

21 世纪医学、生物学的蓝图——人类基因组计划

The Human Genome Project is a Blueprint for the Medicine and Biology in the 21st Century

杨焕明

(国家遗传医学中心,教授 北京 100005)

编者按 本刊曾于1988年3期上刊出了“人类基因组顺序测定蓝图”一文,首次介绍、报导了当时震惊世界科坛的这一空前浩大的生命科学工程。8年来,这项工程进展如何,越来越受到科技界诸方专家、学者的关注。这一动向和这一关注是可以理解的。因为随着生命科学的不断突破和进展,越来越显示出它将成为下一世纪科学的最重要主角之一。而人类基因组图谱工程又是这一主角学科中的前沿主角。本刊这期用了较大的篇幅向读者介绍了人类基因组的研究,即本篇的“21世纪医学、生物学的蓝图——人类基因组计划”和下一篇的“人类基因组和某些功能基因”。这两篇文章虽然专业性偏强了些,篇幅也偏长了些,但对渴望了解这一宏伟生命工程核心、实质及其对未来科学事业深远影响的各方行外专家读者来说,不妨慢慢读来,细细去理解、领会和品味。相信读完后,对我们初步领略这一深奥领域的基本概念、概貌,了解其来龙去脉,了解其科学价值之重大、科学意义之深远,是大有裨益的。

8年前,“科技导报”曾以“人类基因组顺序测定蓝图”为题发表了一篇有关人类基因组的文章,以引起各界的重视。今天,“人类基因组计划”已经成为举世瞩目的全球性的巨大工程。本文作者拟回顾这一计划的历史发展,纵览这一计划十年来的成果,并瞻望这一计划的前景,特别讨论这一计划与我国的关系以及我们应该发挥的作用。

现代自然科学已使人类成为这一星球真正的主宰。自然科学的一切,都是为了人类的幸福与价值。对健康的追求,是人类有史以来最重要的行为。

现代自然科学使人类的健康踏上了新的台阶。人类已根治了天花,战胜了霍乱等瘟疫,控制了麻疯病等不治之症。可是,与现代自然科学在太空、信息、武器、交通等方面的辉煌成绩相比,人类对自身的认识、对自身的保护却越来越不如人意,有20~50%的人每天备受各种慢性病的折磨;我国11%的人患有高血压,4.2%的人为不同程度的残疾病人,2.5%的人患有不同程度的智力低下;非洲和中国的孩子约有1/6患近视、色盲,而成为“第二、第三残废”;肿瘤、心血管等主要死因已成为人类驱除不掉的幽灵,艾滋病等“天外横祸”使人们对未知灾难

又有了新的恐惧。英国的疯牛病又使医学工作者不得不现实地考虑:会不会有一天,天知道是否会有—种不知名的“病”,会象几百年前的感冒一样夺去几百万“健康”人的生命。

人类为健康的追求付出了昂贵的代价。经济上,美国一年的医药费在7000亿美元以上。美国国家医学研究院(NIH)的经费,居国家各研究预算之首,超过100个亿。我国“85”期间每年的医疗费用,也超过1000亿人民币,接近全国财政总收入的20%。

人类的医学已经有了过去难以想象的进步,但是,相对于“摸得到,猜不透”的人体奥秘来说,我们知道的还太少太少。据说在古希腊特尔斐(Delphi)的阿波罗神殿上,就有了“了解自我”的字样。人类社会的发展,就是基于人类的利益而敢于接受挑战。**人类基因组计划**(HGP, Human Genome Project)就是人类认识自身,揭开人体奥秘,奠定21世纪医学、生物学发展基础的一个庞大工程。

在人类自然科学史上,还没有一个科学研究项目能够象“HGP”那样得到全世界各国科学家的认同,能够形成全球性的规模与各个国家的精诚合

作,能够在短短的几年时间内取得比预计更快的发展,又能够如此广泛地波及,乃至冲击社会、立法、伦理诸方面。

在这一人类自然科学史上最重大的合作项目启动 10 周年之际,本人愿以一家一孔之见,对这一计划的纵横始“末”作一回顾与展望。

人类基因组计划历史背景 与“全球化”进程

1. 杜伯克(Dulbecco)“标书”与历史背景

“人类基因组(Human Genome)”是指决定人体的生、长、育、老、病、亡的所有遗传信息的总和。在经典遗传学,是指人体所有基因的总和;在细胞遗传学(研究遗传物质的载体染色体的结构、行为的科学),是指人体细胞中的 24 个不同染色体的总和,即“染色体组”;在分子遗传学(从分子水平研究遗传物质的结构与功能的科学),是指人体细胞中精子或卵子等“单倍体”细胞(人体的大多数体细胞为“双倍体”,即含有两组“份”相似的遗传物质)中的 25 个 DNA 分子(24 个染色体与线粒体的 DNA)的总和。从整个基因组的层次上来研究遗传的科学称“基因组学(genomics)”,它是著名遗传学家麦库西克(V. McKusick)于 1986 年首先提出来的。

尽管类似人类基因组计划的研究在 70 年代即已开始,80 年代在很多国家已形成一定规模,但作为国家级的重大项目还是由美国率先开始的。10 年前,1986 年 3 月 7 日《Science》杂志上发表的一篇短文,被认为是这一项目的“标书”。

这篇短文的作者是著名的生物学家、诺贝尔奖金获得者杜伯克(Dulbecco)。他以“癌症研究的转折点——人类基因组的全序列分析”为题,首先回顾了 70 年代以来癌症研究的进展。这些进展归结为一点,是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。他认为,我们面临两个选择:要么大家各自研究自己感兴趣的基因,即“零敲碎打”途径(piecemeal approach),要么从整体上研究人类的基因组,分析整个人类基因组的序列(Sequencing the whole genome)。

杜伯克阐述了后一策略的必要性与艰巨性,并写道:“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进行这一计划”;“这样的工作是任何一个实验室难以承担的,它应该成为国家级的项目,并且使它成为国际性的项目。人类的 DNA 序列是人类的真谛。这个世界上发生的一切事情,都与这一序列息息相关。”

10 年后的今天,当我们重读这一短文时,不能不钦佩杜伯克的高瞻远瞩,也就不难理解这一“标书”为什么能在全世界引起那么大的反响。

10 年后的今天,有必要简单回顾这一“标书”的技术背景。

80 年代初期,由于生物学技术,特别是 DNA 克隆、测序等技术的进展,生物学、医学的研究正酝酿着新的突破。我们可以举出很多例子:大批肿瘤基因与肿瘤抑制基因的发现,使 70 年代趋于彷徨的肿瘤研究“柳暗花明又一村”;基因克隆技术的突破、遗传工程产业势在必发;基因表达研究技术渐趋完善,“讯号传导”研究初露曙光,神经活动的研究似乎面临新的突破;大规模双向电泳、核磁共振、X 射线衍射等技术的建立与改进,使蛋白质研究方兴未艾……。

上述每一方面都有足够的理由在世界范围内提出一个“计划”,如“肿瘤计划”、“遗传工程计划”、“讯号传导计划”、“脑的十年”、“蛋白质计划”……。这些计划都无可非议,当时也确实有人提出过。但为什么只有“人类基因组计划”能被大家接受并成为国际性重大计划呢?

