



(12)发明专利申请

(10)申请公布号 CN 107109434 A

(43)申请公布日 2017.08.29

(21)申请号 201580069889.1

(22)申请日 2015.10.21

(30)优先权数据

62/067,774 2014.10.23 US

(85)PCT国际申请进入国家阶段日

2017.06.21

(86)PCT国际申请的申请数据

PCT/US2015/056653 2015.10.21

(87)PCT国际申请的公布数据

W02016/064999 EN 2016.04.28

(71)申请人 瑞泽恩制药公司

地址 美国纽约

(72)发明人 Y·沈 D·巴拉科夫 G·陈

J·P·范德尔

(74)专利代理机构 中国国际贸易促进委员会专利商标事务所 11038

代理人 罗菊华

(51)Int.Cl.

C12N 15/85(2006.01)

C12N 15/90(2006.01)

C12N 5/10(2006.01)

权利要求书4页 说明书20页

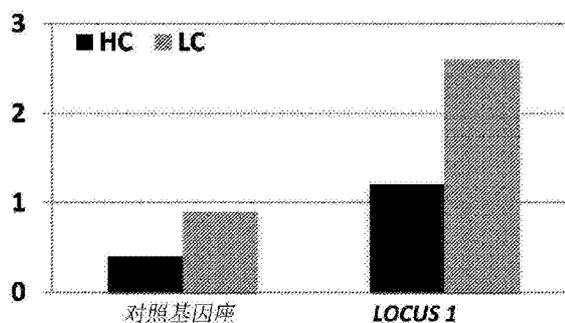
序列表11页 附图5页

(54)发明名称

新颖CHO整合位点和其用途

(57)摘要

提供用于真核表达系统的表达增强核苷酸序列,使得重组蛋白在真核细胞中的表达增强并且稳定。提供用于提供增强的表达的基因组整合位点和其使用方法以便在真核细胞中表达所关注基因。提供用于基因在真核细胞中的增强和稳定表达的染色体基因座、序列和载体。



1. 一种包含整合在基因座内的特定位置处的外源性核酸序列的细胞,其中所述基因座包含与SEQ ID NO:1或SEQ ID NO:4至少90%一致的核苷酸序列。

2. 一种包含整合到第二核酸序列内的特定位置中的第一核酸序列的聚核苷酸,其中所述第二核酸序列包含SEQ ID NO:1或SEQ ID NO:4的核苷酸序列。

3. 根据权利要求2所述的聚核苷酸,其中所述特定位置位于SEQ ID NO:1内的位置或邻近于SEQ ID NO:1内的位置,其中所述位置选自跨越SEQ ID NO:1的编号10-4,000、100-3,900、200-3,800、300-3,700、400-3,600、500-3,500、600-3,400、700-3,300、800-3,200、900-3,100、1,000-3,000、1,100-2,900、1,200-2,800、1,300-2,700、1,200-2,600、1,300-2,500、1,400-2,400、1,500-2,300、1,600-2,200、1,700-2,100、1,800-2,050、1,900-2,040、2,000-2,020、2,002-2,021、2,010-2,015、2,001-2,002、2,009-2,010和2,021-2,022的位置的核苷酸。

4. 根据权利要求3所述的聚核苷酸,其中SEQ ID NO:1内或邻近于SEQ ID NO:1内的位置的所述特定位置选自以下各项组成的群组:跨越SEQ ID NO:1的编号10-500、500-1,000、500-2,100、1,000-1,500、1,000-2,100、1,500-2,000、1,500-2,500、2,000-2,500、2,500-3,000、2,500-3,500、3,000-3,500、3,000-4,000和3,500-4,000的位置的核苷酸。

5. 一种包含经修饰的基因组的经修饰的细胞,其中所述基因组是通过在所述基因组的基因座内插入外源性核酸序列而经修饰的,其中所述基因座包含与SEQ ID NO:1或SEQ ID NO:4至少90%一致的表达增强核苷酸序列。

6. 根据权利要求5所述的经修饰的细胞,其中所述细胞是CHO细胞。

7. 根据权利要求5或6所述的经修饰的细胞,其中所述外源性核酸序列包含一个或多个重组识别序列。

8. 根据权利要求7所述的经修饰的细胞,其中所述外源性核酸序列包含至少两个重组识别序列和置于所述两个重组识别序列之间的可选标记。

9. 根据权利要求7所述的经修饰的细胞,其中所述一个或多个重组识别序列选自以下各项组成的群组:LoxP位点、Lox511位点、Lox2272位点、Lox2372位点、Lox5171位点、Loxm2位点、Lox71位点、Lox66位点、LoxFas位点和frt位点。

10. 根据权利要求5或6所述的经修饰的细胞,其中所述外源性核酸序列包含可操作地连接到所述表达增强核苷酸序列的至少一个外源性所关注基因(GOI)。

11. 根据权利要求10所述的经修饰的细胞,其中所述至少一个外源性GOI是人类基因,并且所述人类基因可操作地连接到外源性启动子。

12. 根据权利要求11所述的经修饰的细胞,其另外包含在所述第一GOI的3'的第二GOI,并且所述第二GOI可操作地连接到外源性启动子。

13. 根据权利要求12所述的经修饰的细胞,其中所述第一GOI编码抗体轻链并且所述第二GOI编码抗体重链。

14. 根据权利要求13所述的经修饰的细胞,其另外包含在所述编码抗体轻链的基因的5'的第一重组酶位点和紧邻所述编码抗体重链的基因并在所述编码抗体重链的基因的3'的第二重组酶位点。

15. 根据权利要求14所述的经修饰的细胞,其中所述第一和第二重组酶识别位点是不同的,并且所述第一和第二重组酶识别位点选自以下各项组成的群组:LoxP位点、Lox511位点、Lox2272位点、Lox2372位点、Lox5171位点、Loxm2位点、Lox71位点、Lox66位点、LoxFas

位点和frt位点。

16. 一种修饰CHO细胞基因组的方法,其包含将包含外源性序列的载体引入到所述CHO细胞中的步骤,其中所述外源性序列整合到所述基因组的包含与SEQ ID NO:1或SEQ ID NO:4至少90%一致的核苷酸序列的基因座中。

17. 一种修饰CHO细胞基因组以表达所关注蛋白的方法,其包含将载体引入到CHO细胞中以便将包含用于表达所述所关注蛋白的序列的外源性核酸引入到所述CHO细胞的基因组中,其中所述载体包含载体,所述载体包含:

- a. 与SEQ ID NO:1或SEQ ID NO:4的核苷酸序列中所存在的序列同源的5'同源臂,
- b. 编码所述所关注蛋白的核酸,和
- c. 与SEQ ID NO:1或SEQ ID NO:4的核苷酸序列中所存在的序列同源的3'同源臂。

18. 根据权利要求17所述的方法,其中所述载体包含至少一个额外载体或mRNA分子。

19. 根据权利要求18所述的方法,其中所述额外载体选自由以下各项组成的群组:腺病毒、慢病毒、逆转录病毒、腺相关病毒、整合性噬菌体载体、非病毒载体、转座子和/或转座酶、整合酶底物和质粒。

20. 根据权利要求17所述的方法,其中所述5'同源臂包含与SEQ ID NO:1或SEQ ID NO:4的核苷酸序列中所存在的序列同源并且与编码所述所关注蛋白的所述核酸相连的序列。

21. 根据权利要求17所述的方法,其中所述3'同源臂包含与SEQ ID NO:1或SEQ ID NO:4的核苷酸序列中所存在的序列同源并且与编码所述所关注蛋白的所述核酸相连的序列。

22. 一种用于修饰CHO细胞基因组以表达所关注蛋白的载体,其包含载体,所述载体包含:

- a. 与SEQ ID NO:1或SEQ ID NO:4的核苷酸序列中所存在的序列同源的5'同源臂,
- b. 编码所述所关注蛋白的核酸,和
- c. 与SEQ ID NO:1或SEQ ID NO:4的核苷酸序列中所存在的序列同源的3'同源臂。

23. 根据权利要求22所述的载体,其中所述载体包含至少一个额外载体或mRNA分子。

24. 根据权利要求23所述的载体,其中所述额外载体选自由以下各项组成的群组:腺病毒、慢病毒、逆转录病毒、腺相关病毒、整合性噬菌体载体、非病毒载体、转座子和/或转座酶、整合酶底物和质粒。

25. 根据权利要求22所述的载体,其中所述5'同源臂包含与SEQ ID NO:1或SEQ ID NO:4的核苷酸序列中所存在的序列同源并且与编码所述所关注蛋白的所述核酸相连的序列。

26. 根据权利要求22所述的载体,其中所述3'同源臂包含与SEQ ID NO:1或SEQ ID NO:4的核苷酸序列中所存在的序列同源并且与编码所述所关注蛋白的所述核酸相连的序列。

27. 一种用于制备所关注蛋白的方法,其包含以下步骤:

- a. 将所关注基因(GOI)引入到细胞中,其中所述GOI整合到包含与SEQ ID NO:1或SEQ ID NO:4的表达增强至少90%一致的核苷酸序列的特定基因座中;
- b. 在允许所述GOI表达的条件下培养(a)的细胞;和
- c. 回收所述所关注蛋白。

28. 根据权利要求27所述的方法,其中所述GOI可操作地连接到所述表达增强序列,并且至少一个重组酶识别位点紧邻所述GOI。

29. 根据权利要求28所述的方法,其中所述GOI编码选自免疫球蛋白或其抗原结合片段

和受体或其配体结合片段的蛋白。

30. 根据权利要求29所述的方法,其中所述免疫球蛋白选自抗体轻链或其抗原结合片段、抗体重链或其抗原结合片段、Fc融合蛋白或Fc受体融合蛋白。

31. 根据权利要求30所述的方法,其中所述至少一个重组酶识别位点选自由以下各项组成的群组:LoxP位点、Lox511位点、Lox2272位点、Lox2372位点、Lox5171位点、Loxm2位点、Lox71位点、Lox66位点、LoxFas位点和frt位点。

32. 根据权利要求31所述的方法,其中所述GOI紧邻所述重组酶识别位点并且在所述重组酶识别位点的5',并且另外包含紧邻所述GOI并且在所述GOI的3'的第二重组酶识别位点。

33. 根据权利要求32所述的方法,其另外包含紧邻所述第二重组酶识别位点并且在所述第二重组酶识别位点的3'的第二GOI。

34. 根据权利要求33所述的方法,其另外包含紧邻所述第二GOI并且在所述第二GOI的3'的第三重组酶识别位点。

35. 根据权利要求34所述的方法,其另外包含在所述第二重组酶识别位点与所述第二GOI之间的至少一个标记基因。

36. 根据权利要求35所述的方法,其中所述至少一个标记基因选自由药物抗性基因和表达报告基因组成的群组。

37. 根据权利要求36所述的方法,其另外包含可操作地连接到所述第一GOI的启动子和可操作地连接到所述第二GOI的启动子。

38. 根据权利要求37所述的方法,其中所述第二和所述第三重组酶识别位点在与所述第一重组酶识别位点相反的方向。

39. 根据权利要求38所述的方法,其中所述第一、第二和第三重组酶识别位点是不同的。

40. 一种用于制备所关注蛋白的方法,其包含:

(a) 将核酸构建体引入到CHO细胞中,所述核酸构建体包含

具有与SEQ ID NO:1或SEQ ID NO:4至少90%一致的核苷酸序列的表达增强序列和编码所关注蛋白的外源性GOI,其中所述GOI可操作地连接到所述表达增强序列;

(b) 在允许所述GOI表达的条件下培养(a)的CHO细胞;和,

(c) 回收所述所关注蛋白。

41. 根据权利要求40所述的方法,其中所述GOI可操作地连接到启动子。

42. 根据权利要求41所述的方法,其中所述GOI编码免疫球蛋白或其抗原结合片段。

43. 根据权利要求42所述的方法,其中所述GOI编码抗体、受体或其配体结合片段、Fc受体融合蛋白或Fc融合蛋白。

44. 一种修饰CHO细胞基因组以整合识别序列的方法,其包含将载体引入到CHO细胞中以便将包含识别序列的外源性核酸引入到所述CHO细胞的基因组中,其中所述载体包含载体,所述载体包含:

a. 与SEQ ID NO:1或SEQ ID NO:4的核苷酸序列中所存在的序列同源的5'同源臂,

b. 包含识别序列的外源性核酸,和

c. 与SEQ ID NO:1或SEQ ID NO:4的核苷酸序列中所存在的序列同源的3'同源臂,

其中所述识别序列整合在所述基因组的具有与SEQ ID NO:1或SEQ ID NO:4至少90%一致的核苷酸序列的基因座内。

45. 根据权利要求44所述的方法,其中所述外源性核酸序列包含至少两个重组识别序列和置于所述两个重组识别序列之间的可选标记。

46. 根据权利要求44或45所述的方法,其中所述载体包含至少一个额外载体或mRNA分子。

47. 根据权利要求46所述的方法,其中所述额外载体包含编码用于整合所述识别序列的位点特异性核酸酶的核酸。

48. 根据权利要求47所述的方法,其中所述位点特异性核酸酶选自由以下各项组成的群组:锌指核酸酶(ZFN)、ZFN二聚体、转录活化因子样效应子核酸酶(TALEN)、TAL效应子结构域融合蛋白或RNA指导的DNA核酸内切酶。

49. 一种用于修饰CHO细胞基因组以整合识别序列的载体,其包含载体,所述载体包含:

- a. 与SEQ ID NO:1或SEQ ID NO:4的核苷酸序列中所存在的序列同源的5'同源臂,
- b. 包含所述识别序列的核酸,和
- c. 与SEQ ID NO:1或SEQ ID NO:4的核苷酸序列中所存在的序列同源的3'同源臂。

50. 根据权利要求49所述的载体,其中所述核酸包含至少两个重组识别序列和置于所述两个重组识别序列之间的可选标记。

51. 根据权利要求49或50所述的载体,其中所述载体包含至少一个额外载体或mRNA分子。

52. 根据权利要求51所述的载体,其中所述额外载体包含编码用于整合所述识别序列的位点特异性核酸酶的核酸。

53. 根据权利要求52所述的载体,其中所述位点特异性核酸酶选自由以下各项组成的群组:锌指核酸酶(ZFN)、ZFN二聚体、转录活化因子样效应子核酸酶(TALEN)、TAL效应子结构域融合蛋白或RNA指导的DNA核酸内切酶。

54. 根据权利要求49所述的载体,其中所述5'同源臂包含与SEQ ID NO:1或SEQ ID NO:4的核苷酸序列中所存在的序列同源并且与包含所述识别序列的所述核酸相连的序列。

55. 根据权利要求49所述的载体,其中所述3'同源臂包含与SEQ ID NO:1或SEQ ID NO:4的核苷酸序列中所存在的序列同源并且与包含所述识别序列的所述核酸相连的序列。

新颖CHO整合位点和其用途

[0001] 相关专利申请的交叉引用

[0002] 本申请要求2014年10月23日提交的美国临时申请第62/067,774号的优先权,所述临时申请的全部内容以引用的方式并入本文中。

[0003] 序列表以引用方式并入

[0004] 创建于2015年10月20日并且经由EFS-Web提交至美国专利商标局的名为32353_T0045US01_SequenceListing.txt的28KB ASCII文本文件中的序列表以引用的方式并入本文中。

技术领域

[0005] 本发明提供重组蛋白在真核细胞中的稳定整合和/或表达。具体来说,本发明包括通过采用表达增强核苷酸序列改善蛋白在真核细胞、尤其中国仓鼠(灰仓鼠(*Cricetulus griseus*))细胞系中的表达的方法和组合物。本发明包括有助于重组介导的盒交换(RMCE)的聚核苷酸和经修饰的细胞。本发明的方法将外源性核酸整合在中国仓鼠细胞基因组的特定染色体基因座以有助于经修饰的细胞增强并且稳定表达重组蛋白。

背景技术

[0006] 细胞表达系统旨在提供用于制备给定蛋白(不论用于研究或治疗用途)的可靠且高效的来源。由于例如哺乳动物表达系统对重组蛋白进行适当的翻译后修饰的能力,哺乳动物细胞中的重组蛋白表达是用于制备治疗性蛋白的优选方法。

