863
生物高技术丛书

# 基因组科学与人类疾病

陈 竺 强伯勤 方福德 主编

科学出版社

## 内 容 简 介

本书是“863”生物高技术丛书之一。书中比较系统地介绍了基因组作图、测序、序列变异及其分析方法、基因表达图谱研究、生物信息学、蛋白质组学、转基因与基因剔除小鼠体系、基因芯片及其应用；以及疾病基因组学，包括单基因遗传病基因和复杂性状疾病如肿瘤、白血病、心血管系统疾病、糖尿病和神经精神疾病等相关基因研究的国内外情况。资料翔实、全面，具有很高的学术参考价值，可供有关专业研究人员、大专院校师生、高新技术行业科技人员和管理人员，以及科技政策决策者参考。

**图书在版编目(CIP)数据**

基因组科学与人类疾病/陈竺、强伯勤、方福德主编. -北京：科学出版社，2001.2
（“863”生物高技术丛书）
ISBN 7-03-008989-8

Ⅰ.基… Ⅱ.①陈…②强…③方… Ⅲ.人类基因-基因组-研究 Ⅳ.R394

中国版本图书馆 CIP 数据核字（2000）第 85273 号

科学出版社 出版
北京东黄城根北街 16 号
邮政编码：100717

新蕾印刷厂 印刷
科学出版社发行 各地新华书店经销

2001 年 2 月第 一 版 开本：787×1092 1/16
2001 年 2 月第一次印刷 印张：18 3/4
印数：1—3000 字数：418 000

**定价：38.00 元**

（如有印装质量问题，我社负责调换〈北燕〉）

## 本书参编作者名单

（按章节顺序排列）


| | |
|---|---|
| 柴建华 | 复旦大学遗传研究所 |
| 杨焕明 | 中国科学院遗传研究所 |
| 于　军 | 中国科学院遗传研究所基因组研究中心 |
| 韩泽广 | 上海第二医科大学 |
| 贺福初 | 军事医学科学院放射医学研究所 |
| 周刚桥 | 军事医学科学院放射医学研究所 |
| 张思仲 | 华西医科大学 |
| 胡赓熙 | 中国科学院上海细胞生物学研究所 |
| 陈润生 | 中国科学院生物物理研究所 |
| 夏其昌 | 中国科学院上海生物化学研究所 |
| 曾　嵘 | 中国科学院上海生物化学研究所 |
| 姚志建 | 中国人类基因组北方研究中心 |
| 成国祥 | 上海市转基因研究中心 |
| 傅继梁 | 第二军医大学 |
| 陆祖宏 | 东南大学 |
| 何农跃 | 东南大学 |
| 赵雨杰 | 东南大学 |
| 孙　啸 | 东南大学 |
| 罗泽伟 | 复旦大学遗传研究所 |
| 张荣梅 | 复旦大学遗传研究所 |
| 陶士珩 | 复旦大学遗传研究所 |
| 刘晓明 | 复旦大学遗传研究所 |
| 夏家辉 | 湖南医科大学 |
| 唐冬生 | 湖南医科大学 |
| 夏　昆 | 湖南医科大学 |
| 赵新泰 | 上海市肿瘤研究所 |
| 王明荣 | 中国医学科学院肿瘤研究所 |
| 王　洁 | 中国医学科学院肿瘤研究所 |
| 吴　旻 | 中国医学科学院肿瘤研究所 |
| 李桂源 | 湖南医科大学 |
| 刘建湘 | 上海第二医科大学，上海市血液病研究所 |

