

微量材料系列性基因表达分析技术的研究*

郑芳¹⁾ 周新^{1)**} 严明²⁾ 叶水清³⁾ 刘芳¹⁾

(¹⁾武汉大学中南医院基因诊断中心, 武汉 430071; ²⁾武汉大学中南医院眼科, 武汉 430071;

³⁾约翰·霍普金斯大学医学院, 巴尔的摩 21218, 美国)

摘要 建立了微量材料 (1 μ g RNA) 系列性基因表达分析技术 (minimal serial analysis of gene expression, miniSAGE), 并用 miniSAGE 技术对冠心病 (coronary artery disease, CAD) 患者和正常人成纤维细胞的基因表达谱进行了分析, 合成了两个高质量的 SAGE 库, 比较了两个库的 SAGE 标签, 为 CAD 发病机理的研究提供了大量表达异常基因的资料. 从而为后续分离与 CAD 发病相关的新基因打下了基础. miniSAGE 技术是一种有力的基因表达分析方法, 可用于微量 RNA (1 μ g) 材料的基因表达分析, 为多因素复杂疾病的研究提供了有力的基因表达分析工具.

关键词 系列性基因表达分析技术, 冠心病, 成纤维细胞

学科分类号 R54.402

当前有两大高通量的基因表达分析技术^[1]: 系列性基因表达分析技术^[2]和 DNA 芯片技术^[3]. 这两大 RNA 分析技术都具有对样品的基因表达谱进行全面分析的能力. 它们有可能为综合性疾病如冠状动脉性疾病、糖尿病和癌症提供一些新的研究方向, 这些流行病是由多种环境和遗传因子所导致的, 其致病机理比较复杂. 而 SAGE 法在发现新基因和提供量化的数据方面, 比 DNA 芯片更有优势. SAGE 法是基于以下两个原理构建的:

a. 在 cDNA 特定区域的一小段寡核苷酸, 足以特异性地代表一个基因转录所形成的 cDNA (转录本), 这一小段寡核苷酸被称为标签; b. SAGE 法中的标签可头尾相连地串联起来形成串联体 (concatemer), 以便于后续的测序分析. SAGE 法的技术路线简要介绍如下: 分离待测标本的总 RNA, 利用生物素标记的引物寡聚 T18, 以 RNA 为模板合成生物素标记的 cDNA. cDNA 片段被生物素标记的 3' 端与包被有链霉抗生物素的试管结合. 用一种“锚定酶” (例如 *Nla* III) 切割 cDNA, 接着, 用反应缓冲液冲洗试管. cDNA 5' 端的 cDNA 片段被冲洗掉, 而 3' 端的 cDNA 片段仍然被固定在试管上. 含有标记酶 (如 *Bsm* FI) 识别位点的两种寡核苷酸 (Linker) 分别连接到这种结合型 cDNA 片段的 5' 端, 这样形成一个 Linker+ 3' 端 cDNA 片段的标记物复合体. 标记酶 (*Bsm* FI), 识别 Linker 的序列, 使 Linker+ 3' 端的 cDNA 标记物复合体从珠状物上消化下来, 形成 “Linker+ 标

签复合体”. 复合体的 3' 端被钝化, 带有两种不同 Linker 的短标签可相互连接形成 102 bp 大小的双标签. 将包含两个标签 (SAGE 双标签) 和两个 Linker 的 “Linker 结合的双标签” 进行 PCR 扩增. “Linker 结合的 SAGE 双标签” 被锚定酶 (*Nla* III) 切割成大约 26~28 个碱基对的寡核苷酸聚合物 (SAGE 双标签), 经纯化、连接形成串联体 (concatemer), 克隆到测序载体. 利用 SAGE 软件对测序结果进行分析, 并将标签分类和量化. 这些标签反映了细胞或组织中转录本的特性和数量, 代表了其基因表达谱.

虽然 SAGE 法已成为一种高通量的基因表达分析技术, 但是 SAGE 法也有其缺点, 它对起始 mRNA 的量要求比较高, 需要 2.5~5.0 μ g mRNA (相当于 250~500 μ g 总 RNA). 这使得 SAGE 法的应用受到限制. 为此, 我们建立了改良的微 SAGE 法. 在这一方法中引入了三个关键的改良步骤: a. 将锁相胶 (phase lock gel, PLG) 用于 DNA 的纯化, 增加了回收 DNA 的纯度和产量; b. 减少 Linker 的用量, 避免了 Linker 的毒害作用; c. 采用 mRNA 俘获试剂盒, 将 mRNA 的分离, cDNA 的合成, cDNA 的酶切, Linker 和结合在试

* 本文部分工作是在美国约翰·霍普金斯大学完成的.

** 通讯联系人.

Tel: 027-87317802, E-mail: zhouxjyk@sina.com

收稿日期: 2001-10-12, 接受日期: 2001-12-29

管壁上的 cDNA 连接, 和 cDNA 标签的释放, 这五个步骤在同一个试管中完成, 减少了反应过程中材料的损耗。

1 材料和方法

1.1 成纤维细胞的培养和总 RNA 的分离提取

研究对象来自美国约翰·霍普金斯大学的冠心病研究人群 (JH-CAD), 选取 CAD⁺ 和 CAD⁻ 先证者。从培养的皮肤活检组织中创建两个研究对象的成纤维细胞系 (JT 和 WY)^[4]。将成纤维细胞接种在组织培养瓶 ($4 \times 10^5/\text{cm}^2$) 中, 在 D-MEM/F-12 胎牛血清培养基 (Cat. No. 11330-032, GIBCOBRL, Grand Island, New York, 10% (体积比)) 中培养。培养基中含青霉素 100 U/ml, 链霉素 100 mg/L, 细胞生长 7 天直到 90% 的细胞融合。根据试剂商的说明往每个培养瓶中加入 2 ml TRIZOL 试剂 (Cat. No. 15596, GIBCOBRL, Grand Island, New York) 进行 RNA 分离。在 DU640 分光光度计 (Beckman Inserument, Inc., USA) 上读取 A_{260} 值测定 RNA 的量。通常在这种情况下, 可以从一个组织培养瓶中获得大约 25~35 μg 的成纤维细胞总 RNA。在由 30 μl (25~35 μg) 总 RNA, 10 μl 10 $\times$ DNase I 缓冲液, 5 μl DNase I (1 U/ μl) (Cat. No. 776-785, Boehringer Mannheim), 55 μl 无 Dnase 和 Rnase 的水组成的逆转录反应体系中, 反应 30 min 可以清除污染的基因组 DNA。总 RNA 在 -80 $^{\circ}\text{C}$ 冷冻保存备用。

1.2 串联的 SAGE 标记物的产生

标准 SAGE 操作程序的最初 5 步: mRNA 制备, cDNA 合成, 锚定酶酶切生物素标记的 cDNA (生物素标记的 cDNA 3' 端与磁性珠状物结合), Linker 与结合的 cDNA 5' 端连接, 标记酶酶切释放 Linker 结合的 cDNA 标记物, 这些都可以使用 mRNA 捕获包 (Cat. No. 1787896, Boehringer Mannheim) 在一个试管中完成。

