

[19]中华人民共和国国家知识产权局

[51]Int. Cl⁶

[12] 发明专利申请公开说明书

[21] 申请号 97198471.9

C12N 15/12

C07K 14/47 C12Q 1/68

A61K 38/18 A01K 67/027

C12N 5/10 C07K 16/18

[43]公开日 1999 年 10 月 20 日

[11]公开号 CN 1232499A

[22]申请日 97.9.29 [21]申请号 97198471.9

[30]优先权

[32]96.10.1 [33]US [31]60/027,633

[32]97.1.16 [33]EP [31]97100583.0

[86]国际申请 PCT/EP97/05355 97.9.29

[87]国际公布 WO98/14568 英 98.4.9

[85]进入国家阶段日期 99.4.1

[71]申请人 居德伦·拉波尔德-赫尔布兰德

地址 联邦德国海德堡

[72]发明人 居德伦·拉波尔德-赫尔布兰德

埃尔科莱·拉奥

[74]专利代理机构 中原信达知识产权代理有限责任公司

代理人 丁业平 王达佐

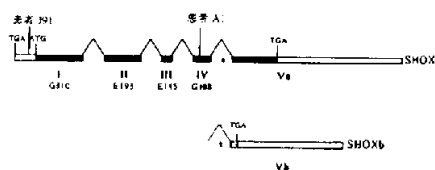
权利要求书 5 页 说明书 69 页 附图页数 5 页

[54]发明名称 人生长基因和身材矮小基因区

[57]摘要

本发明主题是一种分离的人类核酸分子,它编码的多肽包含由 60 个氨基酸组成、具有 SEQIDNO:1 的氨基酸序列的同源异型框区域,该分子具有人体生长调节活性。

在 X 和 Y 染色体上约 500kb 矮小身材关键区内鉴定到三个新基因。至少其中之一导致矮小身材表型。对应该基因的 cDNA 可以用于诊断手段,或者进一步确定矮小身材表型的分子基础。另外,该基因的基因产物的鉴定,提供了新的途径和方法以便开发出矮小身材更好的治疗方法。



权 利 要 求 书

1. 一种分离的人类核酸分子, 该分子编码包含 SEQ ID NO:1 所列氨基酸序列的 60 个氨基酸的同源异型框结构域的多肽, 该多肽具有人体生长调节活性。

2. 根据权利要求 1 的核酸分子, 选自下组:

a) 一种分离的 DNA 分子, 该分子包含: (i) 这样的核苷酸序列, 即编码包含 SEQ ID NO:1 所列 60 个氨基酸的同源异型框结构域的多肽的核苷酸序列, 所述多肽具有人体生长调节活性, 或 (ii) 这样的核苷酸序列, 该序列所编码的多肽包含 SEQ ID NO:1 所列 60 个氨基酸中除了有一或几个氨基酸残基的缺失、增加或取代外的同源异型框结构域, 所述多肽仍具有人体生长调节活性;

b) 一种分离的 DNA 分子, 含有 SHOX ET93[SEQ ID NO:2]和 SHOX ET45[SEQ ID NO:4] 的核苷酸序列或其片段;

c) 能与 a) 或 b) 的 DNA 分子杂交的核酸分子;

d) 核苷酸序列与 a) 或 b) 的核酸分子具有 70%或更高同源性的 DNA 分子。

3. 根据权利要求 2 的 DNA 分子, 该分子所编码的多肽具有在 SEQ ID NO:1 所列 60 个氨基酸的同源异型框结构域之外延伸的 N-末端和/或 C-末端。

4. 根据权利要求 3 的 DNA 分子, 编码长度为 150 到 350 个氨基酸的多肽。

5. 根据权利要求 2 到 4 任一项的 DNA 分子, 还包含 SHOX G310[SEQ ID NO:3]的核苷酸序列。

6. 根据权利要求 2 到 5 任一项的 DNA 分子, 还包含 SHOX

G108[SEQ ID NO:5]的核苷酸序列。

7. 根据权利要求 2 到 6 任一项的 DNA 分子, 还包含 SHOX Va[SEQ ID NO:6]或 SHOX Vb[SEQ ID NO:7]的核苷酸序列。

5

8. 根据权利要求 1 到 4 任一项的 DNA 分子, 编码选自下组的多肽:

- a) 基本具有[SEQ ID NO:11]所列氨基酸序列的转录因子 A;
- b) 基本具有[SEQ ID NO:13]所列氨基酸序列的转录因子 B; 和
- c) 基本具有[SEQ ID NO:15]所列氨基酸序列的转录因子 C。

10

9. 包含 SHOX ET93[SEQ ID NO:2]的核苷酸序列的 DNA 序列。

10. 根据权利要求 9 的 DNA 序列, 还包含 SHOX G310[SEQ ID NO:3]的核苷酸序列。

15

11. 根据权利要求 9 或 10 的 DNA 序列, 还包含 SHOX ET45[SEQ ID NO:4] 的核苷酸序列。

20

12. 根据权利要求 9 到 11 任一项的 DNA 序列, 还包含 SHOX G108[SEQ ID NO:5]的核苷酸序列。

13. 根据权利要求 9 到 12 任一项的 DNA 序列, 还包含 SHOX Va[SEQ ID NO:6]或 SHOX Vb[SEQ ID NO:7]的核苷酸序列。

25

14. 根据权利要求 9 的 DNA 序列, 包含 SHOX ET93[SEQ ID NO:2]和 SHOX ET45[SEQ ID NO:4]的核苷酸序列。

15. 根据权利要求 9 的 DNA 序列, 包含 SHOX ET93[SEQ ID NO:2]、SHOX ET45[SEQ ID NO:4] 和 SHOX G108[SEQ ID NO:5]的核

30

核苷酸序列。

16. 根据权利要求 9 到 15 任一项的 DNA 序列，包含 SHOX G310[SEQ ID NO:3]、SHOX ET93[SEQ ID NO:2]、SHOX ET45[SEQ ID NO:4] 和 SHOX G108[SEQ ID NO:5]的核苷酸序列。

17. 根据权利要求 16 的 DNA 序列，包含 SHOX Va[SEQ ID NO:6]的核苷酸序列。

18. 根据权利要求 16 的 DNA 序列，包含 SHOX Vb[SEQ ID NO:7]的核苷酸序列。

19. 根据权利要求 9 的 DNA 序列，基本由[SEQ ID NO:14]所列的分离出的 PAR1 区基因组序列组成。

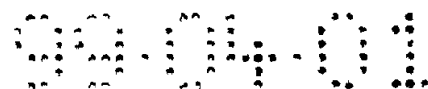
20. 一种包含 SHOX ET92[SEQ ID NO:9]的核苷酸序列的 DNA 序列。

21. 根据权利要求 9 到 20 任一项的 DNA 序列，所述 DNA 是负责调控人体生长的基因组 DNA 或分离的 DNA。

22. 根据权利要求 9 到 21 任一项的 DNA 序列，所述 DNA 是 cDNA。

23. 根据权利要求 22 的 cDNA，基本由 SHOXa [SEQ ID NO:10]或 SHOXb [SEQ ID NO:12]的核苷酸序列组成。

24. 根据权利要求 22 的 cDNA，基本由 SHOT[SEQ ID NO:14]的核苷酸序列组成。



25. 一种具有 SEQ ID NO:11 所列氨基酸序列的人生长蛋白（转录因子 SHOXa）或其功能性片段。

26. 一种具有 SEQ ID NO:13 所列氨基酸序列的人生长蛋白（转录因子 SHOXb）或其功能性片段。

27. 一种具有 SEQ ID NO:15 所列氨基酸序列的人生长蛋白（转录因子 SHOT）或其功能性片段。

10 28. 编码权利要求 25、26 或 27 所述蛋白的 cDNA。

29. 包含权利要求 25 到 27 任一项所述蛋白的药物组合物。

30. 一种治疗矮小身材的方法，包括对需要给药的个体施用治疗有
15 效量的权利要求 25 到 27 所述的蛋白。

31. 权利要求 25 到 27 所述蛋白用于制备治疗矮小身材的药物组合物的用途。

20 32. 权利要求 1 到 24 所述 DNA 序列用于制备治疗由矮小身材基因突变导致的紊乱的药物组合物的用途。

33. 权利要求 1 到 24 任一项所述 DNA 序列用于制备一种试剂盒的用途, 该试剂盒能够鉴定出带有导致人体生长减弱的基因缺陷的个体。

25

34. 权利要求 33 所述 DNA 序列的用途, 用于鉴定矮小身材基因。

35. 基于 RNA 或 DNA 分子确定矮小身材的方法，其中在两个与 SEQ ID NO:2 到 SEQ ID NO:7 中的任一 DNA 序列完全或部分互补的探针存在的条件下，对待测生物样品分子进行扩增，随后用合适的检测系

统进行确定。

36. 权利要求 35 所述的方法用于鉴定带有矮小身材遗传缺陷的人。

5

37. 一种转基因动物，是用包含权利要求 1 到 24 任一项所述 DNA 序列的导致矮小身材的基因转化的。

38. 用权利要求 1 到 24 任一项所述 DNA 序列转化的细胞。

10

39. 鉴定、筛选可用于治疗人矮小身材的药物的检测系统，包含权利要求 38 所述的转化细胞。

15

40. 可用于治疗与矮小身材基因突变有关的紊乱的药物的鉴定、筛选方法，包括提供权利要求 39 所述的检测系统，并在候选药物与所述细胞接触后，确定所述细胞表型的改变或所述细胞表达产物的变化。

41. 包含权利要求 1-8 所述 DNA 分子的表达载体，该载体使所编码多肽得到表达。

20

42. 人生长紊乱的体内基因疗法，所述生长紊乱与 SHOX 或 SHOT 基因内的至少一个突变有关，该方法包括向人细胞导入一种表达质粒，该质粒中权利要求 1-8 任一项所述的 DNA 分子被连接在影响其在人宿主细胞中表达的启动子的下游。

25

43. 根据权利要求 42 所述的方法，用于治疗 Turner 综合症或矮小身材。

30

44. 一种抗体，该抗体是用转录因子 A、B 或 C 或者它们的抗原性片段免疫哺乳动物，并从这些动物中分离出所述抗体而获得的。

说明书

人生长基因和身材矮小基因区

5 本发明涉及新确定的导致人生长失调的人类基因，特别是导致矮小身材或 Turner 综合症的基因的分离、鉴定和特征分析，以及对这些失调病症的诊断和治疗。

10 分离得到的基因组 DNA 或其片段可以用于医药用途，或者作为诊断手段或试剂用于确定或定性分析这些失调所涉及的遗传缺陷。本发明另一个主题是人类生长蛋白（转录因子 A、B、C），它们是上述 DNA 转录为 RNA 或 mRNA 后表达产生的蛋白，这些蛋白可以用于治疗与所述基因内的突变有关的生长失调。本发明还涉及相应的 cDNA 序列，这些 cDNA 序列可以用来制备适用于治疗这些失调病症的重组蛋白。本发
15 明再一个主题是用于表达这些基因的 DNA 序列的质粒载体和合适的含有这些 DNA 的细胞。本发明还有的一个主题涉及分子药物领域，即，采用将本发明的 DNA 连接在能影响它在哺乳动物宿主细胞中的表达的启动子下游而制备得到的表达质粒，从而提供治疗这些生长失调的基因治疗手段和方法。

20

 生长是生物发育的一项基本内容，其调控系统是高度组织化且复杂的。身高是一个多因子性状，受到环境和遗传因子两方面的影响。与身高相关的发育畸形在各种族人群中都是一个普遍现象。生长迟缓导致的矮小身材是见于人类的主要先天性缺陷，其发生率在 100 人中有 3 人。

25

 Turner 综合症是一种常见的染色体紊乱症，在女性中先天性表型的发生率为 1:2500（Rosenfeld 等，1996）。据估计在全部妊娠中有 1-2% 为 45,X 并且这些胎儿 99% 会夭折（Hall 和 Gilchrist, 1990; Robins, 1990）。Turner 综合症（或 Ullrich-Turner 综合症）患者的表型在临床
30 上差异显著(Ullrich,1930;Turner,1938)。但是，矮小身材是总能看到的，

它与性腺发育不全被认为是这种染色体紊乱的最重要症状。Turner 综合症是一种真正的多因子紊乱症。胚胎死亡、矮小身材、性腺发育不全、独特的躯体特征都是由于 X、Y 染色体上的常规基因的单倍体化造成的。这些 X-Y 同源基因的双倍体性是人体正常发育所必需的。预期 Turner 基因（或反 Turner 基因）在女性活性的和失活 X 染色体或 Y 染色体上均表达，以保证准确量的基因产物。因此单倍体性缺陷（由于只有单拷贝的活性基因造成的缺陷），应被考虑是造成这些疾病的遗传机制。

关于矮小身材的产生机制已有多种阐述。据说生长激素和生长激素受体缺乏以及骨骼发育障碍是矮小身材表型的原因（Martial 等,1979; Phillips 等, 1981;Leung 等,1987;Goddard 等,1995）。最近的研究确定：三个人成纤维细胞生长因子受体的编码基因（FGFR1-3）发生突变是多种骨骼发育障碍，包括最常见的侏儒症、软骨发育不全的起因（Shiang 等,1994;Ruosseau 等,1994;Muenke 和 Schell,1995）。一种众所周知且常见（1: 2500 女性）的染色体紊乱，即 Turner 综合症（45, X），也常与矮小身材关联。但是，所有这些已知原因只能解释所有矮小患者的一小部分，而绝大部分矮小身材的原因尚待解释。

性染色体 X、Y 上带有影响身高的基因（Ogata 和 Matsuo,1993），这可以从性染色体异常患者的基因型-表型的相关性推断出来。细胞遗传学研究提供了证据，证明 X 或 Y 染色体短臂末端缺失经常导致该个体的矮小身材（Zuffardi 等,1982;Curry 等,1984）。根据报导，20 多例与 Xp 和 Yp 染色体末端缺失相关的染色体重排，确定导致矮小身材的基因位于假常染色体区域（PAR1）（Ballabio 等,1989,Schaefer 等,1993）。上述定位进一步被缩窄到 PAR1 区域最远端 700bp DNA 范围内，其中带有 DXYS15 侧翼标志序列(Ogata 等,1992;1995)。

哺乳动物的生长调控由一个复杂的系统实施。这就容易想到有多个生长促进基因（蛋白）以高度有序的方式相互作用。其中一个控制身高

的基因被暂时定位于假常染色体区域 PAR1 (Ballabio 等,1989), 已知 X 和 Y 染色体在该区域可随意交换 (参见 Rappold,1993 综述)。完整的 PAR1 区域接近 2700kb。

5 根据对发生缺失的患者进行的研究确定了矮小身材基因的关键区域。当整个 700kb 区域被缺失或该关键区域内的一个特定基因呈现单倍体状态、或被阻断、突变时就会产生矮小身材 (遗传性的矮小身材或 Turner 综合症)。世界范围内 Turner 综合症的发生率是 2500 位女性中有一位; 这种原发性矮小身材在人群中的发生率是 1: 4000-5000。Turner 10 女性和一些矮小身材个体通常接受多年至 10 年以上的非特异性的生长激素 (GH) 治疗, 尽管知道他们的 GH 水平正常, 病因并非是 GH 缺乏。这些患者的治疗费用很昂贵 (估计费用接近 30000USD p.a.)。因此, 应该考虑一种不使用人 GH 的替代疗法解决现存的问题, 即提供一种方法和手段能区别带有基因缺陷的矮小身材患者和该基因没有缺陷的患者; 15 属于基因缺陷的矮小身材患者无论是由于完全的基因缺失 (象 Turner 综合症) 还是点突变 (象遗传性矮小身材的情况), 应能接受这种替代疗法, 而且目前已经可以设计这种方法了。

20 基因型/表型的相关性支持在 Yq 近端和 Yp 远端存在生长基因这一说法。矮小身材也常见于 Xp 末端缺失的个体。近来对假常染色体区域部分单倍体化的男、女患者进行了多方面研究。基于基因型/表型的相关性, 在临近端粒处确定了一个 700kb DNA 的最小共有缺失区域 (Ogata 等,1992; Ogata 等,1995)。显示出人们感兴趣的这个区域位于基因标志 DXYS20 (3cosPP) 和 DXYS15 (113D) 之间, PAR1 区域内的所有候选生长调节基因 (例如, 造血生长因子受体 a; CSF2RA) (Gough 等,1990) 25 根据其物理位置就可以将它们排除在外 (Rappold 等,1992)。这就是说, 生长基因位于 2700 kb PAR1 区域的 700kb 缺失区域中。

30 近来在矮小身材患者中发现他们的性染色体中的假常染色体区域 (PAR1) 发生缺失, 随后确定出 PAR1 区域内的 700kb 的最小共有缺失

区域。利用不同假常染色体标志序列对 AK 和 SS 患者的 DNA 进行 Southern 印迹检测鉴定到远离 DXYS15 约 700kb 处有一个 Xp 末端缺失 (Ogata 等,1992; Ogata 等,1995)。

5 与矮小身材对应的基因区域已被确定是 X、Y 染色体 PAR1 区域内的一个约 500 kb 的区域, 更好的是约 170 kb 的区域。在该区域中已经鉴定出矮小身材基因的三个候选基因。这些基因被定名为 SHOX 基因 (也称为 SHOX93 或 HOX93) (SHOX=含矮小身材同源异型框的基因), pET 92 和 SHOT (第三染色体上的类 SHOX 同源异型框的基因)。

10 SHOX 基因尤其重要, 该基因含有两个分开的剪接位点, 因此有两种变型 (SHOX a 和 b)。在初步研究中, 分析了矮小身材基因核酸序列的必要部分 (SEQ ID No 8)。可以预测出和确定其中的各个外显子或组成部分 (例如, 外显子 I [G310]; 外显子 II[ET93]; 外显子 IV [G108]; pET 92)。所获得的序列信息可以用来设计合适的引物或者能与 SHOX

15 基因的一部分或其片段杂交的核酸探针。SHOX 基因可以用传统的方法进行分离。对矮小身材基因的 DNA 序列进一步进行分析, 可以确定外显子 I-V 的核酸序列 (参见图 1-3)。SHOX 基因含有约 180bp 的一个同源异型框序列 (SEQ ID NO: 1) (参见图 2 和图 3), 从编码第 117 个氨基酸 (Q) 至编码第 176 个氨基酸 (E) 的核苷酸, 即从 CAG (440)

20 到 GAG (619)。通过对 250 个原发性矮小身材个体的筛选, 鉴定该同源异型框序列即为同源异型框—pET 93 (SHOX) 序列, 还分别自一个德国 (A1) 矮小身材患者和一个日本矮小身材患者体内发现了两个点突变。两个点突变在相同的位置, 均导致在第 195 氨基酸处发生蛋白截断, 提示这里可能存在一个突变热点。由于发现了这两个导致蛋白截断的突

25 变是在同一位置, 推测在外显子 4 [G108]中存在一个重组热点。因此可以使用如下的外显子特异性引物, 例如 GCA CAG CCA ACC ACC TAG (正向) 或 TGG AAA GGC ATC ATC CGT AAG (反向)。

30 上述位于 170 kb 间隔区的新的含同源异型框的基因 SHOX, 在选择性剪接后产生两个功能不同的蛋白。突变分析和 DNA 测序用来显示

矮小身材可能起因于 SHOX 的突变。

矮小身材基因关键区域的鉴定和克隆根据本发明如下进行：对 15 名假常染色体区域（PAR1）部分单倍体化的个体作精细的物理图谱研究。通过对照这些个体的身高与缺失断裂点的相关性，确定出约 700 kb 的矮小身材（SS）基因关键区。随后利用 PAR1 区域构建的酵母人工染色体（YACs）(Ried 等,1996)通过粘粒步移将这个关键区克隆生成一个重叠的粘粒连续克隆系。为了找到位于该间隔区的 SS 候选基因，采用多种技术研究位于粘粒 56G10 远端和 51D11 近端之间的约 600 kb 区域。用 cDNA 选择、外显子截留法、CpG 岛克隆，鉴定到两个新基因。

对三个一致性身材矮小的更特殊个体（GA，AT 和 RY）的特征进行研究，可以将矮小身材基因关键间隔区的位置限定到一个更小的 170 kb 的 DNA 间隔区。为了精确定位这些个体的重排断裂点，用来自上述连续克隆系的粘粒在中期染色体上作荧光原位杂交（FISH）。患者 GA，其发生了末端缺失，身高正常，确定出该关键区域的远端界线（断裂点位于粘粒 110E3）；患者 AT，其 X 染色体发生倒位，身高正常，确定出近端界线（断裂点位于粘粒 34F5）。发现发生了末端缺失且矮小身材的患者 RY 的 Y-染色体断裂点也包含在粘粒 34F5 内，提示该区域含有倾向发生染色体重排的序列。

由 Xp/Yp 端粒所限定的整个区域已被克隆生成一组重叠粘粒。用这些粘粒作荧光原位杂交（FISH），研究 6 名 X 染色体发生了重排的患者，其中 3 人身高正常，3 人身材矮小。基因型-表型相关性使我们能将矮小身材基因关键间隔区缩窄到包含对人类生长有重要作用的一个或一些基因的 270 kb 甚至 170 kb DNA 区域。桥接该间隔区的 6 到 8 个粘粒形成的最小盖瓦式路径（tiling path）可以应用于间期和中期 FISH，从而提供了一个有价值的原发性身材矮小症患者的诊断研究工具。

附图简述

图 1 是 SHOX 基因的基因图谱，包括 5 个外显子，标记如下：外显子 I: G310；外显子 II: ET93；外显子 III: ET45；外显子 IV: G108；外显子 Va 和 Vb。其中 Va 和 Vb 源于 SHOX 基因的两个不同剪接位点。外显子 II 和 III 含有 180 个核苷酸的同源异型框序列。

5

图 2 和 3 是 SHOXa 和 SHOXb 的核苷酸序列和推测的氨基酸序列：

SHOXa: 推测翻译起点在第 92 个核苷酸处，而阅读框内第一个终止密码子 (TGA) 在第 968-970 核苷酸，从而形成一个 876bp 的开放阅读框，编码预测为 292 个氨基酸的蛋白 (分别命名为转录因子 A 或 SHOXa 蛋白)。在第 4 核苷酸处的阅读框内 5'终止密码子、起始密码子和推测的终止密码子以粗体表示。同源异型框以框线示出 (从 117 号氨基酸 (Q) 到 176 号氨基酸 (E)，即核苷酸序列是从 CAG 到 GAG)。内含子的位置用箭头示出。位于 3'非翻译区的两个被推测为多聚腺苷酸化信号序列的下面加有下划线。

15

SHOXb: 876bp 的开放阅读框从第 92 号核苷酸对应的第一个蛋氨酸 A 到第 767-769 号核苷酸对应的阅读框内终止密码子为止，形成 675 bp 的开放阅读框，编码预计为 225 个氨基酸的蛋白 (分别为转录因子 B 或 SHOXb 蛋白)。内含子的位置用箭头示出。外显子 I-IV 与 SHOXa 相同，外显子 V 是 SHOXb 特有的。3'非翻译区内的两个被推测为多聚腺苷酸化信号序列的下面加有下划线。

20

图 4 是 SHOT 的核苷酸序列和推测的氨基酸序列。推测翻译起点在第 43 个核苷酸处，而阅读框内第一个终止密码子 (TGA) 在第 613-615 核苷酸处，从而形成一个 573bp 的开放阅读框，编码预测为 190 个氨基酸的蛋白 (分别命名为转录因子 C 或 SHOT 蛋白)。同源异型框由框线示出 (从 11 号氨基酸 (Q) 到第 70 号氨基酸 (E)，即核苷酸序列是从 CAG 到 GAG)。内含子的位置用箭头示出。3'非翻译区内的两个被推测为多聚腺苷酸化信号序列的下面加有下划线。

25

30

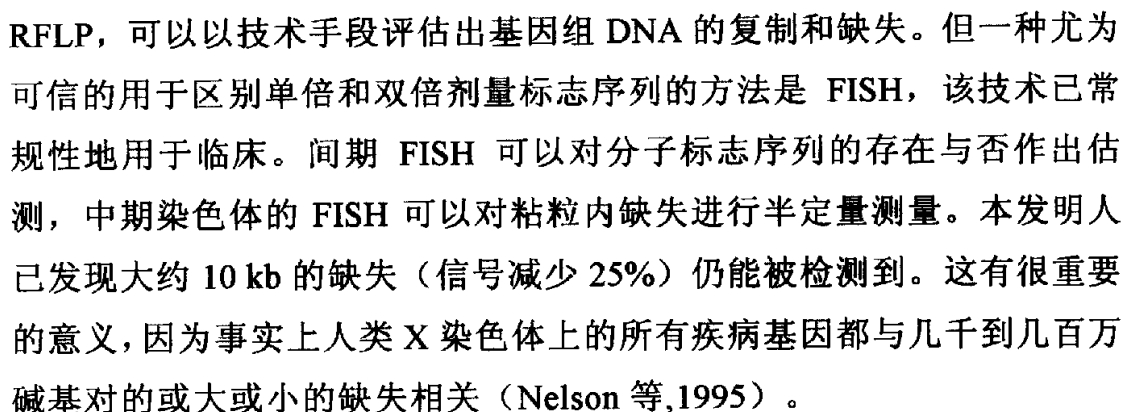
图 5 给出人类 SHOX 基因的外显子/内含子组织方式以及各自在核酸序列中的位置。

序列简述：

- 5 SEQ ID NO. 1: 翻译后的同源异型框结构域 (180 bp) 的氨基酸序列
- SEQ ID NO. 2: SHOX 基因的外显子 II (ET93)
- SEQ ID NO. 3: SHOX 基因的外显子 I (G310)
- SEQ ID NO. 4: SHOX 基因的外显子 III (ET45)
- 10 SEQ ID NO. 5: SHOX 基因的外显子 IV (G108)
- SEQ ID NO. 6: SHOX 基因的外显子 Va
- SEQ ID NO. 7: SHOX 基因的外显子 Vb
- SEQ ID NO. 8: SHOX 基因的初始核苷酸序列
- SEQ ID NO. 9: ET92 基因
- 15 SEQ ID NO. 10: SHOXa 序列 (亦可参见图 2)
- SEQ ID NO. 11: 转录因子 A (亦可参见图 2)
- SEQ ID NO. 12: SHOXb 序列 (亦可参见图 3)
- SEQ ID NO. 13: 转录因子 B (亦可参见图 3)
- SEQ ID NO. 14: SHOX 基因
- 20 SEQ ID NO. 15: SHOT 序列 (亦可参见图 4)
- SEQ ID NO. 16: 转录因子 C (亦可参见图 4)

因为在本发明之前，引起人类生长失调的目标基因 (例如，矮小身材基因区域) 是不为人知的，患者的生物学和临床症状与这个缺失的关系有助于理解这个基因的功能。在本发明中，用荧光原位杂交 (FISH) 检测了 6 名患者的中期和间期淋巴细胞核。其目的是测试所有重叠粘粒能否作为 FISH 探针，以及确定四个病例的断裂点区域，从而确定矮小身材基因最小的关键区域。

30 利用小心控制的定量 PCR 或 Southern 印迹剂量估计或者利用



还可使用本发明的 DNA 序列通过合适的载体系统对动物，例如哺乳动物进行转化。得到的转基因动物随后可用于活体内研究，用于筛选或鉴定能有效治疗涉及矮小身材等疾病的药物。如果动物对所给的候选化合物或试剂呈阳性反应，则可将该化合物或试剂或其衍生物设计成药

物。通过适当的手段，本发明的序列还可以用于其目的在于找到能补偿导致身材矮小的基因缺失的方法的遗传实验（基因被敲除的动物）。

5 本发明另一个方面是，本发明 DNA 序列可以用于转化细胞。可用这些细胞鉴定能有效治疗包括矮小身材等疾病的药物，或筛选这样的化合物或化合物库。在适当的实验体系中，可以确定这些细胞的表型或表达方式的变化，从而在开发药物过程中鉴定到令人感兴趣的候选药剂。

10 本发明的 DNA 序列还可用于设计出合适的引物，该引物能与矮小基因片段或区段在严谨条件下杂交。可以构建合适的引物序列，这些序列可用于诊断由基因缺陷而导致的身材矮小患者。关于这一点，出现在相同位置上的两个突变提示有一个突变热点存在，这是值得人们注意的。

15 总体来说，应该理解的是本发明的 DNA 序列还包括根据遗传密码的简并性变成所示特定序列的 DNA 序列，或者与图示的特定序列在严谨条件下能杂交的 DNA 序列。

本发明具体涵盖以下方面的内容：

20 a)一种分离的人类核酸分子，该分子编码的多肽包括 60 个氨基酸的同源异型框结构域，该结构域具有 SEQ ID NO: 1 的氨基酸序列，该多肽有调节人体生长的活性。

b)一种分离的 DNA 分子，该分子所包含的核苷酸序列基本如图 2、3 或图 4 所示，具体地讲如 SEQ ID NO: 10、SEQ ID NO: 12 或 SEQ ID
25 NO: 15 所示。

c)能与 b)项的 DNA 分子杂交的 DNA 分子。

d) c)中能与 b)项的 DNA 分子在 60-70℃在标准缓冲液中杂交的 DNA 分子

e)包含与图示的 SEQ ID NO: 10、SEQ ID NO: 12 或 SEQ ID NO:
30 15 核苷酸序列有 70%或更高的同源性的核苷酸序列的 DNA 分子，该

DNA 分子编码具有人体生长调节活性的多肽。

f)具有 SEQ ID NO: 11、13 或 16 所列氨基酸序列的人生长蛋白或其功能性片段。

g)用 f)项的人生长蛋白或其抗原性变异体免疫动物后得到的抗体。

h)含有人生长蛋白或其功能性片段的药物组合物，该组合物可用于治疗由人类生长基因的遗传突变导致的失调。

i)一种能有效治疗如上 h)项所述生长失调的物质的筛选方法，包括检测能与上述 a)-e)项的任一 DNA 分子杂交的信使 RNA，以便衡量用该物质处理宿主细胞时引起的 DNA 分子表达水平的提高。

j)含有上述 a)-e) 项中任意核酸分子的表达载体或质粒，该载体或质粒能在哺乳动物细胞中表达所述 DNA 分子。

k)一种在生物组织或体液样品中检测引起矮小身材的一个或一些基因的方法。

在 k)项所述方法中，优选采用本领域技术人员熟知的检测特异核苷酸序列的核酸扩增技术，例如 PCR，其详细方法描述于，例如，本文引作参考的文献中，Mullis 等,1986,冷泉港研讨会定量生物学，51，263-273，及 Saiki 等,1988，科学 239，487-491。要确定的矮小身材核苷酸序列主要是 SEQ ID NO: 2 到 SEQ ID NO: 7 所代表的那些。

