

绵羊 QM 基因的电子克隆及其生物信息学分析

尹珊珊, 吴登俊, 张翔宇, 周明亮 (四川农业大学动物科学与技术学院, 四川雅安 625014)

摘要 [目的]为研究绵羊 QM 基因的生物学功能奠定基础。[方法]利用 EST 数据库和电子克隆等技术克隆绵羊 QM 基因, 通过生物信息学分析预测其序列结构, 检测其在绵羊不同组织中的表达情况并分析其与其他 14 物种 QM 基因的同源性。[结果]首次成功克隆了绵羊 QM 基因, 全长 771 bp, 含有 1 个最长为 645 bp ORF, 编码的蛋白分子量为 24.61 kD, 为碱性蛋白质。QM 蛋白的二级结构含有 6 个 α -螺旋和 10 个 β -折叠, 其余为无规卷曲。QM 基因在绵羊脾脏、脑组织和骨组织中较好表达, 在其余 7 种组织中也均有表达。QM 基因在病变组织中表达量比正常组织高。在 14 个物种中绵羊 QM 基因与牛的进化距离最短。[结论]QM 基因与肿瘤抑制、细胞生长、分化、发育和凋亡等有关, 且有极高保守性。

关键词 绵羊; QM 基因; 电子克隆; 表达序列标签

中图分类号 S826 文献标识码 A 文章编号 0517-6611(2008)14-05767-03

Analysis of Silico Cloning of QM Gene in Sheep and Its Bioinformatics

YIN Shan-shan et al (College of Animal Science and Technology, Sichuan Agricultural University, Ya'an, Sichuan 625014)

Abstract [Objective] The research aimed to lay the foundation for studying the biological functions of QM gene in sheep. [Method] QM gene in sheep was cloned by using EST database, in silico cloning, etc. and its structure sequence was predicted through bioinformatics analysis. Its expression in different tissue of sheep was detected and its homology with 14 other species were compared. [Result] QM gene in sheep was cloned for the 1st time, with total length of 771 bp, containing an ORF with the longest length of 645 bp. Its molecular weight of the coded protein was 24.61 kD, being alkaline. The secondary structure of QM protein contained 6 α -helices and 10 β -folding, others were random coils. QM gene expressed in sheep spleen, brain tissue and bone tissue better, while that also expressed in 7 other tissues. The expression quantity of QM gene in diseased tissues was higher than that in normal tissues. Among 14 species, evolutionary distance of QM gene between sheep and cattle was shortest. [Conclusion] QM gene was related with tumor suppression, cell growth, differentiation, development, death, etc. and it had extremely high conservation.

Key words Sheep; QM gene; In silico cloning; Expressed sequence tag

基因的电子克隆(silico cloning)即利用计算机来协助克隆基因,是与定位克隆、定位候选克隆策略并列的方法之一。EST(Expressed Sequence Tag)表达序列标签是指从不同组织来源的 cDNA 文库中随机挑选克隆,对其 5' 或 3' 端测序后得到的部分 cDNA 序列,长度一般在 300~500 bp^[1]。电子克隆采用生物信息学的方法延伸 EST 序列^[2],以获得基因部分乃至全长的 cDNA 序列。EST 数据库的迅速扩张,已经并将继续导致识别与克隆新基因策略发生革命性变化。

