

第十四章

基因、基因组和人类基因组计划

Chapter 14

Gene, Genome and Human genome project

本章要求

- 1.掌握基因的概念、结构及原核生物与真核生物基因的主要差别。
- 2.掌握基因组的概念及原核生物与真核生物基因组的结构特征。
- 3.掌握人类基因组计划的概念、内容和意义。
- 4.掌握下列概念
可转录序列（结构基因）； 调节序列（调控序列）； 重复序列。

第一节 基因

一、基因的概念

基因（**gene**）是细胞或生物个体贮存和表达遗传信息的基本结构单位。

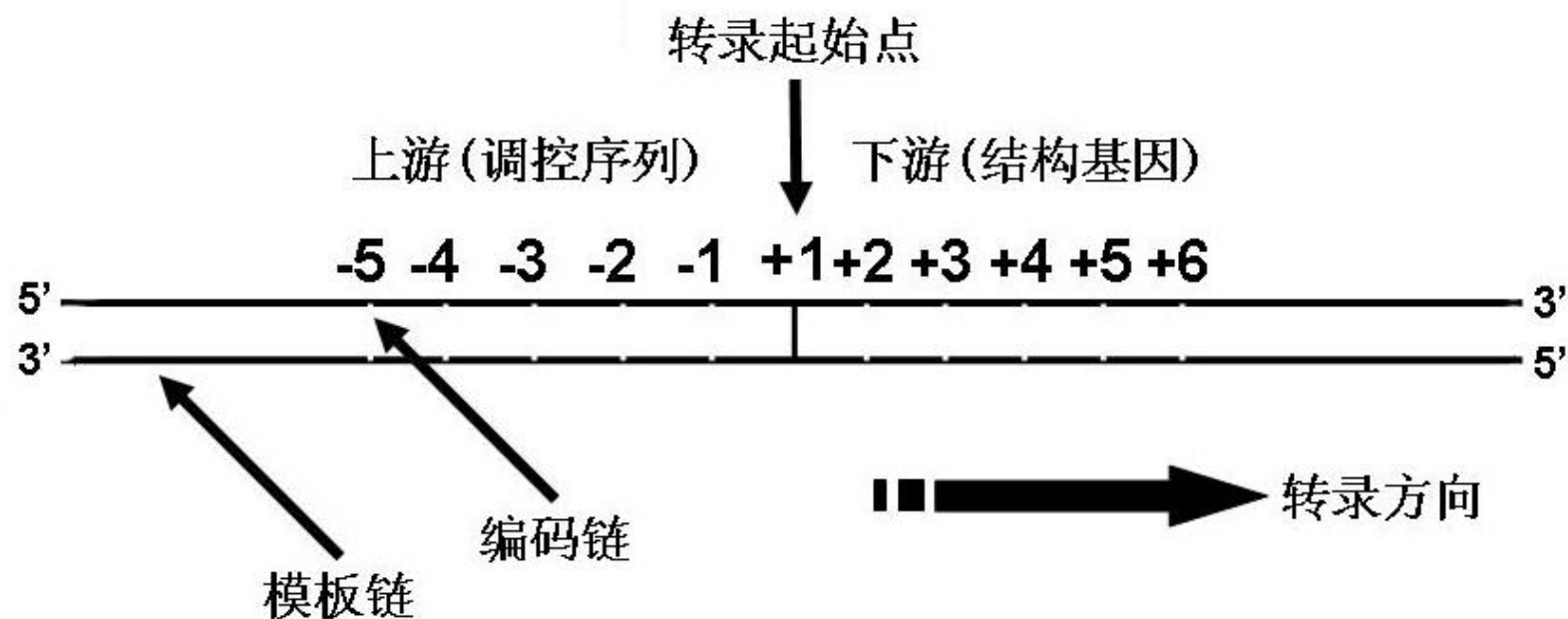
（一）可转录序列

DNA分子中可转录产生**RNA**的序列，习惯上称为结构基因。

(二) 调节序列

DNA分子中具有调节结构基因转录功能的**DNA**序列，习惯上称为调控序列。包括启动子、增强子、沉默子等。

基因的基本结构



二、原核生物基因结构特点

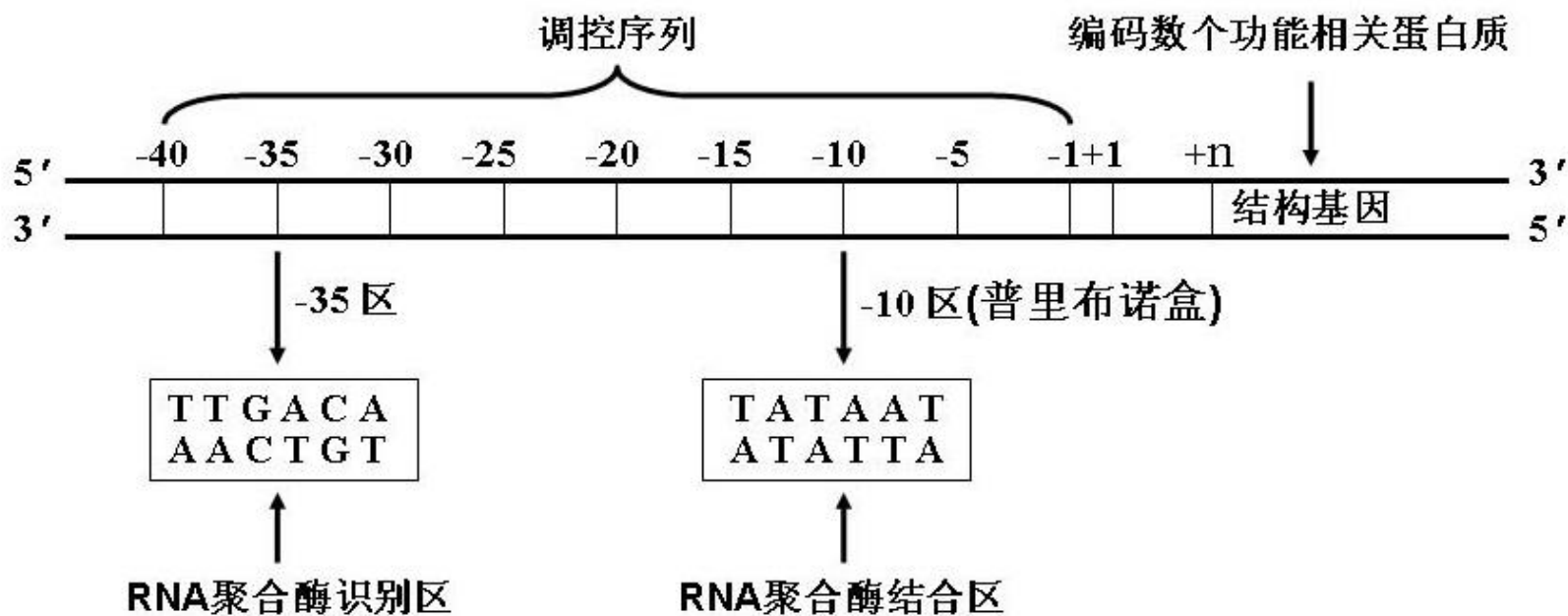
多顺反子转录单位（操纵子）。

