



四川大學  
SICHUAN UNIVERSITY



# 分子考古

Molecular Archaeology

张帆

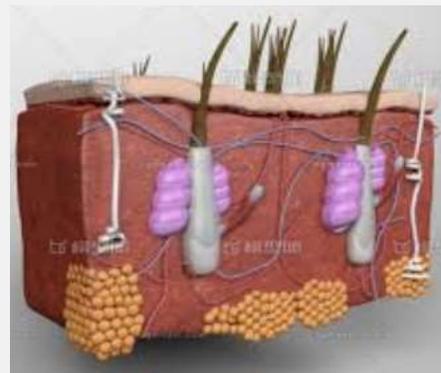
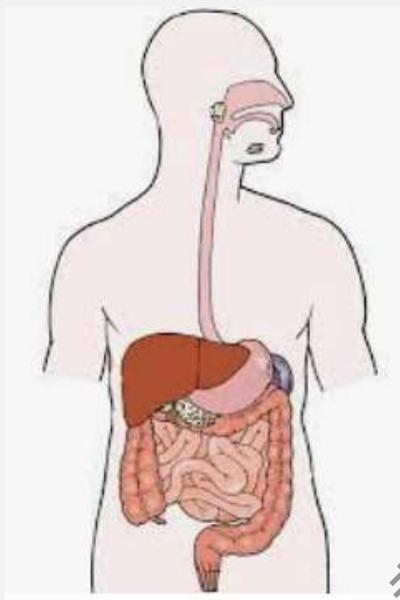
电话: 15948097104

Email:

[fzhang92@scu.edu.cn](mailto:fzhang92@scu.edu.cn)

# 细胞是生命活动的基本单位

人体 → 系统 → 器官 → 组织 → 细胞



细胞膜、细胞质、**细胞核 (DNA)**



## 一、DNA结构

- (1) 核酸分子的组成：**五碳糖、磷酸基团、含氮碱基[ATGCU]**
- (2) 连接方式：**氢键、磷酸二酯键**
- (3) 复制方式：**半保留复制**
- (4) **DNA具有双螺旋结构**  **染色体**



## 二、染色体数目

- 人是2倍体生物 (diploid,  $2n$ )，体细胞内有23对染色体，46条，包括22对常染色体和一对性染色体。
- 性染色体包括：X染色体和Y染色体。
- 男性：22对常染色体+XY。
- 女性：22对染色体+XX。

必需掌握

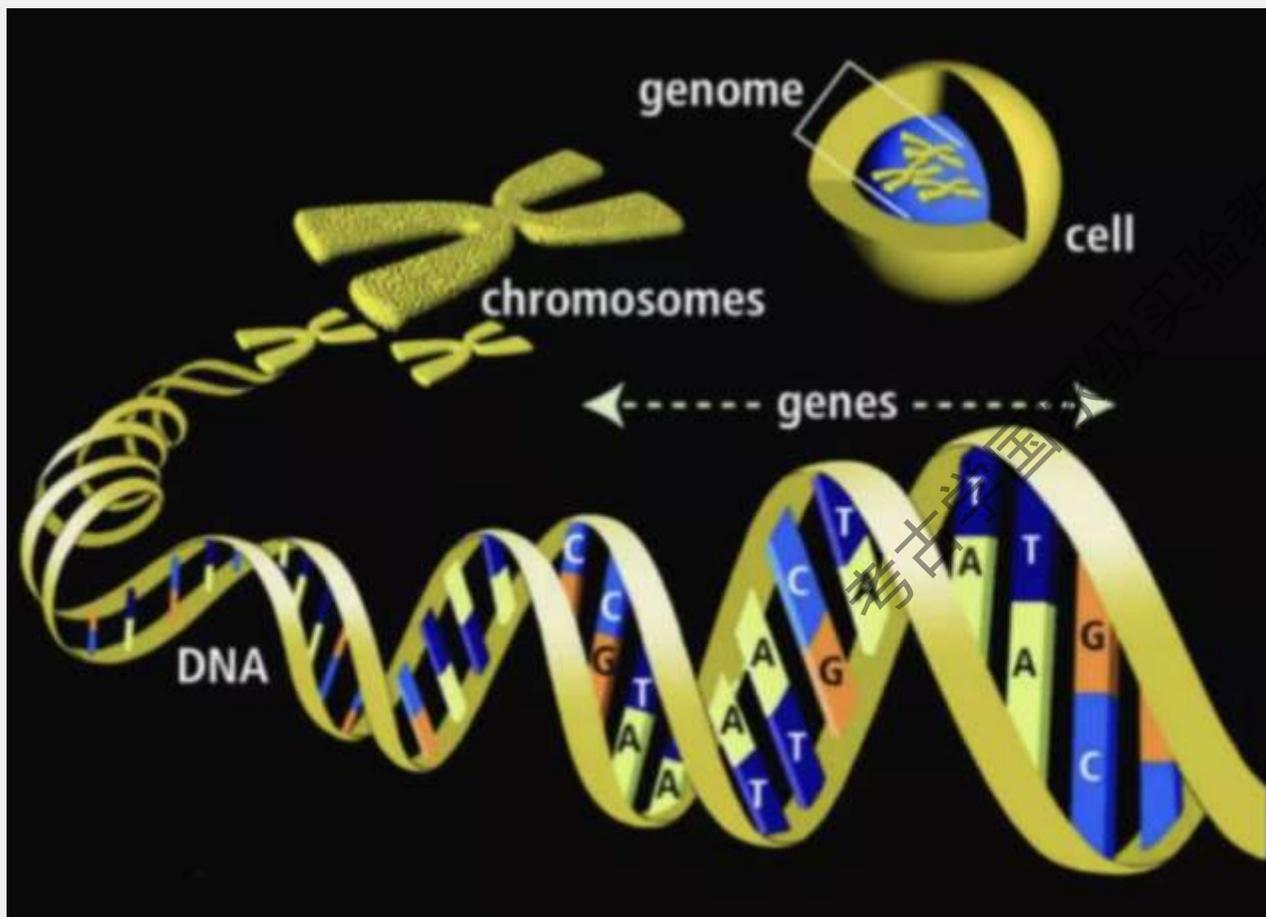




PART 04

# 基因和基因组

## 一、基因 (gene) 的概念



合成有功能的蛋白质或RNA的全部DNA，包括编码蛋白质或RNA的核苷酸序列及保证转录所必需的调控序列。



## 二、基因组 (genome) 的概念

**基因组 (genome)：生物体内所有遗传物质的总和。**

- 生物体遗传信息的“总词典”
- 生物体控制发育的“总程序”
- 生物演化历史的“总档案”

核基因组 (nuclear)

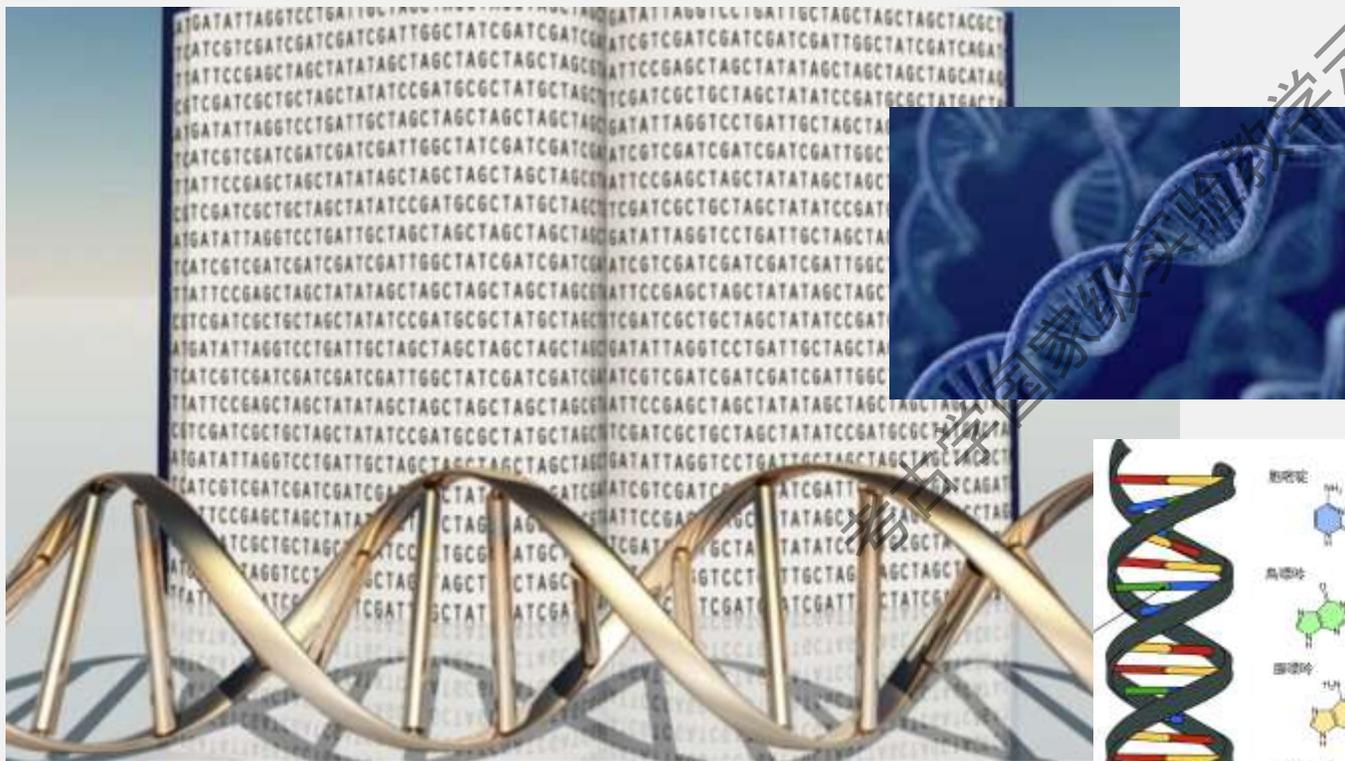
线粒体基因组 (mitochondrial)

叶绿体基因组 (chloroplast)

**genome = GENe + chromosOME**



## 二、基因组 (genome) 的概念



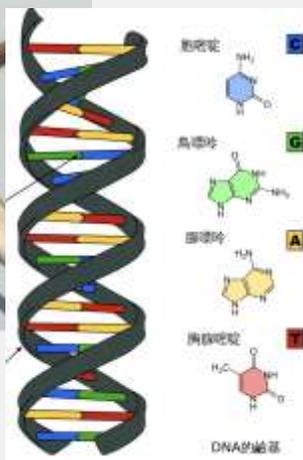
Genome: *Book*

Chromosome: *Chapter*

DNA strands: *Sentences*

Genes: *Words*

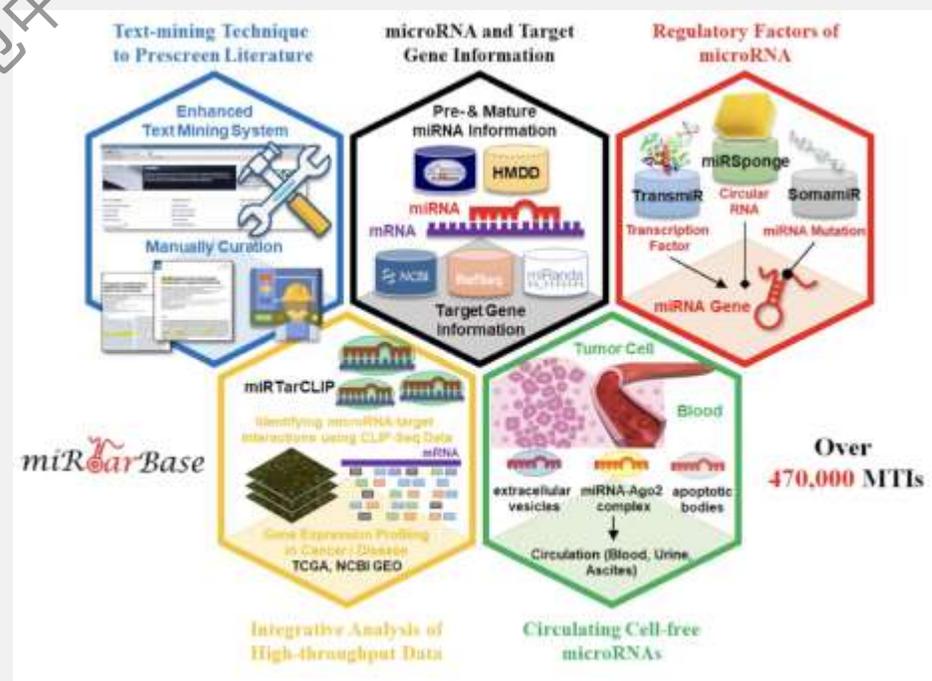
Base-pairs: *Letters*



## 三、基因组学 (genomics) 的概念

基因组学：以生物体全部基因为研究对象，研究基因组结构、功能、演化、组装、编辑等方面的交叉学科。

- 拼接组装生物体全部基因组序列
- 系统解析所有基因功能
- 明确基因间的相互作用
- 揭示基因在生物体生长发育的调控
- 阐明基因组的演化规律



# 基因组学



Digital Health: How Genomic Big Data is Transforming Health Care

HEALTH FUTURES:  
**AI & GENOMICS**

- Transcriptomics (转录组学)
- Proteomics (蛋白质组学)
- Epigenomics (表观基因组学)
- Metagenomics (宏基因组学)
- Pangenomics (泛基因组学)
- Computational Genomics (计算基因组学)
- Comparative Genomics (比较基因组学)
- Functional Genomics (功能基因组学)
- Population Genomics (群体基因组学)
- Systematics Genomics (系统基因组学)
- Evolutionary Genomics (进化基因组学)
- Single-cell Genomics (单细胞基因组学)
- Cancer Genomics (癌症基因组学)
- Pharmacogenomics (药物基因组学)
- Neurogenomics (神经基因组学)
- Health Genomics (健康基因组学)
- Forensic Genomics (法医基因组学)
- Crop Genomics (作物基因组学)
- Archaeogenomics (考古基因组学)
- .....

