

人类新基因 WDR70 的克隆及初步研究

单宏爽¹, 郭丽丽¹, 慈宏亮^{1,2}, 李亦平^{1,2,3}

(1. 北京师范大学 生命科学学院生物医学研究所, 北京 100875; 2. 浙江赛尔生物医学研究院, 浙江 杭州 310006;
3. 美国哈佛大学 Forsyth 研究所, 美国 波士顿 02115)

摘要 运用基于基因组数据库的特定基因同源新基因的克隆策略得到一个人类新基因 WDR70, 该基因编码一个包含 12 个 WD40 结构域的蛋白。WDR70 的 cDNA 序列长 2 266 bp, 预测编码蛋白含 630 个氨基酸, 理论分子量为 70 × 10³ u, 染色体定位为 17p13.1。以小鼠胚胎为模型进行整体原位杂交, 结果显示 WDR70 基因在 8.5 d 小鼠胚胎中没有表达, 而在 9.5 d 和 10.5 d 的小鼠胚胎的脑部有特异表达。由此推断该基因对胚胎期脑部的发育有重要的影响。

关键词 同源分析; WD40 结构域; 生物信息学; 整体原位杂交

中图分类号 Q75 **文献标识码** A **文章编号** 1005 - 7021(2005)01 - 0011 - 05

The Clone of Human Novel Gene WDR 70 and Its Tentative Study

SHAN Hong-shuang¹, GUO Li-li¹, CI Hong-liang^{1,2}, LI Yi-ping^{1,2,3}

(1. Biomed. Res. Inst. Life Sci. Coll. Beijing Normal Univ. Beijing 100875;

2. Zhejiang Cell Biochem. Res. Coll. Hangzhou 310006;

3. The Forsyth Inst., Harvard Univ. Boston, MA 02115, USA)

Abstract A human novel gene WDR70 is obtained with the cloning strategy of novel homologous gene, which is a specific gene based on genome database. It encodes a protein of 12 WD40 domain. The cDNA length of WDR70 is 2 266 bp, and the putative protein encoded by this gene is 630 amino acid residues with a theoretical molecular weight of 70 000 daltons. The WDR70 is mapped to chromosome at 17p13.1. The result of whole-mount *in situ* hybridization showed that it is strongly expressed in the brain of embryo in day 9.5 and 10.5 mice although it isn't expressed in the embryo in day 8.5. Thus, it is concluded that this new gene has great influence upon the development of the embryo brain.

Keywords homological analysis; WD40 domain; bioinformation; whole-mount *in situ*; hybridization

WD 序列重复蛋白 (WD40 repeat) 是指一类含有 4~16 个 WD40 结构域的蛋白。每个 WD40 结构域都含有一个大约包含 40 个氨基酸的保守中心, 这个保守中心通常以色氨酸-天冬氨酸 (WD) 结尾, 这种结构域可以介导蛋白质之间的相互作用^[1]。WD 蛋白家族由许多结构相关、功能不同的蛋白组成, 还可以分为几个亚家族, 它们不但在调节各种细胞功能, 如信号传导、转录调节、细胞增殖等方面起到重要的作用, 而且还与人类的几种疾病有关^[2]。因此克隆和鉴定新的 WD 基

因, 并对其进行研究对扩大和发展 WD 蛋白家族, 以及为后基因组计划补充新的数据都具有重要的价值。随着人类基因组计划中测序工作的基本完成, 以及其他包括大鼠、小鼠和斑马鱼在内的许多物种基因组计划的开展, 及其相关数据库和相应软件的建立与开发, 都为我们利用生物信息学资源进行新基因的克隆研究带来新的机遇。其中, 基于基因组数据库辅助以 EST 数据库, 灵活运用目前的生物信息学软件, 寻找特定基因同源新基因不失为一个准确率高并且快速的新基因克隆的

收稿日期: 2004 - 03 - 31

作者简介: 单宏爽 女, 硕士生。研究方向为分子生物医学。

策略^[3]。2001年2月日本的RIKEN公布了21 076条小鼠全长cDNAs的序列,根据与已知基因的同源性,他们把这些基因分为9类,其中“motif-containing protein”这一类代表着编码具有一定结构域蛋白的573条小鼠新基因^[4]。本文在此基础上,通过同源克隆的方法得到相应的人类新基因,然后通过整体原位杂交实验,对该基因的表达谱进行初步研究。