[0007] 数个细胞系统可用于蛋白的表达,每个系统含有顺式和在一些情况下反式调控元件的各种组合以便在短的培育时间内达到高重组蛋白水平。尽管可获得许多系统,但是仍存在用于表达重组蛋白的整合基因的高效基因转移和稳定性的难题。多个局部遗传因子将不仅决定何时表达所关注的目标基因,而且决定细胞是否可以功能上驱使基因朝着高产性输出转录,或甚至表达是否将长期持续。染色体整合位点,例如中国仓鼠卵巢细胞(CHO)整合位点和特定基因内或相邻的基因座控制区已在所属领域中加以表征(WO2012/138887A1; Li, Q. 等人, 2002Blood. 100:3077-3086)。同样地,通常在内源性蛋白编码区中鉴别靶向调控区。然而,为了长期表达目标转基因,关键考虑因素是对细胞基因的破坏程度最小以避免细胞系表型的变化。

[0008] 工程改造稳定细胞系以容纳额外用于表达的基因,如多特异性抗体中的额外抗体链,是特别具有挑战性的。可能存在整合基因表达水平的大幅变化。整合额外基因可能由于局部遗传环境(即,位置作用)而引起表达的较大变化和不稳定性。因此,所属领域中需要经改善的哺乳动物表达系统。

发明内容

[0009] 在一个方面,本发明提供一种包含整合在基因座内的特定位点处的外源性核酸序列的细胞,其中所述基因座包含与SEQ ID NO:1或SEQ ID NO:4至少90%一致的核苷酸序

列。在一些实施例中,所述基因座包含与SEQ ID NO:1至少90%一致的核苷酸序列。在一些实施例中,所述基因座包含与SEQ ID NO:4至少90%一致的核苷酸序列。

[0010] 在另一个方面,本发明提供一种聚核苷酸,其包含整合到第二核酸序列内的特定位点(例如本发明的基因座)中的第一核酸序列。在一个实施例中,所述第二核酸序列包含SEQ ID NO:1的核苷酸序列。在另一个实施例中,所述第二核酸序列包含SEQ ID NO:4的核苷酸序列。

[0011] 在一个实施例中,所述第二核酸序列是选自与SEQ ID NO:1具有至少90%核酸一致性的核苷酸序列的表达增强序列,或其表达增强片段。在一个实施例中,所述第二核酸序列是选自与SEQ ID NO:4具有至少90%核酸一致性的核苷酸序列的表达增强序列,或其表达增强片段。在另一个实施例中,所述表达增强序列能够增强由外源性核酸序列编码的蛋白的表达。在另一个实施例中,所述表达增强序列能够使由外源性核酸序列编码的蛋白的表达与通常通过随机整合到基因组中所观察到的表达相比表达增强至少约1.5倍到至少约3倍。

[0012] 在另一个实施例中,外源性核酸序列被整合到SEQ ID NO:1或SEQ ID NO:4内的任何位置处的特定位点中。

[0013] 在一些实施例中,处于SEQ ID NO:1内的位置或邻近于SEQ ID NO:1内的位置的特定位点选自由以下各项组成的群组:跨越SEQ ID NO:1的编号10-4,000、100-3,900、200-3,800、300-3,700、400-3,600、500-3,500、600-3,400、700-3,300、800-3,200、900-3,100、1,000-3,000、1,100-2,900、1,200-2,800、1,300-2,700、1,200-2,600、1,300-2,500、1,400-2,400、1,500-2,300、1,600-2,200、1,700-2,100、1,800-2,050、1,850-2,050、1,900-2,040、1,950-2,025、1,990-2,021、2,002-2,021和2,010-2,015的位置的核苷酸。在某些实施例中,处于SEQ ID NO:1内的位置或邻近于SEQ ID NO:1内的位置的特定位点选自由以下各项组成的群组:跨越SEQ ID NO:1的编号1990-1991、1991-1992、1992-1993、1993-1994、1995-1996、1996-1997、1997-1998、1999-2000、2001-2002、2002-2003、2003-2004、2004-2005、2005-2006、2006-2007、2007-2008、2008-2009、2009-2010、2010-2011、2011-2012、2012-2013、2013-2014、2014-2015、2015-2016、2016-2017、2017-2018、2018-2019、2019-2020和2020-2021的位置的核苷酸。

[0014] 在另一个实施例中,处于SEQ ID NO:1内的位置或邻近于SEQ ID NO:1内的位置的特定位点选自由以下各项组成的群组:跨越SEQ ID NO:1的编号10-500、500-1,000、500-2,100、1,000-1,500、1,000-2,100、1,500-2,000、1,500-2,500、2,000-2,500、2,500-3,000、2,500-3,500、3,000-3,500、3,000-4,000和3,500-4,000的位置的核苷酸。在某些实施例中,外源性核酸序列被整合在上文所述特定位点中的任何一个或多个处、内部或附近。

[0015] 在另一个实施例中,外源性核酸序列包含位于如上所述的表达增强序列内的识别位点,其条件是所述表达增强序列包含与SEQ ID NO:1或SEQ ID NO:4的表达增强序列至少约90%一致、至少约91%一致、至少约92%一致、至少约93%一致、至少约94%一致、至少约95%一致、至少约96%一致、至少约97%一致、至少约98%一致或至少约99%一致的序列、其表达增强片段。

[0016] 在一个实施例中,外源性核酸序列包含重组酶识别位点。在一些实施例中,外源性核酸序列另外包含至少一个重组酶识别位点,所述重组酶识别位点包含独立地选自以下各

项的序列:LoxP位点、Lox511位点、Lox2272位点、Lox2372、Lox5171、Loxm2、Lox71、Lox66、LoxFas和frt位点。在一个实施例中,所述重组酶识别位点被整合在表达增强序列内。在另一个实施例中,所述重组酶识别位点在5'方向紧邻基因盒5'端的末端核苷酸,或在3'方向紧邻基因盒3'端的末端核苷酸。在一些实施例中,所述至少一个重组酶识别位点和基因盒被整合在表达增强序列内。

[0017] 在一个实施例中,至少两个重组酶识别位点存在于表达增强序列内。在另一个实施例中,相反方向的两个重组酶识别位点被整合在表达增强序列内。在另一个实施例中,三个重组酶识别位点被整合在表达增强序列内。

[0018] 在一个方面,提供经分离的中国仓鼠卵巢(CHO)细胞,其包含经工程改造的SEQ ID NO:1的表达增强序列或其表达增强片段。在一个实施例中,包含SEQ ID NO:1或SEQ ID NO:4的核苷酸序列的表达增强序列或其稳定变异体经工程改造以整合如上所述的外源性核酸序列。在其它实施例中,本发明提供一种经分离的CHO细胞,其包含插入到包含SEQ ID NO:1或SEQ ID NO:4的表达增强序列或其稳定变异体的基因座中的外源性核酸序列。

[0019] 在一个实施例中,所述CHO细胞另外包含在表达增强序列内的至少一个重组酶识别序列。在另一个实施例中,所述至少一个重组酶识别序列独立地选自LoxP位点、Lox511位点、Lox2272位点、Lox2372、Lox5171、Loxm2、Lox71、Lox66LoxFas和frt位点。在另一个实施例中,所述重组酶识别位点在5'方向紧邻基因盒5'端的末端核苷酸,或在3'方向紧邻基因盒3'端的末端核苷酸。在一些实施例中,所述至少一个重组酶识别位点和基因盒被整合在本文所述的CHO细胞基因组的表达增强序列内。

[0020] 在另一个实施例中,所述至少一个重组识别位点如上所述定位,需要说明的是基因盒包含与SEQ ID NO:1(SEQ ID NO:2)的核苷酸1001至2001包含至少90%一致性、至少约91%一致性、至少约92%一致性、至少约93%一致性、至少约94%一致性、至少约95%一致性、至少约96%一致性、至少约97%一致性、至少约98%一致性或至少约99%一致性的表达增强序列或其表达增强片段。在另一个实施例中,所述至少一个重组识别位点如上所述定位,需要说明的是基因盒包含与SEQ ID NO:1(SEQ ID NO:3)的核苷酸2022至3022包含至少90%一致性、至少约91%一致性、至少约92%一致性、至少约93%一致性、至少约94%一致性、至少约95%一致性、至少约96%一致性、至少约97%一致性、至少约98%一致性或至少约99%一致性的表达增强序列或其表达增强片段。

[0021] 在又一个实施例中,所述至少一个重组酶识别位点被插入在SEQ ID NO:1的核苷酸1990-1991、1991-1992、1992-1993、1993-1994、1995-1996、1996-1997、1997-1998、1999-2000、2001-2002、2002-2003、2003-2004、2004-2005、2005-2006、2006-2007、2007-2008、2008-2009、2009-2010、2010-2011、2011-2012、2012-2013、2013-2014、2014-2015、2015-2016、2016-2017、2017-2018、2018-2019、2019-2020、2020-2021或2021-2022处或所述核苷酸内的CHO细胞基因组中。

[0022] 在另一个实施例中,外源性核酸被插入在SEQ ID NO:1的核苷酸1990-1991、1991-1992、1992-1993、1993-1994、1995-1996、1996-1997、1997-1998、1999-2000、2001-2002、2002-2003、2003-2004、2004-2005、2005-2006、2006-2007、2007-2008、2008-2009、2009-2010、2010-2011、2011-2012、2012-2013、2013-2014、2014-2015、2015-2016、2016-2017、2017-2018、2018-2019、2019-2020、2020-2021或2021-2022处或所述核苷酸内的CHO基因组

中。

[0023] 在另一个实施例中,外源性核酸被插入在SEQ ID NO:1的核苷酸2001-2022处或所述核苷酸内的CHO基因组中。在一些实施例中,外源性核酸被插入在SEQ ID NO:1的核苷酸2001-2002或核苷酸2021-2022处或所述核苷酸内,并且SEQ ID NO:1的核苷酸2002-2021由于插入而缺失。同样,外源性核酸被插入在SEQ ID NO:4的核苷酸9302-9321处或所述核苷酸内的CHO基因组中。在一些实施例中,外源性核酸被插入在SEQ ID NO:4的核苷酸9301-9302或核苷酸9321-9322处或所述核苷酸内,并且SEQ ID NO:4的核苷酸9302-9321由于插入而缺失。

[0024] 在一些实施例中,整合在基因座(如SEQ ID NO:1或SEQ ID NO:4的核苷酸序列)内的特定位置处的外源性核酸序列包含所关注基因(GOI)(例如,编码所关注蛋白或“POI”的核苷酸序列)。在某些实施例中,外源性核酸序列包含一个或多个所关注基因。在一些实施例中,一个或多个所关注基因选自第一GOI、第二GOI和第三GOI组成的群组。

[0025] 在一些实施例中,整合在基因座(如SEQ ID NO:1或SEQ ID NO:4的核苷酸序列)内的特定位置处的外源性核酸序列包含GOI和至少一个重组酶识别位点。在一个实施例中,第一GOI如上所述被插入在SEQ ID NO:1或SEQ ID NO:4的表达增强序列或与SEQ ID NO:1或SEQ ID NO:4具有至少90%核苷酸一致性的表达增强序列或其表达增强片段内,并且所述第一GOI任选地被可操作地连接到启动子,其中所述启动子连接的GOI(或所述GOI)的5'侧翼是第一重组酶识别位点并且3'侧翼是第二重组酶识别位点。在另一个实施例中,第二GOI被插入在第二重组酶识别位点的3',并且所述第二GOI的3'侧翼是第三重组酶识别位点。

[0026] 在又一个实施例中,GOI被可操作地连接到能够驱动GOI表达的启动子,其中所述启动子包含可以由活化因子或抑制因子调控的真核启动子。在其它实施例中,真核启动子被可操作地连接到原核操纵子,并且真核细胞任选地另外包含原核阻遏蛋白。

[0027] 在另一个实施例中,一个或多个可选标记被包括在第一与第二和/或第二与第三重组酶识别位点之间。在一些实施例中,第一和/或第二所关注基因和/或一个或多个可选标记被可操作地连接到启动子,其中所述启动子可以相同或不同。在另一个实施例中,启动子包含真核启动子(如CMV启动子或SV40晚期启动子),其任选地受原核操纵子(如tet操纵子)控制。在其它实施例中,细胞另外包含编码原核阻遏子(如tet阻遏子)的基因。

[0028] 在另一个实施例中,细胞另外包含能够表达重组酶的基因。在一些实施例中,所述重组酶是Cre重组酶。

[0029] 在一个方面,提供一种CHO宿主细胞,其包含选自SEQ ID NO:1或SEQ ID NO:4的表达增强序列、或与SEQ ID NO:1或SEQ ID NO:4具有至少90%核苷酸一致性的表达增强序列、或其表达增强片段,包含第一重组酶识别位点,后面是第一真核启动子、第一可选标记基因、第二真核启动子、第二可选标记基因和第二重组酶识别位点。在更多实施例中,CHO宿主细胞另外提供第三真核启动子、第三标记基因和第三重组酶识别位点。在一个实施例中,表达增强序列如上所述在SEQ ID NO:1或SEQ ID NO:4内。

[0030] 在一个实施例中,第一、第二和第三重组酶识别位点彼此不同。在一些实施例中,重组酶识别位点选自LoxP位点、Lox511位点、Lox2272位点、Lox2372、Lox5171、Loxm2、Lox71、Lox66、LoxFas和frt位点。

[0031] 在一个实施例中,第一可选标记基因是药物抗性基因。在另一个实施例中,药物抗

性基因是新霉素抗性基因或潮霉素抗性基因。在另一个实施例中,第二和第三可选标记基因编码两种不同的荧光蛋白。在一个实施例中,所述两种不同的荧光蛋白选自由以下各项组成的群组:Discosoma珊瑚(DsRed)、绿色荧光蛋白(GFP)、增强型绿色荧光蛋白(eGFP)、蓝绿色荧光蛋白(CFP)、增强型蓝绿色荧光蛋白(eCFP)、黄色荧光蛋白(YFP)、增强型黄色荧光蛋白(eYFP)和远红外荧光蛋白(例如ImKate、mKate2、mPlum、mRaspberry或E2-crimson)。

[0032] 在一个实施例中,第一、第二和第三启动子是相同的。在另一个实施例中,第一、第二和第三启动子彼此不同。在另一个实施例中,第一启动子不同于第二和第三启动子,并且第二和第三启动子是相同的。在更多实施例中,第一启动子是SV40晚期启动子,并且第二和第三启动子各自是人类CMV启动子。在其它实施例中,第一和第二启动子被可操作地连接到原核操纵子。

[0033] 在一个实施例中,宿主细胞系具有外源添加整合到其基因组中、可操作地连接到启动子的编码重组酶的基因。在另一个实施例中,所述重组酶是Cre重组酶。在另一个实施例中,所述宿主细胞具有整合到其基因组中、可操作地连接到启动子的编码调节蛋白的基因。在更多实施例中,所述调节蛋白是tet阻遏蛋白。

[0034] 在一个实施例中,第一GOI和第二GOI编码抗体轻链或其片段,或抗体重链或其片段。在另一个实施例中,第一GOI编码抗体轻链并且第二GOI编码抗体重链。

[0035] 在某些实施例中,第一、第二和第三GOI编码选自由以下各项组成的群组的多肽:第一轻链或其片段、第二轻链或其片段和重链或其片段。在又一个实施例中,第一、第二和第三GOI编码选自由以下各项组成的群组的多肽:轻链或其片段、第一重链或其片段和第二重链或其片段。

[0036] 在一个方面,提供一种用于制备所关注蛋白的方法,其包含(a)将所关注基因(GOI)引入到CHO宿主细胞中,其中所述GOI整合到包含与SEQ ID NO:1或SEQ ID NO:4至少90%一致的核苷酸序列的特定基因座中;(b)在允许所述GOI表达的条件下培养(a)的细胞;和(c)回收所述所关注蛋白。在一个实施例中,所述所关注蛋白选自由以下各项组成的群组:免疫球蛋白的亚单位或其片段,和受体或其配体结合片段。在某些实施例中,所述所关注蛋白选自由以下各项组成的群组:抗体轻链或其抗原结合片段,和抗体重链或其抗原结合片段。

[0037] 在一些实施例中,采用用于重组酶介导的盒交换(RMCE)的靶向载体将GOI引入到细胞中,并且CHO宿主细胞基因组包含在特定基因座内的至少一个外源性识别序列。在其它实施例中,CHO宿主细胞基因组包含在特定基因座内、任选地连接到启动子、IRES和/或聚腺苷酸化(polyA)序列的至少一个外源性识别序列和可选标记。

[0038] 在某些实施例中,CHO宿主细胞基因组包含如上所述的一个或多个重组酶识别位点,并且GOI是经由重组酶识别重组酶识别位点的作用引入到特定基因座中。

[0039] 在另一个实施例中,采用用于同源重组的靶向载体将GOI引入到细胞中,并且其中所述靶向载体包含与特定基因座中所存在的序列同源的5'同源臂、GOI和与特定基因座中所存在的序列同源的3'同源臂。在另一个实施例中,所述靶向载体另外包含两个、三个、四个或五个或更多所关注基因。在另一个实施例中,一个或多个所关注基因被可操作地连接到启动子。