原则上，所有能用于在生物样品中扩增和检测引起生长停滞的基因缺陷的寡核苷酸引物和探针，都适用于扩增矮小身材相关靶序列。特别是表 1 给出的适用于本发明的外显子特异引物对。随后可以进行合适的检测例如放射性或非放射性标记检测。

表 1

外显子	有义引物	反义引物	产物(bp)	Ta(°C)
5'-I(G310)	SP1	ASP1	194	58
3'-I(G310)	SP2	ASP2	295	58
II(ET93)	SP3	ASP3	262	76/72/68
III(ET45)	SP4	ASP4	120	65
IV(G108)	SP5	ASP5	154	62
Va (SHOXa)	SP6	ASP6	265	61

引物中缩写符号的含义:

SP 1 : ATTTCCAATGGAAAGGCGTAAATAAC

SP2 : ACGGCTTTTGTATCCAAGTCTTTTG

SP3 : GCCCTGTGCCCTCCGCTCCC

SP4 : GGCTCTTCACATCTCTCTCTGCTTC

SP5 : CCACACTGACACCTGCTCCCTTTG

SP6 : CCCGCAGGTCCAGGCTCAGCTG

ASP1 : CGCCTCCGCCGTTACCGTCCTTG

ASP 2 : CCCTGGAGCCGGCGCGCAAAG

ASP 3 : CCCCCCCCCGCCCCCGG

ASP 4 : CTTCAGGTCCCCCAGTCCCG

ASP 5 : CTAGGGATCTTCAGAGGAAGAAAAAG

ASP 6 : GCTGCGCGGCGGGTCAGAGCCCCAG

单链 RNA 也可以作为靶序列。RNA 反转录为 cDNA 的方法是众所周知的, 这一技术被描述于 Sambrook 等, 分子克隆: 实验室手册, 纽约, 冷泉港实验室 1989。另外, 优选使用耐热的具有 RT 活性的 DNA 聚合酶进行反转录。

另外, 可以用以前描述过的技术从一群人中挑选其矮小身材是由基因缺陷导致的人, 以便结果进行更特异的药物治疗。

5

10

15

30

基酸序列的一或多个位点中插入或缺失一或多个氨基酸。此处所述多肽一般是指未被修饰生物学活性基本保持不变的等效多肽。

本发明通过以下实施例来说明：

5

实施例 1

患者

6 名患者均为原发性染色体异常。

10

CC 是一名核型为 45,X/46,X pus dic(X) (Xqter→Xp22.3::Xp22.3→Xqter)的女孩。在 6 岁半时最后一次检查，她的身高为 114 厘米（处于第 25-50 百分位的范围内）。其母身高为 155 厘米，其父未做检测。欲知详细内容，请参见 Henke 等,1991。

15

GA 是一名核型为 46,X der X (3pter→3p23::Xp22.3→Xqter)的女孩。在 17 岁时最后一次检查，她的身高正常，为 159 厘米。其母身高为 160 厘米，其父为 182 厘米。欲知详细内容，请参见 Kulharya 等,1995。

20

SS 是一名核型为 46,X rea (X) (Xqter→Xq26::Xp22.3→Xq26:)的女孩。11 岁时身高仍低于日本女孩生长曲线的第 3 百分位；预计其成年身高为 148.5cm，低于其目标身高（163cm）和目标身高范围（155-191cm）。欲知详细内容，请参见 Ogata 等,1992。

25

AK 是一名核型为 46,X rea (X) (Xqter→Xp22.3::Xp22.3→Xp21.3:)的女孩。13 时身高仍低于日本女孩生长曲线的第 2 百分位；预计其成年身高为 142.8cm，低于其目标身高（155.5cm）和目标身高范围（147.5-163.5cm）。欲知详细内容，请参见 Ogata 等,1995。

30

RY：一个环形 Y 患者，检测 100 个淋巴细胞得其核型为 46,X,r(Y)/46,X dic r(Y)/45,X[95:3:2]；其 16 岁时的最终身高为 148 cm；

他的三个兄弟身高均在正常范围，分别为：170 cm（16 岁，兄弟 1），164 cm（14 岁，兄弟 2）、128cm（9 岁，兄弟 3）。这个患者生长迟缓非常严重，可能是在 Yq 还缺失了 GCY 座位。

- 5 AT: 运动失调、X 染色体有倒位的男孩；7 岁身高正常，为 116 cm，其父母分别为 156 cm 和 190cm。

用于突变分析的患者：

- 10 检测 250 名原发性矮小身材个体的 SHOXa 突变。患者的选择遵循如下原则：其年龄对应的身高低于全国身高标准的第三百分位，标准误差为-2（SDS）；没有病原性疾病，特别是：相对妊娠年龄体重（身高）正常，体型正常，没有慢性器官紊乱，饮食正常，没有精神失常，没有骨骼发育不良，没有甲状腺或生长激素缺乏。

- 15 家族 A:

- 20 病例 1 和 2 是一个德国无血缘关系家庭的矮小身材子女。男孩（病例 1）生于在妊娠 38 周时的剖腹产。出生体重为 2660 克，身高为 47cm。他的发育除生长情况外均正常。6.4 岁时检测，他较为矮小（106.8cm，-2.6SDS）和肥胖（22.7kg），但其他方面正常。他的骨骼年龄没有延迟（6 岁），X-射线检查也排除了骨骼发育不良。血清中 IGF-I 和 IGFBP-3 水平和甲状腺参数指示不可能有 GH 或甲状腺激素缺陷。女孩（病例 2）出生于足月剖腹产。出生体重为 2920 克，身高为 47cm。她的发育进程正常，但在 12 个月以前生长不良是明显的（身高：67cm，-3.0 SDS）。4 岁时，其身高为 89.6cm（-3.6SDS）。没有明显的体型畸形或比例失调。她也不肥胖（13 kg）。她的骨骼年龄是 3.5 岁，排除了骨骼发育不良。激素参数正常。有趣的是该男孩和女孩均达到 Turner 综合症女性生长曲线的第 50 百分位。其母亲是家族中最矮的，患有轻度肢根比例失调（142.3cm，-3.8SDS）。她的两姐妹之一（150cm，-2.5SDS）和该家庭的外婆（153cm，-2.0SDS）均矮小但没有比例失调。另一个姐妹身高正常（167cm，+0.4SDS）。该家庭中父亲身高为 166cm（-1.8SDS），
- 25
- 30

外公身高 165cm (-1.9SDS)。另一个患者为日本血统，显示有同样的突变。

实施例 2

5 矮小身材基因的鉴定

A. 原位杂交

a) 荧光原位杂交 (FISH)

用粘粒在 Xp/Yp 假常染色体区域 (PAR1) 进行荧光原位杂交 (FISH)。FISH 研究所用粘粒为 64/75cos(LLNLc110H032)、E22cos (2e2)、F1/14cos (110A7)、M1/70 cos (110E3)、P99F2cos (43C11)、P99cos (LLNLc110P2410)、B6cosb (1CRFc104H0425)、F20cos (34F5)、F21cos (ICRFc104G0411)、F3cos2 (9E3)、F3cos1 (11E6)、P117cos (29B11)、P6cos1 (ICRFc104P0117)、P6cos2 (LLNLc110E0625)和 E4cos (15G7)、按照 Lichter 和 Cremer,1992 方法进行该 FISH 研究。简要步骤是，取一毫克各粘粒克隆用生物素标记，与人中期染色体杂交，所取反应条件能抑制重复 DNA 序列产生的信号。用偶联 FITC 的亲合素检测杂交信号。用冷的电荷耦合器摄影系统拍摄 FITC 图象 (Photometrics, Tucson, AZ)。

20

b) 物理图谱

粘粒来源于 Lawrence Livermore 国家实验室 X-和 Y-染色体库和伦敦皇家癌症研究基金会 (现为柏林 Max Planck 分子遗传学研究所) X 染色体库。用 DXYS15 远端粘粒 E4cos、P6cos2、P6cos1、P117cos 和 F3cos1 可以确定仍有两个拷贝的 E4cos、P6cos2、P6cos1 和一个拷贝的 P117cos 和 F3cos1。患者 AK 和 SS 的断裂点确定位于粘粒 P6cos1 上，相互的最大图距为 10kb。可以得出结论：AK 和 SS 的异常 X 染色体有约 630 kb 的 DNA 缺失。

25

30 其他粘粒来自 ICRF X 染色体特异粘粒库 (ICRFc104)，Lawrence

Livermore X 染色体特异粘粒库 (LLNLc110) 和 Y 染色体特异库 (LLCO3'M) 以及自建的包含整个基因组的粘粒库。用所有已知位于这一区域的探针和用整个 YACs 作为探针进行杂交以鉴定粘粒。为证实重叠序列, 在用已知探针不能证实重叠的情况下, 要用来自于几个粘粒的末端探针。

c) Southern 印迹杂交

用不同假常染色体标志序列做 Southern 印迹杂交, 证明 CC 患者的 X 染色体断裂点位于 DXYS20 (3cosPP) 和 DXYS60 (U7A) 之间 (Henke 等,1991)。为了证实这个发现并精确确定断裂点的位置, 用粘粒 64/75cos、E22cos、F1/14cos、M1/70cos、F2cos、P99F2cos 和 P99cos 做 FISH 探针。可以确定患者 CC 异常 X 染色体上的断裂点在 E22PAC 中, 位于粘粒 64/75cos (一个拷贝) 和 F1/14cos (两个拷贝) 之间。患者 CC 丢失了接近 260-290 kb 的 DNA, 身高仍正常。

Southern 印迹杂交在高度严谨条件下在 Church 缓冲液 (0.5M NaPi pH7.2, 7%SDS, 1mM EDTA) 中进行, 反应温度 65℃, 用 40mM NaPi, 1%SDS 于 65℃洗涤。

d)FISH 检测

生物素标记的粘粒 DNA (插入片段为 32-45 kb) 或粘粒片段 (10-16 kb) 与患者的被激活淋巴细胞的中期染色体杂交, 反应条件如前所述 (Lichter 和 Cremer,1992)。发生了杂交的探针用亲和素偶联的 FITC 检测。

e)PCR 扩增

所有 PCR 均在 50ul 反应体系中进行, 所述 50ul 中含有 100pg-200ng 模板、各为 20pmol 的引物、200uM dNTP's (Pharmacia)、1.5mM MgCl₂、75mM Tris/HCl (pH9)、20mM(NH₄)₂SO₄、0.01%(w/v)Tween20 和 2U 的 Goldstar DNA 聚合酶 (Eurogentec)。热循环用 Thermocycler

GeneE(Techne)做。

f)外显子扩增

5 外显子扩增采用四个粘粒库,各含 4 到 5 个来自粘粒连续克隆系的克隆。每个粘粒库中的粘粒用 *Sau3A* 部分消化。将经凝胶纯化的 4-10kb 片段克隆至 *Bam*HI 消化的 pSPL3B 载体 (Burn 等, 1995) 中, 用如以前所描述方法做外显子扩增 (Church 等, 1994)。

g)基因组测序

10 将两个粘粒 LLOYNCO3'M'15D10 和 LLOYNCO3'M'34F5 经超声破碎的断片分别亚克隆至 M13mp18 载体中。从每个粘粒库至少挑出 1000 个噬菌斑, M13 DNA 的制备和测序用染料-终止剂 (dye-terminators)、耐热测序酶 (Amersham) 和通用 M13-引物 (MWG-BioTech)。在 ABI-377 测序仪上跑胶, 收集数据, 并用 GAP4 程序
15 (Staden) 进行编辑。

20 6 名患者中, GA 的染色体断裂点的特征最不明确。先前用于检测它们是否存在于 X 染色体的最远端标志物是 DXS1060 和 DXS996, 它们与端粒的图距接近 6Mb(Nelson 等,1995)。检测了几个含有 PAR1 内的不同基因序列的粘粒 (MIC2、ANT3、CSF2RA 和 XE7), 结果表明它们全部位于易位染色体上。由矮小身材基因关键区域例如染色体产生的粘粒断定易位断裂点位于粘粒 M1/70cos。正常的和重排 X 染色体之间 M1/70cos 信号强度的定量对比显示这个粘粒的约 70%缺失。

表 2

	CC	GA	AK	SS
64/75 cos	-	-		
E22 cos	-	-		
F1/14 cos	+	-		
M1/70 cos	+	(+)		
F2 cos	+	+		
P99F2 cos	+	+		
P99 cos	+	+		
B6 cos		+		
F20 cos				
F21 cos				
F3 cos2				
F3 cos1			-	-
P117 cos			-	-
P6 cos1			+	+
P6 cos2			+	+
E4 cos			+	+

表 2: 表中总结了用 16 个粘粒在 4 名患者中检测的 FISH 数据

[-]一个拷贝; 表明相应粘粒在重排 X 上缺失, 但在正常 X 染色体上存在。

[+]两个拷贝; 表明相应粘粒在重排 X 和正常 X 染色体上都存在。

[(+)]断裂点区域; 表明 FISH 显示断裂点在粘粒内部。

总之, 用荧光标记的粘粒探针和原位杂交对 6 名 X 染色体重排的患者进行的分子检测表明矮小身材基因关键区可以缩窄到一个 270 kb 间隔区, 该区由患者 GA 着丝粒远端断裂点和患者 AK 和 SS 的着丝粒近端界定。

基因型-表型相关性可以提供一定信息，该相关性已被选择用于描述在人 X 和 Y 染色体上的矮小身材基因关键间隔区。本研究中，用 FISH 检测患者淋巴细胞中期染色体（spreads）和间期细胞核，这些患者的 X 染色体上有缺失和易位，这些断裂点位于 Xp22.3。断裂点在四名患者中的两位（AK 和 SS）中呈现簇状结构，预计是由有染色体重排倾向的序列造成的。发现另一个环形 Y 患者在 270 kb 关键区内有中断，由此使关键间隔区减小到 170 kb。

对照 6 个人的身高与缺失断裂点，可以确定 170 kb 间隔区位于假常染色体区域，该区域的存在与否对身高有显著的影响。这个间隔区由患者 GA 距离端粒 340 kb 的 X 染色体断裂点（Xptel）和患者 AT 及 RY 的 510/520 kb 处的 Xptel 近端断裂点界定。这样将关键间隔区缩小到其原来大小（Ogata 等,1992;Ogata 等,1995）的几乎四分之一。现在可以用 6 到 8 个的一组粘粒做 FISH 试验，用于确定大量原发矮小身材患者中这个基因组位点的分布优势和重要性。

B. 候选矮小身材基因的鉴定

为了找到最小的 170 kb 关键区中的转录单位，对 6 个粘粒（110E3、F2 cos、43C11、P2410、15D10、34F5）做外显子截留和 cDNA 选择研究。三个不同的阳性克隆（ET93、ET45 和 G108）由外显子截留法分离得到，它们又全部定位到粘粒 34F5 上。以前使用 cDNA 选择和过量使用 25 个不同 cDNA 文库的方法经证明无效，这一点提示间隔区的基因的表达丰度极低。

为了查明间隔区内的基因是否有丢失，用任意 M13 方法和染料终止剂化学测定 PAR1 这个区域约 140 kb 的核苷酸序列。挑选出的用于序列分析的粘粒相互重叠程度最低，并且它们组合在一起则跨越整个关键间隔区。DNA 序列分析后，用 1.3c 版的“X Grail”程序和外显子截留法程序 FEXHB 推测蛋白序列，证实了先前克隆的 3 个外显子。除先前分离的外，没有检测到其他编码蛋白质的基因。

C. 矮小身材候选基因 SHOX 的分离

假定三个外显子克隆 ET93、ET45 和 G108 都是同一个基因的一部分，用它们作为探针筛选 14 个不同的 cDNA 文库，这些文库来自 12 个不同的胎儿（肺、肝、脑 1 和 2）和成人组织（卵巢、胎盘 1 和 2、成纤维细胞、骨骼肌、骨髓、脑、脑干、视丘下部、垂体）。在将近 14,000,000 个被铺板克隆中一个克隆也没有检测到。做 3' 和 5' RACE 来分离全长转录产物。做 3' RACE 时，来自外显子 G108 的引物用于胎盘、骨骼肌和骨髓成纤维细胞组织的 RNA，结果显示 G108 在这些组织中表达。所有三种组织均产生两个不同的分别为 1173 和 652bp 的 3' RACE 克隆，提示有两个不同的 3' 外显子存在。命将这两种不同的形式命名为 SHOXa 和 SHOXb。

为了提高分离到表达丰度极低的该基因的完整 5' 外显子的机会，用视黄酸和波佛酯 PMA 处理 Hela 细胞系。用该经诱导的细胞系的 RNA 和胎盘、骨骼肌的 RNA 构建一个“马拉松 cDNA 文库”。从所有三种组织中均分离到相同的 5' RACE cDNA 克隆。

实验步骤：

RT-PCR 和 cDNA 文库构建

从 Clontech 购得人心脏、胰腺、胎盘、骨骼肌、胎儿肾脏和肝的 polyA⁺ RNA。从骨髓成纤维细胞系用 TRIZOL 试剂（Gibco-BRL）按照厂商的说明分离总 RNA。cDNA 第一链的合成采用 Superscript cDNA 第一链合成试剂盒（Gibco-BRL），用寡（dT）-接合体引物（GGCCACGCGTCGACTAGTAC[dT]₂₀N），反应以 100ng polyA⁺ RNA 或者 10ug 总 RNA 起始。cDNA 第一链合成后，将反应体系稀释 1/10。5ul 稀释液用于以下 PCR 试验。

按照厂商描述的马拉松 cDNA 扩增试剂盒（Clontech）使用方法，由胎盘、骨骼肌的 RNA 构建“马拉松 cDNA 文库”。

胎儿脑 (Catalog # HL5015b)、肺 (HL3022)、卵巢 (HL1098a)、垂体 (HL1097 v) 和视丘下部 (HL1172 b) cDNA 文库购自 Clontech。脑、肾脏、肝和肺 cDNA 文库是 Clontech 快速筛选 (型) 人 cDNA 文库组群 (panel) 的一部分。胎儿肌肉 cDNA 文库自 UK 人类基因组图谱计划信息中心获得。

D. 序列分析和 SHOX 基因结构

分析 5' 和 3' RACE 衍生克隆的序列, 得到 SHOXa 和 SHOXb (1349 和 1870bp) 的一个共有序列。鉴定到 1870bp (SHOXa) 和 1349bp (SHOXb) 的唯一开放阅读框, 分别产生 292 (SHOXa) 和 225 (SHOXb) 个氨基酸的蛋白。转录产物 a 和 b 有相同的 5' 末端, 但 3' 末端的外显子不同, 这个发现提示所用的剪接信号是不同的。将两个 cDNA 和来自粘粒 LL0YNCO3'M'15D10、LL0YNC3'M'34F5 的基因组 DNA 序列进行完全的序列对比, 从而建立外显子-内含子结构 (图 4)。基因含有大小在 58bp (外显子 III) -1146 bp (外显子 Va) 的 6 个外显子。外显子 I 含有一个 CpG 岛、起始密码子和 5' 区。每个选择性剪接的外显子 Va 和 Vb 都带有一个终止密码子和 3'-非编码区。

实施例 3

在 160kb 的矮小身材基因关键区域内鉴定到两个 cDNA。这两个 cDNA 对应着基因 SHOX 和 pET92。两者是通过粘粒亚克隆与 cDNA 文库杂交鉴定到的。

完全涵盖所述关键区的一套粘粒克隆给我们提供了鉴定矮小身材发生基因的遗传材料。用外显子截留法和 cDNA 选择技术实施目的在于从该区域分离基因的定位克隆计划。根据它们位于假常染色体区内, 可以设想这些基因逃脱了 X-失活并且具有剂量效应。

对在其丢失或不完善时将导致矮小身材的基因的克隆, 提供了通过

例如单链构象多态性在该基因内进行突变分析的基础，这代表着进一步提高了诊断准确性。另外，这个基因的克隆以及随后的生化性质研究开辟了通往对参与生长调控的生物过程有更深理解的道路。

- 5 本发明的 DNA 序列提供了第一个分子检测手段，在复杂的不同种类的原发矮小身材患者中确认特异性遗传紊乱的个体。

实施例 4

SHOXa 和 SHOXb 的表达方式

- 10 用单个外显子作杂交探针进行的 Northern 印迹分析显示每个外显子表达图案各不相同，这一点强烈提示不同大小和强度的条带代表与其他富含 GC 的基因序列发生交叉杂交形成的产物。为了得到基因 SHOXa 和 SHOXb 的更真实的表达概况，用来自不同组织的 RNA 做 RT-PCR。在骨骼肌、胎盘、胰腺、心脏和骨髓成纤维细胞中观察到 SHOXa 的表
15 达，而 SHOXb 仅在胎儿肾脏、骨骼肌和骨髓成纤维细胞中表达，且在骨髓成纤维细胞中的表达远远超过其他组织，表达最强。

- 20 由胎儿脑、肺和肌肉以及成人脑、肺和垂体制备的几个 cDNA 文库中的 SHOXa 的表达，以及 SHOXb 在测试文库中均不表达，再次证明一种剪接形式（SHOXa）表达更广泛，而另一种（SHOXb）主要以组织特异性方式表达。

- 25 我们用由不同细胞系提取的 RNA 做 RT-PCR 来评测 X 和 Y 染色体上 SHOXa 和 SHOXb 的转录活性，所述细胞系所含有的人染色体只有活性 X、失活 X 或者 Y 染色体。所有细胞系均显示预期中的 119bp（SHOXa）和 541bp（SHOXb）的扩增产物，明确说明 SHOXa 和 SHOXb 都逃脱了 X-失活。

- 30 SHOXa 和 SHOXb 编码新的同源结构域蛋白。SHOX 是从哺乳动物到鱼类和两翼昆虫高度保守的基因。除同源结构域外，5' 最末端和

3'最末端是人和鼠之间可能的保守区，表明它在功能上的重要性。人和鼠之间那些氨基酸区域的差异在进化过程没有发生累积。

实验步骤：

5 a) 5'和 3'RACE

为了克隆 SHOXa 和 SHOXb 转录产物的 5'末端，用我们构建的“马拉松 cDNA 文库”做 5'RACE。所用寡核苷酸引物如下：SHOX B 反向，GAAAGGCATCCGTAAGGCTCCC(位置 697-718，反向链)和接合体引物 AP1。按如下参数完成 PCR：94℃、2 分钟，94℃、30 秒，70℃、30 秒，10 72℃、2 分钟，5 个循环。94℃、30 秒，66℃、30 秒，72℃、2 分钟，5 个循环。94℃、30 秒，62℃、30 秒，72℃、2 分钟，25 个循环。第二轮扩增采用 1/100 的 PCR 产物和以下巢式寡核苷酸引物：SHOX A 反向，GACGCCTTTATGCATCTGATTCTC(位置 617-640，反向链)和接合体引物 AP2。做 35 个 PCR 循环，退火温度为 60℃。

15

为了克隆 SHOXa 和 SHOXb 转录产物的 3'末端，按照以前的描述 (Frohman 等,1988) 用寡聚 (dT) 接合体引导合成的第一链 cDNA 做 3'RACE。所用寡核苷酸引物如下：SHOX A 正向，GAATCAGATGCATAAAGGCGTC(位置 619-640)和寡聚 (dT) 接合体。20 按如下参数完成 PCR：94℃、2 分钟，94℃、30 秒，62℃、30 秒，72℃、2 分钟，35 个循环。二轮扩增采用 1/100 的 PCR 产物和以下巢式寡核苷酸引物：SHOX B 正向，GGGAGCCTTACGGATGCCTTTC (位置 697-718) 和寡聚 (dT) 接合体。做 35 个 PCR 循环，退火温度为 62℃。

25 为了确定 SHOXa 和 SHOXb 转录产物的序列，用一个 5'寡核苷酸引物和一个 3'寡核苷酸引物做 PCR。针对 SHOXa 所用引物如下：G310 正向，AGCCCCGGCTGCTCGCCAGC (位置 59-78) 和 SHOX D 反向，CTGCGCGGCGGGTCAGAGCCCCAG (位置 959-982，反向链)。针对 SHOXb 所用引物如下：G310 正向，AGCCCCGGCTGCTCGCCAGC 和 30 SHOX2A 反向，GCCTCAGCAGCAAAGCAAGATCCC(位置 1215-1238，

e) cDNA 文库的 PCR 筛选

为了检测 SHOXa 和 SHOXb 的表达，用 SHOXa 和 SHOXb 的特异引物对几个 cDNA 文库和第一链 cDNA 进行 PCR 筛选。对 cDNA 文库使用等于 5×10^8 pfu 的 DNA。对 SHOXa，所用引物为 SHOX E 反向，GCTGAGCCTGGACCTGTTGGAAAGG(位置 713-737，反向链) 和 SHOXa 正向。对 SHOXb，所用引物为 SHOX B 正向和 SHOX 2A 反向。两个 PCR 参数均为：94℃、2 分钟，94℃、30 秒，68℃、30 秒，72℃、40 秒，5 个循环。94℃、30 秒，65℃、30 秒，72℃、40 秒，5 个循环。94℃、30 秒，62℃、30 秒，72℃、40 秒，35 个循环。

实施例 5

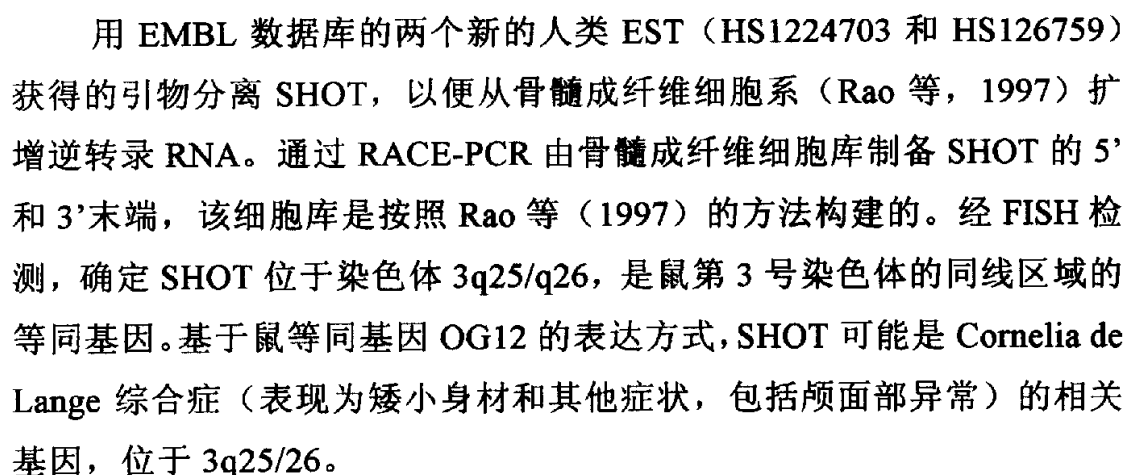
SHOX 和 SHOT 的鼠等同基因 OG12 的表达方式

用第 5p.c. 天到 18,5p.c. 天之间的鼠胚胎以及胎儿和新生鼠做原位杂交确定基因的表达方式。在发育中的肢芽、将来形成鼻子和味觉的鼻中胚层、眼睑、大动脉、发育中的女性生殖腺、发育中的脊髓（仅限于分化中的运动神经细胞）和脑中观察到表达。基于这样的表达方式以及该基因的人等同基因 SHOT 的图谱位置，SHOT 可能是有矮小身材症状的 Cornelia de Lange 综合症的候选基因。

实施例 6

第 3 号染色体上的一个与人生长/矮小身材相关的新的类似 SHOX 同源异形基因 SHOT 的分离

由人体分离到的一个新基因 SHOT(第 3 号染色体上的 SHOX 等同基因)，与鼠 OG12 基因和人 SHOX 基因同源性最高。人 SHOT 基因与鼠 OG12 基因同源性极高，在蛋白水平达到 99% 相同。尽管还未证实，但鉴于 SHOT 和 SHOX 之间的惊人的同源性（仅在同源结构相同），很可能 SHOT 也是一个与矮小身材或人生长相关的基因。



寻找原发性矮小身材患者的突变

对总共 91 个（在 250 名个体中）不相关的男性和女性原发性矮小身材患者（原发性矮小身材在人群中的发生率为 2-2.5%）进行了小片段重排或 SHOXa 基因点突变的检测。设计 6 套 PCR 引物以扩增外显子及外显子侧翼序列和 5'UTR 的一小部分。对最大的外显子，外显子 I，另外制备两个外显子内部引物。用于 PCR 的引物如表 2 所示。

对所有大小在 120 到 295bp 的扩增外显子产物均做单链构象多态性 (SSCP) 检测。只在两名矮小身材患者中 (Y91 和 A1) 识别到条带迁移变化。将 SSCP 模式发生了改变的片段 (独特的 SSCP 构象子) 克隆和测序。为了避免 PCR 和测序时产生的人为假象, 测序用两个独立的 PCR 反应在两条链上进行。患者 Y91 的突变位于 5'UTR 起始密码子 5' 方向 28 bp 处, 包括一个胞嘧啶到鸟嘌呤的置换。为了证实这个突变是否代表一种罕见的多态性, 或者是否是通过例如减弱翻译起始因子的结合来调节基因表达而造成这样的表型, 又检测了其父母和一个姐妹。因为该姐妹和其父身高正常但也有同样的 SSCP 变型 (结果未发表), 这个碱基代表一种罕见的多态性而与表型无关。

将患者 A1 的独特 SSCP 构象子进行克隆和测序，显示有一个胞嘧啶到胸腺嘧啶的转换（核苷酸第 674 位），在预计的 225 和 292 个氨基酸的序列中在氨基酸第 195 位的位点处引入一个终止密码子。为了确定这个无义突变是否与该家族矮小身材有遗传上关联，进行家谱分析。发现所有 6 名矮小个体（身高标准误差低于 2）都有一个异常 SSCP 迁移和胞嘧啶到胸腺嘧啶的转换。身高正常的父亲、一个阿姨和外公都没有这个突变，表明祖母将突变等位基因传给了两个女儿和两个外孙。因此这个家族的突变等位基因和矮小身材间有相关性。

在一个日本血统的矮小身材患者中发现了与以上所述相同的情况。

实施例 8

将本发明的 DNA 序列用于确定这个或这些基因的功能特点。所述 DNA 序列可以作为检索核酸或氨基酸数据库的检索式以确认相关的基因或基因产物。我们已经用 SHOX93 的部分氨基酸序列作为氨基酸数据库的检索式，结果显示与许多已知的同源异型框蛋白高度同源。本发明的 cDNA 可以用来重组制备肽。本领域技术人员所知道的多种表达系统可以用来生产重组蛋白。

利用传统的肽合成方法（Merrifield 蛋白合成方法），合成了具有 CSKSFQKSKDGNGG 序列的肽，并按照标准操作程序从兔和鸡中制备了多克隆抗体。

参考文献

文中引用以下参考文献:

Ashworth A, Rastan S, Lovell-Badge R, Kay G(1991): X 染色体的失活可能解释 XO 人和鼠生存力的差异。自然 351: 406-408。

5

Ballabio A., Bardoni A, Carrozzo R, Andria G, Bick D, Campbell L, Hamel B, Ferguson-Smith MA, Gimelli G, Fraccaro M, Maraschio P, Zuffardi O, Guilo S, Camerino G(1989): 人 X 染色体短臂远端缺失导致的邻近基因综合症。美国科学院学报 86: 10001-10005。

10

Blagowidow N, Page DC, Huff D, Mennuti MT(1989): 一个 Y 染色体性别决定部位缺失的 XY 女性胎儿的 Ullrich-Turner 综合症。美国医学遗传学杂志 34: 159-162。

15

Cantrell MA, Bicknell JN, Pagon RA 等(1989): 46, XY 女性的分子检测和一个新 Y 染色体特异性探针的区域对比。人类遗传学 83: 88-92。

Connor JM, Loughlin SAR(1989): Turner 综合症的分子遗传学。Acta Pediatr. Scand. (增刊) 356: 77-80。

20

Disteche CM, Casanova M, Saal H, Friedman C, Sybert V, Graham J, Thuline H, Page DC, Fellous M(1986): 46,XY 女性的 Y 染色体短臂的小段缺失。美国科学院学报 83: 7841-7844。

25

Ferguson-Smith MA(1965): 性腺发育不全中的核型-表型相关性和该相关性在畸形发病机理中的意义。医学遗传学杂志 2: 142-155。

30

Ferrari D, Kosher RA, Dealy CN(1994): 由于促肿瘤波佛酯处理其软骨分化被抑制的肢间充质细胞继续维持含有同源异型框的基因 MSX1 的表达,但不能进行间隙连接通讯。生物化学和生物物理学研究通讯 205

(1) : 429-434。

Fischer M, Bur-Romero P, Brown LG 等(1990): 人 X-和 Y-染色体上逃脱了 X-失活的同源核糖体蛋白基因以及用于 Turner 综合症的可行性。细胞 63: 1205-1218。

Freund C, Horsford DJ, McInnes RR(1996): 转录因子基因和眼睛的发育: 一个遗传学观点。人类分子遗传学 5: 1471-1488。

Gehring WJ, Qian YQ, Billeter M, Furukubo-Tokunaga K, Schier A F, Resendez-Perez D, Affolter M, Otting G, Wuthrich K(1994):同源结构域-DNA 识别。细胞 78: 211-223。

Gough NM, Gearing DP, Nicola NA, Baker E, Pritchard M, Callen DF, Sutherland GR(1990): 人 GM-CSF 受体基因在 X-Y 假常染色体区域的定位。自然 345: 734-736。

Grumbach MM, Conte FA(1992): 性别分化紊乱。Williams 内分泌学教程, 第八版, Wilson JD, Foster DW 编, 853-952 页, Philadelphia, WB Saunders.

