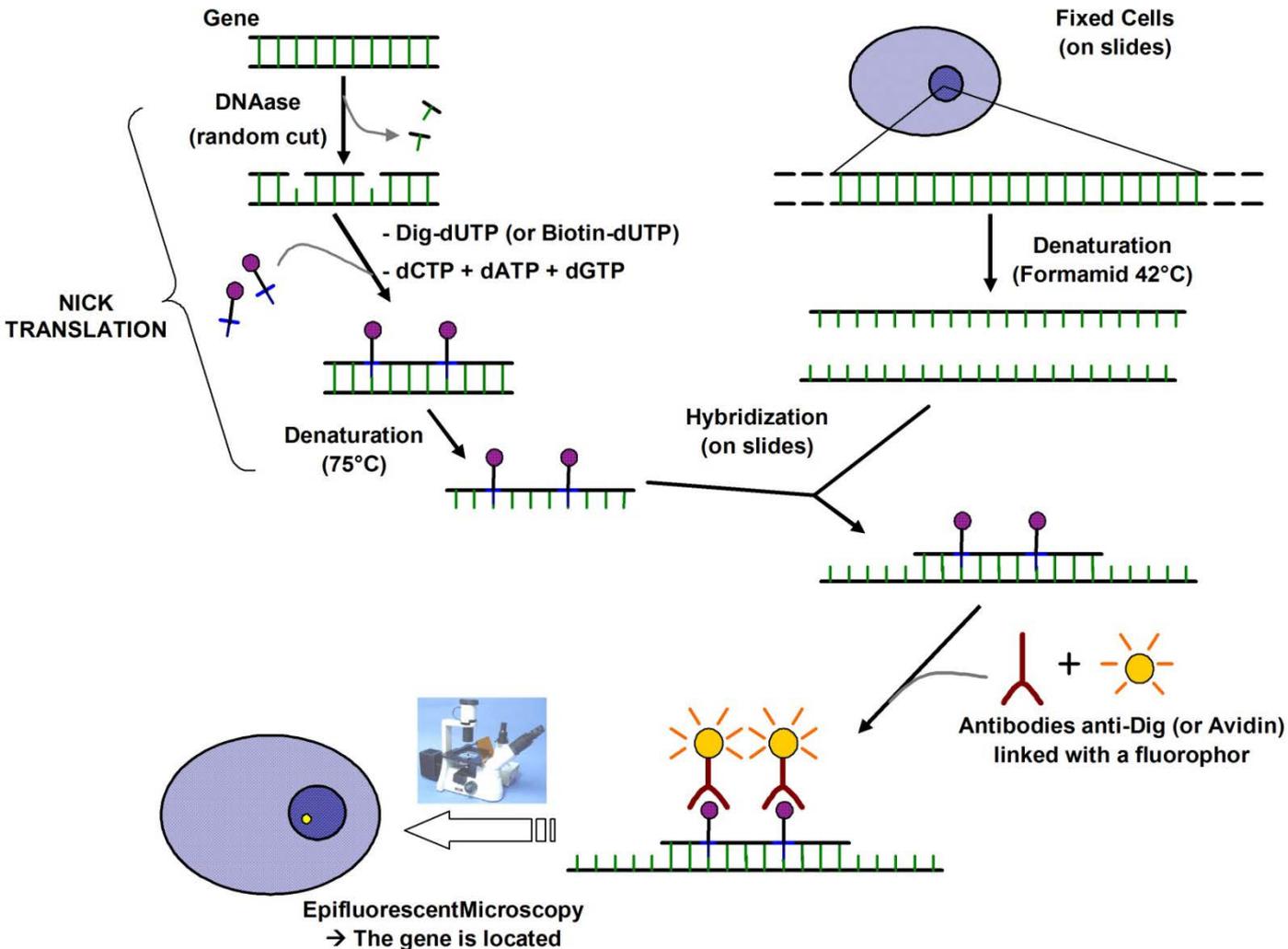
(In Situ Hybridization, ISH)

- 用标记的核酸探针，经放射自显影或非放射检测体系，在组织、细胞、间期核及染色体上对核酸进行定位和相对定量研究的一种手段，分为**RNA**和染色体原位杂交两大类。
- 用放射性或非放射性（如地高辛、生物素等）标记的特异性探针与被固定的组织切片反应，若细胞中存在与探针互补的**mRNA**分子，两者杂交产生双链**RNA**，可通过放射性标记或经酶促免疫显色，对该基因的表达产物做出定性定量分析。
- 用**mRNA**的互补链为探针进行杂交。
- 负对照，**mRNA**的同义链，无杂交信号。

荧光原位杂交

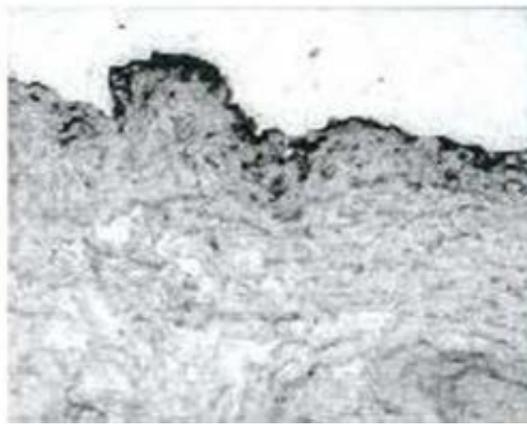
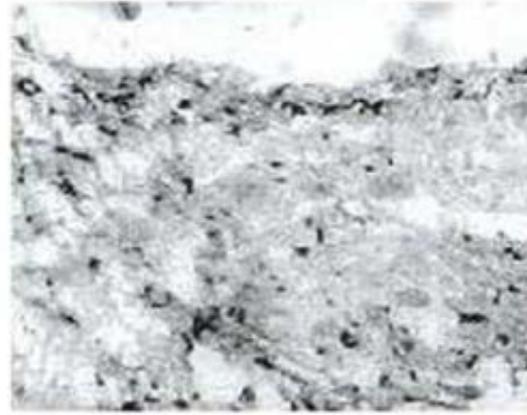
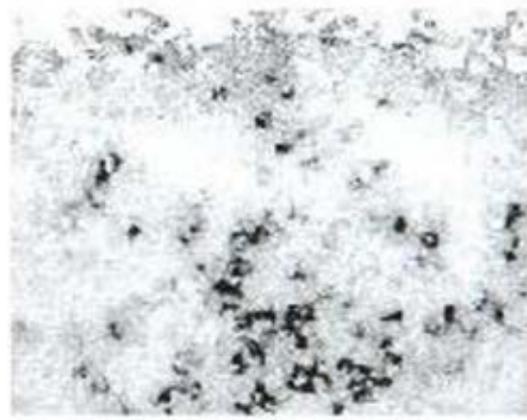
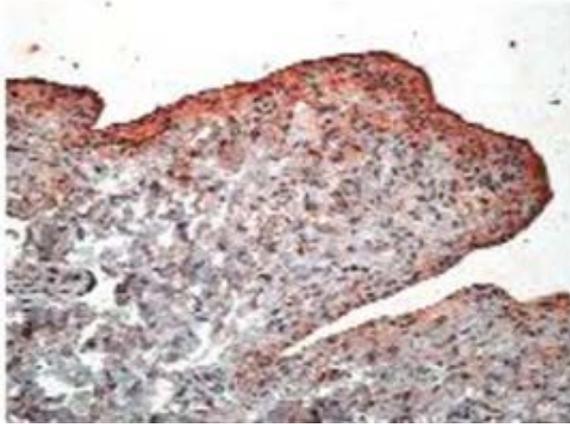
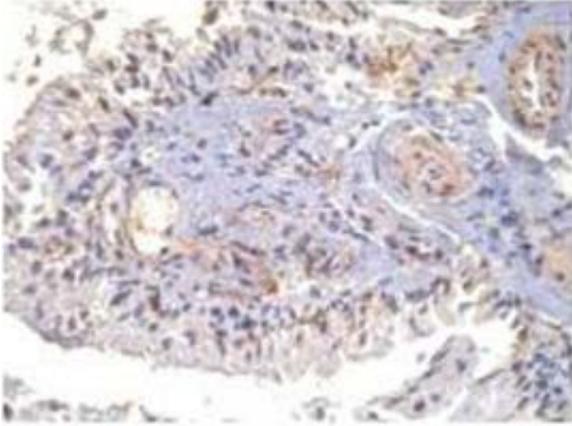
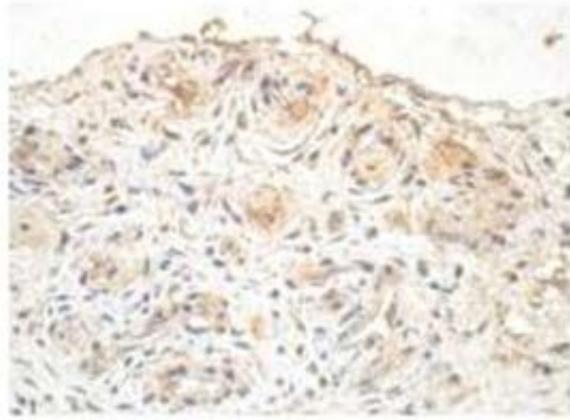
(fluorescence in situ hybridization, FISH)

FISH (Fluorescent In Situ Hybridization)



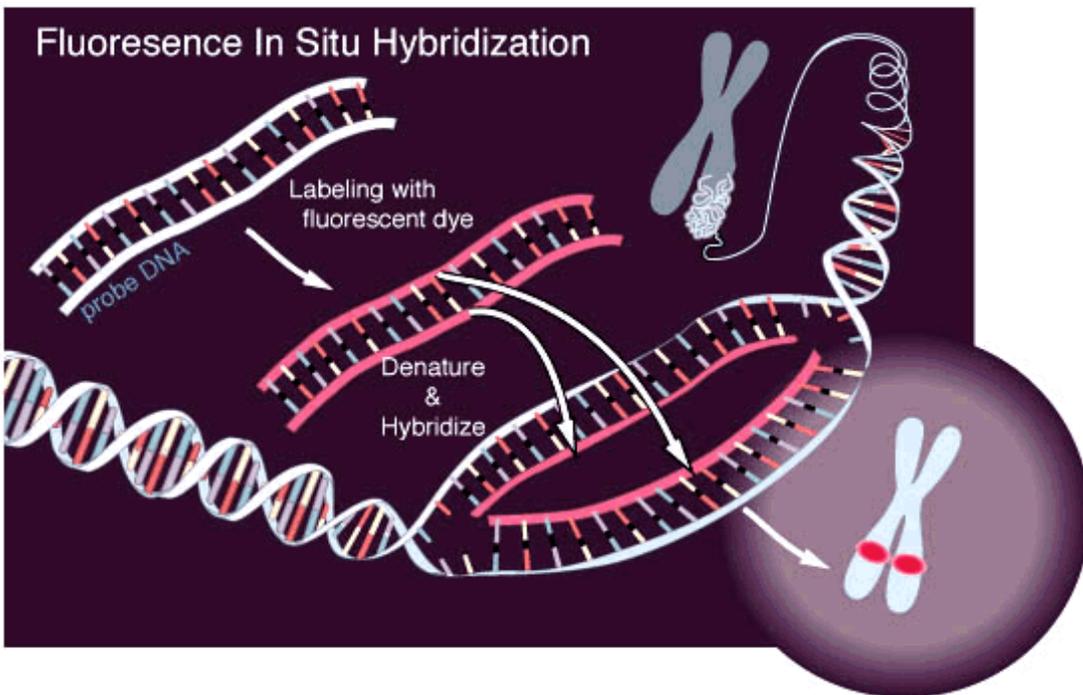
对寡核苷酸探针做特殊的修饰和标记，用原位杂交法与靶染色体或DNA上特定的序列结合，再通过与荧光素分子相耦联的单克隆抗体来确定该DNA序列在染色体上的位置。

地高辛标记

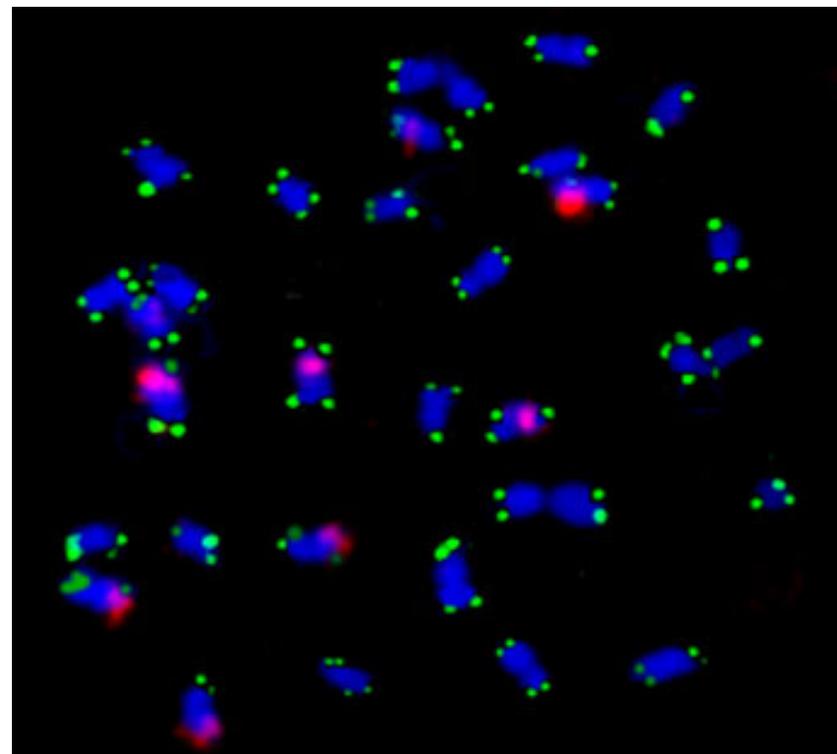


同位素标记

染色体原位杂交

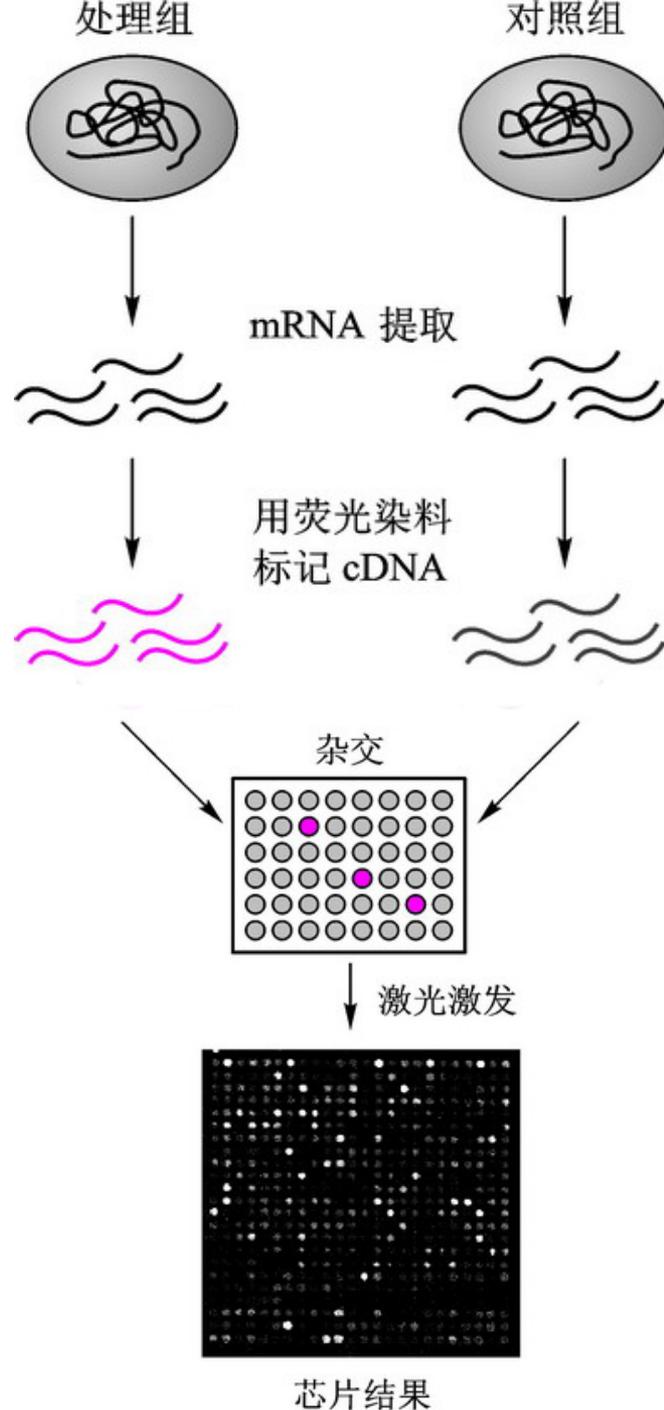


不需要放射性同位素，实验周期短，检测灵敏度高，可同时观察几种DNA探针的定位，得到相应位置和排列顺序的综合信息。



人类染色体端粒DNA的荧光原位杂交

基因芯片及数据分析



基因芯片（**DNA chip**），又称**DNA微阵列（DNA microarray）**技术，是能同时监测大量靶基因表达的实验手段，从而迅速准确地**在基因组水平上阐述不同生物组织或细胞中各种转录本的变化规律**。

基因芯片的应用

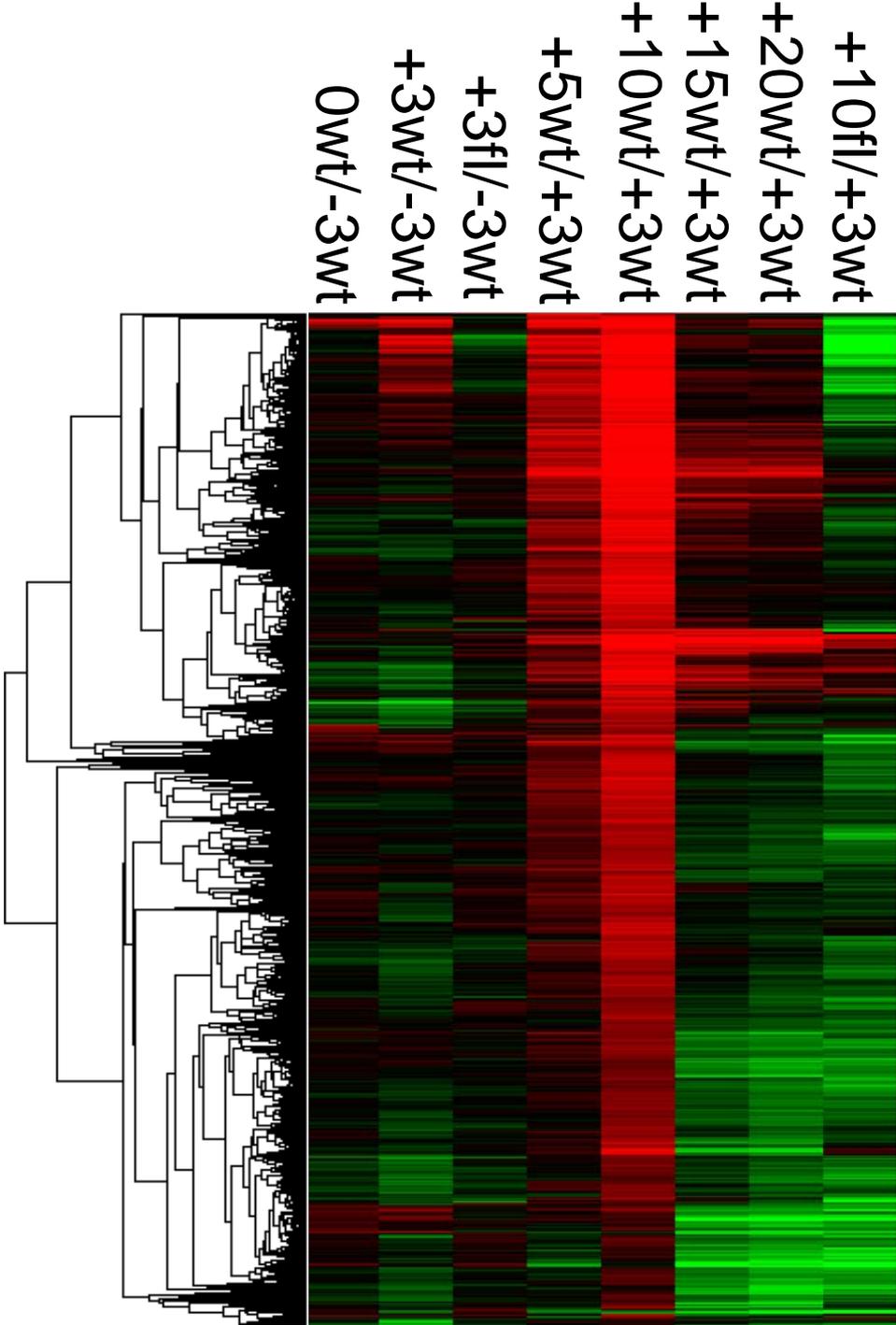
基因芯片可用于进行基因诊断。

先建立正常人特定组织、器官的基因芯片，给出标准杂交信号图，用可疑病人的cDNA做探针与之杂交，确定哪些基因的表达受抑制或激活。

棉花胚珠cDNA芯片差异表达基因聚类分析

红色指示在纤维细胞中特异性表达上调的基因，

绿色指示表达下调的基因。



cDNA测序 (cDNA sequencing)

(1) 表达序列标签测序技术(Expressed Sequence Tag, EST)

EST测序数据是目前数量最多，涉及物种最广的转录组数据，但测序读长较短（每个转录本测定400bp—500bp），测序通量小，测序成本较高，而且无法通过测序同时得到基因表达丰度的信息。

(2) 基因表达系列分析技术

(serial analysis of expression, SAGE)

以DNA序列测定为基础分析全基因组表达模式的技术。

(3) 高通量测序技术

Why cDNA sequencing?

