



(21) 申请号 202080097601.2

(51) Int. Cl.

(22) 申请日 2020.02.28

C12N 15/90 (2006.01)

(85) PCT国际申请进入国家阶段日
2022.08.25

(86) PCT国际申请的申请数据
PCT/CN2020/077211 2020.02.28

(87) PCT国际申请的公布数据
W02021/168799 EN 2021.09.02

(71) 申请人 辉大(上海)生物科技有限公司
地址 200131 上海市浦东新区自由贸易试
验区巴圣路160号7号楼1单元1002室

(72) 发明人 杨辉 胥春龙 周英思 肖庆全

(74) 专利代理机构 北京市金杜律师事务所
11256

专利代理师 陈文平 徐志明

权利要求书5页 说明书67页
序列表41页 附图11页

(54) 发明名称

VI-E型和VI-F型CRISPR-Cas系统及其用途

(57) 摘要

本申请提供了一种新型CRISPR/Cas组合物及其用于靶向核酸的用途。特别地,提供了非天然存在的或工程化的靶向RNA的系统,所述系统包含新型靶向RNA的Cas13e或Cas13f效应蛋白,以及至少一种靶向核酸组分,如指导RNA(gRNA)或crRNA。所述新型Cas效应蛋白是已知Cas效应蛋白中最小的一种,大小为约800个氨基酸,并因此特别适合使用小容量载体(如AAV载体)进行递送。

1. 一种成簇的规律间隔的短回文重复序列 (CRISPR) -Cas 复合物, 所述复合物包含:
 - (1) RNA 指导序列, 其包含能够与靶 RNA 杂交的间隔序列以及在所述间隔序列的 3' 的同向重复 (DR) 序列; 和
 - (2) CRISPR 相关蛋白 (Cas) 或所述 Cas 的衍生物或功能性片段, 所述 CRISPR 相关蛋白具有 SEQ ID NO:1-7 中的任一者的氨基酸序列;其中所述 Cas、所述 Cas 的衍生物和功能性片段能够 (i) 与所述 RNA 指导序列结合, 并且 (ii) 靶向所述靶 RNA,
条件是当所述复合物包含 SEQ ID NO:1-7 中的任一者的 Cas 时, 所述间隔序列与天然存在的噬菌体核酸不是 100% 互补。
2. 如权利要求 1 所述的 CRISPR-Cas 复合物, 其中所述 DR 序列具有与 SEQ ID NO:8-14 中的任一者的二级结构基本相同的二级结构。
3. 如权利要求 1 所述的 CRISPR-Cas 复合物, 其中所述 DR 序列由 SEQ ID NO:8-14 中的任一者编码。
4. 如权利要求 1、2 或 3 所述的 CRISPR-Cas 复合物, 其中所述靶 RNA 由真核 DNA 编码。
5. 如权利要求 4 所述的 CRISPR-Cas 复合物, 其中所述真核 DNA 是非人哺乳动物 DNA、非人灵长类动物 DNA、人 DNA、植物 DNA、昆虫 DNA、鸟 DNA、爬行动物 DNA、啮齿动物 DNA、鱼 DNA、蠕虫/线虫 DNA、酵母 DNA。
6. 如权利要求 1-5 中任一项所述的 CRISPR-Cas 复合物, 其中所述靶 RNA 是 mRNA。
7. 如权利要求 1-6 中任一项所述的 CRISPR-Cas 复合物, 其中所述间隔序列在 15-60 个核苷酸之间、在 25-50 个核苷酸之间、或为约 30 个核苷酸。
8. 如权利要求 1-7 中任一项所述的 CRISPR-Cas 复合物, 其中所述间隔序列与所述靶 RNA 是 90%-100% 互补。
9. 如权利要求 1-8 中任一项所述的 CRISPR-Cas 复合物, 其中所述衍生物包含 SEQ ID NO:1-7 中的任一者的一个或多个残基的保守氨基酸取代。
10. 如权利要求 9 所述的 CRISPR-Cas 复合物, 其中所述衍生物仅包含保守氨基酸取代。
11. 如权利要求 1-10 中任一项所述的 CRISPR-Cas 复合物, 其中所述衍生物在 HEPN 结构域或 RXXXXH 基序中具有与 SEQ ID NO:1-7 中的任一者的野生型 Cas 相同的序列。
12. 如权利要求 1-9 中任一项所述的 CRISPR-Cas 复合物, 其中所述衍生物能够与杂交至所述靶 RNA 的 RNA 指导序列结合, 但由于所述 Cas 的 RNA 酶催化位点突变而不具有 RNA 酶催化活性。
13. 如权利要求 12 所述的 CRISPR-Cas 复合物, 其中所述衍生物具有不超过 210 个残基的 N-末端缺失、和/或不超过 180 个残基的 C-末端缺失。
14. 如权利要求 13 所述的 CRISPR-Cas 复合物, 其中所述衍生物具有约 180 个残基的 N-末端缺失、和/或约 150 个残基的 C-末端缺失。
15. 如权利要求 12-14 中任一项所述的 CRISPR-Cas 复合物, 其中所述衍生物进一步包含 RNA 碱基编辑结构域。
16. 如权利要求 15 所述的 CRISPR-Cas 复合物, 其中所述 RNA 碱基编辑结构域是腺苷脱氨酶, 如双链 RNA 特异性腺苷脱氨酶 (例如, ADAR1 或 ADAR2); 载脂蛋白 B mRNA 编辑酶, 催化多肽样 (APOBEC); 或激活诱导的胞苷脱氨酶 (AID)。

17. 如权利要求16所述的CRISPR-Cas复合物,其中所述ADAR2具有E488Q/T375G双突变或者是ADAR2DD。

18. 如权利要求15-17中任一项所述的CRISPR-Cas复合物,其中将所述碱基编辑结构域与RNA结合结构域,如MS2进一步融合。

19. 如权利要求12-14中任一项所述的CRISPR-Cas复合物,其中所述衍生物进一步包含RNA甲基转移酶、RNA去甲基化酶、RNA剪接修饰子、定位因子或翻译修饰因子。

20. 如权利要求1-19中任一项所述的CRISPR-Cas复合物,其中所述Cas、所述衍生物或所述功能性片段包含核定位信号(NLS)序列或核输出信号(NES)。

21. 如权利要求1-20中任一项所述的CRISPR-Cas复合物,其中靶向所述靶RNA导致对所述靶RNA进行修饰。

22. 如权利要求21所述的CRISPR-Cas复合物,其中所述靶RNA修饰是切割所述靶RNA。

23. 如权利要求21所述的CRISPR-Cas复合物,其中所述靶RNA修饰是腺苷(A)脱氨基为肌苷(I)。

24. 如权利要求1-23中任一项所述的CRISPR-Cas复合物,所述CRISPR-Cas复合物进一步包含靶RNA,所述靶RNA包含能够与所述间隔序列杂交的序列。

25. 一种融合蛋白,所述融合蛋白包含(1)如权利要求1-24中任一项所述的Cas、其衍生物或其功能性片段,和(2)异源功能性结构域。

26. 如权利要求25所述的融合蛋白,其中所述异源功能性结构域包含:核定位信号(NLS)、报告蛋白或检测标记(例如,GST、HRP、CAT、GFP、HcRed、DsRed、CFP、YFP、BFP)、定位信号、蛋白靶向部分、DNA结合结构域(例如,MBP、Lex A DBD、Gal4 DBD)、表位标签(例如,His、myc、V5、FLAG、HA、VSV-G、Trx等)、转录激活结构域(例如,VP64或VPR)、转录抑制结构域(例如,KRAB部分或SID部分)、核酸酶(例如,FokI)、脱氨基结构域(例如,ADAR1、ADAR2、APOBEC、AID或TAD)、甲基化酶、去甲基化酶、转录释放因子、HDAC、具有ssRNA切割活性的多肽、具有dsRNA切割活性的多肽、具有ssDNA切割活性的多肽、具有dsDNA切割活性的多肽、DNA或RNA连接酶、或其任何组合。

27. 如权利要求25或26所述的融合蛋白,其中将所述异源功能性结构域在所述融合蛋白的N-末端、C-末端或内部融合。

28. 一种缀合物,所述缀合物包含与(2)缀合的(1):(1)如权利要求1-24中任一项所述的Cas、其衍生物或其功能性片段,(2)异源功能性部分。

29. 如权利要求28所述的缀合物,其中所述异源功能性部分包含:核定位信号(NLS)、报告蛋白或检测标记(例如,GST、HRP、CAT、GFP、HcRed、DsRed、CFP、YFP、BFP)、定位信号、蛋白靶向部分、DNA结合结构域(例如,MBP、Lex A DBD、Gal4 DBD)、表位标签(例如,His、myc、V5、FLAG、HA、VSV-G、Trx等)、转录激活结构域(例如,VP64或VPR)、转录抑制结构域(例如,KRAB部分或SID部分)、核酸酶(例如,FokI)、脱氨基结构域(例如,ADAR1、ADAR2、APOBEC、AID或TAD)、甲基化酶、去甲基化酶、转录释放因子、HDAC、具有ssRNA切割活性的多肽、具有dsRNA切割活性的多肽、具有ssDNA切割活性的多肽、具有dsDNA切割活性的多肽、DNA或RNA连接酶、或其任何组合。

30. 如权利要求28或29所述的缀合物,其中所述异源功能性部分相对于所述Cas、其衍生物或其功能性片段在N-末端、C-末端或内部缀合。

31. 一种多核苷酸, 所述多核苷酸编码SEQ ID NO:1-7中的任一者、或其衍生物、或其功能性片段、或其融合蛋白, 条件是所述多核苷酸不是SEQ ID NO:15-21中的任一者。

32. 如权利要求31所述的多核苷酸, 所述多核苷酸经密码子优化以在细胞中表达。

33. 如权利要求32所述的多核苷酸, 其中所述细胞是真核细胞。

34. 一种非天然存在的多核苷酸, 所述非天然存在的多核苷酸包含SEQ ID NO:8-14中的任一者的衍生物, 其中所述衍生物(i)与SEQ ID NO:8-14中的任一者相比具有一个或多个(例如, 1、2、3、4、5、6、7、8、9或10个)核苷酸添加、缺失或取代; (ii)与SEQ ID NO:8-14中的任一者具有至少20%、30%、40%、50%、60%、70%、80%、90%、95%或97%的序列同一性; (iii)在严格条件下与SEQ ID NO:8-14中的任一者或(i)和(ii)中的任一者杂交; 或(iv)是(i)-(iii)中的任一者的互补序列, 条件是所述衍生物不是SEQ ID NO:8-14中的任一者, 并且所述衍生物编码RNA(或者是RNA), 所述RNA保持与由SEQ ID NO:8-14编码的任一RNA基本相同的二级结构。

35. 如权利要求34所述的非天然存在的多核苷酸, 其中所述衍生物用作如权利要求1-24中任一项所述的Cas、其衍生物或其功能性片段中的任一者的DR序列。

36. 一种载体, 所述载体包含如权利要求31-33中任一项所述的多核苷酸。

37. 如权利要求36所述的载体, 其中所述多核苷酸与启动子和任选的增强子可操作地连接。

38. 如权利要求37所述的载体, 其中所述启动子是组成型启动子、诱导型启动子、泛素启动子或组织特异性启动子。

39. 如权利要求36-38中任一项所述的载体, 所述载体是质粒。

40. 如权利要求36-38中任一项所述的载体, 所述载体是逆转录病毒载体、噬菌体载体、腺病毒载体、单纯疱疹病毒(HSV)载体、AAV载体或慢病毒载体。

41. 如权利要求40所述的载体, 其中所述AAV载体是血清型AAV1、AAV2、AAV4、AAV5、AAV6、AAV7、AAVrh74、AAV8、AAV9、AAV10、AAV11、AAV12或AAV13的重组AAV载体。

42. 一种递送系统, 所述递送系统包含(1)递送媒介物, 和(2)如权利要求1-24中任一项所述的CRISPR-Cas复合物、如权利要求25-27中任一项所述的融合蛋白、如权利要求28-30中任一项所述的缀合物、如权利要求31-33中任一项所述的多核苷酸、或如权利要求36-41中任一项所述的载体。

43. 如权利要求42所述的递送系统, 其中所述递送媒介物是纳米颗粒、脂质体、外泌体、微泡或基因枪。

44. 一种细胞或其后代, 所述细胞或其后代包含如权利要求1-24中任一项所述的CRISPR-Cas复合物、如权利要求25-27中任一项所述的融合蛋白、如权利要求28-30中任一项所述的缀合物、如权利要求31-33中任一项所述的多核苷酸、或如权利要求36-41中任一项所述的载体。

45. 如权利要求44所述的细胞或其后代, 所述细胞或其后代是真核细胞(例如, 非人哺乳动物细胞、人细胞或植物细胞)或原核细胞(例如, 细菌细胞)。

46. 一种非人多细胞真核生物, 所述非人多细胞真核生物包含如权利要求44或45所述的细胞。

47. 如权利要求46所述的非人多细胞真核生物, 所述非人多细胞真核生物是针对人遗

传障碍的动物(例如,啮齿动物或灵长类动物)模型。

48.一种修饰靶RNA的方法,所述方法包括使所述靶RNA与如权利要求1-24中任一项所述的CRISPR-Cas复合物接触,其中所述间隔序列与所述靶RNA的至少15个核苷酸互补;其中所述Cas、所述衍生物或所述功能性片段与所述RNA指导序列缔合以形成所述复合物;其中所述复合物与所述靶RNA结合;并且其中在所述复合物与所述靶RNA结合后,所述Cas、所述衍生物或所述功能性片段修饰所述靶RNA。

49.如权利要求48所述的方法,其中通过由所述Cas进行切割来修饰所述靶RNA。

50.如权利要求48所述的方法,其中通过由包含双链RNA特异性腺苷脱氨酶的衍生物进行脱氨基来修饰所述靶RNA。

51.如权利要求48-50中任一项所述的方法,其中所述靶RNA是mRNA、tRNA、rRNA、非编码RNA、lncRNA或核RNA。

52.如权利要求48-51中任一项所述的方法,其中在所述复合物与所述靶RNA结合后,所述Cas、所述衍生物和所述功能性片段不会展现出大量的(或可检测的)附带RNA酶活性。

53.如权利要求48-52中任一项所述的方法,其中所述靶RNA在细胞内。

54.如权利要求53所述的方法,其中所述细胞是癌细胞。

55.如权利要求53所述的方法,其中所述细胞受感染原感染。

56.如权利要求55所述的方法,其中所述感染原是病毒、朊病毒、原生动物、真菌或寄生物。

57.如权利要求53-56中任一项所述的方法,其中所述CRISPR-Cas复合物由以下编码:编码SEQ ID NO:1-7中的任一者或其衍生物或功能性片段的第一多核苷酸,以及包含SEQ ID NO:8-14中的任一者和编码能够与所述靶RNA结合的间隔RNA的序列的第二多核苷酸,其中将所述第一多核苷酸和所述第二多核苷酸引入所述细胞中。

58.如权利要求57所述的方法,其中通过相同的载体将所述第一多核苷酸和所述第二多核苷酸引入所述细胞中。

59.如权利要求53-58中任一项所述的方法,所述方法导致以下中的一项或多项:(i)体外或体内诱导细胞衰老;(ii)体外或体内细胞周期停滞;(iii)体外或体内细胞生长抑制;(iv)体外或体内诱导无反应性;(v)体外或体内诱导细胞凋亡;以及(vi)体外或体内诱导坏死。

60.一种在有需要的受试者中治疗病症或疾病的方法,所述方法包括向所述受试者施用包含如权利要求1-24中任一项所述的CRISPR-Cas复合物或编码所述CRISPR-Cas复合物的多核苷酸的组合物;其中所述间隔序列与以下互补:与所述病症或疾病相关的靶RNA的至少15个核苷酸;其中所述Cas、所述衍生物或所述功能性片段与所述RNA指导序列缔合以形成所述复合物;其中所述复合物与所述靶RNA结合;并且其中在所述复合物与所述靶RNA结合后,所述Cas、所述衍生物或所述功能性片段切割所述靶RNA,从而在所述受试者中治疗所述病症或疾病。

61.如权利要求60所述的方法,其中所述病症或疾病是癌症或感染性疾病。

62.如权利要求61所述的方法,其中所述癌症是威尔姆斯肿瘤、尤因肉瘤、神经内分泌肿瘤、胶质母细胞瘤、神经母细胞瘤、黑素瘤、皮肤癌、乳腺癌、结肠癌、直肠癌、前列腺癌、肝癌、肾癌、胰腺癌、肺癌、胆道癌、子宫颈癌、子宫内膜癌、食道癌、胃癌、头颈癌、甲状腺髓样

癌、卵巢癌、神经胶质瘤、淋巴瘤、白血病、骨髓瘤、急性淋巴细胞白血病、急性骨髓性白血病、慢性淋巴细胞白血病、慢性骨髓性白血病、霍奇金淋巴瘤、非霍奇金淋巴瘤或膀胱癌。

63. 如权利要求60-62中任一项所述的方法,所述方法是体外方法、体内方法或离体方法。

64. 一种细胞或其后代,所述细胞或其后代通过如权利要求48-59中任一项所述的方法获得,其中所述细胞和所述后代包含非天然存在的修饰(例如,在所述细胞/后代的转录的RNA中的非天然存在的修饰)。

65. 一种检测靶RNA的存在的方法,所述方法包括使所述靶RNA与包含如权利要求25-27中任一项所述的融合蛋白、或如权利要求28-30中任一项所述的缀合物、或编码所述融合蛋白的多核苷酸的组合物接触,其中所述融合蛋白或所述缀合物包含可检测标记(例如,可通过荧光、RNA印迹或FISH检测的标记)和能够与所述靶RNA结合的复合间隔序列。

66. 一种真核细胞,所述真核细胞包含成簇的规律间隔的短回文重复序列(CRISPR) - Cas复合物,所述CRISPR-Cas复合物包含:

(1) RNA指导序列,其包含能够与靶RNA杂交的间隔序列以及在所述间隔序列的3'的同向重复(DR)序列;和

(2) CRISPR相关蛋白(Cas)或所述Cas的衍生物或功能性片段,所述CRISPR相关蛋白具有SEQ ID NO:1-7中的任一者的氨基酸序列;

其中所述Cas、所述Cas的衍生物和功能性片段能够(i)与所述RNA指导序列结合,并且(ii)靶向所述靶RNA。

VI-E型和VI-F型CRISPR-Cas系统及其用途

背景技术

[0001] CRISPR (成簇的规律间隔的短回文重复序列) 是在原核生物 (如细菌和古细菌) 的基因组中发现的DNA序列家族。这些序列应理解为源自先前已感染原核生物的噬菌体的DNA片段,并在随后的原核生物感染期间用于检测和破坏相似噬菌体的DNA。

[0002] CRISPR相关系统是一组同源基因或Cas基因,其中一些编码具有解旋酶和核酸酶活性的Cas蛋白。Cas蛋白是利用源自CRISPR序列的RNA (crRNA) 作为指导序列以识别并切割与crRNA互补的多核苷酸特定链 (例如,DNA) 的酶。

[0003] CRISPR-Cas系统共同构成原始的原核“免疫系统”,其赋予对外来致病遗传元件 (如存在于染色体外DNA (例如,质粒) 和噬菌体中的那些) 或由外来DNA编码的外来RNA的抗性或获得性免疫。

[0004] CRISPR/Cas系统似乎是在自然界中广泛存在的针对外来遗传物质的原核防御机制,并且见于大约50%的经测序细菌基因组和近90%的经测序古细菌中。这种原核系统后来经过发展形成称为CRISPR-Cas技术的基础,所述技术广泛用在包括人在内的众多真核生物中,用在包括基础生物学研究、生物技术产品开发和疾病治疗在内的多种应用中。

[0005] 原核CRISPR-Cas系统包括一组极其多样的蛋白效应子、非编码元件以及基因座架构,其中一些实例已经过工程化并适于产生重要的生物技术。

[0006] CRISPR基因座结构已在许多系统中进行了研究。在这些系统中,基因组DNA中的CRISPR阵列典型地包含富含AT的前导序列,随后是由独特的间隔序列隔开的短DR序列。这些CRISPR DR序列的大小范围可以是23-55bp,但典型地在从28至37bp的范围。一些DR序列显示出双向对称性(dyad symmetry),这提示RNA中二级结构如茎环(“发夹”)的形成,而其他序列则呈现为非结构化。不同CRISPR阵列中的间隔子大小典型地是32-38bp (范围为21-72bp)。CRISPR阵列中的重复-间隔序列通常少于50个单位。

[0007] 通常在这样的CRISPR重复-间隔阵列旁发现小簇的cas基因。到目前为止,已将所鉴定的93种cas基因基于其编码的蛋白的序列相似性分为35个家族。所述35个家族中的十一个家族形成所谓的cas核心,所述核心包括Cas1至Cas9的蛋白家族。完整的CRISPR-Cas基因座具有至少一个属于cas核心的基因。

[0008] CRISPR-Cas系统可广义地分为两类——1类系统使用多种Cas蛋白的复合物来降解外来核酸,而2类系统使用单个大Cas蛋白用于相同目的。2类系统的单亚基效应组合物为工程化和应用转化提供了更简单的组分集,并且迄今为止已成为用于基因组工程化和其他方面的新颖强大且可编程技术的发现、工程化和优化的重要来源。

[0009] 1类系统进一步分为I型、III型和IV型;并且2类系统分为II型、V型和VI型。这6种系统类型又分为19种亚型。分类也基于存在的cas基因的互补序列。大多数CRISPR-Cas系统具有Cas1蛋白。许多原核生物含有多个CRISPR-Cas系统,表明这些系统相容并且可共享组分。

[0010] 最早且最佳表征的Cas蛋白之一Cas9是2类II型的原型成员,并起源于化脓性链球菌 (*Streptococcus pyogenes*) (SpCas9)。Cas9是DNA内切核酸酶,其由与靶DNA序列互补的

小crRNA分子和单独的反式激活CRISPR RNA (tracrRNA) 激活。crRNA由负责使蛋白与crRNA结合的同向重复 (DR) 序列和间隔序列组成,所述间隔序列可以经工程化为与任何希望的核酸靶序列互补。通过这种方式,可以将CRISPR系统进行编程以通过修饰crRNA的间隔序列来靶向DNA或RNA靶标。已将crRNA和tracrRNA融合以形成实用性更好的单个指导RNA (sgRNA)。当与Cas9结合时,sgRNA与其靶DNA杂交,并指导Cas9切割靶DNA。也已类似地鉴定并使用来自其他物种的其他Cas9效应蛋白,包括来自嗜热链球菌 (*S. thermophilus*) CRISPR系统的Cas9。这些CRISPR/Cas9系统已广泛用在众多真核生物中,包括面包酵母 (酿酒酵母 (*Saccharomyces cerevisiae*))、条件致病病原体白色念珠菌 (*Candida albicans*)、斑马鱼 (zebrafish) (斑马鱼 (*Danio rerio*))、果蝇 (黑腹果蝇 (*Drosophila melanogaster*))、蚂蚁 (跳镰猛蚁 (*Harpegnathos saltator*) 和毕氏卵角蚁 (*Ooceraea biroi*))、蚊 (埃及伊蚊 (*Aedes aegypti*))、线虫 (秀丽隐杆线虫 (*Caenorhabditis elegans*))、植物、小鼠、猴和人胚胎。

[0011] 最近表征的另一Cas效应蛋白是Cas12a (以前称为Cpf1)。Cas12a与C2c1和C2c3都是属于2类V型Cas蛋白的成员,所述蛋白缺乏HNH核酸酶,但具有RuvC核酸酶活性。Cas12a最初在细菌新凶手弗朗西斯菌 (*Francisella novicida*) 的CRISPR/Cpf1系统中表征。它的原名反映其CRISPR-Cas亚型在普雷沃菌属 (*Prevotella*) 和弗朗西斯菌属 (*Francisella*) 谱系中的普遍性。Cas12a显示出与Cas9的几个关键差异,包括:引起双链DNA的“交错 (staggered)”切割,而不是Cas9产生的“钝 (blunt)”切割,依赖于“富含T”的PAM序列 (它为Cas9提供替代性靶向位点),以及只需要CRISPR RNA (crRNA) 而不需要tracrRNA就可成功靶向。Cas12a的小crRNA比Cas9更适合用于多重基因组编辑,因为与Cas9的sgRNA可以包装在一个载体中的数量相比,它们可以包装在一个载体中的数量更多。此外,Cas12a留下的粘性5'突出端可用于DNA组装,所述组装比传统的限制性酶克隆的靶标特异性大得多。最后,Cas12a在其PAM位点下游切割DNA 18-23个碱基对,这意味着在通过NHEJ系统产生双链断裂 (DSB) 后,在DNA修复后,不会破坏核酸酶识别序列,因此Cas12a能够实现多轮DNA切割,与之相反,Cas9切割后可能实现一轮DNA切割,因为Cas9切割序列仅在PAM位点上游的3个碱基对处,并且NHEJ途径典型地导致插入缺失 (indel) 突变,所述插入缺失突变破坏识别序列,从而防止另外的多轮切割。理论上,重复多轮的DNA切割与发生所希望的基因组编辑的机会增加相关。

[0012] 最近,已经鉴定了几种2类VI型Cas蛋白,包括Cas13 (也称为C2c2)、Cas13b、Cas13c和Cas13d,每一种都是RNA指导的RNA酶 (即,这些Cas蛋白使用它们的crRNA来识别Cas9和Cas12a中的靶RNA序列,而不是靶DNA序列)。总体而言,与传统的RNAi和CRISPRi技术相比,CRISPR/Cas13系统可以实现更高的RNA消化效率,同时与RNAi相比,展现出少得多的脱靶切割。

[0013] 这些目前鉴定的Cas13蛋白的一个缺点是它们的尺寸相对较大。Cas13a、Cas13b和Cas13c中的每一个具有多于1100个氨基酸残基。因此,就算可能,也很难将它们的编码序列 (约3.3kb) 和sgRNA以及任何所需的启动子序列和翻译调节序列包装到某些小容量基因疗法载体中,所述小容量基因疗法载体例如目前最有效且最安全的基于腺相关病毒 (AAV) 的基因疗法载体,其具有约4.7kb的包装容量。虽然迄今为止最小的Cas13蛋白Cas13d仅具有约920个氨基酸 (即,约2.8kb的编码序列),并且理论上可以被包装到AAV载体中,但它在基

于单碱基编辑的基因疗法中具有有限的用途,所述基于单碱基编辑的基因疗法依赖于使用具有单碱基编辑功能的基于Cas13d的融合蛋白,如dCas13d-ADAR2DD(其具有约3.9kb的编码序列)。

[0014] 此外,目前已知的Cas13蛋白/系统在通过基于crRNA的靶序列识别激活后全部都具有非特异性/附带(collateral)RNA酶活性。这种活性在Cas13a和Cas13b中特别强烈,并且仍然以可检测的方式存在于Cas13d中。虽然这种特性可以有利地用于核酸检测方法中,但这些Cas13蛋白的非特异性/附带RNA酶活性对基因疗法的使用构成巨大的潜在危险。

发明内容

[0015] 本发明的一个方面提供成簇的规律间隔的短回文重复序列(CRISPR)-Cas复合物,所述复合物包含:(1)RNA指导序列,其包含能够与靶RNA杂交的间隔序列以及在所述间隔序列的3'的同向重复(DR)序列;和(2)CRISPR相关蛋白(Cas)或所述Cas的衍生物或功能性片段,所述CRISPR相关蛋白具有SEQ ID NO:1-7中的任一者的氨基酸序列;其中所述Cas、所述Cas的衍生物和功能性片段能够:(i)与所述RNA指导序列结合,并且(ii)靶向所述靶RNA,条件是当所述复合物包含SEQ ID NO:1-7中的任一者的Cas时,所述间隔序列与天然存在的噬菌体核酸不是100%互补,或者其中所述靶RNA由真核DNA编码。

[0016] 在某些实施方式中,所述DR序列具有与SEQ ID NO:8-14中的任一者的二级结构基本相同的二级结构。

[0017] 在某些实施方式中,所述DR序列由SEQ ID NO:8-14中的任一者编码。

[0018] 在某些实施方式中,所述靶RNA由真核DNA编码。

[0019] 在某些实施方式中,所述真核DNA是非人哺乳动物DNA、非人灵长类动物DNA、人DNA、植物DNA、昆虫DNA、鸟DNA、爬行动物DNA、啮齿动物DNA、鱼DNA、蠕虫/线虫DNA、酵母DNA。

[0020] 在某些实施方式中,所述靶RNA是mRNA。

[0021] 在某些实施方式中,所述间隔序列在15-55个核苷酸之间、在25-35个核苷酸之间、或为约30个核苷酸。

[0022] 在某些实施方式中,所述间隔序列与所述靶RNA是90%-100%互补。

[0023] 在某些实施方式中,所述衍生物包含SEQ ID NO:1-7中的任一者的一个或多个残基的保守氨基酸取代。

[0024] 在某些实施方式中,所述衍生物仅包含保守氨基酸取代。

[0025] 在某些实施方式中,所述衍生物在HEPN结构域或RXXXXH基序中具有与SEQ ID NO:1-7中的任一者的野生型Cas相同的序列。

[0026] 在某些实施方式中,所述衍生物能够与杂交至所述靶RNA的RNA指导序列结合,但由于所述Cas的RNA酶催化位点突变而不具有RNA酶催化活性。

[0027] 在某些实施方式中,所述衍生物具有不超过210个残基的N-末端缺失、和/或不超过180个残基的C-末端缺失。

[0028] 在某些实施方式中,所述衍生物具有约180个残基的N-末端缺失、和/或约150个残基的C-末端缺失。

[0029] 在某些实施方式中,所述衍生物进一步包含RNA碱基编辑结构域。

[0030] 在某些实施方式中,所述RNA碱基编辑结构域是腺苷脱氨酶,如双链RNA特异性腺

苷脱氨酶(例如,ADAR1或ADAR2);载脂蛋白B mRNA编辑酶,催化多肽样(APOBEC);或激活诱导的胞苷脱氨酶(AID)。

[0031] 在某些实施方式中,ADAR2具有E488Q/T375G双突变或者是ADAR2DD。

[0032] 在某些实施方式中,将所述碱基编辑结构域与RNA结合结构域,如MS2进一步融合。

[0033] 在某些实施方式中,所述衍生物进一步包含RNA甲基转移酶、RNA去甲基化酶、RNA剪接修饰子、定位因子或翻译修饰因子。

[0034] 在某些实施方式中,所述Cas、所述衍生物或所述功能性片段包含核定位信号(NLS)序列或核输出信号(NES)。

[0035] 在某些实施方式中,靶向所述靶RNA导致对所述靶RNA进行修饰。

[0036] 在某些实施方式中,所述靶RNA修饰是切割所述靶RNA。

[0037] 在某些实施方式中,所述靶RNA修饰是腺苷(A)脱氨基为肌苷(I)。

[0038] 在某些实施方式中,本发明的CRISPR-Cas复合物进一步包含靶RNA,所述靶RNA包含能够与所述间隔序列杂交的序列。

[0039] 本发明的另一方面提供融合蛋白,所述融合蛋白包含(1)本发明的Cas、其衍生物或其功能性片段,和(2)异源功能性结构域。

[0040] 在某些实施方式中,所述异源功能性结构域包含:核定位信号(NLS)、报告蛋白或检测标记(例如,GST、HRP、CAT、GFP、HcRed、DsRed、CFP、YFP、BFP)、定位信号、蛋白靶向部分、DNA结合结构域(例如,MBP、Lex A DBD、Gal4 DBD)、表位标签(例如,His、myc、V5、FLAG、HA、VSV-G、Trx等)、转录激活结构域(例如,VP64或VPR)、转录抑制结构域(例如,KRAB部分或SID部分)、核酸酶(例如,FokI)、脱氨基结构域(例如,ADAR1、ADAR2、APOBEC、AID或TAD)、甲基化酶、去甲基化酶、转录释放因子、HDAC、具有ssRNA切割活性的多肽、具有dsRNA切割活性的多肽、具有ssDNA切割活性的多肽、具有dsDNA切割活性的多肽、DNA或RNA连接酶、或其任何组合。

[0041] 在某些实施方式中,将所述异源功能性结构域在所述融合蛋白的N-末端、C-末端或内部融合。

[0042] 本发明的另一方面提供缀合物,所述缀合物包含与(2)缀合的(1):(1)本发明的Cas、其衍生物或其功能性片段,(2)异源功能性部分。

[0043] 在某些实施方式中,所述异源功能性部分包含:核定位信号(NLS)、报告蛋白或检测标记(例如,GST、HRP、CAT、GFP、HcRed、DsRed、CFP、YFP、BFP)、定位信号、蛋白靶向部分、DNA结合结构域(例如,MBP、Lex A DBD、Gal4 DBD)、表位标签(例如,His、myc、V5、FLAG、HA、VSV-G、Trx等)、转录激活结构域(例如,VP64或VPR)、转录抑制结构域(例如,KRAB部分或SID部分)、核酸酶(例如,FokI)、脱氨基结构域(例如,ADAR1、ADAR2、APOBEC、AID或TAD)、甲基化酶、去甲基化酶、转录释放因子、HDAC、具有ssRNA切割活性的多肽、具有dsRNA切割活性的多肽、具有ssDNA切割活性的多肽、具有dsDNA切割活性的多肽、DNA或RNA连接酶、或其任何组合。

[0044] 在某些实施方式中,所述异源功能性部分相对于所述Cas、其衍生物或其功能性片段在N-末端、C-末端或内部缀合。

[0045] 本发明的另一方面提供多核苷酸,所述多核苷酸编码SEQ ID NO:1-7中的任一者、或其衍生物、或其功能性片段、或其融合蛋白,条件是所述多核苷酸不是SEQ ID NO:15-21

中的任一者。

[0046] 在某些实施方式中,所述多核苷酸经密码子优化以在细胞中表达。

[0047] 在某些实施方式中,所述细胞是真核细胞。

[0048] 本发明的另一方面提供非天然存在的多核苷酸,所述非天然存在的多核苷酸包含 SEQ ID NO:8-14中的任一者的衍生物,其中所述衍生物(i)与SEQ ID NO:8-14中的任一者相比具有一个或多个(例如,1、2、3、4、5、6、7、8、9或10个)核苷酸添加、缺失或取代;(ii)与SEQ ID NO:8-14中的任一者具有至少20%、30%、40%、50%、60%、70%、80%、90%、95%或97%的序列同一性;(iii)在严格条件下与SEQ ID NO:8-14中的任一者或(i)和(ii)中的任一者杂交;或(iv)是(i)-(iii)中的任一者的互补序列,条件是所述衍生物不是SEQ ID NO:8-14中的任一者,并且所述衍生物编码RNA(或者是RNA),所述RNA保持与由SEQ ID NO:8-14编码的任一RNA基本相同的二级结构。

[0049] 在某些实施方式中,所述衍生物用作本发明的Cas、其衍生物或其功能性片段中的任一者的DR序列。

[0050] 本发明的另一方面提供载体,所述载体包含本发明的多核苷酸。

[0051] 在某些实施方式中,所述多核苷酸与启动子和任选的增强子可操作地连接。

[0052] 在某些实施方式中,所述启动子是组成型启动子、诱导型启动子、泛素启动子或组织特异性启动子。

[0053] 在某些实施方式中,所述载体是质粒。

[0054] 在某些实施方式中,所述载体是逆转录病毒载体、噬菌体载体、腺病毒载体、单纯疱疹病毒(HSV)载体、AAV载体或慢病毒载体。

[0055] 在某些实施方式中,所述AAV载体是血清型AAV1、AAV2、AAV4、AAV5、AAV6、AAV7、AAVrh74、AAV8、AAV9、AAV10、AAV 11、AAV 12或AAV 13的重组AAV载体。

[0056] 本发明的另一方面提供递送系统,所述递送系统包含(1)递送媒介物,和(2)本发明的CRISPR-Cas复合物、本发明的融合蛋白、本发明的缀合物、本发明的多核苷酸、或本发明的载体。

[0057] 在某些实施方式中,所述递送媒介物是纳米颗粒、脂质体、外泌体、微泡或基因枪。

[0058] 本发明的另一方面提供细胞或其后代,所述细胞或其后代包含本发明的CRISPR-Cas复合物、本发明的融合蛋白、本发明的缀合物、本发明的多核苷酸、或本发明的载体。

[0059] 在某些实施方式中,所述细胞或其后代是真核细胞(例如,非人哺乳动物细胞、人细胞或植物细胞)或原核细胞(例如,细菌细胞)。

[0060] 本发明的另一方面提供非人多细胞真核生物,所述非人多细胞真核生物包含本发明的细胞。

[0061] 在某些实施方式中,所述非人类多细胞真核生物是针对人遗传障碍的动物(例如,啮齿动物或灵长类动物)模型。

[0062] 本发明的另一方面提供修饰靶RNA的方法,所述方法包括使所述靶RNA与本发明的CRISPR-Cas复合物接触,其中所述间隔序列与所述靶RNA的至少15个核苷酸互补;其中所述Cas、所述衍生物或所述功能性片段与所述RNA指导序列缔合以形成所述复合物;其中所述复合物与所述靶RNA结合;并且其中在所述复合物与所述靶RNA结合后,所述Cas、所述衍生物或所述功能性片段修饰所述靶RNA。

[0063] 在某些实施方式中,通过由所述Cas进行切割来修饰所述靶RNA。

[0064] 在某些实施方式中,通过由包含双链RNA特异性腺苷脱氨酶的衍生物进行脱氨基来修饰所述靶RNA。

[0065] 在某些实施方式中,所述靶RNA是mRNA、tRNA、rRNA、非编码RNA、lncRNA或核RNA。

[0066] 在某些实施方式中,在所述复合物与所述靶RNA结合后,所述Cas、所述衍生物和所述功能性片段不会展现出大量的(或可检测的)附带RNA酶活性。

[0067] 在某些实施方式中,所述靶RNA在细胞内。

[0068] 在某些实施方式中,所述细胞是癌细胞。

[0069] 在某些实施方式中,所述细胞受感染原感染。

[0070] 在某些实施方式中,所述感染原是病毒、朊病毒、原生动物、真菌或寄生物。

[0071] 在某些实施方式中,所述CRISPR-Cas复合物由以下编码:编码SEQ ID NO:1-7中的任一者或其衍生物或功能性片段的第一多核苷酸,以及包含SEQ ID NO:8-14中的任一者和编码能够与所述靶RNA结合的间隔RNA的序列的第二多核苷酸,其中将所述第一多核苷酸和所述第二多核苷酸引入所述细胞中。

[0072] 在某些实施方式中,通过相同的载体将所述第一多核苷酸和所述第二多核苷酸引入所述细胞中。

[0073] 在某些实施方式中,所述方法导致以下中的一项或多项:(i) 体外或体内诱导细胞衰老;(ii) 体外或体内细胞周期停滞;(iii) 体外或体内细胞生长抑制;(iv) 体外或体内诱导无反应性;(v) 体外或体内诱导细胞凋亡;以及(vi) 体外或体内诱导坏死。

[0074] 本发明的另一方面提供在有需要的受试者中治疗病症或疾病的方法,所述方法包括向所述受试者施用包含本发明的CRISPR-Cas复合物或编码所述CRISPR-Cas复合物的多核苷酸的组合物;其中所述间隔序列与以下互补:与所述病症或疾病相关的靶RNA的至少15个核苷酸;其中所述Cas、所述衍生物或所述功能性片段与所述RNA指导序列缔合以形成所述复合物;其中所述复合物与所述靶RNA结合;并且其中在所述复合物与所述靶RNA结合后,所述Cas、所述衍生物或所述功能性片段切割所述靶RNA,从而在所述受试者中治疗所述病症或疾病。

[0075] 在某些实施方式中,所述病症或疾病是癌症或感染性疾病。

[0076] 在某些实施方式中,所述癌症是威尔姆斯肿瘤、尤因肉瘤、神经内分泌肿瘤、胶质母细胞瘤、神经母细胞瘤、黑素瘤、皮肤癌、乳腺癌、结肠癌、直肠癌、前列腺癌、肝癌、肾癌、胰腺癌、肺癌、胆道癌、子宫颈癌、子宫内膜癌、食道癌、胃癌、头颈癌、甲状腺髓样癌、卵巢癌、神经胶质瘤、淋巴瘤、白血病、骨髓瘤、急性淋巴细胞白血病、急性骨髓性白血病、慢性淋巴细胞白血病、慢性骨髓性白血病、霍奇金淋巴瘤、非霍奇金淋巴瘤或膀胱癌。

[0077] 在某些实施方式中,所述方法是体外方法、体内方法或离体方法。

[0078] 本发明的另一方面提供细胞或其后代,所述细胞或其后代通过本发明的方法获得,其中所述细胞和所述后代包含非天然存在的修饰(例如,在所述细胞/后代的转录的RNA中的非天然存在的修饰)。

[0079] 本发明的另一方面提供检测靶RNA的存在的方法,所述方法包括使所述靶RNA与包含本发明的融合蛋白、或本发明的缀合物、或编码所述融合蛋白的多核苷酸的组合物接触,其中所述融合蛋白或所述缀合物包含可检测标记(例如,可通过荧光、RNA印迹或FISH检测

的标记)和能够与所述靶RNA结合的复合间隔序列。

[0080] 本发明的另一方面提供真核细胞,所述真核细胞包含成簇的规律间隔的短回文重复序列(CRISPR)-Cas复合物,所述CRISPR-Cas复合物包含:(1)RNA指导序列,其包含能够与靶RNA杂交的间隔序列以及在所述间隔序列的3'的同向重复(DR)序列;和(2)CRISPR相关蛋白(Cas)或所述Cas的衍生物或功能性片段,所述CRISPR相关蛋白具有SEQ ID NO:1-7中的任一者的氨基酸序列;其中所述Cas、所述Cas的衍生物和功能性片段能够(i)与所述RNA指导序列结合,并且(ii)靶向所述靶RNA。

[0081] 应理解,本文描述的本发明的任何一个实施方式,包括仅在实施例或权利要求中、或仅在下文的一个方面/部分中描述的那些实施方式,可以与本发明的任何其他一个或多个实施方式组合,除非明确否认或认为不当。

附图说明

[0082] 图1是代表性Cas13e和Cas13f家族成员的基因组基因座的示意图(未按比例绘制)。显示出Cas编码序列(带尖端的长条),随后是多个附近的同向重复序列(DR)(短条)和间隔序列(菱形)。

[0083] 图2显示出与各个Cas13e和Cas13f蛋白相关的DR序列的推定二级结构。它们的编码序列从左到右分别由SEQ ID NO:8-14表示。

[0084] 图3显示出本发明新发现的Cas13e和Cas13f效应蛋白以及相关的先前发现的Cas13a、Cas13b、Cas13c和Cas13d效应蛋白的系统发育树。

[0085] 图4显示出Cas13a-Cas13f蛋白的结构域结构。指示出Cas蛋白的每个代表性成员上两个RXXXXH基序的总体大小和位置。

[0086] 图5显示出Cas13e.1效应蛋白的预测3D结构。

[0087] 图6是示意图,其显示出可以将分别编码以下的三种质粒转染至细胞以表达它们各自的基因产物,从而导致报告mCherry mRNA的降解:(1)Cas13e效应蛋白,(2)指导RNA(gRNA)的编码序列,其可以产生与mCherry mRNA互补、并且可以与所述Cas13e效应蛋白形成复合物的指导RNA,以及(3)mCherry报告基因。

[0088] 图7显示出通过与mCherry mRNA互补的指导RNA进行的mCherry mRNA的敲低,如在荧光显微镜下mCherry的表达降低所证明的。作为阴性对照,不与mCherry mRNA杂交/结合的非靶向(NT)指导RNA未能敲低mCherry的表达。

[0089] 图8显示出在图6的实验中mCherry表达的约75%敲低。

[0090] 图9显示出Cas13e利用具有在3'端的DR序列的指导RNA(与在指导RNA的5'端的DR序列相反)。

[0091] 图10显示出相对于非靶向(NT)对照,间隔序列长度与针对靶RNA的特异性(指导RNA依赖性)RNA酶活性之间的相关性。

[0092] 图11显示出相对于非靶向(NT)对照,间隔序列长度与针对靶RNA的非特异性/附带(非指导RNA依赖性)RNA酶活性之间的相关性。

[0093] 图12显示出dCas13e.1-ADAR2DD融合物具有RNA碱基编辑活性。特别地,可以将分别编码以下的三种质粒转染至细胞以表达它们各自的基因产物:(1)与单碱基RNA编辑器(editor)ADAR2DD融合的dCas13e(RNA酶死亡)蛋白,(2)指导RNA(gRNA)的编码序列,其可以

产生与具有G至A点突变的突变体mCherry mRNA互补、并且可以与所述dCas13e效应蛋白形成复合物的指导RNA,以及(3)编码所述具有G至A点突变的mCherry mRNA的突变体mCherry报告基因。由于点突变,突变体mCherry mRNA通常不能产生荧光mCherry蛋白。在指导RNA与突变体mCherry mRNA结合后,融合的ADAR2DD碱基编辑子将A转化为I(G等同物),从而恢复mRNA编码荧光mCherry蛋白的能力。

[0094] 图13显示出由于成功的RNA碱基编辑而恢复的mCherry表达。在图12的实验中,单独的编码突变体mCherry的质粒(mCherry*)无法表达荧光mCherry。单独的编码dCas13e-ADAR2DD碱基编辑子的质粒也无法表达荧光mCherry。单独的编码gRNA-1或gRNA-2的质粒(其也表达GFP报告子)尽管显著地表达GFP,但也无法表达荧光mCherry。然而,当将所有三种质粒转染至相同的细胞中时,观察到显著的荧光mCherry表达(连同GFP报告子表达)。

[0095] 图14显示出具有提前终止密码子TAG的突变体mCherry基因的相关区段、可与dCas13e-ADAR2DD RNA碱基编辑子复合的两个gRNA的序列、以及“校正的”TGG密码子。

[0096] 图15是示意图(未按比例绘制),其显示出与ADAR2DD RNA碱基编辑子(显示为“ADAR2”)以及其他转录控制元件融合的dCas13e.1的一系列渐进C-末端缺失构建体。

[0097] 图16显示出针对图15中的C-末端缺失突变体系列,mCherry突变体转化回野生型mCherry的百分比结果。

[0098] 图17是示意图(未按比例绘制),其显示出与ADAR2DD RNA碱基编辑子融合的dCas13e.1的一系列渐进C-末端和任选的N-末端缺失构建体。

[0099] 图18显示出针对选定的图17中的C-末端和N-末端缺失突变体,mCherry突变体转化回野生型mCherry的百分比结果。

[0100] 图19显示出编码Cas13a、Cas13b、Cas13d、Cas13e.1和Cas13f.1、mCherry报告基因、以及靶向ANXA4的gRNA编码序列或作为对照的非靶向gRNA的一系列质粒。

[0101] 图20显示出Cas13e.1、Cas13f.1、Cas13a以及Cas13d对ANXA4表达的高效敲低。

具体实施方式

[0102] 1. 概述

[0103] 本文描述的本发明提供新型2类VI型Cas效应蛋白,有时在本文中称为Cas13e和Cas13f。本发明的新型Cas13蛋白比先前发现的Cas13效应蛋白(Cas13a-Cas13d)小得多,使得它们可以很容易地与它们的crRNA编码序列一起包装到小容量基因疗法载体(如AAV载体)中。此外,与Cas13a、Cas13b和Cas13d效应蛋白相比,新发现的Cas13e和Cas13f效应蛋白在敲低RNA靶序列方面更有效,并且在RNA单碱基编辑方面更高效,同时在通过基于crRNA的靶识别激活后展现出可忽略不计的非特异性/附带RNA酶活性,除非间隔序列在特定的窄范围内(例如,约30个核苷酸)。因此,这些新的Cas蛋白非常适合基因疗法。

[0104] 因此,在第一方面,本发明提供Cas13e和Cas13f效应蛋白(如具有SEQ ID NO:1-7的氨基酸序列的那些)、或其直系同源物、同源物、各种衍生物(下文描述)、功能性片段(下文描述),其中所述直系同源物、同源物、衍生物和功能性片段保持SEQ ID NO:1-7的任一种蛋白的至少一种功能。这样的功能包括但不限于结合本发明的指导RNA/crRNA(下文描述)以形成复合物的能力、RNA酶活性、以及在与靶RNA至少部分互补的crRNA的指导下在特定位点处结合并切割所述靶RNA的能力。

[0105] 在某些实施方式中,本发明的Cas13e或Cas13f效应蛋白可以是:(i) SEQ ID NO:1-7中的任一者;(ii) 具有SEQ ID NO:1-7中的任一者的一个或多个氨基酸(例如,1、2、3、4、5、6、7、8、9或10个残基)的添加、缺失和/或取代(例如,保守取代)的衍生物;或(iii) 与SEQ ID NO:1-7中的任一者相比具有至少约80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%的氨基酸序列同一性的衍生物。

[0106] 在某些实施方式中,Cas13e和Cas13f效应蛋白、其直系同源物、同源物、衍生物和功能性片段不是天然存在的,例如与天然存在的序列相比具有至少一个氨基酸差异。

[0107] 在相关方面,本发明提供基于SEQ ID NO:1-7中任一者的Cas13e和Cas13f效应蛋白、或其上述直系同源物、同源物、衍生物和功能性片段的额外的衍生物,所述额外的衍生物包含另一共价或非共价连接的蛋白或多肽或其他分子(如检测试剂或药物/化学部分)。这样的其他蛋白/多肽/其他分子可以通过例如化学偶联、基因融合或其他非共价连接(如生物素-链霉亲和素结合)来连接。这样的衍生的蛋白不影响原始蛋白的功能,如结合本发明的指导RNA/crRNA(下文描述)以形成复合物的能力、RNA酶活性、以及在与靶RNA至少部分互补的crRNA的指导下在特定位点处结合并切割所述靶RNA的能力。

[0108] 例如,这样的衍生可用于添加核定位信号(NLS,如SV40大T抗原NLS)以增强主题Cas13e和Cas13f效应蛋白进入细胞核的能力。这样的衍生也可用于添加靶向分子或部分,以将主题Cas13e和Cas13f效应蛋白引导至特定的细胞或亚细胞位置。这样的衍生还可用于添加可检测标记,以促进主题Cas13e和Cas13f效应蛋白的检测、监测或纯化。这样的衍生可进一步用于添加脱氨基酶部分(如具有腺嘌呤或胞嘧啶脱氨基活性的酶部分)以促进RNA碱基编辑。

[0109] 衍生可以通过在主题Cas13e和Cas13f效应蛋白的N-末端或C-末端处或在内部(例如,通过内部氨基酸的侧链进行内部融合或连接)添加任一额外的部分来进行。

[0110] 在相关的第二方面,本发明提供基于SEQ ID NO:1-7中任一者的主题Cas13e和Cas13f效应蛋白、或其上述直系同源物、同源物、衍生物和功能性片段的缀合物,所述缀合物缀合有如其他蛋白或多肽、可检测标记、或其组合等部分。这样的缀合的部分可包括但不限于定位信号、报告基因(例如,GST、HRP、CAT、GFP、HcRed、DsRed、CFP、YFP、BFP)、标记(例如,荧光染料,如FITC或DAPI)、NLS、靶向部分、DNA结合结构域(例如,MBP、Lex A DBD、Gal4 DBD)、表位标签(例如,His、myc、V5、FLAG、HA、VSV-G、Trx等)、转录激活结构域(例如,VP64或VPR)、转录抑制结构域(例如,KRAB部分或SID部分)、核酸酶(例如,FokI)、脱氨基结构域(例如,ADAR1、ADAR2、APOBEC、AID或TAD)、甲基化酶、去甲基化酶、转录释放因子、HDAC、具有ssRNA切割活性的多肽、具有dsRNA切割活性的多肽、具有ssDNA切割活性的多肽、具有dsDNA切割活性的多肽、DNA或RNA连接酶、其任何组合等。

[0111] 例如,所述缀合物可包括一个或多个NLS,其可以位于或接近N-末端、C-末端、内部、或其组合。缀合可以通过氨基酸(如D或E、或S或T)、氨基酸衍生物(如Ahx、 β -Ala、GABA或Ava)或PEG连接来进行。

[0112] 在某些实施方式中,缀合不影响原始蛋白的功能,如结合本发明的指导RNA/crRNA(下文描述)以形成复合物的能力、RNA酶活性、以及在与靶RNA至少部分互补的crRNA的指导下在特定位点处结合并切割所述靶RNA的能力。

[0113] 在相关的第三方面,本发明提供基于SEQ ID NO:1-7中任一者的主题Cas13e和

Cas13f效应蛋白、或其上述直系同源物、同源物、衍生物和功能性片段的融合物,所述融合物具有如下部分,如定位信号、报告基因(例如,GST、HRP、CAT、GFP、HcRed、DsRed、CFP、YFP、BFP)、NLS、蛋白靶向部分、DNA结合结构域(例如,MBP、Lex A DBD、Gal4 DBD)、表位标签(例如,His、myc、V5、FLAG、HA、VSV-G、Trx等)、转录激活结构域(例如,VP64或VPR)、转录抑制结构域(例如,KRAB部分或SID部分)、核酸酶(例如,FokI)、脱氨基结构域(例如,ADAR1、ADAR2、APOBEC、AID或TAD)、甲基化酶、去甲基化酶、转录释放因子、HDAC、具有ssRNA切割活性的多肽、具有dsRNA切割活性的多肽、具有ssDNA切割活性的多肽、具有dsDNA切割活性的多肽、DNA或RNA连接酶、其任何组合等。

[0114] 例如,所述融合物可包括一个或多个NLS,其可以位于或接近N-末端、C-末端、内部、或其组合。在某些实施方式中,融合不影响原始蛋白的功能,如结合本发明的指导RNA/crRNA(下文描述)以形成复合物的能力、RNA酶活性、以及在与靶RNA至少部分互补的crRNA的指导下在特定位点处结合并切割所述靶RNA的能力。

[0115] 在第四方面,本发明提供分离的多核苷酸,所述分离的多核苷酸包含:(i) SEQ ID NO:8-14中的任一者;(ii)与SEQ ID NO:8-14中的任一者相比具有1、2、3、4或5个核苷酸的缺失、添加和/或取代的多核苷酸;(iii)与SEQ ID NO:8-14中的任一者共享至少80%、85%、90%、95%序列同一性的多核苷酸;(iv)在严格条件下与(i)-(iii)中的任一种多核苷酸或其互补序列杂交的多核苷酸;(v) (i)-(iii)中的任一多核苷酸的互补序列。

[0116] (ii)-(iv)中的任一多核苷酸保持原始SEQ ID NO:8-14的功能,即在主题Cas13e或Cas13f系统中编码crRNA的同向重复(DR)序列。

[0117] 如本文所用,“同向重复序列”可指CRISPR基因座中的DNA编码序列,或指在crRNA中由其编码的RNA。因此,当在RNA分子(如crRNA)的上下文中提到SEQ ID NO:8-14中的任一者时,每个T应理解为代表U。

[0118] 因此,在某些实施方式中,分离的多核苷酸是DNA,所述DNA编码主题Cas13e和Cas13f系统的crRNA的DR序列。

[0119] 在某些其他实施方式中,分离的多核苷酸是RNA,所述RNA是主题Cas13e和Cas13f系统的crRNA的DR序列。

[0120] 在第五方面,本发明提供复合物,所述复合物包含:(i)蛋白组合物,所述蛋白组合物可以是以下中的任一者:主题Cas13e或Cas13f效应蛋白、或其直系同源物、同源物、衍生物、功能性片段、其缀合物、或其融合物;和(ii)多核苷酸组合物,所述多核苷酸组合物包含本发明第4方面中描述的分离的多核苷酸(例如,DR序列)以及与靶RNA的至少一部分互补的间隔序列。在某些实施方式中,所述DR序列在所述间隔序列的3'端。

[0121] 在一些实施方式中,所述多核苷酸组合物是主题Cas13e或Cas13f系统的指导RNA/crRNA,其不包括tracrRNA。

[0122] 在某些实施方式中,为了与Cas13e和Cas13f效应蛋白、其同源物、直系同源物、衍生物、融合物、缀合物或具有RNA酶活性的功能性片段一起使用,所述间隔序列为至少约10个核苷酸,或在10-60、15-50、20-50、25-40、25-50或19-50个核苷酸之间。在某些实施方式中,为了与Cas13e和Cas13f效应蛋白、其同源物、直系同源物、衍生物、融合物、缀合物或不具有RNA酶活性但具有结合指导RNA和与所述指导RNA互补的靶RNA的能力的功能性片段一起使用,所述间隔序列为至少约10个核苷酸,或在约10-200、15-180、20-150、25-125、30-

110、35-100、40-80、45-60、50-55个核苷酸之间或为约50个核苷酸。

[0123] 在某些实施方式中,所述DR序列在15-36、20-36、22-36个核苷酸之间、或约36个核苷酸。在某些实施方式中,指导RNA中的DR序列具有与SEQ ID NO:8-14中的任一者的RNA版本基本相同的二级结构(包括茎、凸起(bulge)和环)。

[0124] 在某些实施方式中,指导RNA比上述任一间隔序列的长度长约36个核苷酸,如在45-96、55-86、60-86、62-86或63-86个核苷酸之间。

[0125] 在第六方面,本发明提供分离的多核苷酸,所述分离的多核苷酸包含:(i)编码以下的多核苷酸:SEQ ID NO:1-7的Cas13e或Cas13f效应蛋白中的任一者、或其直系同源物、同源物、衍生物、功能性片段、融合物;(ii)SEQ ID NO:8-14中的任一者的多核苷酸;或(iii)包含(i)和(ii)的多核苷酸。

[0126] 在一些实施方式中,所述多核苷酸不是天然存在的(naturally occurring/naturally existing),例如不包括SEQ ID NO:15-21。

[0127] 在一些实施方式中,所述多核苷酸经密码子优化以在原核生物中表达。在一些实施方式中,所述多核苷酸经密码子优化以在真核生物中(例如在人或人细胞中)表达。

[0128] 在第七方面,本发明提供载体,所述载体包含或涵盖第六方面的任一多核苷酸。所述载体可以是克隆载体或表达载体。仅举几例,所述载体可以是质粒、噬菌粒或粘粒。在某些实施方式中,所述载体可用于在哺乳动物细胞(如人细胞)中表达多核苷酸、SEQ ID NO:1-7的Cas13e或Cas13f效应蛋白中的任一者、或其直系同源物、同源物、衍生物、功能性片段、融合物;或第4方面的任一多核苷酸;或第5方面的任一复合物。

[0129] 在第八方面,本发明提供宿主细胞,所述宿主细胞包含本发明第4或第6方面的任一多核苷酸和/或第7方面的载体。所述宿主细胞可以是原核生物(如大肠杆菌)或来自真核生物(如酵母、昆虫、植物、动物(例如,哺乳动物,包括人和小鼠))的细胞。所述宿主细胞可以是分离的原代细胞(如用于离体疗法的骨髓细胞)或已建立的细胞系,如肿瘤细胞系、293T细胞或干细胞、iPC等。

[0130] 在相关方面,本发明提供真核细胞,所述真核细胞包含成簇的规律间隔的短回文重复序列(CRISPR)-Cas复合物,所述CRISPR-Cas复合物包含:(1)RNA指导序列,其包含能够与靶RNA杂交的间隔序列以及在所述间隔序列的3'的同向重复(DR)序列;和(2)CRISPR相关蛋白(Cas)或所述Cas的衍生物或功能性片段,所述CRISPR相关蛋白具有SEQ ID NO:1-7中的任一者的氨基酸序列;其中所述Cas、所述Cas的衍生物和功能性片段能够(i)与所述RNA指导序列结合,并且(ii)靶向所述靶RNA。

[0131] 在第九方面,本发明提供组合物,所述组合物包含:(i)选自以下的第一(蛋白)组合物:SEQ ID NO:1-7的Cas13e或Cas13f效应蛋白中的任一者、或其直系同源物、同源物、衍生物、缀合物、功能性片段、融合物;和(ii)包含RNA的第二(核苷酸)组合物,所述RNA涵盖指导RNA/crRNA、特别是间隔序列或其编码序列。所述指导RNA可包含DR序列和可与靶RNA互补或杂交的间隔序列。所述指导RNA可以与(i)的第一(蛋白)组合物形成复合物。在一些实施方式中,所述DR序列可以是本发明第4方面的多核苷酸。在一些实施方式中,所述DR序列可以在所述指导RNA的3'端。在一些实施方式中,所述组合物(如(i)和/或(ii))是非天然存在的或从天然存在的组合物修饰而来。在一些实施方式中,所述组合物的至少一种组分是非天然存在的或从所述组合物的天然存在的组分修饰而来。在一些实施方式中,所述靶序列

是来自原核生物或真核生物的RNA,如非天然存在的RNA。所述靶RNA可以存在于细胞内,如在胞质溶胶中或在细胞器内。在一些实施方式中,所述蛋白组合物可以具有可位于其N-末端或C-末端或内部的NLS。

[0132] 在第十方面,本发明提供组合物,所述组合物包含本发明第7方面的一种或多种载体,所述一种或多种载体包含:(i) 编码以下的第一多核苷酸:SEQ ID NO:1-7的Cas13e或Cas13f效应蛋白中的任一者、或其直系同源物、同源物、衍生物、功能性片段、融合物;任选地与第一调节元件可操作地连接;和(ii) 编码本发明的指导RNA的第二多核苷酸;任选地与第二调节元件可操作地连接。所述第一多核苷酸和所述第二多核苷酸可以在不同的载体上或在相同的载体上。所述指导RNA可与由所述第一多核苷酸编码的蛋白产物形成复合物,并包含DR序列(如第4方面的任一DR序列)和可与靶RNA结合/互补的间隔序列。在一些实施方式中,所述第一调节元件是启动子,如诱导型启动子。在一些实施方式中,所述第二调节元件是启动子,如诱导型启动子。在一些实施方式中,所述组合物(如(i)和/或(ii))是非天然存在的或从天然存在的组合物修饰而来。在一些实施方式中,所述组合物的至少一种组分是非天然存在的或从所述组合物的天然存在的组分修饰而来。在一些实施方式中,所述靶序列是来自原核生物或真核生物的RNA,如非天然存在的RNA。所述靶RNA可以存在于细胞内,如在胞质溶胶中或在细胞器内。在一些实施方式中,所述蛋白组合物可以具有可位于其N-末端或C-末端或内部的NLS。

[0133] 在一些实施方式中,所述载体是质粒。在一些实施方式中,所述载体是病毒载体,所述病毒载体基于逆转录病毒、不能复制的逆转录病毒、腺病毒、不能复制的腺病毒或AAV。在一些实施方式中,所述载体可以在宿主细胞中自我复制(例如,具有细菌复制起点序列)。在一些实施方式中,所述载体可以整合到宿主基因组中并与其一起复制。在一些实施方式中,所述载体是克隆载体。在一些实施方式中,所述载体是表达载体。

[0134] 本发明进一步提供递送组合物,所述递送组合物用于递送本发明第1-第3方面的SEQ ID NO:1-7的Cas13e或Cas13f效应蛋白中的任一者、或其直系同源物、同源物、衍生物、缀合物、功能性片段、融合物;本发明第4和/或第6方面的多核苷酸;本发明第5方面的复合物;本发明第7方面的载体;本发明第8方面的细胞,以及本发明第9和/或第10方面的组合物。递送可以使用媒介物(如一种或多种脂质体、一种或多种纳米颗粒、一种或多种外泌体、一种或多种微泡、基因枪或一种或多种病毒载体)通过本领域已知的任何一种方式,如转染、脂质转染、电穿孔、基因枪、显微注射、超声、磷酸钙转染、阳离子转染、病毒载体递送等来进行。

