

24 人类基因组

王强

November 14, 2018

南京大学生命科学学院

Outline

24.1 人类基因组及其研究

24.2 人类遗传性疾病

24.3 癌基因与恶性肿瘤

24.1 人类基因组及其研究

24.1.1 基因组及基因组学

基因组

一个生物个体中, DNA 分子所携带的遗传信息总和.

- genome
- 核基因组
- 叶绿体基因组
- 线粒体基因组
- 病毒基因组
- 人的基因组 (22+X+Y)

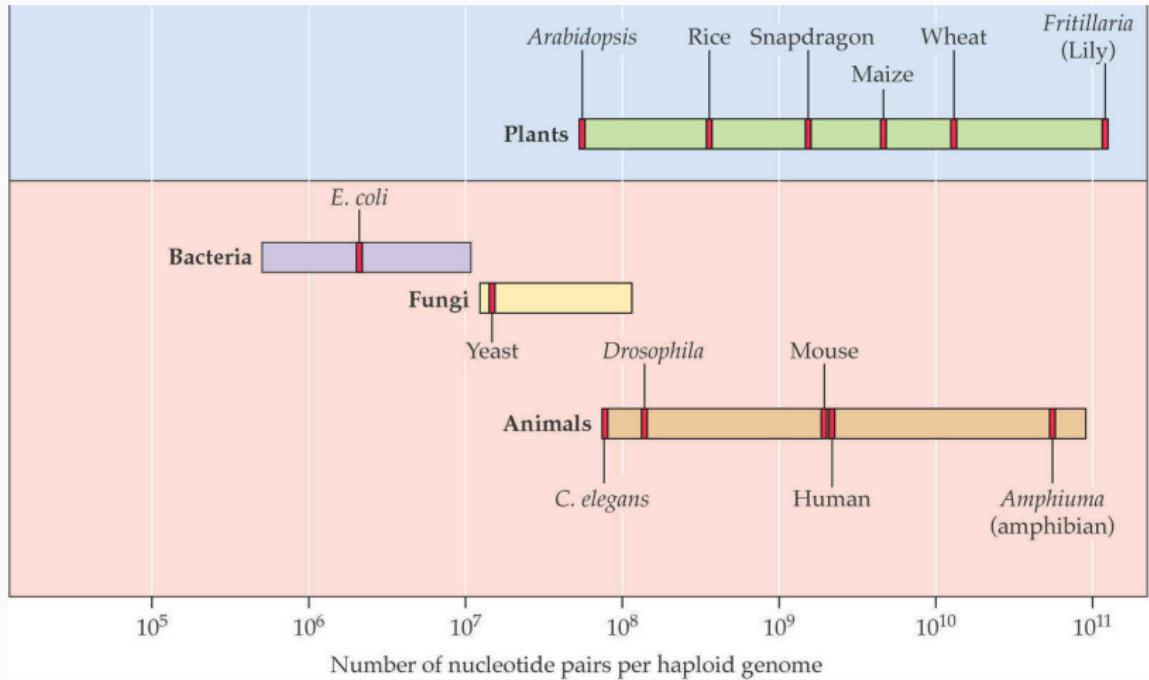


Figure 1. 不同生物基因组碱基对数目比较

基因组学 研究生物体的基因和基因组的结构, 组成和功能.

- 结构基因组学: 研究基因和基因组的结构, 各种元件的序列特征, 基因定位和基因组作图等.
- 功能基因组学: 在基因组水平上阐明 DNA 序列的功能, 着重研究不同的序列结构所具有的不同功能, 基因的表达与调控, 基因和环境之间的相互作用等.
- 多种生物和病毒的基因组全序列测定.

24.1.2 人类基因组计划

- 1985 年, 美国能源部正式提出人类基因组测序.
 - 1990 年, 正式启动人类基因组测序.
1. 绘制人类基因组连锁图
 2. 绘制物理图
 3. 人类基因组测序
 4. 其他物种基因组分析



© #131101 www.photostogo.com

Figure 2. “绘制”

