

[19] 中华人民共和国国家知识产权局



[12] 发明专利申请公布说明书

[21] 申请号 200580044176.6

[51] Int. Cl.

A01K 67/00 (2006.01)

A01K 67/033 (2006.01)

A01K 67/027 (2006.01)

C12N 15/00 (2006.01)

[43] 公开日 2009年3月18日

[11] 公开号 CN 101389214A

[22] 申请日 2005.10.24

[21] 申请号 200580044176.6

[30] 优先权

[32] 2004.10.22 [33] US [31] 60/621,433

[86] 国际申请 PCT/US2005/038582 2005.10.24

[87] 国际公布 WO2006/047603 英 2006.5.4

[85] 进入国家阶段日期 2007.6.21

[71] 申请人 雷维维科公司

地址 美国弗吉尼亚州

[72] 发明人 K·威尔斯 D·阿亚雷斯

[74] 专利代理机构 中国专利代理(香港)有限公司

代理人 权陆军 梁 谋

权利要求书 8 页 说明书 163 页 附图 7 页

[54] 发明名称

具有基因修饰免疫系统的有蹄类动物

[57] 摘要

本发明提供了缺少功能内源免疫球蛋白基因座表达的有蹄类动物、组织和器官以及来自此类动物、组织和器官的细胞和细胞系。本发明还提供了表达异种的(如人)免疫球蛋白基因座的有蹄类动物、组织和器官以及来自此类动物、组织和器官的细胞和细胞系。本发明还提供了有蹄类动物,如猪的重链和轻链免疫球蛋白的基因组 DNA 序列。此类动物、组织、器官和细胞可以用于研究和医疗目的。此外,提供了制备此类动物、器官、组织和细胞的方法。

1.转基因有蹄类动物，其缺少功能性内源免疫球蛋白的任何表达。

2.权利要求1的转基因有蹄类动物，其中该有蹄类动物缺少内源重链免疫球蛋白的任何表达。

3.权利要求1的转基因有蹄类动物，其中该有蹄类动物缺少内源轻链免疫球蛋白的任何表达。

4.权利要求3的转基因有蹄类动物，其中该有蹄类动物缺少内源 κ 链免疫球蛋白的任何表达。

5.权利要求3的转基因有蹄类动物，其中该有蹄类动物缺少内源 λ 链免疫球蛋白的任何表达。

6.权利要求1的转基因有蹄类动物，其中该有蹄类动物选自猪、牛、绵羊和山羊。

7.权利要求6的转基因有蹄类动物，其中该有蹄类动物是猪。

8.权利要求1的转基因有蹄类动物，其中通过核转移产生该有蹄类动物。

9.权利要求1的转基因有蹄类动物，其中该有蹄类动物表达外源免疫球蛋白基因座。

10. 权利要求9的转基因有蹄类动物，其中所述外源免疫球蛋白基因座是重链免疫球蛋白或其片段。

11. 权利要求9的转基因有蹄类动物，其中所述外源免疫球蛋白基因座是轻链免疫球蛋白或其片段。

12. 权利要求11的转基因有蹄类动物，其中所述轻链基因座是 κ 链基因座或者其片段。

13. 权利要求11的转基因有蹄类动物，其中所述轻链基因座是 λ 链基因座或者其片段。

14. 权利要求9的转基因有蹄类动物，其中所述异种基因座是人免疫球蛋白基因座或者其片段。

15. 权利要求9的转基因有蹄类动物，其中人工染色体含有所述异种免疫球蛋白。

15. 权利要求15的转基因有蹄类动物，其中人工染色体包含哺乳动物人工染色体。

16. 权利要求 15 的转基因有蹄类动物, 其中所述哺乳动物人工染色体包含人 14 号染色体、人 2 号染色体和人 22 号染色体的一个或多个或者其片段。

17. 转基因哺乳动物, 其缺少内源入链免疫球蛋白的任何表达。

18. 表达异种免疫球蛋白基因座或者其片段的转基因有蹄类动物, 其中从整合在内源有蹄类动物染色体中的免疫球蛋白基因座表达该免疫球蛋白。

19. 权利要求 18 的转基因有蹄类动物, 其中所述异种免疫球蛋白是人免疫球蛋白或者其片段。

20. 权利要求 18 的转基因有蹄类动物, 其中所述异种免疫球蛋白基因座被后代遗传。

21. 权利要求 18 的转基因有蹄类动物, 其中所述异种免疫球蛋白基因座通过雄性种系被后代遗传。

22. 权利要求 18 的转基因有蹄类动物, 其中所述有蹄类动物是猪、绵羊、山羊或者奶牛。

23. 权利要求 22 的转基因有蹄类动物, 其中所述有蹄类动物是猪。

24. 权利要求 18 的转基因有蹄类动物, 其中通过核转移产生所述有蹄类动物。

25. 权利要求 18 的转基因有蹄类动物, 其中应答对一种或多种抗原的暴露, 所述免疫球蛋白基因座在 B 细胞中表达以产生异种免疫球蛋白。

26. 权利要求 18 的转基因有蹄类动物, 其中人工染色体包含所述异种免疫球蛋白。

27. 权利要求 18 的转基因有蹄类动物, 其中人工染色体包含哺乳动物人工染色体。

28. 权利要求 27 的转基因有蹄类动物, 其中所述人工染色体包含酵母人工染色体。

29. 权利要求 26 的转基因有蹄类动物, 其中所述人工染色体包含人 14 号染色体、人 2 号染色体和人 22 号染色体的一个或多个或者其片段。

30. 来自权利要求 1 的转基因有蹄类动物的转基因有蹄类动物

细胞、组织或者器官。

31. 来自权利要求 18 的转基因有蹄类动物的转基因有蹄类动物细胞、组织或者器官。

32. 权利要求 30 或者 31 的细胞，其中该细胞是体细胞、繁殖细胞或生殖细胞。

33. 权利要求 32 的细胞，其中所述细胞是 B 细胞。

34. 权利要求 33 的细胞，其中所述细胞是成纤维细胞。

35. 包含异种免疫球蛋白基因座的猪动物。

36. 权利要求 35 的猪，其中人工染色体含有所述异种基因座。

37. 权利要求 36 的猪，其中所述人工染色体包含一个或多个异种免疫球蛋白基因座，其经历重排并且应答对一种或多种抗原的暴露可以产生异种免疫球蛋白。

38. 衍生自权利要求 35 的动物的猪细胞。

39. 权利要求 36 的猪细胞，其中该细胞是体细胞、B 细胞或者成纤维细胞。

40. 权利要求 35 的猪，其中所述异种免疫球蛋白是人免疫球蛋白。

41. 权利要求 36 的猪，其中所述一个或多个个人工染色体包含哺乳动物人工染色体。

42. 权利要求 41 的猪，其中所述哺乳动物人工染色体包含人 14 号染色体、人 2 号染色体和人 22 号染色体的一个或多个或者其片段。

43. 产生异种抗体的方法，该方法包括步骤：(a) 对有蹄类动物施用一种或多种目的抗原，所述有蹄类动物的细胞包含一个或多个个人工染色体并且缺少功能性内源免疫球蛋白的任何表达，每个人工染色体包含经历重排的一个或多个异种免疫球蛋白基因座，从而导致产生针对所述一种或多种抗原的异种抗体；和 (b) 从该有蹄类动物回收所述异种抗体。

44. 权利要求 43 的方法，其中所述免疫球蛋白基因座在 B 细胞中经历重排。

45. 权利要求 43 的方法，其中所述异种免疫球蛋白基因座是重链免疫球蛋白或者其片段。

46. 权利要求 43 的方法，其中所述异种免疫球蛋白基因座是轻

链免疫球蛋白或者其片段。

47. 权利要求 43 的方法，其中所述异种基因座是人免疫球蛋白基因座或者其片段。

48. 权利要求 43 的方法，其中人工染色体含有所述异种免疫球蛋白。

49. 权利要求 48 的方法，其中所述人工染色体包含哺乳动物人工染色体。

50. 权利要求 49 的方法，其中所述哺乳动物人工染色体包含人 14 号染色体、人 2 号染色体和人 22 号染色体的一个或多个或者其片段。

51. 分离的核苷酸序列，其包含猪重链免疫球蛋白或者其片段，其中该重链免疫球蛋白包括至少一个连接区和至少一个恒定免疫球蛋白区。

52. 权利要求 51 的核苷酸序列，其中所述重链免疫球蛋白包含至少一个可变区、至少两个多样性区、至少四个连接区和至少一个恒定区。

53. 权利要求 52 的核苷酸序列，其中所述重链免疫球蛋白包含 Seq ID No. 29。

54. 权利要求 51 的核苷酸序列，其中所述重链免疫球蛋白包含 Seq ID No. 4。

55. 权利要求 53 或 54 的核苷酸序列，其中所述序列与 Seq ID Nos 4 或 29 至少 80、85、90、95、98 或 99% 同源。

56. 权利要求 53 或 54 的核苷酸序列，其中所述序列含有 Seq ID No. 4 的至少 17、20、25 或者 30 个连续核苷酸或者 Seq ID No. 29 的残基 1-9,070。

57. 权利要求 53 或 54 的核苷酸序列，其中所述序列包含 Seq ID No. 29 的残基 9,070-11039。

58. 分离的核苷酸序列，其与 Seq ID No. 4 或 29 杂交。

59. 寻靶载体，其包含：

(a) 第一个核苷酸序列，其包含与 SEQ ID No 29 同源的至少 17 个连续核酸；

(b) 选择标记基因；和

(c) 第二个核苷酸序列, 其包含与 SEQ ID No 29 同源的至少 17 个连续核酸, 其不与第一个核苷酸序列重叠。

60. 权利要求 59 的寻靶载体, 其中所述选择标记包含抗生素抗性基因。

61. 权利要求 59 的寻靶载体, 其中所述第一个核苷酸序列代表 5' 重组臂。

62. 权利要求 59 的寻靶载体, 其中所述第二个核苷酸序列代表 3' 重组臂。

63. 用权利要求 59 的寻靶载体转染的细胞。

64. 权利要求 63 的细胞, 其中使得猪重链免疫球蛋白基因座的至少一个等位基因失活。

65. 包含权利要求 64 的细胞的猪动物。

66. 分离的核苷酸序列, 其包含有蹄类动物 κ 轻链免疫球蛋白基因座或者其片段。

67. 权利要求 66 的核苷酸序列, 其中所述有蹄类动物是猪。

68. 权利要求 66 的核苷酸序列, 其中所述有蹄类动物 κ 轻链免疫球蛋白基因座包含至少一个连接区、一个恒定区和/或一个增强子区。

69. 权利要求 66 的核苷酸序列, 其中所述核苷酸序列包含至少五个连接区、一个恒定区和一个增强子区。

70. 权利要求 69 的核苷酸序列, 其包含 Seq ID No. 30。

71. 权利要求 69 的核苷酸序列, 其包含 Seq ID No. 12。

72. 权利要求 70 或 71 的核苷酸序列, 其中所述序列含有 Seq ID No 12 或 30 的至少 17、20、25 或者 30 个连续核苷酸。

73. 分离的核苷酸序列, 其与 Seq ID No 12 或 30 杂交。

74. 寻靶载体, 其包含:

(a) 第一个核苷酸序列, 其包含与 SEQ ID No 30 同源的至少 17 个连续核酸;

(b) 选择标记基因; 和

(c) 第二个核苷酸序列, 其包含与 SEQ ID No 30 同源的至少 17 个连续核酸, 其不与第一个核苷酸序列重叠。

75. 权利要求 74 的寻靶载体, 其中所述选择标记包含抗生素抗

性基因。

76. 权利要求 74 的寻靶载体，其中所述第一个核苷酸序列代表 5' 重组臂。

77. 权利要求 74 的寻靶载体，其中所述第二个核苷酸序列代表 3' 重组臂。

78. 用权利要求 74 的寻靶载体转染的细胞。

79. 权利要求 78 的细胞，其中使得 κ 链免疫球蛋白基因座的至少一个等位基因失活。

80. 包含权利要求 79 的细胞的猪动物。

81. 分离的核苷酸序列，其包含有蹄类动物 λ 轻链免疫球蛋白基因座。

82. 权利要求 81 的核苷酸序列，其中所述有蹄类动物是猪。

83. 权利要求 81 的核苷酸序列，其中所述有蹄类动物是牛。

84. 权利要求 81 的核苷酸序列，其中所述有蹄类动物 λ 轻链免疫球蛋白基因座包含 J 到 C 单位的多联体。

85. 权利要求 81 的核苷酸序列，其中所述有蹄类动物 λ 轻链免疫球蛋白基因座包含至少一个连接区 - 恒定区对和/或至少一个可变区，如通过 Seq ID No. 31 表示的。

86. 权利要求 82 的核苷酸序列，其包含 Seq ID No. 28。

87. 权利要求 83 的核苷酸序列，其包含 Seq ID No. 31。

88. 权利要求 86 或 87 的核苷酸序列，其中所述序列含有 Seq ID No 28 或 31 的至少 17、20、25 或者 30 个连续核苷酸。

89. 分离的核苷酸序列，其与 Seq ID No 28 或 31 杂交。

90. 寻靶载体，其包含：

(a) 第一个核苷酸序列，其包含与 SEQ ID No 28 或 31 同源的至少 17 个连续核酸；

(b) 选择标记基因；和

(c) 第二个核苷酸序列，其包含与 SEQ ID No 28 或 31 同源的至少 17 个连续核酸，其不与第一个核苷酸序列重叠。

91. 权利要求 90 的寻靶载体，其中所述选择标记包含抗生素抗性基因。

92. 权利要求 90 的寻靶载体，其中所述第一个核苷酸序列代表

5' 重组臂。

93. 权利要求 90 的寻靶载体, 其中所述第二个核苷酸序列代表 3' 重组臂。

94. 用权利要求 90 的寻靶载体转染的细胞。

95. 权利要求 94 的细胞, 其中使得 λ 链免疫球蛋白基因座的至少一个等位基因失活。

96. 包含权利要求 95 的细胞的猪动物。

97. 环化至少 100kb DNA 的方法, 其中该 DNA 然后通过位点特异性重组酶整合到宿主基因组中。

98. 权利要求 97 的方法, 其中可以环化至少 100、200、300、400、500、1000、2000、5000、10,000 kb 的 DNA。

99. 权利要求 97 的方法, 其中通过在 DNA 序列的每个末端附着位点特异性重组酶靶位点然后对该 DNA 序列应用位点特异性重组酶, 可以实现 DNA 的环化。

100. 权利要求 97 的方法, 其中所述位点特异性重组酶靶位点是 Lox。

101. 权利要求 97 的方法, 其中人工染色体含有所述 DNA 序列。

102. 权利要求 101 的方法, 其中所述人工染色体是酵母菌人工染色体或者哺乳动物人工染色体。

103. 权利要求 101 的方法, 其中所述人工染色体包含编码人免疫球蛋白基因座或者其片段的 DNA 序列。

104. 权利要求 103 的方法, 其中所述人免疫球蛋白基因座或者其片段包含人 14 号染色体、人 2 号染色体、和/或人 22 号染色体。

105. 缺少内源免疫球蛋白的至少一个等位基因的表达的转基因有蹄类动物, 其中该免疫球蛋白选自重链、 κ 轻链和 λ 轻链或者其任意组合。

106. 权利要求 105 的转基因有蹄类动物, 其中表达异种免疫球蛋白。

107. 产生权利要求 106 的转基因有蹄类动物的方法, 其中将所述缺少内源免疫球蛋白的至少一个等位基因的表达的转基因有蹄类动物与表达异种免疫球蛋白的有蹄类动物育种, 其中所述内源免疫球蛋白选自重链、 κ 轻链和 λ 轻链或者其任意组合。

108. 权利要求 105 - 107 任一项的转基因有蹄类动物，其中该有蹄类动物是猪。

109. 权利要求 106 或 107 的转基因有蹄类动物，其中所述异种免疫球蛋白是人免疫球蛋白基因座或者其片段。

110. 权利要求 109 的转基因有蹄类动物，其中人工染色体含有所述人免疫球蛋白基因座或者其片段。

111. 来自权利要求 105 的有蹄类动物的细胞。

112. 权利要求 1、18、105 或 106 的转基因有蹄类动物，其还包含额外的基因修饰以消除异种抗原的表达。

113. 权利要求 112 的转基因有蹄类动物，其中该有蹄类动物缺少 α -1,3-半乳糖基转移酶基因的至少一个等位基因的表达。

114. 权利要求 112 的转基因有蹄类动物，其中该有蹄类动物是猪。

具有基因修饰免疫系统的有蹄类动物

本申请要求2004年10月22日提交的美国临时申请号60/621,433的优先权，将所述临时申请完整引入本文作为参考。

发明领域

本发明提供了缺少功能内源免疫球蛋白基因座表达的有蹄类动物、组织和器官以及来自此类动物、组织和器官的细胞和细胞系。本发明还提供了表达异种的（如人）免疫球蛋白基因座的有蹄类动物、组织和器官以及来自此类动物、组织和器官的细胞和细胞系。本发明还提供了有蹄类动物，如猪的重链和轻链免疫球蛋白的猪基因组DNA序列。此类动物、组织、器官和细胞可以用于研究和医学治疗。此外，提供了制备此类动物、器官、组织和细胞的方法。

发明背景

抗原是被身体识别为“外来的”试剂或者物质。通常它是作为抗原的更大的外来物质的仅仅一个相对较小的化学组，如细菌的细胞壁的组分。多数抗原是蛋白质，尽管糖类可以作为弱抗原。细菌、病毒和其他微生物通常含有许多抗原，花粉、尘螨、霉菌、食物和其他物质也一样。身体通过产生抗体对抗原产生反应。抗体（也称作免疫球蛋白（Ig））是免疫系统的细胞产生的蛋白质，其结合抗原或者外来蛋白质。抗体在血液的血清中循环以检测外来抗原并构成了血液蛋白质的 γ 球蛋白部分。这些抗体与抗原以高度特异的方式发生化学相互作用，像拼图玩具的两片一样，从而形成抗原/抗体复合体，或者免疫复合体。该结合中和抗原或者引起抗原的破坏。

当脊椎动物首次遇到抗原时，它显示出初次体液免疫应答。如果动物在几天后遇到相同的抗原，那么免疫应答更快速并且具有更大的量级。最初的相遇导致特异免疫细胞（B细胞）克隆增殖和分化。子代淋巴细胞不仅包括效应细胞（产生抗体的细胞），而且包括记忆细胞克隆，所述记忆细胞在最初抗原的随后刺激后保留产生效应细胞和记忆细胞的能力。效应细胞仅存活几天。记忆细胞终生生活着并且可以

通过用相同的抗原再次刺激而再活化。从而，当再次遇到抗原时，它的记忆细胞快速产生效应细胞，所述效应细胞快速产生大量抗体。

通过利用抗体与抗原以高度特异性方式相互作用的独特能力，已经开发抗体作为可以生产并用于诊断和治疗应用的分子。因为它们结合抗原性表位的独特能力，多克隆和单克隆抗体可以用于鉴定携带该表位的分子或者可以自身或者与另一部分结合导向特定部位用于诊断或治疗。可以对实际上任何病原体或者生物学靶产生多克隆和单克隆抗体。术语多克隆抗体指通常含有各种同种型和特异性的病原体特异性抗体的免疫血清。相反，单克隆抗体由单一免疫球蛋白类型组成，代表具有一种特异性的一种同种型。

在 1890 年，Shibasaburo Kitazato 和 Emil Behring 进行了一个基础实验，其证明通过将来自免疫动物的血清转移到非免疫动物，可以将免疫性从一个动物转移到另一个动物。该划时代的实验奠定了将被动免疫引入临床实践的基础。然而，在二十世纪 40 年代，因为与施用异源血清有关的毒性和有效的抗微生物化学疗法的引入，大规模的血清疗法在很大程度上被抛弃。当前，此类多克隆抗体疗法适于治疗相对少的情况下的传染病，如免疫球蛋白缺陷患者中的替代疗法、针对几种病毒（如狂犬病、麻疹、甲型和乙型肝炎、水痘）的暴露后预防，和毒素中和（白喉、破伤风和肉毒中毒）。尽管血清疗法的用途有限，但是在美国，每年需要超过 16 公吨的人抗体用于静脉内抗体治疗。在大多数高度工业化国家的经济中需要相当水平的使用，并且预期在发展中国家中，该需要快速增长。当前，用于被动免疫的人抗体从供体的合并血清得到。从而，用于治疗性和预防性治疗的人抗体的量存在固有的限制。

抗体用于针对生物战争试剂的被动免疫代表了非常有希望的防御策略。对这种生物战争试剂的最终防线是所暴露的个体的免疫系统。当前对生物武器的防御策略包括这样的措施，如增强流行病学监护、接种、和使用抗微生物剂。因为生物战争和生物恐怖主义 (bioterrorism) 的潜在威胁与靶定群体中免疫人数成反比，所以生物制剂是仅对具有很大比例的易感人员的潜在武器。

接种可以减小群体对特定威胁的易感性，前提是存在可以诱导保护性应答的安全疫苗。不幸的是，通过接种产生保护性应答需要的时

间比暴露于疾病和疾病发作之间的时间更长。此外，许多疫苗需要多剂来实现保护性免疫应答，其将限制它们在紧急情况下对于在受攻击后提供快速预防的有用性。此外，并不是所有的疫苗受体都产生保护性应答，即使在接受推荐的免疫方案之后。

当暴露于某些试剂后施用药物时，药物可以提供保护，但是对于生物战争的许多潜在试剂还没有药物。当前，没有可以用于在暴露于预先形成的毒素后预防该疾病的小分子药物。可以提供即刻免疫状态的唯一的当前可利用的干预是用保护性抗体进行被动免疫 (Arturo Casadevall "Passive Antibody Administration (Immediate Immunity) as a Specific Defense Against Biological Weapons" from *Emerging Infectious Diseases*, Posted 09/12/2002)。

除了提供保护性免疫外，现代的基于抗体的疗法组成了对新出现的病原性细菌、真菌、病毒和寄生物的潜在有用的选择(A. Casadevall 和 M. D. Scharff, *Clinical Infectious Diseases* 1995; 150)。恶性肿瘤和癌症患者的治疗(C. Botti 等人, *Leukemia* 1997; Suppl 2:S55-59; B. Bodey, S.E. Siegel, 和 H.E. Kaiser, *Anticancer Res* 1996; 16(2):661)、移植器官的类固醇抗性排斥以及自身免疫病的治疗也可以通过使用单克隆或者多克隆抗体试剂实现(N. Bonnefoy-Berard 和 J.P. Revillard, *J Heart Lung Transplant* 1996; 15(5):435-442; C. Colby, 等人 *Ann Pharmacother* 1996; 30(10):1164- 1174; M.J. Dugan, 等人, *Ann Hematol* 1997; 75(1-2):41 2; W. Cendrowski, *Boll Ist Sieroter Milan* 1997; 58(4):339-343; L.K. Kastrukoff, 等人 *Can J Neurol Sci* 1978; 5(2):175178; J.E. Walker 等人 *J Neurol Sci* 1976; 29(2-4):303309)。

在抗体生产技术中的最近的进展提供了产生人抗体试剂同时避免与人血清疗法有关的毒性的方法。基于抗体的疗法的优点包括多方适应性、低毒性、病原体特异性、免疫功能的增强和有利的药物代谢动力学。

最近才出现了单克隆抗体治疗剂的临床应用。现在单克隆抗体已被批准作为移植、癌症、传染病、心血管疾病和炎症的治疗。更多单克隆抗体处于治疗多种疾病适应症的最后阶段的临床试验中。结果，单克隆抗体代表当前开发的最大类别的药物之一。

尽管当前流行使用单克隆抗体作为治疗剂，但是它们的使用存在一些障碍。例如，单克隆抗体的许多治疗应用需要重复施用，特别对于慢性疾病如自身免疫或癌症。因为小鼠方便用于免疫并且识别大多数抗原作为外来的，所以针对人靶的具有治疗潜力的单克隆抗体通常是鼠来源的。然而，鼠单克隆抗体作为人治疗剂具有内在的缺点。例如，它们需要更频繁的给药以维持单克隆抗体的治疗水平，因为它们在人体中比人抗体具有更短的循环半衰期。更关键的是，反复施用鼠免疫球蛋白产生了人免疫系统将识别鼠蛋白质为外来的，从而产生人抗小鼠抗体应答的可能性，其可以导致严重的变态反应。这种减小的功效和安全性的可能性导致开发许多技术用于减小鼠单克隆抗体的免疫原性。

多克隆抗体对多种抗原性靶非常有效。它们具有靶定并杀死与复杂疾病关联的多种“进化中的靶(evolving target)”的独特能力。同样，在所有药物类别中，多克隆抗体对于在抗原突变事件中保持活性的可能性最高。此外，尽管单克隆抗体对传染物具有有限的治疗活性，但是多克隆抗体可以中和毒素和引导免疫应答以消除病原体，以及生物战争试剂。

开发多克隆和单克隆抗体生产平台以满足将来对生产能力的需要代表了有希望的领域，其是当前多数研究的主题。一个特别有希望的策略是向小鼠或者大的家畜中引入人免疫球蛋白基因。该技术的延伸将包括灭活它们的内源免疫球蛋白基因。大型动物，如羊、猪和牛当前都用于生产血清来源的产品，如超免疫血清和凝血因子用于人用途。这将支持因安全性和伦理学而使用来自此类物种的人多克隆抗体。这些物种的每一种在血清和乳中都天然产生相当大量的抗体。

编码免疫球蛋白的基因的排列

抗体分子从可变基因元件的组合装配而来，且组合种系中许多可变基因元件引起的可能性使得宿主能够合成针对巨大数目抗原的抗体。每个抗体分子由两类多肽链组成：轻(L)链(可以是 κ L-链或者 λ L-链)和重(H)链。重链和轻链结合在一起以限定表位的结合区。单个抗体分子具有L链的两个相同拷贝和两个H链。每条链都包含可变区(V)和恒定区(C)。可变区构成分子的抗原结合部

位。为了实现多样的抗原识别，编码可变区的 DNA 经历基因重排。恒定区氨基酸序列对于抗体的特定同种型以及对于产生抗体的宿主是特异的，并且从而不经历重排。

DNA 重排机制对于重链和轻链基因座的可变区是相似的，尽管仅仅需要一个结合事件来产生轻链基因，而需要两个结合事件来产生完整重链基因。最常见的重排模式涉及两个基因区段之间 DNA 的环出和缺失。这在两个基因区段的编码序列在 DNA 中处于相同方向时发生。在具有相反的转录方向的两个基因区段之间可以发生第二种重组模式。该重组模式较不常见，尽管这种重排可以占据高达所有 V_k 与 J_k 连接的半数；半数人 V_k 基因区段的转录方向与 J_k 基因区段的转录方向相反。

通过分离的基因区段的体细胞重组产生编码完整 V 区的 DNA 序列。免疫球蛋白重链或者轻链的 V 区或者 V 结构域由超过一个基因区段编码。对于轻链，V 结构域由两个单独的 DNA 区段编码。第一个区段编码轻链的前 95-101 个氨基酸并且称作 V 基因区段，因为它编码 V 结构域的大部分。第二个区段编码 V 结构域的剩余部分（最长达 13 个氨基酸）并且称作连接或者 J 基因区段。V 和 J 基因区段的连接产生了编码整个轻链 V 区的连续外显子。为了产生完整的免疫球蛋白轻链信使 RNA，V 区外显子在转录后通过 RNA 剪接连接到 C 区序列。

重链 V 区在三个基因区段中编码。除了 V 和 J 基因区段（称作 V_H 和 J_H 以将它们与轻链 V_L 和 J_L 区别开来），还存在第三个基因区段，称作多样性或者 D_H 基因区段，其位于 V_H 和 J_H 基因区段之间。产生完整重链 V 区的重组过程在两个单独的阶段发生。在第一个阶段， D_H 基因区段连接到 J_H 基因区段；然后 V_H 基因区段重排到 DJ_H 以产生完整的 V_H -区外显子。如对于轻链基因那样，RNA 剪接将装配的 V 区序列连接到邻近的 C 区基因。

抗体谱的多样化在两个阶段发生：主要通过存在于骨髓中的前体 B 细胞中的 Ig V、D 和 J 基因区段的重排（“V(D)J 重组”），然后通过体细胞突变和这些重排的 Ig 基因类别转换重组，此时成熟的 B 细胞被激活。免疫球蛋白体细胞突变和类别转换对于免疫应答的成熟和产生“记忆”应答是重要的。

抗体的基因组基因座非常大并且它们位于不同的染色体上。免疫球蛋白基因区段组织成三簇或者遗传基因座： κ 、 λ 和重链基因座。每个组织稍有不同。例如，在人中，免疫球蛋白基因如下组织。 λ 轻链基因座位于22号染色体上并且一簇 V_λ 基因区段之后是四组 J_λ 基因区段，其每个都连接到单个 C_λ 基因。 κ 轻链基因座在2号染色体上并且 V_κ 基因区段簇之后是一簇 J_κ 基因区段，然后是单个 C_κ 基因。14号染色体上重链基因座的组织与 κ 基因座的相似，具有单独簇的 V_H 、 D_H 和 J_H 基因区段和单独簇的 C_H 基因。重链基因座在一个重要的方面不同：它不是含有单个C区，而是含有轮流排列的一系列C区，每个C区对应于一个不同的同种型。有5种免疫球蛋白重链同种型： IgM 、 IgG 、 IgA 、 IgE 和 IgD 。通常，细胞一次仅表达一种同种型，以 IgM 开始。其他同种型如 IgG 的表达可以通过同种型转换发生。

各种V、D和J基因的连接是完全随机的事件，其对于VDJ(H)产生约50,000个不同的可能的组合，对于VJ(L)产生约1,000个组合。H和L链的随后的随机配对使抗体特异性总数达到约 10^7 种可能性。通过不同遗传区段的不精确连接进一步增加了多样性。重排在两个DNA链上发生，但是仅一条链被转录（由于等位基因排斥）。由于DNA中的不可逆缺失，在B细胞的生命中仅发生一次重排。因此，每个成熟B细胞维持一种免疫特异性并且在后代或者克隆中维持。这组成了克隆选择的分子基础；即每个抗原决定簇触发带有特异性受体分子的B淋巴细胞的原始克隆的应答。V(D)J重组确立的初级B细胞谱主要受到两个紧密连锁的基因：重组激活基因(RAG)-1和RAG-2的控制。

在过去10年里，已经揭示了脊椎动物在Ig基因多样性和抗体谱发育中的相当大的多样性。啮齿类动物和人具有5个重链类别： IgM 、 IgD 、 IgG 、 IgE 和 IgA ，且每种具有 IgG 的四个亚类和 IgA 的一个或两个亚类，而兔具有一个 IgG 重链基因但是对于不同的 IgA 亚类有13种基因(Burnett, R.C 等人 EMBO J 8:4047; Honjo, In Honjo, T, Alt. F.W. T.H. eds, Immunoglobulin Genes p. 123 Academic Press, New York)。猪具有至少6个 IgG 亚类(Kacskovics, I 等人 1994 J Immunol 153:3565)，但是没有 IgD (Butler 等人 1996 Inter. Immunol

8:1897-1904)。编码 IgD 的基因仅已在啮齿类动物和灵长类动物中描述。通过将啮齿类和灵长类动物中看到的模式与对鸡、兔、猪和驯养的牛科(Bovidae)报导的模式比较例证了谱发育机理中的多样性。虽然前面的组具有属于 7 到 10 个家族的大量的 V_H 基因(Rathbun, G. In Hongo, T. Alt. F. W. 和 Rabbitts, T.H., eds, Immunoglobulin Genes, p. 63, Academic press New York), 但后面组的每个成员的 V_H 基因属于单个 V_H 基因家族(Sun, J. 等人 1994 J. Immunol 155:56118; Dufour, V 等人 1996, J Immunol. 156:2163)。除了兔之外, 该家族由小于 25 种基因组成。虽然啮齿类和灵长类动物可以利用 4 到 6 种 J_H 区段, 但鸡中仅一种 J_H 可以用于谱发育(Reynaud 等人 1989 Adv. Immunol. 57:353)。类似地, Butler 等人(1996 Inter. Immunol 8:1897-1904)假定猪可能像鸡一样具有仅一个 J_H 基因。这些物种通常具有较少的 V、D 和 J 基因; 在猪和奶牛中, 存在单个 VH 基因家族, 其由小于 20 种基因区段组成(Butler 等人, Advances in Swine in Biomedical Research, 编著: Tumbleson 和 Schook, 1996; Sinclair 等人, J. Immunol. 159: 3883, 1997)。与 J 和 D 基因区段的较低数目一起, 这导致通过基因重排产生明显较低的多样性。然而, 在这些物种中似乎有更大数目的轻链基因。类似于人和小鼠, 这些物种表达单个 κ 轻链但是多个 λ 轻链基因。然而, 这些似乎不影响通过重排实现的有限的多样性。

因为超过 100 种 V_H 、20 - 30 种 D_H 和 4 到 6 种 J_H 基因区段的组合连接是人中产生抗体谱的主要机制, 所以具有较少的 V_H 、 D_H 或者 J_H 区段的物种必须产生较小的谱或者使用谱发育的备选机制。反刍动物、猪、兔和鸡利用几种机制来产生抗体多样性。在这些物种中, 似乎存在一个重要的次级谱发育, 其在高度特化的淋巴组织如回肠的淋巴集结中发生(Binns 和 Licence, Adv. Exp. Med. Biol. 186: 661, 1985)。次级谱发育在这些物种中通过体细胞突变过程发生, 该过程是随机的并且是没有完全了解的过程。该谱多样化的机制似乎是模板化突变, 或者基因转变(Sun 等人, J. Immunol. 153: 5618, 1994)或者体细胞高变。

基因转变对于一些高等脊椎动物如鸡、兔和奶牛中的抗体多样化是重要的。然而, 在小鼠中, 转变事件在内源抗体基因中似乎不频繁。基因转变是一种不同的多样化机制, 其特征是来自供体抗体 V 基因

区段的同源序列转移到受体 V 基因区段。如果供体和受体区段具有许多序列差别,那么基因转变可以通过单个事件将一组序列改变导入 V 区中。取决于物种,基因转变事件可以在 B 细胞分化期间抗原暴露之前和/或之后发生(Tsai 等人 *International Immunology*, Vol. 14, No. 1, 55-64, 2002 年 1 月)。

体细胞高变实现所有高等脊椎动物物种中的抗体基因多样化。它的特征是向抗体 V(D)J 区段中导入单个点突变。通常,高变似乎通过抗原刺激在 B 细胞中激活。

产生具有人源化免疫系统的动物

为了减小在小鼠中产生的用于人治疗的抗体的免疫原性,已经做了各种努力来在现在称作人源化的方法中用人蛋白质序列替换鼠蛋白质序列。已经构建了转基因小鼠,其自身的免疫球蛋白基因已经被人免疫球蛋白基因功能地替换,从而它们在免疫后产生人抗体。通过使用基因寻靶技术缺失参与小鼠 Ig 基因重排和表达的过程中的关键的顺式作用序列,灭活胚胎干(ES)细胞中的小鼠 Ig 基因,消除了小鼠抗体的产生。通过引入含有人种系构型 H 链和 κ L 链小基因座(minilocus)转基因的兆碱基大小的 YAC,可以恢复这些突变小鼠中的 B 细胞发育。这些转基因小鼠中完全人抗体的表达是占优势的,处于几百 $\mu\text{g/l}$ 血液的水平上。该水平的表达为在表达人 Ig 基因的野生型小鼠中检测到的水平的几百倍,从而表明灭活内源小鼠 Ig 基因以增强小鼠的人抗体产生的重要性。

首次人源化尝试利用分子生物学技术来构建重组抗体。例如,将来自对半抗原特异的小鼠抗体的互补性决定区(CDR)嫁接到人抗体构架上,从而实现 CDR 置换。该新抗体保留 CDR 序列赋予的结合特异性(P. T. Jones 等人 *Nature* 321: 522-525 (1986))。人源化的下一个水平涉及组合完整小鼠 VH 区与人恒定区如 γ_1 (S. L. Morrison 等人, *Proc. Natl. Acad. Sci.*, 81, pp. 6851-6855 (1984))。然而,这些嵌合抗体仍然含有大于 30% 的异种序列,有时相对于完全异种抗体的免疫原性的降低仅仅是微小的(M. Bruggemann 等人, *J. Exp. Med.*, 170, pp. 2153-2157 (1989))。

随后,试图向小鼠中引入人免疫球蛋白基因,从而产生能够用具

有人序列的抗体应答抗原的转基因小鼠(Bruggemann 等人 Proc. Nat'l. Acad. Sci. USA 86:6709-6713 (1989)). 由于人免疫球蛋白基因组基因座的大尺寸, 认为这些尝试受到 DNA 的量的限制, 所述 DNA 可以通过可利用的克隆载体稳定维持。结果, 许多研究人员将注意力集中在产生小基因座上, 该小基因座含有有限数目的 V 区基因并且与天然或者种系构型相比在基因之间具有改变的空间距离(见, 例如, 美国专利号 5,569,825)。这些研究表明在小鼠中产生人序列抗体是可能的, 但是在从这些转基因动物得到结合特异性和效应子功能(同种型)的足够多样性以满足对抗体治疗剂的不断增长的需求方面仍然存在巨大障碍。

为了提供额外的多样性, 已经开展了将人 Ig 基因座的大种系片段加入转基因哺乳动物中的工作。例如, 使用酵母人工染色体(YAC)克隆载体向小鼠中导入在人基因组的未重排种系中发现的以相同间隔排列的多数人 V、D 和 J 区基因和人 C μ 和 C δ 恒定区(见例如, WO 94/02602)。将包含编码人 γ -2 恒定区的序列和类别转换重组所需的上游序列的 22 kb DNA 片段后来附加到前面的转基因中。此外, 使用 YAC 向小鼠中导入包含 V κ 、J κ 和 C κ 区基因的人 κ 基因座的部分, 所述 V κ 、J κ 和 C κ 区基因也在人基因组的未重排的种系中被发现以基本上相同的间隔排列。用基因寻靶灭活鼠 IgH 和 κ 轻链免疫球蛋白基因座并且将这种敲除品种与上面的转基因品种育种以产生小鼠品系, 其具有人 V、D、J、C μ 、C δ 和 C γ_2 恒定区以及人 V κ 、J κ 和 C κ 区基因, 它们都处于灭活的鼠免疫球蛋白背景下(见, 例如, Kucherlapati 等人的 PCT 专利申请 WO 94/02602; 也见 Mendez 等人, Nature Genetics 15:146-156 (1997))。

酵母人工染色体作为克隆载体与内源基因座的基因寻靶和转基因小鼠品种的育种组合提供了抗体多样性问题的一个解决方案。通过该方法得到了几个优点。一个优点是 YAC 可以用于向宿主细胞转移数百 kb DNA。因此, 使用 YAC 克隆载体允许向转基因小鼠中包含相当大部分的完整人 Ig 重链和轻链区, 从而接近在人中可得到的潜在多样性水平。该方法的另一优点是已经表明许多 V 基因在鼠免疫球蛋白生产缺陷的小鼠中恢复完全的 B 细胞发育。这确保为这些重构的小鼠提供必要的细胞以对任何给定免疫原产生强烈的人抗体应答

(见例如, WO 94/02602.; L. Green 和 A. Jakobovits, *J. Exp. Med.* 188:483-495 (1998)). 另一优点是通过利用酵母中的高频率同源重组在 YAC 中缺失或插入序列。这提供了 YAC 转基因的容易的工程改造。

此外, Green 等人 *Nature Genetics* 7:13-21 (1994)描述了产生分别含有人重链基因座和 κ 轻链基因座的 245 kb 和 190kb 大小的种系构型片段的 YAC, 其含有核心可变和恒定区序列。Green 等人的工作最近延伸到导入大于约 80% 的人抗体谱来产生 XenoMouse™小鼠, 所述导入通过引入分别为人重链基因座和 κ 轻链基因座的兆碱基大小的种系构型 YAC 片段来实现。见例如, Mendez 等人 *Nature Genetics* 15:146-156 (1997), Green 和 Jakobovits *J. Exp. Med.* 188:483-495 (1998), 欧洲专利号 EP 0 463 151 B1, PCT 公开号 WO 94/02602, WO 96/34096 和 WO 98/24893。

存在几种策略用于产生生产人抗体的哺乳动物。具体地, 有“小基因座”方法, 其以 GenPharm International, Inc. 和 Medical Research Council 的工作为代表; YAC 导入 Ig 基因座的大的和相当大的种系片段, 其以 Abgenix, Inc (以前 Cell Genesys)的工作为代表。通过使用微细胞融合导入整个或者基本上整个基因座, 以 Kirin Beer Kabushiki Kaisha 的工作为代表。

在小基因座方法中, 通过包括 Ig 基因座的片段(个体基因)模拟外源 Ig 基因座。从而, 一个或多个 V_H 基因、一个或多个 D_H 基因、一个或多个 J_H 基因、 μ 恒定区和第二个恒定区(如 γ 恒定区)形成用于插入动物中的构建体。见例如, 美国专利号 5,545,807, 5,545,806, 5,625,825, 5,625,126, 5,633,425, 5,661,016, 5,770,429, 5,789,650, 5,814,318, 5,591,669, 5,612,205, 5,721,367, 5,789,215, 5,643,763; 欧洲专利号 0 546 073; PCT 公开号 WO 92/03918, WO 92/22645, WO 92/22647, WO 92/22670, WO 93/12227, WO 94/00569, WO 94/25585, WO 96/14436, WO 97/13852 和 WO 98/24884; Taylor 等人 *Nucleic Acids Research* 20:6287-6295 (1992), Chen 等人 *International Immunology* 5:647-656 (1993), Tuailon 等人 *J. Immunol.* 154:6453-6465 (1995), Choi 等人 *Nature Genetics* 4:117-123 (1993), Lonberg 等人 *Nature* 368:856-859 (1994), Taylor 等人 *International*

Immunology 6:579-591 (1994), Tuailon 等人 J. Immunol. 154:6453-6465 (1995), 和 Fishwild 等人 Nature Biotech. 14:845-851 (1996).

在微细胞融合方法中,可以向小鼠中导入部分或者完整人染色体(见,例如,欧洲专利申请号 EP 0 843 961 A1)。使用该方法产生并且含有人 Ig 重链基因座的小鼠将通常具有超过一个,并且可能所有的人恒定区基因。因此,此类小鼠将产生结合特定抗原的具有许多不同恒定区的抗体。

尽管小鼠仍然是关于在人中表达人免疫球蛋白研究最深入的动物,但是最近的技术发展已经允许开始在其他动物如奶牛中应用这些技术。小鼠中的一般方法是基因修饰小鼠的胚胎干细胞以敲除鼠免疫球蛋白,然后向 ES 细胞中插入含有人免疫球蛋白的 YAC。然而,不能从奶牛和其他大的动物如羊和猪中得到 ES 细胞。从而,在甚至可能产生具有免疫球蛋白基因被敲除并且表达人抗体的大型动物之前,必须发生几个根本的发展。用 ES 细胞操作来产生基因修饰动物的备选方案是使用已经基因修饰的体细胞进行克隆。使用用于核转移的基因修饰体细胞进行克隆仅仅在最近才实现。

自从 1997 年宣布多莉(Dolly)(克隆的羊)从成年体细胞出生(Wilmut, L, 等人(1997) Nature 385: 810-813)以来,已经克隆了有蹄类动物,包括牛(Cibelli, J 等人, 1998 Science 280: 1266-1258; Kubota, C. 等人 2000 Proc. Nat'l. Acad. Sci 97: 990-995)、山羊(Baguisi, A. 等人, (1999) Nat. Biotechnology 17: 456-461)和猪(Polejaeva, I.A., 等人 2000 Nature 407: 86-90; Betthausen, J. 等人 2000 Nat. Biotechnology 18: 1055-1059)。

下一个技术进展是技术发展到现在核转移前基因修饰细胞以产生基因修饰动物。关于 PPL 治疗的 PCT 公开号 WO 00/51424 描述了用于核转移的体细胞的定向基因修饰。

这些基础发展之后,已经报导了基因的单个和双等位基因敲除和具有这些修饰的活动的动物的出生。在 2002 到 2004 年之间,三个独立的小组: Immerge Biotherapeutics, Inc.与 University of Missouri 合作(Lai 等人(Science (2002) 295: 1089-1092) & Kolber-Simonds 等人(PNAS. (2004) 101(19):7335-40)), Alexion Pharmaceuticals (Ramsoondar 等人(Biol Reprod (2003)69: 437-445)和 Revivacor, Inc.

(Dai 等人(Nature Biotechnology (2002) 20: 251-255) 和 Phelps 等人(Science (2003) Jan 17; 299(5605):411-4))通过用定向遗传缺失从体细胞进行核转移, 产生了缺少 α -1,3-GT 基因的一个等位基因或者两个等位基因的猪。在 2003 年, Sedai 等人(Transplantation (2003) 76:900-902)报导牛中 α -1,3-GT 基因的一个等位基因的定向破坏, 接着成功地核转移基因修饰细胞的核并产生转基因胎儿。

从而, 现在正刚开始研究在大型动物中敲除免疫球蛋白基因并向它们的细胞中插入人免疫球蛋白基因座的可行性。然而, 由于免疫球蛋白基因的复杂性和物种差异, Ig κ 、 λ 和重链的基因组序列和重排在大多数物种中仍然知之甚少。例如, 在猪中, 仅仅对重链恒定 α 、重链恒定 μ 和重链恒定 δ 描述了部分基因组序列和组织(Brown 和 Butler Mol Immunol. 1994 Jun; 31(8):633-42, Butler 等人 Vet Immunol Immunopathol. 1994 Oct; 43(1-3):5-12, 和 Zhao 等人 J Immunol. 2003 Aug 1; 171(3):1312-8)。

在奶牛中, 已经对免疫球蛋白重链基因座作图(Zhao 等人 2003 J. Biol. Chem. 278:35024-32)并且牛 κ 基因的 cDNA 序列是已知的(见, 例如, 美国专利公开号 2003/0037347)。此外, 已经对约 4.6kb 的牛 μ 重链基因座测序并且通过破坏牛 μ 重链的一个或两个等位基因, 已经产生了重链免疫球蛋白表达降低的转基因小牛。此外, 用牛胎儿成纤维细胞的微细胞介导的染色体转移和核转移技术, 已经向奶牛(TC 奶牛)中导入了含有人 Ig H 链和 κ L 链的完全未排列的序列的哺乳动物人工染色体(MAC)载体(见例如, Kuroiwa 等人 2002 Nature Biotechnology 20:889, Kuroiwa 等人 2004 Nat Genet. Jun 6 Epub, 美国专利公开号 2003/0037347、2003/0056237、2004/0068760 和 PCT 公开号 WO 02/07648)。

尽管在产生表达人免疫球蛋白的牛的方面已经取得了重大进步, 但是在其他大型动物, 如羊、山羊和猪中收效甚微。尽管可以容易地从 Genbank 获得羊、山羊和猪的免疫球蛋白基因的 cDNA 序列信息, 但是免疫球蛋白基因座(其经历大量重排)的独特性质使得需要表征已知的存在于 mRNA(或者 cDNA)的序列之外的序列。因为免疫球蛋白基因座是模块式的并且编码区域是冗余的, 所以缺失已知的编码区不能确保改变基因座的功能。例如, 如果人们缺失重链可变区

的编码区，那么基因座的功能将不被显著改变，因为在该基因座中仍有数百其他功能可变基因。因此，人们必须首先表征基因座以鉴定潜在的“唯一弱点”。

尽管在牛中表达人抗体取得了一定进展，但是对于灭活内源牛 Ig 基因、增加人抗体的表达水平和在其他大型动物，如猪中产生人抗体表达方面仍然存在更大的挑战，所述其他大型动物的免疫球蛋白基因的序列和排列在很大程度上还未知。

因此，本发明的一个目的是提供有蹄类动物免疫球蛋白种系基因序列的排列。

本发明的另一个目的是提供新颖的有蹄类动物免疫球蛋白基因组序列。

本发明的另一个目的是提供缺少重链和/或轻链免疫球蛋白基因的至少一个等位基因的细胞、组织和动物。

本发明的另一个目的是提供表达人免疫球蛋白的有蹄类动物。

本发明的另一个目的是提供产生缺少新的有蹄类动物免疫球蛋白基因序列的至少一个等位基因的细胞、组织和动物和/或表达人免疫球蛋白的方法。

发明概述

本发明首次提供了有蹄类动物免疫球蛋白种系基因序列排列及其新的基因组序列。此外，提供了缺少重链或轻链免疫球蛋白基因的至少一个等位基因的新的有蹄类动物细胞、组织和动物。基于该发现，可以产生完全缺少重链和/或轻链免疫球蛋白基因的至少一个等位基因的有蹄类动物。此外，可以进一步修饰这些有蹄类动物以表达异种的（例如人）免疫球蛋白基因座或者其片段。

在本发明的一方面，提供了缺少功能性内源免疫球蛋白的任何表达的转基因有蹄类动物。在一个实施方案中，该有蹄类动物可以缺少内源重链和/或轻链免疫球蛋白的任何表达。轻链免疫球蛋白可以是 κ 和/或 λ 免疫球蛋白。在额外实施方案中，提供了缺少内源免疫球蛋白的至少一个等位基因的表达的转基因有蹄类动物，其中该免疫球蛋白选自重链、 κ 轻链和 λ 轻链或者其任意组合。在一个实施方案中，通过内源免疫球蛋白基因座的遗传寻靶以防止内源免疫球蛋白的

表达，可以实现功能性内源免疫球蛋白的表达。在一个实施方案中，通过同源重组可以完成遗传寻靶。在另一实施方案中，通过核转移可以产生转基因有蹄类动物。

在其他实施方案中，缺少功能性内源免疫球蛋白的任何表达的转基因有蹄类动物可以进一步基因修饰以表达异种免疫球蛋白基因座。在备选实施方案中，提供了含有异种免疫球蛋白基因座的猪动物。在一个实施方案中，异种免疫球蛋白基因座可以是重链和/或轻链免疫球蛋白或其片段。在另一实施方案中，异种免疫球蛋白基因座可以是 κ 链基因座或者其片段和/或 λ 链基因座或者其片段。在其他实施方案中，人工染色体(AC)可以含有异种免疫球蛋白。在一个实施方案中，AC可以是酵母AC或者哺乳动物AC。在另一实施方案中，异种基因座可以是人免疫球蛋白基因座或者其片段。在一个实施方案中，人免疫球蛋白基因座可以是人14号染色体、人2号染色体和人22号染色体或者其片段。在另一实施方案中，人免疫球蛋白基因座可以包括人免疫球蛋白的任意片段，该片段可以经历重排。在另一实施方案中，人免疫球蛋白基因座可以包括人免疫球蛋白重链和人免疫球蛋白轻链的任意片段，其可以经历重排。在另一实施方案中，人免疫球蛋白基因座可以包括任意人免疫球蛋白基因座或者其片段，其在暴露于抗原后可以产生抗体。在具体实施方案中，应答对一种或多种抗原的暴露，所述外源人免疫球蛋白可以在B细胞中表达以产生异种免疫球蛋白。

在本发明的另一方面，提供了表达异种免疫球蛋白基因座或者其片段的转基因有蹄类动物，其中该免疫球蛋白可以从整合在内源有蹄类动物染色体内的免疫球蛋白基因座表达。在一个实施方案中，提供了来自转基因动物的有蹄类动物细胞。在一个实施方案中，异种免疫球蛋白基因座可以被后代遗传。在另一实施方案中，异种免疫球蛋白基因座可以通过雄性种系由后代遗传。在其他实施方案中，人工染色体(AC)可以含有异种免疫球蛋白。在一个实施方案中，AC可以是酵母AC或者哺乳动物AC。在另一实施方案中，异种基因座可以是人免疫球蛋白基因座或者其片段。在一个实施方案中，人免疫球蛋白基因座可以是人14号染色体、人2号染色体和人22号染色体或者其片段。在另一实施方案中，人免疫球蛋白基因座可以包括人免疫球蛋白

白的任意片段，该片段可以经历重排。在另一实施方案中，人免疫球蛋白基因座可以包括人免疫球蛋白重链和人免疫球蛋白轻链的任意片段，其可以经历重排。在另一实施方案中，人免疫球蛋白基因座可以包括任意人免疫球蛋白基因座或者其片段，其在暴露于抗原后可以产生抗体。在具体实施方案中，应答对一种或多种抗原的暴露，所述外源人免疫球蛋白可以在 B 细胞中表达以产生异种免疫球蛋白。

在本发明的另一方面，提供了编码有蹄类动物免疫球蛋白的重链基因座的新的基因组序列。在一个实施方案中，提供了编码猪重链的分离的核苷酸序列，其包括至少一个可变区、两个多样性区、至少四个连接区和至少一个恒定区，如 μ 恒定区，例如，如 Seq ID No. 29 中表示。在另一实施方案中，提供了分离的核苷酸序列，其包括猪重链基因组序列的至少四个连接区和至少一个恒定区，如 μ 恒定区，如 Seq ID No.4 中表示。在另一实施方案中，提供了核苷酸序列，其包括猪重链基因组序列的第一个连接区的 5'侧翼序列，如 Seq ID No. 1 中表示。此外，提供了核苷酸序列，其包括猪重链基因组序列的第一个连接区的 3'侧翼序列，如 Seq ID No.4 的 3'区中表示。在进一步实施方案中，提供了如 Seq ID No.1、4 或 29 所述的分离的核苷酸序列。还提供了与 Seq ID No.1、4 或 29 至少 80、85、90、95、98 或 99% 同源的核酸序列。此外，提供了含有 Seq ID No.1、4 或 29 的至少 10、15、17、20、25 或 30 个连续核苷酸的核苷酸序列。在一个实施方案中，该核苷酸序列含有 Seq ID No 4 的至少 17、20、25 或 30 个连续核苷酸或者 Seq ID No. 29 的残基 1-9,070。在另一实施方案中，该核苷酸序列含有 Seq ID No. 29 的残基 9,070-11039。还提供了任选在严格条件下与 Seq ID No.1、4 或 29 杂交的核苷酸序列以及其同源核苷酸。

在另一实施方案中，提供了编码有蹄类动物免疫球蛋白的 κ 轻链基因座的新的基因组序列。本发明提供了有蹄类动物 κ 轻链区的首次报导的基因组序列。在一个实施方案中，提供了编码猪 κ 轻链基因座的核酸序列。在另一实施方案中，该核酸序列可以含有 κ 轻链的至少一个连接区、一个恒定区和/或一个增强子区。在另一实施方案中，该核苷酸序列可以包括至少五个连接区、一个恒定区和一个增强子区，例如，如 Seq ID No. 30 中表示。在另一实施方案中，提供了分

离的核苷酸序列，其含有猪基因组 κ 轻链的至少一个、至少两个、至少三个、至少四或者五个连接区和连接区的 3'侧翼序列，例如，如 Seq ID No 12 中表示。在另一实施方案中，提供了猪基因组 κ 轻链的分离的核苷酸序列，其含有第一个连接区的 5'侧翼序列，如 Seq ID No 25 中表示。在另一实施方案中，提供了分离的核苷酸序列，其含有猪基因组 κ 轻链的恒定区的 3'侧翼序列和任选地，增强子区的 5'部分，如 Seq ID Nos. 15、16 和/或 19 中表示。

在另一实施方案中，提供了如 Seq ID Nos 30、12、25、15、16 或 19 中所述的分离的核苷酸序列。还提供了与 Seq ID Nos 30、12、25、15、16 或 19 至少 80、85、90、95、98 或 99% 同源的核酸序列。此外，提供了含有 Seq ID Nos 30、12、25、15、16 或 19 的至少 10、15、17、20、25 或 30 个连续核苷酸的核苷酸序列。还提供了任选在严格条件下与 Seq ID Nos 30、12、25、15、16 或 19 杂交的核苷酸序列，以及其同源核苷酸。

在另一实施方案中，提供了编码有蹄类动物免疫球蛋白的 λ 轻链基因座的新的基因组序列。本发明提供了有蹄类动物 λ 轻链区的首次报导的基因组序列。在一个实施方案中，猪 λ 轻链核苷酸包括 J 到 C 单位的多联体。在具体实施方案中，提供了如 Seq ID No. 28 中描绘的分离的猪 λ 核苷酸序列。在一个实施方案中，提供了包括猪 λ 轻链基因组序列的第一个 λ J/C 区的 5'侧翼序列的核苷酸序列，如 Seq ID No 32 表示。此外，提供了包括猪 λ 轻链基因组序列的 J/C 簇区的 3'侧翼序列，如 λ J/C 下游约 200 个碱基对的核苷酸序列，如 Seq ID No 33 表示。备选地，提供了包括猪 λ 轻链基因组序列的 J/C 簇区的 3'侧翼序列的核苷酸序列，如 Seq ID No 34、35、36、37、38 和/或 39 表示。在另一实施方案中，提供了编码牛 λ 轻链基因座的核酸序列，其可以包括至少一个连接区 - 恒定区对和/或至少一个可变区，如 Seq ID No. 31 表示。在进一步的实施方案中，提供了如 Seq ID Nos 28、31、32、33、34、35、36、37、38 或 39 中所述的分离的核苷酸序列。还提供了与 Seq ID Nos 28、31、32、33、34、35、36、37、38 或 39 至少 80、85、90、95、98 或 99% 同源的核酸序列。此外，提供了含有 Seq ID Nos 28、31、32、33、34、35、36、37、38 或 39 的至少 10、15、17、20、25 或 30 个连续核苷酸的核苷酸序列。还提供了任选在

严格条件下与 Seq ID Nos 28、31、32、33、34、35、36、37、38 或 39 杂交的核苷酸序列，以及其同源核苷酸。

在另一实施方案中，还提供了核酸寻靶载体构建体。该寻靶载体可以设计成在细胞中完成同源重组。可以将这些寻靶载体转化到哺乳动物细胞中，通过同源重组靶定有蹄类动物重链、 κ 轻链或者 λ 轻链基因。在一个实施方案中，寻靶载体可以含有 3'重组臂和 5'重组臂（即侧翼序列），其与有蹄类动物重链、 κ 轻链或者 λ 轻链基因组序列，例如，如上述的 Seq ID Nos. 1、4、29、30、12、25、15、16、19、28 或 31 表示的序列同源。同源 DNA 序列可以包括与基因组序列同源的序列的至少 15 bp、20 bp、25 bp、50 bp、100 bp、500 bp、1 kbp、2 kbp、4 kbp、5 kbp、10 kbp、15 kbp、20 kbp 或 50 kbp。可以设计 3'和 5'重组臂，从而使得它们在基因组序列的至少一个功能可变区、连接区、多样性区和/或恒定区的 3'和 5'末端的侧翼。功能性区域的寻靶使得它失活，其导致细胞不能产生功能性免疫球蛋白分子。在另一实施方案中，同源 DNA 序列可以包括一个或多个内含子和/或外显子序列。除了核酸序列外，表达载体可以含有选择标记序列，如增强的绿色荧光蛋白（eGFP）基因序列、起始和/或增强子序列、聚腺苷酸化尾序列，和/或提供构建体在原核和/或真核宿主细胞中表达的核酸序列。选择标记可以位于 5'和 3'重组臂序列之间。

在一个具体实施方案中，寻靶载体可以含有 5'和 3'重组臂，该重组臂含有与猪免疫球蛋白重链基因座的 J6 区的 3'和 5'侧翼序列同源的序列。因为 J6 区是猪免疫球蛋白重链基因座的唯一的功能连接区，所以这将阻止功能性猪重链免疫球蛋白的表达。在特定实施方案中，寻靶载体可以含有 5'重组臂和 3'重组臂，所述 5'重组臂含有与 J6 区 5'的基因组序列，包括 J1-4 同源的序列，所述 3'重组臂含有与 J6 区 3'的基因组序列，包括 μ 恒定区同源的序列（“J6 寻靶构建体”），见例如图 1。此外，该 J6 寻靶构建体还可以含有位于 5'和 3'重组臂之间的选择标记基因，见例如，Seq ID No 5 和图 1。在其他实施方案中，寻靶载体可以含有 5'重组臂和 3'重组臂，所述 5'重组臂含有与猪重链基因座的多样性区 5'的基因组序列同源的序列，所述 3'重组臂含有与猪重链基因座的多样性区 3'的基因组序列同源的序列。在另一实施方案中，寻靶载体可以含有 5'重组臂和 3'重组臂，所述 5'重组

臂含有与猪重链基因座的 μ 恒定区5'的基因组序列同源的序列,所述3'重组臂含有与猪重链基因座的 μ 恒定区3'的基因组序列同源的序列。

在另一具体实施方案中,寻靶载体可以含有5'重组臂和3'重组臂,它们含有猪免疫球蛋白重链基因座的恒定区的3'和5'侧翼序列同源的序列。因为本发明发现了猪免疫球蛋白 κ 轻链基因座仅有一个恒定区,这将阻止功能性猪 κ 轻链免疫球蛋白的表达。在具体实施方案中,寻靶载体可以含有5'重组臂和3'重组臂,所述5'重组臂含有与恒定区5'的基因组序列(任选包括连接区)同源的序列,所述3'重组臂含有与恒定区3'的基因组序列(任选包括增强子区的至少部分)同源的序列(“ κ 恒定寻靶构建体”),见例如图2。此外,该 κ 恒定寻靶构建体还可以含有位于5'和3'重组臂之间的选择标记基因,见例如Seq ID No 20和图2。在其他实施方案中,寻靶载体可以含有5'重组臂和3'重组臂,所述5'重组臂含有与猪 κ 轻链基因座的连接区5'的基因组序列同源的序列,所述3'重组臂含有与猪 κ 轻链基因座的连接区3'的基因组序列同源的序列。

在另一实施方案中,提供了用于产生寻靶载体的3'和5'序列的引物。寡核苷酸引物能够与猪免疫球蛋白基因组序列,如上述Seq ID Nos. 1、4、29、30、12、25、15、16、19、28或31杂交。在具体实施方案中,引物在严格条件下与如上述的Seq ID Nos. 1、4、29、30、12、25、15、16、19、28或31杂交。另一实施方案提供了能够与猪重链、 κ 轻链或 λ 轻链核酸序列,如上述Seq ID Nos. 1、4、29、30、12、25、15、16、19、28或31杂交的寡核苷酸探针。多核苷酸引物或者探针可以具有与本发明的多核苷酸杂交的至少14个碱基、20个碱基、30个碱基或者50个碱基。探针或引物可以长为至少14个核苷酸,且在具体实施方案中,长为至少15、20、25、28或30个核苷酸。

在一个实施方案中,提供了扩增猪Ig重链的包括功能连接区(J6区)的片段的引物。在一个非限制性实施方案中,所扩增的重链片段由Seq ID No 4表示并且用于扩增该片段的引物可以与J区的部分,如但不限于Seq ID No 2互补,以产生5'重组臂,并且与Ig重链 μ 恒定区的部分,如但不限于Seq ID No 3互补,以产生3'重组臂。在另

一实施方案中，可以将猪 Ig 重链的区域（如但不限于 Seq ID No 4）亚克隆并装配到寻靶载体中。

在其他实施方案中，提供了扩增猪 Ig κ 轻链的包括恒定区的片段的引物。在另一实施方案中，提供了扩增猪 Ig κ 轻链的包括 J 区的片段的引物。在一个非限制性实施方案中，用于扩增该片段的引物可以与 J 区的部分，如但不限于 Seq ID No 21 或 10 互补，以产生 5' 重组臂，并且与恒定区 3' 的基因组序列，如但不限于 Seq ID No 14、24 或 18 互补，以产生 3' 重组臂。在另一实施方案中，可以将猪 Ig 重链的区域（如但不限于 Seq ID No 20）亚克隆并装配到寻靶载体中。

在本发明的另一方面，提供了根据本文描述的方法、序列和/或构建体产生的有蹄类动物细胞，其缺少有蹄类动物重链、 κ 轻链和/或 λ 轻链基因座的功能区的至少一个等位基因。这些细胞可以由于同源重组得到。具体地，通过灭活有蹄类动物重链、 κ 轻链或 λ 轻链基因的至少一个等位基因，可以产生表达有蹄类动物抗体的能力降低的细胞。在其他实施方案中，可以根据本文描述的方法、序列和/或构建体产生缺少有蹄类动物重链、 κ 轻链和/或 λ 轻链基因的两个等位基因的哺乳动物细胞。在进一步的实施方案中，提供了猪动物，其中通过根据本文描述的方法、序列和/或构建体产生的遗传寻靶事件灭活了有蹄类动物重链、 κ 轻链和/或 λ 轻链基因的至少一个等位基因。在本发明的另一方面，提供了猪动物，其中通过遗传寻靶事件灭活了有蹄类动物重链、 κ 轻链和/或 λ 轻链基因的两个等位基因。基因可以通过同源重组被靶定。

在其他实施方案中，可以破坏基因，即，可以改变遗传密码的部分，从而影响基因的该区段的转录和/或翻译。例如，可以通过替代、缺失（“敲除”）或者插入（“敲入”（knock-in））技术发生基因的破坏。可以插入所需蛋白质的额外基因或者调节现有序列转录的调节序列。为了实现有蹄类动物免疫球蛋白基因的多个基因修饰，在一个实施方案中，可以顺序修饰细胞以含有多个基因修饰。在其他实施方案中，可以将动物一起育种以产生含有免疫球蛋白基因的多个基因修饰的动物。作为阐明性实例，缺少表达有蹄类动物重链基因的至少一个等位基因的动物可以进一步基因修饰或者与缺少 κ 轻链基因的至少一个等位基因的动物育种。

在本发明的实施方案中，根据本文描述的方法、序列和/或构建体使得有蹄类动物重链、 κ 轻链或者 λ 轻链基因的等位基因失活，从而不再产生功能性有蹄类动物免疫球蛋白。在一个实施方案中，可以将靶定的免疫球蛋白基因转录成RNA，但是不翻译成蛋白质。在另一实施方案中，可以将靶定的免疫球蛋白基因以失活的截短形式转录。这种截短的RNA可以不被翻译或者可以翻译成无功能的蛋白质。在备选实施方案中，可以将靶定的免疫球蛋白基因以不发生基因的转录的方式灭活。在另一实施方案中，可以转录靶定的免疫球蛋白基因，然后翻译成无功能的蛋白质。

在本发明的另一方面，缺少有蹄类动物重链、 κ 轻链和/或 λ 轻链基因的一个等位基因、任选两个等位基因的有蹄类动物，如猪或者牛的细胞可以用作供体细胞以向受体细胞进行核转移，以产生克隆的转基因动物。备选地，可以在胚胎干细胞中产生有蹄类动物重链、 κ 轻链和/或 λ 轻链基因敲除，其然后用于产生后代。根据本文描述的方法、序列和/或构建体产生的缺少功能性有蹄类动物重链、 κ 轻链和/或 λ 轻链基因的一个等位基因的后代可以育种以进一步通过孟德尔型遗传产生在两个等位基因中都缺少功能性的后代。

在本发明的一方面，提供了破坏有蹄类动物免疫球蛋白基因的表达的方法，其通过(i)分析有蹄类动物重链、 κ 轻链或 λ 轻链基因组基因座的种系构型；(ii)确定该基因座的至少一个功能区域的5'末端和3'末端侧翼的核苷酸序列位置；和(iii)将含有侧翼序列的寻靶构建体转染到细胞中实现，其中，在成功地同源重组后，破坏了免疫球蛋白基因座的至少一个功能区，从而减小或者防止该免疫球蛋白基因的表达。在一个实施方案中，提供了猪重链基因座的种系构型。猪重链基因座含有至少四个可变区、两个多样性区、六个连接区和五个恒定区，例如，如图1图解的。在具体实施方案中，六个连接区的仅一个：J6是功能性的。在另一实施方案中，提供了猪 κ 轻链基因座的种系构型。猪 κ 轻链基因座含有至少六个可变区、六个连接区、一个恒定区和一个增强子区，例如，如图2图解的。在另一实施方案中，提供了猪 λ 轻链基因座的种系构型。

在本发明的其他方面，提供了根据本文描述的方法、序列和/或构建体产生的有蹄类动物和有蹄类动物细胞，它们缺少有蹄类动物重

链、 κ 轻链和/或 λ 轻链基因座的功能区的至少一个等位基因，所述有蹄类动物和有蹄类动物细胞进一步修饰以表达人抗体（即，免疫球蛋白（Ig））基因座的至少部分。在额外实施方案中，提供了表达异种免疫球蛋白的猪动物。该人基因座可以经历重排并在有蹄类动物中表达人抗体分子的不同群体。这些克隆的转基因有蹄类动物提供了人抗体（如多克隆抗体）的可补充的、理论上无限的供应，所述抗体可以用于治疗、诊断、纯化和其他临床上相关的目的。在一个具体实施方案中，人工染色体（AC），如酵母或者哺乳动物人工染色体（YAC或者MAC）可以用于允许人免疫球蛋白基因在有蹄类动物细胞和动物中表达。可以将所有或者部分的人免疫球蛋白基因，如Ig重链基因（人414号染色体）、Ig κ 轻链基因（人2号染色体）和/或Ig λ 轻链基因（22号染色体）插入到人工染色体，其然后可插入到有蹄类动物细胞中。在进一步的实施方案中，提供了有蹄类动物和有蹄类动物细胞，其含有至少一种人抗体基因座的部分或者全部，该基因座经历重排并且表达不同群体的人抗体分子。

在额外实施方案中，提供了产生异种抗体的方法，其中该方法可以包括：（a）对有蹄类动物施用一种或多种目的抗原，所述有蹄类动物的细胞包含一个或多个人工染色体并且缺少功能性内源免疫球蛋白的任何表达，每个人工染色体包含一个或多个经历重排的异种免疫球蛋白基因座，从而导致产生针对所述一种或多种抗原的异种抗体；和/或（b）从有蹄类动物回收所述异种抗体。在一个实施方案中，免疫球蛋白基因座可以在B细胞中经历重排。

在本发明的一方面，通过根据本发明的任意方面的方法可以制备有蹄类动物，如猪或者奶牛。这些克隆的转基因有蹄类动物（例如，猪和牛动物）提供人多克隆抗体的可补充的、理论上无限的供应，所述抗体可以用作治疗、诊断和纯化目的。例如，根据本文描述的方法、序列和/或构建体产生的在血流中产生多克隆人抗体的转基因动物可以用于产生对所需抗原特异的一组不同的人抗体。大量多克隆抗体的可用性也可以用于治疗 and 预防传染病、对生物战争试剂接种、调节免疫系统、除去不希望的人细胞（如癌细胞）和调节特定人分子。

在其他实施方案中，根据本文描述的方法、序列和/或构建体产生的缺少功能性免疫球蛋白表达的动物或细胞可以含有额外的基因

修饰以消除异种抗原的表达。可以修饰此类动物以消除 α -1,3-半乳糖基转移酶基因、CMP-Neu5Ac 羟化酶基因（见例如，USSN 10/863,116）、iGb3 合酶基因（见，例如，美国专利申请 60/517,524）和/或福斯曼(Forssman)合酶基因(见例如，美国专利申请 60/568,922)的至少一个等位基因的表达。在额外实施方案中，本文公开的动物还可以含有基因修饰以表达岩藻糖基转移酶和/或唾液酸转移酶。为了实现这些额外的基因修饰，在一个实施方案中，可以修饰细胞以含有多个基因修饰。在其他实施方案中，动物可以一起育种以实现多个基因修饰。在一个具体实施方案中，根据本文描述的方法、序列和/或构建体产生的缺少功能性免疫球蛋白表达的动物，如猪可以与缺少 α -1,3-半乳糖基转移酶表达的动物如猪育种（如 WO 04/028243 中所述）。

附图简述

图 1 图解了寻靶载体的设计，该寻靶载体破坏猪重链免疫球蛋白基因的连接区的表达。

图 2 图解了寻靶载体的设计，该寻靶载体破坏猪 κ 轻链免疫球蛋白基因的恒定区的表达。

图 3 图解了猪 λ 免疫球蛋白基因座的基因组组织，该基因座包括 J-C 序列的多联体以及包括 JC 区 5'的可变区的侧翼区域。细菌人工染色体(BAC1 和 BAC2)代表可以从 BAC 文库得到的猪免疫球蛋白基因组的片段。

图 4 代表寻靶载体的设计，该寻靶载体破坏猪 λ 轻链免疫球蛋白基因的 JC 簇区域的表达。“SM”代表选择标记基因，其可以用于寻靶载体中。

图 5 图解了向猪 λ 免疫球蛋白基因座的 JC 簇区域 5'的区插入位点特异性重组酶靶或者识别位点的寻靶策略。“SM”代表选择标记基因，其可以用于寻靶载体中。“SSRRS”代表特异性重组酶靶或者识别位点。

图 6 图解了向猪 λ 免疫球蛋白基因座的 JC 簇区域 3'的区插入位点特异性重组酶靶或者识别位点的寻靶策略。“SM”代表选择标记基因，其可以用于寻靶载体中。“SSRRS”代表特异性重组酶靶或者

识别位点。

图 7 图解了 YAC 向宿主基因组的位点特异性重组酶介导的转移。“SSRRS”代表特异性重组酶靶或者识别位点。

详述

本发明首次提供了有蹄类动物免疫球蛋白种系基因序列排列及其新的基因组序列。此外，提供了缺少重链或轻链免疫球蛋白基因的至少一个等位基因的新的有蹄类动物细胞、组织和动物。基于该发现，可以产生完全缺少重链和/或轻链免疫球蛋白基因的至少一个等位基因的有蹄类动物。此外，可以进一步修饰这些有蹄类动物以表达异种的（例如人）免疫球蛋白基因座或者其片段。

在本发明的一方面，提供了缺少功能性内源免疫球蛋白的任何表达的转基因有蹄类动物。在一个实施方案中，该有蹄类动物可以缺少内源重链和/或轻链免疫球蛋白的任何表达。轻链免疫球蛋白可以是 κ 和/或 λ 免疫球蛋白。在额外实施方案中，提供了缺少内源免疫球蛋白的至少一个等位基因的表达的转基因有蹄类动物，其中该免疫球蛋白选自重链、 κ 轻链和 λ 轻链或者其任意组合。在一个实施方案中，通过内源免疫球蛋白基因座的遗传寻靶以防止内源免疫球蛋白的表达，可以实现功能性内源免疫球蛋白的表达。在一个实施方案中，通过同源重组可以完成遗传寻靶。在另一实施方案中，通过核转移可以产生转基因有蹄类动物。

在其他实施方案中，缺少功能性内源免疫球蛋白任何表达的转基因有蹄类动物可以进一步基因修饰以表达异种免疫球蛋白基因座。在备选实施方案中，提供了含有异种免疫球蛋白基因座的猪动物。在一个实施方案中，异种免疫球蛋白基因座可以是重链和/或轻链免疫球蛋白或者其片段。在另一个实施方案中，异种免疫球蛋白基因座可以是 κ 链基因座或者其片段和/或 λ 链基因座或者其片段。在其他实施方案中，人工染色体（AC）可以含有异种免疫球蛋白。在一个实施方案中，AC 可以是酵母 AC 或者哺乳动物 AC。在另一实施方案中，异种基因座可以是人免疫球蛋白基因座或者其片段。在一个实施方案中，人免疫球蛋白基因座可以是人 14 号染色体、人 2 号染色体和人 22 号染色体或者其片段。在另一实施方案中，人免疫球蛋白基因座

可以包括人免疫球蛋白的任意片段，该片段可以经历重排。在另一实施方案中，人免疫球蛋白基因座可以包括人免疫球蛋白重链和人免疫球蛋白轻链的任意片段，其可以经历重排。在另一实施方案中，人免疫球蛋白基因座可以包括任意人免疫球蛋白基因座或者其片段，其在暴露于抗原后可以产生抗体。在具体实施方案中，应答对一种或多种抗原的暴露，所述外源人免疫球蛋白可以在 B 细胞中表达以产生异种免疫球蛋白。

在本发明的另一方面，提供了表达异种免疫球蛋白基因座或者其片段的转基因有蹄类动物，其中该免疫球蛋白可以从整合在内源有蹄类动物染色体内的免疫球蛋白基因座表达。在一个实施方案中，提供了来自转基因动物的有蹄类动物细胞。在一个实施方案中，异种免疫球蛋白基因座可以被后代遗传。在另一实施方案中，异种免疫球蛋白基因座可以通过雄性种系由后代遗传。在其他实施方案中，人工染色体 (AC) 可以含有异种免疫球蛋白。在一个实施方案中，AC 可以是酵母 AC 或者哺乳动物 AC。在另一实施方案中，异种基因座可以是人免疫球蛋白基因座或者其片段。在一个实施方案中，人免疫球蛋白基因座可以是人 14 号染色体、人 2 号染色体和人 22 号染色体或者其片段。在另一实施方案中，人免疫球蛋白基因座可以包括人免疫球蛋白的任意片段，该片段可以经历重排。在另一实施方案中，人免疫球蛋白基因座可以包括人免疫球蛋白重链和人免疫球蛋白轻链的任意片段，其可以经历重排。在另一实施方案中，人免疫球蛋白基因座可以包括任意人免疫球蛋白基因座或者其片段，其在暴露于抗原后可以产生抗体。在具体实施方案中，应答对一种或多种抗原的暴露，所述外源人免疫球蛋白可以在 B 细胞中表达以产生异种免疫球蛋白。

定义

术语“重组 DNA 技术”、“DNA 克隆”、“分子克隆”或者“基因克隆”指将 DNA 序列转移到细胞或者生物中的方法。DNA 片段可以从一种生物转移到自身复制的遗传元件（例如，细菌质粒），其允许 DNA（或者 RNA）序列的任何特定部分的拷贝在许多其他序列中选择并以无限的量产生。质粒和其他类型的克隆载体如人工染色体可以用于拷贝基因和染色体的其他片段以产生足够的相同材料用于进

一步研究。除了细菌质粒（其可以携带最高达 20kb 外源 DNA）外，其他克隆载体包括病毒、粘粒、和人工染色体（例如，细菌人工染色体（BAC）或者酵母人工染色体（YAC））。当染色体 DNA 的片段最终与它的克隆载体在实验室中连接时，将它称作“重组 DNA 分子”。将重组质粒导入合适的宿主细胞后不久，新插入的区段将随着宿主细胞 DNA 一起繁殖。

“粘粒”是人工构建的克隆载体，其携带最高达 45kb 的外来 DNA。它们可以在 λ 噬菌体颗粒中包装以用于转染到大肠杆菌 (*E. coli*) 细胞中。

如本文使用的术语“哺乳动物”（如在“基因修饰的(或者改变的)哺乳动物”中）意在包括任何非人哺乳动物，包括但不限于，猪、羊、山羊、牛、鹿、骡、马、猴、狗、猫、大鼠、小鼠、鸟、鸡、爬行动物、鱼和昆虫。在本发明的一个实施方案中，提供了基因改变的猪和其生产方法。

术语“有蹄类动物”指有蹄的哺乳动物。偶蹄类动物是偶数趾（分趾蹄的）有蹄类动物，包括羚羊、骆驼、奶牛、鹿、山羊、猪和羊。奇蹄类动物是奇数趾的有蹄类动物，其包括马、斑马、犀牛和麂。本文使用的术语有蹄类动物指成年、胚胎或者胎儿有蹄类动物。

本文使用的术语“猪 (porcine)”、“猪动物”、“猪 (pig)”和“猪 (swine)”是一般术语，指相同类型的动物，而不关心性别、大小或者品种。

“同源 DNA 序列或者同源 DNA”是与参考 DNA 序列至少约 80%、85%、90%、95%、98% 或 99% 相同的 DNA 序列。同源序列在严格条件下与靶序列杂交，严格杂交条件包括如果两个序列之间有至少 85%、至少 95% 或者 98% 同一性，那么将允许它们杂交的杂交条件。

“等基因或者基本上等基因 DNA 序列”是与参考 DNA 序列相同或者几乎相同的 DNA 序列。术语“基本上等基因”指与参考 DNA 序列至少约 97-99% 相同，或者与参考 DNA 序列至少约 99.5-99.9% 相同，并且在一些用途中与参考 DNA 序列 100% 相同的 DNA。

“同源重组”指基于序列同源性的 DNA 重组过程。

“基因寻靶”指两个 DNA 序列之间的同源重组，所述序列之一位于染色体上并且另一个不在染色体上。

“非同源的或者随机整合”指 DNA 借以整合到基因组但是不涉及同源重组的任意过程。

“选择标记基因”是一种基因，其表达允许鉴定含有该基因的细胞。选择标记可以是允许细胞在防止或者延缓没有该基因的细胞生长的培养基上增殖的选择标记。实例包括抗生素抗性基因和允许生物在选择的代谢物上生长的基因。备选地，该基因可以通过赋予细胞容易鉴定的表型来促进转化体的视觉筛选。这种可鉴定的表型可以例如，是发光的产生或者有色化合物的产生，或者细胞周围培养基中可检测的改变的产生。

术语“连续的”在本文中以它的标准意思使用，即没有中断或者未中断的。

“严格条件”指这样的条件：（1）利用低离子强度和高温度进行洗涤，例如，0.015 M NaCl/0.0015 M 柠檬酸钠/0.1% SDS 于 50℃，或者（2）在杂交过程中利用变性剂，如甲酰胺。本领域技术人员可以确定并适当改变严格条件以得到清楚和可检测的杂交信号。例如，在杂交和洗涤期间通过升高存在的盐含量、降低温度或者它们的组合一般可以降低严格性。见，例如，Sambrook 等人，*Molecular Cloning: A Laboratory Manual*, Cold Spring Harbour Laboratory Press, Cold Spring Harbour, New York, (1989)。

I. 免疫球蛋白基因

在本发明的一方面，提供了缺少功能性内源免疫球蛋白的任何表达的转基因有蹄类动物。在一个实施方案中，该有蹄类动物可以缺少内源重链和/或轻链免疫球蛋白的任何表达。轻链免疫球蛋白可以是 κ 和/或 λ 免疫球蛋白。在额外实施方案中，提供了缺少内源免疫球蛋白的至少一个等位基因的表达的转基因有蹄类动物，其中该免疫球蛋白选自重链、 κ 轻链和 λ 轻链或者其任意组合。在一个实施方案中，通过内源免疫球蛋白基因座的遗传寻靶以防止内源免疫球蛋白的表达，可以实现功能性内源免疫球蛋白的表达。在一个实施方案中，通过同源重组可以完成遗传寻靶。在另一实施方案中，通过核转移可以产生转基因有蹄类动物。

在本发明的另一方面，提供了破坏有蹄类动物免疫球蛋白基因表

达的方法，其包括 (i) 分析有蹄类动物重链、 κ 轻链或 λ 轻链基因组基因座的种系构型；(ii) 确定该基因座的至少一个功能区域的 5' 末端和 3' 末端侧翼的核苷酸序列位置；和 (iii) 将含有侧翼序列的寻靶构建体转染到细胞中，其中，在成功地同源重组后，破坏该免疫球蛋白基因座的至少一个功能区，从而减小或者防止该免疫球蛋白基因的表达。