自然科学有它自身的规律,有它的内在联系,也有它的发展阶段。只要我们想一想,所有这些计划的“最关键”因素,都需要基因来操作,就会使我们不得不佩服杜伯克“标书”的真知灼见。从某种意义上讲,HGP 是一个“补课计划”。只有得到人类基因组的全部或大部分基因,实施其他计划才有可能。

这里,也许需要讨论另外一个问题。人类基因组那么庞大,又那么复杂,为什么不先从简单的生物着手?为什么不先以更有经济意义的动植物的基因组为研究对象?杜伯克从癌症研究的角度来回答了这一问题,并明确指出:应该先从人类开始。魏特(R. White)及卡斯基(T. Cakey)在《Science》杂志上的一篇专著中更加详尽地回答了为什么人类本身已成为最合适、最高等级的实验系统(the highest order experimental system)的问题。指出这并不仅仅是由于其社会意义——了解人类本身,了解人类的生老病死,更重要的是:(1)现代生物学,尤其是 DNA 技术的发展,为研究人类基因组提供了可能性。(2)没有变异就没有基因的发现,人类本身的众多遗传变异为研究自身提供了最珍贵的材料。所以人类基因组研究的理论与技术进展将更直接、更广泛、更迅速地用于解决其它生物基因组的问题,揭示生命现象的本质;同时,HGP 还将推动生物高新技术的发展,产生重大的经济效益。

2. 美国的人类基因组计划

在杜伯克“标书”之前,对于人类基因组的讨论已经开始。1984 年 12 月 Utah 大学的魏特曾受美国能源部的委托,主持召开了一个小型会议,讨论 DNA 重组技术的发展及测定人类整个基因组的 DNA 序列的意义。1985 年 6 月,在加州 Santa Cruz

举行了一次会议,会上,美国能源部的德里瑟(C. Diisi)与司密斯(D. Smith)提出了美国能源部的“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6月在冷泉港讨论会上,诺贝尔奖金获得者吉尔伯特(W. Gilbert)及伯格(P. Berg)主持了有关“人类基因组计划”的专家会议。1987年初,美国能源部与国家医学研究院(NIH)为“人类基因组计划”下拨了启动经费约550万美元,1987年总额近1.66亿美元。同时,美国开始筹建人类基因组计划实验室。1989年美国成立“国家人类基因组研究中心”。诺贝尔奖金获得者、DNA分子双螺旋模型提出者华森(J. Watson)出任第一任主任。1990年,历经5年辩论之后,美国国会批准美国的“人类基因组计划”于10月1日正式启动。

美国的人类基因组计划总体规划是:拟在15年内至少投入30亿美元,进行对人类全基因组的分析。此计划在1993年作了修订,其主要内容包括:人类基因组的基因图构建与序列分析;人类基因的鉴定;基因组研究技术的建立;人类基因组研究的模式生物;信息系统的建立。此外,还有人类基因组研究的社会、法律与伦理问题,交叉学科的技术训练,技术的转让,研究计划的外延等共9方面的内容。

可以看出,美国计划中“定时定量”的硬任务是第一项,即我们现在所说的3张图:遗传图、物理图、序列图。第二项是在第一项任务的基础上进行的更为长期更为艰巨的任务。随后3项是为第二项任务服务的技术手段。而其他几项,则是对在实施人类基因组计划中,接踵而来的社会问题、经济问题,以及更重要的法律、伦理问题的研究;因为对于人的生物学以及个人遗传信息的了解,可能会给社会以及个人带来复杂的伦理与法律问题。

3. 欧洲各国与日本的人类基因组计划

欧洲最早开始国家级HGP的是意大利,在杜伯克的影响下,意大利的国家研究委员会(National Research Council)在1987年组织了15个(现发展到30个)实验室开始了人类基因组研究。意大利政府将人类基因组研究看成是进入世界行列的途径。意大利HGP的特点是技术多样(YAC、细胞杂交、cDNA等)、区域集中(基本上限于Xq24—qter区域)。

英国的HGP是于1989年2月开始的,特点可归纳为“全国协调、资源集中”。全国协调与资助由ICRP—MRC(帝国癌症研究基金会与国家研究委员会)共同负责,建立了“英国人类基因组资源中心”,以及YAC筛选与克隆、特异细胞株、DNA探针、基因组DNA、cDNA文库、序列分析、信息等“资源中

心”,向全国的有关实验室免费提供技术及实验材料服务。自1993年开始,伦敦的Sanger中心注重改进大规模DNA测序技术,并首先在线虫基因组上积累经验,大大降低了成本。

法国的国家人类基因组计划于1990年6月宣布开始,其计划由科学研究部(Ministry of Research)委托国家医学科学院(IMSERM)制定。法国HGP的主要特点是注重整体基因组(相对于美国的“单个染色体策略”)、cDNA和自动化(原EUREKA计划“机器人”设计能力)。诺贝尔奖金获得者道赛特(J. Dousset)以自己的奖金于1983年底建立了CEPH(人类多态性研究中心),在法国民众的支持下(民间捐助至少为5000万美元),CEPH与相关机构为全世界的人类基因组研究做出了不可磨灭的贡献,特别是在全基因组YAC重叠群与连锁图的构建方面。CEPH家系(80个三代多个体家系)已成为基因组研究的经典材料。

丹麦的HGP是在1991年开始的。作为一个小国,丹麦HGP的特点是:以5个有影响的研究所组成“丹麦人类基因组研究中心”,发挥丹麦的遗传病、流行病学调查登记资料齐全,以及拥有较多遗传病家系的特点,用以定位、克隆具有北欧特点的疾病基因。

此外,欧洲共同体在1990年6月通过了“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,这也是欧洲要在多个领域赶上与超过美国的一个重要方面。

日本的国家级HGP是在美国的推动下,于1990年开始的。特点为:(1)多个部(省)分头资助,如科研省、教育科学文化省、通商产业省、卫生福利省、农林渔业省等。(2)建立“以日本人为材料的”基因组文库与cDNA文库,(3)发展先进技术(如激光显微切割装置)。但与日本的其它领域的领先地位相比,日本的人类基因组仍略逊一筹。值得注意的是,农林渔业省在1990年即开始了“志在必夺”的“水稻基因组计划”,准备在7年之内完成。

此外,加拿大、以色列、瑞典、芬兰、荷兰、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。德国在1995年才开始的“人类基因组计划”,更具有其新的意义与特色。

4. HUGO与UNESCO的努力与人类基因组计划的全球化

人类只有一个基因组。人类基因组的研究成果应该成为人类共同享有的财富。1988年4月,在麦库西克等有远见的西方科学家倡导下,HUGO(国际人类基因组组织)宣告成立。HUGO代表了全世界从事人类基因组研究的科学家,以协调全球范围的人类基因组研究为宗旨,被誉为“人类基因组的联

合国”。

出于同样的考虑,联合国科教文组织(UNESCO)也于1988年10月在西班牙的Valencia召集会议,成立了以格里索里亚(Grisolio)为主席的“UNESCO人类基因组委员会”。1990年又在莫斯科召集了以发展中国家为主体的人类基因组会议,我国著名医学遗传学家吴旻院士出席了此次会议。1992年在巴西召开了“第一次南北基因组会议”,吴旻院士也应邀参加。

