

生命科学的春天^{*}

单祥年

(东南大学医学院, 南京, 210009)

[摘要] 由于中、美、英、法、日、德等六国科学家的共同努力, 启动于 1990 年的人类基因组计划(HGP)在 2000 年 6 月完成了遗传图、物理图、序列图(工作框架), 并于 2001 年 2 月 12 日公布了对人类的基因组(由 31.674 亿个碱基对组成, 含有 3 万到 4 万个结构基因)的分析。这表明了人类结构基因组计划已初步完成, 并启动了后基因组(也即功能基因组)计划, 以揭示人类的出生、年龄、疾病和死亡的秘密。

首先, HGP 促进了医学领域的进步。基因识别、基因治疗、基因药品以及药物基因组学渗透到了医学的每个角落。通过分子生物学诊断和治疗人类疾病将成为普通的临床实践, 人类主宰自己命运的梦想将成为现实, 人们将健康而长寿的生活。

比较基因组学是另一个得益于 HGP 的平稳运行和成功的领域。迄今为止已对 200 多种生物物种绘制了精度不同的遗传图和物理图, 其中一些已经完成了整个基因序列和遗传密码翻译。特别是与人类生活联系紧密的经济类动植物。动植物基因项目的完成将为高质高产地培养繁殖物种提供分子基础和方法。同时, 随着比较基因组学的进步, 一些关系到生物分类和进化的疑点将得以解决。

生命科学所迈出的科技发展的第一步已庄严地展现在人们面前。通过生物工业使我们的国家繁荣昌盛是人们的共同目标。21 世纪是生命科学的世纪, 生命科学的春天将鼓舞每一个从事生命科学研究的科学工作者奋勇向前, 充分施展自己的才华, 迎接未来的挑战。

[关键词] HGP; 后基因组计划; 药物基因组学; 比较基因组学

[中图分类号] Q343.1; [文献标识码] A; [文章编号] 1008-1925(2001)02-0001-11

生命在地球上出现已近 40 亿年, 人类从诞生时就已开始对生命的认识和改造, 但是人类真正有能力认识生命奥秘的真谛是 20 世纪 70 年代出现的 DNA 重组技术和基因扩增技术, 这些技术标志着人类对生命科学研究跨入了一个新的时代。特别是 90 年代启动的人类基因组计划(Human Genome Project, HGP)和在该计划带动和催化下出现的植物基因组计划、动物基因组计划, 这些计划标志着人类已有能力从分子水平上认识生命、分析生命的蓝图、解开生命的奥秘, 伴随这些研究成果而诞生的诸如转基因动物、转基因植物、基因诊断、基因治疗、基因药物学、克隆技术等生命科学的产业化又将带动国民经济的发展。生物世纪的帷幕已经拉开, 决定我们命运的已不再是星宿, 人类掌握自己命运的时代已经到来, 生命科学的春天将鼓舞每一个从事生命科学研究的科学工作者奋勇向前, 充分施展自己的才华, 迎接未来的挑战。

1 人类基因组计划的由来

诺贝尔奖金获得者美国科学家 Dulbecco R 于 1986 年在 Science 上发表题为“癌症研究的转折点——人类基因组的序列分析”的短文, 回顾了 70 年代以来癌症研究的进展, 使人们认识

* 收稿日期: 2000-11-20

作者简介: 单祥年, 1937-, 东南大学医学院教授, 博士生导师, 南京师范大学兼职教授、博士生导师, 德国 Ulm 大学客座教授, 中国遗传学会、中国医学遗传学会理事, 江苏省遗传学会副理事长, 江苏省医学遗传学会主任委员, 江苏省优生优育协会常务理事, 国际人类基因组组织(HUGO)成员, 主要从事细胞遗传学、分子遗传学研究。

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(peicemeal approach)开展研究,要么从整体上研究分析整个人类基因组及其序列。他说“这一计划的意义,可以与征服宇宙的计划媲美,我们应该以征服宇宙的气魄来开展这一计划”,这就是有名的Dulbecco“标书”。后来人们称这一短文是征服人类自身奥秘的宣言,它的伟大意义可以与“曼哈顿原子弹计划”、“阿波罗登月计划”相媲美。Dulbecco的短文在世界范围内引起了巨大的反响,人们为他的高瞻远瞩而敬佩。这一宏伟计划在美国历经5年左右的辩论后,1989年美国国会正式批准美国的“人类基因组计划”,并于1990年10月1日正式启动,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立刻就行动起来,并各具特色。最早开始国家级HGP的是意大利国家研究委员会(National Research Council),先后组织了30个实验室开始人类基因组计划研究,并集中在Xq24-qter区域测序研究。1989年2月开始的英国HGP是由帝国癌症研究基金会和国家医学研究委员会负责,由著名的Sanger中心首先在线虫基因组上进行,他们还建立了大规模DNA测序技术。1990年6月,法国HGP启动,法国科学研究部(Ministry of Research)委托国家医学科学院(INSERM)制定人类基因组计划,特别是Genethon研究中心,该中心是Daussel J用自己的诺贝尔奖金和法国民众捐款的5000万美元建造起来的,该中心构建的YAC(Yeast Artificial Chromosome, YAC)重叠群和微卫星标记为HGP的完成做出了不可磨灭的贡献。德国1995年6月才正式开始HGP,虽然起步较晚,但来势凶猛,对完成21号染色体的大规模测序工作作出了杰出成绩。1995年6月欧共体通过“欧洲人类基因组计划”,资助了23个实验室,并雄心勃勃地要赶超美国。此外参与HGP研究的还有丹麦、俄罗斯、日本、韩国和澳大利亚等。

