新技术 新方法

大规模平行测序技术（MPSS）研究进展

陈 杰*

（第三军医大学新桥医院肿瘤科，重庆 400037）

摘要 大规模平行测序技术（massively parallel signature sequencing，MPSS）是以 DNA 测序为基础的大规模高通量基因分析新技术，通过标签库的建立、微珠与标签的连接、酶切连接反应和生物信息分析等步骤，获得基因表达序列. MPSS 具有能测定表达水平较低、差异较小的基因，不必预先知道基因的序列，自动化和高通量等特点，是值得推广的技术.

关键词 基因，表达分析，大规模平行测序

学科分类号 Q343.1，Q78

大规模平行测序技术（massively parallel signature sequencing，MPSS）是 Brenner 等[1] 于 2000 年建立，由美国 Lynex 公司（www.lynex.com）将其商品化的一种基因克隆新技术. 在发明之初就与 TAKARA 公司联手向全球推广. 其核心技术分别由 MegaClone、MPSS 和生物信息分析三部分组成. 是基于序列分析技术的高通量、高特异性和高敏感性的基因分析技术. 本文就最新建立的大规模平行测序技术做简要介绍，并比较该技术与其他几种常用技术的优缺点.

1 MPSS 概况

基因药物及相关诊断试剂开发的首要条件是要找出导致疾病或与疾病相关的基因. 人类基因组共有 1 万至 5 万个基因，而在人体的任一器官、任一时刻，大约有 1 万至 1.5 万个基因在同时起作用. 找出与疾病直接相关的基因是一个复杂的过程，涉及到基因表达水平的测定、蛋白质表达、抗体产生和测试，以及复杂的实验设计和数理统计. 而其中最重要的一个环节是准确有效地检测基因在不同样品中表达水平的差异.

MPSS 是以基因测序为基础的新技术，其方法学基础是一个标签序列（10～20 bp）含有能够特异识别转录子的信息，标签序列与长的连续分子连接在一起，便于克隆和序列分析. 通过定量测定可以提供相应转录子的表达水平，也就是将 mRNA 的一端测出一个包含 10 至 20 个碱基的标签序列，每一标签序列在样品中的频率（拷贝数）就代表了与该标签序列相应的基因表达水平，所测定的基因表达水平是以计算 mRNA 拷贝数为基础，是一个数字表达系统，只要将病理和对照样品分别进行测定，即可进行严格的统计检验，能测定表达水平较低、差异较小的基因，而且不必预先知道基因的序列，该技术的特点是基因表达水平分析的自动化和高通量.

大规模平行测序技术，其基本方法是从生物样品中提取 mRNA，将 mRNA 分子转换成 cDNA，通过固相克隆将该 cDNA 均匀地加载到特制的小分子载体表面，然后在小分子载体上进行大量的 PCR 扩增. 将所有 cDNA 游离的一端进行精确测序产生 16 至 20 个碱基. 每一特定序列在整个生物样品中所占的比例，就代表了含有该 cDNA 基因在样品中的相对表达水平. 该技术能将一个生物样品中几乎所有表达了的基因全部分别克隆到特制的小分子载体上，然后把几十或上百万个小分子载体放进一个特殊的反应系统内，使所有小分子载体都排列在一个平面上，然后将带特殊荧光标记的 G、A、T、C 单核苷酸按顺序分别加入反应体系中，分别与小分子载体上的 cDNA 进行分子杂交，每次分子杂交后将所有小分子载体进行激光扫描照相. 当加入 G 时，有特殊荧光的小分子载体上所载的 cDNA 在这

* 通讯联系人.

Tel：023-68755646，E-mail：jazz0331@sina.com

收稿日期：2004-02-10，接受日期：2004-03-28

个碱基位置上就是G，当加入A时有荧光，则这个位置就是A，以此类推，只需经过4次反应4次激光扫个碱描照相就可将上百万个cDNA同时将这一位置的序列测出[2].

该技术的特点是：a. 不必事先知道基因的序列，适用于任何生物体及任何性状；b. 基因组覆盖面高，能测量出样品中几乎所有表达了的基因；c. 基因表达水平的测量是通过直接计算样品中cDNA的拷贝数目，属于非连续变量，所以只要有病理和正常个体（或组织）两个样品即可以进行严格的统计检验，能有效地检测差异性中等或较小的基因；d. 实验效率高，只要两个星期即可获得几十万个克隆的16至20个碱基序列.