发展中国家也以积极的态度参与这一国际活动。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。

人类基因组计划内容的演变

尽管人类基因组计划在总体上已被大家认同,但从1985年开始讨论以来,在内容上却经历过重要的历史演变与逐步完善的过程。

1. 从序列分析(sequencing)计划到基因组制图(genome mapping)计划

应该承认,从历史上来说,美国人类基因组计划的原始方案重点确实拘泥于序列分析计划,即一般美国公民所理解的“以30亿美元来搞清楚30亿对核苷酸的序列”。但是,对如此庞大的人类基因组如何着手序列分析?合乎逻辑的思路是首先将人类基因组投上“路标”,分割成一个个区域。经典遗传学家以“经典遗传学”思路,提出以“遗传标记”为路标,以“遗传距离(genetic distance)”为图距的“遗传连锁图”。赞成序列分析计划的分子生物学家提出以一个DNA片段(称为DNA“探针”),或一小段已知序列的DNA片段(STS sequence-tagged site,序列标记的位置)为路标,以物理距离kb(千碱基对)或Mb(百万碱基对)为图距的物理图。而序列分析需要DNA克隆片段来操作,这些片段应相互重叠以覆盖整个基因组,即现在所说的contig(意译为克隆片段重叠群)。持这两种不同观点者曾发生很大分歧,以至于HUGO不得不扮演“媒婆”的角色。在这里我们应认真学习HGP对于不同意见的兼容并蓄的精神。“制图”计划特别是遗传图的构建,原先是作为“序列分析”计划的不同意见提出的,但双方并没有简单否定对方的意见,有关行政决策部门也没有简单地支持一方而摒弃另一方。而最后却是持不同意见的科学家走到了一起,共同制定出一个更加全面、更加科学的计划,使人类基因组计划逐步趋于完善。

2. 从制图计划到基因鉴定计划

1993年美国计划的修订方案增添了一个重要的内容:基因鉴定计划,即鉴定、分离具有重要功能,与重要疾病有关的基因。在策略上,这一计划把

我们所说的基础研究与应用研究结合起来,使人类基因组的研究结果很快转化为社会、经济效益;在客观效果上,争取把长期的科学研究与短期见效的成果结合起来。其最明显的则是新闻效应,能不断吸引民众的注意,争取民众的支持。如“时代”周刊的每年重大科学新闻,无不列入重要基因的成功克隆。

3. 从人类基因组计划到cDNA计划

在人类的基因组中只有2~5%的核苷酸序列是真正为人类的5~10万个基因编码的。首先应注重于cDNA,即来源为蛋白质编码的mRNA逆转录而得到的DNA序列,这个设想曾一度被理解成人类基因组全序列分析计划的反面意见,但没有被主流意见所拒绝,并且也没有因此而放弃“全基因组”策略,而是把cDNA计划作为“基因图”的雏型而纳入整个基因组计划,成为一个组成部分,并作为基因鉴定的重要途径的整体计划。1990年,美国能源部与英国人类基因组资源中心分别开始了cDNA计划。

4. 从人类基因组计划到模式生物计划

进化是生物的基本特点之一,也是生物学的核心。尽管有过先着手研究“人”或“非人”的争论,但人类基因组计划在选择人类作为主攻方向的同时,并不排斥对其他生物基因组的研究。相反把酵母(15Mb)、线虫(28Mb)、果蝇(120Mb)、小鼠(3 000 Mb)列为四大模式生物。其技术意义是从低等入手,建立技术,积累经验。1995年开始的大规模序列分析,便是在酵母、线虫全基因组序列分析的经验基础上进行的。其理论意义则是利用基因组在进化上的连续性,通过研究较为简单的基因组而解答复杂的人类基因组的问题,探索基因组进化过程中“板块”的漂移与组合、基因的分布、基因结构的演变等。德国1995年启动的“人类基因组计划”,又把河豚鱼列入模式动物。这种动物的基因组只有人类基因组的1/7,却含有脊椎动物几乎所有结构基因。它在阐明基因的序列、基因的内含子序列的进化与可能功能方面,将发挥更加重要的作用。

5. 从代表性个体基因组计划到人类基因组多样性计划

基因组的相似性使人类基因组区别于任何“非人”基因组,而人类基因组的群体与个体的多样性,又使我们有了族群(ethnic groups)的区别,并使每个人类的个体成为各不相同的人。基因组计划在开始时,确实选用了—个无名氏男性的基因组材料,引发了当时的“谁应该是亚当”之争。

人类基因组多样性计划(HGDP)有两个内容:一是从个体代表性基因组扩展到各个族群的基因组,即群体多样性的研究。人类的进化史,都刻在现在仍保留的现代人基因组之中,通过HGDP,就有可

能揭开人类“从何而来”的奥秘,并能预测将来“由何而去”,并可为保留人类基因库(gene pool)多样性的呼吁与迫切性提供科学依据;二是从“代表性个性”深入到研究人的差异,以待将来形成每一个人自己的“基因组图”或“序列图”,作为每一个人用于个人保健、治病防病、预测病亡的遗传信息的个人档案。HUGO 已制定了全球性的 HGDP。我国将成为太平洋地区实施这一计划的最重要国家。

“人类基因组计划”的特点与“HGP”精神

10 年后的今天,当我们每天免费分享人类基因组计划所提供的宝贵数据的时候,我们不能不承认这一计划的科学依据与内在优势。人类基因组计划至少有下列 3 个特点,值得所有自然科学家引为借鉴。

1. 在研究策略上,人类基因组计划把遗传学升华为基因学(genetics)和基因组学(genomics)

细胞是生命的基本单位。而今,随着生物学向分子水平发展,使我们认识到各种生命现象都与基因的结构和功能有关。在某种意义上讲,细胞学在细胞水平上曾统一了被分为动物学、植物学、微生物学的“生物学”,而基因组学又在分子水平上统一了整个生物学。现在生物学、医学的各个分支都在讨论有关基因的内容,“基因”已经成为所有医学、生物学家的共同语言。无可否认,人类基因组计划以前的基因学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。

2. 在信息与材料上,实现了全球联网“信息”传递

以一段 DNA 或 cDNA 序列为“电子探针”,就可以实现“电子杂交”的梦想。其原理仍是分子生物学大多数技术的基石——DNA 的“碱基互补”原理,而其“杂交对象”则是国际所有基因数据库贮存的所有生物物种,包括人类在内的所有已知序列,其方法是通过全球联网的电子计算机进行“相似性比较”,其得到的结果是最客观的定量数据。从中可知,某一段序列是否是已知的某基因或一部份,它在多少个物种及组织中存在,与多少个、何个基因有多大程度的相似,有多大可能性是一段编码序列,可以编码几种、多长的“推测的氨基酸序列”,这些基因产物分别可能有几种“推测功能”,可能为哪种类型的蛋白质等。这一“电子杂交”的方法,已经成为基因鉴定、基因分类、基因组比较,以及基因与基因产物的结构、功能预测的最重要的工具之一。