在一个实施方案中，提供了猪重链基因座的种系构型。猪重链基因座含有至少四个可变区、两个多样性区、六个连接区和五个恒定区，例如，如图 1 图解。在特定实施方案中，六个连接区的仅一个：J6 是功能性的。

在另一实施方案中，提供了猪 κ 轻链基因座的种系构型。猪 κ 轻链基因座含有至少六个可变区、六个连接区、一个恒定区和一个增强子区，例如，如图 2 图解。

在另一实施方案中，提供了猪 λ 轻链基因座的种系构型。

提供了如 Seq ID Nos 1-39 中所述的分离的核苷酸序列。还提供了与 Seq ID Nos 1-39 的任一个至少 80、85、90、95、98 或 99% 同源的核酸序列。此外，提供了含有 Seq ID Nos 1-39 任一个的至少 10、15、17、20、25 或 30 个连续核苷酸的核苷酸序列。还提供了任选在严格条件下与 Seq ID Nos 1-39 杂交的核苷酸序列，以及其同源核苷酸。

在核苷酸或者氨基酸序列水平上的同源性或者同一性可以通过 BLAST(Basic Local Alignment Search Tool)分析，使用程序 blastp、blastn、blastx、tblastn 和 tblastx(见例如，Altschul, S. F. 等人, (1997) *Nucleic Acids Res* 25:3389-3402 和 Karlin 等人, (1990) *Proc. Natl. Acad. Sci. USA* 87, 2264- 2268) (它们被定制以用于序列相似性搜索) 所用的算法来确定。BLAST 程序使用的方法是首先考虑查询序列和数据库序列之间的相似性区段(有和没有缺口)，然后评估所鉴定的所有匹配的统计学显著性，并最后仅仅总结满足预先选择的显著性阈值的那些匹配。见例如，Altschul 等人, (1994) (*Nature Genetics* 6, 119-129)。直方图、描述、比对、期望(即，对数据库序列报告匹配的统计学显著性阈值)、截止值(cutoff)、矩阵和滤器(低复杂性) 的搜索参数处于默认设置。blastp、blastx、tblastn 和 tblastx 的默认评

分矩阵是 BLOSUM62 矩阵 (Henikoff 等人, (1992) Proc. Natl. Acad. Sci. USA 89, 10915-10919), 推荐其用于长度超过 85 (核苷酸碱基或者氨基酸) 的查询序列。

猪重链

在本发明的另一方面, 提供了编码有蹄类动物免疫球蛋白的重链基因座的新的基因组序列。在一个实施方案中, 提供了编码猪重链的分离的核苷酸序列, 其包括至少一个可变区、两个多样性区、至少四个连接区和至少一个恒定区, 如 μ 恒定区, 例如, 如 Seq ID No. 29 中表示。在另一实施方案中, 提供了分离的核苷酸序列, 其包括猪重链基因组序列的至少四个连接区和至少一个恒定区, 如 μ 恒定区, 如 Seq ID No.4 中表示。在另一实施方案中, 提供了核苷酸序列, 其包括猪重链基因组序列的第一个连接区的 5'侧翼序列, 如 Seq ID No. 1 中表示。此外, 提供了核苷酸序列, 其包括猪重链基因组序列的第一个连接区的 3'侧翼序列, 如 Seq ID No.4 的 3'区中表示。在进一步实施方案中, 提供了如 Seq ID No.1、4 或 29 所述的分离的核苷酸序列。还提供了与 Seq ID No.1、4 或 29 至少 80、85、90、95、98 或 99% 同源的核酸序列。还提供了任选在严格条件下与 Seq ID No.1、4 或 29 杂交的核苷酸序列以及其同源核苷酸。

此外, 提供了含有 Seq ID No.1、4 或 29 的至少 10、15、17、20、25 或 30 个连续核苷酸的核苷酸序列。在一个实施方案中, 该核苷酸序列含有 Seq ID No 4 的至少 17、20、25 或 30 个连续核苷酸或者 Seq ID No. 29 的残基 1-9,070。在其他实施方案中, 提供了含有 Seq ID No. 29 的至少 50、100、1,000、2,500、4,000、4,500、5,000、7,000、8,000、8,500、9,000、10,000 或 15,000 个连续核苷酸的核苷酸序列。在另一实施方案中, 该核苷酸序列含有 Seq ID No. 29 的残基 9,070-11039。

在进一步实施方案中, 提供了如 Seq ID No.1、4 或 29 所述的分离的核苷酸序列。还提供了与 Seq ID No.1、4 或 29 至少 80、85、90、95、98 或 99% 同源的核酸序列。此外, 提供了含有 Seq ID No.1、4 或 29 的至少 10、15、17、20、25 或 30 个连续核苷酸的核苷酸序列。还提供了任选在严格条件下与 Seq ID No.1、4 或 29 杂交的核苷酸序列以及其同源核苷酸。

在一个实施方案中，提供了编码猪重链的分离的核苷酸序列，其包括至少一个可变区、两个多样性区、至少四个连接区和至少一个恒定区，如 μ 恒定区，例如，如 Seq ID No. 29 中表示。在 Seq ID No. 29 中，重链的多样性区例如通过残基 1089-1099 (D(pseudo))表示，重链的连接区例如通过残基 1887-3352 (例如：J(psuedo): 1887-1931, J(psuedo): 2364-2411, J(psuedo): 2756-2804, J (功能性 J): 3296-3352) 表示，重组信号例如通过残基 3001-3261 (九碱基顺序), 3292-3298 (七碱基序列)表示，恒定区通过下面的残基表示：3353-9070 (J 到 C μ 内含子)、5522 - 8700 (转换区)、9071-9388 (μ 外显子 1)、9389-9469 (μ 内含子 A)、9470-9802 (μ 外显子 2)、9830- 10069 (μ 内含子 B)、10070-10387 (μ 外显子 3)、10388-10517 (μ 内含子 C)、10815-11052 (μ 外显子 4)、11034-11039 (聚腺苷酸信号)。

Seq ID No. 29	<pre> tclagaagacgtggagagaggccagacttcctcggaacagctcaaagag ctctgtcaaagccagatcccatcacacgtggccaccaataggccatgcca gcciccaaaggccgaactgggttctccacggcgacatgaagcctgcagc ctggcttatcctcttccgtgtaagaggcaggcccgggactggacgagg ggctagcagggtgtgtaggcaccttgcgccccccacccggcgaggaacc agagaccctggggctgagagtgaacctcaaacaggatgccccacccctc aggccaccttcaatccagctacactccacctgocattctcctgggca cagggcccaagcccctggatctggccttggctcgaactgacccacgcgc acacacacacttccaaagtgtgtcggctcacccctccccagcgtgtc catgggcagcacggcagtgccgctccggcggtagtgagtgagaggtccc ttcccctccccaggagccccagggtgtgtgagatctggggctcctg tcccttacaccttcatgcccctcccctatacccacctccaggcgggag gcagcgagacctttcccagggactcagccaacgggcacacgggagggca gccctcagcagctggtcccaagaggagggtgggaggtaggccacagct gccacagagagaaccctgacggaccacaggggccaagccagccggaa ccagctccctcgtgggtgagcaatggccaggggccccggccaccacgg ctggccttgcgccagctgagaaticacgtccagtgcaggggagactcaaga cagcctgtcacacagcctcggatctgctcccattcaagcagaaaaagg aaaccgtgcaggcagccctcagcattcaaggattgtgacgaggccaac tattctggcagtgccgattagaatgacctggagaaggcggaagg tctctgtggctctgcggccaacaggccctggctccacctgcccgtgc cagcccgaggggcttgggctgagccagggaaccacagtgetaotgggacc acagtgactgaccaaactcccggccagagcagcccaaggccagccggct ctgcccctggaggactcaccatcagatgcacaaggggcgagtggaag agacgtgtcggccggccatttgggaaggcgaaggacctccaggtgga caggagggtggagcactccaggcaagggaactgggtcccaaggcctggg </pre>
---------------	--

```

gaagggtactggctgggggtagcctggccagggaacggggagcgggg
cggggggctgagcaggggagacctgacctgiggagcagggcaagtcag
gcttcaggcagcagccgacatcccagaccaggagctgagcagggagg
gctgcaagcggggcggggcctgctgctcgggggctcctgggggacg
ctggctctgtttccgtgtcccgcagcacaggccagctcgtgggccta
tgcttaacctgagtgctggggccggggcgtcagggtcgtctcctcag
gggagagtcctccigaggctacgctgggg*ggggactatggcagctcacc
aggggcctggggaccaggggcctggaccaggctgcagcccggaggagg
cagggctctggctcagcatctggccctcgaaatggcagaacccctg
gcgggtgagcagctgagcgggctcagacagacaggggccggcggaaa
ggagaaagtggggcagagcccggaggccagggcccaagggtctggt
ggcaggcctgggtgggacatgggtggtgccaaggctacttagattcgt
ggggccaggggcctcggctaccgctcctcaggtgagcctggtgctga
gtccaagctagcgtgggtggccggggctggcctgctcaggtaggg
caggggctgggatgtgattgtcaaggaggggcaacagggtgcagactg
tgccctggaactgaccactggggcaggggcgtcctggtcagctctcc
tcaggtaagacggcctgtgcccctctcgcgggactggaaaaggaaat
ttcaagattcctgtctgtgtggggccctctggggcccccgggggtg
ctccctcctgccagatggggcctggcctgtggagcacgggctgggca
cacagctcaggtctaggggccagagcccgggctcagggtctgtgtgg
cccggcactggcaggggctcgggtttttggaccccctaagggggc
cacagcactgtaccatctcagctggggcggaggagtcgaggtcacc
gtctctcaggtgagctctgtagccctctcactctctggggggtt
tgctgattttgtggggaaaggagatgcctgggtctcaggtctaaaggt
ctaggccagcgcggggcccaaggaaaggggccgaggggccaggtcggct
cggccaggagcagagcttcagacatctgctcctggcggctcagctca
ggcctttggccgggggggtcagcaccaccaggcctctggctccggag
gtcccggcccggctgcctcaccaggcaccgtgcgggtggcccgggc
tcttgctggccacccttcttaactgggatccgggcttagttgtgcaa
tgtgacaacgggctcgaaggctggggccaggggaccctagctacgagc
ctcgggtgggtgtcccgcacccctcccactttcacgcactcggcgaga
cctgggagtcaggtgtggggacattggaggtcaggaacgggagctg
gggagaggcctctcagcgggggtccagagatggggccctccaaggac
ggcctgcgggggacaagggtctgtggcctggcctggccctcacttg
gggtcagggggggcttcccgggcaagggtcagtcgagggggtgga
attcaggtctgggtcggggtgggggtggcctcatgaacagacag
cccaggcggccgtgtttggccctggggcctggttgaatgcaggt
ctcgggaaagtcaggaggagcctggccagcagagggttccagccctgcg
gcccaggaccctgagacgggagggcattggcgtcagggccagggcc
accccccaGGTTTGTggggcgagcctggagattgcacCACTGTGAT
TACTATGCTATGGATCTCTGGGGCCCAAGCGTTGAAGTCGTCGTCTCCTC
AGgtaagaacggcctccaggcccttaattctgctctgtctgtggc
tttctgactctgacctcgggagcgtctgtgccccccgggatgag
gcccgttccaggagggtcagggaaccaggagcctgtgggaagtctga
cgggggctcaggcgggaaggcccccaccgggggagcccaaggccg
tgggcgagagaccctgagaggtgccttgaggagggtgctcggga
accacgaaccccggggagggcctgctgcaatgcggtctcagacgg
gaggcgttctgcccctaacgctcttcaaggccctgtgggtcgaaga
gccaatgctggagagagagggacaggcctgcccgaactggccgagagcg
ggcagccccggggagagcggggcagcctgggctctgtgaggccag
gtccaaggaggagctgtgtcctcgtgacaggtgcacttgcgaaacct
agaagacggggtatgttggaagcggctcctgatgtaaaaggggaga
ctgtaaaagtgcagagctcctcaagtgttaagggttaaaaggtcaag
igttaaacctttgactgcagttagcaagcgtgcggggagtgatgg
ggtgccaagggtggcgaaggcagtagcaggggcgtgcccgtctcctaatt
cagggcitagtgtgcaataaagctggcctgttttcaaaagcattgg

```

```

tggctgagctgggtggaggaggccggcagccctggccacctgacgc
aggtggcaggaagcaggctggccaaggagctatttaggaagccaagaaa
cacggctgatgaattatagctcttggttccaggagggtggtggcattg
gctttgcgcagccacagaaccgaaagtgccactgagaaaaacaact
ctgcttaattgcatitttctaaaagaagaacagaggctgacggaaac
tggaaagttcctgtttaactactcgaattgagtttcggcttagctta
tcaactgctcacttagattcatttcaagtaaacgtttaagagccgagg
cattcctalcctctictaaggcgttattcctggaggctcattcaccgcca
gcacciccgcctgcctgcaggcattgctgcaccgtcaccgtgacggcgcg
cacgatttcaagtggcccgtccctcgtgattaggacagacggggc
actctgcccagccgtcttggctcagtatctgcaggcgtccgtctcggga
cggagctcaggggaagagcgtgactccagttgaacgtatagtcgggtgc
ttagaggagaccagtcgggtgctgagtcagaaggggcccggggcccga
ggccctgggcaaggacggccgtgacctgcatcacggcccagcgtctag
aggcaggactctggtggagagtgtgagggtgctggggcccctccggagc
tggggccgtgagggtgacgggtggctctcggcggcgggtgtggtgtttct
gcgggattggaggaattctccagtgtgaggagtcgccagtgaccgggc
accaggctggttaagaggaggccgctgctggtgccaagcagctggggagg
gttcggtaaaaggctcggcgttccctttaaaggagactttcctggagg
gcatttagctagtcgggaccgtttcagctcgggaagaggatgctggag
gaggcgatgtcccaggagccgaaggcggcggggagagcccaagggtct
ctctgtcccacagaggcagccactgcccagacagacagggccctt
ccctctgatgacggcaaaaggcctcgtcttgcgggtgctggggggg
agtcccccgaagccgctcaccagaggcctgaggggtgagactgaccga
tgctcttggccggcctggggcggaccgagggggactccgtggaggca
ggcgatggtgctgctgggagggaaccgacctggccgagcccggctg
gcgattcccggcgaaggccctcagccgagggcagtggtgctcggcggaac
cacccttctggccaagccacagggtctcgggaotgtccggggcagc
ctggcgtcccgtggcagggcTGGGCTGACCTGGACTTACCAGACAGAA
CAGGGCTTTCAGGGCTGAGCTGAGCCAGGTTTAGCGAGGCCAAGTGGGGC
TGAACCAGGCTCAACTGGCCTGAGCTGGGTTGAGCTGGGCTGACCTGGGC
TGAGCTGAGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGAC
TGGCTGAGCTGAGCTGGGTTGAGCTGAGCTGAGCTGGCCTGGGTTGAGCT
GGGCTGGGTTGAGCTGAGCTGGGTTGAGCTGGGTTGAGCTGGGTTGATCT
GAGCTGAGCTGGGCTGAGCTGAGCTAGGCTGGGGTGAGCTGGGCTGAGCT
GGTTTGAGTTGGGTTGAGCTGAGCTGAGCTGGGCTGTGCTGGCTGAGCTA
GGCTGAGCTAGGCTAGGTTGAGCTGGGCTGGGCTGAGCTGAGCTAGGCTG
GGCTGATTTGGGCTGAGCTGAGCTGAGCTAGGCTGCGTTGAGCTGGCTGG
GCTGGATTGAGCTGGCTGAGCTGGCTGAGCTGGGCTGAGCTGGCCTGGGT
TGAGCTGAGCTGGACTGGTTTGGCTGGGTCGATCTGGGTTGAGCTGTCC
TGGGTTGAGCTGGGCTGGGTTGAGCTGAGCTGGGTTGAGCTGGGCTCAGC
AGAGCTGGGTTGGGCTGAGCTGGGTTGAGCTGAGCTGGGCTGAGCTGGCC
TGGGTTGAGCTGGGCTGAGCTGAGCTGGGCTGAGCTGGCCTGTGTTGAGC
TGGGCTGGGTTGAGCTGGGCTGAGCTGGATTGAGCTGGGTTGAGCTGAGC
TGGGCTGGGCTGTGCTGACTGAGCTGGGCTGAGCTAGGCTGGGTTGAGCT
GGGCTGAGCTGATCCGAGCTAGGCTGGGCTGGTTTGGGCTGAGCTGAGCT
GAGCTAGGCTGGATTGATCTGGCTGAGCTGGGTTGAGCTGAGCTGGGCTG
AGCTGGTCTGAGCTGGCCTGGGTCGAGCTGAGCTGGACTGGTTTGGCTG
GGTCGATCTGGGCTGAGCTGGCCTGGGTTGAGCTGGGCTGGGTTGAGCTG
AGCTGGGTTGAGCTGGGCTGAGCTGAGGCTGGGTTGAGCTGGGCTGAAC
TAGCCTAGCTAGGTTGGGCTGAGCTGGGCTGGTTTGGGCTGAGCTGAGCT
GAGCTAGGCTGCATTGAGCAGGCTGAGCTGGGCTGAGCAGGCTGGGTTG
AGCTGGGCTAGGTTGGAGCTGAGCTGGGTCGAGCTGAGTTGGGCTGAGCTG
GCCTGGGTTGAGGTAGGCTGAGCTGAGCTGAGCTAGGCTGGGTTGAGCTG
GCTGGGCTGGTTTGCCTGGGTTCAAGCTGGGCCGAGCTGGCCTGGGTTGA
GCTGGGCTCGGTTGAGCTGGGCTGAGCTGAGCCGACCTAGGCTGGGATGA

```

GCTGGGCTGATTTGGGCTGAGCTGAGCTGAGCTAGGCTGCATTGAGCAGG
CTGAGCTGGGCCTGGAGCCTGGCCTGGGGTGAGCTGGGCTGAGCTGGCCT
GAGCTAGGCTGGGTTGAGCTGGCTGGGCTGGTTTGGCGCTGGGTCAAGCTG
GGCCGAGCTGGCCTGGGATGAGCTGGGCCGGTTTGGGCTGAGCTGAGCTG
AGCTAGGCTGCATTGAGCAGGCTGAGCTGGGCTGAGCTGGCCTGGGGTGA
GCTGGGCTGAGCTAAGCTGAGCTGGGCTGGTTTGGGCTGAGCTGGCTGAG
CTGGGTCCTGCTGAGCTGGGCTGAGCTGACCAGGGGTGAGCTGGGCTGAG
TTAGGCTGGGCTCAGCTAGGCTGGGTTGATCTGGCAGGGCTGGTTTGGCG
TGGGTCAAGCTCCCGGAGATGGCCTGGGATGAGCTGGGCTGGTTTGGGC
TGAGCTGAGCTGAGCTGAGCTAGGCTGCATTGAGCAGGCTGAGCTGGGCT
GAGCTGGCCTGGGGTGAGCTGGGCTGGGTGGAGCTGAGCTGGGCTGAACT
GGGCTAAGCTGGCTGAGCTGGATCGAGCTGAGCTGGGCTGAGCTGGCCTG
GGGTTAGCTGGGCTGAGCTGAGCTGAGCTAGGCTGGGTTGAGCTGGCTGG
GCTGGTTTGGCGCTGGGTCAAGCTGGGCCGAGCTGGCCTGGGTTGAGCTGG
GCTGGGCTGAGCTGAGCTAGGCTGGGTTGAGCTGGGCTGGGCTGAGCTGA
GCTAGGCTGCATTGAGCTGGCTGGGATGGATTGAGCTGGCTGAGCTGGCT
GAGCTGGCTGAGCTGGGCTGAGCTGGCCTGGGTTGAGCTGGGCTGGGTTG
AGCTGAGCTGGGCTGAGCTGGGCTCAGCAGAGCTGGGTTGAGCTGAGCTG
GGTTGAGCTGGGCTGAGCTGGGCTGAGCAGAGCTGGGTTGAGCTGAGCTG
GGTTGAGCTGGGCTCGAGCAGAGCTGGGTTGAGCTGAGCTGGGTTGAGCT
GGGCTCAGCAGAGCTGGGTTGAGCTGAGCTGGGTTGAGCTGGGCTGAGCT
AGCTGGGCTCAGCTAGGCTGGGTTGAGCTGAGCTGGGCTGAACTGGGCTG
AGCTGGGCTGAACTGGGCTGAGCTGGGCTGAGCTGGGCTGAGCAGAGCTG
GGCTGAGCAGAGCTGGGTTGGTCTGAGCTGGGTTGAGCTGGGCTGAGCTG
GGCTGAGCAGAGTTGGGTTGAGCTGAGCTGGGTTGAGCTGGGCTGAGCTG
GGCTGGGTTGAGCTGGGTTGAGTTGGGCTGAGCTGGGCTGGGTTGAGCAG
AGCTGGGCTGAACTGGGCTGAGCTGGGCTGAGCAGGAACTGGGTTGATCTG
AATTGAGCTGGGCTGAGCCGGGCTGAGCCGGGCTGAGCTGGGCTAGGTTG
AGCTTGGGTTGAGCTTGCCTCAGCTGGTCTGAGCTAGGTTGGGTGGAGCTA
GGCTGGATTGAGCTGGGCTGAGCTGAGCTGATCTGGCCTCAGCTGGGCTG
AGGTAGGCTGAACTGGGCTGTGCTGGGCTGAGCTGAGCTGAGCCAGTTT
AGCTGGGTTGAGCTGGGCTGAGCTGGGCTGTGTTGATCTTCTGAACTG
GGCTGAGCTGGGCTGAGCTGGCCTAGCTGGATTGAAAGGGGTAAGCTGG
GCCAGGCTGGAAGCTGGGCTGAGCTGAGCTAGGCTGAGCTGAGTTGAAATTG
GTTAAGCTGGGCTGAGATGGGCTGAGCTGGGCTGAGCTGGGCTGAGCTGG
GTCCGACTGGGTTACCTGGGCCACACTGGGCTGAGCTGGGCCGAGCTCG

ataacctggctcaggctgagtcgggtccagcagacatgcgctggccaggc
tgcttgacctggacacgttcgatgagctgcttgggatggtcacctca
gctgagccaggctgctccagctgggctgagctggtgacctgggtgacct
cggtagccaggctgtctgagctccggccaagccaggctgcatcagact
cgccagaccnaaggcctgggccccggctggcaagccaggggcggtgaagg
ctgggctggcaggactgtcccgaaggaggtgcactggagccgcccggga
ccccgaccggcaggacctggaagagcctctcactcccccttctctct
gtccccctcgggtccctcagAGAGCCAGTCTGCCCGAATCTCTACCCCC
TCGTCTCCTGCGTCAGCCCCCGTCCGATGAGAGCCTGGTGGCCCTGGGC
TGCTGGCCCCGGACTTCTGCCCAGCTCCGTACCTTCTCCTGGAACTA
CAAGAACAGCAGCAAGGTCAGCAGCCAGAACATCCAGGACTTCCCCTCC
G
TCCTGAGAGGCGGCAAGTACTTGGCCTCCTCCCGGGTGCTCCTACCCTCT
GTGAGCATCCCCAGGACCCAGAGGCTTCTGGTGTGCGAGGTCCAGCA
CCCCAGTGGCACCAAGTCCGTGTCCATCTCTGGGCCAGgtgagctgggct
ccccctgtgctgtggcggggcggggccgggtgccggggcacagtac
gccccgttctgctgagTCGTAGAGGAGCAGCCCCCGTCTTGAACAT
CTTCGTCCCCACCCGGGAGTCTTCTCCAGTACTCCCAGCGCACGTCCA
AGCTCATCTGCCAGGCTCAGACTTACGCCCAAGCAGATCTCCATGGCC
TGGTTCCTGTGATGGGAAACGGGTGGTGTCTGGCGTCAGCACAGGCCCGT

	<p>GGAGACCCTACAGTCCAGTCCGGTGAACCTACAGGCTCCACAGCATGGTGA CCGTACGGAGTCCGAGTGGCTCAGCCAGAGCGTCTTACCTGCCAGGTG GAGCACAAGGGCTGAACTACGAGAAGAACGCGTCTCTCTGTGCACCTC CAgtgagtgagccctcgggcccggcggggcgggcgggagccacaca cacaccagctgctccctgagccttggcttccgggagtgccaaggcgggg agggctgtgagggcagctggaggcactgtcagctggggcccagcacc ccctaccccggcagggcccgggctccgaggggccccgagtcgagggcc ctgclttggggaagccctacttggccctcagggcgctgacgctccc cccaccaccccgcctagATCCCAACTCTCCATCACCGTCTTCGCCAT CGCCCCCTCCTTCGCTGGCATCTTCTCAACAGTCGGCCAAGCTTTCT GCCTGGTACGGGCTCGTACCAGGGAGAGCCTCAACATCTCCTGGACC CGCCAGGACGGCGAGGTTCTGAAGACCAAGTATCGTCTTCTCTGAGATCTA CGCCAACGGCACCTTCGGCGCCAGGGGCGAAGCCTCCGTCTGCGTGGAGG ACTGGGAGTGGGGGACAGGTTACGTGCACGGTGACCCACACGGACCTG CCCTCGCCGCTGAAGCAGAGCGTCTCCAAGCCCAGAGGtagccctgccc tgccctgctccggcctgtgcttggccgcccggggcgggagccga gcctggcgaggagcgcctcggcccccgcggtcccgaaccacaccct cctgctctctcccagGGATCGCCAGGCACATGCCGTCCGTGTACGTGC TGCCGCGGCCCGGAGGAGCTGAGCCTGCAGGAGTGGGCTCGGTACC TGCTGTGAAGGGCTTCTCCCGGCGGACGTGTTCTGTGACAGTGGCTGCA GAAGGGGAGCCCGTGTCCGCGGACAAGTACGTGACAGCGCGCCGGTG C CCGAGCCCCGAGCCCAAGGCCCCCGCCTCCTACTTCGTGCAGAGCGTCTG ACGGTGAGCGCCAAGGACTGGAGCGACGGGGAGACCTACACCTGCGTGC T GGGCCACGAGGCCCTGCCCCACACGGTGACCGAGAGGACCGTGGACAAG T CCACCGGTAACCCACCCTGTACAACGTCTCCCTGGTCTGTCCGACAAG GCCAGCACCTGCTACTGACCCCTGGCTGCCCGCCGCGGCCGGGGCCAGA GCCCCGGGGCACCATCGCTCTGTGTGGGCTGTGTGCAACCCGACCCGT TCGGGGTGAGCGGTGCGATTTCTGAAAATTAGAAataaaAGATCTCGTGC CG</p>
Seq ID No.1	<p>TCTAgAAGACGCTGGAGAGAGGGCCagACTTCCTCGGAACAGCTCAAAGAG CTCTGTCAAAGCCAGATCCCATCACACGTGGGCACCAATAGGCCATGCCA GCCTCAAAGGGCCGAAC TGGGTTCTCCACGGCGCACATGAAGCCTGCAGC CTGGCTTATCTCTTCCGTGGTGAAGAGGGCAGGCCCGGGACTGGACGAGG GGCTAGCAGGGTGTGGTAGGCACTTGGCCCCCAACCCGGCAGGAACC AGAGACCCTGGGGCTGAGAGTGAAGCCTCAAACAGGATGCCCAACCCCTT AGGCCACCTTCAATCCAGCTACACTCCACCTGCCATTCTCCTCTGGGCA CAGGGCCCAGCCCTGGATCTTGGCCTTGGCTCGACTTGCACCCACGCGC ACACACACTTCTAACGTGCTGTCCGCTCACCCCTCCCCAGCGTGGTC CATGGGCAGCACGGCAGTGGCGTCCGGCGGTAGTGAGTGCAGAGGTCCC TTCCCTCCCCAGGAGCCCCAGGGGTGTGTGCAATCTGGGGGCTCCTG TCCCTTACACTTCATGCCCTCCCCTCATACCCACCTCCAGGCGGGAG GCAGCGAGACCTTTGCCAGGGACTCAGCCAACGGGCACACGGGAGGCC A GOCCTCAGCAGTGGG</p>
Seq ID No.4	<p>GGCCAGACTTCCTCGGAACAGCTCAAAGAGCTCTGTCAAAGCCAGATCCC ATCACACGTGGGCACCAATAGGCCATGCCAGCCTCAAAGGGCCGAAC TGG GTTCTCCACGGCGCACATGAAGCCTGCAGCCTGGCTTATCCTCTTCCGTG GTGAAGAGGCAGGCCCGGACTGGACGAGGGGCTAGCAGGGTGTGGTAG GCACCTTGGCCCCCAACCCGGCAGGAACCAAGAGACCCTGGGGCTGAGA G</p>

TGAGCCTCCAAACAGGATGCCCCACCCTTCAGGCCACCTTTCAATCCAGC
 TACACTCCACCTGOCATTCTCCTCTGGGCACAGGGCCCAGCCCCTGGATC
 TTGGCCTTGGCTCGACTTGCACCCACGCGCACACACACTTCTAACGT
 GCTGTCCGCTACCCCTCCCCAGCGTGGTCCATGGGCAGCACGGCAGTGC
 GCGTCCGGCGGTAGTGAGTGCAGAGGTCCCTTCCCCTCCCCAGGAGCCC
 CAGGGGTGTGTGCAGATCTGGGGGCTCCTGTCCCTTACACCTTCATGCC
 CTCCCCTCATACCCACCCTCCAGGCGGGAGGCAGCGAGACCTTTGCCAG
 GGACTCAGCCAACGGGCACACGGGAGGCCAGCCCTCAGCAGCTGGCTCCC
 AAAGAGGAGGTGGGAGGTAGGTCCACAGCTGCCACAGAGAGAAACCTG
 ACGGACCCACAGGGGCCACGCCAGCCGGAACCAGCTCCCTCGTGGGTGA
 GCAATGGCCAGGGCCCCGCGGCCACCACGGCTGGCCTTGCGCCAGCTGA
 G
 AACTCACGTCCAGTGCAGGGAGACTCAAGACAGCCTGTGCACACAGCCTC
 GGATCTGCTCCATTTCAGCAGAAAAAGGAAACCGTGCAGGCAGCCCTC
 AGCATTTCAGGATTGTAGCAGCGGCCAACTATTCGTCCGCAGTGGCCGA
 TTAGAATGACCGTGGAGAAGGGCGGAAGGGTCTCTCGTGGGCTCTCGCGC
 CAACAGGCCCTGGCTCCACCTGCCCGCTGCCAGCCCGAGGGGGCTTGGGCC
 GAGCCAGGAACCACAGTGTCTACCGGGACCACAGTACTGACCAAACCTCC
 CGGCCAGAGCAGCCCCAGGCCAGCCGGGCTCTCGCCCTGGAGGACTCACC
 ATCAGATGCACAAGGGGGCGAGTGTGGAAGAGACGTGTCCCGGGCCA
 T
 TTGGGAAGGCGAAGGGACCTTCCAGGTGGACAGGAGGTGGGACGCACTC
 C
 AGGCAAGGACTGGGTCCCCAAGGCCTGGGGAAGGGTACTGGCTTGGG
 G
 GTTAGCCTGGCCAGGGAACGGGGAGCGGGGCGGGGGCTGAGCAGGGAG
 G
 ACCTGACCTCGTGGGAGCGAGGCAAGTCAGGCTTCAGGCAGCAGCCGCAC
 ATCCCAGACCAGGAGGCTGAGGCCAGGAGGGGCTTGCAGCGGGGCGGGG
 C
 CTGCCTGGCTCCGGGGGCTCCTGGGGACGCTGGCTCTTGTTCOGTGT
 CCGCAGCACAGGGCCAGCTCGCTGGGCCTATGCTTACCTTGATGTCTGGG
 GCCGGGGCGTCAGGTCGTCTCTCCTCAGGGGAGAGTCCCCTGAGGCTA
 CGCTGGGG*GGGACTATGGCAGCTCCACCAGGGGCTGGGGACCAGGG
 G
 CCTGGACCAGGCTGCAGCCCGGAGGACGGGCAGGGCTCTGGCTCTCCAGC
 ATCTGGCCCTCGGAAATGGCAGAACCCTGGCGGGTGGAGGAGCTGAGA
 G
 CGGGTCAGACAGACAGGGGCCGGCCGGAAGGAGAAGTTGGGGCAGAG
 C
 CCGCCAGGGGCCAGGCCCAAGGTTCTGTGTGCCAGGGCCTGGGTGGGCAC
 ATTGGTGTGGCCATGGCTACTTAGATTTCGTGGGGCCAGGGCATCCTGGTC
 ACCGTCTCCTCAGGTGAGCCTGGTGTCTGATGTCCAGCTAGGCGTGGTG
 GGCCGCGGGTGGCCCTGTCTCAGGCTAGGGCAGGGGCTGGGATGTGATT
 TGTCAAGGAGGGGCAACAGGGTGCAGACTGTGCCCTGAAACTTGACCA
 CTGGGGCAGGGGCGTCTGGTACGTCTCCTCAGGTAAGACGGCCCTGTG
 CCCCTCTCTCGCGGACTGAAAAGGAATTTTCCAAGATTCCTTGGTCTG
 TGTGGGGCCCTCTGGGGCCCCGGGGTGGCTCCCCTCCTGCCAGATGG
 GGCCTCGGCCTGTGGAGCACGGGCTGGGCACACAGCTCGAGTCTAGGGCC
 ACAGAGGCCCGGGCTCAGGGCTCTGTGTGGCCCGGCACTGGCAGGGGG
 C
 TCGGGTTTTTGGACACCCCCTAATGGGGGCCACAGCACTGTGACCATCTT
 CACAGCTGGGGCCGAGGAGTCGAGGTCAACGCTCCTCAGGTGAGTCTC
 GTCAGCCCTCTCTACTCTCTGGGGGGTTTTGCTGCATTTTGTGGGGAA
 AGAGGATGCCTGGGTCTCAGGTCTAAAGGTCTAGGGCCAGCCCGGGGCC
 CAGGAAGGGGCCGAGGGGCCAGGCTCGGCTCGGCCAGGAGCAGAGCTTC

C
AGACATCTCGCCTCCTGGCGGCTGCAGTCAGGCCTTTGGCCGGGGGGTCTC
TCAGCACCACCAGGCCTCTGGCTCCCGAGGTCCCCGGCCCCGGCTGCCT
CACCAGGCACCGTGC CGCGTGGGCCCGGGCTCTTGGTCCGCCACCTTTC
TTAACTGGGATCCGGGCTTAGTTGTGCAATGTGACAACGGGCTCGAAAG
CTGGGGCCAGGGGACCCTAGT*TACGACGCCTCGGGTGGGTGTCCCGCAC
CCCTCCCACTTTCACGGCACTCGGCGAGACCTGGGGAGTCAGGTGTTGG
GGACACTTTGGAGGTCAGGAACGGGAGCTGGGGAGAGGGCTCTGTCAGC
G
GGGTCCAGAGATGGGCCGCCCTCCAAGGACGCCCTGCGCGGGGACAAGG
G
CTTCTTGGCCTGGCCTGGCCGTTCACTTGGGCGTCAGGGGGGGCTTCCC
GGGGCAGGCGGTCAAGTCGAGGCGGGTTGGAATTCGAGTCTGGGTTCCGGG
GTTCCGGGTTCCGCCCTCATGAACAGACAGCCCAGGCGGGCGGTTGTTG
GCCCTGGGGGCTGGTTGGAATGCGAGGTCTCGGGAAGTCAGGAGGGA
G
CCTGGCCAGCAGAGGGTTCAGGCCCTGCGGCCGAGGGACCTGGAGACG
G
GCAGGGCATTGGCCGTCGCAGGGCCAGGCCACACCCCCAGGTTTTTGTG
GGGCGAGCCTGGAGATTGCACCACTGTGATTACTATGCTATGGATCTCTG
GGGCCAGGCGTTGAAGTCGTGCTGTCCTCAGGTAAGAACGGCCCTCCAG
GGCCTTAATTTCTGCTCTCGTCTGTGGGCTTTTCTGACTCTGATCCTCG
GGAGGCGTCTGTGCCCCCCCCGGGGATGAGGCCGGCTTGCCAGGAGGGGT
CAGGACCCAGGAGCCTGTGGGAAGTTCTGACGGGGGCTGCAGGCCGGAA
G
GGCCCCACCGGGGGGCGAGCCCCAGGCCGCTGGGCGGCAGGAGACCCGT
G
AGAGTGCGCCTTGAGGAGGGTGTCTGCGGAACCACGAACGCCCGCCGGG
A
AGGGCTTGCTGCAATGCGGTCTTCAGACGGGAGGCGTCTTCTGCCCTCAC
CGTCTTTCAAGCCCTTGTGGGTCTGAAAAGAGCCATGTCGGAGAGAGAAGG
GACAGGCCTGTCCCGACCTGGCCGAGAGCGGGCAGCCCCGGGGGAGAGC
G
GGGCGATCGGCCTGGGCTCTGTGAGGCCAGGTCCAAGGGAGGACGTGTG
G
TCCTCGTGACAGGTGCACTTGCGAAACCTTAGAAGACGGGGTATGTTGGA
AGCGGCTCCTGATGTTAAGAAAAGGGAGACTGTAAGTGAGCAGAGTCC
TCAAGTGTTAAGGTTTTAAAGGTCAAAGTGTTTTAAACCTTTGTGACT
GCAGTTAGCAAGCGTGC GGGGAGTGAATGGGGTGCCAGGGTGGCCGAGA
G
GCAGTACGAGGGCCGTGCCGCTCTAATTCAGGGCTTAGTTTTGCAGAA
TAAAGTCGGCCTGTTTTCTAAAAGCATTGGTGGTGCTGAGCTGGTGGAGG
AGGCCGCGGGCAGCCCTGGCCACCTGCAGCAGGTGGCAGGAAGCAGGTC
G
GCCAAGAGGCTATTTTAGGAAGCCAGAAAACACGGTCGATGAATTTATAG
CTTCTGGTTTCCAGGAGGTGGTTGGGCATGGCTTTGCCGACGCCACAGA
ACCGAAAGTGCCCACTGAGAAAAAACAACCTCTGCTTAATTTGCATTTTT
CTAAAAGAAGAAACAGAGGCTGACGGAAACTGGAAAGTCTCTGTTTTAAC
TACTCGAATTGAGTTTTCGGTCTTAGCTTATCAACTGCTCACTTAGATTC
ATTTTCAAAGTAAACGTTTAAAGAGCCGAGGCATTCCTATCCTCTTCTAAG
GCGTTATTCCTGGAGGCTCATTACCGCCAGCACCTCCGCTGCCTGCAGG
CATTGCTGTACCGTCACCGTGACGGCGCGCACGATTTTCAGTTGGCCCG
CTTCCCCTCGTGATTAGGACAGACGCGGGCACTCTGGCCCAGCCGTCTTG
GCTCAGTATCTGACGGCGTCCGTCTCGGGACGGAGCTCAGGGGAAGAGCG
TGACTCCAGTTGAACGTGATAGTCGGTGC GTTGAAGAGGAGACCAGTCCG
GTGTCGAGTCAGAAAGGGCCCCGGGCCGAGGCCCTGGGCAGGACGGCC

	<p>C GTGCCCTGCATCACGGGCCAGCGTCCTAGAGGCAGGACTCTGGTGGAGA GTGTGAGGGTGCCTGGGGCCCTCCGGAGCTGGGGCCGTGCGGTGCAGGT TGGGCTCTCGGGCGGGTGTGGCTGTTTCTGCGGGATTGGAGGAATTCT TCCAGTGATGGGAGTCGCCAGTGACCGGGCACCAGGCTGGTAAGAGGGA G GCCGCCGTCTGGCCAGAGCAGCTGGGAGGGTTCGGTAAAAGGCTCGCCC GTTTCCCTTAATGAGGACTTTTCTGGAGGGCATTAGTCTAGTCGGGAC CGTTTTCGACTCGGGAAGAGGGATGCGGAGGAGGGCATGTGCCCAGGAG C CGAAGGCGCCGCGGGGAGAAGCCCAGGGCTCTCCTGTCCCCACAGAGGC G ACGCCACTGCCGACAGACAGAGGGCCTTTCCTCTGATGACGGCAAAGG CGCCTCGGCTCTTGGGGGTGCTGGGGGGAGTCGCCCCGAAGCCGCTCA CCCAGAGGCCTGAGGGGTGAGACTGACCGATGCCTCTTGGCCGGGCTGG GGCCGGACCGAGGGGGACTCCGTGGAGGCAGGGCGATGGTGGCTGCGGG A GGGAACCGACCTGGGCCGAGCCCGGCTTGGCGATTCCCGGGCGAGGGCC CTCAGCCGAGGCGAGTGGGTCCGGCGGAACCACCTTTCTGGCCAGCGCC ACAGGGCTCTCGGGACTGTCCGGGCGACGCTGGGCTGCCCGTGGCAGGC CTGGGCTGACCTGGACTTCACCAGACAGAACAGGGCTTTCAGGGCTGAGC TGAGCCAGGTTTAGCGAGGCCAAGTGGGGCTGAACCAGGCTCAACTGGCC TGAGCTGGGTTGAGCTGGGCTGACCTGGGCTGAGCTGAGCTGGGCTGGC TGGGCTGGGCTGGGCTGGGCTGGGCTGGACTGGCTGAGCTGAGCTGGGTT GAGCTGAGCTGAGCTGGCTGGGTTGAGCTGGGCTGGGTTGAGCTGAGCT GGGTTGAGCTGGGTTGAGCTGGGTTGATCTGAGCTGAGCTGGGCTGAGCT GAGCTAGGCTGGGTTGAGCTGGGCTGAGCTGGTTTGAATGGGTTGAGCT GAGCTGAGCTGGGCTGTGCTGGCTGAGCTAGGCTGAGCTAGGCTAGGTTG AGCTGGGCTGGGCTGAGCTGAGCTAGGCTGGGCTGATTTGGGCTGAGCTG AGCTGAGCTAGGCTGCGTTGAGCTGGCTGGGCTGGATTGAGCTGGCTGAG CTGGCTGAGCTGGGCTGAGCTGGCTGGGTTGAGCTGAGCTGGACTGGTT TGAGCTGGGTCGATCTGGGTTGAGCTGTCTGGGTTGAGCTGGGCTGGGT TGAGCTGAGCTGGGTTGAGCTGGGCTCAGCAGAGCTGGGTTGGGCTGAGC TGGGTTGAGCTGAGCTGGGCTGAGCTGGCCTGGGTTGAGCTGGGCTGAGC TGAGCTGGGCTGAGCTGGCCTGTGTTGAGCTGGGCTGGGTTGAGCTGGC TGAGCTGGATTGAGCTGGGTTGAGCTGAGCTGGGCTGGGCTGTGCTGACT GAGCTGGGCTGAGCTAGGCTGGGTTGAGCTGGGCTGAGCTGATCCGAGCT AGGCTGGGCTGGTTTGGGCTGAGCTGAGCTGAGCTAGGCTGGATTGATCT GGCTGAGCTGGGTTGAGCTGAGCTGGGCTGAGCTGGTCTGAGCTGGCCTG GGTCGAGCTGAGCTGGACTGGTTTGGCTGGGTCGATCTGGGCTGAGCTG GCCTGGGTTGAGCTGGGCTGGGTTGAGCTGAGCTGGGTTGAGCTGGGCTG AGCTGAGGGCTGGGTTGAGCTGGGCTGAACTAGCCTAGCTAGGTTGGGCT GAGCTGGGCTGGTTTGGGCTGAGCTGAGCTGAGCTAGGCTGCATTGAGCA GGCTGAGCTGGGCTGAGCAGGCCTGGGTTGAGCTGGGCTAGGTTGGAGCT G AGCTGGGTCGAGCTGAGTTGGGCTGAGCTGGCCTGGGTTGAGGTTAGGCTG AGCTGAGCTGAGCTAGGCTGGGTTGAGCTGGGCTGGGCTGGTTTGGGCTGG GTCAAGCTGGGCCGAGCTGGCCTGGGTTGAGCTGGGCTCGGTTGAGCTGG GCTGAGCTGAGCCGACCTAGGCTGGGATGAGCTGGGCTGATTTGGGCTGA GCTGAGCTGAGCTAGGCTGCATTGAGCAGGCTGAGCTGGGCTGGAGCCT GGCCTGGGTTGAGCTGGGCTGAGCTGGGCTGAGCTGGGCTGAGCTAGGCTGGGTTGAGCT GGCTGGGCTGGTTTGGGCTGGGTTCAAGCTGGGCCGAGCTGGCCTGGGATG AGCTGGGCCGGTTTGGGCTGAGCTGAGCTGAGCTAGGCTGCATTGAGCAG GCTGAGCTGGGCTGAGCTGGCCTGGGTTGAGCTGGGCTGAGCTAAGCTGA GCTGGGCTGGTTTGGGCTGAGCTGGCTGAGCTGGGCTCTGCTGAGCTGGG CTGAGCTGACCAGGGTTGAGCTGGGCTGAGTTAGGCTGGGCTCAGCTAGG</p>
--	--

<p>CTGGGTTGATCTGGCAGGGCTGGTTTGGCGCTGGGTCAAGCTCCCCGGAGA TGGCCTGGGATGAGCTGGGCTGGTTTGGGCTGAGCTGAGCTGAGCTGAGC TAGGCTGCATTGAGCAGGCTGAGCTGGGCTGAGCTGGCCTGGGGTGAGCT GGGCTGGGTGGAGCTGAGCTGGGCTGAACTGGGCTAAGCTGGCTGAGCTG GATCGAGCTGAGCTGGGCTGAGCTGGCCTGGGGTTAGCTGGGCTGAGCTG AGCTGAGCTAGGCTGGGTTGAGCTGGCTGGGCTGGTTTGGCGTGGGTCAA GCTGGGCCGAGCTGGCCTGGGTTGAGCTGGGCTGGGCTGGGCTGAGCTGAGCTAG GCTGGGTTGAGCTGGGCTGGGCTGAGCTGAGCTAGGCTGCATTGAGCTGG CTGGGATGGATTGAGCTGGCTGAGCTGGCTGAGCTGGCTGAGCTGGGCTG AGCTGGCCTGGGTTGAGCTGGGCTGGGTTGAGCTGAGCTGGGCTGAGCTG GGCTCAGCAGAGCTGGGTTGAGCTGAGCTGGGTTGAGCTGGGGTGAGCTG GGCTGAGCAGAGCTGGGTTGAGCTGAGCTGGGTTGAGCTGGGCTCGAGCA GAGCTGGGTTGAGCTGAGCTGGGTTGAGCTGGGCTCAGCAGAGCTGGGTT GAGCTGAGCTGGGTTGAGCTGGGCTGAGCTAGCTGGGCTCAGCTAGGCTG GGTTGAGCTGAGCTGGGCTGAACTGGGCTGAGCTGGGCTGAACTGGGCTG AGCTGGGCTGAGCTGGGCTGAGCAGAGCTGGGCTGAGCAGAGCTGGGTT G GTCTGAGCTGGGTTGAGCTGGGCTGAGCTGGGCTGAGCAGAGTTGGGTTG AGCTGAGCTGGGTTGAGCTGGGCTGAGCTAGGCTGGGTTGAGCTGGGTTG AGTTGGGCTGAGCTGGGCTGGGTTGAGCGGAGCTGGGCTGAACTGGGCTG AGCTGGGCTGAGCGGAACTGGGTTGATCTGAATTGAGCTGGGCTGAGCCG GGCTGAGCCGGGCTGAGCTGGGCTAGGTTGAGCTGGGTTGAGCTGGCTG AGCTGGTCTGAGCTAGGTTGGGTTGAGCTAGGCTGGATTGAGCTGGGCTG AGCTGAGCTGATCTGGCCTCAGCTGGGCTGAGGTAGGCTGAACTGGGCTG TGCTGGGCTGAGCTGAGCTGAGCCAGTTTGGGTTGAGCTGGGTTGAGCTGGGCTG AGCTGGGCTGTGTTGATCTTCTGAACTGGGCTGAGCTGGGCTGAGCTG GCCTAGCTGGATTGAACTGGGTTAAGCTGGGCTGAGCTGGGCTGAGCTG GCTGAGCTAGGCTGAGCTGAGTTGAATTGGGTTAAGCTGGGCTGAGATGG GCTGAGCTGGGCTGAGCTGGGTTGAGCCAGGTCGGACTGGGTTACCTGG GCCACTGGGCTGAGCTGGGCTGGGCTGAGCTGAGTTAAGCTGGGCTGAG TCCGGTCCAGCAGACATGCGCTGGCCAGGCTGGCTTACCTGGACAGTT CGATGAGCTGCCTTGGGATGGTTACCTCAGCTGAGCCAGGTTGGCTCCAG CTGGGCTGAGCTGGTACCCTGGGTTGACCTCGGTGACCAGGTTGTCTGA GTCCGGCCAAGCCGAGGCTGCATCAGACTCGCCAGACCCAAGCCCTGG CCCCGGCTGGCAAGCCAGGGGCGGTGAAGGCTGGGCTGGCAGGACTGTC CCGGAAGGAGGTGCACGTGGAGCCGCCCGGACCCCGACCGGACGAGCT GGAAAGACGCTCTCACTCCCTTTCTTCTTCTGTCCTTCTGGGCTCA GAGAGCCAGTCTGCCCGAATCTTACCCCTCGTCTCTGCGTCAAGCC CCGTCCGATGAGAGCCTGGTGCCCTGGGCTGCCTGGCCCGGACTTCT GCCAGCTCCGTACCTTCTCTGAA</p>
--

猪 κ 轻链

在另一实施方案中，提供了编码有蹄类动物免疫球蛋白的 κ 轻链基因座的新的基因组序列。本发明提供了有蹄类动物 κ 轻链区的首次报导的基因组序列。在一个实施方案中，提供了编码猪 κ 轻链基因座的核酸序列。在另一实施方案中，该核酸序列可以含有 κ 轻链的至少一个连接区、一个恒定区和/或一个增强子区。在另一实施方案中，该核苷酸序列可以包括至少五个连接区、一个恒定区和一个增强子区，例如，如 Seq ID No. 30 中表示。在另一实施方案中，提供了分离的核苷酸序列，其含有猪基因组 κ 轻链的至少一个、至少两个、至少三个、至少四或者五个连接区和连接区的 3' 侧翼序列，例如，如 Seq ID No 12 中表示。在另一实施方案中，提供了猪基因组 κ 轻链的分离的核苷酸序列，其含有第一个连接区的 5' 侧翼序列，如 Seq ID No

25 中表示。在另一实施方案中，提供了分离的核苷酸序列，其含有猪基因组 κ 轻链的恒定区的 3'侧翼序列和任选地，增强子区的 5'部分，如 Seq ID Nos. 15、16 和/或 19 中表示。

在其他实施方案中，提供了如 Seq ID Nos 30、12、25、15、16 或 19 中所述的分离的核苷酸序列。还提供了与 Seq ID Nos 30、12、25、15、16 或 19 至少 80、85、90、95、98 或 99%同源的核酸序列。此外，提供了含有 Seq ID Nos 30、12、25、15、16 或 19 的至少 10、15、17、20、25 或 30 个连续核苷酸的核苷酸序列。此外，提供了含有 Seq ID No.1、4 或 29 的至少 10、15、17、20、25 或 30 个连续核苷酸的核苷酸序列。在其他实施方案中，提供了含有 Seq ID No. 30 的至少 50、100、1,000、2,500、5,000、7,000、8,000、8,500、9,000、10,000 或 15,000 个连续核苷酸的核苷酸序列。还提供了任选在严格条件下与 Seq ID Nos 30、12、25、15、16 或 19 杂交的核苷酸序列，以及其同源核苷酸。

在一个实施方案中，提供了编码 κ 轻链的分离的核苷酸序列，其包括至少五个连接区、一个恒定区和一个增强子区，例如，如 Seq ID No. 30 中表示。在 Seq ID No. 30 中， κ 轻链的编码区例如通过残基 1-549 和 10026-10549 表示，而内含子序列例如通过残基 550-10025 表示， κ 轻链的连接区例如通过残基 5822-7207 (例如，J1:5822-5859, J2:6180-6218, J3:6486-6523, J4:6826-6863, J5:7170-7207)表示，恒定区通过下面的残基表示：10026-10549 (C 外显子) 和 10026-10354 (C 编码), 10524-10529 (聚腺苷酸信号) 和 11160-11264 (SINE 元件)。

Seq ID No 30	<p>GCGTCCGAAGTCAAAAATATCTGCAGCCTTCATGTATTTCATAGAAACAAG GAATGTCTACATTTTCCAAAGTGGGACCAGAATCTTGGGTTCATGTCTAAG GCATGTGCATTTGCACATGGTAGGCAAAGGACTTTGCTTCTCCCAGCACA TCTTTCTGCAGAGATCCATGGAAAACAAGACTCAACTCCAAAGCAGCAAAAG AAGCAGCAAGTTCTCAAGTGATCTCCTCTGACTCCCTCCTCCCAGGCTAA TGAAGCCATGTTGCCCTGGGGGATTAAGGGCAGGTGTCCATTGTGGCAC CCAGCCCAGAACAAAGCAATTTGATCAGGTTCTGAGCACTCCTGAATGTG GACTCTGGAATTTTCTCCTCACCTTGTGGCATATCAGCTTAAGTCAAGTA CAAGTGACAAACAACATAATCCTAAGAAGAGAGGAATCAAGCTGAAGTC A AAGGATCACTGCCTTGGATTCTACTGTGAATGATGACCTGGAAAATATCC TGAACAACAGCTTCAGGGTGATCATCAGAGACAAAAGTTCAGAGCCAGg tagggaaacctcaagccttcaagagcaaaatcatgccattgggttct taacctctgagtgattactatagttactgtggaggcaagcgctca aatagcctgggtaagtgtcaataaaaaagcaaaagtgttcttga aatgttagacctgaggaaaggaatattgataacttcaataatttcaga atgattatagatgtgcactatgctcagtgctctcccccgcactgac aagcagttagaatttactaagaatctaggttgctggggctacatg ggaatcagctcagtgaaagattgtggaatgattactaaatttcta ttccagcataatccaagaacctcagactagttattgacactgctt ttctccataatccatctcatctccgtccatcatggacacttttagaat gacaggtcctggcagagactcacagatgctctgaacatcctttgcttt caagaatgaacgcacacatactaaaggatctcagtgatccacaatag ttttgccacaatggttcttataaaaagtcttcaaacagcaaat gtttataatagttgttcttataataaitgcatgcttcacttctt ttcttt latgaaatfcccaggctagggtcaaatcagaactacacctactggcct acgccacagccacagcaactcaggatcaagccatgctgggacctcac tacagctcatggcaatgccagatccttaaccaatgagcgagccagggga tcgaacctatgctcctatggatactagcagctcattatccgtgagcc ataacaggaaactcccagtttcttttatacaaatggtagccttat gtttctgaaaaccacaaaatgaatgtattcacataattttaaaggta aataattatgataataagacaatagaaagagaaaacgtcattgcctct ttctccagacaacacgcctccttaattgattgaaagaataactactg agcattgtttagtactctttcagcaattagcctgtattcatagccat acataattcaaitaaatgagatcatgatacacacaatacacaatacaca gcctatagggattttacaatcatcttccacatgactacataaaaacct cctaaaaaasaaaaaaccctactcctcctattggctgtttgtgc tccattaaaaagctctataaattaggttatgatgaggttccatttt ctaccttcaagcaacattcaatgcacagcttatatacacatttgagc ctactttctttcttttttttttttttttttttttttttttttttttt ggctttttgttttcaaggctgcatatggaggttcccaggctagctg tctaatcagaactatagctgctggctacgccacatccacagcaataca gatctgagccatgctgcaactacaccacagctcagcaacgggtgat cctlaaacactgagcaaggccaggatcaaacccataactcatggctc ctagtggattgttaaccactgagccatgatggcaactcctgagcctac ttttcaatcatttccaacctaggacactttttaaagtttcaattttct cccccccccctgtttctgaagtgtttgttccactgggtgactc actcccaggatctatctgagatactgagctaaagtgtatgagctotg aattgaatccaactctgccactcaaggatagggttccgatgtg cccaatgggatcagtggtcctctgagtgccaggacgcaggttccatcc ctggcccagcacagtggttaagaatctggcattgctgcaactgagccat agattcaattgtgctcagatctgatcttggcccaggacatgatg cctcagggcaacaaaaagagaaaagggggtgatagcattagtttcta gatttggggataataataaaagtatccatgtacaatgtatggcattt tgaatgtcaacaaaatttcaactattatggagttccatcatggctca</p>
--------------	--

```

gtggaagggaactgattagcatccatgaggacacaggtcacaaccccgac
ctgctcagtgggcattgctgtgagctgtgcatgggttacagacgaagc
tcggatctggcattgctgtggctgtgtgtaagccagcaactacagctct
cattcagccccagcctgggaacctccatagcctaaaaagacaaaaata
aaatttaaaftaaaaataaagaatgttaactattatgattgtactgct
tgcattactgcaaaagaagtcactttctatactcttaatacttagttg
actgtgtgicagtgaaactatttggacacttaattccactctctcta
tcccaacttgacaactctcttccctctctctgggtgagatccactgctg
actttgctcttaaggaacta gaaaagtgtcagtgacaaaatcaaga
aagttaccttaatcticagaattacaatcttaagttctcttftaaagct
actatttcagtggttagtatttccctgtgcccctacaactatcagct
ctgactatgctgatttcaactattatgttgagttttcccttt
ttccctgtcattctgcaaatgttctgtgagcattgtcaagtgaaagata
ctggactggccttccaataataagacaatgaacatcgggagttctcat
tatggtgcagcagaacgaatccaactggaatgtgaggttcaggttc
gatccctgcccctgtcagtggttaaggatccagcattaccgtgagctg
tgggtgtaggttgacagctggctcagatcctgctgtgtgctgtggc
ataggctggcagctctagctctgattcagccgctagcctgggaacctcca
tgcgccccgagtgacgcccataaaaagcaaaaaaaagaagaagaaga
aagacaatgaacatcaaacagctaacaatccagtagggtagaagaatc
tggcaacagataagagcaltaaatgttctaggtccagtgacctgacctc
tgtctctacacagctgcccactgtgagggagaaggtctctcttgag
ttgagctgaaagacattagtttcaaaaactaatgccagtgatgaa
gggtttccaagcagaggaggtttgtaaaaagctgaaatcacagaa
agactctaaagagtttagatgtgtggagcaacatagctgagatggggc
tggagggttaagagggaacaactatagtaagtgaagctggactcacagc
aaagtgaggacctcagcatccttgatgggttaccatggaaacaccaagg
cacacctgatttcaaaacagcaggcaccctgattcagcccaatgtgaca
tgggtggtaacctctagctctacctgtctgtgacaactgacaaccaac
gaagtttaagctgatttctactctgctgacctgttttttttttca
cgtactctatagctcatgccaaaatagattcaaggtaagacggggcc
ttggttgatatacatgtagtctactgttttgagacaataggtggcaa
ggaaagaggttcaaacaggaaaactctcctaattatgtaactgagaaa
agctaaagagctccataatgacactgaatgaagttcatcttgcaaaag
cttccccccccccaggagactataaaaagtgcaatttttaaatgaa
cttattcaaaacagaatagactcacagacataggaaacgaacagatg
gttaccagggtgaaagggtagggatgggataaataaggagcttgggggt
agcagataccccaggttacacaaaataaacaacaggacactatata
agcacagggaactatagcagtagcttaacataacctataatgaaaaga
atgtgaaaaagaatatatgtatgctgtgtgtgtaacgaatcactttgc
tgaacctgaatcaacataacattgaaatcaactacagtttttttt
tttaagtgcagggtttgtgttttttttttttttttttttttttttttt
ttgtttttgtcttttagggccacaccagacatatgggggttcccagg
ctaggggtctlaattagactacagttgcccgttgcaccacagccacagc
aacatcagatccgagccgacttgcgacttacaccacagctcatggcaat
accagatccttaaccactgagcaaggcccaaggatcgtaccggcaacct
catggttcttagtcagatcttctgtgctgctacaatgggaactcaa
gtcagttttttgtaatgtctgtcttcttcttgaattctatctatcc
tacttcccaataaataaataacataaataaataaataacatacattgtaa
ataactacaatttttaaatgcaagggttttttttttttttttttttttt
ttgtcttttgcctttctagggccctccatggcatatggagggtcc
caggctaggggtcaatcggagctgtagccaccggcctacccaagagcca
cagcaacgaggatccgagccgctgtgcaacctacaccacagctcacgg
caacgcccgatcgttaaccactgagcaaggcagggtatgaaacctgcaa
cctcatgttcttagtcagattcgttaactactgagccacaacggaaact
cctaaagtgcagttttaaatgtgctgtcttcttcttgaatttactc

```

```

aacctactcccaataaataaataaacaataaatcatagacatgg
ttgaattctaaaggaggaccatcaggccctagacagaatacgtcatc
ttctagtattttaaacacactaaagaagcaaacatgctctgccagaga
agcccaggccctccacagctgcttcaaggagtaggctcagtagct
gaaccaaggctctgtctctcctcagggaaaagggttttgcagtgaga
cagcagacagctgtcactgtGTGGACGTTTCGGCCAAGGAACCAAGCTGG
AACTCAAAAGCtaagtcaatccaacgttccttctgtgctgtgtct
tacggctctgtggtctgaaatgattcatgtctgactcttgaacca
gactgacattctccaggcacaactaaagcctgcatcaactggaaaac
tgagggcacatttctggcagaaactaagagtcaggcactgggtgaggaa
aaacttgtagaatgatagttcagaaacttactgggaagcaagcccat
gttctgaacagactctgtcaagggtcaggaggggaaccagttttgta
caggagggaggtgagacgaacccctgtTATATGGTTTCGGCGCGGGGA
CCAAGCTGGAGCTCAAAAGCtaagtgcctttccgactgattcttctg
tttctaattgtggtgcttttgcattttcaggttttcacgaa
ttaagtgcaggaccaaacaattgcttccagattaggtaaccagggga
ggggacattgctgcatgggagaccagagggtgctaattttaacgttc
caagccaaataactggggaaggggcttctgtctctgtgagggtaggtt
ttatagaagtggaggttaagggaatcgtatGTTCACTTTTGGCTC
GGGGACCAAAAGTGGAGCCCAAAAGtagtacatttccatcaattttg
tgagattttgctctgtgtcattgtgcaagttttgacattttggt
tgaatgagccattcccaggacccaagaagatgagaccgaaaagtagaaa
agagccaacttttaagctgagcagacagaccgaattgttgattgtgag
gagagtaggggttttagggagaaagggaacagatcgtggtctttctc
tgaattagcctttcctatgggactggctcagaggggtttttgatgagg
gaagtgtctagagccctaaactgtGGTTGTGTTTCGGTAGCGGGACCAAG
CTGGAAATCAAAAGCtaagtgcacttttctactccttttcttctatc
gggtgtgaaattgggacttttcatgttgagtagagtgaggctcagt
tctgaagagagtgaggactatccaaaatcagaggtagagggtcagaac
agagttgtctcatggaagaacaagacctgtagttgatgaggcagcta
aatgagtcagttgacttggatccaaatggccagactcgtctgtaacca
acaatctaagatgtagcagcaaaaagagatttccatgaggggaaag
taaaattgtaattgtGATCACCTTTGGTGAAGGGACATCCGTGGAG
ATTGAAAGCtaagtatttttctactactcctgaaattgtctaaatg
ccagtgtgacttttagagccttaagtgtcagttttgaaaaatgggta
aaacagagcatttcatattatcagtttcaaaagttaaacicagctc
caaaaatgaaattgtagcaaaaagattaattagccaattgaatgat
tcaaaagaaaaaaatagtlagatgaaaaaggaaictttagctcc
aaagagcaaaagcgaatttaatttcttgaacttgcctaaatctgtaaa
tgattttgttcttacaatttaaaaaggtagagaatgtatttcttag
tctgtttctctctctgtctgataaattattatagagataaaaatgaa
aattaataggatgtgctaaaaaacagtaagaagtagaaaaatatgtt
ttaagttaaagttgccacttaattgagaatcagaagcaatgttatttta
aagctaaaaatgagagataaacgtcaacttaaatctgcagagattc
tatacttgacagatactccttttcaaaaatccaattctatggtaga
ctaaattgaaatgacttccctataatggaggaaaagatggactgacc
ccaaaagctcagattaaagaatctgttaagtgaagaaaataaaaga
actgcaatttttaaggcccatgaattgtagaaaataggaaatatttt
aataagtgattcttttttctgttacttgaagggtttttata
ccgccaaggaggccgtggcaccgctcagtgatctgtagaccatggcg
gectttttcgcgattgaatgaacttggcgggtgggtcccaaggctctgg
tggcagcgcaccagccgctaaaagccgctaaaaactccgctaaaggcca
cagcaacccgcgaccgcccgtcaactgtctgacacagtgatacagat
aatgtcgtacaagaggagatagaatatgacgggcacacgctaattgt
gggaaaaagaggagaagcctgatttttttttagattctagagata
aaatccagattatctcttttaaaaaatttctattaggagatta

```

```

taaagaattaaagctatTTTTAAgtggggtgtaattctttcagtagt
ccttgtcaaatggatttaagtaataagaggcttaatccaaatgagagaaa
tagacgcataacccttcaaggcaaaagctacaagagcaaaaattgaaca
cagcagccagccatclagccactcagattttgatcagtttactgagtt
gaagtaaatacatgaaggataaattgctgataaaaaataagatacagg
tgtacacatcttaagticaaaaattaaaggcttcagtaggattata
ttcacgtatacaagatctaaagcagataaaaatgccattaatggaac
ttaatagaaatatttttaattccttcaictgtgacagaaatttct
aatctgggtcttttaatacacccttgaagagtttagtaattgct
attgccatcgtgttaccagctaatttcaaaagtatacttgagaa
agattattttgtttcaaccacctggcagactattttaggccattt
taaaactctttcaactaagtaatttaactgttcaaccatttagg
ccttttaaaaacttttcatgaattcaaaactcgttaaaagttataag
gigtctggcaagaactccttatcaaatgctaatagtttaactgtta
atgcaggatataaaattaagtgatcaaggctgacccaacaggagat
cttcatagcataattccctccttttttctagaatcatalgatttgc
tgccaaggctatttataatactctggaaaaaaaatagtaatgaagggt
aaaaagagaagaaatacaaacattaaagaaatcggtafttactaactg
cttggttaacatgaaggtttttttttaaggtttctctttataaa
aatctgtcccttttctgtgatttcccaagcaaaagattctgattg
tttttaactcttactctcccaagggcctgaatgccacaaggggg
actccaggaggccatctggcagctcaccgtcagaagtgaagccagc
cagttcctcctggcaggtggccaaaattacagttgaccctcctggtct
ggctgaaccttgcccatatggtgacagccatctggccagggccaggtc
tccctctgaaaccttgggagagagggagagtgctggccgatcacag
atcggaagggggctgactcctcaaccggggtgacactctgcagggggg
tctggcccaaacaccccaagcacgccaggaaggaaggcagcttgg
atcactgccagagctaggagggcaccgggaaatgatctgtccaagac
ccgttctgtcttaactccgagggggtcagatgaagtggtttgttc
ttggcctgaagcactgtttccctcaagaagcggggaacacagaggaa
gagagaaaagatgaactgaacaagcatgcaaggcaaaaaggccttag
atggctgcaggaagttagttctctgattggctccttactggctgtc
atcgcccaaacacgcccccagtggaacttccctgttacttaaaaca
ccattctctgtgttctctcagGGGCTGATGCCAAGCCATCCGTCCT
CATCTTCCC GCCATCGAAGGAGCAGTTAGCGACCCCAACTGTCTGTGG
TGTGCTTGATCAATAACTTCTTCCCCAGAGAAATCAGTGTCAAGTGGAAA
GTGGATGGGGTGGTCCAAAGCAGTGGTCATCCGGATAGTGTACAGAGCA
GGACAGCAAGGACAGCACTACAGCCTCAGCAGCACCCCTCTCGCTGCCA
CGTCACAGTACCTAAGTCATAATTTATATTCCTGTGAGGTCACCCACAAG
ACCCCTGGCCTCCCCTCTGGTCAAAAGCTTCAACAGGAACGAGTGTGAGGC
TtagAGGCCCAAGGCCCTGGCCTGCCCCAGCCCCAGCCCCCTCCCC
ACCTCAAGCCTCAGGCCCTTGCCCCAGAGGATCCTTGGCAATCCCCAGC
CCCTCTTCCCTCCTCATCCCCTCCCCTCTTTGGCTTTAACCGTGTAAAT
ACTGGGGGGTGGGGGAATGAATAaataaaGTGAACCTTTGCACCTGTGA
ttctctctctgtctgatttaagggtgtaaatgitttttcccatta
tagttaacttttaaggaaactacatactgagtgtctaaaaactacccat
cacttataaaattcacgcttctcagtttccctcccctctgtctctc
gtaagacaggcctcctgtaaacccataagcacttctttacaccctctc
ctggccggggtaggagactttttgatgctcccttcaagcaagcctcag
aaccattttgaggggacagttcttcaagtcacat*tcctgtgatctaat
gactttagttaccgaaaagccagctctcaaaaagaagggaacggctaga
aaccaagtcatagaatatatagtataaaatatataatccatata
gtaaaaaacaaaaatgataacagcagtaggtaaacaggcaacaggaa
gttgaagtcattctggcacttcaatttaagggaataggatgccttcat
tacatttaatacaatalacacatggagagcttctatctgccaaagacca
tctgaatgccttccacactcactacaagggttaaaagcattcattacaat

```

	<p>gttgatcaggagttcccgttggtcagcaggtaagaacgtgactgg taaccaggaggatgcgggttgggtcccagcctcgcagtggaataagg atccagtgttctgcaagatcacgggctcagatcccgtgtctatggcta tgggttagggctggtgctgcatgcagccctaattgacccctagcctggg aactgccatagccacatgtgagcccttaaacctaaagaaaaaa gaaaagaatacttacaccaattatagataagagagaagctlaagggtg gcaggcccaaggataaagccctacctgcctatctgacacctgatacaaa ttctgtcttctagggttccaactgcatagaacagagggtcaaacatg ctaccctcccaggactcctccttcaaatgacataaatttgggtccca tctctggggcaaaactcaacaatcaatggcatctctagtaaccaagcaag gctcttctcatgaagcaaacctggaagccagatccatgacccaagg aagtaagacaggtgtactggtgaactgtatcctcaattcaatgct tcaattccaactcccagctcccgttaatacaaccccccttgggaagaga gtccttgcatgtagccacgttaaaagagattatacagaaggctagt gggatgcagtgaaacgggactttcatacttctgggtgaaatgtaaa atgctgcaggcactctagaaataattgccaatttttgaagtaaa caaaatgttatgtgcatctggttattatccccagaaataaaa ttatgtccgcacaaaacgtgtacataatcattataacagcctgtac</p>
--	--

<p>Seq ID No.12</p>	<p>caaggaaaccaagctggaactcaaacgtaagtcaatccaacgttcctcc ttggctgtctgtcttaccggctctgtgctctgaaatgaticatgtgc tgactcttgaaaccagactgacatctccaggcnaaaactaaagcctgt catcaaaactgaaactgagggcacatcttctggcagaactaagagtca ggcactgggtgaggaaaaactgtagaatgatagttcaaaaacttact gggaagcaaaagccatgtctgaacagagctctcaagggtcaggagg ggaaaccagttttgtacaggagggaaggtgagacgaaccctgtatata ggtttcggcgcggggaccagctggagctcaaacgtaagtggcttttcc gactgattcttctgtttctaatgttgggtgctttttgcatctt cagtgitttcatcgaattggtgcaaggcaacaacaattgcctccca gattaggtagcaggaggggacatctgcatgggagaccagagggtggc taattttacgtttccaagccaaaataactggggaaggggcttctgt cctgtgagggtagggtttatagaagtggaagtaaggggaaatgctat ggttcaacttttggctcggggaccaaggtggagoccaaaattgagtacat ttccatcaattaattgtgagatttttctctgtgtgctattgtgcaag ttttgacattttggtgaatgagccattcccagggacccaagaagatga gaccgaaaagtagaaaagagccaacttttaagctgagcagacagaccgaa ttgtgagttgtgaggagagtaggggtttgtaggagaaaggggaaacaga tcgctggcttttctctgaattagccttctcatggactggcttcagag ggggtttttagtagggaggtgtctagagccttaactgtggttgggt cggtagcgggaccaagctggaatcaaacgtaagtgcacttttctactcc ttttctttctataggggtgaaattggggacttttcatgtttggagt atgagttgagggtcagttctgaagagagtgaggactatccaaaactgag gagtaagggtcagaacagagttgtctatggaagaacaagacctagtta gttgatgaggcagtaaatgagcagttgactgggataccaaatggccag acttcgtctgtaaccaacaacttaatgagatgtagcagcaaaaagagatt tccatgaggggaaagtaaaatgttaattgtgatacccttgggtga agggacatccgtggagattgaacgtaagtatttttctactactct gaaattgtctaaatgccagttgacttttagaggcttaagtgtagt ttgtgaaaaatgggtaaacagagcatttcatattatcagttcaa aagtaaacctagctccaaaatgaaattgtagcaaaaagattaattta agccaaatgaaatgattcaaaaggaaaaaaatagtgtagatgaaaaag gaattctacagctccaaagagcaaaagcgaatattttcttgaactt</p>
---------------------	--

	<p>tgccaaatcttgaatgattttgtcttacaatttaaaagggtaga gaaatgtafttcttctgtctgtttctctctctgtctgataaattat atgagataaaaatgaaaatataaggatgtctaaaaatcagtaagaag ttgaaaaataatgttatgtaaaagtccacttaattgaatcaga agcaatgttatttttaagctaaaaatgagagataaactgtcaactta aattctgagagattctatacttgacagatactccittcaaaaatc caattctatggtagactaattgaaatgacttctcataatggagg aaaagatggactgacccaaaagctcagatt*agaaaaacctgttaag *gaaagaaaaaaagaactgcaitttttaaggccatgaattgtaga aaaaaggaaatatttaataagtgatcttittttctgtattac tfgatgggtgtttataccgccaaggaggccgtggcaccgtcagtgat ctgtagaccctaggcggcctttttcgcgattgaatgacctggcgtg ggccccaggctctgtggcagcgcaccgcccgtaaaaagccgctaaa actgccgctaaaggccacagcaaccccggaccgccgtcaactgtgt gacacagtatacagataatgtcgtacagaggagaatagaatatgac gggcacacgctaagtggggaaagaggagagcctgttttttttt tagagattcagagataaaatcccagttatatacttttaaaaaaa ttctattaggagattataaagaatttaagctatttttaagggggt gtaattcttcaagtctctgtcaaatggatttaagtaataaggctt aatccaaatgagaaatagacgcataacccttcaaggcaaaagctaca agagcaaaaatgaacacagcagccacatctagccactcagatttga tcagtttactgagttgaaataatcatgaaggataattgctgata aaaaataagatacaggtgtgacacatcttaagttcagaaattaatg gctcagtaggattatattcacgtatacaagatctcaagcagataaaa atgccattaatggaacthaatagaataatatttaattcctcattc tgtacagaaatttctaactgggtcttttaaccctacccttgaaa gagtttagtaattgclatttgccatcgtctgttactccagtaattca aaagtgatactgagaaagattattttgtttgcaaccacctggcagga ctatttagggccatttaaaactcttcaactaagatttttaactg ttcaaaccttagggccttttaaaaacttttcaatgattcaact cgtaaaagtattaaagggtcttgcaagaacttcttcaaatatgct aatagttaacttgtaatgcaagataaaaatgaatgataaggctt gacccaaacaggagatcttcatagcatattccctcctttttctag aattcatatgattgtgccaaggctattttataatctctggaaaaa aaatagtaagaggtaaaagagaagaataatcagaacttaagaatt cggatitttactaactgcttgtaacatgaaggttttatttttaag gttctatctttataaaaactgttccctttctgctgatttccaagc aaaagattctgattgttttttaactcttctcctcccaagggcct gaaatgccacaaggaggacttccaggaggccatctggcagctctcaccg tcagaagtgagccagccagttcctcctggcagggtgccaataacag ttgacctcctctgtctgctgaaccttgcctcatatggtgacagccatc tggccaggccaggctcctctgaagccttgggaggagaggagagt ggctgcccgatcaca gatcggaaaggctgactcctcaaccgggtgc agactctgagggtgggtctggcccaacaccccaagcagcccaggga aggaaaggcagcttggialcactgccagagctaggagggcaccgggaa aatgactgtccaagaccgtctgtcttaaacctcagggggtcaga tgaagtgttttttctgacctgaagcctggttccctgcaagaagc gggaaacacagagggaaggagaagaatgaactgaacaaagcatgcaag gcaaaaaggccttaggaggtcaggaagttagttctgtcattggc tcttactggctgtgatgccacaacacccaccagtgaggagaact tccctgttactaaacacattctgtctgtcttccctcaggggctgat gccaaagccatccgttcaacttccgccatcgaaggagcagttagcgac cccaactgtctgtgtgtgtgtgca</p>
Seq ID No.15	gatgccaaagccatccgtcttcttcttcccgccatcgaaggagcagttagc

	<p>gacccaactgtctctgtggtgcttgatcaataacttctccccagag aaatcagtgcaagtggaaagtggatgggggtgtccaaagcagtggtcat ccgtagtgtcacagagcaggacagcaaggacagcacctacagcctcag cagcacccctctcgtgccacgtcacagtaagtcataattatatt cctgtgaggtcaccacaagaccctggcctcccctctgtgtcacAAGCTTC AACAGGAACGAGTGTGAGGCTTAGAGGCCACAGGCCCTGGCCTGCCCC CAGCCCCAGCCCCCTCCCCACCTCAAGCCTCAGGCCCTTGCCCCAGAGG ATCCTTGGCAATCCCCAGCCCCCTCTCCCTCCTCATCCCCCTCCCTCT TTGGCTTTAACCGTGTAAATACTGGGGGGTGGGGGAATGAATAAAAG TGAACCTTTGCACCTGTGATTTCTCTCTCTCTGTCTGATTTTAAGGTGTT AAATGTTGTTTTCCCCATTATAGTTAATCTTTTAAGGAACTACATACTGA GTTGCTAAAACTACACCATCACTTATAAAATTCAGCCTTCTCAGTTCT CCCCCTCCCTCCTGTCTCCGTAAGACAGGCCTCCGTGAAACCCATAAGC ACTTCTCTTACACCTCTCCTGGGCCGGGTAGGAGACTTTTIGATGTC CCCTTTCAGCAAGCCTCAGAACCATTTGAGGGGGACAGTTCTTACAGT CACAT*TCCTGTCTAATGACTTTAGTTCCGAAAAGCCAGTCTCTCA AAAAGAAGGGAACGGCTAGAAAACCAAGTCATAGAAATATATGTATAAA ATATATATATATCCATATATGTAAAATAACAAAATAATGATAACAGCATA GGTCAACAGGCAACAGGGAATGTTGAAGTCCATTCTGGCACTTCAATTA AGGGAATAGGATGCCTTACATTAAATAACAATACACATGGAGAGC TTCTATCTGCCAAAGACCATCCTGAATGCCTTCCACACTCACTACAAGG TAAAAGCATTCAATACAATGTTGATCGAGGAGTTCCCGTTGTGGCTCAG CAGGTTAAGAACGTGACTGGTATCCAGGAGGATGCGGGTTTGGTCCCCAG CCTCGCTCAGTGGATTAAGGATCCAGTGTGCTGCAAGATCACGGGCTCA GATCCCGTGTCTATGGCTATGGTGTAGGCTGGTAGCTGCATGCAGCCCT AATTTGACCCCTAGCCTGGGAACCTGCCATAAGCCACATGTGAGGCCCTTA AAACCTAAAAGAAAAAAGAAAAGAAAATATCTTACACCCAATTTATAG ATAAGAGAGAAGCTAAGGTGGCAGGCCAGGATCAAAGCCCTACCTGCCT ATCTTGACACCTGAACAAATCTGTCTTCTAGGGTTCCAACACTGCAT AGAACAGAGGGTCAAACATGCTACCCCTCCAGGGACTCCTCCCTCAAAT GACATAAATTTGTTGCCATCTCTGGGGGCAAACTCAACAATCAATGG CATCTAGTACCAAGCAAGGCTCTTCTCATGAAGCAAAACTCTGAAGCC AGATCCATCATGACCCAAGGAAGTAAAGACAGGTGTTACTGGTTGAACTG TATCCTTCAATTCATATGCTCAATTTCCAACCTCCAGTCCCGTAAATA CAACCCCTTTGGGAAGAGAGTCTTGEAGATGTAGCCAGTTAAAAAGA GATTATACAGAAAGCTAGTGAGGATGCAGTGAAACGGGATCTTTCATAC ATTGCTGGTGGAAATGTAATGCTGCAGGCACTCTAGAAAATAATTTGC CAGTTTTTTGAAAAGCTAAACAAAATAGTTTAGTTGCATTCTGGGTATT TATCCCCAGAAATTAATAATATGTCCGCACAAAAACGTGTACATAATC ATTCATAACAGCCTTGTACGAAAAGCTT</p>
Seq ID No.16	<p>GGATCCTTAACCCACTAATCGAGGATCAAAACAGGCATCCTCATGGACAAT ATGTTGGGTTCTTAGCCTGCTGAGACACAACAGGAACTCCCCTGGCACCA CTTTAGAGGCCAGAGAAACAGCACAGATAAAATTCCTGCCCTCATGAAG CTTATAGTCTAGCTGGGAGATATCATAGGCAAGATAAACACATACAAAT ACATCATCTTAGGTAATAATATACTAAGGAGAAAATTACAGGGGAGAA AGAGGACAGGAATTGCTAGGGTAGGATTATAAGTTCAGATAGTTCATCAG GAACACTGTTGCTGAGAAGATAACATTTAGGTAAAGACCGAAGTAGTAAG GAAATGGACCGTGTGCCTAAGTGGGTAAGACCATCTAGGCAGCAGGAAC AGCGATGAAAGCACTGAGGTGGGTGTTCACTGCACAGAGTTGTTCACTGC ACAGAGTTGTGTGGGAGGGTAGGTCTTGCAGGCTCTTATGGTCACAGG AAGAATTGTTTTACTCCCACCGAGATGAAGGTTGGTGGATTTTGAGCAGA AGAATAATTCTGCCTGGTTTATATATAACAGGATTTCCCTGGGTGCTCTG ATGAGAATAATCTGTACAGGGGTGGGATAGGGAGAGATATGGCAATAGGAG CCTTGGCTAGGAGCCCACGACAATAATCCAAGTGAGAGGTGGTGTGCTGCA</p>

