



2010澳大利亚
青岛（崂山）

主办单位：山东省人民政府

于伟嘉

姚晓华

Bioinformatics Today

生物信息学现状

刘广斌

The University of Southern Queensland
Australia

这里想要讨论的问题太复杂, 太庞大, 而有关文献又浩如烟海, 难以把握。今日所谈难免会使人感到不尽如意或有所偏狭。唯愿今日所讲能给大家提供一点有用的信息。

生物信息学可以简单的描述为生物数学，
生物统计学和计算机科学技术在分子生物学领域的综合利用。

从上个世纪80年代后期开始的分子生物学与信息技术的飞速发展促进了基因组学和基因学的迅速发展,产生出巨量的分子生物学数据和信息！

虽然产生数据是科学研究的一项基本任务，
然而若无高效准确地分析，这些巨量数据
只能是废物。

生物信息学

浩如烟海的资料使人眼花缭乱，甚至莫衷
一是。

于是 应运而生！

如何使用，理解，分析，解读这些信息？

生物信息学历史：
1951年- Pauling
and Corey 提出了
DNA结构的初步设想
(PNAS 1951).

1953年, Watson
and Crick 解释了
DNA的双螺旋结构
(Nature, 171:
737-738, 1953) 。



2003年是生物学研究领域的一个大年：

- 1 DNA双螺旋结构发现50周年。
- 2 人类基因组计划于2003年完成（1986 年启动，历时17年）。从2003年4月12日起，31亿个碱基对的人类DNA序列全部对公众开放。



此后的一系列成就见证着基因组学（Genomics）和蛋白质组学（Proteomics）日新月异的发展。

NCBI genome statistics (17/8/2010)

The Map Viewer provides a wide variety of genome mapping and sequencing data. [More..](#)



▼ Vertebrates	(17)
▼ Mammals	(14)
▶ Primates	(3)
▶ Rodents	(2)
▶ Monotremes	(1)
▶ Marsupials	(1)
▶ Other Mammals	(7)
▶ Other Vertebrates	(3)
▼ Invertebrates	(13)
▶ Insects (B)	(7)
▶ Nematode (B) 线虫	(2)
▶ Echinoderms (B) 棘皮类动物	(1)
▶ Cnidaria 腔肠动物	(1)
▶ Other Invertebrates	(2)
▶ Protozoa (B) 原生动物	(18)
▼ Plants (Q)	(45)
▶ Flowering Plants (B)	(40)
▶ Algae 水藻	(4)
▶ Moss 苔藓	(1)
▼ Fungi (Q) (B)	(17)

Scientific name

Common name

Build

Tools

NCBI genome statistics (16/9/2010)

▼ Vertebrates		(18)
▼ Mammals		(14)
▶ Primates		(3)
▶ Rodents		(2)
▶ Monotremes		(1)
▶ Marsupials		(1)
▶ Other Mammals		(7)
▶ Other Vertebrates		(4)
▼ Invertebrates		(13)
▶ Insects	B	(7)
▶ Nematode	B 线虫	(2)
▶ Echinoderms	B 棘皮类动物	(1)
▶ Cnidaria	腔肠动物	(1)
▶ Other Invertebrates		(2)
▶ Protozoa	B 原生动物	(18)
▼ Plants	Q	(46)
▶ Flowering Plants	B	(41)
▶ Algae	水藻	(4)
▶ Moss	苔藓	(1)
▶ Fungi	Q B	(17) 8

In ensembl database



51 vertebrate species (脊椎动物)



Aedes aegypti

[VectorBase](#) | *Aedes aegypti*



Anopheles gambiae

[VectorBase](#) | *Anopheles gambiae*



Caenorhabditis brenneri

[WormBase](#) | *Caenorhabditis brenneri* PB2801



Caenorhabditis briggsae

[WormBase](#) | *Caenorhabditis briggsae* AF16



Caenorhabditis elegans

[WormBase](#) | *Caenorhabditis elegans* N2



Caenorhabditis japonica

[WormBase](#) | *Caenorhabditis japonica* DF5081



Caenorhabditis remanei

[WormBase](#) | *Caenorhabditis remanei* PB4641



Culex quinquefasciatus

[VectorBase](#) | *Culex quinquefasciatus*



Drosophila ananassae

[FlyBase](#) | *Drosophila ananassae*



Drosophila erecta

[FlyBase](#) | *Drosophila erecta*



Drosophila grimshawi

[FlyBase](#) | *Drosophila grimshawi*



Drosophila melanogaster

[FlyBase](#) | *Drosophila melanogaster*



Drosophila mojavensis

[FlyBase](#) | *Drosophila mojavensis*



Drosophila persimilis

[FlyBase](#) | *Drosophila persimilis*



Drosophila pseudoobscura

[FlyBase](#) | *Drosophila pseudoobscura*



Drosophila sechellia

[FlyBase](#) | *Drosophila sechellia*



Drosophila simulans

[FlyBase](#) | *Drosophila simulans*



Drosophila virilis

[FlyBase](#) | *Drosophila virilis*



Drosophila willistoni

[FlyBase](#) | *Drosophila willistoni*



Drosophila yakuba

[FlyBase](#) | *Drosophila yakuba*



Ixodes scapularis

[VectorBase](#) | *Ixodes scapularis* Wikel



Pediculus humanus

[VectorBase](#) | *Pediculus humanus*



Pristionchus pacificus

[WormBase](#) | *Pristionchus pacificus*

Ensembl Fungi Species 12



Aspergillus clavatus
[CADRE](#) | *Aspergillus clavatus*



Aspergillus flavus
[CADRE](#) | *Aspergillus flavus*



Aspergillus fumigatus A1163
[CADRE](#) | *Aspergillus fumigatus* A1163 A1163



Aspergillus fumigatus Af293
[CADRE](#) | *Aspergillus fumigatus* Af293



Aspergillus nidulans
[CADRE](#) | *Aspergillus nidulans*



Aspergillus niger
[CADRE](#) | *Aspergillus niger*



Aspergillus oryzae
[CADRE](#) | *Aspergillus oryzae*



Aspergillus terreus
[CADRE](#) | *Aspergillus terreus*



Neosartorya fischeri
[CADRE](#) | *Neosartorya fischeri*



Neurospora crassa
 EMBL Nucleotide Sequence Database | *Neurospora crassa*



Saccharomyces cerevisiae
[SGD](#) | *Saccharomyces cerevisiae*



Schizosaccharomyces pombe
[GeneDB](#) [Spombe](#) | *Schizosaccharomyces pombe*

Ensembl Plants Species 8



Arabidopsis lyrata
[Gramene](#) | *Arabidopsis lyrata*



Oryza sativa
[Gramene](#) [MSU](#) | *Oryza sativa* Nipponbare (Japonica rice)