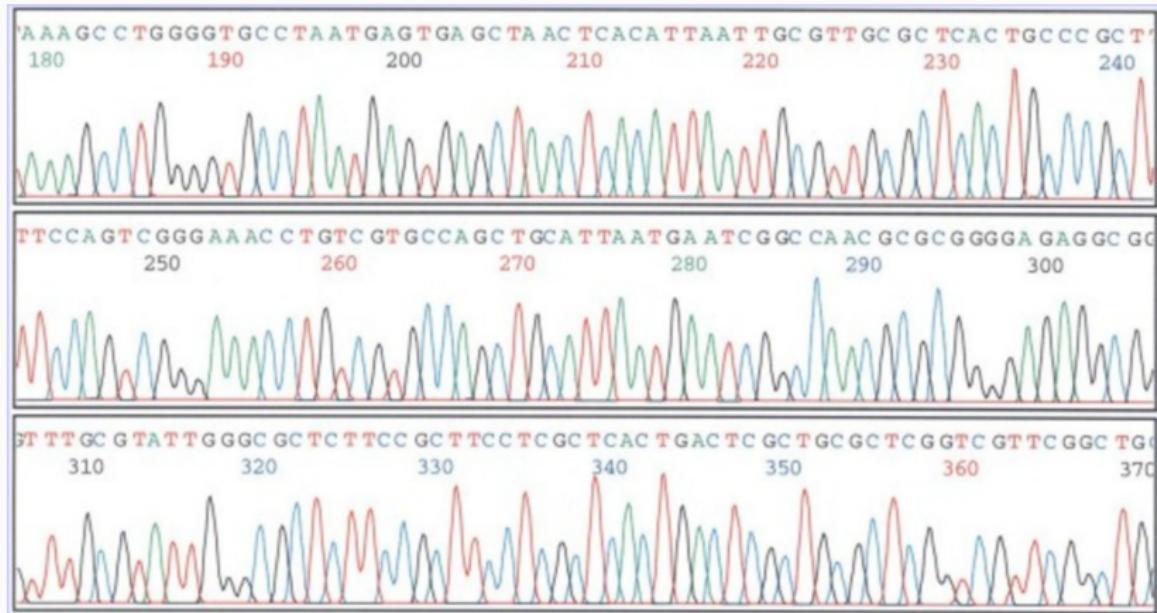


Figure 3. Electropherogram

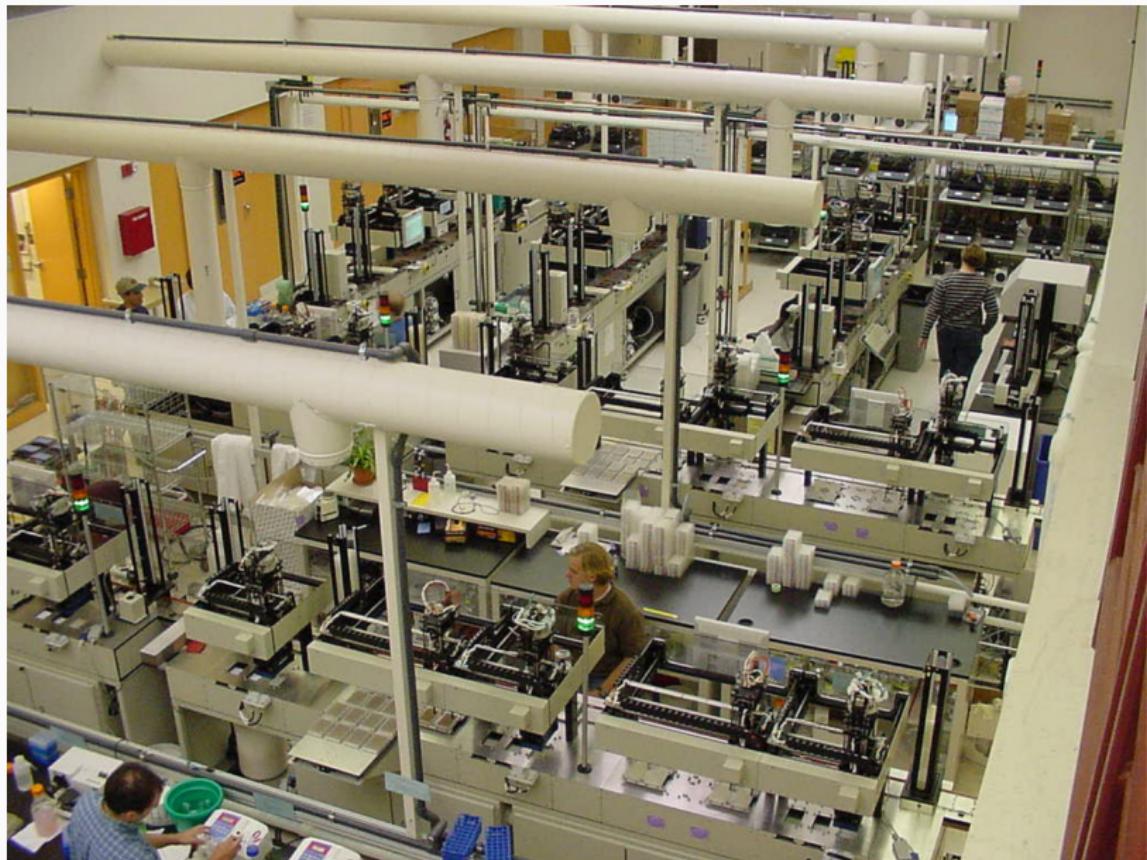


Figure 4. 模板与测序室

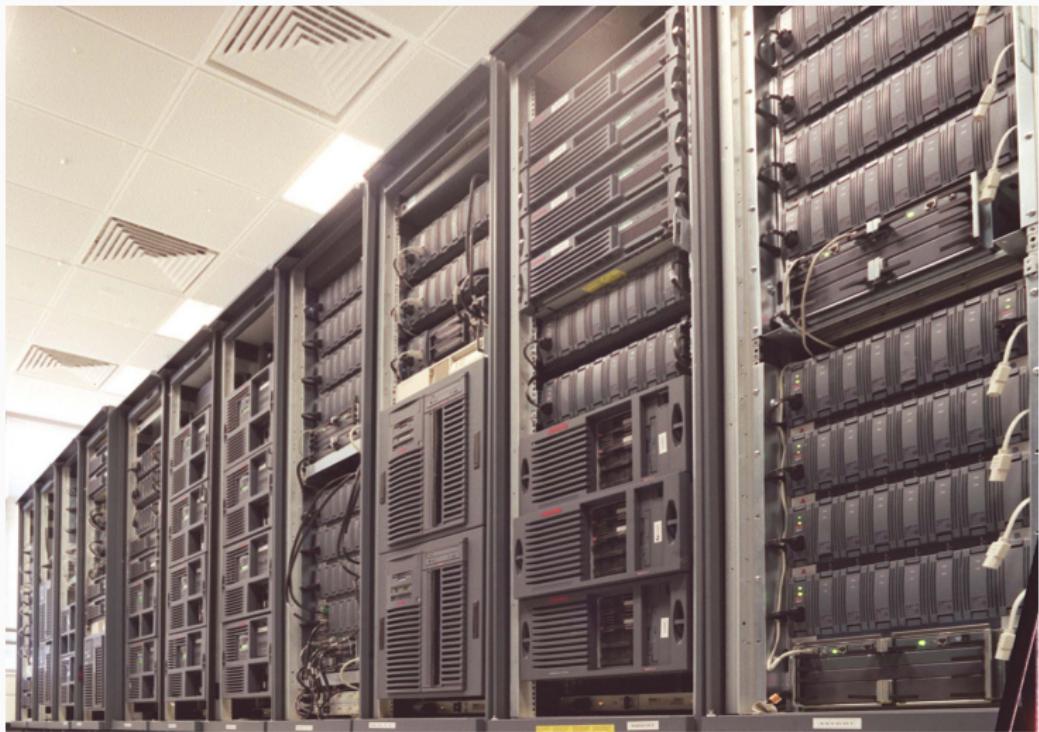


Figure 5. Clusters, Storages

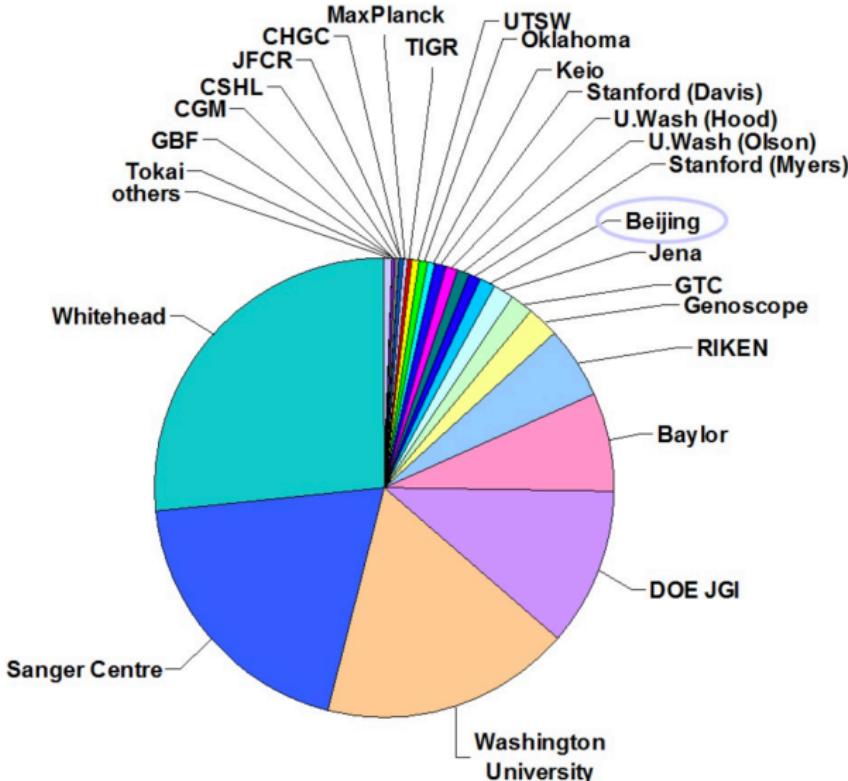


Figure 6. 贡献

基因专利

- Myriad genetics 公司拥有 *BRCA1* 和 *BRCA2* 基因的国际专利, 乳腺癌与卵巢癌的相关基因.
- Myriad 的测试费为 \$869.
- 现在有更便宜的方法, 只要 \$100, 但都侵犯了 Myriad 的专利.

