

一条 KRAB 型锌指蛋白全长新基因的克隆与功能初探

王兆应 康林盛榕 吴海 谢毅 毛裕民

(复旦大学生命科学院遗传学研究所, 遗传工程国家重点实验室, 上海 200433)

摘要: 从人胎脑 cDNA 文库中克隆到一条全长的锌指蛋白新基因的 cDNA, 命名为 *ZNF303*。序列分析表明 *ZNF303* 的 C 末端含有 7 个保守的锌指基序, N 末端含有一个 KRAB (Krüppel-associated box) 结构域。利用肝癌组织表达谱基因芯片杂交证明, 该基因在肝癌组织中的表达量有明显降低。认为这种降低可能跟肝癌的形成和转移有密切的关系。利用辐射杂交基因定位技术, 得出该基因在人类染色体上的位置是 19q13.2。

关键词: 锌指蛋白; KRAB; 肝癌

中图分类号: Q751

文献标识码: A

文章编号: 0253-9772(2000)04-0247-05

Molecular Cloning and Characterization of a Novel Full-length cDNA Encoding Zinc Finger Protein of KRAB Type

WANG Zhao, YING Kang, LIN Sheng-rong, WU Hai, XIE Yi, MAO Yu-min

(Institute of Genetics, Fudan University, Shanghai 200433, China)

Abstract: We have cloned a full-length novel zinc finger cDNA of the Krüppel family from human fetal brain cDNA library. Sequence analysis indicates that *ZNF303* contains 7 highly conserved zinc finger motifs at the C-terminus and a KRAB (Krüppel-associated box) domain at the N-terminus of the deduced protein. Hybridization using gene chip of hepatic cancer tissue demonstrates the expressive amount in hepatic cancer tissue is lower than control. We hypothesize that the decreasing of expression amount is related to the formation and metastasis of hepatic cancer. Finally, we show that *ZNF303* maps on human chromosome 19q13.2 by radiation hybrid.

Key words: zinc finger protein; KRAB; hepatic cancer

DNA 结合蛋白在生物体内有着重要的功能, 在染色体复制、转录、基因激活和 DNA 修复中发挥作用^[1]。其中具备转录作用的 DNA 结合蛋白称为转录因子, 它们往往具有结合 DNA 的特定的基序, 如螺旋-转折-螺旋 (helix-turn-helix)、螺旋-环-螺旋 (helix-loop-helix) 和锌指 (zinc finger) 等。而锌指蛋白 (zinc finger protein, ZFP) 是真核细胞中最普遍存在的一类 DNA 结合蛋白。根据这些 ZFP 锌指结构上的特征, 可分为 C₂H₂ 型、C₃H₄ 型和 C₄H₄ 型等。其中 C₂H₂ 型 ZFP 构成最大的转录因子家族, 在人类基因组中其编码基因约含有 300~700 个成员。这种蛋白质的

结构特点是含有两个或两个以上的锌指单位。锌指单位的基本模式是含有 28~30 个氨基酸, 一致序列为 CX₂₋₄CX₃FX₅LX₂HX₃₋₄H (X 指任何氨基酸)。每两个锌指单位之间的连接区有高度的保守性, 通常为 TGEKPYX。在 C₂H₂ 型 ZFP 中含有功能性结构域, 约有 1/3 的 C₂H₂ 型 ZFP 含有 KRAB (Krüppel-associated box) 型结构域。KRAB 结构域是由大约 75 个氨基酸组成, 往往存在于 ZFP 的氨基端, 由两个连续的模块 A 和 B 构成。研究表明, KRAB 型 ZFP 是起抑制作用的转录因子, KRAB-A 与 KRAB-B 之间有协同作用, 其抑制作用比单独的 KRAB-A 要强^[2]。

收稿日期: 2000-05-26; 修回日期: 2000-06-20

基金项目: 国家科委新药研究与产业化开发项目 (96-901-01-58); 上海市现代生物与新药产业发展基金 (984319121) 资助。本文中的核苷酸序列 GenBank 登录号为 AF265236。

