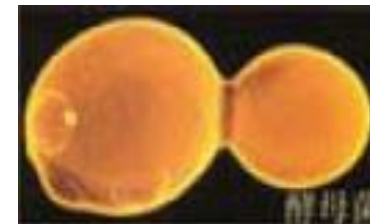


HGP 的主要任务包括人类
基因组的四张图以及六种
模式生物基因组的测序;



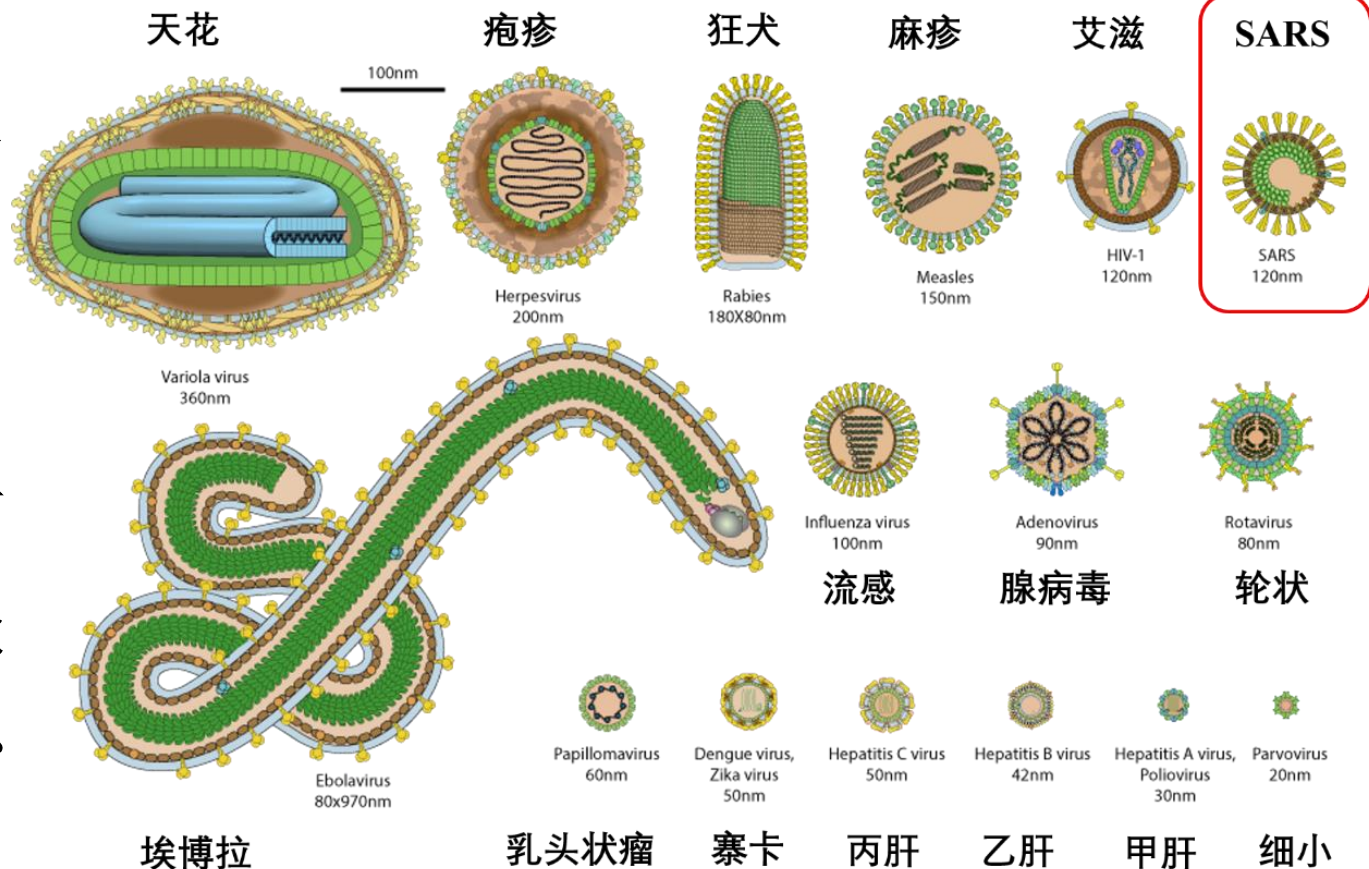
四、模式生物基因组

基因组的组成及特点——病毒

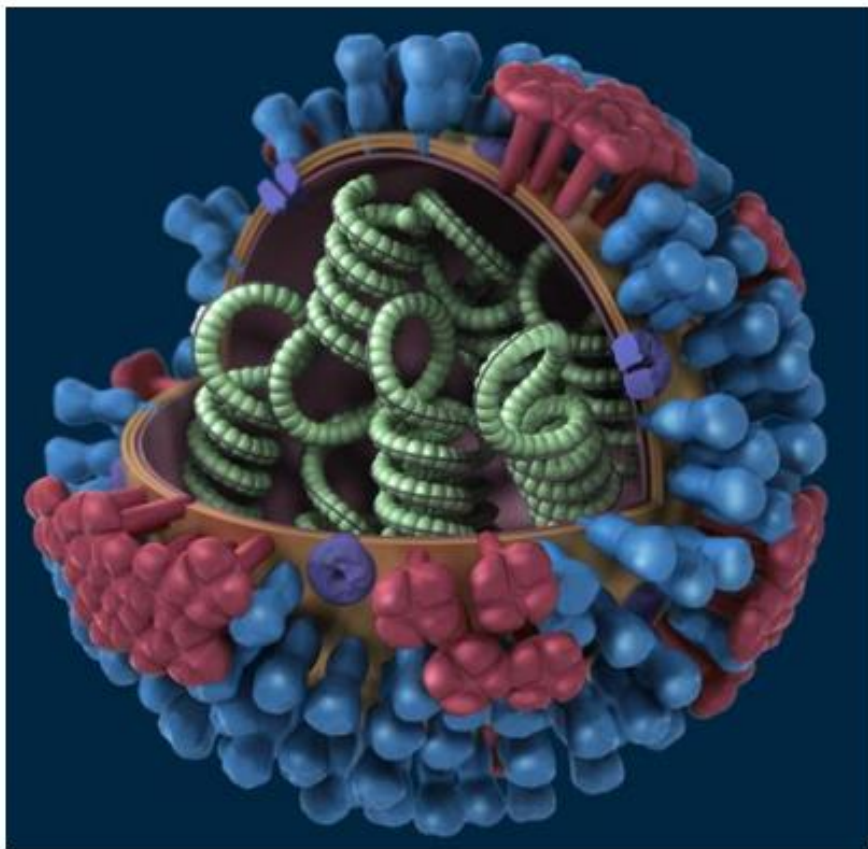
① 病毒基因组容量较小，但组成结构多样化；

② 具重叠基因，
几乎全部序列
用于编码蛋白
质；

③ 部分病毒具有
极高的变异能
力，极具威胁



流感病毒的结构组成与分类依据



核蛋白NP

包裹核酸

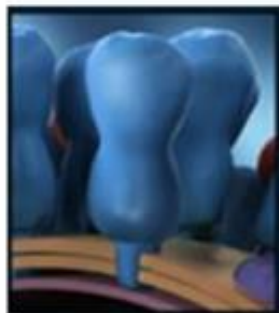
甲、乙、丙型流感



血凝素抗原H

粘住和进入细胞

宿主类型和传染力

