

研究快报

LRRC4 融合蛋白的构建与表达研究*

王洁如¹⁾ 董利¹⁾ 蒋明²⁾ 谭琛¹⁾ 李小玲¹⁾
向娟娟¹⁾ 范松青¹⁾ 彭聪¹⁾ 唐珂¹⁾ 李桂源¹⁾**

(¹⁾中南大学湘雅医学院, 长沙 410078; (²⁾深圳市蓝田区人民法院, 深圳 518081)

摘要 在前期工作中, 采用 EST 介导的定位候选克隆策略, 克隆了一个在脑瘤中表达下调的脑特异表达新基因 LRRC4, 为进一步研究其结构与功能的关系, 构建了含 LRRC4 基因全长编码区的 pGEM-T Easy 质粒, 在此基础上通过亚克隆构建了 LRRC4 融合蛋白的绿色荧光蛋白 (pEGFP-C1) 表达质粒, 瞬时转染哺乳动物细胞, 结果发现表达的 LRRC4 融合蛋白定位于活细胞的细胞膜上. 同时, 构建了 LRRC4 全长和截短型原核表达 pGEX-4T-2 质粒, 成功而高效地在大肠杆菌 BL21 中表达 LRRC4 融合蛋白. 上述工作为制备多抗, 深入研究 LRRC4 基因的功能奠定了基础.

关键词 LRRC4, 融合蛋白表达, EGFP, GST, 生物信息学

学科分类号 Q78

在前期工作中我们运用 EST 介导的定位候选克隆策略, 克隆了一个定位于染色体 7q31~32 区域的脑特异性的富亮氨酸超家族新基因 LRRC4 (GenBank 登录号: AF196976)^[1], 研究发现它在原发性脑瘤中存在表达明显下调或缺失. 最近的研究结果表明 LRRC4 在正常脑组织的表达主要集中在神经胶质细胞, 且体现了一种随发育调节的格局 (准备发表的结果). 结果提示 LRRC4 在神经系统正常功能的维持及疾病形成 (特别是脑胶质瘤) 中发挥重要作用.

本研究构建 LRRC4 融合蛋白的真核表达质粒 pEGFP-C1/LRRC4, 经瞬时转染胶质母细胞瘤细胞株 U251, 可见融合蛋白的绿色荧光集中分布在细胞膜. 证明 LRRC4 在细胞中作为膜蛋白发挥其作用. 同时, 构建 LRRC4 全长及截短型融合蛋白的原核表达质粒 pGEX-4T-2/LRRC4 和 pGEX-4T-2/mLRRC4, 转化大肠杆菌 BL21, 成功地观察到 LRRC4 融合蛋白在原核生物中的表达, 为分离和纯化蛋白质, 制备 LRRC4 抗体, 进一步揭示 LRRC4 基因的功能奠定了基础.

1 材料与方法

1.1 材料

限制性内切酶、连接酶及其他主要试剂为 Promega 公司和华美生物公司产品, advantage2 PCR

试剂盒购自 Clontech 公司, 质粒抽提试剂盒、胶回收试剂盒购于上海华舜公司. U251 细胞株、pEGFP-C1 载体、pGEX-4T-2 载体、大肠杆菌 JM109 和 BL21 菌株由本室保存. PCR 扩增用引物根据 Primer3 程序设计, 由大连宝生物工程公司合成. 转染用脂质体购自美国 GIBCO/BRL 公司.

1.2 生物信息学分析

运用英特网 (Internet) 上的核酸、蛋白质数据库和蛋白质分析软件 (<http://www.ncbi.nlm.nih.gov>, <http://www.expasy.ch>) 等对 LRRC4 编码的蛋白质进行结构和功能预测.

1.3 LRRC4 基因全长阅读框克隆的获得

根据 LRRC4 基因 cDNA 序列设计一组 LRRC4 编码区引物 (left: 5'-CGT GGA TCC ATG AAG CTC TTG TGG CAG-3', right: 5'-CGG GAA TTC TCA TAT TTG AGT TTC CTG-3'), 以人胎脑 cDNA 文库为模板, advantage2 PCR 试剂盒扩增 LRRC4 基因全长编码区. 为方便构建融合蛋白质粒, 在引物两端引入 BamHI 和 EcoRI 酶切位点. PCR 反应循环参数: 94 °C 30 s, 72 °C 4 min,

* 国家“十五”863 计划 (2001AA221031), 国家自然科学基金 (30100191), 湖南省自然科学基金 (00JJY20108) 资助项目.