Hall JG, Gilchrist DM(1990): Turner 综合症及其变型。Pediatr. Clin. North Am. 37:1421-1436.

Henke A, Wapenaar M, van Ommen G-J, Maraschio P, Camerino Q, Rappold GA(1991): 假常染色体区域的缺失帮助定位了三个新标志物以及该缺乏可能对线性生长起到作用。美国人类遗传学杂志 49: 811-819。

Hernandez D, Fisher EMC(1996): Down 综合症遗传学: 揭开一个多因子紊乱的秘密。人类分子遗传学 5: 1411-1416。

Kenyon C(1994): 为什么鸟能飞, 我们不能? 同源基因与进化。细胞 78: 175-180。

5 Krumlauf R(1994): Hox 基因在脊椎动物发育中的作用。细胞 78: 191-201。

Kulharya AS, Roop H, Kukolich MK, Nachtman RG, Belmont JW, Garcia-Heras J(1995): Xpter→Xp22.3 缺失和 3pter→3p23 重复的轻度表型效应。美国医学遗传学杂志 56: 16-21。

10

Lawrence PA, Morata G, (1994): 同源异型框基因: 在果蝇属卵裂和泛型形成中的作用。细胞 78: 181-189。

15 Lehrach H, Drmnac R, Hoheisel JD, Larin Z, Lemon G, Monaco AP, Nizetic D 等, 基因组作图和测序中的杂交指纹。Davies KE, Tilghman S 编, 基因组分析 1990: 39-81, 冷泉港, 纽约。

20 Levilliers J, Quack B, Weissenbach J, Petit C(1989): 人 XY 女性的 X-和 Y-染色体短臂末端部分的交换。美国科学院学报 86: 2296-2300。

Lichter P, Cremer T, 人类细胞遗传学: 一个可行的方法, IRL 1992, Oxford, New York, Tokyo。

25 Lippe BM(1991): Turner 综合症。北美临床内分泌代谢 20: 121-152。

30 Magenis RE, Tochen ML Holahan KP, Carey T, Allen L, Brown MG(1984): Y 染色体短臂部分缺失导致的 Turner 综合症: 男性决定基因的定位。儿科学杂志 105:916-919。

Nelson DL, Ballabio A, Cremers F, Monaco AP, Schlessinger D(1995).
第 6 次国际 X 染色体图谱会议报告。细胞遗传学 71: 308—342。

5 Ogata T, Goodfellow P, Petit C, Aya M, Matsuo N(1992): 一个
DXYS15 远端的 Xp 末端缺失的女孩的矮小身材: 假常染色体区域内的
一个生长基因。医学遗传学杂志 29: 455-459。

10 Ogata T, Tyler-Smith C, Purvis-Smith S, Turner G(1993): Turner 特征
基因在 Yp 染色体上的位置。医学遗传学杂志 30: 918-922。

15 Ogata T, Yoshizawa A, Muroya K, Matsuo N, Fukushima Y, Rappold
GA, Yokoya S(1995): DXYS15 远端的假常染色体区域部分单倍体化女
孩的矮小身材: 假常染色体生长基因关键区域排布的进一步证据。医学
遗传学杂志 32: 831-834。

 Ogata T, Matsuo N(1995): Turner 综合症和女性性染色体失常: 参与
临床特征发育的关键因子的推测。人类遗传学 95: 607-629。

20 Orita M, Suzuki Y, Sekiya T, Hayashi K(1989): PCR 快速灵敏地检
测点突变和多态性。基因组 5: 874-879。

25 Pohlschmidt M, Rappold GA, Krause M, Ahlert D, Hosenfeld D,
Weissenbach J, Gal A(1991): 环形 Y 染色体: 用 DNA 探针确定分子特征。
细胞遗传学 56: 65-68。

 Qiagen(1993)TGGE 手册, Diagen GmbH, TGMA 4112 3/93。

30 Rao E, Weiss B, Mertz A 等(1995): 构建跨越位于假常染色体区域
PAR1 的矮小身材基因候选区域的粘粒连续克隆系。从生命观点看

Turner 综合症：研究和临床发现。第四次国际 Turner 综合症研讨会纪要，Gothenburg, Sweden, 1995 年 5 月 18-21 日。Albertsson-Wikland K, Ranke MB 编辑, 19-24 页, Elsevier.

5 Rao E, Weiss B, Fukami M, Rump A, Niesler B, Mertz A, Muroya K, Binder G, Kirsch S, Winkelmann M, Nordsiek G, Heinrich U, Breuning MH, Ranke MB, Rosenthal A, Ogata T, Rappold GA(1997): 导致原发性矮小身材和 Turner 综合症中生长障碍的包含一个新同源异型框基因的假常染色体缺失。自然遗传学 15: 54-62。

10

Rappold GA(1993): 人类性染色体的假常染色体区域。人类遗传学 92: 315-324。

15

Rappold GA , Willson TA, Henke A, Gough NM(1992): 位于 X-Y 假常染色体区域的人 GM-CSF 受体 α 链基因 CSF2RA 的排列和位置。基因组 14: 455-461。

20

Ried K, Mertz A, Nagaraja R, Trusnich M, Riley J, Anand R, Page D, Lehrach H, Elliso J, Rappold GA(1995): 跨假常染色体区域的酵母人工染色体连续克隆系的特性。基因组 29: 787-792。

25

Robinson A(1990): Turner 综合症的人口统计学和流行率。Turner 综合症, Rosenfeld RG, Grumbach MM 编, 93-100 页, New York, Marcel Dekker

30

Rosenfeld RG(1992): Turner 综合症：医师指南，第二版。Turner 综合症学会。

Rosenfeld RG, Tesch L-G, Rodriguez-Rigau LJ, McCauley E, Albertsson-Wikland K, Asch R, Cara J, Conte F, Hall JG, Lippe B, Nagel

TC, Neely EK, Page DC, Ranke M, Saenger P, Watkins JM, Wilson DM(1994): 对 Turner 综合症个体诊断、治疗和处理的建议。内分泌学家 4 (5) : 351-358。

- 5 Rovescalli AC, Asoh S, Nirenberg M(1996): 四个鼠同源异型框基因的克隆和特性研究。美国科学院学报 93: 10691-10696。

- 10 Schaefer L, Ferrero GB, Grillo A, Bassi MT, Roth EJ, Wapenaar MC, van Ommen GJB, Mohandas TK, Rocchi M, Zoghbi HY, Ballabio A(1993): 人类染色体 Xp22 的高分辨率缺失图谱。自然遗传学 4: 272-279。

 Shalet SM(1993): 生长激素治疗儿童白血病。儿科内分泌学杂志 6: 109-11。

15

 Vimpani GV, Vimpani AF, Lidgard GP, Cameron EHD, Farquhar JW(1977), 严重生长激素缺乏症的流行率。英国医学杂志 2: 427-430。

- 20 Zinn AR, Page DC, Fisher EMC(1993): Turner 综合症: 丢失性染色体的病例。TIG9 (3) : 90-93。

序列表

(1) 一般资料:

(i) 申请人:

- (A) 姓名: Rappold-Hoerbrand, Gudrun, Dr.
(B) 街道: Hausackerweg 14
(C) 城市: Heidelberg
(E) 国家: 德国
(F) 邮政编码: 69118

- (A) 姓名: Rao, Ercole
(B) 街道: Odenwaldstrasse 11
(C) 城市: Riedstadt-Erfelden
(E) 国家: 德国
(F) 邮政编码: 64560

(ii) 发明题目: 人生长基因和身材矮小基因区

(iii) 序列数: 16

(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
(D) 软件: PatentIn Release #1.0, Version #1.30 (EPO)

(vi) 在先申请资料:

- (A) 申请号: US 60/027633
(B) 申请日: 1996 年 10 月 1 日

(vi) 在先申请资料:

- (A) 申请号: EP 97100583.0
(B) 申请日: 1997 年 1 月 16 日

(2) SEQ ID NO: 1 的资料:

(i) 序列特征:

- (A) 长度: 60 个氨基酸
(B) 类型: 氨基酸
(C) 链型: 单链
(D) 拓扑结构: 线性

(ii) 分子类型: 多肽

(xi) 序列描述: SEQ ID NO: 1:

Gln Arg Arg Ser Arg Thr Asn Phe Thr Leu Glu Gln Leu Asn Glu Leu
1 5 10 15
Glu Arg Leu Phe Asp Glu Thr His Tyr Pro Asp Ala Phe Met Arg Glu
20 25 30
Glu Leu Ser Gln Arg Leu Gly Leu Ser Glu Ala Arg Val Gln Val Trp
35 40 45
Phe Gln Asn Arg Arg Ala Lys Cys Arg Lys Gln Glu
50 55 60

(2) SEQ ID NO:2 的资料:

(i) 序列特征:

- (A) 长度: 209 个碱基对
- (B) 类型: 核酸
- (C) 链型: 单链
- (D) 拓扑结构: 线性

(ii) 分子类型: 其他核酸

- (A) 说明: /desc= "外显子 II: ET93"

(v) 片段类型: 线性

(xi) 序列描述: SEQ ID NO:2:

```

GGATTATGA ATGCAAAGAG AAGCGCGAGG ACGTGAAGTC GGAGGACGAG GACGGGCAGA      60
CCAAGCTGAA ACAGAGGCGC AGCCGCACCA ACTTCACGCT GGAGCAGCTG AACGAGCTCG      120
AGCGACTCTT CGACGAGACC CATTACCCCG ACGCCTTCAT GCGCGAGGAG CTCAGCCAGC      180
GCCTGGGGCT CTCCGAGGCG CGCGTGCAAG      209
    
```

(2) SEQ ID NO:3 的资料:

(i) 序列特征:

- (A) 长度: 368 个碱基对
- (B) 类型: 核酸
- (C) 链型: 单链
- (D) 拓扑结构: 线性

(ii) 分子类型: 其他核酸

- (A) 说明: /desc= "外显子 I: G310"

(xi) 序列描述: SEQ ID NO:3:

```

GTGATCCACC CGCGCGCAGG GGCCGTCTTC TCCGCGCGGG GAGACGCGCG CATCCACCAG      60
CCCCGGCTGC TCGCCAGCCC CGGCCCCAGC CATGGAAGAG CTCACGGCTT TTGTATCCAA      120
GTCTTTTGGAC CAGAAAAGCA AGGACGGTAA CGGCGGAGGC GGAGGCGGCG GAGGTAAGAA      180
GGATTCCATT ACGTACCGGG AAGTTTTGGA GAGCGGACTG GCGCGCTCCC GGGAGCTGGG      240
GACGTCGGAT TCCAGCTCC AGGACATCAC GGAGGGCGGC GGCCACTGCC CGGTGCATTT      300
GTTCAAGGAC CACGTAGACA ATGACAAGGA GAAACTGAAA GAATTCGGCA CCGCGAGAGT      360
GGCAGAAG      368
    
```

(2) SEQ ID NO:4 的资料:

(i) 序列特征:

- (A) 长度: 58 个碱基对
- (B) 类型: 核酸
- (C) 链型: 单链
- (D) 拓扑结构: 线性

(ii) 分子类型: 其他核酸

- (A) 说明: /desc= "外显子 III: ET45"

(xi) 序列描述: SEQ ID NO:4:

GTTTGGTTCC AGAACCGGAG AGCCAAGTGC CGCAAACAAG AGAATCAGAT GCATAAAG 58

5 (2) SEQ ID NO:5 的资料:

(i) 序列特征:

(A) 长度: 89 个碱基对

10 (B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑结构: 线性

15 (ii) 分子类型: 其他核酸

(A) 说明: /desc= "外显子 IV: G108"

(xi) 序列描述: SEQ ID NO:5:

20 GCGTCATCTT GGGCACAGCC AACCACCTAG ACGCCTGCCG AGTGGCACCC TACGTCAACA 60

TGGGAGCCTT ACGGATGCCT TTCCAACAG 89

25 (2) SEQ ID NO:6 的资料:

(i) 序列特征:

(A) 长度: 1166 个碱基对

30 (B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑结构: 线性

35 (ii) 分子类型: 其他核酸

(A) 说明: /desc= "外显子: Va"

(xi) 序列描述: SEQ ID NO:6:

40 GTCCAGGCTC AGCTGCAGCT GGAAGGCGTG GCCCAGCGC ACCCGCACCT GCACCCGCAC 60

CTGGCGGCGC ACGCGCCCTA CCTGATGTTT CCCCCGCCGC CCTTCGGGCT GCCCATCGCG 120

45 TCGCTGGCCG AGTCCGCCTC GGCCGCCGCC GTGGTCGCCG CCGCCGCCAA AAGCAACAGC 180

AAGAATTCCA GCATCGCCGA CCTGCGGCTC AAGGCGCGGA AGCAGCGGGA GGCCCTGGGG 240

50 CTCTGACCCG CCGCGCAGCC CCCCCGCCGC CCGGACTCCC GGGCTCCGCG CACCCCGCCT 300

GCACCGCGCG TCCTGCACTC AACCCCGCCT GGAGCTCCTT CCGCGGCCAC CGTGCTCCGG 360

GCACCCCGGG AGCTCCTGCA AGAGGCCTGA GGAAGGAGGC TCCCGGGACC GTCCACGCAC 420

55 GACCCAGCCA GACCCTCGCG GAGATGGTGC AGAAGGCGGA GCGGGTGAGC GGCCGTGCGT 480

CCAGCCCGGG CCTCTCCAAG GCTGCCCGTG CGTCTGGGA CCCTGGAGAA GGGTAAACCC 540

60 CCGCCTGGCT GCGTCTTCCT CTGCTATAAC CTATGCATGC GGTAACTAC ACACGTTTGG 600

AAGATCCTTA GAGTCTATTG AACTGCAAA GATCCCGGAG CTGGTCTCCG ATGAAAATGC 660

CATTTCTTCG TTGCCAACGA TTTCTTTTAC TACCATGCTC CTTCTTCAT CCCGAGAGGC 720

65 TGCGGAACGG GTGTGGATTG GAATGTGGAC TTCGGAATCC CAGGAGGCAG GGGCCGGGCT 780

CTCCTCCACC GCTCCCCCGG AGCCTCCAG GCAGCAATAA GGAAATAGTT CTCTGGCTGA 840

GGCTGAGGAC GTGAACCGCG GGCTTTGGAA AGGGAGGGGA GGGAGACCCG AACCTCCCAC 900
 5 GTTGGGACTC CCACGTTCCG GGGACCTGAA TGAGGACCGA CTTTATAACT TTTCCAGTGT 960
 TTGATTCCCA AATTGGGTCT GGTTTTGTTT TGGATTGGTA TTTTTTTTTT TTTTTTTTTT 1020
 TGCTGTGTTA CAGGATTGAG ACGCAAAGA CTTGCATAAG AGACGGACGC GTGGTTGCAA 1080
 10 GGTGTCATAC TGATATGCAG CATTAACTTT ACTGACATGG AGTGAAGTGC AATATTATAA 1140
 ATATTATAGA TTAATAAAAAA AATAGC 1166

(2) SEQ ID NO:7 的资料:

- 15 (i) 序列特征:
 (A) 长度: 625 个碱基对
 (B) 类型: 核酸
 20 (C) 链型: 单链
 (D) 拓扑结构: 线性
 (ii) 分子类型: 其他核酸
 25 (A)说明: /desc= "外显子: Vb"

(xi) 序列描述: SEQ ID NO:7:

30 ATGGAGTTTT GCTCTTGTCG CCCAGGCTGG AGTATAATGG CATGATCTCG ACTCACTGCA 60
 ACCTCCGCCT CCCGAGTTCA AGCGATTCTC CTGCCTCAGC CTCCCGAGTA GCTGGGATTA 120
 CAGGTGCCCA CCACCATGTC AAGATAATGT TTGTATTTTC AGTAGAGATG GGGTTTGACC 180
 35 ATGTTGGCCA GGCTGGTCTC GAACTCCTGA CCTCAGGTGA TCCACCCGCC TTAGCCTCCC 240
 AAAGTGCTGG GATGACAGGC GTGAGCCCCT GCGCCCGGCC TTTGTAACTT TATTTTTAAT 300
 40 TTTTTTTTTT TTTTAAGAAA GACAGAGTCT TGCTCTGTCA CCCAGGCTGG AGCACACTGG 360
 TGCGATCATA GCTCACTGCA GCCTCAAACCT CCTGGGCTCA AGCAATCCTC CCACCTCAGC 420
 CTCCTGAGTA GCTGGGACTA CAGGCACCCA CCACCACACC CAGCTAATTT TTTTGATTTT 480
 45 TACTAGAGAC GGGATCTTGC TTTGCTGCTG AGGCTGGTCT TGAGCTCCTG AGCTCCAAAG 540
 ATCCTCTCAC CTCCACCTCC CAAAGTGTTA GAATTACAAG CATGAACCAC TGCCCGTGGT 600
 50 CTCCAAAAAA AGGACTGTTA CGTGG 625

(2) SEQ ID NO:8 的资料:

- 55 (i) 序列特征:
 (A) 长度: 15577 个碱基对
 (B) 类型: 核酸
 (C) 链型: 单链
 (D) 拓扑结构: 线性
 (ii) 分子类型: 其他核酸
 (A)说明: /desc= "外显子: HOX93"

- 65 (ix) 特征:
 (A) 名字/关键词: 外显子
 (B) 位置: 1498--1807
 (D) 其他信息: /功能= "外显子 I (G310)的一部分"

- (ix) 特征:
 (A) 名字/关键词: misc-特征
 (B) 位置: 3844--4068
 (D) 其他信息: /功能= "pET92 区 (第一部分)"
- 5 (ix) 特征:
 (A) 名字/关键词: misc-特征
 (B) 位置: 4326--4437
 (D) 其他信息: /功能= "pET92 区 (第二部分)"
- 10 (ix) 特征:
 (A) 名字/关键词: misc-特征
 (B) 位置: 4545--4619
 (D) 其他信息: /功能= "pET92 区 (第三部分)"
- 15 (ix) 特征:
 (A) 名字/关键词: 外显子
 (B) 位置: 5305--5512
 (D) 其他信息: /功能= "外显子 II (ET93) 的一部分"
- 20 (ix) 特征:
 (A) 名字/关键词: 外显子
 (B) 位置: 11620--11729
 (D) 其他信息: /功能= "外显子 IV (G108) 的一部分"
- 25 (ix) 特征:
 (A) 名字/关键词: 外显子
 (B) 位置: 11620--11729
 (D) 其他信息: /功能= "外显子 IV (G108) 的一部分"
- 30 (ix) 特征:
 (A) 名字/关键词: 外显子
 (B) 位置: 11620--11729
 (D) 其他信息: /功能= "外显子 IV (G108) 的一部分"

(xi) 序列描述: SEQ ID NO:8:

CTCTCCCTGT TGTGTCTCTC TTTCTCTCTC TCCATCTCTC TCCGTCTTTC CCCCTCTGTC	60
35 TCTTTCTCTG TCTCCATCCC TCTGTCTCTC CCTTTCTCTC TGTCTTTTCT TGTCTCTCTC	120
TTTCTCTCTC TCTCTCCATC TCTCTCTCTC CCGGTCTCTC TCTCTCCATC TCCCCGTCTC	180
40 TCCGTTTCTC TCTCTGCCTC TCCCTGTCTG TCTCTCTCTT TGTGTGTGTT ACACACACCC	240
CAACCCACCG TCACTCATGT CCCCCACTG CTGTGCCATC TCACACAAGT TCACAGCTCA	300
GCTGTCTATC TGGGTCCCCA GGCCCCGCG GGGAGGAAGA TGCGCCGTGG GGTACGGGA	360
45 GGAAGGGGAC TCCGGGCCTC CTGGTGCCCC ACTTTATTG CAGAAGGTCC TTGGCAGGAA	420
CCGTGACGCG TTTGGTTTCC AGGACTTGGA AAACGAATTT CAGGTCGCGA TGGCGAGCAC	480
50 CGGCTTCCCC TGAAGCACAT TCAATAGCGA GAGGCGGGAG GGAGCGAGCA GGAGCATCCC	540
ACCATGAAAA CCAAAAACAC AAGTATTTTT TTCACCCGGT AAATACCCCA GACGCCAGGG	600
TGACAGCGCG GCGCTAAGGG AGGAGGCCTC GCGCCGGGGT CCGCCGGGAT CTGGCGCGGG	660
55 CGGAAAGAAT ATAGATCTTT ACGAACCAGG TCTCCCGGG ACCTGGGCTT CTTTCTGCGG	720
GCGCTGGAAA CCCGGGAGGC GGCCCCGGGG ATCCTCGGCC TCCGCCGCCG CCGCCTCCCA	780
60 AGCGCCCGCG TCCCGGTTTG GGGACACCCG GCCCCCTCTT CTCACCTTCG GGGATTCTCC	840
AGCCGCGTTC CATCTACCA ACTCTCCATC CAAGGGCGCG CCGCCACCAA CTTGGAGCTC	900
ATCTTCTCCC AAAATCGTGC GTCCCCGGGG CGCCCGGGTC CCCCCCTCG CCATCTCAAC	960
65 CCCGGCGCGA CCCGGGCGCT TCCTGGAAAG ATCCAGGCGC CGGGCTCTGC GCTCCTCCCG	1020
GGAGCGAGGG CGGCCGACA ACTGGGACCC TCCTCTCTCC AGCCGTGAAC TCCTTGCTCTC	1080

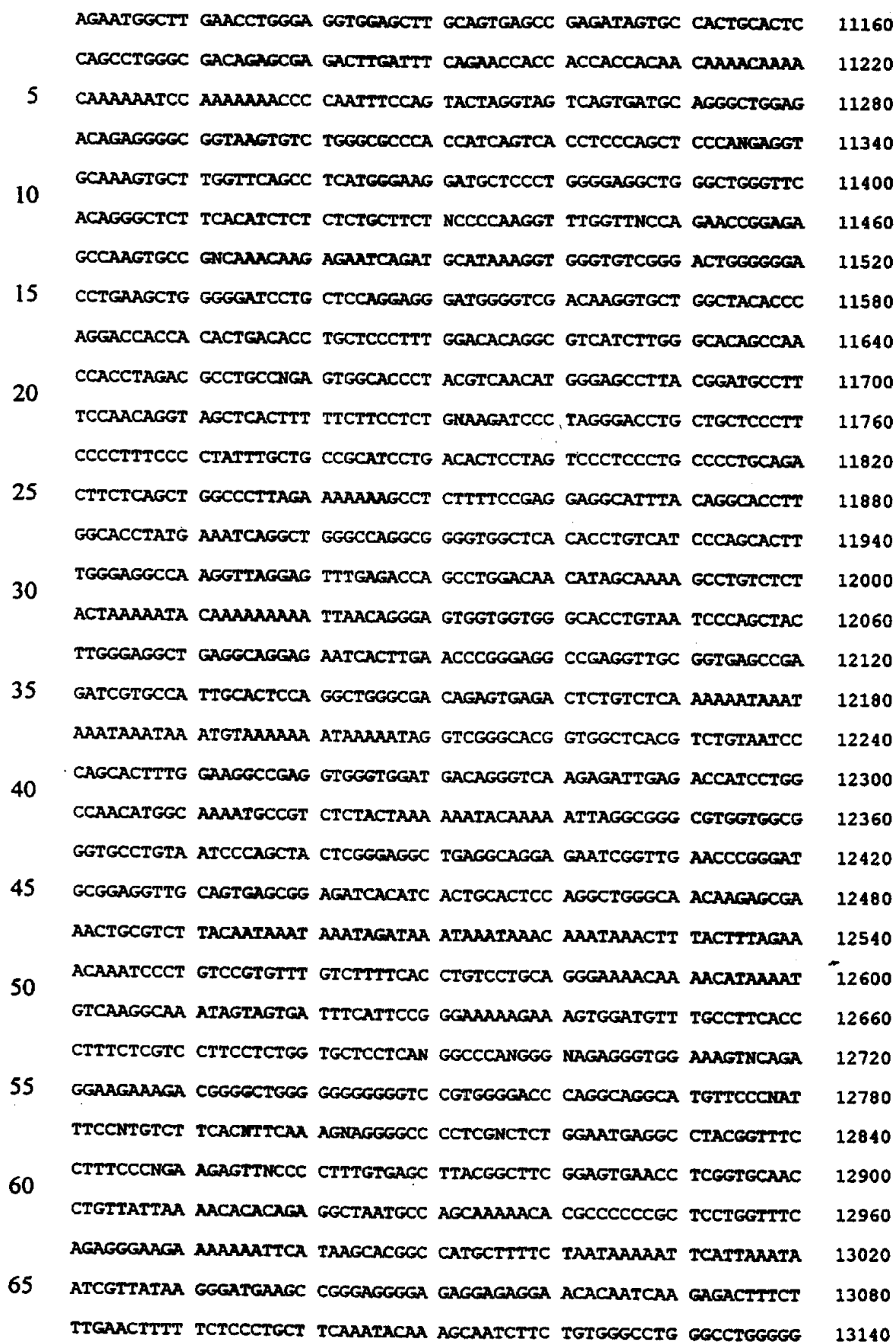
	TCTGTCTCTC TCTGCAGGAA AACTGGAGTT TGCTTTTCCT CCGGCCACGG AAAGAACGCG	1140
5	GGTAACCTGT GTGGGGGGCT CGGGCGCCTG CGCCCCCTC CTGCGCGCGC GCTCTCCCTT	1200
	CCAAAAATGG GATCTTTCCC CCTTCGCACC AAGGTGTACG GACGCCAAAC AGTGATGAAA	1260
	TGAGAAGAAA GCCAATTGCC GGCCTGGGGG GTGGGGGAGA CACAGCGTCT CTGCGTGCCT	1320
10	CCGCCCGCGA GCGCGGAGAC CAGTAATTGC ACCAGACAGG CAGCGCATGG GGGGCTGGGC	1380
	GAGGTGCGCG CGTATAAATA GTGAGATTTC CAATGGAAAG GCGTAAATAA CAGCGCTGGT	1440
15	GATCCACCCG CGCGCACGGG CCGTCCTCTC CGCGCGGGGA GACGCGCGCA TCCACCAGCC	1500
	CCGGCTGCTC GCCAGCCCCG GCGCCAGCCA TGGAAGAGCT CACGGCTTTT GTATCCAAGT	1560
	CTTTTGACCA GAAAAGCAAG GACGGTAACG GCGGAGGCGG AGGCGGCGGA GGTAGAAGG	1620
20	ATTCCATTAC GTACCGGGAA GTTTTGGAGA GCGGACTGGC GCGCTCCCGG GAGCTGGGGA	1680
	CGTCGGATTG CAGCCTCCAG GACATCACGG AGGGCGGCGG CCACTGCCCC GTGCATTGT	1740
25	TCAAGGACCA CGTAGACAAT GACAAGGAGA AACTGAAAGA ATTCGGCACC GCGAGAGTGG	1800
	CAGAAGGTAA GTTCCTTTGC GCGCCGGCTC CAGGGGGGCC CTCCTGGGGT TCGGCGCCTC	1860
	CTCGCCACGG AGTCGGCCCC GCGCGCCCCT CGCTGTGCAC ATTTGCAGCT CCCGTCTCGC	1920
30	CAGGGTAAGG CCCGGGCCGT CAGGCTTTGC CTAAGAAAGG AAGGAAGGCA GGAGTGGACC	1980
	CGACCCGAGA CGCGGGTGGT GGGTAGCGGG GTGCGGGGGG ACCCAGGGAG GGTGCGACGG	2040
35	GGGGCCGCGC GCGTGGGCAC CGACACGGGA AGGTCCCAGG CTGGGGTGGG TCCGGGTGGC	2100
	TGTGCTGAA GCCGTAGGGC CTGAGATGTC TTTTTCATT TCTTTTCTT TCCTTTCCTT	2160
	TTTTTGTTG TTTGTTTGTG TGTTTGAGAC AGAGTCTCGC TCTGTCCCCC AGGCTGGAGT	2220
40	GCAGTGGTGC GATCTCGGCT CACTGCAACC TCCGCCTCCT GGGTTCAAGC GATTCTCCTG	2280
	CCTCAGCCTC CCCAGTAGCT GGGATTACAG GCATGCACCA CCACGCCTGG CTAATTTTGT	2340
45	TGCTTTTAGT AAAGACGGGG ATTCAACATG TTGGCCAGGC TGGTCTCGAA CTCCTGACCT	2400
	CAGGTGATCC ACCCGCCTCG GCCTCCCAAA GTGCTGGGAT GACAGGCGTG AGGCACCGCG	2460
	CCCGGCCTGG GTCCTGACGG CTTAGGATGT GTGTTTCTGT CTCTGCCTGT CTGCCTGTGA	2520
50	TTTACGGTCA CCCAGACGCA CAGAGGAGCC GTCTCCACGC GCCTTCCCAG CGCTCAGCGC	2580
	CTGCCGGGCC CCCGAGATC ACGGGAAGAC TCGAGGCTGC GTGGTAGGAG ACGGGAAGGC	2640
55	CCCGGGTCAG CTCGGTTCTG TTTNCNTTA AGGAACCCCTT CATTATTATT TCATTGTTTT	2700
	CCTTTGAACG TCGAGGCTTG ATCTTGGCGA AAGCTGTTGG GTCCATAAAA ACCACTCCCC	2760
	TGAGCGGAGG TGGCCGGGAT CTGGATGGGG CGCGAGGGGC CCCGGGGAAG CTGGCGGCTT	2820
60	CGCGGGCGCG TCCTAAGTCA AGGTGTGAG AGCGCAGCCG GTTGTGCGCG GCCCGGGGNN	2880
	AGCTCCCCCTC TGGCCCTTCC TCCTGAGACC TCAGTGGTGG GTCGTCCCGT GGTGGAAATC	2940
65	GGGGAGTAAG AGGCTCAGAG AGAGGGGCTG GCGCGGGGGA TCTCTGTGCA CACACGACAA	3000
	CTGGGCGGCA TACATCTTAA GAATAAAATG GGCTGGCTGT GTCGGGGCAC AGCTGGAGAC	3060

