

[19]中华人民共和国专利局

[51]Int.Cl⁶



[12] 发明专利申请公开说明书

C07H 21/02
C07H 21/04 C12P 21/02
C12N 15/63 C12N 15/12
A61K 38/17

[21] 申请号 96121928.9

[43]公开日 1998年5月6日

[11] 公开号 CN 1180709A

[22]申请日 96.10.21

[71]申请人 湖南医科大学

地址 湖南省长沙市北站路22号

[72]发明人 邓汉湘 夏家辉 范朝红

[74]专利代理机构 中国专利代理(香港)有限公司

代理人 姜建成

权利要求书 2 页 说明书 41 页 附图页数 4 页

[54]发明名称 遗传多发性外生骨疣家族的新多核苷酸和多肽基因

[57]摘要

公开了 EXT2 多肽和编码所述酶的 DNA (RNA) 以及用重组技术生产所述多肽的方法。还公开了用所述 EXT2 发展治疗遗传多发性外生骨疣和癌症,如软骨肉瘤等的方法。还公开了检测与核酸序列中的突变和所述多肽浓度改变相关之疾病的诊断方法。

(BJ)第 1456 号

权利要求书

1. 一种含有选自下列组中之成员的分离的多核苷酸：
 - (a) 与编码含有SEQ ID NO: 2氨基酸的多肽的多核苷酸至少有96%的相同性的多核苷酸；
 - (b) 利用遗传密码冗余而编码SEQ ID NO: 2相同氨基酸的多核苷酸；
 - (c) 与(a)或(b)的多核苷酸互补的多核苷酸；和
2. 权利要求1的多核苷酸，其中所述多核苷酸是DNA。
3. 权利要求1的多核苷酸，其中所述多核苷酸是RNA。
4. 权利要求2的多核苷酸含有SEQ ID NO: 1中所列的核苷酸。
5. 权利要求2的多核苷酸，编码含有SEQ ID NO: 2氨基酸的多肽。
6. 含有权利要求2 DNA的载体。
7. 含有权利要求6载体的宿主细胞。
8. 生产多肽的方法，包括从权利要求7的宿主细胞中表达由所述DNA编码的多肽。
9. 生产表达多肽的细胞的方法，包括用权利要求6的载体转化或转染细胞以使所述细胞表达由载体中所含的人cDNA编码的多肽。
10. 含有与SEQ ID NO: 2氨基酸序列至少有96%相同之氨基酸序列的多肽。
11. 含有SEQ ID NO: 2所列氨基酸序列的多肽。
12. 权利要求10多肽的激动剂。
13. 权利要求10多肽的抗体。
14. 权利要求10多肽的拮抗剂。
15. 治疗需要EXT2之患者的方法，包括给所述患者施用治疗有效量的权利要求10的多肽。
16. 权利要求15的方法，其中通过给患者施用编码所述多肽并在体内表达所述多肽的DNA而施用治疗有效量的所述多肽。
17. 治疗需要抑制EXT2多肽之患者的方法，包括给所述患者施用治疗有效量的权利要求16的拮抗剂。
18. 诊断与权利要求10多肽表达相关之疾病或疾病易感性的方法，包括确定编码所述多肽的核酸序列中的突变。

19. 用于疾病诊断的方法，包括分析来自宿主的样品中权利要求10 多肽的水平。

20. 鉴定结合并激活或抑制权利要求10多肽之受体的化合物的方法，包括：
在允许与受体结合的条件下，将在其表面上表达所述多肽之受体的细胞与待筛选的化合物接触，所述受体与能够提供可检测信号以应答化合物与所述受体结合的第二组分有关；和

通过检测是否存在由所述化合物与所述受体相互作用而产生的信号来确定所述化合物是否结合并激活或抑制所述受体。

说明书

遗传多发性外生骨疣家族的新多核苷酸和多肽基因

本发明部分涉及新鉴定的多核苷酸和多肽；所述多核苷酸和多肽的变异体和衍生物；制备上述多核苷酸和多肽其变异体和衍生物的方法；所述多肽的激动剂和拮抗剂；所述多核苷酸，多肽，变异体，衍生物，激动剂和拮抗剂的用途。更具体地从这些和其他方面说，本发明涉及遗传多发性外生骨疣家族的新多核苷酸和多肽，下文称为EXT2。

遗传多发性外生骨疣 (HME或EXT) 是一种常染色体显性病，多发性外生骨疣的特征在于大部分是从长骨的近骨后区产生的。可能涉及的其他骨包括胸和骨盆带，肋骨，而较少发生在椎骨，胸骨，头颅和腕骨和跗骨。另外，还有一个特征是不正常模型，特别是长骨的不正常模型。这会引发有关骨的弯曲，缩短，皮层不规则和干骺端增宽，严重时会导致前臂变形和身材短不成比例。外生骨疣会导致并发症如相邻神经，血管和腱的压迫和刺激以及泌尿和肠梗塞。最严重的并发症是肉瘤变性，这在受影响的个体中有 0.5-2% 的发生率 (Hennekam, R.C.M. Hereditary Multiple Exostoses, J.Med.Genet.28: 262-266, (1991))。Hecht等人 (Hecht, J.T.; Hogue, D.; Strong, L.C.; Hansen, M.F.; Blanton, S.H.; Wagner, M, Hereditary Multiple Exostosis and Chondrosarcoma: Linkage to Chromosome 11 and Loss of Heterozygosity for EXT-Linked Markers on Chromosomes 11 and 8., Am.J.Hum. Genet.56: 1125-1131 (1995)) 报告了多发性外生骨疣中2-5%的软骨肉瘤发生率，与此相比，所有骨癌的随年龄的发生率为1/100,000。多发性外生骨疣是 Langer-Giedion综合症 (150230) 的部分，这种综合症可能是因8q24区的缺失而引起的邻近基因综合症。

遗传连锁研究确定了多发性外生骨疣基因的三个座位。分别将其定位在染色体8, 11和19。

Cook等人 (Cook,A.;Raskind,W.; Blanton,S.H.; Pauli,,R.M.; Gregg,R. G., Francomano, C.A.; Puffenberger,E.; Conrad,E.U. ; Schmale, G. ; Schellenberg,G.; Wijsman,E.; Hecht,J.T.; Wells,D.; Wagner, M. J. , Genetic Heterogeneity in Families with Hereditary Multiple Exostoses., Am.J.Hum.Genet. 53:71-79(1993)) 得出结论约70%的多发性外生骨疣家族与8q24.11-q24.13区中的标记有连锁。研究2个大外生骨疣谱系排除了与8q24标记的连锁, Wu等人 (Wu ,Y.-Q.; Heutink,P.;de Vries,B.B.A. ; Sandkuijl,L.a.;van den Ouweland ,A.M.W.; Niermeijer,M.F; Galjaard.H.; Reyniers ,E.; Willems,P.J; Halley, D.J.J., assignment of a Second Locus for Multiple Esostoses to the Pericentromeric Region of Chromosome 11,Hum.Moled.Genet.3:167-171(1994)) 发现了与11号染色体近短和长臂的微卫星标记的连锁。2点分析的最高优势对数是D11S554, 在 $\theta = 0.03$ 时, 最大优势对数是7.148。

Hecht等人 (1995) 和Raskind 等人 (Raskind ,W.H., Conrad, E. U. ; Chansky, H. , Matsushita , M. Loss of Heterozygosity in Chondrosarcomas for Markers Linded to Hereditary Multiple Exostooses Lodi on Chromosomes 8 and 11,Am J.Hum.Genet.56:1132-1139 (1995)) 提出证据说明在染色体8上的EXT1基因和在染色体11上的EXT2基因有肿瘤抑制功能。他们发现在源于多发性外生骨疣之个体的软骨肉瘤和在散发的软骨肉瘤中, 与这两个基因相连的标记丢失了杂合性。

在美国确定3EXT型频率的大型研究中, Hecht等人 (1995) 发现了一个患有多发性外生骨疣的大的多代家族, 其中一个成员有软骨肉瘤。该家族说明了所述疾病与染色体11标记的连锁。用来自染色体8, 11和19的短串联重复 (STR) 标记比较受影响家族成员的构成型和肿瘤DNA。在肿瘤中观察到染色体8和11标记的杂合性丢失 (LOH) , 但染色体19标记是完整的。Hecht等人 (1995) 在来自所有受影响个体的构成型的DNA和肿瘤样品中, 观察到明显的D11S903缺失。这些结果表明EXT2基因在含D11S903的区内, 两侧为D11S1355和D11S1361。作者相似地分析了来自6个不相关个体的其他构成型和软骨肉瘤DNA, 其中2个有EXT。有EXT之个体的一个肿瘤有染色体8标记的LOH, 发现有散发软骨肉瘤的人有肿瘤特异性LOH和染色体11标记的纯和缺失。这些发现向Hecht等人 (1995)

表明EXT基因可能是肿瘤抑制基因，而且肿瘤发育的引发可能是在多步骤模式之后。Raskind等人(1995)发现在与多发性外生骨疣有关之基因相连的多形座位上，构成型杂合性的丢失。他们检测了在患多发性外生骨疣的人的软骨肉瘤中，与染色体8上的EXT1相连之标记的LOH。他们还发现在17个散发软骨肉瘤中的4个中，有与EXT1相连之标记的LOH，而7个有与EXT2相连之标记的LOH。总之，Raskind等人(1995)在18个肿瘤中的44%上观察到与EXT1或EXT2相连之标记的LOH，而在与EXT3相连的19p上的标记保留了杂合性。这些发现也表明了EXT基因的肿瘤抑制作用。

研究7个扩展的多发性外生骨疣家族，它们都与EXT2座位相连，Wuyts等人(Wuyts,W.;Ramlakhan,S.;Van Hul,W.;Hecht,J.T.;van den Ouweland A.M.W.;Raskind,W.H.;Hofstede,F.C.Reyniers,E.;Wells,D.E.;de Vries,B.;Conrad,E.U.;Hill,A.;Zalatayev,D.;Weissenbach,J.;Wagner,M.J.;Bakker,E.;Halley,D.J.J.;Willems,P.J., Refinement of the Multiple Exostoses Locus(EXT2) to a 3-cM Interval on Chromosome 11, Am. J. Hum. Genet. 57:382-387(1995))将EXT2基因的位置确定在两侧为D11S1355和D11S1361/D11S554的3-cM区。这一发现表明EXT2基因位于11p12-p11。确定的位置排除了许多推测的候选基因，这些基因位于染色体11的着丝粒周围的区域。

McGaughran等人(McGaughran,J.M.;Ward,H.B.;Evans,D.G.R., WAGR Syndrome and Multiple Exostoses in a Patient with Del(11)(p11.2p14.2), J. Med. Genet. 32:823-824(1995))也提供了支持EXT2基因定位的报道，该报道描述了患有多发性外生骨疣和WAGR综合症(Wilms肿瘤，无虹膜，先天畸形和精神发育不全;194070)组合的患者，一种与很多资料相符的因11p13缺失而导致的邻接基因综合症。其患者表现为del(11)(p14.2p11.2)。

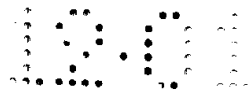
正如由Potocki等人(Potocki,L.;Greenberg,f.;Shaffer,L.G., Interstitial Deletion of 11(p12p11.2):A Rare Chromosomal Syndrome with Mental Retardation, Parietal Foramina, and Multiple exostoses. (abstract), Am. J. Hum. Genet. 57:A123(1995))指出的，对因11p的间隙缺失，del(11)(p12p11.2)而导致的邻接基因综合症的描述，包括以多发性外生骨疣为特征，使得可以确定EXT2的图谱。所述邻接基因综合症的其他特征是精神发育不全和parietal foramina，称为Catlin标记。Potocki和Shaffer(

Potocki, L.; Shaffer, L. G., Interstitial Deletion of 11(p11.2p12) : a Newly Described Contiguous Gene Deletion Syndrome Involving the Gene for Hereditary Multiple Exostoses (EXT2), *Am. J. Med. Genet.* 62: 319-325 (1996)) 报告了在另一有11(p1.2p11.2)缺失的患者中的临床和分子发现。细胞遗传学和分子学分析表明了一个已知与EXT2紧紧相连的标记的新的父本衍生的缺失。所述患者有不正常的面容(两侧内眦愚型, 上睑下垂, 短人中和向下的嘴唇) 精神发育不全, 多发性外生骨疣, 短头畸形和两侧 parietal foramina。

Ahn等人 (Ahn J. et al., Cloning of the Putative Tumor Suppressor Gene for Hereditary Multiple Exostoses, *Nature Genet.*, 11, 137- 143 (1995)) 克隆并表征了以前在2名多发性外生骨疣患者中鉴定的跨染色体断裂点的cDNA。此外, 在2个EXT1家族受影响的成员中, 基因含有移码突变。cDNA含有2238 bp的编码区, 与其他已知的基因序列没有明显的同源性。

Stickens等人 (Stickens D.; Clines, G.; burbee, D.; Ramos, P.; Thonas, s.; Hogue, D.; Hecht, J.T.; Lovett, M.; Evans, G. A., The EXT2 Multiple Exostoses Gene Defines a Family of Putative Tumour suppressor Genes, *Nature Genet.* 14: 25-32 (1996)) 报告了对染色体11上EXT2基因的鉴定和表征。EXT2基因含有7个外显子, 编码718个氨基酸的多肽。通过Northern分析, 他们在几乎所有组织中均检测到一3.5 kb的转录本。另一种剪接在一些组织中产生了一小的3.7 kb转录本。该基因与染色体8上的EXT1基因有显著的序列相似性。由Hecht等人 (1995) 描述的有染色体11连锁的多发性外生骨疣的多代家族中, 其中的一个家族成员发展成了恶性软骨肉瘤。在该家族中, 在所有受影响的成员中均观察到包括多形标记D11S903之区域的显著缺失。在从软骨肉瘤患者衍生的肿瘤组织中, Stickens等人 (1996) 观察到缺失了多形标记D11S903, 而且有其他染色体11标记的LOH。在用于突变的EXT2候选基因的最初研究中, 他们用来自3个不相关家族之受影响成员的DNA完成了SSCP分析。在1个患者的2个等位基因中的1个鉴定了核苷酸784-787的缺失, 导致了EXT2基因产物的移码和成熟前终止。