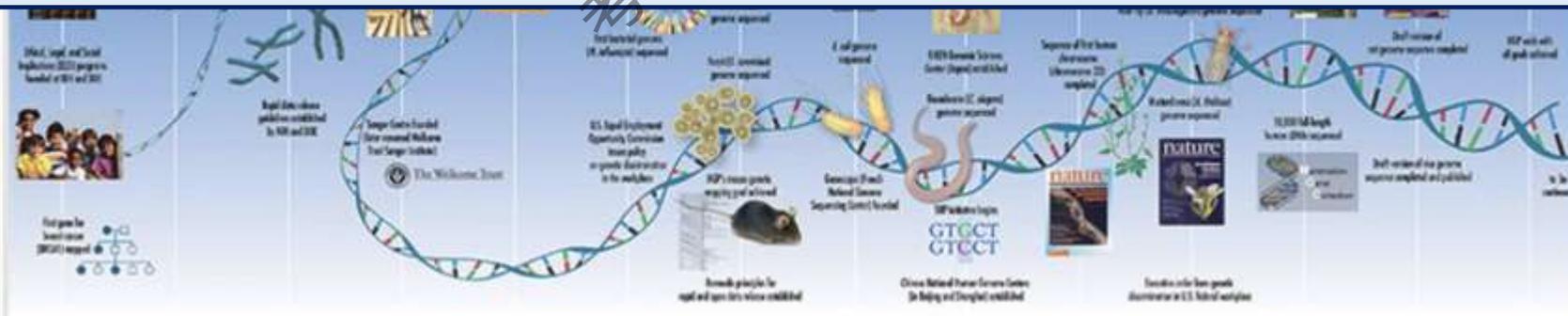


四川大学  
SICHUAN UNIVERSITY

## 四、人类基因组计划 (Human Genome Project)



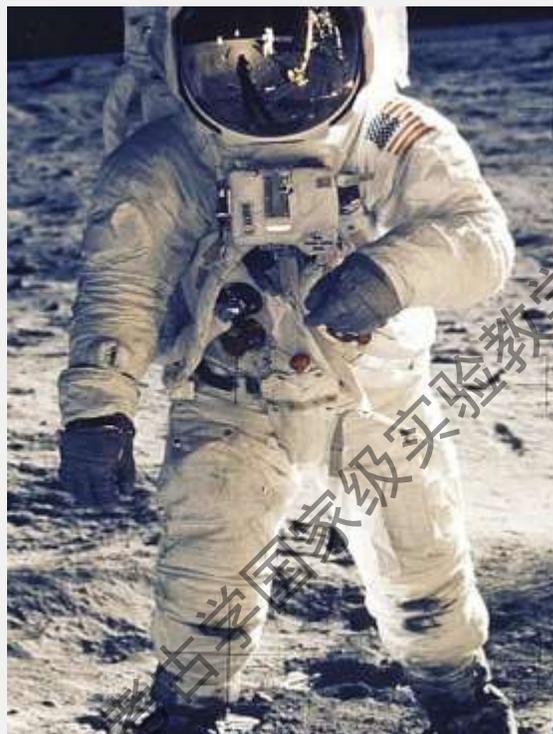
基因组学：伴随人类基因组计划孕育而生





曼哈顿原子弹计划  
(1942-1945)

200亿美元  
1945年美国  
1964年中国



阿波罗登月计划  
(1961-1972)

255亿美元  
1969年美国  
? 中国



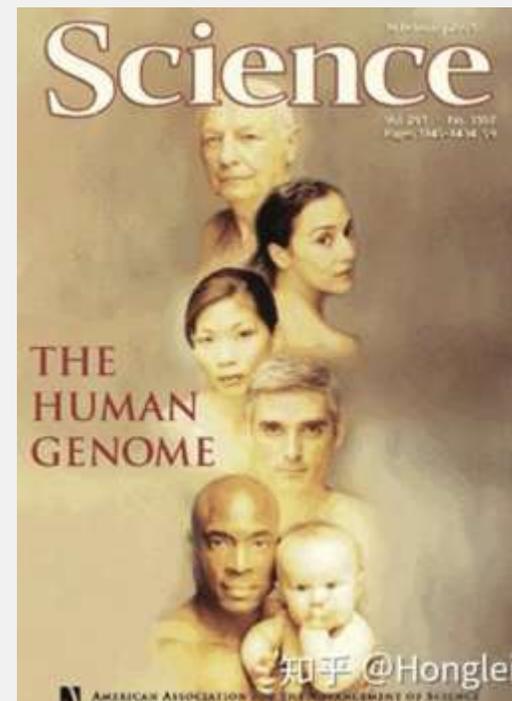
人类基因组计划  
(1990-2003)

30亿美元  
美国、英国、法国  
德国、日本、中国



## 四、人类基因组计划 (Human Genome Project)

- 人类基因组计划 (Human Genome Project, HGP), 1985年提出, 1990年正式启动。
- 美、英、法、德、日、中
- 我国于1999年加入该计划, 承担其中**1%**的任务, 即人类3号染色体上约30Mb的测序任务。



## 四、人类基因组计划 (Human Genome Project)

1984年12月,在美国犹他州盐湖城滑雪胜地阿尔他,由美国能源部资助的一次环境诱变和致癌物防护国际会议上,到会代表在讨论中提出了一个问题:在受原子弹爆炸辐射伤害的人群中,能检测到突变的比率比预期低三分之二。**有什么新办法可以非常有效地,直接地检测出人类基因的突变?**或者说,有没有新方法可在日本广岛,长崎原子弹爆炸后的幸存者及其子女中**直接检测体内突变的基因?**

这是科学家首次讨论人类基因组计划。



## 四、人类基因组计划 (Human Genome Project)

1986年，1975年医学和生理学诺贝尔奖得主，美国Salk Institute研究所癌症研究员杜贝可 (Renato Dulbecco) 在“Science”上发表题为“**癌症研究的转折点：定出人类基因组序列**”一文，引起了美国社会大众的广泛关注，并使基因组测序计划的支持者和反对者进行了一场为数年的争论。

杜贝可提出了两条基因搜寻路线：

- DNA测序
- 基因组作图



## 四、人类基因组计划 (Human Genome Project)

人类基因组计划虽然是一个极富创意的想法，但超出了当时科学技术发展的实际水平，因而遭到各方面的怀疑和反对。主要原因在以下几点：

- **科学上的依据**：人类基因组中编码蛋白质的DNA序列约占总DNA的**2%**，是否值得花很多钱去测序整个人类基因组。
- **测序技术不成熟**：以当时DNA凝胶平板电泳测序水平，一个人每天最多只能完成1000bp测序。按人类基因组DNA总量30亿对碱基数计算，需要1000人工作3000年。
- **挤占大量研究经费**：人类基因组计划预计耗费30亿美金，势必占用生命科学其它研究领域的经费，对生物科学的整体研究产生负面影响



## 四、人类基因组计划 (Human Genome Project)

HGP的最初目标：15年内（1990-2005）投入30亿美元，完成人类24条染色体的30亿个核苷酸序列的分析。

终极目标：解码生命、了解生命、认识种属之间和个体之间存在差异的起因、认识疾病产生的机制以及长寿与衰老等生命现象、为疾病的诊治提供科学依据。



# 人类基因组“工作框架图”完成

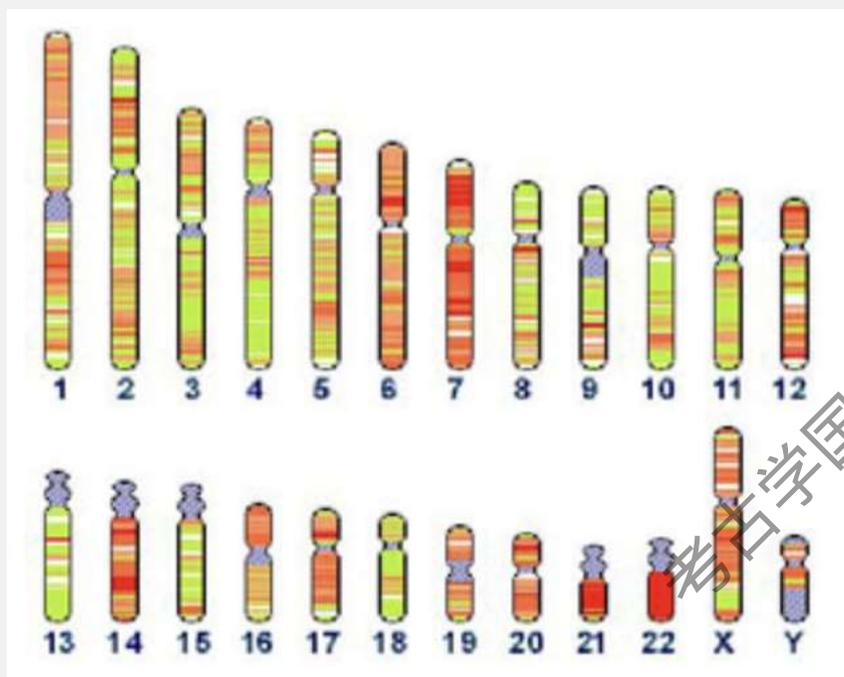
2000年6月26日美国总统克林顿宣布  
人类基因组草图绘制完成



1990-2001年，人  
类基因组测序完  
成度超过90%，比  
预期提前了四年



## 四、人类基因组计划 (Human Genome Project)



截止2006年5月，人类基因组计划已完成22条常染色体和X, Y性染色体的精确测序与初步解读。

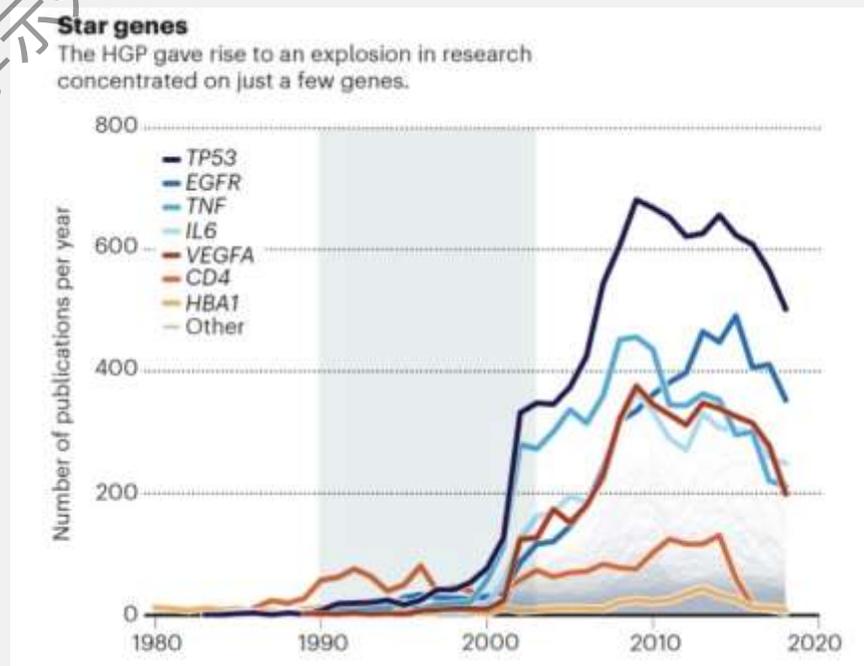
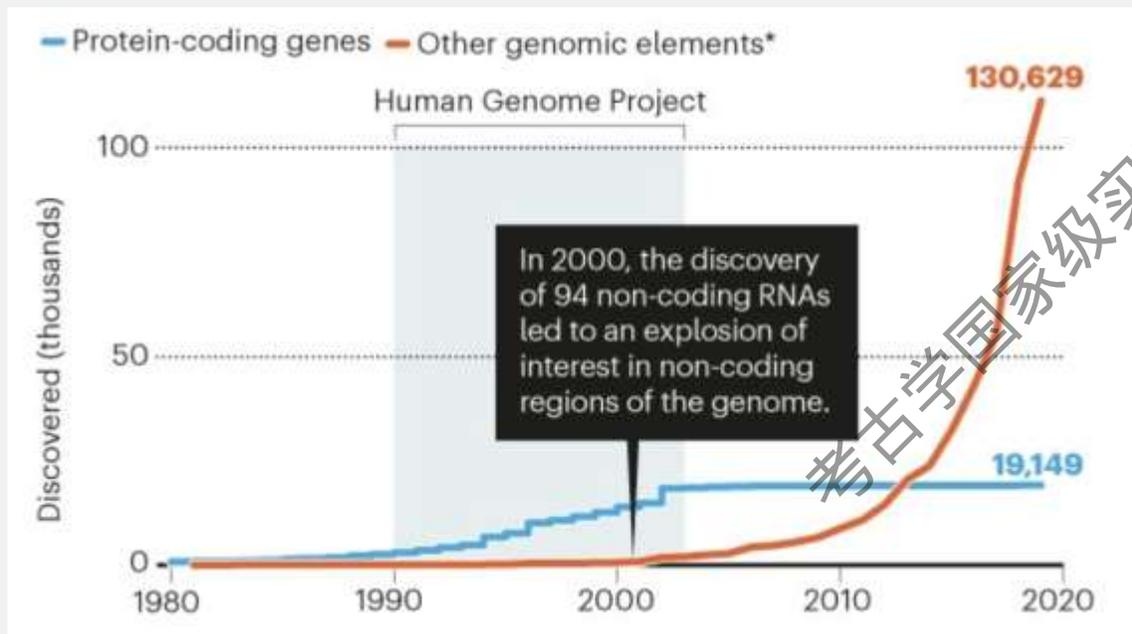
Human Chr.20, completed, February, 2002  
Human Chr.14, completed, January, 2003  
Human Chr.Y, completed, October, 2003  
Human Chr.7, completed, October, 2003  
Human Chr.6, completed, October, 2003  
Human Chr.13, completed, March, 2004  
Human Chr.19, completed, March, 2004  
Human Chr.9, completed, May, 2004  
Human Chr.10, completed, May, 2004  
Human Chr.X, completed, March, 2005  
Human Chr.2-4, completed, April, 2005  
Human Chr.1, completed, May, 2006

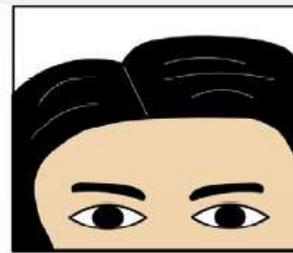
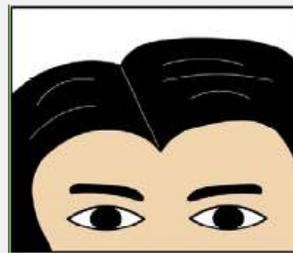
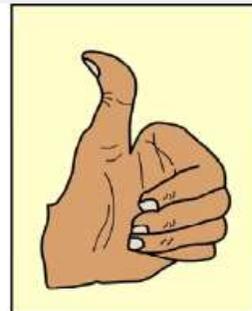
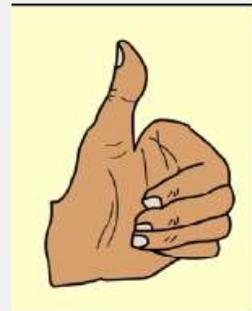
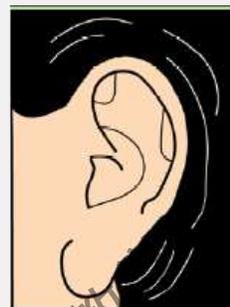
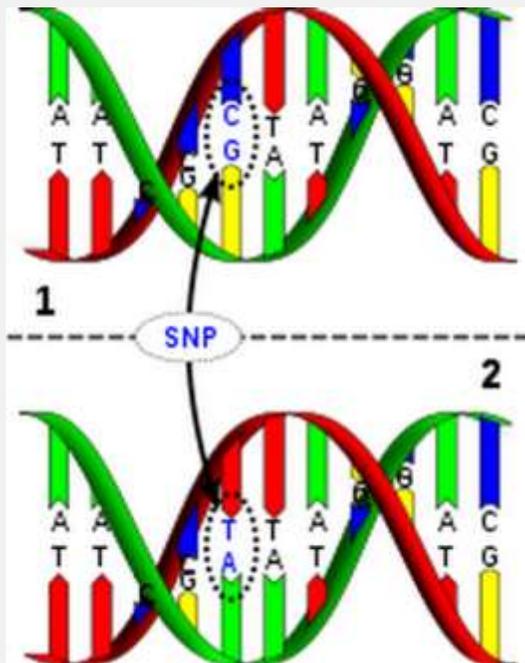