| | |
|---|---|
| 陈　竺 | 上海第二医科大学，上海市血液病研究所，中国人类基因组南方研究中心 |
| 陈赛娟 | 上海第二医科大学，上海市血液病研究所 |
| 朱鼎良 | 上海市高血压研究所 |
| 汤　健 | 北京大学医学部心血管研究所 |
| 方福德 | 中国医学科学院基础医学研究所 |
| 罗　敏 | 上海第二医科大学附属瑞金医院，上海市内分泌研究所 |
| 杜玮南 | 中国医学科学院基础医学研究所 |
| 骆天红 | 上海第二医科大学附属瑞金医院，上海市内分泌研究所 |
| 赵　萸 | 上海第二医科大学附属瑞金医院，上海市内分泌研究所 |
| 黄　薇 | 中国人类基因组南方研究中心 |
| 顾鸣敏 | 中国人类基因组南方研究中心 |
| 王建民 | 中国人类基因组南方研究中心 |
| 金　力 | 复旦大学遗传研究所，中国人类基因组南方研究中心 |
| 沈　岩 | 中国医学科学院基础医学研究所 |
| 贺　林 | 中国科学院上海生命科学中心 |
| 楼　蓉 | 中国科学院上海生命科学中心 |
| 张建刚 | 中国科学院上海生命科学中心 |
| 袁建刚 | 中国医学科学院基础医学研究所 |
| 强伯勤 | 中国医学科学院基础医学研究所 |

# 丛书序 I

生物技术是20世纪末期，在现代分子生物学等生命科学的基础上发展起来的一个新兴独立的技术领域，已被广泛应用于医疗保健、农业生产、食品生产、生物加工、资源开发利用、环境保护，对农牧业、制药业及其相关产业的发展有着深刻的影响，成为全球发展最快的高技术之一。在近20余年的时间里，各种生物新技术不断涌现。70年代创建了重组DNA技术和杂交瘤技术之后，动植物转基因技术、细胞大规模培养技术，以及近几年的基因组学、蛋白组学、生物信息学、组合化学、生物芯片技术和自动化药物筛选技术等相继发展起来。可以说，生物技术的范围在不断地扩展，进入了蓬勃发展的新阶段。

我国的生物技术在“国家高技术研究与发展(863)计划”的支持下，经过15年全国生物技术科技人员的努力拼搏，在农业生物技术和医药生物技术的研究和开发方面都取得了很大的进展。一方面，我们在研究上取得了一批国际影响的创新成果，并获得一批拥有了自己知识产权的专利；另一方面，在开发上已有一批生物技术产品进入市场，还有相当一批产品正在研究开发中；海洋生物技术和环境生物技术也已起步。目前，生物技术研究和产业化已引起了全社会的关注，并将成为我国21世纪的一个新兴支柱产业。

在辞别20世纪，迈入21世纪之际，“863”计划生物领域专家委员会回顾我国生物技术发展历程，展望生物技术发展前景，编写了“‘863’生物高技术丛书”。借此机会，我希望所有从事生物技术研究和开发的科技人员，要进一步团结拼搏，增强创新意识，注重成果转化，为我国生物技术不断发展壮大做出新的贡献！

科学技术部　部长 朱丽兰

2000年7月15日

# 丛书序 Ⅱ

生物技术是20世纪末人类科技史中最令人瞩目的高新技术，为人类解决疾病防治、人口膨胀、食物短缺、能源匮乏、环境污染等一系列问题带来了希望。国际上科学家和企业家公认，信息技术和生物技术是21世纪关系到国家命运的关键技术和作为创新产业的经济发展增长点。

生物技术是指有机体的操作技术。它从史前时代起就一直为人类所开发利用，造福于人类。在我国的悠久历史中，传统的生物技术在经济的发展中一直起重要作用，特别是农业。据传，在石器时代的早期，神农氏曾传授人民如何种植谷物，并实行轮作制度；在石器时代的后期，我国早就善于酒精发酵；在公元前221年的周代后期，我国就能做豆腐并酿制酱油和醋，其所用的基本技术沿用至今。公元前200年，在我国最早的诗集——《诗经》中就提到过采用厌氧菌进行亚麻浸渍处理。早在16世纪，我国的医生就知道，被疯狗咬可以传播狂犬病。公元10世纪，就有了预防天花的活疫苗，到了明朝(1368～1644)，这种疫苗就广泛用于大量人群接种，此后，这种疫苗接种技术通过有名的丝绸之路传入欧洲国家。