用通过显微解剖从11p11-12构建的基因组DNA文库, 从人胎盘cDNA文库中分离了数个cDNA克隆。其中一个克隆与EXT1 cDNA在某些部分享有59%的同源



性。通过荧光原位杂交将该基因定位在染色体11p11, 这是EXT2 基因所处的精确的区域。通过cDNA文库筛选和5'RACE技术, 克隆了全长cDNA序列。预测该cDNA编码728个氨基酸。cDNA和氨基酸序列均与EXT1有显著的相似性。因此, 该基因引起染色体11连锁的多发性外生骨疣。

通过将本发明的cDNA序列与报告的EXT2基因序列相比较, 它们在编码区中部有90 bp的不同, 这导致与Stickens (Stickens, 1996) 报告的有30个氨基酸的不同。

Stickens等人从脑cDNA文库中分离了该cDNA, 而本发明的cDNA是从胎盘cDNA文库中分离的。我们的DNA序列与他们的DNA序列之间的差异表明至少有两种EXT2编码的蛋白质异构体。

这表明EXT2基因作为治疗靶是有用的。显然还需要鉴定并表征在预防, 改善或改正机能障碍或疾病, 包括但不限于遗传性多发性外生骨疣和癌如软骨肉瘤等中起作用的其他基因。

在这样或那样的目的中, 本发明的一个特别的目的是提供了多肽, 所述多肽已经通过图1中所列的氨基酸序列与其他蛋白质如EXT1之已知氨基酸序列的同源性而被鉴定为新EXT2。

本发明的另一目的是提供编码EXT2的多核苷酸, 特别是编码由本文 SEQ ID NO: 2所表示之多肽的多核苷酸。

在本发明该方面特别优选的实施方案中, 所述多核苷酸含有编码图1所列序列中之EXT2的区域。

根据本发明的所述方面, 提供了分离的编码EXT2的核酸分子, 包括mRNAs, cDNAs, 基因组DNAs和片段, 在本发明所述方面的其他实施方案中, 提供了生物学, 诊断学, 临床或治疗有用的其变异体, 类似物或衍生物, 包括所述变异体, 类似物和衍生物的片段。

在本发明该方面特别优选的实施方案中是天然产生的EXT2的等位基因变异体。

本发明的另一方面是特别提供EXT2多肽, 可以将其用于治疗目的, 例如治疗遗传多发外生骨疣和癌, 如软骨肉瘤等。

根据本发明的该方面，提供了新的人源多肽，称为EXT2，以及其生物学，诊断学或治疗有用的片段，变异体和衍生物，所述片段的变异体和衍生物以及前述的类似物。

本发明该方面特别优选的实施方案是由天然产生的EXT2基因的等位基因编码的EXT2的变异体。

根据本发明的另一方面，提供了筛选化合物的方法，所述化合物结合并激活本发明多肽或抑制本发明多肽之激活。

本发明的另一目的是提供生产前述多肽，多肽片段，变异体和衍生物，所述变异体和衍生物的片段以及前述类似物的方法。在本发明该方面优选的实施方案中，提供了生产前述EXT2多肽的方法，包括在宿主中表达EXT2的条件下，培养含有编码外源衍生之EXT2的多核苷酸可表达地掺入其中的宿主细胞，在所述宿主细胞中表达EXT2，然后从宿主细胞中回收表达的多肽。

根据本发明的另一目的，提供了产物，组合物，工艺和除此之外，还提供了利用上述多肽和和多核苷酸进行研究，生物学，临床和治疗等目的的方法。

根据本发明该方面某些优选的实施方案，提供了产物，组合物和方法等以用于下列方面：通过确定EXT2多肽或EXT2编码的mRNA而评价细胞EXT2表达以便通过体外，间接体内或体内将细胞暴露于本文公开的EXT2多肽或多核苷酸而治疗遗传多发性外生骨疣和癌如软骨肉瘤等；检测在EXT2基因中的遗传变异和畸变如缺陷；给生物体施用EXT2多肽或多核苷酸以促进EXT2功能或治疗EXT2机能障碍。

根据本发明另一实施方案，提供了用所述激活化合物刺激本发明的多肽以治疗与EXT2的表达不足相关的疾病的方法。

根据本发明另一方面，提供了用所述抑制化合物治疗与EXT2过度表达相关之疾病的方法。

根据本发明的另一方面，提供了非天然产生的合成的，分离的和/或重组EXT2多肽片段，在本发明EXT2至少一个区有保守氨基酸取代的共有序列片段和/或序列，上述多肽能够从质量和数量上调整EXT2与其受体或配体的结合。

根据本发明的另一方面，提供了合成或重组EXT2多肽，其保守取代和衍生物，其抗体，抗个体基因型抗体，可用作EXT2功能调节剂的组合物和方法，由于用其预期的生物学特性通过与所述多肽结合或调节多肽结合而可用在诊断，

治疗和/或研究应用中。

本发明的另一方面提供被设计用来抑制或模拟各种EXT2或其片段的合成的、分离的或重组多肽。

根据本发明这一方面和其他方面特定的优选的实施方案，提供了与EXT2序列杂交的探针。

在本发明该方面其他优选的实施方案中，提供了抗EXT2多肽的抗体。从这点考虑，在特别优选的实施方案中，所述抗体对于EXT2来说是有高度选择性的。

根据本发明的另一方面，提供了EXT2激动剂。这些优选的激动剂是模拟EXT2多肽并与EXT2结合分子或受体结合，引发或促进EXT2诱导的反应的分子。这些优选的激动剂还是与EXT2或EXT2多肽或其他EXT2活性调节剂相互作用，由此促进EXT2的一种作用或EXT2的一种以上之作用的分子。

根据本发明的另一方面，提供了EXT2拮抗剂。优选的拮抗剂是模拟EXT2多肽以便与EXT2受体或结合分子结合但不引发一种EXT2诱导的反应或一种以上EXT2诱导之反应的那些。优选的拮抗剂还是与EXT2结合或相互作用以便抑制EXT2的一种作用或一种以上EXT2作用或防治EXT2表达分子。

在本发明的另一方面，提供了含有EXT2多核苷酸或EXT2多肽的组合物，以便体外，间接体内或体内施用给细胞或施用给多细胞生物体。在本发明该方面特别优选的实施方案中，所述组合物含有EXT2多核苷酸以便在宿主生物体中表达EXT2多肽从而治疗疾病。从这点考虑，特别优选的是在遗传多发性外生骨疣和癌如软骨肉瘤患者中表达以便治疗与畸变的EXT2内生活性相关的机能障碍。

从下列描述中，本发明的其他目的，特征，优点和方面对于本领域专业人员来说将是显而易见的。但是应当理解，尽管说明了本发明优选的实施方案，但下列描述和具体的实施例只是说明性的。从下列描述的阅读和从本发明公开的其他部分的阅读中，在本发明公开的精神和范围内的各种改变和修饰对于本领域专业人员来说是显而易见的。

下列附图描述了本发明特定的实施方案。他们只是说明性的，而不是为了限制本文公开的本发明。

图1表示EXT2的核苷酸序列和由此推导的氨基酸序列 (SEQ ID NO: 1和2)。

术语

提供下列说明性的解释是为了有利于理解本文经常使用，特别是在实施例中使用的特定术语。提供解释是为了方便而不是为了限制本发明。

DNA的“消化”指用酶，例如但不限于仅作用于DNA中特定序列的限制酶催化裂解DNA。本文所提及的各种限制酶是市售的，而且它们的反应条件，辅助因子和使用的其他要求对于专业人员来说是已知的而且是常规的。

一般为了分析目的，用在约20微升反应缓冲液中约2单位的酶消化1微克的质粒或DNA片段。为了分离DNA片段用于质粒构建，一般在成比例的较大的体积中，用20-250单位的酶消化5-50微克的DNA。

在标准实验室手册（如下文提到的）中描述了对于特定限制酶所适宜的缓冲液和底物的量，而且供应商作了说明。

通常使用在37°C约1小时的保温时间，但条件可以随标准方法，供应商的说明和反应的细节而变化。在消化后，用本领域专业人员熟知的方法，可以分析反应，通过琼脂糖或聚丙烯酰胺凝胶电泳纯化片段。

“遗传元件”主要指含有编码多肽的区域或调节复制，转录，翻译或其他过程之区域的多核苷酸，所述其他区域对于在宿主细胞中表达所述多肽是重要的，或者含有编码多肽的区域以及与其可操作相连之可调节表达区的区域的多核苷酸。

遗传元件可以包含在可作为附加体复制的载体内；即作为生理上独立于宿主细胞基因组的分子。可以将它们含在小染色体内，如在真核细胞中通过氨甲蝶呤选择，在转染的DNA扩增过程中产生的那些。也可以在宿主细胞基因组内包含遗传元件，不是以其自然状态，而是在例如分离，克隆操作后，以纯化的DNA形式或在载体中等导入到宿主细胞中。

“分离的”指“通过人手”从其自然状态改变的；即如果它是天然产生的，则已经将其从其来源环境中改变或除去，或两种情况兼而有之。例如在活动物中以其自然状态存在的天然产生的多核苷酸或天然存在的多肽就不是“分离的”，但从其天然状态共存的物质中分离的相同的多核苷酸或多肽就是“分离的”，本文所用的就是这一术语。例如，就多核苷酸而言，术语分离的指将其从染色体或它天然存在的细胞中分离出来。

作为分离的部分或分离后，可将所述多核苷酸与其他多核苷酸如DNA连接

进行诱变以形成融合蛋白质并用于例如在宿主中增殖或表达。可以将分离的多核苷酸（单独的或与其他多核苷酸，如载体相连的）导入到在培养的或在全生物体中的宿主细胞中。导入到培养或全生物体的宿主细胞中后，所述DNA 仍是可以被分离的，（正如本文所用的术语），因为它们不是以其天然生产的形式或在其天然环境中。相似的，多核苷酸和多肽可以在用于导入多核苷酸或多肽例如至细胞中的组合物，如培养基，制剂，溶液中组合物或溶液以用于化学或酶反应，例如可以不是天然产生的组合物，其中保留了在本文所用术语范围内的分离的多核苷酸或多肽。

“连接”指在两个或更多的多核苷酸，最经常的是双链DNA 之间形成磷酸二酯键的过程。用于连接的技术是本领域熟知的，而且在标准实验室手册和文献，如Sambrook等人（MOLECULAR CLONING, A LABORATORY MANUAL, 2nd Ed.; Cold Spring Harbor Laboratory Press, Cold Spring Harbor）（下文称为Sambrook等人）中描述了用于连接的方法。

“寡核苷酸”指相对短的多核苷酸。该术语常常指单链脱氧核糖核苷酸，但它也指单链或双链核糖核苷酸，RNA: DNA杂交体以及双链DNA等。

通常用化学方法合成寡核苷酸，如单链DNA探针寡核苷酸，如用自动寡核苷酸合成仪合成的那些。但是，也可以用各种其他方法，包括体外重组DNA 介导的技术以及在细胞和生物体中表达DNA来制备寡核苷酸。

首先，得到化学合成的没有5'磷酸的DNA。所述寡核苷酸的5'末端不是通过通常使用DNA连接酶形成重组DNA分子之连接反应形成磷酸二酯键的底物。若需要连接所述寡核苷酸，通过标准技术，如使用激酶和ATP 的那些技术可以加入磷酸。

化学合成的寡核苷酸的3'末端通常有一游离的羟基，存在连接酶，如T4DNA连接酶时，就容易与另一多核苷酸，如另一寡核苷酸的5'磷酸形成磷酸二酯键。正如熟知的，如果需要在连接前通过除去其他多核苷酸的5'磷酸可以选择性地抑制该反应。

“质粒”是稳定遗传的而不是其宿主细胞染色体一部分的遗传元件。它们可以由DNA或RNA组成，可以是线性的或环状的。质粒编码确保其在细胞复制过程中复制并稳定遗传的分子，而且可以编码相当重要的医药，农业和环境的产物。例如它们可以编码能够大大提高致病菌的毒力的毒素。它们还可以编码赋

予抗生素抗性的基因。在分子生物学中，广泛地使用质粒作为载体克隆并表达重组基因。根据本领域熟悉的标准命名惯例，本文用一个小写的“p”，在其前面和/或后面跟着大写字母和/或数字来表示质粒。本文公开的起始质粒可以是市售的，公开可得到的，或可以通过熟知的公开方法的常规应用，从可得到的质粒构建而得。根据本发明可以使用的许多质粒和其他克隆和表达载体是熟知的，而且对于本领域专业人员来说是容易得到的。此外，本领域专业人员可以容易地构建任何适用于本发明的其他质粒。在本发明中所述质粒，以及其他载体的特性，构建和用途从本发明的公开来看，对于本领域专业人员是显而易见的。

“多核苷酸”通常指任何多核糖核苷酸或多脱氧核糖核苷酸，它们可以是未被修饰的RNA或DNA或修饰的RNA或DNA。因此，例如本文所用的多核苷酸之单链和双链DNA，是单链和双链区的混合物的DNA，单链和双链RNA，是单链和双链区之混合物的RNA，可以是单链或更通常是双链或单链和双链区的混合物的含有DNA和RNA杂交体分子。另外，本文所用的多核苷酸指含有RNA或DNA或RNA和DNA的三股链区。在所述区域中的链可以来自相同的分子或来自不同分子。所述区域可以包括一个或多个分子的全部，但更通常是只涉及某些分子的一个区域。三股螺旋区的一种分子常常是寡核苷酸。正如本文所用的，术语多核苷酸还包括含有一个或多个修饰的碱基的上述DNA或RNA。因此带有因稳定性或其他原因而修饰之骨架的DNA或RNA是多核苷酸，正是本文该术语所指的。此外，含有不正常碱基，如次黄苷或修饰的碱基，如三苯甲基化的碱基（仅说了两个实例）的DNA或RNA是多核苷酸，正如本文所用的术语。应理解，已经对DNA和RNA进行了各种修饰，适合于本领域专业人员熟知的有用的目的。术语多核苷酸（正如本文所用的）包括多核苷酸的化学，酶促或代谢修饰形式，以及病毒和细胞，包括特别简单和复杂细胞的DNA和RNA特性的化学形式。