中国的HGP是1994年,在吴院士、强伯勤院士、陈竺院士和杨焕明教授的倡导下,在国家自然科学基金委员会和八六三计划的支持下,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、结构和功能研究”,并于1998年在上海成立了南方基因组中心。1999年在北京先后成立了北方人类基因组中心和中科院遗传所基因组中心以及北京华大基因研究中心。自此我国人类基因组研究有了专门的队伍和机构,特别是有像陈竺、杨焕明、于军、汪健等这样一批从国外归来的热心于中国基因组研究的青年科学家,大大加强了中国人人类基因组研究的力量,使得我国基因组研究出现前所未有的好形势,看到了胜利的曙光。

2 人类基因组计划的内容

人类基因组计划的目标是要在2005年完成遗传图、物理图和序列图。

遗传图(genetic map)又称连锁图(linkage map):关键是要找到多态性标记(Mark),亦即路标(Landmark),有了路标就容易找到要去的地方,这里的地方目标就是基因或基因组中特定的位点。第一代遗传标记主要是利用蛋白质或免疫学标记如ABO血型位点标记9q34、HLA位点标记6q21.3等。70年代后期建立起来的限制性片段长度多态性(Restriction fragment length polymorphism, RFLP)实际上就是酶切位点的有无,这些标记分辨率比较低。第二代遗传标记主要是小卫星(minisatellite)和微卫星(microsatellite),小卫星重复单位长度是6-12个核苷酸,微卫星重复单位长度为2~6个核苷酸,这些卫星的突出特点是高度的多态性,且符合孟德尔遗传规律,一般不受进化上的选择,因此可提供大量的信息。至1996年已建立了6000

多个 STR 遗传标记, 两个标记之间的平均分辨率为 0.7cM (厘摩尔根). 第三代遗传标记为 SNP (single nucleotide polymorphism, SNP), 亦称单核苷酸多态标记, 人类群体中有很高的遗传多态性, 而在大多数基因位点上都会有若干个等位型 (alleles), 对每一个核苷酸来说, 其突变率大约为 10^{-9} 左右, 这就意味着, 每一个核苷酸在任何一代人群中大约每 1×10^9 个个体就会发生一次变异. 由这种方式产生的单碱基变异就产生许多双等位标记 (biallelic marker), 这种标记在人类基因组中可达到 300 万个, 平均约每 1 000 个碱基就会有一个. 3~4 个相邻的这种标记构成的单倍型 (haplotype) 就可以有 8~16 种, 相当于一个微卫星标记形成的多态性. 因此, 由于这种标记数目多, 覆盖密度大, 在基因定位研究中就有着其它标记系统中无可比拟的优越性, 而且, SNP 研究克服了遗传标记分析技术的“瓶颈”——凝胶电泳, 可自动化操作而且为 DNA 芯片 (DNA chips) 技术应用于遗传作图提供了基础.

物理图: 完整的物理图包括人类基因组的不同载体构建的 DNA 克隆片段重叠群, 大片段限制性内切酶切点图, DNA 片段 (探针) 或一段 DNA 序列 (Sequence Tagged Site, STS) 的路标图, 以及基因组中广泛存在的特异性序列 (如 CpG 岛、Alu 序列、isochore 等) 等的标记图, 人类的细胞遗传学图 (如染色体的区带、染色体长度的百分率), 最终在分子水平上与序列图统一.

物理图的基本原理是把庞大的人类基因组先“敲碎”, 再拼接, 以便既能随意研究又能清楚地知道研究内容所处的染色体位置. 物理图以 Mb (1 000 kb)、kb (1 000 bp)、bp 作为图距, 以 DNA 探针的 STS 序列为路标. 至今已测定了 40 000 个以上的 STS, 平均图距 (即分辨率) 可达 100 kb. 因此, 人类这个庞大的基因组已被具有界标的至少分成 40 000 个小区. 其中, 构建物理图的一个重要内容是把含有 STS 对应序列的 DNA 的克隆片段连接成相互重叠的“片段重叠群” (contig), 如用 YAC 作为载体载有人 DNA 片段的文库 (克隆片段长度在 200~2 000 kb) 包含了构建总体覆盖率为 100%、具高度代表性的重叠群. 近年又发展了 BAC (Bacterial Artificial Chromosome, BAC)、PAC (P1 phage + F factor) 和 cosmid 等克隆片段较短的可靠性更高的文库.

序列图: 人类基因组计划的最终目标是要在 15 年内完成测定总长度由 30 亿个核苷酸组成的人类基因组序列图, 而遗传图和物理图的构建就是为序列图建立路标, 因为目前的测序技术还不能对很长的 DNA 分子进行测序, 所以要把庞大的基因组分成若干有路标的区域进行测序. 序列分析需要用一个区域的 DNA 重叠群使测序工作不断延伸, 这中间 STS 则被用作任何两个片段 (上百个 bp) 间重叠区域, 使分别被测的短序列进行正确的拼接, 重叠群所占的比例应尽可能的缩小, 可提高效率, 降低成本.