## 四、人类基因组计划 (Human Genome Project)



## 四、人类基因组计划 (Human Genome Project)





- 数千万个变异
- 每两人差别：~300万SNP



OR6A2嗅觉基因，决定着人类对醛性化学成分的嗅觉敏感程度。而香菜散发的那股令人闻风丧胆的“臭味”就包含这种醛性化学成分。





**DEC2基因：**这个基因调节生物节律，可控制荷尔蒙释放的时间，影响进食和睡眠节律。携带DEC2基因特定突变的人，平均每天只需要睡6.25小时，而未发生这种突变的人，平均需要的睡眠时间为8.06小时；

**ADRB1基因：**这个基因负责编码肾上腺素能 $\beta$ 受体，突变后会导导致受体稳定性下降，功能受到影响。研究发现，发生此突变的小鼠，可以促进自然短睡眠，使大脑更容易被唤醒，从而保持更长时间的清醒；



- ✓ 乳糖不耐受
- ✓ 咖啡因代谢
- ✓ 汗臭味敏感性
- ✓ 轻度脸盲
- ✓ 皮肤弹性
- ✓ 解酒能力
- ✓ 腋臭
- ✓ 背书能力
- ✓ 抗痘能力
- ✓ 甜味敏感度
- ✓ 耳垂形态
- ✓ 少白头
- ✓ 重度痛经(女性特有)
- ✓ 腋下多汗
- ✓ 小腿毛发(男性特有)
- ✓ 招风耳
- ✓ 苦味敏感度
- ✓ 抗压能力
- ✓ 疼痛敏感性
- ✓ 皮肤抗光老化能力
- ✓ 皮肤抗氧化能力
- ✓ 血型
- ✓ 皮肤屏障
- ✓ 口腔溃疡易感性
- ✓ 睡觉多动
- ✓ 招蚊指数
- ✓ 卷发
- ✓ 香菜喜好
- ✓ 吃辣出汗
- ✓ 小臂毛发(女性特有)
- ✓ 美人尖
- ✓ 耳垢类型
- ✓ 情景记忆能力
- ✓ 好奇心指数
- ✓ 骨密度
- ✓ 端粒长度
- ✓ 词汇记忆
- ✓ 香味敏感度
- ✓ 防晒黑能力
- ✓ 脑容量
- ✓ 强光打喷嚏
- ✓ 易困程度
- ✓ 睡觉磨牙
- ✓ 高度近视
- ✓ 小臂毛发(男性特有)
- ✓ 连眉
- ✓ 新生儿体重(女性特有)
- ✓ 长寿可能性
- ✓ 错误学习
- ✓ 音乐感知能力
- ✓ 利他主义
- ✓ 抗寒能力
- ✓ 酒量
- ✓ 皮肤锁水能力
- ✓ 男性早发性脱发(男性特有)
- ✓ 基因身高
- ✓ SARS 冠状病毒易感性
- ✓ 单/双眼皮
- ✓ 雀斑
- ✓ 小腿毛发(女性特有)
- ✓ 希腊脚

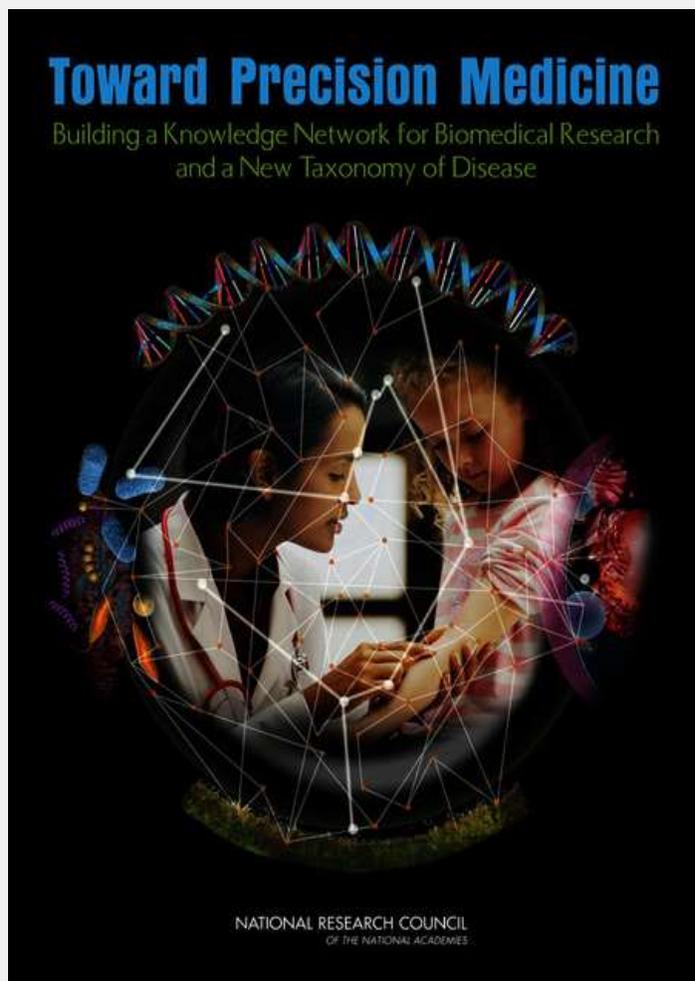
23mofang



## 遗传风险检测内容

- ✓ 心脏骤停
- ✓ 白癜风
- ✓ 嗜睡症
- ✓ 慢阻肺
- ✓ 迟发性阿尔兹海默病
- ✓ 酒精依赖症
- ✓ 继发性肾病综合征
- ✓ 腰椎间盘突出症
- ✓ 遗传过敏性皮炎
- ✓ 痛风
- ✓ 银屑病
- ✓ 毒性弥漫性甲状腺肿
- ✓ 癫痫
- ✓ 不宁腿综合征
- ✓ IgA肾病
- ✓ 十二指肠溃疡
- ✓ 遗传性血色素沉着症 (HFE相关)
- ✓ 慢性肾病
- ✓ 2型糖尿病
- ✓ 强直性脊柱炎
- ✓ 多囊卵巢综合症(女性特有)
- ✓ 缺血性脑卒中
- ✓ 高血压
- ✓ 高甘油三酯血症
- ✓ 多发性硬化症
- ✓ 原发性胆汁性肝硬化
- ✓ 肺纤维化
- ✓ 老年性黄斑变性
- ✓ 慢性萎缩性胃炎
- ✓ 过敏性鼻炎
- ✓ 甲状腺功能减退症
- ✓ 心房颤动
- ✓ 妊娠期糖尿病(女性特有)
- ✓ 冠心病
- ✓ 年龄相关性白内障
- ✓ 类风湿性关节炎
- ✓ 扩张型心肌病
- ✓ 双相情感障碍
- ✓ 脂肪肝
- ✓ 听力减退
- ✓ 哮喘
- ✓ 胆结石
- ✓ 脊柱侧凸
- ✓ 乳糜泻
- ✓ 剥脱性青光眼
- ✓ 帕金森病
- ✓ 肾结石



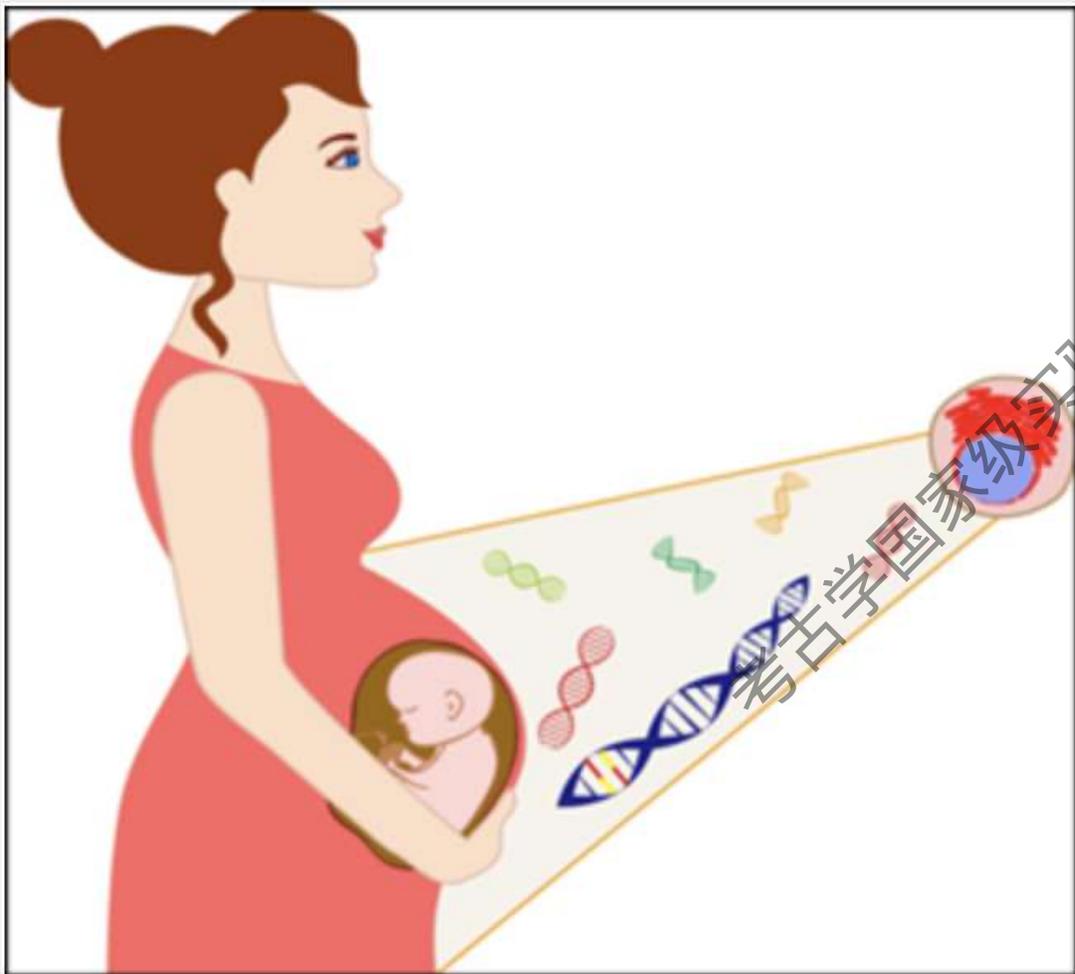


基础	大样本研究获得疾病分子机制的知识体系
依据	组学数据
	患者个体特征 <ul style="list-style-type: none"> <li>• 基因型</li> <li>• 表型</li> <li>• 环境</li> <li>• 生活方式</li> </ul>
手段	<ul style="list-style-type: none"> <li>• 现代遗传学</li> <li>• 分子影像学</li> <li>• 生物信息学</li> <li>• 临床医学</li> </ul>
目标	<ul style="list-style-type: none"> <li>• 精准预防</li> <li>• 精准诊断</li> <li>• 精准治疗</li> </ul>