1.2.1 mRNA 制备: 用 46 μl 溶解缓冲液 (由 mRNA 捕获包提供, 包含低于 2% 的十二硫酸锂) 溶解 1 μg 成纤维细胞总 RNA, 溶解的 RNA 与 4 μl (5 $\mu\text{mol/L}$) 生物素标记的寡 (dT)₂₀ 引物混合, 混合物在 37 $^{\circ}\text{C}$ 退火 5 min。再将 50 μl 的反应混合物均等地分配到两个链霉抗生物素包被的 PCR 试管中, 生物素标记寡 (dT)₂₀ 捕获的 mRNA 在试管壁上固定 3 min (37 $^{\circ}\text{C}$)。溶液中游离型的 RNA 被吸出来, 每次操作用 50 μl 冲洗液冲洗试管 3 次。

1.2.2 cDNA 合成: 用 SuperScript Clone (Cat. No. 18090-019, GIBCOBRL, Grand Island, New York) 体系合成 cDNA。在上述试管中, 用 50 μl 的 1 $\times$ 第一链缓冲液冲洗结合型 mRNA 1 次。接着在 42 $^{\circ}\text{C}$, 含有 4 μl 5 $\times$ 第一链缓冲液, 2 μl 的 0.1 mol/L DTT, 1 μl 的 10 mmol/L dNTPs, 1 μl SuperScript II 逆转录酶 (200 U/ μl) 和 12 μl DEPC 处理水的 20 μl 反应体系中孵育 2 h, 合成第一链 cDNA。除去第一链合成溶液后, 用 50 μl 冲洗液冲洗 PCR 试管 1 次, 加入 50 μl 1 $\times$ 第二链缓冲液。在 16 $^{\circ}\text{C}$ 含有 4 μl 5 $\times$ 第二链缓冲液, 0.4 μl 10 mmol/L dNTPs, 1 μl DNA 多聚酶 I (10 U/ μl ; GIBCOBRL), 2.5 μl T4 DNA 连接酶 (1 U/ μl), 0.5 μl RNaseH (1 U/ μl) 和 11.6 μl 水的 20 μl 反应体系中孵育 2 h, 合成第二链 cDNA。

1.2.3 锚定酶 *Nla* III 酶切生物素标记的 cDNA: 将第二链反应混合物清除, 用 50 μl 冲洗液冲洗 PCR 试管一次, 再加入 50 μl 1 $\times$ NE Buffer 4 (New England Biolabs, Beverly, MA)。在 37 $^{\circ}\text{C}$ 含有 2 μl *Nla* III (10 U/ μl), 2.5 μl BSA (1 g/L), 2.5 μl 10 $\times$ NE Buffer4 和 18 μl 无 DNase 的 25 μl 反应体系中, 锚定酶 *Nla* III 消化 cDNA 达 2 h。*Nla* III 在 65 $^{\circ}\text{C}$ 孵育 20 min 后失活。

1.2.4 Linker 与结合在珠状物上的 cDNA 结合: Linker 1 [5' TTT GGA TTT GCT GGT GCA GTA CAA CTA GGCTTAATAGGGACATG3' 与互补链 5' TCC CTA TTA AGC CTA GTT GTA CTG CAC CAG CAAATCC (amino mod. C7) 3'] 和 Linker 2 [5' TTT CTG CTC GAA TTC AAG CTT CTA ACG ATGTACGGGGACATG3' 与互补链 5' TCC CCG TAC AAT CGT TAG AAG CTT GAA TTC GAGCAG (amino mod. C7) 3'] 被激酶作用退火^[5]。消化反应后的游离型 5' 端 cDNA 被清除后, 用 50 μl 冲洗液和 50 μl 1 $\times$ 连接酶缓冲液 (GIBCOBEL) 冲洗 PCR 试管一次。退火的 Linker 1 和 Linker 2 分别加入到试管 1 和试管 2 中, 每个试管的总体积为 25 μl , 包含 0.5 μl 的 Linker (20 mg/L), 2.5 μl 的 10 $\times$ 连接酶缓冲液和 21 μl LoTE (3 mmol/L Tris-HCl, pH7.5, 0.2 mmol/L EDTA, pH 7.5)。将试管在 50 $^{\circ}\text{C}$ 加热 2 min, 再在室温放置 15 min。往试管中加入 1 μl T4 DNA 连接酶 (5 U/ μl , Cat. No. 15224-041, GIBCOBRL, Grand Island, New York), 混合物在 16 $^{\circ}\text{C}$ 孵育 2 h。

1.2.5 标记酶 *Bsm* FI 释放 cDNA 标记物: 弃掉

上述连接反应溶液后, 用 50 μ l 冲洗液和 50 μ l 1 $\times$ NE Buffer 4 冲洗 PCR 试管. 在 65 $^{\circ}$ C 含有 1 μ l *Bsm*F I (2 U/ μ l Cat. No. 572S, New England Biolabs, Beverly, MA), 2.5 μ l 10 $\times$ NE Buffer 4, 2.5 μ l BSA (1 g/L), 和 19 μ l LoTE 的 25 μ l 反应体系中孵育 1 h, Linker 结合的 cDNA 标记物被释放.

1.3 SAGE 双标签 26~28 碱基聚合物的形成

1.3.1 释放 Linker 结合的 SAGE 标签末端的钝化: *Bsm*F I 消化后, 往含有释放 Linker 结合的 SAGE 标签反应溶液中加入 LoTE, 将溶液总体积补充到 200 μ l 再转入含有轻锁相胶 (Phase Lock GelTM LPLG, Cat. No. 0032005-101, Eppendorf Sci., Inc., Westbury, New York) 的 2 ml 试管中, LPLG 在使用前以 13 000 $\times g$ 离心 20 s. 再加入 200 μ l 苯-氯仿-异戊醇 (PCI, 25:24:1), 将试管摇晃 5~7 次完全混合. 22 $^{\circ}$ C, 13 000 $\times g$ 离心 5 min 使水相和有机相分离. 将上层水相中的 SAGE 标签用吸管移至一洁净试管中, 加入 3 μ l 糖原 (1 g/L), 133 μ l 7.5 mol/L 醋酸铵和 700 μ l 100% 乙醇, 充分振荡混匀. 在 4 $^{\circ}$ C, 13 000 $\times g$ 离心 15 min, 使核酸沉淀 (200 μ l 样品). 吸去上清液, 用 70% 乙醇清洗两次, 室温干燥 10 min, 重新悬浮于 18.5 μ l LoTE 中. 在 37 $^{\circ}$ C, 含有 18.5 μ l Linker 结合的 SAGE 标签, 6 μ l 5 $\times$ 第二链缓冲液, 3 μ l BSA (1 g/L), 0.5 μ l 25 mmol/L dNTPs, 2.0 μ l DNA 多聚酶 (Klenow) 大片段 (1 U/ μ l; Cat. No. 27-0929-01, Amersham Pharmacia Biotech) 的 30 μ l 反应体系中进行钝化末端反应 30 min.

