

## \* 学科发展 \*

# 对开展中国人类基因组多样性研究的思考\*

杜若甫

(遗传研究所 北京 100101)

**摘要** 人类基因组计划是目前世界上宏伟的科学计划。文章对我国开展人类基因组多样性研究的意义、内容、策略、注意事项、预期成果等提出了看法。

**关键词** 人类基因组, 多样性, 研究策略, 中华民族

人类基因组计划是目前世界上对人类影响极大而且极其深远的宏伟科学计划, 进展极其迅速, 至 1996 年 10 月, 制图的基因已达 16 000 个。

众所周知, 每一个人的基因组都不一样, 每一个人群、民族、人种的基因组也都有其特点。目前世界上人类基因组计划所研究的材料只是白种人的基因组, 要真正了解全人类的基因组, 还需研究与比较不同人种、民族、人群的基因组, 了解他们的特点。而研究中华各民族的基因组的特点, 责无旁贷地主要落在中国科学家的肩上。

中国人类基因组多样性的研究, 对于我国防病治病, 提高人民身体素质, 了解中华各民族的源与流, 以及现代人的起源等重大问题, 都具有十分重要的意义。

对中国不同民族与人群中疾病的易感性与抵抗性, 对药物和环境因子的敏感性与抵抗性有关基因的多态性的研究, 对于防病治病及提高我国人民身体素质具有十分重要的意义。例如过去我国对血红蛋白(与先天性贫血症密切相关)、葡萄糖-6-磷酸脱氢酶(与蚕豆病有关)、白细胞抗原系统、乙醛脱氢酶(与酒量有关)、乳糖酶(与对牛奶中的乳糖的消化力有关)等的研究, 都说明我国各个民族与人群之间在这些基因的多态性方面有很大差异, 与蒙古人种等其他民族及白种人、黑种人差异则更大。但是对于许多重要基因在我国各民族与人群间的差异, 至今研究得还很不够。例如, 海南岛的黎族对当地的钩端螺旋体病有很强的抵抗力, 而汉族则极易感染这一疾病。究竟是什么基因在其中起决定性作用, 尚不清楚。

现代人起源是当前世界上尚未解决的重大科学问题。“非洲起源说”认为现代人十多万年前起源于非洲, 而“多地区起源说”则认为亚、非、欧等洲的现代人是当地智人演化而来的。中国的大多数人也认为中国人是 50 万年前北京猿人的后代, 对中国人类基因组多样性的研究所得

结果,与白种人、黑种人的相应结果进行比较分析,可以对这些问题提供更多的证据,并最终与考古学、人类学、语言学等研究成果结合起来,解决这一问题。

中华各民族的源与流是广大中华儿女所十分关心的问题,也为世界上人类学家与民族学家所密切关注。过去一些研究已初步表明藏族起源于北方,新疆维吾尔与哈萨克等民族基本上应归属蒙古人种,这些初步结论在目前都有重要的现实意义。

下面,对于 21 世纪到来以前,如何开展中国人类基因组多样性研究,从四个方面,就其重要性、原有基础、工作内容及应注意之点,提出一些看法。

## 1 继续建立中华民族永生细胞库

中华各民族是若干万年来人类进化所形成的。现有民族与人群是研究中华各族的遗传结构、现代人起源、中华各民族的源与流等非常珍贵的材料。直至我国经济腾飞之前,由于居住地域、宗教信仰、语言、生活习惯等不同,绝大部分民族仍处于相当严格的隔离状态。近年来,随着经济、交通、通讯、文化等各方面迅速发展,人口流动与原人群的分散及民族相互通婚等均迅速增加,各民族原来的群体结构正在消失之中。因此,迫切需要建立我国各民族与人群的永生细胞库,将他们遗传结构的样本保存起来,以供今后千年万代进行研究。

自从中国人类基因组重大项目开始启动以来,我国已有 13 个民族(或人群)建立了永生细胞库,每一民族(或人群)约为 55 个个体。这些民族与人群是:广东广宁汉族与陕西麟游汉族;少数民族有彝族、鄂伦春、鄂温克、藏、黎、傣、景颇、德昂、布朗、哈尼(爱尼支系)、拉祜等 11 个。

汉族是在历史上融合了许多古代少数民族形成的,因而是一个非常异质的群体。汉族明显地以长江为界分为南北两大群,分属南方蒙古人种与北方蒙古人种,而且各地汉族与当地主要少数民族在遗传结构上十分相似,而南、北汉族之间却差异很大。因此,对汉族有必要建立更多一些有代表性的人群的永生细胞库,初步考虑有:山东东部、浙江、四川中部、江西中部、福建中部、吉林、山西北部、河南中部等 8 个人群。

