

差异表达基因的高通量筛选方法

李斐雪 王雁玲*

(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)

摘要 cDNA 阵列、基因芯片技术、iAFLP(introduced amplified fragment length polymorphism)、SAGE(serial analysis of gene expression)技术、cDNA-AFLP、DDRT-PCR (differential display reverse-transcription PCR)、RC4D (RFLP-coupled differential display) 以及抑制消减杂交技术 (suppression subtractive hybridization, SSH) 都是能高效鉴别两个不同细胞群差异表达基因的方法。现对各种方法的原理进行简要介绍, 并对其优缺点进行对比, 并结合作者实验室的研究工作, 列举了 cDNA 阵列和抑制消减杂交技术在生殖生物学领域中的应用情况。

关键词 差异表达基因; 高通量筛选; SAGE; DDRT-PCR; SSH

随着人类基因组全长序列的获得, 基因组研究重点就从基因组学转移到蛋白质组学上来, 以确定每个基因的产物并作出基因表达图谱。在组织的发育与分化过程中, 基因的表达模式也在发生持续的变化, 这种基因表达的时空专一性决定了所有生命过程的有序性, 因此研究在正常发育和病理状况下基因表达的变化情况对诠释这些生命现象有重要意义, 于是差异表达基因筛选技术应运而生。

许多年以来, 分离未知基因的方法仅限于差异筛选 cDNA 文库, 直到 1992 年 Liang 等^[1]发明了一种检测基因转录模式的方法, 即 DDRT-PCR (differential display reverse-transcription PCR), 第一次实现了同时很灵敏地检测到真核细胞中大部分的转录体。后来又先后出现了 RC4D(RFLP-coupled differential display)^[2,3]、cDNA-AFLP^[4]、SAGE(serial analysis of gene expression)^[5]及 cDNA 微阵列 (microarray)即基因芯片技术^[6], 目前基因芯片技术已经得到很大发展, 可以在一张芯片上检测成千上万个基因乃至全部基因组的基因表达情况。以下对现有的差异表达基因显示技术进行简要的总结, 并比较它们的优缺点。

1 基因差异表达的高通量定量显示系统

高通量定量显示系统包括 cDNA 阵列杂交、寡核苷酸阵列杂交、SAGE 和 iAFLP(introduced amplified fragment length polymorphism)技术等, 可以在一次实验中同时确定上千个转录体的丰度。

1.1 cDNA 阵列技术

cDNA 阵列包括 cDNA 阵列膜(macroarray)和 cDNA 微阵列(microarray)两种。macroarray 技术是将变性的双链 DNA 片段以机械方式点到尼龙膜上, 现在一张 12 cm×18 cm 的膜上最多可以点 6000 个点, 将实验组和对照组的 mRNA 通过逆转录掺入 [³²P]-dATP, 分别与阵列膜杂交得到各自特异的杂交模式, 两者对照即得出基因的差异表达情况^[7]。Macroarray 在价格方面有优势, 因为一张尼龙膜可以被重复杂交多达十次, 用少量的膜就可以检测出几种组织或几种不同生理、病理状况下的组织中数千个基因的表达情况。

cDNA 微阵列技术首先出现于 1995 年, 研究者用此技术检测了拟南芥根与叶组织中基因的表达丰度, 发现有 26 个基因在两种组织中的表达量相差 5 倍以上^[6]。cDNA 微阵列技术是将 cDNA 文库中的单个克隆进行 PCR 扩增, 变性获得单链的 DNA 后, 经自动化点样制备阵列, 包括光控固相化学合成与激光共聚焦等在内的多项技术被用于了基因芯片技术, 实验实现了全部自动化, 阵列上探针的密度通常在 30 000/cm²。相比较 macroarray 技术, cDNA 微阵列技术可以在一张芯片上同时分析两种不同的探针, 而且灵敏度达到 macroarray 的 50 倍以上^[8]。但其不足之处在于所点样的序列并不是都是实验需