QM 基因是 Weissman 等首次从人 Wilm's tumor 相关的细胞系中分离出来的一种基因^[3],是一种与肿瘤抑制,细胞生长、分化、发育和凋亡等有关的重要基因。QM 同源基因广泛地存在于动物、植物以及真菌界,目前已在小鼠、鸡、果蝇、牛、酵母、玉米和水稻等生物中克隆了 QM 同源基因,这些基因的核苷酸及其编码的氨基酸序列高度保守。在鸡中, QM 同源基因 Jif-1 可与癌基因 c-Jun 的蛋白产物结合从而抑制 c-Jun 的反式激活转录调节活性,影响细胞生长与增殖^[4];在酵母中, QM 同源基因 GRC5/QSR1 与线粒体呼吸作用有关,而且 QM 基因突变,可导致酵母蛋白质合成下降、生长和分裂阻滞、细胞骨架异常, QM 基因缺失,则导致酵母死亡^[5];在小鼠胚胎发育过程中, QM 参与基因转录和转译后的调节,与组织细胞分化关系密切^[6]。此外, QM 还与 c-Yes 和其他 Src 家族成员相互作用,参与细胞信号的转导。已有研究表明 QM 是一种多功能蛋白,参与细胞生长、分化、发育和凋亡等基本生命活动。迄今,尚未见绵羊中 QM 基因的研究报道。因此,笔者应用生物信息学电子克隆等技术,首次从绵羊中克隆出 QM 基因(GenBank 登陆号: EU143791),

并对其进行了生物信息学分析。

1 材料与方法

1.1 主要数据库 美国国立生物技术信息中心(National Center for Biotechnology Information, NCBI) GenBank 数据库(<http://www.ncbi.nlm.nih.gov/Database/index.html>); 绵羊基因组数据库(<http://www.ncbi.nlm.nih.gov/projects/genome/guide/sheep/>); UniGene(<http://www.ncbi.nlm.nih.gov/sites/entrez?db=unigene>)。

1.2 生物信息学软件 相似性搜索: BLAST(<http://www.ncbi.nlm.nih.gov/BLAST>); 重叠群组装: CAP3, 可以通过 www.genome.cs.mtu.edu/cap/cap3.html 获取; ORF 预测: ORF Finder; 核酸及蛋白质组成分析: Bioedit, DNASTar; 蛋白质二级结构预测: 来自波兰的在线预测程序 Bio-Informatique Lyonnais, (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_or4.html); DDD: <http://www.ncbi.nlm.nih.gov/UniGene/ddd.cgi>? TAX=9940; 序列比对: Clustx, Clustw; 绘制进化树: Clustx, Treeview。

1.3 试验方法

1.3.1 绵羊 QM 基因的电子克隆。考虑到物种间进化上的亲缘性,故采用牛的 QM 基因(GenBank 登陆号: AF143815.1)为探针进行绵羊 QM 基因的克隆。首先进行 BLAST 搜索^[7],输入探针序列,选择绵羊 dbEST 进行 blastn。截至 2007 年 9 月,绵羊 EST 数据库有 186 678 条序列。设定期望值 E 为 0.01, Description 和 Alignment 为 500。将检索到的匹配结果 E 值小于 0.01 的绵羊 mRNA 序列全部下载到本地计算机,建立一个本地绵羊 dbEST 数据库。用 CAP3 序列组装程序将序列组装成重叠群(contigs),以获得的重叠群为被检序列,重复进行 BLAST 检索与序列组装,延伸重叠群序列,重复以上过程,直到没有更多的重叠 EST 检出或者说重叠群序列

作者简介 尹珊珊(1982-),女,四川眉山人,硕士研究生,研究方向:生物信息学。

收稿日期 2008-03-10

不能继续延伸,即获得绵羊 QM 基因 cDNA 全长的基因编码序列^[6]。

1.3.2 绵羊 QM 基因的生物信息学分析。使用 ORF Finder 进行绵羊 QM 基因 ORF 的预测,使用 Bioedit 进行酸碱基组成,加尾信号等分子生物学信息分析及蛋白质组成分析,获得等电点、蛋白质二级结构等。使用 NCBI 提供的数字化差异表达分析(Digital Differential Display, DDD)工具^[9],检测 QM 基因在绵羊身体不同组织的表达情况。

1.3.3 QM 基因同源分析。从 GenBank 中检索到 14 种生物 QM 同源基因序列,将其下载到本地计算机。使用 Clustx 以及 Clustw 软件进行多序列比对。使用 Clustx 和 Treeview 软件进行包括绵羊在内的 15 种生物 QM 基因的分子进化树构建。