	GGCTATGGAC GCCTGTTATG TTTTCATTAC AAAGACGCAG AGAATCTAGC CTCGGCTTTT	3120
	GCTGATTTCGC AAAGTTGAGG TGCGAGGGTG AATGCCCCAA AGGTAATTCT TCCTAAGACT	3180
5	CTGGGGCTAC CTGCTCTCCG GGGCCCTGCA TTTGGGGTGT GGAGTGGCCC CGGGAAATAG	3240
	CCCTTGTTATT CGTAGGAGGC ACCAGGCAGC TTCCCAAGGC CCTGACTTTG TCGAAGCAGA	3300
10	AAGCTGTGGC TACGGTTTAC AAAGCAGTCC CCGGTTTCTG ACCGTCTAAG AGGCAGGAGC	3360
	CCAGCCTGCC TTTGACAGTG AGAGGAGTTC CTCCCTACAC ACTGCTGCGG GCACCCGGCA	3420
	CTGTAATTCA TACACAGAGA GTTGGCCTTC CTGGACGCAA GGCTGGGAGC CGCTTGAGGG	3480
15	CCTGCGTGTA ATTTAAGAGG GTTCGCANGC GCGCGGCGGC CGCTTCTGNT GGGGTTGCTT	3540
	TTTGGTTGTC CTTGNGCAA CACCGTTTTG CTCCTCTNGN AACTCTCTCT TNCTCCCCCN	3600
20	TGGCCNGTNG GACCCGGGNA NGAGCAAAGT GTCCTCCAGA CCNTTTTGAA ANGTGAGAGG	3660
	AAAATAAAGA CCAGGCCAAA NNGACCCAGG GCCACAGGAG AGGAGACAGA GAGTCCCCGT	3720
	TACATTTTNC CCCTTGGCTG GGTGCAGAAA GACCCCGGG CCAGGACTGC CACCCAGGCT	3780
25	ACTATTTATT CATCAGATCC AAGTTAAATC GAGGTTGGAG GGCAGGGGAG AGTCTGAGGT	3840
	TACCGTGGAA GCCTGGAGTT TTTGGGNAAC AGCGTGTCCC CGCCGAGCCT GGGAGCCCGT	3900
30	GGGTTCTGCA AAGCCTGCGG GTGTTTGAGG ACTTTGAAGA CCAGTTTGTC AGTTGGGCTC	3960
	AATTNCCTGG GGTTCAGACT TAGAGAAATG AAGGAGGGAG AGCTGGGGTC GTCTCCAGGA	4020
	AACGATTCAC TTGGGGGGAA GGAATGGAGT GTTCTTGCA GGCATGTCT GTTAGGAGGT	4080
35	GAAACAGAA GTGAAATCCA CGTTGGAGTA AGCGTCCAGC GCTGAATGTA GCTCGGGGTG	4140
	GGGTGGGAGG GCCCTGGTGT GGATCGTGGA AGGNAAGAAA GACAGAACAG GGTGCTAGTA	4200
40	TTTACCCCGT TNCCCTGTAG ACACCCTGGA TTTGTCAGCT TTGCAAGCTT CTTGGTTGCA	4260
	GCGGCCTTGC CTGTGCCCCT TTGAGACTGT TTCCAGACTA AACTTCCAAA TGTCAGCCCC	4320
	TTACCCTTGA CAGCAAGGGA CATCTCATT GGGCATCGCG TGCTTCTCAT CTGTGNTCA	4380
45	GCAGGCCCNG AGATAGGAAN CANGAGGGGC NGTTGGNAGA TGCNCACTTC CACCAGCCCT	4440
	GGGNTTGAAG GGGANGCGAN GGGANGACNA CCTTTTANCT TAAACCCCTN GAGCTTGGTN	4500
50	CAGAGAGGNC TGAATGTCTA AAATGAGGAA GAAAGGTTT TTCACCTGGA AACGCTTGAG	4560
	GGCTGAGTCT TCTGCCNTT CTGACNTCCC CCAGCAAATA CAGACAGGTC ACCAANCTAC	4620
	TGGAGATGAG AAAGTGCCAT TTTTGGCACA CTCGTGTGGG GTAGGTGCCC GACCGCGTGT	4680
55	GAAAAANGTG GGAANNGGAG AGATTCTGN CGCACGGGT TCAGCCCCCA GGCGGGNTG	4740
	GCNGCATTN AGGNTACTCA GACGCGGTTC TGCTGTTCTG CTGAGAAACA GGCTTCGGGT	4800
60	AGGGGCTCCT AGCTCCGCCA GATCGCGGAG GGACCCCCAG CCCTCCTGCG CTGCAGCGGT	4860
	GGGGATAGCG TCTCTCCGTA GGCCTAGAA CTGCAACCCG CCCCAGGTCC TCCCCGTGTC	4920
	CTTCCCGGGC GTCCCGCCGG GGATCCCA CA GTTGGCAGCT CTTCTCAAA TTCTTTCCCT	4980
65	TAAAAATAGG ATTTGACACC CCACTCTCCT TAAAAAATA AAATAAGAAA AAAAGGTTAG	5040
	GTTATGTCAA CAGAGGTGAA GTGGATAATT GAGGAACGA TTCTGAGATG AGGCCAAGAA	5100

	AACAACGCTC GTGCAAAGCC CAGGTTTTTG GGAAAGCAGC GAGTATCCTC CTCGGCTTTT	5160
5	GCGTTATGGA CCCACGCAG TTTTTCGTC AAAGCGCATT GGTTCGAG GGCCCCCTTT	5220
	CCACCGCGGG ATGCACGAAG GGGTTCGCCA CGTTGCGCAA AACCTCCCCG GCCTCAGCCC	5280
	TGTGCCCTCC GCTCCCCACG CAGGGATTTA TGAATGCAAA GAGAAGCGCG AGGACGTGAA	5340
10	GTCGGAGGAC GAGGACGGGC AGACCAAGCT GAAACAGAGG CGCAGCCGCA CCAACTTCAC	5400
	GCTGGAGCAG CTGAACGAGC TCGAGCGACT TTTTGACGAG ACCCATTACC CCGACGCCTT	5460
15	CATGCGCGAG GAGCTCAGCC AGCGCCTGGG GCTTTCGAG GCGCGCGTGC AGGTAGGAAC	5520
	CCGGGGGCGG GGGCGGGGGG CCCGGAGCCA TCGCCTGGTC CTCGGGAGCG CACAGCACGC	5580
	GTACAGCCAC CTGCGCCCGG GCCGCCCGG TCCCCTTCCC GGAGCGCGGG GAGGTTGGGT	5640
20	GAGGGACGGG CTGGGGTTCC TGGACTTTTG GAGACGCCTG AGGCCTGTAG GATGGGTTCA	5700
	TTGCGTTTGT TTTTCACCAA CAGCAAACAA ATATATATAC ATATATATTA TACAAATAAC	5760
25	AAATAAATAT ATATGTTATA CAGATGGGTA TATTGTATAT ATTATAGATA TTTGTTTCGTC	5820
	CTTGGTGCAA AGACACCCGG TGAACCCATA TATTGGCTCC TGA CTGCCTT CGGTTCCCCT	5880
	GGGATTGGTT ATAGGGGCAA CACATGCAAA CAAACTTTC CCTGGATTAT ACTTAGGAGA	5940
30	CGAAGCTACA GATGCGTTTG ATCCAGAGTG TTTTACAAGA TTTTTCATT AAAAAAAAAAT	6000
	GTGTCTTTTG GCCCTGATT CCCCTCCGTC TTCCCGTGTG GCTGCATTGA AAAGGTTTCC	6060
35	TTAGGATGAA AGGAGAGGGG TGTCTCTGT CCCTAGGTGG AGAGAAACAG GGTCTTCTCT	6120
	TTCCTCCGTT TTTTCACCTA CCGTTTCTAT CTCCCTCCTC CCCTCTCCAG CCCTGTCCTC	6180
	TGCTACAAAC CACCCCTCC TCCCTCCGGC TGTGGGAGC GCAGGAGCAC GTTGGGCATC	6240
40	TGGATGAGCG GNAGACTATT AGCGGGGCAC GGGGGCTCCC CGAGGAGCGC GCGAATTCAC	6300
	GCTGCCCCAT GAGACCAGGC ACCGGGGGGC GGAGGGGCCT TGGGTGTCCG CAGAGGGACG	6360
45	GGCGGGCAGA GCCTTCCTCC GCATTCTAAA CATTCACTTA AAGGTATGAG TTTANTTTCA	6420
	GGGGTGCTGC TGGGAGAGCC TCCAAATGGC TTCTTCCAGC CCCTGCCTGA CAGTTCAGCT	6480
	CCCCTGGAAG GTCAACTCCT CTAGTCCTTT CTCTGGTTC TGGGCAGGAC AGAAGTGGGG	6540
50	GGAGGGAGAG AGAGAGAGAG AGAGAGAGAG ACGGTCAGGA TCCCCGACC CTGGGGAACC	6600
	CGTCAAAAAT AAATGAAAT AAGATTGCCG ACCAGAGAGA GAACCGTGAC AAAGCAAACG	6660
55	GCGTTCAAAG CAAAGAGACG AACTGAAAGC CCGTTCCTGT AGGACTGGTT ATGAGGTCAA	6720
	CACATTCAAA CACAGCTTGC TCTGGATTTT GCTGAGCAGA GGAAGATACA GATGCATTG	6780
	ATCCAAAGTG TGTACATCT TTCATTATAT GTGTGTCTAT ATATATAAAC ATATATAAAT	6840
60	ATATAACAT ACATAAATGT ATGTAATAT ATATAATCTA TATACATATA TAAATATATA	6900
	AACACATATA TAATATATAA ATCTATAAAC ATATATAATA TATAACATA AATATATAAA	6960
	CATATATAAT ATATAAATAT ATTAACATAT ATAAATATG TATAAATATA TATAACATA	7020
65	TAAACATATA TAAATATATA AACATATAAA TATATAAACA TATATAAATA TATACAAACA	7080

	TATTGTATAT ATATAAATAT ATATAAAAAC ATATATATAC ATATAAAAAT ATATATAAAC	7140
	ATATATACAT ATAAAGAAAT ATATATAAAC ATATATACAT ATAAATATAC ATATATAAAC	7200
5	ATATATATAC ATAAATATA TATAAACATA TATACATATA AAAATATATA TATATTAACA	7260
	TATATATACA TATAAAAATA TATATATTAA CATATATATA CATATAAAAA TATATATATA	7320
10	TTTTTGGCCC CTGATTCCTT TCGGTTCTTG TGGGATGGGT GATTGAGTCA ACACATTCAA	7380
	ACACAACTTT TCCATCGATG TTGCTTAGGA GATGAGGATA CAGATGCGTT TGATGGAGAG	7440
	GGTTTTACAA GCTCTTTCAT TTAAATATAT ATATATATAT ATATATATTT TTTGGCTCCT	7500
15	GATTCTCTTC CGTCTTCCCA TGTGGCTGCA TTTTAAAGG CTCCCTAAG ATCGTTACGA	7560
	TTAAATCAAC CCTCCCCAGG CATCTTTACC GAGGGCTGTG GTCCCCAAAG CGATACAGCC	7620
20	CAGGAGGGAG AGAGGCTTTG GTGACTTGGA GGAAGGACTG TGTCCTCCTT TAGGGCGTCT	7680
	GTGGCCTCAG TGAGGGAAGG AAGCTGCATC AGACAGGGGT TTCCTCGCTG TCCACCCCTC	7740
	TGGCAGAAGA TGGATTGGGC TGCCCCGNTA TAAATTAATG AAAAGATTAA AGTTTCGCTA	7800
25	AAGGGGACAT CGAGTTTATG TGTCTCTCC TGGTGTCTG TGTGCCNTGG GATNCTGCAA	7860
	TATATCCCAN NGCCCTTGAT GNNNTACTGT TTNCTATAAA AANNTAAATN TACTTGTTNA	7920
30	ATTTAANTTC CNNNACACTA TTTNCTTTCC NNGTNAGTCT NATTANCCGA NCGAGAGCAN	7980
	CGNTTAGTTN CAGCTNGCGG AAAATTGGTT GTGGGGTGTG TGCGGACCCC NGAGNAACGC	8040
	CCNNTAAAT NAAAGACAAA NTCNGGGGAC AAGNCTNGGG GGTATCGNN ATTGCNNAGG	8100
35	GGTCGNCATG AAAANTTTAA CGACGGTAAA TAATAATAAA AANNCAAACA TGGGAATGNC	8160
	AATAAAGAC ATAATTCTCC NNATCGCCGC GGGGGGAAAG GATCCTATAG TAAAGGCGAG	8220
40	TGCGCTTTGA GGGGTCATAA AAATCAATTA GTTCCAACAC CCACGTCCCG CGTTGAGGGG	8280
	ACGGGGACGA GCAGGGACAG AAAAAGAAAC CATATTTGAA TCCCATCTCT CTGTGAATTC	8340
	TTGGGTCACA TGGCTCTCAG TACAGCCCGT CCCGTGCTGT GACCGGATAG AGTTTCAATT	8400
45	TACTGTGGAA ATTTGCTGTA AATAAATTGA GCATCCGATA GAAGCTGTTG CTGATTAACC	8460
	TTTTATTTT AGCGTGGCCC TGCAAAGTCG TATCACCCAG CTGTCAGGCT TCTAATCGAA	8520
50	AGTTATGAGA CCACGGTGAG GGGCAGGCGG TAATTTAATT ACAACAAATA TCTTTGGGTT	8580
	TATGGCGCAG AGCTAAATTA AATGTCATTA TTCACTGTCT GTNAATGGNA AATCAAAANN	8640
	GGAAATCGCA NTTACGGNCA TTTGGGNNA ANGAAAGCGG GGNAGTGCTC TTTAATNGAA	8700
55	NNGAAATAAC TGTCTTAAGC AGTGTCACAC ACTTCACTTA CCATATTTCGN GGCCTNAATT	8760
	GGAANNITGA TCGTNNGAAT CACTCCNAAG ACTNGATTTA TTANGCGCTT CACGNCAGCN	8820
60	NGGCNTAATT CATCNACTTN NGTATTCTTC ATCNNNNATT TTTTTTTTTT CTCTCNNGCC	8880
	GTGTTNNGAA GGGAGAGTGA ATGAGGCTTT CCACGTTTCA GGAGGATTTT CTTTTTTGAA	8940
	AAATGCCCTT CCAGAGGCTT TTGGGTGGCT GGCTTGCTTT CTGGGCCCTG GAGGANGACA	9000
65	GGCGGANGAG TCCAGGTGGG CATGGAGAGG CACAGTGGCA GGTCACCTGG ATGGTCAGTG	9060
	GAGGTGGAGG TCTGAAGGCG CCAGCTTTGG AAATTATTGG TGAATTTCTGA TGTCAGCACC	9120

	AGGNCAGGGG	CCTTTTGGC	GGGGGTGTGA	GGGANGGATG	ANCTTTGCTG	GGAAANNACG	9180
5	GATCAGGTTT	TCCAGGCGCA	CTGCAGCCCG	GTAGGACCCA	CTTTGGAAAT	GAAAAGCCAG	9240
	TNCCGAAAG	CTGGGCTGGA	AGCTTCCGTG	TTGGGTTCAT	GAGCAAGTTC	ACGTTGCGCT	9300
	GTGTAGACTC	CTGGCTGCTC	CCAAACTCTG	AGGGTTTTCT	GAGGTTCCCT	TCATAGGGGC	9360
10	ACCGGCCCTG	GGCCATGCAC	AGTGCGTAAG	GGTGGCTGTG	GGCCGAGGGA	CCCAGCACGT	9420
	GTTTTGCCCA	CAACAGCCGG	AGTGACTGGT	TCACTCACCG	CCTTGGCGGA	GGACGCCTGT	9480
15	TCTCTGGACG	AATCATTTCT	CTTGGGTGGT	GACTGCCTTG	TGGGTCAAGG	TGCAGGTTTT	9540
	CTGCCACAGA	AAACCTGTTA	GGAGGAATTA	AGCGACTAAG	ACTGTCAGGG	AGGTGGTGGT	9600
	GGGGGANGAG	GNAGGGGGTG	GTGTCCAGAT	TACCAGGCAT	AGGCTAAACT	GCCTGCACTC	9660
20	TCCAGCTGGT	CTGTCTGTGG	AGGAGGGGAT	TGTCAATACT	GGGAGAGCAG	AGGAGGCTCG	9720
	TAGGAGGTGA	GAGGGGGTGG	AATTTGCATG	CAAATCTTCA	CATGAGGCCT	GTGTGAATTT	9780
25	CTCCAGCCTC	CTGAGGGTCC	CCTGCGCTAT	TGCACTCAAC	TTCTTGATAG	TTTACCCCAA	9840
	GACTCAGAAG	TCCTTAGAGG	GGCAGAATGC	CCCCACCACA	AAGCCTGCTA	TCCTTGGGCG	9900
	TCCTCAGGAC	CCTTGGTCAT	GAATGGGACC	CTTTCATGTA	TGGGGACCCT	TGGTAATATG	9960
30	AATGGGACGC	CTTCAGCTCC	CCAGGGCTTC	CGAGGAGGCC	GAGAAGGGCA	AAGACACTTC	10020
	CGAGGAGGCC	GAGAAGGGCA	AAGACATTTT	CTGGGCTTGG	TGTGTCAAGA	GCTAGATTGG	10080
35	AGAAGGGGCT	GGATTTGGA	CTCTTTAGCC	ATCAGCTCAC	CCTCTCCGTT	TGTGGCTAAA	10140
	GTCTGAAGGT	GGAAACTTCG	GTTCTCCTAC	AGGGTCTACA	GGAGTTGGGG	GGCGGGGCGC	10200
	CCACACAGAA	CGCTGGAAAG	TTGACAGTC	CACTTCCACT	GGCTCGGAAC	TCACTTTTTC	10260
40	ACCTTAAGTT	CATCAGCGGT	AACGCATAGG	TCTCACTTAG	GCAGGGCAGC	GATGATTAA	10320
	CAATTTCTAC	TTCTAGGTCA	GGTGCGGTGG	CTCACACCTC	TAATCCCAGC	ACTTTGGGAG	10380
45	GCCCAGGAGG	GTGGATCGCT	TGAGGTCAGG	AGTTTGAGAC	CAGCCTGGCC	AACATGGTGA	10440
	AACCCCGTCT	CTACTAAAAT	ACGAAAATTA	GCCAGGCATG	GTGGTGAGCA	CCTGTAATTC	10500
	CAGCTACTCG	GGAGGCTGAG	GCAGGAGAAT	CGCTTGAACC	TGGGAGGTGG	ACGTTGCAGT	10560
50	GAGGTGAGAT	CACACCACTG	CACTCCAGCC	TGGATGAGAG	AGCAAGACTC	TGTCTCAAAA	10620
	ACAAAATAAA	ACAAAAACAA	AACAAAAATC	AAAAAAGAAA	ACCCAATTTC	CAGTTCTAGG	10680
55	CCAGGTGCAG	TGGCTCACGC	CTGTCATCCC	AGCACTTTGG	GAGGCCCAGG	AGGGTGGATC	10740
	GCTTGAGGTC	AGGAGTTTGA	GACCAGCCTG	GCCAACATGG	TGAAACCCCA	TCTTTACTAA	10800
	AAATACAAAC	GTTAGCTGGG	TGTGGTGGTG	TGCGCCTGTA	ATCCCAGCTA	CTCGGGAAGC	10860
60	TGAGGCTGGA	GAATTGCTTG	AATCTGGGAG	GTGGAGGTTG	CAGGGAGGCG	AGATAGTGCC	10920
	ACTGCAGTCC	AGCCTGGACC	AGAGAGCAAG	ACTCCGTCTC	AAAAACAAAA	GAAAGCAAAA	10980
65	ACAAAAACA	AGAGACCAGC	CTGGCCAACA	TGGTGAAACC	GCGTCTTTAC	TAAATACAA	11040
	AATTAGCCGG	GCATGGTGGT	GGGCACCTGT	AGTCCCAGCT	ACTCGGGAGG	CTGAGGCAGG	11100



	GTTCCTCCCT	TTCTCTGCAG	CCCATTGGGA	GGAAGAAAAT	GCTTCCCTGA	ANGTTGCTGC	13200
5	AAAATTGTTT	CTGTTTTTCT	TTCTTTTTTC	TTTTTTTTTT	TTTTTTGAGA	CGGAGTCTCG	13260
	CTCTGTCACC	AGGCTGGAGT	GCAATGGTAT	GATCTCAGCT	CACTGCAACC	TCCACGTTCC	13320
	TGTTTCAAGT	CATTCTCCTG	CCTCAGCCTC	CTGAGTAGCT	GGGACTACAG	GCGCCCGCCA	13380
10	CCACGCCCCG	CTAGTGTGTT	TATTTTTAGA	AAAGACAGGG	TTTCCCCATG	TTGGCCAGGC	13440
	TGGTCTTGAA	CTCCTGTCCT	CAAGTGATCT	GCCTGCCTCG	GCCTCCCAAA	GTGCTGTGTT	13500
15	TCTGTTTTTC	TTCCCCGCT	TTCTTAGGAG	GCCATCGGGA	AGAATAAAAT	GCTTTCCTTG	13560
	AAGTTGATGC	AAAATTGTTT	CTGTTTTTCT	TTTCTCTTTT	CTTTCTTTTT	GAGATGGAGT	13620
	CTCGCTCTTT	CACCCAGGCT	GGAGGGCAGT	GGCGCGACCT	CGGCTCACTG	CAACCTCCGC	13680
20	CTCCCCGGGT	CAAGCGATTG	TCCTGCCTCA	GCCTCCGGAG	TAGCTGGGAT	TACAGGCACC	13740
	TGCCACTATG	CCTGGCTAAT	TTTATTATTT	TTAGTAGAGA	CGGGGTTTCA	CCATGTTGGC	13800
25	CAGGCTGGTC	TCAAACCTCT	GACCTCAGGT	GATCCGCCCG	CCTCGCCTCC	CAAAGTGATG	13860
	GGATGANAG	GNCATNGAGC	NCACCGTGCC	CGGCCCTCTA	ACTCTTTACC	AGACATAAAG	13920
	TCTCCNNTTC	CCCTTTCTAA	ATGTATATAT	TGTGTTTTTA	AAAGTTAACA	GCAGGGATCC	13980
30	CACCTCATTN	CCCCGCTNCT	CTCCCCAAGA	CCTGTCCTGC	ACGTTGCACA	CAGCAGGTGT	14040
	GCCCTGGACA	TATCCCAAAC	CCACGCTGAA	AGAAAGAGGG	TCTCACTACA	CGTATGATAT	14100
35	CTGTGNATCC	TTTAAACATC	TCCGTGGCTT	CCAGGCAACA	CAGCCATAAA	TAGGAATCTC	14160
	ATGTCTGACA	TGATACCGGG	ACCATGTATG	GGNAAATTCT	GGGTGTGAAG	TTCCAGCTAC	14220
	CCCCGCAGAG	GCANCCATTG	CATACCCTCC	AGAAACTCCC	CTGCCGTTNC	AAGCCAAAGA	14280
40	CACAACACAA	ACAGCNTCCG	AGAGAGGGTG	TCATTGAAAA	TCAATACCAT	CATAAGAGCA	14340
	CACAGCACCG	TCTTCTCTT	CTGCCCCTTG	ATACACAATT	ATGAGCAATT	TGCTAACACT	14400
45	GACAACTCGT	GGCAAGAACA	GGTCGTGTTG	ATACGGTTGC	CTCGTGAGGA	CCCATCTGTC	14460
	TTCTGGGGTC	TTGCCTGGAA	CGGAGATCGG	AGTTCAGGGT	GGCTAATAGA	ATCATTACTC	14520
	ACCTAGGGAC	ACAGAAATNAT	GAGGGTTACC	CCCAGTTAAG	TGCATACAGT	CAAACGGACG	14580
50	GCTGCTCTGG	AAGGTACAGT	GACGTGAACA	GCTTTTATGA	AATGCCTAGA	TCTGGACCTT	14640
	CCATACCTGA	GCCACCGTTC	CAAAGCACTG	GGCGTTTTTC	AGATACTTTC	ATGAGAAATG	14700
55	TTGTCAACAC	CGCAAGTTTG	CAGTACACAG	TCTGAAGAT	ATTCTTGTAT	ATGTAGATGT	14760
	CTGTAGATGC	CCTGAAGGTG	TGTAGACTTT	AGACACCCAG	AAGGTGTGTA	GATGTCTGTA	14820
	GACACCTTCT	ATGTGTGTAG	ATGTCTGTAG	ACGCCCTGCA	GGTGTGTAGA	TATATCTAGA	14880
60	TGGTCTGCCT	GTGTATGATA	CAGGCTAAAA	AGACATTTGT	GGTGGACACT	AGTTGATTAT	14940
	TTAGGACTAT	GAGATGGGAA	AGGAAGNAGC	AACCAGCACT	GAAAGGCATG	TGGTGGGTGG	15000
65	GGGGTTGGCA	TTGCAGTGGG	GTCTCTNTGA	NGCAGGTGAC	ACCCACTATA	GGGCTGCCCT	15060
	TGGNATGGAC	GCTTTGTNGA	AGCTGTTTGA	TTTCACCACA	CCAAGCCTGG	AGGCACGGAC	15120

ATTCCAGGAT GGTGAGGAGT CTGCAAAGGA GGAGATTGGA GGAGGTGCAA TATCCCTAGA 15180
 GTACGAGAGA TGAGATAGGA GAGCTGTATA AATAGCACTA CCAGCCGGAT GCGGTGGCTC 15240
 5 ACGCCTGTCA TCCCAGCACT TTAGGAGGCT GAGGCAGGCG GATCACCTGA GGTCAGGAGT 15300
 TCCAGAACAG CCTGGCCAAC ACAATGAAAC CCCATCTTTA CTAAAAATAC AAGATTAGCT 15360
 10 GGGCACGGTG TCTCACGCCT GTCATCCCTG CACTTTGGGA GGTCGAGGTG CGCAGATCAT 15420
 GAGGTCAGTT TGGCCAACGC GGCAGAACCC CGTCTCTACT AAAAATACAA AAAAGTAGCC 15480
 GGGCGTGGTG GTGGGCACCT GTAGTCCCAG CTACTAGGGA GGCTGAGGCA GGAGAATCGC 15540
 15 TTGAACCCGG ATGCGGACAT TGCAGTGAGC CGAGATC 15577

(2) SEQ ID NO:9 的资料:

(i) 序列特征:

- 20 (A) 长度: 753 个碱基对
 (B) 类型: 核酸
 (C) 链型: 单链
 25 (D) 拓扑结构: 线性

(ii) 分子类型: 其他核酸

(A) 说明: /desc= "ET92 基因区段"

(xi) 序列描述: SEQ ID NO:9:

CGTGGAAGCC TGGAGTTTTT GGAACAGCG TGTCCCGGCC GAGCCTGGGA GCCCGTGGGT 60
 35 TCTGCAAAGC CTGCGGGTGT TTGAGGACTT TGAAGACCAG TTTGTCAGTT GGGCTCAATT 120
 CCTGGGGTTC AGACTTAGAG AAATGAAGGA GGGAGAGCTG GGGTCGTCTC CAGGAAACGA 180
 40 TTCCTTGGG GGAAGGAAT GGAGTGTCTT TGCAGGCACA TGTCTGTTAG GAGGTGAAAC 240
 AGAATGTGAA ATCCACGTTG GAGTAAGCGT CCAGCGCTGA ATGTAGCTCG GGGTGGGGTG 300
 GGAGGGCCCT GGTGTGGATC GTGGAAGGAA GAAAGACAGA ACAGGGTGCT AGTATTTACC 360
 45 CCGTTCCCTG TAGACACCCT GGATTTGTCA GCTTTGCAAG CTTCTTGGTT GCAGCGGCCT 420
 TGCCTGTGCC CCTTTGAGAC TGTTTCCAGA CTAACTTCC AAATGTCAGC CCCTTACCCT 480
 TGACAGCAAG GGACATCTCA TTAGGGCATC GCGTGCTTCT CATCTGTGCT CAGCAGGCCC 540
 50 GAGATAGGAA CAGAGGGGCG TTGGAGATGC CACTTCCACC AGCCCTGGGT TGAAGGGGAG 600
 CGAGGGAGAC ACCTTTTACT TAAACCCCTG AGCTTGGTCA GAGAGGCTGA ATGTCTAAAA 660
 55 TGAGGAAGAA AAGGTTTTTC ACCTGGAAAC GCTTGAGGGC TGAGTCTTCT GCCCTTCTGA 720
 CTCCCCCAGC AAATACAGAC AGGTCACCAA CTA 753

