

筛选差别表达基因的方法及其应用新进展*

孙敬^{①②} 陈宏^② 彭景樾^{①**}

(^①中国科学院动物研究所计划生育生殖生物学国家重点实验室 北京 100080;

^②西北农林科技大学动物科技学院 陕西杨凌 712100)

摘要: 筛选差别表达基因是进行生物个体生长、发育调控和病理等研究工作的关键步骤之一,对基因组学、蛋白质组学以及更深层次的生命科学研究都具有重要意义。目前,筛选差别表达基因常用的方法主要有 DDRT-PCR、SSH、SAGE 和基因芯片等方法。本文就这些方法最新的应用进展进行了综述,并重点介绍了近年问世的 Megaclone、Megasort 和 MPSS 等新方法。

关键词: 差别基因表达;大量平行测序技术

中图分类号: Q81 **文献标识码:** A **文章编号:** 0250-3263(2004)05-96-05

The Introduction and Progress of the Methods for Screening Differentially Expressed Genes

SUN Jing^{①②} CHEN Hong^② PENG Jing-Pian^①

(^① *State Key Lab of Reproductive Biology, Institute of Zoology, Chinese Academy of Sciences, Beijing 100080;*

^② *College of Animal Science and Technology, Northwest Science and Technology University of Agriculture and Forestry, Yangling Shanxi 712100, China*)

Abstract: Screening differentially expressed genes, which are significant in genomics, proteomics and further understanding of life, is one of the most important research fields for the biological study of individual growth, development and pathology. At present, several methods for screening differentially expressed genes are frequently adopted, including differential-display reverse transcription-PCR (DDRT-PCR), suppression subtractive hybridization (SSH), serial analysis of gene expression (SAGE) and DNA chip. This review describes latest progress in the application of these methods, and in the meanwhile puts an emphasis on the introduction of recently developed Megaclone, Megasort and massively parallel signature sequencing (MPSS).

Key words: Differentially expressed genes; MPSS

一个有生命的个体在其自身的生长、发育、疾病、衰老及死亡的过程中,要不断适应其周围环境的变化,这必然涉及到在不同的时间、空间以及受到不同的刺激时该个体基因的差别表达,分析基因的差别表达对于了解生命现象的分子机理具有十分重要的意义。目前,常用的筛选差别表达基因的方法主要有 DDRT-PCR、SSH、SAGE 和基因芯片等,这些方法目前在不断改进和完善,通过这些方法虽然不能得到全部差别表达基因,但已有许多重要的基因通过这些技术得到了分离和鉴定。现在,生物体基因组的研究已经进入了后基因组时代,与之相适应又诞生了 Megaclone(巨克隆)、

Megasort(巨筛选)和 MPSS(大量平行测序技术)等新技术,MPSS 结合了现代计算机技术和传统生物学技术,具有高通量、自动化、微型化、分析快速等优点,由于使用成本很高,目前应用受到限制,但相信不久将会成为筛

* 中国科学院知识创新工程重要方向项目(No. KSCX2-SW-201),国家自然科学基金资助项目(No. 30370165);

** 通讯作者, E-mail: pengjp@panda. ioz. ac. cn;

第一作者介绍 孙敬,男,24岁,硕士研究生;研究方向:分子免疫学。

收稿日期:2004-03-20,修回日期:2004-05-20

选差别表达基因不可缺少的技术。

1 抑制性消减杂交技术

抑制性消减杂交技术 (suppression subtractive hybridization, SSH) 是 1996 年由 Diatchenko^[1] 等人发明。该技术将驱动组和实验组进行两次消减杂交以及杂交后再经两次 PCR 扩增,得到的差别基因片段的特异性很高,与传统的差异显示反转录聚合酶链式反应 (differential-display reverse transcription-PCR, DDRT-PCR) 相比^[2],显著地降低了实验中的假阳性率,Stein^[3] 等的研究发现其假阳性率仅为 6%,因而应用广泛。

