

文章编号: 1674 - 7054(2010)02 - 0114 - 07

人类多囊肾病相关蛋白 TMEM130 基因的克隆 及生物信息学分析

齐兴柱^{1,2}, 胡文婷¹, 张俊芳^{2,3}, 黄俊生^{2,3}, 李 军¹

(1. 海南大学 海洋学院, 海南 海口 570228; 2. 海南大学 农学院, 海南 海口 570228;

3. 中国热带农业科学院 环境与植物保护研究所, 海南 儋州 571737)

摘 要: 以递交到 GenBank 上的 TMEM130 蛋白基因序列 (nm - 152913) 为模板设计引物, 以人大脑 cDNA 文库为模板进行巢氏 PCR 扩增, 并将 PCR 产物克隆到 pMD18-T 载体上, 成功获得 TMEM130 蛋白基因编码序列。测序结果显示, 获得的序列为 TMEM130 蛋白基因转录变异体 2, 其编码序列长 1 272 bp。该基因定位在人类第 7 条染色体 7q22.1 区域, 有 3 个转录变异体。生物信息学分析显示该序列可编码 1 个含 423 个氨基酸的蛋白质; Smart mode 程序显示该蛋白质是 1 个跨膜蛋白, 在 100 ~ 250 号氨基酸序列之间有 1 个与多囊肾病 (PKD) 相关的结构域。因此, 鉴定 TMEM130 蛋白是一个与人类多囊肾病相关的蛋白。

关键词: TMEM130 蛋白; 多囊肾病; 生物信息学; 克隆

中图分类号: R 318.04

文献标志码: A

多囊肾病 (Polycystic Kidney Disease, PKD) 是一种肾小管囊肿进行性扩张以及正常肾小管的结构和功能进行性丢失并最终导致终末期肾衰竭的遗传性肾脏疾病, 可分为常染色体显性遗传性多囊肾病 (ADPKD)、常染色体隐性遗传性多囊肾病 (ARPKD) 和成人多囊肾病 (APKD) 等。常染色体显性遗传性多囊肾病 (Autosomal Dominant Polycystic Kidney Disease, ADPKD) 是人类最常见的单基因遗传病之一, 发病率约为 0.1%^[1]。其主要临床特征是双侧肾脏形成多个液性囊肿, 囊肿进行性生长, 导致肾脏结构和功能的损害, 至 60 岁时约 50% 患者发展为终末期肾功能衰竭, 约占终末期肾功能衰竭病因的 10%。多囊肾病患者除肾脏改变外, 还可累及全身多个器官, 如肝、胰、脾囊肿、颅内动脉瘤、高血压^[2]等。常染色体隐性遗传性多囊肾病^[3] (Autosomal Recessive Polycystic Kidney Disease, ARPKD) 属于单基因遗传病, 发病率在 1/40 000 ~ 1/20 000, 临床表现为肾脏集合系统囊样扩张并伴有不同程度的胆囊发育不全、胆管扩张以及肝门静脉周围纤维化。成人多囊肾病 (Adult Polycystic Kidney Disease, APKD) 也是一种常染色体显性遗传病, 发病率高, 大约有 1/1 000 的人为致病基因携带者。多囊肾病为多发病, 也是中年人尿毒症常见病因。为研究人类 TMEM130 蛋白的功能及其与人类多囊肾病的关系, 笔者以人的大脑 cDNA 文库为模板, 成功扩增并克隆了人类多囊肾病相关蛋白 TMEM130 转录变异体 2 的编码序列, 并对该基因进行了相关生物信息学分析和通过软件对该蛋白质的空间结构进行了预测。

1 材料与方法

1.1 人类多囊肾病相关的蛋白 TMEM130 基因编码序列的克隆 通过搜索 <http://www.ncbi.nlm.nih.gov/> 网站的 Nucleotide 数据库, 得到一段人类 cDNA 序列 (登录号 nm - 152913), 以该序列为模板, 利用 Primer Premier 5.0 引物设计软件设计了 2 对 PCR 引物: HTMEM130sen1 (5'-TTCTCGAGCTCCAGCTGCATTC 3'), HTMEM130ant1 (5'-CACTGAGATGGGGTGGGAGG 3'), HTMEM130sen2 (5'-AATGGCCCAGGCAGTGT-

收稿日期: 2010 - 03 - 19

基金项目: 海南省教育厅高等学校科研项目 (Hjkj2010 - 13); 海南省自然科学基金 (808122); 海南热带水生生
物技术重点实验室开放基金 (shkyjj0801)。