2000 年 6 月 26 日, 美国、英国、法国、德国、日本和中国的科学家同时宣布, 历经 10 年, 人类基因组“工作草图”或称“工作框架图”已绘制完毕, 各国的领导人也分别发表了讲话, 这标志着人类基因组计划最初定的遗传图、物理图和序列图的任务已基本完成. 2001 年 2 月 12 号中、美、日、德、法、英等 6 国科学家和美国塞莱拉公司联合公布了人类基因组图谱及初步分析结果, 根据这一结果人类基因组由 31.674 亿个碱基对组成, 含 3 万到 3.5 万个基因, 与蛋白质合成有关的基因只占 2%. 人类的蛋白质有 61% 与果蝇同源, 43% 与线虫同源, 46% 与酵母同源. 这一结果表示人类基因组计划第一阶段任务已基本完成, 比原定的时间提前了近 4 年. 这里特别还提到的是我国以杨焕明为代表的科学家, 他们在一无设备二无资金的情况下大胆勇敢地国际人类基因组计划会议上争得了人类基因组 1% 的测序份额 (3P, 30Mb), 在我国政府和社会的全力支持下, 仅用半年时间出色地完成了任务, 跻身于全世界承担测序任务的 16

个实验室中的前十名,近来就测序的速度、准确性已上升为第六位,为我国争得了荣誉,也为发展中国家争了光.因为在完成测序任务的6个国家中,中国是唯一的发展中国家,受到世界科学家高度的赞扬,他们不畏艰难、顽强拼搏的精神,将载入我国科学史册.

3 后基因组计划

人类基因组计划的完成意味着结构基因组学的结束,那么,是否意味着人类已揭开自我生命的奥秘了呢?不!八字还未见一撇,只能是万里长征才走完第一步.根据2001年公布的最新结果,人类的结构基因只有原来估计数的1/3,只比果蝇基因数目多一倍,一个基因编码一种蛋白质远不能满足人类这样结构复杂的高等生物.因此现在有一种看法,认为一个基因的很多部份在表达时会象洗牌一样有不同的转录本和剪接模式,每种顺序都产生不同的蛋白质,并已在部分基因得到证实.尽管我们了解了人类全部基因组序列,但是我们还无法知道人类全部基因的结构和功能,以及基因与基因之间的相互关系.因此就出现了后基因组计划(post-genome project)或称功能基因组学(functional genomics).实际上人们在完成结构基因组计划的同时就已经在进行后基因组学计划的准备,并已开始了一系列卓有成效的研究.近年诞生的人类基因组多样性计划(Human genome Diversity project)、环境基因组学(Environmental Genomics)、肿瘤基因组解剖学计划(Cancer Genome Anatomy project)以及药物基因组学(pharmacogenomics)就是后基因组计划的代表,其核心问题是人类基因组的多样性、遗传与常见病产生的起因、衰老问题以及基因表达调控的协调作用和所产生蛋白质的功能等.人们预计完成后基因组计划将需要几十年到一百年的时间.

目前世界范围内就后基因组计划展开了激烈的竞争,包括国家和私人企业.结构基因组计划的完成为我们提供了寻找结构基因的极为有利的硬件,而且结构基因组计划是不能专利的,免费对世界开放,在此基础上,谁找到对人类生存攸关的基因,谁就可能申请专利垄断起来,目前的焦点正放在寻找人类6000余种单基因遗传病和常见多发的多基因遗传病的致病基因.基因就是钱,90年代初美国克隆了一个肥胖基因(ob)就以9000万美元卖给一个公司,一个乳腺癌基因(BRCA 1)以200万卖断,高额的利润,诱人的市场前景,使得一些大的私人企业不惜花巨额资金寻找功能基因,如PE(Perkin-Elmer)计划投资3亿美元自己完成人类基因组测序,其目的是不言而喻的.Incyte公司计划投资2亿美元对人类全基因组的蛋白质编码区测序,并把所有数据应用于药物基因组学遗传多样性资料的构建.不难看出,“抢基因”的“世纪大战”已趋于白热化,因为人的基因组只有一套,是一个无法再生的资源,如果把人类基因组比作一个养鱼池,成鱼的数量是恒定的,别人钓走一条,就少一条,你想要就得拿钱买,正因为基因蕴藏着令人眼馋的商业价值,丰厚的回报使“淘金热”中的“挖矿工”不惜工本,对人体基因组苦苦搜索,试图在基因破译方面抢得先机.美国和日本众多公司已提交了数量可观的专利申请,而且这种专利还会与日俱增,谁占有较多的基因专利,谁就将在包括基因药物、基因诊断和基因治疗等商业开发方面占据有利地位.基因并非发明,本不应专利,但由于寻找基因的巨额投入和市场需要,专利无法避免.尽管遭到了许多科学家的反对,但这一“圈地运动”仍然在如火如荼的进行着.面对基因争夺战,一些国外科学家勾结发展中国家一些唯利是图的人,内外勾结,肆意掠夺发展中国家的遗传资源,特别是遗传病家系的标本,因为有了这些标本就可能克隆到新的基因.我国人口占世界的22%,有56个民族和205种语言,是研究人类基因组多样性不可多得的材料.由于发展水平关系,我国迄今仍保存有若干完整的遗传隔离群和疾病大家系,而

发达国家由于迁移习惯、通婚等因素, 基因资源早已被破坏, 因此我国基因资源让很多发达国家垂涎三尺。如果我们不加以保护, 不积极投入人力、物力自己研究、克隆, 我们势必成为国际上基因“狩猎物”。面对日益猖獗的基因海盗, 许多发展中国家已发出怒吼: “你们抢走了我们的黄金, 现在又来抢基因。”并采取了保护措施。我国国务院已于 1998 年 9 月颁布了“人类遗传资源管理暂行办法”, 用法律手段保护我国遗传资源, 但这只能防君子, 难以防小人, 关键是我们应在人类结构基因组计划完成的情况下加强我们自己克隆基因的能力, 一些科学家大声呼吁“自己克隆基因, 克隆自己的基因”。这才是根本解决办法。