1.3.2 Linker 结合的 SAGE 标签相互连接形成 Linker 结合的 SAGE 双标签: 钝化反应后, 如上述完成 PCI 提取和乙醇沉淀, 沉淀重新悬浮在 4 μ l LoTE 中. 在包含 0.8 μ l T4 DNA 连接酶 (5 U/ μ l, Cat. No. 15224-041, GIBCOBRL), 1.2 μ l 5 $\times$ 连接酶缓冲液以及试管 1 中的 2 μ l 钝性末端样品 1 和试管 2 中的 2 μ l 钝性末端样品 2 的 6 μ l 反应体系中, 16 $^{\circ}$ C 过夜完成连接. 同时制备不含连接酶的阴性对照.

1.3.3 Linker 结合的双标签的 PCR 扩增: 连接反应后, 加入 14 μ l LoTE 使体积增加到 20 μ l. 制备 0, 10, 20, 100, 200 倍的稀释样品. 每个稀释的混合物中取出 1 μ l 加入 100 μ l 的 PCR 反应体系, 该体系包括 5U PLATINUM Taq 多聚酶 (Cat. No. 10996-026, GIBCOBRL), 1.5 mmol/L 的 $MgCl_2$,

0.2 mmol/L 的 dNTPs 和 0.8 μ mol/L 生物素标记的 SAGE 引物. 在 96 孔基因扩增仪上, 以 95 $^{\circ}$ C 30 s, 55 $^{\circ}$ C 1 min, 70 $^{\circ}$ C 1 min 为条件扩增, 循环 28 次, 酶的起始活化条件为 95 $^{\circ}$ C 4 min, 最后一次的延伸条件为 70 $^{\circ}$ C 5 min. 用 288 个 PCR 反应建立一个 SAGE 库.

1.3.4 Linker 结合的 SAGE 双标签的分离: 收集 288 个 PCR 试管中的 28.8 ml PCR 产物, 将其均等加到 6 个含有轻 Phase Lock GelTM 的 15 ml 试管中, LPLG 一定要在使用前以 1 500 $\times g$ 离心 1 min. 每个试管中加入 4.8 ml PCI, 上下颠倒 5~7 次使之完全混匀. 在 22 $^{\circ}$ C 以 1 500 $\times g$ 离心 5 min 使水相和有机相分离. 用吸管将上层水相中的“Linker 结合的双标签”移入一洁净试管中. 在 4 $^{\circ}$ C, 以 17 500 $\times g$ 离心 30 min, 使核酸沉淀于乙醇中. 每个试管中的沉淀用 70% 的乙醇清洗两次, 在室温中, 空气干燥 10 min, 重新悬浮于 35 μ l LoTE 中. 最终 6 个试管总共得到 210 μ l 总体积的样品.

在样品中加入 54 μ l 的 1 $\times$ 上样缓冲液 (GLS, Cat. No. G2526, St. Louis, MO). 在两个 12% 聚丙烯酰胺凝胶的 22 个点样孔中, 每孔加入 12 μ l 样品混合物, 每个凝胶以 10 bp 大小的梯带 (GIBCOBRL) 作为标准物. 电泳后, 将凝胶置入 1 $\times$ 流动缓冲液中用 SYBR Green I (1:10 000 稀释) 染色. 将 22 条泳道中, 与 102 bp 标准物作同样的相对运动的凝胶条带切割下来. 在 6 个 1.5 ml 试管的管底, 用一根规格 21 的针打孔. 将 2.5 ml 的试管套在 1.5 ml 的试管外面. 在室温中, 以 14 000 $\times g$ 离心 5 min 使凝胶片段破碎. 加入 300 μ l LoTE 在 65 $^{\circ}$ C 温育 15 min, 使 DNA 从凝胶片段中洗脱出. 将凝胶碎片和 LoTE 的混合物转到 Spin X 柱 (Costar) 内, 在 22 $^{\circ}$ C 以 13 000 $\times g$ 旋转 5 min, 聚丙烯酰胺凝胶会残留到柱子上, 而洗脱液为含 DNA 的 LoTE.

1.3.5 SAGE 双标签的纯化

Spin X 柱中的洗脱液按上述 PCI 纯化, 乙醇沉淀. 沉淀重新悬浮于 70 μ l LoTE 中. 在 37 $^{\circ}$ C 含有 70 μ l 样品, 10 μ l *Nla* III (10 U/ μ l), 10 μ l BSA (1 g/L), 10 μ l 10 $\times$ NE Buffer 4 的 100 μ l 反应体系中, 用锚定酶 *Nla* III (Cat. No. 125S, New England Biolabs, Beverly, MA) 消化“Linker 结合的 SAGE 双标签”1 h. 消化后, PCI 提取、PLG 分离和乙醇沉淀 DNA (离心前用干冰/甲醇低温孵溶 10 min), 沉淀重新悬浮于 35 μ l LoTE 中.

沉淀与 8 μ l 染料混合点样于 4 个点样孔, 每孔加样 10 μ l, 在 12% 聚丙烯酰胺凝胶上用 10 bp DNA 标准物. 电泳后, SAGE 双标签的大小大约为 22~26 bp, 如上述过程洗脱, 仅将温育温度改为 37 $^{\circ}$ C. 将沉淀重新溶解于 100 μ l LoTE 中.

根据 Powell^[6] 的描述, 用链霉抗生素包被的磁珠 (Dyna streptavidin-coated magnetic beads, Dynabeads M-280 streptavidin, Dynal A. S., Oslo, Norway) 清除污染的 Linker. 按照上述步骤进行 PCI 提取、PLG 分离和乙醇沉淀 (离心前用干冰/甲醇孵浴 10 min) 后, 将纯化的 SAGE 双标签重新溶解于 7 μ l LoTE 中.

1.4 SAGE 标签的克隆、测序和分析

1.4.1 SAGE 双标签连接形成串联体 (concatemer): 在 16 $^{\circ}$ C, 含有 1 μ l T4 DNA 连接酶 (5 U/ μ l, Cat. No. 15224-041, GIBCOBRL), 2 μ l 5 $\times$ 连接酶缓冲液, 7 μ l SAGE 双标签的 10 μ l 反应体系中, 纯化的双标签相互连接, 作用 3 h. 连接后, 样本在 65 $^{\circ}$ C 加热 5 min, 与 2.5 μ l 5 $\times$ 上样缓冲液 (GLS) 混合, 点样于 8% 聚丙烯酰胺凝胶的点样孔, 以 100 bp 梯带作为标准物. 用 SYBR Green I 染色后, 切割 600~1200 bp 之间的凝胶带, 回收凝胶中所含的 SAGE 串联体. 串联体如上述 PCI 纯化, 经乙醇沉淀后, 将沉淀重新悬浮于 6 μ l LoTE 中.

1.4.2 串联体 (SAGE 标签串联体) 的克隆和测序: 在含有 1 μ l (1 μ g) pZErO-1 载体 (Cat. No. K2500-01, Invitrogen, Carlsbad, CA), 1 μ l *Sph* I (10 U/ μ l, New England Biolabs), 2 μ l BSA (1 g/L), 2 μ l 10 $\times$ NE Buffer 2 和 14 μ l LoTE 的 20 μ l 反应体系中消化 15 min. 消化后, 按上述进行 PCI 提取、PLG 分离和乙醇沉淀, 将沉淀重新悬浮于 90 μ l LoTE 中. 最后载体的浓度大约是 10 mg/L.

