

胃癌相关 cDNA 片段的快速克隆和表达分析*

李红¹⁾** 王孟薇¹⁾ 王刚石¹⁾ 陈润生²⁾ 凌伦奖²⁾ 王金华²⁾

(¹⁾解放军总医院消化科, 北京 100853; (²⁾中国科学院生物物理研究所, 北京 100101)

摘要 利用差异显示 PCR 技术获得的一条在胃癌和正常组织有差异表达的 EST (W123 (GenBank 登录号为 AF150631)), 通过与 GenBank 的 dbest 库进行电子杂交, 选取了与其同源度高的若干 EST, 在它们共有的保守序列设计了用于扩增的寡聚核苷酸引物, 利用 cDNA 末端快速扩增 PCR (RACE) 技术得到了 7 条带有 polyA 尾的 3'EST, 进行序列分析后, 发现它们均是代表新基因或不同剪接体的 EST, 且具有共同的保守序列, 已登录 GenBank. 采用 RNA 印迹对目的序列进行初步鉴定, 并进行了这些基因的组织分布分析. RACE 技术和生物信息学相结合, 具有快速、高效的特点, 有助于疾病相关基因的克隆.

关键词 cDNA 末端快速扩增 (RACE), 生物信息学, 胃癌, 克隆, cDNA 片段, 表达分析

学科分类号 Q-33, R735.2

随着人类基因组计划、蛋白质组计划和生物芯片技术开展以来, 有关研究成果和数据迅速增长, 如何挖掘 (data mining) 整理和应用开发这些浩如烟海的生物信息, 在当前形势下变得尤为迫切^[1], 一门新兴的生命科学——生物信息学应运而生, 高通量及发展迅速等特点, 使其在新基因的克隆及功能研究方面占有越来越重要的地位^[2], 特别是表达序列标签 (EST) 数据库的迅速扩张, 已导致识别与克隆新基因的策略发生了革命性的变化^[3]. 传统克隆基因的方法已不能满足当前高速发展的生物经济时代, 广泛的 DNA 和蛋白质数据库的创建为新基因的克隆提供了可利用资源, 而 RACE 技术和生物信息学相结合, 具有快速、高效克隆基因的特点, 从胃癌组织出发, 利用该方法快速钓取了胃癌相关多个 cDNA 片段.

1 材料和方法

1.1 标本及总 RNA 的提取

人胃癌癌旁组织, 取自我院外科新鲜手术标本, 病理证实为胃窦部腺癌, 中、低度分化. Trizol 一步法提取人胃癌癌旁组织总 RNA.

1.2 W123 同源 EST 的电子信息杂交与序列组装

以 W123 cDNA 片段为探针, 用 NCBI 提供的 BLASTN (<http://www.ncbi.nlm.nih.gov/blast/>) 程序对 GenBank 的人类 EST 数据库进行电子信息杂交筛选, 选择与 W123 同源度较高 (90% 以上同源) 的 9 条 EST (图 1), 用 DNASTAR 软件将它们进行序列组装, 整合成几条更长的 EST 序列, 在

它们共有的保守序列设计了 2 条用于 3'RACE 扩增的寡聚核苷酸引物 (3'GSP2、3'NGSP2, 图 2).

