

[19] 中华人民共和国国家知识产权局

[51] Int. Cl⁷



[12] 发明专利申请公开说明书

[21] 申请号 02828274.4

C07K 14/52

C12N 15/19

C12N 5/10

C12Q 1/68

C07K 16/28

G01N 33/566

A61K 38/19

[43] 公开日 2005 年 5 月 25 日

[11] 公开号 CN 1620466A

[22] 申请日 2002. 12. 20 [21] 申请号 02828274. 4

[30] 优先权

[32] 2001. 12. 21 [33] GB [31] 0130738. 8

[86] 国际申请 PCT/GB2002/005865 2002. 12. 20

[87] 国际公布 WO2003/055911 英 2003. 7. 10

[85] 进入国家阶段日期 2004. 8. 23

[71] 申请人 阿雷斯贸易股份有限公司

地址 瑞士欧博讷

[72] 发明人 M·D·戴维斯 C·B·菲尔普斯

R·J·法根 C·包尔 M·优克

M·伊伯森

[74] 专利代理机构 上海专利商标事务所有限公司

代理人 周承泽

权利要求书 5 页 说明书 54 页 附图 25 页

[54] 发明名称 胱氨酸结折迭蛋白质

[57] 摘要

本发明涉及新型蛋白质 (INSP002)，它经本发明鉴定为一种分泌性蛋白质，是胱氨酸结折迭细胞因子超家族的 Dan 家族成员，以及涉及该蛋白质和来自编码基因的核酸序列在诊断、预防和治疗疾病中的应用。

I S S N 1 0 0 8 - 4 2 7 4

1. 一种多肽，其特征在于：所述多肽
 - (i) 包括 SEQ ID NO: 6 或者 SEQ ID NO: 8 所示的氨基酸序列；
 - 5 (ii) 是其具有分泌性蛋白质的功能，优选具有胱氨酸结折迭细胞因子超家族成员的功能，更好具有 DAN 亚家族成员的功能，或者具有与(i)的多肽相同的抗原决定簇的片段；或者
 - (iii) 是(i)或(ii)的功能等同物。
2. 如权利要求 1 所述的多肽，其特征在于：所述多肽
 - 10 (i) 由 SEQ ID NO: 6 或者 SEQ ID NO: 8 所示的氨基酸序列组成；
 - (ii) 是其具有分泌性蛋白质的功能，优选具有胱氨酸结折迭细胞因子超家族成员的功能，更好具有 DAN 亚家族成员的功能，或者具有与(i)的多肽相同的抗原决定簇的片段；或者
 - (iii) 是(i)或(ii)的功能等同物。
- 15 3. 一种属于权利要求 1 或 2 (iii)的功能等同物的多肽，其特征在于：所述多肽与 SEQ ID NO: 6 或者 SEQ ID NO: 8 所示的氨基酸序列同源，并且具有胱氨酸结折迭细胞因子的活性，更好是具有 DAN 亚家族成员的活性。
4. 一种属于权利要求 1 至 3 中任何一项所述的片段或功能等同物的多肽，
 - 20 其特征在于：所述多肽与 SEQ ID NO: 6 或者 SEQ ID NO: 8 所示的氨基酸序列，或者与其活性片段的同一性大于 35%，更好是大于 40%、50%、60%、70%、80%、90%、95%、98%或 99%或以上。
5. 一种属于权利要求 1 至 4 中任何一项所述的功能等同物的多肽，其特征在于：所述多肽与具有 SEQ ID NO: 6 或者 SEQ ID NO: 8 所示的氨基酸序列的多肽有明显的结构同源性。
 - 25
6. 一种属于权利要求 1、2 或 4 中任何一项所述的片段的多肽，其特征在于：所述多肽具有与权利要求 1 至 3 中任何一项所述的(i)的多肽相同的抗原决定簇，由 SEQ ID NO: 6 或者 SEQ ID NO: 8 所示氨基酸序列的 7 或以上个氨基酸残基组成。
- 30 7. 一种编码上述权利要求中任何一项所述的多肽的纯化核酸分子。

8. 如权利要求 7 所述的纯化核酸分子, 其特征在于: 所述核酸分子包括 SEQ ID NO: 5 所示的核酸序列或者是其冗余等同物或片段。
9. 如权利要求 7 所述的纯化核酸分子, 其特征在于: 所述核酸分子由 SEQ ID NO: 5 所示的核酸序列组成, 或者是其冗余等同物或片段。
- 5 10. 如权利要求 7 所述的纯化核酸分子, 其特征在于: 所述核酸分子包括 SEQ ID NO: 5 所示的核酸序列的核苷酸 152 至 721(SEQ ID NO: 11)。
11. 如权利要求 7 所述的纯化核酸分子, 其特征在于: 所述核酸分子由 SEQ ID NO: 5 所示的核酸序列的核苷酸 152 至 721(SEQ ID NO: 11)组成。
12. 如权利要求 7 所述的纯化核酸分子, 其特征在于: 所述核酸分子包括 SEQ
10 ID NO: 5 所示的核酸序列的核苷酸 218 至 721(SEQ ID NO: 9)。
13. 如权利要求 7 所述的纯化核酸分子, 其特征在于: 所述核酸分子由 SEQ ID NO: 5 所示的核酸序列的核苷酸 218 至 721(SEQ ID NO: 9)组成。
14. 一种在高度严谨条件下与权利要求 8 至 13 中任何一项所述的核酸分子杂交的纯化核酸分子。
- 15 15. 一种含有权利要求 8 至 14 中任何一项所述的核酸分子的载体。
16. 一种用权利要求 15 所述的载体转化的宿主细胞。
17. 一种配体, 其特征在于: 所述配体特异性地结合权利要求 1 至 6 中任何一项所述的多肽, 最好抑制所述多肽的胱氨酸结折迭细胞因子的活性, 更好是 DAN 亚家族的活性。
- 20 18. 如权利要求 17 中所述的配体, 其特征在于: 所述配体是一种抗体。
19. 一种增加或降低权利要求 1 至 6 中任何一项所述的多肽的表达水平或活性的化合物。
20. 如权利要求 19 所述的化合物, 其特征在于: 所述化合物与权利要求 1 至 6 中任何一项所述的多肽结合, 并且不会诱导该多肽的任何生物作用。
- 25 21. 如权利要求 20 所述的化合物, 其特征在于: 所述化合物是天然或修饰的基质、配体、酶、受体或者结构或功能类似物。
22. 如权利要求 1 至 6 中任何一项所述的多肽、权利要求 7 至 14 中任何一项所述的核酸分子、权利要求 15 所述的载体、权利要求 16 或 17 所述的配体、或者权利要求 19 至 21 中任何一项所述的化合物, 用于治疗或诊断疾病。
30

23. 一种诊断患者疾病的方法，其特征在于：所述方法包括评估所述患者组织中编码权利要求 1 至 6 中任何一项所述的多肽的天然基因的表达水平或者评估权利要求 1 至 6 中任何一项所述的多肽的活性，并将所述表达水平或者活性与对照水平作比较，若该水平与所述对照水平不同，则表示患病。
24. 如权利要求 23 所述的方法，其特征在于：所述方法在体外进行。
25. 如权利要求 23 或 24 所述的方法，其特征在于：所述方法包括以下步骤：
(a) 在适合形成配体-多肽复合物的条件下，将权利要求 17 或 18 所述的配体与生物样品接触；以及(b)检测所述复合物。
26. 如权利要求 23 或 24 所述的方法，其特征在于：所述方法包括以下步骤：
(a) 在允许权利要求 7 至 14 中任何一项所述的核酸分子和核酸探针之间形成杂交复合物的严谨条件下，将患者的组织样品与该探针接触；
(b) 在步骤(a)使用的相同条件下，将对照样品与所述探针接触；以及
(c) 检测所述样品中是否存在杂交复合物，其中，检测到患者样品中杂交复合物的水平与对照样品中杂交复合物水平不同，表示患病。
27. 如权利要求 23 或 24 所述的方法，其特征在于：所述方法包括以下步骤：
(a) 在允许权利要求 7 至 14 中任何一项所述的核酸分子和核酸引物之间形成杂交复合物的严谨条件下，将患者组织的核酸样品与该引物接触；
(b) 在步骤(a)使用的相同条件下，将对照样品与所述引物接触；
(c) 扩增样品中的核酸；以及
(d) 检测患者和对照样品中扩增核酸的水平；其中，检测到患者样品的扩增核酸水平与对照样品的扩增核酸水平明显不同，表示患病。
28. 如权利要求 23 或 24 所述的方法，其特征在于：所述方法包括：
(a) 从有待测试疾病的患者中获得组织样品；
(b) 从所述组织样品中分离权利要求 7 至 14 中任何一项所述的核酸分子；以及
(c) 通过检测表示患病的核酸分子中与疾病相关的突变是否存在，诊断患者疾病。
29. 如权利要求 28 所述的方法，其特征在于：所述方法还包括扩增核酸分子以形成扩增产物，以及检测该扩增产物中是否存在突变。

30. 如权利要求 28 或 29 所述的方法，其特征在于：所述患者是否是否存在突变是这样检测的：将所述核酸分子与在严谨条件下和所述核酸分子杂交的探针接触，形成杂交双链分子，该杂交双链分子在对应于与疾病相关的突变的任何部分具有核酸探针链的未杂交部分；以及检测探针链的未杂交部分是否存在，表示与疾病相关的突变存在或不存。
- 5
31. 如权利要求 23 至 30 中任何一项所述的方法，其特征在于：所述疾病是细胞增殖性疾病、自身免疫/炎症疾病、心血管疾病、神经系统疾病、发育障碍、代谢性疾病、感染或其他病理病征。
32. 如权利要求 1 至 6 中任何一项所述的多肽作为分泌性蛋白质的应用。
- 10 33. 如权利要求 1 至 6 中任何一项所述的多肽作为胱氨酸，优选作为胱氨酸结折迭细胞因子，更好是作为胱氨酸结折迭细胞因子的 DAN 亚家族成员的应用。
34. 一种药物组合物，其特征在于：所述组合物含有权利要求 1 至 6 中任何一项所述的多肽、权利要求 7 至 14 中任何一项所述的核酸分子、权利要求 15 所述的载体、权利要求 17 或 18 所述的配体、或者权利要求 19 至 21 中任何一项所述的化合物。
- 15
35. 一种疫苗组合物，其特征在于：所述组合物含有权利要求 1 至 6 中任何一项所述的多肽或权利要求 7 至 14 中任何一项所述的核酸分子。
36. 如权利要求 1 至 6 中任何一项所述的多肽、权利要求 7 至 14 中任何一项所述的核酸分子、权利要求 15 所述的载体、权利要求 17 或 18 所述的配体、权利要求 19 至 21 中任何一项所述的化合物、或者权利要求 34 所述的药物组合物在制备治疗以下疾病的药物中的应用：细胞增殖性疾病、自身免疫/炎症疾病、心血管疾病、神经系统疾病、发育障碍、代谢性疾病、感染和其他病理病征。
- 20
37. 一种治疗患者疾病的方法，其特征在于：所述方法包括把权利要求 1 至 6 中任何一项所述的多肽、权利要求 7 至 14 中任何一项所述的核酸分子、权利要求 15 所述的载体、权利要求 17 或 18 所述的配体、或者权利要求 19 至 21 中任何一项所述的化合物、或者权利要求 34 所述的药物组合物给予该患者。
- 25
38. 如权利要求 37 所述的方法，其特征在于：当与健康受试者中多肽的天然
- 30

- 基因的表达水平或者活性相比，病患者的该表达水平或活性是较低的，给予该患者的多肽、核酸分子、载体、配体、化合物或组合物是激动剂。
39. 如权利要求 37 所述的方法，其特征在于：当与健康受试者中多肽的天然基因的表达水平或者活性相比，病患者的该表达水平或活性是较高的，给予该患者的多肽、核酸分子、载体、配体、化合物或组合物是拮抗剂。
- 5 40. 一种监测治疗患者疾病的方法，其特征在于：所述方法包括在一段时间内监测患者组织中权利要求 1 至 6 中任何一项所述的多肽的表达水平或活性、或权利要求 7 至 14 中任何一项所述的核酸分子的表达水平，若在该时间内所述表达水平或活性趋向于对照水平，表示该疾病获得缓解。
- 10 41. 一种鉴定有效地治疗和/或诊断疾病的化合物的方法，其特征在于：所述方法包括将权利要求 1 至 6 中任何一项所述的多肽、或者权利要求 7 至 14 中任何一项所述的核酸分子与怀疑对所述多肽或核酸分子有结合亲和力的一或多种化合物接触，选择能特异性地结合所述核酸分子或多肽的化合物。
- 15 42. 一种诊断疾病的试剂盒，其特征在于：所述试剂盒包括第一容器，它含有在严谨条件下与权利要求 7 至 14 中任何一项所述的核酸分子杂交的核酸探针；第二容器，它含有用来扩增该核酸分子的引物；以及该探针和引物的使用说明书，方便诊断疾病。
- 20 43. 如权利要求 42 所述的试剂盒，其特征在于：所述试剂盒还包括装有消化未杂交 RNA 的试剂的第三容器。
44. 一种试剂盒，其特征在于：所述试剂盒包括核酸分子阵列，其中至少一个核酸分子是权利要求 7 至 14 中任何一项所述的核酸分子。
45. 一种试剂盒，其特征在于：所述试剂盒包括一或多种与权利要求 1 至 6 中任何一项所述的多肽结合的抗体；一种用来检测所述抗体与所述多肽之间结合反应的试剂。
- 25 46. 一种转基因或基因剔除非人动物，其特征在于：所述动物被转化为表达权利要求 1 至 6 中任何一项所述的多肽的更高水平、更低水平或缺乏。
47. 一种筛选有效治疗疾病的化合物的方法，其特征在于：所述方法包括将权利要求 46 所述的非人转基因动物与候选化合物接触，并测定该化合物对动物疾病的影响。
- 30

胱氨酸结折迭蛋白质

5 技术领域

本发明涉及新型蛋白质(INSP002)，经本发明鉴定为一种分泌性蛋白质，它是胱氨酸结折迭细胞因子超家族的 Dan 家族的成员，以及涉及该蛋白质和来自编码基因的核酸序列在诊断、预防和治疗疾病中的应用。

本文所引用的出版物、专利和专利申请，均纳入其全文作为参考。

10

发明背景

目前，药物发现方法正蕴酿着一场根本革命，因为功能基因组学年代已经来临。术语“功能基因组学”应用于使用生物信息学工具将功能归因于感兴趣的蛋白质序列的方法。这些工具正日益显示其必要性，因为序列数据产生的速度远远高于研究实验室将功能划分给这些蛋白质序列的能力。

15

随着生物信息学工具的功效和准确性提高，这些工具正快速取代生物化学特性鉴定的常规技术。事实上，鉴定本发明所用的先进生物信息学工具现在能够输出可获得较高置信度的结果。

20

各所研究院和商业机构正在检查已有的序列数据，并且逐渐获得重大发现。然而，仍然有必要鉴定和特性分析其他基因和它们编码的多肽，作为研究和药物发现的目标。

分泌性蛋白质的背景

细胞制造和分泌胞外蛋白质的能力是许多生物过程的中心。通过分泌小泡与质膜融合，酶、生长因子、胞外基质蛋白和信号传导分子全部由细胞分泌出来。多数情况下，但不是所有情况，蛋白质被信号肽导入内质网，并进入分泌小泡。信号肽是顺式作用序列，影响多肽链从细胞质转运至膜结合室如分泌小泡。靶向分泌小泡的多肽或者分泌进入胞外基质，或者留在质膜内。留在质膜内的多肽有一或多个跨膜结构域。在细胞机能中发挥中心作用的分泌性蛋白质的例子是细胞因子、激素、胞外基质蛋白(粘附分子)、蛋白酶及

30

生长和分化因子。

生长因子代表相对较大的一组均有在体内外诱导细胞增殖的性质的多肽。生长因子有两方面与诸如胰岛素或生长激素等典型内分泌激素不同。首先，内分泌激素通常在专有腺体内合成(例如胰岛素在胰腺内合成)，而生长因子一般在多种类型的细胞和组织内合成。其次，典型内分泌激素在合成位点释放进入体液，被血流带到它们的靶组织。在多数情形下，生长因子的特点是它们在被合成的组织内局部地发挥作用(综述参见 Heath, JK. (1993) Growth Factors, Oxford University Press, Oxford, UK, pp. 15-33)。

虽然生长因子之间序列相似性的水平不高，但根据结构和功能相似性可将它们分成超家族。这些超家族的例子包括：(a)造血生长因子，例如生长激素、IL-2、IL-4、G-CSF 和 CNTF，它们全部都具有四螺旋束结构基序；(b) β -三叶家族成员，例如 IL-1 β 、IL-1 α 、FGF 和角化细胞生长因子；(c)EGF-样生长因子，例如 EGF 和 TGF α ，它们全部都有免疫球蛋白样结构域；和(d)胱氨酸结生长因子折迭，包括 NGF、TGF β 、PDGF 和糖蛋白激素。

生长因子位于细胞外，为了发挥生物作用，它们与位于靶细胞质膜上的特异性高亲和力受体相互作用。很多不同生长因子受体的分子特性揭示，它们属于限定家族：酪氨酸激酶受体、G-蛋白相关的 7 次跨膜受体，和丝氨酸/苏氨酸激酶受体。酪氨酸激酶受体的特点在于具有酪氨酸激酶活性的胞外结构域、跨膜结构域和胞内结构域。丝氨酸/苏氨酸激酶生长因子受体与带有胞外结构域、跨膜结构域和胞内结构域的酪氨酸激酶受体相似。胞内结构域具有内在丝氨酸/苏氨酸激酶活性。

生长因子的反调节暗示各种疾病状态，包括但不限于：肿瘤疾病(Bartucci M 等, (2001) Cancer Res. Sep 15; 61(18): 6747-54, Dias S 等, (2001) Proc Natl Acad Sci U S A. Sep 11; 98(19): 10857-62, Djavan B 等, (2001) World J Urol. 19(4): 225-33)、炎症疾病(Fiocchi C. (2001) J Clin Invest. Aug; 108(4): 523-6, Hodge S 等, (2001) Respiriology. Sep; 6(3): 205-211, Fenwick SA 等, (2001) J Anat. Sep; 199(Pt 3): 231-40)、神经系统疾病(Cooper JD 等, (2001) Proc Natl Acad Sci U S A 98(18): 10439-44, Fahnestock M 等, (2001) Mol Cell Neurosci 18(2): 210-20)和代谢疾病(Vickers MH 等, (2001) Endocrinology. 142(9): 3964-73)。

胱氨酸结折迭超家族

胱氨酸结超家族中看到的典型结构基于 6 个胱氨酸的存在，建立 3 个二硫键。其中两个二硫键建立“环状”结构，第 3 个二硫键贯穿其中(Sun 等, 1995)。通常发现胱氨酸结结构域有 6 个以上胱氨酸残基。附加胱氨酸残基
5 正常情况下用于二聚作用过程中建立胱氨酸结结构域内其他二硫键或链间二硫键。

将该胱氨酸结超家族分为亚家族，包括糖蛋白激素(例如卵泡刺激素)、转化生长因子 β (TGF β)蛋白(例如骨形态生成蛋白 4)、血小板分泌的生长因子样(PDGF-样)蛋白(例如血小板分泌的生长因子 A)、神经生长因子(NGF)(例
10 如脑分泌的神经营养因子)和成神经细胞瘤中差异筛选选择的基因畸变(DAN)家族(例如 cerberus)。DAN 亚家族包括 Cer1、Cerberus、Caronte、Drm/Gremlin、PRDC、DAN、Dante 和 CeCan1 (Massague 等, Genes Dev 2000 Mar 15; 14(6): 627-44; Massague & Wotton, EMBO J. 2000 Apr 17; 19(8): 1745-54)。

一般认为，DAN 亚家族成员能够调节蛋白质 TGF β 亚家族成员的作用
15 (Pearce 等, Dev Biol. 1999 May 1; 209(1): 98-110)。更具体地说，DAN 亚家族成员在发育期间能够调节骨形态生成蛋白(BMP)的作用是不可能的。

业已发现，DAN 亚家族成员是骨形态生成蛋白(BMP)的拮抗剂，而骨形态生成蛋白是胱氨酸结超家族的 TGF β 亚家族成员(Stanley 等, Mech Dev. 1998 Oct; 77(2): 173-84; Massague 等, 2000(见上); Massague J & Wotton D, 2002(见
20 上)。BMP 单体在与细胞表面受体相互作用之前，通过它们的胱氨酸结结构域结合而发生同源或异源二聚化。一般认为，DAN 亚家族成员通过它们各自的胱氨酸结结构域能够与 BMP 结合。这防止 BMP 与它的天然二聚配偶子结合，结果是 BMP 不能再与它的细胞表面信号传导受体相互作用。特别针对 DAN、Cer1 和 DRM 进行的实验已经显示，它们抑制 BMP4 的作用(Pearce
25 等, 1999, (见上))。

因为结合研究的结果，使得对 cerberus 功能的理解加深(Piccolo S.等, Nature. 1999 Feb 25; 397(6721): 707-10)。首次对 cerberus 进行的功能研究使用了非洲爪蟾 cerberus 蛋白(cer)。将非洲爪蟾 cerberus mRNA 微量注入非洲爪蟾胚胎中，结果发现 cer 蛋白诱导 Spemann's 组织者的前内胚层中形成
30 异位头(ectopic head) (Bouwmeester 等, Nature 1996 Aug 15; 382(6592): 595-

601; Bouwmeester T., Int J Dev Biol. 200, 145(1 Spec No): 251-8)。由 Piccolo 及合作者进行的结合研究发现, 非洲爪蟾的 cerberus 蛋白通过独立位置结合和抑制 Nodal、BMP 和 Wnt 蛋白质。更具体地说, 他们发现 cerberus 对 Xnr-1(Nodal 家族成员)、BMP4 (BMP 家族成员)和 Xwmt-8 (Wnt 家族成员)有高的特异亲和力和抑制作用。所以, 这一研究将 cerberus 与 DAN 家族的其他成员和发育和组织分化通道联系起来。

由基因 SOST 编码的壳硬蛋白也是 DAN 亚家族的成员(Brumkow 等, 2001, Am. J. Hum. Genet. 68: 577-589)。SOST 与硬化性狭窄有关, 这是常染色体隐性硬化性骨发育不良。硬化性狭窄相关的表型是渐进性骨骼生长过度, 可导致巨人症、面容扭曲及第 7 和第 8 根颅神经受压(Brumkow 等, 2001, (见上))。硬化性狭窄与 SOST 之间的联系通过受疾病感染的家族的纯合子性图谱来确定。Brumkow 及合作者鉴定了硬化性狭窄相关的表型与其他 DAN 亚家族成员相关的作用相类似。由于丧失 TGF β 亚家族成员的负调节因子, 更具体是丧失 BMP 会出现产生硬化性狭窄, 这一启示也使上述联系得到支持。

所以, 要加深理解导致上述疾病状态和相关疾病状态的隐伏通道, 以及研制治疗这些疾病的有效基因或药物疗法, 鉴定分泌性蛋白质, 特别是生长因子, 例如胱氨酸结折迭超家族成员, 尤其是 DAN 亚家族成员具有极为重要的意义。

发明内容

本发明基于如下发现: INSP002 蛋白质起分泌性蛋白质的作用, 此外, 也起胱氨酸结折迭细胞因子超家族的 DAN 亚家族成员的作用。

本发明第一方面提供一种多肽, 该多肽:

- (i) 包括 SEQ ID NO: 2、SEQ ID NO: 4 或 SEQ ID NO: 7 所示的氨基酸序列;
- (ii) 是其具有分泌性蛋白质的功能, 优选具有胱氨酸结折迭细胞因子超家族成员的功能, 更好具有 DAN 亚家族成员的功能, 或者具有与(i)的多肽相同的抗原决定簇的片段; 或者
- (iii) 是(i)或(ii)的功能等同物。

本发明第一方面的多肽优选：

- (i) 包括 SEQ ID NO: 6 或 SEQ ID NO: 8 所示的氨基酸序列；
- (ii) 是其具有分泌性蛋白质的功能，优选具有胱氨酸结折迭细胞因子超家族成员的功能，更好具有 DAN 亚家族成员的功能，或者具有与(i)的多肽
5 相同的抗原决定簇的片段；或者
- (iii) 是(i)或(ii)的功能等同物。

本发明第一方面的另一实施例提供一种多肽，该多肽：

- (i) 由 SEQ ID NO: 2、SEQ ID NO: 4、SEQ ID NO: 6、SEQ ID NO: 7、
或 SEQ ID NO: 8 所示的氨基酸序列组成；
- 10 (ii) 是其具有分泌性蛋白质的功能，优选胱氨酸结折迭细胞因子超家族成员的功能，更好是 DAN 亚家族成员的功能，或者具有与(i)的多肽相同的抗原决定簇的片段；或者
- (iii) 是(i)或(ii)的功能等同物。

下文将具有 SEQ ID NO: 2 所示的序列的多肽称为“INSP002 外显子 1
15 多肽”。将具有 SEQ ID NO: 4 所示的序列的多肽称为“INSP002 外显子 2 多肽”。将具有 SEQ ID NO: 6 所示的序列的多肽称为“INSP002 多肽”。INSP002 外显子 1 多肽的首 22 个氨基酸是信号肽，没有信号序列的 INSP002 多肽序列如 SEQ ID NO: 7 和 SEQ ID NO: 8 所示。下文将具有 SEQ ID NO: 7 所示的序列的多肽称为“没有信号肽的 INSP002 外显子 1 多肽”。将具有 SEQ ID
20 NO: 8 所示的序列的多肽称为“没有信号肽的 INSP002 多肽”。

本发明第一方面的另一实施例提供一种多肽，该多肽：

- (i) 包括或由 SEQ ID NO: 14 所示的氨基酸序列组成；
- (ii) 是其具有分泌性蛋白质的功能，优选胱氨酸结折迭细胞因子超家族成员的功能，更好是 DAN 亚家族成员的功能，或者具有与(i)的多肽相同的
25 抗原决定簇的片段；或者
- (iii) 是(i)或(ii)的功能等同物。

