

# 人类基因组计划与后基因组时代\*

骆建新<sup>1</sup> 郑崛村<sup>1\*\*</sup> 马用信<sup>2</sup> 张思仲<sup>2</sup>

(1 第三军医大学成都军医学院生物化学与分子生物学教研室 成都 610083)

(2 四川大学华西医学中心附属第一医院医学遗传室 成都 610041)

**摘要** 2003年4月14日生命科学诞生了一个新的重要里程碑,人类基因组计划完成,后基因组时代正式来临。着重介绍了人类基因组计划的提出、目标与任务、实施与进展等方面的基本情况,讨论了后基因组时代的时间界定,分析展望了后基因组时代与人类基因组计划密切相关的生物信息学、功能基因组学、蛋白质组学、药物基因组学等几个重要研究领域。

**关键词** 人类基因组计划 后基因组时代

2003年4月14日,美国人类基因组研究项目首席科学家 Collins F 博士在华盛顿隆重宣布:人类基因组序列图绘制成功,人类基因组计划(human genome project, HGP)的所有目标全部实现。这标志“人类基因组计划”胜利完成和“后基因组时代”(post genome era, PGE)正式来临,在举世庆祝“DNA双螺旋结构”提出50周年之际,生命科学诞生了一个新的里程碑。HGP被誉为可与“曼哈顿原子弹计划”、“阿波罗登月计划”相媲美的伟大系统工程,是人类第一次系统、全面地解读和研究人类遗传物质DNA的全球性合作计划。人类基因组序列图的成功绘制是科学史上最伟大的成就之一,奠定了人类认识自我的重要基石,推动了生命与医学科学的革命性进展。在后基因组时代,生命科学关注的范围越来越大,涉及的问题越来越复杂,采用的技术越来越高,取得的成就将越来越多,生命科学及其相关科学将大有作为。

## 1 人类基因组计划

### 1.1 HGP的提出

HGP的提出有两个重要背景。其一,美国(1945年)在日本广岛和长崎投掷的两颗原子弹导致大量幸存者遭受大剂量核辐射,造成受害者DNA结构严重破坏,基因大量突变。美国能源部

(Department of Energy, DOE)用了30多年时间研究“核辐射对人类基因突变作用”,未取得实质的突破性进展,受害者已表现出明显的突变性状,但检测不出其基因突变与对照组存在显著性差异。1984年12月,受美国能源部和国际预防环境诱变剂和致癌剂委员会的委托,犹他大学的 White R 教授在美国犹他州的阿尔塔组织召开了一个小型学术会议<sup>[1]</sup>。主要目的是研讨寻找有效检测人类基因突变的新方法,分析研究核辐射对人类基因突变的作用。与会者普遍认为解决该问题的最好办法是测定受害者及其后代的全基因组序列,并且,首先测定出人类基因组全序列作为“参考文本”。这是历史上第一次提到“测定人类基因组全序列”。其二,美国于1975年投资巨额启动的“肿瘤十年计划”基本上以失败告终<sup>[2]</sup>。1985年,美国DOE在加利福尼亚州的圣克鲁兹会议上第一次对“测定人类基因组全序列”进行了认真讨论,形成了“人类基因组计划”草案,由 Delisi C 和 Smith D 负责可行性论证。1986年3月,在美国新墨西哥州的圣菲会议上进一步讨论了该计划的可行性;1986年3月7日,著名的诺贝尔奖获得者 Dulbecco<sup>[3]</sup>在 Science 杂志上发表题为“癌症研究的转折点—测定人类基因组序列”的文章,高瞻远瞩地率先向世界公开提出人类基因组计划。Dulbecco 在详细讨论癌症研究进展的基础上阐明了测定人类基因组序列的意义,指出不能继续用“零敲碎打”的方法来了解人类的基因,应当从整体上研究和分析人类基因组及其序列,并倡导全世界有能力的科学家共同完成这一国际性

收稿日期:2003-10-29

\* 四川省杰出青年基金资助项目(03ZQ026 056)

\*\* 通讯作者,电子信箱:juecunz@hotmail.com

大课题。1986年5月,美国能源部HGP负责人Smith D在冷泉港会议上正式宣布了“人类基因组启动计划”。1986年8月,美国国家科学研究委员会(National Research Council, NRC)成立了15人专家小组,负责起草一份人类基因组计划的专题报告。1988年2月,一份题为“人类基因组的作图和测序”的专题报告提交给了美国国会。美国国会能源和商业委员会下属的技术评估办公室(Office of Technology Assessment, OTA)就实施该计划的科学和医学价值、所需经费和资助方式、政府各部门和私人机构之间的协调工作以及国际合作与美国在生物技术上的竞争优势等进行了专题研究,并向美国国会提交了一份专题报告,报告指出:HGP势在必行,而且时机已经成熟。此时,拥有雄厚研究实力的生命科学家团体但从未实施过大科学研究的美国国立卫生研究院(National Institutes of Health, NIH)也开始考虑人类基因组测序的问题,并于1988年10月与DOE签署了一项谅解备忘录,联合资助HGP。1988年11月美国国会批准DOE和NIH共同负责HGP。