作者简介: 齐兴柱 (1973 -), 男, 湖南津市人, 海南大学农学院 2009 级在职博士研究生。

GGT 3') ,HTMEM130ant2(5'AGGGGGAGTGCTCACACGGT 3')。通过 NCBI 网站的 UniGene 数据库查询,发现该基因有大量来自人类大脑组织的 EST 序列数据,推测该基因在人类大脑组织中有表达。然后,以人类大脑 cDNA (购自 Clontech 公司) 文库为模板,进行巢氏 PCR 扩增,首先,HTMEM130sen1 和 HTMEM130ant1 为引物进行第 1 次 PCR 扩增,然后以第 1 次 PCR 扩增产物为模板,HTMEM130sen2 和 HTMEM130ant2 为引物进行第 2 次 PCR 扩增。反应体系为:超纯水 16.2 μ L,10 \times ExTaq buffer 2.5 μ L,dNTP 2.5 μ L,正向引物和反向引物各 1.3 μ L,模板 1 μ L,ExTaq DNA 聚合酶 0.2 μ L,总体积 25 μ L。第 1 次 PCR 反应条件为:94 $^{\circ}$ C 变性 4 min,94 $^{\circ}$ C 30 s,52 $^{\circ}$ C 30 s,72 $^{\circ}$ C 2 min,30 个循环,72 $^{\circ}$ C 延伸 8 min;第 2 次 PCR 反应条件为 94 $^{\circ}$ C 变性 4 min,94 $^{\circ}$ C 30 s,60 $^{\circ}$ C 30 s,72 $^{\circ}$ C 2 min,30 个循环,72 $^{\circ}$ C 延伸 8 min。PCR 产物用琼脂糖凝胶电泳进行检测,目的片段经纯化回收试剂盒进行回收。连接反应采用 TAKARA 公司 pMD18-T 载体连接试剂盒进行,连接产物转化大肠杆菌 DH5 α 感受态细胞,提取质粒并经酶切鉴定,然后将克隆成功的样品送上海生工测序。

1.2 TMEM130 基因的生物信息学分析 TMEM130 基因组信息的分析通过 <http://www.ncbi.nlm.nih.gov/> 网站的 Nucleotide、UniGene、UniSTS 数据库进行。同源序列的查询采用 <http://www.ncbi.nlm.nih.gov/> 网站的 blast 在线软件进行。序列比对及进化树分析分别采用 ClustalX 及 MEGA4.0 软件进行。蛋白质序列分析采用 DNASTar 软件进行。蛋白质的结构域三维模型分析采用 SWISS-MODEL 软件进行,并用 WebLab viewer 显示。

2 结果与分析

2.1 TMEM130 基因的克隆及序列分析 以人类大脑 cDNA 文库为模板,PCR 扩增人类 TMEM130 基因的开放阅读框 cDNA 序列,PCR 产物长度在 1 000 ~ 1 500 bp 之间,其附近无非特异性条带(图 1),将产物纯化后克隆到 T 载体上进行测序,测序结果如图 2 所示(只列出 1 个测序反应),经 blast 比对证实克隆片段为人类 TMEM130 基因转录变异体 2 (nm - 152913,blast 比对结果未列出)的开放阅读框的 cDNA 序列。图 3 显示,TMEM130 基因转录变异体 2 的 cDNA 全长 3 029 bp,开放阅读框长 1 272 bp,预测其编码 1 个含 423 个氨基酸的蛋白质。基因组信息查询发现该基因定位在人类第 7 条染色体 7q22.1 区域,在基因组上的跨度为 23.562 kb,包含 8 个外显子(图 4),1 号外显子长 274 bp,2 号外显子长 306 bp,3 号外显子长 160 bp,4 号外显子长 166 bp,5 号外显子长 85