# 精准医学



产前诊断：有用吗？  
智商税？



# 基因、环境和人类疾病

基因型

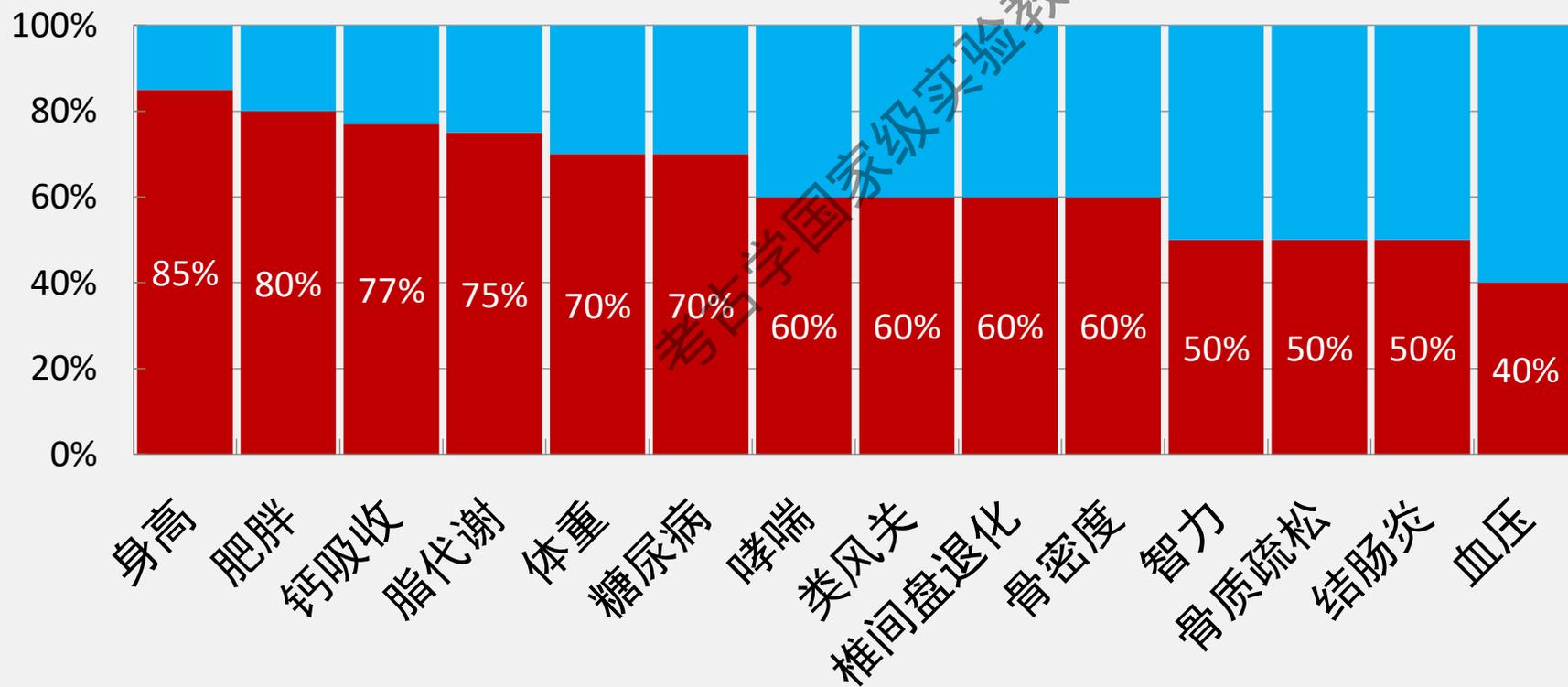
×

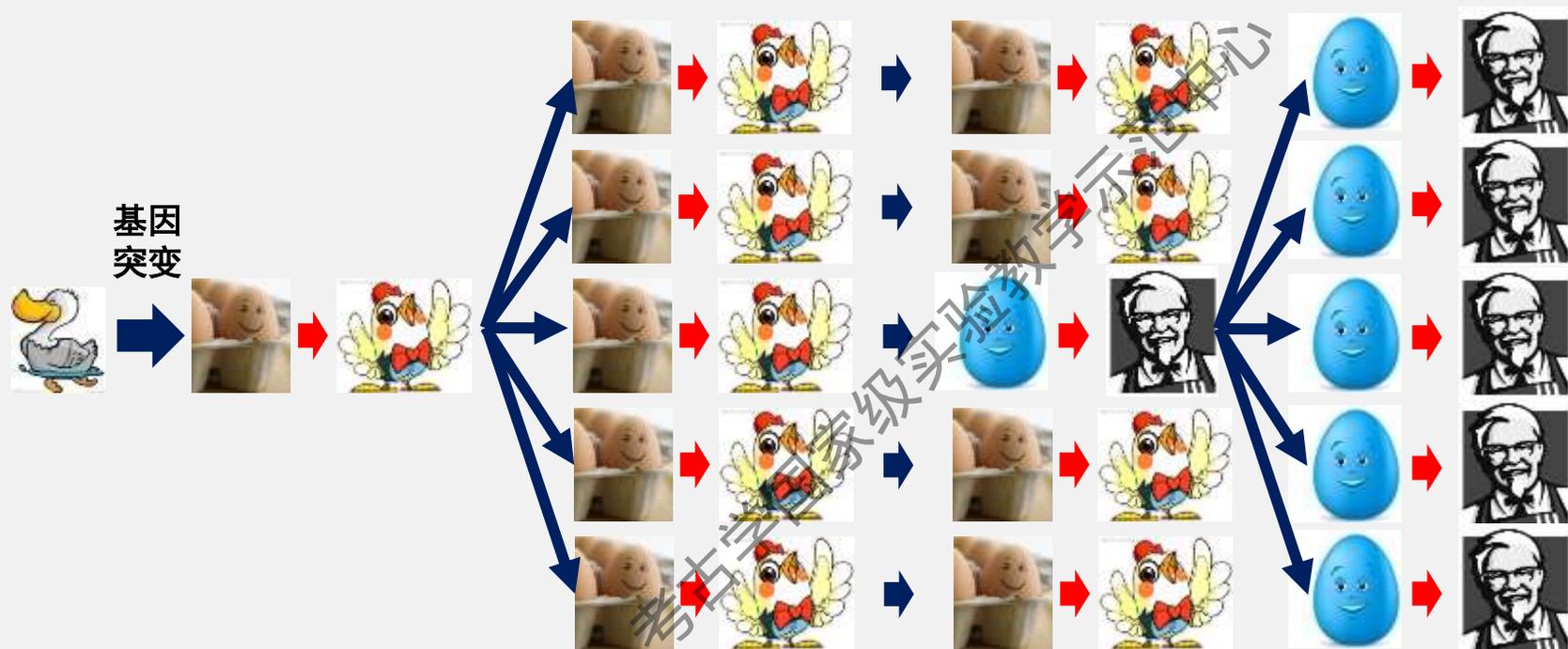
环境

=

表型

$$\text{遗传度} = \frac{\text{遗传方差}}{\text{表型方差}}$$



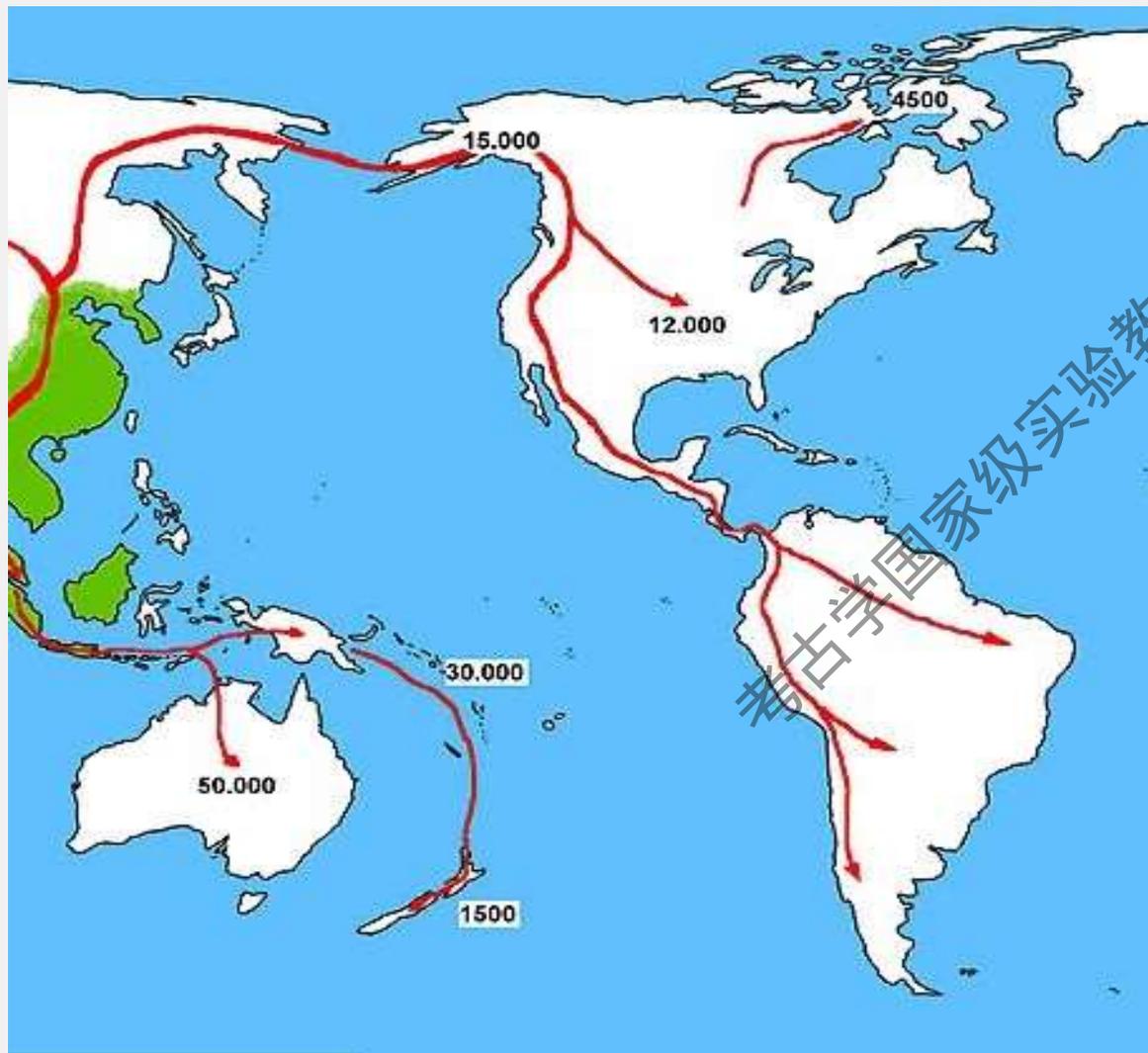


基因突变

自然选择

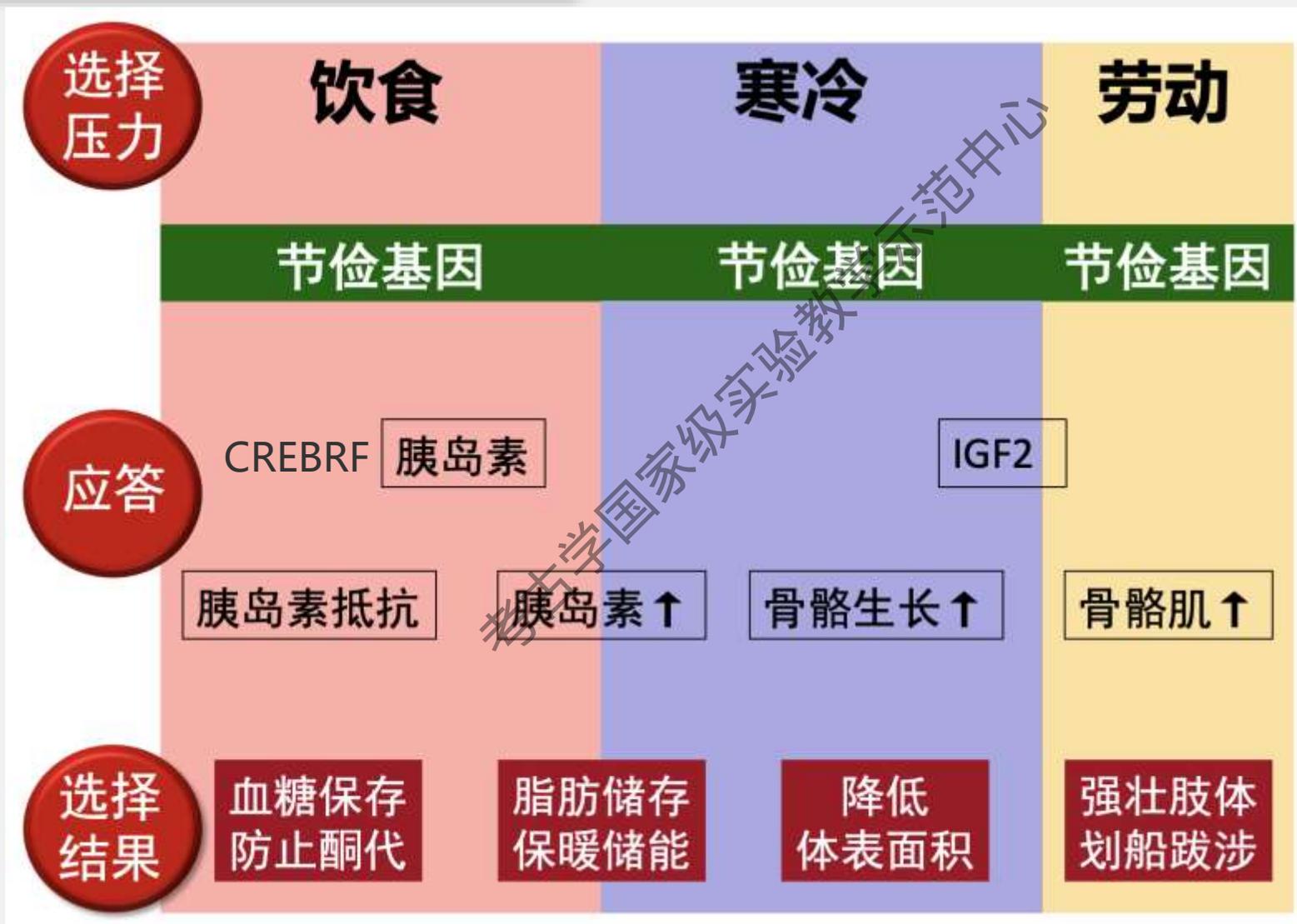
适应

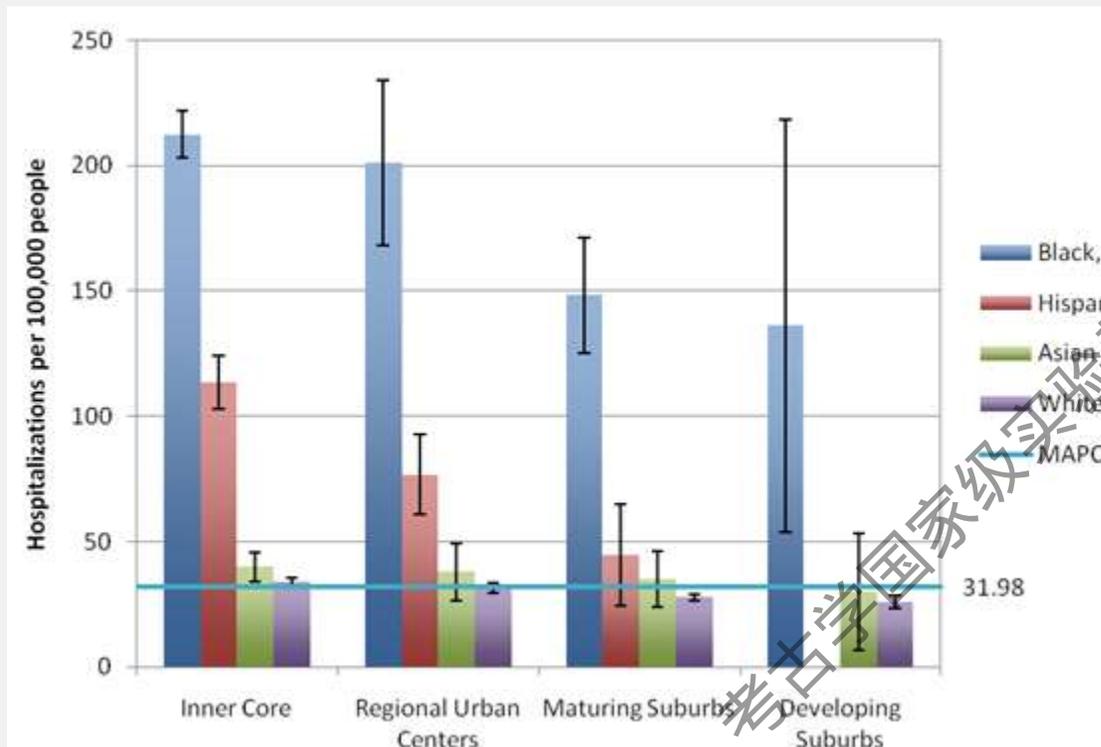




艰难的旅程  
严寒  
食物匮乏

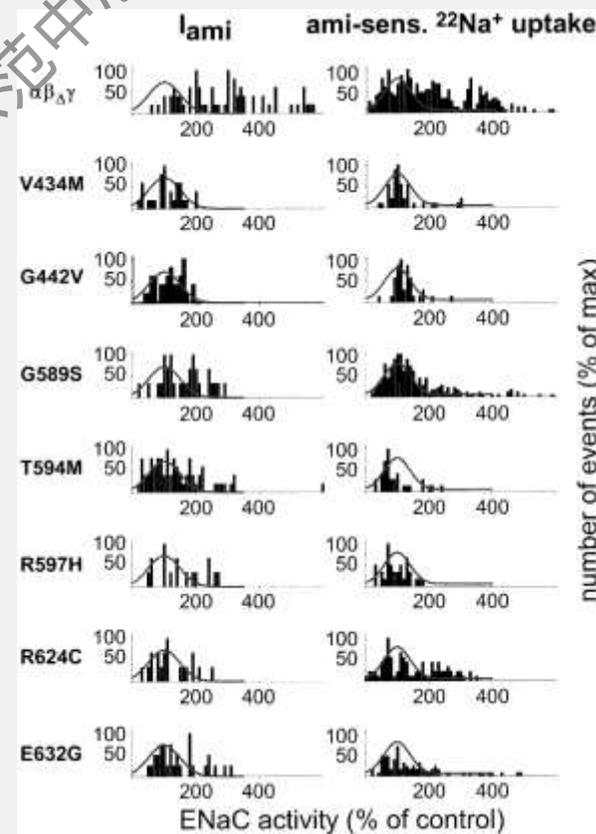
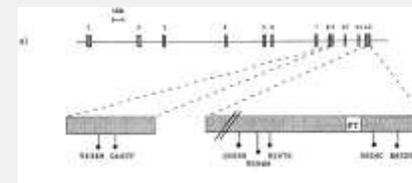
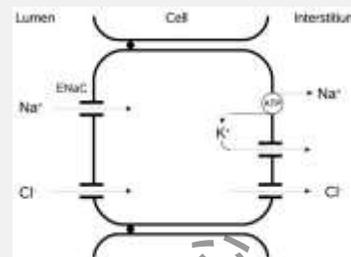




