

# 第一章

## 基因、基因组和基因组学

*Gene, Genome and Genomics*

杜芳



# Activities of Genes

基因——遗传的基本单位，是负载特定生物遗传信息的DNA分子片段，一定条件下表达遗传信息产生特定的生理功能。

## How do genes work?

- ❖ First, they are replicated faithfully;
- ❖ second, they direct the production of RNAs and proteins;
- ❖ third, they accumulate mutations and so allow evolution.



# 第一节 基因 Gene

基因——遗传的基本单位，是负载特定生物遗传信息的DNA分子片段，一定条件下表达遗传信息产生特定的生理功能。

## 一、基因概念的发展

1865 孟德尔的豌豆实验

- 发现了遗传规律  
同对因子分离  
异对因子自由组合
- 首次提出了“遗传因子”（即基因）





现代遗传学之父，奥地利生物学家格雷戈尔·孟德尔

一个遗传因子决定一个性状(1865年)



Friedeich Miescher

1868年，Miescher从脓细胞中分离出细胞核，加酸抽提得到了一种含氮和磷丰富的物质，称其为核素(nuclein). 1872年，他又在鲑鱼精子细胞核中发现了大量的这类物质。故称其为核酸(nucleic acid)。



# 第一节 基因 Gene

基因——遗传的基本单位，是负载特定生物遗传信息的DNA分子片段，一定条件下表达遗传信息产生特定的生理功能。

## 一、基因概念的发展



维尔赫姆·路德维希·约翰逊  
Wilhelm Ludwlg Johannsen,  
1857-1927

1909年Johannsen采用“gene”来描述遗传性状的决定因子，但当时并没有提出基因的定义，甚至反对“基因是具有形态特征的结构”。

谈家桢在留学美国期间应邀为国内科学杂志撰文介绍现代遗传学时把“gene”的汉译名定为“基因”。



# 第一节 基因 Gene

基因——遗传的基本单位，是负载特定生物遗传信息的DNA分子片段，一定条件下表达遗传信息产生特定的生理功能。

## 一、基因概念的发展

1915年至1928年  
摩尔根的果蝇实验



美国遗传学家摩尔根

1933年

诺贝尔生理医学奖

- 进一步证实了孟德尔因子和孟德尔定律
- 两大重要发现  
基因是在染色体上，  
**遗传的基因连锁和互换定律。**
- 建立了作为摩尔根理论主要基础的基因学说。



一个基因控制一个性状  
(1926年)，明确了基因存  
在于染色体上。





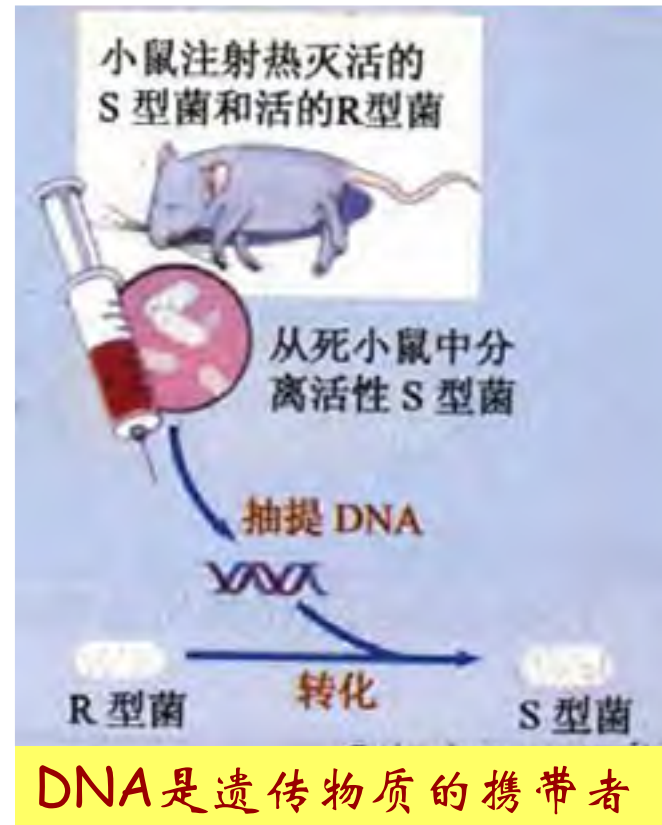
# 第一节 基因 Gene

基因——遗传的基本单位，是负载特定生物遗传信息的DNA分子片段，一定条件下表达遗传信息产生特定的生理功能。

## 一、基因概念的发展



1944年Avery等转化实验证明转移物的化学组成实际上就是脱氧核糖核酸 (DNA)





# 第一节 基因 Gene

基因——是一段有完整功能的DNA序列，可作为一个转录单位，其表达产物通常为1条多肽链或1个RNA分子。

## 一、基因概念的发展

"for their discoveries of split genes"



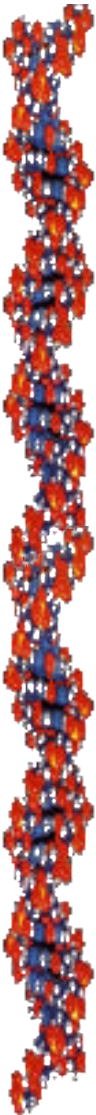
**Richard J. Roberts**

New England Biolabs  
Beverly, MA, USA



**Phillip A. Sharp**

Massachusetts Institute of Technology, Center  
for Cancer Research  
Cambridge, MA, USA



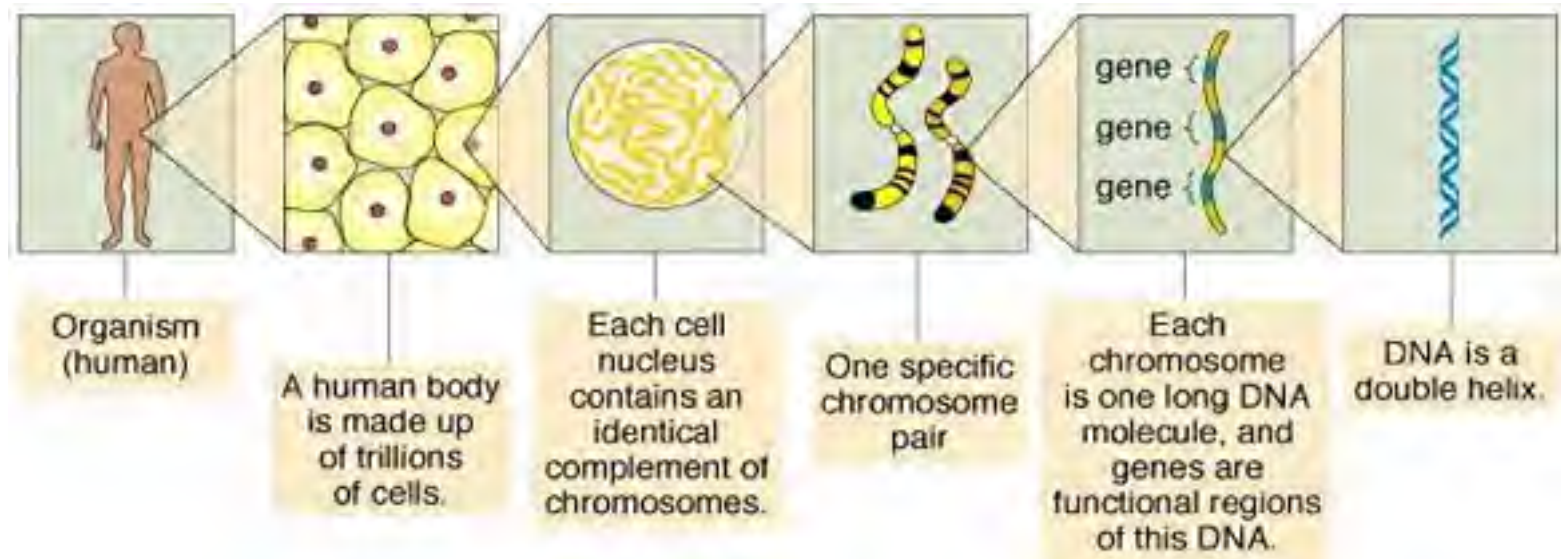


# 第一节 基因 Gene

基因——是一段有完整功能的DNA序列，可作为一个转录单位，其表达产物通常为1条多肽链或1个RNA分子。

## 一、基因概念的发展

基因——生命的最大奥秘



- 基因是所有生物体遗传信息的载体，基因的物质基础是DNA分子。
- 地球上人与人之间99.99%的基因序列相同，人与人之间的差异仅为万分之一。





# 第一节 基因 Gene

**基因**——是DNA分子中含有特定遗传信息的核苷酸序列，负责合成有功能的蛋白质或RNA所必需的DNA片段（部分RNA病毒例外），包括编码序列及调控序列，是遗传物质的最小功能单位。

## 二、 基因的现代概念

### (一) 基因的分类

**包括**

**结构基因：** 转录翻译为多肽链

**调控基因：** 调控结构基因表达的基因

**只转录而不翻译的基因：** 核糖体RNA基因、 小分子RNA



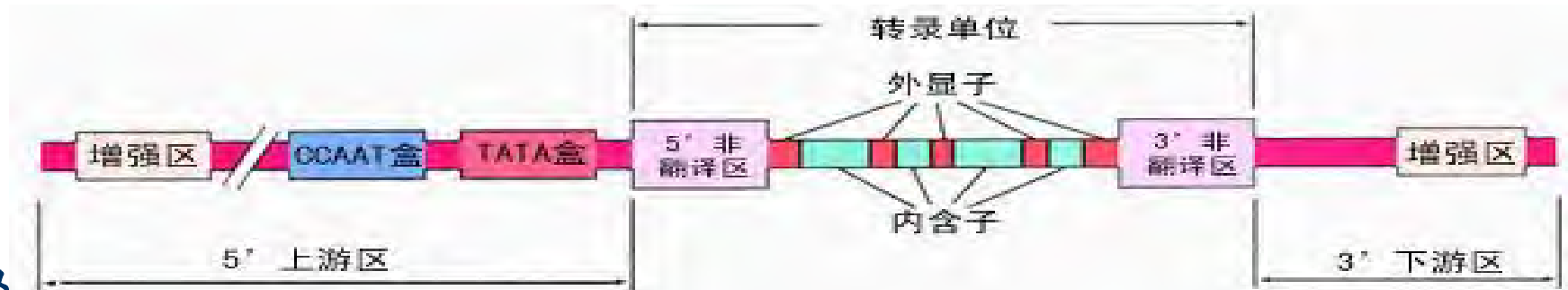
# 第一节 基因 Gene

基因——是DNA分子中含有特定遗传信息的核苷酸序列，负责合成有功能的蛋白质或RNA所必需的DNA片段（部分RNA病毒例外），包括编码序列及调控序列，是遗传物质的最小功能单位。

