

综述与专论

功能基因组学的研究内容与方法

赵剑华 王秀琴 刘芝华 吴

(中国医学科学院 肿瘤研究所, 分子肿瘤学国家重点实验室, 北京 100021)
(中国协和医科大学)

摘要 基因组学的研究已从结构基因组学转向功能基因组学。综述了功能基因组学研究的内容和方法, 主要包括应用微点阵、基因表达系列分析 (SAGE)、蛋白质组、生物信息学等方法来研究基因组表达概况、基因组多样性、模式生物体等。

关键词 人类基因组计划, 功能基因组学, 蛋白质组学, 模式生物体, 生物信息学
学科分类号 Q75

随着人类基因组计划 (HGP) 的顺利进行, 生物医学研究已进入后基因组时代 (post-genome era)^[1]。基因组学的研究从结构基因组学 (structural genomics) 过渡到功能基因组学 (functional genomics)。结构基因组学代表基因组分析的早期阶段, 以建立生物体高分辨遗传、物理和转录图谱为主。功能基因组学代表基因分析的新阶段, 是利用结构基因组学提供的信息, 系统地研究基因功能。它以高通量、大规模实验方法及统计与计算机分析为特征^[2]。

1 基因组的表达

1.1 基因表达概况的研究

基因表达概况是比较不同组织和不同发育阶段, 正常状态和疾病状态, 以及体外培养的细胞中基因表达模式的差异。传统的 RT-PCR、RNase 保护实验、RNA 印迹杂交等方法也能做到这一点, 但 these 方法只能一次做一个或几个基因。高通量的基因表达分析方法需要借助新技术——微点阵 (microarray) 和基因表达系列分析 (serial analysis of gene expression, SAGE) 等方法。

微点阵是指将几百甚至几万个寡核苷酸或 DNA 密集排列在硅片、玻璃片、聚丙烯或尼龙膜等固相支持物上, 作为探针。把要研究的样品 (称为靶 DNA) 标记后与微点阵进行杂交, 用合适的检测系统进行检测。根据杂交信号强弱及探针位置和序列, 即可确定靶 DNA 的表达情况以及突变和多态性的存在。微点阵包括 DNA 芯片 (DNA

chip)、cDNA 表达膜等。

SAGE 是同时、定量分析大量转录本的一种方法。基本原理是从转录本特定位置 (靠近 3' 端) 分离出短核苷酸序列 (9~10 bp), 它包含了能代表该转录本的足够信息。将一个体系的所有转录本中这一短序列分离并连接到一个克隆载体中进行测序, 便可以得到该体系所有转录本的表达情况。这种方法在正常、癌旁、癌组织中基因的差异表达研究方面有独到的优点, 有助于发现肿瘤特异基因^[3]。

1.2 基因产物——蛋白质功能的研究

大部分细胞生命活动发生在蛋白质水平而不是 RNA 水平, 因此即便知道了全部基因的表达概况也难以阐明基因的实际功能。基因在生物体整体的功能最终由其编码的蛋白质在细胞水平上体现。

1.2.1 单个基因的蛋白质体外表达方法: DNA 重组技术的应用使体外合成蛋白质成为可能。克隆化基因有两种表达系统: 原核表达系统和真核表达系统。

1.2.2 蛋白质组 (proteome) 的研究:

a. 蛋白质组和功能蛋白质组的概念。蛋白质组的概念是澳大利亚学者 Wasinger 等^[4]于 1994 年提出, 指的是由基因组编码的全部蛋白质。从这个定义看, 蛋白质组内蛋白质的数目应该等于基因组内编码蛋白质的基因的数目 (准确地说是 ORF 的

数目)。但在生物体内这样的蛋白质组是不存在的,从基因表达的角度看,蛋白质组中蛋白质的数目总是少于基因组中开放读框(ORF)的数目。但从蛋白质修饰的角度看,蛋白质的数目又远远多于基因组中ORF的数目。基因组基本上是固定不变的,而蛋白质组是动态的,具有时空性和可调节性,能反映某基因的表达时间、表达量,以及蛋白质翻译后的加工修饰和亚细胞分布等。因此提出了功能蛋白质组学(functional proteomics)的概念,它是指在特定时间、特定环境^[5]和实验条件下^[6]基因组活跃表达的蛋白质。功能蛋白质组只是总蛋白质组的一部分。