本文所用的“多肽”包括下列描述的所有多肽。多肽的基本结构是熟知的，而且已经在本领域的许多教科书和其他文献中作了描述。在本文中，本文所用的该术语指任何肽或蛋白质，它们含有通过肽键彼此线性相连的两个或多个氨基酸。如本文所用的，该术语指短链，这在本领域中通常指肽，寡肽或寡聚物，该术语也指长链，这在本领域中通常指蛋白质，其中都有许多类型。

应该理解，所述多肽常常含有除通常提及的20种天然产生的氨基酸以外的氨基酸，而且在所给的多肽中，用天然方法，如加工和其他翻译后修饰，或化

学修饰技术 (这些都是本领域熟知的) 可以修饰许多氨基酸, 包括末端氨基酸。甚至多肽中天然发生的普通修饰也是很多的, 难以在此穷举, 但在基本教科书和更详细的专著以及大量的研究文献中均有详细的描述, 因此对于本领域专业人员来说是显而易见的。在本发明多肽中存在的已知的修饰包括, 但不限于, 乙酰化, 酰化, ADP-核糖基化, 酰胺化, 黄素的共价连接, 血红素部分的共价连接, 核苷酸或核苷酸衍生物的共价连接, 脂或脂衍生物的共价连接, phosphatidylinositol的共价连接, 交联, 环化, 二硫键形成, 去甲基化, 共价交联的形成, 胱氨酸的形成, 焦谷氨酸盐的形成, 甲酰化, γ -羧基化, 糖基化, GPI锚形成, 羟基化, 碘化, 甲基化, 豆蔻酰化, 氧化, 蛋白水解加工, 磷酸化, 异戊烯基化, 外消旋化, 氢(氧)硒基化, 硫酸化, 转移RNA介导的将氨基酸加到蛋白质中如精氨酸化和遍在蛋白化。这些修饰对于本领域专业人员是熟知的, 而且在许多科学文献中均有详细的描述。在大多数基本教科书, 如PROTEINS - STRUCTURE AND MOLECULAR PROPERTIES, 2nd Ed., T. E. Creighton, W.H. Freeman and Company, New York, 1993。中描述了几种最常用的修饰, 包括糖基化, 脂附着, 硫酸化, 谷氨酸残基的 γ -羧基化, 羟基化和ADP-核糖基化。还可以得到有关该方面的详细综述。参见例如Wold, F., Posttranslational Protein Modifications: Perspectives and Prospects, pgs.1-12 in POSTTRANSLATIONAL COVALENT MODIFICATION OF PROTEINS, B.C. Johnson, E., Academic Press, New York, 1983; Seifter et al., "Analysis for protein modifications and nonprotein cofactors", Meth.Enzymol., 1990, 182:626-646 and Rattan et al., "Protein Synthesis: Posttranslational Modifications and Aging", Ann.N.Y.Acad. Sci., 1992, 663:48-62。

正如熟知的和上文说明的, 应该理解, 所述多肽并不总是完整的线性。例如, 多肽可以因遍在蛋白化而是分支的, 它们可以是环状的, 有或没有分支, 通常是翻译后加工过程的结果, 包括天然加工和因非自然的人工操作而生产的过程。用非翻译天然方法和完全的合成方法可以合成环状的, 分支的和分支的环状多肽。

修饰可以发生在多肽的任何位置, 包括肽骨架, 氨基酸侧链和氨基或羧基末端。事实上, 通过共价修饰来阻断多肽中的氨基或羧基通常是天然过程, 而

合成的多肽和所述修饰也可以存在于本发明的多肽中。例如在大肠杆菌中产生的多肽的氨基末端残基，在加工前几乎总是N-乙酰甲硫氨酸。

在多肽中发生的修饰经常是它如何制备的一种功能。为了通过在宿主中表达克隆的基因而制备多肽，大部分修饰的性质和程度取决于宿主细胞翻译后修饰的能力和在多肽氨基酸序列中存在的修饰信号。例如，象已知的，在细菌宿主如大肠杆菌中一般不发生糖基化。因此，当需要糖基化时，应在糖基化的宿主，通常是真核细胞中表达多肽。昆虫细胞经常象哺乳动物细胞一样完成翻译后糖基化，由于该原因，已经发展了昆虫细胞表达系统来有效表达哺乳动物蛋白质，所述哺乳动物蛋白质特别地具有天然的糖基化方式。可以将相似的考虑用于其他修饰。

应该理解，可以在给定多肽中的数个位点，以相同或不同的程度存在相同类型的修饰。另外，所给的多肽还可以含有许多类型的修饰。

总之，如本文所用的，术语多肽包括了所有所述修饰，特别是在通过在宿主细胞中表达多核苷酸而合成的多肽中存在的那些。

本文所用的术语多核苷酸或多肽的“变异体”分别是不同于参考多核苷酸或多肽的多核苷酸或多肽。下面和在本发明公开的别处更详细地描述了上述意义上的变异体。

变异体包括核苷酸序列不同于另一种参考多核苷酸的多核苷酸。通常，要限制这种差异以便使参考和变异体的核苷酸序列总的来说很相似，而且在许多区域中相同。

正如下面说明的，在变异体核苷酸序列中的改变是沉默的。即，它们不会改变由所述多核苷酸编码的氨基酸。在将改变限制于这种类型的沉默改变时，变异体将编码与参考有相同氨基酸序列的多肽。下面还说明，在变异体核苷酸序列中的改变将改变由参考多核苷酸编码之多肽的氨基酸序列。如下文所讨论的，所述核苷酸改变会导致由参考序列编码之多肽的氨基酸取代，增加，缺失，融合和截短。

变异体还包括氨基酸序列不同于另一种参考多肽的多肽。通常要限制这种差异以便使参考和变异体的序列总的来说很相似，而且在许多区域中相同。

变异体和参考多肽在氨基酸序列中可以因一个或多个取代，增加，缺失，融合和截短（可以存在任何组合）而有所不同。

本文所用的术语“融合蛋白质”是由两种，经常是不相关的融合的基因或其片段编码的蛋白质。EP-A-0 464533 (加拿大同族2045869) 公开了融合蛋白，它含有免疫球蛋白分子恒定区的不同部分和另一人蛋白质或其部分。在许多情况下，用免疫球蛋白Fc区作为融合蛋白的部分对于用于治疗和诊断是有利的，这样会导致例如改善药物动力学特性 (EP-A 0232 262)。另一方面，为了某些用途，在融合蛋白被表达，检测和纯化后，需要能够检测Fc部分。因此，需要将融合蛋白组分与化学或酶促裂解的连接区相连。这就是证明Fc部分妨碍治疗和诊断应用时的那种情况，例如当将融合蛋白用作免疫抗原时。在药物筛选中，例如已经将人蛋白质，如shIL-5 α 与Fc部分融合以用于高处理量筛选试验以鉴定hIL-5的拮抗剂。参见D. Bennett et al., *Journal of Molecular Recognition*, 1995, 8:52-58; 和K. Johanson et al., *The Journal of Biological Chemistry*, 1995, 270(16):9459-9471。

因此，本发明还涉及遗传工程加工的可溶的融合蛋白，由EXT2或其部分和各种亚类免疫球蛋白 (Ig G, Ig M, IgA, IgE) 重链或轻链恒定区的不同部分组成。优选的免疫球蛋白是人Ig G，特别是Ig G1重链的恒定部分，融合发生在铰链区。在一个实施方案中，通过插入一个可以用血凝集因子Xa裂解的裂解序列就可以简单地除去Fc部分。本发明还涉及通过基因工程方法制备这些融合蛋白的方法，以及其用于诊断和治疗的用途。本发明的另一方面还涉及编码所述融合蛋白的多核苷酸。

膜结合蛋白在形成融合蛋白时是特别有用的。所述蛋白质的特征主要在于拥有三个不同的结构区，细胞外区，跨膜区和胞质区。本发明期望将一个或多个这些区域作为融合蛋白的组分。在WO94/29458和WO94/22914中可以发现所述融合蛋白技术的实例。

“结合分子” (或称为“相互作用分子”或“受体组分因子”) 指与本发明多肽特异性结合或与本发明多肽相互作用的分子，包括受体。这种结合分子是本发明的一部分，结合分子还可以是非天然产生的如与本发明多肽特异性结合的抗体，和抗体衍生的试剂。

正如本领域已知的，两种多肽之间的“相似性”是通过将一种多肽的氨基酸序列和其保守氨基酸取代与第二种多肽的序列进行比较而确定的。此外，本领域还已知的“相同性”指两种多肽或两种多核苷酸序列之间的序列相关程度。

这是通过所述序列两个链之间的匹配相同性而确定的。相同性和相似性都很容易计算 (COMPUTATIONAL MOLECULAR BIOLOGY , Lesk , A.M. ed ., Oxford University Press ,New York ,1988; BIOCOMPUTING : INFORMATICS AND GENOME PROJECTS ,Smith ,D.W. ed.,academic Press , New York , 1993; COMPUTER ANALYSES OF SEQUENCE DATA ,PART I , Griffin , A. M. and Griffin,H.G. ,eds .,Humana Press,New Jersey ,1994;SEQUENCE ANALYSES IN MOLECULAR BIOLOGY ,von Heinje ,G. , Academic Press , 1987; and SEQUENCE ANALYSES PRIMER ,Gribskov ,M. and Devereux ,J., eds . , M Stockton Press ,New York ,1991) 。存在许多测量两种多核苷酸或多肽序列之间相同性和相似性的方法, 术语“相同性”和“相似性”是本领域专业人员熟知的 (Carillo ,H.,and Lipton, D.,SIAM J. Applied math. , 1988, 48: 1073) 。通常用于确定两个序列之间相同性或相似性的方法包括, 但不限于在 Guide to Huge Computers ,Martin J.Bishop ,ed.,Academic Press , San Diego,1994,and Carillo,H.,and Lipton,D.,SIAM J.Applied Math.,1988,48: 1073中公开的那些。设计确定相同性的优选方法是给出待测两个序列之间的最大匹配。用计算机程序也编制了确定相同性和相似性的方法。确定两个序列间相同性和相似性的优选的计算机程序方法包括, 但不限于 GCG 程序包 (Devereux ,J.,et al.,Nucleic acids Research ,1984,12(1) : 387) , BLAST , FASTA (Atschul,S.F.et al.,J.Molec .Biol.,1990,215:403) 。

本发明涉及下文将详细描述的新EXT2多肽和多核苷酸等。特别是, 本发明涉及新EXT2的多肽和多核苷酸, 它与EXT1有氨基酸序列同源性。本发明特别涉及具有图1所示之核苷酸和氨基酸序列的EXT2。

多核苷酸

根据本发明的一个方面, 提供了分离的多核苷酸, 它编码具有图1 推导的氨基酸序列的EXT2多肽。

本发明的多核苷酸可以是RNA形式, 如mRNA, 或通过克隆得到的或通过化学合成技术生产的或其组合方法得到的DNA形式, 包括例如, cDNA和基因组DNA。

所述DNA可以是双链或单链的。单链DNA可以是编码链，也称为有意义链，或者，它也可以是非编码链，也称为反义链。

编码多肽的编码序列可以与图1所示的多核苷酸的编码序列，SEQ ID NO: 1相同。它可以是有不同序列的多核苷酸，由于遗传密码的过多（简并性），它还可以编码图1SEQ ID NO: 2的多肽。

编码图1多肽的本发明多核苷酸包括，但不限于其自身成熟多肽的编码序列；成熟多肽的编码序列和其他的编码序列；成熟多肽的编码序列，带有或不带有前述的其他编码序列并含有其他的非编码序列。其他编码序列的实例包括，但不限于编码前导或分泌序列的序列，如前-或原或前原蛋白质序列。其他非编码序列的实例包括，但不限于内含子和非编码5'和3'序列，如在转录，和mRNA加工中起作用的转录非翻译的序列，包括例如用于mRNA与核糖体结合和稳定性的剪接和多聚腺苷酸化信号。在多肽中也可以掺入能提供其他功能性的编码序列。因此，例如，可以将所述多肽与标记序列如有利于融合多肽纯化的肽融合。在本发明该方面特定优选的实施方案中，标记序列是六聚组氨酸肽，如在pQE载体 (Qiagen, Inc.) 中提供的。如Gentz等人 (Proc. Natl. Acad. Sci., USA, 1989, 86: 821-824) 中所描述的，例如，六聚组氨酸有利于融合蛋白的纯化。在另一实施方案中，标记序列是HA标记。HA标记对应于流感病毒血细胞凝集素衍生的表位，例如，已经由Wilson等人 (Cell, 1984, 37: 767) 作了描述。许多其他的所述标记可以从商业途径获得。

根据前述，本文所用的术语“编码多肽的多核苷酸”包括多核苷酸，根据遗传密码过多，包括编码本发明多肽的任何序列，特别是含有图1, SEQ ID NO: 2所列氨基酸序列的EXT2。该术语还包括多核苷酸，所述多核苷酸包括编码多肽（例如，由内含子间断的）的一个连续区或不连续区以及其他区域，也可含有编码和/或非编码序列。