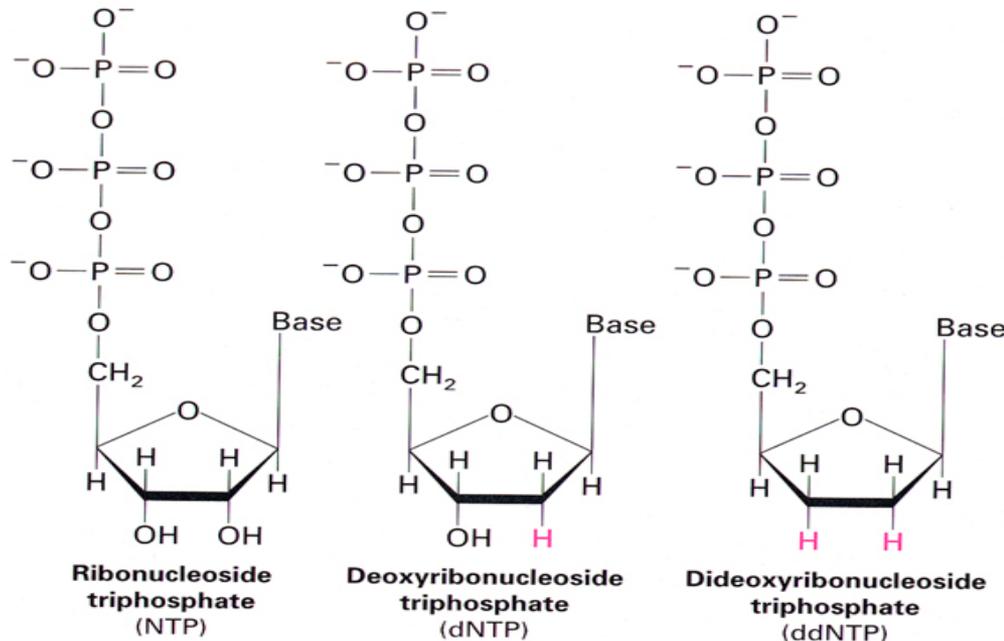
Gain information about the transcriptome

- mRNA populations are variable among cells
- The transcriptome is dynamic and constantly changing
- Cells adapt to environmental, developmental and other signals by modulating their transcriptome
- mRNA populations form an important level of regulation between signal perception and response
- cDNA sequencing allows a direct insight into mRNA populations and allows a dissection of the transcriptome that genomic sequencing alone does not provide
- Sequencing of random cDNA clones prepared from different tissues allows analyses of mRNA abundance

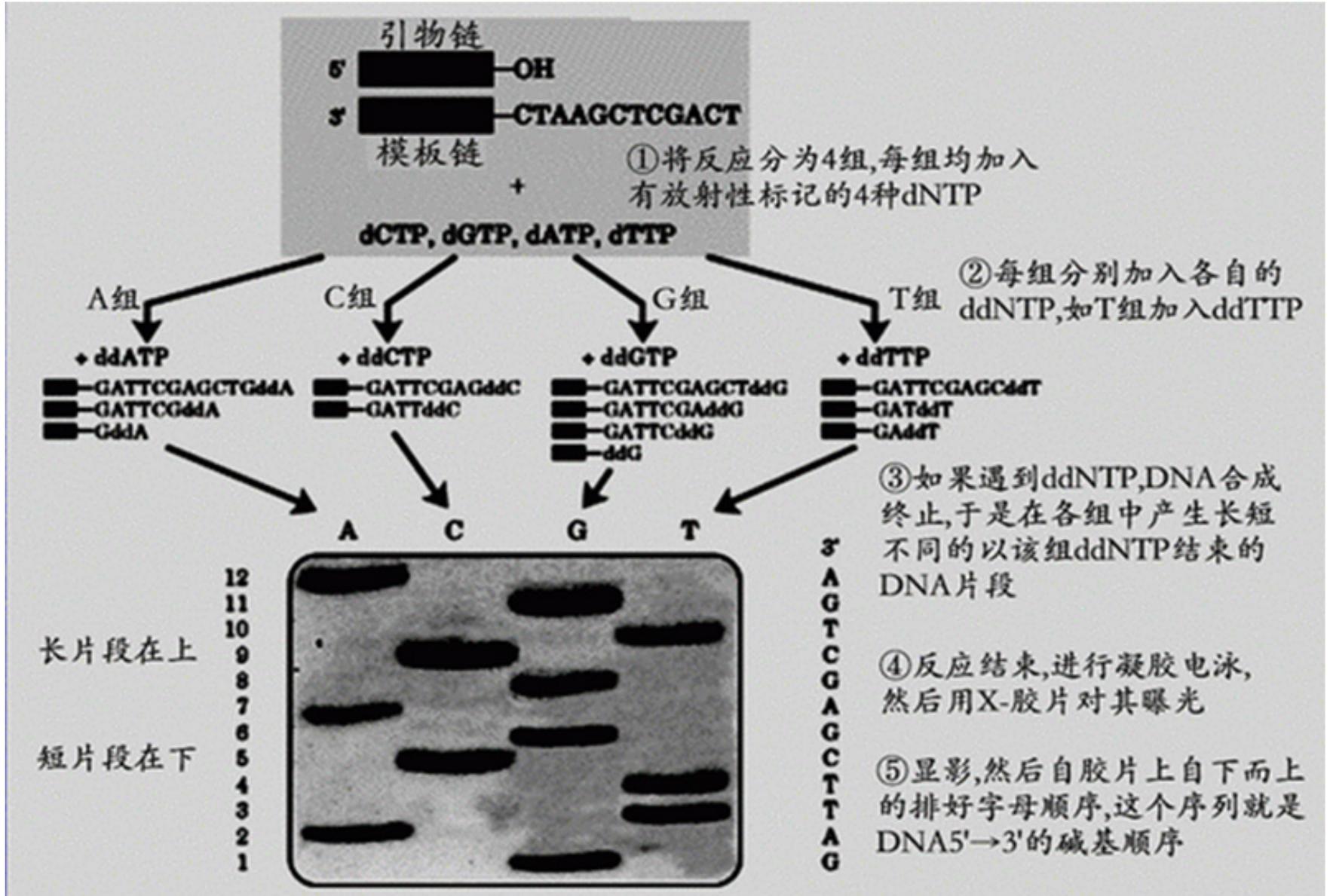
Sanger双脱氧核苷酸链终止法

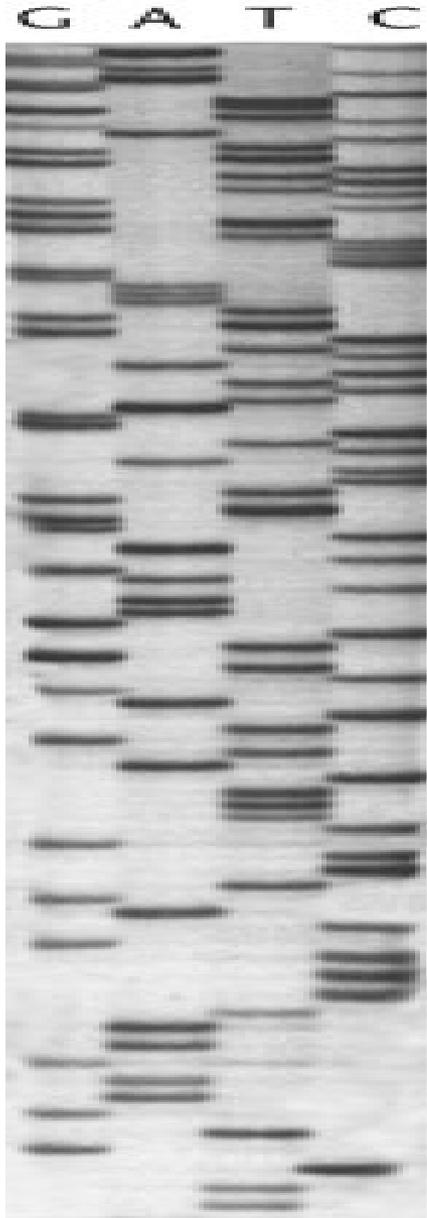
原理：

双脱氧（2', 3'）-核苷酸可以象2'-脱氧核苷酸那样直接掺入新合成的DNA链中，但因3'端不具OH基，DNA链合成至此终止。由于双脱氧核苷酸在每个DNA分子中掺入的位置不同，故可根据不同长度的DNA片段测定出核苷酸序列。



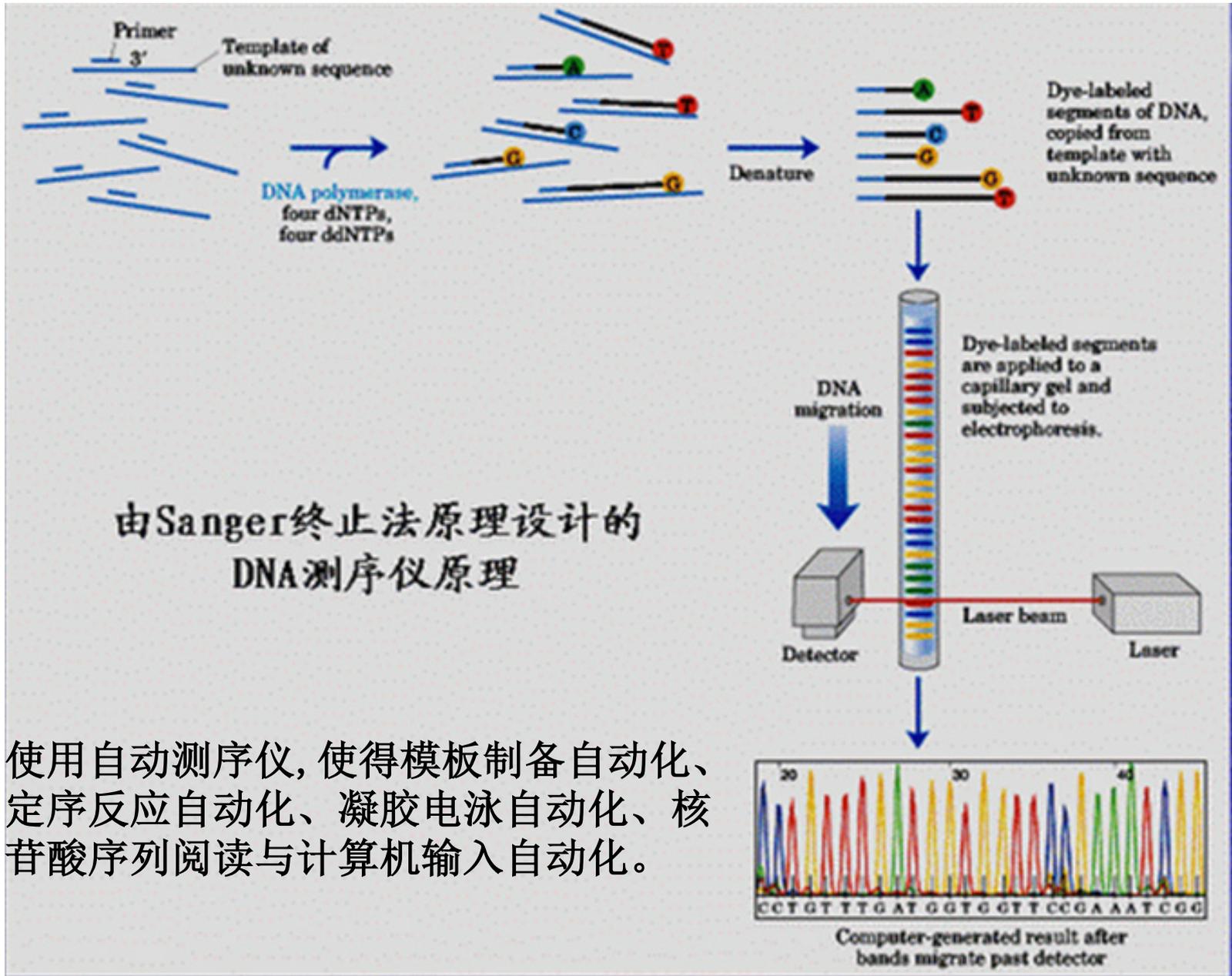
DNA酶法测序步骤





DNA sequencing gel showing the separation of fragments in the four sequencing reactions (one per lane).

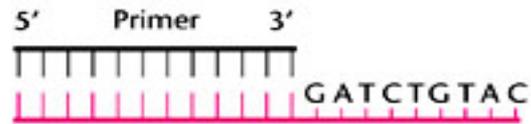
DNA序列测定的自动化



由Sanger终止法原理设计的
DNA测序仪原理

使用自动测序仪,使得模板制备自动化、定序反应自动化、凝胶电泳自动化、核苷酸序列阅读与计算机输入自动化。

1. Primer added



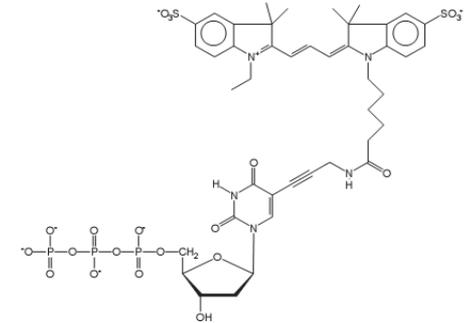
2. Reaction ingredients added

DNA polymerase
dATP
dCTP
dGTP
dTTP

small amount of ddNTPs
with fluorochromes:

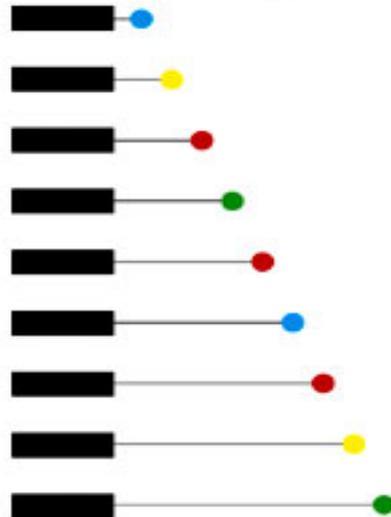
ddATP —●— (red)
ddCTP —●— (blue)
ddGTP —●— (green)
ddTTP —●— (yellow)

荧光素标记核苷酸—应用举例



Cy3-AP3-dUTP

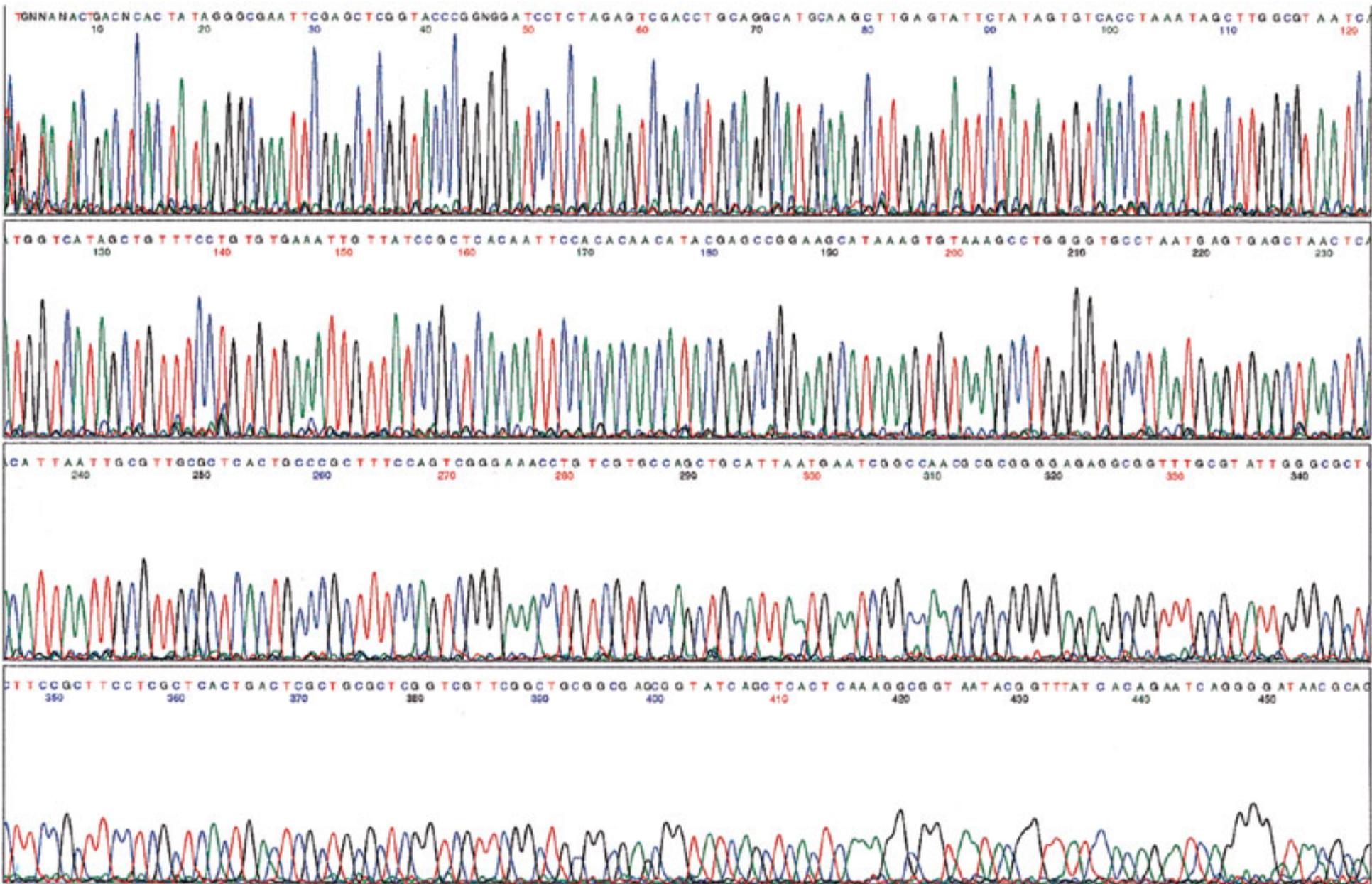
3. Primer extension Chain termination Product recovery



4. Electrophoresis, imaging, data analysis

In DNA sequencing using dideoxynucleotides labeled with fluorescent dyes, all four ddNTPs are added to the same tube, and during primer extension, all combinations of chains are produced.

