(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 119137268 A
(43)申请公布日 2024.12.13

(21)申请号 202380038039.X
(22)申请日 2023.03.01
(30)优先权数据
63/315,159 2022.03.01 US
63/380,178 2022.10.19 US
63/385,171 2022.11.28 US
(85)PCT国际申请进入国家阶段日
2024.11.01
(86)PCT国际申请的申请数据
PCT/US2023/063446 2023.03.01
(87)PCT国际申请的公布数据
WO2023/168242 EN 2023.09.07
(71)申请人 埃皮克里斯皮尔生物技术公司
地址 美国加利福尼亚州
(72)发明人 杨潇 亓磊 文森特·库蒂拉斯
加布里埃拉·阿尔瓦雷斯
塔比莎·蒂乔 丹尼尔·哈特
(74)专利代理机构 北京信慧永光知识产权代理有限责任公司 11290
专利代理师 王芬 杨国强
(51)Int.Cl.
C12N 9/22(2006.01)
C12N 15/113(2006.01)

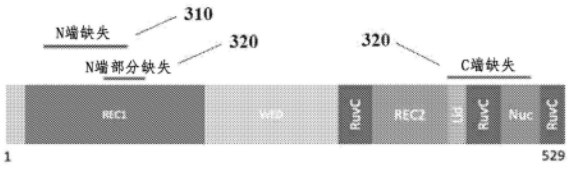
权利要求书4页 说明书210页
序列表（电子公布） 附图14页

(54)发明名称

工程化核酸酶、组合物及其使用方法

(57)摘要

本公开提供了一种或多种工程化核酸酶及其系统、组合物和方法，其中所述一种或多种工程化核酸酶可用于实现对靶多核苷酸序列进行结合、切割和/或编辑。所述一种或多种工程化核酸酶可以是小CRISPR/Cas蛋白的工程化变体。



CN 119137268 A

1.一种工程化多肽,所述工程化多肽包含工程化核酸酶,其中,所述工程化核酸酶包含与SEQ ID NO:1的氨基酸序列至少80%相同的氨基酸序列,其中,与SEQ ID NO:1的多肽序列相比,所述氨基酸序列包含来自氨基酸残基2-100的至少一个缺失。

2.根据权利要求1所述的工程化多肽,其中,与SEQ ID NO:1的氨基酸序列相比,所述至少一个缺失来自氨基酸残基31-40。

3.根据权利要求1所述的工程化多肽,其中,与SEQ ID NO:1的氨基酸序列相比,所述至少一个缺失来自氨基酸残基41-60。

4.根据权利要求1所述的工程化多肽,其中,与SEQ ID NO:1的氨基酸序列相比,所述至少一个缺失来自氨基酸残基61-80。

5.根据权利要求1所述的工程化多肽,其中,与SEQ ID NO:1的氨基酸序列相比,所述至少一个缺失包含至少10个氨基酸残基。

6.根据权利要求1所述的工程化多肽,其中,与SEQ ID NO:1的氨基酸序列相比,所述至少一个缺失包含至少30个氨基酸残基。

7.根据权利要求1所述的工程化多肽,其中,所述工程化核酸酶的所述氨基酸序列与SEQ ID NO:1的氨基酸序列具有至少85%序列同一性。

8.根据权利要求1所述的工程化多肽,其中,所述工程化核酸酶的长度为至多约528个氨基酸。

9.根据权利要求1所述的工程化多肽,其中,与SEQ ID NO:1的氨基酸序列相比,所述工程化核酸酶包含D326或D510处的氨基酸置换,从而降低所述工程化核酸酶的核酸酶活性。

10.根据权利要求1所述的工程化多肽,其中,与SEQ ID NO:1的氨基酸序列相比,所述工程化核酸酶包含选自于由D143R、T147R、K330R和E528R所组成的组中的一个或多个氨基酸置换。

11.根据权利要求1所述的工程化多肽,其中,所述工程化核酸酶包含与SEQ ID NO:11的氨基酸序列具有至少60%序列同一性的氨基酸序列。

12.根据权利要求1所述的工程化多肽,其中,所述工程化核酸酶包含与SEQ ID NO:11的氨基酸序列具有至少80%序列同一性的氨基酸序列。

13.一种用于调节细胞中靶基因的表达和/或活性的方法,所述方法包括:

使所述细胞与工程化多肽接触,所述工程化多肽能够与表现出与所述靶基因的特异性结合的引导核酸分子形成复合体,

其中,所述工程化多肽包含工程化核酸酶,所述工程化核酸酶包含与SEQ ID NO:1的氨基酸序列至少80%相同的氨基酸序列,其中,与SEQ ID NO:1的氨基酸序列相比,所述氨基酸序列包含来自氨基酸残基2-100的至少一个缺失,

其中,在所述接触后,所述复合体与所述靶基因的结合导致对所述靶基因的所述表达和/或活性的调节。

14.根据权利要求13所述的方法,其中,与通过对照复合体接触的对照细胞相比,所述复合体与所述靶基因的所述结合导致对所述细胞中所述靶基因的所述表达和/或活性的增强的调节,所述对照复合体包含(i)含有SEQ ID NO:10的氨基酸序列的对照工程化多肽和(ii)所述引导核酸分子。

15.根据权利要求13所述的方法,其中,与SEQ ID NO:1的氨基酸序列相比,所述至少一

个缺失来自氨基酸残基31-40。

16.根据权利要求13所述的方法,其中,与SEQ ID NO:1的氨基酸序列相比,所述至少一个缺失来自氨基酸残基41-60。

17.根据权利要求13所述的方法,其中,与SEQ ID NO:1的氨基酸序列相比,所述至少一个缺失来自氨基酸残基61-80。

18.根据权利要求13所述的方法,其中,与SEQ ID NO:1的氨基酸序列相比,所述至少一个缺失包含至少10个氨基酸残基。

19.根据权利要求13所述的方法,其中,与SEQ ID NO:1的氨基酸序列相比,所述至少一个缺失包含至少30个氨基酸残基。

20.根据权利要求13所述的方法,其中,所述工程化核酸酶的所述氨基酸序列与SEQ ID NO:1的氨基酸序列至少85%相同。

21.根据权利要求13所述的方法,其中,所述工程化核酸酶的长度为至多约528个氨基酸。

22.根据权利要求13所述的方法,其中,与SEQ ID NO:1的氨基酸序列相比,所述工程化核酸酶包含D326或D510处的氨基酸置换,从而降低所述工程化核酸酶的核酸酶活性。

23.根据权利要求13所述的方法,其中,与SEQ ID NO:1的氨基酸序列相比,所述工程化核酸酶包含选自于由D143R、T147R、K330R和E528R所组成的组中的一个或多个氨基酸置换。

24.根据权利要求13所述的方法,其中,所述氨基酸序列与SEQ IDNO:11的氨基酸序列至少60%相同。

25.根据权利要求13所述的方法,其中,所述氨基酸序列与SEQ ID NO:11的氨基酸序列至少80%相同。

26.一种工程化多肽,所述工程化多肽包含工程化核酸酶,其中,所述工程化核酸酶包含与SEQ ID NO:12的氨基酸序列大于92%相同的氨基酸序列。

27.根据权利要求26所述的工程化多肽,其中,所述工程化核酸酶的所述氨基酸序列与SEQ ID NO:12的氨基酸序列至少约93%相同。

28.根据权利要求26所述的工程化多肽,其中,所述工程化核酸酶的所述氨基酸序列与SEQ ID NO:12的氨基酸序列至少约95%相同。

29.根据权利要求26所述的工程化多肽,其中,所述氨基酸序列的长度为至多约528个氨基酸。

30.根据权利要求26所述的工程化多肽,其中,与SEQ ID NO:1的氨基酸序列相比,所述氨基酸序列包含D326或D510处的氨基酸置换,从而降低所述工程化核酸酶的核酸酶活性。

31.根据权利要求26所述的工程化多肽,其中,与SEQ ID NO:1的氨基酸序列相比,所述氨基酸序列包含选自于由D143R、T147R、K330R和E528R所组成的组中的一个或多个氨基酸置换。

32.根据权利要求26所述的工程化多肽,其中,所述氨基酸序列与SEQ ID NO:11的氨基酸序列至少60%相同。

33.根据权利要求26所述的工程化多肽,其中,所述氨基酸序列与SEQ ID NO:11的氨基酸序列至少80%相同。

34.一种用于调节细胞中靶基因的表达和/或活性的方法,所述方法包括:

使所述细胞与工程化多肽接触,所述工程化多肽能够与表现出与所述靶基因的特异性结合的引导核酸分子形成复合体,

其中,所述工程化多肽包含工程化核酸酶,所述工程化核酸酶包含与SEQ ID NO:12的氨基酸序列大于92%相同的氨基酸序列,

其中,在所述接触后,所述复合体与所述靶基因的结合导致对所述靶基因的所述表达和/或活性的调节。

35.根据权利要求34所述的方法,其中,与通过对照复合体接触的对照细胞相比,所述复合体与所述靶基因的所述结合导致对所述细胞中所述靶基因的所述表达和/或活性的增强的调节,所述对照复合体包含(i)含有SEQ ID NO:10的氨基酸序列的对照工程化多肽和(ii)所述引导核酸分子。

36.根据权利要求34所述的方法,其中,所述工程化核酸酶的所述氨基酸序列与SEQ ID NO:12的氨基酸序列至少约93%相同。

37.根据权利要求34所述的方法,其中,所述工程化核酸酶的所述氨基酸序列与SEQ ID NO:12的氨基酸序列至少约95%相同。

38.根据权利要求34所述的方法,其中,所述氨基酸序列的长度为至多约528个氨基酸。

39.根据权利要求34所述的方法,其中,与SEQ ID NO:1的氨基酸序列相比,所述工程化核酸酶包含D326或D510处的氨基酸置换,从而降低所述工程化核酸酶的核酸酶活性。

40.根据权利要求34所述的方法,其中,与SEQ ID NO:1的氨基酸序列相比,所述工程化核酸酶包含选自于由D143R、T147R、K330R和E528R所组成的组中的一个或多个氨基酸置换。

41.根据权利要求34所述的方法,其中,所述氨基酸序列与SEQ ID NO:11的氨基酸序列至少60%相同。

42.根据权利要求34所述的方法,其中,所述氨基酸序列与SEQ ID NO:11的氨基酸序列至少80%相同。

43.一种组合物,所述组合物包含引导核酸分子,所述引导核酸分子被配置为与Cas蛋白形成复合体,其中,所述引导核酸分子包含:

间隔序列,所述间隔序列表现出与靶基因的特异性结合;和

支架序列,所述支架序列用于与所述Cas蛋白形成所述复合体,其中,所述支架序列包含与SEQ ID NO:597的多核苷酸序列具有至少74%序列同一性的连续多核苷酸序列。

44.根据权利要求43所述的组合物,其中,所述连续多核苷酸序列与SEQ ID NO:597的多核苷酸序列具有至少75%序列同一性。

45.根据权利要求43所述的组合物,其中,所述连续多核苷酸序列与SEQ ID NO:597的多核苷酸序列具有至少80%序列同一性。

46.根据权利要求43所述的组合物,其中,所述连续多核苷酸序列与SEQ ID NO:597的多核苷酸序列具有至少90%序列同一性。

47.根据权利要求43所述的组合物,其中,所述支架序列的长度为至多约110个核苷酸。

48.根据权利要求43所述的组合物,其中,所述支架序列的长度为至多约105个核苷酸。

49.根据权利要求43所述的组合物,所述组合物包含所述Cas蛋白。

50.根据权利要求49所述的组合物,其中,所述Cas蛋白的长度为至多约528个氨基酸。

51.根据权利要求49所述的组合物,其中,所述Cas蛋白的长度为至多约520个氨基酸。

52.一种用于调节细胞中靶基因的表达和/或活性的方法,所述方法包括:

使所述细胞与组合物接触,所述组合物包含引导核酸分子,所述引导核酸分子被配置为与Cas蛋白形成复合体,其中,所述引导核酸分子包含:

间隔序列,所述间隔分子表现出与所述靶基因的特异性结合;和

支架序列,所述支架序列用于与所述Cas蛋白形成所述复合体,其中,所述支架序列包含与SEQ ID NO:597的多核苷酸序列具有至少74%序列同一性的连续多核苷酸序列。

53.根据权利要求52所述的方法,其中,所述复合体与所述靶基因的所述结合导致对所述细胞中所述靶基因的调节,所述调节与通过对照接触的对照细胞中的调节相当,所述对照包含(i)含有SEQ ID NO:500的多核苷酸序列的对照引导核酸分子和(ii)所述Cas蛋白。

54.根据权利要求52所述的方法,其中,所述连续多核苷酸序列与SEQ ID NO:597的多核苷酸序列具有至少75%序列同一性。

55.根据权利要求52所述的方法,其中,所述连续多核苷酸序列与SEQ ID NO:597的多核苷酸序列具有至少80%序列同一性。

56.根据权利要求52所述的方法,其中,所述连续多核苷酸序列与SEQ ID NO:597的多核苷酸序列具有至少90%序列同一性。

57.根据权利要求52所述的方法,其中,所述支架序列的长度为至多约110个核苷酸。

58.根据权利要求52所述的方法,其中,所述支架序列的长度为至多约105个核苷酸。

59.根据权利要求52所述的方法,其中,所述组合物包含所述Cas蛋白。

60.根据权利要求59所述的方法,其中,所述Cas蛋白的长度为至多约528个氨基酸。

61.根据权利要求59所述的方法,其中,所述Cas蛋白的长度为至多约520个氨基酸。

工程化核酸酶、组合物及其使用方法

交叉引用

[0001] 本申请要求于2022年3月1日提交的美国临时申请号63/315,159、于2022年10月19日提交的美国临时申请号63/380,178和于2022年11月28日提交的美国临时申请号63/385,171的权益,通过引用的方式将所述申请整体并入本文。

序列表通过引用的方式并入

[0002] 本申请与电子格式的序列表一起提交。序列表以2023年2月15日创建的大小为426千字节的名为55176-726_601_SL.XML的文件提供。通过引用的方式将电子格式的序列表中的信息整体并入。

背景技术

[0003] 多种核酸酶(例如,核酸内切酶)可用于编辑细胞中的靶序列,或调节细胞中靶基因的表达或活性。例如,可将异源核酸酶引入(例如,递送、表达等)到细胞中,并且异源核酸酶单独或与其它试剂一起可实现靶基因的这种编辑或调节。例如,成簇的规律间隔的短回文重复序列(CRISPR)/CRISPR相关蛋白(Cas)是参与特异性结合、切割和/或编辑靶脱氧核糖核酸(DNA)序列或核糖核酸(RNA)序列(例如,外来DNA序列或RNA序列)的核酸酶家族。这些核酸酶的可编程性质使其可以用作一种通用技术,其彻底改变了靶基因操纵领域,例如,作为基因疗法来治疗或改善受试者的病症(例如,疾病)。

发明内容

[0004] 多种核酸内切酶,例如CRISPR/Cas蛋白(例如,迄今为止使用的Cas12f蛋白)可以具有与Cas9或Cas12a相比更小的尺寸。然而,核酸内切酶的尺寸可能不够小,无法在递送模式(例如病毒载体,如腺相关病毒(AAV)载体)中将其与至少一种另外的试剂(例如,引导RNA、编码治疗性多核苷酸或蛋白质的转基因等)一起包装。因此,本公开的各个方面例如提供了更小但有效结合、切割和/或编辑靶多核苷酸序列的工程化核酸酶、其组合物及其使用方法。

[0005] 本公开的方面提供了工程化多肽,所述工程化多肽包含工程化核酸酶,其中所述工程化核酸酶包含与SEQ ID NO:1的多肽序列至少80%相同的氨基酸序列,其中与SEQ ID NO:1的多肽序列相比,所述氨基酸序列包含来自氨基酸残基2-100的至少一个缺失。

[0006] 本公开的另一方面提供了工程化多肽,所述工程多肽包含工程化核酸酶,其中所述工程化核酸酶包含与SEQ ID NO:12的多肽序列超过92%相同的氨基酸序列。

[0007] 本公开的另一方面提供了工程化多肽,所述工程化多肽包含工程化核酸酶变体,其中所述工程化核酸酶变体包含与SEQ ID NO:12的多肽序列至少70%相同的氨基酸序列,其中与SEQ ID NO:1的多肽序列相比,所述氨基酸序列包含修饰,其中所述修饰包括选自于由A21Q、V23I、N32E、D29E、N33R、E35K、K36Q、I37A、A38G、E40D、K73G、A74T、R75G、K76E、Q83K、G87K、E151A、A340S、H353K、A374K、I387E、N423D、K473Q、T474L、T474R、H497K、L515R、N519T、K521D、K521N、L522I所组成的组中的一个或多个成员,以及来自SEQ ID NO:1的氨基酸残基

400-529的至少一个缺失。

[0008] 本公开的另一方面提供了工程化多肽,所述工程化多肽包含工程化核酸酶变体,其中所述工程化核酸酶变体是嵌合多肽,所述嵌合多肽包含:第一多肽序列,所述第一多肽序列包含与第一Cas蛋白共同的至少3个连续氨基酸残基;以及第二多肽序列,所述第二多肽序列包含与第二Cas蛋白共同的至少3个连续氨基酸残基,其中所述第二Cas蛋白不同于所述第一Cas蛋白,其中所述第一Cas蛋白包含与SEQ ID NO:1的多肽序列至少80%相同的氨基酸序列。

[0009] 本公开的另一方面提供了工程化多肽,所述工程化多肽包含工程化核酸酶变体,其中所述工程化核酸酶是嵌合多肽,所述嵌合多肽包含:第一多肽序列(CP1),所述第一多肽序列包含与第一Cas蛋白的部分共同的至少3个连续氨基酸残基;第二多肽序列(CP2),所述第二多肽序列包含与不同于所述第一Cas蛋白的第二Cas蛋白的部分共同的至少3个连续氨基酸残基;以及第三多肽序列(CPx),所述第三多肽序列包含与以下共同的至少3个连续氨基酸残基:(i)所述第一Cas蛋白的另外的部分,其中第一Cas蛋白的部分和另外的部分在所述第一Cas蛋白中彼此不直接相邻;(ii)所述第二Cas蛋白的另外的部分,其中所述第二Cas蛋白的部分和另外的部分在所述第二Cas蛋白中彼此不直接相邻;或(iii)不同于所述第一Cas蛋白和所述第二Cas蛋白的第三Cas蛋白的部分,其中所述嵌合多肽的长度小于或等于约1,000个氨基酸。