[0040] 在另一个方面,提供一种靶向载体,其中所述靶向载体包含与包含与SEQ ID NO:1

或SEQ ID NO:4至少90%一致的核苷酸序列的基因座中所存在的序列同源的5'同源臂、GOI和与包含与SEQ ID NO:1或SEQ ID NO:4至少90%一致的核苷酸序列的基因座中所存在的序列同源的3'同源臂。在另一个实施例中,所述靶向载体另外包含两个、三个、四个或五个或更多所关注基因。

[0041] 在另一个方面,提供一种用于修饰CHO细胞基因组以整合外源性核酸序列的方法,其包含将包括载体的载具引入到所述细胞中的步骤,其中所述载体包含外源性核酸序列,其中所述外源性核酸整合在所述基因组的包含与SEQ ID NO:1或SEQ ID NO:4至少90%一致的核苷酸序列的基因座内。

[0042] 在一些实施例中,所述载体包含与基因组的包含与SEQ ID NO:1或SEQ ID NO:4至少90%一致的核苷酸序列的基因座中所存在的序列同源的5'同源臂、外源性核酸序列和与基因组的包含与SEQ ID NO:1或SEQ ID NO:4至少90%一致的核苷酸序列的基因座中所存在的序列同源的3'同源臂。

[0043] 在一些实施例中,所述载体中的外源性核酸序列包含

[0044] 一个或多个识别序列。在其它实施例中,外源性核酸包含一个或多个GOI,如可选标记或编码POI的核酸。在又其它实施例中,外源性核酸包含一个或多个GOI和一个或多个识别序列。

[0045] 在一个实施例中,载具包含至少一个额外载体或mRNA。在另一个实施例中,额外载体选自以下各项组成的群组:腺病毒、慢病毒、逆转录病毒、腺相关病毒、整合性噬菌体载体、非病毒载体、转座子和/或转座酶、整合酶底物和质粒。在一些实施例中,额外载体包含编码用于整合外源性核酸序列的位点特异性核酸酶的核苷酸序列。

[0046] 在某些实施例中,位点特异性核酸酶包含锌指核酸酶(ZFN)、ZFN二聚体、转录活化因子样效应子核酸酶(TALEN)、TAL效应子结构域融合蛋白或RNA指导的DNA核酸内切酶。

[0047] 在另一个方面,提供一种用于修饰CHO细胞基因组以整合外源性核酸序列的载具,其中所述载具包括载体,其中所述载体包含与所述基因组的包含与SEQ ID NO:1或SEQ ID NO:4至少90%一致的核苷酸序列的基因座中所存在的序列同源的5'同源臂、外源性核酸序列和与所述基因组的包含与SEQ ID NO:1或SEQ ID NO:4至少90%一致的核苷酸序列的基因座中所存在的序列同源的3'同源臂。

[0048] 在一些实施例中,外源性核酸序列包含一个或多个识别序列。在其它实施例中,外源性核酸包含一个或多个GOI,如可选标记或编码POI的核酸。在又其它实施例中,外源性核酸包含一个或多个GOI和一个或多个识别序列。

[0049] 在又一个方面,提供一种用于修饰CHO细胞基因组以表达治疗剂的方法,所述治疗剂包含用于引入到所述基因组中的载具、包含用于表达所述治疗剂的序列的外源性核酸,其中所述载具包含与SEQ ID NO:1的核苷酸序列中所存在的序列同源的5'同源臂、编码所述治疗剂的核酸和与SEQ ID NO:1或SEQ ID NO:4的核苷酸序列中所存在的序列同源的3'同源臂。

[0050] 在再一个方面,本发明提供一种经修饰的CHO宿主细胞,其包含经修饰的CHO基因组,其中所述CHO基因组是通过在所述基因组的具有与SEQ ID NO:1至少90%一致的核苷酸序列的基因座内插入外源性识别序列而经修饰的。

[0051] 在另一个方面,本发明提供一种经修饰的真核宿主细胞,其包含经修饰的真核基

基因组,其中所述真核基因组在所述基因组的非编码区中的目标整合位点处经修饰以插入外源性核酸。在一些实施例中,外源性核酸是识别序列。在其它实施例中,宿主细胞是哺乳动物宿主细胞,如CHO细胞。在其它实施例中,目标整合位点包含如SEQ ID NO:1的表达增强序列,其条件是所述序列不编码任何内源性蛋白。本发明还提供制备这类经修饰的真核宿主细胞的方法。

[0052] 在上文所述的任何方面和实施例中,表达增强序列可以按与SEQ ID NO:1中一样的指示方向或与SEQ ID NO:1的方向逆向安置。

[0053] 除非另有说明或上下文中显而易见,否则本发明的任何方面和实施例可以与本发明的任何其它方面或实施例结合使用。

[0054] 通过审阅接下来的详细说明,其它目标和优点将变得显而易见。

附图说明

[0055] 图1A和1B.图1A:利用将表达GOI(例如,多链抗体)的核酸分子和选择标记的多个拷贝随机引入到细胞基因组(例如用于鉴别目标基因座的CHO基因组)中的可操作构建体的示意图。示例性构建体包括:重链(HC);第一拷贝选择标记,如:潮霉素抗性基因(Hyg);第一拷贝轻链(LC);第二拷贝选择标记(例如Hyg)、第二拷贝轻链(LC);第三复本选择标记(例如Hyg)。图1B:经由同源重组整合到天然基因座中的实例供体载体鉴别为SEQ ID NO:1.5'和3'同源臂来源于SEQ ID NO:1。

[0056] 图2A至2C说明SEQ ID NO:1的基因座(LOCUS 1)可操作地连接到所关注基因(GOI)与未可操作地连接到LOCUS 1而是连接到对照基因座的相同GOI相比,展现GOI的mRNA表达增强。图2A:编码所关注抗体基因的细胞,即可操作地连接到对照基因座对比LOCUS 1的一个重链(HC)和两个轻链(LC),展现相等数目的基因拷贝。图2B:与对照基因座mRNA相比,表达LOCUS 1中的GOI的mRNA水平较高。图2C:表达LOCUS 1中的GOI的细胞的蛋白效价比由表达对照基因座中的相同GOI的细胞所产生的蛋白效价高3倍。

[0057] 图3A和3B说明整合在LOCUS 1处的包含荧光标记和GOI的实例盒(例如侧接lox位点的mKate与eYFP和GOI交换)与整合在对照基因座处的相同盒(与侧接lox位点的不同荧光标记交换,例如dsRed2)相比,其中这类整合采用Cre重组酶和重组酶介导的盒交换(RMCE)。将这类盒用于实验中以测量GOI的重组效率和转录。

[0058] 图4显示表达LOCUS 1(SEQ ID NO:1)中的GOI的CHO细胞池中所测量的所关注基因(GOI)的mRNA水平与表达在相同调控条件下、但整合在对照基因座(即EESYR)内的相同GOI的CHO细胞池的mRNA相比较。

具体实施方式

[0059] 在描述本发明的方法前,应当理解,本发明不限于所述的特定方法和实验条件,因为这类方法和条件可以变化。还应当理解,本文所用的术语仅用于描述具体实施例的目的,而无意进行限制,因为本发明的范围将仅由所附权利要求书限制。

[0060] 如在本说明书和所附权利要求书中所用,除非上下文另外明确规定,否则单数形式“一个/种(a/an)”和“所述”包括多个提及物。因此,例如提及“一种方法”包括本文所述类型的和/或在阅读本公开后对所属领域的技术人员将变得显而易见的一种或多种方法和/

或一个或多个步骤。

[0061] 除非另有定义或另有规定,否则本文所用的所有技术和科学术语均具有与本发明所属领域的普通技术人员通常理解相同的含义。

[0062] 尽管与本文中描述的那些方法和材料类似或等效的任何方法和材料均可用于本发明的实践或检验,但现在描述特定的方法和材料。本文提及的所有出版物均以全文引用的方式并入本文中。

[0063] 定义

[0064] DNA区当在功能上彼此相关时是可操作地连接的。举例来说,如果启动子能够参与编码序列的转录,那么所述启动子被可操作地连接到所述序列;如果核糖体结合位点经定位以便允许翻译,那么所述核糖体结合位点被可操作地连接到编码序列。一般来说,可操作地连接可以包括但并不要求邻接。就如分泌性前导序列的序列来说,邻接并且适当放置在阅读框中是典型的特征。在所关注基因座的表达增强序列在功能上与所关注基因(GOI)相关的情况下,例如在其存在使得GOI的表达增强和/或整合稳定的情况下,其被可操作地连接到GOI。

[0065] 术语“增强”在用于描述增强的表达时包括例如与相同表达构建体的单一拷贝的随机整合体池相比,超过通常通过将外源性序列随机整合到基因组中或通过整合在不同基因座所观察到的至少约1.5倍增强到至少约3倍增强的表达。采用本发明的序列所观察到的加倍表达增强是与在基本上相同的条件下、在本发明的序列不存在的情况下所测量的相同基因的表达水平相比,例如与整合到相同物种基因组中的另一基因座相比。增强的重组效率包括基因座重组能力的增强(例如,采用重组酶识别位点)。增强是指超过随机重组(例如,不采用重组酶识别位点等)的重组效率,其通常是0.1%。优选增强的重组效率超过随机约10倍,或是约1%。除非规定,否则要求保护的发明不限于特定的重组效率。

[0066] 在关于所关注基因座采用短语“外源添加的基因”或“外源添加的核酸”的情况下,所述短语是指作为自然界中所发现的基因座所关注基因座内不存在的任何DNA序列或基因。举例来说,CHO基因座(例如包含SEQ ID NO:1序列的基因座)内“外源添加的基因”可以是自然界中特定CHO基因座内未发现的仓鼠基因(即,仓鼠基因来自仓鼠基因组的另一基因座)、来自任何其它物种的基因(例如人类基因)、嵌合基因(例如人类/小鼠)或自然界中未发现的任何其它基因存在于所关注CHO基因座内。

[0067] 当描述所关注基因座(如SEQ ID NO:1或SEQ ID NO:4)或其片段时,一致性百分比意味着包括沿着邻接的同源区展示出所列举一致性的同源序列,但在相比较的序列中不具有同源性的间隙、缺失或插入的存在不纳入一致性百分比的计算中。

[0068] 如本文所用,在例如SEQ ID NO:1或其片段与物种同源物之间的“一致性百分比”测定将不包括在比对中物种同源物无同源序列比较(即,SEQ ID NO:1或其片段在那一点处具有插入,或物种同系物具有间隙或缺失,视具体情况而定)的序列比较。因此,“一致性百分比”不包括间隙、缺失和插入的罚分。

[0069] 在核酸序列的情形下,“同源序列”是指基本上与参考核酸序列同源的序列。在一些实施例中,如果两个序列的至少50%、55%、60%、65%、70%、75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或更多的相应核苷酸在相关的残基序列段上是一致的,那么这两个序列被认为是基本上同源的。在一些实施例中,相关序列段是

完全序列。

[0070] “靶向插入”是指用于引导基因或核酸序列插入或整合到基因组的特定位置,即,引导DNA到相连聚核苷酸链中两个核苷酸之间的特定位点的基因靶向方法。也可以对特定基因盒进行靶向插入,所述基因盒包括多个基因、调控元件和/或核酸序列。“插入”和“整合”可互换使用。应理解,基因或核酸序列(例如包含表达盒的核酸序列)的插入可能导致(或可能经工程改造以使得)一个或多个核酸的替代或缺失,这取决于所采用的基因编辑技术。

[0071] “识别位点”或“识别序列”是由核酸酶或其它酶识别以结合并且引导DNA主链的位点特异性裂解的特定DNA序列。核酸内切酶在DNA分子内裂解DNA。识别位点在所属领域中也被称为识别目标位点。

[0072] “重组酶识别位点”是由重组酶,如Cre重组酶(Cre)或翻转酶(Flp)识别的特定DNA序列。位点特异性重组酶在其一个或多个目标识别序列被战略性置于生物体基因组中时,可以进行DNA重排,包括缺失、倒位和易位。在一个实例中,Cre在其DNA目标识别位点loxP处特异性介导重组事件,所述识别位点是由通过8-bp间隔子分隔开的两个13-bp反向重复序列构成。可以采用不止一个重组酶识别位点,例如以有助于重组介导的DNA交换。也可以采用重组酶识别位点(例如lox位点)的变异体或突变体(Araki, N. 等人, 2002, *Nucleic Acids Research*, 30:19, e103)。

[0073] “重组酶介导的盒交换”涉及一种用供体盒精确替换基因组目标盒的方法。通常为了进行此方法所提供的分子组合物包括1) 5'和3'侧接对特定重组酶具有特异性的识别目标位点的基因组目标盒、2) 侧接匹配的识别目标位点的供体盒和3) 位点特异性重组酶。重组酶蛋白在所属领域中是众所周知的(Turan, S. 和 Bode J., 2011, *FASEB J.*, 25, 第4088-4107页)并且能够精确裂解特定识别目标位点内的DNA(DNA的序列)而不增加或丢失核苷酸。常见重组酶/位点组合包括(但不限于)Cre/lox和Flp/frt。

[0074] “载体”是由携带外源性核酸的任何聚核苷酸或聚核苷酸集合组成的用于引入到细胞中的组合物。载体包括通过众所周知的转染方法递送到细胞的载体、质粒和mRNA分子。在一个实例中,引入到细胞中的mRNA可以是瞬时的并且未整合到基因组中,然而,所述mRNA可以携带进行整合过程所必需的外源性核酸。

[0075] 一般说明

[0076] 本发明至少部分地基于基因组中独特序列(即基因座)的发现,所述序列与基因组中的其它区或序列相比展现更高效的重组、插入稳定性和较高水平表达。本发明还至少部分地基于以下发现:当这类表达增强序列被鉴别时,可以在所述序列中或附近外源添加合适的基因或构建体并且外源添加的基因可以被有利地表达或用于另外的基因组修饰。这类被称为表达增强序列的序列被视为稳定的并且不位于基因组的编码区内。这些表达增强和稳定区可以经工程改造以用于未来的克隆或基因组编辑事件。因此,可靠的表达系统被构建到细胞的基因组主链中。

[0077] 本发明还基于外源性基因特异性靶向整合位点。本发明的方法允许细胞基因组高效“转换”到适用的克隆盒中,例如通过采用重组酶介导的盒交换(RMCE)。为此目的,本发明的方法采用细胞基因组重组酶识别位点安置所关注基因,以便产生用于重组蛋白生产的高产细胞系。

[0078] 本发明的组合物也可以被包括在表达构建体中,例如在用于克隆和工程改造新细胞系的表达载体中。包含本发明的聚核苷酸的表达载体可以用于瞬时表达蛋白,或可以通过随机或靶向重组,如同源重组或由识别特定重组位点的重组酶所介导的重组(例如Cre-lox介导的重组)整合到基因组中。包含本发明的聚核苷酸的表达载体还可以用于评定其它DNA序列,例如顺式作用调控序列的功效。

[0079] 整合位点通常是通过随机整合或分析逆转录病毒整合事件来加以鉴别。本文中详细描述 CHO 整合位点是通过随机整合编码多链抗体的DNA并且发现所表达蛋白展现增强的表达来加以鉴别。

[0080] 将包含一个重链(HC)和两个轻链(LC)复本的实例多链抗体随机整合到基因组的含有交替潮霉素抗性基因的表达盒中(参见,例如图1A中所描绘的三个一致的Hyg基因)。通过将表达盒整合在鉴别为SEQ ID NO:1的基因座内产生一个稳定的高表达克隆。

[0081] 与整合到CHO基因组的另一区(对照整合位点)中相比,当整合在SEQ ID NO:1的基因座内时实例多链抗体展现较高表达水平。有趣的是,整合在SEQ ID NO:1相对于对照整合位点内的抗体表达聚核苷酸是基因复本数是相当的,然而整合在SEQ ID NO:1内的抗体表达聚核苷酸的蛋白效价高3倍。

[0082] 使用靶向重组方法将CHO细胞基因组转换到含有重组酶识别位点的克隆构建体中(参见,例如图3A-B)。

[0083] 本质上,在鉴别SEQ ID NO:1的整合位点后,采用基因座中的重组酶识别位点(例如lox位点)引入表达盒,所述表达盒包含可表达的GOI,如可选标记(参见,例如图3A-B),以及任何其它所需元件,例如启动子、增强子、标记、操纵子、核糖体结合位点(例如内部核糖体入口位点)等。