4 机遇与挑战

在结构基因组计划中, 我们出色的完成了 1% 的测序任务, 并培养了一批专业素质高、事业心强、有奉献精神的科技队伍, 表明中国有能力在后基因组计划中和发达国家进行一场基因大战。由于党和国家领导人的高度重视, 科学界和社会的关心和大力支持, HGP 已被确定为国家计划, 中国政府已经提出我国人类遗传资源的管理办法, 对遏制遗传资源的流失和促进平等互利的国际合作起到了关键作用。就我们目前科技队伍的总体水平来看, 仍然是一种有待发挥的潜在优势, 因此我们必须保持清醒的头脑, 积极进取, 突出自身的优势和重点, 以争取在国际上占有一席之地。

中国作为人口大国, 拥有世界上最丰富的人群资源(包括疾病家系、人群和个体), 我国人群的疾病谱广, 且发病率与西方人群可能存在一定的差别。以癌症为例, 中国肝癌、鼻咽癌、食管癌发病率明显高于西方, 而 I 型糖尿病的发病率又明显低于西方。对这些差别的分析表明, 除环境因素外, 遗传因素也有重要的作用。近年来, 心血管病、糖尿病、老年痴呆症、精神分裂症等多基因疾病发病率有逐年升高的趋势, 大城市已接近西方国家的水平。利用这些丰富的遗传资源, 分离我国发病率高、危害性大的重要疾病的致病基因和相关基因, 发展疾病基因组学(Morbid Genomics)将是我国功能基因组学的重点。为此应努力做好疾病遗传资源调查、登记, 血标本和组织标本的采集和保存, 形成现代化管理的样品库。同时要进一步改善现有的技术体系, 争取在肿瘤遗传病和多基因疾病相关位点的定位、克隆、突变检测和大规模测序方面取得突破, 获得一批重要的疾病基因, 进而研究结构和功能。

为了迎头赶上 HGP 的快速发展, 力争在功能基因组学竞争中占据主动地位, 我们还要特别注意生物信息学的研究和应用。当前, 生物信息学已成为基因组学中的带头学科, 它要求数学、物理、计算机科学和生物学的有机结合。与实验生物学相比, 它对设备、试剂方面的投资较小而对人员的智力要求很高, 这一点非常有利于发挥我国多学科协作优势和人才优势, 并形成自己的学科生长点。随着我国 HGP 研究的深入, 建立自己的数据库和服务体系已势在必行, 我国的生物信息学研究一方面可以满足我国在疾病和功能基因的结构、功能研究方面的需要, 通过研究新的算法和软件系统, 为识别和克隆人类新基因及预测其功能作出贡献, 另一方面也加强了对基因组结构组成和进化规律的基础研究。在这些研究中, 一方面要注意编码序列和蛋白质的结构——功能关系、基因和蛋白质的聚类分析、蛋白质高级结构的计算机模拟和预测, 同样要注意非编码序列的结构和功能研究, 因为现在看来, 基因间在外显子和内含子有相互重叠现象, 而且发现了一些基因的内含子有基因调控的功能。

总之, 我国人类基因组学研究方面虽已打响了第一炮, 但任重而道远, 目前我们仍面临前所未有的机遇与挑战, 我国的遗传学界和医学界以及生命科学的其它研究领域的科学家们只

有密切合作, 奋力拼搏, 才能堂堂正正地加入国际“基因俱乐部”。

5 人类基因组计划与医学

人类基因组计划就是人类了解自我、解码生命的过程, 迄今为止, 人类对于自己如何来到这个世界, 又如何一代一代的生生息息知之甚少。中国有这样的一些谚语: “种瓜得瓜, 种豆得豆”, “龙生龙, 凤生凤, 老鼠生儿会打洞”, 表明“遗传”的通俗哲理, 孩子像她(他)们的双亲, 但又不完全像, 每个人都有世界上独一无二的自我。人类疾病更是每一个人时刻关心的问题, 全世界有20%的人备受各种慢性病的煎熬, 中国有15%的人患有高血压病和糖尿病。几乎每家都有一个或知道一个因癌症而失去的亲人。癌症的阴影未去, 艾滋病等新的病魔又在肆虐, 原来已近乎绝迹的一些传染病又卷土重来。我国在“八五”期间的医药卫生费用总额为1 000多亿人民币, 相当于全国财政收入的1/5, 所有这些威胁人类生存的疾病以及人类衰老和死亡奥秘都呼唤我们要尽快探索人类自身的各种奥秘。

50年代的DNA双螺旋模型, 打开了认识生命奥秘的大门; 70年代的DNA克隆技术以及转基因技术, 使我们能直接操作DNA分子; 而跨世纪的人类基因组计划则将人类带进了以DNA序列为基础的、以生物信息学为主导的生命科学与生物产业的新世纪。