在 16 $^{\circ}$ C, 含有 1 μ l T4 DNA 连接酶 (5 U/ μ l, Cat. No. 15224-041, GIBCOBRL), 2 μ l 5 $\times$ 连接酶缓冲液, 1 μ l 消化后载体和 6 μ l 串联体样本的 10 μ l 反应体系中反应 30 min. 完成 *Sph* I 消化后的载体与串联体的连接. 连接后, 经 PCI 提取、PLG 分离和乙醇沉淀来纯化样本. 最后, 将沉淀重新悬浮于 6 μ l LoTE 中.

用 GIBCOBRL Cell-Porator (配置包括电容: 330 μ F; 电阻: 低 (Ω); 电荷率: 快; 电压: 400 V) 将含重组串联体的质粒转染感受态大肠杆

菌 (ElectroMax DH10Bs, Cat. No. 18290-015, GIBCOBRL). 电压可以用 GIBCOBRL Voltage Booster (能承受 4 kV) 提高到 2.5 kV. 转染后, 在含抗生素 Zeocine 的琼脂平板上, 每 10 cm^2 平铺 1/10 (大约 100 μ l) 的转染菌, 37 $^{\circ}$ C 生长 12~16 h. 使用载体特异性的引物, 通过 PCR 来鉴定阳性菌落. 只有大于 650 bp 的 PCR 产物 (至少包含 30 个标记物) 才被选择进行测序分析. 在马里兰州, 巴尔的摩市, 约翰·霍普金斯大学医学院 DNA 分析部, 用 377ABI 自动测序仪可以直接对 PCR 产物进行测序. ABI 仪器利用改良了的四染料荧光标记法和全时间扫描探测进行单泳道测序. 单泳道测序的生化机制与 Sanger-based 法相同.

用 SAGE 软件 3.04 进行序列谱分析^[5].

2 结果

2.1 Linker 结合的 SAGE 双标签的 PCR 扩增

两个“Linker 结合的 SAGE 标签”相连形成 Linker 结合的 SAGE 双标签, 根据 SAGE 法^[7] 不需要额外的 PCR 扩增, 可以精确地对 Linker 结合的 SAGE 双标签进行 PCR 扩增. 图 1 显示一个在 12% 聚丙烯酰胺凝胶上分离, 典型的 Linker 结合的 SAGE 双标签扩增产物的凝胶电泳模式. 从 Linker 结合的 SAGE 双标签扩增产物泳道 (1、2) 中, 可见扩增了的 Linker 结合的 SAGE 双标签的 102 bp 条带, *M* 为 10 bp 的 DNA 标准. 我们用 1/20 比例稀释 Linker 结合的 SAGE 双标签, 以其稀释液为模板作了一个大的 PCR 扩增 (288 个反应体系).

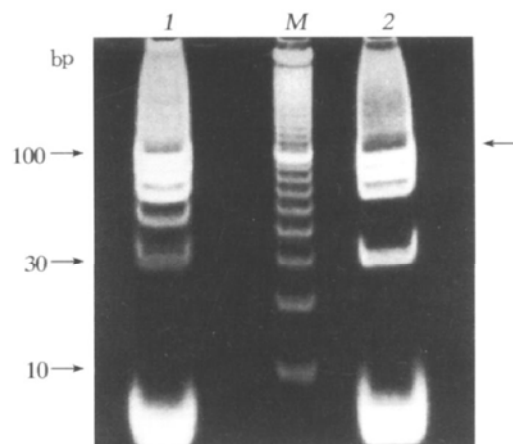


Fig. 1 Linked SAGE ditags

M: 10 bp DNA ladder; 1, 2: linked SAGE ditags.

2.2 SAGE 双标签的纯化和克隆

“SAGE 双标签+ Linker”被锚定酶切割成大约 26~28 bp 的聚合物, 这一聚合物含 2 个 SAGE 标签 (图 2). 这些 SAGE 标签经纯化、连接形成测序串联体 (concatemer) (图 3). 将 600~1 200 bp 大小的串联体从聚丙烯酰胺凝胶上切割下来, 经纯化克隆到测序载体上.

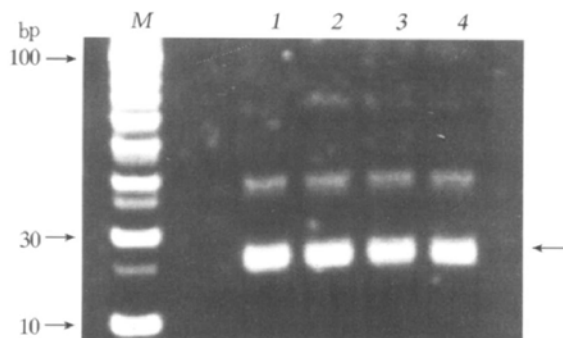


Fig. 2 SAGE ditag

M: 10 bp DNA ladder; 1~4: SAGE ditags.

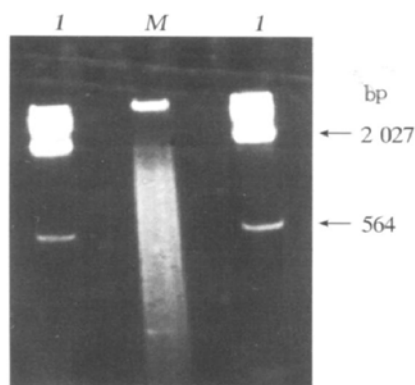


Fig. 3 SAGE concatemers

M: ADNA/*Hind* III fragments; 1: SAGE concatemers.

2.3 插入 pZErO-1 载体中的 SAGE 标签串联体大小

SAGE 标签串联体的插入物大小标志着 SAGE 库的质量. 用一对引物: 5'-GTAAAACGACGGC-CAGT-3' 和 5'-GGAAACAGCTATGACCATG-3', 通过 PCR 扩增, 克隆的插入物 SAGE 标签串联体, 并用 2% 琼脂糖凝胶分离 (图 4). 期望得到 600~

1 200 bp 大小的 SAGE 标签串联体的插入物. 大于 650 bp 的插入物将被测序, 因而在每一个反应体中得到最少 30 个 SAGE 标签, 将 PCR 产物直接测序. 由于本文采用 *Nla* III 为锚定酶, 其识别序列为 CATG. 所以在测序结果的图谱上可见 SAGE 双标签被 CATG 分隔开来 (图 4).

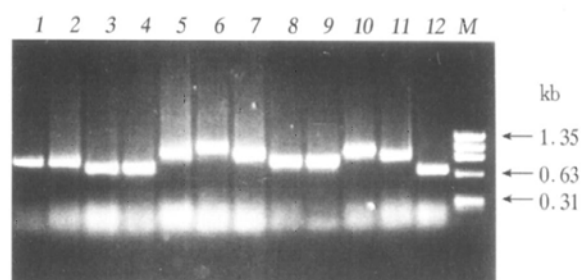


Fig. 4 Insert sizes of SAGE clones

M: ϕ X174 RF DNA/*Hae* III fragments; 1~12: The PCR products of inserts in SAGE clones.

