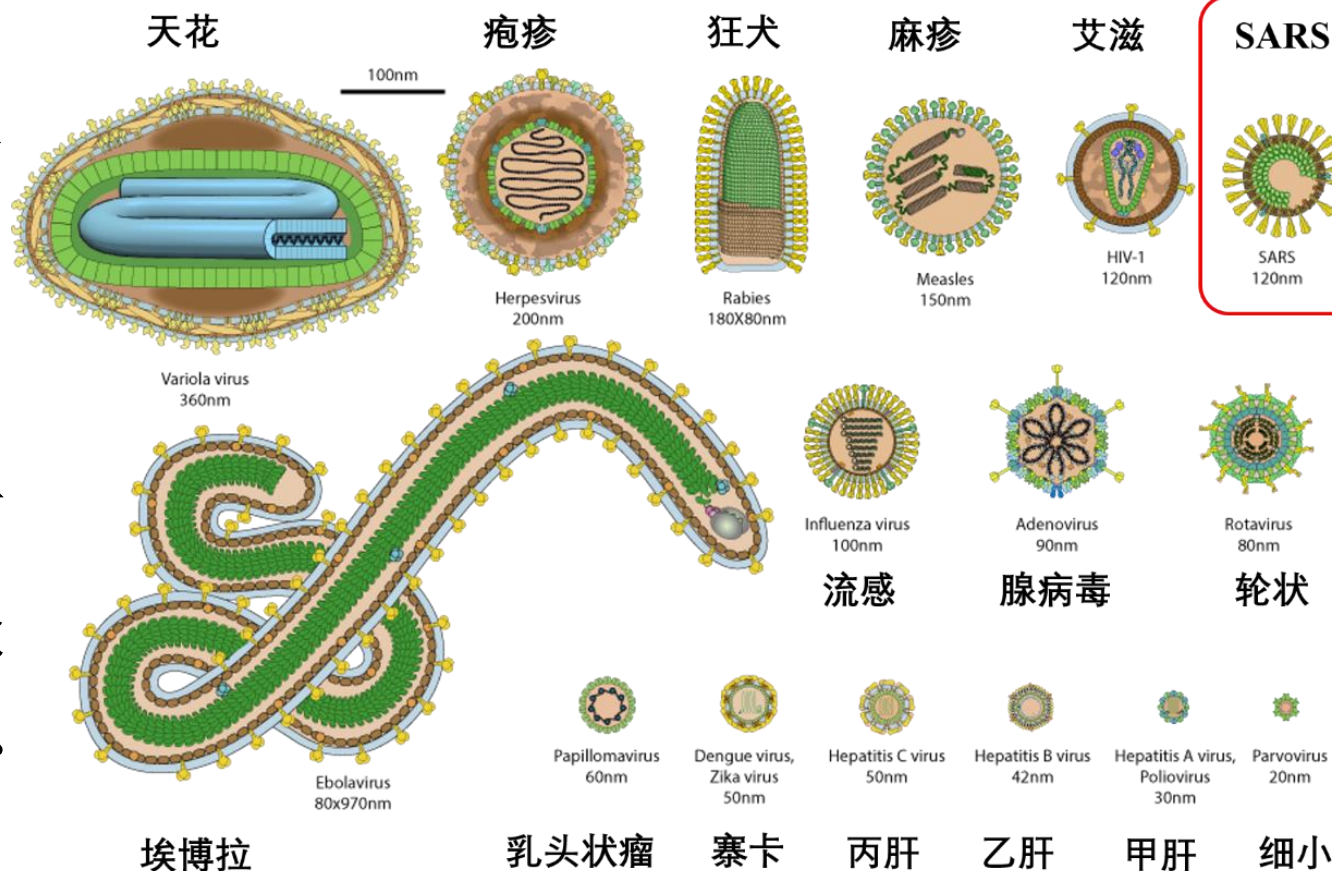


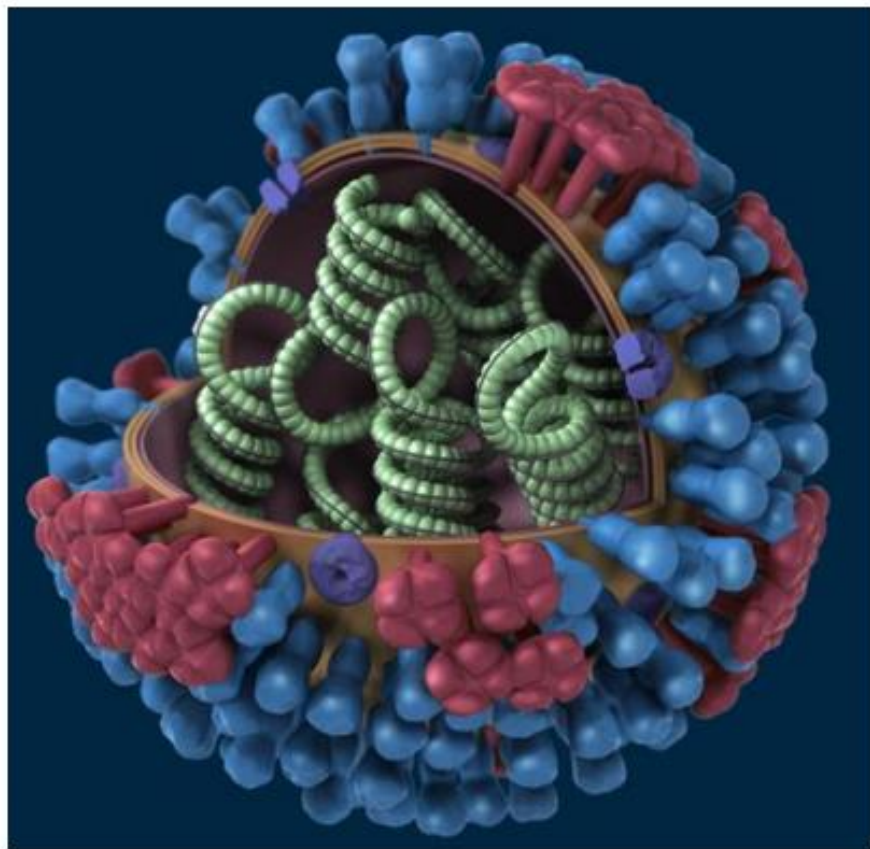
四、模式生物基因组

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

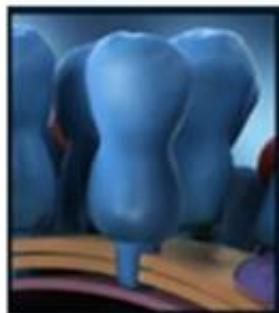
甲、乙、丙型流感



血凝素抗原H

粘住和进入细胞

宿主类型和传染力

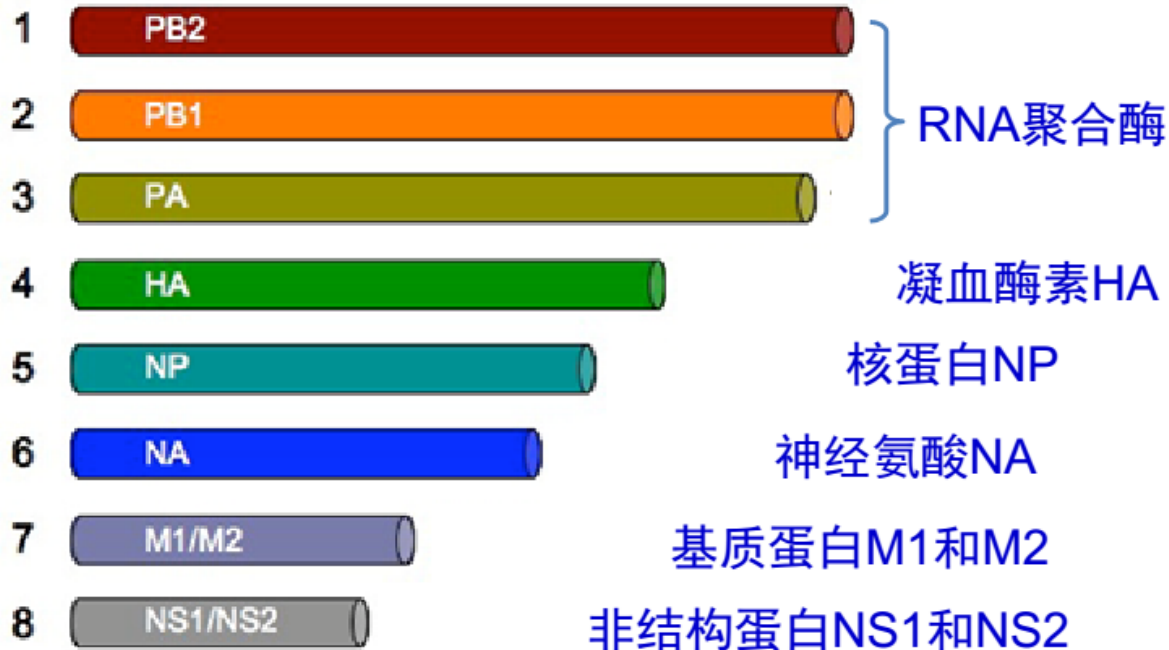
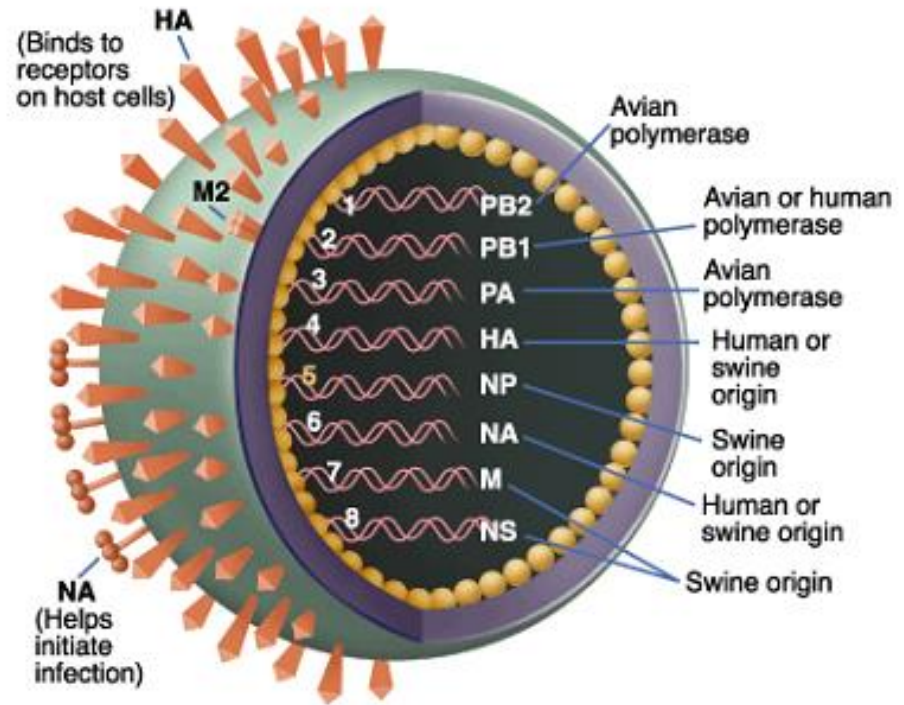