Sorghum bicolor
[Gramene](#) [JGI](#) | *Sorghum bicolor* BTX623



Arabidopsis thaliana
[Gramene](#) | *Arabidopsis thaliana*



Oryza sativa indica group
[Gramene](#) [BGI](#) | *Oryza indica* 93-11 (Indica rice)



Vitis vinifera
[Gramene](#) [Genoscope](#) | *Vitis vinifera*



Brachypodium distachyon
[IBI](#) [Gramene](#) | *Brachypodium distachyon* (L.) Beauv

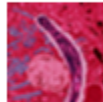


Populus trichocarpa
[Gramene](#) | *Populus trichocarpa*

Ensembl Protists Species 原生生物 6



Dictyostelium discoideum
[DictyBase](#) | *Dictyostelium discoideum*



Plasmodium falciparum
[PlasmoDB](#) | *Plasmodium falciparum*



Plasmodium vivax
[European Nucleotide Archive](#)
[UniProt Knowledgebase](#) | *Plasmodium vivax*



Phaeodactylum tricornutum
[JGI](#) | *Phaeodactylum tricornutum*

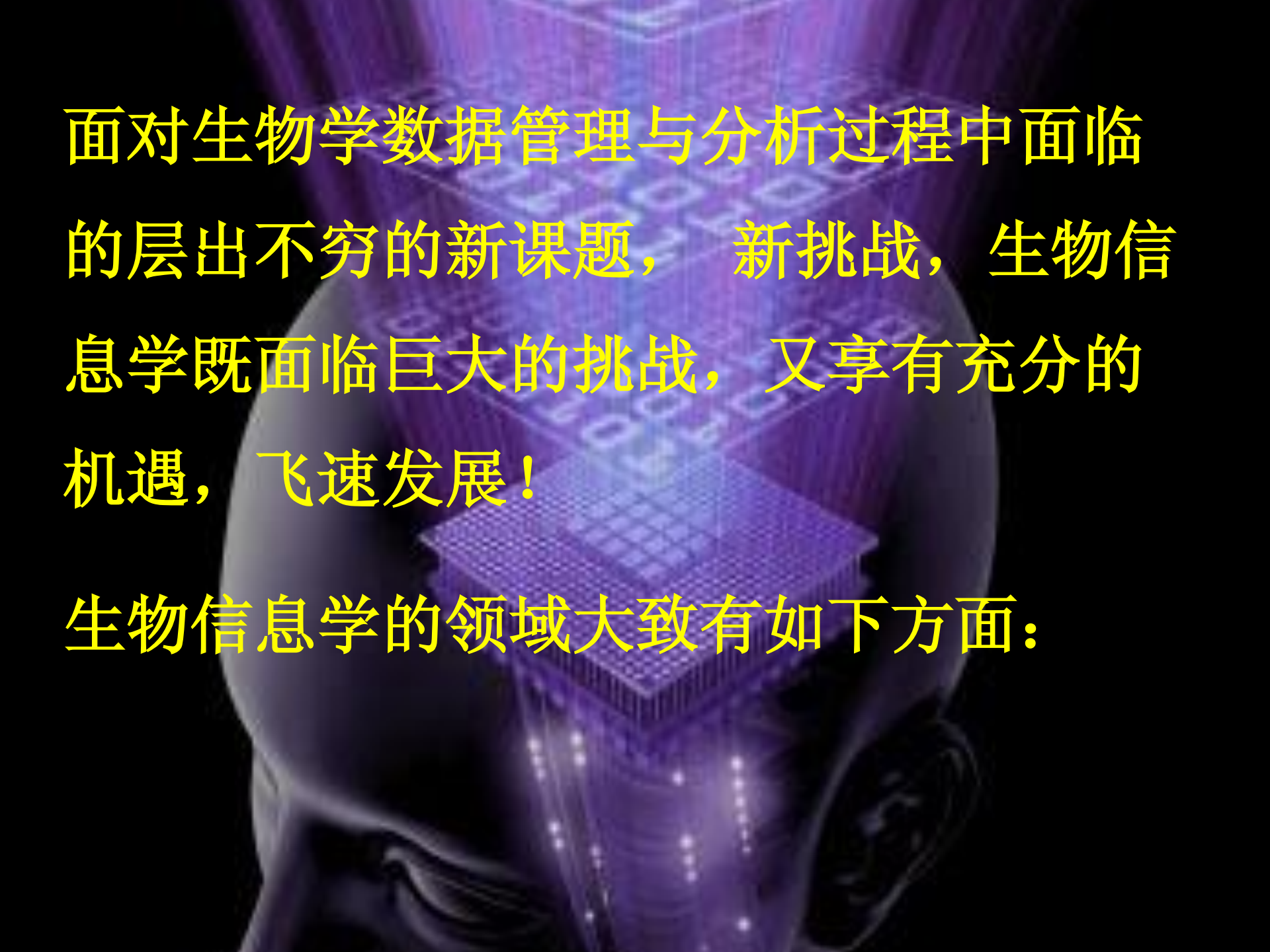


Plasmodium knowlesi
[European Nucleotide Archive](#)
[UniProt Knowledgebase](#) | *Plasmodium knowlesi*