收稿日期: 2003-09-30 接受日期: 2004-03-26

* 通讯作者。Tel: 010-62631832, Fax: 010-62565689, E-mail:

wangyl@panda.ioz.ac.cn

要检测的；此外所需的分析仪器比较复杂。本实验室将 macroarray 和 microarray 技术应用于筛选小鼠子宫动情周期不同阶段和着床起始相关的差异基因，发现它们与细胞增殖和凋亡、免疫调节、精子获能调控以及子宫内膜接受性建立等密切相关，由此更加全面地揭示了子宫多功能性的分子基础。

cDNA 微阵列技术在分析低丰度转录体方面比较有限，要确保某种低丰度转录体包含于 cDNA 微阵列上，需从相应的 cDNA 文库中挑选非常大量的克隆进行扩增点样；例如若要检测到丰度低于 0.001% 的 mRNA 分子，需制备的 cDNA 微阵列至少是标准大小的 20 倍以上。

1.2 寡核苷酸微阵列

寡核苷酸微阵列技术首先在 1980 年被提出来，其原理是通过与足够多的寡核苷酸序列杂交来确定 DNA 的序列，故最初的目的是建立一种经济快速的测序方法^[9]，后来被主要应用于分析不同细胞间基因的差异表达。寡核苷酸微阵列通过直接在支持物表面合成寡核苷酸片段来制备，密度也能达到 30 000/cm²^[10]。与 cDNA 阵列技术相比，寡核苷酸微阵列的优势是可以避免制备并保存许多不同种类的 DNA 群。但要通过寡核苷酸微阵列分析一个基因组的全部基因表达信息，阵列上必须包含该基因组所有可能的寡核苷酸序列，而目前的基因芯片技术还不能使芯片达此种密度，故而要求首先必须知道该基因组的所有序列，这也就限制了此项技术的使用范围。Bonner 等^[11]用寡核苷酸微阵列技术分析了小鼠肺的发育过程中基因表达变化情况，在 12 000 个寡核苷酸序列中发现有 1346 个在各个不同发育阶段的表达量有明显差异，为以后进一步研究差异表达基因的功能打好了基础。

对于一个转录体而言，其浓度与标记荧光的强弱正相关，在上述的几种阵列分析中，各种转录体的浓度便是根据它们与芯片上互补片段杂交后信号的荧光强度来确定的；而由于作为探针的各种转录体的二级结构及退火温度有差异，加之各自不同的复性动力常数，杂交过程中所有转录体的杂交条件都处于最佳是不可能的，因而不可避免会造成一定的实验误差。

1.3 SAGE

SAGE 与微阵列技术一样可以对低丰度转录体进行高通量定量测定和差异分析^[5]。它的工作原理可用图 1 显示，首先通过一系列的限制性内切酶作

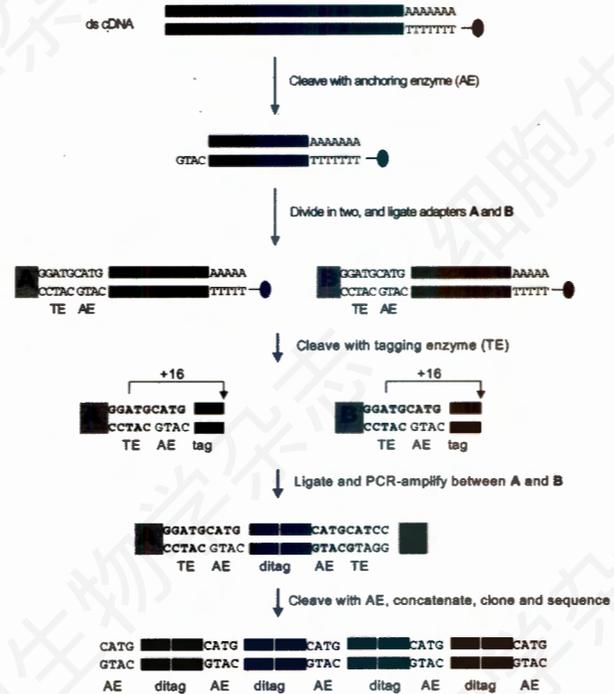


Fig.1 Serial Analysis of Gene Expression (SAGE)

用使每种 cDNA 都产生一 9 bp 的特异寡核苷酸片段，再通过连接、酶切和 PCR 扩增等一系列过程产生这一片段的多联体，将此多联体连接到特异载体，产生一 SAGE 文库。由于这一片段在文库的出现频率与相对应的 mRNA 在组织中的丰度正相关，可对相应的转录体进行定量分析。

