

造血干/祖细胞大规模测序克隆的 一个 ABC 转运体蛋白基因

张庆华 茅 矛 何凯丽 傅 刚 黄秋花 陈 竺

(上海第二医科大学附属瑞金医院 卫生部和上海市人类基因组研究重点实验室, 上海 200025)

A Novel Human ABC Transporter Gene Cloned with Large Scale EST Sequencing

ZHANG Qinghua MAO Mao HE Kaili FU Gang HUANG Qiuhua CHEN Zhu

(Key Laboratory for Human Genome Research, Ruijin Hospital, Shanghai Second Medical University, Shanghai 200025)

ABC 转运体家族是已知的最大基因家族之一⁽¹⁾, 在进化上具高度的保守性, 其成员含有 ATP 结合小匣(ATP binding cassette, ABC)。最近, 我们在对具造血功能的 CD34+造血干/祖细胞(HSPC)的基因表达谱的研究中, 通过大规模的 cDNA 测序, 克隆到一个 ABC 家族的新成员 hABC7, cDNA 2 384bp, 编码 752aa 的蛋白质, 通过同源性比较与小鼠的 mABC7 为直向同源物, 具 N 端的穿膜功能区和 C 端的 ABC 功能区。该基因被 RH 法定位于 Xq21.3, 电子 Northern 显示在多种组织有表达。

1 方法与结果

收集脐带血, 分离出单个核细胞, 经 MACS 分离出 CD34+细胞⁽²⁾, 抽提总 RNA, 选择 CapFinder 系统 (CLONTECH)⁽³⁾, 根据说明进行 PCR 扩增, 构建富含全长 cDNA 的 lambda-ZAP 文库, 经体内剪切为 pBluescript 质粒后随机挑取细菌克隆, 提取质粒, 用通用引物进行测序反应, ABI377 全自动测序, 得到部分 cDNA 表达顺序(EST)。由此得到约 10 000 条 EST 顺序, 经 BLAST 软件查询 GenBank 数据库(截止至 1997 年 9 月 102.0 版), 将这些顺序分为已知和未知基因两大类。其中一个未知基因克隆与小鼠的 mABC7 有高度的相似性⁽⁴⁾, 将此克隆命名为 hABC7, 采用部分缺失及引物延伸等方法测序, hABC7 全长 2 384bp, ORF 为 9~2274, 编码一个 752aa 的可能蛋白质。虽然在 5'上游没有同阅读框架的终止密码, 但第一个 ATG 附近的碱基顺序符合真核生物翻译起始点保守的 Kozak 顺序⁽⁵⁾, 故第一个 ATG 可能即为翻译的起始点(GenBank 登录号 AF03950)。

通过 GCG (Wisconsin)软件包提供的有关软件对 hABC7 的蛋白进行分析显示, hABC7 为 ABC 家族的一个成员, 蛋白的分子结构分为 N 端和 C 端两个区域, N 端具 6 个疏水性的穿膜结构, C 端具典型的保守性的 ABC 结构。与小鼠的 mABC7 相比, hABC7 的 5'端长 184bp, 蛋白结构上 N 端长 123aa。在核苷酸水平, 2 129bp 范围内 91.1%相同, 而蛋白水平 630aa 范围内 93.6%相同⁽⁴⁾。

辐射杂种(Radiation Hybrid, RH)技术是一种新发展的用于基因染色体精确定位的方法⁽⁶⁾, 我们选用 Stanford G3 辐射杂种细胞系嵌盘(Research Genetics, Huntsville, Alabama), 根据 hABC7 的 3'部分顺序, 合成 PCR 引物(5':TTGCTTGCTAACCCTCATAG, 3':CAGTGCAAATATGTAGTCCA), 两次 PCR 扩增结果经 Stanford 大学人类基因组中心(SHGC)的辐射杂交定位服务器(rhserver shgc.stanford.edu)处理, 得到与所定位基因紧密连锁的 LOD 值大于 6.0 (6.61)的顺序标签位点(STS)(AFM303WD1), 结合 SHGC(<http://www.shgc.stanford.edu/Mappinrh/search.html>)和 Genethon(<http://www.cephbh.fr/quickmap.html>)的 RH 图谱和遗传图谱, 将 hABC7 基因定位于 Xq21.3。

用 hABC7 的全长 cDNA 顺序作“探针”，查询公共 EST 数据库(截止至 1997 年 9 月 102.0 版)，进行电子“Northern”，可得到 25 个与之部分序列完全匹配的 EST，即可认为该 EST 为 hABC7 基因的一部分序列，得到该 EST 的组织即表达 hABC7 基因，由此可知有 8 种组织分别表达 hABC7(大肠、胰腺、前列腺、脂肪瘤、肺癌、总胚胎、神经内皮细胞、B 细胞)。