[0084] 用于靶向整合SEQ ID NO:1内的lox位点的实例供体构建体的图示展示于图1B中。所述供体构建体包含由新霉素(neo)抗性基因和内部核糖体进入位点(IRES)驱动的表达盒,其中所述盒包含荧光标记(mKate)并且在5'和3'端侧接有重组酶识别位点和5'和3'同源臂(与SEQ ID NO:1同源)。展示在SEQ ID NO:1的基因座内的插入,其中所述插入使得供体neo/mKate构建体替换包含潮霉素抗性标记的表达盒,其中在SEQ ID NO:1基因座内的表达盒在其5'和3'端侧接与5'和3'同源臂(与SEQ ID NO:1同源)连接的重组酶识别位点(参见图1B)。

[0085] 提供用于将核酸序列稳定整合到真核细胞中的组合物和方法,其中所述核酸序列能够借助于整合在SEQ ID NO:1或其表达增强片段中来增强表达。提供含有SEQ ID NO:1内的重组酶识别序列便于插入GOI的细胞,以便由GOI表达所关注蛋白。还提供靶向与表达构建体(例如表达载体)有关的整合位点和用于将外源性核酸添加到所关注CHO细胞中的组合物和方法。

[0086] CHO整合位点的物理和功能表征

[0087] 凭经验由高水平表达蛋白的细胞系的核酸构建体(包含表达盒)整合位点的上游和下游序列鉴别SEQ ID NO:1的核酸序列(和更广的SEQ ID NO:4的核酸序列)。本发明的核酸序列提供具有与核酸(例如,包含GOI的外源性核酸)增强的表达和稳定性相关联的新功能的序列,并且在不受任何一个理论束缚的情况下,可以与先前关于顺式作用元件(如启动子、增强子、基因座控制区、支架附着区或基质附着区)相同或不同地起作用。SEQ ID NO:1

似乎不具有任何开放阅读框 (ORF), 使得基因座不大可能编码新颖的反式活化蛋白。已在 SEQ ID NO:4 的 3' (下游) 基因组基因座中鉴别出假定的锌指蛋白。

[0088] 鉴别关于包含第一潮霉素 (Hyg) 基因、第一GOI、第二Hyg基因、第二GOI、第三Hyg基因和第三GOI编码序列的表达盒整合在CHO基因组DNA非编码区的独特位点内的表达增强活性。从CHO基因组DNA非编码区鉴别包含例如5'分离的1kb区和3'分离的1kb区的表达载体, 至于表达GOI的表达盒能够在CHO细胞用所述表达载体转染后赋予高水平的重组蛋白表达。

[0089] 本发明涵盖包含逆向SEQ ID NO:1片段或SEQ ID NO:4片段的表达载体。还可以产生本文所述的片段的其它组合。还可以产生的本文所述的片段的其它组合的实例包括含有本文中所公开的表达增强序列的多个拷贝的序列, 或通过将所公开的SEQ ID NO:1片段或SEQ ID NO:4片段与其它核苷酸序列组合以实现调控元件的最优组合而衍生的序列。这类组合可以被连续连接或布置以提供SEQ ID NO:1或SEQ ID NO:4片段的最优间距 (例如通过在所述片段之间引入间隔子核苷酸)。调控元件也可以被布置成提供SEQ ID NO:1片段相对于调控元件的最优间距。

[0090] 本文中所公开的SEQ ID NO:1和SEQ ID NO:4是从CHO细胞分离的。发现其它哺乳动物物种 (如人类或小鼠) 对所鉴别的表达增强区具有有限的同源性, 然而可以在来源于灰仓鼠的其它组织类型或其它同源物种的细胞系中发现同源序列, 并且可以通过所属领域中众所周知的技术分离出来。举例来说, 有人可以通过交叉物种杂交或基于PCR的技术鉴别其它同源序列。另外, 可以通过所属领域中众所周知的定点或随机诱变技术在SEQ ID NO:1、SEQ ID NO:4或其片段中所阐述的核苷酸序列中进行改变。接着可以如本文所述测试所得序列变异体的表达增强活性。在核酸一致性方面与SEQ ID NO:1、SEQ ID NO:4或其片段至少约90%一致的具有表达增强活性的DNA可以通过常规实验分离, 并且预期展现表达增强活性。对于SEQ ID NO:1或SEQ ID NO:4的片段, 一致性百分比是指在SEQ ID NO:1片段或SEQ ID NO:4片段中所发现的参考天然序列部分。因此, SEQ ID NO:1、SEQ ID NO:4或其片段的同源物和其变异体也被本发明的实施例所涵盖。

[0091] 在某些实施例中, SEQ ID NO:1的片段选自由以下各项组成的群组: 跨越SEQ ID NO:1的编号10-4,000、100-3,900、200-3,800、300-3,700、400-3,600、500-3,500、600-3,400、700-3,300、800-3,200、900-3,100、1,000-3,000、1,100-2,900、1,200-2,800、1,300-2,700、1,400-2,600、1,500-2,500、1,600-2,400、1,700-2,300、1,800-2,200、1,900-2,100、2,000-2,000、2,100-2,000、2,200-1,900、2,300-1,800、2,400-1,700、2,500-1,600、2,600-1,500、2,700-1,400、2,800-1,300、2,900-1,200、3,000-1,100、3,100-1,000、3,200-900、3,300-800、3,400-700、3,500-600、3,600-500、3,700-400、3,800-300、3,900-200、4,000-100和4,000-10的核苷酸。在另一个实施例中, SEQ ID NO:1的片段选自由以下各项组成的群组: 跨越SEQ ID NO:1的编号10-500、500-1,000、1,000-1,500、1,500-2,000、2,000-2,500、2,500-3,000、3,000-3,500、3,500-4,000和4,000-500的核苷酸。在某些实施例中, 外源性核酸序列整合在上文所述片段内的特定位点处或所述特定位点附近。

[0092] 在另一个实施例中, 外源性核酸序列定位于如上所述的SEQ ID NO:1或其片段内, 或在与SEQ ID NO:1的表达增强序列或其表达增强片段至少约90%一致、至少约91%一致、至少约92%一致、至少约93%一致、至少约94%一致、至少约95%一致、至少约96%一致、至少约97%一致、至少约98%一致或至少约99%一致的序列内。

[0093] 可以使用本文所提供的方法产生表达增强水平的所关注蛋白的细胞群。表达的绝

对水平将随特定蛋白而变化,取决于细胞如何有效地加工蛋白。通过外源性序列整合在本发明的表达增强序列内所产生的细胞池随时间推移而稳定,并且可以作为稳定细胞系加以处理以用于大部分目的。还可以延迟重组步骤直到稍后在本发明细胞系的发展过程中。

[0094] CHO表达增强基因座和其片段

[0095] 本发明涵盖一种表达增强片段,其核苷酸序列与SEQ ID NO:1或SEQ ID NO:4的核苷酸序列至少约90%一致、至少约91%一致、至少约92%一致、至少约93%一致、至少约94%一致、至少约95%一致、至少约96%一致、至少约97%一致、至少约98%一致或至少约99%一致。本发明包括包含以下片段的载体,所述片段是为了瞬时或稳定转染而包括,跨越SEQ ID NO:1的编号10-4,000、100-3,900、200-3,800、300-3,700、400-3,600、500-3,500、600-3,400、700-3,300、800-3,200、900-3,100、1,000-3,000、1,100-2,900、1,200-2,800、1,300-2,700、1,200-2,600、1,300-2,500、1,400-2,400、1,500-2,300、1,600-2,200、1,700-2,100、1,800-2,050、1,850-2,050、1,900-2,040、1,950-2,025、1,990-2,021、2,002-2,021和2,010-2,015的位置。本发明还包括一种包含这类片段的真核细胞,其中所述片段对于所述细胞来说是外源性的并且被整合到所述细胞基因组中,并且包含这类片段的细胞具有至少一个重组酶识别位点,所述重组酶识别位点是在所述片段内、5'紧靠或3'紧靠所述片段。

[0096] 在一个实施例中,SEQ ID NO:1的表达增强片段的位置位于SEQ ID NO:1内跨越SEQ ID NO:1的编号10-500、500-1,000、500-2,100、1,000-1,500、1,000-2,100、1,500-2,000、1,500-2,500、2,000-2,500、2,500-3,000、2,500-3,500、3,000-3,500、3,000-4,000或3,500-4,000的位置。

[0097] 在支持稳定整合和/或所整合聚核苷酸的增强转录的情况下,基因座插入(即整合)位点相对于例示位点的精确位置不是必需的。实际上,整合位点可以如本文所述在SEQ ID NO:1或SEQ ID NO:1的片段或SEQ ID NO:4或SEQ ID NO:4的片段内或相邻的任何位置。在所关注基因座内或相邻的特定染色体位置是否支持稳定整合和所整合外源性基因的高效转录可以根据所属领域中众所周知的标准程序或本文中例示的方法来确定。

[0098] 本文中所考虑的整合位点位于包含SEQ ID NO:1或SEQ ID NO:4的核苷酸序列的基因座内,或极为接近所关注基因座,例如相对于染色体DNA上SEQ ID NO:1的位置上游(5')或下游(3')小于约1kb、500个碱基对(bp)、250bp、100bp、50bp、25bp、10bp或小于约5bp。在又一些其它实施例中,所采用的整合位点相对于染色体DNA上SEQ ID NO:1或SEQ ID NO:4的位置位于上游(5')或下游(3')约1000、2500、5000或更多个碱基对处。

[0099] 在所属领域中应理解,为了高效复制和转录染色体DNA,采用大基因组区,如支架/基质附着区。支架/基质附着区(S/MAR),也称为支架附着区(SAR)或基质相关或基质附着区(MAR),是核基质附着的真核基因组DNA区。在不受任何一个理论束缚的情况下,S/MAR通常定位到非编码区,使给定转录区(例如染色质结构域)与其相邻者分开,并且还提供用于机器加工和/或结合实现转录的因子(如DNA酶或聚合酶的识别位点)的平台。一些S/MAR已表征为约14-20kb长(Klar等人,2005, Gene 364:79-89)。因而,预期基因整合在LOCUS 1处(SEQ ID NO:1或SEQ ID NO:4内或附近)赋予增强的表达。

[0100] 所属领域的技术人员应认识到,数种元件可以被优化以便在目标基因座处具有高转录活性,从而使得所插入的编码所关注蛋白的基因高度表达。有待考虑的元件包括驱动转录的强启动子、足够的转录机器和具有开放并且可接近的构型的DNA。在所属领域人员的

技能内可以通过靶向在SEQ ID NO:1或SEQ ID NO:4内所选的整合位点而优化在目标基因座处的插入。

[0101] 在一个实施例中,采用SEQ ID NO:1的表达增强序列以增强GOI的表达。图2A显示可操作地连接到SEQ ID NO:1 (LOCUS 1)的GOI与整合在CHO细胞基因组中的不同基因座(对照基因座)的相同GOI相比的结果,各细胞系所测量的基因拷贝数是相等的,但实验显示对于可操作地连接到LOCUS 1的GOI,细胞表达GOI的mRNA水平和蛋白效价高3倍。

[0102] 在各种实施例中,GOI的表达可以通过将GOI放置在SEQ ID NO:1或SEQ ID NO:4内来增强。在各种实施例中,表达增强至少约1.5倍至约3倍或更多。

[0103] 基因修饰目标基因座

[0104] 基因工程改造特定位置(即目标基因座)中的细胞基因组的方法可以用数种方式达成。使用遗传编辑技术将核酸序列稳定整合到真核细胞中,其中所述核酸序列是通常并未在这类细胞中所发现的外源性序列。克隆扩增是为了确保细胞子代将享有经工程改造的细胞系的一致基因型和表现型特征所必需的。在一些实例中,天然细胞是通过同源重组技术来修饰以便将外源性核酸序列整合在SEQ ID NO:1或SEQ ID NO:4内。在其它实例中,提供在SEQ ID NO:1或SEQ ID NO:4内含有至少一个重组酶识别序列的细胞,以便于整合外源性核酸序列或所关注基因。

[0105] 在一些实例中,提供含有第一重组酶识别序列和第二重组酶识别序列的细胞,其中所述第一和第二重组酶识别序列各选自包含以下各项的群组:LoxP、Lox511、Lox5171、Lox2272、Lox2372、Loxm2、Lox-FAS、Lox71、Lox66和其突变体。在这种情况下,如果需要重组酶介导的盒交换(RMCE),那么位点特异性重组酶是Cre重组酶或其衍生物。在其它实例中,第一和第二重组酶识别序列各选自包含FRT、F3、F5、FRT突变体-10、FRT突变体+10和其突变体的群组,并且在这种情境下,如果需要RCME,那么位点特异性重组酶是Flp重组酶或其衍生物。在又一个实例中,所述第一和第二重组酶识别序列各选自包含attB、attP和其突变体的群组,并且在这种情况下,如果需要RMCE,那么位点特异性重组酶是phiC31整合酶或其衍生物。

[0106] 在一个方面,用于将核酸序列稳定整合在SEQ ID NO:1或SEQ ID NO:4或其表达增强片段内的方法和组合物是经由同源重组。所关注核酸分子,即基因或聚核苷酸,可以通过同源重组或通过使用特异性靶向整合位点处的序列的位点特异性核酸酶方法插入到所靶向的基因座(即SEQ ID NO:1)中。关于同源重组,同源聚核苷酸分子(即同源臂)对其并且交换它们的一段序列。如果转基因侧接同源基因组序列,那么可以在此交换期间引入转基因。在一个实例中,可以在整合位点处将重组酶识别位点引入到宿主细胞基因组中。

[0107] 可以通过在染色体DNA中的整合位点处引入断裂来促进真核细胞中的同源重组。模型系统已证明,如果在染色体目标序列中引入双链断裂,那么在基因靶向期间同源重组的频率会增加。这可以通过将某些核酸酶靶向特定整合位点而实现。在目标基因座识别DNA序列的DNA结合蛋白是所属领域中已知的。基因靶向载体也用于促进同源重组。在用于同源指导修复(homology directed repair)的基因靶向载体不存在的情况下,细胞常常会通过非同源末端接合(NHEJ)(其可能在裂解位点处导致多个核苷酸的缺失或插入)来闭合双链断裂。应存在插入或缺失(InDel),因而在断裂位点处随机插入或缺失少量核苷酸并且这些InDel可以移位或破坏目标基因座内基因的任何开放阅读框(ORF)。应理解,鉴别为SEQ ID

NO:1 (或SEQ ID NO:4) 的基因座不是基因编码区。因此,设想在此基因座的插入和/或缺失不破坏内源性基因转录。

[0108] 同源指导修复(或同源指导重组)(HDR)特别适用于在目标基因座插入或整合基因。供体构建体包含如本文所述的来源于SEQ ID NO:1或SEQ ID NO:4的同源臂。

[0109] 基因靶向载体构建和核酸酶选择在本发明所属领域的技术人员的技能内。

[0110] 在一些实例中,具有模块化结构并含有单独锌指域的锌指核酸酶(ZFN)识别目标序列中的特定3-核苷酸序列(例如靶向整合的位点)。一些实施例可以利用具有靶向多个目标序列的单独锌指域的组合的ZFN。

[0111] 转录活化因子样(TAL)效应子核酸酶(TALEN)也可以用于位点特异性基因组编辑。TAL效应子蛋白DNA结合域通常与限制性核酸酶(如FokI)的非特异性裂解域组合使用。在一些实施例中,将包含TAL效应子蛋白DNA结合域和限制性核酸酶裂解域的融合蛋白用于识别和裂解本发明基因座内的目标序列处的DNA(Boch J等人,2009Science 326:1509-1512)。

[0112] RNA指导的核酸内切酶(RGEN)是从细菌适应性免疫机制开发的可编程的基因组工程改造工具。在此系统(成簇规律间隔短回文重复序列(CRISPR)/CRISPR相关性(Cas)免疫反应)中,蛋白Cas9当与两个RNA(其中一个指导靶选择)复合时形成序列特异性核酸内切酶。RGEN由组分(Cas9和tracrRNA)以及靶特异性CRISPR RNA(crRNA)组成。DNA靶裂解的效率以及裂解位点的位置均基于前间区序列邻近基序(PAM)的位置而变化,所述基序是针对靶识别的额外要求(Chen,H.等人,J.Biol.Chem.2014年3月14作为手稿M113.539726在线发表)。

