

文章编号: 1004-0374(2008)04-0589-04

基因功能研究方法的新进展

杜玉梅, 左正宏*

(厦门大学生命科学学院, 厦门 361005)

摘要: 随着生命科学的发展, 研究领域的不断开拓, 越来越多的未知新基因和基因的新功能被科学家们发现, 研究这些未知新基因的功能和已知基因的新功能成为了极其重要的一项内容。本文对基因功能研究的最新方法进行了介绍。

关键词: 基因功能; 功能预测; 生物信息学

中图分类号: Q-31; Q78 **文献标识码:** B

Advances in research methods for gene function

DU Yu-mei, ZUO Zheng-hong*

(School of Life Sciences, Xiamen University, Xiamen 361005, China)

Abstract: With the development of the research areas of life sciences, more and more new genes and new functions of genes were discovered. It would be the main task to study the function of these genes in the next few years. This review introduces several new methods and procedures on the gene function research.

Key words: gene function; function prediction; bioinformatics

1985年, 诺贝尔奖获得者 Dulbecco 第一次提出了人类基因组计划和人类基因组测序的可能性。Dulbecco^[1]指出, 癌症和其他疾病的发生都与基因有关, 并指出测定人类整个基因组序列的途径和重要意义^[1]。1990年, 人类基因组计划正式启动, 在各国科学家的共同努力下, 人类基因组序列图已经完成, 生命科学的研究已进入后基因组时代, 基因组的研究也从结构基因组学转向功能基因组学。目前, 更大的挑战在于如何确定基因的功能和弄清全部的遗传信息, 这将是 21 世纪生命科学研究的重要领域。

本文将对以下研究基因功能的新技术及其新进展作一综述。

1 基因的生物信息学分析

生物信息学以大规模序列信息产出为基本特征, 除了对人类基因的测序外, 还包括了多种模式生物体的基因组测序, 这些序列可以从美国的基因库(Gene Sequence Data Bank, GenBank)、基因组序列数据库(Gene Sequence Data Bank, GSDB)、欧

洲的分子生物学实验室(European Molecular Biology Laboratory, EMBL)和日本的DNA数据库(DNA Data Bank of Japan, DDBJ)中获得。通过序列分析比较工具(如 BLASTn、BLASTx), 可以对基因序列资料中各类信息进行识别和比较, 寻找序列之间的同源性, 得到序列之间的进化关系, 建立基因序列结构和功能的关系。

2 基因的时空表达谱分析

基因的表达在个体发育的不同阶段以及在个体的不同组织和细胞类型中均不相同, 即基因表达的时空性。因此, 在研究一个基因的功能前, 要对基因的时空表达谱进行分析, 包括 mRNA 和蛋白质两个水平上的基因表达谱分析。

2.1 mRNA 水平的表达谱分析 研究 mRNA 水平的基因表达谱分析常用的方法有 Northern blot、原位

收稿日期: 2008-03-30; 修回日期: 2008-04-24

基金项目: 国家自然科学基金项目(40606027)

* 通讯作者: E-mail: zuozhenghong@xmu.edu.cn

杂交、RT-PCR等。Northern blot可对基因进行特异和定量的检测,但测定效率不高,灵敏度也低,不能检出微小的基因表达量,同时实验中使用的放射性物质对人和环境也有危害。作为一种经典的基因表达量分析方法,Northern blot依然被广泛地应用。原位杂交技术由美国耶鲁大学 Gall 和 Pardue 于 1969 年首先创立,广泛用于检测一个特异的 mRNA 在某一种生物体或者组织、细胞里的具体表达位置,对待测核酸分子进行定性、定量及定位分析。RT-PCR 主要有半定量 RT-PCR、实时定量 RT-PCR 以及竞争性定量 RT-PCR: 半定量 RT-PCR 操作简便、快捷,但精确度不高,多用于快速初步分析;实时定量 RT-PCR 是近年来新发展起来的,特异性强、自动化程度高;竞争性 RT-PCR 则是将特异性的目的序列同已知浓度的内标 RNA 一起扩增。通过比较由内标获得的信号和目的模板所获得的信号,确定目的模板的相对含量。