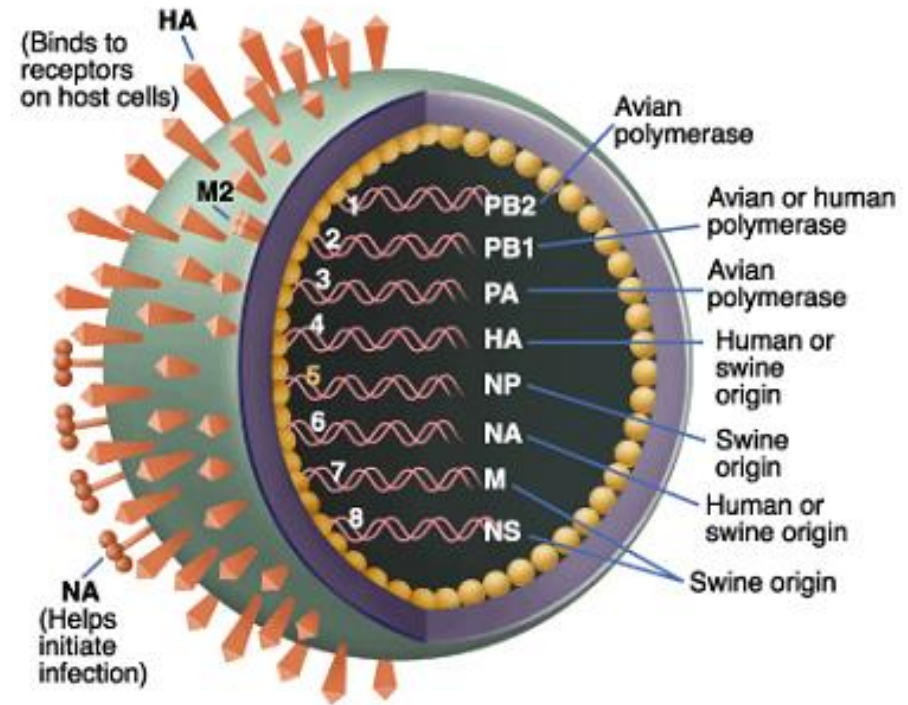


神经氨酸抗原N

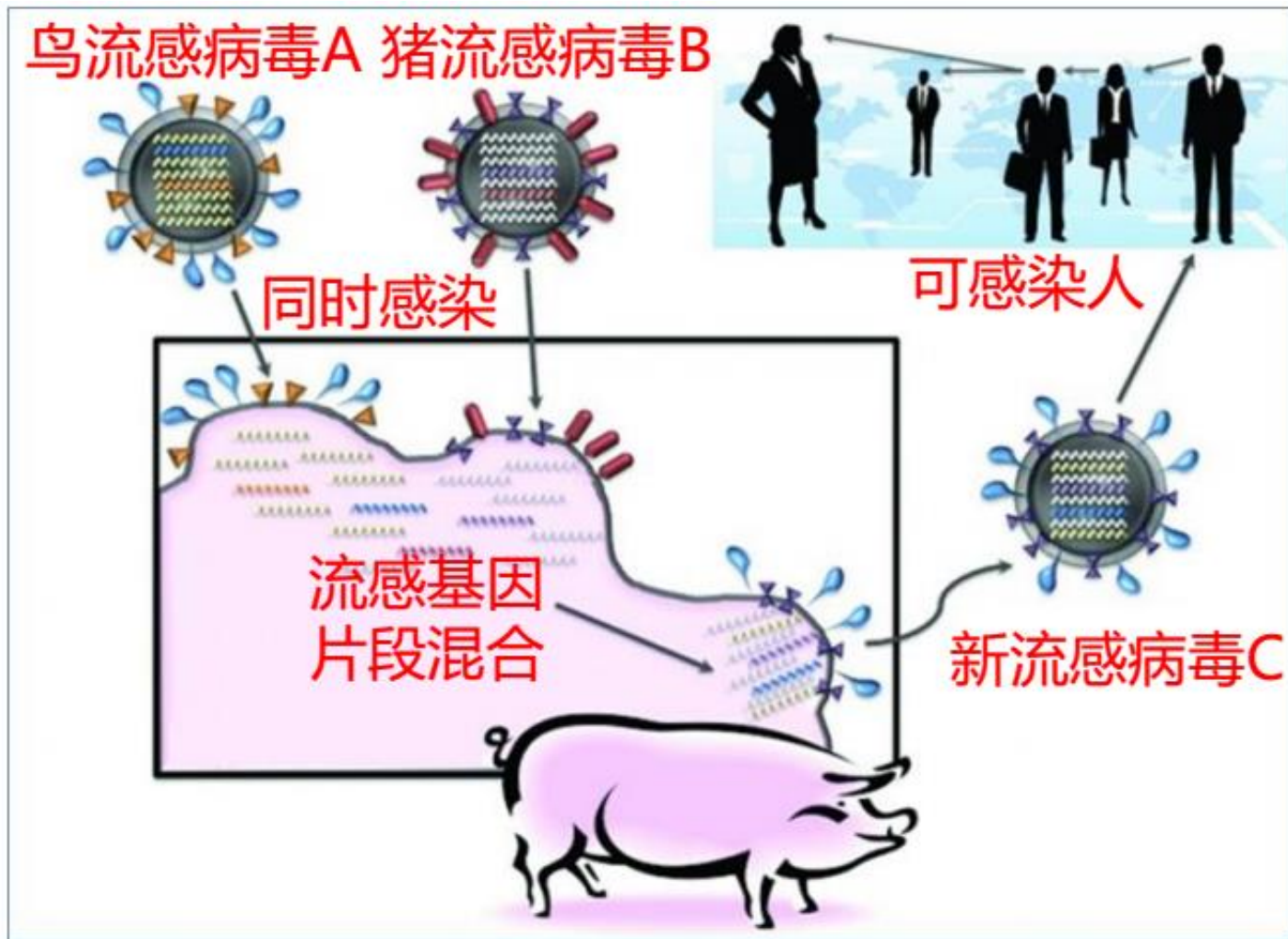
细胞裂解

致死率

H1N1流感病毒为单股RNA病毒，基因组13.6kb，含8个独立的基因片段，编码11个蛋白质

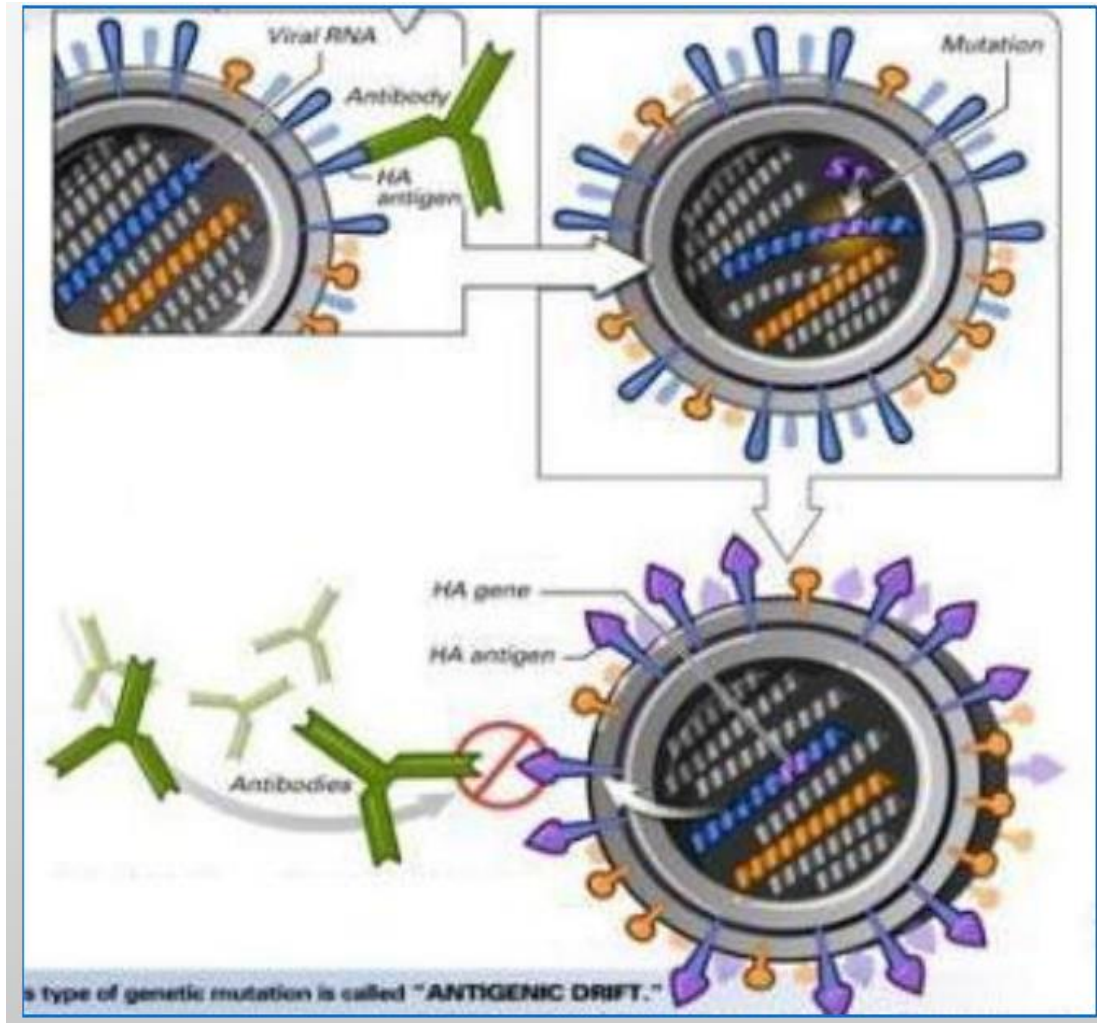


- | | | |
|---|---------|--------------|
| 1 | PB2 | } RNA聚合酶 |
| 2 | PB1 | |
| 3 | PA | |
| 4 | HA | 凝血酶素HA |
| 5 | NP | 核蛋白NP |
| 6 | NA | 神经氨酸NA |
| 7 | M1/M2 | 基质蛋白M1和M2 |
| 8 | NS1/NS2 | 非结构蛋白NS1和NS2 |