该技术的关键是验证数据问题，即如何确定转录子和基因表达水平与标签序列产生的数据之间的关系. 对不同的基因使用正确的标签序列，如果基因与标签序列之间是非特异性和不明确的都将会产生分析错误.

2 MPSS基本步骤

首先用生物素标记的寡核苷酸引物（biotin-labelled oligo-dT primer）将来自细胞或组织的mRNA合成为cDNA双链（图1）[3].

Dpn Ⅱ限制性内切酶（酶切位点为GATC）消化cDNA片段，利用标记的生物素纯化消化的cDNA片段.

将纯化的cDNA片段克隆入包含有32 bp序列的标签（tag）载体中，并通过标签上的PCR引物扩增插入片段. 酶切消化线性化PCR产物，生成含cDNA片段与32 bp标签相连接的产物. 将cDNA模板连接到直径为5μm的微球体上. “克隆”的方法是利用人工设计长度不同的两类互补寡核苷酸（tag和anti-tag），分别将cDNA与tag连接，anti-tag和微球体连接之后，再将cDNA模板通过tag和anti-tag杂交连接与微球体连接起来. 为了能装载下细胞内所有的cDNA模板（若以3^{10}～4^{10}个基因计算），寡核苷酸的数量至少应该要比模板的量多100倍以上，为此Brenner等设计了1.67×10^7个长32 bp的寡核苷酸片段，这样可以保证生物体所有不同的cDNA模板都能与不同的寡核苷酸相连接，而且每一个微球体上也可承载10^4～10^5个相同的cDNA拷贝[3].

与32 bp标签序列互补的序列（anti-tag）杂交连接，而anti-tag预先已经通过共价键与直径为5 μm的微球体连接，这样含cDNA片段与32 bp标签相连接的序列就与微球体相结合.

cDNA序列测定，通过连接接头和ⅡS型限制性酶*Bbv* Ⅰ，进一步消化结合在微球体上cDNA模板，*Bbv* Ⅰ能在距识别位点9个碱基和13个碱基的位置切割cDNA双链，并在cDNA模板上产生4个碱基末端.

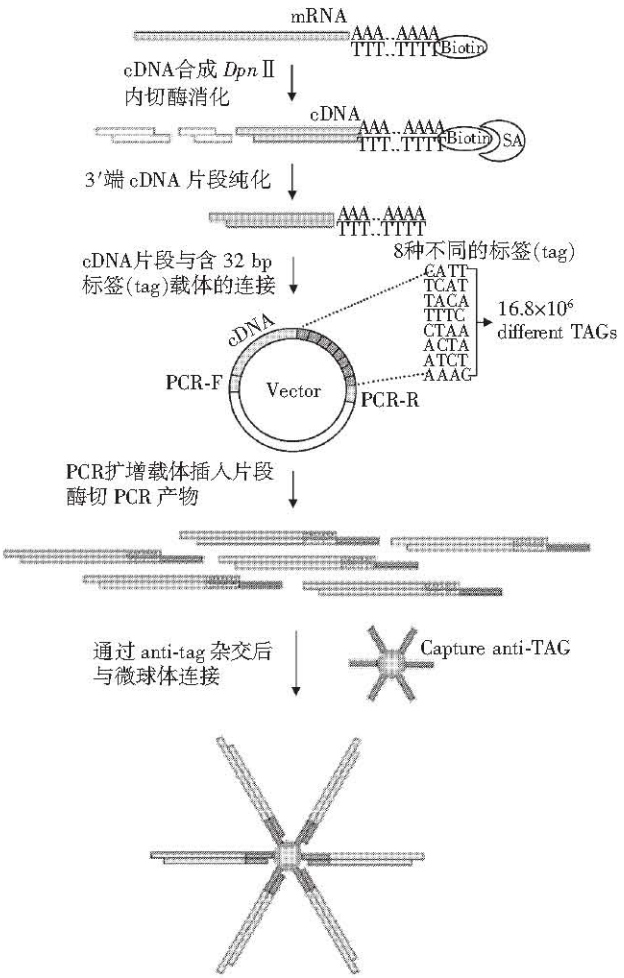

Fig. 1 Attachment of tags to cDNAs[3]

图1 cDNA片段与标签及微球体的结合[3]

洗脱除去寡核苷酸接头，经过*Bbv* Ⅰ酶切后的cDNA模板，进入下一轮分析. 分析所得到的17张荧光显微照片，就可以读出微球体阵列中每一个微球体上长度为17 bp的cDNA模板序列（图2）[3].

