文章编号:1673-1689(2008)01-0122-05

表达基因分析方法

许杨^{1,2}, 阮琼芳^{1,2}, 李燕萍^{1,2}

(1. 南昌大学中德联合研究院, 江西南昌 330047; 2. 食品科学与技术国家重点实验室 南昌大学, 江西南昌 330047)

摘要: 了解不同细胞或同类细胞在不同发育阶段, 不同生理状态下的基因表达状况, 可以为研究生命活动过程提供重要的信息。表达基因分析方法不断完善, 包括抑制消减杂交技术, 基因表达序列分析, 基因表达谱芯片, 电子克隆等。作者主要介绍了这些技术的原理、方法、特点及其应用情况。

关键词: 表达基因; 抑制消减杂交技术; 基因表达序列分析; 基因表达谱芯片; 电子克隆
中图分类号: Q 756 **文献标识码:** A

Analysis Methods of Expression Genes

XU Yang^{1,2}, RUAN Qiong-fang^{1,2}, LI Yan-ping^{1,2}

(1. Sino-Germany Joint Research Institute of Nanchang University, Nanchang 330047, China; 2. State Key Laboratory of Food Science and Technology, Nanchang University, Nanchang 330047, China)

Abstract: Identifying changes in gene expression is of prime interest in molecular biology. A variety of methods have been developed for analyzing differentially expressed genes. There are several methods widely applied, including suppression subtractive hybridization (SSH), serial analysis of gene expression (SAGE), cDNA microarrays, silicon cloning and et al. This manuscript describe the principle, characteristic and application of these methods.

Key words: expression gene; SSH; SAGE; cDNA microarrays; silicon cloning

基因携带着生物体的全部遗传信息, 高等生物大约有 100 000 个不同的基因, 然而, 在任一细胞中表达的基因仅占总数的 15%, 即 15000。基因是按一定的时间和空间顺序进行表达的, 并决定生物的表现型特征。了解不同细胞或同类细胞在不同发育阶段、不同生理状态下的基因表达状况, 可以为研究生命活动过程提供重要的信息, 为医疗诊断、疾病研究及新药设计奠定基础。

最近几年发展起来的方法可以同时两个或多个实验样品之间对所有的差异表达的基因进行

分析, 包括抑制消减杂交技术、基因表达系列分析技术、表达谱基因芯片和电子克隆等。作者主要对这些技术的原理、操作方法、特点及其应用情况做一介绍。

1 抑制消减杂交技术

抑制消减杂交技术 (Suppression Subtractive Hybridization, SSH) 是利用抑制 PCR 和消减杂交的原理来寻找差异表达基因。首先提取受试 (test-

收稿日期: 2007-04-21.

基金项目: 国家自然科学基金项目 (30460006); 长江学者和创新团队发展计划项目 (IRT0540)

作者简介: 许杨 (1951-), 女, 安徽安庆人, 德国博士, 教授, 博士生导师。主要从事微生物及食品生物技术方面的研究。Email: xuyang1951@ yahoo. com. cn

er) 和驱动(driver)的 mRNA 并转录成 cDNA, 将 tester 的 cDNA 分成两份, 在 cDNA 5' 端分别接上接头 1 和接头 2, 再与过量的 driver cDNA (不接接头) 混合, 变性退火, 做第一轮消减杂交。由于 tester 和 driver 中共同表达的基因大多数被中和, 形成杂交体, 使得 tester 中差异表达的 cDNA 大量富集, 第一次杂交后各种单链分子在浓度上基本一致, 再将第一轮消减杂交的两份 DNA 样本混合, 加入过量的 cDNA, 变性退火, 做第二轮消减杂交, 进一步减少了 tester 中非差异表达的 cDNA, 即进一步富集了差异表达的基因。以接头 1 和接头 2 为引物, PCR 扩增将 cDNA 3' 端补平, 按照接头 1 和接头 2 的 5' 端外侧序列设计引物进行 PCR 扩增, 非目的片段因接头产生链内互补“锅柄样”结构而停止扩增, 从而实现选择性抑制非需要序列扩增, 从而有效地扩增低丰度 mRNA 和找出差异表达的基因。