[0010] 本公开的另一方面提供了工程化多肽,所述工程化多肽包含工程化核酸酶变体,其中所述工程化核酸酶变体是嵌合多肽,所述嵌合多肽包含:第一多肽,所述第一多肽包含与第一Cas蛋白共同的至少3个连续氨基酸残基;以及第二多肽,所述第二多肽包含与第二Cas蛋白共同的至少3个连续氨基酸残基,其中所述第二Cas蛋白不同于所述第一Cas蛋白,其中所述第二多肽序列的长度比所述第一多肽序列的长度小约20%。

[0011] 本公开的另一方面提供了工程化多肽,所述工程化多肽包含工程化核酸酶变体,其中所述工程化核酸酶变体:(i)包含与选自表3B、表4B或表5C的成员的多肽序列至少80%相同的氨基酸序列;(ii)不是SEQ ID NO:1-SEQ ID NO:3、SEQ ID NO:10和SEQ ID NO:13-SEQ ID NO:19中的任一个;并且(iii)具有小于或等于约800个氨基酸的长度。

[0012] 本公开的另一方面提供了工程化多肽,所述工程化多肽包含可操作地偶联至基因调节因子的工程化核酸酶变体,其中所述工程化核酸酶变体:(i)包含与SEQ ID NO:1的多肽序列至少70%相同的氨基酸序列;并且(ii)当可操作地偶联至基因调节因子时,与包含可操作地偶联至所述基因调节因子的SEQ ID NO:10的对照工程化多肽诱导的相比,诱导细胞中靶基因的增强的调节。

[0013] 本公开的另一方面提供了控制细胞中的靶基因的方法,所述方法包括使所述细胞与本文公开的工程化多肽中的任一种接触。

[0014] 本公开的另一方面提供了调节细胞中的靶基因的方法,所述方法包括:使所述细胞与包含可操作地偶联至基因调节因子的工程化核酸酶变体的工程化多肽接触,其中所述工程化核酸酶变体包含与SEQ ID NO:1的多肽序列至少70%相同的氨基酸序列,其中与包含可操作地偶联至所述基因调节因子的SEQ ID NO:10的对照工程化多肽导致的相比,所述接触导致所述细胞中靶基因的增强的调节。

[0015] 本公开的另一方面提供了组合物,所述组合物包含引导核酸分子,所述引导核酸

分子被配置为与Cas蛋白形成复合体,其中所述引导核酸分子包含:

[0016] 本公开的另一方面提供了组合物,所述组合物包含引导核酸分子,所述引导核酸分子被配置为与Cas蛋白形成复合体,其中所述引导核酸分子包含:间隔序列,所述间隔序列表现出与靶多核苷酸序列特异性结合;以及支架序列,所述支架序列用于与所述Cas蛋白形成复合体,其中所述支架序列的特征在于:

(i)具有与SEQ ID NO:555的多核苷酸序列具有至少96%序列同一性的连续多核苷酸序列;或具有与SEQ ID NO:597的多核苷酸序列具有至少84%序列同一性的连续多核苷酸序列;

(ii)具有与SEQ ID NO:557的多核苷酸序列具有至少97%序列同一性的连续多核苷酸序列;或具有与SEQ ID NO:598的多核苷酸序列具有至少88%序列同一性的连续多核苷酸序列;

(iii)具有与SEQ ID NO:578的多核苷酸序列具有至少90%序列同一性的连续多核苷酸序列;具有与SEQ ID NO:597的多核苷酸序列具有至少84%序列同一性的连续多核苷酸序列;或具有与SEQ ID NO:599的多核苷酸序列具有至少81%序列同一性的连续多核苷酸序列;

(iv)具有与SEQ ID NO:568的多核苷酸序列具有至少93%序列同一性的连续多核苷酸序列;具有与SEQ ID NO:597的多核苷酸序列具有至少84%序列同一性的连续多核苷酸序列;或具有与SEQ ID NO:600的多核苷酸序列具有至少67%序列同一性的连续多核苷酸序列;或

(v)具有与SEQ ID NO:569的多核苷酸序列具有至少95%序列同一性的连续多核苷酸序列;具有与SEQ ID NO:597的多核苷酸序列具有至少84%序列同一性的连续多核苷酸序列;或具有与SEQ ID NO:601的多核苷酸序列具有至少71%序列同一性的连续多核苷酸序列。

[0017] 本公开的另一方面提供了组合物,所述组合物包含引导核酸分子,所述引导核酸分子被配置为与Cas蛋白形成复合体,其中所述引导核酸分子包含:间隔序列,所述间隔序列表现出与可操作地偶联至靶基因的靶多核苷酸序列特异性结合;以及支架序列,所述支架序列用于与所述Cas蛋白形成复合体,其中所述支架序列表现出与选自表6B、表7B和表8B的成员的多核苷酸序列的至少80%序列同一性,其中所述支架序列与SEQ ID NO:500不同,其中所述复合体与细胞中的靶多核苷酸序列的结合导致所述细胞中所述靶基因的经调节的表达水平,其中(A1)所述复合体导致的靶基因的经调节的表达水平与(A2)包含所述Cas蛋白与含有SEQ ID NO:500的多核苷酸序列的对照引导核酸分子的对照复合体导致的经调节的表达水平相当或优于后者。

[0018] 本公开的另一方面提供了组合物,所述组合物包含引导核酸分子,所述引导核酸分子被配置为与Cas蛋白形成复合体,其中所述引导核酸分子包含:间隔序列,所述间隔序列表现出与可操作地偶联至靶基因的靶多核苷酸序列特异性结合;以及支架序列,所述支架序列用于与所述Cas蛋白形成复合体,其中所述支架序列的长度为至多约158个核苷酸,其中所述复合体与细胞中的靶多核苷酸序列的结合导致所述细胞中所述靶基因的经调节的表达水平,其中(A1)所述复合体导致的所述靶基因的经调节的表达水平与(A2)包含所述Cas蛋白和含有SEQ ID NO:500的多核苷酸序列的对照引导核酸分子的对照复合体导致的

经调节的表达水平相当或优于后者。

[0019] 本公开的另一方面提供了组合物,所述组合物包含编码Cas蛋白和被配置为与所述Cas蛋白形成复合体的引导核酸分子的载体,其中所述载体包含:第一多核苷酸序列,所述第一多核苷酸序列编码所述Cas蛋白;以及第二多核苷酸序列,所述第二多核苷酸序列编码用于与所述Cas蛋白形成复合体的所述引导核酸分子的支架序列,其中所述第一多核苷酸序列的长度与所述第二多核苷酸序列的长度的和为至多约1700个核苷酸。

[0020] 本公开的另一方面提供了控制细胞中的靶基因的方法,所述方法包括使所述细胞与本文公开的组合物中的任一种接触。

[0021] 本公开的另一方面提供了调节细胞中的靶基因的方法,所述方法包括:使所述细胞与包含引导核酸分子和Cas蛋白的复合体接触,其中所述复合体表现出与可操作地偶联至靶基因的靶多核苷酸序列特异性结合,其中所述复合体与所述靶多核苷酸序列的结合导致所述细胞中所述靶基因的经调节的表达水平,其中(A1)所述复合体导致的所述靶基因的经调节的表达水平与(A2)包含所述Cas蛋白和含有SEQ ID NO:500的多核苷酸序列的对照引导核酸分子的对照复合体导致的经调节的水平水平相当或优于后者。

[0022] 本公开的其他方面和优点对于本领域技术人员来说将从以下详细描述中很容易看到,其中仅显示和描述了本公开的说明性实施方式。应认识到,本公开可以有其他不同的实施方式,并且其若干细节可以在各个明显的方面进行修改,所有这些都不背离本公开。因此,附图和描述应被视为说明性的,而不是限制性的。

通过引用并入

[0023] 本说明书中提及的所有出版物、专利和专利申请均通过引用并入本文,其程度与每篇单独的出版物、专利或专利申请均被明确且单独地指明通过引用并入相同。如果通过引用并入的出版物和专利或专利申请与本说明书中包含的公开内容相矛盾,则本说明书旨在取代和/或优先于任何此类矛盾材料。

附图说明

[0024] 本发明的新颖特征特别地在所附权利要求书中阐述。通过参考阐述了利用本发明原理的示例性实施方式的以下详细描述以及附图(本文中也称为“图”),将更好地理解本发明的特征和优点,其中:

[0025] 图1示意性地示出了Un1Cas12f1和AsCas12f之间的结构比对,以鉴定Un1Cas12f1中可能在一个或多个另外的Cas12f同源结构中不保守的一个或多个结构域。

[0026] 图2示意性地示出了Un1Cas12f1和Un2Cas12f1之间的结构比对,以鉴定Un1Cas12f1中可能在一个或多个另外的Cas12f同源结构中不保守的一个或多个结构域。

[0027] 图3A示意性地示出了选择Un1Cas12f1的多个结构域以至少部分地缺失以产生一种或多种工程化核酸酶。图3B示意性地示出了缺失景观方法以产生Un1Cas12f1的工程化核酸酶变体。

[0028] 图4示意性地示出了Cas蛋白(例如,dCas9)的示例缺失景观,以鉴定Cas蛋白中的可以缺失而对Cas蛋白的活性(例如,当dCas9的修饰变体可操作地偶联至基因阻抑因子时诱导转录表达的能力)的影响最小或基本上没有影响的一个或多个结构域。

[0029] 图5显示了本文公开的多种工程化核酸酶变体导致的细胞中内源IFNγ(上图)和

内源CD2(下图)的增强的表达。工程化核酸酶变体被工程化以表现出降低的核酸酶活性(例如,dCas变体),并与基因激活调节因子融合。

[0030] 图6显示了本文公开的多种工程化核酸酶变体导致的细胞中靶基因(例如,GFP)的降低的表达。工程化核酸酶变体被工程化以表现出降低的核酸酶活性(例如,dCas变体),并与基因阻抑调节因子融合。使用了两种不同的基因阻抑调节因子:基因阻抑因子A(上图)和基因阻抑因子B(下图)。

[0031] 图7显示了本文公开的多种工程化核酸酶变体导致的细胞中内源CD2(上图)、内源IFNγ(中图)和内源CXCR4(下图)的增强的表达。工程化核酸酶变体被工程化以表现出降低的核酸酶活性(例如,dCas变体),并与基因激活调节因子融合。

[0032] 图8示意性地示出了被配置为与Cas蛋白形成复合体的引导核酸分子。

[0033] 图9示意性地示出了引导核酸分子的支架区域的区域,该区域可以被修饰(例如,突变或缺失)以对引导核酸分子进行工程化。

[0034] 图10显示了本文公开的多种工程化引导RNA变体导致的细胞中靶基因(例如,GFP)的降低的表达。与对照引导RNA序列("SQ")相比,多种工程化引导RNA变体至少在支架区域中被修饰。用与基因阻抑调节因子偶联的相同dCas蛋白进行降低的基因表达。

[0035] 图11显示了本文公开的另外多种工程化引导RNA变体导致的细胞中靶基因(例如,GFP)的降低的表达。与对照引导RNA序列("SQ")相比,所述另外多种工程化引导RNA变体至少在支架区域中被修饰。用与基因阻抑调节因子偶联的相同dCas蛋白进行降低的基因表达。

[0036] 图12显示了本文公开的另外多种工程化引导RNA变体导致的细胞中内源CD2的增强的表达。与对照引导RNA序列("SQ")相比,所述另外多种工程化引导RNA变体至少在支架区域中被修饰。用与基因激活调节因子偶联的相同dCas蛋白进行增强的基因表达。

[0037] 图13示意性地示出了本文公开的工程化引导核酸分子的实例。

[0038] 图14显示了多种DNA载体的尺寸比较,每种载体编码Cas蛋白和相应的单个引导核酸分子。

[0039] 图15显示了本文公开的截短核酸酶变体t1导致的细胞中内源IFNγ(IFNγ)的增加的表达。工程化核酸酶变体t1与基因激活调节因子VPR融合。

[0040] 图16显示了本文公开的嵌合核酸酶变体导致的细胞中内源CXCR4的减少的表达。工程化核酸酶变体t1与基因阻抑调节因子(ZNF10-KRAB-hDNMT3L)融合。

具体实施方式

[0041] 虽然本文已展示和描述了本发明的各种实施方式,但本领域技术人员将明白,此类实施方式仅以示例的方式提供。本领域技术人员可以想到许多变化、改变和替换,而不会背离本发明。应理解,可以采用本文所述的本发明的实施方式的各种替代方案。

[0042] 每当术语"至少"、"大于"或"大于或等于"位于两个或更多个数值的系列中的第一个数值之前时,术语"至少"、"大于"或"大于或等于"适用于该数值系列中的每个数值。例如,大于或等于1、2或3相当于大于或等于1、大于或等于2、或大于或等于3。

[0043] 每当术语"不大于"、"小于"或"小于或等于"位于两个或多个数值的系列中的第一个数值之前时,术语"不大于"、"小于"或"小于或等于"适用于该数值系列中的每个数值。例

如,小于或等于3、2或1相当于小于或等于3、小于或等于2、或小于或等于1。

[0044] 术语“约”或“大约”通常表示在本领域普通技术人员确定的特定值的可接受误差范围内,这将部分取决于如何测量或确定该值,即测量系统的局限性。例如,根据本领域的实践,“约”可以表示在1个或大于1个标准偏差内。替代地,“约”可以表示给定值的高至20%、高至10%、高至5%或高至1%的范围。替代地,特别是对于生物系统或过程,该术语可以表示在值的数量级内,优选在值的5倍内,更优选在值的2倍内。当在申请和权利要求中描述特定值时,除非另有说明,否则应假定术语“约”表示在特定值的可接受误差范围内。

[0045] 替代项(例如“或”)的使用应理解为表示替代项中的一个、两个或其任何组合。术语“和/或”应理解为表示替代项中的一个或两个。

[0046] 术语“细胞”通常指生物细胞。细胞可以是活的生物体的基本结构、功能和/或生物单位。细胞可以源自任何具有一个或多个细胞的生物体。一些非限制性实例包括:原核细胞、真核细胞、细菌细胞、古细菌细胞、单细胞真核生物体的细胞、原生动物细胞、来自植物的细胞(例如,来自植物作物、水果、蔬菜、谷物、大豆、玉米(corn)、玉蜀黍(maize)、小麦、种子、西红柿、水稻、木薯、甘蔗、南瓜、干草、土豆、棉花、大麻、烟草、开花植物、针叶树、裸子植物、蕨类植物、石松类植物、角苔类植物、苔类植物、藓类植物的细胞)、藻类细胞(例如,布朗葡萄藻(Botryococcus braunii)、莱茵衣藻(Chlamydomonas reinhardtii)、拟微球藻(Nannochloropsis gaditana)、蛋白核小球藻(Chlorella pyrenoidosa)、展枝马尾藻(Sargassum patens C.Agardh)等)、海藻(例如,海带)、真菌细胞(例如,酵母细胞、来自蘑菇的细胞)、动物细胞、来自无脊椎动物(例如,果蝇、刺胞动物、棘皮动物、线虫等)的细胞、来自脊椎动物(例如,鱼、两栖动物、爬行动物、禽类、哺乳动物)的细胞、来自哺乳动物(例如,猪、牛、山羊、绵羊、啮齿动物、大鼠、小鼠、非人类灵长类动物、人类等)的细胞等等。有时细胞并非源自天然生物体(例如,细胞可以是合成的,有时称为人造细胞)。

[0047] 本文所用的术语“核苷酸”通常指碱基-糖-磷酸组合。核苷酸可以包括合成核苷酸。核苷酸可以包括合成核苷酸类似物。核苷酸可以是核酸序列(例如,脱氧核糖核酸(DNA)和核糖核酸(RNA))的单体单元。术语核苷酸可以包括核糖核苷三磷酸,腺苷三磷酸(ATP)、尿苷三磷酸(UTP)、胞苷三磷酸(CTP)、鸟苷三磷酸(GTP),以及脱氧核糖核苷三磷酸,例如dATP、dCTP、dITP、dUTP、dGTP、dTTP或其衍生物。此类衍生物可以包括,例如,[αS]dATP、7-脱氮-dGTP和7-脱氮-dATP,以及赋予含有它们的核酸分子核酸酶抗性的核苷酸衍生物。本文所用的术语核苷酸可指双脱氧核糖核苷三磷酸(ddNTP)及其衍生物。双脱氧核糖核苷三磷酸的示例性实例可以包括但不限于ddATP、ddCTP、ddGTP、ddITP和ddTTP。核苷酸可以是未标记的,或通过众所周知的技术可检测地标记的。标记也可以用量子点进行。可检测标记可以包括例如放射性同位素、荧光标记、化学发光标记、生物发光标记和酶标记。核苷酸的荧光标记可以包括但不限于荧光素、5-羧基荧光素(FAM)、2′7′-二甲氧基-4′5-二氯-6-羧基荧光素(JOE)、罗丹明、6-羧基罗丹明(R6G)、N,N,N′,N′-四甲基-6-羧基罗丹明(TAMRA)、6-羧基-X-罗丹明(ROX)、4-(4′二甲基氨基苯偶氮基)苯甲酸(DABCYL)、级联蓝(Cascade Blue)、俄勒冈绿(Oregon Green)、德克萨斯红(Texas Red)、花青(Cyanine)和5-(2′-氨基乙基)氨基萘-1-磺酸(EDANS)。荧光标记的核苷酸的具体实例可以包括[R6G]dUTP、[TAMRA]dUTP、[R110]dCTP、[R6G]dCTP、[TAMRA]dCTP、[JOE]ddATP、[R6G]ddATP、[FAM]ddCTP、[R110]ddCTP、[TAMRA]ddGTP、[ROX]ddTTP、[dR6G]ddATP、[dR110]ddCTP、[dTAMRA]ddGTP和[dROX]

ddTTP,其可从Perkin Elmer,Foster City,Calif.获得;FluoroLink脱氧核苷酸,FluoroLink Cy3-dCTP、FluoroLink Cy5-dCTP、FluoroLink Fluor X-dCTP、FluoroLink Cy3-dUTP和FluoroLink Cy5-dUTP,其可从Amersham,Arlington Heights,Ill.获得;荧光素-15-dATP、荧光素-12-dUTP、四甲基罗丹明-6-dUTP、IR770-9-dATP、荧光素-12-ddUTP、荧光素-12-UTP和荧光素-15-2′-dATP,其可从Boehringer Mannheim,Indianapolis,Ind.获得;及染色体标记核苷酸,BODIPY-FL-14-UTP、BODIPY-FL-4-UTP、BODIPY-TMR-14-UTP、BODIPY-TMR-14-dUTP、BODIPY-TR-14-UTP、BODIPY-TR-14-dUTP、级联蓝-7-UTP、级联蓝-7-dUTP、荧光素-12-UTP、荧光素-12-dUTP、俄勒冈绿488-5-dUTP、罗丹明绿-5-UTP、罗丹明绿-5-dUTP、四甲基罗丹明-6-UTP、四甲基罗丹明-6-dUTP、德克萨斯红-5-UTP、德克萨斯红-5-dUTP和德克萨斯红-12-dUTP,其可从Molecular Probes,Eugene,Oregon获得。核苷酸也可以通过化学修饰进行标记或标志。化学修饰的单核苷酸可以是生物素-dNTP。生物素化dNTP的一些非限制性实例可以包括生物素-dATP(例如,bio-N6-ddATP、生物素-14-dATP)、生物素-dCTP(例如,生物素-11-dCTP、生物素-14-dCTP)和生物素-dUTP(例如,生物素-11-dUTP、生物素-16-dUTP、生物素-20-dUTP)。