不难看出, 人类基因组计划的建立, 影响最大的莫过于医学领域, 它将使医学发生重大的变化, 人们在了解或克隆到与疾病相关的基因后, 就可以进行基因预测、基因预防、基因诊断和基因治疗, 使医学进入一个崭新的时代——基因医学时代, 将为解决人类的健康做出巨大的贡献。正如DNA双螺旋结构发现者之一, 詹姆斯·沃森曾说过: “过去我们以为自己的命运存在于我们的星座中, 现在我们知道, 在很大程度上, 我们的命运存在于我们的基因中。”例如, 今年英国科学家研究了134个有睾丸癌病史的家族, 发现如果在Y染色体上有TGLT1基因缺陷就容易患睾丸癌。人类另一种绝症白血病, 1999年报导, 约有1/5的病人与AT基因突变有关, 这种基因本来是从一种遗传病叫毛细血管扩张性共济失调症病人体内克隆到, 现发现它只要发生突变从AT→ATM, 就和慢性淋巴细胞白血病有关, 这种白血病还有遗传倾向。类似上述这些与肿瘤发生有关的基因还有很多, 迄今人们发现癌基因110多种, 肿瘤抑制基因十余种, 癌基因和抑癌基因在没有突变和活化以前, 则是我们人体发育、生长、分化不可缺少的基因, 当受到理化因素或生物因子作用发生突变后, 才有可能导致人体发生肿瘤。人类单基因遗传病约6 000种以上, 所谓单基因遗传病就是这种基因在人类基因组中只有一个拷贝, 少数有几个拷贝, 现已克隆到的还不到1 000种, 对这类疾病我们只要检查该基因的状况就可以确定属于哪一种遗传病。如进行性肌营养不良症(DMD、BMD)的基因大约有2 400 kb, 是迄今为止发现的最大的基因, 如果这个基因发生突变或缺失, 就会导致肌营养不良症。被归于人类生命三大杀手之一的高血压病及由它引起的心血管疾病, 现已找到控制人体血压的几个候补基因, 有望在不久的将来有所突破。此外, 像糖尿病、哮喘等威胁人类健康的常见多发病, 也已克隆到若干个与之相关的候补基因, 这类疾病和单基因病有所不同, 导致这些疾病的基因不是一个而是多个, 称为多基因遗传病, 但在多个基因中, 可能只有一个或两个是起关键作用的, 人们把它称之为主效易感基因, 我们可以检查主效基因诊断某种多基因遗传病。总之, 人类的所有疾病都和基因有关, 甚至传染病也不例外, 几个人生活在同一环境中, 有的人患了感冒, 而另外的人则没有, 为什么呢? 这就是对感冒的易感基因因人而异的原因。

既然人类所有疾病都和基因有关, 我们在临床上就可以检查某个病人的基因从而判断该

病人患的是哪一种病, 对于有特征性临床表现的患者则可有目的地检查某个基因, 这就是基因诊断。目前在国内外有条件的医院都已开展了这项工作, 也就是说, 人们对一些疾病已能直接从 DNA 水平上检测人类疾病的基因, 从基因型诊断表现型。与传统的诊断相比较, 基因诊断更准确、快捷, 而且不受基因表达的时空限制, 还可以对传统的诊断方法进行诊断敏感性和标准的评价, 从而最终代替某些已有的诊断方法或促进一些诊断方法的改进或在某些特殊的病种中补充已有的诊断技术。基因诊断的优点还在于它除了可以对现症病人进行诊断外, 还可以进行症状前诊断和产前诊断, 即可以进行基因预测。随着人类结构基因组计划的完成和功能基因组计划的迅速进展, 可提供的疾病基因与日俱增, 新的基因探针将会象雪片一样压向医学界。待人类基因组计划和功能基因组计划完成后, 将来基因诊断可能发展到每个人在出生后不久就会得到一个“基因身份证”, 它录下你所有的遗传信息和基因缺陷, 包括你所有的优秀基因和不良基因。比如何时会患何种病, 不宜吃何种食品, 不宜与哪些人结婚生育后代等信息, 由于 DNA 芯片技术的问世, 这已不再是天方夜谭, 届时人们可以通过基因测试掌握各自的病兆, 达到基因预测、基因预防的目的。

基因治疗是人类基因组计划又一受益领域, 传统的疾病治疗是根据临床特征和经验用药, 只能治表, 而基因治疗是从分子水平上应用 DNA 重组技术将目的基因导入病人体内代替或补偿病人体内已发生突变的基因, 从而使病人症状得到缓解或根治。它的基本原理就是将目的基因 DNA 或 cDNA 重组入特定的载体, 如逆转录病毒、腺病毒、痘苗病毒等, 应用这些病毒能高效感染人细胞的特点把目的基因导入病人体内并在病人体内表达, 这些病毒载体都经过人工改造, 一般不会带来副作用。此外, 物理、化学方法也能导入目的基因, 如显微注射、微粒子轰击法、氯化钙沉淀、电穿孔(细胞打孔)等。目前全世界有临床基因治疗方案 300 多个, 接受治疗的人数 2 500 余人, 其中一半以上是肿瘤病人。对于恶性肿瘤的基因治疗近年发现了很多新的方法如分子化疗法, 又称为“自杀基因”疗法, 将疱疹病毒 TK1 激酶基因通过逆病毒载体导入肿瘤部位, 利用 TK 基因产物能错误地把胸苷类似物 GCV 或 ACV 磷酸化, 以假乱真, 让肿瘤细胞在 DNA 合成时误用假的“胸苷”达到杀死肿瘤细胞的目的。此外还有免疫基因疗法、反义技术等也已在恶性肿瘤基因治疗中应用。