HGP的提出引起了全世界的强烈反响,不仅推动了美国,也推动了全世界HGP的发展。1987年初,美国DOE和NIH为“人类基因组计划”下拨了550万美元的启动经费(全年1.66亿美元);1988年9月,美国NIH成立了“人类基因组研究室”(1989年10月改名为“国家人类基因组研究中心”,1997年1月改名为“国家人类基因组研究所”,诺贝尔奖金获得者、“DNA分子双螺旋模型”提出者Watson J任第一任主任。此外,英国(1989年)、法国(1990年)、日本(1990年)、中国(1993年)、德国(1995年)等国家和地区也开始启动了各自的基因组计划。1988年4月,在MaKusick V等科学家的倡导下还成立了国际人类基因组组织(Human Genome Organization, HUGO),主要负责协调各国科学家共同完成HGP。

HGP的雄心大、规模大、难度大、花钱多,社会大众、政府部门、科学家及其它各界人士都存在不少反对意见,人类基因组计划引起了激烈的大讨论、大辩论<sup>[4]</sup>。经过几年的大讨论、大辩论,人类基因组计划不断成熟、完善。1990年10月1日美国国会正式批准启动人类基因组计划,计划投入30亿美元的资金在15年内完成人类基因组的分析研究,全世界免费共享所有研究成果。

## 1.2 HGP目标与任务

1990年,HGP正式启动时提出了“通过国际合作,在15年内(1990~2005年)投入30亿美元,完成人类全部24条(22条常染色体和X、Y性染色体)染色体的核苷酸( $3 \times 10^9$  bp)序列分析,构建详细的人类基因组遗传图谱和物理图谱,确定人类DNA的全部核苷酸序列,定位约10万个基因,并对其它生物进行类似研究”的总体目标,并制定了具体的五年计划(1990~1995年)。1993年,HGP在广泛征求意见的基础上对计划做了适当修订,制定了第二个五年计划(1993~1998年)<sup>[5]</sup>,包括人类基因组图谱的构建与序列分析;基因的鉴定;基因组研究技术的建立、创新与改进;模式生物基因组的作图和测序;信息系统的建立、信息的储存、处理及相应软件的开发;人类基因组计划有关的伦理、法律和社会问题的研究;研究人员的培训、技术转让及产业开发;研究计划的外延9项内容。对原计划作了4个方面的调整:(1)从序列分析到基因组制图,即在大规模测序之前先作图(遗传图谱和物理图谱)并构建供测序的克隆片段重叠,即相互重叠并覆盖整个基因组的DNA克隆片段;(2)从序列分析和制图到基因鉴定,即在全基因组分析全盘考虑的同时,分离和鉴定具有重要功能、与重要疾病相关的基因;(3)从人类“基因组”计划到人类“cDNA计划”,将“cDNA”计划作为人类基因组整体计划的一个组成部分;(4)从人类基因组到模式生物基因组,重点研究大肠杆菌(4Mb)、酵母(15Mb)、果蝇(120Mb)、小鼠(15Mb)等模式生物,建立新技术,积累经验,为人类基因的功能鉴定提供线索。1998年,HGP制定第三个五年计划(1998~2003年)<sup>[6]</sup>,对HGP内容再次作了调整,主要目标是:(1)得到标记间距为1厘摩的遗传图谱;(2)得到至少有30万个序列标记位点(STS)的物理图谱;(3)2001年得到人类基因组序列的“草稿”,2003年得到“定稿”;(4)测序能力要达到每年500Mb,每个碱基对的的分析费用要少于25美分,支持毛细管阵列电泳、DNA芯片等的测序技术的发展;(5)增加测定人类基因组变异的内容,得到10万个作图定位了的单核苷酸多态性(SNP);(6)得到所有基因的全长cDNA;(7)发展在基因组水平上分析生物功能的技术;(8)完成重点模式生物的全基因组序列。

## 1.3 HGP实施与进展

### 1.3.1 研究策略 人类基因组全序列分析分两大

步骤: 即制图(mapping)和测序(sequencing), 全过程分为 4 个阶段(可交叉进行): (1) 构建 1cM 的遗传图; (2) 构建物理图; (3) 建立重叠克隆系; (4) 完成核苷酸顺序测定。HGP 最终将绘制出 4 张图谱, 即得到遗传图谱、物理图谱、序列图谱和转录图谱, 从分子水平上揭示出人体的奥秘。

**遗传图谱和物理图谱:** 遗传图谱(genetic map)也称连锁图谱(linkage map), 是指通过测量不同性状一起遗传(连锁)的频率而建立的反映基因遗传效应的图谱, 即通过计算连锁的遗传标志的重组率, 以重组率的大小反映遗传标志间的相对距离绘制遗传图谱(1 厘摩表示重组率为 1%)。HGP 利用高度多态性的遗传标记, 如简单串联重复序列(short tandem repeat, STR)和单核苷酸多态性(single nucleotide polymorphism, SNP)标志构建遗传图谱; 物理图谱(physical map)是通过测定遗传标志的排列顺序与位置而绘制成的, 包括限制性酶切图谱、DNA 克隆片段重叠群图、序列标签图以及表达基因的特征性序列标记图等。HGP 在整个基因组染色体每隔一定距离标上序列标签位点(sequence tagged site, STS)之后, 随机将每条染色体酶切为大小不等的 DNA 片段, 以酵母人工染色体(yeast artificial chromosome, YAC)或细菌人工染色体(bacterial artificial chromosome, BAC)等作为“载体”构建 YAC 或 BAC 邻接克隆系, 确定相邻 STS 间的物理联系, 绘制以 Mb、kb、bp 为图距的人类全基因组物理图谱。