[0048] 术语“多核苷酸”、“寡核苷酸”或“核酸”在本文中可互换使用,通常指任意长度的核苷酸的聚合形式,无论是脱氧核糖核苷酸还是核糖核苷酸,或其类似物,无论是单链、双链还是多链形式。多核苷酸可以是细胞外源的或内源的。多核苷酸可以存在于无细胞环境中。多核苷酸可以是基因或其片段。多核苷酸可以是DNA。多核苷酸可以是RNA。多核苷酸可以具有任何三维结构,并且可以执行任何已知或未知的功能。多核苷酸可以包含一种或多种类似物(例如,改变的骨架、糖或核碱基)。如果存在,则可以在聚合物组装之前或之后对核苷酸结构进行修饰。类似物的一些非限制性实例包括:5-溴尿嘧啶、肽核酸、xeno核酸、morpholinos、锁核酸、二醇核酸、苏糖核酸、双脱氧核苷酸、虫草素、7-脱氮-GTP、荧光团(例如,与糖连接的罗丹明或荧光素)、含硫醇核苷酸、生物素连接的核苷酸、荧光碱基类似物、CpG岛、甲基-7-鸟苷、甲基化核苷酸、肌苷、硫尿苷、假尿苷、二氢尿苷、queuosine和wyosine。多核苷酸的非限制性实例包括基因或基因片段的编码区或非编码区、通过链接分析定义的基因位点(基因座)、外显子、内含子、信使RNA(mRNA)、转移RNA(tRNA)、核糖体RNA(rRNA)、短干扰RNA(siRNA)、短发夹RNA(shRNA)、微小RNA(miRNA)、核酶、cDNA、重组多核苷酸、分支多核苷酸、质粒、载体、任何序列的分离DNA、任何序列的分离RNA、无细胞多核苷酸(包括无细胞DNA(cfDNA)和无细胞RNA(cfRNA))、核酸探针和引物。核苷酸序列可以被非核苷酸组分中断。

[0049] 术语“基因”通常指参与编码RNA转录物的核酸(例如,DNA,如基因组DNA和cDNA)及其相应的核苷酸序列。本文中关于基因组DNA所用的术语包括间插非编码区以及调控区,并且可以包括5′和3′端。在一些用途中,该术语涵盖包括5′和3′非翻译区(5′-UTR和3′-UTR)、外显子和内含子的转录序列。在一些基因中,转录区将包含编码多肽的“开放阅读框”。在该术语的一些用途中,“基因”仅包含编码多肽所必需的编码序列(例如,“开放阅读框”或“编码区”)。在一些情况下,基因不编码多肽,例如核糖体RNA基因(rRNA)和转移RNA(tRNA)基因。在一些情况下,术语“基因”不仅包括转录序列,而且还包括非转录区,包括上游和下游调控区、增强子和启动子。基因可以指位于生物体基因组中其天然位置的“内源基因”或天然基因。基因可以指“外源基因”或非天然基因。非天然基因可以指通常不存在于宿主生物

体中但通过基因转移引入宿主生物体的基因。非天然基因也可以指不位于生物体基因组中其天然位置的基因。非天然基因还可以指包含突变、插入和/或缺失的天然存在的核酸或多肽序列(例如,非天然序列)。

[0050] 术语“缺失”通常指从多肽序列中去除(或丢失)一个或多个(或指定数量的)氨基酸(例如,连续或非连续的氨基酸),或从多核苷酸序列(例如,编码多肽序列的序列)中去除(或丢失)一个或多个(或指定数量的)核酸碱基(例如,连续或非连续的核酸碱基)。术语“内部缺失”通常指不包括多肽的N端或C端或多核苷酸的5’端或3’端的缺失。可以通过与参考序列进行比较来鉴定缺失(例如,内部缺失),例如,通过相对于参考序列指定缺失的起始和终止位置。缺失(例如,内部缺失)与置换不同且有区别。例如,与参考多肽序列相比,至少一个氨基酸的缺失的后面没有在与该至少一个氨基酸相同的位置处的至少一个不同氨基酸的插入,使得包含至少一个氨基酸的缺失的经修饰(或工程化)的多肽序列的尺寸(例如,氨基酸残基的数量)比参考多肽序列小被缺失的至少一个氨基酸的尺寸。

[0051] 术语“序列同一性”通常分别指两个多核苷酸或多肽序列的精确核苷酸与核苷酸或氨基酸与氨基酸的对应性。通常,用于确定序列同一性的技术包括确定多核苷酸的核苷酸序列和/或确定由其编码的氨基酸序列,并将这些序列与第二核苷酸或氨基酸序列进行比较。可以通过确定它们的“百分比同一性”来比较两个或更多个序列(多核苷酸或氨基酸)。两个序列(无论是核酸序列还是氨基酸序列)的百分比同一性是两个比对序列之间精确匹配的数目除以较长序列的长度,然后乘以100。百分比同一性也可以例如通过使用可从美国国立卫生研究院获得的高级BLAST计算机程序(包括2.2.9版)比较序列信息来确定。BLAST程序基于Karlin and Altschul,Proc.Natl.Acad.Sci.USA,87:2264-2268(1990)的比对方法,并且如Altschul,等,J.Mol.Biol.,215:403-410(1990);Karlin And Altschul,Proc.Natl.Acad.Sci.USA,90:5873-5877(1993);和Altschul等,Nucleic Acids Res.,25:3389-3402(1997)中讨论的。该程序可用于确定所比较蛋白质的整个长度的百分比同一性。提供默认参数以优化使用短查询序列的搜索,例如使用blastp程序。该程序还允许使用SEG过滤器来屏蔽查询序列的片段,如Wootton and Federhen,Computers and Chemistry 17:149-163(1993)的SEG程序所确定的。期望的序列同一性程度的范围为约50％至100％以及两者之间的整数值。一般而言,本公开涵盖与本文提供的任何序列具有至少50％、至少55％、至少60％、至少65％、至少70％、至少75％、至少80％、至少85％、至少90％、至少95％或至少98％序列同一性的序列。

[0052] 术语“表达”通常指多核苷酸从DNA模板转录(例如转录成mRNA或其他RNA转录本)的一个或多个过程和/或转录的mRNA随后翻译成肽、多肽或蛋白质的过程。转录本和编码的多肽可以统称为“基因产物”。如果多核苷酸源自基因组DNA,则表达可以包括真核细胞中mRNA的剪接。关于表达的“上调”通常指多核苷酸(例如RNA,例如mRNA)和/或多肽序列相对于其在野生型状态下的表达提高的表达水平,而“下调”通常指多核苷酸(例如RNA,例如mRNA)和/或多肽序列相对于其在野生型状态下的表达降低的表达水平。转染基因的表达可以在细胞中瞬时或稳定地发生。在“瞬时表达”期间,转染基因不会在细胞分裂期间转移到子代细胞中。由于其表达仅限于转染细胞,因此该基因的表达会随着时间的推移而消失。相反,当基因与另一个赋予转染细胞选择优势的基因共转染时,可以发生转染基因的稳定表达。此类选择优势可能是对细胞呈现的某种毒素的抵抗力。

[0053] 术语“表达谱”通常指样品(例如细胞)中一个或多个基因的定量(例如丰度)和定性表达。一个或多个基因可以以核酸分子(例如mRNA或其他RNA转录本)的形式表达和确定。替代地或另外,一个或多个基因可以以多肽(例如通过蛋白质印迹测量的蛋白质)的形式表达和确定。基因的表达谱可以定义为基因在一段时间内表达水平的形状(例如,至少或多至约1小时、至少或多至约2小时、至少或多至约3小时、至少或多至约4小时、至少或多至约5小时、至少或多至约6小时、至少或多至约7小时、至少或多至约8小时、至少或多至约9小时、至少或多至约10小时、至少或多至约11小时、至少或多至约12小时、至少或多至约16小时、至少或多至约18小时、至少或多至约24小时、至少或多至约36小时、至少或多至约48小时、至少多至约3天、至少多至约4天、至少或多至约5天、至少或多至约6天、至少或多至约7天、至少或多至约8天、至少或多至约9天、至少或多至约10天、至少或多至约11天、至少或多至约12天、至少或多至约13天、至少或多至约14天等)。替代地,基因的表达谱可以定义为基因在目标时间点的表达水平(例如,在处理细胞以诱导此类表达水平后至少或多至约1小时、至少或多至约2小时、至少或多至约3小时、至少或多至约4小时、至少或多至约5小时、至少或多至约6小时、至少或多至约7小时、至少或多至约8小时、至少或多至约9小时、至少或多至约10小时、至少或多至约11小时、至少或多至约12小时、至少或多至约16小时、至少或多至约18小时、至少或多至约24小时、至少或多至约36小时、至少或多至约48小时、至少或多至约3天、至少或多至约4天、至少或多至约5天、至少或多至约6天、至少或多至约7天、至少或多至约8天、至少或多至约9天、至少或多至约10天、至少或多至约11天、至少或多至约12天、至少或多至约13天或至少或多至约14天测量的基因的表达水平)。

[0054] 术语“肽”、“多肽”或“蛋白”在本文中可互换使用,通常是指由肽键连接的至少两个氨基酸残基的聚合物。该术语不表示聚合物的特定长度,也不旨在暗示或区分肽是使用重组技术、化学或酶合成产生的或是天然存在的。这些术语适用于天然存在的氨基酸聚合物以及包含至少一个修饰的氨基酸的氨基酸聚合物。在一些情况下,该聚合物可以被非氨基酸中断。这些术语包括任意长度的氨基酸链,包括全长蛋白质,以及具有或不具有二级和/或三级结构(例如,结构域)的蛋白。这些术语还包括已经被修饰的氨基酸聚合物,例如通过二硫键形成、糖基化、脂质化、乙酰化、磷酸化、氧化和任何其他操作(例如与标记组分缀合)。本文使用的术语“氨基酸”通常指天然和非天然氨基酸,包括但不限于经修饰的氨基酸和氨基酸类似物。经修饰的氨基酸可以包括天然氨基酸和非天然氨基酸,它们已经经过化学修饰以包括氨基酸上非天然存在的基团或化学部分。氨基酸类似物可以指氨基酸衍生物。术语“氨基酸”包括D-氨基酸和L-氨基酸两者。

[0055] 本文中关于多肽使用的术语“衍生物”、“变体”或“片段”通常指例如通过氨基酸序列、结构(例如,二级和/或三级)、活性(例如,酶活性)和/或功能而与野生型多肽相关的多肽。与野生型多肽相比,多肽的衍生物、变体和片段可以包括一个或多个氨基酸变异(例如,突变、插入和缺失)、截短、修饰或其组合。

[0056] 本文中关于多肽分子(例如,蛋白质)使用的术语“工程化”、“嵌合”或“重组”通常指由于将基因工程技术应用于编码多肽分子的核酸以及表达该多肽分子的细胞或生物体而具有异源氨基酸序列或改变的氨基酸序列的多肽分子。本文中关于多核苷酸分子(例如,DNA或RNA分子)使用的术语“工程化”或“重组”通常指由于应用基因工程技术而具有异源核酸序列或改变的核酸序列的多核苷酸分子。基因工程技术包括但不限于PCR和DNA克隆技

术;转染、转化和其他基因转移技术;同源重组;定点诱变;和基因融合。在一些情况下,工程化或重组多核苷酸(例如,基因组DNA序列)可以通过基因编辑部分进行修饰或改变。

[0057] 例如,本文中公开的工程化核酸酶(例如,工程化Cas蛋白)不是天然存在的核酸酶(例如,不是天然存在的Cas蛋白)。术语“工程化核酸酶”和“工程化核酸酶变体”在本文中可以互换使用。

[0058] 术语“工程化的”和“经修饰的”在本文中可互换使用。术语“工程化”和“修饰”在本文中可互换使用。术语“工程化细胞”或“经修饰的细胞”在本文中可互换使用。术语“工程化特性”和“经修饰的特性”在本文中可互换使用。

[0059] 术语“增强的表达”、“增加的表达”或“上调的表达”通常指在宿主菌株(例如宿主细胞)中,感兴趣的部分(例如,多核苷酸或多肽)的产生水平高于该感兴趣的部分的正常表达水平。正常表达水平可以基本上为零(或空)或高于零。感兴趣的部分可以包括宿主菌株的内源基因或多肽构建体。感兴趣的部分可以包括引入到或进入宿主菌株中的异源基因或多肽构建体。例如,可以将编码感兴趣的多肽的异源基因敲入(KI)宿主菌株的基因组中,以增强宿主菌株中感兴趣的多肽的表达。

[0060] 术语“增强的活性”、“增加的活性”或“上调的活性”通常是指在宿主菌株(例如,宿主细胞)中,感兴趣的部分(例如,多核苷酸或多肽)的活性被修饰到高于感兴趣的部分的正常活性水平的水平。正常活性水平可以基本上为零(或空)或高于零。感兴趣的部分可以包括宿主菌株的多肽构建体。感兴趣的部分可以包括引入到或进入宿主菌株中的异源多肽构建体。例如,可以将编码感兴趣的多肽的异源基因敲入(KI)宿主菌株的基因组中,以增强宿主菌株中感兴趣的多肽的活性。

[0061] 术语“减少的表达”、“降低的表达”或“下调的表达”通常是指在宿主菌株(例如,宿主细胞)中,感兴趣部分(例如多核苷酸或多肽)的产生水平低于感兴趣的部分的正常表达水平。正常表达水平高于零。感兴趣的部分可以包括宿主菌株的内源基因或多肽构建体。在一些情况下,可以在宿主菌株中敲除或敲低感兴趣的部分。在一些实例中,感兴趣的部分的减少的表达可以包括宿主菌株中此类表达的完全抑制。

[0062] 术语“减少的活性”、“降低的活性”或“下调的活性”通常指在宿主菌株(例如,宿主细胞)中,感兴趣的部分(例如多核苷酸或多肽)的活性被修饰到低于感兴趣的部分的正常活性水平的水平。正常活性水平高于零。感兴趣的部分可以包括宿主菌株的内源基因或多肽构建体。在一些情况下,可以在宿主菌株中敲除或敲低感兴趣的部分。在一些实例中,感兴趣的部分的降低的活性可以包括宿主菌株中此类活性的完全抑制。

[0063] 术语“受试者”、“个体”或“患者”在本文中可互换使用,通常指脊椎动物,优选哺乳动物,例如人类。哺乳动物包括但不限于鼠类、猿类、人、农场动物、运动动物和宠物。还包括体内获得或体外培养的生物实体的组织、细胞及其后代。

[0064] 术语“治疗/处理”通常是指获得有益或期望结果(包括但不限于治疗益处和/或预防益处)的方法。例如,治疗/处理可以包括施用本文公开的系统或细胞群。治疗益处是指对所治疗的一种或多种疾病、病症或症状的任何治疗相关改善或影响。对于预防益处,可以将组合物施用于处于发展为特定疾病、病症或症状风险中的受试者,或施用于报告疾病的一种或多种生理症状的受试者,即使该疾病、病症或症状可能尚未表现出来。

[0065] 术语“有效量”或“治疗有效量”通常指组合物,例如包含异源多肽、异源多核苷酸

和/或经修饰的细胞(例如经修饰的干细胞)的组合物的量,其足以在施用于有需要的受试者时产生期望的活性。在本公开的上下文中,术语"治疗有效"通常指对于通过本公开的方法治疗的紊乱,足以延迟其表现、阻止其进展、缓解或减轻其至少一种症状的组合物的量。

概述

[0066] 本公开的多个方面可以提供更小但有效地结合、切割和/或编辑靶多核苷酸序列的工程化核酸酶,其组合物及其使用方法。此类工程化核酸酶(例如,工程化CRISPR/Cas核酸酶)可以例如实现细胞中靶基因(例如,靶内源基因)的表达或活性的操纵,例如,以治疗或改善受试者的病症(例如,疾病)。基因表达可以支持细胞和组织中的多种生理和病理效应,对许多疾病和病症有所贡献,因此利用本公开的工程化核酸酶的组合物和方法可以以期望的方式调节特定基因的表达以具有治疗益处。

工程化核酸酶、其组合物及其方法

[0067] 在一些方面,本公开提供了工程化核酸酶,所述工程化核酸酶包含与SEQ ID NO:1的多肽序列至少50%相同的氨基酸序列。与SEQ ID NO:1(或SEQ ID NO:10)的多肽序列相比(例如,当与之比对时),工程化核酸酶的氨基酸序列可以包含至少一个缺失。所述至少一个缺失可以根据本公开选择性地去除。如本文所公开的,SEQ ID NO:1编码Un1Cas12f1(或Cas14a1)的多肽序列。如本文所公开的,SEQ ID NO:10编码具有降低的核酸酶活性的Un1Cas12f1的工程化变体。因此,如本文公开的工程化核酸酶的氨基酸序列可以是Un1Cas12f1(或其失活变体)的突变序列(或突变变体)。

[0068] 不希望受理论的束缚,本文公开的工程化核酸酶的氨基酸序列的至少一个缺失可以存在于天然Un1Cas12f1核酸酶的一个或多个区域中,所述一个或多个区域在结构上与对照CRISPR/Cas蛋白不对齐。对照CRISPR/Cas蛋白可以来自1类CRISPR系统或2类CRISPR系统(例如,作为野生型CRISPR/Cas蛋白)。1类CRISPR系统可以分为I、III和IV型,2类CRISPR系统可以分为II、V和VI型。在一些情况下,对照CRISPR/Cas蛋白可以是V型Cas蛋白,例如V-A型Cas蛋白、V-B型Cas蛋白、V-C型Cas蛋白、V-D型Cas蛋白、V-E型Cas蛋白、V-F型Cas蛋白、V-G型Cas蛋白、V-H型Cas蛋白、V-I型Cas蛋白、V-J型Cas蛋白、V-K型Cas蛋白或V-U型Cas蛋白。在一些情况下,对照CRISPR/Cas蛋白可以是V-J型蛋白,例如野生型CasΦ(Cas 12J)蛋白。在一些情况下,对照CRISPR/Cas蛋白可以是Un2Cas12f1(SEQ ID NO:2)或AsCas12f(SEQ ID NO:3)。