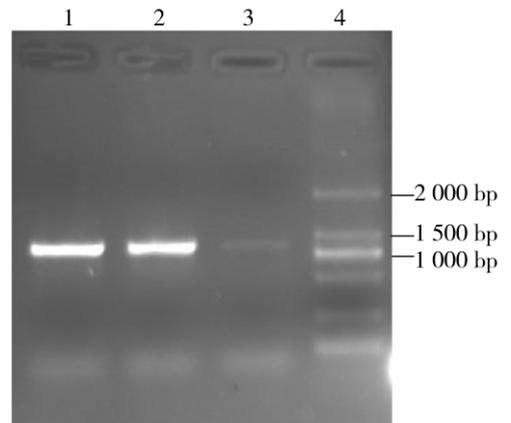


图 1 PCR 产物电泳图
(1~3. 第 2 次 PCR 产物;4. marker)

bp,6 号外显子长 203 bp,7 号外显子长 113 bp,该基因的转录本存在 3 种拼接方式,即 3 种转录变异体 (transcript variant),其中,转录变异体 1 的 8 号外显子长 1 757 bp,转录变异体 2 的 8 号外显子长 1 721 bp,而转录变异体 3 则没有 2 号外显子,其 8 号外显子与转录变异体 2 的 8 号外显子一样长(图 4)。3 种转录变异体的 5'端均有 189 bp 的非翻译区,3'端均有 1 568 bp 左的非翻译区。起始密码子均开始于 190 bp 处,其周围的碱基序列是 GCAATGG,与 Kozak 所描述的真核生物起始翻译序列 APGNNATGG 相一致,表明此 ATG 是 TMEM130 基因编码序列的起始密码子。

A

```
TCTGGCTTGCCTGCCTCTGCCTGGGCCCGGCAGGGGTGGCCGACAGGCCTGTATGAACCTAATCTCACACCGATAGCCCTGCCACCACGGGAGCGGTGGTGACC
ATCTCGGCCAGCCTGGTGGCCAAGGACAACCGCAGCCTGGCCCTGCCCGCTGACGCCACCTCTACCGCTTCCACTGGATCCACACCCCGCTGGTGCTTACTGGCAA
GATGGAGAAGGGTCTCAGCTCCACCATCCGTGTTGTGGCCACGTGCCCGGGAATCCCGGTCTCTGTCTGGGTCACTGCCCTGACTGCTGGATGTGCCAGCCTG
TGGCCAGGGCTTGTGGTCTCCCATCACAGAGTTCCTCGTGGGGGACCTTGTGTGTCACCCAGAACAACCTCCCTACCTGGCCAGCTCCTATCTCACTAAGACC
GTCCTGAAAGTCTCCTTCTCCTCCACGACCCGAGCAACTTCTCAAGACCGCCTGTTTCTCTACAGCTGGGACTTCGGGAGCGGGACCCAGATGGTGACTGAAGA
CTCCGTGGTCTATTATACTATTCCATCATCGGGACCTTACCGTGAAGCTCAAAGTGGTGGCGAGTGGGAAGAGGTGGAGCCGATGCCACGAGGGCTGTGAAGC
AGAAGACCGGGGACTTCTCCGCTCGCTGAAGCTGCAGGAAACCCCTCGAGGCATCCAAGTGTGGGGCCACCCTAATTCAGACCTTCCAAAAGATGACCGTGACC
TTGAACCTTCTGGGGAGCCCTCCTCTACTGTGTGCTGGCGTCTCAAGCCTGAGTGCCTCCCGCTGGAGGAAGGGGAGTGCCACCCTGTGTCCGTGGCCAGCACAGC
GTACAACCTGACCCACACCTTCAGGGACCTGGGGACTACTGCTTACGATCCGGGCCGAGAATATCATCAGCAAGACACATCAGTACCACAAGATCCAGTGTGCC
TCCAGAATCAGCCGGCTGCTTTGCTTCCATGTGCTACACTATCACTGTGATGTTGGGCTTCATCATGTACATGACCCTGGCGATG
```


(*Pan troglodytes* 黑猩猩)、NP_001126556.1(*Pongo abelii* 苏门答腊猩猩)、XP_001494483.2(*Equus caballus* 马)、EFB25467.1(*Ailuropoda melanoleuca* 大熊猫)、XP_001512168.1(*Ornithorhynchus anatinus* 鸭嘴兽)、XP_001111417.1(*Macaca mulatta* 猕猴)、XP_851082.1(*Canis lupus familiaris* 狗)、XP_414751.2(*Gallus gallus* 红原鸡)、XP_002192176.1(*Taeniopygia guttata* 斑胸草雀),但在低等动物、无脊椎动物以及其他低等脊椎动物中没有发现 TMEM130 蛋白的同源蛋白基因。这说明 TMEM130 蛋白基因是一个高度进化的基因。笔者将人类 TMEM130 蛋白质及其在其他物种中的同源蛋白输入 DNASTAR 软件,利用其中的 Clustal V 程序进行同源蛋白序列比对排序,并利用其中的进化树分析程序进行了进化树分析(图 5),表明该蛋白与黑猩猩、苏门答腊猩猩和猕猴的同源蛋白聚为一个大的分支,其中,人类 TMEM130 蛋白与黑猩猩的 MEM130 蛋白亲缘关系最近,其同源性达到 93%,这与整个动物进化分支是一致的。