SSH 技术采用两次杂交, 两次 PCR, 大大降低了假阳性率, 使高、中、低丰度基因均一化, 从而大大提高低丰度基因的检出率。由于 SSH 技术是利用 driver 和 tester 中共同表达的基因形成杂交体, 抑制相同表达基因的扩增来寻找差异表达的基因, 所以决定了 driver 和 tester 具有高度的相似性和良好的遗传背景, 最好是两者仅存在细微的差异。该方法从表型差异入手, 不仅能克隆分析细胞株之间、同种细胞不同条件下基因表达的差异, 也可研究不同组织间、同种组织不同发育阶段以及正常和病变组织之间的基因差异, 并通过这种差异表达的基因进一步研究与功能表型之间的关系。抑制性消减杂交法为寻找差异表达基因和研究已知基因的新生物学功能提供了一个有利工具, 被广泛应用于肿瘤、发育、组织特异性相关差异基因的研究^[2]。Roger^[3]等利用 SSH 技术研究了纤维状神经营养因子对鼠视网膜分化的影响, 获得 24 个差异表达基因。研究表明, 其功能与蛋白质修饰、信号转导和转录调控相关。朱龙付^[4]等利用 SSH 法对接种病原后的棉花 cDNA 进行差减杂交, 共获得了 534 个克隆。对 78 个呈上升表达的克隆进行 Blast 分析比对, 结果表明, 大部分阳性克隆与拟南芥等多个物种不同逆境条件下诱导的相关基因或表达序列相同或同源。但该技术需要以大量纯 mRNA (多达 20 μg) 为材料, 不适合用于活组织采样的样本及胚胎干细胞样本等的研究。

2 基因表达系列分析技术

基因表达系列分析技术(serial analysis of gene

expression, SAGE) 技术是利用 9 个核苷酸长度的序列标签代表特定的 mRNA 转录物, 来研究真核细胞中基因的表达谱信息, 并对细胞内所有表达基因进行定性与定量分析^[1]。首先抽提细胞的总 mRNA 并反转录成双链 cDNA, 用锚定酶酶切后分成两份, 分别同接头 A 和接头 B 连接(接头中包含了标定酶序列), 然后用标定酶酶切成平末端, 从而获得大小为 9 nt 的 DNA 片段。再将连有接头 A 和接头 B 的 DNA 片段混合, 以接头 A 和接头 B 的序列为引物进行 PCR 扩增, 然后用锚定酶酶切, 即可分离得到双标签序列, 将 30~50 份标签序列串联、克隆并测序, 最后利用 SAGE 软件包分析, 即可得知基因表达谱图。

SAGE 技术是一种能检测同一时间、同一细胞中成千上万个表达基因及其表达水平的实验方法, 不仅能分析已知基因组序列的物种, 还可将实验测序结果与 SAGE map 表达数据库中不同来源的已存在文库相比较, 可对大部分基因组序列未知的物种进行研究。构建 SAGE 库需要大约 500 μg 的 mRNA 或 5~50 μg 的总 RNA, 针对 mRNA 需要量大这一特点, MicroSAGE 和 MiniSAGE^[5-6] (MicroSAGE 使用量只需 SAGE 的 1/5000~1/500, MiniSAGE 仅需使用 1 μg 的总 RNA) 技术得以发展, 被广泛用于来源有限的组织, 如骨髓间质干细胞等的研究^[7]。经典的 SAGE 技术是用 9~14 个核苷酸长度的序列做标签, 对于基因复杂程度低的生物来说足以代表特定的 mRNA 转录物, 但对于人类基因组来说, 9~14 个核苷酸序列的标签特异性低^[8], 构建 LongSAGE 文库可以改善这个不足。研究证明, 随着标签长度的增大, 标签的特异性增高, 当长度增加到 21 nt 时, SAGE 标签在人类基因组中的特异性高达 99.83%, 要获得合适长度的标签, 选用适当的锚定酶和标签酶是关键。周村建^[9]等成功构建了白色念珠菌酵母细胞 LongSAGE 文库。从 1 000 个测序文件中, 共获得标签 17 773 个, 代表单一转录本 7 333 种, 其中单一匹配标签占 16.65%, 多重匹配标签占 6.59%, 推测蛋白质标签占 16.27%, 基因组及粘粒等标签占 1.64%, 新标签为 58.86%。标签拷贝数超过 35 的 33 个高表达标签中有 11 个可能属于新的表达序列。

该方法通过快速和详细分析成千上万个表达序列标签来寻找出表达丰度不同的 SAGE 标签序列, 从而全面提供生物体基因表达谱信息, 它还可用来定量比较不同状态下的组织或细胞的所有差异表达基因。它不但能分析转录子的丰富信息, 而