所有这些基因治疗方案, 前提是必须弄清导致人类疾病的是什么基因, 发生什么样的突变, 它的核苷酸顺序如何等, 而人类基因组计划和功能基因组计划就能解决这个问题。

药物基因组学(pharmacogenomics)是伴随人类基因组计划而诞生的另一遗传医学领域。对于医生来说, “对症下药”是一种理想的境界, 然而做起来很不容易, 因为人体的遗传背景千差万别, 而药物都是一种配方、一种规格、统一适用。药物基因组学是基于人类基因组多样性基础上建立起来的, 由于不同个体遗传背景不同, 这种多样性使不同个体对某种药物的敏感性也就不同。药物基因组学就是研究遗传多样性引起对药物或有毒物质反应的差异, 进而研究每个个体基因组上所存在的与药物作用不相同的靶分子情况。例如载脂蛋白 E 位点的等位基因 ApoE4, 不仅与 Alzheimer's 病(一种常染色体显性遗传的老年痴呆症)的发生有关, 还可预测对胆碱酯酶抑制治疗的疗效, 这是通过疾病相关基因来预测药物反应的典型例子。这类多态性研究, 必须大规模地研究基因组对药物反应的标记, 微卫星标记和近年来发展起来的 SNP (single nucleotide polymorphism, SNP), 使这些遗传标记的密度大大增加, SNP 在人基因组中每 1 000 个碱基中就有一个多态, 便于对不同药物敏感性进行关联分析, 也便于自动化操作。总之, 药物基因组学的最终目标是医生将可以为不同病人开出“个性化”的药, 因为我们在得到

影响药物作用的多态性后,可优化药物设计,研制新化合物,以达到一把钥匙开一把锁的目的。

基因药物学则是人类基因组计划又一获益领域。导致人类某种疾病是因为某个基因突变的原因,突变的基因不能正常编码基因产物或编码异常的产物,例如血友病 A 即编码凝血第 Ⅷ 因子的基因发生突变而不能产生正常的凝血第 Ⅷ 因子,使患者凝血机制受损而导致血友病。为了解决这个问题,人们就给患者补充 Ⅷ 因子,早期是给患者输正常人血浆,后来在 Ⅷ 因子基因克隆以后,就可以用基因重组的技术,在体外利用细菌或其它的体外培养细胞来产生人的正常第 Ⅷ 因子,给患者注射,达到缓解症状的目的,这就是基因药物学的梗概。目前市场上大约有几十种基因药物,临床上称之为重组药物。人类基因组计划已进入基因产物与表型相关的功能鉴定阶段,规模大,速度快,伴随而来的竞争最为激烈的就是基因专利,拿到某个基因就能有每年上亿美元的产值,其奥秘就在于可以用 DNA 重组技术将该基因产业化推向医药市场。1999 年,全世界药物产值是 3 000 亿美元,预计 2002 年约升至 4 100 亿美元,其中有很大比例是基因药物投放市场带来的产值。80 年代以来国外兴起的分子农场,也叫生物反应器,也属这个范畴。所谓分子农场就是将人的基因通过转基因技术,转导到动物体内,并使之表达,这样病人可以通过吃动物的乳汁或肉达到治疗疾病的目的。我国上海曾益涛院士成功地将人的第 Ⅷ 凝血因子应用转基因技术导入羊的体内,开辟了我国分子农场的先河。

综上所述,人类基因组计划完成和功能基因组计划的实施,冲击最大的当属医药领域,基因诊断、基因治疗、药物基因组学、基因药物学,无不与人类基因组计划密切相关,生物产业已在世界范围内兴起,来势之凶猛、竞争之剧烈是前所未有的,给我们的机遇与挑战是并存的,要想将来不花巨额外汇购买基因专利或基因药物,我们只有奋起直追,迎头赶上,为中华民族的繁荣昌盛而“解近忧、去远虑”。

6 比较基因组学

异同的比较,历来是遗传学以及整个生物学研究重要的方法。比较基因组学(comparative genomics)的特点是在整个基因组的层次上的比较,如基因组的大小、基因数目的多寡、特定基因的存在或缺失、基因(或标记序列片段)的位置及排列顺序、特定基因或片段的所在组织等,而最重要、也最能体现不同物种基因组的异同的是全基因组核苷酸序列的整体比较。在人类基因组计划完成过程中,模式生物大肠杆菌、酿酒酵母、黑腹果蝇、秀丽线虫、小鼠的基因组研究作出了巨大贡献。此外斑马鱼、河豚鱼、拟南芥等的基因组研究也给人类基因组研究提供很多信息。目前,大肠杆菌、酿酒酵母、秀丽线虫、拟南芥全基因组测序已经完成,小鼠也即将完成,此外还有约 37 种细菌的全基因组测序也已完成,我国已在世界上率先完成了人痢疾杆菌和温泉中一种耐热菌的全基因测序和密码破译。随着不同物种的遗传图、物理图的陆续建立和完善以及测序成本的不断降低,速度越来越快,新的物种全基因测序将不断涌现,目前至少有 40 余种哺乳动物已建立了物理图和遗传图。这将给动物分类进化研究提供一个强有力的工具,以往分子进化大多以线粒体基因序列进行比较,其局限性比较大,也很难得出肯定的结论,如果能进行全基因组比较,很多悬而未决的分类进化上的难题将会迎刃而解。因为在进化上,始祖基因或始祖染色体及始祖基因组往往经历了复制、重排或修饰而产生了生物基因组的多样性,同时也保留了不少保守板块(conserved block),在同一板块的染色体片段中,基因的种类与结构及排列次序在不同物种的基因组中基本一致,可以认为现代的生物基因组是保守板块组成的。一个物种的不同器官之间的差异要比另一物种的同一器官之间的差异大得多。如人的心脏与