具有 SEQ ID NO: 14 所示的序列的多肽是 INSP002 多肽的变体。它与 INSP002 多肽相同，但与 INSP002 多肽相比，它在 107 和 108 位含有两个氨基酸缺失，在 110 位含有一个氨基酸取代。下文将具有 SEQ ID NO: 14 所
30 示的序列的多肽称为“变体 INSP002 多肽”。

本发明第一方面的多肽优选起胱氨酸结折迭细胞因子超家族成员的作用，更好是起 DAN 亚家族成员的作用。术语“胱氨酸结折迭细胞因子”已为本领域所熟悉，应用本领域公知的各种检测法之一，本领域技术人员很容易能够确定一个多肽是否起胱氨酸结折迭细胞因子超家族的成员的作用。

5 具体地说，通过测定它是否是 TGF β 超家族成员的拮抗剂，尤其它是否是 BMP 的拮抗剂，本领域技术人员就能够确定一个多肽是否起 DAN 亚家族的成员的作用。为了测定多肽是否起 BMP 拮抗剂的作用，可以非洲爪蟾胚胎作为一个系统，因为多个 BMP 在非洲爪蟾胚胎中表达(Chang C.等, 1999, Development 126: 3347-3357, Hawley S.等, 1995, Genes Dev. 9: 2923-2935, Hemmati-Brivanlou, A., 和 G. H. Thomsen. 1995, Dev. Genet. 17: 78-89, Jones C. M. 等, 1992, Development 115: 639-647)。BMP-2/4 类别或 BMP-7 类别信号在早期中胚层中过度表达会招致腹部致命，而这些信号的抑制剂(例如 Noggin、Xnr3、Chordin 或 Follistatin)会招致背部致命。所以，利用多肽对胚胎发育的影响可确定该多肽是否是一种 BMP 拮抗剂。

15 本文所用的术语“INSP002 多肽”包括包括 INSP002 外显子 1 多肽、没有信号肽的 INSP002 外显子 1 多肽、INSP002 外显子 2 多肽、INSP002 多肽或者没有信号肽的 INSP002 多肽的多肽，或者由 INSP002 外显子 1 多肽、没有信号肽的 INSP002 外显子 1 多肽、INSP002 外显子 2 多肽、INSP002 多肽、没有信号肽的 INSP002 多肽或者变体 INSP002 多肽组成的多肽。

20 本发明第二方面提供一种编码本发明第一方面的多肽的纯化核酸分子。

纯化核酸分子最好包括 SEQ ID NO: 1(编码 INSP002 外显子 1 多肽)、SEQ ID NO: 3(编码 INSP002 外显子 2 多肽)、SEQ ID NO: 5(编码 INSP002 多肽)、或者 SEQ ID NO: 13(编码变体 INSP002 多肽)所示的核酸序列，或者是这些序列之一的一个冗余等同物或片段。

25 本发明还提供纯化核酸分子，其由 SEQ ID NO: 1(编码 INSP002 外显子 1 多肽)、SEQ ID NO: 3(编码 INSP002 外显子 2 多肽)、SEQ ID NO: 5(编码 INSP002 多肽)、或者 SEQ ID NO: 13(编码变体 INSP002 多肽)所示的核酸序列组成，或者是这些序列之一的一个冗余等同物或片段。

30 根据本发明该方面的一实施例，纯化核酸分子不含有位于编码 INSP002 外显子 1 多肽的核酸序列和编码 INSP002 多肽的核酸序列(SEQ ID NO: 1 和

SEQ ID NO: 5 的核苷酸 1 至 151)上游的 5'非翻译区。根据此实施例, 纯化核酸分子最好包括 SEQ ID NO: 1 的核苷酸 152 至 475 或者 SEQ ID NO: 5 的核苷酸 152 至 721。本发明还提供一种纯化核酸分子, 其由 SEQ ID NO: 1 的核苷酸 152 至 475 或者 SEQ ID NO: 5 的核苷酸 152 至 721 组成。编码不含有 5'非翻译区的 INSP002 多肽的核苷酸序列(SEQ ID NO: 5 的核苷酸 152 至 721)如 SEQ ID NO: 11 所示, 编码不含有 5'非翻译区的 INSP002 外显子 1 多肽的核苷酸序列(SEQ ID NO: 1 的核苷酸 152 至 475)如 SEQ ID NO: 12 所示。

根据本发明该方面的另一实施例, 纯化核酸分子不编码位于 INSP002 外显子 1 多肽和 INSP002 多肽起点的信号肽(SEQ ID NO: 1 和 SEQ ID NO: 5 的核苷酸 152 至 217)。根据此实施例, 纯化核酸分子最好包括 SEQ ID NO: 1 的核苷酸 218 至 475(编码没有信号肽的 INSP002 外显子 1 多肽)或者 SEQ ID NO: 5 的核苷酸 218 至 721(编码没有信号肽的 INSP002 多肽)。本发明还提供一种纯化核酸分子, 其由 SEQ ID NO: 1 的核苷酸 218 至 475(编码没有信号肽的 INSP002 外显子 1 多肽)或者 SEQ ID NO: 5 的核苷酸 218 至 721(编码没有信号肽的 INSP002 多肽)组成。编码成熟 INSP002 多肽的核苷酸序列(SEQ ID NO: 7)如 SEQ ID NO: 9 所示, 编码成熟 INSP002 外显子 1 多肽的核苷酸序列如 SEQ ID NO: 10 所示。

根据本发明该方面的另一实施例, 纯化核酸分子不含有位于编码变体 INSP002 多肽的核酸序列(SEQ ID NO: 13 的核苷酸 1 至 68)上游的 5'非翻译区。根据此实施例, 纯化核酸分子最好包括或由 SEQ ID NO: 13 的核苷酸 69 至 719 组成。编码不含有 5'非翻译区的变体 INSP002 多肽的核苷酸序列(SEQ ID NO: 13 的核苷酸 69 至 719)如 SEQ ID NO: 15 所示。

本发明第三方面提供一种在高度严谨条件下与本发明第二方面的核酸分子杂交的纯化核酸分子。

本发明第四方面提供一种含有本发明第二或第三方面的核酸分子的载体, 例如表达载体。

本发明第五方面提供一种用本发明第四方面的载体转化的宿主细胞。

本发明第六方面提供一种配体, 其特异性地结合, 优选抑制本发明第一方面的多肽的胱氨酸结折迭细胞因子活性。优选地, 配体抑制本发明第一方

面的多肽的功能，所述多肽是胱氨酸结折迭细胞因子的 DAN 亚家族成员。本发明多肽的配体可以有多种形式，包括天然或修饰的基质、酶、受体、有机小分子(例如最大为 2000Da，优选 800Da 或以下的天然或合成有机小分子)、多肽类似物、无机分子、肽、多肽、抗体、上述物质的结构或功能类似物。本发明第七方面提供一种化合物，它能有效地改变编码本发明第一方面的多肽的天然基因的表达，或者能调节本发明第一方面的多肽的活性。

本发明第七方面的化合物可增加(激动)或者减少(拮抗)多肽的基因表达水平或活性。重要的是，对 INSP002 多肽功能的鉴定允许设计能够鉴定有效地治疗和/或诊断疾病的化合物的筛选方法。

10 本发明第八方面提供本发明第一方面的多肽，或者本发明第二或第三方面的核酸分子，或者本发明第四方面的载体，或者本发明第六方面的配体，或者本发明第七方面的化合物在治疗或诊断中的应用。这些分子还可用来制造治疗细胞增殖性疾病、自身免疫/炎症疾病、心血管疾病、神经系统疾病、发育障碍、代谢性疾病、感染和其他病理病征的药物。

15 本发明第九方面提供一种诊断患者疾病的方法，该方法包括评估所述患者组织中编码本发明第一方面的多肽的天然基因的表达水平或者本发明第一方面的多肽的活性，并将所述表达水平或者活性与对照水平作比较，若该水平与所述对照水平不同，则表示患病。这种方法最好在体外进行。可用类似方法监测对患者疾病的治疗，其中多肽或核酸分子的表达水平或活性在一段时间内趋向于对照水平，表示该疾病获得缓解。

20 本发明第九方面和第十方面的障碍或疾病优选涉及胱氨酸结折迭细胞因子，更好是涉及 DAN 亚家族成员的异常水平。这些障碍或疾病也可以涉及胱氨酸结折迭细胞因子，更好是涉及 DAN 亚家族成员的配体的异常水平。例如，这些障碍或疾病可以涉及 TGF β 超家族成员的异常水平。具体地说，25 这些障碍或疾病可以涉及 BMP，例如神经病变、肾病，如糖尿病性神经病变，癌症、伤口愈合、纤维变性、骨质减少、骨质疏松、骨折和硬化性狭窄。

检测本发明第一方面的多肽的优选方法包括以下步骤：(a) 在适合形成配体-多肽复合物的条件下，将本发明第六方面的一种配体，例如抗体与生物样品接触；以及(b)检测所述复合物。

30 本领域的读者将懂得，本发明第九方面的方法有很多种，例如，核酸与

短探针杂交法、点突变分析法、聚合酶链式反应(PCR)扩增法以及使用抗体检测异常蛋白水平的方法。类似方法的使用可以是短期或长期，以治疗患者被监测的疾病。本发明还提供一些用于这些方法中诊断疾病的试剂盒。

本发明第十方面提供本发明第一方面的多肽作为分泌性蛋白质的应用。

- 5 优选地，本发明提供本发明第一方面的多肽作为胱氨酸，更好是作为胱氨酸结折迭细胞因子，特别是作为胱氨酸结折迭细胞因子的 DAN 亚家族成员的应用。

- 10 本发明第十一方面提供一种药物组合物，该组合物含有本发明第一方面的多肽，或者本发明第二或第三方面的核酸分子，或者本发明第四方面的载体，或者本发明第五方面的宿主细胞，或者本发明第六方面的配体，或者本发明第七方面的化合物以及药学上可接受的载体。

- 15 本发明第十二方面提供本发明第一方面的多肽，或者本发明第二或第三方面的核酸分子，或者本发明第四方面的载体，或者本发明第五方面的宿主细胞，或者本发明第六方面的配体，或者本发明第七方面的化合物在制造诊断或治疗疾病的药物中的应用。

本发明第十三方面提供一种治疗患者疾病的方法，该方法包括把本发明第一方面的多肽，或者本发明第二或第三方面的核酸分子，或者本发明第四方面的载体，或者本发明第六方面的配体，或者本发明第七方面的化合物给予该患者。

- 20 本发明第十二方面和第十三方面的疾病优选涉及胱氨酸结折迭细胞因子，更好是涉及 DAN 亚家族成员的异常水平。这些疾病也可以涉及胱氨酸结折迭细胞因子，更好是涉及 DAN 亚家族成员的配体的异常水平。例如，这些疾病可以涉及 TGF β 超家族成员的异常水平。具体地说，这些障碍或疾病可以涉及 BMP，例如神经病变、肾病，如糖尿病性神经病变，癌症、伤口愈合、纤维变性、骨质减少、骨质疏松、骨折和硬化性狭窄。
- 25

- 30 当与健康受试者中编码本发明第一方面的多肽的天然基因的表达水平或者本发明第一方面的多肽的活性相比，病患者的该表达水平或活性较低的，给予该患者的多肽、核酸分子、配体或化合物应该是激动剂。相反，当与健康受试者中所述多肽的天然基因的表达水平或者活性相比，病患者的该表达水平或活性较高的，给予该患者的多肽、核酸分子、配体或化合物应该是拮

抗剂。拮抗剂的例子包括反义核酸分子、核酶和配体，例如抗体。

本发明第十四方面提供转基因或剔除非人动物，它们被转化为表达本发明第一方面的多肽的更高水平、更低水平或缺乏。这些转基因动物是研究疾病极为有用的模型，也可用在鉴定有效治疗或诊断该疾病的化合物的筛选方案中。

以下，给出为实施本发明而可采用的标准技术和步骤的概要。应明白，本发明并不限于所述的这些具体方法、方案、细胞系、载体和试剂。还应明白，本文所用的术语目的仅在于描述具体实施例，该术语不应该被看成为限定本发明的范围。本发明的范围只由所附权利要求书的术语限定。

本说明书中核苷酸和氨基酸使用标准缩写。

除非另有指出，本发明的做法是采用分子生物学、微生物学、重组 DNA 技术和免疫学的常规技术，它们都已为本领域技术人员所掌握。

这些技术在有关文献中已有详细解释。例如，可参考以下特别适用的文献：Sambrook Molecular Cloning; A Laboratory Manual, Second Edition (1989); DNA Cloning, Volumes I and II (D.N Glover ed. 1985); Oligonucleotide Synthesis (M.J. Gait ed. 1984); Nucleic Acid Hybridization (B.D. Hames & S.J. Higgins eds. 1984); Transcription and Translation (B.D. Hames & S.J. Higgins eds. 1984); Animal Cell Culture (R.I. Freshney ed. 1986); Immobilized Cells and Enzymes (IRL Press, 1986); B. Perbal, A Practical Guide to Molecular Cloning (1984); the Methods in Enzymology series (Academic Press, Inc.), especially volumes 154 & 155; Gene Transfer Vectors for Mammalian Cells (J.H. Miller and M.P. Calos eds. 1987, Cold Spring Harbor Laboratory); Immunochemical Methods in Cell and Molecular Biology (Mayer and Walker, eds. 1987, Academic Press, London); Scopes, (1987) Protein Purification: Principles and Practice, Second Edition (Springer Verlag, N.Y.); Handbook of Experimental Immunology, Volumes I-IV (D.M. Weir and C. C. Blackwell eds. 1986)。

本文所用的术语“多肽”包括含有相互间以肽键或修饰肽键连接的两个或以上氨基酸的任何肽或蛋白质，即肽同构物(isostere)。该术语指短链(肽和寡肽)和长链(蛋白质)。

本发明的多肽可以是成熟蛋白质形式，或者可以是前蛋白质(pre-, pro-or

prepro-protein), 其可被前部分(pre-, pro-or prepro-portion)切割激活而产生生活性成熟多肽。这些多肽中的前序列(pre-, pro-or prepro-sequence)可以是先导序列或者分泌序列或者是用来纯化成熟多肽序列的序列。

5 本发明第一方面的多肽可形成融合蛋白的一部分。例如, 它一般适宜包括一或多个附加氨基酸序列, 这些附加氨基酸序列可含有分泌序列或先导序列、前序列(pro-sequences)、有助纯化的序列、或者例如在重组生产中赋予蛋白质更高稳定性的序列。另一个选择或者另一方面是, 成熟多肽可与另一种化合物, 例如与延长该多肽半衰期的化合物(譬如聚乙二醇)融合。

10 多肽可含有由天然过程, 例如翻译后加工, 或者由本领域公知的化学修饰技术修饰的氨基酸, 它们不同于 20 个基因编码的氨基酸。在已知的修饰中, 本发明的多肽通常带有的修饰是糖基化、脂质附着、硫酸化、例如谷氨酸残基的 γ -羧酸化, 羟基化和 ADP-核糖基化。其他可能的修饰包括乙酰化、酰化、酰胺化、黄素的共价附着、血素分子的共价附着、核苷酸或核苷酸衍生物的共价附着、脂质衍生物的共价附着、磷脂酰肌醇的共价附着、交联、15 环化二硫键形成、脱甲基、共价交联形成、半胱氨酸形成、焦谷氨酸盐形成、甲酰化、GPI 锚着形成、碘化、甲基化、豆蔻酰化、氧化、蛋白水解加工、磷酸化、异戊二烯化、外消旋化、硒酰化、转移-RNA 介导的蛋白质中插入氨基酸例如精氨酰化, 以及遍在蛋白化。

20 修饰可发生于多肽的任何位置, 包括肽骨架、氨基酸侧链以及氨基或羧基末端。事实上, 天然存在的肽和合成肽一般通过共价修饰封闭肽的氨基或羧基末端, 或者封闭二者, 这样的修饰存在于本发明的多肽中。

25 发生在多肽内的修饰通常随如何制成该多肽而变化。对于重组制成的多肽, 通过特定宿主细胞翻译后修饰能力以及存在于所述多肽的氨基序列中的修饰信号来确定大部分修饰的性质和程度。例如, 不同类型的宿主细胞之间糖基化型式是不同的。

本发明的多肽可以任何合适的方式制成。这些多肽包括分离的天然存在的多肽(例如自细胞培养基中纯化而来)、重组产生的多肽(包括融合蛋白)、合成产生的多肽或者通过这些方法组合所产生的多肽。

30 本发明第一方面的功能等同多肽可以是与 INSP002 多肽同源的多肽。如果一个多肽的序列与另一个多肽的序列具有足够高程度的同一性或相似性,

本文就用术语“同源的”来形容该两个多肽。“同一性”表示在比对序列的任何特定位置上，两个序列之间的氨基酸残基是相同的。“相似性”表示在比对序列的任何特定位置上，两个序列的氨基酸残基是相似的。同一性和相似性的程度不难计算(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 Analysis of Sequence Data, Part 1, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; Sequence Analysis in Molecular Biology, von Heinje, G., Academic Press, 1987; Sequence Analysis Primer, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991)。

所以，同源多肽包括 INSP002 多肽的天然生物变体(例如，产生多肽的物种内等位基因变体或地理变体)和突变体(例如，含有氨基酸取代、插入或缺失的突变体)。这些突变体可包括其一或多个氨基酸残基被保守性或非保守性氨基酸残基(优选保守性氨基酸残基)取代的多肽，这样一个取代的氨基酸残基可以或不必要是遗传密码编码的残基。典型的取代是在 Ala、Val、Leu 和 Ile 之间；在 Ser 和 Thr 之间；在酸性残基 Asp 和 Glu 之间；在 Asn 和 Gln 之间；在碱性残基 Lys 和 Arg 之间；或者在芳香族残基 Phe 和 Tyr 之间。特别优选的是这样一些变体：有多个氨基酸，即 5 至 10 个氨基酸，1 至 5 个氨基酸，1 至 3 个氨基酸，1 至 2 个氨基酸，或者仅有 1 个氨基酸以任意组合取代、缺失或插入的变体。特别优选的是不改变该蛋白质的性质和活性的沉默取代、插入和缺失。就此而言，特别优选的还有保守性取代。这些突变体也包括其中一或多个氨基酸残基包括一取代基团的多肽。

通常，两个多肽之间的同一性大于 30%，就认为它们在功能上是等同的。本发明第一方面的功能等同多肽与 INSP002 多肽，或其活性片段的序列同一性优选大于 35%。更好的多肽分别具有大于 35%、40%、50%、60%、70%、80%、90%、95%、98%、99%或以上的同一性。

本发明第一方面的功能等同多肽还可以是已用一或多种结构比对技术鉴定的多肽。例如，可以应用 Inpharmatica Genome Threader 技术(它构成用于产生 Biopendium 检索数据库的检索工具的其中一部分，参见共同待审批英国专利申请 PCT/GB01/01105)来鉴定现时具有未知功能的多肽，这些多肽虽

然与 INSP002 多肽具有较低的序列同一性，但由于与 INSP002 多肽序列共享较高水平的结构同源性，故可预计它们有分泌性分子活性。“较高水平的结构同源性”指用 Inpharmatica Genome Threader 预测两个蛋白质确定共享 10%和以上的结构同源性。

- 5 本发明第一方面的多肽还包括 INSP002 多肽的片段、INSP002 多肽的功能等同物的片段，只要那些片段保留了胱氨酸结折迭细胞因子的活性，优选 DAN 胱氨酸结折迭亚家族成员的活性，或者带有与 INSP002 多肽相同的抗原性决定簇。

10 本文所用的术语“片段”指氨基酸序列与 INSP002 多肽或它的其中一种功能等同物的部分而非全部氨基酸序列相同的多肽。这些片段应包括所述序列的至少 n 个保守性氨基酸，取决于特定的序列，n 最好是 7 或以上(例如，8、10、12、14、16、18、20 或以上)。小片段可形成抗原性决定簇。

15 这些片段可以是“独立的(free-standing)”，即不是其他氨基酸或多肽的一部分，也不与其他氨基酸或多肽融合，或者它们可包含在它们构成其中一部分或区域的较大多肽之内。当包含在较大多肽之内时，本发明的片段最好构成单个连续区域。例如，一些优选实施例涉及一个片段，其具有与该片段的氨基末端融合的前多肽区域(pre- and/or pro- polypeptide region)，和/或具有与该片段的羧基末端融合的附加区域。然而，单一个较大多肽内可含有多个片段。

20 本发明的多肽或其免疫原性片段(包含至少一个抗原性决定簇)可被应用来产生对这些多肽有免疫特异性的配体，例如多克隆或单克隆抗体。这些抗体可用于分离或鉴定表达本发明的多肽的克隆，或者用于亲和力层析法纯化这些多肽。抗体还可用来帮助诊断或治疗，以及其他应用，这对本发领域技术人员是显而易见的。

25 术语“免疫特异性”指抗体对本发明的多肽的亲和力远远大于它们对现有领域其他相关多肽的亲和力。本文所用的术语“抗体”指能够与所述抗原性决定簇结合的完整分子及其片段，例如 Fab、F(ab')₂ 和 Fv。所以，这些抗体与本发明第一方面的多肽结合。

30 术语“明显较高的亲和力”指与对已知的分泌性蛋白质的亲和力相比，对本发明的多肽的亲和力有可测量的提高。

优选地，对本发明的多肽的亲合力比已知的分泌性蛋白质，例如胱氨酸结折迭细胞因子，特别是 DAN 亚家族成员至少高 1.5 倍、2 倍、5 倍、10 倍、100 倍、 10^3 倍、 10^4 倍、 10^5 倍或 10^6 倍。

如果需要多克隆抗体，可用本发明第一方面的多肽使一种选定哺乳动物
5 免疫，例如小鼠、兔、山羊或马。用来免疫动物的多肽可通过重组 DNA 技术产生，或者可通过化学方法合成。有需要的时候，可将多肽与一载体蛋白质连接。通过化学方法与多肽偶联的常用载体包括牛血清白蛋白、甲状腺球蛋白和钥孔血蓝蛋白。然后，用偶联的多肽来免疫动物。采集该免疫动物的血清，再按公知的程序，例如免疫亲和力层析法加以处理。

10 本领域技术人员可轻易地生产针对本发明第一方面的多肽的单克隆抗体。应用杂交瘤技术生产单克隆抗体的一般方法已为人所公知(例如，参见 Kohler, G.和 Milstein, C., *Nature* 256: 495-497 (1975); Kozbor 等, *Immunology Today* 4: 72 (1983); Cole 等, 77-96 in *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc. (1985))。

15 可针对各种性质，即同种型、表位、亲和力等，筛选对本发明第一方面的多肽产生的多系列单克隆抗体。单克隆抗体特别用于纯化它们所针对的各种多肽。另一方面，编码感兴趣的单克隆抗体的基因可通过例如本领域公知的 PCR 技术在杂交瘤中分离出来，并可在适当载体中克隆和表达。

还可使用嵌合抗体，其中非人的可变区与人稳定区连接或融合(例如，
20 参见 Liu 等, *Proc. Natl. Acad. Sci. USA*, 84, 3439 (1987))。

可对抗体进行修饰，例如通过使它人源化，以便在某一个体内的免疫原性减弱(参见 ones 等, *Nature*, 321, 522 (1986); Verhoeyen 等, *Science*, 239, 1534 (1988); Kabat 等, *J. Immunol.*, 147, 1709 (1991); Queen 等, *Proc. Natl Acad. Sci. USA*, 86, 10029 (1989); Gorman 等, *Proc. Natl Acad. Sci. USA*,
25 88, 34181 (1991); 以及 Hodgson 等, *Bio/Technology*, 9, 421 (1991))。本文所用的术语“人源化抗体”指非人供体抗体的重链和/或轻链可变区内 CDR 氨基酸和选定的其他氨基酸已经被人抗体的等同氨基酸所取代的抗体分子。所以，人源化抗体与人抗体极为相近，但具有与该供体抗体结合的能力。

另一个选择是，抗体可以是“双特异性”抗体，即它是一种有两个不同
30 抗原结合区的抗体，每一结合区针对不同表位。

可应用噬菌体展示技术选择基因，所述基因编码具有与本发明多肽结合的活性的抗体，多肽来自为具有相关抗体而筛选的以 PCR 技术扩增的人淋巴细胞 V 基因库，或者天然文库(McCafferty, J.等, (1990), Nature 348, 552-554; Marks, J. 等, (1992) Biotechnology 10, 779-783)。这些抗体的亲和力也可通过链改组加以改善(Clackson, T.等, (1991) Nature 352, 624-628)。

以上述技术产生的多克隆抗体或单克隆抗体有额外应用，因为它们可用作免疫测定法、放射性免疫测定法(RIA)或者酶联免疫吸附测定法(ELISA)中的试剂。这些应用可使用可分析检测到的试剂，例如放射性同位素、荧光分子或酶来标记抗体。

10 本发明第二和第三方面的优选核酸分子是编码 SEQ ID NO: 2、SEQ ID NO: 4、SEQ ID NO: 7、SEQ ID NO: 6、SEQ ID NO: 8 和 SEQ ID NO: 14 所示的多肽序列和功能上等同的多肽的那些分子。这些核酸分子可用于本文描述的方法和应用中。本发明的核酸分子最好含有至少 n 个本文所公开的序列中的连续核苷酸，取决于特定的序列，n 是 10 或以上(例如，12、14、15、15 18、20、25、30、35、40 或以上)。

本发明的核酸分子还包括上述核酸分子的补体序列(例如，用于反义或探测目的)。

20 本发明的核酸分子可以是 RNA 形式，如 mRNA，或者是 DNA 形式，包括例如 cDNA、合成 DNA 或基因组 DNA。这样一些核酸分子可通过克隆、化学合成技术或其组合获得。例如，核酸分子可通过应用诸如固相亚磷酰胺化学合成从基因组或 cDNA 文库化学合成，或者从生物体中分离而制备。一般通过在体外或体内转录 DNA 序列产生 RNA 分子。

核酸分子可以是双链或单链。单链 DNA 可以是编码链，也称有义链，或者它可以是非编码链，也称反义链。

25 术语“核酸分子”还包括 DNA 和 RNA 类似物，例如含有修饰骨架的那些类似物，以及肽核酸(PNA)。本文所用的术语“PNA”指反义分子或反基因试剂，其含有长至少 5 个核苷酸的寡核苷酸，所述寡核苷酸与最好位于赖氨酸末端的氨基酸残基肽骨架连接。该末端赖氨酸使组合物具有溶解性。PNA 可被聚乙二醇化，延长它们在细胞内的停留时间，在细胞内它们优先与 30 补体单链 DNA 和 RNA 结合，并终止转录延伸(Nielsen, P.E.等, (1993)