结构基因中无内含子。

启动子具有一致序列如-10区“TATAAT”序列和-35区“TTGACA”序列。

某些原核生物基因另有正调控元件和负调控元件。

原核生物基因的基本结构



三、真核生物基因结构特点

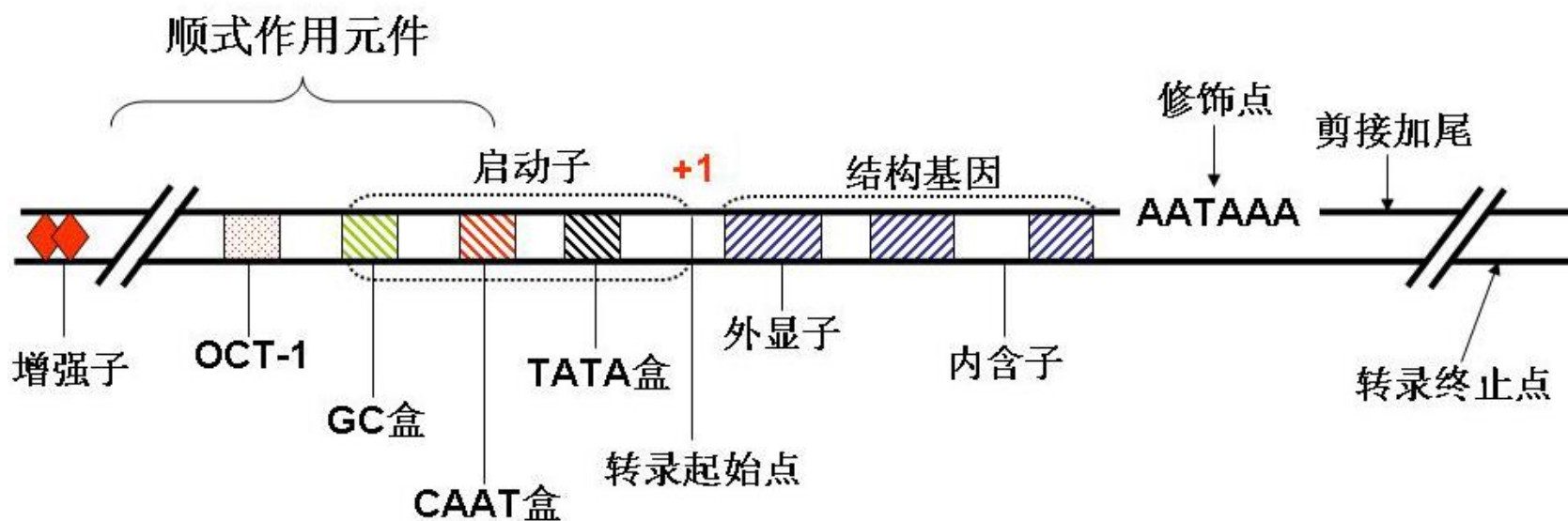
单顺反子转录单位。

结构基因是断裂基因：

外显子（**exon**）、内含子（**intron**）。

顺式作用元件：启动子、增强子、
负调节序列（沉默子）。

真核生物基因的基本结构



第二节 基因组

一、基因组的概念

基因组是细胞或生物体的一套完整单倍体遗传物质的总和。

细胞核基因组、细胞器（线粒体或叶绿体）基因组。

病毒基因约占基因组序列的**95%**；

原核生物基因约占基因组序列的**50%**；

人类基因约占基因组序列的**1.5%**。

基因组中有可移动的**DNA**序列如转座子。

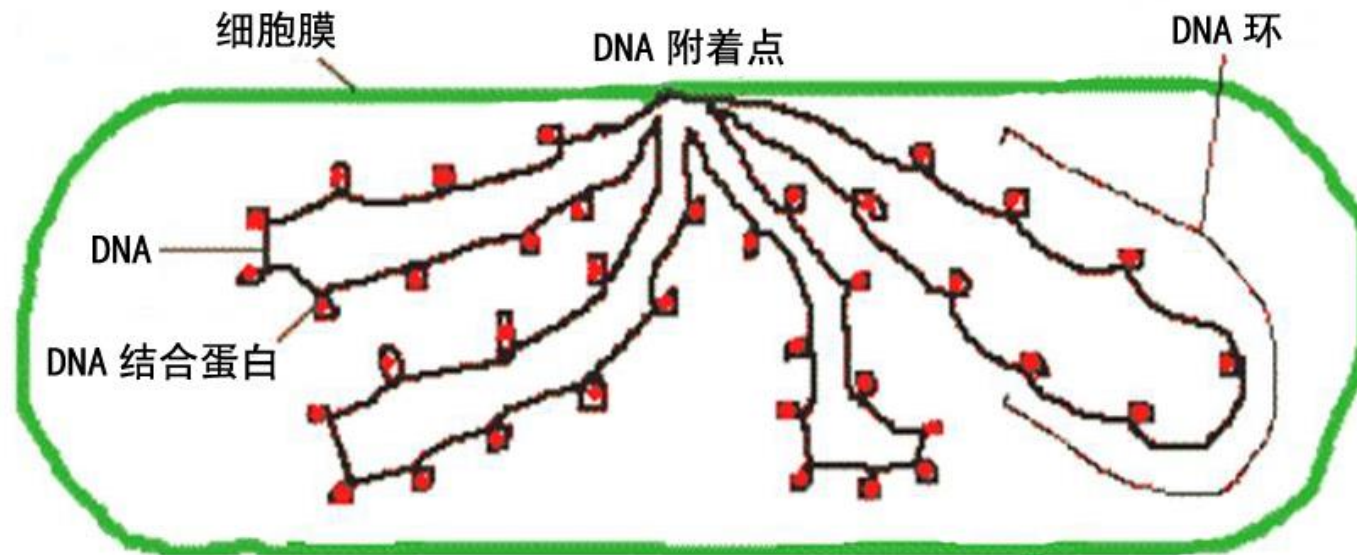
部分病毒基因组由**RNA**组成。

二、原核生物基因组结构特点

原核生物基因组为环状；