Thalassiosira pseudonana
[JGI](#) | *Thalassiosira pseudonana*

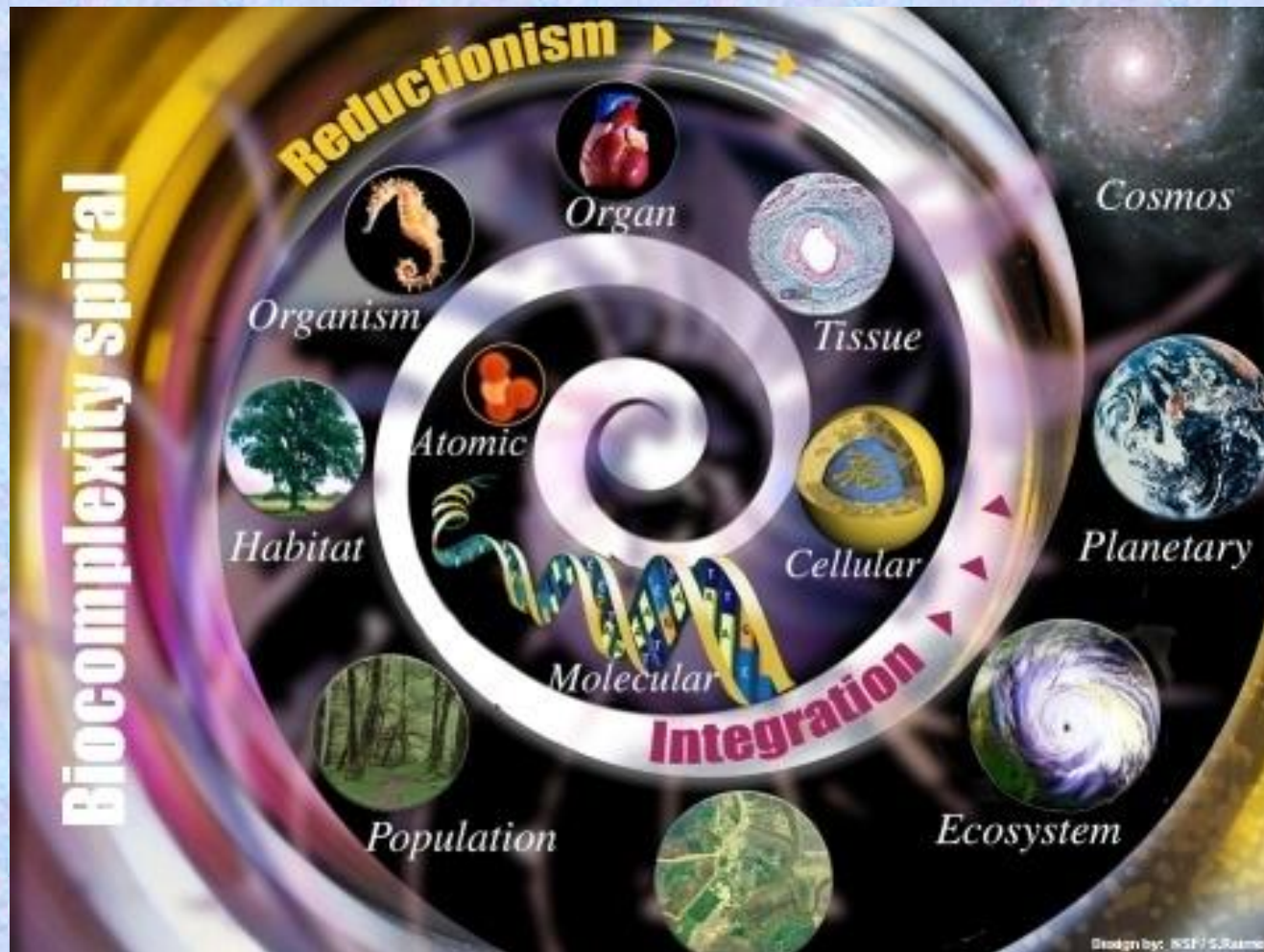
Bacteria (细菌) 183 species



面对生物学数据管理与分析过程中面临的层出不穷的新课题， 新挑战， 生物信息学既面临巨大的挑战， 又享有充分的机遇， 飞速发展！

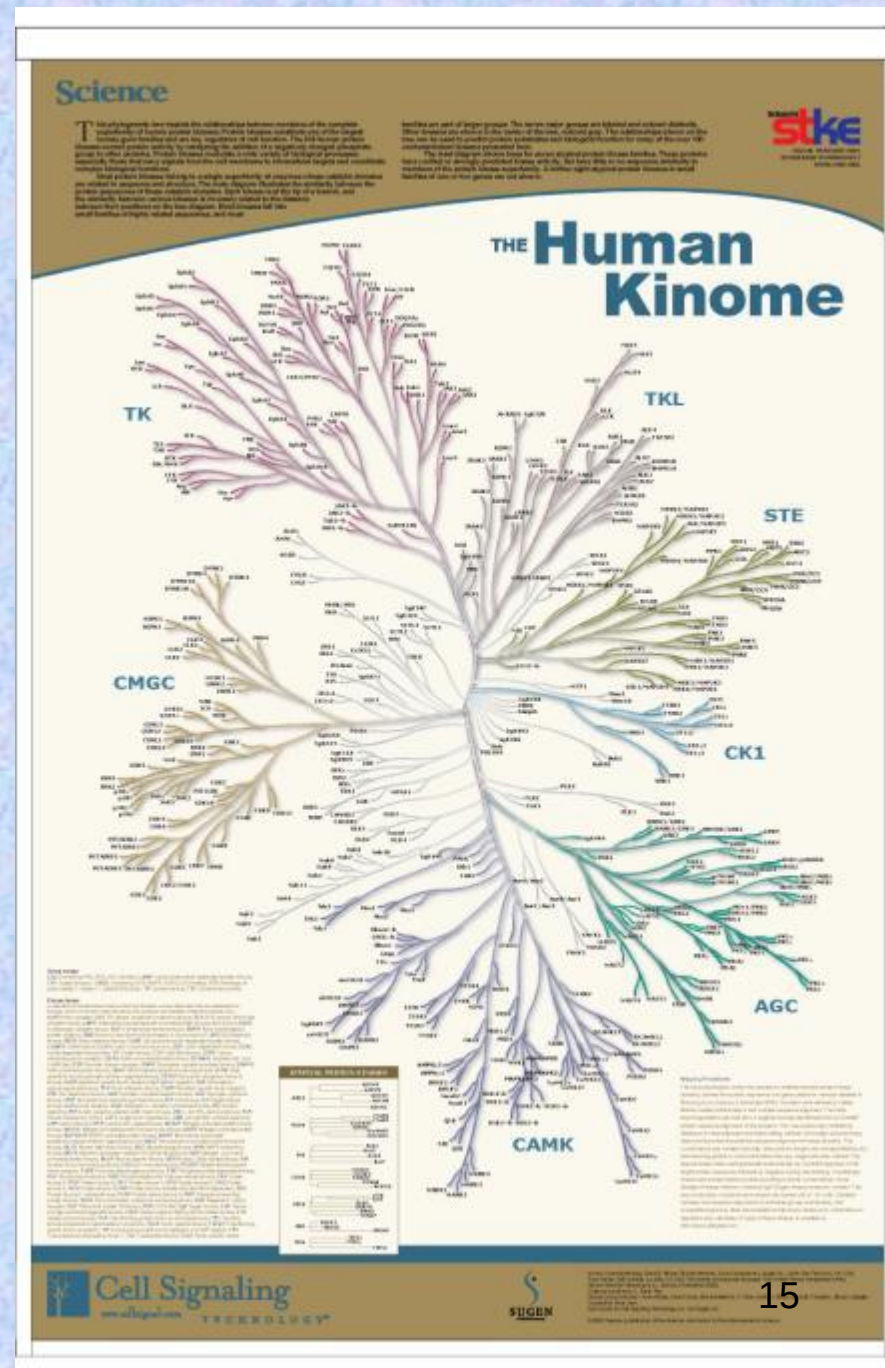
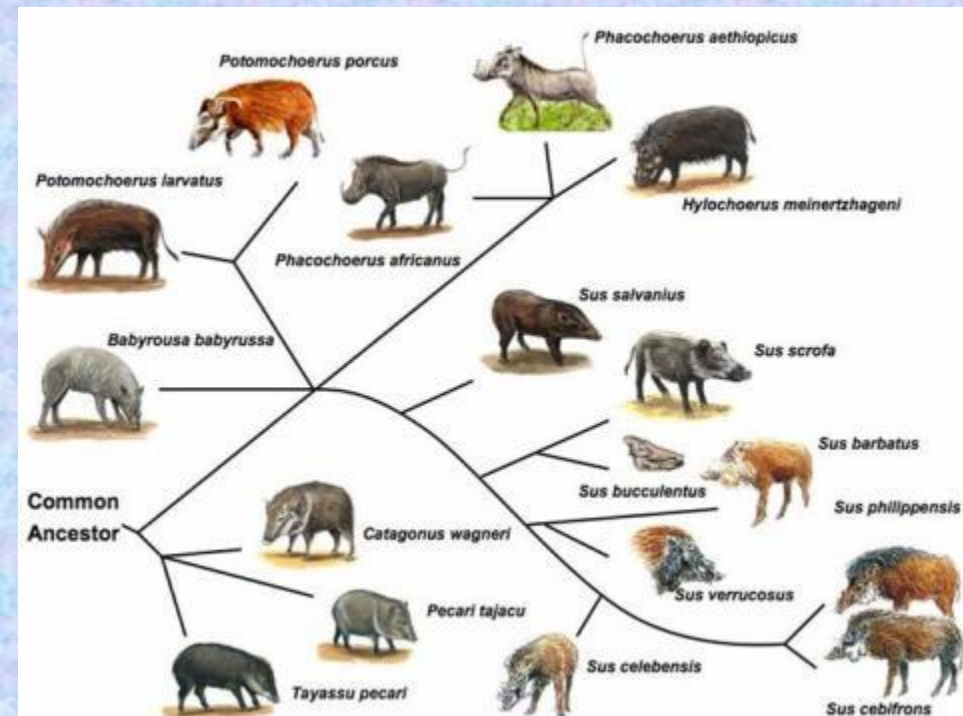
生物信息学的领域大致有如下方面：