[0135] 本发明进一步提供试剂盒,所述试剂盒包含以下中的任一者或多者:本发明第1-第3方面的SEQ ID NO:1-7的Cas13e或Cas13f效应蛋白中的任一者、或其直系同源物、同源物、衍生物、缀合物、功能性片段、融合物;本发明第4和/或第6方面的多核苷酸;本发明第5方面的复合物;本发明第7方面的载体;本发明第8方面的细胞,以及本发明第9和/或第10方面的组合物。在一些实施方式中,所述试剂盒可进一步包括关于如何使用试剂盒组分和/或如何从第3方获得用于与所述试剂盒组分一起使用的其他组分的说明。所述试剂盒的任何组分都可以储存在任何合适的容器中。

[0136] 上文大体描述了本发明,本发明各个方面的更详细描述在下文的单独部分中提供。然而,应理解,为了简洁和减少冗余,本发明的某些实施方式仅在一个部分下描述或仅

在权利要求或实施例中描述。因此,还应理解,本发明的任何一个实施方式,包括仅在一个方面、部分下或仅在权利要求或实施例中描述的那些实施方式,可以与本发明的任何其他实施方式组合,除非特别否认或组合不当。

[0137] 2. 新型2类VI型CRISPR RNA指导的RNA酶及其衍生物

[0138] 在一个方面,本文描述的本发明提供两个新型CRISPR 2类VI型效应子家族,其具有两个严格保守的RX4-6H (RXXXXH) 基序,这是高等真核生物和原核生物核苷酸结合 (HEPN) 结构域的特征。含有两个HEPN结构域的相似CRISPR 2类VI型效应子先前已被表征,并且包括例如CRISPR Cas13a (C2c2)、Cas13b、Cas13c和Cas13d。

[0139] HEPN结构域经证明是RNA酶结构域并赋予结合和切割靶RNA分子的能力。所述靶RNA可以是任何合适形式的RNA,包括但不限于mRNA、tRNA、核糖体RNA、非编码RNA、lncRNA (长链非编码RNA) 和核RNA。例如,在一些实施方式中,Cas蛋白识别并切割位于开放阅读框 (ORF) 的编码链上的RNA靶标。

[0140] 在一个实施方式中,本披露提供两个CRISPR 2类VI型效应子家族,在本文中通常称为VI-E型和VI-F型CRISPR-Cas效应蛋白Cas13e或Cas13f。VI-E型和VI-F型CRISPR-Cas效应蛋白与这些其他系统的效应子的直接比较显示出VI-E型和VI-F型CRISPR-Cas效应蛋白甚至比先前鉴定的最小的VI-D型/Cas13d效应子显著更小(例如,氨基酸少约20%) (参见图4),并且在与其他先前描述的效应蛋白(包括系统发育上最接近的亲属Cas13b)的一对一序列比对中具有小于30%的序列相似性(参见图3)。

[0141] 这两个新鉴定的CRISPR 2类VI型效应子家族可用于多种应用中,并且特别适用于治疗性应用,因为它们比其他效应子(例如,CRISPR Cas13a、Cas13b、Cas13c和Cas13d效应子)显著更小,这允许将编码效应子的核酸及它们的指导RNA编码序列包装到具有大小限制的递送系统(如AAV载体)中。此外,在特定的RNA酶活性激活后,在选定的间隔序列长度范围(如约30个核苷酸,参见图11)处可检测的附带/非特异性RNA酶活性的缺乏使这些Cas效应子在希望不受破坏的靶细胞中较不易于(如果不能免于)产生潜在危险的普遍脱靶RNA消化。另一方面,在其他选定的间隔子长度(如约30个核苷酸)处,这些Cas效应子存在显著的附带RNA酶活性,因此主题Cas效应子也可以用于依赖这样的附带RNA酶活性的效用中。

[0142] 在细菌中,VI-E型和VI-F型CRISPR-Cas系统包括紧邻CRISPR阵列的单个效应子(分别为大约775个残基和790个残基)(参见图1)。所述CRISPR阵列包括长度典型地为36个核苷酸的同向重复(DR)序列,其在序列和二级结构中通常都非常保守(参见图2)。

[0143] 本文提供的数据表明crRNA从5'端加工,使得所述DR序列在成熟的crRNA的3'端处终止。

[0144] Cas13e和Cas13f CRISPR阵列中含有的间隔子最常见的长度为30个核苷酸,大部分长度变化包含在29至30个核苷酸的范围内。然而,可以容忍宽范围的间隔子长度。例如,为了在功能性Cas13e或Cas13f效应蛋白、或其同源物、直系同源物、衍生物、融合物、缀合物或功能性片段中使用,间隔子可以在10-60个核苷酸、20-50个核苷酸、25-45个核苷酸、25-35个核苷酸之间,或为约27、28、29、30、31、32或33个核苷酸。然而,为了在上述任一者的dCas版本中使用,间隔子可以在10-200个核苷酸、20-150个核苷酸、25-100个核苷酸、25-85个核苷酸、35-75个核苷酸、45-60个核苷酸之间,或为约46、47、48、49、50、51、52、53、54或55个核苷酸。

[0145] 示例性VI-E型和VI-F型CRISPR-Cas效应蛋白在下表中提供。

[0146]

Cas13e.1	MAQVSKQTSKKRELSIDEYQGARKWCFTIAFNKALVNRDKNDGLFVESLLRHEKYSKHDWY DEDTRALIKCSTQAANAKAEALRNYFSHYRHSPGCLTFTAEDLRTIMERAYERAIFECRR RETEVIEFPSLFEGDRITTAGVVFVVSFFVERRVLDRLYGAVSGLKKNEGQYKLTRKALS MYCLKDSRFTKAWDKRVLLFRDILAQLGRI PAEAYEYYHGEQGDKKRANDNEGTPKRHKD KFIEFALHYLEAQHSEICFGRRHIVREEAGAGDEHKKHRTKGKVVDVDFSKKDEDQSYISK NNVIVRIDKNAGPRSYRMGLNELKYLVL LSLQGKGDDAIAKLYRYRQHVENILDVVKVTDK DNHVFLPRFVLEQHGIGRKAFAKQRIDGRVKHVRGVWEKKKAATNEMTLHEKARDILQYVNE NCTRSFNPGEYNRLLVCLVGKDVENFQAGLKRLQLAERIDGRVYSIFAQTSTINEMHQVVC DQILNRLCRIGDQKLYDYVGLGKKDEIDYKQKVAWFKEHISIRRGFLRKKFWYDSKKGFAK LVEEHLESGGGQRDVGLDKKYYHIDAIGRFEGANPALYETLARDRLCLMMAQYFLGSRKE LGNKIVWSNDSIELPVEGSGVNEKSIVFSVSDYGKLYVLDDAEFLGRICEYFMPHEKGKIR YHTVYEKGFRAYNDLQKKCVAVLAFEEKVVKAKKMSEKEGAHYIDFREILAQTMCKEAEK TAVNKVRRAFFHHHLKFVIDEFGLFSDVMKKYGIEKEWKFPVK* (SEQ ID NO: 1)
Cas13e.2	MKVENIKEKSKKAMYLINHYEGPKKWCFAIVLNRACDNYEDNPHLFSKSLLEFEKTSRKDW FDEETRELVEQADTEIQPNPNLKPNTTANRKLKDIRNYFSHHYHKNECLYFKNDDPIRCIM EAAYEKSKIYIKGKQIEQSDIPLPELFESSGWITPAGILLLASFFVERGILHRLMGNIGGF KDNERGEYGLTHDIFTTYCLKGSYSIRAQDHDAVMFRDILGYLSRVPTESFORIKQPQIRKE GQLSERKTDKFITFALNYLEDYGLKDLEGCKACFARSKIVREQENVESINDKEYKPHENKK

[0147]

	KVEIHFDQSKEDRFYINRNNVILKIQKKDGHSNIVRMGVYELKYLVLMSLVGKAKEAVEKI DNYIQDLRDQLPYIEGKNKEEIKKEYVRFFPRFIRSHLGLLQINDEEEKIKARLDYVKTWLD KKEKSKELELHKKGRDILRYINERCDRELNRNVYNRILELLVSKDLTGIFYRELEELKRTRR IDKNIVQNLSGQKTINALHEKVCDLVLKEIESLDTENLRKYLGLIPKEEKEVTFKEKVDRI LKQPVIIYKGFLRYQFFKDDKKSFVLLVEDALKEKGGGCDVPLGKEYYKIVSLDKYDKENKT LCETLAMDRCLMMARQYYLSLNAKLAQEAQQIEWKKEDSIELIIFTLNPDQSKQSFYSIR FSVRDFTKLYVTDDPEFLARLCSYFFPVEKEIEYHKLYSEGINKYTNLQKEGIEAILELEK KLIERNRIQSAKNYLSFNEIMNKS GYNKDEQDDLKVRNSLLHYKLIFEKEHLKKFYEVMR GEGIEKKWSLIV* (SEQ ID NO: 2)
Cas13f.1	MNGIELKKEEAAFYFNQAELNLKAIEDNIFDKERRKTLLNNPQILAKMENFIFNFRDVTKN AKGEIDCLLLKRELRFYSHYVHKRDVRELSKGEKPILEKYYQFAIESTGSENVKLEIIE NDAWLADAGVLFFLCIFLKKSQANKLISGISGFKRNDGTGQPRRNLFYFYSIREGYKVVPE MQKHFLFSLVNHLSNQDDYIEKAHQPYDIGEGLFFHRIASTFLNISGILRNMKFYTYQSK RLVEQRGELKREKIDFAWEPPFQNSYFEINGHKG VIGEDDELKELCYAFLIGNQDANKVEG RITQFLEKFRNANSVQQVKDDEMLKPEYFPANYFAESGVGRIKDRVLNRLNKAIKSNKAKK GEIIAYDKMREVMFAFINNSLPVDEKLKPKDYKRYLGMVRFWDREKDNIREFETKEWSKYL PSNFWTAKNLERVYGLAREKNAELFNKLKADVEKMDERELEKYQKINDAKDLANLRLASD FGVKWEEKDWDEYSGQIKKQITDSQKLTIMKQRITAGLKKKHGIENLNLRTIDINKSRKA VLNRIAIPRGFVKRHILGWQSEKVS K KIREAECEILLSKEYEELSKQFFQSKDYDKMTRI NGLYEKNKLIALMAVYLMGQLRILFKEHTKLDDITKTTVDFKISDKVTVKIPFSNYP SLVY TMSSKYVDNIGNYGF SNKDKDKPILGKIDVIEKQRM E F I KEVLGF EKYL FDDKI IDKSKFA DTATHISFAEIVEELVEKGWDKDR LTKLKDARNKALHGEILTGT SFDETKSLINELKK* (SEQ ID NO: 3)
Cas13f.2	MSPDFIKLEKQEAIFYFNQTELNLKAIESNILDQQRMI LLNNPRI LAKVGNFIFNFRDVT KNAKGEIDCLLFKLEELRFYSHYVHTDNVKELSNGEKPLLEYYQIAIQATRS EDVKFEL FETRNNENKITDAGVLFFLCMFLKKSQANKLISGISGFKRNDPTGQPRRNLFYFYSAREGYK ALPDMQKHFLF LTLVNYLSNQDEYISELKQYGEIGQGAFFNRIASTFLNISGISGNTKFYS YQSKRIEQRGELNSEKDSFEWIEPFQNSYFEINGHKG VIGEDDELKELCYALLVAKQDIN AVEGKIMQFLKKFRNTGNLQQVKDDEMLEIEYFPASYFNESKKEDIKKEILGR LDKKIRSC SAKAEKAYDKMKEVMEFINNSLPAAEKLKRKYRRYLKMVR FWSREKGNIEREFRTKEWSK YFSSDFWRKNNLEDVYKLATQKNAELFKNLKAAAEKMGETEF EKYYQQINDVKDLASLRRLT QDFGLKWE EKDWEEYSEQIKKQITDRQKLTIMKQRVTAELKKKHGIENLNLRTIDSNKSR KAVLNRIAIPRGFVKKHILGWQGSEKISK NIREAECKILLSK KYEELSRQFFEAGNFDKLT QINGLYEKNKLTA FMSVYLMGR LNIQLNKHTELGNLKKTEVDFKISDKVTEKIPFSQYPSL VYAMSRKYVDNVDKYKFSHQDKKKPFLGKIDSIEKERIEFIKEVLDFEEYLFKNKVIDKSK FSDTATHISFKEICDEMGGKGCNRNKLTELNNARNAAALHGEIPSETSFREAKPLINELKK*

[0148]

	(SEQ ID NO: 4)
Cas13f.3	MSPDFIKLEKQEAAFYFNQTELNLKAIESNIFDKQQRVILLNNPQILAKVGDFIFNFRDVT KNAKEIDCLLLKLREL<u>RNFYSHY</u>VYVYTDVVKILSNGERPLEKYYQFAIEATGSENVKLEI IESNNRLTEAGVLFFLCMFLKKSQANKLISGISGFKRNDPTGQPRNLFTYFVSUREGYKVV PDMQKHFLFLVLVNHLSGQDDYIEKAQKPYDIGEGLFFHRIASTFLNISGILRNMEFYIYQ SKRLKEQQGELKREKDIFWIEPFQGNSYFEINGNKGII GEDELKELCYALLVAGKDVRVAV EGKITQFLEKFKNADNAQQVEKDEMLDRNNFPANYFAESNIGSIKEKILNRLGKTDDSYNK TGTKIKPYDMMKEVMEFINNSLPADKLRKDYRRYLKMVRIWDSEKDNIREFESKEWSK YFSSDFWMAKNLERVYGLAREKNAELFNKLKAVVEKMDEREFEKYRLINSAEDLASLRLA KDFGLKWEKDWQEYSGQIKKQISDRQKLTIMKQRITAEKKKHGIENLNLRITIDSNKSR KAVLNRIAVPRGFVKEHILGWQGESEKVSCKTREACKILLSKEYEELSKQFFQTRNYDKMT QVNGLYEKNKLLAFMVVYLMERLNILLNKPTLNELEKAEVDFKISDKVMKIPFSQYPSL VYAMSSKYADSVGSYKFENDEKNKPFGLKIDTIEKQRMFEIKVEVLGFEEYLFEEKIIDKSE FADTATHISFDEICNELIKKGWDKDKLTKLDARNAALHGEIPAETSFREAKPLINGLKK*
Cas13f.4	MNI IKLKKEEAAFYFNQITILNLSGLDEIIEKQIPHIISNKENAKKVIDKIFNNRLLKSVE NYIYNFKDVAKNARTEIEAILLKLVEL<u>RNFYSHY</u>VHNDTVKILSNGEKPILEKYYQIAIEA TGSKNVKLVIIENNNCLTDSGVLFLLCMFLKKSQANKLISSVSGFKRNDKEGQPRNLFTY YSVREGYKVVDPDMQKHFLFLFALVNHLSQDDHIEKQQQSDELGKGLFFHRIASTFLNESGI FNKMQFYTYQSNRLKEKRGELKHEKDTFTWIEPFQGNSYFTLNHGKGVISEDQLKELCYTI LIEKQNVDSLEGKIIQFLKKFQNVSSKQQVDEDELLKREYFPANYFGRAGTGTLKEKILNR LDKRMPTSKVTDKAYDKMIEVMEFINMCLPSDEKLRQKDYRRYLKMVRFWNKEKHNIKRE FDSKKWTRFLPTLWLNKRNL EEAYQLARKENKKKLEDMRNQVRSLEKNDLEKYQQINYNVND LENLRLLSQELGVKWQEKDWVEYSGQIKKQISDNQKLTIMKQRITAEKKMHGIENLNLR SIDTNKSQRTVMNRALPKGFVNHIQQNSSEKISKRIREDYCKIELSGKYEELSRQFFDK KNFDKMTLINGLCEKNKLI AFMVIYLLERLGFELKEKTKLGELKQTRMTYKISDKVKEDIP LSYYPKLVYAMNRKYVDNIDSYAFAYESKKA ILDKVDIIEKQRMFEIKQVLCFEEYIFEN RIIEKSKFNDEETHISFTQIHDELIKGRDTEKLSKLKHARNKALHGEIPDGTSFEKAKLL INEIKK* (SEQ ID NO: 6)
Cas13f.5	MNAIELKKEEAAFYFNQARLNISGLDEIIEKQLPHIGSNRENKKTVDMLDNPEVLKKME NYVFNSRDI AKNARGELEALLLKLVEL<u>RNFYSHY</u>VHKDDVKTL SYGEKPLLDKYYEIAIEA TGSKDVRLEIIDDKNKLT DAGVLFLLCMFLKKSEANKLISSIRGFKRNDKEGQPRNLFTY YSVREGYKVVDPDMQKHFLFLTLVNHLSNQDEYISNLRPNQEIGQGGFFHRIASKFLSDSGI LHSMKFYTYRSKRLTEQRGELKPKKDHTWIEPFQGNSYFSVQGGKGVIGEEQLKELCYVL LVAREDFRAVEGKVTQFLKKFQANANNVQVEKDEVLEKEYFPANYFENRDVGRVKDKILNR LKKITESYKAGREVKAYDKMKEVMEFINNCLPTDENLKLKDYRRYLKMVRFWGREKENIK

[0149]

REFDSKKWERFLPRELWQKRNLEDAYQLAKEKNTLNFNKLKTVERMNELEFEKYQQINDA KDLANLRQLARDFGVKWEKDWQEYSGQIKKQITDRQKLTIMKQRITAALKKKQGIENLNL RITTDTNKSRKVVLNRIALPKGFVRKHILKTDIKISKQIRQSQCPIILSNMYMLAKEFFE ERNFDKMTQINGLFEKNVLIAFMIVYLMEQLNLRLGKNTELSNLKKTEVNFTITDKVTEKV QISQYPSLVFAINREYVDGISGYKLPPKKPKPEPPYTFFEKIDAIEKERMEFIKQVLGFEEH LFEKNVIDKTRFTDTATHISFNEICDELIKKGWDENKIIKLKDARNAALHGKIPEDTSFDE AKVLINELKK* (SEQ ID NO: 7)
--

[0150] 在上面的序列中,每个效应子中的两个RX4-6H(RXXXXH)基序加双下划线。在Cas13e.1中,由于基序侧翼的RR和HH序列,C-末端基序可具有两种可能性。在一个或两个这样的结构域处的突变可能产生Cas13e和Cas13f效应蛋白、其同源物、直系同源物、融合物、缀合物、衍生物或功能性片段的RNA酶死亡版本(或“dCas”),同时基本上保持它们结合指导RNA和与所述指导RNA互补的靶RNA的能力。

[0151] Cas效应子的对应DR编码序列在下面列出:

[0152]

Cas13e.1	GCTGGAGCAGCCCCGATTTGTGGGGTGATTACAGC (SEQ ID NO:8)
Cas13e.2	GCTGAAGAAGCCTCCGATTTGAGAGGTGATTACAGC (SEQ ID NO:9)
Cas13f.1	GCTGTGATAGACCTCGATTTGTGGGGTAGTAACAGC (SEQ ID NO:10)
Cas13f.2	GCTGTGATAGACCTCGATTTGTGGGGTAGTAACAGC (SEQ ID NO:11)
Cas13f.3	GCTGTGATAGACCTCGATTTGTGGGGTAGTAACAGC (SEQ ID NO:12)
Cas13f.4	GCTGTGATGGGCCTCAATTTGTGGGGAAGTAACAGC (SEQ ID NO:13)
Cas13f.5	GCTGTGATAGGCCTCGATTTGTGGGGTAGTAACAGC (SEQ ID NO:14)

[0153] 由于DR序列的二级结构(包括茎、凸起和环结构的位置和大小)可能比形成这样的二级结构的特定核苷酸序列更重要,替代性或衍生性DR序列也可用于本发明的系统和方法中,只要这些衍生性或替代性DR序列具有与由SEQ ID NO:8-14中的任一者编码的RNA的二级结构基本类似的二级结构即可。例如,衍生性DR序列可以在一个或两个茎中具有±1或2个碱基对(参见图2),在凸起的一条或两条单链中具有±1、2或3个碱基,和/或在环区中具有±1、2、3或4个碱基。

[0154] 在一些实施方式中,VI-E型和VI-F型CRISPR-Cas效应蛋白包括“衍生物”,所述衍生物具有如下氨基酸序列,所述氨基酸序列与上述SEQ ID NO:1-7中的任一者的氨基酸序列具有至少约80%的序列同一性(例如,81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%)。与SEQ ID NO:1-7中的任一者共享显著的蛋白序列同一性的这样的衍生性Cas效应子保留SEQ ID NO:1-7的Cas的至少一种功能(参见下文),例如与包含SEQ ID NO:8-14的DR序列中的至少一个的crRNA结合并形成复合物的能力。例如,Cas13e.1衍生物可分别与SEQ ID NO:1、2、3、4、5、6或7共享85%的氨基酸序列同一性,并保留分别与具有SEQ ID NO:8、9、10、11、12、13或14的DR序列的crRNA结合并形成复合物的能力。

[0155] 在一些实施方式中,所述衍生物包含保守的氨基酸残基取代。在一些实施方式中,所述衍生物仅包含保守的氨基酸残基取代(即,所述衍生物中的所有氨基酸取代都是保守取代,并且没有不保守的取代)。

[0156] 在一些实施方式中,所述衍生物将不超过1、2、3、4、5、6、7、8、9或10个氨基酸插入

或缺失包含到SEQ ID NO:1-7的任一种野生型序列中。只要保留野生型序列的至少一种功能,插入和/或缺失就可以聚集在一起,或在序列的整个长度上分开。这样的功能可以包括结合指导/crRNA的能力、RNA酶活性、结合和/或切割与指导/crRNA互补的靶RNA的能力。在一些实施方式中,插入和/或缺失不存在于RXXXXH基序中,或距RXXXXH基序5、10、15或20个残基内。

[0157] 在一些实施方式中,所述衍生物保留结合指导RNA/crRNA的能力。

[0158] 在一些实施方式中,所述衍生物保留指导/crRNA激活的RNA酶活性。

[0159] 在一些实施方式中,在所结合的在序列方面与至少一部分靶RNA互补的指导/crRNA存在下,所述衍生物保留结合靶RNA和/或切割所述靶RNA的能力。

[0160] 在其他实施方式中,由于例如RNA指导的RNA酶的一个或多个催化残基的突变,所述衍生物完全或部分丧失指导/crRNA激活的RNA酶活性。这样的衍生物有时称为dCas,如dCas13e.1等。

[0161] 因此,在某些实施方式中,所述衍生物可以经修饰以具有减弱的核酸酶/RNA酶活性,例如,与相应的野生型蛋白相比,核酸酶灭活至少50%、至少60%、至少70%、至少80%、至少90%、至少95%、至少97%或100%。可以通过本领域已知的几种方法减弱核酸酶活性,例如,将突变引入蛋白的核酸酶(催化)结构域中。在一些实施方式中,鉴定出核酸酶活性的催化残基,并且这些氨基酸残基可以被不同的氨基酸残基(例如,甘氨酸或丙氨酸)取代以减弱核酸酶活性。在一些实施方式中,所述氨基酸取代是保守性氨基酸取代。在一些实施方式中,所述氨基酸取代是非保守性氨基酸取代。

[0162] 在一些实施方式中,修饰包含在至少一个HEPN结构域中的一个或多个突变(例如,氨基酸缺失、插入或取代)。在一些实施方式中,在至少一个HEPN结构域中存在一个、两个、三个、四个、五个、六个、七个、八个、九个或更多个氨基酸取代。例如,在一些实施方式中,所述一个或多个突变包含在与以下对应的氨基酸残基处的取代(例如,丙氨酸取代):SEQ ID NO:1的R84、H89、R739、H744、R740、H745,或SEQ ID NO:2的R97、H102、R770、H775,或SEQ ID NO:3的R77、H82、R764、H769,或SEQ ID NO:4的R79、H84、R766A、H771,或SEQ ID NO:5的R79、H84、R766、H771,或SEQ ID NO:6的R89、H94、R773、H778,或SEQ ID NO:7的R89、H94、R777、H782。

[0163] 在某些实施方式中,所述一个或多个突变或所述两个或更多个突变可以在包含HEPN结构域的效应蛋白的催化活性结构域或与HEPN结构域同源的催化活性结构域中。在某些实施方式中,所述效应蛋白包含一个或多个以下突变:R84A、H89A、R739A、H744A、R740A、H745A(其中氨基酸位置对应于Cas13e.1的氨基酸位置)。本领域技术人员将理解,不同的Cas13e和Cas13f蛋白中的对应氨基酸位置可以突变成相同效果。在某些实施方式中,一个或多个突变完全或部分消除蛋白的催化活性(例如,改变的切割速率、改变的特异性等)。

[0164] 其他示例性(催化)残基突变包括:Cas13e.2的R97A、H102A、R770A、H775A,或Cas13f.1的R77A、H82A、R764A、H769A,或Cas13f.2的R79A、H84A、R766A、H771A,或Cas13f.3的R79A、H84A、R766A、H771A,或Cas13f.4的R89A、H94A、R773A、H778A,或Cas13f.5的R89A、H94A、R777A、H782A。在某些实施方式中,本文的任一R和/或H残基可以被G、V或I而不是A替代。

[0165] 与缺乏突变的对应野生型蛋白相比,这些突变中的至少一个的存在导致具有减少

的或减弱的RNA酶活性的衍生物。

[0166] 在某些实施方式中,如本文描述的效应蛋白是“死”效应蛋白,如死Cas13e或Cas13f效应蛋白(即dCas13e和dCas13f)。在某些实施方式中,所述效应蛋白在HEPN结构域1(N-末端)中具有一个或多个突变。在某些实施方式中,所述效应蛋白在HEPN结构域2(C-末端)中具有一个或多个突变。在某些实施方式中,所述效应蛋白在HEPN结构域1和HEPN结构域2中具有一个或多个突变。

[0167] 灭活的Cas或其衍生物或功能性片段可以与一个或多个异源/功能性结构域融合或缔合(例如,经由融合蛋白、接头肽、“GS”接头等)。这些功能性结构域可以具有多种活性,例如,甲基化酶活性、去甲基化酶活性、转录激活活性、转录抑制活性、转录释放因子活性、组蛋白修饰活性、RNA切割活性、DNA切割活性、核酸结合活性、碱基编辑活性和开关活动(例如,光诱导型)。在一些实施方式中,所述功能性结构域是Krüppel相关盒(KRAB)、SID(例如SID4X)、VP64、VPR、VP16、Fok1、P65、HSF1、MyoD1、作用于RNA的腺苷脱氨酶(如ADAR1、ADAR2)、APOBEC、胞苷脱氨酶(AID)、TAD、小型-SOG、APEX和生物素-APEX。

[0168] 在一些实施方式中,所述功能性结构域是碱基编辑结构域,例如,ADAR1(包括野生型或其ADAR1DD版本,具有或不具有E1008Q)、ADAR2(包括野生型或其ADAR2DD版本,具有或不具有一个或多个E488Q突变)、APOBEC或AID。

[0169] 在一些实施方式中,所述功能性结构域可以包含一个或多个核定位信号(NLS)结构域。所述一个或多个异源功能性结构域可以包含至少两个或更多个NLS结构域。所述一个或多个NLS结构域可位于或接近或邻近所述效应蛋白(例如,Cas13e/Cas13f效应蛋白)的末端处,并且如果有两个或更多个NLS,则两者中的每一个可位于或接近或邻近所述效应蛋白(例如,Cas13e/Cas13f效应蛋白)的末端处。

[0170] 在一些实施方式中,至少一个或多个异源功能性结构域可以位于或接近所述效应蛋白的氨基末端处,并且/或者其中至少一个或多个异源功能性结构域位于或接近所述效应蛋白的羧基末端处。所述一个或多个异源功能性结构域可以与所述效应蛋白融合。所述一个或多个异源功能性结构域可以与所述效应蛋白相连。所述一个或多个异源功能性结构域可以通过接头部分与所述效应蛋白连接。

[0171] 在一些实施方式中,存在多个(例如,两个、三个、四个、五个、六个、七个、八个或更多个)相同或不同的功能性结构域。

[0172] 在一些实施方式中,所述功能性结构域(例如,碱基编辑结构域)进一步与RNA结合结构域(例如,MS2)融合。

[0173] 在一些实施方式中,所述功能性结构域与接头序列(例如,柔性接头序列或刚性接头序列)缔合或经由接头序列(例如,柔性接头序列或刚性接头序列)融合。示例性接头序列和功能性结构域序列在下表中提供。

[0174] VI-E型和VI-F型CRISPR Cas效应子的工程化的变体中基序和功能性结构域的氨基酸序列

[0175]

接头 1	GS
接头 2	GSGGGGS
接头 3	GGGSGGGGSGGGGS
ADAR1DD- WT	SLGTGNRCVKGDSLSLKGETVNDCHAEIISRRGFIRFLYSELMKYNSQTAKDSIFEPAKGG EKLQIKKTVSFHLYISTAPCGDGA LFDKSCSDRAMESTESRHYPVFENPKQGKLR TKVENG EGTIPVESSDIVPTWDGIRLGERLRTMSCSDKILRWNVLGLQGALLTHFLQPIY LKSVTLG YLFSQGH LTRAICCRVTRDGS AFEDGLRHPFIVNHPKVGRVSIYDSKRQSGKT KETS VNWC LADGYDLEILDGTRGTVDGPRNELSRVSKKNIFLLFKKLC SFRYRRDLLRLSYGEAKKAAR DYETAKNYFKKGLKDMGYGNWISKPQEEKNF
ADAR1DD- E1008Q	SLGTGNRCVKGDSLSLKGETVNDCHAEIISRRGFIRFLYSELMKYNSQTAKDSIFEPAKGG EKLQIKKTVSFHLYISTAPCGDGA LFDKSCSDRAMESTESRHYPVFENPKQGKLR TKVENG QGTIPVESSDIVPTWDGIRLGERLRTMSCSDKILRWNVLGLQGALLTHFLQPIY LKSVTLG YLFSQGH LTRAICCRVTRDGS AFEDGLRHPFIVNHPKVGRVSIYDSKRQSGKT KETS VNWC LADGYDLEILDGTRGTVDGPRNELSRVSKKNIFLLFKKLC SFRYRRDLLRLSYGEAKKAAR DYETAKNYFKKGLKDMGYGNWISKPQEEKNF
ADAR2DD- WT	QLHL PQVLADAVSRLVLGKFGDLTDNFSSPHARRKVL AGVVM TGT DVKDAKVISVSTGTK CINGEYMSDRGLALNDCHAEIISRRSLLRFLYTQLELYLNNKDDQKRS IFQKSERGGFRLK ENVQFHLYISTSPCGDARIFSPHEPILEEPADRHPNRKARGQLRTKIESGEGTIPVRSNAS IQTWDGVLQGERLLTMS CSDKIARWNVVGIQGSLLSIFVEPIYFSSII LGS LYHGDHLSRA MYQRISNIEDLPPLYTLNKPLLSGISNAEARQPGKAPNFSVNWTVGDSAIEVINATTGKDE LGRASRLCKHALYCRWMRVHGKVP SHLLRSKITKPNVYHESKLA AKEYQA AKARLFTAFIK AGLGAWVEKPTEQDQFSLT
ADAR2DD- E488Q	QLHL PQVLADAVSRLVLGKFGDLTDNFSSPHARRKVL AGVVM TGT DVKDAKVISVSTGTK CINGEYMSDRGLALNDCHAEIISRRSLLRFLYTQLELYLNNKDDQKRS IFQKSERGGFRLK

[0176]

	ENVQFHLYISTSPCGDARIFSPHEPILEEPADRHPNRKARGQLRTKIESGEGTIPVRSNAS IQTWDGVLQGERLLTMS CSDKIARWNVVGIQGSLLSIFVEPIYFSSII LGS LYHGDHLSRA MYQRISNIEDLPPLYTLNKPLLSGISNAEARQPGKAPNFSVNWTVGDSAIEVINATTGKDE LGRASRLCKHALYCRWMRVHGKVP SHLLRSKITKPNVYHESKLA AKEYQA AKARLFTAFIK AGLGAWVEKPTEQDQFSLT
AID-APOB EC1	MDSL L MNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKN GCHVELLF LRYISDWDLDPGR CYRV TWFTSWSPCYDCARHVADFLRGPNLSLRIFTARLYFCEDRKAE PEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAWEGLHENS VRLSRQLRRILLPL YEVDDL RDAFRTLGL
七鳃鳗 _AID-APO BEC1	MTDAEYVRIHEKLDIYTFKKQFFNNKKS VSHRCYVLFELKRRGERRACFWGYAVNKPQSGT ERGIHAEIFSIRKVEEYLRDNPGQFTINWYSSWSPCADCAEKILEWYNQELRGNGHTLKIW ACKLYYEKNARNQIGLWNL RDNGVGLNVMVSEHYQCCRKIFIQSSHNQLNENRWLEKTLKR AEKRRELSIMIQVKILHTTKSPAV
APOBEC1_ BE1	MSSETGPVAVDPTLRRRIEPHEFEVFVFDPRELRKETCLLYEINWGRHSIWRHTSQNTNKH VEVNFIEKFTTERTYFCPNTRCSITWFLSWSPGEC SRAITEFLSRYPHVTLFIYIARLYHH ADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYS PSNEAHWP RYPHLWVRLYVLELY CIILGLPPCLNILRRKQPQLTFFTTIALQSCHYQRLPPHILWATGLK

[0177] 所述一个或多个功能性结构域在灭活的Cas蛋白上的定位允许所述功能性结构域的正确空间取向,从而以所归属的功能性效应影响靶标。例如,如果所述功能性结构域是转录激活子(例如,VP16、VP64或p65),则将所述转录激活子放置成允许其影响所述靶标的转录的空间取向。同样地,将转录阻遏子定位成影响所述靶标的转录,并且将核酸酶(例如,FokI)定位成切割或部分切割所述靶标。在一些实施方式中,所述功能性结构域位于Cas/dCas的N-末端处。在一些实施方式中,所述功能性结构域位于Cas/dCas的C-末端处。在一些实施方式中,将灭活的CRISPR相关蛋白(dCas)修饰为包含在N-末端处的第一功能性结构域和在C-末端处的第二功能性结构域。

[0178] 与一个或多个功能性结构域融合的灭活的CRISPR相关蛋白的多种实例及其使用方法描述于例如国际公布号W0 2017/219027中,将所述文献通过引用以其全文并且特别是关于本文描述的特征并入本文。

[0179] 在一些实施方式中,VI-E型和VI-F型CRISPR-Cas效应蛋白包括上述SEQ ID NO:1-7中的任一者的氨基酸序列。在一些实施方式中,VI-E型和VI-F型CRISPR-Cas效应蛋白不包括上述SEQ ID NO:1-7中的任一者的天然存在的氨基酸序列。

[0180] 在一些实施方式中,可以不使用全长野生型(SEQ ID NO:1-7)或衍生性VI-E型和VI-F型Cas效应子,而使用其“功能性片段”。

[0181] 如本文所用,“功能性片段”是指具有小于全长序列的、SEQ ID NO:1-7中的任一者的野生型蛋白或其衍生物的片段。所述功能性片段中缺失的残基可以在N-末端、C-末端和/或内部。所述功能性片段保留野生型VI-E或VI-F Cas的至少一种功能、或其衍生物的至少一种功能。因此,功能性片段相对于所讨论的功能而特别定义。例如,其中所述功能是结合crRNA和靶RNA的能力的功能性片段,相对于RNA酶功能而言可能不是功能性片段,因为丢失Cas两端的RXXXXH基序可能不会不影响其结合crRNA和靶RNA的能力,但可能消除或破坏RNA酶活性。

[0182] 在一些实施方式中,与全长序列SEQ ID NO:1-7相比,所述VI-E型或VI-F型CRISPR-Cas效应蛋白或其衍生物或其功能性片段缺少来自N-末端的约30、60、90、120、150或约180个残基。

[0183] 在一些实施方式中,与全长序列SEQ ID NO:1-7相比,所述VI-E型或VI-F型CRISPR-Cas效应蛋白或其衍生物或其功能性片段缺少来自C-末端的约30、60、90、120或约150个残基。

[0184] 在一些实施方式中,与全长序列SEQ ID NO:1-7相比,所述VI-E型或VI-F型CRISPR-Cas效应蛋白或其衍生物或其功能性片段缺少来自N-末端的约30、60、90、120、150或约180个残基,并且缺少来自C-末端的约30、60、90、120或约150个残基。

[0185] 在一些实施方式中,所述VI-E型或VI-F型CRISPR-Cas效应蛋白或其衍生物或其功能性片段具有RNA酶活性,例如,指导/crRNA激活的特异性RNA酶活性。

[0186] 在一些实施方式中,所述VI-E型或VI-F型CRISPR-Cas效应蛋白或其衍生物或其功能性片段不具有大量的/可检测的附带RNA酶活性。

[0187] 这里,“附带RNA酶活性”是指在某些其他2类VI型RNA指导的RNA酶(如Cas13a)中观察到的非特异性RNA酶活性。包含Cas13a的复合物,例如,在通过与靶核酸(例如,靶RNA)结合而激活后,产生构象变化,这又导致所述复合物充当非特异性RNA酶,从而切割和/或降解

附近的RNA分子(例如,ssRNA或dsRNA分子)(即,“附带”效应)。

[0188] 在某些实施方式中,由(但不限于)VI-E型或VI-F型CRISPR-Cas效应蛋白或其衍生物或其功能性片段和crRNA构成的复合物在靶标识别后不会展现出附带RNA酶活性。该“无附带”实施方式可包含野生型、工程化的/衍生性效应蛋白或其功能性片段。

[0189] 在一些实施方式中,所述VI-E型或VI-F型CRISPR-Cas效应蛋白或其衍生物或其功能性片段识别并切割所述靶RNA,而无邻近原型间隔子或在原型间隔子侧翼的任何额外要求(即,原型间隔子邻近基序“PAM”或原型间隔子侧翼序列“PFS”的要求)。

[0190] 本披露还提供本文描述的CRISPR相关蛋白的拆分版本(例如,VI-E型或VI-F型CRISPR-Cas效应蛋白)。所述CRISPR相关蛋白的拆分版本可有利于递送。在一些实施方式中,将所述CRISPR相关蛋白拆分成酶的两个部分,所述酶的两个部分合在一起基本上构成有功能的CRISPR相关蛋白。

[0191] 所述拆分能以一个或多个催化结构域不受影响的方式进行。所述CRISPR相关蛋白可以作为核酸酶发挥作用,或者可以是灭活的酶,所述灭活的酶本质上是具有非常小的催化活性或没有催化活性(例如,由于其催化结构域中的一个或多个突变)的RNA结合蛋白。拆分型酶描述于例如Wright等人,“Rational design of a split-Cas9 enzyme complex[拆分型Cas9酶复合物的合理设计],”Proc.Nat’l.Acad.Sci.[美国国家科学院院刊]112(10): 2984-2989,2015中,将所述文献通过引用以其全文并入本文。

[0192] 例如,在一些实施方式中,核酸酶叶(nuclease lobe)和 α -螺旋叶(α -helical lobe)经表达为单独的多肽。尽管所述叶自身并不相互作用,但crRNA将它们募集到三元复合物中,所述复合物重现全长CRISPR相关蛋白的活性并催化位点特异性DNA切割。使用经修饰的crRNA通过防止二聚化来消除拆分型酶的活性,从而允许开发诱导型二聚化系统。

[0193] 在一些实施方式中,可以例如通过采用雷帕霉素敏感性二聚化结构域将拆分型CRISPR相关蛋白与二聚化配偶体融合。这允许生成用于对蛋白活性进行时间控制的化学诱导型CRISPR相关蛋白。因此,所述CRISPR相关蛋白可以通过拆分成两个片段而成为化学诱导性,并且雷帕霉素敏感性二聚化结构域可以用于蛋白的受控重组。

[0194] 拆分点典型地经由计算机模拟设计并克隆到构建体中。在此过程期间,可以将突变引入拆分型CRISPR相关蛋白中,并且可以去除非功能性结构域。

[0195] 在一些实施方式中,所述拆分型CRISPR相关蛋白的两个部分或片段(即,N-末端和C-末端片段)可以形成完整的CRISPR相关蛋白,其包含野生型CRISPR相关蛋白的例如至少70%、至少80%、至少90%、至少95%或至少99%的序列。

[0196] 本文描述的CRISPR相关蛋白(例如,VI-E型或VI-F型CRISPR-Cas效应蛋白)可以设计为自我激活或自我灭活。例如,可以将靶序列引入所述CRISPR相关蛋白的编码构建体中。因此,所述CRISPR相关蛋白可以切割所述靶序列以及编码所述蛋白的构建体,从而自我灭活它们的表达。构建自我灭活CRISPR系统的方法描述于例如Epstein和Schaffer, Mol. Ther. [分子疗法]24:S50,2016中,将所述文献通过引用以其全文并入本文。

[0197] 在一些其他实施方式中,在弱启动子(例如,7SK启动子)的控制下表达的额外的crRNA可以靶向编码所述CRISPR相关蛋白的核酸序列以防止和/或阻断其表达(例如,通过防止所述核酸的转录和/或翻译)。用表达所述CRISPR相关蛋白、所述crRNA、和靶向编码所述CRISPR相关蛋白的核酸的crRNA的载体转染细胞,可导致编码所述CRISPR相关蛋白的核

酸的高效破坏并降低所述CRISPR相关蛋白的水平,从而限制基因组编辑活性。

[0198] 在一些实施方式中,所述CRISPR相关蛋白的基因组编辑活性可以通过哺乳动物细胞中的内源性RNA特征(例如,miRNA)来调节。可以通过在编码所述CRISPR相关蛋白的mRNA的5'-UTR中使用miRNA互补序列来制造CRISPR相关蛋白开关。所述开关选择性地并且高效地响应靶细胞中的miRNA。因此,所述开关可以通过感应异质细胞群内的内源性miRNA活性来对基因组编辑进行差异控制。因此,开关系统可以为基于细胞内miRNA信息的细胞类型选择性基因组编辑和细胞工程化提供框架(参见例如,Hirosawa等人,Nucl.Acids Res.[核酸研究]45(13):e118,2017)。

[0199] 所述CRISPR相关蛋白(例如,VI-E型和VI-F型CRISPR-Cas效应蛋白)可以经诱导表达,例如,它们的表达可以是光诱导的或化学诱导的。这种机制允许激活所述CRISPR相关蛋白中的功能性结构域。光诱导性可以通过本领域已知的各种方法来实现,例如,通过设计如下融合复合物来实现,其中将CRY2 PHR/CIBN配对用于拆分型CRISPR相关蛋白中(参见例如,Konermann等人,“Optical control of mammalian endogenous transcription and epigenetic states[哺乳动物内源性转录和表观遗传状态的光学控制],”Nature[自然]500:7463,2013)。

[0200] 化学诱导性可以例如通过设计如下融合复合物来实现,其中将FKBP/FRB(FK506结合蛋白/FKBP雷帕霉素结合结构域)配对用于拆分型CRISPR相关蛋白中。需要雷帕霉素来形成融合复合物,从而激活所述CRISPR相关蛋白(参见例如,Zetsche等人,“A split-Cas9 architecture for inducible genome editing and transcription modulation[用于诱导型基因组编辑和转录调节的拆分型Cas9架构],”Nature Biotech.[自然生物技术]33:2:139-42,2015)。

[0201] 此外,所述CRISPR相关蛋白的表达可以通过诱导型启动子,例如四环素或强力霉素控制的转录激活(Tet-开和Tet-关表达系统)、激素诱导型基因表达系统(例如,蜕皮素诱导型基因表达系统)和阿拉伯糖诱导型基因表达系统来调节。当作为RNA递送时,RNA靶向效应蛋白的表达可以经由核糖开关进行调节,所述核糖开关可以感应小分子(像四环素)(参见例如,Goldfless等人,“Direct and specific chemical control of eukaryotic translation with a synthetic RNA-protein interaction[通过合成的RNA-蛋白相互作用对真核生物的翻译进行直接和特异性的化学控制],”Nucl.Acids Res.[核酸研究]40:9:e64-e64,2012)。

[0202] 诱导型CRISPR相关蛋白和诱导型CRISPR系统的各种实施方式描述于例如美国专利号8,871,445、美国公布号2016/0208243和国际公布号WO 2016/205764中,将各个文献通过引用以其全文并入本文。

[0203] 在一些实施方式中,所述CRISPR相关蛋白包括至少一个(例如,1、2、3、4、5、6、7、8、9或10个)衔接至所述蛋白的N-末端或C-末端的核定位信号(NLS)。NLS的非限制性实例包括源自以下的NLS序列:SV40病毒大T抗原的NLS,其具有氨基酸序列PKKKRKV;来自核质蛋白的NLS(例如,具有序列KRPAATKKAGQAKKKK的核质蛋白二分NLS);c-myc NLS,其具有氨基酸序列PAAKRVKLD或RQRRNELKRSP;hRNPA1 M9 NLS,其具有序列NQSSNFGPMKGGNFGGRSSGPYGGGGQYFAKPRNQGGY;来自输入蛋白- α 的IBB结构域的序列RMRIZFKNKGKDTAELRRRRVEVSVELRKAKKD EQILKRRNV;肌瘤T蛋白的序列VSRKRPRP和PPKKARED;人p53的序列PQPKKKPL;小鼠c-abl IV

的序列SALIKKKKKMAP;流感病毒NS1的序列DRLRR和PKQKKRK;肝炎病毒 δ 抗原的序列RKLKKKIKKL;小鼠Mx1蛋白的序列REKKKFLKRR;人聚(ADP-核糖)聚合酶的序列KRKGDEV DGVDEVAKKSKK;以及人糖皮质激素受体的序列RKCLQAGMNLEARKTKK。在一些实施方式中,所述CRISPR相关蛋白包含至少一个(例如,1、2、3、4、5、6、7、8、9或10个)附接所述蛋白的N-末端或C-末端的核输出信号(NES)。在优选的实施方式中,附接C-末端和/或N-末端NLS或NES,用于在真核细胞(例如,人细胞)中进行最佳表达和核靶向。

[0204] 在一些实施方式中,本文描述的CRISPR相关蛋白在一个或多个氨基酸残基处突变以改变一种或多种功能性活性。

[0205] 例如,在一些实施方式中,所述CRISPR相关蛋白在一个或多个氨基酸残基处突变以改变其解旋酶活性。

[0206] 在一些实施方式中,所述CRISPR相关蛋白在一个或多个氨基酸残基处突变以改变其核酸酶活性(例如,内切核酸酶活性或外切核酸酶活性)。

[0207] 在一些实施方式中,所述CRISPR相关蛋白在一个或多个氨基酸残基处突变以改变其与指导RNA功能性缔合的能力。

[0208] 在一些实施方式中,所述CRISPR相关蛋白在一个或多个氨基酸残基处突变以改变其与靶核酸功能性缔合的能力。

[0209] 在一些实施方式中,本文描述的CRISPR相关蛋白能够切割靶RNA分子。

[0210] 在一些实施方式中,所述CRISPR相关蛋白在一个或多个氨基酸残基处突变以改变其切割活性。例如,在一些实施方式中,所述CRISPR相关蛋白可以包含一个或多个突变,所述突变使酶不能切割靶核酸。

[0211] 在一些实施方式中,所述CRISPR相关蛋白能够切割与指导RNA杂交的链互补的靶核酸链。

[0212] 在一些实施方式中,本文描述的CRISPR相关蛋白可以经工程化以具有一个或多个氨基酸残基的缺失,以减小酶的大小,同时保留一种或多种所希望的功能性活性(例如,核酸酶活性和与指导RNA功能上相互作用的能力)。截短型CRISPR相关蛋白可以有利地与具有负载限制的递送系统组合使用。

[0213] 在一些实施方式中,本文描述的CRISPR相关蛋白可以与一种或多种肽标签,包括His标签、GST标签、V5标签、FLAG标签、HA标签、VSV-G标签、Trx标签或myc标签融合。

[0214] 在一些实施方式中,本文描述的CRISPR相关蛋白可以与可检测部分,例如GST、荧光蛋白(例如GFP、HcRed、DsRed、CFP、YFP或BFP)或酶(如HRP或CAT)融合。

[0215] 在一些实施方式中,本文描述的CRISPR相关蛋白可以与MBP、LexA DNA结合结构域或Gal4 DNA结合结构域融合。

[0216] 在一些实施方式中,本文描述的CRISPR相关蛋白可以与可检测标记(如荧光染料,包括FITC和DAPI)连接或缀合。

[0217] 在本文的任一实施方式中,本文描述的CRISPR相关蛋白与其他部分之间的连接可以经由共价化学键在所述CRISPR相关蛋白的N-末端或C-末端处,并且有时甚至在内部。所述连接可以通过本领域已知的任何化学连接来实现,所述化学连接例如肽连接、通过氨基酸(如D、E、S、T)的侧链或氨基酸衍生物(Ahx、 β -Ala、GABA或Ava)连接、或PEG连接。

[0218] 3. 多核苷酸

[0219] 本发明还提供核酸,所述核酸编码本文描述的蛋白(例如,CRISPR相关蛋白或辅助蛋白)和指导RNA(例如,crRNA)。

[0220] 在一些实施方式中,所述核酸是合成的核酸。在一些实施方式中,所述核酸是DNA分子。在一些实施方式中,所述核酸是RNA分子(例如,编码所述Cas、其衍生物或功能性片段的mRNA分子)。在一些实施方式中,将所述mRNA加帽、聚腺苷酸化、用5-甲基胞苷取代、用假尿苷取代、或其组合。

[0221] 在一些实施方式中,所述核酸(例如,DNA)与调节元件(例如,启动子)可操作地连接以控制所述核酸的表达。在一些实施方式中,所述启动子是组成型启动子。在一些实施方式中,所述启动子是诱导型启动子。在一些实施方式中,所述启动子是细胞特异性启动子。在一些实施方式中,所述启动子是生物特异性启动子。

[0222] 合适的启动子是本领域已知的并且包括例如pol I启动子、pol II启动子、pol III启动子、T7启动子、U6启动子、H1启动子、逆转录病毒劳斯肉瘤病毒LTR启动子、巨细胞病毒(CMV)启动子、SV40启动子、二氢叶酸还原酶启动子和 β -肌动蛋白启动子。例如,U6启动子可用于调节本文描述的指导RNA分子的表达。

[0223] 在一些实施方式中,一个或多个核酸存在于载体(例如,病毒载体或噬菌体)中。所述载体可以是克隆载体或表达载体。所述载体可以是质粒、噬菌粒、粘粒等。所述载体可以包括一个或多个允许所述载体在目的细胞(例如,细菌细胞或哺乳动物细胞)中繁殖的调节元件。在一些实施方式中,所述载体包括编码本文描述的CRISPR相关(Cas)系统的单个组分的核酸。在一些实施方式中,所述载体包括多个核酸,每个核酸编码本文描述的CRISPR相关(Cas)系统的组分。

[0224] 在一个方面,本披露提供与本文描述的核酸序列具有至少50%、55%、60%、65%、70%、75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%同一性的核酸序列,即,编码如下的核酸序列:Cas蛋白、衍生物、功能性片段、或包括SEQ ID NO:8-14的DR序列的指导/crRNA。

[0225] 在另一方面,本披露还提供编码如下氨基酸序列的核酸序列,所述氨基酸序列与本文描述的氨基酸序列(如SEQ ID NO:1-7)具有至少50%、55%、60%、65%、70%、75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%同一性。

[0226] 在一些实施方式中,所述核酸序列具有至少一部分(例如,至少1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、20、30、40、50、60、70、80、90或100个核苷酸,例如,连续或非连续核苷酸)与本文描述的序列相同。在一些实施方式中,所述核酸序列具有至少一部分(例如,至少1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、20、30、40、50、60、70、80、90或100个核苷酸,例如,连续或非连续核苷酸)与本文描述的序列不同。

[0227] 在相关的实施方式中,本发明提供如下氨基酸序列,所述氨基酸序列具有至少一部分(例如,至少1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、20、30、40、50、60、70、80、90或100个氨基酸残基,例如,连续或非连续氨基酸残基)与本文描述的序列相同。在一些实施方式中,所述氨基酸序列具有至少一部分(例如,至少1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、20、30、40、50、60、70、80、90或100个氨基酸残基,例如,连续或非连续氨基酸残基)与本文描述的序列不同。

[0228] 为了确定两个氨基酸序列或两个核酸序列的同一性百分比,出于最佳比较目的对

序列进行比对(例如,可以在第一和第二氨基酸或核酸序列的一者或两者中引入空位以用于最佳比对,并且出于比较目的可以忽略非同源序列)。一般来说,出于比较目的而比对的参考序列的长度应是参考序列长度的至少80%,并且在一些实施方式中是参考序列长度的至少90%、95%或100%。然后比较对应的氨基酸位置或核苷酸位置处的氨基酸残基或核苷酸。当第一序列中的位置被与第二序列中的对应位置相同的氨基酸残基或核苷酸占据时,则分子在该位置处是相同的。将空位的数量和每个空位的长度考虑在内,两个序列之间的同一性百分比是所述序列共享的相同位置的数量的函数,需要引入所述空位以进行所述两个序列的最佳比对。出于本披露的目的,序列的比较和两个序列之间同一性百分比的确定可以使用具有空位罚分12、空位延伸罚分4、以及移码空位罚分5的B1osum 62评分矩阵来完成。

[0229] 本文描述的蛋白(例如,CRISPR相关蛋白或辅助蛋白)可以作为核酸分子或多肽递送或使用。

[0230] 在某些实施方式中,编码所述CRISPR相关蛋白、其衍生物或功能性片段的核酸分子经密码子优化以在宿主细胞或生物中表达。所述宿主细胞可以包括已建立的细胞系(如293T细胞)或分离的原代细胞。所述核酸可以经密码子优化以用于在任何目的生物(特别是人细胞或细菌)中使用。例如,所述核酸可以针对以下进行密码子优化:任何原核生物(如大肠杆菌(*E. coli*))或任何真核生物,如人和其他非人真核生物,包括酵母、蠕虫、昆虫、植物和藻类(包括粮食作物、稻、玉米、蔬菜、水果、树木、草)、脊椎动物、鱼、非人哺乳动物(例如,小鼠、大鼠、兔子、狗、鸟(如鸡)、牲畜(母牛或牛、猪、马、绵羊、山羊等)、或非人灵长类动物)。密码子使用表易于获得,例如在www.kazusa.or.jp/codon/上可获得的“密码子使用数据库(Codon Usage Database)”中,并且这些表能以多种方式进行调整。参见Nakamura等人,*Nucl. Acids Res.* [核酸研究] 28:292, 2000(将所述文献通过引用以其全文并入本文)。用于密码子优化特定序列以在特定宿主细胞中表达的计算机算法也是可获得的,如基因制造(Gene Forge)(Aptagen公司;宾夕法尼亚州雅各布斯(Jacobus, PA))。

[0231] 在这种情况下,经密码子优化的序列的实例是经优化以在以下中表达的序列:真核生物,例如人(即,经优化以在人中表达),或如本文所讨论的另一真核生物、动物或哺乳动物;参见例如,WO 2014/093622 (PCT/US2013/074667)中的经SaCas9人密码子优化的序列。尽管这是优选的,但应理解其他实例是可能的,并且针对人以外的宿主物种的密码子优化或针对特定器官的密码子优化是已知的。一般来说,密码子优化是指在维持天然氨基酸序列的情况下通过以下方式修饰核酸序列以增强在目的宿主细胞中的表达的方法:用该宿主细胞的基因中更频繁使用或最频繁使用的密码子替代天然序列的至少一个密码子(例如,约或超过约1、2、3、4、5、10、15、20、25、50或更多个密码子)。多种物种对特定氨基酸的某些密码子展现出特定偏倚。密码子偏倚(生物之间密码子使用的差异)通常与信使RNA(mRNA)的翻译效率相关,而所述信使RNA(mRNA)的翻译效率又被认为尤其依赖于经翻译的密码子的特性和特定的转移RNA(tRNA)分子的可获得性。选定的tRNA在细胞中的优势通常反映出肽合成中最频繁使用的密码子。相应地,可以对基因进行定制以基于密码子优化在给定生物中实现最佳基因表达。密码子使用表易于获得,例如在<http://www.kazusa.or.jp/codon/>上可获得的“密码子使用数据库”中,并且这些表能以多种方式进行调整。参见Nakamura, Y.等人“Codon usage tabulated from the international DNA sequence

databases:status for the year 2000[从国际DNA序列数据库中制表的密码子使用:2000年的状态]”Nucl.Acids Res.[核酸研究]28:292(2000)。用于密码子优化特定序列以在特定宿主细胞中表达的计算机算法也是可获得的,如基因制造(Aptagen公司;宾夕法尼亚州雅各布斯)。在一些实施方式中,编码Cas的序列中的一个或多个密码子(例如,1、2、3、4、5、10、15、20、25、50或更多个或所有密码子)对应于特定氨基酸最频繁使用的密码子。

[0232] 4.RNA指导物或crRNA

[0233] 在一些实施方式中,本文描述的CRISPR系统包括至少RNA指导物(例如,gRNA或crRNA)。

[0234] 多种RNA指导物的架构是本领域已知的(参见例如,国际公布号W02014/093622和WO 2015/070083,将各个文献的全部内容通过引用并入本文)。

[0235] 在一些实施方式中,本文描述的CRISPR系统包括多种RNA指导物(例如,一种、两种、三种、四种、五种、六种、七种、八种或更多种RNA指导物)。

[0236] 在一些实施方式中,所述RNA指导物包括crRNA。在一些实施方式中,所述RNA指导物包括crRNA,但不包括tracrRNA。

[0237] 来自多个CRISPR系统的指导RNA的序列在本领域中通常是已知的,参见例如Grissa等人(Nucleic Acids Res.[核酸研究]35(网页服务器议题):W52-7,2007;Grissa等人,BMC Bioinformatics[BMC生物信息学]8:172,2007;Grissa等人,Nucleic Acids Res.[核酸研究]36(网页服务器议题):W145-8,2008;以及Moller和Liang,PeerJ[同行评审科学期刊]5:e3788,2017;在crispr.i2bc.paris-saclayfr/crispr/BLAST/CRISPRsBlast.php处的CRISPR数据库;以及在github.com/molleraj/MetaCRIST处可获得的MetaCRIST)。将所有文献通过引用并入本文。

[0238] 在一些实施方式中,所述crRNA包括同向重复(DR)序列和间隔序列。在某些实施方式中,所述crRNA包含如下同向重复序列、基本上由其组成或由其组成,所述同向重复序列与指导序列或间隔序列(优选地在所述间隔序列的3'端处)连接。

[0239] 一般来说,Cas蛋白与成熟的crRNA形成复合物,所述成熟的crRNA的间隔序列引导所述复合物与靶RNA序列特异性结合,所述靶RNA与所述间隔序列互补和/或与所述间隔序列杂交。所得的复合物包含所述Cas蛋白和与所述靶RNA结合的成熟的crRNA。

[0240] 所述Cas13e和Cas13f系统的同向重复序列通常非常保守,尤其是在末端处,在5'端处的Cas13e的GCTG和Cas13f的GCTGT与在3'端处的Cas13e的CAGC和Cas13f的ACAGC反向互补。这种保守表明潜在地与基因座中的一种或多种蛋白相互作用的RNA茎环结构的强碱基配对。

[0241] 在一些实施方式中,当在RNA中时,同向重复序列包含5'-S1a-Ba-S2a-L-S2b-Bb-S1b-3'的一般二级结构,其中区段S1a和S1b是反向互补序列并形成第一茎(S1),所述第一茎(S1)具有在Cas13e中的4个核苷酸和在Cas13f中的5个核苷酸;区段Ba和Bb不相互碱基配对,并形成对称的或接近对称的凸起(B),并且各具有在Cas13e中的5个核苷酸、以及分别在Cas13f中的5个(Ba)和4个(Bb)或6个(Ba)和5个(Bb)核苷酸;区段S2a和S2b是反向互补序列并形成第二茎(S2),所述第二茎(S2)具有在Cas13e中的5个碱基对和在Cas13f中的6或5个碱基对;并且L是在Cas13e中的8个核苷酸的环和在Cas13f中的5个核苷酸的环。参见图2。

[0242] 在某些实施方式中,S1a具有在Cas13e中的GCUG序列和在Cas13f中的GCUGU序列。

[0243] 在某些实施方式中, S2a具有在Cas13e中的GCCCC序列和在Cas13f中的A/G CCUC G/A序列(其中第一个A或G可以不存在)。

[0244] 在一些实施方式中, 所述同向重复序列包含SEQ ID NO:8-14的核酸序列或由其组成。

[0245] 如本文所用, “同向重复序列”可指CRISPR基因座中的DNA编码序列, 或指在crRNA中由其编码的RNA。因此, 当在RNA分子(如crRNA)的上下文中提到SEQ ID NO:8-14中的任一者时, 每个T应理解为代表U。

[0246] 在一些实施方式中, 所述同向重复序列包含如下核酸序列或由其组成, 所述核酸序列具有SEQ ID NO:8-14的多达1、2、3、4、5、6、7或8个核苷酸的缺失、插入或取代。在一些实施方式中, 所述同向重复序列包含如下核酸序列或由其组成, 所述核酸序列与SEQ ID NO:8-14具有至少80%、85%、90%、95%或97%的序列同一性(例如, 由于SEQ ID NO:8-14中核苷酸的缺失、插入或取代)。在一些实施方式中, 所述同向重复序列包含如下核酸序列或由其组成, 所述核酸序列与SEQ ID NO:8-14中的任一者不同, 但可以在严格杂交条件下与SEQ ID NO:8-14中的任一者的互补序列杂交, 或者可以在生理条件下与SEQ ID NO:8-14中的任一者的互补序列结合。

[0247] 在某些实施方式中, 所述缺失、插入或取代不改变SEQ ID NO:8-14的总体二级结构(例如, 茎和凸起及环的相对位置和/或大小不显著偏离原始茎、凸起和环的相对位置和/或大小)。例如, 所述缺失、插入或取代可以在所述凸起或环区中, 使得所述凸起的总体对称性大致保持相同。所述缺失、插入或取代可以在所述茎中, 使得所述茎的长度不显著偏离原始茎的长度(例如, 在两个茎的每一个中添加或缺失一个碱基对对应于总共4个碱基变化)。

[0248] 在某些实施方式中, 所述缺失、插入或取代导致衍生性DR序列, 所述衍生性DR序列可在一个或两个茎中具有±1或2个碱基对(参见图2), 在所述凸起的一条或两条单链中具有±1、2或3个碱基, 和/或在所述环区中具有±1、2、3或4个碱基。

[0249] 在某些实施方式中, 与SEQ ID NO:8-14中的任一者不同的任一上述同向重复序列保留在所述Cas13e或Cas13f蛋白中作为同向重复序列(作为SEQ ID NO:8-14的DR序列)发挥作用的能力。

[0250] 在一些实施方式中, 所述同向重复序列包含如下核酸或由其组成, 所述核酸具有SEQ ID NO:8-14中的任一者的核酸序列, 且具有初始三个、四个、五个、六个、七个或八个3'核苷酸的截短。

[0251] 在一些实施方式中, 所述Cas蛋白包含SEQ ID NO:1的氨基酸序列, 并且所述crRNA包含同向重复序列, 其中所述同向重复序列包含SEQ ID NO:8的核酸序列或由其组成。

[0252] 在一些实施方式中, 所述Cas蛋白包含SEQ ID NO:2的氨基酸序列, 并且所述crRNA包含同向重复序列, 其中所述同向重复序列包含SEQ ID NO:9的核酸序列或由其组成。

[0253] 在一些实施方式中, 所述Cas蛋白包含SEQ ID NO:3的氨基酸序列, 并且所述crRNA包含同向重复序列, 其中所述同向重复序列包含SEQ ID NO:10的核酸序列或由其组成。