1.3 3'-RACE

SmartTM RACE cDNA Amplification kit 购自 Clontech 公司, 按照说明书进行, 并进行了两点改进: a. 反转录反应加入 40 U RNasin 抑制剂 (Promega); b. 除按说明书设阳性对照外, 又设立了以水为模板反转录的阴性对照. 取总 RNA 1 μ g, 锚定引物 3'CDS (序列为 5'-AAGCAGTGGTAA-CGCAGAGTAC (T)₃₀-3', 10 μ mol/L), 加反转录酶 SUPERSRIPT II (200 U) 于 42°C 孵育 1.5 h 以合成 cDNA. 用 3'GSP2 (其序列为: 5'-TACAT-GGAAACTGAACAACCTGC-3', 10 μ mol/L) 与 UPM 通用引物 (该引物由两条序列组成: 其长序列为 5'-CTAATACGACTCACTATAGGGCAAGCA-GTGGTAACGCAGAGT-3', 0.2 μ mol/L; 其短序列为 5'-CTAATACGACTCACTATAGGGC -3', 1 μ mol/L) 进行第一次扩增. 反应条件分为: 94°C 1 min, 64°C 1 min, 72°C 1.5 min, 33 个循环. 将第一次产物稀释后作为模板用 3'NGSP2 (其序列为: 5'-TCACTAAATGCCCCACAAGAGA-3', 10 μ mol/L) 与巢式引物 NUP (其序列为: 5'-AAGCAGT-GGTAACAACGCAGAGT-3', 10 μ mol/L) 进行第二次扩增, 条件为: 94°C 1min, 64°C 1 min, 72°C 1.5 min, 20 个循环.

* “九五”军队科研基金项目 (96M145).