人的肝脏之间的差异要比人的心脏与猪的心脏之间差异大得多。“进化”这一“千古命题”随着生物学的每一进步都有新的回答,基因组在进化上的整体性与连续性、基因组的纵向或垂直传递(vertical descent)与横向(horizontal)扩展与变异的矛盾与统一,这都是比较基因组研究的对象。在某种意义上可以说进化是一门难以用实验重演或最终证明的科学,而全基因组分析和比较基因组学研究则为进化提供了最本质的材料和根据。

经济动植物的基因组研究是比较基因组学中的重要组成部份,经济植物和家养动物基因组工作已经初步展开。人类已积累了一万年以上丰富的驯化野生动植物的经验,人类生活无时无刻不依赖于自然界和家养的动植物,优良的动植物品种的培养和筛选,有赖于遗传学提供科学、快速的方法和技术。由于植物基因组的研究与人类的关系最大,受到各国的高度重视,美国的食物基因组计划(Food Genome Project),近乎包括所有的经济作物,如小麦、大豆、马铃薯、南瓜、棉花等。日本和中国的水稻基因组计划进展也很快。基因组物理图已经建成,测序工作也已开展,最近袁隆平院士已和华大基因研究中心签订合同,协同研究我国杂交水稻的基因组。水稻基因组约 430 Mb,除本身意义外(世界约 1/3 人口以水稻为主食),还是很有价值的模式生物。植物基因组计划中另一重要的模式生物拟南芥(*Arabidopsisthaliana*)基因组测序已完成,全基因 120 Mb,基因密度 4.8 kb,将为其它植物基因组计划的进行提供重要的信息。

家养动物中的猪和牛已建立较详细的物理图,鸡的基因组研究已取得突破,全基因组大小在 300 Mb 左右,在鸡的 39 个染色体($n=39$)中,有 29 个是小染色体,最近通过对 CpG 岛的分布研究,证明仅占 1/4 基因组 DNA 的这些染色体很可能含有鸡的多数基因。猪是基因组计划中另一倍受重视的家养动物,特别在发展中国家,人们的动物蛋白主要由猪肉提供,我国是世界上生猪的第一生产国和消费大国,猪优良品质资源极为丰富,世界上流行的良种猪都有我国良种猪的基因。以太湖猪为例,它的繁殖率是首屈一指的,被多个国家引进。此外,家猪的基因组无论从基因的数量还是碱基对的数目,和人类都很相似,基因也大部分相同,估计差异不会超过 5%。破译猪的遗传密码,还对医药工业有很大的意义,目前和猪有关的约 40 种产品一直作为药物应用,通过对猪的转基因研究还可人类开辟新的异种器官移植来源。我国以杨焕明教授为首的北京华大基因研究中心,已和丹麦科学家携手启动了猪的基因组计划研究,可望在不久将来有所突破。

水产生物以鱼的基因组研究进展较快,斑马鱼(*Zebrafish, Daniorerio*)的遗传图精度已达 3.3CM,尽管斑马鱼与人类的进化距离约 4.2 亿年左右,但很多片段仍很保守,至少已发现 3 个肯定区段大的同线保守区。如人的 2 号染色体长臂的 10 个基因,在斑马鱼中分布在两个不同的染色体上,提示哺乳动物与斑马鱼的共同祖先在这一染色体上有 10 个基因,后因染色体断裂、融合导致目前的状况。1993 年, Brenner 等人首先提出用河豚鱼(*Fugu rebripes*)作为脊椎动物的模式生物,河豚鱼属硬骨鱼纲,与人的进化距离 450 余万年,它的基因组最主要的特征是小而紧凑,为 260~400Mb,已知基因组计算接近于其它的脊椎动物,基因密度约为 6~8kb,为人的基因密度 7 倍左右,其结构基础是重复序列较少,基因间序列较短,基因的内含子一般较小,但是同源基因的编码序列及基因产物的结构与其他脊椎动物相似程度很大,所以河豚鱼的基因组研究也倍受青睐,是一个理想的模式动物。

迄今为止已对 200 多种生物物种绘制了精度不同的遗传图和物理图。预计,比较基因组学将很快覆盖所有生物。正如细胞学曾在 30 年代统一了生物学一样,基因组学将在新的层次上统一 21 世纪的生物学。目前许多国家都已瞄准生命科学这一朝气蓬勃、发展势头迅猛的领域,

如新加坡最近制订了 21 世纪以生命科学为主的科技发展计划,试图像 90 年代信息学一样再次站在世界科技发展的前列.总之,以生物产业兴国是人们的梦想,它将变为现实,21 世纪是生命科学的世纪,也是我们从事生命科学研究的人大显身手的世纪.

[参考文献]