全球性人类基因组计划的实施与全球化,借助了信息革命。人类基因组计划在人类自然科学史上

第一次建立了遍布全球的完整的数据库与信息网络。遗传图、物理图、序列图的构建能得以提前、完整地、成功地体现了信息高速公路的优越性。人类基因组计划也是自然科学史上第一次将物质的实验试剂变成信息的试剂。研究人类基因组的最重要的试剂——DNA 探针与引物,可以通过国际互连网络直接索取相应的 DNA 序列,并据以在自己的实验室进行合成。

3. 在计划实施上,人类基因组计划是人类历史上第一次由全世界各国科学家一起执行的科研项目

从实施人类基因组计划伊始,第一世界具有远见的科学家即号召全球各政府重视这一项目,并号召全世界科学家共同参与,建议所有的进展、所有的数据应随时公布于众,让全世界免费享用。在实施过程中,各国科学家精诚合作、共享材料、共享数据、共同攻关。这在人类自然科学史上,还是史无前例的。人类基因组计划与另两个有全球性意义的项目,即曼哈顿原子弹计划和阿波罗登月计划相比,更显示了人类的谐同与进步。

人类基因组计划在启动伊始,便重视其可能对社会、法律、伦理方面的冲击,特别注意这一方面的研究,并形成主流意见。特别是 HUGO 的几个重要声明,充分体现了现代自然科学的“求真”、“求善”及对社会的高度责任感。

我们应该颂扬这一“HGP 精神”,使它成为各国、各领域的合作楷模。

人类基因组研究已经取得的成果

人类基因组已经取得的成果可以用 4 张图来概括。

1. 遗传图(连锁图)

遗传图又称连锁图,它是以具有多态性(在一个遗传“位点”上具有多个等位基因)的遗传标记为“路标”,以遗传学距离(即在产生精子或卵子的减数分裂事件中,两个位点之间进行交换、重组的百分率 cM)为图距,反映基因遗传效应的基因组图。因此,建立人类遗传图的关键是必须有足够的高度多态的遗传标记。

经典的遗传标记是蛋白质或免疫学标记,如我们所知的 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,信息量不高,检测技术繁琐等因素,限制了人类基因组的遗传分析工作。DNA 技术的建立提供了新一代遗传标志,从而使经典的遗传学得到了新生。对人类说,这就克服了人类遗传学研究的“瓶颈”,突破了人类不能象“非人类”生物那样进行“选择交配”的客观限制。早在 1978 年当我国把“遗传工程”

列为八大科学项目之一时,我国著名的遗传学家谈家桢院士就精辟地说过:“我国应该补经典遗传学之课”。遗传制图计划的实施使我们不得不钦佩这一论点的正确。13年之后,全世界都在补这一课,也在“补课”过程中“温故而知新”,使经典遗传学发生了根本性的扩充与发展。

第一代的DNA标记是RFLP(Restriction fragment length polymorphism,限制性片段长度多态性)。限制性内切酶是一种能特异性识别并切割DNA链的酶。DNA序列的微小改变,甚至一个核苷酸的变化,就能使一个限制性内切酶的切点丢失或增加,导致酶切片长度的变化。并且可以用凝胶电泳来显示这一长度的“多态性”。RFLP遍布于整个基因组,为利用连锁分析方法,进行疾病表现型的定位提供了很好的标记。其最著名的成功应用便是发病率较高、病情严重的Huntington舞蹈病基因的定位。

RFLP的局限性是很明显的,首先它利用的是DNA的一个“点”上的突变。“点突变”一般只能形成“能切”与“不能切”两种状况,所产生的不同长度的片段(即所谓“等位基因”或“等位片段”)也只有两个,能提供的信息量有限;其次是检测RFLP需要以DNA片段作为探针,并且需要放射性同位素标记与分子杂交技术,技术复杂且不安全。

人们在检测RFLP的过程中,发现其中一种类型是“重复序列”造成的。在人类基因组中存在很多重复序列,即一段序列(一个“重复单位”)可能在基因组中有很多份拷贝,它们可能以正向(头—尾)或反向,头—头(尾—尾)串联成一簇,并分布于基因组的各个部位(即“位点”)。在某一位点上,不同数目的串联重复(VNTR, Various number of tandem repeats)也可提供不同的长度片段。人们把重复单位长度为6~12个核苷酸的VNTR称为小卫星(minisatellite),而把长度2~6个核苷酸的VNTR称为微卫星(micro satellite)或“简短串联重复(STR)”。

STR有两个最突出的优点:首先是它的高度多态性。由于CA等简短重复在进行上不受选择,因而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基因,是其他任何遗传标志所不能比拟的;二是可采用PCR技术(一种利用DNA聚合酶的“连锁反应”进行DNA体外扩增的技术),使操作可以实现机器化、自动化、电脑化以至完全机器人化。1991年,遗传标记开始可以用自动化操作。1992年有了综合机器和电脑的“GENOSCAN”自动系统。1994年美国的MIT(麻省理工学院)人类基因组中心等已经有了一次能进行15万个分析,从加样到结果输出数据分析全部自动化的系统,才使物理图和遗传图的构建达到惊人的神速。

STR作为遗传标记使人类基因组的连锁分析与遗传制图发生了革命性的变化。由法国民众集资支持的CEPH进行开发的STR标记与提供的规范大家系材料,为人类基因组遗传图的建立立下了不朽的功绩。CEPH与美国国家卫生研究院几个中心合作,于1996年初,已经建立了有6000多个以STR为主体的遗传标记,平均分辨率(两个标记之间的平均距离)为0.7cM(分摩,即两个位点之间有1/100的机率进行遗传重组。在人类基因组中,1cM大致对应于1Mb的物理距离)的人类基因遗传图。今年,美国MIT的人类基因组研究中心负责人兰道(E. Lander)又提出了“第三代DNA标记”即“序列多样性标记”。在理论上,由于所有“遗传多态性”的分子基础都是DNA序列多样性,这种新的标记有可能达到人类基因组的多态性的“最大极限”,并可将最高层次的物理图,即DNA“序列图”与“遗传图”及其他物理图最终有机地整合、统一起来;在技术上,它摒弃了遗传标记分析技术的“瓶颈”——凝胶电泳,而把序列分析的最新技术用于统一的“基因组图”的构建。

遗传学界一直认为:基因繁多、世代漫长、个体有限、婚配“无序”的人类很难进行自身遗传制图。所以人类遗传制图也确实一直落后于其它生物。现今,人类居然首先有了一张最详尽的遗传图,这是人类自然科学史上最重要的事件之一,应该说是人类认识自身的一个最重大的飞跃!