Anticancer Drug Des. 8: 53-63)。

编码 SEQ ID NO: 2 所示的多肽的核酸分子可与 SEQ ID NO: 1 所示的核酸分子的编码序列(核苷酸 152 至 475)相同, 如 SEQ ID NO: 12 所示。编码 SEQ ID NO: 7 所示的多肽的核酸分子可与 SEQ ID NO: 1 所示的核酸分子的编码序列(核苷酸 218 至 475)相同, 如 SEQ ID NO: 10 所示。编码 SEQ ID NO: 4 所示的多肽的核酸分子可与 SEQ ID NO: 3 所示的核酸分子的编码序列相同。编码 SEQ ID NO: 6 所示的多肽的核酸分子可与 SEQ ID NO: 5 所示的核酸分子的编码序列(核苷酸 152 至 721)相同, 如 SEQ ID NO: 11 所示。编码 SEQ ID NO: 8 所示的多肽的核酸分子可与 SEQ ID NO: 5 所示的核酸分子的编码序列(核苷酸 218 至 721)相同, 如 SEQ ID NO: 9 所示。编码 SEQ ID NO: 14 所示的多肽的核酸分子可与 SEQ ID NO: 13 所示的核酸分子的编码序列(核苷酸 69 至 719)相同, 如 SEQ ID NO: 15 所示。

这些分子还可以有另一个不同序列, 因为遗传密码简并性, 该不同序列编码 SEQ ID NO: 2、SEQ ID NO: 4、SEQ ID NO: 7、SEQ ID NO: 6、SEQ ID NO: 8 或 SEQ ID NO: 14 所示的多肽。这些编码 SEQ ID NO: 2、SEQ ID NO: 4、SEQ ID NO: 7、SEQ ID NO: 6、SEQ ID NO: 8 或 SEQ ID NO: 14 所示的多肽的核酸分子可包括, 但不限于, 成熟多肽自身的编码序列; 成熟多肽的编码序列加上附加编码序列, 例如编码先导序列或分泌序列(如前多肽序列)的序列; 成熟多肽的编码序列, 加上或不加上前述附加编码序列, 再加上另外一些非编码序列, 包括非编码 5'和 3'序列, 例如在转录(包括终止信号)、核糖体结合和 mRNA 稳定性中发挥作用的转录非翻译序列。核酸分子还包括编码附加氨基酸的附加序列, 例如提供附加功能的那些氨基酸。

本发明第二和第三方面的核酸分子也可编码本发明第一方面的多肽及片段的片段或功能等同物。这样一种核酸分子可以是天然存在的变体, 例如天然存在的等位基因变体, 或者该分子可以是非天然存在的变体。核酸分子的非天然存在的变体可通过诱变技术制成, 包括适用于核酸分子、细胞或生物体的那些技术。

就此而言, 这些变体是通过核苷酸取代、缺失或插入而不同于上述核酸分子的变体。取代、缺失或插入可涉及一或多个核苷酸。变体可在编码区或非编码区或者二者内发生变化。编码区内的变化可产生保守性或非保守性氨

基酸取代、缺失或插入。

本发明的核酸分子也可以采用本领域公知的方法为多种原因而进行改造，这些原因包括修饰克隆、加工、和/或基因产物(多肽)的表达。可用来改造核苷酸序列的技术还包括应用随机断裂所作的 DNA 改组、基因片段和合成寡核苷酸的 PCR 重新装配。可利用定点诱变插入新的限制位点，改变糖基化模式，修改密码子偏向，产生剪接变体，引入突变等等。

编码本发明第一方面的多肽的核酸分子可与异源序列连接，以致该组合核酸分子编码融合蛋白。这样一些组合核酸分子包括在本发明第二和第三方面之内。例如，要在肽库中筛选抑制多肽活性的抑制剂，用这种组合核酸分子表达可被市售抗体识别的融合蛋白是有用的。融合蛋白还可被改造成含有位于本发明多肽的序列与异源蛋白质的序列之间的切割位点，以致该多肽可被切割，并从该异源蛋白质中纯化出来。

本发明的核酸分子还包括一些反义分子，它们与编码本发明的多肽的核酸分子部分互补，因而与编码核酸分子杂交。如本领域普通技术人员所知，这些反义分子，例如寡核苷酸，可设计成识别、特异性地结合以及防止编码本发明多肽的靶核酸的转录(例如，参见 Cohen, J.S., Trends in Pharm. Sci., 10, 435 (1989), Okano, J. Neurochem. 56, 560 (1991); O'Connor, J. Neurochem 56, 560 (1991); Lee 等, Nucleic Acids Res 6, 3073 (1979); Cooney 等, Science 241, 456 (1988); Dervan 等, Science 251, 1360 (1991))。

本文所用的术语“杂交”指两个核酸分子通过氢键相互缔合。通常，将一个分子固定于固相支持物上，而另一个分子在溶液中游离。然后，在有利氢键键合的条件下使两个分子相互接触。影响键合的因素包括：溶剂的类型和体积；反应温度；杂交时间；搅拌；封闭液相分子与固相支持物非特异性附着的试剂(Denhardt's 试剂或 BLOTTO)；分子的浓度；提高分子缔合率的化合物的使用(硫酸葡聚糖或聚乙二醇)；杂交后洗涤条件的严谨程度(参见上述文献 Sambrook 等)。

应用本领域公知的杂交测定法(例如，参见上述文献 Sambrook 等)，可检查一个完全互补分子与靶分子杂交的抑制。遵照 Wahl, G.M.和 S.L. Berger (1987; Methods Enzymol. 152: 399-407)以及 Kimmel, A.R. (1987; Methods Enzymol. 152: 507-511)的教导，在不同严谨程度的条件下，使基本上同源的

分子竞争并抑制完全同源的分子与靶分子的结合。

“严谨程度”指在杂交反应中与不同的分子缔合相比，有利于极度相似的分子缔合的条件。高度严谨杂交条件定义为 42°C 下在一溶液中培养过夜，该溶液含有 50%甲酰胺、5XSSC (150mM 氯化钠, 15mM 柠檬酸三钠)、50mM 5 磷酸钠(pH7.6)、5xDenhardts 溶液、10%硫酸葡聚糖和 20 微克/毫升变性剪切鲑鱼精子 DNA，然后，约 65°C 下在 0.1X SSC 中洗涤过滤器。低度严谨条件涉及在 35°C 下进行的杂交反应(参见上述文献 Sambrook 等)。杂交使用的优选条件是高度严谨杂交条件。

本发明该方面的优选例子是其全长至少有 70%与编码 INSP002 多肽(SEQ 10 ID NO: 2、SEQ ID NO: 4、SEQ ID NO: 7、SEQ ID NO: 6、SEQ ID NO: 8 和 SEQ ID NO: 14)的核酸分子相同的核酸分子以及基本上与这些核酸分子互补的核酸分子。本发明该方面的优选核酸分子含有一区域，该区域的全长至少有 80%与 SEQ ID NO: 1、SEQ ID NO: 10 和 SEQ ID NO: 12 所示的 SEQ ID NO: 2 和 SEQ ID NO: 7 编码序列、与 SEQ ID NO: 3 所示的 SEQ ID NO: 15 4 编码序列、与 SEQ ID NO: 5、SEQ ID NO: 9 和 SEQ ID NO: 11 所示的 SEQ ID NO: 6 和 SEQ ID NO: 8 编码序列、或者与 SEQ ID NO: 13 和 SEQ ID NO: 15 所示的 SEQ ID NO: 14 编码序列相同，或者是一个互补核酸分子。就此而言，特别优选的是其全长至少有 90%，优选至少 95%，更好至少 98%或 99%与所述核酸分子相同的核酸分子。优选的例子是编码基本上保留与 INSP002 20 多肽相同的生物功能或活性的多肽的核酸分子。

本发明还提供一种检测本发明核酸分子的方法，该方法包括以下步骤：
(a) 在杂交条件下将本发明的核酸探针与生物样品接触，形成双链体；以及(b) 检测所形成的任何双链体。

下文另外讨论关于本发明可采用的测定法，上述的核酸分子可用作 25 RNA、cDNA 或基因组 DNA 的杂交探针，分离编码 INSP002 多肽的全长 cDNA 和基因组克隆，以及分离与编码该多肽的基因有高度同一性的同源或直向同源基因的 cDNA 和基因组克隆。

在这点上，可以使用以下技术，其中包括已为本领域技术人员所知的，为了举例说明讨论如下。DNA 的测序和分析方法已为人公知，在本领域中 30 可以得到，事实上这些方法也用于以下讨论的本发明很多实施例中。这些方

法可使用酶，例如 DNA 聚合酶 I 的克列诺片段测序酶(US Biochemical Corp, Cleveland, OH)、Taq 聚合酶(Perkin Elmer)、耐热 T7 聚合酶(Amersham, Chicago, IL)或者这些酶的组合，以及校读外切核酸酶，例如 Gibco/BRL (Gaithersburg, MD)出售的 ELONGASE 扩增系统中找到的那些外切核酸酶。

- 5 优选的测序方法可应用 Hamilton Micro Lab 2200 (Hamilton, Reno, NV)、Peltier Thermal Cycler (PTC200; MJ Research, Watertown, MA)和 ABI 催化剂以及 373 和 377 DNA 测序仪(Perkin Elmer)等仪器进行自动操作。

一种分离编码具有与 INSP002 多肽等同功能的多肽的核酸分子的方法是按照本领域公认的标准程序用天然或人工设计的探针探测基因组或 cDNA 文库(例如，参见"Current Protocols in Molecular Biology", Ausubel 等(eds), Greene Publishing Association and John Wiley Interscience, New York, 1989, 1992)。特别有用的探针是：含有至少 15 个，优选至少 30 个，更好是至少 50 个邻接碱基，这些碱基对应于，或者与适当的编码基因的核酸序列(SEQ ID NO: 1、SEQ ID NO: 3、SEQ ID NO: 5、SEQ ID NO: 9、SEQ ID NO: 10、SEQ ID NO: 11、SEQ ID NO: 12、SEQ ID NO: 13 或 SEQ ID NO: 15)互补。使用可分析检测到的试剂标记探针，方便鉴定。有用的试剂包括，但不限于，放射性同位素、荧光染料和能够催化待检测产物形成的酶。有了这些探针，本领域普通技术人员就能够从人、哺乳动物或其他动物来源中分离出感兴趣的基因组 DNA、cDNA 或 RNA 聚核苷酸编码蛋白质的互补拷贝，并能够在这些来源中筛选出有关序列，例如该家族、类型和/或亚型的额外成员。

多数情况下，分离的 cDNA 序列是不完整的，因为编码多肽的区域通常在 5'端被截短。现时有一些方法可获得全长 cDNA，或者将短 cDNA 延长。应用部分核苷酸序列和本领域公知的方法检测上游序列，例如启动子和调控元件，可将这些序列延长。例如，可采用的一种方法是基于快速扩增 cDNA 末端的方法(RACE; 例如参见 Frohman 等, PNAS USA 85, 8998-9002, 1988)。最近，MarathonTM technology (Clontech Laboratories Inc.)对这种技术进行了改进，例如使较长 cDNA 的检索大大简化。另一种稍稍不同的技术称为“限制位点” PCR，它使用通用探针检索邻接一已知基因座的未知核酸序列(Sarkar, G. (1993) PCR Methods Applic. 2: 318-322)。以基于已知区域的趋异性引物，也可用反向 PCR 来扩增或延长序列(Triglia, T.等, (1988) Nucleic

Acids Res. 16: 8186)。捕捉 PCR 是可使用的另一种方法，它涉及通过 PCR 技术扩增邻接人和酵母人工染色体 DNA 中已知序列的 DNA 片段 (Lagerstrom, M. 等, (1991) PCR Methods Applic., 1, 111-119)。可用来检索未知序列的另一种方法是 Parker, J.D. 等描述的方法(1991, Nucleic Acids Res. 19: 3055-3060)。另外，人们可使用 PCR、嵌套引物和 PromoterFinderTM 文库来查看基因组 DNA (Clontech, Palo Alto, CA)。这种方法不需筛选文库，可用于寻找内含子/外显子接合。

当筛选全长 cDNA 时，最好用已经按大小进行选择包含了较大 cDNA 的文库。优选的还有随机引物文库，因为它们含有较多含基因 5'区的序列。对于寡 d(T)文库不能产生全长 cDNA 的情形，最好使用随机引物文库。基因组文库可用于将序列延伸进入 5'非转录调控区。

在本发明一实施例中，本发明的核酸分子可用于进行染色体定位。在这一技术中，核酸分子特异性靶向，并能与单个人染色体的特定位置杂交。对本发明染色体的相关序列作图谱，是确认那些序列与基因相关性疾病相互关系的重要步骤。一旦将序列与准确染色体位置在图谱反映出来，就可以将染色体上序列的物理位置与基因图谱数据关联起来。这些数据例如可以在人类孟德尔遗传学数据库中找到(Johns Hopkins 大学韦尔奇医学库在线提供)。然后，通过连锁分析(物理相邻基因的共同继承)确定业已映像在同一个染色体区的基因与疾病之间的关系。这为研究人员用定位克隆或其他基因发现技术检索疾病基因提供了有价值的信息。一旦通过遗传连锁分析粗略地把疾病或综合征限制在一特定基因组区，映像在该区域的任何序列代表相关或调控基因，可作进一步研究。核酸分子还可用来检测由于正常、载体或染病个体之间移位、反转等造成染色体位置的差异。

本发明的核酸分子对于组织定位也是有价值的。这些技术通过检测编码多肽的 mRNA 可以确定该多肽在组织中的表达模式。这些技术包括原位杂交技术和核苷酸扩增技术，例如 PCR。从研究中获得的结果可知道生物体中多肽的正常功能。另外，mRNA 的正常表达模式与突变基因编码的 mRNA 的正常表达模式之间的比较研究为理解突变多肽在疾病中的作用提供了有价值的认知。不适当表达可以是时间、空间或量化。

可采取基因沉默方法使编码本发明多肽的基因内源表达下调。一种方法

是 RNA 干扰(Elbashir, SM 等, Nature 2001, 411, 494-498), 可用来使序列特异性翻译后基因沉默。短 dsRNA 寡核苷酸在体外合成, 并导入细胞中。这些 dsRNA 寡核苷酸的序列特异性结合引发靶 mRNA 降解, 从而减少或消除蛋白质表达。

- 5 评估上述基因沉默方法的功效可用 TaqMan 方法在 RNA 水平上测量多肽的表达(例如, Western 印迹)。

本发明的载体包含本发明的核酸分子, 可以是克隆或表达载体。本发明的宿主细胞可以是原核细胞或真核细胞, 它们可用本发明的载体转化、转染或转导。

- 10 本发明的多肽的制备可以重组形式在宿主细胞所含的载体内表达它们的编码核酸分子。这些表达方法已为本领域技术人员公知, 很多方法已由上述的 Sambrook 和 Fernandez 与 Hoeffler 的书(1998, eds. "Gene expression systems, Using nature for the art of expression". Academic Press, San Diego, London, Boston, New York, Sydney, Tokyo, Toronto)详细叙述。

- 15 一般而言, 可使用适合维持、繁殖或表达核酸分子在一所需宿主内产生多肽的任何系统或载体。应用各种公知的常规技术, 例如上述 Sambrook 中描述的技术, 可将适当的核苷酸序列插入一表达系统中。通常, 可使编码基因受控于调控元件, 例如启动子、核糖体结合位点(对于细菌表达), 任选地是操纵子, 以便将编码所希望的多肽的 DNA 序列转录至转化宿主细胞的
20 RNA 内。

- 例如, 适当的表达系统的例子包括染色体系统、附加体系统和病毒衍生系统, 包括例如衍生自以下物质的载体: 细菌质粒、噬菌体、转位子、酵母附加体、插入元件、酵母染色体元件、病毒, 例如杆状病毒、乳多空病毒如 SV40、牛痘病毒、腺病毒、禽痘病毒、假狂犬病毒和逆转录病毒, 或者其组
25 合, 例如从质粒和噬菌体遗传因子所产生的载体, 包括粘粒和噬菌粒。人的人工染色体(HACs)也可用来输送在质粒中所含有和表达的较大 DNA 片段。

- 特别适合的表达系统包括微生物, 例如用重组噬菌体、质粒或粘粒 DNA 表达系统转化的细菌; 用酵母表达载体转化的酵母; 用病毒表达载体(例如杆状病毒)感染的昆虫细胞系统; 用病毒表达载体(例如, 花椰菜花叶病毒
30 CaMV; 烟草花叶病毒 TMV)或者用细胞表达载体(例如, Ti 或 pBR322 质粒)

转化的植物细胞系统；或者动物细胞系统。无细胞翻译系统也可用来产生本发明的多肽。

参照许多标准实验手册，例如 Davis 等(*Basic Methods in Molecular Biology* (1986))和上述 Sambrook 等中描述的方法，可将编码本发明的多肽的核酸分子导入宿主细胞内。特别适合的方法包括磷酸钙转染、DEAE-葡聚糖介导的转染、转位、微量注射、阳离子质粒介导的转染、电穿孔、转导、划痕加载、弹导导入或感染(参见上述 Sambrook 等, 1989; Ausubel 等, 1991; Spector, Goldman & Leinwald, 1998)。在真核细胞中，表达系统可以是暂时的(例如附加体)或者永久的(染色体整合)，取决于该系统的需要。

10 编码核酸分子可以包括或不包括编码调控序列，例如信号肽或先导序列的序列，有需要的时候，例如将翻译的多肽分泌到内质网腔、胞质间隙或胞外环境中。这些信号可以与该多肽同源，或者它们可以是异源信号。先导序列可在翻译后加工中被细菌宿主除去。

除了调控序列之外，希望可以加入能调节相对于宿主细胞生长的多肽表达的调节序列。调节序列的例子是可引起基因的表达对化学和物理刺激，包括调节化合物的存在，或者对各种温度或代谢条件的应答增大或减少的序列。调节序列是载体的那些非翻译区，例如增强子、启动子以及 5' 和 3' 非翻译区。它们与宿主细胞蛋白质相互作用，进行转录和翻译。这些调节序列的强度和特异性可以是不同的。取决于所用的载体系统和宿主，可使用任何数量的适当转录和翻译元件，包括组成型和诱导型启动子。例如在细菌系统内克隆时，可使用诱导型启动子，例如 Bluescript 噬菌粒的杂交 lacZ 启动子 (Stratagene, LaJolla, CA) 或 pSport1™ 质粒 (Gibco BRL) 等等。在昆虫细胞内可用杆状病毒多角体蛋白启动子。由植物细胞基因组(例如，热激、RUBISCO 和贮藏蛋白基因)或者植物病毒(例如，病毒启动子或先导序列)衍生而来的启动子或增强子可克隆到载体中。哺乳动物细胞系统优选使用哺乳动物基因或哺乳动物病毒的启动子。如果需要产生含有序列的多个拷贝的细胞系，可使用基于 SV40 或 EBV 的载体和一适当选择性标记物。

25 构建一个表达载体，以致于使特定核酸编码序列位于该带有适当调节序列的载体内，所述编码序列相对于调节序列的定位和方向使得该编码序列在该调节序列的“控制”下转录，即在调控序列下与 DNA 分子结合的 RNA 聚

合酶转录该编码序列。在一些情况下，有需要将序列修饰成它可以适当方向附着于调控序列，即维持读框。

调控序列和其他调节序列可在插入载体之前先连接到核酸编码序列。另一个选择是，直接将编码序列克隆到已经含有调控序列和适当限制位点的表达载体中。

对于长时间高得率地生产重组多肽，最好有稳定的表达。例如，稳定地表达感兴趣的多肽的细胞系可以用含有病毒来源复制的表达载体和/或同一个或不同载体上的内源性表达元件和选择性基因转化。导入载体之后，允许细胞先在富集培养基中生长 1-2 天，再转移到选择培养基中。选择性标记物的目的是针对选择带来抗性，它的存在可以使成功表达导入序列的细胞生长和恢复。稳定转化的细胞的抗性克隆可应用适合细胞类型的组织培养技术进行增殖。

本领域技术人员已知哺乳动物细胞系可用作表达宿主，而哺乳动物细胞系包括美国典型菌种保藏中心(ATCC)的许多无限增殖化细胞系，包括但不限于，中国仓鼠卵巢(CHO)细胞、HeLa 细胞、幼仓鼠肾(BHK)细胞、猴肾(COS)细胞、C127 细胞、3T3 细胞、BHK 细胞、HEK 293 细胞、Bowes 黑色素瘤细胞和人肝细胞癌(例如 Hep G2)细胞以及其他许多细胞系。

杆状病毒系统中，用于杆状病毒/昆虫细胞表达系统的物质在市场上以试剂盒形式提供，可以购自 inter alia、Invitrogen、San Diego CA (“MaxBac”试剂盒)。这些技术对本领域技术人员而言是众所周知的，在 Summers 和 Smith 的文章(Texas Agricultural Experiment Station Bulletin No. 1555 (1987))中也有所描述。特别适合该系统使用的宿主细胞包括昆虫细胞，例如果蝇 S2 细胞和草地夜蛾 Sf9 细胞。

本领域已经知道很多植物细胞培养基和全植物遗传表达系统。合适的植物细胞遗传表达系统的例子包括美国专利 US 5, 693, 506; US 5, 659, 122; 和 US 5, 608, 143 中描述的系统。植物细胞培养基的遗传表达的另外一些例子在 Zenk, *Phytochemistry* 30, 3861-3863 (1991)中已有描述。

具体地说，可使用原生质体被分离和培养得到全再生植物的所有植物，以致回收到的全植物含有转移基因。特别是所有植物可从培养细胞或组织中再生，包括但不限于甘蔗、糖用甜菜、棉花、水果和其他树木、豆类和蔬菜

的所有主要种属。

特别优选的细菌宿主细胞的例子包括链球菌、葡萄球菌、大肠杆菌、链霉菌和枯草芽孢杆菌细胞。

5 特别适合真菌表达的宿主细胞的例子包括酵母细胞(例如酿酒酵母)和曲霉菌细胞。

本领域已经知道许多选择系统, 它们都可用来回收转化细胞系。例如, 单纯疱疹病毒胸腺嘧啶激酶(Wigler, M.等, (1977) Cell 11: 223-32)和腺嘌呤磷酸核糖转移酶(Lowy, I.等, (1980) Cell 22: 817-23)基因可分别用在 tk-或 apr^t± 细胞中。

10 另外, 选择的基础可以是抗代谢物、抗生素或除草剂抗性; 例如, 二氢叶酸还原酶(DHFR)赋予氨甲蝶呤抗性(Wigler, M.等, (1980) Proc. Natl. Acad. Sci. 77: 3567-70); npt 赋予氨基糖苷新霉素和 G-418 抗性(Colbere-Garapin, F. 等, (1981) J. Mol. Biol. 150: 1-14), als 或 pat 分别赋予绿黄隆和草丁膦乙酰转移酶抗性。此外, 还描述了其他选择性基因, 它们的例子为本领域技术
15 人员所知。

虽然标记基因表达的存在或不存在提示感兴趣的基因也是存在的, 但它的存在和表达有待确认。例如, 如果在一标记基因序列中插入相关的基因, 不存在标记基因功能可以确定含有适当序列的转化细胞。另一个选择是, 在单个启动子控制下将标记基因与编码本发明的多肽的序列串联放置。标记基
20 因应答诱导或选择的表达通常也表示串联基因的表达。

另外, 含有编码本发明的多肽的核酸序列, 并表达所述多肽的宿主细胞可通过本领域技术人员公知的各种程序进行鉴定。这些程序包括但不限于: DNA-DNA 或 DNA-RNA 杂交和蛋白质生物分析, 例如, 荧光激活细胞分类术(FACS)或免疫测定技术(例如, 酶联免疫吸附测定法(ELISA)和放射性免疫
25 测定法(RIA)), 包括用来检测和/或定量分析核酸或蛋白质的膜、溶液或晶片技术(参见 Hampton, R.等, (1990) Serological Methods, a Laboratory Manual, APS Press, St Paul, MN 和 Maddox, D.E. 等, (1983) J. Exp. Med, 158, 1211-1216)。

本领域技术人员公知的各种各样标记和连接技术可用在不同的核酸和氨
30 基酸测定法中。产生标记杂交的装置或者检测与编码本发明的多肽的核酸分

子有关的序列的 PCR 探针包括,寡标记、镍翻译、末端标记或使用标记聚核苷酸的 PCR 扩增。另外,也可将编码本发明的多肽的序列克隆到载体内,以产生 mRNA 探针。这些载体是本领域公知的,可通过商业途径获得,加入诸如 T7、T3 或 SP6 等适当 RNA 聚合酶和标记核苷酸可在体外合成 RNA 5 探针。这些步骤使用市场上供应的各种试剂盒(Pharmacia & Upjohn, (Kalamazoo, MI); Promega (Madison WI); 和 U.S. Biochemical Corp., Cleveland, OH))进行。

易于检测的合适报导分子或标记包括放射性核、酶和萤光、化学发光或发色试剂和物质、协同因子、抑制剂、磁性粒子等等。

10 本发明的核酸分子还可用来制造转基因动物,特别是啮齿动物。这些转基因动物构成本发明的另一方面。这可通过局部修饰体细胞,或者通过种系治疗引入遗传修饰来实现。这些转基因动物特别适用于制作动物模型,分析作为本发明多肽的调节剂的药物分子。

15 从重组细胞培养基中回收和纯化多肽的方法是公知的,包括硫酸铵或乙醇沉淀、酸提取、阳离子或阴离子交换层析法、磷酸纤维素层析法、疏水基相互作用层析法、亲和力层析法、羟磷灰石层析法和血凝素层析法。高效液相层析法特别适用于纯化。当分离和/或纯化期间多肽发生变性时,可应用蛋白质再折迭的公知技术来再产生活性构造。

