

# SNP—为人类基因组描绘新的蓝图

刘万清<sup>1)</sup> 贺林<sup>1,2)</sup>

(1 中国科学院上海生命科学研究中心, 2 中国科学院上海脑研究所, 上海 200031)

## SNP: Tracing New Blueprint for Human Genome

LIU Wan qing<sup>1)</sup> HE Lin<sup>1,2)</sup>

(1 Shanghai Research Center of Life Science, Chinese Academy of Sciences, Shanghai 200031)

(2 Shanghai Brain Research Institute, Chinese Academy of Sciences, Shanghai 200031)

几乎所有的慢性疾病都或多或少与遗传因素有关。在过去的二十年中,对人类遗传学的研究使人们认识到,理清各种疾病中遗传因素所起的作用将对诊断学、治疗学以及预防医学等各个领域带来革命。同时,了解遗传因素在疾病病因学中所占的地位也有利于人们对非遗传的环境因素对人类影响的理解。因此,对各种疾病基因进行定位和克隆就成为人类基因组研究的主流方向。

随着人类基因组计划的实施,遗传图谱、物理图谱以及各种确定基因的技术层出不穷,这对寻找各种遗传病,尤其是孟德尔方式遗传病的基因起到了巨大的推动作用。定位克隆战略自1986年首次提出后即被成功地应用于人类基因的研究上。到1990年,当人类基因组计划启动时,应用这种战略定位克隆基因的数目还很少,但是到1997年时,已有100多个以孟德尔遗传方式为主的重要的遗传病基因通过这种方法确定下来。

然而,目前的疾病基因研究策略仍然存在很多的问题,主要表现在以下几个方面:

(1)对复杂性多基因遗传病缺乏有效的手段能够迅速突破瓶颈,找到致病基因;(2)目前使用的方法如微卫星标记图谱基础上的连锁分析,由于对凝胶电泳等手段的过度依赖使之较难达到完全的自动化,而对于大样本的全基因组扫描,又不是所有的实验室都能够办得到;(3)由于商业利益的驱动,使越来越多的私人研究机构和公司介绍人类基因组的研究并抢占了越来越多的数据,这在一定程度上阻碍了研究者对人类基因组数据的使用;(4)分子遗传学领域中技术的进步,如DNA芯片或微列阵手段的不断成熟,对遗传学研究提出了新的要求和挑战;(5)基因组研究进入中后期,对于致病基因的分离与筛选,由于各种遗传学方法的逐渐成熟以及可利用的基因组数据的增加,要求有大规模迅速检测各种突变的方法与技术来完成大量候选基因中的突变筛选,以迅速获得所要的基因;另外,基因组研究的深入,对人类进化、种群多样性、基因与药物、基因与环境的研究逐渐成为下游研究的主要目标,这些研究方向均与DNA水平的变异直接相关。

在这种情况下,要求有更加完善的理论与各种新技术的结合,才能使遗传学的研究产生新的飞跃。SNP战略的正是在这一过程中脱颖而出,成为人类遗传学界关注的新焦点。

### 1 SNP(single nucleotide polymorphisms)简介

SNP即单核苷酸多态性标记,主要是指由基因组核苷酸水平上的变异引起的DNA序列多态性,包括单碱基的转换、颠换,以及单碱基的插入/缺失等。SNP的位点及其丰富,几乎遍及整个基因组。据估计基因组中大约平均每1000bp就会出现一个SNP,这样SNP在整个基因组的分布就会达到三百万个。

SNP标记在人群中只有两种等位型(allele)故亦称为双等位标记(biallelic marker)。这样在检测时只需一个“+/-”或“全或无”的方式而无需象检测微卫星标记那样对片段的长度作出测量。目前多种非凝胶基础上的SNP的检测手段如DNA芯片以及微列阵的技术已经开发出来并迅速被应用和趋于成熟。使用高密度的DNA芯片或微列阵可以同时对上千个SNP的分型。

SNP 在基因组内可以人为地划分为两种形式: 一即基因编码区的功能性突变, 主要分布于基因编码区(coding region), 故又称其为 cSNP; 第二是遍布于基因组的大量单碱基变异。