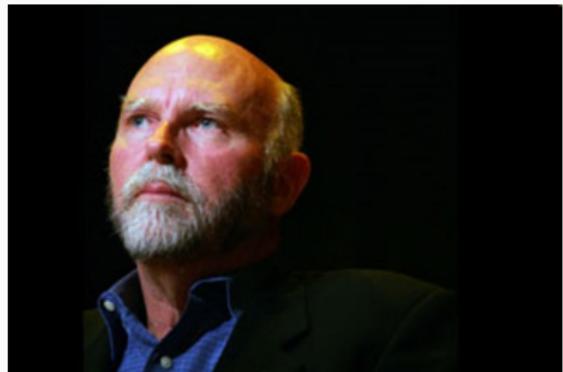


Figure 7. 文特尔, J. Craig Venter

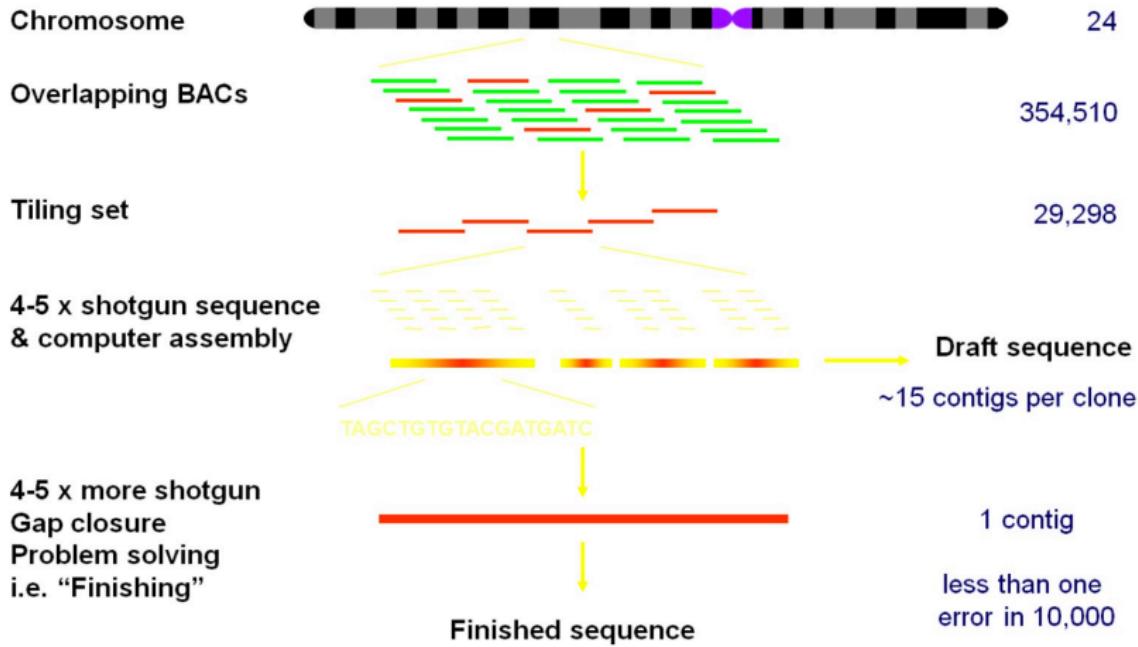
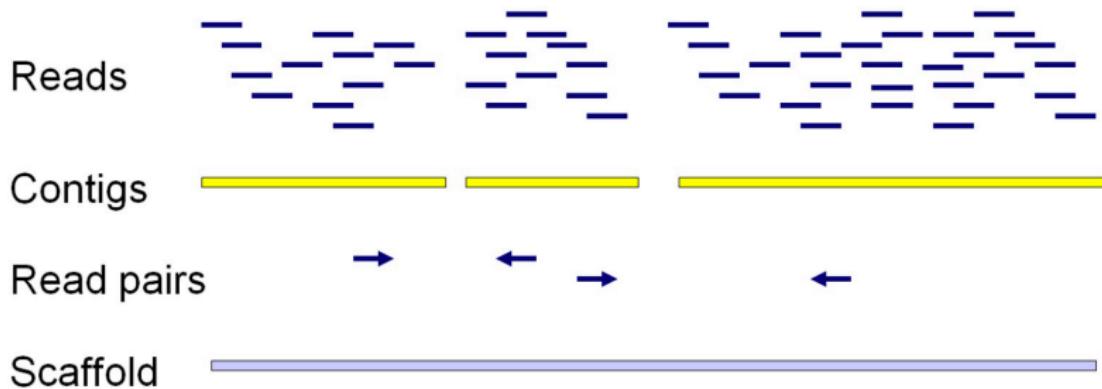


Figure 8. HGP sequencing strategy



Then order scaffolds on the chromosomes
using the HGP clone map and other
publicly available maps

Figure 9. Celera assembly strategy

Celera Corporation (CRA)

Apr 23: 7.50 ↑ 0.25 (3.45%)

Enter Symbol

GET CHART

COMPARE

EVENTS ▾

TECHNICAL INDICATORS ▾

CHART SETTINGS ▾

RESET

Week of Mar 15, 2010 : ■ CRA 6.92



© [type Function] Yahoo! Inc.

2000

2002

2004

2006

2008

2010

1D 5D 1M 3M YTD 6M 1Y 2Y 5Y Max

FROM: Apr 28 1999 TO: Apr 19 2010



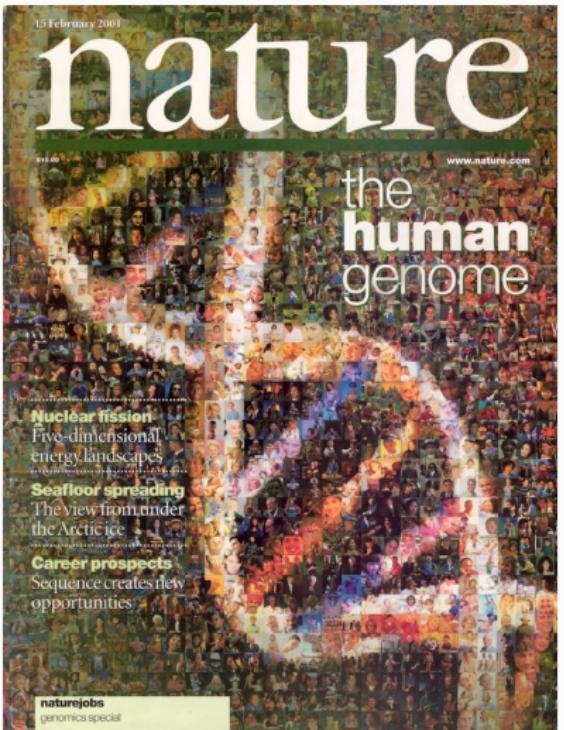
| Print | Share | Send Feedback

Disclaimer. - Quotes delayed at least 15 minutes.