2 讨 论

ABC 转运体家族是已知的最大基因家族之一，其成员的膜功能域都有 6 个由疏水性氨基酸组成的穿膜结构，以及一个在成员中极其保守的 ATP 结合功能域⁽⁸⁾，它们发挥活性的形式是以两个跨膜功能域和两个 ABC 功能域共同作用。在 CFTR、MDR 等大分子中，这些功能域位于同一条肽链分子中⁽⁸⁻¹⁰⁾，而在如 TAP1、TAP2、ALD、ALDR、PMP70 等分子中，仅各有一个这样的功能域，它们以同/异二聚体形式发挥功能，象 hABC7 一样，故 hABC7 也可能以同或异二聚体形式发挥功能。它们多以消耗能量的方式进行物质的跨膜转运，hABC7 对所转运物质的选择性、亲和性以及其它的生物学意义正在进行进一步的研究。

我们实验中克隆到的 hABC7 基因 cDNA 全长 2 384bp，这和 mABC7 以及多个人类的 EST 一致。Savary 等利用数据库中的 EST(T78172) (对应于 hABC7 的 1287~1665)为探针，Northern 显示出数个条带，而主要条带为 3kb，提示 hABC7 存在不同的 3'UTR⁽⁴⁾。

在基因的进化上，基因家族成员多起源于祖先基因的倍增和突变，经基因的转座移位后分布于不同的染色体区域，在许多染色体上都有 ABC 家族成员的分布，我们用 RH 法将 hABC7 定位于 Xq21.3，有人将对应于 hABC7 基因的 ESTU66679 定位与 Xq21-22⁽⁷⁾，这与我们的结果一致。而 Savary 等用对应于 3'端的 EST 作探针，FISH 方法定位于 Xq12-13⁽⁴⁾，造成这种差别的原因有待进一步深入研究，采用的定位方法似乎是关键因素。随着人类基因组计化的进展，越来越多的组织器官都被进行 EST 测序，在公共数据库中已积累越来越多的 EST，故可以用数据库查找的办法来判别基因的组织表达谱，但由于多种文库已经被均化，故尚不能反映该基因的表达水平的差异。由于在多种组织细胞皆有表达，除造血以外，hABC7 可能还有其它的生物学意义。

参 考 文 献

- 1 Henikoff S, Green E A, Pietrokovski S. Gene familis: the taxonomy of protein paralogs and chimeras. *Science*. 1997, 278(5338): 609~614
- 2 石学耕, 周 隽, 史桂英等. 联合使用 MACS、FACS 分离造血干/组细胞及其亚群. *中国癌症杂志*. 1997, (7): 195~199
- 3 Mao M, Fu G, Wu J S *et al*. Identification of genes expressed in human CD34⁺ hematopoietic stem/progenitor cell by expressed sequence tags and efficient full-length cDNA cloning. *Proc. Natl. Acad. Sci. USA*, 1998, 95(14): 8175~8180
- 4 Savary S, Allikmets R, Denizot F *et al*. Isolation and chromosomal mapping of a novel ATP-binding cassette transporter conserved in mouse and human. *Genomics*. 1997, 41(2): 275~278
- 5 Kozak M. Point mutations define a sequence flanking the AUG initiator that modulates translation by eukaryotic ribosomes. *Cell*, 1986, 44(2): 283~292
- 6 Walter M A, Spillet D J, Thomas T *et al*. A method for constructing radiation hybrid maps of whole genomes. *Nat. Genet.*, 1994, 7(1): 22~28
- 7 Allikmets R, Gerrard B, Hutchinson A *et al*. Characterization of the human ABC superfamily: isolation and mapping of 21 new genes using the expressed sequence tags database. *Hum. Mol. Genet.*, 1996, 5(10): 1649~1655
- 8 Hyde S C, Emsley P, Hartshorn M J *et al*. Steuctural model of ATP-binding proteins associated with cystic fibrosis, multidrug resistance and bacterial transport. *Nature*, 1990, 346(6282): 362~365
- 9 Gekeler V, Weger S, Probst H. mdrl/P-glycoprotein gene segments analyzed from various human leukemic cell lines exhibiting different multidrug resistance profiles. *Biochem. Biophys. Res. Commun.*, 1990, 169(2): 796~802
- 10 Lincke C R, Smit J J *et al*. Structure of the human MDR3 gene and physical mapping of the human MDR locus. *J. Biol. Chem.*, 1991, 266(8): 5503~5510