

- 8 Tjuvajev JG et al. *Cancer Res*, 1996; 56: 4087 ~ 4095
- 9 Tjuvajev JG et al. *J Nucl Med*, 1997; 38(5): 239p
- 10 Haberkorn U et al. *J Nucl Med*, 1996; 37: 87 ~ 94
- 11 Tjuvajev JG et al. *J Neuro Oncol*, 1996; 30: p35

(收稿日期: 1999-04-18)

## 基因表达的序列分析

中国医学科学院放射医学研究所(天津, 300192) 杨春综述 卢圣栋\* 审核  
中国协和医科大学

**摘要:** 基因表达的序列分析(SAGE)方法是一种新的研究特定细胞内基因表达水平的方法。它利用  $II_s$  内切酶识别位点与切割位点相距特定碱基数目的特点, 取转录子 3'端特定位置上一定数目的碱基序列来代表此转录子, 并以此短片段出现的次数反映其转录水平。SAGE方法可以进行全基因组基因表达水平的研究, 为基因表达调控等研究提供了新的思路。随着其方法的进一步改进, 必将在后基因组研究中发挥作用。

**关键词:** 基因表达 SAGE DNA芯片 后基因组研究

基因表达水平的分析对于研究了解这一基因的功能起着至关重要的作用。研究人员常常通过测定组织或细胞样本中的 mRNA 水平来衡量基因的表达情况。目前已建立了一些测定 mRNA 水平的方法: Northern blot 和 mRNA spot blot 是较常用的方法, 可以同时多种细胞或组织样本的测定, 但一次只能使用有限数目的探针, 研究有限的基因表达; RT-PCR(逆转录聚合酶链反应)方法也可以用来定量 mRNA, 但其可靠性有争议, 因为它所依赖的扩增反应对不同的模板和不同的引物都会产生变化很大的结果<sup>[1]</sup>; 表达的序列标记(expressed sequence tag, EST)也可以用来定量 mRNA, 但大规模的 EST 测序工作量大, 费用昂贵<sup>[2,3]</sup>。近年发展了一些新的方法, 可用于大规模研究基因表达<sup>[4-6]</sup>, 基因表达的序列分析(serial analysis of gene expression, SAGE)即是其中的一种<sup>[6]</sup>。

### 1 原理

SAGE方法要求从某一转录本(mRNA)

的确定位置分离一小段核苷酸作为标记序列(TAG), 通常为 9~11bp, 给这些标记序列加上一定的标志(Marker), 将其随机连接、扩增、克隆, 选择一定数量的克隆产物进行测序分析。以标志物将标记序列彼此区别, 其标记序列代表相应的转录本, 其重复出现的次数代表转录本拷贝数, 即 mRNA 水平。

### 2 操作

首先, 从细胞或组织中提取全部的 mRNA, 用生物素标记的 Olig(dT)引物将其反转录成 cDNA。选择一种限制性内切酶(称锚酶, 它具有 4bp 识别位点, 在绝大多数转录本中至少有一个酶切位点), 用亲和素磁珠分离 cDNA, 获得距 poly A 位点最近的带唯一锚酶切割位点的断片。将此混合产物一分为二, 补平酶切末端, 分别连上不同的连接子(linker)。连接子中含有  $II_s$  类内切酶识别位点,  $II_s$  类内切酶(称标记酶, tagging enzyme)切割 cDNA 5'方向距识别位点约 20bp 处的核苷键。由标记酶切割产生带不同连接子的

\* 中国医学科学院, 中国协和医科大学基础医学研究所(北京, 100730)