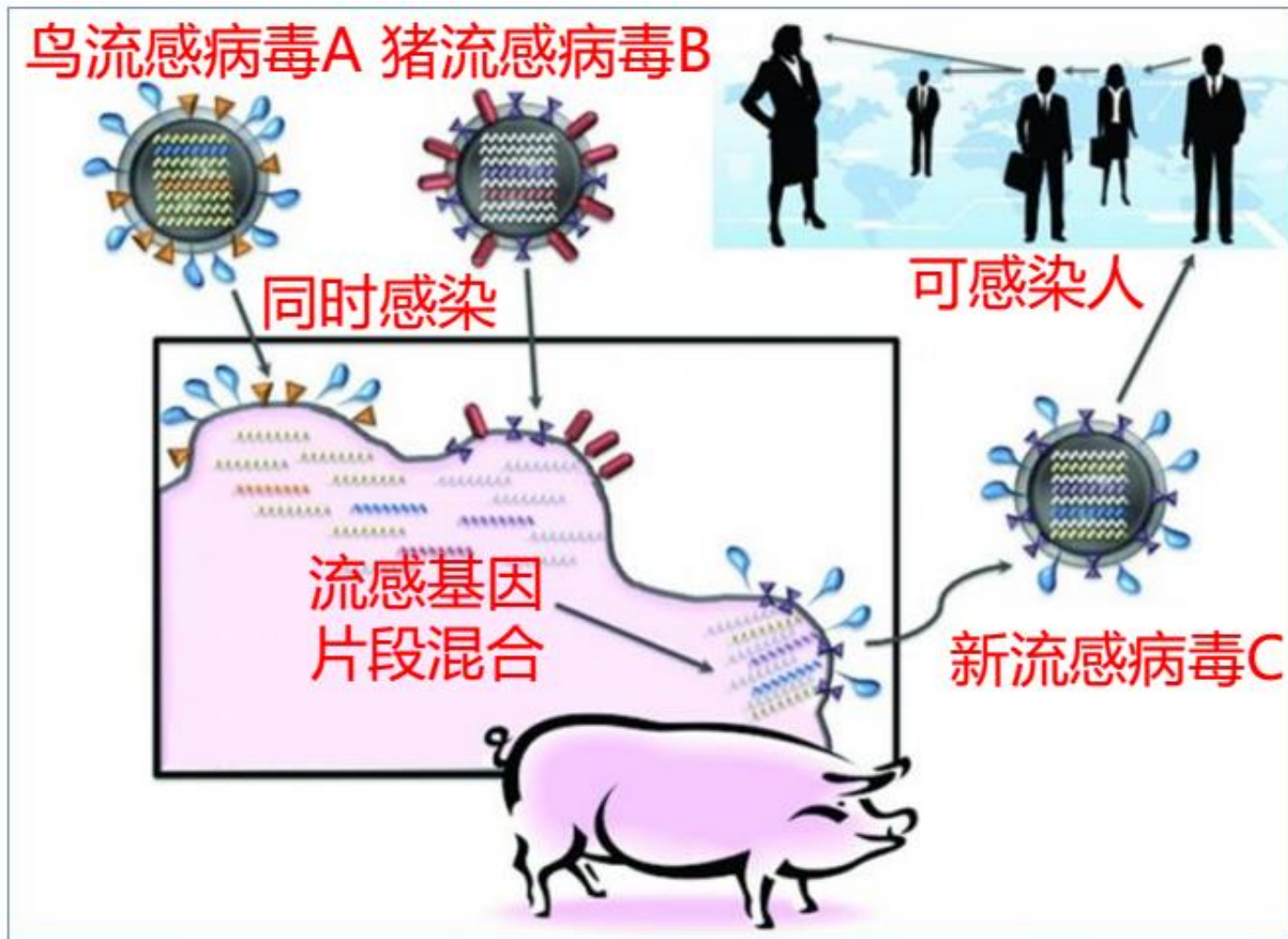
神经氨酸抗原N

细胞裂解

致死率

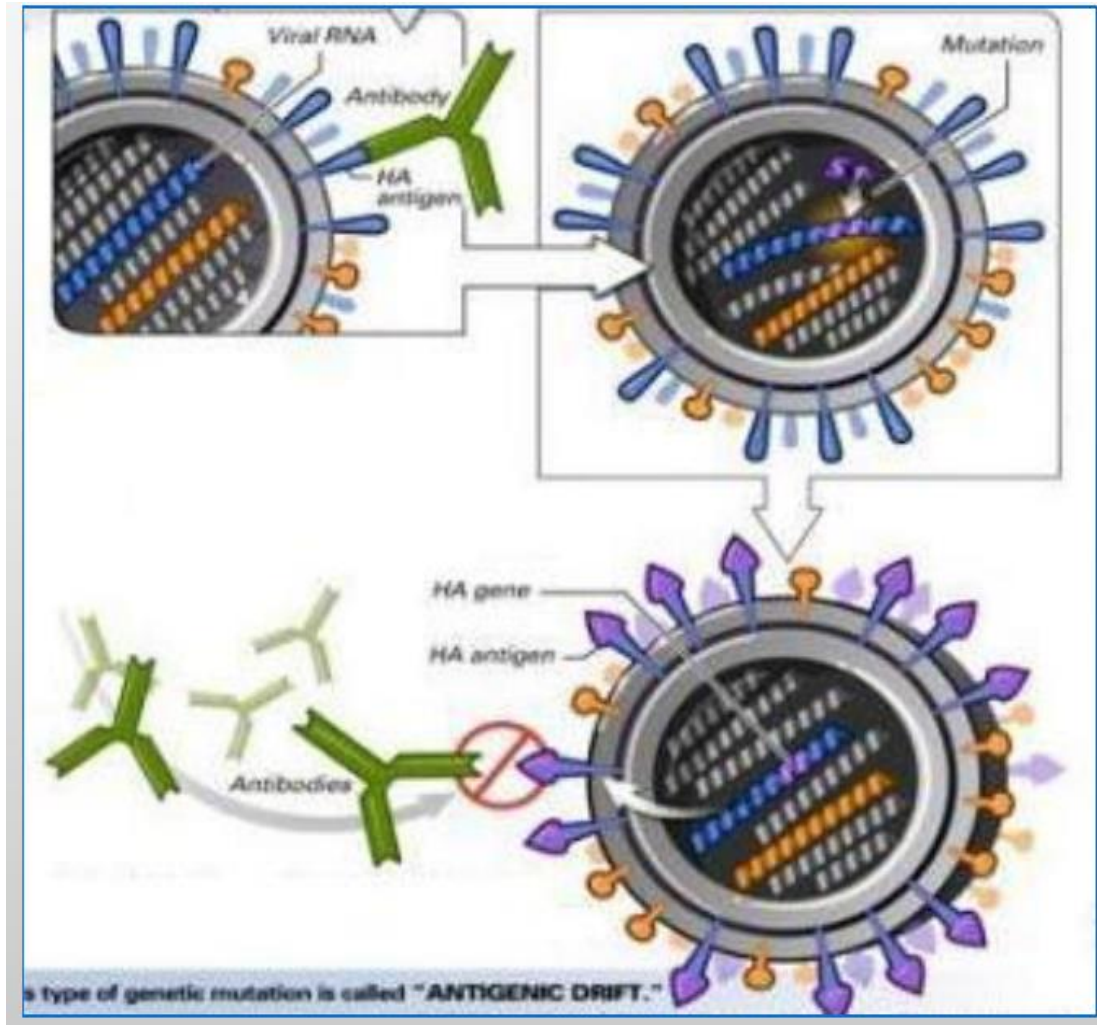
H1N1流感病毒为单股RNA病毒，基因组13.6kb，含8个独立的基因片段，编码11个蛋白质





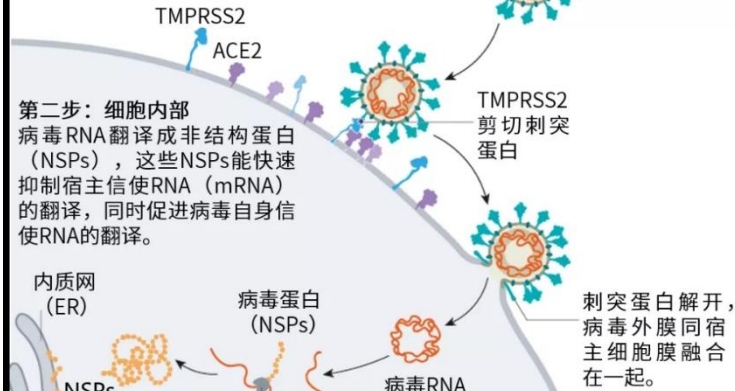
抗原转变 (antigenic shift): 流感病毒的基因序列以独立的片段存在于病毒体内，当不同病毒感染同一宿主细胞时，易发生基因片段之间的重新装配，产生基因组变异。