TTGAAAGCAGGACTAACAAAGACCTGCTGACAGTGTGGATGTAGAAAAAGA
 TAGAGGAGACGAAGGTGCATCTAGGGTTTTCTGCCTGAGGAATTAGAAAG
 ATAAAGCTAAAGCTTATAGAAGATGCAGCGCTCTGGGGAGAAAGACCAGC
 AGCTCAGTTTTGATCCATCTGGAATTAATTTTGGCATAAAGTATGAGGTA
 TGTGGGTAAACATTATTTGTTTTTTTTTCCATGTAGCTATCCAAGT
 TCCCAGCATCATTTATTTTAAAAGACTTTCCTTCCCCTATTGGATTGTT
 TTGGCACCTTCACTGAAGATCAACTGAGCATAAAATTGGGTCTATTTCTA
 AGCTCTTGATTCCATTCCATGACCTATTTGTTTCATCTTACCCAGTAGA
 CACTGCCTTGATGATTAAGCCCTGTTACCATGTCTGTTTTGGACATGG
 TAAATCTGAGATGCCTATTAGCCAACCAAGCAAGCACGGCCCTTAGAGAG
 CTAGATATGAGAGCCTGGAATTCAGACGAGAAAGGTCAGTCTAGAGACA
 TACATGTAGTCCATCACCATGCGGATGGTGTAAAAGCCATCAGACTGC
 AACAGACTGTGAGAGGGTACCAAGCTAGAGAGCATGGATAGAGAAACCCA
 AGCACTGAGCTGGGAGGTGCTCCTACATTAAGAGATTAGTGAGATGAAGG
 ACTGAGAAAGATTGATCAGAGAAGAAGGA₂AATCAGGAAAAATGGTGCTGTC
 cTGA₂AAATCCAAGGGAAGAGATGTTCCAAGAGGAGA₂A₂ACTGATCAGTT
 GTCAGCTAGCGTCAATTGGGATGAAAATGGACCATTGGACAGAGGGATGT
 AGTGGTCAATGGGTGAATAGATAAGAGCAGCTTCTATAGAAATGGCAGGGG
 CAAAATCTCATCTGATCGGCATGGGTTcTAAAGAAAACGGGAAGAAAAA
 ATTGAGTGATGACCAGTCCCTTCAAGTAGAGAGGTgGAAAAGGGAAGGA
 GGAAAATGAGGCCACGACAACATGAGAGAAATGACAGCATTTTTAAAAAT
 TTTTTATTTATTTtATTTATTTATTTTGGCTTTT₂AGGGCTGCCCTGC
 AAcatatggagggtcccagggttaggggtctaatcagagctatagctgcca
 gectacaccacagccatagcaatgccagatctacatgacctacaccacag
 ctacagc₂acg₂cgatccttaaccactgagtgaggccagagatcaaa
 cccat₂tcctatggatactagtcagggtcattaccactgagccaaaatg
 ggaaATCCTGAGTAATGACAGCATTTTTTAATGTGCCAGGAAGCAAAACT
 TGCCACCCGAAATGTCTCTCAGGCATGTGGATTATTTGAGCTGAAAAC
 GATTAAGGCCAAAAAACACAAGAAGAAATGTGGACCTTCCCCAACAGC
 CTA₂AAAAATTTAGATTGAGGGCCTGTTCCAGAAATAGAGCTATTGCCAGA
 CTTGTCTACAGAGGCTAAGGGCTAGGTGTGGTGGGGAAACCCTCAGAGAT
 CAGAGGGACGTTTATGTACCAAGCATTGACATTTCCATCCATGCGAAT
 GGCTTCTTCCCCTCTGTAGCCCCAAACCACCCCAAAATCTTCTTC
 TGTCTTTAGCTGAAGATGGTGTGAAGGTGATAGTTTCAGCCACTTTGGC
 GAGTTCCTCAGTTGTTCTGGGTCTTCTCCTCGGATCCACATTATTCGACT
 GTGTTTGATTTTCTCCTGTTTATCTGTCTCATTGGCACCCATTTTATTCT
 TAGACCAGCCAAAGAACCTAGAAGAGTGAAGGAAAATTTCTTCCACCT
 GACAAATGCTAAATGAGAATCACCgCAGTAGAGGAAAATGATCTGGTgCT
 GCGGGAGATAGAAGAGAAAAATcGCTGGAGAGATGTCACTGAGTAGGTGAG
 ATGGGAAAGGGGGGGCACAGGTGGAGGTGTTGCCCTCAGCTAGGAAGACA
 GACAGTTcacagaagagaagcgggtgtccgtGGACATCTTGCCTCATGGA
 TGAGGAAACCGAGGCTAAGAAAAGACTGCAAAAGAAAGGTAAGGATTGCAG
 AGAGGTGATCCATGACTAAAATCAGTAACCAACCCCAAAACCACCATG
 TTTTCTCCTAGTCTGGCAGTGGCAGGTACTGTGTAGGTTTTCAATATTA
 TTGGTTTGTAAACAGTACCTATTAGGCCTCCATC₂CCTCCTAATACTAA
 CAAAAGTGTGAGACTGGTCAAGTGA₂AAAAATGGTCTTCTTCTATGCAAT
 CTTTCTCAAGAAGATACATAACTTTTTATTTTATCATgGGCTTGAAGAGC
 AAATGAGAAACA₂CCTCCAACCTATGACACCGTAACAAAGTGTTTATGAT
 CAGTGAAGGGCAAGAAAACAAAACATACAC₂GTA₂AAAGACCCCTCCATAATAT
 TGtGGGCTGGCCCA₂CACAGGCCAGGTTGTA₂AAAGCTTTTTATCTTTGA
 TAGAGGAATGGATAGTAATGTTTCAACCTGGACAGAGAT*CATGTTCACT
 GAATCCTTCCAAAAATTCATGGGTAGTTTGAATATAAGGAAAAATAAGAC
 TTAGGATAAATACTTTgTCCA*GATCCAGAGTTAATgCCAAAATCAGTT
 TTCAGACTCCAGCCAGCTGATCAAGGCCTAAACTTTAAAGACACAGTC
 CCTTAATAACTACTATTACAGTTGCACTTTCAgGGCGAAAGACTCATT
 GAATCCTACAATAGAATGAGTTTAGATATCAAATCTCTCAGTAATAGATG

	<p>AGGAGACTAAATAGCGGGCATGACCTGGTCACTTAAAGACAGAATTGAGA TTCAAGGCTAGTGTCTTTCTACCTGTTTTGTTTCTACAAGATGTAGCAA TGCGCTAATTACAGACCTCTCAGGGAAGGAATTCACAACCCTCAGCAAAA ACCAAAGACAAATCTAAGACAACCTAAGAGTGTTGGTTTAAATTTGAAAAA TAACTCACTAACCAAACGCCCTCTTAGCACCCCAATGTCTTCCACCATC ACAGTGCTCAGGCCTCAACCATGCCCAATCACCCAGCCCCAGACTGGT TATTACCAAGTTTCATGATGACTGGCTGAGAAGATCAAAAAAGCAATGA CATCTTACAGGGGACTACCCGAGGACCAAGATAGCAACTGTCATAGCAA CCGTCACTGCTTTGGTCA</p>
<p>Seq ID No.19</p>	<p>ggatcaaacacgcatcctcaiggaacaatggtgggtcttagcctgctg agacacaacaggaactccccfgcaccacttagagggcagagaacagc acagataaaattccctgcccctcatgaagcttatagtctagctggggagat atcataggcaagataaacacatacaatacatcatcttaggtaataat atactaaggagaaaattacaggggagaaaggacaggaattgctagggt aggattataagttcagatagttcatcaggaacactgtgctgagaagata acatttagtaagaccgaagtagtaaggaatggaccgtgctgcttaagt gggtaagaccattctaggcagcaggaacagcgtgaaagcactgaggtg gtgttcactgcacagagtgttcactgcacagagtgtgtggggagggt aggcttgcagggctcttatgggtcacaggaagaattgtttactcccaccg agatgaagggtggtgattttgagcagaagaataattctgctggtttat atataacaggattccctgggtgctctgatgagaataactgtcaggggt gggatagggagagataggaataggagccttggctaggagcccacagaca ataattccaagtgaaggtggtgctgctattgaaagcaggactaacaagac ctgctgacagtggtgatgagaaaagatagaggagacgaaggtgcatct agggtttctgctgaggaattagaagataaagctaaagcttataagaag atgcagcgtctgaggagaaagaccagcagctcagttttgaccatctgg aattaatttggcataaagtagaggtatgtgggtaacatatttggtt ttttttccatgtagctatccaactgtcccagcatatttatttaa agactttcccttcccctattggattgtttggcacccttcaactgagatca actgagcataaanaattgggtctatttcaagctcttgattocattcatga cctatttgcattcttaccagtagacactgctttagattaaagcc cctgtaaccatgctgtttggacatgtaaatctgagatgcttattagc caaccagcaagcaccggcccttagagagctagatagagcctggaatt cagacgagaaaggtcagcttagacatacatgtatgctcaccatg cggatgggtgtaaaagccatcagactgcaacagactgagaggggtacca agctagagagcattgtagagaaacccaagcactgagctgggggtgctc ctacattaagagattagtagatgaaggactgagaagattgatcagagaa gaaggaaatcaggaanaattggtgctgctctgaaaatcnaagggaagagat gttccaaaaggaganaaactgacaggtgtcagctagcgtcaattgggat gaaaatggaccattggacagagggatgtagtgggtcatgggtgaatgat aagagcagcttctatgaatggcaggggcaaaattctcatctgatcggca tgggttctaaagaaaacgggaagaaaaaattgagtcataaccagtcct tcaagtagagagggtgaaaaagggaagggaanaatgaggccacgacaaca tgaagaaatgacagcatttttaaaatttttattttattttat tattttgctttttagggctgcccctgcaacataggggttcccagggt agggtctaaicagagctatagctgccagcctaccacagccatagcaa tgccagatctacatgacctaccacagctcacagcaacgcccgatcctt aacccaactgagtgaggccagagatcaaccatactctatggatactag tcaggttcattaccactgagcaaaatgggaatcctgagtaatgacagc atttttaattgccaaggaaacaaactgcccacccgaaatgtctctca ggcatgtgattttttagctgaaaacgattaaggcccaaaaaacaa gaagaaatgtggaacttccccacagcctaaaaatttagattgagggc ctgttcccaaaatagagctattgccaagctgtctacagaggctaaagggc taggtgtgtgggaaaccctcagagatcagaggacgtttatgtacca</p>

	<p>gcattgacattccaatccatgcgaatggccttctcccctctgtagcc ccaaaccaccacccccaaatctctctgtcttagctgaagatggtgt fgaaggatgatagttcagccattggcaggttcctcagttgtctgggt ctttcctccTgatccacattatcagctgtgttgatttctcctgttta tctgtctcattggcaccatttcaatcttagaccagcccaagaacctag aagagtgaaggaaaatttctccaccctgacaaatgctaaatgagaatca ccgcagtagaggaaaatgatctggtgctgaggagatagaaagaaaaatc gctggagagatgtcactgagtaggtgagatgggaaagggtgacacaggt ggagggttgcctcagctaggaagacagacagttcacagaagagaagcg ggtgtccgtggacatcttgcctcagtgatgagaaaccgaggctaagaaa gactgcaaaagaaaggtaaggattgcagagaggatccatgactaaaa tcacagtaaccaacccccaaaccatgttttctctagtctggcagctg gcaggtactgtgtaggtttcaatattgtgttgaacagtaacctatt aggcctccatcccctccctactaactaacaagaatgtgagactggtcagt gaaaaatggtctctctctatgaatcttctcaagaagatacataact tttattttatcatagctgaagacaaatgagaaacgcctccaacct atgacaccgtaacaaaatgttatgatcagtaggcaagaacaaaac atacacagtaaaaccctccataatattgtgggtggcccaacacagggcca ggtgtaaaagctttttattctttgataggaatggatagtaatgttctc aacctggacagagatcagttcactgaatcctccaaaaatcatgggta gtttgaattataaggaaaaagacttaggataaactttgtccaagat cccaggttaatgcaaaatcagtttccagctccagcagcctgatcaa gagcctaaactttaaaagacacagctcccttaataactatcagctg cacttcagggcgcaaaagactcattgaatcctacaatagaatgattag atatcaaatctctcagtaatagatgaggagactaaatgaggatgacc tggctactaaagacagaattgagattcaaggctagtgttctctacct gtttgtttctacaagatgtagcaatgctcaattacagacctctcaggg aaggaattcaacccctcagcaaaaaccaagacaaatcaagacaacta agagtgtggttaatttgaaaaataactcactaaccaaacgccctct tagcaccccaatgtcttccaccatcacagtgctcaggcctcaacctgcc ccaatcacc</p>
Seq ID No.25	<p>GCACATGGTAGGCAAAAGGACTTTGCTTCTCCCAGCACATCTTTCTGCAGA GATCCATGGAAACAAGACTCAACTCCAAAGCAGCAAAGAAGCAGCAAGTT CTCAAGTGATCTCCTCTGACTCCCTCCTCCCAGGCTAATGAAGCCATGTT GCCCCGTTGGGGATTAAGGGCAGGTGTCCATTGTGGCACCCAGCCCCAAGA CAAGCAATTTGATCAGGTTCTGAGCACTCCTGAATGTGGACTCTGGAATT TTCTCCTCACCTTTGGGCATATCAGCTTAAGTCAAGTACAAGTGACAAAC AACATAATCCTAAGAAGAGAGGAATCAAGCTGAAGTCAAAGGATCACTGC CTTGGATTCTACTGTGAATGATGACCTGGAAAAATCCTGAAACAACAGCT TCAGGGTGATCATCAGAGACAAAAAGTTCCAGAGCCAGGTAGGGAAACCCCT CAAGCCTTGCAAAGAGCAAAATCATGCCATTGGGTTCTTAACCTGCTGAG TGATTTACTATATGTTACTGTGGGAGGCAAAGCGCTCAAATAGCCTGGGT AAGTATGTCAAATAAAAAGCAAAAGTGGTGTCTTCTGAAATGTTAGACCT GAGGAAGGAATATTGATAACTTACCAATAATTTTCAGAATGATTTATAGA TGTGCACTTAGTCAGTGTCTTCCACCCCGCACCTGACAAGCAGTTTAGA ATTTATTCGAAGATCTAGGTTTGTGGGGCTACATGGGAATCAGCTTC AGTGAAGAGTTTGTGGAATGATTCACTAAATTTTCTATTTCCAGCATAA ATCCAAGAACCTCTCAGACTAGTTTATTGACTGCTTTTCTCCATAAT CCATCTCATCTCGTCCATCATGGACACTTTGTAGAATGACAGGTCTCTGG CAGAGACTCAGATGCTTCTGAAACATCCTTTGCCITCAAAGAATGAAC AGCACACATACTAAGGATCTCAGTGATCCACAAATTAGTTTTTGCCACAA TGGTTCTTATGATAAAAAGTCTTTTCAATTAACAGCAAATGTTTTATAATG TTGTTCTGCTTTATAATAATTGCATGCTTCACTTCTTTCTTTCTTTT TTTTCTTTTTTTGCTTTTTAGTGCCGCAAGTgcagcatagaaatttcc caggctaggggtcaaatcagaactacacctactggcctacgccacagcca cagcaactcaggatctaagccatgicgggtgacctacactacagctcatgg</p>

```

caatgccagatccttaaccaatgagcaggccaggatcgaacccatg
cctcatggatactagtcaggctcattatccgctgagccataacaggaact
cccGAGTTTGCTTTTATCAAAAATTGGTACAGCCTTATTGTTTCTGAAAA
CCACAAAATGAATGTATTACATAATTTTAAAAGGTTAAATAATTTATGA
TATACAAGACAATAGAAAAGAGAAAACGTCATTGCCTCTTTCTCCACGAC
AACACGCCTCCTTAATTGATTTGAAGAAATAACTACTGAGCATGGTTAG
TGTACTTCTTTCAGCAATTAGCCTGTATTCATAGCCATACATATTCAATT
AAAATGAGATCATGATATCACACAATACATACCATACAGCCTATAGGGAT
TTTTACAATCATCTCCACATGACTACATAAAAACTACCTAAAAA
AAAAACCCTACTTCATCCTCCTATTGGCTGCTTTGTGCTCCATTA
CTCTATCATAATTAGGTTATGATGAGGATTTCCATTTTCTACCTTTCAAG
CAACATTTCAATGCACAGTCTTATATACACATTTGAGCCTACTTTCTTT
TTCTTTCTTTTTTTGGTTTTTTTTTTTTTTTTTTTTTTTGGTCTTTTTGTG
TTTTCTAAGgctgcatatggagggtccaggctagctgctaatcagaac
tatagctgctggcctacgccacatccacagcaatacaagatctgagccat
gtctgcaacttacaccacagctcacagcaacgggtgatccttaaccact
gagcaaggccaggatcaaacccatAACTTCATGGCTCCTAGTTGGATT
GTTAACCACTGAGCCATGATGGCAACTCCTGAGCCTACTTTTCTAATCAT
TTCCAACCTAGGACACTTTTTTAAGTTTCATTTTTCTCCCCCACCCT
TGTTTTCTGAAGiGTGTTTGCTTCCACTGGGTGACTTCACiCCAGGATC
TCATCTGCAGGATACTGCAGCTAAGTGTATGAGCTCTGAATTTGAATCCC
AACTCTGCCACTCAAAGGGATAGGAGTTTCCGATGTGGCCCAATGGGATC
AGTGGCATCTCTGCAGTGCCAGGACGAggttccatccctggcccagcac
agtggttaagaatcggCATTGCTGCAGCTGAGGCATAGATTTCAATTG
TGCTCAgATCTGATCCTTGCCCAAGGACTGCATATGCCTCAGGGCAAC
CAAAAAGAGAAAAGGGGGTGATAGCATTAGTTTCTAGATTTGGGGGAT
AATTAATAAAGTGATCCATGTACAATGTATGGCATTGTGAAATGCTCA
ACAAATTTCAACTATTATggagttccatcatggctcagtggaagggaat
ctgattagcatccatgaggacacaggtCCAACCCCGACCTTGCTCAGTGG
GCATTGCTGTGAGCTGTGGCATGGGTTACAGACGAAGCTCGGATCTGGCA
TTGCTGTGGCTGTGGTGTAAGCCAgCAActacagctctcattcagccct
agcctgggaacctccatgccTAAAAGACAAAAAATAAAATTTAAATTA
AAAATAAAGAAATGTTAACTATTATGATTGgTACTGCTTGCATTACTGCA
AAGAAAGTCACITTTCTATACTCTTTAATATCTTAGTTGACTGTGTGCTCA
GTGAACATTTTGGACACTTAATTTCCACTCTCTTCTATCTCCAACCTGA
CAACTCTCTTCTCTCTCTGOTGAGATCCACTGCTGACTTTGCTCTTT
AAGGCAACTAGAAAAGTGCTCAGTGACAAAATCAAAGAAAAGTTACCTTAA
TCTTCAGAATTACAATCTTAAGTTCTCTTGTAAGCTTACTATTTAGTG
GTTAGTATTATTCCTTGGTCCCTTACAACCTATCAGCTCTGATCTATTGC
TGATTTTCAACTATTTATTGTTGGAGTTTTTTCCTTTTTTCCCTGTTTCA
TCTGCAAATGTTTGTGAGCATTTGTCAAGTGAAGATACTGGACTGGGCC
TTCCAAATATAAGACAATGAAACATCGGGAGTTCTCATTATGGTGACGA
GAaacgaatccaactaggaatgaggttcagggtcagccctgocct
tgctcagtggttaaggatccagcattaccgtgagctggtgtaggtg
cagacgtggctcagatccctgctgctggtgctggtcagataggctggcag
ctctagctctgaticgaccgctagcctgggaacctccatGCGCCCCGAGT
GCAGCCCTTAAAAAGCAAAAAAAAAAGAAAAGAAAAGACAATGAAA
CATCAAACAGCTAACAAATCCAGTAGGGTAGAAAAGAAATCTGGCAACAGATA
AGAGCGATTAATGTTCTAGTCCAGTGACCTTGCTCTGTGCTCTACAC
AGTCGTGCCACTTGCTGAGGGAGAAGGTCTCTCTTGAGTTGAGTCTGAA
AGACATTAGTTGTTCAAAAATAATGCCAGTGAGTGAAGGTGTTTCCAAG
CAGAGGGAGAGTTTGGTAAAAAGCTGGAAGTCACAGAAAAGACTCTAAAGA
GTTTAGGATGGTGGGAGCAACATACGCTGAGATGGGGCTGGAAGGTTAAG
AGGGAACAACACTATAGTAAGTGAAGCTGGACTCACAGCAAAGTGAGGACC
TCAGCATCCTTGATGGGGTTACCATGGAACACCAAGGCACACCTTGATT
TCCAAAACAGCAGGCACCTGATTACGCCAATGTGACATGGTGGGTACCC

```

```

CTCTAGCTCTACCTGTTCTGTGACAACTGACAAACCAACGAAGTTAAGTCT
GGATTTTCTACTCTGCTGATCCTTGTITTTGTTTCACACGTCATCTATAG
CTTCATGCCAAAATAGAGTTCAAGGTAAGACGCGGGCCTTGGTTTGATAT
ACATGTAGTCTATCTTGTITGAGACAATATGGTGGCAAGGAAGAGGTTCA
AACAGGAAAATACTCTCTAATTATGATTAAGTGAAGAAAGCTAAAGAGTC
CCATAATGACACTGAATGAAGTTCATCATTTGCAAAAGCCTTCCCCCCC
CCCAGGAGACTATAAAAAAGTGCAATTTTTTAAATGAACTTATTTACAAA
ACAGAAAATAGACTCACAGACATAGGAAAACGAACAGATGGTTACCAAGGT
GAAAGGGAGTAGGAGGGATAAATAAGGAGTCTGGGGTTAGCAGATACACC
CCAGTGTACACAAAATAAACCAACAGGGACCTACTATATAGCACAGGGAAC
TATATGCAGTAGCTTACAATAACCTATAATGGAAAAGAATGTAAAAAGA
ATATATGTATGCGTGTGTGTGTAAGTGAATCACTTTGCTGTAACCTGAAT
CTAACATAACATTGTAATCAACTACAGTTTTTTTTTTTTTAAAGTGCAG
GGTTTTGGTGTITTTTTTTTTCATTTTTGTTTTGTTTTGTTTTTGC
TTTTTAGGGCCACCCCAGACATATGGGGGTCCCAGGctAGGGGTcTAA
TTAGAGcTACAGtGCGGGCTTGCaccacagccacagcaacatcagatcc
gagccgcacttgcacttacaccagctcatggcaataccagatccta
accactgagcaaggcccagggatcgtaccgcaacctcatggttcctag
tcagatcattTCTGCTGCGCTACAATGGGAACTCCAAGTGCAGTTTTTT
GTAATGTGCTGTCTTTCTTTGTAATTCATATTCCTACTTCCAATA
AATAAATAAATACATAAATAATAAACATACCATTGTAATCAACTACAAT
TTTTTTAAATGCAGGGTTTTGTTTTTGTTTTTTGTCTTTTTG
CCTTTCTAgggccgctccatggcatatggagttcccaggctaggggt
cgaatcggagctgtagccaccggcctacgccagagccacagcaacgcggg
atccgagcccgcttgcacctacaccagctcaogccaogccggatc
gtaaccactgagcaaggcagggatcgaacctgcaacctcatggttcc
tagtcagattcgttaactactgagccacaacggaacTCCATAAAGTGCAG
TTTTAAATGTGCTTGTCTTTCTTTGTAATTTACACTCAACCTACTTCCC
AATAAATAAATAAATAAACAATAAATCATAGACATGGTTGAATTCTAAA
GGAAGGGACCATCAGGCCTTAGACAGAAATACGTCATCTTCTAGTATTTT
AAAAACAATAAAGAACAAACATGCTCTGCCAGAGAAGCCCAGGGCCT
CCACAGCTGCTTGCAGGGAGTTAGGCTTCAGTAGCTGACCCAAGGCTC
TGTTCTCTCAGGGAAGGGTTTTGTTTCAGTGAGACAGCAGACAGCT
GTCACTGTGgtggacgttcggccaaggaaccaagctggaactcaaacGTA
AGTCAATCCAACGTTCCCTTCCCTGGCTGTCTGTGCTTAAGGTCCTGTG
GGCTCTGAAATGATTCATGTGCTGACTCTCTGAAACCAGACTGACATTCT
CCAGGGCAAAACTAAAGCCTGTCTCAAAACcGAAAACTGAGGGCACATT
TTCTGGCAGAACTAAGAGTCAGGCACTGGGTGAGGAAAACTTGTGAGA
ATGATAGTTTCAGAACTTACTGGGAAGCAAAGCCCATGTTCTGAACAGA
GCTCTGCTCAAGGGTCAGGAGGGGAACCAGTTTTTGTACAGGAGGGAAGT
TGAGACGAACCCCTGTGTAatgtgttcggcggggaccaagctggagc
tcaaacGTAAGTGGCTTTTTCCGACTGATTCCTTGTCTGTTTCTAATTGTT
GGTTGGCTTTTTGTCCATTTTTTCAGTGTITTCATCGAATTAGTTGTCAGG
GACCAAACAAATTGCTTCCCAGATTAGGTACCAGGGAGGGGACATTGCT
GCATGGGAGACCAGAGGGTGGCTAATTTTTAACGTTTCCAAGCCAAAATA
ACTGGGGAAGGGGCTTGTCTGTCTGTGAGGGTAGGTTTTATAGAAGTG
GAAGTTAAGGGGAAATCGCTATGGTtcacttttggctcggggaccaaggt
ggagcccaaatgaGTACATTTTCCATCAATTATTTGTGAGATTTTTGT
CCTGTTGTGCATTTGTGCAAGTTTTTGACATTTTGGTTGAAATGAGCCAT
TCCCAGGGACCCAAAAGGATGAGACCGAAAAGTAGAAAAGAGCCAACCTT
TAAGCTGAGCAGACAGACCGAATTGTTGAGTTTGTGAGGAGAGTAGGGTT
TGTAGGGAGAAAGGGGAACAGATCGCTGGCTTTTTCTCTGAATTAGCCTT
TCTCATGGGACTGGCTCAGAGGGGTTTTTGTAGGGAAGTGTCTAG
AGCCTTAACTGTGGgttggttcgtagcgggaccaagctggaatcaaa
CGTAAGTGCATTTTCTACTCC

```

猪入轻链

在另一实施方案中，提供了编码有蹄类动物免疫球蛋白的入轻链基因座的新的基因组序列。本发明提供了有蹄类动物入轻链区的首次报导的基因组序列。在一个实施方案中，猪入轻链核苷酸包括J到C单位的多联体。在特定实施方案中，提供了如 Seq ID No. 28 中描绘

的分离的猪 λ 核苷酸序列。

在一个实施方案中，提供了包括猪 λ 轻链基因组序列的第一个 λ J/C区的5'侧翼序列的核苷酸序列，如Seq ID No 32表示。此外，提供了包括猪 λ 轻链基因组序列的J/C簇区的3'侧翼序列，如 λ J/C下游约200个碱基对的核苷酸序列，如Seq ID No 33表示。备选地，提供了包括猪 λ 轻链基因组序列的J/C簇区的3'侧翼序列的核苷酸序列，如J/C簇下游约11.8 kb，接近增强子（如Seq JD No. 34表示）， λ 的下游约12Kb，包括增强子区（如Seq ID No. 35表示）， λ 的下游约17.6Kb（如Seq ID No. 36表示）， λ 的下游约19.1Kb（如Seq ID No. 37表示）， λ 的下游约21.3Kb（如Seq ID No. 38表示），和/或 λ 的下游约27Kb（如Seq ID No. 39表示）。

在进一步的实施方案中，提供了如Seq ED Nos 28、31、32、33、34、35、36、37、38或39中所述的分离的核苷酸序列。还提供了与Seq ID Nos 28、31、32、33、34、35、36、37、38或39至少80、85、90、95、98或99%同源的核酸序列。此外，提供了含有Seq ID Nos 28、31、32、33、34、35、36、37、38或39的至少10、15、17、20、25、30、40、50、75、100、150、200、250、500或1,000个连续核苷酸的核苷酸序列。还提供了任选在严格条件下与Seq ID Nos 28、31、32、33、34、35、36、37、38或39杂交的核苷酸序列，以及其同源核苷酸。

Seq ID No.28	<p> CCTTCCTCCTGCACCTGTCAACTCCCAATAAACCGTCCTCCTTGTCAATC AGAAATCATGCTCTCCGCTCACTTGTGTCTACCCATTTTCGGGCTTGCAT GGGGTCATCCTCGAAGGTGGAGAGAGTCCCCCTTGGCCTTGGGGAAGTCG AGGGGGCGGGGGGAGGCCTGAGGCATGTGCCAGCGAGGGGGGTCACTC CACGCCCCTGAGGACCTTCTAGAACCAGGGGCGTGGGGCCACCGCTGAG TGGAAGGCTGTCCAATTTCCCCCGGGCCCCCAGGCTCCCTCCTCCGTGT GGACCTTGTCCAACCTTGACTGGCCCAGCCACTCATGCATTGTTCCCCG AAACCCAGGACGATAGCTCAGCACGCGACAGTGTCCCCCTCTGAGGGCC TCTGTCCATTTACAGGACGACCCGCATGTACAGCGTGACCACTCTGCTCAC GCCCCTCACCACGTCTAGAGCCCCACCCCCAGCCCCATCCTTAGGGGC ACAGCCAGcTCCGACCGCCCCGGGGACACCACCTCTGCCCTTcCCCAG GCCCTCCCTGTCAACAGCACACCACAGGGCCCTCCGTCCCCAGCTGCTC CCTCATCCCTCGGTCCCCTCAGGTAGCCTTCCACCCGCGTGTGTCCCGAG GTCCAGATGCAGCAAGGCCCTGGGACAACGCCAGATCTCTGCTCTcCC CGACCCCTCAGAAGCCAGCCCAGCCTGGCCCCACCACCTGCCTAACg TCCAAGTGTCCATAGGCCTCGGGACCTCCAAGTCCAGGTTCTGCCTCTGG GATTCGGCCATGGGTCTGCCTGGGAAATGATGCACTTGGAGGAGCTCAGC ATGGGATGCGGGACCTTGTCTTAGGGCGTcCCTCAGGATCCCAAGCTG CCCTGTGAGACA CACACAAACACCGCATGCACGCACGCCGGCACACACGCTATTGCAGAGAT GCCACGGTAGCTGTGCCTCGAGGCCGAGTGGAGTGTCTAGAACTCTCGGG GGTCCCCTCTGCAGACGACACTGCTCCATCCCCCGGTGCCCTGAAGGGC TCCTCACTCTCCATCAGGATCTCTCCAAGCTGCTGACCTGGAGAGGAAG GGCCTGGGACAGGGCGGGGACACTCAGACCTCCCTGCTGCCCTCCTCTG CCTGGGCTTGGACGGCTCCCCCTTCCCAGGGTGAAGGTGCAGGTGGGG AGAGGGCACCCCCCTCAGCTCCAGACCCAGACCAGCCCCGTTGGGAGG GGCAGCCTGTGAGCCTCCAGCCAGATGCAGGTGGCCTGGGGTGGGGGTG GAGGGGGCGGGAGGTTTATGTTTGGGGCTGTATCACTGTGTAATATTTTC GGCGGTGGGACCCATCTGACCGTCTCGGTGAGTCTCCCCTTTTCTCTCC TCCTTGGGGATCCGAGTGAATCTGGGTGATCTTCTCTCCGTTCTCCTC CGACTGGGGCTGAGGTCTGAACCTCGGTGGGGTCCGAAGAGGAGGCCCT AGGCCAGGCTCCTCAGCCCCCAGCCCCGACcGGCCTCTTGACACAGGG TCCAGCTAAGGGCAGACATGGAGGCTGCTAGTCCAGGGCCAGGCTGTGAG ACCAAGGGCGCTGCCAAAGGAACCCTTGCCCCAGGGACCTGGGACAA AGCTCCTCACTCAGAGCCTGCAGCCCTGGGGTCTGAGGACAAGGAGGGAC TGAGGACTGGCGTGGGGAGTTCAGGGCGGGGACACCAGGTCCAGGGAGGT GACAAAGGCGCTGGGAGGGGGCGGACGGTGCAGGGGACTCCTCCTGGGGC CTGTGGGCTCGGGTCTTGTGAGGACCTGAGGGACTGAGGGGCCCTG GGCCTAGGGACTTGCagTgAGGGAGGCAGGGAGTGTCCCTTGAACGCTG GCCTCCGCGGGCTGGGTCCCCCTCGTGTCTCCAGCC*GGGAGGACACCCC AGAGCAAGCGCCCCAGGTGGGCGGGGAGGTTCTCTCACAGGGGCAGCTG ACAGATAGAGGCCCGCCAGGCAGATGCTTGTATCCTGGCagTTATACTG GGTTC**GCACAACCTTCCCTGAACAAGGGGCCCTCOGAACAGACACAGA CGCAACCCAGTCCAGCCaggCTCAGCACAgAAAATGCACTGACACCCAAA ACCCTCATCTggggGCTGGCCGGcAtCCCCCCCCAGGACCCAAAGGCCCC TGCCCCCTGGCAGCCCTGGACACGGTCTCTGTGGCCGGTGGGGTcGGG CTGTGGTQACGGTGGCATCGGGGAGCCTGTGCCCCCTCCCTGAAAGGGCG GAGAGGCTCAAGAGGGGACAGAAATGTCTCCCTAGGAAGACCTCGGAC GGGGCGGGGGGGTGTCTCCGACAGACAGATGCCCGGACCGACAGACC TGCCGAGGGAAGAGGGCACCTCGGTGGGTTAGGCTCCAGGCAGCACGAG GGAGCGAGGCTGGGAGGGTGGAGACATGGGAGCCTGAGGAGGAGCTGGAG ACTTCAGCAGGCCCCCCAGCTCCGGGCTTGGGGCTCTGAGATGCTCGGACG CAAGGTGAGTGACCCCACTGTGGCTGACCTGACCTCAgGGgGACAAGGC </p>
--------------	--

TCAGCCTGGGACTCTGTGTCCCCATCGCCTG₆ACAGGGGATTCCCCTGAT
GGACTGAGCCAAACGACCTCCCCTCTCCCCGACCCCCAGGTGAGCCC
AA_gGCCaCTCCCACGGTCAACCTCTTCCCGCCCTCCTCTGAGGAGCTCGG
CACCAACAAGGCCAACCTGGTGTGTCTAATAAGTGACTTCTACCCGGGCG
CCGTGACGGTGACCTGGAAGGCAGGCGCACCAACCGTCACCCAGGGCGTG
GAGACCACCAAGCCCTCGAAAACAGCAACAACAAGTACGCGGCCAGCAG
CTACCTGGCCCTGTCCGCCAGTGAAGTCTTCCAGCGGCTTCACT
GCCAGGTCAACACGAGGGGACCATTGTGGAGAAGACAGTGACGCCCTCC
GAGTGGCCTAGGTCCCCTGGGCCCCACCCCTCAGGGGCTGGAGCCACAG
GACCCCGCGAGGGTCTCCCCGCGACCCCTGGTCCAGCCCAGCCCTTCTC
CTGCACCTGTCAACTCCCAATAAACCGTCCCTCTGTCAATTCAGAAATCA
TGCTCTCCGCTCACTTGTGTCTACCAATTTTCGGGCTTGCATGGGGTCAT
CCTCGAAGGTGGAGAGAGTCCCCCTTGGCCTTGGG_gAAATCGAGGGGGGC
GGGGGAGGGCTGAGGCATGTGCCAGCGAGGGGGGTCACTCCACGCCCC
TGAGGACCTTCTAGAACCAGGGGCGTGGGGCCACCCGACAGTGAAGGC
TGTCCACTTTTCCCCGGGCCCCAGGCTCCCTCCTCCGTGTGGACCTTG
TCCACCTCTGACTGGCCAGCCACTCATGCATTGTTCCCCGAAACCCCA
GGACGATAGCTCAGCACGCGACAGTGTCCCCCTTGAGGGCCTCTGTCCA
TTTCAGGACGACCCGCATGTACAGCGTGACCACTCTGCTCACGCCACTC
ACCACGTCTAGAGCCCCACCCCAAGCCCCATCCTTAGGGGCACAGCCAG
CTCCGACCGCCCGGGGACACCAACCTCTGCCCTTCCCCAGGGCCTCCC
TGTCACACGCACCACAGGGCCCTCCGTCCCGAGACCCTGCTCCCTCATCC
CTGGTCCCCTCAGGTAGCCTTCCACCCCGGTGTGTCCCGAGGTCCAGA
TGCAGCAAGGGCCCTGGGACAACGCCAGATCTGTCTTCCCCGACCCCTC
AGAAGCCAGCCACGCTGGCCACCACTGCCTAACGTCCAAGTGTCTC
CATAGGCTCGGGAcCTC_cA_gTCCAGGTTCTGCCTCTGGGATTCCGCCAT
GGGTCTGCCTGGAATGATGCACCTGGAG_gAgCTCAG_cATGGGATG_cGGAA
CTTGTCTAG_cGCTCCTCAGATCCAcAG_cTGCCTGtGAgAcacacacacac
acacacacacaccAAAcaCG_cATGCACGCACGCCGGCACACACGCTATTA
CAGAGATGGCCACGGTAGCTGTGCCTCGAGGCCGAGTGGAGTGTCTAGAA
CTCTCGGGGGTCCCCTCTGCAGACGACACTGCTCCATCCCCCCCCGTGCC
TGAAGGGCTCCTCACTCTCCATCAGGATCTCTCCAAGCTGCACTGG
AGAGGAAGGGGCTGGGACAGCGGGGACACTCAGACCTCCCTGCTGCC
CTCCTCTGCCTGGGCTTGGACGGCTCCCCCTTCCCAGGGTGAAGGTGC
AGGTGGGAGAGGGCACCCCCCTCACCTCCAGACCCAGACCAGCCCC
GTGGCAGGGGACGCTGTGAGCCTCCAGCCAGATGCAGGTGGCCTGGGGT
GGGGGGTGGAGGGGGCGGAGGTTATGTTGAGGCTGTATTCATCTGTG
TAATA_TTCGGCGGTGGGACCCATCTGACCGTCCCTCGGTGAGTCTCCCC
Tttcttctccttgggatccgagt_{gaa}A_{Tc}TGGGTCTGATCTTCTCTC
CGTTCTCCTCCGACTGGGGCTGAGGTCTGAACCTCGGT_gGGTCCGAAGA
GGAGGCCCTAGGCC*GGCT_cTCAGCCCTCCAGCCCGACCCGCCCTCT
TGACACAGGGTCCAGCTAAGGGCAGACAT***GGCTGCTAGTCCAGGGCC
AGGCT_tTGAGACCCAAGGGCGCTGCCAAGGAACCCCTGCCCCAGGGACC
CTGGGAGCAAAGCTCCTCACTCAGAGCCTGCAGCCCTGG_gGTCTGAGGAC
AAGGAGGGACTGAGGACTGGGCGTGGGGAGTTCAGGC_gGGGACACCCGGT
CCAGGGAGGTGACAAAGGCGCTGGGAGGGGGCGGACGGTGCCGGAGACTC
CTCCTGGGCCCTGTGGCTCGTGGTCTTGTGAGGACCCCTGAGGG*CTGA
GGGGCCCCTGGGCTAGGGACTTGCAGTGAGGGAGGCAGGGAGTGTCCCT
TGAGAACGTGGCCTCCGCGGGCTGGGTCCCCCTCGTGTCTCCAGCAGGGA
GGACACCCAGAGCAAGCGCCCCAGGTGGGCGGGAGGGTCTCCTCACAG
GGCAGCTGACAGATAGAC*G_gcccCGCCAGACAGATGCTTGTATCCTGG
TCag***TACTGGGTTCCG_cACTTCCCTGAACAGGGGCCCTCCGAACAGA
CACAGACGCAGACCaggCTCAGCACAgAAAATGCACTGACACCCAAAACC
CTCATCTG_{gg}GGCCTGGCCGGCATCCCCCCCCAGGACCCAAGGCCCTGC
CCCCGGCAGCCCTGGACACGGTCTCTGTGGGCGGTGGGGT_gGGGCTG
TGGTACGGTGGCATCGGGGAGCCTGTGCCCCCTCCCTGAAAGGGCGGAG

<p>AGGCTCAAGAGGGGACAGAAATGTCCTCCCTAGGAAGACCTCGGACGGG GGCGGGGGGGTGGTCTCCGACAGACAGATGCCCGGGACCGACAGACCTGC CGAGGGAAAGAGGGCACCTCGGTGGGTTAGGCTCCAGGCAGCAGAGGGA GCGAGGCTGGGAGGGTGAGGACATGGGAGCCTGAGGAGGAGCTGGAGACT TCAGCAGGCCCCAGCTCCGGGCTTCGGGCTCTGAGATGCTCGGACGCAA GGTGAGTGACCCACCTGTGGCTGACCTGACCTGACCCAGGGGGACAAG GCTCAGCCTGGGACTCTgTGTCCCATCGCTGACAGGGGATTCCCCTG ATGGACACTGAGCCAACGACCTCCCCTCTCTCCCCGACCCCAAGGTCAGC CCAAGGCCACTCCACGGTCAACCTCTTCCCGCCCTCCTCTGAGGAGCTC GGCACCAACAAGGCCACCTGGTGTGTCTA</p>

Seq ID No.32	<p>GCCACGCCACTCCATCATGCGGGGAGGGATGGGCAGACCCTCCAGAAA GAAGTCCCTGGGGTGCAGGTTAACAGCTTCCCAGACACAGCCAGTACT AGAGTGAGGTGAATAAGACATCCTCCTTGCTTGTGAAATTTAGGAAGTC CCCCAAACATCAGTCATTAAGATAAATAATATTGAATGCACTTTTTTTTT TTATTTTTTTTTTTTGTCTTTTAGGGCTAATCTGCAGCAtatggaagt tccagggtacaagtgaaccagagctgcagctgccagcctacatcacag ccacagcaaccagatccgagccacatctgtactaacactgcagtca cagcaacgccagatccttaaccattgagtgagccaggatcaaaccca catcctcatggatactgtctgttcgtaaacactgagccaCAAGGGGA ACTCCTGAATGCAATATTTTGAATAATTGAAATTAATCTGTCACTCTT CACTTAAGAGTCCCCTTAGATTGGGGAAAATTTAAATATCTGTCTATCTTA GTGCATCTTTGCTCATATGATGTGAATAAAATCCCCAAATCCATATGAAT GAAGCATCAAAATGTACATGAAGTCAGCCTGACCCTGCACTGCCCTCACT TGCTCATGTACCCCCACCTCAAAGGAAGATGCAGAAAAGGATCCAGCC CCTACACCGCCACCTGCCCCACCCTGGAGCCCCTCAGGTCTCCACCT CCTTTTCTGAGCTTCAGTCTTCTGTGGCATTGCCTACCTCTACAGCTGC CCCCTACTAGGCCCTCCCCTGGGGCTGAGCTCCAGGCCTGGACTGGGA AAGTTAGAGGTTAAAGCATGGAAAATCCCAAAGCCACCAGTTCCAGGCT GCCCCCCACCCACCGCCACGTCCAAAAGGGGCATCTTCCAGATCTCT GGCTGGTATTGGTAGGACCCAGGACATAGTCTTTATAACCAATTCTGTGT GTGTCTTAGGAAAGAAactcctctctgtgcttcagttcctcatca taaaAGGAGCAGGCCAGGTTGGAGGGTCTGTGACGTCTGCTGAAGCAGCA GGATTCTCTCCTTTTGTGGAGGAGAACTGATCCTTACCCCCAGGAT CAACAGAGAAGCCAAGGTCTTCAGCCTTCTGGGGACCCCTCAGAGGGAA CTCAGGGCCACAGAGCCAGACCCTGATGCCAGAACCTTTGTATATGCC AGACGGAGACTTCATCCCCCTCCTCCTCAGACCCTCCAGGCCCAACAGT GAGATGCTGAAGATATTAAGAGAAGGGCAAGTCAGeTTAAGTTTGGGGGT AGAGGGGAACAGGGAGTGAGGAGATCTGGCCTGAGAGATAGGAGCCCTGG TGGCCACAGGAGGACTCTTTGGGTCTGTGCGATGGACACAGGGCGGGCC GGGGGCATGTTGGAGCCCGGCTGGTCTTACCAGAGGCAGGGGGCACCT CTGACACGGGAGCAGGGCATGTTCCATACATGACACACCCCTCTGCTCCA GGGCAGGTGGGTGGCGGCACAGAGGAGCCAGGGACTCTGAGCAAGGGGT CACCAGTGGGGCAGTTGGATCCAGACTTCTCTGGGCCAGCGAGGTCTAG CCCTCAGCCGTTCTCTGTCCAGGAGGGGGGTGGGGCAGGCTGGCGGGC AGAGCTCATCCCTCAAGGGTCCAGGGTCTGCCAGACCCAGATTTCCG ACCGCAGCCACCACAAGAGGATGTGGTCTGCTGTGGCAGCTGCCAAGACC TTGCAGCAGGTGCAGGGTGGGGGGTGGGGGCACCTGGGGCAGCTGGGG TCACTGAGTTCAGGGAAAACCCCTTTTTTCCCCTAAACCTGGGGCCATCC CTAGGGGAAAACCAACTTCTGAGCCCTGGGCAGTGGCTGCTGGGAGGGA AGAGCTTCATCCTGGACCCTGGGGGGGAACCCAGCTCCAAAGGTGCAAGG GGCCAGGTCCAAGGCTAGAGTGGGCCAAGCACCGCAATGGCCAGGGAGT GGGGGAGGTGGAGCTGGACTGGATCAGGGCCTCCTTGGGACTCCCTACAC</p>
--------------	--

CCTGTGTGACATGTTAGGGTACCCACACCCCATCACCAGTCAGGGCCTGG
 CCCATCTCCAGGGCCAGGGATGTGCATGTAAGTGTGTGTGAGTGTGTGTG
 TGTGGTGTAGTACACCCCTTGGCATCCGGTTCGAGGCCTTGGGTTCCCTC
 CAAAGTTGCTCTCTGAATTAGGTCAAACCTGTGAGGTCTGTATCGCCATCA
 TCAACTTCGTTCTCCACCTCCCATCATTATCAAGAGCTGGGGAGGGTC
 TGGGATTTCTTCCACCCACAAGCCAAAAGATAAGCCTGCTGGTGATGGC
 AGAAGACACAGGATCCTGGGTCAGAGACAAAAGGCCAGTGTGTACAGCGA
 GAGAGGCAGCCGACTATCAGCTGTACAGAGAGGCCCTTAGTCCGCTGAA
 CTCAGGCCCCAGTACTCCTGTTCCACTGGGCACTGGCCCCCTCCACAG
 CGCCCCAGGCCCCAGGGAGAGGCGTACAGCTTAGAGATGGCCCTGCTG
 AACAGGGAACAAGAACAGGTGTGCCCATCCAGCGCCCCAGGGGTGGGAC
 AGGTGGCTGGATTTGGTGTGAAGCCCTTGAGCCCTGgAACCCAAcCAC
 GCAgGGCAGTTGGTAGATGCCATTTGGGGAGAGGCCACAGGTAAGGGC
 CATGGGCCCTTGAGGGGGCCAGGAGCTGAGGACAGGGACAGAGACGGCCC
 AGGCAGAGGACAGGGCCATGAGGGGTGCACTGAGATGGCCACTGCCAGCA
 GGGCCAGCTGCCAACCCGTCCAGGGAACCTTATTCAGCAGTCAGCTGGAGG
 TGCCATTGACCCTGAGGGCAGATGAAGCCCAGGCCAGGCTAGGTGGGCTG
 TGAAGACCCCAGGGGACAGAGCTCTGTCCCTGGGCAGCACTGGCCTCTCA
 TTCTGCAGGGCTTGACGGGATCCCAAGGCCTGCTGCCCTGATGGTAGTG
 GCAGTACCGCCAGAGCAGGACCCACAGCATGGAAAACCCCAACGGGACGCA
 GCCTGCGGAGCCCAAAAACAGTAAGGAGCCGAAGCAGTCATGGCACGG
 GGAGTGTGGACTTCCCTTTGATGGGGCCAGGCATGAAGGACAGAATGGG
 ACAGCGGCCATGAGCAGAAAATCAGCCGGAGGGGATGGCCTAGGCAGAC
 GCTGGCTTTATTTGAAGTGTGGCATTGTTGCTGGTGTGTATTGTTGGTA
 TTGATTTTATTTAGTATGTCAGTGACATACTGACATATTATGTAACGAC
 ATATTATTATGTGTTTAAAGAAGCACTCCAAGGGAACAGGCTGTCTGTAA
 TGTGTCCAGAGAAGAGAGCAAGAGCTTGGCTCAGTCTCCCCAAGGAGGT
 CAGTTCCTCAACAGGGGTCTAAATGTTTCTGGAGCCAGGCCTGAATCA
 AGGGGgTCATATCTACACGTGGGGCAGACCCATGGACCATTTTCGGAGCA
 ATAAGATGGCAGGGGATACCAAGCTGGTCTTACAGATCCAGGGCTTTG
 ACCTGTGACGCGGGCGCTCCTCCAGGCAAAGGGAGAAGCCAGCAGGAAGC
 TTTCAGAACTGGGGAGAACAGGGTGCAGACCTCCAGGGTCTTGTACAACG
 CACCCTTTATCCTGGGGTCCAGGAGGGGTCAGTGGGGATTTAAGTGGGG
 GACCATCAGAACCAGGTTTGTGTTTGGAAAAATGGCTCCAAGCAGAGA
 CCAGTGTGAGGCCAGATTAGATGATGAAGAAGAGGCCAGTGGAAAGTCGAT
 GGGTGGCCAGGTAGCAAGAGGGCCTATGGAGTTGGCAAGTGAATTTAAG
 TGGTGGCACCAGAGGGCAGATGGGGAGGAGCAGGCAGTGTGACTGT
 CTATAGAAAATCTAAAATGTATACCCTTTTAGCAATATGCAGTGAGTGT
 AAAAGAACACATATATATTTAAATTTGTGTAATCCACTTCTAAGGATTCA
 TCCCAAGGGGGGAAAATAATCAAAGATGTAACCAAAGGTTTACAAAACAAG
 AACTCATCATTAAATCTTCCCTTGTGTTATTTCAACGATATTATTATT
 ACTATTATTATTATTATTATTtTgtcttttgcatttttagggccactc
 ccacggcatagagaggttccaggctaggggtcaaatcggagctacagct
 gccggcctacgccagagccacagcaacgcaggatctgagccacagcaatg
 caggactacaccacagctcatgtaacgctggatccttaaccaatgag
 fgaggccagggatcgaacctgtaactcatggttcctagtcggatcatt
 aaccactgagccacgacaggaactccAACATTATTAATGATGGGAGAAAA
 CTGGAAGTAACCTAAATATCCAGCAGAAAAGGGTGTGGCCAAATACAGCAT
 GGAGTAGCCATCATAAGGAATCTTACACAAGCCTCCAAAATTGTGTTCT
 GAAATTGGGTTTAAAGTACGTTTGCATTTTAAAAAGCCTGCCAGAAAATA
 CAGAAAAATGCTGTGATATGCTCTGGCTGATAGGATTTTGCTTAGTTT
 TAATTTGGCTTTATAATTTCTATAGTTATGAAAATGTTCAAGAAGA
 TATATTTCAATTTAGCTTCTAAAATAATTATAACACAGAAGTAATTTGTG
 CTTTAAAAAATATTCAACACAGAAGTATATAAAGTAAAAATTGaggagt
 tcccatcgtggctcagtgattaacaacccaactgatccatgaggata
 tggattgatccctggccttgcctcagtggttgaggatccagtggtgctg

```

tgagctgtggtgtaggttcagacacagcactctggcgttgctgtgactc
fgcgttagccggcagctacagctccatttggacccctagccctgggaacc
tccatatgcctgagatacggcccTAAAAAGTCAAAAGCCAAAAAATAGT
AAAAATTGAGTGTCTACTTACCACCCCTGCCACATCTTATGTCTAAAA
CCCGTTCTCCAGAGACAAACATCGTCAAGTGGGTCTATATTTCCAGCC
CTCCTCCTGTGTGTATGTCCGTAAAAACACACACACACACACACACG
CACACACACACACAGTATCTAATTAGCATTGGTATTAGTTTTTCAAAG
GGAGGTCACTGCTACCTTTTAGGCGGCAAATAGATTATTTAAACAAATC
TGTTGACATTTTCTATATCAACCCATAAGATCTCCCATGTTCTTGAAAG
GCTTTGTAAGACATCAACATCTGGGTAAACCAGCATGGTTTTTAGGGGT
TGTGTGGATTTTTTTCATATTTTTTAGGGCACACCTGCAgcatatggagg
ttcccaggctaggggtgaatcagagctgtagctggcggcctacaccaca
gccacagcaacgccagatccttaaccactgagaaaggccagggttga
cctgcatacctcatggATGCTGGTCAAGTTTATTTCTGCTGAGCCACAACA
GGAATCCCTGAACCAGAAAGCTTTTAAACATTCCACTTTGCATGGACAT
TTAGATTGTTTCCATTTAAAAATACAAATTACAaggagtcccctgtgg
ctcagtggaacgaattggactaggaacctgaggttcgggttcgatcc
ctggccttgctcgggtgggtaaggatccagcattgatgagataggtg
taggtcgcagagctggctcggatcccagttgctgtgctctgctgctgtagg
ccgcaacaacagctccgattcagcccctagccTGggaacctccatgtgc
cacaggagcagccctaGAAAAGGCAAAAAGACAAAAAATAAAAAATTA
AATGAAAAAATAAAATAAAATACAAATTACAAGAGACGGCTACAAGGAA
ATCCCCAAGTGTGTGCAAAATGCCATATATGTATAAAATGTACTAGTGTCT
CCTCGCGGAAAAGTTGCCTAAAAGTGGGTGGCTGGACAGAGAGGACAGG
CTTTGACATTCTCATAGGTAGTAGCAATGGGCTTCTCAAAATGCTGTTC
AGTTTACACTACCATAGCAAATGACAGTGCCTCTTCTCTCCACCCTTG
CCAATAATGTGACAGGTGGATCTTTTTCTATTTTGTGTATCTGACAAGCA
AAAAATGAGAACAaggagtctctgctggtgagtgagagacaatctgac
taggaacctgaaatttcgggttcaatccctggcctcactcagtaggtaa
aggatccagggttcagtgagctgtgggtaggctgcagacacagtgcaa
attggccctgttgctgtggtgtagggcggcagctatagctccaatt
ggaccctaaccgggaacctcctatgccgtgggtgaggccctAAAAAA
AAGAGTGCAAAAAAATAAGAACAAAAATGATCATCGTTTAATTCT
TTATTTGATCATGGTGAAACTTATTTCCITTTATATTTTATTGACTG
ATTTTATTTCTCCTATGAATTTACCGTCATAGTTTTGCCTGGGTGTTTT
TACTCCGGTTTTAGTTTTGTTGTTGTATTTTCTTAGAGAGCTATAGAA
ACTCTTCATCTATTTGGAATAGTAATTCCTCATTAAAGTATTTGTGCTGCA
AAAAATTTCCCTGATCTGTTTTATGCTTTTGTGTTGGGGTCTTTACG
AGAAAGCCTTTTTAGTTTTACACCTCAGCTTGGTTGTTTTCTTGATTG
TGTCTGTAATCTGCGCCAACATAGGAAACACATTTTACTTTAGTGT
TTTTCTATTTTCTCAAGTACGTCCATTTGTTTTGGTGTCTGATTTTACT
TTGCCTGGGGTTTTGTTTTGTGTGGCAGGAATATAAACTTATGTATTTTC
CAAATGGAGAGCCAATGGTTGTATATTTGTTGAATTCAAATGCAACTTTA
TCAAAACACAAATCATCGATTTATCACAACCTTCTCTGTTTTATTGATC
TAATGATCAATTCCTGTTCCACGCTGTTTTAATTTTATTTAGCTTTGTGGA
TTTTGGTGCCTGGTAGAGAACAAAGCCTCCATTTTTCATTCAAAATAG
TCCCGTCTATTATCTGCCATTGTTGTAGTATTAGACTTTAAATCAATTT
ACTGATTTTCAAAAGTTATTCCTTTGGTGATGTGGAATACTTTATACTTC
ATAAGGTACATGGATTCAATTTGTGGGAATTGATGTCTTTGCTATTGTGG
CCATTTGTCAAGTTGTGTAATTTTTACCCATGCCAACTTTGCATATTGT
ATGTGAGTTTATTTCCAGGGTTTTAATAGGATGTTTATTGAAGTTGTCA
GTGTTTCCACAATTTTCATCGCCTCAGTGCTTACTGTTTGATAAAAGGAA
ACCTACTCACTTTTGCCTATTGCTCTTGTATTCAATCATTTTATGTTAACT
CTTGTGTTAATTTTGTGAGAGTTTTTTCAGCTGACTGTCTGGGGTTTTCTTA
ATAGACTAGCCCTTTGTCTGTAAGAATAATTTTATCGAATTTTCTTAA
CACTCACACTCTCCCCACCCCAACCCCGCTCATCTCCTTTTATTGGGTC

```

	AAATCTGTAGAATACAATAAAAAGTAAGAGTGGGAACCTTAGCCTTTAAGT CGATTTTGCCTTTAAATGTGAATGTTGCTATGTTTCGGGACATTCTCTTT ATCAAGTTGCGGATGTTTCCTTAGATAAATTAACCTTAATAAAAAGACTGGAT GTTTGCTTTCTTCAAATCAGAATTGTGTTGAATTTATATTGCTATTCTGT TTAATTTTGTTCAAAAAATTACATGCACACCTTAAAGATAACCATGAC CAAATAGTCCTCCTGCTGAGAGAAAATGTTGGCCCAATGCCACAGGTTA CCTCCCGACTCAGATAAACTACAATGGGAGATAAAAATCAGATTTGGCAAA GCCTGTGGATTCTTGCCATAACTCTCAGAGCATGACTTGGGTGTTTTTTC CTTTTCTAAGTATTTAATGGTATTTTTGTGTTACAATAGGAAATCTAGG ACACAGAGAGTGATTCATGAGGGGAACGCATTCTGGGATGACTCTAGGC CTCTGGTTTGGGAGAGCTCTATTGAAGTAAAGACAATGAGAGGAAGCAA GTTTGCAGGGAAGTGTGAGGAATTTAGATGGGGAATGTTGGGTTTGAGGT TTCTATAGGGCACGCAAGCAGAGATGCACTCAGGAGGAAGAAGGAGCATA AATCTAGAGGCAAAAAGAGAGGTCAGGACTGGAAATAGAGATGCGAGACA CCAGGGTGGCAGTCAGAGAGCACAGTGTGGGTGAGAAGACAGTGAAGAA CACAAGGGACAGAGAGGGATCTCCAACCTCACTGGGATGAGGGCCTTGT GGCCTTGACCTGAGAGATTTCCAGGAGTTGAGGGTGGGAAGGAGAGGGCT CCTGCACATGTCCTGACATGAAACGGTGCCAGCATATGGGTGCTTGAA GACATTGTTGGACAGATGGATGGATGATGGATGATGGATGAATGGATGA TGGAAAGATGATGGATAAATGGATGATGGATGGATGGATGGACAGAAGGACAAG AGATGGACAGAAAGACAGTGATCTGAGAGAGCAGAGAAGGCTTCATGAAA GGACAGGAACTGAACTGTCTCAGTGGGTGGAGACAATGGTGTAGGGGGTT TCCACATGGAGGCACCAAGGGTTCAGGAATAATCTAGTGTCCACAGGCCCA GGAAGGAAGCTGTCTGCAGGAAATGTTGGGGAAGAAGCTCAGAGTCTTA AATGAGGTCAGGAGTGGTCAGGAGGGTCTGATCAGGTAAGGACTCATGTC CATCATCACATGGTCACCTAAGGGCATGTAGCTCTCAGCATCTCCATCAG GACAGTCTCAGAATGGGGCGGGTACACACTGGGTGACTCAAGCCGTG GGTCATGCCTGCCTCGGACGTGGGCCTGGGCATGGGACACCTCCAGACC ATGGGCCCGCCAGGGCTGCACTGGcctctggtgggctagctaccgctc aagcaacacaggacacagccctacctctgcaaccctgtcccgaacgc ccatctggttcctctccagcccggcccagggaacaggactcaggtgct agcccaatgggttttctgagcctcagtcagcgtgTATTTCTCCGGC AGCGAGACTCAGTTCACCGCCTTAGGttaagtgttctcatgaattct agcagtcctgcactcgtatgcccggaaagtcactttgtcgtggggg ctgttccccgtgcccttgagaatcaaggattgcccactttctctgtg ggggaggtgctggtcttgggtgaccagcaggaggccccaaaagcag gagcagctgctccagAATACTGTCGGCTACAGCTCAAACAGGAGGC CTGGACTGGGGTTTAACCACCAGGGCGGCACGAAGGAGCGAGGCTGGGAG GGTGAGGACATGGGAGCCTGAGGAGGAGCTGGAGACTTCAGCAGGCCCC AGCTCCGGGCTTCGGGCTCTGAGATGCTCGGACGCAAGGTGAGTGACCC ACCTGTGGCTGACCTGACCTCAGGGGGACAAGGCTCAGCCTGAGACTCTG TGTCCCCATCGCCTGCACAGgggattcccctgatggacactgagccaacg acctccgtctctccccagccccaggtcagcccaaggccgcccccaagg tcaacctctcccctctctgaggagctcggcaccacaaggccacc ctggtgtctataagtactctaccgAAGGGCGAATTCCAGCACAC TGGCGGCGTTACTAGTGGATCCGAGCTCGGTACCAAGCTTGATGCATAG CTTGAGTATCTA
--	--

Seq ID No.33	agatcttaaacaccaggagcaaggccaggatcgaaccgcatcctcatg aatcctagtgggtctgtaaccgctgaaccacaatggaactcctGTCT TTCACATTTAATTCACAACCTCTCCAGGATTCTGGGGGTGGGTGGGGAAT CCTAGGTACCCACTGGGAAAAGTAATCCAAGGGGAGAGGCTCACGACTcT AGGGATCGGGGAGGAGGGAAGGTATCTCCAGGAAACTGGCCAGGACAC ATTGGTCTCCGCCCTCCCTTCTCCACTCCTCCTCCAGACAGGACTG
--------------	---

TGCCACCCCTGCCACCTTTCTGGCCAGAAGTGTCCATGGCAGGTGACC
 TTCACATGAGCCCTTCCTCCCTGCCTGCCCTAGTGGGACCCCTCCATACT
 CCCCTGGACCCCGTTGTCTTTCTTTCCAGTGTGGCCCTGAGCATAACT
 GATGCCATCATGGGCTGCTGACCCACCCGGGACTGTGTTGTGCAGTGA
 CACTTCTGTGTCATCAGGGCTTTGTAATTGATAGATAGTGTTCATCATC
 ATTAGGACCGGGTGGCCTCTATGCTCTGTTAGTCTCCAAACACTGATGAA
 AACCTTCGTTGGCATAAGTCCAGCTTCTGTTGCCATCCATAAATCTTG
 ACTTAGGGATGCACATCCTGTCTCCAAGCAACCACCCCTCCCCTAGGCTA
 ACTATAAACTGTCCAATGGCCCTTGTGTGGTGCAGAGTTCATGCTTCC
 AGATCATTCTCTGCTAGATCCATATCTCACCTTGTAAAGTCATCTATAA
 TAACTGATCCATTGATTATTTGCTTCTGTTTTTCCATCTCAAACAGC
 TTCTCAGTTCAGTTCGAATTTTTTATTCCTCCATCCACCATACTTTCC
 TCAGCCTGGGGAACCCCTGCCCCAGTCCCATGCCCTTCTCCTCTG
 CCCAGCTCAGCACCTGCCACCCTCACCTTCTGCTACTCCCTAGGACT
 GGACCATCCACTGGGGCCAGGACACTCCAGCAGCCTTGGCTTCATGGGCT
 CTGAAATCCATGGCCCATCTCTATTCTCACTGGATGGCAGGTTGAGAGA
 TGTGAAAGGTCTAGGAGGAAGCCAGGAAGGAAACTGTTGCATGAAAGGCC
 GGCCTGATGGTTCAGTACTTAAATAATATGAGCTCTGAGCTCCCCAGGAA
 CCAAAGCATGGAGGGAGTATGTGCCTCAGAATCTCTCTGAGATTCAGCAA
 AGCCTTTGCTAGAGGGAAAATAGTGGCTCAACCTTGAGGGCCAGCATCTT
 GCACCACAGTTAAAAGTGGGTATTTGTTTTACCTGAGGCTCAGCATTAT
 GGAACCGGGCTCTGACACAAACACAGGTGCAGCCCGGCAGCCTCAGAAC
 ACAGCAACGACCACAAGCTGGGACAGCTGCCCTGAACGGGGAGTCCACC
 ATGCTTCTGTCTCGGGTACCACCAGGTCACCATCCCTGGGGGAGGTAGTT
 CCATAGCAGTAGTCCCTGATTTCCGCCCTCGGGCGTGTAGCCAGGCAAG
 CTCTGCCTCTGGACCCAGGGTGGACCCTTGTCTCCCACTACCCTGCACA
 TGCCAGACAGTCAAGACCACTCCACCTCTGTCTGAGGCCCCCTTGGGTG
 TCCCAGGGCCCGGAGCTGTCTCTACTCATGGTTCCTCCACTGGGTAC
 AAAAGAGGCGAGGGACACTTTTCTCAGGTTTGCGGCTCAGAAAGGTACCT
 TCTAGGGTTTGTCCACTGGGAGTCACCTCCCTTGCATCTCAATGTCACT
 GGGGAAAAGTGGGTCCCATGGGGGGATTAGTGCCACTGTGAGGCCCTGA
 AGTCTGGGGCTCTAGACACTATGATGATGAGGGATGTGGTAAAAACCC
 CACCCAGCCCTTCTTCCCGGACCCCTGGGCTGTGGCTCCCCATTGCAC
 TTGGGGTCAGAGGGGTGGATGGTGGCTATGGTCAGGCATGTTTCCCATGA
 GCTGGGGGCACCTGGGTGACTTTCTCCTGTGAATCCTGAATTAGCAGCT
 ATAACAAATTGCCAAACTCTTAGGCTTAAAACAACACACATTTATTCT
 CTGGTCCAGGGTCAGAAAGTCCAAAATGAGTCTATAGGCTAAAATTTGA
 GGTGTCTCTGGGTTGAGCTCCTCCTGGAAGCCTTTTCCAGCCTTAGAGT
 CCCAAGTCTTGGCTCTGGGCCCTCCCTCAAGCTTCAAAGCCACAGAAG
 CTTCTAATCTCTCCTTCCCTCTGACCTCTGCTCCCATCCTCATACC
 CTGTCCCTCACTCTGACCCCTCCTGCCTCCCTCTTTCCCTTATAAAGACC
 CTGCATGGGGCCACGGAGATAATCCAGGGTAATCGCCCTCTTCCAGCCC
 TTAATCCATCCCATCTGCAAAAATCCCTGTACCCATAATGGACCTACT
 GATGGTCTGGGGTTAGGACGTGGACAACCTGGGGCCTTATTCATCTGAT
 CAAACTCCAGTTCCAGACCCCAAGACCCCGGGCATTAGGGAAACTTC
 TCCAGTTCCTCTCCCTCTGTCTCCTGCCAGTCTCCAGGATGGGCCACT
 CCCGAGGGCCCTCAGCTCAGGCTCCCCCTCCTTTCTCCTGGCCTCTTG
 TGGCCCCATCTCCTCCTCCGCTCACAGGGAGAGAACTTTGATTTAGCTT
 TGGCTCTGGGGCTTGTCTCCTTCTGGCCATTGGCTGAAGGGCGGGTTTC
 TCCAGGCTTACCTGTCAATCAAAACCGCCCTTGGAGGAAGACCCTAA
 TATGATCCTTACCCTACAGATGGAGACTCGAGGCCAGAGATCCTGAGTG
 ACCTGCTCACATTCACAGCAGGACTGAACCCCAAGTACCTACCCAATC
 CAGGGCTCAGCGCTTTTTTTTTTTTTTTTCTTTTTTgccttttcaggggc
 gctcccgaacatatggagattccaggctaggggtctaattggagcagt
 cgacactggcctaagccasaagccacagcaacaaggcaagccgctctgc
 agcctataccacagctcagcgaatgccggatccttaaccactgagcaa

	<p>agccagggaftgaacctgcaacctcatgttcttagtcaaatgttaac cactgacctatgacgggaactcccAGGGCTCAGCTCTTGACTCCAGGTTCC GCAGCTGCCCTCAAAGCAATGCAACCCTGGCTGGCCCCGCCTCATGCATC CGGCCTCTCCCAAAGAGCTCTGAGCCACCTGGGCCTAGGTCCTCCTC CCTGGGACTCATGGCCTAAGGGTACAGAGTTACTGGGGCTGATGAAGGGA CCAATGGGGACAGGGGCCTCAAATCAAAGTGGCTGTCTCTCATGTCCC TTCTCTCCTCAGGGTCCAAAATCAGGGTCAGGGCOCCAGGGCAGGGGCT GAGAGGGCCTCTTTCTGAAGGCOCTGTCTCAGTGCAAGTTATGGGGGTCT GGGGGAGGGTCAATGCAGGGCTCACCTTCAGTGCCCCAAAGCCTAGAGA GTGAGTGCCTGCCAGTGGCTTCCCAGGCCAATCCCTTGACTGCCTGGGA ATGCTCAAATGCAGGAAGTGTCAACAACCTTCAGTCAGGGGCTGTCTG GGAGGAAAAACAACCTCAGAATTGGGGGTTTCAGGGAAGGCCAAGTCCCAAGC ATAGCAGGAGCTCAGGTGGCTGCAGATGGTGTGAACCCCAAGGAGCAGGAT GGCCGGCACTCCCCCAGACCCTCCAGAGCCCCAGGTTGGCTGCCCTCTT CACTGCCGACACCCCTGGGTCCACTTCTGCCCTTCCCACCTAAAACCTT TAGGGCTCCCCTTTCTCCCAAATGTGAGACATCACACGGCTCCCAGGG AGTGTCCAGAAGGGCATCTGGCTGAGAGGTCTGACATCTGGGAGCCCTCA GGCCCCACAATGGACAGACGCCCTGCCAGGATGTCTGTCAGGGCTGTTA GCTAGGCGGGGTGGAGATGGGGTACTTTGCCCTCAGAGGGCCCCGGCCCC ACCATGAAACCTCAGTGACACCCCATTTCCCTGAGTTACATACCTGTAT CCTACTCCAGTCACTTCCCCACGAACCCCTGGGAGCCAGGATGATGCT GGGGCTGGAGCCACGACCAGCCACGAGTGATCCAGCTCTGCCAATCAGC AGTCATTTCCCAAGTGTCCAGCCCTGCCAGGTCCACTACAGCAGTAAT GGAGGCCCCAGACACCAGTCCAGCAGTTAGAGGGCTGGACTAGCACCAGC TTTCAAGCCTCAGCATCTCAAGGTGAATGGCCAGTGCCCTCCCCGTGGC CATCACAGGATCGCAGATATGACCCTAGGGGAAGAAATATCCTGGGAGTA AGGAAGTGCCCATACTCAAGGATGGCCCTCTGTGACCTAACCTGTCCCT GAGGATTGTACTTCCAGGCGTTAAAACAGTAGAACCCCTGCCCTGTGAACC CCCGCCAAGGGACTGCTTGGGGAGGCCCCCTAAACCAGAACACAGGCACT CCAGCAGGACCTCTGAACTCTGACCACCCCTCAGCAAGTGGCACCCCCCGC AGCTTCCAAGGCAC</p>
--	--

Seq ID No.34	<p>AACAAGATGCTACCCACCAACAAAATTCACCGGAGAAGACAAGGACAGG GGGTTCTGGGGTCTGACAGGGTCACCAAAGAGGGTTCTGGGGCAGCAG CAACTCCAGCCGCTCAGAACAGAGCCTGGAAGCTGTACCTCAGAGCAG AGGCGGAGAGAGAAAAGGGCCTCTTGGTGGGTGAGCAGGAGCAGAGGCTCA GAGGTGGGGTTCAGCCCCCCTTCAACAGGCCAACACAGTGAAGCAGC TGACCCCTCCACCTTGGAGACCCAGACTCCTGTCTCCACGCCACCTTG GTTTTAAGGTAATTTTTATTTATATCAGAGTATGGTTGACTTACAATG TTGTGTTGGTTTCAGGTGTACAGCAGAGTGATTCACTTCTACATAGACTC ATATCTATTCTTCTCAGATTCTTTCCCATATAGGTTATTACAGAATAT TGAGTAGATCCCTGCTGATTACCCATTTTTATAATTGTATATGTTAATCC CAAACCTCCTAATTTATCCCTCCCCAGACTATGATTCTTTATATCTCTATC TGTTTTCTAATCTGTCTCCTCTAAGTCACCCTAGGAGAGCAGAGGGGTCA CGTCTGTCTGTCTCCTGGCCAGCCACCTCTCTCCACCCAGGAATCCCTTG CATTTGGTGCCAAGGGCCCCGGCCCCGCCCTAAAGAGAAAAGGAGAACGGGA TGTGGACAGGACACCGGGCAGAGAGGGACAAGCAGAGGATGCCAGGGTAG GGAGGTCTCCAGGGTGGATGGTGGTCTGTCCGCAGGCAGGATGAGGCAGG AAGGGTGTGGATGTACTCGGTGAGGCTGGCGCATGGCCTGGAGTGTCTG AGCCCTGGGAGGCCTCAGCCCTGGATCAGATCTGTGATTCCAAAGGGCCA CTGCATCCAGAGACCGTTGAGTGGCCATTGTCTGAACCATTTATAGAA CACAGGACAAGCGGTACCTGACTAAGCTGCTCACAGATTCCATGAGGCTG ATGCCAGGGTTGTCAACCCATCTCACAGGCAGGGAAAATGATGCATATAC TGCAGAGCCAGGCAGAGGCCCTCCAGTGCCCCCTCCAGCCTGTGGCCC</p>
--------------	--

	CCCTCCAGTGGCTGGACACTGAGGCCACACTGGGGCACCTGTGGAGATC t
--	--

Seq ID No.35	AGATCTGGCCAGGCCAGAGAAGCCCATGTGGTGACCTCCCTCCATCACTC CACGCCCTGACCTGCCAGGGAGCAGAAAAGTAGGCCAGGGTGGACCCGGT GGCCACCTGCCACCCCATGGCTGGGAGAAGGGAGGGCCTGGGCAAAGGGC CTGGGAAGCCTGTGGTGGGACCCAGACCCAGGGTGGACAGGGAGGGTC CCACACCCACAGCCATTTGCTCCCTCTGTGGGTTCAGTGTCTCATCTC ATCTGTGGGGAGGGGGCTGATAATGAATCTCCCCATTGGGGTGGGCTTG GGGATTAAGGGCCAGTGTCTGTGATATGCCTGGACCATAGTGACCCTCA CCCTCCCCAGCCATTGCTGTACCTTCCGGGCTCTTGCCAGGCCTGCCT GACATGCTGTGTGACCCGGGCAAGATGATCCCCCTTTCTGGGCCCCAGC CTTCTCTCTGCTCCGGAAGTGCTTCTGGGAAACCTGTGGGCTGGATC CTATAGGAAACCTGTCCAATTCCTGGATGCACAGAGGGCAGGGAGGCC TGGGCTGGAGGGGCAGGGAGGCTCGAGGTGGGAGCAGGGTAGGGGCCAG TCCAGGGCAAGGAGGTGGGTGGGTAGGGTG
--------------	---

Seq ID No. 36 :	GATCTGTGTTCCATCTCAGAGCTATCTTAGCAGAGAGGTGCAGGGGCCTC CAGGGCCACCAAAGTCCAGGCTCAGCCAGAGGCAATGGGGTATCGATGAG CTACAGGACACAGGCGTCAGCCAGTGTGAGGGAGAATCACCTTGTGTTGT TTTCTGAGTTCTCTTAAATAGAGTTAATTGGTCTTGGCCTTACGGTTT ACAATAACAACCTGCACCTGTAAACAACGTGAAGAGTACAGAACAACAA TGGGGGAAAACATATTTACCTGAAAGAGCCACCGCTCATATTTTGATGG ATTTCTTCTAGTTTAACTCTGTTTAAATTGTAACCTGTTAAAACAAACA TAAATAAAGAAAATGCATCTGTAAAGTTTAAAAGTCATATCTATGGTGAT GGTTGCAAAACACTGTGAATGTTCACTTTGAAATCGTGAACCTACGTGA TATGCATGTCCCGTTAATTAACCTCACAGGCTCAGAATGTGGTTCATTAT TTCTTTAATTTTCTTTAATTTTATGTCCTCTGTGTGCCCTTAAACCA ACTACTTTTCAGCTCTGCCCTGTTTTGACCTTCACATAGATGACCTTTGT GAGTGTTTTCTTTCTCAACACTGGGTCTGATACCCACCCACGCTGTCTGC TGTCACTGCGGACGTGGAGGGCCACCACCCAGCTATGGCCCCAGCCAGGC CAACACTGGATGAATCTGCCCCAGAGCAGGGCCACCAACACTGGAGGTG CAGAGAGGGTTTCTTACGGGCCATCATTATCCAAGGCATTGTTTCTACTG TAAGCTTTCAAATGCTTCCCCTGATTATTAAGAAATAATAAGATGGG GGGAAAGTACAAGAAGGGAAAGTTCCAGCCCAGCCTGAAGATCGTGCTGG TTGTATCTGGAGCCTGTCTTCTGACAGGCCTCTATTCCAGAGTTA
-----------------	--

Seq ID No. 37 :	GGATCCTAGGGAAGGGAGGGCGGGGCCTGGACAAAGGGGGCCTAAAGGA CATTCTACCTATCCCACTGGACCctctgtctctgaggaggagca gagaggggctgaggcctttccagCTCCTCTGAGTCCCTCCTCOGAG CACCTGGACGGAAGCCCCTCCTCAGGGAGTCTCAGACCCCTCCCCTCCA GCCAGGTTGGCCTGTGTGGAGTCCCCAGTAAGAATAGAATGCTCAGGGCT TCGAGCTGAGCCCTGGCTACTTGGGGGGTGTGGGATTTGGGGTGTGCTG GGCGGGAGCTGGGGTGTCACTAGATGCCAGTAGGCTGTGGGCTCGGGTC TGGGGGTCTGCACATGTGCAGCTGTGGAAAGCCCTATTGGTGGTACCC TCAGACACATATGGCCCCTCAATTTCTGAGACCAGAGACCCAGTCTGGC CTTCCAGAACAGCTGCCCTGGTGGGGGAGATGTAGGGGGCCTTCAGC CCAGGACCCCAACGGCAGGGCCTGAGGCCCCCATCCCTTGTCTGGGC CCAGAGCCTCAGCTATCAGGCCTATCAGAGATCCTGGCTGCCAGCTCAG GTTCCCAGGAGCCAGAGGGAGGCCAGGGGTTACTAGGAAATCCGGAAAAG
-----------------	--

GCCCCTCCAGACCCCGCTCTGGGCAGAACCTGGGAAGGGATGTGGCCAT
CGGGGGATCCCTCCAGGCCATCTCCTCAGATGGGGGCTGGTCGACTAGCT
TCTGAGTCCTCCAAGGAACCGGGTCTTCTAGTCATGACTCTGCCAGAT
GAAGAAGGAGAGCACTTCTCTCCATCAGGAGGATCTGAGCTTCTCTTAAT
TAGAATCAGCTCCTTGGCTTCTACCCCTTAAAAAAGGTACAGAAACTTT
GCACCTTGATCCAGTATCAGGGGAATTTATCAATCAATGTGGGAGAAATT
GGCATCTTTACCACACTGAATCTTTCAATCCATGAATATCCTCTCTCTCT
TCCATGCATAGGTTTTAATAATTCTCAATGGAGTTTAATGTAAGTTTTCC
TCATAGACAATTGCCTTTGGACATCTCTTTAGACTCATCTCTAGTAACT
GATATTCTTAATGCAATTATAAAATGTATCCTGCTTAATGTTATTTTCTA
TTCATTTGCTGTTATATAGAGATACAATGAGTTTCCACATTTGAAACTGG
ATCTGGTAAATTGGCTACCCTTTTTTTATAGATTCTATTAATTTTTATAC
ATTCTGTGGGACTTGGCTACATACTTAATCATGTACCTGTGAAGAATGAC
AATTTGGTTGCTACCCTCCAATTCTTATATGTCCTCAFTTCTTCCCTCT
GCTGGTACTCTGGCAGCAGCAGGGAAGATAATGGGCCTCCTTATCTTGTG
ACAAAAGGATGTTTTAAAGATTTTCGTTATAAAACATAACGCTTTCTGGT
TTTTCTTAAAGATTCTCTCACCAGCTTAAGAAAATTTTCTTATACTCTGT
ATGATAAATGGGTTTTTGACAATCATTGTTGCATTTTACCTAGTGTTTT
CTCTGCATCTTATATGCTTTTTCTCCTTTAATCCTGAAAATGTTTCGA
TTTTCTAACATTGAACCAATCTTACATTCCTGGAATGGATGGACCAGAC
TAGTCCACATGTTTATTCTGCCAATGGCTAGATTTTGTGTTCaatattt
IgttcagaatgittgcatctatattcttGAGTGAGACAGAGCTGCCCTTG
TTAGGTTTTACAACCGAGGTTGTGTTAGCTTCATAAAATGAGACGTTTAT
TCTCTAAAAGAATTGTTTCGCTTCTCTGGATGAATTTGTGTAAGGTTAGA
ATTGCTTACCAGTGAagatctCGGGgCCAGTTCCTTCTTAGGGGAAGATT
TTCAACAATTAAGCTCAATGCCTTTAGAAGAAGTGGAGTTTCTATTATT
TCTTGAGTTAAATATATGTATTTAATTAGACTTTCTAGGAATAGTCTCAT
TTCATCTCAAATAATTGACATATGCTATTAAGCAGATTCTCATGAACCA
TTGTAGGTATCCAGGTCTAGAAAAATGTTCCCTTTGCATCCCTAATGT
GTTTAAATTTTACCTTCTTTCTTTTGTCTTGAGAAATTCACCAAATCAF
TTTTCAATTTCAAGTCATATCCCAAAGCAACCAACTCTCTACCTTCTGTTT
TATCATCCCTGCTGGATTTTTGTTATCTACTTCTCAGTATTTGTTCTTC
CCTTTCTTCTATTCTCATTCATTTTTCCCTTGTTTTCTAACTTTCTGA
GATATATGCTTAGTTCCTTCAATTTGAAGCCTTTTTATTTCTTTTTTTTT
TTTTGGTCTTTTTGTCTTTGTGTGTTGTTGTTGTTGCTATTCTTGGGCCG
CTCCCGCGCATATGGAGGTTCCAGGCTAGGAGTCGAATCGGAGCTGTA
GCCACCGGCTACGCCAGGCCACAGCAATGCGGGATCCGAGCCGCTCT
GCAACCTACACCACAGCTCATGGCAACGCCGGATCGTTAACCCACTGAGC
AAGGGCAGGAACCGAACCCGCAACCTCATGGTTCCTAGTCGGATTCTGTA
CCACTGTGCCACAACAGGAACCTCCGCCTTTTTATTTCTATAAAAATTC
TATGTACATTTTAAAGGTTATAGGTTTCTTCTATGTACCCCATTTGGCTGT
ATCCTCAGGGTCTGTGGAGTGATTCATTATTGTTCAAGTTCAATATGT
CTTCTGATTTTCCAATTTGAATACCTCTCTAAATCAGTAGGTGAATATT
CTTTTTCTTTTTCTTTCTTTCTTTTTTTTTTTTTTTTCTTTCAGCCAGGT
CCATGGCATGCAGAAATTTCCAGGCCAGGAATCAAACCTCTACCATGGCA
GTGACAATGTCCGATCCTTTACCCACTAGGCCACCAGGGAACCTCTGGGAG
CATATGTTTTTATTTCCCGACATCTGAGGATGCCTAGTATGTCTTCATTA
TTGATTTCTAGTTTGCCACTGATTTCTAGTATTTTGCTCATAGAGTGTAT
GCTCAATGGTTTTGGTCAATTTGAAATGATTTAGTCTGCTTTATGACCC
AGTATGTGGTCAGTTTTGTCAATGTTCTTTTTCTGCTTGAAGAGAACCTA
CATGCTGTAACCTCTGGGTGCATGTTCTGTATATAAGTCTATAGGCTGAGC
CGGGGGAGCCTTCTAATCTGCCGTTATCTTCTCGAGTTATTTAGGTAC
TATTTCTTAGCCATAAACCTTTAAATTTCTGATATCAATATAATGACCCCA
GCCCCGTTAGGGTCGGCACTTCATGTTATCTTTTTCCATCCATTTAATCC
CTCCCCACTGTTTTGGCCACACCCGTGGGATATGGGAGTTCCTGGGCCAA
GGATCaGATCTGAGCCGAGCTGCCACCTATGCCACAGCAgcagcaatga