2 结果与分析

2.1 绵羊 QM 基因的序列结构预测 以牛 QM 基因序列为信息探针,通过运行 BLAST 程序搜索绵羊 EST 数据库,得到 244 条同源性≥90%的序列(图 1),然后使用 CAP3 进行拼接,得到 contig0。用 contig0 再次搜索绵羊 dbEST,得到 5 条同源性等于 99%的序列,再次拼接,最终两端不能再延伸,获得一个同源性很高的绵羊 cDNA 序列,命名为 SQM。SQM 全长 771 bp,含有一个最长为 645 bp 的完整的开放阅读框(图 2),5'UTR 长度为 70 bp,第 71 位到 716 位为 ORF,长 645 bp,编码 214 个氨基酸,3'UTR 长度为 55 bp,在 754~759 bp 区域内有一个 AATAAA 加尾信号,证明 SQM 即为绵羊 QM 基因。SQM 基因编码的多肽,通过 Bioedit 软件分析预测其分子量大小为 24.61 kDa,等电点 pI 为 10.6,其中 Asp 和 Glu 酸性氨基酸残基数为 7.9%,碱性氨基酸 Arg 和 Lys 残基数为 20.92%,表明 SQM 蛋白为碱性蛋白质。



图 1 BLASTN 结果 Fig. 1 The results of BLASTN

SQM - 72..677

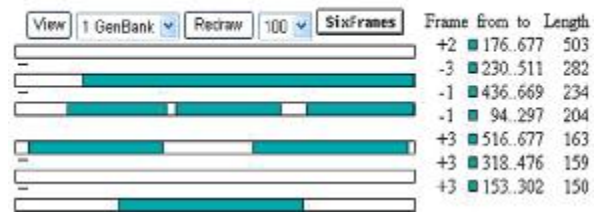


图 2 ORF Finder 结果

Fig. 2 The results of ORF Finder 通过 Bio-Informatique Lyonnais 的在线预测程序,预测 SQM 蛋白二级结构有 6 个 α 螺旋区,富含 10 个 β 折叠片,其余部分为无规卷曲(图 3)。