** 通讯联系人. 解放军总医院南八科, 北京 100853.

Tel: 010-66936747, E-mail: lihongam@hotmail.com

收稿日期: 2001-12-07, 接受日期: 2002-01-31

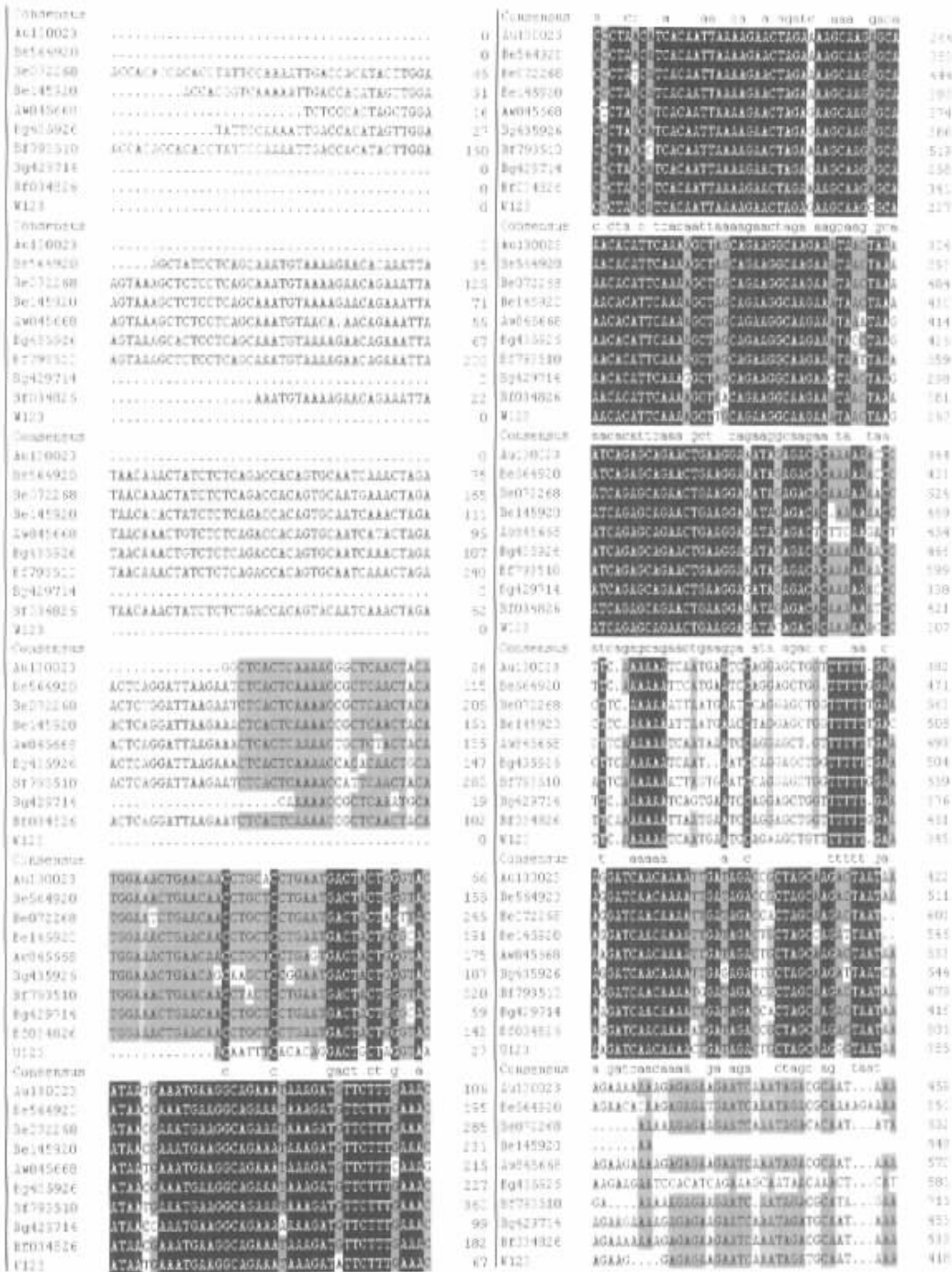


Fig.1 Alignment of nucleotide sequences of a few homologous EST sequences with W123

Identical residues are marked by shading. Just high homologous region is shown in this figure. Numbers of left side are Genebank accessions of ESTs. Numbers of right side are basepairs.

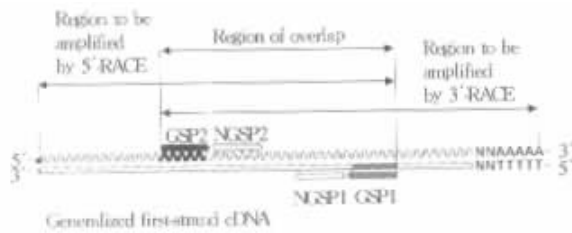


Fig.2 The schematic diagram of RACE

This diagram shows a generalized first-strand cDNA template (contains polyA). region to be amplified by 3' RACE with the gene-specific primers (3'GSP2, 3'NGSP2) and UPM, NUP.

1.4 3'RACE 产物的克隆、测序

将 3'-RACE 所扩增的目的片段以 Wizard PCR 产物纯化试剂盒 (购自 Promega 公司) 进行纯化、回收 , 然后将其克隆到 pGEMT-easy 质粒载体中 (购自 Promega 公司) , 再转染感受态 JM109 菌株 . 经蓝白斑筛选后 , 挑取多于 10 个白色单菌落用载体引物 T₇、SP₆ 进行 PCR 鉴定 , 随机选取 1 个单菌落测序并选取 10 个克隆混合测序 .

1.5 目的序列的生物信息学分析、登录

将测序所得带有 polyA 尾的 7 条 cDNA 片段用 NCBI 提供的 BLASTN 与 GenBank 的 dbEST、nr 数据库进行同源性比较 , 确认代表新基因的 EST 并登录 GenBank (<http://www.ncbi.nlm.nih.gov/submit>) .

1.6 目的序列的 RNA 印迹鉴定

组织标本取自解放军总医院新鲜手术标本 , 病理证实为胃窦部进展期中低分化腺癌 , 分别提取胃癌和其远端正常胃组织总 RNA , 具体方法同 1.1. RNA 转膜与固定 , 在酶切回收克隆插入片段的基础上 , 利用 Promega 公司的随机引物 DNA 标记试剂盒制备同位素标记的探针 , 利用 Clontech 公司的 Extrahybridat 快速杂交液进行预杂交、杂交、洗膜、曝光及显影 .