SNP 开发技术上, 过去的方法一般是建立于凝胶电泳基础上对多个个体进行分析的方法, 如测序、DGGE、SSCP 等, 速度相对较慢且价格昂贵。近来的技术和手段上的进步使得检测成本大大降低而且提高了检出率。估计每年可以检出上千个基因中的 SNP。目前, 几个相对有前景的半自动或全自动地进行大量 SNP 检测的方法已经初露端倪。包括小型测序、多重反向点杂交、DNA 芯片或微阵列, 以及 TaqMAN 的方法<sup>(1)</sup>。

## 2 SNP 开发的紧迫性

由于 SNP 作为一种研究工具具有极大的潜力, 所有 SNP 数据应立即公布和发表, 以让每位研究者都可以自由地使用。但在美国, 一些“专利专家”认为, SNP 尤其是编码序列中的 SNP(cSNP)具有“足够的新意和潜力”, 因而应当可以申请专利。因此, 如果 SNP 的开发得不到足够的指导及经费支持, 大量的 SNP 和 cSNP 就可以被大量的私人研究机构所开发和占有, 这对于其他研究者利用 SNP 就会产生掣肘。对于一些随机选择的 SNP, 这种担心在目前还少一些, 因为这些 SNP 遍及基因组各处; 而对于 200 000 个 cSNP 来说, 由于它们分布于基因编码区, 因而是各个私人研究机构追逐的首要目标。

自 1996 年开始, 美国的几个大的私人公司就先后制订计划进行基因组 SNP 的开发, 并试图将这些数据申请专利。如加利福尼亚的生物芯片制造商 Affimetrix 公司及 MIT Eric Lander 的实验室, 计划在 5 年内投资 4 000 万美元开发 2 000 个以上的 SNP(该组已于 1998 年 5 月份的 Science 杂志上发表了第一张人类 SNP 遗传连锁图谱, 包罗了基因组内的 3 241 个 SNP, 平均分辨率达到 2cM, 已有 2 227 个被定位和作图<sup>(4)</sup>)。而另一项大投资则是芝加哥的制药公司 Abbott 实验室与法国的 Genset 公司联合投资 6 000 多万美元进行 60 000 个 SNP 的开发, 并打算在开发后申请专利, 向其他研究者出售或用于制药研究。鉴于这种形势, 美国政府也加入投资以期达到与这些制药公司竞争的目的, 而防止过多的数据被私人专利后影响公共研究机构的研究。这些数据将转入 Genebank 或其它数据库以供研究者自由使用<sup>(2)</sup>。国立卫生研究院(NIH)已对外公开招标投资支持研究; 国立环境卫生研究院(NIEH)也宣布提供资金支持与环境相互作用基因的变异的研究。其主要目标是: 首先, 开发一套高密度的至少在 100 000 左右的 SNP, 并确定所有已知基因内部的常见 cSNP; 其次, 近来在遗传学研究的理论上, 尤其在对于复杂性状遗传病的研究中, 人们更趋向于利用连锁不平衡和关联分析的战略, 而 SNP 的开发恰好是实施这一策略的理想工具; 第三, 对开发 SNP 和 SNP 分型技术上的提高, 使得 SNP 的检测成本大大降低, 而且提高了检出率。估计其成本约为 100~1 000 美元/SNP 左右, 而这一成本又会随着一些快速检测手段的成熟而继续降低。

由此可见, 虽然人类基因组序列图谱的完成尚需时日, 获得这样一整套 SNP 目前看来还不是太现实, 而到 2005 年人类 DNA 序列图完成后, 这一数据将更容易得到。但无论是理论上的要求还是技术上的进步, SNP 都已成为人类分子遗传学领域的下一个研究热点, 因此, 对 SNP 的开发就具有时间上的紧迫性。

## 3 SNP 的应用

SNP 的用途前景极广, 目前讨论得最多的主要在以下几个方面:

第一, SNP 作为遗传图谱用于致病基因的搜寻。目前, 利用 SNP 对致病基因进行研究的策略主要集中于三方面: 其一, 构筑人类基因组 cSNP 图谱, 进行全基因组的关联分析, 搜寻多基因遗传病相关基因, 以及对环境因素敏感的 DNA 变异位点。其二, 构建人类基因组 SNP 连锁图谱, 对重要疾病进行连锁分析<sup>(3)</sup>; 其三, 结合上述的两种策略, 通过连锁不平衡等手段进行致病基因定位和克隆。

第二, SNP 应用于进化和种群多样性的研究。生物界的进化与进化过程中物种多样性的形成与基因组的突变和突变的选择密切相关。构建整个基因组的 SNP 图谱对于直接研究人类的起源、进化具有极大的意义。

对比人群之间的 SNP 图谱的异同情况可以对人类的起源、迁移等作出估计从而理清人类进化过程中源与流

的问题。有人已利用 DNA 芯片技术对人类线粒体的 SNP 图谱作了研究, 得出人类起源于非洲的结论。

第三, 利用 SNP 标记研究癌细胞的染色体缺失以及基因的杂合缺失。由于 SNP 在基因组中的高密度的特点, 与以前的微卫星或其它遗传标记相比, 利用 SNP 可以在上述的研究中对目的片段或基因作出更加精细的标定, 从而使研究不断深入。

第四, SNP 尤其是 cSNP 与人类的各种遗传特征直接相关, 开发 SNP 就意味着对每个人群甚至每个人进行 DNA 水平的鉴定和区分, 这必将使其在医学领域及法医学领域中得到广泛的应用。例如, 完善的 SNP 图谱将促使诊断学和治疗学产生革命性的进步, 人们可以利用 SNP 对每一种疾病作出更为准确的判断, 从而“对症下药”。

总之, SNP 作为一种新的遗传学战略和研究工具不但可以大大加快人类基因组的研究, 而且对于整个生物学研究和生物产业也将带来巨大的变化, 它的开发将为人类基因组绘制出更加精细、更加绚丽多彩的蓝图。

### 参 考 文 献

- 1 Collins F S, Guyer M S, Chakravarti A. Variation on a theme: Cataloging human DNA sequence variation. *Science*, 1997, 278: 1580~1581
- 2 Marhall E. Human Genome Project: The hunting of the SNP. *Science*, 1997, 278: 2074~2075
- 3 Kruglyak L. The use of a genetic map of biallelic markers in linkage studies. *Nat. Genet.*, 1997, 17: 21~24
- 4 Wang D G, Fan J B, Siao C J *et al.* Large-Scale identification, mapping, and Genotyping of single-nucleotide polymorphisms in the human genome. *Science*, 1998, 280: 1077~1082

1998-06-29 收稿, 1998-08-31 修回.

### Call for Chinese Participants

Dear Chinese geneticists:

The 18th. International Congress of Genetics was successfully held in Beijing, China on 10-15 August 1998. About 2,000 experts from 54 countries and regions attended more than 70 symposiums and discussions, exchanging ideas on the latest research in genetics-related fields such as agriculture, medical treatment, ethics, pharmaceuticals, longevity and environmental protection (China Daily 7,11,17 August 1998; People's Daily 11 August 1998; The New York Time pp.13-14,16 August 1998; Nature 394:707,711,20 August 1998; Science 281:1118-1119,21 August 1998).

To further promote academic exchange between Chinese geneticists and their western counterparts on the issues of the human genome project(HGP)including the advanced genetics technologies and their implication in society [Am J Hum Genet 63(3):688-695,September 1998]), a group of international experts of the World Health Organization on genetics and ethics (WHO 1998) intend to organize an international workshop on ethical, social, legal, and psychological implication of HGP and WHO's guidelines on ethical issues in genetics services in Chengdu, China next year. If you are interested in such workshop, please send a letter saying that you are willing to attend the workshop together with your CV to the following address before 2 January 1999.

Dr. Xin Mao  
 Section of Cancer Genetics  
 Hadow Laboratories  
 Institute of Cancer Research  
 15 Cotswold Road, Sutton  
 Surrey SM2 5NG, UK  
 Tel: 0181-643-8901 x 4661  
 Fax: 0181-770-7290  
 E-mail: xin@icr.ac.uk