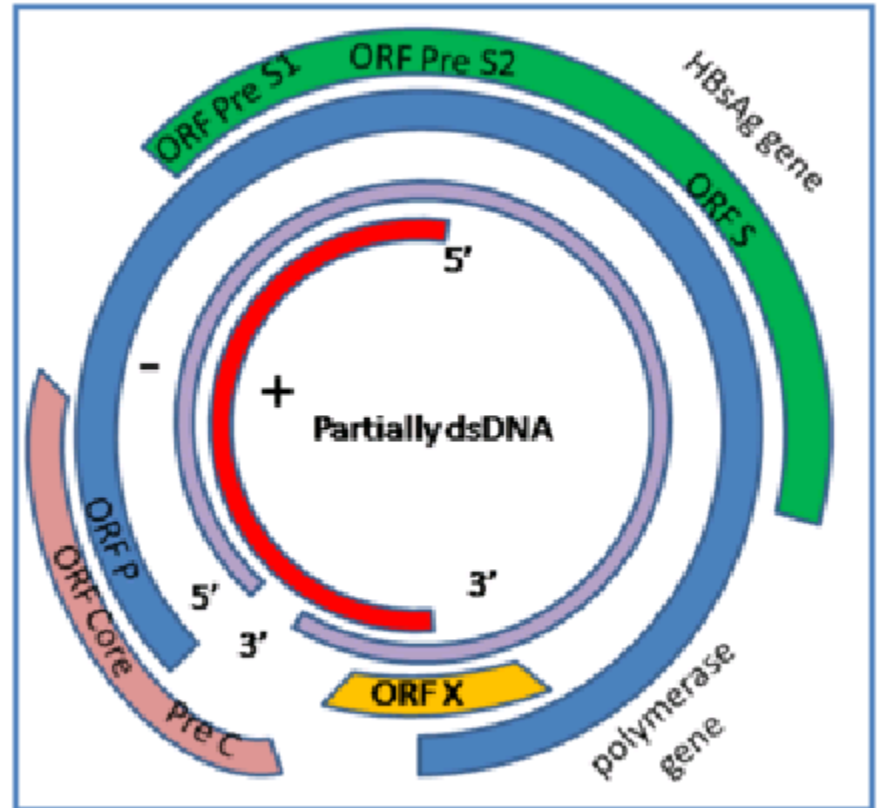
抗原转变 (antigenic shift): 流感病毒的基因序列以独立的片段存在于病毒体内，当不同病毒感染同一宿主细胞时，易发生基因片段之间的重新装配，产生基因组变异。

抗原漂变 (antigenic drift): RNA病毒基因组中没有复制校正系统, 易发生基因的点突变。

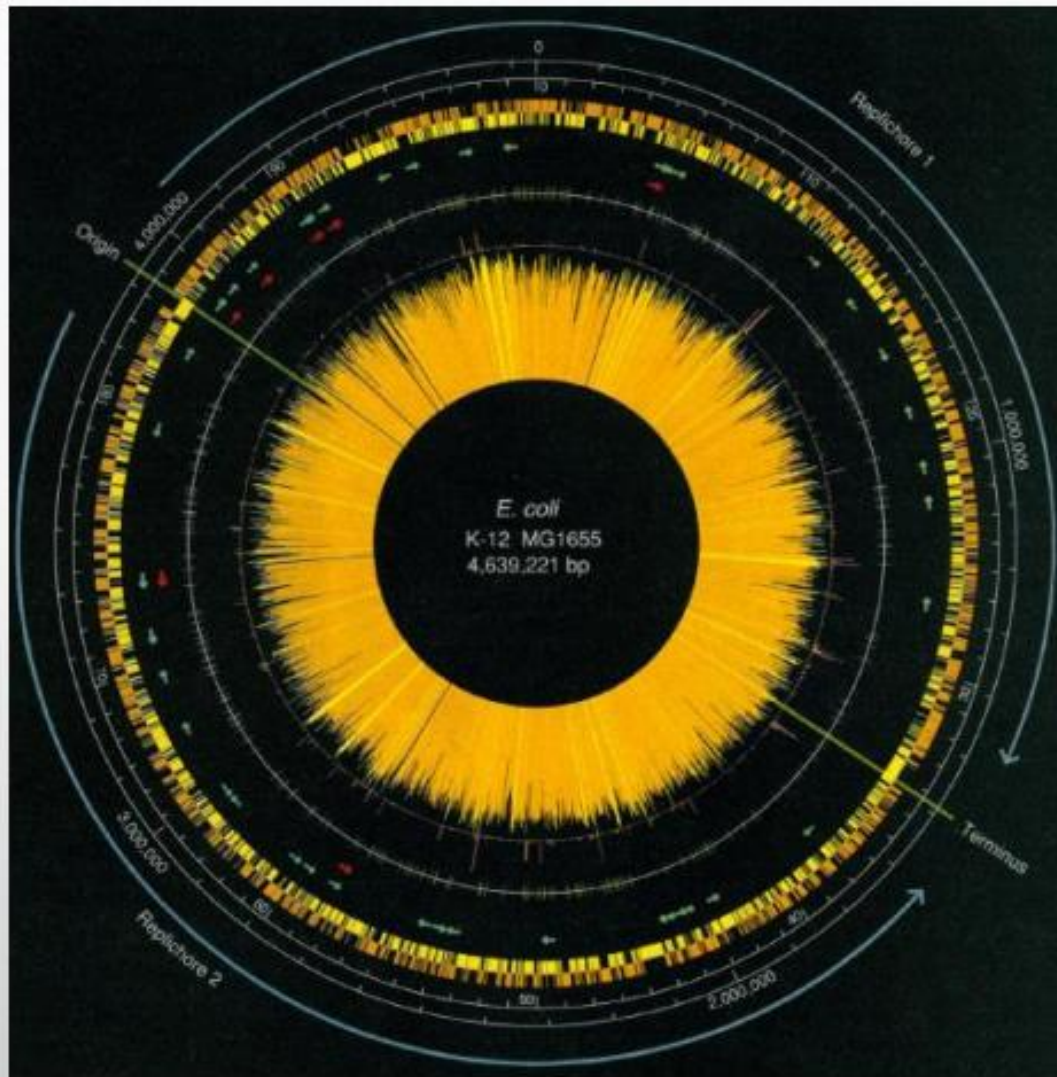


乙型肝炎病毒 Hepatitis B virus

- 不完全双链DNA分子，大小为3.2kb。
- 含有4个基因，分别编码衣壳蛋白S、DNA多聚酶P、核心蛋白C和功能未知蛋白X。
- 存在基因重叠：S基因完全重叠于P基因中，X基因与P、C基因，C基因与P基因都有重叠。



基因组的组成及特点——细菌



Science. 1997 Sep 5;277(5331):1453-62.

