

研究报告

生物信息学辅助定位及延伸辐射诱导
未知表达序列标签*

罗 瑛 隋建丽 铁 轶 周平坤** 孙志贤

(军事医学科学院放射医学研究所, 北京 100850)

摘要 研究辐射诱导的基因表达调控对于认识细胞对辐射损伤的应激反应有重要意义。在低剂量辐射诱导新基因 RIG1 表达序列标签 (expression sequence tag, EST) 片段的基础上, 通过非克隆 cDNA 文库和 RACE (rapid amplification of cDNA end) 技术获得了其 3' 末端。依据实验得到的这两段 EST 序列所提供的信息, 通过生物信息学分析将 RIG1 基因初步定位在 20 号染色体。对 20 号染色体 RIG1 区基因组序列进行外显子扫描, 发现预测的外显子正好与实验得到的 EST 相吻合。利用预测的外显子设计特异引物, 成功地克隆了 RIG1 基因全长序列。同时, 对 20 号染色体 RIG1 区的生物信息学分析表明, 在 RIG1 基因的上游存在启动子区, 从而确定了 RIG1 基因的基因组序列。因此, 通过生物信息学辅助设计实验, 快捷地定位及延伸了未知 EST 片段 RIG1, 基本完成了 RIG1 的全基因、基因组序列及染色体定位研究。

关键词 辐射诱导基因, RIG1, 外显子预测

学科分类号 Q691

辐射诱导的基因表达调控是细胞对辐射损伤应激反应的一个重要方面, 研究这些辐射诱导的功能基因, 对于认识细胞损伤修复的基因调控, 揭示 DNA 损伤修复相关疾病 (如放射病、癌变、衰老) 的发生、发展规律, 寻找防、诊、治的新途径有重要意义。

我们通过差异显示技术, 在 50cGy 照射诱导的正常人胚肺细胞中分离到一新的基因片段表达序列标签 (EST), 长 476 bp, 命名为辐射诱导基因 1 (radiation induced gene 1, RIG1), 已登录 GenBank^[1]。RNA 印迹分析表明, RIG1 基因仅在辐射诱导的细胞中表达, 因此不能通过筛选常规的 cDNA 文库获得全长。同时, 序列分析表明, RIG1 EST 在 GenBank 的 nr 库及 dbEST 库中均无同源序列, 是一个全新的 EST, 不能通过电子克隆的方法进行延伸。鉴于上述信息, 我们选择了 RACE 技术来获取 RIG1 的全基因, 通过改进 cDNA 合成技术建立了非克隆 cDNA 文库^[2]。理论上, 该库含有所有辐射诱导表达基因的 cDNA。以此文库为模板, 设计了巢式 RACE-PCR、生物素磁性分离、生物素标记探针杂交、再次磁性分离的克隆、纯化流程^[3]。应用这套流程成功地克隆到 RIG1 3' 端约 1 245 bp 的序

列, 测序结果表明, 该序列 5' 端包含 RIG1 EST 300 bp 片段, 与我们的设计完全一致; 3' 端具有明显的 polyA 加尾信号及编码区的终止密码子, 表明此序列确为 RIG1 3' 末端。

在此基础上, 依据实验得到的这两段 EST 序列所提供的信息, 通过生物信息学分析辅助设计实验, 快捷地定位及延伸了未知 EST 片段 RIG1。

1 材料与方法

主要软件: 核酸序列同源性分析: Blast2.0 (NCBI, USA); 寻找核酸序列开放阅读框架: ORF Finder (NCBI, USA); 基因预测: GENSCAN (Stanford University, USA), FGENE (Computational Genomics Group, CGG, The Sanger Center, UK)。序列匹配分析: ClustalX 1.8 (EMBL, Germany), Bestfit (Web GCG, WebANGINS, Australia)。转录元件分析: TSSG (Computational Genomics Group, CGG, The Sanger

* 国家自然科学基金资助项目 (39870214)。

** 通讯联系人。

Tel: 010-66931217, E-mail: zhoupk@nic.bmi.ac.cn

收稿日期: 2000-02-28, 接受日期: 2000-03-31

Center, UK), TSSW (Computational Genomics Group, CGG, The Sanger Center, UK)