1953年Watson和Crick提出了脱氧核糖核酸(DNA)的双螺旋结构模型，阐明了它是遗传信息的携带者，从而开辟了现代分子生物学的新纪元。DNA分子是所有生命机体发育和繁殖的蓝本。众所周知，一切生命活动主要是蛋白质的功能，而蛋白质是由基因编码的。60年代初就破译了“遗传密码”。生命现象千姿百态，但生命体的本质却有高度的一致性。它们的蛋白质都是由20种氨基酸以肽链连接而成，核酸都由4种核苷酸以磷酸链构成，其遗传密码在整个生物界也基本一致。于70年代，科学家们发展了一种新技术，也就是众所周知的DNA重组技术。它向人们提供了一种手段，人们可以在试管内，根据人们的意愿来操作基因、改造基因，新的基因信息可以转入一种简单的生命体中，如大肠杆菌，或转入另一种机体，借以提供一种手段来改造谷物和家畜品种，或生产有效药物，制作疫苗和一系列自然蛋白质，或进行基因治疗。显然，新生物技术是一场革命，是生产力的一次解放，被认为是20世纪人类的一项最伟大贡献，它必将深刻地促进世界经济的发展。

广义的新生物技术包括基因工程、细胞工程、发酵工程和酶工程，但新技术的核心是基因工程技术，它能带动其他生物技术的发展，最具有革命性。

近20年来，国际上生物技术飞跃发展，特别是基因操作技术、生物治疗技术、转基因动植物技术、人类和其他生命体基因组工程、基因治疗技术、蛋白质工程技术、生物信息技术、生物芯片技术等。生物技术的创新正在带动着生物技术巨大产业的发展，它包括基因药物、重组疫苗、生物芯片、生物反应器、基因工程抗体、基因治疗与细胞治疗、组织工程、转基因农作物、兽用生物制品、生物技术饲料、胚胎移植工程、基因工程微生物农药、环保、海洋生物技术，以及现代生物技术对发酵、制药、轻工食品等传统产业的改造等领域。

目前，生物技术产业与信息产业相比较还处于发展初期，至1998年全世界共有生物技术公司3600余家，主要集中在美国和欧洲，其中年产值超过10亿美元的有约20家。

生物技术产业在20年中市场总值增加了50多倍；涨幅最快是在近10年，例如美国在1980年生物技术产品的销售额还处于零增长，1991年达到59亿美元，1996年为101亿美元，1998年增至147亿美元；目前，生物技术仍保持25%左右的增长速度，20%左右的融资率和12.5%就业增长率以及8.76%平均股市涨幅。另一方面，也要看到，美国的1300余家生物技术公司中上市公司为300家，而赢利的公司约为20家，这是由于生物技术产品的研究和开发周期较长，因此从整体看生物技术产业还处在投入阶段。从另一方面来看，尽管美国公司的赢利公司不多，但赢利公司的数量却在稳步上升。

1999年全球生物技术产品的总销售额约为500亿美元，而产生的间接经济效益超过3000亿美元，全球有一半以上的人直接享用过生物技术产品。其主要产品为医药产品、农产品和食品。

我国自1986年实施"863"计划以来的15年中，现代生物技术的开发研究与产业化进入飞速发展阶段：二系法杂交稻的开发与推广对我国的粮食增产起了重要作用，2000年已推广5000万亩以上。1993年我国第一例转基因作物抗病毒烟草进入了大田试验。1997年第一例转基因耐贮存番茄获准进行商品化生产，至1999年5月共有6种转基因作物其产品投放市场。2000年我国转基因抗虫棉花种植面积超过550万亩。1990年我国研制了第一例转基因家畜，1991年山羊克隆获得成功，生物技术饲料添加剂已经实现了规模化生产。我国自1989年第一种基因药物——重组α1b干扰素获准投放市场以来，至1999年我国已有18种基因药物和疫苗获准进行商业化生产，另有26种基因药物处于临床前或临床Ⅰ、Ⅱ期试验，我国生物技术医药产业已初具规模。我国已列为人类基因组计划国际大协作的成员国，承担完成1%的任务，美、英、日、法、德、中科学家于2000年6月26日宣布人类基因组全部DNA序列的工作框架图已经完成。我国在国际上首先发现神经性耳聋的基因，基因治疗已有4个项目进入临床试验阶段；生物芯片技术的开发研究与产业化正在与国际上同步发展。15年来我国在生物技术领域中取得的成就是举世瞩目的，同时还培养了一大批中青年科技人才，为下世纪初S-863计划的实施和生物高技术产业化奠定了扎实的基础，也将为下世纪初我国的经济建设做出应有的贡献。