且还能发现新基因,是基因表达中定性和定量研究的一种有效工具,为了解正常基因作用机制,肿瘤发病机制、信号分子调控机制等提供信息。这使之成为从总体上研究基因表达、构建基因表达图谱的首选策略。

3 基因表达谱芯片

表达谱基因芯片(DNA microarrays for gene expression profiles)是基因芯片的一种,对不同细胞或组织来源的 mRNA 在逆转录反应中分别用不同颜色的荧光标记成探针(目前常用 Cy3-dUTP(绿色)标记对照组 mRNA, Cy5-dUTP(红色)标记样品组 mRNA),探针混合后与芯片或微阵列板上的基因进行严格杂交,再通过不同波长的荧光扫描芯片,将扫描所得每一点荧光信号值自动输入计算机并进行信息处理,给出每个点在不同波长下的荧光强度值及其比值(ratio 值),同时计算机还给出直观的显色图。在样品中呈高表达的基因其杂交点呈红色,相反,在对照组中高表达的基因其杂交点呈绿色,在两组中表达水平相当的显黄色,这些信号就代表了样品中基因的转录表达情况^[10]。

基因表达谱可直接检测 mRNA 的种类及丰度,可以同时分析上万个基因的表达变化,来揭示基因之间表达变化的相互关系。基因芯片是将许多的已知序列的 cDNA 或寡聚核苷酸固定在固定相上,最多可以达到 30 000 多个序列,预测人类基因组全序列含有 20 000~25 000 个基因,所以基因芯片一次反应几乎就能够分析整个人的基因^[11]。利用基因芯片可以分析多个样本,同时筛选不同样本(如肿瘤组织、癌前病变和正常组织)之间差异表达的基因,这样可以避免芯片间的变异造成的误差^[12]。张幸燕^[13]等将 512 个人癌基因和抑癌基因的 cDNA 用点样仪点在特制玻片上制成表达谱芯片,对正常人卵巢组织及卵巢癌组织基因表达的差异性进行比较研究,结果发现在卵巢癌组织中下调的基因有 23 个,上调的基因有 15 个,初步筛选出了卵巢癌相关基因。Lowe^[14]等利用胰腺癌、间充质细胞癌等组织的 cDNA 制备基因芯片,筛选到胰腺癌细胞中高表达的基因,为医疗诊断、病理研究及新药设计奠定基础。基因芯片因具有高效率,高通量、高精度以及能平行对照研究等特点,被迅速应用于动、植物和人类基因的研究领域,如病原微生物毒力相关基因的分析、肿瘤相关基因的发现和研发等。

该技术亦存在缺点,如芯片造价昂贵,制备芯

片需要大量的 RNA(每次需 50~200 mg total RNA),对于低拷贝的 mRNA 需要量更大。基因芯片只能分析已知基因的表达情况,无法获得未知基因的表达情况,对于未完成测序工作的生物样本不适用,基因表达谱芯片分析技术依赖于完善的基因数据信息。

4 电子克隆

电子克隆(silicon cloning)是基于生物信息学资源展开的,利用计算机技术,依托现有的网络资源(EST 数据库、核苷酸数据库、蛋白质数据库、基因组数据库等),采用生物信息学方法(包括同源性检索、聚类、序列拼装等),通过 EST 或基因组的序列组装和拼接,利用 RT-PCR 快速地获得全长 cDNA 序列的方法。根据 EST 相关注释,可知基因表达的时空信息,推断基因的功能^[15-16]。

电子克隆是近年来伴随着基因组计划和 EST 计划发展起来的基因克隆的新方法。与传统的基因克隆方法相比,该方法速度快,其中同源性比较、序列拼接组装等工作在计算机上完成,只需要 RT-PCR 序列验证即可;实验成本低;技术要求低,实验步骤简化,只需要掌握 RNA 抽提,反转录,PCR 扩增等分子生物学基本实验。电子克隆依赖 EST 数据库的发展,近年来,由于 DNA 自动测序技术的快速发展,EST 数据库的发展尤为迅速,且趋向于专业化,即建立特定物种和特定内容的数据库。但也因受基因组数据库和 EST 数据库等资源的限制,所以电子克隆更适用于已大量测序的生物(如小鼠、大鼠、水稻、拟南芥等)和人类表达基因的研究。王海涛^[17]等利用电子克隆技术成功克隆了人和大鼠自噬相关新基因 WIPI-3 的 cDNA。张会敏^[18]等利用电子克隆技术从稻瘟菌中克隆到一个新的 I 型烯醇化酶全长 cDNA,是第一个用电子克隆技术从稻瘟菌中克隆到的基因。