1. 发展功能更加强大的软件和网络技术，促成不同数据库的信息交流，挖掘和整合。



2. 创新统计学以及相关数据处理的理论和算法。

Phylogenetic tree (系统发生树)



3。资料的分析和诠释：

如 DNA和蛋白质序列
的分析

多种序列核对比较

功能定位

结构预测

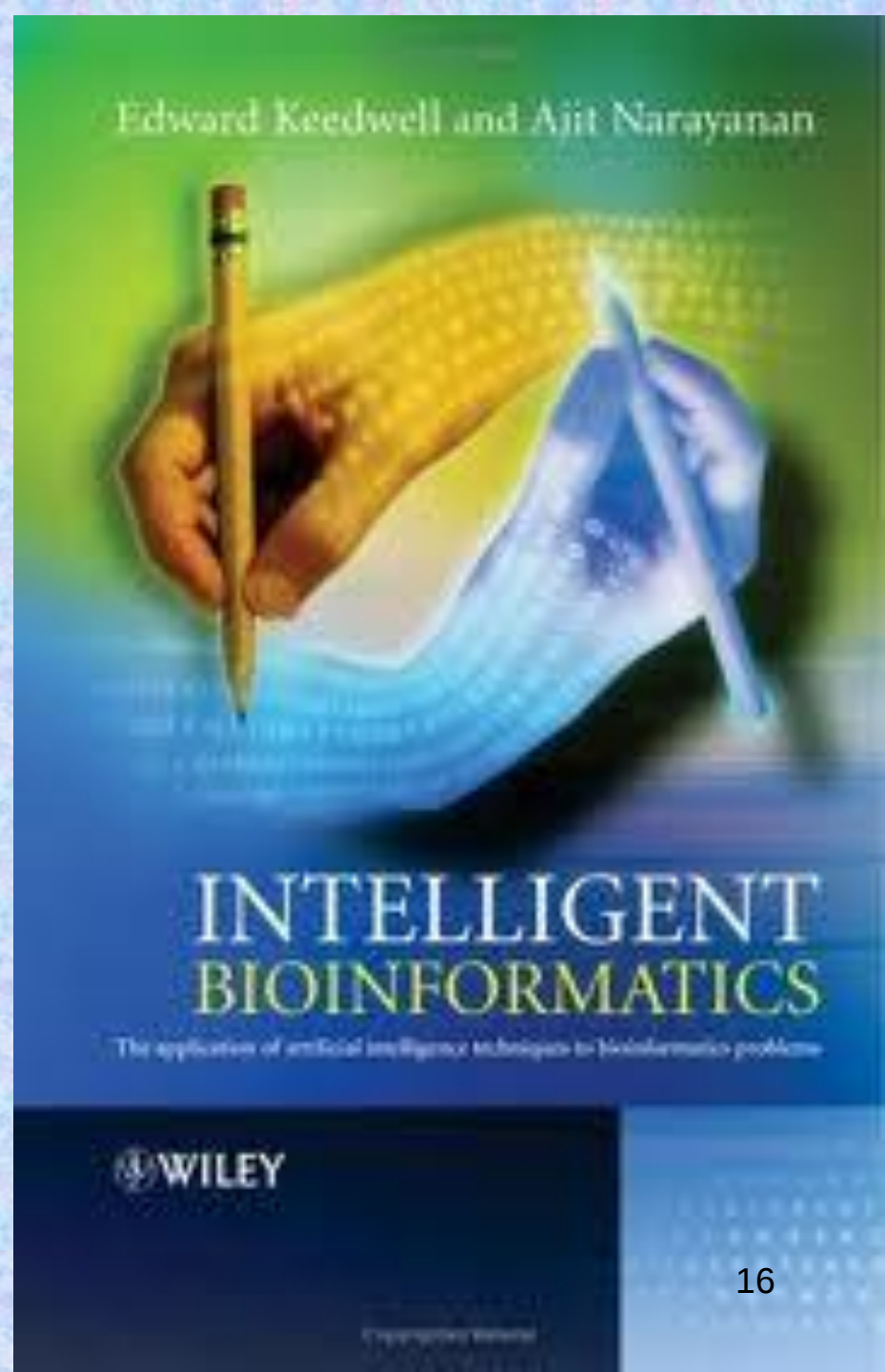
新基因发现

基因表达-微阵列

(microarray)