近年来发展了一些新方法,如表达序列标签串联排列连接(tandem arrayed ligation of expressed sequence tags, TALEST)和 GeneCalling。TALEST 是应用含有 II_s 型限制酶位点的寡核苷酸引物,产生在 mRNA 上固定的短(16bp)ESTs。这些 ESTs 与热变性有关的 GC- 锁状标点序列相邻,因此可串联成长阵列,然后通过高通量 DNA 测序识别、分析^[2]。GeneCalling 法研究的对象是用两种不同限制酶消化的 cDNA 样品。用荧光标记的引物扩增、毛细管电泳分离这些标记的片断,然后同时测定每个片断的精确长度。通过电泳比较两个样品中每个点的强度,自动识别不同表达基因的 cDNA 片段。用大小精确的片段和片段旁侧序列查询特定物种的数据库,得出片段信息,而旁侧序列由限制酶消化而来。查询数据库包括转录子的“in silico”消化片段以及所有预测的基因片段。这种预测称为“GeneCalling”,可瞬时显示基因表达差异的临时列表^[3]。

2.2 蛋白质水平的表达谱分析 蛋白质水平上的表达分析的常用技术是 Western blot 和免疫组化等。Western blot 与 Northern blot 类似,不仅可以进行定量分析,而且还能够检测蛋白质的分子质量大小及其聚体形式。免疫组化是研究细胞内蛋白质定位、定量的重要方法,特异性强、敏感性高、能够准确确定蛋白质是在特定组织中的哪些细胞以及在特定细胞的哪个部位中表达。

3 基因的功能预测

在对基因功能进行实验验证之前,可先对基因

的功能进行合理预测,包括功能学上的预测和结构学上的预测。

3.1 利用生物信息学进行功能学上的预测 通过信号肽序列预测分析,初步判定基因的亚细胞定位后,可对其蛋白质的基本理化性质,如氨基酸组成、等电点等进行分析。再在已知的氨基酸序列的基础上,分析预测其高级结构。之后采用 BLAST 和 FASTA 程序进行序列同源性分析,若新确定的蛋白质与已确定的蛋白质显示出结构和功能上的相似性和同源性,则可以从已知基因的功能来预测未知的基因的功能。通过对蛋白质结构数据库的检索比较,可初步确定新基因是否属于某一基因家族或超家族的新成员,然后运用 ClustalW 和 TreeView 软件来进行多重序列比对和分子进化分析,利用这个基因家族中已知基因的结构、功能和它在生物体中的作用,来预测未知基因的结构与功能。

3.2 从结构学方面预测基因的功能 通过 SMART (<http://smart.embl-heidelberg.de/>)或 PROSITE (<http://www.expasy.org/prosite/>) 数据库可以进行蛋白质功能域的分析,对推测新基因的功能提供有价值的信息。再利用 X 衍射和磁共振的方法,对蛋白质的分子结构进行推导和探索,了解分子之间的识别机制和生化反应的原理,也可以预测蛋白质的功能。

4 基因功能的实验学验证

在对基因的功能进行合理的预测后,需要通过实验来进行研究和验证。通常的研究策略是将基因导入到一个细胞或个体中,通过该基因在体内的表达情况,观察细胞生物学行为或个体表型遗传性状的变化,从而鉴定基因的功能。主要的方法有基因敲除和敲入技术以及人工染色体技术等。此外,一些新的技术,如反义技术、microRNA 技术、基因诱捕技术和微阵列分析等也得到了广泛的应用。

4.1 基因敲除和敲入技术 基因敲除是应用 DNA 同源重组原理发展起来的一门技术。1985 年,首次证实的哺乳动物细胞中同源重组的存在奠定了基因敲除的理论基础。1987 年,Thompson 首次建立了完整的 ES 细胞基因敲除的小鼠模型。此后的几年中,基因敲除技术得到了进一步的发展和完善,目前已成为一种较理想的改造生物遗传物质的实验方法。与基因敲除相反,基因敲入是通过同源重组的方法,将基因的编码序列用另一基因的编码序列进行替换的技术。通过基因敲入,可以让基因在体内表达、研究其功能;也可以与之前的基因进行比较,看其是否具有相同的功能。传统的方法只能在