** 通讯联系人.

Tel: 0731-4805446, E-mail: ligy@public.cs.hn.cn

收稿日期: 2002-03-04, 接受日期: 2002-03-28

30个循环。PCR产物直接克隆入pGEM-T Easy载体，酶切筛选阳性克隆，并进行DNA序列测定，获得重组质粒pGEM-T Easy/LRRC4。

1.4 LRRC4真核表达重组质粒的构建

*EcoR*I 酶切 pGEM-T Easy/LRRC4 质粒，亚克隆入 pEGFP-C1，连接产物转化 JM109 感受态细胞后小量挑选重组质粒，经适当的内切酶消化和 DNA 序列测定，筛选和鉴定表达重组体。

1.5 LRRC4融合蛋白的真核表达

将 2×10^4 个/孔 U251 细胞接种于 6 孔板中，至 80% 细胞融合，按脂质体试剂说明将 pEGFP-C1/LRRC4 表达重组体和空载体 pEGFP-C1 分别导入细胞。每种细胞转染所用质粒为 2 μ g，脂质体为 15 μ l。转染后 6 h 换完全培养基，24~36 h 在荧光显微镜下观察融合蛋白的绿色荧光。

1.6 LRRC4融合蛋白原核表达重组质粒的构建

以 LRRC4 全长阅读框质粒 pGEM-T Easy/LRRC4 质粒为模板，PCR 常规扩增引入酶切位点 *Bam*H I 和 *Eco*R I 的截短型 LRRC4 基因片段，*Bam*H I 和 *Eco*R I 酶切 PCR 产物。同时，*Bam*H I 和 *Eco*R I 消化 pGEM-T Easy/LRRC4，获得 LRRC4 阅读框。两种酶切片段分别定向克隆于 pGEX-4T-2 质粒。连接产物转化 BL21 感受态细

胞。用 *Bam*H I 和 *Eco*R I 双酶切筛选重组体，DNA 序列测定鉴定插入片段。

1.7 LRRC4融合蛋白的原核表达及鉴定

将 pGEX-4T-2/LRRC4、截短的 pGEX-4T-2/LRRC4 重组表达质粒和 pGEX-4T-2 空质粒转化 BL21 感受态细胞，0.5 mmol/L 异丙基硫代- β -D-半乳糖苷酶 (IPTG) 诱导，离心收集菌体，超声处理并煮沸变性，12% SDS-聚丙烯酰胺凝胶电泳 (SDS-PAGE) 检测融合蛋白表达。

2 结 果

2.1 LRRC4基因编码的蛋白质序列特征

开放阅读框架分析 LRRC4 基因含 1 962 bp 完整阅读框架，编码 653 个氨基酸^[1]。预测分子质量：72.7 ku，理论等电点：6.58，为一略偏酸性的蛋白质。含多个保守的富亮氨酸重复结构域 (LRR) 和一个 C2 型免疫球蛋白结构 (IgC2) 因而被人类基因组 HUGO 国际命名委员会命名为 LRRC4。多种程序分析发现，LRRC4 拥有两个疏水区，分别代表一个 N 端信号肽 (氨基酸第 1~42 位) 和 C 端跨膜序列 (氨基酸第 528~550 位)，表明 LRRC4 为一膜蛋白。图 1 为 SMART 预测的 LRRC4 蛋白的结构图。