新药发现与设计

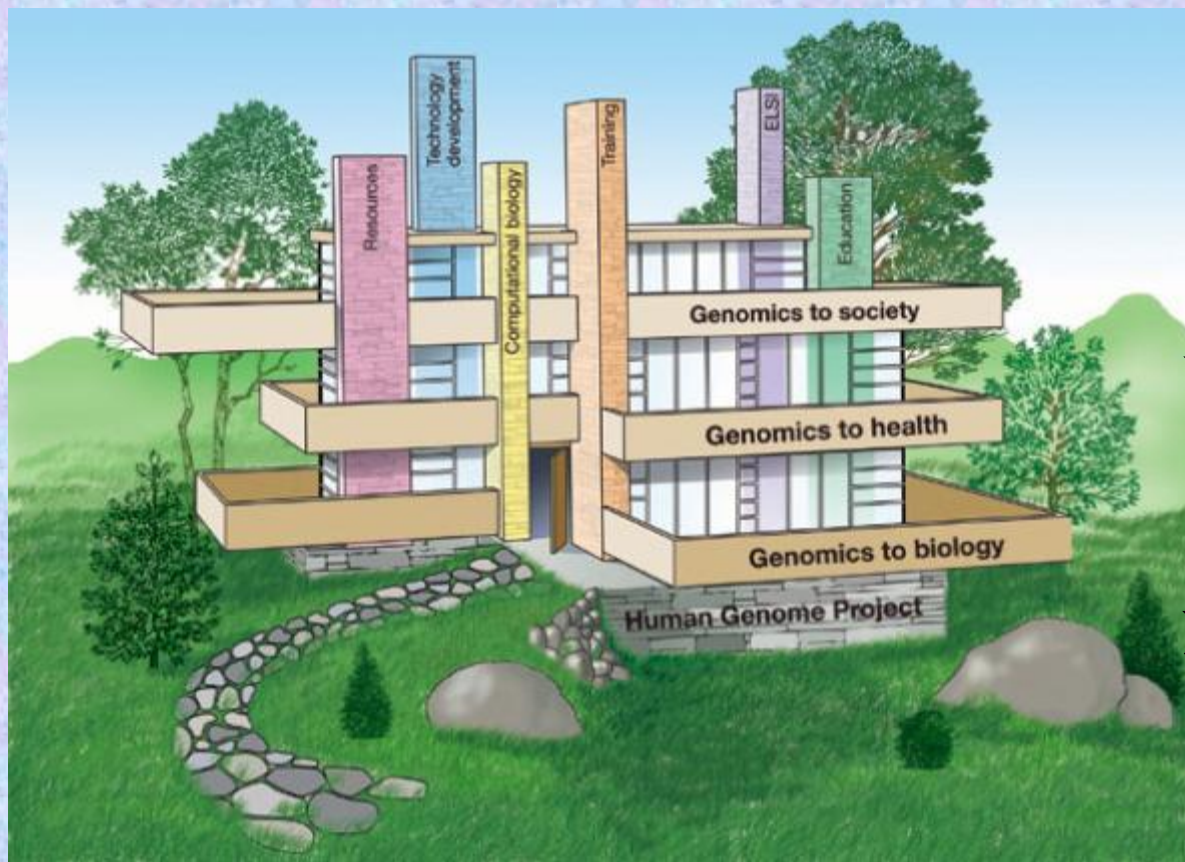
.....



生物信息学决不仅仅是计算生物学，它所涉及的技术领域十分广泛，重要的有：

数据库
互联网
图象处理
人工智能

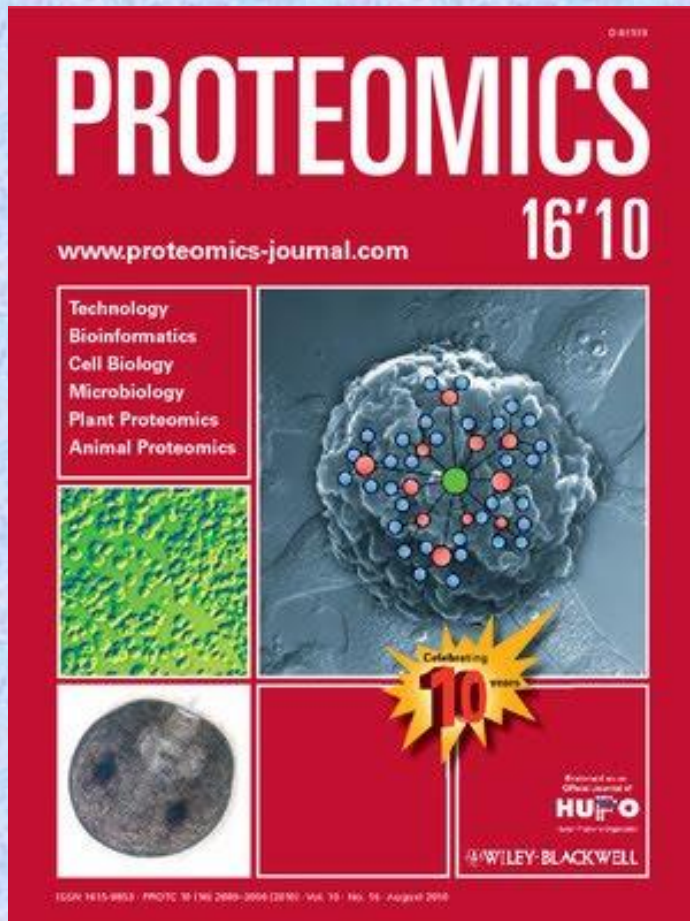
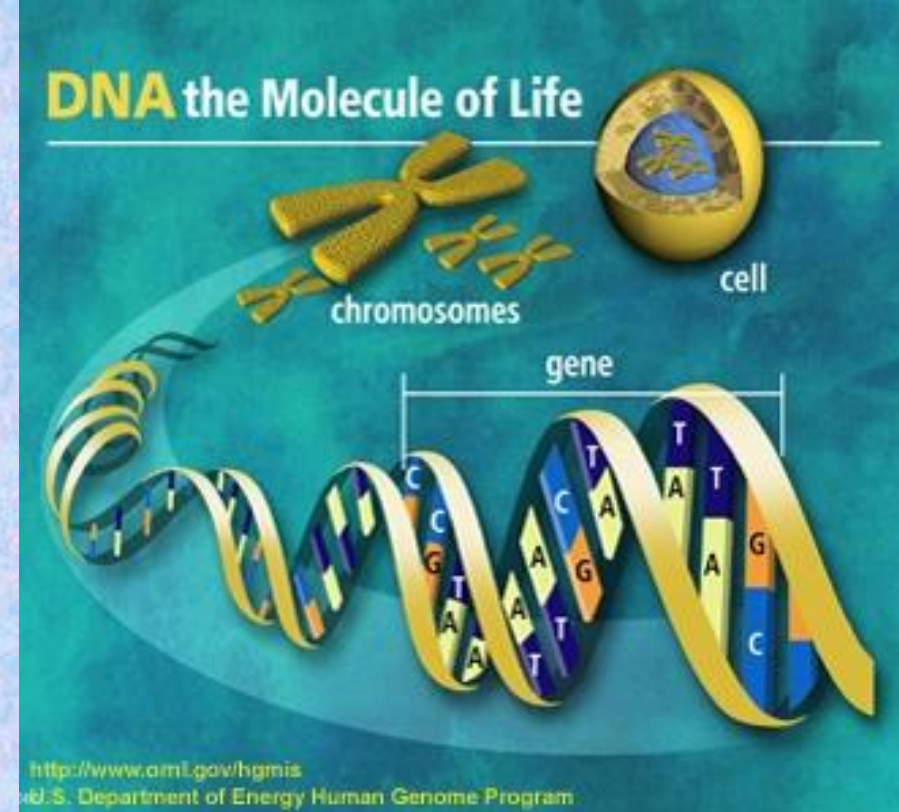
●●●