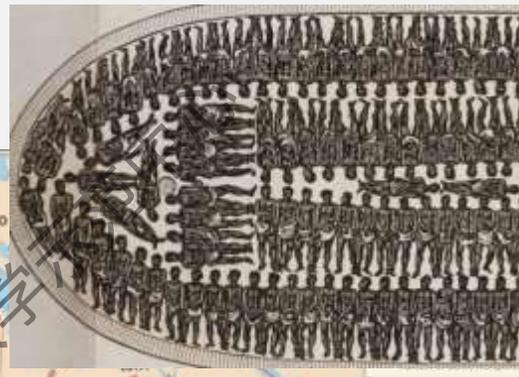
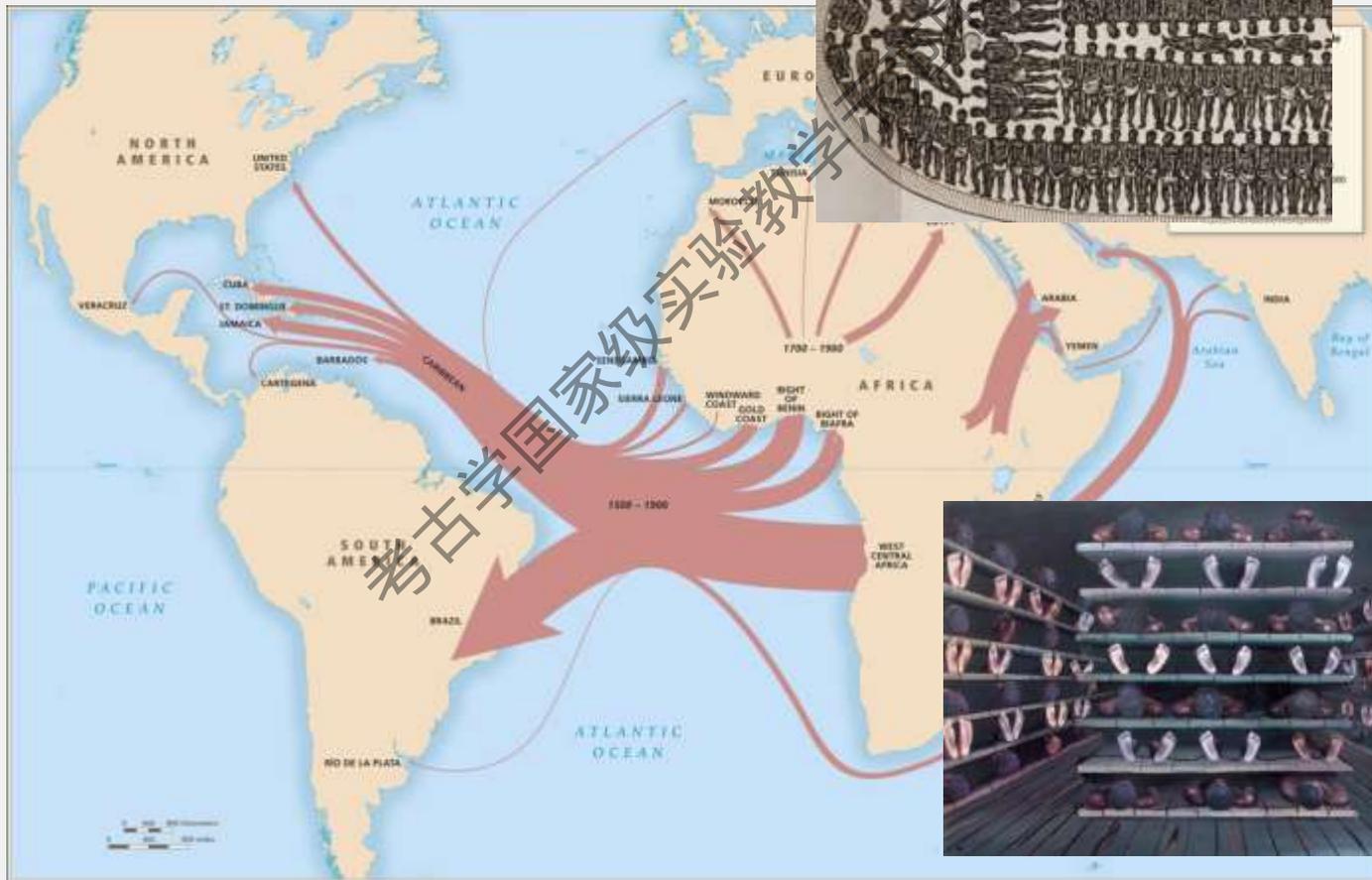


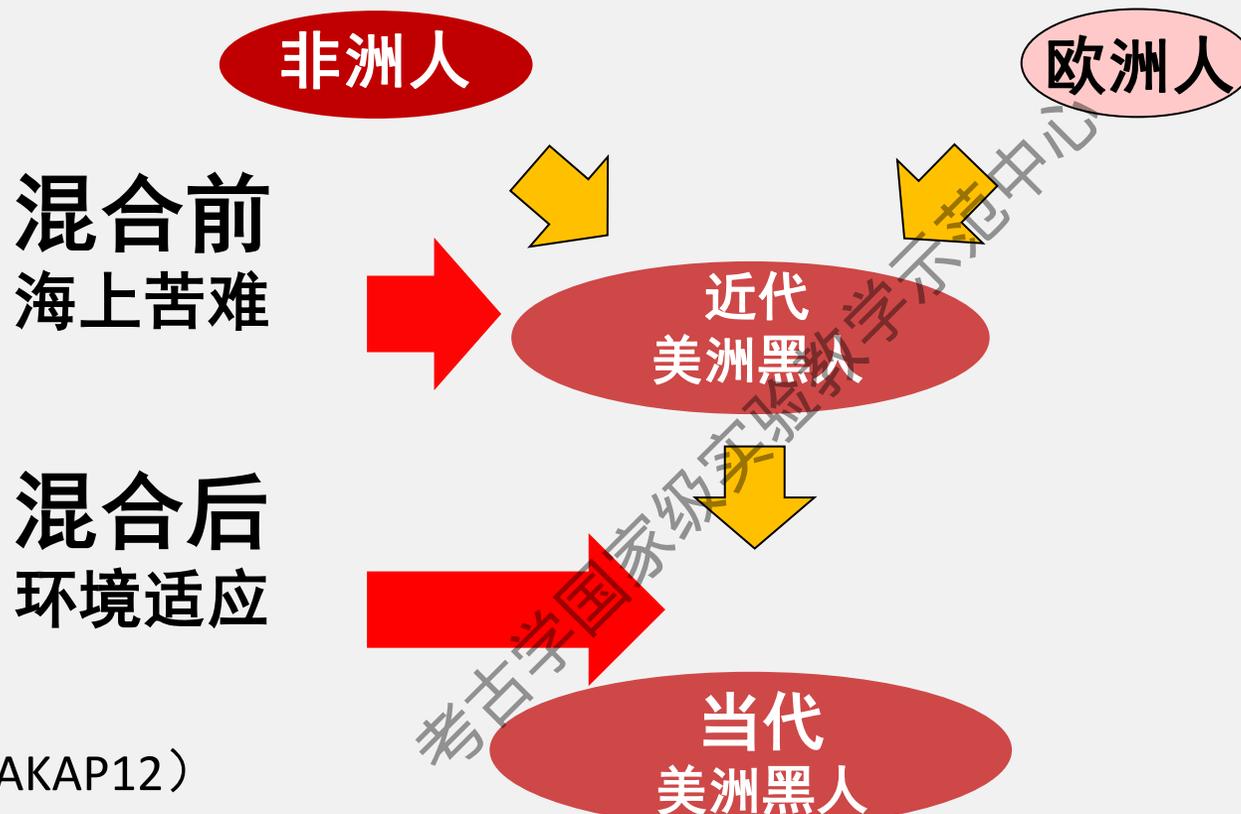
### 美国黑人与高血压：

- 发病率高于其他人群
- 发病早于其他人群



## 西非奴隶贩运路线





- 高血压 (ZP4, AKAP12)
- 前列腺癌 (PSCA)
- 糖尿病 (AKR7A2)
- 抗疟疾能力的丢失 (CD36, HBB, HLA-B)
- 抗流感 (EIF2AK2) (欧洲人的贡献)

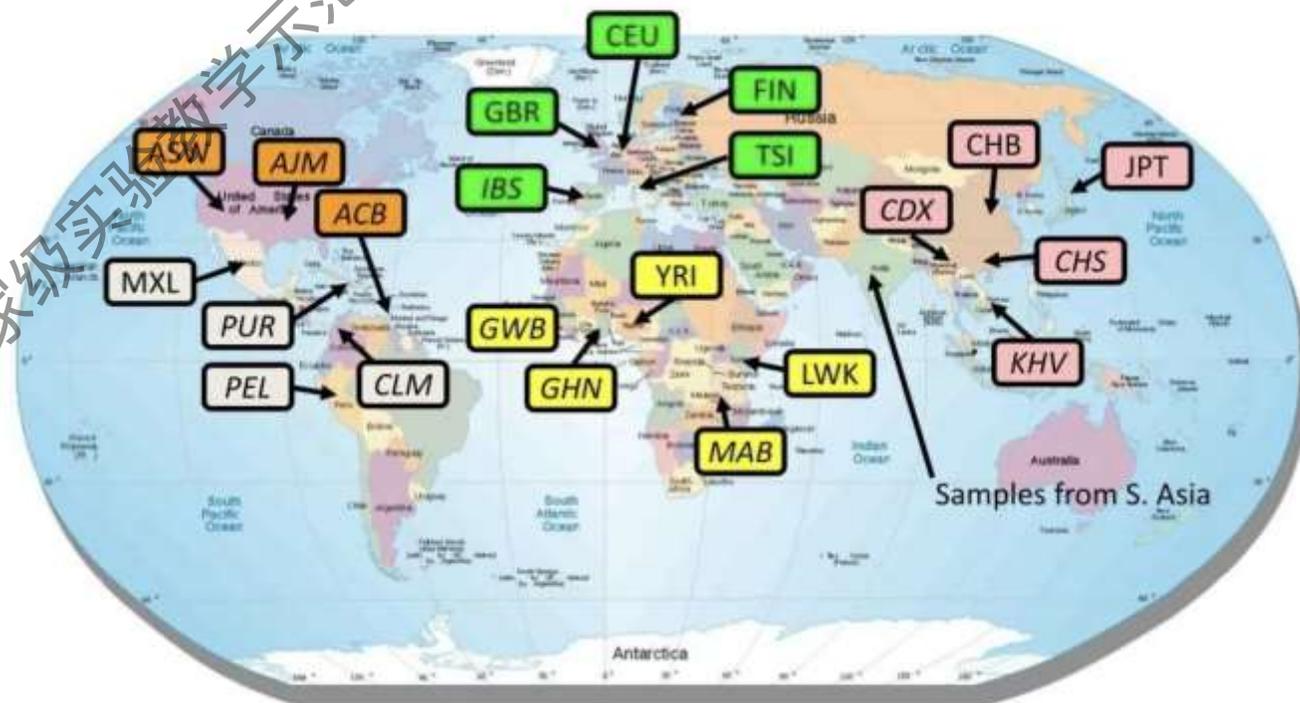
- 样本: 1,890 美洲黑人, 3,320 对照
- 数据: 491,526 位点

*Mol Biol Evol* 2007  
*AJHG* 2012  
*Genome Res* 2012  
*Hum Mol Genet* 2012

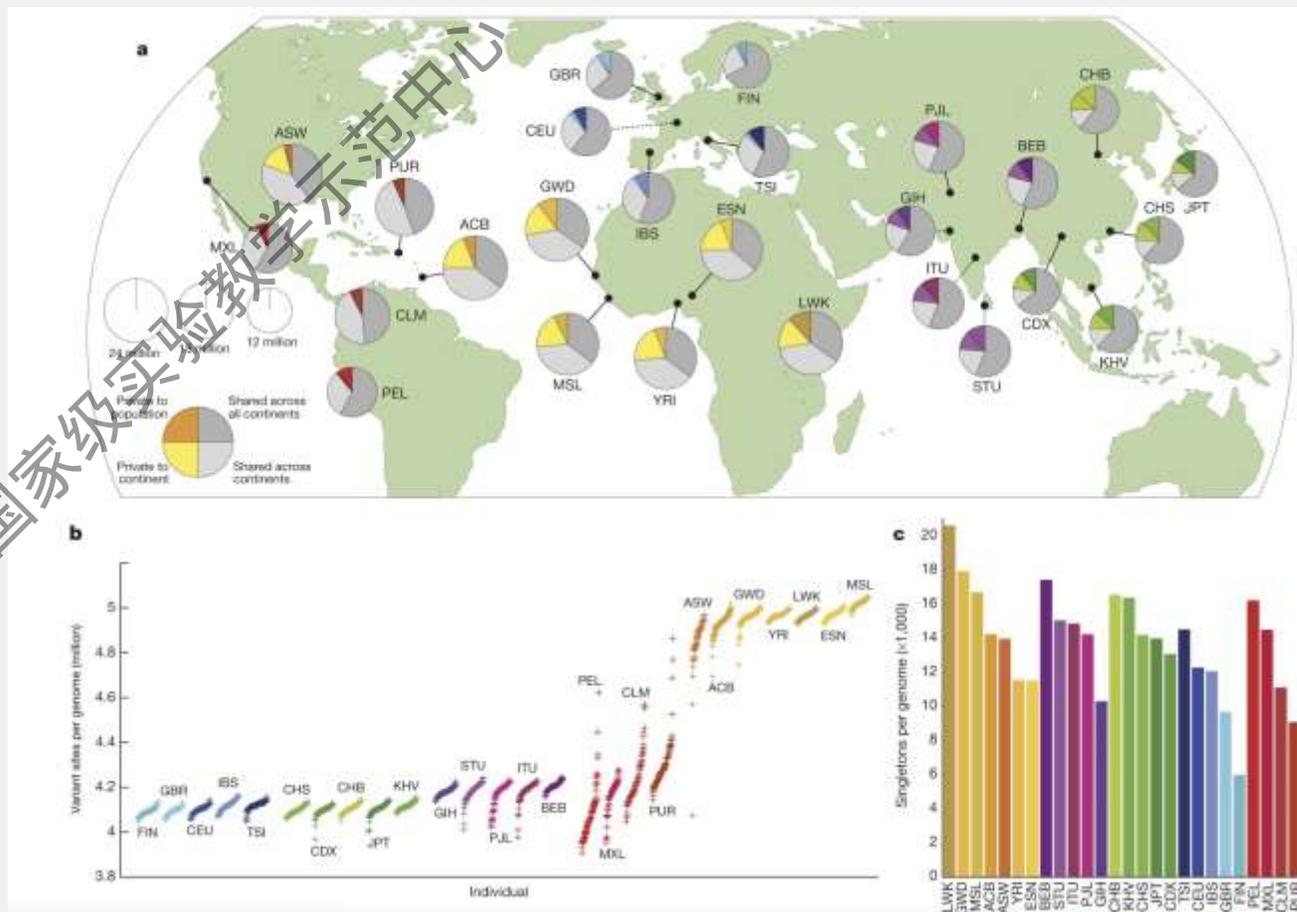


- 1000 Genomes Project (1KG, 2008)