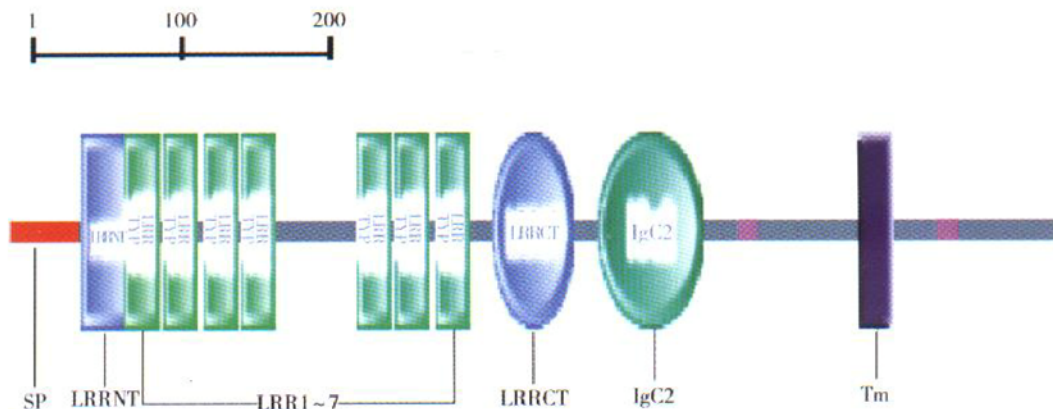


Fig. 1 Structure of deduced LRRC4 protein predicted by SMART program

SP: signal peptide; LRR1~7: seven leucine-rich repeats; LRRNT: leucine rich repeat N-terminal domain; LRRCT: leucine rich repeat C-terminal domain; IgC2: immunoglobulin C2 domain; Tm: transmembrane region.

2.2 LRRC4基因全长阅读框的克隆与鉴定

PCR 产物在 1.5% 琼脂糖凝胶电泳中出现单一条带约 2.0 kb，与设计基因大小一致。PCR 产物回收后直接克隆于 pGEM-T Easy 载体。*Eco*R I 酶切鉴定插入片段大小。如图 2 所示，阳性克隆酶切后出现两条带，分别代表 T 载体和 LRRC4 阅读

框。随机挑选两个阳性克隆进行 DNA 序列测定及决定插入方向。序列分析表明重组质粒包含完整 LRRC4 基因全长阅读框，并且在 5' 端和 3' 端分别引入了酶切位点 *Bam*H I (GGATTC) 和 *Eco*R I (GAATCC)。其中正向克隆起始密码子位于核苷酸第 114~116 位，终止密码子位于核苷酸第 2 073

~ 2 075 位. 选择正向克隆做进一步研究.

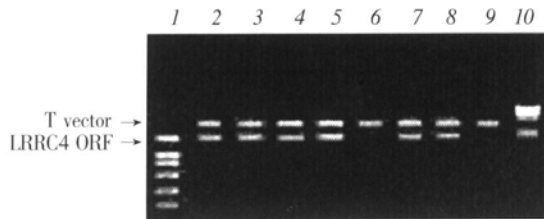


Fig. 2 Identification of pGEM-T Easy/LRRC4 recombinant by *Eco*R I digestion

1: DL 2000 marker; 2~ 5, 7~ 8: positive recombinant; 6, 9: negative recombinant; 10: λ DNA/*Hind* III marker.

2.3 LRRC4 融合表达重组体的构建

*Eco*R I 酶切 pGEM-T Easy/LRRC4 重组质粒获得 LRRC4 全长阅读框, 克隆于经 *Eco*R I 酶切后去磷酸化的 pEGFP-C1 载体. 经 *Bam*H I 酶切鉴

定插入片段大小和方向, 并经过 DNA 序列确证性分析, 获得如图 3a 所示 LRRC4 真核表达质粒 pEGFP-C1/LRRC4.

以 LRRC4 全长阅读框重组质粒 pGEM-T Easy/LRRC4 为模板, PCR 扩增截短型 LRRC4, 获得约 500 bp 的 LRRC4 基因片段 (主要为 IgC2 结构组成), 序列分析表明与设计一致. *Bam*H I 和 *Eco*R I 消化 PCR 产物和 pGEM-T Easy/LRRC4, 分别克隆入经同样酶切修饰的原核表达载体 pEGX-4T-2. 经 *Bam*H I 和 *Eco*R I 酶切鉴定插入片段大小和 DNA 序列分析, 获得如图 3b 和图 3c 所示 LRRC4 原核表达重组质粒 pEGX-4T-2/LRRC4 和截短型重组质粒 pEGX-4T-2/mLRRC4. 两种质粒均含有正确的插入片段, 且阅读框无移码.

