

贺林主编

解

码

生

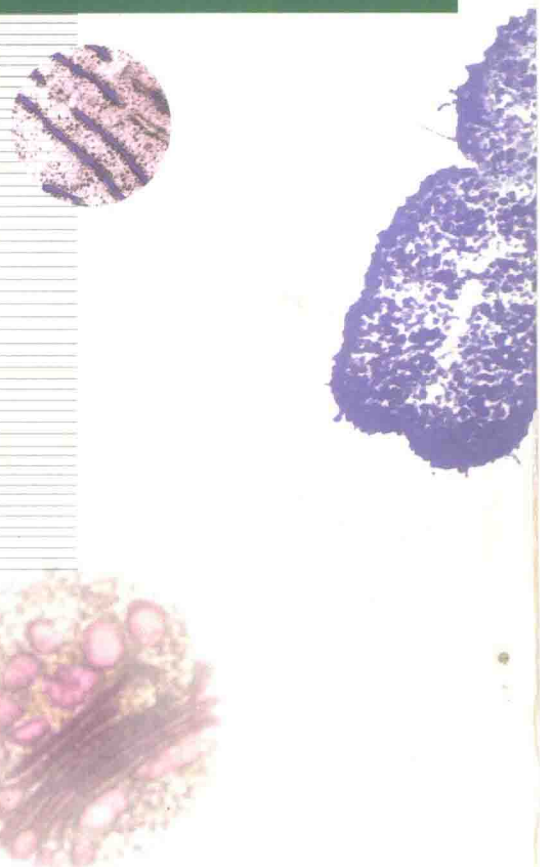
命

人类基因组计划
和后基因组计划



C

G



科学出版社

解码生命

——人类基因组计划和后基因组计划

贺 林 主编

科学出版社

2000

内 容 简 介

有史以来,人们就为认识自身进行着不懈的努力。但真正有系统地“解码生命”是于1990年前后在人类基因组计划启动后才开始的。本书正是在这个基础上,通过较详细地介绍人类基因组计划和后基因组计划从而展示解码生命的思路与过程。

人类基因组计划进展迅速,此书力求从该领域发展的主线入手,尽可能做到准确地反映最新发展动态,突出一个“新”。本书图文并茂,内容基本做到了既能满足业余爱好者和政策制定者对这一领域发展的认识与了解,又使专业工作者能够对这一领域研究的最新动态和最新技术有所掌握,是非常值得向广大生物、医学工作者推荐的一本好书,书中的内容也非常有利于大中学生对21世纪科学发展前沿的理解。

图书在版编目(CIP)数据

解码生命——人类基因组计划和后基因组计划/贺林主编. -北京:科学出版社,2000.4

ISBN 7-03-008072-6

I. 解… II. 贺… III. 人类基因-基因组-科学研究-计划
IV. R394-3

中国版本图书馆CIP数据核字(1999)第67429号

科学出版社出版

北京东黄城根北街16号
邮政编码:100717

中国科学院印刷厂印刷

新华书店北京发行所发行 各地新华书店经售

2000年4月第 一 版 开本:850×1168 1/16
2000年4月第一次印刷 印张:35
印数:1—4 000 字数:809 000

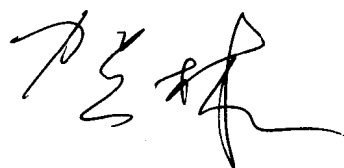
定价:80.00元

(如有印装质量问题,我社负责调换〈环伟〉)

献给我年迈的父母,是他们的力量不断地激励着我在我国本土上献身于中国科学!

献给中国人类基因组计划,是这个计划使我看到了中国科学的光明!

献给我敬爱的老师、同事和学生们,是他们的鼓励和努力使我们的课题组不断进取!

Handwritten signature in black ink, appearing to read 'Li Zhen' (林志).