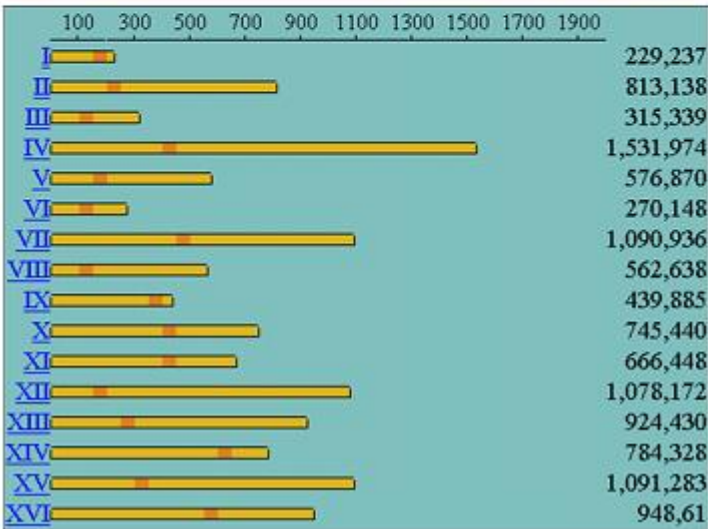
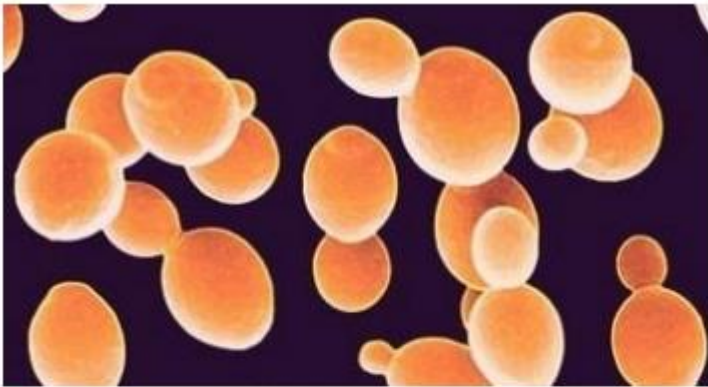
- 闭合环状双链DNA，大小为4.6Mb;
- 87.8%为蛋白编码序列，含4288个蛋白编码基因; 0.8%编码RNA，0.7%为非编码重复序列，余下10.7%为表达调控和未知功能序列;
- 两条DNA链均可作为模板进行基因转录，基因间平均间隔为118bp。



细菌基因组的主要结构和组织特点

- ① 基因组通常是闭合环状双链DNA分子，以超螺旋的组织方式集中在一个较致密的区域，称为类核（nucleoid）。但是也有部分线性基因组，如链霉菌等；
- ② 基因组容量比病毒大，一般为几个Mb；
- ③ 大多数基因为单拷贝基因，出现少数多拷贝基因（如rRNA基因）；含有具有基因表达调控的DNA元件以及少量非编码重复序列，重叠基因较少；
- ④ 具有操纵子结构，数个功能相关的结构基因串联在一起，在同一调节区域的调节下共同转录；
- ⑤ 质粒是细菌染色体外共价闭合的环状DNA分子，1-200kb大小不等，可以独立复制，但不是细菌生存所必需的。

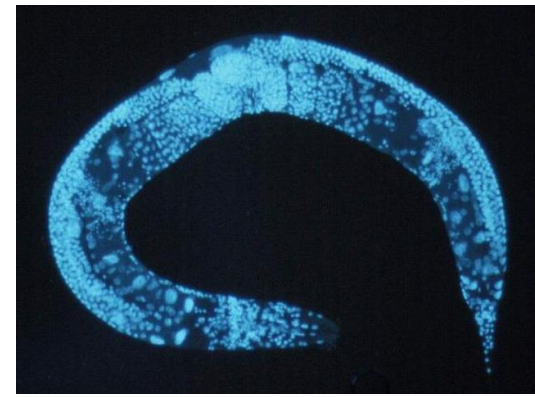
基因组的组成及特点——酿酒酵母 (*Saccharomyces cerevisiae*)



- 为最早完成测序的真核生物基因组（1996年）；
- 基因组全长12,086kb，共16条染色体；
- 编码5885个蛋白编码基因和约455个RNA基因（含140个rRNA基因、275个tRNA基因和40个snRNA基因）；
- 基因组内出现相当数量的冗余序列，包括重复序列和多拷贝基因。

遗传冗余（genetic redundancy）是真核基因组区别于原核基因组的显著特征。

- 1999年，**第一个**多细胞生物秀丽隐杆线虫（*Caenorhabditis elegans*）的全基因组测序完成：在97Mb的基因组中，含有6对染色体，编码19,099个基因；基因在染色体上的密度更低、重复序列的比例也更高。



- 相比简单的单细胞生物酵母，**线虫的基因组编码大量独特的细胞间信号转导蛋白，这也是多细胞生物基因组的一大特点；**



- 2000年，黑腹果蝇（*Drosophila melanogaster*, 140Mb）

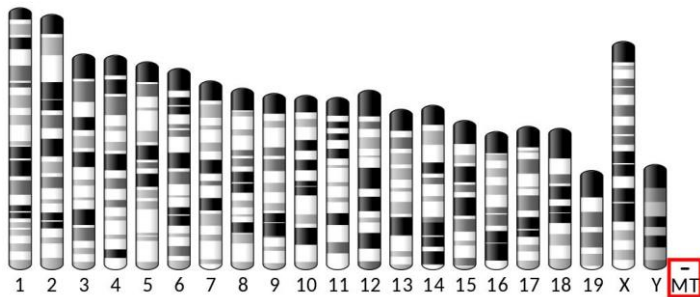
- 2000年，第一个植物拟南芥（*Arabidopsis thaliana*, 115Mb），重复序列仅占基因组13%，多为LTR



基因组的组成及特点——小鼠 (*Mus musculus*)

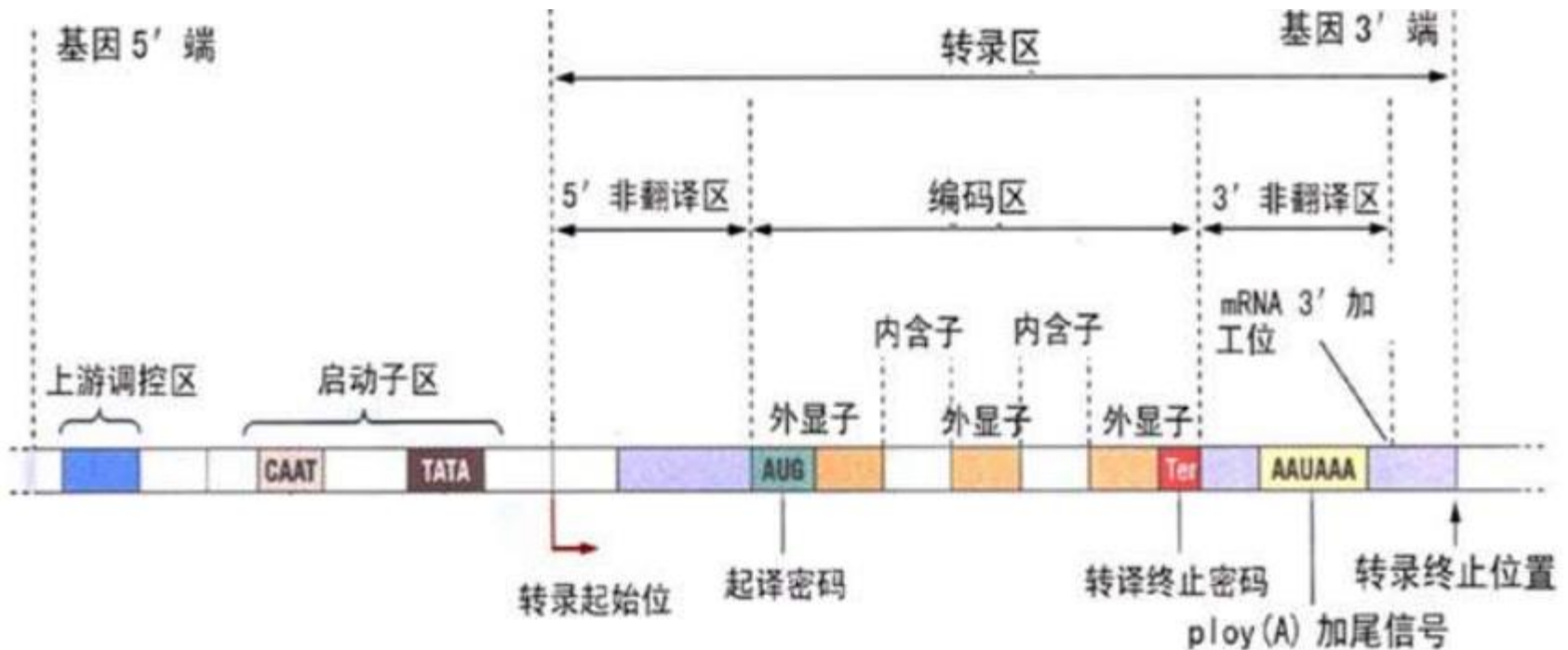


- 继人类后发表的第二个全基因组序列的哺乳动物（2002年）；
- 基因组全长2.5 Gb，单倍体共20条染色体且均为端着丝粒型；
- 小鼠与人类之间有90%以上的相似性，约有40%的小鼠和人类基因组序列高度同源；
- 编码基因约2.3万个，85%的人类基因在小鼠基因组中能找到同源或相似基因，在小鼠和人类中大约有30个印记基因。