多机并
联同步
多课题
运算

人类基因组计划成为许多相关学科的基石 17

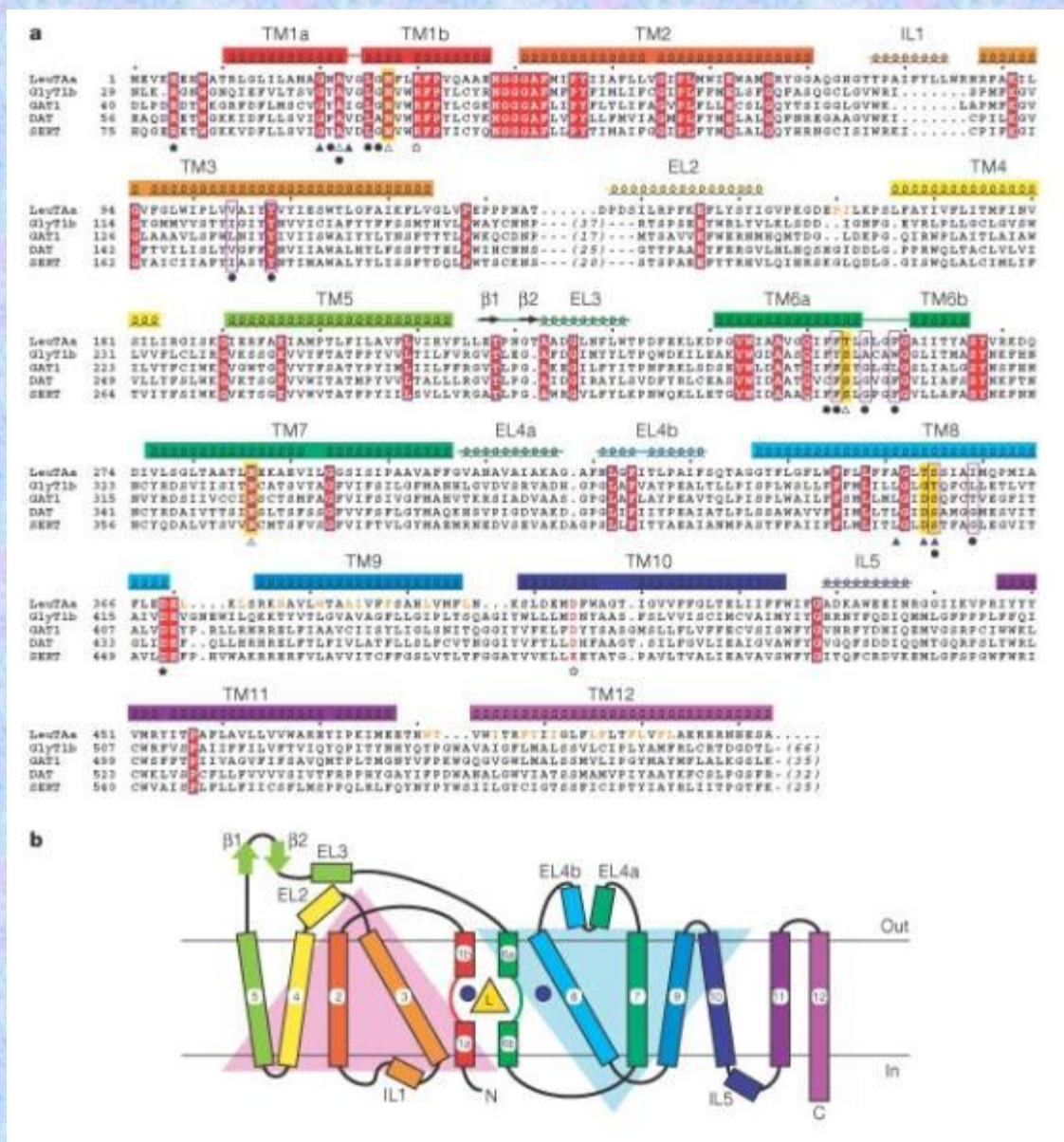
基因组学 (genomics)
研究基因在基因组内
相互作用的学科。



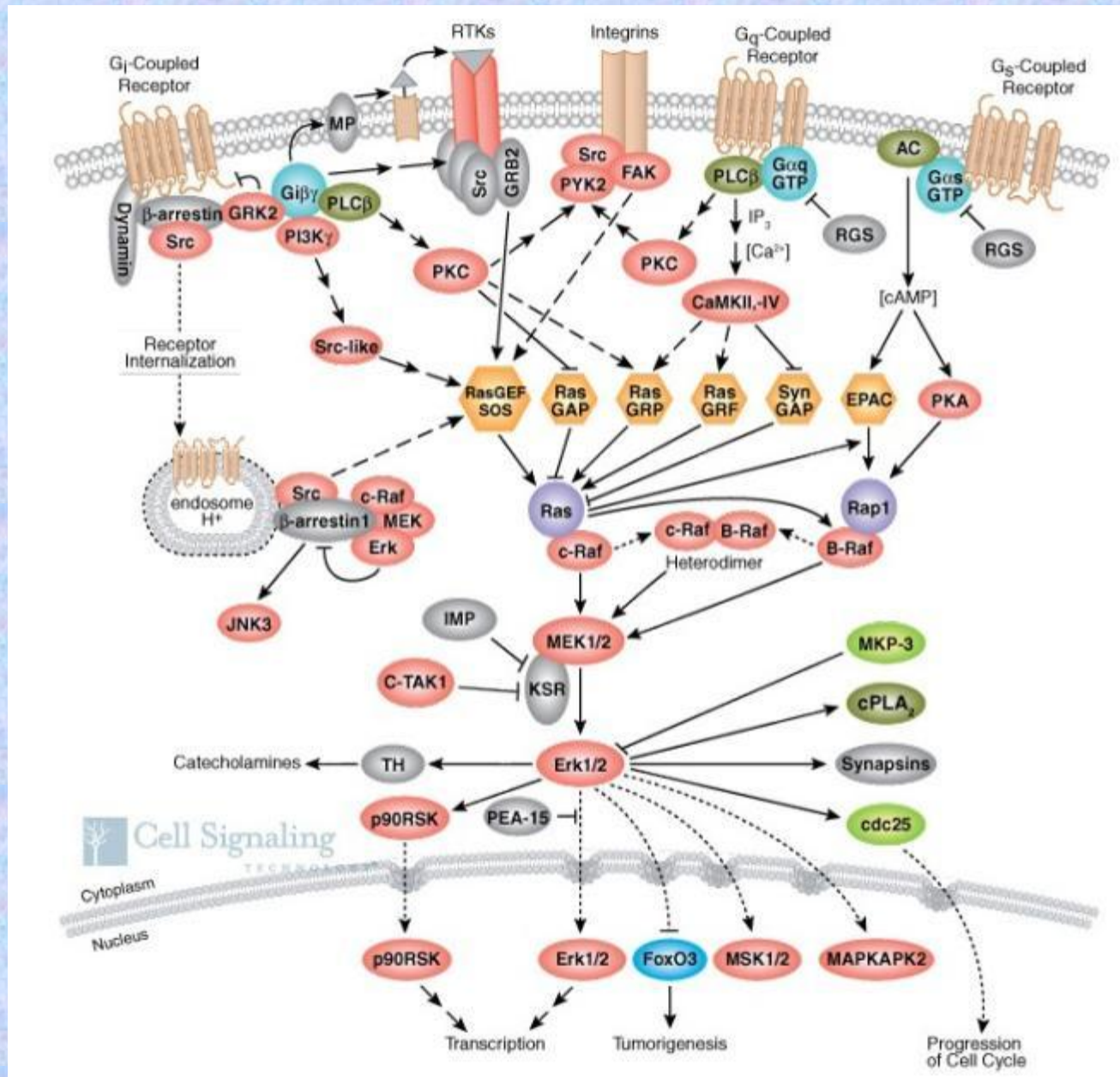
蛋白质组学 (proteomics)
研究所有蛋白质在细胞和
组织中表达的学科。