[0069] 不希望受理论的束缚,本文所公开的工程化核酸酶的氨基酸序列的至少一个缺失可以通过对由SEQ ID NO:1(或SEQ ID NO:10)的多肽序列编码的Cas核酸酶进行缺失景观研究(例如,迭代和/或全面缺失)来确定。

[0070] 不希望受理论的束缚,本文所公开的工程化核酸酶的多种不同变体可以表现出不同的活性(例如,对对照sgRNA的不同结合亲和力、对对照靶基因的不同结合亲和力、不同靶基因切割水平、不同靶基因活化水平、不同靶基因阻抑水平等)。在一些实施方式中,多种不同变体中的第一变体可以包括氨基酸残基1-100处的至少一个缺失(例如,当与SEQ ID NO:1的多肽序列比对时),并且多种不同变体中的第二变体可以包括氨基酸残基101-529处的至少一个缺失(例如,当与SEQ ID NO:1的多肽序列比对时),并且第一变体和第二变体可表现出不同的活性(例如,与第二变体相比,第一变体可实现增强的靶基因活化和/或表达,反之亦然)。替代地,第一变体和第二变体可表现出相当的活性。在一些实施方式中,多种不同

变体中的第一变体可以包括氨基酸残基1-100处的单个缺失(例如,当与SEQ ID NO:1的多肽序列比对时),并且多种不同变体中的第二变体可以包括氨基酸残基1-100处的多个缺失(例如,当与SEQ ID NO:1的多肽序列比对时),并且第一变体和第二变体可表现出不同的活性(例如,与第一变体相比,第二变体可实现增强的靶基因活化和/或表达,反之亦然)。替代地,第一变体和第二变体可表现出相当的活性。

[0071] SEQ ID NO:1(Un1Cas12f1)

```
  1 MAKNTITKTL KLRIVRPYNS AEVEKIVADE KNNREKIALE KNKDKVKEAC
 51 SKHLKVAAYC TTQVERNACL FCKARKLDDK FYQKLRGQFP DAVFWQEISE
101 IFRQLQKQAA EIYNQSLIEL YYEIFIKGKG IANASSVEHY LSDVCYTRAA
151 ELFKNAAIAS GLRSKIKSNF RLKELKNMKS GLPTTKSDNF PIPLVKQKGG
201 QYTGFEISNH NSDFIIKIPF GRWQVKKEID KYRPWEKFDF EQVQKSPKPI
251 SLLLSTQRRK RNKGWSKDEG TEAEIKKVMN GDYQTSYIEV KRGSKIGEKS
301 AWMLNLSIDV PKIDKGVDPS IIGGIDVGVK SPLVCAINNA FSRYSISDND
351 LFHFNKKMFA RRRILLKKNR HKRAGHGAKN KLKPITILTE KSERFRKKLI
401 ERWACEIADF FIKNKVGTVQ MENLESMKRK EDSYFNIRLR GFWPYAEMQN
451 KIEFKLKQYG IEIRKVAPNN TSKTCSKCGH LNNYFNFEYR KKNKFPHFKC
501 EKCNFKENAD YNAALNISNP KLKSTKEEP
```

[0072] SEQ ID NO:10(Un1Cas12f1的失活核酸酶变体,即dCasMINI)

```
  1 MAKNTITKTL KLRIVRPYNS AEVEKIVADE KNNREKIALE KNKDKVKEAC
 51 SKHLKVAAYC TTQVERNACL FCKARKLDDK FYQKLRGQFP DAVFWQEISE
101 IFRQLQKQAA EIYNQSLIEL YYEIFIKGKG IANASSVEHY LSRVCYRRAA
151 ELFKNAAIAS GLRSKIKSNF RLKELKNMKS GLPTTKSDNF PIPLVKQKGG
201 QYTGFEISNH NSDFIIKIPF GRWQVKKEID KYRPWEKFDF EQVQKSPKPI
251 SLLLSTQRRK RNKGWSKDEG TEAEIKKVMN GDYQTSYIEV KRGSKICEKS
301 AWMLNLSIDV PKIDKGVDPS IIGGIAVGVR SPLVCAINNA FSRYSISDND
351 LFHFNKKMFA RRRILLKKNR HKRAGHGAKN KLKPITILTE KSERFRKKLI
401 ERWACEIADF FIKNKVGTVQ MENLESMKRK EDSYFNIRLR GFWPYAEMQN
451 KIEFKLKQYG IEIRKVAPNN TSKTCSKCGH LNNYFNFEYR KKNKFPHFKC
501 EKCNFKENAA YNAALNISNP KLKSTKERP
```

[0073] SEQ ID NO:2(Un2Cas12f1)

```
  1 MEVQKTVMKT LSLRILRPLY SQEIEKEIKE EKERRKQAGG TGELDGGFYK
 51 KLEKKHSEMF SFDRLNLLLN QLQREIAKVY NHAISELYIA TIAQGNKSNK
101 HYISSIVYNR AYGYFYNAYI ALGICSKVEA NFRSNELLTQ QSALPTAKSD
151 NFPIVLHKQK GAEGEDGGFR ISTEGSDLIF EIPIPFYEYN GENRKEPYKW
201 VKKGGQKPVL KLILSTFRRQ RNKGWAKDEG TDAEIRKVTE GKYQVSQIEI
251 NRGKKLGEHQ KWFANFSIEQ PIYERKPNRS IVGGLDVGIR SPLVCAINNS
301 FSRYSVDSND VFKFSKQVFA FRRRLLSKNS LKRKGHGAAH KLEPITEMTE
351 KNDKFRKKII ERWAKEVTNF FVKNQVGIVQ IEDLSTMKDR EDHFFNQYLR
401 GFWPYYQMQT LIENKLKEYG IEVKRVQAKY TSQLCSNPNC RYWNNYFNFE
451 YRKVNKFPKF KCEKCNLEIS ADYNAARNLS TPDIEKFVAK ATKGINLPEK
```

[0074] SEQ ID NO:3(AsCas12f)

```
  1 MIKVYRYEIV KPLDLDWKEF GTILRQLQQE TRFALNKATQ LAWEWMGFSS
 51 DYKDNHGEYP KSKDILGYTN VHGYAYHTIK TKAYRLNSGN LSQTIKRATD
101 RFKAYQKEIL RGDMSIPSYK RDIPLDLIKE NISVNRMNHG DYIASLSLLS
151 NPAKQEMNVK RKISVIIIVR GAGKTIMDRI LSGEYQVSAS QIIHDDRKNK
201 WYLNISYDFE PQTRVLDLNK IMGIDLGVAV AVYMAFQHTP ARYKLEGGEI
251 ENFRRQVESR RISMLRQGKY AGGARGGHGR DKRIKPIEQL RDKIANFRDT
301 TNHRYSRYIV DMAIKEGCGT IQMEDLTNIR DIGSRFLQNW TYYDLQQKII
351 YKAEEAGIKV IKIDPQYTSQ RCSECGNIDS GNRIGQAIFK CRACGYEANA
401 DYNAARNIAI PNIDKIIAES IK
```

[0075] 在整个本公开中,(i)本文公开的工程化核酸酶的氨基酸序列和SEQ ID NO:1的多肽序列之间的序列比较与(ii)本文公开的工程化核酸酶的氨基酸序列和SEQ ID NO:10的多肽序列之间的序列比较相当(例如,基本上相同)。

[0076] 在一些实施方式中,本文公开的工程化核酸酶的氨基酸序列可以与(i)SEQ ID NO:1的多肽序列或(ii)选自表2的任一种Cas蛋白的多肽序列至少或高至约50%、至少或高至约55%、至少或高至约60%、至少或高至约62%、至少或高至约64%、至少或高至约65%、至少或高至约66%、至少或高至约68%、至少或高至约70%、至少或高至约72%、至少或高至约74%、至少或高至约75%、至少或高至约76%、至少或高至约78%、至少或高至约80%、至少或高至约82%、至少或高至约84%、至少或高至约85%、至少或高至约86%、至少或高至约88%、至少或高至约90%、至少或高至约91%、至少或高至约92%、至少或高至约93%、至少或高至约94%、至少或高至约95%、至少或高至约96%、至少或高至约97%、至少或高至约98%、至少或高至约99%相同。例如,工程化核酸酶的氨基酸序列可以与(i)SEQ ID NO:1的多肽序列或(ii)选自表2的任一种Cas蛋白的多肽序列约80%至约100%相同。

[0077] 在一些实施方式中,本文公开的工程化核酸酶的氨基酸序列与(i)SEQ ID NO:1的多肽序列或(ii)选自表2的任一种Cas蛋白的多肽序列不相同。

[0078] 在一些实施方式中,本文公开的工程化核酸酶的氨基酸序列的至少一个缺失可以包括N端缺失、C端缺失和/或内部缺失。例如,所述至少一个缺失可以不包括N端缺失,而是内部缺失和/或C端缺失。

[0079] 在一些实施方式中,与(i)SEQ ID NO:1的多肽序列或(ii)选自表2的任一种Cas蛋白的多肽序列(当适用时)相比,本文公开的工程化核酸酶的氨基酸序列的至少一个缺失可以来自氨基酸残基1-100(例如氨基酸残基2-100)、氨基酸残基101-429、氨基酸残基101-200、氨基酸残基201-300、氨基酸残基301-400、氨基酸残基401-500、或氨基酸残基500-529、和/或氨基酸残基430-529。

[0080] 在一些实施方式中,与(i)SEQ ID NO:1的多肽序列或(ii)选自表2的任一种Cas蛋白的多肽序列(当适用时)相比,本文公开的工程化核酸酶的氨基酸序列的至少一个缺失可以来自:来自氨基酸残基2-100的至少一个缺失。

[0081] 在一些实施方式中,与(i)SEQ ID NO:1的多肽序列或(ii)选自表2的任一种Cas蛋白的多肽序列(当适用时)相比,工程化核酸酶的氨基酸序列的至少一个缺失可以来自氨基酸残基2-98、氨基酸残基2-96、氨基酸残基2-95、氨基酸残基2-94、氨基酸残基2-92、氨基酸残基2-90、氨基酸残基2-88、氨基酸残基2-86、氨基酸残基2-85、氨基酸残基2-84、氨基酸残基2-82、氨基酸残基2-80、氨基酸残基2-78、氨基酸残基2-76、氨基酸残基2-75、氨基酸残基2-74、氨基酸残基2-72、氨基酸残基2-70、氨基酸残基2-68、氨基酸残基2-66、氨基酸残基2-65、氨基酸残基2-64、氨基酸残基2-62、氨基酸残基2-60、氨基酸残基2-58、氨基酸残基2-56、氨基酸残基2-55、氨基酸残基2-54、氨基酸残基2-52、氨基酸残基2-50、氨基酸残基2-48、氨基酸残基2-46、氨基酸残基2-45、氨基酸残基2-44、氨基酸残基2-42、氨基酸残基2-40、氨基酸残基2-38、氨基酸残基2-36、氨基酸残基2-35、氨基酸残基2-34、氨基酸残基2-32、氨基酸残基2-30、氨基酸残基2-28、氨基酸残基2-26、氨基酸残基2-25、氨基酸残基2-24、氨基酸残基2-22、氨基酸残基2-20、氨基酸残基2-18、氨基酸残基2-16、氨基酸残基2-15、氨基酸残基2-14、氨基酸残基2-12、氨基酸残基2-10、氨基酸残基2-8、氨基酸残基2-6、氨基酸残基2-5或氨基酸残基2-4。

[0082] 在一些实施方式中,与(i)SEQ ID NO:1的多肽序列或(ii)选自表2的任一种Cas蛋白的多肽序列(当适用时)相比,工程化核酸酶的氨基酸序列的至少一个缺失可以来自氨基酸残基2-100、氨基酸残基4-100、氨基酸残基5-100、氨基酸残基6-100、氨基酸残基8-100、氨基酸残基10-100、氨基酸残基12-100、氨基酸残基14-100、氨基酸残基15-100、氨基酸残基16-100、氨基酸残基18-100、氨基酸残基20-100、氨基酸残基22-100、氨基酸残基24-100、氨基酸残基25-100、氨基酸残基26-100、氨基酸残基28-100、氨基酸残基30-100、氨基酸残基32-100、氨基酸残基34-100、氨基酸残基35-100、氨基酸残基36-100、氨基酸残基38-100、氨基酸残基40-100、氨基酸残基42-100、氨基酸残基44-100、氨基酸残基45-100、氨基酸残基46-100、氨基酸残基48-100、氨基酸残基50-100、氨基酸残基52-100、氨基酸残基54-100、氨基酸残基55-100、氨基酸残基56-100、氨基酸残基58-100、氨基酸残基60-100、氨基酸残基62-100、氨基酸残基64-100、氨基酸残基65-100、氨基酸残基66-100、氨基酸残基68-100、氨基酸残基70-100、氨基酸残基72-100、氨基酸残基75-100、氨基酸残基76-100、氨基酸残基78-100、氨基酸残基80-100、氨基酸残基82-100、氨基酸残基84-100、氨基酸残基85-100、氨基酸残基86-100、氨基酸残基88-100、氨基酸残基90-100、氨基酸残基92-100、氨基酸残基94-100、氨基酸残基95-100、氨基酸残基96-100、或氨基酸残基98-100。

[0083] 在一些实施方式中,与(i)SEQ ID NO:1的多肽序列或(ii)选自表2的任一种Cas蛋白的多肽序列(当适用时)相比,工程化核酸酶的氨基酸序列的至少一个缺失可以来自氨基

酸残基30-100、氨基酸残基30-98、氨基酸残基30-96、氨基酸残基30-95、氨基酸残基30-94、氨基酸残基30-92、氨基酸残基30-90、氨基酸残基30-88、氨基酸残基30-86、氨基酸残基30-85、氨基酸残基30-84、氨基酸残基30-82、氨基酸残基30-80、氨基酸残基30-78、氨基酸残基30-76、氨基酸残基30-75、氨基酸残基30-74、氨基酸残基30-72、氨基酸残基30-70、氨基酸残基30-68、氨基酸残基30-66、氨基酸残基30-65、氨基酸残基30-64、氨基酸残基30-62、氨基酸残基30-60、氨基酸残基30-58、氨基酸残基30-56、氨基酸残基30-55、氨基酸残基30-54、氨基酸残基30-52、氨基酸残基30-50、氨基酸残基30-48、氨基酸残基30-46、氨基酸残基30-45、氨基酸残基30-44、氨基酸残基30-42、氨基酸残基30-40、氨基酸残基30-38、氨基酸残基30-36、氨基酸残基30-34、或氨基酸残基30-32。

[0084] 在一些实施方式中,与(i)SEQ ID NO:1的多肽序列或(ii)选自表2的任一种Cas蛋白的多肽序列(当适用时)相比,工程化核酸酶的氨基酸序列的至少一个缺失可以包括来自氨基酸残基55-56、氨基酸残基54-57、氨基酸残基54-58、氨基酸残基53-59、氨基酸残基52-60、氨基酸残基51-61、氨基酸残基50-62、氨基酸残基49-63、氨基酸残基48-64、氨基酸残基47-65、氨基酸残基46-66、氨基酸残基45-67、氨基酸残基44-68、氨基酸残基43-69、氨基酸残基42-70、氨基酸残基41-71、氨基酸残基40-72、氨基酸残基39-73、氨基酸残基38-74、氨基酸残基37-73、氨基酸残基36-74、或氨基酸残基35-75的一个或多个缺失(例如基本上所有的氨基酸残基的缺失)。

[0085] 在一些实施方式中,与SEQ ID NO:1的多肽序列相比,工程化核酸酶的氨基酸序列的至少一个缺失可以来自选自于由氨基酸残基2-10、氨基酸残基11-20、氨基酸残基21-30、氨基酸残基31-40、氨基酸残基41-50、氨基酸残基51-60、氨基酸残基61-70、和氨基酸残基71-80所组成的组中的一个或多个成员(例如,单个成员、至少2个成员、至少3个成员、至少4个成员、至少5个成员或更多)。在一些实施方式中,与(i)SEQ ID NO:1的多肽序列或(ii)选自表2的任一种Cas蛋白的多肽序列(当适用时)相比,工程化核酸酶的氨基酸序列的至少一个缺失可以来自选自于由氨基酸残基2-5、氨基酸残基6-10、氨基酸残基11-15、氨基酸残基16-20、氨基酸残基21-25、氨基酸残基26-30、氨基酸残基31-35、氨基酸残基36-40、氨基酸残基41-45、氨基酸残基46-50、氨基酸残基51-55、氨基酸残基56-60、氨基酸残基61-65、氨基酸残基66-70、氨基酸残基71-75、和氨基酸残基76-80所组成的组中的一个或多个成员(例如,单个成员、至少2个成员、至少3个成员、至少4个成员、至少5个成员或更多)。

[0086] 在一些实施方式中,与(i)SEQ ID NO:1的多肽序列或(ii)选自表2的任一种Cas蛋白的多肽序列(当适用时)相比,本文公开的工程化核酸酶的氨基酸序列的至少一个缺失可以来自选自于由氨基酸残基10-90、氨基酸残基15-85、氨基酸残基20-80、氨基酸残基25-75、氨基酸残基30-70、氨基酸残基35-75、氨基酸残基40-70、氨基酸残基45-65、或氨基酸残基50-60所组成的组中的一个或多个成员(例如,两个以上成员、三个以上成员、四个以上成员、五个以上成员等)。

[0087] 在一些实施方式中,与(i)SEQ ID NO:1的多肽序列或(ii)选自表2的任一种Cas蛋白的多肽序列(当适用时)相比,本文公开的工程化核酸酶的氨基酸序列的至少一个缺失可以来自选自于由氨基酸残基20-30、氨基酸残基25-35、氨基酸残基30-40、氨基酸残基35-45、氨基酸残基40-50、氨基酸残基45-55、氨基酸残基50-60、氨基酸残基55-65、氨基酸残基60-70、氨基酸残基65-75、氨基酸残基70-80、氨基酸残基75-85、和氨基酸残基80-90所组成