作者简介: 王兆, 山东枣庄人, 24 岁, 男, 硕士研究生, 研究专业: 分子遗传学。电话: 021-65643723; E-mail: wz921@etang.com

我们从人的胎脑 cDNA 文库中克隆到一条基因,生物信息学分析表明,该基因编码一条新型的 KRAB 型 ZFP,在染色体上位于 19q13.2,命名为 *ZNF303* 基因。通过肝癌组织表达谱基因芯片杂交表明,该基因在肝癌组织中的表达量有明显下调。结合该基因的可能的抑制性转录因子功能,我们推论这种表达量的下降可能跟肝癌有密切的相关性。

1 材料与方 法

1.1 cDNA 文库的构建与测序

人胎脑 Poly A⁺ RNA 购自 Clontech 公司 (Cat.No.6525 - 1),用 SMARTTM PCR(Clontech, Cat.No.K1052 - 1)方法合成双链 cDNA,用 pBlue-script II SK(+)(Stratgene)载体构建 cDNA 文库。挑取 cDNA 克隆在 ABI - 377 型 DNA 测序仪上测定全序列。

1.2 生物信息学分析

采用嘉瑞基因数据库分析管理系统对序列进行管理和生物信息学分析,其中,序列拼接采用 Assembly,同源比较采用 Blast 2.0 版,基因数据库采用 GenBank、SwissPro 和 PDB,功能域数据库采用 prosite pattern、prosite profile 和 Pfam。

1.3 RH(radiation hybrid)基因定位

应用辐射杂交细胞系法定位。PCR 模板使用 Stanford G3 RH 嵌板(Research Genetics 公司),操作按说明书进行。引物为 *ZNF303* 5'ATGCAGGAAATCCTTCAACCAGC3' 和 5'GTGATTTTGAAGCAGGGACGAC3'。PCR 的扩增结果送斯坦福大学人类基因组中心(<http://www-shgc.stanford.edu>)进行统计分析。

1.4 肝癌组织表达谱基因芯片杂交

肝癌和癌旁正常肝组织由上海长海医院提供,抽取 mRNA 并用反转录法标记探针。杂交 BioDoorTM 4000 型基因表达谱芯片(United Gene),杂交参照说明书。用的 ScanArray 3000 扫描芯片,用 ImaGene3.0 软件分析 Cy3 和 Cy5 两种荧光信号的强度和比值。阳性结果的标准为: Cy3 和 Cy5 信号的比值的自然对数的绝对值 > 0.6 α 基因的表达变化在 2 倍以上),Cy3 或 Cy5 信号值其中之一必须 > 600。

2 结 果

2.1 *ZNF303*cDNA 的克隆

ZNF303 经过 Assembly 拼接,对获得的 2978bp cDNA 序列进行分析,该 cDNA 的 337 ~ 1770bp 是一

个完整的开放阅读框,编码 478 个氨基酸。2911bp 处有一个加尾信号(AATAAA)序列(见图 1)。将它的 cDNA 序列送到 GenBank 的 nr 数据库进行查新比较并登录,登录号为 AF265236。把序列信息送往 GDB 数据库登录,以 *ZNF303* 命名。

2.2 *ZNF303* 推导蛋白质序列的生物信息学分析

Motif 分析发现 *ZNF303*₂₈₀ ~ *ZNF303*₄₇₅ 之间含有 7 个连续的锌指结构,每两个锌指之间有典型的 TGEKPYX 结构(见图 1)。用 Expasy 网站(<http://www.expasy.ch>)的 Profilescan 分析发现,在 *ZNF303*₄ ~ *ZNF303*₃₆₆ 之间有典型的 KRAB 结构域。

