

课题综述

SNPs 检测方法在不同种族遗传差异分析的应用

焦伟, 刘斐

基金项目: 广西自然科学基金项目(NO: 桂科自 0640059)

作者单位: 530021 南宁, 广西壮族自治区人民医院科研实验中心

作者简介: 焦伟, 女, 硕士, 主任技师

【摘要】 作为人类可遗传变异中最常见形式, 人类基因组单核苷酸多态性(SNPs)代表了种族间的遗传差异。随着 SNPs 检测技术的快速发展, 专家们在 SNPs 的种族遗传差异的意义及其应用研究中取得了很大进展。本文对 SNPs 的检测方法及其在种族遗传差异分析方面的应用作简要综述。

【关键词】 SNPs; 检测方法; 种族; 遗传差异

【中图分类号】 Q 346.5 **【文献标识码】** A **【文章编号】** 1674-3806(2009)04-0343-03

doi:10.3969/j.issn.1674-3806.2009.04.007

The application of detecting methods of SNPs in analyzing genetic diversity of different race JIAO Wei, LIU Fei. Research Center of Medical Sciences, the People's Hospital of Guangxi Zhuang Autonomous Region, Nan-ning 530021, China

【Abstract】 As the most common form in human genetic variance, human genome single nucleotide polymorphisms (SNPs) stand for genetic diversity of different race. The experts have made great advancement since detecting technology of SNPs develops rapidly. The article concisely reviews for detecting method of SNPs and its application in analyzing genetic diversity of different race.

【Key words】 SNPs; Detecting method; Race; Genetic diversity

人类基因组计划(HGP)揭开了人类遗传信息的秘密, 随着研究的深入, SNPs 的研究应运而生。SNPs 主要是指在基因组水平上由单个核苷酸的变异引起的 DNA 序列多态性, 在人群中的频率大于 1%, 是人类可遗传变异中最常见形式, 占有已知多态性的 90% 以上^[1], 且在不同人群中的分布频率也有差异, 这些差异可以代表某一种族或人群间的遗传差异。因此, SNPs 的研究有助于解释个体间的表型差异、不同群体和个体对疾病的易感性等。随着 HGP 的完成, 人们越来越重视 SNPs 的种族遗传差异的意义及其应用, 相关研究已取得了很大进展。本文对 SNPs 的检测方法及其在种族遗传差异分析方面的应用作简要综述。

1 SNPs 的检测方法

SNPs 可通过电泳、PCR、酶切、直接测序、生物信息学及 DNA 芯片等方法进行检测。尤其随着高密度 DNA 芯片、变性高效液相色谱(DHPLC)、基质辅助激光解吸电离飞行时间质谱(MALDI-TOF MS)等技术的应用, SNPs 技术亦可快速进行自动批量检测, 不仅大大提高了分型速度, 更降低了基因搜索的成本^[2]。对这些方法进行以下归纳并重点介绍 DNA 芯片、DHPLC 和 MALDI-TOF MS 等大规模高通量分析 SNPs 的技术。

1.1 各种技术比较 SNPs 检测方法包括以凝胶电泳、PCR、酶、杂交为基础的分析方法。此外, 针对人类 cDNA 和基因组文库, 科学家发展了基于生物信息学的 SNPs 候选位点搜索方法, 例如 EST 比较法^[3]。PCR-SSCP、RFLP 等必须通过凝胶电泳等进行分析, 只能检测到 SNPs 的一部分。上述方法仅能判断 SNPs 的有无, 而无法确定多态位点的碱基类型。因此, 这些方法发现的 SNPs 要再进行 Sanger 测序确认。对于数据库中已有的 SNPs 在特定人群中序列的验证性分析和频率分析, 使用的技术包括位点特异性探针的应用、Tagman 实验、分子信标(molecular beacons)技术等, 但由于需要荧光标记以及专门的分析仪器, 因此价格较高, 且受仪器设备的限制。现在半自动或全自动地进行大量 SNPs 检测方法包括小型测序、多重反向斑点杂交、DNA 芯片、变性高效液相色谱技术等应用, 极大地提高了在大规模人群中分析 SNPs 差异的工作效率, 并且由于目前大规模 Sanger、焦磷酸等测序技术已经相当成熟, 使采用这些测序进行大规模 SNPs 研究也成为必然趋势。