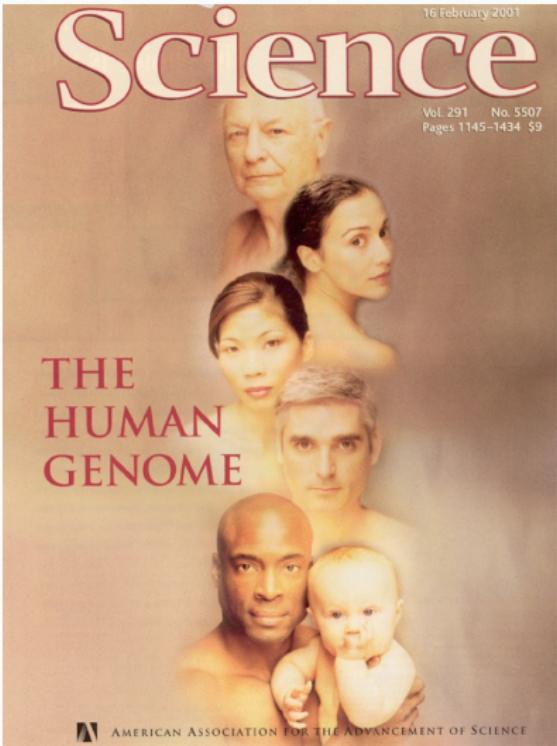
Figure 10. Celera stock

基因组草图, 2001

- International Human Genome Sequencing Consortium
 - ▶ 公共项目
 - ▶ Lander, E. S. et al. Initial Sequencing and Analysis of the Human Genome. *Nature* **409**, 860–921 (2001)
- Celera Genomics, Venter JC et al.
 - ▶ 私有项目
 - ▶ Venter, J. C. et al. The Sequence of the Human Genome. *Science* **291**, 1304–1351 (2001)



(a) Nature



(b) Science

Figure 11. 基因组草图, 2001

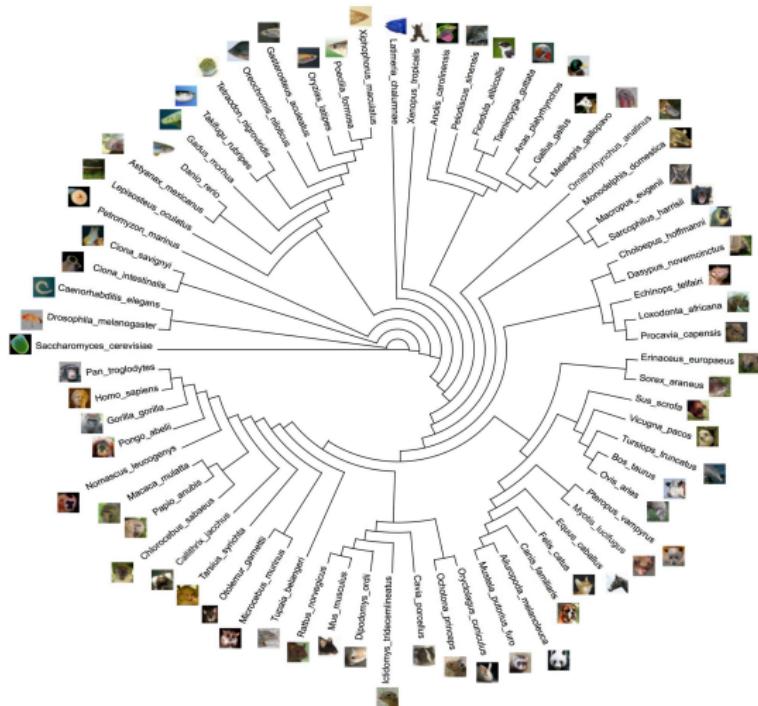


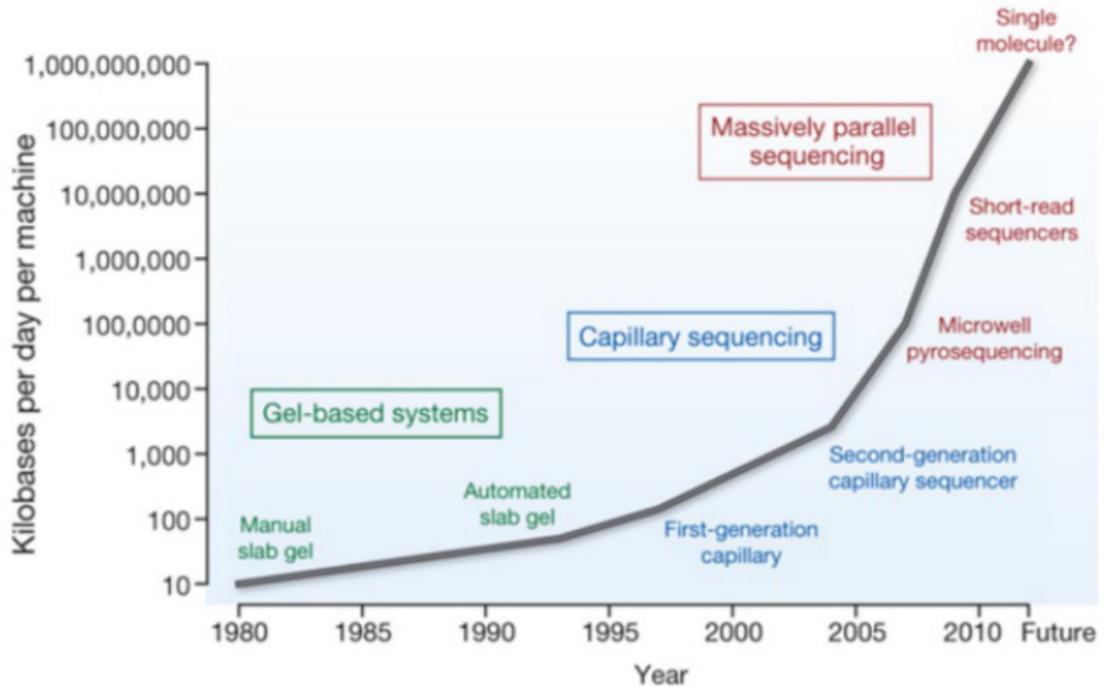
Figure 12. 全基因组测序的生物

24.1.3 基因组研究成果

- 3.2×10^9 bp
- 约 2 万个基因

测序已完成; 进一步的工作是确定各个基因的功能及他们之间的相互关系.

“Informatics is to biology what mathematics is to physics.”



MR Stratton *et al.* *Nature* 458, 719-724 (2009)

Figure 13. 测序能力的增长



NEW HiSeq 2500

Remarkable speed
and flexibility.

MiSeq

Simplicity, integration,
and ease-of-use.

Illumina announces
speed and
performance
enhancements.

Introducing the HiSeq 2500 and Triple the
Output on MiSeq.