2.3 RH 基因定位结果

在 *ZNF303* 基因的 3'UTR 中设计了一对特异的 PCR 引物,分别以 Stanford G3 RH Panel 的 83 个人鼠杂交克隆 DNA 为模板进行 PCR,PCR 扩增结果送斯坦福大学人类基因组中心,分析得到 3 个 LOD 值大于 6 的与该定位基因紧密连锁的 SHGC Marker,它们在染色体上的位置及与 *ZNF303* 基因的距离见表 1。人的 *ZNF303* 基因定位在 D19S425 附近,遗传图谱上人的 *ZNF303* 基因定位在 19 号染色体 19q13.2 区段(见图 2)。

2.4 肝癌组织表达谱基因芯片杂交结果

肝癌组织表达谱芯片杂交结果见表 2。表中的蛋白号是指在芯片上点有的已知基因表达的蛋白,变化率是指用杂交信号跟对照之间的比值,4 次平均所得到的结果。变化率 < 1 指表达下降,变化率 > 1 指表达上升。从基因芯片杂交结果中筛选出 500 多条有明显表达变化的基因,选取其中的 12 条基因跟 *ZNF303* 作对照说明。

3 讨 论

我们在人胎脑 cDNA 文库中筛选到一条 ZFP 基因,命名为 *ZNF303*。生物信息学分析发现,其产物属于 KRAB 型 C₂H₂ 锌指蛋白家族,推测在生物体内也是起到抑制转录作用。而在肝癌组织表达谱基因芯片杂交中显示,肝癌组织中 *ZNF303* 降低表达。在基因芯片所点的大量基因中,产生明显差异表达的仅有约 500,我们认为这 500 条基因是跟肝癌的发生与转移密切相关的。从表 2 中的数据看,癌变后的粘连蛋白受体、干扰素调节因子 - 2、TGF- β 超家族受体 - 1、核因子 Kappa-B 结合亚基(NF κ B) E2F 相关转录因子、DNA 修复解旋酶下调,这也符合癌变后的一般特征。而微体脂肪酰辅酶 A 氧化酶、丙

```

GGAGGCCCTGCTGAGGACTCCGGCAAGTGTGGTCCGGCGACGGCGGGCTAA
GGCCCTGGTCCGGCGCGGTTTGACCACGGCCGGGGCTTGGGCATTTCCTGG
CCTTCTGTGTAGCCGTGTAACCGGGGTGATGACGGCCGACCTCTTGCCA
CTGTTGTGAGAGCGAAGTGGGGCGAGAGCAGCCAGCTACAGTTTTTTTGG
GTTATGTCGTCAAGACCGCCCTTTCAGTGTGCAACCTTGAACAAATGGGA
CACTGCCATCTCTAAGATAAGAACCTGGAAGGGGACTCTGTGGCCATTGGA
AATTGCAGAATAATGCTCAGGTGACATTTAGTGATGTGGCTATAGACTTCTCT
      M S Q V T F S D V A I D F S
CATGAAGAGTGGGCATGCCCTAGATTCCTGCTCAGAGGACTTATACAAGGATGTG
H E E W A C L D S A Q R D L Y K D V
ATGCTCCAGAAATATGAGAACCTGGTCTCTGTAGGCTTTCCGTAACCTAAGCCA
M V Q N Y E N L V S V G L S V T K P
TATGTGATCATGTTATGGAGATGGAAGAGCCCTGGATGATGGAGAAAAAA
Y V I M L L E D G K E P W M M E K K
CTGTCAAAGCTTACCCATTTCTTATCACACTCTGTCTCTCTGTGAAC
L S K A Y P P P L S H S V P A S V N
TTTGGATTCCTGCTCTATTTGAGCATTGTTCAAGTCACTGAAATATTTGAG
F G F S A L F E H C S E V T E I F E
TTGTGAGAACTATGTTTCTGGTGTCTAATTTCTATCCAATTCCTCTAAT
L S E L C V F W V L H F L S N S P N
TCCACTGTAGAAGCTTTTCAAGAAGTAAAAAAGAAAAAAGAAAAAAGAAAA
S T V E A F S R S K K K K K K K K K K
AGGCAGTGCTTTGCTTCTTGATATATTTAGATTGGGAATCAAGATGGAAAA
R Q C F A F L I Y F R L G I K M G K
CAAGGAATATCAACAAGAAGATATTTATATGAAGATTACCCCAACCAAGTA
Q G I I N K E G Y L Y E D S P Q P V
ACAATGGAAGAAAGTTGTAACAAAGTATGAATTTTCAAATTTCTAATAAGAAT
T M E K V V K Q S Y E F S N S N K N
TTGGAATATACAGAATGGCAGACATTTAGAAGCACCTTTCATTCAAAGTCTACT
L E Y T E C D T F R S T F H S K S T
CTTTCAGAACCAAAAACATTTCTGCTGAAGGAATTCACACAAATATGATATA
L S E P Q N N S A E G N S H K Y D I
TTAAGAAGAAATTTATCAAAAAGTCACTTATAAAAAGTGAAGATAAATGGT
L K K N L S K K S V I K S E R I N G