五、基因概念的拓展（在线课程）

- **经典遗传学：** 决定表型的独立的遗传单位，可在世代间传递。
- **分子生物学：** 由不同的DNA片段共同组成的一个完整的独立的表达单元，由**调控区**和**转录区**组成，有特定的表达产物，可以是RNA分子，也可为多肽分子。

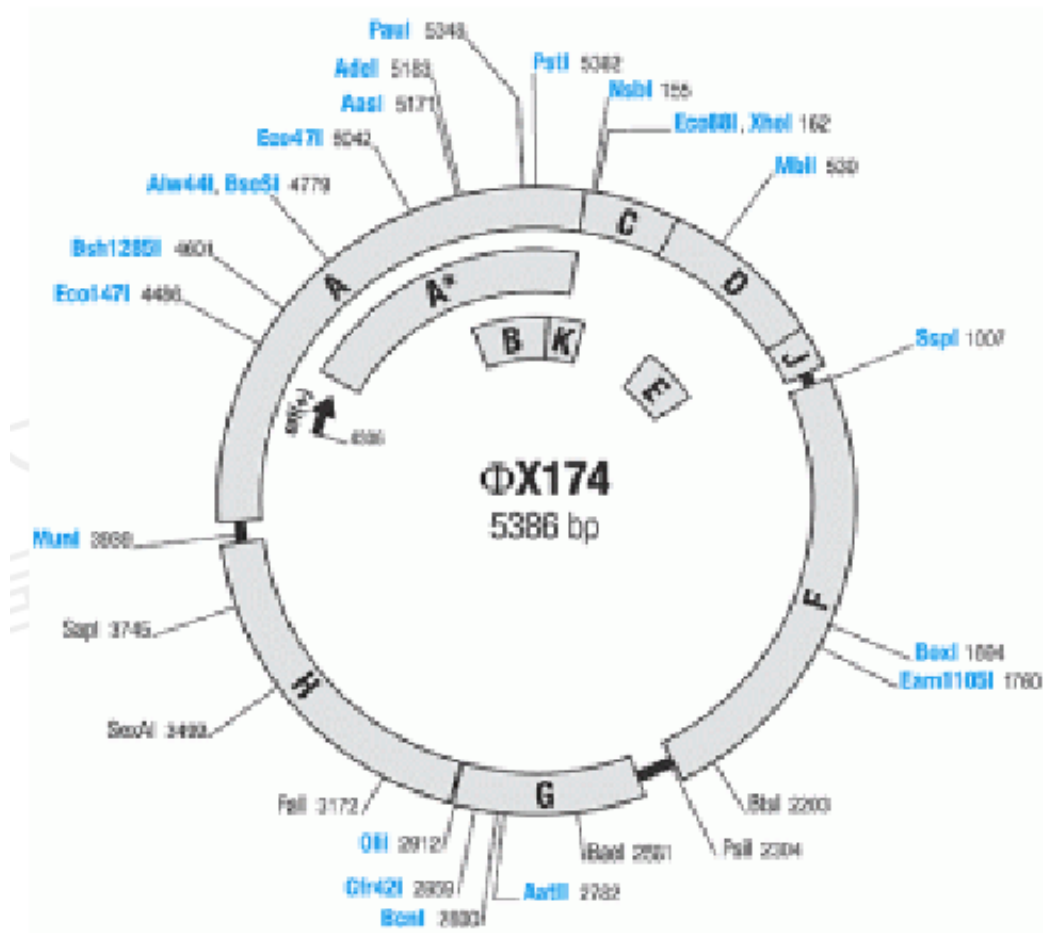


特殊结构基因:



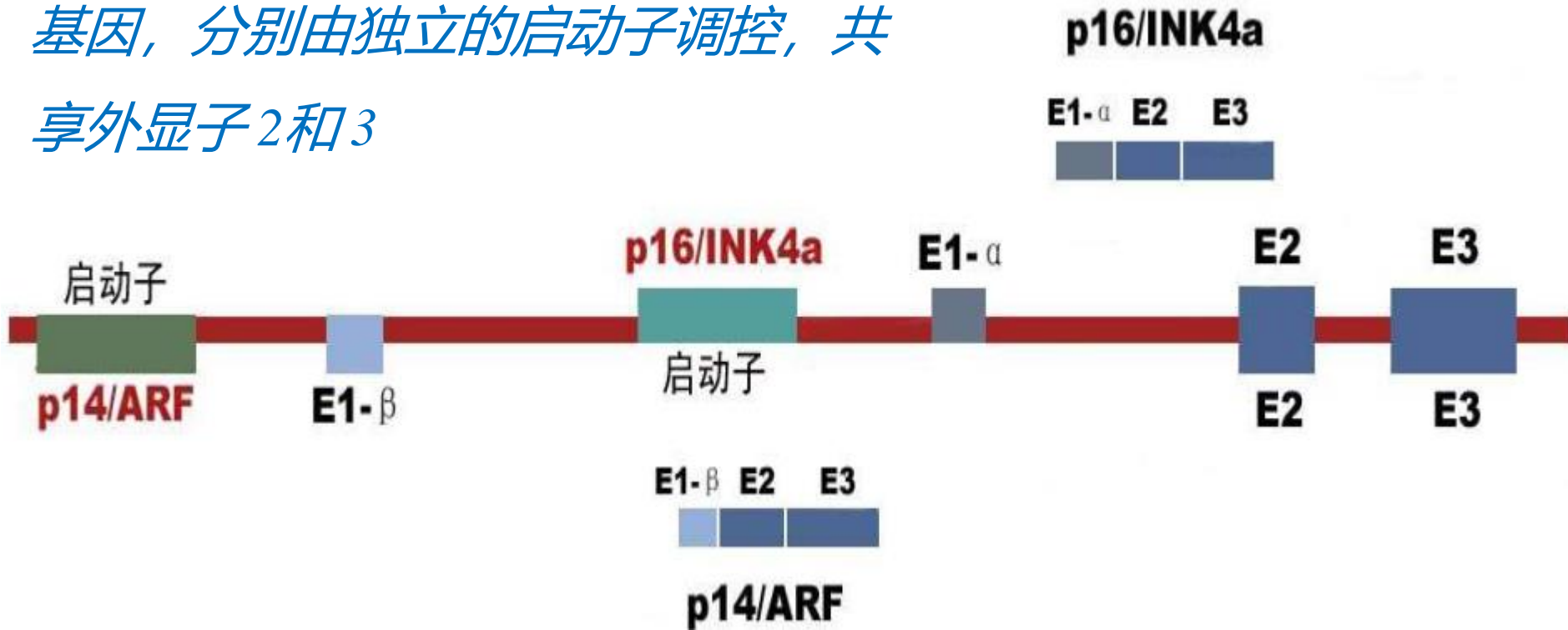
1. **重叠基因(overlapping gene)**: 指两个或两个以上的基因共有一段DNA序列, 或者是一段DNA序列成为两个或两个以上基因的组成部分。