的组中的一个或多个成员(例如,两个以上成员、三个以上成员、四个以上成员、五个以上成员等)。

[0088] 在一些实施方式中,与(i)SEQ ID NO:1的多肽序列或(ii)选自表2的任一种Cas蛋白的多肽序列(当适用时)相比,本文公开的工程化核酸酶的氨基酸序列的至少一个缺失可以来自选自于由氨基酸残基20-25、氨基酸残基25-30、氨基酸残基30-35、氨基酸残基35-40、氨基酸残基40-45、氨基酸残基45-50、氨基酸残基50-55、氨基酸残基55-60、氨基酸残基60-65、氨基酸残基65-70、氨基酸残基70-75、氨基酸残基75-80、和氨基酸残基85-90所组成的组中的一个或多个成员(例如,两个以上成员、三个以上成员、四个以上成员、五个以上成员等)。

[0089] 在一些实施方式中,与(i)SEQ ID NO:1的多肽序列或(ii)选自表2的任一种Cas蛋白的多肽序列(当适用时)相比,本文公开的工程化核酸酶的氨基酸序列的至少一个缺失可以来自氨基酸残基430-529。

[0090] 在一些实施方式中,与(i)SEQ ID NO:1的多肽序列或(ii)选自表2的任一种Cas蛋白的多肽序列(当适用时)相比,工程化核酸酶的氨基酸序列的至少一个缺失可以来自氨基酸残基430-528、氨基酸残基430-526、氨基酸残基430-525、氨基酸残基430-524、氨基酸残基430-522、氨基酸残基430-520、氨基酸残基430-518、氨基酸残基430-516、氨基酸残基430-515、氨基酸残基430-514、氨基酸残基430-512、氨基酸残基430-510、氨基酸残基430-508、氨基酸残基430-506、氨基酸残基430-505、氨基酸残基430-504、氨基酸残基430-502、氨基酸残基430-500、氨基酸残基430-498、氨基酸残基430-496、氨基酸残基430-495、氨基酸残基430-494、氨基酸残基430-492、氨基酸残基430-490、氨基酸残基430-488、氨基酸残基430-486、氨基酸残基430-485、氨基酸残基430-484、氨基酸残基430-482、氨基酸残基430-480、氨基酸残基430-478、氨基酸残基430-476、氨基酸残基430-475、氨基酸残基430-474、氨基酸残基430-472、氨基酸残基430-470、氨基酸残基430-468、氨基酸残基430-466、氨基酸残基430-465、氨基酸残基430-464、氨基酸残基430-462、氨基酸残基430-460、氨基酸残基430-458、氨基酸残基430-456、氨基酸残基430-455、氨基酸残基430-454、氨基酸残基430-452、氨基酸残基430-450、氨基酸残基430-448、氨基酸残基430-446、氨基酸残基430-445、氨基酸残基430-444、氨基酸残基430-442、氨基酸残基430-440、氨基酸残基430-438、氨基酸残基430-436、氨基酸残基430-435、氨基酸残基430-434、或氨基酸残基430-432。

[0091] 在一些实施方式中,与(i)SEQ ID NO:1的多肽序列或(ii)选自表2的任一种Cas蛋白的多肽序列(当适用时)相比,工程化核酸酶的氨基酸序列的至少一个缺失可以来自氨基酸残基430-529、氨基酸残基432-529、氨基酸残基434-529、氨基酸残基435-529、氨基酸残基436-529、氨基酸残基438-529、氨基酸残基440-529、氨基酸残基442-529、氨基酸残基444-529、氨基酸残基445-529、氨基酸残基446-529、氨基酸残基448-529、氨基酸残基450-529、氨基酸残基452-529、氨基酸残基454-529、氨基酸残基455-529、氨基酸残基456-529、氨基酸残基458-529、氨基酸残基460-529、氨基酸残基462-529、氨基酸残基464-529、氨基酸残基465-529、氨基酸残基466-529、氨基酸残基468-529、氨基酸残基470-529、氨基酸残基472-529、氨基酸残基474-529、氨基酸残基475-529、氨基酸残基476-529、氨基酸残基478-529、氨基酸残基480-529、氨基酸残基482-529、氨基酸残基484-529、氨基酸残基485-

529、氨基酸残基486-529、氨基酸残基488-529、氨基酸残基490-529、氨基酸残基492-529、氨基酸残基494-529、氨基酸残基495-529、氨基酸残基496-529、氨基酸残基498-529、氨基酸残基500-529、氨基酸残基502-529、氨基酸残基504-529、氨基酸残基505-529、氨基酸残基506-529、氨基酸残基508-529、氨基酸残基510-529、氨基酸残基512-529、氨基酸残基514-529、氨基酸残基515-529、氨基酸残基516-529、氨基酸残基518-529、氨基酸残基520-529、氨基酸残基522-529、氨基酸残基524-529、氨基酸残基525-529、氨基酸残基526-529、或氨基酸残基528-529。

[0092] 在一些实施方式中,与(i)SEQ ID NO:1的多肽序列或(ii)选自表2的任一种Cas蛋白的多肽序列(当适用时)相比,工程化核酸酶的氨基酸序列的至少一个缺失可以来自选自于由氨基酸残基450-459、氨基酸残基460-469、氨基酸残基470-479、氨基酸残基480-489、氨基酸残基490-499、氨基酸残基500-509、氨基酸残基510-519、和氨基酸残基520-529所组成的组中的一个或多个成员(例如单个成员、至少2个成员、至少3个成员、至少4个成员、至少5个成员或更多)。在一些实施方式中,与(i)SEQ ID NO:1的多肽序列或(ii)选自表2的任一种Cas蛋白的多肽序列(当适用时)相比,工程化核酸酶的氨基酸序列的至少一个缺失可以来自选自于由氨基酸残基450-459、氨基酸残基460-465、氨基酸残基466-469、氨基酸残基470-475、氨基酸残基476-479、氨基酸残基480-485、氨基酸残基486-489、氨基酸残基490-495、氨基酸残基496-499、氨基酸残基500-505、氨基酸残基506-509、氨基酸残基510-515、氨基酸残基516-519、氨基酸残基520-525、和氨基酸残基526-529所组成的组中的一个或多个成员(例如单个成员、至少2个成员、至少3个成员、至少4个成员、至少5个成员或更多)。

[0093] 在一些实施方式中,与(i)SEQ ID NO:1的多肽序列或(ii)选自表2的任一种Cas蛋白的多肽序列(当适用时)相比,工程化核酸酶的氨基酸序列的至少一个缺失可以包括单个氨基酸残基的缺失。在一些实施方式中,与(i)SEQ ID NO:1的多肽序列或(ii)选自表2的任一种Cas蛋白的多肽序列(当适用时)相比,工程化核酸酶的氨基酸序列的至少一个缺失可以包括多个氨基酸残基的缺失。当与(i)SEQ ID NO:1的多肽序列或(ii)选自表2的任一种Cas蛋白的多肽序列(当适用时)比对时,缺失的多个氨基酸残基可以彼此相邻(例如,为连续的)。替代地或另外,当与(i)SEQ ID NO:1的多肽序列或(ii)选自表2的任一种Cas蛋白的多肽序列(当适用时)比对时,缺失的多个氨基酸残基可以彼此不相邻。所述多个氨基酸残基可以包括至少或多至约2个氨基酸残基、至少或多至约3个氨基酸残基、至少或多至约4个氨基酸残基、至少或多至约5个氨基酸残基、至少或多至约6个氨基酸残基、至少或多至约7个氨基酸残基、至少或多至约8个氨基酸残基、至少或多至约9个氨基酸残基、至少或多至约10个氨基酸残基、至少或多至约11个氨基酸残基、至少或多至约12个氨基酸残基、至少或多至约13个氨基酸残基、至少或多至约14个氨基酸残基、至少或多至约15个氨基酸残基、至少或多至约16个氨基酸残基、至少或多至约17个氨基酸残基、至少或多至约18个氨基酸残基、至少或多至约19个氨基酸残基、至少或多至约20个氨基酸残基、至少或多至约22个氨基酸残基、至少或多至约24个氨基酸残基、至少或多至约25个氨基酸残基、至少或多至约26个氨基酸残基、至少或多至约28个氨基酸残基、至少或多至约30个氨基酸残基、至少或多至约32个氨基酸残基、至少或多至约34个氨基酸残基、至少或多至约35个氨基酸残基、至少或多至约36个氨基酸残基、至少或多至约38个氨基酸残基、至少或多至约40个氨基酸残基、至少或

多至约42个氨基酸残基、至少或多至约44个氨基酸残基、至少或多至约45个氨基酸残基、至少或多至约46个氨基酸残基、至少或多至约48个氨基酸残基、至少或多至约50个氨基酸残基、至少或多至约52个氨基酸残基、至少或多至约54个氨基酸残基、至少或多至约55个氨基酸残基、至少或多至约56个氨基酸残基、至少或多至约58个氨基酸残基、至少或多至约60个氨基酸残基、至少或多至约62个氨基酸残基、至少或多至约64个氨基酸残基、至少或多至约65个氨基酸残基、至少或多至约66个氨基酸残基、至少或多至约68个氨基酸残基、至少或多至约70个氨基酸残基、至少或多至约72个氨基酸残基、至少或多至约74个氨基酸残基、至少或多至约75个氨基酸残基、至少或多至约76个氨基酸残基、至少或多至约78个氨基酸残基、至少或多至约80个氨基酸残基、至少或多至约82个氨基酸残基、至少或多至约84个氨基酸残基、至少或多至约85个氨基酸残基、至少或多至约86个氨基酸残基、至少或多至约88个氨基酸残基、至少或多至约90个氨基酸残基、至少或多至约92个氨基酸残基、至少或多至约94个氨基酸残基、至少或多至约95个氨基酸残基、至少或多至约96个氨基酸残基、至少或多至约98个氨基酸残基、或至少或多至约100个氨基酸残基。

[0094] 在一些实施方式中,多个氨基酸残基缺失以生成工程化核酸酶的氨基酸序列可以包括多个非连续氨基酸残基的缺失,例如与(i)SEQ ID NO:1的多肽序列或(ii)选自表2的任一种Cas蛋白的多肽序列(当适用时)相比。在一些情况下,当与(i)SEQ ID NO:1的多肽序列或(ii)选自表2的任一种Cas蛋白的多肽序列(当适用时)比对时,工程化核酸酶的氨基酸序列中多个非连续氨基酸残基的缺失可以包括彼此不连续的第一氨基酸残基(例如,第一单个氨基酸残基或第一组多个氨基酸残基,例如第一组连续氨基酸残基)和第二氨基酸残基(例如,第二单个氨基酸残基或第二组多个氨基酸残基,例如第二组连续氨基酸残基)的缺失。例如,第一氨基酸残基和第二氨基酸残基的位置(例如,当与SEQ ID NO:1的多肽序列比对时)可以相隔至少或多至约1个、至少或多至约2个、至少或多至约3个、至少或多至约4个、至少或多至约5个、至少或多至约10个、至少或多至约15个、至少或多至约20个、至少或多至约30个、至少或多至约40个、至少或多至约50个、至少或多至约60个、至少或多至约70个、至少或多至约80个、至少或多至约90个、至少或多至约100个、至少或多至约120个、至少或多至约150个、至少或多至约200个、至少或多至约250个、至少或多至约300个、至少或多至约350个、至少或多至约400个、至少或多至约450个、至少或多至约500个、至少或多至约510个、至少或多至约520个、或至少或多至约525个氨基酸残基。

[0095] 在一些实施方式中,多个氨基酸残基的缺失以生成工程化核酸酶的氨基酸序列可以包括多个连续氨基酸残基的缺失,例如与(i)SEQ ID NO:1的多肽序列或(ii)选自表2的任一种Cas蛋白的多肽序列(当适用时)相比。在一些情况下,多个连续氨基酸残基的缺失可以包括至少或多至约2个连续氨基酸残基、至少或多至约3个连续氨基酸残基、至少或多至约4个连续氨基酸残基、至少或多至约5个连续氨基酸残基、至少或多至约6个连续氨基酸残基、至少或多至约7个连续氨基酸残基、至少或多至约8个连续氨基酸残基、至少或多至约9个连续氨基酸残基、至少或多至约10个连续氨基酸残基、至少或多至约11个连续氨基酸残基、至少或多至约12个连续氨基酸残基、至少或多至约13个连续氨基酸残基、至少或多至约14个连续氨基酸残基、至少或多至约15个连续氨基酸残基、至少或多至约16个连续氨基酸残基、至少或多至约17个连续氨基酸残基、至少或多至约18个连续氨基酸残基、至少或多至约19个连续氨基酸残基、至少或多至约20个连续氨基酸残基、至少或多至约21个连续氨基

酸残基、至少或多至约22个连续氨基酸残基、至少或多至约23个连续氨基酸残基、至少或多至约24个连续氨基酸残基、至少或多至约25个连续氨基酸残基、至少或多至约26个连续氨基酸残基、至少或多至约27个连续氨基酸残基、至少或多至约28个连续氨基酸残基、至少或多至约29个连续氨基酸残基、至少或多至约30个连续氨基酸残基、至少或多至约31个连续氨基酸残基、至少或多至约32个连续氨基酸残基、至少或多至约34个连续氨基酸残基、至少或多至约35个连续氨基酸残基、至少或多至约36个连续氨基酸残基、至少或多至约37个连续氨基酸残基、至少或多至约38个连续氨基酸残基、至少或多至约39个连续氨基酸残基、至少或多至约40个连续氨基酸残基、至少或多至约45个连续氨基酸残基、至少或多至约50个连续氨基酸残基、至少或多至约55个连续氨基酸残基、至少或多至约60个连续氨基酸残基、至少或多至约65个连续氨基酸残基、至少或多至约70个连续氨基酸残基、至少或多至约75个连续氨基酸残基、至少或多至约80个连续氨基酸残基、至少或多至约85个连续氨基酸残基、至少或多至约80个连续氨基酸残基90、至少或多至约95个连续氨基酸残基、或至少或多至约100个连续氨基酸残基的缺失,例如与(i)SEQ ID NO:1的多肽序列或(ii)选自表2的任一种Cas蛋白的多肽序列(当适用时)相比。

[0096] 例如,缺失可以包括多个连续氨基酸残基的单缺失。在另一个实例中,当与(i)SEQ ID NO:1的多肽序列或(ii)选自表2的任一种Cas蛋白的多肽序列(当适用时)比对时,缺失可以包括第一多个连续氨基酸残基的第一缺失和第二多个连续氨基酸残基的第二缺失,并且第一多个连续氨基酸残基和第二多个连续氨基酸残基可以不连续(例如,可以彼此不相邻)。

[0097] 在一些实施方式中,与(i)SEQ ID NO:1的多肽序列或(ii)选自表2的任一种Cas蛋白的多肽序列(当适用时)相比,本文公开的工程化核酸酶的氨基酸序列可以包括一个或多个异源氨基酸残基(例如,一个或多个多肽序列)的添加。一个或多个异源氨基酸残基可以在工程化核酸酶的至少一个缺失的位置。替代地或另外,一个或多个异源氨基酸残基可以不在工程化核酸酶的至少一个缺失的位置。例如,一个或多个异源氨基酸残基可以在工程化核酸酶的至少一个缺失的位置的上游和/或下游。

[0098] 在一些实施方式中,工程化核酸酶的一个或多个异源氨基酸残基可以包含单个氨基酸残基。在一些实施方式中,工程化核酸酶的一个或多个异源氨基酸残基可以包含多个氨基酸残基,例如,至少或多至约2个氨基酸残基、至少或多至约3个氨基酸残基、至少或多至约4个氨基酸残基、至少或多至约5个氨基酸残基、至少或多至约6个氨基酸残基、至少或多至约7个氨基酸残基、至少或多至约8个氨基酸残基、至少或多至约9个氨基酸残基、至少或多至约10个氨基酸残基、至少或多至约11个氨基酸残基、至少或多至约12个氨基酸残基、至少或多至约13个氨基酸残基、至少或多至约14个氨基酸残基、至少或多至约15个氨基酸残基、至少或多至约16个氨基酸残基、至少或多至约17个氨基酸残基、至少或多至约18个氨基酸残基、至少或多至约19个氨基酸残基、至少或多至约20个氨基酸残基、至少或多至约22个氨基酸残基、至少或多至约24个氨基酸残基、至少或多至约25个氨基酸残基、至少或多至约26个氨基酸残基、至少或多至约28个氨基酸残基、至少或多至约30个氨基酸残基、至少或多至约32个氨基酸残基、至少或多至约34个氨基酸残基、至少或多至约35个氨基酸残基、至少或多至约36个氨基酸残基、至少或多至约38个氨基酸残基、至少或多至约40个氨基酸残基、至少或多至约42个氨基酸残基、至少或多至约44个氨基酸残基、至少或多至约45个氨基

酸残基、至少或多至约46个氨基酸残基、至少或多至约48个氨基酸残基、至少或多至约50个氨基酸残基、至少或多至约52个氨基酸残基、至少或多至约54个氨基酸残基、至少或多至约55个氨基酸残基、至少或多至约56个氨基酸残基、至少或多至约58个氨基酸残基、至少或多至约60个氨基酸残基、至少或多至约62个氨基酸残基、至少或多至约64个氨基酸残基、至少或多至约65个氨基酸残基、至少或多至约66个氨基酸残基、至少或多至约68个氨基酸残基、至少或多至约70个氨基酸残基、至少或多至约72个氨基酸残基、至少或多至约74个氨基酸残基、至少或多至约75个氨基酸残基、至少或多至约76个氨基酸残基、至少或多至约78个氨基酸残基、至少或多至约80个氨基酸残基、至少或多至约82个氨基酸残基、至少或多至约84个氨基酸残基、至少或多至约85个氨基酸残基、至少或多至约86个氨基酸残基、至少或多至约88个氨基酸残基、至少或多至约90个氨基酸残基、至少或多至约92个氨基酸残基、至少或多至约94个氨基酸残基、至少或多至约95个氨基酸残基、至少或多至约96个氨基酸残基、至少或多至约98个氨基酸残基、或至少或多至约100个氨基酸残基。

[0099] 在一些实施方式中,一个或多个异源氨基酸残基的多个氨基酸残基可以是连续氨基酸残基。在一些实施方式中,一个或多个异源氨基酸残基的多个氨基酸残基可以包含多个非连续氨基酸残基。