结构简单，小于真核生物基因组；
附着于细胞膜内表面形成“拟核”；
多为单拷贝基因。

质粒：细菌染色体外的双链环状**DNA**分子。

原核生物基因组拟核结构



质粒



三、真核生物基因组结构特点

细胞核基因组结构庞大，信息量巨大；
线粒体基因组主要呈环状。

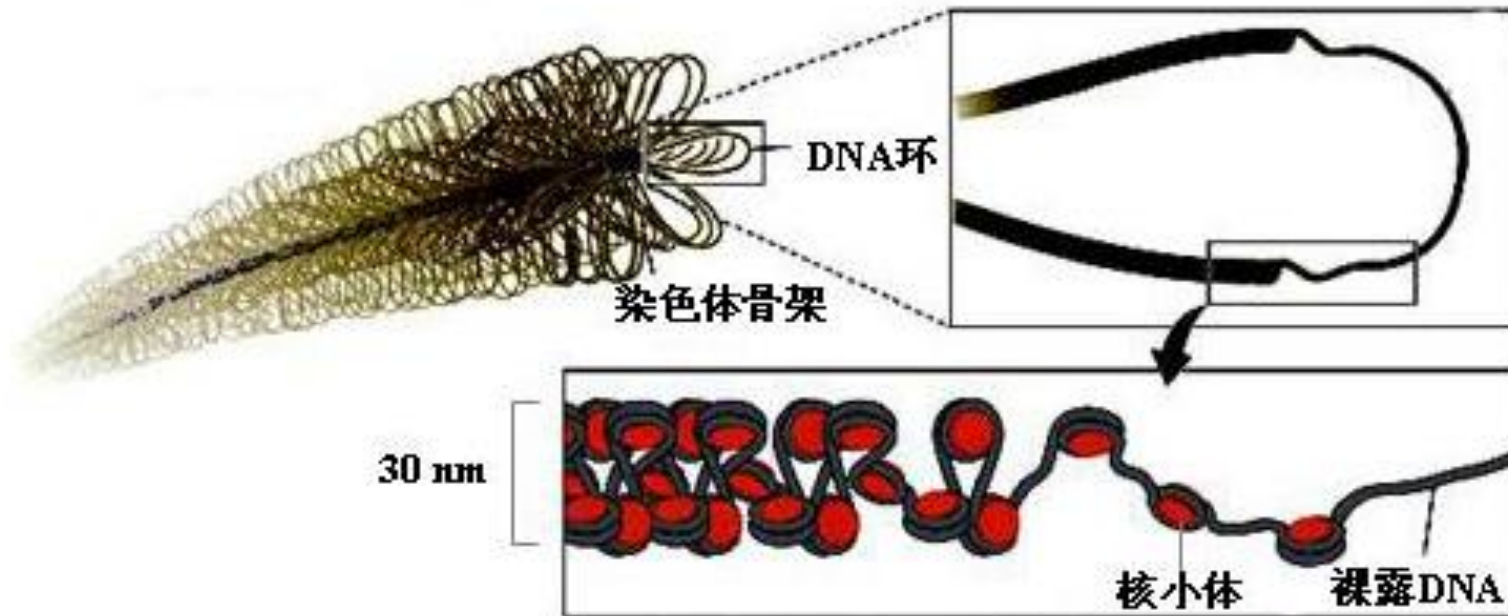
(一) 细胞核基因组

1. 线状DNA主要与组蛋白结合组装成染色质

核基因组线状DNA环绕由组蛋白H2A、H2B、H3和H4各两分子形成的八聚体而形成核小体；H1组蛋白与DNA结合而稳定核小体；核小体折叠呈锯齿状而形成直径为30nm的纤维；30nm纤维盘绕形成辐射状环。

组蛋白修饰（表观遗传学改变）影响染色质结构和基因表达。

核基因组DNA被有序压缩形成染色质



2. 核基因组中存在重复序列

非重复序列即单拷贝序列，基因组DNA序列中只有单一拷贝或少数几个拷贝，多为结构基因，其两侧为间隔序列和散在分布的重复序列。

重复序列主要包括组蛋白基因、免疫球蛋白基因、rRNA基因、tRNA基因和非编码序列；
重复序列的主要功能：调节基因表达。