本丛书是在科学技术部中国生物工程开发中心、"863"计划生物技术领域专家委员会的领导下，由在第一线从事"863"生物高技术研究与开发的科技人员撰写的系列丛书。本丛书包括了农、医生物技术的各个方面，不仅基本上概括了近10年来国际上的研究进展和发展趋势，而且还全面反映了我国"863"计划实施15年来在生物技术领域取得的进展和成果。本丛书的出版无疑将进一步推动我国生物技术开发研究和产业化的进程，促进我国经济的持续发展。同时，本丛书也是培养新一代青年生物技术科学家的重要教科书。

2000年1月16日

# 前 言

自从20世纪90年代初国际上开始实施人类基因组计划以来，我国政府、科技界和企业界对这方面的研究给予了高度重视。1993年国家自然科学基金启动了我国的人类基因组研究，国家“863”计划大力资助基因组及其相关项目研究。国家科技部和北京、上海等地方政府有关主管部门陆续投入经费支持我国的人类基因组研究，设立了若干项目，支持力度逐渐加大。同时，也通过其他各种渠道获得了国内外的资助。所有这些支持有力地推动了我国人类基因组研究工作向前发展。

人类基因组研究是一项庞大的科学工程，它已成为国际科学共同体，参与这个共同体的国家发挥各自的优势，以达到共同协作、优势互补、利益共享之目的。我国是这个共同体的成员之一，不仅承担了1990～2005年第一阶段人类基因组计划中的1%测序任务，在测定十余万个表达序列基础上获得了1000条以上人类新基因的全长cDNA，而且还结合我国的具体国情、需求和优势，启动以疾病基因组学为主要研究内容的功能基因组学研究计划，制订出比较科学、合理和可行的工作目标、策略、途径和方法。在有限的经费投入情况下，我国科技工作者克服困难、努力拼搏、积极进取，做了大量工作，目前已有一批研究论文在国际著名刊物上发表，大量新数据资料在国际数据库注册登录，有些成果申请了专利，还有的研究成果正在转换为实际应用。短短几年中能够取得如此可喜的成绩，说明我国在人类基因组研究领域中正进入实质性进展阶段。

为了回顾、总结和介绍国内外在基因组学研究方面的成果，明确前进的方向，在国家“863”计划的统一筹划下，我们组织了国内从事人类基因组研究的有关专家撰写各相关的内容，编辑成《基因组科学与人类疾病》一书，奉献给广大读者。

需要一提的是，参加编写工作的各位同仁都是在第一线工作的专家，担负着繁重的科研任务，但他们都能在紧张的工作中安排时间完成书搞撰写任务，为本书的顺利出版做出了贡献，我们谨向他们表示衷心的感谢！

人类基因组研究是一个新的科学领域，专业性强，知识积累速度很快，信息量大，“未知数”多，本书只能对其中的若干主要方面的现状和发展趋势进行介绍。而且，由于基因组学各个研究方面进展的不平衡，故各章节在内容的广度和深度的掌握上可能不甚一致，行文风格有所差异，不同的学术观点也得到不同程度的体现，对此我们予以充分尊重，同时希望读者提出意见和建议，以便今后改进。

陈竺　强伯勤　方福德

2000年3月20日

# 目　录

第一章

# 人类基因组物理图谱研究

人类在我们这个地球上已经存在了几十万年至几百万年。人类的现代文明史也已有几千年。很遗憾的是我们对祖先留给自己的“家底”（基因组）一向知之甚少。在古希腊特尔斐城阿波罗神殿上镌刻着的“γνωθισεαυτον”就表示着人类从古时起的一个永恒追求——**认识自我**。尽管人类的这个企望也已经进行了几十年（人类医学遗传学），但真正从 DNA 的分子结构认识基因还是 20 世纪六七十年代以后才开始的。