b. 研究蛋白质组的技术。蛋白质组分析主要涉及两个步骤:蛋白质的分离和蛋白质的鉴定。用于蛋白质分离的技术主要有双向凝胶电泳(2-dimensional gel electrophoresis, 2-DE)。用于蛋白质鉴定的技术有Edman降解法测N端序列、质谱技术(mass spectrometry, MS)和氨基酸组成分析等。其中质谱技术应用广泛,发展较快。

c. 蛋白质组分析的意义及应用。蛋白质组提供了如下信息:从基因序列预测的基因产物是否及何时被翻译;基因产物的相对浓度;翻译后修饰的程度。这三点都无法单独从核酸序列进行精确预测。蛋白质组分析能不依赖于DNA序列信息而进行,而且反过来又有助于对基因组的进一步了解。它有效地补充了差异显示、微点阵、表达序列标签(EST)分析、直接或间接消减杂交、染色体连锁分析及核酸测序等研究基因组的方法^[6]。

与蛋白质组分析相应的一个领域是癌发生学(carcinogenesis)。对一个特定细胞基因组表达的所有蛋白质进行质量和数量分析,可确定全部蛋白质表达谱。Robinson等^[7]用蛋白质组分析的方法鉴定了UVB诱导的黑色素瘤细胞系在不同阶段蛋白质表达模式。

1.3 蛋白质-蛋白质相互作用的研究

对细胞活动的机制和功能深入理解的一个方法是确定蛋白质间的相互作用,因为参与一个共同的生命活动(如信号级联反应或大分子复合物等)的蛋白质经常相互接触。酵母双杂交系统是一种广为应用的鉴定相互作用伙伴的方法。除此之外,还有单杂交系统(one-hybrid system)和三杂交系统(three-hybrid system)及反向(reverse)杂交系统。单杂交系统是将对目的基因表达具有调控作用的某一已知顺式作用元件同报告基因相连后,与一融合

的表达型cDNA文库共转染酵母细胞,根据对报告基因表达活性的检测筛选出与已知顺式作用元件相结合的未知因子的cDNA。三杂交系统主要用来分析蛋白质与RNA、蛋白质与肽配体、蛋白质与有机小配体、蛋白质与蛋白激酶之间复杂的相互作用。反向单/双杂交系统主要是用来鉴定破坏大分子相互作用的因子^[8]。

2 动物体内研究基因功能的方法

2.1 模式生物体和比较基因组学

鉴定基因功能最有效的方法是观察基因表达被阻断或增加后在细胞和整体水平所产生的表型变化^[1]。这就需要建立一个有效的模式生物体(model organism)。迄今为止已有许多生物被作为模型,如大肠杆菌、酿酒酵母、美丽线虫、果蝇、斑马鱼、小鼠等。但在研究人类疾病和哺乳动物发育时,小鼠更有优势。

所有生物都是通过一个共同的进化树联系在一起,因此研究一个生物可为其他生物提供有用的信息。人类基因组计划(HGP)把比较基因组学(comparative genomics)作为今后工作重点之一。在相继完成了流感嗜血杆菌、大肠杆菌、酿酒酵母、美丽线虫等生物的全基因组序列测定后,预计2002年完成果蝇、2005年完成小鼠的全基因组测序,同时逐渐从结构的比较转向功能的比较^[9]。人们已发现几种决定人类遗传性疾病的基因有酵母同源基因,这些基因的功能可以用一系列只能用于酵母的遗传工具来确定。可见,简单模式生物体基因结构和功能的研究为人类复杂基因的研究提供了线索和方法。

2.2 诱变技术在功能基因组学中的应用

2.2.1 定向诱变(targeted mutagenesis):定向诱变是利用同源重组技术,使胚胎干细胞(embryonic stem cell, ES cell)内目的基因产生定点突变。这些突变可进一步用于基因敲除、转基因动物、显性负突变等研究。最近两年发展了许多构建靶结构的新方法,如酿酒酵母中微同源重组(microhomologous recombination)^[10],通过PCR的方法产生一个特定的靶DNA片段,这个片段含有一个两侧带有与酵母基因同源的35~50 bp作为选择性标记,就足以促进酵母的同源重组,而在小鼠ES细胞中,至少需要1.9 kb的连续基因组DNA才能产生有效的同源重组。以往这种方法只用在酵母中,现在也用到小鼠上。

2.2.2 表型诱变 (phenotype-driven mutagenesis):

上述定向诱变方法是用于已知基因的突变, 而表型诱变是用于未知基因. 主要优点是无需知道哪个基因以及这些基因的何种突变导致特定的表型或疾病^[1]. 用表型诱变剂进行诱变后, 可以用筛查整个基因组的办法来寻找新的显性或隐性突变. 该方法需要大量的小鼠杂交群体, 工作量较大, 但这种全基因组扫描法是筛查整个基因组中单一突变的最好方法, 因为任何一个导致一定表型的可能突变都可以被检测出来.