也可利用特异性载体结构来帮助蛋白质纯化,有需要的时候,将编码本 20 发明的多肽的序列与编码有利于可溶性蛋白质纯化的多肽结构域的核酸序列连接。有利于纯化的结构域的例子包括金属螯合肽,例如允许在固定金属上纯化的肌氨酸-色氨酸组件,允许在固定免疫球蛋白上纯化的蛋白 A 结构域,以及用在 FLAGS 延伸/亲和纯化系统(Immunex Corp., Seattle, WA)中的结构域。为了方便纯化,可以采用可切割的接头序列的内含物,例如在纯化结构 25 域与本发明的多肽之间的对 XA 因子或肠激酶有特异性的序列。一种这样的表达载体为融合蛋白的表达提供准备,所述融合蛋白含有本发明的多肽,与硫氧还蛋白或肠激酶切割位点前面的多个肌氨酸残基融合。肌氨酸残基有利于固定金属离子亲和力层析法(IMAC)的纯化,参见 Porath, J.等的描述(Prot. Exp. Purif. 3: 263-281, (1992)),而硫氧还蛋白或肠激酶切割位点为从融合 30 蛋白纯化多肽提供了一种手段。Kroll, D.J.等对含有融合蛋白的载体进行了

讨论(1993; DNA Cell Biol. 12: 441-453)。

如果多肽的表达是用于筛选测定中，通常最好是在它表达的宿主细胞表面上产生该多肽。在此情况下，宿主细胞可在用于筛选测定之前，例如用荧光激活细胞分类术(FACS)或免疫测定技术之前收获。如果多肽分泌到培养基中，可回收该培养基以便回收和纯化表达多肽。如果多肽是在细胞内产生，则回收多肽之前必须先溶解细胞。

本发明的多肽可用于筛选各种药物筛选技术使用的化合物库。这些化合物可以激活(激动)或抑制(拮抗)本发明多肽的基因的表达水平或活性，构成本发明的另一方面。优选的化合物是能有效地改变编码本发明第一方面的多肽的天然基因的表达，或者调节本发明第一方面的多肽的活性。

激动剂或拮抗剂化合物可从，例如，细胞制剂、无细胞制剂、化学库或天然混合产物中分离出来。这些激动剂或拮抗剂可以是天然的或者是经修饰的物质、配体、酶、受体或者结构或功能类似物。关于这些筛选技术的综述，可参见 Coligan 等，Current Protocols in Immunology 1(2): Chapter 5 (1991)。

最有可能成为良好拮抗剂的化合物是与本发明的多肽结合但在结合后又不会诱导该多肽生物作用的分子。潜在的拮抗剂包括有机小分子、肽、多肽和抗体，它们与本发明的多肽结合，由此抑制或消除它的活性。以这种方式，多肽与正常细胞结合分子的结合可被抑制，从而抑制该多肽的正常生物活性。

用在此筛选技术中的本发明的多肽可以在溶液中游离，附着于固相支持物上，留在细胞表面或位于细胞内。一般而言，这些筛选程序可涉及使用表达与测试化合物接触的多肽的适当细胞或细胞膜，观察结合、或功能反应的刺激或抑制。再将与测试化合物接触的细胞和不接触测试化合物的对照细胞的功能反应作比较。借助适当检测系统，这种测定方法可以评估测试化合物是否导致多肽激活产生一个信号。一般情况下，在已知激动剂的存在下分析激活的抑制剂，在测试化合物存在下观察激活对激动剂的影响。

一种鉴定本发明的多肽的激动剂或拮抗剂化合物的优选方法包括：

(a) 在允许与多肽结合的条件下，将在其表面上表达本发明第一方面的多肽的细胞与待筛选化合物接触，所述的多肽与能够在化合物与该多肽结合之后提供可检测信号的第二成分缔合。

(b) 通过测定化合物与多肽相互作用所产生的信号水平来确定该化合物是否结合和激活或抑制该多肽。

另一种鉴定本发明的多肽的激动剂或拮抗剂化合物的优选方法包括：

5 (a) 在允许与多肽结合的条件下，将在其表面上表达本发明第一方面的多肽的细胞与待筛选化合物接触，所述的多肽与能够在化合物与该多肽结合之后提供可检测信号的第二成分缔合。

(b) 通过对化合物与多肽相互作用所产生的信号水平与在该化合物不存在下的信号水平作比较来确定该化合物是否结合和激活或抑制该多肽。

10 在另一优选实施例中，上述的通用方法还包括在多肽的标记或无标记配体存在下对激动剂或拮抗剂进行鉴定。

在另一实施例中，鉴定本发明的多肽的激动剂或拮抗剂的方法包括：

15 在允许结合多肽的条件和候选化合物存在下，测定配体与其表面上带有本发明的多肽的细胞、或者与含有这样一种多肽的细胞膜结合的抑制，并测定与该多肽结合的配体数量。能够引起配体结合减少的化合物就认为是激动剂或拮抗剂。配体最好是带标记的。

更具体地说，筛选多肽拮抗剂或激动剂化合物的方法包括以下步骤：

20 (a) 使标记的配体与在细胞表面上表达本发明的多肽的全细胞，或者含有本发明的多肽的细胞膜培养，
(b) 测定与全细胞或细胞膜结合的标记配体数量，
(c) 在步骤(a)的标记配体和全细胞或者细胞膜的混合物中加入候选化合物，使该混合物达到平衡；
(d) 测定步骤(c)之后与全细胞或细胞膜结合的标记配体数量；以及
(e) 比较步骤(b)和(d)中结合的标记配体之差值，将引起步骤(d)的结合减少的化合物看成激动剂或拮抗剂。

25 在上述测定法中，发现多肽以剂量依赖方式调节各个生理和病理过程。所以，本发明的多肽的“功能等同物”包括在上述测定法中显示剂量依赖方式的任何相同的调节活性的多肽。虽然剂量依赖活性的程度不必与本发明的多肽完全一样，但“功能等同物”在给定的活性测定法中最好基本上具有与本发明的多肽相类似的剂量依赖关系。上述一些实施例可应用简单结合测定
30 法，其中测试化合物与带有多肽的表面的附着用直接或间接与该测试化合物

结合的标记检测，或者可应用涉及与标记竞争物竞争的测定法检测。另一个实施例采用竞争药物筛选测定法，其中能够结合多肽的中和抗体与测试化合物特异性地竞争结合。以此方式可用抗体检测对多肽有特异性结合亲和力的任何测试化合物的存在。

5 测定法也可设计成检测加入的测试化合物对细胞内编码多肽的 mRNA 产生的作用。例如，ELISA 可构建成通过本领域公知的标准方法以单克隆抗体或多克隆抗体测量多肽的分泌或细胞相关水平，这可用来检索抑制或增强在适当操纵的细胞或组织中产生多肽的化合物。然后，测定多肽与测试化合物之间的结合复合物的形成。

10 本发明包括的测定法是在过表达或切除检测中涉及使用本发明的基因和多肽的方法。这些测定法涉及细胞内这些基因/多肽水平的操纵，以及该操纵事件对被操纵细胞生理的影响的评估。例如，这些实验揭示与特定基因/多肽有关联的信号传导和代谢通道的详细信息，产生关于与待研究多肽相互作用的多肽同一性的信息，并提供调节相关基因和蛋白质的方法的思路。

15 可采用另一种药物筛选技术，它高通量筛选对感兴趣的多肽具有适当结合亲和力的化合物(参见国际专利申请 WO84/03564)。在这一方法中，在固相基质上合成大量不同的小测试化合物，它们再与本发明的多肽反应，洗涤。一种固定多肽的方法是用非中和抗体。以本领域公知的方法可检测结合多肽。纯化的多肽也可直接铺在板上，用于上述药物筛选技术中。

20 通过本领域公知的标准受体结合技术，例如配体结合和交联测定法，本发明的多肽可用来鉴定膜结合或可溶性受体，在配体结合和交联测定法中，多肽用放射性同位素标记，以化学方法修饰，或者与有利于它的检测或纯化的肽序列融合，并与推定受体来源(例如，细胞组合物、细胞膜、细胞上清液、组织抽提物或体液)一起培育。结合的效率可用生物物理技术，如表面
25 胞质基因共振和光谱测定。结合测定法用于纯化和克隆受体，但也可鉴定多肽的激动剂和拮抗剂，这些激动剂和拮抗剂与其受体竞争结合多肽。进行筛选测定的标准方法已为本领域所熟悉。

本发明还包括一种筛选试剂盒，它可用在鉴定上述的激动剂、拮抗剂、配体、受体、基质、酶的方法中。

30 本发明包括激动剂、拮抗剂、配体、受体、基质和酶以及通过上述方法

发现的能调节本发明的多肽的活性或抗原性的其他化合物。

本发明还提供一些药物组合物，其含有本发明的多肽、核酸、配体或化合物，以及合适的药物载体。这些组合物适合用作治疗或诊断试剂、疫苗、或其他免疫原性组合物，下文将作详细说明。

- 5 根据本文的术语，当组合物中以 X+Y 的总重量计时至少 85%是 X，认为含有多肽、核酸、配体或化合物[X]的组合物“基本上不含”杂质[本文以 Y 表示]。优选 X 占组合物中 X+Y 总重量的至少约 90%，更好至少约 95%、98%或者甚至 99%(重量)。

- 10 药物组合物最好含有治疗有效量的本发明的多肽、核酸分子、配体或化合物。本文所用的术语“治疗有效量”指需要治疗、缓解、或预防目标疾病或病征，或者具有可检测的治疗或预防作用的治疗药剂的用量。任何一种化合物的有效治疗剂量最初可在例如肿瘤细胞的细胞培养测定法、或者动物模型通常是小鼠、兔、狗、或猪模型中估计。动物模型还可用来确定适当的浓度范围和给药途径。有了这些信息，就可以确定给予人的有用给药剂量和途径。
- 15

- 人受试者的准确有效剂量取决于疾病状态的严重性、受试者的一般健康状况、受试者的年龄、体重和性别、饮食、给药的时间和频率、药物组合、反应灵敏度，以及治疗耐受性/反应。该用量可以常规实验测定，由医生判断。通常，有效剂量从 0.01 毫克/千克至 50 毫克/千克，优选从 0.05 毫克/千克至
- 20 10 毫克/千克。组合物可单独给予患者，或者与其他药剂、药物或激素结合给予。

- 一种药物组合物还可含有药学上可接受的载体，作为治疗试剂给予。这些载体包括抗体和其他多肽、基因和其他治疗试剂，例如脂质体，只要该载体本身不会诱导产生对接受该组合物的个体有害的抗体，而且载体的给予不会产生过高毒性。合适的载体可以是大的代谢缓慢的大分子，例如蛋白质、
- 25 聚糖、聚乳酸、聚羟基乙酸、聚合氨基酸、氨基酸共聚物和失活病毒颗粒。

- 还可使用药学上可接受的盐，例如无机酸盐，如盐酸盐、氢溴酸盐、磷酸盐、硫酸盐等等；以及有机酸盐，如乙酸盐、丙酸盐、丙二酸盐、苯甲酸盐等等。Remington's Pharmaceutical Sciences (Mack Pub. Co., N.J. 1991)对药
- 30 学上可接受的载体作了详尽讨论。

治疗组合物中药学上可接受的载体也可含有水、盐水、甘油和乙醇等液体。另外，这些组合物还可含有辅助物质，例如，湿润剂或乳化剂、pH 缓冲物质等等。这些载体能将药物组合物配制成供患者摄取的片剂、丸剂、糖衣片、胶囊、液体、凝胶、糖浆、膏剂、悬浮液等等。

5 一旦配成以后，本发明的组合物可直接给予受试者。接受治疗的受试者可以是动物；特别是人受试者。

本发明使用的药物组合物可以任何途径给予，包括但不限于，口服、静脉内、肌内、动脉内、脊髓内、鞘内、心室内、透皮或经皮（例如参见 WO98/20734）、皮下、腹腔内、鼻内、肠内、局部、舌下、阴道内或直肠途径。本发明的药物组合物也可用基因枪或无针注射器给予。一般将治疗组合物制成可注射的液体溶液或悬浮液；也可以制成适合注射前溶解或悬浮于液体介质中的固体形式。

通过注射可将组合物直接输送到皮下、腹腔内、静脉内或肌内，或者送到组织间质空间。组合物也可施加在病变部位。剂量治疗可分单剂或多剂。

15 如果本发明的多肽的活性在一特定疾病状态中是过度的，有多种方法可采用。其中一种方法包括把上述抑制剂化合物(拮抗剂)与药学上可接受的载体一起给予受试者，抑制剂化合物的量为通过例如阻断配体、基质、酶、受体的结合，或者抑制第二信号而抑制多肽的功能，从而缓解异常状况的有效量。这些拮抗剂优选是抗体。这些抗体最好是嵌合抗体和/或人源化抗体，如
20 前所述，可使它们的免疫原性减至最低。

另一种方法是给予保留了对上述配体、基质、酶、受体有结合亲和力的多肽的可溶形式。通常，多肽以保留相关部分的片段的形式给予。

在一替换方法中，应用表达封闭技术，例如用内部产生或另外给予的反义核酸分子(如上所述)，抑制编码多肽的基因表达。设计编码多肽的基因的
25 调控区、5'区或调节区(信号序列、启动子、增强子和内含子)的互补序列或反义核酸分子(DNA、RNA 或 PNA)，可获得对基因表达的修饰。类似地，用“三螺旋体”碱基成对方法可实现抑制。三螺旋体成对是有用的，这是因为它能引起抑制双螺旋体充分打开以结合聚合酶、转录因子或调节分子的能力。应用三螺旋 DNA 的治疗取得最新进展，在文献中已有描述(Gee, J.E.等,
30 (1994) In: Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches,

Futura Publishing Co., Mt. Kisco, NY)。互补序列或反义分子还可设计成防止转录子与核糖体结合，从而封闭 mRNA 的翻译。这些寡核苷酸可给予，或在体内表达中原位产生。

另外，使用对它的编码 mRNA 序列有特异性的核糖体，可防止表达本发明的多肽。核糖体是具有催化活性的天然或合成 RNA(例如，参见 Usman, N 等, Curr. Opin. Struct. Biol (1996) 6(4), 527-33)。可设计合成核糖体，在选定位置上特异性切割 mRNA，从而防止 mRNA 翻译为功能多肽。核糖体通常在 RNA 分子中找到，可用天然的磷酸核糖骨架和天然碱基合成。另一个选择是，核糖体可用非天然骨架，例如 2'-氧-甲基 RNA 合成，保护核糖核酸酶免于降解，还可含有经修饰的碱基。

可对 RNA 分子进行修饰，以增加胞内稳定性和延长半衰期。可能的修饰包括但不限于，在分子的 5'和/或 3'末端加入侧翼序列或在分子的骨架内使用硫代磷酸或 2'-氧-甲基而不用磷酸二酯酶连锁。这一概念对于产生 PNA 是固有的，可延伸至所有这些分子中，方法是包含非常规碱基，例如肌苷、核苷(queosine)和高修饰鸟苷(butosine)以及乙酰基-、甲基-、硫代-、及类似修饰形式的腺嘌呤、胞苷、鸟嘌呤、胸腺嘧啶和尿苷，它们不易被内源性内切核酸酶识别。

有一些方法治疗与本发明的多肽表达不足够或其活性相关的异常征状。其中一种方法包括把治疗有效量的能激活该多肽的化合物，即上述激动剂给予患者，缓解异常征状。另外，也可以给予治疗量的多肽与适当药物载体，以恢复多肽的相关生理平衡。

基因疗法可用来在受试者体内通过相关细胞内源性产生多肽。基因疗法以修正治疗基因置换缺陷基因，永久治疗多肽的不适当产生。

本发明的基因疗法可发生于体内或体外。体外基因疗法需要分离和纯化患者的细胞，导入治疗基因，以及将经基因改造的细胞再引入患者体内。与之相反，体内基因疗法不需要分离和纯化患者的细胞。

治疗基因往往被“包装”以方便给予患者。基因输送赋形剂可以是非病毒，例如脂质体，或者复制缺陷型病毒，例如 Berkner, K.L.(Curr. Top. Microbiol. Immunol., 158, 39-66 (1992))所述的腺病毒，或者 Muzyczka, N.(Curr. Top. Microbiol. Immunol., 158, 97-129 (1992))和美国专利 US5, 252, 479 所述的

腺伴随病毒(AVV)载体。例如, 可对编码本发明的多肽的核酸分子进行改造, 以在复制缺陷型逆转录病毒载体中表达。然后分离这种表达构造体, 导入用含有编码多肽的 RNA 的逆转录病毒质粒载体转导的包装细胞内, 使该包装细胞现在能产生含有感兴趣的基因的感染病毒粒子。将这些生产细胞给予患者, 以体内改造细胞和体内表达多肽(参见“Gene Therapy and other Molecular Genetic-based Therapeutic Approaches”(纳入作为文献参考), Human Molecular Genetics (1996), T Strachan 和 A P Read, BIOS Scientific Publishers Ltd)。

另一种方法是给予“裸露 DNA”, 其中治疗基因直接注入血流或肌肉组织。

对于本发明的多肽或核酸分子是引起疾病的试剂的情形, 本发明提出它们可用在疫苗中, 产生针对疾病引发剂的抗体。在上述多肽或核酸分子被上调的情形下, 疫苗的研制可涉及产生抗这些试剂的抗体或 T 细胞(描述见 WO 00/29428)。

本发明的疫苗可以是预防性(即预防感染)或者是治疗性的(即治疗感染后疾病)。这些疫苗包括免疫抗原、免疫原、多肽、蛋白质或核酸, 通常与前述药学上可接受的载体结合, 包括本身不会诱导产生对接受该组合物的个体有害的任何载体。这些载体还可用作免疫刺激剂(“辅助剂”)。此外, 抗原或免疫原可与细菌类毒素例如来自白喉、破伤风、霍乱、幽门螺杆菌和其他致病原的类毒素偶联。

因为多肽可以在胃内分解, 所以含有多肽的疫苗最好肠胃外给予(例如皮下、肌内、静脉内、或皮内注射)。适合肠胃外给予的制剂包括无菌水溶液或非水溶液, 其可含有抗氧化剂、缓冲液、抑菌剂和使制剂与接受者的血液等渗的溶解物, 以及无菌水悬浮液或非水悬浮液, 其可包括悬浮剂或增稠剂。

本发明的疫苗制剂可制成单剂或多剂容器。例如, 密封安瓿和小瓶可储存于冻干条件下, 只要在使用之前加入无菌液体载体。剂量取决于疫苗的比活性, 通过常规实施可轻易测定。

如国际专利申请 WO 98/55607 所述, 与本发明的多肽结合的抗体的基因传送也会受影响。快速注射技术(例如访问 www.powderject.com)可用来配制疫苗。

国际专利申请 WO 00/29428 描述了很多适当的接种疫苗方法和疫苗传送系统。

本发明还涉及本发明的核酸分子作为诊断试剂的应用。检测以本发明的核酸分子为特点，并与功能障碍相关的基因的突变形式提供一种诊断工具，
5 它可加入或限定因为基因的表达不足、过度表达或者空间或时间表达发生改变而引起的疾病或疾病易感性的诊断。通过各种技术可在 DNA 水平上检测携带基因突变的个体。

用于诊断的核酸分子可从受试者的细胞中获得，例如从血液、尿、唾液、组织活体或尸体材料中获得。基因组 DNA 可直接用作检测，或者在分析前
10 用 PCR、连接酶链式反应(LCR)、链置换扩增(SDA)或者其他扩增技术酶解扩增(参见 Saiki 等, Nature, 324, 163-166 (1986); Bej 等, Crit. Rev. Biochem. Molec. Biol., 26, 301-334 (1991); Birkenmeyer 等, J. Virol. Meth., 35, 117-126 (1991); Van Brunt, J., Bio/Technology, 8, 291-294 (1990))。

在一实施例中，本发明该方面提供一种诊断患者疾病的方法，该方法包
15 括评估编码本发明的多肽的天然基因的表达水平，并将所述表达水平与对照水平作比较，若该水平与所述对照水平不同，则表示患病。这种方法包括以下步骤：

- (a) 在允许本发明的核酸分子和核酸探针形成杂交复合物的严谨条件下，将患者的组织样品与探针接触；
- 20 (b) 在与步骤(a)相同的条件下，将对照样品与所述探针接触；
- (c) 以及检测所述样品中是否存在杂交复合物。

其中，检测到患者样品中杂交复合物的水平与对照样品中杂交复合物不同，表示患病。

本发明另一方面包括一种诊断方法，其包括以下步骤：

- 25 (a) 从待测试疾病的患者中获得组织样品；
- (b) 从所述组织样品中分离本发明的核酸分子；
- (c) 通过检测与疾病相关的核酸分子是否存在突变，诊断患者疾病。

为了帮助上述方法检测核酸分子，可包括扩增步骤，例如利用 PCR 技术。

30 将扩增产物大小的变化与正常基因型比较，可检测到缺失或插入。将扩

增 DNA 与本发明的标记 RNA，或者与本发明的标记反义 DNA 序列杂交，可以鉴定点突变。以核糖核酸酶消化或评估熔点温度的差异，发现完全匹配的序列与错配的双链体截然不同。患者是否存在突变可这样检测：将 DNA 与在严谨条件下和该 DNA 杂交的核酸探针接触，形成杂交双链分子，该杂交双链分子在对应于与疾病相关的突变的任何部分具有核酸探针链的未杂交部分；检测探针链的未杂交部分是否存在即表示该 DNA 链的对应部分存在或不存在与疾病相关的突变。这样的诊断特别适用于产前，甚至新生儿测试。

参考基因与“突变体”基因之间的点突变和其他序列差异可通过公知技术鉴定，例如，直接 DNA 测序或单链构象多态性(参见 Orita 等, *Genomics*, 5, 874-879 (1989))。例如，测序引物可与改良 PCR 产生的双链 PCR 产物或单链模板分子一起使用。以使用放射性标记核苷酸的常规程序或者通过使用荧光标记的自动测序程序，进行序列测定。克隆 DNA 节段也可用作检测特异性 DNA 片段的探针。当与 PCR 结合时该方法的灵敏度大大提高。另外，点突变和其他序列变异，例如多态性，可参照上述方法检测，例如使用等位基因特异性寡核苷酸对不同于单个核苷酸的序列进行 PCR 扩增。

在变性试剂存在或不存在下改变 DNA 片段的凝胶电泳迁移率，或直接对 DNA 测序都可以检测到 DNA 序列的差异(例如 Myers 等, *Science* (1985) 230: 1242)。特定位置的序列变化可通过核酸酶保护测定法揭示，例如 RNase 和 S1 保护，或者化学切割方法发现(Cotton 等, *Proc. Natl. Acad. Sci. USA* (1985) 85: 4397-4401)。

除了常规凝胶电泳和 DNA 序列测定之外，微小缺失、非整倍性、易位、倒位等突变也可用原位分析加以检测(例如，参见 Keller 等, *DNA Probes*, 2nd Ed., Stockton Press, New York, N.Y., USA (1993))，换言之，可分析细胞内 DNA 或 RNA 序列的突变，无需分离或固定在膜上。荧光原位杂交(FISH)是目前最常用的方法，就 FISH 的综述已有很多报导(例如，参见 Trachuck 等, *Science*, 250, 559-562 (1990), and Trask *et al.*, *Trends, Genet.*, 7, 149-154 (1991))。

本发明另一实施例中，构建包括本发明核酸分子的寡核苷酸探针阵列，对遗传变体、突变和多态性进行有效筛选。阵列技术方法是公知的，获得广泛应用，可用来解释分子遗传方面的各种问题，包括基因表达、遗传连锁和

遗传变异性(例如, 参见 M.Chee 等, *Science* (1996), Vol 274, pp 610-613)。

在一实施例中, 按照 PCT 申请 WO95/11995 (Chee 等); Lockhart, D. J. 等((1996) *Nat. Biotech.* 14: 1675-1680)); 以及 Schena, M.等((1996) *Proc. Natl. Acad. Sci.* 93: 10614-10619))描述的方法制作和使用阵列。寡核苷酸对的范
5 围从 2 至 1,000,000 以上。应用光引导化学法在基质的指定区域上合成低聚物。基质可以是纸、尼龙或其他种类的膜、过滤器、晶片、玻璃载片或其他任何合适的固相支持物。另外, 参照 PCT 申请 W095/251116 (Balteschweiler 等)中描述的方法, 使用化学偶联程序和喷墨应用装置可在基质表面上合成寡核苷酸。另一方面, 利用真空系统、热学、紫外线、机械或化学结合程序,
10 以类似斑点(或狭线)印迹的“栅格”阵列将 cDNA 片段或寡核苷酸设置和连接于基质表面上。阵列, 例如上述的那些阵列, 可用人手或现有设备(斑点印迹或狭线印迹装置)、材料(任何合适的固相支持物)和机器(包括自动仪器)制作, 它们可含有 8、24、96、384、1536 或 6144 个寡核苷酸, 或者 2 至 1,000,000 以上之间的任何数目, 该数目能够使阵列有效地应用于市场上买到的仪器
15 中。

除了上述讨论的方法之外, 通过包括测定受试者样品中多肽或 mRNA 水平异常增加或减少的方法可以诊断疾病。采用本领域公知的任何方法可以在 RNA 水平上测定表达减少或增加, 以便对聚核苷酸作定量分析, 例如核酸扩增, 如 PCR、RT-PCR、RNase 保护、Northern 印迹和其他杂交方法。

20 可用来确定从宿主获得的样品中本发明的多肽水平的测定技术已为本领域技术人员所知, 在上文提到的一些文献中已有讨论(包括放射性免疫测定法、竞争结合测定法、Western 印迹分析和 ELISA 试验)。本发明该方面提供一种诊断方法, 其包括以下步骤: (a) 在适合形成配体-多肽复合物的条件下, 将上述一种配体与生物样品接触; 以及(b)检测所述复合物。

25 测定多肽水平的方案, 例如 ELISA、RIA 和 FACS 还可为诊断多肽表达的变化或异常水平提供基础。在适合形成复合物的条件下将正常哺乳动物受试者, 优选是人的体液或细胞抽提物与针对多肽的抗体结合, 确立多肽表达的正常或标准值。标准复合物形成的数量可用各种方法定量测定, 例如光度计装置。

30 特异性结合本发明的多肽的抗体可用来诊断以多肽表达为特点的症状或

疾病，或者用在测定法中监测以本发明的多肽、核酸分子、配体和其他化合物治疗的患者。用于诊断目的的抗体可以用上述关于治疗所述的相同方法制成。对多肽的诊断分析包括利用抗体和标记检测人体液或者细胞或组织抽提物的多肽的方法。使用的多肽可以修饰或不作修饰，将它们与报导分子共价或非共价结合加上标记。可以应用本领域已知的各种各样报导分子，其中一些已在上文作了描述。