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。

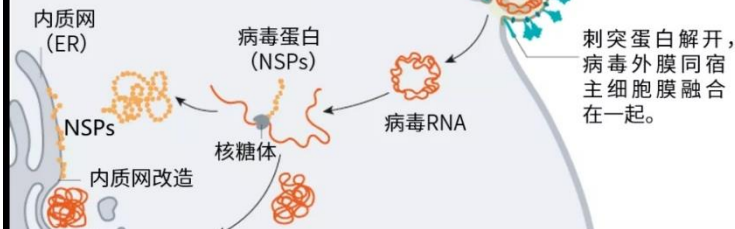




第一步：病毒入胞
新冠病毒的刺突蛋白与名为ACE2的宿主细胞受体结合后，宿主细胞分子TMPRSS2便会剪切刺突蛋白，暴露出的部位让病毒外膜与宿主细胞膜融合。

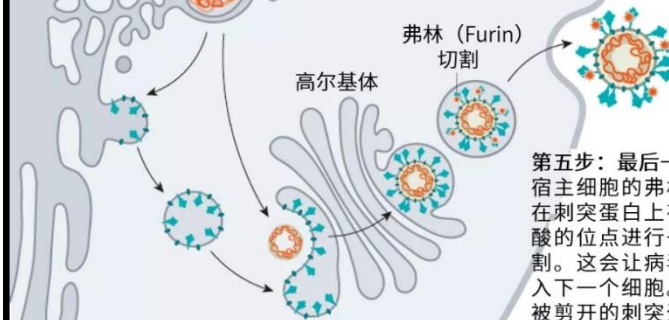


第二步：细胞内部
病毒RNA翻译成非结构蛋白（NSPs），这些NSPs能快速抑制宿主信使RNA（mRNA）的翻译，同时促进病毒自身信使RNA的翻译。

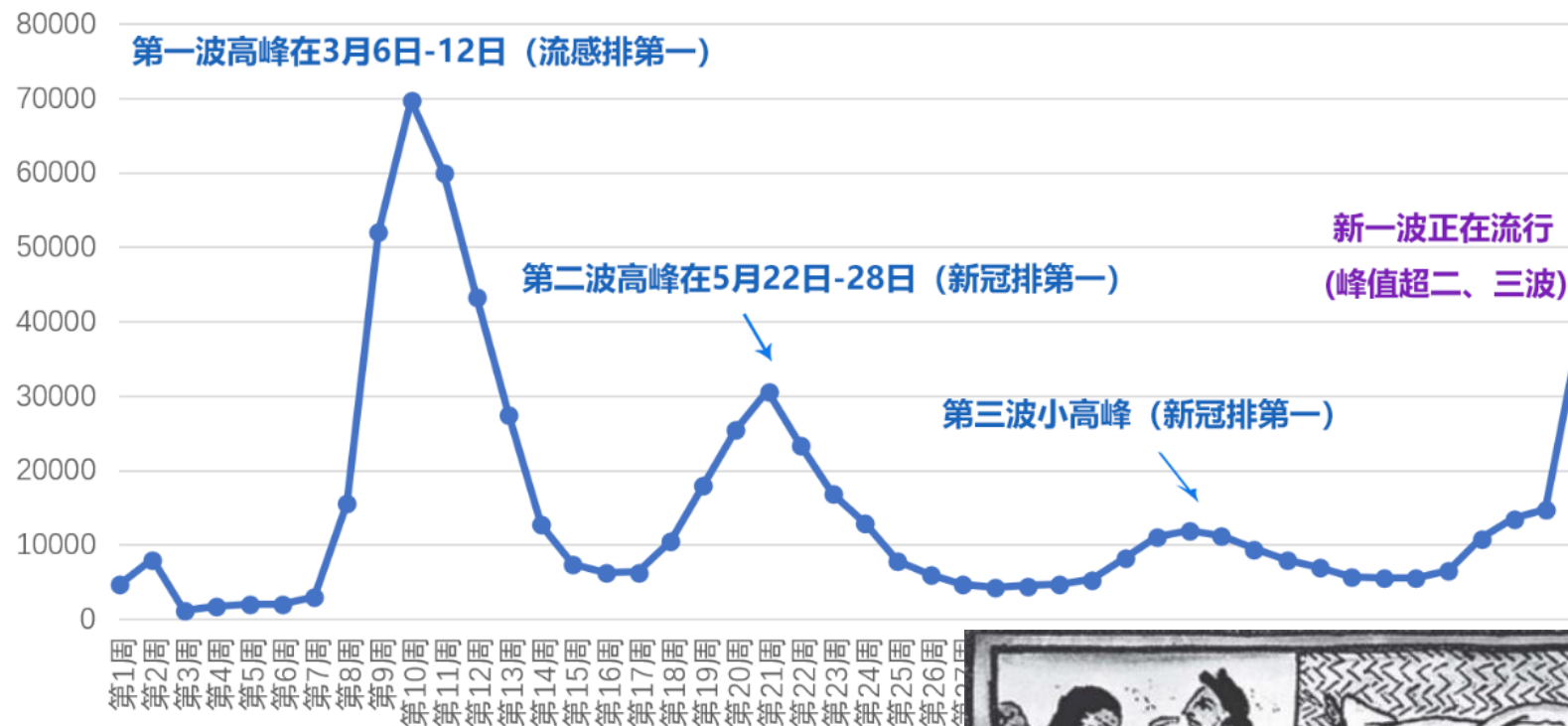


第三步：细胞改造
新冠病毒将细胞的内质网（细胞内膜系统）改造为名为双层膜囊泡（DMVs）的泡状结构。这会给更多的病毒RNA 提供一个避风港，让它们可以安全地复制和翻译。

第四步：出胞
一旦新产生的分子组装成一个完整的病毒颗粒，就会通过名为高尔基体或溶酶体（细胞垃圾桶）的细胞器离开细胞。



第五步：最后一切
宿主细胞的弗林蛋白酶会在刺突蛋白上有五个氨基酸的位点进行关键切割。这会让病毒准备好侵入下一个细胞。变异株中被剪开的刺突蛋白比例更高，能帮助它们更快地感染细胞。



免疫落差（immunity gap）：

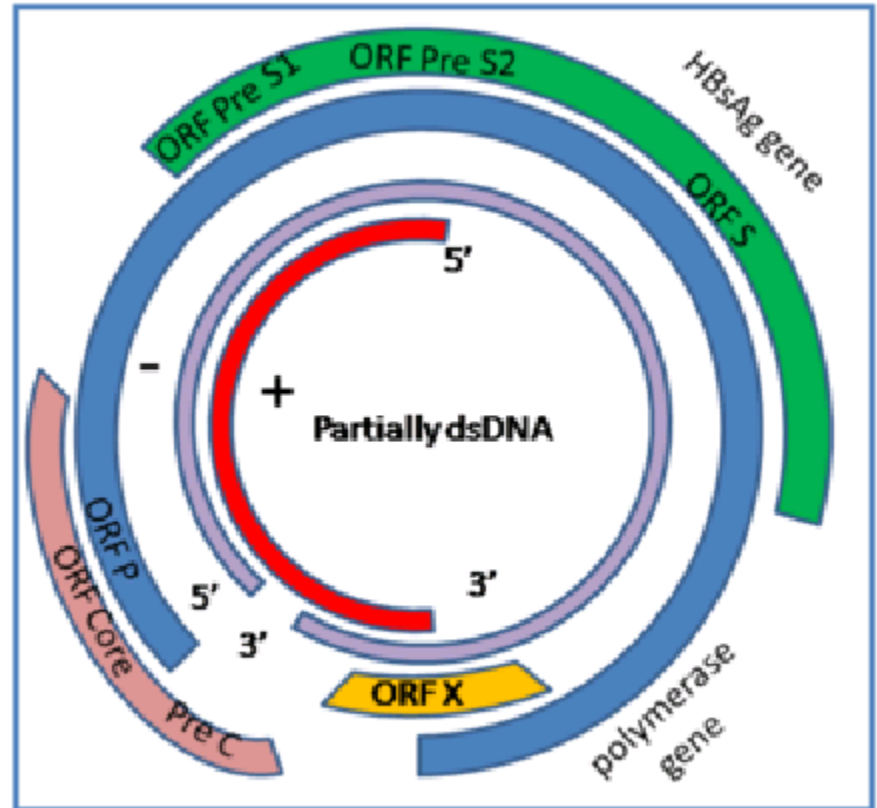
长期低接触某种特定病原体而导致的保护性免疫力不足，这使得更高比例的人口易感染该疾病。