1 材料与方法

1.1 WDR70 基因的电子克隆

从RIKEN的资源网站下载包含WD40功能结构域的小鼠新基因;把该基因作为“钓饵”,从NCBI网站人类基因组数据库中搜寻与其具有同源序列的片段;下载同源序列片段,用GenScan分析这一片段中可能包含的人类新基因候选基因;然后与NCBI网站人类EST数据库进行比较,判断其是否为一真基因,即是否与人类EST具有相同的序列;最后再把它与非冗余序列库进行比较,淘汰已知功能的人类基因,则得到人类新基因WDR70(NCBI登录号:NM-145054)。

1.2 WDR70 的生物信息学分析

1.2.1 WDR70 的核酸序列分析 运用Vector NTI软件对WDR70基因与小鼠同源基因(NCBI登录号:NM-027963)的编码区碱基序列进行同源性比较;在NCBI/Homo Genome数据库做BLAST分析WDR70基因在人基因组上的定位及分布。

1.2.2 WDR70 氨基酸序列及特性的分析 运用NCBI/rpsblast在线分析软件对WDR70的蛋白结构域和可能的功能进行预测;采用ExPASy/ProtParam工具进行蛋白的分子量和等电点分析;ExPASy/Prosite工具进行蛋白质功能位点分析;蛋白质序列的信号肽分析采用CBS的SignalP工具进行;亚细胞定位使用清华大学生物信息学院亚细胞定位数据库进行。

1.3 整体原位杂交

ICR品系的孕鼠购自北京大学医学部实验动物中心;地高辛标记与检测试剂盒购自罗氏(Roche)公司;WDR70(NCBI登录号:NM-145054)克隆购自ATCC,并测序(由上海博亚公司完成)以确定序列的正确性。全胚原位杂交的方法参照文献^[5],以WDR70全长编码区为模板,用地高辛核酸标记试剂盒制备DIG标记的RNA探针,分别与8.5~10.5 dpc的小鼠整体胚胎杂交,后用anti-dig抗体显色。

2 结果

2.1 WDR70 的电子克隆(图1)

从小鼠的包含WD40结构域功能基因入手,采用同源比较得到一个人功能未知基因WDR70。该基因cDNA全长2 266 bp,编码区序列长1 893 bp,含有起始密码子ATG和终止密码子TGA,其ATG起始密码周边序列(AAC ATGG)符合Kozak规律。编码的蛋白质含630个氨基酸。

1 agcaaa

```

7 gtaatcagaacctccaaggatggatgacaaaatttcgcccggaggcccaagtgccggagctggaacttgacgccgtgatcggttcaatg
97 gtccctgttctcctgaaactgttttagtgaataaactaccaactcaagcaaaacaataattacattcaagcatacaacttttctcctaac
187 atggaagcagtagtgttaccttggcagatatttggagtacggaggttgcctgtggaagagggtgcaccatgaaaggtcccaggctcttc
1 M E A V V L P W Q I F G V R E L P V E R G V T M K G P R L F
277 agagcacctggacatgtgccactggctcctcaaatgccatcctgaccaggagcatatgattatcctcttgggtgcacagctcctcattcag
30 R A P G H V P T G L K C H P D Q E H M I Y P L G C T V L I Q
367 gcaataaataactaaagagcagaacttctacagggtcatggcaacaacttctctgcttggccatctccaggctcggagagtacatcgcc
60 A I N T K E Q N F L Q G H G N N V S C L A I S R S G E Y I A
457 tccggacaagtacattcatggggttcaagcgagacatcattttgtgggattataagaacagagagctgcttgcctggctgtcccttcac
90 S G Q V T F M G F K A D I I L W D Y K N R E L L A R L S L H
547 aaaggcaaaattgaagctctggccttttctccaaatgatttacttggatcactaggaggccagatgacggaagtgtggtggtgtgg
120 K G K I E A L A F S P N D L Y L V S L G G P D D G S V V V W