2 结果与讨论

2.1 同源性分析初步定位 RIG1 基因

2.1.1 RIG 476 bp 片段对 GenBank 染色体库进行同源序列检索结果 (图 1).

2.1.2 RIG1 245 bp 片段对 GenBank 染色体库进行同源序列检索结果 (图 2).

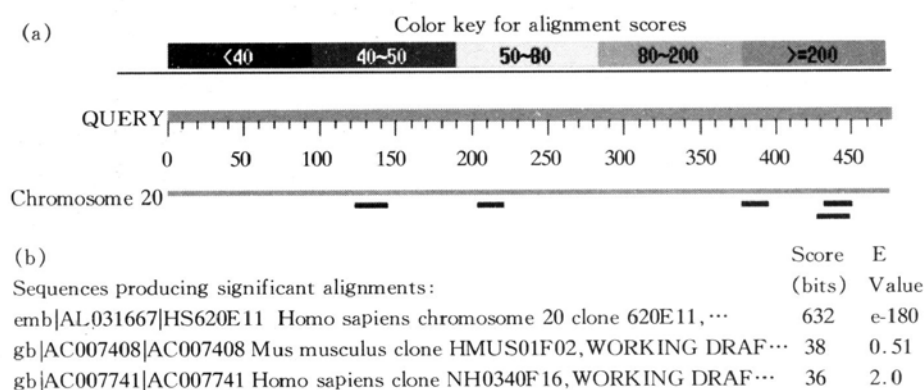


Fig.1 Homologous comparison of 476 bp fragment of RIG1 cDNA against Genbank database

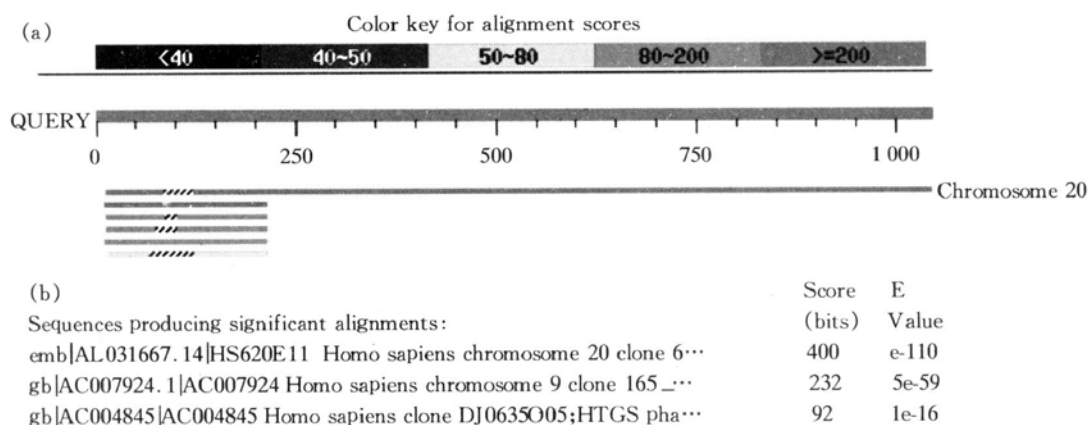


Fig.2 Homologous comparison of 3' part 1 245 bp fragment of RIG1 cDNA against Genbank database

上述结果表明, 原初的 RIG1 476 bp 片段及 RACE 得到的 RIG1 3' 1 245 bp 片段均与 GenBank 的染色体库中 20 号染色体的一个 BAC 克隆 620E11 高度同源, 表明 RIG1 很可能定位在 20 号染色体.