(2) SEQ ID NO:10 的资料:

(i) 序列特征:

- 60 (A) 长度: 1890 个碱基对
 (B) 类型: 核酸
 (C) 链型: 单链
 65 (D) 拓扑结构: 线性

(ii) 分子类型: 其他核酸

(A) 说明: /desc= "SHOXa"

(ix) 特征:

(A) 名字/关键词: CDS

(B) 位置: 91-968

(xi) 序列描述: SEQ ID NO:10:

```

5          GTGATCCACC CGCCGCACGG GCCGTCCTCT CCGCGCGGGG AGACGCGCGC ATCCACCAGC      60
          CCCGGCTGCT CGCCAGCCCC GGCCCCAGCC ATG GAA GAG CTC ACG GCT TTT GTA      114
15          Met Glu Glu Leu Thr Ala Phe Val
              1              5
          TCC AAG TCT TTT GAC CAG AAA AGC AAG GAC GGT AAC GGC GGA GGC GGA      162
20          Ser Lys Ser Phe Asp Gln Lys Ser Lys Asp Gly Asn Gly Gly Gly Gly
              10              15              20
          GGC GGC GGA GGT AAG AAG GAT TCC ATT ACG TAC CGG GAA GTT TTG GAG      210
          Gly Gly Gly Gly Lys Lys Asp Ser Ile Thr Tyr Arg Glu Val Leu Glu
              25              30              35              40
25          AGC GGA CTG GCG CGC TCC CGG GAG CTG GGG ACG TCG GAT TCC AGC CTC      258
          Ser Gly Leu Ala Arg Ser Arg Glu Leu Gly Thr Ser Asp Ser Ser Leu
              45              50              55
          CAG GAC ATC ACG GAG GGC GGC GGC CAC TGC CCG GTG CAT TTG TTC AAG      306
30          Gln Asp Ile Thr Glu Gly Gly Gly His Cys Pro Val His Leu Phe Lys
              60              65              70
          GAC CAC GTA GAC AAT GAC AAG GAG AAA CTG AAA GAA TTC GGC ACC GCG      354
35          Asp His Val Asp Asn Asp Lys Glu Lys Leu Lys Glu Phe Gly Thr Ala
              75              80              85
          AGA GTG GCA GAA GGG ATT TAT GAA TGC AAA GAG AAG CGC GAG GAC GTG      402
40          Arg Val Ala Glu Gly Ile Tyr Glu Cys Lys Glu Lys Arg Glu Asp Val
              90              95              100
          AAG TCG GAG GAC GAG GAC GGG CAG ACC AAG CTG AAA CAG AGG CGC AGC      450
45          Lys Ser Glu Asp Glu Asp Gly Gln Thr Lys Leu Lys Gln Arg Arg Ser
              105              110              115              120
          CGC ACC AAC TTC ACG CTG GAG CAG CTG AAC GAG CTC GAG CGA CTC TTC      498
          Arg Thr Asn Phe Thr Leu Glu Gln Leu Asn Glu Leu Glu Arg Glu Phe
              125              130              135
50          GAC GAG ACC CAT TAC CCC GAC GCC TTC ATG CGC GAG GAG CTC AGC CAG      546
          Asp Glu Thr His Tyr Pro Asp Ala Phe Met Arg Glu Glu Leu Ser Gln
              140              145              150
          CGC CTG GGG CTC TCC GAG GCG CGC GTG CAG GTT TGG TTC CAG AAC CGG      594
55          Arg Leu Gly Leu Ser Glu Ala Arg Val Gln Val Trp Phe Gln Asn Arg
              155              160              165
          AGA GCC AAG TGC CGC AAA CAA GAG AAT CAG ATG CAT AAA GGC GTC ATC      642
60          Arg Ala Lys Cys Arg Lys Gln Glu Asn Gln Met His Lys Gly Val Ile
              170              175              180
          TTG GGC ACA GCC AAC CAC CTA GAC GCC TGC CGA GTG GCA CCC TAC GTC      690
          Leu Gly Thr Ala Asn His Leu Asp Ala Cys Arg Val Ala Pro Tyr Val
              185              190              195              200
65          AAC ATG GGA GCC TTA CGG ATG CCT TTC CAA CAG GTC CAG GCT CAG CTG      738
          Asn Met Gly Ala Leu Arg Met Pro Phe Gln Gln Val Gln Ala Gln Leu

```

	205	210	215	
5	CAG CTG GAA GGC GTG GCC CAC GCG CAC CCG CAC CTG CAC CCG CAC CTG Gln Leu Glu Gly Val Ala His Ala His Pro His Leu His Pro His Leu	786		
	220	225	230	
10	GCG GCG CAC GCG CCC TAC CTG ATG TTC CCC CCG CCG CCC TTC GGG CTG Ala Ala His Ala Pro Tyr Leu Met Phe Pro Pro Pro Pro Phe Gly Leu	834		
	235	240	245	
15	CCC ATC GCG TCG CTG GCC GAG TCC GCC TCG GCC GCC GCC GTG GTC GCC Pro Ile Ala Ser Leu Ala Glu Ser Ala Ser Ala Ala Ala Val Val Ala	882		
	250	255	260	
20	GCC GCC GCC AAA AGC AAC AGC AAG AAT TCC AGC ATC GCC GAC CTG CGG Ala Ala Ala Lys Ser Asn Ser Lys Asn Ser Ser Ile Ala Asp Leu Arg	930		
	265	270	275	280
25	CTC AAG GCG CCG AAG CAC GCG GAG GCC CTG GGG CTC TG ACCCGCCGCG Leu Lys Ala Arg Lys His Ala Glu Ala Leu Gly Leu	978		
	285	290		
30	CAGCCCCCG CGCGCCCGGA CTCCCGGGCT CCGCGCACCC CGCCTGCACC GCGCGTCCTG	1038		
35	CACTCAACCC CGCCTGGAGC TCCTTCCGCG GCCACCGTGC TCCGGGCACC CCGGGAGCTC	1098		
40	CTGCAAGAGG CCTGAGGAGG GAGGCTCCCG GGACCGTCCA CGCACGACCC AGCCAGACCC	1158		
45	TCGCGGAGAT GGTGCAGAAG GCGGAGCGGG TGAGCGGGCCG TGCCTCCAGC CCGGGCCTCT	1218		
50	CCAAGGCTGC CCGTGCGTCC TGGGACCCTG GAGAAGGGTA AACCCCCGCC TGGCTGCGTC	1278		
55	TTCTCTGCT ATACCCTATG CATGCGGTTA ACTACACACG TTTGGAAGAT CCTTAGAGTC	1338		
60	TATTGAACT GCAAAGATCC CGGAGCTGGT CTCCGATGAA AATGCCATTT CTTCGTTGCC	1398		
65	AACGATTTTC TTTACTACCA TGCTCCTTCC TTCATCCCGA GAGGCTGCGG AACGGGTGTG	1458		
70	GATTGGAATG TGGACTTCGG AATCCCAGGA GGCAGGGGCC GGGCTCTCCT CCACCGCTCC	1518		
75	CCCGGAGCCT CCCAGGCAGC AATAAGGAAA TAGTTCTCTG GCTGAGGCTG AGGACGTGAA	1578		
80	CCGCGGGCTT TGGAAAGGGA GGGGAGGGAG ACCCGAACCT CCCACGTTGG GACTCCCACG	1638		
85	TTCCGGGGAC CTGAATGAGG ACCGACTTTA TAACTTTTCC AGTGTTTGAT TCCCAAATG	1698		
90	GGTCTGGTTT TGTTTTGGAT TGGTATTTTT TTTTTTTTTT TTTTTTGCTG TGTTACAGGA	1758		
95	TTCAGACGCA AAAGACTTGC ATAAGAGACG GACGCGTGGT TGCAAGGTGT CATACTGATA	1818		
100	TGCAGCATTA ACTTTACTGA CATGGAGTGA AGTGCAATAT TATAAATATT ATAGATTAAA	1878		
105	AAAAAATAG CA	1890		

55 (2) SEQ ID NO:11 的资料:

(i) 序列特征:

(A) 长度: 292 个氨基酸

(B) 类型: 氨基酸

(D) 拓扑结构: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:11:

Met Glu Glu Leu Thr Ala Phe Val Ser Lys Ser Phe Asp Gln Lys Ser
1 5 10 15

Lys Asp Gly Asn Gly Gly Gly Gly Gly Gly Gly Gly Lys Lys Asp Ser
20 25 30

5 Ile Thr Tyr Arg Glu Val Leu Glu Ser Gly Leu Ala Arg Ser Arg Glu
35 40 45

Leu Gly Thr Ser Asp Ser Ser Leu Gln Asp Ile Thr Glu Gly Gly Gly
50 55 60

10 His Cys Pro Val His Leu Phe Lys Asp His Val Asp Asn Asp Lys Glu
65 70 75 80

15 Lys Leu Lys Glu Phe Gly Thr Ala Arg Val Ala Glu Gly Ile Tyr Glu
85 90 95

Cys Lys Glu Lys Arg Glu Asp Val Lys Ser Glu Asp Glu Asp Gly Gln
100 105 110

20 Thr Lys Leu Lys Gln Arg Arg Ser Arg Thr Asn Phe Thr Leu Glu Gln
115 120 125

Leu Asn Glu Leu Glu Arg Leu Phe Asp Glu Thr His Tyr Pro Asp Ala
130 135 140

25 Phe Met Arg Glu Glu Leu Ser Gln Arg Leu Gly Leu Ser Glu Ala Arg
145 150 155 160

30 Val Gln Val Trp Phe Gln Asn Arg Arg Ala Lys Cys Arg Lys Gln Glu
165 170 175

Asn Gln Met His Lys Gly Val Ile Leu Gly Thr Ala Asn His Leu Asp
180 185 190

35 Ala Cys Arg Val Ala Pro Tyr Val Asn Met Gly Ala Leu Arg Met Pro
195 200 205

Phe Gln Gln Val Gln Ala Gln Leu Gln Leu Glu Gly Val Ala His Ala
210 215 220

40 His Pro His Leu His Pro His Leu Ala Ala His Ala Pro Tyr Leu Met
225 230 235 240

Phe Pro Pro Pro Pro Phe Gly Leu Pro Ile Ala Ser Leu Ala Glu Ser
245 250 255

Ala Ser Ala Ala Ala Val Val Ala Ala Ala Ala Lys Ser Asn Ser Lys
260 265 270

50 Asn Ser Ser Ile Ala Asp Leu Arg Leu Lys Ala Arg Lys His Ala Glu
275 280 285

Ala Leu Gly Leu
290

(2) SEQ ID NO:12 的资料:

(i) 序列特征:

(A) 长度: 1354 个碱基对

(B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑结构: 线性

(ii) 分子类型: 其他核酸

(A) 说明: /desc=SHOXb

(ix) 特征:

(A) 名称/关键词: CDS

(B) 位置: 91..768

5

(xi) 序列描述: SEQ ID NO:12:

	GTGATCCACC CGCCGCACGG GCCGTCCTCT CCGCGCGGGG AGACGCGCGC ATCCACCAGC	60
10	CCCGGCTGCT CGCCAGCCCC GGCCCCAGCC ATG GAA GAG CTC ACG GCT TTT GTA Met Glu Glu Leu Thr Ala Phe Val 295 300	114
15	TCC AAG TCT TTT GAC CAG AAA AGC AAG GAC GGT AAC GGC GGA GGC GGA Ser Lys Ser Phe Asp Gln Lys Ser Lys Asp Gly Asn Gly Gly Gly Gly 305 310 315	162
20	GGC GGC GGA GGT AAG AAG GAT TCC ATT ACG TAC CGG GAA GTT TTG GAG Gly Gly Gly Gly Lys Lys Asp Ser Ile Thr Tyr Arg Glu Val Leu Glu 320 325 330	210
	AGC GGA CTG GCG CGC TCC CGG GAG CTG GGG ACG TCG GAT TCC AGC CTC Ser Gly Leu Ala Arg Ser Arg Glu Leu Gly Thr Ser Asp Ser Ser Leu 335 340 345	258
25	CAG GAC ATC ACG GAG GGC GGC GGC CAC TGC CCG GTG CAT TTG TTC AAG Gln Asp Ile Thr Glu Gly Gly Gly His Cys Pro Val His Leu Phe Lys 350 355 360	306
30	GAC CAC GTA GAC AAT GAC AAG GAG AAA CTG AAA GAA TTC GGC ACC GCG Asp His Val Asp Asn Asp Lys Glu Lys Leu Lys Glu Phe Gly Thr Ala 365 370 375 380	354
35	AGA GTG GCA GAA GGG ATT TAT GAA TGC AAA GAG AAG CGC GAG GAC GTG Arg Val Ala Glu Gly Ile Tyr Glu Cys Lys Glu Lys Arg Glu Asp Val 385 390 395	402
40	AAG TCG GAG GAC GAG GAC GGG CAG ACC AAG CTG AAA CAG AGG CGC AGC Lys Ser Glu Asp Glu Asp Gly Gln Thr Lys Leu Lys Gln Arg Arg Ser 400 405 410	450
	CGC ACC AAC TTC ACG CTG GAG CAG CTG AAC GAG CTC GAG CGA CTC TTC Arg Thr Asn Phe Thr Leu Glu Gln Leu Asn Glu Leu Glu Arg Leu Phe 415 420 425	498
45	GAC GAG ACC CAT TAC CCC GAC GCC TTC ATG CGC GAG GAG CTC AGC CAG Asp Glu Thr His Tyr Pro Asp Ala Phe Met Arg Glu Glu Leu Ser Gln 430 435 440	546
50	CGC CTG GGG CTC TCC GAG GCG CGC GTG CAG GTT TGG TTC CAG AAC CGG Arg Leu Gly Leu Ser Glu Ala Arg Val Gln Val Trp Phe Gln Asn Arg 445 450 455 460	594
55	AGA GCC AAG TGC CGC AAA CAA GAG AAT CAG ATG CAT AAA GGC GTC ATC Arg Ala Lys Cys Arg Lys Gln Glu Asn Gln Met His Lys Gly Val Ile 465 470 475	642
60	TTG GGC ACA GCC AAC CAC CTA GAC GCC TGC CGA GTG GCA CCC TAC GTC Leu Gly Thr Ala Asn His Leu Asp Ala Cys Arg Val Ala Pro Tyr Val 480 485 490	690
	AAC ATG GGA GCC TTA CGG ATG CCT TTC CAA CAG ATG GAG TTT TGC TCT Asn Met Gly Ala Leu Arg Met Pro Phe Gln Gln Met Glu Phe Cys Ser 495 500 505	738
65	TGT CGC CCA GGC TGG AGT ATA ATG GCA TGA TCTCGACTCA CTGCAACCTC Cys Arg Pro Gly Trp Ser Ile Met Ala *	788

510 515

CGCCTCCCGA GTTCAAGCGA TTCTCCTGCC TCAGCCTCCC GAGTAGCTGG GATTACAGGT 848

5 G GCCCACCACC ATGTCAAGAT AATGTTTGTA TTTTCAGTAG AGATGGGGTT TGACCATGTT 908

GGCCAGGCTG GTCTCGAAT CCTGACCTCA GGTGATCCAC CCGCCTTAGC CTCCCAAAGT 968

10 GCTGGGATGA CAGGCGTGAG CCCCTGCGCC CGGCCTTTGT AACTTTIATT TTAATTTTTT 1028

TTTTTTTTTA AGAAAGACAG AGTCTTGCTC TGTACCCAG GCTGGAGCAC ACTGGTGCGA 1088

TCATAGCTCA CTGCAGCCTC AAACCTCTGG GCTCAAGCAA TCCTCCACCC TCAGCCTCCT 1148

15 GAGTAGCTGG GACTACAGGC ACCCACCACC ACACCCAGCT AATTTTTTTG ATTTTACTA 1208

GAGACGGGAT CTGCTTTTGC TGCTGAGGCT GGTCTTGAGC TCCTGAGCTC CAAAGATCCT 1268

20 CTCACCTCCA CCTCCCAAAG TGTTAGAATT ACAAGCATGA ACCACTGCCC GTGGTCTCCA 1328

AAAAAAGGAC TGTTACGTGG AAAAAA 1354

(2) SEQ ID NO:13 的资料:

(i) 序列特征:

- (A) 长度: 226 个氨基酸
(B) 类型: 氨基酸
(D) 拓扑结构: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:13:

35 Met Glu Glu Leu Thr Ala Phe Val Ser Lys Ser Phe Asp Gln Lys Ser
1 5 10 15

Lys Asp Gly Asn Gly Gly Gly Gly Gly Gly Gly Lys Lys Asp Ser
20 25 30

40 Ile Thr Tyr Arg Glu Val Leu Glu Ser Gly Leu Ala Arg Ser Arg Glu
35 40 45

Leu Gly Thr Ser Asp Ser Ser Leu Gln Asp Ile Thr Glu Gly Gly Gly
50 55 60

45 His Cys Pro Val His Leu Phe Lys Asp His Val Asp Asn Asp Lys Glu
65 70 75 80

50 Lys Leu Lys Glu Phe Gly Thr Ala Arg Val Ala Glu Gly Ile Tyr Glu
85 90 95

Cys Lys Glu Lys Arg Glu Asp Val Lys Ser Glu Asp Glu Asp Gly Gln
100 105 110

55 Thr Lys Leu Lys Gln Arg Arg Ser Arg Thr Asn Phe Thr Leu Glu Gln
115 120 125

Leu Asn Glu Leu Glu Arg Leu Phe Asp Glu Thr His Tyr Pro Asp Ala
130 135 140

60 Phe Met Arg Glu Glu Leu Ser Gln Arg Leu Gly Leu Ser Glu Ala Arg
145 150 155 160

65 Val Gln Val Trp Phe Gln Asn Arg Arg Ala Lys Cys Arg Lys Gln Glu
165 170 175

Asn Gln Met His Lys Gly Val Ile Leu Gly Thr Ala Asn His Leu Asp

180 185 190

Ala Cys Arg Val Ala Pro Tyr Val Asn Met Gly Ala Leu Arg Met Pro
195 200 205

5 Phe Gln Gln Met Glu Phe Cys Ser Cys Arg Pro Gly Trp Ser Ile Met
210 215 220

10 Ala *

225

(2) SEQ ID NO:14 的资料:

(i) 序列特征:

- 15 (A) 长度: 32367 个碱基对
(B) 类型: 核酸
(C) 链型: 单链
20 (D) 拓扑结构: 线性

(ii) 分子类型: 其他核酸

(A) 说明: /desc= "粘粒 LLNOYCO3'M'34F5"

25 (xi) 序列描述: SEQ ID NO:14:

TTTCTCTGTC TCCATCCCTC TGTCTCTCCC TTTCTCTCTG TCTTTCCTTG TCTCTCTCTT 60

30 TCTCTCTCTC TCTCCATCTC TCTCTCTCCC TGTCTCTCTC TCTCCATCTC CCCGTCTCTC 120

CGTTTCTCTC TCTGCCTCTC CCTGTCTGTC TCTCTCTTTC TGTGTCTTAC ACACACCCCA 180

35 ACCCACCCTC ACTCATGTCC CCCCCTGCT GTGCCATCTC ACACAAGTTC ACAGCTCAGC 240

TGTCATCCTG GGTCCCCAGG CCCC GCCGGG GAGGAAGATG CGCCGTGGGG TTACGGGAGG 300

AAGGGGACTC CGGGCCTCCT GGTGCCCCAC TTTATTTGCA GAAGGTCCTT GGCAGGAACC 360

40 GTGACGCGTT TGGTTTCCAG GACTTGAAA ACGAATTTCA GGTGCGGATG GCGAGCACCG 420

GCTTCCCCTG AAGCACATTC AATAGCGAGA GCGGGAGGG AGCGAGCAGG AGCATCCCAC 480

45 CATGAAAACC AAAAACACAA GTATTTTTTT CACCCGGTAA ATACCCCAAG CGCCAGGGTG 540

ACAGCGCGGC GCTAAGGGAG GAGGCCTCGC GCCGGGGTCC GCCGGGATCT GGCGCGGGCG 600

GAAAGAATAT AGATCTTTAC GAACCGGATC TCCCGGGGAC CTGGGCTTCT TTCTGCGGGC 660

50 GCTGGAGACC CGGGAGGCGG CCCC GGGGAT CCTCGGCCTC CGCCGCCGCC GCCTCCCAAG 720

CGCCCGCGTC CCGGTTTGGG GACACCCGGC CCCTTCTTCT CACTTTCGGG GATTCTCCAG 780

55 CCGCGTTCCA TCTACCAAC TCTCCATCCA AGGGCGCGCC GCCACCAACT TGGAGCTCAT 840

CTTCTCCCAA GATCGTGCGT CCCC GGGGCG CCCGGGTCCC CCCCCTCGCC ATCTCAACCC 900

CGGCGCGACC CGGGCGCTTC CTGGAAGAT CCAGGCGCCG GGCTCTGCGC TCCTCCCGGG 960

60 AGCGAGGGCG GCCGGACGAC TGGGACCCTC CTCTCTCCAG CCGTGAATC CTTGTCTCTC 1020

TGTCTCTCTC TGCAGGAAA CTGGAGTTTG CTTTCTCTCC GGCCACGGAG AGAACCGGGG 1080

65 TAACCTGTGT GGGGGGCTCG GCGCCTGCG CCCCCTCCT GCGCGCGCGC TCTCCCTTCC 1140

AAAAATGGGA TCTTTCCTCC TCGCACCAA GGTGTACGGA CGCCAAACAG TGATGAAATG 1200

	AGAAGAAAGC	CAATTGCCGG	CCTGGGGGGT	GGGGGAGACA	CAGCGTCTCT	GCGTGCGTCC	1260
	GCCGCGGAGC	CCGGAGACCA	GTAATTGCAC	CAGACAGGCA	GCGCATGGGG	GGCTGGGCGA	1320
5	GGTCGCCGCG	TATAAATAGT	GAGATTTCCA	ATGGAAAGGC	GTAAATAACA	GCGCTGGTGA	1380
	TCCACCCGCG	CGCACGGGCC	GTCCTCTCCG	CGCGGGGAGA	CGCGCGCATC	CACCAGCCCC	1440
10	GGCTGCTCGC	CAGCCCCGGC	CCCAGCCATG	GAAGAGCTCA	CGGCTTTTGT	ATCCAAGTCT	1500
	TTTGACCAGA	AAAGCAAGGA	CGGTAACGGC	GGAGGCGGAG	GCGGCGGAGG	TAAGAAGGAT	1560
	TCCATTACGT	ACCGGGAAGT	TTTGGAGAGC	GGACTGGCGC	GCTCCCGGGA	GCTGGGGACG	1620
15	TCGGATTCCA	GCCTCCAGGA	CATCACGGAG	GGCGGCGGCC	ACTGCCCGGT	GCAATTGTTC	1680
	AAGGACCACG	TAGACAATGA	CAAGGAGAAA	CTGAAAGAAT	TCGGCACC GC	GAGAGTGGCA	1740
20	GAAGGTAAGT	TCCTTTGCGC	GCCGGCTCCA	GGGGGGCCCT	CCTGGGGTTC	GGCGCCTCCT	1800
	CGCCACGGAG	TCGGCCCCGC	GCGCCCCCTG	CTGTGCACAT	TTGCAGCTCC	CGTCTCGCCA	1860
	GGGTAAAGCC	CGGGCCGTC	GGCTTTGCCT	AAGAAAGGAA	GGAAGGCAGG	AGTGGACCCG	1920
25	ACCGGAGACG	CGGGTGGTGG	GTAGCGGGGT	GCGGGGGGAC	CCAGGGAGGG	TCGCAGCGGG	1980
	GGCCGCGCGC	GTGGGCACCG	ACACGGGAAG	GTCCCGGGCT	GGGGTGGATC	CGGGTGGCTG	2040
30	TGCCTGAAGC	CGTAGGGCCT	GAGATGTCTT	TTTCATTTTC	TTTTTCTTTC	CTTTCCTTTT	2100
	TTTGTTTGTT	TGTTTGTGTT	TTTGAGACAG	AGTCTCGCTC	TGTCCCCCAG	GCTGGAGTGC	2160
	AGTGGTGC GA	TCTCGGCTCA	CTGCAACCTC	CGCCTCCTGG	GTTCAAGCGA	TTCTCCTGCC	2220
35	TCAGCCTCCC	CAGTAGCTGG	GATTACAGGC	ATGCACCACC	ACGCCTGGCT	AATTTTGTG	2280
	CTTTTAGTAA	AGACGGGGAT	TCACCATGTT	GGCCAGGCTG	GTCTCGAACT	CCTGACCTCA	2340
40	GGTGATCCAC	CCGCCTCGGC	CTCCCAAAGT	GCTGGGATGA	CAGGCGTGAG	GCACCGCGCC	2400
	CGGCCTGGGT	CCTGACGGCT	TAGGATGTGT	GTTTCTGTCT	CTGCCTGTCT	GCCTTGTATT	2460
	TACGGTCACC	CAGACGCACA	GAGGAGCCGT	CTCCACGCGC	CTTCCAGCG	CTCAGCGCCT	2520
45	GCCGGGCCCC	CGGAGATCAC	GGGAAGACTC	GAGGCTGCGT	GGTAGGAGAC	GGGAAGGCCC	2580
	CGGGTCAGCT	CGGTTCTGTT	TCCTTTAAGG	AACCCTTCAT	TATTATTTCA	TTGTTTTCCT	2640
50	TTGAACGTCG	AGGCTTGATC	TTGGCGAAAG	CTGTTGGGTC	CATAAAAACC	ACTCCCGTGA	2700
	GCGGAGGTGG	CCGGGATCTG	GATGGGGCGC	GAGGGGCCCC	GGGGAAGCTG	GCGGCTTCGC	2760
	GGGCGCGTCC	TAAGTCAAGG	TTGTCAGAGC	GCAGCCGGTT	GTGCGCGGCC	CGGGGGAGCT	2820
55	CCCCCTCTGGC	CCTTCCTCCT	GAGACCTCAG	TGGTGGGTCTG	TCCCGTGGTG	GAAATCGGGG	2880
	AGTAAGAGGC	TCAGAGAGAG	GGGCTGGCCC	CGGGGATCTC	TGTGCACACA	CGACAACTGG	2940
60	GCGGCATACA	TCTTAAGAAT	AAAATGGGCT	GGCTGTGTCTG	GGGCACAGCT	GGAGACGGCT	3000
	ATGGACGCCT	GTTATGTTTT	CATTACAAAG	ACGCAGAGAA	TCTAGCCTCG	GCTTTTGCTG	3060
	ATTGCGCAGAG	TTGAGGTGCG	AGGGTGAATG	CCCCAAAGGT	AATTCTTCCT	AAGACTCTGG	3120
65	GGCTACCTGC	TCTCCGGGGC	CCTGCATTTG	GGGTGTGGAG	TGGCCCCGGG	AAATAGCCCT	3180
	TGTATTGTA	GGAGGCACCA	GGCAGCTTCC	CAAGGCCCTG	ACTTTGTCGA	AGCAGAAAGC	3240

	TGTGGCTACG	GTTTACAAAG	CAGTCCCCCG	TTTCTGACCG	TCTAAGAGGC	AGGAGCCCAG	3300
5	CCTGCCTTTG	ACAGTGAGAG	GAGTTCCTCC	CTACACACTG	CTGCGGGCAC	CCGGCACTGT	3360
	AATTCATACA	CAGAGAGTTG	GCCTTCCTGG	ACGCAAGGCT	GGGAGCCGCT	TGAGGGCCTG	3420
	CGTGTAATTT	AAGAGGGTTC	GCAGCGCCCG	GCGGCCGCTT	CTGTGGGGTT	GCTTTTTTGGT	3480
10	TGTCCTTCGC	AGACACCGTT	TTGCTCCTCT	GAAGTCTCTC	TTCTCCCCCT	GGCCGTGGAC	3540
	CCGGGAGAGC	AAAGTGTCTT	CCAGACCTTT	TGAAAGTGAG	AGGAAAATAA	AGACCAGGCC	3600
15	AAAGACCCAG	GGCCACAGGA	GAGGAGACAG	AGAGTCCCCG	TTACATTTTC	CCCTTGGCTG	3660
	GGTGAGAGAA	GACCCCCGGG	CCAGGACTGC	CACCCAGGCT	ACTATTTATT	CATCAGATCC	3720
	AAGTTAAATC	GAGGTTGGAG	GGCAGGGGAG	AGTCTGAGGT	TACCGTGGAA	GCCTGGAGTT	3780
20	TTTGGGAACA	GCGTGTCCCC	GCCGAGCCTG	GGAGCCCGTG	GGTCTGCAA	AGCCTGCGGG	3840
	TGTTTGAGGA	CTTTGAAGAC	CAGTTTGTCA	GTTGGGCTCA	ATTCCTGGGG	TTCAGACTTA	3900
25	GAGAAATGAA	GGAGGGAGAG	CTGGGGTCGT	CTCCAGGAAA	CGATTCACCT	GGGGGGAAGG	3960
	AATGGAGTGT	TCTTGACAGC	ACATGTCTGT	TAGGAGGTGA	AACAGAATGT	GAAATCCACG	4020
	TTGGAGTAAG	CGTCCAGCGC	TGAATGTAGC	TCGGGGTGGG	GTGGGAGGGC	CCTGGTGTGG	4080
30	ATCGTGGAAG	GAAGAAAGAC	AGAACAGGGT	GCTAGTATTT	ACCCCGTTCC	CTGTAGACAC	4140
	CCTGGATTTG	TCAGCTTTGC	AAGCTTCTTG	GTTGCAGCGG	CCTTGCCTGT	GCCCCTTTGA	4200
35	GACTGTTTCC	AGACTAAACT	TCCAAATGTC	AGCCCCTTAC	CCTTGACAGC	AAGGGACATC	4260
	TCATTAGGGC	ATCGCGTGCT	TCTCATCTGT	GCTCAGCAGG	CCCGAGATAG	GAACAGAGGG	4320
	GCGTTGGAGA	TGCCACTTCC	ACCAGCCCTG	GGTTGAAGGG	GAGCGAGGGA	GACACCTTTT	4380
40	ACTTAAACCC	CTGAGCTTGG	TCAGAGAGGC	TGAATGTCTA	AAATGAGGAA	GAAAAGGTTT	4440
	TTCACCTGGA	AACGCTTGAG	GGCTGAGTCT	TCTGCCCTTC	TGACTCCCCC	AGCAAATACA	4500
45	GACAGGTCAC	CACTACTTGG	AGATGAGAAA	GTGCCATTTT	TGGCACACTC	TGGTGGGGTA	4560
	GGTGCCCGAC	CGCGTGTGAA	AAAGTGGGAA	GGAGAGATTT	CTGCGCACGC	GGTTCAGCCC	4620
	CCAGGCGCGG	TGGCGCATTC	AGGTACTCAG	ACGCGGTTCT	GCTGTTCTGC	TGAGAAACAG	4680
50	GCTTCGGGTA	GGGGCTCCTA	GCTCCGCCAG	ATCGCGGAGG	GACCCCCAGC	CCTCCTGCGC	4740
	TGCAGCGGTG	GGGATAGCGT	CTCTCCGTAG	GCCTAGAATC	TGCAACCCGC	CCGGGGTCTT	4800
55	CCCCGTGTCC	TTCCCGGGCG	TCCCGCCGGG	GATCCACACG	TTGGCAGCTC	TTCCTCAAAT	4860
	TCTTTCCCTT	AAAAATAGGA	TTTGACACCC	CACTCTCCTT	AAAAAAAAAA	AATAAGAAAA	4920
	AAAGGTTAGG	TTATGTCAAC	AGAGGTGAAG	TGGATAATTG	AGGAAACGAT	TCTGAGATGA	4980
60	GGCCAAGAAA	ACAACGCTCG	TGCAAAGCCC	AGGTTTTTGG	GAAAGCAGCG	AGTATCCTCC	5040
	TCGGCTTTTG	CGTTATGGAC	CCCACGCAGT	TTTTGCGTCA	AAGCGCATTG	GTTTTCGAGG	5100
65	GCCCCCTTTC	CACCGCGGGA	TGCACGAGGG	GGTTCGCCAC	GTTGCGCAAA	ACCTCCCCGG	5160
	CCTCAGCCCT	GTGCCCTCCG	CTCCCCACGC	AGGGATTIAT	GAATGCAAAG	AGAAGCGCGA	5220