 LEARN MORE

Figure 14. Illumina

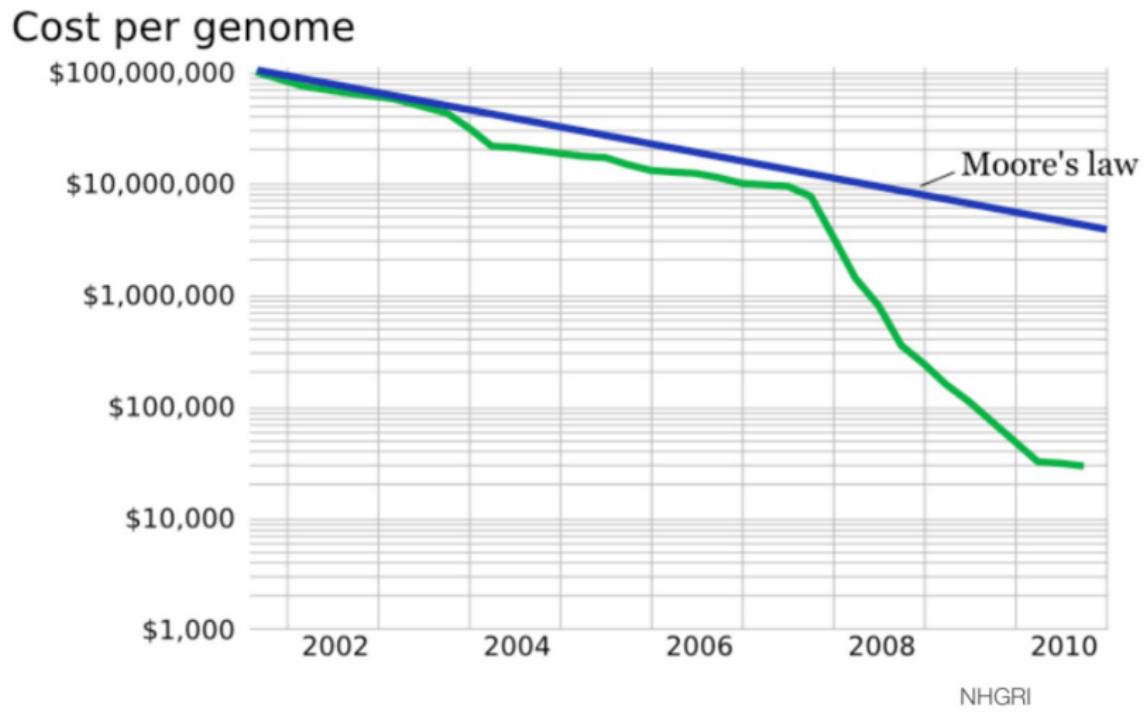


Figure 15. 比摩尔定律更快

1000 Genomes

A Deep Catalog of Human Genetic Variation

Home About Participants Data Contact Wiki

1000 GENOMES PROJECT DATA RELEASE

SNP data downloads and genome browser representing four high coverage individuals

The first set of SNP calls representing the preliminary analysis of four genome sequences are now available to download through the [EBI FTP site](#) and the [NCBI FTP site](#). The README file dealing with the FTP structure will help you find the data you are looking for.

The data can also be viewed directly through the 1000 Genomes browser at <http://browser.1000genomes.org>. Launch the browser and [view a sample region here](#).

More information about the data release can be found in the [data section](#) of this web site.

Download the 1000 Genomes Browser Quick Start Guide

[Quick start \(pdf\)](#)

PRESS RELEASE

WEDNESDAY JUN. 11, 2008
[Three Sequencing Companies Join 1000 Genomes Project](#)

TUESDAY JAN. 22, 2008
[International Consortium Announces the 1000 Genomes Project](#)

LOG IN

Username:

Password:

([I forgot my password](#))

LINKS

 [Download the meeting report](#)

 [View the participants](#)

© 1000 Genomes 2008

Figure 16. 千个基因组计划

Personal Genomes

- Craig Venter
- James Watson
- Stephen Quake
- George Church
- Marjolein Kriek
- Hermann Hauser
- Han Chinese
- Seong-Jin Kim
- Korean AK1
- Yoruban African
NA18507
- 14 others sequenced by Complete Genomics
- Unknown number sequenced by Knome
- 6 genomes sequenced at high depth by the 1000 Genomes Project
- 180 genomes sequenced at low coverage by the 1000 Genomes Project
- Two acute myeloid leukemia patients



(a) 脊椎动物



(b) 英国人健康计划

1001 Genomes
A Catalog of *Arabidopsis thaliana* Genetic Variation



(c) 拟南芥

Figure 17. 其它计划

What's in the NCBI FTP site?

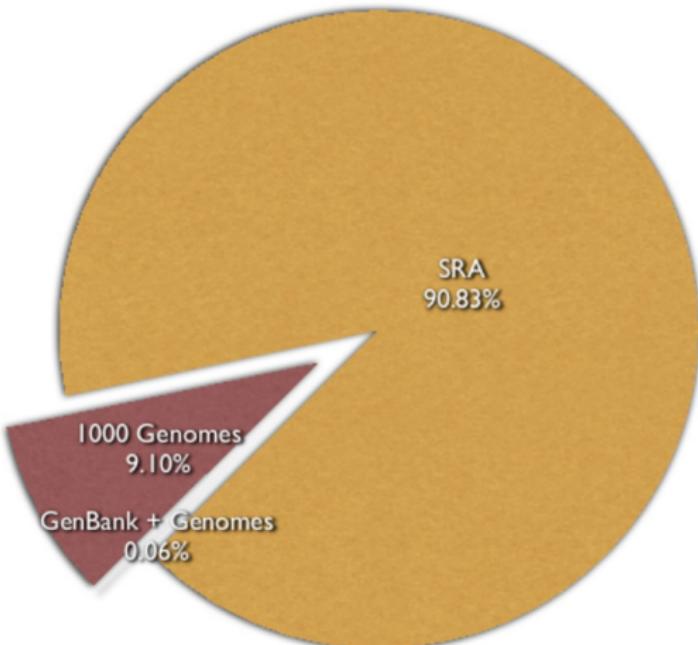


Figure 18. 爆发性增长的数据量



Figure 19. 国际肿瘤基因组计划

ORIGINAL ARTICLE

Whole-Genome Sequencing in a Patient with Charcot–Marie–Tooth Neuropathy

James R. Lupski, M.D., Ph.D., Jeffrey G. Reid, Ph.D., Claudia Gonzaga-Jauregui, B.S.,
David Rio Deiros, B.S., David C.Y. Chen, M.Sc., Lynne Nazareth, Ph.D.,
Matthew Bainbridge, M.Sc., Huyen Dinh, B.S., Chyn Jing, M.Sc.,
David A. Wheeler, Ph.D., Amy L. McGuire, J.D., Ph.D., Feng Zhang, Ph.D.,
Pawel Stankiewicz, M.D., Ph.D., John J. Halperin, M.D., Chengyong Yang, Ph.D.,
Curtis Gehman, Ph.D., Danwei Guo, M.Sc., Rola K. Irikat, B.S., Warren Tom, B.S.,
Nick J. Fantin, B.S., Donna M. Muzny, M.Sc., and Richard A. Gibbs, Ph.D.