Automated DNA sequencing using fluorescent dyes, one for each base



DNA序列测定的自动化

优点:

- (1) 四个反应系统可合并点样，大大节省制胶和点样时间。
- (2) 可同时读出多个样品核苷酸序列，不需干燥凝胶和放射自显影。
- (3) 可连续电泳，可读出较多的核苷酸序列。

缺点:

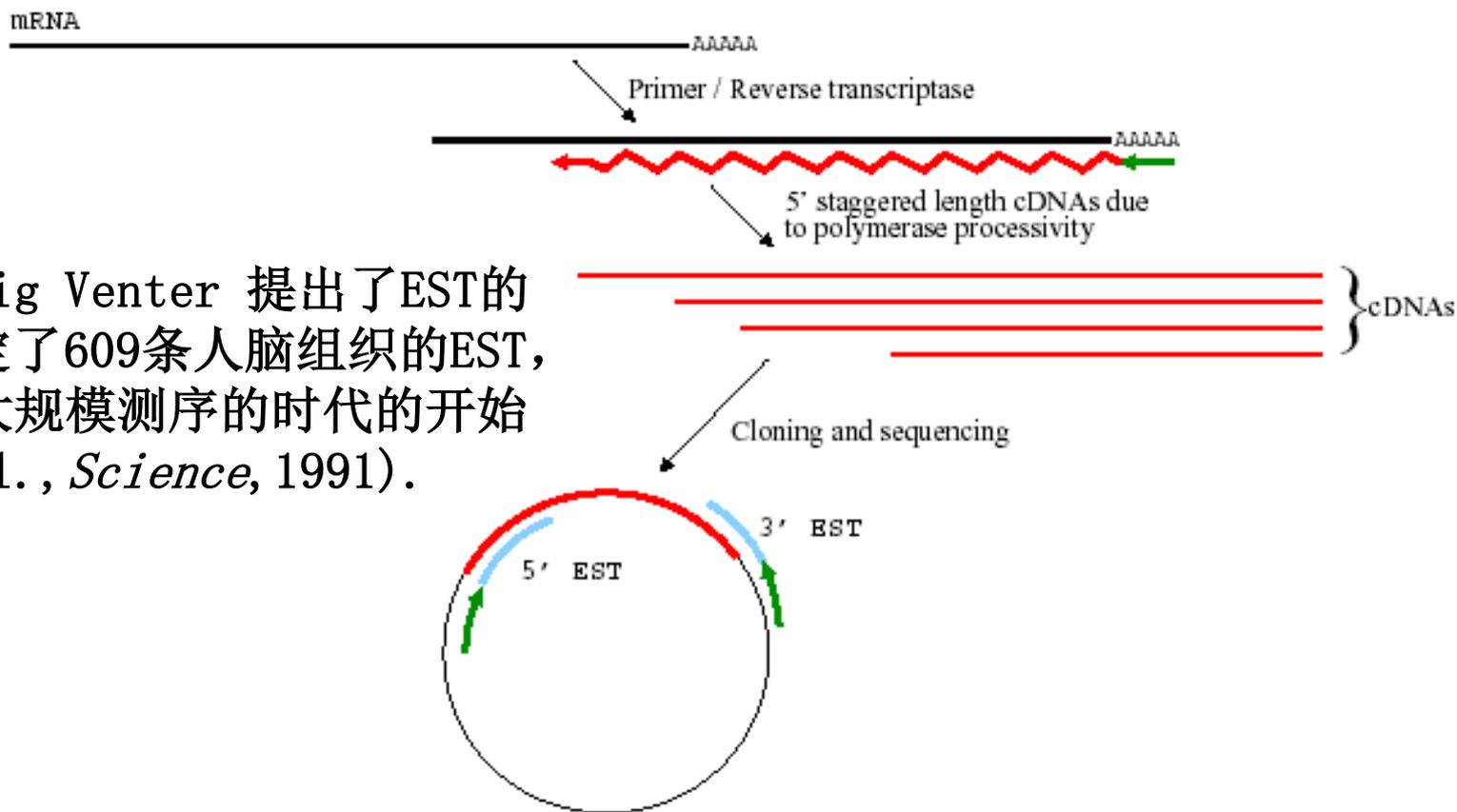
无法保留原始记录，四个反应产物点在一条道上，相互间会发生干扰，导致读序时分辨率下降。

DNA序列分析自动化包括两个方面的内容:

- “分析反应”的自动化;
- “读片过程”的自动化。

什么是 ESTs ?

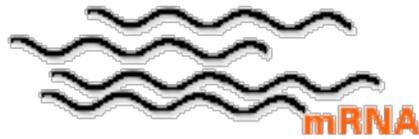
ESTs (Expressed Sequence tags) 是从已建好的cDNA库中随机取出一个克隆, 从5'末端或3'末端对插入的cDNA片段进行一轮单向自动测序, 所获得的约60-500bp的一段cDNA序列。



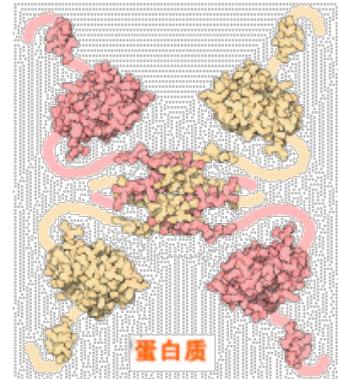
90年代初Craig Venter 提出了EST的概念, 并测定了609条人脑组织的EST, 宣布了cDNA大规模测序的时代的开始 (Adams et al., *Science*, 1991).

EST 技术流程

体外研究：反转录

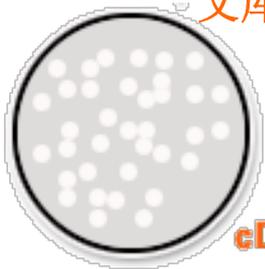


体内：翻译



连接，转化效率问题（基因芯片）
转化

文库构建技术已经成熟



大数据量分析理念已经形成

质量控制和数据分析

测序采样问题（SAGE）

测序成本已经大大降低

EST的应用

1. ESTs与基因识别

- 在同一物种中搜寻基因家族的新成员 (paralogs)。
- 在不同物种间搜寻功能相同的基因 (orthologs)。
- 已知基因的不同剪切模式的搜寻。

2. ESTs与基因图谱的绘制

3. ESTs与基因预测

对预测基因的交替剪切和3' 非翻译区很有效。

4. ESTs与SNPs

不同个体冗余的ESTs可用于发现基因组中转录区域的SNP。

5. 利用ESTs大规模分析基因表达水平

因为EST序列是从某以特定的组织的cDNA文库中随机测序而得到，所以可以用利用未经标准化和差减杂交的cDNA文库EST分析特定组织的基因表达谱。

ESTs数据的不足

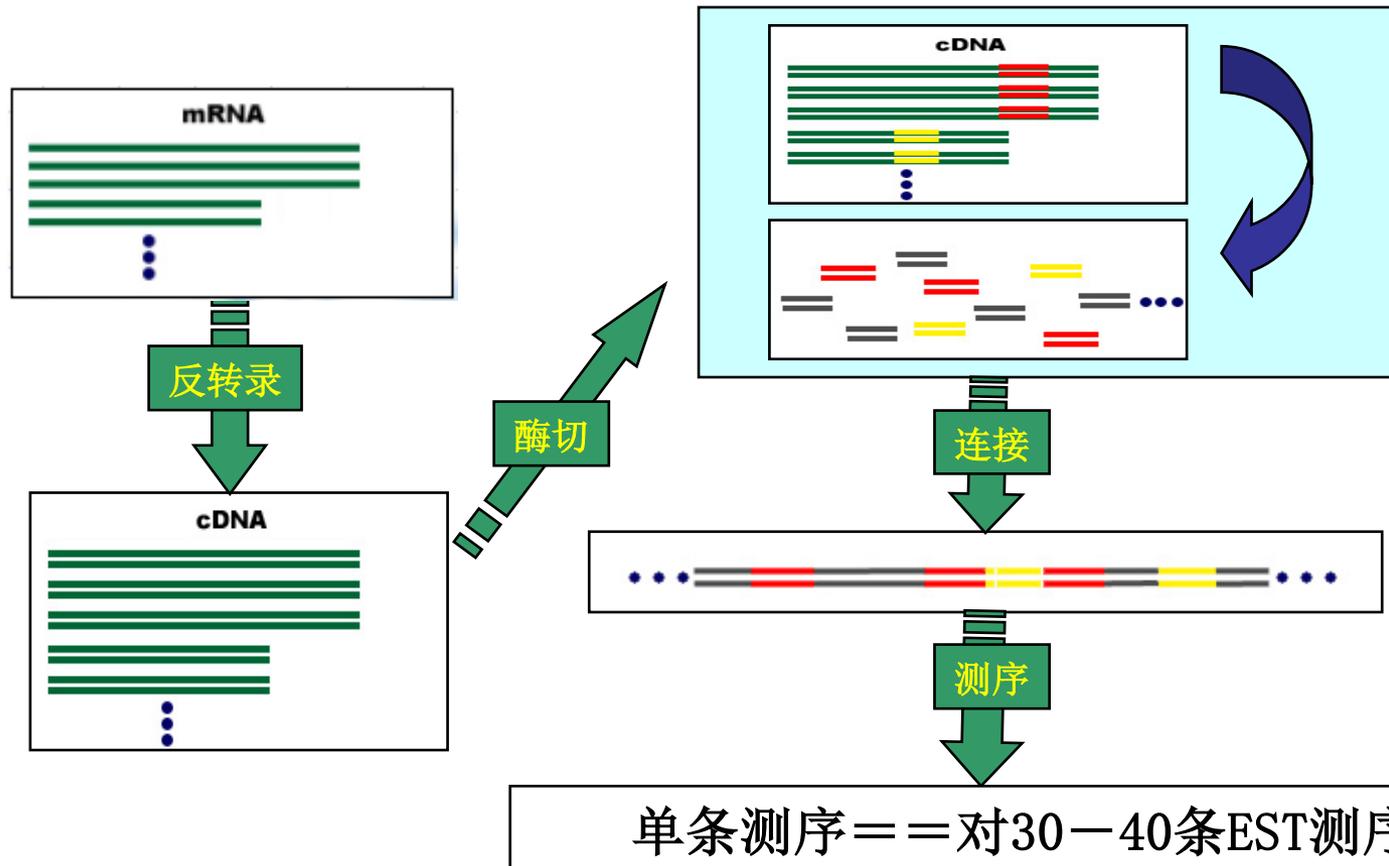
- ◆ ESTs很短，没有给出完整的表达序列；
- ◆ 低丰度表达基因不易获得。
- ◆ 由于只是一轮测序结果，出错率达2%-5%；
- ◆ 有时有载体序列和核外mRNA来源的cDNA污染或是基因组DNA的污染；
- ◆ 有时出现镶嵌克隆；
- ◆ 序列的冗余，导致所需要处理的数据量很大。

基因表达系列分析

(Serial Analysis of Gene Expression, SAGE)

- 基因表达系列分析是一种用于定量，高通量基因表达分析的实验方法(Velculescu et al., 1995)。
- **SAGE**的原理就是分离每个转录本的特定位置（3'端第一个**CATG**位点下游）的较短的单一的序列标签（约**9-14**个碱基对），这些短的序列被连接、克隆和测序，特定的序列标签的出现次数就反应了对应的基因的表达丰度。
- **SAGE**法相对于**EST**测序在通量上大大提高。
- 但过短的序列标签使得序列唯一性降低，即使改进过的**Long SAGE**用**21 bp**标签测序，仍然有约一半的标签无法被准确注释到基因组上。

SAGE技术流程



实验步骤较长要求较高

通过Genebank或EST数据库进行比对，
确认tag代表的基因表达情况。



由于采样量大大提高，可对低表达基因进行分析：
基因表达量分析、寻找新基因等等

几种大规模分析基因表达水平方法的比较

	EST	SAGE	Microarray
发现新基因	是	是	否
有序列	是（可直接进行可变剪切的分析）	否	否
主要问题	采样量	实验过程	重复性

高通量测序技术 (high-throughput sequencing)

又名二代测序（**second-generation sequencing**）或深度测序(**deep sequencing**)，可以一次性测序几十万甚至几百万条，是传统测序技术的革命。

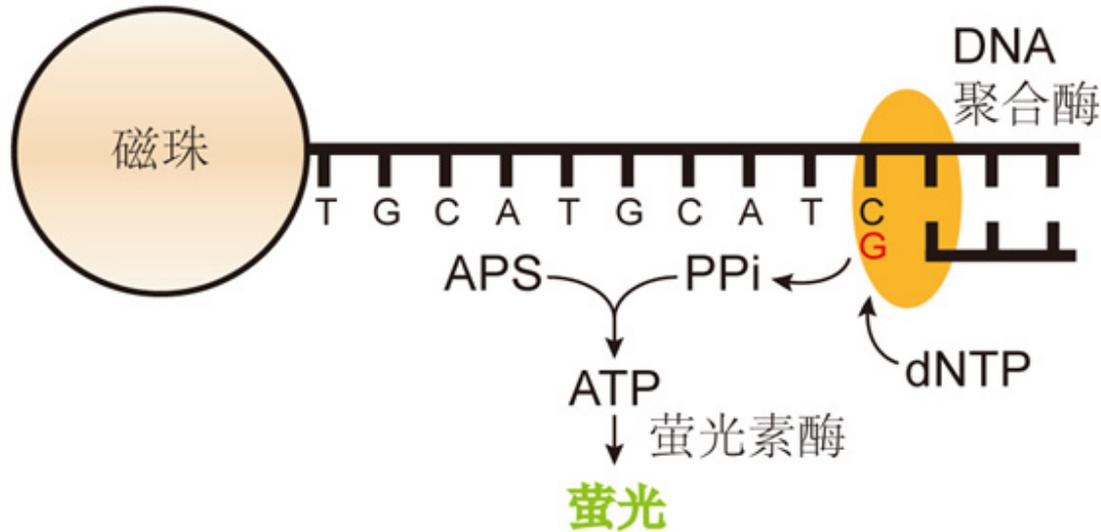
利用高通量测序技术对转录组进行测序分析被称为**RNA-Seq**，对测序得到的大量原始读长（**reads**）进行过滤、组装以及后续的生物信息学分析。

由于测序深度的优势，**RNA-seq**更能全面地揭示生物个体在特定时刻和特定组织的全局基因表达情况，如发现新的转录本、了解基因表达量、挖掘单核苷酸的多态(**SNPs**)、选择性剪接和结构性变异。

454和Solexa高通量测序平台比较

	454	Solexa
读取长度 (bp)	约700	50—150
单次测序数据量	700 Mb	600 Gb
测序周期	23小时	7—14天
测序成本	较高	低

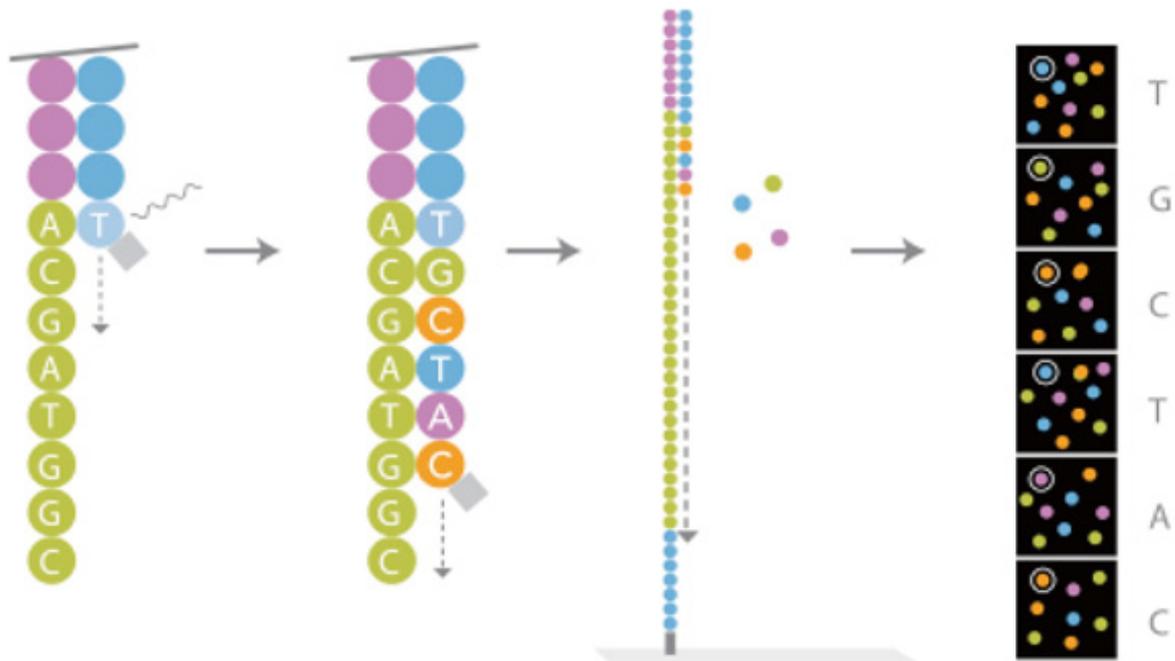
454测序



454系统采用焦磷酸测序(pyro-sequencing): 按T、A、C、G顺序加入单个dNTP与模板配对，释放一分子焦磷酸(PPI)，在ATP硫酸化酶作用下PPI和腺苷酰硫酸(adenosine-5'-phosphosulfate, APS)结合形成ATP，在萤光素酶的催化下产生可见光。

454则由于缺少终止反应的元件，相同碱基的连续掺入常会带来“插入-缺失”类型的测序错误。

Solexa测序系统

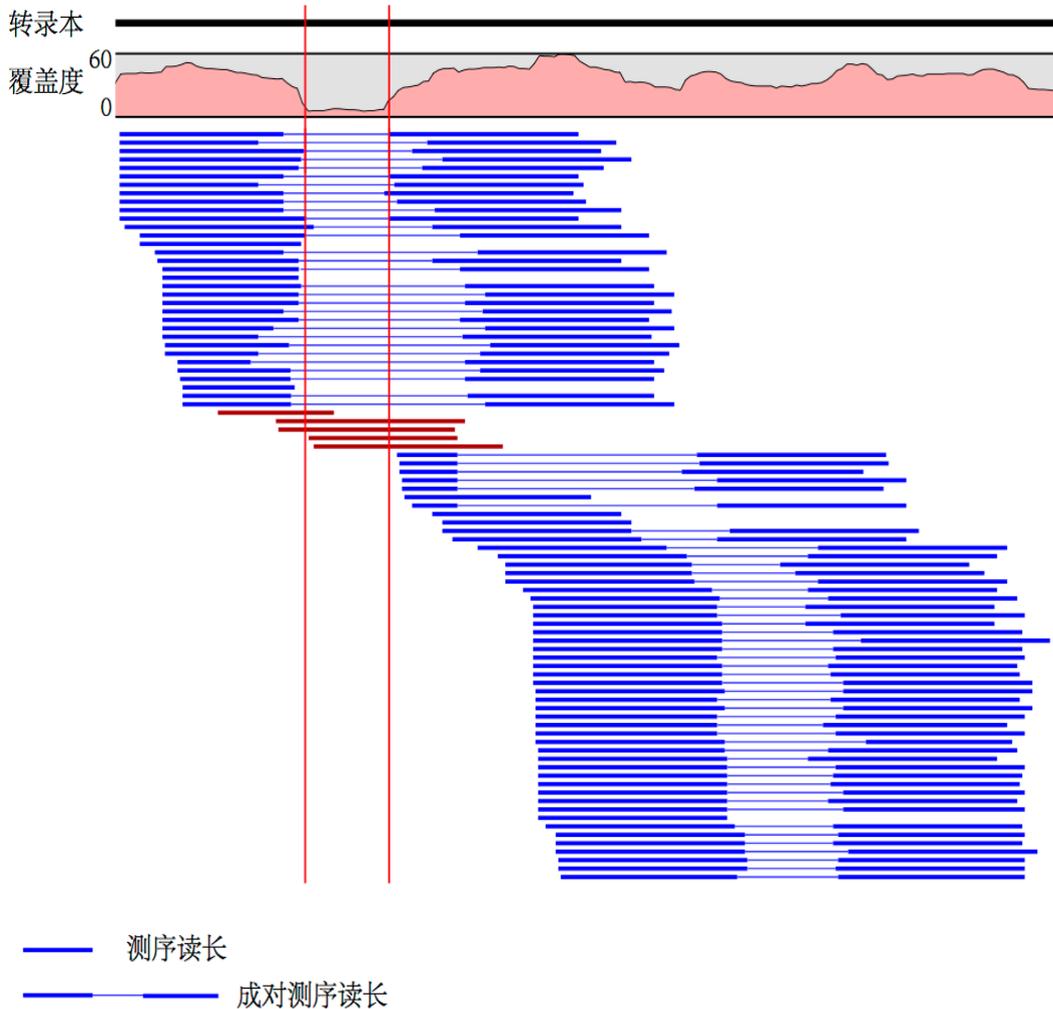


Solexa测序系统采用荧光标记dNTP，其3'羟基末端可被化学切割，每个循环反应只允许掺入一个碱基，由激光扫描反应板表面，读出这一轮反应新加的荧光信号，从而判定碱基种类。之后，经过化学切割恢复3'端粘性，进行下一轮聚合反应。

随着**Solexa**测序反应的进行，已有荧光信号会使新的荧光难以准确分辨，因此该方法的测序读长较短，测序错误主要是**碱基替换**。

DNA序列的从头组装 de novo assembly

在实际组装过程中，图中红色标示区域覆盖度过低，且读长缺乏相对位置信息的区域，其可信度较低，应当剔除，只保留两侧序列。



- 利用大量读长之间重叠覆盖和成对读长（**pair-end reads**）的相对位置关系，组装得到尽可能完整的转录本。
- 以单位长度转录本上覆盖的读长数目（**reads per kilo-base gene per million bases, RPKM**）作为衡量基因表达水平的标准。

RNA-seq的特点及应用

- **数字化信号:**

直接测定每个转录本片段序列,单核苷酸分辨率的精确度,可以检测单个碱基差异、基因家族中相似基因以及可变剪接造成的不同转录本的表达,同时不存在传统微阵列杂交的荧光模拟信号带来的交叉反应和背景噪音问题,能覆盖信号超高的动态变化范围。

- **高灵敏度:**

能够检测到细胞中少至几个拷贝的稀有转录本。

- **任意物种的全基因组分析:**

无需预先设计特异性探针,因此无需了解物种基因信息,能够直接对任何物种进行转录组分析,这对非模式生物的研究尤为重要。

- **更广的检测范围:**

高于6个数量级的动态检测范围,能够同时鉴定和定量稀有转录本和正常转录本;而芯片对过低或过高表达的基因缺乏敏感性,因而动态检测范围小

- **RNA-Seq重复性好,起始样品少,适用于来源极为有限的生物样品分析.**

RNA-seq技术的缺陷

- 测序技术的改进和测序费用的高；
- 生物信息学分析软件的开发不彻底，测序虽然得到了大量的数据，但基于有限的生物信息学分析软件，经常导致人们所对要分析的东西不能得到很好的结果，造成测序数据的浪费；
- 平台测序长度不够大，目前虽然发展起来的 **454 FLX** 测序平台测序长度能达到**700bp**以上，但测序费用的昂贵。

Where is sequencing going?

