

[19]中华人民共和国国家知识产权局

[51]Int. Cl⁶

C07K 14/46

C12N 15/12 A61K 39/395

[12] 发明专利申请公开说明书

[21] 申请号 98103356.3

[43]公开日 1999年3月3日

[11]公开号 CN 1209436A

[22]申请日 98.7.9 [21]申请号 98103356.3

[30]优先权

[32]97.7.9 [33]DE [31]19729211.9

[32]98.2.11 [33]DE [31]19805371.1

[71]申请人 德国赫彻斯特马里奥罗塞尔有限公司

地址 联邦德国法兰克福

[72]发明人 B·弗莱肯斯汀

A·恩塞尔

[74]专利代理机构 中国国际贸易促进委员会专利商标事务所

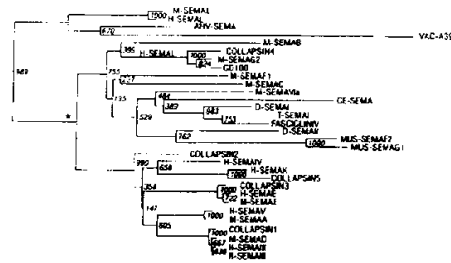
代理人 陈文平

权利要求书 2 页 说明书 127 页 附图页数 18 页

[54]发明名称 人类信息素 L(H-SemaL)和其它物种中相应的信息素

[57]摘要

本发明涉及由一特定的结构域结构所区别的新的信息素和其衍生物,并且涉及编码这些信息素的核酸(DNA,RNA,cDNA)和其衍生物,以及它们的用途。本发明涉及在免疫系统中具有生化功能并且具有迄今未公开和未预料到的结构域结构的信息素(免疫调节信息素)。这些新的信息素被认为是 L 型信息素(Sema-L)。它们包含一个 N 端信号肽、一个特征性 Sema 结构域。并且在蛋白质 C 端区含义一个免疫球蛋白样结构域和代表潜在跨膜结构域的疏水结构域。



权 利 要 求 书

1. 一种包含一个特征性 Sema 结构域的信息素, 其中该蛋白质具有一个 N 端信号肽, 并且在 C 端区具有一个免疫球蛋白样结构域以及一个跨膜结构域。所说的信息素被认为是 L 型信息素(Sema-L), 以及 L 型信息素的衍生物。
2. 如权利要求 1 中所要求的信息素, 其中所说的蛋白质(人类 L 型信息素(H-SemaL))具有氨基酸序列 SEQ ID NO. 3。
3. 如权利要求 1 和 2 中之一个或多个所要求的信息素, 其中所说的蛋白质在 Sema 结构域区域中与 H-SemaL 的 Sema 结构域具有至少 40%的氨基酸等同性。
4. 如权利要求 1 至 2 中之一个或多个所要求的信息素, 其中所说的蛋白质包含部分氨基酸序列 SEQ ID NO. 4 (鼠信息素(M-SemaL))。
5. 一种包含编码如权利要求 1 至 4 中之一个或多个所要求的 L 型信息素的核酸序列的核酸及其衍生物。
6. 如权利要求 5 所要求的核酸, 其中所说的核酸序列是信息素 L 基因。
7. 如权利要求 5 和 6 中之一个或多个所要求的核酸, 其中所说的核酸序列包含 H-SemaL 的基因。
8. 如权利要求 5 所要求的核酸, 其中所说的核酸序列包含 L 型信息素的 cDNA。
9. 如权利要求 8 所要求的核酸, 其中所说的 cDNA 是 H-SemaL 的 cDNA。
10. 如权利要求 8 所要求的核酸, 其中所说的 cDNA 是 M-SemaL 的 cDNA。
11. 一种制备如权利要求 1 至 4 中之一个或多个所要求的 L 型信息素的方法, 该方法包括将编码 L 型信息素或者其衍生物的核酸序列克隆进表达载体并且表达之。
12. 如权利要求 11 所要求的方法, 其中将真核细胞用于表达。
13. L 型信息素或其衍生物或编码 L 型信息素的核酸序列或其衍生物在产生可以用于治疗或预防免疫疾病的药物中的用途。
14. 如权利要求 13 所要求的核酸序列或其衍生物在基因治疗中的用途。
15. L 型信息素或编码 L 型信息素核酸序列在鉴别免疫调节剂的方法中的用途。
16. 一种鉴别免疫调节剂的方法, 该方法包括在特定的条件下将 L 型信息素

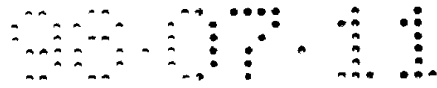
与待测药剂一起温育，在没有待测药剂但是其它方面相同的条件下进行第二批平行培养，然后确定待测药剂的抑制或激活作用。

17. 一种鉴别免疫调节剂的方法，该方法包括在特定的条件下和在待测药剂存在下表达编码L型信息素的核酸序列，并且测定其表达程度。

5 18. 一种制备编码L型信息素的核酸的方法，其中用特异性引物通过聚合酶链式反应扩增这种核酸。

19. 一种信息素抗体，该抗体识别相应于 SEQ ID NO. 4 中的 179-378 位氨基酸的 H-SemaL 的表位，或者识别相应于 SEQ ID NO. 4 中的 480-666 位氨基酸的 H-SemaL 的表位。

10 20. 一种制备如权利要求 19 所要求的信息素抗体的方法，其中作为具有表位标记的融合蛋白表达所说的表位，通过这一表位标记纯化所说的表位，并且将纯化的融合蛋白用于免疫接种。



说明书

人类信息素 L 和其它物种中相应的信息素

5 本发明涉及由一特定的结构域结构识别的新的信息素(semaphorin)和其衍生物, 涉及编码这些信息素的核酸(DNA、RNA、cDNA)和其衍生物, 以及它们的制备和用途。

Kolodkin 首次描述了信息素作为保守基因家族的成员 {Kolodkin 等(1993)细胞 75: 1389-1399}。

10 现在已经克隆出(并且在某些情况下已经描绘出)其它信息素的基因或部分基因。到目前为止, 已经公开了总共 5 种人类信息素(H-SemaIII, H-SemaV, H-SemaIV, H-SemaB 和 H-SemaE) {(Kolodkin 等(1993); Poche 等(1996)癌基因 12: 1289-1297; Sekido 等(1996)美国科学院院报 93: 4120-4125; Xiang 等(1996)基因组 32: 39-48; Hall 等(1996)美国科学院院报 39: 11780-11785; Yamada
15 等(1997)(GenBank 登记号 AB000220)}, 8 种鼠类信息素 (小鼠基因; M-Sema A 到 M-Sema-H) { (Püschel 等(1995)神经元 14: 941-948; Messerschmidt 等(1995)神经元 14: 949-959; Inigaki 等(1995)FEBS 通讯: 370:269-272; Adams 等(1996)Mech. Dev 57: 33-45; Christensen 等(1996)(GenBank 登记号 Z80941, Z93948)}, 5 种鸡精蛋白(小鸡)脑衰蛋白-1 到-5) (Luo 等(1993); Luo 等(1995)神经
20 元 14: 1131-1140 }, 以及来自大鼠(R-Sema-III) {Giger 等(1996)J. Comp. Neurol. 375: 378-392}, 斑马鱼, 昆虫毛(果蝇(果蝇 *D melanogaster*: D-SemaI 和 D-SemaII), 甲虫 *Tribolium confusum*: T-Sema-I), 蚱蜢(*Schistocerca americana*: G-Sema-I) {Kolodkin 等(1993)}, 以及线虫(*C.elegans*: Ce-Sema) {Roy 等(1994)(GenBank 登记号 U15667)} 的基因。此外, 两种痘病毒 (牛痘(ORF-A39)和
25 天花(ORFA39-同源物)) {Kolodkin 等(1993)} 和 alcelaphine 疱疹病毒 1 型(AHV-1)(AHV-Sema) {Ensser 和 Fleckenstein (1995)普通病毒学 76:1063-1067} 具有与信息素同源的基因。

表-1 总结了到目前为止在各种物种中鉴别的信息素。表 1 显示了信息素的名称(第 1 列), 所使用的同义词(第 2 列), 分离特定信息素的物种(第 3 列); 和其中
30 已知的关于所说的编码蛋白质的结构域结构的数据和关于染色体位点的数据(表 1

的第 4 列), 存储在基因数据库(例如, EST(表达序列标记)数据库, EMBL(欧洲分子生物学实验室, Heidelberg)数据库或 NCBI (国家生物技术信息中心, 马里兰, 美国)的基因序列的登记号, 以及这些数据出版的参考信息(表 1 的第 5 列).

到目前为止, 所有公开的信息素基因的基因产物(编码的信息素)都具有一个 N 5 端信号肽, 它在 C 端具有大约 450 至 500 个氨基酸长度的特有 Sema 结构域. 高度保守的氨基酸基元和一些高度保守的半胱氨酸残基被已在 Sema 结构域中找到. 基因产物(信息素)在 C 端序列方面是不同的, 这些序列跟随在 Sema 结构域之后, 并且由一个或多个结构域组成. 例如, 在这些 C 端氨基酸序列中有跨膜结构域(TM), 免疫球蛋白样结构域(Ig) (免疫球蛋白恒定区), 胞质序列(CP), 加工信 10 号(P)(例如有共有序列(RXR), 其中 R 是精氨酸, X 是任何氨基酸)和/或亲水 C 端 (HPC). 到目前为止所公开的信息素根据 C 端的结构域结构的不同可以划分为 5 个不同的子群(I 到 V):

- I 分泌性, 没有其它结构域(例如 ORF-A49)
- II Ig 分泌性(没有跨膜结构, 例如 AHV-Sema)
- 15 III Ig, TM, CP 具有胞质序列, 膜锚着(例如 CD100)
- IV Ig, (P), HPC 具有亲水性 C 端, 分泌性(例如 H-SemaIII, M-SemaD, 脑衰蛋白-I)
- V Ig, TM, CP 具有 C 端 7 血小板反应蛋白基元, 膜锚着 (例如 M-SemaF 和 G)

20 到此为止还没有描述过信息素受体或胞外配体. 已经公开了与信息素-介导作用相关连的胞内杂合三聚体 GTP 结合性蛋白质复合物. 在小鸡中已被鉴别的这些蛋白质复合物的一个组分称为 CRMP(脑衰蛋白应答介导蛋白质), 并且被假定是信息素诱导的胞内信号级联的一个组分(Goshima 等(1995)自然 376: 509-514). 例如, CRMP62 与 unc-33、线虫蛋白质(对轴突定向成长来说是十分重要的)都具有 25 有同源性. 同样地, 已知人类蛋白质 98%的氨基酸与 CRMP62 具有等同性 (Hamajima 等(1996)基因 180: 157-163). 类似地描述了大鼠中的几个 CRMP-相关基因(Wang 等(1996)神经科学 16: 6197-6207).

所说的分泌性或跨膜信息素为增长的神经芽体传达排斥信号. 在神经系统 (CNS)的发育中, 它们起着重要的作用, 并且特别是在肌肉和神经组织中表达

(Kolodkin 等(1993); Luo 等(1993)细胞 75: 217-227).

与密切相关的 M-SemaF 相对比, M-SemaG 的显著表达不仅在所说的 CNS 中而且在淋巴与造血系统的细胞中已经观察到(Furuyima 等(1996)生物化学杂志 271: 33376-33381).

5 近来已经鉴别出两个其它的人类信息素, H-SemaIV 和 H-Sema V(特别是染色体 3p21.3 区域), 其缺失与各种类型的支气管癌有关. H-SemaIV{(Roche 等(1996), Xiang 等(1996), Sekido 等(1996)}与 M-SemaE 在氨基酸水平上大约有 50%的等同性, 而 H-SemaV {Sekido 等(1996)}是 M-SemaA 的直接同系物(86%的氨基酸等同性). 因为在所切除的 3p21.3 基因座上进行 DNA 测序工程期间发现了
10 这些基因(H-SemaIV 和 V), 这两个基因复合物的内含子-外显子结构是已知的. 两个基因都能在各种神经元和非神经元组织中表达.

同样仅仅在前不久, 细胞表面分子 CD100(人)(在活化的 T 细胞上表达并诱导)已经鉴别为是一种信息素(表 1 中同样列出). 它通过 CD40 受体和相应配体 CD40L 帮助与 B 细胞相互作用. CD100 是一个 150 kd (千道尔顿)的膜锚着糖蛋白二聚
15 体. 已经描述了 CD100 的外胞质 C 端 与一迄今未知的激酶的结合 {Hall 等(1996)}. 这就意味 CD100 是第一个并且是迄今为止仅有的已证明在免疫系统的细胞中表达的信息素.

在“rhadinoviruses 的转化基因”方案中, 已经克隆和测序了 alcelaphine 疱疹病毒 1 型(AHV-1)的完全基因组 {Ensser 等(1995)}. AHV-1 是恶性卡他发热、各
20 种与淋巴增殖综合症有关的反刍动物疾病的诱发剂(并且通常是致命的). 经过分析, 发现了一个开放读框, 在病毒基因组的一端具有与痘苗病毒基因远程的但是显著的同源性(ORF-A39 相应于 Ensser 等(1995)普通病毒学 76:1063-1067 的 VAC-A39), 后者已经被归为信息素基因家族. 而 AHV-1 信息素(AHV-Sema)具有非常保守的信息素结构, 痘病毒基因(ORF-A39 和 ORF-A39-同源物, 参见表 1)具
25 有 C 端截短, 即其中仅仅非完全地存在保守 Sema 结构域.

所发现的 AHV-Sema 与 dbEST (EST(表达序列标志)数据库(db))的数据库比较提从人胎盘的 2 种独立 cDNA 克隆都提供了两种 EST 序列(登记号 H02902 , H03806(克隆 151129), 登记号 R33439 和 R33537(克隆 135941)). 与迄今为止所描述的神经元信息素相比, 这些明显地显示了与 AHV-1 的信息素更大的同源性.

70%，优选的是超过 80%，特别优选的是超过 90%的氨基酸等同性。例如可以利用 GAP 程序来确定或计算出百分同源性(GCG 程序包，遗传计算机组(1991))。

本发明这样的实施方案是相应的小鼠信息素(鼠信息素(M-SemaL))。例如，它含有在表 5 中所显示的部分氨基酸序列(鼠信息素(M-SemaL))。

5 本发明也涉及这样的相应信息素，在较少相关的物种(系统发育上相互之间十分远的)中，它与表 4 中显示的 H-SemaL 完全氨基酸序列相比仅仅具有大约 15 至 20%的氨基酸等同性(在蛋白质氨基酸序列的整个长度上考虑)，优选的是 25 至 30%，特别优选的是 35 至 40%，或是更高的等同性。

10 编码 L 型信息素的基因具有外显子-内含子结构。例如，这些基因可能具有 10 到 20 个外显子，优选的是大约 11 至 18 个，特别优选的是 12 至 16 个外显子和相应的数量的内含子。然而，它们也可以具有与 H-SemaL 基因相同数量的外显子和内含子(13 或 15 个外显子，优选的是 14 个外显子)。本发明的一个特定实施方案涉及 H-SemaL 基因。这一基因的长度优选的是 8888 至 10,000 或更多个核苷酸。人类信息素基因优选地包含表 14 所给出的核苷酸序列或在 GenBank®数据库中储
15 存的登记号 AF030697 的核苷酸序列。这些核苷酸序列含有至少 13 个内含子。此外，人类信息素基因在 5' 端具有一个附加序列区域。合适时这一区域进一步含有编码和非编码序列，例如一或两个另外的内含子或外显子。

试图将人类 L 型信息素定位在染色体上，这揭示了将相应的基因定在位置 15q22.3-23。对于 M-SemaL 的基因，相应地定位在位置 9A3.3-B。

20 作为内含子-外显子复合结构的结果，信息素 mRNA 初级转录物的剪接可能发生变化，其结果将导致产生信息素的不同剪接变体。这些剪接变体所翻译的蛋白质是本发明信息素的衍生物。它们的氨基酸序列，实质上其结构域结构相应于本发明描述的 L 型信息素，但是与后者相比较适当地方却被截短了。例如，可以形成完全或部分缺失跨膜结构域的剪接变体，含有一个不完全(或没有)跨膜结构域
25 但是含有信号肽的信息素衍生物可能是具有分泌性的，并且通过这种方式对细胞外部，局部或其它相对较远的距离起作用，例如对其它细胞起作用。例如，另一个剪接变体可以不再含有编码信号肽的序列，同时，在适当时也没有编码代表潜在跨膜结构域的疏水氨基酸序列的序列。一种结果将是其信息素衍生物不能掺入到膜内，也不能进行分泌(除非通过分泌泡)。这样的信息素衍生物可以进入胞内过

程, 例如信号转导过程. 这样就可能以相同的基本分子(L 型信息素)和其衍生物(例如剪接变体)控制和/或协调各种胞内和胞外的过程.

本发明一个特定的实施方案涉及从本发明 L 型信息素产生的信息素衍生物, 但是它含有一个不完全的(或没有)跨膜结构域.

5 本发明的另一个实施方案涉及从本发明 L 型信息素产生的信息素衍生物, 但是它不含有信号肽.

信号肽也可以在翻译后切除. 这导致以截短的结构域结构形成了膜结合性(TM 结构域)或分泌性(没有 TM 结构域的剪接变体)信息素衍生物. 现在一种以这种方式经过翻译后加工的信息素衍生物仅仅含有 Sema 结构域, Ig 结构域. 同时
10 在适当时具有跨膜结构域. 信号肽切割位点可以正好定位在信号肽的末端, 但也可以例如定位在距离氨基端 40 至 50 个氨基酸或更远.

"截短的"(即包含更少的结构域)信息素 L 衍生物可以与其它不衍生自 L 型信息素的信息素的区分是, 在存在的结构域中与 L 型信息素具有很大(>90%)的氨基酸等同性或具有同一氨基酸序列.

15 根据本发明的信息素也可以用其它方法经过翻译后修饰. 例如, 它们可以通过一次, 两次, 三次, 四次, 五次, 六次, 七次, 八次, 九次, 十次或更多次的糖基化(N-和/或 O-糖基化). 于是, 信息素的氨基酸序列可能具有相等数量或更多的潜在糖基化位点的共有序列, 优选的有 5 个这样的位点. 本发明的一个实施方案涉及这样的信息素, 其中糖基化位点位于与 H-SemaL 氨基酸序列(表 4)相一致
20 的位置 105, 157, 258, 330 和 602.

此外, 所说的信息素可以是它们磷酸化衍生物的形式. 信息素可以是不同激酶的底物, 例如氨基酸序列可能有蛋白激酶 C, 酪氨酸激酶和/或肌酸激酶的共有序列. 此外, 信息素的氨基酸序列可能有潜在的肉豆蔻酰化位点的共有序列. 相应的信息素衍生物在这些位点上可以用肉豆蔻酸进行酯化.

25 本发明的 L 型信息素和它们的衍生物可以是单体, 二聚体和/或多聚体的形式. 例如两个或更多信息素或他们的衍生物可以用分子间的二硫键连接在一起. 形成分子内的二硫键也是可能的.

本发明信息素的其它衍生物是融合蛋白. 这一类型的融合蛋白一方面含有 L 型信息素或其部分, 同时, 另外还含有其它的肽或蛋白质或它的一部分. 例如,

肽或蛋白质或其部分可能是表位标记(例如 His 标记(6 X 组氨酸), Myc 标记, 流感标记), 因而可以利用它们, 例如, 用于纯化融合蛋白, 或它们也可以用于标记融合蛋白, 例如 GFP (绿色荧光蛋白质), 例如, L 型信息素衍生物的例子在实施例的构建体中已经说明. 在表 7 至 15 中可以发现这些构建体序列, 并在适当时考虑质粒的注释来解释.

5 本发明进一步涉及编码本发明 L 型信息素和/或它们的衍生物的核酸序列, 优选 DNA 和 RNA 序列, 例如相应的基因, mRNA 的不同剪接变体, 与此相应的 cDNA, 以及其衍生物(例如 DNA 或 RNA 的盐). 本发明的衍生物是经过如分子生物学方法修饰并且适应特定需要的序列或序列片段, 如截短的基因或基因片段
10 (例如启动子序列, 终止子序列), cDNA 或它的嵌合体, 用于表达和克隆的构建体和其盐.

一个实施方案涉及 L 型信息素的基因组序列(基因). 本发明涉及内含子和外显子序列, 以及基因调控序列, 例如启动子, 增强子和沉默子序列.

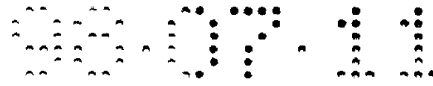
这个实施方案一方面涉及 H-SemaL 的或其衍生物的基因. 本发明一方面涉及
15 包含表 14 中给出的核苷酸序列的基因. 本发明进一步涉及包含以登记号 AF030697 收藏在 GenBank 数据库的核苷酸序列的基因.

这一实施方案进一步涉及 M-SemaL 的基因和它的衍生物.

本发明进一步涉及 H-SemaL 的 cDNA 或它的衍生物(例如 cDNA 片段). 一个特定的实施方案是根据表 2 中的核苷酸序列 H-SemaL 的 cDNA. 本发明进一步涉
20 及以登记号 AF030698 收藏在 GenBank 数据库的 H-SemaL 的 cDNA. 本发明也涉及相应于这些 cDNA 的 mRNA 或它的片段.

本发明进一步涉及 M-SemaL 的 cDNA 或它的衍生物(例如 cDNA 片段). 一个特定的实施方案是在表 3 中所显示的 M-SemaL 的部分 cDNA 序列, 以及包含这部分 cDNA 序列的 cDNA 序列. 另一个本发明的实施方案涉及以登记号 AF030699
25 收藏在 GenBank 数据库的 M-SemaL 的 cDNA. 本发明也涉及相应于这些 cDNA 的 mRNA 或它的部分.

本发明也包含了等位基因和/或与上述的信息素序列相比仅仅轻微不同并编码同一或仅仅轻微修饰之蛋白质(氨基酸序列的不同少于或等于 10%)的基因/mRNA/cDNA 的个体表达形式(衍生物的进一步例子). 在实施例中所表明的构建



体提供了其衍生物的进一步例子。在表 7 至 14 表中描述了这些构建体的序列，可考虑质粒的注释来解释。

本发明进一步涉及包含编码 L 型信息素或其衍生物之 DNA 的质粒。例如，这种类型的质粒可能是适合于例如在大肠杆菌中 DNA 扩增的具有高复制率的质粒。

特定实施方案为具有信息素或其片段或其衍生物的可以在原核和/或真核表达系统中表达的质粒。组成型表达质粒和那些包含诱导型启动子的质粒都是合适的。

本发明也涉及制备编码 L 型信息素或其衍生物的核酸的方法。这些核酸(例如 DNA 或 RNA)可以是合成得来的，例如用化学手段。尤其是，对这些核酸(例如相应的基因或 cDNA 或它的片段)来说利用特异性扩增引物和合适的起始原料作为模板通过 PCR 进行扩增是可能的。(例如来自合适的组织或基因组 DNA 的 cDNA)。制备信息素 L cDNA 和 H-SemaL 基因的特定方法在实施例中描述。

本发明也涉及制备 L 型信息素的方法。例如，信息素 L 或它的衍生物可以通过克隆相应的编码 L 型信息素或其衍生物的核酸序列进入表达载体，并且利用后者重组载体来转化合适细胞来制备。例如，利用原核或真核细胞是可能的。L 型信息素或其衍生物也可以在合适时用化学手段制备。

此外，L 型信息素和其衍生物可以作为与例如蛋白质或肽的融合蛋白表达。所说的蛋白质或肽使得表达的融合蛋白可以被检测，例如作为与 GFP (绿色荧光蛋白质)的融合蛋白。信息素也能作为具有一个，两个，三个或更多个表位标记的融合蛋白进行表达。例如具有 Myc 和/或 His(6 × 组氨酸)和/或流感标记。相应地利用或制备包含编码这些融合蛋白的 DNA 序列的质粒是可能的。例如，编码信息素的序列可以克隆进含有编码 GFP 和/或表位标记(例如 Myc 标记，His 标记，流感标记)的 DNA 序列的质粒。其特定的例子在表中列出的例子和序列中给出，在适当时参考关于质粒的注释。

本发明进一步涉及抗体。具体地说，这些抗体结合或识别 L 型信息素，它的衍生物或部分。其可能的例子是可以在如小鼠，兔，山羊，绵羊，小鸡等等中产生的多克隆或单克隆抗体。

本发明这种主题的特定实施方案包含直接针对如下表位的抗体，所述表位相

应于表 4 中显示的 H-SemaL 序列的 179 至 378 或 480 至 666 位的氨基酸序列。本发明也涉及用于制备特异性抗-信息素 L 抗体的方法该制备中使用包含所说的表位的抗原。

5 本发明也涉及制备抗体的方法，为了这一目的，优选地利用包含特有信息素表位和可以用于重组融合蛋白随后纯化的表位标记的融合蛋白。纯化的融合蛋白可以用于随后的免疫。为了制备重组融合蛋白，应制备相应的重组表达载体，然后用来转化合适的细胞。可以从这种细胞中分离出重组融合蛋白。例如，该方法可以是如同在实施例 8 中所描述的那样。

10 这些抗体可以用来纯化相应的信息素，例如 H-SemaL 和它的衍生物，例如在亲和柱上纯化。或用于蛋白质的免疫检测，例如利用 ELISA，蛋白质印迹法和或免疫组织化学。这些抗体也可以用于分析 H-SemaL 的表达，例如在不同的细胞类型或细胞系中。

15 H-SemaL 的 cDNA 的长度有 2636 个核苷酸(表 2)。H-SemaLcDNA 的基因产物的长度有大约 666 个氨基酸(表 4)，并且显示了典型的 L 型信息素结构域结构。基因产物具有一个 N 端信号肽(第 1 至 44 个氨基酸)，一个 Sema 结构域(第 45 个氨基酸至大约第 545 个氨基酸)，以及一个 Ig (免疫球蛋白)结构域(大约第 550 至 620 个氨基酸)；同时，在 C 端具有一个代表潜在的跨膜结构域的疏水氨基酸序列。这个结构域结构对于信息素以前从来没有描述过。它涉及可能定位在细胞表面并且属于一个新的亚组的膜结合糖蛋白。以这以前未知的结构域结构为基础，现在
20 可以将信息素划分成 VI 个子群

- | | | |
|-----|--------------|--|
| I | 分泌性 | 没有其它结构域(例如 ORF-A49) |
| II | Ig | 分泌性 (没有跨膜结构域)(例如 AHV-Sema) |
| III | Ig, TM, CP | 具有胞质序列, 膜锚着 (例如 CD100) |
| IV | Ig, (P), HPC | 具有亲水 C 端, 分泌性(例如 H-Sema-III, M-SemaD, 25 脑衰蛋白-I) |
| V | Ig, TM, CP | 具有 C 端 7 血小板反应蛋白基元, 膜锚着 (例如 M-SemaF 和 G) |
| VI | Ig, TM | 膜锚着 (例如 H-SemaL, M-SemaL) |

一个 H-SemaL 的非糖基化未加工形式的计算分子量大约是 74.8 kd (74823 道

尔顿)(利用肽-分选程序(Peptide-Sort), GCG 程序包计算)。计算的等电点是 pH =7.56。一个可能的信号肽切割位点定位在 44 和 45 个氨基酸之间(表 3; 以 Signal P 计算(<http://www.cbs.dtu.dk/services/SignalP>), 用于分析信号序列的基于神经网络的一种程序(Nielsen H 等 (1997)蛋白质工程 10: 1-6}), 这给出了加工的蛋白质

5 (没有信号肽)的分子量(MW)是 70.3 kd(70323 道尔顿)和等电点是 pH=7.01。

实际上也同样阐明了基因组的结构。H-SemaL 基因有 13 或 15 或更多个外显子, 优选地 14 个外显子, 以及有 12 或 14 个内含子, 优选地 13 个内含子。由于这种外显子-内含子复合结构, 不同的剪接变体是可能的。在 RNA 印迹法中发现了所转录的 H-SemaL 基因的 mRNA, 特别是在胎盘, 性腺, 胸腺和脾中。在神

10 经组织或肌肉组织中还没有发现任何 mRNA。有证据显示在内皮细胞中的表异性调节性表达。

变异剪接也可以导致具有牵涉到胞内信号转导的胞质内序列的 H-SemaL 形式(例如, 类似于 CD100)。同样地, 变异剪接导致分泌性 H-SemaL 形式(类似于病毒 AHV-Sema)将也是可能的。

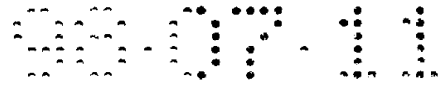
15 借助于 GCG 程序包(遗传学计算机小组(1991)GCG 包的程序手册, 版本 7, 575 Science Drive, 威斯康星, 美国 53711), FASTA (Pearson 和 Lipman (Proc. Natl. Acad. Sci. 85, 2444-2448)和 BLAST 程序(Gish 和 States (1993) Nat. Genet. 3, 266-272; Altschul 等(1990)J. Mol. Biol. 215, 403-410)完成了核苷酸和氨基酸的序列分析。这些程序也与 GenBank(版本 102.0)和 Swiss Prot(版本 34.0)一起用于序

20 列比较。

如 H-SemaL 的糖基化和肉豆蔻酰化翻译后修饰也同样是可能的。借助于 Prosite 程序(GCG 程序包)在 H-SemaL 的氨基酸序列的位置 105, 157, 258, 330 和 602 发现了 N-糖基化位点的共有序列(表 4 中显示), 而肉豆蔻酰化位点的共有序列在位置 114, 139, 271, 498, 499, 502 和 654 发现了(共有序列: G~(E, D, R,

25 K, H, P, F, Y, W)x(S, T, A, G, C, N)~(P))。此外, H-SemaL 的氨基酸序列含有几个不同激酶的潜在的磷酸化的共有序列。因此, 可以认为 H-SemaL 可以是不同激酶的底物, 例如肌酸激酶 2, 蛋白激酶 C 和酪氨酸激酶的磷酸化位点。

推定肌酸激酶 2 的磷酸化位点(共有序列 Ck2: (S, T)x2(D, E))(Prosite, GCG)在氨基酸序列的位置 119, 131, 173, 338, 419 和 481。推定蛋白激酶 C 的磷酸



化位点(共有序列 PkC: (S, T)_x(R, K))(Prosite, GCG)在氨基酸序列的位置 107, 115, 190, 296, 350, 431, 524 和 576。推定酪氨酸激酶的磷酸化位点(共有序列: (R, K)_x{2,3}(D, E){2,3}Y)(Prosite, GCG)在氨基酸序列的位置 205。

以氨基酸的单字母代码表明了所说的共有序列。

- 5 整联蛋白的特征性"RGD"基元(精氨酸-甘氨酸-天冬氨酸)位于位置 267。糖基化位点在病毒 AHV-Sema, H-SemaL 和(就目前所知的)M-SemaL 之间是高度保守的。

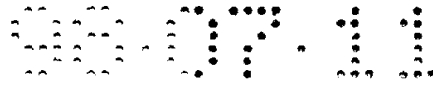
- H-SemaL 的二-或多聚化是可能的, 并且对于其它信息素(如 CD100)已经描述过(Hall 等(1996))。同样地, CD100 分子是 150kd 的膜锚着糖蛋白二聚体。然而, CD100 与本发明的人类信息素(H-SemaL)不是密切相关的。

M-SemaL 的部分 cDNA 序列的长度是 1195 个核苷酸。这种编码蛋白质的序列具有 394 个氨基酸。这 394 个氨基酸相当于 H-SemaL 的 1 至 396 个氨基酸。在 M-SemaL 中的信号肽为从 1 至 44 个氨基酸(恰如在 H-SemaL 中)。Sema 结构域从第 45 个氨基酸开始, 并且延伸到末端或有可能超出表 4 中所显示的序列末端。

- 15 利用 Clustal W 程序 (Thompson 等(1994))进行多重序列对比。利用人工 SEAVIEW (Galtier 等(1996)Comput. Appl. Biosci 12, 543-548)进一步进行序列对比。利用 Clustal W(Thompson 等(1994))来确定系统发育距离。

- 已知蛋白质序列和所说的新的信息素序列的比较以及这些序列的系统发育分析显示了根据它们的系统发育关系可以对基因进行分类。当然一般说来, 相应的信息素亚型的 C 端结构域结构是牵涉到决定相同子群的信息素比不同子群的信息素在系统发育上更密切的原因的一个因素。分离出信息素的物种也有影响, 即相应的物种在系统发育上相互之间是否存在密切的关系。

- 利用 CLUSTAL W 程序 {Thompson J.D 等(1994)核酸研究 22:4673-4680} 对已知信息素氨基酸序列(全序列和/或部分序列, 利用在表 4 和 5 中所显示的 H-SemaL 和 M-SemaL 的氨基酸序列和所有其它以所述登记号入库的序列或从这些序列中衍生的所编码的氨基酸序列)进行系统发育分析(比较图 3), 显示出 H-SemaL 和 M-SemaL 的氨基酸序列相互之间在系统发育上非常密切, 并且形成一个独立的系统发育群。反过来 H-SemaL 和 M-SemaL 与 AHV-Sema 和 Vac-A39 在系统发育上最密切。与任何其它以前所公开的信息素相比, 它们相互之间明显地是更加密切



相关的。这种分析也显示了其它信息素相互之间在系统发育上也密切相关，并且形成具有信息素的独立群。例如，具有分泌性的信息素(例如 H-SemaIII, -IV, -V 和 E)属于一个系统发育群。它们在其它物种中的同系物也属于这个亚族。而人类(跨膜)CD100 与相应的小鼠同系物(M-SemaG2) 和脑衰蛋白 4 一起属于一个系统发育群。

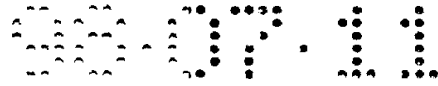
关于完全的氨基酸序列，关于十分密切相关基因(例如，H-和 M-SemaE 或 III/D)在系统发育群之内所观察到的同源性是大约在 90%和 80%之间的氨基酸等同性。而在低相关信息素基因的情况下不到 40%。在 Sema 结构域之内，所观察的氨基酸等同性百分比稍高一点，并且由于它对整个蛋白质的氨基酸序列的高贡献率(蛋白质的 50-80% 属于 Sema 结构域)，这一点显著影响整个等同性。

对于完全蛋白质来计算，H-SemaL 与 AHV-Sema 有 46%的同一性。但是，如果只考虑 Sema 结构域本身，那么氨基酸的等同性是 53%。例如，这高于相关的 M-Sema-B-和-C(完全蛋白质 37%的等同性，Sema 结构域 43%的等同性)，类似于 M-SemaA 和-E (完全蛋白质 43%，Sema 结构域 53%)。在 Sema 结构域区域中部分 M-SemaL 序列(表 6)和 H-SemaL(表 5)之间的氨基酸等同性是 93%，那么可以假定涉及了相应的同源小鼠基因。

与 H-SemaL 相比，在其它物种中相应于 H-SemaL 和 M-SemaL 的信息素在超过 40%的 Sema 结构域之内可以具有氨基酸等同性。在密切相关的脊椎动物(哺乳动物，鸟)中，发现氨基酸等同性高于 70%。

所说的信息素属于一个与病毒的 AHV-Sema 具有比与以前所公开的人和鼠的信息素更大的氨基酸等同性的新的亚族。并且具有一个以前对于人类信息素未公开过的 C 端结构。由于它们的结构域结构，通过是否属于亚组 IV 和/或与 H-SemaL 和 M-SemaL 是同一系统发育群，和/或对于完全氨基酸序列而言，与 H-SemaL 具有至少 30 至 40%的氨基酸等同性。优选的是 50 至 60%，特别优选的是 70 至 80%，或具有更大的等同性。和/或对于 Sema 结构域而言，与 H-SemaL 具有至少 70%的氨基酸等同性。优选的是大于 80%，特别优选的是大于 90%来识别这些新的信息素(亚族成员)。

L 型信息素也有一种不同类型的生化功能。这些信息素的一种新的功能是调节免疫系统。



M-SemaL 基因以发现该基因的其它功能是可能的。如果小鼠没有信息素基因可以生存, 那么它们也阐明了对于炎症疾病的潜在模型系统。如果 M-SemaL 对免疫调节来说是重要的, 那么将要期待大量的这样的小鼠。此外, 可以产生非人类 knock-in 动物(例如小鼠)。例如, 这使得能够以通常的/修饰的 H-SemaL 或修饰 M-SemaL(例如组成型和/或诱导型启动子控制下的新的信息素亚型的整合)来代替 M-SemaL。例如, 为了进一步寻找新的信息素的功能(例如人基因或这些基因衍生物的功能), 可以利用这种类型的动物, 或用于确定和鉴定免疫调节剂。

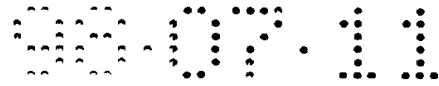
例如, 利用编码 L 型信息素或其衍生物的核酸来产生如重组免疫抑制剂, 其它可溶性蛋白质或从 L 型信息素氨基酸序列(例如 H-SemaL 或相应的核酸, 如基因)衍生出来的肽。以一种类似的方法可能产生具有结构相似性的激动剂。这些免疫抑制剂或激动剂也可以用于自身免疫疾病和炎症疾病和/或器官移植。

利用 L 型信息素的基因治疗, 例如用编码 H-SemaL 或其衍生物的核酸, 例如利用病毒或非病毒的方法。用于自身免疫疾病和炎症疾病器官转移和移植前/间/后以阻止移植排斥。

特别地, 使用新的信息素和/或编码这些信息素的核酸, 以及其衍生物(特别是 H-SemaL, 编码 H-SemaL 的 DNA, 以及其衍生物), 用于药剂筛选(特别是确定和鉴定免疫调节剂)是可能的。

功能 B): H-SemaL 是在细胞表面表达并牵涉到与细胞(例如免疫系统的细胞)的相互作用的附件分子, 例如在信号途径活化中作为附件分子。例如, 病毒基因或病毒基因产物或其它病原基因(例如起源于微生物的)可以充当该附件分子的竞争性抑制剂。具有这种功能的新的信息素同样地用于器官移植, 炎症治疗, 免疫治疗和/或基因治疗领域。

例如, 新的信息素可以用于拮抗剂或抑制剂的筛选。以这种方法鉴别的药剂可以用于例如阻断信息素受体。可溶性和/或分泌性 H-SemaL 拮抗剂或抑制剂例如可以是化学物质或新的信息素或其衍生物本身(例如适合阻断相应受体的其部分/截短形式, 如没有膜结构域或作为 Ig 融合蛋白或从后者衍生的肽)。以这种方法鉴别的特异性拮抗剂和/或抑制剂例如可能具有竞争性作用, 并且用于抑制排斥(例如在器官移植的转基因模型中), 以及用于自身免疫疾病, 炎症和器官移植。编码新的信息素的核酸(例如 DNA)或其借助于分子生物学方法产生的衍生物例如可能用于



产生非人类转基因动物。在这些转基因动物中，H-SemaL 的超量表达能导致提高对自身免疫疾病和/或炎症的易感性。于是这样的转基因动物适合筛选新的特异性免疫调节剂。

5 同样地，这样的核酸可以用于产生非人类“剔除”动物，例如“剔除”小鼠。其中失去了小鼠 M-SemaL 基因。利用这样的“剔除”动物可以找到基因的进一步生化功能。如果没有 M-SemaL 基因的小鼠能够生存，那么它们也为炎症提供了潜在的模型系统。

10 这种 DNA 同样可以用来产生非人类 knock-in 动物，例如小鼠。这使得可以用修饰 M-SemaL 基因/cDNA 或任选修饰的(例如变异)另一物种的 L 型信息素基因/cDNA(例如 H-SemaL)来代替 M-SemaL 基因。这样的转基因动物用来寻找本发明信息素的进一步功能。

15 本发明也涉及 L 型信息素及其衍生物以及编码这些蛋白质的核酸(例如基因/cDNA 和其衍生物)和/或借助于这些信息素鉴别的药剂在生产药物中的应用。例如，生产可以用于基因治疗以及包含 L 型信息素(例如 H-SemaL)表达的激动剂和/或拮抗剂的药物是可能的。为此目的例如也使用病毒的和/或非病毒方法。这些药物例如可以用于自身免疫疾病和炎症，在移植之后和/或期间和/或之前，来阻止器官移植排斥。

编码新的信息素的核酸(例如基因，cDNA 及其衍生物)在分子生物学中也可以作为工具试剂。

20 此外，新的信息素，尤其是 H-SemaL 以及核酸(例如其基因/cDNA)可以用于筛选新的药剂。例如，修饰蛋白质和/或从 H-SemaL 和/或 M-SemaL 衍生的肽在功能化验上可以用来寻找相应的受体和/或它的拮抗剂或激动剂，例如利用 H-SemaL 和同系物的表达构建体。

25 本发明也涉及 L 型信息素或编码 L 型信息素的核酸序列在鉴别药剂(尤其是免疫调节剂)的方法中的应用。

本发明也涉及使用 L 型信息素或其衍生物或编码 L 型信息素的核酸序列或其衍生物鉴别药剂的方法，以鉴别药理活性剂，例如免疫调节剂。例如，本发明涉及这样的方法，其中，在特定条件下 L 型信息素与待测药剂一起培养，平行地，第二批培养中没有待测药剂但是其它条件都相同，然后确定待测药剂的抑制或激

- 70 °C/60s 1 个循环
 Taq 60-60
 96 °C/60s 1 个循环
 96 °C/1 5s-60 °C/20s-70 °C/60s 35 个循环
 5 70 °C/60s 1 个循环
 Taq 62-40
 96 °C/60s 1 个循环
 96 °C/1 5s-62 °C/20s-70 /40s 35 个循环
 70 °C/60s 1 个循环
- 10 用 Taq 聚合酶的 PCR 反应条件:
 50µl 反应混合物中具有 100-200ng 模板, 200µM dNTP, 0.2-0.4 µM 每一引物, 2.5U Ampli-Taq^R, 5µl 10x 供给的反应缓冲液
 使用的程序:
1. XL62-6 (加长模板 PCR System^R, Boehringer Mannheim, 德国)
 15 94 °C/60s 1 个循环
 94 °C/1 5s-62 °C/30s-68 °C/6min 10 个循环
 94 °C/1 5s-62 °C/30s-68 °C/(6min +15s/个循环)25 个循环
 68 °C/7min 1 个循环
 2. XL62-12(加长模板 PCR System^R, Boehringer Mannheim, 德国)
 20 94 °C/60s 1 个循环
 94 °C/1 5s-62 °C/30s-68 °C/1 2min 10 个循环
 94 °C/1 5s-62 °C/30s-68 °C/(12min +5s/个循环)25 个循环
 68 °C/7min 1 个循环
- 用加长模板 PCR 系统的 PCR 反应条件
- 25 50µl 反应混合物中具有 100-200ng 模板, 500µM dNTP, 0.2-0.4 µM 每一引物, 0.75µl 酶混合物, 5µl 10x 供给的 No.2 反应缓冲液.
- 实施例 1:
 从 AHV-Sema 序列开始(Ensser 和 Fleckenstein (1995), 普通病毒学杂志 76 : 1063-1067), 进行 PCR 和 RACE-PCR. 用于这里的起始材料是来自胎盘组织的人
 30 cDNA, 为进行 RACE 扩增其上已连接了衔接子(MarathonTM-cDNA 扩增试剂盒.

Clontech Laboratories GmbH, Tullastraße 4, 69126 Heidelberg, 德国). 最初特异性引物(No.121234+No.121236, 表6)用于扩增长度约为800bp(碱基对)的PCR片段(PCR程序: (Taq60-60)). 克隆与测序该片段 (Taq 染料标记脱氧终止子测序试剂盒, 应用生物系统, Foster City, CA, USA/ Brunnenweg 13, Weil der Stadt).

- 5 PCR产物的测序揭示了与AHV-Sema的DNA序列具有高度同源性的序列, 与两个EST序列相同.

使用引物对(No.121237+No.121239, 表6)鉴别出600bp的PCR片段. 这表明它们是来自相同基因的DNA序列的克隆.

实施例2:

- 10 对实施例1的800bp PCR片段进行放射性标记(通过{Feinberg(1983)Anal. Biochem 132: 6-13}的方法利用 ^{32}P - α -dCTP进行随机引导), 并且作为多组织Northern印迹法(人类多组织Northern印迹法II, Clontech, Heidelberg, 德国)的探针, 这包括来自脾、胸腺、前列腺、睾丸、卵巢、小肠、大肠和白细胞(PBL)的mRNA样品. 这清楚地显示了在脾和性腺(睾丸, 卵巢)中长度约为3.3kb的
- 15 mRNA的表达, 并且在胸腺和肠道中较弱. 主印迹法杂交(许多组织RNA的点印迹法(Human RNA Master Blot™, Clontech))确认了这一结果, 并且在胎盘组织中也显示出强的表达.

- 在42℃严格条件(5xSSC, 50mM磷酸钠 pH 6.8, 50%甲酰胺, 100 $\mu\text{g}/\text{ml}$ 酵母RNA)下杂交16小时. 对印迹严格地冲洗(65℃, 0.2 x SSC, 0.1%SDS)
- 20 并且对Fuji BAS2000 Phosphoimager™暴光.

实施例3:

- 用这种探针筛选在噬菌体 λ gt10(人脾5' STRETCH PLUS cDNA, Clontech)中克隆的来自人脾的cDNA文库, 同时鉴别到一个 λ 克隆. 插入这个克隆的1.6kb长的cDNA利用载体特异性引物No. 207608 +No. 207609(表6)(位于EcoRI克隆
- 25 位点的侧面)由PCR(Expand™长模板PCR系统, Boehringer Mannheim GmbH, Sandhofer Straße 116, 68305 Mannheim)扩增, 同时对产生的PCR片段进行测序. 这个克隆包含cDNA的5'端, 并且在3'方向也扩展已知的cDNA序列. 从新的cDNA部分序列开始, 设计出RACE-PCR的新引物(No. 232643, No. 232644, No. 233084, 表6). 利用一种改进的热循环技术(MJ-Research, Biozym Diagnostik

GmbH, 31833 Hess 的 PTC-200 (Oldendorf), 其具有明显更有效的运行参数(加热和冷却速率), 利用引物 No.232644 和 No.232643 和 AP1 扩增 3' RACE-PCR 产物, 并且将它克隆到载体 pCR2.1 (Invitrogen, De Schelp 12, 9351 NV Leek, 荷兰)中. 对 3' RACE-PCR 产物进行测序, 用这种方法确定 cDNA 的 3' 端. 在 5' 方向的 RACE 扩增(引物 No. 31990 和 No.233084 和 AP1)以几个核苷酸延伸过 cDNA 5' 端, 并且确认在鉴别的 λ 克隆中发现的 H-SemaL 的氨基端.

实施例 4:

从一短的鼠 EST(登记号 260340)开始, 以从中产生的引物 No. 260813 (表 6)和 H-SemaL 特异性引物 No. 121234 (表 6)进行 PCR(条件: Taq52-60)扩增长约 840 bp 的鼠 cDNA 的 DNA 片段, 随后克隆至载体 pCR2.1 中. 包含该 DNA 片段的基因称为 M-SemaL. 产生的 M-SemaL DNA 片段用于从小鼠脾(小鼠脾 5' STRETCH cDNA, Clontech)调查 cDNA 库, 鉴定几个克隆是可能的.

从鼠内皮 cDNA 用引物 No. 260812 和 No. 260813 进行 PCR(Taq60-30)提供了长为 244 碱基对的 PCR 片段. PCR 结果表明在鼠内皮细胞中有明显基线表达. 用细胞因子干扰素- γ 和脂多糖刺激后则下降.

实施例 5:

采用荧光原位杂交(FISH)来调查染色体的定位情况. 为了这一目的, 从人血液样品和小鼠细胞系 BINE 4.8 开始分别制备人和鼠的中期染色体(Keyna 等(1995)免疫学杂志 155, 5536-5542)(Kraus 等(1994)遗传学 23, 272-274). 利用 RNase 和胃蛋白酶处理载玻片(Liehr 等(1995)应用细胞遗传学 21, 185-188). 对于杂交, 利用 120 mg 人切口平移信息素样品和 200 mg 相应的小鼠样品. 每种情况下, 杂交在 37 °C, 20 μ g STD 以及 4.0 μ g COT1-DNA 存在下在湿室中进行 3 天.

用 50% 甲酰胺/2x SSC 冲洗载玻片(45 °C 下 3 次, 每次 5min), 然后, 用 2x SSC 冲洗(37 °C 下 3 次, 每次 5min), 以及用 FITC-亲和素系统(Liehr 等(1995))检测生物素化样品. 利用荧光显微镜测评载玻片. 每个实验以双份进行, 测定了 25 个中期/样品. 表明 H-SemaL 位于染色体 15q23 上. 定位在邻近染色体是 Bardet-Biedls 综合症和 Tay-Sachs 疾病的基因座(氨基己糖苷酶).

实施例 6:

H-SemaL 基因的基因组内含子-外显子结构的大部分已被阐明.