本发明还涉及本文上述多核苷酸的变异体，它编码具有图1推导的氨基酸序列之多肽的片段，类似物和衍生物。所述多核苷酸的变异体可以是天然产生的变异体例如，天然产生的等位基因变异体，或者它可以不是天然产生的变异体。所述多核苷酸的非天然产生的变异体可以通过诱变技术包括用于多核苷酸，细胞或生物体的那些来制备。

从这点看，变异体是通过核苷酸取代，缺失或增加而不同于上述多核苷酸

的变异体。取代，缺失或增加可涉及一个或多个核苷酸。可以在编码或非编码区或这两个区中改变所述变异体。在编码区中的改变可以产生保守的或非保守氨基酸取代，缺失或增加。

从这点看，本发明特别优选的实施方案是编码具有图1所列EXT2 氨基酸序列之多肽的多核苷酸；其变异体，类似物，衍生物和片段，以及所述变异体，类似物和衍生物的片段。

由此看来，特别优选的是编码EXT2变异体，类似物，衍生物和片段，以及所述片段之变异体，类似物和衍生物的多核苷酸，它们含有图1中EXT2 多肽的氨基酸序列，其中有数个，几个5-10，1-5，1-3，2，1 或没有氨基酸残基以任何组合取代，缺失或增加。在这些中特别优选的是沉默取代，增加和缺失，它们不改变EXT2的特性和活性。从这点看，特别优选的是保守取代。最优选的是编码没有取代的具有图1氨基酸序列之多肽的多核苷酸。

本发明的其他优选的实施方案是与编码具有图1氨基酸序列之EXT2 多肽的多核苷酸至少有96%的相同的多核苷酸，和与所述多核苷酸互补的多核苷酸。

此外，在该方面特别优选的实施方案是编码多肽的多核苷酸，所述多肽基本上保持与图1的cDNA所编码之成熟多肽相同的生物学功能或活性。

本发明还涉及与本文上述序列杂交的多核苷酸。在该方面，本发明特别涉及在严格条件下，与本文上述多核苷酸杂交的多核苷酸。如本文所用的，术语“严格条件”指只有在两个序列之间至少96%的相同性，才会发生杂交。

对于本文另外讨论的本发明的多核苷酸检测，例如可以将上文讨论的本发明多核苷酸用作cDNA和基因组DNA的杂交探针以分离编码EXT2的全长cDNA 和基因组克隆，并分离与EXT2基因有高度序列相似性的其他基因的cDNA和基因组克隆。所述探针通常含有至少15个核苷酸。优选的，所述探针至少有30个核苷酸，而且可以至少有50个核苷酸。特别优选的30-50个核苷酸之间。

例如，用已知DNA序列合成寡核苷酸探针，通过筛选可以分离编码EXT2 基因的编码区。然后用与本发明基因序列互补的标记的寡核苷酸筛选cDNA，基因组DNA或mRNA文库以确定与探针杂交的文库成员。

可以用本发明的多核苷酸和多肽作为研究试剂和物质以发现治疗和诊断人疾病的方法，如本文与多核苷酸检测有关的部分将进一步讨论。

多核苷酸可以编码为成熟蛋白质加其余的氨基或羧基末端氨基酸，或成熟

多肽内氨基酸（例如当成熟形式有一个以上多肽链时）的多肽。所述序列可以在从前体到成熟形式的蛋白质加工中起作用，可以有利于蛋白质的开放通行，可以延长或缩短蛋白质的半衰期或有利于蛋白质检测或生产的操作等。如通常在原位的情况下，通过细胞酶可以处理其他氨基酸以使之离开成熟蛋白质。

具有与一个或多个原序列融合之多肽的成熟形式的前体蛋白质可以是失活形式的多肽。当除去原序列时，所述失活的前体通常就被激活。在激活前可以除去一些或所有原序列。通常将所述前体称为原蛋白质。

总之，本发明的多核苷酸可以编码成熟蛋白质，成熟蛋白质加前导序列（可将其称为前蛋白质），具有一个或多个不是前蛋白质前导序列之原序列的成熟蛋白质的前体，或前原蛋白质（是原蛋白质的前体），它含有前导序列和一个或多个原序列，通常在产生活性和成熟形式多肽的加工步骤中将其除去。

多肽

本发明还涉及具有图1，SEQ ID NO: 2推导的氨基酸序列的EXT2多肽。

本发明还涉及这些多肽的片段，类似物和衍生物。当称图1多肽时，术语“片段”，“衍生物”和“类似物”，指基本上保留了与所述多肽相同的生物学功能或活性，即作为EXT2行使功能的多肽，或甚至所述多肽不发挥EXT2多肽之功能时，保留与任何受体或结合分子结合的能力。因此，例如类似物包括通过裂解原蛋白质部分从而被激活以产生有活性的成熟多肽的原蛋白质。

本发明的多肽可以是重组多肽，天然多肽或合成多肽。在特定优选的实施方案中，它是重组多肽。

图1多肽的片段，衍生物或类似物可以是：(i) 其中一个或多个氨基酸残基由保守的或非保守氨基酸残基（优选保守氨基酸残基）取代的一种，而且所述取代的氨基酸残基可以是或不是由遗传密码所编码的；(ii) 其中一个或多个氨基酸残基包括取代基的一种；(iii) 其中成熟多肽与另一种化合物，如提高成熟多肽半衰期的化合物（如聚乙二醇）融合的一种或(iv) 其中其他氨基酸与成熟多肽融合的一种，如用于纯化成熟多肽或原蛋白质序列的前导或分泌序列。从本文的教导中，认为所述片段，衍生物和类似物在本领域熟知的范围内。

从这点看，本发明特别优选的实施方案是具有图1 SEQ ID NO: 2所列EXT2

氨基酸序列的多肽，其变异体，类似物，衍生物和片段，以及所述片段的变异体，类似物和衍生物。就这点而言，本发明特别优选的实施方案还是具有EXT2氨基酸序列的多肽，其变异体，类似物，衍生物和片段，以及保留了该多肽活性/功能的所述片段的变异体，类似物和衍生物。

优选的变异体是通过保守氨基酸取代而不同于参考的那些。所述取代是用另一类似特征的氨基酸取代多肽中一给定氨基酸的那些。在保守取代中通常所见到的是脂肪族氨基酸Ala, Val, Leu和Ile之间的相互取代；羟基残基Ser和Thr的交换，酸性残基Asp和Glu的交换，酰胺残基Asn和Gln之间的取代，碱性残基Lys和Arg的交换，以及芳族残基Phe和Tyr之间的取代。

就这点而言，特别优选的是具有图1EXT2氨基酸序列的变异体，类似物，衍生物和片段，以及所述片段的变异体，类似物和衍生物，其中有数个，几个，5-10，1-5，1-3，2，1或没有氨基酸取代，缺失或增加，或其中的任何组合。其中特别优选的是不会改变所述多肽特性和活性的沉默取代，增加和缺失。

就这点而言，特别优选的是保守取代。最优选的是具有图1，SEQ ID NO: 2之氨基酸序列，没有取代的多肽。

本发明的多肽和多核苷酸优选的以分离的形式提供，而且优选纯化至均一。

本发明的多肽包括SEQ ID NO: 2的多肽（特别是成熟多肽）以及与SEQ ID NO: 2的多肽至少有96%相似性（更优选至少96%相同性）的多肽。

也可以用本发明多肽的片段或部分通过肽合成方法生产对应的全长多肽；因此可以用片段作为中间产物生产全长多肽。可以用本发明多核苷酸的片段或部分合成本发明的全长多核苷酸。片段可以是“游离状态的”，即不是其他氨基酸或多肽的部分，或与其他氨基酸或多肽融合，或它们被包含在较大多肽中，形成一部分或区。当在较大多肽内时，目前所讨论的片段最优选形成一个单独连续区域。但是可以数个片段包含在一个较大的多肽中。例如，特定优选的实施方案涉及本发明EXT2多肽的片段，包含在用于在宿主中表达的前期多肽中，而且含有与EXT2氨基末端融合的异源前-和原-多肽区以及与所述片段羧基末端融合的其他区域。因此，在本文欲指意义的一方面，片段指的是衍生自EXT2的融合多肽或融合蛋白部分。

应理解本发明另外还涉及编码上述片段的核苷酸，与编码所述片段之核苷酸杂交的核苷酸，尤其是那些在严紧条件下杂交的核苷酸，以及用于

扩增编码所述片段之多核苷酸的多核苷酸，例如，P C R 引物。在这点上，优选的多核苷酸是对应于上文讨论的优选片段的那些多核苷酸。

载体，宿主细胞，表达

本发明还涉及含有本发明多核苷酸的载体，用本发明载体进行基因工程处理的宿主细胞以及通过重组技术生产本发明多肽。

通过基因工程操作，使宿主细胞中掺入多核苷酸并表达本发明多肽。例如，使用已知的感染，转导，转染，转载 (transvection) 和转化技术可将多核苷酸导入宿主细胞。可单独导入多核苷酸，或与其他多核苷酸一起导入。这种其他的多核苷酸可单独导入，其导入或与本发明的多核苷酸联合导入。

因此，例如使用标准的共转染和在如哺乳动物细胞中选择的技术，可将本发明的多核苷酸与另一个编码可选择标记的分离多核苷酸一起转染到宿主细胞中。在这种情况下，多核苷酸一般会稳定地掺入宿主细胞基因组。

另外，多核苷酸也可结合到含有可选择标记的载体上以在宿主中增殖。可通过上述技术将载体构建体导入宿主细胞。通常，质粒载体以沉淀物（如磷酸钙沉淀物）中的D N A 形式或以与带电脂质之复合体中的D N A 形式被导入。也可使用电穿孔法将从核苷酸导入宿主。如果载体是病毒，可在体外包装或导入包装细胞，经包装的病毒可被导入细胞。根据本发明的这一方面，适于制备多核苷酸并将多核苷酸导入细胞的多种技术对于本领域技术人员而言是熟知的和常规的。在Sambrook等人的著作中可详细见到这种技术，该著作仅仅是详细描述这些技术的许多实验操作手册的例证而已。

根据本发明的这一方面，载体可以是例如质粒载体，单或双链噬菌体载体，或单链或双链R N A 或D N A 病毒载体。可通过熟知的使D N A 和R N A 导入细胞的技术将这种载体作为多核苷酸，优选作为D N A 导入细胞，对于噬菌体和病毒载体而言，可以并优选通过熟知的感染和转导技术使载体作为经包装或衣壳化的病毒导入细胞。病毒载体可以是有复制活性的或复制缺陷的。在后一种情况下，病毒增殖一般仅在互补的宿主细胞中发生。

在某些方面，载体中优选的是那些可用于表达本发明多核苷酸和多肽的载体。通常，这种载体含有顺式作用的调控区，它可有效用于在宿主中的表达，该调控区与要表达的多核苷酸可操作地相连接。通过导入宿主，适当的反式作用因子可以由宿主提供，由互补载体提供或由载体自身提供。

在某些这方面的优选实施方案中，载体提供了特异性表达。这种特异性表达可以是诱导的表达或仅在某种类型的细胞中的表达或者既是诱导的也是细胞特异性的表达。在诱导的载体中，特别优选的是通过易于操纵的环境因素如温度和营养添加剂的诱导可表达蛋白质的载体。本领域技术人员熟知并常规使用多种适于本发明这一方面的载体，包括用于原核和直核宿主的组成型和可诱导型表达载体。

可在常规的营养培养基中培养经基因工程处理的宿主细胞，所述培养基或可经过修饰以特别适于激活启动子，选择转化体或扩增基因。目前经选择用于表达的宿主细胞所用的培养条件如温度，pH等等一般适于表达本发明的多肽，这一点对于本领域技术人员而言是显而易见的。

可使用多种表达载体表达本发明多肽。这种载体包括衍生自染色体，附加体和病毒的载体，例如，衍生自细菌质粒、噬菌体、酵母附加体、酵母染色体元件，如杆状病毒，乳多空病毒、SV40，痘苗病毒、腺病毒、禽痘病毒、猪 α 疱疹病毒1型和逆转录病毒之病毒的载体，及衍生自它们之组合的载体，如衍生自质粒和噬菌体遗传元件粘粒和噬菌粒的载体。通常，在这一方面，任何适于在宿主中维持、增殖或表达多核苷酸以产生多肽的载体都可用于表达。

可通过多种已知的常规技术中的任一种将适当的DNA序列插入载体。一般通过用一种或多种限制性内切核酸酶裂解DNA序列和表达载体，再使用T4 DNA连接酶将限制性片段连接在一起即可使用于表达的DNA序列与表达载体相连接。用于此目的的限制和连接方法对本领域技术人员而言是已知的和常规的。Sambrook等人的著作中详细描述了在此方面的适当方法，以及使用其他技术构建表达载体的适当方法，这些方法对本领域技术人员而言是熟知的和常规的。

表达载体中的DNA序列与包括例如指导mDNA转录之启动子的适当表达调控序列可操作地连接。这种启动子的代表包括噬菌体入PL启动子，大肠杆菌Lac，trp和tac启动子，SV40早期和晚期启动子和逆转录病毒LTR的启动子，它们仅是列举出的几个熟知的启动子。应理解对本发明此方面有用的大量其他启动子是已知的，而且本领域技术人员可以本文讨论和实施例部分中阐明的方式常规地使用它们。

通常，表达构建体可含有转录起始和终止位点，并在转录区含有供翻译用

的核糖体结合位点。由构建体表达的成熟转录本的编码部分可包括位于开始部分的翻译起始AUG和适当定位于待翻译多肽末端的终止密码子。

另外，构建体也可含有调节以及产生表达的调控区。通常，根据许多一般操作的方法，这一区域通过控制转录来起作用，例子包括阻遏物结合位点和增强子及其他。

用于增殖和表达的载体一般包括选择性标记。选择性标记基因，提供了表型性状以选择转化的宿主细胞，优选的标记包括但不限于用于真核细胞培养的二氢叶酸还原酶或新霉素抗性基因以及用于培养大肠杆菌和其他细菌的四环素或氨基青霉素抗性基因。这种标记也适于扩增。另外，载体也可含有用于此目的的其他标记。