Samples for the 1000 Genomes Project



- 1000 Genomes Project (1KG, 2008)
- Genome-Wide Human Origins Array (HO dataset)



- 1000 Genomes Project (1KG, 2008)
- Genome-Wide Human Origins Array (HO dataset)
- 国际人类表型组计划





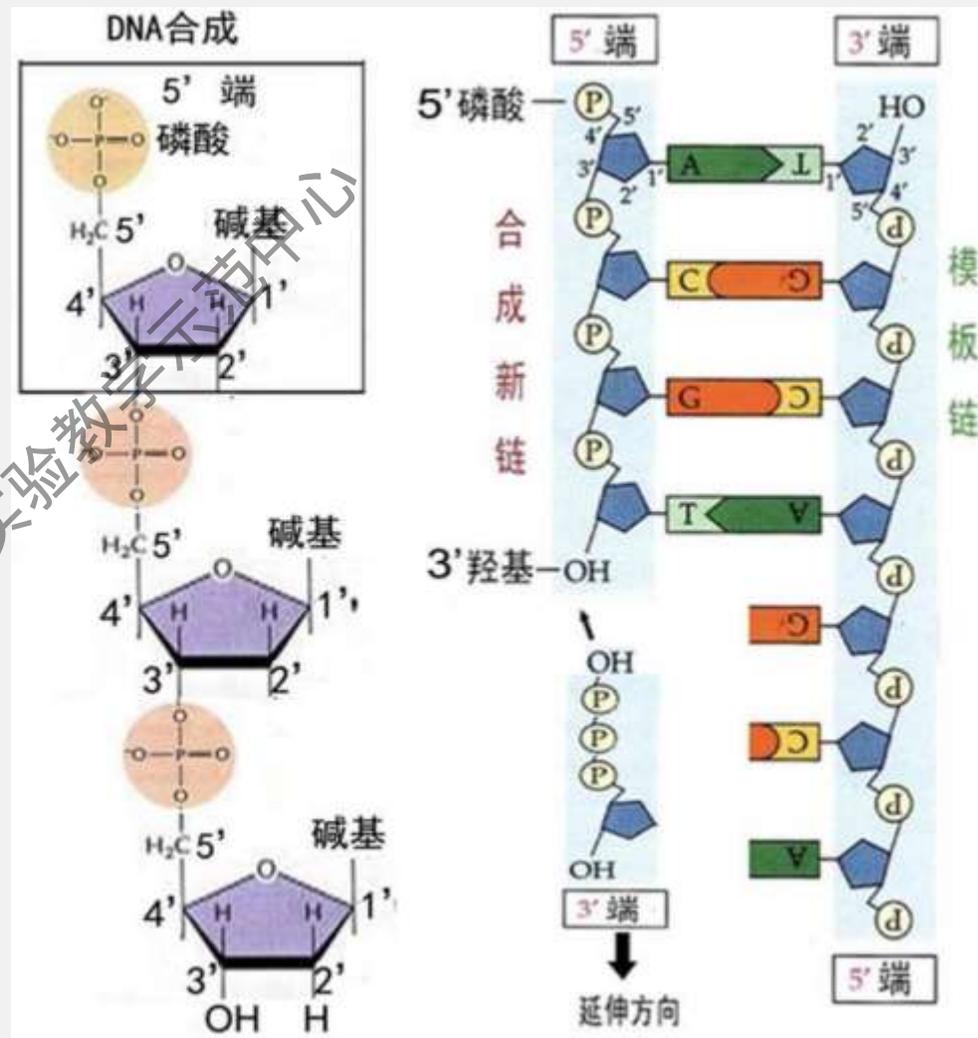
PART 04

# 基因组测序技术

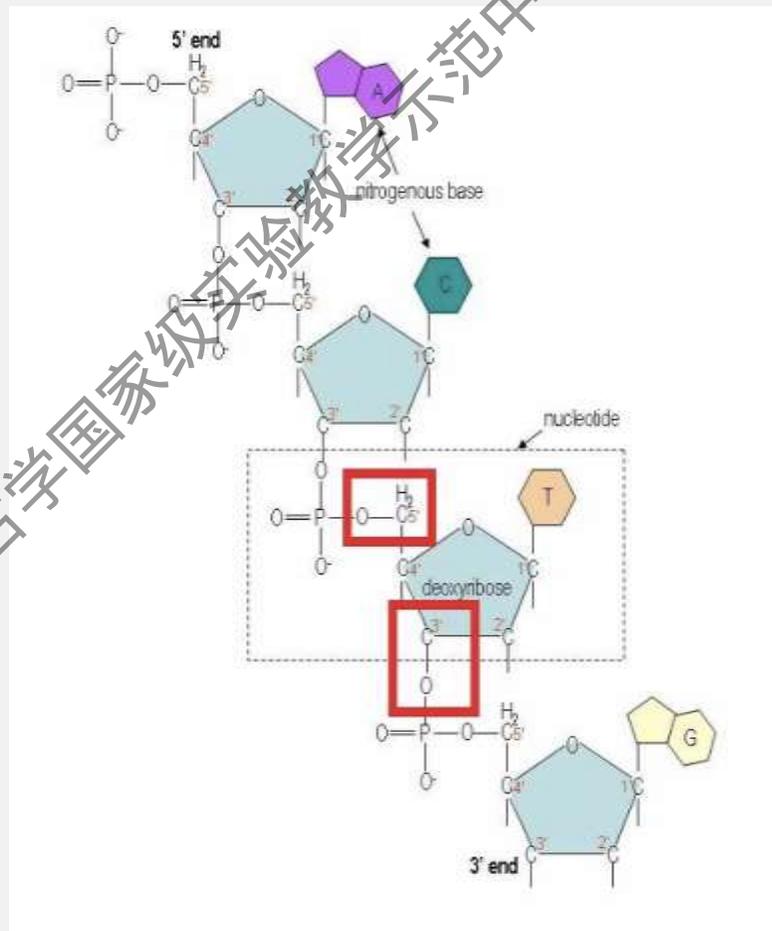
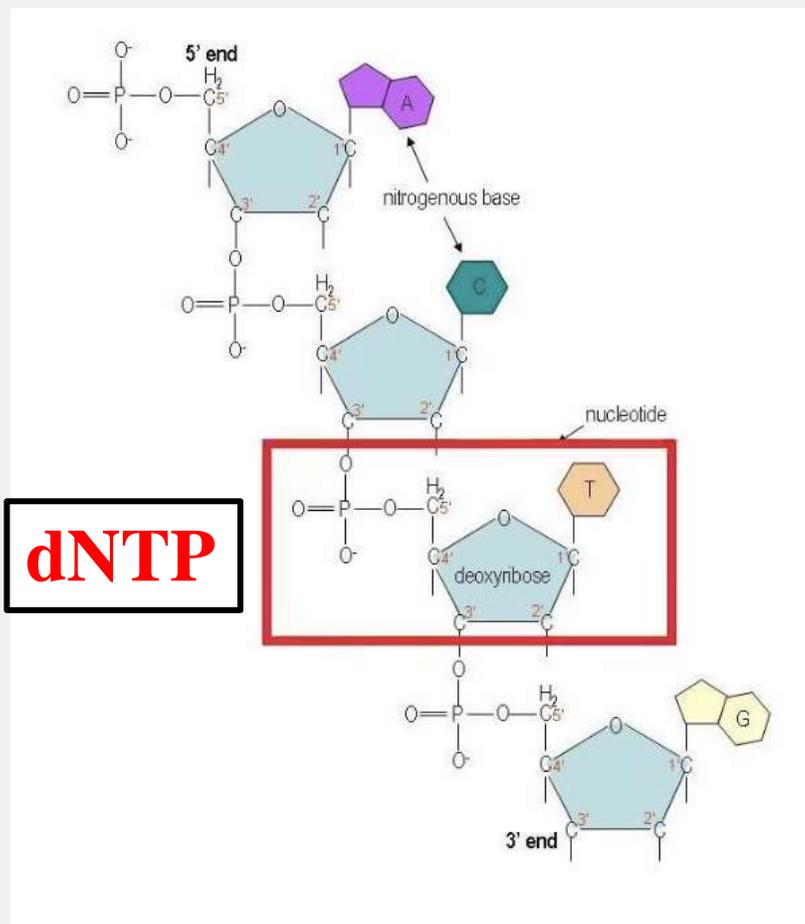
考古学国家级实验教学示范中心

## 一、DNA复制

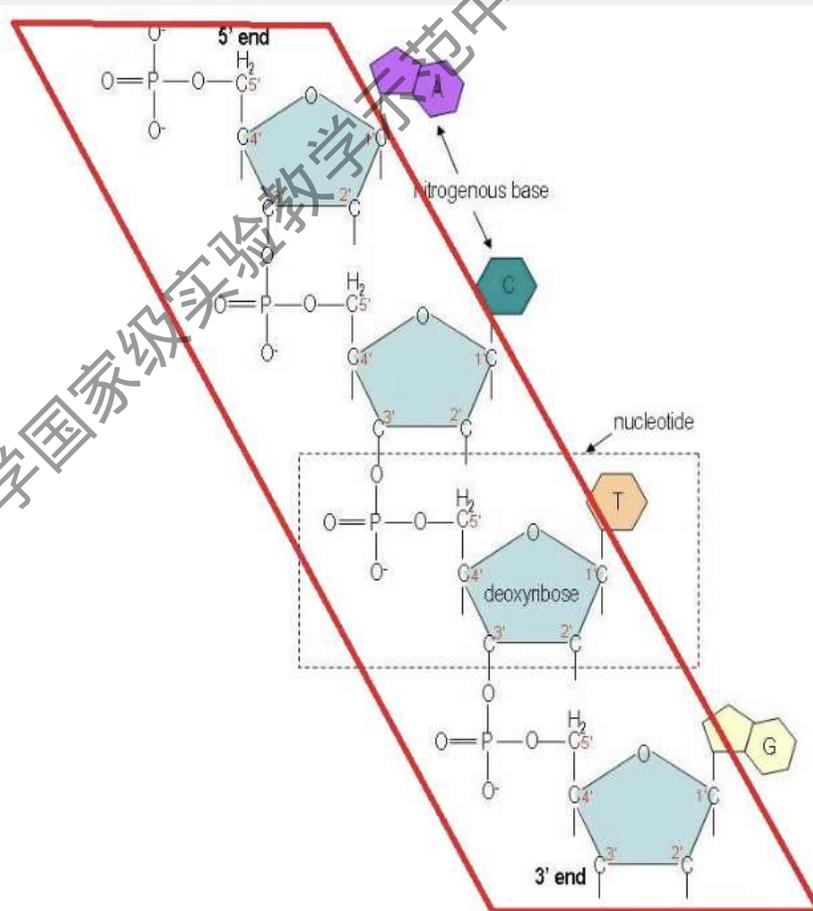
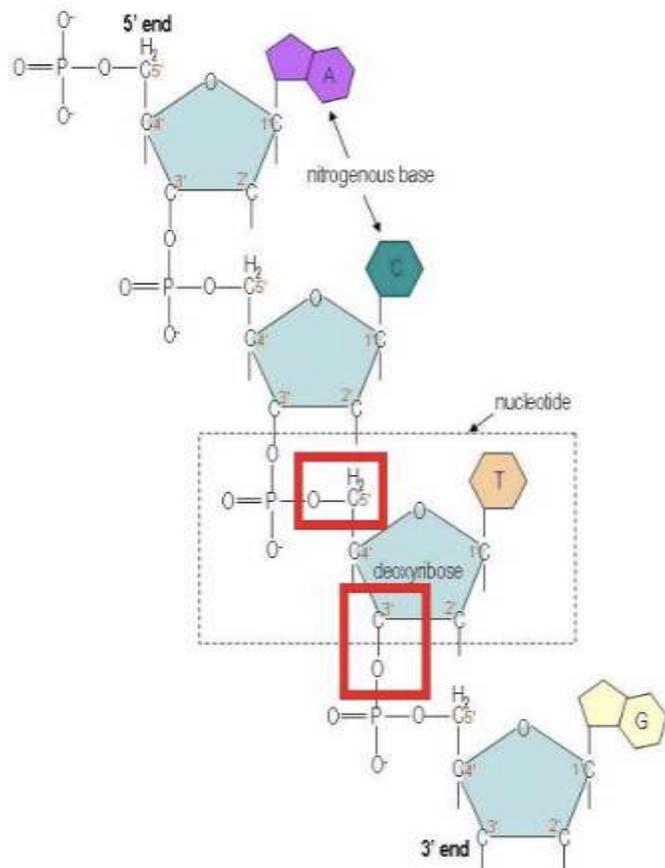
DNA的复制为半保留复制。由于DNA单链复制延伸时只能以核苷酸单体的5'-磷酸基团与前一个核苷酸的3'-羟基合生成磷酸二酯键。因此DNA复制是以5'到3'方式进行。



## 一、DNA复制



## 一、DNA复制



## 一、Sanger测序法

英国生物化学家桑格（Sanger）分别获得1958年和1980年诺贝尔化学奖。他是同一领域内两次获奖的第二人，更关键的是，两次获奖理由都可归结为：测序。1951年，Sanger首次确定了两种蛋白质（牛胰岛素A和B形式）的序列，表明它们是截然不同的分子，为此，他于1958年获得了诺贝尔化学奖。1977年，他引入“双脱氧终止法”，也被称作“桑格法”，因此与另外两名科学家共同获得了1980年诺贝尔化学奖。

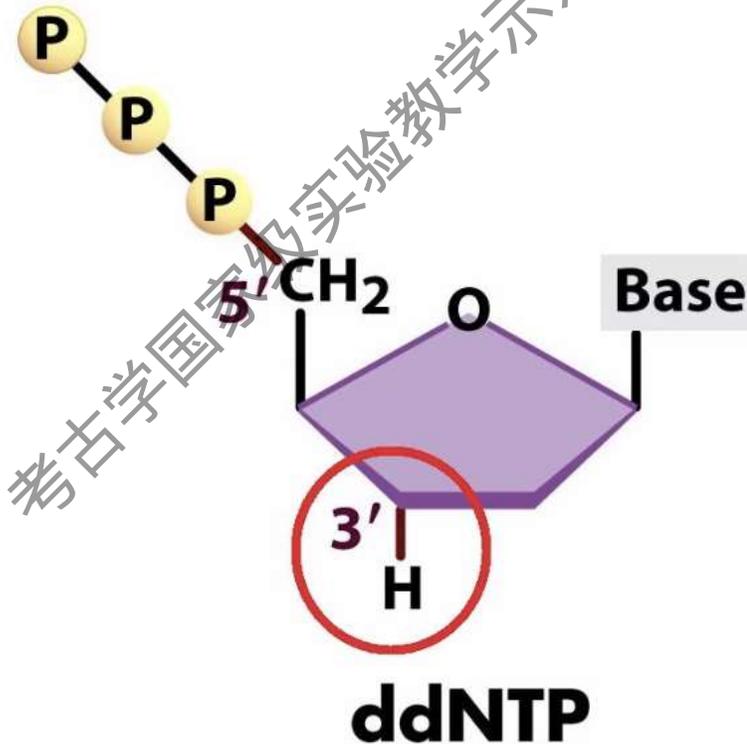
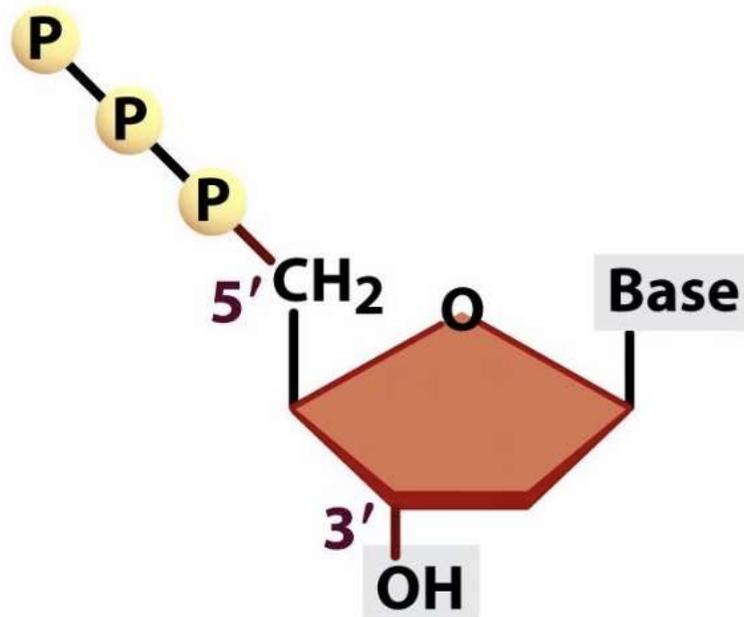