ABE21/07

编写人员

(按姓氏笔画顺序排列)

- 丁远春 中国科学院昆明动物研究所细胞与分子进化开放研究实验室, 昆明 650223
- 于军 Genome Center, University of Washington, Seattle, WA98195-1700, USA; 中国科学院遗传研究所, 北京 100101
(Jun Yu)
- 于常海 香港科技大学生物系, 香港; 中国科学院上海生命科学研究中心, 上海 200031
(Albert C G Yu)
- 王子晖 香港科技大学生化系, 香港
(Jeffrey Tze-Fei Wong)
- 王志珍 中国科学院生物物理研究所生物大分子国家重点实验室, 北京 100101
- 王红艳 中国科学院上海生命科学研究中心, 上海生理研究所, 上海 200031
- 邓可京 复旦大学遗传研究所人类基因组实验室, 上海 200433
- 叶希贤 香港科技大学生化系, 香港
(Hei Yin Yip)
- 冯国鄞 上海精神卫生中心, 上海 200030
- 李胜 中国科学院上海生理研究所, 上海生命科学研究中心, 上海 200031
- 李孝圭 中国科学院上海生理研究所, 上海生命科学研究中心, 上海 200031
- 李涛 Genetics Section, Institute of Psychiatry, London SE5 8AF, UK
(Tao Li)
- 刘万清 中国科学院上海生理研究所, 上海生命科学研究中心, 上海 200031
- 刘晓明 复旦大学遗传研究所群体与数量遗传学实验室, 上海 200433
- 许田 Molecular Oncology and Development, Boyer Center for Molecular Medicine, Yale
(Tian Xu)
University School of Medicine, USA; 复旦大学遗传研究所人类基因组实验室,
上海 200433
- 乔守怡 复旦大学遗传研究所人类基因组实验室, 上海 200433
- 沈伯松 Genetics Section, Institute of Psychiatry, London SE5 8AF, UK
(Pak C Sham)
- 汪建 北京华大基因研究中心, 中国科学院遗传研究所, 北京 100101
- 吴旻 分子肿瘤学国家重点实验室, 北京 100021
- 杨焕明 北京华大基因研究中心, 中国科学院遗传研究所, 北京 100101
- 杨新平 中国科学院上海生理研究所, 上海生命科学研究中心, 上海 200031
- 沈韬 中国科学院上海生理研究所, 上海生命科学研究中心, 上海 200031
- 张亚平 中国科学院昆明动物研究所细胞与分子进化开放研究实验室, 昆明 650223
- 张荣梅 复旦大学遗传研究所群体与数量遗传学实验室, 上海 200433
- 陈仁彪 上海第二医科大学医学遗传学和生物学教研室, 上海 200025
- 陈竺 中国人类基因组南方中心, 上海第二医科大学血液病研究所, 上海 200025
- 陈伟丽 复旦大学遗传研究所人类基因组实验室, 上海 200433

邹承鲁 中国科学院生物物理所生物大分子国家重点实验室,北京 100101
邱仁宗 中国社会科学院应用伦理研究中心,北京 100732
罗泽伟 复旦大学遗传研究所群体与数量遗传学实验室,上海 200433
贺 林 上海交通大学 Bio-X 生命科学研究中心,上海 200030;中国科学院上海生理研究所,上海生命科学研究中心,上海 200031
南新升 MRC Human Genetics Unit,Edinburgh EH4 2XU,U. K.
(Xinsheng Nan)
柏绍春 中国科学院上海生理研究所,上海生命科学研究中心,上海 200031
施小六 湖南医科大学第二附属医院,长沙 410078
赵京华 Genetics Section,Institute of Psychiatry,London SE5 8AF,UK
(JinHua Zhao)
徐 迅 上海专利商标事务所,上海 200233
袁义达 中国科学院遗传研究所,北京 100101
柴建华 复旦大学遗传研究所人类基因组实验室,上海 200433
蒋嵩山 中国科学院上海生理研究所,上海生命科学研究中心,上海 200031
强伯勤 中国人类基因组北方中心,中国协和医科大学基础医学院,北京 100005
薛 红 香港科技大学生化系,香港
(Hannah H Xue)

图表设计、绘制:沈 韬

序

人类基因组计划在 1990 年前后启动,标志着人类在揭示和认识自身的道路上又大大地迈进了一步,同时也标志着系统地进行“解码生命”的正式开始。

这项堪与阿波罗登月计划和曼哈顿原子弹计划相提并论的人类历史上第三大工程,经历了 5 年左右的辩论和酝酿终于以美国人类基因组计划为代表于 1990 年启动,世界许多其他国家的人类基因组计划也先后启动,很快形成了一个世界范围内的大协作。此项目一开始即受到世界各国的密切关注,主要是因为该计划的意义不仅在于它的科学内涵还在于它的直接经济利益,这也是为什么导致了除政府的大量投资外,各类私有企业的介入并投入了巨大的资金,这一状况使得对人类基因组的开发工作大大超过了所计划的进程。