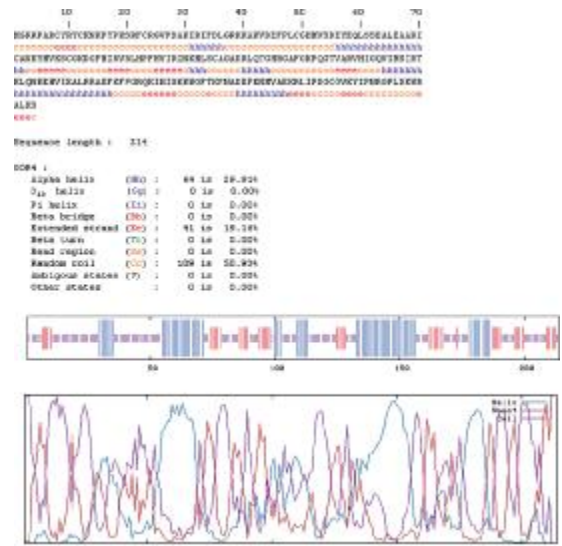


图 3 SQM 蛋白质二级结构 Fig. 3 Secondary structure of SQM protein

2.2 绵羊 QM 基因的数字化差异表达分析 以 SQM 基因的全长 cDNA 对绵羊 EST 数据库进行 BLASTN 搜索,设定期望值,获得 100 条 E=0 的绵羊 EST 序列。对这 100 条序列的来源组织文库进行搜索分析,然后进行数字化差异表达分析(DDD),得到 Oar.13010EST 簇。该 EST 簇与牛 NP001 069503.1protein 有极高的相似性。可以认为是绵羊 QM 基因 EST 文库。该 EST 簇共有 209 条 EST,其中 153 条来自于骨组织,17 条来源于消化道,5 条来源于雌性生殖器,4 条来源于淋巴网状内皮细胞,1 条来源于乳腺,10 条来源于脾脏和脑组织的混合文库,2 条来源于肌肉组织,15 条来源于皮肤组织,1 条来源于后肢。数字化差异表达分析显示,其表达率为 0.003 7、0.001 0、0.000 5、0.000 2、0.000 1、0.000 5、0.003 9、0.000 9、0.000 5。说明 SQM 基因在脾脏、脑组织、骨组织中有较强表达,在其余 7 种组织中也均有表达(图 4)。

A	B	C	D	E	F	G	H	I
0.0037	0.0010	0.0005	0.0002	0.0001	0.0005	0.0039	0.0009	0.0005

注:A 为骨组织;B 为消化值;C 为雌性生殖器;D 为后肢;E 为淋巴网状内皮细胞;F 为乳腺;G 为绵羊脾脏和脑组织混合文库;H 为肌肉组织;I 为皮肤组织。
Note: A.Bone; B. Digestive value; C. Female genital; D. Hindlimb; E. Lymphoreticular; F. Mammary gland; G. Sheep spleen\brain pSport1 library; H. Muscle; I. Skin.