基因组DNA片段扩增从已从PHA-刺激外周淋巴细胞(血液)分离出来的250 mg 人基因组DNA开始。较短的片段利用 Ampli Taq^R (Perkin Elmer)进行扩增, 较长的片段利用加长模板 PCR System^R (Boehringer Mannheim)进行扩增。

到目前为止, 通过PCR 扩增来克隆和鉴别 H-SemaL 的几乎完全的基因组座是可能的。现已可能鉴别总共超过 8888 bp 的基因组序列, 从而基本阐明该基因的内含子-外显子结构。

实施例 7:

表达克隆

因为不能从 lambda-gtl0 cDNA 库分离出来信息素基因完全克隆, 同时也不能经 PCR 获得完全克隆, 所以使用 N 端 DNA 片段的引物 No. 240655 和 No. 121339 以及 C 端 DNA 片段的引物 No. 240656(含有 HindIII 和 PmeI 切割位点)和 No. 121234 通过 PCR(XL62-6)在 2 个重叠亚片段上扩增出 cDNA 的编码区。产生的 DNA 片段(亚片段)克隆到载体 pCR21 中。对两个亚片段完全测序, 最后通过插入 0.6kb C 端 SstI-HindIII 限制片段到质粒(包含 N 端 DNA 片段, 并且已经用限制酶 SstI 与 HindIII 酶切)中来制备完全的 H-SemaL cDNA。利用 EcoRI 切割位点(在 pCR2.1 中)和 HindIII 切割位点(在表 6 的引物 No. 240656 中)从质粒 pCR2.1-H-SemaL(表 7 中显示的序列, SEQ ID NO. 34)上切除完全基因, 并且连接到相应切割的组成性表达载体 pCDNA3.1 (-)MycHisA(Invitrogen)上。EcoRI-ApaI 片段(没有 Myc-His 标记)从所产生的重组质粒 pCDNA3.1(-)H-SemaL-MycHisA(表 8 中显示的序列)中切下, 并且连接到诱导型载体 pIND(同样已用 EcoRI-ApaI 酶切)(蜕皮激素诱导型哺乳动物表达系统, Invitrogen)上。重组质粒称为 pIND-H-SemaLEA (表 11 中显示的序列)。将 pCDNA3.1 (-)H-SemaL-Myc-HisA (表 9 中显示的序列)的 EcoRI-PmeI 片段(具有 Myc-His 标记)插入到 EcoRI-EcoRV-切割的载体 pIND 中。重组质粒称为 pIND-H-SemaL-EE (表 10 中显示的序列)。

通过连接 PCR-扩增 EGFP 读框(来自载体 pEGFP-C1 (Clontech), 利用引物 No. 243068+No. 243069, Taq52-60)到 pCDNA3.1 (-)H-SemaL-MycHisA 的 PmeI 切割位点上制备具有增强绿色荧光蛋白质(EGFP)的 H-SemaL 融合基因, 形成质粒 pCDNA3.1 (-)H-SemaL-EGFP-MycHisA (表 9 中显示的序列)。

在表 7 至 13 表中用小写字母标出了 H-SemaL 的序列, 其部分片段或衍生物,

同时用大写字母标出了质粒序列。

实施例 8：

为了制备 H-SemaL-特异性抗体，将 H-SemaL 的 cDNA 片段整合进原核表达载体，并且在大肠杆菌中表达，以及纯化信息素衍生物。信息素衍生物作为具有 His 标记的融合蛋白来表达。因此，利用包含 His 标记序列并允许信息素 cDNA 片段整合入读框的载体。例如，N 端 6x 组氨酸标记使用镍整合亲和性层析纯化成为可能(Qiagen GmbH, Max-Volmer Straße 4, 40724 Hilden):

1. 利用引物 No. 150788 和 No. 150789 通过 PCR 扩增编码 179-378 位氨基酸的部分 H-SemaLcDNA，并将这条 DNA 片段连接到载体 pQE30 (已用限制性酶 BamHI 和 HindIII 切割)(Qiagen)上(构建体 pQE30-H-SemaL-BH(表 12 中显示的序列))。

2. 从质粒 pCR2.1 上用限制性酶 SstI 与 HindIII 切下编码 C 端 480-666 位氨基酸的 H-SemaLcDNA 部分，并且连接到载体 pQE31(已用 SstI 和 HindIII 切割)(Qiagen)上 (构建体 pQE31-H-SemaL-SH(表 13 中显示的序列))。

15 通过 DNA 测序检查序列在正确读框中的正确整合。通过 Ni²⁺ 亲和性层析纯化由 N 端 6x 组氨酸标记和部分信息素 H-SemaL 组成的融合蛋白。纯化的融合蛋白用于对各种动物(兔、小鸡、小鼠)的免疫。

实施例 9：

各种细胞类型的 FACS 分析(图 4 和 5)

20 在 FACS 缓冲液(含有 5%胎牛血清(FCS)和 0.1%Na 叠氮化钠的磷酸缓冲盐水(PBS)中冲洗细胞(大约 $0.2-0.5 \times 10^6$)，然后在每种情况中与抗血清(在冰上)培养 1 小时。用于对照(上层小鸡预免疫前血清(1: 50))和特异性检测(特异性染色)的初级抗体为 H-SemaL-特异性小鸡抗血清(1: 50)。用通过镍整合亲和性层析(如实施例 8 所描述的)纯化的蛋白质免疫小鸡产生含有抗 H-SemaL 氨基酸(Aa)179-378(具有 N 端 His 标记)抗体的特异性抗血清。所使用的次级抗体是来自兔的 FITC-标记抗小鸡 F(ab')抗体(Dianova Jackson 实验室, Order No. 303-095-006, 汉堡 德国)(1mg/ml)。FITC-标记的兔抗小鼠 IgG 用于 CD100 染色。次级抗体每种情况下在 FACS 缓冲液中以 1: 50 进行稀释。

然后冲洗细胞。在 PBS 中进行重悬浮，并且在 FACS 中分析。利用 FACS-跟

踪仪(Becton-Dickinson)进行FACS 分析。原理: 单细胞悬浮液穿过一个测量通道, 其中以 488 nm 的激光光线照射这些细胞, 从而激发荧光染料(FITC)。测量散射正向光线(正向散射, FSC: 与细胞大小相关), 和其侧向光线(侧向散射, SSC: 与颗粒量相关: 不同细胞类型有差异)以及通道 1 的荧光(FL1)(FITC 发射范围的波长, 530 nm 处最大值)。以这种方法每次测量 10,000 件(细胞)。

点印迹(图 4a-k)(左图): SSC 对 FSC (大小对颗粒量/散射), 右边窗口中分析了(界线内)相似大小和颗粒含量的(均一)细胞群。右边窗口显示了 FL1 的强度(X 轴)对事件数(Y 轴), 也就是说频度分布。

在每个这样的图中, 对照血清的结果(不光滑曲线)叠加在特异性染色(光滑曲线)的结果上面。与对照相比较特异性染色的曲线的右移相当于 H-SemaL 在相应细胞中的表达。一种移出越大意味着更强的表达。

用于 FACS 分析的细胞系:

a) U937 细胞系

15 美国典型培养物保藏中心 ATCC ; ATCC 号: CRL-1593
 名称 U-937
 组织: 淋巴瘤; 组织细胞; 单细胞样
 物种: 人;
 保藏者: H. Koren

b) THP-1 细胞系

20 ATCC 保藏号: TIE-202
 组织: 单核细胞; 急性单核细胞白血病
 物种: 人
 保藏者: S. Tsuchiya

c) K-562 细胞系

25 ATCC 保藏号: CCL-243
 组织: 慢性骨髓性白血病
 物种: 人;
 保藏者: H.T.Holden

d) L-428 细胞系

DSMZ-德意志微生物保藏中心, DSMZ No: ACC 197

细胞类型: 人 Hodgkin 淋巴瘤

e) Jurkat 细胞系

DSMZ-德意志微生物保藏中心, DSMZ No: ACC 282

5 细胞类型: 人 T 细胞白血病

f) Daudi 细胞系

ATCC 保藏号: CCL-213

组织: Burkitt 淋巴瘤; B 淋巴母细胞; B 细胞

物种: 人;

10 保藏者: G. Klein

g) LCL 细胞系

EBV-转化的淋巴母细胞样 B-细胞系.

h) Jiyoye (P-2003)细胞系

ATCC 保藏号: CCL-87

15 组织: Burkitt 淋巴瘤; B 细胞 B 淋巴细胞

物种: 人

保藏者: W. Henle

i) CBL-Mix57

用重组体 H.Saimiri (没有缺失的野生型)转化的人 T-细胞系(从血液中分离)

20 j) CBL-Mix59

用 H.Saimiri (ORF71 缺失)转化的人 T-细胞系(从血液中分离).

实施例 10: 蛋白质凝胶和蛋白质印迹法

25 将可分泌的人 SEMA-L(表 4 的氨基酸 42-649(没有信号肽和跨膜结构域))克隆到质粒 pMelBac-A(Invitrogen De Schelp, Leck, 荷兰, Cv 1950-20)中, 这样就产生了质粒 pMelBacA-H-SemaL(长度 6622bp)(图 8). H-SemaL 衍生物在杆状病毒系统中表达(Bac-N-Blue, Invitrogen). 在从昆虫卵细胞 Sf9(来自 *Spodoptera frugiperda*)和 High Five™(来自粉纹夜蛾 U.S.Pat.No.5,300,435, 从 Invitrogen 购得)中通过具有重组体嗜斑纯化的杆状病毒的感染产生的细胞系中进行表达.

根据制造商的说明进行表达.

30 然后在凝胶中对蛋白质进行分级分离, 并用蛋白质印迹法检测 H-SemaL 衍生

关系在系统树中可以从它们的组群中推出。

图 4：在各种细胞系和各种细胞类型中的 H-SemaL 表达的 FACS 分析(比较实施例 8)。

图 5： CD100 和 H-SemaL 表达的比较分析(比较实施例 9)。

5 图 6： 在 HiFive 和 Sf9 细胞中可分泌的 SEMA-L(H-SemaL)的表达(比较实施例 10)。

在杆状病毒系统(Bac-N-Blue , Invitrogen)中的 pMelBac-A (Invitrogen)的 aa 42-649

用特异性小鸡抗血清 1(1 : 100)和抗-IgY-HRP 偶联物(1 : 3000 , 来自兔

10 Jackson Lab.) 的检测

1, 4, 6 未感染的 HiFive 细胞(无血清)

2, 3, 5, 7, 8 以重组体杆状病毒感染的 HiFive 细胞(无血清)

M Rainbow 分子量标志 (Amersham RPN 756)

9, 10 感染的 Sf9 细胞(包含血清的介质)。

15 图 7： 抗血清的特异性

泳道 1-3： 小鸡 1； 泳道 4-6： 小鸡 2

泳道 1 和泳道 4： 免疫前血清

泳道 2 和 5： 免疫 60 天的血清

泳道 4 和 6： 免疫 105 天的血清

20 用 H-SemaL 的 179-378 位氨基酸(具有氨基端 His 标记)进行免疫(比较实施例 8, 部分 1)。

图 8： 对 pMelBacA-H-SEMAL 的质粒图谱的描绘。

如实施例 11 的描述制备重组质粒。

表1 来自不同物种的信息素的不同亚型

名称	别名	物种	特性	参考文献
H-SemaIII	(H-SemaD)	人	Sec.	(Kokodkin 等, 1993)
CD-100		人	TM, IC, 与 CD45 相连, 在 T 细胞中表达	(Hall 等, 1996)
H-SemaV	(H-SemaA)	人	Sec.; 基因座 3p21.3	(Sekido 等, 1996; Roche 等, 1996)
H-SemaIV	(H-Sema3F)	人	Sec.; 基因座 3p21.3	(Xiang 等, 1996; Sekido 等, 1996)
H-SemaE		人	Sec.; 不同于 M-Sema-E 的 3' 端 (提高对比)	AB000220 (Yamada, 1997 未发表)
H-SemaK	KIAA0311	人	Sec.	(Nagase 等, 1997)
H-SemaL	SEMAL	人	TM, 无 IC	本申请
M-SemaA		小鼠	Sec.	(Puschel 等, 1995)
M-SemaB		小鼠	TM, IC	(Puschel 等, 1995)
M-SemaC		小鼠	TM, IC	(Puschel 等, 1995)
M-SemaD	M-SemaIII	小鼠	Sec.	(Messersmith 等, 1995; Puschel 等, 1995)
M-SemaE		小鼠	Sec., 5' 部分序列	(Puschel 等, 1995)
M-SemaF1	M-SemaF	小鼠	TM, IC	(Inagaki 等, 1995)
M-SemaG2	M-SemaG	小鼠	TM, IC, 在淋巴组织中表达 CD100 的小鼠同系物	(Furuyama 等, 1996)
M-SemaF2	M-SemaF	小鼠	TM, IC, 凝血调节蛋白基因	(Adams 等, 1996)
M-SemaG1	M-SemaG	小鼠	TM, IC, 凝血调节蛋白基因	(Adams 等, 1996)
M-SemaH		小鼠	Sec.	(Christensen, 1996 未发表; Z80941)
M-SemaVia		小鼠	TM, IC	(Zhou 等, 1997)
M-SemaL	Semal	小鼠	部分序列	本申请
豚鼠蛋白-1		小鼠	Sec.	(Luo 等, 1993)
豚鼠蛋白-2		小鼠	Sec.	(Luo 等, 1995)
豚鼠蛋白-3		小鼠	Sec.	(Luo 等, 1995)
豚鼠蛋白-4		小鼠	部分序列	(Luo 等, 1995)
豚鼠蛋白-5		小鼠	Sec.	(Luo 等, 1995)
R-SemaIII		大鼠	Sec.	(Giger 等, 1996)
T-Semal		Tribolium confusum	TM, IC	(Kokodkin 等, 1993)
Ce-Semal		C. elegans	TM, IC	U15667 (Roy, 1994 未出版)
G-Semal	Fasciclin-IV	蜘蛛	TM, IC	(Kokodkin 等, 1993)
D-Semal		果蝇属	TM, IC	(Kokodkin 等, 1993)
D-SemaII		果蝇属	Sec.	(Kokodkin 等, 1993)
AHV-Sema		AHV-1	Sec.	(Ersser 和 Fleckenstein, 1995)
ORF-A39		牛痘	Sec.	(Kokodkin 等, 1993)
ORF-A39 同系物		天花	Sec.	(Kokodkin 等, 1993)

TM: 跨膜区 Sec: 分泌性 IC: 推定的胞内胞浆序列基元



表 2: H-SemaL 的 cDNA 序列(2636 个核苷酸)(SEQ ID NO.: 1)

1 cggggccacg ggatgacgcc tctccgccc ggacgtgccg cccccagcgc
51 accgcgcgcc cgcgccctg gcccgccggc tgggtgggg cttccgctgc
101 ggctgcgct gctgctgctg ctctggcgg cgcgcctc cgcccagggc
151 cacctaaagga gggaccccg catctcgcc gctcggaaag gccatgtagg
201 gcaggaccgg gtggacttg gccagactga gccgcacacg gtgctttcc
251 acgagccagg cagctcctct gtgtgggtgg gaggacgtgg caaggctac
301 ctcttgact tccccaggg caagaacgca tctgtcgca cggtaatat
351 cggctccaca aaggggtcct gtctggataa gcgggactgc gagaactaca
401 tcactctct ggagagggc agtgagggc tctggcctg tggaccaac
451 gcccgccacc ccagctgctg gaacctggtg aatggcactg tgggccact
501 tggcgagatg agaggctacg cccctcag cccggacgag aactccctgg
551 ttctgttga aggggacgag gtgtattcca ccatccgga gcaggaatac
601 aatgggaaga tccctcgggt ccgccgcatc cggggcgaga gtgagctga
651 caccagtgat actgtcatgc agaaccaca gttcatcaa gccaccatcg
701 tgcaccaaga ccaggctac gatgacaaga tctactact cttccgagag
751 gacaatctg acaagaatcc tgaggctcct ctcaatgtg cccgtgtgc
801 ccagttgtc aggggggacc aggtgggga aagttcactg tcagtctca
851 agtgaacac tttctgaaa gccatgctgg tatgcagtga tctgccacc
901 aacaagaact tcaacaggct gcaagacgtc ttctgctcc ctgacccag
951 cggccagtgg agggacacca gggctatgg tgtttctcc aaccctgga
1001 actactcagc cgtctgtg tattccctcg gtgacattga caaggtctc
1051 cgtacctct cactcaagg ctaccactca agcctcca acccgggc
1101 tggcaagtgc ctccagacc agcagccgat accacagag acctccagg
1151 tggctgaccg tcaccagag gtggcgaga gggtgagcc catggggct
1201 ctgaagacgc cattgtcca cttaaatac cactaccaga aagtggcgt
1251 tcaccgatg caagccagcc acggggagac cttcatgtg cttaccta
1301 ctacagacag gggcactat cacaagggtg tgaaccggg ggagcaggag
1351 cacagctcg cctcaacat catggagatc cagccctcc gccgcggc
1401 tgccatccag accatgtcgc tggatgctga gcggaggaag ctgtatgta
1451 gctcccagtg ggaggtgagc caggtgcccc tggacctgtg tgaggtctat
1501 ggcgggggct gccacggtg cctcatgtcc cgagaccct actgoggct
1551 ggaccagggc cgctgcatct ccatctacag ctccgaacgg tcagtgtgc
1601 aatccaltaa tccagccgag ccacacaagg agtgtccaa ccccaacca
1651 gacaaggccc cactgcagaa gtttccctg gcccacaaact ctgctacta
1701 cctgagctgc cccatggaat cccgccacgc cacctactca tggcgccaca
1751 aggagaacgt ggagcagagc tgccaacctg gtcaccagag cccaactgc

1801 atcctgttca tcgagaacct cacggcgag cagtacggcc actactctg
 1851 cgagggcccag gagggtcct acttccgca ggctcagcac tggcagctgc
 1901 tgcccaggga cggcatcatg gccgagcacc tgctgggtca tgcctgtgcc
 1951 ctggctgctt ccctctggct gggggctg cccacactca ctcttgctt
 2001 gctggtccac tagggcctcc cgaggctggg catgcctcag gcttctgag
 2051 cccagggcac tagaacgtct cacactcaga gccggctggc ccgggagctc
 2101 ctgctgccc acttctcca ggggacagaa taaccagtg gaggatgcca
 2151 ggcttgaga cgtccagccg caggcggctg ctgggcccc ggtggcgcac
 2201 ggatggtgag gggctgagaa tgaggcacc gactgtgaag ctggggcatc
 2251 gatgaccaa gactttatct tctgaaaat attttcaga ctctcaaac
 2301 ttgactaat gcagcagtc tccagcccc agagcccatg ggtcggggag
 2351 tgggtttgga taggagagct gggactccat ctgaccctg gggctgaggc
 2401 ctgagtcct ctggactctt ggtaccaca ttgctcctt ccctccctc
 2451 tctcatggt ggggtgctgg tttctctgaa gaccagggc taccctctg
 2501 ccagccctgt cctctgcagc tcctctctg gtcctgggtc ccacaggaca
 2551 gccgcttgc atgtttatg aaggatgtt gcttccgga cggaaggacg
 2601 gaaaaagctc tgaaaaaaaa aaaaaaaaaa aaaaaa

5

10

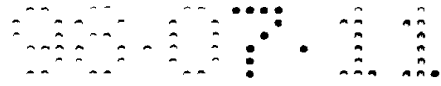
15 表 3: M-SemaL 的 cDNA 的核苷酸序列(部分, 1195 个核苷酸)(SEQ ID NO.:2)

1 cggggctgcg ggatgacgcc tctcctccc ggacgtgccg cccccagcgc
 51 accgcgcgcc cgcgtctca gccctccggc tgggtcggg ctcccgtgc
 101 ggctgcggct tctgctggtg ttctgggtgg ccgcccctc cgccaaggc
 151 cactcgagga gcggacccc catctccgcc gctggaaag ggcaggacca
 201 tgtggacttt agccagcctg agccacacac cgtgctttc catgagccgg
 251 gcagcttctc tctctgggtg ggtggacgtg gcaaggicta ccactcaac
 301 ttccccgagg gcaagaatgc ctctgtcgc acggtgaaca tggctccac
 351 aaaggggtcc tgtcaggaca aacaggactg tgggaattac atcactctc
 401 tagaaaggcg gggtaatggg ctgctggtct gtggcaccaa tgcccgaag
 451 cccagctgct ggaacttggg gaatgacagt gtggtgatgt cacttggta
 501 gatgaaaggc tatccccct tcagcccgga tgagaactcc ctggttctgt
 551 tgaaggaga tgaagtgtac tctaccatcc ggaagcagga atacaacggg
 601 aagatccctc ggttctgacg cattcggggc gagagtgaac tgtacacaag
 651 tgatacagtc atgcagaacc cacagttcat caaggccacc attgtgacc
 701 aagaccaagc ctatgatgat aagatctact acttctccg agaagacaac
 751 cctgacaaga accccgaggc tctctcaat gtgtcccag tagcccagtt

20

25

30



5
 801 gtcaggggg gaccaggtg gtgagagttc gttgtctgtc tccaagtgga
 851 acaccttct gaaagccatg ttgtctgca gcgatgcagc caccaacagg
 901 aactcaatc ggctgcaaga tgtcttctg ctccctgacc ccagtggcca
 951 gtggagagat accagggctt atggcgttt ctccaacccc tggaactact
 1001 cagctgtctg cgtgtattcg ctgggtgaca ttgacagagt ctccgtacc
 1051 tcatcgctca aaggctacca catgggcctt tccaaccctc gacctggcat
 1101 gtgcctcca aaaaagcagc ccataccac agaaacctc caggtagctg
 1151 atagtcaccc agaggtgct cagaggtgg aacctatggg gcccc

表 4: H-SemaL 的氨基酸序列(666 个氨基酸)(SEQ ID NO.: 3)

10
 1 MTPPPPGRAA PSAPRARVPG PPARLGLPLR LRLLLLLWAA AASAQGHLSR
 51 GPRIFAVWKG HVGQDRVDFG QTEPHTVLFH EPGSSSVWVG GRGKVYLFDF
 101 PEGKNASVRT VNIGSTKGSC LDKRDCENYI TLLERRSEGL LACGTNARHP
 151 SCWNLVNGTV VPLGEMRGYA PFSPDENSLV LFEGDEVYST IRKQEYNGKI
 201 PRFRIRIGES ELYTSDTVMQ NPQFIKATIV HQDQAYDDKI YFFREDNPD
 251 KNPEAPLNVS RVAQLCRGDQ GGESSLSVSK WNTFLKAMLV CSDAATNKNF
 301 NRLQDVFLLP DPSGQWRDTR VYGVFSNPWN YSAVCVYSLG DIDKVFRTSS
 15 LKGYHSSLPN PRPGKCLPDQ QPIPTETFQV ADRHPEVAQR VEPMGPLKTP
 401 LFHISKYHYQK VAVHRMQASH GETFHVLYLT TDRGTIHKV EPGEQEHSFA
 451 FNIMEIQPFR RAAAIQTMSL DAERRKLYVS SQWEVSQVPL DLCEVYGGGC
 501 HGCLMSRPY CGWDQGRSIS IYSSERSVLQ SINPAEPHKE CPNPKPKAP
 551 LQKVSLAPNS RYLLSCPMES RHATYSWRHK ENVEQSCEPG HQSPNCILFI
 601 ENLTAQQYGH YFCEAQEGSY FREAQHWQLL PEDGIMAEHL LGHACALAAS
 651 LWLGVLP TLT LGLLVH

20 表 5: M-SemaL 的(部分)氨基酸序列(394 个氨基酸, 相应于 H-SemaL 的 1-396 位)(SEQ ID NO.:4)

25
 1 MTPPPPGRAA PSAPRARVLS LPARFGLPLR LRLLLVFWA AASAQGHRSR
 51 GPRISAVWKG QDHVDFSQPE PHTVLFHEPG SFSVWVGGRG KVYHFNPEP
 101 KNASVRTVNI GSTKGCQDK QDCGNYITLL ERRGNLLVC GTNARKPSCW
 151 NLVNDVMS LGEMKGYAPF SPDENSLVLF EGDEVYSTIR KQEYNGKIPR
 201 FRRIRIGESL YTSDTVMQNP QFIKATIVHQ DQAYDDKIYY FFREDNPKN

251 PEAPLNVS RV AQLCRGDQGG ESSLSVSKWN TFLKAMLVCS DAATNRNFNR
 301 LQDVFLLPDP SGQWRDTRVY GVFSNPWNYS AVCVYSLGDI DRVFRTSSLK
 351 GYHMGLSNPR PGMCLPKKQP IPTETFQVAD SHPEVAQRVE PMGP

表 6: 合成寡核苷酸 (Eurogentec, Seraing, 比利时)

5	引物的编号/名称	引物(合成寡核苷酸)的核苷酸序列
	91506/AP2	actcactatagggctcgagcggc (SEQ ID NO.: 5)
	121234	agccgcacacgggtgctttc (SEQ ID NO.: 6)
	121235/Est 2	gcacagatgcgttctgccc (SEQ ID NO.: 7)
	121236/Est 3	accatagaccctgggtgcc (SEQ ID NO.: 8)
	121237/Est 4	gcagtgatgctgccaccaac (SEQ ID NO.: 9)
10	121238	ccagaccatgctcgtggatg (SEQ ID NO.: 10)
	121239/Est 6	acatgaggcaaccgtggcag (SEQ ID NO.: 11)
	131989/AP1	ccatcctaatacgaactcactatagggc (SEQ ID NO.: 12)
	131990/Est 7	aggtagacctgcccacgtcc (SEQ ID NO.: 13)
	131991	gaacttcaacaggctgcaagacg (SEQ ID NO.: 14)
	131992	atgctgagcggaggaagctg (SEQ ID NO.: 15)
15	131993	ccgccatacacctcacacag (SEQ ID NO.: 16)
	150788	ctggaagcttctgtgggtatcggctgc (SEQ ID NO.: 17)
	150789	tttgatccctggttctgttgaag (SEQ ID NO.: 18)
	167579/cDNA	ttclagaattcagcggccgctttttttttttttttttttttttttttttvn

合成引物(SEQ ID NO.:19)

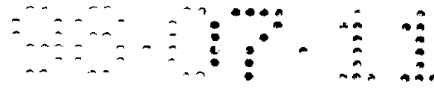
20	168421	ggggaaagttcactgtcagtcctcaag (SEQ ID NO.: 20)
	168422	gggaatacacacagacggctgagtag (SEQ ID NO.: 21)
	207608/	agcaagttcagcctggtaagt (SEQ ID NO.: 22)

λgt10 插入物的扩增

25	207609/	ttatgagtatttctccaggg (SEQ ID NO.: 23)
----	---------	---------------------------------------

λgt10 插入物的扩增

	232643/Est 13	ccaltaatccagccgagccacacaag (SEQ ID NO.: 24)
	232644/Est 14	catclacagctccgaacggctcagtg (SEQ ID NO.: 25)
	233084	cagcgggaagccccaaccgag (SEQ ID NO.: 26)
	240655/hs 5	gggatgacgcctcctccgcccgg (SEQ ID NO.: 27)
30	240656/hs 3	aagcttcacgtggaccagcaagccaagagtg (SEQ ID NO.: 28)
	240657/hs 3c	aagcttttccgtcctccgtccgg (SEQ ID NO.: 29)

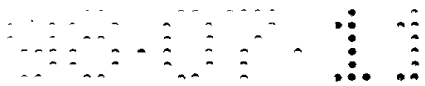


243068 atggtgagcaagggcgaggagctg (SEQ ID NO.: 30)
243069 ctgtacagctcgtccatgccgag (SEQ ID NO.: 31)
260812 GGGTGGTGAGAGTTCGTTGTCTGTC (SEQ ID NO.: 32)
260813 GAGCGATGAGGTACGGAAGACTCTG (SEQ ID NO.: 33)

5

表 7: 重组质粒 pCR2.1-H-SemaL 的核苷酸序列(SEQ ID NO.:34)

1 AGCGCCAAT ACGCAAACCG CCTCTCCCCG CGCGTTGGCC GATTCATTAA
51 TGCAGCTGGCACGACAGGTT TCCCGACTGG AAAGCGGGCA GTGAGCGCAA
101 CGCAATTAAT GTGAGTTAGC TCACTCATTG GGCACCCCAG GCTTTACT
151 TTATGCTTCC GGCTCGTATG TTGTGTGGAA TTGTGAGCGG ATAACAATTT
10 201 CACACAGGAA ACAGCTATGA CCATGATTAC GCCaagcttc acgtggacca
251 gcaagccaag agtgagtgtg ggcagcacc ccagccagag ggaggcagcc
301 agggcacagg catgaccag caggtgctcg gccatgatgc cglcctcggg
351 cagcagctgc cagtgtgag cctcgcggaa gtaggagccc tctgggct
401 cgcagaagta gtggccgtac tgctgcgccg tgaggttctc gatgaacagg
451 atgcagtgg ggctctgtg accaggttc cagctctct ccacgttctc
15 501 ctgtggcgc catgagtagg tggcgtggcg ggattccatg gggcagctca
551 ggtagtagcg agagttggg gccagggaaa ccttctgag tggggcctg
601 tctggttgg ggtgggaca ctctgtgt ggctcggctg gattaatgga
651 ttgcagcact gaccgtcgg agctgtgat ggagatgag cggccctgt
701 cccagccgca gtaggggtc cgggacatga ggcaaccgtg gcagccccg
751 ccatagacct cacacaggtc caggggcacc tggctcacct cccactggga
20 801 gctcacatac agcttctcc gctcagcacc cagcgacatg gctcggatg
851 cagccgcgcg gcggaagggc tggatctcca tgatgtgaa ggcaagctg
901 tgctcctct cccccgttc caccacctg tggatagtc cctgtctgt
951 agttaggtaa agcacaatgaa aggtctccc gtggctggct tgcagcggg
1001 gaacggccac ttctgttag tggatattg agtgaacaa tggcgtctc
1051 agaggccca tgggctccac cctctgcgcc acctctggg gacggctcagc
1101 cacctggaag gtctctgtg gctcggctg ctggctcggg aggcacttc
25 1151 cagcccgcg gttgggaagg ctgagtggt agccctgag tgaggagga
1201 cggagacct tgcataatg accgagggaa tacacacaga cggctgagta
1251 gttccagggg ttggagaaaa caccatagac cctgggttcc ctccactggc
1301 cgctggggtc agggagcagg aagacgtct gcagcctgtt gaagttctg
1351 ttgtggcag calcactgca taccagcatg gcttcagaa aagtttcca
1401 ctggagact gacagtgaac ttccccacc ctggctcccc ctgcacaact
30 1451 gggccacacg ggacacattg agaggagcct caggattct gtcaggattg

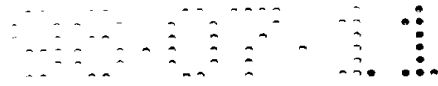


5

1501 tcctctcgga agaagtagta gatctgtca tcgtaagcct ggtcttggtg
1551 cacgatgggtg gctttgatga actgtgggtt ctgcatgaca gtatcactgg
1601 tgtacagctc actctcgccc cggatgcggc ggaaccgagg gatcttccca
1651 ttgtattcct gcttccggat ggtggaatac acctcgtccc ctcaaacag
1701 aaccagggag ttctcgtccg ggctgaaggg ggcgtagcct ccatctcgc
1751 caagtggcac cacagtgcca ttcaccaggf tccagcagct ggggtgccgg
1801 gcgttggtgc cacaggccag cagccccca ctccgcctct ccaggagagt
1851 gatgtagtic tcgcagctcc gcttatccag acaggacccc ttgtggagc
1901 cgatattcac cgtgcgcaca gatgcgttct tgcctcggg gaagtcaaag
1951 aggtagacct tgccacgtcc tcccaccac acagaggagc tgcttggtc
2001 gtggaaaagc accgtgtgag gctcagctg gccaaagtc acccgtcct
2051 gccctacatg gccttccag acggcgaaga tgcggggctc gctccttagg
2101 tggccctggg cggaggcggc ggccgcccag agcagcagca gcagccgag
2151 ccgcagcggg agcccccaacc gagccggcgg gccagggagc cgggcgcgcg
2201 gtgcgctggg ggcggcacgt cggggcggag gaggcgtcat cccaagcga
2251 attcTGCAGA TATCCATCAC ACTGGCGGCC GCTCGAGCAT GCATCTAGAG
2301 GGCCCAATTC GCCCTATAGT GAGTCGTATT ACAATCACT GGCCGTCGTT
2351 TTACAACGTC GTGACTGGGA AAACCCTGGC GTTACCCAAC TTAATCGCCT
2401 TGCAGCACAT CCCCCTTTTCG CCAGCTGGCG TAATAGCGAA GAGGCCCGCA
2451 CCGATCGCCC TTCCAACAG TTGCGCAGCC TGAATGGCGA ATGGGACGCG
2501 CCCTGTAGCGGCGCATTAAAG CGCGGCGGGT GTGGTGGTTA CGCGCAGCGT
2551 GACCGCTACA CTTGCCAGCG CCCTAGCGCC CGCTCCTTTC GCTTCTTCC
2601 CTTCTTTTCT CGCCACGTTT GCGGGCTTTC CCCGTCAAGC TCTAAATCGG
2651 GGGCTCCCTT TAGGGTTCCG ATTTAGAGCT TTACGGCACC TCGACCGCAA
2701 AAAACTTGAT TTGGGTGATG GTTACAGTAG TGGGCCATCG CCCTGATAGA
2751 CGGTTTTTTCG CCCTTTGACG TTGGAGTCCA CGTTCTTTAA TAGTGGACTC
2801 TTGTTCCAAA CTGGAACAAC ACTCAACCCT ATCGCGGTCT ATTCTTTTGA
2851 TTTATAAGGG ATTTTGCCGA TTTCCGCCTA TTGGTTAAAA AATGAGCTGA
2901 TTTAACAAAT TCAGGGCGCA AGGGCTGCTA AAGGAACCGG AACACGTAGA
2951 AAGCCAGTCC GCAGAAACGG TGCTGACCCC GGATGAATGT CAGCTACTGG
3001 GCTATCTGGA CAAGGAAAA CGCAAGCGCA AAGAGAAAGC AGGTAGCTTG
3051 CAGTGGGCTT ACATGGCGAT AGCTAGACTG GCGGTTTTTA TGGACAGCAA
3101 GCGAACCGBAATTGCCAGCT GGGGCGCCCT CTGGTAAGGT TGGGAAGCCC
3151 TGCAAAGTAA ACTGGATGGC TTTCTTGCCG CCAAGGATCT GATGGCGCAG
3201 GGGATCAAGA TCTGATCAAG AGACAGGATG AGGATCGTTT CGCATGATTG
3251 AACAAAGATGG ATTGCACGCA GGTTCTCCGG CCGCTTGGGT GGAGAGGCTA
3301 TTCGGCTATG ACTGGGCACA ACAGACAATC GGCTGCTCTG ATGCCGCCGT
3351 GTTCCGGCTG TCAGCGCAGG GCGGCCCGGT TCTTTTTGTC AAGACCGACC

5

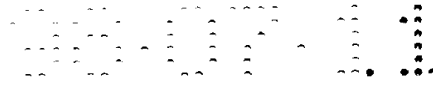
3401 TGTCCGGTGC CCTGAATGAA CTGCAGGACG AGGCAGCGCG GCTATCGTGG
3451 CTGGCCACGA CGGGCGTTCC TTGCGCAGCT GTGCTCGACG TTGTCACTGA
3501 AGCGGGAAGGGACTGGCTGCTATTGGGCGA AGTGCCGGGG CAGGATCTCC
3551 TGTCATCTCG CTTGCTCCT GCCGAGAAAG TATCCATCAT GGCTGATGCA
3601 ATGCGGCGGC TGCATACGCT TGATCCGGCT ACCTGCCCAT TCGACCACCA
3651 AGCGAAACAT CGCATCGAGC GAGCACGTAC TCGGATGGAA GCCGGTCTTG
3701 TCGATCAGGA TGATCTGGAC GAAGAGCATC AGGGGCTCGC GCCAGCCGAA
3751 CTGTTGCGCA GGCTCAAGGC GCGCATGCCG GACGGCGAGG ATCTCGTCGT
3801 GATCCATGGC GATGCCTGCT TGCCGAATAT CATGGTGGAA AATGGCCGCT
3851 TTTCTGGATT CAACGACTGT GGCCGGCTGG GTGTGGCGGA CCGCTATCAG
3901 GACATAGCGT TGGATACCCG TGATATTGCT GAAGAGCTTG GCGGCGAATG
3951 GGCTGACCGC TTCCTCGTGC TTTACGGTAT CGCCGCTCCC GATTGCGCAGC
4001 GCATCGCCTT CTATCGCCTT CTTGACGAGT TCTTCTGAAT TGAAAAAGGA
4051 AGAGTATGAG TATTCAACAT TTCCGTGTCG CCCTTATTCC CTTTTTGGC
4101 GCATTTTGCC TTCCTGTTTT TGCTCACCCA GAAACGCTGG TGAAAGTAA
4151 AGATGCTGAA GATCAGTTGG GTGCACGAGT GGGTTACATC GAACTGGATC
4201 TCAACAGCGG TAAGATCCTT GAGAGTTTTC GCCCCGAAGA ACGTTTTCCA
4251 ATGATGAGCA CTTTTAAAGT TCTGCTATGT CATACTAT TATCCCGTAT
4301 TGACGCCGGG CAAGAGCAAC TCGGTCGCCG GCGCGGTAT TCTCAGAATG
4351 ACTTGTTGA GTACTCACCA GTCACAGAAA AGCATCTTAC GGATGGCATG
4401 ACAGTAAGAG AATTATGCAG TGCTGCCATA ACCATGAGTG ATAACACTGC
4451 GGCCAACCTA CTTCTGACAA CGATCGGAGG ACCGAAGGAG CTAACCGCTT
4501 TTTTGCAAA CATGGGGGAT CATGTAACTC GCCTTGATCG TTGGGAACCG
4551 GAGCTGAATG AAGCCATACC AAACGACGAG AGTGACACCA CGATGCCTGT
4601 AGCAATGCCA ACAACGTTGC GCAAACCTATT AACTGGCGAA CTACTTACTC
4651 TAGCTTCCCG GCAACAATTA ATAGACTGGA TGGAGGCGGA TAAAGTTGCA
4701 GGACCACTTC TCGCTCGGC CTTCCGGCT GGCTGGTTTA TTGCTGATAA
4751 ATCTGGAGCC GGTGAGCGTG GGTCTCGCGG TATCATTGCA GCACTGGGGC
4801 CAGATGGTAA GCCCTCCCGT ATCGTAGTTA TCTACACGAC GGGGAGTCAG
4851 GCAACTATGG ATGAACGAAA TAGACAGATC GCTGAGATAG GTGCCTCACT
4901 GATTAAGCAT TGGTAACTGT CAGACCAAGT TACTCATAT AACTTTAGA
4951 TTGATTTAAA ACTTCATTTT TAATTTAAA GGATCTAGGT GAAGATCCTT
5001 TTTGATAATC TCATGACCAA AATCCCTTAA CGTGAGTTTT CGTTCCACTG
5051 AGCGTCAGAC CCCGTAGAAA AGATCAAAGG ATCTTCTTGA GATCCTTTTT
5101 TTCTGCGCGT AATCTGCTGC TTGCAAACAA AAAAACCACC GCTACCAGCG
5151 GTGGTTTGTG TGCCGGATCA AGAGCTACCA ACTCTTTTTT CGAAGGTAAC
5201 TGGCTTCAGC AGAGCGCAGA TACCAAATAC TGTCCTTCTA GTGTAGCCGT
5251 AGTTAGGCCA CCACTTCAAG AACTCTGTAG CACCGCCTAC ATACCTCGCT



5301 CTGCTAATCC TGTTACCACT GGCTGCTGCC AGTGGCGATA AGTCGTGTCT
 5351 TACCGGGTTG GACTCAAGAC GATAGTTACC GGATAAGGCG CAGCGGTCCG
 5401 GCTGAACGGG GGGTTCGTGC ACACAGCCCA GCTTGGAGCG AACGACCTAC
 5451 ACCGAACTGA GATACCTACA GCGTGAGCAT TGAGAAAGCG CCACGCTTCC
 5501 CGAAGGGAGAAAGGCGGACAGGTATCCGGT AAGCGGCAGG GTCGGAACAG
 5551 GAGAGCGCAC GAGGGAGCTT CCAGGGGGAA ACGCCTGGTA TCTTTATAGT
 5601 CCTGTCCGGT TTCGCCACCT CTGACTTGAG CGTCGATTTT TGTGATGCTC
 5651 GTCAGGGGGG CGGAGCCTAT GGAAAAACGC CAGCAACGCG GCCTTTTAC
 5701 GGTTCCTGGC CTTTGTCTGG CTTTTGCTC ACATGTTCTT TCCTGCGTTA
 5751 TCCCCTGATT CTGTGGATAA CCGTATTACC GCCTTTGAGT GAGCTGATAC
 5801 CGCTCGCCGACCCGAACGACCGAGCGCAG CGAGTCAGTG AGCGAGGAAG
 5851 CGGAAG

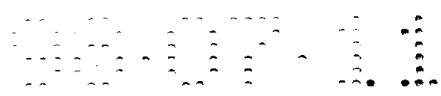
表 8：重组表达质粒 pCDNA3.1(-)H-SemaL-MycHisA 的核苷酸序列(SED ID NO.:35)

1 GACGGATCGG GAGATCTCCC GATCCCCTAT GGTCGACTCT CAGTACAATC
 51 TGCTCTGATG CCGCATAGTT AAGCCAGTAT CTGCTCCCTG CTTGTGTGTT
 101 GGAGGTCGCT GAGTAGTGCG CGAGCAAAAT TTAAGCTACA ACAAGGCAAG
 151 GCTTGACCGA CAATTGCATG AAGAATCTGC TTAGGGTTAG GCGTTTTGCG
 201 CTGCTTCGCG ATGTACGGGC CAGATATACG CGTTGACATT GATTATTGAC
 251 TAGTTATTAA TAGTAATCAA TTACGGGGTC ATTAGTTCAT AGCCCATATA
 301 TGGAGTTCCG CGTTACATAA CTTACGGTAA ATGGCCCGCC TGGCTGACCG
 351 CCCAACGACC CCCGCCATT GACGTCAATA ATGACGTATG TTCCCATAGT
 401 AACGCCAATA GGGACTTTC ATTGACGTCA ATGGGTGGAC TATTACGGT
 451 AAAGTCCCA CTTGGCAGTA CATCAAGTGT ATCATATGCC AAGTACGCC
 501 CCTATTGACG TCAATGACGG TAAATGGCCC GCCTGGCATT ATGCCAGTA
 551 CATGACCTTA TGGGACTTTC CTAATTGGCA GTACATCTAC GTATTAGTCA
 601 TCGCTATTAC CATGGTGATG CGGTTTTGGC AGTACATCAA TGGGCGTGGA
 651 TAGCGGTTTG ACTCACGGGG ATTTCCAAGT CTCCACCCCA TTGACGTCAA
 701 TGGGAGTTTG TTTTGGCACC AAAATCAACG GGAATTTCCA AAATGTCGTA
 751 ACAACTCCGCCCCATTGACGCAAAATGGGCGGTAGGCGTGT ACGGTGGGAG
 801 GTCTATATAA GCAGAGCTCT CTGGCTAACT AGAGAACCCA CTGCTTACTG
 851 GCTTATCGAA ATTAATACGA CTCACTATAG GGAGACCCAA GCTGGCTAGC
 901 GTTAAACGG GCCCTCTAGA CTCGAGCGGC CGCCACTGTG CTGGATATCT
 951 GCAgaattcg gcttgggatg acgcctcctc cgccccggacg tgccgcccc
 1001 agcgcaccgc ggcggcggt ccctggcccc cgggctcggg tggggcttcc



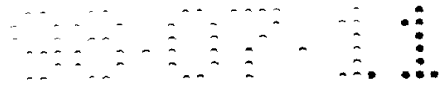
5

2951 ggcttgctgg tccacgtgaa gcttGGGCCC GAACAAAAAC TCATCTCAGA
3001 AGAGGATCTG AATAGCGCCG TCGACCATCA TCATCATCAT CATTGAGTTT
3051 AAACCGCTGA TCAGCCTCGA CTGTGCCTTC TAGTTGCCAG CCATCTGTTG
3101 TTTGCCCTC CCCCCTGCCT TCCTTGACCC TGAAGGTGC CACTCCCCT
3151 GTCCTTCTCT AATAAAATGA GGAAATTGCA TCGCATTGTC TGAGTAGGTG
3201 TCATTCTATTCTGGGGGGTG GGGTGGGGCA GGACAGCAAG GGGGAGGATT
3251 GGAAGACAA TAGCAGGCAT GCTGGGGATG CCGTGGGCTC TATGGCTTCT
3301 GAGGCGGAAAGAACCAGCTG GGGCTCTAGG GGTATCCCC ACGCGCCCTG
3351 TAGCGGCGCATTAAAGCGCGGGGTGTGTT GGTACGCGC AGCGTGACCG
3401 CTACACTTGC CAGCGCCCTA GCGCCCGCTC CTTTCGCTTT CTTCCCTTCC
3451 TTTCTCGCCA CGTTCGCCGG CTTTCCCCGT CAAGCTCTAA ATCGGGGCAT
3501 CCCTTTAGGG TTCCGATTTA GTGCTTTACG GCACCTCGAC CCCAAAAAC
3551 TTGATTAGGG TGATGGTTCA CGTAGTGGGC CATCGCCCTG ATAGACGGTT
3601 TTTCGCCCTT TGACGTTGGA GTCCACGTTT TTAATAGTG GACTCTTGTT
3651 CCAAACCTGGA ACAACTCA ACCCTATCTC GGTCTATTCT TTTGATTTAT
3701 AAGGGATTTT GGGGATTTTC GCCTATTGGT TAAAAATGA GCTGATTTAA
3751 CAAAAATTTA ACGCGAATTA ATTCTGTGGA ATGTGTGTCA GTTAGGGTGT
3801 GGAAAGTCCC CAGGCTCCCC AGGCAGGCAG AAGTATGCAA AGCATGCATC
3851 TCAATTAGTC AGCAACCAGG TGTGGAAAGT CCCCAGGCTC CCCAGCAGGC
3901 AGAAGTATGC AAAGCATGCA TCTCAATTAG TCAGCAACCA TAGTCCCGCC
3951 CCTAACTCCG CCCATCCCGC CCCTAACTCC GCCCAGTTCC GCCCATTCTC
4001 CGCCCCATGG CTGACTAATT TTTTTTATTT ATGCAGAGGC CGAGGCCGCC
4051 TCTGCCTCTG AGCTATTCCA GAAGTAGTGA GGAGGCTTTT TTGGAGGCTT
4101 AGGCTTTTGC AAAAAGCTCC CGGGAGCTTG TATATCCATT TTCGGATCTG
4151 ATCAAGAGAC AGGATGAGGA TCGTTTCGCA TGATTGAACA AGATGGATTG
4201 CACGCAGGTT CTCCGGCCGC TTGGGTGGAG AGGCTATTTC GCTATGACTG
4251 GGCACAACAG ACAATCGGCT GCTCTGATGC CGCCGTGTTT CGGCTGTCAG
4301 CGCAGGGGCG CCCGTTCTT TTTGTCAAGA CCGACCTGTC CCGTGCCCTG
4351 AATGAACTGCAGGACGAGGC AGCGCGGCTA TCGTGGCTGG CCACGACGGG
4401 CGTTCCTTGC GCAGCTGTGC TCGACGTTGT CACTGAAGCG GGAAGGGACT
4451 GGCTGCTATT GGGCGAAGTG CCGGGGCAGG ATCTCCTGTC ATCTCACCTT
4501 GCTCCTGCCG AGAAAGTATC CATCATGGCT GATGCAATGC GGCGGCTGCA
4551 TACGCTTGAT CCGGCTACCT GCCCATTCGA CCACCAAGCG AACATCGCA
4601 TCGAGCGAGC ACGTACTCGG ATGGAAGCCG GTCTTGTGCA TCAGGATGAT
4651 CTGGACGAAG AGCATCAGGG GCTCGCGCCA GCCGAACTGT TCGCCAGGCT
4701 CAAGGCGCGCATGCCCGACG GCGAGGATCT CGTCGTGACC CATGGCGATG
4751 CCTGCTTGCC GAATATCATG GTGGAAAATG GCCGCTTTTC TGGATTCATC
4801 GACTGTGGCCGGCTGGGTGT GGCGGACCGC TATCAGGACA TAGCGTTGGC



5

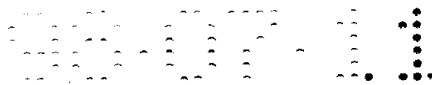
4851 TACCCGTGAT ATTGCTGAAG AGCTTGGCGG CGAATGGGCT GACCGCTTCC
4901 TCGTGCTTTA CGGTATCGCC GCTCCCGATT CGCAGCGCAT CGCCTTCTAT
4951 CGCCTTCTTG ACGAGTTCTT CTGAGCGGGA CTCTGGGGTT CGAAATGACC
5001 GACCAAGCGA CGCCCAACCT GCCATCACGA GATTTCGATT CCACCGCCGC
5051 CTTCTATGAA AGGTTGGGCT TCGGAATCGT TTTCCGGGAC GCCGGCTGGA
5101 TGATCCTCCA GCGCGGGGAT CTCATGCTGG AGTTCTTCGC CCACCCCAAC
5151 TTGTTTATTG CAGCTTATAA TGGTTACAAA TAAAGCAATA GCATCACAAA
5201 TTTCACAAAT AAAGCATTTT TTTCAGTCA TTCTAGTTGT GGTTTGTCCA
5251 AACTCATCAA TGTATCTTAT CATGTCTGTA TACCGTCGAC CTCTAGCTAG
5301 AGCTTGGCGT AATCATGGTC ATAGCTGTTT CCTGTGTGAA ATTGTTATCC
5351 GCTCACAATT CCACACAACA TACGAGCCGG AAGCATAAAG TGAAAGCCT
5401 GGGGTGCCTA ATGAGTGAGC TAACTCACAT TAATTGCGTT GCGCTCACTG
5451 CCCGCTTTCC AGTCGGGAAA CCTGTCGTGC CAGCTGCATT AATGAATCGG
5501 CCAACGCGCG GGGAGAGGCG GTTTGCGTAT TGGGCGCTCT TCCGCTTCTC
5551 CGCTCACTGA CTCGCTGCGC TCGGTCGTTT GGCTGCGGCG AGCGGTATCA
5601 GCTCACTCAA AGGCGGTAAT ACGGTTATCC ACAGAATCAG GGGATAACGC
5651 AGGAAAGAAC ATGTGAGCAA AAGGCCAGCA AAAGGCCAGG AACCGTAAAA
5701 AGGCCGCGTT GCTGGCGTTT TTCCATAGGC TCCGCCCCCC TGACGAGCAT
5751 CACAAAATC GACGCTCAAG TCAGAGGTGG CGAAACCCGA CAGGACTATA
5801 AAGATACCAG GCGTTTCCCC CTGGAAGCTC CCTCGTGCGC TCTCCTGTTT
5851 CGACCCTGCC GCTTACCGGA TACCTGTCCG CCTTTCTCCC TTCGGGAAGC
5901 GTGGCGCTTT CTCAATGCTC ACGCTGTAGG TATCTCAGTT CGGTGTAGGT
5951 CGTTCGCTCC AAGCTGGGCT GTGTGCACGA ACCCCCCGTT CAGCCCGACC
6001 GCTGCGCCTT ATCCGGTAAC TATCGTCTTG AGTCCAACCC GGTAAGACAC
6051 GACTTATCGC CACTGGCAGC AGCCACTGGT AACAGGATTA GCAGAGCGAG
6101 GTATGTAGGC GGTGCTACAG AGTTCTTGAA GTGGTGGCCT AACTACGGCT
6151 AACTAGAAG GACAGTATTT GGTATCTGCG CTCTGCTGAA GCCAGTTACC
6201 TTCGGAAAAA GAGTTGGTAG CTCTTGATCC GGCAAACAAA CCACCGCTGG
6251 TAGCGGTGGT TTTTTGTTT GCAAGCAGCA GATTACGCGC AGAAAAAAG
6301 GATCTCAAGA AGATCCTTTG ATCTTTTCTA CGGGGTCTGA CGCTCAGTGG
6351 AACGAAAAC CACGTTAAGG GATTTTGGTC ATGAGATTAT CAAAAAGGAT
6401 CTTACCTAG ATCCTTTTAA ATTAATAATG AAGTTTTAAA TCAATCTAAA
6451 GTATATATGA GTAACTTGG TCTGACAGTT ACCAATGCTT AATCAGTGAG
6501 GCACCTATCT CAGCGATCTG TCTATTTTCT TCATCCATAG TTGCCTGACT
6551 CCCCCTCGTG TAGATAACTA CGATACGGGA GGGCTTACCA TCTGGCCCCA
6601 GTGCTGCAAT GATACCGCGA GACCCACGCT CACCGGCTCC AGATTTATCA
6651 GCAATAAACC AGCCAGCCGG AAGGGCCGAG CGCAGAAGTG GTCCTGCAAC
6701 TTTATCCGCC TCCATCCAGT CTATTAATTG TTGCCGGGAA GCTAGAGTAA