```

637 agcatagccaagagagatgccatctgtggcagccctgcagccggcctcaatgttggcaatgccaccaatgtgatcttctccaggtgccg
 150 S I A K R D A I C G S P A A G L N V G N A T N V I F S R C R
 727 gatgagatgtttatgactgctggaatgggacaattcgagtatgggaattggatcttccaaatagaaaaatctggccaactgagtgcaa
 180 D E M F M T A G N G T I R V W E L D L P N R K I W P T E C Q
 817 acaggacagttgaaaagaatagtcagatgattggagtgatgatgatagctttttctaccttggcaccacgactggagaatttcta
 210 T G Q L K R I V M S I G V D D D D S F F Y L G T T T G D I L
 907 aaaatgaaccccaggactaaactgctgacagatgttgggcctgcgaaggacaaattcagtttgggagtgctagctatcaggtgcctgaag
 240 K M N P R T K L L T D V G P A K D K F S L G V S A I R C L K
 997 atggggggtttgttgggggctctggagccggactgctggtcttctgtaaagccctggctacaacccatcaagaagattcagttaca
 270 M G G L L V G S G A G L L V F C K S P G Y K P I K K I Q L Q
 1087 ggcggcatcacttctatcacacttcgaggagaaggacaccagtttctctaggaacagaagaatcgacatttatcgtgctagcttcacg
 300 G G I T S I T L R G E G H Q F L V G T E E S H I Y R V S F T
 1177 gatttcaagagacgctcatagcagcttctcactttgatgctgcaaggatattgtctttccatttggcactgctgagctatttgaacc
 330 D F K E T L I A T C H F D A V K D I V F P F G T A E L F A T
 1267 tgtccaagaaggatcaggggtggcagacatcatccaacaggagctgctgaggatcaccgtgcccaacatgacctgccagcgatc
 360 C A K K D I R V W H T S S N R E L L R I T V P N M T C H G I
 1357 gacttcatgagggacggcaaaagcatcatttcagcatggaacagcggtaaaatccgagccttcgccccagagacagggccgactgatgt
 390 D F M R D G K S I I S A W N D G K I R A F A P E T G R L M Y
 1447 gtcattaacaatgctcacaggatcggcgtcactgccatgccaccaccagtgactgtaaaagggtcatcagtgccgggtggggaaggggag
 420 V I N N A H R I G V T A I A T T S D C K R V I S G G G E G E
 1537 gtgagggatggcagatagggctgtagaccagaagctggaggaggccctgaaggacacaagtcacagtgctcattaggggtgaag
 450 V R V W Q I G C Q T Q K L E E A L K E H K S S V S C I R V K
 1627 aggaacaacaggagtgctgctaccgccagcaccgatgggacttgtatcatttgggaccttgtgcgtctcaggaggaaatcagatgatact
 480 R N N E E C V T A S T D G T C I I W D L V R L R R N Q M I L
 1717 gccaacaccttattccagtggtgtgctatcacccctgaggagtccagatcatcaccagcggaacagacagaagattgcttactgggaa
 510 A N T L F Q C V C Y H P E E F Q I I T S G T D R K I A Y W E
 1807 gtatttggagcagtaatacagagaattggaaggttccctgtctgggtcgataaatggcatggatcacacaggaaggggtgcacttt
 540 V F D G T V I R E L E G S L S G S I N G M D I T Q E G V H F
 1897 gtcacaggtggaatgaccatctggtcaaagttgggattataatgaggggtgaagtgactcacgttgggggtgggacacagtggaacatc
 570 V T G G N D H L V K V W D Y N E G E V T H V G V G H S G N I
 1987 acagcgcacccataagtcaggaaatcaatatattgttagtgta agtgccgatggagccattttgggatggaagtacccatatacctcc
 600 T R I R I S P G N Q Y I V S V S A D G A I L R W K Y P Y T S
 2077 tgaagctgatgagatgctctgagccttggcgttgcacgcagtcctggtgaagactgagtttagataaactccaacactagctctcatttc
 *
 2167 tcacagctctgttttgttcttgagtcaattttctcttttctttatagaatgcattttatattcttaaatgcatattaaattgaag
 2257 tatgttcaag

图 1 WDR70 全长 cDNA 序列及所编码的蛋白

起始密码子(ATG),终止密码子(TGA)下划线标出

Fig. 1 The full-length cDNA sequence of *hNA TL* and its coding protein
The initiative code ATG, stop code TGA are underlined

2.2 WDR70 与其他序列的比对结果

WDR70 的编码区与小鼠同源基因的编码区序列一致性(identity)为 86%,氨基酸序列的一致性为 89%。

2.3 WDR70 的染色体定位结果分析

通过 NCBI/Humo Genome 数据库中的 BLAST 分析,得知 WDR70 定位在人的第 17 号染色体短臂上 70 kb 的范围内(图 2)。

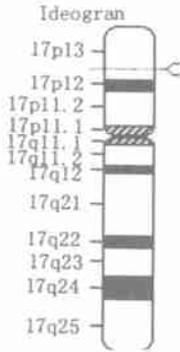


图 2 染色体上的定位
WDR70 基因定位于染色体 17q13.1 区
Fig. 2 The chromosome localization
The WDR70 gene is located on 17q13.1