图 4 数字化差异表达结果 Fig. 4 The results of DDD

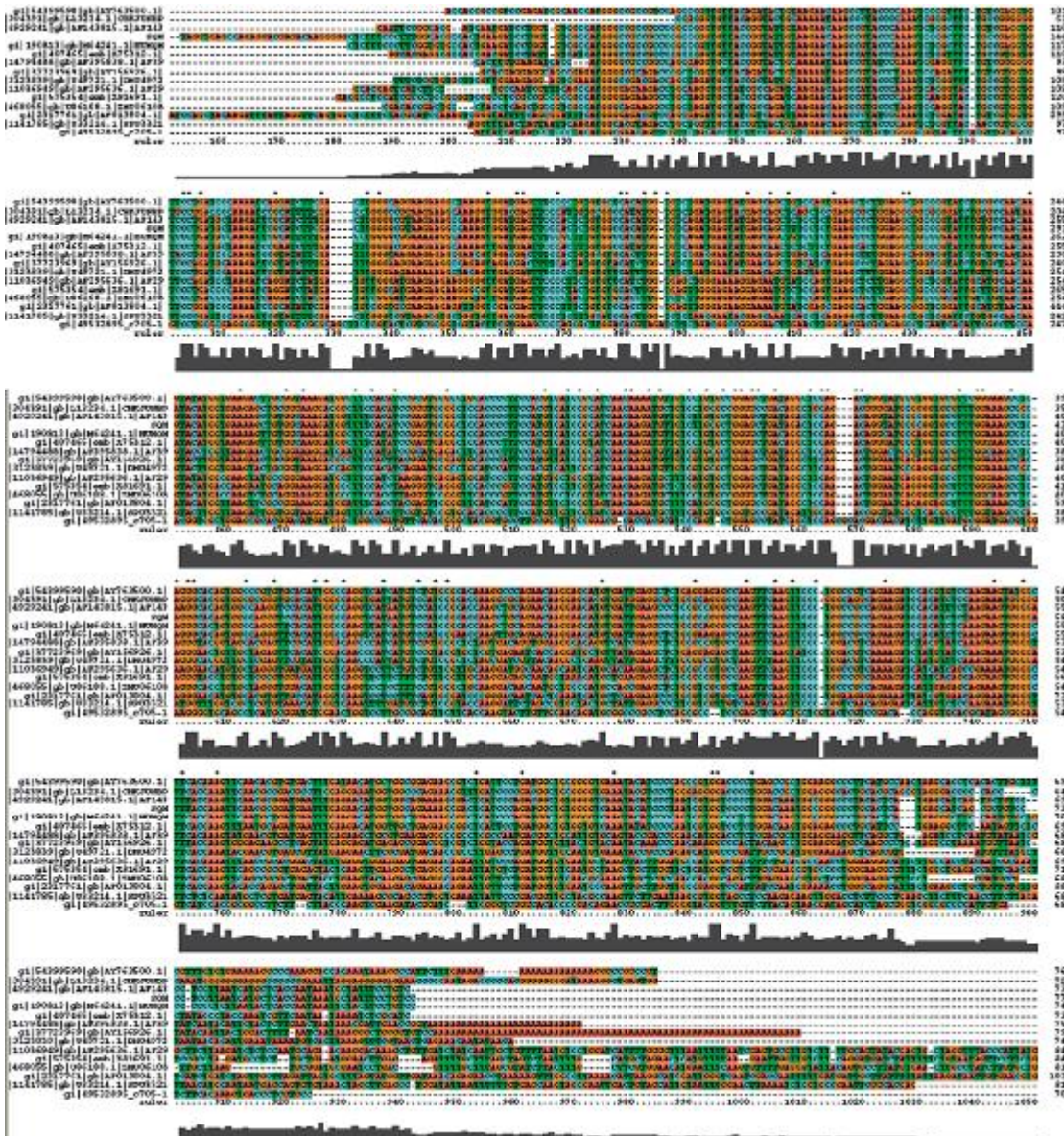


图 5 15 种生物 QM 基因序列比对结果
 Fig. 5 The result of QM gene sequences alignment of 15 species

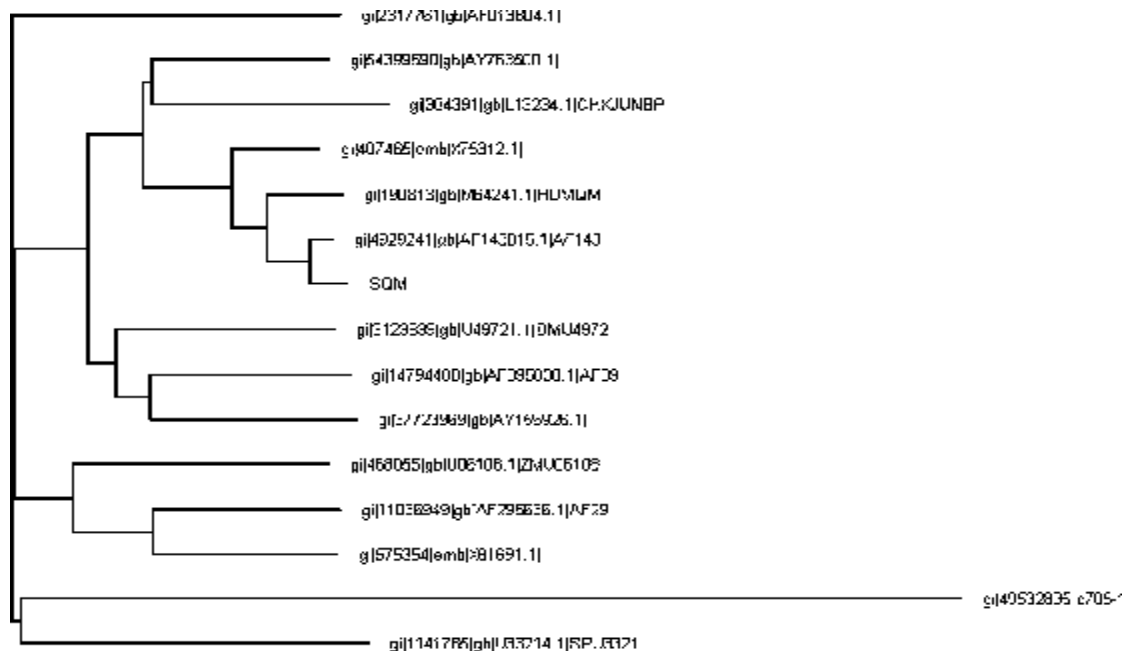


图 6 15 种生物 QM 基因所绘制的进化树
 Fig. 6 Phylogenetic tree of QM gene of 15 species

(下转第 5898 页)