[0113] 用于鉴别SEQ ID NO:1的特异性靶向基因座特有的序列的策略是所属领域中已知的,然而,许多这些序列与CHO基因组的比对显露具有16-17个碱基对匹配的潜在脱靶位点。由SEQ ID NO:5中所阐述的序列(与SEQ ID NO:1的核苷酸1990-2001对应)编码的一个实例20bp指导RNA适用于SEQ ID NO:1或SEQ ID NO:4的RNA指导的CRISPR/Cas基因编辑。包含驱使小引导RNA和tracrRNA(例如SEQ ID NO:6)表达的启动子以及携载在启动子控制下的合适Cas9酶的质粒可以与供体载体(携载侧接5'和3'同源臂的所关注基因)一起共转染以便采用通过此方法的靶向整合。除上文所述的那些以外的各种修饰和RNA分子的变体对于所属领域的技术人员显而易见并且打算属于本发明的范围内。

[0114] 在一些实施例中,用于引入到基因组中的载具,即包含编码所关注基因的序列或识别序列或基因盒的外源性核酸,视具体情况而定包含携载所述外源性核酸的载体和一个或多个额外载体或mRNA。在一个实施例中,所述一个或多个额外载体或mRNA包含编码位点特异性核酸酶的核苷酸序列,所述位点特异性核酸酶包括(但不限于)锌指核酸酶(ZFN)、ZFN二聚体、转录活化因子样效应子核酸酶(TALEN)、TAL效应子结构域融合蛋白和RNA指导的DNA核酸内切酶。在某些实施例中,所述一个或多个载体或mRNA包含具有指导RNA、tracrRNA和编码Cas酶的核苷酸序列的第一载体,以及包含供体(外源性)核苷酸序列的第二载体。这类供体序列包含编码所关注基因的核苷酸序列,或识别序列,或包含打算用于靶向插入的这些外源性元件中的任一个的基因盒。在使用mRNA的情况下,mRNA可以借助于所属领域的技术人员已知的常见转染方法转染到细胞中并且可以编码酶,例如转座酶或核酸内切酶。虽然引入到细胞中的mRNA可以是瞬时的并且未整合到基因组中,但是所述mRNA可以携载对于进行整合来说所必需或有益的外源性核酸。在一些情况下,如果仅需要短期表

来实现GOI的所需整合,那么选择mRNA是为了消除附加聚核苷酸副作用持久的任何风险。

[0115] 另外其它同源重组方法可供技术人员使用,如具有精确DNA结合特异性的BuD衍生的核酸酶(BuDN) (Stella, S.等人Acta Cryst. 2014, D70, 2042-2052)。精确的基因组修饰方法是基于与SEQ ID NO:1内的独特目标序列相容的可用工具来选择,以避免细胞表型被破坏。

[0116] 基因靶向构建体

[0117] 有待整合到宿主基因组中的聚核苷酸序列可以是任何工业上适用于生成细胞表达系统的DNA序列,如识别序列。有待整合到宿主基因组中的聚核苷酸序列可以编码如本文所述的任何治疗上或工业上适用的蛋白。鉴别目标基因座内的目标序列以便整合外源性核酸序列取决于多种因素。根据所采用的同源重组方法,选择与SEQ ID NO:1或SEQ ID NO:4同源的序列正好属于技术人员的技能。位点特异性核酸酶载体在采用时,需要识别打算用于DNA裂解的特定位点的额外组分(序列组合物)。

[0118] 因此,基因靶向构建体通常并入这类核苷酸序列以便于外源性核酸序列靶向整合到所关注基因座中。在一些实施例中,构建体包含第一同源臂和第二同源臂。在其它实施例中,构建体(例如基因盒)包含来源于SEQ ID NO:1或SEQ ID NO:4的同源臂。在一些实施例中,同源臂包含与SEQ ID NO:1或SEQ ID NO:4中所存在的核苷酸序列同源的核苷酸序列。在特定实施例中,构建体包含具有SEQ ID NO:2的核苷酸序列(与SEQ ID NO:1的核苷酸1001-2001对应)的5'同源臂和具有SEQ ID NO:3的核苷酸序列(与SEQ ID NO:1的核苷酸2022-2001对应)的3'同源臂。同源臂,例如第一同源臂(也称为5'同源臂)和第二同源臂(也称为3'同源臂),与基因座内的靶向序列同源。5'到3'的同源臂可以扩增基因座内包含至少1kb、或至少约2kb、或至少约3kb、或至少约4kb、或至少5kb、或至少约10kb的区或靶向序列。在其它实施例中,选择用于第一和第二同源臂的靶向序列的核苷酸总数包含至少1kb、或至少约2kb、或至少约3kb、或至少约4kb、或至少5kb、或至少约10kb。在一些情况下,5'同源臂与3'同源臂(与靶向序列同源)之间的距离包含至少5bp、10bp、20bp、30bp、40bp、50bp、60bp、70bp、80bp、90bp、100bp、200bp、300bp、400bp、500bp、600bp、700bp、800bp、900bp、或至少1kb、或至少约2kb、或至少约3kb、或至少约4kb、或至少5kb、或至少约10kb。在选择SEQ ID NO:2和SEQ ID NO:3作为5'和3'同源臂的情况下,两个同源臂之间的距离可以是20个核苷酸(与SEQ ID NO:1的核苷酸2002-2021对应);并且这类同源臂可以介导外源性核酸序列整合在包含SEQ ID NO:1的基因座内,例如SEQ ID NO:1的核苷酸1990-2021或2002-2021内和同时SEQ ID NO:1的核苷酸2002-2021的缺失。

[0119] 在其它实施例中,构建体包含第一同源臂和第二同源臂,其中所组合的第一和第二同源臂包含替换基因座内的内源性序列的靶向序列。在又其它实施例中,第一和第二同源臂包含整合或插入在基因座内的内源性序列内的靶向序列。

[0120] 经修饰的细胞系是通过在SEQ ID NO:1内的位置整合一个或多个重组酶识别位点来创建。这些经修饰的细胞系还可以包含额外外源性基因用于表达的所关注基因的阴性或阳性选择。

[0121] 本发明提供用于修饰CHO细胞基因组的方法,其包含将一个或多个载体引入到所述细胞中,其中所述一个或多个载体包含具有用于整合的序列的外源性核酸、与SEQ ID NO:1的核苷酸序列中所存在的序列同源的5'同源臂和与SEQ ID NO:1的核苷酸序列中所存

在的序列同源的3'同源臂。在一些实施例中,所述方法另外提供包含核酸酶和用于在整合位点处的位点特异性DNA裂解的组合物中的一个或多个载体。

[0122] 经修饰的细胞系可以作为方便并且稳定的表达系统而用于重组酶介导的盒交换(RMCE)。编码所关注蛋白的核酸序列可以方便地整合到包含SEQ ID NO:1或其表达增强片段、具有至少一个重组酶识别位点的经修饰的细胞中,例如经由RMCE方法。

[0123] 重组表达载体可包含编码蛋白的合成的或cDNA衍生的DNA片段,其可操作地连接到来源于哺乳动物、病毒或昆虫基因的合适的转录和/或翻译调控元件。这类调控元件包括转录启动子、增强子、编码合适的mRNA核糖体结合位点的序列以及控制转录和翻译终止的序列,如下文详细描述。哺乳动物表达载体还可包含非转录元件,如复制起点、其它5'或3'侧翼非转录序列,以及5'或3'非翻译序列,如剪接供体和受体位点。还可并入帮助识别转染子的可选标记基因。

[0124] 荧光标记是适用于识别已经或尚未成功地插入和/或替换的基因盒的可选标记基因,视具体情况而定。荧光标记的实例是所属领域中众所周知的,包括(但不限于)Discosoma珊瑚(DsRed)、绿色荧光蛋白(GFP)、增强型绿色荧光蛋白(eGFP)、蓝绿色荧光蛋白(CFP)、增强型蓝绿色荧光蛋白(eCFP)、黄色荧光蛋白(YFP)、增强型黄色荧光蛋白(eYFP)和远红外荧光蛋白(例如mKate、mKate2、mPlum、mRaspberry或E2-crimson。还参见例如Nagai,T.等人,2002Nature Biotechnology 20:87-90;Heim,R.等人1995年2月23日Nature 373:663-664;和Strack,R.L.等人2009Biochemistry 48:8279-81。

[0125] 适用于转染脊椎动物细胞的表达载体中的转录和翻译控制序列可由病毒来源而提供。举例来说,常用的启动子和增强子来源于病毒,如多瘤病毒、腺病毒2、猿猴病毒40(SV40)和人类巨细胞病毒(CMV)。病毒基因组启动子、控制和/或信号序列可用于驱动表达,所提供的这类控制序列与所选择的宿主细胞相容。还可以使用非病毒细胞启动子(例如 β -球蛋白和EF-1 α 启动子),取决于表达重组蛋白的细胞类型。

[0126] 来源于SV40病毒基因组的DNA序列,例如SV40起点、早期和晚期启动子、增强子、剪接和聚腺苷酸化位点可用于提供对异源DNA序列的表达有用的其它基因元件。早期和晚期启动子是特别有用的,因为二者可容易地从SV40病毒作为还包含SV40病毒复制起点的片段得到(Fiers等人,Nature 273:113,1978)。也可使用较小或较大的SV40片段。通常,包括从位于SV40复制起点中的Hind III位点向BgII位点延伸的大约250bp序列。

[0127] 用于表达多个转录物的双顺反子表达载体先前已有描述(Kim S.K.和Wold B.J.,Cell 42:129,1985)并且可以与本发明的表达增强序列(例如SEQ ID NO:1)或其片段组合使用。其它类型的表达载体也将是有用的,例如描述于美国专利第4,634,665号(Axel等人)和美国专利第4,656,134号(Ringold等人)中的那些。

[0128] 所关注蛋白

[0129] 可使用适于在真核细胞中表达的任何所关注蛋白。举例来说,所关注蛋白包括(但不限于)抗体或其抗原结合片段、嵌合抗体或其抗原结合片段、ScFv或其片段、Fc融合蛋白或其片段、生长因子或其片段、细胞因子或其片段、或细胞表面受体的胞外域或其片段。所关注蛋白可以由单个亚单位组成的简单多肽或包含两个或更多个亚单位的复杂多亚单位蛋白。

[0130] 宿主细胞和转染

277-28)。还可以使用病毒方法进行转染。mRNA递送包括使用TransMessenger™和TransIT®的方法(Bire等人BMC Biotechnology 2013,13:75)。

[0136] 将异源DNA引入到细胞中的一种常用方法是磷酸钙沉淀,如Wigler等人(Proc.Natl.Acad.Sci.USA 77:3567,1980)所述。通过此方法引入到宿主细胞中的DNA经常进行重排,使得此程序适用于单独基因的共转染。

[0137] 聚乙烯诱导的细菌原生质体与哺乳动物细胞的融合(Schaffner等人,(1980) Proc.Natl.Acad.Sci.USA 77:2163)是另一种适用于引入异源DNA的方法。原生质体融合方案经常产生整合到哺乳动物宿主细胞基因组中的质粒DNA的多个拷贝,并且此技术需要选择和扩增标记与GOI在同一质粒上。

[0138] 还可以使用电穿孔将DNA直接引入到宿主细胞的细胞质中,如Potter等人(Proc.Natl.Acad.Sci.USA 81:7161,1988)或Shigekawa等人(BioTechniques 6:742,1988)所述。与原生质体融合不同,电穿孔不需要选择标记和GOI在同一质粒上。

[0139] 已描述适用于将异源DNA引入到哺乳动物细胞中的其它试剂,如Lipofectin™试剂和Lipofectamine™试剂(Gibco BRL,Gaithersburg,Md.)。这两种可商购的试剂均用于形成脂质-核酸复合物(或脂质体),当应用于培养的细胞时,有利于核酸摄入细胞中。

[0140] 在一个实施例中,将一个或多个聚核苷酸引入到细胞中是通过电穿孔、通过胞质内注射、通过病毒感染、通过腺病毒、通过慢病毒、通过逆转录病毒、通过转染、通过脂质介导的转染来介导或经由Nucleofection™介导。

[0141] 用于扩增GOI的方法同样是重组蛋白表达所需的,并且通常涉及使用选择标记(在Kaufman同上中进行了综述)。对细胞毒性药物的抗性是最常用作选择标记的特征,并且可以是显性性状(例如可独立于宿主细胞类型使用)或隐性性状(例如适用于缺乏所选任何活性的特定宿主细胞类型)的结果。数种可扩增标记适用于本发明的表达载体中(例如,如Sambrook,Molecular Biology:A Laboratory Manual,Cold Spring Harbor Laboratory, NY,1989;第16.9-16.14页中所述)。

[0142] 适用于在抗药性哺乳动物细胞中基因扩增的可选标记展示于Kaufman,R.J.,同上的表1中,并且包括DHFR-MTX抗性、P-糖蛋白和多药物抗性(MDR)-各种亲脂性细胞毒性剂(例如阿德力霉素、秋水仙碱、长春新碱)和腺苷脱氨酶(ADA)-Xyl-A或腺苷和2'-脱氧柯福霉素。

[0143] 其它显性可选标记包括来源于微生物的抗生素抗性基因,例如新霉素、卡那霉素或潮霉素抗性。然而,尚未显示这些选择标记可扩增(Kaufman,R.J.,同上)。哺乳动物宿主存在数种合适的选择系统(Sambrook同上,第16.9-16.15页)。也已描述采用两个显性可选标记的共转染方案(Okayama和Berg,Mol.Cell Biol 5:1136,1985)。

[0144] 先前已描述或所属领域已知的有用调控元件也可包括在用于转染哺乳动物细胞的核酸构建体中。所选择的转染方案和选择用于其中的元件将取决于所用宿主细胞的类型。所属领域的技术人员知道许多不同的方案和宿主细胞,并且可基于所用的细胞培养系统的要求来选择用于表达所需蛋白的适当系统。

[0145] 本发明的其它特征在示例性实施例的以下描述过程中将变得显而易见,所述示例性实施例为了说明本发明而给出并且并不打算对本发明进行限制。

[0146] 实例

[0147] 提出以下实例是为了向所属领域的普通技术人员描述如何构造和使用本发明的方法和组合物,而非旨在限制本发明的范围。已努力确保有关所用的数字(例如量、温度等)的准确性,但应考虑某些实验误差和偏差。除非另外指明,否则份数是重量份,分子量是平均分子量,温度是按摄氏度计并且压力是大气压或接近大气压。

[0148] 实例1.所关注基因座的鉴别和整合位点的表征

[0149] 用含有抗体序列和可选抗生素抗性基因作为可选标记的两个质粒转染CHO K1细胞。通过在抗生素存在下扩增细胞进行稳定转染物的选择。用**FASTR®**分选技术分离表达高水平抗体的个别细胞克隆(参见美国专利第8673589B2号)。鉴别展现最高抗体表达水平的数种克隆。

[0150] 使用Covaris Adaptive Focused Acoustics (AFA)TM技术将来自这些克隆的基因组DNA片段化(Fisher,S.等人2011,Genome Biology 12:R1)。使用针对引入到CHO细胞中的全部质粒序列所设计定制的生物素化RNA诱饵(Agilent SureSelectXT#5190-4811)产生并且培育DNA文库(Agilent SureSelectXT#G9612A)。含有质粒序列的基因组DNA片段富含抗生蛋白链菌素磁珠并且对其进行Illumina MiSeq测序以鉴别质粒整合位点。分析含有质粒序列和CHO基因组序列的融合序列并且与CHO基因组比对。单个整合位点是通过南方墨点分析和PCR随后测序来确认。将具有SEQ ID NO:1的核苷酸序列的整合位点鉴别为表达热点(也参见GenBank基因座ID号AFTD01150902.1,nt35529:39558)。分析整合位点以测定其用于进一步生成细胞系的适用性。所期望的是,整合位点位于非编码区中,这样不破坏细胞正常基因组机制(例如蛋白的翻译)或改变细胞表型。

[0151] 根据Blat检索(Kent WJ.,BLAT-the BLAST-like alignment tool.Genome Res.2002年4月;12(4):656-64)比对,SEQ ID NO:1与小鼠和人类基因组序列共有极低同源性。SEQ ID NO:1相对于CHO-1[ATCC]_refseq_transcript(www.chogenome.org)的序列blast揭露所鉴别的基因座序列不含任何已知基因的任何编码区。SEQ ID NO:4的更广序列,其涵盖SEQ ID NO:1,也被鉴别为适于靶向整合的基因座。