随着基因大量测序,阐明基因的结构和功能等问题摆在研究者面前,电子克隆充分利用相关生物信息快速经济地对基因进行克隆及功能分析,符合基因组时代的迅猛发展的要求。

5 结 语

抑制消减杂交技术、基因表达系列分析技术和基因表达谱芯片在新基因的发现方面都各有其独特的优势,均可寻找出差异表达序列,但也有其不足。第一,获得的差异表达序列大部分情况都不是完整的基因。随着基因组数据库和 EST 数据库的

发展,电子克隆的出现弥补了上述方法的不足,但该技术受数据库资源的限制,如仍不适合用电子克隆获 cDNA 全长的物种,还需结合传统方法(如 cDNA 文库的构建或快速末端扩增法)来获得全长基因序列,为进一步的基因功能研究(如基因敲除, siRNA 技术等)提供信息。第二,以 mRNA 为起始材料,且需要量大。MicroSAGE 和 MiniSAGE 的出现解决了传统 SAGE 中 mRNA 需要量大的问题。SSH 与 DNA microarrays 的结合使用^[19],可

对阳性差减产物进行快速、有效的筛选。

如前所述,这些方法各有优缺点,研究者应根据实验目的,充分了解各个技术参数和研究背景等相关信息,对研究方法做合理取舍。随着生物学的发展,研究方法也会不断的完善和推陈出新,更多功能基因将被发现,为研究生命活动过程提供重要的信息,为医疗诊断、疾病研究及新药设计寻找新的契机。

参考文献(References):

- [1] 赵寿元,乔手怡.现代遗传学[M].北京:高等教育出版社,2001.
- [2] 梁莉,丁彦青,石益民.消减抑制杂交及其在肿瘤研究中的应用[J].癌症,2003,22(9):997-1000.
LIANG Li, DING Yan-qiong, SHI Yi-min. Suppression subtractive hybridization and its application in study of tumors [J]. *Chinese Journal of Cancer*, 2003, 22 (9): 997-1000. (in Chinese)
- [3] Jérôme Roger, Olivier Goureau, José-Alain Sahel, et al. Use of suppression subtractive hybridization to identify genes regulated by ciliary neurotrophic factor in postnatal retinal explants[J]. *Molecular Vision*, 2007, 13(2): 206-219.
- [4] 朱龙付,涂礼莉,张献龙,等.黄萎病菌诱导的海岛棉抗病反应的 SSH 文库构建及分析[J].遗传学报,2005,32(5):528-532.
ZHU Long-fu, TU Li-li, ZHANG Xian-long, et al. Construction and analysis of SSH library of gossypium barbadense upon infection with *Verticillium dahliae*[J]. *Acta genetica sinica*, 2005, 32(5): 528-532. (in Chinese)
- [5] Seth Blackshaw, Winston P Kuo, Peter J Park et al. MicroSAGE is highly representative and reproducible but reveals major differences in gene expression among samples obtained from similar tissues[J]. *Genome Biology*, 2003, 4 (3): 10-15.
- [6] Datson NA, Van Der Perk-De Jong J, Van Den Berg MP, et al. MicroSAGE; a modified procedure for serial analysis of gene expression in limited amounts of tissue[J]. *Nucleic Acids Research*, 1999, 27(5): 1305-1307.
- [7] Tuteja R, Tuteja Narendra. Serial analysis of gene expression (SAGE): application in cancer research[J]. *Medical Science Monitor*, 2004, 10 (6): RA132-140.
- [8] Saurabh Saha, Andrew B. Sparks, Carlo Rago, et al. Using the transcriptome to annotate the genome[J]. *Nature Biotechnology*, 2002, 19: 508-512.
- [9] 周村建,郝飞,邓永键,等.白念珠菌酵母相细胞长片段基因表达系列分析文库的构建和初步分析[J].解放军医学杂志,2006,31(10):966-968.
ZHOU Cun-jian, HAO Fei, DENG Yong-jian, et al. Long serial analysis of gene expression in the yeast phase of *Candida albicans*[J]. *Journal of Chinese People's Liberation Army*, 2006, 31 (10): 966-968. (in Chinese)
- [10] Lodish H, Berk A, Zipursky SL, et al. *Molecular cell biology*[M]. New York: Media Connected, 1999.
- [11] Almut Schulze, Julian Downward. Navigating gene expression using microarrays-atechnology review[J]. *Nature Cell Biology*, 2001, (3): 190-195.
- [12] Sofia K Gruvberger-Saal, Heather E Cunliffe, Kristen M Carr, et al. Microarrays in breast cancer research and clinical practice-the future lies ahead[J]. *Endocrine-Related Cancer*, 2006, 13: 1017-1031.
- [13] 张辛燕,李小平,赖娟,等.cDNA 微矩阵筛选卵巢癌相关基因的研究[J].癌症,2003,22(9):943-947.
ZHANG Xing-yan, LI Xiao-ping, LAN Juan, et al. Screening for ovarian cancer-associated genes with cDNA microarrays [J]. *Chinese Journal of Cancer*, 2003, 22 (9): 943-947. (in Chinese)
- [14] Anson W. Lowe, Mari Olsen, Ying Hao, et al. Gene expression patterns in pancreatic tumors, cells and tissues[J]. <http://www.plosone.org>, 2007, 2 (3): e323.
- [15] 邵筱,吴忠道,刘翰腾,等.应用 EST 和电子克隆策略研究血吸虫表达基因谱[J].基础医学与临床,2005,25(7):602-

