

# 基因表达的系列分析方法研究进展

熊勇华\* 许 杨

(中德联合研究院 南昌 30047)

**摘 要** 基因表达的系列分析(SAGE)是探讨组织或器官在不同条件下基因表达丰度以及差异表达的一种有效方法。本文介绍了 SAGE 方法的详细机理并且对 SAGE 方法学的改进进行了综述。

**关键词** SAGE, 修饰 SAGE, 微量 SAGE, rSAGE

中图分类号 Q78 文献标识码 C 文章编号 1000-3061(2002)03-0377-04

基因在不同组织、器官以及不同条件下的差异表达是生物体发育、分化、衰老和抗逆等生命现象的分子基础。探讨基因组在特定时期转录水平上的表达丰度以及差异表达情况,可为研究基因组的功能提供重要的信息,而研究基因组功能是后基因组计划的主要内容之一<sup>[1]</sup>。目前用于基因表达水平以及差异表达基因筛选的常用方法有 Northern 杂交、消减杂交(Subtractive hybridization)、差示筛选(Differential screening)、mRNA 差别显示(Differential display)、表达序列标签(Express sequence tag, EST)以及逆转录-PCR(RT-PCR)等,这些方法在克隆差异表达基因方面曾做出过巨大的贡献,但由于它们在筛选基因方面的随机性使之无法描述基因转录组(Transcriptome)的全貌,同时这些方法只适合于高丰度或表达差异大的基因筛选,而不能胜任对生物体基因转录组进行全面、系统的分析<sup>[2,3]</sup>。

基因表达的系列分析(Serial analysis of gene expression)是以转录子(cDNA)上特定区域 9~11bp 的寡核苷酸序列作为标签(tag)来特异性代表该转录子,然后通过连接酶将多个标签(20~60 个)随机串联并克隆到载体中,建立 SAGE 文库<sup>[4]</sup>。通过对标签的序列分析,可获得基因转录的分布以及表达丰度情况(尤其是可检测到低丰度表达的基因),从而可充分了解基因转录组的全貌。由于该方法不需昂贵仪器,只需拥有 PCR 仪以及测序条件的实验室即可进行操作,因此自 1995 年 Velculescu 等在 Science 首次报道了该研究方法以来,与此有关的研究论文达 349 篇之多(PubMud 数据库检索结果),其研究涵盖了人类疾病,尤其是各类癌症疾病、生物代谢途径、植物和模式生物以及基因转录组等诸多领域<sup>[5,6]</sup>。随着 SAGE 方法广泛深入的应用,许多学者对该方法提出了这样或那样的改进方案,从而使实验结果更加完善,并且拓宽了该方法的应用范围。本文仅就 SAGE 方法学的原理及其研究进展做一简要论述。

## 1 常规 SAGE 方法的操作步骤及原理

该技术原理的基本流程见图 1<sup>[7]</sup>。主要包括以下实验内容。

### 1.1 生物体特定阶段组织或细胞中全部 mRNA 的制备

### 1.2 cDNA 双链的合成

用生物素标记的 Oligo(dT)引物将 mRNA 反转录成双链 cDNA。

### 1.3 锚定酶酶切

锚定酶是一种限制性内切酶,具有 4bp 的识别位点,常用的锚定酶如 *Nla* III,其酶切识别位点为 CATG 4 个碱基序列。酶切后产物通过生物素磁珠分离含 polyA 尾巴的 cDNA 片段,将此产物平均分为 2 份。

### 1.4 cDNA 片段与接头(Linker)的连接

接头包括以下 4 条单链,Linker 1A 5' TTTCGATTTCGCTGCTGCAGTACAACCTAGGCTTAATAGGGACATG 3'; Linker 1B 5' TCCCTATTAAGCCTAGTTGTACTGCACCAGCAAATC 3'; Linker 2A 5' TTTCGCTCGAATTCAGCTTCTAACGATGTACGGGGACATG 3'; Linker 2B 5' TCCCCGTACATCGTTAGAAGCTTGAATTCGAGCAG 3'。将 Linker 2A 以及 Linker 2B 末段去磷酸化,并与相应的 Linker 1A 以及 Linker 1B 进行退火形成接头 Linker 1 和 Linker 2(注:接头末端含有锚定酶以及标签酶酶切位点)。退火产物分别与各自的 cDNA 片段组分连接。