Figure 20. 追踪罕见遗传疾病的致病基因¹

¹ Lupski, J. R. et al. Whole-Genome Sequencing in a Patient with Charcot - Marie - Tooth Neuropathy. *N Engl J Med* 362, 1181–1191 (2010)

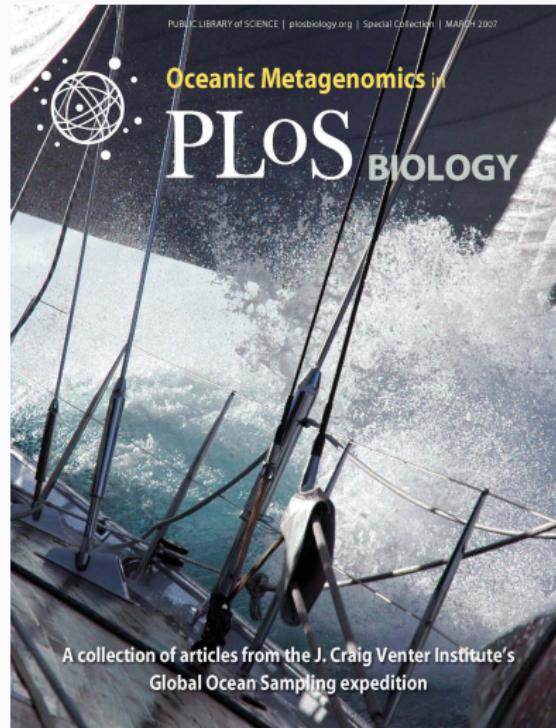


Figure 21. 元基因组 (metagenome)

24.1.4 人类基因组各组成成分的基本特征

1. 基因

- ▶ 断裂基因

2. 基因外 DNA

- ▶ 单拷贝序列和低拷贝数序列占 70–80%
- ▶ 中度和高度重复序列占 20–30%

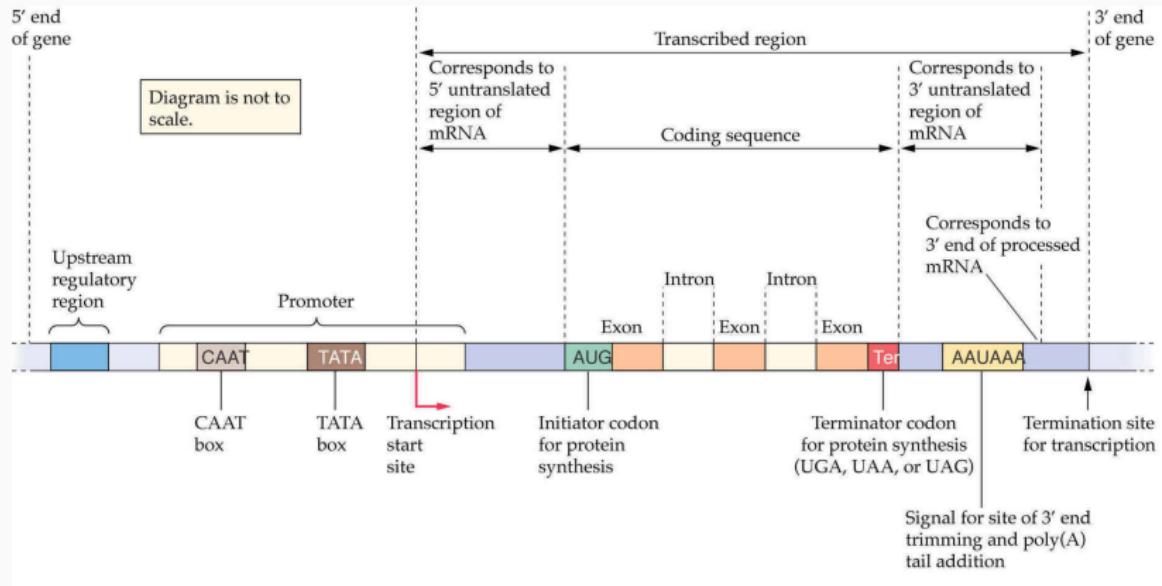


Figure 22. 真核生物基因结构

3. 分散重复序列

- ▶ 中度重复序列, 以散在的方式分布于基因组中.
- ▶ 短分散重复序列.
- ▶ 长分散重复序列: 反转录转座子.

4. 簇状重复序列

- ▶ 卫星DNA, 小卫星DNA 和微卫星DNA.
- ▶ 利用氯化铯密度梯度离心法可以把卫星DNA与其它DNA分开.
- ▶ G+C比例高, 浮力密度大; A+T比例高, 则浮力密度小.
主带, 次带(卫星DNA).