## 一、Sanger测序法

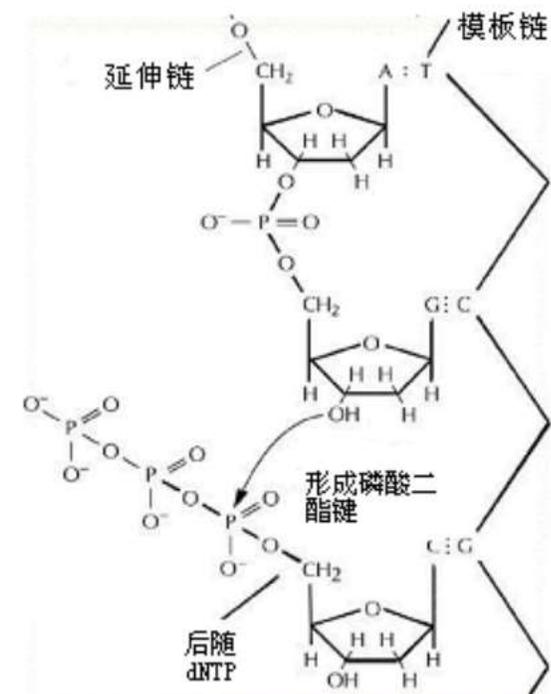
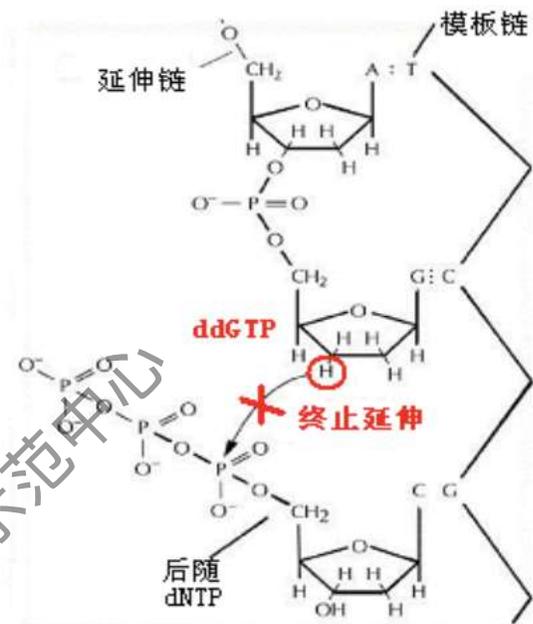
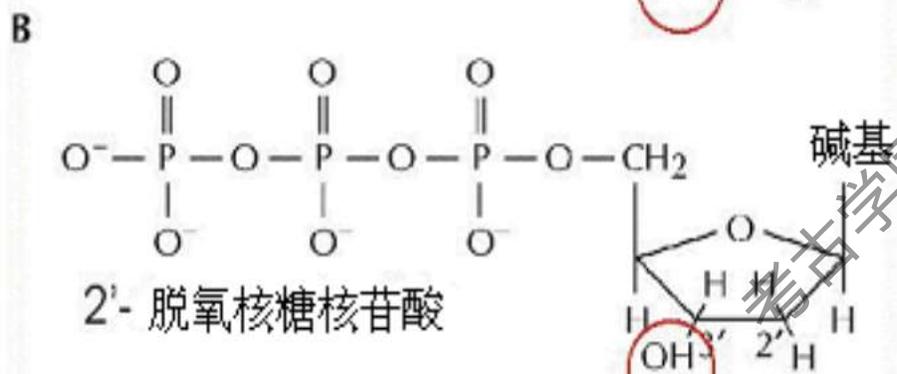
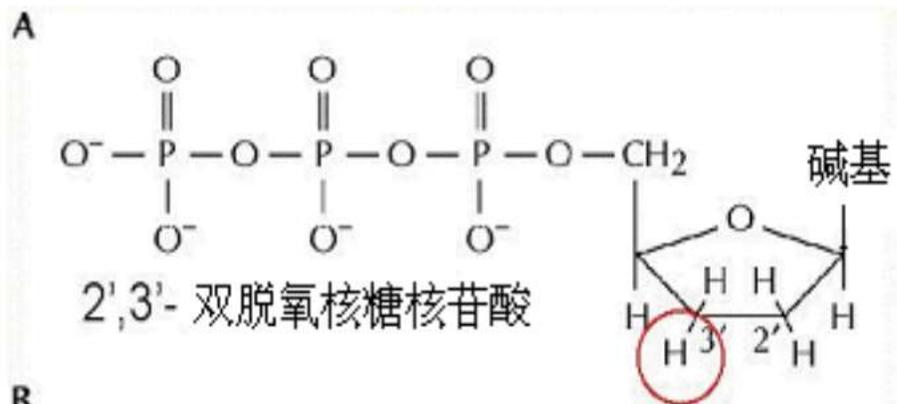


# 第一代测序技术

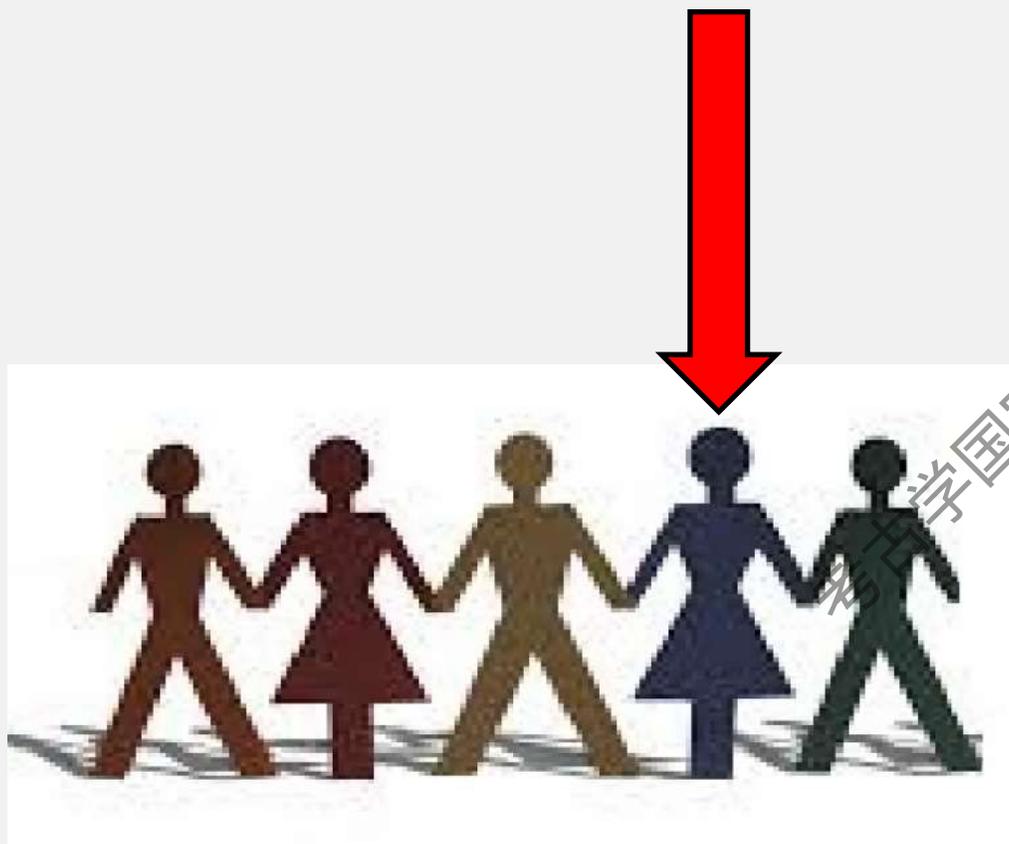
## 一、Sanger测序法：双脱氧核糖核苷酸



# “双脱氧末端终止”的含义



## 一、Sanger测序法



## 一、Sanger测序法



G	GCGAATGCGTCCACACGCTACAGG	G
T	GCGAATGCGTCCACACGCTACAG	T
G	GCGAATGCGTCCACACGCTACAG	G
G	GCGAATGCGTCCACACGCTACAG	G
A	GCGAATGCGTCCACACGCTAC	A
C	GCGAATGCGTCCACACGCTAC	C
A	GCGAATGCGTCCACACGCTA	A
T	GCGAATGCGTCCACACGCT	T
C	GCGAATGCGTCCACACG	C
G	GCGAATGCGTCCACAC	G
C	GCGAATGCGTCCACA	C
A	GCGAATGCGTCCAC	A
A	GCGAATGCGTCCA	A
C	GCGAATGCGTCC	C
A	GCGAATGCGTC	A
C	GCGAATGCGT	C
C	GCGAATGCG	C
T	GCGAATGC	T
G	GCGAATG	G
C	GCGAAT	C
G	GCGAAT	G
T	GCGAAT	T

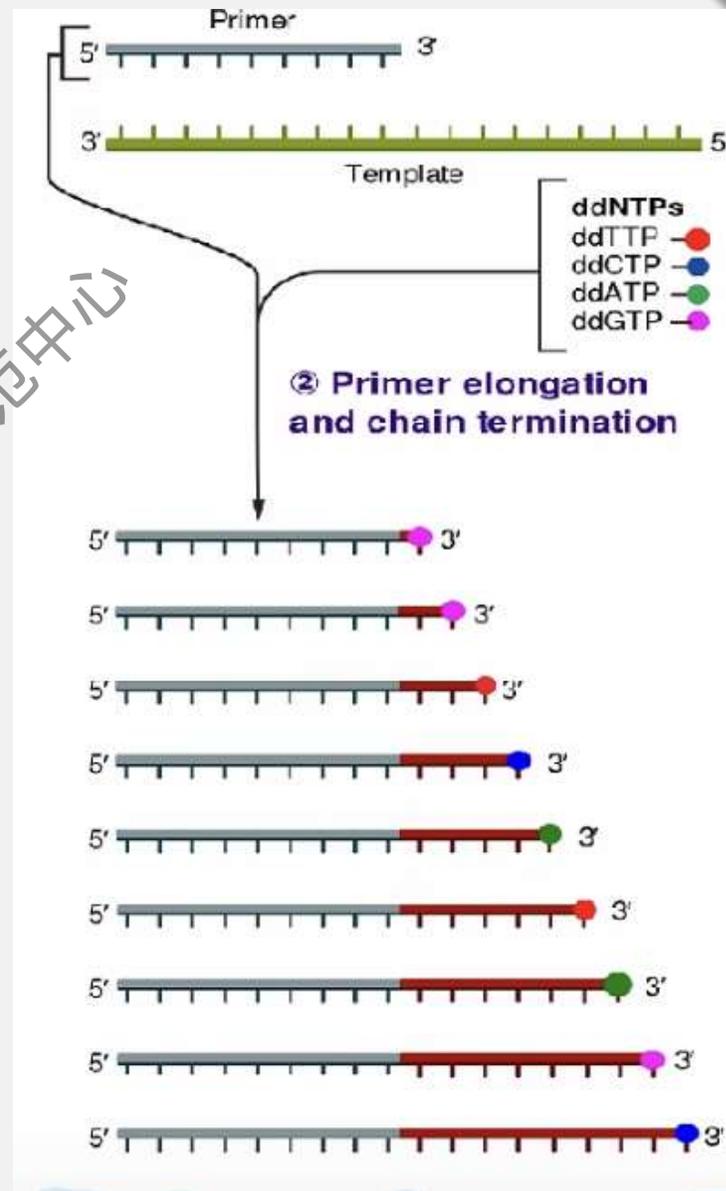
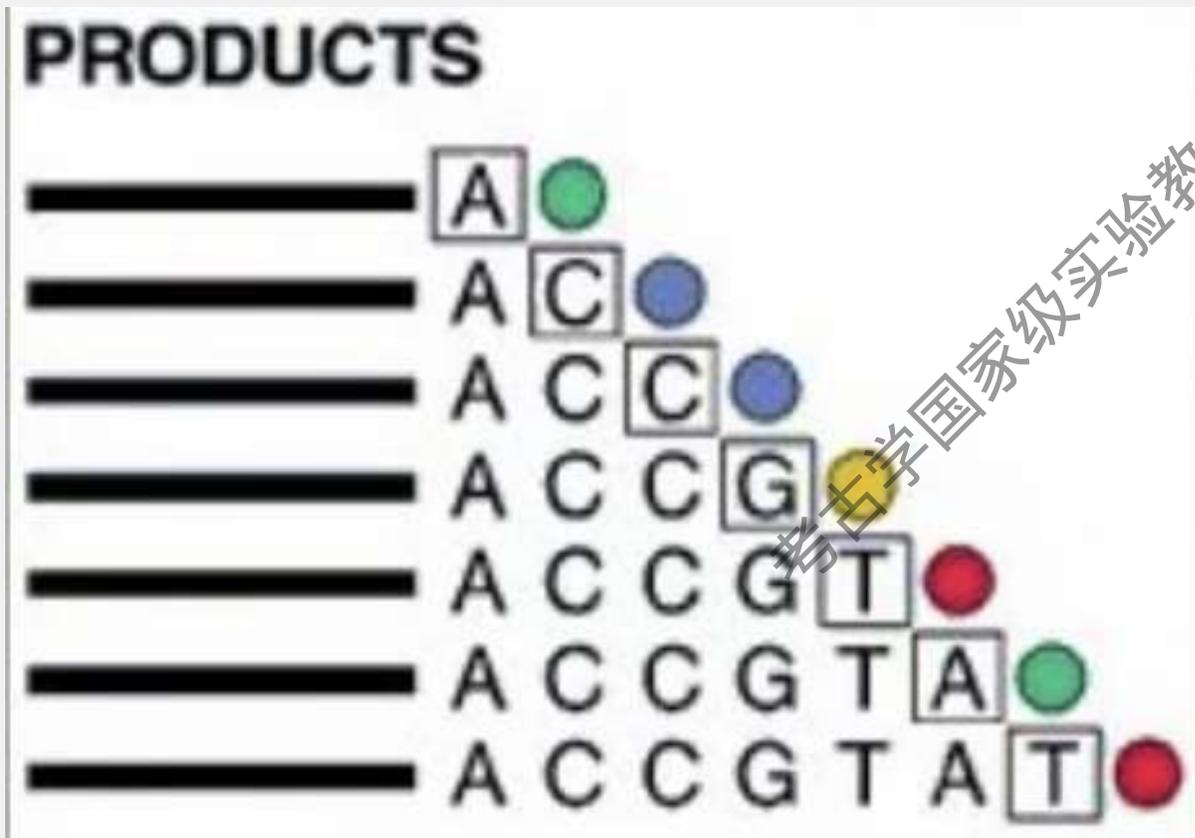
考古学国家级实验教学示范中心