1.2 DNA 芯片 DNA 芯片基本原理是应用已知的核苷酸序列作为探针与标记的靶核苷酸序列进行杂交, 通过对信号的检测进行定性定量分析。DNA 芯片可在微小的硅片、

玻片表面集成大量的分子识别探针,能够在同一时间内平行分析大量基因,进行高自动化、大信息量的检测分析^[4]。Wang等^[5]最早应用高密度基因芯片对2.3Mb人类基因的SNPs进行筛查,确定了3241个SNPs位点,显示出大规模鉴定人类基因型的可能。近年来,Illumina和Affymetrix公司研发的SNP芯片具有高性能、高质量、低成本的优点,更利于大规模人群研究。虽然目前DNA芯片造价昂贵,并需要贵重设备,但随其在疾病预防、临床诊断、个体治疗、药物研发等各个方面的SNP分析中应用及其标准化、批量化生产效率的提高,已显示出广阔的应用前景。

1.3 变性高效液相层析技术 DHPLC亦称为温度调节的杂合双链分析(TMHA),其原理是利用离子对反向高效液相色谱原理,通过一个DNA分离柱,进行核苷酸片段的分离和分析。SNPs的有无最终表现为色谱峰的峰形或数目差异,依据此现象可很容易从色谱图中判断出突变的碱基。该技术具有半自动化、快速、检出率高等特点,是一种高通量、自动化筛查SNPs的有效方法,但它非常精密,对所用试剂和环境要求较高,容易产生误差,不能检测出纯合突变^[6]。

1.4 基质辅助激光解吸电离飞行时间质谱 MALDI-TOF-MS的工作原理是适当大小的有机分子晶体(介质)用接近其吸收光谱的激光瞬时激发,会发生能量转移及解吸过程。MALDI-TOF-MS应用于SNPs分型基于“引物延伸法”,该方法是指延伸引物在待检测SNP位点上延伸一个或若干个碱基,然后根据延伸产物所带荧光的不同或其分子质量的不同而确定其基因型。由于其固有的质量高准确性、灵敏性和可重复性,基于MALDI-TOF-MS的SNPs分析有望替代传统的实验手段,是常规电泳和色谱技术以外的另一种选择^[7]。但是,必须去除样品中的干扰性离子(钠、钾)方能获得富含信息的MALDI-TOF波谱,同时,高通量的自动化样品制备和纯化对于基于MALDI-TOF-MS的基因分型法也是必需的。它适合大规模筛查,但设备昂贵不宜普及^[8]。

2 种族遗传差异的SNPs分析

SNPs是在人类漫长发展中基因组与内外环境交互作用的累积结果,不同的SNPs位点在不同人群中的分布频率及杂合程度不同,在进行疾病分析时提供的信息量可能存在较大差异^[9]。同样,由于研究条件的差异,研究对象的人种、生活地域的不同、生活习惯的不同等因素使研究结果存在差异。在不同的亚群中等位基因的频率可能有差别,例如在白种人中表现多态性的标记位点在中国人中可能没有多态性。我国幅员辽阔,民族众多,一般采集样本的群体很难具有广泛的代表性,这就需要收集多地区、多民族的详细资料,构建信息丰富的SNPs图谱,建立全球性合作,共享DNA样品和细胞标本的公共资源。以下分别从国内外学者对同种族遗传差异SNPs分析的研究进展进行综述。

2.1 国际研究进展

2.1.1 人们借助SNPs对感兴趣的基因或区域进行研究,发展了单体型(haplotype)系统。单体型是指相邻SNPs的等位

位点倾向于以一个整体遗传给后代;位于染色体上某一区域的一组相关联的、连锁不平衡的SNPs等位位点的集合。SNPs直接以序列的差异作为标记,比其他遗传标记更能真实地反映人种、人群及个体之间的遗传差异,但SNP单个基因座的多态性很差,只有二态,为此采用多个SNP基因座进行“单体型”分析尤为重要。

2.1.2 基于以上研究,科学家们在2002年10月开始进行国际人类基因组单体型图计划(HapMap, <http://www.hapmap.org/>),目标是要构建整个人类基因组DNA序列中多态位点的常见模式,至少对百万个SNP进行全基因组规模的基因分型,观察人类“共同的”遗传变异。现在收集了270个祖先分别来自非洲、亚洲和欧洲的四个人群的样本,测序得到了超过1百万个SNP,平均约3000个碱基一个SNP。它包括的4个群体样本具体为:北美血统(CEU)(美国犹他州来自欧洲北部和西部后裔);尼日利亚血统(YK1)(来自非洲尼日利亚的伊巴丹地区优鲁巴人);日本无亲缘关系的东京人(JPJ);中国无亲缘关系的北京人(HCB)。