一个 SAGE 克隆可以含有 20~50 个不同的寡核苷酸片段，所以相比较微阵列的 cDNA 文库，SAGE 文库要小的多。此外，由于所选用的内切酶作用比较广泛，几乎所有的转录体都有该酶的切点，可以产生 SAGE 所必须的 9 bp 特异片段^[12]。如果按每种 cDNA 产生 10 个片段计算，则需要对 150 000 个片段进行测序，如果每个克隆平均含 20 个标记片段，则需要测序分析 7500 个克隆；如此大量繁琐的测序工作加之本身技术的难度和固有的缺陷使这项技术的使用频率要比 DNA 微阵列少很多；然而近些年，人们已经通过简化步骤、提高 PCR 扩增效率等对该方法的修改使这些问题得到很大解决^[13,14]。Berthier 等^[15]用此方法检测了牲畜在感染锥虫病前后基因的表达变化情况，发现了 187 个差异表达的基因，为阐明牲畜对锥虫病有抗性的内在机制提供了新思路。

1.4 iaFLP

iaFLP 也是一种高通量的基因转录模式分析技术，不过它只能用于分析不同组织或细胞间某已知

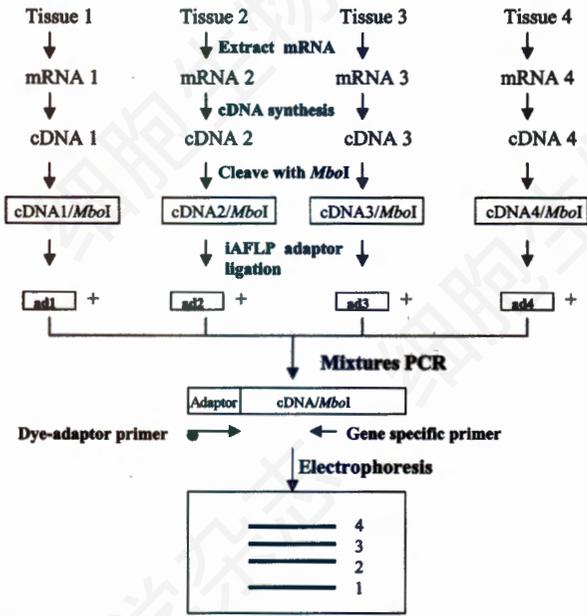


Fig.2 Introduced amplified fragment length polymorphism (iAFLP)

基因的转录体丰度^[16]。它将不同组织来源的 cDNA 群都用限制性内切酶 *Mbo* I 切开，然后分别与不同长度的接头连接，再把连接不同接头的 cDNA 混合在一起，用基因特异的引物与荧光标记的接头引物扩增基因的 3' 端序列，这样由于接头长度不同使同一转录体在不同 cDNA 群中的扩增片段长度不同，聚丙烯酰胺凝胶电泳后，通过与内置的标准比较即可得出转录体丰度(图 2)。这项技术的缺点在于它只能用于已知基因转录体的定量分析，但该技术的灵敏性和准确性特别高，而且具有很好的可重复性。

2 基因差异表达的高通量定性显示系统

定性显示系统包括 cDNA-AFLP、DDRT-PCR、RC4D 和差异筛选技术。它们都只能定性或半定量分析基因的转录模式而不能进行精确的定量分析。

2.1 cDNA-AFLP

cDNA-AFLP 技术由 AFLP 发展而来，经典的 cDNA-AFLP 步骤有 3 步：①限制性内切酶切开 cDNA，连接上寡核苷酸接头；②用 PCR 引物选择性扩增限制性内切酶酶切片段；③电泳分析扩增的片段，电泳条带的强弱与 PCR 模板的量呈正相关，由此可以估计转录体的浓度^[17]。此方法由于采用 PCR 扩增，使得需要的初始 RNA 量较少。但该技术最大的限制就是要求被分析的 cDNA 分子上有合