1.7 目的序列的组织分布

利用互联网癌肿基因组解剖学计划提供的基因表达分析系列 (series analysis of gene expression, SAGE) 数据库进行组织分布分析 . SAGE 表达谱分析数据库 (CGAP database SAGE map—<http://www.ncbi.nlm.nih.gov/SAGE>) 和软件 (Software :virtual northern—<http://www.ncbi.nlm.nih.gov/sagevn.cgi>) 可通过在组织、细胞基因表达文库中钓取基因代表性 tag (一段小于 9 bp 的 DNA 短序列) , 对特定基因的表达以及其表达水平进行分析 , 从而确定其组织分布 .

2 结 果

2.1 3'RACE 产物的序列分析

从上述组装的共有序列约 400 bp 出发 , 设计 3'-GSP 进行 RACE , 共得到 7 条大小不等的片段 , 分别命名为 W13 , W25 , W31 , W41 , W51 , W63 , W74 ; 大小从 195 bp 至 1 200 bp 不等 (图 3) , PCR 扩增重组质粒 , 目的片段大小正确 (图 4) , 对这些序列进行分析 , 它们均是带有 polyA 尾的 cDNA 片段 , polyA 尾 20~30“ A ”不等 , 有的片段可见加

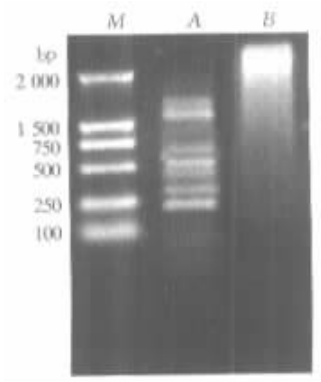


Fig.3 3'-RACE amplification of human gastric cancer paratumor tissue cDNA

M : marker DL 2000 ; A : size distribution of 3' RACE gene-specific products with NUP and 3'GSP2 ; B : size distribution of 3' RACE products amplified with UPM and 3'GSP1 .

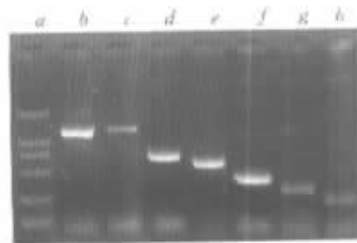


Fig.4 Amplification of recombinant plasmid with T₇ and SP₆ primer

a : marker DL 2000 ; b ~ h : the products of recombinant plasmid .

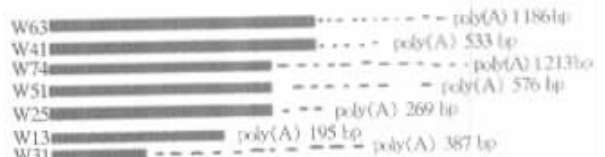


Fig.5 Comparison of similarity from sequence W13 to W74 produced by 3'-RACE

The solid rectangle represent the sequence similarity more than 90% bases . They all have 80 bp common sequence . Among W25 , W51 , W74 , W41 and W63 have 250 common bases . And between W4 and W63 have 310 bp .

尾信号 AATAAA，将这些序列在计算机软件上进行比对 (align) 发现：它们除靠近 5' 端的共有序列外，在 3' 端明显不同 (图 5)，且这些序列不能互相囊括，提示它们可能是不同基因的 3' 端。

2.2 7 条 cDNA 片段核苷酸序列的同源性比较

在 7 个测序克隆中，与 GenBank nr database 进行同源性比较显示，与基因编码序列同源性均小于 40%，因而相应的基因和功能均不清楚，可能代表新的未知基因。

2.3 目的片段在 GenBank 登录结果

7 条 3' EST 分别于 2000-12-10 至 2001-1-7 呈递 GenBank，现已登录，登录号为：AF332603、AF332604、AF332605、AF325202、AF332606、AY032615、AY032616。