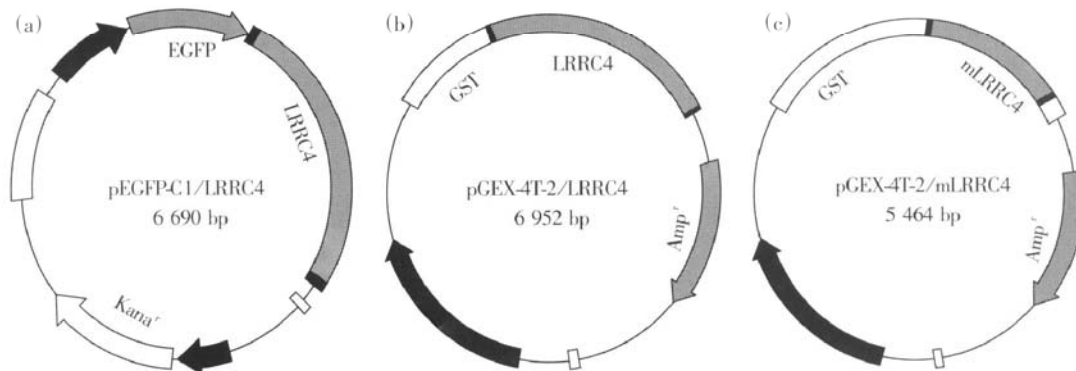


Fig. 3 LRRC4 fusion plasmids

(a) pEGFP-C1/LRRC4; (b) pGEX-4T-2/LRRC4; (c) pGEX-4T-2/mLRRC4.

2.4 LRRC4 融合蛋白在细胞中的表达和定位

分别将重组表达质粒 pEGFP-C1/LRRC4 和 pEGFP-C1 空白质粒经脂质体介导瞬时转染 U251

细胞, 在转染后 24~ 36 h, 荧光显微镜观察绿色荧光的分布, 发现转染空白载体 pEGFP-C1 细胞的荧光在细胞中均匀存在, 而转染重组的 LRRC4 融合

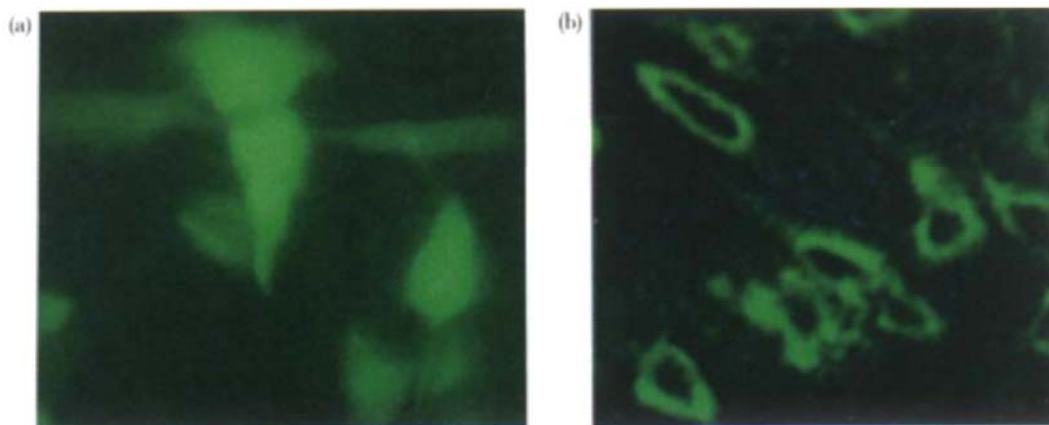


Fig. 4 Expression of LRRC4 fusion protein in U251 cell

(a) expression of pEGFP-C1/U251; (b) expression of pEGFP-C1/LRRC4/U251.

蛋白的细胞发出的荧光集中分布在细胞膜上(图4)。此结果说明 LRRC4 融合蛋白定位于细胞膜,与生物信息学分析结果一致。

2.5 融合蛋白在大肠杆菌中的表达

两种重组质粒和空白载体质粒 pGEX-4T-2 分别转化大肠杆菌 BL21, 经 0.5 mmol/L IPTG 诱导, SDS-PAGE 检测融合蛋白的表达。与未诱导组比较, 在含有两种外源基因表达载体的大肠杆菌中, 与融合蛋白相应的蛋白质分子质量处有特异的