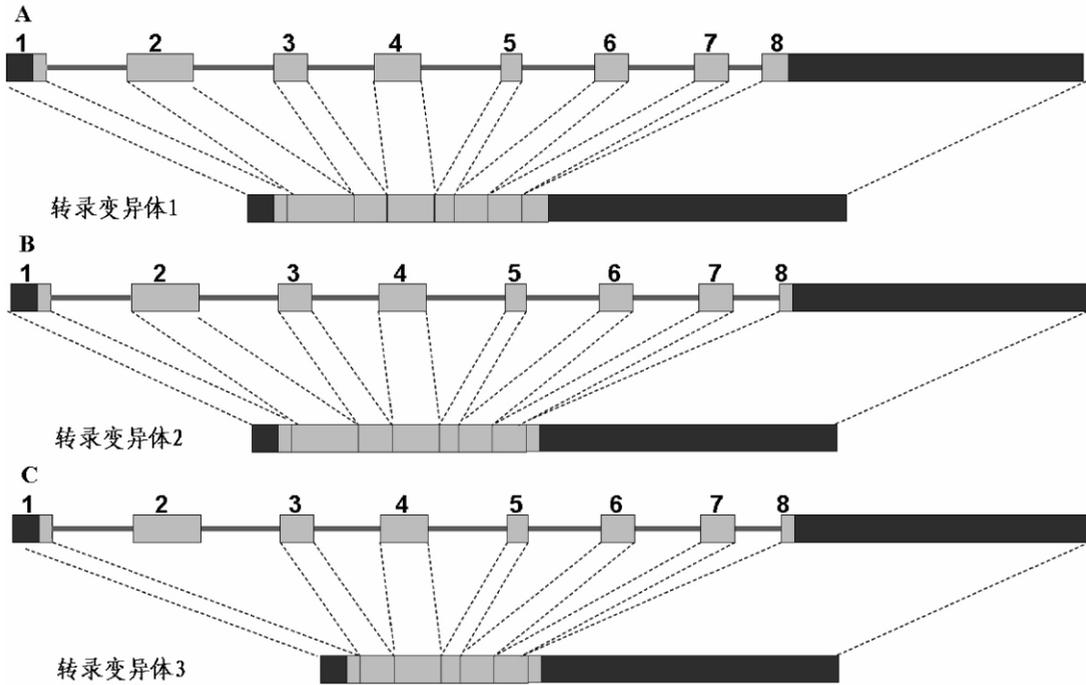


图 4 人类 TMEM130 基因转录本的 3 种拼接方式,结果产生 3 种转录变体。
图中 1 号外显子中的黑色方框表示 5' 非翻译区,8 号外显子中的黑色方框表示 3' 非翻译区,浅黑色方框表示外显子的编码区,虚线条表示内含子。