人类遗传图建立的科学意义与实际意义是不可估量的。6000多个遗传标记作为“路标”(以后在现有成功的技术和经验上可以预期开发更多的遗传标记),已经能够把人的基因组分成6000多个区域,只要以“连锁分析”的方法,找到某一表现型的基因与某一标记邻近(即“紧密连锁”)的证据,就可以把这一基因定位于这一标记所界定的区域。

就疾病来说,遗传分析的最大长处与最主要的特点是摒弃疾病发生中漫长的、多因子参与的、复杂的生理、生化过程,把疾病看成是基因(一个或多个)决定的表现型。然后根据“疾病位点”与选定的“遗传标记”之间的遗传学距离,即两个位点之间可能发生“遗传重组”的机率,确定疾病这一表现型的基因在基因组中的位置。现有分辨率的连锁图(平均分辨率为0.7cM,92%区域的分辨率为5cM)已能满足单基因性状的定位要求,也为人类多基因复杂性状的定位奠定了基础。

2. 物理图

人类基因组的物理图是以一个“物理标记”作为“路标”,以Mb、kb、bp作为图距的基因组图。这一物理标记,以前是一个序列未知的DNA片段(DNA探针),现在是基于DNA探针的序列STS(序列标记位置)。迄今,已经有了15000个STS,平均

图距(即分辨率)已达 200kb。与遗传图一样,现有的物理图已把人类庞大的基因组分成具有界标的 15 000 个小区城。

不同于遗传图,物理图还有一个内容,是要以 DNA 的克隆片段连接成相互重叠的“片段重叠群(contig)”。“相互重叠”的含义是因为这些片段都证明含有同一个 STS 的对应序列。现在,用“酵母人工染色体(YAC)”作为“载体”的 DNA 片段(长度为几百 kb 至 2Mb),已经构建了总体覆盖率达 90%左右的“片段重叠群”。而在很多区域,已经构建了覆盖率为 100%、具有高度代表性的“片段重叠群”,使用的 DNA 片段载体是比 YAC 的 DNA 片段还要可靠(尽管长度要短一些)的“细菌人工染色体”(克隆长度为一至数百 kb)或 cosmid(克隆片段为 40kb 左右)。

以 STS 为路标的物理图与遗传图相互参照,可以把遗传学的信息转化为物理学信息(如某一区域的大小为多少 cM,可基本“折算”成某一区域大小为多少 kb),而“片段重叠群”则提供了该区域可以操作的基因组材料,即相互重叠、覆盖这一区域的 DNA 片段,用这些片段为实验操作材料便可进行这一区域基因组研究或寻找这一区域的基因。

70 年代人类自然科学的重大突破便是 DNA 片段的“克隆”技术,即一般所说的“遗传工程”的基本技术。克隆是指来源同一、相互之间遗传结构相同(或基本相同)的一个群体。DNA 克隆技术的含意是制备大量的完全相同的 DNA 片段。这一技术的核心便是克隆“载体(vector)”的设计与制备。其他技术主要为限制性内切酶与凝胶电泳。用于克隆 DNA 的载体,是人类史上第一个依照人类绘制的蓝图,将有关基因片段“拼接”而创造出来的生物有机体。

第一代载体是质粒(plasmid),这是一个环状的、能在一定条件下自我复制并使所装载的 DNA 片段随之扩增的 DNA 分子。它因制备方便、方法可靠而成为至今技术上最重要、应用最广泛的载体。其缺点是装载能力,即“插入片段”较小,不超过十几个 kb。

第二代载体是经过改造的病毒,主要是 λ 病毒。它有两个主要优点:(1)装载能力达到了 20kb 左右,(2)由于病毒“体外组装”技术的成熟,它具有病毒那样的高效转导能力,进入宿主菌的效率比裸露的质粒 DNA 分子要高得多,因此也得到了广泛的应用,特别是用于来源有限的情况,如以 mRNA 为材料的 cDNA 文库的建立。

第三代载体是兼有病毒与质粒优点的 cosmid。它的命名的含义是具有 λ 病毒的 cos 基因及相关结构的质粒。cos 基因及相关结构使 cosmid 能象 λ 病毒那样被“组装”而有很高的转导能力。它的基本结构与特性是质粒,因而具有质粒的制备方

便等所有特点。它的装载能力可达 30~40kb,迄今还是较为可靠的建立“片段重叠群”的主要载体。

第四代载体是 YAC(yeast artificial chromosome,酵母人工染色体)。YAC 的构建是人类认识真核生物(即主要特征之一是细胞有核膜的结构,因而使 DNA 分子主要集中在细胞核里)染色体的重大飞跃。它是完全由人类将染色体的关键部件——两头的两个与染色体完整性有关的“端粒”,中部的一个与保证染色体正常分离的着丝粒,以及最重要的负责 DNA 与染色体复制的“自主复制中心”——装搭起来的名副其实的“人工染色体”,其行为与天然染色体相似,也能准确复制,并随着宿主细胞(酵母)分裂而分离。其独特优点是装载能力可达 1~2Mb,非前三代载体所能比拟。

YAC 这一大片段 DNA 载体的应用,还有两项重要的“配套”技术,即“稀有”切点限制性内切酶(在基因组中切点较少,因而得到酶切片段较大的限制性内切酶)与脉冲场电泳。在前者的发现中,我国的强伯勤院士曾做出过很大贡献。后者仍采用常规的琼脂糖作为电泳基质,但用特殊装置把分析能力扩大到 1~2Mb。上海陈竺院士引进 CEPH YAC 文库而建立的上海 YAC 分库,以及他所做的复制、保管、建立筛选手段等工作,已经为我国的人类基因组研究作出了重大贡献。

YAC 后载体是在前四代载体的基础上,在理论上博采百家,在技术上加以组合而构建的。现有 BAC(bacterial artificial ohromosome)、P I(一种源于 λ 的载体)、PAC(来源于 P I 的人工染色体)、Fosmid(一种类似于 BAC,来源于大肠杆菌的 F 基因的载体)、MAC(一种类似于 YAC,但以哺乳细胞作为宿主细胞的新型载体)。这些 YAC 后载体各有优缺点。现在 BAC 与 P I 已投入实际应用。其优点是装载能力居中(80~200kb),但方便可靠,没有象 YAC 那样因插入片段超大而带来的“嵌合体”等复杂问题。

这里要特别介绍的是不需要任何宿主菌或宿主细胞的 DNA 片段离体克隆技术——PCR。遗传图与物理图的建立,在很大程度上归功于 1986 年建立的 PCR 技术的应用。PCR 技术的推广,与 1987 年发现的耐热性 DNA 聚合酶与 DNA 引物合成仪的改进与普及分不开(第一台 DNA 合成仪于 1983 年投入使用)。PCR 的设备——热循环机,由电脑控制,机器人操作,数据直接输入并加工。它可以从原先的一次作 24 个反应,发展到 48 个、96 个、96 $\times$ 16 个以及现在的一次完成 15 万个。

染色体是遗传物质最自然、最合理的组装单位,是每一个物种的最重要的遗传特点之一。染色体上所显示的区、带、亚带是基因组中最客观、可靠的遗传与“物理”标记。因此,物理图的所有标记,不

管是基因,还是一个 DNA 片段(探针),还是一段序列(STS),都应该定位于某个染色体上的某个区域、某条带或某条亚带上。“遗传图”上的所有遗传标记也一样。因此,作为人类基因组的最基本层次的“细胞遗传图”既是人类基因组的“物理图”的组成份,又是统一“物理图”与“遗传图”的根本之图。