[0152] 整合位点序列经测定位于CHO和小鼠基因组的非编码区,并且进一步用于下文所述的实验中。

[0153] 实例2.高效并入到宿主细胞整合位点中的外源性DNA

[0154] 通过采用TALE核酸酶(TALEN)将外源性基因靶向插入到鉴别为SEQ ID NO:1的CHO基因组的特定基因座中。TALEN靶向如实例1中的含有随机整合到细胞基因组中的抗体重链和轻链序列的构建体。TALEN靶向抗体表达构建体的三个相同Hyg基因内的位置(参见图1A)。Hyg序列的TALEN目标裂解位点是基于ZiFit.partners.org(ZiFit Targeter版本4.2)。TALEN是基于已知方法(Boch J等人,2009Science 326:1509-1512)所设计。

[0155] 使用标准脂质体方案(LIPOFECTAMINE,Life Technologies,Gaithersburg,Md.)将供体mKate载体(参见图1B)和TALEN编码载体转染到CHO宿主细胞中。培养细胞并且通过FACS分离和分选具有所需特征的稳定克隆。通过南方墨点和PCR确认所需基因座中的单个整合。

[0156] 实例3.经工程改造的细胞在所关注基因座处通过RMCE的靶向重组

[0157] 选择表达高水平荧光基因(例如mKate)的CHO细胞系进行分离,其中所述基因侧接所关注基因座内的lox位点。第二CHO细胞系表达第二荧光基因(dsRed),其中所述基因侧接

位于对照基因座(即EESYR)内的lox位点(美国专利第8389239B2号,2013年3月5日颁布)。

[0158] 经转染的CHO细胞适于在无血清生产培养基中悬浮生长。细胞接着在十厘米板中用供体表达载体和编码Cre重组酶的质粒转染。供体表达载体含有侧接Lox位点的编码Fc融合蛋白的所关注基因(参见图3A或3B)。细胞在转染后在具有400 μ g/ml潮霉素的培养基中培养两周,并且使用流式细胞测量术分离表达eYFP但不表达mKate(或在EESYR基因座整合的情况下,dsRed)的细胞。表达eYFP的细胞在无血清生产培养基中在悬浮培养物中扩增,并且通过qRT-PCR使用标准程序测定编码Fc融合蛋白的各细胞池的mRNA水平(参见图4)。

[0159] 比较细胞池之间的重组交换效率(从表达供体盒标记,即eYFP交换为表达红色标记,即mKate或dsRed的存活细胞群体的百分比)(表1)。在各基因座观察到高重组交换效率。

[0160] 表1:重组效率

[0161]

	红色标记	交换效率(%) (红色标记+/eYFP-)	随机整合(%) (红色标记+/eYFP+)
LOCUS 1 (SEQ ID NO:1)	mKate	72	27
对照基因座 (EESYR)	dsRed	92	7

[0162] 在具有经工程改造的LOCUS1(与对照基因座相比)的细胞池中观察到较高转录率(高1.5倍)(图4)。

[0163] 本发明的范围不受本文所述的特定实施例限制。实际上,除了本文中所述的那些内容之外,所属领域的技术人员根据前述说明和附图将显而易见本发明的各种修改。希望这些修改属于所附权利要求书的范围内。

序列表

<110> Regeneron Pharmaceuticals, Inc.

<120> 新颖CHO整合位点和其用途

<130> 32353 (T0045US01)

<150> 62/067,774

<151> 2014-10-23

<160> 6

<170> PatentIn version 3.5

<210> 1

<211> 4001

<212> DNA

<213> 灰仓鼠

<400> 1

```

ccaagatgcc catcaactga ttaatagatg ataaaattat tgtacatttc agtghtaatat 60
tattcagttt ttaagaaaaa tgaaattatg taataagcat gtaaattgat atatcttgaa 120
acaaccattc cccattatat tacctaaaca ttgaaagtcc aaaatcatat gatcttttta 180
gtggatctac taatcttttg ctatatgtat tttattgaac tacccatgga tgtgagataa 240
ttggtaacaa cagcacatgg gagagcatgg gatcattcaa ggaagattag agagaatgca 300
tttttttagga gataatggag gagcaataga aaggattaaa tgaggttact gatgaaagtg 360
atggttagag aaggcaatat gaggagggat aactagcact tagggccttt tgaaaaagac 420
atagagaaaa tactattgta gaaacttctc ataattgggtg tatagttata tacaccaaag 480
agctcagatg gagttaccct ataattgaaa tattaactac tttttatcac tgtgataaaa 540
catcctgaac agagcaacat agattgggaa gcatttactt tggcttacag ttctaacggg 600
ataaaaattc atgatgaaag aatgaatatg tcagcaaca gcaagtagca tggcctgaga 660
agcaggtgag agctcacatc ttgaagtgta agaattgtgc agagagaaca aactgcaaat 720
gaccagaaaa tgcttttggg tcagagccca taccctctg actgacttct ccagaaattc 780
tgaacaaata aaactcccca aacagagcca taactgaagg tccagtgtct gagactacta 840
gggggtatttc ttattcaaac cactacaatg ggggtggggg agcaatcctc caagtaggca 900
ctacacacag acaataaaaa actctagtaa ctggaatgga ttgacttatt tgaattactt 960
gccagtggag ctacatagag cacaattatt gtattttaat taccctttat gatcttaca 1020
aacttgacag taagatcata ttgctaaaga aaccacatat ttgaatcagg gaacatgggtg 1080
atatctagtt gttcttcaac tggaaacttc atgctttctg cccagcattc atgttgctgg 1140
aaagagcaat gtacactacc agtgtagaaa ttaaateate aatcttatec agatgtggat 1200
cctataagtt acaataaaaa ttagcctgat aagatatecc caccagaaga atattccat 1260
aaatgctatg ggagcaacaa gctatcttct aaattagctt taatcttatt ctacaagaga 1320
gaatccatat ctagaatagt tatagggate aagaacccat ggcttgattg gtcattagcc 1380
caatgggaga tcctaatatt attgttctac aaaatgaaaa taactcctaa tgacttgttg 1440
ctgcagtaat aagttagatg gttgctcaac tctcaacaaga gaagttttgt cttacaataa 1500

```

atggcaatta aagcagcccc acaagattta tatcataccg atctctcat ggcctatgca 1560
tctagaagct aggaaacaaa gaggacccta agagagacat acatggtecc cctggagaag 1620
gggaaggggg caagacctcc aaagctaatt gggagcatgg gggaggggag agggagttag 1680
aagaaagaga aggggataaa aggagggaga ggaggacaag agagagaagg aagatctagt 1740
caagagaaga tagaggagag caagaaaaga gataccatag tagagggagc cttgtatggt 1800
taaatagaaa actggcacta gggattgtc caaagatcca caaggtccaa ctaataatct 1860
aagcaatagt cgagaggcta ccttaaaagc ctttctctga taatgagatt gatgactacc 1920
ttatatacca tcctagagcc ttcattcagt agctgatgga agcagaagca gacatctaca 1980
gctaaacact gagctagttg cagacaggga ggagtgatga gcaaagtcaa gaccaggctg 2040
gagaaacaca cagaaacagc agacctgaaa aaaatgttgc acatggacc cagactgata 2100
gctggggagtc cagcatagga cttttctaga aacctgaat gaggatatca gtttggagggt 2160
ctggttaatc tatggggaca ctggtagtgg atcaatattt atccctagtt catgactgga 2220
atgtgggtac ccattccaca tggaggaatt ctctgtcage ctagacacat gggggagggt 2280
ctaggtcctg ctccaaataa tgtgttagac tttgaagaac tcccttgaga agactcacc 2340
tccctgggga gcagaaaggg gatgggatga gggttggtga gggacaggag aggaggggag 2400
ggtgagggaa ctgggattga caagtaaatg atgcttgttt ctaatttaa tgaataaagg 2460
aaaagtaaaa gaagaaaaga aaacaggcca aaagattata aaagacagag gtggtgggtg 2520
actataaaga aacactatta tctaaataaa aacatgtcag aagcacacat gaacttatag 2580
tgtttatgaa agtatgtata ataactacat aatctcaagc caagaaaaa atatcatctt 2640
tcagtgatga aggtgatttt atttctcca gaattaaagc caaagacct aatgaaagtaa 2700
ttatcttcaa aaggttgaaa atacatactt tgcaatacac agatctgcct agaaatctca 2760
tgttcacaat acacatgatg ctcaattgaa ttccattcaa tgttacagtt tagataaaca 2820
gtttgtagat aaactcacia tgtatcattt ctttttattt ttgaccaa cagcttctca 2880
tctgttattc agaataattc ctgatggca ggatatccat cccaattggg ggaaggggag 2940
aatttgaaga aaacctagac cacatacata ttgccattg ggaacaaaag tctaaaatga 3000
tgttgttcac atcttctcta ctagtctct ccccgctcca aagaacctg gtatatgtgc 3060
ctcattttac agagagagga aagcaggaac tgagcatccc ttacttgcca tctcaaccc 3120
aaaatttgca tcattgctca gctctgcct tctcatatga cagttacaag tcaaggcttc 3180
caaagtccct ctgtcatggt tgggttcaat agtttataca gatgacttca tgtcttcata 3240
tctaattgtc tatatagatt aatattaaac aatgttattt cttaaccac attttaaatt 3300
aatttaaaaa tccattaatt gtgtctataa aatgcagaca gagtgtgag acacaatata 3360
agcctgatga tctgaattg aaactcacac ccaccatg gagaatcaac ttccaaaaat 3420
tttctatta ctccacact tacaccattg tacaacaca ataataatga acaaatgaa 3480
atgaaataaa aaattaagtc tctgtaggta atgctactgt gcagcaaaag taaaaatggc 3540
agcttaagct tgctttatgg ttacaactta ccatctcca ttaattataa ggacttcaat 3600
catggcagaa ctatgtgtt attgtctcag tgtaacctaa ccaggtgttc cagatgttct 3660
taatgtggac acctaaacta ttgatattt gggttaagat ctttccctct ttcagaagaa 3720
acctcaggac agagggaaac ttgtctttta attttgagtc tgtagacttt ttccatttca 3780
aatatacatg aaacaagtga tgaagaaaat taatcaaaag gtgggaattg caatgatatt 3840

aggttcaata ttaagcttca atattatcat ggaatcgctt gttatacaact gagtgtttgg 3900
 caataagga tttttagaag aaggagtttt tattctcaac aggttcetta agtttagctc 3960
 aaataaatct aagcaatcca ctctagaatt aatagtttc c 4001

<210> 2

<211> 1001

<212> DNA

<213> 人工序列

<220>

<223> 合成聚核苷酸

<400> 2

taccctttat gatcttaca aacttgacag taagatcata ttgctaaaga aaccacatat 60
 ttgaatcagg gaacatgggtg atatctagtt gttcttcaac tggaaacttc atgctttctg 120
 cccagcattc atgttctgg aaagagcaat gtacactacc agttagataa ttaaatcatt 180
 aatcttatca agatgtggat cctataagtt acaataaaaa ttagcctgat aagatatccc 240
 caccagaaga atattcacat aatgctatg ggagcaacaa gctatcttct aaattagctt 300
 taatcctatt ctacaagaga gaatccatat ctagaatagt tatagggatc aagaacctat 360
 ggcttgattg gtcataggcc caatgggaga tctaatatt attgttctac aaaatgaaa 420
 taactcctaa tgacttggtg ctgcagtaat aagttagtat gttgctcaac tctacaaga 480
 gaagttttgt cttacaataa atggcaatta aagcagcccc acaagattta tatcattacc 540
 atctcctcat ggctatgca tctagaagct aggaaacaaa gaggacctc agagagacat 600
 acatgggtccc cctggagaag gggaggggg caagacctc aaagctaatt gggagcatgg 660
 gggagggggag agggagttag aagaaagaga aggggataaa agggaggaga ggaggacaag 720
 agagagaagg aagatctagt caagagaaga tagaggagag caagaaaaga gataccatag 780
 tagaggggagc cttgtatgtt taaatagaaa actggcacta gggaattgtc caaagatcca 840
 caaggtccaa ctaataatct aagcaatagt cgagaggcta ccttaaagc ctttctctga 900
 taatgagatt gatgactacc ttatatacca tcttagagcc ttcatccagt agctgatgga 960
 agcagaagca gacatctaca gctaaacact gagctagttg c 1001

<210> 3

<211> 1001

<212> DNA

<213> 人工序列

<220>

<223> 合成聚核苷酸

<400> 3

caaagtcaag accaggctgg agaaacacac agaaacagca gacctgaaa aatgttgca 60
 catggacccc agactgatag ctgggagtcc agcataggac ttttctagaa acctgaatg 120
 aggatatcag tttggaggtc tggttaatct atggggacac tggtagtgga tcaatattta 180
 tccctagttc atgactggaa tttgggtacc cattccacat ggaggaattc tctgtcagcc 240
 tagacacatg ggggagggtc taggtcctgc tccaaataat gtgttagact ttgaagaact 300

cccttgagaa gactcacct ccctggggag cagaaagggg atgggatgag ggttggtgag 360
 ggacaggaga ggaggggagg gtgagggaac tgggattgac aagtaaataa tgcttgtttc 420
 taatttaaat gaataaagga aaagtaaaag aagaaaagaa aacaggccaa aagattataa 480
 aagacagagg tggtaggtga ctataaagaa acactattat ctaaataaaa atatgtcaga 540
 agcacacatg aacttatagt gtttatgaaa gtatgtataa taactacata atctcaagcc 600
 aagaaaaaaaa tatcatcttt cagtgatgaa ggtgatttta tttctcccag aattaaagcc 660
 aaagacctaa tgaaagtaat tatcttcaaa aggttgaaaa tacatacttt gcaatacaca 720
 gatctgccta gaaatctcat gttcacaata cacatgatgc tcaattgaat tccattcaat 780
 gttacagttt agataaacag tttgtagata aactcacaat gtatcatttc tttttatttt 840
 ttgaccaaac agcttctcat ctgttattca gaataattcc tcgatggcag gatatccatc 900
 ccaattgggg gaaggggaga atttgaagaa aacctagacc acatacatat ttgccattgg 960
 gaaacaaagt ctaaaatgat gttgttcaca tttctctac t 1001

<210> 4

<211> 14931

<212> DNA

<213> 灰仓鼠

<220>

<221> misc_feature

<222> (2176)..(2239)

