一个与人鼻咽癌相关新基因的克隆

何志巍 谢 鹭 许亮国 蓝 轲 刘卫东 张 玲
任彩萍 史建凌 周 文 姚开泰*

(湖南医科大学肿瘤研究所卫生部癌变原理重点实验室, 长沙 410078. * 联系人)

摘要 利用高密度 cDNA 微阵列表达检测技术, 筛选出在成人正常鼻咽组织高表达而在成人鼻咽低分化鳞状细胞癌组织中低表达的 EST N27741, 用 RT-PCR 技术证实之, 进而由该 EST 克隆出长度为 1 096 bp 的新基因. 利用生物信息学分析发现, 该序列含有完整的阅读框架, 编码 256 个氨基酸, 在 5′ 端起始密码子上游有终止密码子 TAA, 3′ 端有加尾信号 AATAAA 和 poly A 尾. 该新序列在非冗余的核酸序列数据库中比较, 无同源的已知基因存在, 被认为是一个新的与鼻咽癌相关的基因.

关键词 鼻咽癌 新基因 克隆

鼻咽癌(NPC)是我国南方及东南亚地区发病率较高的一类恶性肿瘤, 病因学研究表明: EB 病毒(EBV)感染、化学促癌物和/或致癌物如 TPA 等在致癌过程中均可能起十分重要的作用[1~3]; 而分子遗传学、分子生物学研究表明 NPC 是一个多基因遗传病[4~6], 但目前仍未找到在鼻咽部特异高或低表达的癌基因或抑癌基因. 本工作采用目前国际上最新的 cDNA 表达微阵列(cDNA expression microarray)技术[7~10], 筛选了成人正常鼻咽组织(NP)和 NPC 组织 5 184 个基因或表达序列标签(expression sequence taq, EST)的表达差异. 将 NP 中高表达而在 NPC 中低表达的 EST N27741 克隆其全长并对其编码蛋白质的氨基酸的理化特性进行了初步预测.

1 材料与方法

(ⅰ) 标本来源. 26 例病理诊断为低分化鳞状细胞癌(NPC)的鼻咽癌组织及 16 例成人慢性鼻咽组织均由湘雅医院耳鼻喉科提供活检组织, 即刻液氮冻存.

(ⅱ) 总 RNA 提取及去 DNA 消化处理. 用 TRIzol (GIBCO BRL)抽提组织总 RNA. DNase Ⅰ消化去除痕量 gDNA.

(ⅲ) α-^{32}P 探针标记、纯化、杂交及结果分析. 参照产品说明书及文献[11]方法进行.

(ⅳ) RT-PCR 证实差异表达的 EST N27741. 用以上两种组织所提取的总 RNA, 分别用 Super Script™ 逆转录系统合成第 1 链 cDNA. 根据该 EST 序列设计出 5′ 引物: ATG GGA TGT TAC ACA CGC CT 和 3′ 引物: CTC TGC TGA GCT GGG TCT TC, 同时以 *GAPDH* 基因引物为内对照, 在 PCR 仪上进行 28 个循环扩增, 取产物于 1%琼脂糖凝胶上电泳检测.

(ⅴ) cDNA 克隆的鉴定. 采用菌落 PCR 法鉴定. 在单克隆平板上取 8 ~ 10 个菌落, 溶于 500 μL 超纯水中, 再从中取出 10 μL 作为模板, 用与上述相同的 EST N27741 引物 PCR 扩增包含该 EST 序列的 cDNA 克隆 IMAGE 255754 (购自 Research Genetics 公司), 于 1%琼脂糖凝胶电泳检测.

(ⅵ) cDNA 克隆插入片段大小的检测. 将 cDNA 克隆 IMAGE 255754, 于含氨苄青霉素的 LB 培养基中扩增, 抽取质粒 DNA 后, 用 *Eco*R Ⅰ和 *Xho* Ⅰ两步法双酶切, 于 1%琼脂糖凝胶上电泳检测.

(ⅶ) 核酸序列测定及同源性比较. 将含有目的片段的 cDNA 阳性克隆转化大肠杆菌 TG1 后，在含有氨苄青霉素的 LB 培养基中扩增，抽取质粒 DNA 后于 API 377 测序仪上，逐段测序. 将得到的新序列于 NCBI 的核酸非冗余序列数据库(nr)中作 Blast 分析，以查询是否存在高度同源的基因.