- Single Molecule
- Single cell
- Real time
- Whole length of a genome
- 100% correct
- Free

What's the difference between RNA-Seq, microarray and RT-PCR ?

Technology	Advantage	Disadvantage
Microarrays	Genome wide, relatively cheap, streamlined handling, oligos	Sequences must be known in advance; limited sensitivity due to hybridization
Quantitative reverse transcription-PCR	High precision and high sensitivity Increasingly multiplexed	Not genome wide; data normalization sensitive to method/choice of reference genes
High-throughput sequencing	Sequences do not need to be known in advance; possibility to sequence very short sequences	Expensive at the moment, few solutions for downstream analysis; direct read out

RNAi (RNA interference) 技术及其应用

RNAi技术利用双链小RNA高效、特异性降解细胞内同源mRNA从而阻断靶基因表达，使细胞出现靶基因缺失的表型。

•RNAi的命名来源于安德鲁·法尔。



Andrew Z. Fire

斯坦福医学院病理学和遗传学教授

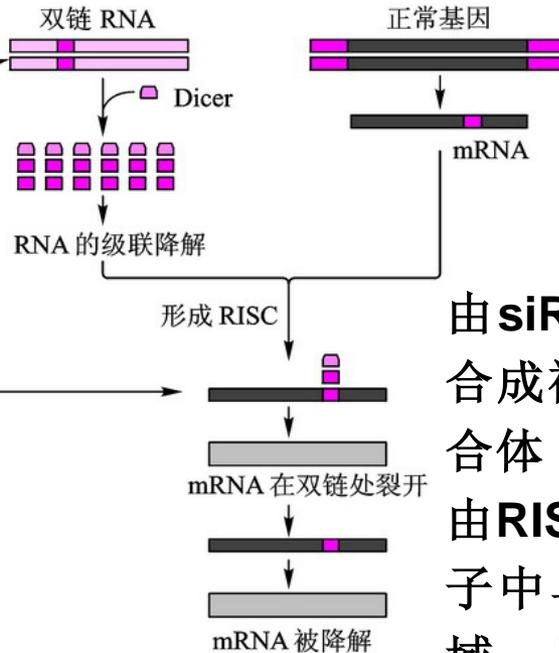
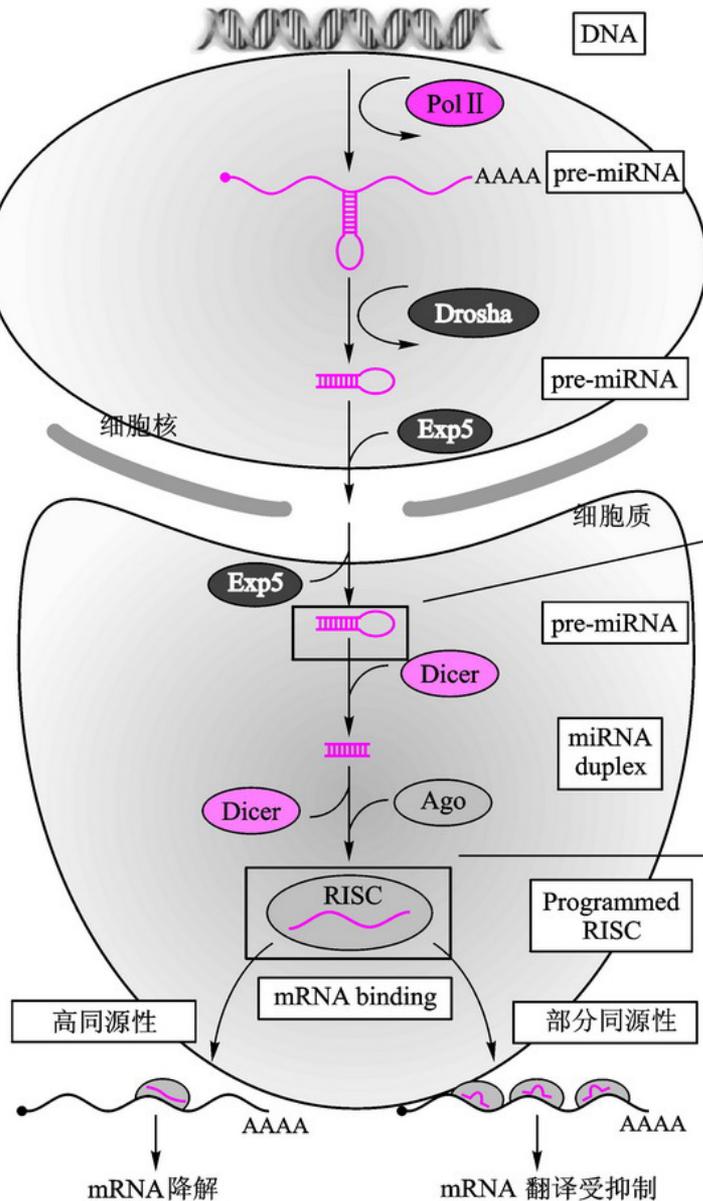


Craig C. Mello

马萨诸塞州医学院分子医学教授

RNAi作用机制

经过**Dicer**（一种具有RNase III活性的核酸酶）的加工，细胞中较长的双链RNA（30个核苷酸以上）首先被降解形成21~25个核苷酸的小分子干扰核糖核酸（**siRNA**, short interfering RNA），并有效地定位目标mRNA。



由**siRNA**中的反义链参与指导合成被称为**RNA**诱导的沉默复合体（**RISC**）的核蛋白体，再由**RISC**介导切割目的**mRNA**分子中与**siRNA**反义链互补的区域，从而实现干扰靶基因表达的功能。

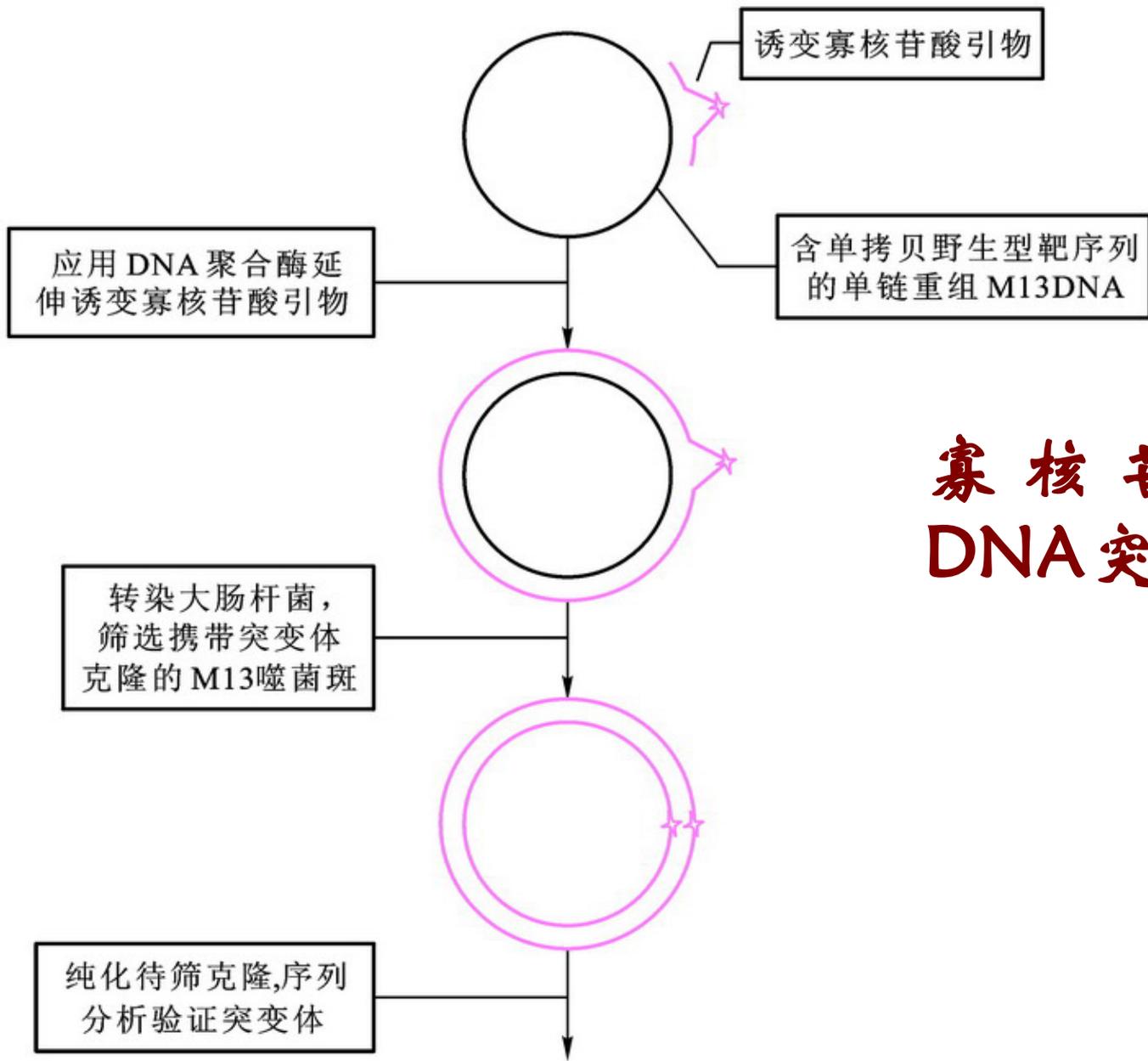
shRNA与siRNA

基因定点突变 (site-directed mutagenesis)

通过改变基因特定位点核苷酸序列来改变所编码的氨基酸序列，用于研究某个（些）氨基酸残基对蛋白质的结构、催化活性以及结合配体能力的影响，也可用于改造**DNA**调控元件特征序列、修饰表达载体、引入新的酶切位点等。

主要采用两种**PCR**方法在基因序列中进行定点突变：

- 重叠延伸技术
- 大引物诱变法

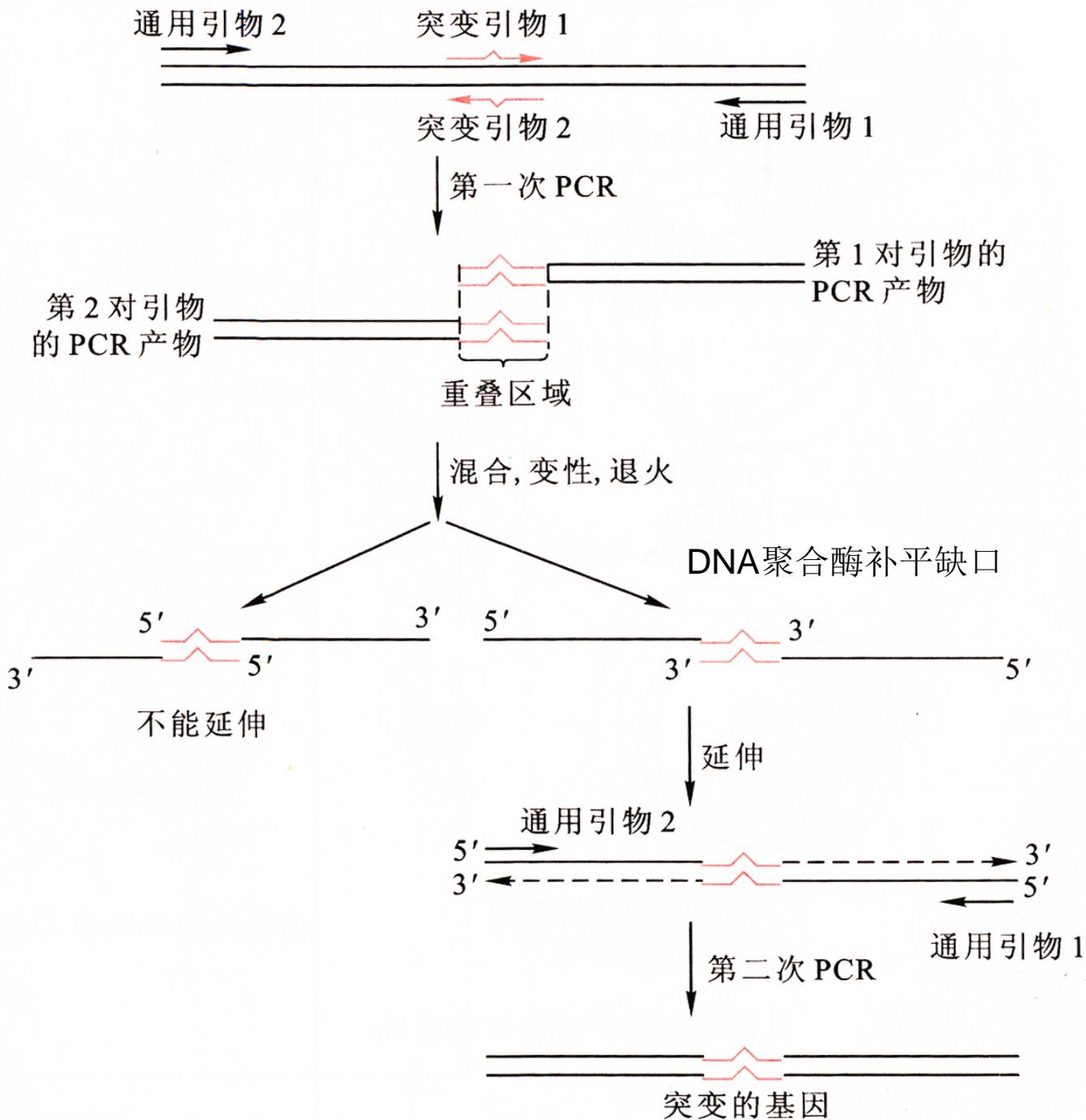


寡核苷酸介导的 DNA 突变技术

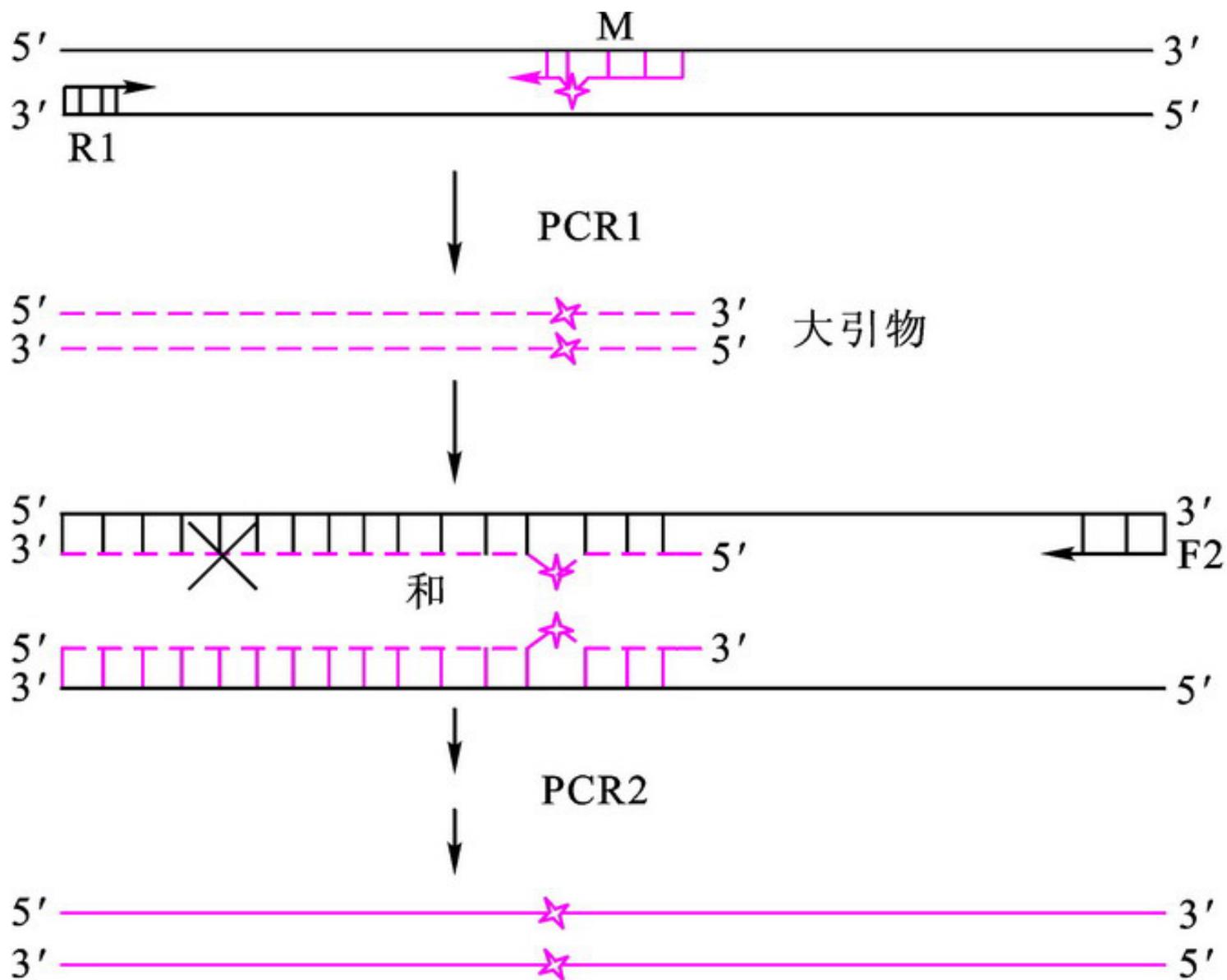
AGGCGGCTA **C** GTGGCCT

AGGCGGCTA **G** GTGGCCT

重叠延伸诱变法



大引物诱变法



PCR介导的定点突变的优点

- 突变体回收率高；
- 能用双链DNA作为模板，可在任何位点引入突变；
- 可在同一试管完成所有反应
- 快速简便，无需在噬菌体M13载体上进行分子克隆。

基因敲除技术

- **经典遗传学 (Forward genetics)** 是从一个突变体的表型出发，研究其基因型，进而找出该基因的编码序列。
- **现代遗传学 (Reverse genetics, 反向遗传学)** 首先从基因序列出发，推测其表现型，进而推导出该基因的功能。



基因敲除的发展史

1952 ↵	<u>BriggsR & KingTC</u> ↵	首次核移植技术成功↵
1981↵	<u>Palmiter RD</u> <u>Brinster RL</u> ↵	转基因技术诞生↵
1981 ↵	<u>EvansMJ</u> & <u>KaufmannMH</u> ↵	胚胎干细胞技术诞生↵
1985 ↵	<u>Smithies</u> & <u>Ronald G</u> ↵	同源重组技术开始应用↵
1987 ↵	<u>Smithies</u> & <u>Capecchi</u> ↵	小鼠中进行基因敲除↵
1997↵	<u>Wilmut I</u> & <u>McWhir J</u> ↵	体细胞克隆绵羊诞生↵
1997 ↵	<u>Schnieke AE</u> & <u>AJ Kind</u> ↵	转基因克隆绵羊诞生↵
2000 ↵	<u>McCreath KJ</u> & <u>Howcroft J</u> ↵	基因打靶绵羊诞生↵
2000 ↵	<u>Akira Onishi et al.</u> ↵	体细胞克隆家猪诞生↵
2002↵	<u>Eiangxue L</u> <u>Donna KS</u> . ↵	基因打靶家猪诞生↵