“人类基因组计划”是 80 年代在全球范围内广泛参与和合作的一项研究计划。目标是全面而透彻地认识人类基因组的正常结构、功能、及基因的异常结构（变异）与人类疾病。这项研究是从 1987 年主要在美国以全基因组 DNA 测序为目标开始进行的（1990 年正式启动）。我们实验室也在国家高技术发展计划（“863”）资助下于同年开始以人类基因组物理图谱分析为目标参于该项研究。英国、法国、意大利、日本、加拿大、澳大利亚、德国等国也都随之参于该项研究。

人类基因组由约 $3\times10^9$碱基对组成，共编码约 8 万个基因，分布在 22 个常染色体，2 个性染色体和 1 个线粒体上。人类基因组研究主要有两个部分或称两个步骤：①人类基因组 DNA 全部序列测定；②基因 DNA 序列的识别和其正常功能，基因变异与人类疾病的研究。开展上述研究必要的第一步就是基因组的“克隆”化和物理图谱构建。人类基因组研究已经进行了 10 年，现在我们已经有了包含有 22 582 个 STS 通过 YAC 构建的物理图谱——STS 图谱（STS content map），分辨率已达 136 kb（http://www.genome.wi.mit.edu/cgi-bin/contig/phys-map,）；X 染色体的物理图谱已包含 STS 2091 个，分辨率达 75 kb（Nagaraja et al.，1997）. 基因组 DNA 已有 468 092 kb 完成测序（NCBI），占人类基因组 DNA 总量的 14.6%，另有草测（draft）序列665 953 kb，占基因组总量的20.8%。22 号染色体 DNA 测序已经完成 33 123 783bp（总长最新估计为 34 Mb），编码序列的计算计分析也已基本完成。在预计的 8 万个基因中，已发现并确定了 7634 个符合孟德尔遗传规律的位点。其中有 5726 个位点已被定位在各染色体上的确定位置(http://www.ncbi.nlm.nih.gov/omim/stats/mimstats.html)。疾病相关基因(包括致病基因、易感基因和抗病基因)有 1385 个(http://bioinfo.weizmann.ac.il/cards-bin/listdiseasecards? type=full)，通过定位克隆(positional cloning)方法克隆的疾病基因已有 108 个(http://genome.nhgri.nih.gov/clone)。全基因组 DNA 序列分析将会在 2003 年之前完成。各种基因的研究还将需要花费更长的时间。

人类基因组的物理图，从广义的角度说，最粗略的图是染色体组型（染色体的细胞遗传学区带）图；最精细的是核苷酸顺序图。这已为大家所熟悉。

## 一、人类基因组 STS 图谱

物理图是以特异的 DNA 序列为标志所展示的染色体图。标志之间的距离或图距以物理距离如碱基对（base pair；bp，kb，Mb）表示。最精细的物理图是核苷酸顺序图，最粗略的物理图是染色体组型图。STS 图谱是最基本和最为有用的染色体物理图谱之一。对人类基因组 DNA 的测序必须从 DNA 克隆开始，对基因组 DNA 每个基因的研究也需要首先得到 DNA 的克隆。曾经有两条路线采用过，“自下而上”（bottom up）和“自上而下”（top down）。20 世纪 80 年代早期之前我们所有的最大容量的克隆载体是黏粒（cosmid），它的最大克隆容量是约 45 kb。那时人们只能设想一个“自下而上”的路线，即在对大量的小片段 DNA 克隆（黏粒、λ 噬菌体等）测序的基础上，根据克隆 DNA 序列重叠关系组装成全染色体 DNA 顺序。此法的重复测序量很大，又一直没有找到更快速的方法，因而效率低，花费多，由于 80 年代后期酵母人工染色体（yeast artificial chromosome，YAC）克隆技术的发展而被放弃，YAC 克隆给物理图带来极大的方便。YAC 是至今最大容量的克隆载体，插入片段可以达到平均 800～1000kb，叫作“mega”（百万碱基）YAC。最大可以达到 2 Mb。于是设想采用“自上而下”的路线。即先构建大 DNA 克隆，并把克隆依染色体排序，这就是“染色体的克隆图”。当把每个克隆测序完成后，就可以“拼装”出整个染色体的 DNA 序列。但是，在 YAC 图谱完成后，发现从 YAC DNA 测序是很困难的，因为纯化单个 YAC 克隆 DNA 要用脉冲电泳技术，有时如果 YAC DNA 的分子量和酵母的一个染色体相当，脉冲电泳也无能为力；另外 YAC 也常有“嵌合现象”，常可造成错误结果。现在实际上用于测序的多为 BAC 或黏粒。于是众多的科学家又转向了 BAC 和黏粒。由于已经有了丰富的 STS 标志资源，BAC 和黏粒的染色体排序变得较为容易。对于单个 BAC 或黏粒的测序，是采取质粒亚克隆和随机测序再拼装的路线，因此，可以说现在的路线是“自上而下”和“自下而上”相结合的。