蛋白质条带, 分别为 99 ku (26 ku+ 72.7 ku) 和 43 ku (26 ku+ 17 ku)。而 GST 蛋白处的蛋白质条带 (26 ku) 消失, 表明新增条带即为 GST/LRRC4 融合蛋白。而 pGEX-4T-2 空质粒转化的细菌则在 GST 蛋白 (26 ku) 相应位置出现新的条带(图5)。LRRC4 全长编码框的融合蛋白量较少, 而截短型的 LRRC4 融合蛋白表达量较高, 且蛋白量随时间延长而增长, 在诱导后 5 h 蛋白质表达量约占细菌总蛋白含量的 15%~20%。

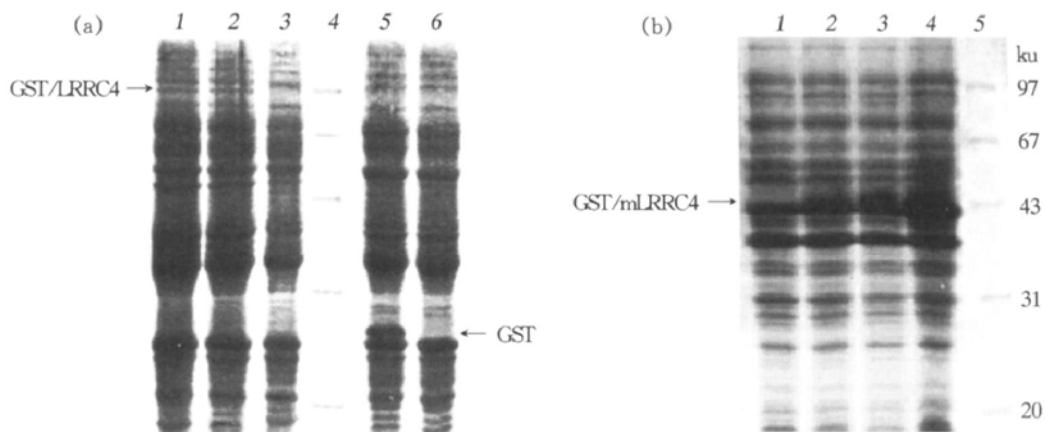


Fig. 5 SDS PAGE analysis of the expression of fusion protein of LRRC4

(a) Expression of LRRC4 and GST fusion protein. 1~3: GST/LRRC4/BL21, 3 h, 2 h and 0 h after induction; 4: low molecular mass protein marker; 5, 6: GST/BL21, 2 h and 0 h after induction. (b) Expression of truncated LRRC4 fusion protein. 1~4: GST/mLRR4/BL21, 0 h, 2 h, 3 h and 5 h respectively after induction; 5: low molecular mass protein marker.

3 讨 论

随着人类基因组草图序列的完成, 生命科学研究的中心逐渐由基因克隆转向基因功能的研究。而生物信息学的飞速发展给基因功能的研究带来了极大的便利。借助于蛋白质序列分析软件的开发及蛋白质相关信息的积累。人们可通过各种软件分析新基因的结构, 新基因编码蛋白质与已知蛋白质的联系, 为进一步研究新基因的功能提供重要的线索。

生物信息学分析 LRRC4 编码蛋白质的分子特征, 表明 LRRC4 含有两个保守的结构域——富亮氨酸重复 (leucine-rich repeats, LRR) 和 IgC2^[11], 是一个在多种模式生物如小鼠、大鼠、果蝇、芥兰、线虫和酵母中高度保守的 LRR 超家族新成员。

LRR 是一类在进化上高度保守的结构域^[2]。它通常由 22~26 个氨基酸残基组成, 其保守的序列为: LxxLxLxxNxL。研究表明所有 LRR 蛋白的功能均与蛋白质相互作用有关, 至少一半的 LRR 蛋白参与信号转导通路^[2-6]。IgC 2 也是一类非常