中国作为一个人口大国、遗传资源大国,同时又是一个经济尚处在发展中的不是很富有的国家,面对这一机遇与挑战应如何对付?其结论是绝不能放弃此一机遇,经费不足就设法走一条经费不足的路。在经过了一段时间的酝酿后,中国人类基因组计划也于 1993 年正式启动。这的确是中国生物学史上的一件大事。经过了 5 年左右的运作,它已从一个萌芽期小规模的模式走到了一个具有南北两个国家级中心和数个具有相当规模的由单位或企业投资组建的中心的格局。

世界和中国的人类基因组计划的形势是喜人而又逼人,面对着这样的机遇与挑战,中国需要一本自己的书介绍和概括人类基因组计划的历史和发展,以期更多的中国人了解和参与此项宏伟的工作。贺林教授所主编的这本《解码生命——人类基因组计划和后基因组计划》正是在这一重要时刻问世的。

此书引用了广而新的参考文献,较准确地反映了最新发展动态。书中的内容基本做到了既能满足业余爱好者和政策制定者对这一领域发展的认识与了解,又使专业工作者能够对这一领域研究的最新动态有所掌握,是非常值得向广大生物、医学工作者推荐的一本好书。

贺林
1999.4.2.

前 言

人类基因组计划在西方世界已深为人知,其重要性已超越了最初的设计。并且,其显著的经济利益驱使了大量公司的涌入。技术在更新,任务在推进,人类基因组计划的最主要目的是解读人类 DNA 的全序列,这也是为了“解码生命”所跨出的最关键一步。由此计划而带来巨大的经济效益的时代已为时不远。

面对这样的现实,中国的人类基因组计划虽然起步晚了一些,但正穷追不舍。除中国科学院已启动了自己的人类基因组计划外,中国还相继在上海和北京各成立南北两大中心,制定了自己的策略。

正是在这样的形势下,中国亟需一本能尽可能全面反应这个领域发展概况的书籍。碰巧在几位博士生和博士后的一再建议下,虽然仍经过了再三犹豫,使我最终还是下了决心于 1998 年春开始动手编写此书。当时犹豫再三的主要原因是:此领域的发展真可谓到了“眼花缭乱”的地步,在你落笔的同时,大量新的内容和数据又涌现出来了;所以永远难以真实反映这一领域发展的最新进展。

然而,没想到这“壮胆子”的行动一开始既得到了该领域多位专家的大力支持,实际上很快扩大到了国内外一些著名人士的参与,为该书的权威性“助了威”。并且,专家们所表示的文责自负的责任心更为内容的严密性及科学性增添了一道“门槛”。有时在章节间会出现某些内容的重叠,但为了保障各章的完整性和连续性,采取了尽量不去删动的方针。

此书的编写与中国人类基因组研究的大好形势是分不开的。如果没有今日中国人类基因组计划的形势,此书也就大大失去了“品尝”的味道。所以大有必要借此机会感谢那些对中国人类基因组计划做出贡献的人们。除了书中的大部分作者外,主要还有这一计划的发起者:谈家桢、吴旻、强伯勤、陈竺、杨焕明等;中国科学院路甬祥、许智宏、陈宜瑜、竺玄、王贵海、田彦、尹萍、袁萍等;中国科技部的朱丽兰、惠永正、刘谦、徐新来等;中国自然科学基金委员会的叶鑫生、童道玉、朱大保、杜生明、王钦南等;中国卫生部刘雁飞等以及对新生事物总是体现出领悟在先的上海科委的华裕达、朱寄萍、李明轩、徐国华、李跃、邵钧、肖惠萍等。此外,除了书中的作者外,还得到了一些知名人士和后起新秀的支持与参与,他们是(按姓氏笔画排列):马永兴、王亚新、王亚辉、王志新、王宗光、王宪、毛裕民、方福德、李林、李伯良、李载平、刘国仰、朱章玉、汤章成、汤建、余龙、严光美、沈岩、吴建屏、杨胜利、杨雄里、张芹、张思仲、罗九甫、林志新、林其谁、金力、施履吉、洪孟民、赵国屏、赵寿元、闻玉梅、胡以平、贺福初、顾牛范、顾建人、徐晋麟、夏家辉、袁建刚、唐建国、黄谷扬、黄薇、曾益新、曾溢滔、傅继梁、谢绳武、谢赏恩(Kathryn Cheah)、敖世洲、褚嘉佑、翟中和、裴钢等,以及一大批在第一线为这一计划做实际工作的实干家们的直接贡献。