[0254] 在一些实施方式中, 所述Cas蛋白包含SEQ ID NO:4的氨基酸序列, 并且所述crRNA包含同向重复序列, 其中所述同向重复序列包含SEQ ID NO:11的核酸序列或由其组成。

[0255] 在一些实施方式中, 所述Cas蛋白包含SEQ ID NO:5的氨基酸序列, 并且所述crRNA包含同向重复序列, 其中所述同向重复序列包含SEQ ID NO:12的核酸序列或由其组成。

[0256] 在一些实施方式中,所述Cas蛋白包含SEQ ID NO:6的氨基酸序列,并且所述crRNA包含同向重复序列,其中所述同向重复序列包含SEQ ID NO:13的核酸序列或由其组成。

[0257] 在一些实施方式中,所述Cas蛋白包含SEQ ID NO:7的氨基酸序列,并且所述crRNA包含同向重复序列,其中所述同向重复序列包含SEQ ID NO:14的核酸序列或由其组成。

[0258] 在经典的CRISPR系统中,指导序列(例如,crRNA)与其对应的靶序列之间的互补程度可以是约50%、60%、75%、80%、85%、90%、95%、97.5%、99%或100%。在一些实施方式中,所述互补程度是90%-100%。

[0259] 指导RNA的长度可以是约5、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、35、40、45、50、75、100、125、150、175、200或更多个核苷酸。例如,为了在功能性Cas13e或Cas13f效应蛋白、或其同源物、直系同源物、衍生物、融合物、缀合物或功能性片段中使用,间隔子可以在10-60个核苷酸、20-50个核苷酸、25-45个核苷酸、25-35个核苷酸之间,或为约27、28、29、30、31、32或33个核苷酸。然而,为了在上述任一者的dCas版本中使用,间隔子可以在10-200个核苷酸、20-150个核苷酸、25-100个核苷酸、25-85个核苷酸、35-75个核苷酸、45-60个核苷酸之间,或为约46、47、48、49、50、51、52、53、54或55个核苷酸。

[0260] 为了减少脱靶相互作用,例如,为了减少指导物与具有低互补性的靶序列相互作用,可以将突变引入所述CRISPR系统中,使得所述CRISPR系统可以区分具有大于80%、85%、90%或95%互补性的靶序列与脱靶序列。在一些实施方式中,所述互补程度为从80%至95%,例如,约83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%或95%(例如,区分具有18个核苷酸的靶标与具有1、2或3个错配的18个核苷酸的脱靶)。相应地,在一些实施方式中,指导序列与其对应的靶序列之间的互补程度大于94.5%、95%、95.5%、96%、96.5%、97%、97.5%、98%、98.5%、99%、99.5%或99.9%。在一些实施方式中,所述互补程度是100%。

[0261] 本领域已知不需要完全的互补性,前提是有足够的互补性发挥作用。可以通过引入错配(例如,间隔序列与靶序列之间的一个或多个错配,如1或2个错配(包括沿着间隔子/靶标的错配的位置))来利用对切割效率的调节。错配(例如,双错配)位于越中心的位置(即,不在3'端或5'端处),切割效率受到的影响越大。相应地,通过选择沿着所述间隔序列的错配位置,可以调节切割效率。例如,如果希望靶标切割小于100%(例如,在细胞群中),可以在所述间隔序列中引入在间隔子和靶序列之间的1或2个错配。

[0262] 已表明VI型CRISPR-Cas效应子采用多于一种RNA指导物,从而使这些效应子以及包括它们的系统和复合物能够实现靶向多个核酸的能力。在一些实施方式中,本文描述的CRISPR系统包括多种RNA指导物(例如,两种、三种、四种、五种、六种、七种、八种、九种、十种、十五种、二十种、三十种、四十种或更多种RNA指导物)。在一些实施方式中,本文描述的CRISPR系统包括单条RNA链或编码单条RNA链的核酸,其中所述RNA指导物串联排列。所述单条RNA链可以包括相同RNA指导物的多个拷贝、不同RNA指导物的多个拷贝、或其组合。本文描述的VI-E型和VI-F型CRISPR-Cas效应蛋白的加工能力使这些效应子能够靶向多个靶核酸(例如,靶RNA)而不丧失活性。在一些实施方式中,所述VI-E型和VI-F型CRISPR-Cas效应蛋白可以与针对不同靶RNA的多种RNA指导物复合进行递送。在一些实施方式中,所述VI-E型和VI-F型CRISPR-Cas效应蛋白可以与多种RNA指导物共同递送,每种RNA指导物对不同的

靶核酸具有特异性。使用CRISPR相关蛋白进行多重复合(multiplexing)的方法描述于例如美国专利号9,790,490B2和EP 3009511 B1中,将各个文献的全部内容通过引用明确并入本文。

[0263] crRNA的间隔子长度可以在约10-60个核苷酸的范围,如15-50个核苷酸、20-50个核苷酸、25-50个核苷酸或19-50个核苷酸。在一些实施方式中,指导RNA的间隔子长度为至少16个核苷酸、至少17个核苷酸、至少18个核苷酸、至少19个核苷酸、至少20个核苷酸、至少21个核苷酸、或至少22个核苷酸。在一些实施方式中,所述间隔子长度为从15至17个核苷酸(例如,15、16或17个核苷酸)、从17至20个核苷酸(例如,17、18、19或20个核苷酸)、从20至24个核苷酸(例如,20、21、22、23或24个核苷酸)、从23至25个核苷酸(例如,23、24或25个核苷酸)、从24至27个核苷酸、从27至30个核苷酸、从30至45个核苷酸(例如,30、31、32、33、34、35、36、37、38、39、40、41、42、43、44或45个核苷酸)、从30或35至40个核苷酸、从41至45个核苷酸、从45至50个核苷酸(例如,45、46、47、48、49或50个核苷酸)或更长。在一些实施方式中,所述间隔子长度为从约15至约42个核苷酸。

[0264] 在一些实施方式中,所述指导RNA的同向重复序列长度为15-36个核苷酸、为至少16个核苷酸、为从16至20个核苷酸(例如,16、17、18、19或20个核苷酸)、为20-30个核苷酸(例如,20、21、22、23、24、25、26、27、28、29或30个核苷酸)、为30-40个核苷酸(例如,30、31、32、33、34、35、36、37、38、39或40个核苷酸)、或为约36个核苷酸(例如,33、34、35、36、37、38或39个核苷酸)。在一些实施方式中,所述指导RNA的同向重复序列长度为36个核苷酸。

[0265] 在一些实施方式中,所述crRNA/指导RNA的总体长度比上文任一间隔序列长度长约36个核苷酸。例如,所述crRNA/指导RNA的总体长度可以在45-86个核苷酸、或60-86个核苷酸、62-86个核苷酸、或63-86个核苷酸之间。

[0266] 所述crRNA序列可以按以下方式修饰:允许在所述crRNA与CRISPR相关蛋白之间形成复合物并与靶标成功结合,同时不允许成功的核酸酶活性(即,没有核酸酶活性/没有导致插入缺失)。这些经修饰的指导序列称为“死crRNA”、“死指导物”或“死指导序列”。关于核酸酶活性,这些死指导物或死指导序列可以是无催化活性的或无构象活性的。死指导序列典型地比导致活性RNA切割的相应指导序列短。在一些实施方式中,死指导物比具有核酸酶活性的相应指导RNA短5%、10%、20%、30%、40%或50%。指导RNA的死指导序列的长度可以为从13至15个核苷酸(例如,长度为13、14或15个核苷酸)、长度为从15至19个核苷酸、或长度为从17至18个核苷酸(例如,长度为17个核苷酸)。

[0267] 因此,在一个方面,本披露提供非天然存在的或工程化的CRISPR系统,所述CRISPR系统包括如本文描述的功能性CRISPR相关蛋白和crRNA,其中所述crRNA包含死crRNA序列,由此所述crRNA能够与靶序列杂交,使得将所述CRISPR系统引导至细胞中的目的基因组基因座而没有可检测的核酸酶活性(例如,RNA酶活性)。

[0268] 对死指导物的详细描述例如在国际公布号WO 2016/094872中进行描述,将所述文献通过引用以其全文并入本文。

[0269] 可以生成作为诱导型系统的组分的指导RNA(例如,crRNA)。所述系统的诱导型性质允许对基因编辑或基因表达进行时空控制。在一些实施方式中,用于所述诱导型系统的刺激包括例如电磁辐射、声能、化学能和/或热能。

[0270] 在一些实施方式中,指导RNA(例如,crRNA)的转录可以通过诱导型启动子,例如四

环素或强力霉素控制的转录激活 (Tet-开和Tet-关表达系统)、激素诱导型基因表达系统 (例如,蜕皮素诱导型基因表达系统) 和阿拉伯糖诱导型基因表达系统来调节。诱导型系统的其他实例包括例如小分子双杂交转录激活系统 (FKBP、ABA等)、光诱导型系统 (光敏色素、LOV结构域或隐花色素) 或光诱导型转录效应子 (LITE)。这些诱导型系统描述于例如WO 2016205764和美国专利号8,795,965中,将所述两个文献通过引用以其全文并入本文。

[0271] 化学修饰可以应用于crRNA的磷酸骨架、糖和/或碱基。骨架修饰 (如硫代磷酸酯) 修饰磷酸骨架上的电荷并有助于寡核苷酸的递送和核酸酶抗性 (参见例如,Eckstein, “Phosphorothioates, essential components of therapeutic oligonucleotides [硫代磷酸酯:治疗性寡核苷酸的必要组分],” Nucl. Acid Ther. [核酸疗法], 24, 第374-387页, 2014); 糖的修饰 (如2'-O-甲基 (2'-OMe)、2'-F和锁核酸 (LNA)) 增强碱基配对和核酸酶抗性两者 (参见例如,Allerson等人 “Fully 2'-modified oligonucleotide duplexes with improved in vitro potency and stability compared to unmodified small interfering RNA [与未经修饰的小干扰RNA相比,完全2'修饰的寡核苷酸双链体具有改善的体外效力和稳定性],” J. Med. Chem. [药物化学杂志] 48.4:901-904, 2005)。化学修饰的碱基 (如2-硫代尿苷或N6-甲基腺苷等) 可允许更强或更弱的碱基配对 (参见例如,Bramsen等人, “Development of therapeutic-grade small interfering RNAs by chemical engineering [通过化学工程开发治疗级小干扰RNA],” Front. Genet. [遗传学前沿], 2012年8月20日; 3:154)。另外, RNA适于5'端和3'端两者与多种功能性部分 (包括荧光染料、聚乙二醇或蛋白) 缀合。

[0272] 可以对化学合成的crRNA分子应用多种修饰。例如,用2'-OMe修饰寡核苷酸以改善核酸酶抗性可以改变沃森-克里克 (Watson-Crick) 碱基配对的结合能。此外,2'-OMe修饰可以影响寡核苷酸与转染试剂、蛋白或细胞中任何其他分子相互作用的方式。这些修饰的效果可以通过经验测试来确定。

[0273] 在一些实施方式中,所述crRNA包括一个或多个硫代磷酸酯修饰。在一些实施方式中,所述crRNA包括用于增强碱基配对和/或增加核酸酶抗性目的的一个或多个锁核酸。

[0274] 这些化学修饰的汇总可见于例如Kelley等人, “Versatility of chemically synthesized guide RNAs for CRISPR-Cas9 genome editing [用于CRISPR-Cas9基因组编辑的化学合成的指导RNA的多功能性],” J. Biotechnol. [生物技术杂志] 233:74-83, 2016; WO 2016205764; 和美国专利号8,795,965B2中; 将各个文献通过引用以其全文并入。

[0275] 可以优化本文描述的RNA指导物 (例如,crRNA) 的序列和长度。在一些实施方式中, RNA指导物的优化长度可以通过鉴定crRNA的加工形式 (即,成熟的crRNA) 或通过crRNA四环的经验长度研究来确定。

[0276] 所述crRNA还可以包括一个或多个适配序列。适配子是具有特定的三维结构并可以与特定的靶分子结合的寡核苷酸或肽分子。所述适配子可以对基因效应子、基因激活子或基因阻遏子具有特异性。在一些实施方式中,所述适配子可以对蛋白具有特异性,而所述蛋白又对特定的基因效应子、基因激活子或基因阻遏子具有特异性并对其进行募集和/或与其结合。所述效应子、激活子或阻遏子能够以融合蛋白的形式存在。在一些实施方式中,所述指导RNA具有对相同的衔接蛋白具有特异性的两个或更多个适配序列。在一些实施方式中,所述两个或更多个适配序列对不同的衔接蛋白具有特异性。所述衔接蛋白可以包括

例如MS2、PP7、Q β 、F2、GA、fr、JP501、M12、R17、BZ13、JP34、JP500、KU1、M11、MX1、TW18、VK、SP、FI、ID2、NL95、TW19、AP205、 ϕ kCb5、 ϕ kCb8r、 ϕ kCb12r、 ϕ kCb23r、7s和PRR1。相应地，在一些实施方式中，所述适配子选自特异性结合如本文描述的任一种衔接蛋白的结合蛋白。在一些实施方式中，所述适配序列是MS2结合环(5'-ggcccAAC AUGAGGAUCACCCAUGUCUGCAGgggcc-3')。在一些实施方式中，所述适配序列是Q β 结合环(5'-ggcccAUGCUGUCUAAGACAGCAUgggcc-3')。在一些实施方式中，所述适配序列是PP7结合环(5'-ggcccUAAGGGUUUAUAUGGAAACCCUUAgggcc-3')。对适配子的详细描述可见于例如Nowak等人, "Guide RNA engineering for versatile Cas9 functionality[针对多种Cas9功能的指导RNA工程化]," Nucl. Acid. Res. [核酸研究], 44 (20): 9555-9564, 2016; 和WO 2016205764中, 将所述文献通过引用以其全文并入本文。

[0277] 在某些实施方式中, 所述方法利用化学修饰的指导RNA。指导RNA化学修饰的实例包括但不限于在一个或多个末端核苷酸处掺入2'-O-甲基(M)、2'-O-甲基3'-硫代磷酸酯(MS)、或2'-O-甲基3'-硫基PACE(MSP)。与未经修饰的指导RNA相比, 这样的化学修饰的指导RNA可以具有增加的稳定性和增加的活性, 尽管中靶相对于脱靶特异性是不可预测的。参见Hendel, Nat Biotechnol. [自然生物技术] 33 (9): 985-9, 2015, 将所述文献通过引用并入。化学修饰的指导RNA可进一步包括但不限于具有硫代磷酸酯键和锁核酸(LNA)核苷酸的RNA, 所述锁核酸(LNA)核苷酸包含在核糖环的2'与4'碳之间的亚甲基桥。

[0278] 本发明还涵盖用于递送多种核酸组分的方法, 其中每种核酸组分对不同的目的靶基因座具有特异性, 从而修饰多种目的靶基因座。复合物的核酸组分可以包含一个或多个蛋白结合RNA适配子。所述一个或多个适配子能够结合噬菌体外壳蛋白。所述噬菌体外壳蛋白可以选自Q β 、F2、GA、fr、JP501、MS2、M12、R17、BZ13、JP34、JP500、KU1、M11、MX1、TW18、VK、SP、FI、ID2、NL95、TW19、AP205、 ϕ Cb5、 ϕ Cb8r、 ϕ Cb12r、 ϕ Cb23r、7s和PRR1。在某些实施方式中, 所述噬菌体外壳蛋白是MS2。

[0279] 5. 靶RNA

[0280] 所述靶RNA可以是任何目的RNA分子, 包括天然存在的和工程化的RNA分子。所述靶RNA可以是mRNA、tRNA、核糖体RNA(rRNA)、微小RNA(miRNA)、干扰RNA(siRNA)、核酶、核糖开关、卫星RNA、微开关、微酶(microzyme)或病毒RNA。

[0281] 在一些实施方式中, 所述靶核酸与病症或疾病(例如, 感染性疾病或癌症)相关。

[0282] 因此, 在一些实施方式中, 本文描述的系统可用于通过靶向这些核酸来治疗病症或疾病。例如, 与病症或疾病相关的靶核酸可以是在患病细胞(例如, 癌细胞或肿瘤细胞)中过表达的RNA分子。所述靶核酸也可以是毒性RNA和/或突变的RNA(例如, 具有剪接缺陷或突变的mRNA分子)。所述靶核酸还可以是对特定微生物(例如, 致病性细菌)具有特异性的RNA。

[0283] 6. 复合物和细胞

[0284] 本发明的一个方面提供CRISPR/Cas13e或CRISPR/Cas13f复合物, 其包含(1)如本文描述的任一Cas13e/Cas13f效应蛋白、其同源物、直系同源物、融合物、衍生物、缀合物、或功能性片段, 和(2)本文描述的任一指导RNA, 每个指导RNA包括设计为与靶RNA至少部分互补的间隔序列和与所述Cas13e/Cas13f效应蛋白、其同源物、直系同源物、融合物、衍生物、缀合物、或功能性片段相容的DR序列。

[0285] 在某些实施方式中, 所述复合物进一步包含所述指导RNA结合的靶RNA。

[0286] 在某些实施方式中,所述复合物不是天然存在的。例如,所述复合物的组分中的至少一种不是天然存在的。在某些实施方式中,所述Cas13e/Cas13f效应蛋白、其同源物、直系同源物、融合物、衍生物、缀合物、或功能性片段不是天然存在的,这是由于例如与野生型蛋白相比,存在至少一个氨基酸突变(缺失、插入和/或取代)。在某些实施方式中,所述DR序列不是天然存在的,即,不是SEQ ID NO:8-14中的任一者,这是由于例如在野生型序列中添加、缺失和/或取代至少一个核苷酸碱基。在某些实施方式中,所述间隔序列不是天然存在的,因为它不存在或不由存在于以下中的任何间隔序列编码:主题Cas13e或Cas13f存在于其中的原核生物的野生型CRISPR基因座。当所述间隔序列与天然存在的噬菌体核酸不是100%互补时,它可能不是天然存在的。

[0287] 在相关方面,本发明还提供细胞,所述细胞包含本发明的任一复合物。

[0288] 在某些实施方式中,所述细胞是原核生物。

[0289] 在某些实施方式中,所述细胞是真核生物。当所述细胞是真核生物时,在所述真核细胞中的复合物可以是分离出Cas13e/Cas13f的原核生物中天然存在的Cas13e/Cas13f复合物。

[0290] 7. 使用CRISPR系统的方法

[0291] 本文描述的CRISPR系统具有多种效用,包括在多种细胞类型中修饰(例如,缺失、插入、易位、灭活或激活)靶多核苷酸或核酸。所述CRISPR系统在以下方面具有广泛的应用:例如DNA/RNA检测(例如,特异性高灵敏度酶促报告子解锁(SHERLOCK))、核酸的跟踪和标记、富集测定(从背景提取所希望的序列)、控制干扰RNA或miRNA、检测循环肿瘤DNA、制备下一代文库、药物筛选、疾病诊断和预后、以及治疗各种遗传障碍。

[0292] DNA/RNA检测

[0293] 在一个方面,本文描述的CRISPR系统可用于DNA或RNA检测中。如实施例所示,当间隔序列为约30个核苷酸时,本发明的Cas13e和Cas13f蛋白在其指导RNA依赖性特异性RNA酶活性激活后展现出非特异性/附带RNA酶活性。因此,本发明的CRISPR相关蛋白可以用CRISPR RNA(crRNA)重新编程以提供用于特定RNA感应的平台。通过选择特定的间隔序列长度,并在识别其RNA靶标后,激活的CRISPR相关蛋白参与附近非靶向的RNA的“附带”切割。这种crRNA编程的附带切割活性允许所述CRISPR系统通过触发程序性细胞死亡或通过经标记的RNA的非特异性降解来检测特定RNA的存在。

[0294] SHERLOCK方法(特异性高灵敏度酶促报告子解锁)提供基于报告RNA的核酸扩增和附带切割的具有渺摩尔(attomolar)灵敏度的体外核酸检测平台,从而允许实时检测靶标。为了实现信号检测,可以将检测与不同的等温扩增步骤组合。例如,重组酶聚合酶扩增(RPA)可以与T7转录偶联,以将扩增的DNA转化为RNA,用于后续检测。通过RPA进行扩增、T7 RNA聚合酶将扩增的DNA转录为RNA、以及通过附带RNA切割介导的报告信号释放检测靶RNA的组合称为SHERLOCK。在SHERLOCK中使用CRISPR的方法详细描述于例如Gootenberg等人“Nucleic acid detection with CRISPR-Cas13a/C2c2[用CRISPR-Cas13a/C2c2进行核酸检测],”Science[科学],2017年4月28日;356(6336):438-442中,将所述文献通过引用以其全文并入本文。

[0295] 所述CRISPR相关蛋白可用于RNA印迹测定中,所述测定使用电泳按大小分离RNA样品。所述CRISPR相关蛋白可用于特异性结合和检测靶RNA序列。所述CRISPR相关蛋白也可以

与荧光蛋白(例如,GFP)融合,并用于跟踪活细胞中的RNA定位。更特别地,可以灭活所述CRISPR相关蛋白,因为它们不再如上所述切割RNA。因此,CRISPR相关蛋白可用于确定RNA或特定剪接变体的定位、mRNA转录物的水平、转录物的上调或下调以及疾病特异性诊断。所述CRISPR相关蛋白可用于(活)细胞中的RNA的可视化,例如使用荧光显微镜检查术或流式细胞术,如荧光激活细胞分选术(FACS),其允许对细胞进行高通量筛选和回收细胞分选后的活细胞。关于如何检测DNA和RNA的详细描述可见于例如国际公布号W02017/070605中,将所述文献通过引用以其全文并入本文。

[0296] 在一些实施方式中,本文描述的CRISPR系统可用于多重抗错荧光原位杂交(multiplexed error-robust fluorescence in situ hybridization,MERFISH)。这些方法描述于例如Chen等人,“Spatially resolved,highly multiplexed RNA profiling in single cells[在单细胞中进行空间分辨的高度多重化RNA分析],”Science[科学],2015年4月24日;348(6233):aaa6090,将所述文献通过在本文中引用以其全文并入本文。

[0297] 在一些实施方式中,本文描述的CRISPR系统可用于检测样品(例如,临床样品、细胞或细胞裂解物)中的靶RNA。当间隔序列具有选择的特定长度(如约30个核苷酸)时,当本文描述的VI-E型和/或VI-F型CRISPR-Cas效应蛋白与靶核酸结合时,所述效应蛋白的附带RNA酶活性受到激活。在与目的靶RNA结合后,所述效应蛋白切割经标记的检测RNA以生成信号(例如,增加的信号或减少的信号),从而允许对样品中的靶RNA进行定性和定量检测。样品中的RNA的特异性检测和定量允许包括诊断在内的多种应用。在一些实施方式中,所述方法包括a)使样品与以下接触:(i)RNA指导物(例如,crRNA)和/或编码所述RNA指导物的核酸,其中所述RNA指导物由同向重复序列和能够与所述靶RNA杂交的间隔序列组成;(ii)VI-E型或VI-F型CRISPR-Cas效应蛋白(Cas13e或Cas13f)和/或编码所述效应蛋白的核酸;和(iii)经标记的检测RNA;其中所述效应蛋白与所述RNA指导物缔合以形成复合物;其中所述RNA指导物与所述靶RNA杂交;并且其中在所述复合物与所述靶RNA结合后,所述效应蛋白展现出附带RNA酶活性并切割所述经标记的检测RNA;以及b)测量通过所述经标记的检测RNA的切割产生的可检测信号,其中所述测量提供对所述样品中单链靶RNA的检测。在一些实施方式中,所述方法进一步包括将所述可检测信号与参考信号进行比较并确定所述样品中靶RNA的量。在一些实施方式中,所述测量使用以下进行:金纳米颗粒检测、荧光偏振、胶体相变/分散、电化学检测和基于半导体的感应。在一些实施方式中,所述经标记的检测RNA包括荧光发射染料对、荧光共振能量转移(FRET)对或猝灭剂/荧光团对。在一些实施方式中,在所述经标记的检测RNA经所述效应蛋白切割后,由所述经标记的检测RNA产生的可检测信号的量减少或增加。在一些实施方式中,所述经标记的检测RNA在经所述效应蛋白切割之前产生第一可检测信号,并且在经所述效应蛋白切割之后产生第二可检测信号。在一些实施方式中,当所述经标记的检测RNA经所述效应蛋白切割时产生可检测信号。在一些实施方式中,所述经标记的检测RNA包含经修饰的核碱基、经修饰的糖部分、经修饰的核酸连接、或其组合。在一些实施方式中,所述方法包括通过使用多个VI-E型和/或VI-F型CRISPR-Cas(Cas13e和/或Cas13f)系统,对样品中的多个独立靶RNA(例如,两个、三个、四个、五个、六个、七个、八个、九个、十个、十五个、二十个、三十个、四十个或更多个靶RNA)进行多通道检测,每个所述系统包括不同的直系同源效应蛋白和对应的RNA指导物,从而允许区分所述样品中的多个靶RNA。在一些实施方式中,所述方法包括使用VI-E型和/或VI-F型CRISPR-Cas

系统的多个实例,对样品中的多个独立靶RNA进行多通道检测,每个所述实例含有具有可区分的附带RNA酶底物的直系同源效应蛋白。使用CRISPR相关蛋白检测样品中的RNA的方法描述于例如美国专利公布号2017/0362644中,将所述文献的全部内容通过引用并入本文。

[0298] 核酸的跟踪和标记

[0299] 细胞过程依赖于蛋白、RNA和DNA间的分子相互作用网络。准确检测蛋白-DNA和蛋白-RNA相互作用是理解这样的过程的关键。体外邻近标记技术采用与报告基团(例如,可光激活基团)组合的亲亲和标签,以在体外标记目的蛋白或RNA附近的多肽和RNA。在UV辐照后,所述可光激活基团与紧邻加标签分子的蛋白和其他分子发生反应,从而标记它们。随后可回收和鉴定经标记的相互作用分子。例如,所述CRISPR相关蛋白可用于将探针靶向选定的RNA序列。这些应用也可以应用于动物模型中,用于疾病或难以培养的细胞类型的体内成像。跟踪和标记核酸的方法描述于例如美国专利号8,795,965、WO 2016205764和WO 2017070605中;将各个文献通过本文引用以其全文并入本文。

[0300] RNA分离、纯化、富集和/或耗竭

[0301] 本文描述的CRISPR系统(例如,CRISPR相关蛋白)可用于分离和/或纯化RNA。可以将所述CRISPR相关蛋白与亲和标签融合,所述亲和标签可用于分离和/或纯化RNA-CRISPR相关蛋白复合物。这些应用例如可用于分析细胞中的基因表达谱。

[0302] 在一些实施方式中,所述CRISPR相关蛋白可用于靶向特定的非编码RNA(ncRNA),从而阻断其活性。在一些实施方式中,所述CRISPR相关蛋白可用于特异性富集特定RNA(包括但不限于增加稳定性等),或替代性地,特异性耗竭特定RNA(例如,特定的剪接变体、同种型等)。

[0303] 这些方法描述于例如美国专利号8,795,965、WO 2016205764和WO 2017070605中;将各个文献通过本文引用以其全文并入本文。

[0304] 高通量筛选

[0305] 本文描述的CRISPR系统可用于制备下一代测序(NGS)文库。例如,为了创建有成本效益的NGS文库,可以使用所述CRISPR系统破坏靶基因的编码序列,并且可以通过下一代测序(例如,在离子激流(Ion Torrent)PGM系统上)同时筛选经所述CRISPR相关蛋白转染的克隆。关于如何制备NGS文库的详细描述可见于例如Bell等人,“A high-throughput screening strategy for detecting CRISPR-Cas9induced mutations using next-generation sequencing[用于使用下一代测序检测CRISPR-Cas9诱导的突变的高通量筛选策略],”BMC Genomics[BMC基因组学],15.1(2014):1002,将所述文献通过引用以其全文并入本文。

[0306] 工程化的微生物

[0307] 微生物(例如,大肠杆菌、酵母和微藻)广泛用于合成生物学。合成生物学的发展具有广泛的效用,包括各种临床应用。例如,可编程CRISPR系统可以用于拆分具有用于靶向细胞死亡的毒性结构域的蛋白,例如使用癌症关联的RNA作为靶转录物。此外,涉及蛋白-蛋白相互作用的途径可以在使用例如与适当效应子(如激酶或酶)的融合复合物的合成生物系统中受到影响。

[0308] 在一些实施方式中,可以将靶向噬菌体序列的crRNA引入微生物中。因此,本披露还提供针对噬菌体感染接种微生物(例如,生产菌株)的方法。

[0309] 在一些实施方式中,本文提供的CRISPR系统可用于对微生物进行工程化,例如以改善产率或改善发酵效率。例如,本文描述的CRISPR系统可用于对微生物(如酵母)进行工程化,以从可发酵糖生成生物燃料或生物聚合物,或降解源自作为可发酵糖的来源的农业废弃物的植物衍生的木质纤维素。更特别地,本文描述的方法可用于修饰生物燃料生产所需的内源性基因的表达和/或修饰可能干扰生物燃料合成的内源性基因。对微生物进行工程化的这些方法描述于例如Verwaal等人,“CRISPR/Cpf1 enables fast and simple genome editing of *Saccharomyces cerevisiae* [CRISPR/Cpf1能实现对酿酒酵母的快速简单的基因组编辑],”*Yeast* [酵母] doi:10.1002/yea.3278, 2017; 和Hlavova等人,“Improving microalgae for biotechnology-from genetics to synthetic biology [改善用于生物技术的微藻——从遗传学到合成生物学],”*Biotechnol. Adv.* [生物技术进展], 33:1194-203, 2015, 将所述两个文献通过引用以全文并入本文。

[0310] 在一些实施方式中,本文提供的CRISPR系统可用于诱导细胞(例如,微生物,如工程化的微生物)的死亡或休眠。这些方法可用于诱导多种细胞类型的休眠或死亡,所述细胞类型包括原核细胞和真核细胞,包括但不限于哺乳动物细胞(例如,癌细胞或组织培养细胞)、原生动物、真菌细胞、受病毒感染的细胞、受细胞内细菌感染的细胞、受细胞内原生动物感染的细胞、受朊病毒感染的细胞、细菌(例如,致病性细菌和非致病性细菌)、以及单细胞和多细胞寄生物。例如,在合成生物学领域,非常希望有控制工程化的微生物(例如,细菌)以防止它们繁殖或传播的机制。本文描述的系统可用作“杀灭开关(kill-switches)”以调节和/或防止工程化的微生物的繁殖或传播。此外,本领域需要现有抗生素治疗的替代物。本文描述的系统还可用于希望杀灭或控制特定微生物群(例如,细菌群)的应用中。例如,本文描述的系统可包括靶向属、种或株特异性的核酸(例如, RNA) 并且可以递送至细胞的RNA指导物(例如, crRNA)。在与所述靶核酸复合和结合后, VI-E型和/或VI-F型CRISPR-Cas效应蛋白的附带RNA酶活性受到激活,从而导致微生物内非靶RNA的切割,最终导致休眠或死亡。在一些实施方式中,所述方法包括使细胞与本文描述的系统接触,所述系统包括VI-E型和/或VI-F型CRISPR-Cas效应蛋白或编码所述效应蛋白的核酸、以及RNA指导物(例如, crRNA) 或编码所述RNA指导物的核酸,其中间隔序列与靶核酸(例如, 属、株或种特异性RNA指导物)的至少15个核苷酸(例如, 16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、35、40、45、50或更多个核苷酸)互补。不希望受任何特定理论的束缚,所述VI-E型和/或VI-F型CRISPR-Cas效应蛋白对非靶RNA的切割可诱导程序性细胞死亡、细胞毒性、细胞凋亡、坏死、坏死性凋亡、细胞死亡、细胞周期停滞、细胞无反应性、细胞生长减少或细胞增殖减少。例如,在细菌中,所述VI-E型和/或VI-F型CRISPR-Cas效应蛋白对非靶RNA的切割可以是抑制细菌的或杀细菌的。

[0311] 在植物中的应用

[0312] 本文描述的CRISPR系统在植物中具有多种效用。在一些实施方式中,所述CRISPR系统可用于对植物基因组进行工程化(例如,改善产量、制造具有希望的翻译后修饰的产物、或引入用于生产工业产物的基因)。在一些实施方式中,所述CRISPR系统可用于将希望的性状引入植物中(例如,对基因组进行或不进行可遗传修饰),或调节植物细胞或整株植物中内源性基因的表达。

[0313] 在一些实施方式中,所述CRISPR系统可用于鉴定、编辑和/或沉默编码特定蛋白

(例如,过敏原蛋白(例如,花生、大豆、扁豆、豌豆、四季豆和绿豆中的过敏原蛋白))的基因。关于如何鉴定、编辑和/或沉默编码蛋白的基因的详细描述在例如以下中描述:Nicolaou等人,“Molecular diagnosis of peanut and legume allergy[花生和豆类过敏的分子诊断],”*Curr.Opin.Allergy Clin.Immunol.*[过敏和临床免疫学的当前观点]11(3):222-8, 2011,和WO 2016205764 A1;将所述两个文献通过引用以全文并入本文。

[0314] 基因驱动

[0315] 基因驱动是有利地偏倚特定基因或一组基因的遗传的现象。本文描述的CRISPR系统可用于建立基因驱动。例如,所述CRISPR系统可以设计为靶向并破坏基因的特定等位基因,从而导致细胞拷贝第二等位基因以固定序列。由于所述拷贝,第一等位基因将转化为第二等位基因,从而增加所述第二等位基因传给子代的机会。关于如何使用本文描述的CRISPR系统建立基因驱动的详细方法描述于例如Hammond等人,“A CRISPR-Cas9 gene drive system targeting female reproduction in the malaria mosquito vector *Anopheles gambiae*[靶向疟疾蚊载体冈比亚按蚊中雌性生殖的CRISPR-Cas9基因驱动系统],”*Nat.Biotechnol.*[自然生物技术]34(1):78-83,2016,将所述文献通过引用以其全文并入本文。

[0316] 混合筛选(Pooled-Screening)

[0317] 如本文所述,混合CRISPR筛选是用于鉴定参与生物机制(如细胞增殖、药物抗性和病毒感染)的基因的强大工具。用本文描述的编码指导RNA(gRNA)的载体的文库批量转导细胞,并且在应用选择性激发之前和之后测量gRNA的分布。混合CRISPR筛选非常适用于影响细胞存活和增殖的机制,并且它们可以扩展至测量单个基因的活性(例如,通过使用工程化的报告细胞系)。一次只靶向一个基因的阵列CRISPR筛选使得使用RNA-seq作为读数成为可能。在一些实施方式中,如本文描述的CRISPR系统可用于单细胞CRISPR筛选中。关于混合CRISPR筛选的详细描述可见于例如Datlinger等人,“Pooled CRISPR screening with single-cell transcriptome read-out[具有单细胞转录组读数的混合CRISPR筛选],”*Nat.Methods.*[自然方法]14(3):297-301,2017,将所述文献通过引用以其全文并入本文。

[0318] 饱和诱变(过度攻击(Bashing))

[0319] 本文描述的CRISPR系统可用于原位饱和诱变。在一些实施方式中,混合指导RNA文库可用于对特定基因或调节元件进行原位饱和诱变。这样的方法可以揭示这些基因或调节元件(例如,增强子)的关键最小特征和离散脆弱性(discrete vulnerabilities)。这些方法描述于例如Canver等人,“BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis[通过Cas9介导的原位饱和诱变进行的BCL11A增强子解析],”*Nature*[自然]527(7577):192-7,2015中,将所述文献通过引用以其全文并入本文。

[0320] RNA相关应用

[0321] 本文描述的CRISPR系统可具有多种RNA相关应用,例如,调节基因表达、降解RNA分子、抑制RNA表达、筛选RNA或RNA产物、确定lincRNA或非编码RNA的功能、诱导细胞休眠、诱导细胞周期停滞、减少细胞生长和/或细胞增殖、诱导细胞无反应性、诱导细胞凋亡、诱导细胞坏死、诱导细胞死亡和/或诱导程序性细胞死亡。对这些应用的详细描述可见于例如WO 2016/205764 A1中,将所述文献通过引用以其全文并入本文。在不同的实施方式中,本文描述的方法可以在体外、在体内或离体进行。

[0322] 例如,可以将本文描述的CRISPR系统向患有疾病或障碍的受试者施用,以靶向处于患病状态中的细胞(例如,癌细胞或受感染原感染的细胞)并诱导所述细胞中的细胞死亡。例如,在一些实施方式中,本文描述的CRISPR系统可用于靶向癌细胞并诱导所述癌细胞中的细胞死亡,其中所述癌细胞来自患有以下的受试者:威尔姆斯肿瘤、尤因肉瘤、神经内分泌肿瘤、胶质母细胞瘤、神经母细胞瘤、黑素瘤、皮肤癌、乳腺癌、结肠癌、直肠癌、前列腺癌、肝癌、肾癌、胰腺癌、肺癌、胆道癌、子宫颈癌、子宫内膜癌、食道癌、胃癌、头颈癌、甲状腺髓样癌、卵巢癌、神经胶质瘤、淋巴瘤、白血病、骨髓瘤、急性淋巴细胞白血病、急性骨髓性白血病、慢性淋巴细胞白血病、慢性骨髓性白血病、霍奇金淋巴瘤、非霍奇金淋巴瘤或膀胱癌。

[0323] 调节基因表达

[0324] 本文描述的CRISPR系统可用于调节基因表达。所述CRISPR系统可以与合适的指导RNA一起使用,以经由RNA加工的控制来靶向基因表达。所述RNA加工的控制可以包括例如RNA加工反应,如RNA剪接(例如,可变剪接)、病毒复制和tRNA生物合成。与合适的指导RNA组合的RNA靶向蛋白也可用于控制RNA激活(RNAa)。RNA激活是小RNA指导的和Argonaute(Ago)依赖性基因调节现象,其中启动子靶向的短双链RNA(dsRNA)在转录/表观遗传水平上诱导靶基因表达。RNAa导致基因表达的促进,因此可以通过破坏或减少RNAa来实现对基因表达的控制。在一些实施方式中,所述方法包括使用RNA靶向CRISPR作为例如干扰核糖核酸(如siRNA、shRNA或dsRNA)的取代物。调节基因表达的方法描述于例如WO 2016205764中,将所述文献通过引用以其全文并入本文。

[0325] 控制RNA干扰

[0326] 对干扰RNA或微小RNA(miRNA)的控制可以通过减少所述干扰RNA或miRNA在体内或体外的寿命来帮助减少脱靶效应。在一些实施方式中,靶RNA可以包括干扰RNA,即,参与RNA干扰途径的RNA,如小发夹RNA(shRNA)、小干扰RNA(siRNA)等。在一些实施方式中,靶RNA包括例如miRNA或双链RNA(dsRNA)。

[0327] 在一些实施方式中,如果选择性地表达RNA靶向蛋白和合适的指导RNA(例如在空间或时间上,在受调节的启动子(例如组织或细胞周期特异性启动子)和/或增强子的控制下),则这可以用于保护细胞或系统(在体内或体外)免受那些细胞中的RNA干扰(RNAi)。在不需要RNAi的邻近组织或细胞中,或者出于对表达和不表达CRISPR相关蛋白和合适的crRNA的细胞或组织进行比较(即,分别为RNAi不受控制和受控制的情况)的目的,这可能是有用的。所述RNA靶向蛋白可用于控制或结合包含RNA或由其组成的分子,如核酶、核糖体或核糖开关。在一些实施方式中,所述指导RNA可以将所述RNA靶向蛋白募集到这些分子中,使得所述RNA靶向蛋白能够与它们结合。这些方法描述于例如WO 2016205764和WO 2017070605中,将所述两个文献通过引用以其全文并入本文。

[0328] 修饰核糖开关和控制代谢调节

[0329] 核糖开关是信使RNA的调节区段,其结合小分子并继而调节基因表达。这种机制允许细胞感应这些小分子的细胞内浓度。特定的核糖开关典型地通过改变该基因的转录、翻译或剪接来调节其相邻基因。因此,在一些实施方式中,可以通过使用与合适的指导RNA组合的RNA靶向蛋白以靶向核糖开关来控制核糖开关活性。这可以通过切割所述核糖开关或与其结合来实现。使用CRISPR系统控制核糖开关的方法描述于例如WO 2016205764和WO 2017070605中,将所述两个文献通过引用以其全文并入本文。

[0330] RNA修饰

[0331] 在一些实施方式中,本文描述的CRISPR相关蛋白可以与碱基编辑结构域,如ADAR1、ADAR2、APOBEC或激活诱导的胞苷脱氨酶(AID)融合,并且可以用于修饰RNA序列(例如,mRNA)。在一些实施方式中,所述CRISPR相关蛋白包括一个或多个突变(例如,在催化结构域中),这使得所述CRISPR相关蛋白不能切割RNA。

[0332] 在一些实施方式中,所述CRISPR相关蛋白可以与RNA结合融合多肽一起使用,所述RNA结合融合多肽包含与RNA结合结构域(如MS2(也称为MS2外壳蛋白)、Q β (也称为Q β 外壳蛋白)或PP7(也称为PP7外壳蛋白))融合的碱基编辑结构域(例如,ADAR1、ADAR2、APOBEC或AID)。所述RNA结合结构域MS2、Q β 和PP7的氨基酸序列在下文提供:

[0333] MS2 (MS2外壳蛋白)

[0334] MASNFTQFVLVDNGGTGDTVAPSNFANGVAEWISSNSRSQAYKVTCSVRQSSAQKRKYTIKVEVPKVA
TQTVGGVELPVAAWRSYLNMEITPIFATNSDCELVKAMQGLLKDGNIPIPSAIAANSGLY

[0335] Q β (Q β 外壳蛋白)

[0336] MAKLETVTGLNIGKDGKQTLVLNPRGVNPTNGVASLSQAGAVPALEKRVTVSVSQPSRNRKNYKVQVKI
QNPTACTANGSCDPSVTRQAYADVTFSFTQYSTDEERAFVRTELAALLASPLLIDAIDQLNPAY

[0337] PP7 (PP7外壳蛋白)

[0338] MSKTIVLSVGEATRTLTEIQSTADRQIFEEKVGPLVGRRLTASLRQNGAKTAYRVNLKLDQADVDCS
TSVCGELPKVRYTQVWSDVTIVANSTEASRKSLYDLTKSLVVQATSEDLVVNLVPLGR

[0339] 在一些实施方式中,所述RNA结合结构域可以与本文描述的系统的crRNA上(例如,当所述crRNA在效应子-crRNA复合物中时)的特定序列(例如,适配序列)或二级结构基序结合,从而将所述RNA结合融合多肽(其具有碱基编辑结构域)募集至所述效应子复合物中。例如,在一些实施方式中,所述CRISPR系统包括CRISPR相关蛋白、具有适配序列(例如,MS2结合环、Q β 结合环或PP7结合环)的crRNA、以及具有与RNA结合结构域融合的碱基编辑结构域的RNA结合融合多肽,所述RNA结合结构域与所述适配序列特异性结合。在该系统中,所述CRISPR相关蛋白与具有所述适配序列的crRNA形成复合物。此外,所述RNA结合融合多肽与所述crRNA结合(经由所述适配序列),从而形成可以修饰靶RNA的三分复合物(tripartite complex)。

[0340] 使用CRISPR系统进行碱基编辑的方法描述于例如国际公布号W02017/219027中,将所述文献通过引用以其全文并且特别是关于其对RNA修饰的讨论并入本文。

[0341] RNA剪接

[0342] 在一些实施方式中,本文描述的灭活的CRISPR相关蛋白(例如,在催化结构域中具有一个或多个突变的CRISPR相关蛋白)可用于靶向RNA转录物上的特定剪接位点并与其结合。所述灭活的CRISPR相关蛋白与RNA的结合可在空间上抑制剪接体与转录物的相互作用,从而能够改变特定转录物同种型的生成频率。这样的方法可用于通过外显子跳跃(exon skipping)来治疗疾病,使得可以在成熟的蛋白中跳过具有突变的外显子。使用CRISPR系统改变剪接的方法描述于例如国际公布号W0 2017/219027中,将所述文献通过引用以其全文并且特别是关于其对RNA剪接的讨论并入本文。

[0343] 治疗性应用

[0344] 本文描述的CRISPR系统可以具有多种治疗性应用。这样的应用可基于主题

CRISPR/Cas13e或Cas13f系统的以下一种或多种体外和体内能力:诱导细胞衰老、诱导细胞周期停滞、抑制细胞生长和/或增殖、诱导细胞凋亡、诱导坏死等。

[0345] 在一些实施方式中,新的CRISPR系统可用于治疗多种疾病和障碍,例如遗传障碍(例如,单基因疾病)、可通过核酸酶活性(例如,Pcsk9靶向或BCL11a靶向)治疗的疾病(杜氏肌营养不良(DMD))、以及多种癌症等。

[0346] 在一些实施方式中,本文描述的CRISPR系统可用于编辑靶核酸以修饰所述靶核酸(例如,通过插入、缺失或突变一个或多个核酸残基)。例如,在一些实施方式中,本文描述的CRISPR系统包含外源性供体模板核酸(例如,DNA分子或RNA分子),所述外源性供体模板核酸包含所希望的核酸序列。在解决用本文描述的CRISPR系统诱导的切割事件后,细胞的分子机制将利用所述外源性供体模板核酸修复和/或解决切割事件。替代性地,细胞的分子机制可以利用内源性模板修复和/或解决切割事件。在一些实施方式中,本文描述的CRISPR系统可用于改变靶核酸,从而导致插入、缺失和/或点突变。在一些实施方式中,所述插入是无痕插入(即,在解决切割事件后将预期的核酸序列插入靶核酸中未导致额外的非预期的核酸序列)。供体模板核酸可以是双链或单链核酸分子(例如,DNA或RNA)。设计外源性供体模板核酸的方法描述于例如国际公布号WO 2016/094874 A1中,将所述文献的全部内容通过引用明确并入本文。

[0347] 在一个方面,本文描述的CRISPR系统可用于治疗由RNA、毒性RNA和/或突变RNA(例如,剪接缺陷或截短)的过表达引起的疾病。例如,毒性RNA的表达可以与核包涵体的形成以及脑、心脏或骨骼肌的迟发型退行性变化相关。在一些实施方式中,所述障碍是强直性肌营养不良。在强直性肌营养不良中,所述毒性RNA的主要致病作用是隔离(sequester)结合蛋白并损害可变剪接的调节(参见例如,Osborne等人,“RNA-dominant diseases[RNA显性疾病],”Hum.Mol.Genet.[人类分子遗传学],2009年4月15日;18(8):1471-81)。遗传学家对强直性肌营养不良(营养不良性肌强直(DM))特别感兴趣,因为它产生极其广泛的临床特征。DM的经典形式(现在称为1型DM(DM1))由编码细胞溶质蛋白激酶的基因DMPK的3'-非翻译区(UTR)中CTG重复序列的扩增引起。如本文描述的CRISPR系统可以靶向过表达的RNA或毒性RNA,例如DMPK基因或DM1骨骼肌、心脏或脑中的任一错误调节的可变剪接。

[0348] 本文描述的CRISPR系统还可以靶向影响RNA依赖性功能的反式作用突变,所述突变导致多种疾病,例如像普拉德-威利综合征(Prader Willi syndrome)、脊髓性肌萎缩(SMA)和先天性角化不良。可以使用本文描述的CRISPR系统治疗的疾病列表汇总于Cooper等人,“RNA and disease[RNA和疾病],”Cell[细胞],136.4(2009):777-793和WO 2016/205764 A1中,将所述两个文献通过引用以全文并入本文。本领域的技术人员将理解如何使用新的CRISPR系统治疗这些疾病。

[0349] 本文描述的CRISPR系统还可用于治疗各种tau蛋白病,包括例如原发性和继发性tau蛋白病,如原发性年龄相关性tau蛋白病(PART)/神经原纤维缠结(NFT)优势型老年性痴呆(其中NFT类似于在阿尔茨海默病(AD)中见到的那些,但没有斑块)、拳击性痴呆(慢性创伤性脑病)和进行性核上性麻痹。tau蛋白病的可用列表和治疗这些疾病的方法描述于例如WO 2016205764中,将所述文献通过引用以其全文并入本文。

[0350] 本文描述的CRISPR系统还可用于靶向破坏顺式作用剪接代码的突变,所述突变可导致剪接缺陷和疾病。这些疾病包括例如,由SMN1基因的缺失导致的运动神经元退行性疾

病(例如,脊髓性肌萎缩)、杜氏肌营养不良(DMD)、17号染色体相关的额颞叶痴呆合并帕金森综合征(FTDP-17)、以及囊性纤维化。

[0351] 本文描述的CRISPR系统可进一步用于抗病毒活性,特别是抗RNA病毒。所述CRISPR相关蛋白可以使用经选择以靶向病毒RNA序列的合适的指导RNA来靶向病毒RNA。

[0352] 本文描述的CRISPR系统还可用于在受试者(例如,人受试者)中治疗癌症。例如,本文描述的CRISPR相关蛋白可以用靶向RNA分子的crRNA编程,所述RNA分子是异常的(例如,包含点突变或者经可变剪接)并见于癌细胞中,以诱导癌细胞中的细胞死亡(例如,经由细胞凋亡)。

[0353] 本文描述的CRISPR系统还可用于在受试者(例如,人受试者)中治疗自身免疫疾病或障碍。例如,本文描述的CRISPR相关蛋白可以用靶向RNA分子的crRNA编程,所述RNA分子是异常的(例如,包含点突变或者经可变剪接)并见于负责引起自身免疫疾病或障碍的细胞中。

[0354] 此外,本文描述的CRISPR系统还可用于在受试者中治疗感染性疾病。例如,本文描述的CRISPR相关蛋白可以用靶向RNA分子的crRNA编程,所述RNA分子由感染原(例如,细菌、病毒、寄生物或原生动物)表达,以靶向并诱导感染原细胞中的细胞死亡。所述CRISPR系统还可用于治疗细胞内感染原感染宿主受试者细胞的疾病。通过对所述CRISPR相关蛋白进行编程以靶向由感染原基因编码的RNA分子,可以靶向受感染原感染的细胞并诱导细胞死亡。

[0355] 此外,体外RNA感应测定可用于检测特定RNA底物。所述CRISPR相关蛋白可用于活细胞中基于RNA的感应。应用的实例是通过感应例如疾病特异性RNA进行的诊断。

[0356] 本文描述的CRISPR系统的治疗性应用的详细描述可见于例如美国专利号8,795,965、EP 3009511、WO 2016205764和WO 2017070605中;将各个文献通过引用以其全文并入本文。

[0357] 细胞及其后代

[0358] 在某些实施方式中,本发明的方法可用于将本文描述的CRISPR系统引入细胞中,并引起所述细胞和/或其后代改变一种或多种细胞产物(如抗体、淀粉、乙醇、或任何其他所希望的产物)的产生。这样的细胞及其后代在本发明的范围内。

[0359] 在某些实施方式中,本文描述的方法和/或CRISPR系统导致细胞的一种或多种RNA产物的翻译和/或转录的修饰。例如,所述修饰可导致RNA产物的转录/翻译/表达增加。在其他实施方式中,所述修饰可导致RNA产物的转录/翻译/表达降低。

[0360] 在某些实施方式中,所述细胞是原核细胞。

[0361] 在某些实施方式中,所述细胞是真核细胞,如哺乳动物细胞,包括人细胞(原代人细胞或已建立的人细胞系)。在某些实施方式中,所述细胞是非人哺乳动物细胞,如来自非人灵长类动物(例如,猴)、母牛/公牛/牛、绵羊、山羊、猪、马、狗、猫、啮齿动物(如兔子、小鼠、大鼠、仓鼠等)的细胞。在某些实施方式中,所述细胞来自鱼(如鲑鱼)、鸟(如禽鸟,包括鸡、鸭、鹅)、爬行动物、贝类(例如,牡蛎、蛤蜊、龙虾、对虾)、昆虫、蠕虫、酵母等。在某些实施方式中,所述细胞来自植物,如单子叶植物或双子叶植物。在某些实施方式中,所述植物是粮食作物,如大麦、木薯、棉花、落花生或花生、玉蜀黍、小米、油棕果、马铃薯、干豆、油菜籽或低芥酸菜籽(canola)、稻、黑麦、高粱、大豆、甘蔗、甜菜、向日葵和小麦。在某些实施方式中,所述植物是谷类(大麦、玉蜀黍、小米、稻、黑麦、高粱和小麦)。在某些实施方式中,所述

植物是块茎(木薯和马铃薯)。在某些实施方式中,所述植物是糖料作物(甜菜和甘蔗)。在某些实施方式中,所述植物是含油作物(大豆、落花生或花生、油菜籽或低芥酸菜籽、向日葵和油棕果)。在某些实施方式中,所述植物是纤维作物(棉花)。在某些实施方式中,所述植物是树木(如桃树或油桃树、苹果树或梨树、坚果树(如杏仁树或核桃树或开心果树)、或柑橘树(例如,橙树、葡萄柚树或柠檬树))、草、蔬菜、水果或藻类。在某些实施方式中,所述植物是茄属植物;芸苔属(*Brassica*)植物;莴苣属(*Lactuca*)植物;菠菜属(*Spinacia*)植物;辣椒属(*Capsicum*)植物;棉花、烟草、芦笋、胡萝卜、卷心菜、西兰花、花椰菜、番茄、茄子、胡椒、生菜、菠菜、草莓、蓝莓、覆盆子、黑莓、葡萄、咖啡、可可等。

[0362] 相关方面提供使用本文描述的CRISPR系统通过本发明的方法修饰的细胞或其后代。

[0363] 在某些实施方式中,所述细胞在体外、在体内或离体进行修饰。

[0364] 在某些实施方式中,所述细胞是干细胞。

[0365] 8. 递送

[0366] 通过本披露和本领域的知识,本文描述的CRISPR系统或本文描述的其任一组分(Cas蛋白、其衍生物、功能性片段或各种融合物或加合物,以及指导RNA/crRNA)、其核酸分子、和/或编码或提供其组分的核酸分子可以通过各种递送系统(如载体,例如质粒和病毒递送载体)使用本领域中任何合适的手段递送。这样的方法包括(但不限于)电穿孔、脂质转染、显微注射、转染、超声、基因枪等。

[0367] 在某些实施方式中,所述CRISPR相关蛋白和/或任一RNA(例如,指导RNA或crRNA)和/或辅助蛋白可以使用合适的载体递送,所述载体例如质粒或病毒载体(如腺相关病毒(AAV)、慢病毒、腺病毒、逆转录病毒载体和其他病毒载体、或其组合)。可以将所述蛋白和一种或多种crRNA包装到一种或多种载体(例如,质粒或病毒载体)中。对于细菌应用,可以使用噬菌体将编码本文描述的CRISPR系统的任一组分的核酸递送至细菌。示例性噬菌体包括但不限于T4噬菌体、Mu、λ噬菌体、T5噬菌体、T7噬菌体、T3噬菌体、Φ29、M13、MS2、Qβ和ΦX174。

[0368] 在一些实施方式中,通过例如肌内注射、静脉内施用、经皮施用、鼻内施用、口服施用或粘膜施用将所述载体(例如,质粒或病毒载体)递送至目的组织。这样的递送可以经由单剂量或多剂量进行。本领域技术人员应理解,本文待递送的实际剂量可取决于多种因素而大幅变化,如载体选择、靶细胞、生物、组织、待治疗受试者的一般状况、所寻求的转化/修饰的程度、施用途径、施用模式、所寻求的转化/修饰的类型等。

[0369] 在某些实施方式中,所述递送经由腺病毒进行,其可以是含有至少 1×10^5 个颗粒(也称为颗粒单位,pu)的腺病毒的单剂量。在一些实施方式中,所述剂量优选地是至少约 1×10^6 个颗粒、至少约 1×10^7 个颗粒、至少约 1×10^8 个颗粒、和至少约 1×10^9 个颗粒的腺病毒。所述递送方法和所述剂量描述于例如WO 2016205764A1和美国专利号8,454,972B2中,将所述两个文献通过引用以全文并入本文。

[0370] 在一些实施方式中,所述递送经由质粒进行。所述剂量可以是足够数量的质粒以引发响应。在一些情况下,质粒组合中质粒DNA的合适量可以是从约0.1至约2mg。质粒将通常包括(i)启动子;(ii)编码靶向核酸的CRISPR相关蛋白和/或辅助蛋白的序列,每个序列与启动子(例如,相同的启动子或不同的启动子)可操作地连接;(iii)可选择标志物;

(iv) 复制起点; 以及 (v) 位于 (ii) 的下游并与其可操作地连接的转录终止子。质粒还可以编码CRISPR复合物的RNA组分, 但这些组分中的一种或多种可以替代地在不同的载体上编码。施用频率在医学或兽医学从业者 (例如, 医师、兽医师) 或本领域技术人员的范围内。

[0371] 在另一实施方式中, 所述递送经由脂质体或脂质转染配制品等进行, 并且可以通过本领域技术人员已知的方法制备。这样的方法描述于例如WO 2016205764和美国专利号5,593,972、5,589,466、和5,580,859中, 将各个文献通过引用以其全文并入本文。

[0372] 在一些实施方式中, 所述递送经由纳米颗粒或外泌体进行。例如, 已表明外泌体在递送RNA方面特别有用。

[0373] 将新的CRISPR系统的一种或多种组分引入细胞中的另外的手段是通过使用细胞穿透肽 (CPP)。在一些实施方式中, 细胞穿透肽与所述CRISPR相关蛋白连接。在一些实施方式中, 所述CRISPR相关蛋白和/或指导RNA与一种或多种CPP偶联以有效地将它们转运到细胞 (例如, 植物原生质体) 内。在一些实施方式中, 所述CRISPR相关蛋白和/或一种或多种指导RNA由一种或多种环状或非环状DNA分子编码, 所述环状或非环状DNA分子与一种或多种CPP偶联用于细胞递送。

[0374] CPP是少于35个氨基酸的短肽, 所述短肽源自能够以非受体依赖性方式跨细胞膜转运生物分子的蛋白或嵌合序列。CPP可以是阳离子肽、具有疏水性序列的肽、两亲性肽、具有富含脯氨酸且抗微生物的序列的肽、以及嵌合肽或二分肽。CPP的实例包括例如Tat (其是1型HIV病毒复制所需的核转录激活蛋白)、穿膜肽、卡波西成纤维细胞生长因子 (FGF) 信号肽序列、整合素 β 3信号肽序列、聚精氨酸肽Args序列、富含鸟嘌呤的分子转运蛋白和甜箭肽。CPP和使用它们的方法描述于例如Hällbrink等人, “Prediction of cell-penetrating peptides [细胞穿透肽的预测],” *Methods Mol. Biol.* [分子生物学方法], 2015; 1324: 39-58; Ramakrishna等人, “Gene disruption by cell-penetrating peptide-mediated delivery of Cas9 protein and guide RNA [通过细胞穿透肽介导的Cas9蛋白和指导RNA的递送来破坏基因],” *Genome Res.* [基因组研究], 2014年6月; 24 (6) : 1020-7; 以及WO 2016205764 A1中; 将各个文献通过引用以其全文并入本文。

[0375] 用于本文描述的CRISPR系统的各种递送方法还描述于例如美国专利号8,795,965、EP 3009511、WO 2016205764和WO 2017070605中; 将各个文献通过引用以其全文并入本文。

[0376] 9. 试剂盒

[0377] 本发明的另一方面提供试剂盒, 所述试剂盒包含本文描述的主题CRISPR/Cas系统的任意两种或更多种组分, 如Cas13e和Cas13f蛋白、其衍生物、功能性片段或各种融合物或加合物、指导RNA/crRNA、其复合物、涵盖它们的载体、或涵盖它们的宿主。

[0378] 在某些实施方式中, 所述试剂盒进一步包括使用其中涵盖的组分的说明, 和/或与可在别处获得的其他组分组合的说明。

[0379] 在某些实施方式中, 所述试剂盒进一步包含一种或多种核苷酸, 例如对应于以下的一种或多种核苷酸: 可用于将指导RNA编码序列插入载体中并将所述编码序列与所述载体的一种或多种控制元件可操作地连接的那些。

[0380] 在某些实施方式中, 所述试剂盒进一步包含一种或多种缓冲液, 所述缓冲液可用于溶解任一所述组分和/或为一种或多种所述组分提供合适的反应条件。这样的缓冲液可

以包括以下中的一种或多种:PBS、HEPES、Tris、MOPS、 Na_2CO_3 、 NaHCO_3 、NaB、或其组合。在某些实施方式中,所述反应条件包括适当的pH,如碱性pH。在某些实施方式中,所述pH在7-10之间。

[0381] 在某些实施方式中,任一种或多种所述试剂盒组分可以储存在合适的容器中。

[0382] 实施例

[0383] 实施例1新型Cas13e和Cas13f系统的鉴定

[0384] 使用了计算流水线技术从基因组和宏基因组来源产生2类CRISPR-Cas系统的扩展数据库。从以下下载了基因组和宏基因组序列:NCBI (Benson等人,2013;Pruitt等人,2012)、NCBI全基因组测序(WGS)、和DOE JGI整合微生物基因组(DOE JGI Integrated Microbial Genomes) (Markowitz等人,2012)。在长度至少5kb的所有重叠群上预测了蛋白(Prodigal (Hyatt等人,2010),匿名模式(anon mode)),并对其进行了去重(即,去除相同的蛋白序列)以构建完整的蛋白数据库。认为大于600个残基的蛋白是大蛋白(LP)。由于目前鉴定出的Cas13蛋白的大小大多数都大于900个残基,为了降低计算的复杂度,只进一步考虑了大蛋白。

[0385] 使用Piler-CR (Edgar, PILER-CR:Fast and accurate identification of CRISPR repeats [PILER-CR:CRISPR重复序列的快速准确鉴定]. BMC Bioinformatics [BMC 生物信息学] 8:18, 2007), 使用所有默认参数鉴定了CRISPR阵列。将位于距CRISPR阵列±10kb内的编码非冗余大蛋白序列的ORF分组为CRISPR近端大蛋白编码簇,并将编码的LP定义为Cas-LP。

[0386] 首先,使用了BLASP在Cas-LP之间进行成对比对,并获得了E值<1E-10的BLASTP比对结果。然后基于所述BLASTP结果使用了MCL进一步聚集Cas-LP,以创建Cas蛋白家族。

[0387] 接下来,使用了BLASTP将Cas-LP与所有LP进行比对,并获得了E值<1E-10的BLASP比对结果。根据所述BLASTP比对结果进一步扩展了Cas-LP家族。获得了所述Cas-LP家族用于进一步分析,扩展后增加不超过一倍。

[0388] 对于候选Cas蛋白的功能性表征,使用了蛋白家族数据库Pfam (Finn等人,2014)、NR数据库、和NCBI中的Cas蛋白注释所述候选Cas蛋白。然后使用MAFFT (Katoh和Standley, 2013) 对每个候选Cas效应蛋白进行了多序列比对。然后使用了JPred和HHpred分析这些蛋白中的保守区,以鉴定具有两个保守RXXXXH基序的候选Cas蛋白/家族。

[0389] 该分析导致鉴定出了七种新型Cas13效应蛋白,它们属于两个新的Cas13家族,不同于先前鉴定出的所有2类CRISPR-Cas系统。这些包括新的Cas13e家族的Cas13e.1 (SEQ ID NO:1) 和Cas13e.2 (SEQ ID NO:2),以及新的Cas13f家族的Cas13f.1 (SEQ ID NO:3)、Cas13f.2 (SEQ ID NO:4)、Cas13f.3 (SEQ ID NO:5)、Cas13f.4 (SEQ ID NO:6) 和Cas13f.5 (SEQ ID NO:7)。