平头末端产物,其中含有标记酶识别位点、锚酶识别切割位点和 cDNA 5'方向由标记酶切割产生的内源性短核苷序列,即标记(tag)序列。将不同连接子引导的标记序列尾尾相连,产生双标记(di-tag),利用与连接子相匹配的引物进行 PCR 扩增。这一反应除了用来扩增双标记序列外,还可对正确连接物进行筛选,去除相同连接子的双标记序列、单标记序列和多标记序列。扩增产物包括两个尾尾相连的标记序列,锚酶识别切割位点(Marker)位于其两侧。测序时,这种安排将产生由 4bp 锚酶识别切割位点相间隔的串联双标记序列。在同一反应体系中,同一标记序列连接的几率很小,即使高丰度的转录子也不例外。PCR 产物再次经锚酶切割,去除连接子及其所含标记酶识别切割位点,剩余断片随机连接,选择一定长度的连接物,克隆、测序,进行计算机辅助分析。

SAGE 方法通过锚酶在转录本上定位并产生标记,通过 II<sub>s</sub> 类内切酶产生一定长度的短核苷序列,这位置一定的 9~11bp 的短核苷序列就可以唯一代表一个转录本。理论上,9bp 核苷酸随机排列可产生 262 144 种不同的组合,即 262 144 种不同的标记序列,这远大于目前所估计的(约 80 000 个)人类基因组数目。Velculescu 等人<sup>[6]</sup>通过对 GenBank 的研究,证实了距 polyA 位点最近的 NlaIII 酶识别切割位点上游 9bp 处的核苷酸序列在 95% 以上已知基因序列中是唯一的。

SAGE 方法减弱了 PCR 反应带来的误差,并可根椐实验需要和实际情况增加测序克隆数目,以进一步提高方法的灵敏度,这是 Northern blot 或 mRNA spot blot 方法所不能达到的。Velculescu 等人<sup>[6]</sup>通过几组对比实验研究了 SAGE 方法的准确性:首先用胰腺组织中丰富表达的四种酶基因为探针与 cDNA 文库杂交,以产生的斑点数代表 mRNA 水平,其结果与由 SAGE 方法产生的相

似;其次,将标记序列(9bp)与标志连接物(4bp)做成放射性标记探针与 cDNA 文库杂交,以 3' 测序保证杂交的准确性,所得结果也十分相似。

### 3 应用

SAGE 方法自 1995 年问世以来,引起了研究人员的广泛关注。利用 SAGE 方法来研究基因表达的工作已深入开展起来<sup>[7-9]</sup>。

Velculescu 等人<sup>[7]</sup>利用 SAGE 方法研究了酵母转录组(transcriptome 代表特定细胞类群中每个表达的基因及其表达水平),收集不同生长时相(对数期, S 期, G<sub>2</sub>/M 期)的细胞总 mRNA,建立 SAGE 文库,共获得了 60 633 个 SAGE 标记序列,其中 56 291 个(93%)与已知基因吻合,这 56 291 个标记序列代表 4 665 个不同基因,占已知酵母基因组数目的 7%;通过 SAGE 方法获得的基因表达水平分布与通过 RNA-DNA 缔合动力学(RNA-DNA reassociation kinetics)方法获得的大体相似,只是前者的低丰度分布的基因远较后者高,这是因为 SAGE 方法可以检测到更低丰度表达的基因;检测到的所有高丰度表达基因(> 60 拷贝/细胞)都编码与能量代谢和蛋白合成有关的酶类,并在不同生长时相中有相同的表达;在这些高表达的基因中(30 个),有 3 个是以前并不知道的。对 SAGE 资料的分析还表明,存在 2648 个基因在可测水平表达,但并不清楚其开放阅读框架(ORF);另外还有 160 个基因在可测水平表达,带有 nonannotated ORF(NORF),其中有一个 NORF 仅在 S 期表达。研究结果表明,SAGE 方法不但可以研究基因表达,而且可以提供大量未知基因的信息,对于发现新基因也有作用。

Zhang 等人<sup>[8]</sup>利用 SAGE 方法研究了人类正常细胞与癌细胞基因表达情况。他们从正常人、结肠癌患者和胰腺癌患者的结肠上皮细胞中共获得 303 706 个标记序列,这