在发育生物学研究方面,Lindsay 等^[4] 以鼠为研究对象,利用 SSH 筛选出了与哺乳动物面部和其它生理结构发育形成有关的基因,其中 15% 可能为新基因;Tetsuya Adachi 等^[5] 在研究精子成熟过程中附睾经 DES 诱导后基因表达变化情况,利用 SSH 筛选出 15 个差别表达基因,在被消减的基因中有一个与 ADAM 7 同源的基因,该基因表达水平在处理显著下降;McClive 等^[6] 利用 SSH 技术研究了睾丸和卵巢差别表达的基因,进而揭示 SRY 在性别决定中的作用。在肿瘤研究方面,Jun Nishikawa 等^[7] 在研究与 EB 病毒有关的癌症中,以经 EB 病毒感染诱发而成的胃癌细胞和正常胃细胞为研究对象,利用该技术筛选二者差别表达基因,揭示了上调基因 hHb1-AN 的重要作用;Katherina 等^[8] 在多形性胶质母细胞瘤的研究中,利用 SSH 筛选脑瘤细胞和正常细胞间的差别表达基因发现,TC-22 基因的表达缺失在病发过程中起重要的作用;Yuan Chen 等^[9] 在肺癌研究中,利用 SSH 技术发现了一种新的细胞转型抑制因子——PDCD4。现在该技术主要应用于上述两方面的研究,由于 SSH 技术的优越性,这项技术在生命科学的其它领域也必将会得到更为广泛的应用。目前,SSH 技术的一个主要障碍是 mRNA 的需要量较大,常达几微克,达不到该要求,则很难检测到低丰度差别表达基因的 cDNA,此外,SSH 要求研究材料的差异不能太大。母胎免疫耐受一直是生殖免疫中一个重要的课题,作者利用 SSH 技术对妊娠过程不同时期子宫、滋养层细胞中 MHC I / II 类抗原及相关细胞因子时空差别表达基因的筛选和甄别,对这些差别表达基因功能的研究无疑将有重要的理论意义和应用价值。目前,我们采用 SSH 技术将大部分驱动组和实验组间共存的基因消减掉,并已构建了消减文库,同时对筛选得到的差别表达基因进行初步的功能研究(尚待发表)。

2 基因表达系列分析

基因表达系列分析 (serial analysis of gene expression,

SAGE) 技术是由 Velculescu 等^[10] 建立起来的一种新的基因表达模式研究技术。它是以待定细胞的总 mRNA 做模板,用 5' 端标记生物素的寡聚 dT 引物反转录合成双链 cDNA,用识别 4 碱基的锚定酶切割双链 cDNA,用链霉素和素磁珠回收双链 cDNA 的 3' 端(每个 3' 片段代表一个 cDNA),并分为两份,分别与两种不同的都带有 II 类型标签酶识别序列的接头 A 和 B 进行连接,再用标签酶消化得到单侧平末端 cDNA,通过补平末端、相互连接以及用对应接头 A 和 B 的引物进行 PCR 扩增,得到绝对数量大大提高的标签二聚体,然后进行第二次锚定酶酶切去除接头序列,并通过连接反应形成标签二聚体的多聚体用于克隆和测序。

Richard Kitching 等^[11] 在导致骨结节形成的形态变化的研究中,以类似成骨细胞并已带有克隆的 MC3T3-E1 细胞为研究材料,利用 SAGE 研究了该细胞在增殖、分化和基质矿化过程中的基因表达变化情况;Meissner 等^[12] 以牛的两种循环 γ 8T 淋巴细胞 [CD3.5 (+) CD8 (-) 和 CD3.5 (-) CD8 (+)] 为研究对象,利用该技术对比分析了两种细胞的基因表达情况,研究了 γ 8T 淋巴细胞亚型之间的差异;Chihiro Ito 等^[13] 利用该技术研究分别经甲基苯丙胺和苯环己哌啶处理的啮齿类动物脑皮层的基因表达变化情况,鉴定了 209 个与人类同源的基因,以期获得关于精神分裂症这种疾病的新认识。这些研究反应出 SAGE 技术的一个突出优点,即可在整体水平对细胞或组织中的大量转录本同时进行分析,从而全面反映细胞或组织中的基因表达情况,同时还可进行不同组织或细胞的所有差别表达基因的比较和研究。