最早的 YAC 是由 Murray 等于 1983 年构建的，克隆中包含着丝粒、端粒、复制子和外源 DNA 顺序，总长约 55kb。Olson 等于 1987 年构建了第一个人类基因组 YAC 分子克隆库，插入片段平均长度为 150kb。目前被各国实验室采用最广的是法国 CEPH（Centre d'Etude du Polymorphism Human）的“mega”（百万碱基）YAC（Chunakov et al.，1992）。其他被采用的 YAC 克隆库还有英国 ICRF（Imperial Cancer Research Fund）（Larin et al.，1991）YAC 库；美国 ICI Pharmaceuticals（Anand et al.，1990）的 ICI YAC 库；美国华盛顿大学的圣路易斯 YAC 库（Brownstein et al.，1989）。我们实验室也构建了人类基因组 YAC 分子克隆库（柴建华和顾扬洪，1993），插入 DNA 片段长度达到“mega”级（戴长虹和柴建华，1995）。图 1.1 是我们的随机 YAC 克隆的脉冲电泳照片。把大片段 DNA 克隆依其在染色体上所在的位置排序（这就是通常所说的 mapping），可以得到相互重叠的一系列克隆，叫做“克隆重叠群”（contig）。选取有关的克隆进行 DNA 测序，可以较为容易地组装成全染色体或基因组 DNA 顺序。现在所采用

的大片段克隆除 YAC 外还有细菌人工染色体（bacterial artificial chromosome，BAC），黏粒也还被采用，这叫做“自上而下”的路线，是已被广为采用的路线。在构建 YAC-STS 图谱之前需要的基本条件是大片段 DNA 克隆库和足够数量的 STS DNA 标志


图 1.1　YAC 脉冲电泳图。以 pBR322 DNA 为探针与 YAC 杂交的放射自显影。

YAC 的克隆容量大是其主要优点。但是它也有重要缺点，即在构建的克隆库中含有大量的嵌合体（chimeric）克隆存在（40%～50%）。嵌合体是由不同染色体来源的 DNA 片段连接在一起构成的克隆。它给研究造成的麻烦很大。我们不能把这种克隆放在任何染色体的任何位置。它也不能被用来进行 DNA 测序或基因的克隆研究。它是由于两个不同 YAC 在同一个细胞内经同源重组产生的。曾有人用具有双重重组缺陷的酵母细胞作为宿主构建 YAC 克隆库，使嵌合率下降至 3%。YAC 的第二个重要缺点是 YAC 克隆 DNA 分离困难。YAC 的大小和性质与酵母染色体无明显差异，不能用简单的方法分离纯化 YAC DNA。在脉冲凝胶电泳上也有时与酵母染色体相重叠，以致很难得到纯化的 YAC DNA，以用于测序和基因的克隆研究。在某种需要时可用 Alu-PCR 方法特异地扩增人类基因组 DNA，作为 YAC 分离的一种补充方法。

BAC 是另一种大片段 DNA 克隆。BAC 是以大肠杆菌 F 因子为基础构建的一种大片段 DNA 克隆载体（Shizuga et al.，1992），其克隆容量可达 300 kb。目前构建的克隆库克隆片段平均长度通常在 125～150kb. 明显地比 YAC 小。但是它是一种环状分子，在大肠杆菌细胞内非常稳定，不会发生缺失或重排，而且容易分离纯化，于是近两年日益被研究者接受。但是 YAC 在数年来已有大量数据积累，人类基因组物理图谱仍然主要是靠 YAC 克隆构建的。BAC 并不能完全替代 YAC。