# 第一代测序技术

## 一、Sanger测序法





## 一、高通量测序（next-generation sequencing, NGS）

四大步骤：

- 样本准备（文库）
- 桥式反应
- 测序

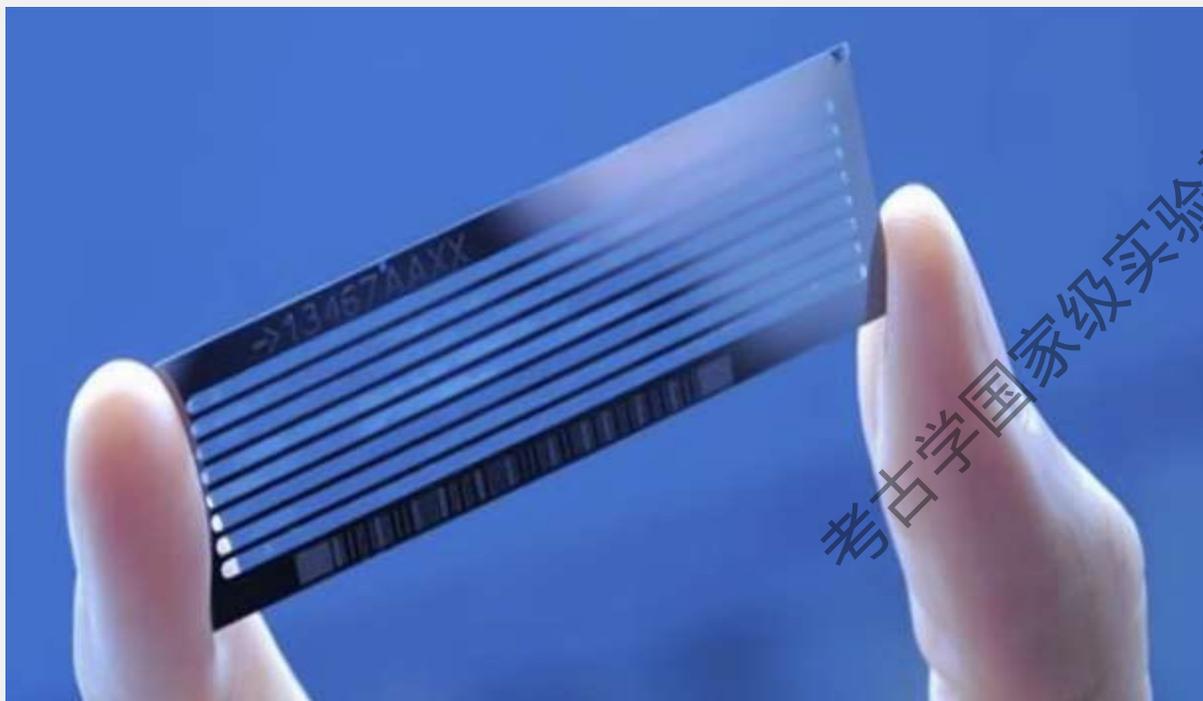
两大特点：

- 边合成边测序
- Paired-End sequencing（双端测序）

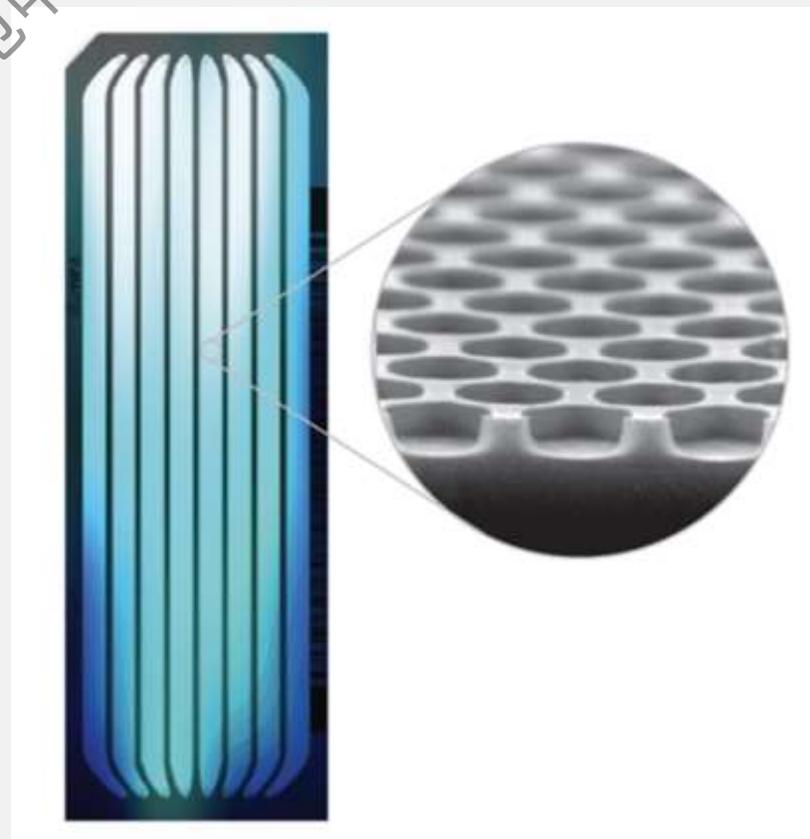
考古学国家级实验教学示范中心



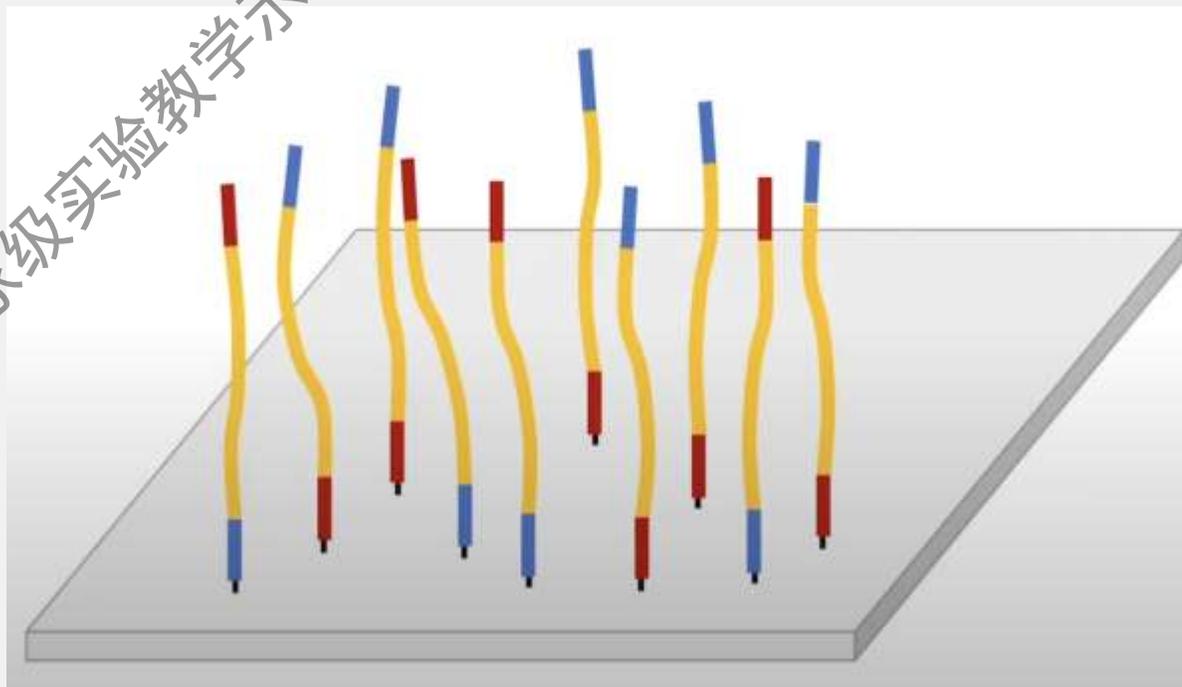
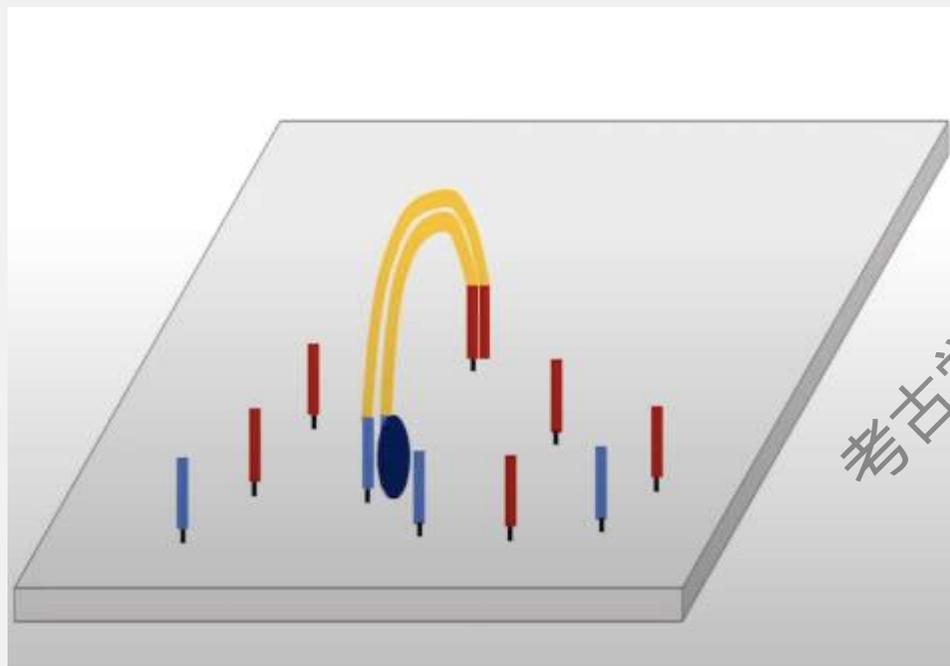
## 二、样本准备（文库）



Flowcell



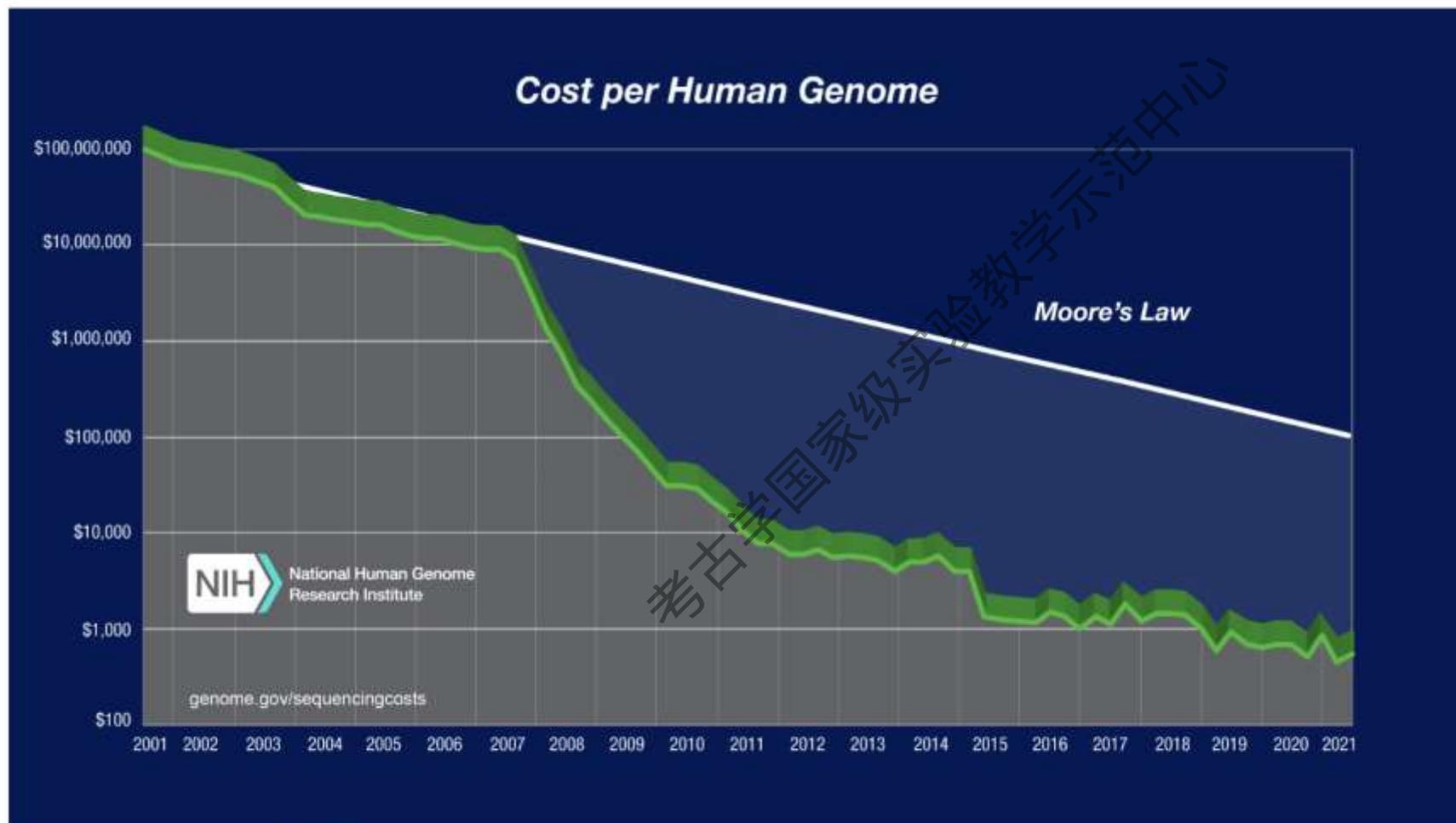
## 三、桥式反应



考古学国家级实验教学示范中心



# 第二代测序技术



Cost per genome data - 2021

100万 - 1千

# 主要内容回顾

- 细胞：人类生命活动的基本单位  
细胞膜、细胞质、**细胞核**、细胞器（**线粒体**）
- DNA：人类遗传信息的载体
- 染色体：高度螺旋化的DNA分子，人类染色体数目
- 基因和基因组：合成具有功能的蛋白质的DNA序列
- 基因的作用：**人类基因组计划、进化、医疗以及自然选择**
- DNA测序：**一代测序、二代测序**