将在受试者活体组织获得的对照样品和患病样品中表达的多肽数量与标准值作比较。标准值与受试者值之间的偏差建立诊断疾病的参数。可用诊断分析区分多肽不存在、存在和过度表达，并监测治疗干预期间多肽水平的调节。这些分析也可用来评估动物研究、临床试验或个体患者的监测治疗中特定治疗方案的疗效。

本发明的一种诊断试剂盒可包括：

- (a) 本发明的核酸分子；
- (b) 本发明的多肽；或者
- 15 (c) 本发明的配体。

本发明一方面提供一种诊断试剂盒，该试剂盒可包括第一容器，其含有在严谨条件下与本发明的核酸分子杂交的核酸探针；第二容器，其含有用来扩增该核酸分子的引物；以及该探针和引物的使用说明书，方便诊断疾病。这种试剂盒还可包括装有消化未杂交 RNA 的试剂的第三容器。

20 本发明另一方面也提供一种诊断试剂盒，该试剂盒可包括核酸分子阵列，其中至少一个核酸分子是本发明的核酸分子。

为了检测本发明的多肽，一种诊断试剂盒可包括一或多种与本发明的多肽结合的抗体；一种用来检测该抗体与多肽之间结合反应的试剂。

25 这些试剂盒都可用于诊断疾病或疾病易感性，特别是细胞增殖性疾病、自身免疫/炎症疾病、心血管疾病、神经系统疾病、发育障碍、代谢性疾病、感染和其他病理病征。这些疾病或障碍优选涉及胱氨酸结折迭细胞因子，更好是涉及 DAN 亚家族成员的异常水平。这些疾病或障碍也可以涉及胱氨酸结折迭细胞因子，更好是涉及 DAN 亚家族成员的配体的异常水平。例如，这些疾病或障碍可以涉及 TGF β 超家族成员的异常水平。具体地说，这些障碍或疾病可以涉及 BMP，例如神经病变、肾病，如糖尿病性神经病变，癌

30

症、伤口愈合、纤维变性、骨质减少、骨质疏松、骨折和硬化性狭窄。

以下，将通过举例说明，特别是结合 INSP002 多肽对本发明的各个方面和实施例进行描述。

应明白，在不脱离本发明的范围的情况下可以作出改型。

5

附图说明

图 1: 用组合的 SEQ ID NO: 2 和 SEQ ID NO: 4 多肽序列通过 BLAST 在 NCBI 非冗余数据库中击中的结果。

10 图 2: 在组合的 SEQ ID NO: 2 和 SEQ ID NO: 4 多肽序列与最接近相关序列智人 cerberus 相关性 1 蛋白质之间通过 BLAST 所产生的比对。

图 3: 2002 年 11 月 26 日就 INSP002 经 BLAST 在 NCBI-nr 与 NCBI-nt 数据库中获得的上部 4 个击中。

图 4: INSP002 与 AK095926.1 的比对。

15 图 5: INSP002 与 IMAGE: 4558384 的比对。

图 6: INSP002 核苷酸序列与翻译。

图 7: INSP002 部分克隆序列与翻译。

图 8: 部分 PCRII-TOPO-INSP002 的图谱。

20 图 9: INSP002 预测序列(上部)与部分克隆序列(下部)的比对，其中图 9a 是核苷酸序列比对，图 9b 是蛋白质序列比对。

图 10: Image: 4558384 的 cDNA 插入片段的核苷酸序列和翻译。

图 11: INSP002 预测序列(上部)与 IMAGE: 4558384 (BC025333.1 (下部)的序列比对，其中图 11a 是核苷酸序列，图 11b 是蛋白质序列。

25 图 12: 通过 PCR 技术在 Image: 4558384 中产生的 INSP002V 的核苷酸序列和翻译。

图 13: pCR4blunt-TOPO-INSP002V 的图谱。

图 14: INSP002 预测序列(上部)与 INSP002V 变体序列(下部)之间的比对，其中图 14a 是核苷酸序列，图 14b 是蛋白质序列。

图 15: 表达质粒 pEAK12d 的图谱。

30 图 16: Gateway 载体 pDONR201 的图谱。

图 17: pEAK12d-INSP002-V-6HIS 的图谱。

图 18: 从心脏克隆的全长 INSP002 的序列。

图 19: 编码从心脏克隆的全长 INSP002 的质粒 pCR4-bluntTOPO-INSP002FL 的图谱。

5

具体实施方式

实施例 1 INSP002 蛋白质与序列数据库的蛋白质的比对

结合 SEQ ID NO: 2 和 SEQ ID NO: 4 产生的多肽序列表示 INSP002 的连续外显子的翻译, 该序列可用作针对 NCBI 非冗余序列数据库的 BLAST 查询序列。上部 10 对匹配包括标注为 *cerberus* 或 *cerberus* 相关蛋白质的序列, 这些蛋白质是胱氨酸结家族成员, 它们全部与查询序列比对的 E 值都很高($2E^{-10}$ 至 $3E^{-06}$)(图 1)。图 2 所示为 INSP002 查询序列与智人 *cerberus* 相关性 1 蛋白质序列之间的序列比对(Feng 等, 2001)。

15 结合 SEQ ID NO: 2 和 SEQ ID NO: 4 产生的多肽序列表示 INSP002 的连续外显子的翻译, 将该序列输入 SignalP V2.0.b2 (Nielsen 等, 1997 Protein Eng 1: 1-6)。程序预计多肽序列有一信号肽。发现该信号肽最可能的切割位点在结合 SEQ ID NO: 2 和 SEQ ID NO: 4 产生的 INSP002 多肽序列的残基 22 和 23 之间。

20 编码多肽 SEQ ID NO: 2 外显子 1 的核苷酸序列 SEQ ID NO: 1 包括 5'非翻译区(5'UTR)和蛋白质编码序列(CDS)。CDS 起点是核苷酸 152。

实施例 2 用 BLAST 重复检索

以结合 SEQ ID NO: 2 和 SEQ ID NO: 4 产生的 SEQ ID NO: 6 多肽序列, 在 2002 年 11 月 26 日对 NCBI-nr 和 NCBI-nt 数据库进行 BLAST 检索。图 3 所示为检索确定的上部四个击中。

检索发现, INSP002 多肽与从心脏克隆并在 2002 年 7 月 16 日保藏的假定的蛋白质 FLJ38607 氨基酸水平以及对应核苷酸序列 AK095926 相同。图 4 所示为 INSP002 查询序列与 AK095926 cDNA 克隆衍生的蛋白质的比对。

30 AK095926 的存在从实验上证实了对 INSP002 所预测的外显子 1 和外显

有 1X AmpliTaq™ 缓冲液、200 μM dNTPs、克隆引物各 50 pmole、2.5 个单位 AmpliTaq™ (Perkin Elmer)和噬菌体文库 DNA 各 100 ng, 程序如下: 94°C, 1 分钟; 94°C 循环 40 次, 1 分钟, $x^{\circ}\text{C}$, 以及 y 分钟, 72°C(其中 x 是最低温度-5°C, $y=1$ 分钟/kb 产物); 然后, 72°C 循环 1 次, 7 分钟, 4°C 持续循环。

5 在 1X TAE 缓冲液(Invitrogen)中的 0.8%琼脂糖凝胶上看到扩增产物, 用 Wizard PCR Preps DNA 纯化系统(Promega)从凝胶纯化得到以预定分子量迁移的 PCR 产物。PCR 产物在 50 微升无菌水中洗脱出来, 或者直接亚克隆或者储存于-20°C。

10 (iii) 用于 PCR 的基因特异性克隆引物

设计长度为 18 至 25 个碱基的 PCR 引物对, 以便运用引物设计软件 (Scientific & Educational Software, PO Box 72045, Durham, NC 27722-2045, USA)扩增虚拟 cDNA 的全长序列。将 PCR 引物优化至温度接近 $55\pm 10^{\circ}\text{C}$, GC 含量为 40-60%。选择对靶序列 INSP002 有高度选择性的引物(对其他模板几乎没有不是非特异性引发)。

(iv) PCR 产物的亚克隆

用购自 Invitrogen Corporation 的 TOPT TA 克隆试剂盒(cat. No.分别是 K4600-01 和 K4575-01), 在生产商指定的条件下将 PCR 产物亚克隆到拓朴异构酶 I 修饰克隆载体(pCR II TOPO)中。简言之, 从人胎肾文库(文库 12)扩增中取 4 微升凝胶纯化的 PCR 产物, 室温下与 1 微升 TOPO 载体和 1 微升盐溶液培养 15 分钟。再按以下方法将反应混合物转化到大肠杆菌菌株 TOP10 (Invitrogen)中: 50 微升 One Shot TOP10 细胞等分试样在冰上解冻, 加入 20 微升 TOPO 反应物。该混合物冰上培育 15 分钟, 再在 42°C 培育热激刚好 30 秒。将样品送回冰上, 加入 250 微升温热 SOC 培养基(室温)。样品在 37°C 摇动培育(220 rpm)1 小时。然后, 将转化混合物铺在含有氨比西林(100 μg/ml)的 L-broth (LB)板上, 37°C 培育过夜。通过集落 PCR 技术鉴定含有 cDNA 插入片段的氨比西林抗性集落。

30 (v) 集落 PCR

用无菌牙签将集落接种在 50 微升无菌水中。然后，按照上述方法，将 10 微升接种物等分试样在 20 微升总反应体积中进行 PCR 扩增，但引物对用 SP6 和 T7。循环条件如下：94°C，2 分钟；94°C 循环 30 次，30 秒，47°C，30 秒，以及 72°C 1 分钟；72°C 循环 1 次，7 分钟。作进一步分析之前将样品保持
5 保持在 4°C(持续循环)。

在 1X TAE 缓冲液中的 1%琼脂糖凝胶上分析 PCR 反应产物。使产生预期 PCR 产物大小(由于多个克隆位点(MCS)而得到 159 bp cDNA+187 bp)的集落生长过夜，生长条件：温度为 37°C，5 毫升含有氨比西林(100 µg/ml)的 L-broth (LB)板，37°C 220 rpm 摇动。

10

(vi) 质粒 DNA 的制备和测序

按照生产商的指示，用 Qiaprep Turbo 9600 自动化系统(Qiagen)或 Wizard Plus SV Minipreps 试剂盒(Promega cat. no. 1460)在 5 毫升培养基中制备小量制备质粒 DNA。将质粒 DNA 在 100 微升无菌水中洗脱。用 Eppendorf BO
15 光度计测定 DNA 浓度。按照生产商的指示，用 BigDyeTerminator 系统(Applied Biosystems cat. no. 4390246)以 T7 和 SP6 为引物对质粒 DNA(200-500 ng)进行 DNA 测序。测序反应用 Dye-Ex 柱(Qiagen)或者 Montage SEQ 968 净化板(Millipore cat. no. LSKS09624)纯化，再用 Applied Biosystems 3700 测序仪分析。

20

(vii) 含有 INSP002 的 cDNA 文库的鉴定

在皮层、结肠、胎肺和胎肾 cDNA 文库(文库 8、9、11 和 12)中鉴定用 INSP002-CP1 和 INSP002-CP2 获得并以正确大小(159 bp)迁移的 PCR 产物。图 7 所示为 pCRII-TOPO 载体中克隆的 PCR 产物的序列，图 8 所示为质粒(质粒 ID
25 13422)的图谱。克隆的部分 cDNA 是 INSP002 外显子 2 的一部分，可参见图 9a 所示的预测 INSP002 核苷酸序列和克隆的部分核苷酸序列的比对，以及图 9b 所示的预测 INSP002 蛋白质序列和克隆的部分核苷酸序列的比对。

实施例 4 从 Image: 4558384 中产生 INSP002 全编码序列

30 眼癌的 Image 克隆 4558384(在质粒 pOTB7 中)购自 Resgen (Invitrogen

Corporation)。将 4558384 大肠杆菌穿刺并分散在含有氨比西林(100 µg/ml)的 LB 板上,37°C 生长过夜。使单氨比西林抗性克隆接种在 5 毫升含氨比西林(100 µg/ml)的 LB 中, 37°C 220 rpm 摇动培育一晚。制备小量制备质粒 DNA, 参照实施例 3 (vi)所述, 用 SP6、T7、M13F、INSP002-CP1 和 INSP002-CP2 引物测序。

插入片段的序列见图 10。Image 4558384 cDNA 与 INSP002 的核苷酸和推定氨基酸序列的比对见图 11。Image 4558384 cDNA 似乎是 INSP002 的剪接变体。它含有一个 87 bp 插入片段, 与 INSP002 预测序列相比, 引入了一个移码和未成熟终止密码子。另外, 3'未翻译序列还含有表示 cDNA 的基因组 DNA 污染的 Alu 复制。Image 4558384 的外显子 2 和 4 等同于 INSP002 预测序列的外显子 1 和 2。然而, Image 455838 在 INSP002 预测序列的外显子 1 和 2 之间插入了一个附加外显子。该附加外显子编码防止胱氨酸结构域翻译的未成熟终止密码子。Image 4558384 的剪接边界如下:

```

15      500      .   :   .   :   .   :   .   :   .   :
      492 GGCTGTGCCCTTCGTTTCAG          ACACGGGAGTCTCGCTATGTTG
          |||||||||||||||||||>>>...>>>|||||||||||||||||||||
10234 GGCTGTGCCCTTCGTTTCAGGTG...TAGACACGGGAGTCTCGCTATGTTG
      外显子 2          外显子 3
20      550      .   :   .   :   .   :   .   :   .   :
      533 CCCAAGCTAGTCTTGAGCTTCTGGCCTCAAGCAATCCTCCCACCTCAGCC
          |||||||||||||||||||>>>...>>>|||||||||||||||||||||
10741 CCCAAGCTAGTCTTGAGCTTCTGGCCTCAAGCAATCCTCCCACCTCAGCC
      600      .   :   .   :   .   :   .   :   .   :
25      583 TCC █████ GTTCTAG          GTGTTCTCCCGGCCCGGCTGCTCAGCCA
          || █████ |||||||>>>...>>>|||||||||||||||||||||
10791 TCC █████ GTTCTAGGTG...CAGGTGTTCTCCCGGCCCGGCTGCTCAGCCA
                                                    外显子 4

```

设计 PCR 策略, 除去基因组 DNA 以产生编码 INSP002 全编码序列(ORF)的全长 cDNA。设计 PCR 引物扩增位于 Image 克隆的 87 bp 插入片段侧翼的 INSP002 的 5'端(上游)和 3'端(下游)。上游序列的反向引物和下游序列的正向引物在它们的 3'和 5'端分别含有互补序列, 产生重迭端, 这样, 可以将每

次反应的 PCR 产物一起混合和退火，从而可以用嵌套上游正向引物和嵌套下游反向引物在第三次 PCR 反应中扩增全长 cDNA。

第一次 PCR 反应扩增所含的 INSP002 5'端(87 bp 插入片段的上游)，50 微升终体积含有：0.5 μ l 10X Platinum Pfx 缓冲液、1.5 μ l dNTPs (10mM)、1 μ l 5 硫酸镁(50 mM)、1.5 μ l INSP002V-5'-F (10 μ M，正向引物)、1.5 μ l INSP002V-5'-R (10 μ M，反向引物)、0.75 μ l Platinum Pfx 和 135 ng IMAGE: 4558384 质粒 cDNA。扩增条件是：94 $^{\circ}$ C 循环 1 次 2 分钟，94 $^{\circ}$ C 循环 30 次 15 秒，68 $^{\circ}$ C 1 分钟；68 $^{\circ}$ C 循环 1 次 7 分钟。在同样条件下进行第二次 PCR 反应，扩增 INSP002 3'端(87 bp 插入片段的下游)，但引物为：正向：10 INSP002V-3'-F，反向：INSP002V-3'-R。

在 1X TAE 缓冲液(Invitrogen)中的 0.8%琼脂糖凝胶上看到扩增产物，用 Wizard PCR Preps DNA 纯化系统(Promega)从凝胶纯化得到以预定分子量(第一次和第二次 PCR 分别是 520 bp 和 448 bp)迁移的 PCR 产物。PCR 产物在 50 微升无菌水中洗脱，用 Eppendorf BO 光度计测定 DNA 浓度。各取 50 ng 纯15 化 PCR 产物作为嵌套式 PCR 的模板，反应在 50 微升体积中进行，含有：0.5 μ l 10X Platinum Pfx 缓冲液、1.5 μ l dNTPs (10mM)、1 μ l 1 硫酸镁(50 mM)、1.5 μ l INSP002V-5'nest-F (10 μ M，嵌套正向引物)、1.5 μ l INSP002V-3'nest-F (10 μ M，嵌套反向引物)。反应混合物在 95 $^{\circ}$ C 加热 3 分钟，加入 0.75 μ l Platinum Pfx 聚合酶。扩增条件如下：94 $^{\circ}$ C 循环 1 次 2 分钟，94 $^{\circ}$ C 循环 30 次 15 秒，61 $^{\circ}$ C20 30 秒和 68 $^{\circ}$ C 1 分钟；68 $^{\circ}$ C 循环 1 次 7 分钟。用 Wizard PCR Preps DNA 纯化系统(Promega)从凝胶纯化得到以预定分子量 719 bp 迁移的 PCR 产物，在 50 微升无菌水中洗脱。参照 1.4 节所述，取 4 μ l 纯化 PCR 产物与 pCR4 blunt TOPO 载体连接。参照 1.5 节所述，通过集落 PCR 测试插入片段的氨比西林抗性集落，引物用 T3 和 T7。获得预定 PCR 产物大小(由于多克隆位点而得到 719 bp+25 106 bp)的集落在 5 毫升含氨比西林(100 μ g/ml)的 LB 中生长，37 $^{\circ}$ C 220 rpm 摇动培育一晚。从 5 毫升培养基中制备小量制备质粒 DNA，参照 1.6 节所述，用 T3 和 T7 引物测序。其中一个所得克隆的序列和对应的质粒图谱(pCR4 blunt TOPO-INSP002V)分别示于图 12 和图 13 中。克隆序列的翻译显示，相对于预测的 INSP002 序列，INSP002V 含有 2 个氨基酸缺失(V 107 和 Q108)和一个氨基酸取代(F110L)。图 14 所示为 INSP002 预测和 INSP002V 的核苷酸30

和氨基酸序列的比对。

当与 INSP002 预测序列相比时，外显子 1 和 2 之间的剪接受体使用 6 bp 上游供体位点。这导致 2 个氨基酸(ValGlu)缺失。使用的剪接受体与 INSP002 预测序列所用的相同，但受体之后发生 2 个测序误差。这导致 1 个氨基酸取代(Phe -> Leu)。

实施例 5 在 HEK293/EBNA 细胞中表达 INSP002V 的质粒的构建

通过 DNA 测序鉴定一个 pCR4 blunt-TOPO 克隆含有 INSP002V 全编码序列(ORF, 可读框) (图 13), 再运用 Gateway™ 克隆方法(Invitrogen)以该克隆将插入片段亚克隆到哺乳动物细胞表达载体 pEAK12d(图 15)中。

(i) 构建 Gateway 相容的与符合读框的 6HIS 标记序列融合的 IPAAA44548 ORF

Gateway 克隆方法的第一阶段涉及二步 PCR 反应, 该反应产生在 5' 末端有 attB1 重组位点和 Kozak 序列, 在 3' 末端有用编码符合读框的 6 组氨酸 (6HIS) 标记、终止密码子和 attB2 重组位点(Gateway 相容的 cDNA)的 INSP002V ORF。第一次 PCR 反应(在 50 微升终体积中)含有: 25 ng pCR4 blunt-TOPO-INSP002V(质粒 13075, 图 13)、1.5 μ l dNTPs (10mM)、5 μ l 10X Pfx 聚合酶缓冲液、1.0 μ l 硫酸镁(50mM)、基因特异性引物各 0.5 μ l (100 μ M) (INSP002V-EX1 和 INSP002V EX2)和 0.5 μ l Platinum Pfx DNA 聚合酶 (Invitrogen)。在 95°C 首次变性步骤进行 PCR 反应 2 分钟, 然后 94°C 循环 12 次 15 秒, 以及 68°C 30 秒。按照生产商的指示, 用 Wizard PCR prep DNA 纯化系统(Promega)直接在反应混合物中纯化 PCR 产物。第二次 PCR 反应(在 50 微升终体积中)含有: 10 μ l 纯化 PCR 产物、1.5 μ l dNTPs (10 mM)、1.0 μ l 硫酸镁(50mM)、5 μ l 10X Pfx 聚合酶缓冲液、Gateway 转化引物各 0.5 μ l (100 μ M) (正向引物: GCP, 反向引物: GCP)和 0.5 μ l of Platinum Pfx DNA 聚合酶。第二次 PCR 反应的条件是: 95°C 1 分钟; 94°C 循环 4 次 15 秒; 45°C 30 秒以及 68°C 3.5 分钟; 94°C 循环 25 次 15 秒; 55°C 30 秒和 68°C 3.5 分钟。参照上述方法纯化 PCR 产物。

(ii) Gateway 相容的 INSP002V ORF 亚克隆到 Gateway 入门载体 pDONR201 和表达载体 pEAK12d

Gateway 克隆方法的第二阶段按如下方法将 Gateway 修饰的 PCR 产物亚克隆到 Gateway 入门载体 pDONR201(Invitrogen, 图 16): 将 5 μ l 纯化的 PCR 产物与 1.5 μ l pDONR201 载体(0.1 μ g/ μ l)、2 μ l BP 缓冲液和 1.5 μ l BP 克隆酶混合物在室温培育 1 小时。加入蛋白酶 K(2 μ g)终止反应, 37°C 培育多 10 分钟。取该反应的一等分试样(2 μ l), 用 Biorad Gene Pulser 电穿孔仪通过电穿孔转化到大肠杆菌 DH10B 细胞中。将转化物铺在 LB-卡那霉素板上。用 Wizard Plus SV Minipreps 试剂盒(Promega cat. no. 1460)在 1-4 个所得集落中制备小量制备质粒 DNA, 在重组反应中使用 1.5 μ l 该质粒洗脱液, 该反应的终体积为 10 毫升, 含有 1.5 μ l pEAK12d 载体(图 9) (0.1 μ g / μ l)、2 μ l LR 缓冲液和 1.5 μ l LR 克隆酶(Invitrogen)。混合物室温培育 1 小时, 加入蛋白酶 K(2 μ g)终止培育, 再在 37°C 培育 10 分钟。取该反应的一等分试样(1 μ l), 通过电穿孔技术转化到大肠杆菌 DH10B 细胞中。

参照上述方法, 进行集落 PCR 鉴定含有正确插入片段的克隆, 但 PCR 的引物用 pEAK12d (正向引物: pEAK12d F, 反向引物: pEAK12d R)。用 Qiaprep Turbo 9600 自动化系统(Qiagen)或人手以 Wizard Plus SV Minipreps 试剂盒(Promega)在含有正确插入片段的克隆中分离出小量制备质粒 DNA, 并用正向引物 pEAK12d F 和反向引物 pEAK12d R 确认序列。

从 500 毫升序列已被确认的克隆培养基中制备用氯化铯梯度纯化的质粒 pEAK12d-INSP002V-6HIS (质粒 ID 号: 13227, 图 17)的大量制备 DNA(Sambrook J.等, Molecular Cloning, a Laboratory Manual, 第二版, 1989, Cold Spring Harbor Laboratory Press), 重悬于 1 μ g/ μ l 无菌水中, -20°C 储存。

(iii) 表达质粒 pEAK12d 的构建

载体 pEAK12d 是哺乳动物细胞表达载体 pEAK12 的 Gateway 克隆系统相容译本(购自 Edge Biosystems), 其中感兴趣的 cDNA 在人 EF1 α 启动子的控制下表达。pEAK12d 按以下方法产生:

用限制酶 HindIII 和 NotI 消化 pEAK12, 用克列诺(New England Biolabs)平端, 并用牛小肠碱性磷酸盐(Roche)脱磷酸。在脱磷酸作用之后, 使载体与

含有 ccdB 基因旁侧的 AttR 重组位点和氯霉素抗性的平端 Gateway 读框盒 C (Gateway 载体转化系统, Invitrogen cat no. 11828-019)连接, 再转化到大肠杆菌 DB3.1 细胞(可使含 ccdB 基因的载体增殖)中。用 Wizard Plus SV Minipreps 试剂盒(Promega)使小量制备 DNA 与几个所得集落分离, 以 AseI/EcoRI 消化
5 鉴定集落, 得到 670 bp 片段, 这表示读框盒以正确方向插入。得到的质粒命名为 pEAK12d(图 15)。

实施例 6 INSP002-SV-6HIS-V1 (质粒号: 13227)在哺乳动物细胞中的表达和纯化

表达埃-巴二氏病毒核抗原的人胚胎肾 293 细胞(HEK293-EBNA, Invitrogen)保持在无 Ex-cell VPRO 血清的培养基悬浮液中(种子储备, 维持培养基, JRH)。转染前 16 至 20 小时(第 1 天), 将细胞接种在 2x T225 烧瓶中(每个烧瓶 50 毫升, 接种在含有 2% FBS 移种培养基(JRH)的 DMEM/F12 (1: 1)中, 密度为 2×10^5 个细胞/毫升)。第二天(转染第 0 天), 用 JetPEI™ 试剂进行转染($2 \mu\text{l}/\mu\text{g}$ 质粒 DNA, PolyPlus-transfection 转染试剂)。每个烧瓶中, 113 μg
15 质粒(No. 13227)与 2.3 μg GFP (荧光报导基因)共转染。再将转染混合物加入 2x T225 烧瓶中, 37°C ($5\% \text{CO}_2$)培育 6 天。为了提高获得更多物质的可能性, 我们在另外 2 个烧瓶中重复这一程序, 总共得到 200 毫升。在第 1 和第 6 天定性荧光检验确认为阳性转染(Axiovert 10 Zeiss)。