(1) 高度重复序列

长度：2~300bp；

最高重复频率： 10^6 ；

反向重复序列：两个序列相同的互补拷贝在
同一DNA链上呈反向排列；

卫星DNA：散在分布，串联排列；

包括微卫星DNA、小卫星DNA、
端粒DNA。

(2) 中度重复序列

长度： $10^2 \sim 10^3\text{bp}$;

最高重复频率： 10^5 ;

短散在分布核元件： $70 \sim 300\text{bp}$ ，基因组内拷贝数达 $10^4 \sim 10^5$ ，如 *Alu* 序列家族。

长散在分布核元件： $5 \sim 7\text{kb}$ ，基因组内拷贝数达 10^4 ，如 *Kpn I* 序列家族。

3. 核基因组中存在多基因家族和假基因

真核生物细胞核基因组中常存在多基因家族和假基因。

(1) 多基因家族

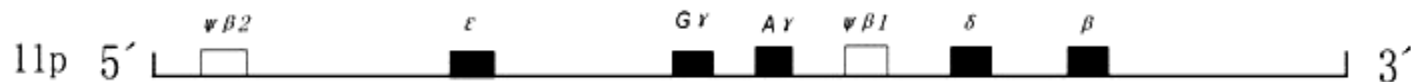
某一祖先基因经过重组和变异所产生的一组基因，序列高度同源、功能相似，成簇分布在同一染色体或不同染色体上，编码一组功能相关的蛋白质。
如人类珠蛋白基因簇。