2.4 SAGE 标签的初步分析

我们用微 SAGE 技术成功合成了两个成纤维细胞 SAGE 库: JT, 来自正常人群; WY, 来自早发冠心病和高 β 脂蛋白血症患者^[4]. 从两个库中分别获得 25 350 和 26 648 个 SAGE 标签. 来自 SAGE 库的标签初步分析见表 1 和表 2. 用 Kinzler 提供的 SAGE 软件 3.04 版, 对照目前的人类基因 SAGE 图谱 (<http://www.sagenet.org/SAGEDatabases/unigene.htm>) 来查询 SAGE 标签^[5].

在 JT 库中, 25 350 个标签代表了 11 428 个基因 (表 1). 其中, 4 084 个 (35.7%) 与基因库中的已知基因相对应, 1 721 个 (15.1%) 与表达序列标记物 (EST) 相匹配, 5 623 个 (47.2%) 不与任何已知的基因和 EST 相配对. 在 WY 库中, 26 648 个标签代表了 11 986 个基因. 其中, 4 400 个 (36.7%) 与已知基因相匹配, 1 867 个 (15.6%) 与 EST 相对应, 5 719 个 (47.7%) 不与任何已知基因或 EST 相匹配.

Table 1 Overall summary of the preliminary analysis of two fibroblast SAGE libraries

	Total tags	Represented genes	Matched genes ¹⁾	Matched to EST	Unmatched genes
JT	25 350	11 428	4 084 (35.7%)	1 721 (15.1%)	5 623 (47.2%)
WY	26 648	11 986	4 400 (36.7%)	1 867 (15.6%)	5 719 (47.7%)

¹⁾ Match to GenBank entries.

表 2 概括了从表达丰度的角度对两个库中 SAGE 标签的初步分析. 每个细胞的基因表达水平范围从 1 拷贝到 86 拷贝, 但大部分基因 (82%) 以 1 拷贝表达. 低丰度表达的基因占 mRNA 总量的 49%, 它们仅占 mRNA 总量的一小部分. 这项

分析与 Zhang 等^[7] 的结论相一致, 他们是以对 303 706 个 SAGE 标签的分析为基础得到的结论.

表 3 列出了 30 个在 WY 文库中高表达的基因, 它们在每个细胞中的表达超过了 10 拷贝.

Table 2 Summary of fibroblast SAGE libraries by abundant classes

Frequency (Copies/ cell)	Total tags				Unique tags			
	No.		%		No.		%	
	JT	WY	JT	WY	JT	WY	JT	WY
> 50	4907	3332	18	13	37	32	0.3	0.3
> 10 and ≤50	4794	4834	18	19	244	238	2.1	2.0
> 5 and ≤10	2302	2031	9	8	308	277	2.7	2.3
> 1 and ≤5	6055	5975	23	24	2249	2261	19.7	18.9
1	8590	9178	32	36	8590	9178	75.2	76.6
Total	26648	25350	100	100	11428	11986	100	100

Table 3 List of top 30 genes from WY library

Tag sequence	Copies	UG No. ¹⁾	Gene description
TGGGGTTTC	85	62954	ferritin, heavy polypeptide 1
TTGGTCCTCT	56	108124	<i>H. sapiens</i> mRNA for homologue to yeast RP L41
GCCCCAATA	43	227751	Human sequence from clone 37E16 on chromosome 22
GCCCCAATA	43	198261	lectin, galactoside binding, soluble, 1 (galectin 1)
TTGGTGAAGG	42	75968	thymosin, beta 4, X chromosome
GTTGTGGTTA	41	75415	beta 2 microglobulin
CCCATCGTCC	26	28968	ESTs, Highly similar to oxidase II, cytochrome [<i>H. sapiens</i>]
ATAAAAAGAA	23	83942	cathepsin K (pseudosynthesis)
AGGCTACGGA	23	119122	<i>H. sapiens</i> mRNA for 23 ku highly basic protein
GTGACCACGG	21	36451	glutamate receptor, ionotropic, N-methyl D-aspartate 2C
AGAGGTGTAG	21	16275	calcium channel, voltage-dependent, beta 3 subunit
TACCATCAAT	20	195188	Human keratinocyte subtraction library mRNA, clone H22a
TACCATCAAT	20	74456	glyceraldehyde 3-phosphate dehydrogenase
CCCTGGGTTTC	20	111334	ferritin, light polypeptide
TGCAGTCACT	19	132826	ESTs
TGCAGTCACT	19	83169	matrix metalloproteinase 1 (interstitial collagenase)
TAGGTTGTCT	18	119252	tumor protein, translationally-controlled 1
AAGACAGTGG	18	184109	ribosomal protein L37a
CCTAGCTGGA	17	182937	peptidylprolyl isomerase A (cyclophilin A)
AGCACCTCCA	16	75309	eukaryotic translation elongation factor 2
TGCACGTTTT	15	169793	ribosomal protein L32
CCAGAACAGA	14	79006	Human thymidylate kinase (CDC8) mRNA, complete cds
CCAGAACAGA	14	111222	ribosomal protein L30
AGCTCTCCCT	13	196322	stromal cell-derived factor 1
AGCTCTCCCT	13	82202	ribosomal protein L17
TTTAACGGCC	12	227211	calcium/calmodulin-dependent protein kinase II γ
CACAAACGGT	12	195453	ribosomal protein S27 (metalloproteinase 1)
GGTCAGTCGG	12	19154	KRAB-associated protein 1
GAGGGAGTTT	11	76064	ribosomal protein L27a
AAGGAGATGG	11	184014	ribosomal protein L31

¹⁾ Unigene Cluster Number.

正如所预料的, 大部分表达丰富的基因是管家基因. 在它们中间, 大约四分之一编码核糖核蛋白; 延长基因 2 在每个细胞中以 16 拷贝表达, 它是蛋白质合成的必要成分. 铁蛋白, 重链多肽 (85 拷贝/细胞) 和轻链多肽 (20 拷贝/细胞) 都与铁的运输相关. 甘油醛-3-磷酸脱氢酶 (20 拷贝/细胞) 参与碳水化合物的代谢. 胸苷酸转录激酶 (14 拷贝/细胞) 参与 DNA 合成. 已知这些基因在大部分组织包括皮肤中高表达. 有些已被报道在成纤维细胞中表达的基因, 也发现在本 SAGE 库中多拷贝表达. 这些基因包括酸性成纤维细胞生长因子

细胞间结合蛋白 (Hs. 7768, 6 拷贝/细胞)、纤维结合素 1 (Hs. 118162, 4 拷贝/细胞)、胰岛素样生长因子结合蛋白 6 (Hs. 106283, 8 拷贝/细胞)、 $\alpha 1$ IV 型胶原蛋白 (Hs. 108885, 7 拷贝/细胞)、去皮质素 (Hs. 76152, 7 拷贝/细胞). 表 4 中概括了一些膜蛋白基因的表达数据. WY 库中, CD151 抗原和连接蛋白 43 (Connexin 43) 基因的转录表达水平比 JT 中高. 表 5 显示了一些分泌蛋白基因的表达数据. WY 中, 假拟蛋白 (hypothetical protein) FLJ20657 和纤连蛋白 1 (fibronectin1) 基因的转录表达水平比 JT 中高.