[0100] 在一些实施方式中,工程化核酸酶的一个或多个异源氨基酸残基可以包括异源多肽序列(例如,对于(i)SEQ ID NO:1的多肽序列或(ii)选自表2的任一种Cas蛋白的多肽序列(当适用时)是异源的)。异源多肽序列可以表现出净正电荷(例如,净+1电荷、净+2电荷、净+3电荷、净+4电荷、净+5电荷等,例如,在约pH 7.4的缓冲液中测量)。异源多肽序列可表现出净负电荷(例如,净-1电荷、净-2电荷、净-3电荷、净-4电荷、净-5电荷等,例如,在约pH 7.4的缓冲液中测量)。异源多肽序列可表现出中性电荷,例如,在约pH 7.4的缓冲液中测量。在一些情况下,异源多肽序列可以包含与SEQ ID NO:11的多肽序列至少约50%、至少约60%、至少约65%、至少约70%、至少约75%、至少约80%、至少约85%、至少约90%、至少约95%、至少约99%或约100%相同的氨基酸序列。

[0101] SEQ ID NO:11(异源多肽)

1 ERRKQAGGTG E

[0102] 在一些实施方式中,本文公开的工程化核酸酶的氨基酸序列可以包含来自(i)与SEQ ID NO:1的多肽序列相比的氨基酸残基2-100和/或(ii)与SEQ ID NO:1的多肽序列相比的氨基酸残基430-529的一个或多个氨基酸残基的缺失,并且所述氨基酸序列可进一步包含与SEQ ID NO:1的多肽序列相比来自如本文公开的氨基酸残基101-429的一个或多个另外的氨基酸残基(例如,至少1个、至少2个、至少3个、至少4个、至少5个、至少6个、至少7个、至少8个、至少9个、至少10个、至少15个、至少20个或更多个另外的氨基酸残基)的缺失。

[0103] 在一些实施方式中,本文公开的工程化核酸酶的氨基酸序列可以包含来自(i)与选自表2的任一种Cas蛋白的多肽序列相比的氨基酸残基2-100和/或(ii)与选自表2的任一种Cas蛋白的多肽序列相比的最后100个C端氨基酸残基的一个或多个氨基酸残基的缺失,并且所述氨基酸序列可进一步包含与选自表2的任一种Cas蛋白的多肽序列相比来自其间的氨基酸残基的一个或多个另外的氨基酸残基(例如,至少1个、至少2个、至少3个、至少4个、至少5个、至少6个、至少7个、至少8个、至少9个、至少10个、至少15个、至少20个或更多个另外的氨基酸残基)的缺失。

[0104] 在一些实施方式中,与SEQ ID NO:1中提供的天然Un1Cas12f1核酸酶相比,本文公开的工程化多肽可以包含一个或多个异源氨基酸残基(例如,除突变之外)。一个或多个异源氨基酸残基的存在可以增强例如工程化多肽的活性、稳定性、表达、与相应的引导核酸分子的结合等。

[0105] 在一些实施方式中,本文公开的工程化多肽(例如,失活的Cas核酸酶变体)可以包含与SEQ ID NO:12的多肽序列至少或多至约50%、至少或多至约55%、至少或多至约60%、至少或多至约65%、至少或多至约70%、至少或多至约75%、至少或多至约80%、至少或多至约85%、至少或多至约90%、至少或多至约91%、至少或多至约92%、至少或多至约93%、至少或多至约94%、至少或多至约95%、至少或多至约96%、至少或多至约97%、至少或多至约98%、至少或多至约99%、或约100%相同的氨基酸序列。

[0106] SEQ ID NO:12(示例的工程化核酸酶多肽)

```
  1 MAKNTITKTL KLRIVRPYNS AEVEKIVADE KERRKQAGGT GELDDKFYQK
 51 LRGQFPDAVF WQEISEIFRQ LQKQAAEIYN QSLIELYYEI FIKGKGIANA
101 SSVEHYLSRV CYRRAAELFK NAAIASGLRS KIKSNFRLKE LKNMKSGLPT
151 TKSDNFPIPL VKQKGGQYTG FEISNHNSDF IIKIPFGRWQ VKKEIDKYRP
201 WEKFDFEQVQ KSPKPISLLL STQRRKRNKG WSKDEGTEAE IKKVMNGDYQ
251 TSYIEVKRGS KICEKSAWML NLSIDVPKID KGVDPSIIGG IAVGVRSPLV
301 CAINNAFSRY SISDNDLFHF NKKMFARRRI LLKKNRHKRA GHGAKNKLKP
351 ITILTEKSER FRKKLIERWA CEIADFFIKN KVGTVQMENL ESMKRKEDSY
401 FNIRLRGFWP YAEMQNKIEF KLKQYGIEIR KVAPNNTSKT CSKCGHLNNY
451 FNFEYRKKNK FPHFKCEKCN FKENAAYNAA LNISNPKLKS TKERP
```

[0107] 在一些实施方式中,本文公开的工程化多肽可以包含与来自表3B(例如,SEQ ID NO:20-SEQ ID NO:27中的一个或多个)、表4B(例如,SEQ ID NO:28-SEQ ID NO:111中的一个或多个)和/或表5C(例如,SEQ ID NO:112-SEQ ID NO:201中的一个或多个)的工程化核酸酶变体的多肽序列至少或多至约50%、至少或多至约55%、至少或多至约60%、至少或多至约65%、至少或多至约70%、至少或多至约75%、至少或多至约80%、至少或多至约85%、至少或多至约90%、至少或多至约91%、至少或多至约92%、至少或多至约93%、至少或多至约94%、至少或多至约95%、至少或多至约96%、至少或多至约97%、至少或多至约98%、至少或多至约99%、或约100%相同的氨基酸序列。

[0108] 在一些实施方式中,本文所公开的工程化多肽可以与以下中的任一个不相同:SEQ ID NO:1和来自表2的任何Cas蛋白(SEQ ID NO:2、SEQ ID NO:3和SEQ ID NO:13-SEQ ID NO:19)。

[0109] 在一些实施方式中,与SEQ ID NO:1或SEQ ID NO:10的多肽序列相比,本文公开的工程化多肽可以包含至少一个氨基酸修饰。所述至少一个氨基酸修饰可以包括选自于由相对于如本文公开的SEQ ID NO:1或SEQ ID NO:10的多肽序列的一个或多个氨基酸残基的任何缺失(例如来自SEQ ID NO:1或SEQ ID NO:10的氨基酸残基400-529的至少一个缺失),以及A21Q、V23I、N32E、D29E、N33R、E35K、K36Q、I37A、A38G、E40D、K73G、A74T、R75G、K76E、Q83K、G87K、E151A、A340S、H353K、A374K、I387E、N423D、K473Q、T474L、T474R、H497K、L515R、N519T、K521D、K521N、L522I所组成的组中的一个或多个成员。与SEQ ID NO:1或SEQ ID NO:10的多

肽序列相比,所述一个或多个成员可以包含选自于由A21Q、V23I、N32E、D29E、N33R、E35K、K36Q、I37A、A38G、E40D、K73G、A74T、R75G、K76E、Q83K、G87K、E151A、A340S、H353K、A374K、I387E、N423D、K473Q、T474L、T474R、H497K、L515R、N519T、K521D、K521N、L522I所组成的组中的至少或多至约1个、至少或多至约2个、至少或多至约3个、至少或多至约4个、至少或多至约5个、至少或多至约6个、至少或多至约7个、至少或多至约8个、至少或多至约9个、至少或多至约10个、至少或多至约11个、至少或多至约12个、至少或多至约13个、至少或多至约14个、至少或多至约15个、至少或多至约20个氨基酸、至少或多至约25个氨基酸、或至少或多至约30个氨基酸修饰。在一些情况下,所述至少一个氨基酸修饰可以包含选自表5A的至少一组修饰。在一些情况下,所述至少一个氨基酸修饰可以包含选自表5B的至少一种修饰组合。例如,选自表5B的至少一种修饰组合可能不是cA2.55或cA2.84。

[0110] 在一些实施方式中,与SEQ ID NO:1或SEQ ID NO:10的多肽序列相比,本文公开的工程化多肽可以包含至少一个氨基酸修饰。所述至少一个氨基酸修饰可以包含选自于由A21Q、V23I、D29E、N33R、E40D、Q83K、G87K、E151A、A340S、H353K、A374K、I387E、N423D、K473Q、T474L、T474R、H497K、L515R、N519T、K521D、K521N和L522I所组成的组中的一个或多个成员(例如,至少1、2、3、4、5、6、7、8、9、10、11、12、13、14、15或更多个成员)。

[0111] 在一些实施方式中,与SEQ ID NO:1或SEQ ID NO:10的多肽序列相比,本文公开的工程化多肽可以包含至少一个氨基酸修饰。所述至少一个氨基酸修饰可以包含选自于由N32E、N33R、E35K、K36Q、I37A、A38G、K73G、A74T、R75G和K76E所组成的组中的一个或多个成员(例如,至少1、2、3、4、5、6、7、8、9或10个成员)。

[0112] 在一些实施方式中,本文公开的工程化多肽的工程化核酸酶变体可以是包含源自不同Cas蛋白的不同多核苷酸序列结构域的嵌合多肽。所述嵌合多肽可以包含第一多肽序列和第二多肽序列,第一多肽序列包含与第一Cas蛋白共同的至少3个连续氨基酸残基,第二多肽序列包含与第二Cas蛋白共同的至少3个连续氨基酸残基,其中第二Cas蛋白不同于第一Cas蛋白。第一Cas蛋白和第二Cas蛋白可以是不同的天然存在的Cas蛋白。第一Cas蛋白和第二Cas蛋白可以具有彼此相差不超过500个、不超过400个、不超过350个、不超过300个、不超过250个、不超过200个、不超过150个、不超过140个、不超过130个、不超过120个、不超过110个、不超过100个、不超过90个、不超过80个、不超过70个、不超过60个、不超过50个、不超过40个、不超过30个或不超过20个氨基酸残基的尺寸(或氨基酸序列长度)。

[0113] 在一些情况下,第一多肽序列的长度可以与第二多肽序列的长度基本上相同。替代地,第一多肽序列的长度可以与第二多肽序列的长度不同。第二多肽序列的长度可以小于或等于第一多肽序列的长度的约90%、小于或等于约85%、小于或等于约80%、小于或等于约75%、小于或等于约70%、小于或等于约65%、小于或等于约60%、小于或等于约55%、小于或等于约50%、小于或等于约45%、小于或等于约40%、小于或等于约35%、小于或等于约30%、小于或等于约25%、小于或等于约20%、小于或等于约18%、小于或等于约16%、小于或等于约15%、小于或等于约14%、小于或等于约12%、小于或等于约10%、小于或等于约9%、小于或等于约8%、小于或等于约7%、小于或等于约6%、小于或等于约5%、小于或等于约4%、小于或等于约3%、小于或等于约2%、或小于或等于约1%。第二多肽序列的长度可以为第一多肽序列的长度的至少或多至约1%、至少或多至约2%、至少或多至约3%、至少或多至约4%、至少或多至约5%、至少或多至约6%、至少或多至约7%、至少或多

至约8%、至少或多至约9%、至少或多至约10%、至少或多至约12%、至少或多至约14%、至少或多至约15%、至少或多至约16%、至少或多至约18%、至少或多至约20%、至少或多至约25%、至少或多至约30%、至少或多至约35%、至少或多至约40%、至少或多至约45%、至少或多至约50%、至少或多至约55%、至少或多至约60%、至少或多至约70%、至少或多至约80%、至少或多至约90%、或至少或多至约95%。

[0114] 在一些情况下,第一Cas蛋白和第二Cas蛋白可以是选自SEQ ID NO:1和来自表2的任何Cas蛋白(SEQ ID NO:2、SEQ ID NO:3和SEQ ID NO:13-SEQ ID NO:19)的两个不同成员。第一Cas蛋白或第二Cas蛋白可以包含与选自SEQ ID NO:1和来自表2的任何Cas蛋白(SEQ ID NO:2、SEQ ID NO:3和SEQ ID NO:13-SEQ ID NO:19)的成员的多肽序列至少或多至约5%、至少或多至约10%、至少或多至约15%、至少或多至约20%、至少或多至约30%、至少或多至约40%、至少或多至约50%、至少或多至约60%、至少或多至约70%、至少或多至约75%、至少或多至约80%、至少或多至约85%、至少或多至约90%、至少或多至约92%、至少或多至约95%、至少或多至约96%、至少或多至约97%、至少或多至约98%、至少或多至约99%、或基本上约100%相同的氨基酸序列。例如,此类成员可以是SEQ ID NO:1或SEQ ID NO:2。

[0115] 在一些情况下,工程化核酸酶变体的嵌合多肽的第一多肽可以包含与第一Cas蛋白共同的至少或多至约4个连续氨基酸残基、至少或多至约5个连续氨基酸残基、至少或多至约6个连续氨基酸残基、至少或多至约7个连续氨基酸残基、至少或多至约8个连续氨基酸残基、至少或多至约9个连续氨基酸残基、至少或多至约10个连续氨基酸残基、至少或多至约11个连续氨基酸残基、至少或多至约12个连续氨基酸残基、至少或多至约13个连续氨基酸残基、至少或多至约14个连续氨基酸残基、至少或多至约15个连续氨基酸残基、至少或多至约18个连续氨基酸残基、至少或多至约20个连续氨基酸残基、至少或多至约30个连续氨基酸残基、至少或多至约40个连续氨基酸残基、或至少或多至约50个连续氨基酸残基。

[0116] 在一些情况下,工程化核酸酶变体的嵌合多肽的第二多肽可以包含与第二Cas蛋白共同的至少或多至约4个连续氨基酸残基、至少或多至约5个连续氨基酸残基、至少或多至约6个连续氨基酸残基、至少或多至约7个连续氨基酸残基、至少或多至约8个连续氨基酸残基、至少或多至约9个连续氨基酸残基、至少或多至约10个连续氨基酸残基、至少或多至约11个连续氨基酸残基、至少或多至约12个连续氨基酸残基、至少或多至约13个连续氨基酸残基、至少或多至约14个连续氨基酸残基、至少或多至约15个连续氨基酸残基、至少或多至约18个连续氨基酸残基、至少或多至约20个连续氨基酸残基、至少或多至约30个连续氨基酸残基、至少或多至约40个连续氨基酸残基、或至少或多至约50个连续氨基酸残基。

[0117] 在一些情况下,在工程化核酸酶变体的嵌合多肽的第二多肽中,第一多肽的长度可以比第二多肽的长度大至少或多至约1个氨基酸、至少或多至约2个氨基酸、至少或多至约5个氨基酸、至少或多至约10个氨基酸、至少或多至约15个氨基酸、至少或多至约20个氨基酸、至少或多至约25个氨基酸、至少或多至约30个氨基酸、至少或多至约35个氨基酸、至少或多至约40个氨基酸、至少或多至约45个氨基酸、至少或多至约50个氨基酸、至少或多至约55个氨基酸、至少或多至约60个氨基酸、至少或多至约70个氨基酸、至少或多至约80个氨基酸、至少或多至约90个氨基酸、至少或多至约100个氨基酸、至少或多至约110个氨基酸、至少或多至约120个氨基酸、至少或多至约130个氨基酸、至少或多至约140个氨基酸、至少

或多至约150个氨基酸、至少或多至约200个氨基酸、至少或多至约250个氨基酸、至少或多至约300个氨基酸、至少或多至约350个氨基酸、至少或多至约400个氨基酸、至少或多至约450个氨基酸、或至少或多至约500个氨基酸。

[0118] 在一些情况下,第一多肽可以源自第一Cas蛋白的N端50%、N端45%、N端40%、N端35%、N端30%、N端25%、N端20%、N端15%、N端10%或N端5%。替代地或另外,第一多肽可以源自第一Cas蛋白的C端50%、C端45%、C端40%、C端35%、C端30%、C端25%、C端20%、C端15%、C端10%或C端5%。在一些情况下,第二多肽可以源自第二Cas蛋白的N端50%、N端45%、N端40%、N端35%、N端30%、N端25%、N端20%、N端15%、N端10%或N端5%。替代地或另外,第二多肽可以源自第二Cas蛋白的C端50%、C端45%、C端40%、C端35%、C端30%、C端25%、C端20%、C端15%、C端10%或C端5%。

[0119] 在一些情况下,第一多肽可以源自来自第一Cas蛋白的N端的前5个氨基酸残基、前10个氨基酸残基、前15个氨基酸残基、前20个氨基酸残基、前30个氨基酸残基、前40个氨基酸残基、前50个氨基酸残基、前60个氨基酸残基、前70个氨基酸残基、前80个氨基酸残基、前90个氨基酸残基、前100个氨基酸残基、前150个氨基酸残基、前200个氨基酸残基、前250个氨基酸残基或前300个氨基酸残基。替代地或另外,第一多肽可以源自来自第一Cas蛋白的C端的前5个氨基酸残基、前10个氨基酸残基、前15个氨基酸残基、前20个氨基酸残基、前30个氨基酸残基、前40个氨基酸残基、前50个氨基酸残基、前60个氨基酸残基、前70个氨基酸残基、前80个氨基酸残基、前90个氨基酸残基、前100个氨基酸残基、前150个氨基酸残基、前200个氨基酸残基、前250个氨基酸残基或前300个氨基酸残基。

[0120] 在一些情况下,第二多肽可以源自来自第二Cas蛋白的N端的前5个氨基酸残基、前10个氨基酸残基、前15个氨基酸残基、前20个氨基酸残基、前30个氨基酸残基、前40个氨基酸残基、前50个氨基酸残基、前60个氨基酸残基、前70个氨基酸残基、前80个氨基酸残基、前90个氨基酸残基、前100个氨基酸残基、前150个氨基酸残基、前200个氨基酸残基、前250个氨基酸残基或前300个氨基酸残基。替代地或另外,第二多肽可以源自来自第二Cas蛋白的C端的前5个氨基酸残基、前10个氨基酸残基、前15个氨基酸残基、前20个氨基酸残基、前30个氨基酸残基、前40个氨基酸残基、前50个氨基酸残基、前60个氨基酸残基、前70个氨基酸残基、前80个氨基酸残基、前90个氨基酸残基、前100个氨基酸残基、前150个氨基酸残基、前200个氨基酸残基、前250个氨基酸残基或前300个氨基酸残基。