- 5386 bp
- 11 genes
- > 2300 aa
- 几个基因具有不同程度的重叠
- 重叠的基因具有不同的阅读框



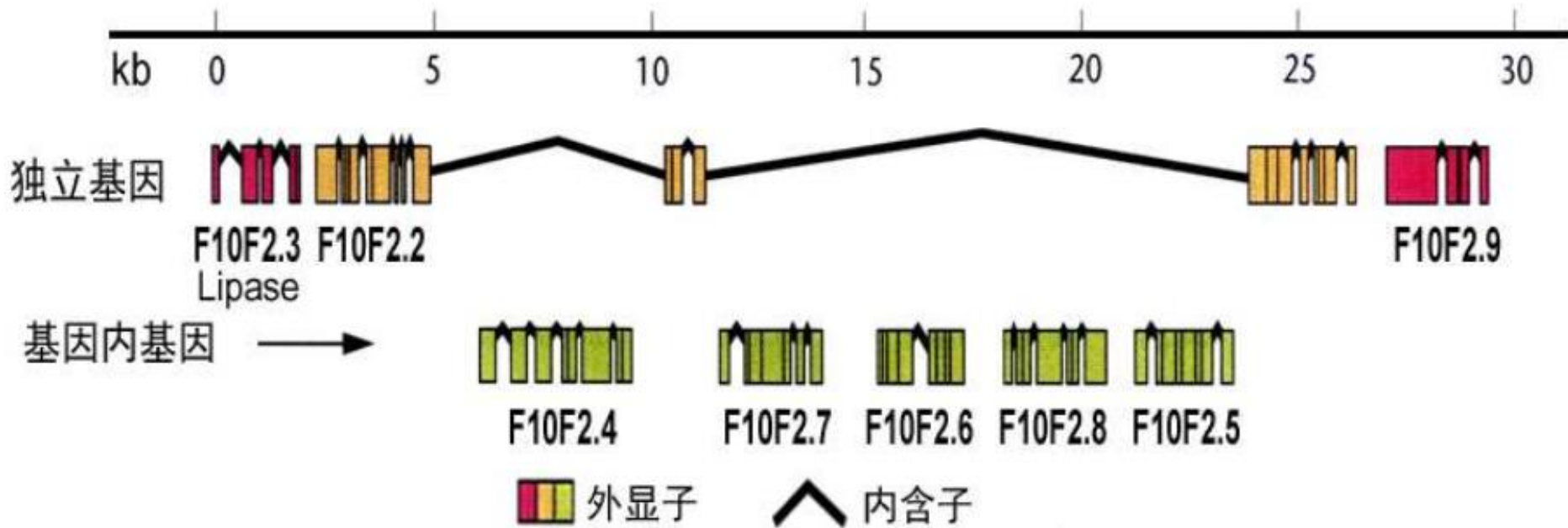
人类的重叠基因

人类基因p14/ARF和p16/INK4a为重叠基因，分别由独立的启动子调控，共享外显子2和3



2. 基因内基因

如：线虫内含子中的独立基因

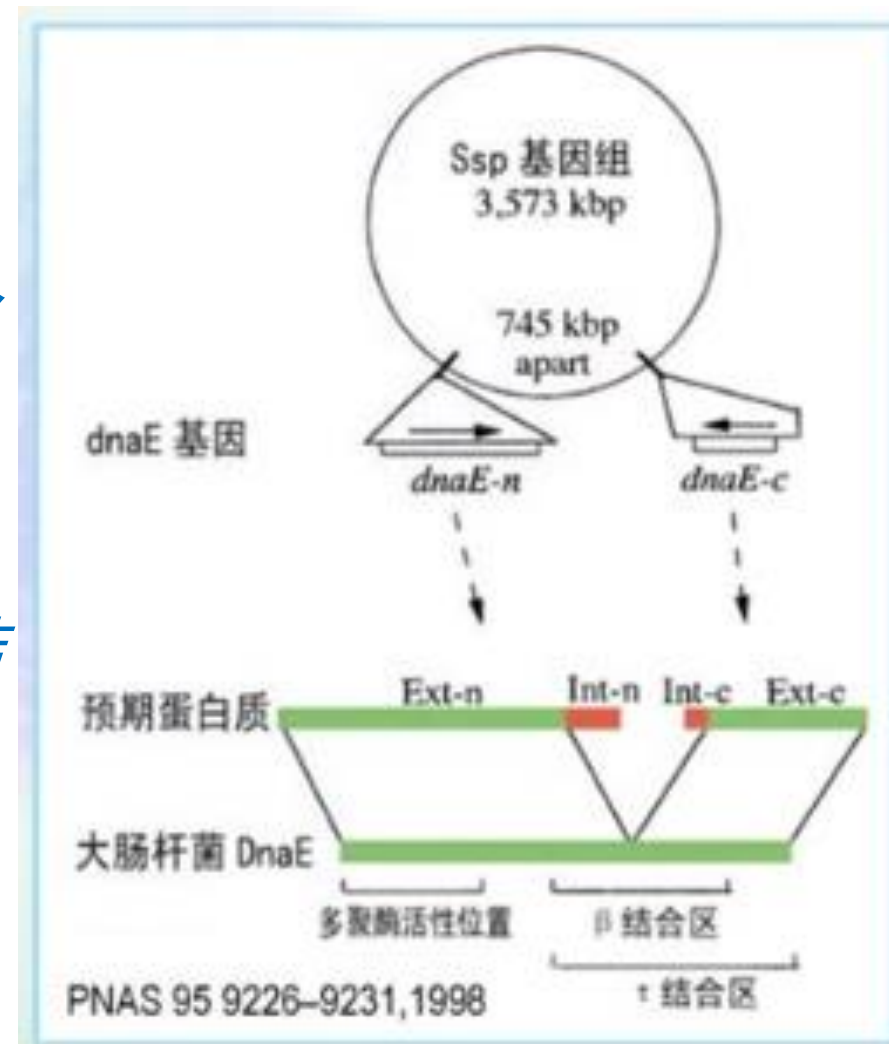


线虫基因组中一段30kb的区段中含有8个基因，其中FGAM(甘氨酸酰胺合成酶)基因长25kb,有21个内含子。内含子9中含有一个独立的基因，内含子11中含有4个独立的基因。基因内基因分别具有独立的外显子和内含子。

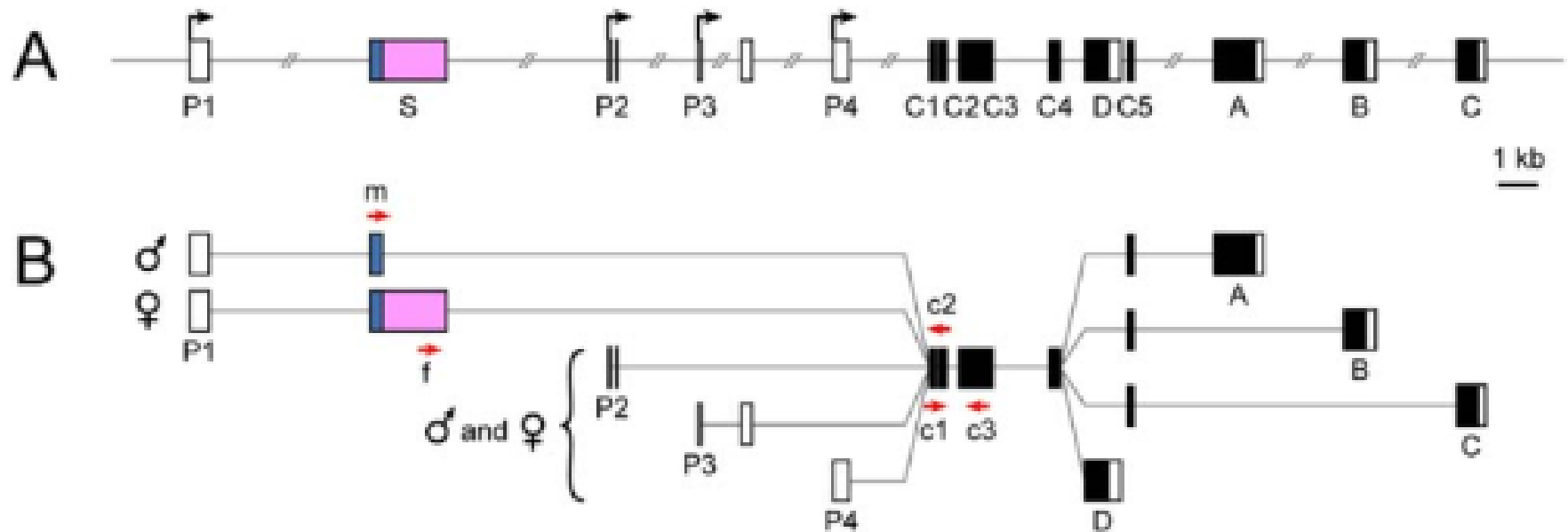
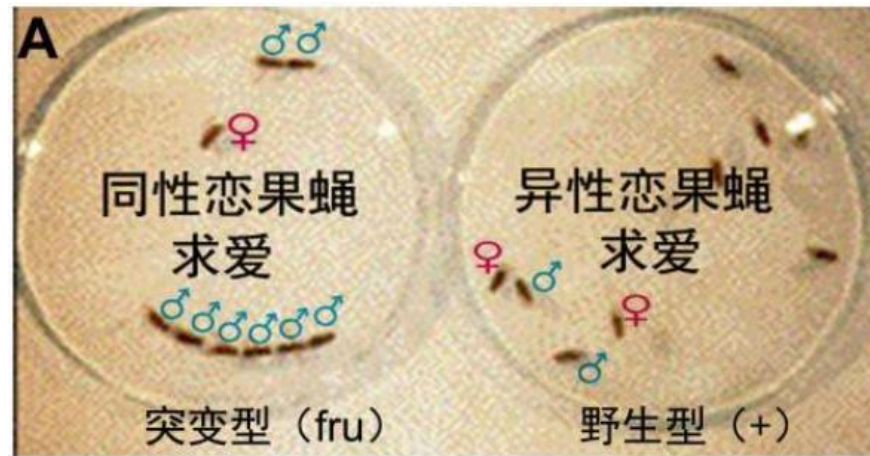
3. 分裂基因（由两段非连续DNA组成的基因）