些序列代表了 49 000个不同的基因,其表达水平从 1到 5 300个拷贝/细胞不等。占 mRNA总重量 75%的转录本含有 5个以上拷贝/细胞,而 86%的转录本其表达水平低于 5个拷贝/细胞。大量的基因处于低水平表达,用其它方法,如差异显示(differential display)就很难区分。许多 SAGE标记序列代表未曾描述过的转录本,只有 54%的转录本与 GenBank相符,这些相符的转录本中的 20%与已知的 mRNA配对外,另外的 80%只与未被了解的 EST序列配对,它们占高丰度表达基因(> 500拷贝/细胞)的 98%,占低丰度表达基因(< 5拷贝/细胞)的 51%。这说明,由其它方法发现的 mRNA序列或 EST序列中绝大多数是高表达基因,而对更多的低水平表达的基因我们还知之甚少, SAGE方法可以帮助我们去发现这些低水平表达的未知基因。比较正常结肠上皮和原发性结肠癌细胞基因表达情况发现,许多转录本的水平是相似的,但有 289个转录本表现异常,其中 181个转录本在癌细胞中表达降低(平均 10倍), 108个转录本在癌细胞中表达升高(平均 13倍)。比较结肠癌细胞系和结肠组织中基因表达水平表明,在 181个低水平表达的基因中, 103个基因在细胞系中也呈低水平表达,而在 108个高水平表达的基因中, 48个基因在细胞系中也呈高水平表达。这表明,尽管许多在活体组织中存在的正常细胞与癌细胞基因表达的差异在离体条件下也存在,微环境对基因表达还是有影响的。分析胰腺来源的 SAGE资料表明,另有 136个基因可能与肿瘤组织的发生相关而与细胞来源无关。利用 SAGE方法,还研究了同种组织不同个体间异质性的问题:同一器官来源的肿瘤组织间存在生物学特性的异质性。这似乎也可作为个体差异提供一些佐证。

随着 SAGE受到更多研究人员的关注,在方法上也有了改进。Powell在双标记序列 PCR扩增反应时引入生物素标记的引物,使

过量的引物及不正确的连接物可以更加有效地分离,提高了双标记序列的产量及克隆产物的信息含量<sup>[11]</sup>。针对基因组中大多数基因是低水平表达,最近, Wang和 Powell<sup>[12]</sup>在 SAGE方法的基础上引入差减克隆方法,形成 IPGI(integrated procedure for gene identification)。此法除去了许多高丰度表达基因,使发现新基因的能力大大增强。同时,此方法强调了与已有的 EST信息文库的紧密联系。

SAGE方法与同它几乎同时出现。现在已广为重视的 DNA芯片(DNA chip)法在研究基因表达差异方面有异曲同工的效果。DNA芯片方法主要基于杂交和微点阵原理,它以荧光强度来代表 mRNA水平,常常以相对值的方式出现。SAGE和 DNA芯片法都可以进行大量基因表达水平的研究,相比较, DNA芯片法速度更快,自动化程度也更高,但 DNA芯片法通常研究的是已知基因或序列,同时目前 DNA芯片法还存在一些方法上的制约,如杂交背景高、特异性低、寡核苷酸配对杂交等<sup>[13]</sup>, DNA芯片法对仪器设备的要求也很高,需要激光共聚焦显微镜、DNA合成仪、高精度光栅等;而任何一家实验室只要会 PCR和手工测序,就可以利用 SAGE方法进行研究,使用一台自动测序仪,一个工作人员在一个月内就可完成 20 000个 SAGE标记序列的分析<sup>[14]</sup>。

对单个基因的分析 and 了解最终要在综合的、整体条件下加以深化和印证。SAGE方法摆脱了传统上对单个基因的研究,为从整体上全面地研究转录组的生物学动态变化,提供了一种在更广范围内研究物质基础和生物学特性的方法,便于人们更加深入地了解基因的功能和生命的本质。