贺林

2000.1

目 录

序 (谈家桢)

前言

第 1 篇 人类基因组计划核心篇

| | |
|-----------------------|-----|
| 第 1 章 人类基因组计划与解码生命 | 3 |
| 第 2 章 基因组的结构与组织 | 22 |
| 第 3 章 遗传图谱 | 52 |
| 第 4 章 物理图谱 | 84 |
| 第 5 章 序列图谱/大规模测序与测序策略 | 123 |
| 第 6 章 基因图谱 | 140 |

第 2 篇 人类基因组计划延伸篇

| | |
|--------------------------|-----|
| 第 7 章 模式生物 | 185 |
| 第 8 章 比较基因组学 | 202 |
| 第 9 章 基因突变研究 | 225 |
| 第 10 章 基因的修饰、调控与疾病发生的关系 | 242 |
| 第 11 章 基因组与环境以及药物开发的联系 | 265 |
| 第 12 章 基因组多样性 | 286 |
| 第 13 章 后基因组与蛋白质组 | 322 |
| 第 14 章 生物信息学及其在基因组研究中的应用 | 333 |

第 3 篇 专利、伦理及实践篇

| | |
|----------------------|-----|
| 第 15 章 专利的功效 | 355 |
| 第 16 章 人类基因组研究的伦理学问题 | 379 |
| 第 17 章 中国人类基因组计划的切入口 | 390 |

| | |
|---------------|-----|
| 附录 1 相关的资料和方法 | 405 |
|---------------|-----|

| | |
|---------------|-----|
| 附录 2 常用的微卫星标记 | 427 |
|---------------|-----|

| | |
|---------------------------------------|-----|
| 附录 3 人类基因组测序参加单位索引(HGSI Participants) | 531 |
|---------------------------------------|-----|

| | |
|--------------|-----|
| 英汉术语和机关名称对照表 | 533 |
|--------------|-----|

第 1 篇

人类基因组计划 核心篇

人类基因组计划与解码生命

- | | |
|----------------------|----------------|
| 1.1 引言 | 1.3.2 物理图谱 |
| 1.1.1 问题的提出 | 1.3.3 序列图谱 |
| 1.1.2 历史的回顾 | 1.3.4 基因图谱 |
| 1.1.3 Dulbecco 短文的功绩 | 1.4 重要性和经济利益 |
| 1.2 世界的行动 | 1.5 “土地”的分割 |
| 1.3 任务与进展 | 1.5.1 现状 |
| 1.3.1 遗传图谱 | 1.5.2 遗传资源至关重要 |
| 1.3.1.1 第 1 代标记 | 1.6 人类基因组计划的延伸 |
| 1.3.1.2 第 2 代标记 | 1.7 面对的基因专利问题 |
| 1.3.1.3 第 3 代标记 | 1.8 伦理问题不容忽视 |
| 1.3.1.4 意义 | 参考文献 |

1.1 引言

1.1.1 问题的提出

尽管生物机体的尺寸有限,但并未能为研究工作带来任何容易之处。人们经过了不懈的努力,渴望解开生命之谜这个多年的愿望并未向前推进多少,谜仍是个谜! 以往研究的艰履或失败教训使人们头脑开始清醒地认识到,任何仅依靠单一学科如细胞学、发育学、肿瘤学、人类遗传学或分子生物学的独自努力都无济于事,都太局限了,难以完成人类对自身的认识和保护。美国曾投巨资但基本上以失败告吹了的肿瘤十年计划也说明了这个问题。所以,要知道某事物的局部作用机制最好先知道全局的看法逐渐主导了人们的认识(Dulbecco R, 1986)。在绕了一大段弯路后,人们回过头来决定开始进行人的所有基因即基因组的研究,全面探讨这个“摸得到,猜不透”的人体奥秘,由此形成了基因组学(genomics)和人类基因组计划(Human Genome Project, HGP),其最终目的是对生命进行系统地科学地解码,以此达到了解和认识生命的起源,种间和个体间存在差异的起因,疾病产生的机制以及长寿与衰老等生命现象(Lander ES, 1996)。人类基因组计划以前的遗传学或称基因学(genetics)偏重于单个基因的研究,而人类基因组计划则是把目光投向整个基因组的所有基因,从整体水平去考虑基因的存在、基因的结构与功能、基因之间的相互关系等。随着数理化、信息和材料等学科的渗透以及具有时代特征的工业化技术管理模式的引进, HGP 真正成为了生命科学领域的第一项大科学工程,其规模和意义远远超过阿波罗(Apollo)登月计划和曼哈顿