细胞遗传学的发展,主要经历了染色体分析、常规带型分析、高分辨带型分析等阶段。而在这些历史过程中,有很多我们中国人作出的贡献,如徐道觉发现的“低渗”法使三维的染色体成为平铺的二维图像,使细胞遗传学走出了“黑暗时期”;庄有兴等确定人类体细胞的染色体正确数目(46条);高法恬的人鼠杂种及染色体显微技术等重大贡献;湖南、北京国家医学遗传中心所分析的数以千计的染色体改变,以及笔者与同事们一起应用“高分辨带”对首例男性 X 染色体缺失及另一例男性 X 染色体长臂的一个微小缺失的分析(这些研究为无脉络症及 X 染色体失活等重要位点的定位提供了细胞遗传学依据而被反复引用)、对 5 号染色体短臂“自然发生断裂点”的分析等。

80 年代出现的 FISH(荧光原位杂交)及有关技术(见封二图),能使 DNA 片段直接定位于染色体的某个位置,因而在分子遗传学与细胞遗传学之间建立了桥梁,使细胞遗传学进入了分子细胞遗传学的新阶段。90 年代初期,建立的间期细胞核及染色体纤维 FISH 技术,使原先限于细胞有丝分裂中期的经典细胞遗传学(即染色体学)发展成为可以研究全细胞周期各个时期的遗传物质的“名副其实”的细胞遗传学。笔者与同事们在国内创立的单倍体(精子)与多倍体染色体纤维 FISH 技术,即属于“全周期”FISH 技术。

体细胞遗传学是细胞遗传学的一个分支,主要是由活细胞冷冻技术、组织与体细胞离体培养技术、细胞杂交与染色体分离等技术组成。人与鼠类的细胞在体外以人工技术诱导而融合成新的“人—鼠”杂种细胞,使人类首次打破了自然界的“种间屏障”。根据人类染色体在“人—鼠杂种细胞”中逐步丢失的现象,我们有了新的定位在细胞水平上表现的表现型(如某些蛋白质,特别是酶)相关基因的方法,并建立了含有人的单个染色体甚至一个染色体片段的“人—鼠”杂种细胞系,为人类基因组以及 DNA 片段的定位做出了重大贡献。当今在建立以 STS 为标记的“物理图”中所用的“辐射杂种(RH, radiation hybrids)”便是将单个人类染色体以辐射打成碎片,再将其整合进鼠类染色体而建立的新一代含有人类微小染色体片段的“人—鼠”杂种细胞。

因此,完整的各个层次上的“物理图”包括人类基因组的细胞遗传学图(即某染色体的某区、某带、某亚带,或以人为的等分图即以染色体依长度的百

分率作为“标记”),人类基因组的“大片段”限制性内切酶切点图, DNA 片段(探针)或一段 DNA 序列(STS)的“标记”图,以及基因组中广泛存在的特征性序列(如 CpC 序列、Alu 序列、isochore 等)的“标记”图,不同载体 DNA 克隆片段重叠群“图”,最终的分子水平上的序列图。

3. 序列图:

人类基因组的核苷酸序列图也就是分子水平的最高层次的、最详尽的物理图。测定总长 1 米,由 30 亿核苷酸组成的人类基因组序列图是人类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。

就历史来说,连锁图与物理图的构建是作为序列图的绘制的手段的,只有把庞大的基因组分成若干有标记的区域,才可能着手序列分析。STS 也就是作为一个区域的起始标记而出现的。序列分析需要一个区域的“片段重叠群”作为材料(称为模板)。由于目前阶段测序技术的限制,所需模板的长度只能为几百 kb 的质粒克隆片段。从某种意义上说,测序策略的基本点便是 DNA 小片段重叠群的构建,以及尽可能地降低重叠部分所占的比例。

测序的技术主要采用“酶促末端终止法”。对这一技术的改进与自动化是测序的主要难关,现已有不少新的,可能为未来很有前途的设想与方案,如寡核苷酸反向杂交法、毛细管电泳法、焦磷酸产物收集分析法、末端核苷酸残基收集分析法,以至于用 Scanning tunneling 电镜直视观察法等。

1977 年,人类完成对自然界第一个基因组,即 Φ X174 噬菌体(全长 5.3kb)的全序列分析,用了差不多 5 年时间。1987 年,第一台自动序列分析仪投入使用。1990 年,完成了长 230kb 的人的细胞巨大病毒(HCMC)的全序列分析。人类基因组计划在开始启动时,估计至少要 30 亿美元的投入,按照当时的估算需 15 年完成,工作确实有很大的难度。

近几年来,由于策略的成熟,寡核苷酸人工合成与自动测序仪的改进,测定速度比以前提高了好几个数量级。美国的圣路易华盛顿大学与英国的伦敦 Sanger 中心预计,仅凭他们两个实验室的经济、技术实力与经验,即使没有新的技术突破,于 3~5 年内完成全基因组的 90%以上的测序任务仍是很乐观的。

4. 转录图

迄今认为所有生物的性状,包括疾病都是由结构或功能蛋白质决定的。而已知的所有蛋白质都是由 RNA 聚合酶指导合成的带有多 A(多聚腺苷酸)尾巴的 mRNA(信使 RNA)编码的。如果能把 mRNA(或据 mRNA 的信息人工合成的 cDNA)先分离,就等于抓住了“大头”。因此,一张人类基因组的转录图或 cDNA(包括 cDNA 片段,即 EST)可表达序列

“标记”图,就是人类基因组图的雏形。

转录图的另一优点是:整个人类基因组中,只有2~3%的序列直接为蛋白质编码,在人体每一种特定组织中只有10%的基因表达,即只有不足1万个不同类型的mRNA(只有在胎儿脑组织中,可能有30~60%的基因表达)。根据mRNA的特点,可能用一种通用性引物(与多A尾巴互补的寡T或克隆载体的相关序列引物)以自动或手工测序仪,一次测定mRNA双端尾侧的几百个bp。这一测定EST(可表达序列)序列的策略已被好几个实验中心采用。现在国际数据库中所贮存EST的数量正以每日1 000多个的速率增长,至1996年3月,至少已有60万个,序列总长近100Mb,与1万多个已知基因有不同程度的相似。经比较、组合,去掉相同的序列,现已具有2.7万个各不相同的独立单位。

转录图有它本身特定的意义。首先由于cDNA是有组织与时间的特异性的,有人提出可以绘制一张反映在正常或受控条件中表达的全基因的时空图。在将来的数据库中,我们可以了解某一基因在不同时间不同组织、不同水平的表达;也可以了解一种组织中不同时间、不同基因中不同水平的表达,还可以了解某一特定时间、不同组织中不同基因不同水平的表达。有了“正常”的转录图,就奠定了构建特定生理条件下(如受外源的病原、药物、食物、精神的刺激)与“异常”(病理情况)下cDNA图的基础,为步入21世纪的医学—基因医学绘制了新的蓝图。

除这一特定意义外,转录图还有多方面的意义:(1)它能为估计人类基因,或基因家族(指结构或功能相似的几个基因)的数目,提供较为可靠的依据。而相应的准确数目只有在所有基因都克隆、鉴定后才能知道。(2)它能提供不同组织(空间),在不同发展阶段(时间)基因表达的数目、种类及结构、功能的信息,如已知cDNA的序列中,40%的基因与细胞的结构、细胞的基本能量代谢、细胞内环境自我调节(homeostasis)有关,22%与核酸、蛋白质的合成、加工、代谢有关,12%与细胞内或细胞间的“讯号转导”有关。(3)提供了功能基因的“标记”,因为一个EST本身就是某一特定基因的编码序列的一部分,提供了克隆、分析功能基因的一个起点。(4)它本身就具有一定的经济价值,如作为基因诊断或基因克隆的一种工具。(5)这是序列分析效益最高、收获最快的方案。正因为如此,在cDNA图的构建,特别是EST的分离方面的竞争十分剧烈,另外,关于EST的序列是否应公之于众,EST能否专利的问题的争论已近乎白炽化。