	<p> tggatcttaaccactgcaccacactggggattgaaccaagcctcagc agcaaccaagctactgcagagacaacaccagatccttaacctgtgtgc catagcgggaaTTCCATCCATTTACTTTCAAGCCAGCTGAATAACCTAG CCCACCATGCCTGGACATGGGTGCTCTGCTTCAAATGATTTTGTTCAGTC AGCATCCATCTCTGAAATGTGTGCCAAGCATTATATGCATGCAAGAGTC ATGTTGGCACTTCTATCATTCCAACAGTTCAGTAGCCTTTGTATCATGA CATTCTTGGCCTTTCTCTACAATATTTGAGGCTGAGCAGACTGGCCGT GCCCTGTCCATGCTTCCAGAGCCTGTGTGCAGACTTCTGCTCTAGACAG AGACAGCTAACCATCCTGCAGTGCCAGAAAACCCAACCTCAAAGACCCTC AAGTAAGGAAGGATTTATTGGCTCACGTAATCTGGAATCCAGGCATGGGG TATTCAGGGCCACCTGAACCAGAGGCCCTGGCCCTGTTCTCTAAGCTTCT TCCTGCCCTGCCCTCGTTCTGGAAGTGACCCTGAAGGACAGCAATGAAGG GCAGCTCCCCAGGGACAGATGACTGAGAGGTCCATTTCAAGTCCAACCT GGCCTAGATTGAGAGGCAGCAAGAAATATGGACCTACAGTGAGTCACAGG ATTTACCAGTGGTTTGGCTGGGTTGTGAGTGTACAGGCTAAACATTTGG GTCCCTCCAAAATTAACATGTTGCCACTTAACCACAAAATCatggtat ttgggggtggggcccttgaggtaattagggttagaaAGAATGAAGAGGG GGCCCTTGTGATGGGACTAGTGCCCTTATAGAGAGAGAAGAGAGAGGG </p>
--	---

Seq ID No.39	<p> CACCTCATCCCCAACCACTGGATGGTGGCAAGTGGCAGGCTGAGAGGCT GCATATGAGCTCATCAAGAGGGTCCCCACCCACAGAGGCTGACCCAGCT GCCACTGCCACCTAGTGGCTGATCGGCCAAGAGCAGGAGCCCCAGGGGCA GCTCCATTCCCTGGGGCGGCCAGGGAACCACTGGTGGTAGGACAATTCC ATTGCACCTCATCCATCAGGAAAAGGTTTGCCTCCCTGGCAGTAATGCA TCTTCCATAACATGGTCCCTGGCCTCTTGGAAATGGCTTGGCCACCGTCA TGGCCTCACCACAAGCCTTGTGCTCAGCAAGGAACCTATTCCACAGC AAAGGACTTGCAGCCTGGAATGAACTGGTCTGACTACATACCCCACTGCC CAGAAGTAGGTGGTCTATTGCAAAGTGGAGTGGCTTACCCAAGACTCAGT TGTGCCAAAGTTGAGAGATAGCATCCTAAAATATGGGCTTATGTCTCACT GGCTGAGGTTTATTCTTTGAATCAAAGACAAATTATATGGTGTGGTCCCC CAGAGATAGAATACATGAGTCTGGGAATCAAGGGATAGAAGTAAGAAGAG ATTTTGTCAACATTAATCCCAATAACTCGCCAAAGAATATTTGCTTTCT GTCTTGGCAGCTCTGCTGCTTTGGCAATAACTTCTAGAAATAATAATGTCT CCACCAGGGGACTCCACAACGGTTCCATTGATTTGAAGCCAATAAGGCCAGA GGAGGGGCTGCCTTACTGGTCCGACTGGTCAGCCCTGATTACTAAGGAGA AATCAGGCAACTTCAACAAAATAAGGCAGGGGGGACTTTGTCTAGAACC CAAAGCACTAAGCATCTTAGTACTTTTTAGTTCTCAGAGCCTCCAAGAAC AAAGATTTAGCCCTCAGCACCACCAGGTAAAGAACAGGTAAATCCAGCT GAGGACAAGAGAAATATTGAATGGATAGAGGAAGAAAGAAATTATAGATA TCAACTATGGCCTCATGACTAGAGTCTCCAGATTAAGCGGAATAAAAATA CAGATGATTGATCTGAACATCAGGCCAAAACAACGAACAACAGTTAAGT GCGACCTAGGCAATATTTGGGACATACTTATACTAAAATTTTTTCGCTAT TTGAGCATCCTGTATTTTATCTGGCAACTTTATTCATCCCTAGCGAAAAA GGAAGTGTGGTAACTTAGTGTATTTTTACTTTGCTCATTATTGTGTATAT ACCTACTTGTATTTATCAATCATATTTACTCTGTTCTCAGTATTACTTIA TATAGCAGTTGGTGGTGTAGGTTAGCAACATATTCAGTGGAACTGTGACT GAATTTGAGGAGAAATTAACAGAGTTGGCTGTGGCTACAATAACCCCTCG GGACATGTGTCCCCTCATTTTGGGGAGATGGTTtagatctCTGGGTAATG TTAGGGCATCTGAGCCAGAAACCAAGATTTTGCCAGCTGGTGCAATGTCA GATTTTACCAGCAGAGGGTGCCAGAGGAATGCGGCAAAACCCGAGTGCCA GAAAGCACCTCCCTGTTTTCCAGCTTTTCTTCTTTTTATTTATTTATT TACGGCCCAGGAGTCCGTAATAGCGCTGAGGATGGCCAGGCTCTTCTCA GCAGCCCTGACTGACTAGTTCAGCAATGCGCTCAGGCCCATCTGGCCAC CGGGCAGCCTCTTCTGTGGTAGCTCCAGCCTCAGCCAGTGCAAAAAGGCTA </p>
--------------	--

	CCCTACACTGGCGCCACTTCTACAATCAGCACTGGCCACACCCTCCACGC CATCCGGCACGGAGCCAGGTGATCTGCCGGCCAGATTGCAGTTTCGTGCTG CCTGAGTCCAGGTGATTACACTGGCTGCATCTTTTCTTTCTGGACCAITC atccatttttt
--	--

牛λ轻链

在另一实施方案中，提供了编码牛λ轻链基因座的核酸序列，其可以包括至少一个连接区 - 恒定区对和/或至少一个可变区，如 Seq ID No. 31 表示。在 Seq ID No. 31 中，牛λ C 可以在残基 993-1333 处找到，J-C 对可以在残基 33848-35628 的互补序列处找到，其中 C 是 33848-34328 的互补序列，而 J 是 35599-35628 的互补序列，V 区可以在残基 10676-10728、11092-11446、15088-15381、25239-25528、29784-30228 和 51718-52357(或者其互补序列)处找到。Seq ID No. 31 可以见 Genbank 编号 AC117274。还提供了载体和/或寻靶构建体，其含有 Seq ID No. 31 的全部或者部分，如 Seq ID No. 31 的至少 100、250、500、1000、2000、5000、10000、20000、500000、75000 或 100000 个连续核苷酸，以及含有被破坏的牛λ基因的 ceels 和动物。

Seq ID No 31	<p>1 tgggtctat gccaccagc ttggtctctg atggtaactt gaggccocca tctcatggca 61 aagagggaac tggattgcag atgagggacc gtgggcagac atcagagga cacagaacct 121 tcaaggctgg ggaccagagt cagagggcca ggaaggctg gggacctgg gcttagggat 181 ccgggtcagg gactcggcaa aggtggaggg ctcccgaagg cctccatgg gggacctgc 241 agatcctgg ccggccagg acccaggaa agtgcaggg gaagacggg gaggagaagg 301 tgctgaact agaactggg aaagatatg gaggcagga tgcaggggac accgactct 361 gactctcag gacacactc tcagaagcag ggtccctga agaagcagag agacaggta 421 cagggcagga aacctccaga cccaagaaga ctacagagg aacctgagct cagatctgc 481 gatggggga ccgaggacag gcagacagg tccccotoga ccagcacaga gctccaagg 541 gacacagact tggagacca cggacgctt cgggcaaagg ctgcaacaca catgtagct 601 caaaatatac ctggactgac tcacaggagg ccagggagg cacatcatc actcagggga 661 cagactgcca gccccaggca gacccatca accgtcagac gggcaggcaa ggagagtgag 721 ggtcagatgt ctgtgtgga aaccaagaac caggagctt caggacagcg ctggcaggg 781 tccaggctca ggctttcca ggaagatgg gaggctctg agaaaaccc acccccttc 841 cctggcacag gccctctggc tcacagtgt gcctggactc ggggtcctgc tgggtctca 901 aaggatcctg tgccccctg tgacacagac tcaggggctc ccatgacgg caccagacct 961 ctgatttgg tcttctccc ctgcccact ttgcaggta gcccaagtc acaccctgg 1021 tcacctgtt cccgccctc aaggaggagc tcagcaccia caaggccacc ctggtgtgc 1081 tcatcagca ctctaccg gtagcgtga ccgtgtcta gaaggcagac ggcagacca 1141 tcaccgcaa cgtggagacc acccgggct ccaaacagag caacagcaag tacgcgcca 1201 gcagctact gagectgat ggcagcact ggaatcga aggagttac agctgcagg 1261 tcacgcaga ggggagcacc gtgacgaaga cagtgaagc tcagagtgt cttagggccc 1321 tgggccccca cccggaaag ttctaccctc ccacctggt tccccctagc ccttctct</p>
-----------------	---

1381	gcacacaatc agctcttaat aaaatgtcct caitgtcatt cagaaatgaa tgctctctgc
1441	tcattttgt tgatacattt ggtgccctga gtcacgttat ctcaaaagga acaaaatcct
1501	cttagccitt gggaatcagg agagagggtg gaagcttggg ggtttggga gggatgatt
1561	cactgtatc cagaatcccc cagagaacat tctggaacag gggatggggc cactgcagga
1621	gtggaagtct gtccaccctc cccatcagcc gccatgctc ctctctgtg tggaccgtgt
1681	ccagctctga tgggtcacgc aacacactct ggttgccacg ggcccagggc agtatctcg
1741	ctccctcac tgggtctca gcaatcacat ctggaagctg ctctgtctca agcggccctc
1801	tgccactta gatgatgacc cccctgaagt catgctgttt ttggctgaaa ccccaccctg
1861	gtgattccca gtcgtcacag ccaagactcc ccccgactcg accttccaa gggcactacc
1921	ctctgccct ccccagggc tcccctcac agcttcagg ggaccggcaa gcccacaacc
1981	ctgtcactc atctcacagt tcccagggt cgcctctcc ccactgcat ggcaggaggg
2041	tcccagctga ctctgaggtc tctgaccagc ccagctctgc tctgaccc cttaaaactc
2101	agcccaccac ggagcccagc accatctcag gtccaagtgg ccgttttgt tgatgggttc
2161	cgtagctca agcccagaat caggtaggg aggtctgtgc gtgtctatct ctgacctgg
2221	gtgtttctt aggagctcag aatgggagct gatacacgga taggctgtgc taggactcc
2281	cacgggacca cactgagca ccgttagaca cacacacaca cacacacaca cacacacaca
2341	cacacacgag tcaclacaaa cacggcatg ttggtggac gcatctctag gaccagaggc
2401	gcttcagaa tccgcatgg cctcactctg cggagaccac agctccatcc cctccgggt
2461	gaaaaccgtc tctcaccct cccaccggg tgaccocaa agctgtcac gaggagcccc
2521	cactctctc aggagaagt ccttgggacc cgtgtgaca cccagccgtc cctctgccc
2581	ctcccggcc tggagatgg cggcgccca ttcccaggg gtgaactcac aggacgggag
2641	gggtctctc cctcaccgc ccggagggtc aaaccgccc ttgaccagg agggggcgg
2701	acctggggt ccgagtgcag ctgcaggcgg gccccgggg gtggcggggc tggcggcagg
2761	gttatgtg gaggctgt cactgtcgt gtttctgg ttgaggacc cagctggcca
2821	tccgggtga gctccccct tccagcttc cggagtcagg agtgacaat gggtagattc
2881	ttgttttt ctaccatc tgggctgag gctctctca ccttagcct gtaacctcc
2941	ccctttagc ctgtccctc tgggtctct cactttctc tgaggagacg ttactgtc
3001	accacgcaa gccagagaa tatccagatg ggcaggcaa ttgggacgg caagctagt
3061	caccctcta ccttggctc cccgcccct ccgataatg tctgagctgc ctcttgat
3121	gttcacctt ctgagactg gaggcaaga acccctccc caaaaggag gagaccgac
3181	cccagtcag atgaactgc tctgaggga ccttggagt aagtgggtc tggcgggac
3241	ctgactcatt gcagactgat gcccaggca ggtgagagg tcatggcgc cgacaccagc
3301	agctgcagg agcacaggc gggggcaagt catgcagaca ggacaggacg tctgacctg
3361	aagatcaga gtgacagcc ggggggggg ccggagctcc cgagattagg gcttggctc
3421	taacgggac caggagggtc caggggcca cccagccct ctccctcac ccaatcaact
3481	tcaataaaa cgtctctat tcttcaaa aaacctgct ctctctcat gttttcct
3541	gccccctt taactgcaa cctctcagg attctggaac tgggtgggg nnnnnnnnn
3601	nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn
3661	nnnnnnnnn nnnnnnnnn nnnnnnnnn agcttatgt gtcggcagg ggttagtaag
3721	atcaaaagt ctaaaatga taaagccgc atgatatac agtttgata aaaaatgat
3781	ggaaaagta gaaaggacg gaggggggt aggcggaaga aagggggag aaggaaaaa
3841	aaataagaga gaggaacaa gaaaggagg ggggcccgt atgggggtg gatagaat
3901	aataatgga gtaagagta gcgggtgct gtaattcc ggggggata gaaaaaaa
3961	aaaaaaatg tgcgggtgg cggtagtat ggagattta taaattat gtgtggaata
4021	atgagcggg gtggcgggc aaggcgagag taaaagggg cgagagaaa aaattaggat
4081	ggaatatg ggtaaatt taaatagagg gtgatatg ttgattgag caagatata
4141	atagatgg tgggggaaa gagacaagg tgagcgcaa aacccctcc cgtatcatt
4201	gcctcttc cttaccac tcttcaac tcttttca gaacctga gcggtcagg
4261	ccgggctgg ggtgggala cccgggagg gctgcgct cctcttgc aggggggtc
4321	gaggagtgg agctgagca ggagactgc aggtggaga gatgctgt gactctctc
4381	cttttgaac tcaagtcac agtccagac ccactgaat ggcctaata ccatattt
4441	ctggggagag agttagagc gagcactga ggcgactca tctatctac agggcccca
4501	ctgcaggga cttgtgtgt gtcgtctc ttgctcagt gtgtccgact cttatgact
4561	tcatgactg taacctgca gctctctc tccgtggaat tctccagca agaactggt
4621	agtggtagc cactctac tccggggat ctctgacc caagaatca acctgagt
4681	cccgcattg aggcacttc ttctgtct gagccaccg ggaagccct taagtggag

4741	atctaaatag agtgttagg agtataagag aaaggaagga cgtctataca agatccttcg
4801	gttcctgtaa ctacgactcg agttaacaag cccgtgtgga gtgagtgcc agtaattatt
4861	gctaacctgt ttcttact cactgagcca ggtalccigt gagacggcat acitaccctc
4921	tctctgcat tctcgggat ggagctgtgc ggtggcctct aggactacca catcgaccag
4981	gtcagaccca gggacagagg atgtctgaga tgcactgaga agttgtcag cctaggtctt
5041	caccacaca gactgtgctg tctctacca cgtaatctt cctgtccaaa gaactggtta
5101	aacgctcctg aagcgtatc tggctgctt caaaaagtgc ctcttctt tataagttcc
5161	gccaatcctg gactttgccc caggccagtc tacttatt gtgggaaagg ttttttgg
5221	ctttttgt taaactctg cagaaattgc ttacatttt ggtgtgcaat ggctcactct
5281	tacggttcta gctgtattca aagggtgctc tttcttgt ttttaaagct tttgaacgt
5341	ggaccatttt taaagcttt attaacgtc taacatcgt tctggtttat tttgtgtg
5401	tctggccatg aggcctacgg gcttagctc cctaccagg gtccaacca catccttgc
5461	actggacggc aaggcttaa cctttgaacc accagagagc ttctgaaagg gctgtcttt
5521	ctccaatct cttgtctccc tgcctgctgg tagggattca gcaccctgc aatagccctg
5581	tctgtctta ggggctcagt agccttctg cctgggtgtg gagctgggtg tgaagagag
5641	cttcatggt ttggacaga cctacgactc agaggtaga ctccatctta gcgctgta
5701	gacctcttc caacaaccac ccccaccacc ctggaccact gatcaggaga gatgattctc
5761	tctctatca tcaacgtgt cagtcctaaa ctgcaccctg gcctgtcata gatgtagcag
5821	gtaagcaata aatattgtt gaatgtaag tgaattgaaa taacataagt gaaaaagaaa
5881	acacttaaaa acatgtgtt ttataattc acagtaaca tataatcatt gtgaaaaaa
5941	atcgaagag tggcgggggc caagtgaana ccaactccc tggatgtcc acccggccgg
6001	gtagcccccag gtaagaggig cggacacgga tggccctgta gacacagaga cacacgtca
6061	tatgtgggt cttgtctgt gacctctgg ggalgatgt attttcacga tgcattcaa
6121	acctctacc acaccattt tagagggtcg tctatgtaa atcagttcac tctttgtt
6181	tctgatttg aaagtgtcac attctcag aaatgagaag gaacaggcgc gcataaggaa
6241	gaaagtaaac actgtgccc gcttccaggg ggcaactcagc gtgtgtgtg gcacgtggc
6301	agcttttct ctgtgacgt catggcctt tcccaagggt gggctcagat aagaccgct
6361	ccatccctt gtcctgtcc cgtccctca cgttgaacc caccacggc agtctccga
6421	ggcctttgg gctgtggac gtaggctgt gtagcatgc tctgtgtgg gaccagggc
6481	tggcagcac gtttcccctg ggtcccgggc cagtgaggag ctccaagga gcagggtctg
6541	tggccaaaag ggcagtgcgt cccgagccca tggacaagg gatacattc ctgtgaagg
6601	gtggactgc gttccctgg gccccttgg agtcatgggc agtggggagg cctctgctca
6661	ccccgttcc caccatggc tcatgtgca gccaggagcg cctggggctg gacgccgag
6721	ggcggagccc ctccctgctg tctgacggg ctggtgacc ctgcccctc ctccctgggg
6781	ccctgtgac cgcggggccc accccggcca gtctgagat tccctgggg tccagccctc
6841	caggatccca ggaccagga tggcaaggat gttgaggagg cagctagggg gcagcatcg
6901	gcccagaccg gggctgggca gggctgggc gcaggcgggt ggggggtct gcaccccc
6961	acctgcnagc tgcnnnncn ttgnmncg tcttccctgn tctgtgtc tcccggcgg
7021	ggggccccc ctggtctgt ttgtccccc tcccgtccc tccccctt ttccgtct
7081	ctccccctt ttatctgccc cctgtgtgc gttttttc cgtccctt ttgttttt
7141	gtcttttt ttccccct ctctccctt gctctttt tcatctgct gttttctg
7201	tccctccct ctccccccg ctttttcc ctgtctctt ttgtgtct cctctctac
7261	ccccctgca gctatttt ttatatac catttcccc tagtatttg ccccgetta
7321	ctctccctc attttatt tcttcttt aactaaaac accgtgtgt tataagttt
7381	aaccttttt gcaccgcca caatgcaac ttacgcacg cccccctgt cagctcctt
7441	aaatacctt gctactgccc cccctcttg tataatacg cgtcacgtgg tcaaccatta
7501	taacctctc accacctac cacatttcc tcnnnnnnn nnnnnnnnn nnnnnnnnn
7561	nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn
7621	nnnnnnnnn nntgaaaa agaaaaggct gggcaggtt taatatggg ggttggag
7681	ggaatgaaa tgcattggag tggtycaac aaatgaaa gtctcaggag cgtctctcc
7741	ccatcaggag ctggaagaa gtggaagca agcaaggaat tctgtgatg gccagaggc
7801	agggcaggg agctgcaag actgccgct gttgtgact gncctctcc ggtgcat
7861	gttagcaggg agcattaca ctatgtctt gtttctaa ctattctta ctattgta
7921	gttcaaggat catgtgac tcttgaac ccagggact cagcccga ggctctctg
7981	tccatgggat ttcagggca agaactgtg aggtgtgac catttctc accatgggat
8041	ctcccagc cagaaatgga acccgatcg cctctgtgc atgggctg ctgcctaca

8101	ggcagatatt tgacgtctga gccaacaggg aggacagacg gtaattatac caaccattga
8161	aagaggaant acacactaat ctttatcaaa atctttcaaa cagtagagga gaaaggatac
8221	tccttagatt attccataaa gtfggaatta cgcttatcaa taaagacatt acaagaaaag
8281	aaagtgaagc cccaaatgcc ttataaatat acaagaaaaa aicctttaag atattagcca
8341	acttaataca caaaaaatgt atcaaaagtc caagtaacat tcaccccagg aatgcaagt
8401	tggttcagcc taagacaatc agtcatgagt ataccacgga aacaaatfaa agagaaaaa
8461	cattaaatct cacaaatggt gcagaaaaag atttggcaat atcgaacatc ttttcagac
8521	caaaggaaaa aaaagaaca aaacaccaga aaattctgtg tagaagaat atatctaac
8581	ccaatgaagg gcattatga aaaaccaca gcatacatca cactccatga gaaagactga
8641	aagctttccc cactgccatt gaactctgtc ctggaaatc tagtcacagc gacagaaaa
8701	gagaaagaaa taacggccgt ctaaactggt aggaagaaa caaagcgtct ctattctctg
8761	ggcgcataat acaatataga caaattcta aagtcacaaa aaattcttag agctcataat
8821	gaatccagaa atcgctcagg gctcaagatt cagatgcaaa aatcgtctgg gttttgatc
8881	accaacaaac aattccatta acaataaac caaggaatta atttaacta gaagagaaaa
8941	gaaactgitta cagagagtha taaaacattt ggtgatgaaa ttaaataaga gtaaatcata
9001	tagaaacacc gttcgtgtht tggagacctt atgtcataaa cgtggcaaca cagagacgcc
9061	tcacggggaa ccttgagcct cctctccaa acaggcctgc tcatcattc acaggtaacc
9121	tgagacccta aagcttgact ctgaggcaet ttgaggccat gaagagagca gtagctctc
9181	ccatgggacc gacagtcaag gccacaggaa tgaccacctg gacagatgac tccccgct
9241	catcagcagt cggctcagag tggccaccag ggggcagcag agagtgcctc aacactgcac
9301	ctggagatga ggcaacctgg gcatcaggtg cccatgcagg ggctggatac ccacacctc
9361	cacctgagga caggggcccg cttctgtgg tctgcccctc tcaggatgca cagactccac
9421	cctctcgtct gcatgaca gcctctgtcc ticttgagg acaagctcca ccttccccat
9481	cctccccag gggcctgggg ccaacagtgt tctctctgt ccactccagg aacacagagc
9541	caagagattt atttcttia attagaaaa ctatttgtat tctgcaatt cccagtaac
9601	tgaaggcaac ttaaaaaat gtatttctg gacttccctg gtgggccagt gcttagactc
9661	tgactccca gtcgatggg cctgggtica atccctgctc aggaactac atcccacagg
9721	ctgcaataa gatcctgcat gccaccogt gcaggcaag aaacaagtgt tgggtatgca
9781	tgatttcaac gtgaggtgtt tctataatt acagccagta tctgtctta cacttagica
9841	tictttgac cacatgatc gtcgatgcc cagaccacac acaggaatac tgaggcccag
9901	caccaccgg ctgccagaa cctcatgcc aagggtggac acttacagga cctcagggga
9961	ccttaagaa cgcctcgtc tcttgcaag ggagcaggtt taagcatgac tctgtccctc
10021	gggagctgtg tctggctgc gtcacacc tgtgtgtgg gcctgtgag ggtcacctc
10081	caggggccc ctgaggctcag aagaacctc ccttaaaagt tctagaggtg gactgaaac
10141	cagaccaca tgtgaactgc acccaaaac agtgaaggat gagacactc aagctctgg
10201	gtgaaatfaa ggccttccc ctgaaccagg atggagcaga ggaaggactt gcttccagg
10261	aaacctgac gtcaccagc tgaactggc cggggtcatg gcaggccca ggaactttg
10321	gtgcaagga ctcaaggctc ctgaaaata cagtccac cctgagccc tcagtgaaa
10381	ggcttctct cccaggagt ggcgaaggac ccagattggg gtggagctgt cccccagac
10441	cctgagacca gcaggtcag gagcagccc ggcctgaggg gagtgtgagg gacttccc
10501	ccgtctcaaa ccgtgtagc cctgggctga gcctctcca ccacggctgc aggcagccc
10561	caccccacc cccagccctg gctcggactg attgtatcc ccagcagca gggataaga
10621	caggcctggg aggagccctg cccagcctgg gtttggcag cagactcagg gcgcctccac
10681	catggcctgg acccctctct cctggcctc ctggtcact gcacaggtga gccccaggt
10741	ccaccaccc cagccagaa ctggggaca ggcctggccc tgaactgag ctcaagtggga
10801	tctgcccgtg agggcaggag gctcctgggg ctgctcagg gtggcagct gaggggctg
10861	aaatccccct ctgtctcac tctaggtca gcctgaggg ctgtcctgc cagggaaagg
10921	gggtctct tctcagag actccatcca ccaggacat gagccgggg tctgagact
10981	gacggggagg gtctcctgg gggccagaga atctttgca cttaactgc atcaggcagg
11041	gggttctgt tcttaggtc ttacgtcca gctacctc cttctctc ctgcaggcgc
11101	tgigtctcc tacgactga ctcactcacc cccggcatc atgtccccg gacagcggc
11161	caggatcacg tgitggggc ccagcgttg aggtganaat gttgagtgg accagcagaa
11221	gccaggccag gcctgtgcg tggctctta tggtagcat aaccgacca cgggggtccc
11281	tgaccagtic tctggccca actcaggaa catggcacc ctgccatca gcggggccc
11341	ggccaaggat gagggcact attactgca ctgtgggac agcagcagta acaatctca
11401	cagtgcaca ggcagcggg aaggagatg caaacccct gcctggccc cggggcccag

11461	cctcctcgga gcagctgcag gtcccgtga ggccegggtc cctctgtget cagggcctct
11521	gttcatcttg ctgagcagcg gcaagtgggc attggtcca agtcttgggg gcatacagc
11581	acccttgagc cagagggtta ggggttaggg ttgggttag gctgtcctga gtcctaggac
11641	agccgtgtcc cctgtccatg ctcagcttct ctcaggactg gtgggaagat tccagaacca
11701	ggcaggaaac cgtcagtcgc ttgtggccgc tgatcagge agccaticg gtcagcctac
11761	cggatcgtcc agcacigaga cccggggcct ccttgaggg caggaggtgg gactgcagcc
11821	cggccccac accgtcacc caaacctcg gagaaccgc cccccagga cgctgcccc
11881	tttcaacct gacatccga cattttcatic agaactctg caaaatatic acaccgtcc
11941	ttatgcaca ttccicagaa gctaaaagt atcatgctt gtaaccact ctcttaaat
12001	attcttctt aacgtccatc ttccctgctc cttagacgc ttctatccc acatgtctta
12061	ctgcctttgg tctgtctgtg tattttctt tttttttt ttatttga atattttg
12121	gttacaatgt tgaattgaa ttgtttctg tigtacaaca atgtgaatta gttatacatg
12181	tcctgaggag gggcggtgc gtgggtgca gaggggcag aggagctact ccacgttaa
12241	ggtcaggagg ggcggccgtg aggagatacc cctctgcca gtaagagaa acccaagtaa
12301	gacggtaggt gtgcgagag ggcatcagag ggcagacaca ctgaaacct aatcacagaa
12361	actagccaat gtatcacac ggaccacagc ctggtctaac tcatgaaac taagccatgc
12421	ccatggggcc aaccaagatg ggcgggtcat gtgccatgg ggccaacca gatggcgggg
12481	tcatggtgaa gaggctgat ggaatgtgt ccactggaga agggaaaggc aaaccactc
12541	agtattcttg ccttgagagc cccatgaaca gtatgaaaag gcaaatgat aggatactga
12601	aagaggaaat ccccaggta gtaggtgcc aatgtctac tggagatcag tggagaaata
12661	actccagaaa gaatgaagg atggagccaa agcaaaaaca ataccagat gtgtagtga
12721	ctggtagatg aagcaaggc caatgatga aagagcaata ttcatagga acctggaatg
12781	ttaagtccaa gannnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn
12841	nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnagaattt
12901	gagcattact ttactagct gtgagacgag tgcaatttg cgttagttg agcattctt
12961	ggcattgctt tctttgga ttggaatgaa aactgacctg ttccaggcct gtggccactg
13021	ctgagtttcc caaattgct ggcgtatga gtcatcact ttaacagcat catctttag
13081	gatttgaat agctcaactg gaattctatc actttagcta atccattca ttgctttg
13141	tttagtgat gcttccaaag gccccctgg ctttacttc ctgagctct ggtctgtgtg
13201	agtatcaca ccgctgtgat tctctggct atgaagctt tttgtatg tcttcttag
13261	gaacagat atgatctcc atcctgcat ctggtatat ctgagagaagc actgactccc
13321	ttcatggtga cgtcagatcc tcatgactaa caaatggcct ttgtaagat gagtgcctca
13381	tggtattgag cccccctg accaagaact tatgactgac cccccact gcccagggtg
13441	cctctcgaag cgtctgagat gccctctcc aggtctcact cctcatttg cccccataa
13501	aacttaact gcagcttcc agctgtgcat ctgttttag ttgacagtac aaatataatg
13561	gaaaatttaa attaaatata atctatggg agaattccaa acatcttatg agggagagag
13621	agggagagaa aggaagaag aagaagcagg aggaggagga gagttagaa acagggggag
13681	ggcggcagg agacagagg gaggacacc agggaaagg gaggaaaggc agtgcagta
13741	gagagaggcc agagttcatic agagctgga ctgcagccc aatcccagc gtgtgtccc
13801	aagcagggga gagctgagc caggcggaga cagagctgtg tctccagtc tegtggcgt
13861	gacctggagc tgtgtgtca gccccctga ccccagcctg gccctgctg ttgtcggagg
13921	cagtatctt ggacacagt tctgagctc tgtctgaaat cctgtggag ggcactca
13981	ggcggacct cgcctggccc cacctggatc tgcaggcca gcccagatg gggcttctg
14041	cctggaactg agcagctgga gggcgctct cccccagca gtggagcggc cccagggcg
14101	ctcagagctg ccggggggac acagagctg tctgagacce aggcctctc tccaggggt
14161	cccctaagg tcttctggc caggctcaga gccggatga gcacagctt gattcagact
14221	ttcagactg gtgctgcat cctggggac agaggctgg gtcttaacct ggggtcaga
14281	ggcagagac ggagcccagc tgaocctgg gactggcct cctctgtgt cccccctgg
14341	cagtacagc tccccggac gtgactctg aggagacag ctggggcctg gctgtcagga
14401	gggggttca gaggccacac tcagaggagg agacctggc ctgctgggt tgtactgag
14461	ttttgggtt cctctaggag actctggccc tgcaggcct gcaaggtcat cctatgtga
14521	gcaggactc acaagattga tgaactgaat cctctaggag aggtgtggtt gtgaggggc
14581	agcattctag aaccaacagc gtgtcaggt agctggacc ggtctagtg gggcgggca
14641	ggcactcag ggccactag ggtctgggg gattcaatg tccccagc actgggtct
14701	ccatcagaat cccagactc acaaggcagt ttggggatg aggtcaggac gtgagggcca
14761	cagagaggtg gtgatggct agacaagtcc ttacagaga gactccagg gccatgata

14821	agatggatgg gtctgtattg tcagttccc cacatcaaca ccgtggtccc gccagcccat
14881	aatgctctgt ggatgcccc gtgcagagcc tacctggagg cccgggaggc gggggccgct
14941	gggggctcag ctccgggga accgggccc gctgtcccct gctgtgtcca cagtctccc
15001	gggggtggag gagagtgtga gcaggacagg aggggtttgt tctacttcc ctggctgtct
15061	gtgtcactgg gaacattgta actgccactg gccacgcaca gacagtaata gtcggcttca
15121	tcctggcac ggaccccact gatggtcaag atggctgtt tgcggagct ggagccagag
15181	aactggtcag ggaacctga gcgcgccta ctgtcttat aaatgaccag cttaggggcc
15241	tggcccggct tctgtggtt ccactgagta taitgtcat ccagcagctc ccccagcag
15301	gtgatcttgg ccgtctgtcc caaggccact gacctgaag tcaactgtgt cagttcatag
15361	gagaccacgg agcctggaag agaggaggga gaggggatga gaaggaagga ctcttcccc
15421	aagtgagaag ggccttccc ctgagggtgt gtctgggctg agctctgggt ttgaggcagg
15481	ctcagctctg agtctgagg gaccaggccc ggggtgctgt gctggggggc cgcacctgtg
15541	cagagagtga ggaggggcag caggagagg gtcaggcca tggtgacgt gccccagct
15601	ctgctctga gcccccagca gtctgggct ctctgagacc cttattccc tctcagagct
15661	tgcaggggc cagtgggt ttgggttat gcaattcac cccccgggg cccctcactc
15721	agaggcgggg tcaccacacc atcagccctg tctgtccca gcttctctt cggtcttca
15781	ctgtgcaca tcagactgt cctcaggac tgaggtcact gtcacctcc ctgtgtctga
15841	ccacatgacc actgtccca gcccccctgc ctgtgtctt gggctccca gttggggctg
15901	cagcttgcca gcgtctggc cgtggactgc ggcattggt cctggggtt actgtgtatg
15961	tgacctcag aggtgtcac tagtctgag gggatggct gtcagctct gacttctgc
16021	caagcgtgc tccctggaca cctgtggac cacaggctg gtcacctga agccccgct
16081	ggcagccca gctctgacc tctgtctct ggcgcgctc tctgtcccc tctgtgtac
16141	cccatgtct gctctagca gactgtgat tctcagcat aactgattac tcttccagt
16201	actttcatgt cctgtgacg gctgagta gcatttca cactagaaa ccacagctct
16261	cctgtgtaa gtatcacac tctctctgt gggactttt taaaagattc tgcagccagg
16321	agtcattggt ggtcttagt gagaatgct ggaicagaga gacctgataa ccgatgtga
16381	gaggggaacc tgaagatct tcagttcagt tcaattcagt caticagttg tctccagctg
16441	ttgggatcc catgactgc cacagccag tccctcctgt ccatcaccaa ctctgaagc
16501	ttgtcaaac tcatgcat caagtggag atgcttca accatctat cctctgtat
16561	ctcctccc tccccttc aacttccc agcattagg tctttcctt gactcagtc
16621	tgcacatg gttgccaagt ttggagtt cagttcagc atcagctct tcaatgaata
16681	gtaaggactg atttctta gtaggactg gttgatac ctgacgtc aaggactct
16741	caagagtct ctcaaacct gcagttaaa gccatcaat ctctggtct cagcttctt
16801	ttgtgtaaa ctctcact catacatgac taccgaaat acattagtc tglagaacca
16861	gttgggct tcccagtg ctctagtgt aaagaatag cctgccaact cagaagatgt
16921	aagagatg gttcaatct tgggtcggga agatcccctg gagaaggcca tgaacacca
16981	ctcagatatt ttgcttga gaatccatg gacagagaag cctggtggac tgcagtcct
17041	ggagctcac agagtcagac acgactgaag caacttagct acttggaaa gagcatgac
17101	gaagctgt aaaaaacag tcaagaagtc ttgttttg aagtttact gagaagttg
17161	atgactgt ccaacactc ctctcagttg aaaagatcag aagcgttga tcaaatggtg
17221	gtcaatact tggatgct ccaacaggt atactgcag atggaatga aggcagttt
17281	tggggtaact ggaggacaag atgagatcat acacttggaa cactgtctgg catcaaggc
17341	gtgtacagta aacattagct gttaitgca aataaattc agcttgaat acccaaatca
17401	gatggactc taaagccc tgaagtgtaa aatcagggt gtgcagcca aacttccat
17461	ttgactcatt atgattcca tgtcacaaga ctgaaagtc acttctct cagcagaaga
17521	gaaggtagaa caattaac ttttttga gtgtcaagg aattttgtt acactgtaa
17581	gtcagtga aaatgaagc tttcattg tggaaaat taaatgta aattgaaat
17641	tttaaaatt atctgggt agttttgt ttocagtgt catgcatgga tgtgaggt
17701	ggactataa gaaagctgag cgtgaagaa ttaattctt tgaactgtg cactggaga
17761	gactcttag agtcccctg tctgcaagga gatcaacca gtccatccta aaggaaatca
17821	gtcctgaata ttactggaa gactgatgc tgaagctgaa actccaatc ttggccacc
17881	tgatgtga aactgacta tatgaaaga ctgatgct gggaaagatt gaagtggtga
17941	ggagaagggg acgacagagg atgagatgct gaaatggcat caccactct atggacatga
18001	gtctgaataa gctctggag ttgttagtg acaggaggc cctggagtgc tgcagtcct
18061	gggattgcaa agagttggac atgactagt gactgaactg aactgattt ggtaacagat
18121	atgagaatta tataattaa atctaaact ttggtattc ttctttggc gtttcaaaa

18181	gagctgccc tctgtaac tataaac cttttgaga attactaat tgataatgt
18241	cacaagttat ccaatttc cttacttta gttgcagta taagaatcc catttgatt
18301	atcatgriat agtatctgca actctaatag ttcagtctg acaaaatitt attttatta
18361	aaaaattgg catacagtaa aatttcaaac aaatacaat tctcccttc agtttaaaa
18421	acaaaacaaa acaaaagtaa tattagtaa aaaaatccgg gaagaatcca agcatttaa
18481	attgcatcac atttctatgc tagacaagct gatataaagt tataattaat aaaggattgg
18541	actatnaaac tctttacata tgaggtaaca tggctctcta gcaaacatt taaaaatg
18601	ttgtgggtaa attattgtg tccftaaaga aataaaaaga cataagcgtg agcaattgn
18661	nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn
18721	nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnna aaatggataa ggggggagga
18781	catgggtagg ggagcgcgat ggaggaagta aggtggctga gggagtggg gggggaataa
18841	gtgggtaaaa gggaaagcgg cggaaggagg gggaaagcagg agagagggtt gggcgtcaga
18901	tcggggggag gggatgagg gagagggaat ggtagacggg ggggtggag cataaggaa
18961	aagatgggg ggggaaaagt tagaagaaga atgagggat aggcggaaag ggaagagaaa
19021	tgggagaaga acagaaaaat agggggagg ggggcgtaaa gaggggggg gaggcaggt
19081	gtggagatga cagatacggg gaatgcccc gtataaaga gtatatggcg tgggcgaga
19141	aggctgcat cctgtggag gggggacgag gagaaccctt cggctatag ggaggattcg
19201	gggggatcgt tcgggaaggc agtcagcaca gcaccacca aggggtcagg gatgatctg
19261	gggtccaaa gaagaggccc aatcccgtt cttggcagca aggagccctg gagactggg
19321	agtgtccagg aactgacc aggggttca ggaaccaga agtgtgctg tgaagtgtg
19381	ttttgtggg ggcaggtcc agagcttga gcaaaaagc ggccatggcc tgtggaggc
19441	caaccacgt gatctttt aaaaagttt tttttgatg tggaccatt taaagtctt
19501	cattgaatt gctacaat ttttctgt ttatgctctg gtttctcg ctgcaaggtt
19561	tgtgtatcg tctctctca accaggactg aaccacagc cctctactg gaaggcgaag
19621	tctaaccca gatgccagg aacgtccctt cctcactga tctaatcaa gacctcatt
19681	aaggaaaaac cgagatcaa agtccccca ggaggactcg gtgggagga gagagccaag
19741	caticagcac tcagtcagc acggcgcctt cctgtccag ggcgaggctt cggccgaag
19801	accaccggag accctgtcgg atcaccagt aggatttga ggaattcaa ctactttt
19861	aaatctgt ctcaaggctg ttacaaggc actttaccag taacttaaaa gttgaaagg
19921	acttcccagg cggcaattgc ggtgaagaac ccgcccctg gtttaggag acataagaa
19981	tggtgtag atcccgtgt caggaggatt cccctgaga aggaatggc aaccactcc
20041	agtattctg cctgaaagc ctcacggaca gaggaggctg gcgggctaca gtccacggg
20101	tcgcacaga ctgaatcagc ttacttcaa gttgagacag gaagaggcag tgaactgtg
20161	caaaacacc caccatgct cccaggggac ctgcagcgt ctggttcatg agctgtcta
20221	acaaaaatca acccaacgag aggccagac agagggaagc tgagtatc aaacacggg
20281	atgatgtga ggagataatc caggaaagga cctgccaagc ccatgacaga ccggtgtct
20341	gtctagggc cgtctggca gagcagtga gggcoctcc agaccgcc agctccagc
20401	ccgctgggg gctacagggt gggctgagc tcaaggact ctgctgtgag cccacgtca
20461	aggaggatca cctgtttgt tttctagt tcttataaa tagccttat ggtcctgtt
20521	ctttgtttt aaaaatacaa ctgtctccg taacaacgt gaaaaaaac aaacaggagg
20581	aaaacaagc agcccggca ttccaccgg aagagccgc tctaacatt tgacgggtg
20641	ccttctatt taaccctgt ttattgtaa actgtaaaa ccatcata aataaata
20701	aggctctgt gaagttaaa aagtaagcat ggcggggcg atggctgtc cacaccgtg
20761	acgctggtt caaaacgta aatttaggg accccctgtt ggtccagtg gtgagattt
20821	gcttccattg caggagcctt ggtttgatc cctggttgg gaactaagat cccatgct
20881	gtatggagt gccaaaaa aattttgta aatggtagt tttagtgac gtaatttc
20941	cattgatga cttcacagc tcagatgag ccaggccctt aggaagccc agtccaccg
21001	tctttactt tcttagag ttatgctt tctgttctg oecttaaac caocatgtt
21061	caaccctc tgatttga cttataata aagttaggct gttttcag aaacttct
21121	cagtattct taataacta aatgaaaga attgaaaa agacagaca ctgtatctg
21181	cataactgaa tcaattgtt gtacacctga aactcagtg cagccgtca gtcgtctcg
21241	accctgcag cccacggact gcagcagcg ggttccctg cccatcaca actccggag
21301	ttactcaaa cacatgctc tgactcggg gatgcccgc aaccgtctca tctctgtc
21361	tccctctc ctcccctt caactttc cagcatcagg gtttttcaa atgagtcagt
21421	tcttcacac agtgggcag agtattggg ttccagctt agcatcagc cttccaaga
21481	cccccatc ctgaagctaa cacagtcta atccactgt ctgcaacatg aaagaaaaac

21541	acattittta agtttaggct ggtgtgtot tctctctc aacctgegt ctgacccac
21601	ccacctgcc cagcactgca ttcccgtgg acaggaggcc cctgcccca cagtgcgtg
21661	ccggccggtc actgccgagc agacctgcc gccagagtg gggcccctgg cactggggac
21721	aaggcagggg cctccagg gccggctact gtccactgt cctactggt ttgtttca
21781	aagtgaggc agcgtaat ticcctgatt ataaaagaa gtacacaggt tctccaaa
21841	taaacaggg gaaaagtata aagaatgaa gtcccagca cagcctggag atcacgccg
21901	gtcacctgg ggtgtctc caggctggac ctacattc acgacacat cagaaggctg
21961	cgagatctac ccagaaggct gggtagatg gggataggc agtgacaaa agtagacaga
22021	gagatatac gacagatgat ggatagac agctaacac accgagcag gggacagacg
22081	gatggaagac accatcctt gtactgacc acacaccac atgggtgtg tgagccggt
22141	gtcactctg tgaacctgt gctctcaca caccagctg gccctccag ccccagctc
22201	ccacacagca gactcccgc tcatccca ggcaggaat ccaccacaa ctgggggtga
22261	ccctcccgc aggaaggct tctgtctaa ggccttga gcaagtaca gacctctc
22321	tgggaagaca gcgcaaac gcctacccc cagagcccag gaggaccct ggtctctag
22381	gaaggagca cgggctgg acggggagc gccccagac gctgcccac acctgtcca
22441	cctactcct gctctctc gaggcgggc gcagagagg gccctgagc ctctccag
22501	tctgggag caccactgg gcctgaaca ggcagaagc cccctctca aggtgtccc
22561	agaccctc cctccactc cgggtgct gtctctggc agcaggagc cccagtga
22621	agacacagt ccaggctgt atctggccc ctggctctc tggcagtg ggggtggg
22681	gtcctggga ggcatgagt gctggggc gggctgtga aagcacctc aggtcagtg
22741	gctgtgtg gggctctgc agtccgcac gggtagact gtcccaggc acaggagcc
22801	tgtcagtg tccaagag caggccaaa ggaagggtt cgggcccctc tggctctca
22861	gtctctgag ccggggacc cagctggcc ttgtaggg ggcgattga ggtacacg
22921	atccaaaaga aaacacac ctacaggga agagtctga ggaggagaga gctacacaga
22981	gggtctcac actcggaca ctgctggag tctgagact cagtgccgg gcacagtga
23041	cgaggagag acggaacc caaggacac ggcgccgat ggccagagc acacgacgt
23101	cccatgagg ccggctctc agcgcaggg gactctca ttaaggctc tctgtaata
23161	gtgaggaga ctggcccct ggtgggaa acttagcca gaagaacgc tgcctggcc
23221	caaaggata nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn
23281	nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn tgcctttgc
23341	caccaggag ggaggagc tggatctg gttgcttg gtttaagg atcccccac
23401	tccctttta gccactcct gtcctggca ttcttaaga ctggagctc caaagattg
23461	gacacactga gcagtgaa tgcactgagc taagaaaag tctttgatt cctccaaaca
23521	aaacacact gcttggga ctctcttg ttgttaca aatgtctgt cctctgttc
23581	tcctggccag ctctgggtg tcatctgac ctgacgaag caaagggagc ctggaccctc
23641	aaaatctga ggaccagca cccctcatt acacctct tcccgcgga cggggcact
23701	gttccgct ctggcgtat gttaagca cgggtgata ctggggagc ttactctgt
23761	tcctttct ctgggtgac accaccatc gcacactct gctgaatg gaacatttg
23821	gtgattgat gtggccaga ctccccaac gaatgtact tcaagggtt ttctcttt
23881	tatatttgc tttgtgat agacacagga tccatcag tctatagt gagaagtaa
23941	aaaccactc agccttagt gcatggag ctatagtaa gatagcact tagccgaaa
24001	tggaaattc agccagaat taaaagcgt gtctggaag gagaagagg actcagccc
24061	gagcacact cccacgct gacccagg ctctgacag tctaccgco ggggtctca
24121	tgggacagc catgcagcc acgatccgt tgaagatt ctgcttcc catcacatg
24181	gcaattgac gcttctct tctctaca tggattta ctttatcc agacatttg
24241	gttctctc tgaattgc caattgaca gatcttaca gtattcta accacataga
24301	attcggcag ggggtggg gacaggga gggggggg agagtggg ggggggctg
24361	caccgagc catctgggt ctagctccc tgcgggat agacctgt cccctcagt
24421	gacagcac agtctctc tctgaactgc caggagct cctgcaatt acttaatga
24481	aggcatctaa ttaggatt tgggtgaca tttaattt aagtgtga gcagtata
24541	tagtcatat cattiatag ttctgtat ttactgct aaagggtt tgggtttct
24601	ttttttta aaagctaaa tctgtttt aatccatgg aatacaaaa aaaaaagtct
24661	gtagaatatt taaagagt aggctttgt tggaaatg agccttgc tccactgac
24721	cgaacgtaa taacattgt agaagagc cagagtga ggtacctt ttattgagt
24781	gacatgac caccatgc gtgattat ggctggagt tagagacgg ccatgtgg
24841	ctaacctct tattctgt ctacgctt gataataat cagaagctt ctctgaag

24901	agtgggtca gctgtcagac tcttaggtgt ctacctcgag cagggtggg attaatgca
24961	gcagccagta gatacgggat ggggcaagag gtcacctgt cctttgtg ctgctggag
25021	agaggctgt cctgggtcca gtggggccaa agctgtgact ttgtaccac aggatgttc
25081	tgacctgcc ttgggtccc tgagggtgga gggacagcag ggtctcccc gttcctggc
25141	cggagaagga cccccacc ctgtctct gacatcccc caggacttgc ccggagtag
25201	gttttcagg atgggcatcc gggcccacc ctgactcctg gagctggccg gctagagctt
25261	gctcagaat gaggcctgg ccatgggce cctgaaggag ctgcccgtca agctctccc
25321	gaggctgtt acggcggcct ttgccaggag gcacacccat gccgtgaagg cgtatgtgca
25381	ggcctggccc ttcccctacc tcccgatggg ggccctgatg aaggactacc agcctcatc
25441	ggagaccttc caggctgtac ttgatggcct ggacctcctg ctgtgagg aggtccggc
25501	taggtaaggt cgacctggca gactggggg gcctggggtg tgagcaagat gcagccaggc
25561	caggaagatg aggggtcacc tgggaacagg cgttgggtgt acaggactgg ttgaggctca
25621	gaggggacaa aaggcacgtg ggccctcccc ccagtgtccc ttaagtggg aaccaaggg
25681	gccccggaag ccggaggagc tctgtgtgt ggagtgcaga gccctcggc ggtcctgat
25741	cccgtcgac tctgcacagc tcagcgtgtg ccccgggcc cgttagcgg tggaagctgc
25801	aggtctgga ctgtcccg aacgccacc agggacttct ggacctgtg gtcggcctc
25861	aaggccagcg tgtctcact gctggagccc ggtcagccc agccatgca gaagaggagc
25921	agggtagagg gtccagggg tggggctga agcctgtgc gggcccttg gaggtgctg
25981	tgacctgtg cctcaaggag gacacgtgg acgagacct ctgctacctg ctgaagaagg
26041	ccaagcagag gaggagcctg ctgcacctgc gctgccagaa gctgaggatc ttgccatgc
26101	ccatgcagag catcaggagg atcctgagge tgggtcagct ggactccatc caggacctgg
26161	aggtgaactg cacctggaag ctggctgggc cggatgggca acctgcggc ctgctgtgt
26221	ctgtcatgc cctgttccc cgcaccgcc ccgaccggga ggagcactgc gttggccagc
26281	tcaccgcca gtctcagc ctgcccacc tgcaggagct ctacctggc tccatctct
26341	tctcaaggg cccgtctac caggctcga ggtgaggcgt gggccagct ccaagacca
26401	gagcaggcct ctctgttc gtcccgtg gggacattgc cagggtgcc ggcactcgg
26461	aagctctac gatgccacc ctctgacct gggcatctg tcaggctact tccctggtta
26521	gggtcagagg cgtggcctag gttaatgct gtaaaaggg actcctttct gggagtcgc
26581	atagtgggg ctgtgtgta tgccttggg aattcttcc gagagatga tctctagct
26641	gagataatga cagataacta agcgagaagg accgtccatc aggtgtgagg ttgaagtc
26701	aaagctctgt ctctccctc cactgccc tctgtcctg agctgttta ggctccagg
26761	gagctgggg aagtgggta tctggagat gacaagaagg gacaggagg ggaanaatgt
26821	ggctcctaag cagtcagag aagagaaaa gtaataaag cattattgt aaagtggct
26881	cagctcttt aagccaaat tataattata atttctct aagactctg aatacatag
26941	aaatctcag taacaggta ttgctctgc ttgaacacag tgataaaagc tggaggatg
27001	cagcctaact tgtctgtg atgagttgt atgattccc ttttggcag ctgcaaac
27061	caagcattag gaataaat gtctactgag aaccccgag aagaaagaa agaaaaaaa
27121	aaagaatgt aggtgtgt ggcagggtt tggcccctga atactggg gatgtcacc
27181	caggatcac gttactgc tgggacccc agcccctgt cactgcac cagcctgctg
27241	ttgaattcc cggatcnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn
27301	nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnncaat
27361	tcgagctcg taccacaag gtccgtctg tcaaggctat ggttttcca gtgtcatgt
27421	atgatgtga gagtggact gtaagaag ctgagtgcca aagaattatt ctttgtact
27481	gggtgttga gaagactct gagatccct tgaactgcaa ggagatcaa ccagtcggt
27541	ctaaggaga tcagtctga atgtctatg gaaggactga tctgaagct gaaactcaa
27601	tactttgcc acctgactg aagattgac tcatggaaa agaccatgat gctgagagga
27661	atgggggca gtaggagaag gggacagcag aggatgatg gctggatg catoccaa
27721	tcgatngac atgatttg taaactca ggattgtg atgacttg aggcctgtg
27781	tgtgggatt catgggtcg cagatcga catgactgag cactgaact gaactgaact
27841	gagctgaaga gctacctgt accagactc ctgagctct cctcaggcc tggctgta
27901	ggccccagg taccgtct cctcttca tccatctt taccagagg ctgggagtg
27961	ggtgagtga gttcttgt atctagaat tctgatgc accctcagag tgcaattag
28021	ctccagaga ctgactcca agattcatt tttctttt ctctttat atactacct
28081	ctctgagca gagacctat gtcaggaga agggactct gctcttca gccctttgt
28141	ctccaagac ccacagggg aggtgctct gcttactga gccggaagt tcaattgct
28201	atgcttcca gaaacacccc cccccaga gacccccaga aataagtgga acagcactt

28261	gttcccaga caagtgggac acacgttatg aaccacctca gtgattaaaa tagtaacctc
28321	tggtatgtg tattactgg agaaggaaac ggcaacctac tccactatic ctgcctagaa
28381	aattccatgg gagagaagcc aggcaggcta cagtcacagg ggtcacagag actgaacata
28441	cacaagcaca tggagtgta tttgacgta ttttaaatt tgtcagttc aacatggagt
28501	acaagaaitc aaatcgtgaa gtcaatgac caagaaacca gaagaaatca ctgtgtgtg
28561	atctctgtgg aggtaacatg ggtacctgtg ctctgacctc cacagcctct ggctctctct
28621	ctacatgtac atacacatat atttccatgt atgtatgtat tggagagatt tcacatacgt
28681	ctcaccagtc cacagcccc gcgttccctg atgcccagaa catctgtgat agctgtgagt
28741	attgtacca gataagatct tccaggttcc tgcactcaca ttggtatca ggctctctctg
28801	atccagcatt tctcagctaa gattcctgt gactcctggc tgcagaatct tctgcaaaag
28861	tcccacagag aggagtgta tcactgtaca caggaggggc gtggttctct agtgtgagaa
28921	aaagtaactc agcccgtcac aggcagctga atgtacctga gacagtaatc agtatactg
28981	agaaatcaca gctctgctag aggcagcaca tggggtagcc agcagggggc agcagagcac
29041	ggccaggagc cgcaggtcag aggctgggct gcccaagcgg gccttcaggg gaaccagccc
29101	tgcgggtcca cagggtcca gggagcagcg ctggcagga agtcaggacc ggacagggca
29161	tccccagcagg actagtacc acctctgagg gtcacatcca cagtgaacc cagagacca
29221	tgcctcagtc cacggccagg acgctgccag gctgaccgcc ccactgggga gtccagggga
29281	gaccacagcc cggggggctt gggacagta tcatgtgttc agacacagag aagggtgacg
29341	tgacctcagt ccttgaggac aagtctgatg tgcagactg agaagccgag gaggagctg
29401	gggacagaca gggctgatgg tgggtgacc ccgctctca gtgagggggc cccgggggtg
29461	aattgcata aacccaagcc ctactgccc ccacaaagct ctgagagggg ataaaggggc
29521	tgggagacc cagcactgct ggggctcag aggcagagct cggggcggct ccaccatggc
29581	ctgggcccct ctctactgc cctctctac tctctgcga ggtgcccgc cccagcctg
29641	gtcccaagt gaccaggcct caggctggcc tctcagctca gcacaggggc tctgacagg
29701	aatggggccc gctgggagga gacgctctc ccacactccc ctctctctc tctctctag
29761	gtaccctggc ttctctcag ctgactcagc cgctgctgt tccgtgtcc ttggacaga
29821	cgccagcat cacctgccag ggagacgact tagaaagcta ttatgtcac tggaccagc
29881	agaagccaag ccaggccccc tgtgtgtgtc attatgagt ctagtgagag acctcaggg
29941	atcctgacc ggttctctgg ctccagctca gggaaacagc ccacctgac catcagggg
30001	gcccagactg aggcagagcc cgaactattc tctcagctat atgacagcag cgggtactc
30061	cacagtaca cagacagac gggaggtgag acacaaacct tccagtcctg ctacgctct
30121	ctccagccc cgggaggact ggggcacag caggacagg cctggcccgg tccccggg
30181	gtgagcccc caggggccc cgcctcccgg cctccagcc aggcctgca caggggcgtt
30241	agcagtggac gatgggctgg caggcccgc tgtgtcggg tctgggctgt ggagtacct
30301	ggagaacgga ggcctggatg aggactaaca gaggacaga gactcagtc taatggccc
30361	tgggtgcca tgtatgctg gctggacct cagcagccaa aatctctgg atgacctca
30421	gaactcca gatccagtc cagctggctt tagaaagct taggagtgta acaagtggg
30481	tgaggctac catggtgacc tggaccagaa ctctgagac ccatggcacc ccactcagt
30541	actctccct ggaaatccc atggacggag ggcctggaa ggctcagcc catgggctg
30601	ctaagagtc gacacgactg agcagctca ctctccctt tcacttcat gcattggaga
30661	aggaaatggc aaccagtc agtctcctg cctggaaat cccagggaca ggggagcctg
30721	gtgggctgcc atccatggg ccacacagag tcaacacga ctgaagcaac ttacagcag
30781	cagcagcagc ccaataaac tcagcttaag taatggcctc taatggaco ctattgcaa
30841	ataaggtcca ctgcgtgca ctctgttag gactcagtt cctgattgtg gagggtccc
30901	acaagacgtg tgtgtatatt ggtgtgccg gaaaacagtg tcaatgtgag catcccagc
30961	ctacacct cctactcca ctattccatt gctctgag gtattaagca taaaggtaa
31021	gggtctatt agatggaaga ggagtgaata ctctctgtg cttaacat accaagtacc
31081	atcaagttc ttctatta ttaactgtg tttatcag aaatgtcta tgtagaagca
31141	tccggagat agccatgtt acagacggg aagetgagc atgaagtct cagcaoctg
31201	ttcacgta gacctgaac gggcagagc cggcagcaa caagttcct ctcccagc
31261	gcccgtctt caccgcttc ctatggctt cactgtgt tcttaacta agctctccc
31321	aacctgtgg agacaggatt agagactta ggagaaaaga ccaggaacat cccacaccg
31381	accgagtg gaccataaga caagctttg taaggacaga accagcaggt gctctcagc
31441	agccagggag agacctgca ccaaaaaca tattgtaga tctgacctt ggactctga
31501	cctccagaaa tgtgaaaag aacctgtgg ggttaatca actcaccgtt gttattgtt
31561	tatgactgcc tgagtaaga aggagtggg aacctgtg ttaggtgtt tatgaaat

31621	aagctctgtt tctctgaat aaattccaa gggataatt cctaggtgtt agggtaactg
31681	ccacaaatct aggcagctta ttaaaaaaca aagatatcac ttgccagca aaggctcata
31741	tagtcaaat atggtttta tagtagcat gtatggatgt aaaagtggga tcataaagaa
31801	ggctgagcac cagagaattg atccctcaa atcgtgggtc tggagaagac tctgagagt
31861	ccctggaca gcaaggagat ccaaccagtc aatcctaag gaaatgaact gtgaatattc
31921	actggaagga ctgatctga agctgaagat ccaatacttt gccacctga tgcgaagagt
31981	tgactcattg gaaaagacc tgatgctga aagctgagg gcagagagag aagaggggcg
32041	cagaggatga gacgggtgga tggcatcact gactcaatgg acatgagtt gagccaactc
32101	tgggagacag tgaaggatag ggaaggctgg cgtggtacag tgcattcggc cacaaagagt
32161	ctgacacatc ttatgtactc aacaacgaca gcaacacagg caacacgc ttatgtgat
32221	aagcggcaga actgtttcc aggggtccgn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn
32281	nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn
32341	nnnnnnnnng tacgattca gctcggacc tgacatttg agtcacgta tgagcagctg
32401	tttcggctc ttcagggatt gggagcatt tctgttggg ttgctcag ataattagt
32461	tcagactag gttcttctt tccaggccac.gagcgacatg tticagggtg agatgacgtg
32521	gtgggggatg ggcggccaag cccccactgg ggggggaggg attctgtgt gggcaggagt
32581	tggcagatc cctgaactga tgacctgca tccaggtgac aagaaccggg gcatattatt
32641	cctctgctt ctcattgat gtctcgggtt ctcatgatg aaaacatag acaatacagg
32701	ggagttagat ttggcgggc acaactctgg gtgggggacc cgttggcatt gtgccagca
32761	ggccatcaa gatgaggcg acttgggtgg tcccctctc ccttgggctc ttatgttcc
32821	cctcatgaa atggatcag gcagcagca tggacacacc cgacctgac tctctcacc
32881	tctctctg tgatttggg tgggatacc aggcataag acctggggcg ggggacatc
32941	actctctgc agcaggagc cgcagagtc ctccgtccat gaggacttc tccctgggt
33001	gacctgagg actgctggg gctgaagctg gaggcacagg cgggtgca ggcagggtc
33061	ctgaggacga cagagccagt ggggctgca cctgagcag atggccctc gcccggggc
33121	ctgacttgt ggtccagct gcaggttcg tcaagtgag cactacgta tgggggaggo
33181	gcccctggca gggatgggg gtgctgactc ctccagatt ccgacctct ggggacatc
33241	tggccacatc taagcctgg caagactgg gtccatcagt ctaactctc tctgaagtc
33301	caatggactc tctcatgag gcagtcactg gatgacctt ttatccccg ttgtctctt
33361	tccgctgac ctggctctc tgaccacctc ccagcccccc aacatacagg aagatggcac
33421	ctgtctctg cagagctaa tccacctg gcttggcttc agatgcttac agtctctctg
33481	cgggaggccc cgtccccac taggccccaa gcctgcccgt tgagtctcag tctcacctg
33541	aacctctc atttctccc agtctcagc tcccaacccc agaggtatcc cctgcccc
33601	tcaaggccct tgcctctc tgggggatg ggggtatgg gaggcaagc ctgatcccc
33661	gagcctgac cgtgacaat gtccgtctt ggcacatgc tcccctgct ctcagactc
33721	cctgttccc ggggatgggt tgcgtgatg acaagtgat ggcctctcag gtcacacctg
33781	tcccctccc aaggactga ccttaacc ccagactcgg ccagaccag aagcacttc
33841	agacatgctg gctgataat gagaaggctt ttatcagga gaaacaggaa caggagggg
33901	ggagaggccc ctgtgtgag gcgacctggg tagggctca ggggtccatg gagaggtgg
33961	ggaggggtg tggccagag ggcacctgag ggtggggctc cagggcccta agaacagct
34021	gaggtctca ctgtctgt caggtgctc cctctgtgag tgacctgca gctgtaactg
34081	ctttcattt tccagctgt gcccctcagg ctcatgctt gctggcccg tatttctgt
34141	tgctctgtt ggaggcccgg gtgtctcca cgttgcgggt gatgtgctg cogtctgct
34201	tccaggccc ggtcacgta cccgggtag agtctctgat gagacacacc aggggtgct
34261	tgttggcct gactctctg tggggggcg gaaacagggt gaccgagggt gcggacttg
34321	gctgacctg gtggacagag gagagggtt aagacgccc ggaggttctc acctgtccc
34381	cacgtagcc ctgttgcct tctgtgccc ctccgacct tgcctcagc ccttggggg
34441	cagacagccc ctgagaagc attgcaatc acttccaag tgaccagca aactgtgct
34501	cagatcccc ggtgagacc agggctgctc tctcctctc tctggcccc gggagtctg
34561	gtctctctt ggcactgacc cctgagccc tcaaccccct ccagaccct cctgacctt
34621	ccgctatgc agcccagggt cctctctctg gaaccgggt cccccccc acctgccagg
34681	acggtctga tgggagatgt ggggacaagc gtctagggt catgtcggga gccggggccg
34741	ggctctctc tctgcccc ccccagctc agctctctg gccaaagccc ggggtctctc
34801	tgaggtctg ctgtctacc tccgcccc ctgagtgca gggccccctg cctcactgc
34861	ctcagggga cgtgcccc acacagcacc tcaaaagacc ccgattctgt gggagtca
34921	gcccgttca tatctctaa gtccaatgt cgtctgagg ccagcggagg ccgacctc

34981	gacagggtg accctgggt cccagggat caggtctcc agactgacga gtttctccc
35041	catgggacc gctcttct gaccgtgtc ctgagatcct ctggcagct tccccctct
35101	cagctgtgc caccggccc ctacggccc agcggggcag accctctct ctctgccctc
35161	cagggcttc ctacagctg cctctgtgt tctggggcc tggcatagc ccccgccag
35221	ccccaaact cctgtctgc cccccgtg gggcatggg ctacagcac agagcccggg
35281	gcttgagat gccctagtc agcaccagc tctggcccgc accccagcgt ctgccctga
35341	agaggggaa aagtcctgc attcctggac caaacaccag ccccgccgc ccgactggcc
35401	ccattggag gtggccact ggaigtctt gctggttacc ccaagacaa ccgctctcc
35461	ctccggccc caccgagaaa ggtggggtc gcccctaag gccgggggga cagagaggaa
35521	gctgccccca gagcaagaga agtgacttc ccgagagagc agagggtag agaggtggg
35581	gtaggtgag agccactac ccaggacgt gaccaggtc ccgcccta agacaaaata
35641	cagagactaa gtctggacc aaaaccgcc gggacagcgc ctggggcctg tccccgggg
35701	gggtgggccc gagcgggaa ctgtggggc tgacggggc agggctgag ccggtgggg
35761	tggtctctc gctgaggggt gttgtggag cagccttca gaggccagg gacctgtgt
35821	cctggagtg cctgtgccc agccccctg ccgagggcagc agccacacac gccctgggg
35881	tcaccagtg cccctcaet cggaggtgt cctggccacc actgagcct tagcgtgag
35941	ggagacgtg agcgcgcgt ctgtcgggg cggcagagga gtacggcct gcttggacc
36001	tgccagccg ctctggcct cactgaagg ccttgggtg tcttcccc acagcttca
36061	cagtccagc agccagctc ctctgggg ctgtggacc cgggctatc ctacggcccc
36121	aagtgggaa cctgcccct ttctccacc caccgagatg cagttcagt tgtctctc
36181	aatgaacatt ctctgtgc agtactgt cttctgtac atctgtgt ccatccatg
36241	atccaaalc catccatca tccatccc agccatcat ctgtatcca acatccatc
36301	ttccatcat tgcatacca tcttccatc tgcactgt ctgtccaca gtggccatca
36361	agcaccctc tccaagccc tgttcacac gctgggact ggtgggggga gccctgccc
36421	tccaccctc ccaictctc-tgaaactct ggggtcaagt ctaacaagt cccatccgt
36481	ctagtctgag gtccccccg agcctctct tccactct ctgtcttga cccactgt
36541	gcactggac gaaccaccg gcccttga tccctgttc ctctgacc tctttttt
36601	ggctctgat ttatacat ctgctctt ggaggctct cagctgagt gtcccaca
36661	cgctcagac tcagatctt ccatgaaac tctcccag tcttgcaga cctgtcccc
36721	cacattgtc tcaatcggg agatttctc acaagccaga gccctgact catccataa
36781	tgctgccc tcatgtgc agctctgt tctaccata accaaalc ccttaaaa
36841	tctcagaaga acaaaaaag caccagatg gcactgtcag agttatgat gacaagaatc
36901	ctcagttcag ttactact cagctgtc cgactctt gcacccatg aatgcagca
36961	cgccagcct cctgtccat cccaactcc cggagttac tcagactac gtccattgag
37021	tcagtatgc catccagca tctatctc tctgctccc tctctctt cccccaatc
37081	ctcccagca tcagattt ttcaatgag tcaactctc gctgaggtg accaaagtac
37141	tggatttca gctcagat cttctctc aaagaatcc caggctgat ctcttcaga
37201	atgactggt tggatctct tacagtcca gggacttca agactctt ccaacacc
37261	agtcaaaag cctcaattt tggcctca gcccttca cagttcaact ctacatcca
37321	tacatgaca cagaaaaac cataacctg actagatga ccttgttg caaagtaatg
37381	tctctctt taaatgct atctaggtg ctcaactt tcttccaag aagtaagt
37441	ctttaatt catggctga atcaacatc gcagtatt tggagccca aaaaataag
37501	tctgccact ttccactgt tccccatc atttccatg aagtatgg accagatgce
37561	atgatctt ttctgaat gtagctt aagccaact ttactctc actttact
37621	tcatcaagag gcttttag tctctcac ttctgcat aagggtgtg tcatgcat
37681	atctgaggt attgatatt ctctggca tctgattcc agttgtgt tctccagtc
37741	cagtttct catgatgac tctcatata agtaataa gcagggtgat aatatacagc
37801	ctgactac tctttctt attggaac agctgtgt tcatgtca gtttaactg
37861	tgtctctg acctcaca agatttca agagcaggt cagggtctt ggtatcca
37921	tcttctcag aattttcac agttatgt gatccaca gcaaaaggt tggcatagt
37981	caataagca gaaatagat ttcttcaa actctctg ttttccatg atccagcaga
38041	tgtggcaat tcatctc gttctctc ctttctaa accagctga acatcaggaa
38101	gttccaggt catgtatgc tgaacctg ctggagaat ttgagcatt cctttctag
38161	cggtgagat gagtcaatt gtggcagc ttgagcatt ttggcatt cctttctg
38221	ggattgaaat gaaaactgac ctgtccagg cctgtgcca ctgtgagt ttcccaatt
38281	gctggcatat tgaatgagc actttacag catatctt caggattga aatgctcca

38341	ctggaatcc atcacctca ctgctttgt ttgtatgat gctcttaag gcccactga
38401	cttcacattc caggatgct ggctctagat gaggatcac accatcgtga ttatctgggt
38461	cgtaagatc tttttgtac agtctctctg tctattctg ccacctctc ttaatactt
38521	ctgctctctg taggccata cegttctgt cctgcctat cgagccctc cctcctacg
38581	tagagactt aagcaggaag gtagccctg ctgcactggg tccagcatgc ttttaattc
38641	gcagtggaa tctgggtca tgaattgtt taaggatgc gcatacatt ttgaagcaa
38701	aattaacag gacagcagt taaagcagt acttattct gattaagaa agcaaatatc
38761	cagcctgta ctaagftat taactaaga aacatctca acitaaata cagtatctcc
38821	tgaacttac agcatgctc acattaaag gcaaacatc tttagagcc agggctcca
38881	cgcttactt tattattaa tatatgctc agatcaagc ccatgacaca aaatgggggg
38941	aagagtgta ggttaggaa aaatgagata aaattggtt tgcagggta tgggctagt
39001	tacttaaaa aaaaaacaa aacaagctc agatgaactg aaggactat agaactgta
39061	caagagtaa cctgtgatg aatacaagca ggctgggcaa aactcagcag gtttctct
39121	atacaggcag taatgattga gaatacгаа cggcgggagc gcttacaacc tcgataacg
39181	ttctatata agccctagga atgaactaa cacggnnnnn nnnnnnnnn nnnnnnnnn
39241	nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn
39301	nnnnnnnnn nnnngctcc cccaccctc cctctctcc ccccaccac cagtgccca
39361	ggtctctgc ccagagagct gaagatgcca gcaggccgc tgcctcctc gctcgcctg
39421	ccgggctcg ctgccgtct gctgcccag cacacagatg cagcccagc tctcctgce
39481	accgctccc ccagggcagg actctcccac aacaccaagg gcgtctctg gtcaggatg
39541	gccctctg aggtgtaag tctctccgg ggctgagacg aatggccgg agatcaaac
39601	gagccaagg ccgccaggc gctggcgcg gggcaccat ggtcagagc ggcccagctc
39661	ctcctccc tccctccc cctctctct tctctccc gctatctc ttttactct
39721	gcaattaga aatgataccg aaggcaaac accgtccc cgtgtgct gcttaaac
39781	ctttatctc ttatctata gcgttccaa gtttctgc taagtatg aaggaacct
39841	accacaagc agcaagctc ccagacct cgcctgtca actgggaatg taatgtct
39901	ttcaaggac ctaagttct atgtcaaaa ccgtgtgtg ttcttttg gagtgaacct
39961	aggccactg ttgtctgcc ttcaaaagca ttctaaaca ctctccaga cccagggctt
40021	ggcttactt tccagaaat ccaagacag aacttggaa acctgatga gaagccctg
40081	gagcacaga gggccgggg tactgaggt aggtggggg ctccgtctg atgacacgg
40141	cttctactt ctatcgttg ccgtccaga tctctccac ctccggagct ttacagagg
40201	tcacgtgtt ggccaggtc gtgtatcat gatctgcaac cagagcggg gctgcgtca
40261	gcccggggc gggcagcag caggagcagc caggagcgc agcacaccg gctctca
40321	tgaggaggt ggggaagcg gctgtgacc tcacactgc ccgatgtgg cctctca
40381	aggccggcc tggacctgg ctctctcag aggcctgct gggccctcg cacagctc
40441	agccacagg cctcttggga caggaggct ccagagtgc ccggccggc ggaagaggt
40501	tgacacctc gcagtcaca acacgaagc aggtggagat gggatgagg atgagaaca
40561	ctttctttt aaacaagag ccagagagt tggaaagag tctgcaac gcaatgaa
40621	ctctggccc ggtggcagc ggcctggga gccgagttc tggcaatc gaccacagt
40681	tgctaggga gccgggtga gacggagggt taggggaag cggctccca gggagcgcg
40741	ggcccgggt gcccaaggt ccagggggc aagcagct agggcgcag ggtatgtac
40801	cgccactga cccggcgcg gaggccagg gaggggctg aaggtcacag cagtgtgtg
40861	acaagaggt ccggctctg cgttaaaga acgctggga cagaccaga cagcgcacg
40921	gacacacta taccggacgg actgogagt gcagcgcgc gcacacac acacacaca
40981	cacacacac cacacggcc gggacacact cataccgac ggactgoga gtgcacgcg
41041	acacacacac ccaccacac cacaccacc acacacac ccaccacac cacacacac
41101	cacacacac ccaccacac ccaccacac ccaccacac ccaccacac cacaccaca
41161	cacacacac cacacacac cacacacac gcccggtgc cccagggca cacagcagg
41221	agcaaatg cacagacac agagcagcg ctgcccacc ggtgccaga ccaggccca
41281	cgcatggt tggggcggg gacggggag ggcgggagc aacggnnnn nnnnnnnnn
41341	nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn
41401	nnnnnnnnn nnnnnnnnn nnnngtatt aaagaagcc ggagcgaga tatgacggca
41461	agaggatga gttggggcg gggcaagat aaagagcgc gacgtagag gggatcgat
41521	tgtatcgg aagcagacg aggagtatg ccgtattga ttgalagca gaggaacgt
41581	aggaggggg ggggagaga ggggaggtg gggggtgt ggtgggagg gaactttaa
41641	aaaaagagg gagagttga ggggggata aacggcgtt aaaaaaac aattgaaat

41701	taccagggtg gggcgccag gggggtgatt cattcttga gggggcaaca tatggggggt
41761	ggctgtcgcg gattaggaga aataaatat caggggtgat taagtittg gcgttgggga
41821	ataatgaat aagaatcaaa tatgaatcgc gtggcatcg ttaccatcg ggggaacat
41881	ttccatgca aggaacaagg atgigagaat gcgtccgtc gaaccaccgt cccgggtcc
41941	cagtaggact cgcgagctg atagttgcc gagcaacagt taaggagca gaagctgcta
42001	caaaaccacc acctgcaaa gtagggtc caatlacgga gtgcgcctc tgggtgtcgg
42061	tccaaacctt tggaaaggac ctggaataa gtgctacca ccagatatta atataaacc
42121	acctggccag gagaggcagg cgctgtggc acaggaagt tccccagact cagtcatca
42181	ggtaaataat atttgggac ctcccggaa atccagtgt taggactct cggttcaat
42241	cctggtcggg gaactaagat cccacaagt acaagacat gccaaaita aaaaagaaa
42301	aaagagagag aatatattg tgcaatagt ttagaattg aaaitaagt cctgcccacc
42361	ccccccccc aatctgatg aataaagcat tgaatagta agtgaagca ggctctgaca
42421	tgactgatg tgactacct taagcaacc ccaccatag actggtcggg ttccaggag
42481	ttcagggggt gccaggaaga tggagtccag cccctgccc cccccccac cactctcc
42541	actggagcgg cctccccac ctcccacccc tccgacccc gtaccccc acccctgccc
42601	ccaggtctcc cctgtctgt gtctgagct cacactttt gggcagtct tcccttaca
42661	gtgtttct gtgtcccgt accggcccc tcccctctgt tcagttcagt tcagtcctc
42721	agtcattct gactttgt gacccatgg actgcagcac accaggctc cctggccatc
42781	accaacccc agaacttact caaactatg tccatcagc cagtatgcc atccaacat
42841	ctcatctct gtgacccc tctctggcc tcaatcttc ccagcatcag ggtctttcc
42901	aatgagtcag ttctttgat caggtagca aagtattga gtttcagct cagcatcatt
42961	tcttcaatg aatattcagg actcatttc ttgggatga actggttga tctctctga
43021	gtccaaggga ctcaagag tcttctcaa caccacagt caaaagcacc aattctcag
43081	gtcagctc tctttagt ccaacttca catccatag tgaccactg aaaaaccata
43141	gctcagca gatggaact tggggcaaa gtaatgtct tcttttga tatgtctct
43201	aggttgtca taactttct tccaaggag aagcgtctt taattcatg gtgcagica
43261	ccatctcag tgaatttgg agcccaaga aataaagtct gtactgtt ccactgttc
43321	cccgtctatt taacggagg aaatttcca gagccccag gttccaggct gggcccacc
43381	ccactccat gtcccagaga gctgttctt cccaggctc cggctggcgc tggtaagtc
43441	caggatag tcttaccac aagtgtgt gtgtcttag aagaaactc tcccctctg
43501	tgctctgt cctcatccg cagaagtac tgccaggtc gggagtctgt gacttcca
43561	gaagccggag gattttctc ccatttctg aaagagagct cgggggggg gaagctctg
43621	cacccttag ataccagag gagccagggt ctccagggt cccggggacc cctcagtg
43681	ggctcaggaa ccacagagcc agaccctgt tcaaaaacc tgtcacacc tccagatgac
43741	ctttgtccc ttggctccc ctcaaatgt ccaagcccc acagtgaagc gttaaaga
43801	aggatccacc aggtttagt tggggagga ggaagtggg gagctgggg agggcctggg
43861	cctgggagc aggaatccac catggttca ggcagggtct ctggggctg cgggtggag
43921	agcggcagg agcagacaga ggtgactga cagcacac ccttccact caaggagggt
43981	gggcaggggc ggggcacaga ggaacaagag accctgagaa ggggtccacc gagcagctg
44041	ctggaccag acatcttga gccagctga atccagctt aagccatgt cagcccaggc
44101	agggtatagg gcaggactga gtggagtggc cagagctga gctgcatggg ctgggaaggc
44161	cctgcccgtc cctgagggt cccccagggt ctgcccagc tcaatttcc gaccgagca
44221	cacacaggag gaagtgtcg ggggtgagt gcccagagg tctggcagg tccaggggtg
44281	gggaagggg gcagctggag tccccgtg attcaggga cgttccctt tcttccctga
44341	aacctgggg tgtcccggg gccaccgag cctccaggca gggggggc ccagcccaca
44401	atatgtaga agagcaggc ccaggctga gagagcgaag caccatgtg gggagaagt
44461	agactgac ggggccccta ggggtccc cggacctga cggcagcgt cagggcacc
44521	gaacccatt gctgtcagt gtggccagt gtccaaggcc aggatgtgt gtgtgtgt
44581	gtcgtgctg gcgtgctgt gtgtgtcgt gtgtgcgct gcgtgctgt gtgtgtgt
44641	gcgtgctgt gcgtgctag acgtgtgct gcgtgctgc gtgctgctgt gtgtgtgc
44701	acggcgcag cccagctca gcactggacc aggcagcct ggattctcc aaactgct
44761	tgtgatttg gtcaacctg gaggcttga taccgcat ccatcgcce cctctgccc
44821	ctctacac cgtgttgt gtattatc agagcttgg aggtctggg aggtatccc
44881	acctgccag taaacctga gctgcccga atgcactga tgcgggcaga cccgagcgc
44941	tgtccggag acgaaggca gctgtcacc ccgcccagc ggcagctgg ccacaagcat
45001	catccaagca gtgttctt gagccccag ggtgatgca aaggagccag gagacacct