3 基因芯片法

将寡聚核苷酸、基因组 DNA、cDNA 或肽核苷酸固定在硅片、玻璃片、塑料片、凝胶或尼龙膜等固相介质上,可制成 DNA 微阵列 (DNA microarray),这种技术又称为基因芯片技术。在用于分析差别表达基因时,将不同组织或细胞来源的 mRNA 制成带有标记的 cDNA 探针,分别与基因芯片杂交,再根据信号的分布可找到差别基因,并进行克隆和进一步的分析。

由于这种方法是将大量的 cDNA 同时固定在同一支持物上,因而可对某种组织或细胞内的大量基因进行平行检测,因此能较好地反映出该组织或细胞整体基因的表达情况,并能同时发现很多差别表达基因。该方法的灵敏性很高,便于重复,而且芯片业已商品化,分析技术自动化。但是,利用基因芯片进行研究的实验成本很高,一次实验至少需要上万元,这是更多的

实验室利用该技术的一个障碍。

目前该技术在很多研究领域里得到了应用:在酵母基因组功能研究中,Audrey 等^[14]利用基因芯片技术揭示了酵母为适应环境压力和饥饿时自身基因组遗传信息表达的重新编程的机制;在细胞凋亡的研究中, Tomoyuki Morita 等^[15]将鼠的肝脏分别切除 90% 和 95% 后,在不同的时间分别提取两种动物模型肝细胞的 RNA,制备探针,与芯片杂交,发现切除 95% 后的肝细胞中 p21 CDK 抑制子、Fas 和 IL-18 等基因的表达发生上调,而 Bcl-2, HSP 和谷胱甘肽-S-转移酶的基因的表达发生下调,正是这些基因表达的变化导致肝细胞不断凋亡以及再生功能的下降,最终引起鼠的死亡;在癌症研究方面, Stremmel 等^[16]报道在用 D 基因芯片技术分析正常与癌变结肠直肠组织时,找出的 6 000 个基因中的 100~500 个基因的表达水平发生改变,其中绝大多数基因与细胞生长和分化调控有关,他们还提出基因芯片技术的将来发展方向之一,是制作专门具有癌细胞基因表达模式特征和专门鉴定细胞类型这两类芯片,以方便于癌症的早期诊断与治疗;在病毒学方面, Byeong-Sun Choi 等^[17]利用乳突淋瘤病毒芯片分析了韩国从事商业性服务的妇女中该病毒的基因型分布情况,以期发现更为有效的对付该病毒的疫苗;在精神分裂症研究中, Carmen Chung 等^[18]利用基因芯片技术比较了患精神分裂症的海马组织与正常组织间的基因表达变化,发现了 4 个上调基因(chondrex, histamine-releasing factor, HERC2 和 heat-shock 70)。

4 Megaclone、Megasort 和 MPSS

2000 年,美国科学家 Brenner 等^[19]将由细胞中的 mRNA 反转录来的 cDNA 与带有约 1 670 万种不同 tag (tag 为一由 32 碱基排列组合成的 DNA 短链)的 DNA 载体库相连接,制备带有单链 tag 的 cDNA 文库。在此过程中,每种载体只带有一种 tag,且每个 tag 只与一种 cDNA 分子相对应;然后将带有单链 tag 的 cDNA 与固定有 antitag 的 microbeads 进行杂交,microbead 是一种利用环氧丙酰胺甲基丙烯酸盐制成的直径为 5 μm 的微粒体,每个 microbead 上固定有一种 antitag $10^4 \sim 10^5$ 个。由于 tag 与 antitag 二者序列互补,从而将 cDNA 固定到特定的 microbeads 上,并且一个 microbead 上只固定一种 cDNA 分子,这种新的构建 cDNA 文库的方法称为 Megaclone 技术。这项技术与传统的生物克隆技术相比,它的突出优点是对固定到 microbeads 上的 DNA 无需做任何进一步的加工处理就可直接进行生化分析,而且可利用根据特殊性质设计的探针对固定到 microbeads