[0390] MAQVSKQTSKKRELSIDEYQGARKWCFTIAFNKALVNRDKNDGLFVESLLRHEKYSKHDWYDEDTRALIKCSTQAANAKAEALRNYFSHYRHSPGCLTFTADELRTIMERAYERAIFECRRRETEVIEFPSLFEGDRITTAGVVFVVSFFVERRVLDRLYGAVSGLKKNEGQYKLTRKALSMYCLKDSRFTKAWDKRVLLFRDILAQLGRIPAEAYEY YHGEQGDKKRANDEGTNPKRHKDKFIEFALHYLEAQHSEICFGRRHIVREEAGAGDEHKKHRTKGKVVVDFSKKD EDQSYIISKNNVIVRIDKNAGPRSYMGLNELKYLVLSSLQKGDDAIAKLYRYRQHVENILDVVKVTDKDNHVFLLPRFVLEQHGIGRKAFKQRIDGRVKHVRGVWEKKKAATNEMTLHEKARDILQYVNENCTRSFNPGEYNRLLVCLVGK

DVENFQAGLKRQLAERIDGRVYSIFAQTSTINEMHQVVCDQILNRLCRIGDQKLYDYVGLGKKDEIDYKQKVAWF
KEHISIRRGFLRKKFWYDSKKGFAKLVEEHLESGGQRDVGLDKKYYHIDAIGRFEGANPALYETLARDRLCLMMA
QYFLGSRVKELGNKIVWSNDSIELPVEGSGVNEKSIVFSVSDYGKLYVLDDAEFLGRICEYFMPHEKGKIRYHTVY
EKGFRAYNDLQKKCVEAVLAFEEKVVKAKKMSEKEGAHYIDFREILAQTMCKEAEKTAVNKVRRAFFHHHLKFVID
EFGLFSDVMKKYKIEKEWKFPVK* (SEQ ID NO:1)

[0391] MKVENIKEKSKKAMYLINHYEGPKKWCFAIVLNRACDNYEDNPHLFSSLSLEFEKTSRKDWFEETRE
LVEQADTEIQPNPNLKPNTTANRKLKDIRNYFSHHYHKNCLYFKNDPIRCIMEAAEYKSKIYIKGKQIEQSDIP
LPELFESSGWITPAGILLASFFVERGILHRLMGNIGGFKDNRGEYGLTHDIFTTYCLKGSYSIRAQDHDAMFRD
ILGYLSRVPTESFQRIKQPQIRKEGQLSERKTDKFITFALNYLEDYGLKDLEGCKACFARSKIVREQENVESINDK
EYKPHENKKKVEIHFDQSKEDRFYINRNNVILKIQKKDGHSNIVRMGVYELKYLVLMSLVGKAKEAVEKIDNYIQD
LRDQLPYIEGKNKEEIKKEYVRFFPRFIRSHLGLLQINDEEKIKARLDYVKTWLDKKEKSKELELHKKGRDILRYI
NERCDRELNRNVYNRIELLEVSVDLTGFYRELEELKRTRRIDKNIVQNLSGQKTINALHEKVCDLVLKEIESLDTE
NLRKYLGLIPKEEKEVTFKEKVDRILKQPVIIYKGLRYQFFKDDKKSFVLLVEDALKEGGGCDVPLGKEYYKIVS
LDKYDKENKTL CETLAMDRCLMMARQYYLSLNAKLAQEAQQIEWKKEDSIELIIFTLKNPDQSKQSFSIRFSVRD
FTKLYVTDDPEFLARLCSYFFPVEKEIEYHKL YSEGINKYTNLQKEGIEAILELEKKLIERNRIQSAKNYLSFNEI
MNKSGYNKDEQDDLKKVRNSLLHYKLIFEKEHLKKFYEVMRGEGIEKKWSLIV* (SEQ ID NO:2)

[0392] MNGIELKKEEAIFYFNQAEINLKAIEDNIFDKERRKTLLNNPQILAKMENFIFNFRDVTNAKGEIDC
LLLKLRELNRNFYSHYVHKRDVRELSKGEKPILEKYYQFAIESTGSENVKLEI IENDAWLADAGVLFFLCIFLKKSQ
ANKLISGISGFKRNDGTGQPRRNLFTYFSIREGYKVPPEMQKHLLFSLVNHLSNQDDYIEKAHQPYDIGELFFH
RIASTFLNISGILRNMKFYTYQSKRLVEQRGELKREKDIFAWEEPFQGN SYFEINGHKG VIGEDELKELCYAFLIG
NQDANKVEGRITQFLEKFRNANSVQQVKDDEMLKPEYFPANYFAESGVGRIDRVLNRLNKAIKSNKAKKGEI IAY
DKMREVMFINSPLVDEKLKPKDYKRYLGMVRFWDREKDNKREFETKEWSKYLP SNFWTAKNLERVYGLAREKN
AELFNKLKADVEKMDERELEKYQKINDAKDLANLRLASDFGVKWEKDWDEYSGQIKKQITDSQKLTIMKQRITA
GLKKKHG IENLNL RITIDINKSRKAVLNRIAIPRGFVKRHILGWQSEKVSKKIREAECEILLSKEYEELSKQFFQ
SKDYDKMTRINGLYEKNKLIALMAVYLMGQLRILFKEHTKLDDITKT TVDFKISDKVTVKIPFSNYP SLVYTMSSK
YVDNIGNYGF SNKDKDKPILGKIDVIEKQRM EFIKEVLGF EKYLFDDKIIDKSKFADTATHISFAEIVEELVEKGW
DKDRLTKLKDARNKALHGEILTGTSFDETKSLINELKK* (SEQ ID NO:3)

[0393] MSPDFIKLEKQEAIFYFNQTELNLKAIESNILDKQRMILLNNPRILAKVGNFIFNFRDVTNAKGEI
DCLLFKLEELNRNFYSHYVHTDNVKELSNGEKPLLERYYYQIAIQATRS EDVKFELFETR NENKITDAGVLFFLCMFL
KKSQANKLISGISGFKRNDPTGQPRRNLFTYFSAREGYKALPDMQKHLLFTLVNYLSNQDEYISELKQYGEIGQG
AFFNRIASTFLNISGISGNTKFYSYQSKRIKEQRGELNSEKDSFEWIEPFQGN SYFEINGHKG VIGEDELKELCYA
LLVAKQDINAVEGKIMQFLKKFRNTGNLQQVKDDEMLEIEYFPASYFNESKKEDIKKEILGRLDKKIRSCSAKAEK
AYDKMKEVMEFINNSLPAAEKLKRKYRRYLKMVRFSREKGNIEREFRTKEWSKYFSSDFWRKNNLEDVYKLATQ
KNAELFKNLKAAAEKMGTEFEKYQQINDVKDLASLRLTQDFGLKWEKDWEEYSEQIKKQITDRQKLTIMKQRV
TAE LKKKHG IENLNL RITIDSNSRKAVLNRIAIPRGFVKKHILGWQSEKISKNI REAECKILLSKKYEELSRQF
FEAGNFDKLTQINGLYEKNKL TAFMSVYLMGR LNIQLNKHTELGNLKKTEVDFKISDKVTEKIPFSQYPSLVYAMS
RKYVDNVDKYKFSHQDKKKPFLGKIDSIEKERIEFIKEVLDFEEYLFKNKVIDKSKFSDTATHISFKEICDEMGGK
GCNRNKL TELNNARNAALHGEIPSETSFREAKPLINELKK* (SEQ ID NO:4) _MSPDFIKLEKQEAIFYFNQTEL
NLKAIESNIFDKQQRVILLNNPQILAKVGDFIFNFRDVTNAKGEIDCLLLKLRELNRNFYSHYVYTDDVKILSNGE

RPLLEKYYQFAIEATGSENVKLEIIESNNRLTEAGVLFFLCMFLKKSQANKLISGISGFKRNDPTGQPRRNLFITYF
SVREGYKVVPMQKHFLFLVNLHLSGQDDYIEKAQKPYDIGEGLFFHRIASTFLNISGILRNMEFYIYQSKRLKE
QQGELKREKDIFPWIEPFQGNSYFEINGNKGIIGEDELKELCYALLVAGKDVRAVEGKITQFLEKFNADNAQQVE
KDEMLDRNNFPANYFAESNIGSIKEKILNRLGKTDDSYNKTGTIKIPYDMMKEVMEFINNSLPADKLRKDYRRY
LKMVRIWDSEKDNIREFESKEWSKYFSSDFWMAKNLERYVGLAREKNAELFNKLKAVVEKMDEREFEKYRLINSA
EDLASLRRLAKDFGLKWEKDWQEYSGQIKKQISDRQKLTIMKQRITAEKKKKHGIENLNLRTIDSNSRKAVALN
RIAVPRGFVKEHILGWQSEKVSCKTREAACKILLSKEYEELSKQFFQTRNYDKMTQVNGLYEKNKLLAFMVVYLM
ERLNILLNKPTELNELEKAEVDFKISDKVMAKIPFSQYPSLVYAMSSKYADSVGSYKFENDEKNKPFLGKIDTIEK
QRMEFIKEVLGFEEYLFEEKIIDKSEFADTATHISFDEICNELIKKGWDKDKLTKLKDARNAALHGEIPAETSFRE
AKPLINGLKK* (SEQ ID NO:5)

[0394] MNIIKLKKEEAAFYFNQITILNLSGLDEIEEKQIPHIIISNKENAKKVIDKIFNNRLLLKSVENYIYNFK
DVAKNARTEIEAILLKLVELRNFYSHYVHNDTVKILSNGEKPILEKYYQIAIEATGSKNVKLVIENNCLTDSGV
LFLLCMFLKKSQANKLISSVSGFKRNDKEGQPRRNLFITYYVREGYKVVPMQKHFLFLVNLHLSQDDHIEKQQ
QSDDELGKGLFFHRIASTFLNESGIFNKMQFYTYQSNRLKEKRGELKHEKDTFTWIEPFQGNSYFTLNHGKGVISED
QLKELCYTILIEKQNVDSLEGKIIQFLKKFQNVSSKQVDEDELKREYFPANYFGRAGTGTLEKILNRLDKRMD
PTSKVTDKAYDKMIEVMEFINMCLPSDEKLQKDYRRYLKMVRFWNKEKHNIREFDSKKWTRFLPTELWNKRNL
EAYQLARKENKKKLEDMRNQVRSLKENDLEKYQQINYNLNLRLLSQELGVKWQEKDWVEYSGQIKKQISDNQK
LTIMKQRITAEKMMHGIENLNLRLISIDTNKSRQTMNRIALPKGFVNHIQQNSSEKISKRIREDYCKIELSGKY
EELSRQFFDKKNFDKMTLINGLCEKNKLI AFMVIYLLERLGFELKEKTKLGELKQTRMTYKISDKVKEDIPLSYYP
KLVYAMNRKYVDNIDSYAFAAYESKKAIDKVDIEKQRMEFIKQVLCFEEYIFENRIIEKSKFNDEETHISFTQI
HDELIKGRDTEKLSKLHARNKALHGEIPDGTSEKAKLLINEIKK* (SEQ ID NO:6)

[0395] MNAIELKKEEAAFYFNQARLNISGLDEIEEKQLPHIGSNRENAKKTVDMILDNPEVLKKMENYVFNSR
DIAKNARGELEALLKLVELRNFYSHYVHKDDVKTLSEGEKPLLDKYYEIAIEATGSKDVRLEIIDDKNKLTDAGV
LFLLCMFLKKESEANKLISSIRGFKRNDKEGQPRRNLFITYYVREGYKVVPMQKHFLFLVNLHLSNQDEYISNLR
PNQEIQQGGFFHRIASKFLSDSGILHSMKFYTYRSKRLTEQRGELKPKKDHTWIEPFQGNSYFSVQGGKGVIGEE
QLKELCYVLLVAREDFRAVEGKVTQFLKKFQANNVQVQVEKDEVLEKEYFPANYFENRDVGRVKDKILNRLKKITE
SYKAKGREVKAYDKMKEVMEFINNCLPTDENLKLKDYRRYLKMVRFWGREKENIKREFDSKKWERFLPRELWQKR
LEDAYQLAKEKNTLFLNKLKTTVERMNELEFEKYQQINDAKDLANLRQLARDFGVKWEKDWQEYSGQIKKQITDR
QKLTIMKQRITAALKKKQGIENLNLRLITTDTNKSRKVVLNRIALPKGFVRKHILKTDIKISKQIRQSCPIILSNN
YMKLAKEFFEERNFDKMTQINGLFEKNVLI AFMIVYLMEQLNRLGKNTLSNLKKTEVNFTITDKVTEKVQISQY
PSLVFAINREYVDGISGYKLPPKKPKEPPYTFEIKIDAIEKERMEFIKQVLGFEEHLFEKNVIDKTRFTDTATHIS
FNEICDELIKKGWDENKIIKLDARNAALHGKIPEDTSFDEAKVLINELKK* (SEQ ID NO:7)

[0396] 编码对应的在各pre-crRNA序列中的同向重复 (DR) 序列的DNA分别为SEQ ID NO:
8-14。

[0397] GCTGGAGCAGCCCCGATTTGTGGGGTGATTACAGC (SEQ ID NO:8)

[0398] GCTGAAGAAGCCTCCGATTTGAGAGGTGATTACAGC (SEQ ID NO:9)

[0399] GCTGTGATAGACCTCGATTTGTGGGGTAGTAACAGC (SEQ ID NO:10)

[0400] GCTGTGATAGACCTCGATTTGTGGGGTAGTAACAGC (SEQ ID NO:11)

[0401] GCTGTGATAGACCTCGATTTGTGGGGTAGTAACAGC (SEQ ID NO:12)

[0402] GCTGTGATGGGCCTCAATTTGTGGGGAAGTAACAGC (SEQ ID NO:13)

[0403] GCTGTGATAGGCCTCGATTTGTGGGGTAGTAACAGC (SEQ ID NO:14)

[0404] Cas13e.1、Cas13e.2、Cas13f.1、Cas13f.2、Cas13f.3、Cas13f.4和Cas13f.5蛋白的天然(野生型)DNA编码序列分别为SEQ ID NO:15-21。

[0405] ATGGCGCAAGTGTCAAAGCAGACTTCGAAAAAGAGAGAGTTGTCTATCGATGAATATCAAGGTGCTCG
GAAATGGTGTTTTACGATTGCCTTCAACAAGGCTCTTGTGAATCGAGATAAGAACGACGGGCTTTTGTGCGAGTCG
CTGTTACGCCATGAAAAGTATTCAAAGCAGACTGGTACGATGAGGATACACGCGCTTTGATCAAGTGTAGCACAC
AAGCGGCAATGCGAAGGCCGAGGCGTTAAGAACTATTTCTCCCACTATCGACATTCGCCCCGGTGTCTGACATT
TACAGCAGAAGATGAGTTGCGGACAATCATGGAAGGGCGTATGAGCGGGCGATCTTTGAATGCAGGAGACGCGAA
ACTGAAGTGATCATCGAGTTTCCCAGCCTGTTCTGAAGGCGACCGGATCACTACGGCGGGGTTGTGTTTTTCGTTT
CGTTCTTTGTTGAACGGCGGGTGTGGATCGTTTGTACGGTGCGGTAAGTGGGCTTAAGAAAAACGAAGGACAGTA
CAAGCTGACTCGGAAGGCGCTTTCGATGTATTGCCTGAAAGACAGTCGTTTCACGAAGGCGTGGGACAAACGCGTG
CTGCTTTTCAGGGATATACTCGCGCAGCTTGGACGCATCCCTGCGGAGGCGTATGAATACTACCACGGAGAGCAGG
GCGACAAGAAAAGAGCAAAACGACAATGAGGGGACGAATCCGAAACGCCATAAAGACAAGTTCATCGAGTTTGCCT
GCATTATCTGGAGGCGCAACACAGTGAGATATGCTTCGGGCGGCGACACATTGTCAGGGAGGAGGCCGGGGCAGGC
GACGAACACAAAAAGCACAGGACCAAAGGCAAGGTAGTTGTGACTTTTCAAAAAAGACGAAGATCAGTCATACT
ATATCAGTAAGAACAATGTTATCGTCAGGATTGATAAGAATGCCGGCCTCGGAGTTATCGCATGGGGCTTAACGA
ATTGAAATACCTTGTATTGCTTAGCCTTCAGGGAAAGGGCGACGATGCGATTGCAAACTGTACAGGTATCGGCAG
CATGTGGAGAACATTCTGGATGTAGTGAAGTTCACAGATAAGGATAATCACGTCTTCCTGCCGCGATTTGTGCTGG
AGCAACATGGGATTGGCAGGAAAGCTTTAAGCAAAGAATAGACGGCAGAGTAAAGCATGTTTCGAGGGGTGTGGGA
AAAGAAGAAGGCGGCGACCAACGAGATGACACTTCACGAGAAGGCGCGGGACATTCTTCAATACGTAAATGAAAT
TGCACGAGGTCTTTCAATCCCGGCGAGTACAACCGGCTGCTGGTGTGTCTGGTTGGCAAGGATGTTGAGAATTTTC
AGGCGGGACTGAAACGCCTGCAACTGGCCGAGCGAATCGACGGGCGGGTATATTCAATTTTTGCGCAGACCTCCAC
AATAAACGAGATGCATCAGGTGGTGTGTGATCAGATTCTCAACAGACTTTGCCGAATCGGCGATCAGAAGCTCTAC
GATTATGTGGGGCTTGGGAAGAAGGATGAAATAGATTACAAGCAGAAGGTTGCATGGTTCAAGGAGCATATTTCTA
TCCGCAGGGGTTTCTTGCGCAAGAAGTCTGGTATGACAGCAAGAAGGGATTCGCGAAGCTTGTGGAAGAGCATTT
GGAAAGCGGCGGCGACAGAGGGACGTTGGGCTGGATAAAAAGTATTATCATATTGATGCGATTGGGCGATTTCGAG
GGTGCTAATCCAGCCTTGTATGAAACGCTGGCGCGAGACCGTTTGTGTCTGATGATGGCGCAATACTTCCTGGGGA
GTGTACGCAAGGAATTGGGTAATAAAATTGTGTGGTCTGAATGATAGCATCGAGTTGCCCGTGGAGGGCTCAGTGGG
TAACGAAAAAAGCATCGTCTTCTCAGTGAGTGATTACGGCAAGTTATATGTGTTGGATGACGCTGAGTTTCTTGGG
CGGATATGTGAGTACTTTATGCCGCACGAAAAAGGGAAGATACGGTATCATACAGTTTACGAAAAAGGGTTTAGGG
CATATAATGATCTGCAGAAGAAATGTGTCGAGGCGGTGCTGGCGTTTGAAGAGAAGGTTGTCAAAGCCAAAAAGAT
GAGCGAGAAGGAAGGGGCGCATTATATTGATTTTCGTGAGATACTGGCACAACAATGTGTAAAGAGGCGGAGAAG
ACCGCCGTGAATAAGGTGCGTAGAGCGTTTTTCCATCATCATTTAAAGTTTGTGATAGATGAATTTGGGTTGTTTA
GTGATGTTATGAAGAAATATGGAATTGAAAAGGAGTGGAAGTTTCCTGTAAATGA (SEQ ID NO:15)

[0406] ATGAAGGTTGAAAATATTAAGAAAAAAGCAAAAAAGCAATGTATTTAATCAACCATTATGAGGGACC
CAAAAAATGGTGTTTTGCAATAGTTCTGAATAGGGCATGTGATAATTACGAGGACAATCCACACTTGTTTTCCAAA
TCACTTTTGAATTTGAAAAACAAGTCGAAAAGATTGGTTTACGAAGAAAACACGAGAGCTTGTTGAGCAAGCAG
ATACAGAAATACAGCCAAATCCTAACCTGAAACCTAATACAACAGCTAACCGAAAACTCAAAGATATAAGAACTA

TTTTTCGCATCATTATCACAAGAACGAATGCCTGTATTTTAAGAACGATGATCCCATACGCTGCATTATGGAAGCG
GCGTATGAAAAATCTAAATTTATATCAAAGGAAAGCAGATTGAGCAAAGCGATATACCATTGCCGAATTGTTTG
AAAGCAGCGGTTGGATTACACCGGCGGGGATTTTGTTACTGGCATCCTTTTTTGTGTAACGAGGGATTCTACATCG
CTTGATGGGAAATATCGGAGGATTTAAAGATAATCGAGGCGAATACGGTCTTACACACGATATTTTACCACCTAT
TGTCTTAAGGGTAGTTATTCAATTCGGGCGCAGGATCATGATGCGGTAATGTTTCAGAGATATTCTCGGCTATCTGT
CACGAGTTCCCACTGAGTCATTTACAGCGTATCAAGCAACCTCAAATACGAAAAGAAGGCCAATTAAGTGAAAGAAA
GACGGACAAATTTATAACATTTGCACTAAATTATCTTGAGGATTATGGGCTGAAAGATTTGGAAGGCTGCAAAGCC
TGTTTTGCCAGAAGTAAAAATTGTAAGGGAACAAGAAAATGTTGAAAGCATAAATGATAAGGAATACAAACCTCACG
AGAACAAAAAGAAAGTTGAAATTCACCTTCGATCAGAGCAAAGAAGACCGATTTTATATTAATCGCAATAACGTTAT
TTTGAAGATTCAGAAGAAAGATGGACATTCCAACATAGTTAGGATGGGAGTATATGAACTTAAATATCTCGTTCTT
ATGAGTTTAGTGGGAAAAGCAAAGAAGCAGTTGAAAAAATTGACAACTATATCCAGGATTTGCGAGACCAGTTGC
CTTACATAGAGGGGAAAAATAAGGAAGAGATTAAAGAATACGTCAGGTTCTTTCCACGATTTATACGTTCTCACCT
CGGTTTACTACAGATTAACGATGAAGAAAAGATAAAAGCTCGATTAGATTATGTTAAGACCAAGTGGTTAGATAAA
AAGGAAAAATCGAAAGAGCTTGAACCTCATAAAAAAGGACGGGACATCCTCAGGTATATCAACGAGCGATGTGATA
GAGAGCTTAACAGGAATGTATATAACCGTATTTTAGAGCTCCTGGTCAGCAAAGACCTCACTGGTTTTTATCGTGA
GCTTGAAGAACTAAAAAGAACAAGGCGGATAGATAAAAAATATTGTCCAGAATCTTCTGGGCAAAAAACCATTAAT
GCACTGCATGAAAAGGTCTGTGATCTGGTGTGAAGGAAATCGAAAGTCTCGATACAGAAAATCTCAGGAAATATC
TTGGATTGATACCCAAAGAAGAAAAAGAGGTCACTTTCAAAGAAAAGGTGCGATAGGATTTTGAAACAGCCAGTTAT
TTACAAAGGGTTTCTGAGATACCAATTCTTCAAAGATGACAAAAAGAGTTTTGTCTTACTTGTGTAAGACGCATTG
AAGGAAAAAGGAGGAGGTTGTGATGTTCTCTTGGGAAAGAGTATTATAAAATCGTGTCACTTGATAAGTATGATA
AAGAAAATAAAACCTGTGTGAAACTCTGGCGATGGATAGGCTTTGCCATTATGATGGCAAGACAATATTATCTCAG
TCTGAATGCAAACTTGCACAGGAAGCTCAGCAAATCGAATGGAAGAAAGAAGATAGTATAGAATTGATTATTTTC
ACCTTAAAAAATCCCGATCAATCAAAGCAGAGTTTTTCTATACGGTTTTTCGGTCAGAGATTTTACGAAGTTGTATG
TAACGGATGATCCTGAATTTCTGGCCCGGCTTTGTTCTACTTTTTCCAGTTGAAAAAGAGATTGAATATCACAA
GCTCTATTGAGAAGGGATAAATAAATACACAAACCTGCAAAAAGAGGGAATCGAAGCAATACTCGAGCTTGAAAAA
AAGCTTATTGAACGAAATCGGATTCAATCTGCAAAAAATTATCTCTCATTTAATGAGATAATGAATAAAAGCGGTT
ATAATAAAGATGAGCAGGATGATCTAAAGAAGGTGCGAAATTTCTCTTTGCATTATAAGCTTATCTTTGAGAAAGA
ACATCTCAAGAAGTTCTATGAGGTTATGAGAGGAGAAGGGATAGAGAAAAAGTGGTCTTTAATAGTATGA (SEQ ID
NO: 16)

[0407] ATGAATGGCATTGAATTAaaaaaagaagaAGCAGCATTTTATTTTAATCAGGCAGAGCTTAATTTAAA
AGCCATAGAAGACAATATTTTTGATAAAGAAAGACGAAAGACTCTGCTTAATAATCCACAGATACTTGCCAAAATG
GAAAAATTCATTTTCAATTCAGAGATGTAACAAAAATGCAAAAGGGGAAATTGACTGCTTGCTGTTGAAACTAA
GAGAGCTGAGAACTTTTACTCGCATTATGTCCACAAACGAGATGTAAGAGAATTAAGCAAGGGCGAGAAACCTAT
ACTTGAAAAGTATTACCAATTTGCGATTGAATCAACCGGAAGTGAAAATGTTAACTTGAGATAATAGAAAACGAC
GCGTGGCTTGACAGATGCCGGTGTGTTGTTTTCTTATGTATTTTTTTGAAGAAATCTCAGGCAAATAAGCTTATAA
GCGGTATCAGCGGTTTTAAAAGAAACGATGATACCGGTGAGCCGAGAAGGAATTTATTTACCTATTTTCAGTATAAG
GGAGGGATACAAGGTTGTTCCGGAATGCAGAAACATTTCTTTTGTCTTCTGTTAATCATCTCTCTAATCAA
GATGATTATATTGAAAAAGCGCATCAGCCATACGATATAGGCGAGGGTTTATTTTTTCATCGAATAGCTTCTACAT
TTCTTAATATAAGTGGGATTTTAAGAAATATGAAATTCTATACCTATCAGAGTAAAAGGTTAGTAGAGCAGCGGG

AGAACTCAAACGAGAAAAGGATATTTTTGCGTGGAAGAACCGTTTCAAGGAAATAGTTATTTTGAAATAAATGGT
CATAAAGGAGTAATCGGTGAAGATGAATTGAAGGAATATGTTATGCATTTCTGATTGGCAATCAAGATGCTAATA
AAGTGAAGGCAGGATTACACAATTTCTAGAAAAGTTTAGAAATGCGAACAGTGTGCAACAAGTTAAAGATGATGA
AATGCTAAAACCAGAGTATTTTCCTGCAAATTATTTTGCTGAATCAGGCGTCGGAAGAATAAAGGATAGAGTGCTT
AATCGTTTGAATAAAGCGATTAAAAGCAATAAGGCCAAGAAAGGAGAGATTATAGCATACGATAAGATGAGAGAGG
TTATGGCGTTCATAAATAATTCTCTGCCGGTAGATGAAAAATTGAAACCAAAAGATTACAAACGATATCTGGGAAT
GGTTCGTTTCTGGGACAGGGAAGATAACATAAAGCGGGAGTTTCGAGACAAAAGAATGGTCTAAATATCTTCCA
TCTAATTTCTGGACGGCAAAAAACCTTGAAAGGTCTATGGTCTGGCAAGAGAGAAAAACGCAGAATTATTCAATA
AACTAAAAGCGGATGTAGAAAAAATGGACGAACGGGAACCTTGAGAAGTATCAGAAGATAAATGATGCAAAGGATTT
GGCAAATTTACGCCGGCTTGCAAGCGACTTTGGTGTGAAGTGGGAAGAAAAAGACTGGGATGAGTATTCAGGACAG
ATAAAAAACAAATTACAGACAGCCAGAACTAACAATAATGAAGCAGCGGATAACCGCAGGACTAAAGAAAAAGC
ACGGCATAGAAAATCTTAACCTGAGAATAACTATCGACATCAATAAAAGCAGAAAGGCAGTTTTGAACAGAATTGC
GATTCCGAGGGGTTTTGTAAAAAGGCATATTTTAGGATGGCAAGAGTCTGAGAAGGTATCGAAAAAGATAAGAGAG
GCAGAATGCGAAATTCTGCTGTCGAAAGAATACGAAGAATATCGAAACAATTTTCCAAAGCAAAGATTATGACA
AAATGACACGGATAAATGGCCTTTATGAAAAAACAACTTATAGCCCTGATGGCAGTTTATCTAATGGGGCAATT
GAGAATCCTGTTTAAAGAACACACAAAACCTTGACGATATTACGAAAACAACTGTGGATTTCAAAATATCTGATAAG
GTGACGGTAAAAATCCCCTTTCAAATTATCCTTCGCTCGTTTATACAATGTCCAGTAAGTATGTTGATAATATAG
GGAATTATGGATTTTCCAACAAAGATAAAGACAAGCCGATTTTAGGTAAGATTGATGTAATAGAAAAACAGCGAAT
GGAATTTATAAAGAGGTTCTTGGTTTTGAAAAATATCTTTTTGATGATAAAATAATAGATAAAGCAAATTTGCT
GATACAGCGACTCATATAAGTTTTGCAGAAATAGTTGAGGAGCTTGTTGAAAAAGGATGGGACAAAGACAGACTGA
CAAACTTAAAGATGCAAGAAATAAAGCCCTGCATGGTGAAATACTGACGGGAACCAGCTTTGATGAAACAAAATC
ATTGATAACGAATTAAAAAAATGA (SEQ ID NO:17)

[0408] ATGTCCCCAGATTTTCATCAAATTAGAAAAACAGGAAGCAGCTTTTTACTTTAATCAGACAGAGCTTAA
TTTAAAGCCATAGAAAGCAATATTTTAGACAAACAACAGCGAATGATTCTGCTTAATAATCCACGGATACTTGCC
AAAGTAGGAAATTTCAATTTTCAATTCAGAGATGTAACAAAAATGCAAAAGGAGAAATAGACTGTCTGCTATTTA
AACTGGAAGAGCTAAGAACTTTTACTCGCATTATGTTTCATACCGACAATGTAAAGGAATTGAGTAACGGAGAAAA
ACCCCTACTGGAAGATATTATCAAATCGCTATTCAGGCAACCAGGAGTGAGGATGTAAAGTTCGAATTGTTTGAA
ACAAGAAACGAGAATAAGATTACGGATGCCGGTGATTGTTTTTCTTATGTATGTTTTTAAAAAATCACAGGCAA
ACAAGCTTATAAGCGGTATCAGCGGCTTCAAAGAAATGATCCAACAGGCCAGCCGAGAAGAACTTATTTACCTA
TTTCAGTGCAAGAGAAGGATATAAGGCTTTCCTGATATGCAGAAACATTTTCTCTTTTTACTCTGGTTAATTAT
TTGTGAATCAGGATGAGTATATCAGCGAGCTTAAACAATATGGAGAGATTGGTCAAGGAGCCTTTTTTAATCGAA
TAGCTTCAACATTTTTGAATATCAGCGGGATTTAGGAAATACGAAATTCTATTCGTATCAAAGTAAAGGATAAA
AGAGCAGCGAGGCGAACTCAATAGCGAAAAGGACAGCTTTGAATGGATAGAGCCTTTCCAAGGAAACAGCTATTTT
GAAATAAATGGGCATAAAGGAGTAATCGGCGAAGACGAATTAAGAAGCTTTGTTATGCATTGTTGGTTGCCAAGC
AAGATATTAATGCCGTTGAAGGCAAAATTATGCAATTCCTGAAAAAGTTTAGAAATACTGGCAATTTGCAGCAAGT
TAAAGATGATGAAATGCTGGAAATAGAATATTTTCCCGCAAGTTATTTAATGAATCAAAAAAGAGGACATAAAG
AAAGAGATTCTTGGCCGGCTGGATAAAAAGATTGCTCCTGCTCTGCAAAGGCAGAAAAAGCCTATGATAAGATGA
AAGAGGTGATGGAGTTTATAAATAATTCTCTGCCGGCAGAGGAAAAATTGAAACGCAAAGATTATAGAAGATATCT
AAAGATGGTTCGTTTCTGGAGCAGAGAAAAAGGCAATATAGAGCGGGAATTTAGAACAAAGGAATGGTCAAAATAT

TTTTCATCTGATTTTTGGCGGAAGAACAATCTTGAAGATGTGTACAACTGGCAACACAAAAAACGCTGAAGTGT
TCAAAAATCTAAAAGCGGCAGCAGAGAAAATGGGTGAAACGGAATTTGAAAAGTATCAGCAGATAAACGATGTAAA
GGATTTGGCAAGTTTAAGGCGGCTTACGCAAGATTTTGGTTTGAAGTGGGAAGAAAAGGACTGGGAGGAGTATTCC
GAGCAGATAAAAAACAAATTACGGACAGGCAGAACTGACAATAATGAAACAAAGGGTTACGGCTGAACTAAAGA
AAAAGCACGGCATAGAAAATCTTAATCTGAGAATAACCATCGACAGCAATAAAAGCAGAAAGGCGGTTTTGAACAG
AATAGCAATTCCAAGAGGATTTGTAAAAAACATATTTTAGGCTGGCAGGGATCTGAGAAGATATCGAAAAATATA
AGGGAAGCAGAATGCAAAATCTGCTATCGAAAAATATGAAGAGTTATCAAGGCAGTTTTTTGAAGCCGGTAATT
TCGATAAGCTGACGCAGATAAATGGTCTTTATGAAAAGAATAAACTTACAGCTTTTATGTCAGTATATTTGATGGG
TCGGTTGAATATTCAGCTTAATAAGCACACAGAACTTGAAATCTTAAAAAACAGAGGTGGATTTTAAGATATCT
GATAAGGTGACTGAAAAAATACCGTTTTCTCAGTATCCTTCGCTTGTCTATGCGATGTCTCGCAAATATGTTGACA
ATGTGGATAAATATAAATTTTCTCATCAAGATAAAAAGAAGCCATTTTTAGGTAAAATTGATTCAATTGAAAAAGA
ACGTATTGAATTCATAAAAGAGGTTCTCGATTTTGAAGAGTATCTTTTTAAAAATAAGGTAATAGATAAAAGCAAA
TTTTCCGATACAGCGACTCATATTAGCTTTAAGGAAATATGTGATGAAATGGGTAAAAAAGGATGTAACCGAAACA
AACTAACCGAACTTAACAACGCAAGGAACGCAGCCCTGCATGGTGAAATACCGTCGGAGACCTCTTTTCGTGAAGC
AAAACCGTTGATAATGAATTGAAAAAATGA (SEQ ID NO:18)

[0409] ATGTCCCAGATTTTCATCAAATTAGAAAAACAAGAAGCAGCTTTTTACTTTAATCAGACAGAGCTTAA
TTTAAAGCCATAGAAAGCAATATTTTCGACAAACAACAGCGAGTGATTCTGCTTAATAATCCACAGATACTTGCC
AAAGTAGGAGATTTTATTTTCAATTCAGAGATGTAACAAAAACGCAAAAGGAGAAATAGACTGTTTGCTATTGA
AACTAAGAGAGCTGAGAACTTTTACTCACACTATGTCTATACCGATGACGTGAAGATATTGAGTAACGGCGAAAG
ACCTCTGCTGGAAAAATATTATCAATTTGCGATTGAAGCAACCGGAAGTGAAATGTTAACTTGAAATAATAGAA
AGCAACAACCGACTTACGGAAGCGGGCGTGCTGTTTTTCTTGTGTATGTTTTTGAAAAAGTCTCAGGCAAATAAGC
TTATAAGCGGTATCAGCGTTTTTAAAAGAAATGACCCGACAGGTCAGCCGAGAAGGAATTTATTACCTACTTCAG
TGTAAGGGAGGGATACAAGGTTGTGCCGATATGCAGAAACATTTTCTTTTGTGTTGTTCTTGTCAATCATCTCTCT
GGTCAGGATGATTATATTGAAAAGGCGCAAAAGCCATACGATATAGGCGAGGGTTTATTTTTTCATCGAATAGCTT
CTACATTTCTTAATATCAGTGGGATTTTAAAGAAATATGGAATTCTATATTTACCAGAGCAAAAGACTAAAGGAGCA
GCAAGGAGAGCTCAAACGTGAAAAGGATATTTTCCATGGATAGAGCCTTTCCAGGGAAATAGTTATTTTGAAATA
AATGGTAATAAAGGAATAATCGGCGAAGATGAATTGAAAGAGCTTTGTTATGCGTTGCTGGTTGCAGGAAAAGATG
TCAGAGCCGTCGAAGGTAAAATAACACAATTTTTGGAAGGTTTAAAAATGCGGACAATGCTCAGCAAGTTGAAAA
AGATGAAATGCTGGACAGAAACAATTTTCCCGCCAATTATTTGCGCGAATCGAACATCGGCAGCATAAAGGAAAAA
ATACTTAATCGTTTGGGAAAAACTGATGATAGTTATAATAAGACGGGGACAAAGATTAAACCATACGACATGATGA
AAGAGGTAATGGAGTTTATAAATAATTCTCTTCCGGCAGATGAAAAATTGAAACGCAAAGATTACAGAAGATATCT
AAAGATGGTTTCGTATCTGGGACAGTGAGAAAGATAATATAAAGCGGGAGTTTGAAAGCAAAGAATGGTCAAAATAT
TTTTCATCTGATTTCTGGATGGCAAAAAATCTTGAAGGGTCTATGGGTTGGCAAGAGAGAAAAACGCCGAATTAT
TCAATAAGCTAAAAGCGGTTGTGGAGAAAATGGACGAGCGGAATTTGAGAAGTATCGGCTGATAAATAGCGCAGA
GGATTTGGCAAGTTTAAGACGGCTTGCAGAAAGATTTTGGCCTGAAGTGGGAAGAAAAGGACTGGCAAGAGTATTCT
GGGCAGATAAAAAACAAATTTCTGACAGGCAGAACTGACAATAATGAAACAAAGGATTACGGCTGAACTAAAGA
AAAAGCACGGCATAGAAAATCTCAATCTTAGAATAACCATCGACAGCAATAAAAGCAGAAAGGCAGTTTTGAACAG
AATCGCAGTTCCAAGAGGTTTTGTGAAAGAGCATATTTTAGGATGGCAGGGGTCTGAGAAGGTATCGAAAAAGACA
AGAGAAGCAAAGTGCAAAATCTGCTCTCGAAAGAATATGAAGAATTATCAAAGCAATTTTTCCAAACCAGAAATT

ACGACAAGATGACGCAGGTAAACGGTCTTTACGAAAAGAATAAACTCTTAGCATTTATGGTCGTTTATCTTATGGA
GCGGTTGAATATCCTGCTTAATAAGCCCACAGAACTTAATGAACTTGAAAAAGCAGAGGTGGATTTCAGATATCT
GATAAGGTGATGGCCAAAATCCCGTTTTTCACAGTATCCTTCGCTTGTTGACGCGATGTCCAGCAAATATGCTGATA
GTGTAGGCAGTTATAAAATTTGAGAATGATGAAAAAACAAGCCGTTTTTAGGCAAGATCGATACAATAGAAAAACA
ACGAATGGAGTTTATAAAAAGAAGTCCTTGGTTTTGAAGAGTATCTTTTTGAAAAAGAAGATAATAGATAAAAGCGAA
TTTGCCGACACAGCGACTCATATAAGTTTTGATGAAATATGTAATGAGCTTATTAAGGATGGGATAAAGACA
AACTAACCAAACTTAAAGATGCCAGGAACGCGGCCCTGCATGGCGAAATACCGGCGGAGACCTCTTTTCGTGAAGC
AAAACCGTTGATAAATGGATTGAAAAATGA (SEQ ID NO:19)

[0410] ATGAACATCATTAATTAAGGCTGCGTTTTATTTAATCAGACGATCCTCAATCTTTC
AGGGCTTGATGAAATTATTGAAAAACAAATCCGCACATAATCAGCAACAAGGAAAATGCAAAGAAAGTGATTGAT
AAGATTTTCAATAACCGCTTATTATTAAGGTGTGGAGAATTATATCTACAACTTAAAGATGTGGCTAAAAACG
CAAGAACTGAAATTGAGGCTATATTGTTGAAATTAGTAGAGCTACGTAATTTTACTCACATTACGTTTATAATGA
TACCGTCAAGATACTAAGTAACGGTGAAAAACCTATACTGGAAAAATATTATCAAAATGCTATAGAAGCAACCGGA
AGTAAAAATGTTAACTTGAATCATAGAAAACAACAAGTGTCTCACGGATTCTGGCGTGCTGTTTTGCTGTGTA
TGTTCTTAAGGCTTACAGGCAACAAAGCTTATAAGTTCCGTTAGTGGTTTTAAAGGAATGATAAAGAAGGACA
ACCGAGAAGAAATCTATTCATTATTATAGTGTGAGGGAGGGATATAAGGTTGTGCCTGATATGCAGAAGCATTTTC
CTTCTATTGCTCTGGTCAATCATCTATCTGAGCAGGATGATCATATTGAGAAGCAGCAGCAGTCAGACGAGCTCG
GTAAGGGTTTGTTCATCGTATAGCTTCGACTTTTTTAAACGAGAGCGGCATCTTCAATAAAATGCAATTTTA
TACATATCAGAGCAACAGGCTAAAAGAGAAAAGAGGAGAAGTCAAACACGAAAAGGATACCTTTACATGGATAGAG
CCTTTTCAAGGCAATAGTTATTTACGTTAAATGGACATAAGGGAGTGATTAGTGAAGATCAATTGAAGGAGCTTT
GTTACACAATTTTAATTGAGAAGCAAAACGTTGATTCCTTGGAAGGTAAAATTATACAATTTCTCAAAAAATTTCA
GAATGTCAGCAGCAAGCAGCAAGTTGACGAAGATGAATTGCTTAAAGAGAATATTTCCCTGCAAATTACTTTGGC
CGGGCAGGAACAGGGACCTAAAAGAAAAGATTCTAAACCGGCTTGATAAGAGGATGGATCCTACATCTAAAGTGA
CGGATAAAGCTTATGACAAAATGATTGAAGTGATGGAATTTATCAATATGTGCCTTCCGCTGATGAGAAGTTGAG
GCAAAAGGATTATAGACGATACTTAAAGATGGTTCGTTTCTGGAATAAGGAAAAGCATAACATTAAGCGCGAGTTT
GACAGTAAAAAATGGACGAGTTTTTGCCGACGGAATTGTGGAATAAAAGAAATCTAGAAGAAGCCTATCAATTAG
CACGGAAGAGAAACAAAAGAACTTGAAGATATGAGAAATCAAGTACGAAGCCTTAAAGAAAATGACCTTGAAAA
ATATCAGCAGATTAATTACGTTAATGACCTGGAGAATTTAAGGCTTCTGTACAGGAGTTAGGTGTGAAATGGCAG
GAAAAGGACTGGGTGAAATATCCGGGCAGATAAAGAAGCAGATATCAGACAATCAGAACTTACAATCATGAAAC
AAAGGATTACCGCTGAACTAAAGAAAATGCACGGCATCGAGAATCTTAATCTTAGAATAAGCATTGACACGAATAA
AAGCAGGCAGACGGTTATGAACAGGATAGCTTTGCCCAAAGGTTTTGTGAAGAATCATATCCAGCAAAATTCGTCT
GAGAAAATATCGAAAAGAATAAGAGAGGATTATTGTAAGTTGAGCTATCGGGAATATGAAGAATTTCAAGGC
AATTTTTTGATAAAAAGAATTCGATAAGATGACACTGATAAACGGCCTTTGTGAAAAGAACAACTTATCGCATT
TATGGTTATCTATCTTTTGGAGCGCTTGGATTTGAATTAAGGAGAAAAACAAATTAGGCGAGCTTAAACAAACA
AGGATGACATATAAAATATCCGATAAGGTAAAAGAAGATATCCCGCTTTCCTATTACCCCAAGCTTGTGTATGCAA
TGAACCGAAAATATGTTGACAATATCGATAGTTATGCATTTGCGGCTTACGAATCCAAAAAGCTATTTTGATAA
AGTGGATATCATAGAAAAGCAACGTATGGAATTTATCAACAAGTTCTCTGTTTTGAGGAATATATTTTCGAAAAT
AGGATTATCGAAAAAGCAAATTTAATGACGAGGAGACTCATATAAGTTTTACACAAATACATGATGAGCTTATTA
AAAAAGGACGGGACACAGAAAACTCTCTAACTCAAACATGCAAGGAATAAAGCCTTGACGGCGAGATTCTGA

TGGGACTTCTTTTGAAAAAGCAAAGCTATTGATAAATGAAATCAAAAAATGA (SEQ ID NO:20)

[0411] ATGAATGCTATCGAACTAAAAAAGAGGAAGCAGCATTTTATTTAATCAGGCAAGACTCAACATTTCCAGGACTTGATGAAATTATTGAAAAGCAGTTACCACATATAGGTAGTAACAGGGAGAATGCGAAAAAACTGTTGATATGATTTTGGATAATCCCGAAGTCTTGAAGAAGATGGAATTTATGTCTTTAACTCACGAGATATAGCAAAGAACGCAAGAGGTGAACTTGAAGCATTGTTGTTGAAATTAGTAGAACTGCGTAATTTTATTACATTATGTTTCATAAAGATGATGTTAAGACATTGAGTTACGGAGAAAAACCTTTACTGGATAAATATTATGAAATTGCGATTGAAGCGACCGGAAGTAAAGATGTCAGACTTGAGATAATAGATGATAAAAAATAAGCTTACAGATGCCGGTGTGCTTTTTTTATTGTGTATGTTTTTGAAAAAATCAGAGGCAAACAACTTATCAGTTCAATCAGGGGCTTTAAAAGAAACGATAAAGAAGGCCAGCCGAGAAGAAATCTATTCACCTACTACAGTGTGAGAGGGATATAAGGTTGTGCCTGATATGCAGAAACATTTCTTTTATTACACTGGTTAACCATTGTCAAATCAGGATGAATACATCAGTAATCTTAGGCCGAATCAAGAAATCGGCCAAGGGGGATTTTTCCATAGAAATAGCATCAAAATTTTTGAGCGATAGCGGGATTTACATAGTATGAAATTCTACCTACCGGAGTAAAAGACTAACAGAACAACGGGGGAGCTTAAGCCGAAAAAAGATCATTTTACATGGATAGAGCCTTTTCAGGGAAACAGTTATTTTTCAGTGCAGGGCCAAAAAGGAGTAATTGGTGAAGAGCAATTAAAGGAGCTTTGTTATGTATTGCTGGTTGCCAGAGAAGATTTTAGGGCCGTTGAGGGCAAAGTTACACAATTTCTGAAAAAGTTTCAGAATGCTAATAACGTACAGCAAGTTGAAAAAGATGAAGTGCTGGAAAAAGAATATTTTCCTGCAAATTATTTTGAAATCGAGACGTAGGCAGAGTAAAGGATAAGATACTTAATCGTTTGAAAAAATCACTGAAAGCTATAAAGCTAAAGGGAGGGAGGTTAAAGCCTATGACAAGATGAAAGAGGTAATGGAGTTTATAATAATTGCCTGCCAACAGATGAAAAATTTGAAACTCAAAGATTACAGAAGATATCTGAAAATGGTTCGTTTCTGGGGCAGGGAAAAGGAAAATATAAAGCGGGAATTTGACAGTAAAAAATGGGAGAGGTTTTTGCCAAGAGAAGTCTGGCAGAAAAAGAAACCTCGAAGATGCGTATCACTGGCAAAAAGAGAAAAACACCGAGTTATTCAATAAATTGAAAACAACTGTTGAGAGAATGAACGAAGTGAATTCGAAAAGTATCAGCAGATAAACGACGCAAAAGATTTGGCAAATTTAAGGCAACTGGCGCGGGACTTCGGCGTGAAGTGGGAAGAAAAGGACTGGCAAGAGTATTCGGGGCAGATAAAAAAACAATTACAGACAGGCAAAAACCTTACAATAATGAAACAAAGGATTACTGCTGCATTGAAGAAAAAGCAAGGCATAGAAAATCTTAATCTTAGGATAACAACCGACACAATAAAAGCAGAAAGGTGGTATTGAACAGAATAGCGCTACCTAAAGGTTTTGTAAGGAAGCATATCTTAAAAACAGATATAAAGATATCAAAGCAAATAAGGCAATCACAATGTCCTATTATACTGTCAAACAATTATATGAAGCTGGCAAGGAATCTTTGAGGAGAGAAATTTTGATAAGATGACGCAGATAAACGGGCTATTTGAGAAAAATGTACTTATAGCGTTTATGATAGTTTATCTGATGGAACAACCTGAATCTTCGACTTGGTAAGAATACGGAACCTTAGCAATCTTAAAAAACGGAGGTTAATTTTACGATAACCGACAAGGTAACGGAAAAAGTCCAGATTTTCGCAGTATCCATCGCTTGTTTTCCGCCATAAACAGAGAATATGTTGATGGAATCAGCGTTATAAGTTACCGCCCCAAAAACCGAAAGAGCCTCCGTATACTTTCTTCGAGAAAAATAGACGCAATAGAAAAAGAACGAATGGAATTCATAAAACAGGTCCTCGGTTTCGAAGAACATCTTTTTGAGAAGAATGTAATAGACAAAACCTCGCTTTACTGATACTGCGACTCATATAAGTTTTAATGAAATATGTGATGAGCTTATAAAAAAAGGATGGGACGAAAACAAAATAATAAACTTAAAGATGCGAGGAATGCAGCATTGCATGATAAGATACCGGAGGATACGTCTTTTGATGAAGCGAAAGTACTGATAAATGAATTAAAAAATGA (SEQ ID NO:21)

[0412] 针对另外的功能性实验生成了七种Cas13e和Cas13f蛋白(即,Cas13e.1、Cas13e.2、Cas13f.1、Cas13f.2、Cas13f.3、Cas13f.4和Cas13f.5),它们的经人密码子优化的编码序列分别为SEQ ID NO:22-28。

[0413] ATGGCCCAGGTGAGCAAGCAGACCTCCAAGAAGAGGGAGCTGAGCATCGACGAGTACCAGGGCGCCCGGAAGTGGTGCTTCACCATTGCCTTCAACAAGGCCCTGGTGAACCGGGACAAGAACGACGGCCTGTTTCGTGGAAGC

CTGCTGAGACACGAGAAGTACAGCAAGCACGACTGGTACGACGAAGATACCCGGGCCCTGATCAAGTGCAGCACCC
AGGCCGCCAACGCCAAGGCTGAAGCCCTGCGGAACCTACTTCAGTCACTACCGGCATAGCCCTGGCTGCCTGACCTT
CACCGCCGAGGACGAACTGCGGACCATCATGGAGAGAGCCTATGAGCGGGCCATCTTCGAGTGCAGAAGAAGAGAG
ACAGAGGTGATCATCGAGTTTCCCAGCCTGTTTCGAGGGCGACCGGATCACCAACCGCCGGCGTGGTGTTCGTGA
GCTTTTTCGTGGAAAGAAGAGTGCTGGATCGGCTGTATGGAGCCGTGTCCGGCCTGAAGAAGAATGAGGGACAGTA
CAAGCTGACCCGGAAGGCCCTGAGCATGTACTGCCTGAAGGACAGCAGATTACCAAGGCCTGGGATAAGCGGGTG
CTGCTGTTTCAGAGACATCCTGGCCCAGCTGGGAAGAATCCCCGCCGAGGCCTACGAGTACTACCACGGCGAGCAGG
GTGATAAGAAGAGAGCTAACGACAATGAGGGCACAAATCCCAAGCGGCACAAGGACAAGTTCATCGAATTTGCACT
GCACTACCTGGAAGCCCAGCACAGCGAGATCTGCTTCGGCAGACGCCACATCGTGCGGGAAGAGGCCGGCGCCGGC
GATGAGCACAAGAAGCACCGGACCAAGGGAAAGGTGGTGGTGGACTTCAGCAAGAAGGACGAGGACCAGAGCTACT
ATATCTCCAAGAACAACGTGATCGTGC GGATCGACAAGAACGCCGGCCCTAGAAGCTACCGGATGGGCCTGAACGA
GCTGAAGTACCTCGTGCTGCTGAGCCTGCAGGGGAAGGGCGACGATGCCATCGCCAAGCTGTACAGATACAGACAG
CACGTGGAGAACATCCTGGATGTGGTGAAGGTGACCGATAAGGATAACCACGTGTTCTGCCCCGCTTCGTGCTGG
AGCAGCACGGCATCGGCAGAAAGGCCTTCAAGCAGCGGATCGATGGACGGGTGAAGCACGTGCGGGGCGTGTGGGA
GAAGAAGAAGGCCGCCACCAATGAAATGACCCTGCACGAGAAGGCCAGAGACATCCTGCAGTACGTGAACGAAAAC
TGCACCCGGTCCTTCAACCCTGGCGAATACAACAGACTGCTGGTGTGCCTGGTGGGCAAGGACGTGGAGAACTTTC
AGGCCGGCCTGAAGCGGCTGCAGCTGGCCGAAAGGATCGATGGCCGGGTGTA CTCCATCTTCGCCCAGACCAGCAC
CATCAATGAGATGCACCAGGTGGTGTGCGACCAGATCCTGAACCGGCTGTGCAGAATCGCGCACCAGAAGCTGTAC
GATTACGTGGGACTGGGCAAGAAGGACGAAATCGACTACAAGCAGAAGGTGGCCTGGTTCAAGGAGCACATCAGCA
TCCGGAGAGGATTCTTGAGAAAGAAGTTCTGGTACGATAGCAAGAAGGGATTTCGCAAAGCTGGTGGAGGAACACCT
GGAGTCCGGCGGGCGGCCAGCGCGACGTGGGCCTGGACAAGAAGTACTACCACATCGACGCCATCGGCAGATTTCGAG
GGCGCCAACCCCGCCCTGTACGAGACCCTGGCCAGAGATCGGCTGTGCCTCATGATGGCCCAGTACTTCCTGGGCA
GCGTGAGAAAGGAAGTGGGCAACAAGATTGTGTGGAGCAACGACAGCATCGAACTGCCTGTGGAAGGCTCTGTGGG
AAATGAGAAGAGCATCGTGTTCCTCCGTGTCTGACTACGGCAAGCTGTACGTGCTGGACGATGCCGAATTCCTGGGC
CGGATCTGCGAATACTTCATGCCCCACGAAAAGGGCAAGATCCGGTACCACACAGTGTACGAAAAGGGCTTTAGAG
CATAACGACCTGCAGAAGAAGTGCGTGGAGGCCGTGCTGGCTTTTCGAAGAGAAGGTGGTGAAGGCCAAGAAGAT
GAGCGAGAAGGAAGGCGCCCACTACATCGACTTCGGGAGATCCTGGCCCAGACCATGTGCAAGGAGGCCGAGAAG
ACCGCAGTGAACAAGGTGAGACGCGCCTTCTTCCACCACCACCTGAAGTTCGTGATTGACGAGTTCGGCCTGTTCA
GCGACGTGATGAAGAAGTACGGCATCGAGAAGGAATGGAAGTTCCTGTCAAGTAA (SEQ ID NO:22)

[0414] ATGAAGGTGGAGAACATCAAGGAAAAGTCCAAGAAGGCTATGTATCTGATCAACCACTATGAAGGCC
TAAGAAGTGGTGCTTCGCCATCGTGCTGAATAGGGCCTGCGACAACCTATGAGGATAACCCCCACCTGTTTCAGCAAG
AGCCTGCTGGAATTTGAAAAGACCAGCAGAAAGGACTGGTTTCGACGAGGAGACCAGGGAAGTGGTGGAGCAGGCCG
ACACCGAGATCCAGCCCAACCCCAACCTGAAGCCTAACACCACCGCCAACAGAAAGCTGAAGGACATCCGGAACCTA
CTTCAGCCACCACTACCACAAGAATGAGTGCCTGTACTTCAAGAACGACGACCCTATCCGGTGCATCATGGAGGCA
GCCTACGAGAAGTCCAAGATCTACATCAAGGGCAAGCAGATTGAGCAGTCCGACATCCCCCTCCCTGAGCTGTTTG
AGTCTAGCGGCTGGATCACCCAGCCGGCATCCTGCTGCTGGCCAGCTTCTTTGTGGAGAGAGGCATTCTGCACAG
ACTGATGGGCAACATCGGCGGCTTCAAGGACAACCGGGGCGAATACGGACTGACCCACGATATCTTACCACCTAC
TGCCTGAAGGGCAGCTACTCCATCAGAGCCCAGGACCACGACGCCGTGATGTTTCAGAGACATCCTGGGCTACCTGA
GCAGAGTGCCGACCGAGAGCTTTCAGCGCATCAAGCAGCCACAGATCAGAAAGGAGGGGCAGCTGAGCGAGCGGAA

GACAGACAAGTTTATCACCTTCGCCCTGAACTACCTGGAAGATTATGGACTGAAGGATCTGGAAGGCTGCAAGGCC
TGCTTCGCCCCGAGCAAGATCGTGAGAGAGCAGGAGAACGTGGAAAGCATCAATGACAAGGAGTACAAGCCTCACG
AAAACAAGAAGAAGGTGGAAATCCACTTCGATCAGTCTAAGGAAGACCGGTTCTACATCAACCGGAACAACGTGAT
CCTGAAGATCCAGAAGAAGGACGGCCACAGCAACATCGTGAGAATGGGCGTGACGAGCTGAAGTATCTGGTGCTG
ATGTCCCTGGTGGGCAAGGCCAAGGAAGCCGTGGAGAAGATCGACAACCTACATCCAGGATCTGAGAGACCAGCTGC
CCTACATCGAGGGCAAGAACAAGGAAGAAATCAAGGAGTACGTGAGATTCTTCCCCAGATTTCATCAGATCCCACCT
GGGCTGCTGCAGATTAACGATGAGGAGAAGATCAAGGCCCGGCTGGACTATGTGAAGACAAAGTGGCTGGACAAG
AAGGAGAAGTCCAAGGAGCTGGAGCTGCACAAGAAGGGCCGGGATATCCTGCGGTACATCAACGAGCGGTGCGACC
GGGAGCTGAACCGGAACGTGTACAACCGGATCCTGGAGCTGCTGGTGAGCAAGGACCTGACCGGCTTCTACCGGGA
GCTGGAGGAGCTGAAGCGGACCAGACGGATCGATAAGAACATTGTGCAGAACCTGTCCGGCCAGAAGACCATCAAC
GCCCTGCACGAAAAGGTGTGCGATCTCGTGCTGAAGGAGATCGAGAGCCTGGACACCGAGAACCTGCGGAAGTACC
TGGGCCTGATCCCCAAGGAGGAGAAGGAAGTGACCTTTAAGGAGAAGGTGGACAGGATCCTGAAGCAGCCGGTGAT
CTACAAGGGCTTCTGCGGTACCAAGTCTTCAAGGACGACAAGAAGAGCTTCGTGCTGCTGGTGAAGACGCCCTG
AAGGAGAAGGGAGGCGGCTGCGACGTGCCCCCTGGGCAAGGAGTACTACAAGATCGTGCCCTGGACAAGTATGACA
AGGAAAATAAGACCCTGTGCGAGACCCTGGCAATGGATAGACTGTGCTGATGATGGCCCGGAGTATTACCTGAG
CCTGAACGCCAAGCTGGCCCAGGAGGCCAGCAGATCGAATGGAAGAAGGAGGATAGCATTGAGCTGATCATCTTC
ACACTGAAGAATCCTGACCAGTCCAAGCAGAGCTTCTCCATCCGGTTCAGCGTGCGGGACTTCACCAAGCTGTACG
TGACCGACGACCCCGAATTCCTGGCCCGGCTGTGCAGCTACTTCTTCCCCGTGGAGAAGGAGATCGAATACCACAA
GCTGTACTCTGAAGGCATTAACAAGTACACCAACCTGCAGAAGGAGGGGATCGAAGCCATCCTGGAGCTGGAGAAG
AAGCTGATCGAAAGAAACCGGATCCAGTCCGCCAAGAAGTACCTGAGCTTTAACGAAATCATGAACAAGAGCGGCT
ACAACAAGGATGAGCAGGATGACCTGAAGAAGGTGAGGAAGTCCCTGCTGCACTACAAGCTGATCTTCGAAAAGGA
GCACCTGAAGAAGTTCTATGAAGTGATGCGGGGCGAGGGAATCGAGAAGAAGTGGTCCCTGATCGTGTA (SEQ ID
NO: 23)

[0415] ATGAATGGCATCGAGCTGAAGAAGGAAGAAGCCGCTTCTACTTCAATCAGGCCGAGCTGAACCTGAA
GGCCATTGAGGACAACATCTTCGACAAGGAGAGACGGAAGACACTGCTGAACAACCCCCAGATCCTGGCCAAGATG
GAGAACTTTATCTTCAATTTCCGGGACGTGACCAAGAACGCCAAGGGCGAAATCGACTGCCTGCTGCTGAAGCTGA
GAGAGCTGCGGAACCTTTTACAGCCACTACGTGCACAAGCGGGACGTGAGAGAACTGAGCAAGGGCGAGAAGCCGAT
CCTGGAGAAGTACTACCAGTTCGCCATCGAATCCACCGGCTCTGAGAACGTGAAGCTCGAAATCATCGAAAACGAC
GCCTGGCTGGCCGACGCCGGCGTGCTGTTCTTCTGTCATCTTCTGAAAGAAGAGCCAGGCAAACAAGCTGATCA
GCGGCATCAGCGGCTTCAAGAGAAACGACGACACCGGCCAGCCTCGGAGAAACCTGTTACCTACTTCTCCATCCG
GGAGGGCTACAAGGTGGTGCCCGAAATGCAGAAGCACTTCTGCTGTTCTCCCTGGTGAACCACCTGAGCAACCAG
GACGATTATATCGAAAAGGCCACCAGCCCTACGACATCGGCGAGGGCCTCTTCTCCACCGGATTGCCAGCACCT
TCCTGAACATCTCCGGAATCCTGAGAAACATGAAGTTCTACACCTATCAGAGCAAGAGACTGGTGGAGCAGAGAGG
CGAGCTGAAGCGGGAAAAGGACATCTTCGCCTGGGAAGAACCGTTTCAGGGCAATTCCTACTTTGAGATCAACGGC
CACAAGGGCGTGATTGGCGAAGACGAGCTGAAGGAGCTGTGCTACGCCTTCTGATCGGCAACCAGGACGCCAACA
AGGTGGAGGGCCGGATCACCCAGTTCCTGGAGAAGTTCAGAAAACGCCAACAGCGTGCAGCAGGTGAAGGACGACGA
GATGCTGAAGCCTGAATATTTCCCCGCCAACTACTTTGCCGAGAGCGGCGTGGGCCGGATCAAGGACCGGGTGCTG
AACAGACTGAACAAGGCCATCAAGAGCAACAAGGCCAAGAAGGGCGAGATCATCGCCTATGACAAGATGAGAGAAG
TGATGGCTTTCATCAATAACTCTCTGCCCCGTGGACGAGAAGCTGAAGCCCAAGGATTACAAGAGATACCTGGGCAT