45061	cgctccaag ctgggggacc ccaggctgt tatgccggac agtaaacacg ttcagctcgg
45121	gagggaagg gttcccctac ctccagggt ttctcattcc acaaacatcc aaagacaatc
45181	cataccgaag gcgatccgtg cctttgcicc tgagacgtgc ggaagcacag agatccacag
45241	acactgtctc ccaggatcct atgtatgtaa aggaaccgaa gtcccaggct gtgtgtctgg
45301	taccacatcc cacggaacag gctggactga tttcaccaa atgtagcaga aacgtaagg
45361	agtatcagct tcaaaaatag agggccagac atgtctgaga agtcccctcc agaaaagtc
45421	cittggggtc ctccccaga gttgctgaaa cagagaaccg gaagggctgc agagctgaac
45481	ttaacaact ggatcgcaa ggtccgtctc atcagagcga tggttttcc agtggctatg
45541	tatgatgag agagttggac cataaagaaa gctgagcgc gaagaatcga tgcctttgaa
45601	cictgggtgt ggagaagact cttagagtc ccttggactg caaggagatc caaccagtca
45661	atcctaagg aaatcaatcc tgaatatica tgggaaggac tgatgctgaa gctgaaactc
45721	caatacttg gccactgat gcaagaact gactactgg aaaaacctg atgctgggaa
45781	aggtigaagg caggaggaga aggggtcgc agaggatgag atggttgggt ggcacacc
45841	accatggac tcaatggaca tgggtttgag taaactctgg gatttggta tggacagaga
45901	atcctggcat cctgctgccc atggggctat agagagtcag acacaactga cgcactgaca
45961	gaactgaagc aactggcaag ccggagggtg ggtgccgct gcgatgagc ggaacgtgca
46021	acctgccag tggagctct cctacacca gactcctgac ggcactggga ccttagccct
46081	ccacggcctc tccaggcca cgagacacc tcacagagca gagaagcga acagagctgg
46141	tgtcagaac caggccccgg ggtggggcg gggctgtgg gcagcctta gtgagaagcc
46201	cttagccct ggaaccagag cagagcagaa cagtggcag agggccccct gggagaggcc
46261	cccccccag agtaccggcc ctggccctg gggagaggc cggctctgg ggcagggaca
46321	gaaggccag gcagaggatg gggccctgg gacggggcgc accaaaacag cccctgccag
46381	caagggaag ctggggcact ttgaccccc tccaaggagg agcccacac agcgcactg
46441	cccaaggtg ccttggccct gggggcacat gaggocagg ccaggccagg gggccatga
46501	ggcccccag ggtcagtga gtgtcccag gcagccctgg cctctcacc tcttggccct
46561	ggcctttat cccgtggcg cccacggcct gctgccccc acagcggcgc ctacagcac
46621	agccccccg atggaagccc cgtcaggaaa gagccctgg agcctcagg acaggttaag
46681	gcagaggag tcatggtgca ggaagtgg gcttccctc gatggacc aggggtgat
46741	gaccgcagg gcggggaac agaaggaaa ccagctggag agaaggacc tgggcagacg
46801	tggctgacg cacagcctg accctgggcc cagtgtgct tigtgtggg tttatttt
46861	aatttgtat tgatgctg tttatctgt ggagctttg ccgccctgag aattgtacc
46921	cgtggctgt gtcctctg cctcacccc gcctctgtg cagggcagac acggcgcaac
46981	gggcagggc gtgccagga ggcactgca tttggggc agcggcccac caaggcaggt
47041	ctgcctctc cccctctac aggcagcgc agaggtccag agaggtgagg caagctgcc
47101	aatgtcac acgacacagg cgcagtcaca gactgtaga aatccggga ctgacaggg
47161	accagagtgt cotgtgtt taaaaaac gcccaagaa agaggcaagt ctgcaaggcg
47221	tcccgggag gcagcaggg ctggctcgg tctccccaa ggaggccagc tctcagcga
47281	gttctctaa gtctaacgg agccaagcct gaaccaagg ggtcacgtgc agtatggga
47341	cactgacctg ggtggggga gctccagca aaggagtag gggggccaag gaggagag
47401	gggtgcacag gctgcagg agctccaga gctggggaaa acgggttca gaccagggg
47461	tcatgccac cctcctta tctgggac cggggcaggt atgaggat ttatgtcgg
47521	ggctgtcagg gtcagttcg tctgtggaa aaattgttc agatcagaga ccagcgtgag
47581	gtcaggttag aggatggaga agaagctgt aagggtgat ggagagcgg gggacgctc
47641	tgggtatca ggcaccgaga tggccatgg aatccgagg cgaatttaca gtacgtctg
47701	cagagggctg tgggggga acaggcactg tcatgaaatg gctacaaaa tctaaaatgt
47761	gcacctttt cggcaatag cagcaagtc taaaagaaa cgcatttct taaaatgctg
47821	taattccgt ttaggaatt cactggggg cgggggaaca atcaaaaaga tgtaccaaa
47881	ggtttacaag ccagggaatc aactcgtta tcatggaga aaaccgaaa taacctaat
47941	atccaacaga aagggtgta tgaagcag catggcact ccaccgcaag gaatcctaac
48001	acaaacttc aaaaatata ttctgacgtt ggtttttaa agcatgctg cacittcaa
48061	agctgtcag aaaaataga aatatgcaa taatgtct ctagccaat ttttaatt
48121	ttctttata atttataa gttataatg tatgaatat aatgataaa ttataacta
48181	taaaaagtt atgaaatgt tcacaagaag atatacatgt aattttatct tctacaatc
48241	ttttatc cagaataac gtcttttaa aaagattgag cacagaagcg tataangta
48301	aaattgag ttctgtca ccaaccac gcttacctt aaaccctt cccagcgag
48361	agacagttc atgtggctg gtacactct ggccttctc ctaggcatgt atgtccctga

48421	aaactcacac acacggctaa tggctgctggg attttagtiti tcaaacgga ctcatactct
48481	gcctatgagc ctgcaactat ttatcagtc tgttgagatt ttctataca gccacatgg
48541	atcccgcagc ttctctgaat ggctctgtat gaattcaag ttggaagaa gcagcgtgc
48601	tttaacatt cgcctattaa tggacgttg gggtgttcc actacaaan nnnnnnnnn
48661	nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnn
48721	nnnnnnnnnn nnnnnnnnnn nnnnnnnnnng atacaattcg agctcgttac cctggcttga
48781	actatatgaa cagagaacga tgagaacagt ttctcaact tggaaacgtt aacatttgg
48841	gctaaatgat tctttttgt gtggagtgg cctatgaata gaggatata gcagcatcat
48901	ttaaccttta ctcactacat accgttagca actacatcct ctccattgt gtaacataa
48961	actgtctccg gacatggaca agtgtgccc tgggatgggt ggaatgacct ttgttaaga
49021	accactgggt cagagattca tagattttg tctgttgac ttttaaaa tacatcttg
49081	tttttattt attggtttc gctcttact ttatgattac ctcccttta ctggggctt
49141	ccctgataga tttcccttc tggctcagct ggtaaagaat ctgccctgca tgcaggagac
49201	ctgggtcag tccctgggtt ggaggatcc cctggagagg agaagggcta cccacccag
49261	tattctggcc tggaggattc catggagtgt atagtcctg gggtcgcaga gtggacatg
49321	actgagtac ttcacacac acatagtcc ctggtagctc agctagtaa gaatcccacc
49381	cgcaatgcag gagaccccg tcaattcct gggtccgga gattccctt ttttactcc
49441	ataagatctt atctggggac aaaactaaca gctatgccag accttctgga catcaggaa
49501	ctgtaggggt gtggactgga cagatgtgt ttttctcca aacacaaaca tacatctga
49561	tacatgata tggagagagg gggaggagg ctgtgagct ccaggggacc gtgcaacct
49621	gtgacattca tggaggcgt tgcgggtgat cactacacag ttctcttc tggttctg
49681	gtcaattgac ttcaaatc caattcctt acttcaatt agactgagg aattttacac
49741	tattgaaga catatgata catgagtat gticagccc atgagggtc attttgtg
49801	tccacttgc ctggaacaa agttggactg atttactct aggggtcct ggggtgttt
49861	ctggaggaca ggagcattg aaccaaggc ctccgtgaag catgagctc tctcaggtg
49921	gaccaggag gaacgcaagg ccgaggaagg cagactctcc tcttccctaa cccgaggtc
49981	ctgctcaga aagggacaat ataagacta gaagaaaga aagaacatca gctgtggag
50041	gittgtctc tggagcagat tcacacgtg aggcctatgt gcaggaaic taggtgaaac
50101	agagcagca cccatgtgt tggaaaat taaattaca ttgcagta cgaattgtt
50161	taagccagac agggtagcac agcaagica ccatgtgtc acctgttiti tgaaggag
50221	agagaacttg ctggcacatt caggaaggc cgtgtctag ctttggagg acactgagag
50281	gocacaagca gatgtgagg accagggtct cgggcagagg gatcaatca ctgcttca
50341	cttttgcac atctgtgtc tgtccatct ggccagagta gticagctt cagatctgg
50401	agttccactt gtagaatac caatctgggt cattttaaa cctctctgg ttctactaa
50461	tgtttttaa atctcttgg ctcaagaaa aaaataaca taattttaa ggggtgttg
50521	gggcttgc tataaagac attatctgg ccaattcaga gcatgttga attaatcat
50581	ttctgctta ctatagctc tattttctg atctttaca gtaatttiti gitaggaac
50641	gggtactgt aatatttct tgttgaatac gggatcttg tatttttcc taattttt
50701	ttttttca ttttggtt tacctcagg aaagtcacta ggactcagga aagtccttg
50761	tccgctgtt attcagctt ctacctggg gccagggcag cgttctctt gggctaagt
50821	tcccacaac cggggccagt tctctcact ctccacctg aggcctaat gaggagctc
50881	cctgcgctg agcagccggc cctctgtga cgtgcgtgt tctctggca tggcgtccg
50941	gttctcttg aggttccgt ctccttgc tcactgtcc ccgacatga gctctcaggc
51001	tcaaagcagt gtcgcagtg tgcagacct ctgttagct ctctctctt caggactct
51061	ccctctagat gtgttttc ttttgcctc ttgacctc gctctgaac caggctgtt
51121	gtctagtg atctctggag ggaagcctg gaggctggac gggctcgc tgcgtgtg
51181	tgacaggtgt gggctcggg cgggcctgc agtctctt gaccagacc gggactggc
51241	tccgggctc aggcatact gactgaatct ccttcacaga gggctcagg cctggggg
51301	ggaaccgtct ctgcaatgac agccccctcc agggaggca cagcgggag ctgcccaggc
51361	tccagccta gtggagggtc ggggagcca ggggagcgc ctgacggccc cacaccggc
51421	caggctgtt tcttctgtt tctgagctc aacagaagct ccgaggagct gggcagttt
51481	ctgaattct cccggagtt tggctctga gtctctgtc agcaccgtat ggacatccag
51541	agtcactag cagtggtct tctctctg tctgtcttc atcaggctt ttgtccaggt
51601	caccacagg ccaaccagg gacagcttg tcccgcagc ccaatctcc tgcggagcc
51661	cctgtcagc ctgccgaagg gccggaggc cgggggaacc gggccaggcc tctccctgt
51721	gtgtccacag tctcccggt gctggaggag agcgtgagca ggacggagg gtttctct

51781	cactccccg tctgtctgtg tcaactgtgag gattatcact gctgtcagct gactgacagt
51841	aatagtggc ctcgtctctg gtcgggccc cgtgatggc cagcgtggc gttttccctg
51901	agctggagcc agagaaccgg tcagagatcc ctgaggcccg ctacactatc ttataaatga
51961	ccctcacagg gccctggccc ggcctctgct ggtaccactg agtatattgt tcatccagca
52021	ggtccccga gcagggaic ttggccgtct gtcccaaggc cactgacact gaagtcggct
52081	gggtcagttc ataggagacc acggagccgg aagagaggag ggagagggga tgagaaagaa
52141	ggacccttc cccgggcac ccaccctgag gcggtgctg gagtgacac tgggttcggg
52201	gcaggcccc gccagggctc ctgtgtggcc ggagcctgag gccagggccg gggggccgca
52261	ccgtgcaga gagtggagag gggcagcagg agagggtcc aggccatggt ggatgcgcc
52321	cgagctctgc ctctgagccc gcagcagcac tggcctctc gagaccctt atccctctc
52381	agagctttgc aggggccagt gagggtttg gttatgcaa atccacccc gggggcccct
52441	cactgagagg cggggcacc acaccatcag cctgtctgt cccagcttc ctctcggt
52501	tctcagctc gcacatcaga ctgtcctca gggactgag tcaactgac ctccccctc
52561	tctgaccaca tgaccactg cccaagcccc ccggcctgtg gtcctccctg gactccccg
52621	tggggcgtc agcctggcag cctcctggcc gggactgag gcatgggtc ctgggttca
52681	ctgtgatgt gacccicaga ggtgtcact agtctgagg ggatggctg tccagctctg
52741	acttctgcc aagcgtgct ccttggacag ctgtggacce gcagggtgc tccccgaa
52801	gtccccctg ggcagcccag cctctgacct gctgctctg gccagctct gctgccccct
52861	gctgtggag gacgatcagg gcagcggctc cctccccca ggtcaccaca aggccctgt
52921	cagcagagag ggtgtggacc tgggagttca gccctgctg gccagcact agaggccgc
52981	tgcaccggga agttgctgt ctgtgacct gctcagggc ggagatgacc gcgccgtcc
53041	tttggttgt tagtggagt gagggtccg gatgactca gccgtaaac gccaggtcc
53101	gtagcaacct gtgcgatcc cccggggacc caggcctct tgtgtgtg taccaggtt
53161	ggcactagc ccaccccagg agggcactc gctgatggt ttcctggcag ttgagtcat
53221	ttgagaact acatcattt catcaccaca tctcaccac cagtaccac accaccatca
53281	ccattccac atctcttc tcttttct ttatgcatc tcacaatcc acaccctca
53341	agagtttga ttgtagcat atttactta gcacagtgt cctctttta ggaaactggg
53401	ggtctctgc tgatacccc gggaacccat ccagaaatt tactgatgc tgaacccctg
53461	cgtttgatt ctgcccagg agaccctagg gcccaaat tctctgaac ctcccatag
53521	taacaacac tcatggccc ttttatac ttaatttga aaatatcct tgaagtgt
53581	acctaccct acatttaca gcaggtaag ctgcttcca ttgagagca agtccccaga
53641	tcaataaga gaatgggatg aaccaggat ggggccagg gtcctggat tcagactca
53701	gccgttagg acagaactg actaggtac aagtggcgg ggtgggggg caatctggg
53761	ggaactgtg cacccccagg gctcgggccc atccccca cctctggt tcatcagta
53821	gccccicag cctgcgtgt gaggaggcca gggaagctat ggtccagtc atgctggga
53881	atatgtggg ctgggtgct gctggctct aggggtctg ccaggtctg ctgctctgc
53941	tggcagtga taattgctc tcatctctc gagaagcac gagtacagg tctctaatg
54001	ccagctatt ggaggaggca gtgagcact caccctgc agacatctc ggaggcatca
54061	gtgtcctgt aggtgtctc ggggttgg ccgggggacc tgagattcag ccaatgactc
54121	tcagagggc cagctgtgg tgcagggca gggctggcg gtggaggata cctaccaga
54181	gcaaaataa gagatcacc aaccgataa aattgacta cacccttgg totggacat
54241	tctgtctga aattctgt ggacaggaca cagtccctg ataaaggat tctatctg
54301	ctgtgcaat agagctgtc acacgttgg ctggacatg taalccttg aacatggtat
54361	taattctgt tcaatacat ctgaaggat tttgcatca ataaacctaa ggtatitgc
54421	cctgcaatt cctgtctg tagtctctc gtaggctg gaagggtaa ccagcttcc
54481	aaatcgagt aggaattcc ctattctc cactgtctaa tagacttca taagattgt
54541	gtaattcct cttaaatcg ctgctataa cactctgt gccaccgta ctgaatttt
54601	tgttaggat attttaaac aagcattta atgatttcc cttaattt cggctgtct
54661	gggtctgt gctgtgtgc ggcgttctc cgtgtggcc agtggggcg ctgctctgc
54721	gttgcgaag tgggtctc gactgagtg gcttctctc ttgcagagcg cggctccag
54781	ggcctcagg ctgcgtggc tgcggcacgt gggcicagta gtcctgggc acaggtcag
54841	cagccttca gacgtttg tcccagatg gtgggtggt cgaaccggg tcccctgct
54901	tgcaaggtg attctcacc gctggaccac cagcagctt cctggaggt tttaattat
54961	gatttaagc tctcattaga tctctctca ctttctat tcttttga gicagttga
55021	tactttgt gtgtctgaa gtttccat ttatccaag tcatctaat tttgataga
55081	caattattg ttgactatc aattgtgt ttacaattt gagagcattg tctgcaatt

55141	ccitctatct gcaagattgg taataataic tcccaagagg agtcacaac tgaatgaga
55201	ttanatacag gcttttttt taaaagaatg aacttatgtt gttgccttc tcatagatct
55261	tacticttag catgactgta ctactgact ggggcgttt catgtctgtg tggagagcta
55321	ccattaglac ttctatgce ccaaagacat cgggctoctg ggcacagtga aaacactcct
55381	ttctgtgget atttgcaa atatggccta gcttagcgtc ataagggatc acagctgaca
55441	acigctggaa cagagggaca tgcgaagcaa cgtgagggct ggaacctgga gggctctc
55501	tggggacagt ttaaccagct ataattgaca ttccagatc tgggacatgg agctgtgaac
55561	tggaccaatg actgtcattt ttggaagaga aatcccagga gagaagggtc caggggaaac
55621	tgaggccgca tgcagtgcct caggacaggg gacaccttct ccagcagagc agggggggcc
55681	gcccaggccg cctgcagtga ttccaccagg aggagatgca ttccctgcaga cctctgacag
55741	cacggccctc tctctagaca cagggtcaca cccggggccc tggaacctt tgaaccctta
55801	aacctttct ttctgacca cctgacagc agtctagctc agaacagaca tcttctttt
55861	cagcaggaaa atectttcc tegttagg gagcagctgg caccggagga gctgagtctt
55921	ttaaacacag gctgctgaa cctcaggat gacctgcagc tctcagagg aggctggagt
55981	gtgatagctc actctaatgt tactaaaagg aacatattgg acacccctc tctgaaaaat
56041	ttccctctg cctctatct cttagtccac ttatgccc tttactgct ttctattta
56101	ctactictaa cgccaacctt tctatttcc cctccagtt taacacggtt ttccctccac
56161	ccgctctct taatctcaga agattctgoc tttctctta ttatcacacg cccctacttt
56221	ttatttttt tctaccoco ctittatcc ctcccctct cactctctat ttaattacat
56281	cttaactaca cgcctctgc tatcttgaa tctatccaaa tattttccc ttatataca
56341	ctcaggccg agcgggtaac ttattataat ttctttatg cgcctacctt atttcccttt
56401	attictaait atctatatat acccatgcaa ttccgnnnnn nnnnnnnnnn nnnnnnnnnn
56461	nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn
56521	nnnnnnnnnn nnnntgggt gtacgtata gactaaacc gcataagaa gtgggtcaat
56581	ctatggctgt gagaggcaga aaataatatt atcatatata atttatgta taacacactg
56641	agggtgtgg ctctagaat agtgcggagc gggagaaagg tgggaaggag aagacacaag
56701	agagagatgt tgcctcgcg ggtggatgg gcggaggat agaagaataa aaagaggaga
56761	ggtatagagg gggcggggg gcataacgtg tggggggta aatagtaggc ggtattatg
56821	aaaaaaagaa agacgggggg ggcggtaaca tagaatcgc aaaaaagta tactatgaac
56881	gggattagg gagaagggt gggggcgtg ggtgcccgg gaaagagggt tgtgtataat
56941	tgtatggag tgtatttga atatatata atgtaatagg gagttaatt agtgaattg
57001	tggagtatt atattgggt gtgggggaca tggcaaatg atgacggga taaaaaagt
57061	aaagcaagag gggaggggaa aataagggg gggagaagg cgaagaaaat aagaggaaga
57121	agaagaacg ggggtggcg gcggggggg cgcgcctct giatctgct tttttgtt
57181	gtcgtggtt gtcgcctct tgtgggctc gggcggggtg tgcggaaaa aaaaaggcg
57241	ggaggcccg gccccggca cgcggcacc cgcgggctc ctgctctc ctccgcagc
57301	tccgggggtc ggtgagcctg gcctctcgg gcgcgggccc ctagctgtgt gcctctgca
57361	gaatcggagc cgtctggca gcacgcggag ggcgcggca agggccacg gacggacct
57421	caaaggccg gcgggagc gcgaagcga accgaggcg gctctgggat cggccagagc
57481	ctctccccc ctccgcgtg gccccaggt cgcgggtga ctggggcgg tacaagcac
57541	tcacccctg cccgcccga gaaagcctc caggactct acagagcacc gccaggagg
57601	catccggtc cccctcgc tcagtcagt tctcagtc tctcaactc ttgcgaccc
57661	catgactgc agcacccaa gcttccctgt ccaaccaca ctccggagt ttactcaac
57721	tcaictattg agtcaigtat gccatcaac cgtctatcc tctgtgtcc cctctctc
57781	ccacttcaa tcttccag catcagggtc ttctttatg agccagtct tcaatcagg
57841	tgtcagagt atggagttt cagctcagc atcagctct ccaatgaaca ctcaggactg
57901	atttcttta ggtggactg gctggatgca gcgccagaca ccgaccgct ttaccctg
57961	tgtctttcc aatgctgtc cctcggggc ctagggcat tgtgagggt tgaatcctg
58021	tggccttga tttaacct tagtccagg tccaggcag gccatccgg attcaggatg
58081	ctccagcc ctccaggaat ggcaggttt catggtcct tctgagtgg tctgagtgg
58141	tcatattgt gcccttggca gggaggctc ctgacttcc tatctcaca tcaigtccc
58201	caaccoccaa gagaggcctc ttggcccagg gactgcagg aggatgaat caggagcaga
58261	agcatgggt aggggctca ggtggcaga gaggcccct ctgtaggag gaacggcaag
58321	caggaggga acaggggac cggcagtgc tggcaagctg ggtgatgca cgactacgc
58381	ccgaccacac agtctctca gccagcccga gaagcagggc cctcccctga ccccatctg
58441	ggcctgggtc tcagtttct cctccctgca atggggtag tgttgcctc caggagagg

58501	gagcatgtaa aggtggccac tctctctgg cagacatgcc aggcctgggc cagcctccac
58561	ccctttgctc ctgcagcccc tctgacctg ctctgttig ccacaccggc ccctcctggg
58621	ctgatcaggg cccccctct gcaggaagcc ctctgggaca agcccgact gctgtaactg
58681	tgctttcca ctgtacctg caacgtggga gctgttact taaaactccc atgactggtg
58741	gattgccggg ccccagaaca agggcacgca tccctggagg cctcagagac cattaagggt
58801	agttaaacat tttacttta tgcatttca tgtgtacag aaagaaaaaa aatgtalcat
58861	cagttcatca aatccatgat tcttgacca atattgctaa gatgaggctg aaataggcat
58921	tccattttt aaaaaactga atcactctga agaacagat ggcaggcttc cctgggtgtc
58981	cgggtgtaa cagtcacatg tccagtct gggggcatgg gticgatccc tgaaaattt
59041	aaaaaggaag aaaaagatgg cccccctgc cctgggattc tccaggcaag aacactggag
59101	tgggtgcca tttctctc cagtcatga aagggaaaag ggaagtga gtcgctcagt
59161	cggtgcgac tcttagaac cccatggact gcagcctacc agactcctcc gtcctggga
59221	tttccaggc aagagtactg gactgggtg ccattgctt cccaggcaa acggcctgct
59281	actgtactg ctgtaaatc gcttcagtg tctcaactc tgtgcaccc catagacggc
59341	agcccaccag gctccccgt cctgggatt cccaggcaa gaacactgga gtgggggtgc
59401	attgcctca gctgtctgt gctgtgcta agtcgttca gtcgtgtcc actctgtgtg
59461	accgataga cggcagccca ccaggctccc cctgctctgg gattctcag gcaagaacac
59521	tggagtgggt tgcatttcc tctccaatg catgaaagtg aaaagitaa gtgaantgc
59581	tcagtgtgt ccgacttta gtaaccaat ggactgcagc ctaccagggt cctccatcca
59641	tgggatttc caggcaagag tacitggagt ggggtccatt cggcctaggg agtgagaat
59701	cacggctgc tccccctc tgcctctc ggggtctctg tggagcctcc ctggagagge
59761	cgcggcggct cggggactg gagggggagg ggggggtgag tcagccggtg gccctccct
59821	cgctgccgt ctctccctt tttaggcaca agctgggcgc ccttttagg cgcagcctca
59881	ccctgcgggc cactgccgt gtttggctc cccggagata aacagatig cctgcacccc
59941	gggtcatcac aaggattgia tgaccgttc ccagtgtct caccaccctc cctctgatc
60001	tcagagacgc gccctgcct caggaggctg ctatccag gccaaagggc ggcgtgggt
60061	ccccagccc ccgcacagac actgctctt gaccaccctc tcccaacagc ttactgcca
60121	agaaggcctc ctgaccctc atcctgccc gtgttttga gaaagctca tctggccct
60181	cttctcggg gctcagitt cccccctgt gaactggcgg attctgcaa gctgacgtc
60241	tggccagcgc cctccccgt gccagtctc cccgggacac agctgaatgt cctgtctgg
60301	gatgcacct cccaagtgg cctgtcagga ggcggggcgc agcagggaaa cccgactct
60361	ctcagacgge ccatgcatt ggggacgtg aggcgccggag cagcggcacc ctctggcca
60421	gggtcattct cccgccccc cccgtccctc gggcctccg agaccgcagc ccggcccgc
60481	ccgggaagga ccgatccgc gggccgggc acccccctc cctggccgcg ggcgcgggc
60541	gagtgcagaa caaaagcgg ggcgggggc gggcggggg cggggcggag gatataagg
60601	gcggcggccc ggcgcacccc agcaggcct gcaccccgg ggggatggo tggggcggc
60661	ggcctcgcg ggcggcctc gcgcgcctt ttgttttg tgagggtat ggggcggc
60721	gccccgtae atttttcat ttataattg gtattagcta gcgagtgaa ccacaccct
60781	attccactat agccaatit tgcgggggca tcttacata cagactgcc cgcctctat
60841	tccgtacag catacagat cgtctctta ctacagact agtgattat gctatagta
60901	cacaaaaaga acggttgtt cggcgtaat gttcattit cctcctctgt tctcctgac
60961	cactcaatt acaccaacac tctactatt aatcacgta ttgtacgcca cctccgccc
61021	gcgaactaaa agaattgca gatattctga agataaaac gttcattgt acgccccgcg
61081	cgcttcgct atattactt tagaactct tctcggccc agcagttat cccccccgc
61141	aactagatgt cgccttaata ttgttctaa ccgttttga tctaacgat aggcgggaaa
61201	ggtagacatt cagccgtac gacaactaaa atcgacgagc acaggctatt tatctgcga
61261	ccacacgccc gcgtataca nacctgaaa ttactaaca tcgagagtaa gggcacagag
61321	cgaatacaa gcggcgtgt gggaggtgt tctgtatga attcgacct cgcgcggccg
61381	ctctgtgcg tcnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn
61441	nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nngatataa
61501	tattaataa cagcggatag atgtgtgtaa gggaggagg gcataagaa taaagagag
61561	gcggcgggag agaaatagag tagaggagga tgagagaaa aagaaagca gcgtaggac
61621	aacggcgggt gggtagtag ataaagttag tctatatt tgagtaaag aagggtagat
61681	ggagtataa gaagtaagga gaggagagg cggcggagag agagagtga aagaaaata
61741	gtggccaaag gcggggggg tgagaagcag tagaagagaa gatagagaag ggggaaaag
61801	aggaaaatga ggattagaac aagtaggaca ggatagatgt gaaaaatgag atcaggctca

61861	ggaggagaaa aagtagaac tggggcgtga ttgtaaaaaa gggaggcccgc gatggggcag
61921	caccataagc gaagagatga attaatgaaa gcaaggcagg gagaatcaaa tgagtgggt
61981	ggaggaaagga ggctgtgact tccttcgctg ccggaagag aactagaata gcctcgggct
62041	gtggggggag gtaagataa agtgactict gggccctggg ggaggcccag gattttctac
62101	cgagctgagc tgggtgccic tcccaaatgc ccaaccccc gagagtgcag gggagagcac
62161	agcctggcca aacctgggca gggcacacgt gtccctcacc ccacagtggt cagagccca
62221	gcgtgtgcc tgcgttggc gggaaacaca gaccctaca cccacacaa ggtccggcc
62281	gccttcaaat aacagcagcc gtgccctctg ggccggtgac ccggacacag agagatgaag
62341	tccgcatctc tcagagtgcg ctgtccctcg ccggtcagg ccgggtccc ctgtctct
62401	gaggtacca ggaggatig catgtgggtc tcaggacac aggttcagt atgtgacaga
62461	gggtagtggg tcccagcagg gccggtctt ggaccggtt tctgaaaag ccagttggcg
62521	acctggggtc acagcaaac tgatcctgtt tggccaggag tctccagtg acggcctccc
62581	ccagaacatc gggcccagtg ggggctccag ggggtagact tgcctccag ctacgcccg
62641	tgcttgaca agtccatgat ttgtaaaat taatttgtt tggatggagt tgatttagt
62701	gtgtgtgagt ttctgtggc cagcaaacg aatcagttac gcatacacat gtatccagct
62761	cttctacga ttctgtccc atataggica ttatggggg tcaggtagag ctctctgtc
62821	tacgcagtac ggccctatic agttcagctc agtcgtgtcc gactcctgt gacccatgg
62881	actgcagcac gccaggctcc cctgtccatc accaactctt ggagcttatt caaactcatg
62941	tccatcgagc cgggtgatcc atccaacct ctcatcctct gtcttccct ctctctctc
63001	cttcagctt tcccagcacc ccttagaaa gggatggca aaccacttcg gtattcttc
63061	cttgagaacc ccatgaacag tacgaaaagt ccttattag ttctatttt atatagca
63121	gtgcacacgt gtcagccca atctcgcaat ttatcacc cctccggcgc cpatgttag
63181	tcattttgt ttctacatc tgcgactcta ttctgtttt gtaaacagt tcaattcac
63241	cacttttta gatttgcac atactggca agcccacagc aacatgctc aatggtgaa
63301	gactgaaagc atttctcta agatcaaaa caagacgagg atgtccactc actcctttt
63361	tactcaacac agcccigaac gtcctagcca tggcaatcag agaagagaaa gaaattaagg
63421	aatccaaatt ggaaaagaag aagtaaaact cactctttgc aatgacatg acacttatac
63481	ccagaaaac ctagagatgc taccagataa ctattagagc tcatcagta attgttga
63541	ggatacaaaa ttaatacaca gaattctct gcattcctat agactgaca caaagatct
63601	gagagagaaa ttaagaaac catcccacgg catgaaaag agtaaaatc ctaggataa
63661	agctacctaa agaggcaaaa gacctgtact cagaaaacta taaaactg acaaggaaa
63721	tcagacgaca cagagagaga gagatccac gctcttggat gagaagaatc gatagtgtga
63781	caatgactat actaccaga gaacataca gattcagtac aacccctatc aaattccaa
63841	tggcatttt cacagaatca gaattagaac aaaaagttt acaagtcca gggaaacaag
63901	aaagatccta aagagccaga gcaatctga gaaagaaaa tggagctgga agagtacggc
63961	tccttgatt ctgactgtt atacaagct ggcattgatt ttaacagcag ggggtgaaat
64021	gaacttttc acaaacaga tgggtgggtg ggcttccctg gtggtcagc tggtaagaa
64081	tcctctgca acgcagaga cctgggttcg atccctaggc tgggaagatc cctggagaa
64141	gggaaaggct acccactcca gtattctggc ctggaatatt ccaaggacca tatagccat
64201	gggtttgcaa agagtggac acgactgagc gacttccat cctggaaacg tccattgtg
64261	gacggtgaac tgggttgtc caagctcagg gtaaccgtt cctgagtac tgacactct
64321	tcctatgggt taaaatgtg gcccaaggc caggaccaga ccccgagtc agccagcag
64381	acctgtgca gcccagcga gtgtgtggc gccgtgagt tcttgccc catggccctc
64441	gactggagcc cctggagtga gccctccc tcccagccc tgaagagctg ggtcagccc
64501	taaccatttc ccaccagtg acagatccgc ctgtgtgaa acctgctctt gtcccagg
64561	aacctggcag gactcagga gaattctca gggcggccac agatcagggt ctggggggc
64621	aggctgggt ccagcagag cctgtgccc actcccggg aagagcagct gatgtcagc
64681	atgaccacc aggcaccga cgcgtcttg cacacagcc gccccctcat ggtgacactc
64741	tttctgtg gccatctc gccccctag gtccctctg ctcccagct cctggcctgg
64801	gaacctctc ccccccgg ggagctcagg cctgtgtcc actgagcact ccatcccgg
64861	gactgtgtg atcaccaga cctgcacccc ctctgggtc tcaccagat gggcaactcc
64921	tgcccattca gcaccagcc tctgggtac aatcggggg aggagggaga agcctgggccc
64981	agaccccag tggctccc aaggaggaca gaaagctgc cgtggccag ccgagagcag
65041	ctcttgaga gactgggac cccagaccac ctgtgagcca cccgaggt ctctgtcac
65101	acgggccacc agcccagcact tagtgggac gagggtgagt gggtagggc caggtgcacc
65161	agggcaagt ggtgagccc gattggacag ggtgagggg tgaggcccag gtgaccagg

65221	gccccatgtgg gtagggccc ggtagaccag agtagcggg tagggcccag gtagacagg
65281	cgagcgggtg aggccagggt ggacagggcg agcgggtgag gccggggtg acagggcgag
65341	cgggtagggc ccgggtggac agggcgagcg ggtgagggccc gggtagacag ggcgagtggt
65401	tgaggcccgg gtagaccagg gcgagtggt gaggcccggg tagacagggc gagggggtga
65461	ggcccgggtg gaccagggcg agtagggtag gccaggtgg acagggtag tggtagggc
65521	ccagtagac caggggccc agcaaaagccc cggctcagca gtagttct gaggcccac
65581	tgctgcagg gacctagcg atggttaggc agccctgtg ggggctccc actggggaca
65641	gcatcagag agcgagtggt cccctggaga aacagccagg gcatggccgg gcgcccgtcc
65701	aggctgccc aggggccaca gctgagcccc gaggcgcca gggccggga cagccctgat
65761	tctgggttg gggctgggg ccagagtccc ctctgtcag ctggccggt gacagtgcg
65821	ctcgtctccc tggggcccg gaggggacgg taggtgaa aatggacgt tggggtctc
65881	tgggtgtgac agtgtcgc attggcactg gctgttgg gccagcagc ctacggccc
65941	caccccggg gctcccacg gcccgcac cctacccca cgcagctggc ctggcaaac
66001	caagaggccc tagcggcga aatagccagg aaaccccac cagcccca gccctggcag
66061	gctgctc cctcctccc gggtagggg aggggttct ccagttctg aagctccac
66121	cagcccagct ggagaaggc ccacatcca gaccaggc cggccaggcc cctgttcca
66181	ggcctggcg cctgagacca cgtcctcag aagcggcctc tctatcca ctagctgtg
66241	totggatcc tggagtcac gggcctctc ggggcccag gagccatct aagtccagg
66301	ctcagagctg aggtctcgc gggacacaga gtagctggg ctggcctagg gaccgcggt
66361	cacactccc ctgcccctc tacttggga ctcttggcg gtagggactg agccaagt
66421	ggggatggg agaaaaatgg gacccctac gatcactgc ctgggagccc tgggtcgtc
66481	ggagtaaca tgggtgact cgaagcacag ctgttccca ctaggctca caggctctt
66541	ctccagggga cgggacctca gatggcagc cactatcca tccccacga ggctcagc
66601	ggtcctctc caggggacgg gacctagat ggccagtcac tcatcatic cccatgagt
66661	ctacagggt cctctccag gggacggac ctcatggc cagtcacta tccattccc
66721	acgaggctc acagggtct tctccaggg acgggacccc agatggcca gtcactc
66781	catcctctg tgcaccatc cgtccaacca tacccttcc ctccatccat ctgaaagct
66841	cctgagggc tccccggga cccagcctg atgggcccct cagctgctca tcccaggcca
66901	gtcaggccc gcacagtaaa ggccaaagtc agacctgaa ggtgcttct taccacggg
66961	aggaggggg ctgtgacac agggcgccc atgcccctgc cagctgccc cccgtctg
67021	gcccagatc tagggcaac ggggggacg aggtgggac agacagggca cgtggggg
67081	ccagctccg cctgctcgg ggtgagcaga ctgcccctc caccaggt acaggtctc
67141	ctgatgccc ctgcccctc tgcctctc tccgctcca atcagagagg tcccggcatt
67201	ccagggtcc gtagctctc tgggaataa aggtgggaa caagtccc gcacgtctc
67261	ctgagcccac ccccaaac acacaaaaa atccctccac cgttggact taccagctc
67321	gtcttaggg gactgccc ggggtcccc agcccagga agccagggc caggcctga
67381	agtcacagc cataacaca tctcagctg cacagagaga cagtcttg tagcacggt
67441	ccccacctg cgagcctgga gagtgtgccc ctgcccctc ccagccggg tagtggct
67501	cagcaaccg tctcactcc cagcgcctg gctcccctg tggcccagg tcaagctctg
67561	ggggtgaag taagtcagg agcctatcc atgcccagc cggagcccac agcgcctca
67621	agaaatgct ctctccca ttaggaaca ttagtggaa agacaagagc tggggggtc
67681	tgggtctct gggatcaga tgaagggtc tgggagcag agcagctca ggcacccaa
67741	aacaaggccc aggagctgga ctcccaggc taggggacg agggaaggaa ggcctctg
67801	ggggttggc tagcaaaag caccaggtg ggggtgagc acccctggc tggcacac
67861	agccccac tgcagtact tcccctgg agacctgg ctcccctc cggcctgac
67921	tccatctg ctaccacc agaaatcct gagtgggtg ccatgtgact gggcctgcc
67981	ctgggagga aggagatca gacagacag atgcccagg agagagggc gacagagga
68041	tctgggag gggcccggg agcctggg ggcaggggg caggatct ccagggtga
68101	cggcctgt ctatctcg tagcacaga gggccgggt gtcccaggc tgggaacca
68161	gcagagggc agggacggg ctcaaaggc ccaaaggcc agcctgacc agacctgtg
68221	gtccagaag cagctgccc ctgaggccc tagtggccc cgtgtcccga accaccgtg
68281	aaacatggg cacagctcc caggcgagc cactctgccc tcccggagg ctcccagcg
68341	gctcctgct ccatccca gggaggaaa ccgaggcca gatagcaac atcccggca
68401	gcaggtcaa gccagcccct ggggtccc ctcccggct ggggctccc ctctcagg
68461	tggaaaccg agggcacaca ggggtccat ggggtgccc tctccagc cctgacac
68521	ccgggggta ccccgcctc tcatccca gccctggag gcccggaca cccgtgat

68581	gacccccgc tetatcatcc cagccctggg ggacagatgg gaggcccaag cgtggacccc
68641	ctggccaccc cctaccccac agccgggagg agccgggagc tggtgccaa gggcctagag
68701	gagccagann nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn
68761	nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnca atatagaggg
68821	ggtgggataa agggtaatat gatgttagg tagttagat taattagaa gggttggat
68881	aaagattaat aaaattacaa gcgtacatat cgtgtgagtg tgggtgataa tatttgtga
68941	tgtggggaat agaagtgagt gtgagttaga ticaagatgt aagtgtgca atacaggtct
69001	gagcgatttg aatggaagtg aaaaaagcg tgtgtgga ggaggcggga gaggaaata
69061	gtgtgggga agaaaagaag gctagtgggt aaagaatat cagtggcgg tgacgaaag
69121	aagaactagg aagaattaat ataaaaata agggaggatt aaaaaataa gagggaggag
69181	gtaacgaaa tagttagta agaaaagaat ggagagtgga gtaagataa ataaggagt
69241	aatgggagtg agggagata aataaaaaa tggtaggga aatagagta gaatgagaac
69301	aagaatgaaa aaggagtgta aggggggtaa aaaaagtaa gtgaaaaa gaggaaaaa
69361	aaggagaaga taaaaaata aataaaaaa aggaaaaaa agaaaaaag aaagaagggt
69421	taaaggacga aaagaaggga agagaaaaa aatagtttaa gtggggagg gtaaaaaa
69481	attaataag taatatggt tgggtgcaa aaaaaaaa aatigtgt gtgatgaga
69541	agaaaaaaa aaagaagaaa gggaaaagca aaagaagg agaaaaaag acaaccccac
69601	cgcccggcg catggagggt gaggatggcg cagcccgcg gatggcacag catcacgca
69661	atctaaaac gtttcagac cgtgtcact tcaccgcgc cgcgccgc cggccctcc
69721	tcccgcctg accgcggacc cccaccgca cggggagcc taccaccac cggggagcg
69781	tccgccacg taagtcagg actgcccga agacgcgc gggtaaaa gtttactt
69841	catgacataa gcgagtggt tgaacagg ttacaaacc ctctgaga cgcacctta
69901	ggttaggt ttgttttt accatgtac gatcaacta ttcttctt cttctaca
69961	gtgctagtc gccccagag cgagggtat cttgtaca gagacctg gaacatcgg
70021	aggtagttt ccacctagg gtaagcgag aaggctcatt acgaggcgg gggctcctg
70081	gggaaggca gggccctgg gcagaggct tgcacctca gtgacagca gaccacgcg
70141	ggctcagc cgcgggctc tgaagcagg caaagcccga tctgtgaca tcagggttc
70201	cgcagcagc aaggtctgg ccgacctgg cccactgga ggggtaagc tctcctcc
70261	gacgacgca ccaagtcag gaaggccac gcagacact gtgagacag gccccccg
70321	agctgccga gaagctcga cttgacta aagatctct gcgcgtcca aaagtgaag
70381	gcctctctt ctttactt aagacttga tattttacg atglaataa taccagaag
70441	ggctttaa ttacagaca ttaggataa ttccccct agccctgt gcttgtta
70501	gtaacgaa tcacaaccga aataccaag gaatttcca aagagtca aaagcgtta
70561	tcagcaata ctgactgt gcatacata tcaactccc aaacaatag ctcctgtgc
70621	cagtactca aagtactact tactgacga aaacaatct agtctaagc ttttcaaa
70681	gaaactccac tctccgca acittcaga aacaaccact cgtacagtg gcaggggacc
70741	gtgctggac tgggtctgg ctcctctgt gaccaggca cactgcccc tctcggct
70801	ccctagcct cttgacaaat gttcagc tgaagatc accccagag ggaccact
70861	ctctattc ccagtaact accccattt aggagttt ttgtgacag ttctgcat
70921	ttcatgat ttagaggtt acataatcag gctctgaa cagcatgaga gacgtgcca
70981	caaggctct cctgacact gcgcagggg caggcgagt tatctgctt gacgtggt
71041	accatcagg ggtaacaca gttccagga cgttttgac aagacactga cccgatgac
71101	cccactacca cgtgaggt cctgagcc tcccagcct ccaggccct cccgaggtc
71161	ctcggaaact tagggactc ggtctgccc cctgggtt cctgacca gctttgcc
71221	cctctgacc caggttccc aaatgaaaa cgaaggtg ggtatgaa cctcctggg
71281	tcctctcagc tgtcctctg catgtgatg acgctgccc atcgggggg gcaggactg
71341	ggcagctgc gcacccctc caaggctgt accccagat ggtgtgggc cctgtggca
71401	cgtctgctc agcgaactc ctgaaacca gcctctcc tctccggg gcaaccggc
71461	cgggagccaa gcaccactc cgtcagaga gctctgct gtgagtggac gccagttag
71521	ctctgaacc tcccaggcc tctgaggc tgaacattt aaaaagc cccggacggc
71581	aactgctct cctcctgccc gctgtgctc cataaactgc atctaggac aaatctct
71641	actcaccagg gctgaacag aagactgca ctatcttt caaatctag gtgtctaca
71701	ggcaagtc cagaactgt ctggcctaag catctatca gatcctgag acaagagctg
71761	tggagccaa cctggagca gagctctc cgtctgccc acctggacc gcgtccacc
71821	cagtaaacgc agcttgatt tcaaaagta ccaccgact agagccaatg ctaaacggac
71881	cactttct gccattaga tgggtgaag gtttttaa tcaatctgcc agtcaccaca

71941	tgccgcctct gtgccacag gctggcgaag accttictga gctacggcat gtggcaggca
72001	gcggcacctc icttcagtac ggccagctgt caaggggagc gttictgtga tgatgtgaaa
72061	atacattgca tcggccccg tgttcatga acacgggtga ggaaggaaa cacacaaagt
72121	tctgatgca ctgacagcac gggtctcata actcaataca agtcagacaa accacaggga
72181	gtcacaggga atcccaatag cctcatctag tgtgaccatc atgaggctta attattcag
72241	tgtattcaat cataaagagg gggaanaatt gtaaaaaaa aaaaaagaa agagtgaat
72301	gtgtaatact gaaaactgtt gctaggaga gcaagcattg gcgtttgtaa ctgcttgac
72361	lccccaagac ccacactgc ctgctacaa aaggaggga ctgctgtca gtacttgac
72421	accgaactg cggatttga attaaaaa gtgtgtgtgg acacagcaca agccagagac
72481	tgccaaagt tgaggacac tggagaact taatatactt ggtgatgct gccagtaca
72541	gtcagtacc agctgattca atagatgcc gaaaggcac ctittaggta aggatgagg
72601	ggctctggc tegtactt gcactaac agagtatg cgagatacc gaagtccag
72661	gtccctcca ttgctgat gatctagctc agggacggct ggccctagc catccaaaa
72721	tcaagcatg ttctccaac ctgtcttc gctgataatg gaagtcaga acgcccacc
72781	gcccactca aagtcanaa acaccaagc ggtgagtc cactaagctc ggtgttcca
72841	atcagcgggt tcaggatcc agctgggca atgaggagg gaggctgca gggatccaa
72901	acctgcccc gtgcgagca agggataacc caacacccc ttctgtacg tccgctgga
72961	gttgtggaac tcagcgcgga cccggggca ccgcgacccc cgggacctg gccgcccgg
73021	gcatccccg tcccgggaca cgggtaagc tcccacaaact gccggacgc gggcggggc
73081	ttctcgcca cgcgccgata ggccacgcc aaggacaagg atgtctgtc ccagacggc
73141	ggggcgggn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn
73201	nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn
73261	ggcggggcg gggctccgc cgcgcgtata ggacggtgt cggccggcct gggctccgg
73321	cgggaatgac cccgccctc cccgatccc gcagcccgc cggcggccc tctccgcgc
73381	accgcctgc gcaccgccc cctcggccg cggccccgc cccgccccg tgggcccag
73441	cggcctgat ggcgcagat gcgaccacc cggccggagt ggcgctggc tggctgtgg
73501	gccacgtct ggcagcgt ctgaccggag ccttcagtgg gggagctca gagcccccc
73561	agctctgag ccagcaggtg agcaaggct caggggaac tgggcccga cacagagcc
73621	cagcaagaag gatcactg gtcactggc tgtggcctg ggtcctacc aggcgggctc
73681	tcccaacca tcccctgag ccaaggtccc tagaacccc tggcagaca ccaaccagc
73741	cttaaatat ggggaaaca aggtgcttag ggtcagaga tagccctagg tcccacac
73801	ctagtagaag ggaggctgt tggagtctt gagtccgc tctccacc cccggaggc
73861	ccctctga gcccaagggt gactgtagt cagtacitt gggcctggc acctgtacc
73921	cactggcac cccaccagt ctgaccaca ttgggctta gtacgggt caggatcat
73981	gaggatcaat gtgctgagc caggaggtg ttagaacctg tggcctgga gttcataca
74041	gcactgccct ggcctttct agaccatgt cccgctctt gccccacctg cccctgtcc
74101	cgcacccac cagcagggc agggctctg agaggctgt gggctacc tattcaggg
74161	atgagccgc taagacctg ggcactgc ccgtaggga cccctgagg accaggccg
74221	gggctctgc ggagggcag ccgccccc cagcttggg gtcctctcc ggtgcccag
74281	cccagctga tccgctgce tcccagctg tcccaggc cccggagcgc gcccccgc
74341	agcccctga gatgggccc tgtgctatg agatcaggca gttctggac tctccaca
74401	cccagagca cctgacctg tigtgggct tcagcaggc cctgaagcag tcaagtaca
74461	accagctga gggctgctg cccgactgc gccagggtg gaaggcgg cccagctcc
74521	cactctctg gggctctcc gctattcca ggtgctctg cactcccat gttctccga
74581	ttctctctg tctctctct cctctggt gctccttgc ctcccagg ctcccctc
74641	tcctgtgtc tctgtctct cccggcgt cctgtacct cggcctgacc tctctctct
74701	acaggtctga gctccctcc ctaagagacc agagcagatt ggttgccag cctgcaccc
74761	acctgaccc cctccacc gacagccga ccatgactg agatgtacc cacagctg
74821	ggaaccagag tgaggagg gtcctcaac cacagatga cctgagatga aaagtgcaa
74881	taanaacct ttatttagc cgaacctgt gtgtctctc ttgttgact gttctgggg
74941	ggcgggggg agggagatg aagtcacct gcggggggg gtgccaccc ttacgtgct
75001	gccccctg gggagggtga cctgtcatc ctgctaac cgacggcag cgcagaccg
75061	atggtaggc actaactgt gacctaac ctaaggcgc tccgactcc gccagctga
75121	gacctggag gagcgtgcc cctctctc gttctgggg gccctcgtt ggcctcagc
75181	tcgtctgt accctcccc tctgtgat gcaattccc cgttaattga gaticagc
75241	gaggatgct tgggcttt gacctgacc gcatgagc aggtcacggc cagccccct

75301	ggatctcagt ccagctcggc cgcitggccg tgacgttcca ggtcacaggg cctgccggca
75361	cagaggagca ggcccttcag tgccgtcag cactcggagc tgctgccccc gctgagtca
75421	ctcagtgtct acgcacagag cgcacctgt gtaccaggcc ctattccacg tccccagtc
75481	accgagcccc cagggctggg ggggaactgc cctcgggtac actgtgtccc gtcacgtggc
75541	ttacgtgtg tctctgaggg aggctggcat tgcggtccac ctctcagcac aaacatctgt
75601	ccccctggaa gggggctcca ttctgggtg cgagcagccc cctggggctc gtgtctctc
75661	cttacctgce tcaaggcccc ggctcctggg tcttgagacg caggagagccc acccctcggg
75721	gctgtggagg gggacctgc ttctggagge cagccggagg gccacggcgc cgcctcggc
75781	cgctgccctg agggagcagg cccgacgcca gcgcggtccc tctgtaggc cggggaacc
75841	ctgctgagg gtgcgggtgg gcagggtccc ctgccccag gctctcctgt gtgagtaca
75901	ctcaccagcc agctctggat gccaccatc cgggtctcc aggaggcact catagcgggt
75961	gggttccct cctcccccc tctgtggagg gaggagctc gatcactggg aggctggtgg
76021	tccgtaccgc cccccccac tctggacgtg ttactacc ccgctgggc tcaggacagg
76081	gcattggatg ggaaggacag ggtgggtcc tggccaggct gggggctctg cagggtatgg
76141	gtccccctgt ctctcttat attcaactg cactgcaggg gggcgcaaat ctggacccc
76201	acttactgat gatctgac aggcataagg tccccctcc tgcagcgggg gctggcccac
76261	ggaggcgcgt ggggaaggcc cctctccag cccctcggc aggtcacca ggtgccatc
76321	ctcagccagc agggcgacgc tgcctggag ggcggagagg gaggcagggc agggctggtg
76381	cgacccccgc tggggcgggg gggccctcag ccggtctcc agcacccttg ctgcccccc
76441	tcaccgtcag ggggcacctg gccgctctgc ctacggggg cggtagggg cccaaggcca
76501	caccaggtgt tcaccagctc ccagcagctg gctgtggag aggggcagag gtggcgcat
76561	ggcaccgcc tccccccag accaggatgc tctgcttcc tccccccat ctccccagc
76621	atctgaagga ctcttccct caccatgcag ccccgctcc accagaagct caggttccc
76681	gcccccttc cccgaagctg caggaccct gaccagcga gagatggac agttggaaca
76741	cagctcccc cagcagcggc acagcagctg tgtggcccag aagagcccc ctgttccct
76801	caagcaactc ccatggatg tcatccatg gacaccctc tccccacac gccctctctg
76861	ttccccctc caaggcagag ggaacgcacc cccactgtc tgctaggaca ggggaccca
76921	cttaccctcg aacatcacct tgataacat ggcctgtgtg gggacagatc cctccgacc
76981	ccaacttcg acctgggaa ggaactggg tggagctga ctgcagggtg gggccctgtg
77041	ggaggtgtac gggtagagag ggtgatggg gggtagctc aagcggagct cctgtctag
77101	tccaggcgtt ccttgcagct agtccaggat cctcagcctt ctccccctca ctggatcagg
77161	gaagactgag gtctccctcc ctgcccccc acccagctc caagctgttc tctgtggag
77221	tggagctgc caagaggtct gagcggccag taccgggta acggggttg tggagggtcc
77281	gggcatccc ggtcagggc tctagtggg gctggagcct cgggcccaga gctgtccaga
77341	gaccagtcc ctcccaccgc cgcgccccg aaggagagac agagctcca ggcggggagt
77401	cggaggttcc tggaggggga gcatctcaa ctctgcaggc ccccttcca ggcgcactcc
77461	cggctcccc gctctctgc cctgtctt gttgaagt gattggcata cagttcacag
77521	ccactcttg gagtgtctc cacactaag atacagaaca tctcctctg ccccccaac
77581	tcccagccag gctgtcacga agaggaggc ggcgcaggg gcaggccct gcactctgc
77641	gttggggtc cacagggtc gtccccgtg cggtagcccc ttctctcac gccaggagg
77701	tccccctg tggaggtgct gttgatccc tgcctgctg ctctttggg ttgttcccgc
77761	atggggatg gatgaagagg ccagtacaga cactgccag caggctctg ggtgaacag
77821	catctattc tcttctga gggcagatc tggagtggg gtcgggacc gtcggggag
77881	agtatcttc tgttctaag aagctgccg gttctcagt gctgtcacc atgtcacggo
77941	ccctctgtg gcttgactc aggagacct ctctcagcg gccctcccc ccagggtgtc
78001	aggcatctg tgcctctg gggcagagc tcagcggcg aggcgggag aggccagat
78061	cccagcagc cccaccagc ttgctctgt tcccggca tcatagctg gagaagggc
78121	aaggagcacc ggtgaagcc ccactggag gacgcactc gatggcagca ggtgtcaga
78181	ggtggcccc ggcagactc ccagagca caggccagt ctctctcc aggcaccac
78241	tgtgtctgg gaccagatc ctgcagcag gtcgggagcg gctgtccca gattccggcc
78301	tgcacctg gctccagca caccctgt ttgtcaagg gttttgtct ttcagccgc
78361	cgaggaggga gcttttgc tgcagtica cagaagtgc ataaagggg gccacagtg
78421	ggagcttat aacattgtg cggagggtg taacagtica gggaggcact tgaggagcc
78481	tctaggggc atggagatg tctaaaatt ggtctggta caggctacag agatgtgtg
78541	gtgtgtgtg gtgtgtgtg aaacctcg agccacagt gtgagctg tcatgtgac
78601	cgtacacagg agacctggt ggaagcagc cacctgctt gactgcact gttgattcc

78661	agctcctgcc ctcaggcggc cctgcggggc ccactggctg acggggagac ggcaccgccc
78721	tccccgctg tcagggtggg gggcctgacg attgcatgt cgtgtcaggg tccagcggcc
78781	tcccttgcgt ggaggtcccg aagcacctgg agcgcgccc gcagaacagc ggactcctgc
78841	ctgcctccct gcctctggcc atggcctgcc cgcctctggc cctctttctg ctggggccc
78901	tcttgccagg tgagccctcc caaggcctgc ctcacctagg ggtgtgtaag acagcacggg
78961	gctctagaag taatcgcgg ggaagtaaat cgtagtgggc aggggggatg gtttccgaag
79021	gggccctgag ggggacagga gacctggcct cagttcccc actggtgagt gaccagatag
79081	ccagggtacc ttggactct gactctgggg ggctctcaga gactggctc ctactcagtt
79141	ttcagagggg gaagctggtg tggcctgtc actgccctgc agggcctcag ggacaagcta
79201	tccctgagga ggtctccagc agtcagtggc cggaggctga gccgatgat atagtaacag
79261	cccaggcggc ctcttggggg tggtcagcct gtaccaggt ttggacagc cgaagtgc
79321	ctaagtgatg ggggtctgca gagcaaggga tgagggtggg cagcaggagg acccagagcc
79381	caccagccca cctctgaat tctggacct tagctcatg tggctcctg ggaagacggg
79441	gcttaagggt tgcctgctt gtggcccaca cagtctgat tccacagcac tgctctgag
79501	cttttgggag cagattccc cggggagtct gaccaggtct ttgtgggca gggcctggag
79561	ggaaggggcc caggccagac ctgagtgtgt gtctctcagc ctcccagcca gccctgacca
79621	agccagaagc actgctggtc ttccaggac aagtggccca actgtctgc acgatcagcc
79681	cccattacgc catcctggg gacctggcg tctctggta tcagcagcga gcaggcagcg
79741	ccccccgct gctctctac taccgtcagc aggagcacca acaccgggccc cccggcattc
79801	cggaccgctt ctctcagct gcggatcag cccacaacac ctgcatctg accatcagcc
79861	ccgtgcagcc cgaagatgac gccgattat actgctttg gggtagcta ttctaggggt
79921	gtggatgag tctctccgt ctgctgcca ctctactcc tgacctggg accctctctc
79981	tgagcctcag tttctcct ctgtgaaat ggttaaac actcaccatg tcaacaataa
80041	ctctctgag ggtatgaga tccctgtgce tgggggtg gggtaggga tggctctggg
80101	gattactgca gaagaggaag cacctgagac ccttggctg gggcccagcc tccccaccag
80161	ccccagggg cccagactgg tggctctgc ctctctgta cgggaggagc tggagtgaga
80221	gaaaaaggaa ccagccttg ctggctccgg ctctcagtg ctggtgggt tccaacactc
80281	aacgagggga ctgaccggg tctcgggag cccctgcta ctctggggt gggcaagggg
80341	gcaggtgta gtgtgtgt ggggtcaga cctcagagg cacctgaagg caggtgggca
80401	gagggcagg gaggcatgg cagcagcct cctgggtag agaggcaggc ttcccaccag
80461	aagcagaact tagccctgg aggggggtg gggggtgaa gaacacagct ctctctctc
80521	ccggtctc taagagcgc cacatgaaca gggggactac ccatcagatg nnnnnnnnn
80581	nnnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn
80641	nnnnnnnnnn nnnnnnnnn nnnnnnnnn agagggtgg tgggtggaat ttaataatg
80701	ggtgcgctg gagcgtggc ggcgcatca aggcgctcat ctaaaatagt ggataggggg
80761	tgtgtgaca ataacgggtg gtgatgtg ttaccgggg gtcaaatagt tctagttg
80821	tatgtctt ctgatggg ttccggctg tggacctag cctttagtat tgggggggg
80881	aaaagcagtg agggtagtag ggatgggaaa tattgtgga ggtctttg tgggtattt
80941	tttgatata tgttgggtg tggagtgtg ggttgggtg aatttcgct gcggtatgtg
81001	ttttttct tttctgtc gtgggtggg ttggtgtg ctttgtgtg gtgtggtt
81061	gtggtataa aaaaaatgt tggttgtct cagcttagcc ctataacggt cggctttgt
81121	tctgtttg tcttgggg tgagcggatg gctcgggct cctgtctccg cggcgcggcc
81181	tgcgcgccc tctgtctcc gctgtctg ctgctgtgc tcccgcgcc gccgtgtgt
81241	ctggcccggg ccccgggcc gccggtagt gccgcgct ctcagcccc cccgcccgc
81301	cccgcctcc acgccaggg gcgccggct gcagagctgg atccaagggg gtgcccgga
81361	gtggccggc cggcccggt accccgaac gctgtctgg tccccggg gtgtgtgga
81421	tagtgactt cccgtccct gaagtatgca agtgaagccg gcgccggat cgtcgggct
81481	ggtgtgag cggcgggac tggctgggc gctagacgca cgcgccagc cccccagct
81541	ccagacctg cactccgag cccgcccgg cgcgctccg ggtgtgtg tgtgtgag
81601	gggagggaca gcgggagtg ctacagggt cccgactcag cgcagggaca aagaccggg
81661	ggtccccagc tggcctcagc cgcaggtgt gtggcctcgg tgagcacacc tccaggcggg
81721	aggggtgag gaagcgtgt ggggagggca tgggggtct gagcctgga gagacggatg
81781	ctaccgctg gacctgta gtggcggat tgggaggcta tggaaacagg aggcagccta
81841	agcgtgag ctcgggtg gccctgggg ggtgtagg ggggagcc cctgtgtgtg
81901	ccagcctgc tgtccctaa aggtcgcgc cccccact gctgggctt cgggggacca
81961	gtcacagct agctactc agggcacag cccccgga gcccgccca cggggtgtg

82021	ccgctgagcc tccagcctgt cggggcaggg gtgggggca gggatgggg cgttagcggg
82081	gttgggggca gacgccagg cagactctt gggcacagct ccggtgacaa gggaggtctg
82141	gcaagcctgg gcccctctg tccagccacg ccagctctgc cctggccagt ctgcccct
82201	ggcagtgctg gggatggaag ggggagcggg tacctcagtc tgggggcct gcctctccc
82261	cagccccgc cggcccccta ggctagggg cagagctag ggtcaccct ggggagctgc
82321	tgaatcccg ggttaggaa ccggaggac ctggccttt gaaccagtg gccctagtg
82381	agccctccgg cgctcggta gccctaccc ccagcctgt ccagtgggc gggtgggagg
82441	cgacagtgcc cactgctgg ctgaacagc tctcaggga ggccaggaga gctgggcaca
82501	cggacaggt ccaicacct gagctgccac tgtccactt gtgcggggtc agggggggtc
82561	tgaagcggg tgatctgt cacgccacag atatcaggg ggcactcgg gtcgcctcg
82621	acatgctat ccttgacgg ctgttgca ggcggggaag gctctgtaa tattatcca
82681	teccagtc aagcttccag ggtgatga agcccccg cccgccact gtgggggacc
82741	ccgcttccc tctggagcc agcggggta gggggggg gagatggacc tgcctgcca
82801	ggagcaggg ctgtactct ggcaggtcac ttagctctc ttagcctcag gggggggc
82861	tgaggtgtg cggatctct ctgccttcc ccagcctga ccagtctc cccctggg
82921	tcgctcctg cccaccag agggggggc tatggggacc tggccgatg gcaggcaggc
82981	cggagagggc atgcccgtc cagcctgccc cagccttcc cagtcaggg gccccgcca
83041	ctcccagcg ctggctgct cccatttcc cgttgca ggtggccga ggcctgaccg
83101	agccttggc tcagctggga gactgaatt cccaagcaat tctcaagga tgtgtaggc
83161	tgtgtgtg tgctatcc ggagaggtg ggtgagcga ctggccact ccgcccagg
83221	caggcccag gagacgtg ctgacagca ggcaggcct caaggaggac gacagccat
83281	ctcaggaat tgggtttg agacaagca cagctgggg ggtgggggg ccatgggtg
83341	ggagcctga tcccaggct taggtccagc tctggctcc ctgcccgtg gacctggg
83401	caagacctg acctctg gccccgtc tcccctgg aggtggggc atgctgctc
83461	cccaatccc caggctgt gatgagcag acgaggtgt tctcatcc caccctg
83521	cttccagca gcccggggc gggggggg tgggactg cgcaccagg tgggatcag
83581	gcttgagc tagggaggc ccccagccc caggccaga aggacacgg gagacagaat
83641	gcaggaggg ggcagagcagg gggccagcg tgggaaact gaggccaaga gctgtggc
83701	gatgtctc aggaaggac ctgctgct ggggctgga tctagagcc tccaggagc
83761	gtaccatga cgtggcagg gaaccggag cccggctg caggtggacc cggcgcagt
83821	cacttctc ctctgccc gagactcc ttccagctg cgctcctg tctaatg
83881	aagctggg gctggggg caggtgggg ctgactcca ggtggggg ggcaggaat
83941	tggcagagca gcgtccaga gtgggaga ccagccatg gggggact tctcatgcc
84001	tgctgcccc aaggcgta tagagagag tggttacc ctcccatg gcccgttc
84061	catgaacag atggaaagt ggggctgag agaaggctg gactgccc ggtctcct
84121	ggcatggaac tgggctgct gactcagg ccgggatct cgtctgca ctgacacgc
84181	caggatcag ggtctggc ctgacatg cgcctctg gggcaagaga ggaagcacg
84241	ctggcctgc ctgaccc caccacc gttgctgt gctaggcc tctggggc
84301	agggagagg ggaattca ctgctgga gctaggccc tggctctct gggctcgg
84361	gggaacaatg cagccctgt ctctgagg aggtcctg gacctcacc aggttgagg
84421	aaaggattc tctctctt ggaggtcac gagccgacat ggggaggagc agggcaggc
84481	ccggggcca cactcagt gtgacctg gactgtgt ctcccactg acgtgggg
84541	tgggggtg gggccgggg ggtaccag accctgcc ccaattgt tggagacag
84601	cggtaactg gctattcc ctctctc tctgttcc ctggggga cagctcctc
84661	ctggggag gggacccca gctgaaga cagagcag ctgggctag ggtgtgct
84721	ggagcaga gacctctg ctctgctg tggctatc tgggctct gtagtggca
84781	ctgggggg agcctggg gtctgtgt gactctgg ggtccagg cctgggag
84841	tggcagag ctgagcca gggggggc ctggggccc gctctgcc ccaaatag
84901	ctgggaag ccacagggc actgagcaga caggccggc cagacgggc ctgagctc
84961	cggctctcc ccagctcc ctgtaccct cacctggc cgggggtgc agggcccc
85021	ctgtgtct cctgtctt gcaggctg cccacggct ctccctgct ctctgact
85081	ccgcttcc caggcaggg aaccgcacc tccagctgg gacggggga ggtgtatg
85141	gccagctc aatccccct ccaccggag agcgtgtc agggccctg gcagggtgg
85201	gaccgacat ctggaaact ccagccacc ccaccatg agggggaca tacagacc
85261	acggagctg tgcctcgt gcagaaact gagaacacc agccgggc aaacataat
85321	aactaataa taaaagttt aaagatgct actaaaaa acaagtgct cccagctc

85381	ggacccagct tcccgggtgcc ctgagtggtg ccggccctgt gctgagcatg gcctggtgg
85441	ttacccccca gatccacact aaagggtggg aicaccoccta ctagttaggt gagcagatgc
85501	agggggggag ggcggcagcc cctccatgct ggtgggtggc cgtgggggtg gtctgggca
85561	ggagccagct cacggagctg gagaggacag acctgggggg ttgggggggc ccaggaagaa
85621	acgcaggggg agaggtgtct gccgggggtg ggggtccctt cgaggctgtg cgtgaagagg
85681	gcaggcgggc ctgcagcccc acctaccctt ccccgccca aacggcggga gtaagtgacc
85741	ctgggcacct ggggcccccc aggagggggc gggaggccit gggatcagca tctggacgcc
85801	agtcagcccc cgcagagcgc ccatgctccc cgaaggcctc cgctggagtg aggtgctgt
85861	gacaccaca ccgctgaccc gggcctctct cccgctcagg atccccccg ccgccacccc
85921	gtgagcagag ggccacagcc ctggcccagc gccctcccc acagtgacgc ccccggcctg
85981	gocaccagg aggccctccc gctgtctggc cggcccagac ctccccctg cggtgtcct
86041	gacctcccc atggggcag tgcccgaac cgacagaag ggttctgtct gtcgggggg
86101	cgctgggaga agacggacct cacctacagg tagggccagt ggccacagc tggccttga
86161	tctccactg ctgtctgaga cacgctggag ctggggggag ggcagatccc tatggccaac
86221	aggctggagt gtcccccaac tccgtgccc actgctcaac accccaacc cacactaga
86281	tgcactccca tgcctccct tgggagcagc gtctccacac ccacctggcc accccacaca
86341	cccgtggggc acggccgta gtacccacg caacctctgc gggcaccgtg ctggggcca
86401	ggccctggga ctctcagta gggagggcaga cacggccctt cctccggggg agcgagggtg
86461	tcccacgcc cggttcagct ctgacaccgc actcgggacc ctacagggga gggaccact
86521	ggggcaggcc aggtgacgc tgggtgacc tggcccctg ggcctgagc tacacttct
86581	ctagtggggc gcgaagatgg gtgtggtgc ccacgtctt gcagcgggga ctctggggc
86641	cagggaagtg tctggggcg gtagcctggg gagcaggaag ggcaggctt ggggtccaag
86701	gctccccac ggtcaggctt gggagggggc ctggggctc ttgggtcctt tccgccagt
86761	gcagaccctc gggccacct aagggcacac agaccacaca aagctgtgcc catgcatgt
86821	ggggagtgtt ggcaccctc agagcacact gggccacat cacgcacgcc tgcacctca
86881	ctgtcatcc ggggaaact ctggcccga cagccagcgg ggcgtgacct acccctgtag
86941	ccagaccag gcccccctca cggcccctgt cctcccagg atctccggt tccatggca
87001	gctgtcggg gaacaggctc ggcagacgtt ggcggaggcc ctccaggtt ggagcagat
87061	cacaccgct accctcacg aggtgcacga gggccgccc gacatgta tgcactcac
87121	caggtaagg ggggctgag ggcacccca cctgggaag gaaaccata tgcggcagc
87181	cactgactt gccctaccc acccccagc aggtactggc acggggaca totgccctt
87241	gatggacct ggggcatct gcccacgcc ttctccca agaccaccg agaagggat
87301	gtccacttg actatgata gacctgacc atcggggaca accaggtag ggcctggggc
87361	cccacttcc gaggggccc tgtcaggcc cgggagcgg gccggggctc tgcctccgt
87421	ggggagctg cgcattggc gctgtctcc ctctccagg cacgcatc ctgagggtg
87481	cggcacaga gttggccac gtctgggc tgcagcac gacagctgc aaggccctga
87541	tgtccctt ctacacttc cgtaccac tgacctcag ccaagacgc cgcagggca
87601	tccagcag gtacggccg cctcagtag ctcccagtc caggcctcc gacctggcc
87661	ctggcacgg ggcggacacc aacgagatc cgcgctgga ggtgaggccc tctcccc
87721	gcccaggct gctctgag ctccaatg gctctctt aaccttgc tctaccca
87781	gcccagccc ccaccggat cctgcccgt ctctttgac gcagcccca ccatccgtg
87841	cgactctt ttctcaag caggcttgt gtggcggct cgcggggcc gctgcagcc
87901	tggctacct ggcgtggct ctgcacct gcaggggtg ccagccctg tggatgagc
87961	cttcaggac gccaggccc acalciggt ctccaaggt gagtgggagc cggctcac
88021	tcaggagact gcaggagcc aggaacgta tggcaaggg tagggacaga cagactgat
88081	gagcagatg acagacggag ggggtcccg agtttgggg ccaggaaga gcgtgactca
88141	ctccttggg cacagctgg agcttctg gaggagcgg ttctgaagc gggagttaga
88201	taaaagtat tgcacccat gaagcagtg tcatcttgc cctagagac aggctctg
88261	ggctcagag tggtaagt acccatga gggcacagct tggagaatg cggagggat
88321	gtgactcag tgtccagag atgggacct gtagcatgcc aaggggcagg gcctgtgcc
88381	tgagagctg cactgggtg ggcagccaag tgcaggatg gagcggggc ccaggtggcc
88441	tcctgtctc tcagaagac ctctccatg tatactccc agogccctg gcattgcca
88501	gtctcttct tggggcagg agtaccnag agccattat actggcctt tgttttat
88561	ggacaacgaa actgaggctg ggaaggctc aggtggtgt gttggcgaa gttggccgt
88621	gggcagccct gttcagcac acaccccc cccaccgtt ctccaacag agctcagtc
88681	tgggttatg acggtgaga gccggtctg gcccggcc cctctccga gctgggctg

88741	caggggtccc cgatccatgc cgccctggtg tggggctccg agaagaaca gatctacttc
88801	ticcgaagtg gggactactg gcgctccag cccagcgccc gccgcgtgga cagccctgtg
88861	ccgcgccggg tcaccgactg gcgaggggtg ccctcggaga tcgacggggc cticcaggat
88921	gctgaagtg tgcagggggc aggcctctg cccagcccc tccattccg cccctctcc
88981	tgccaaggac tgtcctaact cctgtgctc catcttggg gctgtgggca ccaggcacgg
89041	catggagact gaggcccgtg cccaggtccc ttgatgtgg ctagtgaat cagtcggagg
89101	ciccagcttc tgcaggctg ggtggcagct cagaccagac cctgagggca ggcagaagg
89161	ctcggccaag ggtagaaga cctcggggt tcttgggtg ctacagact aaagcgtctg
89221	cctgcaatgc gggagacctg gattgatcc ctgggtcagg gagatccct ggagaaggaa
89281	atggcaatgc cctcgggtac tgtgcctgg aaaatccat ggacagagca gcctggaagc
89341	tccatgggtg cggaagagt cagacacaa ggagcagact cactgttta agggccacct
89401	gaggcttca ggttcaagg aaccagcag tggccaaggc ctgtcccat cctctgtcc
89461	acttaccagg cctgacctc cctgtctcc aggccttgc ctactctg cgtggccgc
89521	tctactgaa gttgacccc gtaaggta aagccttga gggcttccc cggctcgtgg
89581	gccccgact ctacagctg actgagctg ccaacactt ccgctatca ccgctggct
89641	gtctcaggc cctgacacct ccacacagga gaccgtggcc gtgctgtgg ctgtagttac
89701	caggcagggc acggagtgc ggctgctatg ggggcaaggc agggcgctgc caccaggact
89761	gcaggagggc ccacggcggc cgtggcact gccagcagct gctcagact gggcaggggg
89821	gctctggcat ggaggctgag ggtggttgg gctgtgctc acgacgctg tgcaggtcac
89881	atggaaccca gctgccatg gctccatcc acaccctca ggtcgggccc tcagcagggc
89941	tggggagct gagccctca ccgtctcgc tgtgggtcc catagggggc tggcacgtg
90001	gttcagggt cctgcgctc ctgctccca cagggttgg ctctcgtag gtctgctt
90061	ccagtttgg gttctggag acctattccc caagatcctg gccaaaaggc caggcagct
90121	ggtgggggtg ctctctcca gagacctgc acctggggg cccagcata cctcagctc
90181	atcaggggc agatcctcca aagccatga aatgttaca gttgtataa agctttttg
90241	ttttcatt ttaaccgac tgcattaaa cacggtcgt tctacctgc ctgctgggt
90301	gtctctgta gtcaaggcc agtatagggt ggaactggac caggagttg ggaggcttg
90361	ctgggaccc gctcagccc ctggtctca gggctgggtg ttggtcagg gctccccctg
90421	ctccttca tctgtctga atgcctacg ttgcttaca gttctctcc catctccca
90481	cgccctctc agaccgtcgt ccaccaagtg ctgtcacgt ttccgatcc agccactgc
90541	aggacacaga accgaactca aggttactgt gctgactcc tcactctctg ggttactt
90601	gcctgccacc ctacagagc caaggatccg cctgtgatc aggagtagt gaagtgcctc
90661	agccgagtc gactcttgc aaccctatg gactgtacc taccagctc ctctctctat
90721	gggatittc agccaagat gctggagtgg gttgccatt ccttccag gggatctcc
90781	caacctggt ctccgcata gcaggcagac tcttactgt ctgaccacc aggcaatga
90841	ggagacctg tttcagctc tgggtggga agatccctg gagaaggaa tgacaactg
90901	cttcaatt ctgtattgg gaatccatg gacaaaggc cctggaggcc tacagccat
90961	aggtgtcaa gagacagac tgagcaagtc acacacacag agccctact ggtatctat
91021	agcggacct catagctgc atgtatcagg tgtggcatg ggcagccatc agcagggggc
91081	cattctgac ccactgctt gttccaccg atacagggt gcttctgt gttcgggccc
91141	cactcggctg tcagcggcca agggcagggc tctcggagg cacagggcac agagttaagg
91201	aggggatggg gacgttagct cctccccag tctcagcggc tgcagcaggc aaaacaaag
91261	ctaggatcc tgcacaacc ggtagtctt gccatgctc gcccatccc cagagccaca
91321	agaacgggag ctgggggtg gcccgagct gggatactgg tccctgggccc cggccatgtg
91381	ctcggccgca cagctctc cggcgggga aactgaggca cggcgctc cggcttctc
91441	ccccttcc gggctcgc tctctctc tcaccaggc agtattccag cccggctgt
91501	gagacggaga agggcggct tcgagtcagg gcccgggctg ttattctgc cgtgagcgg
91561	ccttccctg tacctcact tgagggcg cgggaaagg cagaaacgg gccgagctc
91621	cttaagggg cctgtgggg cgcggccggc cctttgtcc ggttggggc ggcggcagc
91681	cgcgctcag cgtcaagcc cgcggctgc cactgaggc ggcctgctg tctctcgg
91741	cggcggggc ggcggggc gaggaggca acccaatctg gcttggcaag agactgagnn
91801	nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn
91861	nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn
91921	ggacgtacac cgcggcctgc gtctcacca ccggcggct ggtaacgcc cccgggggt
91981	gccaaggta cgaattgacc ctccccccc cgaacctgt ccctagggt ggggggtcg
92041	ggggcaggt tctaagatc cctggtccg cagcagctgg aactctcag tccctccag

92101	ctctacttca acccgcacct cgtgtccgg aagtccagg tgaggccgcc ccgccccttg
92161	cacttctgtg cccaaccct cccgccagc gctggcctga ccgccccca ccccgccac
92221	cccacgagg ttggaggct catccaac ttctttct tgggcccct gggattcage
92281	ttcttctca acatgctct cgtatcct gcgccgtgt ggaagcggga ggaaggcggg
92341	gcgggggacc gggcgggagg cagcggccc cgggaagctg agaccctcca aggggcacgc
92401	ttctatacc aaagccgag gticcgtac tggccatgc tggaggagg ctcctccgc
92461	ggccgcacgg ccgacttct cttcatgtt ctctcgggg ggtcctgat gactgtacc
92521	ttccgggct cggggaccta tgggtccgg cctctgttg cctgaggcc ctgctgagc
92581	gcatgccaca gaggagagt tgcgacccc agctgagggt gttttgagc gtacatcacg
92641	tgctagctg cagggtccc tgcgaactc cagggtaca cccaaaatc cacaggcgag
92701	ggtgccagg ggctgagtc tgaatgcagg tagccaggag gatctagggc tggcccggg
92761	ggctggggg aagtggagag gcaggccga tcaggggccc cctggaggcc accgtttgt
92881	cttagagtgg gaagcgaac caactgctt gagggttca ggggttagg aagttagagg
92881	ggccctggg agggcacaag acctgactc tggccagct actgggctc ctgggtagc
92941	ttcttctct gggccaggcc ctacggcca tctgtgtga cgtgtggagc ccgcccagcc
93001	ctggggtgag ggtcaactc ttggcctc tcaccttca ggcgccgtc ctgcccggg
93061	cgctcatggg ctttcaatg ctgtgggca actccatct ggtggacct ctgggtgagc
93121	ctgtgtcca gggagcctgc ccaaagctg gtgtgtggg ccagagccct ggtcctctcc
93181	ccgccccac ccctctccc cactctggc gccccatcc ttccagccc tccaacaagt
93241	cagcctatag gtttacta ttgagcctg acccatttg tgacgtgt gtggggccc
93301	accggtagg gatgggtggc tcagggtgcc tgcacacgc tccacttct ctgagctct
93361	caggctgac tctcccag gttctgcta ctctgggcca agcctggccc cagctgggc
93421	tggctggcg tcagggcat cagacccca tctttgggg gcttcagggc tgtggagggt
93481	ggcctggca ttgggcctc tcccacagg attgctgtg gccacgtca ctactctg
93541	gaggagctc tcccaccca gcctggaggc aagaggctc tctgacccc cagcttctg
93601	tgagtctga cagcctccc cacccttcc ccagatggc tcttcccc atgaggggg
93661	gggaccctgc cagctggcg tcagctggg ctctcccca caggaaactg ctactggatg
93721	ccccagagg ggaacccaat tactgccc tcccgggga gcagccaggc cccctgcagc
93781	agtggagag acctcacca gagccgggtc cccaccccc acccctggcc tgcacagcag
93841	ctcctgtcc tggaggcgg gcttggccc agggcccgc cctgaataa acaagtgacc
93901	tgcagcctgt tggccacagc actggcttc ctggcgggc cagcctctcc acgggggca
93961	gggtctgtg gccgagacc agggccacca agcctgactg gctctccgac ccagaacatt
94021	ggcacagctg gaggcccaga gagggtccag aacctgccc ctgcccagca gaactctgag
94081	cacagagggc agcctgtctg gggttctat cctgcccct cctgtccct aattcagct
94141	ccactgatgg ggtcacatc tcaggggcgg ggtgggact gggatgttg gttgtctga
94201	gcttggcgg tggggcccct cctgtccga actagcaacc ccaaaggga cctctgttc
94261	attcccagc caggccactg aaggacgggc cagggtcaga agaggccag gcccttctg
94321	tgactccgaa gcctcaagtg tcagtttg caagttccag tggctgagg agaggcctct
94381	gggaagctc gccctgccc ttgcagctg agggcggcag gagcctacc tggccccag
94441	ctcacgggca ttggaggacc agtccgacg gtggttact cctgggtcgg caccagccgc
94501	cgccggctgt ccttttaca gaggataaa gtactgctc tggagtggga cttaatgtt
94561	gtcatgaaac ctctggcca gcagcgggt ccgcagtggg tggcaggta aggccctcc
94621	ccggccctct ccaggcagg gcccctggc cagcaggga ggcaggcagt gcatcccc
94681	actggctctg ggctcaggc taactctgc tgtggccga acatctccc cagtgtgga
94741	gcccagtgc cgtgaggcca gctgggctg aaaccttct ctctgaagcc ccgtgtccc
94801	ctggccctgt atggaggca gaggctggag ccgaagtcc taggatgtgc ttgcgagacc
94861	cccagccca gggcggaggc ccatctcagc ccaccccga actggaacc ctggagctc
94921	tggccctct ggtgtgaggc cctgtctat gcaacctcag cctgcccagc aacggaagt
94981	gcaggcccg ggcccacggg cttaacgaa ctgggcttg gtcacctgc gggcctggc
95041	ccaggaggaa gaccacggg ccacctctc ggtgccacg tccaggtac gtggggacc
95101	gtccatgta cagaagatgc aggttacc ggtgagctg ccggggccc tccagagca
95161	ccagccggg gtggagggtg gcccagctc tctgtcagg cactgtgtc tggagggtc
95221	ggccgggca gtgccacca gctgcagcag gacagggtgg cacaggccca ccagcagtc
95281	ccgacggga tggcccctg caagggccag agaagccag ctctgtgtg gggctggg
95341	tggactgac aggtggcct gccctctgc cccactact tccagccac ccgggactcc
95401	aaggactgc tgagctggc aggtgggag ccgaggggag tcaactgct cgtggggca