但是,要完成人类基因组的转录图,还需要将这数以几十万计的EST在人的基因组中定位。EST的大规模快速分离,自己给自己制造了“瓶颈”;越

来越多的EST急待定位。由于技术的限制,现在直接定位一个EST在基因组中的位置的成本约为140~170美元。但是,间接定位,即通过EST来源的基因组片段来确定EST的基因组的位置的技术,已取得了重大进展,约有70%可通过与基因组片段的序列比较而间接地定位。由相互重叠的EST组成的重叠片段群THC(“tentative human consensus” expressed sequence,暂定人类一致性表达序列群)已达3万多个,它们在基因组中的位置已大致或接近确定。

在人类历史上,格雷(Henry Gray)的第一张人体解剖图不仅解开了人体之谜,也奠定了近代医学的基础。而人类基因组计划所提供的这4张图,组成了人类的第二张解剖图,它揭示了人类基因组之谜,也将奠定21世纪生物学、医学进一步发展的基础,必将带来生物学、医学新的飞跃。

基因鉴定——人类基因组计划的重要内容与“副产品”

如前所述,基因鉴定在1993年被美国正式列入人类基因组计划的重要项目之一。基因鉴定,特别是重大疾病基因的克隆与鉴定,一直是最能反映人类基因组计划的社会与经济意义的。能够反映自然科学基础研究变为实际应用与经济价值的实例,才能吸引民众注意力、动员民众的参与感、争取民众的支持。

经典的分离疾病相关基因方法的起点是“功能”。首先根据疾病发生过程中生理与生化变化,寻找与这些变化相关的蛋白质(或酶),再对蛋白质结构,特别是一级结构即氨基酸序列进行分析。依据推演的该基因的DNA核苷酸序列,来得到cDNA或基因。但由于大多数疾病的生理与生化变化非常复杂,蛋白质功能的分化更为复杂,因此经典方法(一般统称为“功能克隆”)具有很大局限性。

分离疾病基因的关键是建立基因与表现型(疾病)之间的关系。伴随人类基因组计划而发展起来的“定位克隆”策略,则完全根据人类基因组计划所提供的连锁图与物理图,首先以连锁分析(或特异性染色体改变)确定疾病相关基因在人类基因组中的位置(定位),进而以“位置”为起点在初步定位的区域里搜索可能存在的基因,再进而鉴定这些候选基因的疾病相关性改变,如点突变、缺失及其他重排、实核苷酸重复、组织特异改变、离体或活体功能鉴定,来分离、克隆这一疾病相关基因。

运用“定位克隆”(曾被称为“反向遗传学”方法)策略的首次成功的例子是DMD(杜氏肌营养不良症)基因的克隆。而完全运用“连锁分析”与“定位克隆”分离得到的第一个基因,是白种人中发病率

最高(约为 1/2 000)的遗传病,囊性纤维变(CF)的基因。这一基因的成功克隆,使人们更进一步地认识到遗传图的重要意义,以及疾病大家系在基因克隆中的价值。随之分离到的最重要基因,是历经 10 年艰辛,才从连锁定位到成功分离的 Huntington 舞蹈病基因 Huntingtin(1993),以及对辐射敏感、细胞周期的重要调控因子之一、与多种癌症发生有关的 A—T(共济失调与毛细血管扩张综合症)基因 ATM(1995)。

以“定位克隆”途径克隆到的经典遗传病基因已达 70 多个。罕见的经典遗传病基因的克隆,其意义并不局限于这一罕见疾病本身的研究诊断与治疗价值,更重要的是这些基因的发现。如着色性干皮病(XP)基因的克隆,使我们得到了与 DNA 损伤修复有关的基因,因而在揭开癌症发生的奥秘中发挥了很大的作用;视网膜母细胞瘤(Rb)基因的克隆,使我们得到了一个在细胞周期中起主要调控作用的基因;ATM 基因的克隆帮助我们了解了人体的抗辐射、抗癌机制。正象有人说的:没有一种疾病“罕见”到没有克隆的价值。

对于“与疾病相关的染色体改变”来说,有时一个家系的一个病人便起了很大的作用。如 DMD 是一种单基因引起的经典的 X 染色体“连锁”遗传病。但后来发现了一个 DMD 病人(还伴有其他病症)在 X 染色体的短臂 2 区 1 带(Xp21)有一个很小的“缺失”区域,哈佛大学的一个小组利用这一病人的 DNA,以“定位克隆”的途径克隆到了 DMD 基因。另一例子是神经纤维变(Neurofibromatosis, NFI),也是一种经典的单基因遗传病,有人发现了一个典型的病人,其带有一个涉及 17 号染色体的“易位”(一个染色体的一段接到了另一个染色体上的改变),这就提供了 NFI 的相关基因位于 17 号染色体上的线索,据此线索最终克隆了 NFI 基因。

“定位克隆”的问世,开辟了人类基因克隆的新纪元。它使经典的“功能克隆”所不可能克隆的重大疾病基因逐渐被克隆,它的策略已相当成熟。可以说,在用遗传方法或细胞遗传学方法将疾病基因初步定位后,任一疾病基因的克隆都可以用这两个策略如愿以偿。但是,目前“定位克隆”的策略还需时很长,耗资很大。Huntingtin 基因克隆历时 10 年,ATM 基因克隆经历了 8 年,这还是在以“连锁分析”将基因所在的区域准确地基本定位之后所需的年月。而策略的正确则举足轻重。乳腺癌基因 1(brcal)在两年之内,便由美国几个公司与实验室合作克隆成功。而一直研究这一肿瘤,搜集了最多的家系,并将该基因区域准确定位的另外几个著名实验室则在竞争中失利。将“疾病相关基因”定位则有待于诊断明确的大家系,确定疾病相关染色体畸变材料的发现以及高超的策略。当然,我们也不应夸

大这一策略的艰难性。相反,由于技术的改进与新的思路的出现,有时很容易克隆。其中最快捷的实例是荷兰的韦门加(Wijmenga)到一个实验室访问,利用该实验室的家系材料,在 6 个星期之内便将 FSHD (Facioscapulo humeral muscular dystrophy)区域定位到 4 号染色体上;而另一著名的实力雄厚的实验室虽然努力,历时近 10 年还是在竞争中败北。