人类珠蛋白基因簇

α 珠蛋白基因簇



β 珠蛋白基因簇



胚胎血红蛋白: $\zeta_2 \epsilon_2$, $\alpha_2 \epsilon_2$, $\zeta_2 \gamma_2$

胎儿血红蛋白: $\alpha_2 \gamma_2$

成人血红蛋白: $\alpha_2 \beta_2$, $\alpha_2 \delta_2$

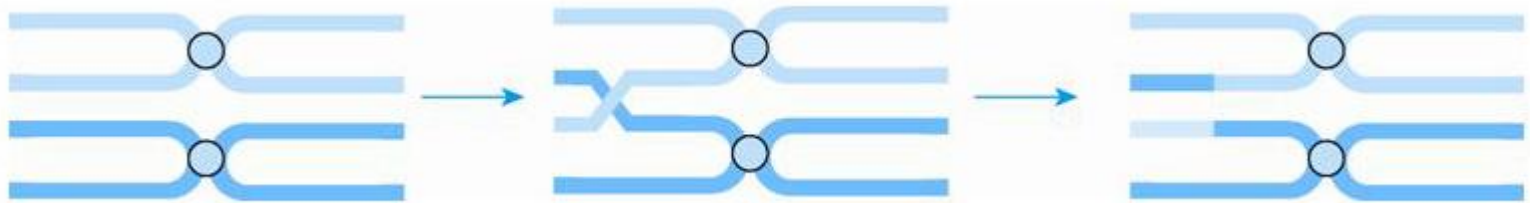
(2) 假基因

在长期的进化过程中，某些功能基因经缺失、易位或点突变等而改变结构，不再表达有活性的产物而成为假基因（ ψ ）。如珠蛋白基因族中的 $\psi\zeta$ 、 $\psi\alpha2$ 、 $\psi\alpha1$ 、 $\psi\rho$ 、 $\psi\beta2$ 、 $\psi\beta1$ 等假基因。