上的所有克隆同时进行分析,因而该技术应用范围很广泛,例如,可根据克隆的序列信息、丰度和它们连接特殊基团的能力等来鉴定该克隆的特性,或对它们进行分离等。

值得注意的是,在设计 tag 库时须遵守以下 6 条原则^[19]:(1) tag 库必须有足够的多样性,以使每一种 cDNA 分子只与一种 tag 相对应;(2)在经酶切或其它处理后,tag 必须仍与 cDNA 分子相连接;(3)所有 tag 与 antitag 形成的双螺旋的熔点必须相同;(4)完全配对的 tag/antitag 双螺旋与任何仅有一个错配的 tag/antitag 双螺旋二者间的熔点之差对所有序列均应相同,且较大,以有利于正确配对的形成;(5)必须有一种对构建这种 tag 库实际且有效的途径;(6)应用简便。

利用 Megasort^[19] 技术可以筛选两种生物样品 (Sample A 和 Sample B) 之间有表达差别的基因。方法是将欲比较的两种样品 (Sample A 和 Sample B) 的 mRNA 利用上述的 Megaclone 技术并经变性处理分别构建单链 cDNA microbeads Library,用兰色 (fluorescein, FAM) 和红色 (Cy5) 两种荧光对这两种样品的 cDNA 分别进行标记,等量混合后与制备的 cDNA microbeads Library 进行竞争性杂交,最后利用荧光激活分选仪 (fluorescence-activated cell sorter, FACS) 将两种荧光强度有差别的 microbeads 单个分离出来,进行基因的克隆、鉴定等分析。

大量平行测序技术 (massively parallel signature sequencing, MPSS) 也由 Brenner 等人^[20] 于 2000 年发明。这项技术的系统主要包括:(1) Flow cell 板, microbeads 在 flow cell 中紧密排列形成 microbeads 阵列;(2) Reagent reservoirs、Reagent pump 和 valve block, 提供反应所需的各种试剂,每种试剂的选择、流速以及 Flow cell 中的温度均由计算机程序控制;(3) 荧光信号收集和处理系统,包括计算机,配有 75 W 氙气弧光灯的共聚焦荧光显微镜,滤光轮,4 百万象素的 CCD 照相机和赛贝克效应板等。

MPSS 的主要流程包括:(1) 将 cDNA 模板利用 Megaclone 技术制作 microbeads,并在 cDNA 的 5' 末端加上 *Dpn* II 酶切位点;(2) 用 *Dpn* II 酶切 microbeads 上的 cDNA 形成粘性末端,并与含有 II 型限制性内切酶 *Bbv* I 识别位点的起始接头相连接,然后将 microbeads 充填至 Flow cell 板上;(3) 用 *Bbv* I 酶切产生四碱基末端,并与含有 *Bbv* I 识别位点的编码接头 (encoded adaptor, EA) 杂交,同时每种接头 3' 端带有一个解码连接位点 (decoder binding site, DBS),可与带有荧光标记的解码探针 (decoder probe, DP) 杂交产生荧光信号,该信号可反映出接头 5' 端与模板 cDNA 杂交的不同位置上的碱基

信息,进而获得模板的序列信息;(4)加入可产生红黄蓝绿杂交信号的四套 DP 进行杂交,杂交后洗去上一轮杂交的探针,显微拍照记录荧光信号,并输入计算机存储;(5)用 *Bbv* I 进一步消化 cDNA 模板,产生紧接着的四碱基末端,重复(3)至(5)4~5次,分析得到的荧光照片即可鉴定 microbeads 上长度为 16~20 bp 的 cDNA 模板序列。