随着人类基因组计划的进展,“候选克隆”也将日趋重要。“候选克隆”以已知的“候选基因”为起点,在将某一疾病基因的位置定到某一区域后,再将这些区域的所有功能未知的基因、cDNA 或 EST 作为现成的“候选者”而直接以与“定位结构”相同的策略来进行疾病相关改变的鉴定。“候选克隆”的另一思路是根据某一疾病的已知病理、生理知识,将功能可能相关的已知基因作为“候选者”,同样直接进行疾病相关改变的筛选与鉴定,以这一途径克隆的基因已达 50 多个。随着 4 张图的发展,特定区域的功能未定的基因、全长 cDNA 或 EST 越来越多,“候选克隆”还将变得更加重要。虽曾有以此途径“三分钟”克隆一个基因之说,但是在很大程度上依赖于机遇,当然更有赖于基因组计划的全面进展、基因组信息的全面阐明来提供更多的“候选者”。

随着“第二张解剖图”的趋于详尽,人类基因组计划的基因组数据已成功地用于克隆常见的复杂性状的多基因病。近年来的大肠癌易感基因、乳腺癌易感基因、肥胖基因的克隆,哮喘与糖尿病基因的定位,都是基于人类基因组计划的进展而采用定位克隆的策略。使用近 400 个遗传标记进行的“全基因组扫描”,已在 I 型糖尿病等复杂性状的定位上作了很有意义的尝试,将发展成为克隆复杂性状相关基因的重要方法之一。人类基因组计划的成果,将成为现代生物学、医学用之不竭的源泉。

不管是“定位克隆”还是“候选克隆”,单基因还是“多基因”病,遗传病的大家系或具有与疾病相关的染色体改变的病人都是关键,与哮喘有关的基因定位便是一个很好的例子。

南大西洋的 Tristan da Cunha 的 Loneliest 岛上仅有 301 个居民,全部是近亲婚配的堂表兄弟姐妹,1/3 的人患有严重哮喘。这一家系,连同后来包括中国浙江象山的一个大家系在内的几个大家系,定位了与哮喘相关的至少两个位点。正如有人所说的,找到隔离群与大家系,已成为“如今人类遗传的全部所在(This is human genetics is at the moment)”。在芬兰发现的诸多大家系,已成功用于至少 12 个基因的克隆,且 90% 的患者都有相同的疾病相关突变,为基因克隆作出了重大贡献,也为利

(下转第 29 页)

国较有影响的各学科全国性学会。这些全国性学会也都分别参加了各该学科的国际性学会,而各学科的国际性学会联合会 ICSU 就是这些国际性学会的联合组织。我国派有正式代表参加 ICSU 的活动。

关于论文作者的荣誉与责任

大家已经习惯于根据一位科学家所发表的论文评价这位科学家的学术水平,并以此作为提职晋级、评定各种奖励和荣誉的依据,这些无疑都是正确和必要的。但是人们在注意到论文作者荣誉的同时,常常会忽视论文作者应负的责任。国外有的刊物,在接受一篇论文发表时要求每一位作者在投稿信上签名,以示负责。虽然每一位作者的侧重点不同,但同样都应应对论文负责,导师或实验室负责人更应该全面负责。

国外有些实验室有一种陋习,就是无论对某项工作是否确有贡献,导师在所有学生的论文上挂名,实验室负责人在该室所有论文上挂名。更有所谓互送论文的做法,即虽然完全没有参加工作,但此次我让你在论文上挂名,以换取下次你让我在你的论文上挂名。据说还有一个专用术语,谓之“搭车”。据 SCI 统计,发表论文的世界记录是每年 948 篇,平均每天 2.6 篇(包括节假日在内)。如果说其中不少是仗势搭车,恐怕是不过分的。这种做法,强调了作者荣誉的一面,而忽略了论文作者应对论文负责的一面。还应当指出,借口“负责”而在自己没有实质性贡献的论文上署名,也是不应该的。更有甚者,如果只享受论文作者拥有的荣誉,而对论文中所产生的问题拒绝承担责任,则是一种不负责任的行为。应当指出,这种不正之风已经在我国学术界有所表现。

美国洛克菲勒(Rockefeller)大学校长、诺贝尔奖获得者巴尔的摩(Baltimore)教授,由于他署名的一篇论文被揭发数据有弄虚作假情况,美国有关部门照样进行严格调查,经调查属实。虽然巴尔的摩本人对具体工作中的弄虚作假并不知情,但作为作者之一,仍不得不对此负责并引咎辞去洛克菲勒大学校长职务。我国是否也有大大小小的,严重程度各不相同的“Baltimore 事件”呢?各种严重违反科学道德的事情,我国不仅有,而且还不少。不幸的

是,虽然被揭发对象远没有巴尔的摩的声望,但我国对于不良行为的揭发者却没有给以应该给予的公正对待,对所揭发的问题经常采取捂盖子的做法,揭发文章常被禁止发表,以致完全剥夺了对问题进一步讨论的可能性。揭发者搞不好虽然不致于丢掉饭碗,但穿小鞋的味道也是不好受的。应该严肃指出,对违反科学道德的行为采取保护或捂盖子的做法,只能助长科学界的歪风邪气,甚至把青年科学家引入歧途。关于这一点,难道我们不能在“Baltimore 事件”中,从正反两方面,吸取什么有益的教训吗?

避免行政干扰

正确评价一项科研成果首先要判断结论是否成立,这包括实验路线方法是否合理,实验数据是否确切可靠,是否有必要的对照实验,证据是否充分等。前已提到基础研究贵在创新,这就必须充分了解国内外这一领域的工作基础和现状,并实事求是地在论文中确切交代前人已做过的工作,是否确为国际上首创。这些高度专业的问题只有依靠科学本身,依靠同行科学家的集体审议,才能做出正确判断。任何一位领导,即使本人也是有高度成就的科学家,也无法对所有专业的每一个细节问题都了解清楚,因此在这一过程中务必坚决避免行政干扰,我们衷心希望各级领导同志能对此严格把关。

当前我国科学界的一种不正之风,就是一旦取得一些结果,首先想到的不是同行意见,而是领导看法;不是在国内外核心刊物上发表正式的研究论文,以求在发表前先经过同行审稿人的严格评审,发表后再听取广大同行的意见,而是首先用夸张的语言向领导报喜,一旦通过领导批准,就迫不及待地通过新闻媒介广泛宣传,取得廉价的新闻价值,以此为谋取名利的手段。这不是严肃的科学态度,也不是一个正派的科学家所应该做的。

建国 40 余年来,经领导推荐、报刊宣传,有的甚至自己宣称将获诺贝尔奖等的所谓“科学成果”曾屡见不鲜。但是科学是严肃的,时间是无情的,这些被吹捧上天的工作现在绝大多数都已成为过眼烟云,昔日的宣传徒然成为人们偶而提及的笑柄而已。

(责任编辑 蔡德诚)

(上接第 18 页)

用“隔离群”中可能存在“连锁不平衡”的价值,提供了很好的例证。这些家系最后都追踪到同一个祖先,其“始祖效应(Founder effort)”非常显著。“连锁不平衡”区域竟大至 9~13cM。

我国地广人多,由于历史、地理、经济、文化等原因,形成了中华民族的 56 个族群与众多的隔离

群、近交系与大家系,以及已经发现与急待发现的遗传病家系和染色体改变病例。这些都是人类基因克隆、基因组多样性研究的最为难得的珍贵资源和材料,我们应该发挥自己的聪明才智,利用这些珍贵材料为全人类做出贡献。

(责任编辑 王宏章 蔡德诚)