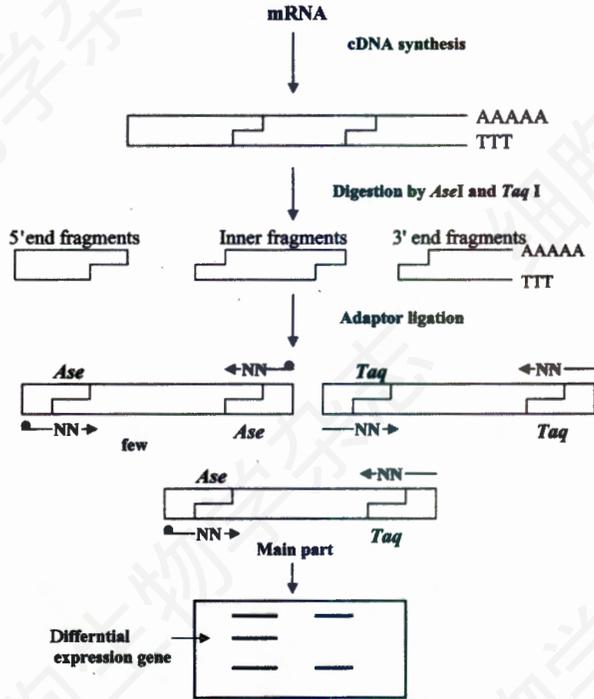


Fig.3 cDNA-AFLP

适的限制性内切酶位点，所以往往需要选用多种不同的酶以确保每一种 cDNA 分子都能被检测到。在 PCR 扩增过程中会产生三种产物，因为在基因序列中可能会存在一种酶的两个切点，使酶切片段的两端连接同样的引物，但这种几率很低(图 3)。

2.2 DDRT-PCR

DDRT-PCR 技术最早于 1992 年出现^[1]，可以用于分离在不同的真核细胞中差异表达的 cDNA 并加以克隆。其原理是将两种细胞的 mRNA 逆转录后进行 PCR 扩增。PCR 3' 端引物序列是针对 mRNA 的 poly(A) 尾设计的，一般是 11 个 T 再加上两个碱基，这样 12 种 3' 端引物($T_{11}AA$, $T_{11}AC$, $T_{11}AG$, $T_{11}AT$, $T_{11}CA$, $T_{11}GA$, $T_{11}CC$, $T_{11}CG$, $T_{11}CT$, $T_{11}GC$, $T_{11}GG$, $T_{11}GT$)就可以与所有 mRNA 的 poly(A) 尾匹配；5' 端引物是随机引物，一般为 10 个碱基，因此产生一些不同长度的 cDNA 片段，电泳后比较两者的差别而得到差异表达基因的 cDNA。

这个方法存在许多严重的缺陷，它的 5' 端随机引物一般常有 2~3 个碱基不能与 cDNA 模板完全匹配，而且 PCR 反应中随机性、偶然性比较大，容易造成非特异性扩增而造成高的假阳性率^[18]，这就使下游的筛选工作很巨大^[19]。理论上此方法可以检测到 95% 以上的转录体，但由于引物序列的随机性

和竞争性模板结合位点的存在, 很难确定实际的效率^[20]。此外, 由于 PCR 扩增从 mRNA 的 3' 端开始, 所以一些扩增子(amplicon)不含有义序列, 而只含有 3' 端的非翻译区, 从而在基因文库中找不到同源序列。最后, 扩增信号的强弱与开始时 mRNA 的浓度不存在精确的正比关系。尽管有上述缺陷, 由于其实验步骤较简单, 此方法在实际工作中应用仍较多, 例如用于筛选在肿瘤发生、愈伤过程以及胚胎植入过程中起重要作用的基因^[21,22]等。

2.3 RC4D

RC4D 由 cDNA-AFLP 与 DDRT-PCR 结合发展而来, 与其他技术不同, 它只能用于分析属于同一个家族的基因, 并且需要已知一个基因家族特异的核苷酸序列(family specific domain, FSD)。鉴于同一基因家族的基因通常含有保守的区域即 FSD, RC4D 技术就根据这一点设计长的 FSD 特异引物代替 DDRT-PCR 中的随机引物, 使其特异扩增同一基因家族的基因^[23]; 由于扩增产物的长度大略相同, 所以就引入了 AFLP 方法, 通过限制性内切酶切割后, 扩增子的长度由原来的 1 kb 变为几百个碱基不等, 这样同一家族的不同基因就可通过聚丙烯酰胺凝胶电泳很好地区分开来。此技术中由于采用长引物与严格退火条件使之比 DDRT-PCR 有更好的可重复性, 并且可以半定量分析转录体的浓度。RC4D 技术的主要缺陷是: 被分析基因家族的成员在 3' 端的上游要存在一保守区, 同时在保守区内不能含有限制性内切酶位点, 否则会因丢失引物的结合区而不被检测到。