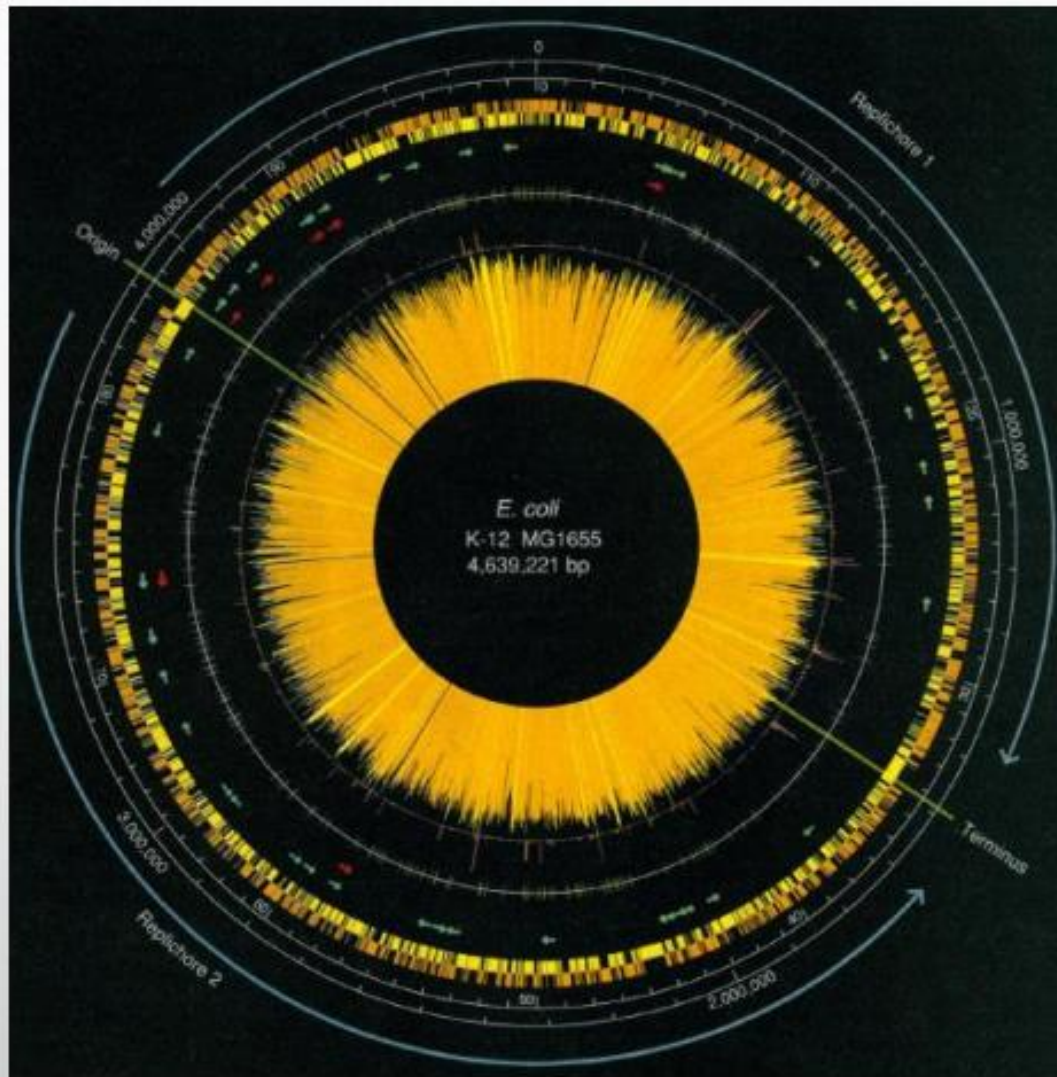
天花感染印第安人（《疾病改变历史》）

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.

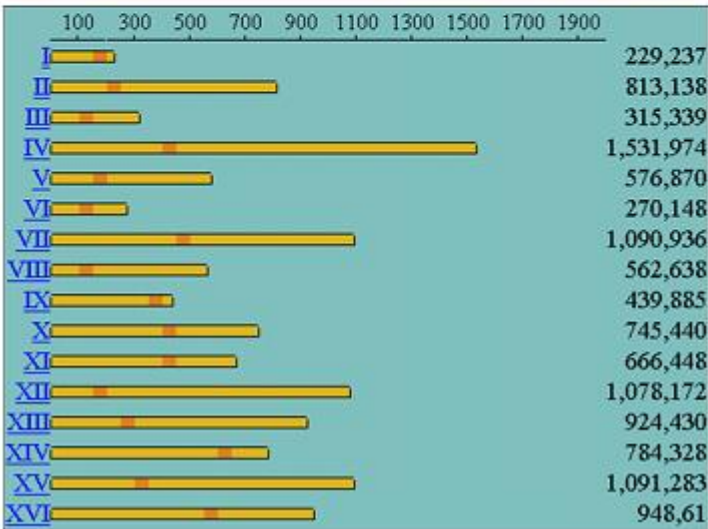
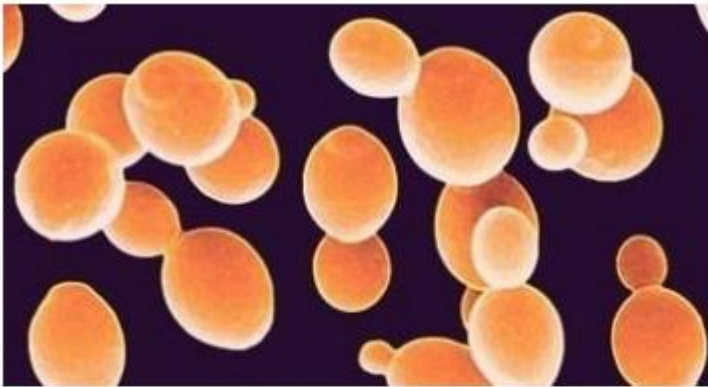
- 闭合环状双链DNA，大小为4.6Mb;
- 87.8%为蛋白编码序列，含4288个蛋白编码基因; 0.8%编码RNA，0.7%为非编码重复序列，余下10.7%为表达调控和未知功能序列;
- 两条DNA链均可作为模板进行基因转录，基因间平均间隔为118bp。



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。

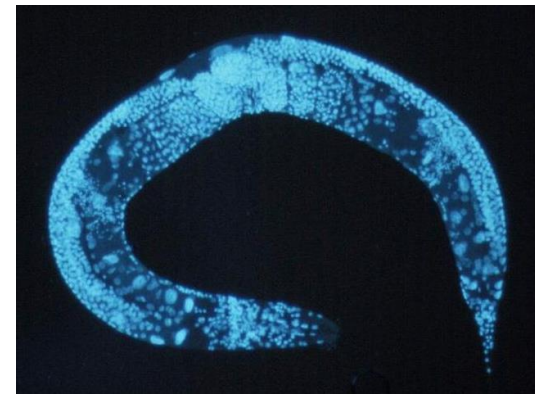
基因组的组成及特点——酿酒酵母 (*Saccharomyces cerevisiae*)



- 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）

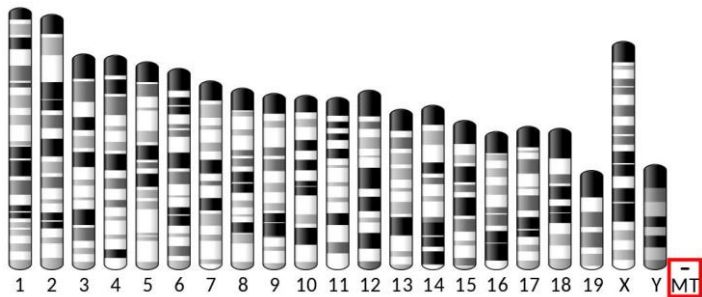
- 2000年，第一个植物拟南芥（*Arabidopsis thaliana*, 115Mb），重复序列仅占基因组13%，多为LTR



基因组的组成及特点——小鼠 (*Mus musculus*)

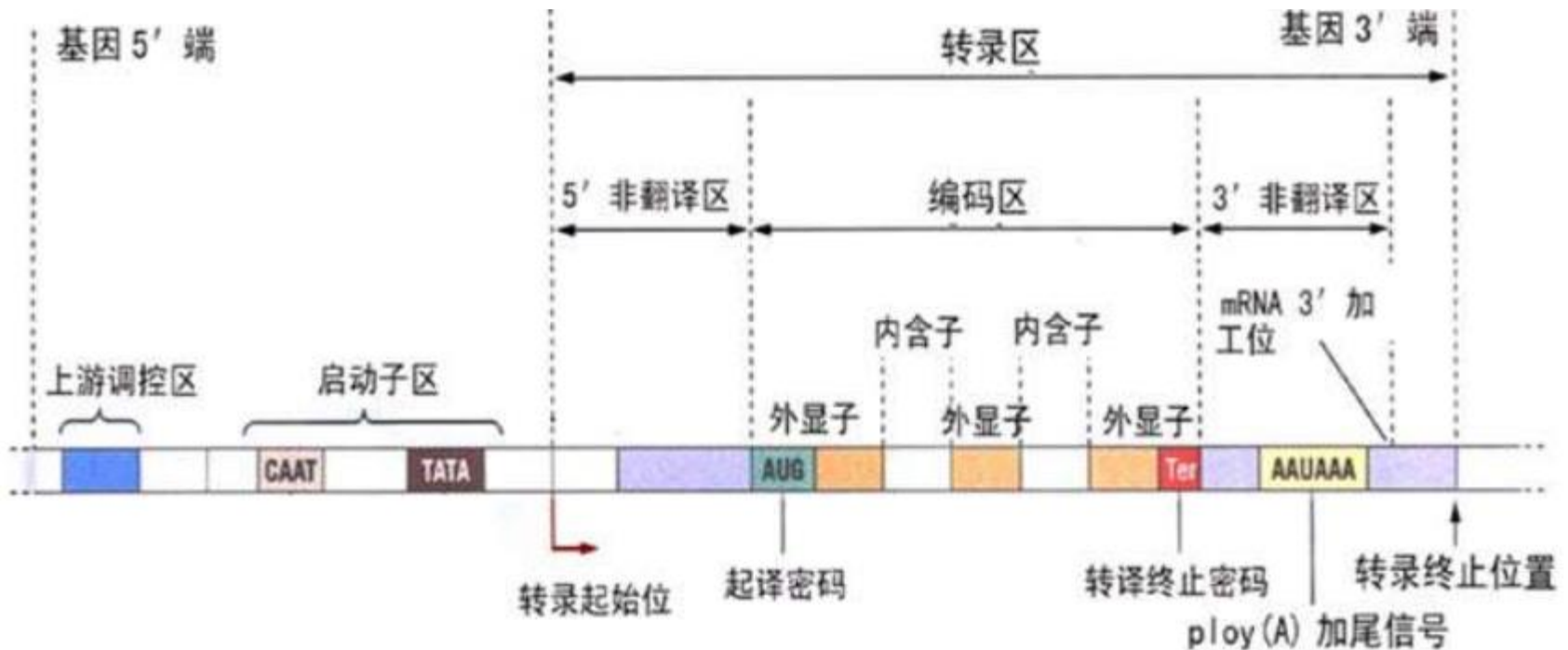


- 继人类后发表的第二个全基因组序列的哺乳动物（2002年）；
- 基因组全长2.5 Gb，单倍体共20条染色体且均为端着丝粒型；
- 小鼠与人类之间有90%以上的相似性，约有40%的小鼠和人类基因组序列高度同源；
- 编码基因约2.3万个，85%的人类基因在小鼠基因组中能找到同源或相似基因，在小鼠和人类中大约有100+个印记基因。