Table 4 Membrane proteins

Tag sequence	JT	WY	Description	Unigene
TTGCCCAGGC	0	1	interleukin enhancer binding factor 3	256583
GCAAAACCCC	0	2	CGF 119 protein	25615
GTGGTGGGTG	0	1	urea transporter	250592
AGCCACCACG	1	0	acetyl-LDL receptor (SREC)	198599
AAGATTGGTG	4	2	(p24) CD9 antigen	1244
AAGATTGGGG	5	4	CD44 antigen	169610
TGCCTCTGCC	7	17	CD151 antigen	75564
CCCCCACCTA	1	1	intestinal membrane A4 protein	245613
TGTTCTGGAG	2	14	connexin 43	74471

Table 5 Secreted proteins

Tag sequence	JT	WY	Description	Unigene
GCAAAACCCC	0	2	CGF 119 protein	25615
GCGAAACCT	0	1	erbB3 (soluble form)	209559
ATGTGAAGAG	10	20	hypothetical protein FLJ20657	164256
CCACTACACT	0	1	TRAIL (TNF ligand superfamily, member 10)	134610
TGCACGTTTT	39	58	LIM and cysteine rich domains 1	279943
GCCCCAATA	161	168	beta-galactoside binding protein	275964
ATCTTGTTAC	10	24	fibronectin 1	118162
AACGCGGCCA	3	4	macrophage migration inhibitory factor	231494

一些凋亡相关蛋白的基因表达数据在表 6 中列出. 发现 WY 中 DNA 断裂因子 (DNA fragmentation factor)、TRAIL (TNF 配体超家族 (ligand superfamily, 第 10 号) 和 Fas 活化的丝氨酸/苏氨酸激酶 (Fas-activated serine/threonine kinase) 的转录表达拷贝数比 JT 中高. 对于脂代谢问题, 我们概括了一些脂代谢中的关键基因 (表 7). WY

中, 低密度脂蛋白受体 (low density lipoprotein, LDLR)、羟甲基戊二酸单酰 CoA 还原酶 (HMG CoA reductase, HMGCR)、脂酶 A、乙酰 CoA 羧化酶的转录表达水平比 JT 中低. 但是 WY 中在脂肪酸 β 氧化途径中某些基因的转录表达拷贝数比 JT 中高 (表 8). 这些基因中大部分在 WY 或 JT 库的细胞表达水平是相等的. 与 JT 相比, WY 中

0.16% 的基因表达水平提高了 5 倍多, 而 0.10% 的基因表达水平下降了 10 倍. 其中, 在 WY 中表达水平比 JT 中高 10 倍以上的两个标记物 WY13 和 WY12, 分别与已知基因血红素加氧酶 (heme

oxygenase) 和 HL14 匹配. 在 WY 中表达水平比 JT 低至少 8 倍的前 10 个转录本中的两个标签 JT11 和 JT8, 分别与已知基因 Calcyin 和 Sea urchin homolog 匹配.

Table 6 Apoptosis associated proteins

Tag sequence	JT	WY	Description	Unigene
GCAAAACCCC	0	2	CGF119 protein	25615
AACCTGGGAG	2	4	DNA fragmentation factor, 45 ku, A subunit: KIAA0563	281819
GCGGACGAGG	1	1	programmed cell death 5	166468
CCACTACACT	0	1	TRAIL (TNF ligand superfamily, member 10) tyrosine 3-monooxygenase/tryptophan	134610
TGAGAGGGTG	3	3	5-monooxygenase	74405
CAACTTAGTT	1	12	death associated protein 6	180224
GAGGCGCTGG	0	2	BCL2-antagonist of cell death	76366

Table 7 Key genes of lipid metabolism

Unigene No.	Description	Symbol	Tag	JT	WY
Hs. 213289	LDL receptor	LDLR	AGGTCAGGAG	2	6
Hs. 11899	HMGCoA reductase	HMGCR	GAAGTGGCAG	4	0
Hs. 1644	7 alpha hydroxylase	CYP7A1	TAGTCATCAA	0	0
Hs. 14553	ACAT	SOAT1	GAAGATATTC	1	1
Hs. 85226	lipase A (Ch. esterase)	LIPA	AGAAGTGTCC	1	0
Hs. 28867	diacylglycerol O-acyltransferase	DGAT	TACGTCCACG	0	0
Hs. 7232	acetyl-CoA carboxylase	ACACA	CAAGGCTACT	0	1
Hs. 242908	LCAT	LCAT	CTCCTCACCT	17	19

Table 8 Fatty acid β -oxidation pathway

Unigene	Description	Symbol	Tag	JT	WY
Hs. 1209	acyl-CoA dehydrogenase	A-CADL	TGCCTGGCAT	0	2
Hs. 1531	enoyl-CoA hydratase	EHHADH	ACAGGAAAAC	0	1
Hs. 75860	hydroxyacyl-CoA dehydrogenase	HADHA	CCTCGCTCAG	0	1
Hs. 166160	Acetyl-CoA acyltransferase 1	ACAA1	AATTTAAAAA	0	1

3 讨 论

人类基因组 DNA 序列草图于 2000 年 6 月 26 日宣告完成, 然而人类基因组 DNA 序列仅代表基因复杂性的一个层次, 第二个更重要的层次就是这些基因在有机体内有时有序地表达^[1]. 而且现在, 只获得了人类基因组序列的草图, 其中包括许多间隙, 一些未知基因; 详细的, 完全的基因组序列可能还需一段时间来完成. 在不同细胞类型和相同细

胞类型的不同生理状态下, 其基因表达的总体概况都有助于确认可能参与不同疾病过程的候选基因, 另外, 基因在某些疾病组织异常表达的特性可能有助于发现一些能作为诊断标志, 预后指标或治疗目标的基因, 许多方法如 cDNA 敲除^[8], mRNA 差异显示^[9], EST^[10] 仅能分析有限数量的基因而且不能进行定量分析. 近来发展的以杂交为基础的 cDNA 芯片^[3], 寡核苷酸^[11] 芯片技术都局限于对已知基因的分析. 而杂交类检测方法都很相似, 不

能进行定量的分析, 只能进行定性的分析. SAGE 技术是一种更有效地从总体上评价基因表达特征的技术, 它能确认相关疾病的新基因, 提供出量化的数据.