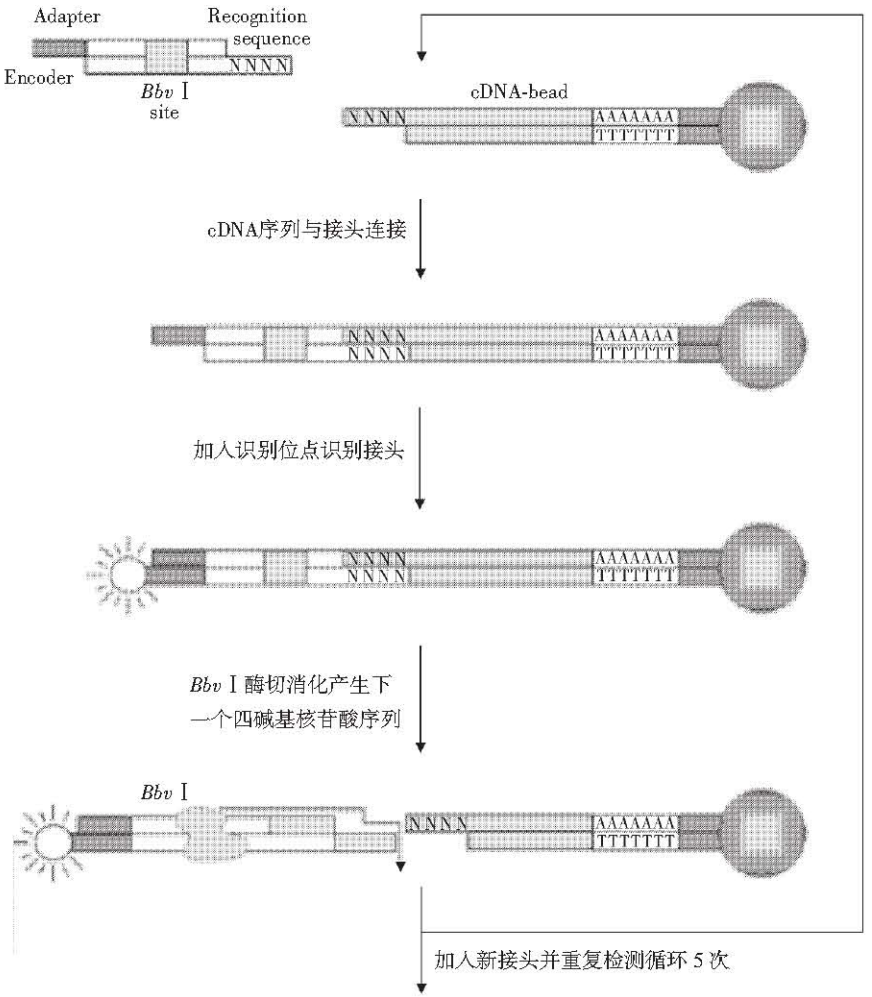

Fig. 2 Determine the 17mer-signature of each cDNA-bead[3]

图2 每个 **cDNA-bead** 序列中 **17 bp** 标签的产生[3]

3 MPSS 应用

根据 MPSS 技术的原理可以知道，MPSS 一方面可提供某一 cDNA 在体内特定发育阶段的拷贝数，另一方面还可测定出相应 cDNA 17 bp 的序列，所以这就为在转录水平上进行基因表达分析提供了强有力的定性和定量手段，很明显，这一技术首先可以应用于不同丰度基因的差异表达分析，制作基因转录图谱，这无疑将加速新基因克隆和基因功能的分析. MPSS 所获得的基因序列可提供 PCR 引物，可通过比较 GenBank EST 数据库等进行基因定位，也可转化为分子标记构建遗传图谱等等，因此该技术可广泛用于动植物体分类学和遗传学，功能基因组学，蛋白质组学等研究.