(ⅷ) 编码氨基酸及二级结构预测. 在 Internet 网上，采用 http://pbil.ibcp.fr/cgi-bin/primanal_color.pl 分析软件分析该序列所编码的氨基酸序列；分子量、等电点分析网址为：http://www.expasy.ch/cgi-bin/protparam；采用以下网址预测此氨基酸序列二级结构、亲水性和疏水性特征：1) http://pbil.ibcp.fr/cgi-bin/secpred_sopm.pl; 2) http://www.expasy.ch/cgi-bin/protscale.pl; 3) http://azusa.proteome.bio.tuat.ac.jp/; 4)http://www.enzim.hu/hmmtop/.

2 结果

2.1 差异表达 EST N27741 的进一步鉴定

在采用 cDNA 微阵列对 NPC 及正常鼻咽组织的基因表达谱研究的基础上[11]，进而筛选出在 NPC 低表达而正常鼻咽高表达的 EST N27741. 为进一步证实该 EST 在组织中表达差异，以看家基因 *GAPDH* 为内对照的半定量 RT-PCR 法，检测了 14 例 NPC 和 2 例正常鼻咽组织，其中 12 例均为明显低表达或无表达，仅 2 表达与 GAPDH 相近或较高，而 2 例正常鼻咽组织中均明显较高表达，如图 1.

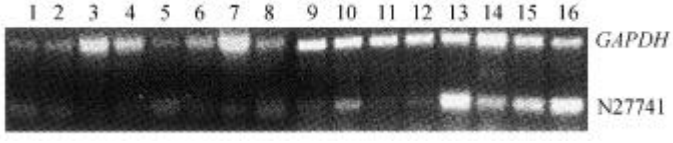

图 1 RT-PCR 检测 14 例 NPC 和 2 例 NP 组织中 EST N27741 的差异表达

1~14 为 14 例 NPC, 15, 16 为 2 例 NP

2.2 cDNA 克隆的鉴定及插入片段大小的检测

用菌落 PCR 法鉴定 12 个菌落，发现它们均存在相同的 PCR 扩增产物，说明该 cDNA 克隆为单克隆来源的阳性克隆. 再将包含 EST N27741 序列的 cDNA 克隆 IMAGE 255754，经 LB 培基氨苄抗性筛选扩增，抽取质粒 DNA，*Eco*RⅠ和 *Xho*Ⅰ两步法双酶切后得知该质粒的插入片段为 1.1 kb.

2.3 新基因 YH1 cDNA 序列测定

将 cDNA 克隆 IMAGE 255754 于氨苄抗性的 LB 培养基中扩增，抽取质粒 DNA，API 377 测序仪逐段测序得到序列如图 2.

我们将含有完整的阅读框架的该序列命名为 *YH1* 基因，其起始密码子位于 62 ~ 64 位核苷酸，终止密码子位于 829 ~ 831 位核苷酸，编码 256 个氨基酸，起始密码子前-27 至-25 位有一个终止密码子. 编码区后有 AATAAA 加尾信号及 poly A 尾，该新基因 *YH1* 在 GenBank 的收录号码为：AF158745.