**序列图谱(sequence map)与转录图谱(transcriptional map):** 序列图谱是人类基因组在分子水平上最高层次、最为详尽的物理图。测定 30 亿对核苷酸组成的基因组全部 DNA 序列是基因组计划中最为明确、最为艰巨的定时、定量、定质的硬任务。在遗传图谱和物理图谱基础上, 精细分析各克隆的物理图谱, 将其切割成易于操作的小片段, 构建 YAC 或 BAC 文库, 得到 DNA 测序模板, 测序得到各片段的碱基序列, 再根据重叠的核苷酸顺序将已测定序列依次排列, 获得人类全基因组的序列图谱; 转录图谱又称表达图谱(expression map), 是一种根据组织细胞中可表达片段标签(expressed sequence tags, EST)绘制的图谱, 是将 mRNA 逆转录合成的 cDNA 或 EST 的部分 cDNA 片段作为“探针”与基因组 DNA 进行分子杂交, 标记转录基因, 绘制出可表达基因转录图, 最终绘制出人体所有组织、

所有细胞以及所有发育阶段的全基因组转录图谱。

HGP 在制图的基础上测序, 最后获得 4 张图谱, 核心是获得高质量的基因组序列图。首先力争获得覆盖人类全基因组序列 90%、精确率为 95% 的“工作草图”; 在此基础上, 查漏补缺, 获得覆盖率 99%、精确率 99.99% 的“精细图”; 最后获得覆盖率 100%、精确率 99.99% 的“完成图”。

**1.3.2 国际合作与私立公司竞争** HGP 的最大特点是“全球化”。整个人类基因组计划主要由美国、英国、日本、法国、德国和中国 6 个国家(20 个测序中心)的 1100 名生物科学家、计算机专家在 HUGO 的统一协调下“精诚合作、共享材料、共享数据、共同攻关”完成的。美国有 12 家研究中心参与 HGP, 占有国家研究的国家贡献率为 54%, 主要负责 2、3、4、5、7、8、9、10、11、12、15、16、17、18、19、X 和 Y 染色体测序分析; 英国于 1989 年 2 月开始实施 HGP, Sanger 研究中心(世界最大的测序中心)负责测序工作, 其国家贡献率为 33%, 主要负责 1、6、9、10、11、13、20、22 和 X 染色体的测序分析; 日本于 1990 年 2 月开始 HGP, 有 RIKEN 等 2 家研究中心参与, 国家贡献率为 7%, 主要负责 2、6、8、11、18、21 和 22 号染色体的测序分析; 法国的 HGP 于 1990 年 6 月开始, Genoscope 研究中心参与, 国家贡献率为 2.8%, 主要负责 14 号染色体的分析排序; 德国的 HGP 于 1995 年 5 月开始, 有 IMB 等 3 家研究中心参与, 国家贡献率为 2.2%, 主要负责 8、9、17、21 和 X 染色体的部分测序分析; 中国的 HGP 是在国家自然科学基金委员会、国家高技术研究发展计划(863 计划)等共同资助下于 1993 年开始, 1999 年 7 月注册加入国际人类基因组计划, 1999 年 10 月 1 日正式启动 HGP, 北京华大研究中心、国家南方基因研究中心等 3 家测序中心(1 家注册)参与, 国家贡献率为 1%, 主要负责人类 3 号染色体短臂从 D3S3610 至端粒的 30Mb 区域上 3000 万个碱基对序列的测定分析。

此外, 加拿大、丹麦、以色列、瑞典、芬兰、挪威、澳大利亚、新加坡、原苏联及原东德等也都开始了不同规模、各有特色的人类基因组研究; 印度、巴西、墨西哥、智利、肯尼亚等国也以不同的方式参与这一全球范围的合作与竞争。

1998 年 5 月 11 日, 世界上最大的 DNA 自动测序仪生产商美国 PE Biosystems 公司与 Venter C 教授共同组建了赛莱拉(Celera)私立公司, 公开与政

府投资的 HGP 展开竞争, 宣称 3 年内投资 3 亿美元, 以“全基因组鸟枪法”(Whole genome shotgun, WGS)<sup>[7]</sup> 完成人类基因组测序, 专利 200~400 个重要基因, 并将所有序列信息保密 3 个月。该公司号称拥有自己研制的 300 台最新毛细管自动测序仪 (ABI 3700)、“全球第三”的超大型计算机和超过全球所有序列组装解读力量总和的研究实力。Celera 公司参与竞争大大加速了 HGP 的进程, 该公司运用的自动操作和测序技术与 HGP 基本相似, 但测序的方法不同, Celera 公司运用全基因组鸟枪法, 而 HGP 运用“分级鸟枪法”(Hierarchical shotgun, HS)<sup>[8]</sup>, 即基于 BAC 连续克隆系的测序方法。两种方法各有优缺点, 分级鸟枪法具有每条单独序列是确定已知的优势, 但需要构建基因组的大片段的 YAC 或 BAC 文库, 而全基因组鸟枪法不需要建立文库, 直接将染色体 DNA 随机拆分为不同大小的片段, 最后将每个片段的序列拼接起来得到全基因组序列, 但该方法的计算处理难度大, 精确度相对较差<sup>[9]</sup>。