2.4 RNA 印迹分析结果

RNA 印迹 (图 6) 显示 7 条 cDNA 中有 6 条在胃癌及正常对应组织表达，其中 W25 探针未见阳性结果。W31、W41、W63 号探针可见 3 个转录本，大小从 1.2~6.0 kb 不等，W13 有 2 个转录本，大小为 2.0 kb、1.4 kb，W51 号片段有一个转录本，约 3.5 kb，W74 号探针显影转录本主要集中在 3.0~5.0 kb 之间。用 Image J 软件扫描各转录本灰度，用公式 $A_{\text{mean}} = \lg_{10} (255 / \text{Mean grey})$ 计算强度，以 A_{mean} 作为量化指标，分析各转录本的表达丰度。其中，W13、W31、W41、W51、W74 在胃癌组织的表达强度高于正常组织，W63 则与之相反，在胃癌组织下调 (图 7)。

2.5 基因电子表达谱分析结果

7 条 cDNAs 中，除 W25 未见阳性杂交结果外，其余 6 个探针均在多种肿瘤和/或正常 SAGE 文库中有不同程度表达，SAGE 文库总 tags 数见表 1。

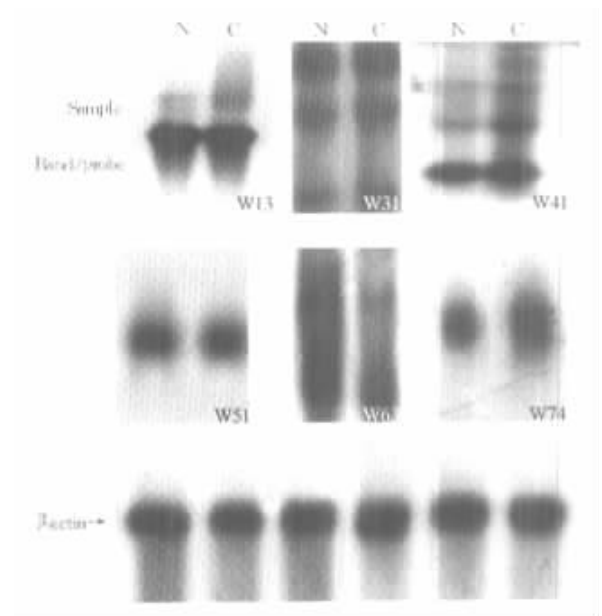


Fig.6 Verification of differential gene expression by Northern blot analysis using SMART RACE PCR-amplified cDNA probe

The northern blot contains normal (N) and tumor (C) stomach total RNA 30 μg per lane. Hybridization to a cDNA probe for β -actin was used to quantify variation in the amount of nucleic acid loaded.

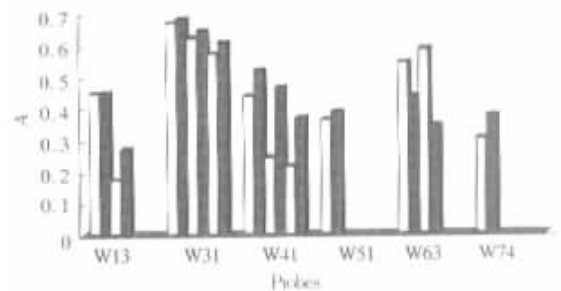


Fig.7 Relative amounts of cDNA probe expression

□: normal; ■: cancer.

Table 1 Total tags in different SAGE libraries for six cDNA probes

SAGE libraries Probe \ tag	prostate		colon		nerve		breast		ovary		pancreas	
	N	C	N	C	N	C	N	C	N	C	N	C
W13	15	82	0	35	32	37	26	0	0	0	0	0
W31	380	1 907	598	2 415	1 476	5 373	382	1 548	362	1 057	366	599
W41	0	0	0	17	0	95	0	83	20	44	0	29
W51	0	22	0	16	22	106	0	14	0	0	0	29
W63	76	152	19	142	10	289	148	96	0	23	61	95
W74	0	0	0	39	47	16	0	16	0	0	0	0

N: normal; C: cancer.