### 1.5 标签酶酶切释放标签序列

标签酶是一种 II S 限制酶,其切割位点距识别位点 10~14bp。常用的标签酶如 *Bsm*FI,采用 Klenow 大片段将标签序列末端补平。

### 1.6 连接形成双标签(Ditags)

将含有 Linker 1 和 Linker 2 的标签在连接酶作用下连接形成含 Linker1、2 的双标签。

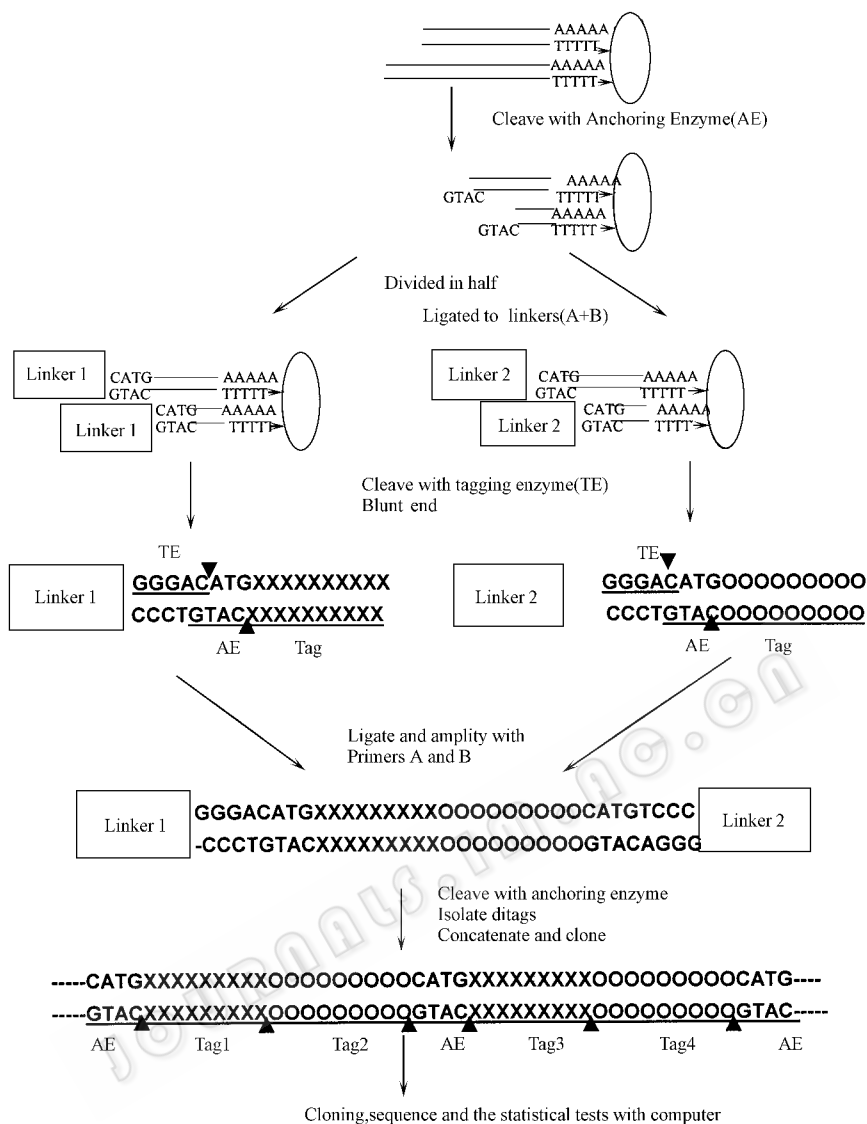


图 1 常规 SAGE 方法图解

Fig.1 Schematic of common SAGE

### 1.7 PCR 扩增(Ditags)

PCR 扩增引物依据 Linker 1 和 Linker 2 序列设计,分别为 Primer 1 :5' GGATTGCTGGTGCAGTACA 3' ; Primer 2 :5' CT-GCTCGAATTCAAGCTTCT 3'。只有含 Linker 1 和 Linker 2 的双标签才能得到有效的扩增。采用 12% 的聚丙烯酰胺凝胶电泳纯化 PCR 产物,回收 102bp 的扩增片段。