6751 GTAGTTCGCC AGTTAATAGT TTGCGCAACG TTGTTGCCAT TGCTACAGGC
6801 ATCGTGGTGT CACGCTCGTC GTTTGGTATG GCTTCATTCA GCTCCGGTTC
6851 CCAACGATCA AGGCGAGTTA CATGATCCCC CATGTTGTGC AAAAAAGCGG
6901 TTAGCTCCTT CGGTCCTCCG ATCGTTGTCA GAAGTAAGTT GGCCGCAGTG
6951 TTATCACTCA TGGTTATGGC AGCACTGCAT AATTCTCTTA CTGTCAATGC
5 7001 ATCCGTAAGA TGCTTTTCTG TGAAGTGGTGA GACTCAACC AAGTCATTCT
7051 GAGAATAGTG TATGCGGCGA CCGAGTTGCT CTTGCCCGGC GTCAATACGG
7101 GATAATACCG CGCCACATAG CAGAAGTTTA AAAGTGCTCA TCATTGAAA
7151 ACGTTCTTCG GGGCGAAAAC TCTCAAGGAT CTTACCGCTG TTGAGATCCA
7201 GTTCGATGTA ACCCACTCGT GCACCCAACCT GATCTTCAGC ATCTTTTACT
7251 TTCACCAGCG TTTCTGGGTG AGCAAAAACA GGAAGGCAAA ATGCCGCAAA
7301 AAAGGGAATA AGGGCGACAC GGAAATGTTG AATACTCATA CTCTTCCTTT
10 7351 TTCAATATTA TTGAAGCATT TATCAGGGTT ATTGTCTCAT GAGCGGATAC
7401 ATATTTGAAT GTATTTAGAA AAATAAACA ATAGGGGTTC CGCGCACATT
7451 TCCCGAAAA GTGCCACCTG ACGTC

表 9: 重组质粒 pcDA3.1-SemaL-EGFP-MychisA 的核苷酸序列(SEQ ID NO.:36)

15 1 GACGGATCGG GAGATCTCCC GATCCCCTAT GGTCGACTCT CAGTACAATC
51 TGCTCTGATG CCGCATAGTT AAGCCAGTAT CTGCTCCCTG CTTGTGTGTT
101 GGAGGTGCGT GAGTAGTGCG CGAGCAAAAT TTAAGCTACA ACAAGGCAAG
151 GCTTGACCGA CAATTGCATG AAGAATCTGC TTAGGGTTAG GCGTTTTGCG
201 CTGCTTCGCG ATGTACGGGC CAGATATACG CGTTGACATT GATTATTGAC
251 TAGTTATTAA TAGTAATCAA TTACGGGGTC ATTAGTTCAT AGCCCATATA
20 301 TGGAGTCCG CGTTACATAA CTTACGGTAA ATGGCCCGCC TGGCTGACCG
351 CCAACGACC CCCGCCATT GACGTCAATA ATGACGTATG TTCCCATAGT
401 AACGCCAATA GGGACTTTC ATTGACGTCA ATGGGTGGAC TATTTACGGT
451 AAAGTCCCA CTTGGCAGTA CATCAAGTGT ATCATATGCC AAGTACGCCC
501 CCTATTGACG TCAATGACGG TAAATGGCCC GCCTGGCATT ATGCCAGTA
551 CATGACCTTA TGGGACTTTC CTAATTGGCA GTACATCTAC GTATTAGTCA
25 601 TCGCTATTAC CATGGTGTG CCGTTTTGGC AGTACATCAA TGGGCGTGGA
651 TAGCGGTTTG ACTCACGGGG ATTTCCAAGT CTCCACCCCA TTGACGTCAA
701 TGGGAGTTTG TTTTGGCACC AAAATCAACG GGAATTTCCA AAATGTCGTA
751 ACAACTCCGC CCCATTGACG CAAATGGGCG GTAGGCGTGT ACGGTGGGAG
801 GTCTATATAA GCAGAGCTCT CTGGCTAACT AGAGAACCCA CTGCTTACTG
851 GCTTATCGAA ATTAATACGA CTCACTATAG GGAGACCCAA GCTGGCTAGC
30 901 GTTAAACGG GCCCTCTAGA CTCGAGCGGC CGCCACTGTG CTGGATATCT



5

951 GCAgaattcg gcttgggatg acgcctctc cgcccggacg tccgcccc
1001 agcgcaccgc ggcgccgct ccctggcccg ccggctcggg tggggctcc
1051 gctgcggctg cggctgctg tctgctctg ggcggccgcc gcctccgccc
1101 agggccacct aaggagcggg ccccgcaict tcccgctg gaaaggccat
1151 gtagggcagg accgggtgga cttggccag actgagccgc acacggtgct
1201 ttccacgag ccaggcagct cctctgtg ggtgggagga cgtggcaagg
1251 tctacctt tgactcccc gagggcaaga acgcatctg ggcacggtg
1301 aatatggct ccacaaagg gtcctgctg gataagcggg actgcgagaa
1351 ctacatcact ctctggaga ggcggagtga gggctgctg gcctgtggca
1401 ccaacgcccg gcaccccagc tgctggaacc tggtaatgg cactgtggtg
1451 ccactggcg agatgagagg ctacgcccc ttcagcccgg acgagaactc
1501 cctggtctg ttgaagggg acgaggtgta ttccaccatc cggaagcagg
1551 aatacaatgg gaagatccct cggttccgcc gcatccgggg cgagagtgag
1601 ctgtacacca gtgatactgt catgcagaac ccacagtca tcaaagccac
1651 catcgtcac caagaccagg ctacgatga caagatctac tacttctcc
1701 gagaggacaa tctgacaag aatcctgagg ctctctcaa tgttcccgt
1751 gtggcccagt tgtcagggg ggaccagggt ggggaaagt cactgtcagt
1801 ctccaagtgg aacactttc tgaagccat gctggtatg agtgatgct
1851 ccaccaaaa gaacttaac aggtgcaag acgtctct gtcctctgac
1901 cccagcggcc agtggagggg caccagggtc tatggtgtt tctcaaccc
1951 ctggaactac tcagccgtct gtgtgtatt cctcgggtgac attgacaagg
2001 tctccgtac ctctcactc aagggtacc actcaagcct tccaacccg
2051 cggcctggca agtgctccc agaccagcag ccgataccca cagagacct
2101 ccagggtgct gaccgtcacc cagaggtggc gcagagggg gagcccatgg
2151 ggcctctgaa gacgccattg ttccactta aataccacta ccagaaagt
2201 gccgttacc gcatgcaagc cagccacggg gagaccttc atgtgctta
2251 cctaactaca gacaggggca ctatccaaa ggtggtggaa ccgggggagc
2301 aggagcacag ctgccttc aacatcatgg agatccagcc ctccgcccg
2351 gcggctgcca tccagaccat gtcgctggat gctgagcggg ggaagctgta
2401 tgtgagctcc cagtgggagg tgagccagg gcccctggac ctgtgtgagg
2451 tctatggcgg gggctgccac ggttcctca tgtcccgaga ccctactgc
2501 ggctgggacc agggccgctg catctcatc tacagctccg aacggtcagt
2551 gctgcaatcc attaatccag ccgagccaca caaggagtgt cccaacccca
2601 aaccagacaa ggcccactg cagaaggtt cctggcccc aaactctgc
2651 tactacctga gctgccccat ggaatcccgc cagccacct actcatggcg
2701 ccacaaggag aacgtggagc agagctgca acctggtcac cagagcccca
2751 actgatcct gttcatcgag aacctacgg cgcagcagta cggccactac
2801 ttctgcgagg cccaggagg ctctacttc cgcgaggctc agcactggca



2851 gctgctgccc gaggacggca tcatggccga gcacctgctg ggatcatgct
2901 gtgccctggc tgcctccctc tggctggggg tgctgccac actcactctt
2951 ggcttgctgg tccacATGGT GAGCAAGGGC GAGGAGCTGT TCACCGGGGT
3001 GGTGCCCATC CTGGTCGAGC TGGACGGCGA CGTAAACGGC CACAAGTTCA
3051 GCGTGTCCGGCGAGGGCGAGGGCGATGCCA CCTACGGCAA GCTGACCCTG
3101 AAGTTCATCT GCACCACGG CAAGCTGCCC GTGCCCTGGC CCACCCTCGT
3151 GACCACCCTG ACCTACGGCG TGCAGTGCTT CAGCCGCTAC CCCGACCACA
3201 TGAAGCAGCA CGACTTCTTC AAGTCCGCCA TGCCCGAAGG CTACGTCCAG
3251 GAGCGCACCA TCTTCTTCAA GGACGACGGC AACTACAAGA CCCGCGCCGA
3301 GGTGAAGTTCGAGGGCGACA CCCTGGTGAA CCGCATCGAG CTGAAGGGCA
3351 TCGACTTCAA GGAGGACGGC AACATCCTGG GGCACAAGCT GGAGTACAAC
3401 TACAACAGCC ACAACGTCTA TATCATGGCC GACAAGCAGA AGAACGGCAT
3451 CAAGGTGAAC TTCAAGATCC GCCACAACAT CGAGGACGGC AGCGTGCAGC
3501 TCGCCGACCA CTACCAGCAG AACACCCCCA TCGGCGACGG CCCCCTGCTG
3551 CTGCCCGACA ACCACTACCT GAGCACCCAG TCCGCCCTGA GCAAAGACCC
3601 CAACGAGAAG CGCGATCACA TGGTCCTGCT GGAGTTCGTG ACCGCCGCCG
3651 GGATCACTCT CGGCATGGAC GAGCTGTACA Aggtgaagct tGGGCCCGAA
3701 CAAAACTCA TCTCAGAAGA GGATCTGAAT AGCGCCGTGC ACCATCATCA
3751 TCATCATCAT TGAGTTTAAA CCGCTGATCA GCCTCGACTG TGCCTTCTAG
3801 TTGCCAGCCA TCTGTTGTTT GCCCCTCCCC CGTGCCTTCC TTGACCCTGG
3851 AAGGTGCCAC TCCCCTGTC CTTTCCTAAT AAAATGAGGA AATTGCATCG
3901 CATTGTCTGA GTAGGTGTCA TTCTATTCTG GGGGGTGGGG TGGGGCAGGA
3951 CAGCAAGGGGGAGGATTGGGAAGACAATAG CAGGCATGCT GGGGATGCGG
4001 TGGGCTCTATGGCTTCTGAG GCGGAAAGAA CCAGCTGGGG CTCTAGGGGG
4051 TATCCCCACGCGCCCTGTAG CGGCGCATTAGCGCGGCGG GTGTGGTGGT
4101 TACGCGCAGC GTGACCGCTA CACTTGCCAG CGCCCTAGCG CCCGCTCCTT
4151 TCGCTTTCTT CCCTTCCTTT CTCGCCACGT TCGCCGGCTT TCCCCGTCAA
4201 GCTCTAAATC GGGGCATCCC TTAGGGTTC CGATTTAGTG CTTTACGGCA
4251 CCTCGACCCC AAAAACTTG ATTAGGGTGA TGGTTCACGT AGTGGGCCAT
4301 CGCCCTGATA GACGGTTTTT CGCCCTTGA CGTTGGAGTC CACGTTCTTT
4351 AATAGTGGAC TCTTGTTCCA AACTGGAACA AACTCAACC CTATCTCGGT
4401 CTATCTTTT GATTTATAAG GGATTTTGGG GATTTCCGCC TATTGGTTAA
4451 AAAATGAGCT GATTTAACA AAATTTAACG CGAATTAATT CTGTGGAATG
4501 TGTGTCAGTT AGGGTGTGGA AAGTCCCCAG GCTCCCCAGG CAGGCAGAAG
4551 TATGCAAAGC ATGCATCTCA ATTAGTCAGC AACCAGGTGT GGAAAGTCCC
4601 CAGGCTCCCC AGCAGGCAGA AGTATGCAA GCATGCATCT CAATTAGTCA
4651 GCAACCATAG TCCCGCCCT AACTCCGCC ATCCCGCCCC TAACTCCGCC
4701 CAGTTCGCC CATTCTCCGC CCCATGGCTG ACTAATTTT TTTATTTATG

5

5

4751 CAGAGGCCGA GGCCGCCTCT GCCTCTGAGC TATTCCAGAA GTAGTGAGGA
4801 GGCTTTTTTG GAGGCCTAGG CTTTTGCAA AAGCTCCCGG GAGCTTGTAT
4851 ATCCATTTTC GGATCTGATC AAGAGACAGG ATGAGGATCG TTTGCGATGA
4901 TTGAACAAGA TGGATTGCAC GCAGGTTCTC CGGCCGCTTG GGTGGAGAGG
4951 CTATTGCGCT ATGACTGGGC ACAACAGACA ATCGGCTGCT CTGATGCCCG
5001 CGTGTTCCGG CTGTCAGCGC AGGGGCGCCC GTTCTTTTT GTCAAGACCG
5051 ACCTGTCCGG TGCCCTGAAT GAACTGCAGG ACGAGGCAGC GCGGCTATCG
5101 TGGCTGGCCA CGACGGGCGT TCCTTGCGCA GCTGTGCTCG ACGTTGTCAC
5151 TGAAGCGGAAGGGACTGGCTGCTATTGGG CGAAGTGCCG GGGCAGGATC
5201 TCCTGTCATC TCACCTTGCT CCTGCCGAGA AAGTATCCAT CATGGCTGAT
5251 GCAATGCGGC GGCTGCATAC GCTTGATCCG GCTACCTGCC CATTGACCA
5301 CCAAGCGAAA CATCGCATCG AGCGAGCACG TACTCGGATG GAAGCCGGTC
5351 TTGTCGATCA GGATGATCTG GACGAAGAGC ATCAGGGGCT CGCGCCAGCC
5401 GAACTGTTCC CCAGGCTCAA GGC GCGCATG CCCGACGGCG AGGATCTCGT
5451 CGTGACCCAT GCGGATGCCT GCTTGCCGAA TATCATGGTG GAAAATGGCC
5501 GCTTTTCTGG ATTCATCGAC TGTGGCCGGC TGGGTGTGGC GGACCGCTAT
5551 CAGGACATAG CGTTGGCTAC CCGTGATATT GCTGAAGAGC TTGGCGGCGA
5601 ATGGGCTGAC CGCTTCCTCG TGCTTTACGG TATCGCCGCT CCCGATTGCG
5651 AGCGCATCGC CTTCTATCGC CTTCTTGACG AGTTCTTCTG AGCGGGACTC
5701 TGGGGTTCGA AATGACCGAC CAAGCGACGC CCAACCTGCC ATCACGAGAT
5751 TTCGATTCCA CCGCCGCTT CTATGAAAGG TTGGGCTTCG GAATCGTTTT
5801 CCGGGACGCCGGCTGGATGA TCCTCCAGCG CGGGGATCTC ATGCTGGAGT
5851 TCTTCGCCCA CCCCACTTG TTTATTGCAG CTTATAATGG TTACAAATAA
5901 AGCAATAGCA TCACAAATTT CACAAATAAA GCATTTTTTT CACTGCATTC
5951 TAGTTGTGGT TTGTCCAAAC TCATCAATGT ATCTTATCAT GTCTGTATAC
6001 CGTCGACCTC TAGCTAGAGC TTGGCGTAAT CATGGTCATA GCTGTTTCCT
6051 GTGTGAAATT GTTATCCGCT CACAATTCCA CACAACATAC GAGCCGGAAG
6101 CATAAAGTGT AAAGCCTGGG GTGCCTAATG AGTGAGCTAA CTCACATTAA
6151 TTGCGTTGCG CTCACTGCCC GCTTTCCAGT CGGGAAACCT GTCGTGCCAG
6201 CTGCATTAAT GAATCGGCCA ACGCGCGGGG AGAGGCGGTT TCGTATTGG
6251 GCGCTCTTCC GCTTCCTCGC TCACTGACTC GCTGCGCTCG GTCGTTCCGGC
6301 TGCGGCGAGC GGTATCAGCT CACTCAAAGG CGGTAATACG GTTATCCACA
6351 GAATCAGGGG ATAACGCAGG AAAGAACATG TGAGCAAAAG GCCAGCAAAA
6401 GGCCAGGAAC CGTAAAAAGG CCGCGTTGCT GCGTTTTTTC CATAGGCTCC
6451 GCCCCCTGA CGAGCATCAC AAAAATCGAC GCTCAAGTCA GAGGTGGCGA
6501 AACCCGACAG GACTATAAAG ATACCAGGCG TTTCCCCTG GAAGCTCCCT
6551 CGTGCGCTCT CCTGTTCCGA CCCTGCCGCT TACCGGATAC CTGTCCGCCT
6601 TTCTCCCTTC GGAAGCGTG GCGCTTCTC AATGCTCACG CTGTAGGTAT

5

6651 CTCAGTTCGG TGTAGGTCGT TCGCTCCAAG CTGGGCTGTG TGCACGAACC
6701 CCCCCTTCAG CCGGACCGCT GCGCCTTATC CGGTAAGTAT CGTCTTGAGT
6751 CCAACCCGGT AAGACACGAC TTATCGCCAC TGGCAGCAGC CACTGGTAAC
6801 AGGATTAGCA GAGCGAGGTA TGTAGGCGGT GCTACAGAGT TCTTGAAGTG
6851 GTGGCCTAAC TACGGCTACA CTAGAAGGAC AGTATTTGGT ATCTGCGCTC
6901 TGCTGAAGCC AGTTACCTTC GGAAAAAGAG TTGGTAGCTC TTGATCCGGC
6951 AAACAAACCA CCGCTGGTAG CGGTGGTTTT TTTGTTTGCA AGCAGCAGAT
7001 TACGCGCAGA AAAAAAGGAT CTCAAGAAGA TCCTTTGATC TTTTCTACGG
7051 GGTCTGACGC TCAGTGGAAC GAAAACTCAC GTTAAGGGAT TTTGGTCATG
7101 AGATTATCAA AAAGGATCTT CACCTAGATC CTTTTAAATT AAAAATGAAG
7151 TTTTAAATCA ATCTAAAGTA TATATGAGTA AACTTGGTCT GACAGTTACC
7201 AATGCTTAAT CAGTGAGGCA CCTATCTCAG CGATCTGTCT ATTTTCGTTCA
7251 TCCATAGTTG CCTGACTCCC CGTCGTGTAG ATAACTACGA TACGGGAGGG
7301 CTTACCATCT GGCCCCAGTG CTGCAATGAT ACCGCGAGAC CCACGCTCAC
7351 CGGCTCCAGA TTTATCAGCA ATAAACCAGC CAGCCGGAAG GGCCGAGCGC
7401 AGAAGTGGTC CTGCAACTTT ATCCGCCTCC ATCCAGTCTA TTAATTGTTG
7451 CCGGGAAGCT AGAGTAAGTA GTTCGCCAGT TAATAGTTTG CGCAACGTTG
7501 TTGCCATTGC TACAGGCATC GTGGTGTAC GCTCGTCGTT TGGTATGGCT
7551 TCATTCAGCT CCGGTTCCCA ACGATCAAGG CGAGTTACAT GATCCCCCAT
7601 GTTGTGCAAA AAAGCGGTTA GCTCCTTCGG TCCTCCGATC GTTGTGAGAA
7651 GTAAGTTGGC CGCAGTGTTA TCACTCATGG TTATGGCAGC ACTGCATAAT
7701 TCTCTTACTG TCATGCCATC CGTAAGATGC TTTTCTGTGA CTGGTGAGTA
7751 CTCAACCAAG TCATTCTGAG AATAGTGTAT GCGGCGACCG AGTTGCTCTT
7801 GCCCGGCGTC AATACGGGAT AATACCGCGC CACATAGCAG AACTTTAAAA
7851 GTGCTCATCA TTGGAAAACG TTCTTCGGGG CGAAAACCTCT CAAGGATCTT
7901 ACCGCTGTTG AGATCCAGTT CGATGTAACC CACTCGTGCA CCCAACTGAT
7951 CTTCAGCATC TTTTACTTTC ACCAGCGTTT CTGGGTGAGC AAAAACAGGA
8001 AGGCAAAATG CCGCAAAAAA GGGAATAAGG GCGACACGGA AATGTTGAAT
8051 ACTCATACTC TTCCTTTTTT AATATTATTG AAGCATTAT CAGGGTTATT
8101 GTCTCATGAG CGGATACATA TTTGAATGTA TTTAGAAAAA TAAACAAATA
8151 GGGGTTCCGC GCACATTTCC CCGAAAAGTG CCACCTGACG TC

1751 tccactctaa ataccctac cagaaagtgg ccgltcaccg catgcaagcc
 1801 agccacgggg agaccttca tgtgcttac ctaactacag acaggggcac
 1851 latccacaag gtggtggaac cgggggagca ggagcacagc ttcgccttca
 1901 acatcatgga galccagccc tccgcccgc cggctgcat ccagaccatg
 1951 tcgctggatg ctgagcggag gaagctgtat gtgagctccc agtgggaggt
 2001 gagccagggt cccctggacc tgtgtgaggt ctatggcggg ggctgccacg
 2051 gttgcctcat gtcccagac ccctactgcg gctgggacca gggccgctgc
 2101 atctcatct acagctcca acggtcagtg ctgcaatcca ttaalccagc
 2151 cgagccacac aaggagtgtc ccaaccccaa accagacaag gccccactgc
 2201 agaaggttc cctggccca aactctcgt actacctgag ctgccccatg
 2251 gaatccgcc acgccacct cctatggcgc cacaaggaga acglggagca
 2301 gagctgcgaa cctggtcacc agagcccaa ctgcatctg ttcacgaga
 2351 acctcacggc gcagcaglac ggccactact tctgcgagc ccaggagggc
 2401 tctacttcc gcgaggctca gcactggcag ctgctgccc aggacggcat
 2451 catggccgag cacctgctg gtcctgctg tcccctggct gcctccctct
 2501 ggctgggggt gctgccaca ctactcttg gcttgctggt ccacgtgaag
 2551 cttGGGCCC TTTAAACCCG CTGATCAGCC TCGACTGTGC CTTCTAGTTG
 2601 CCAGCCATCT GTTGTTTGCC CCTCCCCCGT GCCTTCCTTG ACCCTGGAAG
 2651 GTGCCACTCC CACTGTCCTT TCCTAATAAA ATGAGGAAAT TGCATCGCAT
 2701 TGTCTGAGTA GGTGTCATTC TATTCTGGGG GGTGGGGTGG GGCAGGACAG
 2751 CAAGGGGGAGGATTGGGAAGACAATAGCAG GCATGCTGGG GATGCGGTGG
 2801 GCTCTATGGC TTCTGAGGCG GAAAGAACCA GCTGGGGCTC TAGGGGGTAT
 2851 CCCACGCGCCCTGTAGCGG CGCATTAAAGC GCGGCGGGTG TGGTGGTTAC
 2901 GCGCAGCGTG ACCGCTACAC TTGCCAGCGC CCTAGCGCCC GTCCTTTCCG
 2951 CTTTCTTCCC TTCTTTTCTC GCCACGTTCC CCGGCTTTCC CCGTCAAGCT
 3001 CTAATCGGG GCATCCCTTT AGGTTCCGA TTTAGTGCTT TACGGCACCT
 3051 CGACCCCAA AACTTGATT AGGGTGATGG TTCACGTAGT GGGCCATCGC
 3101 CCTGATAGAC GGTTTTTCGC CCTTTGACGT TGGAGTCCAC GTTCTTTAAT
 3151 AGTGGACTCT TGTTCCAAAC TGAACAACA CTCAACCCTA TCTCGGTCTA
 3201 TTCTTTTGAT TTATAAGGGA TTTTGGGGAT TTCGGCCTAT TGGTTAAAAA
 3251 ATGAGCTGAT TTAACAAAAA TTTAACGCGA ATTAATTCTG TGAATGTGT
 3301 GTCAGTTAGG GTGTGGAAAG TCCCAGGCT CCCAGGCAG GCAGAAGTAT
 3351 GCAAAGCATG CATCTCAATT AGTCAGCAAC CAGGTGTGGA AAGTCCCAG
 3401 GCTCCCCAGC AGGCAGAAGT ATGCAAAGCA TGCATCTCAA TTAGTCAGCA
 3451 ACCATAGTCC CGCCCTAAC TCCGCCATC CCGCCCTAA CTCCGCCAG
 3501 TTCCGCCAT TCTCCGCCCC ATGGCTGACT AATTTTTTTT ATTTATGCAG
 3551 AGGCCGAGGC CGCCTCTGCC TCTGAGCTAT TCCAGAAGTA GTGAGGAGGC
 3601 TTTTTTGAG GCCTAGGCTT TTGCAAAAAG CTCCCGGGAG CTTGTATATC

3651 CATTTCGGA TCTGATCAAG AGACAGGATG AGGATCGTTT CGCATGATTG
 3701 AACAAAGATGG ATTGCACGCA GGTTCCTCCGG CCGCTTGGGT GGAGAGGCTA
 3751 TTCGGCTATG ACTGGGCACA ACAGACAATC GGCTGCTCTG ATGCCGCCGT
 3801 GTTCCGGCTG TCAGCGCAGG GCGGCCCGGT TCTTTTTGTC AAGACCGACC
 3851 TGTCCGGTGC CCTGAATGAA CTGCAGGACG AGGCAGCGCG GCTATCGTGG
 3901 CTGGCCACGA CGGGCGTTCC TTGCGCAGCT GTGCTCGACG TTGTCACTGA
 3951 AGCGGGAAGGGACTGGCTGCTATTGGGCGA AGTGCCGGGG CAGGATCTCC
 4001 TGTCATCTCA CCTTGCTCCT GCCGAGAAAG TATCCATCAT GGCTGATGCA
 4051 ATGCGGCGGC TGCATACGCT TGATCCGGCT ACCTGCCCAT TCGACCACCA
 4101 AGCGAAACAT CGCATCGAGC GAGCACGTAC TCGGATGGAA GCCGGTCTTG
 4151 TCGATCAGGA TGATCTGGAC GAAGAGCATC AGGGGCTCGC GCCAGCCGAA
 4201 CTGTTCCGCA GGCTCAAGGC GCGCATGCC GACGGCGAGG ATCTCGTCGT
 4251 GACCCATGGC GATGCCTGCT TGCCGAATAT CATGGTGGAA AATGGCCGCT
 4301 TTTCTGGATT CATCGACTGT GGCCGGCTGG GTGTGGCGGA CCGCTATCAG
 4351 GACATAGCGT TGGCTACCCG TGATATTGCT GAAGAGCTTG GCGGCGAATG
 4401 GGCTGACCGC TTCCTCGTGC TTTACGGTAT CGCCGCTCCC GATTGCGAGC
 4451 GCATCGCCTT CTATCGCCTT CTTGACGAGT TCTTCTGAGC GGGACTCTGG
 4501 GGTTCGAAAT GACCGACCAA GCGACGCCA ACCTGCCATC ACGAGATTC
 4551 GATTCCACCG CCGCCTTCTA TGAAAGGTTG GGCTTCGGAA TCGTTTTCCG
 4601 GGACGCCGGC TGGATGATCC TCCAGCGCGG GGATCTCATG CTGGAGTTCT
 4651 TCGCCCACCC CAACTTGTTT ATTGCAGCTT ATAATGGTTA CAAATAAAGC
 4701 AATAGCATCA CAAATTTAC AAATAAAGCA TTTTTTTCAC TGCATTCTAG
 4751 TTGTGGTTTG TCCAAACTCA TCAATGTATC TTATCATGTC TGTATACCGT
 4801 CGACCTCTAG CTAGAGCTTG GCGTAATCAT GGTCATAGCT GTTTCCTGTG
 4851 TGAAATTGTT ATCCGCTCAC AATTCCACAC AACATACGAG CCGGAAGCAT
 4901 AAAGTGTAAG GCCTGGGGTG CCTAATGAGT GAGCTAACTC ACATTAATTG
 4951 CGTTGCGCTC ACTGCCCGCT TTCCAGTCGG GAAACCTGTC GTGCCAGCTG
 5001 CATTAAATGAA TCGGCCAACG CGCGGGGAGA GGCGGTTTGC GTATTGGGCG
 5051 CTCTTCCGCT TCCTCGCTCA CTGACTCGCT GCGCTCGGTC GTTCGGCTGC
 5101 GGCGAGCGGT ATCAGCTCAC TCAAAGGCGG TAATACGGTT ATCCACAGAA
 5151 TCAGGGGATA ACGCAGGAAA GAACATGTGA GCAAAGGCC AGCAAAGGC
 5201 CAGGAACCGT AAAAAGGCCG CGTTGCTGGC GTTTTTCCAT AGGCTCCGCC
 5251 CCCCTGACGA GCATCACAAA AATCGACGCT CAAGTCAGAG GTGGCGAAAC
 5301 CCGACAGGAC TATAAAGATA CCAGGCGTTT CCCCTGGAA GCTCCCTCGT
 5351 GCGCTCTCCT GTTCCGACCC TGCCGCTTAC CGGATACCTG TCCGCCTTTC
 5401 TCCCTTCGGG AAGCGTGGCG CTTTCTCAAT GCTCACGCTG TAGGTATCTC
 5451 AGTTCGGTGT AGGTCGTTCC CTCCAAGCTG GGCTGTGTGC ACGAACCCCC
 5501 CGTTCAGCCC GACCGCTGCG CTTATCCGG TAACTATCGT CTTGAGTCCA

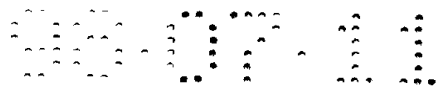
5551 ACCCGGTAAG ACACGACTTA TCGCCACTGG CAGCAGCCAC TGGTAACAGG
 5601 ATTAGCAGAG CGAGGTATGT AGGCGGTGCT ACAGAGTTCT TGAAGTGGTG
 5651 GCCTAACTAC GGCTACACTA GAAGGACAGT ATTTGGTATC TGCCTCTGCTGC
 5701 TGAAGCCAGT TACCTTCGGA AAAAGAGTTG GTAGCTCTTG ATCCGGCAAA
 5 5751 CAAACCACCG CTGGTAGCGG TGGTTTTTTT GTTTGAAGC AGCAGATTAC
 5801 GCGCAGAAAA AAAGGATCTC AAGAAGATCC TTTGATCTTT TCTACGGGGT
 5851 CTGACGCTCA GTGGAACGAA AACTCACGTT AAGGGATTTT GGTCATGAGA
 5901 TTATCAAAAA GGATCTTCAC CTAGATCCTT TAAATTAATAA AATGAAGTTT
 5951 TAAATCAATC TAAAGTATAT ATGAGTAAAC TTGGTCTGAC AGTTACCAAT
 6001 GCTTAATCAG TGAGGCACCT ATCTCAGCGA TCTGTCTATT TCGTTCATCC
 10 6051 ATAGTTGCCT GACTCCCCGT CGTGTAGATA ACTACGATAC GGGAGGGCTT
 6101 ACCATCTGGC CCCAGTGCTG CAATGATACC GCGAGACCCA CGCTCACCGG
 6151 CTCCAGATTT ATCAGCAATA AACCAGCCAG CCGGAAGGGC CGAGCGCAGA
 6201 AGTGGTCTCG CAACTTTATC CGCCTCCATC CAGTCTATTA ATTGTTGCCG
 6251 GGAAGCTAGA GTAAGTAGTT CGCCAGTTAA TAGTTTGCGC AACGTTGTTG
 6301 CCATTGCTAC AGGCATCGTG GTGTCACGCT CGTCGTTTGG TATGGCTTCA
 15 6351 TTCAGCTCCG GTTCCCAACG ATCAAGGCGA GTTACATGAT CCCCCATGTT
 6401 GTGCAAAAAA GCGGTTAGCT CCTTCGGTCC TCCGATCGTT GTCAGAAGTA
 6451 AGTTGGCCGC AGTGTTATCA CTCATGGTTA TGGCAGCACT GCATAATTCT
 6501 CTTACTGTCA TGCCATCCGT AAGATGCTTT TCTGTGACTG GTGAGTACTC
 6551 AACCAAGTCA TTCTGAGAAT AGTGATGCG GCGACCGAGT TGCTCTTGCC
 6601 CGGCGTCAAT ACGGGATAAT ACCGCGCCAC ATAGCAGAAC TTTAAAAGTG
 20 6651 CTCATCATTG GAAAACGTTT TTCGGGGCGA AACTCTCAA GGATCTTACC
 6701 GCTGTTGAGA TCCAGTTCGA TGTAACCCAC TCGTGCACCC AACTGATCTT
 6751 CAGCATCTTT TACTTTCACC AGCGTTTCTG GGTGAGCAAA AACAGGAAGG
 6801 CAAAATGCCG CAAAAAGGG AATAAGGGCG ACACGGAAAT GTTGAATACT
 6851 CATACTCTTC CTTTTCAAT ATTATTGAAG CATTATCAG GGTTATTGTC
 6901 TCATGAGCGG ATACATATTT GAATGTATTT AGAAAAATAA ACAAATAGGG
 25 6951 GTTCCGCGCA CATTCCCCG AAAAGTGCCA CCTGACGTCG ACGGATCGGG

表 11：重组质粒 pIND-H-SemaL-EA 的核苷酸序列(SEQ ID NO.:38)

1 AGATCTCGGC CGCATATTAA GTGCATTGTT CTCGATACCG CTAAGTGCAT
 51 TGTTCTCGTT AGCTCGATGG ACAAGTGCAT TGTTCTCTTG CTGAAAGCTC
 101 GATGGACAAG TGCATTGTTT TCTTGCTGAA AGCTCGATGG ACAAGTGCAT
 30 151 TGTTCTCTTG CTGAAAGCTC AGTACCCGGG AGTACCCTCG ACCGCCGGAG

201 TATAAATAGA GCGCTTCGT CTACGGAGCG ACAATTCAAT TCAAACAAGC
 251 AAAGTGAACA CGTCGCTAAG CGAAAGCTAA GCAAATAAAC AAGCGCAGCT
 301 GAACAAGCTA AACAATCTGC AGTAAAGTGC AAGTTAAAGT GAATCAATTA
 351 AAAGTAACCA GCAACCAAGT AAATCAACTG CAACTACTGA AATCTGCCAA
 401 GAAGTAATTA TTGAATACAA GAAGAGAACT CTGAATACTT TCAACAAGTT
 5 451 ACCGAGAAAG AAGAACTCAC ACACAGCTAG CGTTTAAACT TAAGCTTGGT
 501 ACCGAGCTCG GATCCACTAG TCCAGTGTGG TGgaattcgg cttgggatga
 551 cgcctcctcc gcccgagct gccgccccca ggcaccgcg cggccgctc
 601 cctggcccg cggctcgggt ggggcttccg ctgcccgtgc ggctgctgct
 651 gctgctctgg gcggccgccc cctccgccc gggccaccta aggagcggac
 701 cccgcatctt cgccgtctgg aaaggccatg tagggcagga cggggtggac
 751 ttggccaga ctgagccgca cacgggtgct ttccacgagc caggcagctc
 801 ctctgtgtgg gtgggaggac gtggcaaggt ctaccttt gacttccccg
 851 agggcaagaa cgcatctgtg cgcacgggta atatcggtc cacaaagggg
 901 tcctgtctgg ataagcggga ctgcgagaac tacatcactc tcctggagag
 951 ggggagttag gggctgctgg cctgtggcac caacgcccgg caccacagct
 1001 gctggaacct ggtgaatggc actgtggtgc cacttggcga gatgagaggc
 1051 tacgccccct tcagcccgga cgagaactcc ctggtctgt tgaagggga
 1101 cgagggtgat tccacatcc ggaagcagga atacaatggg aagatccctc
 1151 ggttccgccc catccggggc gagagtgagc tgtacaccag tgatactgtc
 1201 atgcagaacc cacagttcat caaagccacc atcgtgcacc aagaccaggg
 1251 ttacgatgac aagatctact acttctccg agaggacaat cctgacaaga
 1301 atcctgaggc tcctctcaat gtgtcccgtg tggcccagtt gtgcaggggg
 1351 gaccaggggt gggaaagttc actgtcagtc tccaagtga acacttttct
 1401 gaaagccatg ctggtatgca gtgatgctgc caccaacaag aacttcaaca
 1451 ggctgcaaga cgtcttctg ctccctgacc ccagcggcca gtggagggac
 1501 accagggctc atggtgttt ctccaacccc tgaactact cagccgtctg
 1551 tgtgtattcc ctggtgaca ttgacaaggt cttccgtacc tcctcactca
 1601 agggctacca ctcaagcct cccaaccgc ggcctggcaa gtgcctcca
 1651 gaccagcagc cgataccac agagaccttc caggtggctg accgtcacc
 1701 agaggtggcg cagaggggtg agccatggg gcctctgaag acgccattgt
 1751 tccactctaa ataccactac cagaaagtgg cgttcaccg catgcaagcc
 1801 agccacgggg agaccttca tgtgcttac ctaactacag acaggggac
 1851 tatccacaag gtggtggaac cgggggagca ggagcacagc ttgccttca
 1901 acatcatgga gatccagccc ttccgcccgc cggctgcat ccagaccatg
 1951 tcgctggatg ctgagcggag gaagctgat gtgagctccc agtgggaggt
 2001 gagccaggtg cccctggacc tgtgtaggt ctatggcggg ggctgccacg
 2051 gttgcctcat gtcccagac ccctactgcg gctgggacca gggccgctgc

2101 atctccatct acagctccga acggtcagtg ctgcaatcca ttaatccagc
 2151 cgagccacac aaggagtgtc ccaaccccaa accagacaag gccccactgc
 2201 agaaggttc cctggcccca aactctcgt actacctgag ctgccccatg
 2251 gaatccgcc acgccaccta ctcatggcgc cacaaggaga acgtggagca
 2301 gagctgcaa cctggtcacc agagcccaa ctgcatcctg tcatogaga
 2351 acctcacggc gcagcagtac ggccactact tctgcgagggc ccaggagggc
 2401 tcctactcc gcgaggctca gcaactggcag ctgctgcccg aggacggcat
 2451 catggccgag cacctgctgg gtcatgcctg tgccctggct gccctcctct
 2501 ggctgggggt gctgccaca ctactcttg gcttctggt ccacgtgaag
 2551 cttGGGCCCG AACAAAACT CATCTCAGAA GAGGATCTGA ATAGCGCCGT
 2601 CGACCATCAT CATCATCATC ATTGAGTTTA TCCAGCACAG TGGCGGCCGC
 2651 TCGAGTCTAG AGGGCCCGTT TAAACCCGCT GATCAGCCTC GACTGTGCCT
 2701 TCTAGTTGCC AGCCATCTGT TGTTTGCCCC TCCCCGTGC CTTCTTGAC
 2751 CCTGGAAGGT GCCACTCCA CTGTCCTTTC CTAATAAAT GAGGAAATTG
 2801 CATCGCATTG TCTGAGTAGG TGTCATTCTA TTCTGGGGGG TGGGGTGGGG
 2851 CAGGACAGCAAGGGGGAGGA TTGGGAAGAC AATAGCAGGC ATGCTGGGGA
 2901 TCGGGTGGGC TCTATGGCTT CTGAGGCGGA AAGAACCAGC TGGGGCTCTA
 2951 GGGGGTATCCCCACGCGCCCTGTAGCGGCG CATTAAAGCGC GGCGGGTGTG
 3001 GTGGTTACGC GCAGCGTGAC CGCTACACTT GCCAGCGCCC TAGCGCCCGC
 3051 TCCTTTCGCT TTCTTCCCTT CCTTTCTCGC CACGTTGCGC GGCTTTCGCC
 3101 GTCAAGCTCT AAATCGGGGC ATCCCTTTAG GGTCCGATT TAGTGCTTTA
 3151 CGGCACCTCG ACCCCAAAAA ACTTGATTAG GGTGATGGTT CACGTAGTGG
 3201 GCCATCGCCC TGATAGACGG TTTTTCGCCC TTTGACGTTG GAGTCCACGT
 3251 TCTTTAATAG TGGACTCTTG TTCCAAACTG GAACAACACT CAACCCTATC
 3301 TCGGTCTATT CTTTTGATTT ATAAGGGATT TTGGGGATT CGGCCTATTG
 3351 GTTAAAAAAT GAGCTGATTT AACAAAAATT TAACGCGAAT TAATTCTGTG
 3401 GAATGTGTGT CAGTTAGGGT GTGGAAAGTC CCCAGGCTCC CCAGGCAGGC
 3451 AGAAGTATGC AAAGCATGCA TCTCAATTAG TCAGCAACCA GGTGTGGAAA
 3501 GTCCCCAGGC TCCCCAGCAG GCAGAAGTAT GCAAAGCATG CATCTCAATT
 3551 AGTCAGCAAC CATAGTCCCG CCCCTAACTC CGCCCATCCC GCCCCTAACT
 3601 CCGCCCAGTT CCGCCCATT CCGCCCCAT GGCTGACTAA TTTTTTTTAT
 3651 TTATGCAGAG GCCGAGGCCG CCTCTGCCTC TGAGCTATTC CAGAAGTAGT
 3701 GAGGAGGCTT TTTTGGAGGC CTAGGCTTTT GCAAAAAGCT CCCGGGAGCT
 3751 TGTATATCCA TTTTCGGATC TGATCAAGAG ACAGGATGAG GATCGTTTCG
 3801 CATGATTGAA CAAGATGGAT TGCACGCAGG TTCTCCGGCC GCTTGGGTGG
 3851 AGAGGCTATT CGGCTATGAC TGGGCACAAC AGACAATCGG CTGCTCTGAT
 3901 GCCGCCGTGT TCCGGCTGTC AGCGCAGGGG CGCCCGTTT TTTTGTCAA
 3951 GACCGACCTGTCCGGTGCCC TGAATGAACT GCAGGACGAG GCAGCGCGGC



4001 TATCGTGGCT GGCCACGACG GCGTTCCTT GCGCAGCTGT GCTCGACGTT
4051 GTCACTGAAGCGGGAAGGGA CTGGCTGCTA TTGGGCGAAG TGCCGGGGCA
4101 GGATCTCCTG TCATCTCACC TTGCTCCTGC CGAGAAAGTA TCCATCATGG
4151 CTGATGCAAT GCGGCGGCTG CACACGCTTG ATCCGGCTAC CTGCCCATTC
4201 GACCACCAAG CGAAACATCG CATCGAGCGA GCACGTAATC GGATGGAAGC
4251 CGGTCTTGTC GATCAGGATG ATCTGGACGA AGAGCATCAG GGGCTCGCGC
4301 CAGCCGAAGTTCGCCAGG CTCAAGGCGC GCATGCCCGA CGGCGAGGAT
4351 CTCGTCGTGA CCCATGGCGA TGCCTGCTTG CCGAATATCA TGGTGGAAAA
4401 TGGCCGCTTT TCTGGATTCA TCGACTGTGG CCGGCTGGGT GTGGCGGACC
4451 GCTATCAGGA CATAGCGTTG GCTACCCGTG ATATTGCTGA AGAGCTTGGC
4501 GGCGAATGGG CTGACCGCTT CCTCGTGCTT TACGGTATCG CCGCTCCCGA
4551 TTCGCAGCGC ATCGCCTTCT ATCGCCTTCT TGACGAGTTC TTCTGAGCGG
4601 GACTCTGGGG TTCGAAATGA CCGACCAAGC GACGCCCAAC CTGCCATCAC
4651 GAGATTCGA TTCCACCGCC GCCTTCTATG AAAGGTTGGG CTTCGGAATC
4701 GTTTCCGGG ACGCCGGCTG GATGATCCTC CAGCGCGGGG ATCTCATGCT
4751 GGAGTTCTTC GCCACCCCA ACTTGTTTAT TGCAGCTTAT AATGGTTACA
4801 AATAAAGCAA TAGCATCACA AATTCACAA ATAAAGCATT TTTTCACTG
4851 CATTCTAGTT GTGGTTTGTG CAAACTCATC AATGTATCTT ATCATGTCTG
4901 TATACCGTCG ACCTCTAGCT AGAGCTTGGC GTAATCATGG TCATAGCTGT
4951 TTCCTGTGTG AAATTGTTAT CCGCTCACAA TTCCACACAA CATACGAGCC
5001 GGAAGCATAA AGTGTAAGC CTGGGGTGCC TAATGAGTGA GCTAACTCAC
5051 ATTAATTGCG TTGCGCTCAC TGCCCGCTTT CCAGTCGGGA AACCTGTCGT
5101 GCCAGCTGCA TTAATGAATC GGCCAACGCG CGGGGAGAGG CGGTTTGGCT
5151 ATTGGGCGCT CTTCCGCTTC CTCGCTCACT GACTCGCTGC GCTCGGTCGT
5201 TCGGCTGCGG CGAGCGGTAT CAGCTCACTC AAAGGCGGTA ATACGGTTAT
5251 CCACAGAATC AGGGGATAAC GCAGGAAAGA ACATGTGAGC AAAAGGCCAG
5301 CAAAAGGCCA GGAACCGTAA AAAGGCCGCG TTGCTGGCGT TTTTCCATAG
5351 GCTCCGCCCC CCTGACGAGC ATCACAAAAA TCGACGCTCA AGTCAGAGGT
5401 GGCGAAACCC GACAGGACTA TAAAGATACC AGGCGTTTCC CCCTGGAAGC
5451 TCCCTCGTGC GCTCTCCTGT TCCGACCCTG CCGCTTACCG GATACCTGTC
5501 CGCCTTTCTC CCTTCGGGAA GCGTGGCGCT TTCTCAATGC TCACGCTGTA
5551 GGTATCTCAG TTCGGTGTAG GTCGTTGCT CCAAGCTGGG CTGTGTGCAC
5601 GAACCCCCCG TTCAGCCCGA CCGCTGCGCC TTATCCGGTA ACTATCGTCT
5651 TGAGTCCAAC CCGGTAAGAC ACGACTTATC GCCACTGGCA GCAGCCACTG
5701 GTAACAGGAT TAGCAGAGCG AGGTATGTAG GCGGTGCTAC AGAGTTCTTG
5751 AAGTGGTGGC CTAACACGG CTACACTAGA AGGACAGTAT TTGGTATCTG
5801 CGCTCTGCTG AAGCCAGTTA CCTTCGGAAA AAGAGTTGGT AGCTCTTGAT
5851 CCGGCAAACA AACCACCGCT GGTAGCGGTG GTTTTTTGT TTGCAAGCAG

5

5901 CAGATTACGC GCAGAAAAA AGGATCTCAA GAAGATCCTT TGATCTTTTC
 5951 TACGGGGTCT GACGCTCAGT GGAACGAAAA CTCACGTTAA GGGATTTTGG
 6001 TCATGAGATT ATCAAAAAGG ATCTTCACCT AGATCCTTTT AAATTA AAAA
 6051 TGAAGTTTTA AATCAATCTA AAGTATATAT GAGTAACTT GGTCTGACAG
 5 6101 TTACCAATGC TTAATCAGTG AGGCACCTAT CTCAGCGATC TGTCTATTTT
 6151 GTTCATCCAT AGTTGCCTGA CTCCCCGTCG TGTAGATAAC TACGATACGG
 6201 GAGGGCTTAC CATCTGGCCC CAGTGCTGCA ATGATACCGC GAGACCCACG
 6251 CTCACCGGCT CCAGATTTAT CAGCAATAAA CCAGCCAGCC GGAAGGGCCG
 6301 AGCGCAGAAG TGGTCCTGCA ACTTTATCCG CCTCCATCCA GTCTATTAAT
 6351 TGTTGCCGGG AAGCTAGAGT AAGTAGTTCG CCAGTTAATA GTTTGCGCAA
 10 6401 CGTTGTTGCC ATTGCTACAG GCATCGTGGT GTCACGCTCG TCGTTTGGTA
 6451 TGGCTTCATT CAGCTCCGGT TCCAACGAT CAAGGCGAGT TACATGATCC
 6501 CCCATGTTGT GCAAAAAGC GGTTAGCTCC TTCGGTCCTC CGATCGTTGT
 6551 CAGAAGTAAG TTGGCCGCGAG TGTTATCACT CATGGTTATG GCAGCACTGC
 6601 ATAATTCTCT TACTGTCATG CCATCCGTAA GATGCTTTTC TGTGACTGGT
 6651 GAGTACTCAA CCAAGTCATT CTGAGAATAG TGTATGCCGC GACCGAGTTG
 15 6701 CTCTTGCCCG GCGTCAATAC GGGATAATAC CGCGCCACAT AGCAGAACTT
 6751 TAAAAGTGCT CATCATTGGA AAACGTTCTT CGGGGCGAAA ACTCTCAAGG
 6801 ATCTTACCGC TGTTGAGATC CAGTTCGATG TAACCCACTC GTGCACCCAA
 6851 CTGATCTTCA GCATCTTTTA CTTTCACCAG CGTTTCTGGG TGAGCAAAAA
 6901 CAGGAAGGCA AAATGCCGCA AAAAAGGGAA TAAGGGCGAC ACGGAAATGT
 6951 TGAATACTCA TACTCTTCCT TTTTCAATAT TATTGAAGCA TTTATCAGGG
 20 7001 TTATTGTCTC ATGAGCGGAT ACATATTTGA ATGTATTTAG AAAAATAAAC
 7051 AAATAGGGGT TCCGCGCACA TTTCCCGAA AAGTGCCACC TGACGTCGAC
 7101 GGATCGGG

表 12：重组质粒 pQE30-H-SemaL-BH 的序列(SEQ ID NO.:39)

1 CTCGAGAAAT CATAAAAAAT TTATTTGCTT TGTGAGCGGA TAACAATTAT
 25 51 AATAGATTCA ATTGTGAGCG GATAACAATT TCACACAGAA TTCATTAAG
 101 AGGAGAAATT AACTATGAGA GGATCGCATC ACCATCACCA TCACGGAtcc
 151 ctggttctgt ttgaagggga cgaggtgtat tccaccatcc ggaagcagga
 201 atacaatggg aagatccctc ggttccgccc catccggggc gagagtgagc
 251 tglacaccag tgatactgic atgcagaacc cacagttcat caaagccacc
 301 atcgtgcacc aagaccaggc ttacgatgac aagatclact acttcttccg
 30 351 agaggacaat cctgacaaga atcctgaggc tctctcaat ggttcccggtg