	GGACGTGAAG TCGGAGGACG AGGACGGGCA GACCAAGCTG AACAGAGGC GCAGCCGCAC	5280
	CAACTTCACG CTGGAGCAGC TGAACGAGCT CGAGCGACTC TTCGACGAGA CCCATTACCC	5340
5	CGACGCCTTC ATGCGCGAGG AGCTCAGCCA GCGCCTGGGG CTCTCCGAGG CGCGCGTGCA	5400
	GGTAGGAACC CGGGGGCGGG GCGGGGGGGC CCGGAGCCAT CGCCTGGTCC TCGGGAGCGC	5460
10	ACAGCACGCG TACAGCCACC TGC GCCCGGG CCGCCGCCGT CCCCTTCCCG GAGCGCGGGG	5520
	AGGTTGGGTG AGGGACGGGC TGGGGTTCCT GGACTTTTGG AGACGCCTGA GGCCTGTAGG	5580
	ATGGGTTTCAT TCGTTTTGT TTTACCAAC AGCAAACAA TATATATACA TATATATTAT	5640
15	ACAAATAACA AATAAATATA TATGTTATAC AGATGGGTAT ATTGTATATA TTATAGATAT	5700
	TTGTTTCGTCC TTGGTGCAAA GACACCCGGT GAACCCATAT ATTGGCTCCT GACTGCCTTC	5760
20	GGTTCCTCTG GGATTGGTTA TAGGGGCAAC ACATGCAAA AAACTTTCC CTGGATTATA	5820
	CTTAGGAGAC GAAGCTACAG ATGCGTTTGA TCCAGAGTGT TTTACAAGAT TTTTCATTTA	5880
	AAAAAAAATG TGTCTTTTGG CCCCTGATTC CCTCCGTCT TCCCGTGTGG CTGCATTGAA	5940
25	AAGGTTTCCT TAGGATGAAA GGAGAGGGGT GTCCTCTGTC CCTAGGTGGA GAGAAACAGG	6000
	GTCTTCTCTT TCCTCCGTTT TTTACCTAC CGTTTCTATC TCCCTCCTCC CCTCTCCAGC	6060
30	CCTGTCTCTT GCTACAAACC ACCCCCTCCT CCTCCGGCT GTGGGGAGCG CAGGAGCAGC	6120
	TTGGGCATCT GGATGAGCGG AGACTATTAG CGGGGCACGG GGGCTCCCCG AGGAGCGCGC	6180
	GAATTCACGC TGCCCCATGA GACCAGGCAC CGGGGGGCGG AGGGGCCCTG GGTGTCCGCA	6240
35	GAGGGACGGG CGGGCAGAGC CTTCTCCGC ATTCTAAACA TTCACTTAAA GGTATGAGTT	6300
	TATTTTCAGGG GTGCTGCTGG GAGAGCCTCC AAATGGCTTC TTCCAGCCCC TGCCTGACAG	6360
40	TTCAGCTCCC CTGGAAGGTC AACTCCTCTA GTCCTTTCTC CTGGTTCTGG GCAGGACAGA	6420
	AGTGGGGGGA GGGAGAGAGA GAGAGAGAGA GAGAGAGACG GTCAGGATCC CCGGACCCTG	6480
	GGGAACCCGT CAAAAATAAA TGAAATTAAG ATTGCCGACC AGAGAGAGAA CCGTGACAAA	6540
45	GCAAACGGCG TTCAAAGCAA AGAGACGAAC TGAAAGCCCG TTCCCGTAGG ACTGGTTATG	6600
	AGGTCAACAC ATTCAAACAC AGCTTGCTCT GGATTTTGCT GAGCAGAGGA AGATACAGAT	6660
50	GCATTTGATC CAAAGTGTGT TACATCTTTC ATTATATGTG TGTCTATATA TATAAACATA	6720
	TATAAATATA TAAACATACA TAAATGTATG TAAATATATA TAATCTATAT ACATATATAA	6780
	ATATATAAAC ACATATATAA TATATAAATC TATAAACATA TATAATATAT AAACATAAAT	6840
55	ATATAAACAT ATATAATATA TAAATATATT AACATATATA AAATATGTAT AAATATATAT	6900
	AAACATATAA ACATATATAA ATATATAAAC ATATAAATAT ATAAACATAT ATAAATATAT	6960
60	ACAAACATAT TGTATATATA TAAATATATA TAAAAACATA TATATACATA TAAAAATATA	7020
	TATAAACATA TATACATATA AAGAAATATA TATAAACATA TATACATATA AATATACATA	7080
	TATAAACATA TATATACATA AAATATATAT AAACATATAT ACATATAAAA ATATATATAT	7140
65	ATTAACATAT ATATACATAT AAAAATATAT ATATTAACAT ATATATACAT ATAAAAATAT	7200
	ATATATATTT TTGGCCCCCTG ATTCCCTTCG GTTCCTGTGG GATGGGTGAT TGAGTCAACA	7260

	CATTCAAACA CAACTTTTCC ATCGATGTTG CTTAGGAGAT GAGGATACAG ATGCGTTTGA	7320
5	TGGAGAGGGT TTTACAAGCT CTTTCATTTA AATATATATA TATATATATA TATATTTTTT	7380
	GGCTCCTGAT TCTCTTCCGT CTTCCCAITG GGCTGCATTT TAAAAGGCTT CCCTAAGATC	7440
	GTTACGATTA AATCAACCCT CCCCAGGCAT CTTTACCGAG GGCTGTGGTC CCCAAAGCGA	7500
10	TACAGCCCAG GAGGGAGAGA GGCTTTGGTG ACTTGGAGGA AGGACTGTGT CCTCCTTAG	7560
	GGCGTCTGTG GCCTCAGTGA GGGGAAGGAAG CTGCATCAGA CAGGGGTTTC CTCGCTGTCC	7620
15	ACCCCTCTGG CAGAAGATGG ATTGGGCTGC CCCGTATAAA TTAATGAAAA GATTAAAGTT	7680
	TCGCTAAAGG GGACATCGAG TTTATGTGTC ATCTCCTGGT GTCTGTGTGC CTGGGATCTG	7740
	CAATATATCC CAGCCCTTGA TGTACTGTTT CTATAAAAAT AAATTACTTG TAATTTAATT	7800
20	CCACACTATT TCTTCCGTA GTCTATTACC GACGAGAGCA CGTTAGTTCA GCTGCGGAAA	7860
	ATTGGTTGTG GGGTGTGTGC GGACCCCGAG AACGCCCTAA AATAAGACA AATCGGGGAC	7920
25	AAGCTGGGGG TTATCGATTG CAGGGGTCGC ATGAAAATT AACGACGGTA AATAATAATA	7980
	AAAACAAACA TGGGAATGCA ATAAAAGACA TAATTCTCCA TCGCCGCGGG GGGAAAGGAT	8040
	CCTATAGTAA AGGCGAGTGC GCTTTGAGGG GTCATAAAAA TCAATTAGTT CCAACACCCA	8100
30	CGTCCCGCGT TGAGGGGACG GGGACGAGCA GGGACAGAAA AAGAAACCAT ATTTGAATCC	8160
	CATCTCTCTG TGAATCTTG GGTACATGC GTCTCAGTAC AGCCCGTCCC GTGCTGTGAC	8220
35	CGGATAGAGT TTCAATTTAC TGTGGAATT TGCTGTAAAT AAATTGAGCA TCCGATAGAA	8280
	GCTGTTGCTG ATTAACCTTT TATTTTTAGC GTGGCCCTGC AAAGTCGTAT CACCCAGCTG	8340
	TCAGGCTTCT AATCGAAAGT TATGAGACCA CGGTGAGGGG CAGGCGGTAA TTTAATTACA	8400
40	ACAAATATCT TTGGGTTTAT GGCGCAGAGC TAAATTAAAT GTCATTATTC ACTGTCTGTA	8460
	ATGGAATCA AAAGGAAATC GCATTACGGC ATTTGGGAAA GAAAGCGGGG AGTGCTCTTT	8520
45	AATGAAGAAA TAACTGTCTT AAGCAGTGC ACACACTTCA CTTACCATAT TCGGGCCTAA	8580
	TTGGAATGGA TCGTGAATCA CTCCAAGACT GATTTATTAG CGCTTCACGC AGCGGCTAAT	8640
	TCATCACTTG TATCTTTCAT CATTTTTTTT TTTCTCTCG CCGTGTTGAA GGGAGAGTGA	8700
50	ATGAGGCTTT CCACGTTTCA GGAGGATTTT CTTTTTTGAA AAATGCCCTT CCAGAGGCTT	8760
	TTGGGTGGCT GGCTTGCTTT CTGGGCCCTG GAGGAGACAG GCGGAGAGTC CAGGTGGGCA	8820
55	TGGAGAGGCA CAGTGGCAGG TCACCTGGAT GGTCACTGGA GGTGGAGGTC TGAAGGCGCC	8880
	AGCTTTGGAA ATTATTGGTG AATTTGATG TCAGCACCAG GCAGGGGCCT TTTTGGCGGG	8940
	GGTGTGAGGG AGGATGACTT TGCTGGGAAA CAGGATCAGG TTCTCCAGGC GCACTGCAGC	9000
60	CCGGTAGGAC CCACTTTGGA AATGAAAAGC CAGTTCCGAA AGCTGGGCTG GAAGCTTCCG	9060
	TGTTGGGTTT AAGAGCAAGT TCACGTTGCG CTGTGTAGAC TCCTGGCTGC TCCCAAATC	9120
65	TGAGGGTTTT CTGAGGTTCC CTTCAIAGGG GCACCGGCCC TGGGCCATGC ACAGTGCCTA	9180
	AGGGTGGCTG TGGGCCGAGG GACCCAGCAC GTGTTTTGCC CACAACAGCC GGAGTGACTG	9240

	GTTCACTCAC	CGCCTTGGCG	GAGGACGCCT	GTTCTCTGGA	CGAATCATT	CTCTTGGGTG	9300
	GTGACTGCCT	TGTGGGTCAA	GGTGCAGGTT	TTCTGCCACA	GAAAACCTGT	TAGGAGGAAT	9360
5	TAAGCGACTA	AGACTGTCAG	GGAGGTGGTG	GTGGGGGAGA	GGAGGGGGTG	GTGTCCAGAT	9420
	TACCAGGCAT	AGGCTAAACT	GCCTGCACTC	TCCAGCTGGT	CTGTCTGTGG	AGGAGGGGAT	9480
10	TGTCAATACT	GGGAGAGCAG	AGGAGGCTCG	TAGGAGGTGA	GAGGGGGTGG	AATTTCATG	9540
	CAAATCTTCA	CATGAGGCCT	GTGTGAATTT	CTCCAGCCTC	CTGAGGGTCC	CCTGCGCTAT	9600
	TGCACTCAAC	TTCTTGATAG	TTTACCCCAA	GACTCAGAAG	TCCTTAGAGG	GGCAGAATGC	9660
15	CCCCACCACA	AAGCCTGCTA	TCCTTGGGCG	TCCTCAGGAC	CCTTGGTCAT	GAATGGGACC	9720
	CTTTCATGTA	TGGGGACCC	TGGTAATATG	AATGGGACGC	CTTCAGCTCC	CCAGGGCTTC	9780
20	CGAGGAGGCC	GAGAAGGGCA	AAGACACTTC	CGAGGAGGCC	GAGAAGGGCA	AAGACATTTT	9840
	CTGGGCTTGG	TGTGTCAAGA	GCTAGATTGG	AGAAGGGGCT	GGATTGGAA	CTCTTTAGCC	9900
	ATCAGCTCAC	CCTCTCCGTT	TGTGGCTAAA	GTCTGAAGGT	GGAAACTTCG	GTTCTCCTAC	9960
25	AGGGTCTACA	GGAGTTGGGG	GGCGGGGCGC	CCACACAGAA	CGCTGGAAAG	TTGACAGTC	10020
	CACTTCCACT	GGCTCGGAAC	TCACTTTTTC	ACCTTAAGTT	CATCAGCGGT	AACGCATAGG	10080
30	TCTCACTTAG	GCAGGGCACG	GATGATTTAA	CAATTTCTAC	TTCTAGGTCA	GGTGGGGTGG	10140
	CTCACACCTC	TAATCCCAGC	ACTTTGGGAG	GCCCAGGAGG	GTGGATCGCT	TGAGGTCAGG	10200
	AGTTTGAGAC	CAGCCTGGCC	AACATGGTGA	AACCCCGTCT	CTACTAAAAT	ACGAAAATTA	10260
35	GCCAGGCATG	GTGGTGAGCA	CCTGTAATTC	CAGCTACTCG	GGAGGCTGAG	GCAGGAGAAT	10320
	CGCTTGAAAC	TGGGAGGTGG	ACGTTGCAGT	GAGGTGAGAT	CACACCACTG	CACTCCAGCC	10380
40	TGGATGAGAG	AGCAAGACTC	TGTCTCAAAA	ACAAAATAAA	ACAAAAACAA	AACAAAAATC	10440
	AAAAAAGAAA	ACCCAATTTC	CAGTTCTAGG	CCAGGTGCAG	TGGCTCACGC	CTGTCATCCC	10500
	AGCACTTTGG	GAGGCCCAGG	AGGGTGGATC	GCTTGAGGTC	AGGAGTTCGA	GACCAGCCTG	10560
45	GCCAACATGG	TGAAACCCCA	TCTCTACTAA	AAATACAAAC	GTTAGCTGGG	TGTGGTGGTG	10620
	TGCGCCTGTA	ATCCCAGCTA	CTCGGGAAGC	TGAGGCTGGA	GAATTGCTTG	AATCTGGGAG	10680
50	GTGGAGGTTG	CAGGGAGGCG	AGATAGTGCC	ACTGCAGTCC	AGCCTGGACC	AGAGAGCAAG	10740
	ACTCCGTCTC	AAAAACAAAA	GAAAGCAAAA	ACAAAAACA	AGAGACCAGC	CTGGCCAACA	10800
	TGGTGAAACC	GCGTCTCTAC	TAAAATACAA	AATTAGCCGG	GCATGGTGGT	GGGCACCTGT	10860
55	AGTCCCAGCT	ACTCGGGAGG	CTGAGGCAGG	AGAATGGCTT	GAACCTGGGA	GGTGGAGCTT	10920
	GCACTGAGCC	GAGATAGTGC	CACTGCACTC	CAGCCTGGGC	GACAGAGCGA	GAATTGATTT	10980
60	CAGAACCAAC	ACCACCACAA	CAAAACAAAA	CAAAAAATCC	AAAAAAACCC	CAATTTCCAG	11040
	TACTAGGTAG	TCAGTGATGC	AGGGCTGGAG	ACAGAGGGGC	GGTAAGTGTC	TGGGCGCCCA	11100
	CCATCAGTCA	CCTCCCAGCT	CCCAGAGGTG	CAAAGTGCTT	GGTTCAGCCT	CATGGGAAGG	11160
65	ATGCTCCCTG	GGGAGGCTGG	GCTGGGTTCA	CAGGGCTCTT	CACATCTCTC	TCTGCTTCTC	11220
	CCCAAGGTTT	GGTTCAGAA	CCGGAGAGCC	AAGTGCCGCA	AACAAGAGAA	TCAGATGCAT	11280

	AAAGGTGGGT	GTCTGGGACTG	GGGGGACCTG	AAGCTGGGGG	ATCCTGCTCC	AGGAGGGATG	11340
5	GGGTCGACGA	GGTGTGGCT	ACACCCAGGA	CCACCACACT	GACACCTGCT	CCCTTTGGAC	11400
	ACAGGCGTCA	TCTTGGGCAC	AGCCAACCAC	CTAGACGCCT	GCCGAGTGGC	ACCCTACGTC	11460
	AACATGGGAG	CCTTACGGAT	GCCTTTCCAA	CAGGTAGCTC	ACTTTTTCTT	CCTCTGAAGA	11520
10	TCCCTAGGGA	CCTGCTGCTC	CCTTCCCCTT	TCCCCTATTT	GCTGCCGCAT	CCTGACACTC	11580
	CTAGTCCCCTC	CCTGCCCTG	CAGACTTCTC	AGCTGGCCCT	TAGAAAAAAA	GCCTCTTTTC	11640
15	CGAGGAGGCA	TTTACAGGCA	CCTTGGCACC	TATGAAATCA	GGCTGGGCCA	GGCGGGGTGG	11700
	CTCACACCTG	TCATCCCAGC	ACTTTGGGAG	GCTGAGGAGG	GTGCATCACC	TGAGATCAGG	11760
	AGTTCAAGAC	CAGCCTGGCC	AACTTAACGA	AACCCCGTCT	ATTAAAAATA	CAAAATGGGT	11820
20	GTGGTGGCTC	ACGCCTGTCA	TCCCAGCACT	TTGGGAGGCC	GAGGCAGGTG	GATCACCTGA	11880
	GGTCAGGAAT	TCGAGACCAG	CCTGACCAAC	ATGCTGAAAC	CCCGTCTCTA	CTGAAACAC	11940
25	AAAGCTTAGC	CGGGCGTGGT	GGTGACACAC	TGTGATCCCA	GGTACTTGGG	AGGGAGAATC	12000
	ACTTGAACCT	GGGAGGTGGA	GGTTGCCGTG	AGCCAATATC	GCGCCACTGC	ACTCCACTCT	12060
	GGGTGACAGA	GTGAGACTCC	AAGACTCCAT	CTCAAAAAAA	AAAAAAAAAA	TCAGGCTGTA	12120
30	AAAATCCACT	TTTGGGAAGG	TGAACACACA	CAAGCCCAAA	CAGAAATCTG	ACAAAAACCA	12180
	GAGGGGTGAA	AAGTCCACAC	AGTCAGGCAC	CCCCACCTGG	CTTGCTGCCT	GGTTAAGAAG	12240
35	GGCGCAGATG	CCTGTGCCTG	GATACCAGAG	ATGGGACAGA	CACCCATTCC	CTTTTCATCA	12300
	CCACCCCCGA	GTGCCCCGAG	GCCTGGGGCG	TCTGCCTGGC	CCCTGGCCCC	TGGCTTGGGC	12360
	TCTGCACCTC	TGAAGTGGAG	ACACCCACT	CAGCTCCCCA	CTTACTTTGG	AGTGAGCAGC	12420
40	GCTTGGGTGC	CCAGCGTGGA	TTTGGGGCTT	CCAGGGAGTC	GGGGTTCGGT	CGCGGAGCCC	12480
	AAGCTTCCCA	AGGGCGCCCC	CGCCCTGCCC	TGGCTTAGTG	GTGGGGATGG	GATGGGGGGA	12540
45	AACGGGGAGC	TGCGTGGAAG	GAGGTGAAGG	GTACACAGGAG	GAGAGAGCGC	AGCGCCCACG	12600
	TGCGCCCTGC	CTGAACGCGC	AGCGCAGCGC	CCGGCTGCGG	TGCCCCCTGC	CCCTTCGGTC	12660
	CCTAATTTGG	GGATCGGGAG	TGCATGCGCG	GGCGGAACGG	GCTTGGGGGG	GGGGCTCTGG	12720
50	CAGGGCGGAC	GCGTGGCCTC	CCTTCTTCAC	CGTTTTATT	CAAGGGGACA	GGCTGGGGAT	12780
	TGTATTTGGG	CGCGTGTGTT	GCTGAGGGTG	CAGGGACTTG	GGGGGTGGCG	GTGGGGAGCG	12840
55	CGGAAGGTAT	AAACGTATAA	ATCATAAGTA	AACAACTCAG	AAATGGACCC	CGAGCGCTGG	12900
	TCGCCGCTAG	CTCTCCAGCT	CTCCCTGGCC	CAGGCCCGAA	GGAGAGGGGT	CCGCATCCCT	12960
	CCGCGGTTCT	CCTCTCCTGG	GTACCTGGCC	TTGAGGTGGG	GGAACGAGCC	TACTTCTTGT	13020
60	ACCGTCTTTT	GCCGACGGCG	GGACCCAGTG	AAATTAGGCC	GTTGGAGCCC	GCAGGCCTGC	13080
	CTGGCTTTGC	GCACCGGAGT	CTTGGGGACC	TGGTGTCCCC	GGGAAAAACT	TGGGGACCTG	13140
65	GTATCCCCGG	GAGAGGCTTG	GGGACCTGGT	GTCCCGGGAG	AGGCTTGGGT	ACCTGGTTTC	13200
	TCTGGAAGAG	GCTTGGACAC	CTGGTGTCTT	GGGAGGGCCT	TTGGGACCTG	GTGTCCTGGG	13260

	AGAGGCTTGG	AGATCTGTTG	TCCTGGGAGA	GGCTTGGGGA	CCTGGTGTCC	CTGGAGAGGC	13320
	TTGGGGACCT	GGTGACCTTG	GAGAGGCTTG	GAGACCTGGT	GTTCTGGGAG	AGGCTTGGGG	13380
5	ACCTGGTGTT	CTGGGAGAGG	CTTGGGGACC	TGGTGTCTCT	GGAAGAGGCT	TGGACACCTG	13440
	GTGACCCGGG	AGGGCCTTGG	GGATCTGGTG	TCCCGGGAGA	GCCTTGGGGA	CCTGGTGTCC	13500
10	TGGGAGAGGC	TTGGGGACCT	GGTGACCTTG	GAGAGGCTTG	GGGACCTGGT	GTCCTGAGAG	13560
	AGCCTTGGGG	ATCTGGTGTC	CCAGGAGAGG	CTTGGGGACC	TGGTGTCTCT	GGAAGAGGCT	13620
	TGGACACCTG	GTGTCTGGG	GAGAGGCTTG	GGGACCTGGT	GTCCTGGGAG	AGGCTTGGGG	13680
15	ACCTGGTGTC	CTGGGAGAGG	CTTGGAGATC	TGGTGAGCCG	GGAGAGGCTT	GGGGACCTGG	13740
	TGTCCCGGGA	GAGGCTTGGG	GACTTGGTGT	CCCGGGAGAG	GCTTGAACAC	CTGGTGTCCC	13800
20	AGGAGAGGCT	TGGGGACCTG	GTGACCTTGG	AGAGGCCTGG	GGACCTGGTG	ACCCGGGAGA	13860
	GCCTTGGGGA	CCTGGTGTCC	TGGGGAGAGC	CTTGGGGACC	TGGTGACCTT	GGAGAGGCTT	13920
	GGGGACCTGG	TGTCTCGGGA	GTGCCTTGGG	GACCTAGTGA	CCCGGGAGAG	GCTTGGGGAC	13980
25	CTGGTGTCCC	GGGAGAGGCT	TGGGGACCTG	GTGTCTGGG	AGAGCCTTGG	GGATCTGGTG	14040
	TCCTGGGGAG	AGGCTGGGGG	ACCTGGTGTC	TCGGGAGAGA	GCCTTGGGGA	CCTGGTGACC	14100
30	CGGGAGAGGC	TTGGACACCT	GGTGTCCCGG	GAGAGGCTTG	GGGACCTGGT	GACCCGGGAG	14160
	AGCCTTGGGG	ACCTGGTGTC	CTGGGGAGAG	GCTGGGGGAC	CTGGTGTCTC	GGGAGAGAGC	14220
	CTTGGGGACC	TGGTGACCCG	GGAGAGGCTT	GGACACCTGG	TGTCCCGGGA	GAGGCTTGGG	14280
35	AGCCTGGTGT	CCCGGGAGAG	CCTTGGGGAC	CAGGTGACCT	TGGAGAGGCT	TGGGGACCTG	14340
	GTGATCTTGG	AGAGGCTTGG	GGACCTGGTG	TCTCGGGAGA	GGTTACGGGG	GCTGGTTGGG	14400
40	GGAGAGAAAG	TTGTGAGCCA	AAGTCCCTGA	ATCCCTGCGA	AAAGAGCGCA	TCGGGAGCTC	14460
	CCCCTGAGGG	CGTTCCATTT	GTGGACCCCC	CTCCCATGCG	CTTTCAGGGG	AGCTGTTCCG	14520
	ATTCCCTTGG	CCCGGCTCCC	GCGGATGCAT	CCAGTGGCAG	CGCCAATTCT	GGGCCAGGGG	14580
45	GAAGGAGGAA	AGGCGGGTGT	GGGGTGGTCT	CCACGGCTGG	AGAAGGGGCG	ACGCTCCCTA	14640
	GGGGAGAAGA	GGCACGTTGG	GGGTTTCCGG	GGGCGCGGGG	CGGAGCAGGC	CCCCCAGTCC	14700
50	CCATCCTGCG	CCCTCACCCC	GCCGGGTCCG	CTCCCGCAGG	TCCAGGCTCA	GCTGCAGCTG	14760
	GAAGGCGTGG	CCCACGCGCA	CCCGCACCTG	CACCCGCACC	TGGCGGCGCA	CGCGCCCTAC	14820
	CTGATGTTCC	CCCCGCCGCC	CTTCGGGCTG	CCCATCGCGT	CGCTGGCCGA	GTCCGCCTCG	14880
55	GCCGCCGCCG	TGGTCGCCGC	CGCCGCCAAA	AGCAACAGCA	AGAATTCCAG	CATCGCCGAC	14940
	CTGCGGCTCA	AGGCGCGGAA	GCACGCGGAG	GCCCTGGGGC	TCTGACCCGC	CGCGCAGCCC	15000
60	CCCGCGCGCC	CGGACTCCCG	GGCTCCGCGC	ACCCCGCCTG	CACCGCGCGT	CCTGCACTCA	15060
	ACCCCGCCTG	GAGCTCCTTC	CGCGGCCACC	GTGCTCCGGG	CACCCCGGGA	GCTCCTGCAA	15120
	GAGGCCTGAG	GAGGGAGGCT	CCCGGGACCG	TCCACGCACG	ACCCAGCCAG	ACCCTCGCGG	15180
65	AGATGGTGCA	GAAGGCGGAG	CGGGTGAGCG	GCCGTGCGTC	CAGCCCGGGC	CTCTCCAAGG	15240
	CTGCCCGTGC	GTCCTGGGAC	CCTGGAGAAG	GGTAAACCCC	CGCCTGGCTG	CGTCTTCCTC	15300