### 1.8 分离纯化双标签

采用锚定酶酶切回收的扩增产物,酶切产物采用 12% 的聚丙烯酰胺凝胶电泳回收 26bp 的片段,即得双标签。

### 1.9 双标签随机连接形成串联子(Concatemers)

纯化的双标签采用连接酶使之成锁链状连接形成不同大小的串联子。8% 的聚丙烯酰胺凝胶电泳回收一定片段长度的串联子。

### 1.10 克隆串联子 测序

将串联子克隆至含 *Sph* I 酶切位点的 pZero 载体中,测定串联子的序列。

### 1.11 计算机辅助分析

测序结果采用 SAGE 软件包进行分析,获得标签序列及其丰度信息,每个标签可通过与 Genbank 数据库或者 EST 数据库数据进行对比,从而可确认其所代表的基因。

## 2 常规 SAGE 方法的技术改进

### 2.1 提高 *Nla* III 的酶切效率

获得高产率的双标签(24 ~ 26bp),是关系到 SAGE 文库完整性的关键。Angelastro 等在实验中发现 *Nla* III 对电泳法回收的 102bp 的双标签片段酶切效率较低,往往导致整个实验的失败。分析原因可能是由于聚丙烯酰胺凝胶中溶出的污染物严重地抑制了 *Nla* III 酶的活性。因此认为对电泳法回收的双标签片段进行纯化是必要的。他们采用了 Qiaquick 试剂盒(一种 DNA 快速纯化试剂盒)以及 Clontech spe10 凝胶过滤离心柱纯化双标签片段,有效地提高了 *Nla* III 的酶切效率。酶切效率为 80% ~ 90% 之间,是未经纯化组的 2 倍。

倍<sup>[8]</sup>。

## 2.2 提高串联子的连接效率

在 Velculescu 等描述的常规 SAGE 方法中,PCR 扩增产物经 *Nla* III 的酶切后,采用了 12% 的聚丙烯酰胺凝胶电泳回收双标签(26bp),而 Powell 等认为该步骤无法避免在回收片段中存在接头污染的可能性,如果存在该种污染,在下一步连接反应中接头的粘性末端将会连接在串联子上,从而使连接反应终止。于是他们在实验中提出了一种新的方案,即在 PCR 扩增的引物中掺入含生物素标记的核苷酸,制备出生物素标记的引物。扩增产物经 *Nla* III 的酶切后,采用磁珠法除去反应物中接头片段,从而获得不含接头污染的双标签(26bp)<sup>[9]</sup>。并将其结果与常规 SAGE 方法进行对比,发现其方法双标签产率在 800~2000ng 之间,克隆子克隆片的平均长度为 394~514bp,每个克隆平均含有的标签数为 30~39 个,而常规 SAGE 方法双标签产率为 400ng,克隆子克隆片的平均长度为 284bp,每个克隆平均只含 21 个标签。

在常规 SAGE 方法中,连接产物经 8% 聚丙烯酰胺凝胶分离,其片段长度平均分布在 300 至几千 bp 之间,将 700~1000bp、1000~1600bp、1600~2500bp 3 组不同长度的片段分别克隆至 pZero 载体上,结果测得插入载体的片段长度分别为 137±47bp、208±100bp、542±105bp。造成连接产物在克隆过程中降解的原因,目前尚无定论。有些学者在实验技术上提出了一些有效的改进措施,使克隆效率得以大大提高。Kenzelmann 等在常规 SAGE 方法中增加了一个加热步骤,从而有效地提高了后期的克隆效率,其改进方案如下:双标签

(26bp)经连接反应过夜(16℃),后升温至 65℃保持 15min,立即置于冰上保持 10min,其余步骤如常规 SAGE 方法。增加该步骤后,同上 3 组不同长度的接头插入 pZero 载体后获得的克隆片段长度分别为 669±119bp、993±342bp、1455±225bp<sup>[10]</sup>。

## 3 microSAGE 和 miniSAGE

尽管 SAGE 方法在许多领域得到了广泛的应用,但由于该方法需要较多的出发材料来提供充足的 mRNA(2.5~5μg),因而使之用于构建某些微量样本来源的 SAGE 文库受到极大的限制,因此许多科学家在常规 SAGE 方案上提出了不同的改进方案,从而建立了多种微量 SAGE 的实验方法。