2.2 基因组序列的外显子预测

利用这一信息, 结合生物信息学辅助手段^[4], 我们对 20 号染色体 RIG1 区基因组序列进行外显子扫描, 其中一组预测的外显子如下 (图 3):

Gn.Ex	Type	S	.Begin	...End	.Len	Fr	Ph	I/Ac	Do/T	CodRg	P...	Tscr...
1.01	Intr	+	4458	4698	241	2	1	59	41	139	0.216	2.50
1.02	Intr	+	6197	6543	347	0	2	50	9	238	0.370	6.29
1.03	Intr	+	7377	7568	192	2	0	94	56	64	0.834	2.57
★1.04	Intr	+	8744	8908	165	1	0	108	84	295	0.997	30.34
1.05	Intr	+	10193	10372	180	1	0	41	87	331	0.988	27.44
1.06	Intr	+	11436	11546	111	2	0	100	87	94	0.999	10.16
●1.07	Intr	+	13230	13432	203	2	2	82	55	258	0.778	19.06
●1.08	Intr	+	14147	14346	200	2	2	63	94	189	0.963	15.07
1.09	Intr	+	15067	15226	160	2	1	110	89	119	0.992	12.32
1.10	Intr	+	18201	18290	90	0	0	58	103	119	0.985	8.59
★1.11	Intr	+	20404	20507	104	1	2	112	66	137	0.944	12.80

Fig.3 Prediction of exons of RIG1 gene located on chromosome 20

分析图 3 所列出的外显子, 发现其中预测的第七、第八外显子 (1.07、1.08, ●号标出) 正好是 RIG1 476 bp EST 与染色体序列匹配的位置, 而 1.09~ 1.11 外显子 (3' 端外显子) 也与 RIG1 1 245 bp 片段与染色体匹配的位置相吻合. 表明该外显子预测结果比较可靠, 可用于设计特异引物, 克隆 RIG1 476 bp EST 的 5' 端未知片段.

2.3 利用预测外显子克隆 RIG1 基因中段序列

在确定了上述外显子预测可信的基础上, 在靠近 RIG1 基因 3'、5' 端选取两个 $P > 0.9$ 的外显子

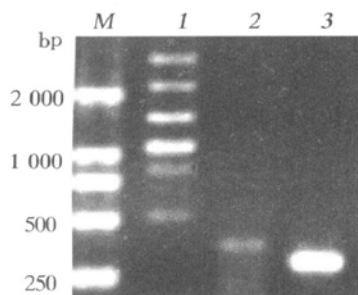


Fig.4 Amplification of medium part of RIG1 cDNA using the predicted or specific primers

M: molecular markers; 1: the PCR product produced by using the non-cloned cDNA library as template and the primers annealing with the predicted 3' and 5' sequences of RIG1; 2, 3: the PCR product produced by using the fourth band cDNA of 1 as template and the specific primers of RIG1 cDNA.

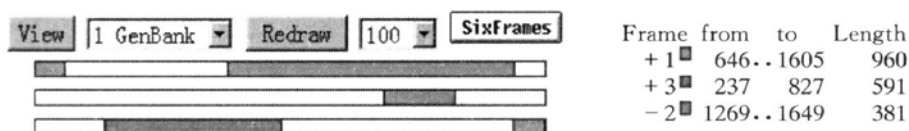


Fig.5 ORF prediction of RIG1 gene

综上所述, 我们在实验获得 RIG1 EST 片段的基础上, 通过生物信息学分析获得大量预测信息, 依据预测信息辅助设计实验, 较为简便、快捷地完成了 RIG1 的全基因、基因组序列及染色体定位研究. 研究表明, RIG1 基因组序列长约 38 kb, 其中包含 11 个外显子, 表达为 2 196 bp 的基因. 分析表明, 该基因的启动子较弱, 可能导致基因的低表达, 这与该基因仅在辐射诱导后细胞中被检测到的现象是一致的. 另外, 具有多个外显子和弱启动子的基因可能有较复杂的基因表达调控机制, 对 RIG1 基因表达调控元件研究具有较好的前景. 对 RIG1 基因的结构研究, 为下一步研究其在细胞辐

(图 3 中 1.04、1.11, ★号标出), 依据它们提供的序列设计特异引物, 使我们成功地扩增到 RIG1 基因中段约 1.2 kb 序列 (图 4).

将此片段克隆、测序. 分析结果表明, 该序列同样与 20 号染色体的 620E11 克隆高度同源. 同时, 该序列与 RIG1 476 bp、1 245 bp 序列匹配很好.