- [1] 贺林. 解码生命[M]. 北京: 科学出版社, 2000
- [2] 杨焕明. 生命大解密[M]. 北京: 中国青年出版社, 2000
- [3] 于军, 杨焕明. 关于我国人类基因组计划策略性发展的几点意见[J]. 遗传, 1998, 20(6): 1 ~ 5
- [4] 陈竺, 张思仲. 我国人类基因组研究面临的机遇与挑战[J]. 中华医学遗传学杂志, 1998, 15(4): 195 ~ 197
- [5] Collins F, Galas D. A new five-year plan for U. S Human Genome project 1998– 2003[J]. Science, 1998, 282: 682 ~ 698
- [6] Fields C, Adams D. How many genes in the human genome[J]. Nature Genetics, 1994, 7: 345
- [7] Hieter P, Boguski M. Functional genomics: it's all how you read it[J]. Science, 1997, 278(5338): 601 ~ 602
- [8] Miklos G L, Rubin G M. The role of the genome project in determining gene function: insights from model organisms[J]. Cell, 1996, 86(4): 521 ~ 529
- [9] Pennisi E. NRC Oks long-delayed survey of human genome diversity[J]. Science, 1997, 278(5338): 568
- [10] Rowen L, Mahairas G, Hood L. Sequence of the human genome[J]. Science, 1997, 278(5338): 605 ~ 607
- [11] Schuler G D. A gene map of the human genome[J]. Science, 1996, 274(5287): 540 ~ 546
- [12] Weber J L, Myers E W. Human whole-genome shotgun sequencing[J]. Genome Res, 1997, 7: 401 ~ 409
- [13] Barbujani. Apportionment of human DNA diversity[J]. Proc Natl Acad Sci 1997, 94: 4516
- [14] O'Brien, S J. The promise of comparative genomics in mammals[J]. Science, 1999, 286: 458 ~ 481
- [15] Dean M. Polymorphism A mixture Typing in Human Ethnic population[J]. Am J Hum Genet, 1994, 55(4): 788 ~ 808
- [16] Corder E H. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families Science[J]. 1993, 261(5123): 921 ~ 923
- [17] Housman D, Ledley F D. Why a pharmacogenomics? Why now[J]. Nat Biotechnology, 1998, 16: 492 ~ 493
- [18] Marshall A. Genset-Abbott deal heralds pharmacogenomics era[J]. Nat Biotechnology, 1997, 15: 829 ~ 830
- [19] Patrick W. Genetic variation as a Guide to Drug Development[J]. Science, 1998, 281: 1820 ~ 1821
- [20] Brenner S. Characterization of the pufferfish(Fugu) genome as a compact model vertebrate genome[J]. Nature, 1993, 366(6452): 265 ~ 268
- [21] Bult C T. Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii[J]. Science, 1998, 273(5278): 1058 ~ 1073
- [22] Copeland N G. A genetic linkage map of the mouse: current application and future prospects[J]. Science, 1993, 262(5130): 57 ~ 66
- [23] DeBry R W. Human/mouse homology relationships[J]. Genomics, 1996, 33(3): 337 ~ 351
- [24] Ecker V R. Genome sequencing, Genes blossom from a weed[J]. Nature, 1998, 39(6666): 438 ~ 439
- [25] Huynen M A. Differential genome display[J]. Trends Genet, 1997, 13(10): 389 ~ 390
- [26] Judson H F. The Eighth Day of Creation: Makers of the Revolution in Biology[M]. 1998, Expanded Edition. New York cold Spring Harbor Laboratory press.
- [27] Kappes S M. A second-generation linkage map of the bovine genome[J]. Genome Res, 1997, 7(3): 235 ~ 249
- [28] Konin E V. Prokaryotic genomics: the emerging paradigm of genome-based microbiology[J]. Curr Opin

Genet Dev, 1997, 7(6) : 757 ~ 763

- [29] Lyons L A. Comparative anchor tagged sequences for integrative mapping of mammalian genomes[J]. Nature Genet, 1997, 15(1) : 47 ~ 56
- [30] McQueen H A. CpG islands of chicken one concentrated on microchromosomes [J]. Nat Genet, 1996, 12(3) : 321 ~ 324
- [31] Mckusick V. Genomics: structural and functional studies of genomes[J]. Genomics, 1997, 45(2) : 244 ~ 249
- [32] Pennisi E. Genomics data shake tree of life[J]. Science, 1998, 280(5364) : 672 ~ 674
- [33] E S Force A. Vertebrate genome evolution and the Zebrafish gene map[J]. Nat Genet, 1998, 18(4) : 345 ~ 349
- [34] Sasaki T. He Japanese rice genome research program[J]. Genome Res, 1996, 6(8) : 661 ~ 666
- [35] 单祥年. 临床基因诊断[M]. 南京: 南京师范大学出版社, 1999

The Spring of Life Science

Shan Xiangnian

(Medical College of Southeast University, Nanjing, 210009, PRC)

Abstract: Owing to the joint effort, by Chinese, American, English, French, Japanese and German scientists, Genetic and physical maps as well as sequence map have been completed the June of 2000 by HGP started in 1990. A preliminary analysis has also been made by HGP that human genome consists of 316. 74Mb and 30 to 40 thousand structural genes. It declares both the accomplishment of human structural genomics and the beginning of post-genomic project, i. e. functional genomics. Scientists are now working hard to revealing the mysteries of human's birth, age, illness, and death.

First many changes have taken place in the medical field driven and prompted by HGP. Gene diagnosis, gene therapy, gene pharmacy, and pharmaceutical genomics have penetrated every corner of medicine. It will be a common clinical practice to diagnose and cure human's disease by molecular biology. The dream that human can dominate his own fate will come true and people will live a much longer and healthier life.

Comparative genomics is another field which benefits a lot from HGP's smooth operation and success. So far there are more than 200 species for which genetic and physical maps have been drawn with different precision. The whole genomic sequencing and codon translation have been accomplished for some of them, especially the economical animals and plants closely linked with human's life. The completion of the animals and plant's genomic project will provide molecular bases and methods for culturing breeds with top quality and high production. At the same time, with the progress of comparative genomics some doubts concerning biological sorting and evolution will be solved.

The development of science and technology which gives first place to life science has been grandly presented to public. It is people's objective to make our country prosperous by biological industry. The twenty-first century is the century of life science, and the spring of life science will inspire every one engaged in life science research to forge ahead, give full play to his talent and meet the future's challenges.

Key words: HGP, Post-genome project, pharmacogenomics, Comparative genomics

[责任编辑: 严海琳]