GGTGAGATTCTGGGATAGAGAAAAGGACAATATCAAGCGCGAGTTCGAAACGAAGGAGTGGAGCAAGTATCTGCCC
TCCAACCTTCTGGACCGCCAAGAACCTGGAGAGAGTGTACGGACTGGCCCGGAAAAAGAACGCAGAGCTGTTTAACA
AGCTGAAGGCCGACGTGGAGAAGATGGACGAAAGAGAGCTGGAAAAGTATCAGAAGATCAACGACGCCAAGGATCT
GGCCAACCTGCGGCGGCTGGCCAGCGACTTCGGAGTGAAGTGGGAGGAGAAGGATTGGGACGAGTACTCCGGCCAG
ATCAAGAAGCAGATCACAGATTCCCAGAAGCTGACCATCATGAAGCAGAGAATCACAGCCGGCCTGAAGAAGAAGC
ACGGCATCGAAAACCTGAACCTGAGGATCACCATCGACATCAACAAGTCCAGAAAAGGCCGTGCTGAATCGGATCGC
CATCCCCAGAGGATTTGTGAAGCGGCACATCCTGGGCTGGCAGGAATCCGAGAAGGTGAGCAAGAAGATCAGAGAA
GCCGAATGCGAGATTCTGCTGAGCAAGGAGTACGAGGAGCTGAGCAAGCAGTTCTTTCAGAGCAAGGACTACGACA
AGATGACCCGCATCAACGGCCTGTACGAGAAGAATAAGCTGATCGCCCTGATGGCCGTGTATCTGATGGGGCAGCT
GAGAATCCTGTTCAAGGAGCACACCAAGCTGGACGACATCACCAAGACCACCGTGGATTTCAAGATCAGCGACAAG
GTGACCGTGAAGATCCCCTTCTCCAATATCCCTCCCTGGTGTACACCATGAGCAGCAAGTACGTGGACAATATCG
GCAACTACGGCTTCAGCAACAAGGACAAGGATAAGCCCATTCTGGGCAAGATCGACGTGATCGAGAAGCAGCGGAT
GGAGTTTATCAAGGAGGTGCTGGGATTTCGAGAAGTACCTGTTTGACGATAAGATCATCGACAAGAGCAAGTTTCGCC
GACACCGCCACCCACATCAGCTTTGCCGAAATCGTGAAGAAGTGGTGGAGAAGGGCTGGGACAAGGACCGGCTGA
CGAAGCTGAAGGATGCCCGGAACAAGGCCCTGCACGGCGAGATCCTGACCGGCACCAGCTTCGACGAGACAAAGTC
CCTGATCAACGAGCTGAAGAAGTAA (SEQ ID NO:24)

[0416] ATGAGCCCTGATTTTCATCAAGCTGGAGAAGCAGGAAGCAGCCTTCTACTTTAACCAGACCGAGCTGAA
CCTGAAGGCCATCGAATCCAATATCCTGGATAAGCAGCAGAGAATGATCCTGCTGAACAACCCAGAATCCTGGCC
AAGGTGGGCAACTTCATCTTCAATTTCCGGGACGTGACCAAGAACGCAAAGGGCGAAAATCGACTGCCTGCTGTTCA
AGCTGGAGGAACTGCGGAACTTCTACAGCCACTACGTGCACACCGATAACGTGAAGGAACTGTCCAACGGAGAGAA
GCCTCTGCTGGAGCGGTACTACCAGATCGCCATCCAGGCCACAAGAAGCGAGGACGTGAAGTTCGAGCTGTTTCGAG
ACCAGGAACGAGAACAAGATCACCGACGCAGGCGTGCTGTTCTTCTGTGCATGTTCTCTGAAGAAGAGCCAGGCTA
ATAAGCTGATTTCCGGCATCAGCGGCTTCAAGCGGAACGACCCACCGGCCAGCCAGACGGAACCTCTTTACCTA
CTTCTCTGCCCCGGGAGGGCTACAAGGCCCTGCCTGACATGCAGAAGCACTTCCTGCTGTTACCCCTGGTGAACCTAC
CTGAGCAACCAGGACGAGTACATCTCCGAGCTGAAGCAGTACGGAGAGATCGGACAGGGAGCCTTCTTCAACAGAA
TCGCCAGCACCTTCCTGAACATCAGCGGCATCAGCGGCAACACCAAGTTCTACAGCTACCAGAGCAAGAGAATCAA
GGAGCAGCGGGGCAACTGAACAGCGAAAAGGACAGCTTCGAGTGGATCGAGCCCTTTCAGGGCAACTCTTATTTT
GAGATCAACGGCCACAAGGGCGTGATCGGCGAAGACGAGCTGAAGGAGCTGTGCTACGCCCTGCTGGTGGCCAAGC
AGGACATCAATGCCGTGGAGGGAAAGATCATGCAGTTCCTGAAGAAGTTCAGGAACACCGGCAACCTGCAGCAGGT
GAAGGACGACGAGATGCTGGAATCGAGTACTTTCCCGCCAGCTACTTCAACGAGAGCAAGAAGGAGGACATCAAG
AAGGAGATCCTGGGCAGACTGGACAAGAAGATCCGGTCCTGCAGCGCCAAGGCCGAGAAGGCCTACGACAAGATGA
AGGAGGTGATGGAGTTTATCAATAACAGCCTGCCCGCCAGGAGAAGCTGAAGAGGAAGGACTACCGCAGATACCT
GAAGATGGTGAGATTCTGGTCCAGAGAAAAGGGCAACATCGAGAGAGAGTTCAGAACCAAGGAGTGGTCCAAGTAC
TTCAGCAGCGACTTCTGGAGAAAAGAACAATCTGGAGGATGTGTACAAGCTGGCCACCCAGAAGAACGCCGAGCTGT
TCAAGAATCTGAAGGCCGCCGCCGAGAAGATGGGCGAAACAGAATTCGAAAAGTACCAGCAGATCAACGATGTGAA
GGACCTGGCCAGCCTGAGACGGCTGACCCAGGATTTCCGCCTGAAGTGGGAGGAGAAGGATTGGGAGGAGTACAGC
GAACAGATCAAGAAGCAGATCACCGACCGGCAGAAGCTGACAATCATGAAGCAGCGGGTGACCGCCGAGCTGAAGA
AGAAGCACGGCATCGAGAATCTGAACCTCAGAATTACCATCGATTCCAACAAGAGCAGAAAGGCCGTGCTGAACAG
AATCGCCATTCCCCGGGGCTTCGTGAAGAAGCACATTCTGGGCTGGCAGGGCAGCGAAAAGATCAGCAAGAATATC

CGGGAGGCCGAGTGCAAGATCCTGCTGTCCAAGAAGTATGAGGAGCTGTCTCGGCAGTTCTTTGAGGCTGGCAACT
TCGACAAGCTGACCCAGATCAACGGCCTGTACGAAAAGAATAAGCTGACCGCCTTCATGTCCGTCTACCTGATGGG
CAGACTGAACATCCAGCTGAACAAGCACACGGAGCTGGGAAATCTGAAGAAGACCGAGGTGGACTTCAAGATTTCC
GACAAGGTGACAGAAAAGATCCCCCTTCTCCCAGTACCCTAGCCTGGTGTACGCTATGAGCCGGAAGTACGTGGACA
ACGTGGACAAGTACAAGTTCAGCCACCAGGACAAGAAGAAGCCCTTCTCTGGGCAAGATCGACAGCATCGAAAAGGA
GAGAATCGAATTCATCAAGGAGGTGCTGGACTTCGAAGAGTACCTGTTAAGAACAAGGTGATCGACAAGAGCAAG
TTCAGCGATACCGCCACCCATATCTCTTTCAAGGAAATCTGCGACGAGATGGGCAAGAAGGGCTGCAACCGCAACA
AGCTGACCGAGCTGAATAACGCTAGAAACGCCGACTGCACGGAGAAATCCCCAGCGAGACCAGCTTCCGGGAGGC
CAAGCCCCTGATCAACGAACTGAAGAAGTAA (SEQ ID NO:25)

[0417] ATGAGCCCTGACTTCATCAAGCTGGAAAAGCAGGAAGCCGCTTCTACTTTAATCAGACCGAGCTGAA
CCTGAAGGCCATCGAGAGCAACATCTTCGACAAGCAGCAGCGGTGATCCTGCTGAATAACCCCCAGATCCTGGCC
AAGGTGGGCGACTTCATCTTCAACTTCCGGGACGTGACCAAGAACGCCAAGGGAGAAATCGACTGCCTGCTGCTGA
AGCTGCGGGAGCTGAGAACTTCTACAGCCACTATGTGTACACCGACGACGTGAAGATCCTGAGCAACGGCGAGAG
GCCCCCTGCTGGAGAAGTACTACCAGTTTGCCATCGAGGCCACCGGATCTGAGAATGTGAAGCTGGAGATCATCGAG
AGCAACAACCGGTGACCGAAGCGGGCGTGTGTTCTTCTGTGCATGTTCTTGAAGAAGAGCCAGGCCAACAAGC
TGATTTCCGGCATCTCCGATTCAAGCGCAACGACCCTACCGGACAGCCTCGGCGGAACCTGTTACCTACTTTAG
CGTGCGGGAGGGCTACAAGGTGGTGCCCGACATGCAGAAGCACTTCTGCTGTTCTGTGCTGGTGAACCACCTGTCC
GGCCAGGATGACTATATTGAGAAGGCCCAGAAGCCCTACGACATCGGCGAAGGCCTGTTCTTCCACAGAATCGCCA
GCACCTTTCTCAACATCAGCGGCATCCTGAGAAACATGGAATTCTACATCTACCAGAGCAAGCGGCTGAAGGAGCA
GCAGGGAGAGCTGAAGAGAGAGAAGGACATCTTCCCTTGGATCGAGCCTTCCAGGGCAACAGCTACTTTGAGATC
AACGGAAACAAGGGCATCATCGGCGAGGACGAACTGAAGGAACTGTGCTACGCCCTGCTGGTGGCCGGCAAGGACG
TGAGAGCCGTGGAAGGAAAGATCACCCAGTTCCTGGAGAAGTTCAAGAACGCCGATAACGCCCAGCAGGTGGAGAA
GGATGAAATGCTGGACCGGAACAACCTCCCTGCCAATTACTTTGCCGAAAGCAACATCGGCAGCATCAAGGAAAAG
ATCCTGAATAGACTGGGCAAGACCGACGACTCCTACAACAAGACCGGCACCAAGATCAAGCCCTACGACATGATGA
AGGAGGTGATGGAGTTCATCAATAATTCTCTGCCCCCGCATGAGAAGCTGAAGCGGAAGGACTACCGGAGATACCT
GAAGATGGTCCGATCTGGGACAGCGAAAAGGACAATATCAAGCGGGAGTTTGAGAGCAAGGAATGGAGCAAGTAT
TTCAGCAGCGACTTCTGGATGGCCAAGAACCTGGAAAGAGTGACGGCCTGGCCAGGGAAAAGAACGCCGAGCTGT
TTAACAAGCTGAAGGCCGTGGTGGAGAAGATGGACGAGCGGGAGTTCGAAAAGTACCGGCTGATCAACAGCGCCGA
AGACCTGGCCAGCCTGCGGAGACTGGCCAAGGACTTCGGCCTGAAGTGGGAGGAGAAGGACTGGCAGGAGTATTCT
GGCCAGATCAAGAAGCAGATCTCCGACAGACAGAAGCTGACAATTATGAAGCAGCGGATCACAGCCGAACTGAAGA
AGAAGCACGGAATCGAGAACCTGAATCTGCGGATCACCATCGACAGCAACAAGTCCAGAAAGGCCGTGCTGAACCG
GATCGCCGTGCCCCGGGGCTTCGTGAAGGAACACATCTGGGCTGGCAAGGCTCTGAAAAGGTGAGCAAGAAGACC
AGAGAAGCCAAGTGCAAGATCCTGCTGAGCAAGGAGTACGAGGAACTGAGCAAGCAGTTCTTTTCAGACACGGAATT
ACGACAAGATGACCCAGGTGAACGGCCTGTACGAGAAGAACAAGCTGCTGGCCTTCATGGTGGTGTACCTGATGGA
GAGACTGAACATCCTGCTGAACAAGCCCACAGAGCTGAACGAACTGGAAAAGGCCGAAGTGGACTTCAAGATCTCC
GACAAGGTGATGGCCAAGATCCCTTTCTCTCAGTACCCAGCCTGGTGTATGCAATGAGCTCCAAGTACGCCGACA
GCGTGGGCTCTTACAAGTTCGAAAACGACGAGAAGAACAAGCCCTTTCTGGGCAAGATCGACACAATCGAGAAGCA
GAGAATGGAGTTCATCAAGGAGGTGCTGGGCTTCGAGGAATACCTGTTTCGAGAAGAAGATCATCGATAAGAGCGAA
TTCGCCGACACCGCCACCCACATCAGCTTCGACGAGATCTGCAACGAGCTGATCAAGAAGGGCTGGGACAAGGACA

AGCTGACCAAGCTGAAGGACGCCCGAACGCCGCCCTGCACGGCGAGATCCCCGCCGAGACCAGCTTCCGGGAGGC
CAAGCCCCTGATTAACGGCCTGAAGAAGTAA (SEQ ID NO:26)

[0418] ATGAACATCATCAAGCTGAAGAAGGAGGAAGCCGCCTTTTACTTTAACCAGACAATCCTGAATCTGAG
CGGCCTGGACGAGATCATCGAGAAGCAGATCCCCACATCATCTCCAATAAGGAAAACGCCAAGAAGGTGATTGAT
AAGATCTTCAATAACAGACTGCTGCTGAAGAGCGTGAAAACTATATCTACAACCTCAAGGACGTGGCCAAGAACG
CCCGGACCGAAATCGAAGCCATCCTGCTGAAGCTGGTGGAGCTGAGAACTTCTACTCCCCTACGTGCACAACGA
CACCGTGAAGATCCTGTCCAATGGCGAGAAGCCCATCCTGAAAAAGTACTACCAGATCGCCATCGAAGCCACCGGC
TCTAAGAACGTGAAGCTGGTCATTATCGAAAACAACAACCTGCCTGACCGACTCCGGCGTGCTGTTCTGCTGTGCA
TGTTCTCTGAAGAAGAGCCAGGCCAACAAGCTGATTAGCAGCGTGAGCGGCTTTAAGCGGAACGACAAGGAAGGCCA
GCCCAGAAGGAACCTCTTTACTTACTATAGCGTGAGGGAAGGCTACAAGGTGGTGCCAGACATGCAGAAGCACTTC
CTGCTGTTCCGCCCTGGTCAACCACCTGTCCGAGCAGGACGACCACATCGAGAAGCAGCAGAGCGACGAGCTGG
GCAAGGGCCTGTTCTTCCACAGAATCGCCAGCACATTCTGAATGAAAGCGGCATCTTCAACAAGATGCAGTTTTTA
CACCTACCAGAGCAATCGGCTGAAGGAGAAGCGGGGCGAGCTGAAGCACGAGAAGGACACCTTCACCTGGATCGAG
CCTTTCCAGGGAAACAGCTACTTCACCCTGAACGGGCACAAGGGCGTGATCAGCGAGGATCAGCTGAAGGAAGTGT
GCTACACAATCCTGATCGAGAAGCAGAACGTGGACAGCCTGGAGGGCAAGATCATTAGTTCTCTGAAGAAGTTTCA
GAACGTGTCTAGCAAGCAGCAGGTGGATGAGGACGAGCTGCTGAAGCGGGAATACTTCCCCGCCAACTACTTCGGC
CGGGCCGGCACCGGCACCCCTGAAGGAGAAGATCCTGAACCGGCTGGACAAGCGGATGGACCCACCAGCAAGGTGA
CCGACAAGGCCTATGACAAGATGATCGAGGTGATGGAGTTCATCAACATGTGCCTGCCAGCGACGAGAAGCTGCG
GCAGAAGGATTACCGGAGATATCTGAAGATGGTCAGATTCTGGAACAAGGAGAAGCACAACATCAAGAGAGAATTC
GACAGCAAGAAGTGGACCAGATTCTGCCCACCGAGCTGTGGAATAAGCGGAACCTGGAGGAAGCCTACCAGCTGG
CCCGGAAGGAGAACAAGAAGAAGCTGGAGGACATGAGGAATCAGGTGAGGAGCCTGAAGGAGAACGACCTGGAGAA
GTACCAGCAGATCAACTATGTGAACGACCTGAAAAACCTGCGGCTGCTGTCCCAAGAGCTGGGCGTGAAGTGGCAG
GAGAAGGACTGGGTGGAATACAGCGGCCAGATCAAGAAGCAGATCAGCGATAACCAGAAGCTGACAATCATGAAGC
AGAGAATCACCGCCGAGCTGAAGAAGATGCACGGCATCGAGAACCTGAACCTGAGAATCAGCATCGACACCAACAA
GTCCCGGCAGACTGTGATGAACAGAATTGCCCTGCCCAAGGGCTTCGTGAAGAACCACATTCAGCAGAACAGCAGC
GAGAAGATCAGCAAGAGAATCAGAGAGGACTACTGCAAGATCGAGCTGTCCGGCAAGTACGAAGAGCTGAGCAGAC
AGTTTTTTCGACAAGAAGAACTTTGACAAGATGACCCTGATCAACGGACTGTGCGAGAAGAATAAGCTCATCGCCTT
CATGGTGATTTACCTGCTGGAGCGGCTGGGCTTCGAGCTGAAGGAGAAGACCAAGCTGGGCGAGCTGAAGCAGACC
CGGATGACATATAAGATCAGCGACAAGGTGAAGGAGGACATCCCCCTCTCCTACTACCCCAAGCTGGTGTACGCCA
TGAATCGGAAGTATGTGGACAACATCGATAGCTACGCCTTCGCCGCTACGAGTCTAAGAAGGCCATCCTGGACAA
GGTGGACATCATTGAGAAGCAGAGAATGGAATTCATCAAGCAGGTGCTGTGCTTCGAGGAATACATCTTCGAGAAC
AGAATCATCGAGAAGAGCAAGTTCAACGATGAGGAGACCCACATCAGCTTCACCCAGATCCACGACGAAGTATCA
AGAAGGGCAGAGATACCGAAAAGCTGAGCAAGCTGAAGCACGCCAGAAAACAAGGCCCTGCACGGCGAGATCCCCGA
CGGGACCAGCTTTGAGAAGGCCAAGCTGCTGATCAACGAAATCAAGAAGTAA (SEQ ID NO:27)

[0419] ATGAACGCCATCGAGCTGAAGAAGGAAGAGGCCGCCTTCTACTTCAACCAGGCCAGACTGAACATCTC
TGGCCTGGACGAAATCATCGAGAAGCAACTGCCACACATCGGCTCTAACAGAGAGAACGCCAAGAAGACTGTGGAC
ATGATCCTGGATAACCCCGAGGTGCTGAAGAAGATGGAAAACTACGTGTTCAACTCCCGCGATATTGCCAAGAATG
CCCGGGGCGAGCTGGAGGCCCTGCTGCTGAAGCTGGTGCAGCTGAGAACTTCTATAGCCACTACGTGCACAAGGA
CGACGTCAAGACACTGAGCTACGGTGAGAAGCCTCTGCTGGATAAGTACTACGAGATCGCCATCGAAGCCACCGGA

TCCAAGGACGTGCGGCTGGAGATCATTGACGACAAGAATAAGCTGACCGACGCCGGAGTGCTGTTCTGCTGTGCA
TGTTCTCTGAAGAAGAGCGAGGCTAACAAGCTGATTTCCAGCATCCGGGGCTTCAAGAGGAACGACAAGGAGGGCCA
GCCTAGAAGAAACCTGTTACCTACTACAGCGTGAGAGAGGGCTATAAGGTGGTGCCCGACATGCAGAAGCACTTT
CTGCTGTTACCCCTGGTGAACCACCTGTCCAATCAGGACGAGTACATCTCCAACCTGCGCCCAAACCAGGAAATCG
GCCAGGGCGGATTTTTCCACCGGATCGCCAGCAAGTTCCTGAGCGACAGCGGAATCCTGCACAGCATGAAGTTCTA
CACATACAGATCCAAGCGGCTGACCGAGCAGCGGGGAGAGCTGAAGCCCAAGAAGGACCACTTTACATGGATCGAG
CCTTTCCAGGGCAATTCTTACTTCAGCGTGAGGGCCAGAAGGGCGTGATCGGAGAGGAGCAGCTCAAGGAGCTGT
GCTACGTGCTGCTGGTGGCCCGGGAGGACTTCAGAGCCGTGGAGGGCAAGGTGACCCAGTTCTGAAGAAGTTCCA
GAATGCCAATAACGTGCAGCAGGTGGAGAAGGACGAGGTGCTGGAAGAGGAGTACTCCCCGCCAACTACTTTGAG
AACCGGGACGTGGGAAGAGTCAAGGACAAGATCCTGAACAGACTGAAGAAGATCACCGAGAGTTATAAGGCCAAGG
GTAGAGAGGTGAAGGCCCTACGACAAGATGAAGGAAGTGATGGAGTTCATCAACAACCTGCCCGCCACCGATGAAAA
CCTGAAGCTGAAGGACTACCGGCGGTACCTGAAGATGGTGAGATTCTGGGGCAGAGAGAAGGAAAACATCAAGCGG
GAGTTCGACTCCAAGAAGTGAGGAGCGTTTCTCCCCGGGAGCTGTGGCAGAAGAGAAACCTGGAGGACGCCTACC
AGCTCGCCAAGGAGAAGAACACAGAGCTGTTCAACAAGCTGAAGACCACCGTGGAGAGAATGAACGAACCTGGAGTT
CGAGAAGTACCAGCAGATCAATGACGCCAAGGACCTGGCCAACCTGAGACAGCTGGCCAGAGACTTTGGAGTGAAG
TGGGAGGAAAAGGACTGGCAGGAATACTCTGGACAGATCAAGAAGCAGATCACCGACCGGCAGAAGCTGACCATCA
TGAAGCAGCGGATCACCGCCGCCCTGAAGAAGAAGCAGGGAATCGAAAACCTGAACCTGAGAATCACAACAGATAC
GAATAAGAGCAGGAAGGTGGTGCTGAACCGGATCGCACTGCCCAAGGGATTTCGTCAGAAAGCACATCCTGAAGACC
GACATCAAGATCAGCAAGCAGATCCGGCAGAGCCAGTGCCCTATCATCCTGTCTAACAACCTACATGAAGCTGGCCA
AGGAGTTCTTTGAAGAGCGGAACCTCGATAAGATGACCCAGATCAATGGCCTGTTCGAGAAGAACGTGCTGATCGC
CTTCATGATCGTGACCTGATGGAGCAGCTGAACCTGAGACTGGGCAAGAACACCGAGCTGTCCAACCTGAAGAAG
ACCGAGGTGAACCTTTACCATCACCGACAAGGTGACCGAGAAGGTGCAAATCTCCAGTACCCAGCCTGGTGTTCG
CCATTAACCGGGAGTACGTGGACGGCATCAGCGGCTACAAGCTGCCCCCAAGAAGCCCAAGGAACCTCCCTACAC
CTTCTTCGAAAAGATCGACGCCATCGAAAAGGAGCGGATGGAATTCATCAAGCAGGTGCTGGGCTTCGAGGAGCAC
CTCTTCGAAAAGAACGTGATCGACAAGACCCGTTTACCGACACCGCCACCCACATCAGCTTCAATGAGATCTGCG
ATGAGCTGATCAAGAAGGGCTGGGACGAAAACAAGATCATCAAGCTGAAGGATGCACGGAACGCTGCCCTGCACGG
CAAGATCCCTGAAGATACCTCCTTTGACGAAGCCAAGGTGCTGATCAACGAACCTGAAGAAGTAA (SEQ ID NO:
28)

[0420] 图1显示了七种CRISPR/Cas13e和Cas13f基因座结构。

[0421] 使用RNAfold对pre-crRNA中的七个DR序列的RNA二级结构进行了进一步分析。图2显示了结果。很明显，全部都共享了非常保守的二级结构。

[0422] 例如，在Cas13e家族中，每个DR序列形成由以下组成的二级结构：4碱基对茎（5'-GCUG-3'）、随后是5+5个核苷酸的对称的凸起（不包括4个茎核苷酸）、进一步地随后是5碱基对茎（5'-GCC C/U C-3'）和末端8碱基环（5'-CGAUUUGU-3'，不包括2个茎核苷酸）。

[0423] 同样地，在Cas13f家族中，除了一个例外（Cas13f.4），每个DR序列形成由以下组成的二级结构：5碱基对茎（5' GCUGU3'）、随后是5+4个核苷酸的接近对称的凸起（不包括4个茎核苷酸）、进一步地随后是6碱基对茎（5' A/G CCUCG3'）和末端5碱基环（5' AUUUG3'，不包括2个茎核苷酸）。唯一的例外是Cas13f.4的DR，其中第二茎短1个碱基对，并在第一凸起处添加了2个额外的碱基以形成大致对称的6+5凸起。

[0424] 使用MAFFT对Cas13e和Cas13f蛋白以及先前鉴定的Cas13a、Cas13b、Cas13c和Cas13d家族蛋白进行的多序列比对揭示了Cas13e和Cas13f蛋白与Cas13b蛋白在系统发育树上相对最接近(图3)。

[0425] 此外,在RXXXXH基序相对于Cas蛋白的N-末端和C-末端的位置方面,与Cas13a、Cas13c和Cas13d相比,Cas13e和Cas13f蛋白以及Cas13b蛋白(在较小程度上)具有更接近它们的N-末端和C-末端的RXXXXH基序(参见图4)。

[0426] 然后使用了I-TASSER预测Cas13e蛋白的3D结构,随后使用PyMOL对预测结构进行可视化。尽管两个RXXXXH基序非常接近Cas13e.1的N-末端和C-末端,但它们在3D结构中非常接近(图5)。

[0427] 实施例2Cas13e是效应RNA酶

[0428] 为了确认新鉴定的Cas13e蛋白是在CRISPR/Cas系统中发挥作用的有效RNA酶,将Cas13e.1编码序列经密码子优化用于人表达(SEQ ID NO:22),并克隆到具有GFP基因的第一质粒中。同时,将靶向报告基因(mCherry)mRNA的指导RNA(gRNA)的编码序列克隆到具有GFP基因的第二质粒中。gRNA由侧翼为Cas13e.1的两个同向重复序列的间隔编码区(SEQ ID NO:29)组成。GFP和mCherry报告基因的序列分别为SEQ ID NO:30-31。

[0429] GCTGGAGCAGCCCCGATTTGTGGGGTGATTACAGCGGTCTTCGATATTCAAGCGTCGGAAGACCTGCTGGAGCAGCCCCGATTTGTGGGGTGATTACAGC(SEQ ID NO:29)

[0430] ATGGTGAGCAAGGGCGAGGAGGATAACATGGCCATCATCAAGGAGTTCATGCGCTTCAAGGTGCACATGGAGGGCTCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCTACGAGGGCACCCAGACGCCAAGCTGAAGGTGACCAAGGGTGGCCCCCTGCCCTTCGCTGGGACATCCTGTCCCCTCAGTTCATGTACGGCTCCAAGGCCTACGTGAAGCACCCCGCCGACATCCCCGACTACTTGAAGCTGTCCCTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCGAGTTCATCTACAAGGTGAAGCTGCGCGGCACCAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACCATGGGCTGGGAGGCCTCCTCCGAGCGGATGTACCCCGAGGACGGCGCCCTGAAGGGCGAGATCAAGCAGAGGCTGAAGCTGAAGGACGGCGGCCACTACGACGCTGAGGTCAAGACCACCTACAAGGCCAAGAAGCCCGTGACGTGCCGGCGCCTACAACGTC AACATCAAGTTGGACATCACCTCCCACAACGAGGACTACACCATCGTGGAACAGTACGAACGCGCCGAGGGCCGCC ACTCCACCGGCGGCATGGACGAGCTGTACAAGTAA(SEQ ID NO:30)

[0431] ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGACCAACCGCAAGCTGCCCGTGCCCTGGCCACCCCTCGTGACCACCTGACCTACGGCGTGAGTGTCTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATC GAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAAGCGCATCAAGGTGAAGTTCAGATCCGCCACAACATCGAGGACGGCAGCGTGACGTGCGCGACCACTACCAGCAGAACACCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCAC TACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCTGCTGGAGTTCGTGACCGCCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTGA(SEQ ID NO:31)

[0432] 将HEK293T细胞按照标准方案在24孔组织培养板中培养,并用于使用LIPOFECTAMINE® 3000和P3000™试剂进行的三重质粒转染,以引入分别编码Cas13e.1

蛋白、靶向mCherry的gRNA、和mCherry编码序列的三种质粒。在阴性对照实验中,没有使用编码靶向mCherry的gRNA的质粒,而是使用了编码非靶标gRNA的对照质粒。Cas13e.1和gRNA质粒中存在GFP编码序列,因此GFP的表达可用作转染成功/效率的内部对照。参见图6中的示意图。然后将经转染的HEK293T细胞在37℃、5%CO₂下孵育约24小时,然后在荧光显微镜下对细胞进行检查。

[0433] 如图7所示,经靶向mCherry的gRNA转染的细胞和经对照非靶向(NT) gRNA转染的细胞在明场显微镜下具有等同的生长和形态,并且两者中的GFP表达大致等同。然而,基于流式细胞术分析,来自mCherry表达的RFP信号显著降低了高达75%(图8)。这表明Cas13e可以利用靶向mCherry的gRNA有效地敲低mCherry mRNA水平,从而敲低mCherry蛋白表达。

[0434] 实施例3Cas13e的sgRNA的有效方向

[0435] 由于Cas13e系统理论上可以利用DR+间隔子(5' DR)或间隔子+DR(3' DR)取向,本实验旨在确定Cas13e使用的正确取向。

[0436] 使用与实施例2中类似的三重转染实验设置,发现只有3' DR取向(间隔子+DR)支持显著的mCherry敲低。这表明Cas13e使用其crRNA,所述crRNA具有在间隔子的3'端的DR序列。参见图9。

[0437] DR+间隔子(5' DR)和间隔子+DR(3' DR)的sgRNA分别为SEQ ID NO:32和33。

[0438] GCTGGAGCAGCCCCGATTTGTGGGGTGATTACAGCGGTCTTCGATATTCAAGCGTCGGAAGACCT (SEQ ID NO:32)

[0439] GGTCTTCGATATTCAAGCGTCGGAAGACCTGCTGGAGCAGCCCCGATTTGTGGGGTGATTACAGC (SEQ ID NO:33)

[0440] 实施例4间隔序列长度对Cas13e.1的特异性活性和附带活性的影响

[0441] 为了研究间隔序列长度对Cas13e.1的特异性活性和附带活性的影响,设计了一组靶向mCherry报告基因的sgRNA,间隔序列长度为20nt、25nt、30nt、35nt、40nt、45nt或50nt (SEQ ID NO:34-40)。

[0442] TTGGTGCCGCGCAGCTTCAC (SEQ ID NO:34)

[0443] TTGGTGCCGCGCAGCTTCACCTTGT (SEQ ID NO:35)

[0444] TTGGTGCCGCGCAGCTTCACCTTG TAGATG (SEQ ID NO:36)

[0445] TTGGTGCCGCGCAGCTTCACCTTG TAGATGAACTC (SEQ ID NO:37)

[0446] TTGGTGCCGCGCAGCTTCACCTTG TAGATGAACTCGCCGT (SEQ ID NO:38)

[0447] TTGGTGCCGCGCAGCTTCACCTTG TAGATGAACTCGCCGTCCTGC (SEQ ID NO:39)

[0448] TTGGTGCCGCGCAGCTTCACCTTG TAGATGAACTCGCCGTCCTGCAGGGA (SEQ ID NO:40)

[0449] 使用与实施例2中类似的三重转染实验设置,通过流式细胞术分析了mCherry和GFP基因的敲低效率。

[0450] mCherry和GFP敲低实验的结果分别显示了Cas13e.1的特异性活性和非特异性活性(附带活性)。发现间隔子长度在约30nt至约50nt之间的Cas13e.1具有高特异性活性。参见图10。同时,当间隔子长度为约30nt时,Cas13e.1具有最高的非特异性活性。参见图11。

[0451] 实施例5使用dCas13e.1-ADAR2DD融合物进行的单碱基RNA编辑

[0452] 为了测试Cas13e是否可用于RNA单碱基编辑,通过对两个RXXXXH基序进行突变以消除RNA酶活性来生成dCas13e.1。然后将具有E488Q和T375G双突变的高保真ADAR2DD突变

体与dCas13e.1的(C-末端)融合,以创建命名为dCas13e.1-ADAR2DD的推定A至G单碱基RNA编辑器。参见SEQ ID NO:41中的编码序列。

[0453] ATGCCCAAGAAGAAGCGGAAGGTGGCCCAGGTGAGCAAGCAGACCTCCAAGAAGAGGGAGCTGAGCAT
CGACGAGTACCAGGGCGCCCGGAAGTGGTGCTTACCATTGCCTTCAACAAGGCCCTGGTGAACCGGGACAAGAAC
GACGGCCTGTTCTGTGAAAAGCCTGCTGAGACACGAGAAGTACAGCAAGCAGACTGGTACGACGAAGATACCCGGG
CCCTGATCAAGTGAGCAGCCAGGCCGCCAACGCCAAGGCTGAAGCCCTGGCGAACTACTTCAGTGCTTACCGGCA
TAGCCCTGGCTGCCTGACCTTACCAGCCGAGGACGAAGTGCAGGACCATCATGGAGAGAGCCTATGAGCGGGCCATC
TTCGAGTGCAGAAGAAGAGAGACAGAGGTGATCATCGAGTTTCCCAGCCTGTTTCGAGGGCGACCGGATCACCACCG
CCGGCGTGGTGTTCCTGAGCTTTTCCTGGAAGAAGAGTGCTGGATCGGCTGTATGGAGCCGTGTCCGGCCT
GAAGAAGAATGAGGGACAGTACAAGCTGACCCGGAAGGCCCTGAGCATGTACTGCCTGAAGGACAGCAGATTACCC
AAGGCCTGGGATAAGCGGGTGCTGCTGTTTCAGAGACATCCTGGCCCAGCTGGGAAGAATCCCCGCCGAGGCCTACG
AGTACTACCACGGCGAGCAGGGTGATAAGAAGAGAGCTAACGACAATGAGGGCACAAATCCCAAGCGGCACAAGGA
CAAGTTCATCGAATTTGCACTGCACTACCTGGAAGCCCAGCACAGCGAGATCTGCTTCGGCAGACGCCACATCGTG
CGGGAAGAGGCCGGCGCCGGCGATGAGCACAAGAAGCACCGGACCAAGGGAAAGGTGGTGGTGGACTTCAGCAAGA
AGGACGAGGACCAGAGCTACTATATCTCCAAGAACAACGTGATCGTGCGGATCGACAAGAACGCCGGCCCTAGAAG
CTACCGGATGGGCCTGAACGAGCTGAAGTACCTCGTGCTGCTGAGCCTGCAGGGGAAGGGCGACGATGCCATCGCC
AAGCTGTACAGATACAGACAGCACGTGGAGAACATCCTGGATGTGGTGAAGGTGACCGATAAGGATAACCACGTGT
TCCTGCCCCGCTTCGTGCTGGAGCAGCACGGCATCGGCAGAAAGGCCTTCAAGCAGCGGATCGATGGACGGGTGAA
GCACGTGCGGGCGTGTGGGAGAAGAAGAAGGCCGCCACCAATGAAATGACCCTGCACGAGAAGGCCAGAGACATC
CTGCAGTACGTGAACGAAAAGTGCACCCGGTCTTCAACCCTGGCGAATACAACAGACTGCTGGTGTGCCTGGTGG
GCAAGGACGTGGAGAAGTTTCAGGCCGGCCTGAAGCGGCTGCAGCTGGCCGAAAGGATCGATGGCCGGGTGTACTC
CATCTTCGCCCAGACCAGCACCATCAATGAGATGCACCAGGTGGTGTGCGACCAGATCCTGAACCGGCTGTGCAGA
ATCGGCGACCAGAAGCTGTACGATTACGTGGGACTGGGCAAGAAGGACGAAATCGACTACAAGCAGAAGGTGGCCT
GGTTCAAGGAGCACATCAGCATCCGGAGAGGATTCTGAGAAAGAAGTTCTGGTACGATAGCAAGAAGGGATTTCGC
AAAGCTGGTGGAGGAACACCTGGAGTCCGGCGGGCGCCAGCGGACGTGGGCCTGGACAAGAAGTACTACCACATC
GACGCCATCGGCAGATTCGAGGGCGCCAACCCCGCCCTGTACGAGACCCTGGCCAGAGATCGGCTGTGCCTCATGA
TGGCCAGTACTTCTGGGCAGCGTGAGAAAGGAAGTGGGCAACAAGATTGTGTGGAGCAACGACAGCATCGAACT
GCCTGTGGAAGGCTCTGTGGGAAATGAGAAGAGCATCGTGTTCTCCGTGTCTGACTACGGCAAGCTGTACGTGCTG
GACGATGCCGAATTCCTGGGCCGGATCTGCGAATACTTCATGCCCCACGAAAAGGGCAAGATCCGGTACCACACAG
TGTACGAAAAGGGCTTTAGAGCATACAACGACCTGCAGAAGAAGTGCGTGGAGGCCGTGCTGGCTTTTCAAGAGAA
GGTGGTGAAGGCCAAGAAGATGAGCGAGAAGGAAGGCGCCACTACATCGACTTCCGGGAGATCCTGGCCCAGACC
ATGTGCAAGGAGGCCGAGAAGACCGCAGTGAACAAGGTGGCGGCTGCCTTCTTCGCTGCGCACCTGAAGTTCGTGA
TTGACGAGTTCGGCCTGTTTCAGCGACGTGATGAAGAAGTACGGCATCGAGAAGGAATGGAAGTTCCTGTCAAGCC
CAAGAAGAAGCGGAAGGTGGGTGGAGGCGGAGGTTCTGGGGGAGGAGGTAGTGGCGGTGGTGGTTCAGGAGGCGGC
GGAAGCCAGCTGCATTTACCGCAGGTTTTAGCTGACGCTGTCTCACGCCTGGTCCTGGGTAAGTTTGGTGACCTGA
CCGACAACCTTCTCCTCCCTCACGCTCGCAGAAAAGTGCTGGCTGGAGTCGTCATGACAACAGGCACAGATGTTAA
AGATGCCAAGGTGATAAGTGTTCACAGGAGGCAAATGTATTAATGGTGAATACATGAGTGATCGTGGCCTTGCA
TTAAATGACTGCCATGCAGAAATAATATCTCGGAGATCCTTGCTCAGATTTCTTTATACACAACCTGAGCTTTACT
TAAATAACAAAGATGATCAAAAAAGATCCATCTTTCAGAAATCAGAGCGAGGGGGGTTTAGGCTGAAGGAGAATGT

CCAGTTTCATCTGTACATCAGCACCTCTCCCTGTGGAGATGCCAGAATCTTCTCACCACATGAGCCAATCCTGGAA
GAACCAGCAGATAGACACCCAAATCGTAAAGCAAGAGGACAGCTACGGACCAAAATAGAGTCTGGTCAGGGGACGA
TTCCAGTGCCTCCAATGCGAGCATCCAAACGTGGGACGGGGTGCTGCAAGGGGAGCGGCTGCTCACCATGTCCTG
CAGTGACAAGATTGCACGCTGGAACGTGGTGGGCATCCAGGGATCACTGCTCAGCATTTTCGTGGAGCCCATTAC
TTCTCGAGCATCATCCTGGGCAGCCTTTACCACGGGGACCACCTTTCCAGGGCCATGTACCAGCGGATCTCCAACA
TAGAGGACCTGCCACCTCTCTACACCCTCAACAAGCCTTTGCTCAGTGGCATCAGCAATGCAGAAGCACGGCAGCC
AGGGAAGGCCCCCAACTTCAGTGTCAACTGGACGGTAGGCGACTCCGCTATTGAGGTCATCAACGCCACGACTGGG
AAGGATGAGCTGGGCCGCGCTCCCGCCTGTGTAAGCACGCGTTGTACTGTCGCTGGATGCGTGTGCACGGCAAGG
TTCCCTCCCACTTACTACGCTCCAAGATTACCAAGCCCAACGTGTACCATGAGTCCAAGCTGGCGGCAAAGGAGTA
CCAGGCCGCCAAGGCGCGTCTGTTCACAGCCTTCATCAAGGCGGGGCTGGGGCCTGGGTGGAGAAGCCCACCGAG
CAGGACCAGTTCTCACTCACGTACCCATACGACGTACCAGATTACGCTTAA (SEQ ID NO:41)

[0454] 为了用作推定RNA碱基编辑子的靶标,野生型mCherry编码序列经突变以产生提前终止密码子TAG(参见SEQ ID NO:42中的粗体且加双下划线的序列),使得在没有通过RNA碱基编辑子将A校正为G的情况下不会产生功能性mCherry蛋白。参见图12和14。然后设计gRNA以实现所希望的A至G编辑(图12和14),并且使用标准方案将编码dCas13e.1-ADAR2DD碱基编辑子的CX530质粒、编码sgRNA的CX537/Cx538质粒、和编码突变mCherry基因的CX337质粒三重转染到HEK293T细胞中。将经转染的HEK293T细胞在37℃、5%CO₂下孵育24小时,然后使细胞经受流式细胞术以分离具有校正的mCherry mRNA并表达mCherry蛋白的细胞。参见说明性图——图12。图13显示了流式细胞术分析的结果。

[0455] 很明显,gRNA-1 (SEQ ID NO:43) 和gRNA-2 (SEQ ID NO:44) 两者都成功地校正了TAG提前终止密码子以生成功能性mCherry蛋白。

ATGGTGAGCAAGGGCGAGGAGGATAACATGGCCATCATCAAGGAGTTCATGCGCTTCAAGGTGCACATGGAGGGC
TCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCTACGAGGGCACCCAGACCGCCAAG
CTGAAGGTGACCAAGGGTGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCTCAGTTTCATGTACGGCTCCAAG
GCCTACGTGAAGCACCCCGCCGACATCCCCGACTACTTGAAGCTGTCCTTCCCCGAGGGCTTCAAGTAGGAGCGC
GTGATGAACTTCGAGGACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCGAGTTCATCTAC
AAGGTGAAGCTGCGCGGCACCAACTTCCCCTCCGACGCCCCGTAATGCAGAAGAAGACCATGGGCTGGGAGGCC
TCCTCCGAGCGGATGTACCCGAGGACGGCGCCCTGAAGGGCGAGATCAAGCAGAGGCTGAAGCTGAAGGACGGC
GGCCACTACGACGCTGAGGTCAAGACCACCTACAAGGCCAAGAAGCCCGTGACGCTGCCCGCGCCTACAACGTC
AACATCAAGTTGGACATCACCTCCCACAACGAGGACTACACCATCGTGGAACAGTACGAACGCGCCGAGGGCCGC
CACTCCACCGCGGCATGGACGAGCTGTACAAGTAA (SEQ ID NO: 42)

[0457] caagtagtcggggatgtcggggggtgcttcacCtaggccttgagccgtGCTGGAGCAGCCCCCGAT
TTGTGGGGTGATTACAGC (SEQ ID NO: 43)

[0458] cggggatgtcggggggtgcttcacCtaggccttgagccgtacatgaacGCTGGAGCAGCCCCCGAT
TTGTGGGGTGATTACAGC (SEQ ID NO: 44)

[0459] 实施例6使用缩短的dCas13e.1-ADAR2DD融合物进行的单碱基RNA编辑

[0460] 为了确定可用于RNA单碱基编辑的dCas13e.1的最小尺寸,生成了一系列五种表达dCas13e.1的渐进增大的C-末端缺失的构建体,每个构建体的C-末端减少30个残基(即,30个、60个、90个、120个和150个残基缺失)。所得的构建体用于创建在各自的C-末端处与高保真adar2 (ADAR2DD) 融合的dCas13e.1的编码序列。将这些构建体克隆到Vys15 (“V15”) 至

Vysz-19 (“V19”) 质粒中 (图15), 用于与实施例4中的实验类似的实验。在所有这些构建体中, 融合蛋白从CMV启动子 (pCMV) 和增强子 (eCMV) 表达, 并且紧接进一步增强蛋白表达的内含子的下游。两个核定位序列 (NLS) 位于融合物的dCas13e.1部分的N-末端和C-末端处, 并且ADAR2结构域 (如ADAR2DD) 通过NLS接头与C-末端NLS融合, 并由HA标签在C-末端处加标签。在EFS启动子 (pEFS) 独立控制下的EGFP编码序列存在于所有质粒的聚A添加序列的下游。

[0461] 有趣的是, 发现渐进C-末端缺失稳定地增加了融合编辑子中的RNA碱基编辑活性, 使得具有150个C-末端残基缺失的编辑子 (在V19中) 展现了最高的碱基编辑活性。参见图16。然而, C-末端缺失180个残基似乎已经消除了碱基编辑活性, 这表明Cas13e.1的C-末端的最大/最佳缺失可能在150-180个残基之间。

[0462] 基于这一发现, 针对具有150个C-末端残基缺失的dCas13e.1生成了一系列N-末端缺失突变体。生成了七个这样的N-末端缺失突变体, 分别具有30个、60个、90个、120个、150个、180个和210个残基缺失。参见图17。图18中的结果显示出在以下突变体中观察到最好的RNA编辑活性: 所述突变体具有180个N-末端残基缺失和150个C-末端残基缺失, 即从775个残基的Cas13e.1蛋白中总共缺失330个残基, 以产生用于生成ADAR2DD融合物的445个残基的最佳dCas13e.1。

[0463] 实施例7使用不同的Cas13蛋白进行的哺乳动物内源性mRNA敲低效率比较

[0464] 该实验表明, Cas13e和Cas13f蛋白 (尤其是Cas13f.1) 高效敲低了哺乳动物内源性靶mRNA, 比先前鉴定出的Cas13蛋白好。

[0465] 特别地, 构建了五种质粒, 每种质粒表达Cas13蛋白中的一种, 所述Cas13蛋白即Cas13e.1 (SEQ ID NO:22)、Cas13f.1 (SEQ ID NO:23)、LwaCas13a (SEQ ID NO:45)、PspCas13b (SEQ ID NO:46)、和RxCas13d (SEQ ID NO:47)。每种质粒还编码了mCherry报告基因, 以及侧翼为两个天然DR序列的各个Cas13蛋白的sgRNA/crRNA编码序列。这些sgRNA设计为具有靶向ANXA4 mRNA的间隔序列。参见SEQ ID NO:48-50。作为阴性对照, 构建了5种额外的质粒, 每种质粒编码非靶向sgRNA/crRNA, 而不是靶向ANXA4的sgRNA/crRNA (“对照NT构建体”)。参见图19。

[0466] ATGCCCAAGAAGAAGCGGAAGGTGGGATCCATGAAAGTGACCAAGGTCGATGGCATCAGCCACAAGAA
GTACATCGAAGAGGGCAAGCTCGTGAAGTCCACCAGCGAGGAAAACCGGACCAGCGAGAGACTGAGCGAGCTGCTG
AGCATCCGGCTGGACATCTACATCAAGAACCCGACAACGCCTCCGAGGAAGAGAACCGGATCAGAAGAGAGAACC
TGAAGAAGTTCTTTAGCAACAAGGTGCTGCACCTGAAGGACAGCGTGCTGTATCTGAAGAACCGGAAAGAAAAGAA
CGCCGTGCAGGACAAGAACTATAGCGAAGAGGACATCAGCGAGTACGACCTGAAAAACAAGAACAGCTTCTCCGTG
CTGAAGAAGATCCTGCTGAACGAGGACGTGAAGTCTGAGGAACTGGAAATCTTTCGGAAGGACGTGGAAGCCAAGC
TGAACAAGATCAACAGCCTGAAGTACAGCTTCAAGAGACAAGGCCAACTACCAGAAGATCAACGAGAACAACGT
GGAAAAAGTGGGCGGCAAGAGCAAGCGGAACATCATCTACGACTACTACAGAGAGAGCGCCAAGCGCAACGACTAC
ATCAACAACGTGCAGGAAGCCTTCGACAAGCTGTATAAGAAAGAGGATATCGAGAACTGTTTTCTGATCGAGA
ACAGCAAGAAGCACGAGAAGTACAAGATCCGCGAGTACTATCACAAGATCATCGGCCGGAAGAACGACAAAGAGAA
CTTCGCCAAGATTATCTACGAAGAGATCCAGAACGTGAACAACATCAAAGAGCTGATTGAGAAGATCCCCGACATG
TCTGAGCTGAAGAAAAGCCAGGTGTTCTACAAGTACTACCTGGACAAAAGAGGAACTGAACGACAAGAATATTAAGT
ACGCCTTCTGCCACTTCGTGGAAATCGAGATGTCCAGCTGCTGAAAACTACGTGTACAAGCGGCTGAGCAACAT

CAGCAACGATAAGATCAAGCGGATCTTCGAGTACCAGAATCTGAAAAAGCTGATCGAAAACAACTGCTGAACAAG
CTGGACACCTACGTGCGGAACTGCGGCAAGTACAACCTACTATCTGCAAGTGGGCGAGATCGCCACCTCCGACTTTA
TCGCCCCGGAACCGGCAGAACGAGGCCTTCCTGAGAAACATCATCGGCGTGTCCAGCGTGGCCTACTTCAGCCTGAG
GAACATCCTGGAACCGGAGAACGAGAACGATATCACCGGCCGGATGCGGGCAAGACCGTGAAGAACAACAAGGGC
GAAGAGAAATACGTGTCCGGCGAGGTGGACAAGATCTACAATGAGAACAAGCAGAACGAAGTGAAAGAAAATCTGA
AGATGTTCTACAGCTACGACTTCAACATGGACAACAAGAACGAGATCGAGGACTTCTTCGCCAACATCGACGAGGC
CATCAGCAGCATCAGACACGGCATCGTGCACCTTCAACCTGGAAGTGAAGGCAAGGACATCTTCGCCTTCAAGAAT
ATCGCCCCCAGCGAGATCTCCAAGAAGATGTTTCAGAACGAAATCAACGAAAAGAAGCTGAAGCTGAAAATCTTCA
AGCAGCTGAACAGCGCCAACTGTTCAACTACTACGAGAAGGATGTGATCATCAAGTACCTGAAGAATACCAAGTT
CAACTTCGTGAACAAAAACATCCCCTTCGTGCCAGCTTCACCAAGCTGTACAACAAGATTGAGGACCTGCGGAAT
ACCCTGAAGTTTTTTTGGAGCGTGCCCAAGGACAAAGAAGAGAAGGACGCCAGATCTACCTGCTGAAGAATATCT
ACTACGGCGAGTTCCTGAACAAGTTCGTGAAAACTCCAAGGTGTTCTTTAAGATCACCAATGAAGTGATCAAGAT
TAACAAGCAGCGGAACAGAAAACCGGCCACTACAAGTATCAGAAGTTCGAGAACATCGAGAAAACCGTGCCCGTG
GAATACCTGGCCATCATCCAGAGCAGAGAGATGATCAACAACCAGGACAAAGAGGAAAAGAATACCTACATCGACT
TTATTCAGCAGATTTTCTGAAGGGCTTCATCGACTACCTGAACAAGAACAATCTGAAGTATATCGAGAGCAACAA
CAACAATGACAACAACGACATCTTCTCCAAGATCAAGATCAAAAAGGATAACAAAGAGAAGTACGACAAGATCCTG
AAGAATATGAGAAGCACAATCGGAACAAAGAAATCCCTCACGAGATCAATGAGTTCGTGCGCGAGATCAAGCTGG
GGAAGATTCTGAAGTACACCGAGAATCTGAACATGTTTTACCTGATCCTGAAGCTGCTGAACCACAAAGAGCTGAC
CAACCTGAAGGGCAGCCTGGAAGTACCAGTCCGCCAACAAAGAAGAAACCTTCAGCGACGAGCTGGAAGTATC
AACCTGCTGAACCTGGACAACAACAGAGTGACCGAGGACTTCGAGCTGGAAGCCAACGAGATCGGCAAGTTCCTGG
ACTTCAACGAAAACAAAATCAAGGACCGGAAAGAGCTGAAAAAGTTCGACACCAACAAGATCTATTTGACGGCGA
GAACATCATCAAGCACCGGGCCTTCTACAATATCAAGAAATACGGCATGCTGAATCTGCTGAAAAGATCGCCGAT
AAGGCCAAGTATAAGATCAGCCTGAAAGAACTGAAAGAGTACAGCAACAAGAAGATGAGATTGAAAAGAACTACA
CCATGCAGCAGAACCTGCACCGGAAGTACGCCAGACCCAAGAAGGACGAAAAGTTCAACGACGAGGACTACAAAGA
GTATGAGAAGGCCATCGGCAACATCCAGAAGTACCCACCTGAAGAACAAGGTGGAATTCAATGAGCTGAACCTG
CTGCAGGGCCTGCTGCTGAAGATCCTGCACCGGCTCGTGGGCTACACCAGCATCTGGGAGCGGGACCTGAGATTCC
GGCTGAAGGGCGAGTTTCCCGAGAACCCTACATCGAGGAAATTTTCAATTTGACAACTCCAAGAATGTGAAGTA
CAAAAGCGGCCAGATCGTGAAAAGTATATCAACTTCTACAAAGAACTGTACAAGGACAATGTGAAAAGCGGAGC
ATCTACTCCGACAAGAAAGTGAAGAACTGAAGCAGGAAAAAAGGACCTGTACATCCGGAACCTACATTGCCACT
TCAACTACATCCCCACGCCGAGATTAGCCTGCTGGAAGTGTGAAAAACCTGCGGAAGCTGCTGTCTACGACCG
GAAGCTGAAGAACGCCATCATGAAGTCCATCGTGACATTCTGAAAAGATACGGCTTCGTGGCCACCTTCAAGATC
GGCGCTGACAAGAAGATCGAAATCCAGACCCTGGAATCAGAGAAGATCGTGACCTGAAGAATCTGAAGAAAAAGA
AACTGATGACCGACCGGAACAGCGAGGAACTGTGCGAACTCGTGAAAGTCATGTTTCGAGTACAAGGCCCTGGAATGA
(SEQ ID NO:45)

[0467] ATGCCAAGAAGAAGCGGAAGGTGGTCGACAACATCCCCGCTCTGGTGGAACCAGAAGAAGTACTT
TGGCACCTACAGCGTGATGGCCATGCTGAACGCTCAGACCGTGTGGACCACATCCAGAAGGTGGCCGATATTGAG
GGCGAGCAGAACGAGAACAACGAGAATCTGTGGTTTCACCCCGTGATGAGCCACCTGTACAACGCCAAGAACGGCT
ACGACAAGCAGCCCGAGAAAACCATGTTTCATCATCGAGCGGCTGCAGAGCTACTTCCATTCTGAAGATCATGGC
CGAGAACCAGAGAGAGTACAGCAACGGCAAGTACAAGCAGAACCGCGTGGAAGTGAACAGCAACGACATCTTCGAG

GTGCTGAAGCGCGCCTTCGGCGTGCTGAAGATGTACAGGGACCTGACCAACCACTACAAGACCTACGAGGAAAAGC
TGAACGACGGCTGCGAGTTCCTGACCAGCACAGAGCAACCTCTGAGCGGCATGATCAACAATACTACACAGTGGC
CCTGCGGAACATGAACGAGAGATACGGCTACAAGACAGAGGACCTGGCCTTCATCCAGGACAAGCGGTTCAAGTTC
GTGAAGGACGCCTACGGCAAGAAAAAGTCCCAAGTGAATACCGGATTCTTCCTGAGCCTGCAGGACTACAACGGCG
ACACACAGAAGAAGCTGCACCTGAGCGGAGTGGGAATCGCCCTGCTGATCTGCCTGTTCTCTGGACAAGCAGTACAT
CAACATCTTTCTGAGCAGGCTGCCCATCTTCTCCAGCTACAATGCCCAGAGCGAGGAACGGCGGATCATCATCAGA
TCCTTCGGCATCAACAGCATCAAGCTGCCCCAAGGACCGGATCCACAGCGAGAAGTCCAACAAGAGCGTGGCCATGG
ATATGCTCAACGAAGTGAAGCGGTGCCCCGACGAGCTGTTCAACAACACTGTCTGCCGAGAAGCAGTCCCGGTTTCA
AATCATCAGCGACGACCACAATGAAGTGCTGATGAAGCGGAGCAGCGACAGATTCTGTCCTCTGCTGCTGCAGTAT
ATCGATTACGGCAAGCTGTTTCGACCACATCAGGTTCCACGTGAACATGGGCAAGCTGAGATACCTGCTGAAGGCCG
ACAAGACCTGCATCGACGGCCAGACCAGAGTCAGAGTGATCGAGCAGCCCCTGAACGGCTTCGGCAGACTGGAAGA
GGCCGAGACAATGCGGAAGCAAGAGAACGGCACCTTCGGCAACAGCGGCATCCGGATCAGAGACTTCGAGAACATG
AAGCGGGACGACGCCAATCCTGCCAACTATCCCTACATCGTGACACCTACACACACTACATCCTGGAAAACAACA
AGGTGAGATGTTTATCAACGACAAAGAGGACAGCGCCCCACTGTCTGCCCCGTGATCGAGGATGATAGATACGTGGT
CAAGACAATCCCCAGCTGCCGGATGAGCACCTTGAAATTCCAGCCATGGCCTTCCACATGTTTCTGTTTCGGCAGC
AAGAAAACCGAGAAGCTGATCGTGACGTGCACAACCGGTACAAGAGACTGTTCCAGGCCATGCAGAAAGAAGAAG
TGACCGCCGAGAATATCGCCAGCTTCGGAATCGCCGAGAGCGACCTGCCTCAGAAGATCCTGGATCTGATCAGCGG
CAATGCCACGGCAAGGATGTGGACGCCTTCATCAGACTGACCGTGACGACATGCTGACCGACACCGAGCGGAGA
ATCAAGAGATTCAAGGACGACCGGAAGTCCATTTCGGAGCGCCGACAACAAGATGGGAAAGAGAGGCTTCAAGCAGA
TCTCCACAGGCAAGCTGGCCGACTTCCTGGCCAAGGACATCGTGCTGTTTCAGCCCAGCGTGAACGATGGCGAGAA
CAAGATCACCGCCTGAACTACCGGATCATGCAGAGCGCCATTGCCGTGTACGATAGCGGCGACGATTACGAGGCC
AAGCAGCAGTTCAAGCTGATGTTTCGAGAAGGCCCGGCTGATCGGCAAGGGCACAACAGAGCCTCATCCATTTCTGT
ACAAGGTGTTTCGCCCCGAGCATCCCCGCCAATGCCGTGAGTTCTACGAGCGCTACCTGATCGAGCGGAAGTTCTA
CCTGACCGGCCTGTCCAACGAGATCAAGAAAGGCAACAGAGTGATGTGCCCTTCATCCGGCGGGACCAGAACAAG
TGGA AACACCCGCCATGAAAACCTGGGCAGAATCTACAGCGAGGATCTGCCCCGTGAACTGCCAGACAGATGT
TCGACAATGAGATCAAGTCCCACCTGAAGTCCCTGCCACAGATGGAAGGCATCGACTTCAACAATGCCAACGTGAC
CTATCTGATCGCCGAGTACATGAAGAGAGTGCTGGACGACGACTTCCAGACCTTCTACCAGTGGAACCGCAACTAC
CGGTACATGGACATGCTTAAGGGCGAGTACGACAGAAAGGGCTCCCTGCAGCACTGCTTCACCAGCGTGGAAGAGA
GAGAAGGCCTCTGGAAGAGCGGGCTCCAGAACAGAGCGGTACAGAAAGCAGGCCAGCAACAAGATCCGCAGCAA
CCGGCAGATGAGAAACGCCAGCAGCGAAGAGATCGAGACAATCCTGGATAAGCGGTGAGCAACAGCCGGAACGAG
TACCAGAAAAGCGAGAAAAGTGATCCGGCGCTACAGAGTGCAGGATGCCCTGCTGTTTCTGCTGGCCAAAAAGACCC
TGACCGAACTGGCCGATTTTCGACGGCGAGAGGTTCAA ACTGAAAGAAATCATGCCCCGACGCCGAGAAGGGAATCCT
GAGCGAGATCATGCCATGAGCTTCACCTTCGAGAAAGGCGGCAAGAAGTACACCATCACCAGCGAGGGCATGAAG
CTGAAGAACTACGGCGACTTCTTTGTGCTGGCTAGCGACAAGAGGATCGGCAACCTGCTGGA ACTCGTGGGCAGCG
ACATCGTGTCCAAAGAGGATATCATGGAAGAGTTCAACAAATACGACCAGTGCAGGCCCCGAGATCAGCTCCATCGT
GTTCAACCTGGAAGAGTGGGCCTTCGACACATAACCCGAGCTGTCTGCCAGAGTGACCGGGAAGAGAAGGTGGAC
TTCAAGAGCATCCTGAAAATCCTGCTGAACAACAAGAACATCAACAAAGAGCAGAGCGACATCCTGCGGAAGATCC
GGAACGCCTTCGATCACAACAATTACCCGACAAAGGCGTGGTGAAATCAAGGCCCTGCCTGAGATCGCCATGAG
CATCAAGAAGGCCTTTGGGGAGTACGCCATCATGAAGGGATCCCTTCAATGA (SEQ ID NO: 46)

[0468] ATGCCTAAAAAGAAAAGAAAGGTGGGTCTGGTATCGAGAAGAAGAAGAGCTTCGCCAAGGGCATGGG
AGTGAAGAGCACCTGGTGTCCGGCTCTAAGGTGTACATGACCACATTTGCTGAGGGAAGCGACGCCAGGCTGGAG
AAGATCGTGAGGGCGATAGCATCAGATCCGTGAACGAGGGAGAGGCTTTCAGCGCCGAGATGGCTGACAAGAACG
CTGGCTACAAGATCGGAAACGCCAAGTTTTCCACCCAAAGGGCTACGCCGTGGTGGCTAACAACCCACTGTACAC
CGGACCAGTGCAGCAGGACATGCTGGGACTGAAGGAGACACTGGAGAAGAGGTACTIONCGGCGAGTCCGCCGACGGA
AACGATAACATCTGCATCCAGGTCATCCACAACATCCTGGATATCGAGAAGATCCTGGCTGAGTACATCACAAACG
CCGCTTACGCCGTGAACAACATCTCCGGCCTGGACAAGGATATCATCGGCTTCGGAAAGTTTTCTACCGTGTACAC
ATACGACGAGTTCAAGGATCCAGAGCACCACGGGCCGCTTTTAACAACAACGACAAGCTGATCAACGCCATCAAG
GCTCAGTACGACGAGTTCGATAACTTTCTGGATAACCCAGGCTGGGCTACTTCGGACAGGCTTTCTTTTCTAAGG
AGGGCAGAACTACATCATCAACTACGGAACGAGTGTTACGACATCCTGGCCCTGCTGAGCGGACTGAGGCACTG
GGTGGTGCACAACAACGAGGAGGAGTCTCGGATCAGCCGCACCTGGCTGTACAACCTGGACAAGAACCTGGATAAC
GAGTACATCTCCACACTGAACTACCTGTACGACAGGATCACCAACGAGCTGACAAAACAGCTTCTCCAAGAACTCTG
CCGCTAACGTGAACTACATCGCTGAGACCCTGGGCATCAACCCAGCTGAGTTCGCTGAGCAGTACTTCAGATTTTC
CATCATGAAGGAGCAGAAGAACCTGGGCTTCAACATCACAAAGCTGAGAGAAGTGATGCTGGACAGAAAGGATATG
TCCGAGATCAGGAAGAACCACAAGGTGTTTCGATTCTATCAGAACCAAGGTGTACACAATGATGGACTTTGTGATCT
ACAGGTACTACATCGAGGAGGATGCCAAGGTGGCCGCTGCCAACAAGAGCCTGCCCCGACAACGAGAAGTCTCTGAG
CGAGAAGGATATCTTCGTGATCAACCTGAGAGGCTCCTTTAACGACGATCAGAAGGACGCTCTGTACTACGATGAG
GCCAACAGGATCTGGAGAAAGCTGGAGAACATCATGCACAACATCAAGGAGTTCGGGGAAACAAGACCCGCGAGT
ACAAGAAGAAGGACGCTCCAAGGTGCCTAGGATCCTGCCTGCTGGAAGGGACGTGAGCGCCTTCAGCAAGCTGAT
GTACGCCCTGACAATGTTTCTGGACGGAAGGAGATCAACGATCTGCTGACCACACTGATCAACAAGTTCGACAAC
ATCCAGTCTTTTCTGAAAGTGATGCCTCTGATCGGCGTGAACGCTAAGTTCGTGGAGGAGTACGCCTTCTTTAAGG
ACAGCGCCAAGATCGCTGATGAGCTGCGGCTGATCAAGTCCTTTGCCAGGATGGGAGAGCCAATCGCTGACGCTAG
GAGAGCTATGTACATCGATGCCATCCGGATCCTGGGAACCAACCTGTCTTACGACGAGCTGAAGGCTCTGGCCGAC
ACCTTCAGCCTGGATGAGAACGGCAACAAGCTGAAGAAGGGCAAGCACGGAATGCGCAACTTCATCATCAACAACG
TGATCAGCAACAAGCGGTTTCACTACCTGATCAGATACGGCGACCCAGCTCACCTGCACGAGATCGCTAAGAACGA
GGCCGTGGTGAAGTTCGTGCTGGGACGGATCGCCGATATCCAGAAGAAGCAGGGCCAGAACGGAAAGAACAGATC
GACCGCTACTACGAGACCTGCATCGGCAAGGATAAGGGAAAGTCCGTGTCTGAGAAGGTGGACGCTCTGACCAAGA
TCATCACAGGCATGAACTACGACCAGTTCGATAAGAAGAGATCTGTGATCGAGGACACCGGAAGGGAGAACGCCGA
GAGAGAGAAGTTTAAGAAGATCATCAGCCTGTACCTGACAGTGATCTACCACATCCTGAAGAACATCGTGAACATC
AACGCTAGATACGTGATCGGCTTCCACTGCGTGGAGCGGATGCCAGCTGTACAAGGAGAAGGGATACGACATCA
ACCTGAAGAAGCTGGAGGAGAAGGGCTTTAGCTCCGTGACCAAGCTGTGCGCTGGAATCGACGAGACAGCCCCGA
CAAGAGGAAGGATGTGGAGAAGGAGATGGCCGAGAGAGCTAAGGAGAGCATCGACTCCCTGGAGTCTGCTAACCTT
AAGCTGTACGCCAACTACATCAAGTACTCCGATGAGAAGAAGGCCGAGGAGTTCACCAGGCAGATCAACAGAGAGA
AGGCCAAGACCGCTCTGAACGCCTACCTGAGGAACACAAAGTGAACGTGATCATCCGGGAGGACCTGCTGCGCAT
CGATAACAAGACCTGTACACTGTTCCGGAACAAGGCTGTGCACCTGGAGGTGGCTCGCTACGTGCACGCCTACATC
AACGACATCGCCGAGGTGAACTCCTACTTTTCAGCTGTACCACTACATCATGCAGAGGATCATCATGAACGAGAGAT
ACGAGAAGTCTAGCGGCAAGGTGTCTGAGTACTTCGACGCCGTGAACGATGAGAAGAAGTACAACGATAGACTGCT
GAAGCTGCTGTGCGTGCCTTTTCGGATACTGTATCCACGGTTTAAGAACCTGAGCATCGAGGCCCTGTTTCGACCGC
AACGAGGCTGCCAAGTTTGATAAGGAGAAGAAGAAGGTGAGCGGCAACTCCTGA (SEQ ID NO: 47)

[0469] ATGGCCCTTCGCAGCTCTTGACGTCATAC (SEQ ID NO:48)

[0470] TTAGGCAGCCCTCATCAGTGCCGGCTCCCT (SEQ ID NO:49)

[0471] GGCCAGGATCTCAATTAGGCAGCCCTCATC (SEQ ID NO:50)

[0472] 如在实施例4中,将五种编码Cas13/sgrNA的质粒转染到HEK293细胞中。在培养24小时后,通过流式细胞术分离表达mCherry的细胞,并使用RT-PCR确定ANXA4 mRNA的表达,以评估与由编码Cas13/NT的质粒转染的对照细胞相比的敲低效率。

[0473] 图20显示出Cas13b仅具有边缘ANXA4 mRNA敲低,而Cas13e.1、Cas13f.1和Cas13d均具有超过80%的靶ANXA4 mRNA敲低。在它们当中,Cas13e.1似乎具有最强的敲低效率。

序列表

<110> 辉大(上海)生物科技有限公司 (HUIGENE THERAPEUTICS CO., LTD.)