3 基因组多样性

人类是一个具有多态性的群体, 不同群体和个体在生物学性状及在对疾病的易感性/抗性上存在差别. 在全基因组测序基础上进行基因组水平再测序来直接识别序列变异, 以进行多基因疾病及肿瘤相关基因的研究, 将成为功能基因组时代的热点.

4 生物信息学在功能基因组学中的应用

生物信息学 (bioinformatics) 是用数理和信息科学的观点、理论和方法去研究生命现象, 组织和分析呈指数增长的生物学数据的一门学科. 研究 DNA 和蛋白质, 以计算机为主要工具, 发展各种软件, 对日益增长的 DNA 和蛋白质的序列和结构进行收集、整理、储存、发布、提取、加工、分析和发现. 生物信息学由数据库、计算机网络和应用软件三大部分组成. 结构基因组学提供了巨大的 DNA 和蛋白质数据, 功能基因组学的一个任务就是如何充分利用数据库去研究基因功能.

综上所述, 虽然 HGP 最初目标的实现已指日可待, 结构基因组研究已接近尾声, 但后基因组时代的功能基因组学已经来临, 其内容丰富繁杂, 需要不断地创造新方法、开发新技术去完成这个跨世纪任务.

参 考 文 献

- 1 Woychik R P, Klebig M L, Justice M J, *et al.* Functional genomics in the post-genome era. *Mut Res*, 1998, **400** (1~2): 3~14
- 2 Hieter P, Boguski M. Functional genomics: its all how you read it. *Science*, 1997, **278** (5338): 601~602
- 3 Velculescu V E, Zhang L, Vogelstein B, *et al.* Serial analysis of gene expression. *Science*, 1995, **270** (5235): 484~487
- 4 Wasinger V C, Cordwell S J, Cerpa Poljak A, *et al.* Progress with gene-product mapping of the Mollicutes: *Mycoplasma genitalium*. *Electrophoresis*, 1995, **16** (7): 1090~1094
- 5 Cordwell S J, Basseal D J, Bjellqvist B, *et al.* Characterisation of basic proteins from *Spiroplasma melliferum* using novel immobilised pH gradients. *Electrophoresis*, 1997, **18** (8): 1393~1398
- 6 Humphery-Smith L, Cordwell S J, Blackstock W P. Proteome research: complementarity and limitations with respect to the RNA and DNA worlds. *Electrophoresis*, 1997, **18** (8): 1217~1242
- 7 Robinson E S, Dooley T P, Williams K L. UV-induced melanoma cell lines and their potential for proteome analysis: a review. *J Exp Zool*, 1998, **282** (1~2): 48~53
- 8 Brachmann R K, Boeke J D. Tag games in yeast: the two-hybrid system and beyond. *Curr Opin Biotechnol*, 1997, **8** (5): 561~568
- 9 Collins F S, Patrinos A, Jordan E, *et al.* New goals for the U. S. human genome project: 1998~2003. *Science*, 1998, **282** (5389): 682~689
- 10 Manivasakam P, Weber S C, McElver J, *et al.* Micro-homology mediated PCR targeting in *Saccharomyces cerevisiae*. *Nucl Acids Res*, 1995, **23** (14): 2799~2800

Content and Methods of Functional Genomics.
ZHAO Jian-Hua, WANG Xiu-Qin, LIU Zhi-Hua,
WU Min (*National Laboratory of Molecular
Oncology, Institute of Oncology, Chinese Academy
of Medical Sciences & Peking Union Medical
College, Beijing 100021, China*).

Abstract The main direction of study in genomics has been transferred from structural genomics to functional genomics. Content and methods of functional genomics are reviewed, including gene expression profile, genomic diversity, model organism, genomic comparison and evolution studied by microarray, serial analysis of gene expression, proteomics, bioinformatics, and so on.

Key words human genome project, functional genomics, proteomics, model organism, bioinformatics