(Manhattan)原子弹计划。HGP的正式启动也就标志着解码生命的真正开始!因此,HGP也就很自然地成为人们关注的焦点。

1.1.2 历史的回顾(Cantor C,1990;Watson J,1990)

对人类基因组的研究在70年代已具有一定的雏形,在80年代在许多国家已形成一定规模,并在以下的几个事件的影响下形成了投资额最多、最具规模的美国人类基因组计划。

1984年在Utah州的Alta,White R和Mendelsonhn M受美国能源部(DOE)的委托主持召开了一个小型专业会议讨论测定人类整个基因组的DNA序列的意义和前景(Cook-Deegan RM,1989)。1985年5月在加州Santa Cruz由美国能源部的Sinsheimer RL主持的会议上提出了测定人类基因组全序列的动议,由此形成了美国能源部的“人类基因组计划”草案。1986年3月,在新墨西哥州的Santa Fe讨论了这一计划的可行性,随后美国能源部宣布实施这一草案。1986年著名遗传学家McKusick V提出从整个基因组的层次研究遗传的科学称“基因组学”。1986年3月7日,诺贝尔奖获得者Dulbecco R在*Science*杂志上发表的一篇有关开展人类基因组计划的短文。1986年6月在美国冷泉港,另两位诺贝尔奖获得者Gilbert W及Berg P主持了有关“人类基因组计划”的专家会议。1987年初,美国能源部与国家健康研究院(NIH)为“人类基因组计划”下拨了启动经费约550万美元(1987年全年1.66亿美元),并开始筹建人类基因组计划实验室。1988年2月,国家科学研究委员会(NRC)的专家撰写了“人类基因组的作图与测序(mapping and sequencing the human genome)”的报告,全面地介绍了有关这项史无前例的、看起来似“胆大妄为”计划的内容(National Research Council,1988)。同年,美国成立了“国家人类基因组研究中心”,由因提出DNA分子双螺旋模型的贡献而获诺贝尔奖的沃森(Watson J)出任第一任主任。

1.1.3 Dulbecco短文的功绩

Dulbecco R于1986年在*Science*杂志上发表的题为“癌症研究的转折点——人类基因组的全序列分析”的短文,回顾了70年代以来癌症研究的进展,使人们认识到包括癌症在内的人类疾病的发生,都与基因直接、间接有关;同时,他指出要么仍处在用“零敲碎打”的方法(piecemeal approach)开展研究,要么从整体上研究和分析整个人类基因组及其序列。

Dulbecco R在他的文章中还指出:“这一计划的意义,可以与征服宇宙的计划媲美。我们也应该以征服宇宙的气魄来开展这一计划”,并且谈到:“这样的工作是任何一个实验室难以单独承担的项目。这个世界上发生的一切事情,都与这一人类的DNA序列息息相关”。今天当我们重读这一短文时仍感到Dulbecco的胆识和引起全世界巨大反响的意义所在。这也是为什么许多单项领域的进展都有足够的理由在世界范围内提出一个“计划”,如“遗传工程计划”、“肿瘤计划”、“脑计划”、“蛋白质计划”、“信号传导计划”等时,人们最终选择和接受“人类基因组计划”作为全球性重大计划的原因所在。这看来是由于自然科学的自身规律,内在联系和发展阶段等因素导致了人们的最后选择。如果仔细想一下,所有这些计划的最关键因素都离不开基因的作用,就会倍加感到Dulbecco短文的高瞻远瞩,只有弄清楚了人类基因组的全部或大部分基因的工作情况,实施其他计划才成为可能。