1.3.3 研究进展 1987 年, 美国马萨诸塞人类基因合作组的研究人员以限制性片段多态性 (RFLP) 为遗传标志绘制了第一代人类遗传图谱<sup>[10]</sup>, 为 HGP 基因组作图奠定了基础。1992 年, 法国 Genethon 实验室绘制出以 STR 为遗传标志的第二代人类遗传图谱<sup>[11]</sup>, 并于 1994 年得到更加完善的遗传图谱<sup>[12]</sup> (提前 1 年完成人类遗传图谱的绘制目标), 1996 年得到完整的人类基因组遗传图谱<sup>[13]</sup>。2000 年, 美国 Whitehead Institute/MIT 中心得到以 SNP 为遗传标志的第三代人类基因组遗传图谱<sup>[14]</sup>。1995 年, 美国 Whitehead Institute/MIT 绘制出 STS 为物理标记的第一张物理图谱<sup>[15]</sup>, 该图包括 15000 多个标记位点, 分辨率为 200kb, 覆盖了人类基因组 94% 的区域。1997 年, 美国斯坦福人类基因组中心以 YAC 为载体绘制出 STS 为标志的物理图谱<sup>[16]</sup>。1998 年, 英国 SANGER 等中心联合绘制出包含 30 000 个 STS 标签、3 万个基因的物理图谱<sup>[17]</sup>, 几乎覆盖了人类全基因组序列 (精确度提高 2~3 倍)。2001 年美国斯坦福人类基因组中心进一步利用辐射杂种细胞系技术 (radiation hybrid, RH), 以 BAC 为载体绘制出高精度的物理图谱<sup>[18]</sup>。

1991 年, Adams 等<sup>[19]</sup>建立了研究组织细胞中基因表达的 EST 法, Sudo 等<sup>[20]</sup> (1994 年) 从人胚肺的

cDNA 文库中得到 2058 个 EST 标签, Touchman 等<sup>[21]</sup> (1997 年) 得到人体 7 号染色体 2006 个 EST 标签; 1995 年, Velculescu 等<sup>[22]</sup>建立了基因表达的系列分析新方法 (serial analysis of gene expression, SAGE), Lal 等<sup>[23]</sup> (1999 年) 从 25 个不同的肿瘤或正常组织的 SAGE 文库中得到 10 万个转录标签; 2000 年, Dias 等<sup>[24]</sup>建立了开放阅读框 EST 法 (open reading frame expressed sequences tags, ORESTES) 绘制转录图谱, 并于 2001 年得到 70 万个 ORESTES 标签的转录图谱<sup>[25]</sup>。在 HGP 的前半段时间, 人类全基因组序列的测序工作进展比较缓慢, 1997 年 HGP 仅完成 5% 的测序任务。1998 年 5 月 Celera 公司的竞争参与加速了 HGP 的进程, 1999 年 12 月, 英、美、日等国的研究小组共同完成了人类第 22 号染色体基因全序列<sup>[26]</sup> (首次破译完整的人类染色体全基因全序列); 2000 年 4 月 13 日, 美国 JGI 中心完成第 5、第 16 和第 19 号染色体的遗传密码草图; 2000 年 5 月, 德、日两国共同破译了人类第 21 号染色体基因全序列<sup>[27]</sup>; 2000 年 6 月 26 日, 国际人类基因组测序联盟与 Celera 公司联合发布了“人类基因组工作草图” (work draft); 2001 年 2 月 12 日又分别在 Nature 和 Science 杂志上公布了人类基因组“精确图” (准确度达到 99.99%) 及初步分析结果<sup>[28, 29]</sup>, 两组数据存在一定差异, 但大部分高度吻合<sup>[30, 31]</sup>, 发现人类基因数目约为 3.4 万至 3.5 万个 (仅比果蝇多 2 万个, 远小于原来估计的 10 万个基因); 2003 年 4 月 14 日 Collins F 博士在华盛顿隆重宣布 HGP 完成, 得到了人类基因组“完成图” (包括 99% 的人类基因组序列, 准确度为 99.99%)。迄今为止, Nature 杂志发表了 6 条 (7、14、20、21、22 和 Y) 染色体序列的完成情况及基本信息<sup>[26, 27, 32~35]</sup>, 其余染色体将陆续发表。

中国作为 HGP 第六个成员国 (唯一的发展中国家) 于 1999 年 10 月 1 日正式启动 HGP, 2000 年 4 月, 完成了 1% 人类基因组“工作框架图”; 2001 年 8 月 26 日绘制完成“中国卷”, 提前 2 年获得精确度达 99.99% 的“完成图”序列。所有 BAC 序列都经过指纹图谱的验证, 共测定 31.6 Mb 的序列, 识别 122 个基因, 其中 86 个是已知基因 (55 个为功能明确的基因, 8 个为疾病相关基因), 在 31 个基因中找到了 75 种不同的剪切方式, 发现了 1760 个新的 SNP, 分析了“完成图”中重复序列、CpG 岛和 GC 含量。