基因敲除技术

1

基因敲除定义

2

基因敲除的基本技术路线

3

基因敲除的原理和方法

4

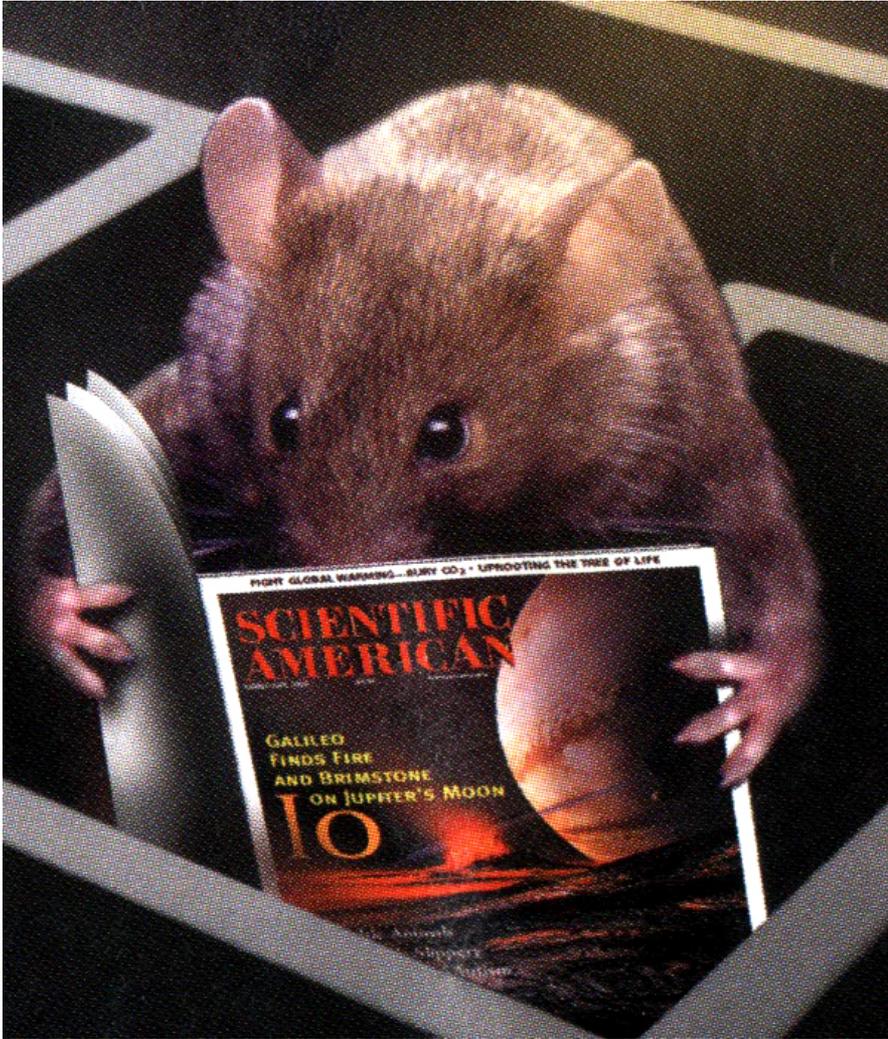
基因敲除的应用、前景及缺陷

基因敲除 (gene knock-out) 又称**基因打靶**，通过外源DNA与染色体DNA之间的同源重组，进行精确的定点修饰和基因改造，具有专一性强、染色体DNA可与目的片段共同稳定遗传等特点。

基因敲除分为下列**2种**：

- 1. 完全基因敲除**：指通过同源重组法完全消除细胞或者动植物个体中的靶基因活性；
- 2. 条件型基因敲除**（又称不完全基因敲除）：指通过定位重组系统实现特定时间和空间的基因敲除。

高等动物基因敲除技术



真核生物基因敲除的技术路线主要包括构建重组基因载体，用电穿孔、显微注射等方法把重组**DNA**导入胚胎干细胞纯系中，使外源**DNA**与胚胎干细胞基因组中相应部分发生同源重组，将重组载体中的**DNA**序列整合到内源基因组中并得以表达。

基因敲除基本步骤

•基因载体的构建

(1)因基因打靶的目的不同，可分为**替换型载体**和**插入型载体**。

(2)打靶载体分为全基因敲除，条件性基因敲除，基因敲进，诱导性基因敲除等。

•ES细胞的获得

利用129，**c57BL/6**遗传背景的ES细胞进行基因打靶。

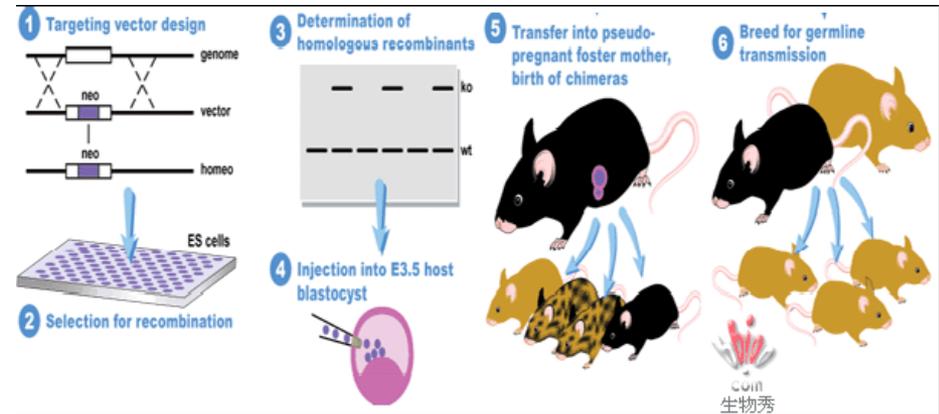
•目的基因导入

通过电穿孔法和显微注射导入同源的胚胎干细胞中。

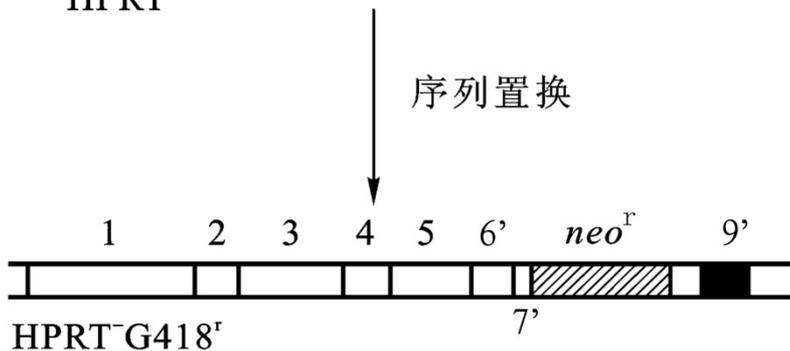
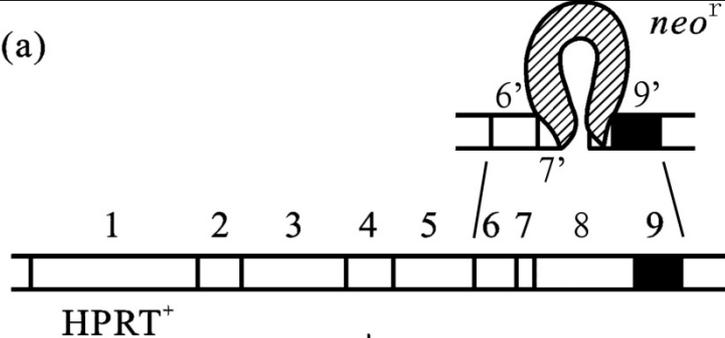
•选择筛选已击中的细胞

•由于基因转移的同源重组自然发生率极低，动物的重组概率为 $10^{-2} \sim 10^{-5}$ ，植物的概率为 $10^{-4} \sim 10^{-5}$ 。目前常用的方法是正负筛选法（PNS法），标记基因的特异位点表达法以及PCR法。

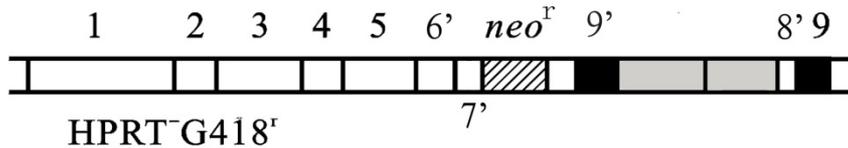
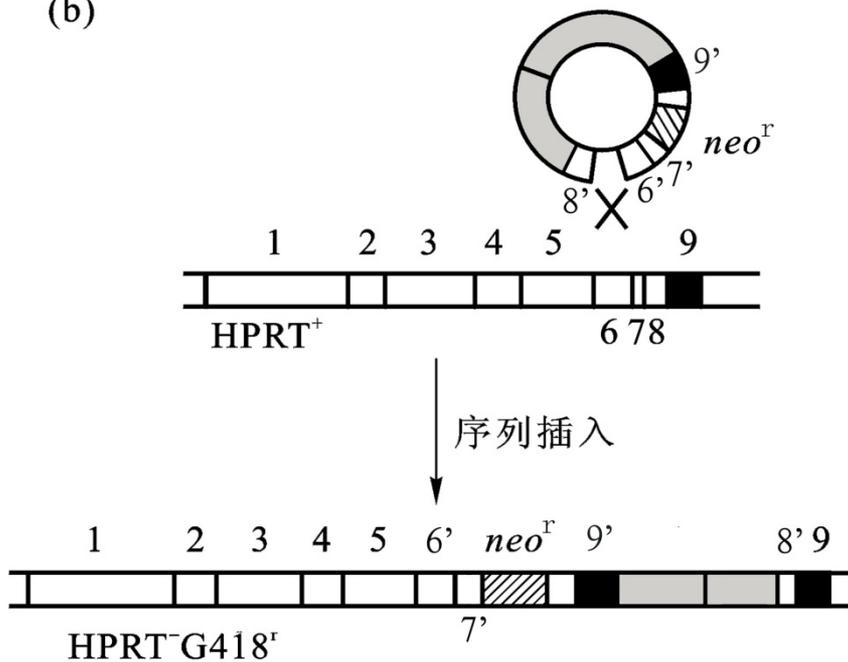
•得到纯合体，表型研究



(a)



(b)



基因载体的构建

- 把目的基因和与细胞内靶基因特异片段同源的DNA分子都重组到带有标记基因(如*neo*基因, TK基因等)的载体上, 成为重组载体。
- 用取代型(替换型)或插入型载体进行完全基因敲除实验。

常用筛选方法

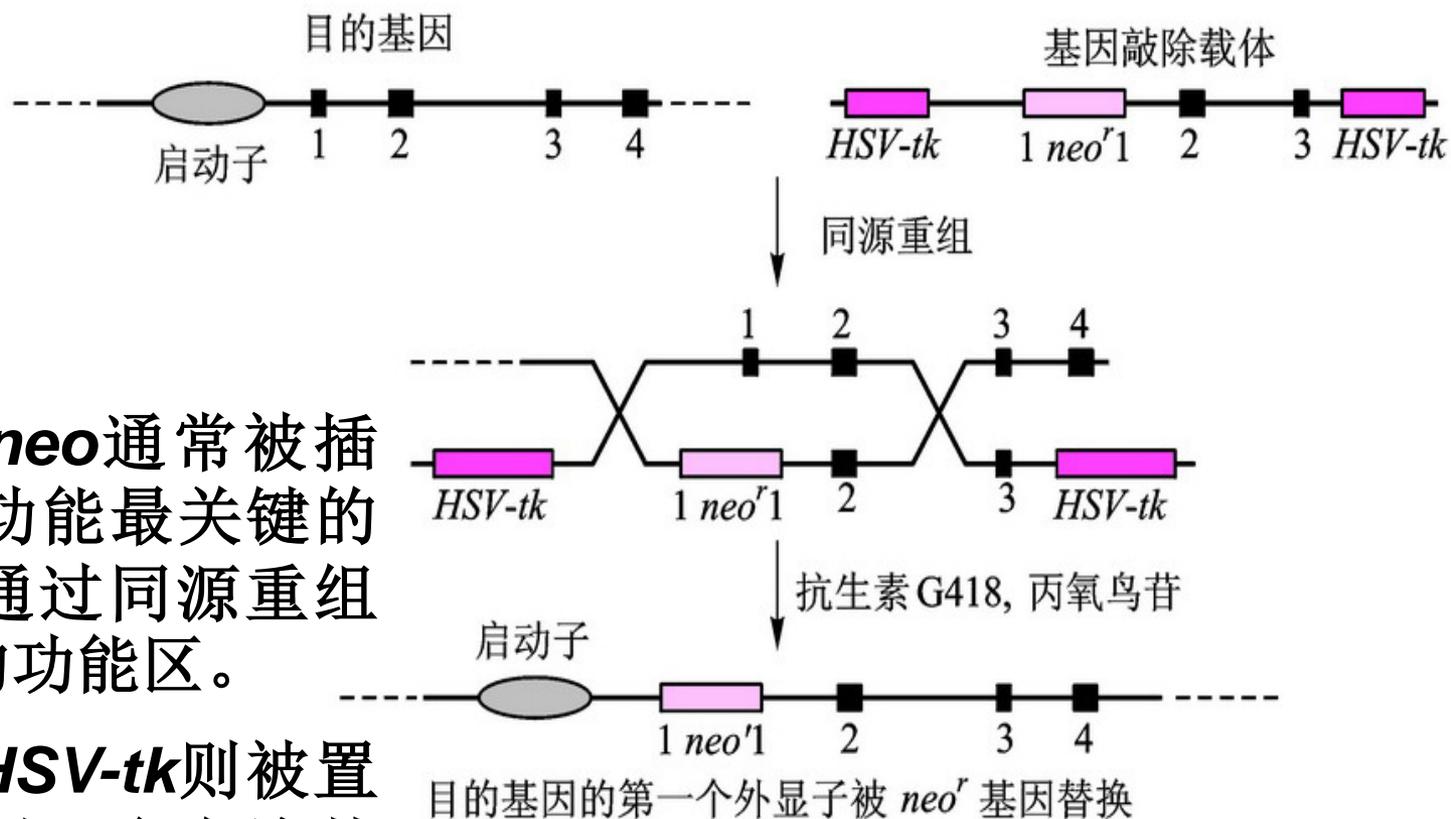
1. 遗传筛选法

- a. 正向筛选
 - (1) 启动子缺失筛选法
 - (2) Poly- A缺失筛选法
- b. 正负双向筛选系统

2. PCR筛选方法

其原理是通过在目的突变基因序列中引入特定的PCR引物序列，利用PCR方法直接检测转化细胞的DNA结构，然后根据特定扩增DNA片段的有无来鉴别中靶阳性细胞克隆。

正负筛选法 (PNS法) 筛选已发生同源重组的细胞



正向选择基因 *neo* 通常被插入载体靶 DNA 功能最关键的外显子中，或通过同源重组法置换靶基因的功能区。

负向选择基因 *HSV-tk* 则被置于目的片段外侧，含有该基因的重组细胞不能在选择培养基上生长。如果细胞中发生了随机重组，负向选择基因就可能被整合到基因组中，导致细胞死亡。

基因敲除的原理和方法

1. 利用基因同源重组进行基因敲除

- 利用同源重组构建基因敲除动物模型的基本步骤
- 条件性基因敲除法
- 诱导性基因敲除法

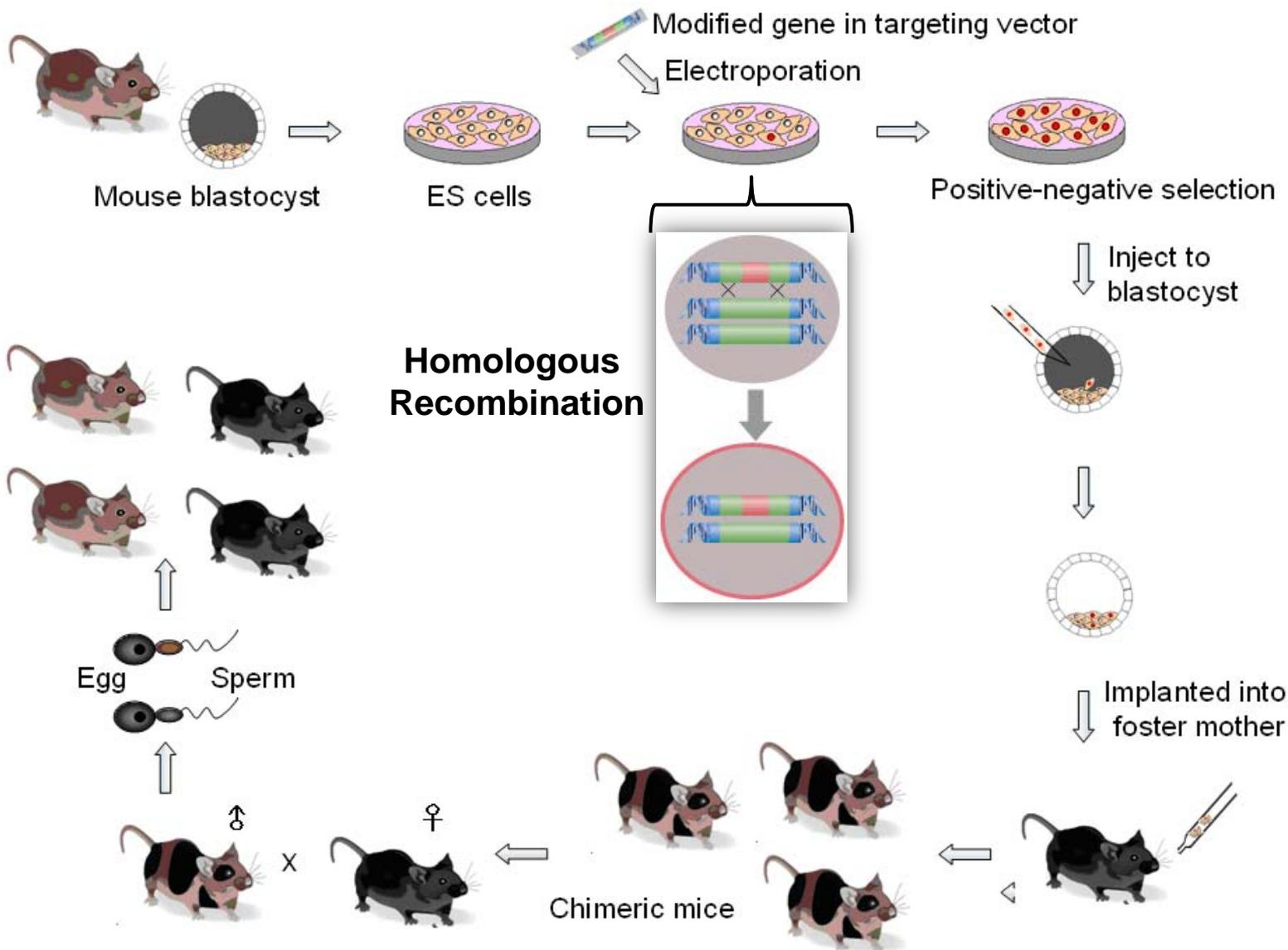
2. 利用随机插入突变进行基因敲除

基因捕获法

3. 核酸酶介导的基因编辑

- 锌指核酸酶（ZFNs）
- TALENs
- CRISPR/Cas9

基因同源重组法敲除靶基因的基本步骤

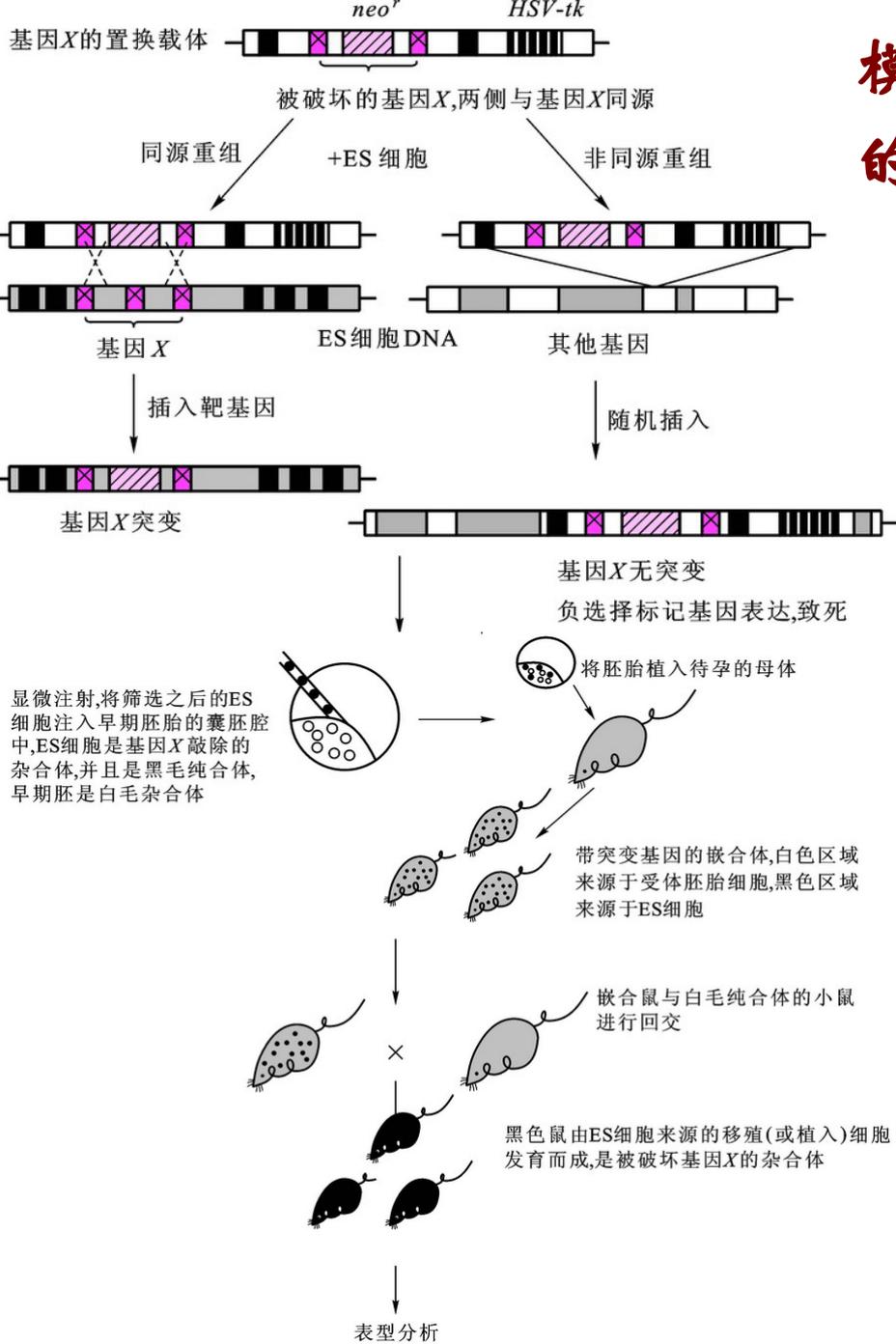


模式动物小鼠中完全基因敲除的主要技术策略与应用

显微注射命中率较高，技术难度相对大些。

电穿孔法命中率比显微注射低，操作使用方便。

胚胎干细胞（ES细胞）分离和体外培养的成功奠定了哺乳动物基因敲除的技术基础。



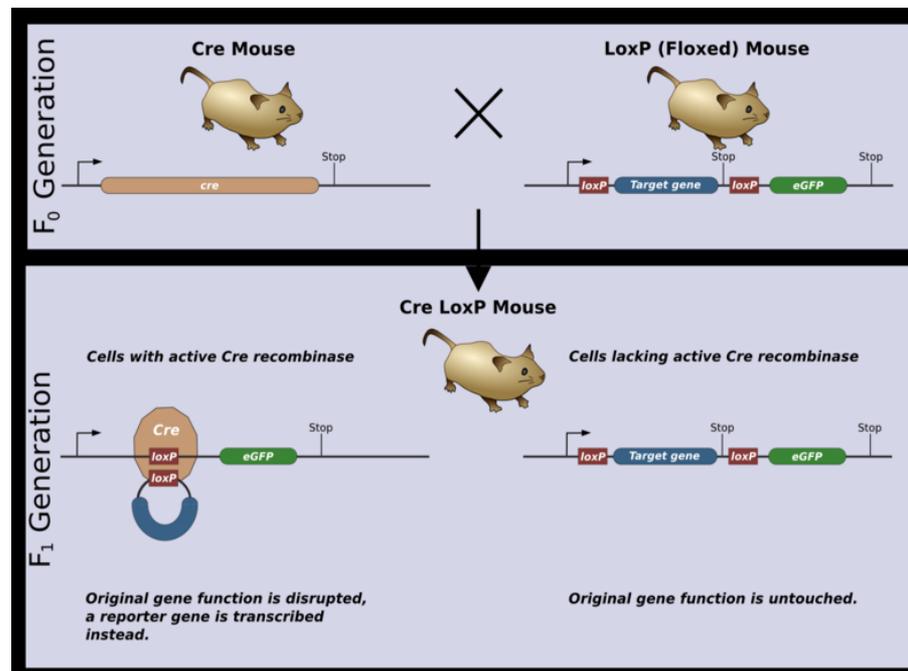
条件性基因敲除法

条件性基因打靶(**conditional gene targeting**)可定义为将某个基因的修饰限制于小鼠某些特定类型的细胞或发育的某一特定阶段的一种特殊的基因打靶方法。