1.2 世界的行动

历经5年左右的辩论后,美国国会正式批准美国的“人类基因组计划”于1990年10月1日正式启动。其规模在世界上是最大的,总体计划是在15年内投入至少30亿美元进行人类全基因组的分析。

在Dulbecco短文的影响下,整个西欧几乎全部都动了起来,并各具特色。最早开始国家级HGP的是意大利的国家研究委员会(National Research Council)。1987年,该委员会组织了15个,以后发展到30个实验室开始了人类基因组计划的研究。其特点是技术多样(YAC、杂种细胞、cDNA等)、区域集中(基本上限于Xq24-qter区域)。

1989年2月开始的英国HGP的特点是:帝国癌症研究基金会与国家医学研究委员会(ICRP-MRC)共同负责全国协调与资金调控;剑桥附近的Sanger中心注重首先在线虫基因组上积累经验,改进大规模DNA测序技术;同时,建立了YAC库的筛选与克隆、特异细胞系、DNA探针、基因组DNA、cDNA文库、比较生物基因组DNA序列、信息分析等的“英国人类基因组资源中心”,并向全国的用于研究目的会员免费提供技术及实验材料服务。可谓“资源集中,全国协调”。

1990年6月法国的国家人类基因组计划开始启动。科学研究部(Ministry of Research)委托国家医学科学院(INSERM)制定人类基因组计划,主要特点是注重整体基因组、cDNA和自动化。诺贝尔奖金获得者Dausset J于1983年用自己的奖金建立的CEPH(人类多态性研究中心)和法国民众捐款至少5000万美元建造的Genethon为全世界的HGP做出了不可磨灭的贡献,特别是在全基因组YAC重叠群、微卫星标记(遗传图)的构建以及驰名世界的用作基因组研究的经典材料CEPH家系(80个3代多个体家系)的影响方面产生了巨大的影响。

1995年6月德国才正式开始了HGP。虽然起步晚,但来势迅猛。他们先后成立了资源中心和基因扫描定位中心,并已开始了对21号染色体的大规模测序工作。

此外,1990年6月欧洲共同体通过了“欧洲人类基因组研究计划”,主要资助23个实验室重点用于“资源中心”的建立与运转。这也是欧洲计划在多个领域赶超美国的一个重要方面。另外,榜上有名的还有丹麦、俄罗斯等国,以及亚太地区的日本、韩国、澳大利亚等。

中国HGP是于1994年初在吴旻院士、强伯勤院士、陈竺院士和杨焕明教授的倡导下启动的。最初由国家自然科学基金委员会和“863”高科技计划的支持下,先后启动了“中华民族基因组中若干位点基因结构的研究”和“重大疾病相关基因的定位、克隆、结构与功能研究”(陈竺等,1998)。经过了几年的摸索已逐渐形成了自己的体系,并进一步看清了开展这项工作的重要性。1998年在国家科技部的领导和牵线下,决定在北京和上海各成立一个中心。上海中心以陈竺院士挂帅于同年的3月份正式揭牌,并于10月份改名为中国南方基因中心。同时,决定成立若干中国人类遗传资源保护中心,国家卫生部起到牵头的作用。中国科学院和中国医学科学院以及一些大学、研究所也为中国的HGP做了大量的工作和注入了资金。1998年中国科学院遗传所由杨焕明和余军教授组织,以及1999年由强伯勤院士挑头在北京又先后成立了中国科学院北京人类基因组中心和北方人类基因组中心。尽管它们显

得有些姗姗来迟了,但使我国在人类基因组的研究实力豁然间得到了进一步的增强。

然而,除了早期的政府介入外,时至今日,世界上几乎所有大大小小的医药公司都卷入了这场所谓的 HGP,无形中已形成了一场“抢基因”大战。为了迎接挑战,由 Collins F 领导的美国国家人类基因组计划和 Morgan M 负责的由 Wellcome 慈善基金会所资助的英国人类基因组计划决定于 2000 年春天完成人类基因组具有 90% 序列的“工作框架图”(Pennisi E, 1999), 而到 2003 年完成具有 99.99% 的高精度的序列图 (Marshall E, 1999)。这里包含着两层意思,第一层也是主要的一层,为了迎接公司方面的挑战,特别是对付像 Celera Genomics 这类的公司所号称的将于 2001 年测完人类基因组的全部序列 (Marshall E and Pennisi E, 1998) 的“宣战”;第二层意思则是于 2003 年庆贺 DNA 双螺旋模型提出 50 周年。