除四张图谱以外, HGP 在生物信息科学、数据处理、知识产权及社会伦理学研究、特别是基因专利申请、基因诊断、基因治疗对保险、就业影响等多方面都取得了较大的进展。

#### 1.4 HGP 精神

HGP 给人类带来极大物质利益的同时, 也给人类带来社会文化高度文明的精神享受。HGP 第一精神是“协作”精神, HGP 是人类历史上第一次由全世界科学“精诚合作、共享材料、共享数据、共同攻关”完成的史无前例的国际性大课题, 比“曼哈顿原子弹计划”和“阿波罗登月计划”的“协作性”更强, 它代表一种进步文化, 将成为未来国际大课题的楷模。其次是 HGP 对社会的“高度负责”精神, HGP 十分重视人类基因组的研究对社会、法律、伦理等问题的影响, 在整个人类基因组计划的实施过程中充分体现了“求真”、“求善”的社会高度责任感。

## 2 后基因组时代

### 2.1 时间界定

“后基因组时代”到底从何时算起至今尚无定论, “基因组时代”与“后基因组时代”就其研究手段、研究内容本身存在着交叉、重叠, 没有严格的界限。有人将“新千年”公布人类基因组“工作草图”的 2000 年 6 月 26 日作为后基因组时代到来的标志, 有人将“新世纪”国际人类基因组测序联盟与 Celera 公司分别在 Nature 和 Science 杂志上公布了人类基因组“精确图”及分析数据的 2001 年 2 月 12 日作为后基因组时代到来的标志。目前, 普遍将 Collins F 宣布“人类基因组序列图绘制成功, 人类基因组计划的所有目标全部实现”的 2003 年 4 月 14 日作为后基因组时代正式来临的标志。其理由有三: 其一, 2003 年 4 月 HGP 任务全部完成。其二, 1990 年正式启动 HGP 时指出“2005 年完成人类基因组计划时, 生命科学便进入后基因组时代”。其三, 2003 年正好是“DNA 双螺旋结构”提出 50 周年。

### 2.2 重要研究领域

2.2.1 生物信息学 生物信息学(Bioinformatics)是应用计算机技术研究生物信息的一门新生学科, 是生物学、数学、物理学、计算机科学等众多学科交叉的新兴学科。将生物遗传密码与电脑信息相结合, 通过各种程序软件计算、分析核酸、蛋白质等生物大分子的序列, 揭示遗传信息, 并通过查询、搜

索、比较、分析生物信息, 理解生物大分子信息的生物学意义。生物信息学为分子生物学家提供了一条寻找和研究新基因的新思路, 在 HGP 实施过程中起着非常重要的作用, 从高度自动化的大规模测序、数据的获取与分析处理、序列片段的拼接、新基因的发现、基因功能的预测直到基因的分子进化等研究的各个环节都有生物信息学的功劳。著名科学家 Crick F (“DNA 双螺旋结构”提出者) 指出: 计算机技术的进展使基因组的大规模测序成为可能, 推动生物学成为一门“大科学”。离开生物信息学要完成 HGP 是难以想象的, 后基因组时代没有生物信息学生命科学也是不可思议的。完成 HGP、获得生命“天书”之后, 人类亟需破译基因组所蕴涵的功能信息, 解码生命。在后基因组时代生物信息学将更加举足轻重, 要读懂“天书”, 仅仅依靠传统的实验观察手段无济于事, 必须借助高性能计算机和高效数据处理的算法语言, 只有如此, “天书”才能发挥它应有的价值。生命科学的革命性巨变已把生物信息学推到了前台, 生物信息技术已成为后基因组时代的核心技术之一, 在“蛋白质组学”、“功能基因组学”、“药物基因组学”等领域必将更有用武之地, 必将对生命科学(尤其是医学)的发展产生无法估计的巨大影响, 生物信息学在后基因组时代将有长足发展。

2.2.2 功能基因组学 功能基因组学(Functional Genomics)是指在全基因组序列测定的基础上, 从整体水平研究基因及其产物在不同时间、空间、条件的结构与功能关系及活动规律的学科。功能基因的研究是后基因组时代的关键点, 疾病基因组学研究已成为后基因组时代的主旋律<sup>[36]</sup>。获得基因组的结构信息只是认识基因组的第一步, 弄清基因相应的功能及实际应用才是关键所在, 没有基因的功能分析, 基因的序列不过是 4 种核苷酸的排列组合, 只有了解基因的功能, “天书”才真正有意义。根据基因的 mRNA 表达水平绘制每种组织细胞中的基因表达谱、细胞不同发育阶段的基因表达谱、正常和病理状态下的基因表达谱、治疗条件下的基因表达谱是功能基因组学的重要基础。HGP 在基因表达谱方面已取得一定进展<sup>[25]</sup>, 但人类基因有 90% 的功能尚不明确, 功能基因组学将借助生物信息学的技术平台, 利用先进的基因表达技术及庞大的生物功能检测体系, 从浩瀚无垠的基因库筛选并确知某一特定基因的功能, 并通过比较分析基因及