2.4 编码氨基酸及二级结构预测

采用最新发展起来的生物信息学分析方法，在国际互联网上分析该序列所编码的氨基酸序列及其分子量、等电点等生物学特征如下：

```
GCACGAGGGA GACCAGGACA GCTGCTGAGA CCTCTAAGAA GTCCAGATAC   50
TAAGAGCAAA GATGTTTCAA ACTGGGGGCC TCATTGTCTT CTACGGGCTG  100
TTAGCCCAGA CCATGGCCCA GTTTGGAGGC CTGCCCGTGC CCCTGGACCA  150
GACCCTGCCC TTGAATGTGA ATCCAGCCCT GCCCTTGAGT CCCACAGGTC  200
TTGCAGGAAG CTTGACAAAT GCCCTCAGCA ATGGCCTGCT GTCTGGGGGC  250
CTGTTGGGCA TTCTGGAAAA CCTTCCGCTC CTGGACATCC TGAAGCCTGG  300
AGGAGGTACT TCTGGTGGCC TCCTTGGGGG ACTGCTTGGA AAAGTGACGT  350
CAGTGATTCC TGGCCTGAAC AACATCATTG ACATAAAGGT CACTGACCCC  400
CAGCTGCTGG AACTTGGCCT TGTGCAGAGC CCTGATGGCC ACCGTCTCTA  450
TGTCACCATC CCTCTCGGCA TAAAGCTCCA AGTGAATACG CCCCTGGTCG  500
GTGCAAGTCT GTTGAGGCTG GCTGTGAAGC TGGACATCAC TGCAGAAATC  550
TTAGCTGTGA GAGATAAGCA GGAGAGGATC CACCTGGTCC TTGGTGACTG  600
CACCCATTCC CCTGGAAGCC TGCAAATTTC TCTGCTTGAT GGACTTGGCC  650
CCCTCCCCAT TCAAGGTCTT CTGGACAGCC TCACAGGGAT CTTGAATAAA  700
GTCCTGCCTG AGTTGGTTCA GGGCAACGTG TGCCCTCTGG TCAATGAGGT  750
TCTCAGAGGC TTGGACATCA CCCTGGTGCA TGACATTGTT AACATGCTGA  800
TCCACGGACT ACAGTTTGTC ATCAAGGTCT AAGCCTTCCA GGAAGGGGCT  850
GGCCTCTGCT GAGCTGGGTC TTCCCCCAAC AGAACTATTT CTTGCTGCTC  900
AATCCATTCC CTCTGGCCCA GCTTCCCAGT GCTCACAGAT GGCTGGCCCA  950
TGTGCTGGAA GATGACACAG TTGCCTTCTC TCCGAGGAAC CTGCCCCCTC 1000
TCCTTTCCCA CCAGGCGTGT GTAACATCCC ATGTGCCTCA CCTAATAAAA 1050
TGGCTCTTCT TATGCAAAAA AAAAAAAAAA AAAACTCGAG GGGGGG     1096
```

图 2 *YH1* **基因全长** cDNA **序列**

```
M F Q T G G L I V F Y G L L A Q T M A Q F G G L P V P L D Q T L P L N V N P A L P L S P T
G L A G S L T N A L S N G L L S G G L L G I L E N L P L L D I L K P G G G T S G G L L G G
L L G K V T S V I P G L N N I I D I K V T D P Q L L E L G L V Q S P D G H R L Y V T I P L G
I K L Q V N T P L V G A S L L R L A V K L D I T A E I L A V R D K Q E R I H L V L G D C T
H S P G S L Q I S L L D G L G P L P I Q G L L D S L T G I L N K V L P E L V Q G N V C P L V
N E V L R G L D I T L V H D I V N M L I H G L Q F V I K V
```

分子量为 26.7 ku, **等电点为** 5.42, **采用自身优化预测法**(self-optimized prediction method, SOPM)[12]**预测此序列二级结构表明**, **该序列中的** 33.2%**的氨基酸组成**α-**螺旋结构**(alpha helix), **而在近羧基端存在** 3 **个该结构**, 21.48%**的氨基酸组成延伸主链**(extended strand), 36.72%**的氨基酸组成无规卷曲**(random coil), 8.59%**氨基酸组成**β-**转角**(beta turn), **如图**3. **再分别采用** SOPMA[13], PHD[14] **及** MLRC[15]**等多种神经网络运算预测法所得到的结果与此相似**. **综合得到的结果是**: 30.08%**的氨基酸组成** α-**螺旋结构**, 18.36%**的氨基酸组成延伸主链**, 45.70%**的氨基酸组成无规卷曲**, **而** 5.86%**氨**

基酸处于模糊状态(ambiguous states). 除以上网址外, 还用 http:// www.biokemi.su.se/server/DAS/ 及 http://www.enzim.hu/hmmtop/分析亲水性和疏水性特征, 发现在氨基端的第 4 位至 24 位的 21 个氨基酸, 构成了一个α-超螺旋结构, 疏水指数为 0.608 的可溶性蛋白质; 用 Kyte J 法[16]预测在氨基端同样是一个疏水指数为 2.555 的疏水结构域, 可能具有分泌型的特征, 如图 4.