五、基因概念的拓展

- **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。

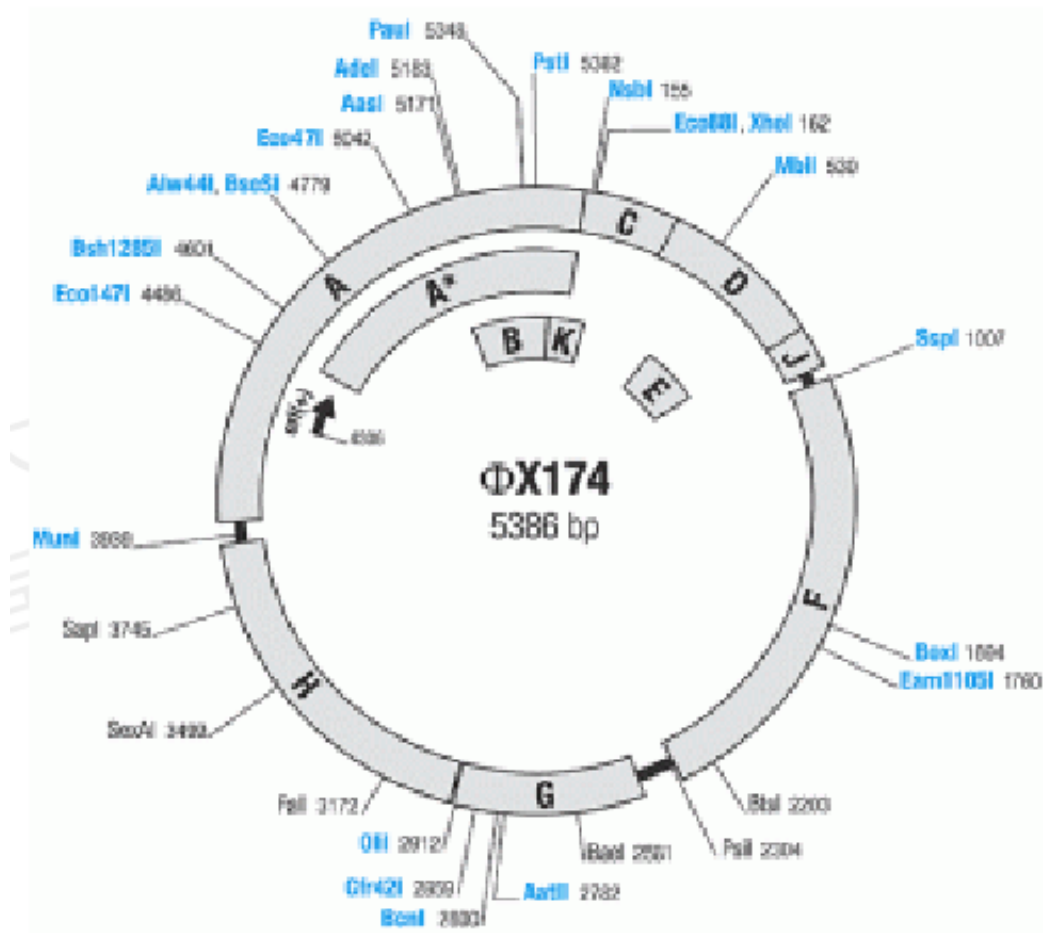


特殊结构基因:



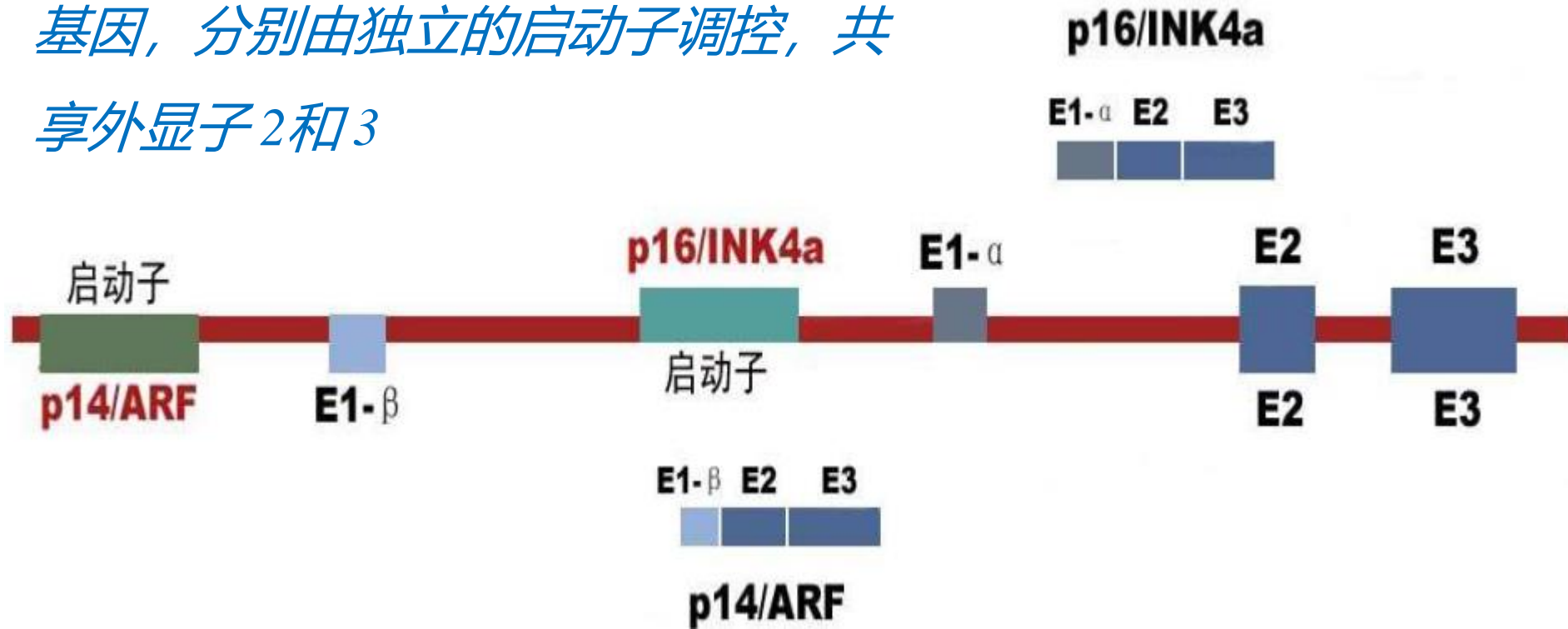
1. **重叠基因(overlapping gene)**: 指两个或两个以上的基因共有一段DNA序列, 或者是一段DNA序列成为两个或两个以上基因的组成部分。