## 二、 基因的现代概念

### (二) 基因的结构

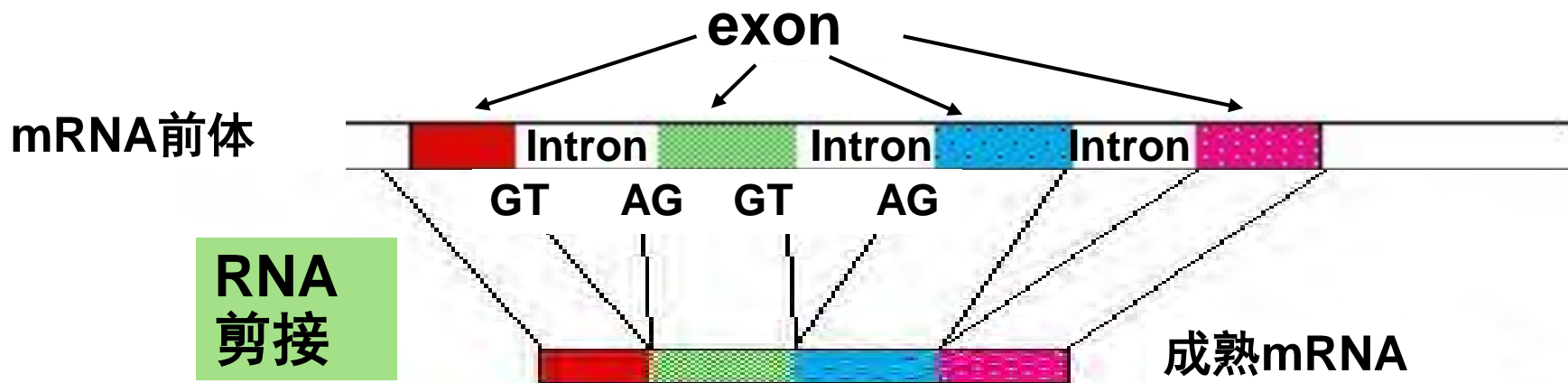
{	包括	编码序列：	外显子(exon)
		插入序列：	内含子(intron)
		侧翼序列：	含有调控序列





## 结构基因的结构特点

- ❖ 原核生物：连续
- ❖ 真核生物：断裂基因(由外显子和内含子组成)
- ❖ 病毒：连续 / 断裂



**GT-AG法则**:真核基因中RNA剪接的识别信号

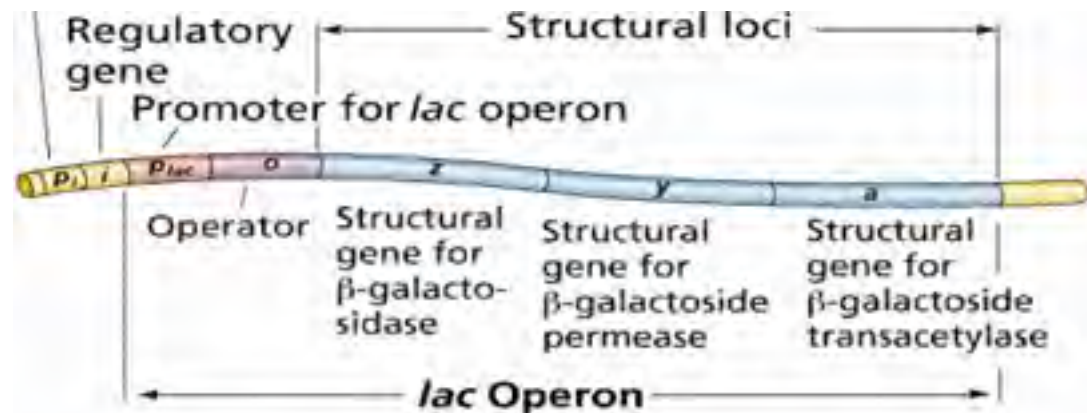
内含子的5'端以GT开始, 3'端以AG结束



## 调控基因的结构特点

### ❖ 原核生物基因的调控序列

- 启动子 (*promoter*)
- 终止子 (*terminator*)
- 操纵元件 (*operator*)



### ❖ 真核生物基因的调控序列

- 顺式作用元件 (*cis-acting element*) 影响基因表达的非编码的DNA序列。

启动子和上游启动元件: *promoter, upstream promoter element*

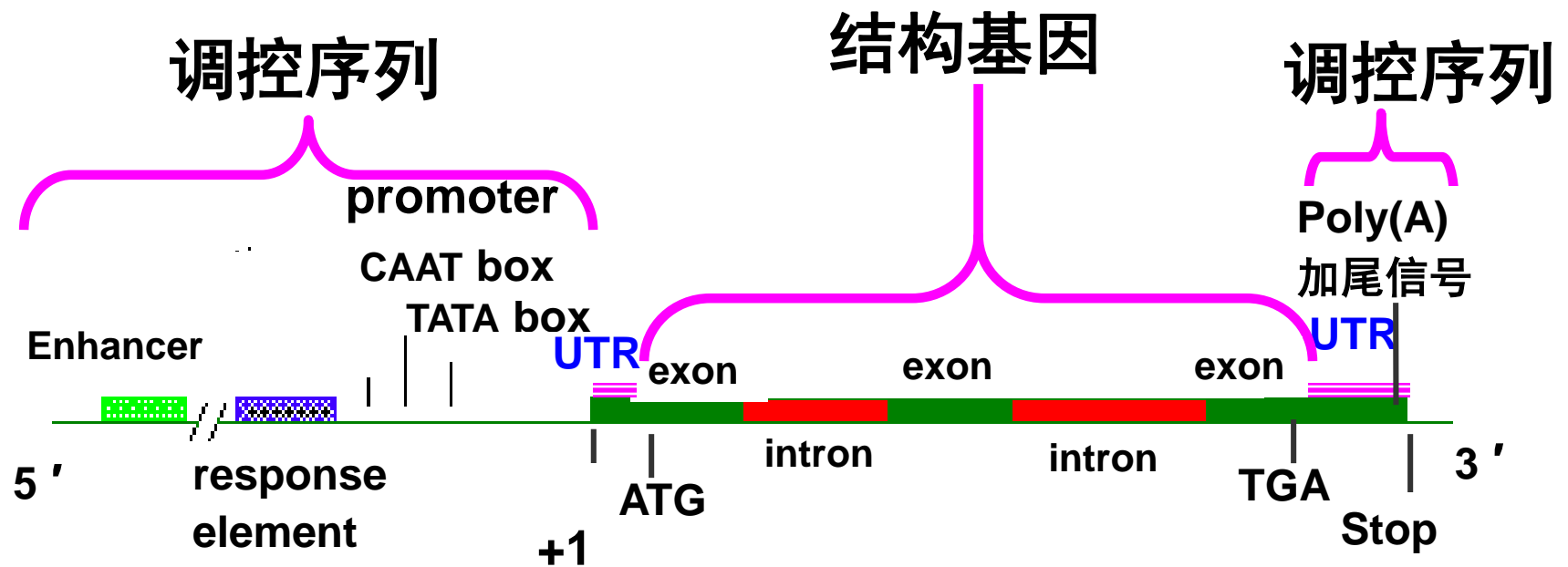
增强子和沉默子: *enhancer, silencer*

反应元件 *response element*

Poly(A)加尾信号



# 真核生物的基因结构



开放阅读框: open reading frame , ORF

非翻译区: untranslated regions , UTR





# 真核生物基因的调控序列

❖ **启动子和上游启动元件**：基因转录起始点上游100到200bp，是RNA聚合酶结合位点周围的一组转录控制序列，至少包括一个转录起始点以及一个以上的功能组件。

如TATA盒(TATA Box)：位于-25~-30bp，启动基因转录；

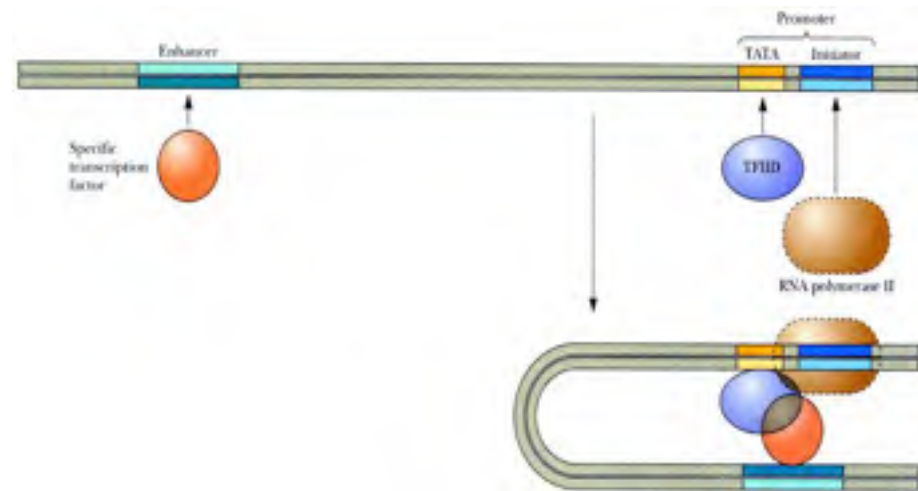
CAAT盒(CAAT Box)：位于-70~-80bp，决定启动子转录效率；

GC盒(GC Box)：位于-35bp，与转录因子Sp1结合，促进转录的过程。

❖ **增强子**：与TF特异结合，增强转录活性，基因任意位置都有效，无方向性。

❖ **Poly(A)加尾信号**：

含有II类启动子的基因末端存在终止信号：保守的AATAAA顺序及下游GT或T富含区。





# 第一节 基因 Gene

**基因**——是DNA分子中含有特定遗传信息的核苷酸序列，负责合成有功能的蛋白质或RNA所必需的DNA片段（部分RNA病毒例外），包括编码序列及调控序列，是遗传物质的最小功能单位。

## 二、 基因的现代概念

### （三） 基因的功能

**包括** { **生物学功能**：如蛋白激酶对蛋白质的磷酸化修饰  
**细胞学功能**：参与细胞间/内信号传递

- 基因的功能体现于基因的表达，基因的活化有严格的程序；
- 鉴定基因功能的技术：基因芯片、Western blot、模式生物等；
- 基因控制生物的性状：控制酶的合成、结构蛋白的成分。



# 第二节 基因组 Genome

**基因组——一个细胞或生物体中，一套完整单倍体的遗传物质的总和。**

- ❖ 不同生物基因组的大小和复杂程度各不相同；
- ❖ 基因组的功能：贮存和表达遗传信息。

不同生物基因组蕴含遗传信息量有差别。病毒<细菌<真核生物  
单倍体基因组中的全部DNA量——C值。

- ❖ 基因组中不同的区域具有不同的功能：
  - ◎ 有些区域编码蛋白质的结构基因
  - ◎ 有些区域复制及转录的调控信号
  - ◎ 有些区域的功能尚不清楚