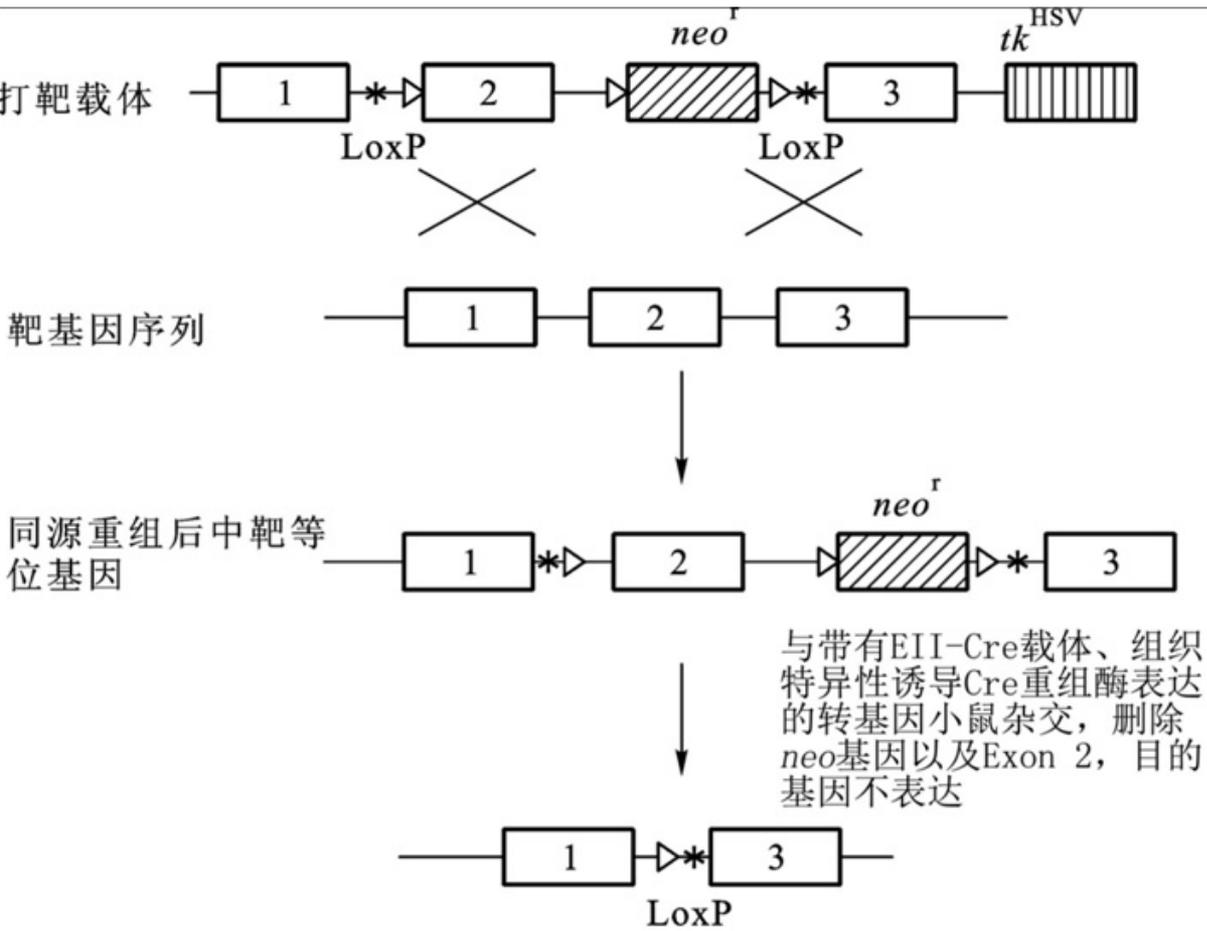
它实际上是在常规的基因打靶的基础上,利用重组酶介导的位点特异性重组技术,在对小鼠基因修饰的时空范围上设置一个可调控的“**按钮**”,从而使对小鼠基因组的修饰的范围和时间处于一种**可控状态**。

四种条件型定位重组系统:

- 噬菌体的**Cre/Loxp**系统 (广泛)
- Gin/Gix**系统
- 酵母细胞的**FLP/FRT**系统
- 酵母细胞的**R/RS**系统



条件性基因敲除 (Conditional Knockout) 策略



Cre的表达特性决定了靶基因的修饰(切除)特性

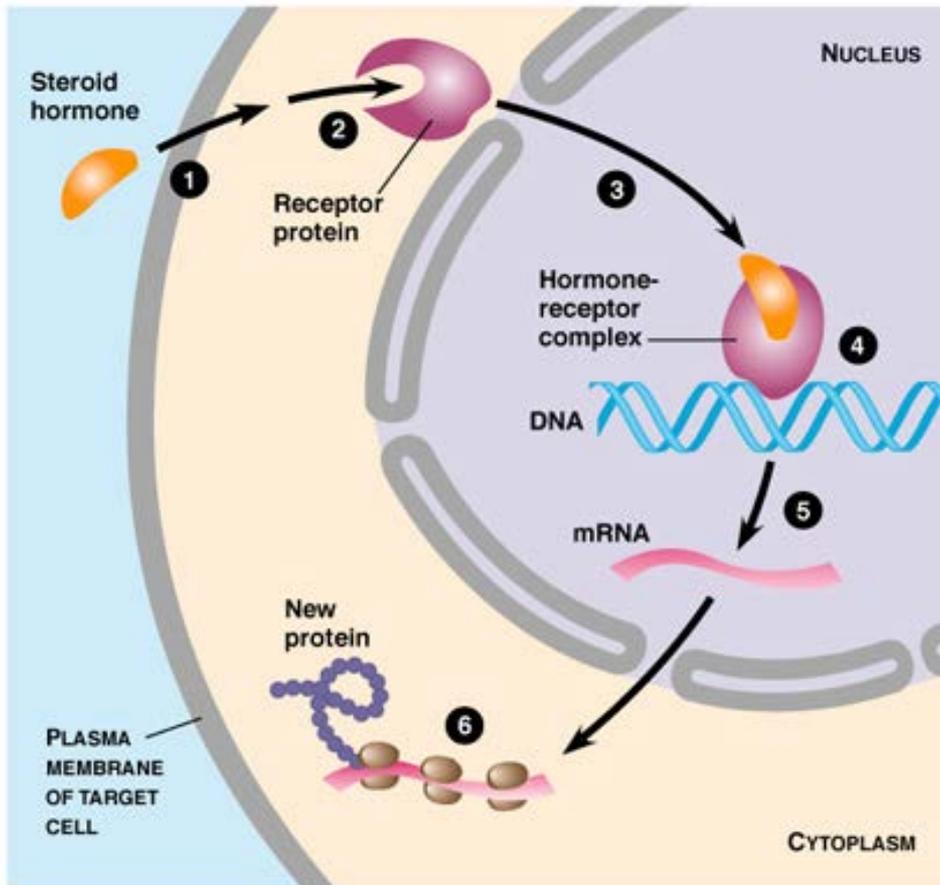
- 在基因组的靶位点上装上两个同向排列的1oxP, 并以此两侧装接上loxP的(“loxP floxed”)ES 细胞产生“loxP floxed”小鼠(表型仍同野生型的一样);

- 通过将“loxP floxed”小鼠与Cre转基因鼠杂交, Cre基因表达产生的Cre重组酶就会介导靶基因两侧的1oxP间发生切除反应, 将一个loxP和靶基因切除, 产生靶基因发生特定方式(如特定的组织特异性)修饰的条件性突变小鼠

诱导性基因敲除

- 是以**Cre/loxP** 系统为基础，但却是利用控制**Cre** 表达的启动子的活性或所表达的**Cre** 酶活性具有可诱导的特点，通过对诱导剂给予时间的控制，从而在**loxP** 动物的一定发育阶段和一定组织细胞中实现对特定基因进行遗传修饰之目的的基因敲除技术。
- 人们可以通过对诱导剂给予时间的预先设计的方式来对动物基因突变的时空特异性进行人为控制、以避免出现死胎或动物出生后不久即死亡的现象。
- 常见的几种诱导性类型如下：
四环素诱导型；干扰素诱导型；激素诱导型；腺病毒介导型。

诱导性基因敲除 (inducible Knockout)



Steroid Hormone Action (类固醇激素反应)

Inducible Cre Recombinase
-----CreERT2



ERT2: Mutated estrogen binding domain

Steroid: 4-OH tamoxifen

诱导性基因敲除优点

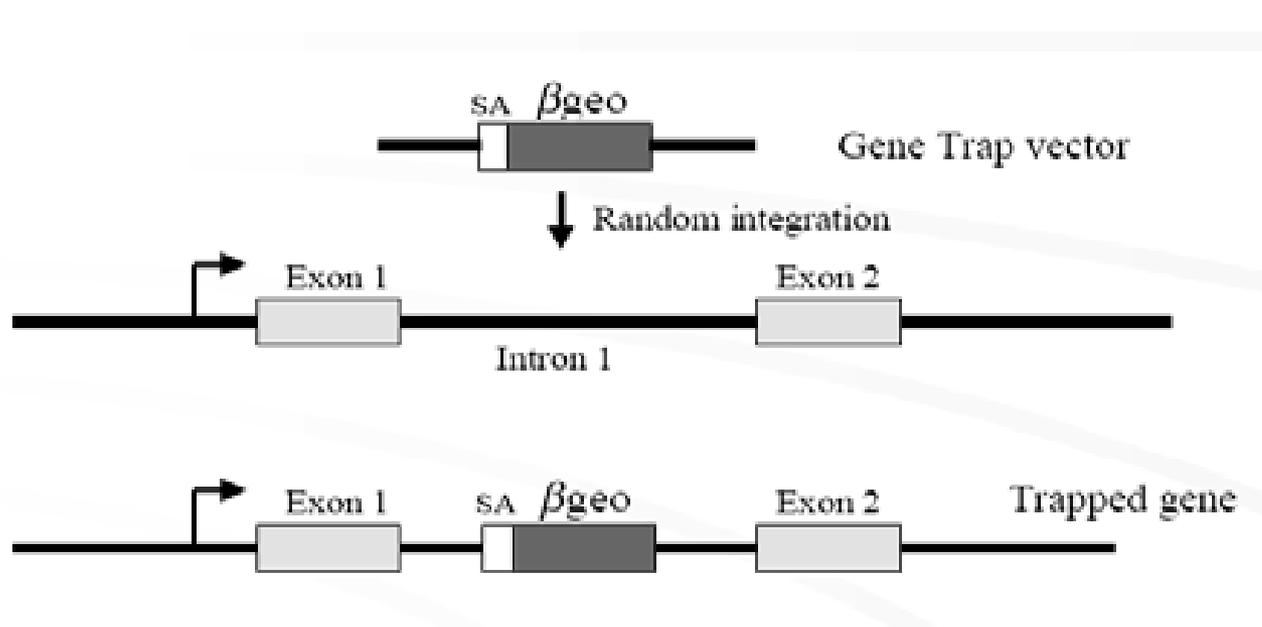
- ① 诱导基因突变的时间可人为控制；
- ② 可避免因基因突变而致死胎的问题
- ③ 在2个loxP位点之间的重组率较高；
- ④ 如用病毒或配体/DNA复合物等基因转移系统来介导Cre的表达，则可省去建立携带Cre的转基因动物的过程。

利用随机插入突变进行基因敲除 ---基因捕获技术

原理:

利用某些能随机插入基因序列的病毒，细菌或其他基因载体，在目标细胞基因组中进行随机插入突变，建立一个携带随机插入突变的细胞库，然后通过相应的标记进行筛选获得相应的基因敲除细胞。

基因捕获法



- 基因捕获载体还包括一个无启动子的报道基因，通常是**neo**基因，**neo**基因插入到**ES**细胞染色体组中，并利用捕获基因的转录调控元件实现表达的**ES**克隆可以很容易地在含**G418**的选择培养基中筛选出来，从理论上讲，在选择培养基中存活的克隆应该**100%**地含有中靶基因。
- 中靶基因的信息可以通过筛选标记基因侧翼**cDNA**或染色体组序列分析来获得。

基因捕获法的优缺点

优点：

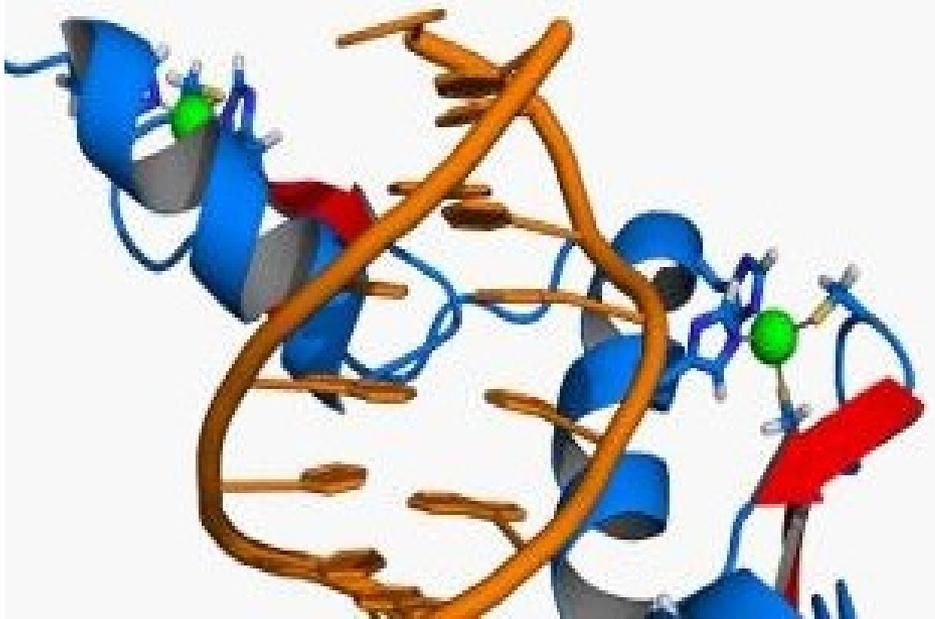
利用基因捕获可以建立一个携带随机插入突变的**ES** 细胞库，节省大量筛选染色体组文库以及构建特异打靶载体的工作及费用，更有效和更迅速地进行小鼠染色体组的功能分析。

缺点：

- 只能剔除在**Es** 细胞中表达的基因。单种的细胞类型中表达的基因数目约为 10^4 ，现在的基因捕获载体从理论上来讲应能剔除所有在**ES**细胞表达的基因，因此，在**ES** 细胞中进行基因捕获还是大有可为的。
- 用基因捕获法进行基因剔除的另一个缺点是无法对基因进行精细的遗传修饰。

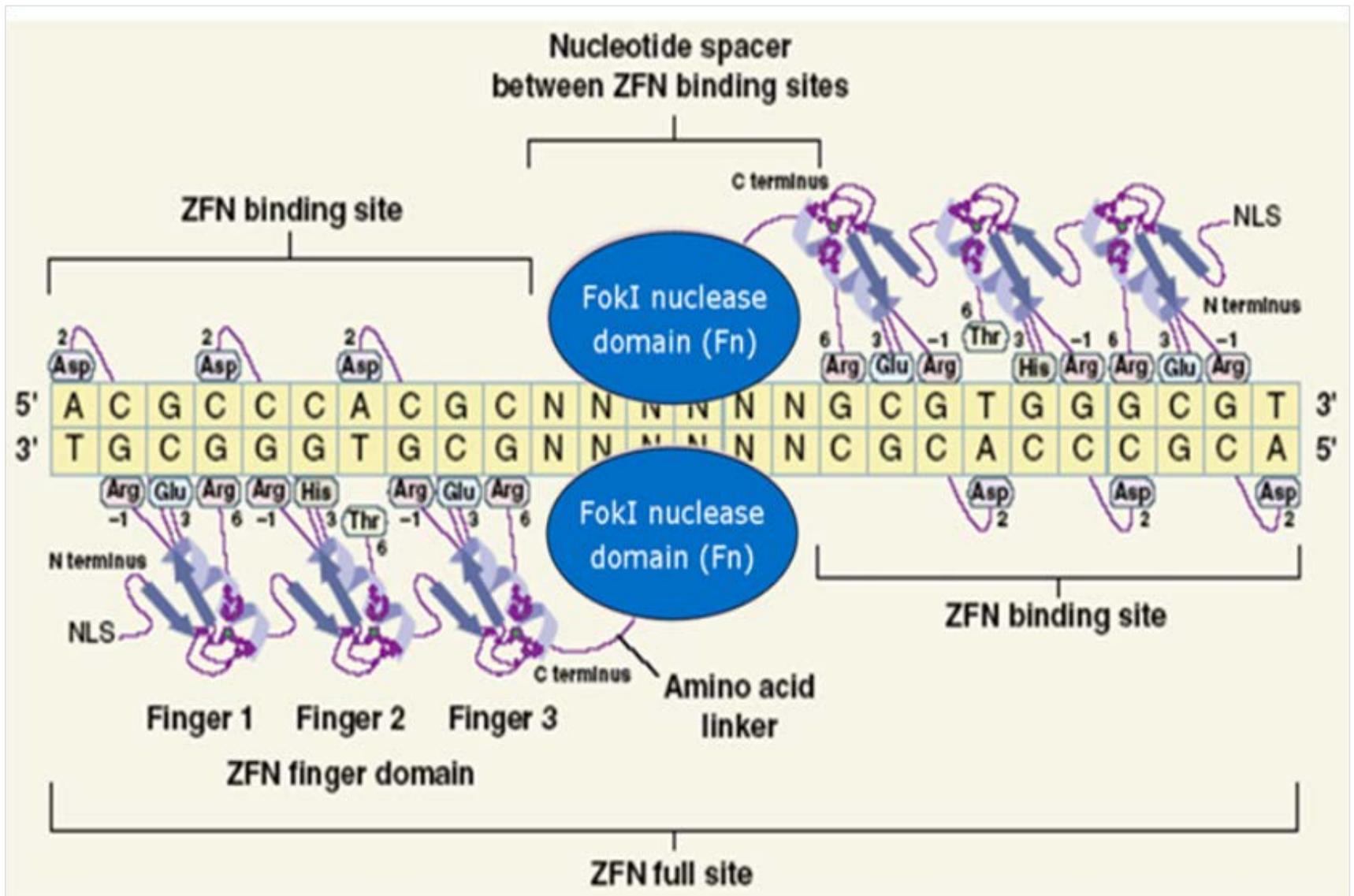
锌指核酸酶

(zinc finger nuclease, ZFNs)