2.4 差异筛选技术

最早用于分离差异表达 cDNA 分子的方法是差异筛选质粒或噬菌体 cDNA 文库。理论上它可以用来筛选任何差异表达的基因, 而实际运用中, 只有当被筛选基因的 mRNA 在一种细胞总 mRNA 中的含量 $\geq 0.05\%$, 而在另一种细胞中的含量 $\leq 0.01\%$ 时, 对这两种细胞进行基因表达的差异筛选才最有效^[24]。

随后又发展起来消减杂交(subtractive hybridization)技术, 让来自一种细胞的 cDNA, 与来自另一种细胞的大约 20 倍量的 mRNA 彻底杂交, 这样在第一种细胞中表达而在第二种细胞中不表达的基因就不会形成 DNA-RNA 杂交体而保持为单链 DNA, 可以提高低丰度 mRNA 的检测效率, 被检测 mRNA 的丰度范围为 1/500 到 1/20 000。但此方

法有一些局限性, 首先由于需要构建文库和进行大量的克隆筛选, 所以此方法通常被认为是效率低而且工作量极大的工作, 尤其当所寻找基因的表达丰度极低, 这可能是不可避免的; 但是如果所寻找的基因在实验组表达很高, 而在对照组表达非常低, 或反之, 则需要筛选的克隆数不用很多。其次, 消减杂交需要的 mRNA 量往往较高, 而且会出现一定的假阳性。此外, 一次消减杂交只能向一个方向消减, 得到实验组中高表达的差异基因, 如要想确定实验组中低表达或对照组中高表达的基因, 就需要进行反方向消减实验。

2.5 抑制消减杂交

抑制消减杂交(suppression subtractive hybridization, SSH)它最早出现于 1996 年, 是消减杂交进一步改进后建立的一种简便而高效的差异表达基因筛选方法。该方法是以抑制 PCR 为基础的 cDNA 消减杂交, 其原理见图 4, 抑制 PCR 就是利用非目标序列片段两端的长反向重复序列在退火时

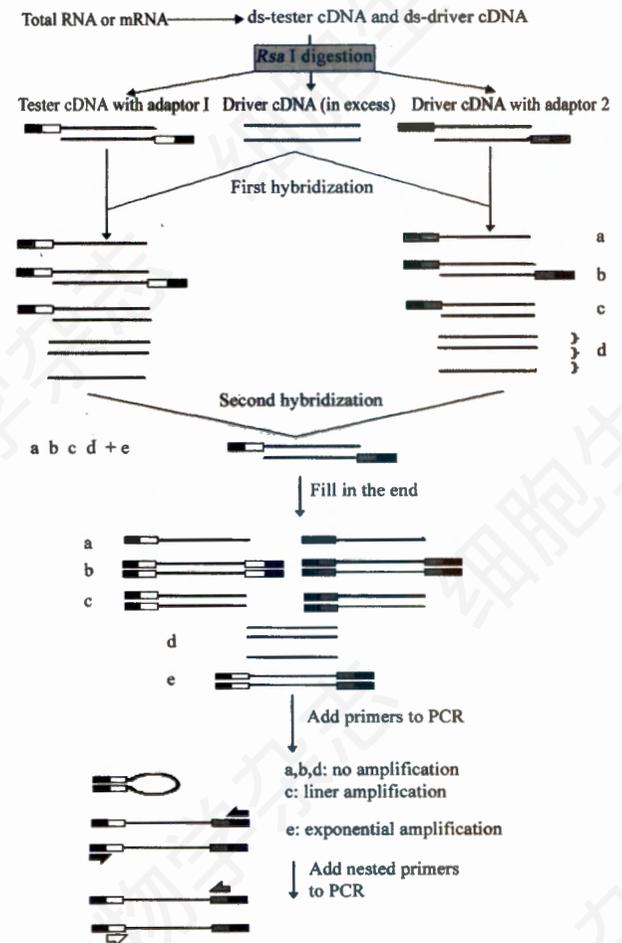


Fig.4 Suppression subtractive hybridization (SSH)

产生“锅柄”结构,无法与引物配对,从而选择性地抑制非目标序列的扩增。同时由于 SSH 采用两步杂交和两步 PCR,保证了该方法具有较高的特异性,大大降低假阳性率。而且此方法的均等化和目标片段的富集,保证了低丰度 mRNA 也有被检出的可能,从而具有比 DDRT-PCR 方法更高的敏感性。