401 tggcccagtt gtgcaggggg gaccaggggtg gggaaagttc actgtcagtc
 451 tccaagtgga acacttttct gaaagccatg ctggtatgca gtgatgctgc
 501 caccaacaag aacttcaaca ggctgcaaga cgtcttctcg ctccctgacc
 551 ccagcggcca gtggaggac accaggtct atggtgttt ctccaacccc
 601 tggaaactact cagccgtctg tgtgtattcc ctgggaca ttgacaaggt
 651 ctccgtacc tctcactca agggctacca ctcaagcctt cccaaccgcg
 701 ggctggcaa gtgctccca gaccagcagc cgataccac agaAAGCTTA
 751 ATTAGCTGAG CTTGGACTCC TGTTGATAGA TCCAGTAATG ACCTCAGAAC
 801 TCCATCTGGA TTTGTTCAGA ACGCTCGGTT GCCGCCGGGC GTTTTTTATT
 851 GGTGAGAATC CAAGCTAGCT TGGCGAGATT TTCAGGAGCT AAGGAAGCTA
 901 AAATGGAGAA AAAAATCACT GGATATACCA CCGTTGATAT ATCCCAATGG
 951 CATCGTAAAG AACATTTTGA GGCATTTTCA TCAAGTTGCTC AATGTACCTA
 1001 TAACCAGACC GTTCAGCTGG ATATTACGGC CTTTTTAAAG ACCGTAAAGA
 1051 AAAATAAGCA CAAGTTTTAT CCGGCCTTTA TTCACATTCT TGCCCGCCTG
 1101 ATGAATGCTC ATCCGGAATT TCGTATGGCA ATGAAAGACG GTGAGCTGGT
 1151 GATATGGGAT AGTGTTTACC CTTGTTACAC CGTTTTCCAT GAGCAAAGCTG
 1201 AACGTTTTTC ATCGCTCTGG AGTGAATACC ACGACGATTT CCGGCAGTTT
 1251 CTACACATAT ATTCGCAAGA TGTGGCGTGT TACGGTGAAA ACCTGGCCTA
 1301 TTTCCCTAAA GGGTTTATTG AGAATATGTT TTTCGTCTCA GCCAATCCCT
 1351 GGGTGAGTTT CACCAGTTTT GATTAAACG TGGCCAATAT GGACAACCTC
 1401 TTCGCCCCCG TTTTCACCAT GGGCAAATAT TATACGCAAG GCGACAAGGT
 1451 GCTGATGCCG CTGGCGATTC AGGTTTATCA TGCCGTCTGT GATGGCTTCC
 1501 ATGTCGGCAG AATGCTTAAT GAATTACAAC AGTACTGCGA TGAGTGGCAG
 1551 GGCGGGGCGT AATTTTTTTA AGGCAGTTAT TGGTGCCCTT AAACGCCTGG
 1601 GGTAATGACT CTCTAGCTTG AGGCATCAA TAAAACGAAA GGCTCAGTCG
 1651 AAAGACTGGG CCTTTCGTTT TATCTGTTGT TTGTCCGGTGA ACGCTCTCCT
 1701 GAGTAGGACA AATCCGCCGC TCTAGAGCTG CCTCGCGCGT TTCGGTGATG
 1751 ACGGTGAAAA CCTCTGACAC ATGCAGCTCC CGGAGACGGT CACAGCTTGT
 1801 CTGTAAGCGGATGCCGGGAGCAGACAAGCC CGTCAGGGCG CGTCAGCGGG
 1851 TGTTGGCGGGTGTCCGGGCG CAGCCATGAC CCAGTCACGT AGCGATAGCG
 1901 GAGTGTATAC TGGCTTAACT ATGCGGCATC AGAGCAGATT GACTGAGAG
 1951 TGCACCATAT GCGGTGTGAA ATACCGCACA GATGCGTAAG GAGAAAATAC
 2001 CGCATCAGGC GCTCTTCCGC TTCCTCGCTC ACTGACTCGC TGCGCTCGGT
 2051 CTGTCCGGCTG CGGCGAGCGG TATCAGCTCA CTCAAAGGCG GTAATACGGT
 2101 TATCCACAGA ATCAGGGGAT AACGCAGGAA AGAACATGTG AGCAAAAGGC
 2151 CAGCAAAGG CCAGGAACCG TAAAAGGCC GCGTTGCTGG CGTTTTTCCA
 2201 TAGGCTCCGC CCCCTGACG AGCATCACA AAATCGACGC TCAAGTCAGA
 2251 GGTGGCGAAA CCCGACAGGA CTATAAGAT ACCAGGCGTT TCCCCCTGGA

5

2301 AGCTCCCTCG TGCCTCTCC TGTTCGACC CTGCCGCTTA CCGGATACCT
 2351 GTCCGCCTTT CTCCCTTCGG GAAGCGTGGC GCTTTCTCAA TGCTCAGCT
 2401 GTAGGTATCT CAGTTCGGTG TAGGTCGTTT GCTCCAAGCT GGGCTGTGTG
 2451 CACGAACCCC CCGTTCAGCC CGACCGCTGC GCCTTATCCG GTAACATATCG
 2501 TCTTGAGTCC AACCCGGTAA GACACGACTT ATCGCCACTG GCAGCAGCCA
 2551 CTGGTAACAG GATTAGCAGA GCGAGGTATG TAGGCGGTGC TACAGAGTTC
 2601 TTGAAGTGGT GGCCTAACTA CGGCTACACT AGAAGGACAG TATTTGGTAT
 2651 CTGCGCTCTG CTGAAGCCAG TTACCTTCGG AAAAAGAGTT GGTAGCTCTT
 2701 GATCCGGCAA ACAACCACC GCTGGTAGCG GTGGTTTTTT TGTTTGCAAG
 2751 CAGCAGATTA CGCGCAGAAA AAAAGGATCT CAAGAAGATC CTTTGATCTT
 2801 TTCTACGGGG TCTGACGCTC AGTGGAACGA AACTCACGT TAAGGGATT
 2851 TGGTCATGAG ATTATCAAAA AGGATCTTCA CCTAGATCCT TTAAATTA
 2901 AAATGAAGTT TAAATCAAT CTAAAGTATA TATGAGTAA CTTGGTCTGA
 2951 CAGTTACCAA TGCTTAATCA GTGAGGCACC TATCTCAGCG ATCTGTCTAT
 3001 TTCGTTTATC CATAGCTGCC TACTCCCCG TCGTGTAGAT AACTACGATA
 3051 CGGGAGGGCT TACCATCTGG CCCAGTGCT GCAATGATAC CGCGAGACCC
 3101 ACGCTACCG GCTCCAGATT TATCAGCAAT AAACCAGCCA GCCGGAAGGG
 3151 CCGAGCGCAG AAGTGGTCCT GCAACTTTAT CCGCCTCCAT CCAGTCTATT
 3201 AATTGTTGCC GGAAGCTAG AGTAAGTAGT TCGCCAGTTA ATAGTTTGGC
 3251 CAACGTTGTT GCCATTGCTA CAGGCATCGT GGTGTCACGC TCGTCGTTTG
 3301 GTATGGCTTC ATCAGCTCC GGTCCCAAC GATCAAGGCG AGTTACATGA
 3351 TCCCCATGT TGTGCAAAA AGCGGTTAGC TCCTTCGGTC CTCCGATCGT
 3401 TGTCAGAAGT AAGTTGGCCG CAGTGTTATC ACTCATGGTT ATGGCAGCAC
 3451 TGCATAATTC TCTTACTGTC ATGCCATCCG TAAGATGCTT TTCTGTGACT
 3501 GGTGAGTACT CAACCAAGTC ATTCTGAGAA TAGTGTATGC GGCGACCGAG
 3551 TTGCTCTTGC CCGGCGTCAA TACGGGATAA TACCGCGCCA CATAGCAGAA
 3601 CTTTAAAAGT GCTCATCATT GGAAAACGTT CTTCGGGGCG AAAACTCTCA
 3651 AGGATCTTAC CGCTGTTGAG ATCCAGTTCG ATGTAACCCA CTCGTGCACC
 3701 CAACTGATCT TCAGCATCTT TTTACTTTCAC CAGCGTTTCT GGGTGAGCAA
 3751 AAACAGGAAG GCAAAATGCC GCAAAAAGG GAATAAGGGC GACACGGAAA
 3801 TGTTGAATAC TCATACTCTT CCTTTTTCAA TATTATTGAA GCATTTATCA
 3851 GGGTTATTGT CTCATGAGCG GATACATATT TGAATGTATT TAGAAAAATA
 3901 AACAAATAGG GGTCCGCGC ACATTTCCC GAAAAGTGCC ACCTGACGTC
 3951 TAAGAAACCA TTATTATCAT GACATTAACC TATAAAAATA GGCGTATCAC
 4001 GAGGCCCTTT CGTCTTCAC

表 13: 重组质粒 pQE31-H-SemaL-SH 的序列(SEQ ID NO.:40)

1 CTCGAGAAAT CATAAAAAAT TTATTTGCTT TGTGAGCGGA TAACAATTAT
51 AATAGATTCA ATTGTGAGCG GATAACAATT TCACACAGAA TTCATTAAG
101 AGGAGAAATT AACTATGAGA GGATCGCATC ACCATCACCA TCACACGGAT
5 151 CCGCATGCga gctcccagtg ggaggtgagc cagggtcccc tggacclgtg
201 tgaggtctat ggcgggggct gccacggttg cctcatgtcc cgagaccct
251 actcgggctg ggaccagggc cgctgcatct ccatctacag ctccgaacgg
301 tcagtgtctc aatccattaa tccagccgag ccacacaagg agtgtccaa
351 ccccaaacca gacaaggccc cactgcagaa ggtttccctg gcccact
401 ctgctacta cctgagctgc cccatggaat cccgccacgc cacctactca
451 tggcgccaca aggagaacgt ggagcagagc tgccaacctg gtcaccagag
501 ccccaactgc atcctgttca tcgagaacct cacggcgcag cagtacggcc
551 actactctg cgaggcccag gagggctcct acttccgca ggctcagcac
601 tggcagctgc tgcccagga cggcatcatg gccgagcacc tgcctgggtca
651 tgcctgtgcc ctggctgct cctctggct ggggggtgct cccactca
701 ctcttgctt gctggccac gtaagcttA ATTAGCTGAG CTTGGACTCC
751 TGTTGATAGA TCCAGTAATG ACCTCAGAAC TCCATCTGGA TTTGTTTCTGGA
801 ACGCTCGGTT GCCGCCGGGC GTTTTTTATT GGTGAGAATC CAAGCTAGCT
851 TGGCGAGATT TTCAGGAGCT AAGGAAGCTA AAATGGAGAA AAAAATCACT
901 GGATATACCA CCGTTGATAT ATCCCAATGG CATCGTAAAG AACATTTTGA
951 GGCATTTTCTG TCAGTTGCTC AATGTACCTA TAACCAGACC GTTCAGCTGG
1001 ATATTACGGC CTTTTTAAAG ACCGTAAAGA AAAATAAGCA CAAGTTTTAT
1051 CCGGCCTTTA TTCACATTCT TGCCCGCCTG ATGAATGCTC ATCCGGAATT
1101 TCGTATGGCA ATGAAAGACG GTGAGCTGGT GATATGGGAT AGTGTTTACC
1151 CTTGTTACAC CGTTTTCCAT GAGCAAATG AAACGTTTTT ATCGCTCTGG
1201 AGTGAATACC ACGACGATTT CCGGCAGTTT CTACACATAT ATTCGCAAGA
1251 TGTGGCGTGT TACGGTGAAA ACCTGGCCTA TTTCCCTAAA GGGTTTTATTG
1301 AGAATATGTT TTTCTGCTCA GCCAATCCCT GGGTGAGTTT CACCAGTTTT
1351 GATTTAAACG TGGCCAATAT GGACAACTT TCGCCCCCG TTTTACCAT
1401 GGGCAAATAT TATACGCAAG GCGACAAGGT GCTGATGCCG CTGGCGATTG
1451 AGGTTTATCA TGCCGTCTGT GATGGCTTCC ATGTCGGCAG AATGCTTAAT
1501 GAATTACAAC AGTACTGCGA TGAGTGGCAG GCGGGGGCGT AATTTTTTTA
1551 AGGCAGTTAT TGGTGCCCTT AAACGCCTGG GGTAATGACT CTCTAGCTTG
1601 AGGCATCAA TAAAACGAAA GGCTCAGTCG AAAGACTGGG CCTTTCGTTT
1651 TATCTGTTGT TTGTCGGTGA ACGCTCTCCT GAGTAGGACA AATCCGCCCG
1701 TCTAGAGCTG CCTCGCGCGT TTCGGTGATG ACGGTGAAAA CCTCTGACAC

5

1751 ATGCAGCTCC CGGAGACGGT CACAGCTTGT CTGTAAGCGG ATGCCGGGAG
1801 CAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTGGCGGG TGTCGGGGCG
1851 CAGCCATGAC CCAGTCACGT AGCGATAGCG GAGTGTATAC TGGCTTAACT
1901 ATGCGGCATC AGAGCAGATT GTECTGAGAG TGCACCATAT GCGGTGTGAA
1951 ATACCGCACA GATGCGTAAG GAGAAAATAC CGCATCAGGC GCTCTTCCGC
2001 TTCCTCGCTC ACTGACTCGC TGCCTCGGT CTGTCCGGCTG CGGCGAGCGG
2051 TATCAGCTCA CTCAAAGGCG GTAATACGGT TATCCACAGA ATCAGGGGAT
2101 AACGCAGGAA AGAACATGTG AGCAAAAGGC CAGCAAAAGG CCAGGAACCG
2151 TAAAAAGGCC GCGTTGCTGG CGTTTTTCCA TAGGCTCCGC CCCCTGACG
2201 AGCATCACAA AAATCGACGC TCAAGTCAGA GGTGGCGAAA CCGACAGGA
2251 CTATAAAGAT ACCAGGCGTT TCCCCCTGGA AGCTCCCTCG TGCCTCTCC
2301 TGTTCCGACC CTGCCGCTTA CCGGATACCT GTCCGCCTTT CTCCCTCGG
2351 GAAGCGTGGC GCTTTCTCAA TGCTCACGCT GTAGGTATCT CAGTTCGGTG
2401 TAGGTCGTTT GCTCCAAGCT GGGCTGTGTG CACGAACCCC CCGTTCAGCC
2451 CGACCGCTGC GCCTTATCCG GTAACATGTT TCTTGAGTCC AACCCGGTAA
2501 GACACGACTT ATCGCCACTG GCAGCAGCCA CTGGTAACAG GATTAGCAGA
2551 GCGAGGTATG TAGGCGGTGC TACAGAGTTC TTGAAGTGGT GGCCTAACTA
2601 CGGCTACACT AGAAGGACAG TATTTGGTAT CTGCGCTCTG CTGAAGCCAG
2651 TTACCTTCGG AAAAAGAGTT GGTAGCTCTT GATCCGGCAA ACAAACCACC
2701 GCTGGTAGCG GTGGTTTTTT TGTTTGCAAG CAGCAGATTA CGCGCAGAAA
2751 AAAAGGATCT CAAGAAGATC CTTTGATCTT TTCTACGGGG TCTGACGCTC
2801 AGTGGAACGA AAACACAGT TAAGGGATTT TGGTCATGAG ATTATCAAAA
2851 AGGATCTTCA CCTAGATCCT TTTAAATTA AAATGAAGTT TAAATCAAT
2901 CTAAAGTATA TATGAGTAAA CTTGGTCTGA CAGTTACCAA TGCTTAATCA
2951 GTGAGGCACC TATCTCAGCG ATCTGTCTAT TTCGTTATC CATAGCTGCC
3001 TGAATCCCG TCGTGTAGAT AACTACGATA CGGGAGGGCT TACCATCTGG
3051 CCCAGTGCT GCAATGATAC CGCGAGACCC ACGCTACCG GCTCCAGATT
3101 TATCAGCAAT AAACCAGCCA GCCGGAAGGG CCGAGCGCAG AAGTGGTCCT
3151 GCAACTTTAT CCGCCTCCAT CCAGTCTATT AATTGTTGCC GGAAGCTAG
3201 AGTAAGTAGT TCGCCAGTTA ATAGTTTGCG CAACGTTGTT GCCATTGCTA
3251 CAGGCATCGT GGTGTCACGC TCGTCGTTTG GTATGGCTTC ATTCAGCTCC
3301 GGTTCCCAAC GATCAAGGCG AGTTACATGA TCCCCATGT TGTGCAAAAA
3351 AGCGGTTAGC TCCTTCGGTC CTCCGATCGT TGTCAGAAGT AAGTTGGCCG
3401 CAGTGTTATC ACTCATGGTT ATGGCAGCAC TGCATAATC TCTTACTGTC
3451 ATGCCATCCG TAAGATGCTT TTCTGTGACT GGTGAGTACT CAACCAAGTC
3501 ATTCTGAGAA TAGTGTATGC GCGGACCGAG TTGCTCTTGC CCGGCGTCAA
3551 TACGGGATAA TACCGCGCCA CATAGCAGAA CTTTAAAGT GCTCATCATT
3601 GGAAAACGTT CTTCGGGGCG AAAACTCTCA AGGATCTTAC CGCTGTTGAG

3651 ATCCAGTTCG ATGTAACCCA CTCGTGCACC CAACTGATCT TCAGCATCTT
 3701 TTA CTTTCAC CAGCGTTTCT GGGTGAGCAA AAACAGGAAG GCAAATGCC
 3751 GCAAAAAGG GAATAAGGGC GACACGGAAA TGTTGAATAC TCATACTCTT
 3801 CCTTTTTCAA TATTATTGAA GCATTTATCA GGGTTATTGT CTCATGAGCG
 3851 GATACATATT TGAATGTATT TAGAAAATA AACAAATAGG GGTTCCGCGC
 5 3901 ACATTTCCCC GAAAAGTGCC ACCTGACGTC TAAGAAACCA TTATTATCAT
 3951 GACATTAACC TATAAAAATA GCGTATCAC GAGGCCCTTT CGTCTTCAC

表 14: 人信息素 L 基团的(部分)核苷酸序列(8888 个核苷酸)(SEQ ID NO.:41)

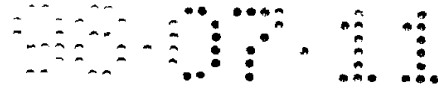
GAGCCGCACACGGTGCTTTTCCACGAGCCAGGCAGCTCCTCTGTGTGGGTGGGAGGACGT
 10 GGCAAGGTCTACCTCTTTGACTTCCCCGAGGGCAAGAACGCATCTGTGCGCACGGTGAGC
 CTCTCTCTTCCCCAACACCCCCCTACCCTCTTATCTCCCCTCTGGCCCTGCCAAGGGT
 CCTCAGGGAATCCGAGGGAGCTGGCTTCTTCTAAACTGCCCCACCTCCGTATCCTA
 TAAATGGCTCCTGGGGGAGGCTCCCTAAAGGTAGTCCAGATTGGAGTGGGGAGCTGGGGC
 GGTGTGGAGAAAAACAGGAGCTAATGGGCCTGGCCAGCTGGGCAGCGCTGCTGCGGAAAG
 CCCAGGCTGGAAGCTGGGCCCCAGAGCCCATGCCTGGTCTTCTGAACCCTCTGGGCCTCA
 15 GCTCTGGATATGAGACCCTGTTTGACCTCAGGTAGATCACTCACCCCTCTCAGAGCCCCAG
 TTGCTCATCTGTGAGATGAGAATAATGGTTGCTTCCCTTTGGGGCTTATCCTGAGGCTGTG
 TGGAAAGCATTTCAGGGGTACCTCACCCCTGGCAGATTGAACTAATGCTTCTCCCCTTCC
 CCAGGTGAATATCGGCTCCACAAAGGGGTCCTGTCTGGATAAGCGGGTGAGCGGGGGAGG
 GATCTGGAGGGGTCTGAGCCACTTGGTAAAGGGAGAGGAGACCCTGAGGGTCTAAGGAAG
 GAAGCATGGCCCTGCCCCACGAGTCCCAGACTGATGGGGAGACGTGGTCCTCTGTGCTTA
 20 GGGGATGGCGTCAGCTGCACACACTCTGGGCTGTCCCGGGAGGCTGTCACCTATGCTAAG
 CCCTTCTGACACCTTCTTCCCTGATCCTGGGGTCTAGTGCTAGGCTTGCCAGGGCCTT
 CCAGCAACCAATTTCTCTCCTCCCTTCTCTTCCCCGGGCAGGACTGCGAGAACTACAT
 CACTCTCCTGGAGAGGCGGAGTGAGGGGCTGCTGGCCTGTGGCACCAACGCCCGGCACCC
 CAGCTGCTGGAACCTGGTGAGAAGGCTGCTCCCCATGTGCCTGATCAGCTCACCTTCTAC
 TGCCTGGGCTTCTGCCCTCATGGTGGGAAGGAGATGGCGAGACTCCAATGCTGGCCTTG
 CCCTGGGAGGATGGGGCTCCTGGCCGAGAACTGGCCGTCATGGGAGGCAGTGGCTGTGG
 25 GATTATGTGGCCATCCAACCCTCTGGATCTCCACAGGTGAATGGCACTGTGGTGCCACT
 TGCGGAGATGAGAGGCTACGCCCTTCCAGCCCGGACGAGAACTCCCTGGTTCTGTTTGA
 AGGTTGGGGCATGCTTCGGAACCTGGGCTGGGAGCAGGATGGTCAGCTCTTGTCCAGTGT
 CCGGAGGAGGGACTTCCAGGAGCTGCCTGCCCTTACTCATTTCTCCCTCCCACTGACCCC
 AGGGGACGAGGTGTATTCCACCATCCGGAAGCAGGAATACAATGGGAAGATCCCTCGGTT
 CCGCCGCATCCGGGGCGAGAGTGAGCTGTACACCAGTGATACTGTCATGCAGAGTGAGTC
 30

5

AGGCTCCGGCTGGGCTGAGGGTGGGCAAGGGGTGTGAGCACTTAAGGTGGCAGATGGGA
TCCTGATGTTTCTGGGAGGGCTCCCTGAGGGCCGCTGGGGCCATGCAGGAAAGCAGGACC
TTGGTATAGGCCTGAGAAGTTAGGGTTGGCTGGGAGCAGAGGAACAGACAAGGTATAGCA
GTGGGATGGGCCAGCCCTCTTCAGGAACACAAACAGAGGGAGCCCCAGACCCAGTGCAG
GGTCCCCAGGAGCCAAAGTTTATCCTCTGCTGAGTTCACGTGGAGGCAGCCCCCAACTC
CCTCCTCATCAGGGCTCTGCCAATTGAGCAGAAGTGACATAGGGGCCCCAGGGACCTTC
CCCCACTCCCCAGGCATGAAGTCATTGCTCCTGGGCCGATGACATCTTTGTAGGAAGAGG
GCAAAACAGGTGTGGGGTGGAGGTGCAGGGTCTAGGGCCCCTCGGGGAGTTGGACCTGAT
GTTATGAGTCCTATTCCAGATCTGATTTGCCATGGTTTGTGCAGACCCGAAGGAGGGAGG
AGAGTGTGCAGGGTTGGAATGGTCTCCCGGGCAAGCTTCCAGCCTTACGCCATTGCGCT
TCTGTGCCCTGGCAGACCCACAGTTCATCAAAGCCACCATCGTGCACCAAGACCAGGCTT
ACGATGACAAGATCTACTACTTCTTCCGAGAGGACAATCCTGACAAGAATCCTGAGGCTC
CTCTCAATGTGTCCCGTGTGGCCAGTTGTGCAGGGTGAACACGGGCGTGAGGGCTGCTG
GCTACGTGTCTGTGCATGAATAGGCCTGAGTGAGGGTGAAGTCTGTGTGTCCGTGTGCAT
GTAGAAGTTGTGTGGATGTATGAGTGGGTCTGTGTGTCAGGGACTGTGGGAGCAGCTGTGTG
TGCATGGAGCATCATGTGTCTGTGTGTGGGTAAAGGTGGCTGAGCTCCTGTGCACGTATG
ATGGCGTGTGAGCGTGTGTATGATGGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT
GTGTGAATGTGCTGTGCCACGTATGTGGGTGCGTGAGTCAGTAAATGTGTGTCTGAGTCC
GTCTGCTCTGTGGGGACCTGGCACTCTCACCTGCCCTGACCCTGGGCACTGCTGGCCCTG
GGCTCTGGATCAGCCAGGCCTGCTTGCAGGAGTCTCATCTGGAGACCTGCCCTGAGTCCT
GGGGCACCCCCGGCAGGTCTGGCCCCCTGCAGCCTGCCTTCCCTCCTCTGGGCCAGGTG
TTGATATTGCTGGCAGTGGTTTCTGGGGTGTGTGGGGAAGCCCGGGCAGGTGCTGAGGG
GCCTCTTCTCCCCTCTACCCTTCCAGGGGGACCAGGGTGGGGAAAGTTCACTGTCAGTCT
CCAAGTGGAACACTTTTCTGAAAGCCATGCTGGTATGCAGTGTGCTGCCACCAACAAGA
ACTTCAACAGGCTGCAAGACGTCTTCTGCTCCCTGACCCAGCGGCCAGTGGAGGGACA
CCAGGGTCTATGGTGTCTTCTCCAACCCCTGGTGTGAGTGGCCCTTGTCTGGGGCCGGGGC
TGCCATTGGTTCAGTGTCCAGTAGGGACAGGAGGCCTTGGGCCCTGCTGAGGGCCTCCCT
GGTGTGGCAGGAGCAGGGGCTGCAGGCTCAAGAGGCTGGGCTGTTGCTGGGTGTGGGGTG
GGGGGACAGCCAGTGCAGTGTATGTACTGTTGTGTGAGTGTGCTGCACTCATGGGTGTG
TGTGCATGCCCTATATGCACACTCATGACTGCACTTGTGCCTGTGTGTCCCACCACCTGC
TTGTGCCGAGAGTGGACACTGGGCCAGGAGGAAGCTGCTGAAGCATCTCTCGGGGAGCT
GGGTGCTATTACACCTGCTCAGGCACTGCCTGAGCCCGATAATTACACTTCTTAATCAC
TCTCATTGATTGAACACACGGCAGGCGGAAGTGTGGGTGTGTGTGGGGAGAGTTAGGGA
TAGAGTGGAGGAAGCCAAGACCCTGCTCTGTGGCTCCTGGGTGAGTGGGTCCCCAGGCT
GGGAAGGGGTTGGGGGTCTGGCCTCCTGGGGCATCAGCACCCACAGCCTGTGCCCAGGG
AGGGCTAGAGAACTGCTCAGCCTATGATGGGGTTCCTCCTGCCTTGGGGTGGGTAGAGC
AGATGGCCTCTAGACTCA.GTGATTCTGTAACAGGATAACAAGTTTGTGGTTTTAAATTGCA
GCACAAAGAAATTAGGCTGAACTCCTCTCCTTCCCTCCTCCTCCATCCCTCCCCATTTTCAG

5

TGGTGGTTGGCAACTCAGTGCCAGGCACAAGGCTGGCCTGGGTGAGTGGAGGTGGATGGG
TGGGTTCTGGGCCCCCATGAGCTGGTCTCCATGTCACTGCAGGAACTACTCAGCCGTC
TGTGTGATTCCCTCGGTGACATTGACAAGGTCTTCCGTACCTCCTCACTCAAGGGCTAC
CACTCAAGCCTTCCCAACCCGCGGCCTGGCAAGGTGAGCGTGACACCAGCCGTGGCCCAG
GCCAGCCCTCCTTCTGCCTCACCTCCCACCACCCCACTGACCTGGGCCTGCTCTCCTTG
CCCAGTGCCTCCAGACCAGCAGCCGATACCCACAGAGACCTTCCAGGTGGCTGACCGTC
ACCAGAGGTGGCGCAGAGGGTGGAGCCCATGGGGCCTCTGAAGACGCCATTGTTCCACT
CTAAATACCACTACCAGAAAAGTGGCOGTCCACCGCATGCAAGCCAGCCACGGGGAGACCT
TTCATGTGCTTTACCTAACTACAGGTGAGAGGCTACCCCGGGACCCTCAGTTTGCTTTGT
AAAAACGGGCATGAAAGGTGTAAGGAATAATGTAGTTAACATCTGGTTGGATCTTTACAT
GTGGAAGGAATAATTGAGTGACTGGAGTTGTCAGGGGTTAATGTGTGTGGGTGTGGAAGA
GCCAGGCAGGGAGAGCTTCCTGGAGGAGGTAGGGGCAAGAGGGAAAGGGGGATGGGAGAA
AAGCAAGCACTGGGATTTGGAGGCGGAAATCTGGAGAGTCTGAGCAAAGCCAGGTGCACC
TTTGGTCCAGATGTCTGACTCAGGGAAGAAGATGGTAGGAAGAGACGTGGCAAATGAGGA
GGAGGGGCCTGAACCACAGGGATACTGGCCTCTGCCAGGCAGAATGAGGGAGTCAGGCC
TGCGCCTGTCTTTGGGATTGTGCAGGTGAGAAGAAACATTTGAGGAGTTGATGGGGCACA
AATTAGGTATGGGGAAGGAGTTCCAGGGGGCAGAACCTTTGCCATCTCACAGAGGACAGG
GGCAGCTTCTCTTCCCTGGAGTAGGCCCTGCTGGGGGAAGCTGGGTGGAATGCCGTG
GGAGATGCTCCTGCTTTCTGGAAAGCCACAGGACACGGAGGAGCCAGTCCTGAGTTGGGT
TTGTGCGCAGCTTCCCATGCCAGCTGCCTTCCCTGAGACTGGAAGGGCCTCTAGCACCCC
TGGGGCCATTCAATTCAGGCCCAGGCGCCCAACCTCAGTTGTTACATTCCCCATGTGAT
CTCCTGTTGCTGCTTACCTTGGGACTGTCTCGGCTTTGGTGACCTTGTAGGAAACTGGA
ACCCAGCACCATTTGTTGGCTCCTGGAAGCCTTGGGGAGAGGAATTTCCACAGGGCAG
GGCCTGGGTCCCTGATTCCCTGCCTCTTACTCCCTATTCATCCCGGCTACACCTTGGGC
CCCCATCCTTGCTTGGCTCCAGTACTGGCTGGCACAGCTGTTGTGGTCATCCAGGGATGG
CAGGGCACTGGGGAACAGAAGAGAGAGGTACACAGTGCGGAACTGGGAGCAGGAGCTAG
GACAAGGAAGGCTGGACTTGGGCCATGGATTCCCTTCCCTGCAGACTTGGGAAGTGAGCAC
ACTTGAGTGATTAGAGAAGGTGTCTTCGTTCTAAGGGCAGTGGAGGAGGCACCATTTTG
AGCCTGCATCATTGATTTGGGCTAGATTGAAAATAGAGCTTTCTAAGTCCTCTGCAG
AGAATGGGAGGCTCTCACAACCTGGGAGAAGTATTGGCTCTTTTCTGAGAATTTTGCCAA
GGGTATGCTGTTACTGGGGCTGGTTTGAAGGAGTATAGGGCATTATGTCTGTGAAGGCA
GTGGCTGGGGTGGGGCCTTATCAGGCCCAAGGAGCATCTGGCCACATCTCAGAGTCCACA
GATGAGGATCACGGATGTGTAGAGGAAACATCCTAGGCAGGCAATCATCTGACTGCTTTT
TTGGGGCAGGTGATGCCCTGGGAAATTGGGAGGGAGGGAGAGAGGGAGGTAGGCTATTCT
AGAAACTGGGAGAGCAGGTGAGGTAGGATTGGGAGGACCAGGGGTCCCATTTGG
TCCCTAATTGAGAACGGAGAGAGCATTGGTCTAGGAGGCAGGCAGCTCGGTTATAAGACC
TTGGGAACTCTTGATTTAGAATCCAAGATCCTTTTTAGATCTAGGATTTTATAAAATTA
GATATCCCCTAAGATCAAATGCAACGTGGAGTCCTGAATTGGATCCTAGAACAGAAGAAG



5

GACATTTGTGGAAAACTAGTGAAATCCAAATAAAGTCTGTAGTTTTGTTAATAGTAATG
CACCAATGTCAGTTGCCTAGTTGTGACAAATATACCGTGGTTATGTAAGATGGTAACATT
AGGGGGAAGTGGAGAAGGGTAGATTGGAGCTCTCTGACTATCTTTGCAACTTTTCTGGG
AATCTAAAATTAAGTCCAAAATAAAAAAAAAAATGATTTAAAGTAAATATATCCCTAAGA
GTCCAGGAGGCAGGGGAGTTGTAGAAGCAGCTGAGTGGTTGGGTCTGACAGATTTGGTT
CCAAGTCCGTCTCTGCTGCTCACCAGCTGTGTGACCTTGAGCAAGTGGCTTAGCCTTTCT
GAGCCTGATTTCTTATCTGTGGAGTGGGGAAGATGACAGCCACCTCGCAGGGCTGTGGA
GGTTAAACGAGGTGATGCATGGACAGCAGCCGCACTGACCTTGCTGGTGTGGGGCTCCT
GCTTCTGTTCTTCCCGTGCAGCCTTGGGAATGTTGGAGGCCGTATCCAGGGACCCCTGGG
CCTCCTGGGATGGCCTCTCTGGATCAGCCTTGAAGGTTCCAGGCTGCCCTTAGGCTCCC
ACATTCTTCCCAGTCACGCTCTCCTCGCCCTGCCACACCAGTCTGTGACCCTTGCCCT
GAGTTGTGACTTCCCACCCCTCCCCGGCCTAGAGGAAAGCTGCCTGGCCCTCAGTGGGA
CTCCCGCCCACTGACCCTCTGTCCACCATACACAGACAGGGGCACTATCCACAAGGTGGT
GGAACCGGGGAGCAGGAGCACAGCTTCGCCTTCAACATCATGGAGATCCAGCCCTTCCG
CCGCGCGGCTGCCATCCAGACCATGTCGCTGGATGCTGAGCGGGTGAAGCTTCCCCACT
GCGTCCCATGGGCTATGCAGTGACTGCAGCTGAGGACAGGGCTCCTTTGCATGTGATTTG
TGTGTTCTTTAAGAGCTTCTAGGCCTTAGGGCCTGGACATTTAGGACTGAGTGTGGGGT
GGGGCCCGGGCCTGACCCAATCCTGCTGTCTTCCAGAGGAAGCTGTATGTGAGCTCCCA
GTGGGAGGTGAGCCAGGTGCCCTGGACCTGTGTGAGGTCTATGGCGGGGGCTGCCACGG
TTGCCTCATGTCCCGAGACCCCTACTGCGGCTGGGACCAGGGCCGCTGCATCTCCATCTA
CAGCTCCGAACGGTACGTTGGCCGGGATCCCTCCGTCCCTGGGACAAGGTGGGCATGGGA
CAGGGGGAGGTGTTGTCGGGCTGGAAGAGGTGGCGGTACTGGGCCTTTCTTGTGGGACCT
CCTCTCTACTGGAAGTGCAGTGGGGTAAGGATATGAGGGTCAAGTCTGCAGCCTTGAT
CTGCTGATCCTCTTTCGTCTTCCCACTCCAGGTCAAGTCTGCAATCCATTAATCCAGCC
GAGCCACACAAGGAGTGTCCCAACCCCAAACAGGTACCTGATCTGGCCCTGCTGGCGGC
TGTGGCCCAATGAGTGGGGTACTGCCCTGCCCTGATTGTCTGCTGAGGGAAACATGG
CCTTGTCTGTGGGCCCCAGGTACATGGGGCAGGATACAGTCTGCAGAGGGAGCCCTCT
TGGTGGGATGAGCGAGACGGGAGAAAAAGGAGGACGCTGAGGGCTGGGTTCCCCACGTT
CATTGAGAAGCCTTGTCTGGGATCCCAGTCCGTGGGGAGGACACATCCTCCCCTGGGAG
CTCTTTGTCCCTCCTCACGGCTGCTTCCCCACTGCCTCCCCAGACAAGGCCCCACTGCAG
AAGGTTTCCCTGGCCCCAACTCTCGCTACTACCTGAGCTGCCCATGGAATCCCGCCAC
GCCACCTACTCATGGCGCCACAAGGAGAACGTGGAGCAGAGCTGCGAACCTGGTCAACCAG
AGCCCCAACTGCATCCTGTTTCATCGAGAACCTCACGGCGCAGCAGTACGGCCACTACTTC
TGCGAGGCCAGGAGGGCTCCTACTTCCGCGAGGCTCAGCACTGGCAGCTGCTGCCCGAG
GACGGCATCATGGCCGAGCACCTGCTGGGTGATGCCTGTGCCCTGGCCGCCTCCCTCTGG
CTGGGGGTGCTGCCCACTCACTCTTGGCTTGTGGTCCAAGTGGCCCTCCCGAGGCTG
GGCATGCCTCAGGCTTCTGCAGCCAGGGCACTAGAACGTCTCACACTCAGAGCCGGCTG
GCCCGGGAGCTCCTTGCCTGCCACTTCTTCCAGGGGACAGATAAACCAGTGGAGGATGC

CAGGCCTGGAGACGTCCAGCCGCAGGCGGCTGCTGGGCCCCAGGTGGCGCACGGATGGTG
 AGGGGCTGAGAATGAGGGCACCGACTGTGAAGCTGGGGCATCGATGACCCAAGACTTTAT
 CTTCTGGAAAATATTTTTCAGACTCCTCAAACCTTGACTAAATGCAGCGATGCTCCCAGCC
 CAAGAGCCCATGGGTCGGGGAGTGGGTTTGGATAGGAGAGCTGGGACTCCATCTCGACCC
 TGGGGCTGAGGCCTGAGTCCTTCTGGACTCTTGGTACCCACATTGCCTCCTTCCCCTCCC
 5 TCTCTCATGGCTGGGTGGCTGGTGTTCCTGAAGACCCAGGGCTACCCTCTGTCCAGCCCT
 GTCCTCTGCAGCTCCCTCTCTGGTCCTGGGTCCCACAGGACAGCCGCCTTGCATGTTTAT
 TGAAGGATGTTTGCTTTCCGGACGGAAGGACGGAAAAAGCTCTGAAAAAAAAAAAAAAAAA
 AAAAAAA

表 15: pMelBacA-H-SEMAL 的核苷酸序列(6622bp)(SEQ ID NO.:42)

10

1 GATATCATGG AGATAATTAA AATGATAACC ATCTCGCAA TAAATAAGTA

51 TTTTACTGTT TTCGTAACAG TTTTGTAATA AAAAAACCTA TAAATATGAA

101 ATTCTTAGTC AACGTTGCC TGTGTTTTAT GGTCGTATAC ATTTCTTACA

15

151 TCTATGCGGA TCGATGG

gga tccgccagg gccacctaag gagcggacc

201 *cgcatctcg ccgtctgaa aggccatgta gggcaggacc ggttgactt*

20

251 *tgccagact gagccgcaca cgggtcttt ccacgagcca ggcagctct*

301 *ctgtgtgggt gggaggacgt ggcaaggtct acctcttga ctccccgag*

351 *ggcaagaacg catctgtcg cacggtgaat atcggctcca caaaggggtc*

25

401 *ctgtctggat aagcgggact gcgagaacta catcactctc ctggagaggc*

451 *ggagtgaggg gctgctggcc tgtggacca acgcccggca cccagctgc*

501 *tggaacctgg tgaatggcac tgtggtgcca ctggcgaga tgagaggcta*

30

551 *tgccccctc agcccggacg agaactccct ggtctgttt gaaggggacg*

5

601 aggtgtattc caccatccgg aagcaggaat acaatgggaa gatccctcgg
651 ttccgccgca tccggggcga gagtgagctg tacaccagtg atactgtcat
701 gcagaacca cagttcatca aagccacat cgtgcaccaa gaccaggctt
751 acgatgaca gatctactac ttctccgag aggacaatcc tgacaagaat
801 cctgaggctc ctctcaatgt gtcccgtg gcccagttgt gcagggggga
851 ccaggggagg gaaagttcac tgcagctc caagtggaac actttctga
901 aagccatgct ggtatgcagt gatgctgcca ccaacaagaa ctcaacagg
951 ctgcaagag tcttctgct cctgacccc agcggccagt ggagggacac
1001 cagggctctat ggtgtttct ccaaccctg gaactactca gccgtctgtg
1051 tgtattccct cggtgacatt gacaaggtct tccgtacct ctactcaag
1101 ggctaccact caagcctcc caaccgggg cctggcaagt gcctcccaga
1151 ccagcagccg ataccacag agacctcca ggtggctgac cgtcacccag
1201 aggtggcgca gagggaggag cccatggggc ctctgaagac gccattgtc
1251 cactctaaat accactacca gaaagtggc gttcacgca tgcaagccag
1301 ccacggggag accttcatg tgctttacct aactacagac aggggcacta
1351 tccacaaggt ggtggaaccg ggggagcagg agcacagctt cgcctcaac
1401 atcatggaga tccagccctt ccgccgcgcg gctgccatcc agaccatgtc
1451 gctggatgct gagcggagga agctgtatgt gagctcccag tgggaggatg
1501 gccaggtgcc cctggacctg tgtgaggtct atggcggggg ctgccacggt

1551 **tgccatcatg cccgagacct ctactgcggc tgggaccagg gccgctgcat**

1601 **ctccatctac agctccgaac ggtcagtgct gcaatccatt aatccagccg**

1651 **agccacacaa ggagtgtccc aaccccaaac cagacaaggc cccactgcag**

5

1701 **aaggttccc tggcccaaaa ctctcgctac tacctgagct gccccatgga**

1751 **atcccgccac gccacctact catggcgcca caaggagaac gtggagcaga**

1801 **gctgcgaacc tggcaccag agccccaact gcatcctgtt catcgagaac**

1851 **ctcacggcgc agcagtacgg ccactacttc tgcgaggccc aggagggtc**

1901 **ctactccgc gaggctcagc actggcagct gctgcccag gacggcatca**

1951 **tggccgagca cctgctgggt catgcctgtg ccctggctgc ctgaattc**

GA

2001 **AGCTTGGAGT CGACTCTGCT GAAGAGGAGG AAATTCTCCT TGAAGTTTCC**

2051 **CTGGTGTTCA AAGTAAAGGA GTTTGCACCA GACGCACCTC TGTTCACTGG**

2101 **TCCGGCGTAT TAAACACGA TACATTGTTA TTAGTACATT TATTAAGCGC**

2151 **TAGATTCTGT GCGTTGTTGA TTTACAGACA ATTGTTGTAC GTATTTAAT**

2201 **AATTCATTAA ATTTATAATC TTTAGGGTGG TATGTTAGAG CGAAAATCAA**

2251 **ATGATTTTCA GCGTCTTTAT ATCTGAATTT AAATATTAAA TCCTCAATAG**

2301 **ATTTGTAAAA TAGGTTTCGA TTAGTTTCAA ACAAGGGTTG TTTTCCGAA**

2351 **CCGATGGCTG GACTATCTAA TGGATTTTCG CTCAACGCCA CAAAACCTGC**

2401 **CAAATCTTGT AGCAGCAATC TAGCTTTGTC GATATTCGTT TGTGTTTTGT**

2451 **TTTGTAATAA AGGTTTCGACG TCGTTCAAAA TATTATGCGC TTTTGTATTT**

5

2501 CTTTCATCAC TGTCGTTAGT GTACAATTGA CTCGACGTAA ACACGTTAAA
2551 TAAAGCCTGG ACATATTTAA CATCGGGCGT GTTAGCTTTA TTAGGCCGAT
2601 TATCGTCGTC GTCCCAACCC TCGTCGTTAG AAGTTGCTTC CGAAGACGAT
2651 TTTGCCATAG CCACACGACG CCTATTAATT GTGTCGGCTA ACACGTCCGC
2701 GATCAAATTT GTAGTTGAGC TTTTGGGAAT TATTTCTGAT TGCGGGCGTT
2751 TTTGGGCGGG TTTCAATCTA ACTGTGCCCG ATTTTAATTC AGACAACACG
2801 TTAGAAAGCG ATGGTGCAGG CGGTGGTAAÇ ATTCAGACG GCAAATCTAC
2851 TAATGGCGGC GGTGGTGGAG CTGATGATAA ATCTACCATC GTGGAGGCG
2901 CAGGCGGGGC TGGCGGCGGA GGCGGAGGCG GAGGTGGTGG GGTGATGCA
2951 GACGGCGGTT TAGGCTCAAA TTGTCTCTTT CAGGCAACAC AGTCGGCACC
3001 TCAACTATTG TACTGGTTTC GGGCGTATGG TGCACTCTCA GTACAATCTG
3051 CTCTGATGCC GCATAGTTAA GCCAGCCCCG ACACCCGCCA ACACCCGCTG
3101 ACGCGCCCTG ACGGGCTTGT CTGCTCCCGG CATCCGCTTA CAGACAAGCT
3151 GTGACCGTCT CCGGGAGCTG CATGTGTCAG AGGTTTTTAC CGTCATCACC
3201 GAAACGCGCG AGACGAAAGG GCCTCGTGAT ACGCCTATTT TTATAGGTTA
3251 ATGTCATGAT AATAATGGTT TCTTAGACGT CAGGTGGCAC TTTTCGGGGA
3301 AATGTGCGCG GAACCCCTAT TTGTTTATTT TTCTAAATAC ATTCAAATAT
3351 GTATCCGCTC ATGAGACAAT AACCTGATA AATGCTTCAA TAATATTGAA
3401 AAAGGAAGAG TATGAGTATT CAACATTTCC GTGTCGCCCT TATTCCCTTT

5

3451 TTTGCGGCAT TTTGCCTTCC TGTTTTTCT CACCCAGAAA CGCTGGTGAA
 3501 AGTAAAAGAT GCTGAAGATC AGTTGGGTGC ACGAGTGGGT TACATCGAAC
 3551 TGGATCTCAA CAGCGGTAAG ATCCTTGAGA GTTTTCGCCCG CGAAGAACGT
 3601 TTTCCAATGA TGAGCACTTT TAAAGTTCTG CTATGTGGCG CGGTATTATC
 3651 CCGTATTGAC GCCGGGCAAG AGCAACTCGG TCGCCGCATA CACTATTCTC
 3701 AGAATGACTT GGTTGAGTAC TCACCAGTCA CAGAAAAGCA TCTTACGGAT
 3751 GGCATGACAG TAAGAGAATT ATGCAGTGCT GCCATAACCA TGAGTGATAA
 3801 CACTGCGGCC AACTTACTTC TGACAACGAT CGGAGGACCG AAGGAGCTAA
 3851 CCGCTTTTTT GCACAACATG GGGGATCATG TAACTCGCCT TGATCGTTGG
 3901 GAACCGGAGC TGAATGAAGC CATAACAAAC GACGAGCGTG ACACCACGAT
 3951 GCCTGTAGCA ATGGCAACAA CGTTGCGCAA ACTATTAACCT GGCGAACTAC
 4001 TTA CTCTAGC TTCCCGGCAA CAATTAATAG ACTGGATGGA GGCGGATAAA
 4051 GTTGCAGGAC CACTTCTGCG CTCGGCCCTT CCGGCTGGCT GGTTTATTGC
 4101 TGATAAATCT GGAGCCGGTG AGCGTGGGTC TCGCGGTATC ATTGCAGCAC
 4151 TGGGGCCAGA TGGTAAGCCC TCCCGTATCG TAGTTATCTA CACGACGGGG
 4201 AGTCAGGCAA CTATGGATGA ACGAAATAGA CAGATCGCTG AGATAGGTGC
 4251 CTCACTGATT AAGCATTGGT AACTGTCAGA CCAAGTTTAC TCATATATAC
 4301 TTTAGATTGA TTAAAACTT CATTTTTAAT TTAAAAGGAT CTAGGTGAAG
 4351 ATCCTTTTTG ATAATCTCAT GACCAAATC CCTTAACGTG AGTTTTCGTT

5

4401 CCACTGAGCG TCAGACCCCG TAGAAAAGAT CAAAGGATCT TCTTGAGATC
4451 CTTTTTTTCT GCGCGTAATC TGCTGCTTGC AAACAAAAAA ACCACCGCTA
4501 CCAGCGGTGG TTTGTTTGCC GGATCAAGAG CTACCAACTC TTTTCCGAA
4551 GGTAAGTGGC TTCAGCAGAG CGCAGATACC AAATACTGTT CTTCTAGTGT
4601 AGCCGTAGTT AGGCCACCAC TTCAAGAACT CTGTAGCACC GCCTACATAC
4651 CTCGCTCTGC TAATCCTGTT ACCAGTGGCT GCTGCCAGTG GCGATAAGTC
4701 GTGTCTTACC GGGTTGGACT CAAGACGATA GTTACCGGAT AAGGCGCAGC
4751 GGTCGGGCTG AACGGGGGGT TCGTGCACAC AGCCCAGCTT GGAGCGAACG
4801 ACCTACACCG AACTGAGATA CCTACAGCGT GAGCTATGAG AAAGCGCCAC
4851 GCTTCCCGAA GGGAGAAAGG CGGACAGGTA TCCGGTAAGC GGCAGGGTCCG
4901 GAACAGGAGA GCGCACGAGG GAGCTTCCAG GGGGAAACGC CTGGTATCTT
4951 TATAGTCCTG TCGGGTTTCG CCACCTCTGA CTTGAGCGTC GATTTTTGTG
5001 ATGCTCGTCA GGGGGGCGGA GCCTATGGAA AAACGCCAGC AACGCGGCCT
5051 TTTTACGGTT CCTGGCCTTT TGCTGGCCTT TTGCTCACAT GTTCTTTCTT
5101 GCGTTATCCC CTGATTCTGT GGATAACCGT ATTACCGCCT TTGAGTGAGC
5151 TGATACCGCT CGCCGCAGCC GAACGACCGA GCGCAGCGAG TCAGTGAGCG
5201 AGGAAGCATC CTGCACCATC GTCTGCTCAT CCATGACCTG ACCATGCAGA
5251 GGATGATGCT CGTGACGGTT AACGCCTCGA ATCAGCAACG GCTTGCCGTT
5301 CAGCAGCAGC AGACCATTTT CAATCCGCAC CTCGCGGAAA CCGACATCGC