# 第二节 基因组 Genome

基因组——一个细胞或生物体中，一套完整单倍体的遗传物质的总和。

## 一、原核生物基因组的特点



1. 基因组通常由环状双链DNA组成，相对较小( $4.6 \times 10^6$ bp, 约4000个基因), 细菌基因组包括细菌染色体DNA和质粒DNA, 前者存在于类核(nucleoid)区。
2. 只有一个Ori., 具操纵子结构: 功能上相关的几个结构基因往往串联在一起, 连同上游共同的调控区(启动子和操纵基因)及下游的转录终止信号构成的基因表达单位。基因开放时, 转录为一条多顺反子mRNA, 然后分别翻译合成各自的蛋白肽链。
3. 编码序列一般不重叠。结构基因是连续的, 无内含子, 转录后不需剪接加工。



# 第二节 基因组 Genome

基因组——一个细胞或生物体中，一套完整单倍体的遗传物质的总和。

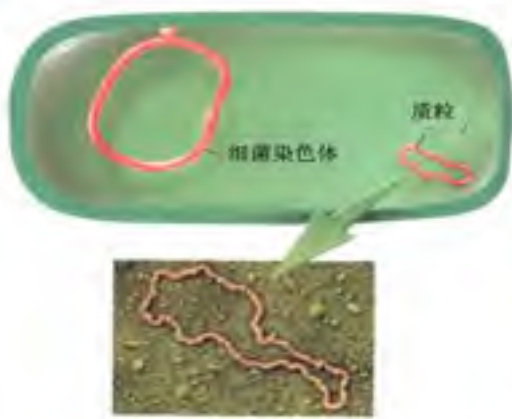
## 一、原核生物基因组的特点

4. 基因组中重复序列很少，多为单拷贝，编码rRNA基因是多拷贝。
5. 有编码同工酶的同基因 (isogene, 结构不完全相同, 而功能相同的基因)
6. 细菌基因组中存在可移动的DNA序列(插入序列和转座子)
7. 编码区所占比例较高(~50%), 小部分的非编码区主要是基因表达的调控序列。
8. 多种功能识别区域, 如复制起始区、复制终止区、转录起始和终止区, 这些区域往往含有反向重复序列。
9. 不同的原核生物基因组GC含量变化很大(25%-75%), 可用于识别细菌种类。





大肠杆菌 (*Escherichia coli*)



细菌染色体DNA和质粒DNA

质粒是细菌染色体外的、具有自主复制能力的闭合环状双链DNA分子。小的为2-3kb，大的数千个bp。

特点：

- ❖ 宿主细胞内独立自主复制；
- ❖ 质粒可以转移；
- ❖ 携带某些遗传信息，会赋予宿主细胞新的遗传性状；
- ❖ 可能发生质粒间或质粒与染色体之间的重组。

DNA replication



*Barbara McClintock* 1902-1992

Cold Spring Harbor Laboratory

Cold Spring Harbor, NY, USA

1983年诺贝尔生理学或医学奖

**转座/移位 (transposition) :**  
指染色体因子从原来的位置释放，插入到一个新位置上。

**转座因子 transposable element**

是一种可以由染色体的一个位置转移到另外位置的遗传因子，也就是一段可以发生转座的DNA。

转座子对基因组而言是一个不稳定因素，它可导致宿主序列删除、倒位或易位，在基因组中成为“可移动的同源区”。



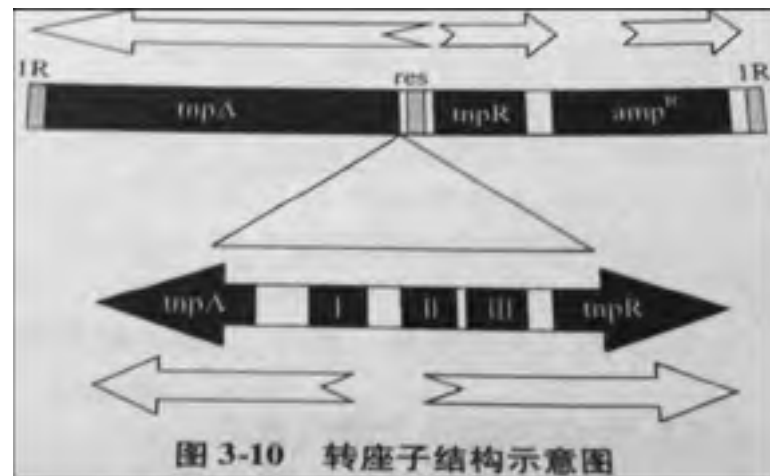
# 原核生物转座因子

- ❖ **插入序列**(insertion sequences, IS)小于2000bp, 只有转座相关基因
- ❖ **转座子** (transposon, Tn) 2-20kb, 常带有抗性基因(Kan<sup>R</sup>、ter<sup>R</sup>)等其它基因
- ❖ **Mu噬菌体** (MU)

是一种大肠杆菌的温和型噬菌体, 能像转座子那样在细胞内DNA上的不同位置间转移, 并没有一定的整合位置, 可引起突变。

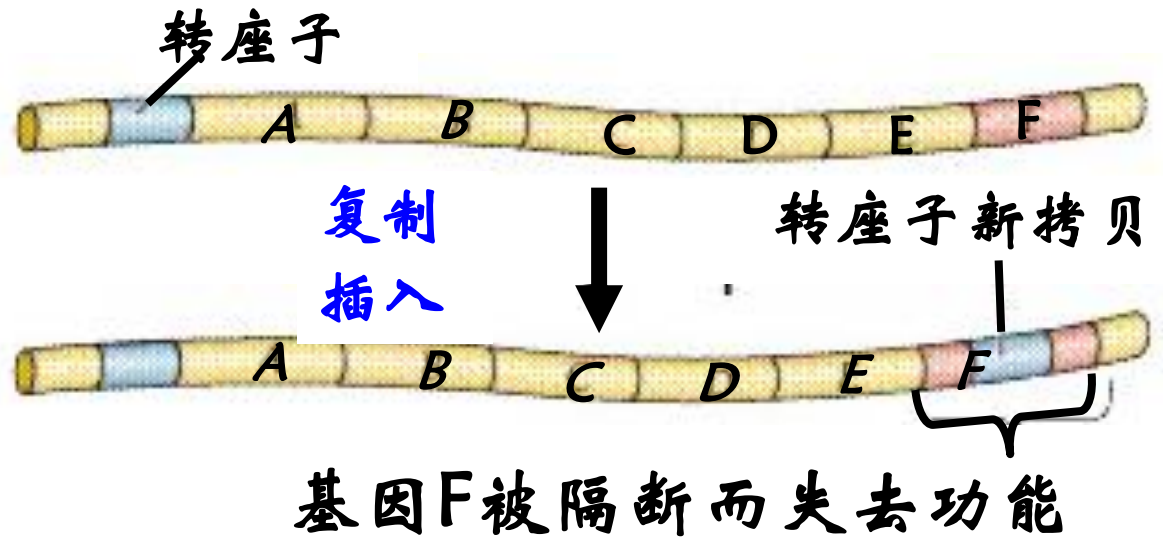
**转座的结果使靶点序列倍增;**  
**转座引起插入突变;**  
**携带标志基因使受体增添新基因。**

- ❖ **转座因子共同结构**  
**两末端具有反向重复序列**

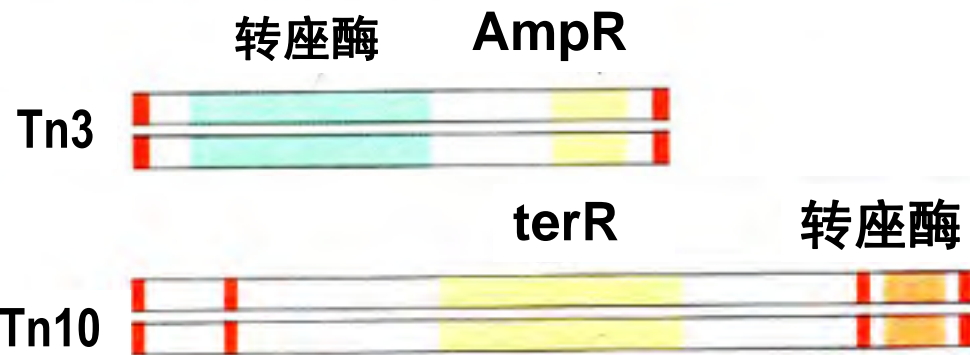




# 转座引起插入突变



# 携带标志基因使受体增添新基因





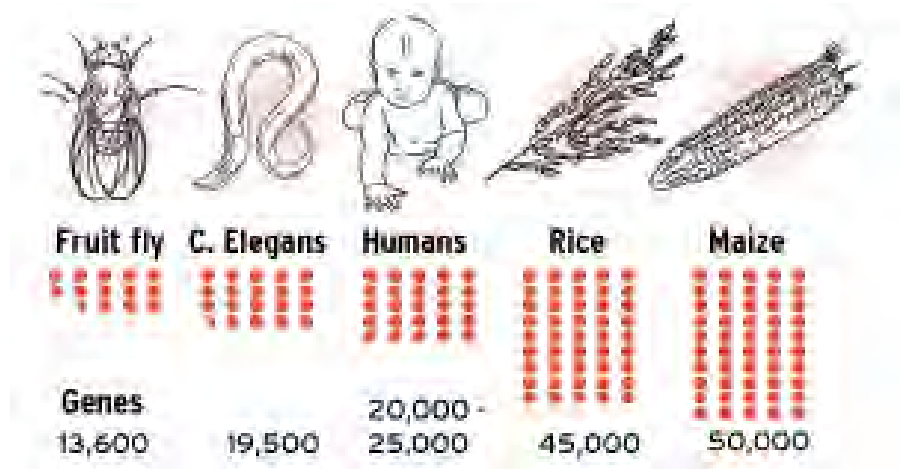
# 第二节 基因组 Genome

基因组——一个细胞或生物体中，一套完整单倍体的遗传物质的总和。

## 二、真核生物基因组的特点

### (一) 真核生物基因组 > 原核生物基因组

以染色体形式存于细胞核内，除配子细胞外，体细胞基因组是双份的。并非生物越高等，基因组越大。



{ 染色体DNA  
线粒体DNA

{ 复杂多样的结构  
复杂精细的调控机制





# 第二节 基因组 Genome

## (二) 真核生物基因组特点

1. 真核基因组较庞大( $3 \times 10^9$ bp, 约2.5万个基因), 非编码序列多于编码序列, 基因组包括染色体DNA和线粒体DNA, 前者被包裹在核膜内。
2. 基因组结构中有多个Ori., 但每个复制子长度较小。
3. 基因是不连续的断裂基因 (包含外显子和内含子等)。
4. 转录单位一般是单顺反子的, 即一个基因一种mRNA一种蛋白质, 但蛋白质的最终产物可因剪接方式的不同而有差异。
5. DNA序列组织的可变性: 从胚胎到成人并非一成不变。如B细胞成熟过程中Ig基因结构的重排等。