在第 6 天(收获天), 将四个烧瓶中的上清液(200 ml)合并, 离心(4°C ,
20 400g), 并置于带有独特鉴定器的器皿内。

保留一等分试样(500 μl)对 6His 标记蛋白作 QC(内部生物加工 QC)。

纯化方法

含有 C 端带 6His 标记的重组蛋白质的 200 毫升培养基样品用冷冻缓冲液 A(50 mM 磷酸二氢钠; 600 mM 氯化钠; 8.7%(重量/体积)甘油, pH 7.5)稀
25 释至终体积 400 毫升。样品经 0.22 μm 无菌过滤器(Millipore, 500 毫升过滤装置)过滤, 4°C 储存于无菌方形培养瓶中(Nalgene)。

4°C 下用连接到样品自动装载机(Labomatic)的 VISION 工作站(Applied Biosystems)进行纯化。纯化程序由两个连续步骤组成: 在装有镍离子的 Poros
30 20 MC (Applied Biosystems)柱(4.6x50 mm, 0.83 ml)上进行金属亲和力层析,

再在 Sephadex G-25 培养(Amersham Pharmacia)柱(1.0x10 cm)上进行凝胶过滤。

第一个层析步骤的金属亲和柱用 30 柱体积的 EDTA 溶液(100 mM EDTA; 1 M NaCl; pH 8.0)再生, 用 15 柱体积的 100 mM 硫酸镍溶液洗涤重新装载镍离子, 用 10 柱体积的缓冲液 A 和 7 柱体积的缓冲液 B(50 mM 磷酸二氢钠; 600 mM 氯化钠; 8.7%(重量/体积)400 mM 甘油; 咪唑, pH 7.5)洗涤, 最后用 15 柱体积含 15 mM 咪唑的缓冲液 A 平衡。200 毫升样品用 Labomatic 的样品装载机转移到 200 毫升定量环, 然后以 10 ml/min 的速率装到镍金属亲和柱上。一旦要将 400 毫升样品装在柱上, 重复转移和装载离子的步骤。装载离子的步骤结束后, 柱再用 12 柱体积的缓冲液 A 和 28 柱体积含 20 mM 咪唑的缓冲液 A 洗涤。在 20 mM 咪唑洗涤期间, 附着比较松散的污染蛋白质从柱中洗脱出来。重组的 His 标记蛋白质最后用 10 柱体积的缓冲液 B 洗脱, 流速为 2 ml/min, 洗脱出来的蛋白质收集得到 1.6 毫升馏分。

第二层析步骤的 Sephadex G-25 凝胶过滤柱用 2 毫升缓冲液 D (1.137 M 氯化钠; 2.7 mM 氯化钾; 1.5 mM 磷酸二氢钾; 8 mM 磷酸二氢钠; pH 7.2)再生, 然后用 4 柱体积的缓冲液 C (137 mM 氯化钠; 2.7 mM 氯化钾; 1.5 mM 磷酸二氢钾; 8 mM 磷酸二氢钠; 20 % (重量/体积)甘油; pH 7.4)平衡。从镍柱洗脱出来的峰值馏分自动通过装在 Sephadex G-25 柱并连接到 VISION 的整合型样品装载机, 蛋白质用缓冲液 C 洗脱, 流速为 2 ml/min。去盐后样品回收得到 2.2 毫升馏分。该馏分经过 0.22 μ m 无菌离心过滤器(Millipore)过滤, 冻存于-80°C。取样品的一等分试样以使用抗 His 抗体的 SDS-PAGE (4-12 % NuPAGE 凝胶; Novex)的 Western 印迹进行分析。

在电穿孔之后, 在 4°C、290 mA、1 小时内将蛋白质从凝胶电转移到硝基纤维素膜上。室温下用 5%在缓冲液 E(137 mM 氯化钠; 2.7 mM 氯化钾; 1.5 mM 磷酸二氢钾; 8 mM 磷酸二氢钠; 0.1 % Tween 20, pH 7.4)中的奶粉封闭纤维素膜 1 小时, 再于 4°C 在 2.5%在缓冲液 E 中的奶粉中与 2 种兔多克隆抗 His 抗体(G-18 和 H-15, 各 0.2 μ g/ml; Santa Cruz)混合物培养过夜。室温再培养 1 小时之后, 膜用缓冲液 E(3x10 分钟)洗涤, 再与结合 HRP 的第二抗兔抗体(DAKO, HRP 0399)室温培养 2 小时, 第二抗体用含 2.5%奶粉的缓冲液 E 稀释至 1/3000。用缓冲液 E(3x10 分钟)洗涤之后, 膜用 ECL 试剂盒

(Amersham Pharmacia)显影 1 分钟。然后，将膜与超膜(Hyperfilm) (Amersham Pharmacia)接触，使超膜显影，对 Western 印迹影像作目测分析。

实施例 7 从心脏克隆全长 INSP002

5 按如下方法从心脏 cDNA 克隆 INSP002 的全长编码序列

(i) 心脏 cDNA 模板的产生

人心脏总 RNA 购自 Clontech。用 Agilent 2100 生物分析仪分析 RNA 的质量和浓度。

对于 cDNA 的合成，12 μ l 反应混合物含有：1 μ l oligo (dT)₁₅ 引物 (500
10 μ g/ml, Promega cat. no. C 1101)、2 μ g 总 RNA、1 μ l dNTPs (10 mM)。混合物在 65°C 加热 5 分钟，然后置于冰上冷冻。加入下列试剂：4 μ l 5X first strand buffer、2 μ l DTT (0.1 M)、1 μ l RNaseOut 重组核糖核酸酶抑制剂(40 单位/微升, Promega, cat. no. N 2511), 加入 1 μ l (200 单位)Superscript II (Invitrogen cat. no. 18064-014)之前在 42°C 培育 2 分钟。该混合物在 42°C 培育 50 分钟，再
15 在 70°C 加热 15 分钟。为了去掉 RNA 模板，加入 1 μ l (2 单位)大肠杆菌核糖核酸酶 H(Invitrogen cat. no.18021-014)，反应混合物再在 37°C 培育 20 分钟。用无菌水将终反应混合物稀释至 200 微升，-80°C 储存。

(ii) 通过 PCR 技术克隆 INSP002 的全长编码序列

20 在 50 微升 PCR 反应混合物中通过 PCR 技术从人心脏 cDNA 中克隆 INSP002 的全长编码序列，所述反应混合物含有 5 μ l 心脏 cDNA、5 μ l 10X Pfx 缓冲液、1.5 μ l dNTPs (10 mM)、1 μ l 硫酸镁(50 mM)、1.5 μ l 基因特异性正向引物 INSP002-FL-F (10 μ M)、1.5 μ l 基因特异性反向引物 INSP002-FL-R (10 μ M)和 0.5 μ l Platinum Pfx DNA 聚合酶(Invitrogen)。循环条件是：94°C 循环
25 4 次 4 分钟；94°C 循环 35 次 15 秒；55°C 30 秒；68°C 1 分钟；68°C 循环 1 次 10 分钟，然后 4°C 保持循环。

在 1X TAE 缓冲液(Invitrogen)中的 0.8%琼脂糖凝胶上看到扩增产物，用 Qiagen MinElute 凝胶抽提试剂盒(Qiagen)从凝胶纯化得到以预定分子量(589 bp)迁移的 PCR 产物。PCR 产物在 50 微升 10 mM Tris-HCl pH 8.5 中洗脱，
30 参照上述方法(1.4 节)，亚克隆到 pCR4 blunt TOPO 载体中。含有正确大小插

入片段(由于多克隆位点而得到 589 bp+ 106 bp)的集落在 5 毫升含氨比西林 (100 µg/ml)的 L-Broth (LB)中生长, 37°C 220 rpm 摇动培育一晚。按照生产商的指示, 用 Qiaprep Turbo 9600 自动化系统(Qiagen)或 Wizard Plus SV Minipreps 试剂盒(Promega cat. no. 1460)从 5 毫升培养基中制备小量制备质粒 DNA, 参照 1.6 节所述, 用 T3 和 T7 引物(表 III)对 200-500 ng 小量制备 DNA 测序。图 18 所示为克隆序列。图 19 所示为获得的质粒 pCR4-blunt TOPO-INSP002FL (质粒号: 13514)的图谱。

表I 人cDNA文库

| 文库 | 组织/细胞来源 | 载体 | 宿主菌 | 供应商 | Cat. no. |
|----|------------------------|---------|---------------|------------|----------|
| 1 | 人胎儿脑 | Zap II | XL1-Blue MRF' | Stratagene | 936206 |
| 2 | 人卵巢 | GT10 | LE392 | Clontech | HL1098a |
| 3 | 人垂体 | GT10 | LE392 | Clontech | HL1097a |
| 4 | 人胎盘 | GT11 | LE392 | Clontech | HL1075b |
| 5 | 人睾丸 | GT11 | LE392 | Clontech | HL1010b |
| 6 | Human sustanta nigra | GT10 | LE392 | in house | |
| 7 | 人胎儿脑 | GT10 | LE392 | in house | |
| 8 | 人脑皮质 | GT10 | LE392 | in house | |
| 9 | 人结肠 | GT10 | LE392 | Clontech | HL1034a |
| 10 | 人胎儿脑 | GT10 | LE392 | Clontech | HL1065a |
| 11 | 人胎儿肺 | GT10 | LE392 | Clontech | HL1072a |
| 12 | 人胎儿肾 | GT10 | LE392 | Clontech | HL1071a |
| 13 | 人胎儿肝脏 | GT10 | LE392 | Clontech | HL1064a |
| 14 | 人骨髓 | GT10 | LE392 | Clontech | HL1058a |
| 15 | 人外周血单核细胞 | GT10 | LE392 | Clontech | HL1050a |
| 16 | 人胎盘 | GT10 | LE392 | in house | |
| 17 | 人SHSYSY | GT10 | LE392 | in house | |
| 18 | 人U373细胞系 | GT10 | LE392 | in house | |
| 19 | 人CFPoc-1细胞系 | Uni Zap | XL1-Blue MRF' | Stratagene | 936206 |
| 20 | 人视网膜 | GT10 | LE392 | Clontech | HL1132a |
| 21 | 人膀胱 | GT10 | LE392 | in house | |
| 22 | 人血小板 | Uni Zap | XL1-Blue MRF' | in house | |
| 23 | 人成神经细胞瘤Kan + TS | GT10 | LE392 | in house | |
| 24 | 人支气管平滑肌 | GT10 | LE392 | in house | |
| 25 | 人支气管平滑肌 | GT10 | LE392 | in house | |
| 26 | 人胸腺 | GT10 | LE392 | Clontech | HL1127a |
| 27 | 人脾5'延伸 | GT11 | LE392 | Clontech | HL1134b |
| 28 | 人外周血单核细胞 | GT10 | LE392 | Clontech | HL1050a |
| 29 | 人睾丸 | GT10 | LE392 | Clontech | HL1065a |
| 30 | 人胎儿脑 | GT10 | LE392 | Clontech | HL1065a |
| 31 | Human substancia Nigra | GT10 | LE392 | Clontech | HL1093a |
| 32 | 人胎盘#11 | GT11 | LE392 | Clontech | HL1075b |
| 33 | 人胎儿脑 | GT10 | LE392 | Clontech | custom |
| 34 | 人胎盘#59 | GT10 | LE392 | Clontech | HL5014a |
| 35 | 人垂体 | GT10 | LE392 | Clontech | HL1097a |

| | | | | | |
|----|-------------|------------|---------------|------------|---------|
| 36 | 人胰腺 #63 | Uni Zap XR | XL1-Blue MRF' | Stratagene | 937208 |
| 37 | 人胎盘 #19 | GT11 | LE392 | Clontech | HL1008 |
| 38 | 人肝5'延伸 | GT11 | LE392 | Clontech | HL1115b |
| 39 | 人子宫 | Zap-CMV XR | XL1-Blue MRF' | Stratagene | 980207 |
| 40 | 人肾大插入cDNA文库 | TriplEx2 | XL1-Blue | Clontech | HL5507u |

表 II INSP002 克隆引物

| 引物 | 序列 (5'-3') |
|-------------------------|---|
| INSP002-CP1 | CTC AGC CAT ACG CCT CCG AA |
| INSP002-CP2 | GCT GAG CTG CCA GTG AGA CA |
| INSP002V-5'-F(正向) | ACC TGG AAG GAA GCG ACT GCA CTG A |
| INSP002V-5'-R(反向) | GCA GCC GGG CCG GGA GAG AAC GAA GGG CAC AGC CTT A |
| INSP002V-3'-F(正向) | AGG CTG TGC CCT TCG TTC TCT CCC GGC CCG GCT GCT C |
| INSP002V-3'-R(反向) | ACT CCA GGA CGG GCA CTG TGT CTA C |
| INSP002V-5'nest-F(嵌套正向) | GTC GAC TGC TAG TGA CCT TGA G |
| INSP002V-3'nest-R(嵌套反向) | ACA TCA TCC AGG TCC ACG TCT T |

5

表 III INSP002 亚克隆和测序的引物

| 引物 | 序列 (5'-3') |
|--------------|--|
| SP6 | ATT TAG GTG ACA CTA TAG |
| T7 | TAA TAC GAC TCA CTA TAG GG |
| T3 | ATT AAC CCT CAC TAA AGG GA |
| M13F | TGT AAA ACG ACG GCC AGT |
| pEAK12-F(正向) | GCC AGC TTG GCA CTT GAT GT |
| pEAK12-R(反向) | GAT GGA GGT GGA CGT GTC AG |
| INSP002V-EX1 | AA GCA GGC TTC GCC ACC ATG CTC CTT GGC CAG CTA TC |
| INSP002V-EX2 | <i>GTG ATG GTG ATG GTG</i> TGC TTT TGG GCT GCA GTG AC |
| GCP(正向) | G GGG ACA AGT TTG TAC AAA AAA GCA GGC TTC <u>GCC ACC</u> |
| GCP(反向) | GGG GAC CAC TTT GTA CAA GAA AGC TGG GTT TCA <i>ATG GTG ATG GTG</i> <i>ATG GTG</i> |

10

表 IV 克隆全长 INSP002 的 PCR 引物

| 引物 | 序列 (5'-3') |
|--------------|----------------------------|
| INSP002-FL-F | GAT GCT CCT TGG CCA GCT AT |
| INSP002-FL-R | CCA TCC ACG ATG CTC AGT TC |

下划线序列 = Kozak 序列

粗体 = 终止密码子

斜线序列 = His 标记

15

序列表

注意：对于外显子-外显子连接编码的氨基酸,它们将被指为 more 5'外显子。

SEQ ID NO:1 (INSP002 -核苷酸序列外显子 1)

```

1 AAATGCCTCC CAGGCTATCC AGGAGGGGCC AAGAGATTAA AAGCAGGTTC
51 AGAAGGCTCA GATGCCACTC ACCAGACAGC AGGGTCGACT GCTAGTGACC
5 101 TTGAGCCCAG TCCGGACAGA CAGACAGGCA GACAGACGCA CGGACAAGCA
151 GATGCTCCTT GGCCAGCTAT CCACTCTTCT GTGCCTGCTT AGCGGGGCCC
201 TGCCTACAGG CTCAGGGAGG CCTGAACCCC AGTCTCCTCG ACCTCAGTCC
251 TGGGCTGCAG CCAATCAGAC CTGGGCTCTG GGCCAGGGG CCCTGCCCCC
301 ACTGGTGCCA GCTTCTGCC TTGGGAGCTG GAAGGCCTTC TTGGGCCTGC
10 351 AGAAAGCCAG GCAGCTGGGG ATGGGCAGGC TGCAGCGTGG GCAAGACGAG
401 GTGGCTGCTG TGA CTCTGCC GCTGAACCCT CAGGAAGTGA TCCAGGGGAT
451 GTGTAAGGCT GTGCCCTTCG TTCAG

```

SEQ ID NO:2 (INSP002 -蛋白质序列外显子 1)

```

1 MLLGQLSTLL CLLSGALPTG SGRPEPQSPR PQSWAAANQT WALGPGALPP
15 51 LVPASALGSW KAFLGLQKAR QLGMRLQRG QDEVAAVTLP LNPQEVIQGM
101 CKAVPFVQ

```

SEQ ID NO:3 (INSP002 -核苷酸序列外显子 2)

```

1 GTGTTCTCCC GGCCCGGCTG CTCAGCCATA CGCCTCCGAA ATCATCTGTG
20 51 CTTTGGTCAT TGCTCCTCTC TCTACATCCC TGGCTCGGAC CCCACCCAC
101 TAGTCCTGTG CAACAGCTGT ATGCCTGCTC GCAAGCGTTG GGCACCCGTG
151 GTCCCTGTGGT GTCTCACTGG CAGCTCAGCC TCCCGTCGAC GGGTGAAGAT
201 ATCCACCATG CTGATCGAGG GGTGTCCTG CAGCCCAAAA GCATGA

```

25 SEQ ID NO:4 (INSP002 -蛋白质序列外显子 2)

```

1 VFSRPGCSAI RLRNHLFCFH CSSLYIPGSD PTPLVLCNSC MPARKRWAPV
51 VLWCLTGSSA SRRRVKISTM LIEGCHCSPK A

```

SEQ ID NO:5 (INSP002 -核苷酸序列)

```

30 1 AAATGCCTCC CAGGCTATCC AGGAGGGGCC AAGAGATTAA AAGCAGGTTC
51 AGAAGGCTCA GATGCCACTC ACCAGACAGC AGGGTCGACT GCTAGTGACC
101 TTGAGCCCAG TCCGGACAGA CAGACAGGCA GACAGACGCA CGGACAAGCA
151 GATGCTCCTT GGCCAGCTAT CCACTCTTCT GTGCCTGCTT AGCGGGGCCC
201 TGCCTACAGG CTCAGGGAGG CCTGAACCCC AGTCTCCTCG ACCTCAGTCC

```

251 TGGGCTGCAG CCAATCAGAC CTGGGCTCTG GGCCAGGGG CCCTGCCCC
 301 ACTGGTGCCA GCTTCTGCCC TTGGGAGCTG GAAGGCCTTC TTGGGCTGC
 351 AGAAAGCCAG GCAGCTGGGG ATGGGCAGGC TGCAGCGTGG GCAAGACGAG
 401 GTGGCTGCTG TGA CTCTGCC GCTGAACCCT CAGGAAGTGA TCCAGGGGAT
 5 451 GTGTAAGGCT GTGCCCTTCG TTCAGGTGTT CTCCCGGCC GGCTGCTCAG
 501 CCATACGCCT CCGAAATCAT CTGTGCTTTG GTCATTGCTC CTCTCTCTAC
 551 ATCCCTGGCT CGGACCCAC CCCACTAGTC CTGTGCAACA GCTGTATGCC
 601 TGCTCGCAAG CGTTGGGCAC CCGTGGTCCT GTGGTGTCTC ACTGGCAGCT
 651 CAGCTCCCG TCGACGGGTG AAGATATCCA CCATGCTGAT CGAGGGGTGT
 10 701 CACTGCAGCC CAAAAGCATG A

SEQ ID NO:6 (INSP002 –蛋白质序列)

1 MLLGQLSTLL CLLSGALPTG SGRPEPQSPR PQSWAANQT WALGPGALPP
 15 51 LVPASALGSW KAFGLQKAR QLGMGRLQRG QDEVAAVTLP LNPQEVIOGM
 101 CKAVPFVQVF SRPGCSAIRL RNHLCFGHCS SLYIPGSDPT PLVLCNSCMP
 151 ARKRWAPVVL WCLTGSSASR RRVKISTMLI EGCHCSPKA

SEQ ID NO: 7 (INSP002 – 没有信号肽的蛋白质序列外显子 1)

20 1 RPEPQSPRPQ SWAANQTWA LGPGALPLV PASALGSWKA FLGLQKARQL
 51 GMGRLQRGQD EVAAVTLPLN PQEVIOGMCK AVPFVQ

SEQ ID NO: 8 (INSP002 –没有信号肽的蛋白质序列)

25 1 RPEPQSPRPQ SWAANQTWA LGPGALPLV PASALGSWKA FLGLQKARQL
 51 GMGRLQRGQD EVAAVTLPLN PQEVIOGMCK AVPFVQVFSR PGCSAIRLRN
 101 HLCFGHCSSL YIPGSDPTPL VLCNSCMPAR KRWAPVVLWC LTGSSASRRR
 151 VKISTMLIEG CHCSPKA

30

SEQ ID NO: 9 (编码没有信号肽的 INSP002 蛋白质序列的核苷酸序列 – 成熟 INSP002 多肽)

35 1 AGGCCTGAAC CCCAGTCTCC TCGACCTCAG TCCTGGGCTG CAGCCAATCA GACCTGGGCT
 61 CTGGGCCAG GGGCCCTGCC CCCACTGGTG CCAGCTTCTG CCCTTGGGAG CTGGAAGGCC
 121 TTCTTGGGCC TGCAGAAAGC CAGGCAGCTG GGGATGGGCA GGCTGCAGCG TGGGCAAGAC
 181 GAGGTGGCTG CTGTGACTCT GCCGCTGAAC CCTCAGGAAG TGATCCAGGG GATGTGTAAG
 241 GCTGTGCCCT TCGTTCAGGT GTTCTCCCGG CCCGGCTGCT CAGCCATACG CCTCCGAAAT

301 CATCTGTGCT TTGGTCATTG CTCCTCTCTC TACATCCCTG GCTCGGACCC CACCCCACTA
 361 GTCCTGTGCA ACAGCTGTAT GCCTGCTCGC AAGCGTTGGG CACCCGTGGT CCTGTGGTGT
 421 CTCACTGGCA GCTCAGCCTC CCGTCGACGG GTGAAGATAT CCACCATGCT GATCGAGGGG
 461 TGTCCTGCA GCCCAAAGC ATGA

5

SEQ ID NO: 10 (编码没有信号肽的 INSP002 外显子 1 蛋白质序列的核苷酸序列 – 成熟 INSP002 外显子 1 多肽)

1 AGGCCTGAAC CCCAGTCTCC TCGACCTCAG TCCTGGGCTG CAGCCAATCA GACCTGGGCT
 61 CTGGGCCAG GGGCCCTGCC CCCACTGGTG CCAGCTTCTG CCCTTGGGAG CTGGAAGGCC
 10 121 TTCTTGGGCC TGCAGAAAGC CAGGCAGCTG GGGATGGGCA GGCTGCAGCG TGGCAAGAC
 181 GAGGTGGCTG CTGTGACTCT GCCGCTGAAC CCTCAGGAAG TGATCCAGGG GATGTGTAAG
 241 GCTGTGCCCT TCGTTCAG

SEQ ID NO: 11 (编码没有 5'非翻译区的 INSP002 蛋白质的核苷酸序列)

15 1 ATGCTCCTT GGCCAGCTAT CCACTCTTCT GTGCCTGCTT AGCGGGGCCCT
 51 GCCTACAGG CTCAGGGAGG CCTGAACCCC AGTCTCCTCG ACCTCAGTCCT
 101 GGGCTGCAG CCAATCAGAC CTGGGCTCTG GGCCAGGGG CCCTGCCCCCA
 151 CTGGTGCCA GCTTCTGCCC TTGGGAGCTG GAAGGCCTTC TTGGGCTGCA
 201 GAAAGCCAG GCAGCTGGGG ATGGGCAGGC TGCAGCGTGG GCAAGACGAGG
 20 251 TGGCTGCTG TGA CTCTGCC GCTGAACCCCT CAGGAAGTGA TCCAGGGGATG
 301 TGTAAGGCT GTGCCCTTCG TTCAGGTGTT CTCCCAGGCC GGCTGCTCAGC
 351 CATA CGCCT CCGAAATCAT CTGTGCTTTG GTCATTGCTC CTCTCTTACA
 401 TCCCTGGCT CGGACCCAC CCCACTAGTC CTGTGCAACA GCTGTATGCCT
 451 GCTCGCAAG CGTTGGGCAC CCGTGGTCTT GTGGTGTCTC ACTGGCAGCTC
 25 501 AGCCTCCCG TCGACGGGTG AAGATATCCA CCATGCTGAT CGAGGGGTGTC
 551 ACTGCAGCC CAAAAGCATG A

SEQ ID NO: 12 (编码没有 5'非翻译区的 INSP002 外显子 1 蛋白质的核苷酸序列)

30 1 ATGCTCCTT GGCCAGCTAT CCACTCTTCT GTGCCTGCTT AGCGGGGCCCT
 51 GCCTACAGG CTCAGGGAGG CCTGAACCCC AGTCTCCTCG ACCTCAGTCCT
 101 GGGCTGCAG CCAATCAGAC CTGGGCTCTG GGCCAGGGG CCCTGCCCCCA
 151 CTGGTGCCA GCTTCTGCCC TTGGGAGCTG GAAGGCCTTC TTGGGCTGCA
 35 201 GAAAGCCAG GCAGCTGGGG ATGGGCAGGC TGCAGCGTGG GCAAGACGAGG
 251 TGGCTGCTG TGA CTCTGCC GCTGAACCCCT CAGGAAGTGA TCCAGGGGATG
 301 TGTAAGGCT GTGCCCTTCG TTCAG

SEQ ID NO: 13 (编码变体 INSP002 多肽的核苷酸序列)

1 GTCGACTGCT AGTGACCTTG AGCCCAGTCC GGACAGACAG ACAGGCAGAC AGACGCACGG
 61 ACAAGCAGAT GCTCCTTGGC CAGCTATCCA CTCTTCTGTG CCTGCTTAGC GGGGCCCTGC
 5 121 CTACAGGCTC AGGGAGGCCT GAACCCAGT CTCCTCGACC TCAGTCCTGG GCTGCAGCCA
 181 ATCAGACCTG GGCTCTGGGC CCAGGGGCC TGCCCCACT GGTGCCAGCT TCTGCCCTTG
 241 GGAGCTGGAA GGCCTTCTTG GGCCTGCAGA AAGCCAGGCA GCTGGGGATG GGCAGGCTGC
 301 AGCGTGGGCA AGACGAGGTG GCTGTGTGTA CTCTGCCGCT GAACCTCAG GAAGTGATCC
 361 AGGGGATGTG TAAGGCTGTG CCCTTCGTTT TCTCCCGCC CGGCTGCTCA GCCATACGCC
 10 421 TCCGAAATCA TCTGTGCTTT GGTCAATGCT CCTCTCTCTA CATCCCTGGC TCGGACCCCA
 481 CCCCACTAGT CCTGTGCAAC AGCTGTATGC CTGCTGCAA GCGTTGGGCA CCCGTGGTCC
 541 TGTGGTGTCT CACTGGCAGC TCAGCCTCCC GTCGACGGGT GAAGATATCC ACCATGCTGA
 601 TCGAGGGGTG TCACTGCAGC CAAAAGCAT GAACTGAGCA TCTGGATGGG TGCACGGAGA
 661 CACGCACCTT GGAGAAATGA GGGGAGATGG ACCAAGAAAG ACGTGGACCT GGATGATGT
 15