然而, SAGE 需要毫克单位级的起始 mRNA. 在有些情况下因为仅能得到少量的起始 mRNA 而使这种方法的使用受限. 为了克服这个缺点, 几个实验室均试图从少量 mRNA 中获得 SAGE 库, 一般有两种方法: SAGE-Lite^[12] 和 PCR-SAGE^[13], 均用改良的 Clontech SMART 系统 PCR 扩增起始 cDNA; MicroSAGE^[14] 和 SADE^[15] 采用两步双标签的扩增法, 先将双标签扩增 28 个循环, 然后凝胶纯化, 再扩增 12~18 个循环, 这些额外的 PCR 扩增均可能产生扩增偏差及对 SAGE 法的定量分析的特性造成影响. Spinella 等^[16] 已证实任何 SAGE 标签的序列都可能影响 PCR 扩增产物, PCR 过程有利于短的模板扩增, 故长的模板可能在最后分析中具低于实际表达水平的拷贝; 某一模板被有效扩增的较多或较少取决于不同的模板序列. 双标签的两步 PCR 扩增法经发现可产生大量相同的双标签, 这些相同的双标签被认为是非特异性过度扩增的产物^[12], 尽管在最后用 SAGE 软件分析时可除去, 但是不必要地增加了测序成本及工作量. 本文建立的 mini SAGE 法可避免附加的 PCR 扩增步骤, 能由 1 μ g 总 RNA 来合成 SAGE 库, 将所需的起始 RNA 量降低了约 250~500 倍.

用 mini SAGE 技术已成功制备了两个高质量的 SAGE 库, 下列参数可支持这一结论: a. 从有限的 SAGE 标签分析中获得的典型的成纤维细胞基因表达谱; b. 在前 50 种丰富表达的基因中没有发现与 rRNA 序列相对应的标签, 在其他已发表的 3 种方法中, rRNA 在前 50 种丰富表达的标记物中的百分比分别为 4%^[12], 12%^[17] 和 24%^[18]; c. miniSAGE 库仅包含 3% 左右的重复双标记物, 其结果优于已发表的其他方法的结果 (Peters 等^[12] 8.6%; Velculescu 等^[19] 8.3%). 插入载体的 SAGE 标记物串联体约 650 bp, 含至少 30 个标签. 并已产生足够的克隆, 据估计每个 SAGE 库可产生 5 000 多个克隆. 因为 SAGE 串联体重质质粒转染的 *E. coli* 被平铺到 20 个含 Zeocine 的 LB 琼脂板, 每个约含 250 多个克隆, 因此, 我们能对每个库中的 150 000 个标签进行测序.

mini SAGE 中, 对标准 SAGE 法作了一些调整, 其中较值得一提的是, 在 SAGE 操作过程中

的 7 个苯酚提纯 DNA 步骤中用 PLG (Eppendorf) 纯化 DNA. 在苯酚抽提 DNA 后, 很难使位于上层水相的核酸从水相与有机相分界处的变性蛋白质中游离回收. 在进行苯酚提取 DNA 时, PLG 通过离心力作用位于有机物相与水相之间, 形成一条隔离带, 有机物相及分界处物质被有效地隔离于 PLG 下. 然后用吸管将上层含核酸的水相回收至一新试管中. 使用 PLG 能从水相、有机物相反应液中多回收 20%~30% 的核酸, 在 mini SAGE 的 7 步中加入 PLG 的方法明显增加了 DNA 的回收量和纯度. 其次, 与 microSAGE 法^[14] 相比较, 我们在 cDNA 连接反应中只加入了约 1/25 的 Linker (每次反应 10 ng 而代替 micro SAGE 法中的每次反应 250 ng). 如在 microSAGE 中所提到的, 2 200 个标签中有 109 个标签 (约 5%) 是 Linker 序列. Linker 双标签所占的百分比太高可能引起每个克隆测序所获得的 SAGE 标签的数量减少. Datson 等^[14] 提出, 含 Linker 的双标签被优先扩增从而干扰真正的 SAGE 双标签的扩增, 过多的 Linker 可能妨碍 SAGE 双标签连接为串联体, 减少 Linker 的数量可减少其对 SAGE 双标签扩增的干扰, 增加 SAGE 双标签的产量. 再次, 和 microSAGE 一样, miniSAGE 使用 mRNA 捕获试剂盒 (Boehringer Mannheim). 将 mRNA 的分离, cDNA 的合成, cDNA 的酶切与连接, cDNA 标记物从链霉抗生素珠上释放前 5 个步骤在一个试管中进行, 明显降低了这些连续过程中由于转管所导致材料丢失.

建立两个 SAGE 库 (表 1), 在这两个库中约 36% 的标签与 GenBank 中已知基因相对应, 约 16% 的标签与表达序列标记物 (EST) 相对应, 47% 的标签为未知序列, 代表新的基因, 这就意味着体内大多数基因仍未被识别. 细胞中基因表达水平从 1 拷贝到 86 拷贝不等, 而且大多数基因 (80%) 在每个细胞中呈低水平表达, 它们的 mRNA 仅占总 mRNA 量的一小部分 (表 2), 但却占 mRNA 种类的大部分, 与 Zhang 等^[19] 的发现一致.

在 WY 库中, 细胞中表达多于 10 个拷贝的前 30 个转录本列于表 3, 大多数都代表管家基因, 一些已知在成纤维细胞中表达的基因也被发现在本 SAGE 库中以多个拷贝表达.

膜蛋白、CD151 抗原和 Connexin 43 在 WY 中比 JT 中表达水平高, 研究发现 CD151 对于神经轴

突的数目、长度和生长率都有促进作用^[20]，它与 CAD 的关系尚不清楚，而 Connexin43 表达增强促进造骨细胞系的增殖和分化^[21]。分泌蛋白、hypotetical 蛋白 FLJ20657 和 fibronectin-1 在 WY 中也呈上调表达；在 WY 中与凋亡相关的蛋白质也以较高的水平，表达如酪氨酸-3-单氧化酶/色氨酸-5-单氧化酶、程序性细胞死亡因子-5 和 CGF-119 蛋白。脂类代谢的关键基因的 HMGCoA 还原酶、LIPA 和乙酰 CoA 羟化酶都呈下调表达，而 LDLR 基因呈上调表达，参与脂肪酸 β 氧化途径的基因也以较高水平表达^[22]。上述证据提示在 CAD 成纤维细胞中存在复杂的代谢失调。

在 WY 或 JT 库中，大多数基因在细胞中表达水平相当，在 WY 中有 0.16% 的基因表达水平高出 JT 库中约 5 倍多，而 WY 中有 0.10% 的基因表达水平比 JT 库低约 5 倍。在 WY 中表达增加的两种转录本 WY13（意指在 WY 中表达拷贝是 JT 中的 13 倍），WY12（意指在 WY 中表达拷贝是 JT 中的 12 倍），分别与血红素加氧酶和 HL14 对应。血红素加氧酶氧化血红素转变为胆红素、CO 和游离铁。在受到不同刺激如缺氧、氧中毒、重金属中毒等血红素加氧酶的活性上调。血红素加氧酶诱导升高，可能保护机体避免进一步损伤^[23]。HL14 基因编码的 β -半乳糖苷结合蛋白，近来认为是一种与鼠具同源性的细胞生长抑制因子^[24]。

在 WY 中表达降低的两种转录本，JT11 与 JT8（分别代表在 JT 中的表达拷贝是 WY 中的 11 倍与 8 倍的标记物），分别与已知基因 Calcycin 和 Seurchin homlog (hsn) 基因相对应，hsn mRNA 在所有已分析的人类组织中表达有差异。在生长活化的肾癌细胞系和活化的淋巴细胞中 hsn 呈高表达水平，表明了 hsn 在增殖中的作用^[25]。