其表达的状态, 确定出基因的功能内涵, 揭示生命奥秘, 甚至开发出基因产品。功能基因组学在后基因组时代占有重要位置, 其研究成果直接给人类健康带来福音。功能基因组学在后基因组时代将取得丰硕成果, 人们的生活质量将大幅度提高。

**2.2.3 蛋白质组学** 蛋白质组是指一个基因组、一种生物或一种组织细胞所表达的全套蛋白质, 蛋白质组学( Proteomics) 是以蛋白质组为研究对象的新的研究领域, 主要研究细胞内蛋白质的组成及其活动规律, 建立完整的蛋白质文库。蛋白质组学分为表达蛋白质组学( 建立细胞、组织中蛋白定量表达图谱, 或扫描 EST 图) 和细胞图谱蛋白质组学( 确定蛋白质在亚细胞结构中的位置, 纯化细胞器或用质谱仪鉴定蛋白复合物组成, 确定蛋白质- 蛋白质的相互作用)。HGP 已经确定人类 3 万多个基因在 23 对染色体上的位置及其碱基排列顺序, 后基因组时代科学家将盘点人类蛋白质组里所有蛋白质, 研究其生理功能。基因研究是 20 世纪生命科学的主线, HGP 使其达到“登峰造极”, HGP 的实现为未来生命科学研究奠定了坚实的基础, 但基因的重要作用最终需要蛋白质来体现, 蛋白质组学与基因组学同等重要, 甚至更为重要。进入 21 世纪, 蛋白质组学的基础与应用研究正以指数增长, 蛋白质研究规模与深度到了前所未有的程度, 生命科学已从核酸时代回归蛋白质时代, 对生命奥秘的探索由基因、核酸层次深入到蛋白质层次, 蛋白质组学将成为后基因组时代的重要支柱之一, 其发展不可限量, 蛋白质组学的深入研究将带来巨大的经济效益。

目前, 美国、英国、日本等国家纷纷投入巨资资助蛋白质组研究。英国科学家欲绘制人类蛋白质图谱<sup>[13]</sup>, 试图找到与人类疾病有关的蛋白质互动模式; 美国 DOE 在 HGP 之后又启动了“生命基因组计划”( the genomes to life program)<sup>[37]</sup>, 其主要目的在于“识别执行关键生命功能的多蛋白质复合物, 分析其基因调控网络特征”; Celera 公司首席科学家 Venter C 指出: 人类基因组、蛋白质组和药物是生命科学研究的三个阶段, Celera 公司现已进入蛋白质组研究的第二阶段。公司增添大批蛋白质鉴别和分析设备用于蛋白质组图谱的绘制。HGP 首席科学家 Collins F 指出: 有关人类蛋白质组的研究将是艰巨的, 目前仅仅是迈出第一步, 蛋白质组技术还远远落后于破译基因组的手段, 后基因组时代的

蛋白质组研究将大有作为, 蛋白质组研究将带来药物开发的实质性突破, 研制出治疗包括癌症和艾滋病等在内的多种疾病的药物。

**2.2.4 药物基因组学** 尽管人类的基因 99.99% 是相同的, 但在药物作用机制、药物代谢转化、药物毒副作用等方面都存在着个体差异。药物基因组学( Pharmacogenomics) 以提高药物疗效与安全性为目的, 研究影响药物作用, 药物吸收、转运、代谢、清除等过程中基因差异, 通过对疾病相关基因、药物作用靶点、药物代谢酶谱、药物转运蛋白基因多态性等方面研究, 寻找新的药物先导物和新的给药方式, 并指导临床用药。药物发现经历了从自然界发现药物、随机筛选发现药物到以机制和靶结构为基础的新药发现和开发过程, 但未在根本上从基因分子水平了解疾病发病的实质以及药物与基因组之间的相互作用关系, 长期缺乏有效治疗诸如遗传病、肿瘤等疾病的药物。HGP 使药物的研究和开发进入了一个新的阶段, 尽管药物基因组学还是一门十分年轻的学科, 与医疗实践应用还有相当一段距离, 但人们对其前景充满信心。药物作用靶点与发病机制的关系、以基因为基础设计新药( 如反义药物)、单核苷酸多态性与疾病及药物敏感性的关系等方面的成就表明药物基因组学的价值已初露端倪。HGP 的完成还掀起新一轮的基因热潮和生物科技竞赛, 各大跨国制药公司看好基因药物市场, 已投入大量资金, 利用基因研究开发新药物, 抢占基因药物市场。随着药物基因组学的发展成熟和新基因的更多发现, 将会有更多的新药出现, 生物医药产业必将成为未来经济的支柱产业之一。

**2.2.5 HGP 的伦理、法律和社会问题** 基因组研究是一把双刃剑, 给人类带来福音的同时也可能带来灾难。HGP 一开始就非常注重与基因组研究相关的伦理、法律和社会问题( ethical, legal, and social issues, ELSI) 的研究。ELSI 是 HGP 的重要组成部分, 基因信息利用的公平性问题, 隐私和保密问题, 个体基因差异而引起的心理影响和伤害问题, 遗传检测和人口普查涉及的问题, 生殖问题, 处理或预防基因缺陷的基因治疗问题, 基因改进问题, 临床质量控制的标准和标准的执行问题, 知识产权和数据、资料的利用问题, 自由意志和基因决定论、疾病和健康等问题都在 HGP 之列。1988 年 9 月, 美国 NIH 的人类基因组研究室第一任主任 Waston J 宣布 HGP 预算经费的 3% 用于 ELSI 研究, 1989 年 1