Hoth 等[4]用 MPSS 克隆出细胞分裂素上调基因 823 和下调基因 917. Christensen 等[5]用 MPSS 分析了单叶 ROB 基因家族的保守亚群和发育调节基因. Jongeneel 等[2]用 MPSS 分析了 HB4a（正常乳腺上皮细胞）和 HCT-116（结肠腺癌细胞）两株细胞的转录子特征. 每株细胞获得了 10^7 个序列标签，建立了一个基因表达短标签的分析平台. 每个细胞株单拷贝表达基因数量为 10 000 ~ 15 000 之间. 两株细胞中绝大多数转录子都可以在已知基因和多聚 A 变异体上找到对应的位置，从表达序列标签上克隆的基因，大约 8 000 个两株细胞能公共表达，而 6 000 个分别特异表达. Potschka 等[6]以大鼠颞叶癫痫模型为研究对象，采用 MPSS 技术克隆大鼠癫痫特异表达基因，结果提示，在海马回中有 263 个特异表达基因，其中，最有意义的是知觉早期基因 Homer1A，其功能与谷氨酸受体修饰有

关，在癫痫大鼠海马回中过表达. Hoth 等[7]在研究基因组 ABA（phytohormone abscisic acid，ABA）反应基因在拟南芥（*Arabidopsis thaliana*）和 abi1-1 突变株中的表达差异时，应用了 MPSS 技术，结果提示，在 ABA 处理的野生株中发现的 1 354 个上调和下调基因，在这些 ABA 反应基因中大多数编码信号传导组分. 在 abi1-1 突变的对 ABA 无反应的对照组中，84.5% 的克隆基因表达减弱，6.9% 基因表达消失，而 8.6% 的基因仍然有一定的调节作用. 因此作者认为与其他几种基因表达分析方法相比较，MPSS 具有高度特异性和敏感性，是在拟南芥野生株中克隆到大量 ABA 反应基因的主要手段.

4 MPSS 技术的特点

目前，应用于基因克隆的技术有 DNA 芯片（DNA microarrays）、基因表达系列分析（serial analysis of gene expression，SAGE）、定量 PCR（quantitative RT-PCR）、差异显示 RT-PCR（differential-display RT-PCR）、抑制消减杂交（subtractive suppress hybridization，SSH）和大规模平行测序（massively parallel signature sequencing，MPSS）等方法. 每一种方法，都有自身的优点和不足. 考察每一种方法的优劣，应该从该技术的特异性、敏感性、可信性、技术难度和运作成本等方面考虑. 尤其是建立的基因表达数据库是否有利于下一步生物信息学的分析.

基于序列分析的技术有基因表达系列分析、大规模平行测序（MPSS）和表达标签序列分析等方法. 相对而言，基于杂交技术的 DDRT-PCR、SSH 和 RNA 点杂交等技术具有可靠、前期操作简单、通量低、后期生物信息学处理较容易、实验成本低等特点. 而基于序列分析的 SAGE、EST 和 MPSS 等具有自动化程度高、通量大、生物信息学处理困难和运作成本高等特点. 所以在整体基因和基因组分析中很难说那一种技术占有绝对的优势，研究者可以根据各自的实验目标选择一种适当的方法.

MPSS 分析系统对基因表达分析过程，诸如微球体阵列的制作，反应液的供排、各种反应条件的控制，图像的处理和数据的分析已经完全自动化，能够在很小的一块微球体阵列上，通过常规的分子生物学手段：连接、酶切、萤光成像等简单几个步骤就可以同时分析数以万计的基因数目，这大大超过了基因的 EST、RNAse 保护分析、DDRT RT-PCR 分析（这几种方法一次只能检测很少的基因表达情况），甚至超过了 SAGE 的一次性分析能力，同时不需要耗费时间做大量的 PCR 实验，不需要对 cDNA 模板做特殊的处理，也不用对探针序列进行提前选择，因此 MPSS 技术分析样品基因表达的操作简便，速度快，时间短. 更为重要的是 MPSS 可根据荧光信号对基因表达水平做定量的分析，能提供基因末端序列信息，这是 MPSS 与 RT-PCR、SAGE 等常规方法不同之处. 另外，MPSS 对基因末端序列与常规测序不同的是，它不需要进行基因片段的分离、克隆再逐一测序，而是具备了 cDNA 芯片、cDNA 微阵列荧光分析法直接读出序列的优点，可同时获得大量 cDNA 末端序列，从而简化了测序过程，这符合后基因组时代基因功能分析的高通量、自动化、微型化的要求[8,9].

MPSS 与基因芯片技术相比较，有下列优点：a. 可以避免在 cDNA 芯片技术中出现的高度同源序列的交叉杂交. 因此可以保证基因的高度特异性. 97.2% 的标签中，17 bp 长度的标签已经足够鉴别基因组中相关的基因. 如此高的鉴别率，cDNA 芯片技术很难达到；b. MPSS 的高分辨率可以检测很低表达水平的基因；c. MPSS 技术检测基因不需要预先知道该基因的相关信息，可以应用于任何生物体的基因表达检测，而 cDNA 芯片技术需要将已知基因片段作为探针固定在片基上[10].