SEQ ID NO: 14 (变体 INSP002 多肽)

1 MLLGQLSTLL CLLSGALPTG SGRPEPQSPR PQSWAAANQT WALGPGALPP LVPASALGSW
 61 KAFLGLQKAR QLGMRLQRG QDEVAAVTLP LNPQEVIQGM CKAVPFVLSR PGCSAIRLRN
 121 HLCFGHCSSL YIPGSDPTPL VLCNSCMPTLR KRWAPVVLWC LTGSSASRRR VKISTMLIEG
 20 161 CHCSPKA

SEQ ID NO: 15 (编码没有 5'非翻译区的变体 INSP002 多肽的核苷酸序列)

1 ATGCTCCTTG GCCAGTATC CACTCTTCTG TGCCTGCTTA GCGGGGCCCT GCCTACAGGC
 61 TCAGGGAGGC CTGAACCCCA GTCTCCTCGA CCTCAGTCCT GGGCTGCAGC CAATCAGACC
 25 121 TGGGCTCTGG GCCCAGGGC CCTGCCCCA CTGGTGCCAG CTTCTGCCCT TGGGAGCTGG
 181 AAGGCCTTCT TGGGCCTGCA GAAAGCCAGG CAGCTGGGA TGGCAGGCT GCAGCGTGGG
 241 CAAGACGAGG TGCTGTGTG GACTCTGCCG CTGAACCTC AGGAAGTGAT CCAGGGGATG
 301 TGTAAGGCTG TGCCCTTCGT TCTCTCCCGG CCCGGCTGCT CAGCCATACG CCTCCGAAAT
 361 CATCTGTGCT TTGGTCATG CTCTCTCTC TACATCCCTG GCTCGGACCC CACCCCACTA
 30 421 GTCCTGTGCA ACAGCTGTAT GCCTGCTCGC AAGCGTTGGG CACCCGTGGT CCTGTGGTGT
 481 CTCCTGGCA GCTCAGCCTC CCGTCGACGG GTGAAGATAT CCACCATGCT GATCGAGGGG
 541 TGTCACTGCA GCCCAAAGC A

查询= INSP002

(189 个字母)

数据库: ncbi-nt

1,021,646 个序列; 共 4,527,726,597 个字母

检索.....完成

| 产生明显比对的序列: | 评分 (bits) | E 值 |
|---|--------------|-------|
| gi 15217064 gb AF400435.1 AF400435 智人 cerberus 相关的... | 69 | 2e-10 |
| gi 11427432 ref XM_005320.1 智人 cerberus 1 (非洲爪蟾... | 69 | 2e-10 |
| gi 4885134 ref NM_005454.1 智人 cerberus 1 (非洲爪蟾... | 69 | 2e-10 |
| gi 5802091 gb AF139721.1 AF139721 红原鸡 cerberus 同源物... | 68 | 3e-10 |
| gi 5902633 gb AF179484.1 AF179484 红原鸡 caronte (CAR) mR... | 67 | 5e-10 |
| gi 2853615 gb AF035579.1 AF035579 小鼠 cerberus 同源物 ... | 61 | 5e-08 |
| gi 6753409 ref NM_009887.1 小鼠 cerberus 1 同源物(非洲爪蟾... | 61 | 5e-08 |
| gi 2654026 gb AF031896.1 AF031896 小鼠 cerberus 相关的 ... | 61 | 5e-08 |
| gi 1513087 gb U64831.1 XLU64831 非洲爪蟾头部诱导因子... | 57 | 9e-07 |
| gi 6679017 ref NM_008675.1 小鼠成神经细胞瘤,抑制... | 55 | 3e-06 |

图 1

```
>gi|15217064|gb|AF400435.1|AF400435 智人 cerberus 相关的 1
      (CER1) mRNA, 完整序列
      长度 = 804

      评分 = 68.9 bits (167), 预测 = 2e-10
      同一性 = 36/102 (35%), 正的 = 53/102 (51%)
      框 = +1

      查询: 87  VTLPLNPQEVIQGMCKAVPFVQVFSRPGCSAIRLRNHLFCGHCSSLYIPGSDPTPLVLCN 146
              V LP+  EV  C+ VPF Q  +  GC  +  ++N+LCFG C S++ PG+  C+

      题述: 442  VILPIKSHEVHWETCRTVPFSQTITHEGCEKVVVQNNLCFGKCGSVHFPGAAQHSHTSCS 621

      查询: 147  SCMPARKRWAPVVLWCLTGSSASRRVKISTMLIEGCHCSPK 188
              C+PA+  + L C  SS  +K+  ML+E C C  K

      题述: 622  HCLPAKFTTMHLPLNCTELSSV----IKV-VMLVEECQCKVK 732
```

图 2

i) ncbi-nr

数据库: 全部非冗余 GenBank CDS

翻译+PDB+SwissProt+PIR+PRF

1,242,768 个序列; 共 395,571,179 个字母

检索.....完成

| 产生明显比对的序列: | 评分 (bits) | E 值 |
|--|--------------|-------|
| ref NP_689867.1 假拟蛋白质 FLJ38607 [智人] >g... | 394 | e-109 |
| ref XP_113990.1 假拟蛋白质 XP_113990 [智人] | 221 | 5e-57 |
| ref XP_164658.1 与假拟蛋白质 FLJ38607 [智人] 相似... | 70 | 1e-11 |
| ref NP_005445.1 cerberus 1; cerberus 相关的 1; cerberus 1 (非洲爪蟾... | 69 | 3e-11 |
| gb AAD51610.1 AF139721_1 cerberus 同源物 [红原鸡] | 68 | 6e-11 |

ii) ncbi-nt

数据库: All GenBank+EMBL+DDBJ+PDB sequences (但没有 EST、STS、GSS 或者相 0、1 或 2 HTGS 序列)

1,469,831 个序列; 共 7,238,625,236 个字母

检索.....完成

| 产生明显比对的序列: | 评分 (bits) | E 值 |
|---|--------------|-------|
| ref NM_152654.1 智人假拟蛋白质 FLJ38607 (FLJ... | 394 | e-108 |
| dbj AK095926.1 智人 cDNA FLJ38607 fis, 克隆 HEART2004941 | 394 | e-108 |
| gb BC025333.1 智人, 克隆 IMAGE:4558384, mRNA, 部分编码序列 | 221 | e-103 |
| ref XM_113990.1 智人 LOC199699 (LOC199699), mRNA | 221 | e-103 |
| gb AC092069.2 智人染色体 19 克隆 CTC-425F1, 互补... | 222 | 1e-56 |

图 3

```
>dbj|AK095926.1| 智人 cDNA FLJ38607 fis, 克隆 HEART2004941
      长度 = 1731

评分 = 394 bits (1013), 预测 = e-108
同一性 = 189/189 (100%), 正的 = 189/189 (100%)
框 = +2

查询: 1  MLLGQLSTLLCLLSGALPTGSGRPEPQSPRPQSWAAANQTWALGPGALPPLVPASALGSW 60
      MLLGQLSTLLCLLSGALPTGSGRPEPQSPRPQSWAAANQTWALGPGALPPLVPASALGSW
题述: 44  MLLGQLSTLLCLLSGALPTGSGRPEPQSPRPQSWAAANQTWALGPGALPPLVPASALGSW 223

查询: 61  KAFLGLQKARQLGMGRLQRGQDEVAAVTLPLNPQEVIQGMCKAVPFVQVFSRPGCSAIRL 120
      KAFLGLQKARQLGMGRLQRGQDEVAAVTLPLNPQEVIQGMCKAVPFVQVFSRPGCSAIRL
题述: 224 KAFLGLQKARQLGMGRLQRGQDEVAAVTLPLNPQEVIQGMCKAVPFVQVFSRPGCSAIRL 403

查询: 121 RNHL CFGHCSSLYIPGSDPTPLVLCNSCMPARKRWAPVVLWCLTGSSASRRRVKISTMLI 180
      RNHL CFGHCSSLYIPGSDPTPLVLCNSCMPARKRWAPVVLWCLTGSSASRRRVKISTMLI
题述: 404 RNHL CFGHCSSLYIPGSDPTPLVLCNSCMPARKRWAPVVLWCLTGSSASRRRVKISTMLI 583

查询: 181 EGCHCSPKA 189
      EGCHCSPKA
题述: 584 EGCHCSPKA 610
```

图 4

```

>gb|BC025333.1| 智人, 克隆 IMAGE:4558384, mRNA, 部分编码序列
      长度 = 1746

评分 = 221 bits (562), 预测(2) = e-103
同一性 = 108/108 (100%), 正的 = 108/108 (100%)
框 = +1

查询: 1  MLLGQLSTLLCLLSGALPTGSGRPEPQSPRPQSWAAANQTWALGPGALPPLVPASALGSW 60
      MLLGQLSTLLCLLSGALPTGSGRPEPQSPRPQSWAAANQTWALGPGALPPLVPASALGSW
题述: 187 MLLGQLSTLLCLLSGALPTGSGRPEPQSPRPQSWAAANQTWALGPGALPPLVPASALGSW 366

查询: 61  KAFGLQKARQLGMGRLQRGQDEVAAVTLPLNPQEVIQGMCKAVPFVQ 108
      KAFGLQKARQLGMGRLQRGQDEVAAVTLPLNPQEVIQGMCKAVPFVQ
题述: 367 KAFGLQKARQLGMGRLQRGQDEVAAVTLPLNPQEVIQGMCKAVPFVQ 510

评分 = 178 bits (451), 预测(2) = e-103
同一性 = 81/81 (100%), 正的 = 81/81 (100%)
框 = +2

查询: 109 VFSRPGCSAIRLRNHLFCFGHCSSLYIPGSDPTPLVLCNSCMPARKRWAPVVLWCLTGSSA 168
      VFSRPGCSAIRLRNHLFCFGHCSSLYIPGSDPTPLVLCNSCMPARKRWAPVVLWCLTGSSA
题述: 596 VFSRPGCSAIRLRNHLFCFGHCSSLYIPGSDPTPLVLCNSCMPARKRWAPVVLWCLTGSSA 775

查询: 169 SRRRVKISTMLIEGCHCSPKA 189
      SRRRVKISTMLIEGCHCSPKA
题述: 776 SRRRVKISTMLIEGCHCSPKA 838

```

图 5

```

1  AAATGCCTCC CAGGCTATCC AGGAGGGGCC AAGAGATTAA AAGCAGGTTC AGAAGGCTCA
61  GATGCCACTC ACCAGACAGC AGGGTCGACT GCTAGTGACC TTGAGCCCAG TCCGGACAGA
121  CAGACAGGCA GACAGACGCA CGGACAAGCA GATGCTCCTT GGCCAGCTAT CCACTCTTCT
      m l l g q l s t l

181  GTGCCTGCTT AGCGGGGCCC TGCCTACAGG CTCAGGGAGG CCTGAACCCC AGTCTCCTCG
      l c l l s g a l p t g s g r p e p q s p

241  ACCTCAGTCC TGGGCTGCAG CCAATCAGAC CTGGGCTCTG GGCCCAGGGG CCCTGCCCCC
      r p q s w a a a n q t w a l g p g a l p

301  ACTGGTGCCA GCTTCTGCCC TTGGGAGCTG GAAGGCCTTC TTGGGCCTGC AGAAAGCCAG
      p l v p a s a l g s w k a f l g l q k a

361  GCAGCTGGGG ATGGGCAGGC TGCAGCGTGG GCAAGACGAG GTGGCTGCTG TGA CTCTGCC
      r q l g m g r l q r g q d e v a a v t l

421  GCTGAACCCT CAGGAAGTGA TCCAGGGGAT GTGTAAGGCT GTGCCCTTCG TTCAGGTGTT
      p l n p q e v i q g m c k a v p f v q v

481  CTCCCGGCCG GGCTGCTCAG CCATACGCCT CCGAAATCAT CTGTGCTTTG GTCATTGCTC
      f s r p g c s a i r l f n h l c f g h c
      INSP002-CP1

541  CTCTCTCTAC ATCCCTGGCT CGGACCCAC CCCACTAGTC CTGTGCAACA GCTGTATGCC
      s s l y i p g s d p t p l v l c n s c m

601  TGCTCGCAAG CGTTGGGCAC CCGTGGTCCT GTGGTGTCTC ACTGGCAGCT CAGCCTCCCG
      p a r k r w a p v v l w c l t g s s a s
      INSP002-CP2

661  TCGACGGGTG AAGATATCCA CCATGCTGAT CGAGGGGTGT CACTGCAGCC CAAAAGCATG
      r r r v k i s t m l i e g c h c s p k a

721  AACTGAGCAT CGTGGATGGG TGCACGGAGA CACGCACCTT GGAGAAATGA GGGGAGAT
      -

```

图 6

```
1 CTCAGCCATA CGCCTCCGAA ATCATCTGTG CTTTGGTCAT TGCTCCTCTC TCTACATCCC  
  s a i r l r n h l c f g h c s s l y i  
61 TGGCTCGGAC CCCACCCAC TAGTCCTGTG CAACAGCTGT ATGCCTGCTC GCAAGCGTTG  
  p g s d p t p l v l c n s c m p a r k r  
121 GGCACCCGTG GTCCTGTGGT GTCTCACTGG CAGCTCAGC  
  w a p v v l w c l t g s s
```

图 7

分子: 产物 1, 4109 bps DNA Circular
 文件名: ~7879480.cm5

说明: NoName 与 PCR II-TOPO-open 连接

分子性质:

| 类型 | 起始位 | 终止位 | 名称 | 说明 |
|-----|------|------|----------------|------------|
| 区域 | 1 | 336 | | LacZa' |
| 标记物 | 239 | | SP6 | |
| 基因 | 337 | 495 | INSP002 部分编码序列 | 插入的 PCR 产物 |
| 区域 | 496 | 747 | | 'LacZa |
| 标记物 | 584 | | C T7 | |
| 区域 | 749 | 1163 | | f1 ori |
| 基因 | 1497 | 2291 | KanR | 卡那霉素抗性基因 |
| 基因 | 2309 | 3169 | AmpR | 氨比西林抗性基因 |
| 基因 | 3314 | 3987 | | pUC ori |

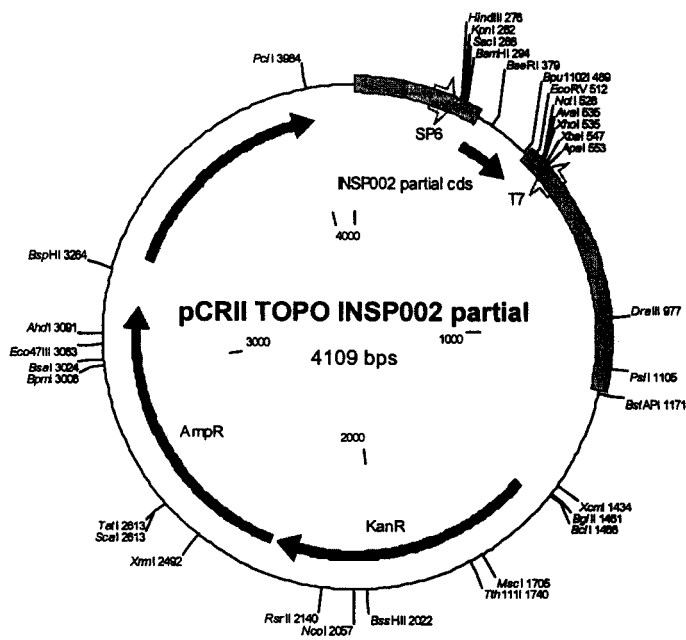


图 8

图 9a: 核苷酸序列比对

```

          10      20      30      40      50      60
INSPpr CAGAAATGCCTCCAGGCTATCCAGGAGGGGCCAAGAGATTAAGCAGGTTCAGAAGGC
:
Serono C-----

          70      80      90      100     110     120
INSPpr TCAGATGCCACTCACCAGACAGCAGGGTCGACTGCTAGTGACCTTGAGCCCAGTCCGGAC
Serono -----

          130     140     150     160     170     180
INSPpr AGACAGACAGGCAGACAGACGCACGGACAAGCAGATGCTCCTTGGCCAGCTATCCACTCT
Serono -----

          190     200     210     220     230     240
INSPpr TCTGTGCCTGCTTAGCGGGCCCTGCCTACAGGCTCAGGGAGGCCTGAACCCAGTCTCC
Serono -----

          250     260     270     280     290     300
INSPpr TCGACCTCAGTCCTGGGCTGCAGCCAATCAGACCTGGGCTCTGGGCCAGGGCCCTGCC
Serono -----

          310     320     330     340     350     360
INSPpr CCCACTGGTGCCAGCTTCTGCCCTTGGGAGCTGGAAGGCCTTCTTGGGCCTGCAGAAAGC
Serono -----

          370     380     390     400     410     420
INSPpr CAGGCAGCTGGGGATGGGCAGGCTGCAGCGTGGGCAAGACGAGGTGGCTGCTGTGACTCT
Serono -----

          430     440     450     460     470     480
INSPpr GCCGCTGAACCCCTCAGGAAGTGATCCAGGGGATGTGTAAGGCTGTGCCCTTCGTTCAGGT
Serono -----

          490     500     510     520     530     540
INSPpr GTTCTCCCGCCCCGGCTGCTCAGCCATACGCCTCCGAAATCATCTGTGCTTTGGTCATTG
:
Serono -----TCAGCCATACGCCTCCGAAATCATCTGTGCTTTGGTCATTG
          10      20      30      40

          550     560     570     580     590     600
INSPpr CTCCTCTCTCTACATCCCTGGCTCGGACCCACCCACTAGTCTGTGCAACAGCTGTAT
:
Serono CTCCTCTCTCTACATCCCTGGCTCGGACCCACCCACTAGTCTGTGCAACAGCTGTAT

          50      60      70      80      90      100

          610     620     630     640     650     660
INSPpr GCCTGCTCGCAAGCGTTGGGCACCCGTGGTCTGTGGTGTCTCACTGGCAGCTCAGCCTC
:
Serono GCCTGCTCGCAAGCGTTGGGCACCCGTGGTCTGTGGTGTCTCACTGGCAGCTCAGC---
          110     120     130     140     150

          670     680     690     700     710     720
INSPpr CCGTCGACGGGTGAAGATATCCACCATGCTGATCGAGGGGTGCTACTGCAGCCCAAAGC
Serono -----

```

图 9b: 蛋白质序列比对

```

Serono -----
INSP00 MLLGQLSTLLCLLSGALPTGSGRPEPQSPRPQSWAAANQTWALGPGALPPLVPASALGSW
          10      20      30      40      50      60

Serono -----SAIRL
          :::::
INSP00 KAFLGLQKARQLGMGRLQRGQDEVAAVTLPLNPQEVIQGMCKAVPFVQVFSRPGCSAIRL
          70      80      90      100     110     120

          10      20      30      40      50
Serono RNHLCFGHCSSLYIPGSDPTPLVLCNSCMPARKRWAPVVLWCLTGSSA-----
          ::::::::::::::::::::::::::::::::::::::::::::::::::::::
INSP00 RNHLCFGHCSSLYIPGSDPTPLVLCNSCMPARKRWAPVVLWCLTGSSASRRRVKISTMLI
          130     140     150     160     170     180

Serono -----
INSP00 EGCHCSPKA

```

图 9

1501 *TTTGTATTTT TAGTAGAGAC AGAGTTTCAC CATATTAGCC TGGCTGGTCT CAAACTCCTG*
1561 *GCCTCAGGTG ATCTGCCAC CTCGGCCTCC CAAAGTGCTG GGATCAAATC CACTGTTAAT*
1621 *CATTAGGCTG AACTGTCTCT TATAGAATGA GGTCAAAGAC ACTCCAGTT GCAGGGAGGG*
1681 *TAGATGGCCC CACCCAGACC GAGAGACACA GTGATGACCT CAGCCTAGGG ACMCCAAAA*
1741 *AAAAAAAAA AAAAAA*

87 bp 插入片段与 INSP002 预测序列相比较的位置阴影框显示。

3'非编码序列的 Alu 序列以斜线表示。

图 10

图 11a: 核苷酸序列

```

INSP00 -----CAGAAATGCCTCCAGGCTATCCAGGAG
BC0253 GGGAGACCTGGAAGGAAGCGACTGCACTGATCCAGAAATGCCTCCAGGCTATCCAGGAG
          10      20      30      40      50      60
          30      40      50      60      70      80
INSP00 GGGCCAAGAGATTAAAAGCAGGTTCAGAAGGCTCAGATGCCACTCACCAGACAGCAGGGT
BC0253 GGGCCAAGAGATTAAAAGCAGGTTCAGAAGGCTCAGATGCCACTCACCAGACAGCAGGGT
          70      80      90      100     110     120
          90     100     110     120     130     140
INSP00 CGACTGCTAGTGACCTTGAGCCAGTCCGGACAGACAGACAGGCAGACAGACGCACGGAC
BC0253 CGACTGCTAGTGACCTTGAGCCAGTCCGGACAGACAGACAGGCAGACAGACGCACGGAC
          130     140     150     160     170     180
          150     160     170     180     190     200
INSP00 AAGCAGATGCTCCTTGCCAGCTATCCACTCTTCTGTGCCTGCTTAGCGGGGCCCTGCCT
BC0253 AAGCAGATGCTCCTTGCCAGCTATCCACTCTTCTGTGCCTGCTTAGCGGGGCCCTGCCT
          190     200     210     220     230     240
          210     220     230     240     250     260
INSP00 ACAGGCTCAGGGAGGCTGAACCCAGTCTCCTCGACCTCAGTCCTGGGCTGCAGCCAAT
BC0253 ACAGGCTCAGGGAGGCTGAACCCAGTCTCCTCGACCTCAGTCCTGGGCTGCAGCCAAT
          250     260     270     280     290     300
          270     280     290     300     310     320
INSP00 CAGACCTGGGCTCTGGGCCAGGGGCCCTGCCCCACTGGTGCCAGCTTCTGCCCTTGGG
BC0253 CAGACCTGGGCTCTGGGCCAGGGGCCCTGCCCCACTGGTGCCAGCTTCTGCCCTTGGG
          310     320     330     340     350     360
          330     340     350     360     370     380
INSP00 AGCTGGAAGGCCTTCTTGGGCTGCAGAAAGCCAGGCAGCTGGGGATGGGCAGGCTGCAG
BC0253 AGCTGGAAGGCCTTCTTGGGCTGCAGAAAGCCAGGCAGCTGGGGATGGGCAGGCTGCAG
          370     380     390     400     410     420
          390     400     410     420     430     440
INSP00 CGTGGGCAAGACGAGGTGGCTGCTGTGACTCTGCCGCTGAACCCCTCAGGAAGTGATCCAG
BC0253 CGTGGGCAAGACGAGGTGGCTGCTGTGACTCTGCCGCTGAACCCCTCAGGAAGTGATCCAG
          430     440     450     460     470     480
          450     460     470
INSP00 GGGATGTGTAAGGCTGTGCCCTTCGTTTCAG-----
BC0253 GGGATGTGTAAGGCTGTGCCCTTCGTTTCAGACACGGGAGTCTCGCTATGTTGCCCAAGCT
          490     500     510     520     530     540
          480
INSP00 -----GTGTT
BC0253 AGTCTTGAGCTTCTGGCCTCAAGCAATCCTCCCACCTCAGCCTCCTGAGTTCTAGGTGTT
          550     560     570     580     590     600
          490     500     510     520     530     540
INSP00 CTCCCGCCCGGCTGCTCAGCCATACGCCTCCGAAATCATCTGTGCTTTGGTCATTGCTC
BC0253 CTCCCGCCCGGCTGCTCAGCCATACGCCTCCGAAATCATCTGTGCTTTGGTCATTGCTC
          610     620     630     640     650     660
          550     560     570     580     590     600
INSP00 CTCTCTACATCCCTGGCTCGGACCCACCCCACTAGTCCTGTGCAACAGCTGTATGCC
BC0253 CTCTCTACATCCCTGGCTCGGACCCACCCCACTAGTCCTGTGCAACAGCTGTATGCC
          670     680     690     700     710     720

```

```

          610      620      630      640      650      660
INSP00 TGCTCGCAAGCGTTGGGCACCCGTTGGTCTGTGGTGTCTCACTGGCAGCTCAGCCTCCCG
          :
BC0253 TGCTCGCAAGCGTTGGGCACCCGTTGGTCTGTGGTGTCTCACTGGCAGCTCAGCCTCCCG
          730      740      750      760      770      780

          670      680      690      700      710      720
INSP00 TCGACGGGTGAAGATATCCACCATGCTGATCGAGGGGTGTCACTGCAGCCAAAAGCATG
          :
BC0253 TCGACGGGTGAAGATATCCACCATGCTGATCGAGGGGTGTCACTGCAGCCAAAAGCATG
          790      800      810      820      830      840

INSP00 A-----
          :
BC0253 AACTGAGCATCTGGATGGGTGCACGGAGACACGCACCTTGGAGAAATGAGGGGAGATGGA
          850      860      870      880      890      900

INSP00 -----
BC0253 CCAAGAAAGACGTGGACCTGGATGATGTA TCTCTGGGTCAAGAGACCAGGGATGCAGGGTT
          910      920      930      940      950      960

INSP00 -----
BC0253 AGGCAGACAGGTCCCAGAGTCCTCACCCCTGCTCCCAGACAGTAGACACAGTGCCCGTC
          970      980      990      1000     1010     1020

INSP00 -----
BC0253 CTGGAGTTGCACCACTGATAGTACAGCACACAATGATTGACAAC TCACTTTTTTTTTT
          1030     1040     1050     1060     1070     1080

INSP00 -----
BC0253 TTTTGTAGATGGAGTCTCGCTCTGTGCGCCAGGCTGGAGTGCAGTGGCGCAATCTCAGCT
          1090     1100     1110     1120     1130     1140

INSP00 -----
BC0253 CACTGCAAGCTCCACCTCCCGGTTTATGCCATTCTCCTGTCTCAGCCTCCCGAGTAGCT
          1150     1160     1170     1180     1190     1200

INSP00 -----
BC0253 GGGACTACAGGCACCCGCCAACACGCCCGGCTAATTTTTTTGTATTTT TAGTAAAGACAGG
          1210     1220     1230     1240     1250     1260

INSP00 -----
BC0253 GTTTCACCGTGTAGCCAGGATGGTCTCTATCTCCTGACCTCGTGATCTGCCTGCCTTGG
          1270     1280     1290     1300     1310     1320

INSP00 -----
BC0253 CCTTATTATTTTTTTTTTAAGGACAGAGTCTCTCTGTCAACCAGGCTGGAGTGCAATG
          1330     1340     1350     1360     1370     1380

INSP00 -----
BC0253 GCGCGATCTTGGCTCACTGTAACTTCCACTTGCCAGGCTCAAGCAGTTCTCCTGCCTCAG
          1390     1400     1410     1420     1430     1440