使用适于表达本文所需多肽的多种已知技术可将载体导入适当的宿主，所述载体含有本文所述的选定多核苷酸序列以及适当的启动子和其他适当的调控序列。适当宿主的代表性例子包括细菌细胞，如大肠菌，链霉菌和鼠伤寒沙门氏菌（*Salmonella typhimurium*）细胞；真菌细胞，如酵母细胞；昆虫细胞如果蝇S2和草地夜蛾sf9细胞；动物细胞如CHO，COS和Bowes黑素瘤细胞；和植物细胞。多种表达构建体的宿主是已知的，本发明的公开可使本领域技术人员能够常规选择用于表达根据本发明此方面多肽的宿主。

更具体地说，本发明还包括如表达构建体的重组构建体，它含有一个或多个上述序列。此构建体含有载体，如质粒或病毒载体，其中已插入了本发明的序列。此序列可以正方向或反方向插入。在此方面某些优选的实施方案中，构建体进一步包含调控序列，包括例如与此序列可操作连接的启动子。本领域技术人员已知大量适当的载体和启动子，有很多市售载体适用于本发明。

例如可提供下列市售载体，优选用于细菌的载体是得自Qiagen的PQE70，PQE60和PQE-9；得自Stratagene的PBS载体，Phagescript载体，Bluescript载体，pNH8A，pNH16a，pNH18A，pNH46A；以及得自Pharmacia的ptr099a，pKK223-3，pKK233-3，pDR540，pRIT5。优选的真核载体是得自Stratagene的PWLNEO，pSV2CAT，pOG44，pXT1和PSG；和得自Pharmacia的pSVK3，pBPV，pMSGT和pSVL。这些载体仅通过阐明多种商业可用已知载体的方式被列出，本领域技术人员可将这些载体用于本发明。应理解本发明在此方面可使用适于例如在宿主中导入、维持、增殖或表达本发明多核苷酸或多肽的任

何其他质粒或载体。

可使用含有缺乏启动子区域之报道转录单元的载体从任何所需基因中选择启动子区域，所述单元例如位于导入候选启动子片段，即含有启动子三片段的限制性位点下游的氯霉素乙酰转移酶（“CAT”）转录单元。已知在CAT基因上游的限制性位点处将含有启动子的片段导入载体可导致CAT活性的产生，此活性可通过标准的CAT检测法检测。适于此目的的载体是已知的并易于使用。这种载体的两个例子包括pKK232-8和pCM7。因此，用于表达本发明多核苷酸的启动子不仅包括已知的和易于使用的启动子，也包括通过使用报道基因的上述技术易于得到的启动子。

适于表达本发明多核苷酸和多肽的已知细菌启动子是E.Coli *Lac I* 和 *Lac Z* 启动子，T3和T7启动子，*gpt*启动子， λ PR，PL启动子和启动子。

适于此方面的已知真核启动子是CMV立即早期启动子，HSV胸苷激酶启动子，早期和晚期SV40启动子，如劳氏肉瘤病毒("RSV")的逆转录病毒的LTR启动子以及如小鼠金属硫蛋白-I启动子的金属硫蛋白启动子。

选择适当载体和启动子以在宿主细胞中表达是已知的方法，用于构建表达载体，将载体导入宿主并在宿主中表达的必需技术对本领域技术人员而言是常规技术。

本发明还涉及含有上述构建体的宿主细胞。宿主细胞可以是高等真核细胞，如哺乳动物细胞，低等真核细胞，如酵母细胞或原核细胞如细菌细胞。

构建体导入宿主细胞受磷酸钙转染，DEAE-葡聚糖介导的转染，阳离子脂质介导的转染，电穿孔、转导、感染或其他方法的影响。许多标准实验操作手册中都描述了这类方法。

可以常规方式使用宿主细胞中的构建体从而产生由重组序列编码的基因产物。另外，也可通过常规的肽合成仪合成产生本发明多肽。

可在适当启动子的控制之下，在哺乳动物细胞、酵母菌、细菌或其也细胞中表达成熟的蛋白质。使用无细胞翻译系统也可产生这种蛋白质，该系统使用了衍生自本发明DNA构建体的RNA。Sambrook等人描述了原核和真核宿主使用的适当克隆和表达载体。

一般而言，重组表达载体可包括复制起点，衍生自指导下游结构序列转录之高表达基因的启动子和暴露于载体之后允许分离含有载体之细胞的选择性

标记。适当的启动子衍生自编码如3-磷酸甘油酸激酶 (“PGK”) 的糖酵解酶, α -因子, 酸性磷酸酶和热休克蛋白等的基因。选择性标记包括E. coli的氨基青霉素抗性基因和啤酒糖酵母 (S. cerevisiae) 的trp1 基因。

通过在载体中插入增强了序列可增强高等真核生物对编码本发明多肽之DNA的转录。增强子是DNA的顺式作用元件, 通常约为10至300 bp, 它在给定的宿主细胞类型中起到增加启动子转录活性的作用。增强子的例子包括SV40增强子, 它位于bp100至270处之复制起点的晚期侧, 还包括细胞肥大病毒早期启动子增强子, 位于复制起点晚期侧的多瘤增强子和腺病毒增强子。

通常, 可使用标准技术将编码本发明多肽之异源结构序列的本发明多核苷酸插入载体中以使它与启动子可操作地连接以供表达。将多核苷酸放在适当的位置以使转录起始位点大致位于核糖体结合位点的5'端, 核糖体结合位点位于AUG的5'方向。所述AUG可启动欲表达之多肽的翻译。通常, 没有以起始密码子 (常为AUG) 开始, 并位于核糖体结合位点和起始密码子之间的其他开放阅读框。而且多肽末端常具有翻译终止密码子, 在转录区的3'末端大致分布有多聚腺苷酸化信号和转录终止信号。

可在经表达的多肽中掺入适当的分泌信号以将经翻译的蛋白质分泌到内质网腔, 壁膜间隙或细胞外环境。此信号可与多肽内源或异源。

此多肽可以修饰的形式 (如融合蛋白) 被表达, 它不仅包括分泌信号还包括另外的异源功能性区域。因此, 例如可在多肽的N-末端加上附加的氨基酸区域, 尤其是带电荷的氨基酸以在纯化或随后的处理和贮存过程中改善它在宿主细胞中的稳定性和持续性。也可以在多肽中增加一个区域以便于纯化。在最终制备多肽之前可除去这一区域。在多肽中加入肽部分以引起分泌或外泌, 改善稳定性和便于纯化等等都是本领域中熟知和常规技术。

用于增殖, 维持或表达本发明多核苷酸和多肽的适当原核宿主包括大肠杆菌, 枯草芽孢杆菌和鼠伤寒沙门氏菌。假单胞菌, 链霉菌, 和葡萄球菌的许多种也是适当的宿主。另外, 本领域技术人员已知的许多其他宿主也可用于本发明。

作为一个代表性而非限制性的例子, 供细菌使用的有用表达载体可含有衍生自市售质粒的选择性标记和细菌复制起点, 所述市售质粒含有已知克隆载体

pBR322 (ATCC 37017) 的基因元件。 这各商购载体包括例如 pKK223-3 (Pharmacia Fine Chemicals, Uppsala, Sweden) 和 GEM1 (Pramesa Biotech, Madison, WI, USA)。 在这些载体中, pBR322 “骨架” 部分与适当的启动子和欲表达的结构序列相连接。

转化适当的宿主菌株之后, 使宿主菌株生长至适当的细胞密度。 当选定的启动子为诱导型时, 可通过适当方法 (例温度变化或暴露于化学诱导剂) 诱导启动子并将细胞再培养一段时间。 典型地再通过离心收集细胞, 通过物理或化学方法破碎细胞, 保留最终得到的粗提取作进一步纯化。

可通过任何便利的方法破碎表达蛋白质所用的微生物细胞, 包括反复冻融, 超声处理, 机械破碎或使用细胞裂解剂。 本领域技术人员已知这些方法。

多种哺乳动物细胞培养物系统也可用于表达, 哺乳动物表达系统的例子包括 C127, 3T3, CHO, HeLa, 人肾293 和 BHK 细胞系以及猴肾成纤维细胞的 COS-7 细胞系, 它们描述于 Gluzman et al., Cell, 1981, 23: 175。

哺乳动物表达载体可含有复制起点, 适当的启动子和增强子, 和任何必需的核糖体结合位点, 多聚腺苷酸化位点, 剪接供体和受体位点, 转录终止序列以及表达所必需的5' 侧翼的非转录序列。 在某些优选的实施方案中, 将衍生自 SV40 剪接位点和 SV40 多聚腺苷酸化位点的 DNA 序列用于必需的非转录基因元件。

可通过已知方法从重组的细胞培养物中回收并纯化 EXT2 多肽, 所述方法包括硫酸铵或乙醇沉淀, 酸提取, 阴离子或阳离子交换层析, 磷酸纤维素层析, 疏水作用层析, 亲和层析, 羟基磷灰石层析和凝集素层析。 最优选使用高效液相层析 (“HPLC”) 纯化。 当多肽在分离和或纯化过程中变性可使用已知的再折叠蛋白质技术以重新产生活性构型。

本发明多肽包括天然纯化的多肽, 通过化学合成方法产生的多肽以及通过重组技术由包括例如细菌、酵母菌, 高等植物, 昆虫和哺乳动物细胞的原核或真核宿主产生的多肽。 依据重组生产方法中所用的宿主, 本发明多肽可以是糖基化的或非糖基化的。 另外, 在某些情况下作为宿主介导过程的结果。 本发明多肽也包括起始的经修饰甲硫氨酸残基。

根据本发明, EXT2 多核苷酸和多肽可用于多种应用, 尤其是那些利用

此酶化学和生物特性的应用。其他的应用涉及细胞、组织和生物体疾病的诊断和治疗。下列讨论中将进一步阐述本发明的这些方面。

多核苷酸测定

本发明还涉及使用E X T 2 多核苷酸以检测互补的多核苷酸，从而用作例如诊断试剂。与机能障碍相关的E X T 2 突变型的检测可提供一种诊断工具，此工具可增加或限定对疾病的诊断或对疾病的易感性，所述疾病是由E X T 2 的过少表达，过多表达或经改变的表达引起的。通过多种技术可在DNA水平上检测到携有E X T 2 基因突变的个体。诊断用的核酸可得自病人的细胞，如血液，尿液，唾液、组织活检或尸检材料。可直接将基因组DNA用于检测，或者也可在分析前通过使用P C R 来酶促扩增基因组DNA。P C R (Saiki et al., Nature, 1986, 324: 163-166)。也可以类似方式使用RNA或cDNA。作为例子，可使用与编码E X T 2 之核酸互补的P C R 引物来鉴定和分析E X T 2 的表达和突变。例如，通过扩增产物与正常基因型比较后从大小的变化可检测缺失和插入。通过使扩增的DNA与经过放射性标记的E X T 2 RNA或经放射性标记的E X T 2 反义DNA序列杂交可鉴定点突变。通过RNase A消化或解链温度的差异可区分精确匹配的序列和不匹配的双螺旋。

通过直接的DNA测序也可揭示出参考基因和具有突变之基因之间的序列差异。另外，也可使用克隆的DNA片段作为探针以检测特殊的DNA片段。通过适当使用P C R 或其他扩增方法可大大增强这种方法的敏感性。例如，由经修饰的P C R 产生的双链P C R 产物或单链模板分子使用了测序引物。通过带有放射性标记核苷酸的常规方法或通过带有荧光标记物的自动测序方法可进行序列测定。

通过用或不用变性剂，检测凝胶中DNA片段电泳泳动率的变化可达到基于DNA序列差异的基因测定。通过高分辨的凝胶电泳，肉眼即可观察到小的序列缺失和插入。在变性的甲酰胺梯度凝胶上可区分序列不同的DNA片段，在此凝胶中，不同DNA片段根据其特异的解链或部分解链温度，其泳动被阻滞在凝胶中的不同位置(例见Myers et al., Science, 1985, 230: 1242)。

通过核酸酶保护测定法，如RNase和S1保护或化学裂解法也可揭示出特异位置处的序列变化（例Cotton et al., Proc. Natl Acad. Sci, USA, 1985, 85: 4397-4401）。

因此，通过例如杂交，RNase保护，化学裂解，直接DNA测序或使用限制性酶（如限制性片段长度多态性（“RFLP”）和基因组DNA的Southern印迹的方法可检测特殊的DNA序列。

根据本发明的另一方面，提供了诊断或确定遗传多发性外生骨疣和癌症（如软骨肉瘤等）之易感性的方法。EXT2基因的突变或能是对遗传多发性外生骨疣和癌症（如软骨肉瘤等）之易感性的象征；在确定这种易感性的试验中可使用上述核酸序列。因此，例如可通过试验确定本文所述的EXT2基因中的突变，如取代，缺失，截短，插入、移码等，这种突变是遗传多发性外生骨疣和癌症（如软骨肉瘤等）之易感性的象征。

本发明提供了诊断疾病的方法，具体地说，提供了诊断遗传多发性外生骨疣和癌症（如软骨肉瘤等）的方法；包括测定患者样品中具有图1，SEQ ID NO: 1之序列的多核酸非正常减少或增加的表达水平。使用本领域熟知的任何定量多核苷酸的方法，如PCR，RT-PCR，PNase保护，Northern印迹和其他杂交法都可测量多核苷酸表达的减少或增加。