95461	ggaggggcgg tccacagggc tgagccctga gctgaacct ggccctgctc gtggttgg
95521	gggtgggggg gtccagtggc gccctagccc tctgaggcc cagctgggac gtgcgcggc
95581	gagggcgagg ggccagccca tgccatgctg tccccgttc tcagctccat gctaccact
95641	tgaagaaca gaacctgtg ccttttatt tagaaagtgt tcttgccct gccggggct
95701	tctatacaa aaacaacac agctcaact ggctctctc gaccagagac gggcgggtgg
95761	gactggggct cagcagacgg aatgtgtccc cggcggcggg agaccaggag gccctggcc
95821	cgctctcag gacggctggg ctgtccccc ctgttccct ccgagccaga agatggagga
95881	gaggtgggct gatctccaga tgcctccctgg gagccaagcg ccacgggggtg gtcaccaggc
95941	cggggccgtg ttggccagac gccctatccg cctgtgggag ggggagggca gcaacccccg
96001	gatctctcag gcaaccgagt gaggaggcag gagccccag cccctccctc ggccgctctg
96061	ctgctgggg cctgaagtc gtcctctgtc tgcccccct ccccaggag agtgagcctg
96121	ttctggctg tgtcagacc tgcggaggg ccagcctgc ccggggccct gctctgctg
96181	gaagggcctg gggcagcacc ttgtgtccg gtcctgttc cggatcttct tctccatctc
96241	tgcatccgtc agggctcca gcagcgggca ccactggca gctgctctg tcttccggat
96301	ggcaatctcc accgtgggca gggggcttc actgtggagg acgagagagg tagacggctc
96361	acagagcagc tgcaggagag gcccttagaa agcagtgtcc accccgtgc gggcagacag
96421	gacatggagc ctgtttctg caccggctc ccgacacagg gcggccgggc acgtctccaa
96481	catggcatct ccgggtctg atgtggggag ggtccacag gacagtctg caggccagc
96541	catctccagt gacttctg ggaggaggag gcccgtccc cccgtcagt gtcaggaga
96601	aaggagagca aaggagtcca tccaccagg agtggagtc caggccccct gccctgacca
96661	gccctgagg gggccctcgg cccatctac agggggccag aatccataag cctgactgc
96721	tccaccggg gggccctcaa agacggcct agactcctc cgaggggccac ctgcacacc
96781	tctgggaaag tggactcagg gctgggggtc agcctcgtg agcccgcaa gctgggggac
96841	tcttggcca gctgtgct ctgccaggag ccaggcccag cctgcccggc agcctcagcc
96901	acgcccctac cccctcctc cgcgggcca cgtggcctc cgggtctct ccttggcct
96961	cctgtgggc cactgtgtc cagccccagc agtggcctg ccaggagccc tgcagagca
97021	gccccagag ggaggagggg gcccgggga acagcacagg aacaacaga cccctggct
97081	tagtttagc tctcatctg gaaaatggg acagtctct tctgagagg ggttcagag
97141	gaccactgce atgcaacc cagcacacac ccactgcgtg ggggctcgg cccgagccgg
97201	tgccccag tccagcgtg gtggctgggc cggcccagcc accctgcca cagtctctc
97261	ccagccgggc ggtgtgctg cagtccaga gccagcactg cagacccaa tctactctc
97321	cagttgccc gctcccagct gccctcctg ggggcagcag acacgaaagt caccagccc
97381	acggccagc gagcaaac gcttctctc taacaagt cgggtcccgg aggccctgtg
97441	ttacctccc tgtgtctcg ggaagatgc atcccagggg gttgtctaa accaaggct
97501	gctcgggcca ggcttgaag gaggggcctg gaggcaggag cccacctta cgggcactg
97561	gcttctggg tctcaagcc ggctgggacc ctgcatccc accaccgcc aggtgcaagc
97621	agggaggccc tctggagga gccagaggc ctggagggtc gctctcagc tgacctact
97681	ttacaacct cacaggctc gcaggccagc tgggagcat ggctgtccc tctgttaga
97741	tgagaacaag actcaggga tgalcccc tgaactccc caaccaggag gagacaaaac
97801	tgggtgtcgc cctctgctt aagatcaact gactctggac aaggggcca gccaccga
97861	tgggaaagg gcagctctc caacaaggc tcttgggac ggaccggca ggccatggt
97921	tctcagctat gacaccgca gcacaagcac ccagagaaa acagctaac tggcactgt
97981	cacacaagt aactccaac ccaagaaaac cacaaaaagc ctgctgatct tcatatgt
98041	gggaagggac ctgtatctg aatgtataac gaactctga aagtgaaag ttttagtca
98101	tcagtctgt cagctcttg caaccctat gagggtagc tccaggctc ctctgccc
98161	ggatctctc aggcaagaat actggagtg gttccatgc ctctctcag ggtatcttc
98221	caaccaggg atgaaactg tctctctct gcactggcag gcgggtctt taccagtgc
98281	gccactgag tagaaact ccagtgccc ttagtctcag agcaggagg actcggccca
98341	ggctgtgag gggacctct ccagctccc tctgcacag cagttagagg tctgtctga
98401	gtcagctcc agggatgag gacttgggt cagatcaact cccaggacct caggatctg
98461	tctgggaagc gaggctccc aggttggccc caggcccgt ggctcagct cgtgagcct
98521	gcgtggacag gtcctatg cagccctccc acgggactc gggcgcgcc tggaccggg
98581	ggctgccagt ggtcggggg gggccctgt ggggctgt cctctcttg ctccagctc
98641	taggaacatg gtggcgtg cctctgggg tttctggaga agcagctgag atgcaaacg
98701	ccccagcgc tccctcagct gttccctgc acgggtggcc cttgtgtac ggcctcctg
98761	cagggacgt gacagctca gcagccgctc aaaaccac ggggacgtg gcagctcag

98821	cagccgcgta aagcctgaca tccaattigg aagcctcccg cagtggaga gggcccggg
98881	gacggggctg cccggggcga gctccaccgg gtcgggggtc acgaggagcc caccgcgctc
98941	cccgocacca gcacctggga ccagataccc tcccgcctct gaggggggcc tgaacgccgc
99001	ccccccccc gggggggccc accgcctgct cgtggactga acaagaggcg gcagtggcct
99061	ccagacccc tggggggagg gcagacctgt ccgagactga gcacaagtcc agggaatgag
99121	caagggtctc agtaatgcc ccaccgggac gggacgggag gaggcgacag aggccgctga
99181	ggtgcggggc agccctcagt agctggcacc aaggccccag gcagtcccgg ggcacccccg
99241	cagggggcgg gggcgaccac cggcccggag ccaggcagtc ccggggcacc cctgcagcgg
99301	gcggggcgca ccaccggccc gagccctacc tgaaggcgta ggtctctga tgcagctca
99361	gctgtccccg gatgctgtag gcgatggggg tgacgaactc cccgccaccg cccagctcgg
99421	agcacagctt cagagcgaac ttctggggcg agttctctt ctccgacatg tcccactga
99481	actgttccac caaggagatg ttcccacgt ggatgticag ctggcccggg agcacagaca
99541	tgagccagag cggcccctc tggggccagg ccgcaccctc accaccctt ctcgccgaa
99601	catccccgcc tgttcttg ccgcgccct gtgtgtctac ttgggtaag gaaaacaacc
99661	cccctctc tgaaaagggt taactagcga ggaagatgct ctgtaactg gaaaactccc
99721	tacaaagaa gcttggatct gatggctca ctggtgaatt ccaccaaca ttcaagcac
99781	taacacat cctatcaaa tctgccaaa aaactgaaa ggaaggaca catcataact
99841	ccctgccttg ataccaaagc cagacaaaga tactacgaga aaggaaagg gcagaccggc
99901	acttactgtg gacattgatg tgaacctca gcagacaga gcaaaactac atccaccagc
99961	acgtcagaag aatcacacac cgtataaat gatggatga tgacacaacc acattataa
100021	cggggggct tactctgggt atgtaaggac ggctcagtaa gaaaaccggt caatgccatg
100081	aaccactiga acagagtga ggacaaaaac cacacagta tctgataat tggaggaaa
100141	tcattagaca aactcaacg tgccttcag ataaaagcac tcagttaact aagatcagat
100201	ggaaccaca tcaacaagat taattcagtc aaaaaatca ctgcaagat caccacaat
100261	ggcagaagac tggtaacttt tctctaaaga tcaggaacga gccaaagata cccagcttg
100321	ccactttgt tcaatagc gttggaatt ctactcagtg cagtgcagtc gctcagctg
100381	gtccgactct ttgcacccc atggatcaca gcagccagg cctccctgic catccaac
100441	tcccgagtt caccacaact catgtcact gactcagta tccatccag ccatctc
100501	ctctgtctc ccttctct cctgcctca atccctcca gcagtaggc aagaaaaata
100561	aatcaaaagg atccacctg aatggaaga gtaaaactat ctctggccg agatgtaca
100621	atcttatg cagagttaa gatgtaaca aaatactatt agaactatg aatgaatca
100681	gcaaggtacc aggatacaa gtcaactgc aaaaatcagc cgtatttca catgtaaca
100741	ctgcacaatc tgaagaaga aggatgaaca aattacaata acataaaaa gaataaaatc
100801	cttagaaatt aactgatca aagagatga caatgaaca tataaaat actgaagaa
100861	attgaagata taaataatg gaaaacatc ctatgtccat ggatggaag acttaaaat
100921	attaagctt caaggctatg gttttccag tggctatga tggatgtgag agttggacta
100981	taagaaagc tgagaccga agaagtatg ctittgaact gtggtgttg agaagactct
101041	tgagggtcc ttgactgca aggagatcca accagttccat cctaaaggag atcagctctg
101101	ggtgttcat ggaaggactg atgtaaacg tgaacicca atactttgg cacctgatc
101161	gaagactga ctattgaa aagaccctga tctgggtaa gattgaggc gggaggggaa
101221	gggacaaca gaggatgaga tggttgatg gcacaccga ctcaatggac atgggtttg
101281	gtgactctg gaagtgtg atggacagg aggcctggcg tctgcggtt catgggttg
101341	tgaggagtgc gacagactg agcagctga ctgaactga catgaatcc caagcaatc
101401	tacaaagca atgtaatcc ctatcaaat cccaatgca ttctgcaga aacaggaaa
101461	aaaactttaa aattcatatg gaatctaaag aaaaacaaag gatgtctgt caaaacaatg
101521	acgaaaagaa caacaaagct ggaagactca cacttctga tticagaact tactgcaag
101581	atacaaat gaaaacactg tggactaac gtaaaagcag acacgtggc caacgggaca
101641	gcccagaat aaactctca atnagcagtc aatgatit caacagat gccagacca
101701	ctcagtgaag gaaagtgtt gcaaccaacg gtttgggaa aaaagaacc acatcgaaa
101761	gaatgaagtg ggaccctac ccagcccat ctacagaat caactcaaa cagacagac
101821	atatggctca agccataaaa cgtcagaaa aacagagcaa agcttatga tgttgattt
101881	ggcgtgatt tctcagatg gactcaaa gcatgggta taagcgaaaa aataaactgg
101941	acttaccaaa aatacaacac ttctatgat ccaaggacac taccgacagc ataacaaggc
102001	agcccaggga aaggaggaaa catccgaaa tcacagatc tgggaacaga ccgctccctg
102061	tgagatacag ggaaccgata aaaacaagaa aacagcaaaa cccggactca aaaatgggaa
102121	ggactccagc agacacagga gcagacaag ccgccagcag gtcactaat gcgaagcaag

102181	gcccgc aaag gcccglatcc aaggetgtgg ttttccagt ggtcatgtag gaaagagagc
102241	tgatcgtaa gaaagctgag cgtgaagaa ttgattgaac tttgtgttg gagaagactc
102301	ttgagatcc ctggactgc aagatcaaac cagtccatc tgaaggagat cagtcccgaa
102361	tagtactga aggactgat ctgtagctcc aatactttg ccacctgatt cgaagaactg
102421	actcattggc aaagaccctg atgctgggaa agattgaagg caggaggaga aggggacgac
102481	agaggatgag atggttggat ggcacactg actccatgga catgagcttg ggcaagctcc
102541	ggagagagt gaaggacagg gaagcctggc gtgctgcagc ccgtgggtcc caaatctttg
102601	gaccaagcga ctgaacaata acaaatcaac agggaaatgc aatacaaac cacagtgaga
102661	tactgtcac caccaggcag gcgtttca gcggggttcg gggcagggtg tgcctcttc
102721	ttctgaacg ccccaggac gcgggggct gctgagacag catggggtgt gcttggccta
102781	gcctgccat gacaagagt gcagtgtgt cgcctcact gcacctcc tgcctgtccc
102841	accagctggc ccaccttgg gaccaccag ctcccgctcc gtggacggca agggcgagc
102901	agcgccgga cagcccaga acgtgggtcc ctctcagaa gtggccgtt gcccttctg
102961	ggacaagcc ccaagagac agtcttcag agccctccc cacaacacgg accccagaca
103021	ggctcctgt gaggctcca cgcacctcg cactcgcga ccccagga caaggcaggc
103081	ccgtcgggg tgaggagcc cctacctga taatgacgg ctggtctgac tggcttcca
103141	ggatcctgc cgtgggtg gactgatct gctgtctgt ggcagaggca atgcttggca
103201	cgaatgtcag tgggtcaga tccaggtct cacagagaat ctctgagaac atctccgggg
103261	tcactcgtt ctctgaacg atgacggagc ggggaacc ccagtggacc acagggccta
103321	cgctcagct gctcagccc gccctcccc agccttgcct cctctgccac gcccccccg
103381	ggtgagaca ggaacccct gacgacgca gacagagct agtgcagcc agccaggggc
103441	gcggacggac caticatgt ccaggtaa gcatccgca gcttctccc gtaactcc
103501	atgtcagtc gatggggac cagcacctc ggctgggagc cgttctgtg galcaggtt
103561	ggctcgtgt cgtcgaagc tgaaggagc cggcccgct ctcagcaag cgggtgggc
103621	ccctgtccc agggcctcc tctctgacc actgttgcct gagacctcc cagagaggac
103681	ctgtccacta cggccgggc cggcagaaac aggcctggc ggggtccac ggggaggga
103741	gggagctgc cactcggca cgggacaag ctcagaggtt ccctcagga agagaggtt
103801	aagcccaga gcaggcagga ttctccagc agctgtggg aaaaaaggt atgtccaga
103861	gaagaaacc tgaacaan gcccggggc aggggggtt aggagctgt tggagagcag
103921	tgaaggggc ctggcggct ggggggtgt ggggagctc ggtggccaag caccagggc
103981	teccacctg cagcctggc cccagggag cccagagga cggagagca ggcagctccg
104041	cactcacacc tgccttag gatgggaag aggaagaga cggggctgc gggggcaag
104101	gaaccagc agccccgt tagaccggg ggcgagaacc acttccaag aacgcagggg
104161	cgcaatgat gaacaatgg tagcagccc caggcggag gcccggtgc caggccctt
104221	caccagagc ggaaggtcc ctctgtcg cggccatgc ggttctgtt gatgtgtg
104281	gagcaggca cggcgtccag gtgtgtgag ctgtgggca ggtgggca ccaatgctg
104341	ttctcttg ccttctg cctggagac acagagccc gtccgtcag cctatggcc
104401	aaaagccgc cccagccgc caggtgtg ccagtggac cccgccatc cctctgggc
104461	ccagccccc atgggacct ctgtgcgcc agctccggg tggttacc ccaggtcca
104521	agcggcacct gctcgggtc accagtitta gggagagag agaggcagg gcccagcc
104581	cagtctgta gctgcacc ccaggctcca agcggcacct gctcgggtc accagtitta
104641	gggagagg agaggcagg gcccagcc cagtctgta gctgcacc ccaggctcca
104701	agcggcacct gctcgggtc accagtitta gggagagag agaggcagg gcccagcc
104761	cagtctgta gctgcacc ccaggctcca agcggcacct gctcgggtc accagtitta
104821	gggagagg agaggcagg gcccagcc cagtctgta gctgcacc gtctatgt
104881	ctggctgg cactcagaa agagggtcag ggtcacgg ggggtgggc cagatttc
104941	aggagagccc cagggcagc agagaggag ctcaggtaa tggttggca ggggcccag
105001	gtggagaca cagagagg cccgattgg ggggtgccc taagcaggt gctgggagc
105061	cctggggt tgcacctt cagcaggtt gttattcag acgttggc cagcctgaa
105121	caggtaatc ctctggctc cggccttca gggatgaaa gatactag aaaggggac
105181	tcaagtaac tcaaggaact cggctccc agtggggagc ccttctcc aattacatg
105241	ggcgcttac tacaggaaa ataccgaag ccgtttgag ctgaggtcc cggccgggc
105301	tgctcgtt tgagactgt cgtaccctt gggccatc cctgtggcc aaggggcaa
105361	tcagtcgg gactgcag caacctct cagccctgc ccacagctt caccatcgt
105421	gacgtccac cctggagaa cctgacctt cccgggtt cgcgtaaa acgcacctc
105481	caggatggg ggcagagga gaggccttg cctttact cctctctg agcgggggc

105541	cctcgcacc cagtccccg gccagggag gcccttggg gtggggcagg gagggatcca
105601	cacaccaagg ggagccagga ccccccaaa tctgtgccc tgcctgata cccgagacct
105661	ggggaaacgg gggactggg ctgatgggg caggaccaag aactgaggcg gtgagacggg
105721	gtcccacca caggccatct ggctggcagt tictactcc ggctgcagg ccaagaggga
105781	aaaggtgccc cactcagatc agggcctcc cgtcccagg gagggcctac aaggtcagat
105841	ccittgtaac ttccacgggc aaaactggct tcttggcct gtgcccggcg catggcgtg
105901	gaccaccaca ccttcccca ctgagtcc agcgggagct gtcaccagg tcccccagg
105961	ccagccccc cccgccacct tgcagtagcc tctcgtacc aggccaggc tggccggtcg
106021	accctcctg cctgatggc tcaagtggac aatgcgagtc acgttcagc acgtgagttg
106081	gacgggcagc gccacgggg gtccgggcat ccgagtcca ccaactagcc tccctccgc
106141	tgacagagg tctgtcaag agcctgggg gccatccagc cctgtccga cctggccggt
106201	gtggaagagg ggggtgcca cccctctgg ggggtggct gggcgtggg caggccctc
106261	ctaagagtgg agcccactgg tggtttct gcagccccc ctcacacag cagttctac
106321	tgcccagtaa caggaggcta ctgcctagc tctcctcc gtgtgatga ctcaaccagg
106381	agcgttcag gcccacaca gggttctgg ctgctcatg aggatctca agccccatc
106441	acgtcatgt aatctctcc gtaactct ctagggaag ccgctatcc tgcctctc
106501	accgaccac cagggcgaga aaagccatct ccagcgtca catccacaat gggccaggcc
106561	gtgagcacac caccttctc gggaggtgt gggggcgggn nnnnnnnnn nnnnnnnnn
106621	nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn
106681	nnnnnnnnn nnnnnnnng cgcgcccc cccccggg cgcgggacc ccggcggcg
106741	gccccggcg ctggagcag gtgcggggc gcggccgctc gtgagcctc agcccggagg
106801	acgggcccc gggcggcc cgggtcccag gccctgggag ccccggagg cagagtcca
106861	gaggccgga ggaccggga agcccgaga gaggtggaa gcacggggt ccagccctag
106921	gccatttcc ccccaagcc atcggtaaa ccattgtgg cccagataa aagcgtgcc
106981	aacttttca ccccgggga gactttagc ggtagtgc ccctagggg aatgpaaaa
107041	ccaggattta ccagggtgg gtgggtcaca actgccaga tctgagaaa ggggggtcag
107101	tggggcggga agattagtg ggagaggag ttccagaacc caagggaatg aaacgagct
107161	tgagggtgt taccagcag ccgcccctg ccccgtgagt gagcgaaggc tggcccctt
107221	atgtacat ctccagctc ttgctaga aacctagat ttaatact gtggcagctg
107281	agtcaaaaa taaggaaa cccactct tgagagccag gcacaaggc tctgtgacg
107341	ggtctccagg ctgccatt gcagctctg aaacggagg ttttcgaga aggaggtctt
107401	gggtgctct ccagaattg agggggggc gggggaagt aggaccaga agagaggct
107461	tggcccctg caaggagtc actggacct ggagctgaag ccagcaga aactggaac
107521	tcaaatctg tctcgtgc agccaagg cctatgatt tcttggcga cgttcagcat
107581	cttagagga gctggcggg gaggcggta gttcgtggc ggttcagca gggcaggaag
107641	gtgaggaacc tgaggctgt cagagagctg gttggagtga tgcctcgg tggaccgct
107701	ggagaaggcc tgagttaga aggtctaag ttaacggga aggggtggc cagggtgaa
107761	atgggtggg aagttgagg aggggggca gtggagatg ggggtgag gaatggagt
107821	gagcttagc gcttgagga tactgcagt ctgtcttt ttccacct ggctgaaat
107881	tcactgaaa caaacaacc ctgctctgt gacagctag aggggtggga gggaggctta
107941	agagggagg gacgtgctg tgcctatgg cgaatcatg ggtgtcgg cagaagcaa
108001	cacagtatg aattaccct caattaaga tcaagtaca cttaaaacc ccaaacaca
108061	catgtaatg cagtagact ccagtaaca tticagtaag aagaticaac tgggaatgag
108121	ttccgctg actatcctga tgaattccc gtgtctctt gaggccatt ccttgaac
108181	ttcgtgtt ggggaagct gcccttgtat ggagtcctga ggagtaaat agacggcct
108241	gtagaaggcc tagtagtcc ttgcacggc cagatgctca ataacctga gtttcacca
108301	ttatgtacc tcaagagtct ccttgagct tgcacggtt ctgaatggg tcttgggg
108361	ctccctggg gtcaccat ggggtggg gctgagtg ggtgtccc cctctgctt
108421	gtccctgtg gaaccccc tccaccoga gcagcttgc tttgtctt tgtttgtt
108481	tatatctct agattgtgt tcagtctc agtctgtcc aactctoga cccatggac
108541	tcagcacac caggcctct gccctacca tctccggag ctgtcaca ctcctgtca
108601	ttgagttgt gatccgtcc aacctctg tctctgtc tcccctct ctttgacct
108661	cagcttcc cagcatcagg gcttttcca atgagtcag tcttgactc aggtgcca
108721	gtattggag tcaacttca ttacagtc ttcaaatga taticagggt tgaattctt
108781	taggattgag tgactgac tcttgcagt ccaaggact ctcaagatc ttaacacca
108841	cagttcaaaa gcactgctc ttggcactc agcctctt atgatcaac gccacatc

108901	gtacatgact actggaaaaa ctttgctca gagataaitg acttgattga atacaaagt
108961	ctttggcaaa aataaaaagt gtggcaagca gtactgacac aaaagcaagt ggctttcct
109021	ccgttgagtc attattat tcagtggtg tgctcgtga gagacggagc ggctgtgctg
109081	ggagctgggg ctccacttc agaggagccc cggacctgcc ctggggagt tcacaggcag
109141	tgctcgggg ggtcctgcca ggacgcctgc cctgcgagtg cccagtgctg tgatggatgc
109201	gtgtccgca tctcggcca ctggggccac gtcccgaga ttgtccgggt ctgagggtgc
109261	agagaagagg aggcatttg actgagctg gaaaaatgag catgtggcca cgtgagaagc
109321	cagtggtag gggaccagt aggcggagga aagagcgct caccagagt gggagctgg
109381	aagcatgagg gtgtgtgaa gcagaggccg gggacagggc cgcagggccg gccatggagg
109441	gctgtggctg ctgcaggctc ctgagaagg ggacgtgcc atcatgacc gggttaggtg
109501	ttgacctg gtgtccact agaggacaga tgtgtgggg gggagctgga gatgggcatc
109561	catcgggagt cagcctggag agagcagag accccgtcag tggccctca gacgtggat
109621	ggggcggatg ttgggaagat ctgactcct gggtccgggt ggggctccgg gctggagggg
109681	tgcgcccac cagacacagg aggcnaacag atgccctc ccagcaagac cccagccca
109741	gcacctccg gggcggact ccgccctct tccagaatgg ctccctgct gtctcggcc
109801	atcttccg tgccctgagc ctctagagtc tggacaccag cgtccgctt gcgttgtt
109861	ctgggaagt ictggctgt ctctactca cccaggaccg tctcagagg caaggtgtg
109921	tccttggtc catctgctt ggggtccgg tctcgtgc ttacctgt gatgtgacag
109981	tgctctgt ttctttca gaaccgaga gcagctgtg gtgtccaga cagaccagc
110041	cgttgggatg acgggcccct ctgtgagat cccccggcc gccaaagtgg gtgagcttt
110101	cgtgttgcc ggcgggctgg acatgcagg agacctgtt cgggaggagg acctggggg
110161	cccccttc caggggagg ctctggagca gatggcgtc atctcaagg agatccctt
110221	cggggagcaa ggcaggagc aggcagata cggggggac ttcatctgt gctccagccc
110281	tgctccgt cagagcttc cccgggaga cagggccag gacgatgagc tctcggccc
110341	gacctctc cagaaccag acccgactg gtaccgata cgggagagc gggaaagccg
110401	cgatccgct gccaggagg cggtgggcag ggtgacttg gggctgagg ggcggccc
110461	gaccggcag cccccaagc cctacgctg tggggagtg ggcaaggcct tcagccagag
110521	ctcgcacctg ctccggacc tgggtatca caccggggag aagcctatg agtgcggcga
110581	gtcggcaag gccctcagc agagctgca cctgctcgg caccaggca tcacacogg
110641	ggagaagccg tacgagtg gcgagtgcc caaggcttc cggcagagct cggccctg
110701	gcagcagcg aagcgcaca gcggaggcg gccgtactg tcccgagat gcggcaagg
110761	ctcagccc agctccagc tgcgcaagc cagcgcata cacaccggg agaagcccta
110821	cgcgtgccc gagtgccga aggcctcaa ccagagctg ggcctgagc agcaccgca
110881	gattccactg ctgcagagg cgcagcctg cagctgtgc gggaaagcct tctgccacg
110941	ctcgcacctg ctgggcacc agcgcctca cacgggcaag aagcctagc cctgcggga
111001	ctcggcaag gccctcagc agagctcaa cctcatgag caccgcaaga cgcacaggg
111061	cgaagggccc taccgtgccc acaagtccg caaggcttc agccagagct cggcgctat
111121	cagcaccag cgcaccaca cggcgagag gccctacgag tgcggcagt gcggcaagg
111181	ctccgccac agctcggcc tcatccaga ccagcagc cacacggcc gcaagcccta
111241	cgtgtgcaac gagtgcgca aggcctccg ccaccgctg gcgctatg agcactaca
111301	gacgcacaag cgcagcggc ctacgagtg caaccgctg gccaaagcct tccggggag
111361	ctcgcacctg ctccgccacc agaaggtca cggggggac aagctctag gtccgccc
111421	ggcgaggga cgcgggccc ggcgccccg gcccagcggg tggacctgg gggcagccc
111481	gacggcggaa tccgggccc ctctctctg ccgtgaccc ggggggtgg tttgccc
111541	cattcgttt tctaaagt cagacgaata cagctcag ggacgaagt gggtaagc
111601	ccgggagac gtccggcag ctctaagtc agacactga agaagtgaag cggactgca
111661	gccgtacag cccggggaag atgagicca agtcgaggt caccitggc actgcagggt
111721	cgtcggcgg tggggcggag cgggtcagg aggcctctc ctgggcttg ggtggcagg
111781	gaggaccocg cgcctcag cctcggcct ggttggctg agggcggcc tgggttagg
111841	ccctccagc gaggtggagg cgtcggcgg ctacgccag cacaggacc tggcagagg
111901	agtaccctc gccagacc ggcgtcagg ctgggccc tgcgggccc cgttctgtg
111961	gctggcagc ctgcgccc tccaggatg aagggttcc ggtctgaag gctgggtca
112021	gggtccagct ctggcccct ctgcttgg gtctggagg aagcccaag gctccgttc
112081	ccctccagc aggtggggac gttgggaat ccacatccc ctggggggtg tgtgtgtg
112141	tcaaggctc ccattcagac tgggactgg cactacgag ctttgcaac tggcaactga
112201	ggacggagac ccagggtgac accccacct ctgtcgggc cccccggca ggggagacac

112261	aggcccgctct gggtcccaag atggcagggc cctccccct ccagcttgct cctgggtgt
112321	ggtgcctggg gctacagcga cccctccgg tccccgggc cagttcagct gggcatcctc
112381	agggcggggc tetgagggg ccaigtctc agagctcctc ctctccac cagtagcagg
112441	cgggcggcca gctcccaggc agccccctgg catgcctag gtgcacacct gccccgtgtg
112501	accacgaag gctgaagggt ggccatcca gttagtcc ctgccccctgg cccaggaatg
112561	ggctcgggca gggccgcatc tggctgccc agaagcgtct gtcccctggc tetgggagt
112621	ggcgggtgtc tctgtactg tccctgcag ggccccctag cactgctcgg ggaggagggt
112681	ggcgaactg atttgaagt ttacatgtc tgcggccga gctctagag cccgtcagg
112741	icatgctggt tattcagca gatggggctt ggctcggcag ctaggatggt cctgaataaa
112801	aatgggaagg ccagagctgt tctccatca gcaggcttg cagctgggga cgttgaagg
112861	acaggctgc tggctgggg agaccagctc tctcagccc ctgctgccc tgggggtact
112921	aaaccagccc ctgtgtgccc ccatctagt ggcagcccc ctggaggatc gccatcact
112981	igtgagaat gagagaatgc tgcaccccc gcttgggca gggggacagg gccccctaa
113041	atctacctc ttgcccacc cccgggacc cctcagcctt ggccaggact gctctactg
113101	ggcagggcag tcatccactt ccaaccttg cctctcctc cgcgcctgt gctcccagg
113161	aaatgtttt attttttcc aagcatcact tgcacacgt caccacttc ctaaaacca
113221	ccctccgga gtctctgct cgtaaatgc cggttcagc caacctgggt cgcceccaa
113281	gcccagaag cctgctgagc cccgcgcctc ccagctact cagctcgc tcaagctct
113341	aaacgggac ctctccccc ccaccccc cctttctt tetgattat gtaacaggc
113401	aggtaagact cctctctga aggggtgaca gactcacaca aaacctgggt cagaccaggc
113461	aagtgccttt tticagaagt gtgagcggaa cctagcttc agctcatgct cttctctgt
113521	ttcttatgt gttctaagtc cttgacttg gctcccaga cagcgactt gtaagaggcc
113581	gtcctggtag cattgaatt gtctcagat ttctgtctg gatttgit tattgtcta
113641	gtttccctt ctttagcag acgtttga ctgtgtaa gctccagtc ttggtctgt
113701	ttactaatca aatgtttg tcaatgata tgtattctc tctttctt atcttttg
113761	ttcctaata ttaacactt acattctaa gattaaitat ttggtaait aataatttt
113821	aacattcta gtaacgtgg gtactgggt ctgtttgt ttctgtag ttacagctt
113881	ttctctcta tactgtgac gtctgggtt ttittgctc ttggaatt cccctgacc
113941	ccattatat tatttaait agtattttt aataattaa aattagtgt ttaattaa
114001	ccctaatct aaccccagtg atgactgct cagtcatgc tttactat tatgtctgg
114061	tgtcaggatt ttaagtgc catagacatt ctctagcct gaatatata tcaatttat
114121	acagcattg tctactctc agaacgtgt ttacactctc tcaattcgt ttgtacctc
114181	agtcttatg ttattttgt ccagtcgca ctgtctcaa ctgtctcc ctctagggtg
114241	tgaggagcc tggcagccc tgagcatcc ggggtccggg gtctgtggcc caggcgcaca
114301	gcaaaacct gtgggtgtgt gcacggctgg gctctccgg gaggaacct gtggccacc
114361	ggtagtagg agcgtggtt tactgtgca caccacggtc tggtttgg tcttttccc
114421	tgacgtgtt ctgtttgccc ttggttcta tctgttta tgaagtccgt ttacgcttg
114481	ttagtcagc cgttatctg atagacagg tctacgtat caagtatta cgtatttg
114541	agcagatgc tattaacag agatgaactg agaacctgt cctttgatg cctctttgc
114601	ctctttaat gctctagct tcaactctc ttcccaaac attataatg aaaccttg
114661	cttttttt ttaatttc attgcatga gatttatt agctcggcat ttatttta
114721	aaatttgtt atatatatt gctatatc tgaactat aaacagcaa ttattggat
114781	ttgcttctg attcttctg taattctct tacaangaa gttctctat gagnaactt
114841	gctgttaga gtgaggcatg attatttcc agcttagtat gtttgggtc ggttaacccc
114901	caaaggicat gctcatccc gcccatctc tgtgattat tctccgagtg tggagcggc
114961	tgtctaggcc gacgagagc ccaccatcg gcacacctgc cctctctgt ctggtcagtg
115021	ccggctctg tctgagtc actctgatg tcaaggctg gtcttcagc gacctggct
115081	gtacacgga ggtgtgatg gcactgcca gcccatggg gcttggagg ctaaaggatg
115141	cacacctgc tggcagactg agggcacagg tttttcac actgtcagc ttggaata
115201	ttctttgat ttctacctt aactccaaa ggccgtcaa cataagctag atgtactgt
115261	ggtgctgat tacatttag aaaagtcca gcaatacca cgagatgcag caaagaacta
115321	gacctacag atcaggccc ctgcataagg gagcccac agctgtgga gacggggacc
115381	ctctcccag tctgtctgt cccaggatg tccccacc cgcacctct ctcccctgc
115441	cctctgttg tggggccgg ccaccatcac agctgcagag cctcaagaag ggggtcggc
115501	tggccactc cgtggcagg gggacagag ggcaggagct taccgagggt gcagtgtct
115561	cgatcatgct cagctggccc ctgctgggtc ggggggacag ttactggga ggcaggagc

115621	cccactacag ctgccaggac ttctcagagg tgacaagggg gticagtcac ctcagcccag
115681	gtgaaacca aatgccctct tgcgcggctc ctggggccac gcgaggtic gctgggatca
115741	caggatctg gatgtgtcg ccatggacat gcaccacctt cgggggtaa ggggtggga
115801	aaggcagccc ctttctttg ggggacccc tctcagttg ctgatacca ggaacccaaa
115861	tcagaaggtg gtctgggggt gctgagcagg gtgtctcta caccacagg cacacactca
115921	cacagcctcc aggactccag tggggctgag cgtggagac tcacccactg ttgctacccc
115981	cccacccaag gccatcccag aacagctgcc tgcgtctca cggctggccc ctcccctcg
116041	gtctaacca gtgtgggtg gccggcctgg ggtctccacc tgcctctgc tgttccctgg
116101	gctgtggct gtctgcagat gcggggccct ggcccggaga agcccctca gggcccagag
116161	gacgggagt gagcggggag gtgagccccg gactctgag gggccagagg caaaaactg
116221	ggctgttcc ctggaaggca gtttccatg aaaccttca tataggccc cccagacgat
116281	cagcctcacc tctactggg attctcccc gtgagcaat gtgattggg tctacatgga
116341	ccgggactt ctgttgaat tataacttt cccccctgc cctccaggg atctgaaaa
116401	tggaggcctg ggctagacgg aagcttctc caagattct tattgaagg attcgaagag
116461	aaacagggtg tcagtaact gtgggggatg gagggtgag cgtactgtt aacggttta
116521	ctgtgtcac gggaccagt ttgatgtt tcccctcaa gaagcagacc caaacccga
116581	gatgctgagg ttgacagcac agagcgggt catccacaag gcaaccaggc agggagacca
116641	gagacgctt ggaatctgc tccatggg cagggctgg gtgtcacgg atgaagacca
116701	agcagcaggt gccgtggggc gtggggagcc tgcggaaagc gatggacaag gtgcggacc
116761	gcggtccgg cgtggagccc aagctccgc tctgcgtgc agcggagct gggggcggg
116821	cttcaggga ccgcgaccg cgcaccagg gagggtccg ggtccacca gtctaacag
116881	ctcagctcca gctagacgcc gctgagtcg gctttctaga gagcaacccc ggcgggtatt
116941	ttatggttt gcttctga ttgaggaca cgcagctct agaaccct tgaatgtgc
117001	gggcaggcgg aatggattg actgacacg atctgcagt tcaccatctc agggccgcc
117061	ctcaccacca cctatctgc caaaggggg gctcgggtg tgagatcgg gccacagtg
117121	cactagacgg tgggtacgc ctgctctga gggaccagg gccatctc acaccgccac
117181	tggcccctg gctcaataa aggaaggaaa gcgggaaaag gcctttctg ccgcggtggc
117241	ctcgcgctt cctccatgc calctgtgg cagagcccg catggcccc gctgcacaga
117301	aacctcgtg tccgttggg tcccctcc ttgacccca gagagcccc tccgtccaaa
117361	atgaaaaaca gctctccca agagtcaaa taatcacag caatgtgtt aattctct
117421	cggatccact cacagttca cggaccatc tgcatacct tacaactc tacataaagc
117481	aactactgaga agaaaagac gtggtgata aatacaagg catacaaaa taaggagcaa
117541	agaaaaaga cagtctcgc agttctgtt tttctatc tcatgagtag gatgacagat
117601	aaaacacaga atgcccagtg aataattta gtctaagt gtcccata ctgcctaact
117661	ttcaaatca acctatttt taaaatat atttttgt ggtactcat cagttcatgc
117721	accaaagcct ttgttctg actcctaac tttagcccc tctggggtga ggagacccc
117781	taacctcag agcccatac acagtccct tggactaga cctttctg cccatcacag
117841	ctgaccgaa gggccagccc atggccagc ctgcgcccc ctggcgga caactctgcg
117901	ggcagcccc ggagcccagg tgcgacccc cgtctctgg cgcctctag tgtgaaaga
117961	ctctctctg gtgtcccag tcaatggct gttttatt agagaagat ctgcgctgac
118021	gatgatgat gtctttacc gggaggcac ttggggcgc gtcggctcag gggccagct
118081	attagctgc atgcgccc caagcctgc gtccccctga gccgggtcag ctgtggctg
118141	tctgacacg ggttcccc agtctctgc ccgtgtccc tccaggta gtgtccagc
118201	ttcccctt ggttggac ttgtcagc gtctcagc atggagggc gaccctaaag
118261	gatgtattga gcatctcag cactgtct ccccaggt tctgtctcag cagtgaagt
118321	accgggaaaa gggctgtct tgggtctt tcagaggct gggtagacc aaagtctt
118381	agaagattca ccattcagg gactcaaga caaaactag ggttcagca atctgtggg
118441	gattcggcg tgagggaat ctgactcta catgtaatg ttactact gtagggaa
118501	attttccc cctacaaca gcaggccaa atactgagat gicaggttg catcaagag
118561	cgggttctc cacaaggca ccagagaac cctggaatc tgcctccc cgggcacagg
118621	ctgggtgct acggatgaag accaagcagc aggtggcgt gggagtggg agcctggga
118681	aagcagatga caaggtcga gacctccg cgcgagctg agcggagct tccagggaca
118741	cgccgccag cccagtggga ggtcagcgg tccatcagt cctaacagt cagctcaac
118801	tagacgtgc tgactctgc ttctagaga acactccgg cgggtattt atgttttg
118861	cttctgact ggaggcgt caagcttaa aacacctg attagtggg ggaggcggaa
118921	tggattgac tgaicagac ccgacgttc accatctcag gggcccct caccctcc

118981	tacctacca aaggTggggg catcgggtct gagatctggg gtgacacata aaatcaggTg
119041	aagtcttagg acaggggggc gattccaggT cctagggTgc agaaaaaac tacctggccc
119101	cgggctagac agcgtggagg gcgTggcccg ggcTggTgca cagaagTggc ccccaactgg
119161	tcagaaggTg Tgggagccca gggctgTct actgcagaag gggTgcctgT gtggacagag
119221	Tgggctgta gtgcTgctg aactggTccg TcagggtcTg Tgagcagaca cgggccatca
119281	Tcactggctc ctgTctcga tagaaggag ggaaccagg aaagcaagg cgtttatgg
119341	ccgctttgt gtttcgctt cctcTgac cgtcTgccg cagaacgTgg catcacatcc
119401	gctggccaaa cctcggggTc cggctggat gTcccacTc TgTctcTgga gatcTcaact
119461	ctcagcagT cccctgggga caatgTcag aagatgcac ctgaccTgg agctcggTgg
119521	agaggTgTcc ctggTtctt TccgagTg ctggagTgg agTgcctca TgtTggctg
119581	ggaacgggag gaaggaaaca gTcTcagT gagatctct agacagactg Tccctgctct
119641	Tgcaaaatt cagaagattg TctTaaTaa atattccatt TttgTatgc ccttaggtct
119701	attccagac actTaaata Tattgaaaga ctTaaatT Ttatataaaa atattatta
119761	TgactgTat aaaaggaaaca gTtagaactg gactTggaac aacagactgg Ttccaaatg
119821	gaaaaggagT agTcaaggc TgtatattgT caccctgctt attTaaTaa TatcagagT
119881	acatcagTg aaactgTgg ctggagaaa cacaagTgTg aatcaagatt gccgggagaa
119941	atatcaataa cctcagatgT gcagatgaca ccaccctTat ggcagaaagT gaagaggaaac
120001	Tcaaaagcct ctgTatgaa gTgaaaggag agagcgaaaa agTtgctTaa agctcaaca
120061	Ttagaaaac gaagatcagT gcactgTgTc ccatcactTc atggaatag atggggaaac
120121	agTgagaca gTgcagact TtattTgg gggTccaat gaaTaaaaa gacgctTact
120181	TctTggaagg aaagTatga ccaactTga cagcatata aaaagcagag acactactTt
120241	gccagcaagT gTccTctag Tcaagctat ggtTttcca gTgTcTatg atggatgTg
120301	gagTggact gTgaaagg ctgagcaccg aagaagTat gctTttgaa TgtgTgTg
120361	gagaagact Tgagaggcc ctggactgc aaggatcc aaccagTcca TgTaaagg
120421	gatcacccc TggTggTca Tggaaaggac TgatgTgaa gctgaaactc cagTactTg
120481	gctacTaat cgaagagct gactcattgT aaagaccT gatgctyTgga aagattgaa
120541	gTgggaggag aaggggacaa cagaggatga gatgTtTgga Tgcatcact gactcagTg
120601	acTgagTct gatgaaTgTc TgggagTgg Tgatggccag ggaggocctg gcTgctggc
120661	ggTcatggg gTgcTcaaga gTcggccatg actgagTgac Tgaactgaa Tgatccagaa
120721	attTaaatT aatataTaa ccaatccat gcagacaatT atagcatat attataatg
120781	cataattata agcaagTata TgtatattT ataagTtT ataTgTat TaaTgcaag
120841	TatataTat TaaTgata attTgaaTg gaagTaaTt TgggctTcc TggTggTca
120901	gacagTaaag aactgocTg cagTacagga gaccggTcTc gatccctgTg TgggTaaT
120961	Tccctggaga agggaaTggc aaccaactcc aacatgTtTg cctggagaaT Tccatggaca
121021	gaggagoccc gaaggTgca gTccatgggg Tgcaaaag ctgTataTaa cagagTgact
121081	aaacatgTg TaaTaaTaa Ttaactata TattgTatT atattTaaT acatattcag
121141	atattataTaa TaaTgaaTaa catattatc atgTattTaa atactgTat aaacataaat
121201	TaaTaaTaa atTtTcagcc ctTggctTg gggTgTgTt Tgtggactc TttgTctac
121261	TgttctgTaa gTggagctct cccctccaa accagctTtT gaaTgactgT gTaaagcaat
121321	gaaTaaTaa agcatcagga agatgcaac agagctgTca TtctTcag agggTgTgct
121381	TgagTgTgTgT gcaagTccc gagaatgTgT acagattTaa atagTctatt TaaTatgTg
121441	TagcaaatT acgagTgTcT attTcaagT TaaagactT ctggTctct cagTcagTt
121501	cagTgctTgT gTgTgTccg actTttTgT acccattgga cgcagcagc ccaggcctcc
121561	ctgTccatca ccaactcctg gactTcactc aaactcagT ccatcagTc gTgTatgcca
121621	Tcaaacate TcactcTgT gctTcccctT cctcTcccac ctTcaactT Tccagcact
121681	aggTctTcT ccagTgagTc agTctTtTc atcaggTggc cagagTgTgT gactTcagc
121741	Tcagcactg gTcctTcaa TgaatattT gactgattT cctTaggat TgactgTtTg
121801	gatctcctTg cagTcaagg gactTcaag agTctTcTcc aacagcagT TctatgaaT
121861	gaaTgcaaa Tgaatagaga ataacattT cTgagTata TttTaccatt gcaTaaata
121921	TatcagctTg Tgagaacag actTgTccc aggggagagg gTggTgagg atggagTgg
121981	agTtTngat cancagaagc gactgTtat atagaagatg gataaaagg atacacaaca
122041	atgTctTact gTgTggcacc gggaccTata TtcaTgact Tgtgagaac catatcagc
122101	aagactgagg aaagTatT atatgTat gTactgagT TgctTgctg TacaagaTaa
122161	attaaTcaaa catTgTaaT cTatattTca atagaatcca ccccccaaa TatataagTt
122221	TctTggagT ggagacggca accactTca TtctTgac ccaatattT TgctTgagg
122281	atccatgga TagaggatcTg caaagactcTg gacataacc agcagTaaac actTtcccTt

122341	tcaaatgtgt aggtttacta gcgigaatct acagagatgc ccaagacatt cgtttatgag
122401	gaaactcca cacgcagctt cactgagaat tattaaacct attaaagggg gagagcgcca
122461	ggatattcat ggattgaaag aticgatgtg gtcaagtgc cagtttccc caaactgatt
122521	ggtaaatcc ccaggagctg gctcaaggcg caaaattccc ttacctttt titaagagac
122581	gaagccaagg agccgattct gggtgagaga cgtcaggtc ctccggggg agagcagccc
122641	tcttctccc ggtgcctgg gcagtttga ggccacgacc agaaggactt ggctccctgt
122701	gtcgcgact cagaagtct cctctcctc ccaaggactc agaagctggg cgtcctgccc
122761	gcagcagagg aggcagcctg gaggggcccc gcgggcacag cggtcgggt ttcagccgag
122821	ttgccccccc cggccctcta cctgggcctt gccgccggc tccggggccc gccgtgccc
122881	ccgtggccc aaggcctgc tgcoccccg ctggaagtgc tgaccggag gaaggggccc
122941	agacggaggg actcggagcc tccgagtac accctgggac tccgagcgt ggagccctggc
123001	gtcaccccag gcaggggcag tgggggcccc gggcggggtc agggcctcc cccggttctc
123061	atttgacacc gcgggggtgc gctgggcaca gtgtccagg gccacgtcc gagcaggggc
123121	gcgatgcagg cccgggcgcg gctgtcccg ggcgcgagtc cagctgctt gcagaggtgg
123181	cgccaggtcg cagtaccct cacagagacg cccactctg cggctccagg tggcctgtg
123241	ccccccagaa gtgtgacct gtgcaccggg aaggcacagg gccccccagc catgtctgg
123301	atggaagagc cggaaaccgc ccatgccctt cctcgtgac cggcaggcac ccgccgtgtg
123361	tccacacgt gagccctcg gctcccctg ctgacatac acccaggacc tgagtgtca
123421	ggaagttaga agggcaggt gtgtgacac gatgccatc agcalcact gagacctgg
123481	acaaacctca ggggcccagc ctgctctgt aggccccag gccgggccc tccccgacc
123541	ctgcccctga atccggccac actgcccgc ttctgtctc tgcggctgt cagacagccc
123601	tgagcccagg gctgtgac tgcctgccc ttctgccagg actgctctc cccaggctct
123661	tgctggggt cccctctc attcggggg ggcctctct gttcagtgc tcaagtgtc
123721	ccagctttg caaccocat gactgcagca cggcaggct cctgtctt cactagctc
123781	tgagtttgc tcaactcat gtccattgag tcagtgatc tatcaacca tctatcct
123841	tgctcccac ttctctct gctctaac ttcccagca tcagggtct tccaatgag
123901	ttagctctc gcatcaggag gccaaagt tggagctca gcatcagtc ttccagtga
123961	tatgcaggt tgattccct tagaattgac tgggtgac tcttctct ccagagaact
124021	ctcaagagtc ttcccagca ccacagctgg agagcatcag ttctcagtc atcaggttc
124081	ttatagccc agctctaca tgggtacag actattgaa aaccatagc ttgattaga
124141	tggacctca ttgcaaat gatggccct cattggccct gcttttaac acaccatca
124201	ggtttctgt agcttctt ccaagagca aactctttt aatttctgg ctgcagtaac
124261	catccatgt gatittggg cccaagaaa taaaatctg cactgttcc acttttccc
124321	cttctattg ctatgaagt aggggactgg atgccatgt cttagttaa accagcagt
124381	gtaccccga ccgcttct tctaaagag ctcatcac ctcccactg aatgcaatg
124441	gtgctgtc cgcctgctt acctcctgg actttgctg aggtctgt ctctgaggc
124501	ctgcccgtat ccccaggcc cagagcagt ctgggcttc agtccgata gggactatg
124561	gtgtgactg gatgtgctt gcttctct gggaaagaga gacctggcc tgggaacga
124621	ggggacctg gtgaccgga tctctcct cgggagagga gccaaagcag tggacacagg
124681	tcagtgtc ttgctctgt gtggcaggt tccgctgt gtctgtc tggcattc
124741	gggtttctg tgaaccagc cctcccctc ctgataccc atcccatcag cacagaggag
124801	actgggctg gggactctt ggtctgaga ttctctcc catgtgac ccccctctg
124861	ggggagcag gcaccgtg tgaggagggt ggaagcttt caagacccc agctttctg
124921	tcccagggg ctctggcagg gccitggag ctggaatgag ctggaatct ggccagtgg
124981	ggtttccctg gtgtaaaga accgcctgc ccatgcagca ggcataag agcgggctc
125041	gatactggg tgggaagat cccctacag agggcatgc aaccactcc agtattctt
125101	cctgaagaat ccttgaca gaggagcct gtggctaca gtctctggg tggcaaggag
125161	tggacacga ctgaagcag ttaccatga cgcacgggg gtccagggtc agggccgccc
125221	tgttacctg ctgtgtgac ttggccagt cacaccccc aggcctgaa agagaacagt
125281	cttcccagc tgggcatcc aggtcttac agactgctt gtgacttg tgaacttggc
125341	ctgtggccg cttagggcg ctgtccccc gccctatgt cgtgcaagc atgtgacat
125401	gttcgatac gtgtgtcat ctgtggggg cgcacgtgc gggacacgg gcacgggtc
125461	aggaaccgag cccggacacc tccagtgcc ccgagatc cgtcaggtg gggctgtgc
125521	tccgtgtgt ggtgaccoc cctcccccc gcgaactgg tcatagta ccgctggct
125581	ggctctctg gctcagccat cctgcccc ggtcagctc ccgacagcc cagcttagg
125641	ccccagcgt ggaccaggc ccccaggccc cggcctgta gatggacct ccgtctggg

	125701 ggctcattct gctcccggag gcctggcagg cccctcctct ttggcattgc ataccctcgc 125761 attggggtgg gtaagcacag taccctatgc ctgtggcccc gtgggagcgg cctgctcagg 125821 gaggccggag cctcagctac agggctgtca caccgggctg cagaggaaga agacgggagc 125881 gaggcctaca ggaacctagc caggccctgg ccactgagc cgacaggagc ctggccagag 125941 gcctgcacag gacggggtgg cgggggggt ggggtgggt gctgggcccc gtggcctga 126001 ctgcagacc caggggctcc tcagcttaga acggccaagc ctgagtcctg ggggtgcagg 126061 tcaggggg
--	--

引物

在另一实施方案中,提供了用于产生寻靶载体的3'和5'序列的引物。寡核苷酸引物能够与猪免疫球蛋白基因组序列,如上述 Seq ID Nos. 1、4、29、30、12、25、15、16、19、28 或 31 杂交。在具体实施方案中,引物在严格条件下与如上述的 Seq ID Nos. 1、4、29、30、12、25、15、16、19、28 或 31 杂交。另一实施方案提供了能够与猪重链、 κ 轻链或 λ 轻链核酸序列,如上述 Seq ID Nos. 1、4、29、30、12、25、15、16、19、28 或 31 杂交的寡核苷酸探针。多核苷酸引物或者探针可以具有与本发明的多核苷酸杂交的至少 14 个碱基、20 个碱基、30 个碱基或者 50 个碱基。探针或引物可以长为至少 14 个核苷酸,且在具体实施方案中,长为至少 15、20、25、28 或 30 个核苷酸。

在一个实施方案中,提供了扩增猪 Ig 重链的包括功能连接区(J6 区)的片段的引物。在一个非限制性实施方案中,所扩增的重链片段可由 Seq ID No 4 表示并且用于扩增该片段的引物可以与 J 区的部分,如但不限于 Seq ID No 2 互补,以产生 5'重组臂,并且与 Ig 重链 μ 恒定区的部分,如但不限于 Seq ID No 3 互补,以产生 3'重组臂。在另一实施方案中,可以将猪 Ig 重链的区域(如但不限于 Seq ID No 4)亚克隆并装配到寻靶载体中。

在其他实施方案中,提供了扩增猪 Ig κ 轻链的包括恒定区的片段的引物。在另一实施方案中,提供了扩增猪 Ig κ 轻链的包括 J 区的片段的引物。在一个非限制性实施方案中,用于扩增该片段的引物可以与 J 区的部分,如但不限于 Seq ID No 21 或 10 互补,以产生 5'重组臂,并且与恒定区 3'的基因组序列,如但不限于 Seq ID No 14、24 或 18 互补,以产生 3'重组臂。在另一实施方案中,可以将猪 Ig 重链的区域(如但不限于 Seq ID No 20)亚克隆并装配到寻靶载体中。

II. 免疫球蛋白基因的遗传寻靶

本发明提供了被基因修饰而灭活免疫球蛋白基因,如上述免疫球蛋白基因的细胞。可以被基因修饰的动物细胞可以从多种不同器官和组织得到,所述器官和组织如但不限于,皮肤、间充质、肺、胰、心脏、肠、胃、膀胱、血管、肾、尿道、生殖器官,和胚胎、胎儿或成年动物的完整或者部分的解聚制剂。在本发明的一个实施方案中,细胞可以选自,但不限于,上皮细胞、成纤维细胞、神经细胞、角质形成细胞、造血细胞、黑素细胞、软骨细胞、淋巴细胞(B和T)、巨噬细胞、单核细胞、单核的细胞、心肌细胞、其他肌细胞、卵泡细胞、卵丘细胞、表皮细胞、内皮细胞、胰岛细胞、血细胞、血前体细胞、骨细胞、骨前体细胞、神经元干细胞、原始干细胞、肝细胞、角质形成细胞、脐静脉内皮细胞、主动脉内皮细胞、微血管内皮细胞、成纤维细胞、肝星状细胞、主动脉平滑肌细胞、心肌细胞、神经元、Kupffer细胞、平滑肌细胞、施旺细胞、和上皮细胞、红细胞、血小板、嗜中性粒细胞、淋巴细胞、单核细胞、嗜酸性粒细胞、嗜碱性粒细胞、脂肪细胞、软骨细胞、胰岛细胞、甲状腺细胞、甲状旁腺细胞、腮腺细胞、肿瘤细胞、神经胶质细胞、星形胶质细胞、红细胞、白细胞、巨噬细胞、上皮细胞、体细胞、垂体细胞、肾上腺细胞、毛细胞、膀胱状细胞、肾细胞、视网膜细胞、杆细胞、视锥细胞、心细胞、起搏细胞、脾细胞、抗原呈递细胞、记忆细胞、T细胞、B细胞、浆细胞、肌细胞、卵巢细胞、子宫细胞、前列腺细胞、阴道上皮细胞、精细胞、睾丸细胞、生殖细胞、卵细胞、间质细胞、管周细胞、支持细胞、黄体细胞、颈细胞、子宫内膜细胞、乳房细胞、滤泡细胞、粘液细胞、纤毛细胞、非角化上皮细胞、角化上皮细胞、肺细胞、杯形细胞、柱状上皮细胞、鳞状上皮细胞、骨细胞、成骨细胞和破骨细胞。在一个备选实施方案中,可以使用胚胎干细胞。可以使用胚胎干细胞系或者可以从宿主如牛动物新鲜得到胚胎干细胞。细胞可以在合适的成纤维细胞-饲养层上生长或者在白血病抑制因子(LIF)存在下生长。

在具体实施方案中,细胞可以是成纤维细胞;在一个具体实施方案中,细胞可以是胎儿成纤维细胞。成纤维细胞是合适的体细胞类型,因为它们可以从发育中的胎儿和成年动物大量得到。这些细胞可以在体外容易地增殖,具有快的倍增时间,并且可以克隆繁殖以用于

基因寻靶程序中。

寻靶构建体

同源重组

在一个实施方案中，免疫球蛋白基因可以通过同源重组遗传靶定在细胞中。同源重组允许内源基因的位点特异性修饰，且从而可以向基因组中工程改造新的改变。在同源重组中，引入的 DNA 与基因组中含有基本同源的 DNA 序列的位点相互作用并整合到该位点中。在非同源（“随机”或者“不法”）整合中，引入的 DNA 在基因组中的同源序列中没有发现但是整合到别处的许多可能的位置之一。通常，用高等真核生物细胞进行的研究已经揭示同源重组的频率远小于随机整合的频率。这些频率的比率对于“基因寻靶”具有直接的含义，所述基因寻靶取决于通过同源重组（即，外源“寻靶 DNA”与基因组中对应的“靶 DNA”之间的重组）的整合。

许多文献描述了同源重组在哺乳动物细胞中的用途。这些文献示例为 Kucherlapati 等人，*Proc. Natl. Acad. Sci. USA* 81:3153-3157, 1984; Kucherlapati 等人，*Mol. Cell. Bio.* 5:714-720, 1985; Smithies 等人，*Nature* 317:230-234, 1985; Wake 等人，*Mol. Cell. Bio.* 8:2080-2089, 1985; Ayares 等人，*Genetics* 111:375-388, 1985; Ayares 等人，*Mol. Cell. Bio.* 7:1656-1662, 1986; Song 等人，*Proc. Natl. Acad. Sci. USA* 84:6820-6824, 1987; Thomas 等人，*Cell* 44:419-428, 1986; Thomas 和 Capecchi, *Cell* 51: 503- 512, 1987; Nandi 等人，*Proc. Natl. Acad. Sci. USA* 85:3845-3849, 1988; 和 Mansour 等人，*Nature* 336:348-352, 1988. Evans 和 Kaufman, *Nature* 294:146-154, 1981; Doetschman 等人，*Nature* 330:576-578, 1987; Thoma 和 Capecchi, *Cell* 51:503-512, 1987; Thompson 等人，*Cell* 56:316-321, 1989。

本发明可以使用同源重组以灭活细胞、如上述细胞中的免疫球蛋白基因。DNA 可以包含特定基因座上基因的至少部分，其中在天然基因的至少一个拷贝、任选两个拷贝中引入改变，从而防止功能性免疫球蛋白的表达。所述改变可以是插入、缺失、替换或者其组合。当向被灭活的基因的仅一个拷贝中引入改变时，将具有靶基因的单个未突变的拷贝的细胞扩增并进行第二次寻靶步骤，其中该改变可以与第一次改变相同或不同，通常不同，并且当涉及缺失或者替换时，第二

次改变可以与最初引入的改变的至少部分重叠。在该第二次寻靶步骤中，可以使用具有相同的同源性臂但是含有不同的哺乳动物选择标记的寻靶载体。对所得转化体筛选功能性靶抗原的不存在并且可以进一步筛选细胞的 DNA 以确保不存在野生型靶基因。备选地，通过将突变杂合的宿主育种实现表型的纯合性。

寻靶载体

在另一实施方案中，还提供了核酸寻靶载体构建体。可以设计寻靶载体来完成细胞中的同源重组。可以将这些寻靶载体转化到哺乳动物细胞中并通过同源重组靶定有蹄类动物重链、 κ 轻链或者 λ 轻链基因。在一个实施方案中，寻靶载体可以含有 3'重组臂和 5'重组臂（即，侧翼序列），其与有蹄类动物重链、 κ 轻链或者 λ 轻链基因组序列，例如，如上述的 Seq ID Nos. 1、4、29、30、12、25、15、16、19、28 或 31 表示的序列同源。同源 DNA 序列可以包括与基因组序列同源的序列、尤其是连续序列的至少 15 bp、20 bp、25 bp、50 bp、100 bp、500 bp、1 kbp、2 kbp、4 kbp、5 kbp、10 kbp、15 kbp、20 kbp 或 50 kbp。可以设计 3'和 5'重组臂从而使得它们在基因组序列的至少一个功能性可变区、连接区、多样性区和/或恒定区的 3'和 5'末端的侧翼。功能区的寻靶使得它失活，其导致细胞不能产生功能性免疫球蛋白分子。在另一实施方案中，同源 DNA 序列可以包括一个或多个内含子和/或外显子序列。除了核酸序列外，表达载体可以含有选择标记序列，如增强的绿色荧光蛋白（eGFP）基因序列，起始和/或增强子序列，聚腺苷酸尾序列，和/或提供构建体在原核和/或真核宿主细胞中表达的核酸序列。选择标记可以位于 5'和 3'重组臂序列之间。

通过向细胞中导入 DNA 可以产生细胞的靶定基因座的修饰，其中所述 DNA 与靶基因座具有同源性并且包括标记基因，从而允许选择包含整合的构建体的细胞。靶载体中的同源 DNA 将与染色体 DNA 在靶基因座重组。标记基因的两边侧翼可以是同源 DNA 序列、3'重组臂和 5'重组臂。构建寻靶载体的方法已经在本领域中描述，参见例如，Dai 等人，*Nature Biotechnology* 20: 251-255, 2002; WO 00/51424。

可以制备各种构建体用于在靶基因座发生同源重组。构建体可以

包括与靶基因座同源的序列的至少 50 bp、100 bp、500 bp、1kbp、2 kbp、4 kbp、5 kbp、10 kbp、15 kbp、20 kbp 或 50 kbp。该序列可以包括免疫球蛋白基因的任何连续序列。

各种考虑可以涉及确定靶 DNA 序列的同源性程度，如靶基因座的大小、序列的可用性、在靶基因座上双交换事件的相对效率和靶序列与其他序列的相似性。

寻靶 DNA 可以包括这样的序列，其中基本上等基因的 DNA 将所需序列修饰置于基因组中将修饰的对应的靶序列两侧。基本上等基因的序列与对应的靶序列可具有至少约 95%、97-98%、99.0-99.5%、99.6-99.9% 或 100% 同一性（除了所需序列修饰外）。在具体实施方案中，寻靶 DNA 和靶 DNA 可以共有一段 100% 相同的至少约 75、150 或者 500 个碱基对的 DNA 序列。因此，寻靶 DNA 可以来自与将靶定的细胞系密切相关的细胞；或者寻靶 DNA 可以来自与被靶定的细胞相同的细胞系或者动物的细胞。

猪重链寻靶

在本发明的具体实施方案中，提供了用来靶定猪重链基因座的寻靶载体。在一个具体实施方案中，寻靶载体可以含有 5' 和 3' 重组臂，该重组臂含有与猪免疫球蛋白重链基因座的 J6 区的 3' 和 5' 侧翼序列同源的序列。因为 J6 区是猪免疫球蛋白重链基因座的唯一的功能性连接区，所以这将阻止功能性猪重链免疫球蛋白的表达。在特定实施方案中，寻靶载体可以含有 5' 重组臂和 3' 重组臂，所述 5' 重组臂含有与 J6 区 5' 的基因组序列，任选包括 J1-4 同源的序列，所述 3' 重组臂含有与 J6 区 3' 的基因组序列，包括 μ 恒定区同源的序列（“J6 寻靶构建体”），见例如图 1。此外，该 J6 寻靶载体还可以含有位于 5' 和 3' 重组臂之间的选择标记基因，见例如，Seq ID No 5 和图 1。在其他实施方案中，5' 寻靶臂可以含有 J1 5' 的序列，如 Seq ID No. 1 和/或 Seq ID No 4 中描绘。在另一实施方案中，5' 寻靶臂可以含有 J1、J2 和/或 J3 5' 的序列，如 Seq ID No 4 的约残基 1-300、1-500、1-750、1-1000 和/或 1-1500 所示。在进一步的实施方案中，5' 寻靶臂可以含有恒定区 5' 的序列，如 Seq ID No 4 的约残基 1-300、1-500、1-750、1-1000、1-1500 和/或 1-2000 或 Seq ID No 4 的任一片段和/或 Seq ID No

4的任何连续序列或者其片段所示。在另一实施方案中,3'寻靶臂可以含有恒定区3'的序列和/或包括恒定区的序列,如Seq ID No 4的残基7000-8000和/或8000-9000或其片段。在其他实施方案中,寻靶载体可以含有Seq ID No 4的任何连续序列或者其片段。在其他实施方案中,寻靶载体可以含有5'重组臂和3'重组臂,所述5'重组臂含有与猪重链基因座多样性区5'的基因组序列同源的序列,所述3'重组臂含有与猪重链基因座多样性区3'的基因组序列同源的序列。在进一步的实施方案中,寻靶载体可以含有5'重组臂和3'重组臂,所述5'重组臂含有与猪重链基因座 μ 恒定区5'的基因组序列同源的序列,所述3'重组臂含有与猪重链基因座 μ 恒定区3'的基因组序列同源的序列。

在进一步的实施方案中,寻靶载体可以包括,但不限于下面序列的任一种:重链的多样性区如通过Seq ID No 29的残基1089-1099(D(pseudo))表示的,重链的连接区例如通过Seq ID No 29的残基1887-3352(例如:J(psuedo):Seq ID No 29的1887-1931, J(psuedo):Seq ID No 29的2364-2411, J(psuedo):Seq ID No 29的2756-2804, J(功能性J):Seq ID No 29的3296-3352)表示,重组信号例如通过Seq ID No 29的残基3001-3261(九碱基顺序),Seq ID No 29的3292-3298(七碱基序列)表示,恒定区通过下面的残基表示:Seq ID No 29的3353-9070(J到C μ 内含子)、Seq ID No 29的5522-8700(转换区)、Seq ID No 29的9071-9388(μ 外显子1)、Seq ID No 29的9389-9469(μ 内含子A)、Seq ID No 29的9470-9802(μ 外显子2)、Seq ID No 29的9830-10069(μ 内含子B)、Seq ID No 29的10070-10387(μ 外显子3)、Seq ID No 29的10388-10517(μ 内含子C)、Seq ID No 29的10815-11052(μ 外显子4)、Seq ID No 29的11034-11039(聚腺苷酸信号),或者其任意片段或者组合。此外,Seq ID No 29的至少约17、20、30、40、50、100、150、200或300个核苷酸的任意连续序列或者其片段和/或组合可以用作重链寻靶载体的寻靶序列。可以理解通常当设计寻靶构建体时,一个寻靶臂将在另一个寻靶臂的5'。

在其他实施方案中,设计用来破坏猪重链基因的表达的寻靶载体可以含有重组臂,如3'或者5'重组臂,其靶定重链的恒定区。在一个实施方案中,重组臂可以靶定 μ 恒定区,如上述或者Sun & Butler

Immunogenetics (1997) 46: 452-460 中描述的 C μ 序列。在另一实施方案中,重组臂可以靶定 δ 恒定区,如 Zhao 等人(2003) J Immunol 171: 1312-1318 中公开的序列,或者靶定 α 恒定区,如 Brown & Butler (1994) Molec Immunol 31: 633-642 中公开的序列。

Seq ID No.5	<p>GGCCAGACTTCCTCGGAACAGCTCAAAGAGCTCTGTCAAAGCCAGATCCC ATCACAGTGGGCACCAATAGGCCATGCCAGCCTCCAAGGGCCGAACCTGG GTTCTCCACGGCGCACATGAAGCCTGCAGCCTGGCTTATCCTCTTCCGTG GTGAAGAGGCAGGCCCGGACTGGACGAGGGGCTAGCAGGGTGTGGTAGG CACCTTGGCCCCCCCCACCCCGGCAGGAACCAAGAGACCCCTGGGGCTGAGAG TGAGCCTCCAAACAGGATGCCCAACCTTCAGGCCACCTTTCAATCCAGC TACATCCACCTGCCATTTCTCTCTGGGCACAGGGCCAGCCCTGGATC TTGGCCTTGGCTCGACTTGACCCACGCGCACACACACTTCTTAACGT GCTGTCCGCTCACCCCTCCCAGCGTGGTCCATGGGCAGCACGGCAGTGC GCGTCCGGCGGTAGTGAGTGCAGAGGTCCCTTCCCCTCCCCAGGAGCCC CAGGGGTGTGTGACAGATCTGGGGGCTCCTGTCCCTTACACCTTCATGCC CTCCCCTCATACCCACCTCCAGGCGGGAGGCAGCGAGACCTTTGCCAG GACTCAGCCAACGGGCACACGGGAGGCCAGCCCTCAGCAGCTGGCTCCC AAAGAGGAGGTGGGAGGTAGGTCCACAGCTGCCACAGAGAGAAACCTTGA CGGACCCACAGGGGCCACGCCAGCCGGAACAGCTCCCTCGTGGGTGAG CAATGGCCAGGGCCCCCGCGGCCACCACGGCTGGCCTTGGCCAGCTGAG AACTCAGTCCAGTGCAGGGAGACTCAAGACAGCCTGTGCACACAGCCTC GGATCTGCTCCATTTCAAGCAGAAAAAGGAAACCGTGCAGGCAGCCCTC AGCATTCAAGGATTGTAGCAGCGGCCAACTATTCTGTCGGCAGTGGCCGA TTAGAATGACCGTGGAGAAGGGCGGAAGGGTCTCTCGTGGGCTCTGCGGC CAACAGGCCCTGGCTCCACCTGCCCGCTGCCAGCCGAGGGGCTTGGGCC GAGCCAGGAACACAGTGTCAACGGGACCCACAGTGAATGACCAAACTCC CGGCCAGAGCAGCCCCAGGCCAGCCGGGCTCTCGCCCTGGAGCACTACC ATCAGATGCACAAGGGGGCGAGTGTGGAAGAGACGTGTGCCCCGGGCCAT TTGGAAAGGCGAAGGGACCTTCCAGGTGGACAGGAGGTGGGACGCACTCC AGGCAAGGGACTGGGTCCCCAAGGCCTGGGGAAGGGTACTGGCTTGGGG GTTAGCCTGGCCAGGGAACGGGGAGCGGGGCGGGGGCTGAGCAGGGAGG ACCTGACCTCGTGGGAGCGAGGCAAGTCAGGCTTCAGGCAGCAGCCGCGAC ATCCAGACCAGGAGGCTGAGGCAGGAGGGGCTTGCAGCGGGGCGGGGGC CTGCCTGGCTCCGGGGGCTCCTGGGGGACGCTGGCTCTTGTTCCTGTGTC CCGACGACAGGGCCAGCTCGCTGGGCTATGCTTACCTTGTGTCTGGG GCCGGGGCTCAGGGTCTGCTCTCCTCAGGGGAGAGTCCCCTGAGGCTA CGCTGGG*GGGACTATGGCAGCTCCACCAGGGGCTGGGACCAGGGG CCTGGAACCAGGCTGCAGCCCGGAGGACGGGCAGGGCTCTGGCTCTCCAGC ATCTGGCCCTCGGAAATGGCAGAACCCCTGGCGGTGAGCGAGCTGAGAG CGGGTCAGACAGACAGGGGCCCGCCGGAAGGAGAAATTGGGGCAGAGC CCGCCAGGGCCAGGCCAAGGTTCTGTGTGCCAGGGCCTGGTGGGCAC ATTGGTGTGCCATGGCTACTTAGACGCGTGATCAAGGGCGAATCCAGC AACTGGCGCGCTTACTAGTggatcccggcgcgccctaccggtagggg aggcgttttccaagcagctggagcatgctgcttagcagccccgctg ggcacttggcgctacacaagtggccttggcctgcacacattccacac caccgtagggccaaccggctcgtctttgtggcccttcgcccac cttctactctcccctagtcagggaagttccccccgcccgcagctcgg tctgtaggagcgtgacaaatggaatgacagctcactagctcgtgca gatggacagcaccgctgagcaatggaagcggtaggcttggggcagcg gccaatagcagcttggctccttgccttctggctcagagctgggaag gggtgggtccggggcgggctcagggcgggctcagggcgggcgggcg cccgaaggtctccgggaagcccggcattctgcagcttcaaaagcgcag tctggcgctgttctctctctctcaictccggccttgcacctgag ccaataggatogccattgaacaagatggattgcagcaggttctccg ggccttgggtggagaggctattggctatgactgggcacaacagacaat cggctgctctgatccgcccgtttccggctgtagcgcagggcgcccgg ttcttttgcagaccgacctgcccggctccctgaatgaactgagcagc gaggcagcgcggctatcgtgctggccacgacggcgttcttggcagc tgtctcagcgttgcactgaagcgggaaggactggctgctattggcg aagtccggggcagatctcctctcaictcactgctcctgcccagaaa gtatccatcatggctgatcaatgcccggctgcatacgttgatccggc tacctgccattcgaccaccaagcgaacatcgcagcagcagcagta ctggatggaagcggcttctgcaatcaggatgactggacgaagagcat caggggctcgcgccagccgaactgttcgccaggtcaaggcgcgatgcc</p>
-------------	---

cgacggcgaggatctcgtcgtgacccatggcgatgctgcttgcgaata
tcattgggaaaatggccgtttctggattcagactgtggcggctg
gggtggcgatcgtatcaggacatagcgttggctacccgtgatattg
tgaagagcttggcggaatgggtgaccgcttctcgtgctttacgga
tcgcccctccgattcgcagcgcacgccttctatcgccttcttgacgag
ttctctgagggatcaattcTCTAGATGCATGCTCGAGCGGCCGCCAGT
GTGATGGATATCTGCAGAAATTCGCCCTCCAGGCGTTGAAGTCGTCGTGT
CCTCAGGTAAGAACGGCCCTCCAGGGCCTTTAATTTCTGCTCTCGTCTGT
GGGCTTTTCTGACTCTGATCCTCGGGAGGCGTCTGTGCCCCCCCCGGGA
TGAGGCCGGCTTGCCAGGAGGGGTGAGGACCAGGAGCCTGTGGGAAGTT
CTGACGGGGGCTGCAGGCGGGAAGGGCCCCACCGGGGGGCGAGCCCCAGG
CCGCTGGGCGGACAGGAGACCCGTGAGAGTGCGCCCTTGAGGAGGGTGTCTG
CGGAACACGAACGCCCGCCGGAAGGGCTTGCTGCAATGCGGTCTTCAG
ACGGGAGGCGTCTTCTGCCCTCACCGTCTTTCAAGCCCTGTGGGTCTGA
AAGAGCCATGTCGGAGAGAGAAGGGACAGGCCTGTCCCGACCTGGCCGAG
AGCGGGCAGCCCCGGGGGAGAGCGGGCGATCGGCCTGGGCTCTGTGAGG
CCAGGTCCAAGGGAGGACGTGTGGTCTCTGTGACAGGTGCACTTGCGAAA
CCTTAGAAGACGGGGTATGTTGGAAGCGGCTCCTGATGTTTAAAGAAAAGG
GAGACTGTAAGTGAGCAGAGTCTCAAGTGTGTTAAGGTTTTAAAGGTC
AAAGTGTTTTAAACCTTTGTGACTGCAGTTAGCAAGCGTGCGGGGAGTGA
ATGGGGTGCCAGGGTGCCGAGAGGCAGTACGAGGGCCGTGCCCTCTCT
AATTCAGGGCTTAGTTTTGCAGAATAAAGTCGGCCTGTTTTCTAAAAGCA
TTGGTGGTGCTGAGCTGGTGGAGGAGGCCGCGGGCAGCCCTGCCACCTG
CAGCAGGTGGCAGGAAGCAGGTCCGCCAAGAGGCTATTTTAGGAAGCCAG
AAAACACGGTCGATGAATTTATAGCTTCTGGTTTCCAGGAGGTGGTTGGG
CATGGCTTTGCGCAGCGCCACAGAACCGAAAGTGCCCACTGAGAAAAAAC
AACTCCTGCTTAATTTGCATTTTTCTAAAAGAAGAAAACAGAGGCTGACGG
AACTGGAAGTTCTGTTTTAACTACTCGAATTGAGTTTTCGGTCTTAG
CTTATCAACTGCTCACTTAGATTCATTTCAAAGTAAACGTTTAAAGACC
GAGGCATTCCTATCCTCTTCTAAGGCGTTATTCCTGGAGGCTCATTACCC
GCCAGCACTCCGCTGCCTGCAGGCATTGCTGTCACCGTCAACCGTGACGG
CGCGCACGATTTTCAGTTGGCCCGCTTCCCCTCGTGATTAGGACAGACGC
GGGCACTCTGGCCAGCCGTCTTGGCTCAGTATCTGCAGGCGTCCGTCTC
GGGACGGAGCTCAGGGGAAGAGCGTGACTCCAGTTGAACGTGATAGTCGG
TGCGTTGAGAGGAGACCCAGTCCGGTGTGAGTCAAGAGGGGCCCGGGG
CCGAGGCCCTGGGCAGGACGGCCCGTGCCTGCATCACGGGCCAGCGTC
CTAGAGGCAGGACTCTGGTGGAGAGTGTGAGGGTGCCTGGGGCCCTCCG
GAGCTGGGGCCGTGCGGTGCAGGTTGGGCTCTCGGCGCGGTGTTGGCTGT
TTCTGCGGGATTTGGAGGAATTCTTCCAGTGATGGGAGTCCGCACTGACC
GGGACCAAGGCTGTAAGAGGGAGGCCCGCGTCTGGCCAGAGCAGCTGG
GAGGGTTCGGTAAAAGGCTCGCCCGTTTCTTTAATGAGGACTTTTCTG
GAGGGCATTTAGTCTAGTCGGGACCGTTTTTCCACTCGGGAAGAGGGATGC
GGAGGAGGGCATGTGCCAGGAGCCGAAGGCGCCGCGGGGAGAAGCCAG
GGCTCTCCTGTCCCCACAGAGGGACGCCACTGCCGACAGACAGAGGGC
CTTCCCTCTGATGACGGCAAGGGCCCTCGGCTCTTGGGGGTGCTGGG
GGGGAGTCCCGGAAAGCCGCTCACCCAGAGGCCTGAGGGGTGAGACTGA
CCGATGCCTCTTGGCCGGGCTGGGGCCGACCGAGGGGGACTCCGTGGA
GGCAGGGCGATGGTGGCTGCGGGAGGGAACCGACCCTGGGCCGAGCCCG
CTTGGCGATTCCCGGGCAGGGCCCTCAGCCGAGGCGAGTGGGTCCCGCG
GAACCACTTTTCTGGCCAGCGCCACAGGGCTCTCGGGACTGTCCGGGGC
GACGCTGGGCTGCCCGTGGCAGGCCTGGGCTGACCTGGACTTACCAGAC
AGAACAAGGGCTTTCAGGGCTGAGCTGAGCCAGGTTTACGAGGCCAAGTG
GGGCTGAACAGGCTCAACTGGCCTGAGCTGGGTTGAGCTGGGCTGACCT
GGGCTGAGCTGAGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCT
GGACTGGCTGAGCTGAGCTGGGTTGAGCTGAGCTGAGCTGGCCTGGGTTG
AGCTGGGCTGGGTTGAGCTGAGCTGGGTTGAGCTGGGTTGAGCTGGGTTG