24.2 人类遗传性疾病

24.2.1 染色体病

Polytene chromosomes have many bands

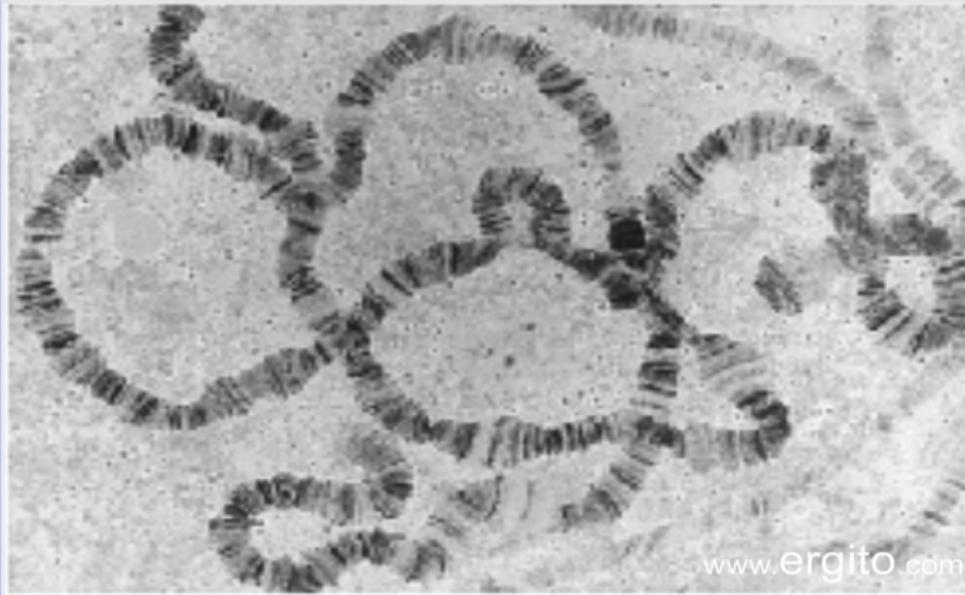


Figure 23. 果蝇的多线染色体

The X chromosome has many G-bands

©virtualtext www.ergito.com

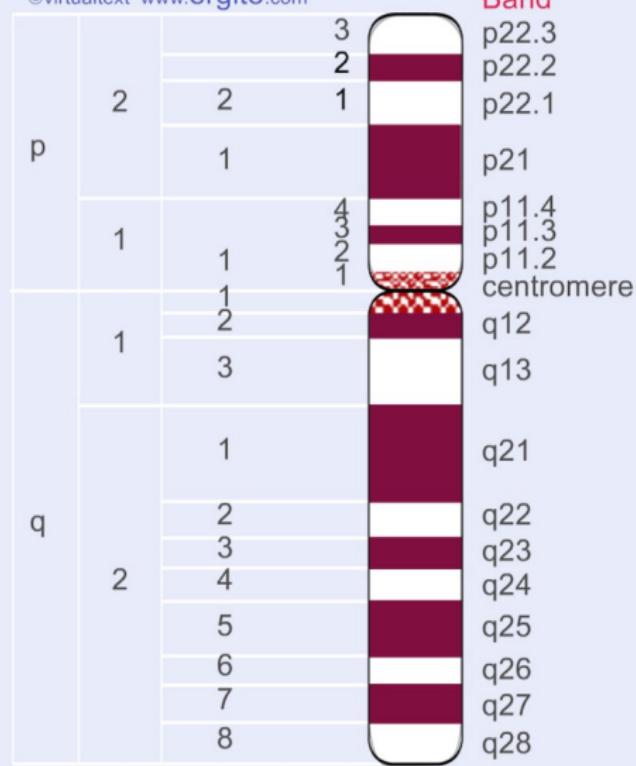
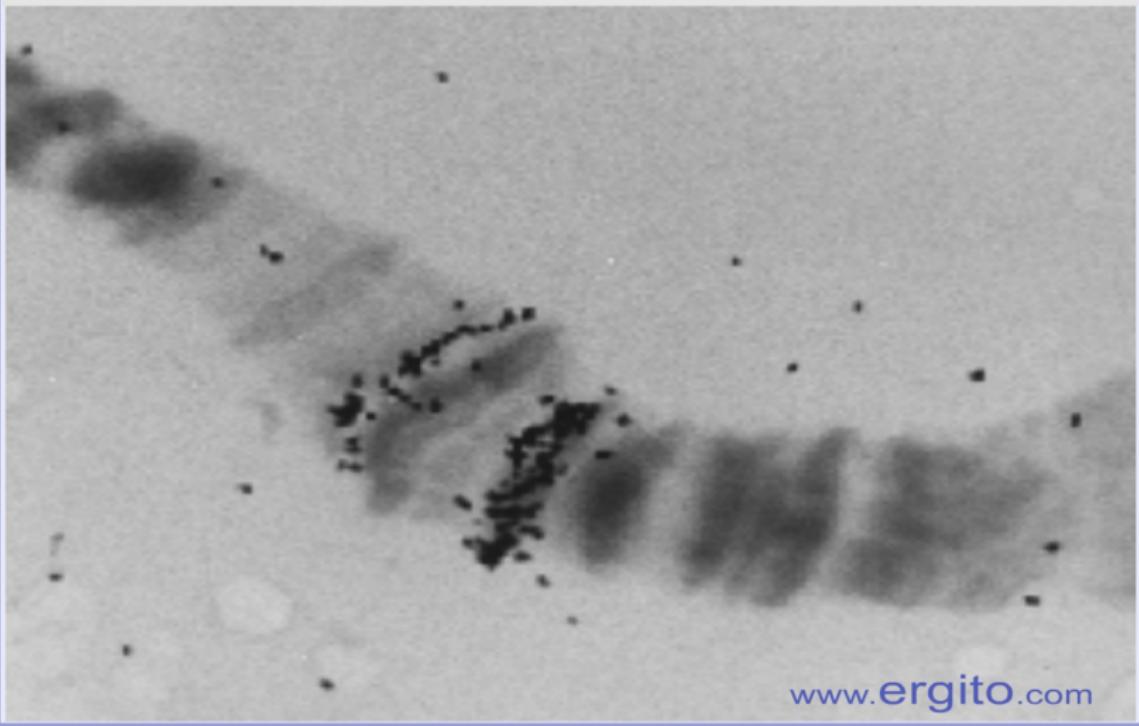


Figure 24. 人类 X 染色体

A band is identified by *in situ* hybridization



www.ergito.com

Figure 25. 原位杂交鉴定染色体带

1. 染色体的结构变异

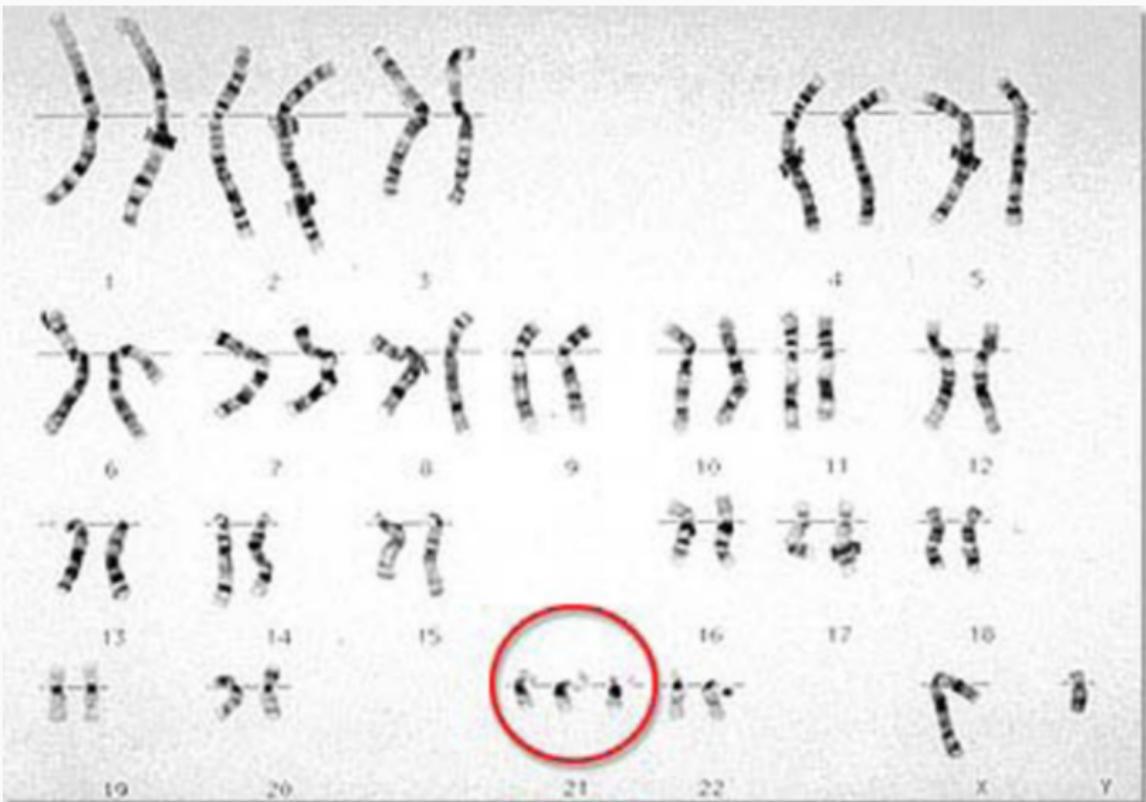
- ▶ 缺失
- ▶ 重复
- ▶ 倒位
- ▶ 易位



Figure 26. 5号染色体短臂缺失

2. 染色体数目变异

- ▶ 整倍体
 - 三倍体无籽西瓜
- ▶ 非整倍体
 - Klinefelter 综合症 (XXY)
 - 唐氏综合症 (21 三体)