总之，建立了一种 mini SAGE 方法，仅用 1 μ g 总 RNA 来构建 SAGE 库而不需要额外的 PCR 扩增，获得两个高质量的 SAGE 库。利用 mini SAGE 技术给冠心病 (CAD) 的病理和发病机理研究提供了大量表达异常基因的信息。

参 考 文 献

- Velculescu V E. Tantalizing transcriptomes—SAGE and its use in global gene expression analysis. *Science*, 1999, **286** (5444): 1491~1492
- Velculescu V E, Zhang L, Vogelstein B, *et al.* Serial analysis of gene expression. *Science*, 1995, **270** (5235): 484~487
- Schena M, Salon D, Heller R, *et al.* Parallel human genome analysis: microarray-based expression monitoring of 1000 genes. *Proc Natl Acad Sci USA*, 1996, **93** (20): 10614~10619
- Kwiterovich Jr P O, Motevalli M, Miller M. Acylation-stimulatory activity in hyperapob fibroblasts: Enhanced cholesterol esterification with a new serum basic protein. *Proc Natl Acad Sci USA*, 1990, **87** (22): 8980~8984
- Velculescu V E, Zhang L, Zhou W, *et al.* Serial Analysis of Gene Expression: Detailed Protocol (Version 1.0d), Nov. 22, 1999. Available from Johns Hopkins Oncology Center and Howard Hughes Medical Institute, 424 North Bond Street, Baltimore, MD 21231, USA, Fax: (410) 955~0548
- Powell J. Enhanced concatemer cloning—a modification to the SAGE (serial analysis of gene expression) technique. *Nucleic Acids Res*, 1998, **26** (14): 3445~3446
- Zhang L, Zhou W, Velculescu V E, *et al.* Gene expression profiles in normal and cancer cells. *Science*, 1997, **276** (5316): 1268~1272
- Hedrick S M, Cohen D I, Nielsen E A, *et al.* Isolation of cDNA clones encoding T cell-specific membrane-associated proteins. *Nature*, 1984, **308** (5955): 149~153
- Liang P, Pardee A B. Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science*, 1992, **257** (5072): 967~971
- Adams M D, Kerlavage A R, Fleischmann R D, *et al.* Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence. *Nature*, 1995, **377** (6547, Suppl): 173~174
- Lockhart D J, Dong H, Byrne M C, *et al.* Expression monitoring by hybridization to high density oligonucleotide arrays. *Nat Biotechnol*, 1996, **14** (13): 1675~1680
- Peters D G, Kassam A M, Yonas H, *et al.* Comprehensive transcript analysis in small quantities of mRNA by SAGE-Lite. *Nucleic Acids Res*, 1999, **27** (24): e39 (i~vi)
- Neilson L, Andalibi A, Kang D, *et al.* Molecular phenotype of the human oocyte by PCR-SAGE. *Genomics*, 2000, **63** (1): 13~24
- Datson N A, Jong J, van der Perk-de, *et al.* MicroSAGE: a modified procedure for serial analysis of gene expression in limited amounts of tissue. *Nucl Acids Res*, 1999, **27** (5): 1300~1307
- Virlon B, Cheval L, Buhler J M, *et al.* Serial microanalysis of renal transcriptomes. *Proc Natl Acad Sci USA*, 1999, **96** (26): 15286~15291
- Spinella D G, Bernardino A K, Redding A C, *et al.* Tandem arrayed ligation of expressed sequence tags (TALEST): a new method for generating global gene expression profiles. *Nucleic Acids Res*, 1999, **27** (18): e22
- Welle S, Bhatt K, Thornton C A. Inventory of high-abundance mRNAs in skeletal muscle of normal men. *Genome Res*, 1999, **9** (5): 506~513
- de Waard V, van den Berg B M, Veken J, *et al.* Serial analysis of gene expression to assess the endothelial cell response to an atherogenic stimulus. *Gene*, 1999, **226** (1): 1~8
- Zhang L, Zhou W, Velculescu V E, *et al.* Gene expression profiles in normal and cancer cells. *Science*, 1997, **276** (5316): 1268~1272
- Stipp C S, Hemler M E. Transmembrane-4 superfamily proteins CD151 and CD81 associate with alpha 3 beta 1 integrin, and selectively contribute to alpha 3 beta 1-dependent neurite outgrowth. *J Cell Sci*, 2000, **113** (Pt 11): 1871~1882
- Gramsch B, Gabriel H D, Wiemann M, *et al.* Enhancement of connexin 43 expression increases proliferation and differentiation of an osteoblast-like cell line. *Exp Cell Res*, 2001, **264** (2): 397~407

- 22 Bennett M J, Rinaldo P, Strauss A W. Inborn errors of mitochondrial fatty acid oxidation. *Crit Rev Clin Lab Sci*, 2000, **37** (1): 1~ 44
- 23 Amersi F, Buelow R, Kato H, *et al.* Upregulation of heme oxygenase 1 protects genetically fat Zucker rat livers from ischemia/reperfusion injury. *J Clin Invest*, 1999, **104** (11): 1631~ 1639
- 24 Goldstone S D, Lavin M F. Isolation of a cDNA clone, encoding a human beta-galactoside binding protein, overexpressed during glucocorticoid-induced cell death. *Biochem Biophys Res Commun*, 1991, **178** (2): 746~ 750
- 25 Duh F M, Latif F, Weng Y, *et al.* cDNA cloning and expression of the human homolog of the sea urchin fascin and *Drosophila* singed genes which encodes an actin-bundling protein. *DNA Cell Biol*, 1994, **13** (8): 821~ 827

Serial Analysis of Gene Expression From Minimal Amount of RNA Material^{*}

ZHENG Fang¹⁾, ZHOU Xin^{1)**}, YAN Ming²⁾, YE Shu-Qing³⁾, LIU Fang¹⁾

¹⁾ Gene Diagnosis Centre of Zhongnan Hospital, Wuhan University, Wuhan 430071, China;

²⁾ Department of Ophthalmology, Zhongnan Hospital, Wuhan University, Wuhan 430071, China;

³⁾ Johns Hopkins University School of Medicine, Baltimore, Maryland, 21218, USA)

Abstract A miniSAGE (minimal serial analysis of gene expression) technique was established for determining the gene expression pattern from minimal amount of starting material (1 μ g RNA). The gene profiles of two fibroblast cell lines were analyzed. One was from a coronary artery disease (CAD) patient, another one was from a normal person. Two high quality SAGE libraries were synthesized. Comparing the SAGE tags in the two libraries, abounding data about the differential expression genes was provided for the research on the mechanism of CAD. The new gene related with CAD could be isolated based on these data. The miniSAGE technique is a powerful gene expression analysis tool for study on complex disease, and can be applied on only 1 μ g RNA.

Key words serial analysis of gene expression, coronary artery disease, fibroblast cell

^{*} Part of this work was performed in Johns Hopkins University, Baltimore, Maryland, 21218, USA.

^{**} Corresponding author. Tel: 86-27-87317802, E-mail: zhouxjyk@sina.com

Received: October 12, 2001 Accepted: December 29, 2001