# 第二节 基因组 Genome

## (二) 真核生物基因组特点

6. 非编码序列存在大量重复序列(约占50%), 重复序列具多态性。

- 1) 高度重复序列 (10~300bp, 重复约 $10^6$ 次, 约占10%~60%, 较集中)
- 2) 中度重复序列 (100~500bp,  $10^3$ ~ $10^5$ 次, 约占10%~40%, 散在分布)
- 3) 低度重复序列 (单拷贝序列, 约占50~80%, 部分用于编码蛋白质)



# (1) 高度重复序列

❖ **卫星DNA satellite DNA**: 非编码区的串联重复序列，有固定的2~10bp重复单位。可用密度梯度离心将其与主体DNA分开，故称卫星DNA。在人基因组中约占5%。

■ **大卫星 (macro-satellite) DNA**

■ **小卫星 (minisatellite) DNA**  $A_{260}$  主带 1,700 1,692

总长度0.1~20kb，重复长度15~70bp，位于染色体近端粒处



是常 | 高度可变的小卫星DNA:

❖ 重复单位9~24 bp，呈高度多态性

核心序列 GGGCAGGAXG

分布广泛且排列。约占5%，易于重复性，1/3为回文结构。

端粒DNA: 2~20kb，端粒，(TTAGGG)<sub>n</sub>，用于亲子鉴定

功能: 染色体复制，末端保护



## (1) 高度重复序列

### ❖ 高度重复序列的功能

- a. 参与复制水平的调节
- b. 参与基因表达的调控
- c. 参与转位作用
- d. 与进化有关
- e. DNA指纹
- f.  $\alpha$  卫星DNA成簇分布在染色体着丝粒附近，可能与染色体减数分裂时染色体配对有关。



## (2) 中度重复序列 ( $10^4 \sim 10^5$ , 约占基因组的10~40%)

- 短分散重复片段(SINES): 300-500 bp, 拷贝数 $10^5$
- 长分散重复片段(LINES): 3500-5000 bp, 拷贝数 $10^4$

**Alu家族:** 有两个130bp的重复序列组成, 中间有31bp间隔序列, 重复单位300bp, 重复30~50万次, 因在170bp处有一AluI的酶切位点(AG/CT)而得名。

Alu序列是灵长类基因组特有, 可作为天然标记。也可能在hnRNA转录和加工中起作用; 遗传重组及染色体不稳定性有关; 形成Z-DNA。

还有: **KpnI家族:** 人类和灵长类DNA经Kpn I酶解后, 产生4个片段(1.2、1.5、1.8、1.9kb), 故被命名为Kpn I家族。散在分布的, 功能尚不清楚。

**Hinf家族:** 人类基因组经Hinf酶解后, 产生319bp长的串联重复序列。分散分布, 50-100个拷贝。



## (2) 中度重复序列 ( $10^3 \sim 10^5$ , 约占基因组的10~40%)

### ❖ 中度重复序列的功能

- 编码tRNA、rRNA、组蛋白、Ig的结构基因
- 种特异性, 可作为区分不同种哺乳动物的探针
- 与基因调控相关序列





### (3) 低度重复序列 (单拷贝序列)

- ❖ 在基因组中只出现一次或数次，占基因组的50~80%，一小部分用于编码不同功能的蛋白质。其它功能不清。
- ❖ 单拷贝序列的两侧往往为散在分别的重复序列。





## (4) 重复序列多态性

- ❖ 人类基因组中，平均约200对碱基可发生一对变异(称为中性突变)，其导致个体间核苷酸序列的差异，称为**基因多态性**。
  - **DNA位点多态性** DNA site polymorphism
  - **可变数串联重复序列多态性** Variable number of tandem repeats, VNTRs (高频率发生在小卫星和微卫星DNA中)
  - **限制性片段长度的多态性** Restriction fragment length polymorphism, RFLP
  - **单核苷酸多态性** Single-nucleotide polymorphism, SNP



## 限制性片段长度多态性(RFLP)

- ❖ DNA多态性发生在限制性内切酶识别位点上，酶解该DNA片段就会产生长度不同的片断。
- ❖ 按孟德尔方式遗传，在某一特定家族中，如果某一致病基因与特异的多态性片断紧密连锁，就可用这一多态性片断为一种“遗传标志”，来判断家庭成员或胎儿是否为致病基因的携带者。
- ❖ 甲型血友病、囊性纤维病变和苯丙酮尿症等均可借助这一方法得到诊断。





# HindIII

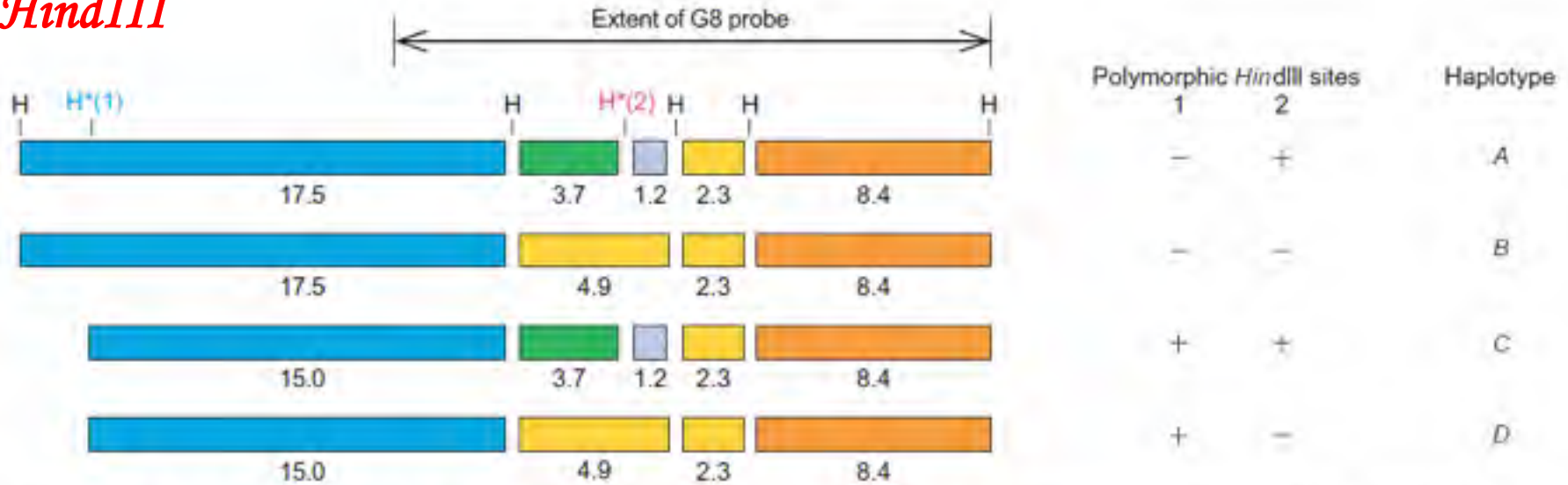


Figure 24.3 The RFLP associated with the Huntington disease gene. The *Hind*III sites in the region that hybridizes to the G8 probe are shown. The families studied show polymorphisms in two of these sites, marked with an asterisk and numbered 1 (blue) and 2 (red). Presence of site 1 results in a 15-kb fragment plus a 2.5-kb fragment that is not detected because it lies outside the region that hybridizes to the G8 probe. Absence of this site results in a 17.5-kb fragment. Presence of site 2 results in two fragments of 3.7 and 1.2 kb. Absence

of this site results in a 4.9-kb fragment. Four haplotypes (A–D) result from the four combinations of presence or absence of these two sites. These are listed at right, beside a list of polymorphic *Hind*III sites and a diagram of the *Hind*III restriction fragments detected by the G8 probe for each haplotype. For example, haplotype A lacks site 1 but has site 2. As a result, *Hind*III fragments of 17.5, 3.7, and 1.2 are produced. The 2.3- and 8.4-kb fragments are also detected by the probe, but we ignore them because they are common to all four haplotypes.

- ❖ A family living around Lake Maracaibo in Venezuela
- ❖ G8 probe to detect a PRLP that is very tightly linked to Huntington Disease

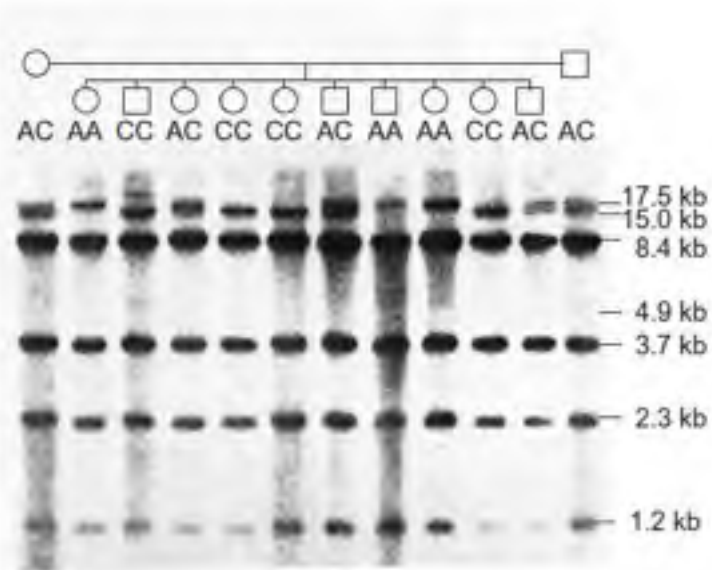
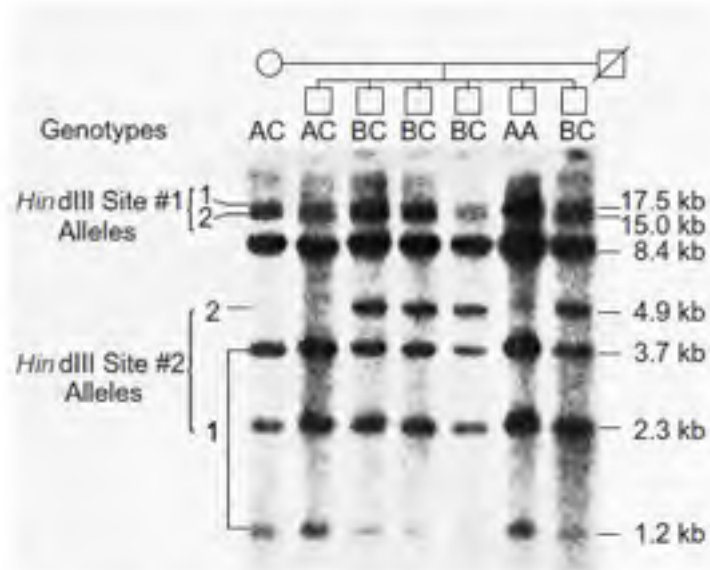


Figure 24.4 Southern blots of *Hind*III fragments from members of two families, hybridized to the G8 probe. The bands in the autoradiographs represent DNA fragments whose sizes are listed at right. The genotypes of all the children and three of the parents are shown at top. The fourth parent was deceased, so his genotype could

not be determined. (Source: Gusella, J.F., N.S. Wexler, P.M. Conneally, S.L. Naylor, M.A. Anderson, R.E. Taub, et al., A polymorphic DNA marker genetically linked to Huntington's disease, *Nature* 306:236. Copyright © 1983 Macmillan Magazines Limited.)