```

```

INSP00 -----
BC0253 CCTCCTGAGTAGCTGGGACTACAGGCACCCGCCACCATGCCCAGCTAATTTTTGTATTTT
      1450      1460      1470      1480      1490      1500

INSP00 -----
BC0253 TAGTAGAGACAGAGTTTTACCATATTAGCCTGGCTGGTCTCAAACCTCCTGGCCTCAGGTG
      1510      1520      1530      1540      1550      1560

INSP00 -----
BC0253 ATCTGCCACCTCGGCCTCCCAAAGTGCTGGGATCAAATCCACTGTTAATCATTAGGCTG
      1570      1580      1590      1600      1610      1620

INSP00 -----
BC0253 AACTGTCTCTTATAGAATGAGGTCAAAGACACTCCAGTTGCAGGGAGGGTAGATGGCCC
      1630      1640      1650      1660      1670      1680

INSP00 -----
BC0253 CACCCAGACCGAGAGACACAGTGATGACCTCAGCCTAGGGACACCAAAAAAAAAAAAAAA
      1690      1700      1710      1720      1730      1740

INSP00 -----
BC0253 AAAAAA

```

图 11b: 蛋白质序列

```

4558384_OR      MLLGQLSTLLCLLSGALPTGSGRPEPQSPRPQSWAAANQTWALGPGALPPLVPASALGSW
INSP002_PR      MLLGQLSTLLCLLSGALPTGSGRPEPQSPRPQSWAAANQTWALGPGALPPLVPASALGSW
                  *****

4558384_OR      KAFLGLQKARQLGMGRLQRGQDEVAAVTLPLNPQEVIQGMCKAVPFVQTRE-----
INSP002_PR      KAFLGLQKARQLGMGRLQRGQDEVAAVTLPLNPQEVIQGMCKAVPFVQVFSRPGCSAIRL
                  *****

4558384_OR      -----SRYVAQAS---LELLASNPP-----TSAS-----
INSP002_PR      RNHLCFGHCSSLYIPGSDPTPLVLCNSCMPARKRWAPVVLWCLTGSSASRRRVKISTMLI
                  * * . * * * * * .***

4558384_OR      -----
INSP002_PR      EGCHCSPKA

```

图 11

```

1  GTCGACTGCT AGTGACCTTG AGCCCAAGTCC GGACAGACAG ACAGGCAGAC AGACGCACGG
61  ACAAGCAGAT GCTCCTTGGC CAGCTATCCA CTCTTCTGTG CCTGCTTAGC GGGGCCCTGC
      m l l g   q l s   t l l   c l l s   g a l

121  CTACAGGCTC AGGGAGGCCT GAACCCAGT CTCCTCGACC TCAGTCCTGG GCTGCAGCCA
      p t g   s g r p   e p q   s p r   p q s w   a a a

181  ATCAGACCTG GGCTCTGGGC CCAGGGGCC TGCCCCACT GGTGCCAGCT TCTGCCCTTG
      n q t   w a l g   p g a   l p p   l v p a   s a l

241  GGAGCTGGAA GGCCTTCTTG GGCCTGCAGA AAGCCAGGCA GCTGGGGATG GGCAGGCTGC
      g s w   k a f l   g l q   k a r   q l g m   g r l

301  AGCGTGGGCA AGACGAGGTG GCTGCTGTGA CTCTGCCGCT GAACCTCAG GAAGTGATCC
      q r g   q d e v   a a v   t l p   l n p q   e v i

361  AGGGGATGTG TAAGGCTGTG CCCTTCGTTT TCTCCCGGCC CGGCTGCTCA GCCATACGCC
      q g m   c k a v   p f v   l s r   p g c s   a i r

421  TCCGAAATCA TCTGTGCTTT GGTCAATGCT CCTCTCTCTA CATCCCTGGC TCGGACCCCA
      l r n   h l c f   g h c   s s l   y i p g   s d p

481  CCCCACTAGT CCTGTGCAAC AGCTGTATGC CTGCTCGCAA GCGTTGGGCA CCCGTGGTCC
      t p l   v l c n   s c m   p a r   k r w a   p v v

541  TGTGGTGTCT CACTGGCAGC TCAGCCTCCC GTCGACGGGT GAAGATATCC ACCATGCTGA
      l w c   l t g s   s a s   r r r   v k i s   t m l

601  TCGAGGGGTG TCACTGCAGC CCAAAGCAT GAACTGAGCA TCTGGATGGG TGCACGGAGA
      i e g   c h c s   p k a

661  CACGCACCTT GGAGAAATGA GGGGAGATGG ACCAAGAAAG ACGTGGACCT GGATGATGT

```

图 12

分子: pCR4 blunt TOPO INSP002V, 4676 bps DNA Circular
 文件名: pCR4 blunt TOPO INSP002V.cm5
 说明: INSP002 嵌套正反引物的产物与 pCR4Blunt-TOPO 的连接
 备注: 质粒 ID 13075

分子性质:

| 类型 | 起始位 | 终止位 | 名称 | 说明 |
|-----|------|------|--------------|-----------------|
| 区域 | 2 | 216 | | lac 启动子区域 |
| 区域 | 205 | 221 | | M13 反向引发位点 |
| 区域 | 217 | 294 | | LacZa-ccdB 基因融合 |
| 标记物 | 243 | | T3 | |
| 区域 | 262 | 294 | | 多接头 |
| 区域 | 294 | 294 | | TOPO 克隆位点 |
| 区域 | 295 | 1013 | | INSP002V PCR 产物 |
| 基因 | 363 | 923 | INSP002V ORF | |
| 区域 | 1014 | 1529 | | LacZa-ccdB 基因融合 |
| 区域 | 1014 | 1031 | | 多接头 |
| 区域 | 1014 | 1014 | | TOPO 克隆位点 |
| 标记物 | 1066 | | C T7 | |
| 区域 | 1089 | 1074 | C | -20M13 正向引发位点 |
| 基因 | 1878 | 2672 | KanR | |
| 区域 | 2864 | 2868 | | 核糖体结合位点 |
| 基因 | 2876 | 3736 | AmpR | |
| 区域 | 3881 | 4554 | | pUC 起点 |

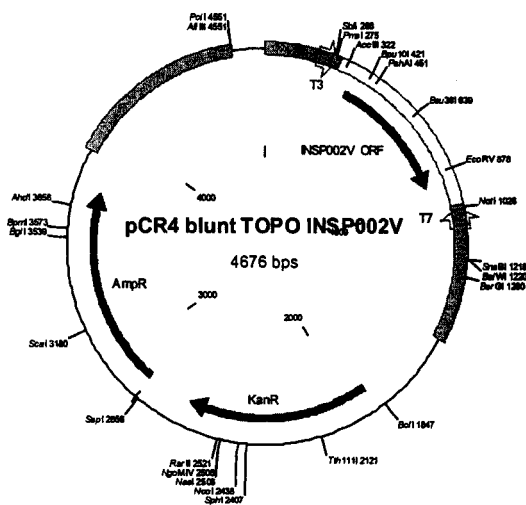


图 13

图 14a: 核苷酸序列

```

INSPpr CAGAAATGCCTCCCAGGCTATCCAGGAGGGGCCAAGAGATTAAAAGCAGGTTCAGAAGGC
INSPen -----
          70      80      90      100     110     120
INSPpr TCAGATGCCACTCACCAGACAGCAGGGTCGACTGCTAGTGACCTTGAGCCCAGTCCGGAC
          :
INSPen -----GTCGACTGCTAGTGACCTTGAGCCCAGTCCGGAC
          10      20      30
          130     140     150     160     170     180
INSPpr AGACAGACAGGCAGACAGACGCACGGACAAGCAGATGCTCCTTGGCCAGCTATCCACTCT
          :
INSPen AGACAGACAGGCAGACAGACGCACGGACAAGCAGATGCTCCTTGGCCAGCTATCCACTCT
          40      50      60      70      80      90
          190     200     210     220     230     240
INSPpr TCTGTGCCTGCTTAGCGGGGCCCTGCCTACAGGCTCAGGGAGGCCTGAACCCCAGTCTCC
          :
INSPen TCTGTGCCTGCTTAGCGGGGCCCTGCCTACAGGCTCAGGGAGGCCTGAACCCCAGTCTCC
          100     110     120     130     140     150
          250     260     270     280     290     300
INSPpr TCGACCTCAGTCCTGGGCTGCAGCCAATCAGACCTGGGCTCTGGGCCAGGGGCCCTGC
          :
INSPen TCGACCTCAGTCCTGGGCTGCAGCCAATCAGACCTGGGCTCTGGGCCAGGGGCCCTGC
          160     170     180     190     200     210
          310     320     330     340     350     360
INSPpr CCCACTGGTGCCAGCTTCTGCCCTTGGGAGCTGGAAGGCCTTCTTGGGCCTGCAGAAAGC
          :
INSPen CCCACTGGTGCCAGCTTCTGCCCTTGGGAGCTGGAAGGCCTTCTTGGGCCTGCAGAAAGC
          220     230     240     250     260     270
          370     380     390     400     410     420
INSPpr CAGGCAGCTGGGGATGGGCAGGCTGCAGCGTGGGCAAGACGAGGTGGCTGCTGTGACTCT
          :
INSPen CAGGCAGCTGGGGATGGGCAGGCTGCAGCGTGGGCAAGACGAGGTGGCTGCTGTGACTCT
          280     290     300     310     320     330
          430     440     450     460     470     480
INSPpr GCCGCTGAACCCTCAGGAAGTGATCCAGGGGATGTGTAAGGCTGTGCCCTTCGTTCCAGGT
          :
INSPen GCCGCTGAACCCTCAGGAAGTGATCCAGGGGATGTGTAAGGCTGTGCCCTTCGTTCC----
          340     350     360     370     380     390

```

```

          490      500      510      520      530      540
INSPpr GTTCTCCCGCCCGGCTGCTCAGCCATACGCCTCCGAAATCATCTGTGCTTTGGTCATTG
          ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
INSPen --TCTCCCGCCCGGCTGCTCAGCCATACGCCTCCGAAATCATCTGTGCTTTGGTCATTG
          400      410      420      430      440

          550      560      570      580      590      600
INSPpr CTCCTCTCTCTACATCCCTGGCTCGGACCCACCCCACTAGTCCTGTGCAACAGCTGTAT
          ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
INSPen CTCCTCTCTCTACATCCCTGGCTCGGACCCACCCCACTAGTCCTGTGCAACAGCTGTAT
          450      460      470      480      490      500

          610      620      630      640      650      660
INSPpr GCCTGCTCGCAAGCGTTGGGCACCCGTGGTCCCTGTGGTGTCTCACTGGCAGCTCAGCCTC
          ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
INSPen GCCTGCTCGCAAGCGTTGGGCACCCGTGGTCCCTGTGGTGTCTCACTGGCAGCTCAGCCTC
          510      520      530      540      550      560

          670      680      690      700      710      720
INSPpr CCGTCGACGGGTGAAGATATCCACCATGCTGATCGAGGGGTGTCAGTGCAGCCCAAAGC
          ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
INSPen CCGTCGACGGGTGAAGATATCCACCATGCTGATCGAGGGGTGTCAGTGCAGCCCAAAGC
          570      580      590      600      610      620

INSPpr ATGA-----
          ::::
INSPen ATGAAGTGCATCTGGATGGGTGCACGGAGACACGCACCTTGGAGAAATGAGGGGAGAT
          630      640      650      660      670      680

INSPpr -----

INSPen GGACCAAGAAAGACGTGGACCTGGATGATGT
          690      700      710

```

图 14b: 蛋白质序列

```

INSP002V    23 MLLGQLSTLLCLLSGALPTGSGRPEPQSPRPQSWAAANQTWALGPGALPPLVPASALGSW
INSP002      1 MLLGQLSTLLCLLSGALPTGSGRPEPQSPRPQSWAAANQTWALGPGALPPLVPASALGSW
*****

INSP002V    83 KAFLGLQKARQLGMGRLQRGQDEVAAVTLPLNPQEVIIQGMCKAVPF--VLSRPGCSAIRL
INSP002     61 KAFLGLQKARQLGMGRLQRGQDEVAAVTLPLNPQEVIIQGMCKAVPFVQVFSRPGCSAIRL
*****

INSP002V    141 RNHLCFGHCSSLYIPGSDPTPLVLCNSCMPARKRWAPVVLWCLTGSSASRRRVKISTMLI
INSP002     121 RNHLCFGHCSSLYIPGSDPTPLVLCNSCMPARKRWAPVVLWCLTGSSASRRRVKISTMLI
*****

INSP002V    201 EGCHCSPKA
INSP002     181 EGCHCSPKA
*****

```

图 14

分子: pEAK12 d, 8760 bps DNA Circular
 文件名: pEAK12DEST.cm5

说明: 哺乳动物细胞表达载体(质粒 ID 11345)

分子性质:

| 类型 | 起始位 | 终止位 | 名称 | 说明 |
|-----|------|------|-----------|---------------|
| 区域 | 2 | 595 | | pmb-ori |
| 基因 | 596 | 1519 | Amp | |
| 区域 | 1690 | 2795 | EF-1alpha | |
| 区域 | 2703 | 2722 | | pEAK12F 引物的位置 |
| 区域 | 2796 | 2845 | | MCS |
| 标记物 | 2855 | | attR1 | |
| 基因 | 3256 | 3915 | CmR | |
| 基因 | 4257 | 4562 | ccdB | |
| 标记物 | 4603 | | C attR2 | |
| 区域 | 4733 | 4733 | | MCS |
| 区域 | 4734 | 5162 | | 聚腺苷酸/剪接 |
| 区域 | 4819 | 4848 | C | pEAK12R 引物的位置 |
| 基因 | 5781 | 5163 | C PUR | PUROMYCIN |
| 区域 | 6005 | 5782 | C tK | tK 启动子 |
| 区域 | 6500 | 6006 | C Ori P | |
| 基因 | 8552 | 6500 | C EBNA-1 | |
| 区域 | 8553 | 8752 | sv40 | |

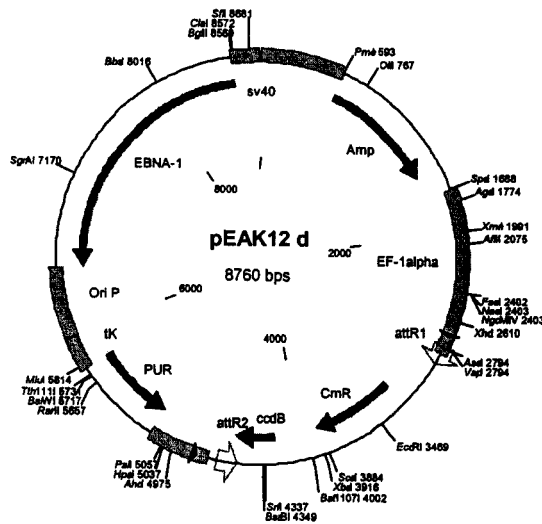


图 15

分子: pDONR201, 4470 bps DNA Circular
 文件名: pDONR201.cm5

说明: Gateway 入门载体 (Invitrogen) - 质粒 ID 13309

分子性质:

| 类型 | 起始位 | 终止位 | 名称 |
|----|------|------|---------|
| 区域 | 332 | 563 | attP1 |
| 基因 | 959 | 1264 | ccdB |
| 区域 | 2513 | 2744 | attP2 |
| 基因 | 2868 | 3677 | KanR |
| 区域 | 3794 | 4467 | pUC ori |

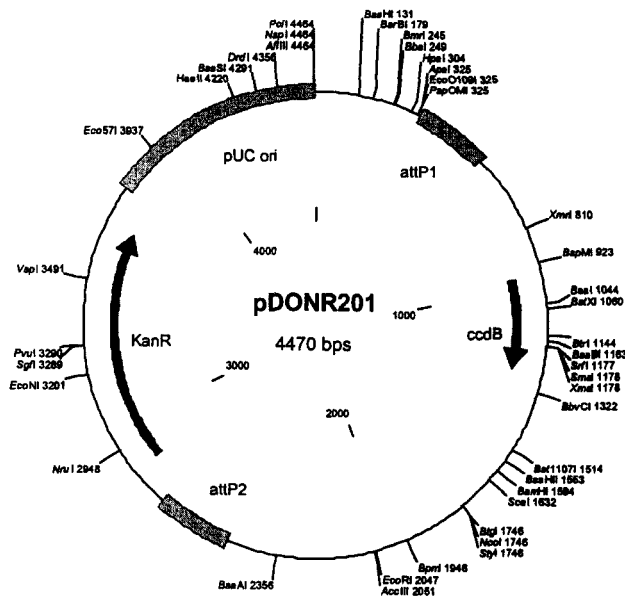


图 16

分子: pEAK12d-INSP002-V-6HIS, 7528 bps DNA Circular
 文件名: pEAK12d-INSP002V-6HIS.cm5
 说明: cDNA 插在 attR1 和 attR2 位点之间
 备注: 质粒 ID 13227

分子性质:

| 类型 | 起始位 | 终止位 | 名称 | 说明 |
|----|------|------|-----------|-------------|
| 区域 | 2 | 595 | | pmb-ori |
| 基因 | 596 | 1519 | AmpR | |
| 区域 | 1690 | 2795 | EF-1alpha | 启动子 |
| 区域 | 2703 | 2722 | | pEAK12F |
| 区域 | 2796 | 2845 | | MCS '1 |
| 区域 | 2855 | 2874 | | attB1 |
| 基因 | 2888 | 3469 | INSP002-V | |
| 区域 | 3474 | 3495 | | attB2 |
| 区域 | 3501 | 3501 | | 'MCS |
| 区域 | 3502 | 3930 | | 聚腺苷酸/剪接 |
| 区域 | 3616 | 3597 | C | pEAK12R |
| 基因 | 4549 | 3931 | C | PUROMYCIN R |
| 区域 | 4773 | 4550 | C | tK 启动子 |
| 区域 | 5268 | 4774 | C Ori P | |
| 基因 | 7320 | 5268 | C EBNA-1 | |
| 区域 | 7321 | 7520 | sv40 | |

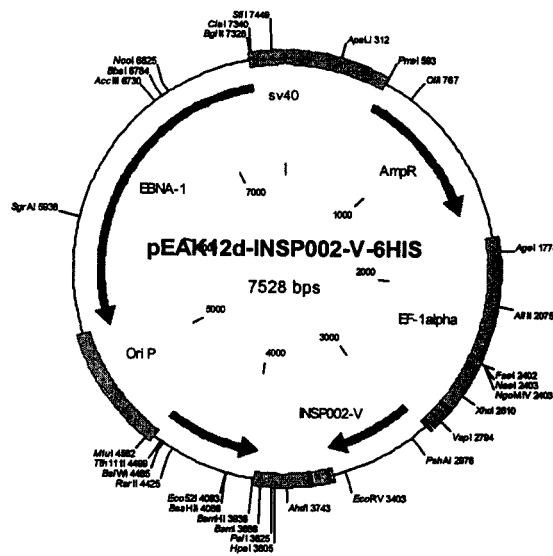


图 17

```

1  GATGCTCCTT GGCCAGCTAT CCACTCTTCT GTGCCTGCTT AGCGGGGCC TGCCTACAGG
   m l l g q l s t l l c l l s g a l p t
   INSP002-FL-F
61  CTCAGGGAGG CCTGAACCCC AGTCTCCTCG ACCTCAGTCC TGGGCTGCAG CCAATCAGAC
   g s g r p e p q s p r p q s w a a a n q
121 CTGGGCTCTG GGCCAGGGG CCCTGCCCCC ACTGGTGCCA GCTTCTGCCC TTGGGAGCTG
   t w a l g p g a l p p l v p a s a l g s
181 GAAGGCCTTC TTGGGCCTGC AGAAAGCCAG GCAGCTGGGG ATGGGCAGGC TGCAGCGTGG
   w k a f l g l q k a r q l g m g r l q r
241 GCAAGACGAG GTGGCTGCTG TGA CTCTGCC GCTGAACCCT CAGGAAGTGA TCCAGGGGAT
   g q d e v a a v t l p l n p q e v i q g
301 GTGTAAGGCT GTGCCCTTCG TTCAGGTGTT CTCCCGGCC GGCTGCTCAG CCATACGCCT
   m c k a v p f v q v f s r p g c s a i r
361 CCGAAATCAT CTGTGCTTTG GTCATTGCTC CTCTCTCTAC ATCCCTGGCT CGGACCCAC
   l r n h l c f g h c s s l y i p g s d p
421 CCCACTAGTC CTGTGCAACA GCTGTATGCC TGCTCGCAAG CGTTGGGCAC CCGTGGTCCT
   t p l v l c n s c m p a r k r w a p v v
481 GTGGTGTCTC ACTGGCAGCT CAGCCTCCCG TCGACGGGTG AAGATATCCA CCATGCTGAT
   l w c l t g s s a s r r r v k i s t m l
541 CGAGGGGTGT CACTGCAGCC CAAAAGCATG AACTGAGCAT CGTGGATGG
   i e g c h c s p k a
   INSP002-FL-R

```

图 18

分子: pCR4blunt TOPO INSP002FL, 4546 bps DNA Circular
 文件名: pCR4 blunt TOPO INSP002 FL.cm5,
 说明: NoName 与 pCR4Blunt-TOPO open*的连接

分子性质:

| 类型 | 起始位 | 终止位 | 名称 | 说明 |
|-----|------|------|---------|---------------------|
| 区域 | 2 | 216 | | lac 启动子区域 |
| 区域 | 205 | 221 | | M13 反向引发位点 |
| 区域 | 217 | 294 | | LacZa-ccdB 基因融合 |
| 标记物 | 243 | | T3 | |
| 区域 | 262 | 294 | | 多接头 |
| 区域 | 294 | 294 | | TOPO 克隆位点 |
| 区域 | 295 | 883 | | INSP002-F3R3 PCR 产物 |
| 基因 | 296 | 865 | INSP002 | INSP002 ORF |
| 区域 | 884 | 1399 | | 'LacZa-ccdB 基因融合 |
| 区域 | 884 | 901 | | 多接头 |
| 区域 | 884 | 884 | | 'TOPO 克隆位点 |
| 标记物 | 936 | | C T7 | T7 启动子 |
| 区域 | 959 | 944 | C | -20M13 正向引发位点 |
| 基因 | 1748 | 2542 | KanR | 卡那霉素抗性基因 ORF |
| 区域 | 2734 | 2738 | | 核糖体结合位点 |
| 基因 | 2746 | 3606 | AmpR | 氨比西林抗性基因 ORF |
| 区域 | 3751 | 4424 | pUC ori | pUC 起点 |

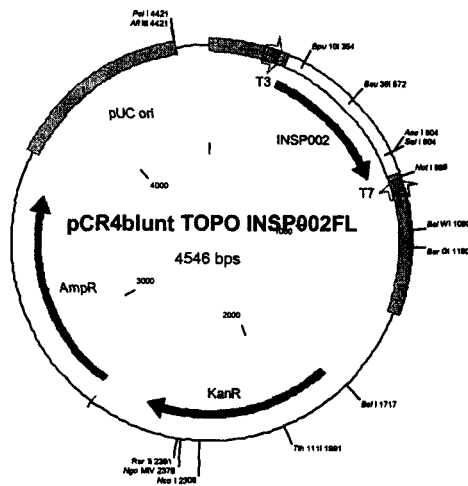


图 19

| | | | |
|----------------|---|---------|------------|
| 专利名称(译) | 胱氨酸结折迭蛋白质 | | |
| 公开(公告)号 | CN1620466A | 公开(公告)日 | 2005-05-25 |
| 申请号 | CN02828274.4 | 申请日 | 2002-12-20 |
| [标]申请(专利权)人(译) | 阿雷斯贸易股份有限公司 | | |
| 申请(专利权)人(译) | 阿雷斯贸易股份有限公司 | | |
| 当前申请(专利权)人(译) | 阿雷斯贸易股份有限公司 | | |
| [标]发明人 | MD戴维斯 CB菲尔普斯 RJ法根 C包尔 M优克 M伊伯森 | | |
| 发明人 | M·D·戴维斯 C·B·菲尔普斯 R·J·法根 C·包尔 M·优克 M·伊伯森 | | |
| IPC分类号 | A01K67/027 A61K38/00 A61K39/00 A61P3/00 A61P9/00 A61P25/00 A61P29/00 A61P31/04 A61P35/00 A61P37/06 C07H21/04 C07K14/475 C07K14/52 C07K16/18 C12N1/15 C12N1/19 C12N1/21 C12N5/10 C12N15/09 C12N15/19 C12P21/02 C12Q1/68 G01N33/15 G01N33/50 G01N33/53 C07K16/28 G01N33/566 A61K38/19 | | |
| CPC分类号 | A61K38/00 A61K39/00 A61P3/00 A61P9/00 A61P25/00 A61P29/00 A61P31/04 A61P35/00 A61P37/06 C07K14/52 | | |
| 代理人(译) | 周承泽 | | |
| 优先权 | 2001030738 2001-12-21 GB | | |
| 外部链接 | Espacenet SIPO | | |

摘要(译)

本发明涉及新型蛋白质(INSP002)，它经本发明鉴定为一种分泌性蛋白质，是胱氨酸结折迭细胞因子超家族的Dan家族成员，以及涉及该蛋白质和来自编码基因的核酸序列在诊断、预防和治疗疾病中的应用。