5

5351 AGGCTTCTGC TTCAATCAGC GTGCCGTCGG CGGTGTGCAG TTCAACCACC
5401 GCACGATAGA GATTCGGGAT TTCGGCGCTC CACAGTTTCG GGTTCGAC
5451 GTTCAGACGT AGTGTGACGC GATCGGTATA ACCACCACGC TCATCGATAA
5501 TTTCACCGCC GAAAGGCGCG GTGCCGCTGG CGACCTGCGT TTCACCCTGC
5551 CATAAAGAAA CTGTTACCCG TAGGTAGTCA CGCAACTCGC CGCACATCTG
5601 AACTTCAGCC TCCAGTACAG CGCGGCTGAA ATCATCATT AAGCGAGTGG
5651 CAACATGGAA ATCGCTGATT TGTGTAGTCG GTTTATGCAG CAACGAGACG
5701 TCACGGAAAA TGCCGCTCAT CCGCCACATA TCCTGATCTT CCAGATAACT
5751 GCCGTCCTC CAACGCAGCA CCATCACCGC GAGGCGGTTT TCTCCGGCGC
5801 GTAAAAATGC GCTCAGGTCA AATTCAGACG GCAAACGACT GTCCTGGCCG
5851 TAACCGACCC AGCGCCCGTT GCACCACAGA TGAAACGCCG AGTTAACGCC
5901 ATCAAAAATA ATTTCGCTCT GGCCTTCCTG TAGCCAGCTT TCATCAACAT
5951 TAAATGTGAG CGAGTAACAA CCCGTCCGAT TCTCCGTGGG AACAAACGGC
6001 GGATTGACCG TAATGGGATA GGTCACGTTG GTGTAGATGG GCGCATCGTA
6051 ACCGTGCATC TGCCAGTTTG AGGGGACGAC GACAGTATCG GCCTCAGGAA
6101 GATCGCACTC CAGCCAGCTT TCCGGCACCG CTTCTGGTGC CGGAAACCAG
6151 GCAAAGCGCC ATTCGCCATT CAGGCTGCGC AACTGTTGGG AAGGGCGATC
6201 GGTGCGGGCC TCTTCGCTAT TACGCCAGCT GGCGAAAGGG GGATGTGCTG
6251 CAAGGCGATT AAGTTGGGTA ACGCCAGGGT TTTCCAGTC ACGACGTTGT

6301 AAAACGACGG GATCTATCAT TTTAGCAGT GATTCTAATT GCAGCTGCTC

6351 TTTGATACAA CTAATTTTAC GACGACGATG CGAGCTTTTA TTCAACCGAG

5

6401 CGTGCAATGTT TGCAATCGTG CAAGCGTTAT CAATTTTCA TTATCGTATT

6451 GTTGACATC AACAGGCTGG ACACCACGTT GAACTCGCCG CAGTTTTGCG

6501 GCAAGTTGGA CCCGCCGCGC ATCCAATGCA AACTTTCCGA CATTCTGTTG

6551 CCTACGAACG ATTGATTCTT TGTCCATTGA TCGAAGCGAG TGCCTTCGAC

6601 TTTTCGTGT CCAGTGTGGC TT

序 列 表

(1) 一般信息:

(i) 申请人:

- 5 (A) 名称: Hoechst Marion Roussel Deutschland GmbH
 (B) 街道:
 (C) 城市: Frankfurt
 (D) 州:
 (E) 国家: 德国
 10 (F) 邮区代码(ZIP): 69926
 (G) 电话: 069-305-7072
 (H) 传真: 069-35-7175
 (I) 电传:

(ii) 发明名称: 人类信息素 L 和其它物种中相应的信息素

15 (iii) 序列数: 44

(iv) 计算机可读形式:

- (A) 介质类型: 软盘
 (B) 计算机: IBM PC 兼容机
 (C) 操作系统: PC-DOS/MS-DOS
 20 (D) 软件: PatentIn Release#1.0 , 版本#1.25 (EPO)

(2) SEQ ID NO: 1 的信息:

(i) 序列特征:

- (A) 长度: 2636 个碱基对
 25 (B) 类型: 核酸
 (C) 链型: 单链
 (D) 拓扑结构: 线型

(ii) 分子类型: DNA(基因组的)

(ix) 特征:

- 30 (A) 名称/关键词: 外显子

	GGCGGGGGCT GCCACGGTTG CCTCATGTCC CGAGACCCCT ACTGCGGCTG GGACCAGGGC	1560
	CGCTGCATCT CCATCTACAG CTCCGAACGG TCAGTGCTGC AATCCATTAA TCCAGCCGAG	1620
	CCACACAAGG AGTGTCCCAA CCCCAAACCA GACAAGGCC CACTGCAGAA GGTTCCTCTG	1680
5	GCCCCAACT CTCGCTACTA CCTGAGCTGC CCCATGGAAT CCCGCCACGC CACCTACTCA	1740
	TGGCGCCACA AGGAGAACGT GGAGCAGAGC TCGAACCTG GTCACCAGAG CCCCAACTGC	1800
	ATCCTGTTCA TCGAGAACCT CACGGCGCAG CAGTACGGCC ACTACTTCTG CGAGGCCAG	1860
	GAGGGCTCCT ACTTCCGCGA GGCTCAGCAC TGGCAGCTGC TGCCCGAGGA CGGCATCATG	1920
	GCCGAGCACC TGCTGGGTCA TGCCTGTGCC CTGGCTGCCT CCCTCTGGCT GGGGGTGCTG	1980
10	CCCACACTCA CTCTTGGCTT GCTGGTCCAC TAGGGCCTCC CGAGGCTGGG CATGCCTCAG	2040
	GCTTCTGCAG CCCAGGGCAC TAGAACGTCT CACACTCAGA GCCGGCTGGC CCGGGAGCTC	2100
	CTTGCCCTGCC ACTTCTTCCA GGGGACAGAA TAACCCAGTG GAGGATGCCA GGCCTGGAGA	2160
	CGTCCAGCCG CAGGCGGCTG CTGGGCCCCA GGTGGCGCAC GGATGGTGAG GGGCTGAGAA	2220
	TGAGGGCACC GACTGTGAAG CTGGGGCATC GATGACCCAA GACTTTATCT TCTGGAAAAT	2280
15	ATTTTTCAGA CTCCTCAAAC TTGACTAAAT GCAGCGATGC TCCCAGCCCA AGAGCCCATG	2340
	GGTCGGGGAG TGGGTTTGGG TAGGAGAGCT GGGACTCCAT CTCGACCCTG GGGCTGAGGC	2400
	CTGAGTCCTT CTGGACTCTT GGTACCCACA TTGCCTCCTT CCCCTCCCTC TCTCATGGCT	2460
	GGGTGGCTGG TGTTCTGAA GACCCAGGGC TACCCTCTGT CCAGCCCTGT CCTCTGCAGC	2520
	TCCCTCTCTG GTCCTGGGTC CCACAGGACA GCCGCCTTGC ATGTTTATTG AAGGATGTTT	2580
20	GCTTTCGGA CGGAAGGACG GAAAAAGCTC TGAAAAAAA AAAAAAAA AAAAA	2636

(2) SEQ ID NO: 2 的信息:

(i) 序列特征:

- 25 (A) 长度: 1195 个碱基对
- (B) 类型: 核酸
- (C) 链型: 单链
- (D) 拓扑结构: 线型

(ii) 分子类型: DNA(基因组的)

30 (ix) 特征:

(A) 名称/关键词: 外显子

(B) 位置: 1..1195

(xi) 序列描述: SEQ ID NO: 2:

	CGGGGCTGCG GGATGACGCC TCCTCCTCCC GGACGTGCCG CCCCAGCGC ACCGCGCGCC	60
5	CCCGTCCTCA GCCTGCCGGC TCGGTTGGGG CTCCCGCTGC GGCTGCGGCT TCTGCTGGTG	120
	TTCTGGGTGG CCGCCGCCTC CGCCCAAGGC CACTCGAGGA GCGGACCCCG CATCTCCGCC	180
	GTCTGGAAAG GGCAGGACCA TGTGGACTTT AGCCAGCCTG AGCCACACAC CGTGCTTTTC	240
	CATGAGCCGG GCAGCTTCTC TGTCTGGGTG GGTGGACGTG GCAAGGTCTA CCACTTCAAC	300
	TTCCCCGAGG GCAAGAATGC CTCTGTGCGC ACGGTGAACA TCGGCTCCAC AAAGGGGTCC	360
10	TGTCAGGACA AACAGGACTG TGGGAATTAC ATCACTCTTC TAGAAAGGCG GGGTAATGGG	420
	CTGCTGGTCT GTGGCACCAA TGCCCGAAG CCCAGCTGCT GGAACTGGT GAATGACAGT	480
	GTGGTGATGT CACTTGGTGA GATGAAAGGC TATGCCCCCT TCAGCCCGGA TGAGAACTCC	540
	CTGGTTCTGT TTGAAGGAGA TGAAGTGAC TCTACCATCC GGAAGCAGGA ATACAACGGG	600
	AAGATCCCTC GGTTTCGACG CATTCGGGGC GAGAGTGAAC TGTACACAAG TGATACAGTC	660
15	ATGCAGAACC CACAGTTCAT CAAGGCCACC ATTGTGCACC AAGACCAAGC CTATGATGAT	720
	AAGATCTACT ACTTCTTCCG AGAAGACAAC CCTGACAAGA ACCCCGAGGC TCCTCTCAAT	780
	GTGTCCCGAG TAGCCAGTT GTGCAGGGGG GACCAGGGTG GTGAGAGTTC GTTGTCTGTC	840
	TCCAAGTGGG ACACCTTCCT GAAAGCCATG TTGGTCTGCA GCGATGCAGC CACCAACAGG	900
	AACTTCAATC GGCTGCAAGA TGTCTTCCTG CTCCCTGACC CCAGTGGCCA GTGGAGAGAT	960
20	ACCAGGGTCT ATGGCGTTTT CTCCAACCCC TGGAACTACT CAGCTGTCTG CGTGTATTCTG	1020
	CTTGGTGACA TTGACAGAGT CTTCCGTACC TCATCGCTCA AAGGCTACCA CATGGGCCTT	1080
	TCCAACCCTC GACCTGGCAT GTGCCTCCCA AAAAAGCAGC CCATACCCAC AGAAACCTTC	1140
	CAGGTAGCTG ATAGTCACCC AGAGGTGGCT CAGAGGGTGG AACCTATGGG GCCCC	1195
25		

(2) SEQ ID NO: 3 的信息:

(i) 序列特征:

(A) 长度: 666 个氨基酸

(B) 类型: 氨基酸

(C) 链型: 单链

30

(D) 拓扑结构: 线型

(ii) 分子类型: 蛋白质

(ix) 特征:

(A) 名称/关键词: 蛋白质

(B) 位置: 1.666

(ix) 特征:

(A) 名称/关键词: 蛋白质

(B) 位置: 1.666

(xi) 序列描述: SEQ ID NO: 3 :

5

Met Thr Pro Pro Pro Pro Gly Arg Ala Ala Pro Ser Ala Pro Arg Ala
1 5 10 15

Arg Val Pro Gly Pro Pro Ala Arg Leu Gly Leu Pro Leu Arg Leu Arg
20 25 30

Leu Leu Leu Leu Leu Trp Ala Ala Ala Ala Ser Ala Gln Gly His Leu
35 40 45

15

Arg Ser Gly Pro Arg Ile Phe Ala Val Trp Lys Gly His Val Gly Gln
50 55 60

Asp Arg Val Asp Phe Gly Gln Thr Glu Pro His Thr Val Leu Phe His
65 70 75 80

Glu Pro Gly Ser Ser Ser Val Trp Val Gly Gly Arg Gly Lys Val Tyr
85 90 95

20

Leu Phe Asp Phe Pro Glu Gly Lys Asn Ala Ser Val Arg Thr Val Asn
100 105 110

Ile Gly Ser Thr Lys Gly Ser Cys Leu Asp Lys Arg Asp Cys Glu Asn
115 120 125

Tyr Ile Thr Leu Leu Glu Arg Arg Ser Glu Gly Leu Leu Ala Cys Gly
130 135 140

25

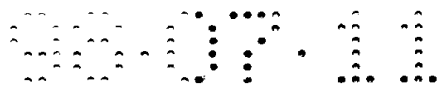
Thr Asn Ala Arg His Pro Ser Cys Trp Asn Leu Val Asn Gly Thr Val
145 150 155 160

Val Pro Leu Gly Glu Met Arg Gly Tyr Ala Pro Phe Ser Pro Asp Glu
165 170 175

Asn Ser Leu Val Leu Phe Glu Gly Asp Glu Val Tyr Ser Thr Ile Arg
180 185 190

30

Lys Gln Glu Tyr Asn Gly Lys Ile Pro Arg Phe Arg Arg Ile Arg Gly
195 200 205



Trp Asp Gln Gly Arg Cys Ile Ser Ile Tyr Ser Ser Glu Arg Ser Val
 515 520 525
 Leu Gln Ser Ile Asn Pro Ala Glu Pro His Lys Glu Cys Pro Asn Pro
 530 535 540
 Lys Pro Asp Lys Ala Pro Leu Gln Lys Val Ser Leu Ala Pro Asn Ser
 545 550 555 560
 5 Arg Tyr Tyr Leu Ser Cys Pro Met Glu Ser Arg His Ala Thr Tyr Ser
 565 570 575
 Trp Arg His Lys Glu Asn Val Glu Gln Ser Cys Glu Pro Gly His Gln
 580 585 590
 Ser Pro Asn Cys Ile Leu Phe Ile Glu Asn Leu Thr Ala Gln Gln Tyr
 595 600 605
 10 Gly His Tyr Phe Cys Glu Ala Gln Glu Gly Ser Tyr Phe Arg Glu Ala
 610 615 620
 Gln His Trp Gln Leu Leu Pro Glu Asp Gly Ile Met Ala Glu His Leu
 625 630 635 640
 Leu Gly His Ala Cys Ala Leu Ala Ala Ser Leu Trp Leu Gly Val Leu
 645 650 655
 15 Pro Thr Leu Thr Leu Gly Leu Leu Val His
 660 665

(2) SEQ ID NO: 4 的信息:

- (i) 序列特征:
 - (A) 长度: 394 个氨基酸
 - (B) 类型: 氨基酸
 - (C) 链型: 单链
 - (D) 拓扑结构: 线型
- (ii) 分子类型: 蛋白质
- (ix) 特征:
 - (A) 名称/关键词: 蛋白质
 - (B) 位置: 1..394

(xi) 序列描述: SEQ ID NO: 4:

Met Thr Pro Pro Pro Pro Gly Arg Ala Ala Pro Ser Ala Pro Arg Ala
 1 5 10 15
 Arg Val Leu Ser Leu Pro Ala Arg Phe Gly Leu Pro Leu Arg Leu Arg
 20 25 30
 30 Leu Leu Leu Val Phe Trp Val Ala Ala Ala Ser Ala Gln Gly His Ser
 35 40 45

0071

Arg Ser Gly Pro Arg Ile Ser Ala Val Trp Lys Gly Gln Asp His Val
 50 55 60
 Asp Phe Ser Gln Pro Glu Pro His Thr Val Leu Phe His Glu Pro Gly
 65 70 75 80
 Ser Phe Ser Val Trp Val Gly Gly Arg Gly Lys Val Tyr His Phe Asn
 85 90 95
 5 Phe Pro Glu Gly Lys Asn Ala Ser Val Arg Thr Val Asn Ile Gly Ser
 100 105 110
 Thr Lys Gly Ser Cys Gln Asp Lys Gln Asp Cys Gly Asn Tyr Ile Thr
 115 120 125
 Leu Leu Glu Arg Arg Gly Asn Gly Leu Leu Val Cys Gly Thr Asn Ala
 130 135 140
 10 Arg Lys Pro Ser Cys Trp Asn Leu Val Asn Asp Ser Val Val Met Ser
 145 150 155 160
 Leu Gly Glu Met Lys Gly Tyr Ala Pro Phe Ser Pro Asp Glu Asn Ser
 165 170 175
 Leu Val Leu Phe Glu Gly Asp Glu Val Tyr Ser Thr Ile Arg Lys Gln
 180 185 190
 Glu Tyr Asn Gly Lys Ile Pro Arg Phe Arg Arg Ile Arg Gly Glu Ser
 195 200 205
 15 Glu Leu Tyr Thr Ser Asp Thr Val Met Gln Asn Pro Gln Phe Ile Lys
 210 215 220
 Ala Thr Ile Val His Gln Asp Gln Ala Tyr Asp Asp Lys Ile Tyr Tyr
 225 230 235 240
 Phe Phe Arg Glu Asp Asn Pro Asp Lys Asn Pro Glu Ala Pro Leu Asn
 245 250 255
 Val Ser Arg Val Ala Gln Leu Cys Arg Gly Asp Gln Gly Gly Glu Ser
 260 265 270
 20 Ser Leu Ser Val Ser Lys Trp Asn Thr Phe Leu Lys Ala Met Leu Val
 275 280 285
 Cys Ser Asp Ala Ala Thr Asn Arg Asn Phe Asn Arg Leu Gln Asp Val
 290 295 300
 Phe Leu Leu Pro Asp Pro Ser Gly Gln Trp Arg Asp Thr Arg Val Tyr
 305 310 315 320
 Gly Val Phe Ser Asn Pro Trp Asn Tyr Ser Ala Val Cys Val Tyr Ser
 325 330 335
 25 Leu Gly Asp Ile Asp Arg Val Phe Arg Thr Ser Ser Leu Lys Gly Tyr
 340 345 350
 His Met Gly Leu Ser Asn Pro Arg Pro Gly Met Cys Leu Pro Lys Lys
 355 360 365
 Gln Pro Ile Pro Thr Glu Thr Phe Gln Val Ala Asp Ser His Pro Glu
 370 375 380
 30 Val Ala Gln Arg Val Glu Pro Met Gly Pro
 385 390

(2) SEQ ID NO: 5 的信息:

(i) 序列特征:

(A) 长度: 23 个碱基对

(B) 类型: 核酸

5 (C) 链型: 单链

(D) 拓扑结构: 线型

(ii) 分子类型: DNA(基因组的)

(ix) 特征:

(A) 名称/关键词: 外显子

10 (B) 位置: 1..23

(xi) 序列描述: SEQ ID NO: 5 :

ACTCACTATA GGGCTCGAGC GGC

23

(2) SEQ ID NO: 6 的信息:

15 (i) 序列特征:

(A) 长度: 20 个碱基对

(B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑结构: 线型

20 (ii) 分子类型: DNA(基因组的)

(ix) 特征:

(A) 名称/关键词: 外显子

(B) 位置: 1..20

(xi) 序列描述: SEQ ID NO: 6 :

25 AGCCGCACAC GGTGCTTTTC

20

(2) SEQ ID NO: 7 的信息:

(i) 序列特征:

(A) 长度: 20 个碱基对

30 (B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑结构: 线型

(ii) 分子类型: DNA(基因组的)

(ix) 特征:

5

(A) 名称/关键词: 外显子

(B) 位置: 1..20

(xi) 序列描述: SEQ ID NO: 7:

GCACAGATGC GTTCTTGCCC

20

10 (2) SEQ ID NO: 8 的信息:

(i) 序列特征:

(A) 长度: 20 个碱基对

(B) 类型: 核酸

(C) 链型: 单链

15

(D) 拓扑结构: 线型

(ii) 分子类型: DNA(基因组的)

(ix) 特征:

(A) 名称/关键词: 外显子

(B) 位置: 1..20

20

(xi) 序列描述: SEQ ID NO: 8:

ACCATAGACC CTGGTGTCCC

(2) SEQ ID NO: 9 的信息:

(i) 序列特征:

25

(A) 长度: 20 个碱基对

(B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑结构: 线型

(ii) 分子类型: DNA(基因组的)

30

(ix) 特征:

(A) 名称/关键词: 外显子

(B) 位置: 1..20

(xi) 序列描述: SEQ ID NO: 9:

GCAGTGATGC TGCCACCAAC 20

5

(2) SEQ ID NO: 10 的信息:

(i) 序列特征:

(A) 长度: 20 个碱基对

(B) 类型: 核酸

10

(C) 链型: 单链

(D) 拓扑结构: 线型

(ii) 分子类型: DNA(基因组的)

(ix) 特征:

(A) 名称/关键词: 外显子

15

(B) 位置: 1..20

(xi) 序列描述: SEQ ID NO: 10:

CCAGACCATG TCGCTGGATG 20

(2) SEQ ID NO: 11 的信息:

20

(i) 序列特征:

(A) 长度: 20 个碱基对

(B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑结构: 线型

25

(ii) 分子类型: DNA(基因组的)

(ix) 特征:

(A) 名称/关键词: 外显子

(B) 位置: 1..20

(xi) 序列描述: SEQ ID NO: 11:

30 ACATGAGGCA ACCGTGGCAG 20

(2) SEQ ID NO: 12 的信息:

(i) 序列特征:

(A) 长度: 27 个碱基对

(B) 类型: 核酸

5 (C) 链型: 单链

(D) 拓扑结构: 线型

(ii) 分子类型: DNA(基因组的)

(ix) 特征:

(A) 名称/关键词: 外显子

10 (B) 位置: 1..27

(xi) 序列描述: SEQ ID NO: 12 :

CCATCCTAAT ACGACTCACT ATAGGGC

27

(2) SEQ ID NO: 13 的信息:

15 (i) 序列特征:

(A) 长度: 20 个碱基对

(B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑结构: 线型

20 (ii) 分子类型: DNA(基因组的)

(ix) 特征:

(A) 名称/关键词: 外显子

(B) 位置: 1..20

(xi) 序列描述: SEQ ID NO: 13 :

25 AGGTAGACCT TGCCACGTCC

20

(2) SEQ ID NO: 14 的信息:

(i) 序列特征:

(A) 长度: 23 个碱基对

30 (B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑结构: 线型

(ii) 分子类型: DNA(基因组的)

(ix) 特征:

5 (A) 名称/关键词: 外显子

(B) 位置: 1..23

(xi) 序列描述: SEQ ID NO: 14 :
GAACTTCAAC AGGCTGCAAG ACG

23

10 (2) SEQ ID NO: 15 的信息:

(i) 序列特征:

(A) 长度: 20 个碱基对

(B) 类型: 核酸

(C) 链型: 单链

15 (D) 拓扑结构: 线型

(ii) 分子类型: DNA(基因组的)

(ix) 特征:

(A) 名称/关键词: 外显子

(B) 位置: 1..20

20 (xi) 序列描述: SEQ ID NO: 15 :

ATGCTGAGCG GAGGAAGCTG

20

(2) SEQ ID NO: 16 的信息:

(i) 序列特征:

25 (A) 长度: 20 个碱基对

(B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑结构: 线型

(ii) 分子类型: DNA(基因组的)

30 (ix) 特征:

(A) 名称/关键词: 外显子

(B) 位置: 1..20

(xi) 序列描述: SEQ ID NO: 16:

CCGCCATACA CCTCACACAG

20

5

(2) SEQ ID NO: 17 的信息:

(i) 序列特征:

(A) 长度: 28 个碱基对

(B) 类型: 核酸

10

(C) 链型: 单链

(D) 拓扑结构: 线型

(ii) 分子类型: DNA(基因组的)

(ix) 特征:

(A) 名称/关键词: 外显子

15

(B) 位置: 1..28

(xi) 序列描述: SEQ ID NO: 17:

CTGGAAGCTT TCTGTGGGTA TCGGCTGC

28

(2) SEQ ID NO: 18 的信息:

20

(i) 序列特征:

(A) 长度: 25 个碱基对

(B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑结构: 线型

25

(ii) 分子类型: DNA(基因组的)

(ix) 特征:

(A) 名称/关键词: 外显子

(B) 位置: 1..25

(xi) 序列描述: SEQ ID NO: 18:

30 TTTGGATCCC TGGTCTGTT TGAAG

25

(2) SEQ ID NO: 19 的信息:

(i) 序列特征:

(A) 长度: 50 个碱基对

(B) 类型: 核酸

5 (C) 链型: 单链

(D) 拓扑结构: 线型

(ii) 分子类型: DNA(基因组的)

(ix) 特征:

(A) 名称/关键词: 外显子

10 (B) 位置: 1..50

(ix) 特征:

(A) 名称/关键词: 外显子

(B) 位置: 1..50

(xi) 序列描述: SEQ ID NO: 19:

15 TTCTAGAATT CAGCGGCCGC TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT
50

(2) SEQ ID NO: 20 的信息:

(i) 序列特征:

20 (A) 长度: 27 个碱基对

(B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑结构: 线型

(ii) 分子类型: DNA(基因组的)

25 (ix) 特征:

(A) 名称/关键词: 外显子

(B) 位置: 1..27

(xi) 序列描述: SEQ ID NO: 20:

GGGGAAAGTT CACTGTCAGT CTC CAAG

27

30

(2) SEQ ID NO: 21 的信息:

(i) 序列特征:

(A) 长度: 26 个碱基对

(B) 类型: 核酸

5 (C) 链型: 单链

(D) 拓扑结构: 线型

(ii) 分子类型: DNA(基因组的)

(ix) 特征:

(A) 名称/关键词: 外显子

10 (B) 位置: 1..26

(xi) 序列描述: SEQ ID NO: 21 :

GGGAATACAC ACAGACGGCT GAGTAG 26

(2) SEQ ID NO: 22 的信息:

15 (i) 序列特征:

(A) 长度: 22 个碱基对

(B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑结构: 线型

20 (ii) 分子类型: DNA(基因组的)

(ix) 特征:

(A) 名称/关键词: 外显子

(B) 位置: 1..22

(xi) 序列描述: SEQ ID NO: 22 :

25 AGCAAGTTCA GCCTGGTTAA GT 22

(2) SEQ ID NO: 23 的信息:

(i) 序列特征:

(A) 长度: 21 个碱基对

30 (B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑结构: 线型

(ii) 分子类型: DNA(基因组的)

(ix) 特征:

5 (A) 名称/关键词: 外显子

(B) 位置: 1..21

(xi) 序列描述: SEQ ID NO: 23 :

TTATGAGTAT TTCTTCCAGG G

21

10 (2) SEQ ID NO: 24 的信息:

(i) 序列特征:

(A) 长度: 26 个碱基对

(B) 类型: 核酸

(C) 链型: 单链

15 (D) 拓扑结构: 线型

(ii) 分子类型: DNA(基因组的)

(ix) 特征:

(A) 名称/关键词: 外显子

(B) 位置: 1..26

20 (xi) 序列描述: SEQ ID NO: 24 :

CCATTAATCC AGCCGAGCCA CACAAG

26

(2)SEQ ID NO: 25 的信息:

(i) 序列特征:

25 (A) 长度: 25 个碱基对

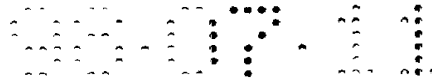
(B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑结构: 线型

(ii) 分子类型: DNA(基因组的)

30 (ix) 特征:



(A) 名称/关键词: 外显子

(B) 位置: 1..25

(xi) 序列描述: SEQ ID NO: 25 :

CATCTACAGC TCCGAACGGT CAGTG 25

5

(2) SEQ ID NO: 26 的信息:

(i) 序列特征:

(A) 长度: 20 个碱基对

(B) 类型: 核酸

10

(C) 链型: 单链

(D) 拓扑结构: 线型

(ii) 分子类型: DNA(基因组的)

(ix) 特征:

(A) 名称/关键词: 外显子

15

(B) 位置: 1..20

(xi) 序列描述: SEQ ID NO: 26 :

CAGCGGAAGC CCCAACCGAG 20

(2) SEQ ID NO: 27 的信息:

20

(i) 序列特征:

(A) 长度: 23 个碱基对

(B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑结构: 线型

25

(ii) 分子类型: DNA(基因组的)

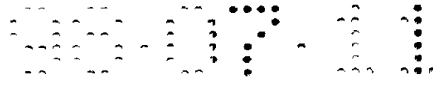
(ix) 特征:

(A) 名称/关键词: 外显子

(B) 位置: 1..23

(xi) 序列描述: SEQ ID NO: 27 :

30 GGGATGACGC CTCCTCCGCC CGG 23



(2) SEQ ID NO: 28 的信息:

(i) 序列特征:

(A) 长度: 31 个碱基对

(B) 类型: 核酸

5 (C) 链型: 单链

(D) 拓扑结构: 线型

(ii) 分子类型: DNA(基因组的)

(ix) 特征:

(A) 名称/关键词: 外显子

10 (B) 位置: 1..31

(xi) 序列描述: SEQ ID NO: 28 :

AAGCTTCACG TGGACCAGCA AGCCAAGAGT G 31

(2) SEQ ID NO: 29 的信息:

15 (i) 序列特征:

(A) 长度: 25 个碱基对

(B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑结构: 线型

20 (ii) 分子类型: DNA(基因组的)

(ix) 特征:

(A) 名称/关键词: 外显子

(B) 位置: 1..25

(xi) 序列描述: SEQ ID NO: 29 :

25 AAGCTTTTTC CGTCCTTCCG TCCGG 25

(2) SEQ ID NO: 30 的信息:

(i) 序列特征:

(A) 长度: 24 个碱基对

30 (B) 类型: 核酸



(A) 名称/关键词: 外显子

(B) 位置: 1..25

(xi) 序列描述: SEQ ID NO: 32 :

GGGTGGTGAG AGTTCGTTGT CTGTC

25

5

(2) SEQ ID NO: 33 的信息:

(i) 序列特征:

(A) 长度: 25 个碱基对

(B) 类型: 核酸

10

(C) 链型: 单链

(D) 拓扑结构: 线型

(ii) 分子类型: DNA(基因组的)

(ix) 特征:

(A) 名称/关键词: 外显子

15

(B) 位置: 1..25

(xi) 序列描述: SEQ ID NO: 33 :

GAGCGATGAG GTACGGAAGA CTCTG

25

(2) SEQ ID NO: 34 的信息:

20

(i) 序列特征:

(A) 长度: 5856 个碱基对

(B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑结构: 线型

25

(ii) 分子类型: DNA(基因组的)

(ix) 特征:

(A) 名称/关键词: 外显子

(B) 位置: 1..5856

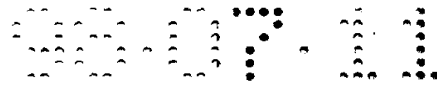
(xi) 序列描述: SEQ ID NO: 34 :

30



5

AGCGCCCAAT ACGCAAACCG CCTCTCCCCG CGCGTTGGCC GATTCATTAA TGCAGCTGGC 60
ACGACAGGTT TCCCRACTGG AAAGCGGGCA GTGAGCGCAA CGCAATTAAT GTGAGTTAGC 120
TCACTCATTAA GGCACCCCAG GCTTTACTACT TTATGCTTCC GGCTCGTATG TTGTGTGGAA 180
TTGTGAGCGG ATAACAATTT CACACAGGAA ACAGCTATGA CCATGATTAC GCCAAGCTTC 240
ACGTGGACCA GCAAGCCAAG AGTGAGTGTG GGCAGCACCC CCAGCCAGAG GGAGGCAGCC 300
AGGGCACAGG CATGACCCAG CAGGTGCTCG GCCATGATGC CGTCCTCGGG CAGCAGCTGC 360
CAGTGCTGAG CCTCGCGGAA GTAGGAGCCC TCCTGGGCCT CGCAGAAGTA GTGGCCGTAC 420
TGCTGCGCCG TGAGGTTCTC GATGAACAGG ATGCAGTTGG GGCTCTGGTG ACCAGGTTTCG 480
CAGCTCTGCT CCACGTTCTC CTTGTGGCGC CATGAGTAGG TGGCGTGGCG GGATTCCATG 540
GGGCAGCTCA GGTAGTAGCG AGAGTTTGGG GCCAGGGAAA CCTTCTGCAG TGGGGCCTTG 600
TCTGGTTTGG GGTGGGACA CTCCTTGTGT GGCTCGGCTG GATTAATGGA TTGCAGCACT 660
GACCGTTCGG AGCTGTAGAT GGAGATGCAG CGGCCCTGGT CCCAGCCGCA GTAGGGGTCT 720
CGGGACATGA GGCAACCGTG GCAGCCCCCG CCATAGACCT CACACAGGTC CAGGGGCACC 780
TGGCTCACCT CCCACTGGGA GCTCACATAC AGCTTCCTCC GCTCAGCATC CAGCGACATG 840
GTCTGGATGG CAGCCGCGCG GCGGAAGGGC TGGATCTCCA TGATGTTGAA GGCGAAGOTG 900
TGCTCCTGCT CCCCCGGTTC CACCACCTTG TGGATAGTGC CCCTGTCTGT AGTTAGGTAA 960
AGCACATGAA AGGTCTCCCC GTGGCTGGCT TGCATGCGGT GAACGGCCAC TTTCTGGTAG 1020
TGGTATTTAG AGTGGAACAA TGGCGTCTTC AGAGGCCCCA TGGGCTCCAC CCTCTGCGCC 1080
ACCTCTGGGT GACGGTCAGC CACCTGGAAG GTCTCTGTGG GTATCGGCTG CTGGTCTGGG 1140
AGGCACTTGC CAGGCCGCGG GTTGGGAAGG CTGAGTGGT AGCCCTTGAG TGAGGAGGTA 1200
CGGAAGACCT TGTCAATGTC ACCGAGGGAA TACACACAGA CGGCTGAGTA GTTCCAGGGG 1260
TTGGAGAAA CACCATAGAC CCTGGTGTCC CTCCACTGGC CGCTGGGGTC AGGGAGCAGG 1320
AAGACGTCTT GCAGCCTGTT GAAGTTCTTG TTGGTGGCAG CATCACTGCA TACCAGCATG 1380
GCTTTCAGAA AAGTGTTCCA CTGGAGACT GACAGTGAAC TTTCCCCACC CTGGTCCCCC 1440
CTGCACAACCT GGGCCACACG GGACACATTG AGAGGAGCCT CAGGATTCTT GTCAGGATTG 1500



	TCCTCTCGGA AGAAGTAGTA GATCTTGTCA TCGTAAGCCT GGTCTTGGTG CACGATGGTG	1560
	GCTTTGATGA ACTGTGGGTT CTGCATGACA GTATCACTGG TGTACAGCTC ACTCTCGCCC	1620
	CGGATGCGGC GGAACCGAGG GATCTTCCCA TTGTATTCTT GCTTCCGGAT GGTGGAATAC	1680
	ACCTCGTCCC CTTCAAACAG AACCAGGGAG TTCTCGTCCG GGCTGAAGGG GGCGTAGCCT	1740
5	CTCATCTCGC CAAGTGGCAC CACAGTGCCA TTCACCAGGT TCCAGCAGCT GGGGTGCCGG	1800
	GCGTTGGTGC CACAGGCCAG CAGCCCCTCA CTCCGCCTCT CCAGGAGAGT GATGTAGTTC	1860
	TCGCAGTCCC GCTTATCCAG ACAGGACCCC TTTGTGGAGC CGATATTCAC CGTGCGCACA	1920
	GATGCGTTCT TGCCCTCGGG GAAGTCAAAG AGGTAGACCT TGCCACGTCC TCCCACCCAC	1980
	ACAGAGGAGC TGCCTGGCTC GTGGAAAAGC ACCGTGTGCG GCTCAGTCTG GCCAAAGTCC	2040
	ACCCGGTCCT GCCCTACATG GCCTTTCAG ACGGCGAAGA TGCGGGGTCC GCTCCTTAGG	2100
	TGGCCCTGGG CGGAGGCGGC GGCCGCCAG AGCAGCAGCA GCAGCCGCAG CCGCAGCGGA	2160
	AGCCCCAACC GAGCCGGCGG GCCAGGGACG CGGGCGCGCG GTGCGCTGGG GGCGGCACGT	2220
	CCGGGCGGAG GAGGCGTCAT CCCAAGCCGA ATTCTGCAGA TATCCATCAC ACTGGCGGCC	2280
	GCTCGAGCAT GCATCTAGAG GGCCCAATTC GCCCTATAGT GAGTCGTATT ACAATTCACT	2340
	GGCCGTGTT TTACAACGTC GTGACTGGGA AAACCCTGGC GTTACCCAAC TTAATCGCCT	2400
	TGCAGCACAT CCCCCTTTCG CCAGCTGGCG TAATAGCGAA GAGGCCCGCA CCGATCGCCC	2460
	TTCCCAACAG TTGCGCAGCC TGAATGGCGA ATGGGACGCG CCCTGTAGCG GCGCATTAG	2520
	CGCGGCGGGT GTGGTGGTTA CGCGCAGCGT GACCGCTACA CTGCCAGCG CCCTAGCGCC	2580
	CGTCTCTTC GCTTTCTTCC CTTCCTTCT CGCCACGTT GCCGGCTTTC CCCGTCAAGC	2640
	TCTAATCGG GGGCTCCCTT TAGGGTCCG ATTTAGAGCT TTACGGCACC TCGACCGCAA	2700
	AAAACCTGAT TTGGGTGATG GTTCACGTAG TGGGCCATCG CCCTGATAGA CGGTTTTTCG	2760
	CCCTTGACG TTGGAGTCCA CGTCTTTAA TAGTGGACTC TTGTCCAAA CTGGAACAAC	2820
	ACTCAACCCT ATCGCGGTCT ATTCTTTGA TTTATAAGGG ATTTGCCGA TTTCGGCCTA	2880
	TTGGTTAAAA AATGAGCTGA TTTAACAAT TCAGGGCGCA AGGGCTGCTA AAGGAACCGG	2940
	AACACGTAGA AAGCCAGTCC GCAGAAACGG TGCTGACCCC GGATGAATGT CAGCTACTGG	3000
	GCTATCTGGA CAAGGAAAA CGCAAGCGCA AAGAGAAAGC AGGTAGCTTG CAGTGCGCTT	3060
	ACATGGCGAT AGCTAGACTG GGCGTTTTA TGGACAGCAA GCGAACCGGA ATTGCCAGCT	3120
	GGGGCGCCCT CTGGTAAGGT TGGGAAGCCC TGCAAAGTAA ACTGGATGGC TTTCTTGCCG	3180
	CCAAGGATCT GATGGCGCAG GGGATCAAGA TCTGATCAAG AGACAGGATG AGGATCGTTT	3240



5

CGCATGATTG	AACAAGATGG	ATTGCACGCA	GGTTCTCCGG	CCGCTTGGGT	GGAGAGGCTA	3300
TTCGGCTATG	ACTGGGCACA	ACAGACAATC	GGCTGCTCTG	ATGCCGCCGT	GTTCGGGCTG	3360
TCAGCGCAGG	GGCGCCCGGT	TCTTTTTGTC	AAGACCGACC	TGTCCGGTGC	CCTGAATGAA	3420
CTGCAGGACG	AGGCAGCGCG	GCTATCGTGG	CTGGCCACGA	CGGGCGTTCC	TTGCGCAGCT	3480
GTGCTCGACG	TTGTCACTGA	AGCGGGAAGG	GACTGGCTGC	TATTGGGCGA	AGTGCCGGGG	3540
CAGGATCTCC	TGTCATCTCG	CCTTGCTCCT	GCCGAGAAAG	TATCCATCAT	GGCTGATGCA	3600
ATGGCGCGGC	TGCATACGCT	TGATCCGGCT	ACCTGCCCAT	TCGACCACCA	AGCGAAACAT	3660
CGCATCGAGC	GAGCACGTAC	TCGGATGGAA	GCCGGTCTTG	TCGATCAGGA	TGATCTGGAC	3720
GAAGAGCATC	AGGGGCTCGC	GCCAGCCGAA	CTGTTCCGCA	GGCTCAAGGC	GCGCATGCCC	3780
GACGGCGAGG	ATCTCGTCGT	GATCCATGGC	GATGCCTGCT	TGCCGAATAT	CATGGTGGAA	3840
AATGGCCGCT	TTTCTGGATT	CAACGACTGT	GGCCGGCTGG	GTGTGGCGGA	CCGCTATCAG	3900
GACATAGCGT	TGGATACCCG	TGATATTGCT	GAAGAGCTTG	GCGGCGAATG	GGCTGACCGC	3960
TTCTCGTGC	TTTACGGTAT	CGCCGCTCCC	GATTGCGAGC	GCATCGCCTT	CTATCGCCTT	4020
CTTGACGAGT	TCTTCTGAAT	TGAAAAGGA	AGAGTATGAG	TATTCAACAT	TTCCGTGTGC	4080
CCCTTATTCC	CTTTTTTGCG	GCATTTTGCC	TTCTGTGTTT	TGCTCACCCA	GAAACGCTGG	4140
TGAAAGTAAA	AGATGCTGAA	GATCAGTTGG	GTGCACGAGT	GGGTTACATC	GAACTGGATC	4200
TCAACAGCGG	TAAGATCCTT	GAGAGTTTTC	GCCCCGAAGA	ACGTTTTCCA	ATGATGAGCA	4260
CTTTTAAAGT	TCTGCTATGT	CATACACTAT	TATCCCGTAT	TGACGCCGGG	CAAGAGCAAC	4320
TCGGTCCGGG	GGCGCGGTAT	TCTCAGAATG	ACTTGGTTGA	GTACTCACCA	GTCACAGAAA	4380
AGCATCTTAC	GGATGGCATG	ACAGTAAGAG	AATTATGCAG	TGCTGCCATA	ACCATGAGTG	4440
ATAACACTGC	GGCCAACTTA	CTTCTGACAA	CGATCGGAGG	ACCGAAGGAG	CTAACCGCTT	4500
TTTTGCACAA	CATGGGGGAT	CATGTAAGTC	GCCTTGATCG	TTGGGAACCG	GAGCTGAATG	4560
AAGCCATACC	AAACGACGAG	AGTGACACCA	CGATGCCTGT	AGCAATGCCA	ACAACGTTGC	4620
GCAAACATAT	AACTGGCGAA	CTACTTACTC	TAGCTTCCCG	GCAACAATTA	ATAGACTGGA	4680
TGGAGGCGGA	TAAAGTTGCA	GGACCACTTC	TGCGCTCGGC	CCTTCCGGCT	GGCTGGTTTA	4740
TTGCTGATAA	ATCTGGAGCC	GGTGAGCGTG	GGTCTCGCGG	TATCATTGCA	GCACTGGGGC	4800
CAGATGGTAA	GCCCTCCCGT	ATCGTAGTTA	TCTACACGAC	GGGGAGTCAG	GCAACTATGG	4860
ATGAACGAAA	TAGACAGATC	GCTGAGATAG	GTGCCTCACT	GATTAAGCAT	TGGTAACTGT	4920
CAGACCAAGT	TTACTCATAT	ATACTTTAGA	TTGATTTAAA	ACTTCATTTT	TAATTTAAAA	4980

	GGATCTAGGT GAAGATCCTT TTTGATAATC TCATGACCAA AATCCCTTAA CGTGAGTTTT	5040
	CGTCCACTG AGCGTCAGAC CCCGTAGAAA AGATCAAAGG ATCTTCTTGA GATCCTTTTT	5100
	TTCTGCGCGT AATCTGCTGC TTGCAAACAA AAAAACCACC GCTACCAGCG GTGGTTTGTT	5160
	TGCCGGATCA AGAGCTACCA ACTCTTTTTTC CGAAGGTAAC TGGCTTCAGC AGAGCGCAGA	5220
5	TACCAAATAC TGTCCTTCTA GTGTAGCCGT AGTTAGGCCA CCACTTCAAG AACTCTGTAG	5280
	CACCGCCTAC ATACCTCGCT CTGCTAATCC TGTTACCAGT GGCTGCTGCC AGTGGCGATA	5340
	AGTCGTGTCT TACCGGGTTG GACTCAAGAC GATAGTTACC GGATAAGGCG CAGCGGTCGG	5400
	GCTGAACGGG GGGTTCGTGC ACACAGCCCA GCTTGGAGCG AACGACCTAC ACCGAACTGA	5460
	GATACCTACA GCGTGAGCAT TGAGAAAGCG CCACGCTTCC CGAAGGGAGA AAGGCGGACA	5520
10	GGTATCCGGT AAGCGGCAGG GTCGGAACAG GAGAGCGCAC GAGGGAGCTT CCAGGGGGAA	5580
	ACGCCTGGTA TCTTTATAGT CCTGTCCGGT TTCGCCACCT CTGACTTGAG CGTCGATTTT	5640
	TGTGATGCTC GTCAGGGGGG CGGAGCCTAT GGAAAAACGC CAGCAACGCG GCCTTTTTAC	5700
	GGTTCCCTGGC CTTTTGCTGG CTTTTGCTC ACATGTTCTT TCCTGCGTTA TCCCCTGATT	5760
	CTGTGGATAA CCGTATTACC GCCTTTGAGT GAGCTGATAC CGCTCGCCGC AGCCGAACGA	5820
15	CCGAGCGCAG CGAGTCAGTG AGCGAGGAAG CGGAAG	5856

(2) SEQ ID NO: 35 的信息:

(i) 序列特征:

- (A) 长度: 7475 个碱基对
- (B) 类型: 核酸
- (C) 链型: 单链
- (D) 拓扑结构: 线型

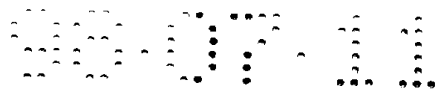
(ii) 分子类型: DNA(基因组的)

(ix) 特征:

- (A) 名称/关键词: 外显子
- (B) 位置: 1..7475

(xi) 序列描述: SEQ ID NO: 35:

	GACGGATCGG GAGATCTCCC GATCCCCTAT GGTCGACTCT CAGTACAATC TGCTCTGATG	60
	CCGCATAGTT AAGCCAGTAT CTGCTCCCTG CTTGTGTGTT GGAGGTCGCT GAGTAGTGCG	120
	CGAGCAAAT TTAAGCTACA ACAAGGCAAG GCTTGACCGA CAATTGCATG AAGAATCTGC	180
30	TTAGGGTTAG GCGTTTTGCG CTGCTTCGCG ATGTACGGGC CAGATATACG CGTTGACATT	240

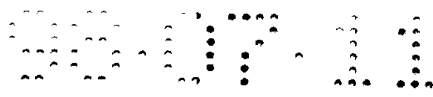


5

GATTATTGAC	TAGTATTAA	TAGTAATCAA	TTACGGGGTC	ATTAGTTCAT	AGCCCATATA	300
TGGAGTTCGG	CGTTACATAA	CTTACGGTAA	ATGGCCCGCC	TGGCTGACCG	CCCAACGACC	360
CCCGCCCAT	GACGTCAATA	ATGACGTATG	TTCCCATAGT	AACGCCAATA	GGGACTTTCC	420
ATTGACGTCA	ATGGGTGGAC	TATTTACGGT	AAACTGCCCA	CTTGGCAGTA	CATCAAGTGT	480
ATCATATGCC	AAGTACGCCC	CCTATTGACG	TCAATGACGG	TAAATGGCCC	GCCTGGCATT	540
ATGCCCAGTA	CATGACCTTA	TGGGACTTTC	CTACTTGCCA	GTACATCTAC	GTATTAGTCA	600
TCGCTATTAC	CATGGTGATG	CGGTTTTGGC	AGTACATCAA	TGGGCGTGGA	TAGCGGTTTG	660
ACTCACGGGG	ATTTCCAAGT	CTCCACCCCA	TTGACGTCAA	TGGGAGTTTG	TTTTGGCACC	720
AAAATCAACG	GGACTTTCCA	AAATGTCGTA	ACAACCTCCGC	CCCATTGACG	CAAATGGGCG	780
GTAGGCGTGT	ACGGTGGGAG	GTCTATATAA	GCAGAGCTCT	CTGGCTAACT	AGAGAACCCA	840
CTGCTTACTG	GCTTATCGAA	ATTAATACGA	CTCACTATAG	GGAGACCCAA	GCTGGCTAGC	900
GTTTAAACGG	GCCCTCTAGA	CTCGAGCGGC	CGCCACTGTG	CTGGATATCT	GCAGAATTCG	960
GCTTGGGATG	ACGCCTCCTC	CGCCCGGACG	TGCCGCCCCC	AGCGCACCGC	GCGCCCGCGT	1020
CCCTGGCCCC	CGGGCTCGGT	TGGGGCTTCC	GCTGCGGCTG	CGGCTGCTGC	TGCTGCTCTG	1080
GGCGGCCCGC	GCCTCCGCCC	AGGGCCACCT	AAGGAGCGGA	CCCCGCATCT	TCGCCGTCTG	1140
GAAAGGCCAT	GTAGGGCAGG	ACCGGGTGGG	CTTTGGCCAG	ACTGAGCCGC	ACACGGTGCT	1200
TTTCCACGAG	CCAGGCAGCT	CCTCTGTGTG	GGTGGGAGGA	CGTGGCAAGG	TCTACCTCTT	1260
TGACTTCCCC	GAGGGCAAGA	ACGCATCTGT	GCGCACGGTG	AATATCGGCT	CCACAAAGGG	1320
GTCCTGTCTG	GATAAGCGGG	ACTGCGAGAA	CTACATCACT	CTCCTGGAGA	GGCGGAGTGA	1380
GGGGCTGCTG	GCCTGTGGCA	CCAACGCCCG	GCACCCACG	TGCTGGAACC	TGGTGAATGG	1440
CACTGTGGTG	CCACTTGGCG	AGATGAGAGG	CTACGCCCCC	TTCAGCCCGG	ACGAGAACTC	1500
CCTGGTTCTG	TTGAAGGGG	ACGAGGTGTA	TTCCACCATC	CGGAAGCAGG	AATACAATGG	1560
GAAGATCCCT	CGGTTCCGCC	GCATCCGGGG	CGAGAGTGAG	CTGTACACCA	GTGATACTGT	1620
CATGCAGAAC	CCACAGTTCA	TCAAAGCCAC	CATCGTGCAC	CAAGACCAGG	CTTACGATGA	1680
CAAGATCTAC	TACTTCTTCC	GAGAGGACAA	TCCTGACAAG	AATCCTGAGG	CTCCTCTCAA	1740
TGTGTCCCGT	GTGGCCCAGT	TGTGCAGGGG	GGACCAGGGT	GGGGAAAGTT	CACTGTCACT	1800
CTCCAAGTGG	AACACTTTTC	TGAAAGCCAT	GCTGGTATGC	AGTGATGCTG	CCACCAACAA	1860
GAACTTCAAC	AGGCTGCAAG	ACGTCTTCTT	GCTCCCTGAC	CCCAGCGGCC	AGTGGAGGGA	1920
CACCAGGGTC	TATGGTGTTT	TCTCCAACCC	CTGGAECTAC	TCAGCCGTCT	GTGTGTATTTC	1980