病猴第3天开始饮水,第4天饲喂其香蕉,第5天见其肛门周围有少量稀粪,第6天排出大量稀粪,伤口愈合良好,恢复正常饲喂。7d后康复出院。一个月后随访,一切正常。

3 小结与讨论

(1)依据该病的发生部位、外观和特征性的临床表现,极易作出诊断。但必须判断脱出肠管中是否有套叠现象,若忽视诊断,仅将脱出肠管整复,容易造成不良后果。鉴别单纯性直肠脱和套叠性直肠脱的方法:一是触压早期脱出的肠管,前者整体空虚感强,后者可触及一段坚实、无弹性的香肠样肠管。此外,也可进行消化道灌服硫酸钡X射线鉴别,能够对肠套叠作出准确诊断。若为轻微套叠,手术时可将套叠部位用手指适当挤压还纳入腹腔,然后及时插入钢针进行手术,可有效防止套叠复发。

(2)对于直肠脱,应根据具体情况及早进行治疗。对于脱出过多、整复有困难、脱出的直肠发生坏死、穿孔或有套叠而不能复位的病例,多采用手术切除法,有直肠部分切除术,黏膜下层切除术^①。对于发病初期或黏膜性脱垂的病例可使用单纯的肠管整复术,而且应尽可能在直肠壁及肠周围蜂窝组织未发生水肿以前施行。

(下转第5769页)

另外,通过分析获得的100条序列的具体来源组织器官,发现其在病变组织中的表达量比正常组织中的表达量有所提高。

2.3 QM基因同源分析 在GenBank上检索到14种生物的QM同源序列,它们是:人、家鼠、斑马鱼、牛、珍珠贝、家蚕、鸡、菱形斑纹蛾、果蝇、酵母、火炬松、玉米、水稻、非洲油棕榈。通过Clustx进行多序列比对,并用Treeview绘制进化树(图5、6)。来源相同的物种在进化树上往往聚集成簇,序列间的遗传距离对应于进化树中连接2个物种水平线的长度。结果显示,QM基因可归为4大类,包括脊椎动物、无脊椎动物、酵母和植物,其中绵羊QM基因和牛QM基因在进化上的距离最短,同时也发现QM基因具有极大的保守性。

3 结论与讨论

(1)生物信息学是在生命科学的研究中,以计算机为工具对生物信息进行储存、检索和分析的科学。它是当今生命科学和自然科学的重大前沿领域之一,通过检索数据库进行核酸序列同源性检索,在电子基因定位、电子延伸、电子克隆和电子表达以及蛋白质功能分析、基因鉴定等方面起到了重要作用。充分利用生物信息学资源和技术研究基因组结构与功能、发现新的功能基因、揭示基因起源与分子进化规律等已成为后基因组时代的核心任务。

(2)目前,通过数据库查询、cDNA文库直接测序、mRNA差别显示(DDRT-PCR)、代表性差异分析(RDA-PCR)和抑制差减杂交(SSH)等方法获得的EST数据越来越庞大,其在基因结构与功能研究中的应用日益广泛。传统的测序技术通常要通过筛选cDNA文库,或者通过RACE等试验手段去获得新基因的亚克隆,并对亚克隆进行排序,这些工作需要大量的人力物力。相比较而言,利用基因组数据库电子克隆就成为功能基因克隆的新途径,生物信息工作者可以利用来自世界各地不同实验室的序列数据进行电子克隆及生物信息学分析,并将结果上传到互联网上共享,这样可避

(3)钢针“十”字交叉固定肠管前,也可先向直肠内插入一根长短适宜的橡胶管或塑料管,这样有利于后面的缝合,层次分明、不易混淆。也可用带胶套的肠钳夹住脱出的肠管,兼有固定和止血作用。