GGAAACAACTTTTAAATCTAATAAAAAGTGGGCAGCCTTCAACCAAGAGCAA
G K K L L N S N K S G A A F N Q S K
TCTCTACCTTCCCAGACTTGAATAGAGAGAAAATCTATACATGCGAGTGAA
S L T L P Q T C [ N R E K I Y T ] C S E
TGTGGAAAGCCCTTGGCAACAGTCAATCTCAGTCGCACTGGAGAAATTCAT
C G K A F G K Q S I L S R H W R I H
ACAGGAGAGAAGCCCTATGAATGCTGGAATGTGGGAAGACTTTAGCCATGCT
[ T G E K P Y E ] C R E C G K T F S H G
TCATCCCTTACACGACATCAGATAAGCCATAGTGGAGAGAAACCTTACAAATGC
S S L T R H Q I S H [ S G E K P Y K ] C
ATTGAATGTGGAAAGCCCTTTAGCCATGGCTCATCACTACTACTAACCATCAGAGC
I E C G K A F S H G S S L T N H Q S
ACTCACACGGGAGAGAAACCGTATGAATGTATGAACCTGTGGAAAGCTTTTATGT
T H [ T G E K P Y E ] C M N C G K S F S
CGTGTGTCCTTCTCATTACAGATCTAAGAATTCATACGCAAGAAAAACCGTAT
R V S L L I Q H L R I H [ T Q E K R Y ]
GAGTCTGTATATGTGAAAGCCCTTCAATCATAGTTCGTCTCTCATTACCATT
[ E ] C R I C G K A F I H S S S L I H H
CAGAAAAGCCATACTGGAGAGAAGCCCTTATGAATGTAGAGAATGTGGGAAGCT
Q K S H [ T G E K P Y E ] C R E C G K A
TTCTGTGTAGTACACCTTACTCAACATCAAGAATTCACAGTATGAAGAAA
F C C S S H L T Q H Q R I H [ S M K K ]
AAATATGAATGCAAAATGTCTCAAGGCTTTAGTAGTCTCATTCTTCTGT
[ K Y E ] C N K C L K V F S S F S F L V
CAACATCAGACTATCATACTGAAGAAAACCGTTTGAAGTTTGAAGAAATGCAGG
Q H Q S I H T E E K P F E V *
AAATCCTTCAACCAAGCTTGAATCACTGAATATGCATTTGAGAAATCACATTAGA
TTGAAACCTTACGAATGCAGTATATGTGGAAAGCCCTTTAGTCATAGGTCGTCC
CTGCTTCAACATCAGATTTTCATCTGAGAGAAAACCTTACGAATGTATTTAAA
TGTGGGAAGCCCTTCACTGTAGTCAAACTTACTGTACATCAGAGAATTCAT
ACTGGAGAAAAGCCATATAAATGTAGTGTGGAAAGCTTTTAGCAAAAGCC
TCGAATCTTACTGCCCATCAAGAGTACATAAATGGAGAGAAAACCAATAGTGTG
GTAAGTGTGAAAAGCCCTTGTAGTATGAATCCCTATACATGTGAGAAATCT
TACAGAAGAGAAGCAGTGTATATCAGGTAACCTTCAATCATAGATCTCCCTT
ATTTAACATCAGAAAATGTATCTGGGAAAAGTGTGATGAAGGTGGTGAACA
TGGAGACTTTTAGCAATGATGCAGATTTTTTTATAGAGTTTACTGTAGAG
AAATCATATGAGTCAATAAATGTGGAAAGCCCTTGTGAGTATTAATCCCTTA
ATTGACCNTAAGTATACTACACTAGGAAAATCTGTGTACATGTAGCAATGT
GGGAAAGACTATAGGCAATAGGAATCTCTGCAAACTCTACAGGAGAAAAGTT
GTATGAATGTGGAACTTTTGAATAAGGAATTTTCCAGTCCAAGTGCAT
CCCTTATTCTATAGGAAATAAAGTGGAGAAAATCTCATTAAAGAGATGCAGCA
AAGTGTCACTAAGAGTGTATCTTGGCAGACATAAGAAGATGAATGTAGAG
CAACCTGAAGGATTTAGAAATACATATAAATCTTTCAGTATGCTATTTGTA
AACAGGATATATAGGAGACAAATAAACATAAGTATGCAATTTCTTAGAGCAGT
AGCTTGCAGTTTCAAGTGTACTTACTTAGAAAATCTTTTACTAGTGGGCAT
GTGAAGATATTTAGTCAACCCAGAGGCCAGTAAATGTATAATGTTAAAAATT
AAAGCTGCAAAAAGAAATAAATGCTGTTAATAAAAAATTTGGCATCTAATAAAA
TCATTTGTATATCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAA

```

图 1 ZNF303 cDNA 序列及推导蛋白质氨基酸序列

阴影部分所示为终止密码子；□中所示为连续锌指单位之间的保守序列；—上方所示为加尾信号 AATAAA；
←上方所示为 KRAB 结构域。本序列 GenBank 登录号为：AF265236。

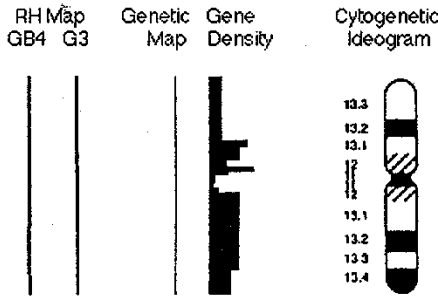


图2 ZNF303 在 19 号染色体上的定位

酰辅酶 A 羧化酶、谷胱甘肽转移酶 (GST1)、微体 phytanoyl 辅酶 A 羟化酶 (PAHX)、甲基丙二酸单酰辅酶 A 变位酶、铁传递蛋白受体 (TFR2) 有明显上调, 跟肝脏组织癌变后生理生化变化密切相关。所以我们有理由相信, ZNF303 在肝癌组织中表达的下调变化跟肝癌有密切的关系。无论 ZNF303 是肝癌形成的原因还是结果, 它都是与肝癌相关的重要一环。如果寻找出 ZNF303 作用的上游和下游基因, 将

其精确定位, 对于阐明肝癌发生或发生以后的生理病理变化将有很大的帮助。

许多 ZFP 在人类基因组上是成簇排列的, 而且它们倾向于分布到端粒区或脆性染色体区域, 易于产生断裂和染色体易位, 因此往往跟发育疾病和肿瘤的的产生密切相关^[3]。现在已经发现一些疾病跟 ZFP 的关系密切^[4-7]。在 19p12 区域, 存在着超过 40 个 ZFP 基因, 每一个基因两侧含有高度同源的 β -卫星 DNA, 在进化上可能来源于一个“种子”ZFP 基因的扩增。扩增的原因可能是最早 ZFP 基因插入到中心粒附近的 β -卫星 DNA 中, 其后在选择压力下, ZFP 基因增多起来^[8]。ZNF303 位于 19q13.2, 有超过 15 个不同的 ZFP 基因存在于 350~450kb 的区域内, 而且这些 ZFP 之间的间隔区比较小, 它们在染色体上以“头-尾”的方式连接在一起。在进化上, 这个区域的基因簇的形成可能与 19p12 区有类似的机制。
(下转第 251 页)