STS（sequence tagged sites）作为一种基因组染色体位标是由 Olson 等于 1989 年提出的。STS 是从单拷贝 DNA 序列发展而来的一对 PCR 引物，因此 STS 是基于 PCR 方法的 DNA 标志，操作简单快速。每个 STS 在基因组内有特定而单一的同源位置。STS 在染色体上的定位在前几年主要用 FISH（flurescent *in situ* hybridization）方法，但是

FISH方法较为费时费力。近年新发展的辐射杂种细胞法（radiation hybrid cell panel）则较为方便和快速。对每一个选用的STS可以用PCR方法筛选YAC克隆库，可以得到包含有同一位点的一组YAC克隆。这一组YAC克隆叫做“克隆重叠群”。一个YAC可以包含有几个STS。STS在YAC DNA上的次序可以其在不同YAC中的存在于以测定，但无法给出其相互间的距离。这种DNA标志间的紧密连锁关系是“自下而上”的信息。对于相距大于1Mb的DNA标志，它们一般不可能同时存在于一个YAC内，则要靠YAC重叠群来测定。两个相邻的DNA标志不能单独依靠在同一个YAC中的存在予以证实，因为有很多YAC可能是嵌合体，或可能存在实验错误。如果在2或3个YAC中都得到同一的结果，则此结果是较为可信的。一个STS可以同时存在于一组YAC内。一个克隆重叠群可以包含多个STS和多个相互重叠的YAC克隆。靠这种相互的连锁关系，我们就可以测定一个小区间内若干个DNA标志的连锁关系。如果我们有足够多的STS位标并在染色体上的分布达到足够密度，我们就可以根据每个YAC所含有的STS把各YAC排列在染色体上，得到一个相互重叠排列的染色体的YAC图谱。这就是通常所说的“STS mapping”的含意。这个图谱就叫作YAC-STS物理图谱。STS位标中有些是来自非编码区的单拷贝DNA顺序，它们只起位标作用。有些还有多态性，这部分STS的引物顺序仍然是具有单拷贝性质，但是扩增产物中通常包含着串联短重复序列（1～6bp，short tandem repeats，STR），这类STS又可称为微卫星（microsatellite）DNA，这种短重复序列的重复次数在个体间常有差异，即所谓”多态性”，因而这类DNA标志除可以用作染色体位标外，还可用于个体鉴定，在刑事侦察中有重要用途。也用于基因的连锁分析定位。还有一些本身就是基因，它们中有些可能原本是从cDNA发展而来的，称之为EST（expressed sequence tags）。STS在染色体上的排列顺序就构成了染色体的STS图谱。依据这种图谱可以构建一个依染色体连续排列的YAC图谱，它对于下一步寻找和克隆基因十分有用。Collins和Galas提出的1994～1998年的人类基因组研究目标中物理图谱的指标是STS分辨率达到100 kb（1993）。HUGO（Human Genome Organization）组织各国有关科学家成立各染色体专题委员会（http://gdbwww.gdb.org/gdb/hugoEditors.html），负责收集和编辑有关图谱资料。各染色体的STS图谱已经基本完成，分辨率分别达到或超过100 kb。各染色体STS内容日益丰富，这些图谱都可以通过WWW在GDB（genome data base）数据库（http://gdbwww.gdb.org/）中找到。由于篇幅十分庞大，而且内容频繁扩充和更新，这种图谱多不印刷发表，而是由INTERNET提供读者随时查阅。美国DOE（Department of Energy）和NIH（National Institute of Health）分别支持22个Human Genome Center，它们是大规模进行人类基因组研究的主要力量。我们实验室对人X染色体短臂Xp11.21～p21.3长约35Mb区段进行了YAC-STS物理图谱分析（魏勇等，1995；魏勇等，1996；缪为民等，1997；魏勇等，1997），图1.2是我们实验室构建的人X染色体短臂该区段的YAC-STS图谱。

人类基因组物理图谱的分辨率在过去的十年中以每年45%的速度增长。如果按这个增长指数继续增长，到全基因组DNA测序完成时，物理图中将会包含有上百万个STS标志。那时分辨率可达3kb。