	TGCTATACCC	TATGCATGCG	GTAACTACA	CACGTTTGGA	AGATCCTTAG	AGTCTATTGA	15360
5	AACTGCAAAG	ATCCCGGAGC	TGGTCTCCGA	TGAAAATGCC	ATTTCTTCGT	TGCCAACGAT	15420
	TTTCTTTACT	ACCATGCTCC	TTCCTTCATC	CCGAGAGGCT	GCGGAACGGG	TGTGGATTG	15480
	AATGTGGACT	TCGGAATCCC	AGGAGGCAGG	GGCCGGGCTC	TCCTCCACCG	CTCCCCCGGA	15540
10	GCCTCCCAGG	CAGCAATAAG	GAAATAGTTC	TCTGGCTGAG	GCTGAGGACG	TGAACCGCGG	15600
	GCTTTGGAAA	GGGAGGGGAG	GGAGACCCGA	ACCTCCCACG	TTGGGACTCC	CACGTTCCGG	15660
15	GGACCTGAAT	GAGGACCGAC	TTTATAACTT	TTCCAGTGTT	TGATTCCCAA	ATTGGGTCTG	15720
	GTTTTGTTTT	GGATTGGTAT	TTTTTTTTTT	TTTTTTTTTT	GCTGTGTTAC	AGGATTGAGA	15780
	CGCAAAAGAC	TGCATAAGA	GACGGACGCG	TGGTTGCAAG	GTGTCATACT	GATATGCAGC	15840
20	ATTAACCTTA	CTGACATGGA	GTGAAGTGCA	ATATTATAAA	TATTATAGAT	TAAAAAATAA	15900
	ATAGCCGTGC	ACTCTTGACC	CCGTCAACGT	CCAACGTGGA	AAAGGCGTTA	CCTCTTCTCC	15960
25	CAGCGCTGGC	CGCCTGGCCA	CTGAGGGCCC	TTTGCAAAAA	TCACGGGTGT	AGAGATGGCC	16020
	CTGGGCGCGC	TGGGAGTGTG	GTTGTGTTTC	TGAAGGGGAT	AAAAGAGGGC	ACGGTGGTGC	16080
	CAAGATATCA	GTTTGGTACC	TGAGCTGTTT	CTGGTTGGGA	AGCGTAAAAG	CCAGGGAGAG	16140
30	ATCCAGAGAG	TTTTCAAGTT	TTTGAGATG	TAGGTGGTTC	CAGCTTTTCT	TTCTCCCTTA	16200
	CTCCATCTTC	TGCGTTCCCC	CAGTCTTTT	ATTTCTTTGT	TTTTTATTTT	TGAGACAGAG	16260
35	ACTTGCTTTG	TCGCCCAGGC	TGGAGTGCCG	TGGCGCAATG	TCAGCTCACT	GCCACCTCCA	16320
	CCTCCCGGGT	TCAAGCGATG	CTCCTGCCTC	AGCCTCCCGA	GTAGCTGGGA	CTACAGGCAC	16380
	CTGCCACCAC	CCCCGGCTAA	TTTTTTGTAT	TTATAGTAGA	GACGGGGTTT	CACCGTGTG	16440
40	GCCAGGCTCG	TCTCGAACTC	CTGACCTCAG	GTGATCTGCC	CGCCTCGGCC	TCCCAACGTG	16500
	CCCCCAGTTT	TATAAACAGC	AGATAGCAAC	TTGTCTGCAC	AGCTGGCATG	GGCTGGACAG	16560
45	TTGCTTGAAA	TGACCTAACC	AAAAACATTC	AAGGGTTCTG	CCCCCAGATT	TCGGGAGATC	16620
	CACGTTCCAT	GTTCTGATTG	GTTTTCTGGG	AACACAGCAA	GGGGTTTGGT	GACCTCCGAG	16680
	AAGATCCATC	TGCATGATTG	GCATTAGTTA	CCACAGCCTG	CCCAGAGAGA	AACTATCTTC	16740
50	TCCCAACATT	TACTAACATC	CACTGGTCAA	CTCTCTTATT	TCCATAACAC	ATTTGCATCT	16800
	TTCTGGATTG	AAGCTTGGTG	GTTTTCTTTC	CTAATTCTG	ATTAGATAC	TTCTCCCTGA	16860
55	GGTGGGGATA	AAAGAAAAAA	AAAAACAAC	TTCTTTTTTT	CTCCGCATA	ACACTTTCTA	16920
	TCTTGTCACT	GAGCTGAACT	GTAGATCCAT	TTGGACCCGT	CTCATTTGTA	TCTTCTGATA	16980
	TTCTTTATAC	AAACCAAAAG	TCCCCTTCAA	CATTTTTTAT	GTCAAAATGT	TACAACCGCT	17040
60	GTAAAATGAC	GGAGAGAGAG	AGAAAGAATC	CCAGACATTA	ACGGTATTAG	AGAGTTTGCC	17100
	TCATTCAATC	ATTTTTCTTA	AAAGCTGGAA	ATTAAAAAAA	AAAAGAGAG	AGAGAGGCTT	17160
65	TAATAGTTAA	GCTGAAATTT	TTATCGAAAA	GAAGAATTGC	ATTTTGAATC	TTTGGGAAGT	17220
	AGGTTCAATC	ATCAGAGTAT	GTAACCTTTT	GGAAAAGTGG	TTGGTAAGAT	ATGTACAGCC	17280

	CTAGATTTTT TTTTTTTTAA CCAAAAAGGC TGACTAATTT TGAAAAATCG AAACATAACA	17340
	GTGTGTCATC ATTTCTCTCC AAGAAAAAGC TCACTCCACG TGAGTAGAAA GACATCTACC	17400
5	TGGTCCCTGT AGAATCTGAA CGTTTCTCTT TAGAGACGGA ATTTCAATCT TGTGCCCAG	17460
	GCTGGAGTGC AGTGGCACAA TCTCGGCTCA CCGCAACCTC CGCCTCCCGG GTTCAAGCCA	17520
10	TTCTCCTGCC TCAGTCTCCC GAGTAGCTGG GATTACAGGC ACCTGCCACC AGGCCTGGGT	17580
	AACTTTCTGG TATTTTTAGT AGAGACAGGG TTTCAGCCTC CCGAGTAGCT GGGATTACAG	17640
	GCACCTGCCA CCAGGCCTGG GTAACTTTCT GGTATTTTAA GTAGAGACAG GGTTCAGCC	17700
15	TCCCGAGTAG CTGGGATTAC AGGCACCTGC CACCAGGCCT GGGTAACTTT CTGGTAGTTT	17760
	TAGTAGAGAC AGGGTTTCGG CCTCCCGAGT AGCTGGGATT ACAGGCACCT GCCACCAGGC	17820
20	CTGGGTAACT TTCTGGTATT TTTAGTAGAG ACAGGGTTTC GGCCTCCCGA GTAGCTGGGA	17880
	TTACAGGCAC CTGCCACCAG GCCTGGGTAA CTTTCTGGTA TTTTGTAGTAC AGACAGGGTT	17940
	TCGGCCTCCT GAGTAGCTGG GATTACAGGC ACCTGCCACC AGGCCTGGGT AACTTTCTGG	18000
25	TAGTTTTAGT AGAGACAGGG TTTCAGCCTC CCGAGTAGCT GGGATTACAG GCACCTGCCA	18060
	CCAGGCCTGG GTAATTTTTT TGCATTTTTG GTAGAGACAG GTTTTGGCCG TGTGGCCCCG	18120
30	GCTGGTCTCA AACTCCTGAC CTCAGGTTGA CCTGCCCGCT TTGTCCCTCG CAAAGTGCTG	18180
	GGATTACAGG CGTGAGCCAC CACACCTGGC CTGAATCTGA ACTTTTAAAA GGGAGTTACT	18240
	GACTCTCAAC TGTGCGGGGA CGGTTTCACT TTGATTTAAT ATGGAAAGAG GGCCAAGTGT	18300
35	CATCCTCACA AATGGGTCCC CGAAGCAGAT CAAACGCAGA GAACTGTGAG GGTGGGACAC	18360
	GAGTGTCTGT GGACACTGGC TGCCTTTGGC TTTTCTCTCG CGAGAGAAGT TGGGTGACTT	18420
40	TCTGTAGGTG GATGAGTGAT CCCTGAATGA GTGTGGGGTA CGTGTATGCT AGCTGCTTCT	18480
	TTCTCCCTGA AACTCTCGGA TGGGAAGGAAG TAAGAAATTC AGCTTGGGCT GTGACCAGTT	18540
	CTCACCACCA ACGCCCTCTT CTCTCTCCCT TCTCCTTCTT TCCTTCCTTC CTTCTTTTCT	18600
45	TTCTTTTTCT TTCTTTCTCT CTTTCTTTCT TTTCTTTCTT TCTGTTTCTT TCCTTTTTAT	18660
	CTTTCTCTCT TTTTCTTTCT CTTTCTCTTT TTTGTTTCTT TCTTTCTTTT TCTTTCTTTC	18720
50	TTTTTCTTTC TTCTTTCTTT CTTCTGATGA GTCTCACTCT GTCACCCAGG CTGGAGTGCA	18780
	GTGGTGCAAT CCCAGCTCAC TGCATCTCT ACCTCCTGGC TTCAAGAAAT TCTCTGCCT	18840
	CAGCCTCCCA AGTAGCTGGG ATGACAGGCA CCCACCACCA TTCCCGGATA ATTTTGTAT	18900
55	TTTTTAGTAG AGACTGGGTT TCGCCATGTT GGCCAGGCTG GTCTTGAACCT CCTGACCTCA	18960
	CATGATCCAC CCGCCTCAGC CTCCCAGAGT GCTGGGATTA CGGGGTGAGG CACCGCGCCC	19020
60	GGCCTCCTCT CTCTTTTCTT GAGATGTTTA GGAAGGACTG GGCTGATGGG GACCCTCTGT	19080
	ATGTGATGTG CGTGGGTTTG GTTTCCCGGA AGGCCCTCCA GAGACACGTT TCGTGAACA	19140
	TTCAGCATGG AAACAACATA CGTCTCTCCA CAGGAGGTGA GAAATTGAAT TTATGGGGTG	19200
65	GGTGTACGCT GGCGATTCTT GGTGCTTTTT GCTCAAACA AGGTTCTTTT GAAAGTCACG	19260
	TTCTGCTTTT CCCTGTGGCT TCCCGGTGAG CTCGCTCGCA GAGCAAGGAA TACCACCCAG	19320

	AGAGCAACGT GGGCTGTGTT CCGTTGTAAC GCCGTTGCAG AGAGAGGATT TGGTGTGTGA	19380
5	GATCCGTACC AGCTCCAGCA CACTGATAGG AACACGTTGC TGGCCGAAC GAACGATGCT	19440
	GGGTTGGGTC CTGATTGATA CGTATTTTCT TCCCTCCTCT CCCCAAAAC TGGCCAAATA	19500
	GTCCGTGGAG GGTGTGTCAGT CGCCGCAGTT GAGCAAAAAA CACTTCTTCC TTTGAGTGGC	19560
10	TGTTCTGGTG AAATCTGTTT CTGACATATC CACTTTTCTC TCTCTTTTCT CTCTCTCTGA	19620
	CTGCCAAGCA CCCACAGGGA GAAGGAATTG GATGTATCGG ATGTTGGTAT TAGATTTTCT	19680
15	TTCTCCGTTT GAGTCTCTGA CTGGTGCATA CTTTGCAAAG GTGTGTTCCCT GGCAATTGCC	19740
	AAGAGTTAGA AAAATGCACC TTCTCTGGTG GCCGTTGGGG TGTTGTTTCA CAGGCAGTGG	19800
	TGACAGGGCC CCTTGGCTGT GGCTGTCTTC TCCAGCGCCG TGGATAAAGA GACGGGACAG	19860
20	ATTCTGTGCC TCTGTACGAT TTAGAGCGTA ACTGACCGCG TCCAACACCC GTTTTCCAC	19920
	TTACAAAGCT GGTGGTGCGA CGGGCTTGGT GTCTCCCGTA CGGGAAGGAG GCCTTTGGGC	19980
25	CGCTCCAAAG ACGCCCTGTC GTAGGAATGG CCTCTCCATC CCGCCAAAGT CCAGCCAGGC	20040
	CCCCGAAATG GTCCCATTTT CTTGGAAGCC TGAGTTTCTG TTCTGGTCTT GCTGCTGTCC	20100
	TTGGCCACGT CAGCACGTGG GAGCATCTGT GGATACCGCA GAGTCTGGGG ACAGCTGGGC	20160
30	GTTTAACCGA AATGAAGCCG AGACGGGTTT CAGGTTTTGG TGCCAAGCTC TGGTGAGGAT	20220
	GAAAGGGAAA TACCAGAGTC CTCTGTCTTC GCCTCTGGGT TTCATGCTGA CCTTTCTAAC	20280
35	ATTTGTTTTT CCCTAAGAAC AAGCAGAAGC CTCCAGCTCC CTTTAGCTCC ACAGTTTTCC	20340
	CGGGGACATA GCGAGGATGG CACACGGCAG CCACTCCCAC GACACACATT TCGGAGGCAC	20400
	TTTGCTGGAA GCCGCTTGTC TCCTCCAGCT TTGGGAGGTC TGGGGAGGAG AGAGGCTTTC	20460
40	GGTGGACACG TTTGACATTA AAAAAAAAAA AAAAAAAAAA AAAAAACTG GTGCCTAATT	20520
	TATTAAAGAG AATTAGCTTA GCGAGTATAT GCTGATATTC TTCGACACAC GTGGGTAAGT	20580
45	TGATGCCATT TATAAATGTT TTATTGAAAT TTGATATTTA ATGAGAAGCC GGTAAAGGAA	20640
	TGTAGACAAT ATCCCGTTTC AAAGCTATGA AATGTGCTAT TTATTGAAAG GGGATGTGGC	20700
	TTCACGAGTT CAGCCCATTG TACGTGCAGG TCCCGTGGGA AGGAGGCAAA AGCCCCTGCT	20760
50	TCTTACTTTG TGATGTATGT GCATTGTGTA TTTATTTTTT TTTCCTTGGT CGGACGTTCA	20820
	TAAATATGTA CTATTTTAAAT TATGTCGAGT GTAAATTTGA CATCGCGTTG CATTATTTTT	20880
55	TATATTTCTG AAAACTGTTG CTTTTCTTTT TTCCCTCCCC CATTGACGAC ATAGCGGCCC	20940
	CCGCGTCCGG GTTACAAATA CATCTACAGA TATTTTCAGG GATTGCTTCA GATGAAACA	21000
	AATCACACAC CGTTTCCCAA ACCAACAGTC TTCACATTTT TATCCCTCTG TTATTGTCCG	21060
60	CAGGCGGTGA GGGGTAGAAA AAAAACAAAC AAACAAACAG AAAAAAAAAA CAAAAAAAAA	21120
	CACCCTGAGT TTCTCTGGTG ACGCCCTCAT TCTCCTAACG TTCAATAATC TCAATGTTGA	21180
65	GTTGCAGCAA CAGACTGTAT TTTTGTGACG CCCCCTAGTA TGAATGTACA TCTTGTAATA	21240
	CTGAGATATA AATAAATTTA TAAATATTTG TATTCAAGTG TTAATAAAAA AAAAATTTCT	21300

	AACCTCTCCC	CTGAGGACAG	GCTTATTGGA	AAAAAAAAAA	AAAAAAAAAA	ATCCTGAGTC	21360
	GGCCGTGGCT	GAACACAGAG	TGTTGTTCTG	CTCCGTGCAT	TTCCAGGGTG	GGTACCCAGT	21420
5	GTTGCCCCC	AGCCTTAGAT	CGGGAGGTAC	CATTGACTTT	TGCTTGATC	CCATCCCCTT	21480
	CCTTTACTGA	AACCTACCTC	CCCCTTCTC	AGCCAACGTC	CCCCCAGAAG	GTGGCAAAAA	21540
10	AAACAGAGGA	AAAAGCCCTG	ATTTGAATCA	AGTCAGAGCT	GCTAATCTC	CACTTTCTTT	21600
	AATTAATTAA	TTTATTTTTT	TTTTTGAGAC	TGAGTCTCGC	TCTGTCGCCC	AGGCCGGAGG	21660
	AGTGACAGGG	CGCGATCTCG	GCTCACCGCG	ACCTCCGCCT	CCCGGGTTCA	AGCGACTCTC	21720
15	CTGCCTCAGC	CTCCCGAGTA	GCTGGGATGA	CAGTCACCTG	CACCACCGCG	CCCGGCTCAT	21780
	TTTTGTATTT	TTAGTAGCAA	TGGGGTTTCA	CCGTGTTGGT	CAGGCTGGTC	TCGAACCTCT	21840
20	GACCTCGTGA	TCCACCCGCG	TCTGGGCCCC	GCCGGTGATG	TGTGTGCTTT	TAACTTTTAT	21900
	TTTGTTCCAG	TTTTCGACAG	TGGCACGGAT	TTCCAGCAC	GGTCTTGCAA	GGATGATTGA	21960
	GTCATTTTTG	AGACAAAAAA	TATAATAATA	ATAAATGGAA	AAAGAAATCG	ACTTTTAAAA	22020
25	ATGACAAATT	TTTTTTTTTT	TTTTTGCAT	AGATTTTTCT	CTCTTTATGT	AAAGGAAAGT	22080
	TCATGATTGG	ATTTGGCCGG	CCTGACTGCT	TCCCGGCTGT	GATAAAAAAC	ACATGTGAGC	22140
30	TGGGAGGGAA	GTGGGGGAGG	GACACAGCTG	CCCACACAGG	GTTCCCACCG	CGGTTACAGG	22200
	GTGGGCAGTG	CTGGGGGAGC	TTTCTCTGTG	GGGGGCTCAG	AGCCTGAGGA	CAGGTGAGCC	22260
	TCTCCGACAC	CTCCCCAGTT	GCCTGGAGTC	TAAACCGTCC	GTTGTCTGTA	CCGTCCGTTT	22320
35	TTCCTGCTGA	CTCCTGGTAG	TTCCTGAAAG	CTTCTCTTGG	CCAGAGAAGG	GGTTTCAGAG	22380
	GCCGTGTGTC	CAGGCCATTG	TGCAAAAGTC	AACCTGACCG	TTCCTTTCCT	TTTCTGGCCT	22440
40	GCGTGGTCTG	AAGCTCAGAG	CCCTCTCTTC	ACCCAGCCTG	TGTGTGTCTT	GCCGGACAGA	22500
	AGAAAAATGG	TGCTTTTTGC	GTGTTAGCAG	AGGTGCTTTT	CATGGCTGAC	CTCAACGCGT	22560
	CCATCTCCAG	CCTTGACCAA	GCTGTTTTTT	AGGGGCAAAC	GCAGGCAAGT	TCTGAATGCA	22620
45	CACAGTTATT	TCATGGTTAA	ACTATTCAGC	TTTGGCCGGG	CGCAGTGTGG	CTCTCACGCC	22680
	TGTCATCCCA	GCACTTTGGG	AGGCCGAGGC	GGGTGGATCA	CCTGAGGTCA	GGAGTTCGAG	22740
50	ACCAGCCTGG	CCAACACGGT	GAAACTCTAT	CTCTACTAAA	AATACAAAAA	TTAGCCGGGC	22800
	GTGGTGGTGT	GTATCTGTAA	TCCCAGCTAC	TCAGGAAGCT	GAGGCAGGAG	AATCGCTTGG	22860
	ACCCAGGAGG	CGGAGGTTGC	ACTGAGCCGA	GATCGCGCCA	TTGCACTCCA	GCCTGGGCGA	22920
55	CAGAGCCAGA	CGCTGTCTCA	AAAAAATGAA	TAATAAAATA	AAATAACAGG	AACTAAATAA	22980
	AATAAAACGT	TCAGCTTTGT	TCTGCAAATC	CACCTCTATT	GTTTTACGTG	GTTTGAGAGA	23040
60	CTCTGTCCCT	TAGAAATAGA	TGTTTGTTCG	CAATTGTAAT	GAATCTGTTT	CAAAAATGAA	23100
	CAGAATATTC	AAATGGTTTG	AGAGATCTTT	TCCCTTAGAA	ATAGCTTGTT	GCCAATCACA	23160
	AAGAATGTTT	TTCAAAAATG	AATGGAATCT	TCCTGGATAT	CGCTTCCAGA	TCTTCATTTT	23220
65	TTTTGCATAG	TTCAACCTGA	AAAGTAAGTG	TCTCAGCCCT	GAATTTCTTT	CTGATTTTTT	23280
	CATGGGTTGT	CTTGCAGACT	TCTCTGGACT	TGACCACATT	TAAAAAATAA	AAAATTAACT	23340

	TTTTACACAG GACACGGTTT CAATAGGAAT GAGATCTTTG AGTTTTTATG TAACAGATTC	23400
5	TTACCATCAG TTCTCAGATT CCCAAATTAC ACACAAAAAG CCACGGACTT CGCCTCCTGC	23460
	TAACATGTCC TTCTGTTTCT GAGGCTTCTG TTGGTGTTAG ACTTTCATGT TTAATAGCAG	23520
	ACAATGTAGG GATTTAAAGA AAAATGCAGA GAAAGCAAAA ACACTGACCA AACACACGGA	23580
10	GATAAGCTTT CTAAAGCCTT TGTTCCTTGA GTTGTCGTTA AAAAAAAAAA GTTGTTTTAA	23640
	ACTTTGCAAG CATGCCTATA TTGAACTCAT AAGCAAGAGA GCCAAGAAAA ATAGTGTCGG	23700
15	TCGTCTACTC TACACGTTTT CCCAAAACAG ACGTATTTTA ATTTCTTTTG TTTGAACTCA	23760
	CAGATGCTGA GAGTTAAAAG TTAAATTTTT GTCATGAACA ATAGTGGCCA AAACCACAGT	23820
	TACTTTTGCA CTATAGCATA ATAAGAAAAA TACAGGCTGG GCTCGGTGGC TCACACCTGT	23880
20	AATCAAAGCA CTTTTGGAGG CGAAACAGCC AGATCCCTTG AGCCCAGGAG ATTGAGACCA	23940
	GCCTGGGCAA CATAGCGAGA CCCTCATCTC TACAAAAAAG GTTTGTTACA TATGTAACAA	24000
25	ACCTGCACAT TGTGCACATG TACCCTAAAA CTAAAGTAT AATAATAAAA AAATTAAAAA	24060
	AAAATTCACC AATCAACTGC CTGCTGGTGC CTTCAAGAGA CTCACCTAAC ACATAAGGAC	24120
	TTGCATAAAC TTATAAAACA ATTCAATGGA AGAATCCTTG AAAGTATTCT GAGAAGACAG	24180
30	TATAATAAAC TGATTTCTAA AAAGGCTATA AAAAATTGAA TAAATCATTG TTGGGCATCC	24240
	TGTGCTGAAA TATAATGCAG CCAATAAAAA TTACAAAATG AATAAACATT TTATAACAAT	24300
35	AAAAAAAGT CAAATAATTA GGCAGGCATG GTGGTGCTCT CCTACGGTTG AAGCTATTCA	24360
	GCAGGCAAGA GGATACTTTG TTTTGTGTTT TTAATTTTTT TTGAGACAGA GTCTCGCTCT	24420
	GTGCGCAGGC TGGAGTGCAG TGGCGTGATC TCAGCTCACT GTAATTTCTG CCTCCCGGGT	24480
40	TCAAGCGATT TTCCTGCCCC AGCCTCCCGA GTAGCTGGGA TTACAGGTGC CCGCCACCAC	24540
	ACCTGGCTAA TTTCTTTTGT ATTTTGTAGT GAGACGAGGT TTCCCATGT TGGCCAGGCT	24600
45	GGTTTTGAGC TCCCGACCTC GGGTGATCCA CCCGCCTCAG CCTCCCAAAG TGCTGGGATG	24660
	ACAGGCGTGA GCCACCGCGC CTGGCCCAGG AGGATTATTT GATCCCAGGA GGTGGAGGCT	24720
	GCAGGAAGCC ATGATTGCAC CACTGCACTC CAGCCTGGCT GACAGAGTGA GACCACATCT	24780
50	CTAAATAAAT GAATAAATAC AGGCAGAAAC TTTTTTGTG TTGTTTTGAT GGAGTCTTGC	24840
	TCTGTCACCA GGCAGGAGTG CAGTGGTGCC ATCTCAGCTC ACTGCAACCT CCACCTCCTG	24900
55	GGTTCAAGCA ATCCTCCTGC CTCAGCCTCC CGAGTAGCTG GGATTACAGG TGCCCGCCAC	24960
	CACGCCCGGC TAATTTTTTG TATGTTTAGT AGAGACGGGA TTTCACCGTG TTAGCCAGGA	25020
	TGGTCTTGAT CTCTTGACTT TGTGATCTGC CTGCCTCAGC CTCCCAAAGT GCTGGGATTA	25080
60	CAGGCATGAG CCCAGGAGTT CAAGACCAGC CTCAGCAACA AAGTGAGACC TTTTCTCTCC	25140
	AAAAAATCAA AAATTTAGCC AGCTGTGGTG GCTCCTGCCC GTGATCCCAG TACTGTGGGA	25200
65	GGCTGAGGCA GAATTGCTTG AGCCCAGGAG TTCGAGACCA ACCTCAGCAA AAAGGACTCT	25260
	CTCTCTCTCT CTCTCTCTCT CTCTCTCTCT CTCTCTATAT ATATATATAT ATATATATAT	25320

	GAGTTTCAAA AATTGCTGGG TGACCAGCTC ATCTACTGGT TTTCCCTTG GGAAAGTGAA	25380
	ATTGTCATGT ATTGAAGATT TCCAAGGAAG TTGTATTGAA TGAGAAACAA ACTCAATCTG	25440
5	TTCGTGTTTA AAGAGCTGCA GTGCGTTTGC TGTGTTTCCC ATAAACTGC ACTTCCAAAA	25500
	GACACGCTGA GAAAGGAGAC CAGGATTGT AATTCAGAAA TTGGAAAGCA AGTTAGGCTG	25560
10	GACGTGGTAG CTCATGCTTG TTGTAATCTC AGCACTCTGG GAGGCTGAGG CAGGAGGATC	25620
	ACTTGAGCCC AGGAGTTCAA GACCAGCCCG TGCCACATGG TGAAACCCTG TCTCTCCAAA	25680
	AAATAAAACA TTTAGCCAGA TGTGGTGACT CATGCCTGTA ATCCCGGTAT TCTGGGAGGC	25740
15	TGAGGCAGAG TTGCTTGAGC CCAGGAGTTC AAGACCAGCC TCGGCAACAA AGTGAGACCC	25800
	TGTCTCTCCA AAAAATAAAA CATTTAGCCA GCTGTGGTGA CTCATGCCTG TAATCTCAGT	25860
20	ACTCTGGGAG GCTGGGGCAG AATGGCTTGA GCCCAGGAGT TCGAGACCAA CCTCAGCAAC	25920
	AAAGTGAGAT CTTGTTTCTC CAAAAATCA AAAATTTAGC CAGCTGTGCT GGCTCATGCC	25980
	TGTAATCCCG GTACTCTGGG AGGCTGAGGC AGAATCGTTT GAGCCCAGGA GTTCGAGACC	26040
25	AACCTCAGCA ACAAAGTGAG ATCTTGTTTC TCCAAAAAAA TCAAAAATTT AGCCAGCTGT	26100
	GCTGGCTGGT GCCTGTAATC CCGGTACTCT GGGAGGCTGA GCGGGAATTG CTTGAGCCCA	26160
30	GGAGTTCAAG ACCAGCCTCA GCAACAAAGT GAGATCTTGT TTCTCCAAAA AATAAACAT	26220
	TTAGTCAGCT GTGGTGGCTC AAGCCTGTGA TCCCAGCATT TTGGGAGGCC GAGGCGGGCG	26280
	GATCACGAGG TCATGAGATC GAGACCATCC TGGCTAACAC GGTGAAACCC CGTCTCTACT	26340
35	AAAAATACAA AGAAAATTAG CCGGGCGTGG TGGCGGGCGC CTGTAGTCCC AGCTACTCAG	26400
	GAGGCTGAGG CAGGAGAATG CCGTGAGCCT GGGAGGCGGA CCATGCAGTG AGTCAAGATC	26460
40	GCGCCACTGC CCTCCAGCCT GGGCCACAGA GCAAGACTCC GTCTCAAAAA AAAAAAAAAA	26520
	AAAACTGCTG CCCAACCTGT GTTTGCACCA CTGCCCTCCA GCCTGGGCAA CAGAGCAAGA	26580
	CTCCGTCTCA AAAAAAAAAA AATGCTGCCC AAGCTGTGTT TGCACCACTG CCCTCCAGCC	26640
45	TGGGCAACAG AGCAAGACTC CGTCTCAAAA AAAAAAAAAA AAAATGCTGC CCAAGCTGTG	26700
	TTTGCACCAC TGCCCTCCAG CCTGGGCAAC AGAGCAAGAC TCTGTCTCAA AAAAAAAAAA	26760
50	AATGCTGCCC AAGCTGTGTT TGCACCACTG CCCTCCGGCC TGGGCAACAG AGCAAGACTC	26820
	CGTCTCAAAA AAAAAAAAAA AATGCTGCCC AAGCTGTGTT TGCACCACTG CCCTCCAGCC	26880
	TGGGCAACAA AGCAAGCCTC AGCTTTCTGC CATCTCCACA ACCAAGAAAG CAATTCACAC	26940
55	AGAAATCAGT GCATCGTGCA GTGACCTCTT CAGAAAACCA ATGAGTTTTC CACCTGAGGA	27000
	ACTGTTTCTG AGCCCCATTC AGAAAAACAC ATCCCTGTAA CTGCAGGGCA GATTIACCTA	27060
60	CTGTATGCCT GTTTAAATAA AGCTTCCAGC CTCTGCATGG GGTCTGTCTG GAAGCTCCTG	27120
	TATCTGTCCC ACATTCTTGG AATCACAATG CACCCCTGGG AGGAAGATAT GTATTTAAAG	27180
	GGAGTGGATG TTATGGTGAG AAAATGCTGC CCATCCTTCT AGAAGACAAA AGCCACACAA	27240
65	AATACATCAC AAGAACCAGT TTTTTCAGA GAAGAACCTG CACAAAGAAC CTGCTCCCCC	27300
	CACACCCCCA CACACAGGTG AATTAACAGG ATGTATGTTT TATCATAAAA GCACAGGTTT	27360