2.1.3 由于样本量巨大且检测位点繁多,科学家们通过不断改进的测序技术,更加有效、高速、节约成本地实施该计划。近期Nature介绍的两项重大协作研究工作:对于来自西非Yoruba族群的一个人的基因组定性研究和一个汉族人个人基因组的报告,展示了最新大规模并行合成DNA测序技术的威力,这两项研究为了解基因组随种族变化而变化的性质提供了线索^[10,11]。

2.2 国内研究进展

2.2.1 至今为止在我国分析各族人群SNPs遗传差异尚处于初期阶段,大都采用以凝胶电泳为基础的RFLP、SSP-PCR、巢式PCR等,这些方法各有优势,能够简便、低廉、可靠地用于比较汉族及各少数民族之间特定基因SNPs位点的基因型和/或单型体的异同。目前分别在贵州、云南、广西、东北等地共16个民族进行研究^[12~15],与同地区各民族甚至是国际上的数据进行比较,取得了一定成果。他们发现在各民族之间SNPs位点存在一定差异:贵州省布依族、仡佬族、仫佬族、毛南族、壮族5个民族之间有密切联系,且与国内其他地域有较大的遗传差异^[12];云南的白族、佤族、拉祜族普通人群之间,甘露聚糖结合凝集素MBL基因各单体型和各基因型的分布也存在差异^[13];与冠心病发生密切的ABCA1基因1883M在广西壮族人群与汉族人群之间没有差异^[14],但其等位基因频率与德国人群^[16]的分布特征均存在显著性的差异;东北少数民族中ApoE基因2440SNP位点在7个少数民族间存在差异,并且与国外文献报道及国际SNP数据库比较有很大的差异^[15]。

2.2.2 学者们因地制宜地建立以上的SNPs检测方法,在一定范围有效地对少量人群SNPs位点差异进行研究,但不适用于种族遗传差异的大规模研究。近年来,身为国际稀有血型交换组成员和从事输血医学基础研究的朱自严教授对血型抗原基因的SNPs差异进行比较,寻找不同民族的稀有血型频率和分布情况。前期工作采用序列特异性引物PCR

(PCR-SSP)及 RFLP 结合 DNA 测序,更大范围对上海和广东的汉族以及新疆的维吾尔族、回族等人群的 Del、Kell、Jk 等稀有血型相关基因进行分型及 SNPs 分析^[17-19]。目前正逐步在广西、新疆等各大少数民族聚集地进行大规模研究,这为各民族稀有血型相关基因遗传差异分析及筛选打下了坚实的基础。

2.2.3 我们实验室结合朱自严教授的课题基础,在国内首先采用基于单碱基延伸的 SNPs 分型芯片,对广西壮族地区壮族个体和汉族个体的染色体基因组上 10 个 SNP 位点进行了分型,其中 9 个位于 6 个血小板同种抗原基因中,1 个位于基因间区。结果显示广西原住汉族人在等位基因频率上未检测到与当地壮族有显著性的差异位点,但在基因型频率上,其中的两个位点有显著差异。结合 Hapmap 计划第二期公布的四个人群的 SNPs 分型数据,分析这六个人群的遗传结构,结果显示广西壮族人与广西汉族、HCB 及 JPJ 的遗传结构相近,但与祖先来自 CEU 以及 YK1 有显著差异的遗传成分存在。得到这样结果可能是样本采集量太小导致不能完全反应整个群体的客观事实,但更可能是壮汉两族由于历史上的多次基因交流导致其遗传信息在很大程度上是相近的,因此今后有必要扩大规模来考察壮汉族遗传上的多态性(待发表)。