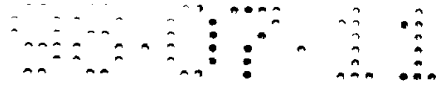


	ACAACACTCA ACCCTATCTC GGTCTATTCT TTTGATTTAT AAGGGATTTT GGGGATTTTCG	3720
	GCCTATTGGT TAAAAAATGA GCTGATTTAA CAAAAATTTA ACGCGAATTA ATTCTGTGGA	3780
	ATGTGTGTCA GTTAGGGTGT GGAAAGTCCC CAGGCTCCCC AGGCAGGCAG AAGTATGCAA	3840
	AGCATGCATC TCAATTAGTC AGCAACCAGG TGTGGAAAGT CCCCAGGCTC CCCAGCAGGC	3900
5	AGAAGTATGC AAAGCATGCA TCTCAATTAG TCAGCAACCA TAGTCCCGCC CCTAACTCCG	3960
	CCCATCCCGC CCCTAACTCC GCCCAGTTCC GCCCATTCTC CGCCCCATGG CTGACTAATT	4020
	TTTTTTATTT ATGCAGAGGC CGAGGCCCGCC TCTGCCTCTG AGCTATTCCA GAAGTAGTGA	4080
	GGAGGCTTTT TTGGAGGCCT AGGCTTTTGC AAAAAGCTCC CGGGAGCTTG TATATCCATT	4140
	TTCGGATCTG ATCAAGAGAC AGGATGAGGA TCGTTTCGCA TGATTGAACA AGATGGATTG	4200
	CACGCAGGTT CTCCGGCCGC TTGGGTGGAG AGGCTATTCG GCTATGACTG GGCACAACAG	4260
	ACAATCGGCT GCTCTGATGC CGCCGTGTTT CCGCTGTCAG CGCAGGGGCG CCCGGTTCTT	4320
	TTGTCAAGA CCGACCTGTC CGGTGCCCTG AATGAACTGC AGGACGAGGC AGCGCGGCTA	4380
	TCGTGGCTGG CCACGACGGG CGTTCCTTGC GCAGCTGTGC TCGACGTTGT CACTGAAGCG	4440
	GGAAGGGACT GGCTGCTATT GGGCGAAGTG CCGGGGCAGG ATCTCCTGTC ATCTCACCTT	4500
	GCTCCTGCCG AGAAAGTATC CATCATGGCT GATGCAATGC GCGGGCTGCA TACGCTTGAT	4560
	CCGGCTACCT GCCCATTGCA CCACCAAGCG AAACATCGCA TCGAGCGAGC ACGTACTCGG	4620
	ATGGAAGCCG GTCTTGTCGA TCAGGATGAT CTGGACGAAG AGCATCAGGG GCTCGCGCCA	4680
	GCCGAACTGT TCGCCAGGCT CAAGGCGCGC ATGCCCGACG GCGAGGATCT CGTCGTGACC	4740
	CATGGCGATG CCTGCTTGCC GAATATCATG GTGGAAAATG GCCGCTTTTC TGGATTCATC	4800
	GACTGTGGCC GGCTGGGTGT GCGGGACCGC TATCAGGACA TAGCGTTGGC TACCCGTGAT	4860
	ATTGCTGAAG AGCTTGGCGG CGAATGGGCT GACCGCTTCC TCGTGCTTTA CGGTATCGCC	4920
	GCTCCCGATT CGCAGCGCAT CGCCTTCTAT CGCCTTCTTG ACGAGTTCTT CTGAGCGGGA	4980
	CTCTGGGGTT CGAAATGACC GACCAAGCGA CGCCCAACCT GCCATCACGA GATTCGATT	5040
	CCACCGCCGC CTTCTATGAA AGGTTGGGCT TCGGAATCGT TTTCCGGGAC GCCGGCTGGA	5100
	TGATCCTCCA GCGCGGGGAT CTCATGCTGG AGTTCTTCGC CCACCCCAAC TTGTTTATTG	5160
	CAGCTTATAA TGGTTACAAA TAAAGCAATA GCATCACAAA TTTCACAAAT AAAGCATTTT	5220
	TTTCACTGCA TTCTAGTTGT GGTITGTCCA AACTCATCAA TGTATCTTAT CATGTCTGTA	5280
	TACCGTCGAC CTCTAGCTAG AGCTTGGCGT AATCATGGTC ATAGCTGTTT CCTGTGTGAA	5340
	ATTGTTATCC GCTCACAATT CCACACAACA TACGAGCCGG AAGCATAAAG TGTAAGCCT	5400



5

GGGGTGCCTA ATGAGTGAGC TAACTCACAT TAATTGCGTT GCGCTCACTG CCCGCTTTC 5460
AGTCGGGAAA CCTGTCGTGC CAGCTGCATT AATGAATCGG CCAACGCGCG GGGAGAGGCG 5520
GTTTGCATAT TGGGCGCTCT TCCGCTTCTT CGCTCACTGA CTCGCTGCGC TCGGTCGTTC 5580
GGCTGCGGCG AGCGGTATCA GCTCACTCAA AGGCGGTAAT ACGGTTATCC ACAGAATCAG 5640
GGGATAACGC AGGAAAGAAC ATGTGAGCAA AAGGCCAGCA AAAGGCCAGG AACCGTAAAA 5700
AGGCCGCGTT GCTGGCGTTT TTCCATAGGC TCCGCCCCCC TGACGAGCAT CACAAAAATC 5760
GACGCTCAAG TCAGAGGTGG CGAAACCCGA CAGGACTATA AAGATACCAG GCGTTTCCCC 5820
CTGGAAGCTC CCTCGTGC GC TCTCCTGTTT CGACCCTGCC GCTTACCGGA TACCTGTCCG 5880
CCTTTCTCCC TTCGGGAAGC GTGGCGCTTT CTCAATGCTC ACGCTGTAGG TATCTCAGTT 5940
CGGTGTAGGT CGTTCGCTCC AAGCTGGGCT GTGTGCACGA ACCCCCGTT CAGCCCGACC 6000
GCTGCGCCTT ATCCGGTAAC TATCGTCTTG AGTCCAACCC GGTAAGACAC GACTTATCGC 6060
CACTGGCAGC AGCCACTGGT AACAGGATTA GCAGAGCGAG GTATGTAGGC GGTGCTACAG 6120
AGTTCTTGAA GTGGTGGCCT AACTACGGCT AACTAGAAG GACAGTATTT GGTATCTGCG 6180
CTCTGCTGAA GCCAGTTACC TTCGGAAAAA GAGTTGGTAG CTCTTGATCC GGCAAAACAA 6240
CCACCGCTGG TAGCGGTGGT TTTTTGTTT GCAAGCAGCA GATTACGCGC AGAAAAAAG 6300
GATCTCAAGA AGATCCTTTG ATCTTTTCTA CGGGGTCTGA CGCTCAGTGG AACGAAAAC 6360
CACGTTAAGG GATTTTGGTC ATGAGATTAT CAAAAGGAT CTCACCTAG ATCCTTTTAA 6420
ATTAAAAATG AAGTTTTAAA TCAATCTAAA GTATATATGA GTAAACTTGG TCTGACAGTT 6480
ACCAATGCTT AATCAGTGAG GCACCTATCT CAGCGATCTG TCTATTTTCTG TCATCCATAG 6540
TTGCCTGACT CCCCCTCGTG TAGATAACTA CGATACGGGA GGGCTTACCA TCTGGCCCCA 6600
GTGCTGCAAT GATACCGCGA GACCCACGCT CACCGGCTCC AGATTTATCA GCAATAAACC 6660
AGCCAGCCGG AAGGGCCGAG CGCAGAAGTG GTCCTGCAAC TTTATCCGCC TCCATCCAGT 6720
CTATTAATTG TTGCCGGGAA GCTAGAGTAA GTAGTTCGCC AGTTAATAGT TTGCGCAACG 6780
TTGTTGCCAT TGCTACAGGC ATCGTGGTGT CACGCTCGTC GTTTGGTATG GCTTCATTCA 6840
GCTCCGGTTC CCAACGATCA AGGCGAGTTA CATGATCCCC CATGTTGTGC AAAAAAGCGG 6900
TTAGCTCCTT CGGTCTCCG ATCGTTGTCA GAAGTAAGTT GGCCGAGTG TTATCACTCA 6960
TGGTTATGGC AGCACTGCAT AATCTCTTA CTGTCATGCC ATCCGTAAGA TGCTTTTCTG 7020
TGA CTGGTGA G TACTCAACC AAGTCATTCT GAGAATAGTG TATGCGGCCA CCGAGTTGCT 7080



CTTGCCCGGC GTCAATACGG GATAATACCG CGCCACATAG CAGAACTTTA AAAGTGCTCA 7140
 TCATTGGAAA ACGTTCTTCG GGGCGAAAAC TCTCAAGGAT CTTACCGCTG TTGAGATCCA 7200
 GTTCGATGTA ACCCACTCGT GCACCCAAC TATCTTCAGC ATCTTTTACT TTCACCAGCG 7260
 TTTCTGGGTG AGCAAAAACA GGAAGGCAAA ATGCCGCAAA AAAGGGAATA AGGGCGACAC 7320
 5 GGAAATGTTG AATACTCATA CTCTTCCTTT TTCAATATTA TTGAAGCATT TATCAGGGTT 7380
 ATTGTCTCAT GAGCGGATAC ATATTTGAAT GTATTTAGAA AAATAAACAA ATAGGGGTTT 7440
 CGCGCACATT TCCCCGAAA GTGCCACCTG ACGTC 7475

(2) SEQ ID NO: 36 的信息:

(i) 序列特征:

- 10 (A) 长度: 8192 个碱基对
- (B) 类型: 核酸
- (C) 链型: 单链
- (D) 拓扑结构: 线型

(ii) 分子类型: DNA(基因组的)

15 (ix) 特征:

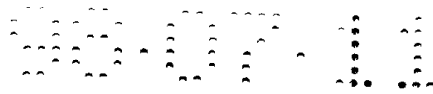
- (A) 名称/关键词: 外显子
- (B) 位置: 1..8192

(xi) 序列描述: SEQ ID NO: 36 :

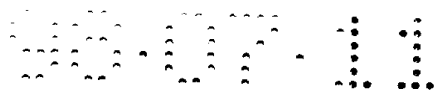
GACGGATCGG GAGATCTCCC GATCCCCTAT GGTCCGACTCT CAGTACAATC TGCTCTGATG 60
 20 CCGCATAGTT AAGCCAGTAT CTGCTCCCTG CTTGTGTGTT GGAGGTGCGT GAGTAGTGCG 120
 CGAGCAAAAT TTAAGCTACA ACAAGGCAAG GCTTGACCGA CAATTGCATG AAGAATCTGC 180
 TTAGGGTTAG GCGTTTTGCG CTGCTTCGCG ATGTACGGGC CAGATATACG CGTTGACATT 240
 GATTATTGAC TAGTTATTAA TAGTAATCAA TTACGGGGTC ATTAGTTCAT AGCCCATATA 300
 TGGAGTCCG CGTTACATAA CTTACGGTAA ATGGCCCGCC TGGCTGACCG CCCAACGACC 360
 25 CCCGCCATT GACGTCAATA ATGACGTATG TTCCCATAGT AACGCCAATA GGGACTTTCC 420
 ATTGACGTCA ATGGGTGGAC TATTTACGGT AACTGCCCA CTTGGCAGTA CATCAAGTGT 480
 ATCATATGCC AAGTACGCCC CCTATTGACG TCAATGACGG TAAATGGCCC GCCTGGCATT 540
 ATGCCAGTA CATGACCTTA TGGGACTTTC CTACTIONGCA GTACATCTAC GTATTAGTCA 600
 TCGCTATTAC CATGGTGATG CGGTTTTGGC AGTACATCAA TGGGCGTGGG TAGCGGTTTG 660
 30 ACTCACGGGG ATTTCCAAGT CTCCACCCCA TTGACGTCAA TGGGAGTTG TTTTGGCACC 720



GGGCTGCCAC GGTGCCTCA TGTCCCGAGA CCCCTACTGC GGCTGGGACC AGGGCCGCTG 2520
CATCTCCATC TACAGCTCCG AACGGTCAGT GCTGCAATCC ATTAATCCAG CCGAGCCACA 2580
CAAGGAGTGT CCCAACCCCA AACCAGACAA GGCCCCACTG CAGAAGGTTT CCCTGGCCCC 2640
AAACTCTCGC TACTACCTGA GCTGCCCCAT GGAATCCCGC CACGCCACCT ACTCATGGCG 2700
5 CCACAAGGAG AACGTGGAGC AGAGCTGCGA ACCTGGTCAC CAGAGCCCCA ACTGCATCCT 2760
GTTTCATCGAG AACCTCACGG CGCAGCAGTA CGGCCACTAC TTCTGCGAGG CCCAGGAGGG 2820
CTCCTACTTC CGCGAGGCTC AGCACTGGCA GCTGCTGCCC GAGGACGGCA TCATGGCCGA 2880
GCACCTGCTG GGTTCATGCCT GTGCCCTGGC TGCCTCCCTC TGGCTGGGGG TGCTGCCAC 2940
ACTCACTCTT GGCTTGCTGG TCCACATGGT GAGCAAGGGC GAGGAGCTGT TCACCGGGGT 3000
GGTGCCCATC CTGGTCGAGC TGGACGGCGA CGTAAACGGC CACAAGTCA GCGTGTCCGG 3060
CGAGGGCGAG GCGGATGCCA CCTACGGCAA GCTGACCCTG AAGTTCATCT GCACCACCGG 3120
CAAGCTGCCC GTGCCCTGGC CCACCCTCGT GACCACCCTG ACCTACGGCG TGCAGTGCTT 3180
CAGCCGCTAC CCCGACCACA TGAAGCAGCA CGACTTCTTC AAGTCCGCCA TGCCCGAAGG 3240
CTACGTCCAG GAGCGCACCA TCTTCTTCAA GGACGACGGC AACTACAAGA CCCGCGCCGA 3300
GGTGAAGTTC GAGGGCGACA CCCTGGTGAA CCGCATCGAG CTGAAGGGCA TCGACTTCAA 3360
GGAGGACGGC AACATCCTGG GGCACAAGCT GGAGTACAAC TACAACAGCC ACAACGTCTA 3420
TATCATGGCC GACAAGCAGA AGAACGGCAT CAAGGTGAAC TTCAAGATCC GCCACAACAT 3480
CGAGGACGGC AGCGTGCAGC TCGCCGACCA CTACCAGCAG AACACCCCA TCGGCGACGG 3540
CCCCGTGCTG CTGCCCGACA ACCACTACCT GAGCACCCAG TCCGCCCTGA GCAAAGACCC 3600
CAACGAGAAG CGCGATCACA TGGTCCTGCT GGAGTTCGTG ACCGCCGCCG GGATCACTCT 3660
CGGCATGGAC GAGCTGTACA AGGTGAAGCT TGGGCCCCGA CAAAACTCA TCTCAGAAGA 3720
GGATCTGAAT AGCGCCGTCG ACCATCATCA TCATCATCAT TGAGTTTAAA CCGCTGATCA 3780
GCCTCGACTG TGCCTTCTAG TTGCCAGCCA TCTGTTGTTT GCCCCTCCCC CGTGCCTTCC 3840
TTGACCCTGG AAGGTGCCAC TCCCACTGTC CTTTCCTAAT AAAATGAGGA AATTGCATCG 3900
CATTGTCTGA GTAGGTGTCA TTCTATTCTG GGGGGTGGGG TGGGGCAGGA CAGCAAGGGG 3960
GAGGATTGGG AAGACAATAG CAGGCATGCT GGGGATGCGG TGGGCTCTAT GGCTTCTGAG 4020
GCGGAAAGAA CCAGCTGGGG CTCTAGGGGG TATCCCCACG CGCCCTGTAG CGGCGCATT 4080
AGCGCGGCGG GTGTGGTGGT TACGCGCAGC GTGACCGCTA CACTTGCCAG CGCCCTAGCG 4140
CCCGCTCCTT TCGCTTTCTT CCCTTCCTTT CTCGCCACGT TCGCCGGCTT TCCCCGTCAA 4200



GCTCTAAATC GGGGCATCCC TTTAGGGTTC CGATTTAGTG CTTTACGGCA CCTCGACCCC 4260
AAAAAAGTTG ATTAGGGTGA TGGTTCACGT AGTGGGCCAT CGCCCTGATA GACGGTTTTT 4320
CGCCCTTTGA CGTTGGAGTC CACGTTCTTT AATAGTGGAC TCTTGTTCCA AACTGGAACA 4380
5 ACACTCAACC CTATCTCGGT CTATTCTTTT GATTTATAAG GGATTTTGGG GATTTCGGCC 4440
TATTGGTTAA AAAATGAGCT GATTTAACAA AAATTTAACG CGAATTAATT CTGTGGAATG 4500
TGTGTCAGTT AGGGTGTGGA AAGTCCCAG GCTCCCAGG CAGGCAGAAG TATGCAAAGC 4560
ATGCATCTCA ATTAGTCAGC AACCAGGTGT GGAAAGTCCC CAGGCTCCCC AGCAGGCAGA 4620
AGTATGCAAA GCATGCATCT CAATTAGTCA GCAACCATAG TCCCGCCCCT AACTCCGCCC 4680
ATCCCGCCCC TAACTCCGCC CAGTTCCGCC CATTCTCCGC CCCATGGCTG ACTAATTTTT 4740
TTTATTTATG CAGAGGCCGA GGCCGCCTCT GCCTCTGAGC TATTCCAGAA GTAGTGAGGA 4800
GGCTTTTTTG GAGGCCTAGG CTTTTGCAAA AAGCTCCCGG GAGCTTGTAT ATCCATTTTC 4860
GGATCTGATC AAGAGACAGG ATGAGGATCG TTTCCGATGA TTGAACAAGA TGGATTGCAC 4920
GCAGGTTCTC CGGCCGCTTG GGTGGAGAGG CTATTCGGCT ATGACTGGGC ACAACAGACA 4980
ATCGGCTGCT CTGATGCCGC CGTGTTCGGG CTGTCAGCGC AGGGGCGCCC GGTCTTTTTT 5040
GTCAAGACCG ACCTGTCCGG TGCCCTGAAT GAACTGCAGG ACGAGGCAGC GCGGCTATCG 5100
TGGCTGGCCA CGACGGGCGT TCCTTGCGCA GCTGTGCTCG ACGTTGTCAC TGAAGCGGGA 5160
AGGGACTGGC TGCTATTGGG CGAAGTGCCG GGGCAGGATC TCCTGTCATC TCACCTTGCT 5220
CCTGCCGAGA AAGTATCCAT CATGGCTGAT GCAATGCGGC GGCTGCATAC GCTTGATCCG 5280
GCTACCTGCC CATTGACCA CCAAGCGAAA CATCGCATCG AGCGAGCACG TACTCGGATG 5340
GAAGCCGGTC TTGTCGATCA GGATGATCTG GACGAAGAGC ATCAGGGGCT CGCGCCAGCC 5400
GAACTGTTCC CCAGGCTCAA GCGCGCATG CCCGACGGCG AGGATCTCGT CGTGACCCAT 5460
GGCGATGCCT GCTTGCCGAA TATCATGGTG GAAAATGGCC GCTTTTCTGG ATTCATCGAC 5520
TGTGGCCGGC TGGGTGTGGC GGACCGCTAT CAGGACATAG CGTTGGCTAC CCGTGATATT 5580
GCTGAAGAGC TTGGCGGCGA ATGGGCTGAC CGCTTCCTCG TGCTTTACGG TATCGCCGCT 5640
CCCGATTGCG AGCGCATCGC CTTCTATCGC CTTCTTGACG AGTTCTTCTG AGCGGGACTC 5700
TGGGGTTCGA AATGACCGAC CAAGCGACGC CCAACCTGCC ATCACGAGAT TCGATTCCA 5760
CCGCCGCCTT CTATGAAAGG TTGGGCTTCG GAATCGTTTT CCGGGACGCC GGCTGGATGA 5820
TCCTCCAGCG CGGGGATCTC ATGCTGGAGT TCTTCGCCCA CCCCAACTTG TTTATTGCAG 5880



	CTTATAATGG	TTACAAATAA	AGCAATAGCA	TCACAAATTT	CACAAATAAA	GCATTTTTTT	5940
	CACTGCATTC	TAGTTGTGGT	TTGTCCAAAC	TCATCAATGT	ATCTTATCAT	GTCTGTATAC	6000
	CGTCGACCTC	TAGCTAGAGC	TTGGCGTAAT	CATGGTCATA	GCTGTTTCCT	GTGTGAAATT	6060
	GTTATCCGCT	CACAATTCCA	CACAACATAC	GAGCCGGAAG	CATAAAGTGT	AAAGCCTGGG	6120
5	GTGCCTAATG	AGTGAGCTAA	CTCACATTA	TTGCGTTGCG	CTCACTGCCC	GCTTTCCAGT	6180
	CGGGAACCT	GTCGTGCCAG	CTGCATTAAT	GAATCGGCCA	ACGCGCGGGG	AGAGGCGGTT	6240
	TGCGTATTGG	GCGCTCTTCC	GCTTCCTCGC	TCACTGACTC	GCTGCGCTCG	GTCGTTCCGGC	6300
	TGCGGCGAGC	GGTATCAGCT	CACTCAAAGG	CGGTAATACG	GTTATCCACA	GAATCAGGGG	6360
	ATAACGCAGG	AAAGAACATG	TGAGCAAAAG	GCCAGCAAAA	GGCCAGGAAC	CGTAAAAAGG	6420
	CCGCGTTGCT	GGCGTTTTTC	CATAGGCTCC	GCCCCCCTGA	CGAGCATCAC	AAAAATCGAC	6480
	GCTCAAGTCA	GAGGTGGCGA	AACCCGACAG	GACTATAAAG	ATACCAGGCG	TTTCCCCCTG	6540
	GAAGCTCCCT	CGTGCGCTCT	CCTGTTCCGA	CCCTGCCGCT	TACCGGATAC	CTGTCCGCCT	6600
	TTCTCCCTTC	GGGAAGCGTG	GCGCTTCTC	AATGCTCAGC	CTGTAGGTAT	CTCAGTTCGG	6660
	TGTAGGTCGT	TCGCTCCAAG	CTGGGCTGTG	TGCACGAACC	CCCCGTTCAG	CCCGACCGCT	6720
	GCGCCTTATC	CGGTAACTAT	CGTCTTGAGT	CCAACCCGGT	AAGACACGAC	TTATCGCCAC	6780
	TGGCAGCAGC	CACTGGTAAC	AGGATTAGCA	GAGCGAGGTA	TGTAGGCGGT	GCTACAGAGT	6840
	TCTTGAAGTG	GTGGCCTAAC	TACGGCTACA	CTAGAAGGAC	AGTATTTGGT	ATCTGCGCTC	6900
	TGCTGAAGCC	AGTTACCTTC	GGAAAAAGAG	TTGGTAGCTC	TTGATCCGGC	AAACAAACCA	6960
	CCGCTGGTAG	CGGTGGTTTT	TTTGTTTGCA	AGCAGCAGAT	TACGCGCAGA	AAAAAAGGAT	7020
	CTCAAGAAGA	TCCTTTGATC	TTTTCTACGG	GGTCTGACGC	TCAGTGGAAC	GAAAACTCAC	7080
	GTTAAGGGAT	TTTGGTCATG	AGATTATCAA	AAAGGATCTT	CACCTAGATC	CTTTTAAATT	7140
	AAAAATGAAG	TTTTAAATCA	ATCTAAAGTA	TATATGAGTA	AACTTGGTCT	GACAGTTACC	7200
	AATGCTTAAT	CAGTGAGGCA	CCTATCTCAG	CGATCTGTCT	ATTTCGTTCA	TCCATAGTTG	7260
	CCTGACTCCC	CGTCGTGTAG	ATAACTACGA	TACGGGAGGG	CTTACCATCT	GGCCCCAGTG	7320
	CTGCAATGAT	ACCGCGAGAC	CCACGCTCAC	CGGCTCCAGA	TTTATCAGCA	ATAAACCAGC	7380
	CAGCCGGAAG	GGCCGAGCGC	AGAAGTGGTC	CTGCAACTTT	ATCCGCCTCC	ATCCAGTCTA	7440
	TTAATTGTTG	CCGGGAAGCT	AGAGTAAGTA	GTTCCGCCAGT	TAATAGTTTG	CGCAACGTTG	7500
	TTGCCATTGC	TACAGGCATC	GTGGTGTCAC	GCTCGTCGTT	TGGTATGGCT	TCATTCAGCT	7560
	CCGGTTCCCA	ACGATCAAGG	CGAGTTACAT	GATCCCCCAT	GTTGTGCAAA	AAAGCGGTTA	7620

AGATCTCGGC CGCATATTAA GTGCATTGTT CTCGATACCG CTAAGTGCAT TGTTCTCGTT 60
 AGCTCGATGG ACAAGTGCAT TGTTCTCTTG CTGAAAGCTC GATGGACAAG TGCATTGTTC 120
 TCTTGCTGAA AGCTCGATGG ACAAGTGCAT TGTTCTCTTG CTGAAAGCTC AGTACCCGGG 180
 AGTACCCTCG ACCGCCGGAG TATAAATAGA GGCGCTTCGT CTACGGAGCG ACAATTC AAT 240
 5 TCAAACAAGC AAAGTGAACA CGTCGCTAAG CGAAAGCTAA GCAAATAAAC AAGCGCAGCT 300
 GAACAAGCTA AAC AATCTG C AGTAAAGTGC AAGTTAAAGT GAATCAATTA AAAGTAACCA 360
 GCAACCAAGT AAATCAACTG CAACTACTGA AATCTGCCAA GAAGTAATTA TTGAATACAA 420
 GAAGAGAACT CTGAATACTT TCAACAAGTT ACCGAGAAAG AAGAACTCAC ACACAGCTAG 480
 CGTTTAAACT TAAGCTTGGT ACCGAGCTCG GATCCACTAG TCCAGTGTGG TGG AATTCCG 540

10

(2) SEQ ID NO: 37 的信息:

(i) 序列特征:

- (A) 长度: 7000 个碱基对
- (B) 类型: 核酸
- (C) 链型: 单链
- (D) 拓扑结构: 线型

15

(ii) 分子类型: DNA(基因组的)

(ix) 特征:

- (A) 名称/关键词: 外显子
- (B) 位置: 1..7000

20

(xi) 序列描述: SEQ ID NO: 37:

GCTCCTTCGG TCCTCCGATC GTTGTCAGAA GTAAGTTGGC CGCAGTGTTA TCACTCATGG 7680
 TTATGGCAGC ACTGCATAAT TCTCTTACTG TCATGCCATC CGTAAGATGC TTTTCTGTGA 7740
 CTGGTGAGTA CTC AACCAAG TCATTCTGAG AATAGTGTAT GCGGCGACCG AGTTGCTCTT 7800
 GCCCGGCGTC AATACGGGAT AATACCGCGC CACATAGCAG AACTTTAAAA GTGCTCATCA 7860
 25 TTGGAAAACG TTCTTCGGGG CGAAA AACTCT CAAGGATCTT ACCGCTGTTG AGATCCAGTT 7920
 CGATGTAACC CACTCGTGCA CCCAACTGAT CTCAGCATC TTTTACTTTC ACCAGCGTTT 7980
 CTGGGTGAGC AAAACAGGA AGGCAAAATG CCGCAAAAAA GGAATAAGG GCGACACGGA 8040
 AATGTTGAAT ACTCATACTC TTCCTTTTTC AATATTATTG AAGCATTAT CAGGGTTATT 8100
 30 GTCTCATGAG CGGATACATA TTTGAATGTA TTAGAAAAA TAAACAAATA GGGGTTCGCG 8160
 GCACATTTC CCGAAAAGTG CCACCTGACG TC 8192

103

CTTGGGATGA CGCCTCCTCC GCCCGGACGT GCCGCCCCCA GCGCACCGCG CGCCCGCGTC 600

CCTGGCCCCG CGGCTCGGTT GGGGCTTCCG CTGCGGCTGC GGCTGCTGCT GCTGCTCTGG 660

GCGGCCGCGG CCTCCGCCCA GGGCCACCTA AGGAGCGGAC CCCGCATCTT CGCCGTCTGG 720

AAAGGCCATG TAGGGCAGGA CCGGGTGGAC TTTGGCCAGA CTGAGCCGCA CACGGTGCTT 780

5 TTCCACGAGC CAGGCAGCTC CTCTGTGTGG GTGGGAGGAC GTGGCAAGGT CTACCTCTTT 840

GACTTCCCCG AGGGCAAGAA CGCATCTGTG CGCACGGTGA ATATCGGCTC CACAAAGGGG 900

TCCTGTCTGG ATAAGCGGGA CTGCGAGAAC TACATCACTC TCCTGGAGAG GCGGAGTGAG 960

GGGCTGCTGG CCTGTGGCAC CAACGCCCGG CACCCCAGCT GCTGGAACCT GGTGAATGGC 1020

ACTGTGGTGC CACTTGGCGA GATGAGAGGC TACGCCCCCT TCAGCCCGGA CGAGAACTCC 1080

CTGGTTCTGT TTGAAGGGGA CGAGGTGTAT TCCACCATCC GGAAGCAGGA ATACAATGGG 1140

AAGATCCCTC GGTTCGCGG CATCCGGGGC GAGAGTGAGC TGTACACCAG TGATACTGTC 1200

ATGCAGAACC CACAGTTCAT CAAAGCCACC ATCGTGCACC AAGACCAGGC TTACGATGAC 1260

AAGATCTACT ACTTCTTCCG AGAGGACAAT CCTGACAAGA ATCCTGAGGC TCCTCTCAAT 1320

GTGTCCCGTG TGGCCCAAGT GTGCAGGGGG GACCAGGGTG GGGAAAGTTC ACTGTCAGTC 1380

TCCAAGTGGA ACACTTTTCT GAAAGCCATG CTGGTATGCA GTGATGCTGC CACCAACAAG 1440

AACTTCAACA GGCTGCAAGA CGTCTTCCTG CTCCCTGACC CCAGCGGCCA GTGGAGGGAC 1500

ACCAGGGTCT ATGGTGT TTT CTCCAACCCC TGGAATACT CAGCCGTCTG TGTGTATTCC 1560

CTCGGTGACA TTGACAAGGT CTTCGGTACC TCCTCACTCA AGGGCTACCA CTCAAGCCTT 1620

CCCAACCCGC GGCCTGGCAA GTGCCTCCCA GACCAGCAGC CGATACCCAC AGAGACCTTC 1680

CAGGTGGCTG ACCGTCACCC AGAGGTGGCG CAGAGGGTGG AGCCCATGGG GCCTCTGAAG 1740

ACGCCATTGT TCCACTCTAA ATACCACTAC CAGAAAGTGG CCGTTCACCG CATGCAAGCC 1800

AGCCACGGGG AGACCTTTCA TGTGCTTTAC CTAATACTACAG ACAGGGGCAC TATCCACAAG 1860

GTGGTGGAAC CGGGGGAGCA GGAGCACAGC TTCGCCTTCA ACATCATGGA GATCCAGCCC 1920

TTCCGCCGCG CGGCTGCCAT CCAGACCATG TCGCTGGATG CTGAGCGGAG GAAGCTGTAT 1980

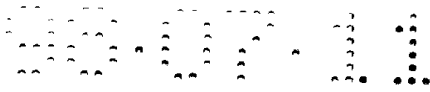
GTGAGCTCCC AGTGGGAGGT GAGCCAGGTG CCCCTGGACC TGTGTGAGGT CTATGGCGGG 2040

GGCTGCCACG GTTGCCTCAT GTCCCAGAC CCCTACTGCG GCTGGGACCA GGGCCGCTGC 2100

ATCTCCATCT ACAGCTCCGA ACGGTCAGTG CTGCAATCCA TTAATCCAGC CGAGCCACAC 2160

AAGGAGTGTC CCAACCCCAA ACCAGACAAG GCCCCACTGC AGAAGGTTTC CCTGGCCCCA 2220

AACTCTCGCT ACTACCTGAG CTGCCCCATG GAATCCCGCC ACGCCACCTA CTCATGGCGC 2280



	CACAAGGAGA	ACGTGGAGCA	GAGCTGCGAA	CCTGGTCACC	AGAGCCCCAA	CTGCATCCTG	2340
	TTCATCGAGA	ACCTCACGGC	GCAGCAGTAC	GGCCACTACT	TCTGCGAGGC	CCAGGAGGGC	2400
	TCCTACTTCC	GCGAGGCTCA	GCACTGGCAG	CTGCTGCCCG	AGGACGGCAT	CATGGCCGAG	2460
5	CACCTGCTGG	GTCATGCCTG	TGCCCTGGCT	GCCTCCCTCT	GGCTGGGGGT	GCTGCCACCA	2520
	CTCACTCTTG	GCTTGCTGGT	CCACGTGAAG	CTTGGGCCCG	TTTAAACCCG	CTGATCAGCC	2580
	TCGACTGTGC	CTTCTAGTTG	CCAGCCATCT	GTTGTTTGCC	CCTCCCCCGT	GCCTTCCTTG	2640
	ACCCTGGAAG	GTGCCACTCC	CACTGTCCTT	TCCTAATAAA	ATGAGGAAAT	TGCATCGCAT	2700
	TGTCTGAGTA	GGTGCATTC	TATTCTGGGG	GGTGGGGTGG	GGCAGGACAG	CAAGGGGGAG	2760
	GATTGGAAG	ACAATAGCAG	GCATGCTGGG	GATGCGGTGG	GCTCTATGGC	TTCTGAGGCG	2820
	GAAAGAACCA	GCTGGGGCTC	TAGGGGGTAT	CCCCACGCGC	CCTGTAGCGG	CGCATTAAGC	2880
	GCGGCGGGTG	TGGTGGTTAC	GCGCAGCGTG	ACCGCTACAC	TGCCAGCGC	CCTAGCGCCC	2940
	GCTCCTTTCG	CTTTCTTCCC	TTCCTTTCTC	GCCACGTTTCG	CGGGCTTTC	CCGTCAAGCT	3000
	CTAAATCGGG	GCATCCCTTT	AGGGTCCGA	TTTAGTGCTT	TACGGCACCT	CGACCCCAA	3060
	AAACTTGATT	AGGGTGATGG	TTCACGTAGT	GGGCCATCGC	CCTGATAGAC	GGTTTTTCGC	3120
	CCTTTGACGT	TGGAGTCCAC	GTTCTTTAAT	AGTGGACTCT	TGTTCCAAAC	TGGAACAACA	3180
	CTCAACCCTA	TCTCGGTCTA	TTCTTTTGAT	TTATAAGGGA	TTTTGGGGAT	TTCCGGCCTAT	3240
	TGGTTAAAAA	ATGAGCTGAT	TTAACAAAAA	TTAACGCGA	ATTAATTCTG	TGGAATGTGT	3300
	GTCAGTTAGG	GTGTGGAAAG	TCCCCAGGCT	CCCCAGGCAG	GCAGAAGTAT	GCAAAGCATG	3360
	CATCTCAATT	AGTCAGCAAC	CAGGTGTGGA	AAGTCCCCAG	GCTCCCCAGC	AGGCAGAAGT	3420
	ATGCAAAGCA	TGCATCTCAA	TTAGTCAGCA	ACCATAGTCC	CGCCCCTAAC	TCCGCCCATC	3480
	CGGCCCTTAA	CTCCGCCCAG	TTCCGCCCAT	TCTCCGCCCC	ATGGCTGACT	AATTTTTTTT	3540
	ATTTATGCAG	AGGCCGAGGC	CGCCTCTGCC	TCTGAGCTAT	TCCAGAAGTA	GTGAGGAGGC	3600
	TTTTTTGGAG	GCCTAGGCTT	TTGCAAAAAG	CTCCCGGGAG	CTTGATATATC	CATTTTCGGA	3660
	TCTGATCAAG	AGACAGGATG	AGGATCGTTT	CGCATGATTG	AACAAGATGG	ATTGCACGCA	3720
	GGTTCTCCGG	CCGCTGGGT	GGAGAGGCTA	TTCGGCTATG	ACTGGGCACA	ACAGACAATC	3780
	GGCTGCTCTG	ATGCCGCCGT	GTTCCGGCTG	TCAGCGCAGG	GGCGCCCGGT	TCTTTTTGTC	3840
	AAGACCGACC	TGTCCGGTGC	CCTGAATGAA	CTGCAGGACG	AGGCAGCGCG	GCTATCGTGG	3900
	CTGGCCACGA	CGGGCGTTC	TTGCGCAGCT	GTGCTCGACG	TTGTCACTGA	AGCGGGAAGG	3960



5
10
15
20

TGAAGCCAGT TACCTTCGGA AAAAGAGTTG GTAGCTCTTG ATCCGGCAAA CAAACCACCG 5760
CTGGTAGCGG TGGTTTTTTT GTTTCGAAGC AGCAGATTAC GCGCAGAAAA AAAGGATCTC 5820
AAGAAGATCC TTTGATCTTT TCTACGGGGT CTGACGCTCA GTGGAACGAA AACTCACGTT 5880
AAGGGATTTT GGTCATGAGA TTATCAAAAA GGATCTTCAC CTAGATCCTT TTAAATTAAA 5940
AATGAAGTTT TAAATCAATC TAAAGTATAT ATGAGTAAAC TTGGTCTGAC AGTTACCAAT 6000
GCTTAATCAG TGAGGCACCT ATCTCAGCGA TCTGTCTATT TCGTTCATCC ATAGTTGCCT 6060
GACTCCCCGT CGTGTAGATA ACTACGATAC GGGAGGGCTT ACCATCTGGC CCCAGTGCTG 6120
CAATGATACC GCGAGACCCA CGCTCACCGG CTCCAGATTT ATCAGCAATA AACCAGCCAG 6180
CCGGAAGGGC CGAGCGCAGA AGTGGTCCTG CAACTTTATC CGCCTCCATC CAGTCTATTA 6240
ATTGTTGCCG GGAAGCTAGA GTAAGTAGTT CGCCAGTTAA TAGTTTGCGC AACGTTGTTG 6300
CCATTGCTAC AGGCATCGTG GTGTCAGCTT CGTCGTTTGG TATGGCTTCA TTCAGCTCCG 6360
GTTCCCAACG ATCAAGGCGA GTTACATGAT CCCCCATGTT GTGCAAAAAA GCGGTTAGCT 6420
CCTTCGGTCC TCCGATCGTT GTCAGAAGTA AGTTGGCCGC AGTGTTATCA CTCATGGTTA 6480
TGGCAGCACT GCATAATTCT CTTACTGTCA TGCCATCCGT AAGATGCTTT TCTGTGACTG 6540
GTGAGTACTC AACCAAGTCA TTCTGAGAAT AGTGATGCG GCGACCGAGT TGCTCTTGCC 6600
CGGCGTCAAT ACGGGATAAT ACCGCGCCAC ATAGCAGAAC TTTAAAAGTG CTCATCATTG 6660
GAAAACGTTT TTCGGGGCGA AACTCTCAA GGATCTTACC GCTGTTGAGA TCCAGTTCGA 6720
TGTAACCCAC TCGTGCACCC AACTGATCTT CAGCATCTTT TACTTTCACC AGCGTTTCTG 6780
GGTGAGCAAA AACAGGAAGG CAAAATGCCG CAAAAAAGGG AATAAGGGCG ACACGGAAAT 6840
GTTGAATACT CATACTCTTC CTTTTTCAAT ATTATTGAAG CATTATCAG GGTATTGTC 6900
TCATGAGCGG ATACATATTT GAATGTATTT AGAAAAATAA ACAAATAGGG GTTCCGCGCA 6960
CATTCCCCG AAAAGTGCCA CCTGACGTCG ACGGATCGGG 7000

(2) SEQ ID NO: 38 的信息:

(i) 序列特征:

- 25 (A) 长度: 7108 个碱基对
(B) 类型: 核酸
(C) 链型: 单链
(D) 拓扑结构: 线型

(ii) 分子类型: DNA(基因组的)

- 30 (xi) 序列描述: SEQ ID NO: 38 :



AGATCTCGGC CGCATATTAA GTGCATTGTT CTCGATACCG CTAAGTGCAT TGTTCTCGTT 60
AGCTCGATGG ACAAGTGCAT TGTTCTCTTG CTGAAAGCTC GATGGACAAG TGCATTGTTC 120
TCTTGCTGAA AGCTCGATGG ACAAGTGCAT TGTTCTCTTG CTGAAAGCTC AGTACCCGGG 180
5 AGTACCCTCG ACCGCCGGAG TATAAATAGA GCGCCTTCGT CTACGGAGCG ACAATTCAAT 240
TCAAACAAGC AAAGTGAACA CGTCGCTAAG CGAAAGCTAA GCAAATAAAC AAGCGCAGCT 300
GAACAAGCTA AACAATCTGC AGTAAAGTGC AAGTTAAAGT GAATCAATTA AAAGTAACCA 360
GCAACCAAGT AAATCAACTG CAACTACTGA AATCTGCCAA GAAGTAATTA TTGAATACAA 420
GAAGAGAACT CTGAATACTT TCAACAAGTT ACCGAGAAAG AAGAACTCAC ACACAGCTAG 480
CGTTTAAACT TAAGCTTGGT ACCGAGCTCG GATCCACTAG TCCAGTGTGG TGGAATTCCG 540
CTTGGGATGA CGCCTCCTCC GCCCGGACGT GCCGCCCCCA GCGCACCCGG CGCCCGCGTC 600
CCTGGCCCCG CGGCTCGGTT GGGGCTTCCG CTGCGGCTGC GGCTGCTGCT GCTGCTCTGG 660
GCGGCCGCCG CCTCCGCCCA GGGCCACCTA AGGAGCGGAC CCCGCATCTT CGCCGTCTGG 720
AAAGGCCATG TAGGGCAGGA CCGGGTGGAC TTTGGCCAGA CTGAGCCGCA CACGGTGCTT 780
TTCCACGAGC CAGGCAGCTC CTCTGTGTGG GTGGGAGGAC GTGGCAAGGT CTACCTCTTT 840
GACTTCCCCG AGGGCAAGAA CGCATCTGTG CGCACGGTGA ATATCGGCTC CACAAAGGGG 900
TCCTGTCTGG ATAAGCGGGA CTGCGAGAAC TACATCACTC TCCTGGAGAG GCGGAGTGAG 960
GGGCTGCTGG CCTGTGGCAC CAACGCCCGG CACCCCAGCT GCTGGAACCT GGTGAATGGC 1020
ACTGTGGTGC CACTTGGCGA GATGAGAGGC TACGCCCCCT TCAGCCCGGA CGAGAACTCC 1080
CTGGTTCTGT TTGAAGGGGA CGAGGTGTAT TCCACCATCC GGAAGCAGGA ATACAATGGG 1140
AAGATCCCTC GGTTCCGCCG CATCCGGGGC GAGAGTGAGC TGTACACCAG TGATACTGTC 1200
ATGCAGAACC CACAGTTCAT CAAAGCCACC ATCGTGCACC AAGACCAGGC TTACGATGAC 1260
AAGATCTACT ACTTCTTCCG AGAGGACAAT CCTGACAAGA ATCCTGAGGC TCCTCTCAAT 1320
GTGTCCCGTG TGGCCAGTT GTGCAGGGGG GACCAGGGTG GGGAAAGTTC ACTGTCAGTC 1380
TCCAAGTGGA ACACTTTTCT GAAAGCCATG CTGGTATGCA GTGATGCTGC CACCAACAAG 1440
AACTTCAACA GGCTGCAAGA CGTCTTCCTG CTCCCTGACC CCAGCGGCCA GTGGAGGGAC 1500
ACCAGGGTCT ATGGTGTTTT CTCCAACCCC TGGAACTACT CAGCCGTCTG TGTGTATTCC 1560
CTCGGTGACA TTGACAAGGT CTTCCGTACC TCCTCACTCA AGGGCTACCA CTCRAGCCTT 1620
CCCAACCCGC GGCCTGGCAA GTGCCTCCCA GACCAGCAGC CGATACCCAC AGAGACCTTC 1680

108

CAGGTGGCTG ACCGTCACCC AGAGGTGGCG CAGAGGGTGG AGCCCATGGG GCCTCTGAAG 1740
ACGCCATTGT TCCACTCTAA ATACCACTAC CAGAAAGTGG CCGTTCACCG CATGCAAGCC 1800
AGCCACGGGG AGACCTTTCA TGTGCTTTAC CTAACTACAG ACAGGGGCAC TATCCACAAG 1860
5 GTGGTGGAAAC CGGGGGAGCA GGAGCACAGC TTCGCCTTCA ACATCATGGA GATCCAGCCC 1920
TTCCGCCCGC CGGCTGCCAT CCAGACCATG TCGCTGGATG CTGAGCGGAG GAAGCTGTAT 1980
GTGAGCTCCC AGTGGGAGGT GAGCCAGGTG CCCCTGGACC TGTGTGAGGT CTATGGCGGG 2040
GGCTGCCACG GTTGCCTCAT GTCCCGAGAC CCCTACTGCG GCTGGGACCA GGGCCGCTGC 2100
ATCTCCATCT ACAGCTCCGA ACGGTCAGTG CTGCAATCCA TTAATCCAGC CGAGCCACAC 2160
AAGGAGTGTC CCAACCCCAA ACCAGACAAG GCCCCACTGC AGAAGGTTTC CCTGGCCCCA 2220
AACTCTCGCT ACTACCTGAG CTGCCCCATG GAATCCCGCC ACGCCACCTA CTCATGGCGC 2280
CACAAGGAGA ACGTGGAGCA GAGCTGCGAA CCTGGTCACC AGAGCCCCAA CTGCATCCTG 2340
TTCATCGAGA ACCTCACGGC GCAGCAGTAC GGCCACTACT TCTGCGAGGC CCAGGAGGGC 2400
TCCTACTTCC GCGAGGCTCA GCACTGGCAG CTGCTGCCCC AGGACGGCAT CATGGCCGAG 2460
CACCTGCTGG GTCATGCCTG TGCCCTGGCT GCCTCCCTCT GGCTGGGGGT GCTGCCCACA 2520
CTCACTCTTG GCTTGCTGGT CCACGTGAAG CTTGGGCCCC AACAAAACT CATCTCAGAA 2580
GAGGATCTGA ATAGCGCCGT CGACCATCAT CATCATCATC ATTGAGTTTA TCCAGCACAG 2640
TGGCGGCCCG TCGAGTCTAG AGGGCCCGTT TAAACCGCT GATCAGCCTC GACTGTGCCT 2700
TCTAGTTGCC AGCCATCTGT TGTTTGCCCC TCCCCCGTGC CTCCTTGAC CCTGGAAGGT 2760
GCCACTCCCA CTGTCCTTTC CTAATAAAAT GAGGAAATTG CATCGCATTG TCTGAGTAGG 2820
TGTCATTCTA TTCTGGGGGG TGGGGTGGGG CAGGACAGCA AGGGGGAGGA TTGGGAAGAC 2880
AATAGCAGGC ATGCTGGGGA TGCGGTGGGC TCTATGGCTT CTGAGGCGGA AAGAACCAGC 2940
TGGGGCTCTA GGGGGTATCC CCACGCGCCC TGTAGCGGCG CATTAGCGC GCGGGGTGTG 3000
GTGGTTACGC GCAGCGTGAC CGCTACACTT GCCAGCGCCC TAGCGCCCGC TCCTTTCGCT 3060
TTCTTCCCTT CCTTCTCGC CACGTTCCGC GGCTTCCCC GTCAAGCTCT AAATCGGGGC 3120
ATCCCTTTAG GGTTCGATT TAGTGCTTTA CGGCACCTCG ACCCCAAAAA ACTTGATTAG 3180
GGTGATGGTT CACGTAGTGG GCCATCGCCC TGATAGACGG TTTTTCGCCC TTTGACGTTG 3240
GAGTCCACGT TCTTTAATAG TGGACTCTTG TTCCAAACTG GAACAACACT CAACCCTATC 3300
TCGGTCTATT CTTTTGATT ATAAGGGATT TTGGGGATT CGGCCTATTG GTTAAAAAT 3360



5

GAGCTGATTT	AACAAAAATT	TAACGCGAAT	TAATTCTGTG	GAATGTGTGT	CAGTTAGGGT	3420
GTGGAAAGTC	CCCAGGCTCC	CCAGGCAGGC	AGAAGTATGC	AAAGCATGCA	TCTCAATTAG	3480
TCAGCAACCA	GGTGTGGAAA	GTCCCCAGGC	TCCCCAGCAG	GCAGAAGTAT	GCAAAGCATG	3540
CATCTCAATT	AGTCAGCAAC	CATAGTCCCG	CCCCTAACTC	CGCCCATCCC	GCCCCTAACT	3600
CCGCCAGTT	CCGCCATTC	TCCGCCCAT	GGCTGACTAA	TTTTTTTTAT	TTATGCAGAG	3660
GCCGAGGCCG	CCTCTGCCTC	TGAGCTATTC	CAGAAGTAGT	GAGGAGGCTT	TTTTGGAGGC	3720
CTAGGCTTTT	GCAAAAAGCT	CCCGGGAGCT	TGTATATCCA	TTTCGGATC	TGATCAAGAG	3780
ACAGGATGAG	GATCGTTTCG	CATGATTGAA	CAAGATGGAT	TGCACGCAGG	TTCTCCGGCC	3840
GCTTGGGTGG	AGAGGCTATT	CGGCTATGAC	TGGGCACAAC	AGACAATCGG	CTGCTCTGAT	3900
GCCGCCGTGT	TCCGGCTGTC	AGCGCAGGGG	CGCCCGGTTT	TTTTGTCAA	GACCGACCTG	3960
TCCGGTGCCC	TGAATGAACT	GCAGGACGAG	GCAGCGCGGC	TATCGTGGCT	GGCCACGACG	4020
GGCGTTCCTT	GCGCAGCTGT	GCTCGACGTT	GTCACTGAAG	CGGGAAGGGA	CTGGCTGCTA	4080
TTGGGCGAAG	TGCCGGGGCA	GGATCTCCTG	TCATCTCACC	TTGCTCCTGC	CGAGAAAGTA	4140
TCCATCATGG	CTGATGCAAT	GCGGCGGCTG	CATACGCTTG	ATCCGGCTAC	CTGCCCATTC	4200
GACCACCAAG	CGAAACATCG	CATCGAGCGA	GCACGTACTC	GGATGGAAGC	CGGTCTTGTC	4260
GATCAGGATG	ATCTGGACGA	AGAGCATCAG	GGGCTCGCGC	CAGCCGAACT	GTTCCGCCAGG	4320
CTCAAGGCCG	GCATGCCCGA	CGGCGAGGAT	CTCGTCGTGA	CCCATGGCGA	TGCCTGCTTG	4380
CCGAATATCA	TGGTGGAAAA	TGGCCGCTTT	TCTGGATTCA	TCGACTGTGG	CCGGCTGGGT	4440
GTGGCGGACC	GCTATCAGGA	CATAGCGTTG	GCTACCCGTG	ATATTGCTGA	AGAGCTTGGC	4500
GGCGAATGGG	CTGACCGCTT	CCTCGTGCTT	TACGGTATCG	CCGCTCCCGA	TTCCGAGCGC	4560
ATCGCCTTCT	ATCGCCTTCT	TGACGAGTTC	TTCTGAGCGG	GACTCTGGGG	TTCGAAATGA	4620
CCGACCAAGC	GACGCCCAAC	CTGCCATCAC	GAGATTCGA	TCCACCGCC	GCCTTCTATG	4680
AAAGTTGGG	CTTCGGAATC	GTTTTCCGGG	ACGCCGGCTG	GATGATCCTC	CAGCGCGGGG	4740
ATCTCATGCT	GGAGTTCTTC	GCCCACCCCA	ACTTGTTTAT	TGCAGCTTAT	AATGGTTACA	4800
AATAAAGCAA	TAGCATCACA	AATTCACAA	ATAAAGCATT	TTTTTCACTG	CATTCTAGTT	4860
GTGGTTTGTG	CAAACATC	AATGTATCTT	ATCATGTCTG	TATACCGTCG	ACCICTAGCT	4920
AGAGCTTGGC	GTAATCATGG	TCATAGCTGT	TTCTGTGTG	AAATTGTTAT	CCGCTCACAA	4980
TTCCACACAA	CATACGAGCC	GGAAGCATAA	AGTGTAAGC	CTGGGGTGCC	TAATGAGTGA	5040
GCTAACTCAC	ATTAATTGCG	TTGCGCTCAC	TGCCCGCTTT	CCAGTCGGGA	AACCTGTCGT	5100