这项技术存在着工作量大、步骤繁琐等缺点,同时还需要较为昂贵的仪器及相应的软件协同操作,因而国内外与之相关的应用报道还较少。但这项技术每套系统每天的碱基分析通过量可与目前高通量商业测序仪相媲美,它的每套系统每天的标签分析通过量超过测序仪 10 倍之多^[20]。传统的测序仪处理几千个模板给出含有几百个碱基的序列,而 MPSS 系统可处理几百万个模板后给出仅含有几十个碱基的序列,即利用 MPSS 可产生大量短的标签序列。同时,MPSS 无需进行常规的基因分离提取和模板的处理等步骤,也不需复杂的机器人电子系统,就可在分子水平上通过为数不多的反应对几百万个模板进行同时分析,对基因表达的全面综合分析是最佳的选择。Stefan Hoth 等^[21]以转基因植物 *Arabidopsis thaliana* 树苗为研究对象,通过诱导异戊烯转移酶表达 24 h 后,利用 MPSS 分析受其调控的基因表达情况,鉴定出 823 个上调基因和 917 个下调基因,同时对对比分析分别经 6 h 和 24 h 诱导后的基因表达情况,发现不同的基因群表现出相类似的表达调控途径。MPSS 技术不仅可提供某一 cDNA 在体内特定发育阶段的拷贝数,同时可测定出相应 cDNA 16~20 bp 的序列,为在转录水平上进行基因表达分析提供了一种强有力的定性和定量手段。

5 小结

分子生物学、发育生物学、医学等学科迅猛发展,差别表达基因的筛选起着举足轻重的作用。SSH 技术与 DDRT-PCR 相比,因其假阳性率低且一次实验可筛选出大量的差别表达基因,应用十分广泛。在实际研究中,每位科研工作者可根据各自感兴趣的目标基因选择合适的筛选方法,如在研究个别的差别表达基因时,可选用 SSH 和基因芯片等技术,但是后者成本较高且在一些实验中其信号较难识别与鉴定,结果不能令人十分满意;如要了解基因组的整体表达时,则可选用 SAGE、基因芯片等技术,SAGE 已建立了数据库,研究人员上网即可将实验结果与数据库中已有信息进行比较分析,既方便又快捷。对于新近问世的 MPSS 技术由于使用成本很高,较多实验室还无法承受。伴随着生命