除了更常规的凝胶电泳和DNA测序外，原位分析也可检测到突变。

多肽的测定

本发明还涉及检测细胞组织中EXT2蛋白质水平的诊断试验。这种试验可以是定量或定性的，因此，例如，可使用根据本发明用于检EXT2蛋白之过少表达（与正常的对照组织样品相比）的诊断试验来检测遗传多发性外生骨疣和癌症（如软骨肉瘤等）的存在。本领域技术人员熟知用于测定宿主样品中蛋白质（如本发明的EXT2蛋白）水平的测定技术，这种测定方法包括放射性免疫测定，竞争性竞争试验，Western印迹分析和酶联免疫吸附试验（ELISA），其中常优选ELISA。ELISA试验首先包括制备EXT2特异性抗体，优选为单克隆抗体。此外正常需制备与单克隆抗体结合的报道抗体。报道抗体与可检测的试剂结合，如放射性、荧光或酶试剂、如辣根过氧化物酶）。

为了进行E L I S A，可从宿主中取出样品，并在可束缚样品中蛋白质的固体支撑物，如聚苯乙烯平皿中保温。然后通过与非特异性蛋白质，如牛血清白蛋白一起保温即可封闭平皿上任何空余的蛋白质结合位点。在平皿中保温单克隆抗体，其时单克隆抗体与结合剂聚苯乙烯平皿上的任何E X T 2 蛋白结合，用缓冲液洗去未结合的单克隆抗体。在平皿中加入与辣根过氧化物酶相联的报道抗体，可使报道抗体与结合剂E X T 2 蛋白上的任何单克隆抗体结合，然后洗去未结合的报道抗体，在平皿中加入过氧化物酶活性试剂，包括比色底物。通过一级和二级抗体与E X T 2 相连的固定化过氧化物酶产生了有色反应产物。在给定时间内产色的量表示样品中所存在E X T 2 蛋白的量，典型地，通过参照标准曲线可得到定量的结果。

也可使用竞争性试验，其中固体支持物结合的E X T 2 特异性抗体与经标记的E X T 2 和宿主的样品从固体支持物上通过。检测到的与固体支撑物结合之标记物的量与样品中E X T 2 的量相关。

抗体

也可以将多肽，其片段或其他衍生物或其类似物，或表达它们的细胞用作免疫原以产生其抗体。例如，这些抗体可以是多克隆或单克隆抗体。本发明也包括嵌合、单链和人源化抗体，以及F ab片段或F ab表达文库的产物。也可使用本领域已知的各种方法生产这种抗体和片段。

通过本领域技术人员已知的多种方法可得到抗对应于本发明序列之多肽而产生的抗体。例如，在一个实施方案中，将多肽直接注射给动物，优选注射给非人类动物。再将如此得到的抗体与多肽自身结合。在此实施方案中，甚至也可使用仅编码多肽片段的序列以产生与整个天然多肽结合的抗体。可使用这种抗体从表达所述多肽的组织中分离多肽。

为了制备单克隆抗体，可使用能提供由连续细胞系培养产生的抗体的任何技术，例如杂交瘤技术 (Kohler, G. and Milstein, C., Nature, 1975

256: 495-497), trioma技术, 人B-细胞杂交瘤技术 (Kozbor et al., Immunology Today, 1983, 4: 72) 和EBv-杂交瘤技术 (Cole et al., MONOCLONAL ANTIBODIES AND CANCER THERAPY, pg, 77-96, Alan R. Liss, Inc., 1985)。

经描述用于产生单链抗体的技术（美国专利4,946,778）也适于产生本发明免疫原性多肽产物的单链抗体。也可使用转基因小鼠，或包括其他哺乳动物的其他生物来表达本发明免疫原性多肽产物的人源化抗体。

通过将抗体与固体支持物结合以通过亲和层析分离和/或纯化，即可使用上述抗体分离或鉴定表达多肽的克隆或纯化本发明多肽。

可使用抗E X T 2 的抗体来抑制遗传多发性外生骨疣和癌症，如软骨肉瘤及其他。

E X T 2 结合分子和测定

也可使用E X T 2 以分离与其相互作用的蛋白质；这种相互作用可能是干扰的靶子。E X T 2 和其他因子之间蛋白质-蛋白质相互作用的抑制剂会导致能调节E X T 2 活性的药剂的产生。

因此，本发明也提供了鉴定E X T 2 之结合分子的方法。通过本领域技术人员已知的多种方法如配体淘选和F A C S 分类法即可鉴定编码E X T 2 结合分子之蛋白质的基因。这种方法描述于多种实验手册中，如Coligan et al., CURRENT PROTOCOLS IN IMMUNOLOGY, 第5章, 1991。

例如，酵母双杂合体系使用转录激活剂活性的重建可提供体内检测第一种试验蛋白质和第二种试验蛋白质之间相互作用的方法。美国专利5,283,173 中公开了此方法；可使用Clontech和Stratagene的试剂。简而言之，使E X T 2 c D N A 与G a l 4 转录因子D N A 结合区融合并在酵母细胞中表达。将得自感兴趣细胞的c D N A 文库成员与G a l 4 的反式激活区融合，表达能与E X T 2 相互作用的蛋白质的c D N A 克隆会导致G a l 4 活性的重建和如G a l 4 -l a c Z 的报道基因表达的反式激活作用。

另一种方法涉及用重组 E X T 2 筛选 λ g t 1 1 λ Z A P (Stratagene) 或等价的c D N A 表达文库。重组的E X T 2 蛋白或其片段与如F L A G、H S V 或G S T 的小肽标记物融合。所述肽标记物具有便利的激酶（如心肌肌酸激酶）磷酸化位点或者它们可被生物来酰化。重组E X T 2 可被 32 [P]磷酸化，也可以未经标记的形式使用之，用链霉亲和素或抗此标记物的抗体可检测之。从感兴趣的细胞中制得 λ g t 1 1 c D N A 表达文库，并与重组E X T 2 一起保温，洗涤并分离与E X T 相互作用的c D N A 克隆。此方

法是本领域技术人员常规使用的方法，例见Sambrook et al.

另一种方法是筛选哺乳动物表达文库，在此方法中，在载体的哺乳动物启动子和多聚腺苷酸化位点之间克隆cDNA并瞬时转染到COS或293细胞中。48小时后，通过将固定的和经洗涤的细胞与经标记的EXT2一起保温可检测结合蛋白。在优选的实施方案中，EXT2被碘化，通过放射自显影术可检测任何结合的EXT2，例见Sims et al., Science, 1988, 241: 585-589和McMahan et al., EMBO J., 1991, 10: 2821-2832。以这种方式，可选择含有编码感兴趣之结合蛋白的cDNA库，通过进一步次分每个库，接着进行瞬时转染、结合和放射自显影的循环可分离感兴趣的cDNA。另外，通过将完整的cDNA文库转染到哺乳动物细胞中并在含有与平板结合之EXT2的平皿中淘选细胞可分离感兴趣的cDNA。裂解洗涤后仍附着的细胞，分离质粒DNA，在细菌中扩增，重复转染和淘选的循环直至得到单个cDNA克隆。见Seed et al., Proc. Natl. Acad. Sci. USA 1987, 84: 3365和Aruffo et al., EMBO J., 1987, 6: 3313。如果结合蛋白是可分泌的，则一旦建立了结合或中和试验以测定瞬时转染细胞的上清液即可通过类似的建库策略得到其cDNA。Wong et al., Science, 1985, 228: 810-815中公开了筛选上清液的一般方法。

另一种方法涉及直接从细胞中分离与EXT2相互作用的蛋白质。制备EXT2与GST或小的肽标记物的融合蛋白并固定在小珠上。制备得自感兴趣细胞的生物合成标记或未标记的蛋白质提取物，与小珠一起保温并用缓冲液洗涤。从小珠上特异性地洗脱与EXT2相互作用的蛋白质，通过SDS-PAGE分析之。通过微量测序得到结合对象的一级氨基酸序列数据。任性地，可用能诱导功能性反应（如细胞蛋白质的酪氨酸磷酸化）的试剂处理细胞。这种试剂的例子可以是生长因子或细胞因子，如白细胞介素-2。

另一种方法是免疫亲和纯化。将重组EXT2与标记或未标记的细胞提取物上起保温，并用抗人抗体免疫沉淀。用蛋白质A-Sepharose回收免疫沉淀物并通过SDS-PAGE分析。通过生物来酰化作用标记未标记的蛋白质，并在含链霉亲和素的SDS凝胶上检测。通过微量测序分析结合对象蛋白质。另外，在微量测序前可使用本领域技术人员已知的标准生化纯化步骤。

另一种方法涉及筛选结合对象的肽文库。使用重组的经标记的E X T 2 从与E X T 2 相互作用的肽或磷酸肽文库中选择肽。测定肽的序列可鉴定出在相互作用的蛋白质中可能会发现的共有肽序列。

激动剂和拮抗剂 - 测定和分子

在筛选可激活此多肽 (激动剂) 或抑制此多肽激活 (拮抗剂) 的化合物的方法中, 可使用本发明的E X T 2 。

例如, 可以表达与E X T 2 结合分子 (如受E X T 2 调节的信号或调节途径分子) 的细胞中制备细胞区室, 如膜或其制品 (如膜制品)。在可能是E X T 2 激动剂或拮抗剂的候选分子缺乏或存在的情况下, 将制品与经标记的E X T 2 一起保温。经标记的配体结合的减少可反映出候选分子与结合分子结合的能力。无结合分子, 即会不诱导E X T 2 对E X T 2 结合分子结合作用的分子可能是最好的拮抗剂。结合良好并产生与E X T 2 相同或密切相关之作用的分子是激动剂。

例如, 通过测定候选分子与细胞或适当的细胞制品相互作用之后第二信使系统的活性, 将其作用效果与E X T 2 的或能产生与E X T 2 相同之作用的分子的作用效果相比较, 可测定潜在的激动剂和拮抗剂的E X T 2 一样作用。在此方面有用的第二信使系统包括但不限于c A M P 鸟苷酸环化酶, 离子通道或磷酸肌醇水解第二信使系统。

E X T 2 拮抗剂测定法的另一个例子是竞争性试验, 即在适当的竞争性抑制试验的条件下, 使E X T 2 和潜在的拮抗剂与膜结合的E X T 2 受体分子或重组的E X T 2 受体分子结合。E X T 2 可被如放射性标记以使与受体分子结合的E X T 2 分子数目能被精确测定从而是估计出潜在拮抗剂的效力。

潜在拮抗剂的例子包括抗体, 或在某些情况下还包括能与多肽结合但不会产生可抑制多肽活性之第二信使反应的寡核苷酸。

潜在的拮抗剂也包括已丧失酶活性的与E X T 2 密切相关的蛋白质, 即多肽片段。

潜在的拮抗剂也包括通过使用反义技术制备的反义构建体。使用反义技术可通过三股螺旋的形成或反义D N A 或R N A 来控制基因的表达, 两种方法都基于多核苷酸与D N A 或R N A 的结合。例如, 可使用编码本发明成熟多肽的

多核苷酸序列的5' 编码部分来设计一个长度约为10至40个碱基对的反义RNA寡核苷酸。经设计DNA寡核苷酸与涉及转录的基因区互补(三股螺旋, 见Lee et al., Nucl. Acids Res., 1979, 6: 3073; Cooney et al., Science, 1988, 241: 456; and Dervan et al., Science, 1991, 251: 1360), 从而可防止EXT2的转录和产生。反义RNA寡核苷酸与mRNA体内杂交, 并阻断了mRNA分子翻译成酶(反义-见Okano, J. Neurochem., (1991) 56: 560; Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression, CRC Press, Boca Raton, FL(1988))。上述寡核苷酸也可被传递到细胞中以使反义RNA或DNA在体内表达从而抑制EXT2的产生。

另一种潜在的拮抗剂是小分子, 它与多肽结合, 使得它不能与结合分子接近从而抑制了正常的生物活性。小分子的例子包括但不限于小肽或肽样分子。

潜在的拮抗剂也包括EXT2的可溶性形式, 例如多肽片段, 它与结合分子结合, 从而防止结合分子与膜结合的EXT2相互作用。

在哺乳动物宿主中, EXT2无处不在, 它对包括多种病理的各种生物学功能起作用。因此, 希望能发现一方面可刺激酶活性, 另一方面, 可抑制EXT2功能的化合物和药物。

可将EXT2拮抗剂用于此类遗传多发性外生骨疣和癌症(如软骨肉瘤等)的多种治疗和预防目的。

本发明还提供了治疗与EXT2活性过量相关之非正常状况的方法, 包括给受试者施用有效量的上述抑制剂化合物(拮抗剂)以及药用可接受的载体以抑制酶的激活, 或通过抑制第二信号从而减轻非正常状况。

通常, 可将EXT2激动剂用于治疗和预防遗传多发性外生骨疣和癌症(如软骨肉瘤等)的目的。

本发明还涉及治疗与EXT2及其活性过少表达相关之非正常状况的方法, 包括给受试者施用治疗有效量的可激活多肽的化合物(激动剂), 从而减轻遗传多发性外生骨疣和癌症(如软骨肉瘤等)的非正常状况。

组合物和试剂盒

可将可溶性形式的E X T 2 以及可激活或抑制此酶的化合物与适当的药用载体联合使用。这种组合物含有治疗有效量的多肽或化合物，以及药物可接受的载体或赋形剂。这种载体包括但不限于盐水、缓冲盐水、葡萄糖、水，甘油，乙醇及它们的组合。配方应适合施用的方式。本领域技术人员可常规进行根据施用方式选择适当载体。

本发明进一步涉及药物成套试剂或试剂盒，该试剂盒含有一个或多个容器，所述容器充满一种或多种上述本发明组合物成分。

施用

可单独或与其他化合物（如治疗化合物）一起使用本发明多肽和其他化合物。

可以任何有效、便利的方式施用药物组合物，包括例如通过局部、口服、肛门，阴道，静脉内，腹膜内，肌内，皮下、鼻内或皮内途径等施药。

通常以有效治疗或预防特异性症状的量施用药物组合物。一般以至少约 $10 \mu\text{g} / \text{kg}$ 体重的量施用组合物。大多数情况下以不超过每天的 $8 \text{mg} / \text{kg}$ 体重的量施用组合物，大多数情况下优选的施药剂量是每天约 $10 \mu\text{g} / \text{kg}$ 至约 $1 \text{mg} / \text{kg}$ 体重。应理解考虑到症状，其严重程度，施药方式，复杂情况等，通过每种治疗方式和疗法的标准方法可确定最适的剂量。