3 讨 论

虽然现在已经肯定了很多基因在胃癌中的异常变化,对胃癌的诊断和治疗有一定的参考意义,但它们最初发现时均非通过研究胃癌所得,所以从胃癌组织出发,加强胃粘膜癌变过程中基因型变化的研究,可以为明确胃癌的发生机制及早期诊断提供进一步线索.而生物学技术和信息学资源相结合,为快速基因克隆提供了可能^[4,5].在进行差异片段 W123 生物信息学分析时发现, W123 有若干个同源 ESTs,它们在 5'端序列一致性很好,蛋白质的编码序列可能有高度的保守性,但 3'端差异却很明显,这与多基因家族(multifungene family)的特点相吻合^[6,7],而且癌基因及抑癌基因多以基因家族的形式存在.根据生物信息学的提示,试图通过实验方法来抓住这一可能的基因家族同源候选成员,为基因的克隆摸索新的途径.

RACE 技术的主要优点是快速、灵敏,短时间内即可获得新基因的 cDNA 全长序列,而且 RACE 能产生大量独立克隆^[8],可用来证实核苷酸序列,如果一个基因是多基因家族的一个成员,用基因特异引物(GSP)可能同时扩增几个高度同源的 cDNAs^[9].本研究以发现的这一保守区域设计了钓取这一可能的家族保守性引物,在胃癌癌旁组织中进行 3'-RACE 扩增,共钓取了 7 条 cDNA 片段,经测序分析,它们均是带 polyA 尾的 3' EST,因 3' UTRs(3'untranslated regions)很少有内含子和如编码区内的保守序列,更能准确地代表基因^[10],经用 NCBI 提供的 BLASTN 对 GenBank 的 nr 数据库进行电子信息杂交,这 7 条均是代表新基因的 EST,现已在 GenBank 中登记注册.

为了对我们得到的 7 条候选同源基因与胃癌的关系作出进一步确证,使用 RNA 印迹分析了它们在胃癌组织及其远端正常组织的表达状况,6 条 cDNA 分别在胃癌及正常组织有不同程度的表达,其中 5 条 cDNA(W13、W31、W41、W51 和 W74)在胃癌组织中表达高于正常胃粘膜.为了更全面检测所克隆的基因在其他肿瘤的表达丰度及其功能进行预测,我们把克隆所得的 cDNA 放到了 CGAP(cancer genomic anatomy project)的大背景下,利用它的近 100 个 SAGE 文库,通过 virtual Northern 程序,来分析这些基因的表达谱.其中 W31、W41、W51 在所有肿瘤的 SAGE 文库中表达高于正常的文库,这些肿瘤文库来源于前列腺

癌、结肠癌、神经系统肿瘤、乳腺癌、卵巢癌和胰腺癌,结合它们在胃癌中的表达情况,提示它们很可能是在肿瘤的发生、发展过程中,起作用的相互协调共表达的基因簇.而 W63 在正常乳腺文库中有 148 个 tags,在乳腺癌文库中有 96 个 tags, W13 在前列腺癌、结肠癌中呈高表达,而在乳腺的正常组织呈高表达,结合它们在胃癌组织中的表达情况,它们在肿瘤形成中的作用,还有待于进一步研究、探讨.

利用 RACE 技术与开发 EST 库结合,可快速发现多个基因,在此基础上再寻找基因家族同源成员,是完善、归类同源基因家族新成员的一种颇具规模化的方法,且随着信息库容量和基因代表性片段的增长,这种功能将会越来越强大.这里,我们所作的只是将实验技术和信息学相结合的一种尝试,得到的结果也还需要进一步验证整理,但有一点可以肯定的是,我们正在摒弃以前单基因的思维模式,从基因组的整体水平出发,寻找结构或功能相似的基因簇,和它们之间的相互作用,把“基因集合(gene cluster)”作为一个复杂系统来一起考虑,所以把克隆的 W13 等几个 cDNA 片段联系起来进行研究,对搞清楚胃癌发生中的基因调控网络,将提供很好的理论思路和实践证明.