- 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



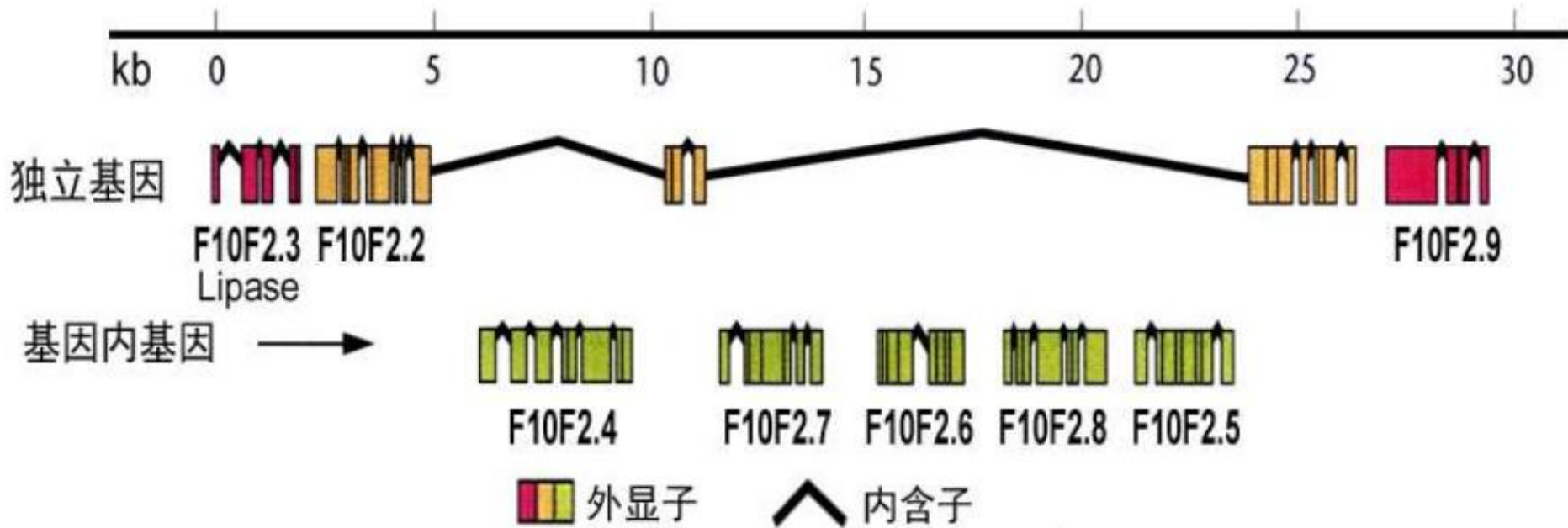
人类的重叠基因

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

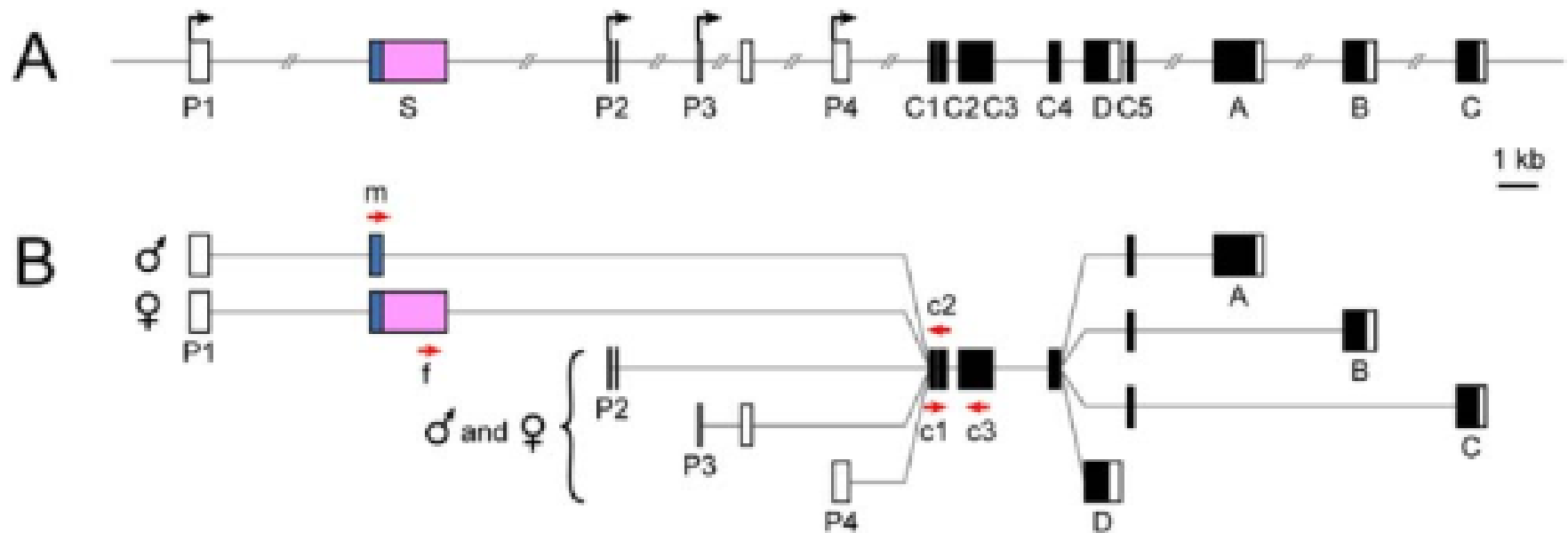
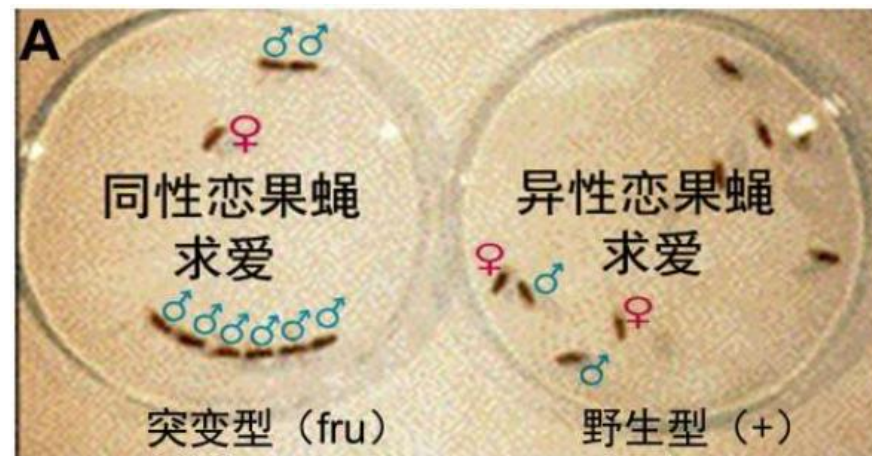


线虫基因组中一段30kb的区段中含有8个基因，其中 *FGAM*(甘氨酸酰胺合成酶) 基因长25kb,有21个内含子。内含子9中含有一个独立的基因，内含子11中含有4个独立的基因。基因内基因分别具有独立的外显子和内含子。

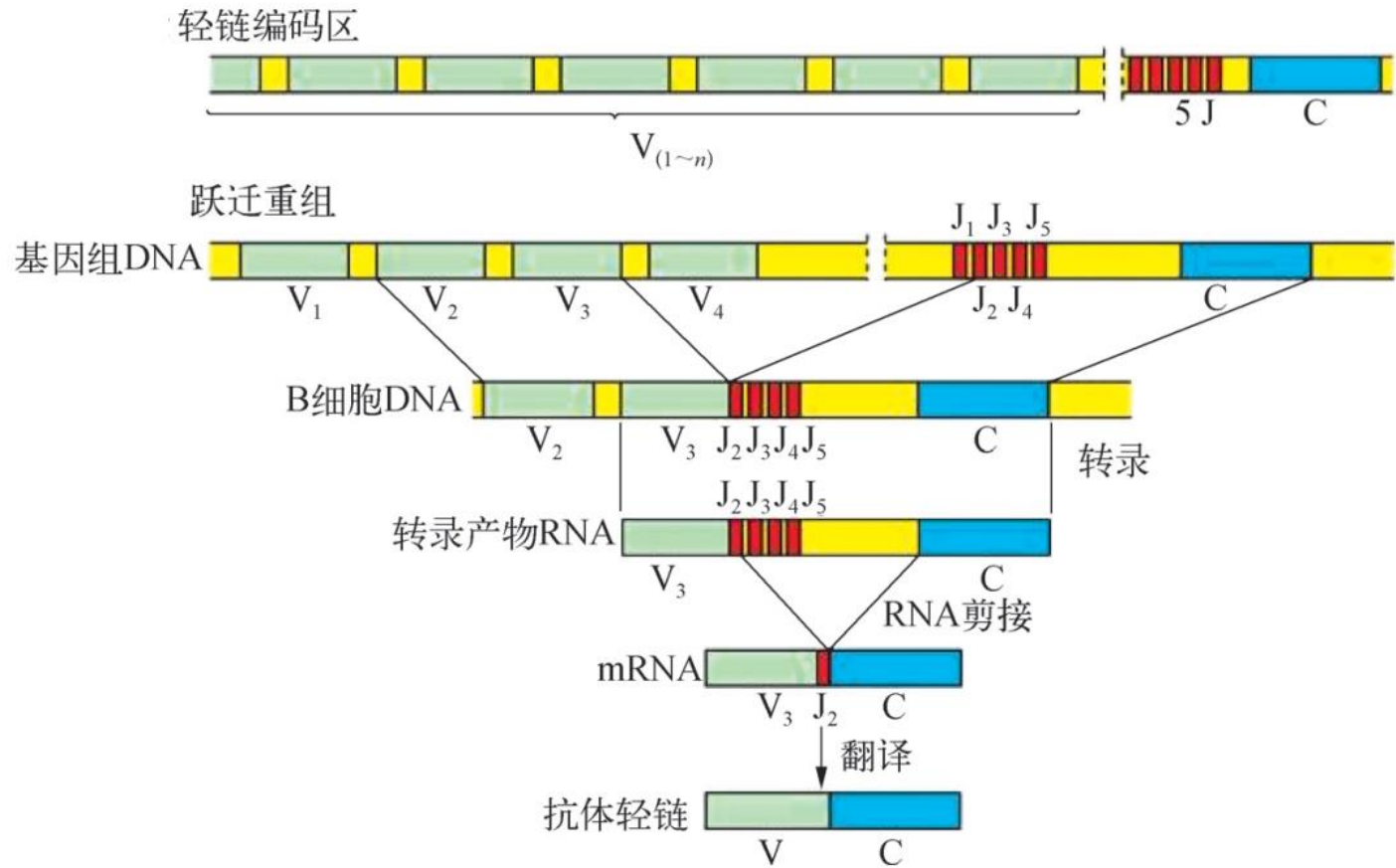
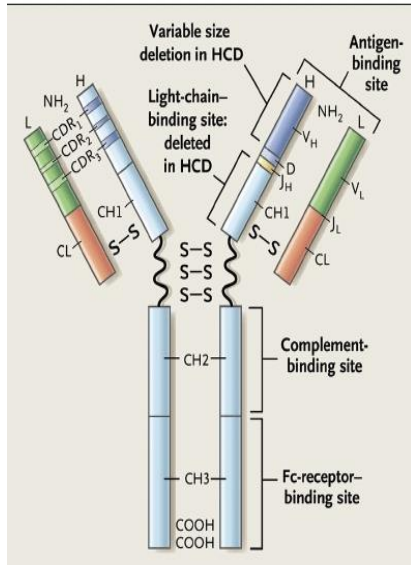
PCC6803基因组中编码DNA聚合酶III α 亚基的基因dnaE由两个分开表达的区段组成, dnaE-n和dnaE-c, 位于两条相反的DNA链的不同区段。两个独立结构的区段转录后翻译成两段独立多肽链, 经反式剪接组成活性蛋白。



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的方式进行运动，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.