蓝细菌*Synechocystis sp.*

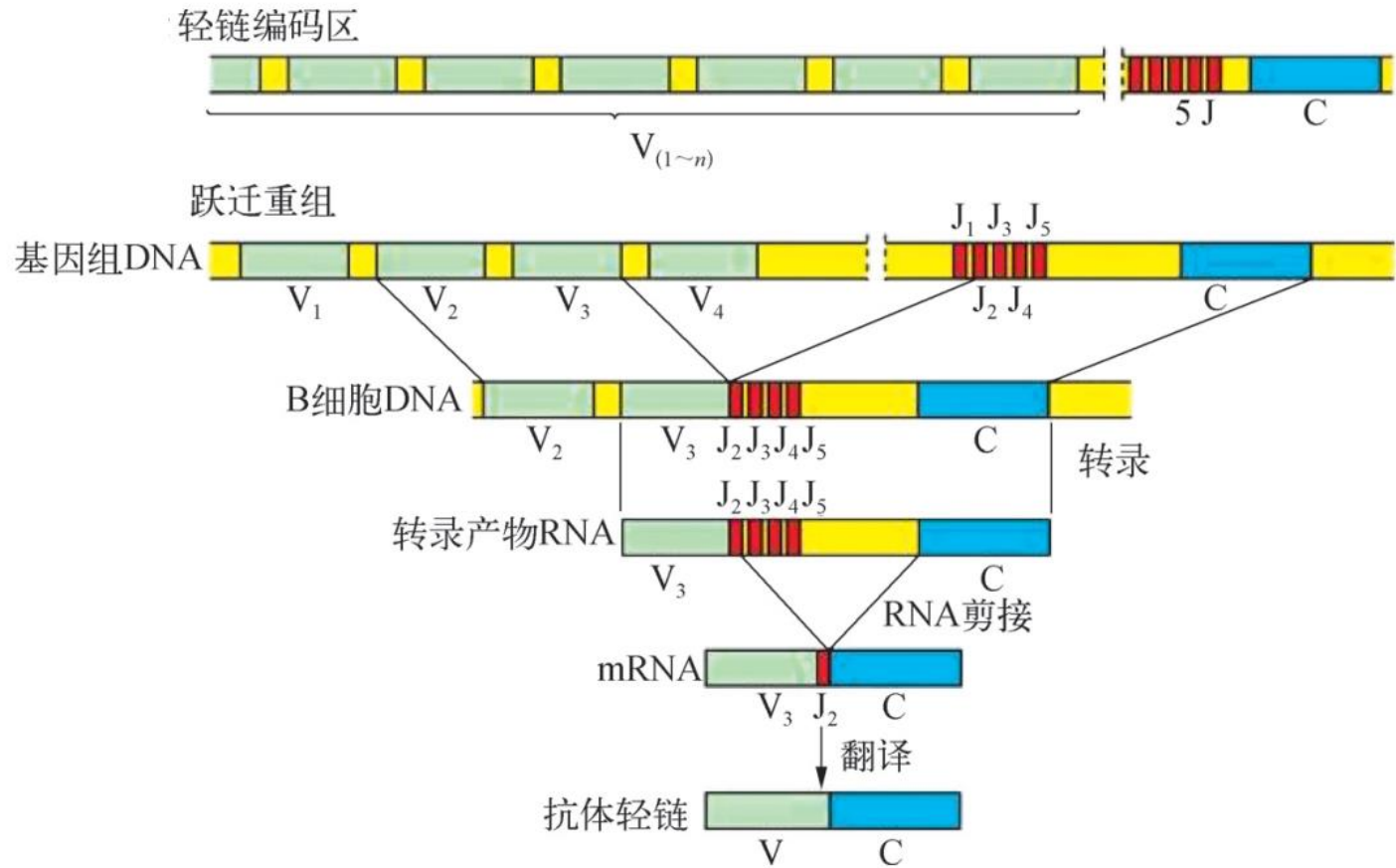
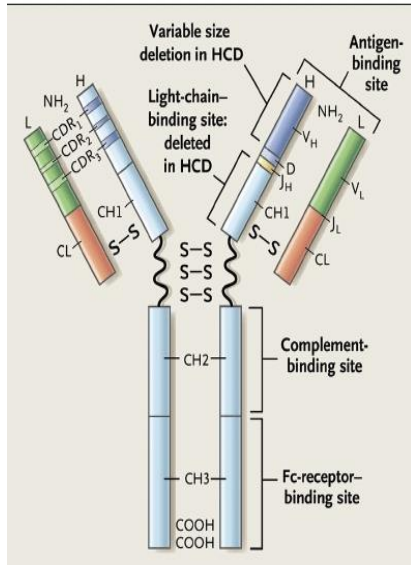
*PCC6803*基因组中编码DNA聚合酶III α 亚基的基因*dnaE*由两个分开表达的区段组成, *dnaE-n*和*dnaE-c*, 位于两条相反的DNA链的不同区段。两个独立结构的区段转录后翻译成两段独立多肽链, 经反式剪接组成活性蛋白。



4. 多启动子基因



5. 基因可重排



免疫球蛋白轻链的基因重排

6. 基因可移动



- Barbara McClintock
(1902-1992)
- 玉米籽粒色斑不稳定遗传
- 1951年，首次提出转座的概念，认为染色体上排列成串的基因并不是固定的，基因可以在染色体上以不规则地方式进行运动，即**转座 (transposition)**。
- 1983 Nobel Prize