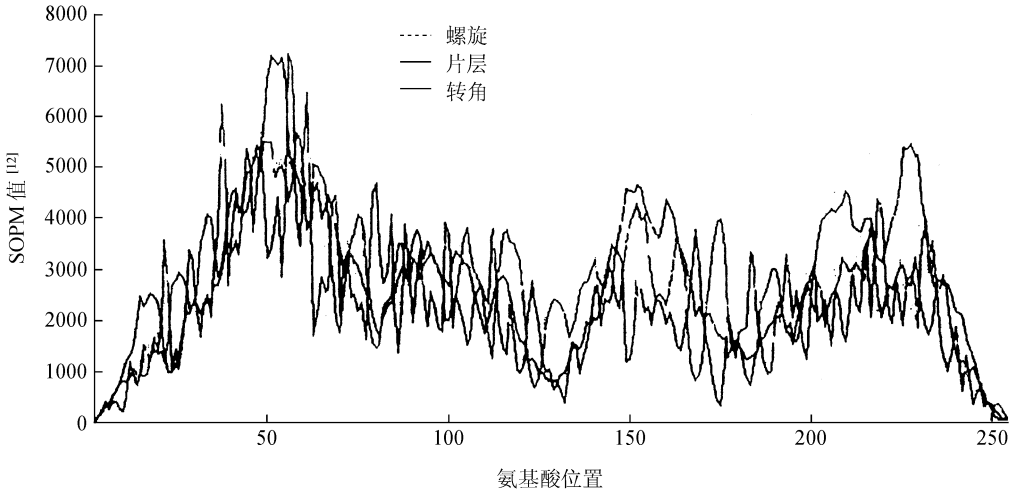

图 3 新基因 *YH1* 所编码氨基酸的二级结构

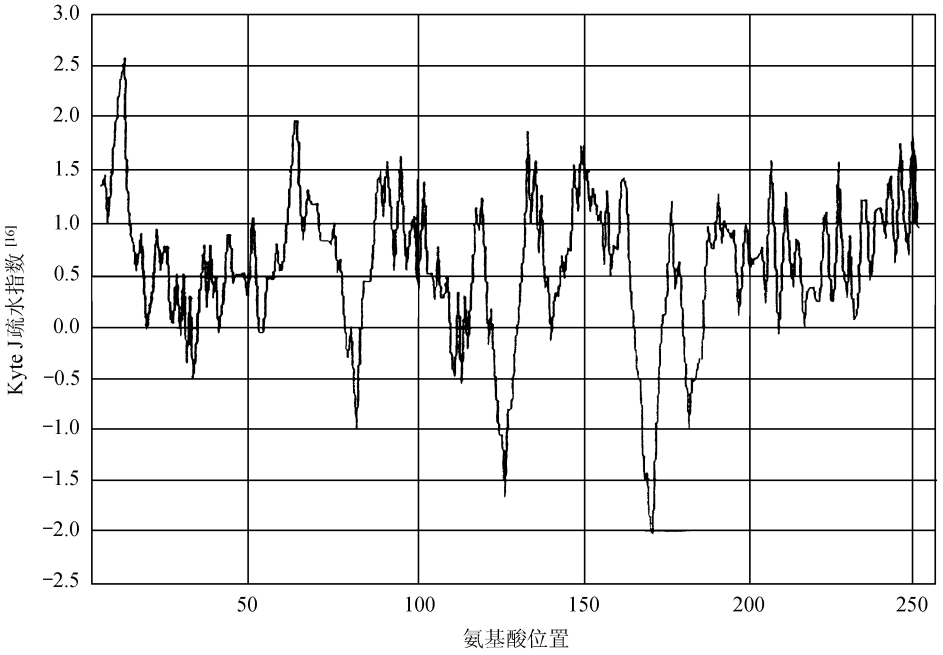

图 4 新基因 *YH1* 所编码氨基酸的亲、疏水性特征

3 讨论

高密度 cDNA 表达阵列已成功地应用于人类多种疾病的基因表达谱的研究, 并获得相应

基因的表达谱及发现了新的起作用的疾病相关基因[7~10]. 我们采用含有 5 184 个点阵的高密度 cDNA 表达阵列, 比较研究了成人鼻咽组织和鼻咽低分化鳞状细胞癌组织基因的表达差异[11]. 从这 5184 个点阵代表不同的表达序列标签 EST (expression sequence tag)或基因中, 筛选出在鼻咽组织高表达而在鼻咽癌组织低表达的 EST N27741, 进而用 RT-PCR 证实与杂交结果相符. 在 14 例 NPC 组织中, 低或无表达的占绝大多数(12 例), 仅 2 例表达较高, 这可能与组织取材或肿瘤状态有关, 而在 2 例 NP 组织中均有明显高表达, 这表明该 EST 确实在 NPC 中较 NP 低表达. 在 IMAGE cDNA 文库中查找到包含有该 EST 序列的 cDNA 克隆, 用该 EST 所设计的引物进行 PCR 扩增加以鉴定, 用双酶切及测序发现该 cDNA 序列长度在 1.1 kb, 含完整的阅读框及基因的序列特征, 在 NCBI 的核酸非冗余序列数据库中比较, 未发现与之高度同源的基因, 为新的鼻咽癌相关基因, GenBank 收录号为 AF158745.