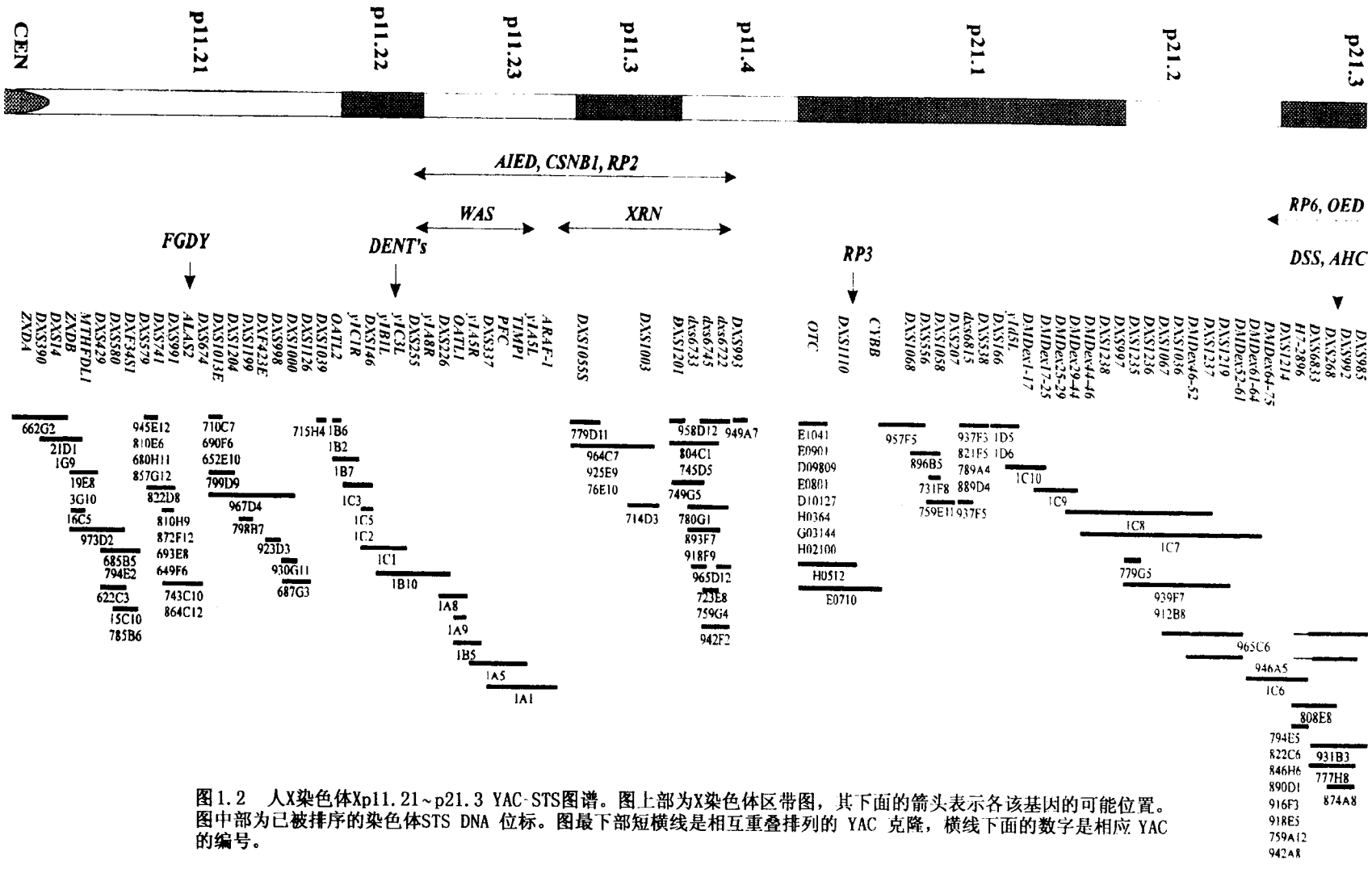


图1.2 人X染色体Xp11.21~p21.3 YAC-STS图谱。图上部为X染色体区带图，其下面的箭头表示各该基因的可能位置。图中部为已被排序的染色体STS DNA 位标。图最下部短横线是相互重叠排列的 YAC 克隆，横线下面的数字是相应 YAC 的编号。