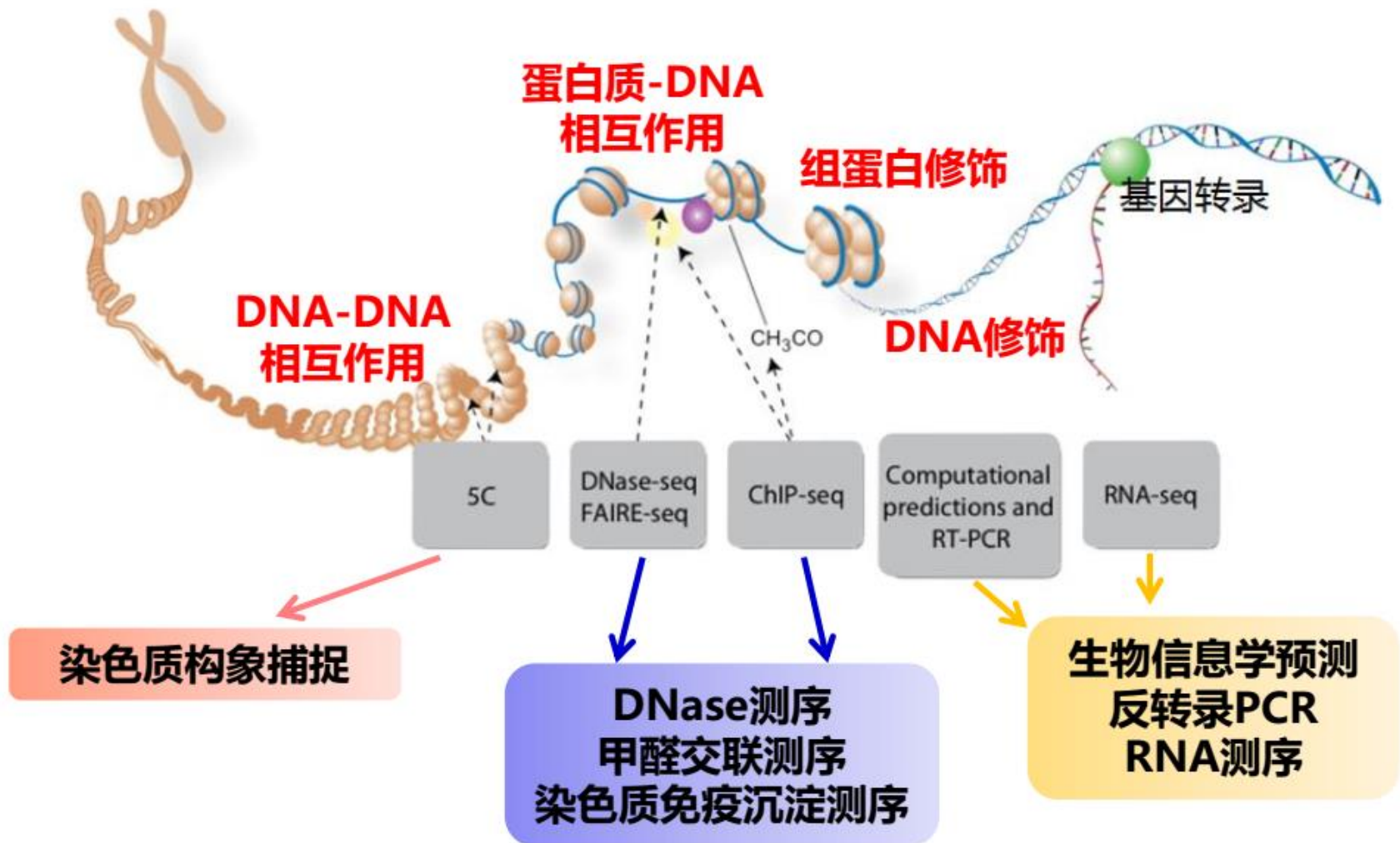
End of the beginning OR Beginning of the end

海量数据的注释工作

基础向临床的转化研究

1. Encyclopedia of DNA Elements, **ENCODE**

(DNA元件的百科全书计划, 2003)





<https://www.encodeproject.org/>

ENCODE Data Encyclopedia Materials & Methods Help

Search the ENCODE Portal ENCODE ☐ SCREEN ☒

Functional genomics

Functional characterization

Encyclopedia of elements

Rush Alzheimer's

EN-TE_x

Deeply profiled cell lines

Protein knockdown (Degron)

Computational and integrative products

Human donors

2. 国际千人基因组计划

- 由英国、美国和中国发起主导，研究人员对27个种群的2500个样本进行研究、构建人类基因组遗传多态性图谱；
- 该图谱包含约15万个SNPs，100万个插入/缺失，2万个结构变异；
- 平均每个人大概携带250-300个失去功能的突变，其中50到100个与遗传病有关；
- 某些疾病的发病率在种族之间存在明显差异。如高血压发病率在白种人中为5-7%，而在黑人中可高达20-30%，黄种人最低；



<https://www.genomicsengland.co.uk>



Genomic Medicine

Our Initiatives

Patients and Participants

Research and Partnerships

News and Events

About Us

Overview

Getting your results

Additional findings

Participant resources

Document library

The impact of the project

The 100,000 Genomes Project was a British initiative to sequence and study the role our genes play in health and disease. Recruitment was completed in December 2018, although research and analysis is still ongoing.

Our participants have already helped us find actionable results for many patients with rare diseases and cancer.

Key facts



18.5%

of data from the
Project turned into
actionable findings



85K+

participants'
genomes sequenced
for the Project



100K+

genomes sequenced
by December 2018



The future of health begins with you.

The *All of Us* Research Program is inviting one million people across the U.S. to help build one of the most diverse health databases in history. We welcome participants from all backgrounds. Researchers will use the data to learn how our biology, lifestyle, and environment affect health. This may one day help them find ways to treat and prevent disease.

[JOIN NOW!](#)

3. 泛基因组 (pangenome)：一个物种内所有基因信息的总和，比单一参考基因组含有更多的遗传信息，体现物种遗传多样性。

泛基因组的分析内容（了解）：

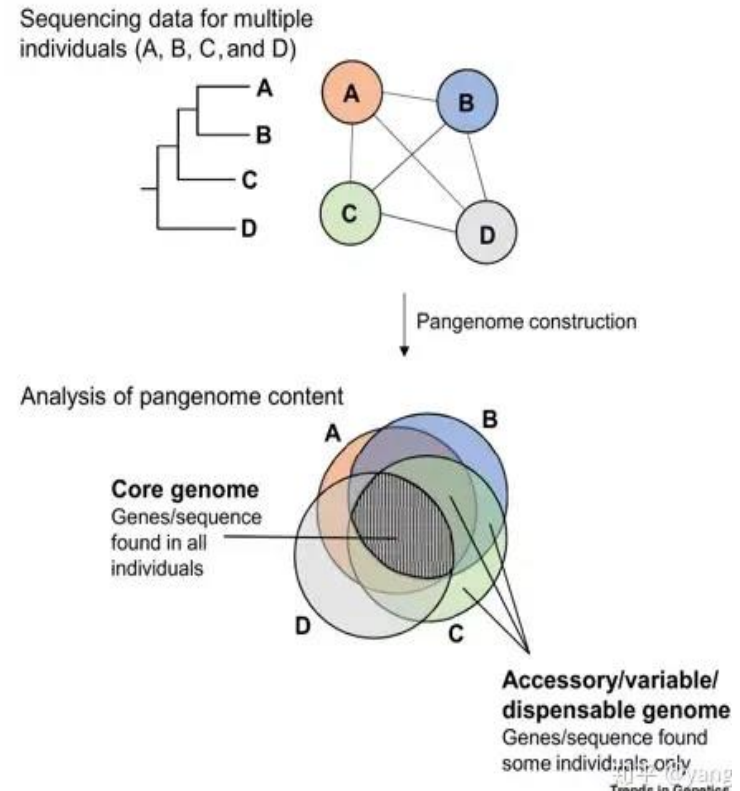
1. 物种特征分析：核心基因组的大小（物种里共有多少基因/基因家族）、泛基因组的大小（群体变异性）；

2. 获得物种全面且准确的变异信息 (SNP、Indel、CNV、PAV)。

3. 鉴定出参考序列中不存在的基因，进行新基因的功能分析。

4. 对物种进行广而深入的进化研究，以了解物种的起源与演化。

5. 泛基因组进阶分析：调控区差异分析、核心基因和可变基因的调控网络构建



Chinese Pangenome Consortium

[Home](#) | [Phase I](#) | [Participants](#) | [Data](#) | [Publications](#) | [Conference Talks](#) | [About](#)

CPC

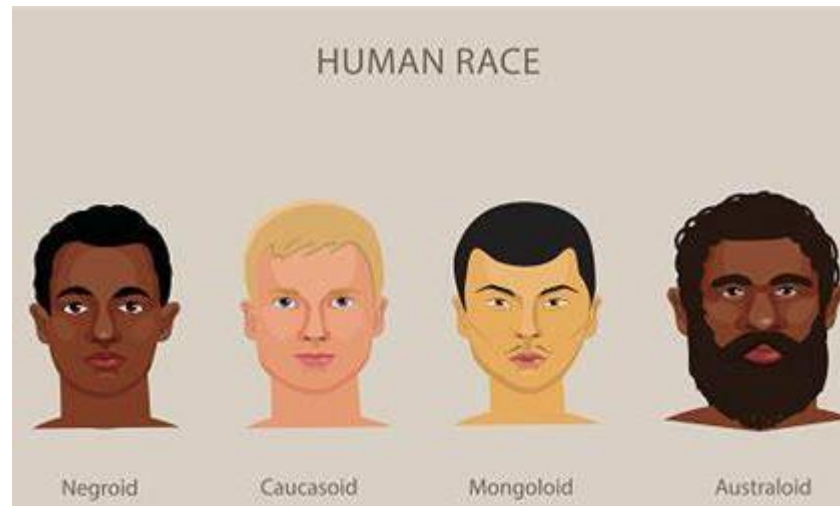
An effort toward understanding genomic diversity and structural variation of natural populations in China and neighboring countries