2.4 WDR70 的保守结构域

使用 NCBI/rpsblast 中不同的数据库进行蛋白保守结构域的预测,测出几个高度保守的 WD40 重复结构域和几个含 WD40 重复结构域的区域。综合以上结果,得知 WDR70 基因共包含 12 个 WD40 结构域(图 3)。

2.5 其他生物信息学分析

WDR70 蛋白的理论分子量为 70×10^3 u,等电点为 8.20,没有前导肽区段,是个非分泌型蛋白,亚细胞定位结果为细胞质。另外,WDR70 蛋白还含有 5 个可能的 N-糖基化位点,7 个可能的 PKC 磷酸化位点,6 个可能的酪氨酸激酶磷酸化位点,14 个可能的 N-十四烷酰化位点。

2.6 整体原位杂交结果

为了观察 WDR70 在胚胎发育时的表达分布,用不同时期的小鼠胚胎做了整体原位杂交。发现在小鼠 9.5 dpc 和 10.5 dpc 的胚胎当中,该基因在脑部有较为活跃的表达;而 8.5 dpc 胚胎中,该基因没有表达(图 4)。7.5~10.5 dpc 正是小鼠胚胎脑部发育形成的关键时期,提示我们该基因对早期脑的发育有很重要的作用。

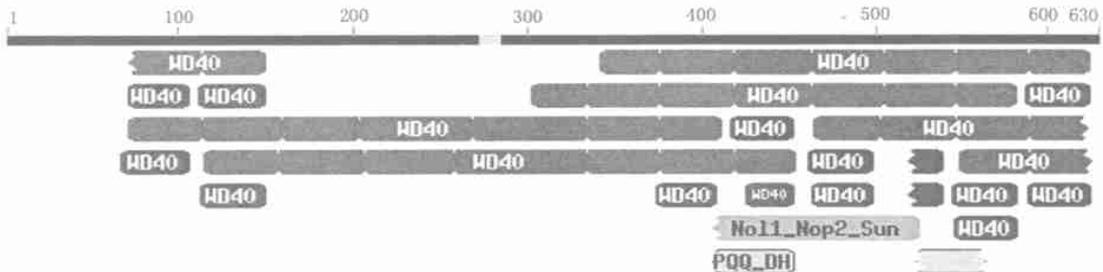


图 3 WDR70 蛋白一级结构功能结构域预测
Fig. 3 The predicted domains of WDR70 protein

3 讨论

自从 1986 年人们初次认识 WD40 结构域以来^[6],人们从原核生物到真核生物(尤其是真核生物)的许多蛋白中都发现了 WD 重复基元,而所有 WD 重复家族成员的一个共同的潜在的功能就是调整多蛋白复合物的装配^[1]。这些重复的结构域作为蛋白相互作用的支架,可以同时与几个不同的蛋白之间发生相互作用,例如 G 蛋白

的 亚基、TAFII 转录因子和 E3 泛素连接酶。由于 WD 重复蛋白是一个广泛而不断发展的家族,WDR 重复蛋白家族履行了一个广泛的细胞功能。最近有研究表明:基于基因组测序的完成,从当前包含 WD 重复序列的蛋白中已鉴定出 30 多个功能亚家族^[7],这些亚家族从功能上可以粗略的分为:信号转导(如 G 蛋白、RACK1、Striatin、STE4、LIS1、MSH1、PR55、PLAP、RbAp-48),RNA 合成/加工(如 TAFs、TUP1、CstF),染色质聚集

(CAF-1、HIR1、HIR2),囊泡运输(, 包被蛋白、SEC13),细胞骨架装配(Arp2、Arp3),细胞周期调控(CDC4、CDC20、CDC40、p55、Coronin、Mad2),细胞凋亡(Apaf-1、Dark、CED-4)等,另外还有新发现的功能未知的WD重复蛋白(WDR1、WDR3、WDR4、WDR6、WDR10)。最近有研究表明至少有4个WD基因与人类遗传病有关。Lissencephaly疾病、Cockayne氏综合征、Allgrove综合征都与其相关的WD基因突变有关^[2]。因此,克隆和鉴定WD蛋白并对其进行研究,对更好地了解 and 掌握WD蛋白家族的性质及功能有一定的价值;同时也为后基因组时代对功能基因的研究提供有价值的参考资料。