(二) 染色体重组

同源染色体之间可发生等量信息交换；
具有不同等位基因的同源染色体之间的信息交换
可致遗传连锁的差别。

同源染色体之间的信息交换



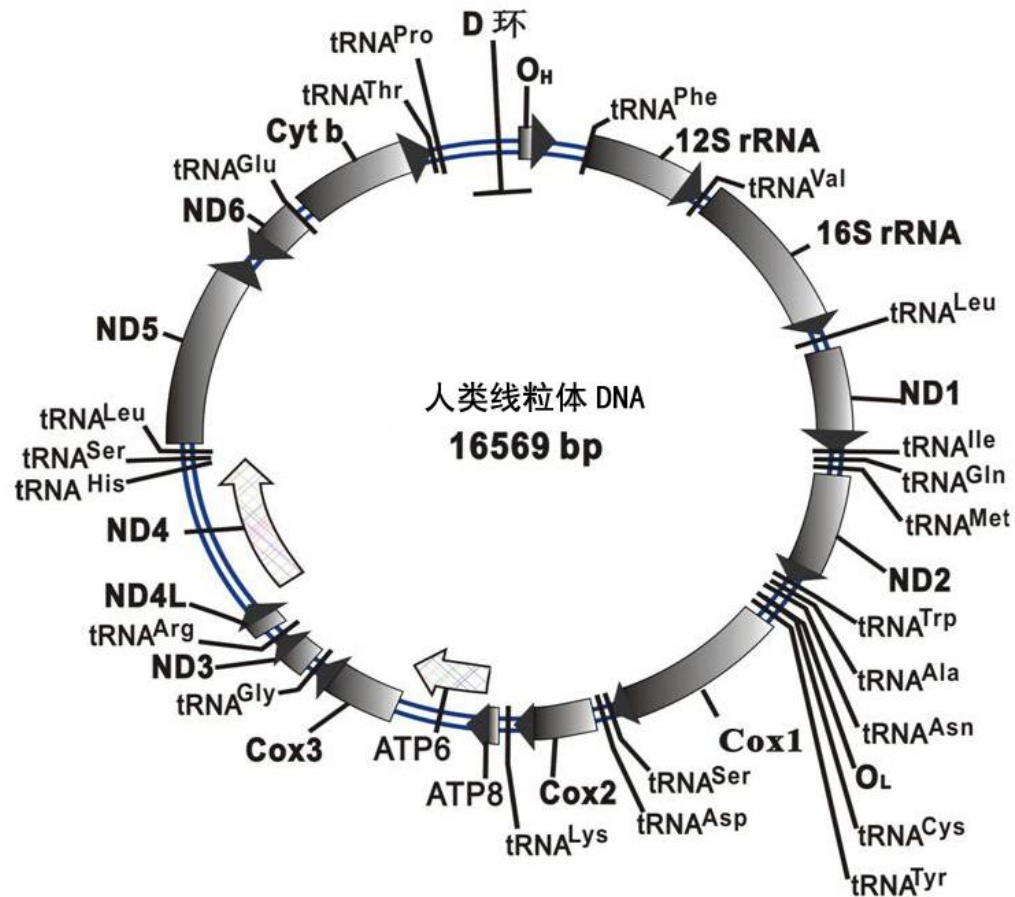
（三）线粒体基因组

线粒体**DNA**是核外遗传物质，可编码线粒体蛋白质。线粒体**DNA**是双链环状分子，无内含子，含少量非转录序列。

人类线粒体基因组含**37**个编码基因；另有约**900**种线粒体蛋白质由核基因组编码。

线粒体基因所携带的遗传密码与标准遗传密码有差别。人类线粒体**DNA**全部来自卵细胞，线粒体遗传性状由女性传递。

人类线粒体基因组结构



四、转座元件