重要的功能域, 它在多种组织中参与细胞间的粘附和发育中的定向诱导^[7,8]。目前只发现少数几种分子同时含有这两种结构域, 如 trk 受体^[9]、LIG-1^[10]、peroxidasin^[11,12]、Pal^[13] 和 ZfNLRR^[14]。对它们的研究表明它们均在神经系统的信号途径中扮演重要角色。如 trk 受体能与多种神经营养因子如 NGF, BDNF, NT-3 和 NT-4 等结合, 活化细胞内的多条信号通路如 Ras/ERK、NF-kappaB、P13K/Akt 和 MAPK 等, 影响神经系统的发育和功能^[9]。LIG-1 在神经胶质细胞中特异表达^[10], peroxidasin^[11,12] 在果蝇神经系统的发育中发挥重要作用, 而 Pal 在视网膜的感光细胞中表达, 作为营养因子或粘附分子的受体而存在^[13]。ZfNLRR, 一个包含 IgC2 的神经亮氨酸分子, 作为一种神经系统特异性的黏附分子或可溶性受体参与神经系统损伤后的再生^[14]。

pGFP-C1 是一种穿梭载体。当插入的外源基因与其阅读框一致时, 形成的表达重组体能在巨细胞病毒启动子的启动下, 表达 GFP 和外源基因的融合蛋白。该融合蛋白保留了 GFP 的荧光特性,

因而能方便地进行活体定位观察。本研究通过构建 pEGFP-C1/LRRC4 表达重组体, 瞬时转染 U251 细胞, 成功地观察到 LRRC4 融合蛋白在活的哺乳动物细胞中的表达及细胞膜的定位, 结合 LRRC4 基因表达和氨基酸序列特征, LRRC4 可能作为一种粘附分子或受体参与多种信号途径, 而影响中枢神经系统的发育和正常功能维持。筛选 LRRC4 相互作用的蛋白质, 研究 LRRC4 在神经系统的信号途径中的位置, 将有助于揭示脑的发育及脑瘤分子机制的阐明。

基因的作用最终通过其编码的蛋白质分子完成, 因此分离蛋白质及相应结构域, 制备单抗是澄清一个基因功能及其活性部位的重要手段。本研究在构建 LRRC4 全长融合蛋白表达重组体的同时, 根据 LRRC4 的二级结构特征和亲、疏水性特点, 选择了一段主要由 IgC2 组成, 且具良好抗原性的序列构建截短的 LRRC4 融合表达载体。考虑到目的蛋白是新基因编码的蛋白质, 选用简便、高效及生产成本低的原核表达系统, 观察融合蛋白的表达。SDS-PAGE 结果表明 LRRC4 融合蛋白及其截短型融合蛋白均能在大肠杆菌 BL21 中表达。在含空载体的大肠杆菌中, GST 相应分子质量 (26 ku) 处有 GST 蛋白的表达, 在含有外源基因表达载体的大肠杆菌中, 与融合蛋白相应的蛋白质分子质量处有特异的蛋白质条带出现, 而 GST 蛋白处的条带消失。与 LRRC4 全长编码区相比, 截短型融合蛋白能在大肠杆菌中高效表达。此结果为分离与纯化 LRRC4 融合蛋白奠定了基础, 截短型融合质粒的构建也有利于研究 LRRC4 发挥作用的功能活性部位。进一步优化实验条件获得高产量的 LRRC4 全长蛋白及各种截短型 LRRC4 蛋白的工作正在进行中。