<120> VI-E 型和 VI-F 型 CRISPR-Cas 系统及其用途

<130> 201175

<160> 50

<170> SIPOSequenceListing 1.0

<210> 1

<211> 775

<212> PRT

<213> 宏基因组 (metagenomic)

<400> 1

[0001]

```

Met Ala Gln Val Ser Lys Gln Thr Ser Lys Lys Arg Glu Leu Ser Ile
1          5          10          15
Asp Glu Tyr Gln Gly Ala Arg Lys Trp Cys Phe Thr Ile Ala Phe Asn
          20          25          30
Lys Ala Leu Val Asn Arg Asp Lys Asn Asp Gly Leu Phe Val Glu Ser
          35          40          45
Leu Leu Arg His Glu Lys Tyr Ser Lys His Asp Trp Tyr Asp Glu Asp
          50          55          60
Thr Arg Ala Leu Ile Lys Cys Ser Thr Gln Ala Ala Asn Ala Lys Ala
65          70          75          80
Glu Ala Leu Arg Asn Tyr Phe Ser His Tyr Arg His Ser Pro Gly Cys
          85          90          95
Leu Thr Phe Thr Ala Glu Asp Glu Leu Arg Thr Ile Met Glu Arg Ala
          100          105          110
Tyr Glu Arg Ala Ile Phe Glu Cys Arg Arg Arg Glu Thr Glu Val Ile
          115          120          125
Ile Glu Phe Pro Ser Leu Phe Glu Gly Asp Arg Ile Thr Thr Ala Gly
          130          135          140
Val Val Phe Phe Val Ser Phe Phe Val Glu Arg Arg Val Leu Asp Arg
145          150          155          160
Leu Tyr Gly Ala Val Ser Gly Leu Lys Lys Asn Glu Gly Gln Tyr Lys
          165          170          175
Leu Thr Arg Lys Ala Leu Ser Met Tyr Cys Leu Lys Asp Ser Arg Phe
          180          185          190
Thr Lys Ala Trp Asp Lys Arg Val Leu Leu Phe Arg Asp Ile Leu Ala
          195          200          205
Gln Leu Gly Arg Ile Pro Ala Glu Ala Tyr Glu Tyr Tyr His Gly Glu
          210          215          220
Gln Gly Asp Lys Lys Arg Ala Asn Asp Asn Glu Gly Thr Asn Pro Lys
225          230          235          240
Arg His Lys Asp Lys Phe Ile Glu Phe Ala Leu His Tyr Leu Glu Ala
          245          250          255
Gln His Ser Glu Ile Cys Phe Gly Arg Arg His Ile Val Arg Glu Glu

```

	260	265	270
	Ala Gly Ala Gly Asp Glu His Lys Lys His Arg Thr Lys Gly Lys Val		
	275	280	285
	Val Val Asp Phe Ser Lys Lys Asp Glu Asp Gln Ser Tyr Tyr Ile Ser		
	290	295	300
	Lys Asn Asn Val Ile Val Arg Ile Asp Lys Asn Ala Gly Pro Arg Ser		
	305	310	315
	Tyr Arg Met Gly Leu Asn Glu Leu Lys Tyr Leu Val Leu Leu Ser Leu		
	325	330	335
	Gln Gly Lys Gly Asp Asp Ala Ile Ala Lys Leu Tyr Arg Tyr Arg Gln		
	340	345	350
	His Val Glu Asn Ile Leu Asp Val Val Lys Val Thr Asp Lys Asp Asn		
	355	360	365
	His Val Phe Leu Pro Arg Phe Val Leu Glu Gln His Gly Ile Gly Arg		
	370	375	380
	Lys Ala Phe Lys Gln Arg Ile Asp Gly Arg Val Lys His Val Arg Gly		
	385	390	395
	Val Trp Glu Lys Lys Lys Ala Ala Thr Asn Glu Met Thr Leu His Glu		
	405	410	415
	Lys Ala Arg Asp Ile Leu Gln Tyr Val Asn Glu Asn Cys Thr Arg Ser		
	420	425	430
	Phe Asn Pro Gly Glu Tyr Asn Arg Leu Leu Val Cys Leu Val Gly Lys		
	435	440	445
	Asp Val Glu Asn Phe Gln Ala Gly Leu Lys Arg Leu Gln Leu Ala Glu		
	450	455	460
[0002]	Arg Ile Asp Gly Arg Val Tyr Ser Ile Phe Ala Gln Thr Ser Thr Ile		
	465	470	475
	Asn Glu Met His Gln Val Val Cys Asp Gln Ile Leu Asn Arg Leu Cys		
	485	490	495
	Arg Ile Gly Asp Gln Lys Leu Tyr Asp Tyr Val Gly Leu Gly Lys Lys		
	500	505	510
	Asp Glu Ile Asp Tyr Lys Gln Lys Val Ala Trp Phe Lys Glu His Ile		
	515	520	525
	Ser Ile Arg Arg Gly Phe Leu Arg Lys Lys Phe Trp Tyr Asp Ser Lys		
	530	535	540
	Lys Gly Phe Ala Lys Leu Val Glu Glu His Leu Glu Ser Gly Gly Gly		
	545	550	555
	Gln Arg Asp Val Gly Leu Asp Lys Lys Tyr Tyr His Ile Asp Ala Ile		
	565	570	575
	Gly Arg Phe Glu Gly Ala Asn Pro Ala Leu Tyr Glu Thr Leu Ala Arg		
	580	585	590
	Asp Arg Leu Cys Leu Met Met Ala Gln Tyr Phe Leu Gly Ser Val Arg		
	595	600	605
	Lys Glu Leu Gly Asn Lys Ile Val Trp Ser Asn Asp Ser Ile Glu Leu		
	610	615	620
	Pro Val Glu Gly Ser Val Gly Asn Glu Lys Ser Ile Val Phe Ser Val		
	625	630	635
	Ser Asp Tyr Gly Lys Leu Tyr Val Leu Asp Asp Ala Glu Phe Leu Gly		
	645	650	655
	Arg Ile Cys Glu Tyr Phe Met Pro His Glu Lys Gly Lys Ile Arg Tyr		
	660	665	670

His Thr Val Tyr Glu Lys Gly Phe Arg Ala Tyr Asn Asp Leu Gln Lys
 675 680 685
 Lys Cys Val Glu Ala Val Leu Ala Phe Glu Glu Lys Val Val Lys Ala
 690 695 700
 Lys Lys Met Ser Glu Lys Glu Gly Ala His Tyr Ile Asp Phe Arg Glu
 705 710 715 720
 Ile Leu Ala Gln Thr Met Cys Lys Glu Ala Glu Lys Thr Ala Val Asn
 725 730 735
 Lys Val Arg Arg Ala Phe Phe His His His Leu Lys Phe Val Ile Asp
 740 745 750
 Glu Phe Gly Leu Phe Ser Asp Val Met Lys Lys Tyr Gly Ile Glu Lys
 755 760 765
 Glu Trp Lys Phe Pro Val Lys
 770 775

<210> 2
 <211> 805
 <212> PRT
 <213> 宏基因组 (metagenomic)

<400> 2
 Met Lys Val Glu Asn Ile Lys Glu Lys Ser Lys Lys Ala Met Tyr Leu
 1 5 10 15
 Ile Asn His Tyr Glu Gly Pro Lys Lys Trp Cys Phe Ala Ile Val Leu
 20 25 30
 [0003] Asn Arg Ala Cys Asp Asn Tyr Glu Asp Asn Pro His Leu Phe Ser Lys
 35 40 45
 Ser Leu Leu Glu Phe Glu Lys Thr Ser Arg Lys Asp Trp Phe Asp Glu
 50 55 60
 Glu Thr Arg Glu Leu Val Glu Gln Ala Asp Thr Glu Ile Gln Pro Asn
 65 70 75 80
 Pro Asn Leu Lys Pro Asn Thr Thr Ala Asn Arg Lys Leu Lys Asp Ile
 85 90 95
 Arg Asn Tyr Phe Ser His His Tyr His Lys Asn Glu Cys Leu Tyr Phe
 100 105 110
 Lys Asn Asp Asp Pro Ile Arg Cys Ile Met Glu Ala Ala Tyr Glu Lys
 115 120 125
 Ser Lys Ile Tyr Ile Lys Gly Lys Gln Ile Glu Gln Ser Asp Ile Pro
 130 135 140
 Leu Pro Glu Leu Phe Glu Ser Ser Gly Trp Ile Thr Pro Ala Gly Ile
 145 150 155 160
 Leu Leu Leu Ala Ser Phe Phe Val Glu Arg Gly Ile Leu His Arg Leu
 165 170 175
 Met Gly Asn Ile Gly Gly Phe Lys Asp Asn Arg Gly Glu Tyr Gly Leu
 180 185 190
 Thr His Asp Ile Phe Thr Thr Tyr Cys Leu Lys Gly Ser Tyr Ser Ile
 195 200 205
 Arg Ala Gln Asp His Asp Ala Val Met Phe Arg Asp Ile Leu Gly Tyr
 210 215 220
 Leu Ser Arg Val Pro Thr Glu Ser Phe Gln Arg Ile Lys Gln Pro Gln
 225 230 235 240

[0004]

```

Ile Arg Lys Glu Gly Gln Leu Ser Glu Arg Lys Thr Asp Lys Phe Ile
      245                      250                      255
Thr Phe Ala Leu Asn Tyr Leu Glu Asp Tyr Gly Leu Lys Asp Leu Glu
      260                      265                      270
Gly Cys Lys Ala Cys Phe Ala Arg Ser Lys Ile Val Arg Glu Gln Glu
      275                      280                      285
Asn Val Glu Ser Ile Asn Asp Lys Glu Tyr Lys Pro His Glu Asn Lys
      290                      295                      300
Lys Lys Val Glu Ile His Phe Asp Gln Ser Lys Glu Asp Arg Phe Tyr
      305                      310                      315                      320
Ile Asn Arg Asn Asn Val Ile Leu Lys Ile Gln Lys Lys Asp Gly His
      325                      330                      335
Ser Asn Ile Val Arg Met Gly Val Tyr Glu Leu Lys Tyr Leu Val Leu
      340                      345                      350
Met Ser Leu Val Gly Lys Ala Lys Glu Ala Val Glu Lys Ile Asp Asn
      355                      360                      365
Tyr Ile Gln Asp Leu Arg Asp Gln Leu Pro Tyr Ile Glu Gly Lys Asn
      370                      375                      380
Lys Glu Glu Ile Lys Glu Tyr Val Arg Phe Phe Pro Arg Phe Ile Arg
      385                      390                      395                      400
Ser His Leu Gly Leu Leu Gln Ile Asn Asp Glu Glu Lys Ile Lys Ala
      405                      410                      415
Arg Leu Asp Tyr Val Lys Thr Lys Trp Leu Asp Lys Lys Glu Lys Ser
      420                      425                      430
Lys Glu Leu Glu Leu His Lys Lys Gly Arg Asp Ile Leu Arg Tyr Ile
      435                      440                      445
Asn Glu Arg Cys Asp Arg Glu Leu Asn Arg Asn Val Tyr Asn Arg Ile
      450                      455                      460
Leu Glu Leu Leu Val Ser Lys Asp Leu Thr Gly Phe Tyr Arg Glu Leu
      465                      470                      475                      480
Glu Glu Leu Lys Arg Thr Arg Arg Ile Asp Lys Asn Ile Val Gln Asn
      485                      490                      495
Leu Ser Gly Gln Lys Thr Ile Asn Ala Leu His Glu Lys Val Cys Asp
      500                      505                      510
Leu Val Leu Lys Glu Ile Glu Ser Leu Asp Thr Glu Asn Leu Arg Lys
      515                      520                      525
Tyr Leu Gly Leu Ile Pro Lys Glu Glu Lys Glu Val Thr Phe Lys Glu
      530                      535                      540
Lys Val Asp Arg Ile Leu Lys Gln Pro Val Ile Tyr Lys Gly Phe Leu
      545                      550                      555                      560
Arg Tyr Gln Phe Phe Lys Asp Asp Lys Lys Ser Phe Val Leu Leu Val
      565                      570                      575
Glu Asp Ala Leu Lys Glu Lys Gly Gly Gly Cys Asp Val Pro Leu Gly
      580                      585                      590
Lys Glu Tyr Tyr Lys Ile Val Ser Leu Asp Lys Tyr Asp Lys Glu Asn
      595                      600                      605
Lys Thr Leu Cys Glu Thr Leu Ala Met Asp Arg Leu Cys Leu Met Met
      610                      615                      620
Ala Arg Gln Tyr Tyr Leu Ser Leu Asn Ala Lys Leu Ala Gln Glu Ala
      625                      630                      635                      640
Gln Gln Ile Glu Trp Lys Lys Glu Asp Ser Ile Glu Leu Ile Ile Phe

```

```

        645                650                655
Thr Leu Lys Asn Pro Asp Gln Ser Lys Gln Ser Phe Ser Ile Arg Phe
        660                665                670
Ser Val Arg Asp Phe Thr Lys Leu Tyr Val Thr Asp Asp Pro Glu Phe
        675                680                685
Leu Ala Arg Leu Cys Ser Tyr Phe Phe Pro Val Glu Lys Glu Ile Glu
        690                695                700
Tyr His Lys Leu Tyr Ser Glu Gly Ile Asn Lys Tyr Thr Asn Leu Gln
705                710                715                720
Lys Glu Gly Ile Glu Ala Ile Leu Glu Leu Glu Lys Lys Leu Ile Glu
        725                730                735
Arg Asn Arg Ile Gln Ser Ala Lys Asn Tyr Leu Ser Phe Asn Glu Ile
        740                745                750
Met Asn Lys Ser Gly Tyr Asn Lys Asp Glu Gln Asp Asp Leu Lys Lys
        755                760                765
Val Arg Asn Ser Leu Leu His Tyr Lys Leu Ile Phe Glu Lys Glu His
        770                775                780
Leu Lys Lys Phe Tyr Glu Val Met Arg Gly Glu Gly Ile Glu Lys Lys
785                790                795                800
Trp Ser Leu Ile Val
        805

```

<210> 3

<211> 790

<212> PRT

[0005] <213> 宏基因组 (metagenomic)

<400> 3

```

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

```

	180	185	190
	Leu Val Asn His Leu Ser Asn Gln Asp Asp Tyr Ile Glu Lys Ala His		
	195	200	205
	Gln Pro Tyr Asp Ile Gly Glu Gly Leu Phe Phe His Arg Ile Ala Ser		
	210	215	220
	Thr Phe Leu Asn Ile Ser Gly Ile Leu Arg Asn Met Lys Phe Tyr Thr		
	225	230	235
	Tyr Gln Ser Lys Arg Leu Val Glu Gln Arg Gly Glu Leu Lys Arg Glu		
	245	250	255
	Lys Asp Ile Phe Ala Trp Glu Glu Pro Phe Gln Gly Asn Ser Tyr Phe		
	260	265	270
	Glu Ile Asn Gly His Lys Gly Val Ile Gly Glu Asp Glu Leu Lys Glu		
	275	280	285
	Leu Cys Tyr Ala Phe Leu Ile Gly Asn Gln Asp Ala Asn Lys Val Glu		
	290	295	300
	Gly Arg Ile Thr Gln Phe Leu Glu Lys Phe Arg Asn Ala Asn Ser Val		
	305	310	315
	Gln Gln Val Lys Asp Asp Glu Met Leu Lys Pro Glu Tyr Phe Pro Ala		
	325	330	335
	Asn Tyr Phe Ala Glu Ser Gly Val Gly Arg Ile Lys Asp Arg Val Leu		
	340	345	350
	Asn Arg Leu Asn Lys Ala Ile Lys Ser Asn Lys Ala Lys Lys Gly Glu		
	355	360	365
	Ile Ile Ala Tyr Asp Lys Met Arg Glu Val Met Ala Phe Ile Asn Asn		
	370	375	380
[0006]	Ser Leu Pro Val Asp Glu Lys Leu Lys Pro Lys Asp Tyr Lys Arg Tyr		
	385	390	395
	Leu Gly Met Val Arg Phe Trp Asp Arg Glu Lys Asp Asn Ile Lys Arg		
	405	410	415
	Glu Phe Glu Thr Lys Glu Trp Ser Lys Tyr Leu Pro Ser Asn Phe Trp		
	420	425	430
	Thr Ala Lys Asn Leu Glu Arg Val Tyr Gly Leu Ala Arg Glu Lys Asn		
	435	440	445
	Ala Glu Leu Phe Asn Lys Leu Lys Ala Asp Val Glu Lys Met Asp Glu		
	450	455	460
	Arg Glu Leu Glu Lys Tyr Gln Lys Ile Asn Asp Ala Lys Asp Leu Ala		
	465	470	475
	Asn Leu Arg Arg Leu Ala Ser Asp Phe Gly Val Lys Trp Glu Glu Lys		
	485	490	495
	Asp Trp Asp Glu Tyr Ser Gly Gln Ile Lys Lys Gln Ile Thr Asp Ser		
	500	505	510
	Gln Lys Leu Thr Ile Met Lys Gln Arg Ile Thr Ala Gly Leu Lys Lys		
	515	520	525
	Lys His Gly Ile Glu Asn Leu Asn Leu Arg Ile Thr Ile Asp Ile Asn		
	530	535	540
	Lys Ser Arg Lys Ala Val Leu Asn Arg Ile Ala Ile Pro Arg Gly Phe		
	545	550	555
	Val Lys Arg His Ile Leu Gly Trp Gln Glu Ser Glu Lys Val Ser Lys		
	565	570	575
	Lys Ile Arg Glu Ala Glu Cys Glu Ile Leu Leu Ser Lys Glu Tyr Glu		
	580	585	590

Glu Leu Ser Lys Gln Phe Phe Gln Ser Lys Asp Tyr Asp Lys Met Thr
 595 600 605
 Arg Ile Asn Gly Leu Tyr Glu Lys Asn Lys Leu Ile Ala Leu Met Ala
 610 615 620
 Val Tyr Leu Met Gly Gln Leu Arg Ile Leu Phe Lys Glu His Thr Lys
 625 630 635 640
 Leu Asp Asp Ile Thr Lys Thr Thr Val Asp Phe Lys Ile Ser Asp Lys
 645 650 655
 Val Thr Val Lys Ile Pro Phe Ser Asn Tyr Pro Ser Leu Val Tyr Thr
 660 665 670
 Met Ser Ser Lys Tyr Val Asp Asn Ile Gly Asn Tyr Gly Phe Ser Asn
 675 680 685
 Lys Asp Lys Asp Lys Pro Ile Leu Gly Lys Ile Asp Val Ile Glu Lys
 690 695 700
 Gln Arg Met Glu Phe Ile Lys Glu Val Leu Gly Phe Glu Lys Tyr Leu
 705 710 715 720
 Phe Asp Asp Lys Ile Ile Asp Lys Ser Lys Phe Ala Asp Thr Ala Thr
 725 730 735
 His Ile Ser Phe Ala Glu Ile Val Glu Glu Leu Val Glu Lys Gly Trp
 740 745 750
 Asp Lys Asp Arg Leu Thr Lys Leu Lys Asp Ala Arg Asn Lys Ala Leu
 755 760 765
 His Gly Glu Ile Leu Thr Gly Thr Ser Phe Asp Glu Thr Lys Ser Leu
 770 775 780
 Ile Asn Glu Leu Lys Lys
 [0007] 785 790

<210> 4
 <211> 792
 <212> PRT
 <213> 宏基因组 (metagenomic)

<400> 4
 Met Ser Pro Asp Phe Ile Lys Leu Glu Lys Gln Glu Ala Ala Phe Tyr
 1 5 10 15
 Phe Asn Gln Thr Glu Leu Asn Leu Lys Ala Ile Glu Ser Asn Ile Leu
 20 25 30
 Asp Lys Gln Gln Arg Met Ile Leu Leu Asn Asn Pro Arg Ile Leu Ala
 35 40 45
 Lys Val Gly Asn Phe Ile Phe Asn Phe Arg Asp Val Thr Lys Asn Ala
 50 55 60
 Lys Gly Glu Ile Asp Cys Leu Leu Phe Lys Leu Glu Glu Leu Arg Asn
 65 70 75 80
 Phe Tyr Ser His Tyr Val His Thr Asp Asn Val Lys Glu Leu Ser Asn
 85 90 95
 Gly Glu Lys Pro Leu Leu Glu Arg Tyr Tyr Gln Ile Ala Ile Gln Ala
 100 105 110
 Thr Arg Ser Glu Asp Val Lys Phe Glu Leu Phe Glu Thr Arg Asn Glu
 115 120 125
 Asn Lys Ile Thr Asp Ala Gly Val Leu Phe Phe Leu Cys Met Phe Leu
 130 135 140

[0008]

```

Lys Lys Ser Gln Ala Asn Lys Leu Ile Ser Gly Ile Ser Gly Phe Lys
145                150                155                160
Arg Asn Asp Pro Thr Gly Gln Pro Arg Arg Asn Leu Phe Thr Tyr Phe
                165                170                175
Ser Ala Arg Glu Gly Tyr Lys Ala Leu Pro Asp Met Gln Lys His Phe
                180                185                190
Leu Leu Phe Thr Leu Val Asn Tyr Leu Ser Asn Gln Asp Glu Tyr Ile
                195                200                205
Ser Glu Leu Lys Gln Tyr Gly Glu Ile Gly Gln Gly Ala Phe Phe Asn
                210                215                220
Arg Ile Ala Ser Thr Phe Leu Asn Ile Ser Gly Ile Ser Gly Asn Thr
225                230                235                240
Lys Phe Tyr Ser Tyr Gln Ser Lys Arg Ile Lys Glu Gln Arg Gly Glu
                245                250                255
Leu Asn Ser Glu Lys Asp Ser Phe Glu Trp Ile Glu Pro Phe Gln Gly
                260                265                270
Asn Ser Tyr Phe Glu Ile Asn Gly His Lys Gly Val Ile Gly Glu Asp
                275                280                285
Glu Leu Lys Glu Leu Cys Tyr Ala Leu Leu Val Ala Lys Gln Asp Ile
                290                295                300
Asn Ala Val Glu Gly Lys Ile Met Gln Phe Leu Lys Lys Phe Arg Asn
305                310                315                320
Thr Gly Asn Leu Gln Gln Val Lys Asp Asp Glu Met Leu Glu Ile Glu
                325                330                335
Tyr Phe Pro Ala Ser Tyr Phe Asn Glu Ser Lys Lys Glu Asp Ile Lys
                340                345                350
Lys Glu Ile Leu Gly Arg Leu Asp Lys Lys Ile Arg Ser Cys Ser Ala
                355                360                365
Lys Ala Glu Lys Ala Tyr Asp Lys Met Lys Glu Val Met Glu Phe Ile
                370                375                380
Asn Asn Ser Leu Pro Ala Glu Glu Lys Leu Lys Arg Lys Asp Tyr Arg
385                390                395                400
Arg Tyr Leu Lys Met Val Arg Phe Trp Ser Arg Glu Lys Gly Asn Ile
                405                410                415
Glu Arg Glu Phe Arg Thr Lys Glu Trp Ser Lys Tyr Phe Ser Ser Asp
                420                425                430
Phe Trp Arg Lys Asn Asn Leu Glu Asp Val Tyr Lys Leu Ala Thr Gln
                435                440                445
Lys Asn Ala Glu Leu Phe Lys Asn Leu Lys Ala Ala Ala Glu Lys Met
                450                455                460
Gly Glu Thr Glu Phe Glu Lys Tyr Gln Gln Ile Asn Asp Val Lys Asp
465                470                475                480
Leu Ala Ser Leu Arg Arg Leu Thr Gln Asp Phe Gly Leu Lys Trp Glu
                485                490                495
Glu Lys Asp Trp Glu Glu Tyr Ser Glu Gln Ile Lys Lys Gln Ile Thr
                500                505                510
Asp Arg Gln Lys Leu Thr Ile Met Lys Gln Arg Val Thr Ala Glu Leu
                515                520                525
Lys Lys Lys His Gly Ile Glu Asn Leu Asn Leu Arg Ile Thr Ile Asp
                530                535                540
Ser Asn Lys Ser Arg Lys Ala Val Leu Asn Arg Ile Ala Ile Pro Arg

```

	545		550		555		560
	Gly Phe Val Lys Lys His Ile Leu Gly Trp Gln Gly Ser Glu Lys Ile						
		565		570		575	
	Ser Lys Asn Ile Arg Glu Ala Glu Cys Lys Ile Leu Leu Ser Lys Lys						
		580		585		590	
	Tyr Glu Glu Leu Ser Arg Gln Phe Phe Glu Ala Gly Asn Phe Asp Lys						
		595		600		605	
	Leu Thr Gln Ile Asn Gly Leu Tyr Glu Lys Asn Lys Leu Thr Ala Phe						
		610		615		620	
	Met Ser Val Tyr Leu Met Gly Arg Leu Asn Ile Gln Leu Asn Lys His						
	625		630		635		640
	Thr Glu Leu Gly Asn Leu Lys Lys Thr Glu Val Asp Phe Lys Ile Ser						
		645		650		655	
	Asp Lys Val Thr Glu Lys Ile Pro Phe Ser Gln Tyr Pro Ser Leu Val						
		660		665		670	
	Tyr Ala Met Ser Arg Lys Tyr Val Asp Asn Val Asp Lys Tyr Lys Phe						
		675		680		685	
	Ser His Gln Asp Lys Lys Lys Pro Phe Leu Gly Lys Ile Asp Ser Ile						
		690		695		700	
	Glu Lys Glu Arg Ile Glu Phe Ile Lys Glu Val Leu Asp Phe Glu Glu						
	705		710		715		720
	Tyr Leu Phe Lys Asn Lys Val Ile Asp Lys Ser Lys Phe Ser Asp Thr						
		725		730		735	
	Ala Thr His Ile Ser Phe Lys Glu Ile Cys Asp Glu Met Gly Lys Lys						
		740		745		750	
[0009]	Gly Cys Asn Arg Asn Lys Leu Thr Glu Leu Asn Asn Ala Arg Asn Ala						
		755		760		765	
	Ala Leu His Gly Glu Ile Pro Ser Glu Thr Ser Phe Arg Glu Ala Lys						
		770		775		780	
	Pro Leu Ile Asn Glu Leu Lys Lys						
	785		790				
	<210> 5						
	<211> 792						
	<212> PRT						
	<213> 宏基因组 (metagenomic)						
	<400> 5						
	Met Ser Pro Asp Phe Ile Lys Leu Glu Lys Gln Glu Ala Ala Phe Tyr						
	1	5	10	15			
	Phe Asn Gln Thr Glu Leu Asn Leu Lys Ala Ile Glu Ser Asn Ile Phe						
		20	25	30			
	Asp Lys Gln Gln Arg Val Ile Leu Leu Asn Asn Pro Gln Ile Leu Ala						
		35	40	45			
	Lys Val Gly Asp Phe Ile Phe Asn Phe Arg Asp Val Thr Lys Asn Ala						
		50	55	60			
	Lys Gly Glu Ile Asp Cys Leu Leu Lys Leu Arg Glu Leu Arg Asn						
	65	70	75	80			
	Phe Tyr Ser His Tyr Val Tyr Thr Asp Asp Val Lys Ile Leu Ser Asn						
		85	90	95			
	Gly Glu Arg Pro Leu Leu Glu Lys Tyr Tyr Gln Phe Ala Ile Glu Ala						

				100					105					110		
	Thr	Gly	Ser	Glu	Asn	Val	Lys	Leu	Glu	Ile	Ile	Glu	Ser	Asn	Asn	Arg
		115						120				125				
	Leu	Thr	Glu	Ala	Gly	Val	Leu	Phe	Phe	Leu	Cys	Met	Phe	Leu	Lys	Lys
		130					135				140					
	Ser	Gln	Ala	Asn	Lys	Leu	Ile	Ser	Gly	Ile	Ser	Gly	Phe	Lys	Arg	Asn
	145				150				155					160		
	Asp	Pro	Thr	Gly	Gln	Pro	Arg	Arg	Asn	Leu	Phe	Thr	Tyr	Phe	Ser	Val
				165				170					175			
	Arg	Glu	Gly	Tyr	Lys	Val	Val	Pro	Asp	Met	Gln	Lys	His	Phe	Leu	Leu
		180						185				190				
	Phe	Val	Leu	Val	Asn	His	Leu	Ser	Gly	Gln	Asp	Asp	Tyr	Ile	Glu	Lys
		195					200				205					
	Ala	Gln	Lys	Pro	Tyr	Asp	Ile	Gly	Glu	Gly	Leu	Phe	Phe	His	Arg	Ile
		210				215					220					
	Ala	Ser	Thr	Phe	Leu	Asn	Ile	Ser	Gly	Ile	Leu	Arg	Asn	Met	Glu	Phe
	225				230				235					240		
	Tyr	Ile	Tyr	Gln	Ser	Lys	Arg	Leu	Lys	Glu	Gln	Gln	Gly	Glu	Leu	Lys
			245					250				255				
	Arg	Glu	Lys	Asp	Ile	Phe	Pro	Trp	Ile	Glu	Pro	Phe	Gln	Gly	Asn	Ser
		260						265				270				
	Tyr	Phe	Glu	Ile	Asn	Gly	Asn	Lys	Gly	Ile	Ile	Gly	Glu	Asp	Glu	Leu
		275						280				285				
	Lys	Glu	Leu	Cys	Tyr	Ala	Leu	Leu	Val	Ala	Gly	Lys	Asp	Val	Arg	Ala
		290					295				300					
[0010]	Val	Glu	Gly	Lys	Ile	Thr	Gln	Phe	Leu	Glu	Lys	Phe	Lys	Asn	Ala	Asp
	305				310				315					320		
	Asn	Ala	Gln	Gln	Val	Glu	Lys	Asp	Glu	Met	Leu	Asp	Arg	Asn	Asn	Phe
			325					330				335				
	Pro	Ala	Asn	Tyr	Phe	Ala	Glu	Ser	Asn	Ile	Gly	Ser	Ile	Lys	Glu	Lys
		340						345				350				
	Ile	Leu	Asn	Arg	Leu	Gly	Lys	Thr	Asp	Asp	Ser	Tyr	Asn	Lys	Thr	Gly
		355						360				365				
	Thr	Lys	Ile	Lys	Pro	Tyr	Asp	Met	Met	Lys	Glu	Val	Met	Glu	Phe	Ile
		370					375				380					
	Asn	Asn	Ser	Leu	Pro	Ala	Asp	Glu	Lys	Leu	Lys	Arg	Lys	Asp	Tyr	Arg
	385				390				395					400		
	Arg	Tyr	Leu	Lys	Met	Val	Arg	Ile	Trp	Asp	Ser	Glu	Lys	Asp	Asn	Ile
			405					410				415				
	Lys	Arg	Glu	Phe	Glu	Ser	Lys	Glu	Trp	Ser	Lys	Tyr	Phe	Ser	Ser	Asp
		420						425				430				
	Phe	Trp	Met	Ala	Lys	Asn	Leu	Glu	Arg	Val	Tyr	Gly	Leu	Ala	Arg	Glu
		435					440					445				
	Lys	Asn	Ala	Glu	Leu	Phe	Asn	Lys	Leu	Lys	Ala	Val	Val	Glu	Lys	Met
		450					455				460					
	Asp	Glu	Arg	Glu	Phe	Glu	Lys	Tyr	Arg	Leu	Ile	Asn	Ser	Ala	Glu	Asp
	465				470				475					480		
	Leu	Ala	Ser	Leu	Arg	Arg	Leu	Ala	Lys	Asp	Phe	Gly	Leu	Lys	Trp	Glu
			485					490				495				
	Glu	Lys	Asp	Trp	Gln	Glu	Tyr	Ser	Gly	Gln	Ile	Lys	Lys	Gln	Ile	Ser
		500						505				510				

[0011]

```

Asp Arg Gln Lys Leu Thr Ile Met Lys Gln Arg Ile Thr Ala Glu Leu
515                      520                      525
Lys Lys Lys His Gly Ile Glu Asn Leu Asn Leu Arg Ile Thr Ile Asp
530                      535                      540
Ser Asn Lys Ser Arg Lys Ala Val Leu Asn Arg Ile Ala Val Pro Arg
545                      550                      555                      560
Gly Phe Val Lys Glu His Ile Leu Gly Trp Gln Gly Ser Glu Lys Val
565                      570                      575
Ser Lys Lys Thr Arg Glu Ala Lys Cys Lys Ile Leu Leu Ser Lys Glu
580                      585                      590
Tyr Glu Glu Leu Ser Lys Gln Phe Phe Gln Thr Arg Asn Tyr Asp Lys
595                      600                      605
Met Thr Gln Val Asn Gly Leu Tyr Glu Lys Asn Lys Leu Leu Ala Phe
610                      615                      620
Met Val Val Tyr Leu Met Glu Arg Leu Asn Ile Leu Leu Asn Lys Pro
625                      630                      635                      640
Thr Glu Leu Asn Glu Leu Glu Lys Ala Glu Val Asp Phe Lys Ile Ser
645                      650                      655
Asp Lys Val Met Ala Lys Ile Pro Phe Ser Gln Tyr Pro Ser Leu Val
660                      665                      670
Tyr Ala Met Ser Ser Lys Tyr Ala Asp Ser Val Gly Ser Tyr Lys Phe
675                      680                      685
Glu Asn Asp Glu Lys Asn Lys Pro Phe Leu Gly Lys Ile Asp Thr Ile
690                      695                      700
Glu Lys Gln Arg Met Glu Phe Ile Lys Glu Val Leu Gly Phe Glu Glu
705                      710                      715                      720
Tyr Leu Phe Glu Lys Lys Ile Ile Asp Lys Ser Glu Phe Ala Asp Thr
725                      730                      735
Ala Thr His Ile Ser Phe Asp Glu Ile Cys Asn Glu Leu Ile Lys Lys
740                      745                      750
Gly Trp Asp Lys Asp Lys Leu Thr Lys Leu Lys Asp Ala Arg Asn Ala
755                      760                      765
Ala Leu His Gly Glu Ile Pro Ala Glu Thr Ser Phe Arg Glu Ala Lys
770                      775                      780
Pro Leu Ile Asn Gly Leu Lys Lys
785                      790

```

<210> 6

<211> 799

<212> PRT

<213> 宏基因组 (metagenomic)

<400> 6

```

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

```

[0012]

```

Tyr Asn Phe Lys Asp Val Ala Lys Asn Ala Arg Thr Glu Ile Glu Ala
65          70          75          80
Ile Leu Leu Lys Leu Val Glu Leu Arg Asn Phe Tyr Ser His Tyr Val
          85          90          95
His Asn Asp Thr Val Lys Ile Leu Ser Asn Gly Glu Lys Pro Ile Leu
          100         105         110
Glu Lys Tyr Tyr Gln Ile Ala Ile Glu Ala Thr Gly Ser Lys Asn Val
          115         120         125
Lys Leu Val Ile Ile Glu Asn Asn Asn Cys Leu Thr Asp Ser Gly Val
          130         135         140
Leu Phe Leu Leu Cys Met Phe Leu Lys Lys Ser Gln Ala Asn Lys Leu
145          150          155          160
Ile Ser Ser Val Ser Gly Phe Lys Arg Asn Asp Lys Glu Gly Gln Pro
          165          170          175
Arg Arg Asn Leu Phe Thr Tyr Tyr Ser Val Arg Glu Gly Tyr Lys Val
          180         185         190
Val Pro Asp Met Gln Lys His Phe Leu Leu Phe Ala Leu Val Asn His
          195         200         205
Leu Ser Glu Gln Asp Asp His Ile Glu Lys Gln Gln Gln Ser Asp Glu
          210         215         220
Leu Gly Lys Gly Leu Phe Phe His Arg Ile Ala Ser Thr Phe Leu Asn
225          230          235          240
Glu Ser Gly Ile Phe Asn Lys Met Gln Phe Tyr Thr Tyr Gln Ser Asn
          245         250         255
Arg Leu Lys Glu Lys Arg Gly Glu Leu Lys His Glu Lys Asp Thr Phe
          260         265         270
Thr Trp Ile Glu Pro Phe Gln Gly Asn Ser Tyr Phe Thr Leu Asn Gly
          275         280         285
His Lys Gly Val Ile Ser Glu Asp Gln Leu Lys Glu Leu Cys Tyr Thr
          290         295         300
Ile Leu Ile Glu Lys Gln Asn Val Asp Ser Leu Glu Gly Lys Ile Ile
305          310          315          320
Gln Phe Leu Lys Lys Phe Gln Asn Val Ser Ser Lys Gln Gln Val Asp
          325         330         335
Glu Asp Glu Leu Leu Lys Arg Glu Tyr Phe Pro Ala Asn Tyr Phe Gly
          340         345         350
Arg Ala Gly Thr Gly Thr Leu Lys Glu Lys Ile Leu Asn Arg Leu Asp
          355         360         365
Lys Arg Met Asp Pro Thr Ser Lys Val Thr Asp Lys Ala Tyr Asp Lys
          370         375         380
Met Ile Glu Val Met Glu Phe Ile Asn Met Cys Leu Pro Ser Asp Glu
385          390          395          400
Lys Leu Arg Gln Lys Asp Tyr Arg Arg Tyr Leu Lys Met Val Arg Phe
          405         410         415
Trp Asn Lys Glu Lys His Asn Ile Lys Arg Glu Phe Asp Ser Lys Lys
          420         425         430
Trp Thr Arg Phe Leu Pro Thr Glu Leu Trp Asn Lys Arg Asn Leu Glu
          435         440         445
Glu Ala Tyr Gln Leu Ala Arg Lys Glu Asn Lys Lys Lys Leu Glu Asp
          450         455         460
Met Arg Asn Gln Val Arg Ser Leu Lys Glu Asn Asp Leu Glu Lys Tyr

```

	465		470		475		480
	Gln Gln Ile Asn Tyr Val Asn Asp Leu Glu Asn Leu Arg Leu Leu Ser						
		485		490		495	
	Gln Glu Leu Gly Val Lys Trp Gln Glu Lys Asp Trp Val Glu Tyr Ser						
		500		505		510	
	Gly Gln Ile Lys Lys Gln Ile Ser Asp Asn Gln Lys Leu Thr Ile Met						
		515		520		525	
	Lys Gln Arg Ile Thr Ala Glu Leu Lys Lys Met His Gly Ile Glu Asn						
		530		535		540	
	Leu Asn Leu Arg Ile Ser Ile Asp Thr Asn Lys Ser Arg Gln Thr Val						
	545		550		555		560
	Met Asn Arg Ile Ala Leu Pro Lys Gly Phe Val Lys Asn His Ile Gln						
		565		570		575	
	Gln Asn Ser Ser Glu Lys Ile Ser Lys Arg Ile Arg Glu Asp Tyr Cys						
		580		585		590	
	Lys Ile Glu Leu Ser Gly Lys Tyr Glu Glu Leu Ser Arg Gln Phe Phe						
		595		600		605	
	Asp Lys Lys Asn Phe Asp Lys Met Thr Leu Ile Asn Gly Leu Cys Glu						
		610		615		620	
	Lys Asn Lys Leu Ile Ala Phe Met Val Ile Tyr Leu Leu Glu Arg Leu						
	625		630		635		640
	Gly Phe Glu Leu Lys Glu Lys Thr Lys Leu Gly Glu Leu Lys Gln Thr						
		645		650		655	
	Arg Met Thr Tyr Lys Ile Ser Asp Lys Val Lys Glu Asp Ile Pro Leu						
		660		665		670	
[0013]	Ser Tyr Tyr Pro Lys Leu Val Tyr Ala Met Asn Arg Lys Tyr Val Asp						
		675		680		685	
	Asn Ile Asp Ser Tyr Ala Phe Ala Ala Tyr Glu Ser Lys Lys Ala Ile						
		690		695		700	
	Leu Asp Lys Val Asp Ile Ile Glu Lys Gln Arg Met Glu Phe Ile Lys						
		705		710		715	
	Gln Val Leu Cys Phe Glu Glu Tyr Ile Phe Glu Asn Arg Ile Ile Glu						
		725		730		735	
	Lys Ser Lys Phe Asn Asp Glu Glu Thr His Ile Ser Phe Thr Gln Ile						
		740		745		750	
	His Asp Glu Leu Ile Lys Lys Gly Arg Asp Thr Glu Lys Leu Ser Lys						
		755		760		765	
	Leu Lys His Ala Arg Asn Lys Ala Leu His Gly Glu Ile Pro Asp Gly						
		770		775		780	
	Thr Ser Phe Glu Lys Ala Lys Leu Leu Ile Asn Glu Ile Lys Lys						
		785		790		795	
	<210>	7					
	<211>	803					
	<212>	PRT					
	<213>	宏基因组 (metagenomic)					
	<400>	7					
	Met Asn Ala Ile Glu Leu Lys Lys Glu Glu Ala Ala Phe Tyr Phe Asn						
		1		5		10	
	Gln Ala Arg Leu Asn Ile Ser Gly Leu Asp Glu Ile Ile Glu Lys Gln						

	20	25	30
	Leu Pro His Ile Gly Ser Asn Arg	Glu Asn Ala Lys Lys Thr Val Asp	
	35	40	45
	Met Ile Leu Asp Asn Pro Glu Val	Leu Lys Lys Met Glu Asn Tyr Val	
	50	55	60
	Phe Asn Ser Arg Asp Ile Ala Lys Asn Ala Arg	Gly Glu Leu Glu Ala	
	65	70	75
	Leu Leu Leu Lys Leu Val Glu Leu Arg Asn Phe Tyr Ser His Tyr Val		
	85	90	95
	His Lys Asp Asp Val Lys Thr Leu Ser Tyr Gly Glu Lys Pro Leu Leu		
	100	105	110
	Asp Lys Tyr Tyr Glu Ile Ala Ile Glu Ala Thr Gly Ser Lys Asp Val		
	115	120	125
	Arg Leu Glu Ile Ile Asp Asp Lys Asn Lys Leu Thr Asp Ala Gly Val		
	130	135	140
	Leu Phe Leu Leu Cys Met Phe Leu Lys Lys Ser Glu Ala Asn Lys Leu		
	145	150	155
	Ile Ser Ser Ile Arg Gly Phe Lys Arg Asn Asp Lys Glu Gly Gln Pro		
	165	170	175
	Arg Arg Asn Leu Phe Thr Tyr Tyr Ser Val Arg Glu Gly Tyr Lys Val		
	180	185	190
	Val Pro Asp Met Gln Lys His Phe Leu Leu Phe Thr Leu Val Asn His		
	195	200	205
	Leu Ser Asn Gln Asp Glu Tyr Ile Ser Asn Leu Arg Pro Asn Gln Glu		
	210	215	220
[0014]	Ile Gly Gln Gly Gly Phe Phe His Arg Ile Ala Ser Lys Phe Leu Ser		
	225	230	235
	Asp Ser Gly Ile Leu His Ser Met Lys Phe Tyr Thr Tyr Arg Ser Lys		
	245	250	255
	Arg Leu Thr Glu Gln Arg Gly Glu Leu Lys Pro Lys Lys Asp His Phe		
	260	265	270
	Thr Trp Ile Glu Pro Phe Gln Gly Asn Ser Tyr Phe Ser Val Gln Gly		
	275	280	285
	Gln Lys Gly Val Ile Gly Glu Glu Gln Leu Lys Glu Leu Cys Tyr Val		
	290	295	300
	Leu Leu Val Ala Arg Glu Asp Phe Arg Ala Val Glu Gly Lys Val Thr		
	305	310	315
	Gln Phe Leu Lys Lys Phe Gln Asn Ala Asn Asn Val Gln Gln Val Glu		
	325	330	335
	Lys Asp Glu Val Leu Glu Lys Glu Tyr Phe Pro Ala Asn Tyr Phe Glu		
	340	345	350
	Asn Arg Asp Val Gly Arg Val Lys Asp Lys Ile Leu Asn Arg Leu Lys		
	355	360	365
	Lys Ile Thr Glu Ser Tyr Lys Ala Lys Gly Arg Glu Val Lys Ala Tyr		
	370	375	380
	Asp Lys Met Lys Glu Val Met Glu Phe Ile Asn Asn Cys Leu Pro Thr		
	385	390	395
	Asp Glu Asn Leu Lys Leu Lys Asp Tyr Arg Arg Tyr Leu Lys Met Val		
	405	410	415
	Arg Phe Trp Gly Arg Glu Lys Glu Asn Ile Lys Arg Glu Phe Asp Ser		
	420	425	430

[0015]

Lys Lys Trp Glu Arg Phe Leu Pro Arg Glu Leu Trp Gln Lys Arg Asn
 435 440 445
 Leu Glu Asp Ala Tyr Gln Leu Ala Lys Glu Lys Asn Thr Glu Leu Phe
 450 455 460
 Asn Lys Leu Lys Thr Thr Val Glu Arg Met Asn Glu Leu Glu Phe Glu
 465 470 475 480
 Lys Tyr Gln Gln Ile Asn Asp Ala Lys Asp Leu Ala Asn Leu Arg Gln
 485 490 495
 Leu Ala Arg Asp Phe Gly Val Lys Trp Glu Glu Lys Asp Trp Gln Glu
 500 505 510
 Tyr Ser Gly Gln Ile Lys Lys Gln Ile Thr Asp Arg Gln Lys Leu Thr
 515 520 525
 Ile Met Lys Gln Arg Ile Thr Ala Ala Leu Lys Lys Lys Gln Gly Ile
 530 535 540
 Glu Asn Leu Asn Leu Arg Ile Thr Thr Asp Thr Asn Lys Ser Arg Lys
 545 550 555 560
 Val Val Leu Asn Arg Ile Ala Leu Pro Lys Gly Phe Val Arg Lys His
 565 570 575
 Ile Leu Lys Thr Asp Ile Lys Ile Ser Lys Gln Ile Arg Gln Ser Gln
 580 585 590
 Cys Pro Ile Ile Leu Ser Asn Asn Tyr Met Lys Leu Ala Lys Glu Phe
 595 600 605
 Phe Glu Glu Arg Asn Phe Asp Lys Met Thr Gln Ile Asn Gly Leu Phe
 610 615 620
 Glu Lys Asn Val Leu Ile Ala Phe Met Ile Val Tyr Leu Met Glu Gln
 625 630 635 640
 Leu Asn Leu Arg Leu Gly Lys Asn Thr Glu Leu Ser Asn Leu Lys Lys
 645 650 655
 Thr Glu Val Asn Phe Thr Ile Thr Asp Lys Val Thr Glu Lys Val Gln
 660 665 670
 Ile Ser Gln Tyr Pro Ser Leu Val Phe Ala Ile Asn Arg Glu Tyr Val
 675 680 685
 Asp Gly Ile Ser Gly Tyr Lys Leu Pro Pro Lys Lys Pro Lys Glu Pro
 690 695 700
 Pro Tyr Thr Phe Phe Glu Lys Ile Asp Ala Ile Glu Lys Glu Arg Met
 705 710 715 720
 Glu Phe Ile Lys Gln Val Leu Gly Phe Glu Glu His Leu Phe Glu Lys
 725 730 735
 Asn Val Ile Asp Lys Thr Arg Phe Thr Asp Thr Ala Thr His Ile Ser
 740 745 750
 Phe Asn Glu Ile Cys Asp Glu Leu Ile Lys Lys Gly Trp Asp Glu Asn
 755 760 765
 Lys Ile Ile Lys Leu Lys Asp Ala Arg Asn Ala Ala Leu His Gly Lys
 770 775 780
 Ile Pro Glu Asp Thr Ser Phe Asp Glu Ala Lys Val Leu Ile Asn Glu
 785 790 795 800
 Leu Lys Lys

<210> 8

<211> 36

	<212> DNA	
	<213> 宏基因组 (metagenomic)	
	<400> 8	
	gctggagcag ccccgattt gtggggtgat tacagc	36
	<210> 9	
	<211> 36	
	<212> DNA	
	<213> 宏基因组 (metagenomic)	
	<400> 9	
	gctgaagaag cctccgattt gagaggtagt tacagc	36
	<210> 10	
	<211> 36	
	<212> DNA	
	<213> 宏基因组 (metagenomic)	
	<400> 10	
	gctgtgatag acctcgattt gtggggtagt aacagc	36
	<210> 11	
	<211> 36	
	<212> DNA	
[0016]	<213> 宏基因组 (metagenomic)	
	<400> 11	
	gctgtgatag acctcgattt gtggggtagt aacagc	36
	<210> 12	
	<211> 36	
	<212> DNA	
	<213> 宏基因组 (metagenomic)	
	<400> 12	
	gctgtgatag acctcgattt gtggggtagt aacagc	36
	<210> 13	
	<211> 36	
	<212> DNA	
	<213> 宏基因组 (metagenomic)	
	<400> 13	
	gctgtgatgg gcctcaattt gtggggaagt aacagc	36
	<210> 14	
	<211> 36	
	<212> DNA	
	<213> 宏基因组 (metagenomic)	

<400> 14	
gctgtgatag gcctcgattt gtgggtagt aacagc	36
<210> 15	
<211> 2328	
<212> DNA	
<213> 宏基因组 (metagenomic)	
<400> 15	
atggcgcaag tgtcaaagca gacttcgaaa aagagagagt tgtctatcga tgaatatcaa	60
ggtgctcgga aatggtgttt tacgattgcc ttcaacaagg ctcttgtgaa tcgagataag	120
aacgacgggc tttttgtcga gtcgctgtta cgccatgaaa agtattcaaa gcacgactgg	180
tacgatgagg atacacgcgc ttgatcaag tgtagcacac aagcggccaa tgcgaaggcc	240
gaggcgtaa gaaactattt ctcccactat cgacattcgc ccgggtgtct gacatttaca	300
gcagaagatg agttgcggac aatcatggaa agggcgatg agcgggcgat ctttgaatgc	360
aggagacgcg aaactgaagt gatcatcgag ttcccagcc gtctcgaagg cgaccggatc	420
actacggcgg gggttgtgtt ttctgttcg ttctttgttg aacggcgggt gctggatcgt	480
ttgtacggtg cggtaagtgg gcttaagaaa aacgaaggac agtacaagct gactcggaag	540
gcgctttcga tgtattgcct gaaagacagt cgtttcacga aggcgtggga caaacgcgtg	600
ctgcttttca gggatatact cgcgacgctt ggacgcattc ctgcggaggc gtatgaatac	660
taccacggag agcagggcga caagaaaaga gcaaacgaca atgaggggac gaatccgaaa	720
cgccataaag acaagttcat cgagtttgca ctgcattatc tggaggcgca acacagttag	780
atatgcttcg ggccggcgaca cattgtcagg gaggaggccg gggcaggcga cgaacacaaa	840
aagcacagga ccaaaggcaa ggtagttgtc gacttttcaa aaaaagacga agatcagtca	900
tactatatca gtaagaacaa tgttatcgtc aggattgata agaatgccgg gcctcggagt	960
tatcgcatgg ggcttaacga attgaaatac ctgtatttgc ttagcettca gggaaagggc	1020
gacgatgcga ttgcaaaact gtacaggatg cggcagcatg tggagaacat tctggatgta	1080
gtgaaggtca cagataagga taatcacgtc ttctgcccgc gatttgtgct ggagcaacat	1140
gggattggca ggaaagcttt taagcaaaga atagacggca gaggtaaagca tgttcgaggg	1200
gtgtgggaaa agaagaaggc ggccgaccaac gagatgacac ttacagagaa ggccgaggac	1260
attcttcaat acgtaaatga aaattgcacg aggtctttca atcccggcga gtacaaccgg	1320
ctgctgggtg gtctggttgg caaggatgtt gagaatttcc aggcgggact gaaacgcctg	1380
caactggccg agcgaatcga cgggcgggta tattcaattt ttgcgcagac ctccacaata	1440
aacgagatgc atcaggttgt gtgtgatcag attctcaaca gactttgccg aatcggcgat	1500
cagaagctct acgattatgt ggggcttggg aagaaggatg aaatagatta caagcagaag	1560
gttgcattgt tcaaggagca tatttctatc cgcaggggtt tcttgcgcaa gaagttcttg	1620
tatgacagca agaagggtt cgcgaaagct gtggaagagc atttgaaag cgccggcgga	1680
cagagggacg ttgggctgga taaaaagtat tatcatattg atgcgattgg gcgattcgag	1740
ggtgctaate cagccttgta tgaacgctg gcgcgagacc gtttgtgtct gatgatggcg	1800
caatacttcc tggggagtgt acgcaaggaa ttgggtaata aaatttgtgt gtcgaatgat	1860
agcatcgagt tgcccgtgga gggctcagtg ggtaacgaaa aaagcatcgt cttctcagtg	1920
agtgattacg gcaagttata tgtgttggat gacgctgagt ttcttgggcg gatatgtgag	1980
tactttatgc cgcacgaaaa agggaaagata cggtatcata cagtttacga aaaagggttt	2040
agggcataata atgatctgca gaagaaatgt gtcgaggcgg tgcctggcgtt tgaagagaag	2100
gttgtcaaag ccaaaaagat gagcgagaag gaaggggcgc attatattga tttcgtgag	2160
atactggcac aaacaatgtg taaagaggcg gagaagaccg ccgtgaataa ggtgcgtaga	2220
gcgtttttcc atcatcattt aaagtttgtg atagatgaat ttgggttgtt tagtgatgtt	2280
atgaagaaat atggaattga aaaggagtgg aagtttcttg ttaaatga	2328
<210> 16	
<211> 2418	

<212> DNA

<213> 宏基因组 (metagenomic)

<400> 16

[0018]

```

atgaaggttg aaaatattaa agaaaaaagc aaaaaagcaa tgtatttaaat caaccattat      60
gagggaccca aaaaatggtg ttttgcaata gttctgaata gggcatgtga taattacgag      120
gacaatccac acttgttttc caaatcactt ttggattttg aaaaaacaag tcgaaaagat      180
tggtttgacg aagaaacacg agagcttggt gagcaagcag atacagaaat acagccaaat      240
cctaacctga aacctaatc aacagctaac cgaaaactca aagatataag aaactatttt      300
tcgcatcatt atcacaagaa cgaatgcctg tattttaaga acgatgatcc catacgtgc      360
attatggaag cggcgtatga aaaatctaaa atttatatca aaggaaagca gattgagcaa      420
agcgatatac cattgccga attgtttgaa agcagcgggt ggattacacc ggcggggatt      480
ttgttactgg catccttttt tgttgaacga gggattctac atcgcttgat gggaaatate      540
ggaggattta aagataatcg aggcgaatac ggtcttacac acgataatct taccacctat      600
tgtcttaagg gtagttatc aattcgggcg caggatcatg atcgcgtaat gtccagagat      660
attctcggct atctgtcacg agtcccact gagtcatctt agcgtatcaa gcaacctcaa      720
atacgaagaa aaggccaatt aagtgaagaa aagacggaca aatttataac atttgacta      780
aattatcttg aggattatgg gctgaagat ttggaaggct gcaaagcctg ttttgccaga      840
agtaaaattg taagggaaca agaaaatgtt gaaagcataa atgataagga atacaacct      900
cacgagaaca aaaagaaagt tgaattcac ttcgatcaga gcaaagaaga ccgattttat      960
attaatcgca ataacgttat tttgaagatt cagaagaaag atggacattc caacatagtt     1020
aggatgggag tatatgaact taaatatctc gttcttatga gtttagtggg aaaagcaaaa     1080
gaagcagttg aaaaaattga caactatac caggatttgc gagaccagtt gccttacata     1140
gaggggaaaa ataaggaaga gattaagaa tacgtcaggt tctttccacg atttatacgt     1200
tctcacctcg gtttactaca gattaacgat gaagaaaaga taaaagctcg attagattat     1260
gttaagacca agtggttaga taaaaggaa aaatcgaaag agcttgaact tcataaaaaa     1320
ggacgggaca tcctcaggta tatcaacgag cgatgtgata gagagcttaa caggaatgta     1380
tataaccgta ttttagagct cctggtcagc aaagacctca ctggttttta tcgtgagctt     1440
gaagaactaa aaagaacaag gcgtagatag aaaaatattg tccagaatct tcttgggcaa     1500
aaaaccatta atgcactgca tgaagagtc tgtgatctgg tgctgaagga aatcgaaagt     1560
ctcgatacag aaaatctcag gaaatatctt ggattgatac ccaaagaaga aaaagaggtc     1620
actttcaaag aaaaggtcga taggattttg aaacagccag ttattttaca agggtttctg     1680
agataccaat tcttcaaaga tgacaaaaag agttttgtct tacttgttga agacgcattg     1740
aaggaaaaag gaggaggttg tgatgttcct cttgggaaag agtattataa aatcgtgtca     1800
cttgataagt atgataaaga aaataaaacc ctgtgtgaaa ctctggcgat ggataggctt     1860
tgcttatgta tggcaagaca atattatctc agtctgaatg caaaactgc acaggaagct     1920
cagcaaatcg aatggaagaa agaagatagt atagaattga ttattttcac cttaaaaaat     1980
cccgatcaat caaagcagag tttttctata cggttttcgg tcagagattt tacgaagtgt     2040
tatgtaacgg atgatcctga atttctggcc cggctttgtt cctacttttt ccagtttgaa     2100
aaagagattg aatatcaca gctctattca gaagggataa ataaatacac aaacctgcaa     2160
aaagagggaa tcgaagcaat actcgagctt gaaaaaagc ttattgaacg aaatcggatt     2220
caatctgcaa aaaattatct ctcatttaat gagataatga ataaaagcgg ttataataaa     2280
gatgagcagg atgatctaaa gaaggtgcga aattctcttt tgcattataa gcttatcttt     2340
gagaaagaac atctcaagaa gttctatgag gttatgagag gagaagggat agagaaaaag     2400
tggtctttta tagtatga
2418

```

<210> 17

<211> 2373

<212> DNA

<213> 宏基因组 (metagenomic)