[0121] 在一些情况下,本文公开的工程化多肽的工程化核酸酶变体可以包含第三多肽序列,所述第三多肽序列包含与第一Cas蛋白共同的至少3个连续氨基酸残基(或更多,如本文公开的)。在这种情况下,第一多肽序列和第三多肽序列在嵌合多肽内可以彼此连续或不连续。替代地或另外,第三多肽序列可以包含与第二Cas蛋白共同的至少3个连续氨基酸残基(或更多,如本文公开的)。第一多肽序列和第三多肽序列在嵌合多肽中可以彼此不连续。替代地,第一多肽序列和第三多肽序列在嵌合多肽中可以彼此连续。第二多肽序列和第三多肽序列在嵌合多肽中可以彼此不连续。替代地,第二多肽序列和第三多肽序列在嵌合多肽中可以彼此连续。另外替代地或另外,第三多肽序列可以包含与第三Cas蛋白共同的至少3个连续氨基酸残基,所述第三Cas蛋白不同于第一Cas蛋白和第二Cas蛋白。例如,第一Cas蛋白、第二Cas蛋白和第三Cas蛋白可以是选自SEQ ID NO:1和选自表2的任何Cas蛋白的三个不同成员。

[0122] 在一些情况下,工程化核酸酶变体的嵌合多肽的第三多肽可以包含与第三Cas蛋白共同的至少或多至约4个连续氨基酸残基、至少或多至约5个连续氨基酸残基、至少或多至约6个连续氨基酸残基、至少或多至约7个连续氨基酸残基、至少或多至约8个连续氨基酸残基、至少或多至约9个连续氨基酸残基、至少或多至约10个连续氨基酸残基、至少或多至约11个连续氨基酸残基、至少或多至约12个连续氨基酸残基、至少或多至约13个连续氨基酸残基、至少或多至约14个连续氨基酸残基、至少或多至约15个连续氨基酸残基、至少或多至约18个连续氨基酸残基、至少或多至约20个连续氨基酸残基、至少或多至约30个连续氨基酸残基、至少或多至约40个连续氨基酸残基、或至少或多至约50个连续氨基酸残基。

[0123] 在一些情况下,第三多肽可以源自来自第三Cas蛋白的N端的前5个氨基酸残基、前10个氨基酸残基、前15个氨基酸残基、前20个氨基酸残基、前30个氨基酸残基、前40个氨基酸残基、前50个氨基酸残基、前60个氨基酸残基、前70个氨基酸残基、前80个氨基酸残基、前90个氨基酸残基、前100个氨基酸残基、前150个氨基酸残基、前200个氨基酸残基、前250个氨基酸残基或前300个氨基酸残基。替代地或另外,第三多肽可以源自来自第三Cas蛋白的C端的前5个氨基酸残基、前10个氨基酸残基、前15个氨基酸残基、前20个氨基酸残基、前30个氨基酸残基、前40个氨基酸残基、前50个氨基酸残基、前60个氨基酸残基、前70个氨基酸残基、前80个氨基酸残基、前90个氨基酸残基、前100个氨基酸残基、前150个氨基酸残基、前200个氨基酸残基、前250个氨基酸残基或前300个氨基酸残基。

[0124] 在一些情况下,在本文公开的工程化核酸酶变体的嵌合多肽中,第一多肽序列、第二多肽序列和第三多肽序列可以从嵌合多肽的N端到C端以以下结构中的任一种排列:CP1-CP2-CPx(I)、CP1-CPx-CP2(II)、CP2-CP1-CPx(III)、CP2-CPx-CP1(IV)、CPx-CP1-CP2(V)、或CPx-CP2-CP1(VI),其中“-”可以是氨基酸接头(例如,包含一个或多个氨基酸序列)或直接共价键。本文公开的氨基酸接头可以包含单个氨基酸、至少或多至约2个氨基酸、至少或多至约3个氨基酸、至少或多至约4个氨基酸、至少或多至约5个氨基酸、至少或多至约8个氨基酸、至少或多至约10个氨基酸、至少或多至约12个氨基酸、至少或多至约15个氨基酸、至少或多至约16个氨基酸、或至少或多至约20个氨基酸。氨基酸接头可以包含至少一个甘氨酸、至少一个丝氨酸或至少一个甘氨酸-丝氨酸二肽。

[0125] 在一些实施方式中,本文公开的工程化核酸酶(例如,本文公开的嵌合多肽)的氨基酸序列的长度可以为至多528个氨基酸、至多527个氨基酸、至多526个氨基酸、至多525个氨基酸、至多524个氨基酸、至多523个氨基酸、至多522个氨基酸、至多521个氨基酸、至多520个氨基酸、至多519个氨基酸、至多518个氨基酸、至多517个氨基酸、至多516个氨基酸、至多515个氨基酸、至多514个氨基酸、至多513个氨基酸、至多512个氨基酸、至多511个氨基酸、至多510个氨基酸、至多509个氨基酸、至多508个氨基酸、至多507个氨基酸、至多506个氨基酸、至多505个氨基酸、至多504个氨基酸、至多503个氨基酸、至多502个氨基酸、至多501个氨基酸、至多约500个氨基酸、至多约495个氨基酸、至多约490个氨基酸、至多约485个氨基酸、至多约480个氨基酸、至多约475个氨基酸、至多约470个氨基酸、至多约465个氨基酸、至多约460个氨基酸、至多约455个氨基酸、至多约450个氨基酸、至多约445个氨基酸、至多约440个氨基酸、至多约435个氨基酸、至多约430个氨基酸、至多约425个氨基酸、至多约420个氨基酸、至多约415个氨基酸、至多约410个氨基酸、至多约405个氨基酸、至多约400个氨基酸、至多约395个氨基酸、至多约390个氨基酸、至多约385个氨基酸、至多约380个氨基

酸、至多约375个氨基酸、至多约370个氨基酸、至多约365个氨基酸、至多约360个氨基酸、至多约355个氨基酸、至多约350个氨基酸、至多约345个氨基酸、至多约340个氨基酸、至多约335个氨基酸、至多约330个氨基酸、至多约325个氨基酸、至多约320个氨基酸、至多约315个氨基酸、至多约310个氨基酸、至多约305个氨基酸、或至多约300个氨基酸。

[0126] 在一些实施方式中,包含本文公开的氨基酸序列的工程化核酸酶的长度可以为至多约1000个氨基酸、至多约950个氨基酸、至多约900个氨基酸、至多约850个氨基酸、至多约800个氨基酸、至多约750个氨基酸、至多约700个氨基酸、至多约650个氨基酸、至多约640个氨基酸、至多约630个氨基酸、至多约620个氨基酸、至多约610个氨基酸、至多约600个氨基酸、至多约590个氨基酸、至多约580个氨基酸、至多约570个氨基酸、至多约560个氨基酸、至多约550个氨基酸、至多约540个氨基酸、至多约530个氨基酸、至多约520个氨基酸、至多约510个氨基酸、至多约500个氨基酸、至多约490个氨基酸、至多约480个氨基酸、至多约470个氨基酸、至多约460个氨基酸、至多约450个氨基酸、至多约440个氨基酸、至多约430个氨基酸、至多约420个氨基酸、至多约410个氨基酸、至多约400个氨基酸、至多约350个氨基酸、或至多约300个氨基酸。

[0127] 在一些实施方式中,本文公开的工程化核酸酶变体的至少一部分可以源自(例如,通过工程化)天然存在的Cas蛋白(例如,本文所述的第一Cas蛋白、第二Cas蛋白或第三Cas蛋白)。在一些情况下,天然存在的Cas蛋白的长度可以为至多约800个氨基酸、至多约750个氨基酸、至多约700个氨基酸、至多约650个氨基酸、至多约600个氨基酸、至多约550个氨基酸、至多约540个氨基酸、至多约530个氨基酸、至多约510个氨基酸、至多约500个氨基酸、至多约490个氨基酸、至多约480个氨基酸、至多约470个氨基酸、至多约460个氨基酸、至多约450个氨基酸、或至多约400个氨基酸。天然存在的Cas蛋白例如可以是来自SEQ ID NO:1和选自表2的任何Cas蛋白的成员。

[0128] 在一些实施方式中,本文公开的工程化核酸酶变体的嵌合多肽的第一多肽序列、第二多肽序列和/或第三多肽序列可以不源自Cas12a。在一些实施方式中,工程化核酸酶变体的嵌合多肽可以不源自Cas12a。在一些实施方式中,工程化核酸酶变体的嵌合多肽可以完全源自一个或多个Cas12f型直系同源物(例如,SEQ ID NO:1或选自表2的Cas蛋白)。

[0129] 在一些实施方式中,包含本文公开的氨基酸序列的工程化核酸酶可以进行突变和/或修饰以产生核酸酶缺陷蛋白或相对于野生型Cas蛋白具有降低的核酸酶活性的蛋白。核酸酶缺陷蛋白可以保留结合靶基因(例如DNA)的能力,但可能缺乏核酸切割活性或具有降低的核酸切割活性。在一些实施方式中,与由SEQ ID NO:1编码的Cas核酸酶或选自表2的Cas蛋白相比,包含本文公开的氨基酸序列的工程化核酸酶可以表现出降低的核酸酶活性(例如,核酸酶缺陷或核酸酶无效)。降低的核酸酶活性可以为由SEQ ID NO:1编码的Cas核酸酶或选自表2的Cas蛋白的活性的至多约95%、至多约90%、至多约80%、至多约70%、至多约60%、至多约50%、至多约40%、至多约30%、至多约20%、至多约10%、至多约5%、至多约1%、至多约0.5%、至多约0.1%或更小。在一些情况下,与SEQ ID NO:1的多肽序列相比,工程化核酸酶可以包含D326和/或D510处的置换。例如,D326和/或D510置换可以是丙氨酸替换(例如,D326A和/或D510A)。

[0130] 在一些实施方式中,本文公开的工程化核酸酶的氨基酸序列可以包含天然氨基酸序列中的一个或多个置换,其中这些置换中的至少一些的位置遵循一个或多个特定规则,

这些规则被确定为对工程化核酸酶具有令人惊讶的优势。在一些情况下,基于其产生工程化核酸酶的变体的能力(例如,其可能能够在真核细胞内发挥作用)来选择特定置换规则。根据这些特定规则,天然氨基酸序列中的一个或多个置换的全部或部分(1)在天然氨基酸序列的(D/E/K/N)X(R/F)(E/K)N基序内或不超过该基序下游30个氨基酸,(2)在天然氨基酸序列的位置241处或不超过该位置上游或下游30个氨基酸,(3)在天然氨基酸序列的位置516处或不超过该位置上游或下游30个氨基酸,和/或(4)在天然氨基酸序列中具有带电氨基酸。

[0131] 在一些实施方式中,与(i)SEQ ID NO:1的多肽序列或(ii)本文公开的不具有一个或多个置换的工程化核酸酶的氨基酸序列相比,本文公开的工程化核酸酶的氨基酸序列可以在(D/E/K/N)X(R/F)(E/K)N基序内或不超过该基序上游和/或下游的阈值长度(例如,30个氨基酸残基)的氨基酸位置处包含一个或多个置换。在一些情况下,一个或多个置换中的至少一个可例如在基序内或不超过基序的28个氨基酸、26个氨基酸、24个氨基酸、22个氨基酸、20个氨基酸、18个氨基酸、16个氨基酸、14个氨基酸、12个氨基酸或10个氨基酸。在一些情况下,一个或多个置换中的至少一个可以是置换为R、A、S或G氨基酸残基。在一些情况下,一个或多个置换可以包括在选自于由D143、T147、E151和K154所组成的组中的一个或多个位置处的置换(例如,D143R、T147R、E151R和/或K154R)。在一些情况下,一个或多个置换可以包括在选自于由N504、E507、N516、N519、E527和E528所组成的组中的一个或多个位置处的置换(例如,N504R、E507R、N516R、N519R、E527R和/或E528R)。在一些情况下,一个或多个置换可以包括在选自于由K11、K73、D143、T147、E151、K154、E241、D318、K330、K457、E425、E462、N504、E507、N516、N519、E527和E528所组成的组中的一个或多个位置处的置换(例如,K11R、K73R、D143R、T147R、E151R、K154R、E241R、D318R、K330R、E425N、K457R、E462R、N504R、E507R、N516R、N519R、E527R和/或E528R)。

[0132] 在一些实施方式中,本文公开的包含(D/E/K/N)X(R/F)(E/K)N基序上游和/或下游的一个或多个置换的工程化核酸酶的氨基酸序列可表现出阳离子电荷(例如,正电荷),其比缺少所述一个或多个置换的工程化核酸酶的对照氨基酸序列大至少或多至约1个阳离子电荷、至少或多至约2个阳离子电荷、至少或多至约3个阳离子电荷、至少或多至约4个阳离子电荷、至少或多至约5个阳离子电荷、至少或多至约6个阳离子电荷、至少或多至约7个阳离子电荷、至少或多至约8个阳离子电荷、至少或多至约9个阳离子电荷、至少或多至约10个阳离子电荷、至少或多至约11个阳离子电荷、至少或多至约12个阳离子电荷、至少或多至约13个阳离子电荷、至少或多至约14个阳离子电荷、至少或多至约15个阳离子电荷、至少或多至约16个阳离子电荷、至少或多至约17个阳离子电荷、或至少或多至约18个阳离子电荷。

[0133] 不希望受理论的束缚,本文所公开的包含(D/E/K/N)X(R/F)(E/K)N基序上游和/或下游的一个或多个置换的工程化核酸酶的氨基酸序列可以表现出对(i)引导核酸序列(例如,引导RNA序列)和/或(ii)Cas/引导核酸复合体的靶多核苷酸序列(例如,细胞中的靶基因,如靶内源基因)的增强的(例如,更高的)结合亲和力,与(A)缺少所述一个或多个置换的工程化核酸酶的对照氨基酸序列和/或(B)由SEQ ID NO:1编码的CRISPR/Cas蛋白相比,高至少或多至约1%、至少或多至约5%、至少或多至约10%、至少或多至约15%、至少或多至约20%、至少或多至约25%、至少或多至约30%、至少或多至约40%、至少或多至约50%、至少或多至约60%、至少或多至约70%、至少或多至约80%、至少或多至约90%、至少或多至

约100％、至少或多至约110％、至少或多至约120％、至少或多至约125％、至少或多至约150％、至少或多至约200％、至少或多至约250％、至少或多至约300％、至少或多至约350％、至少或多至约400％、至少或多至约450％、或至少或多至约500％，如通过表面等离子体共振(SPR)或等温滴定量热法(IRC)测定所确定的。

[0134] 不希望受理论的束缚，本文所公开的包含(D/E/K/N)X(R/F)(E/K)N基序上游和/或下游的一个或多个置换的工程化核酸酶的氨基酸序列可以表现出对(i)引导核酸序列(例如，引导RNA序列)和/或(ii)Cas/引导核酸复合体的靶多核苷酸序列(例如，细胞中的靶基因，如靶内源基因)的增强的(例如，更高的)结合亲和力，与(A)缺少所述一个或多个置换的工程化核酸酶的对照氨基酸序列和/或(B)由SEQ ID NO:1编码的CRISPR/Cas蛋白相比，高至少或多至约0.1倍、至少或多至约0.2倍、至少或多至约0.5倍、至少或多至约1倍、至少或多至约2倍、至少或多至约3倍、至少或多至约4倍、至少或多至约5倍、至少或多至约6倍、至少或多至约7倍、至少或多至约8倍、至少或多至约9倍、至少或多至约10倍、至少或多至约15倍、至少或多至约20倍、至少或多至约25倍、至少或多至约30倍、至少或多至约35倍、或至少或多至约40倍，如通过表面等离子体共振(SPR)或等温滴定量热法(IRC)测定所确定的。

[0135] 在一些实施方式中，本公开提供了包含本文公开的工程化核酸酶的系统。在一些实施方式中，该系统可以包括重组表达(或产生)形式的工程化核酸酶。在一些实施方式中，该系统可以包括至少编码工程化核酸酶的一个或多个多核苷酸。在一些实施方式中，该系统可以包括至少包含(例如，工程化以包含，例如转染或转导以表达)所述工程化核酸酶的细胞(或细胞群)。

工程化核酸酶的使用方法

[0136] 在一些实施方式中，本文公开的工程化核酸酶可用于实现靶多核苷酸序列的结合、切割和/或编辑，例如，以调控由靶多核苷酸序列编码或可操作地偶联至靶多核苷酸序列的多肽(例如，蛋白质)的靶多核苷酸序列的表达和/或活性水平。在一些情况下，可将包含工程化核酸酶的异源多肽引入细胞(例如，哺乳动物细胞)中，以实现细胞的靶多核苷酸序列(例如，细胞的内源基因或异源基因)的结合、切割和/或编辑。在一些实施方式中，本文公开的工程化核酸酶或包含所述工程化核酸酶的蛋白质可称为致动部分(actuator moiety)。

[0137] 在一些实施方式中，本文公开的工程化核酸酶可保留由SEQ ID NO:1的多肽序列编码的核酸酶的核酸酶活性或选自表2的Cas蛋白的核酸酶活性的至少一部分(例如，基本上全部)。

[0138] 在一些实施方式中，本文公开的工程化核酸酶可以为核酸酶缺陷型。在一些实施方式中，工程化核酸酶可以为核酸酶无效的DNA结合蛋白，其不诱导靶DNA序列的转录激活或阻抑，除非其与本公开的一种或多种异源基因效应因子以复合体存在。在一些实施方式中，工程化核酸酶可以为核酸酶无效的DNA结合蛋白，其可诱导靶DNA序列的转录激活或阻抑(例如，其可以通过本文提供的异源基因效应因子的存在而改变或增强)。术语“基因效应因子(gene effector)”和“基因调节因子(gene modulator)”在本文中可互换使用。术语“基因效应多肽”和“基因调节多肽”在本文中可互换使用。

[0139] 在一些实施方式中，本文公开的工程化核酸酶可以是RNA核酸酶，例如工程化(例如，可编程或可靶向)RNA核酸酶。在一些实施方式中，本文公开的工程化核酸酶可以是核酸

酶无效的RNA结合蛋白,其不诱导靶RNA序列的转录激活或阻抑,除非其与本公开的一种或多种异源基因效应因子以复合体存在。在一些实施方式中,本文公开的工程化核酸酶可以是核酸酶无效的RNA结合蛋白,其可以诱导靶RNA序列的转录激活或阻抑(例如,其可以通过本文提供的异源基因效应因子的存在而改变或增强)。

[0140] 在一些实施方式中,工程化核酸酶可以是核酸引导的靶向系统。在一些实施方式中,工程化核酸酶可以是DNA引导的靶向系统。在一些实施方式中,工程化核酸酶可以是RNA引导的靶向系统。核酸引导的靶向系统可以包括并且利用例如促进CRISPR-Cas系统(例如其核酸酶缺陷形式,如dCas9或dCas14)与靶基因(例如靶内源基因)或靶基因调控序列特异性结合的引导核酸序列。例如,靶基因可以是表1中列出的基因中的任一个,并且靶基因调控序列可以可操作地偶联至表1中列出的基因中的任一个。结合特异性可以通过使用引导核酸(例如单个引导RNA(sgRNA)或其一部分)来确定。在一些实施方式中,使用不同的sgRNA允许本公开的组合物和方法与不同的靶基因(例如,靶内源基因)或靶基因调控序列一起使用(例如,靶向它们)。