基因组中插入较小的 DNA 片段。Venken 和 Bellen^[4] 将含有 DNA 片段的 P 转座因子转入到质粒里, 质粒能够较 P 转座因子自身更稳定地携带大片断 DNA。这种方法可将 20kb 到 133kb 的 DNA 敲入果蝇基因组。这一突破使生物学家向果蝇体内敲入大片的 DNA 成为可能。

4.2 人工染色体的转导 质粒载体不仅承载能力有限, 且表达水平低、缺乏组织特异性等。因此, 克隆大片段 DNA 常用酵母人工染色体(YAC)、细菌人工染色体(BAC)等。它们不仅可产生较高的表达水平和组织特异性, 还可精确地调节重组。自 20 世纪末, 又出现了一种全新的载体系统——人类人工染色体(HAC)。与其他基因载体相比, HAC 能携带包含完整基因或多个基因以及基因的所有外显子和附近染色体区域的调控区的大片段 DNA, 为目的基因提供了一个与其在正常染色体上一致的环境, 保证了转基因在正常细胞中时空性的表达。同时, HAC 不整合到基因组中, 从而不产生宿主基因组本身的插入突变和转基因沉默等现象, 能够使基因表达的时限增长。Shitara 等^[5] 就将这种非插入型的 HAC 载体成功应用于人端粒逆转录酶(hTERT)的研究。

4.3 反义技术 反义技术包括三类: 反义寡核苷酸、核酶和小干扰 RNA。反义寡核苷酸包括反义 DNA 和反义 RNA。1978 年, Zameenik 利用人工合成的与劳氏肉瘤病毒(RSV)的 mRNA 互补的 DNA 来抑制 RSV 增殖, 阻止了 RSV 使鸡红细胞癌变, 使其成为研究反义 DNA 的第一人。目前反义 DNA 用于基因功能研究效果较好, 也可用自动 DNA/RNA 合成仪很方便地合成, 但因合成后的未修饰寡核苷酸对核酸酶的抵抗力较弱, 不易透过细胞膜, 与靶序列的亲合力也较低, 故经常需要修饰。反义 RNA 则是通过与靶 mRNA 形成较稳定的二聚体来抑制靶基因的表达, 其作用机理可能在 DNA 复制、转录及翻译水平上抑制靶基因的表达。核酶是一类具有催化活性的特殊的 RNA 分子, 具有高度专一内切核酸酶活性。单个核酶分子可以结合多个 mRNA 分子并使之在特定部位断裂, 而核酶本身具有不易受 RNase 攻击的较稳定的空间结构, 使得催化效率比反义 RNA 高。核酶已经成功地应用于培养细胞中基因表达的阻断, 主要靶基因包括 *HIV*、*c2fos*、*bcr2abl* 和 *H2ras* 等^[6]。小干扰 RNA 可通过双链 RNA(dsRNA)介导特异性的降解靶 mRNA, 导致转录水平或转录后水平的基因沉默。1995 年, Guo 和 Kempheus^[7] 在研究秀丽新小杆线虫(*C. elegans*)的

par1 基因功能时最先发现这种现象, 但直到 1998 年, Fire 等^[8] 才解释了这种现象, 他们发现将 dsRNA 注入线虫后可以有效地引起序列特异性的基因抑制, 并将这种转录后水平的基因沉默机制称为 RNAi(RNA interference)。RNAi 广泛存在于自然界中, 其高度的序列专一性和高效的干扰能力, 可以使特定的基因表达降低或沉默, 是研究基因功能的强有力的工具。Xia 等^[9] 利用 RNAi 降低 Sp1 和 CREB 的表达, 同时通过 EMSA 位点的突变和缺失, 从而验证了 cAMP 反应位点(CRE)和 Sp1 结合位点在调节人类 SNF2L 基因基底活性中的作用。