<223> n是a、c、g、t或核苷酸缺失

<400> 4

catgtacact tatgcaagta tgatatggcc caacacagta ttttacacca atttttatct 60
 ataaaatata catgtacatc aaaatatatt attaataata acatcattat tttttctttc 120
 caagtaataa acacatacac tgaaattttg gttcttgtgg ataatttttaa tgaacagga 180
 aatgcaaatt tatcttagca tgtttacttc actttctttg catagataac cagtaatcac 240
 attgatggat catgtagtga aatgtatttt taggtatcta aggaattttg gcttcgtttt 300
 gtgcttgttg aactgaatt ctattcctaa caacagtggt taaggattct gtctgatttc 360
 ttttaccagt atttgcctat ttgcatttc tttattatc atggctgctg ttctagaaag 420
 tggaaggtag tgtgtcaagt ctgtttaaca tgtttcctg atgatcagtg tcttaacacc 480
 tctctgagta catgttggcc aatgtcgttt ctagacceat ctattcttgc ttgacttate 540
 ctggtacatg cctgccaaga aatttctect cactctttct gtctcttctc tgatttactt 600
 gatgtgtgga tttcacattg atcatatgga aatagaagat acaattttct ttattcacag 660
 tttggaagac tttcaatctc atagatcate attatttttt gctactgttc cctatgctat 720
 ggtgaaattt ccatttgaat aattgettaa acaattaaca agaaagaatc tatttttact 780
 tgcaataact tccatttcag aacatttact aactgttac tatatccaaa aactagtttt 840
 atatatcatg tgagaaatga ctaattcata atttgccat gacatttttt tcagaaacag 900
 aaaaagtgac caatacatac acaatgctat aaatattaag acttcagcaa attaaatatt 960
 tattcatgat atcacataaa attcatttat tatgttttat ttaaagtgtt ttttaaaaca 1020
 gtggtatcac taaatattaa gttagatgtg tttatgtgct taatgaattt atattttaga 1080

atgttataag ttgtatatag tcaaataatgt aataaatttt attttttagg tttttctcat 1140
taaggatatt taattttggg tcccttttcc agagtgactc tagctcatga tgagttgaca 1200
taaaaactaa acagtacaaa atgtacattg cattcagtat tgcacttgat ctttgactg 1260
aagtttgagt cagttcatac atttagtact tgggaagtac attaagctaa ctttcattgc 1320
tctggcaaaa tgctcgataa gataagagtc tattgtggaa agccatggca gcaggaaagt 1380
aagactgctg atgatgttta atccatagtc aagacgcaga aggagatgaa tgctggatc 1440
caacattttt tgctgttcat tttctctaga accctagtc ataaagatgt atgacttgca 1500
ttcaaaatgc gtcccttca gttgttcaac tttctgtaa atatccttc aggcatgtct 1560
agaagattgt ttcgcaaata cttctcaate cattcaagtt gatagtgcag attaatact 1620
gcagaataaa agcctgtaac ttggctcacg tgccaaggaa tatgcacact cctgacacat 1680
caataagtaa atcaaagtgt agcttttgc ttttaacattg ccagacttat gtaatgttct 1740
gcacgttctt cctccatcac tttttattct aatgggtgtt ccttgacatt gaatcacgct 1800
gtggaagctg cttagaatta acattgaaat ctactgatat atttatgatg cagcaattta 1860
gatttactat tttacttaga attttttata attgagagaa tataatattt tcacagttat 1920
ctatctgctg taaatagagg attttaaaaa aaatctctat aacttttttt tacaacacac 1980
agtaaaatta agttaaatt taataaagtc actatgttga tttcaaagtg tgctacgccc 2040
acgggtgta cgcagggtga gcagaagatg ccactaaggt gggctaaggc cgatgggttg 2100
gggtctgcgc tccctggaga tgagccccag gcggttccct ggcaatcagc tgcgatcatg 2160
atgcccgatg agccannnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 2220
nnnnnnnnn nnnnnnnnc tgggtgactt tatggaaaga atttgataga tttcatgatg 2280
tagaagaatt ttattaggct tattttacag gagactaaga ccctgggacc taaagatata 2340
tgggtctctga gaatcaggaa atgggtagag acgtggttga tggatgaga cagattttag 2400
agaactctta gatcatgggc aatgaccgca atctgatgct tagaatagat catctataaa 2460
caattatgct gttcttttc tttctgttgt atgatctgat gatgtagccc ccttgccaag 2520
ttccctgate ccccttgcca agttccctga ttgtaacagt atataagcat tgcttgagag 2580
catattcaac tacattgagt gtgtctgtct gtcatttctc cgccgattcc tgatttctcc 2640
ttgagccttt tccctgttc tccctcggtc ggtggtctcc acgagaggcg gtccgtggca 2700
aaagtgtata aatgttctaa aacatttgaa ctctaaaaca tgcaaatga aaaattaaaa 2760
taaataaaca tgaaaattaa aatatattag ctgctaaaag ttaaacaata ctatataata 2820
ttttgttatt agaattcaaa atcacattag ttggatttaa tttgaacatt gcattcttcc 2880
aataataatt tcaataaaaa aagtttccc atgatagtag aaaataataa catatgtatc 2940
tatctattta tttaactaca catatatagc atttgttca actaaaataa atgaatgagc 3000
aaagcaccta agtaattggg gtctattata tttatgaage caatagttc aaataatta 3060
tcatgcataa ggaggtattg caaatgttaa acctttttg aacagatat tcccagttac 3120
agaaattata atttctaate tttctataa gtagaatgat gataattaat ataggcatt 3180
tgtaaataat gttcagatta aaatattctc tatttcaacta gagaagaatg atattaaatg 3240
tattatattt tatttcccat ttgtttgca ccaactattct atatccctca gcagtttaaa 3300
ttgtttcac catatgtgtg tgtgtttgta tcttaaatat ggcactaaa ttagaataat 3360
ttaatataaa tctttaggag aaaagatatt gaattatttt atgttgatag gaaaatatct 3420

ttttaattgtc caagaatact ttttcttcta ttttaggact gatcagacce aggactaata 3480
 ttttataatgt actaattcta tgtacccaaa tatgttatta tctcatgaat tctgtctcaa 3540
 tattgaggta ataaaaatag tccatcatga actttaaaat taaaataatg attaattaat 3600
 ttttattcat attttgttg tatgaatggg tatacatcac atgtgtgcct ggtgactgtg 3660
 aatgtcagga gaaggtatga aagccactgg aattggaata agagataata tttgagatgt 3720
 tatgtgggtg ctgagaatta gacgcaagcc atcttcaaga atagccagca tactatacca 3780
 ctgagtaatc cattcatccc tcaataatta tctttgtaga cagtaaataat atttctaac 3840
 tataaatgac cagaaaaatt aatgtattat taatgaagac attcatctca tgtgacacac 3900
 ttcacctgtc taaatcagta acactctctc cactaattaa gattttctaa gtgcatgaca 3960
 cttactattt ctaaagctgt ccaatggggg ccagteccca gtcagcacc agtgagataa 4020
 tccatgaatg catttatac ttaggaaaaa ttcttatcta tgtagtattt agaacatttt 4080
 catgtgaggg gataaacaag gaagcacaga tgctttctga tagaaacttt ctctttaatt 4140
 catctagaaa aaaaaaacct ctcaggaaaa tctctcttgc tctctceca atgctctatt 4200
 cagcatcttc tccctactta attctagatc tttttcteta tgectcttg ctgctgcct 4260
 gctggctctg ctctatgect ccccatgtea cttttcttg ctatctacc gttaccttct 4320
 ctgectcact ctctgecttc ttctctgett ctccatggc caggctctgg acaattatag 4380
 ttatatgtta cattctcata acacatgata tgcacatag tttctctcag gctagggata 4440
 tcacaatgac tggccaatga gcaagtggcc ttgcatgtag ctctaagttg gtgatggttc 4500
 ccagacagta agtagccatt tggttgaaat ttgaggttg gtagtacatg aagactgaat 4560
 tttcttcaaa ctctggcctt gaaatagtaa aacaacacct atgaaaatga cgacctgtat 4620
 ttgtctttag aggcaaccac atattgtctg caggcctgc tttgaatttg ctctgaagtt 4680
 agcttgtttg tgtaaaagga agaatcctat atcagcctga gaaatgtaa atatcctagc 4740
 atttcaagtc atcaaaatta tatggagagt ataaatcacc cttctgacta ttcatagtca 4800
 tatttgtgtc caccaagtat aaaacacact accaaagggc tgtggaaaa atcgccataa 4860
 ctgttcttat tagggaggca tagcagtggg acctgaggaa gttacagcaa caaccagtca 4920
 tccagtcaat aacccatgg ctttgccact tggaggtacc caataatggt tggctttgcc 4980
 gagtaggact ccaacaaatt cagagggtca atttttaa atgctggtgtc actgctgaac 5040
 agtcccattg ccctctgcat aattccaca tggaaagctt ttactactga ttgccaatca 5100
 ttaaacagcc tactcagcat aaacaggtat gatattatc tgcattttgt tacattacta 5160
 gatgaattcc tatttcttcc tacaatagtg gaactgaaaa aagatacaca atcactactac 5220
 ccctctacta atcttatgac ttatatcatt tcaattttca gaccataatg caaactattg 5280
 accaaaaacat gtgaagatga aaaatagaaa tgtagaataa tattacatat aaaaagaaaa 5340
 ggcggactta ttttgtttta tttcttagca tgcatagcaa tacatgattt gaggtttata 5400
 taataaaggg acaataaate ttcaagaaac ttaccctac tgaattaaa tattaagaa 5460
 ggtcacacat ttactcaaat atattagact actgggcaaa tagacatgaa aagtagagtt 5520
 aatattgagg taggecttct gtgaaatgtc taaggaaatt atgtttcata cagtgtgtaa 5580
 ccaagtggga atcatatcag aaagcagtca aaagcttata ttacaagtaa cagatgcttg 5640
 gttatatgac ctcccagagc ttgactgtct atacacaaaa agtgggtgta ataaaactgt 5700
 aatttgggct atgtttttt aaatggcttc accaacatga aaggaaggga atgagcatgt 5760

catggatgct tagagattat gcttcagca agaagaattg agctttgget cttattacag 5820
aaacatgaca aggtgtgagt tttatttatt agaaattata taatatttta agctggggac 5880
taaaaatttt attgaaacaa acaggcaagg gataggcatg tactagaagc aaaaatagga 5940
tgtcaatgct gtaatgttat tttttggacc aaaatagtat ttcctataga aatgacaatg 6000
atccttaggtt attattcttc ataaagatga caagttcaca agatateccta gttcattaaa 6060
atcgttttag tcatttaata gagtgtctgtg atagattaca caaaggaaag cacttacgat 6120
gagaaataat gatatccaca attattttct taattcttag aacatttcta ttgttatatc 6180
tcaatctcag aagccactta ttgctttatt attgaaacat atgaaattgt aagttatata 6240
ttgtctatgg tgacatttca aagaacatgt gacgtacagt gtagcacaga taaagaacat 6300
aactgcagct gaatcagtaa ctaaacttac atacattaaa tctgcatgtg tggcaacagt 6360
gtgtgcacta ccaaaggatg tactaatgct cacgacactc ccctatgtec ccctttgttc 6420
atcattacat cataggtcta ttttgtttgc ttttgaaatc tagaccaagt cttttgtgtc 6480
tttccaagca cagagctcat taatttacct catagacttg ttaaacttct tctggttcat 6540
caattgaata gaaatactca ctactaatta tgtgagacc tgccagtacc atagcacatg 6600
gataattttt acataaaaaca tgcatacaag taagattatt cagactgaac atgaatttta 6660
gagaaatcag gaaggagtat atgggagtgg ttggagttag actagagaaa tgtaattaaa 6720
ctataatctc aatacaaaaga tctactaagc aaaaaacatg aacattgtc attcaagtga 6780
aacatcagtc ttcaaatgg aaagatattt ttactaggaa aatgtctggt agatggttat 6840
tatctagaaa acacaaaaat tagaaaacgg taaactttaa taaaaagaat aatacaatga 6900
gactacatga aaagtctta actaatgaaa caaatatctt gaaacttttt tcttaaaagt 6960
ttaatatcaa taaccatcat ggaaattcaa attaaaacta tttacatatt acccctgaaa 7020
taataactaa tacccaataa aaataatata aacaaaaaat ggcaatgcat gccatcatgg 7080
atgtgggaga gagaatgttc attgcagttc tgaatggata ctggtgccac cacggtgaaa 7140
atctctgtat aggtccttc aaaagctgaa aatagacata tcacaagacc tgccacacat 7200
ttttcaagca aatacccaaa ggactctacc tgactgcaga gacactttct cataaaatat 7260
tattgttgat ctattcataa tatctggaaa atagaaacag ccaagatgcc catcaactga 7320
ttaatagatg ataaaattat tgtacatttc agtghtaatat tattcagttt ttaagaaaaa 7380
tgaaattatg taataagcat gtaaatggat atatcttgaa acaaccattc ccattatat 7440
tacctaaaca ttgaaagtcc aaaatcatat gatcttttta gtggatctac taatcttttg 7500
ctatatgtat tttattgaac taccatgga tgtgagataa ttgtaacaa cagcacatgg 7560
gagagcatgg gatcattcaa ggaagattag agagaatgca ttttttagga gataatggag 7620
gagcaataga aaggattaaa tgaggttact gatgaaagtg atggttagag aaggcaatat 7680
gaggagggat aactagcact tagggccttt tgaaaaagac atagagaaaa tactattgta 7740
gaaacttcct ataattggtg tatagttata tacaccaaag agctcagatg gagttacct 7800
ataatggaaa tattaactac tttttatcac tgtgataaaa catcctgaac agagcaacat 7860
agattgggaa gcatttactt tggettacag ttctaacggg ataaaaatc atgatgaaag 7920
aatgaatatg tcagcaaaaca gcagtagcaa tggcctgaga agcaggtgag agctcacatc 7980
ttgaagtgta agaattgtag agagagaaca aactgcaaat gaccagaaaa tgcttttgga 8040
tcagagecca taccctctg actgacttct ccagaaatc tgaacaaata aaactcecca 8100

aacagagcca taactgaagg tccagtgtct gagactacta ggggtatttc ttattcaaac 8160
cactacaatg ggggtggggg agcaatcctc caagtaggca ctacacacag acaataaaaa 8220
actctagtaa ctggaatgga ttgacttatt tgaattactt gccagtggag ctacatagag 8280
cacaattatt gtatttaa ataccctttat gatcttacia aacttgacag taagatcata 8340
ttgctaaaga aaccacatat ttgaatcagg gaacatgggtg atatctagtt gttcttcaac 8400
tggaaacttc atgctttctg cccagcattc atgttgctgg aaagagcaat gtacactacc 8460
agtgtagaaa ttaaatcatc aatcttatca agatgtggat cctataagtt acaataaaaa 8520
ttagcctgat aagatatccc caccagaaga atattcacat aatgctatg ggagcaacaa 8580
gctattttct aaattagctt taatcctatt ctacaagaga gaatccatat ctagaatagt 8640
tatagggatc aagaacccat ggcttgattg gtcataggcc caatgggaga tcctaattatt 8700
attgttctac aaaaatgaaaa taactcctaa tgacttggtg ctgcagtaat aagttagtat 8760
gttgctcaac tctacaaga gaagttttgt cttacaataa atggcaatta aagcagcccc 8820
acaagattta tatcataacc atctctcat ggctatgca tctagaagct aggaacacaa 8880
gaggacccta agagagacat acatggctcc cctggagaag gggaggggg caagacctcc 8940
aaagctaatt gggagcatgg gggaggggag agggagttag aagaaagaga aggggataaa 9000
aggagggaga ggaggacaag agagagaagg aagatctagt caagagaaga tagaggagag 9060
caagaaaaga gataccatag tagagggagc cttgtatggt taaatagaaa actggcacta 9120
gggaattgtc caaagatcca caaggtccaa ctaataatct aagcaatagt cgagaggcta 9180
cctaaaagc ctttctctga taatgagatt gatgactacc ttatatacca tcctagagcc 9240
ttcatccagt agctgatgga agcagaagca gacatctaca gctaaacact gagctagtgtg 9300
cagacagga ggagtgatga gcaaagtcaa gaccaggctg gagaaacaca cagaaacagc 9360
agacctgaaa aaaaatgttg acatggacce cagactgata gctgggagtc cagcatagga 9420
cttttctaga aaccctgaat gaggatatca gtttgagggt ctgggtaate tatggggaca 9480
ctggtagtgg atcaatattt atccctagtt catgactgga atttgggtac ccattccaca 9540
tggaggaatt ctctgtcagc ctacacacat gggggagggt ctaggctctg ctcaaataa 9600
tgtgttagac tttgaagaac tcccttgaga agactcacc tccctgggga gcagaaagg 9660
gatgggatga gggttggtga gggacaggag aggggggag ggtgagggaa ctgggattga 9720
caagtaaatg atgcttgttt ctaatttaaa tgaataaagg aaaagtaaaa gaagaaaaga 9780
aaacaggcca aaagattata aaagacagag gtggtgggtg actataaaga aacactatta 9840
tctaaataaa aatatgtcag aagcacacat gaacttatag tgtttatgaa agtatgtata 9900
ataactacat aatctcaage caagaaaaaa atatcatctt tcagtgatga aggtgatttt 9960
atttctccca gaattaaage caaagacctc atgaaagtaa ttatcttcaa aaggttgaaa 10020
atacatactt tgcaatacac agatctgect agaaatctca tgttcacaat acacatgatg 10080
ctcaattgaa ttccattcaa tgttacagtt tagataaaca gttttagat aaactcaca 10140
tgtatcattt cttttattt ttgacaaa cagcttctca tctgttattc agaataatte 10200
ctcgatggca ggatatccat cccaattggg ggaaggggag aatttgaaga aaacctagac 10260
cacatacata ttgccattg ggaaacaaa tctaaaatga tgttgttcac atcttctcta 10320
ctagctctct ccccgctcca aagaacctg gtatatgtgc ctcatcttac agagagagga 10380
aagcaggaac tgagcatccc ttacttgcca tctcaacce aaaatttgca tcattgctca 10440