### 3.1 microSAGE

Datson 等以 325μm 的小鼠脑组织切片(最多含 10<sup>5</sup> 个细胞,1~5ng mRNA)为出发材料建立了 microSAGE 文库,该文库共含 1792 个转录子标签,代表了 1242 个不同转录子信息<sup>[11]</sup>。我们在 SAGE 的专业网站(<http://WWW.sagenet.org/>)也可以看到 Brad St. Croix 等刊登的该方法的详细步骤。microSAGE 与常规 SAGE 方法不同之处见表 1。该方法在实验操作过程中,采用了偶联生物素的 PCR 薄壁管,使实验中许多步骤能够在 PCR 管中完成,因而大大减少了实验操作过程中样品的流失,同时在后期采用两次 PCR 反应,从而使微量的双标签(102bp)模板得到了比一次 PCR 反应更大量的扩增。以上即是 microSAGE 与常规 SAGE 方法之间的本质区别。

表 1 普通 SAGE 和微量 SAGE 方法的主要区别

Table 1 Main difference between common SAGE and microSAGE<sup>[11]</sup>

	Common SAGE	microSAGE
A mount of input material	2.5~5μg mRNA	Total RNA from a single punch of a 300 μm tissue slice (10 <sup>5</sup> cells; ±1~5ng mRNA)
Capture of cDNA by:	Streptavidin-coated magnetic beads	Streptavidin-coated PCR tube
Multiple-tube reaction vs. single-tube reaction	· Subsequent reactions performed in multiple tubes · Multiple phenol-chloroform extraction and ethanol precipitation steps between cDNA synthesis and tag release	· Single-tube reaction from RNA isolation to tag release · Easy change of buffers between subsequent steps · No phenol-chloroform extraction and ethanol precipitation step until tag release · Fewer manipulations reduces loss of material
PCR	25~28 cycles of PCR	28 cycles followed by a limited number of cycles of re-PCR on excised ditag(8~15cycles)

### 3.2 miniSAGE

该方法与 microSAGE 有异曲同工之效,采用该法,建立 SAGE 文库仅需 1μg 总 RNA,相当于常规 SAGE 方法的 1/250~1/500。该法与 microSAGE 不同之处在于未采用两次 PCR 进行双标签(102bp)的扩增,因而有效地减少了两次 PCR 反应所造成的转录组标签发生丰度偏移的潜在可能性。两种方法的其余实验步骤基本相似。Ye 等采用此法以微量人表

皮活检标本的成纤维细胞为原料,建立了 SAGE 文库,共获得 3838 个转录组标签,代表了 2308 个不同转录子信息<sup>[12]</sup>。

## 4 RAST-PCR 和 rSAGE

人类基因组大约含有 10 万个功能基因,其中在 Genbank 或 EST 数据库中已经注册的基因占基因组总基因数目的 40%。因此通过 SAGE 方法所获得的转录组的标签中,总有部

分标签未能与数据库中已知基因相匹配。然而获取这些未知标签的信息对于充分了解全基因组的功能以及发现新基因具有较大的参考价值。由于 SAGE 文库所获得仅是 9-11bp 的标签(片段太短),无法作为探针或引物从其基因组中调出全基因。Anke van den Berg 等依据逆转录-PCR(RT-PCR)原理建立了一种快速分析未知 SAGE 标签的方法,叫做 RAST-PCR(Rapid analysis of unknown SAGE tags),该课题组以此法获得了 6 个未知标签的 cDNA 大片段,这些片段长度分别为 204bp、112bp、104bp、214bp、82bp、186bp<sup>[13]</sup>。Yu 在 SAGE 专业网页上刊登了一种叫做逆转录-SAGE(Reverse SAGE-rSAGE)的实验方法<sup>[14,15]</sup>。该方法与 RAST-PCR 方法相比较,除了引物设计略有不同外,其余基本一致。它们的实验步骤主要如下:1. 以 5'末端含 M13 尾巴的 Oligo(dT)<sub>9</sub> 为引物逆转录合成 cDNA (引物:5'-生物素 CCGGGCGCGCCGTA AAAACGACGGC-CAGT<sub>1-19</sub>,该引物中含有一个 Asc I 的酶切位点)2. *Not* III 消化,磁珠法分离 cDNA 片段(该步骤同常规 SAGE)3. 以常规 SAGE 方法中的 linker2 与上步所得 cDNA 片段连接4. *Asc* I 酶切释放磁珠上 3'cDNA 片段5. 根据 linker2 和逆转录引物序列设计 PCR 扩增的双引物,其中 linker2 的 3'末端含某一待扩增的标签序列,例如:该引物可设计为:5'-CTGCTCGAAT-TCAAGCTTCT-9 bp 标签序列6. PCR 扩增,获得 cDNA 大片段。