采用生物信息学分析方法, 对该基因所编码氨基酸序列进行相关分析, 如用 SOPM, SOPMA, PHD 及 MLRC 等多个网点多种方法对该基因编码氨基酸的二级结构进行预测所得到的结果相似. 该基因是共编码 256 个氨基酸、分子量为 26.7 ku、等电点为 5.42 的酸性蛋白质, 所含有的多个α-螺旋主要在羧基端. 其中约 30%的氨基酸组成α-螺旋结构, 18%的氨基酸组成延伸主链, 45%的氨基酸组成无规卷曲. 在氨基端有一超螺旋结构组成的疏水区, 为一跨膜可溶性蛋白质. 目前采用这些神经网络运算法对氨基酸的二级结构进行预测, 大大提高了预测的精确度. 利用该基因的编码序列稳定转化低分化人鼻咽癌细胞系, 发现细胞生长增殖受到明显抑制, 而目前正在进行克隆形成率及裸鼠体内致瘤作用等功能研究, 以进一步了解该基因在鼻咽癌形成中的作用.

致谢 本工作为国家自然科学基金(批准号: 39900174)及重点基金(批准号: 39730200)和中华医学会基金(批准号: 96-655)资助项目.

参 考 文 献

1 游绍进, 姚开泰, 曹 亚, 等. Epstein-Barr 病毒潜伏状态与鼻咽癌的关系. 中华肿瘤杂志, 1996, 18(1): 23 ~ 26

2 Yeung W M, Zong Y S, Chiu C T, et al. Epstein-Barr virus carriage by nasopharyngeal carcinoma *in situ*. Int J Cancer, 1993, 53: 746 ~ 751

3 Sun Y. Molecular oncology of human nasopharyngeal carcinoma. The Cancer Journal, 1995, 8(6): 325 ~ 331

4 Lo K W, Tsao S, Leung S, et al. Detailed deletion mapping on the short arm of chromosome 3 in nasopharyngeal carcinoma. Intern J Oncol, 1994, 4: 1359 ~ 1364

5 Huang D P, Lo K W, Choi P H, et al. Loss of heterozygosity on the short arm of chromosome in nasopharyngeal carcinoma. Cancer Genet Cytogenet, 1991, 54: 91 ~ 99

6 Huang D P, Lo K W, van Hasselt C A, et al. A region of homozygous deletion on chromosome 9p21-22 in primary nasopharyngeal carcinoma. Cancer Res, 1994, 54: 4003 ~ 4006

7 Zhang L, Zhou W, Velculescu V E, et al. Gene expression profiles in normal and cancer cells. Science, 1997, 273: 1268 ~ 1272

8 Saizieu A D, Certa U, Warrington J, et al. Bacterial transcript imaging by hybridization of total RNA to oligonucleotide arrays. Nature Biotechnology, 1998, 16: 45 ~ 49

9 Schena M, Shalon D, Heller R, et al. Parallel human genome analysis: microarray-based experssion monitoring of 1000 genes. Proc Natl Acad Sci USA, 1996, 93: 10614 ~ 10619

10 Heller R A, Schena M, Chai A, et al. Discovery and analysis of inflammatory disease-releated genes using cDNA

microarrays. Proc Natl Acad Sci USA, 1997, 94: 2150 ~ 2156

11 **何志巍, 任彩萍, 许亮国, 等. 用微阵列研究鼻咽癌与正常鼻咽组织基因表达谱. 科学通报**, 1999, 44(23): 2528 ~ 2531

12 Geourjon C, Deleage G. A self-optimized method for protein secondary structure prediction. Protein Eng, 1994, 7(2): 157 ~ 160

13 Geourjon C, Deleage G. Significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. Comput Appl Biosci, 1995, 11(6): 681 ~ 684

14 Rost B, Sander C. Prediction of protein secondary structure at better than 70% accuracy. J Mol Biol, 1993 , 232(2): 584 ~ 599

15 Guermeur Y, Geourjon C, Gallinari P, et al. Improved performance in protein secondary structure prediction by inhomogeneous score combination. Bioinformatics, 1999, 15(5): 413 ~ 421

16 Kyte J, Doolittle R F. Amino acid scale: hydropathicity. J Mol Biol, 1982, 157: 105 ~ 132

(1999-11-09 **收稿**, 2000-03-17 **收修改稿**)