1.3 任务与进展

HGP 的基本任务可以用 4 张图谱来概括(图 1.1)。对于遗传图谱、物理图谱和序列图谱的完成已有具体的时间和任务指标(表 1.1)。图 1.2 概要地介绍了进行这几张图谱研究的相互关系,但在本节中仅对 4 张图谱做一概念上的介绍,详细内容将留在以后的相关章节中阐述。

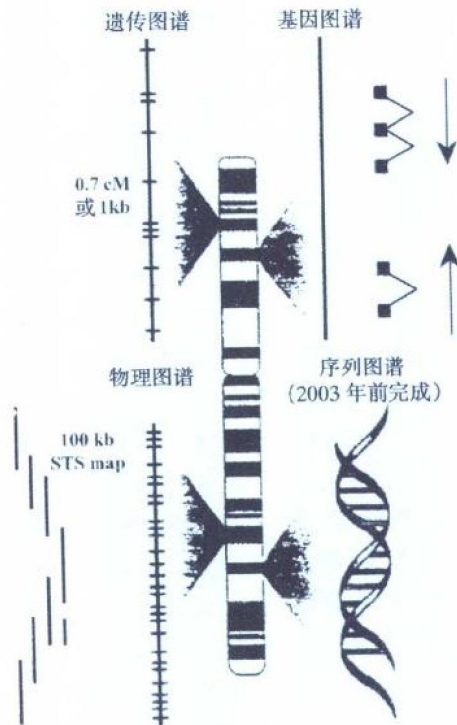


图 1.1 HGP 计划所从事的 4 张图谱
(引自 Science, 1996, 略加修改)

表 1.1 人类基因组计划 1993 ~ 1998 年的目标和截止到 1998 年 10 月全球定量完成目标的情况以及 1998 至 2003 年的新目标

| 研究项目 | 1993 ~ 1998 目标 | 1998 年 10 月完成情况 | 1998 ~ 2003 目标 |
|---------|--|---|---|
| 遗传图谱 | 平均分辨率为 2 至 5cM | 1994 年 9 月发表了 1cM 连锁图 | 已完成 |
| 物理图谱 | 制作 30 000STSs 图 | 完成 52 000STSs 图 | 已完成 |
| 序列图谱 | 1998 年以前完成各类生物的 80Mb 序列 | 人类 180Mb 序列和非人类 111Mb 序列 | 2001 年底前完成人类 1/3 序列;2003 年底以前完成人类全基因组序列 |
| 测序技术 | 重大的改进和革新 | 50 美分一个碱基,达到测 90Mb/年的能力 毛细管阵列电泳启用;显微装配技术可行 | 通过集成化与自动化取得 25 美分一个碱基,每年达到测 500Mb 的能力 支持创新 |
| 人类序列多样性 | 没设目标 | 没设目标 | 制作 100 000SNPs 图 发展技术 |
| 基因图谱 | 发展技术 | 完成 30 000EST 图 | 全长 cDNA |
| 功能分析 | 没设目标 | 没设目标 | 建立基因组为一体的技术 |
| 模式生物 | 大肠杆菌全序列 酵母菌全序列 线虫大部分序列 果蝇测序开始 | 1997 年 7 月发表 1996 年 4 月公开 完成 80% 完成 9% | 1998 年 12 月完成 2002 年完成 建立广泛的基因组资源 2005 年前基本完成全序列 测序和工作框架图 |
| | 小鼠:制作 10 000STS 的图谱 | 完成 12 000STS 的图谱 | |