实用举例:

1. 您在一条深海未知名生物体内发现了一个新蛋白序列。该序列和某已知蛋白结构上十分相似, 但又不完全一样。



2。通过对大量微阵列资料的分析,结果指出某7个蛋白的结构变异与乳腺癌某特殊性状间具有相关关系。



生物信息学的三个研究层面：

1. 单一基因/蛋白的分析， 例如：

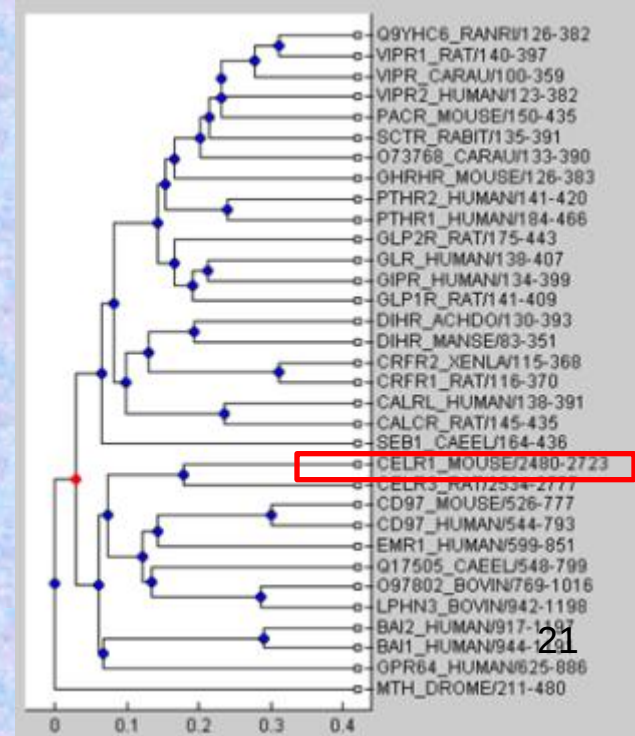
- 某基因在系统发生树上的位置， 与其他基因的相似性和在进化上的关系

- 该基因特殊功能集团的位置与特性

- 该基因的物理性状，
结合位置， 调控位置

- 其亚细胞功能位置预测

- 次级和三级结构的预测



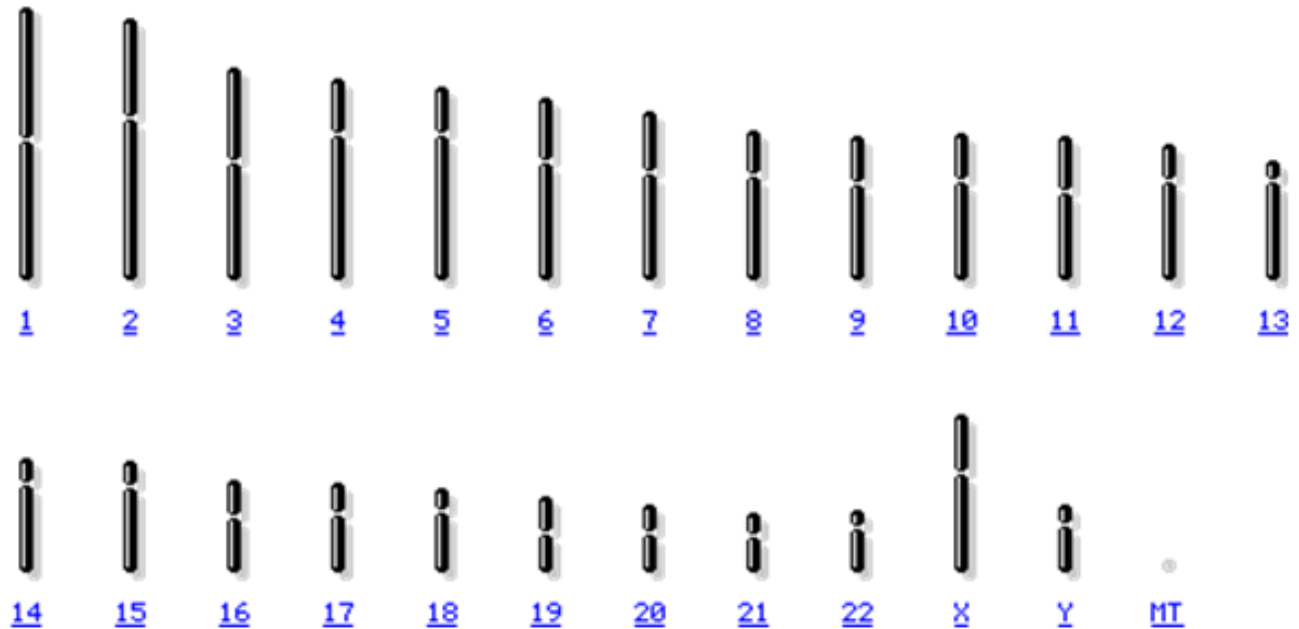
2. 全基因组分析

如：那些基因家族存在，那些缺如？

这些基因在染色体上的位置， 功能关系，
进化学上的特性

Homo sapiens (human) genome view

[Build 37.1 statistics](#) [Switch to previous build](#)