转座元件指能够在一个**DNA**分子内部或两个**DNA**分子之间移动的**DNA**片段。

真核生物与原核生物的基因组转座元件结构相似。转座元件的两端具有反向或同向重复序列，中间部分有编码转位酶的结构基因。

转座元件能够反复插入到基因组中许多位点的特殊**DNA**序列中。

真核生物基因组中的一些转座元件是通过逆转录转座。

第三节 人类基因组计划

一、人类基因组计划的概念

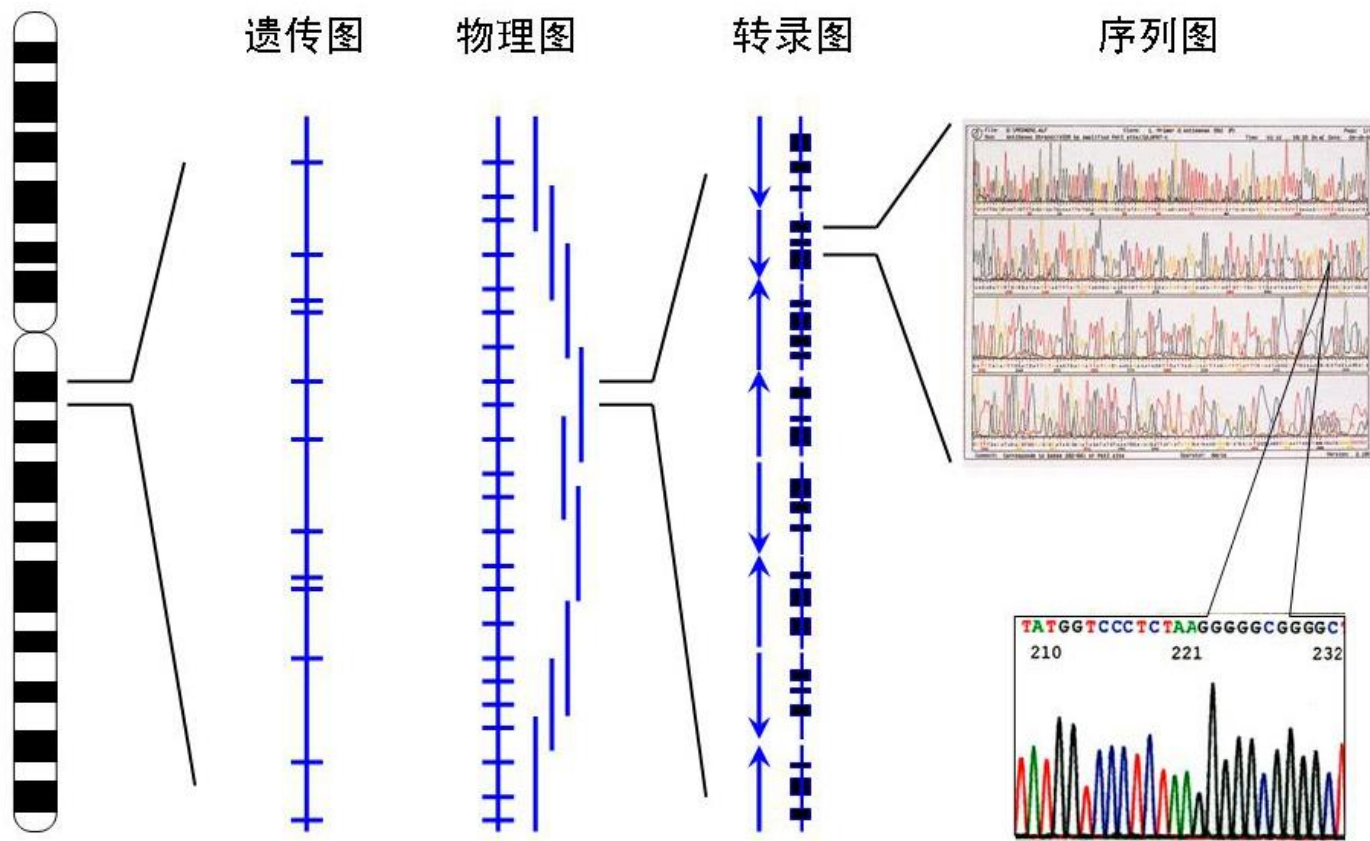
1985年美国科学家提出人类基因组计划，1989年美国成立国家人类基因组研究中心，1990年美国国会正式批准启动人类基因组计划，由美国、英国、日本、法国、德国和中国科学家参与完成。