111

GCCAGCTGCA TTAATGAATC GGCCAACGCG CGGGGAGAGG CGGTTTGCGT ATTGGGCGCT 5160
CTTCCGCTTC CTCGCTCACT GACTCGCTGC GCTCGGTCGT TCGGCTGCGG CGAGCGGTAT 5220
CAGCTCACTC AAAGGCGGTA ATACGGTTAT CCACAGAATC AGGGGATAAC GCAGGAAAGA 5280
5 ACATGTGAGC AAAAGGCCAG CAAAAGGCCA GGAACCGTAA AAAGGCCGCG TTGCTGGCGT 5340
TTTTCCATAG GCTCCGCCCC CCTGACGAGC ATCACAAAAA TCGACGCTCA AGTCAGAGGT 5400
GGCGAAACCC GACAGGACTA TAAAGATACC AGGCGTTTCC CCCTGGAAGC TCCCTCGTGC 5460
GCTCTCCTGT TCCGACCCTG CCGCTTACCG GATACCTGTC CGCCTTTCTC CCTTCGGGAA 5520
GCGTGGCGCT TTCTCAATGC TCACGCTGTA GGTATCTCAG TTCGGTGTAG GTCGTTGCGT 5580
CCAAGCTGGG CTGTGTGCAC GAACCCCCG TTCAGCCCGA CCGCTGCGCC TTATCCGGTA 5640
ACTATCGTCT TGAGTCCAAC CCGGTAAGAC ACGACTTATC GCCACTGGCA GCAGCCACTG 5700
GTAACAGGAT TAGCAGAGCG AGGTATGTAG GCGGTGCTAC AGAGTTCTTG AAGTGGTGGC 5760
CTAACTACGG CTACACTAGA AGGACAGTAT TTGGTATCTG CGCTCTGCTG AAGCCAGTTA 5820
CCTTCGGAAA AAGAGTTGGT AGCTCTTGAT CCGGCAAACA AACCACCGCT GGTAGCGGTG 5880
GTTTTTTTGT TTGCAAGCAG CAGATTACGC GCAGAAAAAA AGGATCTCAA GAAGATCCTT 5940
TGATCTTTTC TACGGGGTCT GACGCTCAGT GGAACGAAAA CTCACGTTAA GGGATTTTGG 6000
TCATGAGATT ATCAAAAAGG ATCTTCACCT AGATCCTTTT AAATTAAAAA TGAAGTTTAA 6060
AATCAATCTA AAGTATATAT GAGTAACTT GGTCTGACAG TTACCAATGC TTAATCAGTG 6120
AGGCACCTAT CTCAGCGATC TGTCTATTTT GTTCATCCAT AGTTGCCTGA CTCCCCGTCG 6180
TGTAGATAAC TACGATACGG GAGGGCTTAC CATCTGGCCC CAGTGCTGCA ATGATACCGC 6240
GAGACCCACG CTCACCGGCT CCAGATTTAT CAGCAATAAA CCAGCCAGCC GGAAGGGCCG 6300
AGCGCAGAAG TGGTCTGCA ACTTTATCCG CCTCCATCCA GTCTATTAAT TGTTGCCGGG 6360
AAGCTAGAGT AAGTAGTTCG CCAGTTAATA GTTTGCGCAA CGTTGTTGCC ATTGCTACAG 6420
GCATCGTGGT GTCACGCTCG TCGTTTGGTA TGGCTTCATT CAGCTCCGGT TCCCAACGAT 6480
CAAGGCGAGT TACATGATCC CCCATGTTGT GCAAAAAAGC GGTTAGCTCC TTCGGTCCTC 6540
CGATCGTTGT CAGAAGTAAG TTGGCCGCAG TGTTATCACT CATGGTTATG GCAGCACTGC 6600
ATAATTCTCT TACTGTGATG CCATCCGTAA GATGCTTTTC TGTGACTGGT GAGTACTCAA 6660
CCAAGTCATT CTGAGAATAG TGTATGCGGC GACCGAGTTG CTCTTGCCCG GCGTCAATAC 6720
GGGATAATAC CGCGCCACAT AGCAGAAGTT TAAAAGTGCT CATCATTGGA AAACGTTCTT 6780

	CGGGGCGAAA ACTCTCAAGG ATCTTACCGC TGTTGAGATC CAGTTCGATG TAACCCACTC	6840
	GTGCACCCAA CTGATCTTCA GCATCTTTA CTTTCACCAG CGTTTCTGGG TGAGCAAAAA	6900
	CAGGAAGGCA AAATGCCGCA AAAAAGGGAA TAAGGGCGAC ACGGAAATGT TGAATACTCA	6960
	TACTCTTCCT TTTTCAATAT TATTGAAGCA TTTATCAGGG TTATTGTCTC ATGAGCGGAT	7020
5	ACATATTTGA ATGTATTTAG AAAAATAAAC AAATAGGGGT TCCGCGCACA TTTCCCCGAA	7080
	AAGTGCCACC TGACGTCGAC GGATCGGG	7108

(2) SEQ ID NO: 39 的信息:

(i) 序列特征:

- (A) 长度: 4019 个碱基对
- (B) 类型: 核酸
- (C) 链型: 单链
- (D) 拓扑结构: 线型

(ii) 分子类型: DNA(基因组的)

(ix) 特征:

- (A) 名称/关键词: 外显子
- (B) 位置: 1..4019

(xi) 序列描述: SEQ ID NO: 39 :

	CTCGAGAAAT CATAAAAAAT TTATTGCTT TGTGAGCGGA TAACAATTAT AATAGATTCA	60
	ATTGTGAGCG GATAACAATT TCACACAGAA TTCATTAAAG AGGAGAAATT AACTATGAGA	120
20	GGATCGCATC ACCATCACCA TCACGGATCC CTGGTTCTGT TTGAAGGGGA CGAGGTGTAT	180
	TCCACCATCC GGAAGCAGGA ATACAATGGG AAGATCCCTC GGTCCGCCG CATCCGGGGC	240
	GAGAGTGAGC TGTACACCAG TGATACTGTC ATGCAGAACC CACAGTTCAT CAAAGCCACC	300
	ATCGTGCACC AAGACCAGGC TTACGATGAC AAGATCTACT ACTTCTTCCG AGAGGACAAT	360
	CCTGACAAGA ATCCTGAGGC TCCTCTCAAT GTGTCCCGTG TGGCCAGTT GTGCAGGGGG	420
25	GACCAGGGTG GGGAAAGTTC ACTGTCAGTC TCCAAGTGG AACTTTTCT GAAAGCCATG	480
	CTGGTATGCA GTGATGCTGC CACCAACAAG AACTTCAACA GGCTGCAAGA CGTCTTCCTG	540
	CTCCCTGACC CCAGCGGCCA GTGGAGGGAC ACCAGGGTCT ATGGTGTTTT CTCCAACCCC	600
	TGGAACTACT CAGCCGTCTG TGTGTATTCC CTCGGTGACA TTGACAAGGT CTTCCGTACC	660
	TCCTCACTCA AGGGCTACCA CTCAAGCCTT CCCAACCCGC GGCCTGGCAA GTGCCTCCCA	720
30	GACCAGCAGC CGATACCCAC AGAAAGCTTA ATTAGCTGAG CTTGGACTCC TGTTGATAGA	780

TCCAGTAATG	ACCTCAGAAC	TCCATCTGGA	TTGTTTCTAGA	ACGCTCGGTT	GCCGCCGGGC	840
GTTTTTTATT	GGTGAGAATC	CAAGCTAGCT	TGGCGAGATT	TTCAGGAGCT	AAGGAAGCTA	900
AAATGGAGAA	AAAAATCACT	GGATATACCA	CCGTTGATAT	ATCCCAATGG	CATCGTAAAG	960
AACATTTTGA	GGCATTTCAG	TCAGTTGCTC	AATGTACCTA	TAACCAGACC	GTTTACGCTGG	1020
ATATTACGGC	CTTTTTAAAG	ACCGTAAAGA	AAAATAAGCA	CAAGTTTTAT	CCGGCCTTTA	1080
TTCACATTCT	TGCCCCCCTG	ATGAATGCTC	ATCCGGAATT	TCGTATGGCA	ATGAAAGACG	1140
GTGAGCTGGT	GATATGGGAT	AGTGTTTACC	CTTGTTACAC	CGTTTTCCAT	GAGCAAACCTG	1200
AAACGTTTTT	ATCGCTCTGG	AGTGAATACC	ACGACGATTT	CCGGCAGTTT	CTACACATAT	1260
ATTCGCAAGA	TGTGGCGTGT	TACGGTGAAA	ACCTGGCCTA	TTTCCCTAAA	GGGTTTATTG	1320
AGAATATGTT	TTTCGTCTCA	GCCAATCCCT	GGGTGAGTTT	CACCAGTTTT	GATTTAAACG	1380
TGGCCAATAT	GGACAACCTC	TTCGCCCCCG	TTTTCACCAT	GGGCAAATAT	TATACGCAAG	1440
GCGACAAGGT	GCTGATGCCG	CTGGCGATTG	AGGTTTATCA	TGCCGTCTGT	GATGGCTTCC	1500
ATGTCGGCAG	AATGCTTAAT	GAATTACAAC	AGTACTGCGA	TGAGTGGCAG	GGCGGGGCGT	1560
AATTTTTTTA	AGGCAGTTAT	TGGTGCCCTT	AAACGCCTGG	GGTAATGACT	CTCTAGCTTG	1620
AGGCATCAAA	TAAAACGAAA	GGCTCAGTCG	AAAGACTGGG	CCTTTCGTTT	TATCTGTTGT	1680
TTGTCCGGTGA	ACGCTCTCCT	GAGTAGGACA	AATCCGCCGC	TCTAGAGCTG	CCTCGCGCGT	1740
TTCCGGTGATG	ACGGTGAAAA	CCTCTGACAC	ATGCAGCTCC	CGGAGACGGT	CACAGCTTGT	1800
CTGTAAGCGG	ATGCCGGGAG	CAGACAAGCC	CGTCAGGGCG	CGTCAGCGGG	TGTTGGCGGG	1860
TGTCGGGGCG	CAGCCATGAC	CCAGTCACGT	AGCGATAGCG	GAGTGTATAC	TGGCTTAACT	1920
ATGCCGCATC	AGAGCAGATT	GTAATGAGAG	TGCACCATAT	GCGGTGTGAA	ATACCGCACA	1980
GATGCGTAAG	GAGAAAATAC	CGCATCAGGC	GCTCTTCCGC	TTCCTCGCTC	ACTGACTCGC	2040
TGCGCTCGGT	CTGTCCGGCTG	CGGCGAGCGG	TATCAGCTCA	CTCAAAGGCG	GTAATACGGT	2100
TATCCACAGA	ATCAGGGGAT	AACGCAGGAA	AGAACATGTG	AGCAAAGGC	CAGCAAAGG	2160
CCAGGAACCG	TAAAAGGCC	GCGTTGCTGG	CGTTTTTCCA	TAGGCTCCGC	CCCCCTGACG	2220
AGCATCACAA	AAATCGACGC	TCAAGTCAGA	GGTGGCGAAA	CCCGACAGGA	CTATAAAGAT	2280
ACCAGGCGTT	TCCCCCTGGA	AGCTCCCTCG	TGCGCTCTCC	TGTTCCGACC	CTGCCGCTTA	2340
CCGGATACCT	GTCCGCCTTT	CTCCCTTCGG	GAAGCGTGGC	GCTTTCTCAA	TGCTCACGCT	2400
GTAGGTATCT	CAGTTCGGTG	TAGGTCGTTT	GCTCCAAGCT	GGGCTGTGTG	CACGAACCCC	2460
CCGTTACGCC	CGACCGCTGC	GCCTTATCCG	GTAACATATCG	TCTTGAGTCC	AACCCGGTAA	2520

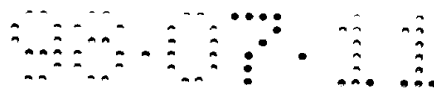
	GACACGACTT ATCGCCACTG GCAGCAGCCA CTGGTAACAG GATTAGCAGA GCGAGGTATG	2580
	TAGGCGGTGC TACAGAGTTC TTGAAGTGGT GGCCTAACTA CGGCTACACT AGAAGGACAG	2640
	TATTTGGTAT CTGCGCTCTG CTGAAGCCAG TTACCTTCGG AAAAAGAGTT GGTAGCTCTT	2700
	GATCCGGCAA ACAAACCACC GCTGGTAGCG GTGGTTTTTT TGTTTGCAAG CAGCAGATTA	2760
5	CGCGCAGAAA AAAAGGATCT CAAGAAGATC CTTTGATCTT TTCTACGGGG TCTGACGCTC	2820
	AGTGGAAACGA AAACTCACGT TAAGGGATTT TGGTCATGAG ATTATCAAAA AGGATCTTCA	2880
	CCTAGATCCT TTAAATTAA AATGAAGTT TAAATCAAT CTAAAGTATA TATGAGTAAA	2940
	CTTGGTCTGA CAGTTACCAA TGCTTAATCA GTGAGGCACC TATCTCAGCG ATCTGTCTAT	3000
10	TTCGTTTCATC CATAGCTGCC TGACTCCCCG TCGTGTAGAT AACTACGATA CGGGAGGGCT	3060
	TACCATCTGG CCCCAGTGCT GCAATGATAC CGCGAGACCC ACGCTCACCG GCTCCAGATT	3120
	TATCAGCAAT AAACCAGCCA GCCGGAAGGG CCGAGCGCAG AAGTGGTCCT GCAACTTTAT	3180
	CCGCCTCCAT CCAGTCTATT AATGTTGCC GGAAGCTAG AGTAAGTAGT TCGCCAGTTA	3240
	ATAGTTTGCG CAACGTTGTT GCCATTGCTA CAGGCATCGT GGTGTCACGC TCGTCGTTG	3300
15	GTATGGCTTC ATTCAGCTCC GGTCCCAAC GATCAAGGCG AGTTACATGA TCCCCCATGT	3360
	TGTGCAAAAA AGCGGTTAGC TCCTTCGGTC CTCGGATCGT TGTGAGAAGT AAGTTGGCCG	3420
	CAGTGTTATC ACTCATGGTT ATGGCAGCAC TGCATAATTC TCTTACTGTC ATGCCATCCG	3480
	TAAGATGCTT TTCTGTGACT GGTGAGTACT CAACCAAGTC ATTCTGAGAA TAGTGTATGC	3540
	GGCGACCGAG TTGCTCTTGC CCGGCGTCAA TACGGGATAA TACCGCGCCA CATAGCAGAA	3600
20	CTTTAAAAGT GCTCATCATT GGAAAACGTT CTTCGGGGCG AAAACTCTCA AGGATCTTAC	3660
	CGCTGTTGAG ATCCAGTTCG ATGTAACCCA CTCGTGCACC CAACTGATCT TCAGCATCTT	3720
	TTACTTTCAC CAGCGTTTCT GGGTGAGCAA AAACAGGAAG GCAAAATGCC GCAAAAAAGG	3780
	GAATAAGGGC GACACGAAA TGTTGAATAC TCATACTCTT CCTTTTTCAA TATTATTGAA	3840
	GCATTTATCA GGGTTATTGT CTCATGAGCG GATACATATT TGAATGTATT TAGAAAAATA	3900
25	AACAAATAGG GGTCCGCGC ACATTTCCCC GAAAAGTGCC ACCTGACGTC TAAGAAACCA	3960
	TTATTATCAT GACATTAACC TATAAAAATA GCGGTATCAC GAGGCCCTTT CGTCTTCAC	4019

(2) SEQ ID NO: 40 的信息:

(i) 序列特征:

(A) 长度: 3999 个碱基对

(B) 类型: 核酸



(C) 链型: 单链

(D) 拓扑结构: 线型

(ii) 分子类型: DNA(基因组的)

(ix) 特征:

(A) 名称/关键词: 外显子

(B) 位置: 1..3999

(xi) 序列描述: SEQ ID NO: 40 :

5 CTCGAGAAAT CATAAAAAAT TTATTTGCTT TGTGAGCGGA TAACAATTAT AATAGATTCA 60
 ATTGTGAGCG GATAACAATT TCACACAGAA TTCATTAAAG AGGAGAAATT AACTATGAGA 120
10 GGATCGCATC ACCATCACCA TCACACGGAT CCGCATGCCA GCTCCCAGTG GGAGGTGAGC 180
 CAGGTGCCCC TGGACCTGTG TGAGGTCTAT GGCGGGGGCT GCCACGGTTG CCTCATGTCC 240
 CGAGACCCCT ACTGCGGCTG GGACCAGGGC CGCTGCATCT CCATCTACAG CTCCGAACGG 300
 TCAGTGCTGC AATCCATTAA TCCAGCCGAG CCACACAAGG AGTGTCCCAA CCCCAAACCA 360
 GACAAGGCC CACTGCAGAA GGTTCCTG GCCCCTACT CTCGCTACTA CCTGAGCTGC 420
15 CCCATGGAAT CCCGCCACGC CACCTACTCA TGGCGCCACA AGGAGAACGT GGAGCAGAGC 480
 TGCGAACCTG GTCACCAGAG CCCCAACTGC ATCCTGTTCA TCGAGAACCT CACGGCGCAG 540
 CAGTACGGCC ACTACTTCTG CGAGGCCCGAG GAGGGCTCCT ACTTCCGCGA GGCTCAGCAC 600
 TGCCAGCTGC TGCCCGAGGA CGGCATCATG GCCGAGCACC TGCTGGGTCA TGCCTGTGCC 660
 CTGGCTGCCT CCCTCTGGCT GGGGGTGCTG CCCACACTCA CTCTTGGCTT GCTGGTCCAC 720
20 GTGAAGCTTA ATTAGCTGAG CTTGGACTCC TGTTGATAGA TCCAGTAATG ACCTCAGAAC 780
 TCCATCTGGA TTTGTTTCTG ACGCTCGGTT GCCGCCGGGC GTTTTTTATT GGTGAGAATC 840
 CAAGCTAGCT TGGCGAGATT TTCAGGAGCT AAGGAAGCTA AAATGGAGAA AAAAATCACT 900
 GGATATACCA CCGTTGATAT ATCCCAATGG CATCGTAAAG AACATTTTGA GGCATTTTCA 960
 TCAGTTGCTC AATGTACCTA TAACCAGACC GTTCAGCTGG ATATTACGGC CTTTTTAAAG 1020
25 ACCGTAAAGA AAAATAAGCA CAAGTTTTAT CCGGCCTTTA TTCACATTCT TGCCCGCCTG 1080
 ATGAATGCTC ATCCGGAATT TCGTATGGCA ATGAAAGACG GTGAGCTGGT GATATGGGAT 1140
 AGTGTTCACC CTTGTTACAC CGTTTTCCAT GAGCAAACCTG AAACGTTTTT ATCGCTCTGG 1200
 AGTGAATACC ACGACGATTT CCGGCAGTTT CTACACATAT ATTCGCAAGA TGTGGCGTGT 1260
 TACGGTGAAA ACCTGGCCTA TTTCCCTAAA GGGTTTATTG AGAATATGTT TTTCTGTTCA 1320
30 GCCAATCCCT GGGTGAGTTT CACCAGTTTT GATTTAAACG TGGCCAATAT GGACAACTTC 1380

5001

5

TTCGCCCCCG TTTTCACCAT GGGCAAATAT TATACGCAAG GCGACAAGGT GCTGATGCCG 1440
CTGGCGATTG AGGTTTCATCA TGCCGTCTGT GATGGCTTCC ATGTCGGCAG AATGCTTAAT 1500
GAATTACAAC AGTACTGCGA TGAGTGGCAG GCGGGGGCGT AATTTTTTTA AGGCAGTTAT 1560
TGGTGCCCTT AAACGCCTGG GGTAATGACT CTCTAGCTTG AGGCATCAAA TAAAACGAAA 1620
GGCTCAGTCG AAAGACTGGG CCTTTCGTTT TATCTGTTGT TTGTCGGTGA ACGCTCTCCT 1680
GAGTAGGACA AATCCGCCGC TCTAGAGCTG CCTCGCGCGT TTCGGTGATG ACGGTGAAAA 1740
CCTCTGACAC ATGCAGCTCC CGGAGACGGT CACAGCTTGT CTGTAAGCGG ATGCCGGGAG 1800
CAGACAAGCC CGTCAGGGCG CGTCAGCGGG TGTGGCGGG TGTCGGGGCG CAGCCATGAC 1860
CCAGTCACGT AGCGATAGCG GAGTGTATAC TGGCTTAACT ATGCGGCATC AGAGCAGATT 1920
GTACTGAGAG TGCACCATAT GCGGTGTGAA ATACCGCACA GATGCGTAAG GAGAAAATAC 1980
CGCATCAGGC GCTCTCCCGC TTCCTCGCTC ACTGACTCGC TCGCCTCGGT CTGTCCGGCTG 2040
CGGCGAGCGG TATCAGCTCA CTCAAAGGCG GTAATACGGT TATCCACAGA ATCAGGGGAT 2100
AAGGCAGGAA AGAACATGTG AGCAAAGGC CAGCAAAGG CCAGGAACCG TAAAAGGCC 2160
GCGTTGCTGG CGTTTTTCCA TAGGCTCCGC CCCCTGACG AGCATCACAA AAATCGACGC 2220
TCAAGTCAGA GGTGGCGAAA CCCGACAGGA CTATAAAGAT ACCAGGCGTT TCCCCCTGGA 2280
AGCTCCCTCG TCGCCTCTCC TGTTCGGACC CTGCCGCTTA CCGGATACCT GTCCGCCTTT 2340
CTCCCTTCGG GAAGCGTGGC GCTTTCTCAA TGCTCACGCT GTAGGTATCT CAGTTCGGTG 2400
TAGGTGCTTC GCTCCAAGCT GGGCTGTGTG CACGAACCCC CCGTTCAGCC CGACCGCTGC 2460
GCCTTATCCG GTAACATCG TCTTGAGTCC AACC CGGTAA GACACGACTT ATCGCCACTG 2520
GCAGCAGCCA CTGGTAACAG GATTAGCAGA GCGAGGTATG TAGGCGGTGC TACAGAGTTC 2580
TTGAAGTGGT GGCCTAACTA CGGCTACACT AGAAGGACAG TATTTGGTAT CTGCGCTCTG 2640
CTGAAGCCAG TTACCTTCGG AAAAAGAGTT GGTAGCTCTT GATCCGGCAA ACAAACCACC 2700
GCTGGTAGCG GTGGTTTTTT TGTGTTGCAAG CAGCAGATTA CGCGCAGAAA AAAAGGATCT 2760
CAAGAAGATC CTTTGATCTT TTCTACGGGG TCTGACGCTC AGTGAACGA AAACCTCACGT 2820
TAAGGGATTT TGGTCATGAG ATTATCAAAA AGGATCTTCA CCTAGATCCT TTAAATTAA 2880
AAATGAAGTT TAAATCAAT CTAAAGTATA TATGAGTAAA CTTGGTCTGA CAGTTACCAA 2940
TGCTTAATCA GTGAGGCACC TATCTCAGCG ATCTGTCTAT TTCGTTTCATC CATAGCTGCC 3000
TGACTCCCCG TCGTGTAGAT AACTACGATA CGGGAGGGCT TACCATCTGG CCCAGTGCT 3060
GCAATGATAC CGCGAGACCC ACGCTCACCG GCTCCAGATT TATCAGCAAT AAACCAGCCA 3120



GCCGGAAGGG CCGAGCGCAG AAGTGGTCCT GCAACTTTAT CCGCCTCCAT CCAGTCTATT 3180
 AATTGTTGCC GGAAGCTAG AGTAAGTAGT TCGCCAGTTA ATAGTTGCG CAACGTTGTT 3240
 GCCATTGCTA CAGGCATCGT GGTGTCACGC TCGTCGTTTG GTATGGCTTC ATTCAGCTCC 3300
 GGTTCCCAAC GATCAAGGCG AGTTACATGA TCCCCATGT TGTGCAAAAA AGCGGTTAGC 3360
 5 TCCTTCGGTC CTCGATCGT TGTCAGAAGT AAGTTGGCCG CAGTGTATC ACTCATGGTT 3420
 ATGGCAGCAC TGCATAATC TCTTACTGTC ATGCCATCCG TAAGATGCTT TTCTGTGACT 3480
 GGTGAGTACT CAACCAAGTC ATTCTGAGAA TAGTGTATGC GCGACCGAG TGCTCTTGC 3540
 CCGGCGTCAA TACGGGATAA TACCGCGCCA CATAGCAGAA CTTTAAAAGT GCTCATCATT 3600
 GGAAAACGTT CTTCGGGGCG AAAACTCTCA AGGATCTTAC CGCTGTTGAG ATCCAGTTCG 3660
 10 ATGTAACCCA CTCGTGCACC CAACTGATCT TCAGCATCTT TACTTTTAC CAGCGTTTCT 3720
 GGGTGAGCAA AAACAGGAAG GCAAAATGCC GCAAAAAGG GAATAAGGGC GACACGGAAA 3780
 TGTTGAATAC TCATACTCTT CCTTTTTCAA TATTATTGAA GCATTTATCA GGGTTATTGT 3840
 CTCATGAGCG GATACATATT TGAATGTATT TAGAAAAATA AACAAATAGG GGTTCCGCGC 3900
 ACATTTCCCC GAAAAGTGCC ACCTGACGTC TAAGAAACCA TTATTATCAT GACATTAACC 3960
 15 TATAAAAATA GCGGTATCAC GAGGCCCTTT CGTCTTCAC 3999

(2) SEQ ID NO: 41 的信息:

(i) 序列特征:

(A) 长度: 8888 个碱基对

(B) 类型: 核酸

20 (C) 链型: 单链

(D) 拓扑结构: 线型

(ii) 分子类型: DNA(基因组的)

(ix) 特征:

(A) 名称/关键词: 外显子

25 (B) 位置: 1..8888

(xi) 序列描述: SEQ ID NO: 41 :

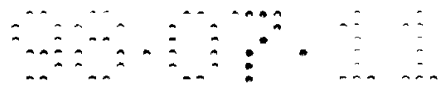
GAGCCGCACA CCGTGCTTTT CCACGAGCCA GGCAGCTCCT CTGTGTGGGT GGGAGGACGT 60
 GGCAAGGTCT ACCTCTTGA CTTCCCCGAG GGCAAGAACG CATCTGTGCG CACGGTGAGC 120
 CTCTCTCTTC CCCCAACACC CCCCTACCC TCTTATCTCC CCTCTGGCCC TGCCAAGGGT 180

30

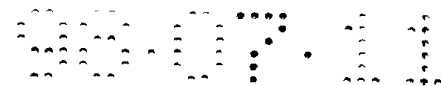


5

CCTCAGGGAA	TCCGAGGGAG	CTGGCTTCTC	TTCCTAAACT	GCCCCACCT	CCGTATCCTA	240
TAAATGGCTC	CTGGGGGAGG	CTCCCTAAAG	GTAGTCCAGA	TTGGAGTGGG	GAGCTGGGGC	300
GGTGTGGAGA	AAAACAGGAG	CTAATGGGCC	TGGCCAGCTG	GGCAGCGCTG	CTGCGGAAAG	360
CCCAGGCTGG	AAGCTGGGCC	CCAGAGCCCA	TGCCTGGTCT	TCTGAACCCT	CTGGGCCTCA	420
GCTCTGGATA	TGAGACCCTG	TTGACCTCA	GGTAGATCAC	TCACCCTCTC	AGAGCCCCAG	480
TTGCTCATCT	GTCAGATGAG	AATAATGGTT	GCTTCCTTTG	GGGCTTATCC	TGAGGCTGTG	540
TGGAAAGCAT	TTCAGGGGTA	CCTCACCCCT	GGCAGATTGA	ACTAATGCTT	CTCCCCCTCC	600
CCAGGTGAAT	ATCGGCTCCA	CAAAGGGGTC	CTGTCTGGAT	AAGCGGGTGA	GCGGGGGAGG	660
GATCTGGAGG	GGTCTGAGCC	ACTTGGTAAA	GGGAGAGGAG	ACCCTGAGGG	TCTAAGGAAG	720
GAAGCATGGC	CCTGCCCCAC	GAGTCCCAGA	CTGATGGGGA	GACGTGGTCC	TCTGTGCTTA	780
GGGGATGGCG	TCAGCTGCAC	ACACTCTGGG	CTGTCCCGGG	AGGCTGTCAC	CTATGCTAAG	840
CCCTTCTGAC	ACCTTCTTCC	CTGATCCTGG	GGGTCCTAGT	GCTAGGCTTG	CCAGGGCCTT	900
CCAGCAACCA	ATTCTCTCTC	TCCCTTCTCT	CTTCCCCGGG	CAGGACTGCG	AGAACTACAT	960
CACTCTCCTG	GAGAGGCGGA	GTGAGGGGCT	GCTGGCCTGT	GGCACCAACG	CCCGGCACCC	1020
CAGCTGCTGG	AACCTGGTGA	GAAGGCTGCT	CCCCATGTGC	CTGATCAGCT	CACCTTCTAC	1080
TGCGTGGGCT	TCTGCCCTC	ATGGTGGGAA	GGAGATGGCG	AGACTCCAAT	GCTGGCCTTG	1140
CCCTGGGAGG	ATGGGGCTCC	TGGCCGAGAA	ACTGGCCGTC	ATGGGAGGCA	GTGGCTGTGG	1200
GATTATGTGG	CCATCCAACC	CTCTGGATCT	CCCACAGGTG	AATGGCACTG	TGGTGCCACT	1260
TGGCGAGATG	AGAGGCTACG	CCCCCTTCAG	CCCGGACGAG	AACTCCCTGG	TTCTGTTTGA	1320
AGGTTGGGGC	ATGCTTCGGA	ACTGGGCTGG	GAGCAGGATG	GTCAGCTCTT	TGTCCAGTGT	1380
CCGGAGGAGG	GACTTCCAGG	AGCTGCCTGC	CCTTACTCAT	TTCTCCCTCC	CACTGACCCC	1440
AGGGGACGAG	GTGTATTCCA	CCATCCGGAA	GCAGGAATAC	AATGGGAAGA	TCCCTCGGTT	1500
CCGCCGCATC	CGGGGCGAGA	GTGAGCTGTA	CACCAGTGAT	ACTGTCATGC	AGAGTGAGTC	1560
AGGCTCCGGC	TGGGCTGAGG	GTGGGCAAGG	GGGTGTGAGC	ACTTAAGGTG	GCAGATGGGA	1620
TCCTGATGTT	TCTGGGAGGG	CTCCCTGAGG	GCCGCTGGGG	CCATGCAGGA	AAGCAGGACC	1680
TTGGTATAGG	CCTGAGAAGT	TAGGGTTGGC	TGGGAGCAGA	GGAACAGACA	AGGTATAGCA	1740
GTGGGATGGG	CCCAGCCCTC	TTCAGGAACA	CAAACAGAGG	GAGCCCCAGA	CCCAGTGCAG	1800
GGTCCCCAGG	AGCCAAAGTT	TATCCTCTGC	TGAGTTCACG	TGGAGGCAGC	CCCCCAACTC	1860
CCTCCTCATC	AGGGCTCTGC	CAATTGAGCA	GAAGTGACAT	AGGGGCCCCC	AGGGACCTTC	1920



CCCCACTCCC CAGGCATGAA GTCATTGCTC CTGGGCCGAT GACATCTTTG TAGGAAGAGG 1980
GCAAAACAGG TGTGGGGTGG AGGTGCAGGG TCTAGGGCCC CTCGGGGAGT TGGACCTGAT 2040
GTTATGAGTC CTATTCCAGA TCTGATTTGC CATGGTTTGT GCAGACCCGA AGGAGGGAGG 2100
5 AGAGTGTGCA GGGTTGGAAT GGTCTCCCGG GCAAGCTTCC CAGCCTTACG CCCATTGCTT 2160
TCTGTGCCCT GGCAGACCCA CAGTTCATCA AAGCCACCAT CGTGCACCAA GACCAGGCTT 2220
ACGATGACAA GATCTACTAC TTCTTCCGAG AGGACAATCC TGACAAGAAT CCTGAGGCTC 2280
CTCTCAATGT GTCCCGTGTG GCCCAGTTGT GCAGGGTGAA CACGGGCGTG AGGGCTGCTG 2340
GCTACGTGTC TGTGCATGAA TAGGCCTGAG TGAGGGTGAG TTCTGTGTGT CCGTGTGCAT 2400
GTAGAAGTTG TGTGGATGTA TGAGTGGGTC TGTGTCAGGG ACTGTGGGAG CAGCTGTGTG 2460
TGCATGGAGC ATCATGTGTC TGTGTGTGGG TAAAGGTGGC TGAGCTCCTG TGCACGTATG 2520
ATGGCGTGTG AGCGTGTGTA TGATGGGGTG TGTGTGTGTG TGTGTGTGTG TGTTTTGCCT 2580
GTGTGAATGT GCTGTGCCAC GTATGTGGGT GCGTGAGTCA GTAAATGTGT GTCTGAGTCC 2640
GTCTGCTCTG TGGGGACCTG GCACTCTCAC CTGCCCTGAC CCTGGGCACT GCTGGCCCTG 2700
GGCTCTGGAT CAGCCAGGCC TGCTTGCCAGG AGTCTCATCT GGAGACCTGC CCTGAGTCCCT 2760
GGGGCACCCC CGGCAGGTCC TGGCCCCTCG CAGCCTGCCT TCCTCCTCTG GGCCAGGTG 2820
TTGATATTGC TGGCAGTGGT TTCTTGGGGT GTGTGGGGAA GCCCGGGCAG GTGCTGAGGG 2880
GCCTCTTCTC CCCTCTACCC TTCCAGGGGG ACCAGGGTGG GGAAAGTTCA CTGTCAGTCT 2940
CCAAGTGGAA CACTTTTCTG AAAGCCATGC TGGTATGCAG TGATGCTGCC ACCAACAAGA 3000
ACTTCAACAG GCTGCAAGAC GTCTTCTGTC TCCCTGACCC CAGCGGCCAG TGGAGGGACA 3060
CCAGGGTCTA TGGTGTTTTC TCCAACCCT GGTGAGTGGC CCTTGTCTCTG GGGCCGGGGC 3120
TGGCATTGGT TCAGTGTCCA GTAGGGACAG GAGGCCTTGG GCCCTGCTGA GGGCCTCCCT 3180
GGTGTGGCAG GAGCAGGGGC TGCAGGCTCA AGAGGCTGGG CTGTTGCTGG GTGTGGGGTG 3240
GGGGGACAGC CAGTGCATG TATGTACTGT TGTGTGAGTG AGTCTGCACT CATGGGTGTG 3300
TGTGCATGCC CTATATGCAC ACTCATGACT GCACTTGTGC CTGTGTGTCC CACCACCTGC 3360
TTGTGCCGAG AGTGGACACT GGGCCCAGGA GGAAGCTGCT GAAGCATCTC TCGGGGAGCT 3420
GGGTGCTATT ACACCTGCTC AGGCACTGCC TGAGCCCGAT AATTCACACT TCTTAATCAC 3480
TCTCATTGAT TGAACACAG GCAGGCGGAA GTGTTGGGTG TGTGTGGGGA GAGTTAGGGA 3540
TAGAGTGGAG GAAGCCAAGA CCCTGCTCTG TGGCTCCTGG GTGAGTGGGT CCCCCAGGCT 3600

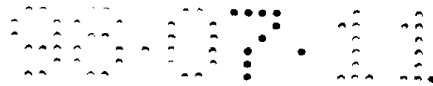


GGGAAGGGGT TGGGGTCTG GCCTCCTGGG GCATCAGCAC CCCACAGCCT GTGCCCAGGG 3660
AGGGCTAGAG AACTGCTCAG CCTATGATGG GGTTCCTCCT GCCTTGGGGT TGGGTAGAGC 3720
AGATGGCCTC TAGACTCAGT GATTCTGTAA CAGGATACAA GTTTGTGGTT TTA AATTGCA 3780
GCACAAAGAA ATTAGGCTGA ACTCCTCTCC TTCCTCCTCT CCATCCCTCC CCATTTTCAG 3840
TGGTGGTTGG CAACTCAGTG CCAGGCACAA GGCTGGCCTG GGTGAGTGA GGTGGATGGG 3900
TGGGTTCTGG GCCCCCATT GAGCTGGTCT CCATGTCACT GCAGGAACTA CTCAGCCGTC 3960
TGTGTGTATT CCCTCGGTGA CATTGACAAG GTCTTCCGTA CCTCCTCACT CAAGGGCTAC 4020
CACTCAAGCC TTCCCAACCC GCGGCCTGGC AAGGTGAGCG TGACACCAGC CGTGGCCCAG 4080
GCCCAGCCCT CTTTCTGCCT CACCTCCCAC CACCCCCTG ACCTGGGCCT GCTCTCCTTG 4140
CCCAGTGCCT CCCAGACCAG CAGCCGATAC CCACAGAGAC CTTCCAGGTG GCTGACCGTC 4200
ACCCAGAGGT GGCAGCAGAG GTGGAGCCCA TGGGGCCTCT GAAGACGCCA TTGTTCCACT 4260
CTAAATACCA CTACCAGAAA GTGGCCGTCC ACCGCATGCA AGCCAGCCAC GGGGAGACCT 4320
TTCATGTGCT TTACCTAACT ACAGGTGAGA GGCTACCCCG GGACCCTCAG TTTGCTTTGT 4380
AAAAACGGGC ATGAAAGGTG TAAGGAATAA TGTAAGTTAAC ATCTGGTTGG ATCTTTACAT 4440
GTGGAAGGAA TAATTGAGTG ACTGGAGTTG TCAGGGGTTA ATGTGTGTGG GTGTGGAAGA 4500
GCCAGGCAGG GAGAGCTTCC TGGAGGAGGT AGGGGCAAGA GGGAAAGGGG GATGGGAGAA 4560
AAGCAAGCAC TGGGATTTGG AGGCGGAAAT CTGGAGAGTC TGAGCAAAGC CAGGTGCACC 4620
TTTGGTCCAG ATGTCTGACT CAGGGAAGAA GATGGTAGGA AGAGACGTGG CAAATGAGGA 4680
GGAGGGGCCT GAACCACAGG GATACTGGCC TCTGCCAGGC AGAATGAGGG AGTCAGGCCC 4740
TGCGCCTGTC TTTGGGATTG TGCAGGTGAG AAGAAACATT TGAGGAGTTG ATGGGGCACA 4800
AATTAGGTAT GGGGAAGGAG TTCCAGGGGG CAGAACCTTT GCCATCTCAC AGAGGACAGG 4860
GGCAGCTTCT CTTCTTCCCT GGAGTAGGCC CTGCTGGGGG AAGCTGGGTG GAATGCCGTG 4920
GGAGATGCTC CTGCTTTCTG GAAAGCCACA GGACACGGAG GAGCCAGTCC TGAGTTGGGT 4980
TTGTGCGAGC TTCCCATGCC AGCTGCCTTC CTTGAGACTG GAAAGGGCCT CTAGCACCCC 5040
TGGGGCCATT CAATTCAGGC CCAGGCGCC AACCTCAGTT GTTCACATTC CCCATGTGAT 5100
CTCCTGTTGC TGCTTCACCT TGGGACTGTC TCGGCTTTGG TGACCTTGTA GGAAACTGGA 5160
ACCCAGCAC CATGTTTGG CTCCTGGAAG CTTGGGGAG AGGAATTTCC CACAGGGCAG 5220
GGCCTGGGTC CTGATTCCT GCCTCTTTAC TCCCTATTCA TCCCGGCTAC ACCCTTGGGC 5280
CCCCATCCTT GCTTGGCTCC AGTACTGGCT GGCACAGCTG TTGTGGTCAT CCAGGGATGG 5340

5

CAGGGCACTG	GGGAACAGAA	GAGAGAGGTC	ACACAGTGCG	GAACTGGGAG	CAGGAGCTAG	5400
GACAAGGAAG	GCTGGACTTG	GGCCATGGAT	TCCCTTCCTG	CAGACTTGGG	AAGTGAGCAC	5460
ACTTGAGTGA	TTAGAGAAGG	TGTCTTCGTT	CTAAGGGCAG	TGGAGGAGGC	ACCATTTTGG	5520
AGCCTGCATC	ATTCGTATTT	GGGCTAGATT	GAAAAATAGA	GCTTTCTAAG	TCCTCTGCAG	5580
AGAATGGGAG	GCTCTCACAA	CTGGGAGAAG	TATTGGCTCT	TTTCTGAGA	ATTTTGCCAA	5640
GGGTATGCTG	TTACTGGGGC	TGGTTTGAA	GGAGTATAGG	GCATTATGTC	TGTGAAGGCA	5700
GTGGCTGGGG	TGGGGCCTTA	TCAGGCCCAA	GGAGCATCTG	GCCACATCTC	AGAGTCCACA	5760
GATGAGGATC	ACGGATGTGT	AGAGGAAACA	TCCTAGGCAG	GCAATCATCT	GACTGCTTTT	5820
TTGGGGCAGG	TGATGCCCTG	GGAAATTGGG	AGGGAGGGAG	AGAGGGAGGT	AGGCTATTCT	5880
AGAAACTGGG	AGAGCAGGTG	AGGTAGGATT	GGGAGGACCA	GGGGTCAGGG	TCCCCATTGG	5940
TCCCTAATTG	AGAACGGAGA	GAGCATGGT	CTAGGAGGCA	GGCAGCTCGG	TTATAAGACC	6000
TTGGGAACTC	TTGATTTAGA	ATCCAAGATC	CTTTTTAGAT	CTAGGATTTT	ATAAAATTAA	6060
GATATCCCCT	AAGATCAAAT	GCAACGTGGA	GTCCTGAATT	GGATCCTAGA	ACAGAAGAAG	6120
GACATTTGTG	GAAAAACTAG	TGAAATCCAA	ATAAAGTCTG	TAGTTTGTGTT	AATAGTAATG	6180
CACCAATGTC	AGTTGCCTAG	TTGTGACAAA	TATACCGTGG	TTATGTAAGA	TGGTAACATT	6240
AGGGGGAACT	GGAGAAGGGT	AGATTGGAGC	TCTCTGTACT	ATCTTTGCAA	CTTTTCTGGG	6300
AATCTAAAAT	TACTCCAAAA	TAAAAAATA	ATGTATTTAA	AGTAAATATA	TTCCCTAAGA	6360
GTCCAGGAGG	CAGGGGAGTT	GTAGAAGCAG	CTGAGTGGTT	GGGTTCTGAC	AGATTTGGTT	6420
CCAACTCGGT	CTCTGCTGCT	CACCAGCTGT	GTGACCTGA	GCAAGTGGCT	TAGCCTTTCT	6480
GAGCCTGATT	TCCTTATCTG	TGGAGTGGGG	AAGATGACAG	CCACCTCGCA	GGGCTGTGGA	6540
GGGTAAACG	AGGTGATGCA	TGGACAGCAG	CCGCACTGAC	CTTGCTGGTG	TGGGGCTCCT	6600
GCTTCTGTTC	TTCCCGTGCA	GCCTTGGGAA	TGTTGGAGGC	CGTATCCAGG	GACCCCTGGG	6660
CCTCCTGGGA	TGGCCTCTCT	GGATCAGCCT	TGGAAGGTTT	CAGGCTGCCC	TTAGGCTCCC	6720
ACATTCTTCC	CCAGTCACGC	TCTCCTCGCC	CTGCCACAC	CAGTCCTGTG	ACCCTTGCCT	6780
GAGTTGTGAC	TTCCCACCCC	TCCCCGGCCT	AGAGGAAAGC	TGCCTGGCCC	CTCAGTGGGA	6840
CTCCCGCCCA	CTGACCCTCT	GTCCACCATA	CACAGACAGG	GGCACTATCC	ACAAGGTGGT	6900
GGAACCGGGG	GAGCAGGAGC	ACAGCTTCGC	CTTCAACATC	ATGGAGATCC	AGCCCTCCG	6960
CCGCGCGGCT	GCCATCCAGA	CCATGTCGCT	GGATGCTGAG	CGGGTGAGCC	TTCCCCACT	7020

	GGTCCCATG	GGCTATGCAG	TGACTGCAGC	TGAGGACAGG	GCTCCTTTGC	ATGTGATTG	7080
	TGTGTTCTTT	TAAGAGCTTC	TAGGCCTTAG	GGCCTGGACA	TTAGGACTG	AGTGTGGGGT	7140
5	GGGGCCCGGG	CCTGACCCAA	TCCTGCTGTC	CTTCCAGAGG	AAGCTGTATG	TGAGCTCCCA	7200
	GTGGGAGGTG	AGCCAGGTGC	CCCTGGACCT	GTGTGAGGTC	TATGGCGGGG	GCTGCCACGG	7260
	TTGCCTCATG	TCCCGAGACC	CCTACTGCGG	CTGGGACCAG	GGCCGCTGCA	TCTCCATCTA	7320
	CAGCTCCGAA	CGGTACGTTG	GCCGGGATCC	CTCCGTCCCT	GGGACAAGGT	GGGCATGGGA	7380
	CAGGGGGAGG	TGTTGTCGGG	CTGGAAGAGG	TGGCGGTACT	GGGCCTTTCT	TGTGGGACCT	7440
	CCTCTCTACT	GGAAGTGCAC	TAGGGGTAAG	GATATGAGGG	TCAGGTCTGC	AGCCTTGTAT	7500
	CTGCTGATCC	TCTTTCGTCC	TTCCCACTCC	AGGTCAGTGC	TGCAATCCAT	TAATCCAGCC	7560
	GAGCCACACA	AGGAGTGTCC	CAACCCCAA	CCAGGTACCT	GATCTGGCCC	TGCTGGCGGC	7620
	TGTGGCCCAA	TGAGTGGGGT	ACTGCCCTGC	CCTGATTGTC	CTGGTCTGAG	GGAAACATGG	7680
	CCTTGTCTCTG	TGGGCCCCAG	GTACATGGGG	CAGGATACAG	TCCTGCAGAG	GGAGCCCTCT	7740



TGGTGGGATG AGCGAGACGG GAGAAAAAAG GAGGACGCTG AGGGCTGGGT TCCCCACGTT 7800
CATTCAGAAG CCTTGTCTG GGATCCCAGT CGGTGGGGAG GACACATCCT CCCCTGGGAG 7860
CTCTTTGTCC CTCCTCACGG CTGCTTCCCC ACTGCCTCCC CAGACAAGGC CCCACTGCAG 7920
5 AAGGTTTCCC TGGCCCCAAA CTCTCGCTAC TACCTGAGCT GCCCCATGGA ATCCCGCCAC 7980
GCCACCTACT CATGGCGCCA CAAGGAGAAC GTGGAGCAGA GCTGCGAACC TGGTCACCAG 8040
AGCCCCAACT GCATCCTGTT CATCGAGAAC CTCACGGCGC AGCAGTACGG CCACTACTTC 8100
TGCAGAGCCC AGGAGGGCTC TACTTCCGC GAGGCTCAGC ACTGGCAGCT GCTGCCCGAG 8160
GACGGCATCA TGGCCGAGCA CCTGCTGGGT CATGCCTGTG CCCTGGCCGC CTCCTCTGG 8220
10 CTGGGGGTGC TGCCCACT CACTCTTGGC TTGCTGGTCC ACTAGGGCCT CCCGAGGCTG 8280
GGCATGCCTC AGGCTTCTGC AGCCAGGGC ACTAGAACGT CTCACACTCA GAGCCGGCTG 8340
GCCCCGGGAGC TCCTTGCTG CCACTTCTTC CAGGGGACAG AATAACCCAG TGGAGGATGC 8400
CAGGCCTGGA GACGTCCAGC CGCAGGCGGC TGCTGGGCCC CAGGTGGCGC ACGGATGGTG 8460
AGGGGCTGAG AATGAGGGCA CCGACTGTGA AGCTGGGGCA TCGATGACCC AAGACTTTAT 8520
15 CTTCTGGAAA ATATTTTCA GACTCCTCAA ACTTGACTAA ATGCAGCGAT GCTCCCAGCC 8580
CAAGAGCCCA TGGGTCGGGG AGTGGGTTTG GATAGGAGAG CTGGGACTCC ATCTCGACCC 8640
TGGGGCTGAG GCCTGAGTCC TTCTGGACTC TTGGTACCCA CATTGCCTCC TTCCCCTCCC 8700
TCTCTCATGG CTGGGTGGCT GGTGTTCCCTG AAGACCCAGG GCTACCCTCT GTCCAGCCCT 8760
GTCCTCTGCA GCTCCCTCTC TGGTCCTGGG TCCCACAGGA CAGCCGCCTT GCATGTTTAT 8820
20 TGAAGGATGT TTGCTTCCG GACGGAAGGA CGGAAAAAGC TCTGAAAAA AAAAAAAA 8880
AAAAAAA 8888

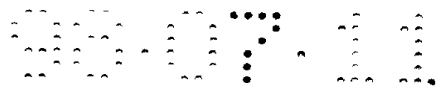
(2) SEQ ID NO: 42 的信息:

(i) 序列特征:

- 25 (A) 长度: 6622 个碱基对
(B) 类型: 核酸
(C) 链型: 单链
(D) 拓扑结构: 线型

(ii) 分子类型: 基因组 DNA

30 (ix) 特征:



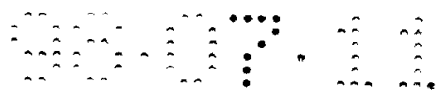
(A) 名称/关键词: 外显子

(B) 位置: 1..6622

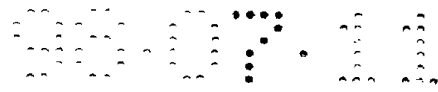
(xi) 序列描述: SEQ ID NO: 42 :