当然该技术同 DNA 芯片技术一样，需要较为昂贵的硬件和相配套的软件协同运做. 目前国内外的相关应用报道较少，因此目前还亟需降低仪器检测的成本，加强推广和普及工作. 总之，MPSS 技术是基因表达定性和定量研究的一种有效工具，它能在短时间内检测细胞或组织内全部基因的表达情况，并能通过与已知基因数据库进行比对，定量显示出基因在细胞或组织内的表达状况，是功能基因组研究和基因发现的有力工具，对于致病基因的识别、药物在组织中的药效分析、揭示基因与疾病之间的传导通路. 揭示基因在疾病中的作用都是非常有价值的，而这些与疾病相关的基因将是非常有价值的药靶. 随着 MPSS 技术的不断发展，相信该技术必将在各种生物基因组功能方面及其相关领域研究中发挥巨大的作用.

参 考 文 献

1 Brenner S, Johnson M, Bridgham J, *et al*. Gene expression analysis by massively parallel signature sequencing（MPSS）on microbead arrays. Nat Biotechnol, 2000, **18**（6）: 630～634

2 Jongeneel C V, Iseli C, Stevenson B J, *et al*. Comprehensive sampling of gene expression in human cell lines with massively parallel signature sequencing. Proc Natl Acad Sci USA, 2003, **100** (8): 4702 ~4705
3 Jeannette R, Eddy B, Jingzhong L, *et al*. Massively parallel signature sequencing (MPSS) as a tool for in-depth quantitative gene expression profiling in all organisms. Briefings in Functional Genomics and Protemics, 2002, **1** (1): 95 ~104
4 Hoth S, Ikeda Y, Morgante M, *et al*. Monitoring genome-wide changes in gene expression in response to endogenous cytokinin reveals targets in *Arabidopsis thaliana*. FEBS Lett, 2003, **554** (3): 373 ~380
5 Christensen T M, Vejlupkova Z, Sharma Y K, *et al*. Conserved subgroups and developmental regulation in the monocot rop gene family. Plant Physiol, 2003, **133** (4): 1791 ~1808
6 Potschka H, Krupp E, Ebert U, *et al*. Kindling-induced overexpression of homer 1A and its functional implications for epileptogenesis. Eur J Neurosci, 2002, **16** (11): 2157 ~2165
7 Hoth S, Morgante M, Sanchez J P, *et al*. Genome-wide gene expression profiling in *Arabidopsis thaliana* reveals new targets of abscisic acid and largely impaired gene regulation in the abi1-1 mutant. J Cell Sci, 2002, **115** (24): 4891 ~4900
8 Pollock J D. Gene expression profiling: methodological challenges, results, and prospects for addiction research. Chem Phys Lipids, 2002, **121** (1 ~2): 241 ~56
9 Brenner S, Williams S R, Vermaas E H, *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): 1665 ~1670
10 Blohm D H, Guiseppi-Elie A. New developments in microarray technology. Curr Opin Biotechnol, 2001, **12** (1): 41 ~47

A Novel Gene Identification Approach: Massively Parallel Signature Sequencing

CHEN Jie*

(*Cancer Treatment Center, Xinqiao Hospital, The Third Military Medical University, Chongqing* 400037, China)

Abstract Massively parallel signature sequencing, MPSS, is an open platform that reveals the expression level of virtually all genes expressed in a sample by counting the number of individual mRNA molecules produced from each gene. The MPSS process involves cloning each mRNA molecule onto the surface of a 5 μm bead. The DNA combitag sequence is attached to a fragment of cDNA. The cDNA library is hybridized to beads. After hybridization, each of the beads displays amplified copies of one and only one starting mRNA molecule. MPSS has a routine sensitivity of a few molecules of mRNA per cell and the results are in a digital format that simplifies data management and analysis. MPSS results will be particularly useful for generating the type of complete data sets that will help to identify the functionally important genes in the sample of interest.

Key words gene, expression analysis, massively parallel signature sequencing (MPSS)

* Corresponding author. Tel: 86-23-68755646, E-mail: jazz0331@sina.com

Received: February 10, 2004 Accepted: March 28, 2004