4.4 microRNA 成熟的 microRNA 长约 19–25 nt, 是一类重要的内源性单链的非编码小分子 RNA, 可调节与其序列互补 mRNA 的表达。1993 年, 由 Lee 等在线虫中首次发现。microRNA 在不同物种间具有高度的保守性, 表达具有细胞或组织特异性, 与细胞的生长、增殖、分化和衰亡等相关。它可以在转录和翻译水平来抑制基因的表达。根据最近的研究推测, 目前已经有多达 200 个基因可以通过 microRNA 进行调节^[10,11]。Stewart 等^[12] 在研究猪胎儿血红蛋白中 short hairpin RNA(shRNA)的干扰效应时, 结合 microRNA 技术和 RT-PCR 检测技术, 来进一步验证和研究 shRNA 的干扰效应。可以说, microRNA 技术和 RT-PCR 检测技术的结合, 使得基因功能的研究技术得到了新的发展^[12,13]。

4.5 基因诱捕技术和基因诱捕数据库 基因诱捕技术是近几年发展起来的, 是基因打靶技术的进一步发展, 已成为研究基因功能及分析其生物学现象的重要工具。其基本原理是: 通过物理、化学、生物等方法, 将带有外源基因的 DNA 载体导入 ES 细胞中, 使内源基因突变, 并在被诱捕序列启动子的转录控制下表达插入的报告基因(常为新霉素或半乳糖苷酶基因)以鉴定突变^[14]。无启动子、增强子的报告基因在 ES 细胞中通过同源重组得到重组子后, 分析不同发育阶段、不同组织器官中报告基因的表达情况, 可以研究重组部分内源基因的表达特性。基因诱捕载体在整合位点可利用内源基因调控元件模仿内源基因表达, 使其表达终止, 从而可以阐明内源基因的功能, 因此广泛应用于基因功能的研究。Uchihashi 等^[15] 利用该技术来研究与软骨分化相关因子的作用。目前, 公共投资的大规模基因诱捕计划在一些国家开展, 基因诱捕也臻于完善, 基因诱捕组织也逐渐增多。近年来, 全世界主要的基因诱捕组织联合成立了国际基因诱捕联盟(IGTC), 并为此

开发了一个以用户为导向的网站(<http://www.genetrap.org>), 其目的在于集中接纳所有可以公共获得的基因诱捕资源^[16]。IGTC 数据库及网站的建立标志着大规模小鼠基因诱捕的一个重大进展。通过基因诱捕计划的世界性合作, 发展了标准化的鉴定及注释程序用以进行诱捕数据分析, 为基因功能研究人员在第一时间提供了资讯平台。

4.6 微阵列分析 微阵列(microarray)是近年来发展起来的可用于大规模快速检测基因差异表达、基因组表达谱、DNA 序列多态性、疾病相关基因的一项研究基因功能的新技术, 于1984年由 Geysen 等首次开发出。它包括 cDNA 微阵列和 DNA 芯片, 其原理为: 将成千上万条 DNA 片段(cDNA、表达序列标签等)按横行纵列方式在固相支持物上有序点样。检测时, 先用来自不同生理状态和发育阶段的 mRNA 作为模板, 以放射性同位素或荧光标记的 dNTP 为底物反转录合成 cDNA, 再用所得 cDNA 与微阵列或 DNA 芯片进行杂交, 然后通过计算机对结果进行判读和处理, 从而判断待测样品中基因是否存在或者存在多少。

微阵列分析是一种新的大规模检测基因表达的技术, 具有高通量分析的优点。采用微阵列分析, 可以进行 DNA 或 RNA 表达水平的高效快速的检测^[17,18]。微阵列分析打破了以往“一种疾病一个基因”的研究模式, 通过对个体在不同生长发育阶段或不同生理状态下大量基因表达的平行分析, 研究相应基因在生物体内的功能, 阐明不同层次多基因协同作用的机理, 已成为后基因组时代生命科学研究强有力的工具。