ATCTGAGCTGAGCTGGGCTGAGCTGAGCTAGGCTGGGGTGAGCTGGGCTG
AGCTGGTTTGGTTGGGTTGAGCTGAGCTGAGCTGGGCTGTGCTGGCTGA
GCTAGGCTGAGCTAGGCTAGGTTGAGCTGGGCTGGGCTGAGCTGAGCTAG
GCTGGGCTGATTTGGGCTGAGCTGAGCTGAGCTAGGCTGCGTTGAGCTGG
CTGGGCTGGATTGAGCTGGCTGAGCTGGCTGAGCTGGGCTGAGCTGGCCT
GGGTTGAGCTGAGCTGGACTGGTTTGGCTGGGTCGATCTGGGTTGAGCT
GTCCTGGGTTGAGCTGGGCTGGGTTGAGCTGAGCTGGGTTGAGCTGGGCT
CAGCAGAGCTGGGTTGGGCTGAGCTGGGTTGAGCTGAGCTGGGCTGAGCT
GGCCTGGGTTGAGCTGGGCTGAGCTGAGCTGGGCTGAGCTGGCCTGTGTT
GAGCTGGGCTGGGTTGAGCTGGGCTGAGCTGGATTGAGCTGGGTTGAGCT
GAGCTGGGCTGGGCTGTGCTGACTGAGCTGGGCTGAGCTAGGCTGGGCTG
AGCTGGGCTGAGCTGATCCGAGCTAGGCTGGGCTGGTTTGGGCTGAGCTG
AGCTGAGCTAGGCTGGATTGATCTGGCTGAGCTGGGTTGAGCTGAGCTGG
GCTGAGCTGGTCTGAGCTGGCCTGGGTCGAGCTGAGCTGGACTGGTTTGA
GCTGGGTCGATCTGGGCTGAGCTGGCCTGGGTTGAGCTGGGCTGGGTTGA
GCTGAGCTGGGTTGAGCTGGGCTGAGCTGAGGGCTGGGGTGAGCTGGGCT
GAACTAGCCTAGCTAGGTTGGGCTGAGCTGGGCTGGTTTGGGCTGAGCTG
AGCTGAGCTAGGCTGCATTGAGCAGGCTGAGCTGGGCTGAGCAGGCTGG
GGTGAGCTGGGCTAGGTTGAGCTGAGCTGGGTCGAGCTGAGTTGGGCTGA
GCTGGCCTGGGTTGAGGTTAGGCTGAGCTGAGCTGAGCTAGGCTGGGTTGA
GCTGGCTGGGCTGGTTTGGGCTGGGTCAGGCTGGGCTGGGCTGGGCTGGG
TTGAGCTGGGCTGGTTGAGCTGGGCTGAGCTGAGCCGACCTAGGCTGGG
ATGAGCTGGGCTGATTTGGGCTGAGCTGAGCTGAGCTAGGCTGCATTGAG
CAGGCTGAGCTGGGCTGGAGCCTGGCCTGGGCTGAGCTGGGCTGAGCTG
CGCTGAGCTAGGCTGGGTTGAGCTGGCTGGGCTGGTTTGGGCTGGGTCAG
GCTGGGCTGAGCTGGGCTGGGATGAGCTGGGCTGGTTTGGGCTGAGCTGA
GCTGAGCTAGGCTGCATTGAGCAGGCTGAGCTGGGCTGAGCTGGGCTGGG
GTGAGCTGGGCTGAGCTAAGCTGAGCTGGGCTGGTTTGGGCTGAGCTGGC
TGAGCTGGGCTGGTTGAGCTGGGCTGAGCTGAGCAGGGTGAGCTGGGCT
TGAGTTAGGCTGGGCTCAGCTAGGCTGGGTTGATCTGGCAGGGCTGGTTT
GCGCTGGGTCAGGCTCCCGGGAGATGGCCTGGGATGAGCTGGGCTGGTTT
GGGCTGAGCTGAGCTGAGCTGAGCTAGGCTGCATTGAGCAGGCTGAGCTG
GGCTGAGCTGGCCTGGGTTGAGCTGGGCTGGGTTGAGCTGAGCTGGGCTG
AACTGGGCTAAGCTGGGCTGAGCTGGATCGAGCTGAGCTGGGCTGAGCTGG
CCTGGGTTAGCTGGGCTGAGCTGAGCTGAGCTAGGCTGGGTTGAGCTGG
CTGGGCTGGTTTGGGCTGGGTCAGGCTGGGCTGGGCTGGGCTGGGTTGAG
CTGGGCTGGGCTGAGCTGAGCTAGGCTGGGTTGAGCTGGGCTGGGCTGAG
CTGAGCTAGGCTGCATTGAGCTGGGCTGGGATGGATTGAGCTGGGCTGAGCT
GGCTGAGCTGGGCTGAGCTGGGCTGAGCTGGCCTGGGTTGAGCTGGGCTGG
GTTGAGCTGAGCTGGGCTGAGCTGGGCTCAGCAGAGCTGGGTTGAGCTGA
GCTGGGTTGAGCTGGGTTGAGCTGGGCTGAGCAGAGCTGGGTTGAGCTGA
GCTGGGTTGAGCTGGGCTCGAGCAGAGCTGGGTTGAGCTGAGCTGGGTTG
AGCTGGGCTCAGCAGAGCTGGGTTGAGCTGAGCTGGGTTGAGCTGGGCTG
AGCTAGCTGGGCTCAGCTAGGCTGGGTTGAGCTGAGCTGGGCTGAACTGG
GCTGAGCTGGGCTGAACTGGGCTGAGCTGGGCTGAGCTGGGCTGAGCAGA
GCTGGGCTGAGCAGAGCTGGGTTGCTGAGCTGGGTTGAGCTGGGCTGA
GCTGGGCTGAGCAGAGTTGGGTTGAGCTGAGCTGGGTTGAGCTGGGCTGA
GCTAGGCTGGGTTGAGCTGGGTTGAGTTGGGCTGAGCTGGGCTGGGTTGA
GCGGAGCTGGGCTGAACTGGGCTGAGCTGGGCTGAGCGGAACTGGGTTGA
TCTGAATTGAGCTGGGCTGAGCCGGGCTGAGCCGGGCTGAGCTGGGCTAG
GTTGAGCTGGGTTGAGCTGGGCTCAGCTGGTCTGAGCTAGGTTGGGCTGA
GCTAGGCTGGATTGAGCTGGGCTGAGCTGAGCTGATCTGGCCTCAGCTGG
GCTGAGGTTAGGCTGAACTGGGCTGTGCTGGGCTGAGCTGAGCTGAGCCAG
TTTGAAGCTGGGTTGAGCTGGGCTGAGCTGGGCTGTGTTGATCTTCTGA
ACTGGGCTGAGCTGGGCTGAGCTGGCCTAGCTGGATTGAACGGGGGTAAG
CTGGGCCAGGCTGGACTGGGCTGAGCTGAGCTAGGCTGAGCTGAGTTGAA

	<p>TTGGGTTAAGCTGGGCTGAGATGGGCTGAGCTGGGCTGAGCTGGGTTGAG CCAGGTCGGACTGGGTTACCCTGGGCCACACTGGGCTGAGCTGGGCGGAG CTCGATTAACCTGGTCAGGCTGAGTCGGGTCCAGCAGACATGCGCTGGCC AGGCTGGCTTGACCTGGACACGTTTCGATGAGCTGCCTTGGGATGGTTAC CTCAGCTGAGCCAGGTGGCTCCAGCTGGGCTGAGCTGGTGACCCTGGGTG ACCTCGGTGACCAGGTTGTCCTGAGTCCGGGCCAAGCCGAGGCTGCATCA GACTCGCCAGACCCAAGGCCTGGGCCCGGCTGGCAAGCCAGGGCGGTG AAGGCTGGGCTGGCAGGACTGTCCCGAAGGAGGTGCACGTGGAGCCGCC CGGACCCCGACCGGCAGGACCTGAAAGACGCCTCTCACTCCCTTTCTC TTCTGTCCCCTCTCGGGTCTCAGAGAGCCAGTCTGCCCCGAATCTCTAC CCCCCTGCTCCTGCGTCAGCCCCCGTCCGATGAGAGCCTGGTGCCCT GGGCTGCCTGGCCCGGACTTCTGCCAGCTCCGTCACCTTCTCTGGAA</p>
--	---

猪 κ 链寻靶

在本发明的具体实施方案中，提供了靶定猪 κ 链基因座的寻靶载体。在一个具体实施方案中，寻靶载体可以含有 5'重组臂和 3'重组臂，它们含有猪免疫球蛋白 κ 链基因座的恒定区的 3'和 5'侧翼序列的同源序列。因为本发明发现猪免疫球蛋白 κ 轻链基因座仅有一个恒定区，所以这将阻止功能性猪 κ 轻链免疫球蛋白的表达。在具体实施方案中，寻靶载体可以含有 5'重组臂和 3'重组臂，所述 5'重组臂含有与恒定区 5'的基因组序列，任选包括连接区同源的序列，所述 3'重组臂含有与恒定区 3'的基因组序列，任选包括增强子区的至少部分同源的序列（“ κ 恒定寻靶构建体”），见例如图 2。此外，该 κ 恒定寻靶构建体还可以含有位于 5'和 3'重组臂之间的选择标记基因，见例如 Seq ID No 20 和图 2。在其他实施方案中，寻靶载体可以含有 5'重组臂和 3'重组臂，所述 5'重组臂含有与猪 κ 轻链基因座的连接区 5'的基因组序列同源的序列，所述 3'重组臂含有与猪 κ 轻链基因座的连接区 3'的基因组序列同源的序列。在其他实施方案中，寻靶载体的 5'臂可以包括 Seq ID No 12 和/或 Seq ID No 25 或者其任何连续序列或者片段。在另一实施方案中，寻靶载体的 3'臂可以包括 Seq ID No 15、16 和/或 19 或者其任何连续序列或者片段。

在进一步的实施方案中，寻靶载体可以包括，但不限于下面序列的任一种： κ 轻链的编码区例如通过 Seq ID No 30 的残基 1-549 和 Seq ID No 30 的 10026-10549 表示，而内含子序列由例如 Seq ID No 30 的残基 550-10025 表示， κ 轻链的连接区例如由 Seq ID No 30 的残基 5822-7207（例如，J1: Seq ID No 30 的 5822-5859，J2: Seq ID No 30 的 6180-6218，J3: Seq ID No 30 的 6486-6523，J4: Seq ID No 30 的 6826-6863，J5: Seq ID No 30 的 7170-7207）表示，恒定区由下面的残基

表示: Seq ID No 30 的 10026- 10549 (C 外显子) 和 Seq ID No 30 的 10026-10354 (C 编码), Seq ID No 30 的 10524-10529 (聚腺苷酸信号) 和 Seq ID No 30 的 11160-11264 (SINE 元件)或者其任意片段或者组合。此外, Seq ID No 30 的至少约 17、20、30、40、50、100、150、200 或 300 个核苷酸的任何连续序列或者其片段和/或组合可以用作重链寻靶载体的寻靶序列。可以理解通常当设计寻靶构建体时, 一个寻靶臂将在另一个寻靶臂的 5'。

Seq ID No.20	ctcaaacgtaagtggcttttccgactgattcttctgctttctaatgt tggttggctttttgiccatttticagtgtttccatcgaattagttgtcag ggaccaacaacaattgccttccagattaggtaccaggaggaggacattgc tgcattgggagaccagagggtgctaaatttaacgtttccaagccaaat aactggggagggggcttgcctctgtgagggtagggttttatagaagt ggaagttaagggaatcgctatggttcacttttggctcggggaccaaaag tggagcccaaatgagtacatttccatcaattatttggagattttg tctgttgtgtcatttggcaagttttgacattttggitgaatgagcca ttcccaggacccaaaaggatgagaccgaaaagtgaaaaagagccaact ttaagctgagcagacagaccgaattgtgatttggaggagagtagggt tigtagggagaaaagggaacagatcgctgcttttctgaaatagcct ttctcatgggactggcttcagaggggtttttgatgagggaagtgttcta gagcctaactgtgggttgttctgtagcgggaccaagctggaatcaa acgtaagtgcacttttctactcttttctttatcgggtgtgaaat tggggacttttcatgttggagtatgagttgaggtcagttctgaagagag tgggactcatcaaaaactgaggagtaagggtcagaacagagttgtctc atggaagaacaaaagacctagttatgtaggagcctaataatgagtcagt tgcattgggatccaaatggccagacttctctgtaaccaacaatcfaatg agatgtagcagcaaaaagagatttccattgaggggaaagtaaaattgta atatttggatcacctttgtgaaggacatccgtggagattgaacgtaa gtaatttttctactaccttctgaaatttgcataatgccagttgtgac ttttagaggcttaagtgtcagttttgtgaaaatggtaacaagagcat ttcattttatcatcagtttcaaaaagttaaacctcagctcaaaaatgaat ttgtagcaaaaagattaatttaagccaaattgaatgattcaaggaaaa aaaaattagtgtatgaaaaaggaaattctacagctcacaagagcaaaa gcgaattaatttttgaactttgccaatctgtaaatgattttgtt ctttacaatttaaaaagggttagaaaatgtatttcttagctgttttctc tcttctgtctgataaattattatagagataaaaatgaaatataatgga tgtgctaaaaatcagtaagaagttgaaaaatatagtttatgtaaaag
--------------	--

GCAGCCAATATGGGATCGGCCATTGAACAAGATGGATTGCACGCAGGTTCTCCGGCCGCTTGGGTGGAGAGGCTATTCGGCTATGACTGGGCACAACAGACAATCGGCTGCTCTGATGCCGCCGTGTTCCGGCTGTCAGCGCAGGGGCGCCGGTCTTTTTGTCAAGACCGACCTGTCCGGTGCCTGAATGAACTGCAAGACGAGGCAGCGCGCTATCGTGGCTGGCCACGACGGGGCTTCCTTGGCCAGCTGTCTCGACGTTGTCACTGAAGCGGGAAGGGACTGGCTGTATTGGCGAAGTGCCGGGGCAGGATCTCCTGTCTCACCCTTGTCTCTGCCGAGAAATCCATCATGGCTGATGCAATGCGGGCGGTGCATACGCTTGATCGGCTACCTGCCCATTGACCACCAAGCGAAACATCGCATCGAGCGAGCACGTACTCGGATGGAAGCCGGTCTTGTCAATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGAACTGTTCCGCCAGGCTCAAGGCGCGCATGCCCGACGGCGAGGATCTCGTCTGACCCATGGCGATGCCTGCTTGGCCGAAATCATGGTGGAAAATGGCCGCTTTTCTGGATTTCAGACTGTGGCCGCTGGGTGTGGCCGATCGCTATCAGGACATAGCGTTGGCTACCCGTGATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCTCGTGTCTTACGGTATCGCCGCTCCCGATTTCGAGCGCATCGCCTTCTATCGCCTTCTTGAAGTCTTCTGAGGGGATCAATTCTCTAGAGCTCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCTTCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCCTGTCCTTTCCTAATAAAATGAGGAAATGTCATCGCATTGTCTGAGTAGGTGTCACTTCTATTCTGGGGGGTGGGTGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGATGCCGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACCAGCTGGGGCGGCCctcgagcggcccgagtgatgatatactgcagaattcggccttggatcaaacacgcatccatggacaataigtgggtcttagcctgctgagacacaacaggaactccctggcaccacttagaggccagagaacagcacagataaaatccctgccctcatgaagcttatagctagctgggagatataaggcaagataaacacatacaantacatcttaggtaataatataactaaggagaaaattacagggagaaagaggacaggaattgctaggtaggattaaagttcagatagttcatcaggaacactgttctgagaagatacatttagtaagaccgaagtagtaagaaatggaccgtgtgctaaagtggttaagaccattctaggcagcaggaacagcagtaaaagcactgaggtaggttcaactgcacagagttgctgagggaggtaggtctgaggtcttatggtcacaggaagaattgtttactccaccgagatgaaggtgggtgattttgagcagaagaataatctgcctggttatataaacaggattccctgggtctctgatgagaataatctgtcagggtgggatagggagagatagcaataggaccttggctaggagccacgacataatccaagtgaaggtggtgctgcatgaaagcaggactaa caagacctgctgacagtgtgatgtagaaaagataaggagacgaaggtgcatctagggtttctgcctgaggaatagaagataaagctaaagcttagaagatgcagcgtctgggggaaagaccagcagctcagttgatccatctggaattaatitggcataaagtagaggtatgtgggttaacattatgttttttttttccatgtagctatccaactgtcccagcatcatttatitaaaagactttcccttcccctattgattgtttggcaccttcaactgagatcaactgagcataaaatgggtctatttctaagctcttgattccattccatgacctattgttcatctttacccagtagacactgacctgatgattaaagcccctgtaccatgctgttttgacatggtaaatctgagatgcctattagccaaccaagcaagcacggcccttagagagctagatgagagcctggaattcagacgagaaggcagtcctagacatacatgtagtccatcaccatgcggatgggttaaagccatcagactgcaacagactgtgagaggtaaccaagctagagacatggatagagaaccaagcactgagctgggaggtctctacatthaagagattgtagatgaaggactgagaagattgacagagaaagaaatcagaaatgggtgctgctgaaatccaagggaagagatgttccaaagaggagaactgacagttgctagctagcgtcaatfggatgaaaatggaccattggacagaggatgtagtgggtcatgggtgatagataagagcagcttctatagaatggcaggggcaaaatctcatctgaticgcatgggttctaaagaaacgggaaagaaaaatgagtgcatgacca

```

gtccctcaagtagagagggtgaaaaggaggagaaatgaggccacg
acaacatgagagaaatgacagcatttttaaaattttttattttta
ttatttttttcttttagggctgcccctgcaacataggggttcc
caggftagggtctaatcagagctatagctgccagcctacaccacagca
tagcaatgccagatctacatgacctacaccacagctcacagcaacgcccgg
atccttaaccactgagtgaggccagagatcaaaccataccttatgga
tactagtcagggttaccactgagccaaaatgggaaatcctgagtaat
gacagcatttttaattgtccaggaagcaaaactgccaccccgaatgt
ctctcaggcatggtatttttagctgaaaacgattaaaggcccaaaaa
acacaagaagaatgtggacctcccccaacagcctaaaaatttagatt
gagggccigtcccagaatagagctatgcccagactgtctacagaggct
aagggttaggtgtgtgggaaacctcagagatcagaggacggttatg
taccaagcattgacattccatctcattgcaatggccttctccccctct
gtagcccaaacaccaccccccaaaatcttctctgtcttagctgaaga
tgggttgaaaggtatagttcagccacttggcgagttcctcagttgtt
ctgggtcttctccTgatccacattatcagctgtgtttgattttctc
tgttatctgtctcattggcaccattcattcttagaccagcccaaga
acctagaagagtgaaggaaaatttccacctgacaaatgctaaatga
gaatcaccgagtagaggaatgatctgtgtcgggagatagaagag
aaaatcgtcggagagatgtcactgagtaggtgagatgggaaagggtgac
acagggtgaggtgtgcccctcagctaggaagacagacagttcacagaaga
gaagcgggtgtccgtggacatcttgcctcagatgaggaaccggaggct
aagaaagactgcaaaagaaggtatgagattgcagagaggtcagccatga
ctaaatcacagtaaccaaccccccaaccaccatgttttctctagtctgg
cacgtggcaggtagtctgttaggtttcaatattgtgttgaacagta
cctattaggcctccatccccctcttaactaacaanaagtgtagacig
gtcagtgaaaaatggctctttctctatgaattttctcaagagatac
ataactttttttatcataggttgagagcaaatgagaacagcctc
caacctatgacaccgtaacaaaatgtttatgacagtgaaaggcaagaaa
caaacatacacagtaaaagacctccataatattgggtggcccaacac
aggccaggttgaaaagcctttttattttgatagaggaatgatagtaa
tgttcaacctggacagagatcattgttcacigaaatcctccaanaatca
tgggtagtgtgaattataaggaaaataagacttaggataaactttgtc
caagatcccagagttatgccaanaatcagtttcagactccaggcagcct
gatcaagagcctaacttaagacacagctcccttaataactactattca
cagttgcactttcaggcgcgcaagactcattgaatcctacaatagaatga
gtttagatcaaatctctcagtaafagatgaggagactaaatagcgggc
atgacctggtcacttaagacagaattgagattcaaggctaggttcttt
ctacctgtttgtttctacaagatgtagcaatgocgtaattacagacctc
tcagggaaggaa

```

猪λ链寻靶

在本发明的具体实施方案中，提供了靶定猪重链基因座的寻靶载体。在一个实施方案中，通过设计寻靶构建体可以靶定λ，该构建体含有位于第一个 JC 簇 5' 的序列的 5' 臂和含有最后一个 JC 簇 3' 的序列的 3' 臂，从而防止 λ 基因座的功能性表达（见图 3-4）。在一个实施方案中，寻靶载体可以含有 Seq ID No 28 的任意连续序列（如约 17、20、30、40、50、75、100、200、300 或 5000 个核苷酸的连续序列）或者其片段。在一个实施方案中，5' 寻靶臂可以含有 Seq ID No 32，其包括猪 λ 轻链基因组序列的第一个 λ J/C 区的 5' 侧翼序列或者其任何连续序列（如约 17、20、30、40、50、75、100、200、300 或 5000 个核苷酸的连续序列）或者其片段（也见例如图 5）。在另一实施方案中，3' 寻靶臂可以含有但不限于一个或多个下面的：Seq ID No.

33, 其包括猪 λ 轻链基因组序列的 J/C 簇区的 3'侧翼序列, 从 λ J/C 下游约 200 个碱基对起; Seq ID No. 34, 其包括猪 λ 轻链基因组序列的 J/C 簇区的 3'侧翼序列, J/C 簇下游约 11.8 Kb, 接近增强子; Seq ID No. 35, 其包括 λ 下游约 12Kb, 包括增强子区; Seq ID No 36, 其包括 λ 下游约 17.6Kb; Seq ID No 37, 其包括 λ 下游约 19.1Kb; Seq ID No 38, 其包括 λ 下游约 21.3Kb; 和 Seq ID No 39, 其包括 λ 下游约 27 Kb, 或者 Seq ID Nos 32-39 的任意连续序列 (如约 17、20、30、40、50、75、100、200、300 或 5000 个核苷酸的连续序列) 或者其片段 (也见例如图 6)。可以理解通常当设计寻靶构建体时, 一个寻靶臂将在另一个寻靶臂的 5'。

在额外实施方案中, λ 基因座的寻靶构建体可以含有位点特异性重组酶位点, 如 lox。在一个实施方案中, 寻靶臂可以将位点特异性重组酶位点插入到靶定的区域中。然后, 位点特异性重组酶可以被激活和/或应用于细胞, 从而两个位点特异性重组酶位点之间的间插核苷酸序列可以被切除 (见, 例如, 图 6)。

选择标记基因

可以设计 DNA 构建体来修饰内源的靶免疫球蛋白基因。用于靶定构建体的同源序列可以具有一个或多个缺失、插入、替代或其组合。改变可以是插入与靶基因的上游序列在阅读框融合的选择标记基因。

合适的选择标记基因包括, 但不限于: 赋予在某些培养基底物上生长的能力的基因, 如 tk 基因 (胸苷激酶) 或者 hppt 基因 (次黄嘌呤磷酸核糖基转移酶), 其赋予在 HAT 培养基 (次黄嘌呤、氨基蝶呤和胸苷) 上生长的能力; 细菌 gpt 基因 (鸟嘌呤/黄嘌呤磷酸核糖基转移酶), 其允许在 MAX 培养基 (霉酚酸、腺嘌呤和黄嘌呤) 上生长。见, 例如, Song, K-Y., 等人 Proc. Nat'l Acad. Sci. U.S.A. 84:6820-6824 (1987); Sambrook, J., 等人, Molecular Cloning-A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, N. Y. (1989), 第 16 章。选择标记的其他实例包括: 赋予对化合物如抗生素的抗性的基因, 赋予在所选的底物上生长能力的基因, 编码产生可检测信号如发光的蛋白质 (如绿色荧光蛋白、增强的绿色

荧光蛋白(eGFP))的基因。多种此类标记是已知的和可得到的,包括例如,抗生素抗性基因,如新霉素抗性基因(neo)(Southern, P., 和 P. Berg, *J. Mol. Appl. Genet.* 1:327-341 (1982)); 和潮霉素抗性基因(hyg)(*Nucleic Acids Research* 11:6895-6911 (1983), 和 Te Riele, H., 等人, *Nature* 348:649-651 (1990)). 其他选择标记基因包括: 乙酰羟酸合酶(AHAS)、碱性磷酸酶(AP)、 β 半乳糖苷酶(LacZ)、 β 葡萄糖醛酸糖苷酶(GUS)、氯霉素乙酰转移酶(CAT)、绿色荧光蛋白(GFP)、红色荧光蛋白(RFP)、黄色荧光蛋白(YFP)、青色荧光蛋白(CFP)、辣根过氧化物氧化酶(HRP)、萤光素酶(Luc)、胭脂碱合酶(NOS)、章鱼碱合酶(OCS)、和其衍生物。可以得到多种选择标记,它们赋予对氨苄青霉素、博来霉素、氯霉素、庆大霉素、潮霉素、卡那霉素、林可霉素、氨基蝶呤、膦丝菌素、嘌呤霉素和四环素的抗性。

掺入抗生素抗性基因和阴性选择因子的方法是本领域技术人员熟悉的(见,例如,WO 99/15650; 美国专利号 6,080,576; 美国专利号 6,136,566; Niwa 等人, *J. Biochem.* 113:343-349 (1993); 和 Yoshida 等人, *Transgenic Research* 4:277-287 (1995)).

还可以使用选择标记的组合。例如,为了靶定免疫球蛋白基因,可以将 neo 基因(有或者没有其自身启动子,如上文讨论)克隆到与该免疫球蛋白基因同源的 DNA 序列中。为了使用标记的组合,可以克隆 HSV-tk 基因从而使得它在寻靶 DNA 的外面(如果需要,可以在相反侧翼放置另一种选择标记)。向将要靶定的细胞导入 DNA 构建体后,可以在合适的抗生素上选择细胞。在该具体实例中,通过同源重组最可能产生抗 G418 和更昔洛韦的那些细胞,在同源重组中,neo 基因已经重组到免疫球蛋白基因中,但是 tk 基因已经失去,因为它位于双交换区的外面。

缺失可以为约 50bp,更通常至少约 100bp,且通常不超过约 20kbp,其中缺失可以通常包括编码区的至少部分,包括一个或多个外显子的部分、一个或多个内含子的部分,并且可以包括或不包括侧翼非编码区、尤其 5'非编码区(转录调节区)的部分。从而,同源区可以延伸到编码区外进入 5'非编码区或者备选地进入 3'非编码区。插入可以通常不超过 10kbp,通常不超过 5kbp,通常至少 50bp,更通常至少 200bp。

同源区可以包括突变，其中突变可以还灭活靶基因，以提供移码，或者改变关键氨基酸，或者突变可以校正功能异常的等位基因等等。突变可以是轻微改变，不超过同源侧翼序列的约5%。当需要基因的突变时，可以向内含子或外显子中插入标记基因。

可以根据本领域已知的方法制备构建体，可以将各个片段聚集在一起，导入合适的载体，克隆，分析然后进一步操作，直到已得到了所希望的构建体。可以对序列进行各种修饰，以允许限制酶切分析、切除、探针的鉴定，等等。如需要，可以导入沉默突变。在各个阶段，可以使用限制酶切分析、测序、用聚合酶链反应扩增、引物修复、体外诱变等等。

构建体可以用细菌载体制备，所述细菌载体包括原核生物复制系统，例如，大肠杆菌可识别的起点，可以在每个阶段克隆并分析构建体。可以使用标记，其与将用于插入的标记相同或不同，并且可以在导入靶细胞前除去。一旦含有该构建体的载体已经完成，它就可以被进一步操作，如通过缺失细菌序列、线性化、在同源序列中引入短缺失。最后操作后，可以将构建体导入细胞。

本发明还包括含有免疫球蛋白基因的序列的重组构建体。构建体包含载体，如质粒或者病毒载体，已经向其中以正向或者反向插入了本发明的序列。构建体还可以包括调节序列，包括例如，与该序列可操作地连接的启动子。许多合适的载体和启动子是本领域技术人员已知的，并且可以商购得到。提供下面的载体作为实例。细菌的：pBs、pQE-9 (Qiagen)、phagescript、PsiX174、pBluescript SK、pBsKS、pNH8a、pNH16a、pNH18a、pNH46a (Stratagene)；pTrc99A、pKK223-3、pKK233-3、pDR540、pRIT5 (Pharmacia)。真核生物的：pWLneo、pSv2cat、pOG44、pXT1、pSG (Stratagene) pSVK3、pBPv、pMSG、pSVL (Pharmacia)，病毒来源的载体（M13载体、噬菌体1载体、腺病毒载体和逆转录病毒载体）、高、低和可调节拷贝数载体、对于组合用于单个宿主中具有相容复制子的载体(pACYC184和pBR322)和真核附加型复制载体(pCDM8)。其他载体包括原核表达载体，如pcDNA II、pSL301、pSE280、pSE380、pSE420、pTrcHisA、B和C、pRSET A、B和C (Invitrogen, Corp.)、pGEMEX-1、和pGEMEX-2 (Promega, Inc.)、pET载体(Novagen, Inc.)、pTrc99A、

pKK223-3、pGEX 载体、pEZZ18、pRIT2T、和 pMC1871 (Pharmacia, Inc.)、pKK233-2 和 pKK388-1 (Clontech, Inc.)、和 pProEx-HT (Invitrogen, Corp.)和其变体和衍生物。其他载体包括真核生物表达载体, 如 pFastBac、pFastBacHT、pFastBacDUAL、pSFV、和 pTet-Splice (Invitrogen)、pEUK-Cl、pPUR、pMAM、pMAMneo、pBI101、pBI121、pDR2、pCMVEBNA、和 pYACneo (Clontech)、pSVK3、pSVL、pMSG、pCH110、和 pKK232-8 (Pharmacia, Inc.)、p3'SS、pXT1、pSG5、pPbac、pMbac、pMC1neo、和 pOG44 (Stratagene, Inc.)、和 pYES2、pAC360、pBlueBacHis A、B、和 C、pVL1392、pBlueBacIII、pCDM8、pcDNA1、pZeoSV、pcDNA3 pREP4、pCEP4、和 pEBVHis (Invitrogen, Corp.) 和其变体和衍生物。可以使用的额外的载体包括: pUC18、pUC19、pBlueScript、pSPORT、粘粒、噬菌粒、YAC's (酵母人工染色体)、BAC's (细菌人工染色体)、P1 (大肠杆菌噬菌体)、pQE70、pQE60、pQE9 (quagan)、pBS 载体、PhageScript 载体、BlueScript 载体、pNH8A、pNH16A、pNH18A、pNH46A (Stratagene)、pcDNA3 (Invitrogen)、pGEX、pTrsfus、pTrc99A、pET-5、pET-9、pKK223-3、pKK233-3、pDR540、pRIT5 (Pharmacia)、pSPORT1、pSPORT2、pCMVSPORT2.0 和 pSV-SPORT1 (Invitrogen)、pTrxFus、pThioHis、pLEX、pTrcHis、pTrcHis2、pRSET、pBlueBacHis2、pcDNA3.1/His、pcDNA3.1(-)/Myc-His、pSecTag、pEBVHis、pPIC9K、pPIC3.5K、pAO815、pPICZ、pPICZ□、pGAPZ、pGAPZ□、pBlueBac4.5、pBlueBacHis2、pMelBac、pSinRep5、pSinHis、pIND、pIND(SP1)、pVgRXXR、pcDNA2.1、pYES2、pZErO1.1、pZErO-2.1、pCR-Blunt、pSE280、pSE380、pSE420、pVL1392、pVL1393、pCDM8、pcDNA1.1、pcDNA1.1/Amp、pcDNA3.1、pcDNA3.1/Zeo、pSe、SV2、pRc/CMV2、pRc/RSV、pREP4、pREP7、pREP8、pREP9、pREP 10、pCEP4、pEBVHis、pCR3.1、pCR2.1、pCR3.1-Uni、和 pCRBac (来自 Invitrogen); □ExCell、□gt11、pTrc99A、pKK223-3、pGEX-1□T、pGEX-2T、pGEX-2TK、pGEX-4T-1、pGEX- 4T-2、pGEX-4T-3、pGEX-3X、pGEX-5X-1、pGEX-5X-2、pGEX-5X-3、pEZZ18、pRIT2T、pMC1871、pSVK3、pSVL、pMSG、pCH110、pKK232-8、pSL1180、pNEO、和 pUC4K

(Pharmacia) ; pSCREEN-1b(+), pT7Blue(R), pT7Blue-2, pCITE-4abc(+), pOCUS-2, pTAg, pET-32LIC, pET-30LIC, pBAC-2cp LIC, pBACgus-2cp LIC, pT7Blue-2 LIC, pT7Blue-2, □SCREEN-1, □BlueSTAR, pET-3abcd, pET-7abc, pET9abcd, pET11abcd, pET12abc, pET-14b, pET-15b, pET-16b, pET-17b, pET-17xb, pET-19b, pET-20b(+), pET-21abcd(+), pET-22b(+), pET-23abcd(+), pET-24abcd(+), pET-25b(+), pET-26b(+), pET-27b(+), pET-28abc(+), pET-29abc(+), pET-30abc(+), pET-31b(+), pET-32abc(+), pET-33b(+), pBAC-1, pBACgus-1, pBAC4x-1, pBACgus4x-1, pBAC-3cp, pBACgus-2cp, pBACsurf-1, plg, Signal plg, pYX, Selecta Vecta-Neo, Selecta Vecta-Hyg, 和 Selecta Vecta-Gpt (来自 Novagen); pLexA, pB42AD, pGBT9, pAS2-1, pGAD424, pACT2, pGAD GL, pGAD GH, pGAD10, pGilda, pEZM3, pEGFP, pEGFP-1, pEGFP-N, pEGFP-C, pEBFP, pGFPuv, pGFP, p6xHis-GFP, pSEAP2-Basic, pSEAP2-Contral, pSEAP2-启动子, pSEAP2-增强子, p□gal-基本, p□gal-对照, p□gal-启动子, p□gal-增强子, pCMV□, pTet-Off, pTet-On, pTK-Hyg, pRetro-Off, pRetro-On, pIRES1neo, pIRES1hyg, pLXSN, pLNCX, pLAPSN, pMAMneo, pMAMneo-CAT, pMAMneo-LUC, pPUR, pSV2neo, pYEX4T-1/2/3, pYEX-S1, pBacPAK-His, pBacPAK8/9, pAcUW31, BacPAK6, pTriplEx, □gt10, □gt11, pWE15, 和 □TriplEx(来自 Clontech); Lambda ZAP II, pBK-CMV, pBK-RSV, pBluescript II KS +/-, pBluescript II SK +/-, pAD-GAL4, pBD-GAL4 Cam, pSurfsript, Lambda FIX II, Lambda DASH, Lambda EMBL3, Lambda EMBL4, SuperCos, pCR-Script Amp, pCR-Script Cam, pCR-Script Direct, pBS +/-, pBC KS +/-, pBC SK +/-, Phagescript, pCAL-n-EK, pCAL-n, pCAL-c, pCAL-kc, pET-3abcd, pET-11abcd, pSPUTK, pESP-1, pCMVLacI, pOPRSVI/MCS, pOPI3 CAT, pXT1, pSG5, pPbac, pMbac, pMC1neo, pMC1neo Poly A, pOG44, pOG45, pFRT□GAL, pNEO□GAL, pRS403, pRS404, pRS405, pRS406, pRS413, pRS414, pRS415, 和 pRS416 来自 (Stratagene) 和其变体或衍生物。还可以使用双杂交和反向双杂交

载体, 例如, pPC86、pDBLeu、pDBTrp、pPC97、p2.5、pGAD1-3、pGAD10、pACT、pACT2、pGADGL、pGADGH、pAS2-1、pGAD424、pGBT8、pGBT9、pGAD-GAL4、pLexA、pBD-GAL4、pHISi、pHISi-1、placZi、pB42AD、pDG202、pJK202、pJG4-5、pNLexA、pYESTrp 和其变体或衍生物。可以使用任何其他质粒和载体, 只要它们在宿主中可复制和存活。

可以用于允许 DNA 构建体进入宿主细胞的技术包括, 例如, 磷酸钙/DNA 共沉淀、DNA 向细胞核中的显微注射、电穿孔、细菌与完整细胞的原生质体融合、转染或者本领域技术人员已知的任意其他技术。DNA 可以是单链或者双链的、线性或环状的、松弛或者超螺旋 DNA。对于用于转染哺乳动物细胞的各种技术, 见例如, Keown 等人, *Methods in Enzymology* Vol. 185, pp. 527-537 (1990)。

在一个具体实施方案中, 通过用含有从等基因 DNA 分离的免疫球蛋白基因序列的敲除载体转染原代胎儿成纤维细胞, 可以产生杂合或纯合的敲除细胞。在另一实施方案中, 载体可以结合启动子捕获策略, 其中使用例如 IRES (内部核糖体进入位点) 来启动 *Neor* 基因的翻译。

位点特异性重组酶

在额外实施方案中, 寻靶构建体可以含有位点特异性重组酶位点, 如 lox。在一个实施方案中, 寻靶臂可以将位点特异性重组酶靶位点插入到靶定的区域中, 从而使得一个位点特异性重组酶靶位点位于第二个位点特异性重组酶靶位点的 5'。然后, 该位点特异性重组酶被激活和/或应用于细胞, 从而使得切除两个位点特异性重组酶位点之间的间插核苷酸序列。

位点特异性重组酶包括识别并结合短核苷酸位点或者序列特异性重组酶靶位点, 即, 重组酶识别位点, 并催化与这些位点有关的核酸重组的酶或者重组酶。这些酶包括重组酶、转座酶和整合酶。序列特异性重组酶靶位点的实例包括, 但不限于, lox 位点、att 位点、dif 位点和 frt 位点。位点特异性重组酶的非限制性实例包括, 但不限于, 噬菌体 P1 Cre 重组酶、酵母 FLP 重组酶、Inti 整合酶、λ 噬菌体、phi 80、P22、P2、186 和 P4 重组酶、Tn3 解离酶、Hin 重组酶和 Cin

重组酶、大肠杆菌 xerC 和 xerD 重组酶、苏云金芽孢杆菌 (*Bacillus thuringiensis*) 重组酶、TpnI 和 β -内酰胺酶转座子, 和免疫球蛋白重组酶。

在一个实施方案中, 重组位点可以是噬菌体 P1 的 Cre 重组酶识别的 lox 位点。Lox 位点指噬菌体 P1 的 cre 基因产物 Cre 重组酶可以催化位点特异性重组事件所处的核苷酸序列。多种 lox 位点是本领域已知的, 包括天然存在的 loxP、loxB、loxL 和 loxR, 以及许多突变体或者变体 lox 位点, 如 loxP511、loxP514、lox.DELTA.86、lox.DELTA.117、loxC2、loxP2、loxP3 和 lox P23。Lox 位点的额外实例包括, 但不限于 loxB、loxL、loxR、loxP、loxP3、loxP23、lox Δ 86、lox Δ 117、loxP511 和 loxC2。

在另一实施方案中, 重组位点是非 Cre 的重组酶识别的重组位点。在一个实施方案中, 重组酶位点可以是啤酒糖酵母 (*Saccharomyces cerevisiae*) 的 2 μ 质粒的 FLP 重组酶识别的 FRT 位点。FRT 位点指酵母 2 μ 质粒的 FLP 基因产物 FLP 重组酶催化位点特异性重组所处的核苷酸序列。非-Cre 重组酶的额外实例包括, 但不限于位点特异性重组酶, 包括: λ 噬菌体的 Int 重组酶识别的 att 位点 (例如, att1、att2、att3、attP、attB、attL 和 attR)、解离酶家族识别的重组位点和苏云金芽孢杆菌的转座酶识别的重组位点。

在本发明的具体实施方案中, 寻靶构建体可以含有: 与本文描述的猪免疫球蛋白基因同源的序列、选择标记基因和/或位点特异性重组酶靶位点。

同源重组的细胞的选择

然后可以在合适选择的培养基上培养细胞以鉴定提供合适的整合的细胞。插入免疫球蛋白基因的选择标记基因的存在确立了靶构建体向宿主基因组中的整合。这些显示出所需表型的细胞可以通过限制酶切分析、电泳、DNA 印迹分析、聚合酶链反应等等进一步分析 DNA, 以确定发生了同源还是非同源重组。这可以通过下述方法确定: 使用插入片段的探针, 然后对插入片段侧翼的 5' 和 3' 区测序以确定延伸到构建体的侧翼区之外的免疫球蛋白基因的存在或者当导入缺失时, 鉴定该缺失的存在。还可以使用与构建体内的序列互补和与构建体外面的序列互补和在靶基因座的引物。这样, 如果已经发生了

同源重组，那么人们仅可以得到在互补链中存在两个引物的 DNA 双链体。通过阐明引物序列或者预期大小的序列的存在，来支持同源重组的发生。

用于筛选同源重组事件的聚合酶链反应是本领域已知的，见例如，Kim 和 Smithies, *Nucleic Acids Res.* 16:8887-8903, 1988; 和 Joyner 等人, *Nature* 338:153-156, 1989. Thomas 和 Capecchi, 同上, 1987; Nicholas 和 Berg (1983) 在 *Teratocarcinoma Stem Cell*, eds. Siver, Martin 和 Strikland (Cold Spring Harbor Lab., Cold Spring Harbor, N.Y. (pp. 469-497); 和 Linney 和 Donerly, *Cell* 35:693-699, 1983 已经显示，驱动新霉素基因的突变的肿瘤增强子和胸苷激酶启动子的特定组合在胚胎干细胞和 EC 细胞中是有效的。

从第一轮寻靶得到的细胞系对于靶定的等位基因可能是杂合的。可以以多种方式实现纯合性，其中两个等位基因都被修饰。一种方法是生长许多其中一个拷贝已经被修饰的细胞，然后将这些细胞使用不同选择标记进行另一轮的寻靶。备选地，根据传统孟德尔遗传学，通过对修饰的等位基因杂合的动物育种得到纯合子。在一些情况下，需要具有两个不同修饰的等位基因。这可以通过连续轮的基因寻靶或者通过育种杂合子（每个杂合子携带一个所需的经修饰等位基因）来实现。

已经经历同源重组的细胞的鉴定

在一个实施方案中，选择方法可以直接检测免疫球蛋白基因的缺失，该缺失是由于通过同源重组对该免疫球蛋白基因的定向敲除，或者基因中的导致无功能或者不表达的免疫球蛋白的突变。通过抗生素抗性选择已经最常用于筛选（见上文）。该方法可以检测寻靶载体上抗性基因的存在，但是不直接指出整合是定向的重组事件还是随机整合。某些技术，如聚腺苷酸和启动子捕获技术，增加了定向事件的概率，但是仍然不给出所需表型的直接证据，已经得到了免疫球蛋白基因表达缺陷的细胞。此外，负选择形式可以用于选择定向整合；在这些情况中，细胞致死因子的基因以这样的方式插入，从而使得仅定向事件允许细胞避免死亡。通过这些方法选择的细胞然后可以测定其基因破坏、载体整合和最终测定免疫球蛋白基因缺失。在这些情况中，

因为选择是基于检测寻靶载体整合并且不是基于改变的表型，所以仅仅可以检测定向敲除，而检测不到点突变、基因重排或者截短或者其他此类修饰。

可以进一步表征被认为缺少功能性免疫球蛋白基因表达的动物细胞。此类表征可以通过下面的技术实现，包括，但不限于：PCR分析、DNA印迹分析、RNA印迹分析、特异凝集素结合测定法，和/或测序分析。

如技术中描述的PCR分析可以用于确定寻靶载体的整合。在一个实施方案中，扩增引物可以在抗生素抗性基因中产生并且延伸到载体序列外的区域中。可以用DNA印迹分析表征基因座中的总体修饰，如寻靶载体向免疫球蛋白基因座中的整合。然而，RNA印迹分析可以用于表征从每个等位基因产生的转录物。

此外，从RNA转录物产生的cDNA的测序分析地可以用于确定免疫球蛋白等位基因中任何突变的精确位置。

在本发明的另一方面，提供了根据本文描述的方法、序列和/或构建体产生的缺少有蹄类动物重链、 κ 轻链和/或 λ 轻链基因座的功能区的至少一个等位基因的有蹄类动物细胞。这些细胞可以由于同源重组得到。具体地，通过灭活有蹄类动物重链、 κ 轻链或 λ 轻链基因的至少一个等位基因，可以产生表达猪抗体的能力降低的细胞。在其他实施方案中，可以根据本文描述的方法、序列和/或构建体产生缺少有蹄类动物重链、 κ 轻链和/或 λ 轻链基因的两个等位基因的哺乳动物细胞。在进一步的实施方案中，提供了猪动物，其中通过根据本文描述的方法、序列和/或构建体产生的遗传寻靶事件灭活有蹄类动物重链、 κ 轻链和/或 λ 轻链基因的至少一个等位基因。在本发明的另一方面，提供了猪动物，其中通过遗传寻靶事件灭活了有蹄类动物重链、 κ 轻链和/或 λ 轻链基因的两个等位基因。基因可以通过同源重组靶定。在其他实施方案中，可以破坏基因，即，可以改变遗传密码的部分，从而影响基因的该区段的转录和/或翻译。例如，可以通过替代、缺失（“敲除”）或者插入（“敲入”）技术发生基因的破坏。可以插入所需蛋白质的额外基因或者调节现有序列转录的调节序列。

在本发明的实施方案中，根据本文描述的方法、序列和/或构建

体使得有蹄类动物重链、 κ 轻链或者 λ 轻链基因的等位基因失活，从而不再能产生功能性有蹄类动物免疫球蛋白。在一个实施方案中，可以将靶定的免疫球蛋白基因转录成 RNA，但是不翻译成蛋白质。在另一实施方案中，可以将靶定的免疫球蛋白基因以失活的截短形式转录。这种截短的 RNA 可以不被翻译或者可以翻译成无功能的蛋白质。在备选实施方案中，可以将靶定的免疫球蛋白基因以这样的方式灭活，从而使得不发生基因的转录。在另一实施方案中，可以转录靶定的免疫球蛋白基因，然后翻译成无功能的蛋白质。

III. 含有人免疫球蛋白基因的人工染色体的插入 人工染色体

本发明的一方面提供了根据本文描述的方法、序列和/或构建体产生的有蹄类动物和有蹄类动物细胞，它们缺少有蹄类动物重链、 κ 轻链和/或 λ 轻链基因座的功能区的至少一个等位基因，所述有蹄类动物和有蹄类动物细胞进一步修饰以表达人抗体（即，免疫球蛋白（Ig））基因座的至少部分。该人基因座可以经历重排并在有蹄类动物中表达人抗体分子的不同群体。这些克隆的转基因有蹄类动物提供了人抗体（如多克隆抗体）的可补充的、理论上无限的供应，所述抗体可以用于治疗、诊断、纯化和其他临床上相关的目的。

在一个具体实施方案中，人工染色体（AC）可以用于实现向有蹄类动物细胞和动物中转移人免疫球蛋白基因。AC 允许含有单个或多个基因的兆碱基大小的 DNA 片段的定向整合。因此，AC 可以将同源 DNA 导入所选细胞中以用于产生通过异源 DNA 编码的基因产物。在一个实施方案中，具有整合的人免疫球蛋白 DNA 的一个或多个 AC 可以用作将人 Ig 基因导入有蹄类动物（如猪）的载体。

AC 首先在 1983 年在酵母中构建，是人造的线性 DNA 分子，其从负责天然染色体的正确复制和分配的基本顺式作用 DNA 序列元件构建 (Murray 等人 (1983), Nature 301:189-193)。染色体需要至少三个元件来发挥作用。具体地，人工染色体的元件包括至少：（1）自主复制序列（ARS）（具有复制起点的性质，其是用于启动 DNA 复制的位点），（2）着丝粒（动粒装配的位点，其负责有丝分裂和减数分裂时复制的染色体的正确分配），和（3）端粒（在线性染色体

末端的特化结构，其功能为稳定末端并且促进 DNA 分子的最末端的完全复制）。

在一个实施方案中，人 Ig 可以维持为与有蹄类动物染色体 DNA 分离的独立的单位（附加体）。例如，附加体载体含有细胞内 DNA 复制和载体维持所需的必需 DNA 序列元件。附加体载体可以通过商业途径获得（见，例如，Maniatis, T. 等人，*Molecular Cloning, A Laboratory Manual* (1982) pp. 368-369）。AC 可以稳定复制并且与内源染色体一起分离。在备选实施方案中，人 IgG DNA 序列可以整合到有蹄类动物细胞的染色体中，从而允许新信息复制和分配到细胞的子代作为天然染色体的部分（见，例如，Wigler 等人 (1977), *Cell* 11:223)。可以将 AC 易位或者插入到有蹄类细胞的内源染色体中。两个或更多个 AC 可以同时或者顺序地导入宿主细胞中。

此外，AC 可以提供基因组外基因座，用于含有单个或多个基因，包括可操作地连接一个启动子的单个基因的多个拷贝或者连接到单独的启动子的每个拷贝或者几个拷贝的兆碱基大小的 DNA 片段的定向整合。AC 可以允许含有单个或多个人免疫球蛋白基因的兆碱基大小的 DNA 片段的定向整合。通过在本领域技术人员已知的条件下培养具有双着丝粒染色体（即，具有两个着丝粒的染色体）的细胞，从而染色体分裂形成微型染色体和以前的双着丝粒染色体，可以产生 AC。

AC 可以从人（人人工染色体：“HAC”）、酵母（酵母人工染色体：“YAC”）、细菌（细菌人工染色体：“BAC”）、噬菌体 P1 衍生的人工染色体：“PAC”）和其他哺乳动物（哺乳动物人工染色体：“MAC”）构建。AC 基于着丝粒的来源命名（例如，YAC、BAC、PAC、MAC、HAC）。例如，YAC 可以从啤酒糖酵母得到着丝粒。另一方面，MAC 包括活性哺乳动物着丝粒，而 HAC 指包括人着丝粒的染色体。此外，还可以构建植物人工染色体（“PLAC”）和昆虫人工染色体。AC 可以包括来自染色体的负责复制和维持的元件。因此，AC 能够稳定维持大的基因组 DNA 片段，如人 Ig DNA。

在一个实施方案中，提供了含有 YAC 的有蹄类动物。YAC 是基因工程改造的环形染色体，其含有来自酵母如啤酒糖酵母染色体的元件，和外来 DNA 的区段，该区段比常规克隆载体（例如，质粒、粘

粒)接受的区段大得多。YAC 允许外源 DNA 的极大区段(Schlessinger, D. (1990), Trends in Genetics 6:248-253)在哺乳动物细胞和动物中增殖(Choi 等人 (1993), Nature Gen 4:117-123)。YAC 转基因方法非常有效并且通过有效操作克隆的 DNA 的能力得到极大的增强。酵母的主要技术优点是通过 DNA 介导的转化和同源重组可以进行的特定基因组修饰的容易性(Ramsay, M. (1994), Mol Biotech 1:181-201)。在一个实施方案中, 具有整合的人 Ig DNA 的一个或多个 YAC 可以用作将人 Ig 基因导入有蹄类动物(如猪)的载体。

YAC 载体含有用于在酵母中复制的特定结构元件, 包括: 着丝粒、端粒、自主复制序列(ARS)、酵母选择标记(例如, *TRP1*、*URA3* 和 *SUP4*), 和用于插入外源 DNA 的大于 50kb 的大区段的克隆位点。标记基因可以允许选择携带 YAC 的细胞并作为特定限制性内切核酸酶合成的位点。例如, *TRP1* 和 *URA3* 基因可以用作双重选择标记来确保仅维持完整的人工染色体。可以在着丝粒的两边携带酵母选择标记, 并且协助体内端粒形成的两个序列是分离的。然而, 酵母细胞的总 DNA 的仅 1% 的部分是复制所必需的, 包括染色体的中心(着丝粒, 其作为姐妹染色单体之间的附着位点和有丝分裂期间纺锤丝附着的位点)、染色体的末端(端粒, 其作为维持真核生物染色体末端必需的序列), 和 DNA 的另一个短序列, 其称作 ARS, 作为 DNA 区段, 其中双螺旋可以解旋并且开始自己复制。

在一个实施方案中, YAC 可以用于克隆最高达约 1、2 或者 3Mb 免疫球蛋白 DNA。在另一实施方案中, 至少 25、30、40、50、60、70、75、80、85、90 或 95 kb。

酵母整合质粒、复制载体(其是 YAC 的片段)也可以用于表达人 Ig。酵母整合质粒可以含有提供复制起点和用于在细菌(例如, 大肠杆菌)中生长的药物抗性基因的细菌质粒序列、用于选择酵母中的转化体的酵母标记基因, 和用于插入 Ig 序列的限制位点。宿主细胞可以通过将该质粒直接整合在染色体中而稳定获得该质粒。酵母复制载体也可以用于表达人 Ig 作为酵母中的游离质粒环。通过加入着丝粒序列可以稳定酵母或者含有 ARS 的载体。YAC 具有着丝粒和端粒区, 并且可以用于克隆 DNA 的非常大的片段, 因为重组体基本上作为酵母染色体维持。

YAC 如在美国专利号 6,692,954、6,495,318、6,391,642、6,287,853、6,221,588、6,166,288、6,096,878、6,015,708、5,981,175、5,939,255、5,843,671、5,783,385、5,776,745、5,578,461 和 4,889,806; 欧洲专利号 1 356 062 和 0 648 265; PCT 公开号 WO 03/025222、WO 02/057437、WO 02/101044、WO 02/057437、WO 98/36082、WO 98/12335、WO 98/01573、WO 96/01276、WO 95/14769、WO 95/05847、WO 94/23049 和 WO 94/00569 中公开的提供。

在另一实施方案中, 提供了含有 BAC 的有蹄类动物。BAC 是在细菌如大肠杆菌中发现的基于 F 的质粒, 其可以向宿主细胞中转移约 300kb 的外来 DNA。一旦 Ig DNA 已经克隆到宿主细胞中, 新插入的区段就可以随着质粒的剩余部分一起复制。结果, 可以在非常短的时间内产生外来 DNA 的数十亿拷贝。在具体实施方案中, 具有整合的人 Ig DNA 的一个或多个 BAC 用作向有蹄类动物 (例如猪) 中导入人 Ig 基因的载体。

BAC 克隆系统基于大肠杆菌 F 因子, 其复制受到严格控制并且从而确保大构建体的稳定维持 (Willets, N., 和 R. Skurray (1987), *Structure and function of the F-factor and mechanism of conjugation. In Escherichia coli and Salmonella Typhimurium: Cellular and Molecular Biology* (F. C. Neidhardt, Ed) Vol.2 pp 1110-1133, Am. Soc. Microbiol., Washington, D.C.). BAC 已经广泛用于从各种真核生物物种克隆 DNA (Cai 等人 (1995), *Genomics* 29:413-425; Kim 等人 (1996), *Genomics* 34:213-218; Misumi 等人 (1997), *Genomics* 40:147-150; Woo 等人 (1994), *Nucleic Acids Res* 22:4922-4931; Zimmer, R. 和 Gibbins, A.M. V. (1997), *Genomics* 42:217- 226)。F 质粒的低发生率可以减小 DNA 片段之间的重组潜力并且可以避免克隆的细菌基因的致命超表达。BAC 可以将单个拷贝载体中的人免疫球蛋白基因稳定维持在宿主细胞中, 甚至在连续生长 100 或更多世代之后。

BAC(或者 pBAC)载体可以容纳约 30 到 300 kb 对范围的插入片段。一种特定类型的 BAC 载体 pBeloBac11 使用 lacZ 基因的互补, 通过颜色区分含有插入片段的重组分子与携带 BAC 载体的菌落。当将 DNA 片段克隆到 pBeloBac11 的 lacZ 基因中时, 插入激活导致转

化后在 X-Gal/IPTG 板上形成白色菌落(Kim 等人 (1996), *Genomics* 34:213-218)以容易鉴定阳性克隆。

例如, BAC 可以如美国专利号 6,713,281、6,703,198、6,649,347、6,638,722、6,586,184、6,573,090、6,548,256、6,534,262、6,492,577、6,492,506、6,485,912、6,472,177、6,455,254、6,383,756、6,277,621、6,183,957、6,156,574、6,127,171、5,874,259、5,707,811 和 5,597,694; 欧洲专利号 0 805 851; PCT 公开号 WO 03/087330、WO 02/00916、WO 01/39797、WO 01/04302、WO 00/79001、WO 99/54487、WO 99/27118 和 WO 96/21725 中公开的提供。

在另一实施方案中, 提供了含有噬菌体 PAC 的有蹄类动物。在具体实施方案中, 具有整合的人 Ig DNA 的一种或多种噬菌体 PAC 用作向有蹄类动物(例如猪)中导入人 Ig 基因的载体。例如, PAC 可以如美国专利号 6,743,906、6,730,500、6,689,606、6,673,909、6,642,207、6,632,934、6,573,090、6,544,768、6,489,458、6,485,912、6,469,144、6,462,176、6,413,776、6,399,312、6,340,595、6,287,854、6,284,882、6,277,621、6,271,008、6,187,533、6,156,574、6,153,740、6,143,949、6,017,755 和 5,973,133; 欧洲专利号 0 814 156; PCT 公开号 WO 03/091426、WO 03/076573、WO 03/020898、WO 02/101022、WO 02/070696、WO 02/061073、WO 02/31202、WO 01/44486、WO 01/07478、WO 01/05962 和 WO 99/63103 中公开的提供。

在另一实施方案中, 提供了含有 MAC 的有蹄类动物。MAC 具有高有丝分裂稳定性、一致和受调节的基因表达、高克隆能力和非免疫原性。哺乳动物染色体可以包含约 50 到 250 Mb 大小的 DNA 的连续线性链。DNA 构建体可以还含有 DNA 构建体在酵母细胞中增殖必需的一种或多种序列。DNA 构建体还含有编码选择标记基因的序列。DNA 构建体可以能够在用该 DNA 构建体转化的细胞中作为染色体进行维持。MAC 提供基因组外特异整合位点用于导入编码目的蛋白质的基因和允许兆碱基大小的 DNA 整合, 从而例如, 编码整个代谢途径的基因、非常大的基因[例如, 囊性纤维化(CF)基因(~ 600 kb)]或者几种基因[例如, 用于制备多价疫苗的一系列抗原]可以稳定导入细胞中。

提供了哺乳动物人工染色体[MAC]。还提供了其他高等真核生物

物种，如昆虫和鱼的人工染色体，它们使用本文提供的 MACS 产生。用于产生和分离此类染色体的方法。还提供了使用 MAC 从其他物种如昆虫和鱼物种构建人工染色体的方法。人工染色体是完全功能性的稳定染色体。提供了两种类型的人工染色体。一种类型在本文中称作 SATAC[随体人工染色体]是稳定的异染色质染色体，且另一类型是基于常染色质的扩增的微型染色体。如本文使用的，前面的双着丝粒染色体是当双着丝粒染色体断裂并且获得新的端粒时产生的染色体，从而产生两个染色体，每个染色体具有一个着丝粒。每个片段都可以是可复制的染色体。

还提供了其他高等真核生物物种，如昆虫和鱼的人工染色体，它们使用本文提供的 MACS 产生。在一个实施方案中，提供了 SATAC[随体人工染色体]。SATAC 是稳定的异染色质染色体。在另一实施方案中，提供了微型染色体，其中微型染色体基于常染色质的扩增。

在一个实施方案中，通过将细胞与双着丝粒染色体在染色体断裂以形成微型染色体和以前的双着丝粒染色体的条件下培养，产生了人工染色体。在一个实施方案中，SATAC 可以从微型染色体片段产生，见例如，美国专利号 5,288,625。在另一实施方案中，SATAC 可以从以前的双着丝粒染色体的片段产生。SATAC 可以由短卫星 DNA 的重复单位组成并且可是完全异染色质的。在一个实施方案中，不存在异源或者外来 DNA 的插入，SATAC 不含有遗传信息。在其他实施方案中，提供了各种大小的 SATAC，它们通过在选择条件下反复培养并亚克隆含有从以前的双着丝粒染色体产生的染色体的细胞来提供。这些染色体可以基于大小为 7.5 到 10Mb 的重复单位，或者兆复制子 (megareplicon)。这些兆复制子可以是卫星 DNA 的串联块，其侧翼为异源的非卫星 DNA。扩增可以产生相同的染色体区段（每个称作扩增子）的串联排列，所述区段含有边界为异源[“外来”]DNA 的两个倒置的兆复制子。反复细胞融合，在选择培养基上生长和/或 BrdU[5-溴脱氧尿苷]处理，或者其他基因组去稳定剂或者试剂，如电离辐射，包括 X 射线处理，以及亚克隆导致得到这样的细胞系，其携带稳定的异染色质或者部分异染色质的染色体，包括 150-200 Mb “腊肠 (sausage)” 染色体、500-1000 Mb 千兆染色体 (gigachromosome)、稳

定的 250-400 Mb 兆染色体(megachromosome)和多种从其得到的较小的稳定染色体。这些染色体基于这些重复单位并且可以包括所表达的人免疫球蛋白 DNA (也见美国专利号 6,743,967)。

在其他实施方案中, MAC 可以如美国专利号 6,743,967、6,682,729、6,569,643、6,558,902、6,548,287、6,410,722、6,348,353、6,297,029、6,265,211、6,207,648、6,150,170、6,150,160、6,133,503、6,077,697、6,025,155、5,997,881、5,985,846、5,981,225、5,877,159、5,851,760 和 5,721,118; PCT 公开号 WO 04/066945、WO 04/044129、WO 04/035729、WO 04/033668、WO 04/027075、WO 04/016791、WO 04/009788、WO 04/007750、WO 03/083054、WO 03/068910、WO 03/068909、WO 03/064613、WO 03/052050、WO 03/027315、WO 03/023029、WO 03/012126、WO 03/006610、WO 03/000921、WO 02/103032、WO 02/097059、WO 02/096923、WO 02/095003、WO 02/092615、WO 02/081710、WO 02/059330、WO 02/059296、WO 00/18941、WO 97/16533 和 WO 96/40965 中公开的提供。

在本发明的另一方面, 提供了含有 HAC 的有蹄类动物和有蹄类动物细胞。在具体实施方案中, 具有整合的人 Ig DNA 的一种或多种 HAC 用作向有蹄类动物(如猪)中导入人 Ig 基因的载体。在具体实施方案中, 具有整合的人 Ig DNA 的一种或多种 HAC 用于通过核转移产生有蹄类动物(如猪), 其应答免疫而表达人 Ig 并且经历亲和力成熟。

各种方法可以用于产生表达人抗体(“人 Ig”)的有蹄类动物。这些方法包括例如, 在其胚胎阶段或者在其出生后向有蹄类动物中插入含有重链和轻链 Ig 基因的 HAC 或者向有蹄类动物插入人 B 细胞或者 B 细胞前体(例如, 免疫缺陷的或者免疫抑制的有蹄类动物)(见, 例如, 2000 年 11 月 17 日提交的 WO 01/35735, 2002 年 3 月 20 日提交的 US 02/08645)。在任一情况下, 在有蹄类动物中可以同时存在产生人抗体的细胞和产生有蹄类动物抗体的 B 细胞。在含有 HAC 的有蹄类动物中, 单个 B 细胞可以产生含有有蹄类动物和人重链和轻链蛋白质的组合的抗体。在其他实施方案中, HAC 的总大小为至少约 1、2、3、4、5、6、7、8、9 或 10 Mb。

例如, HAC 可以如美国专利号 6,642,207、6,590,089、6,566,066、

6,524,799、6,500,642、6,485,910、6,475,752、6,458,561、6,455,026、6,448,041、6,410,722、6,358,523、6,277,621、6,265,211、6,146,827、6,143,566、6,077,697、6,025,155、6,020,142 和 5,972,649; 美国专利申请号 2003/0037347; PCT 公开号 WO 04/050704、WO 04/044156、WO 04/031385、WO 04/016791、WO 03/101396、WO 03/097812、WO 03/093469、WO 03/091426、WO 03/057923、WO 03/057849、WO 03/027638、WO 03/020898、WO 02/092812、和 WO 98/27200 中公开的提供。

可以插入人免疫球蛋白序列以用于本发明的 AC 的额外实例包括例如, BACs (例如, pBeloBAC11 或 pBAC108L; 见, 例如, Shizuya 等人 (1992), Proc Natl Acad Sci USA 89(18):8794-8797; Wang 等人 (1997), Biotechniques 23(6):992-994)、噬菌体 PACs、YACs (见, 例如, Burke (1990), Genet Anal Tech Appl 7(5):94-99), 和 MACs (见, 例如, Vos (1997), Nat. Biotechnol. 15(12):1257-1259; Ascenzioni 等人 (1997), Cancer Lett 118(2):135-142), 如 HACs, 也见, 美国专利号 6,743,967、6,716,608、6,692,954、6,670,154、6,642,207、6,638,722、6,573,090、6,492,506、6,348,353、6,287,853、6,277,621、6,183,957、6,156,953、6,133,503、6,090,584、6,077,697、6,025,155、6,015,708、5,981,175、5,874,259、5,721,118 和 5,270,201; 欧洲专利号 1 437 400、1 234 024、1 356 062、0 959 134、1 056 878、0 986 648、0 648 265 和 0 338 266; PCT 公开号 WO 04/013299、WO 01/07478、WO 00/06715、WO 99/43842、WO 99/27118、WO 98/55637、WO 94/00569 和 WO 89/09219。额外的实例包括例如在 PCT 公开号 WO 02/076508、WO 03/093469、WO 02/097059; WO 02/096923; US 公开号 US 2003/0113917 和 US 2003/003435; 和美国专利号 6,025,155 中提供的那些 AC。

在本发明的其他实施方案中, AC 通过在每世代中的雄性配子发生传递。AC 可以是整合的或非整合的。在一个实施方案中, AC 可以通过基本上所有分裂的细胞中的有丝分裂传递。在另一实施方案中, AC 可以提供人免疫球蛋白核酸序列的不依赖于位置的表达。在具体实施方案中, AC 可以通过每个雄性和雌性配子发生而具有至少 10% 的传递效率。在一个具体实施方案中, AC 可以是环状的。在另

一具体实施方案中，非整合 AC 可以是 2000 年 3 月 27 日以保藏号 LMBP 5473 CB 保藏在比利时微生物协调保藏中心——BCCM 的非整合 AC。在额外实施方案中，提供了产生 AC 的方法，其中鉴定含有通过雄性配子发生传递的外源核酸的有丝分裂稳定的单位；并且有丝分裂稳定的单位中的进入位点允许人免疫球蛋白基因整合到该单位中。

在其他实施方案中，提供了 AC，其包括：功能性着丝粒、选择标记和/或独特克隆位点。在其他实施方案中，AC 可以显示出一种或多种下面的性质：它可以作为独立的染色体稳定分离，免疫球蛋白序列可以以受控的方式插入并且可以从 AC 表达，它可以通过雄性和雌性种系有效传递和/或转基因动物可以在大于约 30、40、50、60、70、80 或 90% 的它的细胞中具有该染色体。

在具体的实施方案中，AC 可以从成纤维细胞（如任何哺乳动物或者人成纤维细胞）分离，其中它是有丝分裂稳定的。将 AC 转移到仓鼠细胞中后，可以插入 lox(如 loxP)位点和选择标记位点。在其他实施方案中，AC 可以维持有丝分裂稳定性，例如，不存在选择时，每次有丝分裂的损失小于约百分之 5、2、1、0.5 或 0.25。也见 US 2003/0064509 和 WO 01/77357。

异种免疫球蛋白基因

在本发明的另一方面，提供了表达异种免疫球蛋白基因座或者其片段的转基因有蹄类动物，其中该免疫球蛋白可以从整合在内源有蹄类动物染色体的免疫球蛋白基因座表达。在一个实施方案中，提供了从转基因动物得到的有蹄类动物细胞。在一个实施方案中，该异种免疫球蛋白基因座被后代遗传。在另一实施方案中，该异种免疫球蛋白基因座可以通过雄性种系由后代遗传。在其他实施方案中，人工染色体(AC)可以含有异种免疫球蛋白。在一个实施方案中，AC 可以是酵母 AC 或者哺乳动物 AC。在进一步的实施方案中，异种基因座可以是人免疫球蛋白基因座或者其片段。在一个实施方案中，人免疫球蛋白基因座可以是人 14 号染色体、人 2 号染色体和人 22 号染色体或其片段。在另一实施方案中，人免疫球蛋白基因座可以包括人免疫球蛋白的任意片段，该片段可以经历重排。在另一实施方案中，人免疫球

蛋白基因座可以包括人免疫球蛋白重链和人免疫球蛋白轻链的任意片段，其可以经历重排。在另一实施方案中，人免疫球蛋白基因座可以包括任意人免疫球蛋白基因座或者其片段，其在暴露于抗原后可以产生抗体。在具体实施方案中，应答对一种或多种抗原的暴露，所述外源人免疫球蛋白可以在 B 细胞中表达以产生异种免疫球蛋白。

在其他实施方案中，缺少功能性内源免疫球蛋白的任何表达的转基因有蹄类动物可以进一步基因修饰以表达异种免疫球蛋白基因座。在备选实施方案中，提供了含有异种免疫球蛋白基因座的猪动物。在一个实施方案中，异种免疫球蛋白基因座可以是重链和/或轻链免疫球蛋白或者其片段。在另一个实施方案中，异种免疫球蛋白基因座可以是 κ 链基因座或者其片段和/或 λ 链基因座或者其片段。在其他实施方案中，人工染色体 (AC) 可以含有异种免疫球蛋白。在一个实施方案中，AC 可以是酵母 AC 或者哺乳动物 AC。在另一实施方案中，异种基因座可以是人免疫球蛋白基因座或者其片段。在一个实施方案中，人免疫球蛋白基因座可以是人 14 号染色体、人 2 号染色体和人 22 号染色体或者其片段。在另一实施方案中，人免疫球蛋白基因座可以包括人免疫球蛋白的任意片段，该片段可以经历重排。在另一实施方案中，人免疫球蛋白基因座可以包括人免疫球蛋白重链和人免疫球蛋白轻链的任意片段，其可以经历重排。在另一实施方案中，人免疫球蛋白基因座可以包括任意人免疫球蛋白基因座或者其片段，其在暴露于抗原后可以产生抗体。在具体实施方案中，应答对一种或多种抗原的暴露，所述外源人免疫球蛋白可以在 B 细胞中表达以产生异种免疫球蛋白。

在其他实施方案中，缺少功能性内源免疫球蛋白的任何表达的转基因有蹄类动物可以进一步基因修饰以表达异种免疫球蛋白基因座。在备选实施方案中，提供了含有异种免疫球蛋白基因座的猪动物。在一个实施方案中，异种免疫球蛋白基因座可以是重链和/或轻链免疫球蛋白或者其片段。在另一个实施方案中，异种免疫球蛋白基因座可以是 κ 链基因座或者其片段和/或 λ 链基因座或者其片段。在其他实施方案中，人工染色体 (AC) 可以含有异种免疫球蛋白。在一个实施方案中，AC 可以是酵母 AC 或者哺乳动物 AC。在另一实施方案中，异种基因座可以是人免疫球蛋白基因座或者其片段。在一

个实施方案中，人免疫球蛋白基因座可以是人 14 号染色体、人 2 号染色体和人 22 号染色体或者其片段。在另一实施方案中，人免疫球蛋白基因座可以包括人免疫球蛋白的任意片段，该片段可以经历重排。在另一实施方案中，人免疫球蛋白基因座可以包括人免疫球蛋白重链和人免疫球蛋白轻链的任意片段，其可以经历重排。在另一实施方案中，人免疫球蛋白基因座可以包括任意人免疫球蛋白基因座或者其片段，其在暴露于抗原后可以产生抗体。在具体实施方案中，应答对一种或多种抗原的暴露，所述外源人免疫球蛋白可以在 B 细胞中表达以产生异种免疫球蛋白。

在另一实施方案中，提供了含有异种免疫球蛋白基因座的猪动物。在一个实施方案中，异种免疫球蛋白基因座可以是重链和/或轻链免疫球蛋白或者其片段。在另一个实施方案中，异种免疫球蛋白基因座可以是 κ 链基因座或者其片段和/或 λ 链基因座或者其片段。在其他实施方案中，人工染色体 (AC) 可以含有异种免疫球蛋白。在一个实施方案中，AC 可以是酵母 AC 或者哺乳动物 AC。在另一实施方案中，异种基因座可以是人免疫球蛋白基因座或者其片段。在一个实施方案中，人免疫球蛋白基因座可以是人 14 号染色体、人 2 号染色体和人 22 号染色体或者其片段。在另一实施方案中，人免疫球蛋白基因座可以包括人免疫球蛋白的任意片段，该片段可以经历重排。在另一实施方案中，人免疫球蛋白基因座可以包括人免疫球蛋白重链和人免疫球蛋白轻链的任意片段，其可以经历重排。在另一实施方案中，人免疫球蛋白基因座可以包括任意人免疫球蛋白基因座或者其片段，其在暴露于抗原后可以产生抗体。在具体实施方案中，应答对一种或多种抗原的暴露，所述外源人免疫球蛋白可以在 B 细胞中表达以产生异种免疫球蛋白。

可以将人免疫球蛋白基因，如 Ig 重链基因 (人 14 号染色体)、Ig κ 链基因 (人 2 号染色体) 和/或 Ig λ 链基因 (22 号染色体) 插入到 AC 中，如上述。在具体实施方案中，人重链、 κ 和/或 λ Ig 基因的任意部分可以插入到 AC 中。在一个实施方案中，核酸可以与来自人的天然存在的核酸的对应区域至少 70、80、90、95 或 99% 相同。在其他实施方案中，有蹄类动物产生超过一类的人抗体。在各种实施方案中，有蹄类动物产生超过一种不同的人 Ig 或者抗体。在一个实施

方案中，可以导入含有人 Ig 重链基因和 Ig 轻链基因两者的 AC，如自主人人工染色体（“AHAC”，在体内通过内源表达的限制性内切核酸酶转化成线性人染色体的环状重组核酸分子）。在一个实施方案中，人重链基因座和轻链基因座在不同的染色体臂（即，在着丝粒的不同侧）上。在一个实施方案中，重链可以包括 μ 重链，且轻链可以是 λ 或 κ 轻链。Ig 基因可以在一个或超过一个的 AC 中同时或顺序导入。

在具体实施方案中，有蹄类动物或者有蹄类动物细胞表达编码人 Ig 基因的全部或者部分的一种或多种核酸，所述人 Ig 基因经历重排并且表达超过一种的人 Ig 分子，如人抗体蛋白质。从而，编码人 Ig 链或者抗体的核酸为其未重排的形式（即，该核酸没有经历 V(D)J 重组）。在具体实施方案中，编码抗体轻链的 V 基因区段的所有核酸区段可以与编码 J 基因区段的所有核酸区段通过一个或多个核苷酸隔开。在具体实施方案中，编码抗体重链的 V 基因区段的所有核酸区段可以与编码 D 基因区段的所有核酸区段通过一个或多个核苷酸隔开，和/或编码抗体重链的 D 基因区段的所有核酸区段可以与编码 J 基因区段的所有核酸区段通过一个或多个核苷酸隔开。向含有未重排的人 Ig 基因的转基因有蹄类动物施用抗原后，接着在人 Ig 基因座中核酸区段重排并产生与抗原反应的人抗体。

在一个实施方案中，AC 可以表达人染色体的部分或者片段，其含有免疫球蛋白基因。在一个实施方案中，AC 可以表达人轻链基因座的至少 300 或者 1300kb，如 Davies 等人 1993 *Biotechnology* 11: 911-914 中所述。

在另一实施方案中，AC 可以表达人 22 号染色体的部分，其含有至少 λ 轻链基因座，包括 V_λ 基因区段、 J_λ 基因区段和单个 C_λ 基因。在另一实施方案中，AC 可以表达至少一个 V_λ 基因区段、至少一个 J_λ 基因区段和 C_λ 基因。在其他实施方案中，AC 可以含有 λ 基因座的部分，如 Popov 等人 *J Exp Med.* 1999 May 17;189(10):1611-20 中描述。

在另一实施方案中，AC 可以表达人 2 号染色体的部分，其含有至少 κ 轻链基因座，包括 V_κ 基因区段、 J_κ 基因区段和单个 C_κ 基因。在另一实施方案中，AC 可以表达至少一个 V_κ 基因区段、至少一个 J

κ 基因区段和 C_{κ} 基因。在其他实施方案中，含有 κ 轻链基因座的部分的 AC 可以是例如 Li 等人 2000 J Immunol 164: 812- 824 和 Li S Proc Natl Acad Sci U S A. 1987 Jun;84(12):4229-33 中描述的那些。在另一实施方案中，提供了含有约 1.3 Mb 人 κ 基因座的 AC，如 Zou 等人 FASEB J. 1996 Aug;10(10):1227-32 中所述。

在其他实施方案中，AC 可以表达人 14 号染色体的部分，其含有至少人重链基因座，包括 V_H 、 D_H 、 J_H 和 C_H 基因区段。在另一实施方案中，AC 可以表达至少一个 V_H 基因区段、至少一个 D_H 基因区段、至少一个 J_H 基因区段和至少一个 C_H 基因区段。在其他实施方案中，AC 可以表达人重链基因座的至少 85kb，如 Choi 等人 1993 Nat Gen 4:117-123 和/或 Zou 等人 1996 PNAS 96: 14100-14105 中所述。

在其他实施方案中，AC 可以表达重链和轻链基因座的部分，如至少 220、170、800 或 1020 kb，如 Green 等人 1994 Nat Gen 7:13-22; Mendez 等人，1995 Genomics 26: 294-307; Mendez 等人 1997 Nat Gen 15: 146-156; Green 等人 1998 J Exp Med 188: 483-495 和/或 Fishwild 等人 1996 Nat Biotech 14: 845-851 中公开的。在另一实施方案中，AC 可以表达兆碱基量的人免疫球蛋白，如 Nicholson J Immunol. 1999 Dec 15;163(12):6898-906 和 Popov Gene. 1996 Oct 24;177(1-2):195-201 中所述。此外，在一个具体实施方案中，来自人 14 号染色体（包含 Ig 重链基因）、人 2 号染色体（包含 Ig κ 链基因）和人 22 号染色体（包含 Ig λ 链基因）的 MAC 可以同时或者连续导入，如 Robl 等人的美国专利公开号 2004/0068760 中所述。在另一实施方案中，MAC 的总大小小于或者等于约 10、9、8 或者 7 兆碱基。

在具体实施方案中，可以将种系构型的人免疫球蛋白的人 V_h 、人 D_h 、人 J_h 区段和人 μ 区段插入 AC，如 YAC，从而使得 V_h 、 D_h 、 J_h 和 μ DNA 区段形成免疫球蛋白谱，其含有对应于人 DNA 区段的部分，例如，如 Babraham Institute 的美国专利号 5,545,807 所述。此类 AC 在插入有蹄类动物细胞和产生有蹄类动物后可以产生重链免疫球蛋白。在一个实施方案中，这些免疫球蛋白可以形成功能性重链-轻链免疫球蛋白。在另一实施方案中，这些免疫球蛋白可以以允许从有蹄类动物的合适的细胞或者体液回收的量表达。此类免疫球蛋白可以插入酵母人工染色体载体，如 Burke, D T, Carle, G F 和 Olson, M

V (1987) "Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors" *Science*, 236, 806-812 所述, 或者通过导入染色体片段 (如 Richer, J 和 Lo, C W (1989) "Introduction of human DNA into mouse eggs by injection of dissected human chromosome fragments" *Science* 245, 175-177 所述)。

关于含有人免疫球蛋白基因的特定 AC 的额外信息可以见例如 Giraldo & Montoliu (2001) *Transgenic Research* 10: 83-103 和 Peterson (2003) *Expert Reviews in Molecular Medicine* 5: 1-25 的最近综述。

AC 转移方法

可以首先将人免疫球蛋白基因插入 AC 中然后将含有人免疫球蛋白的 AC 插入有蹄类动物细胞中。备选地, 可以将 AC 转移到中间哺乳动物细胞, 如 CHO 细胞, 之后插入到有蹄类动物细胞中。在一个实施方案中, 中间哺乳动物细胞也含有 AC 并且第一种 AC 可以插入到哺乳动物细胞的该 AC 中。具体地, 可以将酵母细胞中含有人免疫球蛋白基因或者其片段的 YAC 转移到含有 MAC 的哺乳动物细胞。YAC 可以插入到 MAC 中。然后 MAC 转移到有蹄类动物细胞。可以通过同源重组将人 Ig 基因插入 AC 中。所得的含有人 Ig 基因的 AC 可以接着导入有蹄类动物细胞中。可以通过本文描述的或者本领域已知的技术选择一种或多种有蹄类动物细胞, 其含有包含人 Ig 的 AC。

本文提供了用于导入 AC 的合适宿主, 其包括但不限于任何动物或植物、其细胞或组织, 包括但不限于: 哺乳动物、鸟类、爬行动物、两栖类、昆虫、鱼、蜘蛛、烟草、番茄、小麦、单子叶植物、双子叶植物和藻类。在一个实施方案中, 可以通过本领域已知的任意试剂浓缩 AC (Marschall 等人, *Gene Ther.* 1999 Sep;6(9):1634-7), 所述试剂包括但不限于, 精胺、亚精胺、聚氮丙啶, 和/或聚赖氨酸, 之后导入细胞中。通过细胞融合或者微细胞融合或者分离后通过本领域技术人员已知的任意方法可以导入 AC, 所述方法包括但不限于: 直接 DNA 转移、电穿孔、核转移、微细胞融合、细胞融合、原生质球融

合、脂质介导的转移、脂转染、脂质体、微粒轰击、显微注射、磷酸钙沉淀和/或任意其他合适的方法。用于将 DNA 导入细胞的其他方法包括核显微注射、电穿孔、细菌与完整细胞的原生质体融合。也可以使用聚阳离子，如 1,5-二甲基-1,5-二氮十一亚甲基聚甲溴化物和聚鸟氨酸。关于用于转化哺乳动物细胞的各种技术，见例如，Keown 等人 *Methods in Enzymology* (1990) Vol.185, pp. 527-537; 和 Mansour 等人 (1988) *Nature* 336:348-352。

通过直接 DNA 转化、在细胞或者胚胎中显微注射、植物的原生质体再生、电穿孔、微粒枪和本领域技术人员已知的其他此类方法可以导入 AC (见例如，Weissbach 等人 (1988) *Methods for Plant Molecular Biology*, Academic Press, N.Y., Section VIII, pp. 421-463; Grierson 等人 (1988) *Plant Molecular Biology*, 2d Ed., Blackie, London, Ch. 7-9; 还见例如，美国专利号 5,491,075; 5,482,928; 和 5,424,409; 还见例如，美国专利号 5,470,708,。)

在具体实施方案中，可以使用含有人 Ig 基因的一种或多种分离的 YAC。可以通过本领域已知的任意试剂浓缩分离的 YAC(Marschall 等人，*Gene Ther.* 1999 Sep;6(9):1634-7)，所述试剂包括但不限于，精胺、亚精胺、聚氮丙啶，和/或聚赖氨酸。然后通过本领域中已知的任意方法（例如，显微注射、电穿孔、脂质介导的转染等等）将浓缩的 YAC 转移到猪细胞。备选地，可以通过精子介导的基因转移或者胞质内精子注射技术（ICSI）介导的基因转移将浓缩的 YAC 转移到卵母细胞。在一个实施方案中，可以用原生质球融合将含有人 Ig 基因的 YAC 转移到猪细胞。