<400> 17	
atgaatggca ttgaattaaa aaaagaagaa gcagcatittt attttaatca ggcagagctt	60
aatttaaaag ccatagaaga caatatittt gataaagaaa gacgaaagac tctgcttaat	120
aatccacaga tacttgccaa aatggaaaat ttcattttca atttcagaga tgtaacaaaa	180
aatgcaaaaag gggaaattga ctgcttgctg ttgaaactaa gagagctgag aaacttttac	240
tgcattatg tccacaaacg agatgtaaga gaattaagca agggcgagaa acctatactt	300
gaaaagtatt accaatittg gattgaaatca accggaagtg aaaatgttaa acttgagata	360
atagaaaacg acgcgtggct tgcagatgcc ggtgtgttgt ttttcttatg tatttttttg	420
aagaaatctc aggcaataa gcttataagc ggtatcagcg gttttaaaag aaacgatgat	480
accggtcagc cgagaaggaa tttatttacc tatttcagta taaggaggagg atacaagggt	540
gttccggaag tgcagaaaca tttctttttg ttttctcttg ttaatcatct ctctaataca	600
gatgattata ttgaaaaagc gcatcagcca tacgatatag gcgagggttt attttttcat	660
cgaatagctt ctacatttct taatataagt gggattttta gaaatatgaa attctatacc	720
tatcagagta aaaggttagt agagcagcgg ggagaactca aacgagaaaa ggataatttt	780
gcgtgggaag aaccgtttca aggaaatagt tattttgaaa taaatggtca taaaggagta	840
atcggatgaag atgaattgaa ggaactatgt tatgcatttc tgattggcaa tcaagatgct	900
aataaagtgg aaggcaggat tacacaattt ctagaaaagt ttagaaatgc gaacagtgtg	960
caacaagtta aagatgatga aatgctaaaa ccagagtatt ttcttgcaaa ttattttgct	1020
gaatcaggcg tcggaagaat aaaggataga gtgcttaatc gtttgaataa agcgattaaa	1080
agcaataagg ccaagaaagg agagattata gcatacgata agatgagaga ggttatggcg	1140
ttcataaata atttctctgc ggtagatgaa aaattgaaac caaaagatta caaacgatat	1200
ctgggaatgg ttctgttctg ggacaggga aaagataaca taaagcggga gttcgagaca	1260
aaagaatggt ctaaatatct tccatctaatt ttctggacgg caaaaaacct tgaaagggtc	1320
tatggtctgg caagagagaa aaacgcagaa ttattcaata aactaaaagc ggatgtagaa	1380
aaaaatggacg aacgggaact tgagaagtat cagaagataa atgatgcaaa ggatttgcca	1440
[0019] aatttacgcc ggcttgcaag cgactttggt gtgaagtggg aagaaaaaga ctgggatgag	1500
tattcaggac agataaaaaa acaattaca gacagccaga aactaacaat aatgaagcag	1560
cggataaccg caggactaaa gaaaaagcac ggcatagaaa atcttaacct gagaataact	1620
atcgacatca ataaaagcag aaaggcagtt ttgaacagaa ttgcgattcc gaggggtttt	1680
gtaaaaaggc atatttttagg atggcaagag tctgagaagg tatcgaaaaa gataagagag	1740
gcagaatgcg aaattctgct gtcgaaagaa tacgaagaac tatcgaaaca atttttccaa	1800
agcaaagatt atgacaaaat gacacggata aatggccttt atgaaaaaaa caaacttata	1860
gccctgatgg cagtttatct aatggggcaa ttgagaatcc tgtttaaaga acacacaaaa	1920
cttgacgata ttacgaaaac aactgtggat ttcaaaatat ctgataaggt gacggtaaaa	1980
atcccccttt caaattatcc ttctgctggt tatacaatgt ccagtaagta tgttgataat	2040
ataggaatt atggattttc caacaagat aaagacaagc cgatttttagg taagattgat	2100
gtaatatagaa aacagcgaat ggaatttata aaagaggttc ttggttttga aaaatatctt	2160
tttgatgata aaataataga taaaagcaaa tttgctgata cagcgactca tataagtttt	2220
gcagaaatag ttgaggagct tgttgaaaaa ggatgggaca aagacagact gacaaaactt	2280
aaagatgcaa gaaataaagc cctgcatggt gaaatactga cggaaccag ctttgatgaa	2340
acaaaatcat tgataaacga attaaaaaaa tga	2373
<210> 18	
<211> 2379	
<212> DNA	
<213> 宏基因组 (metagenomic)	
<400> 18	
atgtccccag atttcatcaa attagaaaaa caggaagcag ctttttactt taatcagaca	60
gagcttaatt taaaagccat agaaagcaat atttttagaca aacaacagcg aatgattctg	120
cttaataatc cacggatact tgccaaagta ggaaatttca ttttcaattt cagagatgta	180

	acaaaaaatg caaaaaggaga aatagactgt ctgctatitta aactggaaga gctaagaaac	240
	ttttactcgc attatgttca taccgacaat gtaaaggaat tgagtaacgg agaaaaaccc	300
	ctactggaaa gatatatca aatcgctatt caggcaacca ggagtgagga tgtaagttc	360
	gaattgtttg aaacaagaaa cgagaataag attacggatg ccggtgtatt gtttttctta	420
	tgtatgtttt taaaaaatc acaggcaaac aagcttataa gcggtatcag cggttcaaa	480
	agaaatgac caacaggcca gccgagaaga aacttattta cctatttcag tgcaagagaa	540
	ggatataagg ctttgccctga tatgcagaaa cattttcttc tttttactct ggtaattat	600
	ttgtcgaatc aggatgagta tatcagcgag cttaaacaat atggagagat tggtaagga	660
	gcctttttta atcgaatagc ttcaacattt ttgaatatca gcgggatttc aggaaatacg	720
	aaattctatt cgtatcaaag taaaaggata aaagagcagc gaggcgaact caatagcgaa	780
	aaggacagct ttgaatggat agagcctttc caaggaaaca gctattttga aataaatggg	840
	cataaaggag taatcgcgca agacgaatta aaagaacttt gttatgcatt gttggtgcc	900
	aagcaagata ttaatgccgt tgaaggcaaa attatgcaat tccgaaaaa gtttagaaat	960
	actggcaatt tgcagcaagt taaagatgat gaaatgctgg aaatagaata tttcccgcga	1020
	agttatttta atgaatcaaa aaaagaggac ataaagaaag agattcttgg ccggtggat	1080
	aaaaagattc gctcctgctc tgcaaaggca gaaaaagcct atgataagat gaaagaggtg	1140
	atggagttta taaataattc tctgccgca gaggaataat tgaaacgcaa agattataga	1200
	agatatctaa agatggttcg tttctggagc agagaaaaag gcaatataga gcgggaattt	1260
	agaacaaagg aatgttcaaa atatttttca tctgattttt ggcggaagaa caatcttgaa	1320
	gatgtgtaca aactggcaac acaaaaaaac gctgaactgt tcaaaaatct aaaagcggca	1380
	gcagagaaaa tgggtgaaac ggaatttgaa aagtatcagc agataaacga tgtaaaggat	1440
	ttggcaagtt taaggcggct tacgcaagat tttggtttga agtgggaaga aaaggactgg	1500
	gaggagtatt ccgagcagat aaaaaaacia attacggaca ggcagaaact gacaataatg	1560
	aaacaaaggg ttacggctga actaaagaaa aagcacggca tagaaaatct taatctgaga	1620
[0020]	ataaccatcg acagcaataa aagcagaaag gcggttttga acagaatagc aattccaaga	1680
	ggatttghta aaaaacatat tttagcttgg cagggatctg agaagatata gaaaaatata	1740
	agggaaagcag aatgcaaaat tctgctatcg aaaaaatatg aagagttatc aaggcagttt	1800
	tttgaagccg gtaatttcga taagctgacg cagataaatg gtctttatga aaagaataaa	1860
	cttacagctt ttatgtcagt atatttgatg ggtcggttga atattcagct taataagcac	1920
	acagaacttg gaaatcttaa aaaaacagag gtggatttta agatatctga taagggtgact	1980
	gaaaaaatac cgttttctca gtatccttcg cttgtctatg cgatgtctcg caaatatgtt	2040
	gacaatgttg ataaatataa attttctcat caagataaaa agaagccatt tttaggtaaa	2100
	attgattcaa ttgaaaaaga acgtattgaa ttcataaaag aggttctcga tttgaagag	2160
	tatcttttta aaaataaggt aatagataaa agcaaatttt ccgatacagc gactcatatt	2220
	agctttaagg aaatatgtga tgaaatgggt aaaaaaggat gtaaccgaaa caaactaacc	2280
	gaacttaaca acgcaaggaa cgcagccctg catggtgaaa taccgtcgga gacctctttt	2340
	cgtgaagcaa aaccgttgat aaatgaattg aaaaaatga	2379
<210>	19	
<211>	2379	
<212>	DNA	
<213>	宏基因组 (metagenomic)	
<400>	19	
	atgtccccag atttcatcaa attagaaaaa caagaagcag ctttttactt taatcagaca	60
	gagcttaatt taaaagccat agaaagcaat attttcgaca aacaacagcg agtgattctg	120
	cttaataatc cacagatact tgccaaagta ggagatttta ttttcaattt cagagatgta	180
	acaaaaaacg caaaaaggaga aatagactgt ttgctattga aactaagaga gctgagaaac	240
	ttttactcac actatgtcta taccgatgac gtgaagatat tgagtaacgg cgaaagacct	300
	ctgctggaag aatattatca atttgcgatt gaagcaaccg gaagtgaaga tgtaaaactt	360
	gaaataatag aaagcaacaa ccgacttacg gaagcgggag tgctgttttt cttgtgtatg	420

	tttttgaaaa agtctcagcg aaataagctt ataagcggta tcagcggttt taaaagaaat	480
	gacccgacag gtcagccgag aaggaattta tttacctact tcagtgtgtaag ggagggatac	540
	aaggttgtgc cggatatgca gaaacatttt cttttgtttg ttcttgtcaa tcatctctct	600
	ggtcaggatg atttatattga aaaggcgcaa aagccatacg atataggcga gggtttatit	660
	tttcatcgaa tagcttctac atttcttaat atcagtggga ttttaagaaa tatggaattc	720
	tatatattacc agagcaaaaag actaaaggag cagcaaggag agctcaaacg tgaaaaggat	780
	atttttccat ggatagagcc ttccaggga aatagttatt ttgaaataaa tggtaataaa	840
	ggaataatcg gcgaagatga attgaaagag ctttgttatg cgttgctggt tgcaggaaaa	900
	gatgtcagag ccgtcgaagg taaaataaca caatttttg aaaagttaa aaatgcggac	960
	aatgctcagc aagttgaaaa agatgaaatg ctggacagaa acaattttcc cgccaattat	1020
	ttcgcgaat cgaacatcgg cagcataaag gaaaaaatac ttaatcgtt ggaaaaact	1080
	gatgatagtt ataataagac ggggacaaag attaaaccat acgacatgat gaaagaggta	1140
	atggagttaa taaataattc tttccggca gatgaaaaa tgaaacgcaa agattacaga	1200
	agatatctaa agatggttcg tatctgggac agtgagaaag ataataaaa gcgggagttt	1260
	gaaagcaaaag aatgtgcaaa atatitttca tctgatttct ggatggcaaa aaatcttgaa	1320
	agggtctatg ggttggaag agagaaaaac gccgaattat tcaataagct aaaagcggtt	1380
	gtggagaaaa tggacgagcg ggaatttgag aagtatcggc tgataaatac cgcagaggat	1440
	ttggcaagtt taagacggct tgcgaaagat ttggcctga agtgggaaga aaaggactgg	1500
	caagagtatt ctgggcagat aaaaaaaca atttctgaca ggcagaaact gacaataatg	1560
	aaacaaagga ttacggctga actaaagaaa aagcacggca tagaaaatct caatcttaga	1620
	ataaccatcg acagcaataa aagcagaaag gcagttttga acagaatcgc agttccaaga	1680
	ggttttgtga aagagcatat tttaggatgg caggggtctg agaaggtatc gaaaaagaca	1740
	agagaagcaa agtgcaaaat tctgtctctg aaagaatatg aagaattatc aaagcaattt	1800
	ttccaaacca gaaattacga caagatgacg caggtaaacg gtcttttacga aaagaataaa	1860
	ctcttagcat ttatggctgt ttatcttatg gagcggttga atatcctgct taataagccc	1920
[0021]	acagaactta atgaacttga aaaagcagag gtggatttca agatatctga taaggatgat	1980
	gccaaaatcc cgttttcaca gtatccttcg cttgtgtacg cgatgtccag caaatatgct	2040
	gatagtgtag gcagttataa atttgagaat gatgaaaaa acaagccgtt tttaggcaag	2100
	atcgatacaa tagaaaaaca acgaatggag ttataaaaag aagtccttgg tttgaagag	2160
	tatctttttg aaaagaagat aatagataaa agcgaatttg ccgacacagc gactcatata	2220
	agttttgatg aaatatgtaa tgagcttatt aaaaaaggat gggataaaga caaactaacc	2280
	aaacttaaag atgccaggaa cgcggccctg catggcgaaa taccggcgga gacctctttt	2340
	cgtgaagcaa aaccgttgat aaatggattg aaaaaatga	2379
	<210> 20	
	<211> 2400	
	<212> DNA	
	<213> 宏基因组 (metagenomic)	
	<400> 20	
	atgaacatca ttaaattaaa aaaagaagaa gctgcgtttt attttaatca gacgacctc	60
	aatctttcag ggcttgatga aattattgaa aaacaaattc cgcacataat cagcaacaag	120
	gaaaatgcaa agaaagtgat tgataagatt ttcaataacc gcttattatt aaaaagtgtg	180
	gagaattata tctacaactt taaagatgtg gctaaaaacg caagaactga aattgaggct	240
	atattgttga aattagtaga gctacgtaat ttttactcac attacgttca taatgatacc	300
	gtcaagatac taagtaacgg tgaaaaacct atactggaaa aatattatca aattgctata	360
	gaagcaaccg gaagtaaaaa tgttaaacct gtaatcatag aaaacaacaa ctgtctcacg	420
	gattctggcg tgcgtttttt gctgtgtatg ttcttaaaaa aatcacagge aaacaagctt	480
	ataagttccg ttagtggttt taaaaggaat gataaagaag gacaaccgag aagaaatcta	540
	ttcacttatt atagtgtgag ggaggatatt aaggttgtgc ctgatatgca gaagcatttc	600
	ctctatttcg ctctggctca tcatctatct gagcaggatg atcatattga gaagcagcag	660

	cagtcagacg agctcggtaa gggtttgttt ttccatcgta tagcttcgac ttttttaaac	720
	gagagcggca tcttcaataa aatgcaatth tatacatatc agagcaacag gctaaaagag	780
	aaaagaggag aactcaaaca cgaaaaggat acctttacat ggatagagcc ttttcaaggc	840
	aatagttatt ttacgttaaa tggacataag ggagtgtatta gtgaagatca attgaaggag	900
	ctttgttaca caattttaat tgagaagcaa aacgttgatt ccttggaagg taaaattata	960
	caatttctca aaaaaattca gaatgtcagc agcaagcagc aagttgacga agatgaattg	1020
	cttaaaagag aatatttccc tgcaaattac tttggccggg caggaacagg gaccctaaaa	1080
	gaaaagattc taaaccggct tgataagagg atggatccta catctaaagt gacggataaa	1140
	gcttatgaca aaatgattga agtgaatgaa tttatcaata tgtgccttc gctgtatgag	1200
	aagttgaggc aaaaggatta tagacgatac ttaaagatgg ttcgtttctg gaataaggaa	1260
	aagcataaca ttaagcgcga gtttgacagt aaaaaatgga cgaggttttt gccgacggaa	1320
	ttgtggaata aaagaaatct agaagaagcc tatcaattag cacggaaaga gaacaaaaag	1380
	aaacttgaag atatgagaaa tcaagtacga agccttaaaag aaaatgacct tgaataatat	1440
	cagcagatta attacgttaa tgacctggag aatttaaggc ttctgtcaca ggagttaggt	1500
	gtgaaatggc aggaaaagga ctgggttgaa tattccgggc agataaagaa gcagatatca	1560
	gacaatcaga aacttacaat catgaaacaa aggattaccg ctgaactaaa gaaaatgcac	1620
	ggcatcgaga atcttaatct tagaataagc attgacacga ataaaagcag gcagacggtt	1680
	atgaacagga tagctttgcc caaaggtttt gtgaagaatc atatccagca aaattcgtct	1740
	gagaaaatat cgaaaagaat aagagaggat tattgtaaaa ttgagctatc gggaaaatat	1800
	gaagaacttt caaggcaatt ttttgataaa aagaatttcg ataagatgac actgataaac	1860
	ggcctttgtg aaaagaacaa acttatcgca tttatggtta tctatctttt ggagcggctt	1920
	ggatttgaat taaaggagaa aacaaaatta ggcgagctta aacaaacaag gatgacatat	1980
	aaaatatccg ataaggtaaa agaagatata ccgctttcct attaccccaa gcttgtgtat	2040
	gcaatgaacc gaaaatatgt tgacaatata gatagttatg catttgcggc ttacgaatcc	2100
	aaaaaagcta ttttgataaa agtggatata atagaaaagc aacgtatgga atttatcaaa	2160
[0022]	caagttctct gttttgagga atatatthtc gaaaatagga ttatcgaaaa aagcaaattt	2220
	aatgacgagg agactcatat aagttttaca caatacatg atgagcttat taaaaagga	2280
	cgggacacag aaaaactctc taaactcaaa catgcaagga ataaagcctt gcacggcgag	2340
	attcctgatg ggacttcttt tgaaaaagca aagctattga taaatgaaat caaaaaatga	2400

<210> 21

<211> 2412

<212> DNA

<213> 宏基因组 (metagenomic)

<400> 21

atgaatgcta tcgaactaaa aaaagaggaa gcagcatttt attttaatca ggcaagactc	60
aacatttcag gacttgatga aattattgaa aagcagttac cacatatagg tagtaacagg	120
gagaatgcga aaaaaactgt tgatatgatt ttggataatc ccgaagtctt gaagaagatg	180
gaaaattatg tctttaactc acgagatata gcaaagaacg caagaggtga acttgaagca	240
ttgttgttga aattagtaga actgcgtaat ttttattcac attatgttca taaagatgat	300
gttaagacat tgagttacgg agaaaaacct ttactggata aatattatga aattgcgatt	360
gaagcgaccg gaagtaaaga tgtcagactt gagataatag atgataaaaa taagcttaca	420
gatgccggtg tgcttttttt atttgttatg tttttgaaaa aatcagaggc aaacaaactt	480
atcagttcaa tcaggggcctt taaaagaaac gataaagaag gccagccgag aagaaatcta	540
ttcacttact acagtgtcag agagggatat aaggtttgtc ctgatatgca gaaacatttt	600
cttttattca cactggttaa ccatttgtca aatcaggatg aatacatcag taatcttagg	660
ccgaatcaag aaatcggcca agggggattt ttccatagaa tagcatcaaa atttttgagc	720
gatagcggga ttttcatatg tatgaaattc tacacctacc ggagtaaaag actaacagaa	780
caacgggggg agcttaagcc gaaaaaagat catthttacat ggatagagcc ttttcaggga	840
aacagttatt tttcagtga gggccaaaaa ggagtaattg gtgaagagca attaaaggag	900

ctttgttatg	tattgctggt	tgccagagaa	gatttttaggg	cgttgaggg	caaagttaca	960
caatttctga	aaaagtttca	gaatgctaata	aacgtacagc	aagttgaaaa	agatgaagtg	1020
ctggaaaaag	aataatttcc	tgcaaatat	tttggaaaac	gagacgtagg	cagagtaaa	1080
gataagatac	ttaatcgttt	gaaaaaaac	actgaaagct	ataaagctaa	agggagggag	1140
gttaaacct	atgacaagat	gaaagaggta	atggagtta	tataataattg	cctgccaca	1200
gatgaaaatt	tgaactcaa	agattacaga	agatatctga	aaatggttcg	tttctggggc	1260
agggaaaagg	aaaatataaa	gcgggaattt	gacagtaaaa	aatgggagag	gtttttgcc	1320
agagaactct	ggcagaaaag	aaacctcgaa	gatgcgtatc	aactggcaaa	agagaaaaac	1380
accgagttat	tcaataaatt	gaaaacaact	gttgagagaa	tgaacgaact	ggaattcgaa	1440
aagtatcagc	agataaacga	cgcaaaagat	ttggcaaatt	taaggcaact	ggcgcgggac	1500
ttcggcgtga	agtgggaaga	aaaggactgg	caagagtatt	cggggcagat	aaaaaaacaa	1560
attacagaca	ggcaaaaact	tacaataatg	aaacaaagga	ttactgctgc	attgaagaaa	1620
aagcaaggca	tagaaaatct	taatcttagg	ataacaaccg	acaccaataa	aagcagaaag	1680
gtggtattga	acagaatagc	gctacctaaa	ggttttgtta	ggaagcatat	cttaaaaaca	1740
gatataaaga	tatcaaagca	aataaggcaa	tcacaatgtc	ctattatact	gtcaacaat	1800
tatatgaagc	tggaacaagga	attctttgag	gagagaaatt	ttgataagat	gacgcagata	1860
aacgggctat	ttgagaaaaa	tgtacttata	gcgtttatga	tagtttatct	gatggaacaa	1920
ctgaatcttc	gacttggtta	gaatacggaa	cttagcaatc	ttaaaaaac	ggaggttaat	1980
tttacgataa	ccgacaaggt	aacggaaaaa	gtccagattt	cgcagtatcc	atcgcttggt	2040
ttcgccataa	acagagaata	tgttgatgga	atcagcggtt	ataagttacc	gccccaaaaa	2100
ccgaaagagc	ctccgtatac	tttcttcgag	aaaatagacg	caatagaaaa	agaacgaatg	2160
gaattcataa	aacaggctct	cggtttcgaa	gaacatcttt	ttgagaagaa	tgtaatagac	2220
aaaactcgct	ttactgatac	tgcgactcat	ataagtttta	atgaaatatg	tgatgagctt	2280
ataaaaaaag	gatgggacga	aaacaaaata	ataaaactta	aagatgcgag	gaatgcagca	2340
ttgcatggta	agataccgga	ggatacgtct	tttgatgaag	cgaaagtact	gataaatgaa	2400
[0023]	ttaaaaaat	ga				2412

<210> 22

<211> 2328

<212> DNA

<213> 人工序列(Artificial Sequence)

<220>

<221> 尚未归类的特征

<222> (1).. (2328)

<223> 经人密码子优化的编码序列

<400> 22

atggcccagg	tgagcaagca	gacctccaag	aagagggagc	tgagcatcga	cgagtaccag	60
ggcgcccgga	agtgtgtcct	caccattgcc	ttcaacaagg	ccttggtgaa	ccgggacaag	120
aacgacggcc	tgttcgtgga	aagcctgctg	agacacgaga	agtacagcaa	gcacgactgg	180
tacgacgaag	atacccgggc	cctgatcaag	tgcagcaccc	aggccgcca	cgccaaggct	240
gaagccctgc	ggaactactt	cagtcactac	cgcatagacc	ctggctgcct	gaccttcacc	300
gccgaggacg	aactgcggac	catcatggag	agagcctatg	agcgggccat	cttcgagtgc	360
agaagaagag	agacagaggt	gatcatcgag	tttcccagcc	gttcgaggg	cgaccggtac	420
accaccgccc	gcgtggtgtt	tttcgtgagc	tttttcgtgg	aaagaagagt	gctggatcgg	480
ctgtatggag	ccgtgtccgg	cctgaagaag	aatgagggac	agtacaagct	gacccggaag	540
gccctgagca	tgtactgcct	gaaggacagc	agattcacca	aggcctggga	taagcgggtg	600
ctgctgttca	gagacatcct	ggcccagctg	ggaagaatcc	ccgccgaggc	ctacgagtac	660
taccacggcg	agcagggtga	taagaagaga	gctaacgaca	atgagggcac	aatcccaag	720
cggcacaagg	acaagttcat	cgaatttgca	ctgcactacc	tggaagccca	gcacagcgag	780

	atctgcttcg gcagacgcc aatcgtgcgg gaagaggccg gcgccggcga tgagcacaag	840
	aagcaccgga ccaagggaaa ggtggtggtg gacttcagca agaaggacga ggaccagagc	900
	tactatatct ccaagaacaa cgtgatcgtg cggatcgaca agaacgccgg ccctagaagc	960
	taccggatgg gcctgaacga gctgaagtac ctcgtgctgc tgagcctgca ggggaagggc	1020
	gacgatgcca tcgccaagct gtacagatac agacagcacg tggagaacat cctggatgtg	1080
	gtgaagggtg ccgataagga taaccacgtg ttcctgcccc gcttcgtgct ggagcagcac	1140
	ggcatcggca gaaaggcctt caagcagcgg atcgatggac gggatgaagca cgtgcggggc	1200
	gtgtgggaga agaagaaggc cgccaccaat gaaatgaccc tgcacgagaa ggccagagac	1260
	atcctgcagt acgtgaacga aaactgcacc cggctcctca accctggcga atacaacaga	1320
	ctgctggtgt gcctgggtgg caaggacgtg gagaacttc aggccggcct gaagcggctg	1380
	cagctggccg aaagatcga tggccgggtg tactccatct tcgccagac cagcaccatc	1440
	aatgagatgc accaggtggt gtgcgaccag atcctgaacc ggctgtgcag aatcggcgac	1500
	cagaagctgt acgattacgt gggactgggc aagaaggacg aaatcgacta caagcagaag	1560
	gtggcctggt tcaaggagca catcagcatc cggagaggat tcctgagaaa gaagtcttg	1620
	tacgatagca agaagggtt cgcaaagctg gtggaggaac acctggagtc cggcggcggc	1680
	cagcgcgacg tgggcctgga caagaagtac taccacatcg acgccatcgg cagattcgag	1740
	ggcgccaacc ccgccctgta cgagaccctg gccagagatc ggctgtgcct catgatggcc	1800
	cagtacttcc tgggcagcgt gagaaggaa ctgggcaaca agattgtgtg gagcaacgac	1860
	agcatcgaac tgccctgtga aggctctgtg ggaaatgaga agagcatcgt gttctccgtg	1920
	tctgactacg gcaagctgta cgtgctggac gatgccgaat tcctgggccc gatctgcgaa	1980
	tacttcatgc ccacgaaaa gggcaagatc cggatccaca cagtgtacga aaaggcctt	2040
	agagcatata acgacctgca gaagaagtgc gtggaggccc tgctggcctt cgaagagaag	2100
	gtgtgaagg ccaagaagat gagcgagaag gaaggcggcc actacatcga cttccgggag	2160
	atcctggccc agaccatgtg caaggaggcc gagaagaccg cagtgaacaa ggtgagacgc	2220
	gccttcttcc accaccacct gaagtctgtg attgacgagt tcggcctgtt cagcgacgtg	2280
[0024]	atgaagaagt acggcatcga gaaggaatgg aagttccctg tcaagtaa	2328

<210> 23

<211> 2418

<212> DNA

<213> 人工序列(Artificial Sequence)

<220>

<221> 尚未归类的特征

<222> (1).. (2418)

<223> 经人密码子优化的编码序列

<400> 23

atgaagggtg agaacatcaa gaaaaagtcc aagaaggcta tgtatctgat caaccactat	60
gaaggcccta agaagtgggtg cttegccatc gtgctgaata gggcctgcga caactatgag	120
gataaccccc acctgttcag caagagcctg ctggaatttg aaaagaccag cagaaaggac	180
tggttcgacg aggagaccag ggaactgggtg gagcaggccc acaccgagat ccagcccaac	240
cccaacctga agcctaacac caccgccaac agaaagtga aggacatccg gaactacttc	300
agccaccact accacaagaa tgagtgcctg tacttcaaga acgacgaccc tatccggtgc	360
atcatggagg cagcctacga gaagtccaag atctacatca agggcaagca gattgagcag	420
tccgacatcc cctccctga gctgtttgag tctagcggct ggatcacccc agccggcatc	480
ctgctgctgg ccagcttctt tgtggagaga ggcatctgc acagactgat gggcaacatc	540
ggcggettca aggacaaccg gggcgaatac ggactgaccc acgatatctt caccacctac	600
tgctgaagg gcagctactc catcagagcc caggaccacg acgcctgat gttcagagac	660
atcctgggct acctgagcag agtgccgacc gagagcttcc agcgcatcaa gcagccacag	720

	atcagaaagg aggggcagct gagcgagcgg aagacagaca agtttatcac cttcgccctg	780
	aactacctgg aagattatgg actgaaggat ctggaaggct gcaaggcctg cttcgcccgg	840
	agcaagatcg tgagagagca ggagaacgtg gaaagcatca atgacaagga gtacaagcct	900
	cacgaaaaca agaagaaggt ggaaatccac ttcgatcagt ctaaggaaga ccggttctac	960
	atcaaccgga acaacgtgat cctgaagatc cagaagaagg acggccacag caacatcgtg	1020
	agaatgggcg tgtacgagct gaagtatctg gtgctgatgt ccctgggtgg caaggccaag	1080
	gaagccgtgg agaagatcga caactacatc caggatctga gagaccagct gccttacatc	1140
	gagggcaaga acaaggaaga aatcaaggag tacgtgagat tcttccccag attcatcaga	1200
	tcccacctgg gcctgctgca gattaacgat gaggagaaga tcaaggcccg gctggactat	1260
	gtgaagacaa agtggctgga caagaaggag aagtccaagg agctggagct gcacaagaag	1320
	ggccgggata tcttgcggtg catcaacgag cgggtgcgacc gggagctgaa ccggaacgtg	1380
	tacaaccgga tcttggagct gctggtgagc aaggacctga ccggttctta ccgggagctg	1440
	gaggagctga agcggaccag acggatcgat aagaacattg tgcagaacct gtccggccag	1500
	aagaccatca acgccctgca cgaaaagggtg tgcgatctcg tgctgaagga gatcgagagc	1560
	ctggacaccg agaacctgcg gaagtacctg ggctgatcc ccaaggagga gaagggaagt	1620
	acctttaagg agaaggtgga caggatcctg aagcagccgg tgatctacaa ggcttctctg	1680
	cggtagcagt tcttcaagga cgacaagaag agcttctgtc tgctggtgga agacgccctg	1740
	aaggagaagg gagcgggctg cgacgtgccc ctgggcaagg agtactacaa gatcgtgtcc	1800
	ctggacaagt atgacaagga aaataagacc ctgtgcgaga ccctggcaat ggatagactg	1860
	tgctgatga tggcccggca gtattacctg agcctgaacg ccaagctggc ccaggaggcc	1920
	cagcagatcg aatggaagaa ggaggatagc attgagctga tcatcttcac actgaagaat	1980
	cctgaccagt ccaagcagag cttctccatc cggttcagcg tgcgggactt caccaagctg	2040
	tacgtgaccg acgaccccga attcctggcc cggctgtgca gctacttctt ccccgtaggag	2100
	aaggagatcg aataccacaa gctgtactct gaaggcatta acaagtacac caacctgcag	2160
	aaggagggga tcgaagccat cctggagctg gagaagaagc tgatcgaaag aaaccgcatc	2220
[0025]	cagtcgcca agaactacct gagctttaac gaaatcatga acaagagcgg ctacaacaag	2280
	gatgagcagg atgacctgaa gaaggtgagg aactccctgc tgcactacaa gctgatcttc	2340
	gaaaaggagc acctgaagaa gttctatgaa gtgatgcggg gcgagggaat cgagaagaag	2400
	tggtccctga tcgtgtaa	2418

<210> 24

<211> 2373

<212> DNA

<213> 人工序列(Artificial Sequence)

<220>

<221> 尚未归类的特征

<222> (1).. (2373)

<400> 24

atgaatggca tcgagctgaa gaaggaagaa gccgccttct acttcaatca ggccgagctg	60
aacctgaagg ccattgagga caacatcttc gacaaggaga gacggaagac actgctgaac	120
aacccccaga tcttggccaa gatggagaac tttatcttca atttccggga cgtgaccaag	180
aacgccaagg gcgaaatcga ctgcctgctg ctgaagctga gagagctgcg gaacttttac	240
agccactacg tgcacaagcg ggacgtcaga gaactgagca agggcgagaa gccgatcctg	300
gagaagtact accagttcgc catcgaatcc accggtcttg agaactgaa gctcgaaatc	360
atcgaaaacg acgcctggct ggccgacgcc ggctgtctgt tcttctcttg catcttctctg	420
aagaagagcc aggcacaaca gctgatcagc ggcatcagcg gcttcaagag aaacgacgac	480
accggccagc ctcggagaaa cctgttcacc tacttctcca tccgggaggg ctacaaggtg	540
gtccccgaaa tgcagaagca cttctgtctg ttctctcttg tgaaccacct gagcaaccag	600
gacgattata tcgaaaagcg ccaccagccc tacgacatcg gcgagggcct cttcttccac	660

	cggattgcc	gcaccttcc	gaacatctcc	ggaatcctga	gaaacatgaa	gttctacacc	720
	tatcagagca	agagactggt	ggagcagaga	ggcggagctga	agcgggaaaa	ggacatcttc	780
	gcctgggaag	aaccgtttca	gggcaattcc	tactttgaga	tcaacggcca	caaggcgctg	840
	attggcgaag	acgagctgaa	ggagctgtgc	tacgccttcc	tgatcggcaa	ccaggacgcc	900
	aacaaggtgg	agggccggat	caccagttc	ctggagaagt	tcagaaacgc	caacagcgctg	960
	cagcaggtga	aggacgacga	gatgctgaag	cctgaatatt	tccccgcaa	ctactttgcc	1020
	gagagcggcg	tgggcccggat	caaggaccgg	gtgctgaaca	gactgaacaa	ggccatcaag	1080
	agcaacaagg	ccaagaaggg	cgagatcatc	gcctatgaca	agatgagaga	agtgatggct	1140
	ttcatcaata	actctctgcc	cgtggacgag	aagctgaagc	ccaaggatta	caagagatac	1200
	ctgggcatgg	tgagattctg	ggatagagaa	aaggacaata	tcaagcgcca	gttcgaaacg	1260
	aaggagtgga	gcaagtatct	gccctccaac	ttctggaccg	ccaagaacct	ggagagagtg	1320
	tacggactgg	cccgggaaaa	gaacgcagag	ctgtttaaca	agctgaaggc	cgacgtggag	1380
	aagatggacg	aaagagagct	ggaaaagtat	cagaagatca	acgacgcaa	ggatctggcc	1440
	aacctgcggc	ggctggccag	cgacttcgga	gtgaagtggg	aggagaagga	ttgggacgag	1500
	tactccggcc	agatcaagaa	gcagatcaca	gattcccaga	agctgaccat	catgaagcag	1560
	agaatcacag	ccggcctgaa	gaagaagcac	ggcatcgaaa	acctgaacct	gaggatcacc	1620
	atcgacatca	acaagtccag	aaaggccgtg	ctgaatcgga	tcgccatccc	cagaggattt	1680
	gtgaagcggc	acatcctggg	ctggcaggaa	tccgagaagg	tgagcaagaa	gatcagagaa	1740
	gccgaatgcg	agattctgct	gagcaaggag	tacgaggagc	tgagcaagca	gttctttcag	1800
	agcaaggact	acgacaagat	gacccgcata	aacggcctgt	acgagaagaa	taagctgate	1860
	gccttgatgg	ccgtgtatct	gatggggcag	ctgagaatcc	tgttcaagga	gcacaccaag	1920
	ctggacgaca	tcaccaagac	caccgtggat	ttcaagatca	gcgacaaggt	gaccgtgaag	1980
	atcccccttct	ccaactatcc	ctccctggtg	tacaccatga	gcagcaagta	cgtggacaat	2040
	atcggcaact	acggcttcag	caacaaggac	aaggataagc	ccattctggg	caagatcgac	2100
	gtgatcgaga	agcagcggat	ggagtttatc	aaggaggtgc	tgggattcga	gaagtacctg	2160
[0026]	tttgacgata	agatcatcga	caagagcaag	ttcgccgaca	ccgccaccca	catcagcttt	2220
	gccgaaatcg	tggaagaact	ggtggagaag	ggctgggaca	aggaccggct	gacgaagctg	2280
	aaggatgccc	ggaacaaggc	cctgcacggc	gagatcctga	ccggcaccag	cttcgacgag	2340
	acaaagtccc	tgatcaacga	gctgaagaag	taa			2373

<210> 25

<211> 2379

<212> DNA

<213> 人工序列(Artificial Sequence)

<220>

<221> 尚未归类的特征

<222> (1).. (2379)

<223> 经人密码子优化的编码序列

<400> 25

atgagccctg	atttcatcaa	gctggagaag	caggaagcag	ccttctactt	taaccagacc	60
gagctgaacc	tgaaggccat	cgaatccaat	atcctggata	agcagcagag	aatgacacctg	120
ctgaacaacc	ccagaatcct	ggccaagggtg	ggcaacttca	tcttcaattt	ccgggacgtg	180
accaagaacg	caaagggcga	aatcgactgc	ctgctgttca	agctggagga	actgcggaac	240
ttctacagcc	actacgtgca	caccgataac	gtgaaggaa	tgtccaacgg	agagaagcct	300
ctgctggagc	ggtactacca	gategccatc	caggccacaa	gaagcgagga	cgtgaagtgc	360
gagctgttcg	agaccaggaa	cgagaacaag	atcaccgacg	caggcgtgct	gttcttctctg	420
tgcatgttcc	tgaagaagag	ccaggctaat	aagctgattt	ccggcatcag	cggttcaag	480
cggaacgacc	ccaccggcca	gcccagacgg	aacctcttta	cetacttctc	tgcccgggag	540

	ggctacaagg ccctgcctga catgcagaag cacttccctg tgttcaccct ggtgaactac	600
	ctgagcaacc aggacgagta catctccgag ctgaagcagt acggagagat cggacaggga	660
	gccttcttca acagaatcgc cagcaccttc ctgaacatca gcggcatcag cggcaacacc	720
	aagtcttaca gctaccagag caagagaatc aaggagcagc ggggcgaact gaacagcgaa	780
	aaggacagct tcgagtggat cgagcccttt cagggcaact cttattttga gatcaacggc	840
	cacaagggcg tgatcggcga agacgagctg aaggagctgt gctacgccct gctggtaggc	900
	aagcaggaca tcaatgccgt ggagggaag atcatgcagt tcctgaagaa gttcaggaa	960
	accggcaacc tgcagcaggt gaaggacgac gagatgctgg aaatcgagta cttccccgc	1020
	agctacttca acgagagcaa gaaggaggac atcaagaagg agatcctggg cagactggac	1080
	aagaagatcc ggtcctgcag cgccaaggcc gagaaggcct acgacaagat gaaggaggtg	1140
	atggagttaa tcaataacag cctgcccgcc gaggagaagc tgaagaggaa ggactaccgc	1200
	agatacctga agatggtag attctgtgac agagaaaagg gcaacatcga gagagagttc	1260
	agaaccaagg agtgggtcaa gtacttcagc agcgaacttc ggagaaagaa caatctggag	1320
	gatgtgtaca agctggccac ccagaagaac gccgagctgt tcaagaatct gaaggccgcc	1380
	gccgagaaga tgggcgaaac agaattcgaa aagtaccagc agatcaacga tgtgaaggac	1440
	ctggccagcc tgagacggct gaccaggat ttcggcctga agtgggagga gaaggattgg	1500
	gaggagtaca gcgaacagat caagaagcag atcaccgacc ggcagaagct gacaatcatg	1560
	aagcagcggg tgaccgccga gctgaagaag aagcacggca tcgagaatct gaacctcaga	1620
	attaccatcg attccaacaa gagcagaaag gccgtgctga acagaatcgc cattccccgg	1680
	ggcttcgtga agaagcacat tctgggctgg cagggcagcg aaaagatcag caagaatate	1740
	cgggaggccg agtgcagat cctgctgtcc aagaagtatg aggagctgtc tcggcagttc	1800
	tttgaggctg gcaacttcga caagctgacc cagatcaacg gcctgtacga aaagaataag	1860
	ctgaccgcct tcatgtccgt ctacctgatg ggcagactga acatccagct gaacaagcac	1920
	acggagctgg gaaatctgaa gaagaccgag gtggacttca agatttccga caaggtgaca	1980
	gaaaagatcc ctttctccca gtacctagc ctggtgtacg ctatgagccg gaagtacgtg	2040
[0027]	gacaacgtgg acaagtacaa gttcagccac caggacaaga agaagccctt cctgggcaag	2100
	atcgacagca tcgaaaagga gagaatcgaa ttcataagg aggtgctgga cttcgaagag	2160
	tacctgttta agaacaagg gatcgacaag agcaagtcca gcgataccgc caccatatac	2220
	tctttcaagg aaatctgcga cgagatgggc aagaagggtc gcaaccgcaa caagctgacc	2280
	gagctgaata acgctagaaa cgccgcactg cacggagaaa tccccagcga gaccagcttc	2340
	cgggaggcca agcccctgat caacgaactg aagaagtaa	2379

<210> 26

<211> 2379

<212> DNA

<213> 人工序列(Artificial Sequence)

<220>

<221> 尚未归类的特征

<222> (1)..(2379)

<223> 经人密码子优化的编码序列

<400> 26

atgagccctg acttcatcaa gctggaaaag caggaagccg ctttctactt taatcagacc	60
gagctgaacc tgaaggccat cgagagcaac atcttcgaca agcagcagcg ggtgatcctg	120
ctgaataacc ccagatcct ggccaagggtg ggcgacttca tcttcaactt ccgggacgtg	180
accaagaacg ccaagggaga aatcgactgc ctgctgctga agctgcggga gctgagaaac	240
ttctacagcc actatgtgta caccgacgac gtgaagatcc tgagcaacgg cgagaggccc	300
ctgctggaga agtactacca gtttgccatc gaggccaccg gatctgagaa tgtgaagctg	360
gagatcatcg agagcaacaa ccggtgacc gaagcgggcg tgctgttctt cctgtgcatg	420

	ttcctgaaga agagccaggc caacaagctg atttccggca tctccggatt caagcgcaac	480
	gaccctaccg gacagcctcg gcggaacctg ttcacctact ttagcgtgcg ggagggctac	540
	aaggtggtgc cgcacatgca gaagcacttc ctgctgttcg tgctggtgaa ccacctgtcc	600
	ggccaggatg actatattga gaaggcccag aagccctacg acatcggcga aggcctgttc	660
	ttccacagaa tcgccagcac ctttctcaac atcagcggca tcctgagaaa catggaattc	720
	tacatctacc agagcaagcg gctgaaggag cagcaggag agctgaagag agagaaggac	780
	atcttccctt ggatcgagcc ttccagggc aacagctact ttgagatcaa cggaaacaag	840
	ggcatcatcg gcgaggacga actgaaggaa ctgtgctacg ccctgctggt ggccggcaag	900
	gacgtgagag ccgtggaagg aaagatcacc cagttcttg agaagttaa gaacgccgat	960
	aaccccagc aggtggagaa ggatgaaatg ctggaccgga acaacttccc tgccaattac	1020
	tttgccgaaa gcaacatcgg cagcatcaag gaaaagatcc tgaatagact ggcaagacc	1080
	gacgactcct acaacaagac cggcaccaag atcaagccct acgacatgat gaaggaggtg	1140
	atggagtcca tcaataattc tctgccgcc gatgagaage tgaagcgga ggactaccgg	1200
	agatacctga agatggtccg gatctgggac agcgaagg acaatatcaa gcgggagttt	1260
	gagagcaagg aatggagcaa gtatttcagc agcgacttct ggatggccaa gaacctggaa	1320
	agagtgtacg gcctggccag ggaaaagaac gccgagctgt ttaacaagct gaaggccgtg	1380
	gtggagaaga tggacgagcg ggagttcgaa aagtaccggc tgatcaacag cgccgaagac	1440
	ctggccagcc tgcggagact ggccaaggac ttcggcctga agtgggagga gaaggactgg	1500
	caggagtatt ctggccagat caagaagcag atctccgaca gacagaagct gacaattatg	1560
	aagcagcgga tcacagccga actgaagaag aagcacgga tcgagaacct gaatctgcgg	1620
	atcaccatcg acagcaacaa gtccagaaag gccgtgctga accggatcgc cgtgccccgg	1680
	ggcttcgtga aggaacacat cctgggcttg caaggctctg aaaaggtgag caagaagacc	1740
	agagaagcca agtgcaagat cctgtgagc aaggagtacg aggaactgag caagcagttc	1800
	tttcagacac ggaattacga caagatgacc caggtgaacg gcctgtacga gaagaacaag	1860
	ctgtctgcct tcatggtggt gtacctgatg gagagactga acatcctgct gaacaagccc	1920
[0028]	acagagctga acgaactgga aaaggccgaa gtggacttca agatctccga caaggtgatg	1980
	gccaaagatcc ctttctctca gtacccagc ctggtgatg caatgagctc caagtacgcc	2040
	gacagcgtgg gctcttaca gttcgaaaac gacgagaaga acaagccctt tctgggcaag	2100
	atcgacacaa tcgagaagca gagaatggag ttcatcaagg aggtgctggg cttcgaggaa	2160
	tacctgttcg agaagaagat catcgataag agcgaattcg ccgacaccgc caccacatc	2220
	agcttcgacg agatctgcaa cgagctgac aagaagggtc gggacaagga caagctgacc	2280
	aagctgaagg acgcccggaa cgccgccctg cagggcgaga tccccgccga gaccagcttc	2340
	cgggaggcca agcccctgat taacggcctg aagaagtaa	2379

<210> 27

<211> 2400

<212> DNA

<213> 人工序列(Artificial Sequence)

<220>

<221> 尚未归类的特征

<222> (1)..(2400)

<223> 经人密码子优化的编码序列

<400> 27

atgaacatca tcaagctgaa gaaggaggaa gccgcctttt actttaacca gacaatcctg	60
aatctgagcg gccctggacga gatcatcgag aagcagatcc cccacatcat ctccaataag	120
gaaaacgcca agaaggtgat tgataagatc ttcaataaca gactgtgct gaagagcgtg	180
gaaaactata tctacaactt caaggacgtg gccaagaacg cccggaccga aatcgaagcc	240
atctgtctga agctggtgga gctgagaaac ttctactccc actacgtgca caacgacacc	300

	gtgaagatcc tgtccaatgg cgagaagccc atcctggaaa agtactacca gatcgccatc	360
	gaagccaccg gctctaagaa cgtgaagctg gtcattatcg aaaacaacaa ctgcctgacc	420
	gactccggcg tgcgttccct gctgtgcatg ttcctgaaga agagccaggc caacaagctg	480
	attagcagcg tgagcggctt taagcggaaac gacaaggaag gccagcccag aaggaacctc	540
	tttacttact atagcgtgag ggaaggctac aaggtggtgc cagacatgca gaagcattc	600
	ctgctgttcg ccttgggtcaa ccacctgtcc gagcaggacg accacatcga gaagcagcag	660
	cagagcgacg agctgggcaa gggcctgttc ttccacagaa tcgccagcac attcctgaat	720
	gaaagcggca tcttcaacaa gatgcagttt tacacctacc agagcaatcg gctgaaggag	780
	aagcggggcg agctgaagca cgagaaggac accttcacct ggatcgagcc ttccagggga	840
	aacagctact tcacctgaa cgggcacaag ggcgtgatca gcgaggatca gctgaaggaa	900
	ctgtgtaca caatcctgat cgagaagcag aacgtggaca gcctggaggg caagatcatt	960
	cagttcctga agaagtttca gaacgtgtct agcaagcagc aggtggatga ggacgagctg	1020
	ctgaagcggg aatacttccc cgccaactac ttcgccggg ccggcaccgg caccctgaag	1080
	gagaagatcc tgaaccggct ggacaagcgg atggacccca ccagcaaggt gaccgacaag	1140
	gcctatgaca agatgatcga ggtgatggag ttcatacaaa tgtgcctgcc cagcgacgag	1200
	aagctgcggc agaaggatta ccggagatat ctgaagatgg tcagattctg gaacaaggag	1260
	aagcacaaca tcaagagaga attcgacagc aagaagtga ccagattcct gccaccgag	1320
	ctgtggaata agcggaaacct ggaggaaagg taccagctgg cccggaagga gaacaagaag	1380
	aagctggagg acatgaggaa tcaggtgagg agcctgaagg agaacgacct ggagaagtac	1440
	cagcagatca actatgtgaa cgacctggaa aacctgcggc tgcgtgtcca agagctgggc	1500
	gtgaagtggc aggagaagga ctgggtggaa tacagcggcc agatcaagaa gcagatcagc	1560
	gataaccaga agctgacaat catgaagcag agaatacccg ccgagctgaa gaagatgcac	1620
	ggcatcgaga acctgaacct gagaatcagc atcgacacca acaagtcccg gcagactgtg	1680
	atgaacagaa ttgccctgcc caagggttc gtgaagaacc acattcagca gaacagcagc	1740
	gagaagatca gcaagagaat cagagaggac tactgcaaga tcgagctgtc cgcaagtac	1800
[0029]	gaagagctga gcagacagtt ttctgacaag aagaactttg acaagatgac cctgatcaac	1860
	ggactgtgcg agaagaataa gctcatcgcc ttcattgtga ttacctgct ggagcggctg	1920
	ggcttcgagc tgaaggagaa gaccaagctg ggcgagctga agcagacccg gatgacatat	1980
	aagatcagcg acaaggtgaa ggaggacatc cccctctcct actaccccaa gctgggtgtac	2040
	gccatgaatc ggaagtatgt ggacaacatc gatagctacg ccttcgccgc ctacgagtct	2100
	aagaaggcca tcctggacaa ggtggacatc attgagaagc agagaatgga attcatcaag	2160
	caggtgctgt gcttcgagga atacatcttc gagaacagaa tcatcgagaa gagcaagttc	2220
	aacgatgagg agaccacat cagcttcacc cagatccacg acgaactgat caagaagggc	2280
	agagataccg aaaagctgag caagctgaag cagccagaa acaaggccct gcacggcgag	2340
	atccccgacg ggaccagctt tgagaaggcc aagctgtctg tcaacgaaat caagaagtaa	2400
	<210> 28	
	<211> 2412	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<221> 尚未归类的特征	
	<222> (1).. (2412)	
	<223> 经人密码子优化的编码序列	
	<400> 28	
	atgaacgcca tcgagctgaa gaaggaagag gccgccttct acttcaacca ggccagactg	60
	aacatctctg gcctggacga aatcatcgag aagcaatgc cacacatcgg ctetaacaga	120
	gagaacgcca agaagactgt ggacatgac ctggataacc ccgaggtgct gaagaagatg	180

	gaaaactacg tgttcaactc ccgcgatatt gccagaatg cccggggcga gctggaggcc	240
	ctgctgctga agctggctga gctgagaaac ttctatagcc actacgtgca caaggacgac	300
	gtcaagacac tgagctacgg tgagaagcct ctgctggata agtactacga gatcgccatc	360
	gaagccaccg gatccaagga cgtgcggctg gagatcattg acgacaagaa taagctgacc	420
	gacgccggag tgctgttcct gctgtgcatg ttctgaaga agagcgagge taacaagctg	480
	atttcagca tccggggctt caagaggaac gacaaggagg gccagcctag aagaaacctg	540
	ttcacctact acagcgtgag agagggtat aagggtgggc ccgacatgca gaagcacttt	600
	ctgctgttca ccctggtgaa ccacctgtcc aatcaggacg agtacatctc caacctgcgc	660
	ccaaccagg aaatcgccca gggcggtatt ttccaccgga tcgccagcaa gttcctgagc	720
	gacagcgga tcctgcacag catgaagtcc tacacatata gatccaagcg gctgaccgag	780
	cagcggggag agctgaagcc caagaaggac cactttacat ggatcgagcc ttccagggc	840
	aattctact tcagcgtgca gggccagaag ggcgtgatcg gagaggagca gctcaaggag	900
	ctgtgctacg tgctgctggt ggcccgggag gacttcagag ccgtggaggg caaggtgacc	960
	cagttcctga agaagttcca gaatgccaat aacgtgcagc aggtggagaa ggacgaggtg	1020
	ctggaaaagg agtacttccc cgccaactac ttgagaacc gggacgtggg aagagtcaag	1080
	gacaagatcc tgaacagact gaagaagatc accgagagtt ataaggccaa gggtagagag	1140
	gtgaaggcct acgacaagat gaaggaaagt atggagtcca tcaacaactg cctgcccacc	1200
	gatgaaaacc tgaagctgaa ggactaccgg cggtagctga agatgggtgag attctgggga	1260
	agagagaagg aaaacatcaa gcgggagttc gactccaaga agtgggagcg cttctctccc	1320
	cgggagctgt ggcagaagag aaacctggag gacgcctacc agctcgccaa ggagaagaac	1380
	acagagctgt tcaacaagct gaagaccacc gtggagagaa tgaacgaact ggagtctgag	1440
	aagtaccagc agatcaatga cgccaaggac ctggccaacc tgagacagct ggccagagac	1500
	tttgagtgat agtgggagga aaaggactgg caggaaact ctggacagat caagaagcag	1560
	atcaccgacc ggcagaagct gaccatcatg aagcagcgga tcaccgccgc cctgaagaag	1620
	aagcagggaa tcgaaaacct gaacctgaga atcacaacag atacgaataa gagcaggaag	1680
[0030]	gtgtgtctga accggatcgc actgcccagg ggattcgtca gaaagcacat cctgaagacc	1740
	gacatcaaga tcagcaagca gatccggcag agccagtgcc ctatcatcct gtctaacaac	1800
	tacatgaagc tggccaagga gttctttgaa gagcggaaact tcgataagat gaccagatc	1860
	aatggcctgt tcgagaagaa cgtgtgatc gccttcatga tcgtgtacct gatggagcag	1920
	ctgaacctga gactgggcaa gaacaccgag ctgtccaacc tgaagaagac cgaggtgaac	1980
	tttaccatca ccgacaaggt gaccgagaag gtgcaaatct ccagtagacc cagcctggtg	2040
	ttcgccatta accgggagta cgtggacggc atcagcggt acaagctgcc cccaagaag	2100
	cccaaggaac ctccctacac cttcttcgaa aagatcgacg ccacgaaaa ggagcggatg	2160
	gaattcatca agcaggtgct gggcttcgag gagcaccctc tcgaaaagaa cgtgatcgac	2220
	aagaccgggt ttaccgacac cgccaccac atcagcttca atgagatctg cgatgagctg	2280
	atcaagaagg gctgggacga aaacaagatc atcaagctga aggatgcacg gaacgctgcc	2340
	ctgcacggca agatccctga agatacctcc tttagcgaag ccaaggtgct gatcaacgaa	2400
	ctgaagaagt aa	2412
<210>	29	
<211>	102	
<212>	DNA	
<213>	人工序列(Artificial Sequence)	
<220>		
<221>	尚未归类的特征	
<222>	(1)..(102)	
<223>	gRNA	
<400>	29	

gctggagcag cccccgattt gtgggggtgat tacagcggtc ttcgatattc aagcgtcgga	60
agacctgctg gagcagcccc cgatttgtgg ggtgattaca gc	102

<210> 30
 <211> 711
 <212> DNA
 <213> 人工序列(Artificial Sequence)

<220>
 <221> 尚未归类的特征
 <222> (1).. (711)
 <223> GFP 报告基因

<400> 30	
atggtgagca agggcgagga ggataacatg gccatcatca aggagtcat gcgcttcaag	60
gtgcacatgg agggctccgt gaacggccac gaggttcgaga tcgagggcga gggcgagggc	120
cgcccttacg agggcaccca gaccgccaag ctgaaggatga ccaagggtgg cccctgccc	180
ttcgctggg acatcctgtc cctcagttc atgtacggct ccaaggccta cgtgaagcac	240
cccgcgcaca tccccgacta cttgaagctg tccttccccg agggttcaa gtgggagcgc	300
gtgatgaact tcgaggacgg cggcgtggtg accgtgacct aggactctc cctgcaggac	360
ggcgagttca tctacaaggt gaagctgcgc ggcaccaact tcccctccga cggccccgta	420
atgcagaaga agaccatggg ctgggaggcc tcctccgagc ggatgtacct cgaggacggc	480
gacctgaagg gcgagatcaa gcagaggctg aagctgaagg acggcggcca ctacgacgct	540
gaggtcaaga ccacctataa ggccaagaag cccgtgcagc tgcccggcgc ctacaacgtc	600
aacatcaagt tggacatcac ctcccacaac gaggactaca ccacgtgga acagtacgaa	660
cgcgcgagg gccgcactc caccggcggc atggacgagc tgtacaagta a	711

[0031]

<210> 31
 <211> 720
 <212> DNA
 <213> 人工序列(Artificial Sequence)

<220>
 <221> 尚未归类的特征
 <222> (1).. (720)
 <223> mCherry 报告基因

<400> 31	
atggtgagca agggcgagga gctgttcacc ggggtggtgc ccatcctggt cgagctggac	60
ggcgacgtaa acggccacaa gttagcgtg tccggcgagg gcgagggcga tgccacctac	120
ggcaagctga cctgaagtt catctgcacc accggcaage tgcccgtgcc ctggcccacc	180
ctctgacca cctgacctc cggcgtgcag tgcttcagcc gctaccccga ccacatgaag	240
cagcacgact tcttcaagtc cgccatgccc gaaggtacg tccaggagcg caccatcttc	300
ttcaaggacg acggcaacta caagaccgc gccgaggatga agttcgaggg cgacacctg	360
gtgaaccgca tcgagctgaa gggcatcgac ttcaaggagg acggcaacat cctggggcac	420
aagctggagt acaactataa cagccacaac gtctatatca tggccgacaa gcagaagaac	480
ggcatcaagg tgaacttcaa gatccgccac aacatcgagg acggcagcgt gcagctcgcc	540
gaccactacc agcagaacac ccccatcggc gacggccccg tgctgtgcc cgacaaccac	600
tacctgagca ccagtcgac cctgagcaaa gacccaacg agaagcgcga tcacatggtc	660

	ctgctggagt tcgtgaccgc cgccgggatc actctcggca tggacgagct gtacaagtga	720
	<210> 32	
	<211> 66	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<221> 尚未归类的特征	
	<222> (1).. (66)	
	<223> SgRNA	
	<400> 32	
	gctggagcag ccccgattt gtggggtgat tacagcggtc ttcgatattc aagcgtcgga	60
	agacct	66
	<210> 33	
	<211> 66	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<221> 尚未归类的特征	
	<222> (1).. (66)	
[0032]	<223> SgRNA	
	<400> 33	
	ggtcttcgat attcaagcgt cggaagacct gctggagcag ccccgattt gtggggtgat	60
	tacagc	66
	<210> 34	
	<211> 20	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<221> 尚未归类的特征	
	<222> (1).. (20)	
	<223> SgRNA	
	<400> 34	
	ttggtgccgc gcagcttcac	20
	<210> 35	
	<211> 25	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	

	<220>	
	<221> 尚未归类的	
	<222> (1)..(25)	
	<223> SgRNA	
	<400> 35	
	ttggtgccgc gcagcttcac cttgt	25
	<210> 36	
	<211> 30	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<221> 尚未归类的特征	
	<222> (1)..(30)	
	<223> SgRNA	
	<400> 36	
	ttggtgccgc gcagcttcac cttgtagatg	30
	<210> 37	
	<211> 35	
[0033]	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<221> 尚未归类的特征	
	<222> (1)..(35)	
	<223> SgRNA	
	<400> 37	
	ttggtgccgc gcagcttcac cttgtagatg aactc	35
	<210> 38	
	<211> 40	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<221> 尚未归类的特征	
	<222> (1)..(40)	
	<400> 38	
	ttggtgccgc gcagcttcac cttgtagatg aactcgccgt	40
	<210> 39	
	<211> 45	

	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<221> 尚未归类的特征	
	<222> (1).. (45)	
	<223> SgRNA	
	<400> 39	
	ttgtgtgccgc gcagcttcac cttgtagatg aactcgccgt cctgc	45
	<210> 40	
	<211> 50	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<221> 尚未归类的特征	
	<222> (1).. (50)	
	<223> SgRNA	
	<400> 40	
	ttgtgtgccgc gcagcttcac cttgtagatg aactcgccgt cctgcaggga	50
[0034]	<210> 41	
	<211> 3615	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<221> 尚未归类的特征	
	<222> (1).. (3615)	
	<223> dCas13e.1-ADAR2DD	
	<400> 41	
	atgcccaaga agaagcggaa ggtggcccag gtgagcaagc agacctcaa gaagagggag	60
	ctgagcatcg acgagtacca gggcgcccgg aagtgtgtgt tcaccattgc cttcaacaag	120
	gccctggtga accgggacaa gaacgacggc ctgttcgttg aaagcctgct gagacacgag	180
	aagtacagca agcacgactg gtacgacgaa gatacccggg ccctgatcaa gtgcagcacc	240
	caggccgccca acgccaaggc tgaagccctg gcgaactact tcagtgttta ccggcatagc	300
	cctggctgcc tgaccttcac cgccgaggac gaactgcgga ccatcatgga gagagcctat	360
	gagcgggcca tcttcgagtg cagaagaaga gagacagagg tgatcatcga gtttcccagc	420
	ctgttcgagg gcgaccgatg caccaccgcc ggcgtggtgt ttttcgtgag ctttttcgtg	480
	gaaagaagag tgctggatcg gctgtatgga gccgtgtccg gcctgaagaa gaatgaggga	540
	cagtacaagc tgacccgga ggcctgagc atgtactgcc tgaaggacag cagattcacc	600
	aaggcctggg ataagcgggt gctgctgttc agagacatcc tggcccagct gggaagaatc	660
	cccgccgagg cctacgagta ctaccacggc gagcagggtg ataagaagag agctaacgac	720
	aatgagggca caaatcccaa gcggcacaag gacaagtcca tcgaatttgc actgcactac	780

	ctggaagccc agcacagcga gatctgcttc ggcagacgcc acatcgtgcg ggaagaggcc	840
	ggcgccggcg atgagcacia gaagcaccgg accaaggga aggtggtggt ggacttcagc	900
	aagaaggacg aggaccagag ctactatata tccaagaaca acgtgatcgt gcggatcgac	960
	aagaacgccg gccctagaag ctaccggatg ggcctgaacg agctgaagta cctcgtgctg	1020
	ctgagcctgc aggggaagg cgacgatgcc atcgccaagc tgtacagata cagacagcac	1080
	gtggagaaca tcctggatgt ggtgaagggt accgataagg ataaccacgt gttcctgccc	1140
	cgcttcgtgc tggagcagca cggcatcggc agaaaggcct tcaagcagcg gatcgatgga	1200
	cgggtgaagc acgtgcgggg cgtgtgggag aagaagaagg ccgccaccaa tgaatgacc	1260
	ctgcacgaga aggccagaga catcctgcag tacgtgaacg aaaactgcac ccggtccttc	1320
	aacctggcg aatacaacag actgctggtg tgcctggtgg gcaaggacgt ggagaacttt	1380
	cagcccgcc tgaagcggt cagctggcc gaaaggatcg atggccgggt gtactccatc	1440
	ttcggccaga ccagcaccat caatgagatg caccaggtgg tgtgcgacca gatcctgaac	1500
	cggctgtgca gaatcgcgga ccagaagctg tacgattacg tgggactggg caagaaggac	1560
	gaaatcgact acaagcagaa ggtggcctgg ttcaaggagc acatcagcat ccggagagga	1620
	ttcctgagaa agaagttctg gtacgatagc aagaagggat tcgcaaagct ggtggaggaa	1680
	cacctggagt ccggcggcgg ccagcgcgac gtgggcctgg acaagaagta ctaccacatc	1740
	gacgccatcg gcagattcga gggcgccaac ccgcccctgt acgagaccct ggccagagat	1800
	cggctgtgcc tcatgatgac ccagtacttc ctgggcagcg tgagaaagga actgggcaac	1860
	aagattgtgt ggagcaacga cagcatcgaa ctgectgtgg aaggtctctg gggaaatgag	1920
	aagagcatcg tgttctccgt gtctgactac ggcaagctgt acgtgctgga cgatgccgaa	1980
	ttcttgggcc ggatctcgca atacctcatg cccacgaaa agggcaagat ccggtaccac	2040
	acagtgtacg aaaaggcctt tagagcatac aacgacctgc agaagaagtg cgtggaggcc	2100
	gtgtctggctt tcgaagagaa ggtggtgaag gccagaaga tgagcgagaa ggaaggcgcc	2160
	cactacatcg acttcggga gatcctggcc cagaccatgt gcaaggaggc cgagaagacc	2220
	gcagtgaaca aggtggcggc tgccttcttc gctgcgcacc tgaagttcgt gattgacgag	2280
[0035]	ttcggcctgt tcagcgacgt gatgaagaag tacggcatcg agaaggaatg gaagtccct	2340
	gtcaagccca agaagaagcg gaaggtgggt ggaggcggag gttctggggg aggaggtagt	2400
	ggcgggtggtg gttcaggagg cggcggaagc cagctgcatt taccgcaggt tttagctgac	2460
	gctgtctcac gcctggtcct gggtaagttt ggtgacctga ccgacaactt ctctccccct	2520
	cacgtctgca gaaaagtgtt ggctggagtc gtcattgaca caggcacaga tgttaaagat	2580
	gccaaaggtga taagtgttct tacaggaggc aatgtatta atggtgaata catgagtgat	2640
	cgtggccttg cattaaatga ctgccatgca gaaataatat ctggagatc cttgctcaga	2700
	tttctttata cacaacttga gctttactta aataacaaag atgatcaaaa aagatccatc	2760
	tttcagaaat cagagcgagg ggggtttagg ctgaaggaga atgtccagtt tcattctgtac	2820
	atcagcacct ctccctgtgg agatgccaga atcttctcac cacatgagcc aatcctggaa	2880
	gaaccagcag atagacaccc aaatcgtaaa gcaagaggac agctacggac caaaatagag	2940
	tctggtcagg ggacgattcc agtgcgctcc aatgcgagca tccaaacgtg ggacggggtg	3000
	ctgcaagggg agcggctgct caccatgtcc tgcagtgaca agattgcacg ctggaacgtg	3060
	gtgggcatcc agggatcact gctcagcatt ttcgtggagc ccatttactt ctcgagcate	3120
	atcctgggca gcctttacca cggggaccac ctttcagggt ccatgtacca gcgcatctcc	3180
	aacatagagg acctgccacc tctctacacc ctcaacaagc ctttgctcag tggcatcagc	3240
	aatgcagaag cacggcagcc aggggaaggcc cccaacttca gtgtcaactg gacggtaggc	3300
	gactccgcta ttgaggtcat caacgccacg actgggaagg atgagctggg ccgcgcgtcc	3360
	cgctgtgtga agcacgcgtt gtactgtcgc tggatgctgtg tgcacggcaa ggttccctcc	3420
	cacttactac gctccaagat taccaagccc aacgtgtacc atgagtccaa gctggcggca	3480
	aaggagtacc aggccgcaa ggccgctctg ttcacagcct tcatcaaggc ggggctgggg	3540
	gcctgggttg agaagccac cgagcaggac cagttctcac tcacgtaccc atacgacgta	3600
	ccagattacg cttaa	3615