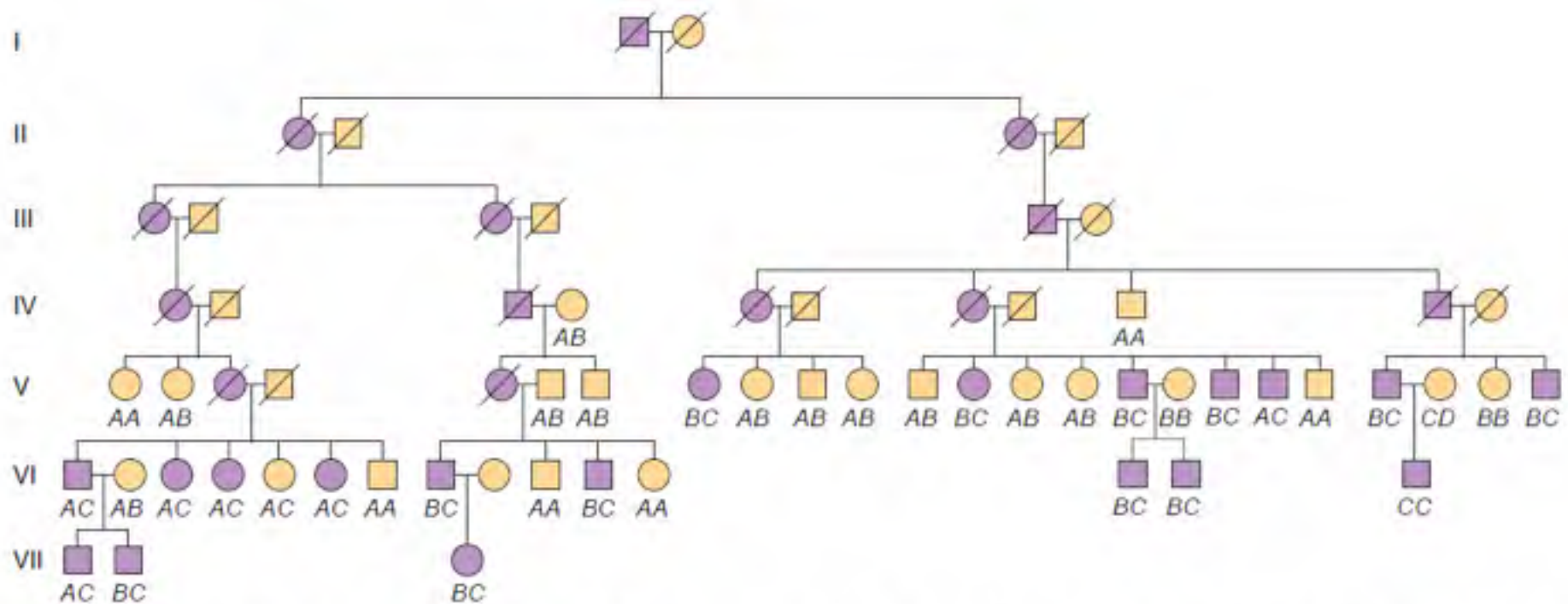


Figure 24.5 Pedigree of the large Venezuelan family with Huntington disease. Family members with confirmed disease are represented by purple symbols. Notice that most of the individuals with the C haplotype already have the disease, and that no sufferers

of the disease lack the C haplotype. Thus, the C haplotype is strongly associated with the disease, and the corresponding RFLP is tightly linked to the Huntington disease gene.





## 单核苷酸多态性 (SNP)

- ❖ 不同个体基因组DNA序列中在单个核苷酸上的差别
- ❖ 定义SNP位点的变异在人群中出现的频率大于1%，人类基因组中SNP的数目约为3百万~1千万。
- ❖ SNP是基因组多样性的主要表现。遗传信息变异是所有基因组的共同特征。不同个体在疾病易感性、反应性和其他性状上的差别，最常见与SNP变异有关。

**SNP决定你的与众不同!**

## Polymorphism

"Poly" many "morpho" form



遗传信息变异与疾病易感性



# 第二节 基因组 Genome

## (二) 真核生物基因组特点

7. 存在多基因家族和超基因家族，有假基因。

(1) 多基因家族 (multigene family):

是指一组具有类似功能，核苷酸序列又有同源性的基因。

假基因 (pseudo gene,  $\Psi$ ):

在多基因家族中，不产生有功能基因产物的基因。即序列与有功能的基因结构相似，但不能转录，或转录后生成无功能的基因产物。

产生原因是基因在进化过程中复制时发生突变，或插入了mRNA逆转录的cDNA，缺失基因启动子序列所致。

假基因往往缺少正常基因的内含子，两侧有顺向重复序列，大多没有调控区。



# 第二节 基因组 Genome

## (二) 真核生物基因组特点

### 8. 基因类型多样

#### (1) 跳动(跃)基因 (可转移的DNA成分、转座子) :

可在DNA分子间进行转移的DNA片段。与一般转移概念不同, 转移后仍保留原来位置上的DNA序列, 只是把一个新拷贝插入到另外的位置上。真核和原核细胞都有。

#### (2) 非剪接基因 (连续基因) :

原核和真核细胞都有。真核rRNA基因也是非剪接基因。

#### (3) 自私DNA (selfish DNA) :

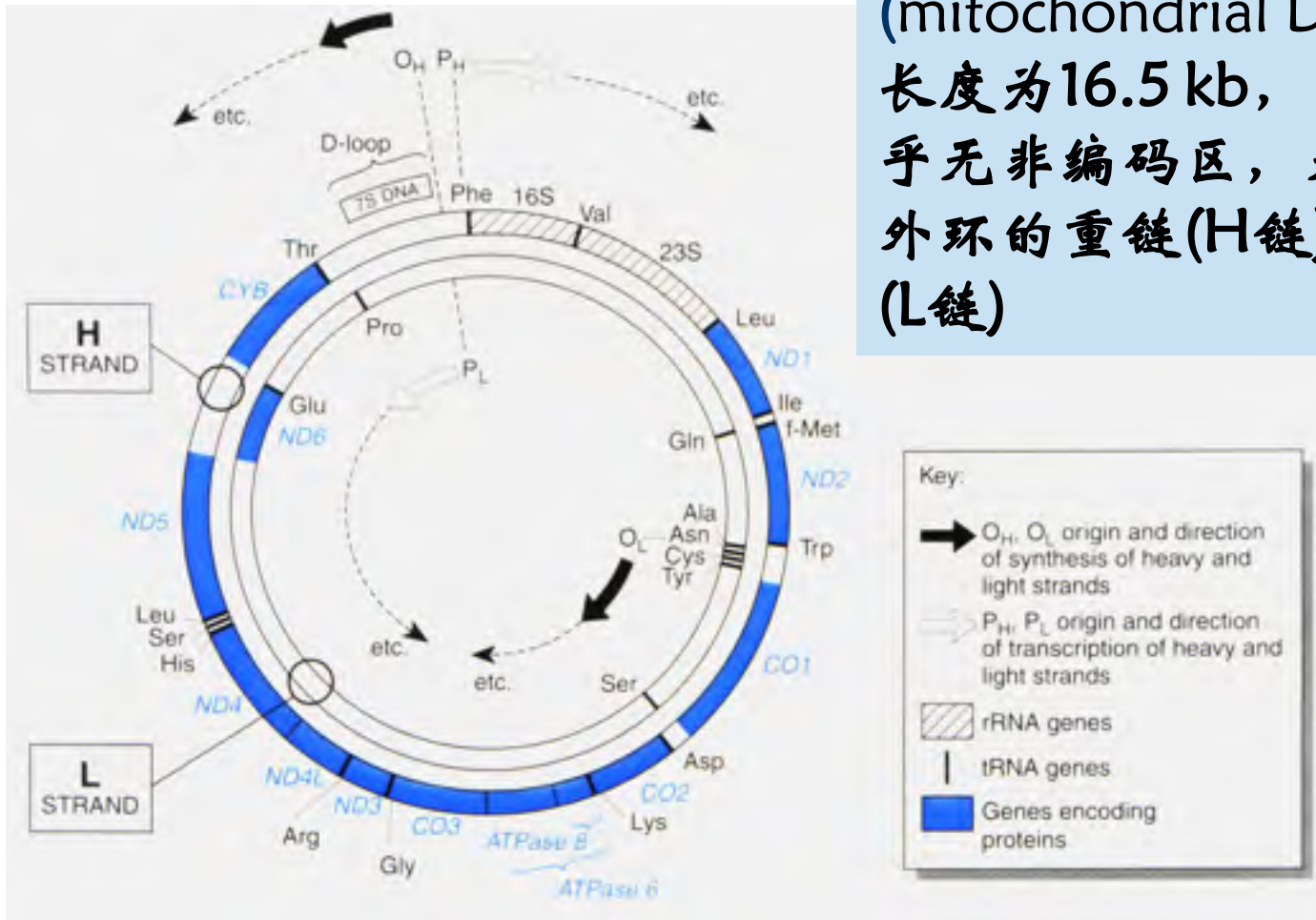
指极少转录成mRNA并翻译的非编码序列, 包括分散的高度、中度重复序列, 内含子和间隔序列等, 除了复制自己, 它们对细胞存活、代谢等几无贡献, 故称为自私DNA。且有些成分还通过转录成mRNA, 生成cDNA, 再转位插入到基因组, 颇象机体内寄生虫的繁殖、生活, 故也有人称之为寄生DNA。



# 第二节 基因组 Genome

## (三) 线粒体基因组特点

闭环双链的**线粒体DNA** (mitochondrial DNA, mtDNA), 长度为16.5 kb, 结构紧凑, 几乎无非编码区, 无内含子。外环的重链(H链)和内环的轻链(L链)