4. 中国科学家的重要贡献

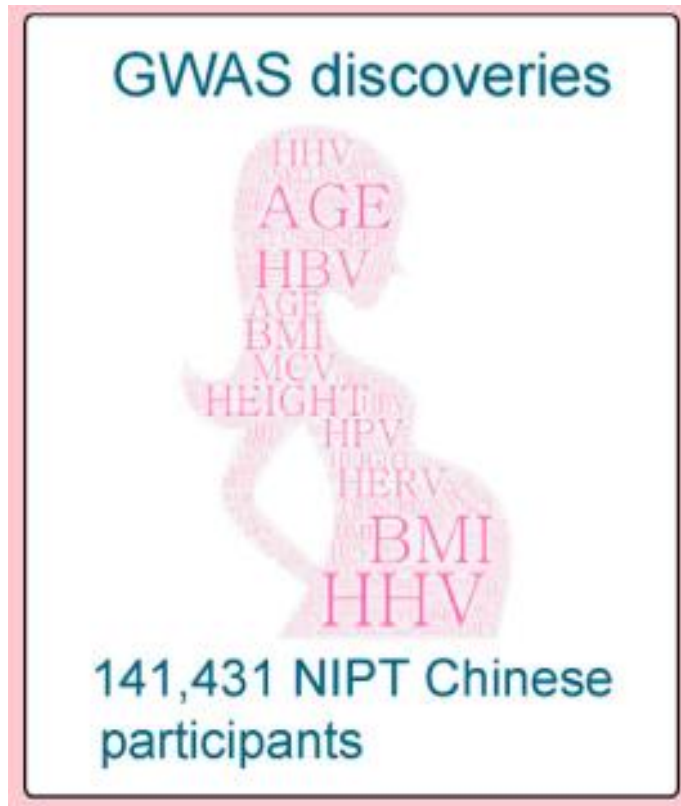


- 全球人类基因组测序计划的1%;
- 发布水稻、家蚕、鸡等基因组序列;
- 2006年, 第一个100%中国人基因组图谱----“炎黄一号”测序计划启动;
- 2012年, 提出百万人基因组计划 (Genome of 1 Million, G1M), 旨在揭秘中国人群中遗传特征和基因差异。



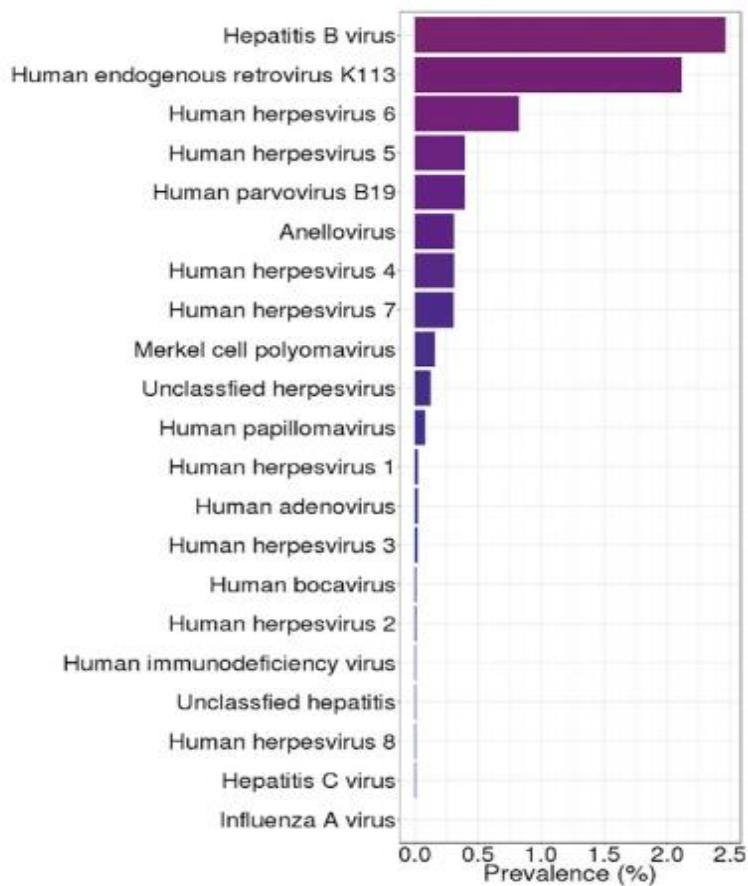
	Celera (2001)	HUGO (2004)	YH 1.0 (2015)
测序对象	混合人种	高加索人种	蒙古人种
测序方法	全基因组鸟枪法	逐个克隆法	二代测序+全基因组鸟枪法
测序速度	快	慢	快
单碱基准确度	低	高	高
覆盖深度	5.1 倍	6-10 倍	36 倍
总长度	2.91 Gb	2.8 Gb	5.15 Gb
缺口	数千个	341 个	基本完整

Large-scale analysis of genome sequences from non-invasive prenatal testing in Chinese women yields insights into phenotypic trait associations, viral infection patterns, and population history.

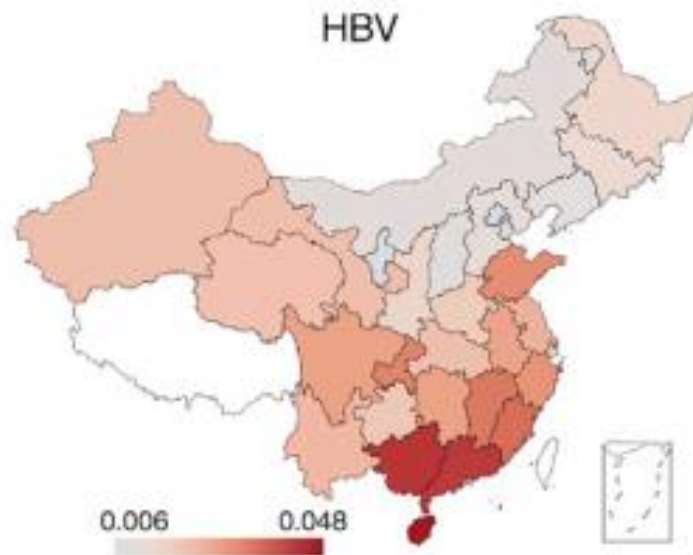


Cell. 2018 Oct 4;175(2):347-359.

A



4) 人群中病毒感染率



<https://db.cngb.org/cmdb/>

CHINESE MILLIONOME DATABASE

SEARCH FOR YOUR INTEREST

Search for a gene or variant or region



Please Log In before searching



ARTICLE OPEN

The ChinaMAP analytics of deep whole genome sequences in 10,588 individuals

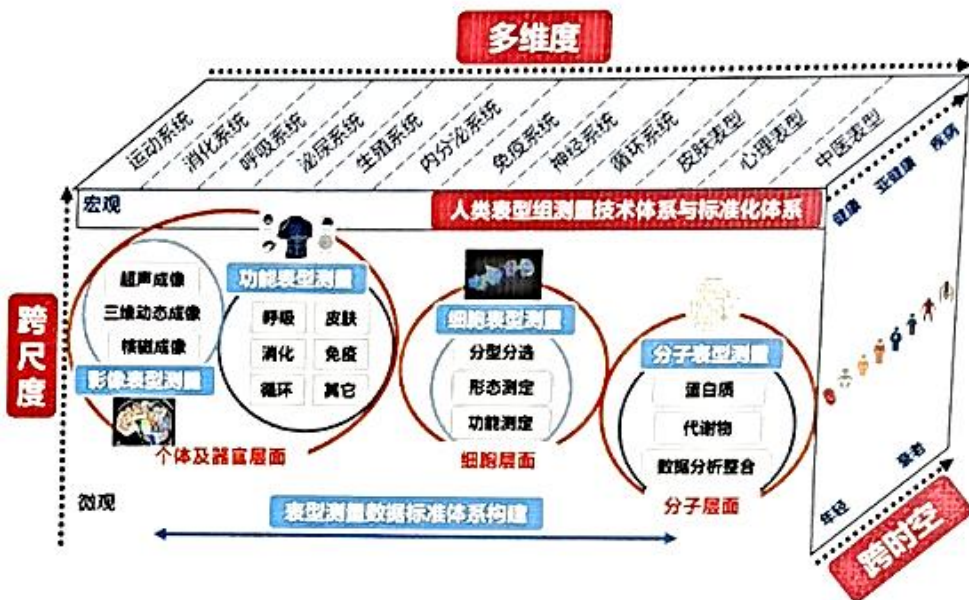
Yanan Cao^{1,2}, Lin Li^{1,2}, Min Xu¹, Zhimin Feng¹, Xiaohui Sun¹, Jieli Lu¹, Yu Xu¹, Peina Du¹, Tiange Wang¹, Ruying Hu³, Zhen Ye³, Lixin Shi⁴, Xulei Tang⁵, Li Yan⁶, Zhengnan Gao⁷, Gang Chen⁸, Yinfei Zhang⁹, Lulu Chen¹⁰, Guang Ning¹, Yufang Bi¹, Weiqing Wang¹ and The ChinaMAP Consortium

- **中国代谢组分析计划 ChinaMAP (China Metabolic Analytics Project)**
- **数据来源于三个大型队列:** 1) The China Noncommunicable Disease Surveillance 2010, 150,000 participants; 2) the Risk Evaluation of cancers in Chinese diabetic Individuals: 250,000 participants; 3) the Community based Cardiovascular Risk During Urbanization in Shanghai: 50,000 participants. **总计45万名参与者。**
- **随机选取了10588位参与者进行深度测序(40X), 参与者来自全国27个地区, 包括汉族、回族、满族、苗族、蒙古族、彝族、藏族和壮族8个族裔, 参与者平均年龄54岁, 女性占64.8%。**



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



5. 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。



在基因组研究的发展和应用过程中，是否也会出现一些问题，甚至是弊端？我们该如何对待？