在本发明的其他实施方案中，含有人 Ig 的 AC 可以插入到成年、胎儿或者胚胎有蹄类动物细胞中。有蹄类动物细胞的额外实例包括未分化的细胞，如胚胎细胞（如胚胎干细胞）、分化的或者体细胞，如上皮细胞、神经细胞、表皮细胞、角质形成细胞、造血细胞、黑素细胞、软骨细胞、B 淋巴细胞、T 淋巴细胞、红细胞、巨噬细胞、单核细胞、成纤维细胞、肌细胞、来自雌性生殖系统的细胞，如乳腺、卵丘、卵泡或者输卵管细胞、生殖细胞、胎盘细胞，或者来自任何器官，如膀胱、脑、食管、输卵管、心脏、肠、胆囊、肾、肝、肺、卵巢、胰、前列腺、脊髓、脾、胃、睾丸、胸腺、甲状腺、气管、输尿管、

尿道、和子宫的细胞或者本文中描述的任何其他细胞类型。

位点特异性重组酶介导的转移

在本发明的具体实施方案中，含有人免疫球蛋白基因的 AC 向猪细胞，如本文描述的或者本领域已知的那些细胞的转移可以通过位点特异性重组酶介导的转移来完成。在一个具体实施方案中，可以将 AC 转移到猪成纤维细胞。在另一具体实施方案中，AC 可以是 YAC。

在本发明的其他实施方案中，含有位点特异性重组酶靶位点的环化 DNA，如 AC 可以转移到在其基因组中具有位点特异性重组酶靶位点的细胞系中。在一个实施方案中，细胞的位点特异性重组酶靶位点可以位于外源染色体。外源染色体可以是不整合在宿主的内源基因组中的人工染色体。在一个实施方案中，AC 可以通过种系传递转移到后代。在一个具体实施方案中，含有人免疫球蛋白基因或者其片段的 YAC 可以通过位点特异性重组酶环化，且然后转移到含有 MAC 的宿主细胞中，其中 MAC 含有位点特异性重组酶位点。然后现在含有人免疫球蛋白基因座或者其片段的该 MAC 可以与猪细胞，如但不限于，成纤维细胞融合。然后该猪细胞可以用于核转移。

在本发明的一些实施方案中，在插入有蹄类动物细胞中前，含有人免疫球蛋白基因或者其片段的 AC 可以转移到哺乳动物细胞，如 CHO 细胞。在一个实施方案中，中间哺乳动物细胞也含有 AC 并且第一种 AC 可以插入到哺乳动物细胞的该 AC 中。具体地，可以将酵母细胞中含有人免疫球蛋白基因或者其片段的 YAC 转移到含有 MAC 的哺乳动物细胞。YAC 可以插入到 MAC。然后 MAC 可转移到有蹄类动物细胞。在具体实施方案中，含有人 Ig 基因或者其片段的 YAC 可以含有位点特异性重组酶靶基因。YAC 可以首先通过应用合适的位点特异性重组酶环化，然后插入到含有其自身的位点特异性重组酶靶位点的哺乳动物细胞中。然后，可以应用位点特异性重组酶以将 YAC 整合到中间哺乳动物细胞中的 MAC 中。位点特异性重组酶可以顺式或者反式应用。具体地，位点特异性重组酶可以反式应用。在一个实施方案中，位点特异性重组酶可以通过位点特异性重组酶表达质粒，如 Cre 表达质粒的转染来表达。此外，YAC 的一个端粒区也可以用选择标记，如本文描述的或者本领域已知的选择标记改

进 (retrofit)。然后将中间哺乳动物细胞的 MAC 内的人 Ig 基因或者其片段转移到有蹄类动物细胞, 如成纤维细胞中。

备选地, 含有人 Ig 基因或者其片段的 AC, 如 YAC 可以含有任选位于每个端粒附近的位点特异性重组酶靶位点。YAC 可以首先通过应用合适的位点特异性重组酶环化, 然后插入到在其基因组内含有其自身位点特异性重组酶靶位点的有蹄类动物细胞中。备选地, 有蹄类动物细胞可以含有其自身的 MAC, 该 MAC 含有位点特异性重组酶靶位点。在该实施方案中, YAC 可以直接插入有蹄类动物细胞的内源基因组中。在具体实施方案中, 有蹄类动物细胞可以是成纤维细胞或者可以用于核转移的任何其他合适的细胞。见例如, 图 7; Call 等人, *Hum Mol Genet.* 2000 Jul 22;9(12): 1745-51。

在其他实施方案中, 提供了环化至少 100kb DNA 的方法, 其中 DNA 然后通过位点特异性重组酶整合到宿主基因组中。在一个实施方案中, 可以环化至少 100、200、300、400、500、1000、2000、5000、10,000 kb 的 DNA。在另一实施方案中, 可以环化至少 1000、2000、5000、10,000 或 20,000 兆碱基的 DNA。在一个实施方案中, 通过将位点特异性重组酶靶位点附着在 DNA 序列的每个末端, 然后应用位点特异性重组酶以导致 DNA 的环化可以完成 DNA 的环化。在一个实施方案中, 位点特异性重组酶靶位点可以是 lox。在另一实施方案中, 位点特异性重组酶靶位点可以是 Flt。在一些实施方案中, DNA 可以是人工染色体, 如 YAC 或者本文描述的或者本领域已知的任何 AC。在另一实施方案中, AC 可以含有人免疫球蛋白基因座或者其片段。

在另一优选的实施方案中, YAC 可以转化成或者整合在人工哺乳动物染色体内。哺乳动物人工染色体可以转移到或者包含在猪细胞内。可以通过本领域已知的任意方法将人工染色体导入猪基因组内, 所述方法包括但不限于直接注射中期染色体、脂质介导的基因转移或者微细胞融合。

位点特异性重组酶包括识别并结合短核酸位点或者序列特异性重组酶靶位点, 即, 重组酶识别位点, 并催化与这些位点有关的核酸重组的酶或者重组酶。这些酶包括重组酶、转座酶和整合酶。序列特异性重组酶靶位点的实例包括, 但不限于, lox 位点、att 位点、dif

位点和 *frt* 位点。位点特异性重组酶的非限制性实例包括，但不限于，噬菌体 P1 Cre 重组酶、酵母 FLP 重组酶、Inti 整合酶、 λ 噬菌体、*phi* 80、P22、P2、186 和 P4 重组酶、Tn3 解离酶、Hin 重组酶和 Cin 重组酶、大肠杆菌 *xerC* 和 *xerD* 重组酶、苏云金芽孢杆菌重组酶、TpnI 和 β -内酰胺酶转座子，和免疫球蛋白重组酶。

在一个实施方案中，重组位点可以是噬菌体 P1 的 Cre 重组酶识别的 *lox* 位点。*Lox* 位点指噬菌体 P1 的 *cre* 基因产物 Cre 重组酶可以催化位点特异性重组事件所处的核苷酸序列。多种 *lox* 位点是本领域已知的，包括天然存在的 *loxP*、*loxB*、*loxL* 和 *loxR*，以及许多突变体或者变体 *lox* 位点，如 *loxP511*、*loxP514*、*lox.DELTA.86*、*lox.DELTA.117*、*loxC2*、*loxP2*、*loxP3* 和 *lox P23*。*Lox* 位点的额外实例包括，但不限于 *loxB*、*loxL*、*loxR*、*loxP*、*loxP3*、*loxP23*、*lox Δ 86*、*lox Δ 117*、*loxP511* 和 *loxC2*。

在另一实施方案中，重组位点是非 Cre 的重组酶识别的重组位点。在一个实施方案中，重组酶位点可以是啤酒糖酵母的 2*pi* 质粒的 FLP 重组酶识别的 FRT 位点。FRT 位点指酵母 2 μ 质粒的 FLP 基因产物 FLP 重组酶催化位点特异性重组所处的核苷酸序列。非-Cre 重组酶的额外实例包括，但不限于位点特异性重组酶，包括： λ 噬菌体的 Int 重组酶识别的 *att* 位点（例如，*att1*、*att2*、*att3*、*attP*、*attB*、*attL* 和 *attR*）、解离酶家族识别的重组位点和苏云金芽孢杆菌的转座酶识别的重组位点。

IV. 产生基因修饰动物

在本发明的额外方面，通过本领域技术人员已知的任意方法可以产生含有本文描述的基因修饰的有蹄类动物。此类方法包括，但不限于：核转移、胞质内精子注射技术、合子直接修饰和精子介导的基因转移。

在另一实施方案中，克隆此类动物如猪的方法包括：将卵母细胞去核、将卵母细胞与来自细胞的供体核融合，所述细胞中至少一种免疫球蛋白基因的至少一个等位基因已经被灭活，和将核转移得到的胚胎植入替代母亲中。

备选地，通过灭活胚胎干细胞中免疫球蛋白基因的两个等位基

因，提供了产生缺少功能性免疫球蛋白的任何表达的活动物的方法，所述细胞然后可以用于产生后代。

在另一方面，本发明还提供了产生活动物，如猪的方法，其中已经使得免疫球蛋白基因的两个等位基因都失活。在一个实施方案中，通过使用来自细胞的供体细胞核克隆产生了动物，所述细胞中免疫球蛋白基因的两个等位基因都被灭活。在一个实施方案中，通过遗传寻靶事件灭活免疫球蛋白基因的两个等位基因。

通过直接修饰合子可以产生基因改造动物。对于哺乳动物，可以将经修饰的合子导入能够怀有该动物至足月的假孕雌性的子宫中。例如，如果需要缺少免疫球蛋白基因的完整动物，那么来自该动物的胚胎干细胞可以被寻靶并且以后导入胚胎中以使修饰的细胞生长成嵌合动物。对于胚胎干细胞，可以使用胚胎干细胞系或者新鲜得到的干细胞。

在本发明的合适的实施方案中，全能细胞是胚胎干(ES)细胞。从胚胎分离ES细胞、确立ES细胞系和它们的随后培养可通过常规方法进行，如 Doetchmann 等人, *J. Embryol. Exp. Morph.* 87:27-45 (1985); Li 等人, *Cell* 69:915-926 (1992); Robertson, E. J. "Tetracarcinomas and Embryonic Stem Cells: A Practical Approach," ed. E. J. Robertson, IRL Press, Oxford, England (1987); Wurst 和 Joyner, "Gene Targeting: A Practical Approach," ed. A. L. Joyner, IRL Press, Oxford, England (1993); Hogen 等人, "Manipulating the Mouse Embryo: A Laboratory Manual," eds. Hogan, Beddington, Costantini and Lacy, Cold Spring Harbor Laboratory Press, New York (1994); and Wang 等人, *Nature* 336:741-744 (1992)所述。在本发明的另一合适的实施方案中，全能细胞是胚胎生殖干(EG)细胞。胚胎生殖干细胞是与ES细胞在功能上等价的未分化的细胞，即，它们可以在体外培养和转染，然后有助于嵌合体的体细胞和生殖细胞谱系(Stewart 等人, *Dev. Biol.* 161:626-628 (1994))。通过培养原生殖细胞——配子的祖先与生长因子：白血病抑制因子、青灰因子和碱性成纤维细胞生长因子的组合得到EG细胞(Matsui 等人, *Cell* 70:841-847 (1992); Resnick 等人, *Nature* 359:550-551 (1992))。可以使用 Donovan 等人, "Transgenic Animals,

Generation and Use," Ed. L. M. Houdebine, Harwood Academic Publishers (1997)和其中引用的原始文献中描述的方法进行 EG 细胞的培养。

用于本发明的四倍体胚泡可以通过天然合子产生和发育得到,或者通过已知的二细胞胚胎的电融合和随后的培养得到,如 James 等人, *Genet. Res. Camb.* 60:185-194 (1992); Nagy 和 Rossant, "Gene Targeting: A Practical Approach," ed. A. L. Joyner, IRL Press, Oxford, England (1993); 或 Kubiak 和 Tarkowski, *Exp. Cell Res.* 157:561-566 (1985)所述。

可以通过本领域已知的任意方法进行 ES 细胞或者 EG 细胞向胚泡中的导入。用于本发明目的的合适的方法是如 Wang 等人, *EMBO J.* 10:2437-2450 (1991)所述的显微注射方法。

备选地,通过修饰的胚胎干细胞可以产生转基因动物。可以根据常规技术将基因修饰胚胎干细胞注射到胚泡然后在雌性宿主哺乳动物中使其足月。然后可以使用诸如 PCR 或者 DNA 印迹的技术,对杂合后代筛选靶基因座位点上所述改变的存在。与相同物种的野生型宿主交配后,可以将所得嵌合子代杂交以得到纯合宿主。

用寻靶载体转化胚胎干细胞以改变免疫球蛋白基因后,可以将细胞铺平板在合适的培养基,如胎牛血清增强的 DMEM 中的饲养层上。使用选择培养基可以检测含有构建体的细胞,并且菌落生长足够的时间后,可以挑取菌落并分析同源重组的发生。可以使用聚合酶链反应,引物在构建体序列内和外,但是在靶基因座上。然后显示出同源重组的那些菌落可以用于胚胎操作和胚泡注射。可以从超排卵的雌性得到胚泡。然后可以将胚胎干细胞胰蛋白酶消化并将经修饰的细胞加入含有胚泡的小滴。可以将至少一个经修饰的胚胎干细胞注射到胚泡的囊胚腔中。注射后,至少一个胚泡可以返回到假孕雌性的每个子宫角。然后允许雌性足月并对所得的胎仔(litter)筛选具有所述构建体的突变细胞。从转化的 ES 细胞选择胚泡的不同家系。通过提供不同表型的胚泡和 ES 细胞,可以容易地检测嵌合后代,且然后进行基因型分型以探测经修饰的免疫球蛋白基因的存在。

在其他实施方案中,精子介导的基因转移可以用于产生本文描述的基因修饰有蹄类动物。本文描述的用于消除内源免疫球蛋白的表达

或者插入异种免疫球蛋白基因的方法和组合物可以用于通过本文描述或者本领域已知的任意技术基因修饰精细胞。基因修饰精子然后通过人工受精、胞质内精子注射技术或者任何其他已知的技术用于使雌性受体怀孕。在一个实施方案中，精子和/或精子头可以与外源核酸温育足够的时间段。足够的时间段包括例如，约 30 秒到约 5 分钟，通常约 45 秒到约 3 分钟，更一般约 1 分钟到约 2 分钟。在具体实施方案中，异种的，如人免疫球蛋白基因在如本文描述的有蹄类动物中的表达可以通过胞质内精子注射技术完成。

精细胞作为基因转移的载体的潜在用途首次由 Brackett 等人, *Proc. Natl. Acad. Sci. USA* 68:353-357 (1971) 提出。接着报导了卵母细胞与已经通过裸 DNA 温育的精子的体外受精后产生转基因小鼠和猪 (见, 例如, Lavitrano 等人, *Cell* 57:717-723 (1989) 和 Gandolfi 等人 *Journal of Reproduction and Fertility Abstract Series* 4, 10 (1989)), 尽管其他实验室不能重复这些实验 (见例如, Brinster 等人 *Cell* 59:239-241 (1989) 和 Gavora 等人, *Canadian Journal of Animal Science* 71:287-291 (1991))。从这以后, 几个报导描绘了鸡 (见, 例如, Nakanishi 和 Iritani, *Mol. Reprod. Dev.* 36:258-261 (1993)); 小鼠 (见, 例如, Maione, *Mol. Reprod. Dev.* 59:406 (1998)); 和猪 (见, 例如, Lavitrano 等人 *Transplant. Proc.* 29:3508-3509 (1997); Lavitrano 等人, *Proc. Natl. Acad. Sci. USA* 99:14230-5 (2002); Lavitrano 等人, *Mol. Reprod. Dev.* 64:284-91 (2003)) 的成功精子介导的基因转移。类似的技术也在 2002 年 4 月 23 日授权的美国专利号 6,376,743、2001 年 11 月 22 日公开的美国专利公开号 20010044937 和 2002 年 8 月 8 日公开的 20020108132 中描述。

在其他实施方案中, 胞质内精子注射可以用于产生本文描述的基因修饰有蹄类动物。这可以如下完成: 将外源核酸和精子共同插入 (coinserting) 未受精的卵母细胞的细胞质中以形成转基因的受精的卵母细胞, 并允许转基因的受精的卵母细胞发育成转基因胚胎, 且如果需要, 发育成活的后代。精子可以是膜破坏的精子头或者去膜 (demembrated) 的精子头。共同插入步骤可以包括将精子与外源核酸预温育足够的时间段的子步骤, 如约 30 秒到约 5 分钟, 通常约 45 秒到约 3 分钟, 更通常约 1 分钟到约 2 分钟。通过显微注射可以将精子

和外源核酸共同插入卵母细胞中。与精子混合的外源核酸可以含有超过一种的转基因，以产生对本文描述的超过一种的转基因的胚胎。胞质内精子注射可以通过本领域已知的任一种技术完成，见，例如，美国专利号 6,376,743。在具体实施方案中，异种的，如人免疫球蛋白基因在如本文描述的有蹄类动物中的表达可以通过胞质内精子注射完成。

本领域已知的任何额外的技术都可以用于向动物中导入转基因。此类技术包括，但不限于前核显微注射（见，例如，Hoppe, P. C. 和 Wagner, T. E., 1989, 美国专利号 4,873,191）；向种系的逆转录病毒介导的基因转移（见，例如，Van der Putten 等人，1985, Proc. Natl. Acad. Sci., USA 82:6148-6152）；胚胎干细胞中的基因寻靶（见，例如，Thompson 等人，1989, Cell 56:313-321；Wheeler, M. B., 1994, WO 94/26884）；胚胎的电穿孔（见，例如，Lo, 1983, Mol Cell. Biol. 3:1803-1814）；细胞枪（cell gun）；转染；转导；逆转录病毒感染；腺病毒感染；腺伴随病毒感染；脂质体介导的基因转移；裸 DNA 转移；和精子介导的基因转移（见，例如，Lavitrano 等人，1989, Cell 57:717-723）等等。关于这种技术的综述，见例如，Gordon, 1989, Transgenic Animals, Intl. Rev. Cytol. 115:171-229。在具体实施方案中，通过这些技术可以完成异种的如人免疫球蛋白基因在如本文描述的有蹄类动物中的表达。

体细胞核转移以产生克隆的转基因后代

在本发明的另一方面，缺少有蹄类动物重链、 κ 轻链和/或 λ 轻链基因的一个等位基因、任选地两个等位基因的有蹄类动物，如猪或者牛细胞可以用作供体细胞以核转移到受体细胞中而产生克隆的转基因动物。备选地，可以在胚胎干细胞中产生有蹄类动物重链、 κ 轻链和/或 λ 轻链基因敲除体，其然后用于产生后代。根据本文描述的方法、序列和/或构建体产生的缺少功能性有蹄类动物重链、 κ 轻链和/或 λ 轻链基因的单个等位基因的后代可以育种以进一步通过孟德尔型遗传产生在两个等位基因中缺少功能性的后代。

在另一实施方案中，本发明提供了产生缺少功能性 α -1,3-GT 的任何表达的活猪的方法，其通过将 α -1,3-GT 基因杂合的公猪与对 α -1,3-GT 基因杂合的母猪育种实现。在一个实施方案中，由于 α -

1,3-GT 基因的一个等位基因的基因修饰以防止该等位基因表达, 所以猪是杂合的。在另一实施方案中, 由于 α -1,3-GT 基因的一个等位基因中存在点突变, 所以猪是杂合的。在另一实施方案中, 点突变可以是在 α -1,3-GT 基因的外显子 9 的第二个碱基处 T 到 G 的点突变。在一个具体实施方案中, 提供了产生缺少功能性 α -1,3-GT 的任何表达的猪动物的方法, 其中将在 α -1,3-GT 基因的外显子 9 的第二个碱基处含有 T 到 G 的点突变的公猪与在 α -1,3-GT 基因的外显子 9 的第二个碱基处含有 T 到 G 的点突变的母猪育种, 或反过来也一样。

本发明提供了通过体细胞核转移克隆缺少功能性免疫球蛋白基因的动物如猪的方法。通常, 通过核转移方法可以产生动物, 所述方法包括下面的步骤: 得到将用作供体核来源的所需的分化细胞; 从动物得到卵母细胞; 将所述卵母细胞去核; 将所需的分化细胞或者细胞核转移到去核的卵母细胞中, 例如, 通过融合或者注射, 以形成 NT 单位; 激活所得的 NT 单位; 并将所述培养的 NT 单位转移到宿主动物, 从而 NT 单位发育成胎儿。

核转移技术或者核移植技术是本领域已知的(Dai 等人 *Nature Biotechnology* 20:251-255; Polejaeva 等人, *Nature* 407:86-90 (2000); Campbell 等人, *Theriogenology*, 43:181 (1995); Collas 等人, *Mol. Report Dev.*, 38:264-267 (1994); Keefer 等人, *Biol. Reprod.*, 50:935-939 (1994); Sims 等人, *Proc. Natl. Acad. Sci., USA*, 90:6143-6147 (1993); WO 94/26884; WO 94/24274, 和 WO 90/03432, 美国专利号 4,944,384 和 5,057,420)。

将已经修饰以改变免疫球蛋白基因的供体细胞核转移到受体卵母细胞。该方法的用途不限于具体供体细胞类型。供体细胞可以如本文描述, 也见例如, Wilmut 等人, *Nature* 385 810 (1997); Campbell 等人, *Nature* 380 64-66 (1996); Dai 等人, *Nature Biotechnology* 20:251-255, 2002 或 Cibelli 等人, *Science* 280 1256-1258 (1998)。可以使用可成功地用于核转移的正常核型的所有细胞, 包括胚胎、胎儿和成年体细胞。胎儿成纤维细胞是尤其有用的一类供体细胞。通常, 用于核转移的合适的方法在 Campbell 等人, *Theriogenology* 43 181 (1995), Dai 等人 *Nature Biotechnology* 20:251-255, Polejaeva 等人, *Nature* 407:86-90 (2000), Collas 等人, *Mol. Reprod. Dev.* 38

264-267 (1994), Keefer 等人, Biol. Reprod. 50 935-939 (1994), Sims 等人, Proc. Nat'l. Acad. Sci. USA 90 6143-6147 (1993), WO-A-9426884、WO-A-9424274、WO-A-9807841、WO-A-9003432、美国专利号 4,994,384 和美国专利号 5,057,420 中描述。还可以使用分化或者至少部分分化的供体细胞。供体细胞还可以但不必须是培养的并且可以是休眠的。休眠的核供体细胞是可以被诱导以进入休眠或者在体内以休眠状态存在的细胞。现有技术方法也已经在克隆程序中使用胚胎细胞类型(Campbell 等人, (Nature, 380:64-68, 1996)和 Stice 等人, (Biol. Reprod., 20 54:100-110, 1996)。

可以从多种不同的器官和组织得到体细胞核供体细胞, 所述器官和组织如, 但不限于, 皮肤、间充质、肺、胰、心脏、肠、胃、膀胱、血管、肾、尿道、生殖器官, 和胚胎、胎儿或成年动物的完整或者部分的解聚制剂。在本发明的合适的实施方案中, 核供体细胞选自上皮细胞、成纤维细胞、神经细胞、角质形成细胞、造血细胞、黑素细胞、软骨细胞、淋巴细胞(B和T)、巨噬细胞、单核细胞、单核的细胞、心肌细胞、其他肌细胞、 \square xtended \square 细胞、卵丘细胞、表皮细胞或内皮细胞。在另一实施方案中, 核供体细胞是胚胎干细胞。在具体实施方案中, 成纤维细胞可以用作供体细胞。

在本发明的另一实施方案中, 本发明的核供体细胞是动物的生殖细胞。动物物种的处于胚胎、胎儿或者成年阶段的任何生殖细胞都可以用作核供体细胞。在合适的实施方案中, 核供体细胞是胚胎生殖细胞。

核供体细胞可以停滞在细胞周期的任意阶段(G₀、G₁、G₂、S、M)以便确保与受体细胞协调。本领域已知的任意方法都可以用于操作细胞周期阶段。控制细胞周期阶段的方法包括, 但不限于通过培养的细胞的接触抑制诱导的G₀休眠、通过除去血清或者其他必需营养物质诱导的G₀休眠、通过衰老诱导的G₀休眠、通过加入特定生长因子诱导的G₀休眠; 通过物理或者化学手段如热激、高压或者用化学制剂、激素、生长因子或者其他物质处理诱导的G₀或者G₁休眠; 通过用干扰复制过程的任意点的化学试剂处理进行S期控制; 通过用荧光激活细胞分选、有丝分裂抖落、用微管破坏剂或破坏有丝分裂进展的任何化学制剂处理进行M期控制(也见 Freshney, R. I., "Culture

of Animal Cells: A Manual of Basic Technique," Alan R. Liss, Inc, New York (1983).

分离卵母细胞的方法是本领域公知的。基本上,这可以包括从动物的卵巢或者生殖道分离卵母细胞。卵母细胞的容易获得的来源是屠宰场材料。对于如基因工程、核转移和克隆的技术的组合,卵母细胞必须通常在体内成熟,之后这些细胞才可以用作核转移的受体细胞,并且之后它们通过由精细胞受精以发育成胚胎。该方法通常需要从哺乳动物卵巢,如在屠宰场得到的牛卵巢收集不成熟(前期 I)卵母细胞,并在受精或者去核前在成熟培养基中成熟,直到卵母细胞达到中期 II 阶段,其对于牛卵母细胞通常在抽吸后约 18-24 小时发生。该时间段称作“成熟时间段”。在一些实施方案中,从育肥小母猪(gilt)得到卵母细胞。“育肥小母猪”是从未有过后代的母猪。在其他实施方案中,从母猪得到卵母细胞。“母猪”是以前产生后代的母猪。

中期 II 阶段卵母细胞可以是受体卵母细胞,在该阶段认为卵母细胞可以被或者被足够“激活”以与处理受精精子一样处理导入的细胞核。已经在体内成熟的中期 II 阶段卵母细胞已经成功地用于核转移技术。基本上,可以在动情期开始后或者人绒毛膜促性腺激素(hCG)或者类似的激素注射后 35 到 48、或者 39 到 41 小时从非超排卵的或者超排卵的动物通过外科手术收集成熟的中期 II 卵母细胞。可以将卵母细胞置于合适的培养基,如透明质酸酶(hyalurodase)溶液中。

固定时间的成熟时间段(其为约 10 到 40 小时,约 16-18 小时,约 40-42 小时或约 39-41 小时)后,可以对卵母细胞去核。在去核前,可以取出卵母细胞并将其放置在合适的培养基,如含有 1mg/ml 透明质酸酶的 HECM 中,之后取出卵丘细胞。可以对带状卵母细胞筛选极体,并且然后将如通过极体的存在选择的中期 II 卵母细胞用于核转移。如下所述去核。

通过已知的方法,如美国专利号 4,994,384 中所述的进行去核。例如,可以将中期 II 卵母细胞置于 HECM 中,任选含有 7.5 微克/微升松胞菌素 B,用于立即去核,或者可以置于合适的培养基,如胚胎培养基如 CR1aa 加 10% 动情期奶牛血清中,且然后去核,如不超过 24 小时后,或者不超过 16-18 小时后。

可以通过显微外科手术完成去核,其中使用微量移液管以取出极

体和相邻的细胞质。然后可以筛选卵母细胞以鉴定已经成功去核的细胞。一种筛选卵母细胞的方法是用溶于 HECM 中的 1 微克/毫升 33342 Hoechst 染料对卵母细胞染色，并然后在紫外线照射下观察卵母细胞小于 10 秒。然后可以将成功去核的卵母细胞置于合适的培养基，如 CR1aa 加 10% 血清中。

然后可以将与去核的卵母细胞相同物种的单个哺乳动物细胞转移到用于产生 NT 单位的去核卵母细胞的卵周隙。哺乳动物细胞和去核的卵母细胞可以用于根据本领域已知的方法产生 NT 单位。例如，可以通过电融合融合细胞。通过提供足够导致质膜的瞬时破裂的电脉冲完成电融合。该质膜的破裂非常短，因为膜快速重新形成。从而，如果诱导两个相邻的膜破裂并且在重新形成后，脂质双层混合，那么两个细胞之间可以打开小的通道。由于这种小开口的热力学不稳定性，它变大，直到两个细胞变成一个。见例如，Prather 等人的美国专利号 4,997,384。可以使用多种电融合介质，包括例如，蔗糖、甘露糖醇、山梨糖醇和磷酸缓冲溶液。使用仙台病毒作为融合剂也可以完成融合(Graham, Wister Inot. Symp. Monogr., 9, 19, 1969)。同样，不使用电穿孔融合，可以将细胞核直接注射到卵母细胞中。见，例如，Collas 和 Barnes, Mol. Reprod. Dev., 38:264-267 (1994)。融合后，将所得融合的 NT 单位置于合适的培养基中直到活化，例如，CR1aa 培养基。通常融合之后很短时间内，如融合后小于 24 小时后，或者约 4 - 9 小时后，或者最佳地融合后 1 - 2 小时实现活化。在具体实施方案中，在融合后至少 1 小时和成熟后 40 - 41 小时发生活化。

通过已知方法可以活化 NT 单位。此类方法包括例如，在亚生理学 (sub-physiological) 温度，大体上通过对 NT 单位应用冷，或者实际上凉的温度休克来培养 NT 单位。这可以最方便地通过在室温下培养 NT 来进行，所述室温相对于胚胎通常暴露的生理学温度条件是冷的。备选地，通过应用已知的活化剂可以实现活化。例如，已经表明在受精期间通过精子穿透卵母细胞可以活化预融合卵母细胞以在核转移后产生更大数目的活的妊娠和多个遗传上相同的小牛。同样，处理如电和化学休克可以用于在融合后活化 NT 胚胎。见例如，Susko-Parrish 等人的美国专利号 5,496,720。通过使用 ECM2001 Electroculture Manipulator (BTX Inc., San Diego, CA)应用 5V 的 AC 脉

冲 5 秒接着 1.5 kV/cm 的两次每次 60 微秒的 DC 脉冲,可以诱导融合和活化。此外,通过升高卵母细胞中二价阳离子的水平,并减小卵母细胞中细胞蛋白质的磷酸化,可以同时或者顺序实现活化。这可以通过通常通过将二价阳离子导入卵母细胞细胞质来实现,所述二价阳离子为例如镁、镉、钡或者钙,例如,以离子载体的形式。提高二价阳离子水平的其他方法包括使用电休克、用乙醇处理和用笼状(caged)螯合剂处理。通过已知方法可以减小磷酸化,例如,通过加入激酶抑制剂,如丝氨酸-苏氨酸激酶抑制剂,如 6-甲基-氨基嘌呤、星形孢菌素、2-氨基嘌呤和鞘氨醇。备选地,通过向卵母细胞中导入磷酸酶,如磷酸酶 2A 和磷酸酶 2B 可以抑制细胞蛋白质的磷酸化。

活化的 NT 单位,或者“融合的胚胎”然后可以在合适的体外培养基中培养直到产生细胞集落。适于胚胎的培养和成熟的培养基是本领域公知的。可以用于胚胎培养和维持的已知培养基的实例包括 Ham's F-10+10%胎牛血清(FCS)、组织培养基-199(TCM-199)+10%胎牛血清、Tyrodes-清蛋白-乳酸-丙酮酸(TALP)、Dulbecco 氏磷酸缓冲盐水(PBS)、Eagle 氏和 Whitten 氏培养基,和在一个具体实例中,活化的 NT 单位可以在 NCSU-23 培养基中于约 38.6°C 在 5% CO₂ 的湿润气氛中培养约 1-4 小时。

之后,可以洗涤培养的一个或多个 NT 单位,然后将其置于有孔板中所含的合适的培养基中,所述有孔板可以含有合适的汇合饲养层。合适的饲养层包括例如成纤维细胞和上皮细胞。将 NT 单位在饲养层上培养直到 NT 单位达到适于转移到受体雌性,或者适于得到可以用于产生细胞集落的细胞的大小。可以培养这些 NT 单位直到至少约 2 到 400 个细胞、约 4 到 128 个细胞或者至少约 50 个细胞。

然后将活化的 NT 单位转移(胚胎转移)到母猪的输卵管中。在一个实施方案中,母猪可以是动情期同步化的受体育肥小母猪。可以使用杂种育肥小母猪(大白/Duroc/Landrace)(280-400 lbs)。通过经口施用混合到饲料中的 18-20 mg Regu-Mate (Altrenogest, Hoechst, Warren, NJ)可以将育肥小母猪同步化为受体动物。Regu-Mate 可以喂饲 14 个连续日。最后的 Regu-Mate 处理后约 105 小时,可以肌肉(i.m.)施用 1000 单位的人绒毛膜促性腺激素(hCG, Intervet America, Millsboro, DE)。hCG 注射后约 22-26 小时可以进行胚胎转移。在一

个实施方案中，可以使妊娠达到足月，并导致生出活的后代。在另一实施方案中，妊娠可以早期结束并且收获胚胎细胞。

育种得到所需纯合敲除动物

在另一方面，本发明提供了产生缺少功能免疫球蛋白基因的任何表达的活的动物的方法，其通过将对该免疫球蛋白基因杂合的雄性与对该免疫球蛋白基因杂合的雌性育种而实现。在一个实施方案中，由于免疫球蛋白基因的一个等位基因的基因修饰以防止该等位基因的表达，所以动物是杂合的。在另一实施方案中，由于在 α -免疫球蛋白基因的一个等位基因中的点突变的存在，所以动物是杂合的。在其他实施方案中，此类杂合敲除体可以与表达异种免疫球蛋白如人免疫球蛋白的有蹄类动物育种。在一个实施方案中，通过将缺少内源免疫球蛋白的至少一个等位基因的表达的转基因有蹄类动物与表达异种免疫球蛋白的有蹄类动物育种可以得到动物，其中该内源免疫球蛋白选自重链、 κ 轻链和 λ 轻链或者其任意组合。在另一实施方案中，通过将缺少重链、 κ 轻链和 λ 轻链的一个等位基因的表达的转基因有蹄类动物与表达异种如人免疫球蛋白的有蹄类动物育种可以得到动物。在进一步的实施方案中，通过将缺少重链、 κ 轻链和 λ 轻链的一个等位基因的表达并且表达异种如人免疫球蛋白的转基因有蹄类动物与缺少重链、 κ 轻链和 λ 轻链的一个等位基因的表达并表达异种如人免疫球蛋白的另一个转基因有蹄类动物育种，以产生缺少重链、 κ 轻链和 λ 轻链的两个等位基因的表达并表达异种如人免疫球蛋白的纯合的转基因有蹄类动物，可得到动物。还提供了产生此类动物的方法。

在一个实施方案中，从携带免疫球蛋白基因双敲除的供体细胞的核转移产生的性成熟的动物可以育种并且测试它们后代的纯合敲除。然后这些纯合敲除动物可以育种以产生更多动物。

在另一个实施方案中，来自性成熟的双敲除动物的卵母细胞可以使用来自两个遗传上不同的猪品系的野生型精子在体外受精，并将胚胎植入合适的替代动物中。可以对来自这些交配的后代测试敲除的存在，例如，通过cDNA测序，和/或PCR可以测试它们。然后，在性成熟时，来自这些胎仔每一只的动物可以交配。在根据本发明的该方

面的一些方法中,妊娠可以早期结束从而可以分离胎儿成纤维细胞并进一步在表型和/或基因型上表征。缺少免疫球蛋白基因的表达的成纤维细胞可以用于根据本文描述的方法进行核转移以产生携带所需双敲除的多个妊娠和后代。

额外的基因修饰

在其他实施方案中,根据本文描述的方法、序列和/或构建体产生的缺少功能性免疫球蛋白的表的动物或细胞可以含有额外的基因修饰以消除异种抗原的表达。通过进一步基因修饰从本文描述的转基因细胞和动物得到的细胞或者通过将本文描述的动物与已经进一步基因修饰的动物育种,可以进行额外的基因修饰。可以修饰此类动物以消除 α -1,3-半乳糖基转移酶基因、CMP-Neu5Ac 羟化酶基因(见例如, USSN 10/863,116)、iGb3 合酶基因(见,例如,美国专利申请 60/517,524),和/或福斯曼合酶基因(见例如,美国专利申请 60/568,922)的至少一个等位基因的表达。在额外实施方案中,本文公开的动物还可以含有基因修饰以表达岩藻糖基转移酶、唾液酸转移酶和/或葡糖基转移酶家族的任意成员。为了实现这些额外的基因修饰,在一个实施方案中,可以修饰细胞以含有多个基因修饰。在其他实施方案中,动物可以一起育种以实现多个基因修饰。在一个具体实施方案中,根据本文描述的方法、序列和/或构建体产生的缺少功能性免疫球蛋白表的动物,如猪可以与缺少 α -1,3-半乳糖基转移酶表的动物如猪育种(如 WO 04/028243 中所述)。

在另一个实施方案中,可以消除或者减少负责异种移植物排斥的额外基因的表达。此类基因包括,但不限于 CMP- NEUAc 羟化酶基因、异红细胞糖苷脂(isoGloboside)3 合酶基因,和福斯曼合酶基因。此外,负责补体介导的裂解的抑制的编码补体相关蛋白质的基因或者 cDNA 还可以在本发明的动物和组织中表达。此类基因包括,但不限于 CD59、DAF、MCP 和 CD46(见,例如 WO 99/53042; Chen 等人 *Xenotransplantation*, Volume 6 Issue 3 Page 194 - August 1999,其描述了表达 CD59/DAF 转基因的猪; Costa C 等人, *Xenotransplantation*. 2002 Jan; 9(1):45-57, 其描述了表达人 CD59 和 H 转移酶的转基因猪; Zhao 等人; Diamond LE 等人 *Transplantation*.

2001 Jan 15; 71(1):132-42, 其描述人 CD46 转基因猪。

额外的修饰可以包括表达组织因子途径抑制剂 (TFPI)、肝素、抗凝血酶、蛭素、TFPI、壁虱抗凝肽, 或者蛇毒液因子, 如 WO 98/42850 和美国专利号 6,423,316, 标题 "Anticoagulant fusion protein anchored to cell membrane" 中所述; 或者下调细胞的细胞粘着分子表达的化合物, 如抗体, 如 WO 00/31126, 标题为 "Suppression of xenograft rejection by down regulation of a cell adhesion molecules" 所述, 和其中阻止信号 2 的共刺激的化合物, 如通过对器官受体使用来自异种供体生物的 CTLA-4 的可溶形式的, 如 WO 99/57266, 标题 "Immunosuppression by blocking T cell co-stimulation signal 2 (B7/CD28 interaction)" 所述。

在下面的非限制性实施例中更详细地描述本发明的一些方面。

实施例

实施例 1: 猪重链寻靶和缺少重链表达的猪动物的产生

从 3X 冗余猪 BAC 文库分离猪 Ig 重链基因座的部分。通常, 通过断裂猪总基因组 DNA, 其然后可以用于得到代表完整动物的基因组的至少三倍的 BAC 文库, 可以产生 BAC 文库。然后通过对本文描述的猪重链免疫球蛋白有选择性的探针杂交, 选择含有猪重链免疫球蛋白的 BAC。

来自克隆的序列 (Seq ID 1) 用于产生与 J 区的部分互补的引物 (该引物由 Seq ID No. 2 表示)。单独地, 设计与 Ig 重链 μ 恒定区的部分互补的引物 (该引物由 Seq ID No. 3 表示)。这些引物用于扩增猪 Ig 重链的片段 (由 Seq ID No. 4 表示), 其使功能性连接区 (J 区) 和足够的侧翼区用以设计和构建寻靶载体。为了维持该片段和该片段的亚克隆处于天然状态, 将含有这些片段的大肠杆菌 (Stable 2, Invitrogen cat #1026-019) 维持在 30°C。亚克隆 Seq ID No. 4 的区域并将其用于装配如 Seq ID No. 5 所示的寻靶载体。将该载体转染到猪胎儿成纤维细胞中, 随后将其用 G418 选择。通过 PCR 筛选所得集落以检测可能的寻靶事件 (Seq ID No. 6 和 Seq ID No. 7, 5' 筛选引物; Seq ID No. 8 和 Seq ID No. 9, 3' 筛选引物)。关于寻靶的示例性图解见图 1。通过 DNA 印迹验证寻靶。使用靶定的胎儿成纤维细胞作为核

供体通过核转移产生小猪。

核转移

将靶定的胎儿成纤维细胞用作核供体细胞。通过本领域公知的方法进行核转移（见，例如，Dai 等人，*Nature Biotechnology* 20: 251-255, 2002；和 Polejaeva 等人，*Nature* 407:86-90, 2000）。

hCG 注射后 46-54 小时用含有牛血清清蛋白（BSA;4gl⁻¹）的预热的 Dulbecco 氏磷酸缓冲盐水（PBS）通过输卵管的反向冲洗收集卵母细胞（如 Polejaeva, I.A., 等人（*Nature* 407, 86-90 (2000)中所述）。如 Polejaeva, I.A., 等人（*Nature* 407, 86-90 (2000)）所述，在成熟后 40 到 42 小时开始体外成熟的卵母细胞（BioMed, Madison, WI）的去核。将回收的卵母细胞在含有 4gl⁻¹ BSA 的 PBS 中于 38℃ 洗涤，并转移到 38℃ 的无钙的磷酸缓冲的 NCSU-23 培养基中用于转运到实验室。为了去核，我们将卵母细胞在含有 5 μg ml⁻¹ 松胞菌素 B (Sigma) 和 7.5 μg ml⁻¹ Hoechst 33342 (Sigma) 的无钙的磷酸缓冲的 NCSU-23 培养基中在 38℃ 温育 20 分钟。然后使用 18 μM 玻璃移液管 (Humagen, Charlottesville, Virginia) 直接在第一个极体下抽吸少量的细胞质。我们将抽吸的核体暴露于紫外线以证实存在赤道板。

为了核转移，将单个成纤维细胞置于透明带下与每个去核的卵母细胞接触。使用 ECM2001 Electrocell Manipulator (BTX Inc., San Diego, CA)，通过应用 5V 的 AC 脉冲 5 秒，接着 1.5 kV/cm 的两次每次 60 微秒的 DC 脉冲诱导融合和活化。融合的胚胎在 NCSU-23 培养基中于 38.6℃ 在 5% CO₂ 的湿润气氛中培养 1-4 小时，然后转移到动情期同步化的受体育肥小母猪的输卵管中。通过经口施用混合到其饲料中的 18-20 mg Regu-Mate (Altrenogest, Hoechst, Warren, NJ) 将杂种育肥小母猪（大白/Duroc/landrace）（280-400 lbs）同步化为受体。Regu-Mate 喂饲 14 个连续日。最后的 Regu-Mate 处理后 105 小时，肌肉施用人绒毛膜促性腺激素（hCG, 1000 单位；Intervet America, Millsboro, DE）。hCG 注射后 22-26 小时进行胚胎转移。

核转移从四胎仔产下 18 个健康小猪。这些动物具有一个功能性野生型 Ig 重链基因座和一个破坏的 Ig 重链基因座。

Seq ID 2:来自 Butler 亚克隆用

ggccagacttctcggaacagctca

于扩增 J 到 C 重链的引物 (637Xba5')	
Seq ID 3:用于扩增 J 到 C 重链 的 C 的引物 (JM1L)	tccaggagaaggtgacggagct
Seq ID 6:用于 5'筛选的重链 5' 引物 (HCKOXba5'2)	tctagaagacgctggagagaggccag
Seq ID 7: 用于 5'筛选的重链 3'引物(5'arm5')	taaagcgcagctccagactgcctt
Seq ID 8: 用于 3'筛选的重链 5'引物(NEO4425)	catcgccttctatcgccttctt
Seq ID 9: 用于 3'筛选的重链 3'引物(650+CA)	Aagtactgccgcctctcagga

细胞和猪组织样品的 DNA 印迹分析。将细胞或者组织样品在裂解缓冲液(10 mM Tris, pH 7.5, 10mM EDTA, 10mM NaCl, 0.5% (w/v) 十二烷基肌氨酸钠, 1 mg/ml 蛋白酶 K)中于 60℃ 过夜裂解并用乙醇沉淀 DNA。然后将 DNA 用 NcoI 或 XbaI 消化, 这取决于所用的探针, 并在 1% 琼脂糖凝胶上分离。电泳后, 将 DNA 转移到尼龙膜上并用洋地黄毒苷标记的探针 (SEQ ID No 41 用于 NcoI 消化, SEQ ID No 40 用于 XbaI 消化) 探测。使用化学发光底物系统 (Roche Molecular Biochemicals) 检测条带。

用于重链 DNA 印迹的探针:

HC J 探针 (用于 XbaI 消化):

```
CTCTGCACTCACTACCGCCGGACGCGCACTGCCGTGCTGCCATGGACCA
CGCTGGGGAGGGGTGAGCGGACAGCACGTTAGGAAGTGTGTGTGTGCGCG
TGGGTGCAAGTCGAGCCAAGGCCAAGATCCAGGGGCTGGGCCCTGTGCC
AGAGGAGAATGGCAGGTGGAGTGTAGCTGGATTGAAAGGTGGCCTGAAGG
GTGGGGCATCCTGTTTGGAGGCTCACTCTCAGCCCCAGGGTCTCTGGTTC
CTGCCGGGGTGGGGGGCGCAAGGTGCCTACCACACCCTGCTAGCCCCTCG
TCCAGTCCCAGGGCCTGCCTCTTACCACGGAAGAGGATAAGCCAGGCTGC
AGGCTTCATGTGCGCCGTGGAGAACCCAGTTCGGCCCTTGGAGG (Seq ID No 40)
```

HC μ 探针 (用于 NcoI 消化):

GGCTGAAGTCTGAGGCCTGGCAGATGAGCTTGGACGTGCGCTGGGGAGTA
 CTGGAGAAGGACTCCCGGGTGGGGACGAAGATGTTCAAGACGGG3GGCTG
 CTCCTCTACGACTGCAGGCAGGAACGGGGCGTCACTGTGCCGGCGGCACC
 CGGCCCCGCCCCGCCACAGCCACAGGGGGAGCCCAGCTCACCTGCCCCA
 GAGATGGACACGGACTTGGTGCCACTGGGGTGCTGGACCTCGCACACCAG
 GAAGGCCTCTGGGTCTGGGGGATGCTCACAGAGGGTAGGAGCACCCGGG
 AGGAGGCCAAGTACTTGCCGCCTCTCAGGACGG (Seq ID No 41)

实施例 2: 猪 κ 轻链寻靶和产生缺少 κ 轻链表达的猪

从 3X 冗余的猪 BAC 文库分离猪 Ig κ 链基因座的部分。通常，通过断裂猪总基因组 DNA，其然后可以用于得到代表完整动物的基因组的至少三倍的 BAC 文库，可以产生 BAC 文库。然后通过如本文描述的猪 κ 轻链免疫球蛋白选择性的探针杂交，可以选择含有猪 κ 轻链免疫球蛋白的 BAC。

使用与 J 区的部分互补的引物（该引物通过 Seq ID No. 10 表示）和与 κ C 区的部分互补的引物（该引物通过 Seq ID No. 11 表示）扩增猪 Ig 轻链 κ 的片段。将所得的扩增引物克隆到质粒载体中并在 Stable2 细胞中于 30°C 维持 (Seq ID No. 12)。示意性说明见图 2。

单独地，使用与 C 区的部分互补的引物 (Seq ID No. 13) 和与 κ 增强子区的区域互补的引物 (Seq ID No. 14) 扩增猪 Ig 轻链 κ 的片段。将所得的扩增引物通过限制酶断裂并将产生的 DNA 片段克隆，在 Stable2 细胞中于 30°C 维持并测序。作为测序的结果，装配了两个非重叠的重叠群 (Seq ID No. 15, 扩增引物的 5' 部分; 和 Seq ID No. 16, 扩增引物的 3' 部分)。来自下游重叠群的序列 (Seq ID No. 16) 用于设计一组引物 (Seq ID No. 17 和 Seq ID No. 18)，它们用于扩增增强子附近的连续片段 (Seq ID No. 19)。Seq ID No. 12 和 Seq ID No. 19 中每一个的亚克隆用于构建寻靶载体 (Seq ID No. 20)。将该载体转染到猪胎儿成纤维细胞中，其随后用 G418 进行选择。通过 PCR 筛选所得集落以检测潜在的寻靶事件 (Seq ID No. 21 和 Seq ID No. 22, 5' 筛选引物; 和 Seq ID No. 23 和 Seq ID No. 43, 3' 筛选引物, 和 Seq ID No. 24 和 Seq ID No. 24, 内源筛选引物)。通过 DNA 印迹证实寻靶。通过克隆额外的 κ 序列促进 DNA 印迹策略设计，所述序列对应于种系 κ 转录物的模板 (Seq ID No. 25)。通过核转移产生胎儿猪。

核转移

将靶定的胎儿成纤维细胞用作核供体细胞。通过本领域公知的方法进行核转移（见例如，Dai 等人，*Nature Biotechnology* 20: 251-255, 2002; 和 Polejaeva 等人，*Nature* 407:86-90, 2000）。

hCG 注射后 46-54 小时用含有牛血清清蛋白（BSA; 4g l^{-1} ）的预热的 Dulbecco 氏磷酸缓冲盐水（PBS）通过输卵管的反向冲洗收集卵母细胞（如 Polejaeva, I.A., 等人（*Nature* 407, 86-90 (2000) 中所述）。如 Polejaeva, I.A., 等人（*Nature* 407, 86-90 (2000)）所述，在成熟后 40 到 42 小时开始体外成熟的卵母细胞（BioMed, Madison, WI）的去核。将回收的卵母细胞在含有 4g l^{-1} BSA 的 PBS 中于 38°C 洗涤，并转移到 38°C 的无钙的磷酸缓冲的 NCSU-23 培养基中用于转运到实验室。为了去核，我们将卵母细胞在含有 $5\ \mu\text{g ml}^{-1}$ 松胞菌素 B (Sigma) 和 $7.5\ \mu\text{g ml}^{-1}$ Hoechst 33342 (Sigma) 的无钙的磷酸缓冲的 NCSU-23 培养基中在 38°C 温育 20 分钟。然后使用 $18\ \mu\text{M}$ 玻璃移液管 (Humagen, Charlottesville, Virginia) 直接在第一个极体下抽吸少量的细胞质。我们将抽吸的核体暴露于紫外线以证实存在赤道板。

为了核转移，将单个成纤维细胞置于透明带下与每个去核的卵母细胞接触。使用 ECM2001 Electrocell Manipulator (BTX Inc., San Diego, CA)，通过应用 5V 的 AC 脉冲 5 秒，接着 $1.5\ \text{kV/cm}$ 的两次每次 60 微秒的 DC 脉冲诱导融合和活化。融合的胚胎在 NCSU-23 培养基中于 38.6°C 在 $5\% \text{CO}_2$ 的湿润气氛中培养 1-4 小时，然后转移到动情期同步化的受体育肥小母猪的输卵管中。通过经口施用混合到其饲料中的 18-20 mg Regu-Mate (Altrenogest, Hoechst, Warren, NJ) 将杂种育肥小母猪（大白/Duroc/landrace）（280-400 lbs）同步化为受体。Regu-Mate 喂饲 14 个连续日。最后的 Regu-Mate 处理后 105 小时，肌肉内施用人绒毛膜促性腺激素（hCG, 1000 单位；Intervet America, Millsboro, DE）。hCG 注射后 22-26 小时进行胚胎转移。

使用 κ 靶定的核转移从 5 胎仔产下 33 个健康猪。这些猪具有一个功能性野生型猪 Ig 轻链 κ 等位基因和一个破坏的 Ig 轻链 κ 等位基因。

Seq ID 10: κ J 到 C 5' 引物 (kjc5'1)	caaggaqaccaagctggaactc
Seq ID 11: κ J 到 C 3' 引物	tgatcaagcacaccacagagacag

(kjc3'2)	
Seq ID 13: κ C 到 E 5'引物 (porKCS1)	gatgccaagccatccgtctcatc
Seq ID 14: κ C 到 E 3'引物 (porKCA1)	tgaccaaagcagtgtagcggtgc
Seq ID 17: 用于扩增增强子区的 κ 5'引物 (K3'arm1S)	ggatcaaacacgcatcctcatggac
Seq ID 18: 用于扩增增强子区的 κ 3'引物 (K3'arm1A)	ggtgattgggcatggtgagg
Seq ID 21: κ 筛选, 5'引物,5' (kappa5armS)	cgaaccctgtgtatatagtt
Seq ID 22: κ 筛选, 3'引物,5' (kappaNeoA)	gagatgaggaagaggagaaca
Seq ID 23: κ 筛选, 5'引物,3' (kappaNeoS)	gcattgtctgagtaggtgcatt
Seq ID 24: κ 筛选, 3'引物,5' (kappa5armProbe3')	cgcttctgcaggaacacgat
Seq ID 43: κ 筛选, 3'引物 (kappa3armA2)	GTCCTTGGTTTTTGCTGAGGGTT

细胞和猪组织样品的 DNA 印迹分析。将细胞或者组织样品在裂解缓冲液(10 mM Tris, pH 7.5, 10mM EDTA, 10mM NaCl, 0.5% (w/v) 十二烷基肌氨酸钠, 1 mg/ml 蛋白酶 K)中于 60℃ 过夜裂解并用乙醇沉淀 DNA。然后将 DNA 用 SacI 消化, 并在 1% 琼脂糖凝胶上分离。电泳后, 将 DNA 转移到尼龙膜上并用洋地黄毒苷标记的探针 (SEQ ID No 42) 探测。使用化学发光底物系统(Roche Molecular Biochemicals) 检测条带。

用于 κ DNA 印迹的探针:

Kappa5ArmProbe 5'/3'

```

gaagtgaagccagccagttcctcctgggcaggtggccaaaattacagttg
accctcctggtctggctgaaccttgcctcatatggtgacagccatctgg
ccagggccaggtctcctctgaagccttggaggagagggagagtggc
tggccgatcacagatcggaaggggctgactcctcaaccggggtgcaga
ctctgcagggtgggtctgggcccaacacacccaaagcacgccaggaagg
aaaggcagcttggatcactgccagagctaggagagggcaccgggaaat
gatctgtccaagaccgttcttcttaactccgagggggtcagatga
agtggttttgttcttggcctgaagcatcgtgttcctgcaagaagcgg (SEQ ID No 42)

```

实施例 3 猪 λ 基因座的表征

为了破坏或者使猪 λ 丧失功能，已经设计了寻靶策略，其允许除去或者破坏 λ 基因座的区域，包括 J 到 C 表达盒的多联体。可以产生含有猪基因组的部分的 BAC 克隆。从 3X 冗余的猪 BAC 文库分离猪 Ig λ 链基因座的部分。通常，通过断裂猪的总基因组 DNA，其然后用于得到代表完整动物的基因组的至少三倍的 BAC 文库，可以产生 BAC 文库。然后通过对如本文描述的猪 λ 链免疫球蛋白选择性的探针杂交，可以选择含有猪 λ 链免疫球蛋白的 BAC。

含有 λ J-C 侧翼区的 BAC 克隆（见图 3）可以独立地断裂并亚克隆到质粒载体中。通过 PCR 已经筛选了个体亚克隆中 J 到 C 内含子的部分的存在。我们已通过从一个 C 区到下一个 C 区扩增克隆了这些表达盒的几个。通过使用方向允许在任何一个 C 区内背驰延伸的引物（Seq ID 26 和 Seq ID 27）完成了该扩增。为了得到成功的扩增，延伸的产物与来自相邻 C 区（与相同的 C 区相对）的延伸产物会聚。该策略主要产生了从一个 C 延伸到相邻 C 的扩增引物。然而，一些扩增引物是越过相邻 C 到位置远离该相邻 C 的下一个 C 扩增的结果。这些多基因扩增引物含有部分 C、下一个 J-C 单位的 J 和 C 区、第三个 J-C 单位的 J 区，和第三个 J-C 单位的 C 区的部分。Seq ID 28 是一个这样的扩增引物并且代表必须除去或者破坏的序列。

已经克隆的其他猪 λ 序列包括：Seq ID No. 32，其包括猪 λ 轻链基因组序列的第一个 λ J/C 区的 5'侧翼序列；Seq ID No. 33，其包括猪 λ 轻链基因组序列的 J/C 簇区的 3'侧翼序列，从 λ J/C 下游约 200 个碱基对开始；Seq ID No. 34，其包括猪 λ 轻链基因组序列的 J/C 簇区的 3'侧翼序列，J/C 簇下游约 11.8 Kb，接近增强子；Seq ID No. 35，其包括 λ 下游约 12Kb，包括增强子区；Seq ID No 36，其包括 λ 下游约 17.6Kb；Seq ID No 37，其包括 λ 下游约 19.1Kb；Seq ID No 38，其包括 λ 下游约 21.3Kb；和 Seq ID No 39，其包括 λ 下游约 27 Kb。

Seq ID 26: λ C 到 C 扩增引物的 5'引物 (lam C5')	ccttcctcctgcacctgcaac
Seq ID 26: λ C 到 C 扩增引物的 3'引物 (lam C3')	tagacacaccagggtggccttg

实施例 4 产生 λ 基因的寻靶载体

在一个实施方案中，已经设计并构建了一种载体，其中一个寻靶臂与 J1 的上游区同源，而另一个臂与最后一个 C 下游的区域同源(见图 4)。一个寻靶载体设计成靶定 J1 的上游。该寻靶载体利用选择标记，可以选择所述选择标记的存在或不存在。可以使用本文描述的和本领域已知的阳性和阴性选择标记的任何组合。使用由单纯疱疹胸苷激酶 (TK) 的编码区和 Tn5 氨基糖苷磷酸转移酶 (Neo 抗性) 基因组成的融合基因。该融合基因侧翼是本文描述的或者本领域已知的任何位点特异性重组酶 (SSRRS) 的识别位点，如 lox 位点。在通过使用 G418 选择分离靶定的细胞后，以反式提供 Cre 以缺失标记基因 (见图 5)。通过加入本领域已知的可以被 TK 代谢成毒性产物的任何药物，如更昔洛韦来选择已缺失所述标记基因的细胞。然后将所得基因型用第二种载体靶定。第二种寻靶载体 (图 6) 设计成靶定最后一个 C 的下游并且使用阳性/阴性选择系统，其仅一侧是特异性重组位点 (lox)。该重组位点与第一个寻靶事件距离很远。在分离靶定的基因型后，Cre 再次以反式提供以介导从第一个寻靶事件中提供的重组位点到第二个寻靶事件中递送的重组位点的缺失。将除去整个 J 到 C 簇。通过施用更昔洛韦再次选择合适的基因型。

在另一实例中，用插入寻靶载体独立地破坏每个 C 区。将设计并装配插入寻靶载体以破坏单独的 C 区基因。在 J-C 簇中有至少 3 个 J 到 C 区。我们将通过设计靶定第一个和最后一个 C 区的载体开始该过程并且将在寻靶载体中包括位点特异性重组位点。一旦产生了两个插入，就将用位点特异性重组酶缺失间插区。

实施例 5: 重链单敲除猪与 κ 单敲除体猪的杂交

为了产生具有一个破坏的 Ig 重链基因座与一个破坏的 Ig 轻链 κ 等位基因两者的猪，将单敲除动物杂交。第一次妊娠产生了四个胎

儿,其中两个通过 PCR 和 DNA 印迹对重链和 κ 寻靶事件筛选都是阳性的(引物见实施例 1 和 2)。将胎儿成纤维细胞分离、增殖并冷冻。 κ 单敲除体与重链单敲除体交配导致的第二次妊娠产生了四头健康的小猪。

含有重链单敲除和 κ 链单敲除的胎儿成纤维细胞将用于进一步寻靶。此类细胞将用于通过本文描述的方法和组合物来靶定 λ 基因座。所得后代将对于重链、 κ 链和 λ 链将是杂合的敲除体。这些动物将与含有如本文描述的人 Ig 基因的动物进一步杂交,且然后与其他单 Ig 敲除动物杂交,以产生具有人 Ig 置换基因的猪 Ig 双敲除动物。

本发明已经参考其优选实施方案进行了描述。根据前面本发明的详细描述,本发明的变形和修饰将对本领域技术人员是显而易见的。

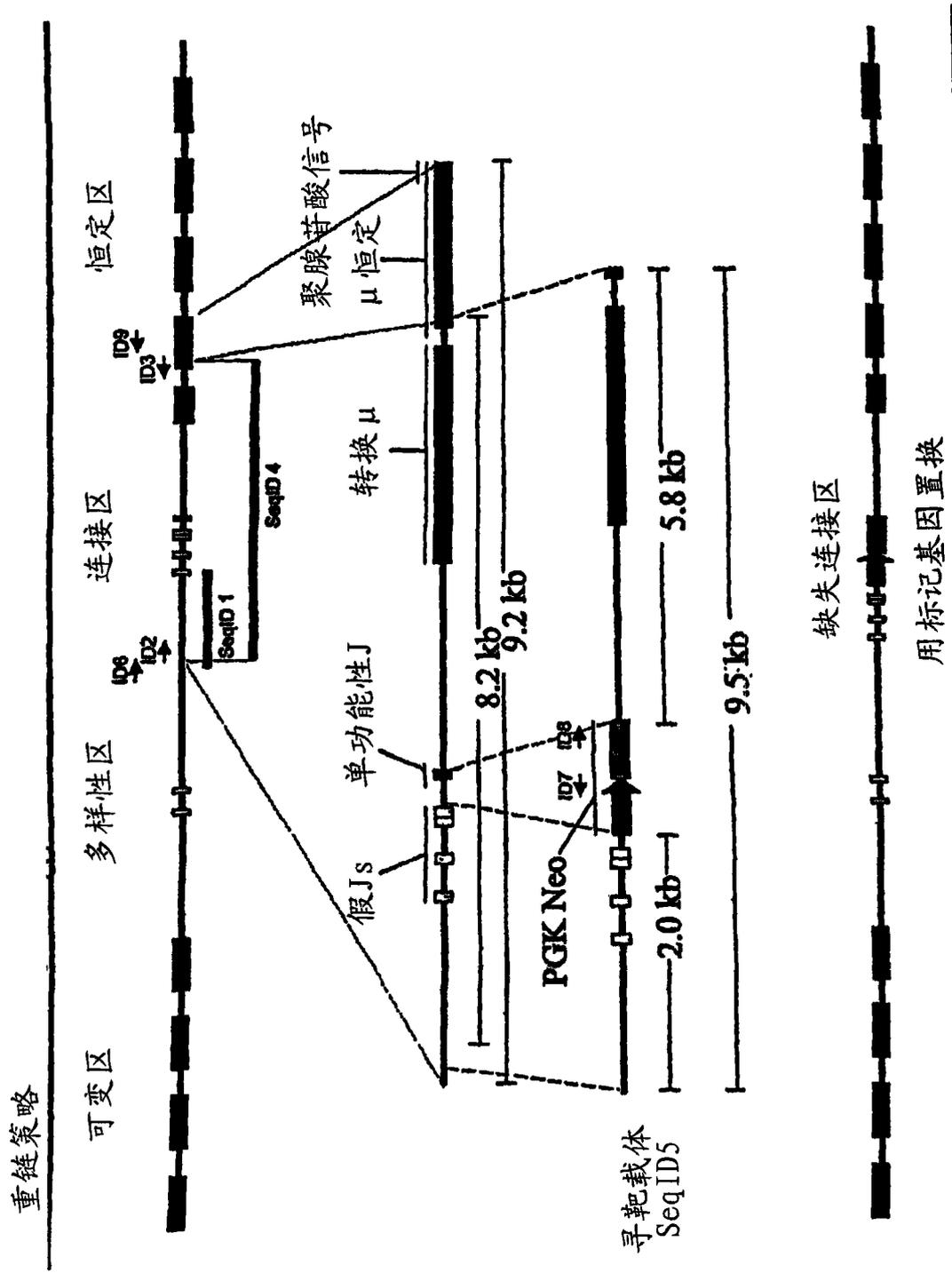


图 1

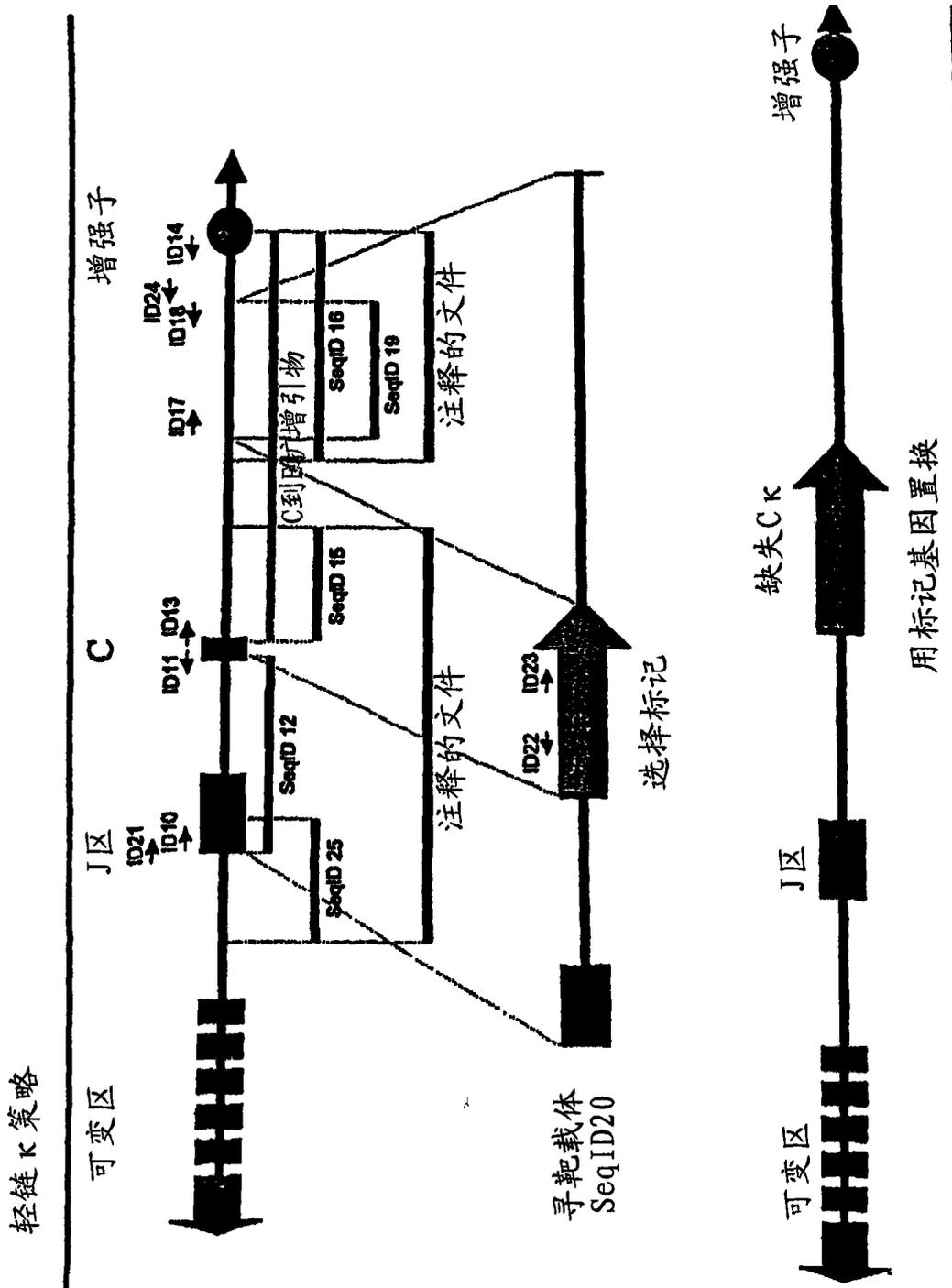


图 2

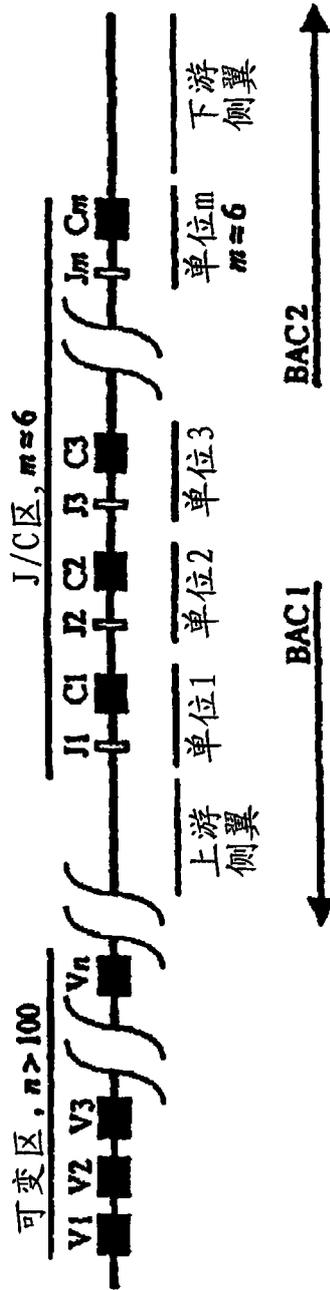


图 3

具有侧接JC簇的猪入重叠群的人序列AC002060
和NG_000002的图示

大约比例: 1个块为约1kb

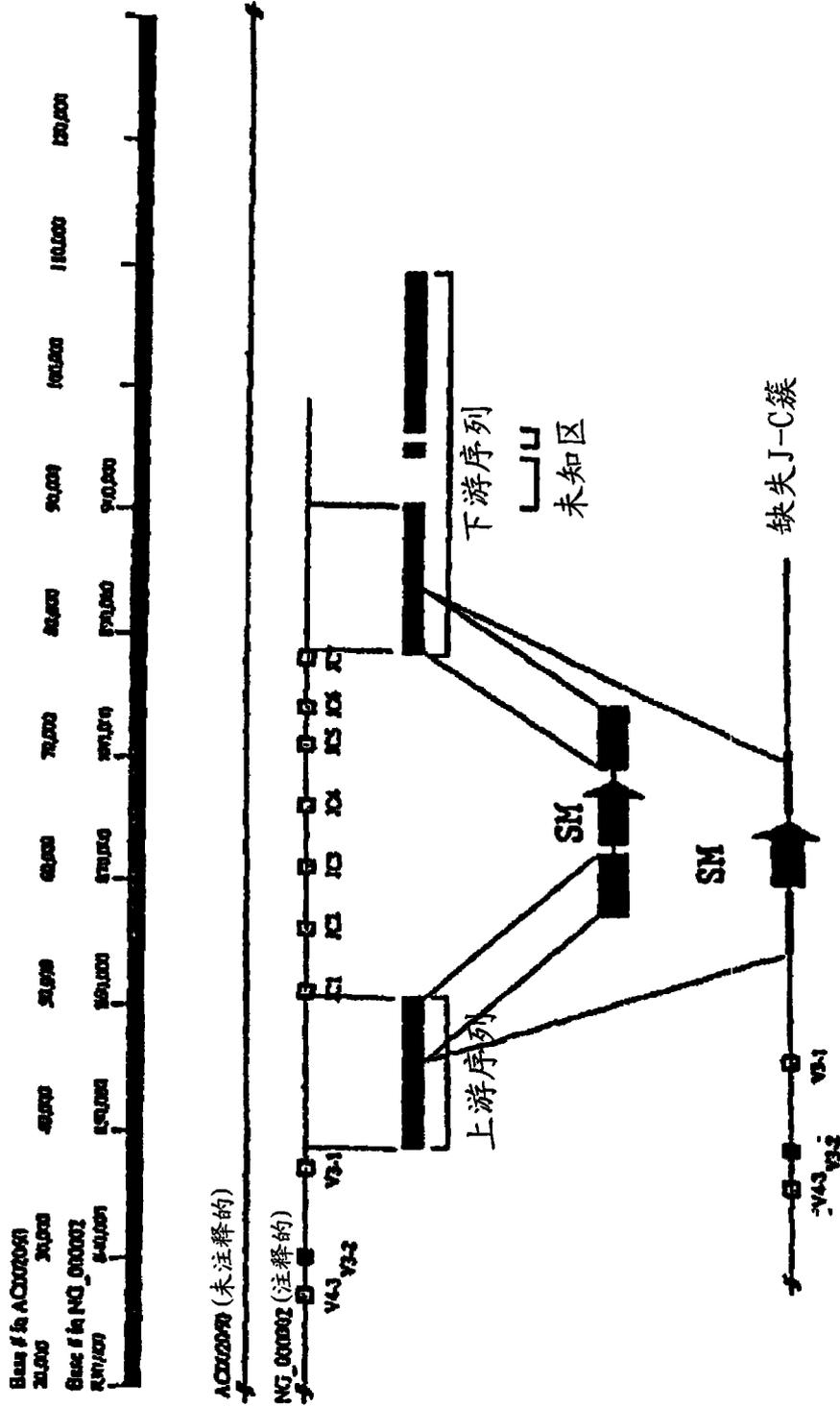


图 4

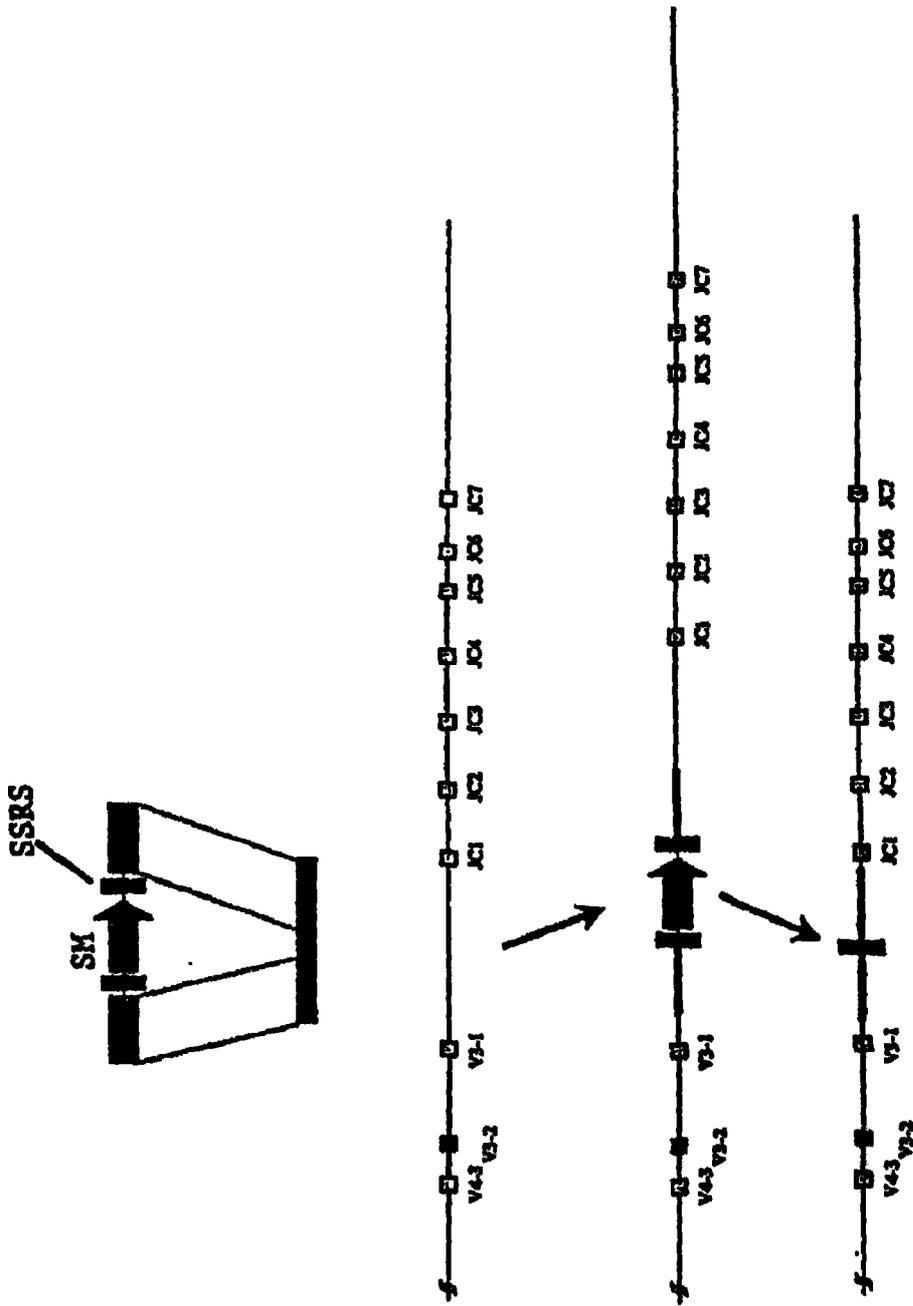


图 5

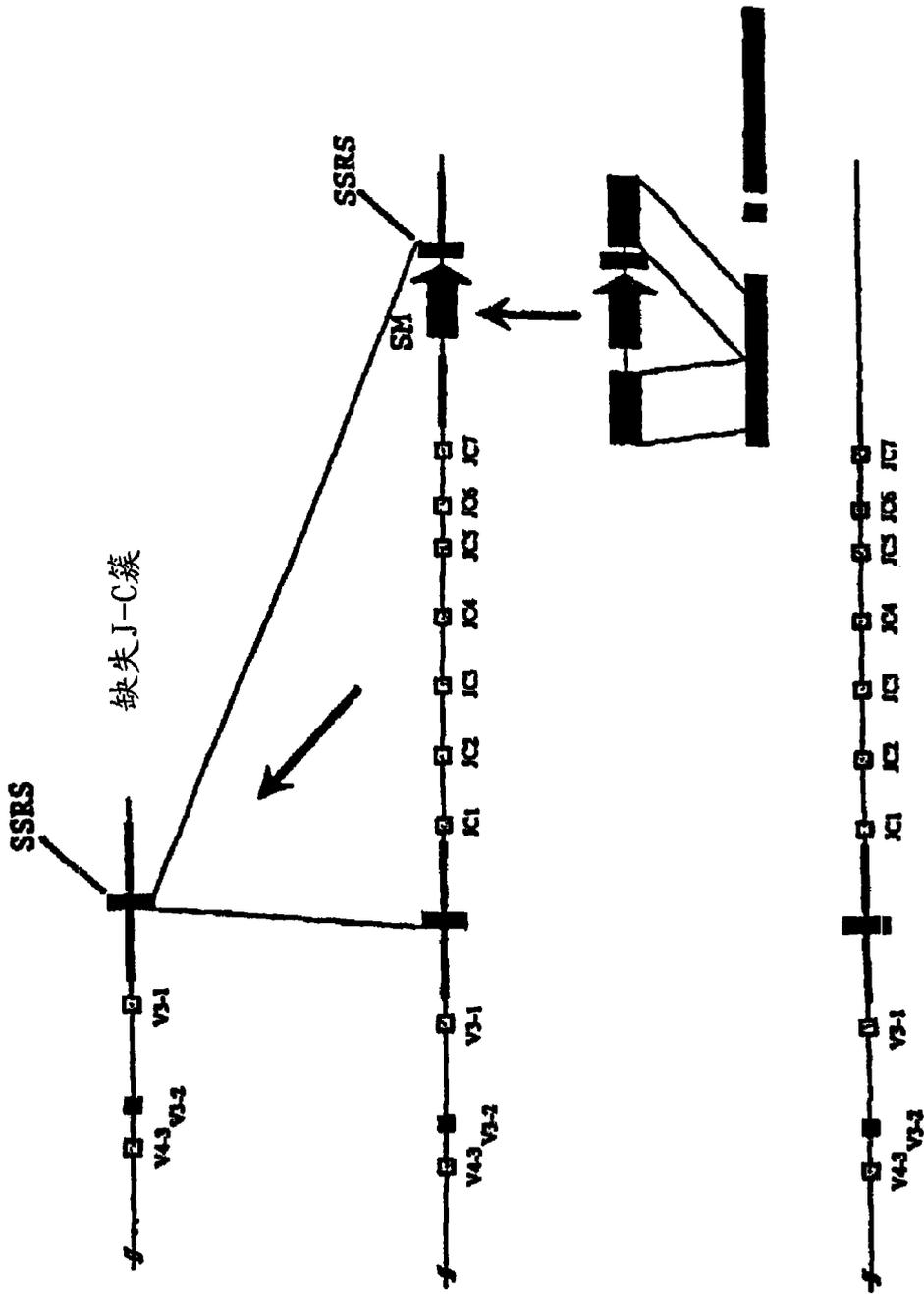
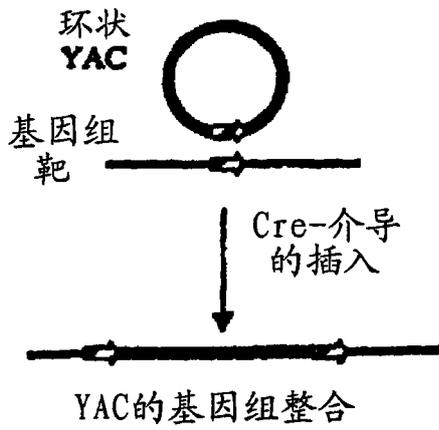


图 6

A.



B.

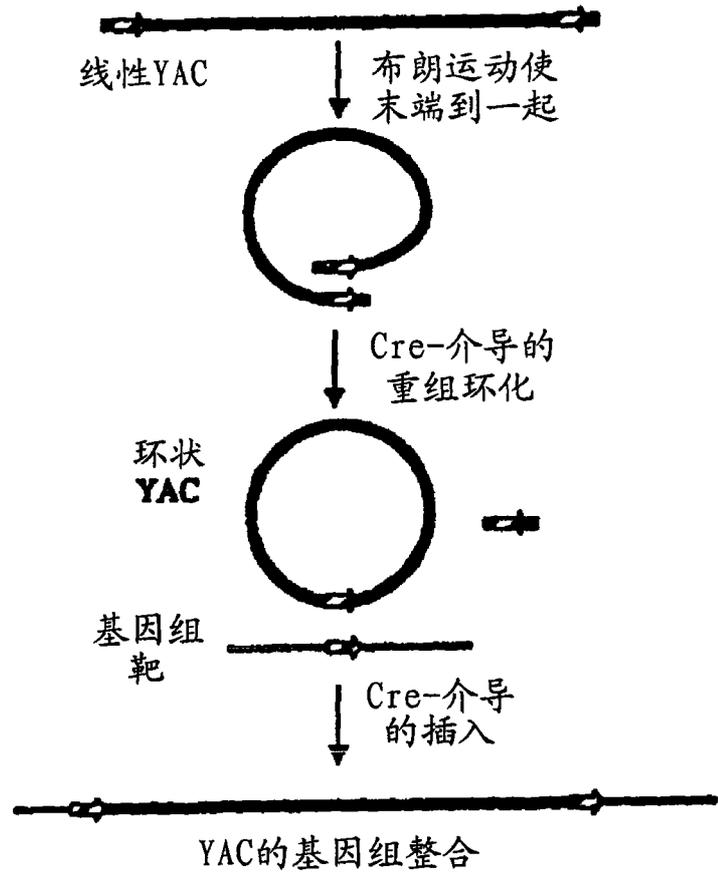


图 7