# 第二节 基因组 Genome

## (三) 线粒体基因组特点

动物细胞的mtDNA编码2个rRNA、22个tRNA和多种呼吸链复合物的酶蛋白：

- ❖ **tRNA基因**：啤酒酵母mtDNA中有24个tRNA基因，粗链孢霉菌的mtDNA中有40个tRNA基因，而人类mtDNA中有22个tRNA基因。
- ❖ **rRNA基因**：人类mtDNA中有一个拷贝的16S及12SrRNA基因。
- ❖ **细胞色素氧化酶基因**：细胞色素氧化酶有七个亚基，其中三个亚基mtDNA编码，四个亚基由细胞核DNA编码。
- ❖ **ATP酶基因**：ATP酶(MW: 340 KDa)，含有十个亚基，其中四个由mtDNA编码。
- ❖ **细胞色素还原酶 (b,c复合物) 基因**：酶的七个亚基中有一个由mtDNA编码。
- ❖ 另外，还有一些**抗药性基因**也在mtDNA上。



# 第二节 基因组 Genome

## 三、病毒基因组的特点

### 1. 病毒基因组大小差别很大

最小的乙肝病毒：3kb，仅编码4种蛋白质；

最大的痘病毒：可达300kb，有几百个基因。

### 2. 病毒基因组由一种核酸组成：可以是DNA或RNA

DNA或RNA病毒可以是单链或双链，可以是环状或线性的。

### 3. 多数RNA病毒基因组是连续的RNA单链组成

少数RNA病毒基因组是由不连续的几条链组成(节段性)，

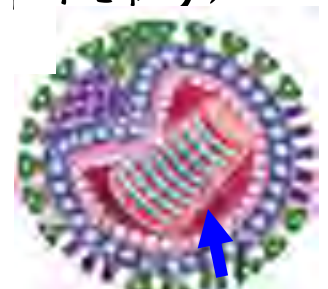
每段RNA分子都有编码蛋白质的信息。

☹️ 流感病毒基因组由8条RNA分子构成；

☹️ 呼肠孤病毒基因组由10个双链RNA片段。

大多数DNA病毒的基因组是双链DNA分子

武汉大学基础医学院微生物学系 多数RNA病毒的基因组是单链RNA分子



8节段-ssRNA

禽流感病毒





# 第二节 基因组 Genome

## 三、病毒基因组的特点

4. 病毒基因组的功能单位(转录单位)形成多顺反子mRNA  
可被一起转录成为含有多个mRNA的分子,称为多顺反子mRNA,然后再加工成各种蛋白质的模板mRNA.
5. ❖ 病毒基因组基因组编码的蛋白质( $>90\%$ )是在一个启动子启动作用部位或编码蛋白质的顺读框内,是基因集成各种顺读框(即)
  - ❖ 牛痘病毒基因组DNA中比翻译的核苷酸用0.78%的Kb mRNA中,然后再翻译成各种蛋白质,其中I、F、G及H
6. 病毒基因组外存在基因的,重叠蛋白与病毒的装配有关, E蛋白是基因组DNA解旋酶编码的蛋白质分子,这种结构使较小的基因组能够携带较多的遗传信息。

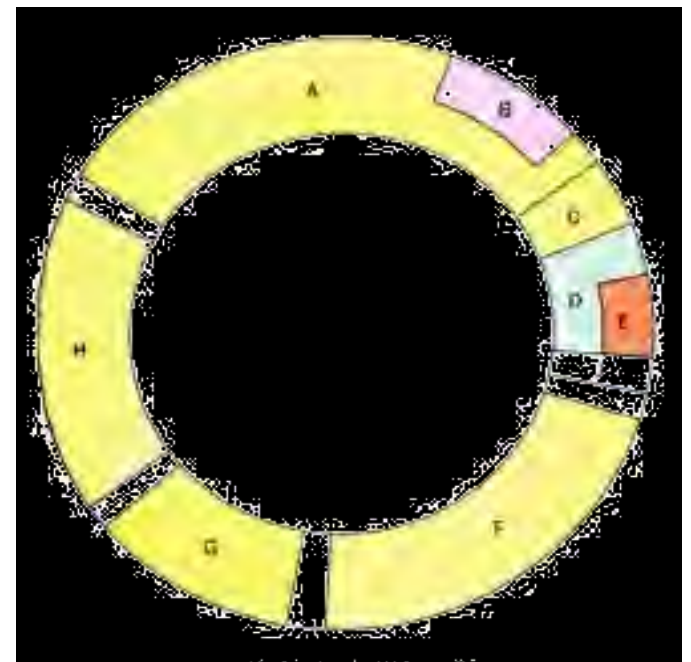


# 基因重叠：同一段DNA片段能够编码两种甚至三种蛋白质 仅见于线粒体和质粒DNA

❖ 基因重叠有以下几种情况：

- 1) 完全重叠
- 2) 部分重叠
- 3) 两个基因只有一个碱基重叠

一个基因终止密码子的最后一个碱基是  
另一个基因起始密码子的第一个碱基



噬菌体  $\phi$ X174 基因图



# 第二节 基因组 Genome

## 四、逆转录病毒基因组的特点

### 1. 逆转录病毒基因组的结构特征：

RNA病毒类似于真核生物的mRNA，  
5'有帽子，3'有polyA尾巴。

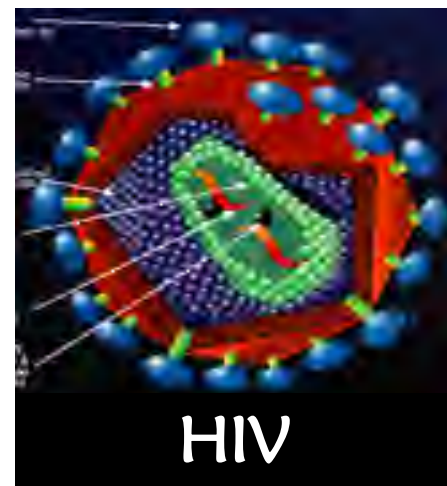
基因组有两个拷贝。

🏠 编码区：均含有gag、pol、env三个结构基因

🏠 非编码区：与基因组复制和基因表达有关。

### 2. 原病毒基因组的转录和翻译

逆病毒基因组的转录和翻译需经过DNA中间体（原病毒）  
才能完成。





# 第三节 基因组学 Genomics

## 基因组学是研究基因组的科学

- **背景：**1985年提出，1990年开始实施人类基因组计划 (Human Genome Project, HGP)，产生基因组学。
- **概念：**以分子生物学技术、计算机技术和信息网络技术为研究手段，以生物体内全部基因为研究对象，在全基因背景下和整体水平上探索生命活动的内在规律及其内外环境影响机制的科学。
- **目的：**阐明基因组结构、结构与功能的关系以及基因与基因之间相互作用。
- **分类：**

{	结构基因组学	{	肿瘤基因组学
	功能基因组学		植物基因组学
			药物基因组学
			环境基因组学



# 第三节 基因组学 Genomics

## 一、基因组学研究内容

### (一) 结构基因组学

以全基因组测序为目标的基因结构研究，  
弄清基因组中全部基因的位置和结构，  
为基因功能的研究奠定基础。

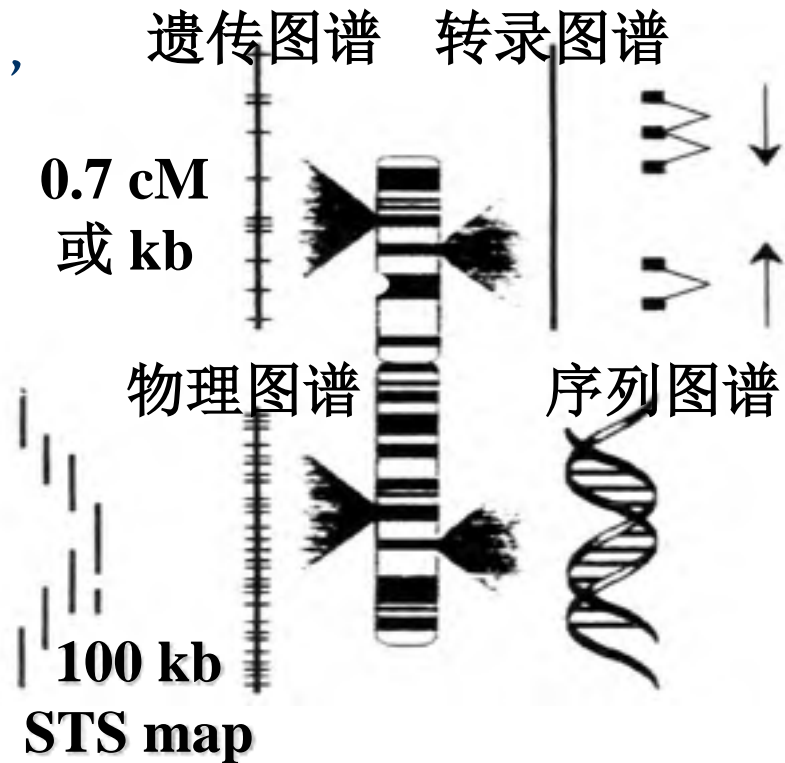
其包括建立高分辨的遗传图谱

物理图谱

转录图谱

序列图谱

(人类基因组的四张图谱)





## ❖ 遗传图谱 (genetic linkage map, 连锁图谱)

——孟德尔的“新生”

指基因或DNA标记在染色体上的相对位置与遗传距离。是以DNA多态性遗传标记为界标，以遗传学距离（在减数分裂事件中两个位点之间进行交换、重组的百分率，1%的重组率称为1厘摩cM）为图距的基因组图。

人类单倍体基因组

遗传标志

限制性片段长度多态性 (RFLP)

短串联重复序列STR (short tandem repeat, SSLP)

单核苷酸多态性 (SNP)