现在,酵母基因组已完全测序<sup>[15]</sup>,人类基因组 DNA测序工作将于 2003年完成<sup>[16]</sup>,对基因结构和功能及其生物学特性的研究将成为下一世纪生物学研究的主要内容。

SAGE方法可以定量研究不同细胞(包括组织、器官、种属)、不同发育阶段、不同进化程度及不同生理病理状态下基因表达的差异,是全面综合了解基因表达与生命活动关系的强有力的工具,在后基因组研究中必将发挥更大的作用

参 考 文 献

- 1 Veres G et al. Science, 1987; 237: 415~ 419
- 2 Lee NH et al. Proc Natl Acad Sci USA, 1995; 92 8303~ 8307
- 3 Fung YW and Liew CC. J Mol Cell Cardiol, 1996; 28 1241~ 1249
- 4 Nguyen C et al. Genomics, 1995; 29 207~ 216
- 5 Schena M et al. Science, 1995; 270 467~ 470
- 6 Velculescu VE et al. Science, 1995; 270 484~

- 487
- 7 Velculescu VE et al. Cell, 1997; 88 243~ 251
- 8 Zhang L et al. Science, 1997; 276 1268~ 1272
- 9 Madden SL et al. Oncogene, 1997; 15 1079~ 1085
- 10 Schena M et al. Proc Natl Acad Sci USA, 1996; 93 10614~ 10619
- 11 Powell J. Nucl Acids Res, 1998; 26 3445~ 3446
- 12 Wang SM and Powell J. Proc Natl Acad Sci USA, 1998; 95 11909~ 1191
- 13 Coffen A. Nature, 1997; 385 202~ 203
- 14 Adams M D. Bioassays, 1996; 18 261~ 282
- 15 Coffen A et al. Science, 1996; 274 246~ 247
- 16 Collins FS. Proc Natl Acad Sci USA, 1995; 92 10821~ 10823

(收稿日期: 1999-04-23)

## 乏氧组织显像剂研究最新进展

中国医学科学院放射医学研究所(天津, 300192) 李 莉综述 李美佳 郑妙 \* 审校  
中国协和医科大学

摘 要: 乏氧组织显像剂能选择性地滞留在乏氧组织或细胞中,并通过核医学显像探测组织缺氧及其程度。乏氧显像能直接提供任何器官中组织存活但有功能障碍的依据,这些信息在临床诊断决策中起到重要作用,尤其是在心、脑血管疾病和实体瘤的诊断方面。本文简要介绍乏氧组织显像剂与乏氧显像的特点及其研究的最新进展。

关键词: 乏氧组织显像剂 乏氧显像

应用乏氧组织显像剂探测缺氧、缺血组织是目前国际上放射性新药研究的一个热点。组织乏氧是临床上许多疾病的一个显著特征,了解组织氧水平对这些疾病的诊断十分重要。

采用非侵入性方法确定乏氧组织一直受到广泛重视,到目前为止,比较成熟的方法有磁共振光谱法(MRS)、核医学技术及电子自旋共振(ESR)光谱法等<sup>[1]</sup>。其中,核医学技术是利用乏氧组织显像剂进行SPECT或PET,简便易行,非常适于临床推广应用。

### 1 乏氧组织显像剂与乏氧显像

乏氧组织显像剂是一类阳性显像剂,它们能选择性地滞留在乏氧组织或细胞中,并通过核医学显像探测组织缺氧及其程度。理想的乏氧组织显像剂应具有如下特点<sup>[2]</sup>:①能迅速集中于感兴趣的乏氧组织,病灶本底的比值大于3:1,并能在短时间内提供足够的光子流以得到高质量的图像(要求不超过病人能承受的辐射剂量);②能在血液中迅速清除,若要在注射后1小时成功显像,它在血

\* 天津医科大学第二附属医院核医学科(天津, 300211)