基因治疗

通过在治疗方法中表达多肽（通常指的是“基因治疗”）可根据本发明使用E X T 2 多核苷酸，多肽，激动剂和拮抗剂（都是多肽）。

因此，例如可用多核苷酸（如DNA或RNA）改造来自病人的细胞以编码来自间接体内的多肽。然后将经改造的细胞提供给用多肽治疗的病人。在此实施方案中，通过例如使用含有编码本发明多肽之RNA的逆转录病毒质粒载体可在间接体内选细胞。本领域已知这种方法，从本文的教导中可看出它们在本发明中的应用是显而易见的。

类似地，可通过本领域已知的方法在体内改造细胞以在体内表达多肽。例如，可改造本发明多核苷酸以在上述的复制缺陷型逆转录病毒载体中表达。然

后分离逆转录病毒表达构建体，并导入经逆转录病毒质粒载体转导的包装细胞，所述载体含有编码本发明多肽的RNA，包装细胞从而可产生含有感兴趣基因的感染性病毒颗粒。可以给病人施用这些生产细胞以在体内改造细胞并在体内表达多肽。鉴于本发明的教导，这些和其他施用本发明多肽的方法对于本领域技术人员而言是显而易见的。

本文上述逆转录病毒质粒载体之来源的逆转录病毒包括但不限于Moloney鼠白血病病毒，脾坏死病毒，劳氏肉瘤病毒，Harvey肉瘤病毒，禽白血病病毒，Gibbon Ape 白血病病毒，人免疫缺损病毒，腺病毒，骨髓增殖肉瘤病毒和乳腺癌病毒。在优选的实施方案中，逆转录病毒质粒载体衍生自Moloney鼠白血病病毒。

这种载体会包括一种或多种启动子以表达多肽。可以使用的适当的启动子包括但不限于逆转录病毒LTR；SV40启动子；和人肥大细胞病毒（CMV）启动子（描述于Miller et al., Biotechniques, 1989, 7: 980-990）。也可使用如真核细胞启动子的细胞启动子，它们包括但不限于组蛋白，RNA聚合酶III和 β -肌动蛋白启动子，可以使用的其他病毒启动子包括但不限于腺病毒启动子，胸苷激酶（TK）启动子和B19细小病毒启动子。鉴于本文所含的教导，适当启动子的选择对本领域技术人员而言是显而易见的。

编码本发明多肽的核酸序列被置于适当启动子的控制之下，可使用的适当启动子包括但不限于腺病毒启动子，如腺病毒主要晚期启动子；或异源启动子，如肥大细胞病毒（CMV）启动子；呼吸道合胞病毒（RSV）启动子，可诱导的启动子，如MMT启动子，金属硫蛋白启动子；热休克启动子；白蛋白启动子；A_{pi}AI启动子；人珠蛋白启动子；病毒胸苷激酶启动子，如单纯疱疹病毒胸苷激酶启动子；逆转录病毒LTR（包括上述的经修饰的逆转录病毒LTR）； β -肌动蛋白启动子；和人生长激素启动子。启动子也可以是能控制编码多肽之基因的天然启动子。

使用逆转录病毒质粒载体转导包装细胞系以形成生产细胞系。可被转染的包装细胞的例子包括但不限于PE501，PA317，Y-2，Y-AM，PA12，T19-14X，VT-19-17-H2，YCRE YCRIP，Gp+E-86，Gp+envAm12和DAN细胞系（描述于Miller, A.,

Human Gene Therapy, 1990, 1: 5-14)。可通过本领域已知的任何方法将载体转导到包装细胞内。这种方法包括但不限于电穿孔, 脂质体的使用和CaPO₄沉淀。在一个替代方法中, 将逆转录病毒质粒载体包熏在脂质体中或与脂质偶联, 然后给宿主施用。

生产细胞系会产生感染性逆转录病毒载体颗粒, 它包括编码多肽的核酸序列。然后可使用这种逆转录病毒载体颗粒在体外或体内转导真核细胞。经转导的真核细胞可表达编码多肽的核酸序列。可被转导的真核细胞包括但不限于胚胎干细胞, 胚胎癌细胞以及造血干细胞, 肝细胞, 成纤维细胞, 成肌细胞, 角质形成细胞, 内皮细胞和支气管上皮细胞。

实施例

E X T 2 的基因治疗表达

通过皮肤活检从受试者体内得到成纤维细胞, 将所得组织置入组织培养基中并分离成小片。在组织培养瓶的潮湿表面放置组织小碎片, 每瓶中约放10片。将培养瓶倒转过来, 紧密靠近并置于室温中过夜, 在室温下放24小时后, 翻转培养瓶(组织碎片固定在瓶底), 加入新鲜培养基(例如Ham氏F12培养基, 它含有10%FBS, 青霉素和链霉素)。然后将组织在37℃下保温约1周。此时, 加入新鲜培养基, 随后每隔几天更换一次。再培养两周后, 出现单层成纤维细胞, 用胰蛋白酶消化单层并扩大到大瓶规模。

用限制性酶消化用于基因治疗的载体以克隆需表达的片段, 用牛小肠磷酸酶处理经消化的载体以防止自身连接。在琼脂糖凝胶上分级分离去磷酸化的线性载体并纯化之。

分离能表达活性E X T 2 的E X T 2 c D N A, 如有必要可修饰片段的末端以克隆到载体上。例如, 可用D N A 聚合酶处理5'突出端以产生钝端。使用S 1 核酸酶可除去3'突出端。用T 4 D N A 连接酶可将接头连接到钝端上。

将等量的Moloney 鼠白血病病毒线性骨架和E X T 2 片段相混合并使用T 4 D N A 连接酶连接。使用连接混合物转化大肠杆菌, 再将细菌涂布在含卡那霉素的琼脂上, 卡那霉素类型和限制性分析可确证载体已正确地插入基因。

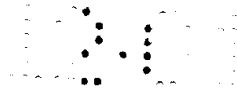
在含有10%牛血清(CS), 青霉素和链霉素的Dulbecco氏修改的Eagles培养基(DMEM)中, 组织培养包装细胞, 使其生长后铺满密度。通过标准技术将含有EXT2基因的载体导入包装细胞。从现在被称作生产细胞的包装细胞中收集含有EXT2基因的感染性病毒颗粒。

在生产细胞中加入新鲜培养基, 保温适当时间之后, 从铺满生产细胞的平板中收集培养基。通过Millipore滤器过滤含有感染性病毒颗粒的培养基以除去脱附的生产细胞。然后使用经过滤的培养基感染成纤维细胞。从成纤维细胞的汇合品平板中除去培养基, 并快速换上经过滤的培养基。在培养基中可包括溴化己二甲铵(Aldrich)以便于转导。适当保温后除去培养基并换上新鲜培养基。如果病毒的滴度高, 实际上所有的成纤维细胞将被杂而不需要选择。如果病毒的滴度低, 则有必要使用具有选择性标记(如neo或his)的逆转录病毒载体以选择出被转导的细胞用于扩充。

或者单独或者在微载体珠(如cytodex 3珠)上已生长而汇合后将经改造的成纤维细胞注射到大鼠体内。经注射的成纤维细胞产生了EXT2产物, 蛋白质的生物学作用传递给宿主。

显然, 也可不以上述描述部分和实施例所具体描述的那样实施本发明。

鉴于上文的教导, 本发明的各种修饰和变化都是可能的, 因此它们在所附权利要求的范围之内。



序列表

(1) 一般信息

- (i) 申请人: 湖南医科大学
- (ii) 发明名称: EXT2
- (iii) 序列数目: 2
- (iv) 通讯地址:
 - (A) 地址: SmithKline Beecham Corporation
 - (B) 街道: 709 Swedeland Road
 - (C) 城市: King of Prussia
 - (D) 州: PA
 - (E) 国家: USA
 - (F) 邮政编码: 19406
- (v) 计算机可读形式:
 - (A) 介质形式: 磁盘
 - (B) 计算机: IBM Compatible
 - (C) 操作系统: DOS
 - (D) 软件: FastSEQ Version 1.5
- (vi) 本申请信息:
 - (A) 申请号:
 - (B) 提交日:
 - (C) 分类:
- (vii) 在先申请信息:
 - (A) 申请号:
 - (B) 提交日:
- (viii) 代理人/代理机构信息:
 - (A) 姓名: William T. Han
 - (B) 登记号: 34, 344
 - (C) 参考/档案号: ATG50019
- (ix) 电讯信息:
 - (A) 电话: (610) 270-5219
 - (B) 电话传真: (610) 270-5090
 - (C) 电报:



(2) SEQ ID NO: 1信息:

(i) 序列特征:

- (A) 长度: 3003 个碱基对
- (B) 类型: 核酸
- (C) 链数: 单链
- (D) 拓扑结构: 线性

(ii) 分子类型: cDNA

(iii) 假设: 无

(iv) 反义: 无

(v) 片段类型:

(vi) 原始来源:

(xi) 序列描述: SEQ ID NO: 1

```
CTCGCCAGCC CAGACTCGGC CCTGGCAGTG GCGGCTGGCG ATTCGGACCG ATCCGACCTG      60
GGCGGAGGTG GCCC GCGCCC CGCGGCATGA GCCGGTGACC AAGCTCGGGG CCGAGCGGGA      120
GGCAGCCGTG GCCGAGGAGT GTGAGGAAGA GGCTGTCTGT GTCATTATGT GTGCGTCGGT      180
CAAGTATAAT ATCCGGGGTC CTGCCCTCAT CCCAAGAATG AAGACCAAGC ACCGAATCTA      240
CTATATCACC CTCTTCTCCA TTGTCCTCCT GGGCCTCATT GCCACTGGCA TGTTTCAGTT      300
TTGGCCCCAT TCTATCGAGT CCTCAAATGA CTGGAATGTA GAGAAGCGCA GCATCCGTGA      360
TGTGCCGGTT GTTAGGCTGC CAGCCGACAG TCCCATCCCA GAGCGGGGGG ATCTCAGTTG      420
CAGAATGCAC ACGTGT TTTG ATGTCTATCG CTGTGGCTTC AACCCAAAGA ACAAATCAA      480
GGTGTATATC TATGCTCTGA AAAAGTACGT GGATGACTTT GGCGTCTCTG TCAGCAACAC      540
CATCTCCCGG GAGTATAATG AACTGCTCAT GGCCATCTCA GACAGTGACT ACTACACTGA      600
TGACATCAAC CGGGCCTGTC TGT TTTGTTCC CTCCATCGAT GTGCTTAACC AGAACACACT      660
GCGCATCAAG GAGACAGCAC AAGCGATGGC CCAGCTCTCT AGGTGGGATC GAGGTACGAA      720
TCACCTGTTG TTCAACATGT TGCCTGGAGG TCCCCAGAT TATAACACAG CCCTGGATGT      780
CCCCAGAGAC AGGGCCCTGT TGGCTGGTGG CGGCTTTTCT ACGTGGACTT ACCGCAAGG      840
CTACGATGTC AGCATTCCTG TCTATAGTCC ACTGTCAGCT GAGGTGGATC TTCCAGAGAA      900
AGGACCAGGT CCACGGCAAT ACTTCCTCCT GTCATCTCAG GTGGGTCTCC ATCCTGAGTA      960
CAGAGAGGAC CTAGAAGCCC TCCAGGTCAA ACATGGAGAG TCAGTGTTAG TACTCGATAA     1020
ATGCACCAAC CTCTCAGAGG GTGTCCTTTC TGTCCGTAAG CGCTGCCACA AGCACCAGGT     1080
CTTCGATTAC CCACAGGTGC TACAGGAGGC TACTTCTGT GTGGTTCTTC GTGGAGCTCG     1140
GCTGGGCCAG GCAGTATTGA GCGATGTGTT ACAAGCTGGC TGTGTCCCGG TTGTCATTGC     1200
AGACTCCTAT ATTTTGCTT TCTCTGAAGT TCTTGACTGG AAGAGAGCAT CTGTGGTTGT     1260
ACCAGAAGAA AAGATGTCAG ATGTGTACAG TATTTTGCAG AGCATCCCC AAAGACAGAT     1320
TGAAGAAATG CAGAGACAGC TCTTCATGGA ACCAGTCAGG AGAGAGAACT GGTCAGCTGC     1380
TAATCACCAA ATGAACTCCC TGATCTGGCC TAGGGAACAG TGGGATTCAC AGATTATCAA     1440
TGACCGGATC TATCCATATG CTGCCATCTC CTATGAAGAA TGGAATGACC CTCCTGCTGT     1500
GAAGTGGGGC AGCGTGAGCA ATCCACTCTT CCTCCCGCTG ATCCCACCAC AGTCTCAAGG     1560
```