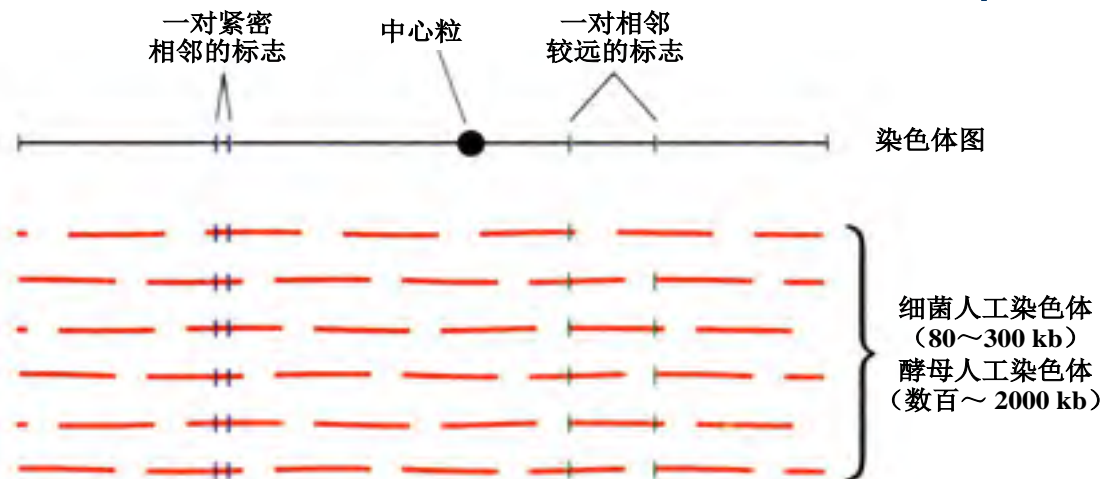




## ❖ 物理图谱 (physical map) - 路标与路轨

表示基因组中不同基因或DNA标记之间的实际距离。以**定位的序列标记位点STS**作为路标，以**DNA实际长度**即bp、kb、Mb (百万碱基对) 为图距的基因组图谱。是进行DNA序列分析和基因组组织结构研究的基础。

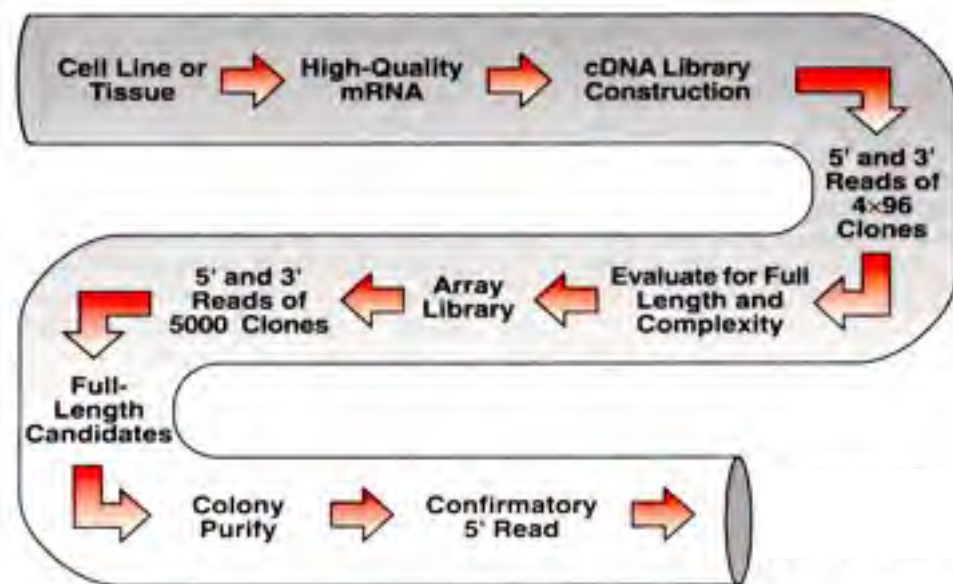
- 物理图谱首先是利用**限制性内切酶**将染色体切成片段，再根据重叠序列把片段连接成染色体，确定遗传标志之间的**物理距离**。
- 作图标志是已知序列的DNA片段，称为STS (sequence tagged site) 。





## ❖ 转录图谱（表达图谱）——生命的乐谱

转录图谱是在识别基因组所包含的蛋白质编码序列的基础上绘制的结合有关基因序列、位置及表达模式等信息的图谱。以表达序列标签EST（长度为300~500 bp的cDNA）为位标，根据转录顺序的位置和距离绘制的图谱，它是染色体DNA某一区域内所有可转录序列的分布图（基因图谱）。





## ❖ 序列图谱(分子水平的物理图谱)——重中之重

以某一染色体上所含的**全部碱基顺序**绘制的图谱。是转录序列、调节序列和功能未知序列的**总和**。

**DNA sequencing** (DNA序列分析)是一个包括制备DNA片段化及碱基分析、DNA信息翻译的多阶段的过程。通过测序得到基因组的序列图谱。



2003.5.28—2003.6.2

冷泉港-人类基因组完成图发布会

终极目标?





# 基因和基因组数据库

<http://www.ncbi.nih.gov/genome>

The screenshot shows the NCBI homepage with the following elements:

- Header:** NCBI logo, National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health.
- Navigation:** PubMed, All Databases, BLAST, OMIM, Books, TaxBrowser, Structure.
- Search:** Search for NCBI Web Site.
- Left Sidebar:** SITE MAP, Alphabetical List Resource Guide, About NCBI: An introduction to NCBI, GenBank: Sequence submission support and software, Literature databases: PubMed, OMIM, Books, and PubMed.
- Main Content:**
  - What does NCBI do?** Established in 1988 as a national resource for molecular biology information, NCBI creates public databases, conducts research in computational biology, develops software tools for analyzing genome data, and disseminates biomedical information - all for the better understanding of molecular processes affecting human health and disease. [More...](#)
  - New Global NCBI Search Engine** NCBI's growing number of Entrez databases can now be searched at once! [Go](#)
  - Entrez Gene** You can now use Entrez to search for information centered on the concept of a gene, and connect to many sources of related information both within and outside NCBI.
- Hot Spots:**
  - Assembly Archive
  - Clusters of orthologous groups
  - Coffee Break, Genes & Disease, NCBI Handbook
  - Electronic PCR
  - Entrez Home
  - Entrez Tools
  - Gene expression omnibus (GEO)
  - Human genome resources
  - Malaria genetics &



## DDBJ (DNA Data Bank of Japan)

创建于1984年，由日本国家遗传学研究所负责管理。

DDBJ的网址是：<http://www.ddbj.nig.ac.jp/>

## EMBL (European Molecular Biology Laboratory)

创建于1984年，<http://www.ebj.ac.uk/>

1988年，EMBL、NCBI与DDBJ共同成立了国际核酸序列联合数据库中心，可进行核苷酸序列检索和序列相似性咨询。



# 第三节 基因组学 Genomics

## 一、基因组学研究内容

### (二) 功能基因组学 (后基因组学)

在基因组水平上阐明DNA序列的功能，从基因整体水平上对基因的活动规律进行系统研究。

1. 基因表达及调控
2. 人类基因信息的识别和鉴定
3. 基因功能信息的提取和鉴定
  - 1) 人类基因突变体的系统鉴定
  - 2) 基因表达图谱的绘制
  - 3) “基因改变—功能改变”的鉴定
  - 4) 蛋白质水平，修饰状态和相互作用的检测

### 4. 测序和基因多样性分析







# 第三节 基因组学 Genomics

## 一、基因组学研究内容

### (三) 比较基因组学 comparative genomics

基于基因组图谱和测序基础上，对已知的基因和基因组结构进行比较，以了解基因的功能、表达机理和物种进化的学科。

将人类基因组与模式生物基因组进行比较研究，根据同源性方法分析人类基因的功能，疾病分子机制、阐明物种进化关系和基因组内在结构。



# 第三节 基因组学 Genomics

## 一、基因组学研究内容

### (四) 基因组学的研究技术和方法

#### 结构基因组学

- ❖ 脉冲场凝胶电泳
- ❖ 毛细管电泳
- ❖ 基因芯片技术
- ❖ 全基因组随机测序

#### 功能基因组学

- ❖ 基因转导技术
- ❖ 反义核苷酸技术
- ❖ 核酶技术
- ❖ 肽核酸(PNA)技术
- ❖ RNA干扰技术
- ❖ 基因敲除技术
- ❖ 基因表达系列分析(SAGE)
- ❖ cDNA芯片
- ❖ 反向遗传学
- ❖ 蛋白质组研究
- ❖ 生物信息学技术



# 第三节 基因组学 Genomics

## 二、人类基因组计划及其进展

### (一) HGP

#### 人类基因组研究——揭开生命的奥秘



由美国科学家于1985年率先提出、于1990年正式启动的。

- ❖ 1984年12月Utah州的Alta,White R受美国能源部的委托,主持召开了一个小型会议,讨论DNA重组技术的发展及测定人类整个基因组的DNA序列的意义。
- ❖ 1985年6月,在美国加州举行了一次会议,美国能源部提出了“人类基因组计划”的初步草案。
- ❖ 1986年3月,诺贝尔奖获得者Renato Dulbecco发表短文——《肿瘤研究的转折点:人类基因组测序》(Science, 231: 1055-1056)
- ❖ 1986年6月,在新墨西哥州讨论了这一计划的可行性。随后美国能源部宣布实施这一草案。
- ❖ 1987年初,美国能源部与国家医学研究院(NIH)下拨了启动经费约550万美元,1987年总额近1.66亿美元。同时,美国开始筹建实验室。
- ❖ 1989年美国成立“国家人类基因组研究中心”。诺贝尔奖金获得者J.Waston为第一任主任。
- ❖ 1990年,历经5年辩论之后,美国国会批准美国的“人类基因组计划”于10月1日正式启动。总体规划是:拟在15年内至少投入30亿美元,进行对人类全基因组的分析。



# 第三节 基因组学 Genomics

## 二、人类基因组计划及其进展

### (二) HGP的研究进展

目标:

完成人类基因组全部DNA序列测定

在2003年6月26日由美国总统克林顿与英国首相布莱尔联合宣布完成，这一天也因此成为人类历史上“值得载入史册的一天”

主要成果:

- 30亿bp的DNA包含人类基因约3-3.5万个;
- 人类基因组中存在“热点”和大片“荒漠”。
- 35.3%的基因组包含重复序列
- 地球上人与人具有99.99%的相同基因，SNP对“个性”起决定作用。



1999年12月完成人类第22号染色体测序



# 第三节 基因组学 Genomics

## 二、人类基因组计划及其进展

### (三) 我国对HGP的贡献

- ❖ 1994年启动,先后开展“中华民族基因组中若干位点基因结构的研究”和“重大疾病相关基因的定位克隆、结构和功能研究”;
- ❖ 1998年在上海成立南方基因中心;
- ❖ 1999年在北京成立北方人类基因组中心;
- ❖ 1999年7月完成人类3号染色体短臂上约30Mb的测序,约占人类基因组的1%。
- ❖ 2001年8月,HGP中国部分"完成图"提前两年绘就。



Small steps, big gains. CAS's Yang Huanming with a map showing the Beijing center's progress toward sequencing its allotted 1% slice of the human genome project.