<210> 42

<211> 711

	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<221> 尚未归类的特征	
	<222> (1).. (711)	
	<223> 突变 mCherry	
	<400> 42	
	atggtgagca agggcgagga ggataacatg gccatcatca aggagtcat gcgcttcaag	60
	gtgcacatgg agggctccgt gaacggccac gagttcgaga tcgagggcga gggcgagggc	120
	cgccctacg agggcaccca gaccgccaag ctgaaggtga ccaagggtgg cccctgccc	180
	ttcgctggg acatcctgtc ccctcagttc atgtacggt ccaaggccta cgtgaagcac	240
	cccgcgaca tccccgacta cttgaagctg tcttccccg agggcttcaa gtaggagcgc	300
	gtgatgaact tcgaggacgg cggcgtgggtg accgtgacct aggactctc cctgcaggac	360
	ggcgagttca tctacaaggt gaagctgcgc ggcaccaact tccctccga cggccccgta	420
	atgcagaaga agaccatggg ctgggaggcc tcttcgagc ggatgtacct cgaggacggc	480
	gccctgaagg gcgagatcaa gcagaggctg aagctgaagg acggcggcca ctacgacgt	540
	gaggtcaaga ccacctacaa ggccaagaag cccgtgcagc tgcccggcgc ctacaacgtc	600
	aacatcaagt tggacatcac ctcccacaac gaggactaca ccatcgtgga acagtacgaa	660
	cgcgccgagg gccgccactc caccggcggc atggacgagc tgtacaagta a	711
	<210> 43	
	<211> 86	
[0036]	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<221> 尚未归类的特征	
	<222> (1).. (86)	
	<223> gRNA	
	<400> 43	
	caagtagtcg gggatgtcgg cggggtgctt cacctaggcc ttgagccgt gctggagcag	60
	ccccgattt gtggggtgat tacagc	86
	<210> 44	
	<211> 86	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<221> 尚未归类的特征	
	<222> (1).. (86)	
	<223> gRNA	
	<400> 44	
	cggggatgtc ggcggggtgc ttcacctagg ccttggagcc gtacatgaac gctggagcag	60

	ccccgattt gtgggtgat tacagc	86
	<210> 45	
	<211> 3489	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<221> 尚未归类的特征	
	<222> (1).. (3489)	
	<223> LwaCas13a	
	<400> 45	
	atgcccaga agaagcggaa ggtgggatcc atgaaagtga ccaaggtcga tggcatcagc	60
	cacaagaagt acatcgaaga gggcaagctc gtgaagtcca ccagcgagga aaaccggacc	120
	agcgagagac tgagcgagct gctgagcatc cggctggaca tctacatcaa gaaccccgac	180
	aacgcctccg aggaagagaa ccgcatcaga agagagaacc tgaagaagtt ctttagcaac	240
	aagtgctgc acctgaagga cagcgtgctg tatctgaaga accgaaaaga aaagaacgcc	300
	gtgcaggaca agaactatag cgaagaggac atcagcgagt acgacctgaa aaacaagaac	360
	agcttctccg tgctgaagaa gatcctgctg aacgaggacg tgaactctga ggaactggaa	420
	atctttcgga aggacgtgga agccaagctg aacaagatca acagcctgaa gtacagcttc	480
	gaagagaaca aggccaacta ccagaagatc aacgagaaca acgtggaaaa agtgggcggc	540
	aagagcaagc ggaacatcat ctacgactac tacagagaga gcgccaagcg caacgactac	600
	atcaacaacg tgcaggaagc cttcgacaag ctgtataaga aagaggatat cgagaaactg	660
[0037]	tttttctga tgcagaacag caagaagcac gagaagtaca agatccgcga gtactatcac	720
	aagatcatcg gccggaagaa cgacaaagag aacttcgcca agattatcta cgaagagatc	780
	cagaacgtga acaacatcaa agagctgatt gagaagatcc ccgacatgtc tgagctgaag	840
	aaaagccagg tgtttctaca gtactacctg gacaaagagg aactgaacga caagaatatt	900
	aagtacgcct tctgccactt cgtggaaatc gagatgtccc agctgctgaa aaactacgtg	960
	tacaagcggc tgagcaacat cagcaacgat aagatcaagc ggatcttcga gtaccagaat	1020
	ctgaaaaagc tgatcgaaaa caaactgctg aacaagctgg acacctacgt gcggaactgc	1080
	ggcaagtaca actactatct gcaagtgggc gagatcgcca cctccgactt tatcgcccg	1140
	aaccggcaga acgaggcctt cctgagaaac atcatcggcg tgtccagcgt ggccacttc	1200
	agcctgagga acatcctgga aaccgagaac gagaacgata tcaccggccg gatgcggggc	1260
	aagaccgtga agaacaacaa gggcgaagag aaatacgtgt ccggcgaggt ggacaagatc	1320
	tacaatgaga acaagcagaa cgaagtgaag gaaaatctga agatgttcta cagctacgac	1380
	ttcaacatgg acaacaagaa cgagatcgag gactttctcg ccaacatcga cgaggccatc	1440
	agcagcatca gacacggcat cgtgcacttc aacctggaac tggaaggcaa ggacatcttc	1500
	gccttcaaga atatcgcccc cagcgagatc tccaagaaga tgtttcagaa cgaaatcaac	1560
	gaaaagaagc tgaagctgaa aatcttcaag cagctgaaca gcgccaacgt gttcaactac	1620
	tacgagaagg atgtgatcat caagtacctg aagaatacca agttcaactt cgtgaacaaa	1680
	aacatccctt tegtgccag cttcaccaag ctgtacaaca agattgagga cctgcggaat	1740
	acctgaagt ttttttgag cgtgcccaag gacaaagaag agaaggacgc ccagatctac	1800
	ctgtgaaga atatctacta cggcgagtgc ctgaacaagt tcgtgaaaaa ctccaaggtg	1860
	ttctttaaga tcaccaatga agtgatcaag attaacaagc agcggaacca gaaaaccggc	1920
	cactacaagt atcagaagtt cgagaacatc gagaaaaccg tgcccgtgga atacctggcc	1980
	atcatccaga gcagagagat gatcaacaac caggacaag aggaaaagaa tacctacatc	2040
	gactttatc agcagatctt cctgaagggc ttcactgact acctgaacaa gaacaatctg	2100
	aagtatatcg agagcaacaa caacaatgac aacaacgaca tcttctccaa gatcaagatc	2160
	aaaaagata acaaaagaaa gtacgacaag atcctgaaga actatgagaa gcacaatcgg	2220

```

aacaagaaaa tccctcacga gatcaatgag ttcgtgcgcg agatcaagct ggggaagatt 2280
ctgaagtaca ccgagaatct gaacatgttt tacctgatcc tgaagctgct gaaccacaaa 2340
gagctgacca acctgaaggg cagcctggaa aagtaccagt ccgccaacaa agaagaaacc 2400
ttcagcgacg agctggaact gatcaacctg ctgaacctgg acaacaacag agtgaccgag 2460
gacttcgagc tggaaagcaa cgagatcggc aagttcctgg acttcaacga aaacaaaatc 2520
aaggaccgga aagagctgaa aaagttcgac accaacaaga tctatttcga cggcgagaac 2580
atcatcaagc accgggcctt ctacaatatc aagaaatacg gcatgctgaa tctgctggaa 2640
aagatcgccg ataaggcaa gtataagatc agcctgaaag aactgaaaga gtacagcaac 2700
aagaagaatg agattgaaaa gaactacacc atgcagcaga acctgcaccg gaagtacgcc 2760
agaccaaga aggacgaaaa gttcaacgac gaggactaca aagagtatga gaaggccatc 2820
ggcaacatcc agaagtacac ccacctgaag aacaaggtgg aattcaatga gctgaacctg 2880
ctgcagggcc tgcctgctgaa gatcctgcac cggtcctgg gctacaccag catctgggag 2940
cgggacctga gattccggct gaaggcgagc ttccccgaga accactacat cgaggaaatt 3000
ttcaatttcg acaactccaa gaatgtgaag taaaaagcg gccagatcgt ggaaaagtat 3060
atcaacttct acaaagaact gtacaaggac aatgtggaaa agcggagcat ctactccgac 3120
aagaaagtga agaaactgaa gcaggaaaaa aaggacctgt acatccggaa ctacattgcc 3180
cacttcaact acatccccc cgccgagatt agcctgctgg aagtgtgga aaacctgcgg 3240
aagctgctgt cctacgaccg gaagctgaag aacgccatca tgaagtccat cgtggacatt 3300
ctgaaagaat acggcttcgt ggccaccttc aagatcggcg ctgacaagaa gatcgaaatc 3360
cagaccctgg aatcagagaa gatcgtgcac ctgaagaatc tgaagaaaaa gaaactgatg 3420
accgaccgga acagcgagga actgtgcgaa ctcgtgaaag tcatgttcga gtacaaggcc 3480
ctggaatga 3489

```

<210> 46

<211> 3312

[0038]

<212> DNA

<213> 人工序列(Artificial Sequence)

<220>

<221> 尚未归类的特征

<222> (1).. (3312)

<223> PspCas13b

<400> 46

```

atgcccaga agaagcgga ggtggtcgac aacatccccg ctctggtgga aaaccagaag 60
aagtactttg gcacctacag cgtgatggcc atgctgaacg ctacagaccg gctggaccac 120
atccagaagg tggccgatat tgaggcgag cagaacgaga acaacgagaa tctgtggttt 180
caccctgtga tgagccacct gtacaacgcc aagaacggct acgacaagca gcccgagaaa 240
accatgttca tcacgagcg gctgcagagc tacttcccat tctgaagat catggccgag 300
aaccagagag agtacagcaa cggcaagtac aagcagaacc gcgtggaagt gaacagcaac 360
gacatcttcg aggtgctgaa gcgcgccttc ggctgtctga agatgtacag ggacctgacc 420
aaccactaca agacctacga ggaaaagctg aacgacggct gcgagttcct gaccagcaca 480
gagcaacctc tgagcggcat gatcaacaac tactacacag tggccctgcg gaacatgaac 540
gagagatacg gctacaagac agaggacctg gccttcatcc aggacaagcg gttcaagttc 600
gtgaaggacg cctacggcaa gaaaaagtcc caagtgaata ccgattctt cctgagcctg 660
caggactaca acggcgacac acagaagaag ctgcacctga gcggagtggg aatcgccctg 720
ctgatctgcc tgttcctgga caagcagtac atcaacatct ttctgagcag gctgcccac 780
ttctccagct acaatgccc gagcgaggaa cggcggatca tcatcagatc cttcggcatc 840
aacagcatca agctgcccga ggaccggatc cacagcgaga agtccaacaa gagcgtggcc 900
atggatatgc tcaacgaagt gaagcgggtc cccgacgagc tgttcacaac actgtctgcc 960

```

	gagaagcagt cccggttcag aatcatcagc gacgaccaca atgaagtgct gatgaagcgg	1020
	agcagcgaca gattcgtgcc tctgctgctg cagtatatcg attacggcaa gctgttcgac	1080
	cacatcaggt tccacgtgaa catgggcaag ctgagatacc tgctgaaggc cgacaagacc	1140
	tgcatcgacg gccagaccag agtcagagtg atcgagcagc ccctgaacgg cttcggcaga	1200
	ctggaagagg ccgagacaat gcggaagcaa gagaacggca ccttcggcaa cagcggcatc	1260
	cggatcagag acttcgagaa catgaagcgg gacgacgcca atcctgccaa ctatccctac	1320
	atcgtggaca cctacacaca ctacatcctg gaaaacaaca aggtcgagat gtttatcaac	1380
	gacaagagg acagcgccc actgctgccc gtgatcgagg atgatagata cgtggtcaag	1440
	acaatcccca gctgccgat gagcaccctg gaaattccag ccatggcctt ccacatgttt	1500
	ctgttcggca gcaagaaaac cgagaagctg atcgtggacg tgcacaaccg gtacaagaga	1560
	ctgttccagg ccatgcagaa agaagaagt accgccgaga atatcgccag cttcgggaatc	1620
	gccgagagcg acctgcctca gaagatcctg gatctgatca gcggcaatgc ccacggcaag	1680
	gatgtggacg ccttcacag actgaccgtg gacgacatgc tgaccgacac cgagcggaga	1740
	atcaagagat tcaaggacga ccggaagtcc attcggagcg ccgacaacaa gatgggaaag	1800
	agaggcttca agcagatctc cacaggcaag ctggccgact tcctggccaa ggacatcgtg	1860
	ctgtttcagc ccagcgtgaa cgatggcgag aacaagatca ccggcctgaa ctaccggatc	1920
	atgcagagcg ccattgccgt gtacgatagc ggcgacgatt acgaggccaa gcagcagttc	1980
	aagctgatgt tcgagaaggc ccggtgatc ggcaagggca caacagagcc tcatccattt	2040
	ctgtacaagg tgttcgcccg cagcatcccc gccaatgccg tcgagttcta cgagcgctac	2100
	ctgatcgagc ggaagttcta cctgaccggc ctgtccaacg agatcaagaa aggcaacaga	2160
	gtggatgtgc ccttcacccg gcgggaccag aacaagtgga aaacacccgc catgaaaacc	2220
	ctgggcagaa tctacagcga ggatctgccc gtggaactgc ccagacagat gttcgacaat	2280
	gagatcaagt cccacctgaa gtccctgcc aagatggaag gcatcgactt caacaatgcc	2340
	aacgtgacct atctgatcgc cgagtacatg aagagagtgc tggacgacga cttccagacc	2400
[0039]	ttctaccagt ggaaccgcaa ctaccggtac atggacatgc ttaagggcga gtacgacaga	2460
	aagggtccc tgcagcactg cttcaccagc gtggaagaga gagaaggcct ctggaaagag	2520
	cgggcctcca gaacagagcg gtacagaaaag caggccagca acaagatccg cagcaaccgg	2580
	cagatgagaa acgccagcag cgaagagatc gagacaatcc tggataagcg gctgagcaac	2640
	agccggaacg agtaccagaa aagcgagaaa gtgatccggc gctacagagt gcaggatgcc	2700
	ctgtgtttc tgctggccaa aaagaccctg accgaactgg ccgatttcga cggcgagagg	2760
	ttcaaaactga aagaaatcat gcccgacgcc gagaagggaa tcctgagcga gatcatgccc	2820
	atgagcttca ccttcgagaa aggcggcaag aagtacacca tcaccagcga gggcatgaag	2880
	ctgaagaact acggcgactt ctttgtgctg gctagcgaca agaggatcgg caacctgctg	2940
	gaactcgtgg gcagcgacat cgtgtccaaa gaggatatca tggaaagatt caacaaatac	3000
	gaccagtga gggcgagat cagctccatc gtgttcaacc tggaaaagtg ggccttcgac	3060
	acataccccg agctgtctgc cagagtggac cgggaagaga aggtggactt caagagcatc	3120
	ctgaaaatcc tgctgaacaa caagaacatc aacaaagagc agagcgacat cctcggaag	3180
	atccggaacg ccttcgatca caacaattac cccgacaaag gcgtggtgga aatcaaggcc	3240
	ctgcctgaga tcgcatgag catcaagaag gcctttgggg agtacgccat catgaaggga	3300
	tccttcaat ga	3312

<210> 47

<211> 2934

<212> DNA

<213> 人工序列(Artificial Sequence)

<220>

<221> 尚未归类的特征

<222> (1).. (2934)

<223> RxCas13d

[0040]	<400> 47					
	atgcctaaaa	agaaaagaaa	ggtgggttct	ggtatcgaga	agaagaagag	cttcgccaaag 60
	ggcatgggag	tgaagagcac	cctgggtgtcc	ggctctaagg	tgtacatgac	cacatttgct 120
	gagggaagcg	acgccagget	ggagaagatc	gtggagggcg	atagcatcag	atccgtgaac 180
	gagggagagg	ctttcagcgc	cgagatggct	gacaagaacg	ctggctacaa	gatcggaaac 240
	gccaaagttt	cccacccaaa	gggctacgcc	gtgggtggcta	acaacccact	gtacaccgga 300
	ccagtgcagc	aggacatgct	gggactgaag	gagacactgg	agaagaggta	cttcggcgag 360
	tccgccgacg	gaaacgataa	catctgcatc	caggtcatcc	acaacatcct	ggatatcgag 420
	aagatcctgg	ctgagtacat	cacaaacgcc	gcttacgccg	tgaacaacat	ctccggcctg 480
	gacaaggata	tcatcggtt	cggaaagttt	tctaccgtgt	acacatacga	cgagttcaag 540
	gatccagagc	accaccgggc	cgcttttaac	aacaacgaca	agctgatcaa	cgccatcaag 600
	gctcagtacg	acgagttcga	taactttctg	gataacccca	ggctgggcta	cttcggacag 660
	gctttctttt	ctaaggaggg	cagaaactac	atcatcaact	acggaaacga	gtgttacgac 720
	atcctggccc	tgctgagcgg	actgaggcac	tgggtgggtg	acaacaacga	ggaggagtct 780
	cggatcagcc	gcacctggct	gtacaacctg	gacaagaacc	tggataacga	gtacatctcc 840
	acactgaact	acctgtacga	caggatcacc	aacgagctga	caaacagctt	ctccaagaac 900
	tctgccgcta	acgtgaacta	catcgctgag	acctggggca	tcaaccagcg	tgagttcgct 960
	gagcagtact	tcagattttc	catcatgaag	gagcagaaga	acctgggctt	caacatcaca 1020
	aagctgagag	aagtgatgct	ggacagaaag	gatatgtccg	agatcaggaa	gaaccacaag 1080
	gtgttcgatt	ctatcagaac	caaggtgtac	acaatgatgg	acttttgtat	ctacaggtac 1140
	tacatcgagg	aggatgccaa	ggtggccgct	gccaaacaaga	gcctgcccga	caacgagaag 1200
	tctctgagcg	agaaggatat	cttcgtgata	aacctgagag	gctcctttaa	cgacgatcag 1260
	aaggacgctc	tgtactacga	tgaggccaac	aggatctgga	gaaagctgga	gaacatcatg 1320
	cacaacatca	aggagttccg	gggaaacaag	acccgcgagt	acaagaagaa	ggacgctcca 1380
	aggctgccta	ggatcctgcc	tgctggaagg	gacgtgagcg	ccttcagcaa	gctgatgtac 1440
	gccttgacaa	tgtttctgga	cggaaaggag	atcaacgata	tgtgaccac	actgatcaac 1500
	aagttcgaca	acatccagtc	ttttctgaaa	gtgatgcctc	tgatcggcgt	gaacgctaag 1560
	ttctgtggagg	agtacgcctt	ctttaaggac	agcgccaaga	tcgctgatga	gctcgggctg 1620
	atcaagtcct	ttgccaggat	gggagagcca	atcgctgacg	ctaggagagc	tatgtacatc 1680
	gatgccatcc	ggatcctggg	aaccaacctg	tcttacgacg	agctgaaggc	tctggccgac 1740
	acctcagcc	tggatgagaa	cggcaacaag	ctgaagaagg	gcaagcacgg	aatgcgcaac 1800
	ttcatcatca	acaacgtgat	cagcaacaag	cggtttact	acctgatcag	atacggcgac 1860
	ccagctcacc	tgcacgagat	cgctaagaac	gaggccgtgg	tgaagttcgt	gctgggacgg 1920
	atcgccgata	tccagaagaa	gcagggccag	aacggaaaga	accagatcga	ccgtactac 1980
	gagacctgca	tcggcaagga	taagggaag	tccgtgtctg	agaagggtga	cgctctgacc 2040
	aagatcatca	caggcatgaa	ctacgaccag	ttcgataaga	agagatctgt	gatcagggac 2100
	accggaagg	agaacgccga	gagagagaag	tttaagaaga	tcatcagcct	gtacctgaca 2160
	gtgatctacc	acatcctgaa	gaacatcgtg	aacatcaacg	ctagatacgt	gatcggcttc 2220
	cactgcgtgg	agcgcgatgc	ccagctgtac	aaggagaagg	gatacgacat	caacctgaag 2280
	aagctggagg	agaagggtct	tagctccgtg	accaagctgt	gcgctggaat	cgacgagaca 2340
	gcccccgaca	agaggaagga	tgtggagaag	gagatggccg	agagagctaa	ggagagcatc 2400
	gactccctgg	agtctgctaa	ccctaagctg	tacgccaact	acatcaagta	ctccgatgag 2460
	aagaaggccg	aggagtccac	caggcagatc	aacagagaga	aggccaagac	cgctctgaac 2520
	gcctacctga	ggaacacaaa	gtggaacgtg	atcatccggg	aggacctgct	gcgcatcgat 2580
	aacaagacct	gtacactgtt	ccggaacaag	gctgtgcacc	tggagggtgg	tcgctacgtg 2640
	cacgcctaca	tcaacgacat	cggcgagggt	aactcctact	ttcagctgta	ccactacatc 2700
	atgcagagga	tcatcatgaa	cgagagatac	gagaagtcta	gcggcaagg	gtctgagtac 2760
	ttcgacgccg	tgaacgatga	gaagaagtac	aacgatagac	tgtgaagct	gctgtgcgtg 2820
	cctttcggt	actgtatccc	acggtttaag	aacctgagca	tcgaggccct	gttcgaccgc 2880
	aacgaggctg	ccaagtttga	taaggagaag	aagaagggtga	gcggcaactc	ctga 2934

	<210> 48	
	<211> 30	
	<212> DNA	
	<213> 智人(Homo sapiens)	
	<400> 48	
	atggcccttc gcagctcttg cacgtcatac	30
	<210> 49	
	<211> 30	
	<212> DNA	
[0041]	<213> 智人(Homo sapiens)	
	<400> 49	
	ttaggcagcc ctcattcagtg ccggtccct	30
	<210> 50	
	<211> 30	
	<212> DNA	
	<213> 智人(Homo sapiens)	
	<400> 50	
	ggccaggatc tcaattaggc agccctcatc	30

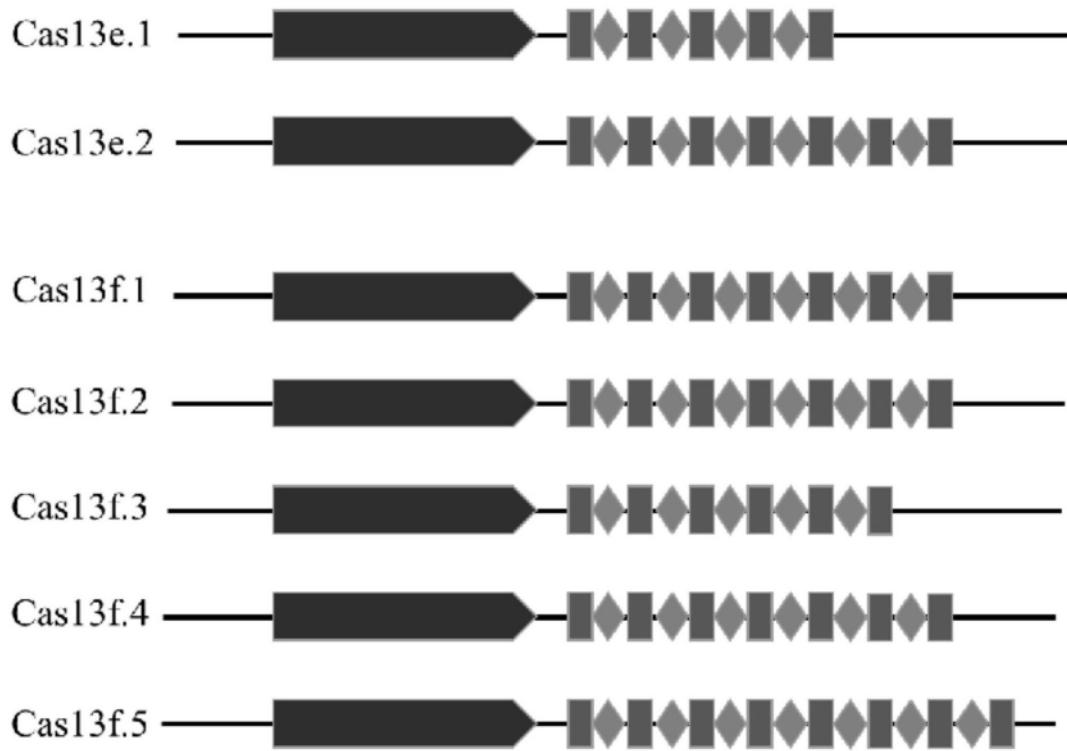


图1

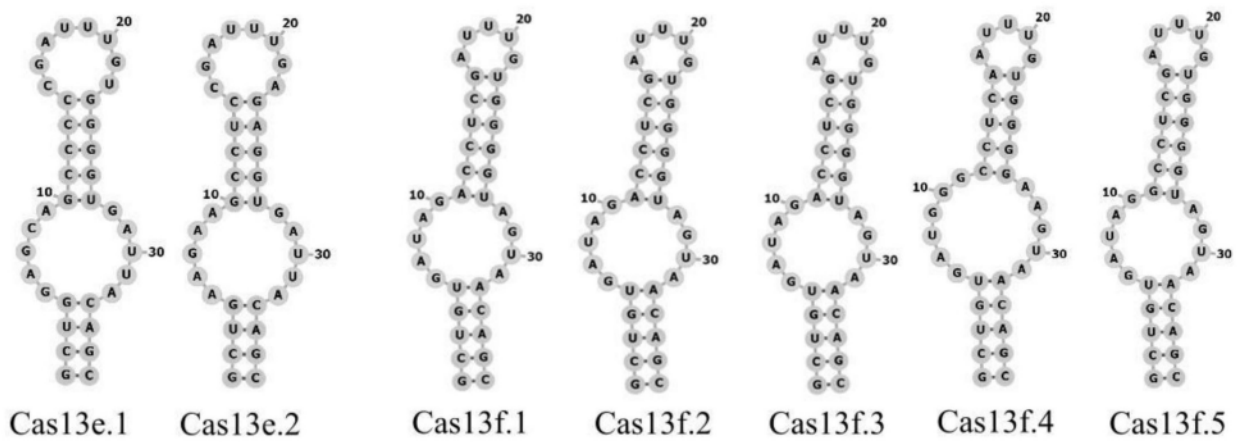


图2

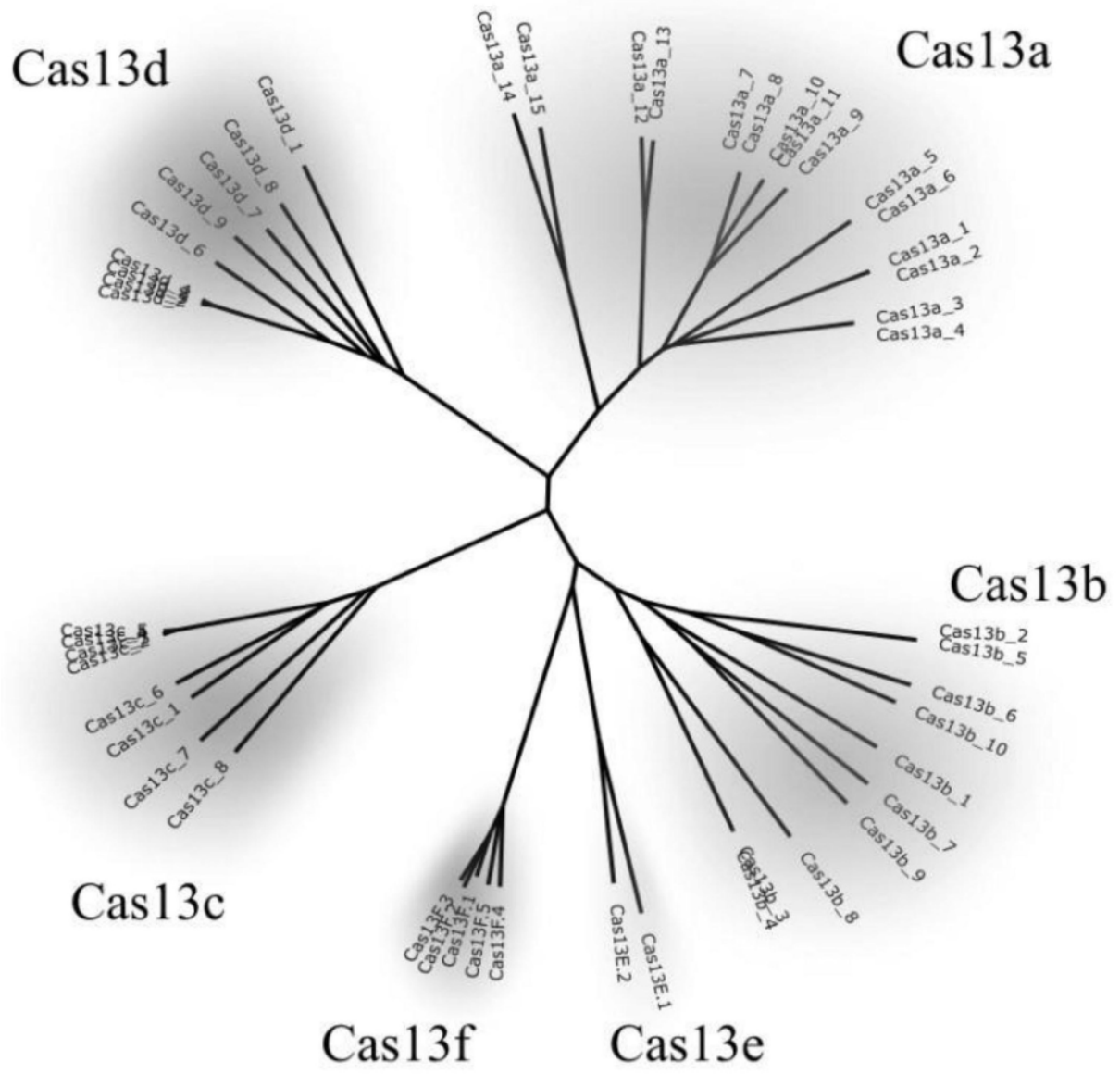


图3

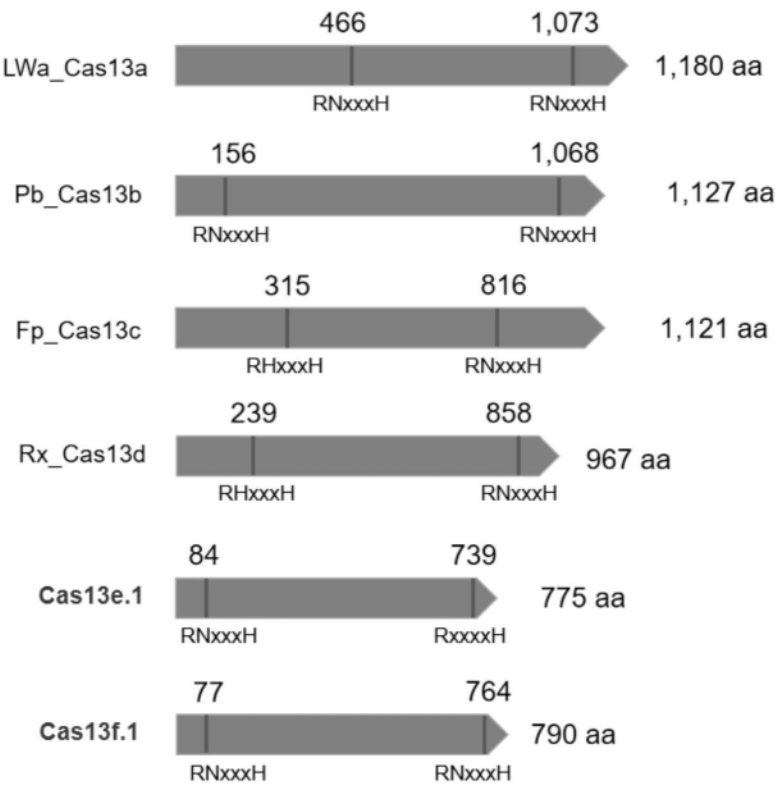


图4

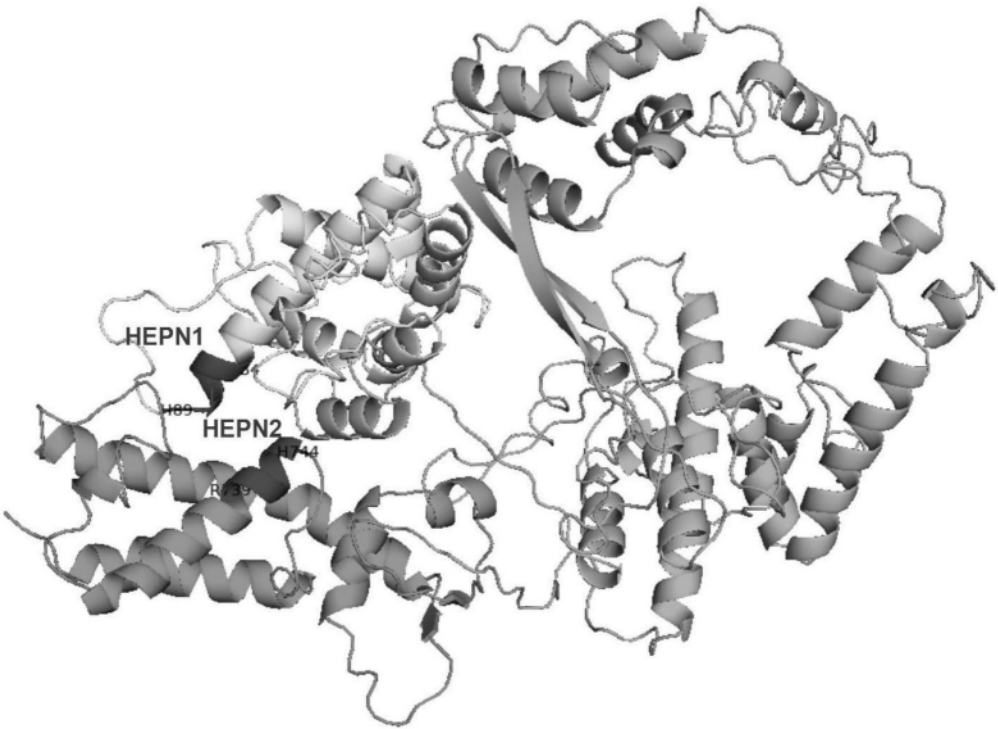


图5

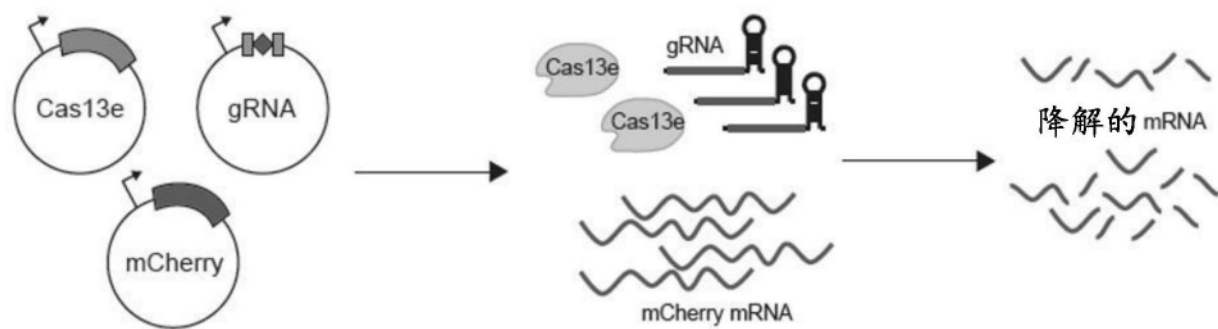


图6

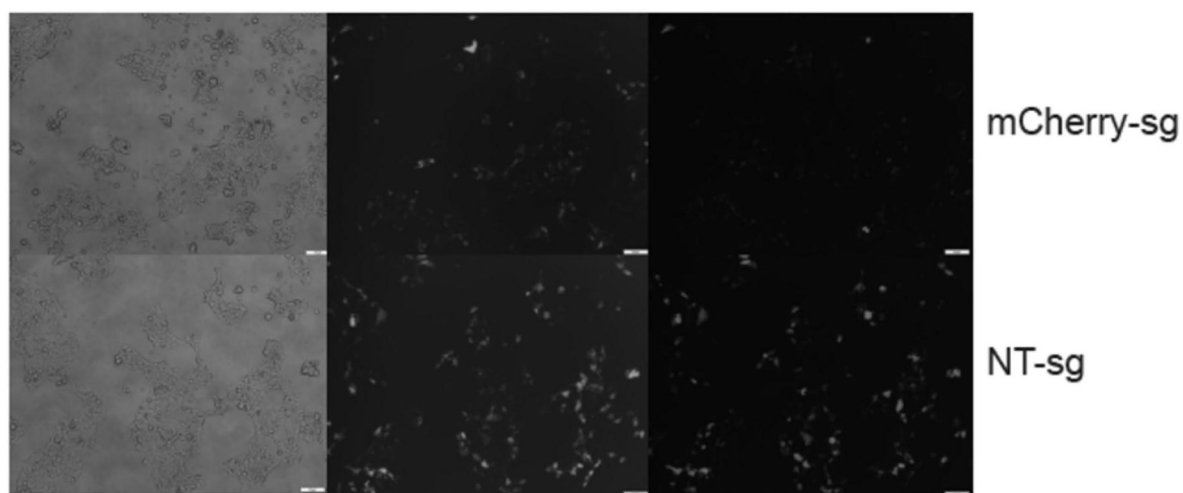


图7

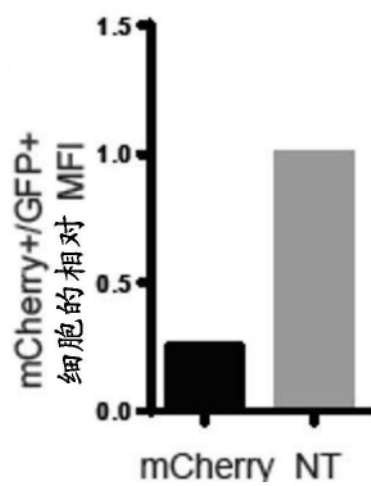


图8

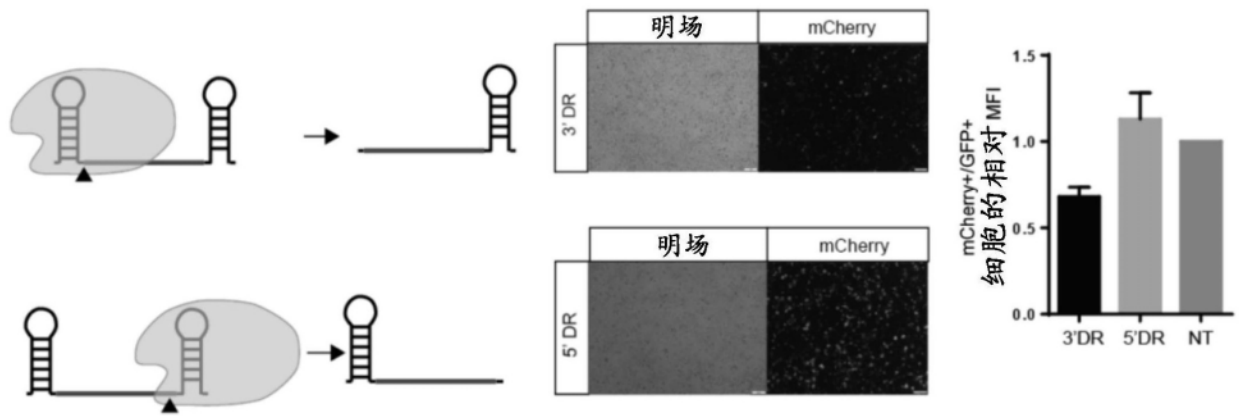


图9

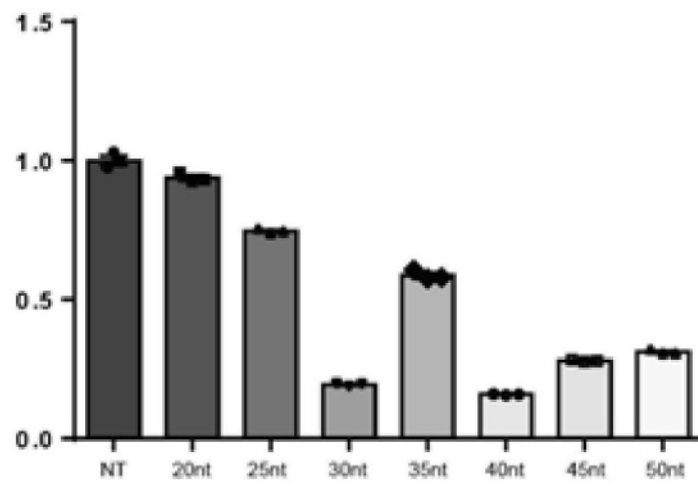


图10

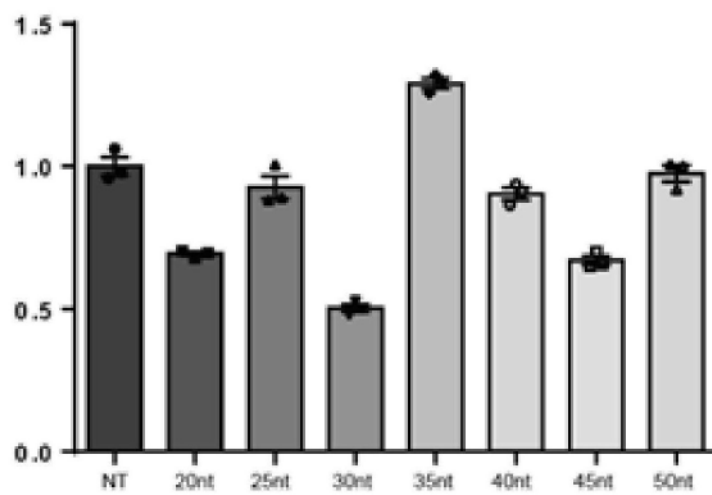


图11

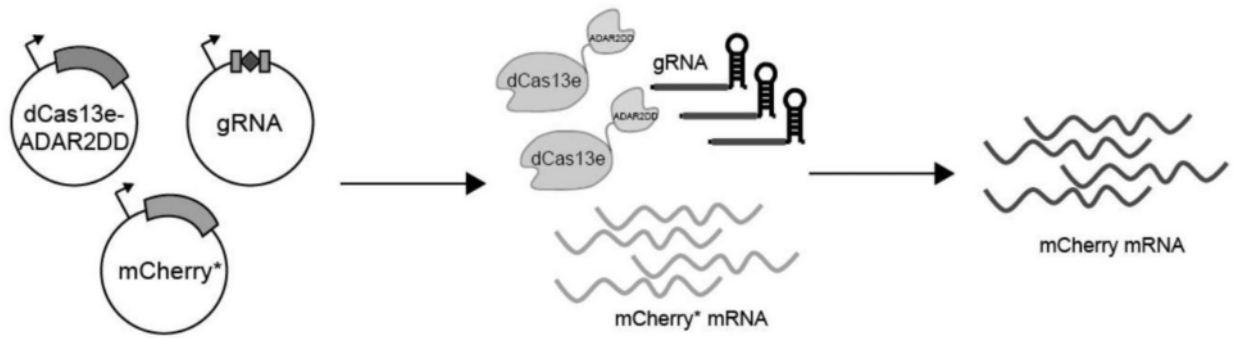


图12

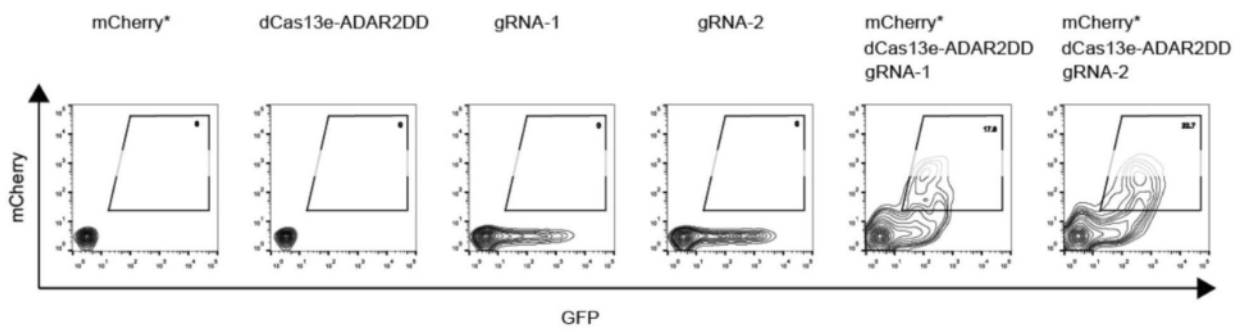


图13

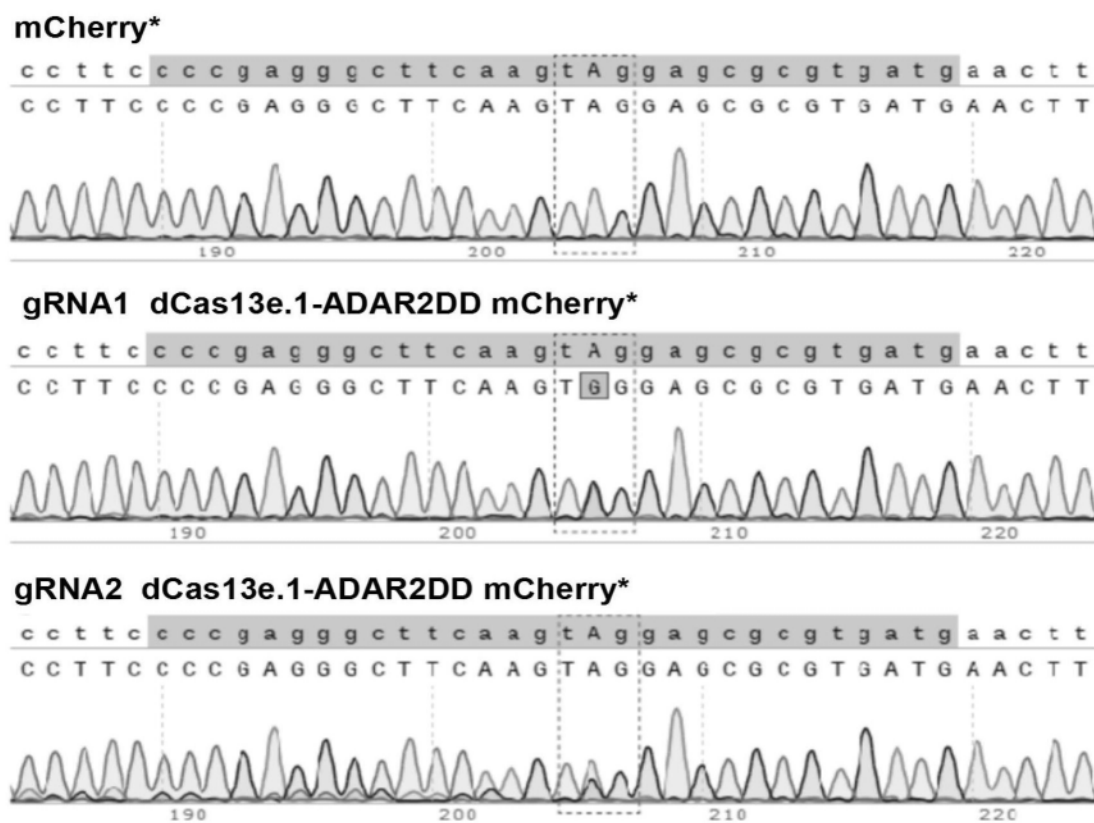


图14

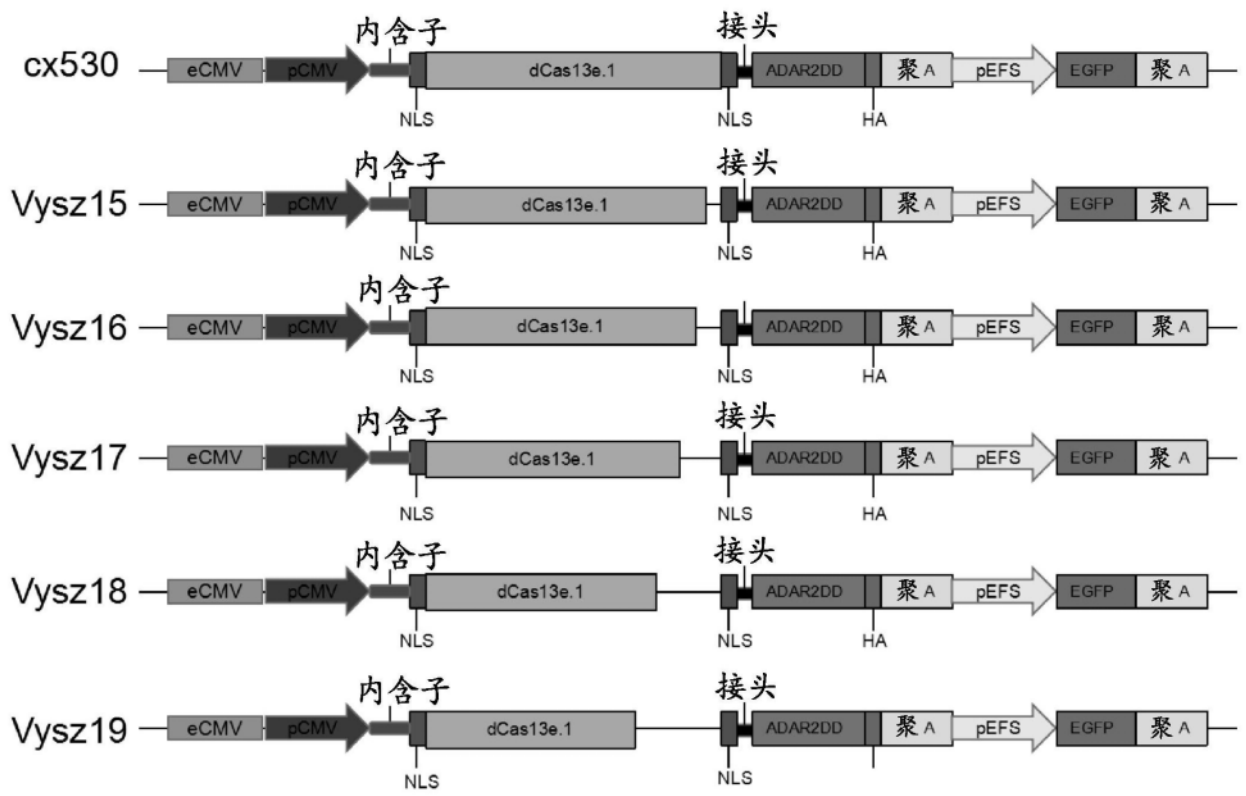


图15

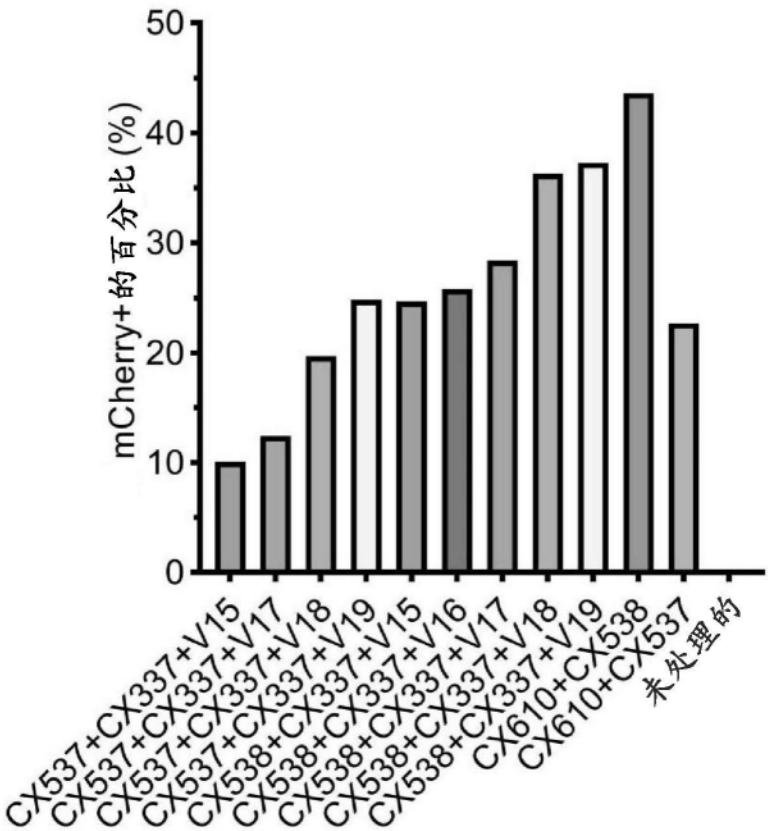


图16

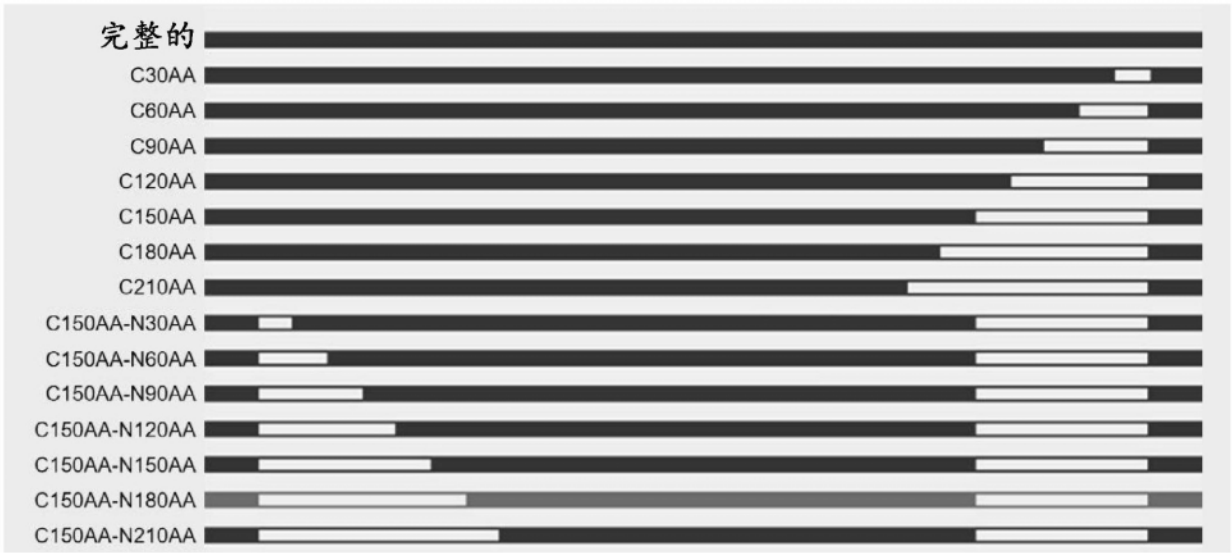


图17

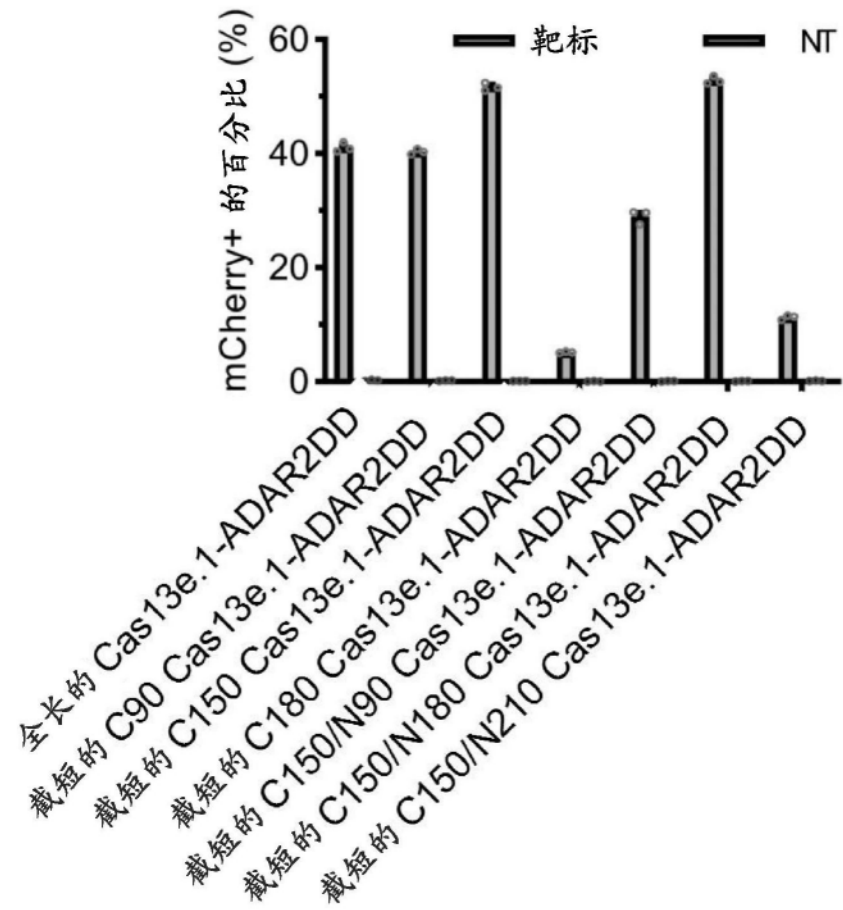


图18

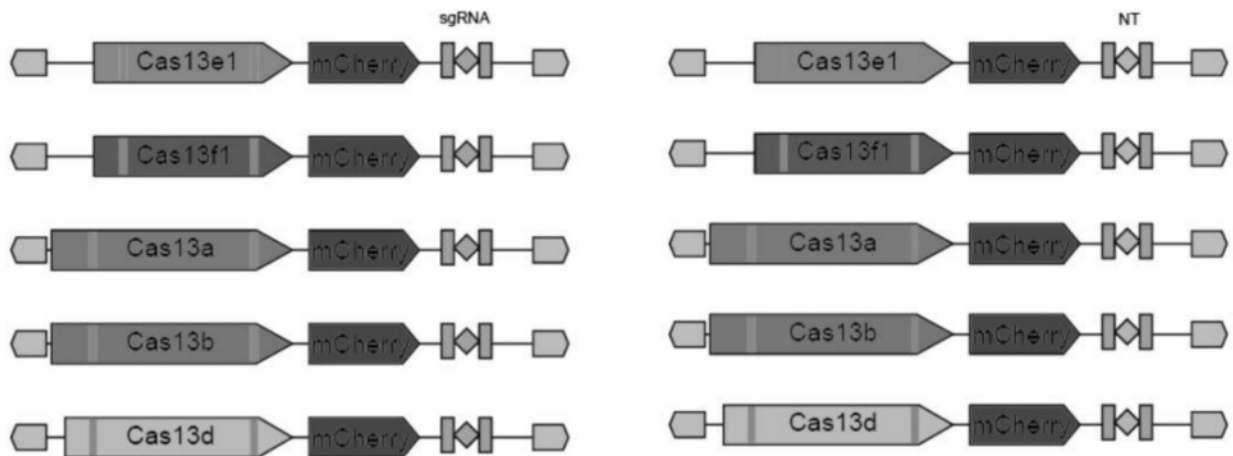


图19

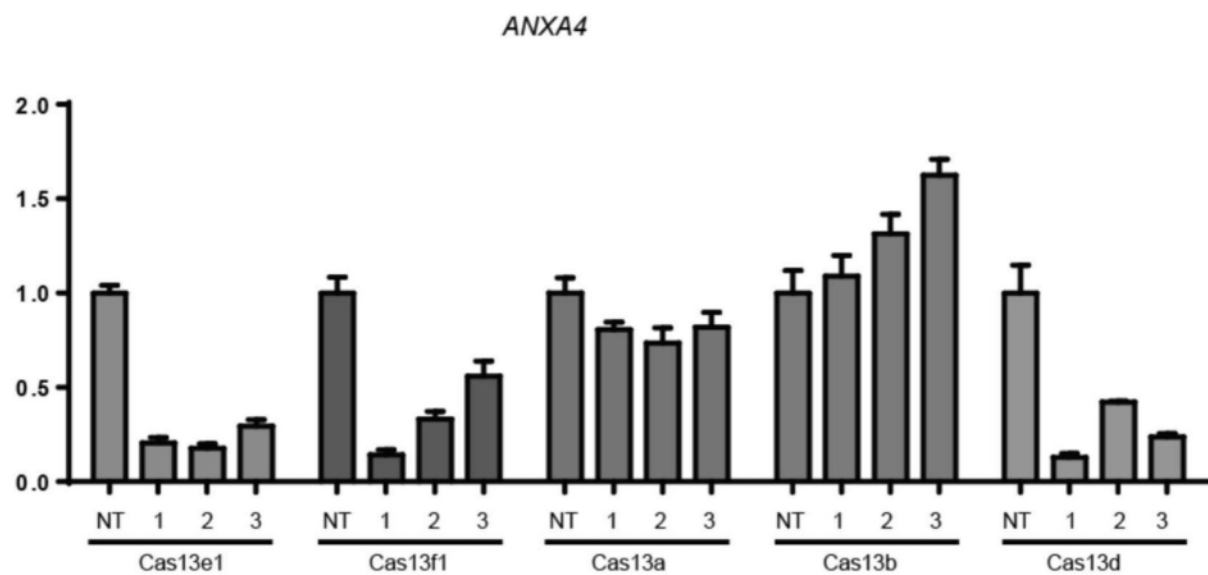


图20