结合到DNA(橙色)上的锌指核酸酶(蓝色)，图片来自维基共享资源，Thomas Splettstoesser

- 又名锌指蛋白核酸酶，是一种人工改造的核酸内切酶，由一个**DNA识别域**和一个**非特异性核酸内切酶**构成，其中DNA识别域赋予特异性，在DNA特定位点结合，而非特异性核酸内切酶赋予剪切功能，**两者结合就可在DNA特定位点进行定点断裂。**



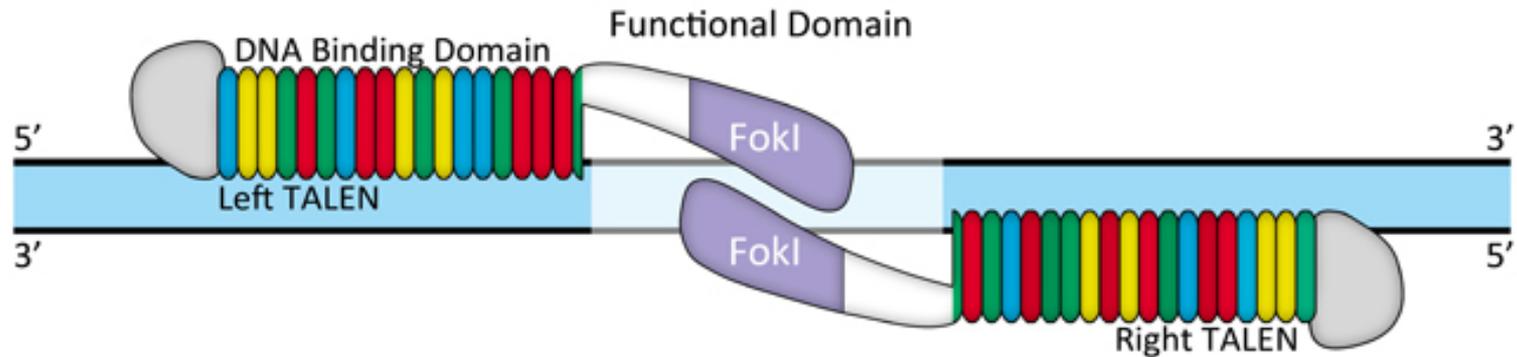
利用ZFNs进行基因打靶的实验步骤

- (1) 确定ZFN靶位点的DNA序列
- (2) 设计和选择可识别靶位点的锌指蛋白(ZFPs)
- (3) 利用已选择和设计的锌指蛋白组装ZFNs
- (4) 单独将ZFNs转入正常细胞诱导目标基因发生DSB, 则可激活NHEJ修复机制, 或许将ZFNs和与目标基因同源的供体DNA一同转入正常细胞, 则可以激活HR介导的DSB修复机制, 并产生出变异群体
- (5) 检测与研究供体DNA模板在特异位点所引起的基因变异与基因修复情况
- (6) 还需要通过Southern杂交来检测供体DNA序列并没有整合到细胞基因组的其他位点

ZFNs的优缺点

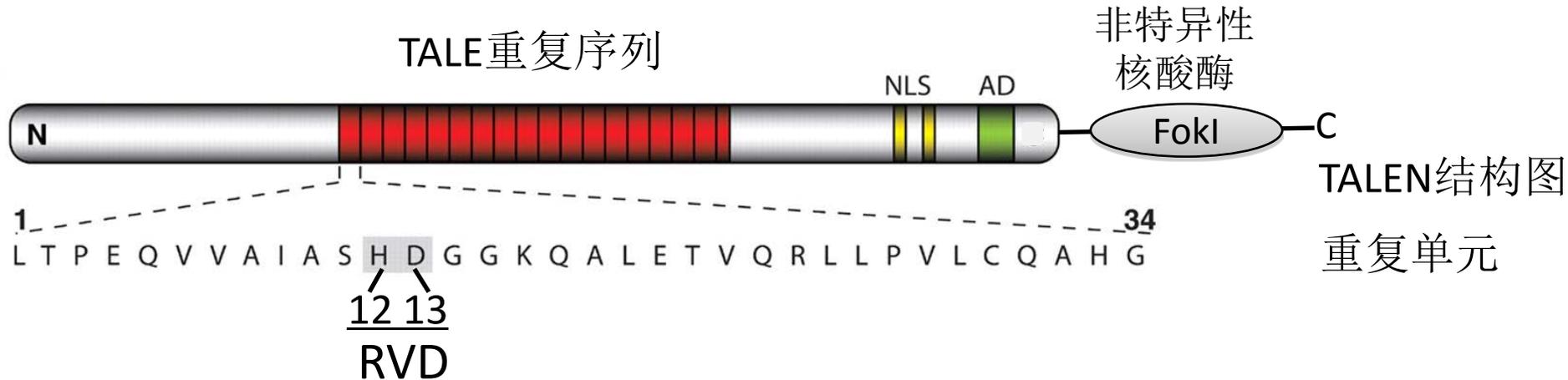
- 由于**ZFNs**的合成组装技术难度大，一般实验室难以实施；
- **ZFNs**易于对基因组进行非特异性切割，或对靶点**DNA**切割效率低，一直限制其进入实际应用。

TALEN介导的基因组定点修饰技术



TALENs是一种可靶向特异DNA序列的酶，它借助于TAL效应子，一种由植物细菌分泌的天然蛋白，来识别特异性DNA碱基对。TAL效应子可被设计识别和结合所有的目的DNA序列。对TAL效应子附加一个核酸酶就生成了TALENs。TAL效应核酸酶可与DNA结合并在特异位点对DNA链进行切割，从而导入新的遗传物质。

转录激活因子样效应物核酸酶的结构 (TALENs)



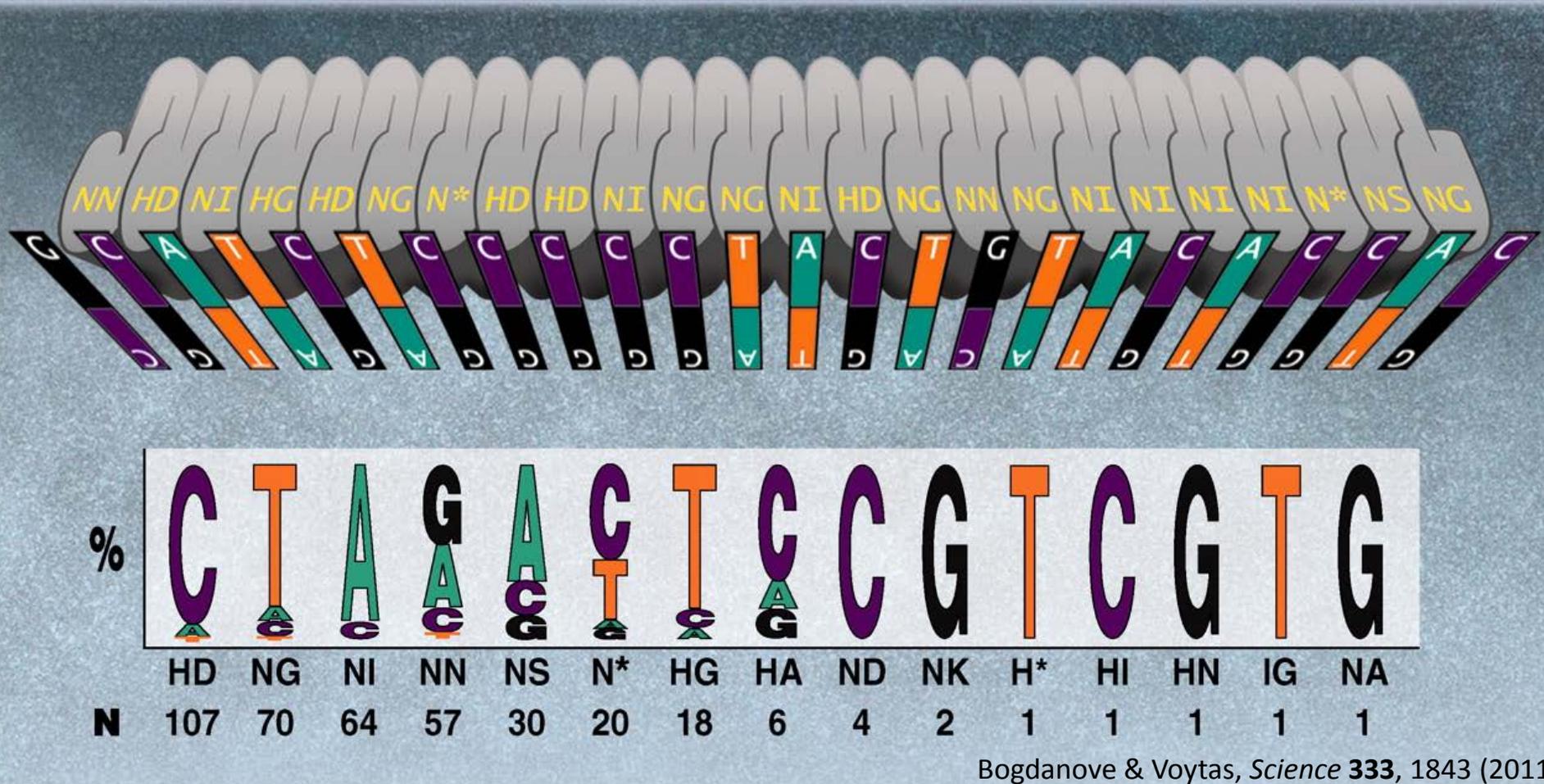
HD: C NG: T NI: A NN: G/A

RVD和
对应碱基类型

TALE 的重复单元跟靶序列具有很好的一一对应性

- **RVD: Repeat variable di-residue /重复可变双残基**
- **NLS: Nuclear localization signal /核定位信号**
- **AD: Activation domain /转录激活结构域**

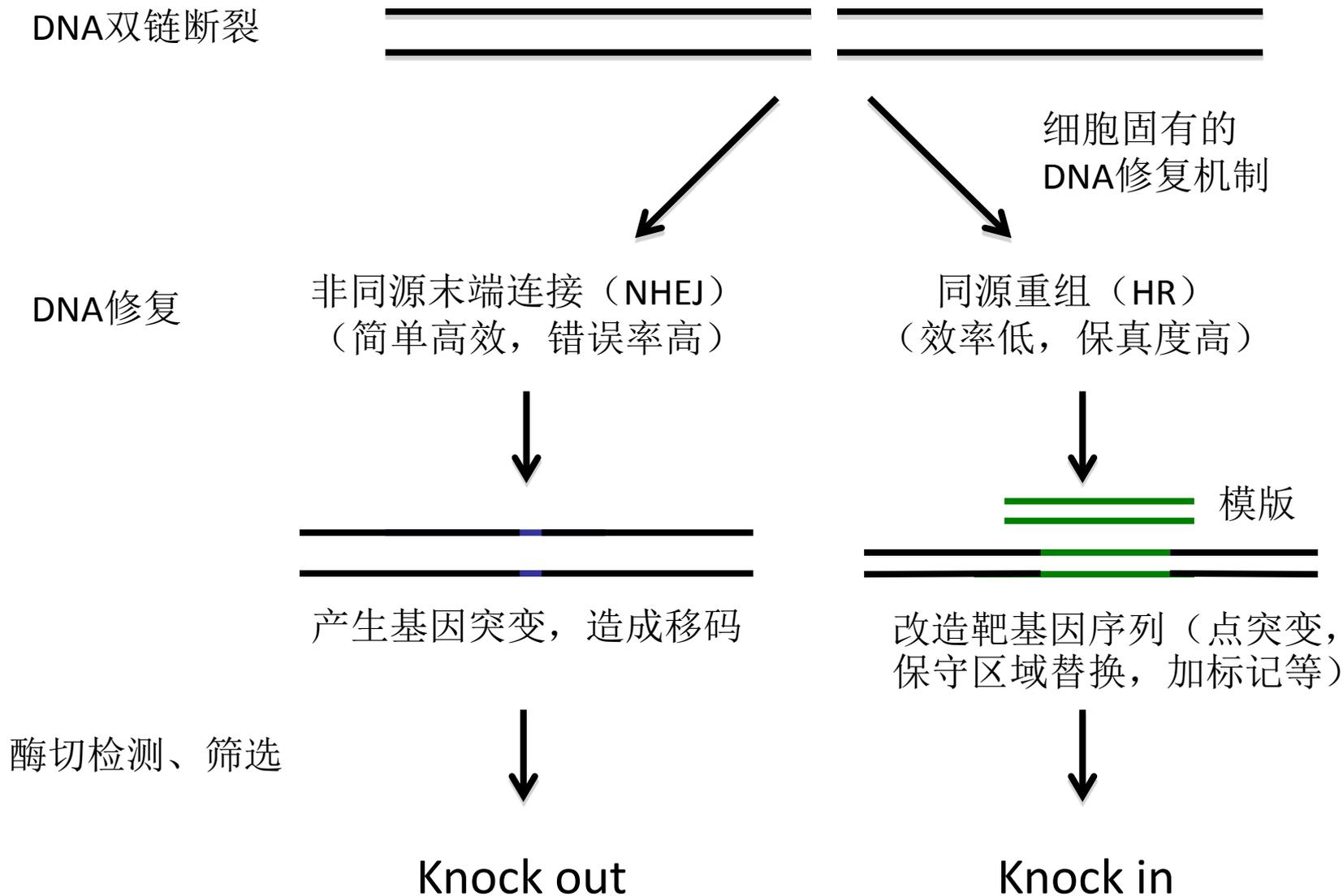
TALE识别DNA序列示意图



Bogdanove & Voytas, *Science* **333**, 1843 (2011)

- TALEN 靶位点的挑选原则：
TALEN 靶位点5'端的前一位(第0位) 碱基应为胸腺嘧啶(T)
- 设计TALEN靶点网站: <https://tale-nt.cac.cornell.edu/>

TALEN介导的基因组定点修饰流程



TALEN的优缺点

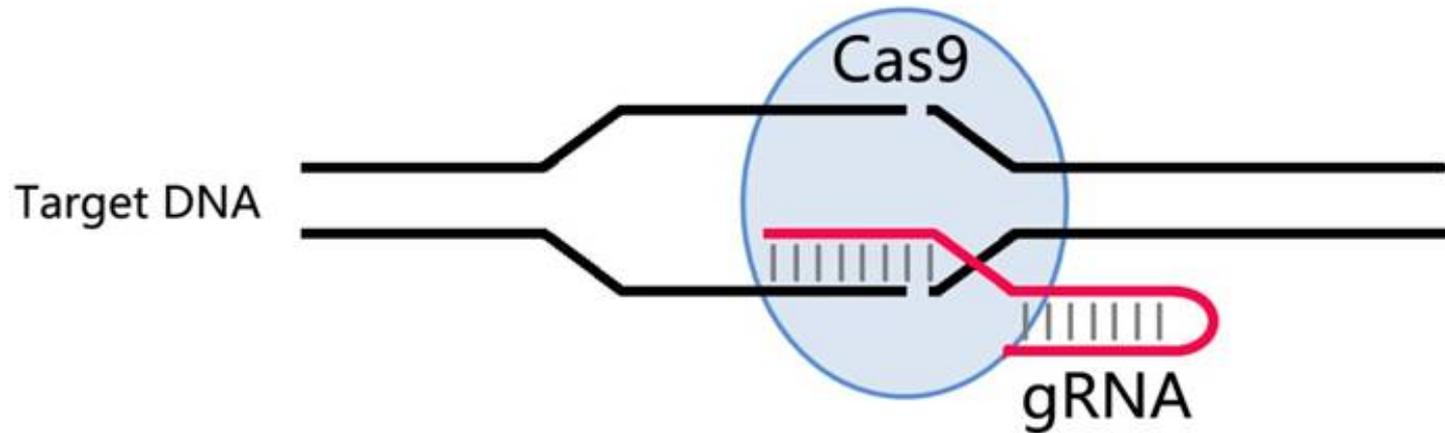
优点:

- 毒性较低，靶向任意基因，无细胞类型限制；
- 靶点在基因组上序列高度特异，脱靶情况少；
- 成对的TALE识别模块保证了基因敲除靶点的准确性；
- 可应用于基因敲除、基因敲入、基因修饰、基因激活或抑制等等；
- 灵活的结合域和功能域设计，可选择TALEN，TALE-TF或其他TAL效应子。

缺点:

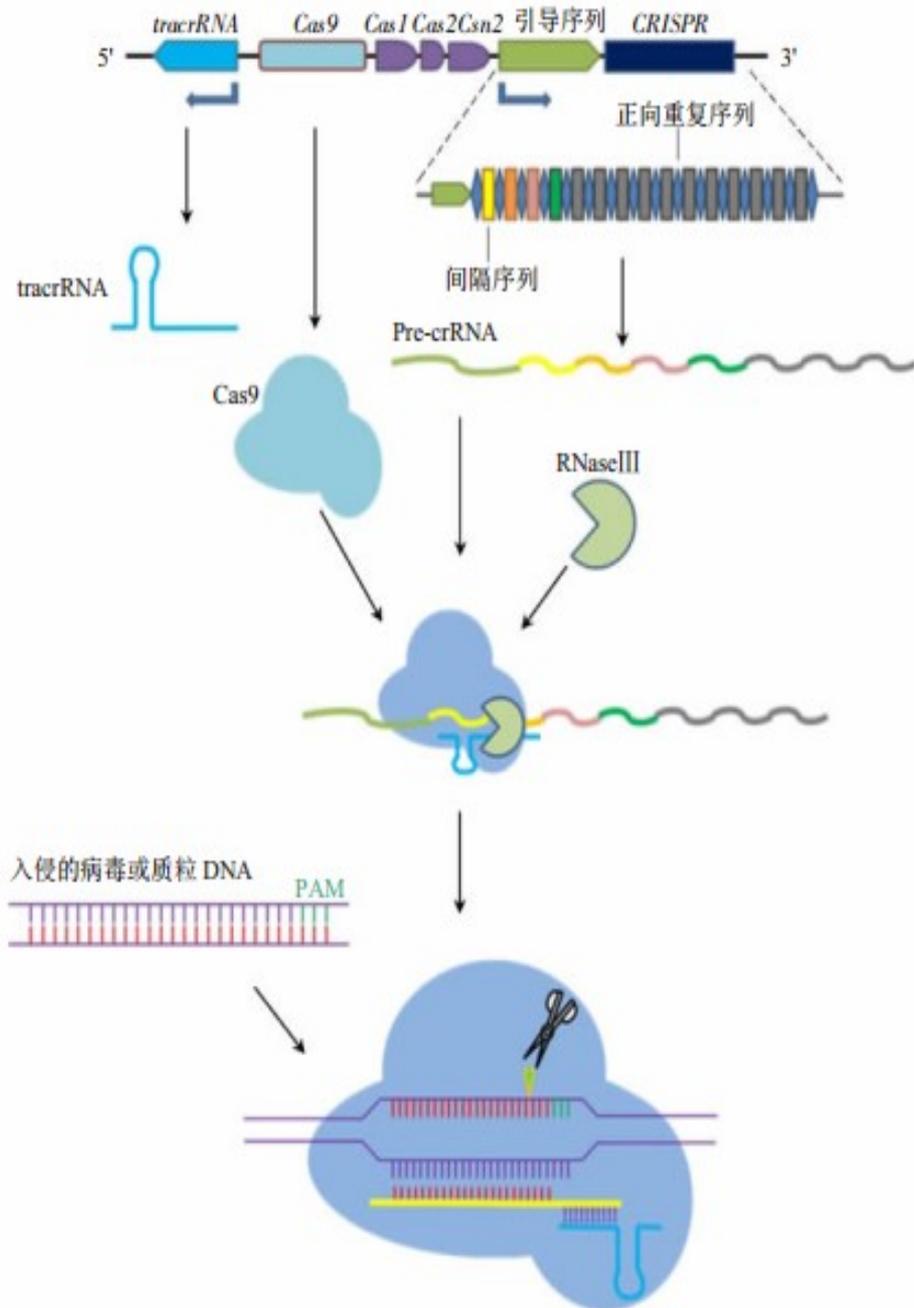
模块组装过程繁琐。

CRISPR-Cas9 基因敲除技术



2013年1月份，美国两个实验室发表了基于 **CRISPR-Cas9** 技术在细胞系中进行基因敲除的新方法，一种全新的人工核酸内切酶 **clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated (Cas)9** 出现，它主要是基于细菌的一种由 **RNA** 介导的可遗传的获得性免疫系统改造而成，其特点是制作简单、成本低、作用高效。