综上所述,SAGE 方法不仅可以定量分析生物体在不同时期转录组的全部信息,而且是全面了解生物体在不同条件下基因表达的一种强有力工具,随着更多的研究人员对 SAGE 方法的关注,该方法将会在更多的领域中发挥其独特优势,为后基因组计划做出更大的贡献。

#### REFERENCES(参考文献)

[1] LI Z Y(李子银),CHNE S Y(陈受宜). Progress in functional genomics in plants. *Hereditas*(遗传)2000 **22**(1):57~60

[2] YANG Q(杨春),LU S D(卢圣栋). Serial analysis of gene expression. *Foreign Medical Science, Radiology Unclear medicine Fascicule* (国外医学,放射医学核医学分册),1999 **23**(3):131~134

[3] ZHOU J(周津),WANG X Q(王秀琴),LIU Z H(刘芝华) *et al.* The technique of serial analysis of gene expression. *Chin J Med Genet* (中华医学遗传学杂志),1999 **16**(6):397~399

[4] Yamashita T, Hashimoto S H, Shuichi K *et al.* Comprehensive gene expression profile of a normal human liver. *Biochemical and Biophysical Research Communication*, 2000 **269**:110~116

[5] ZHANG L, ZHOU W, Velculescu V E *et al.* Gene expression profiles in normal and cancer cells. *Science*, 1997 **276**:1268~1272

[6] Madden S L, Wang C J, Landes G. Serial analysis of gene expression from gene discovery to target identification. *Drug Discov Today*, 2000 **5**(9):415~425

[7] Velculescu V E, ZHANG L, Vogelstein B *et al.* Serial analysis of gene expression. *Science*, 1995 **270**:484~487

[8] Angelastro J M, Klimaschewski L P, Vitolo O V. Improved *Not* III digestion of PAGE-purified 102bp ditags by addition of a single purification step in both the SAGE and microSAGE protocols. *Nucleic Acids Research* 2000 **28**(12):e62

[9] Powell J. Enhance concatemer cloning--a modification to the SAGE (Serial Analysis of Gene Expression) technique. *Nucleic Acids Research*, 1998 **26**(14):3445~3446

[10] Kenzelmann M, Muhlemann K. Substantially enhanced cloning efficiency of SAGE(Serial Analysis of Gene Expression) by adding a heating step to the original protocol. *Nucleic Acids Research*, 1999 **27**(3):917~918

[11] Datson N A, van der Perk-de Jong J, van den Berg M P. MicroSAGE: a modified procedure for serial analysis of gene expression in limited amounts of tissue. *Nucleic Acids Research*, 1999 **27**(5):1300~1307

[12] YE S Q, ZHANG L Q, ZHENG F *et al.* MiniSAGE: gene expression profiling using serial analysis of gene expression from 1 $\mu$ g total RNA. *Analytical Biochemistry*, 2000 **287**:144~152

[13] van den Berg A, van der Leij J, Poppema S. Serial analysis of gene expression: rapid RT-PCR analysis of unknown SAGE tags. *Nucleic Acids Research*, 1999 **27**(17):e17

[14] Neilson L, Andalibi A, Kang D. Molecular phenotype of the human oocyte by PCR-SAGE. *Genomics*, 2000 **63**(1):13~24

[15] Polyak K, XIA Y, Zweier J L *et al.* A model for p53-induced apoptosis. *Nature*, 1997 **389**(6648):300~305

## Technological Advances of Serial Analysis of Gene Expression

XIONG Yong-Hua\* XU Yang

(Sino-German Joint Research Institute, Nanchang 330047, China)

**Abstract** Serial analysis of gene expression (SAGE) is an effective method of determining gene expression profiles of tissues and organs under different conditions. In this paper, the detail protocol of SAGE was introduced and some modified procedure of SAGE was reviewed.

**Key words** SAGE, modified SAGE, microSAGE, rSAGE