# “炎黄一号”

(07.10.11)



## 炎黄计划(06年)

### 第一步

绘制中国人的基因组参考图，命名为“炎黄一号”，现已完成。

### 第二步

再绘制99个中国人的个体全基因组序列图，构成中国人群的遗传和多态性标准图谱，成为基因与医疗和健康的关键组成部分，简称“炎黄99”。

### 第三步

在上述基础上开展大众的基因与健康的预测、监测及个体化诊断和治疗，实现解读基因。

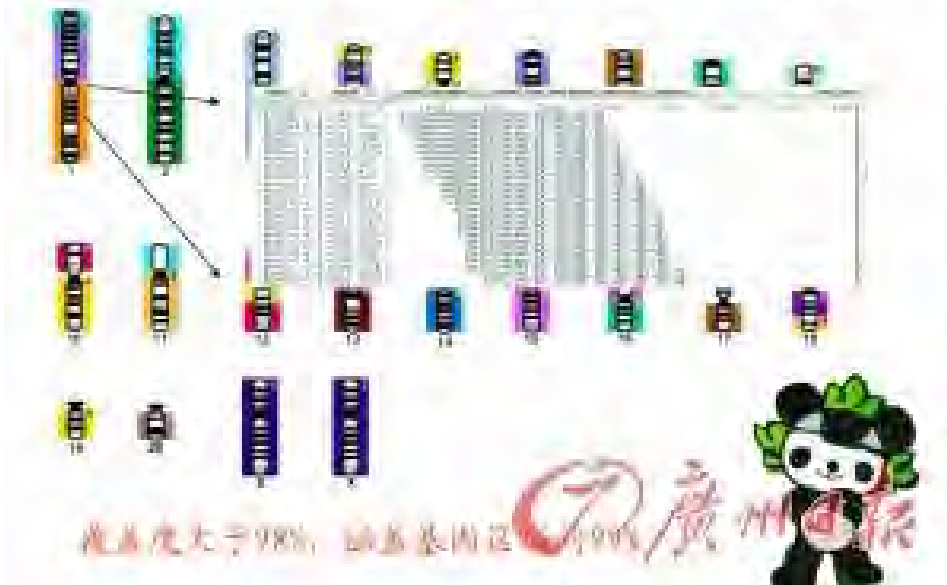
- ❖ 2008年1月，华大基因研究院和国际同行提出“国际千人基因组计划”。
- ❖ 万种微生物基因组计划，微生物的基因比人的要少得多，大约有自己的千分之一左右，现在华大基因已经做了400种左右





# 世界首张大熊猫基因组序列图谱 (09.10.11)

## “大熊猫基因组”基因组框架图完成



**大熊猫基因组与狗的基因组在结构上最为接近，与人也有较大的相似性，在哺乳动物中与小鼠差异较大。大熊猫的基因组与人的基因组大小相似，大熊猫共有21对染色体，约为30亿个碱基对，包含2-3万个基因。**



# 第三节 基因组学 *Genomics*

## 二、人类基因组计划及其进展

### (四) HGP 的意义与影响

- ❖ 生物的进化史，都刻写在各基因组的“天书”上；
- ❖ 发展生物芯片等生物技术研究新功能基因；
- ❖ 对细胞、胚胎、组织工程的推动：干细胞、克隆技术、器官再造。
- ❖ 发现新的致病基因，建立疾病模型；
- ❖ 基因诊断：发展一些复杂疾病的基因诊断方法，如“肿瘤基因组解剖计划”；亲子鉴定；疾病易感基因的识别；
- ❖ 转基因食品与基因工程药物（生物产业的社会和经济效益）
- ❖ 筛选新药和药物的靶点，针对基因蛋白产物的药物设计；
- ❖ 基因治疗
- ❖ 人类基因组多样性形成个体化医学及个体化药物治疗：药物基因组学





# 第三节 基因组学 Genomics

## 二、人类基因组计划及其进展

### (五) HGP带来的负面作用与影响

- ❖ 侏罗纪公园不再是科幻
- ❖ 种族选择性灭绝性生物武器
- ❖ 基因专利战、基因资源掠夺战



“基因是人类共同财产”  
VS “人类基因组实际上是个人”

- ❖ 人类基因组 DNA 序列数据的自由共享？  
新发现的基因是否可以申请专利？

优生涉及种族问题？ “遗传歧视”

基因注  
定论？  
人隐私！





# 第三节 基因组学 Genomics

## 三、基因组变异的生理病理意义

### (一) 基因组在进化过程中发生变异

基因组的序列变异：突变、插入、缺失、不同数目串联重复及SNP；寡核苷酸微阵列分析(ROMA)检测基因组中“拷贝数多态性”

### (二) 染色体DNA变异可导致疾病发生

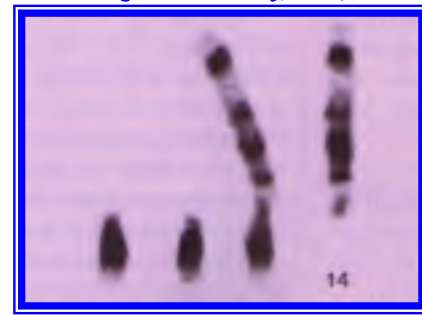
染色体数目的变异；结构与排列的变异

### (三) 线粒体DNA突变可引起线粒体基因病

mtDNA无修复系统及His保护，突变频率高，如 Leber遗传性视神经病、家族性唐氏症

### (四) 易感基因与环境的相互作用

环境相关疾病易感基因的等位多态性。





# 第三节 基因组学 Genomics

## 三、基因组变异的生理病理意义

### (五) 所有的疾病，都可以说是基因病

■ 基因相关论  
**单基因病**

■ 基因修饰论  
变基因的表

■ 基因的多态  
**多基因疾病**

**线粒体基因**

**基因组与癌症**

❖ 基因组某**单个基因座**上存在缺陷基因所引起的一类遗传病。（**6000**余种）红绿色盲、血友病等。

❖ 一种基因型可产生不同的表现型，或一种表现型可由**其二基因型**所影响 即基因型与表型非“一

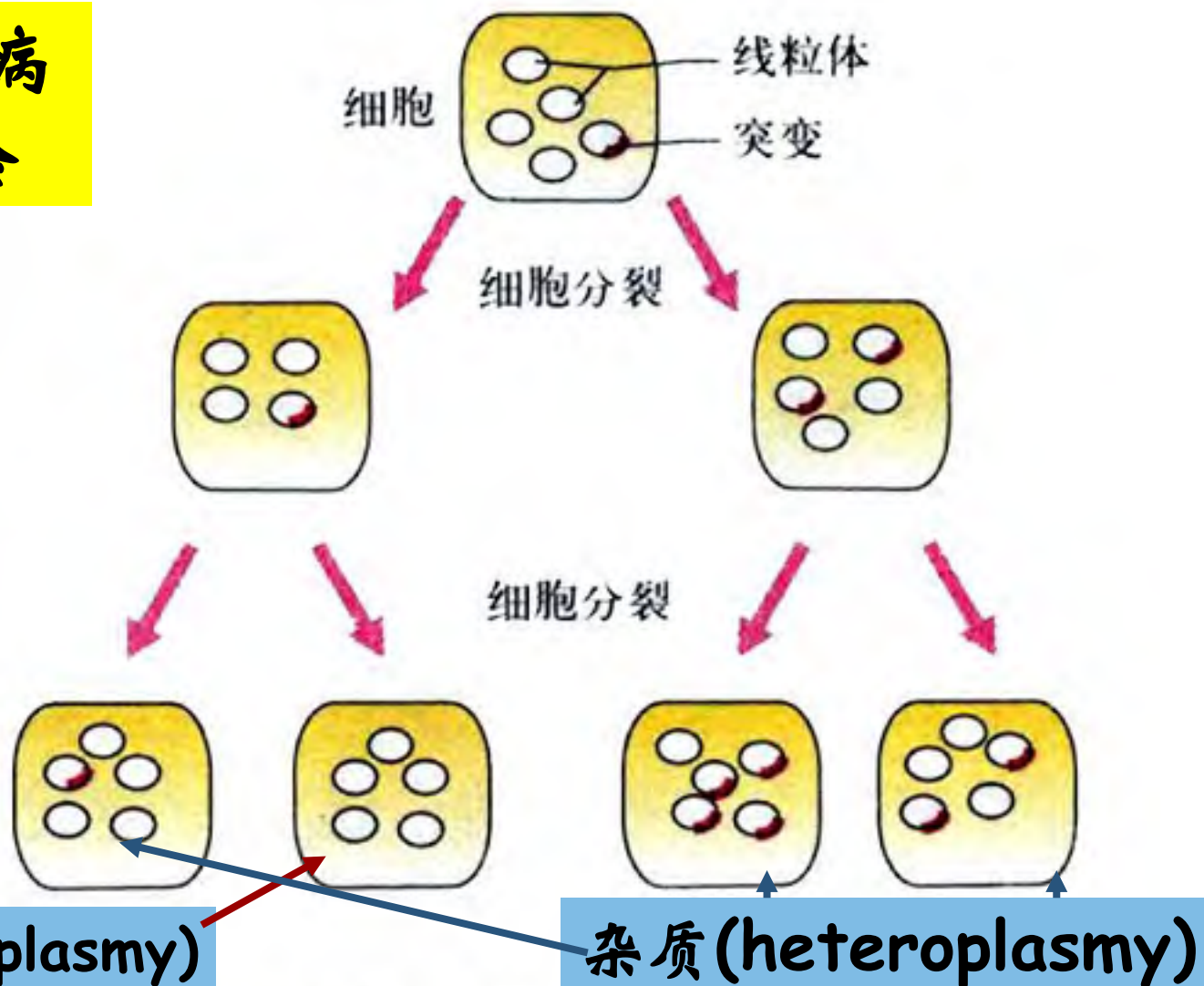
❖ 突变基因在mtDNA上，其传递和表达完全不同于核基因突变引起的遗传病，而成为一组独特的遗传病。

❖ 遗传特点:母系遗传(突变的mtDNA通过卵子细胞质的线粒体传给子代); 数量特征: 突变的mtDNA数量超过阈值时, 会出现临床症状。突变mtDNA所占比例似与临床症状的表现程度相关。





# 线粒体基因病的数量概念



线粒体突变导致的异质性胞质





# 基因组与癌症研究

- ❖ 癌症是最常见的基因病。
- ❖ 癌基因和抑癌基因
- ❖ 全部基因组序列对比
- ❖ 获得癌基因活化和抑癌基因失活：基因片段丢失，重排，碱基替换，小片段插入或缺失，扩增或甲基化。





# 小结

- ❖ 医学分子生物学研究对象和研究内容
- ❖ 基本概念：基因，基因组，断裂基因，ORF，外显子，内含子，多基因家族，基因超家族，假基因，DNA多态性，卫星DNA，RFLP，SNP，转座子，基因重叠，转位因子，结构基因组学，功能基因组学
- ❖ 原核生物基因组的结构特点
- ❖ 真核生物基因组的结构特点
- ❖ 结构基因组学四大图谱