3. 有关功能的基因/基因组分析 如：

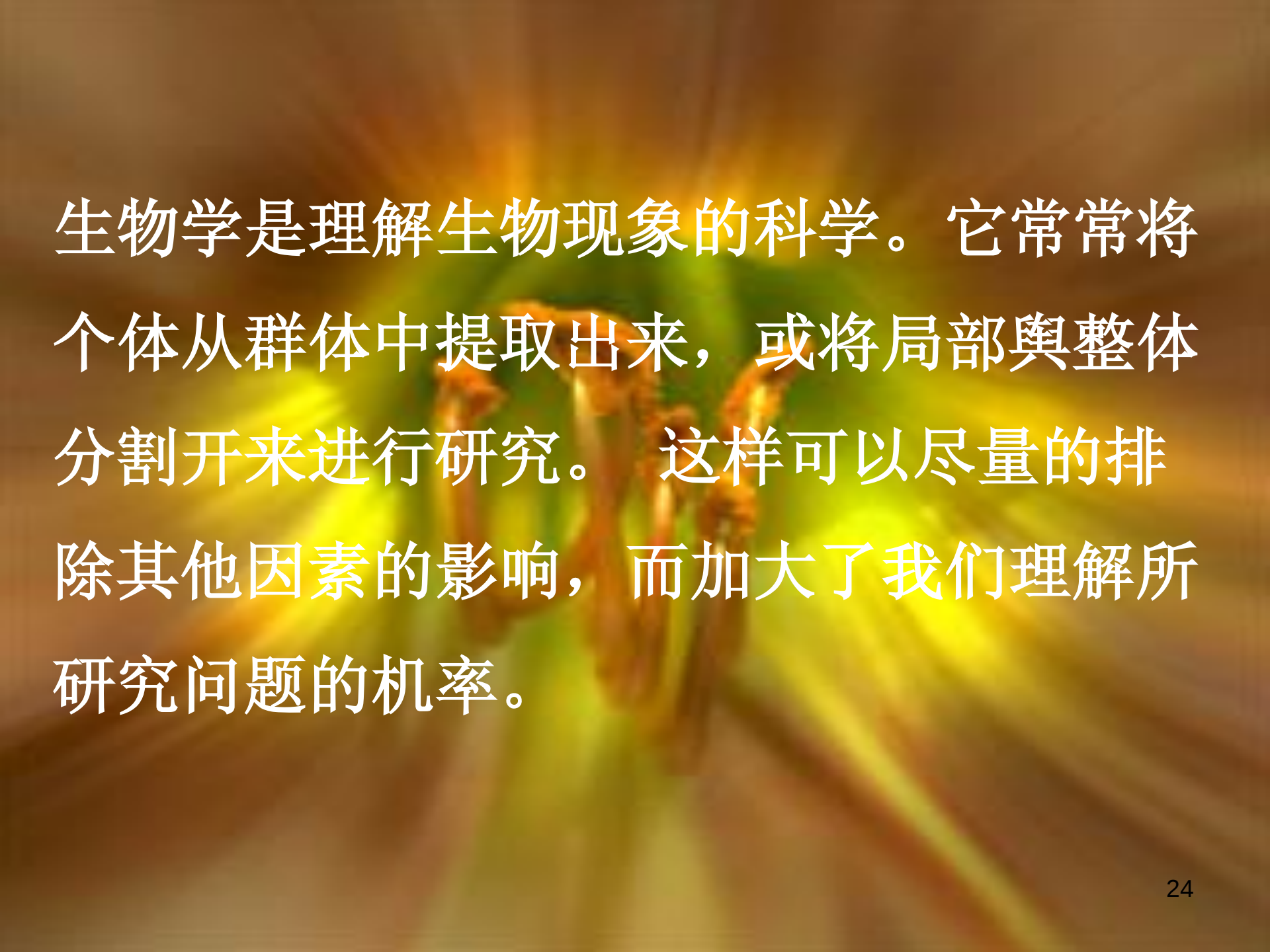
某些基因在微阵列中的表达水平，
信使核糖核酸的测量水平
蛋白组学
蛋白产物的测量
生物化学通路的分析比较
基因突变的基因型与表现型
识别鉴定基本基因与特殊功能基因

相应



Analysis of the
multivariate s

Incubation with secondary,
fluorescent Cy5-coupled
antibodies



生物学是理解生物现象的科学。它常常将个体从群体中提取出来，或将局部与整体分割开来进行研究。这样可以尽量的排除其他因素的影响，而加大了我们理解所研究问题的机率。

因为要深入理解每一个特殊的生物系统，必须从整体着手，将各个系统与结构联结起来。



生命是一个长期动态进化进程的结果，它涵盖了非常宽广的时间与空间尺度。

生物信息学有关的科学研究前景，人才资源结构和工业应用前景。

这些年有许多公司，大学，研究机构在这个领域里慷慨投资，希望获得丰厚的回报。数年下来虽有佼佼者尽显风流，然而也有许许多多的淘金者黯然退出，或空手而归，或损失惨重。

悉尼大学 BioManager的例子

原因：政府资源雄厚，对手强大，有太多的免费资源。

人类资源结构

生物信息学牵涉到了太多的领域, 一个重要的问题是: 如何去训练生物信息学专业人员? 如何使未来的生物信息学家们获得多方面的知识和技能 - 作为生物信息学家必需的知识结构?

具体说来就是如何将这种有机结合了生物学， 数学， 计算机和网络科学的知识结构提供给学生，

甚麽时候合适， 应该是本科生还是研究生？

印度的生物信息学教育

互联网和生物信息学

人人都有巨大的创造性。处于不同国度，地域的人们只要有了相关的基本知识和技能，通过互联网就能涉足生物信息学，甚至成为巨匠！

研究资料 and 结果通过互联网对公众开放，将始终扮演着一个重要且基本的角色，它不仅催化发现，也将帮助发展中国家相关领域科研的成长。

如今我们面临的资料浩如烟海，为解决同一问题不同的研究者开发出了许多有用的工具。

知道哪些是最好的，快速的找到所需要的工具，正确的使用它们，正确无误的分析解读结果，是生物信息学者的一项基本功。

结论

随着生物信息学和基因学的发展，医学实践已经进入了一个新的纪元。许多源于生物信息学的新视野将深深地影响疾病的预防，诊断和治疗。

通过这些工具，我们可以建立新的理论，发现新的药物，矫正和疾病相关的基因位点以及和病理有关的基因表达，等等。

真的，在不远的将来，我们会早期发现将来可能患冠心病的孩子，从而在他的孩童时期就采取预防措施。每一个人在决定组成家庭前会知道他患隐性遗传病的可能性大小。甚至在一些延迟表现的症候群出现以前就能分析该个体的基因概貌。

相似的预防措施可以而且已经在实践中采用了。经由互联网这一切正在发生和发展着。每天都有新的突破。在不久的将来，这些工具将为医学实践提供无价的帮助。

谢谢！