科学的发展,这些技术将会不断完善,新的更方便、更准确的基因筛选技术也将会不断涌现,对生命科学更深层次的研究都会起到巨大的推动作用。

参 考 文 献

- [1] Diatchenko L, Lau Y F, Chenchik A, et al. Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. *Proc Natl Acad Sci, USA*, 1996, **93**(12):6 025 ~ 6 030.
- [2] Liang P, Pardee A B. Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science*, 1992, **257**:967 ~ 971.
- [3] OD von Stein, Thies W G, Hofmann M. A high throughput screening for rarely transcribed differentially expressed genes. *Nucl Acids Res*, 1997, **25**:2 598 ~ 2 602.
- [4] Lindsay F Fowles, Jennifer S Bennetts, Jennifer L, et al. Genomic screen for genes involved in mammalian craniofacial development. *Genesis*, 2003, **35** (2):73 ~ 87.
- [5] Tetsuya Adachi, Yoshiharu Matsuno, Atsushi Sugimura, et al. ADAM 7 (a disintegrin and metalloprotease 7) mRNA is suppressed in mouse epididymis by neonatal exposure to Diethylstilbestrol. *Molecular Reproduction and Development*, 2003, **64**(4):414 ~ 421.
- [6] McClive P J, Hurley T M, Sarraj M A, et al. Subtractive hybridisation screen identifies sexually dimorphic gene expression in the embryonic mouse gonad. *Genesis*, 2003, **37** (2):84 ~ 90.
- [7] Jun Nishikawa, Csaba Kiss, Shousuke Imai, et al. Upregulation of the truncated basic hair keratin 1 (hHb1-AN) in carcinomacells by Epstein-Barr virus (EBV). *International Journal of Cancer*, 2003, **107**(4):597 ~ 602.
- [8] Katherina O Shostak, Vladimir V Dmitrenko, Oleg M, et al. Downregulation of putative tumor suppressor gene TSC-22 in human brain tumors. *Journal of Surgical Oncology*, 2003, **82** (1):57 ~ 64.
- [9] Yuan Chen, Thomas Knösel, Glen Kristiansen, et al. Loss of PDCD4 expression in human lung cancer correlates with tumour progression and prognosis. *The Journal of Pathology*, 2003, **200**(5):640 ~ 646.
- [10] Victor E Velculescu, Lin Zhang, Bert Vogelstein, et al. Serial analysis of gene expression. *Science*, 1995, **270**:484 ~ 487.
- [11] Richard Kitching, Shirley Qi, Vincent Li, et al. Coordinate gene expression patterns during osteoblast maturation and retinoic acid treatment of MC3T3-E1 cells. *Journal of Bone and Mineral Metabolism*, 2002, **20**(5):269 ~ 280.
- [12] Meissner N, Radke J, Hedges J F, et al. Serial analysis of gene

- expression in circulating gamma delta T cell subsets defines distinct immunoregulatory phenotypes and unexpected gene expression profiles. *Journal of Immunology*, 2003, **170** (1): 356 ~ 364.
- [13] Chihiro Ito, Yuta Ouchi. Toward schizophrenia genes: genetics and transcriptome. *Drug Dev Res*, 2003, **60**: 111 ~ 118.
- [14] Audrey P Gasch, Margaret Werner-Washburne. The genomics of yeast responses to environmental stress and starvation. *Functional & Integrative Genomics*, 2002, **2** (4 ~ 5): 181 ~ 192.
- [15] Tomoyuki Morita, Shinji Togo, Toru Kubota, *et al.* Mechanism of postoperative liver failure after excessive hepatectomy investigated using a cDNA microarray. *Journal of Hepato-Biliary-Pancreatic Surgery*, 2002, **9**(3): 352 ~ 359.
- [16] Stremmel C, Wein A, Hohenberger W, *et al.* DNA microarrays: a new diagnostic tool and its implications in colorectal cancer. *International Journal of Colorectal Disease*, 2002, **17**(3): 131 ~ 136.
- [17] Byeong-Sun Choi, Okjin Kim, Mi Sun Park, *et al.* Genital human papillomavirus genotyping by HPV oligonucleotide microarray in Korean commercial sex workers. *Journal of Medical Virology*, 2003, **71**(3): 440 ~ 445.
- [18] Carmen Chung, Teresa Talerico, Philip Seeman. Schizophrenia hippocampus has elevated expression of chondrex glycoprotein gene. *Synapse*, 2003, **50**(1): 29 ~ 34.
- [19] Sydney Brenner, Steven R Williams, Eric H Vermaas, *et al.* *In vitro* cloning of complex mixtures of DNA on microbeads: Physical separation of differentially expressed cDNAs. *Proc Natl Acad Sci, USA*, 2000, **97** (4): 1 665 ~ 1 770.
- [20] Sydney Brenner, Maria Johnson, John Bridgham, *et al.* Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nature Biotechnology*, 2000, **18**: 630 ~ 634.
- [21] Stefan Hoth, Yoshihisa Ikeda, Michele Morgante, *et al.* Monitoring genome-wide changes in gene expression in response to endogenous cytokinin reveals targets in *Arabidopsis thaliana*. *FEBS Letters*, 2003, **554** (3): 373 ~ 380.