606.

SHAO Xiao, WU Zhong-dao, LIU Han-teng, et al. Application of EST and silico clone for the analysis of the gene expression profile of *Schistosoma japonicum*[J]. **Basic and Clinical Medicine**, 2005, 25 (7): 602-606. (in Chinese)

[16] Rex T Nelson, Randy Shoemaker. Identification and analysis of gene families from the duplicated genome of soybean using EST sequences[J]. **BMC Genomics**, 2006, 7: 204.

[17] 王海涛, 畅继武, 马小兵, 等. 人和大鼠自噬相关新基因 WIPI-3 的电子克隆和序列分析[J]. 现代生物医学进展, 2006, 6 (4): 16-21.

WANG Hai-tao, CHANG Ji-wu, MA Xiao-bing, et al. Identification and characterization of human and rat autophagy-related novel gene WIPI3 in silico[J]. **Progress in Modern Biomedicine**, 2006, 6 (4): 16-21. (in Chinese)

[18] 张会敏, 姜明国, 冯友军. 稻瘟菌 I 型烯醇化酶基因全长 cDNA 的电子克隆[J]. 生物信息学, 2006, 4(2): 57-61.

ZHANG Hui-min, JIANG Min-guo, FENG You-jun. In silico cloning of MgEno-1 cDNA from *Magnaporthe grisea*[J]. **China Journal of Bioinformatics**, 2006, 4 (2): 57-61. (in Chinese)

[19] Knut Muller, Hilde Heller, Walter Doerfler. Foreign DNA integration[J]. **J Biological Chemistry**, 2001, 276 (17): 14271-14278.

(责任编辑:李春丽)

《食品与生物技术学报》2008 年征订启事

《食品与生物技术学报》(双月刊)是教育部主管、江南大学主办的有关食品、生物工程及其相关研究的专业性学术期刊,为全国中文核心期刊、中国科技核心期刊、中国期刊方阵双效期刊,目前被美国化学文摘(CA)等国内外 10 余家著名检索系统收录。主要刊发食品科学与工程,食品营养学,粮食、油脂及植物蛋白工程,制糖工程,农产品及水产品加工与贮藏,微生物发酵,生物制药工程等专业最新科研成果(新理论、新方法、新技术)的学术论文、试验报告、反映学科前沿研究动态的高质量综述文章,同时兼发动物营养与饲料工程、环境生物技术等方面的最新研究成果。

本刊读者对象:相关领域的高等院校、科研院所、企事业单位的教学、科研等专业技术人员,专业管理人员以及有关院校师生。本刊热忱欢迎广大读者订阅。

《食品与生物技术学报》,A4(大 16K)开本,128 页,每册定价 15.00 元,全年 6 期;本刊邮发代号:28-79,全国各地邮局均可订阅。

《食品与生物技术学报》编辑部