参 考 文 献

- 1 Allen J F. Bioinformatics and discovery: induction beckons again. *Bioessays*, 2001, **23**(1): 104~107
- 2 Board P, Tetlow N, Blackburn A, *et al.* Database analysis and gene discovery in pharmacogenetics. *Clin Chem Lab Med*, 2000, **38**(9): 863~867
- 3 Adams M D, Kelley J M, Gocayne J D, *et al.* Complementary DNA sequencing: expressed sequence tags and Human Genome Project. *Science*, 1991, **252**(5103): 1651~1656
- 4 Rossi D I, Vicari A P, Franz-Baccon K, *et al.* Identification through bioinformatics of two new macrophage proinflammatory human chemokines: MIP-3 alpha and MIP-3 beta. *J Immunol*, 1997, **158**(3): 1033~1036
- 5 Scohy S, Gabant P, Van Reeth, *et al.* Identification of KLF13 and KLF14, a novel members of the SP/XKLF transcription factor family. *Genomics*, 2000, **70**(1): 93~101
- 6 Koonin E V. Cell cycle and apoptosis: possible roles of Gadd 45 and MyD118 proteins inferred from their homology to ribosomal proteins. *J Mol Med*, 1997, **75**(4): 236~238
- 7 Hardiman G, Kastelein R A, Bazan J F. Isolation, characterization and chromosomal location of human WNT10B. *Cytogenet Cell Genet*, 1997, **77**(27): 278~282
- 8 Frohman M A, Dush M K, Martin G R. Rapid production of full-length cDNA from rare transcripts: amplification using a single gene-specific oligonucleotide primer. *Proc Natl Acad Sci USA*, 1988, **85**(23): 8998~9002
- 9 Dieffenbach C W, Dveksler G S. PCR Primer: A Laboratory Manual. New York: Cold Spring Harbor Laboratory Press, 1995.

8.61~8.63

10 Makalowski W , Zhang J , Boguski M S. Comparative analysis of

1196 orthologous mouse and human full-length mRNA and protein sequences. *Genome Res* , 1996 , 6 (9) : 846~857

Rapid Cloning and Tissue Expressive Analyses of cDNA Fragments in Human Gastric Cancer*

LI Hong^{1)**} , WANG Meng-Wei¹⁾ , WANG Gang-Shi¹⁾ ,
CHEN Run-Sheng²⁾ , LING Lun-Jiang²⁾ , WANG Jin-Hua²⁾

(¹⁾*The Digestive Department of The PLA General Hospital , Beijing 100853 , China ;*

(²⁾*Institute of Biophysics , The Chinese Academy of Sciences , Beijing 100101 , China)*

Abstract Previously , an EST (expression sequence tag) fragment , W123 (Genbank accession No. AF150631) differentially expressed between gastric cancer and normal tissues , was cloned using differential-display PCR technique. A few homologous EST sequences were captured when processing similarity search in human EST database. To search for additional W123 homologous genes in gastric tissue , a primer for 3'-RACE (rapid amplification of cDNA end) from highly conserved region among the above ESTs was designed and seven EST fragments with poly (A) tail were cloned. Compared with ESTs in GenBank , the seven EST fragments represented novel genes with a common sequence. Northern blot was applied to detect the expression of these genes between gastric cancer and normal tissues. The results suggested that a combination of bioinformatics and RACE technique is a rapid and effective method for seeking for disease related genes in specific tissue.

Key words rapid amplification of cDNA end (RACE) , bioinformatics , gastric cancer , clone , cDNA fragment , expressive analysis

* This work was supported by grants from The " Nine-fifth " Medical Research Foundation in PLA (96M145).

** Corresponding author. Tel : 86-10-66936747 , E-mail : lihongam@hotmail.com

Received : December 7 , 2001 Accepted : January 31 , 2002