G TTCACCGCC ATAGTCCTCA CCTACGACCG AGTAGAGAGC CTCTTCCGGG TCATCACTGA 1620
AGTGTC CAAG GTGCCAGTC TATCCAACT ACTTGTCGTC TGGAATAATC AGAATAAAAA 1680
CCCTCCAGAA GATTCTCTCT GGCCCAAAT CCGGGTTCCA TTAAAAGTTG TGAGGACTGC 1740
TGAAAACAAG TTAAGTAACC GTTTCTTCCC TTATGATGAA ATCGAGACAG AAGCTGTTCT 1800
GGCCATTGAT GATGATATCA TTATGCTGAC CTCTGACGAG CTGCAATTTG GTTATGAGGT 1860
CTGGCGGGAA TTTCTTGACC GGTGGTGGG TTACCCGGAT CGTCTGCATC TCTGGGACCA 1920
TGAGATGAAT AAGTGGAAGT ATGAGTCTGA GTGGACGAAT GAAGTGTCCA TGGTGTCTAC 1980
TGGGGCAGCT TTTTATCACA AGTATTTTAA TTACCTGTAT ACCTACAAA TGCCTGGGGA 2040
TATCAAGAAC TGGGTAGATG CTCATATGAA CTGTGAAGAT ATTGCCATGA ACTTCCTGGT 2100
GGCCAACGTC ACGGGAAAAG CAGTTATCAA GGTAACCCCA CGAAAGAAAT TCAAGTGTCC 2160
TGAGTGCACA GCCATAGATG GGCTTTCCT AGACCAAACA CACATGGTGG AGAGGTCAGA 2220
GTGCATCAAC AAGTTTGCTT CAGTCTTCGG GACCATGCCT CTCAAGGTGG TGGAACACCG 2280
AGCTGACCCT GTCCTGTACA AAGATGACTT TCCTGAGAAG CTGAAGAGCT TCCCCAACAT 2340
TGGCAGCTTA TGAAACGTGT CATTGGTGGG GGTCTGAATG TGAGGCTGGG ACAGAGGGAG 2400
AGAACAAGGC CTCCAGCAC TCTGATGTCA GAGTAGTAGG TTAAGGGTGG AAGGTTGACC 2460
TACTTGATC TTGGCATGCA CCCACCTAAC CCACTTTCTC AAGAACAAGA ACCTAGAATG 2520
AATATCCAAG CACCTCGAGC TATGCAACCT CTGTTCTTGT ATTTCTTATG ATCTCTGATG 2580
GGTTCTTCTC GAAAATGCCA AGTGAAGAC TTTGTGGGCA TGCTCCAGA TTTAAATCCA 2640
GCTGAGGCTC CCTTTGTTTT CAGTTCATG TAACAATCTG GAAGGAACT TCACGGACAG 2700
GAAGACTGCT GGAGAAGAGA AGCGTGTAG CCCATTTGAG GTCTGGGGAA TCATGTAAAG 2760
GGTACCCAGA CCTCACTTTT AGTTATTTAC ATCAATGAGT TCTTTCAGGG AACCAAACCC 2820
AGAATTCCGGT GCAAAAGCCA AACATCTTGG TGGGATTTGA TAAATGCCTT GGGACCTGGA 2880
GTGCTGGGCT TGTGCACAGG AAGAGACCA GCCGCTGAGT CAGGATCCTG TCAGTTCAT 2940
GAGCTATTCC TCTTTGTTT GGCTTTTGA TATGATTAAT ATTATTTTTT ATTCCTTTTA 3000
AAA 3003

(2) SEQ ID NO: 2信息:

(i) 序列特征:

- (A) 长度: 728 个氨基酸
- (B) 类型: 氨基酸
- (C) 链数: 单链
- (D) 拓扑结构: 线性

(ii) 分子类型: 肽

(iii) 假设: 无


(iv) 反义: 无

(v) 片段类型: N-末端

(vi) 原始来源:

(xi) 序列描述: SEQ ID NO: 2:

Met Cys Ala Ser Val Lys Tyr Asn Ile Arg Gly Pro Ala Leu Ile Pro
1 5 10 15



Lys Met Ser Asp Val Tyr Ser Ile Leu Gln Ser Ile Pro Gln Arg Gln
 370 375 380
 Ile Glu Glu Met Gln Arg Gln Leu Phe Met Glu Pro Val Arg Arg Glu
 385 390 395 400
 Asn Trp Ser Ala Ala Asn His Gln Met Asn Ser Leu Ile Trp Pro Arg
 405 410 415
 Glu Gln Trp Asp Ser Gln Ile Ile Asn Asp Arg Ile Tyr Pro Tyr Ala
 420 425 430
 Ala Ile Ser Tyr Glu Glu Trp Asn Asp Pro Pro Ala Val Lys Trp Gly
 435 440 445
 Ser Val Ser Asn Pro Leu Phe Leu Pro Leu Ile Pro Pro Gln Ser Gln
 450 455 460
 Gly Phe Thr Ala Ile Val Leu Thr Tyr Asp Arg Val Glu Ser Leu Phe
 465 470 475 480
 Arg Val Ile Thr Glu Val Ser Lys Val Pro Ser Leu Ser Lys Leu Leu
 485 490 495
 Val Val Trp Asn Asn Gln Asn Lys Asn Pro Pro Glu Asp Ser Leu Trp
 500 505 510
 Pro Lys Ile Arg Val Pro Leu Lys Val Val Arg Thr Ala Glu Asn Lys
 515 520 525
 Leu Ser Asn Arg Phe Phe Pro Tyr Asp Glu Ile Glu Thr Glu Ala Val
 530 535 540
 Leu Ala Ile Asp Asp Asp Ile Ile Met Leu Thr Ser Asp Glu Leu Gln
 545 550 555 560
 Phe Gly Tyr Glu Val Trp Arg Glu Phe Pro Asp Arg Leu Val Gly Tyr
 565 570 575
 Pro Asp Arg Leu His Leu Trp Asp His Glu Met Asn Lys Trp Lys Tyr
 580 585 590
 Glu Ser Glu Trp Thr Asn Glu Val Ser Met Val Leu Thr Gly Ala Ala
 595 600 605
 Phe Tyr His Lys Tyr Phe Asn Tyr Leu Tyr Thr Tyr Lys Met Pro Gly
 610 615 620
 Asp Ile Lys Asn Trp Val Asp Ala His Met Asn Cys Glu Asp Ile Ala
 625 630 635 640
 Met Asn Phe Leu Val Ala Asn Val Thr Gly Lys Ala Val Ile Lys Val
 645 650 655
 Thr Pro Arg Lys Lys Phe Lys Cys Pro Glu Cys Thr Ala Ile Asp Gly
 660 665 670
 Leu Ser Leu Asp Gln Thr His Met Val Glu Arg Ser Glu Cys Ile Asn
 675 680 685
 Lys Phe Ala Ser Val Phe Gly Thr Met Pro Leu Lys Val Val Glu His
 690 695 700
 Arg Ala Asp Pro Val Leu Tyr Lys Asp Asp Phe Pro Glu Lys Leu Lys
 705 710 715 720



Ser Phe Pro Asn Ile Gly Ser Leu
725

说明书附图

图 1. EXT 2 cDNA 序列和预测的氨基酸序列

ctgccagcccagactcggcctggcagtgggcgtggcgattcggaccgatccgacctgggagggtggcccgcgc
cccgcggcatgagccggtgaccaagctcggggccgagcgggaggcagccgtggccgaggagtgtgaggaagaggctgt
ctgtgtcatt

atg tgt gcg tgc gtc aag tat aat atc cgg ggt cct gcc ctc atc cca aga atg aag
M C A S V K Y N I R G P A L I P R M K

acc aag cac cga atc tac tat atc acc ctc ttc tcc att gtc ctc ctg ggc ctc att
T K H R I Y Y I T L F S I V L L G L I

gcc act ggc atg ttt cag ttt tgg ccc cat tct atc gag tcc tca aat gac tgg aat
A T G M F Q F W P H S I E S S N D W N

gta gag aag cgc agc atc cgt gat gtg ccg gtt gtt agg ctg cca gcc gac agt ccc
V E K R S I R D V P V V R L P A D S P

atc cca gag cgg ggg gat ctc agt tgc aga atg cac acg tgt ttt gat gtc tat cgc
I P E R G D L S C R M H T C F D V Y R

tgt ggc ttc aac cca aag aac aaa atc aag gtg tat atc tat gct ctg aaa aag tac
C G F N P K N K I K V Y I Y A L K K Y

gtg gat gac ttt ggc gtc tct gtc agc aac acc atc tcc cgg gag tat aat gaa ctg
V D D F G V S V S N T I S R E Y N E L

ctc atg gcc atc tca gac agt gac tac tac act gat gac atc aac cgg gcc tgt ctg
L M A I S D S D Y Y T D D I N R A C L

ttt gtt ccc tcc atc gat gtg ctt aac cag aac aca ctg cgc atc aag gag aca gca
F V P S I D V L N Q N T L R I K E T A

caa gcg atg gcc cag ctc tct agg tgg gat cga ggt acg aat cac ctg ttg ttc aac
Q A M A Q L S R W D R G T N H L L F N

atg ttg cct gga ggt ccc cca gat tat aac aca gcc ctg gat gtc ccc aga gac agg
M L P G G P P D Y N T A L D V P R D R

gcc ctg ttg gct ggt ggc ggc ttt tct acg tgg act tac cgg caa ggc tac gat gtc
 A L L A G G G F S T W T Y R Q G Y D V

agc att cct-gtc tat agt cca ctg tca gct gag gtg gat ctt cca gag aaa gga cca
 S I P V Y S P L S A E V D L P E K G P

ggt cca cgg caa tac ttc ctc ctg tca tct cag gtg ggt ctc cat cct gag tac aga
 G P R Q Y F L L S S Q V G L H P E Y R

gag gac cta gaa gcc ctc cag gtc aaa cat gga gag tca gtg tta gta ctc gat aaa
 E D L E A L Q V K H G E S V L V L D K

tgc acc aac ctc tca gag ggt gtc ctt tct gtc cgt aag cgc tgc cac aag cac cag
 C T N L S E G V L S V R K R C H K H Q

gtc ttc gat tac cca cag gtg cta cag gag gct act ttc tgt gtg gtt ctt cgt gga
 V F D Y P Q V L Q E A T F C V V L R G

gct cgg ctg ggc cag gca gta ttg agc gat gtg tta caa gct ggc tgt gtc ccg gtt
 A R L G Q A V L S D V L Q A G C V P V

gtc att gca gac tcc tat att ttg cct ttc tct gaa gtt ctt gac tgg aag aga gca
 V I A D S Y I L P F S E V L D W K R A

tct gtg gtt gta cca gaa gaa aag atg tca gat gtg tac agt att ttg cag agc atc
 S V V V P E E K M S D V Y S I L Q S I

ccc caa aga cag att gaa gaa atg cag aga cag ctc ttc atg gaa cca gtc agg aga
 P Q R Q I E E M Q R Q L F M E P V R R

gag aac tgg tca gct gct aat cac caa atg aac tcc ctg atc tgg cct agg gaa cag
 E N W S A A N H Q M N S L I W P R E Q

tgg gat tca cag att atc aat gac cgg atc tat cca tat gct gcc atc tcc tat gaa
 W D S Q I I N D R I Y P Y A A I S Y E

gaa tgg aat gac cct cct gct gtg aag tgg ggc agc gtg agc aat cca ctc ttc ctc
 E W N D P P A V K W G S V S N P L F L

ccg ctg atc cca cca cag tct caa ggg ttc acc gcc ata gtc ctc acc tac gac cga
 P L I P P Q S Q G F T A I V L T Y D R

gta gag agc ctc ttc cgg gtc atc act gaa gtg tcc aag gtg ccc agt cta tcc aaa



V E S L F R V I T E V S K V P S L S K

cta ctt gtc gtc tgg aat aat cag aat aaa aac cct cca gaa gat tct ctc tgg ccc
L L V V W N N Q N K N P P E D S L W P

aaa atc cgg gtt cca tta aaa gtt gtg agg act gct gaa aac aag tta agt aac cgt
K I R V P L K V V R T A E N K L S N R

ttc ttc cct tat gat gaa atc gag aca gaa gct gtt ctg gcc att gat gat gat atc
F F P Y D E I E T E A V L A I D D D I

att atg ctg acc tct gac gag ctg caa ttt ggt tat gag gtc tgg cgg gaa ttt cct
I M L T S D E L Q F G Y E V W R E F P

gac cgg ttg gtg ggt tac cgg gat cgt ctg cat ctc tgg gac cat gag atg aat aag
D R L V G Y P D R L H L W D H E M N K

tgg aag tat gag tct gag tgg acg aat gaa gtg tcc atg gtg ctc act ggg gca gct
W K Y E S E W T N E V S M V L T G A A

ttt tat cac aag tat ttt aat tac ctg tat acc tac aaa atg cct ggg gat atc aag
F Y H K Y F N Y L Y T Y K M P G D I K

aac tgg gta gat gct cat atg aac tgt gaa gat att gcc atg aac ttc ctg gtg gcc
N W V D A H M N C E D I A M N F L V A

aac gtc acg gga aaa gca gtt atc aag gta acc cca cga aag aaa ttc aag tgt cct
N V T G K A V I K V T P R K K F K C P

gag tgc aca gcc ata gat ggg ctt tca cta gac caa aca cac atg gtg gag agg tca
E C T A I D G L S L D Q T H M V E R S

gag tgc atc aac aag ttt gct tca gtc ttc ggg acc atg cct ctc aag gtg gtg gaa
E C I N K F A S V F G T M P L K V V E

cac cga gct gac cct gtc ctg tac aaa gat gac ttt cct gag aag ctg aag agc ttc
H R A D P V L Y K D D F P E K L K S F

ccc aac att ggc agc tta tga
P N I G S L Stop

aacgtgtcattgggtggaggctgtaaatgtgaggctgggacagagggagagaacaaggcctcccagcactctgatgtcaga
gtagtaggttaagggtggaagggtgacactacttgatcttggcatgacccacctaaccactttctcaagaacaagaac



ctagaatgaatatccaagcacctcgagctatgcaacctctgttcttgtatttcttatgatctctgatgggttcttctgaa
aatgccaagtggaagactttgtgggcatgctcccagatttaaaccagctgaggctcccttggtttcagttccatgtaa
caatctggaaggaaacttcacggacaggaagactgctggagaagagaagcgtgtagccatttgaggtctggggaatc
atgtaaagggtaccagacctcacttttagttattacatcaatgagttcttccaggaaccaaaccagaattcgggtgc
aaaagccaaacatcttgggtgggatttgataaatgccttgggacctggagtgctgggcttggcacaggaagagcaccag
ccgctgagtcaggatcctgtcagttccatgagctattcctcttggtttgcttttggatgatgataaaattatttttatt
ccttttaaaa