5

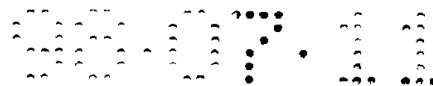
GATATCATGG	AGATAATTAA	AATGATAACC	ATCTCGCAA	TAAATAAGTA	TTTACTGTT	60
TTCGTAACAG	TTTGTAAATA	AAAAAACCTA	TAAATATGAA	ATTCTTAGTC	AACGTTGCCC	120
TTGTTTTTAT	GGTCGTATAC	ATTTCTTACA	TCTATGCGGA	TCGATGGGGA	TCCGCCCAGG	180
GCCACCTAAG	GAGCGGACCC	CGCATCTTCG	CCGTCTGGAA	AGGCCATGTA	GGGCAGGACC	240
GGGTGGACTT	TGGCCAGACT	GAGCCGCACA	CGGTGCTTTT	CCACGAGCCA	GGCAGCTCCT	300
CTGTGTGGGT	GGGAGGACGT	GGCAAGGTCT	ACCTCTTTGA	CTTCCCCGAG	GGCAAGAACG	360
CATCTGTGCG	CACGGTGAAT	ATCGGCTCCA	CAAAGGGGTC	CTGTCTGGAT	AAGCGGGACT	420
GCGAGAACTA	CATCACTCTC	CTGGAGAGGC	GGAGTGAGGG	GCTGCTGGCC	TGTGGCACCA	480
ACGCCCCGCA	CCCCAGCTGC	TGGAACCTGG	TGAATGGCAC	TGTGGTGCCA	CTTGGCGAGA	540
TGAGAGGCTA	TGCCCCCTTC	AGCCCCGACG	AGAACTCCCT	GTTTCTGTTT	GAAGGGGACG	600
AGGTGTATTC	CACCATCCGG	AAGCAGGAAT	ACAATGGGAA	GATCCCTCGG	TTCCGCCGCA	660
TCCGGGGCGA	GAGTGAGCTG	TACACCAGTG	ATACTGTCAT	GCAGAACCCA	CAGTTCATCA	720
AAGCCACCAT	CGTGCACCAA	GACCAGGCTT	ACGATGACAA	GATCTACTAC	TTCTTCCGAG	780
AGGACAATCC	TGACAAGAAT	CCTGAGGCTC	CTCTCAATGT	GTCCCGTGTG	GCCCAGTTGT	840
GCAGGGGGGA	CCAGGGTGGG	GAAAGTTCAC	TGTCAGTCTC	CAAGTGGAAC	ACTTTTCTGA	900
AAGCCATGCT	GGTATGCAGT	GATGCTGCCA	CCAACAAGAA	CTTCAACAGG	CTGCAAGACG	960
TCTTCTGCT	CCCTGACCCC	AGCGGCCAGT	GGAGGGACAC	CAGGGTCTAT	GGTGTTTTCT	1020
CCAACCCCTG	GAACTACTCA	GCCGTCTGTG	TGTATTCCCT	CGGTGACATT	GACAAGGTCT	1080
TCCGTACCTC	CTCACTCAAG	GGCTACCACT	CAAGCCTTCC	CAACCCGCGG	CCTGGCAAGT	1140
GCCTCCAGA	CCAGCAGCCG	ATACCCACAG	AGACCTTCCA	GGTGGCTGAC	CGTCACCCAG	1200
AGGTGGCGCA	GAGGGTGGAG	CCCATGGGGC	CTCTGAAGAC	GCCATTGTTC	CACTCTAAAT	1260
ACCACTACCA	GAAAGTGGCC	GTTCAACGCA	TGCAAGCCAG	CCACGGGGAG	ACCTTTCATG	1320
TGCTTTACCT	AACTACAGAC	AGGGGCACTA	TCCACAAGGT	GGTGGAACCG	GGGGAGCAGG	1380
AGCACAGCTT	CGCCTTCAAC	ATCATGGAGA	TCCAGCCCTT	CCGCCGCGCG	GCTGCCATCC	1440
AGACCATGTC	GCTGGATGCT	GAGCGGAGGA	AGCTGTATGT	GAGCTCCCAG	TGGGAGGTGA	1500
GCCAGGTGCC	CCTGGACCTG	TGTGAGGTCT	ATGGCGGGGG	CTGCCACGGT	TGCCTCATGT	1560
CCCGAGACCC	CTACTGCGGC	TGGGACCAGG	GCCGCTGCAT	CTCCATCTAC	AGCTCCGAAC	1620
GGTCAGTGCT	GCAATCCATT	AATCCAGCCG	AGCCACACAA	GGAGTGTCCT	AACCCCAAAC	1680



	CAGACAAGGC CCCACTGCAG AAGGTTTCCC TGGCCCCAAA CTCTCGCTAC TACCTGAGCT	1740
	GCCCCATGGA ATCCCGCCAC GCCACCTACT CATGGCGCCA CAAGGAGAAC GTGGAGCAGA	1800
	GCTGCGAACC TGGTCACCAG AGCCCCAACT GCATCCTGTT CATCGAGAAC CTCACGGCGC	1860
	AGCAGTACGG CCACTACTTC TCGGAGGCC AGGAGGGCTC TACTTCCGC GAGGCTCAGC	1920
5	ACTGGCAGCT GCTGCCCGAG GACGGCATCA TGGCCGAGCA CCTGCTGGGT CATGCCTGTG	1980
	CCCTGGCTGC CTGAATTCGA AGCTTGGAGT CGACTCTGCT GAAGAGGAGG AAATTCTCCT	2040
	TGAAGTTTCC CTGGTGTTC AAGTAAAGGA GTTTGCACCA GACGCACCTC TGTTCACTGG	2100
	TCCGGCGTAT TAAAACACGA TACATTGTTA TTAGTACATT TATTAAGCGC TAGATTCTGT	2160
	GCGTTGTTGA TTTACAGACA ATTGTTGTAC GTATTTTAAT AATTCATTAA ATTTATAATC	2220
	TTTAGGGTGG TATGTTAGAG CGAAAATCAA ATGATTTTCA GCGTCTTTAT ATCTGAATTT	2280
	AAATATTAAA TCCTCAATAG ATTTGTAAAA TAGGTTTCGA TTAGTTTCAA ACAAGGGTTG	2340
	TTTTTCCGAA CCGATGGCTG GACTATCTAA TGGATTTTCG CTCAACGCCA CAAACTTGC	2400
	CAAATCTTGT AGCAGCAATC TAGCTTTGTC GATATTCGTT TGTGTTTTGT TTTGTAATAA	2460
	AGGTTTCGACG TCGTTCAAAA TATTATGCGC TTTTGTATTT CTTTCATCAC TGTCGTTAGT	2520
	GTACAATTGA CTCGACGTAA ACACGTTAAA TAAAGCCTGG ACATATTTAA CATCGGGCGT	2580
	GTTAGCTTTA TTAGGCCGAT TATCGTCGTC GTCCCAACCC TCGTCGTTAG AAGTTGCTTC	2640
	CGAAGACGAT TTTGCCATAG CCACACGACG CCTATTAATT GTGTCGGCTA ACACGTCCGC	2700
	GATCAAATTT GTAGTTGAGC TTTTGGGAAT TATTTCTGAT TCGGGCGGTT TTTGGGCGGG	2760
	TTTCAATCTA ACTGTGCCCC ATTTAATTC AGACAACACG TTAGAAAGCG ATGGTGCAGG	2820
	CGGTGGTAAC ATTTCAGACG GCAAATCTAC TAATGGCGGC GGTGGTGGAG CTGATGATAA	2880
	ATCTACCATC GGTGGAGGCG CAGGCGGGGC TGGCGGCGGA GCGGAGGCG GAGGTGGTGG	2940
	CGGTGATGCA GACGGCGGTT TAGGCTCAAA TTGTCTCTTT CAGGCAACAC AGTCGGCACC	3000
	TCAACTATTG TACTGGTTTC GGGCGTATGG TGCACCTCTCA GTACAATCTG CTCTGATGCC	3060
	GCATAGTTAA GCCAGCCCCG ACACCCGCCA ACACCCGCTG ACGCGCCCTG ACGGGCTTGT	3120
	CTGCTCCCGG CATCCGCTTA CAGACAAGCT GTGACCGTCT CCGGGAGCTG CATGTGTCAG	3180
	AGGTTTTTAC CGTCATCACC GAAACGCGCG AGACGAAAGG GCCTCGTGAT ACGCCTATTT	3240
	TTATAGGTTA ATGTATGAT AATAATGGTT TCTTAGACGT CAGGTGGCAC TTTTCGGGGA	3300
	AATGTGCGCG GAACCCCTAT TTGTTTATTT TTCTAAATAC ATTCAAATAT GTATCCGCTC	3360
	ATGAGACAAT AACCCCTGATA AATGCTTCAA TAATATTGAA AAAGGAAGAG TATGAGTATT	3420
	CAACATTTCC GTGTCGCCCT TATTCCTTTT TTTGCGGCAT TTTGCCTTCC TGTTTTTGCT	3480
	CACCCAGAAA CGCTGGTGAA AGTAAAAGAT GCTGAAGATC AGTTGGGTGC ACGAGTGGGT	3540



TACATCGAAC TGGATCTCAA CAGCGGTAAG ATCCTTGAGA GTTTTCGCCC CGAAGAACGT 3600
TTTCCAATGA TGAGCACTTT TAAAGTTCTG CTATGTGGCG CGGTATTATC CCGTATTGAC 3660
GCCGGGCAAG AGCAACTCGG TCGCCGCATA CACTATTCTC AGAATGACTT GGTGAGTAC 3720
TCACCAGTCA CAGAAAAGCA TCTTACGGAT GGCATGACAG TAAGAGATT ATGCAGTGCT 3780
5 GCCATAACCA TGAGTGATAA CACTGCGGCC AACTTACTTC TGACAACGAT CGGAGGACCG 3840
AAGGAGCTAA CCGCTTTTTT GCACAACATG GGGGATCATG TAACTCGCCT TGATCGTTGG 3900
GAACCGGAGC TGAATGAAGC CATAACAAAC GACGAGCGTG ACACCACGAT GCCTGTAGCA 3960
ATGGCAACAA CGTTGCGCAA ACTATTAAC TGGCAACTAC TTACTIONTACTG TTTCCCGCAA 4020
CAATTAATAG ACTGGATGGA GCGGATAAA GTTGCAGGAC CACTTCTGCG CTCGGCCCTT 4080
CCGGCTGGCT GTTTTATTGC TGATAAATCT GGAGCCGGTG AGCGTGGGTC TCGCGGTATC 4140
ATTGCAGCAC TGGGGCCAGA TGGTAAGCCC TCCCGTATCG TAGTTATCTA CACGACGGGG 4200
AGTCAGGCAA CTATGGATGA ACGAAATAGA CAGATCGCTG AGATAGGTGC CTCACTGATT 4260
AAGCATTGGT AACTGTCAGA CCAAGTTTAC TCATATATAC TTTAGATTGA TTTAAAACCT 4320
CATTTTTAAT TTAAAAGGAT CTAGGTGAAG ATCCTTTTTG ATAATCTCAT GACCAAATC 4380
CCTTAACGTG AGTTTTCGTT CCACTGAGCG TCAGACCCCG TAGAAAAGAT CAAAGGATCT 4440
TCTTGAGATC CTTTTTTTCT GCGCGTAATC TGCTGCTTGC AAACAAAAAA ACCACCGCTA 4500
CCAGCGGTGG TTTGTTTGCC GGATCAAGAG CTACCAACTC TTTTCCGAA GGTAACGGC 4560
TTCAGCAGAG CGCAGATACC AAATACTGTT CTTCTAGTGT AGCCGTAGTT AGGCCACCAC 4620
TTCAAGAACT CTGTAGCACC GCCTACATAC CTCGCTCTGC TAATCCTGTT ACCAGTGGCT 4680
GCTGCCAGTG GCGATAAGTC GTGTCTTACC GGGTTGGACT CAAGACGATA GTTACCGGAT 4740
AAGGCGCAGC GGTGCGGCTG AACGGGGGGT TCGTGACAC AGCCAGCTT GGAGCGAACG 4800
ACCTACACCG AACTGAGATA CCTACAGCGT GAGCTATGAG AAAGCGCCAC GCTTCCGAA 4860
GGGAGAAAGG CGGACAGGTA TCCGGTAAGC GGCAGGGTCG GAACAGGAGA GCGCAGGAG 4920
GAGCTTCCAG GGGGAAACGC CTGGTATCTT TATAGTCTG TCGGGTTTCG CCACCTCTGA 4980
CTTGAGCGTC GATTTTTGTG ATGCTCGTCA GGGGGGCGGA GCCTATGGAA AAACGCCAGC 5040
AACGCGGCCT TTTTACGTT CCTGGCCTTT TGCTGGCCTT TTGCTCACAT GTTCTTTCCT 5100
GCGTTATCCC CTGATTCTGT GGATAACCGT ATTACCGCCT TTGAGTGAGC TGATACCGCT 5160
CGCCGCAGCC GAACGACCGA GCGCAGCGAG TCAGTGAGCG AGGAAGCATC CTGCACCATC 5220
GTCTGCTCAT CCATGACCTG ACCATGCAGA GGATGATGCT CGTGACGGTT AACGCCTCGA 5280
ATCAGCAACG GCTTGCCGTT CAGCAGCAGC AGACCATTTT CAATCCGCAC CTCGCGGAAA 5340
CCGACATCGC AGGCTTCTGC TTCAATCAGC GTGCCGTCCG CGGTGTGCAG TTCAACCACC 5400
GCACGATAGA GATTGCGGAT TTCGGCGCTC CACAGTTTCG GGTTTTCGAC GTTCAGACGT 5460



AGTGTGACGC GATCGGTATA ACCACCACGC TCATCGATAA TTTACCCGCC GAAAGGCGCG 5520
GTGCCGCTGG CGACCTGCGT TTCACCCTGC CATAAAGAAA CTGTTACCCG TAGGTAGTCA 5580
CGCAACTCGC CGCACATCTG AACTTCAGCC TCCAGTACAG CGCGGCTGAA ATCATCATT 5640
5 AAGCGAGTGG CAACATGGAA ATCGCTGATT TGTGTAGTCG GTTTATGCAG CAACGAGACG 5700
TCACGGAAAA TGCCGCTCAT CCGCCACATA TCCTGATCTT CCAGATAACT GCCGTCACTC 5760
CAACGCAGCA CCATCACCGC GAGGCGGTTT TCTCCGGCGC GTAAAAATGC GCTCAGGTCA 5820
AATTCAGACG GCAAACGACT GTCCTGGCCG TAACCGACCC AGCGCCCGTT GCACCACAGA 5880
TGAAACGCCG AGTTAACGCC ATCAAAAATA ATTCGCGTCT GGCCTTCCTG TAGCCAGCTT 5940
10 TCATCAACAT TAAATGTGAG CGAGTAACAA CCCGTCGGAT TCTCCGTGGG AACAAACGGC 6000
GGATTGACCG TAATGGGATA GGTACAGTTG GTGTAGATGG GCGCATCGTA ACCGTGCATC 6060
TGCCAGTTTG AGGGGACGAC GACAGTATCG GCCTCAGGAA GATCGCACTC CAGCCAGCTT 6120
TCCGGCACCG CTTCTGGTGC CGGAAACCAG GCAAAGCGCC ATTCGCCATT CAGGCTGCGC 6180
AACTGTTGGG AAGGGCGATC GGTGCGGGCC TCTTCGCTAT TACGCCAGCT GCGCAAAGGG 6240
GGATGTGCTG CAAGGCGATT AAGTTGGGTA ACGCCAGGGT TTTCCCAGTC ACGACGTTGT 6300
15 AAAACGACGG GATCTATCAT TTTTAGCAGT GATTCTAATT GCAGCTGCTC TTTGATACAA 6360
CTAATTTTAC GACGACGATG CGAGCTTTTA TTCAACCGAG CGTGCATGTT TGCAATCGTG 6420
CAAGCGTTAT CAATTTTTC AATATCGTATT GTTGCACATC AACAGGCTGG ACACCACGTT 6480
GAACTCGCCG CAGTTTTGCG GCAAGTTGGA CCCGCCGCGC ATCCAATGCA AACTTTCCGA 6540
CATTCTGTTG CCTACGAACG ATTGATTCTT TGTCCATTGA TCGAAGCGAG TGCCTTCGAC 6600
20 TTTTTCGTGT CCAGTGTGGC TT 6622

(2) SEQ ID NO: 43 的信息:

(i) 序列特征:

(A) 长度: 31 个碱基对

25 (B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑结构: 线型

(ii) 分子类型: 基因组 DNA

(ix) 特征:

30 (A) 名称/关键词: 外显子



(B) 位置: 1..31

(xi) 序列描述: SEQ ID NO: 43 :

CCGGATCCGC CCAGGGCCAC CTAAGGAGCG G 31

5 (2) SEQ ID NO: 44 的信息:

(i) 序列特征:

(A) 长度: 29 个碱基对

(B) 类型: 核酸

(C) 链型: 单链

10 (D) 拓扑结构: 线型

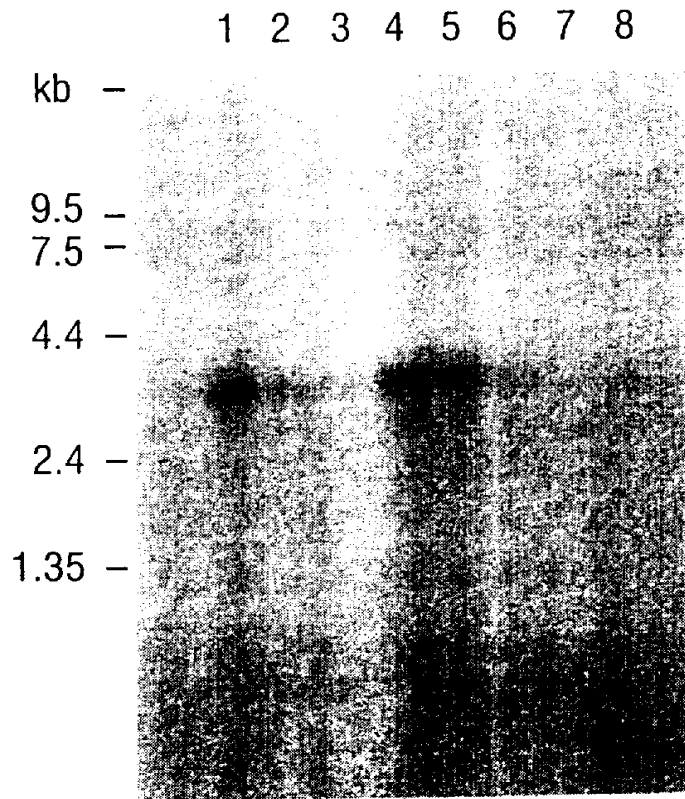
(ii) 分子类型: 基因组 DNA

(xi) 序列描述: SEQ ID NO: 44 :

CTGAATTCAG GAGCCAGGGC ACAGGCATG 29

说明书附图

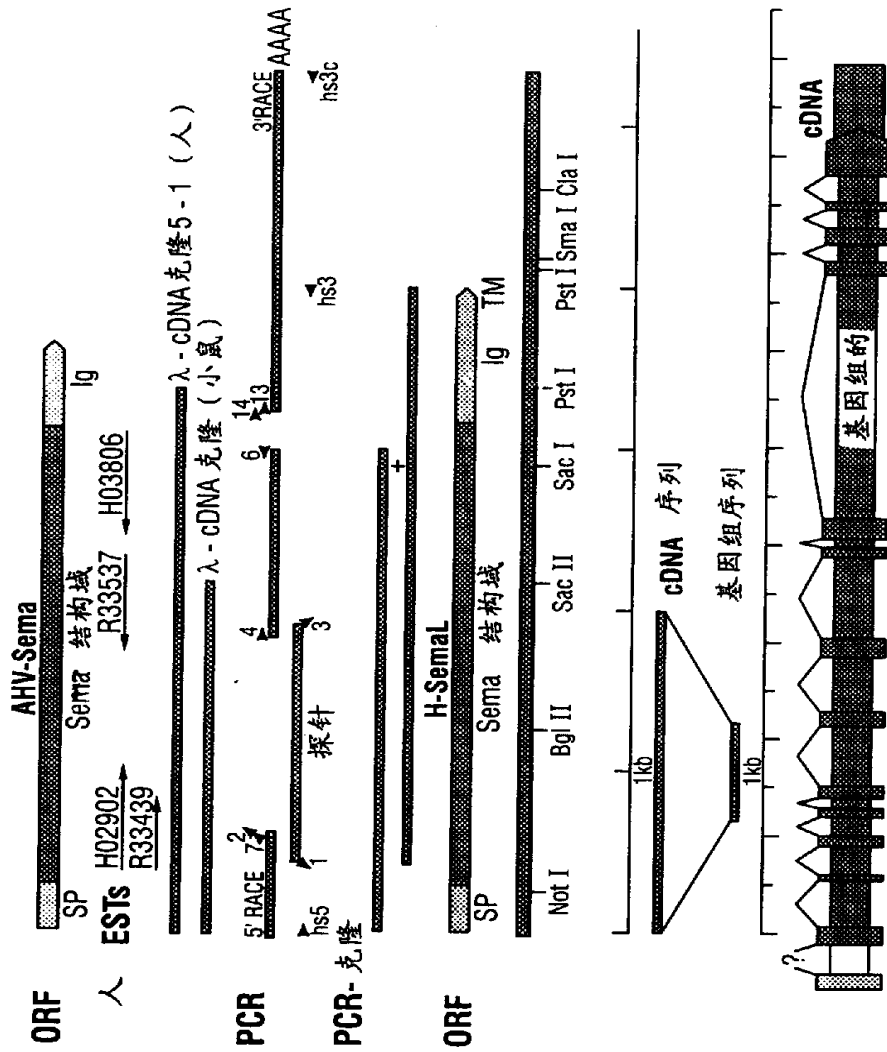
图 1



β -肌动蛋白
对照



图 2



1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100

图 4a (U937)

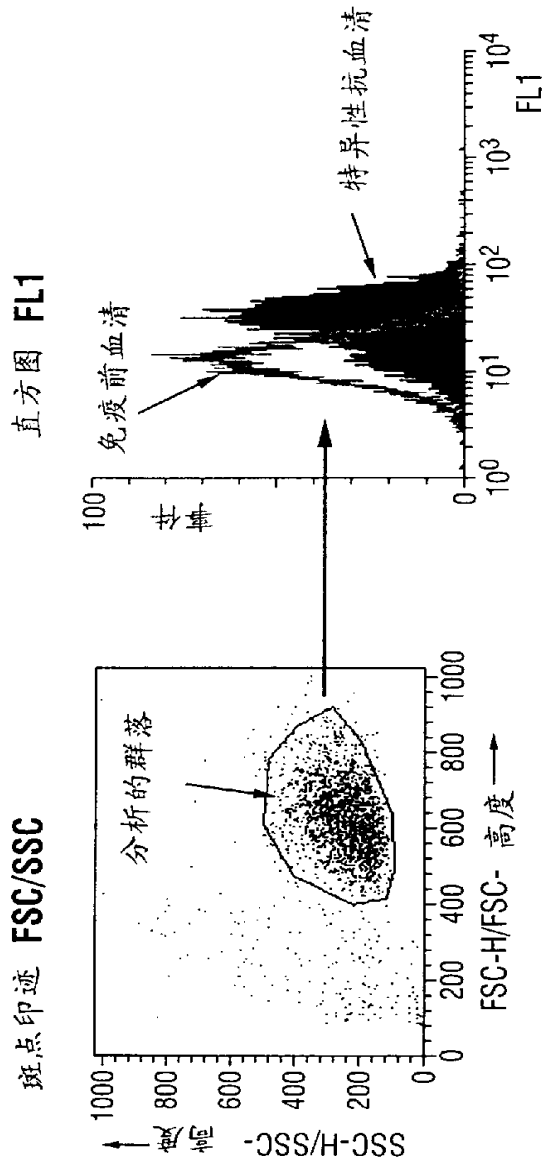


图 4b (THP-1)

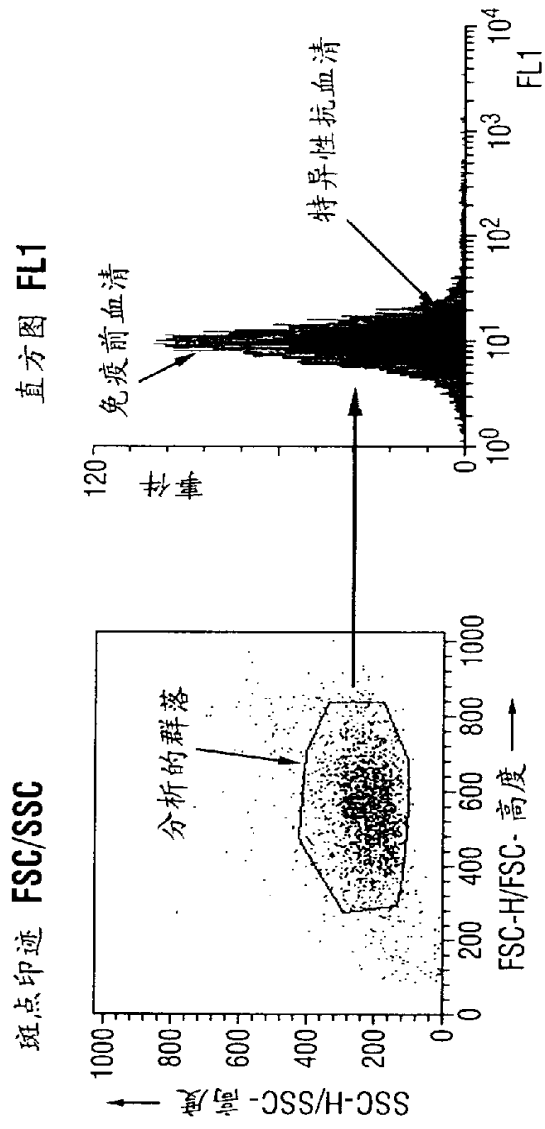


图 4C (K562)

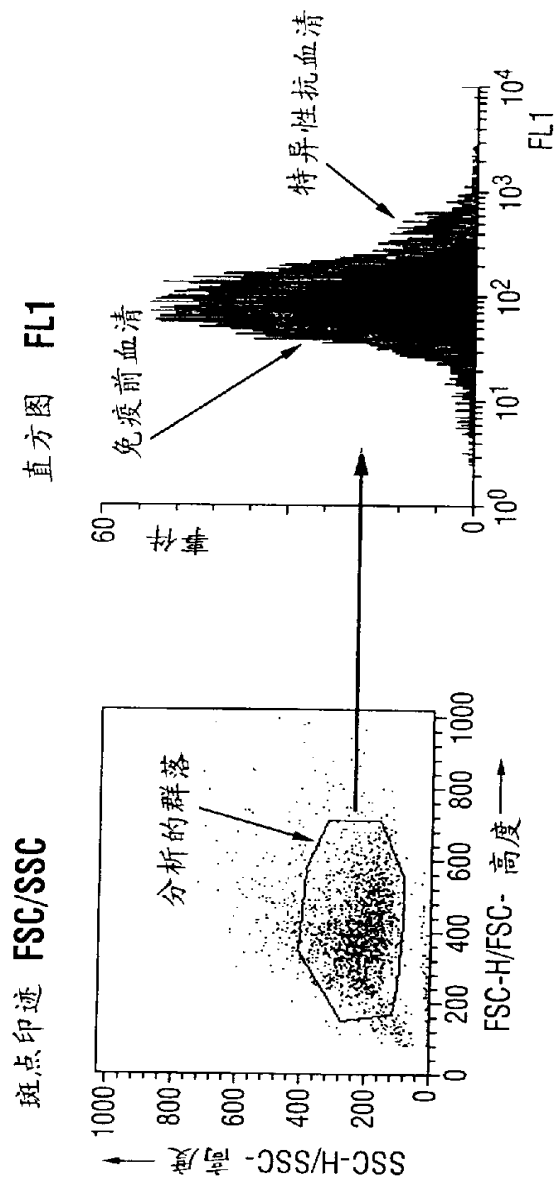


图 4d (L-428)

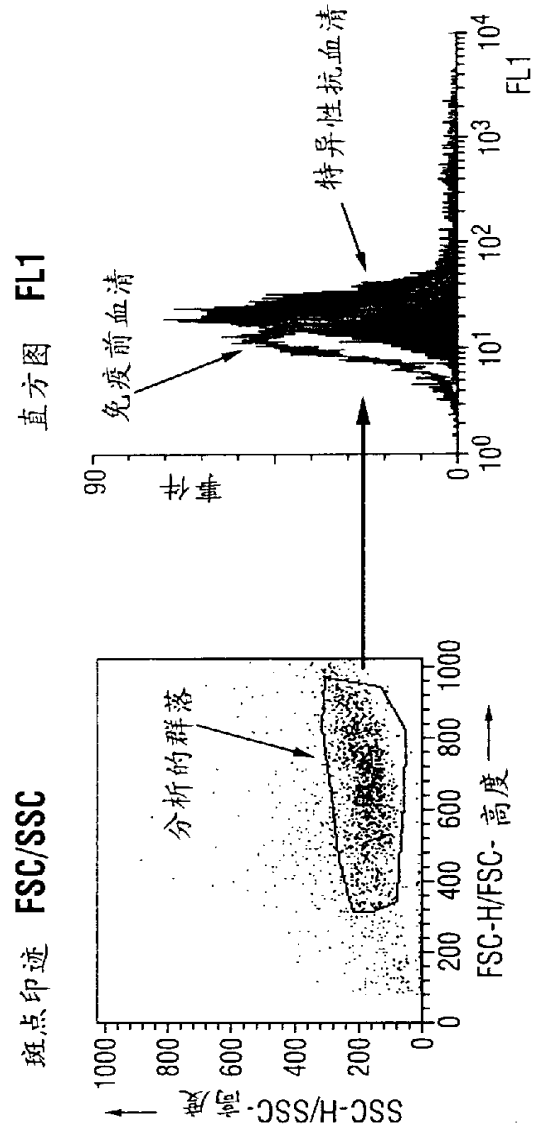


图 4e (Jurkat)

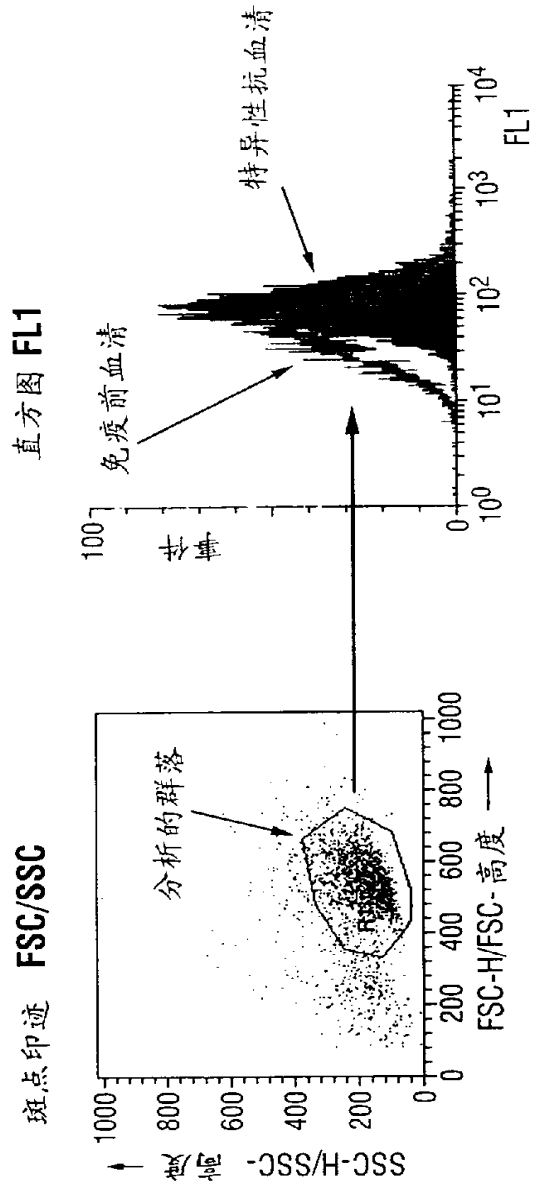


图 4f (Daudi)

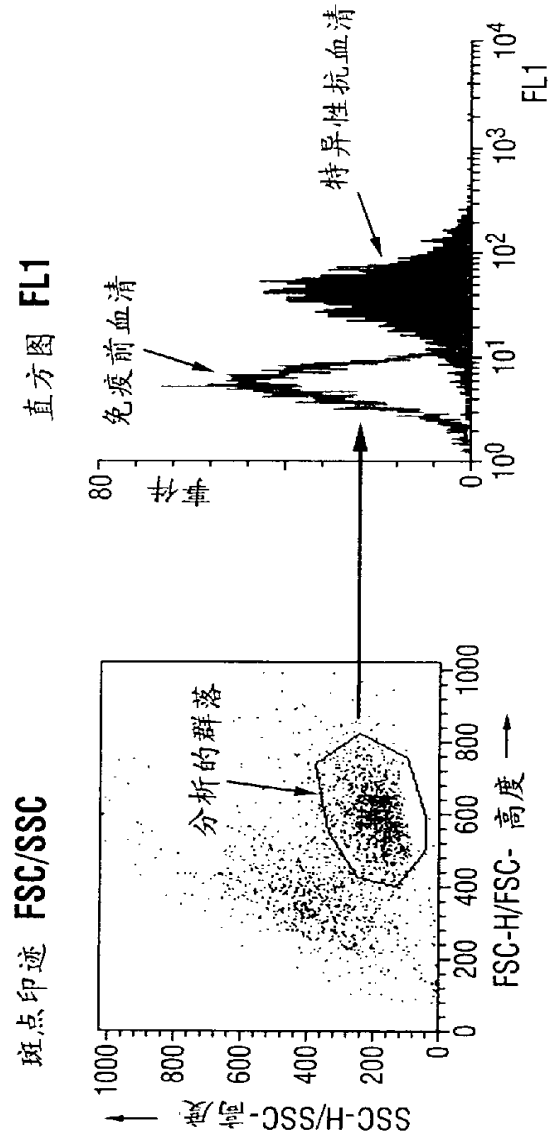


图 4g (LCL EBV-Transformed B-cells)

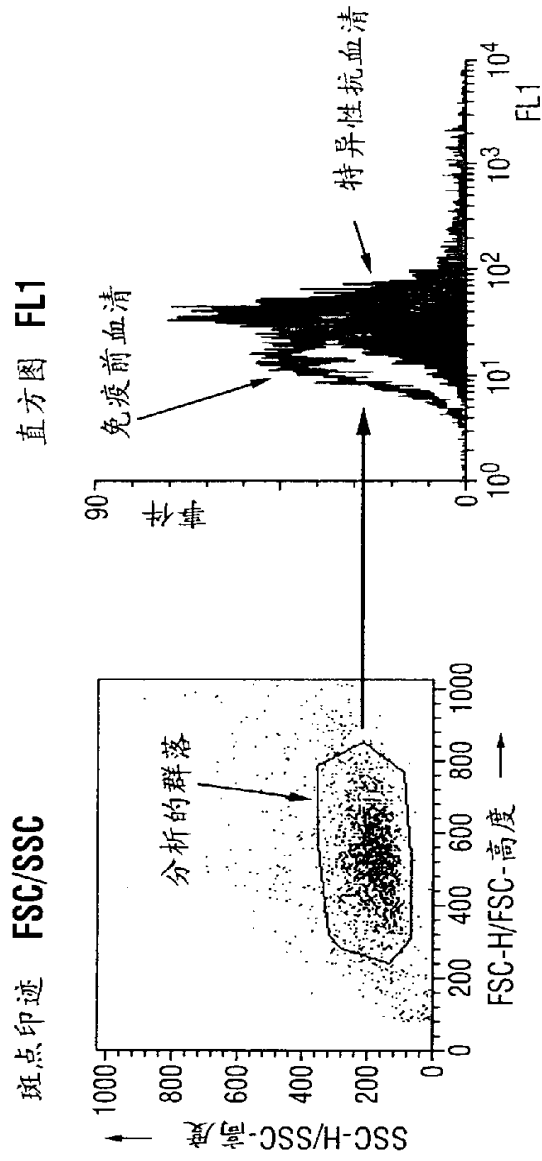


图 4h (Jiyoye)

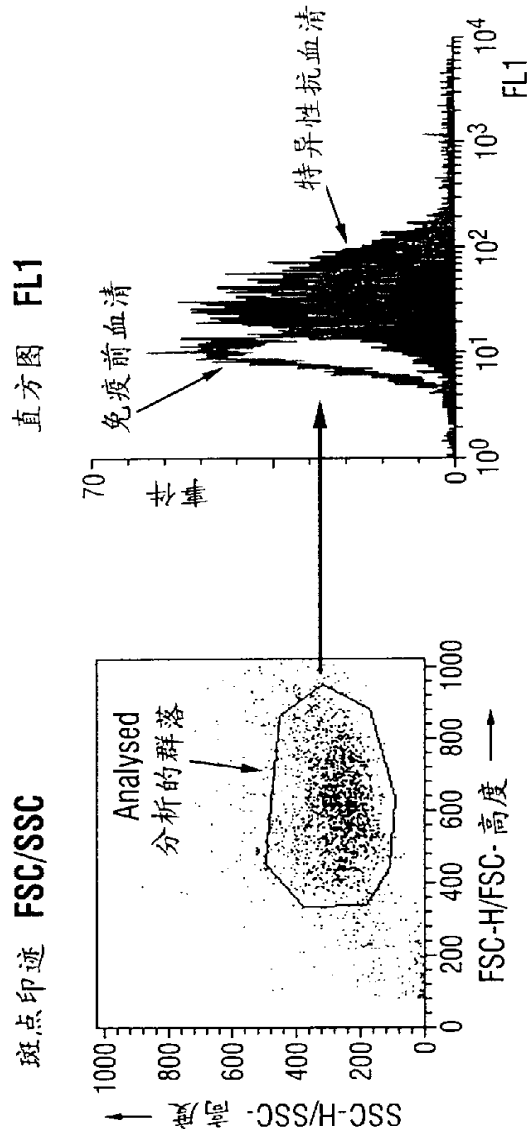


图 4 i (CBL-Mix57)

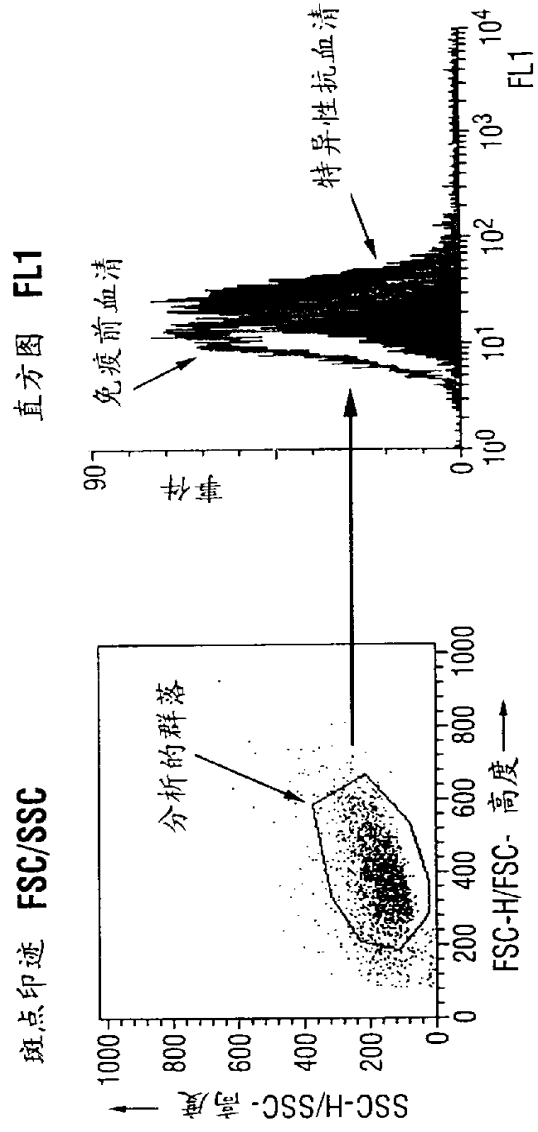


图 4k (CBL-Mix59)

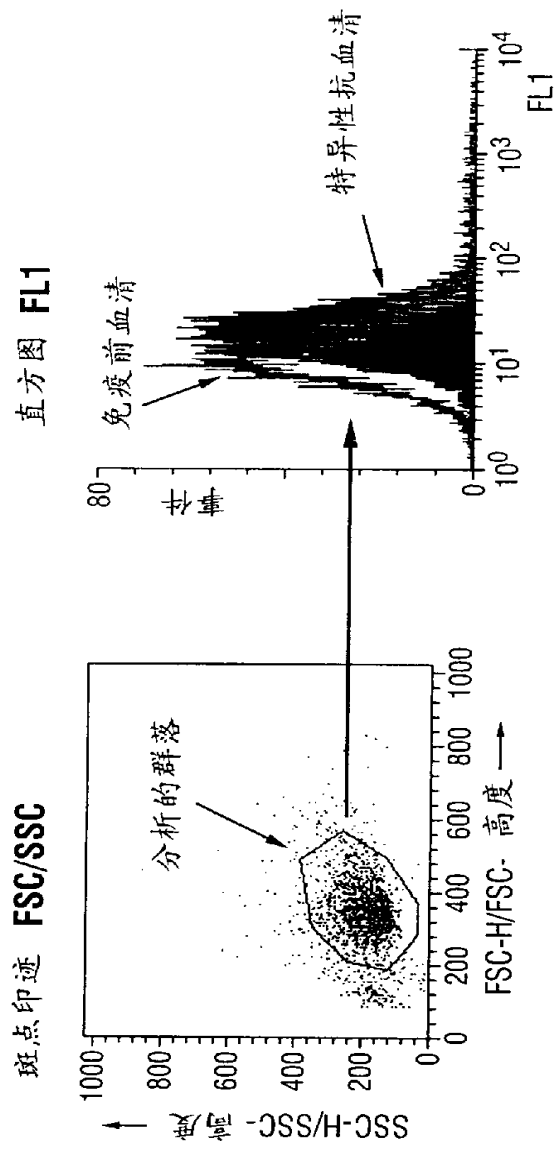


图 5 (HVS CBL-Mix57)

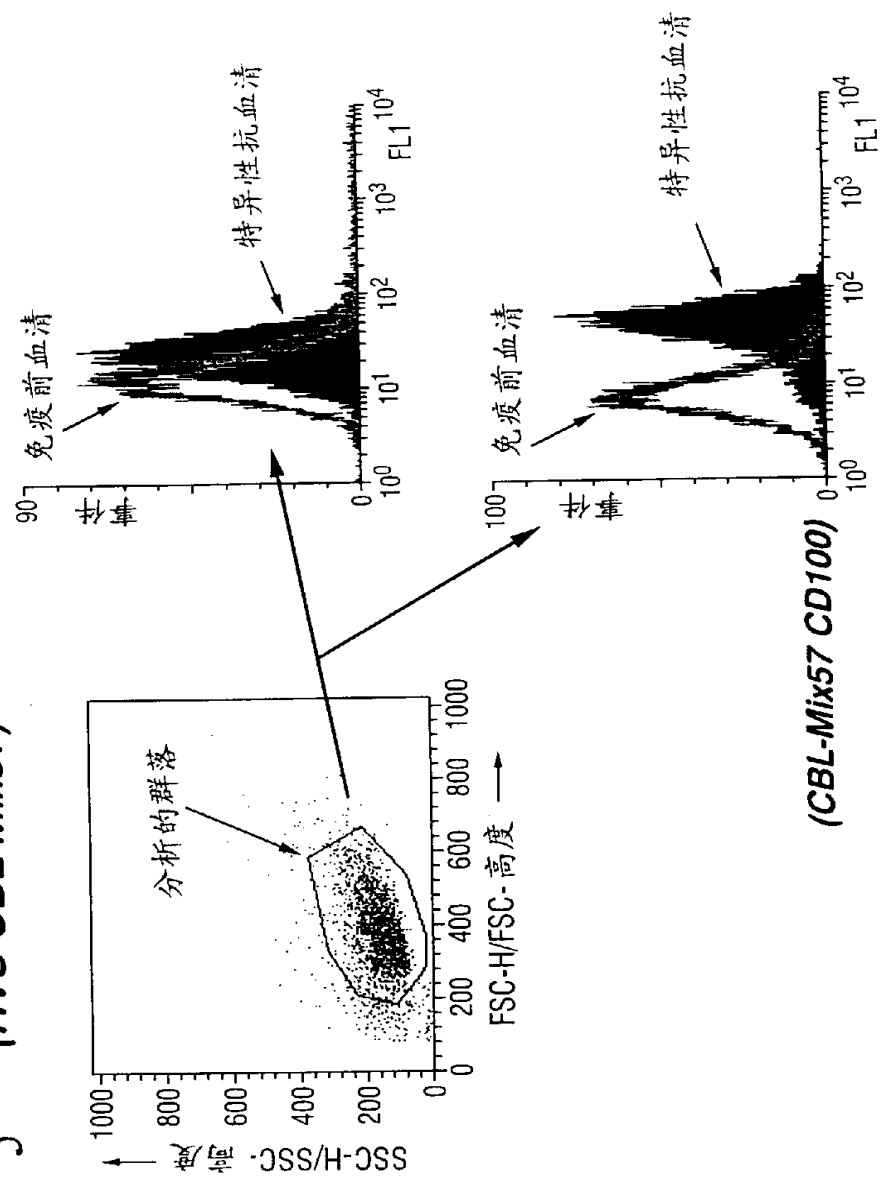


图 5 (续) (CBL-Mix59)

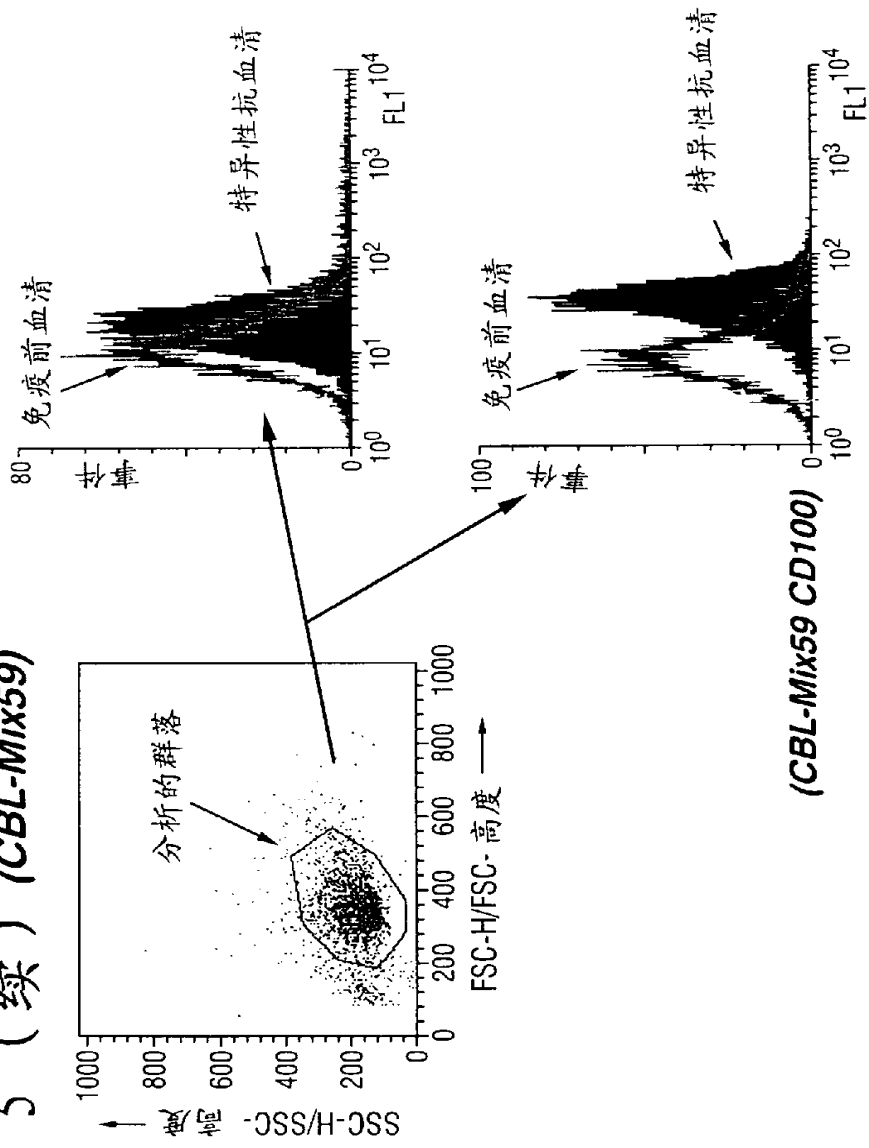


图 6

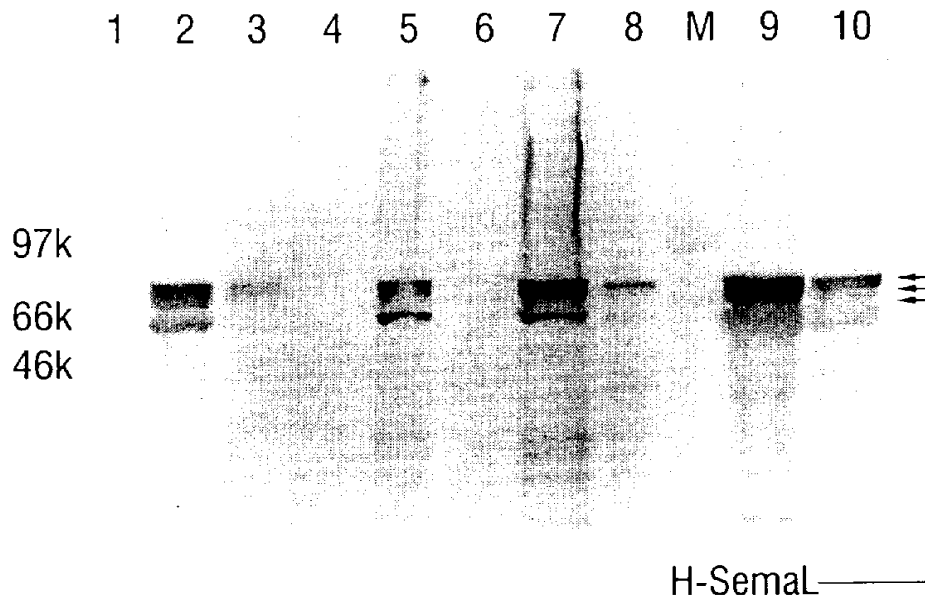


图 7

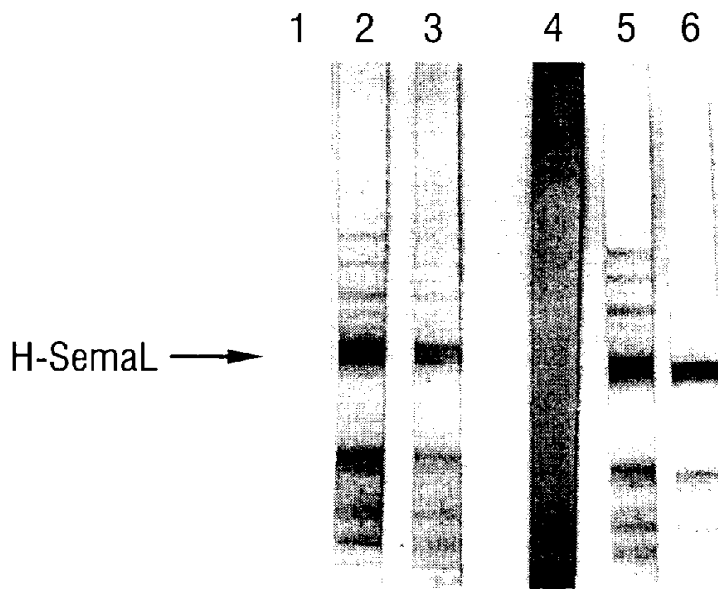


图 8