(4)手术最后闭合肛门时,要不松不紧,既能防止直肠再度脱出,又不影响正常排便。

(5)检查缝合效果也很重要,缝合完毕时用手指伸入直肠触摸钢针,若能顺利摸到两根钢针说明达到预期目的;也可将直肠回复入腹腔后,用手指触摸直肠壁,若其自动展开,触诊与正常状态相似,平滑顺畅,则达到预期目的。

参考文献

- [1] 侯加法.小动物疾病学[M].北京:中国农业出版社,2002:9-72.
- [2] 彭广能.兽医外科学[M].四川:四川科学技术出版社,2004:225-229.
- [3] 李志霞,韩志刚.直肠脱垂的外科治疗[J].中华胃肠外科杂志,2004(4):115-121.
- [4] 林德贵.家畜外科手术学[M].北京:中国农业出版社,2002:9-72.
- [5] 汪世昌.家畜外科学[M].北京:农业出版社,1999:3-115.
- [6] AMERICAN College of Veterinary Surgeons. Veterinary surgery[J]. Philadelphia: W B Saunders, 2001, 30(3):33-36.
- [7] 邢兰君.手术治疗猪直肠脱[J].畜牧与兽医,2005(10):63-64.
- [8] 王承东,李德生,何长贵,等.大熊猫直肠脱及继发肠梗阻护理探讨[J].四川畜牧兽医,2007(5):25-28.

免重复多余的劳动。

(3)在该研究中,通过利用日益增长的绵羊EST数据库,根据物种间同源基因的保守特性,以牛的QM基因cDNA序列为模板,通过EST拼接等生物信息学手段,成功地电子克隆出绵羊QM基因。数字化差异显示分析是运用计算机技术对代表某一基因的、源于不同组织、不同生理状态或在特定发育阶段表达的EST序列数据进行分析,从而绘制出该基因在不同时空差异表达的图谱,笔者利用该方法,分析了SQM的组织表达规律,再次证明QM基因是一种与肿瘤抑制、细胞生长、分化、发育和凋亡等有关的重要基因。

参考文献

- [1] ADAMS M D, KELLEY J M, GOCAYNE J D, et al. Complementary DNA sequencing: Expressed sequence tags and human genome project[J]. Science, 1991, 225: 1651-1656.
- [2] JONATHAN PEVSNER.生物信息学[M].孙之荣,译.北京:化学工业出版社,2006,17-19.
- [3] WEISSMAN B, STANBRIDE E, FARMER A, et al. Molecular characterization of QM, a novel gene with properties consistent with tumor suppressor function [J]. Cold Spring Harb Symp Quant Biol, 1994, 59: 573-576.
- [4] FARMER A A, LOFTUS T M, MILLS A A, et al. Extreme evolutionary conservation of QM, a novel c-Jun associated transcription factor[J]. Hum Mol Genet, 1994, 3: 723-728.
- [5] KOLLER H T, KLADE T, ELLINGER A, et al. The yeast growth control gene GRC5 is highly homologous to the mammalian putative tumor suppressor gene QM[J]. Yeast, 1996, 12(1): 53-56.
- [6] MILLS A A, MILLS M J, GARDINER D M, et al. Analysis of the pattern of QM expression during mouse development[J]. Differentiation, 1999, 64(3): 161-171.
- [7] ALTSCHUL, STEPHEN F, THOMAS L, et al. Gapped BLAST and PSI-BLAST a new generation of protein database search programs [J]. Nucleic Acids Res, 1997, 25: 3389-3402.
- [8] GILL R W, SANSEAU P. Rapid in silicon cloning of genes using expressed sequence tags (ESTs) [J]. Biotechnol Annu Rev, 2000, 5: 25-44.
- [9] DAVID MEINTRUP, ELLEN REISINGER. A statistical model providing comprehensive predictions for the mRNA differential display[J]. Bioinformatics, 2005, 21(15): 3880.