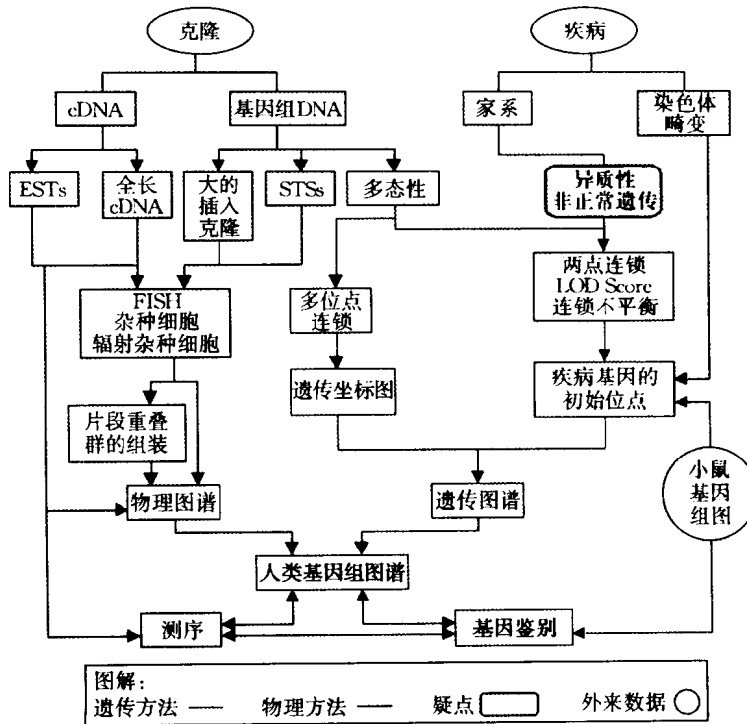


图 1.2 研究人类基因组计划所从事 4 张图谱的常规策略

人类基因组计划的进展是顺利的,由于第1个5年计划的提前完成,导致了1993年对计划的重新修订(Collins F and Galas D,1993)。从最近对任务完成情况的核查结果分析(Collins F *et al.*,1998),目前的5年计划(1993~1998)已成功地完成了全部的主要目标。由于来自私有公司压力越来越强,因此又制定了1998年至2003年新的5年计划的详细目标(表1.1),并对原计划2005年完成测定人类基因组全序列的目标作了进一步的修订,即提前到2003年底前完成这一任务。实际上,在此章刚刚落笔的同时,人类基因组序列图的完成时间又被作了进一步修改(如前所述)。

1.3.1 遗传图谱

遗传图又称连锁图。既然是图,在制图时就应设“标记”,标记越细找到要去的地方就会越容易。在过去的若干年里,标记已有过几次从“粗”到“细”的大的演变。

1.3.1.1 第1代标记

第1代标记是一类经典的遗传标记。最初主要是利用蛋白质或免疫学的标记,如ABO血型位点标记、HLA位点标记等。但由于已知多态的蛋白质很少,等位基因的数目有限且无法获得足够的信息量和检测技术的繁琐等因素,限制了人类基因组的遗传分析工作,这促使人们设法从DNA上寻找标记。

70年代中后期建立起来的限制性片段长度多态性(RFLP)方法被认为是第2代标记或第1代的DNA标记,所以仍习惯上统称为第1代标记。这类标记在整个基因组中确定的位点数目可达 10^5 以上。该系统一经建立就广泛应用到基因组的研究中。其工作原理是用限制性内切酶特异性切割DNA链,由于DNA的一个“点”上的变异所造成的“能切”与“不能切”的两种状况可产生不同长度的片段(即等位片段),再通过凝胶电泳来显示这一长度的“多态性”,并从片段多态性的信息与疾病表型间的关系进行连锁分析,找到致病基因。最成功应用是在发病率较高、病情严重的Huntington舞蹈症基因的定位。虽然,RFLP遍布于整个基因组,但有其局限性,即由于酶切只能产生2到3个片段,所以可提供的信息量有限。并且,有时还需用放射性同位素标记的DNA片段为探针检测RFLP,因而又存在着工作环境和费用等问题。

1.3.1.2 第2代标记

1985年英国Leicester大学的Jeffreys A与其同事在人的肌球蛋白基因中发现了一些短的简单重复单位,他称之为“小卫星中心(minisatellite core)”。随后的许多研究逐渐证明,在人类基因组中分布有大量的此类长度的多态性标记,它们可能以正向(头-尾)或反向(头-头或尾-尾)串联成簇,并分布于基因组的各个部位。在某一位点上,数量可变的串联重复(variable number of tandem repeats, VNTR)可提供不同长度的片段,其重复单位长度为6~12个核苷酸。1989年另一类称作微卫星标记(microsatellite marker)系统被发现和建立,它们的重复单位长度为2~6个核苷酸,有时又被称作“简短串联重复(STR)”。STR的最突出的优点是:高度多态性,并且由于CA等简短重复不受进化上的选择,以致于在同一位点中数目变化很大,在群体中可形成多达几十种的等位片段(多态