2.4 RIG1 全基因及基因组序列的获取及分析

RIG1 基因中段序列的成功克隆在实验上验证了生物信息学分析辅助设计的可行性. 因此, 基因预测与实验得到的 EST 相互印证, 将 RIG1 基因定位在人 20q11.2~ 12 区.

进一步对基因组序列进行分析, 在 RIG1 基因中段序列的上游发现了启动子区, 使我们确定了 RIG1 基因的全基因组序列.

启动子区的确定与外显子预测共同分析, 发现了 RIG1 基因第一个外显子, 并由此成功地扩增到 RIG1 基因 5' 端片段. 至此, 基本确定了 RIG1 基因的全长 cDNA 序列, 长约 2 196 bp, 这与早期 RNA 印迹分析中确定 RIG1 基因约为 2 kb 的结果相吻合^[1]. 对 RIG1 转录子进行六种可能读框的查找, 发现在 +1、+3 读框分别有长 960 bp, 编码 320 氨基酸及长 591 bp, 编码 197 氨基酸的完整开放阅读框架 (图 5), 哪一个自然界真实存在的编码框架, 有待于进一步的研究确定.

射应激反应中的功能及表达调控模式奠定了基础.

参 考 文 献

- 周平坤, 孙国敬, 隋建丽等. 一种快速 mRNA 差异显示方法及应用于分离辐射诱导转录子. 生物化学杂志, 1997, 13 (6): 622~ 625
Zhou P K, Sun G J, Sui J L, et al. Chinese Biochemical Journal, 1997, 13 (6): 622~ 625
- 罗 琰, 隋建丽, 铁 铁, 等. 自建非克隆 cDNA 文库从 EST 片段快速克隆全长 cDNA. 军事医学科学院院刊, 1999, 23 (3): 167~ 167
Luo Y, Sui J L, Tie Y, et al. Bull Acad Mil Med Sci, 1999, 23 (3): 167~ 167
- Chenchik A, Zhu Y, Diatchenko L, et al. Generation and use of high quality cDNA from small amounts of total RNA by SMART

PCR. Siebert P, Larrick J, Eds. In: RT-PCR Methods for Gene Cloning and Analysis. MA: Bio Techniques Books, 1998. 305~319

4 Burg C, Karlin S. Prediction of complete gene structures in human genomic DNA. J Mol Biol, 1997, **268** (1): 78~ 81

Chromosome Location and Elongation of Radiation-induced Expressed Sequence Tag by the Aid of Bioinformatics^{*}

LUO Ying, SUI Jian-Li, TIE Yi, ZHOU Ping-Kun^{**}, SUN Zhi-Xian

(Institute of Radiation Medicine, Beijing 100850, China)

Abstract Regulation of gene expression is one of the most important responses of cells to DNA damage induced by radiation. A novel expressed sequence tag (EST) fragment had been cloned from human embryo lung cells induced by 50cGy radiation and named RIG1. To clone the full-length cDNA of RIG1, a non-cloned cDNA library of human embryo lung cells induced by low dose irradiation had been established. This library was used as template in enhanced nest RACE PCR and biotin-labeled probe was used for further purification. The 3' flanking sequence of this EST was cloned and sequenced with this set of technology. It was illuminated by homology analysis that this 3' flanking sequence and the original EST are well aligned with a BAC clone of 20th chromosome and the predicted exons' sequence of this chromosome is well consistence with the real EST. Thus the RIG1 can be roughly located in 20th chromosome. By use of the exons' sequence predicted from chromosome sequence by GENSCAN, full-length of RIG1 gene has been cloned. Chromosome location of RIG1 gene is further determined by this successful verification of Bioinformatics prediction by experiment. By the same step, genome sequence of RIG1 has been determined. Therefore, by the combined use of Bioinformatics analysis, the full-length cDNA sequence and genome sequence of RIG1 gene are obtained and the predicted protein sequence is determined.

Key words radiation induced gene, RIG1, exon prediction

^{*} This work was supported by a grant from National Nature Science Foundation of China (39870214).

^{**} Corresponding author. Tel: 86-10-66931217, E-mail: zhoupk@nic.bmi.ac.cn

Received: February 2, 2000 Accepted: March 31, 2000