人类基因组计划旨在测定人类**24**条染色体约**30**亿个碱基的精确序列，发现所有基因并确定其在染色体上的位置和核苷酸排列顺序；人类基因组计划、曼哈顿原子弹计划、阿波罗登月计划被誉为世界三大科学工程。

二、人类基因组计划的研究内容

人类基因组计划要完成四张图谱即遗传图、物理图、转录图和序列图，最终确定人类基因组**DNA**序列所含**30**亿个核苷酸的排列顺序。

人类基因组计划要完成的四张图谱



(一) 遗传图

遗传图又称连锁图，即以多态性标记为“路标”、以遗传学距离厘摩为“图距”的基因组图。

(二) 物理图

物理图是指标示各遗传标志之间物理距离的图谱。

人类基因组计划中的物理图是指染色体上限制性内切核酸酶识别位点或序列标签位点等的位置图。

(三) 转录图

转录图又称**cDNA**图或表达序列图，是以表达序列标签为标志绘制的图谱。转录图包括了几乎所有基因表达的**mRNA**序列，可以了解不同基因的时空特异性表达。

(四) 序列图

序列图即基因组**DNA**的核苷酸排列顺序图。2001年2月公布了初步的人类基因组图谱；2006年5月，英、美科学家最终完成了1号染色体全部**DNA**序列测定，宣布人类基因组**DNA**序列测定圆满完成。

三、人类基因组的特征

人类基因组序列中功能性蛋白质的编码基因为3万~4万个。

人类基因组序列中约有100个编码基因是经逆转录转座机制进入DNA结构的。

人类基因组含有数量众多的单核苷酸多态位点。

人类基因组序列中基因家族种类丰富。

四、人类基因组计划的意义

促进了生物技术的进步；
促进了医学研究的发展；
促进了模式生物基因组研究；
促进了学科交叉与重组；
创建了生物信息学；
引发了基因争夺、基因歧视等社会问题。

思考题：

1. 解释可转录序列和调节序列。
2. 试述真核生物基因和原核生物基因的结构特点。
3. 试述真核生物基因组的结构特点。
4. 试述真核生物基因组重复序列的特征。
5. 试述人类基因组计划的研究内容和意义。