在汉族中,还有一些特殊的人群,如浙江绍兴,广东、贵州的部分,海南之临高人及西部讲春话的汉族等,对这些人群是否需建立永生细胞库,可放到三四年后考虑。

在少数民族中,还有 44 个民族的永生细胞库有待建立。但其中塔塔尔族人数少且居住分散;俄罗斯族与赫哲族人数少、分散,而且有相当一部分已与汉族等其它民族通婚;保安族人数少,而且长期与当地回族及东乡等族通婚;西藏的门巴、珞巴两族与云南的独龙族聚居地交通不便;台湾的高山族(实际上是 9 个民族)目前取样困难。以上共计 8 个民族的永生细胞库可暂缓建立。其余 36 个少数民族的永生细胞库应争取在三年内建立。

还有一些“未识别民族”,如西藏的澄人与夏尔巴人,四川阿坝地区的白马藏族(氐人),云南的科木人、芒人,广西的莱人(可能是留在大陆的黎族)等等,也只能留待以后考虑是否应建立他们的永生细胞库。

这样,在 21 世纪到来之前再建立汉族与少数民族的 44 个人群的永生细胞库,即可基本建成中华民族的永生细胞库。

中华民族的永生细胞库是我国人类学、医学、遗传学、民族学等研究的宝贵财富,必须保

存好,同时将其充分地利用起来,发挥其作用。因此,必须建立和逐步完善我国永生细胞库保存和利用的制度,并帮助有关单位建立和完善保存的必要条件。为此,建议:

(1) 尽快召集有关单位开会,专门讨论一次我国永生细胞库及永生细胞株的保存与利用的有关问题,拟出初步制度、办法,必要时签订协议书。

(2) 帮助保存单位建立保存所需之基本条件。从长远看,还需落实今后长期保存的日常开支与人员的经费。

## 2 积极开展我国人群 DNA 多态性的研究

DNA 多态性研究有两个方面:一是结构基因易变区的 DNA 序列分析,另一是 DNA 链上非编码区的序列,尤其是微 DNA 或小 DNA 序列的研究。

DNA 多态研究的主要优点是其精确性。过去对人类基因组多样性研究所用的一些遗传标记,有的是用免疫化学方法来定型的,如红细胞血型与白细胞抗原系统;有的是主要用电泳方法来分型的,如许多蛋白质与酶。但是有些基因 DNA 分子水平上的变异型的产物,用免疫化学或电泳等方法都无法与其它变异型区分出来。而今天,许多结构基因中的重要易变区段,都已可在 DNA 水平上进行鉴别,因此,使分型的精确性达到前所未有的最高水平。

我国过去已对许多重要的血型系统、酶、蛋白质的多态性用电泳或免疫化学等手段进行过大量的研究,因此,对于结构基因的 DNA 多态性的研究首先可以在此基础上进行,使过去的分型更加精确、可靠。其次,由于世界上人类基因组计划的飞速发展,至今已有成千上万个新的基因被制图、测序,因此,已可对许多新发现的基因进行 DNA 多态性研究。可见,结构基因 DNA 多态性研究这一领域的前景是十分广阔的。

其次,对于非编码区的 DNA 多态性的研究,目前国际上也正在蓬勃开展,它对于了解现代人的起源与进化,研究民族的源与流,都是十分有用的。除了其精确性这一优点之外, DNA 中非编码区的多态性研究还有两个方面的突出优点:其一是它们绝大部分是中性变异,不受选择的影响,可以更准确地追踪其谱系。其二是线粒体 DNA 是母系遗传的,研究线粒体 DNA 的多态性可以追踪其母系谱系;Y 染色体是父系遗传的,研究 Y-DNA 的多态性可以追踪其父系谱系。而且无论线粒体 DNA 和 Y 染色体上拟常染色体区的非重组区,都是单倍性的,在世代传递中不发生重组,因此,可以比常染色体上的 DNA 序列更精确地追踪祖先系谱。

关于研究的非编码区 DNA 遗传标记数量,从理论上讲当然越多越好,结果越准确,但是由于人力、物力及时间的限制,研究的标记不可能太多。在目前,为研究现代人起源与进化和中华各民族的源与流,如果能有 15—20 个 DNA 遗传标记的多态性,就已较理想,其中线粒体 DNA 与 Y-DNA 均应各有 3—5 个,其余的是常染色体上的标记。常染色体上的标记可优先考虑容易做的,可进行多位点复合扩增的一些短串联重复序列,它们可同时扩增、分型,结果也比较可靠,多态性强。

在定下遗传标记后,应尽量坚持用相同的标记研究一系列人群,至少在几年内不变,以求积累中国人群的一整套数据。这样,这些数据才可以相互比较、分析,才能发挥出最大的作用,体现其价值。切忌这几个人群用这几个标记,那几个人群用那几个标记,数据零散,利用价值就低。

每一人群的个体数,最好有 200 人以上,至少应在 50 人以上。尤其 Y 染色体与线粒体 DNA 均为单倍体,人数更不应太少。

为节省研究费用与人力、物力,研究的人群应尽量是已保存有样本的人群或为建立永生细胞库而取样的人群。但从重要性方面考虑,在我国各民族中,应首先着重研究下列民族:汉族(应考虑有不同地区汉族人群)以及维吾尔、哈萨克、蒙古、回(宁夏回族)、满、朝鲜、鄂温克(索伦人)、藏、彝(原黑彝)、白、傈僳、哈尼(爱尼支系)、布朗、景颇(景颇支系)、傣(水傣)、佤、苗(青苗)、黎(本地黎)、土家、水、羌、瑶(盘瑶)、纳西、壮、侗等 25 个少数民族。