受精是新个体发育的时空点,从一个形态单纯的单细胞受精卵,发育成能独立生活的个体动

物,形态上出现一系列的变化,而更重要的是基因表达。基因组是机体内惟一确定的,为所有各类细胞共同拥有,但基因组内各个基因表达的选择性和程度随时间、位置和环境条件的不同而发生改变。小鼠胚胎于7.5 d神经板形成,8~9 d神经沟与神经褶形成,并发育成预定的脑区,随即形成3个原始脑泡:前脑、中脑和菱脑。其中8.5 d形成神经管,9 d中脑神经褶愈合,9.5 d前神经孔关闭,10 d后神经孔关闭。第10 d具有端脑、间脑、中脑、后脑和未脑5个脑泡^[8,9]。7.5~10.5 d是小鼠胚胎脑发育的重要时期,在该时期其独特的基因时空表达谱,而WDR70在小鼠胚胎8.5 d未发现表达,但9.5、10.5 d在脑部有特异表达,说明WDR70对胚胎脑的发育有重要影响,并为早期脑发育的基因表达谱提供新的信息。

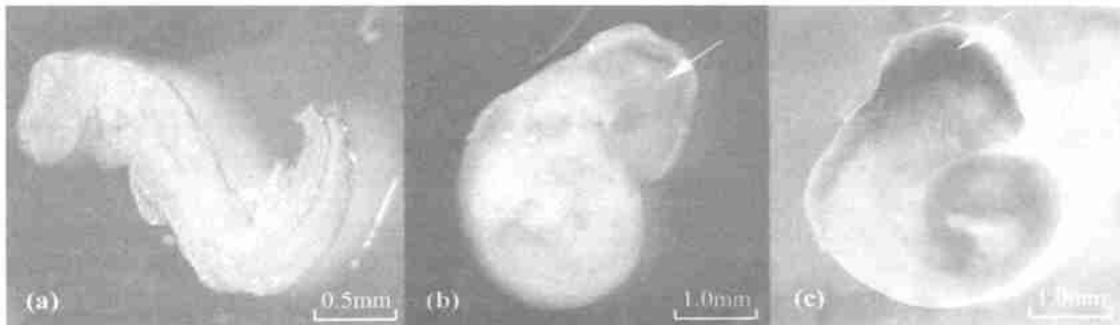


图4 小鼠胚胎整体原位杂交结果

箭头所指为胚胎头部特异性染色部位。(a)8.5 dpc;(b)9.5 dpc;(c)10.5 dpc

Fig.4 Whole-mount in situ hybridization of embryo mice

The arrow indicates WDR70 specifically expressed in mouse embryo brain. (a)8.5 dpc;(b)9.5 dpc;(c)10.5 dpc

参考文献:

- [1] Smith T. F., Gaitatzes C., Saxena K. and Neer E. J. The WD repeat: a common architecture for diverse functions[J]. Trends Biochem. Sci., 1999, 24: 181 - 185.
- [2] D. Li and R. Robert. WD-repeat proteins: structure characteristics, biological function, and their involvement in human diseases[J]. Cell. Mol. Life Sci., 2001, 58: 2 085 - 2 097.
- [3] 李铁石, 常智杰, 傅新元, 等. 基于 Internet 网生物信息资源特定基因同源新基因克隆策略[J]. 生命科学, 2002, 14(4): 234 - 237.
- [4] Kawai J., Shinagawa A., Shibata K., Yoshino M., *et al.* Functional annotation of a full-length mouse cDNA collection[J]. Nature, 2001, 409(6 821): 685 - 690.
- [5] Wilkinson D. G. In situ hybridization. In: Stern CD, Holland PWH (eds.) Essential Development Biology: A Practical Approach[M]. IRL Press, Oxford, U K, 1993, 257 - 274.
- [6] Fong H. K., Hurley J. B., Hopkins R. S., *et al.* Repetitive segmental structure of the transducin beta subunit: homology with the CDC4 gene and identification of related mRNAs[J]. Proc. Natl. Acad. Sci. USA, 1986, 83: 2 162 - 2 166.
- [7] Yu L., Gaitatzes C., Neer E., *et al.* Thirty-plus functional families from a single motif[J]. Protein Sci, 2000, 9: 2 470 - 2 476.
- [8] 秦鹏春, 李云龙, 谭景和, 等. 哺乳动物胚胎学[M]. 北京: 科学出版社, 2001, 238.
- [9] 俞慧珠, 叶百宽. 小白鼠胚胎发生[M]. 北京: 科学出版社, 1985, 86 - 96.