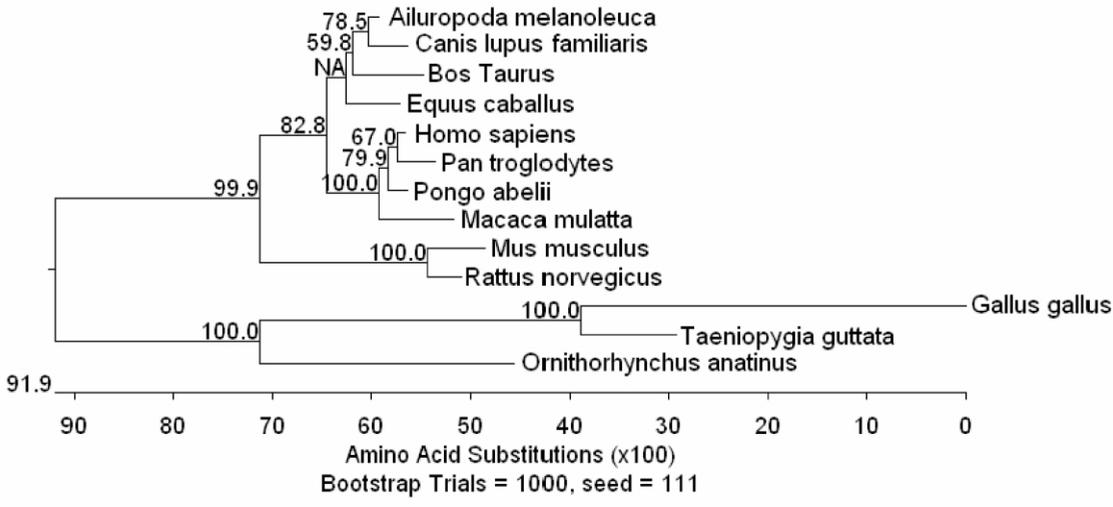


图 5 人类 TMEM130 转录变体 2 的蛋白及其同源蛋白的进化树

2.3 TMEM130 蛋白的结构域分析 笔者将克隆后测序并比对正确的 TMEM130 蛋白基因 cDNA 序列输入 DNASTar 软件中的 EDITSEQ 程序,预测到该编码序列的编码蛋白,由 423 个氨基酸组成,分子质量为 46 988.79 D,等电点为 6.684。将该蛋白质输入到 NCBI 蛋白质数据库中进行 blast 搜索,发现该蛋白质的 100~250 号氨基酸组成结构是一个与人类多囊肾病即 PKD 相关的结构域。图 6 为该蛋白质与典型的人类 PKD 蛋白质结构域保守序列的比对结果。笔者将该蛋白序列输入在线 SWISS-MODEL 软件,以 1B4R_A 序列为模板分析蛋白质的结构域三维模型,并用 WebLab viewer 显示,获得了该蛋白质的三维空间模型^[4-6](图 7)。图 8 显示的是典型 Cn3D4.1 软件显示的典型 PKD 结构域的三维空间模型示意图,其蛋白质的二级结构含有 6 个 β 折叠区域 β 折叠区域之间也有 α 螺旋,其超二级结构属于 3 种基本结构中的 $\alpha\beta\alpha$ 型。而图 7 显示的 TMEM130 转录变异体 2 蛋白质的三维空间模型图中也有 3 个 β 折叠区域。从 3 个 β 折叠的排列和走向看,与图 8 中 β 折叠区域非常类似。因此,笔者从 3 个方面 [1) 蛋白质的序列组成; 2) 与典型 PKD 结构域的序列比较(图 6); 3) TMEM130 转录变异体 2 蛋白的三维空间模型与典型 PKD 结构域的三维空间模型比较(图 7~8)] 进行预测,鉴定该蛋白质是一个与人类多囊肾病相关的蛋白质。将该蛋白质序列输入到 smart mode 在线软件,预测到该蛋白质 1~24 号氨基酸为信号肽,同时,12~31 及 340~362 号氨基酸为该蛋白的跨膜序列,因此该蛋白为膜蛋白。这与它的 PKD 结构域及作为细胞膜上的蛋白质与蛋白质或蛋白质与糖类相互作用的配基位点是一致的。