CRISPR-Cas9 基因敲除技术原理



crRNA (CRISPR-derived RNA) 通过碱基配对与 *tracrRNA* (trans-activating RNA) 结合形成 *tracrRNA*/crRNA 复合物，此复合物引导核酸酶 **Cas9** 蛋白在与 crRNA 配对的序列靶位点剪切双链 DNA。而通过人工设计这两种 RNA，可以改造形成具有引导作用的sgRNA (short guide RNA)，以引导 **Cas9** 对 DNA 的定点切割。

CRISPR/Cas9的特点和优势

1、操作简单，靶向精确性更高。

sgRNA靶向序列和基因组序列必须完全匹配，Cas9才会对DNA进行剪切。编码sgRNA的序列不超过100bp，因此比构建TALENs和ZFNs更简单方便，用于CRISPR的sgRNA识别序列仅需20个核苷酸。

2、CRISPR/Cas9系统是由RNA调控的对DNA的修饰，其基因修饰可遗传。

3、基因修饰率高，基因调控方式多样，例如敲除、插入、抑制、激活等

4、可实现对靶基因多个位点同时敲除。

5、无物种限制，已被应用于多种生物，包括人、小鼠、大鼠、斑马鱼、秀丽隐杆线虫、植物及细菌。

6、实验周期短，最快仅需2个月，节省大量时间和成本。

TALEN 及 CRISPR-Cas9 体系应用对比

性能	TALEN	CRISPR-Cas9
识别类型	蛋白质-DNA	RNA-DNA
甲基化敏感性	敏感	不敏感
染色质结构敏感性	敏感	敏感
脱靶效应	较少观察到脱靶效应	潜在脱靶效应高于 TALENs 及 ZFNs
多靶点	较少使用	可用

Nature: 基因组编辑新技术或超越CRISPR

能够将缺陷基因的功能性拷贝插入到患者的基因组中，对于许多从事遗传疾病治疗的临床医生来说是一个诱人的目标。现在，斯坦福大学医学院的研究人员设计出了一种新方法来实现这种遗传技能。这一研究发现有可能给称作为**CRISPR/Cas9**的基因组技术提供了一种替代选择。

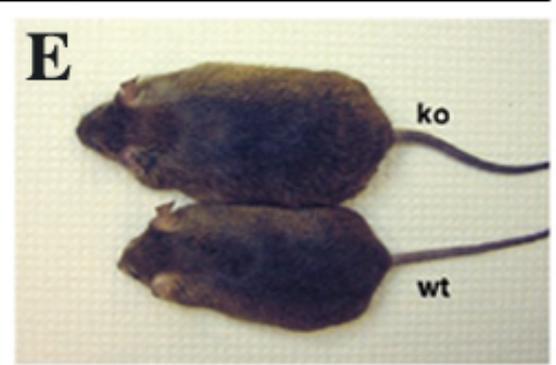
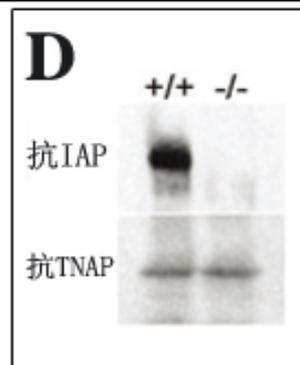
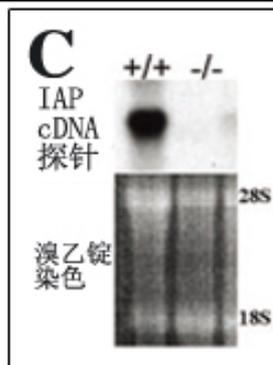
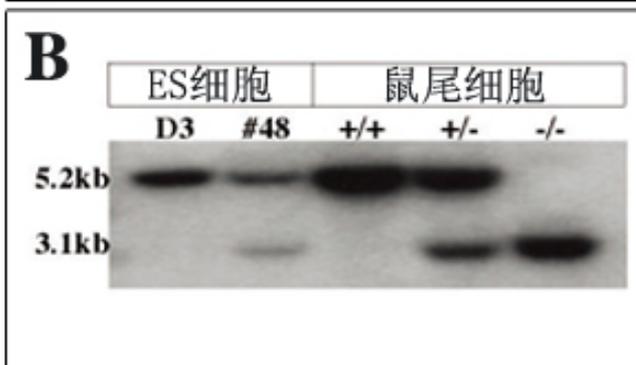
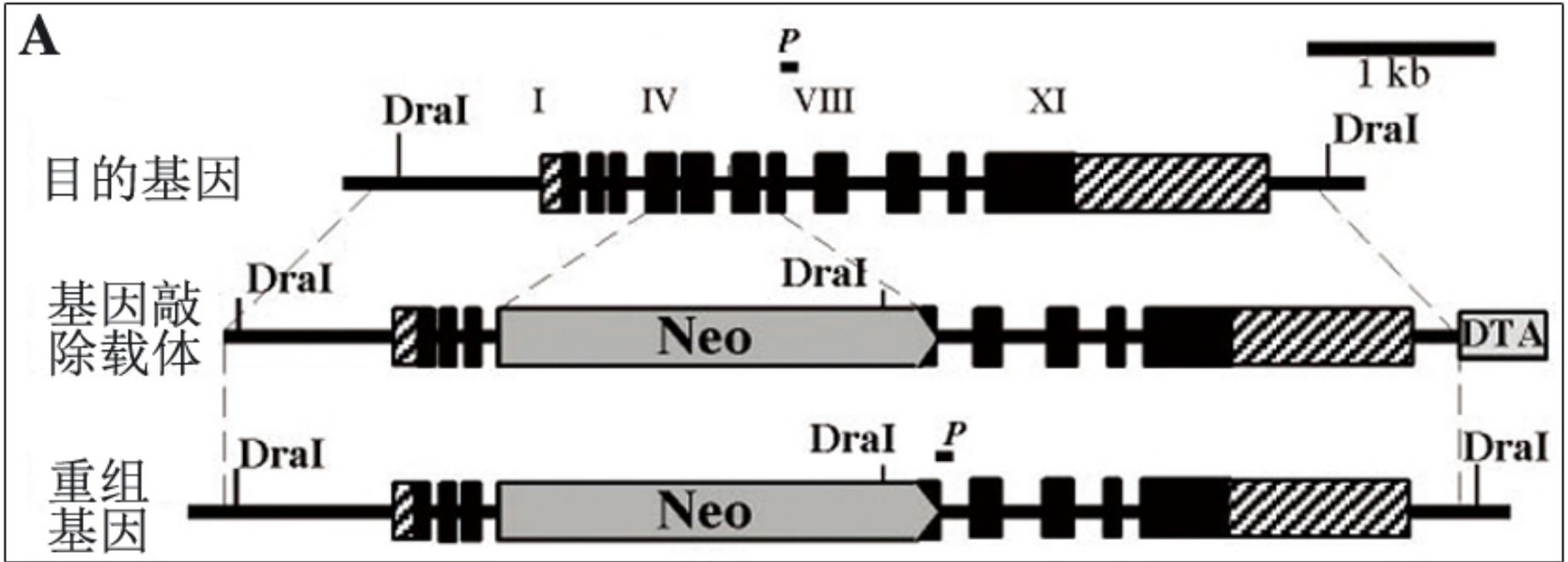
这种方法不同于其他一些广受欢迎的技术，它**不需要共传递内切核酸酶在特异位置剪切受体的DNA**。它也不依赖于共同插入称作为启动子的遗传“**on**”开关来激活新基因表达。

这些差异有可能使得这一新技术更为安全及持久。利用这一技术，斯坦福大学的研究人员通过插入血友病小鼠缺失的一个凝血因子的编码基因治愈了这些动物。

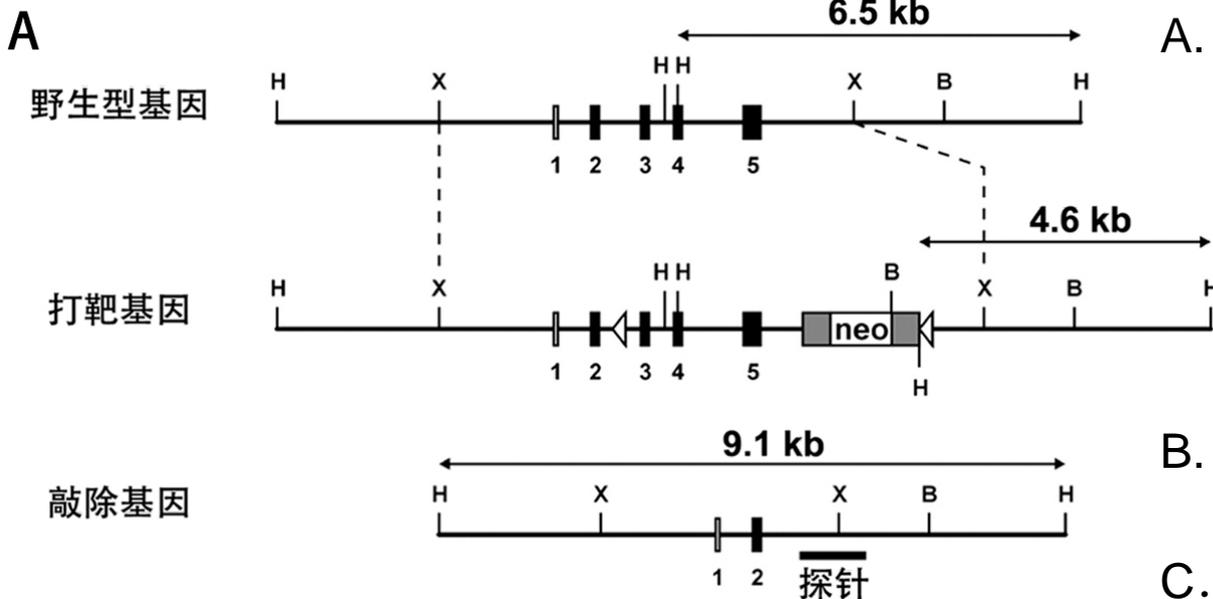


Promoterless gene targeting without nucleases ameliorates haemophilia B in mice.
Nature. 2014, Oct 29. [doi:10.1038/nature13864](https://doi.org/10.1038/nature13864)

肠碱性磷酸酶 (Intestinal Alkaline Phosphatase, IAP) 基因敲除加速小鼠肥胖



肝组织特异性表达Cre-LoxP系统条件型敲除小鼠40S核糖体蛋白S6基因导致肝细胞分裂增殖受阻

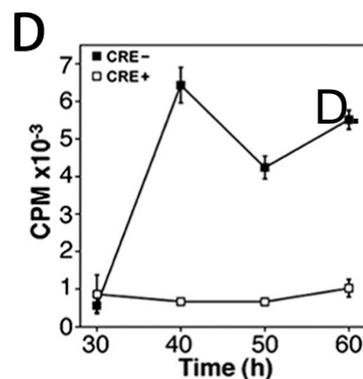
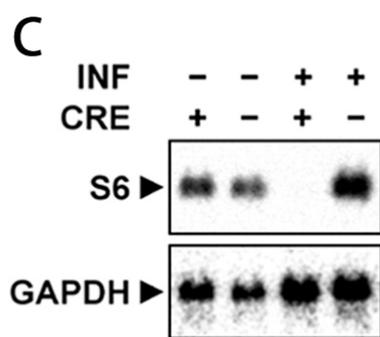
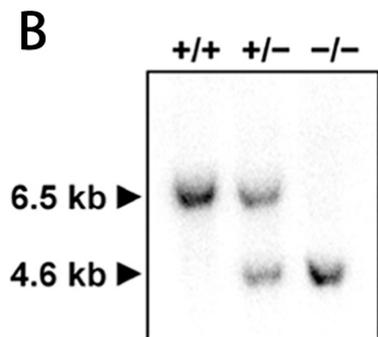


A. 构建带有S6基因的LoxP打靶载体的ES细胞，经过杂交筛选，获得纯合小鼠，与带有肝组织特异性、受INF- α 诱导的Mx-Cre转基因小鼠杂交，删除neo和外显子3-5，得到肝组织特异性敲除S6基因的小鼠。

B. Southern杂交表明，纯合体小鼠内有DNA片段缺失。

C. Northern杂交表明，INF- α 诱导后Cre小鼠体内S6基因不表达。

D. [^3H] 胸腺嘧啶能被S6基因未敲除小鼠吸收，却不能掺入S6基因敲除小鼠肝组织，表明细胞不能进入S期，不能正常分裂增殖。



基因敲除技术的应用、前景

- ① 建立生物模型。
- ② 疾病的分子机理研究和疾病的基因治疗。
- ③ 提供廉价的异种移植器官。
- ④ 免疫学中的应用。
- ⑤ 改造生物、培育新的生物品种。

基因敲除技术缺陷

基因敲除技术始终存在着一个难以克服的缺点，即敲掉一个基因并不一定就能获知该基因的功能，其原因包括：

- 许多基因在功能上是冗余的，敲掉一个在功能上冗余的基因，并不能造成容易识别的表型，因为基因家族的其他成员可以提供同样的功能；
- 对于某些必需基因，敲除后会造成细胞的致死性，也就无法对这些必需基因进行相应的研究了。

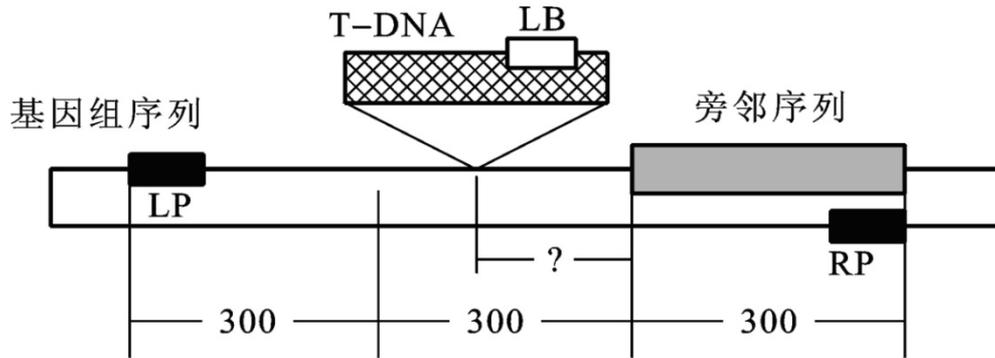
植物基因敲除技术

T-DNA插入失活技术是目前在植物中使用最为广泛的基因敲除手段。

利用根癌农杆菌**T-DNA**介导转化，将带有报告基因的**DNA**序列整合到基因组**DNA**上，如果这段**DNA**插入到目的基因内部或附近，就会影响该基因的表达，从而使该基因“失活”。

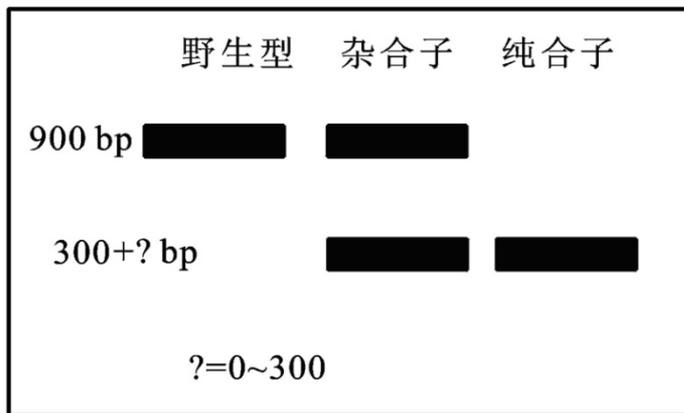
植物基因敲除及突变体筛选

A



a. 引物设计。LP和RP分别代表插入基因两端的引物，LB是指载体上的引物，目的基因片段900bp。

B



b, PCR产物电泳结果。分别代表野生型、杂合子和纯合子PCR条带。

预祝大家期中考试
取得好成绩！



Single-Stranded DNA
to be Sequenced



Add:

DNA Polymerase I

dATP

dGTP

dCTP

dTTP

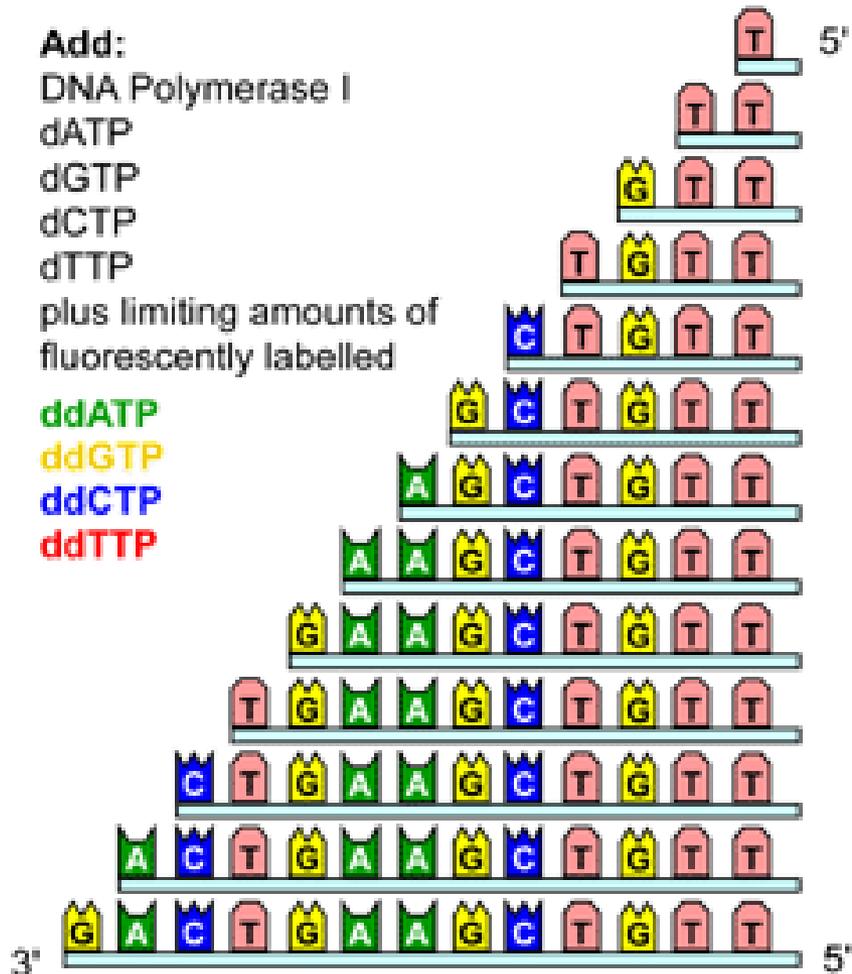
plus limiting amounts of
fluorescently labelled

ddATP

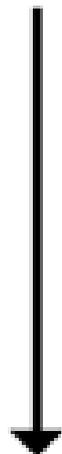
ddGTP

ddCTP

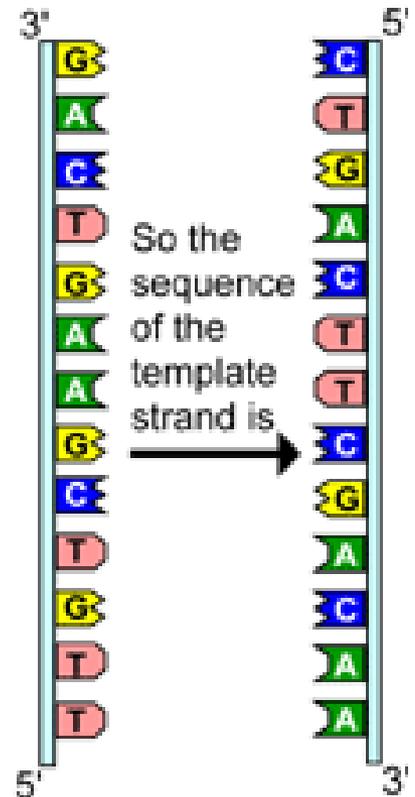
ddTTP

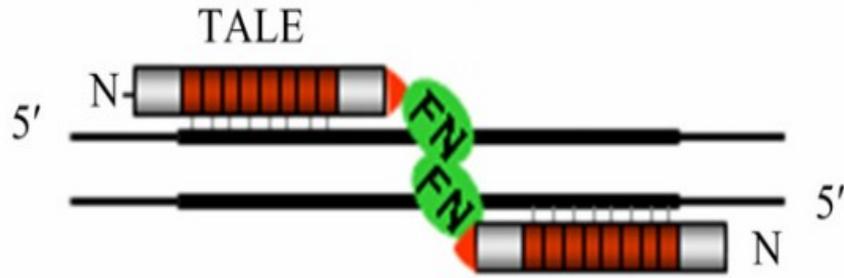


Larger
fragments



Smaller
fragments





↓ TALEN pair induces DNA breakage



34 aa repeats



LTPEQVVVAIASNGGGKQALETVQRLLPVLCQAHG

TAL模块及对应碱基

- NG = T
- HD = C
- NI = A
- NN = G/A