medgen.genetics.utah.edu

Figure 27. 21 染色体三体

24.2.2 单基因病

1. 隐性遗传病

- ▶ 囊状纤维化病
- ▶ 近亲结婚的后代易得

2. 显性遗传病

- ▶ 舞蹈病

3. X 连锁遗传病

- ▶ 色盲
- ▶ 血友病

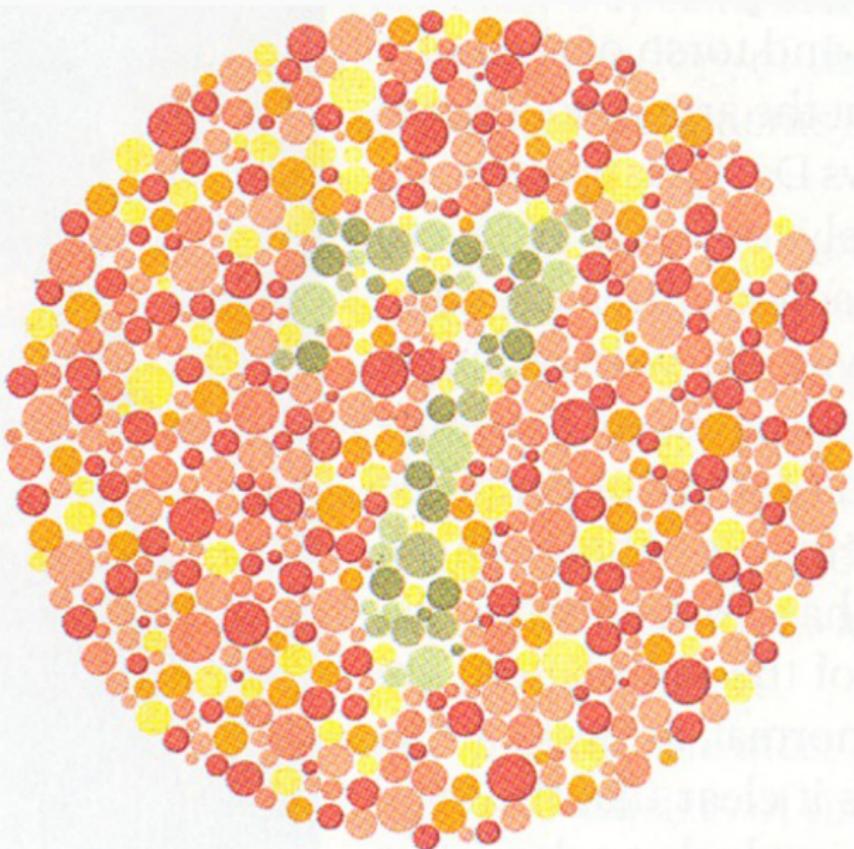


Figure 28. 色盲症辨别

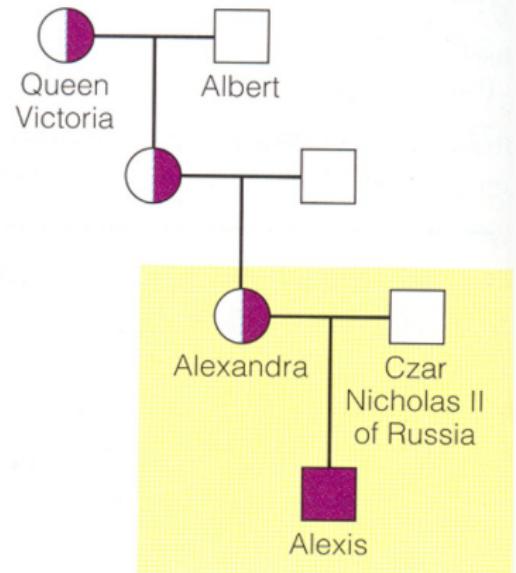


Figure 29. 血友病遗传谱系

24.2.3 多基因遗传病

- 哮喘, 原发性高血压, 糖尿病.
- 微效基因的作用累加
 - ▶ 遗传因素和环境因素共同作用
 - ▶ 有家族聚集倾向

24.3 癌基因与恶性肿瘤

肿瘤

正常的细胞增殖和凋亡失控, 扩张性增生的细胞群形成肿块.

恶性肿瘤 (癌症).

实际上是一类遗传病.

24.3.1 原癌基因和抑癌基因

1. 病毒癌基因和原癌基因

- ▶ 1910年, Peyton Rous 鸡肉瘤
- ▶ 1970年, Martin 等确定反转录病毒病毒基因与细胞癌变有关
- ▶ 1976年, 分离出该基因 *src*

2. 抑癌基因

- ▶ 使细胞不癌变或机体不长癌

24.3.2 癌症的遗传学基础

1. 基因突变
2. 多次遗传改变的致癌作用
 - ▶ 结肠癌: 内璧细胞非正常分裂 → 良性肿瘤 → 癌

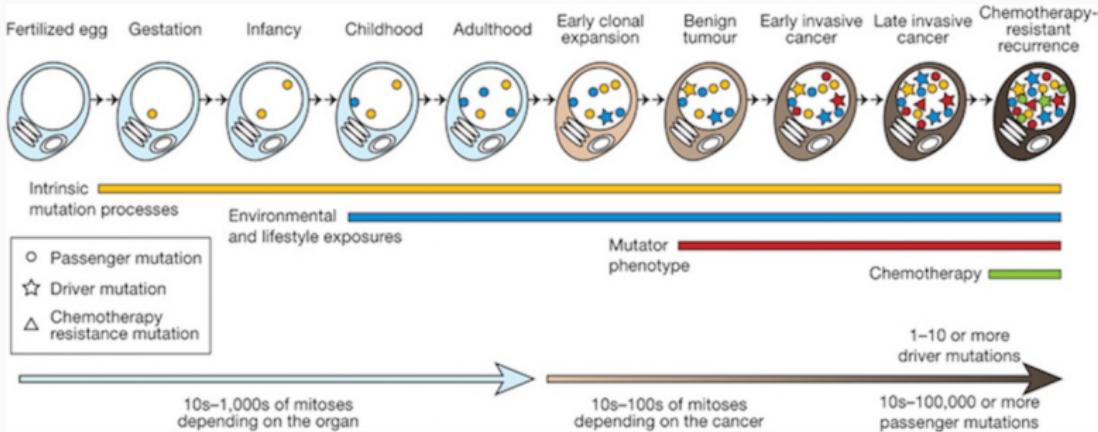


Figure 30. 体细胞突变的积累

24.3.3 改变生活方式能降低癌症的危险性

能够改变 DNA 的结构, 引起 DNA 损伤的物质, 可以使细胞癌变.

在生活中应该避免接触或摄入这些物质.