为了在较短时间(例如三年内)能以相同的一批遗传标记研究较多人群,每一人群有足够的个体数,最好全国开展协作,参加的单位有一个统一的计划,大家分工负责去进行。而且为了保证实验结果的准确性与可比性,最好单位间按标记分工,而不是按民族或人群分工,即全部人群的同一标记由一个单位负责分析。

### 3 继续进行中国人基因组经典遗传标记的多样性研究

所谓经典遗传标记指的是血型、酶、蛋白质等过去长期以来所研究的指标,甚至可包括耳垢、苯硫脲味盲等指标。这些指标与 DNA 多态性相比,虽然其精确性略低一些,但对研究疾病、环境因子的敏感性或抵抗力、现代人的起源、中华民族的源与流等,仍是非常有用的。因为这些标记已有大量民族与人群进行过研究,已积累有极丰富的资料。如果某一民族或人群过去未研究过这些指标,只要现在研究,就可与过去已积累的大量资料相比较,进行分析,对该民族或人群与国内其它民族与人群的亲缘关系得出看法。因此,对这些经典指标仍应继续进行研究。尤其,建立中华少数民族的永生细胞库的工作还正在开展,在获得的血液样本中,就可能有一些民族或人群过去未曾被很好研究过,可以利用建库的样本进行这一方面的研究。

与少数民族相比,过去对汉族各地人群的经典标记研究反而不够,只对一些常见的红细胞血型、白细胞抗原系统及血红蛋白等研究得较多。因此,在建立永生细胞库时拟采取样本的汉族 8 个人群,如有条件,都应对其经典标记进行研究。

在少数民族中,经典遗传标记方面研究得比较不够的有柯尔克孜、乌兹别克、哈萨克、塔吉克、裕固、锡伯、东乡、土、撒拉、仡佬、佤、水、毛南、基诺、景颇、京等族。而独龙、俄罗斯、赫哲、门巴、珞巴、高山、保安、塔塔尔等则因取样困难,暂时不可能研究。

有的少数民族人口多、分支多,而且分支间隔离较严重,差异较大,如苗族、蒙古族、藏族、鄂温克、黎族、瑶族、彝族、回族、壮族、怒族、哈尼族、傣族等,因此值得对他们的分支进行深入研究,对一些“未识别民族”也值得进行研究,有待今后做深入的研究工作。

### 4 对 DNA 多样性大规模快速研究技术的探索

为研究中国人基因组的多样性,必须进行大量的个体样本分析。而目前,分子遗传学的进步非常迅速,新方法、新技术层出不穷。因此,很有必要由一两个单位专门负责,对国际上有可能用于人类基因组多样性大规模、快速分析的方法技术进行观察、引进、探索与研究,并随时介绍到国内,加以推广应用。同时,我们自己也应开展这方面的研究,以求对这方面的方法、技

术进行改进、发展与创新。

在以上四方面工作中,第一和第二两方面是重点。预期在 2000 年初,可以得到下列主要成果:

(1)基本建成中华各民族的永生细胞库(10 个汉族人群、47 个少数民族,共 57 个群体),建立保存永生细胞库的基本条件和保存利用的制度。

(2)获得 40 个民族(或人群)15—20 个遗传标记的整套基因频率数据,同时获得经典遗传标记方面一批过去尚缺的人群基因频率数据,结合过去已有的近 2 000 套中国人群基因频率数据,进行计算分析,对中华各民族的源与流,对中国人究竟来自非洲还是北京猿人的后代,北方蒙古人种与南方蒙古人种是何时分开等问题,提出新的或更准确、全面的看法。同时,获得一批与疾病、环境因子的敏感性有关基因中我国特有或较多的一些等位基因在 DNA 序列水平的分型结果。

———— \* ————— \* ————— \* —————

## \* 简讯 \*

### 国家科委表彰水下机器人深海考察成功

**本刊讯** 国家科委于 1997 年 9 月 3 日召开了 6 000 米水下机器人深海资源考察汇报表彰会。受表彰的单位有:中国科学院沈阳自动化研究所、声学研究所和中国船舶科研中心、哈尔滨工业大学、上海交通大学、华东船舶工程学院、海军航海保证部。

考察工作是在今年 5 月 21 日至 6 月 27 日在檀香山东南 1 350 公里处的太平洋中部进行的,考察结果取回了近百兆位的数据、16.5 平方公里海底声学扫描以及录象带、照片等丰富的科学资料。试验表明,我国研制的 6 000 米水下机器人性能卓越,可靠性强,能据行赋予的使命。它的研制试验成功,使我国水下机器人从有缆到无缆、从潜深数百米到 6 000 米,成为世界上少数拥有这项高技术的国家之一。

(袁勤)