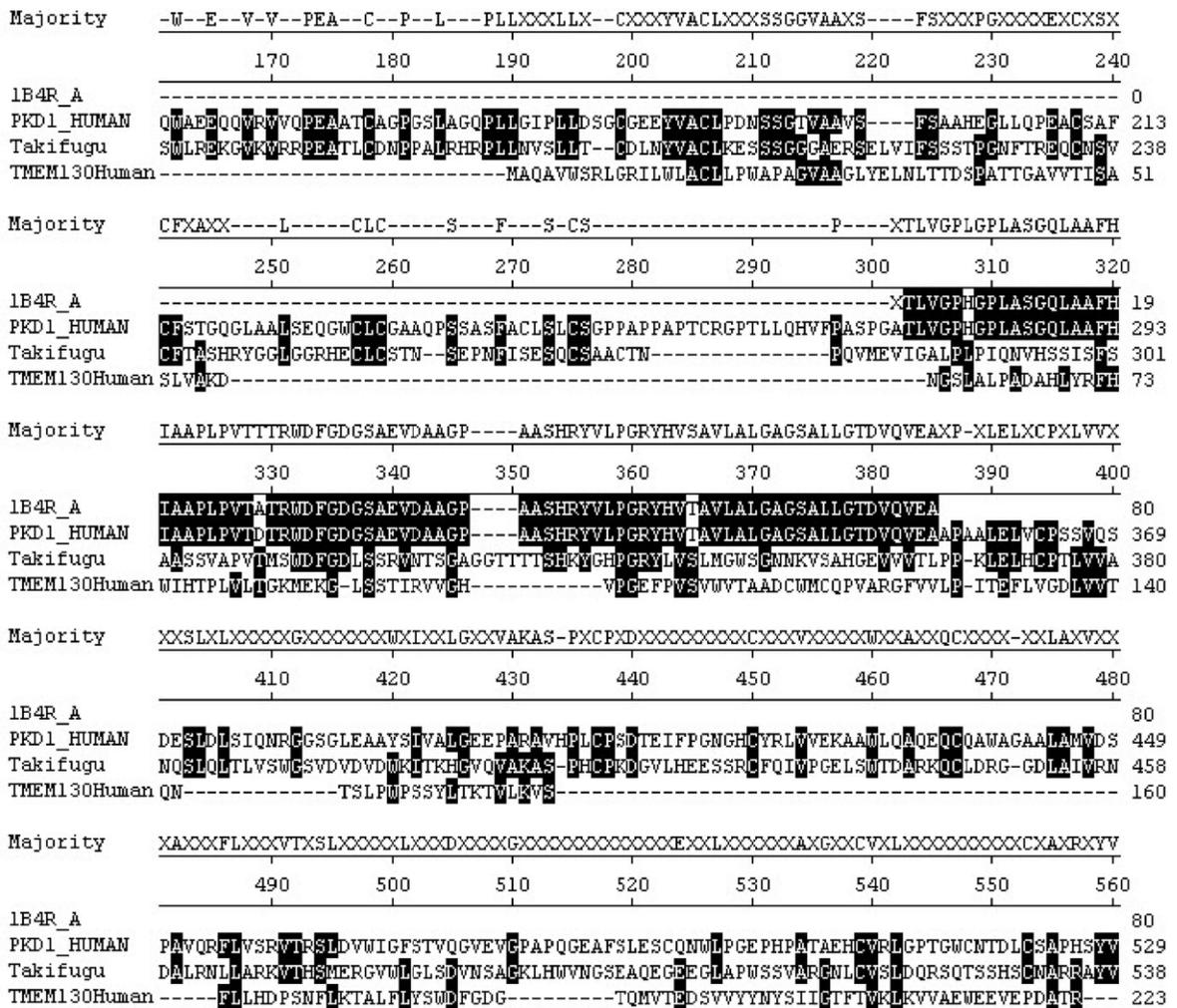


图 6 人类 TMEM130 蛋白与人类 PKD1 蛋白(PKD1_HUMAN) 人 PKD 标准序列(1B4R_A) 及暗纹东方鲀(Takifugu fasciatus) PKD1 蛋白的 PKD 结构域序列比较 (被鉴定的保守氨基酸由黑色阴影部分标记)

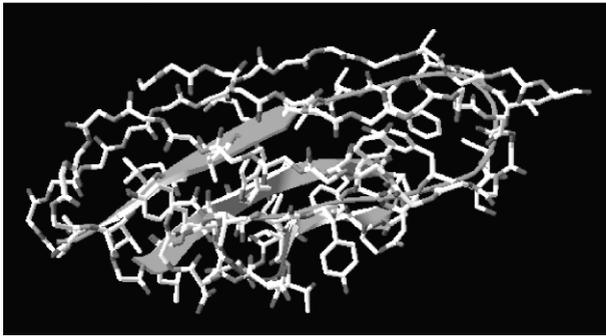


图 7 人类 TMEM130 蛋白质结构的三维空间模型
(箭头表示 β -折叠, 其他部分为不规则区域的氨基酸残基及其侧链)

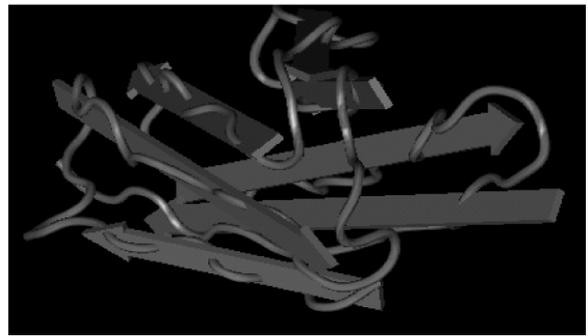


图 8 人类 TMEM130 蛋白质 PKD 结构域的超二级结构 Cn3D 图像
(图中箭头表示 β -折叠, 其它部分为 α -螺旋或不规则结构)