gctctgccct tctcatatga cagttacaag tcaaggettc caaagtcct ctgtcatggt 10500
tggtgtcaat agtttataca gatgacttca tgtcttcata tctaatgtct tatatagatt 10560
aatattaaac aatgttattt ctctaaccac attttaaatt aatttaaaaa tccattaatt 10620
gtgtctataa aatgcagaca gagtgctgag acacaatata agcctgatga tctgaatttg 10680
aaactcacac ccaccacatg gagaatcaac ttccaaaaat tttcctatta cttccacact 10740
tacaccattg tacaacaca ataataatga acaaaatgaa atgaaataaa aaattaagtc 10800
tctgtaggta atgctactgt gcagcaaaag taaaaatggc agcttaagct tgctttatgg 10860
ttacacttta ccatcttcca ttaattataa ggacttcaat catggcagaa ctatgctgtt 10920
attgtctcag tgtaacctaa ccaggtgttc cagatgttct taatgtggac acctaaacta 10980
tttgatattt gggtaagat ctttccctct ttcagaagaa acctcaggac agagggaatc 11040
ttgtctttta attttgagtc tgtagacttt ttccatttca aatatacatg aaacaagtga 11100
tgaagaaaaat taatcaaaaag gtgggaattg caatgatatt aggttcaata ttaagcttca 11160
atattatcat ggaatgcct gttatacaact gagtgtttgg caataaggga tttttagaag 11220
aaggagtttt tattctcaac aggttcetta agtttagctc aaataaatct aagcaatcca 11280
ctctagaatt aaatagtttc ctaagggcac agctatgaat agagctcaat ttacatataa 11340
aattttgttc accatttatg tcattccagt tttcattagt acaaggaaaa taaaaatat 11400
ttagatgtca atatcaagt aatagttcat ctctttttt aatataatc acctaaatca 11460
ccattttctc agaaaaatct ggccgaagt tctgtctgga acttcaacat gaaaaatag 11520
cacagcttgc tattataaat cctagttgat ttttaagatt catgtctggt gtctgactca 11580
gaggggcccag aggctagaca aatatttttt gaatcttcat tgtgaagatt tttaatgatt 11640
attttaatat aaataacaaa gatgatggat aatgtaactt tgtacagttc atagacgctg 11700
aactactttg tgcttaaaat gttagtccc tatcataaat gataggtgat aagtgtatgt 11760
ttaatacttt ccctctgagc tatattcatg tactagagaa ttatttttaa catgaaaaga 11820
ctgtgtttat agtctcagct cctgagaact ggtccaacct taggcaggtg aatgccagga 11880
gcaacgtttt tcttctacag aggatgcttt gctgccaagc aacctggttg tgtggaaatg 11940
ttcctttttt aatcaagttt aaaggtctt catcatgctg ttgctccaca tattttcagg 12000
ttagagcttg gtccttgag tattatcttt taccagaaaa ttcatagtat tctttcaata 12060
actaacaact aaacttttcg ataaaaaga attggaattt caatttttaa gcctgagtaa 12120
aattcttgtg aatcaggata ttttatttta agtcttatct tttaaaagt tattttattt 12180
tttaaaaaat tataatatac tttcataatt tcctctctc acttttcttt acaaacactt 12240
ctatagatca ccatgtgttt ttttttttac atttatggcc ttttctgtt cattgttatt 12300
acatacaaat agtcttgct atagaagaac accacaattt gttacctgat acaaattat 12360
caacccttaa aacctacaaa ctattgatat tactgaaaag actatactta tagatgtaaa 12420
gatatatgtg tgtgcacata tatagataca catatatgta ggatttttaa ttttagattt 12480
tagacatcaa aattatttat atgactgaga aactagacac tataaatgag cattcagtat 12540
tcaacaccgt gatttttagat attgtcacia tgacagaaaa ttttcttata gaaaatttta 12600
agttttgtga ttgctctgtg cacttagtga agtctcacag aaaaagaatc atagtatttt 12660
tagtttataa taaaaagtac atataattaa aatggttggc acaaaacaac atttgagcat 12720
ttttcctatt tactatcaag tagtatcatt ttgaaataat aatttgacta gtttcaaaaa 12780

tgaaaacaaa atttaaacta aatgcctaata ctagecctgat aacatttttta tgaatgaaat 12840
 tattcaatag tgttatcaat taggggccca aaacttttcc taaaataaaa cttttaattt 12900
 ttttccattt ttatttaaata tagaaacaaa attgttttac atgtaaataca gagtttcctc 12960
 accctcccct tctcccgtc cctcactaac accctacttg tcccatacca tttctgctcc 13020
 ccagggaggg tgaggccttc catggggaaa cttcagagtc tgtctatcct ttcggatagg 13080
 gcctaggccc tcaccattt gtctaggcta aggcacacaa agtttactcc tatgctagtg 13140
 ataagtactg atctactaca agagacacca tagatttctt aggccttctc actgacaccc 13200
 atgttcatgg ggtctggaac aatcatatgc tagtttctta ggtatcagtc tggggacat 13260
 gagctcccc ttgttcaggt caactgttcc tgtgggttcc accaccctgg tcttgactgc 13320
 tttgctcacc actcctccct ttctgtaact gggttccagt acaattccgt gtttagctgt 13380
 ggggtgtctac ttctacttcc atcagcttct gggatggagc ctctaggata gcatacaatt 13440
 agtcatcacc tcattatcag ggaagggcat ttaaagtagc ctctccattg ttgcttggat 13500
 tgtagttgg tgctatctt gtagatctct ggacatttcc ctagtgccag atatctctt 13560
 aaacctacaa gactacctt attatggtat ctcttttctt gctctctctt attcttccag 13620
 acaaaatctt cctgctccct tatattttcc tctcccctcc tcttctccc ttctcattct 13680
 cctagatcca tcttcccctc ccccatgctc ccaagagaga tgttgctcag gagatcttgt 13740
 tccttaacc ttttcttggg gatctgtctc tcttaggggt gtccttgttt cctagcttct 13800
 ctggaagtgt ggattgtaag ctggtaataca tttgctccat gtctaaaatc catatatgag 13860
 tgatgtttgt ctttttgta ctgggttacc tcaactcaaaa tggtttcttc catatgtctg 13920
 tggatttcaa tagcacaac aacatacagt atcttggggc aacactaac aaacaagtga 13980
 aagaccagta tagcaagaac tttgagtta aagaaagaaa ttaaagaaga taccagaaaa 14040
 tggaaagatc tcccatgctc tttgatagc agaatacaca tagtaaaaat ggcaatcttg 14100
 ccaaaatcca tctacagact caatgcaatc cccattaaat accagcacac ttcttcacag 14160
 acctgaaaga ataactta actttatatg gagaacaaa agaccagga taggccaac 14220
 aacctgtac aatgaaggca cttccagagg catcccctc cctgacttca agctctatta 14280
 tagagtaata atcctgaaaa cagcttggta atggcacaac aatagacagg tagaccaatg 14340
 gaattgagtt gaaaaccctg atattaacc acatatctat gaacacctga ctttgacaaa 14400
 gaagctaagg ttatacaatg taagaagaa agcatcttca acaaatctg ctggcataac 14460
 tggatgctgg catgtagaag actgcagata gatccatgct taatgcatg cacaaaactt 14520
 aagtccaaat ggatcaaaaa cctcaacata aatccagcca cactgaacct catagaagag 14580
 aaagtgggaa gtatccttga ataaattggt acaggagacc acatcttga cttaacacca 14640
 gtagcacaga caatcagatc aataatcaat aatgggacc tctgaaact gagaagcttc 14700
 tgtaaggcaa tggataagtc aacaggacaa aatggcagcc cacggaatgg gaaaagatat 14760
 tcaccaatcc tatatctgac agagggetgc tctctatttg caaagaacac aataagctag 14820
 tttttaaacc accaattaat ccgattataa agttgggtag agaactaat aaagaattgt 14880
 taacagagca atctaacttg gcagaaaagac acataagaaa gtgctacca t 14931

<210> 5

<211> 20

<212> DNA

<213> 人工序列

<220>

<223> 合成聚核苷酸

<400> 5

tgagctagtt gcagacaggg 20

<210> 6

<211> 79

<212> RNA

<213> 人工序列

<220>

<223> 合成聚核苷酸

<400> 6

guuuuagagc uagaauagca aguuuuuuua aggcugagucc guuaucaacu ugaaaaagug 60

gcaccgaguc ggugcuuuu 79

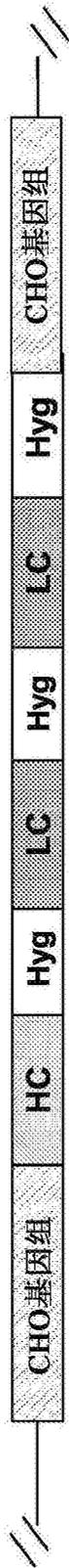


图1A. 在SEQ ID NO:1基因座处的整合

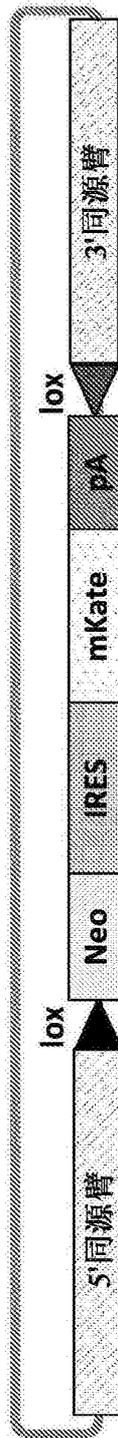


图1B. 荧光标记供体载体:

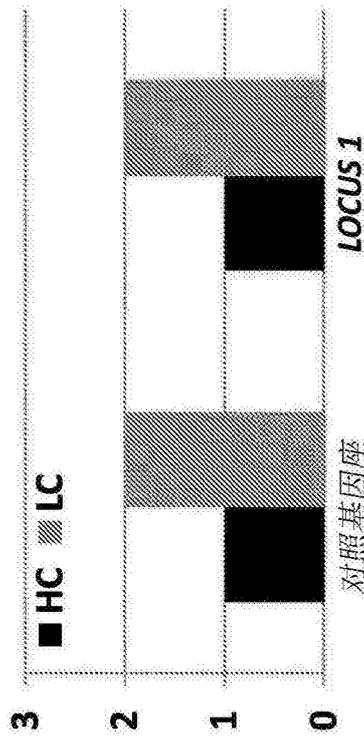


图2A.通过qPCR的基因拷贝数

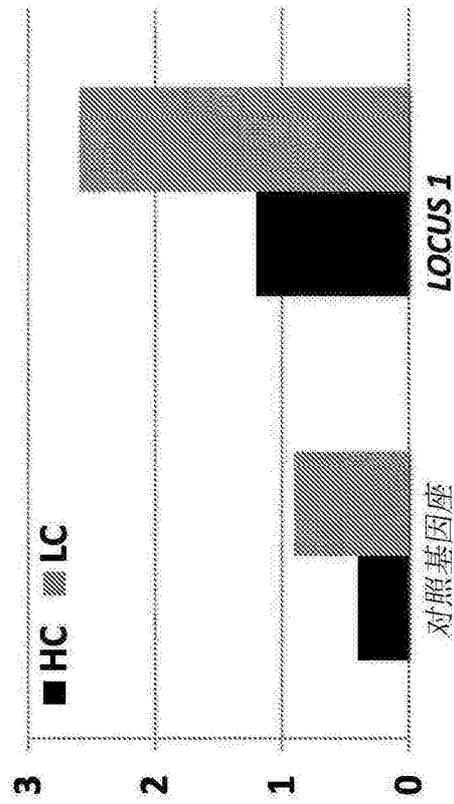


图2B.通过qPCR的相对mRNA水平

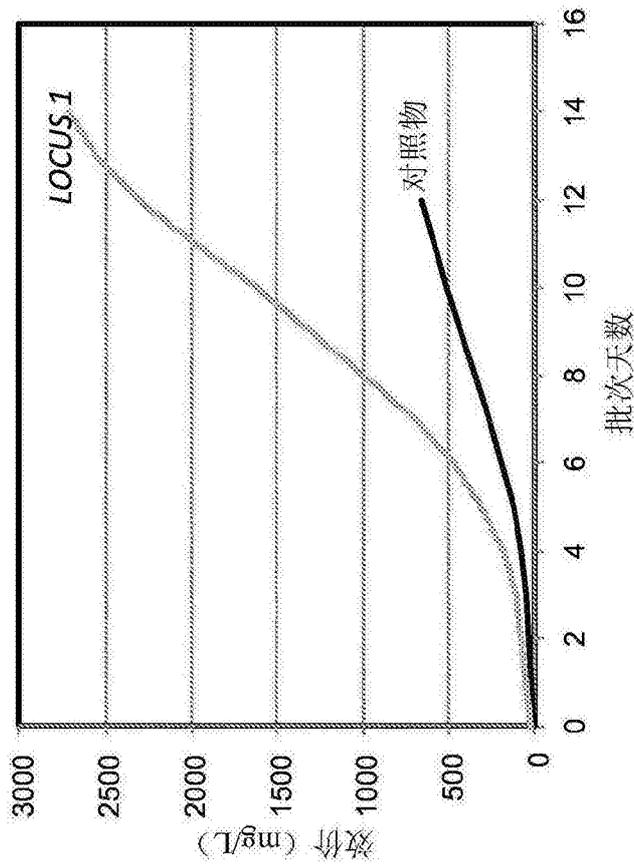


图2C. 对照基因座对比LOCUS 1生产线的生物反应器效价

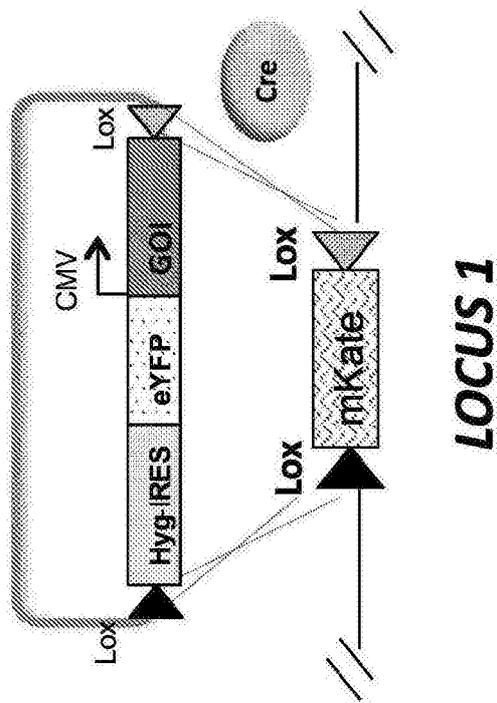


图3A

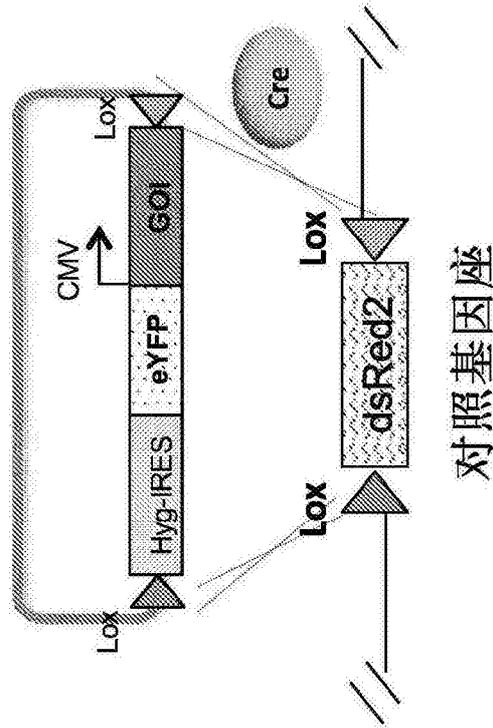


图3B

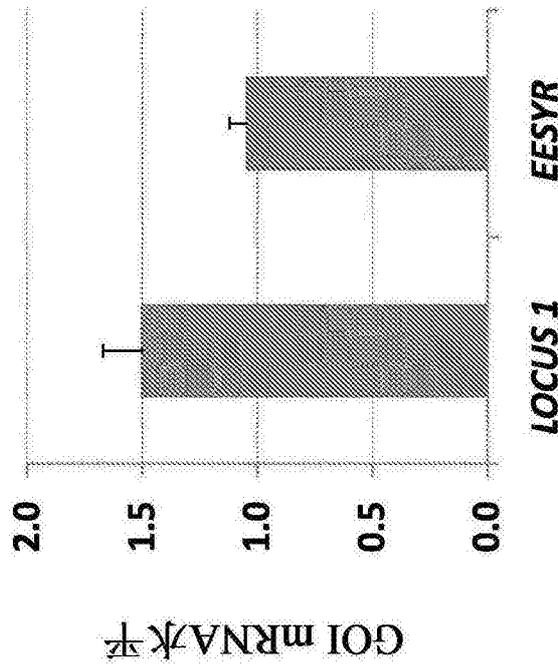


图4. 经工程改造的LOCUS 1支持RMCE和高水平转录