月,人类基因组顾问委员会成立了 ELSI 工作组,1990 年确定了 ELSI 的目标与任务,提出了具体的研究方案。1997 年 7 月,美国 DOE 和 NIH 还成立了“ELSI 计划与评价工作组”(2000 年 2 月被“ELSI 研究顾问”取代),负责评述与分析 ELSI 研究方案,起草 5 年计划。HUGO 也多次召开国际会议专题讨论 ELSI,协调世界各国的 ELSI 研究。2000 年,ELSI 计划与评价工作组报告了 HGP 的 ELSI 研究进展。目前,ELSI 研究取得较大进展,完成了预期的目标与任务<sup>[38~40]</sup>。然而,ELSI 涉及面广、问题复杂、影响深远,不仅是自然科学问题,更是社会科学问题,它是一个涉及政治、经济、法律、伦理道德、社会教育与心理等多领域的庞大复杂的科学体系,已取得的 ELSI 研究成果仅仅解决了其中部分问题,甚至是极少部分,还有大量问题需要研究解决。后基因组时代 ELSI 显得更为重要,必须高度重视、大力加强 ELSI 研究。

### 3 结束语

HGP 对生命科学的发展产生了深远影响,HGP 的完成为人类进一步认识自我、探索生命奥秘、增进健康、提高生活质量奠定了坚实基础。但是,人类基因组全序列仍只是一部遗传结构的“天书”,目前还只能看懂其中的小部分,还有很多艰巨工作要做。HGP 完成的人类全基因序列是 99%,而不是 100%,准确度约为 99.99%,也不是 100%。在现有的技术条件下,人类全基因序列中还有 1% 无法完成,还有待于新技术、新理论的建立。基因组序列测定完成是破解人类遗传奥秘的历史性开端,而不是结束,后基因组时代生命科学将大有作为。

中国是世界上生物资源最为丰富的国家之一,也是基因资源最为丰富的国家之一,人类基因组计划需要中国,后基因组时代也同样需要中国。中国以 HGP 的“1% 任务”跻身于世界生命科技竞争行列,在后基因组时代中国将更加努力,国家已投入 6 亿元人民币开展重大疾病、重要生理功能相关功能基因、中华民族单核苷酸多态性的开发应用、与人类重大疾病及重要生理功能相关的蛋白质、重要病原真菌功能基因组等方面的研究与开发,进一步完善我国生物技术创新体系。中国应在大规模测序能力的基础上,结合自身特点和优势,开拓创新、形成特色,力争在激烈的国际竞争中进入世界先进行列,在生命科学的一些重要领域占据一席之地。

### 参考文献

- [1] Cook Deegan R M. The Alta summit, December 1984. *Genomics*, 1989, 5(3): 661~ 663
- [2] Geelhoed G W. The message in the bottle: a value predicament in cancer research. *Del Med J*, 1985, 57(2): 95~ 98
- [3] Dulbecco R. A turning point in cancer research: sequencing the human genome. *Science*, 1986, 231(4742): 1055~ 1056
- [4] Roberts L. The human genome. Controversial from the start. *Science*, 2001, 291(5507): 1182~ 1188
- [5] Collins F, Galas D. A new five year plan for the U. S. Human Genome Project. *Science*, 1993, 262(5130): 43~ 46
- [6] Collins F, Patrinos A, Jordan E, et al. New Goals for the U. S. Human Genome Project: 1998 ~ 2003. *Science*, 1998, 282(5389): 682~ 689
- [7] Weber J L, Myers E W. Human Whole Genome Shotgun Sequencing. *Genome Res*, 1997, 7(5): 401~ 409
- [8] Green P. Against a whole genome shotgun. *Genome Res*, 1997, 7(5): 410~ 417
- [9] Waterston R H, Lander E S, Sulston J E. On the sequencing of the human genome. *Proc Natl Acad Sci U S A*, 2002, 99(6): 3712~ 3716
- [10] Donis Keller H, Green P, Helms C, et al. A genetic linkage map of the human genome. *Cell*, 1987, 51(2): 319~ 337
- [11] Weissbach J, Gyapay G, Dib C, et al. A second-generation linkage map of the human genome. *Nature*, 1992, 359(6398): 794 ~ 801
- [12] Gyapay G, Morissette J, Vignal A, et al. The 1993~ 94 Genethon human genetic linkage map. *Nat Genet*, 1994, 7(2): 246~ 339
- [13] Dib C, Faure S, Fizames C, et al. A comprehensive genetic map of the human genome based on 5, 264 microsatellites. *Nature*, 1996, 380(6570): 152~ 154
- [14] Altshuler D, Pollara V J, Cowles C R, et al. An SNP map of the human genome generated by reduced representation shotgun sequencing. *Nature*, 2000, 407(6803): 513~ 516
- [15] Hudson T J, Stein L D, Gerety S S, et al. An STS based map of the human genome. *Science*, 1995, 270(5244): 1945~ 1954
- [16] Stewart E A, McKusick K B, Aggarwal A, et al. An STS-based radiation hybrid map of the human genome. *Genome Res*, 1997, 7(5): 422~ 433
- [17] Deloukas P, Schuler G D, Gyapay G, et al. A physical map of 30, 000 human genes. *Science*, 1998, 282(5389): 744~ 746
- [18] Olivier M, Aggarwal A, Allen J, et al. A high resolution radiation hybrid map of the human genome draft sequence. *Science*, 2001, 291(5507): 1298~ 1302
- [19] Adams M D, Kelley J M, Gocayne J D, et al. Complementary DNA sequencing: expressed sequence tags and human genome project. *Science*, 1991, 252(5013): 1651~ 1656
- [20] Sudo K, Chinen K, Nakamura Y. 2058 expressed sequence tags (ESTs) from a human fetal lung cDNA library. *Genomics*, 1994, 24(2): 276~ 279
- [21] Touchman J W, Bouffard G G, Weintraub L A, et al. 2006 expressed sequence tags derived from human chromosome 7