5 展望

基因功能的研究是科学研究的重要内容, 也是一项复杂的工程。除了以上的研究方法外, 还有许多其他的方法可用于基因功能的研究。在实际工作中, 研究者需要根据具体情况制定某一特定基因功能的研究方案, 且对一个特定基因功能进行全面、系统的研究。相信通过广大研究者的努力创新、国际间的广泛合作以及新技术新方法的开创和应用, 还会有更多更好的方法出现。在不久的将来, 我们能够更全面地了解基因组的功能, 完成生命周期表的制作, 认识人类基因组中约 3 万个基因是如何在人类的生长、发育、疾病、衰老、死亡等过程中发挥功能及相互协调, 解密生命的奥秘。

[参 考 文 献]

- [1] Dulbecco R. A turning point in cancer research: equencing the human genome. *Science*, 1986, 231(4742): 1055-6
- [2] Spinella DG, Bernardino AK, Redding AC, et al. Tandem arrayed ligation of expressed sequence tags(TAL EST): a new method for generating global gene expression profiles. *Nucleic Acids Res*. 1999, 27(18): e22
- [3] Shimkets RA, Lowe DG, Tai JT, et al. Gene expression analysis by transcript profiling coupled to a gene database query. *Nat Biotechnol*, 1999, 17(8): 798-03
- [4] Venken KJ, Bellen HJ. Transgenesis upgrades for *Drosophila melanogaster*. *Development*, 2007, 134(20): 3571-84
- [5] Shitara S, Kakeda M, Nagata K, et al. Telomerase-mediated life-span extension of human primary fibroblasts by human artificial chromosome (HAC) vector. *Biophys Res Commun*, 2008, 369(3): 807-11
- [6] Rossi JJ. Ribozyme therapy for HIV infection. *Adv Drug Deliv Rev*, 2000, 44(1): 71-8
- [7] Guo S, Kempheus KJ. A gene required for establishing polarity in *C. elegans* embryos, encodes a putative Ser/Thr kinase that is asymmetrically distributed. *Cell*, 1995, 81(4): 611-20
- [8] Fire A, Xu SQ, Montgomery MK, et al. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature*, 1998, 391(6669): 806-11
- [9] Xia Y, Jiang BC, Zou YX, et al. Sp1 and CREB regulate basal transcription of the human SNF2L gene. *Biochem Biophys Res Commun*. 2008, 368(2): 438-44
- [10] Lewis BP, Burge CB, Bartel DP. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell*, 2005, 120(1): 15-20
- [11] Krek A, Grun D, Wolf R, et al. Combinatorial microRNA target predictions. *Nat Genet*, 2005, 37(5): 495-500
- [12] Stewart CK, Li JL, Golovan SP. Adverse effects induced by short hairpin RNA expression in porcine fetal fibroblasts. *Biochem Biophys Res Commun*, 2008, 370(1): 113-7
- [13] Asirvatham AJ, Gregorie CJ, Hu ZH, et al. MicroRNA targets in immune genes and the Dicer/Argonaute and ARE machinery components. *Mol Immunol*, 2008, 45(7): 1995-2006
- [14] Takeuchi T. A gene trap approach to identify genes that control development. *Dev Growth Differ*, 1997, 39(2): 127-34
- [15] Uchihashi T, Kimata M, Tachikawa K, et al. Involvement of nuclear factor I transcription/ replication factor in the early stage of chondrocytic differentiation. *Bone*, 2007, 41(6): 1025-35
- [16] Nord AS, Chang PJ, Conklin BR, et al. The international gene trap consortium website: a portal to all publicly available gene trap cell lines in mouse. *Nucleic Acids Res*, 2006, 34(Database issue): D642-8
- [17] Koga Y, Matsuzaki A, Suminoe A, et al. Expression of cytokine-associated genes in dendritic cells (DCs): Comparison between adult peripheral blood- and umbilical cord blood-derived DCs by cDNA microarray. *Immunol Lett*, 2008, 116(1): 55-63
- [18] Huggins CE, Domenighetti AA, Ritchie ME, et al. Functional and metabolic remodelling in GLUT4 - deficient hearts confers hyper-responsiveness to substrate intervention. *Mol Cell Cardiol*, 2008, 44(2): 270-80