六、基因组后续计划



除去1.5%的蛋白质编码序列，
剩余98.5%是垃圾DNA吗？



人与人之间的不同（疾病易
感性）又是由什么决定的？



利用个人基因组测序为个性
化医疗提供方案设计的可能？



NEWS FEATURE

NATURE Vol 445 | 24 June 2010



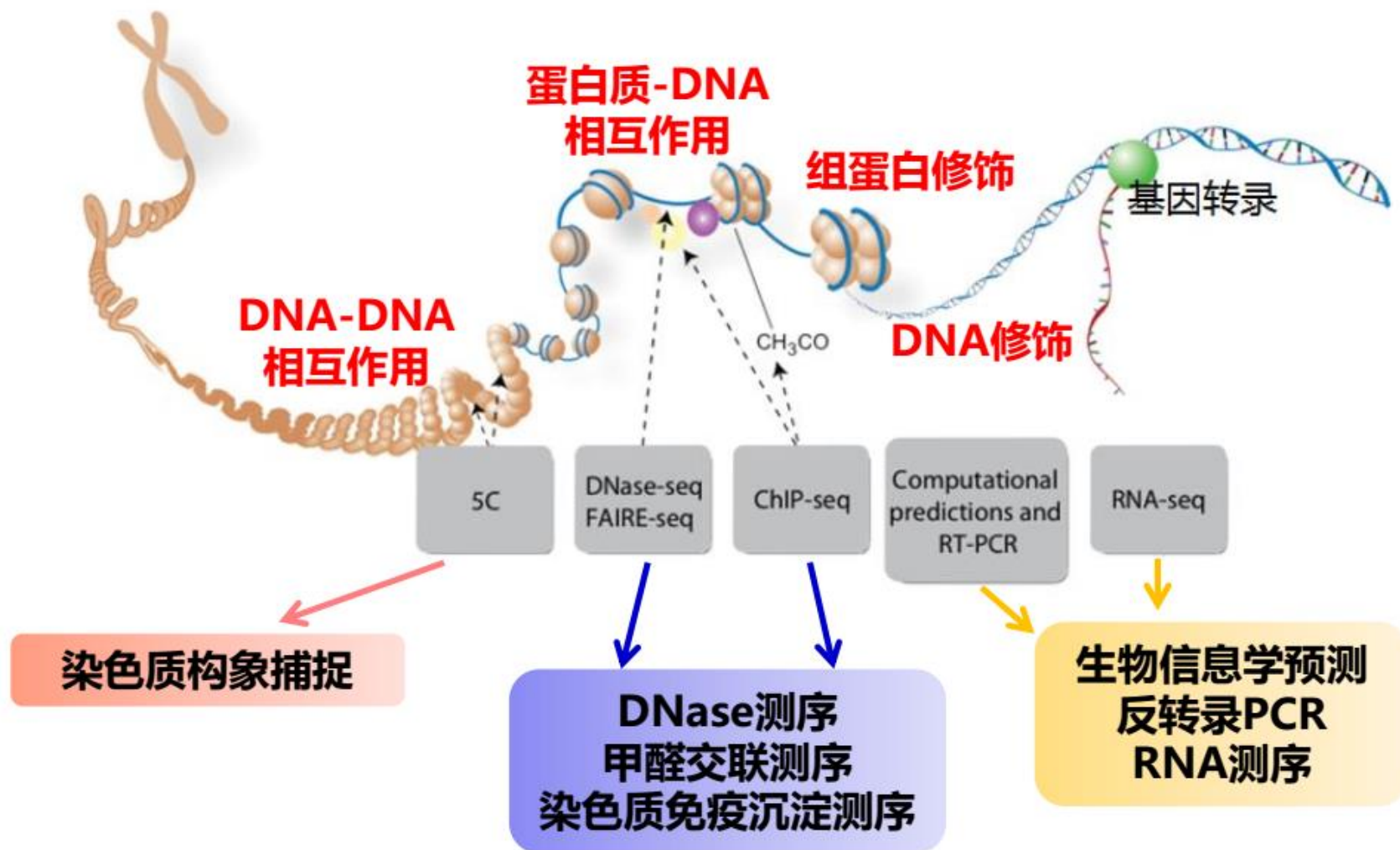
SCIENCE AFTER THE SEQUENCE

The completion of the draft human genome sequence was announced ten years ago. *Nature's* survey of life scientists reveals that biology will never be the same again. **Declan Butler** reports.

End of the beginning OR Beginning of the end

海量数据的注释工作

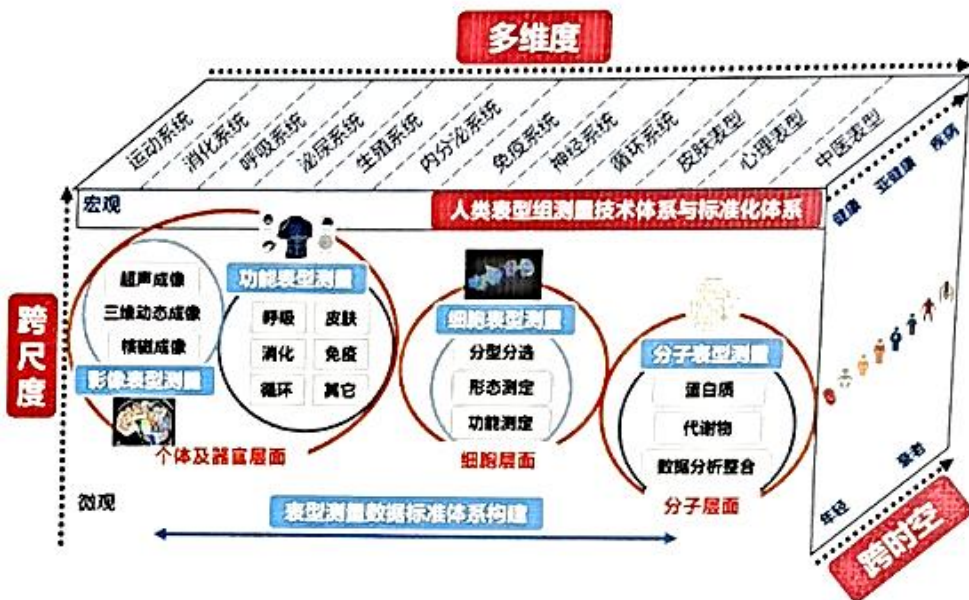
基础向临床的转化研究





国际人类表型组研究协作组

- 覆盖六大洲20个国家、23家机构
- 对人类的各种表型进行全尺度、全周期精密系统测量，从而解析基因-表型-环境之间以及宏观-微观表型之间的关联，进一步破解各种表型和人类健康与疾病的关系。
- 初步绘制了基于上海核心队列的全球首张人类表型组参比导航图



覆盖15个领域、2万个表型词条



公众号



...



复旦大学人类
表型组研究院

关注

HGP的意义和影响



一、创造了一种新的文化：合作

HGP是人类自然科学史上第一次影响最大的多国参与的国际合作计划，开辟了作为“国际化”的一个重要组成——国际科研合作的新篇章；

二、催生了一门新的学科：组学

HGP是基因组学的第一次实践，使基因组学成为科学并形成了自己的特点：从全基因组规模的广度和核苷酸水平的深度来研究生物学的所有问题

三、提供了一个新的技术：测序

HGP的运行过程，就是测序技术发展的过程。测序技术此后的发展，也应归功于HGP和基因组学和其他“组学”的推动。

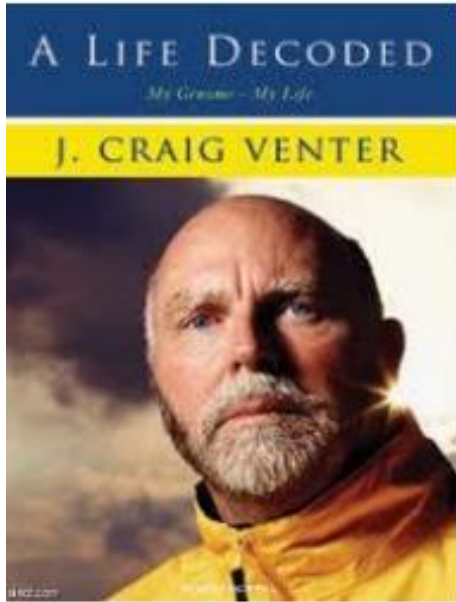


测序技术应用的六个方面

- 外显子和全外显子组测序——单基因性状与遗传病
- 全基因组测序——复杂性状与常见疾病
- 单细胞测序——基因组异质性
- META基因组测序——微生物及病原基因组
- 痕量DNA测序——无创检测、法医鉴定和古DNA研究



● 个体基因组分析：基于个体遗传信息的健康保健与精准医疗



第一个公开发表的个人基因组，和参考序列相比较：

- 有1.2%的差异, 涉及410万处, 78%是SNP; 44%基因存在杂合变异;
- 发现一段在参考序列中不存在, 但在黑猩猩中存在的序列
- DAT1基因带有10次重复的序列。小儿多动症?
- GSTM1杂合缺失, GSTP1携带ile105突变。哮喘?
- ApoE为E3/E4型; SORL1存在突变; 老年痴呆?
- CFN带有杂合突变。黄斑变性?

PLoS Biology 2007, 5(10):e254



癌症基因检测



- 多名一级亲属（母亲，父亲，姐妹，兄弟，孩子）患有癌症
- 父系或者母系一方的许多亲戚患有相同类型的癌症
- 家族史中的癌症，与某个基因突变相关
- 有与遗传性癌症相关的病变（例如，结肠息肉很多）
- 一个或多个已经进行过基因检测的家庭成员携带已知的增加患癌风险的基因突变



临床级：

- ✓ 产前诊断：三体综合征、单基因遗传病
- ✓ 新生儿筛查（出生后3-7天）：苯丙酮尿症、甲状腺功能减低
- ✓ 癌症患者：早期确诊、检测癌症相关基因突变、指导临床用药
- ✓ 癌症风险评估：前列腺癌、乳腺癌、结直肠癌等

消费级：结果仅供参考