表 1 ZNF303 基因的 RH 定位数据及结果

SHGC64867
000000010000000000000000000010100000001001001001000000001001000000100011001000000
SHGC56626
000000010000000000000000000010100010001001000001000000001001000000100011001000000
SHGC4701
000000010000000000000000000010100010001000011001000000001001000000100010001000000
“0”表示 PCR 阴性结果; “1”表示 PCR 阳性结果。

标记名称	染色体序号	Lod 值	与 ZNF303 的距离 (cR)
SHGC - 64867	19	11.48	14
SHGC - 56626	19	11.48	14
SHGC - 4701	19	11.48	14

表 2 肝癌组织表达谱基因芯片杂交结果

号码	蛋白号	蛋白名称	变化率(平均)
1	BAA13220	黏连蛋白受体	0.01413286
2	CAA34073	干扰素调节因子 - 2	0.04933759
3	AAA16160	TGF-B超家族受体 - I	0.07448689
4	AAA36360	核因子 Kappa-B 结合亚基 (NF κ B)	0.170136
5	ZNF303	锌指蛋白 303	0.28984
6	AAA58440	E2F 相关转录因子	0.25513863
7	AAA52396	DNA 修复解旋酶	0.3191441
8	AAB30019	微体脂肪酰辅酶 A 氧化酶	4.28825844
9	CAA51825	丙酰辅酶 A 羧化酶	5.68593917
10	AAA59203	谷胱甘肽转移酶 (GST1)	6.154975
11	AAB81834	微体 phytanoyl 辅酶 A 羟化酶 (PAHX)	24.81242
12	AAA59569	甲基丙二酸单酰辅酶 A 变位酶	41.050156
13	NM-003218	铁传递蛋白受体 (TFR2)	86.02217

(上接 250 页)

参考文献:

- [1] Carl O Pabo, Robert T Sauer. Transcription factors: structure families and principles of DNA recognition[J]. *Annu Rev Biochem*, 1992, 61: 1053 ~ 95.
- [2] H. Vissing, W. K. H. Meyer, *et al.* Repression of transcriptional activity by heterologous KRAB domains present in zinc finger proteins [J]. *FEBS Letters*, 1995, 369, 153 ~ 157.
- [3] Tommerup N, Vissing H. Isolation and fine mapping of 16 novel human zinc finger-encoding cDNAs identify putative candidate genes for developmental and malignant disorders[J]. *Genomics*, 1995, 27, 259 ~ 264.
- [4] Niels Tommerup, Lissi Aagaard, *et al.* A zinc-finger gene ZNF141 mapping at 4p16.3/ D4S90 is a candidate gene for the Wolf-Hirschhorn (4p-) syndrome[J]. *Human Molecular Genetics*, 1993, 2: 10, 1571 ~ 1575.
- [5] Andrea Vortkamp, Manfred Gessler, *et al.* GLI3 zinc-finger gene interrupted by translocation in Greig syndrome families[J]. *Nature*, 1991, 352: 8, 539 ~ 540.
- [6] Call, K. M., Glaser, *et al.* Isolation and characterization of a zinc finger polypeptide gene at the human chromosome 11 Wilms'tumor locus[J]. *Cell* 1990, 60, 509 ~ 520.
- [7] Benoit Grondin, Martine Bazinet, *et al.* The KRAB zinc finger gene ZNF74 encodes an RNA-binding protein tightly associated with the nuclear matrix[J]. *J Cell Biol*, 1996, 271, 26, 15458 ~ 15467.
- [8] Evan E. Eichler, Susan M. Hoffman, *et al.* Complex β -satellite repeat structures and the expansion of the zinc finger gene cluster in 19p12 [J]. *Genome Research*, 1998, 791 ~ 808.