参 考 文 献

1 王洁如, 钱 骏, 李桂源, 等. 富亮氨酸重复超家族新成员

LRRC4 的克隆及在脑瘤中的表达分析. 生物化学与生物物理进展, 2002, 29 (2): 233~ 239

WANG J R, QIAN J, LI G Y, *et al.* Prog Biochem Biophys, 2002, 29 (2): 233~ 239

- 2 Kobe B, Deisenhofer J. The leucine rich repeat: a versatile binding motif. Trends in Biochemical Science, 1994, 19 (10): 415~ 421
- 3 Kobe B, Deisenhofer J. A structural basis of the interactions between leucine rich repeats and protein ligands. Nature, 1995, 374 (6518): 183~ 186
- 4 Kobe B, Kajava A V. The leucine rich repeat as a protein recognition motif. Curr Opin Struct Biol, 2001, 11 (6): 725~ 732
- 5 Fetchko M, Huang W, Li Y, *et al.* Drosophila Gp150 is required for early ommatidial development through modulation of Notch signaling. EMBO J, 2002, 21 (5): 1074~ 1083
- 6 Dong J, Li C, Schade A J, *et al.* Mutation in the leucine rich repeat of platelet glycoprotein Ib alpha results in defects in its interaction with immobilized von Willebrand factor under flow. Chin Med J (Engl), 2000, 113 (8): 693~ 698
- 7 Williams A F, Barclay A N. The immunoglobulin superfamily domains for cell surface recognition. Annu Rev Immunol, 1988, 6: 381~ 405
- 8 Mendoza H L, Fave I. Physiological aspects of the immunoglobulin superfamily in invertebrates. Dev Comp Immunol, 1999, 23 (4~ 5): 359~ 374
- 9 Potapoutian A, Reichardt L F. Trk receptors: mediators of neurotrophin action. Curr Opin Neurobiol, 2001, 11 (2): 272~ 280
- 10 Yutake S, Naoya S, Masaya T, *et al.* cDNA cloning of a novel membrane glycoprotein that is expressed specifically in glial cells in the mouse brain (LIG-1, a protein with leucine rich repeats and immunoglobulin like domains). J Biol Chem, 1996, 271 (37): 22657~ 22662
- 11 Nelson R E, Fessler L I, Takagi Y, *et al.* Peroxidase: a novel enzyme matrix protein of *Drosophila* development. EMBO J, 1994, 13 (15): 3438~ 3447
- 12 Sritunyalucksana K, Wongsuebsantati K, Johansson M W, *et al.* Peroxinectin, a cell adhesive protein associated with the proPO system from the black tiger shrimp, *Penaeus monodon*. Dev Comp Immunol, 2001, 25 (5-6): 353~ 363
- 13 Gomi F, Imaizumi K, Yoneda T, *et al.* Molecular cloning of a novel membrane glycoprotein, pal, specifically expressed in photoreceptor cells of the retina and containing leucine rich repeat. J Neurosci, 2000, 20 (9): 3206~ 3213
- 14 Bormann P, Roth L W, Andel D, *et al.* ZfNLRR, a novel leucine-rich repeat protein is preferentially expressed during regeneration in zebrafish. Mol Cell Neurosci, 1999, 13 (3): 167~ 179

Preliminary Study of LRRC4 Protein: Bioinformatic Analysis, Fusion Expression in Eukaryote and Prokaryote*

WANG Jie-Ru¹⁾, DONG Li¹⁾, JIANG Ming²⁾, TAN Chen¹⁾, LI Xiao-Ling¹⁾,
XIANG Juan-Juan¹⁾, FAN Song-Qing¹⁾, PENG Cong¹⁾, TANG Ke¹⁾, LI Gui-Yuan^{1)**}

¹⁾Xiangya School of Medicine, Central South University, Changsha 410078, China;

²⁾People Hospital of Shenzhen Lantian District, Shenzhen 518081, China)

Abstract In previous study, a novel gene, LRRC4, a member of leucine-rich repeat (LRR) superfamily was cloned. Expression analysis indicated that LRR may play an important role in the central nervous system. To investigate the function and the structure-function relationship of LRRC4, full length coding region was amplified and subcloned into pGEM T Easy vector. Further, the recombinant plasmid, pEGFP-C1/LRRC4, was constructed and transfected transiently into U251 cell. Under the fluorescence microscope, the green fluorescence produced by LRRC4 fusion protein was observed on the cytoplasmic membrane. Consistent to prediction by bioinformatics, this result indicated that product of LRRC4 is a membrane protein. In addition, the recombinant of LRRC4, pGEX-4T-2/LRRC4 and truncated LRRC4 recombinant, pGEX-4T-2/mLRRC4, were constructed and transformed into *E. coli* BL21. Induced by 0.5 mmol/L IPTG, The band corresponding to fusion protein were observed in SDS-PAGE as expected. Together with bioinformatic analysis of LRRC4 protein, these results establish the basis for functional study of LRRC4.

Key words LRRC4, fusion protein expression, EGFP, GST, bioinformatics

* This work was supported by grants from The State 863 High-Tech R&D Program (2001AA221031); The National Natural Science Foundation of China (30100191) and Hunan Province Natural Science Foundation (00JJY20108).

** Corresponding author. Tel: 86-731-4805446, E-mail: ligy@public.cs.hn.cn

Received: March 4, 2002 Accepted: March 28, 2002