GTTTCCTATG CACTCTCTGA GGATTGGCC ATATGCAAAG ATGTACAAAA ACCTTCTCTT 27420
 TCCCCAGGGA ACCGTAACCC GTCTGAAAAG ATGCCCTTCT CAGAAGCGAG TTGAACGATT 27480
 5 GTTGGAAAAG ATAAAATACG ACGTGACAC ACACAGTAGA GAAATGTCAC CCATGCAAAT 27540
 TATGTGTTTG AATGGAACAC ATTCAGGAAG CTAAATGGGG TATGACCACA CATTGGGGTT 27600
 10 GATTTATTTG ACGAGTGGAA GGGCAGATG GAAATGAATA CTGCTGTTT CCTTTGGAAG 27660
 GCCATATATG GGAATACCAA GAGGATTACT TTGGAAGTTT AGCTTCTCCA GGTGGTCTCT 27720
 CTCTCTCTCT CTTTTTTTGA GACAGAGTCT CACTCTGTCA CCCAGGCTGC AGTGCAATGG 27780
 15 CGTGCTCTCG GCTCACTGCA ACCTCAGCCT CCCAGGTACA AGCGATTCTC CTGCCTCAGC 27840
 CTCCCAGATA GCTGGGATCA CAGGTGTGCA CCACCACGCC TGGCTAATGT TTGTATTTTC 27900
 20 AGTAGAGATG AGGTTTTACC ATGTTGGCCA GGCTGGTCTT GAACTCCTGA CCTCAGGTGA 27960
 TCCGCCTGCC TCGGCCTCCC AAAGTGCTGG GATGACAGAC ATGAGCTAGC ACGCCCGGCC 28020
 CCAGGTGGTC TTTTAGCGG GTATTAAGC AGCTTCTCT CTGAGCCTTA AACCATGAAG 28080
 25 ATAGACAGAC TCAGTGATG GGTTTAGAG TTGTAATTT ATAAAAATAA GAAAAAGTCG 28140
 ACCTATCATT GATGGTTAGT ATTTTTGTG GCAGTTGCAT GCAATATTAG GATAAGGCAT 28200
 30 GTTCTCAAAA AGAACTCTT TTTTTTTTT TTTGAGACGG AGTCTCGCTC TGTCACCCAG 28260
 GCTGGAGTGC AGTGGCACGA TCTCCGCTCA CTGCAAGCTC CTCTTCCCGG GTTCACGCCA 28320
 TTCTCCTGCC TCAGCCTCCC CAGTAGCTGG GACTACAGGC GCCCCGCCACC ACGCCCGGCT 28380
 35 AATTTTTTGT ATTTTATAGTA GAGACGGGGT TTCACCATGT TAGCCAGGAA GGTCTCGATC 28440
 TCCTGACCTC ATGATCCGTC CGCCTCAGCC TCCCAAAGTG CTGGGACTAC AGGCGTGAGC 28500
 40 CACTGCACTT GGCCTTTTTT TTTTTTAGA TGGAGTTTTG CTCTTGTCGC CCAGGCTGGA 28560
 GTATAATGGC ATGATCTCGA CTCACTGCAA CCTCCGCCTC CCGAGTTCAA GCGATTCTCC 28620
 TGCCTCAGCC TCCCAGTAG CTGGGATTAC AGGTGCCAC CACCATGTCA AGATAATGTT 28680
 45 TGTAATTTCA GTAGAGATGG GGTGACCA TGTTGGCCAG GCTGGTCTCG AACTCCTGAC 28740
 CTCAGGTGAT CCACCCGCCT TAGCCTCCA AAGTGCTGGG ATGACAGGCG TGAGCCCCCTG 28800
 50 CGCCCGGCTT TTGTAACCTT ATTTTAAAT TTTTTTTTTT TTTAAGAAAG ACAGAGTCTT 28860
 GCTCTGTAC CCAGGCTGGA GCACACTGGT GCGATCATAG CTCACTGCAG CCTCAAATC 28920
 CTGGGCTCAA GCAATCCTCC CACCTCAGCC TCCTGAGTAG CTGGGACTAC AGGCACCCAC 28980
 55 CACCACACCC AGCTAATTTT TTTGATTTT ACTAGAGACG GGATCTTGCT TTGCTGCTGA 29040
 GGCTGGTCTT GAGCTCCTGA GCTCCAAAGA TCCTCTCACC TCCACCTCCC AAAGTGTAG 29100
 60 AATTACAAGC ATGAACCACT GCGGTGGTC TCCAAAAAA GGAATGTTAC GTGGATGTTT 29160
 TAGCTTCCTG TTCTCGTCTT TTTTGTGTA ATTGTACAGT TTGAGGGTGT GTGTGCGTGT 29220
 GCGCACGTGT GTGTGTGCAG TCTCCTGATT TCATGTATTT AATGTTATT ACCACCACCT 29280
 65 CCATCTCTCA TTCCTTCTTA CCTCACTGT GTAAAGATAC ATGTTGTTTT TAAATTTTAT 29340

	GTATTTATAT TTATTTATTT GTATTTCTGA GACAGAGTCT CACTCTGTTG CCCAGGCTAG	29400
	TGGCATGATC TCAGCTCACA GCAACCTTTG CCTCCTGGGT TCAAGCGATT CTCCTGCCTC	29460
5	AGCCTCCCGA GTAGCTGAGA TTACAGGCAC ACACCACCAC ACCCGGCTAG TTTTGTTTTG	29520
	AGACGGAGTC TCGCTCTGTT GCAGGCTGCA GTGCAGTGGC GTGATCCTGG CTCACTGCAA	29580
10	CCTCTGCCTC CTGGATTCAA GCGATTCTCC TGCCTCAGCC TCCCAAGTAG CTGGGATTAC	29640
	AGGCGCCAC CGCCACACCT GGCTAATTTT TTATTGGTAG TAGAGACGGG GTTTCTCCAT	29700
	GTTGACCAGA CTGGTCTTGA ACTCCCAACC TCGGGTGATC CACCCACCTG GGCCTCCCAA	29760
15	AGTGCTGGGA TGACAGGCGA GGGCCACCGC GTCCAGCCTT CTTCTTCTTC TTCTTTTTTT	29820
	TTTTTTTAAG ATGGAGTTTC ACTCTGTTGC CCAGGCTGGA GTGCAGTGGT GCAATCTCGG	29880
20	CTCCCTGCAA CCTCCACCTC CCAGGTTCAA GAAATCTTTT TGCCTCAGCC TCCCGAGTAG	29940
	CTGGGACTAC AGGTGCCCCG CACCACACCC ACCTAATGTT TGTATTTTTT TGGTAGAGAC	30000
	GGGGCTTCAC CACATTGGCC AGGCTGGTCT TGAACCTCTG ACTTCAGATG ATCCTCCTGC	30060
25	CTCAGCCTCC CAGAGTGTTG GGATTACAGG CGTGAGCCAC GGTGCCCGGC CAGACGTCAT	30120
	GTCTTAGGAA ATCAGAAAGT GGGTAGTTTC CGCACTCTGA GGAGAAAAAG AGACGTCCGG	30180
30	CGAAGAGAAA GGAGAGTGAA AGGATGTCTC CTCTTGCTG TAGCCTGTTC TCAATCGTGA	30240
	GTGAGCCAAT TGCCAGAAAC TGAGGGTGCT TCATTTGGCC AGGCAAGCTT CTCAACAGAA	30300
	TGTCTAAGTA CTTGTTAATG CTGAGAAGCT CTCCAAGCTA CTGCACTCCA GCCTGGGTGA	30360
35	CAGAGCACGA CCTTGCTCTGA AAACAATTAA TTAATCAATT AATTAATATA ATGAAATCAT	30420
	ACTGAACTCA GGAGACCATT GGGGTGGGCA GGGCTGGGGT TGGAAAGGAA CATAAATAT	30480
40	GGTGCAATGG ACTTTGCTCC AGTCTCCCTC CCCATCTCTT CTCGCCAAGA GTCTCTGGAG	30540
	GGAGCATGGG GAAGATGCTT TGGGAATCTG TAACTTCTTG TCTTGTAAC AGAATATCTA	30600
	AGTAATTGTT AATGCTGAGA AGTTATAGAT TTCCAAAGCC TTTCTCCAGG CTACGGACAA	30660
45	GGGTCATGGG TTAICTAGTG TTACAGAAAG AATGACATGG AGATGTTTGT TACATCTTAA	30720
	GGAACCATGA GGGGCCAGAG TATTTTACTC TAAGTGTAGA TGGTACATTG GCCACGCCTG	30780
50	TCCCAACACC ACCAATGGTG GCACCTAACT TTTGTGTTTG TGCCCCACAT TTCTTCTTCT	30840
	TTTCTGACGT AAATGCAAGT GATATTCCTT GGAAACCATG CTGCAGCAAG AGGCCATCTG	30900
	ACTACTAGTG ATACCCTGTA GCTCACCTAC AGCAGCTCAC TTGAAGCAGC TCACCCATAG	30960
55	CTCAGGTATA GCTCACCTGC AGCGGCTCAC CTGTAGCTCA CGTGTAGCTC ACTTGTAGCA	31020
	GCTCACTGGT AGCTCACCTG CAGCAGCTCA CCTGTACCTC ACCTGTACCT CACCTGCAGC	31080
60	AGCTCACCTG TAGCTCACCT GTACGTGAGC CACCGTACCC GGCCAGCAAG ACCCCATTTC	31140
	TAAATAAAT ACACAAAAAT TAGCCGGACG CGGTGGCGCG TGTCTGTAGT TGTAGCTACT	31200
	CAGGAGGCTG AGGTGGGAGG ATTGCTGGAG GCTGGGAGGT AGAGGCTGCA GTGAACCGTG	31260
65	ATCCAGCCAC TGTACTCTAG CCTGGATGAC ATAGCAAAAC CTTGTCTCAA AAAACAAAAA	31320
	CAAAAAACAA AACAAAGAAA CAAACAAAAA ACCCACACAC ACCGGAAAC AAAACAAAAA	31380

	GCAAAAAGGA AAGAAAAGAG AGCCAGGTCC CAAATATATA TTTCCTTGGA GAACCATTG	31440
5	CAAAGAGCAC ACTTAAAGGCC GGGCGCGGTG GCTCACGCCT GTCATCCCGG CACTTTGGGA	31500
	GGCCGAGGTG GGTGGATCAC GAGGTTGGGA GATCGAGACC ATCCTGGCCA ACATGGCGAA	31560
	ACCCCATCTC TACTAAAAAT ACAAAAAATC AGCCAGGTGC TGAGGCAGGT GCCTGTAGTC	31620
10	CCAGCCACTC AGGAGGCTGA GGCAGGAGAA TGGCATGAAC CTGGGAGGTG GAGGTTGCAG	31680
	TGAGCCGAGA TCGCGCCCCT GCACTCCAGC CTGGGCGACA GAGCGAGACT CCTTCTCAAA	31740
15	TAAATAAATA AATAATAAC AAAGAGCAAA CTTAAAATTG TCTCAGAAAT CCCACGGGAT	31800
	ATTGGATCTC CCTCATGCCT ATCTGATGAC ACTTTGAGTG TCTGGGGCCC CGTGCCTATT	31860
	TTCTGGGGTT CCCAGAAGCT GCCGTTCTGA AAGTGTGGCT CTCGGGGACG TGGCACAGGT	31920
20	GTGGATGTCT GTTTTAAATG TCAGGCGTTT GGACGTTGAG GAACGTGAGG CTGAAGGTCG	31980
	CCTTCGCCGA CCCCTGAGT TTAGGGTCCT GCCTTTTAAA ATCTTCCGAG CACTCTGTTG	32040
25	TTCACGCAAG CGTCCCATCT GTTTGGGTGG CCGTGCCGTC TGCATCTGTC TCGAACCTTC	32100
	ACAGCTTTGC AGAATATCCT GTTCTCAAT ACGGATGGAG AAACACGAGA CGCGTTTTCT	32160
	GGGTTATTTT AGCCGTCACG GAGAACCCCA GACTCATGTG TGCTAATGAC CTCATTAATG	32220
30	ATACTCTGAG GCAGACAGCC CTGCCTGATC TTAACAACAT TTTTAAATT TCTTTTTTGT	32280
	TTGTTGTTGT TACAGCATCA TTCATATAAC GTAGGAAACC GTGATCAGTA GCTTTTAGGA	32340
35	TATTTGCAAC AGGGTGTAAC ADAAABD	32367
	(2) SEQ ID NO:15 的资料:	
	(i) 序列特征:	
	(A) 长度: 806 个碱基对	
40	(B) 类型: 核酸	
	(C) 链型: 单链	
	(D) 拓扑结构: 线性	
45	(ii) 分子类型: 其他核酸	
	(A) 说明: /desc= "SHOT"	
	(ix) 特征:	
50	(A) 名称/关键词: CDS	
	(B) 位置: 43..615	
	(xi) 序列描述: SEQ ID NO:15:	
55	GTGTCCCCGG AGCTGAAAGA TCGCAAAGAG GATGCGAAAG GGATGGAGGA CGAAGGCCAG	60
	ACCAAAATCA AGCAGAGGCG AAGTCGGACC AATTTACCC TGGAACTACT CAATGAGCTG	120
	GAGAGGCTTT TTGACGAGAC CCACTATCCC GACGCCTTCA TGCAGAGGA ACTGAGCCAG	180
60	CGACTGGGCC TGTCGGAGGC CCGAGTGCGG GTTTGGTTTC AAAATCGAAG AGCTAAATGT	240
	AGAAAACAAG AAAATCAACT CCATAAAGGT GTTCTCATAG GGGCCGCCAG CCAGTTTGAA	300
65	GCTTGTAGAG TCGCACCTTA TGTCAACGTA GGTGCTTTAA GGATGCCATT TCAGCAGGTT	360
	CAGGCGCAGC TGCAGCTGGA CAGCGCTGTG GCGCACGCGC ACCACCACCT GCATCCGCAC	420

CTGGCCGCGC ACGCGCCCTA CATGATGTTC CCAGCACCGC CCTTCGGACT GCCGCTCGCC 480
 5 ACGCTGGCCG CGGATTCGGC TTCCGCCGCC TCGGTAGTGG CGGCCGCAGC AGCCGCCAAG 540
 ACCACCAGCA AGGACTCCAG CATCGCCGAT CTCAGACTGA AAGCCAAAAA GCACGCCGCA 600
 GCCCTGGGTC TGTGACVCCA ACGCCAGCAC CAATGTCGCG CCTGTCCCGC GGCACCTCAGC 660
 10 CTGCASNCCC TNDDKANMCG TTRCTYHTCM ATTACACTTT GGGACCYCGG GDBAGVCCTT 720
 TTNNAGACTT YVATKGGSCW CSCTGGBCCC TBRKGAVVAC TTGSGHYCGR GAACCGAKHT 780
 15 GCCCABAYGA GGACCRGTTT GGAADG 806

(2) SEQ ID NO:16 的资料:

(i) 序列特征:

- 20 (A) 长度: 190 个氨基酸
 (B) 类型: 氨基酸
 (C) 链型: 单链
 (D) 拓扑结构: 线性

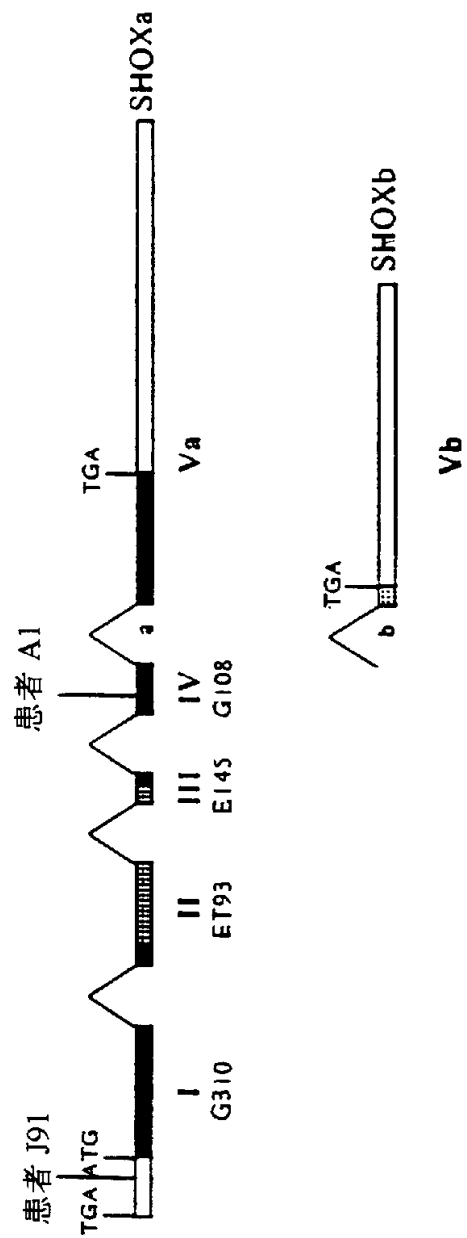
25 (ii) 分子类型: 肽

(xi) 序列描述: SEQ ID NO:16:

30 Met Glu Asp Glu Gly Gln Thr Lys Ile Lys Gln Arg Arg Ser Arg Thr
 1 5 10 15
 35 Asn Phe Thr Leu Glu Gln Leu Asn Glu Leu Glu Arg Leu Phe Asp Glu
 20 25 30
 Thr His Tyr Pro Asp Ala Phe Met Arg Glu Glu Leu Ser Gln Arg Leu
 35 40 45
 40 Gly Leu Ser Glu Ala Arg Val Gln Val Trp Phe Gln Asn Arg Arg Ala
 50 55 60
 45 Lys Cys Arg Lys Gln Glu Asn Gln Leu His Lys Gly Val Leu Ile Gly
 65 70 75 80
 Ala Ala Ser Gln Phe Glu Ala Cys Arg Val Ala Pro Tyr Val Asn Val
 85 90 95
 50 Gly Ala Leu Arg Met Pro Phe Gln Gln Val Gln Ala Gln Leu Gln Leu
 100 105 110
 Asp Ser Ala Val Ala His Ala His His His Leu His Pro His Leu Ala
 115 120 125
 55 Ala His Ala Pro Tyr Met Met Phe Pro Ala Pro Pro Phe Gly Leu Pro
 130 135 140
 60 Leu Ala Thr Leu Ala Ala Asp Ser Ala Ser Ala Ala Ser Val Val Ala
 145 150 155 160
 Ala Ala Ala Ala Ala Lys Thr Thr Ser Lys Asp Ser Ser Ile Ala Asp
 165 170 175
 65 Leu Arg Leu Lys Ala Lys Lys His Ala Ala Ala Leu Gly Leu
 180 185 190

说明书附图

图 1



SHOXa

图 2

```

1   GTGATCCACCCGCGCGCACGGGCGCTCCTCTCCGCGCGGGGAGACGCGCGCATCCACCAG
61  CCCC GGCTGCTCGCCAGCCCCGGCCCCAGCCATGGAAGAGCTCACGGCTTTTGTATCCAA
    M E E L T A F V S K
121 GTCTTTTGACCAGAAAAGCAAGGACGGTAACGGCGGAGGCGGAGGCGGCGGAGGTAAGAA
11  S F D Q K S K D G N G G G G G G G G G K K
181 GGATTCCATTACGTACCGGGAAGTTTGGAGAGCGGACTGGCGCGCTCCCGGGAGCTGGG
31  D S I T Y R E V L E S G L A R S R E L G
241 GACGTGCGATTCCAGCCTCCAGGACATCACGGAGGGCGGCGGCCACTGCCCGGTGCATTT
51  T S D S S L Q D I T E G G G H C P V H L
301 GTTCAAGGACCACGTAGACAATGACAAGGAGAACTGAAAGAATTCGGCACCGCGAGAGT
71  F K D H V D N D K E K L K E F G T A R -V
361 GGCAGAAGGGATTATGAATGCAAAGAGAAGCGCGAGGACGTGAAGTCGGAGGACGAGGA
91  A E G I Y E C K E K R E D V K S E D E D
421 CGGGCAGACCAAGCTGAAACAGAGGGCGAGCCGACCAACTTCACGCTGGAGCAGCTGAA
111 G - Q T K L K Q - R - R - S R T N F T L E - Q L N
481 CGAGCTCGAGCGACTTTTTGACGAGACCCATTACCCCGACGCCTTCATGCGCGAGGAGCT
131 E L E R L F D E T H Y P D A F M R E E L
541 CAGCCAGCGCCTGGGGCTTTCCGAGGCGCGCTGCGAGTTTGGTTCCAGAACC GGAGAGC
151 S Q R L G L S E A R V Q V W F Q N R R A
601 CAAGTGCCGCAAACAAGAGATCAGATGCATAAAGCGTCATCTTGGGCACAGCCAACCA
171 K C R K Q E N Q M H K G V I L G T A N H
661 CCTAGACGCCTGCCGAGTGGCACCCCTACGTCAACATGGGAGCCTTACGGATGCCTTTCCA
191 L D A C R V A P Y V N M G A L R M P F Q
721 ACAGTTCAGGCTCAGCTGCAGCTGGAAGGCGTGGCCCCACGCGCACCCGCGACCTGCACCC
211 Q V Q A Q L Q L E G V A H A H P H L H P
781 GCACCTGGCGGCGCACGCGCCCTACCTGATGTTCCCCCGCGCCCTTCGGGCTGCCCCAT
231 H L A A H A P Y L M F P P P P F G L P I
841 CGCGTCGCTGGCCGAGTCCGCCTCGGCCGCGCGCGTGGTGGCGCGCGCGCCAAAAGCAA
251 A L S A E S A S A A A V V A A A A K S N
901 CAGCAAGAATTCCAGCATCGCCGACCTGCGGCTCAAGGCGCGGAAGCACGCGGAGGCCCT
271 S K N S S I A D L R L K A R K H A E A L
961 GGGGCTCTGACCCGCGCGCAGCCCCCGCGCGCGCGGACTCCCGGGCTCCGCGCACCCC
291 G L *
1021 GCCTGCACCGCGCGTCTGCACTCAACCCCGCCTGGAGCTCCTTCCGCGGCCACCGTGCT
1081 CCGGCCACCCCGGGAGCTCCTGCAAGAGGCCTGAGGAGGGAGGCTCCCGGGACCGTCCAC
1141 GCACGACCCAGCCAGACCCTCGCGGAGATGGTGCAGAAGGCGGAGCGGGTGAGCGGCCGT
1201 GCGTCCAGCCCGGGCCTCTCCAAGGCTGCCCCGTGCGTCTTGGGACCCTGGAGAAGGGTAA
1261 ACCCCCGCCTGGCTGCGTCTTCTCTGCTATACCCTATGCATGCGGTAACTACACACGT
1321 TTGGAAGATCCTTAGAGTCTATTGAACTGCAAAGATCCCGGAGCTGGTCTCCGATGAAA
1381 ATGCCATTTCTTCTGTGCCAACGATTTTCTTTACTACCATGCTCCTTCCTTCATCCCGAG
1441 AGGCTGCGGAACGGGTGTGGATTGTAATGTGGACTTCGGAATCCCAGGAGGCAGGGGCCG
1501 GGCTCTCCTCCACCGCTCCCCCGGAGCCTCCCAGGCAGCAATAAGGAAATAGTTCTCTGG
1561 CTGAGGCTGAGGACGTGAACCGCGGGCTTTGGAAAGGGAGGGGAGGGAGACCCGAACCTC
1621 CCACGTTGGGACTCCACGTTCCGGGGACCTGAATGAGGACCGACTTTATAACTTTTCCA
1681 GTGTTTGATTCCCAAATTGGGTCTGGTTTTGTTTTGGATTGGTATTTTTTTTTTTTTTTT
1741 TTTTGTGTGTGTACAGGATTCAGACGCAAAAGACTTGCATAAGAGACGGACGCGTGGTT
1801 GCAAGGTGTCTACTGATATGCAGCATTAACCTTTACTGACATGGAGTGAAGTGAATATT
1841 ATAAATATTATAGATTAAAAAAAATAGC [A]

```

```

1      GTGATCCACCCGCGCGCACGGGCCGTCCTCTCCGCGCGGGGAGACGCGCGCATCCACCAG
61     CCCCCGGCTGCTCGCCAGCCCCGGCCCCAGCCATGGAAGAGCTCACGGCTTTTGTATCCAA
1      M E E L T A F V S K
121    GTCTTTTGACCAGAAAAGCAAGGACGGTAACGGCGGAGGCGGCGGAGGTAAGAA
11     S F D Q K S K D G N G G G G G G G G K K
181    GGATTCCATTACGTACCGGGAAGTTTGGAGAGCGGACTGGCGCGCTCCCGGGAGCTGGG
31     D S I T Y R E V L E S G L A R S R E L G
241    GACGTCGGATTCCAGCCTCCAGGACATCACGGAGGGCGGCGGCCACTGCCCGGTGCATTT
51     T S D S S L Q D I T E G G G H C P V H L
301    GTTCAAGGACCACGTAGACAATGACAAGGAGAACTGAAAGAATTGGGCACCGCGAGAGT
71     F K D H V D N D K E K L K E F G T A R V
361    GGCAGAAGGGATTATGAATGCAAAGAGAAGCGCGAGGACGTGAAGTCGGAGGACGAGGA
91     A E G I Y E C K E K R E D V K S E D E D
421    CGGGCAGACCAAGCTGAAACAGAGGCGCAGCCGACCAACTTCACGCTGGAGCAGCTGAA
111    G Q T K L K Q R R S R T N F T L E Q L N
481    CGAGCTCGAGCGACTTTTGGACGAGACCCATTACCCCGACGCTTCATGCGCGAGGAGCT
131    E L E R L F D E T H Y P D A F M R E E L
541    CAGCCAGCGCCTGGGGCTTTCCGAGGCGCGCGTGCAGGTTTGGTTCCAGAACCGGAGAGC
151    S Q R L G L S E A R V Q V W F Q N R R A
601    CAAGTGGCGCAAACAAGAGATCAGATGCATAAAGGCGTCATCTTGGGCACAGCCAACCA
171    K C R K Q E N Q M H K G V I L G T A N H
661    CCTAGACGCCTGCCGAGTGGCACCCTACGTCAACATGGGAGCCTTACGGATGCCTTTCCA
191    L D A C R V A P Y V N M G A L R M P F Q
721    ACAGATGGAGTTTGTCTCTTGTGCCCCAGGCTGGAGTATAATGGCATGATCTCGACTCAC
211    Q M E F C S C R P G W S I M A *
781    TGCAACCTCCGCCTCCCGAGTTCAAGCGATTCTCCTGCCTCAGCCTCCCGAGTAGCTGGG
841    ATTACAGGTGCCCCACCACCATGTCAAGATAATGTTTGTATTTTCAGTAGAGATGGGGTTT
901    GACCATGTTGGCCAGGCTGGTCTCGAACTCCTGACCTCAGGTGATCCACCCGCCTTAGCC
961    TCCCAAAGTGCTGGGATGACAGGCGTGAGCCCCTGCGCCCGGCCTTTGTAACTTTATTTT
1021  TAATTTTTTTTTTTTAAAGAAAGACAGAGTCTTGCTCTGTCACCCAGGCTGGAGCACA
1081  CTGGTGCGATCATAGCTCACTGCAGCCTCAAACCTCCTGGGCTCAAGCAATCCTCCACCT
1141  CAGCCTCCTGAGTAGCTGGGACTACAGGCACCCACCACACCCAGCTAATTTTTTTTGA
1201  TTTTACTAGAGACGGATCTTGCTTTGCTGCTGAGGCTGGTCTTGAGCTCCTGAGCTCC
1261  AAAGATCCTCTCACCTCCACCTCCCAAAGTGTTAGAATTACAAGCATGAACCACTGCCCC
1321  TGGTCTCCAAAAAAGGACTGTTACGTGG [A]

```

图 4

GTGTCCCCGAGCTGAAAGATCGCAAAGAGGATGCGAAAGGGATGGAGGACGAAGGCCAG
M E D E G Q

ACCAAAATCAAGCAGAGGGCGAAGTCGGACCAATTTACCCTGGAACAACTCAATGAGCTG
T K I K

GAGAGGCTTTTTGACGAGACCCACTATCCCGACGCCTTCATGCGAGAGGAACTGAGCCAG

CGACTGGGCCTGTGCGAGGCCCGAGTGCAGGTTTGGTTTCAAAATCGAAGAGCTAAATGT

AGAAAAACAAGAAATCAACTCCATAAGGTGTTCTCATAGGGGGCCGCCAGCCAGTTTGAA
N Q L H K G V L I G A A S Q F E

GCTTGTAGAGTCGCACCTTATGTCAACGTAGGTGCTTTAAGGATGCCATTTACAGCAGGAT
A C R V A P Y V N V G A L R M P F Q Q D

AGTCATTGCAACGTGACGCCCTTGCCCTTTCAGGTTTCAGGCGCAGCTGCAGCTGGACAGC
S H C N V T P L P F Q V Q A Q L Q L D S

GCTGTGGCGCACGCGCACCACCACCTGCATCCGCACCTGGCCGCGCACGCGCCCTACATG
A V A H A H H H L H P H L A A H A P Y M

ATGTTCCAGCACCGCCCTTCGGACTGCGGCTCGCCACGCTGGCCGCGGATTTCGGCTTCC
M F P A P P F G L P L A T L A A D S A S

GCCGCCTCGGTAGTGGCGGCCGAGCAGCCGCCAAGACCACCAGCAAGGACTCCAGCATC
A A S V V A A A A A A K T T S K D S S I

GCCGATCTCAGACTGAAAGCCAAAAAGCACGCCGAGCCCTGGGTCTGTGACGCCAACGC
A D L R L K A K K H A A A L G L *

CAGCACCAATGTGCGCCTGTCCCGCGGCACTCAGCCTGCACGCCCTCCGCGCCCCGCTG
CTTCTCCGTTACCCCTTTGAGACCTCGGGAGCCGCCCTCTTCCCGCCTCACTGACCATC
CCTCGTCCCCTATCGCATCTTGGACTCGGAAAGCCAGACTCCACGCAGGACCAGGGATCT
CACGAGGCACGCAGGCTCCGTGGCTCCTGCCGTTTTCTACTCGAGGGCCTAGAATTGG
GTTTTGTAGGAGCGGGTTTGGGGGAGTCTGGAGAGAGACTGGACAGGGTAGTGCTGGAAC
CGCGGAGTTTGGCTCACCGCAAAGCTACAACGATGGACTCTTGATAGAAAAAAAATC
TTGTTAACAATGAAAAATGAGCAACAAAAAATCGAAAGACAAACGGGAGAGAAAAAG
AGGAAGGCAACTTATTTCTTAAGTCTATTTGGCAGAAGCTGAAATTGGAGAACCAAGGA
GCAAAAACAAATTTTAAATTAAGTATTTTATACATTTAAAAATATGAAAAACAACCC
AGACGATTCTCGAGAGACTGGGGGGAGTTACCAACTTAAATGTGTGTTTTAAAAATGCG
CTAAGAAGGCAAGCAGAAAGAAGAGGTATACTTATTTAAAAAATAAGATGAAAAAAGT
GCGCAGGTGGGAAGTTACAGGTTTTGAAACTGACCTTTTTCTGCGAAGTTCACGTTAAT
ACGAGAAATTTGATGAGAGAGGCGGCCCTCTTTTACGTTGAATCAGATGCTTTGAGTTT
AAACCCACCATGTATGGAAGAGCAAGAAAAGAGAAAAATATTAAACGAGGAGAGAGAAAA
ATAATGGCAAACTGTCTGGACTGCTGACAGTAAATTCGGTTTGCATGGAAAAA
AAAAAAAAAAAAAAAAA

人 SHOX 基因的外显子/内含子结构

外显子	cDNA-a	cDNA-b	基因组 DNA	外显子大小	内含子/外显子	外显子/内含子
I	UTR-368	UTR-368	1-368	368	GTGATCCACC	GTGGCAGAAAGgtaagtccct
II	369-577	369-577	3817-4025	209	ccccacgcagGGATTATGA	GCGGTGCAGgtaggaaccc
III	578-635	578-635	9851-9908	58	tctccccaaagTTTGGTTCC	ATGCATAAAAGgtgggtgtcg
IV	636-724	636-724	10029-10117	89	ttggacacagGCGTCATCTT	TTTCCAACAGgtagctcact
Va	725-1890	----	13364-14529	1166	gctccccgcagGTCCAGGCTC	AAAAAATAGC
Vb	----	725-1349	27154-27778	625	ttttttttttagATGGAGTTTT	TGTTACGTGG
polyA	> 1891	> 1350	----			

外显子大小以碱基对表示；外显子序列用大写字母表示；供体和受体的剪接位点加下划线。基因组和 cDNA 序列可以从 GenBank 编号 no.xy/获得。

图 5

SHOX