- enriched cDNA libraries. *Genome Res*, 1997, 7(3): 281~ 292
- [22] Velculescu V E, Zhang L, Vogelstein B, et al. Serial analysis of gene expression. *Science*, 1995, 270(5235): 484~ 487
- [23] Lal A, Lash A E, Altschul S F, et al. A public database for gene expression in human cancers. *Cancer Res*, 1999, 59(21): 5403~ 5407
- [24] Dias Neto E, Correa R G, Verjovski-Almeida S, et al. Shotgun sequencing of the human transcriptome with ORF expressed sequence tags. *Proc Natl Acad Sci U S A*, 2000, 97(7): 3491~ 3496
- [25] Camargo A A, Samaia H P, Dias Neto E, et al. The contribution of 700,000 ORF sequence tags to the definition of the human transcriptome. *Proc Natl Acad Sci U S A*, 2001, 98(21): 12103~ 12108
- [26] Dunham I, Shimizu N, Roe B A, et al. The DNA sequence of human chromosome 22. *Nature*, 1999, 402(6761): 489~ 495
- [27] Hattori M, Fujiyama A, Taylor T D, et al. The DNA sequence of human chromosome 21. *Nature*, 2000, 405(6784): 311~ 319
- [28] Lander E S, Linton L M, Birren B, et al. Initial sequencing and analysis of the human genome. *Nature*, 2001, 409(6822): 860~ 921
- [29] Venter J C, Adams M D, Myers E W, et al. The sequence of the human genome. *Science*, 2001, 291(5507): 1304~ 1351.
- [30] Aach J, Bulyk M L, Church G M, et al. Computational comparison of two draft sequences of the human genome. *Nature*, 2001, 409(6822): 856~ 859
- [31] Li S, Cutler G, Liu J J, et al. A comparative analysis of HGSC and Celera human genome assemblies and gene sets. *Bioinformatics*, 2003, 19(13): 1597~ 1605
- [32] Deloukas P, Matthews L H, Ashurst J, et al. The DNA sequence and comparative analysis of human chromosome 20. *Nature*, 2001, 414(6866): 865~ 871
- [33] Hillier L W, Fulton R S, Fulton L A, et al. The DNA sequence of human chromosome 7. *Nature*, 2003, 424(6945): 157~ 164
- [34] Heilig R, Eckenberg R, Petit J L, et al. The DNA sequence and analysis of human chromosome 14. *Nature*, 2003, 421(6923): 601~ 607
- [35] Skaletsky H, Kuroda Kawaguchi T, Minx P J, et al. The male specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature*, 2003, 423(6942): 825~ 837
- [36] Butcher S P. Target discovery and validation in the post-genomic era. *Neurochem Res*, 2003, 28(2): 367~ 371
- [37] Frazier M E, Johnson G M, Thomassen D G, et al. Realizing the potential of the genome revolution: the genomes to life program. *Science*, 2003, 300(5617): 290~ 293
- [38] Zneimer S M. The human genome project: exploring its progress and successes and the ethical, legal, and social implications. *Clin Leadersh Manag Rev*, 2002, 16(3): 151~ 157
- [39] Mowat D. Ethical, legal and social issues surrounding the human genome project. *Intern Med J*, 2002, 32(3): 89~ 90
- [40] Clayton E W. Ethical, legal, and social implications of genomic medicine. *N Engl J Med*, 2003, 349(6): 562~ 569

## Human Genome Project and the Post Genome Era

Luo Jianxin<sup>1</sup> Zheng Juecun<sup>1</sup> Zhang Sizhong<sup>2</sup>

(1 Department of Biochemistry and Molecular Biology, Chengdu Medical College, Third Military Medical University Chengdu 610083)

(2 Department of Medical Genetics, The First Hospital, West China Medical Center, Sichuan University Chengdu 610041)

**Abstract** A new great milestone of life science was born on 14 April 2003, which enunciated the human genome project (HGP) had been completed and life science is entering a new period named the post genome era (PGE). The propose, target, mission, execution and progress of the HGP are summarized respectively, the time mark of PGE, the prospect of important domain related to the human genome project in the post genome era, such as bioinformatics, functional genomics, proteomics, pharmacogenomics, and ethical, legal, and social issues are discussed and depicted.

**Key words** Human genome project Post genome era