3 展望

随着 SNP 检测方法的不断开发和改进,人们对不同种族 SNPs 差异的认识逐渐深入。这为 SNPs 应用于疾病防治、临床诊断、输血等医疗卫生方面的发展提供了广阔的空间和深厚的基础,尤其是国家之间及其内部各民族之间存在的 SNPs 差异对于以上研究方向具有重要意义。开发大规模筛查人类 SNPs 位点的技术不断进步,其中 SNPs 分型芯片的应用引人注目,然而目前它仍存在许多问题:首先,该方法成本较高,大多是一次性使用,检测杂交的仪器昂贵,所以要转化为实践运用还很困难;其次,SNPs 的杂交检验对技术要求较高,需要注意杂交的灵敏度和准确度,这就限制了现场办公的便捷性。而且,目前把该技术分析 SNPs 位点用于民族 SNPs 分析的实例还较少。相信随着科技的进步,它在种族遗传实践中的普遍应用也终将实现。

参考文献

- Collins FS, Brooks LD, Chakravarti A. A DNA polymorphism discovery resource for research on human genetic variation[J]. Genome Res, 1998, 8(12): 1229-1231.
- 汪维鹏,倪坤仪,周国华. 单核苷酸多态性检测方法的研究进展[J]. 遗传, 2006, 28(1): 117-126.
- Huntley D, Baldo A, Johri S, et al. SEAN: SNP prediction and display program utilizing EST sequence clusters[J]. Bioinformatics, 2006, 22(4): 495-496.
- Marshall A, Hodgson J. DNA chips: an array of possibilities[J]. Nat

Biotechnol, 1998, 16(1): 727-731.

- Wang DG, Fan JB, Siao CJ, et al. Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome[J]. Science, 1998, 280(5366): 1077-1082.
- Betz B, Flori AR, Seifert HH, et al. Denaturing high-performance liquid chromatography (DHPLC) as a reliable high-throughput pre-screening method for aberrant promoter methylation in cancer[J]. Hum Mutat, 2004, 23(6): 612-620.
- Griffin TJ, Smith LM. Single-nucleotide polymorphism analysis by MALDI-TOF mass spectrometry[J]. Trends Biotechnol, 2000, 18(2): 77-84.
- Bocker S. SNP and mutation discovery using base-specific cleavage and MALDI-TOF mass spectrometry[J]. Bioinformatics, 2003, 19 (Suppl 1): i44-i53.
- Goddard KA, Hopkins PJ, Hall JM, et al. Linkage disequilibrium and allele-frequency distributions for 114 single-nucleotide polymorphisms in five population[J]. Am J Hum Genet, 2000, 66(1): 216-234.
- Bentley DR, Balasubramanian S, Swerdlow HP, et al. Accurate whole human genome sequencing using reversible terminator chemistry [J]. Nature, 2008, 456(7218): 49-51.
- Wang J, Wang W, Li R, et al. The diploid genome sequence of an Asian individual[J]. Nature, 2008, 456(7218): 60-65.
- 刘炬, 单可人, 齐晓岚, 等. 贵州布依族、仡佬族、仫佬族、毛南族、壮族 Y-SNP 的初步研究[J]. 遗传, 2006, 28(11): 1350-1354.
- 吕成伟, 葛争鸣, 李江川, 等. 中国白族、佤族和拉祜族人群 MBL 基因启动子区 SNP 的研究[J]. 细胞与分子免疫学杂志, 2005, 21(1): 110-113.
- 马义丽, 黄健, 戴玲, 等. ABCA1 基因 1883M 在广西汉族和壮族人群中的多态性分布[J]. 长江大学学报, 2007, 4(3): 88-90.
- 刘岸, 高凌寒, 陈峰, 等. 东北地区 7 个少数民族群体中 ApoE 基因单核苷酸多态性的研究[J]. 哈尔滨医科大学学报, 2005, 39(3): 205-207.
- Evans D, Beil F U. The association of the R219K polymorphism in the ATP-binding cassette transporter 1 (ABCA1) gene with coronary heart disease and hyperlipidaemia[J]. J Mol Med, 2003, 81(4): 264-270.
- 王玲玲, 杨颖, 朱自严. 中国人群 Kell 血型 K 阳性的分子特征及分型试剂盒的研制和应用[J]. 临床输血与检验, 2008, 10(2): 110-113.
- 邓诗杭, 曾玫玫, 严康峰, 等. 番禺地区无偿献血人群中 Jk(a-b-) 表型的筛选与研究[J]. 中国输血杂志, 2007, 20(1): 9-12.
- 李勤, 叶璐夷, 郭忠慧, 等. 中国人群中 Del 表型分子背景研究[J]. 中华医学遗传学杂志, 2006, 23(5): 486-491.

[收稿日期 2009-03-08][本文编辑 宋卓孙 刘京虹]