SSH 方法需要的 mRNA 量较多,通常为几微克,若是 mRNA 量不够,那么低丰度的差异表达基因很可能会检测不到。目前一个有效的解决办法是将高保真全长 cDNA 合成技术与 SSH 相结合,使一些量少而且很珍贵的实验材料也可以用 SSH 检测。本实验室近年在这方面开展了一些有意义的工作,我们将 Clontech 公司的 SMART™ PCR cDNA Synthesis 试剂盒与 PCR-Select™ cDNA Subtraction 试剂盒联合使用,筛选着床起始期恒河猴着床点与非着床点子宫内膜差异表达的基因。SMART cDNA 合成技术可以确保从 0.05~1 μg 的总 RNA 也可以得到扩增量的高质量全长双链 cDNA,而且很好地保持了原初 mRNA 的相对丰度,使差异筛选所需的 RNA 量可以降低很多。在恒河猴着床位点的实验材料非常稀少而难得的情况下,我们成功筛选到一些与胚胎植入启动相关的恒河猴新基因^[25]。

参考文献 (References)

- [1] Liang P *et al.* *Science*, 1992, **257**: 967
- [2] Fischer A *et al.* *Proc Nat Acad Sci USA*, 1995, **92**: 5331
- [3] Vos P *et al.* *Nucleic Acids Research*, 1995, **23**:4407
- [4] Bachem CW *et al.* *Plant J*, 1996, **9**: 745
- [5] Velculescu VE *et al.* *Science*, 1995, **270**: 484
- [6] Schena M *et al.* *Science*, 1995, **270**: 467
- [7] Chen JJ *et al.* *Genomics*, 1998, **51**: 313
- [8] Kehoe DM *et al.* *Trends Plant Sci*, 1999, **4**: 38
- [9] Chetverin AB *et al.* *Biotechnology*, 1994, **12**: 1093
- [10] Yershov G *et al.* *Proc Nat Acad Sci USA*, 1996, **93**:4913
- [11] Bonner AE *et al.* *J Med Genet*, 2003,**40**: 408
- [12] Matsumura H *et al.* *Plant J*, 1999, **20**: 719
- [13] Kannbley U *et al.* *Biotechniques*, 2003, **34**: 1212
- [14] Du Z *et al.* *Biotechniques*, 2003, **35**: 66
- [15] Berthier D *et al.* *Genet Sel Evol*, 2003, **35**: S35
- [16] Kawamoto S *et al.* *Genome Research*, 1999, **9**: 1305
- [17] Bachem CWB *et al.* *Plant Mol Biol Rep*, 1998, **16**: 157
- [18] Malhotra K *et al.* *Nucleic Acids Research*, 1998, **26**: 854
- [19] Poirier GM *et al.* *Nucleic Acids Research*, 1997, **25**: 913
- [20] Wan JS *et al.* *Nat Biotechnol*, 1996, **14**: 1685
- [21] Welss T *et al.* *Int J Cancer*, 2003, **104**: 66
- [22] Soo C *et al.* *Plast Reconstr Surg*, 2002, **110**: 787
- [23] Theissen G *et al.* *Methods Mol Biol*, 1997, **85**: 123
- [24] Sambrook J *et al.* *Molecular Cloning: A Laboratory Manual*, 2nd ed., New York: Cold Spring Harbor Laboratory Press, 1989
- [25] 孙晓阳等. *生物化学与生物物理进展*, 2003, **30**: 889

High Throughput Screening Methods Used to Identify Differentially Regulated Genes

Fei-Xue Li, Yan-Ling Wang*

(State Key Laboratory of Reproductive Biology, Institute of Zoology, Chinese Academy of Sciences, Beijing 100080, China)

Abstract cDNA array, gene chip, iAFLP (introduced AFLP), SAGE (serial analysis of gene expression), cDNA-AFLP, DDRT-PCR, RC4D (RFLP-coupled differential display) and SSH (suppression subtractive hybridization) are techniques being used extensively to identify differentially regulated genes in organisms. In this review, these display systems are evaluated and compared. The general principles on which the different techniques are based and which determine their potential and their limitations are described. Some examples are given about the utilization of these methods.

Key words differentially expressed gene; high throughput Screening; SAGE; DDRT-PCR; SSH

Received: September 30, 2003 Accepted: March 26, 2004

*Corresponding author. Tel: 86-10-62631832, Fax: 86-10-62565689, E-mail: wangyl@panda.ioz.ac.cn