3 讨论

Blast NCBI 的 EST 数据库显示 EST 数据库中有大量来自人类大脑组织的 TMEM130 转录变异体 2 的 mRNA。为了研究人类 TMEM130 蛋白的功能及其与人类多囊肾病的关系, 以人的大脑 cDNA 文库为模板, 成功扩增并克隆了人类多囊肾病相关蛋白 TMEM130 转录变异体 2 的编码序列。序列分析显示, 笔者设计的 PCR 引物在基因的第 1 个外显子和第 8 个外显子上, PCR 扩增的结果只得到转录变异体 2, 因此笔者推测: (1) TMEM130 蛋白 (GI: 23097273) 在人类大脑中有较多表达; (2) 在该基因的 3 种转录变异体中, 转录变异体 2 是最主要的 1 种转录本。

目前, 多个与人类多囊肾病相关的基因都已经被克隆和进行了序列分析^[7-9], 人类多囊肾病发病的分子机制并没有得到详细的阐明。人类 TMEM130 蛋白是否参与到多囊肾病发病的分子机制还有待进一步研究。

参考文献:

- [1] BOUCHER C, SANDFORD R. Autosomal dominant polycystic kidney disease (ADPKD, MIM173900, PKD1 and PKD2 genes, protein products known as polycystin-1 and polycystin-2) [J]. *Eur J Hum Genet*, 2004, 12(5): 347-354.
- [2] ROMAO E A, MOYSES N, TEIXEIRA S R, *et al.* Renal and extrarenal manifestations of autosomal dominant polycystic kidney disease [J]. *Braz J Med Biol Res*, 2006, 39(4): 533-538.
- [3] 江雨. 常染色体隐性遗传性多囊肾病基因缺陷的研究进展 [J]. *中国优生与遗传杂志* 2007, 15(2): 114-117.
- [4] ARNOLD K, BORDOLI L, KOPP J, *et al.* The SWISS-MODEL Workspace: A web-based environment for protein structure homology modelling [J]. *Bioinformatics* 2006 22: 195-201.
- [5] SCHWEDE T, KOPP J, GUEX N, *et al.* SWISS-MODEL: an automated protein homology-modeling server [J]. *Nucleic Acids Research*, 2003, 31: 3381-3385.
- [6] GUEX N, PEITSCH M C. SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modelling [J]. *Electrophoresis*, 1997, 18: 2714-2723.
- [7] WARD C J, HOGAN M C, ROSSETTI S, *et al.* The gene mutated in autosomal recessive polycystic kidney disease encodes a large, receptor-like protein [J]. *Nat Genet*, 2002, 30(3): 259.
- [8] MCCONNELL R S, RUBINSZTEIN D C, FANNIN T F, *et al.* Autosomal dominant polycystic kidney disease unlinked to the PKD1 and PKD2 loci presenting as familial cerebral aneurysm [J]. *J Med Genet*, 2001, 38(4): 238-240.
- [9] TORRESMJ, RODRIGUEZ PEREZ J C, HERNANDEZ SOCORRO C R, *et al.* Molecular diagnosis of adult dominant polycystic kidney disease in the Canary Islands [J]. *Nefrologia*, 2006, 26(6): 666-672.

Cloning and Bioinformatics Analysis of Human Polycystic Kidney Disease-related Protein TMEM130

QI Xing-zhu^{1,2}, HU Wen-ting¹, ZHANG Jun-fang^{2,3}, HUANG Jun-sheng^{2,3}, LI Jun¹

(1. Ocean College, Hainan University, Haikou 570228, China; 2. Agricultural College, Hainan University, Danzhou 571737, China;

3. Environment and Plant Protection Institute, CATAS, Danzhou 571737, China)

Abstract: According to TMEM130 gene sequence (NM-152913) information, human brain cDNA library was used as the template and the specific primers were designed. The nPCR amplification product was separated by agarose gel and the band was cloned into PMD18T-vector and sequenced. The analysis results showed that the sequence is the transcript variant 2 of TMEM130, as a transmembrane protein, which contains an open reading frame of 1272 nucleotides and encodes a protein with 423 amino acids and contains a polycystic kidney disease (PKD)-related domain between 100 and 250 amino acid.

Key words: TMEM130 protein; polycystic kidney disease; bioinformatics; cloning