[0141] 在一些实施方式中,工程化核酸酶可以与引导核酸(例如引导RNA或其一部分)形成复合体。在一些实施方式中,工程化核酸酶可以与单个引导核酸(例如单个引导RNA(sgRNA))形成复合体。在一些实施方式中,工程化核酸酶可以是RNA结合蛋白(RBP),其任选地与引导核酸(例如引导RNA(例如,sgRNA))复合,所述引导核酸能够与Cas蛋白形成复合体。在一些实施方式中,工程化核酸酶可以是核酸酶无效的DNA结合蛋白,其可以诱导靶DNA序列的转录激活或阻抑。在一些实施方式中,工程化核酸酶可以是源自RNA的核酸酶无效的RNA结合蛋白。

[0142] 本公开的组合物和方法中使用的引导核酸可以为例如至少8个、至少9个、至少10个、至少11个、至少12个、至少13个、至少14个、至少15个、至少16个、至少17个、至少18个、至少19个、至少20个、至少21个、至少22个、至少23个、至少24个、至少25个、至少26个、至少27个、至少28个、至少29个、至少30个、至少31个、至少32个、至少33个、至少34个、至少35个、至少36个、至少37个、至少38个、至少39个或至少40个核苷酸。

[0143] 在一些实施方式中,在本公开的组合物和方法中使用的引导核酸为至多10个、至多11个、至多12个、至多13个、至多14个、至多15个、至多16个、至多17个、至多18个、至多19个、至多20个、至多21个、至多22个、至多23个、至多24个、至多25个、至多26个、至多27个、至多28个、至多29个、至多30个、至多31个、至多32个、至多33个、至多34个、至多35个、至多36个、至多37个、至多38个、至多39个或至多40个核苷酸。

[0144] 在一些实施方式中,本公开的组合物和方法中使用的引导核酸为约8至约40个核苷酸、约10至约40个核苷酸、约11至约40个核苷酸、约12至约40个核苷酸、约13至约40个核苷酸、约14至约40个核苷酸、约15至约40个核苷酸、约16至约40个核苷酸、约17至约40个核苷酸、约18至约40个核苷酸、约19至约40个核苷酸、约20至约40个核苷酸、约22至约40个核苷酸、约24至约40个核苷酸、约26至约40个核苷酸、约28至约40个核苷酸、约30至约40个核苷酸、约8至约30个核苷酸、约10至约30个核苷酸、约11至约30个核苷酸、约12至约30个核苷酸、约13至约30个核苷酸、约14至约30个核苷酸、约15至约30个核苷酸、约16至约30个核苷酸、约17至约30个核苷酸、约18至约30个核苷酸、约19至约30个核苷酸、约20至约30个核苷酸、约22至约30个核苷酸、约24至约30个核苷酸、约26至约30个核苷酸、约28至约30个核苷

酸、约8至约25个核苷酸、约10至约25个核苷酸、约11至约25个核苷酸、约12至约25个核苷酸、约13至约25个核苷酸、约14至约25个核苷酸、约15至约25个核苷酸、约16至约25个核苷酸、约17至约25个核苷酸、约18至约25个核苷酸、约19至约25个核苷酸、约20至约25个核苷酸、约22至约25个核苷酸、约24至约25个核苷酸、约8至约20个核苷酸、约10至约20个核苷酸、约11至约20个核苷酸、约12至约20个核苷酸、约13至约20个核苷酸、约14至约20个核苷酸、约15至约20个核苷酸、约16至约20个核苷酸、约17至约20个核苷酸、约18至约20个核苷酸、约19至约20个核苷酸、约8至约18个核苷酸、约10至约18个核苷酸、约11至约18个核苷酸、约12至约18个核苷酸、约13至约18个核苷酸、约14至约18个核苷酸、约15至约18个核苷酸、约16至约18个核苷酸、约8至约16个核苷酸、约10至约16个核苷酸、约11至约16个核苷酸、约12至约16个核苷酸、约13至约16个核苷酸、约14至约16个核苷酸、或约15至约16个核苷酸。在一些实施方式中，引导核酸可以是引导RNA或其一部分。

[0145] 可以对本文公开的工程化核酸酶进行修饰，以增强本公开的组合物和方法(例如，作为本文公开的复合体的一部分)对基因表达的调控。可以对工程化核酸酶进行修饰，以提高或降低核酸结合亲和力、核酸结合特异性、酶活性和/或与其他因子(例如异二聚化或寡聚化结构域)的结合和诱导配体。还可以对工程化核酸酶进行修饰，以改变蛋白质的任何其他活性或性质，例如稳定性。例如，可以对工程化核酸酶的一个或多个核酸酶结构域进行修饰、缺失或失活，或者可以截短工程化核酸酶的至少一部分，以去除对蛋白质或复合体的期望功能不重要的结构域。可以对工程化核酸酶进行修饰，以调节(例如，增强或降低)该工程化核酸酶通过包含异源基因效应因子的本公开的复合体调控基因表达的活性。

[0146] 例如，可将工程化核酸酶偶联(例如，融合、共价偶联或非共价偶联)至异源基因效应因子(例如，表观遗传修饰结构域、转录激活结构域和/或转录抑制结构域)。工程化核酸酶可以偶联(例如，融合、共价偶联或非共价偶联)至本文公开的寡聚化或二聚化结构域(例如，异二聚化结构域)。工程化核酸酶可以偶联(例如，融合、共价偶联或非共价偶联)至提供提高或降低的稳定性的异源多肽。工程化核酸酶可以偶联(例如，融合、共价偶联或非共价偶联)至可以促进工程化核酸酶或包含工程化核酸酶的复合体的降解的序列。工程化核酸酶可以偶联(例如融合、共价偶联或非共价偶联)至编辑、突变或修饰(直接或间接)靶多核苷酸序列的基因编辑部分(例如异源蛋白或其结构域或功能片段)。

[0147] 工程化核酸酶可以偶联(例如融合、共价偶联或非共价偶联)至任何合适数量的伴侣，例如至少一个、至少两个、至少三个、至少四个、或至少五个、至少六个、至少七个或至少八个伴侣。在一些实施方式中，本公开的工程化核酸酶偶联(例如融合、共价偶联或非共价偶联)至至多两个、至多三个、至多四个、至多五个、至多六个、至多七个、至多八个或至多十个伴侣。在一些实施方式中，本公开的工程化核酸酶偶联(例如融合、共价偶联或非共价偶联)至1-5、1-4、1-3、1-2、2-5、2-4、2-3、3-5、3-4或4-5个伴侣。在一些实施方式中，本公开的工程化核酸酶偶联(例如融合、共价偶联或非共价偶联)至一个伴侣。在一些实施方式中，本公开的工程化核酸酶偶联(例如融合、共价偶联或非共价偶联)至两个伴侣。在一些实施方式中，本公开的工程化核酸酶偶联(例如融合、共价偶联或非共价偶联)至三个伴侣。在一些实施方式中，本公开的工程化核酸酶偶联(例如融合、共价偶联或非共价偶联)至四个伴侣。在一些实施方式中，本公开的工程化核酸酶偶联(例如融合、共价偶联或非共价偶联)至五个伴侣。在一些实施方式中，本公开的工程化核酸酶偶联(例如融合、共价偶联或非共价偶

联)至六个伴侣。

[0148] 本文公开的工程化核酸酶可以是融合蛋白,例如,包含工程化核酸酶和本文公开的伴侣中的一个或多个的融合物。融合结构域或异源多肽可以位于工程化核酸酶的N端、C端或内部。

[0149] 工程化核酸酶的伴侣(例如,共价或非共价地偶联至本文公开的工程化核酸酶的核酸酶缺陷或无效变体)可以是转录效应因子(例如,转录激活因子或转录阻抑因子)。转录效应因子对于本文提供的细胞可以是异源的。

[0150] 在一些实施方式中,转录效应因子可以是组蛋白表观遗传修饰剂(或组蛋白修饰剂)。在一些情况下,组蛋白表观遗传修饰剂可以通过甲基化(例如,组蛋白甲基化修饰剂,如氨基酸甲基转移酶,例如KRAB)调节组蛋白。在一些情况下,组蛋白表观遗传修饰剂可以通过乙酰化调节组蛋白。在一些情况下,组蛋白表观遗传修饰剂可以通过磷酸化调节组蛋白。在一些情况下,组蛋白表观遗传修饰剂可以通过ADP-核糖基化调节组蛋白。在一些情况下,组蛋白表观遗传修饰剂可以通过糖基化调节组蛋白。在一些情况下,组蛋白表观遗传修饰剂可以通过SUMO化调节组蛋白。在一些情况下,组蛋白表观遗传修饰剂可以通过泛素化调节组蛋白。在一些情况下,组蛋白表观遗传修饰剂可以通过重塑组蛋白结构(例如通过ATP水解依赖性过程)调节组蛋白。

[0151] 在一些实施方式中,转录效应因子可以是基因表观遗传修饰剂(或基因修饰剂)。在一些情况下,基因修饰剂可以通过甲基化(例如,基因甲基化修饰剂,如DNA甲基转移酶或DNMT)调节基因。在一些情况下,基因修饰剂可以通过乙酰化调节基因。

[0152] 在一些实施方式中,转录效应因子来自相关组蛋白乙酰转移酶家族。组蛋白乙酰转移酶的非限制性实例包括GNAT亚家族、MYST亚家族、p300/CBP亚家族、HAT1亚家族、GCN5、PCAF、Tip60、MOZ、MORF、MOF、HBO1、p300、CBP、HAT1、ATF-2、SRC1和TAFII250。

[0153] 在一些实施方式中,转录效应因子可以包括表观遗传修饰剂。在一些实施方式中,转录效应因子包括组蛋白表观遗传修饰剂(例如,组蛋白赖氨酸甲基转移酶、组蛋白赖氨酸去甲基化酶或DNA甲基化酶)。表观遗传修饰剂的非限制性实例可以包括EZH亚家族、非SET亚家族、其他SET亚家族、PRDM亚家族、SET1亚家族、SET2亚家族、SUV39亚家族、SYMD亚家族、ASH1L、EHMT1、EHMT2、EZH1、EZH2、MLL、MLL2、MLL3、MLL4、MLL5、NSD1、NSD2、NSD3、PRDM1、PRDM10、PRDM11、PRDM12、PRDM13、PRDM14、PRDM15、PRDM16、PRDM2、PRDM4、PRDM5、PRDM6、PRDM7、PRDM8、PRDM9、SET1、SET1L、SET2L、SETD2、SETD3、SETD4、SETD5、SETD6、SETD7、SETD8、SETDB1、SETDB2、SETMAR、SUV39H1、SUV39H2、SUV420H1、SUV420H2、SYMD1、SYMD2、SYMD3、SYMD4和SYMD5。

[0154] 可以用作融合伴侣以提高转录的蛋白质(或其片段)的实例包括但不限于:转录激活因子,例如VP16、VP64、VP48、VP160、p65亚结构域(例如来自NFkB),以及EDLL的激活结构域和/或TAL激活结构域(例如用于植物中的活性)、SET1A、SET1B、MLL1至5、ASH1、SYMD2、NSD1、JHDM2a/b、UTX、JMJD3、GCN5、PCAF、CBP、p300、TAF1、TIP60/PLIP、MOZMYST3、MORFMYST4、SRC1、ACTR、PI 60、CLOCK、10-11易位(TET)双加氧酶1(TET1CD)、TET1、DME、DML1、DML2、ROS1等。此类基因激活调节因子的另一实例是VP64-p65-Rta融合多肽(VPR)。

可以用作融合伴侣以降低转录的蛋白质(或其片段)的实例包括但不限于:转录阻抑因子,例如Kruppel相关盒(KRAB或SKD);KOX1阻抑结构域;Mad mSIN3相互作用结构域

(SID);ERF阻抑因子结构域(ERD)、SRDX阻抑结构域(例如用于植物中的阻抑)等;组蛋白赖氨酸甲基转移酶,例如Pr-SET7/8、SUV4-20H1、RIZ1等;组蛋白赖氨酸去甲基化酶,例如JMJD2A/JHDM3A、JMJD2B、JMJD2C/GASC1、JMJD2D、JARJD 1A/RBP2、JARID1B/PLU-1、JARID 1C/SMCX、JARIDID/SMCY等;组蛋白赖氨酸去乙酰化酶,例如HDAC1、HDAC2、HDAC3、HDAC8、HDAC4、HDAC5、HDAC7、HDAC9、SIRT1、SIRT2、HDAC11等;DNA甲基化酶,例如Hhal DNA m5c-甲基转移酶(M.Hhal)、DNA甲基转移酶1(DNMT1)、DNA甲基转移酶3a(DNMT3a)、DNA甲基转移酶3b(DNMT3b)、METI、DRM3(植物)、ZMET2、CMT1、CMT2(植物)等;以及外周募集元件,例如核纤层蛋白A、核纤层蛋白B等。

[0155] 在多个方面,本文提供的工程化核酸酶可实现靶多核苷酸的编辑或突变,如本文所述。在一些实施方式中,编辑或突变靶多核苷酸序列涉及将靶多核苷酸中的一个或多个核苷酸改变为一个或多个不同的核苷酸。在一些实施方式中,编辑或突变靶多核苷酸涉及将鸟嘌呤(G)改变为不同的核苷酸。在一些情况下,鸟嘌呤(G)可以改变为腺嘌呤(A)。在一些情况下,鸟嘌呤(G)可以改变为胸腺嘧啶(T)。在一些情况下,鸟嘌呤(G)可以改变为胞嘧啶(C)。在一些情况下,鸟嘌呤(G)可以改变为尿嘧啶(U)。在一些情况下,鸟嘌呤(G)可以改变为肌苷(I)。在一些实施方式中,编辑或突变靶多核苷酸涉及将胞嘧啶(C)改变为不同的核苷酸。在一些情况下,胞嘧啶(C)可以改变为鸟嘌呤(G)。在一些情况下,胞嘧啶(C)可以改变为腺嘌呤(A)。在一些情况下,胞嘧啶(C)可以改变为胸腺嘧啶(T)。在一些情况下,胞嘧啶(C)可以改变为尿嘧啶(U)。在一些情况下,胞嘧啶(C)可以改变为肌苷(I)。在一些实施方式中,编辑或突变靶多核苷酸涉及将胸腺嘧啶(T)改变为不同的核苷酸。在一些情况下,胸腺嘧啶(T)可以改变为胞嘧啶(C)。在一些情况下,胸腺嘧啶(T)可以改变为鸟嘌呤(G)。在一些情况下,胸腺嘧啶(T)可以改变为腺嘌呤(A)。在一些情况下,胸腺嘧啶(T)可以改变为尿嘧啶(U)。在一些情况下,胸腺嘧啶(T)可以改变为肌苷(I)。在一些实施方式中,编辑或突变靶多核苷酸涉及将腺嘌呤(A)改变为不同的核苷酸。在一些情况下,腺嘌呤(A)可以改变为鸟嘌呤(G)。在一些情况下,腺嘌呤(A)可以改变为胞嘧啶(C)。在一些情况下,腺嘌呤(A)可以改变为胸腺嘧啶(T)。在一些情况下,腺嘌呤(A)可以改变为尿嘧啶(U)。在一些情况下,腺嘌呤(A)可以改变为肌苷(I)。在一些实施方式中,编辑或突变靶多核苷酸涉及将尿嘧啶(U)改变为不同的核苷酸。在一些情况下,尿嘧啶(U)可以改变为鸟嘌呤(G)。在一些情况下,尿嘧啶(U)可以改变为胞嘧啶(C)。在一些情况下,尿嘧啶(U)可以改变为胸腺嘧啶(T)。在一些情况下,尿嘧啶(U)可以改变为腺嘌呤(A)。在一些情况下,尿嘧啶(U)可以改变为肌苷(I)。在一些实施方式中,编辑或突变靶多核苷酸涉及将肌苷(I)改变为不同的核苷酸。在一些情况下,肌苷(I)可以改变为鸟嘌呤(G)。在一些情况下,肌苷(I)可以改变为胞嘧啶(C)。在一些情况下,肌苷(I)可以改变为胸腺嘧啶(T)。在一些情况下,肌苷(I)可以改变为腺嘌呤(A)。在一些情况下,肌苷(I)可以改变为尿嘧啶(U)。

[0156] 在一些实施方式中,编辑或突变靶多核苷酸涉及将一个或多个点突变引入靶多核苷酸。在一些实施方式中,编辑或突变靶多核苷酸涉及将一个或多个缺失(例如,一个或多个核苷酸的缺失)引入靶多核苷酸。在一些实施方式中,编辑或突变靶多核苷酸涉及将一个或多个插入(例如,一个或多个核苷酸的插入)引入靶多核苷酸。在一些实施方式中,编辑或突变靶多核苷酸涉及将一个或多个倒置(例如,两个或更多个核苷酸的倒置)引入靶多核苷酸。在一些实施方式中,编辑或突变靶多核苷酸涉及将一个或多个易位(例如,一个或多个

核苷酸的易位)引入靶多核苷酸。在一些实施方式中,编辑或突变靶多核苷酸涉及将一个或多个转座引入靶多核苷酸。

[0157] 在一些情况下,本文所述的工程化核酸酶可以偶联至影响靶多核苷酸编辑或突变的伴侣(例如,基因编辑部分),如本文所述。在一些情况下,本文提供的核酸酶缺陷或核酸酶无效的工程化核酸酶共价或非共价地偶联至影响靶多核苷酸编辑或突变的基因编辑部分(例如,蛋白质,或其功能结构域或功能片段)。在一些实施方式中,影响靶多核苷酸编辑或突变的基因编辑部分(例如,蛋白质,或其功能结构域或功能片段)是将一个或多个核苷酸改变为不同核苷酸的基因编辑部分。在一些实施方式中,影响靶多核苷酸编辑或突变的基因编辑部分(例如,蛋白质,或其功能结构域或功能片段)是将鸟嘌呤(G)改变为不同核苷酸的基因编辑部分。在一些情况下,影响靶多核苷酸编辑或突变的基因编辑部分将鸟嘌呤(G)改变为胞嘧啶(C)。在一些情况下,影响靶多核苷酸编辑或突变的基因编辑部分将鸟嘌呤(G)改变为胸腺嘧啶(T)。在一些情况下,影响靶多核苷酸编辑或突变的基因编辑部分将鸟嘌呤(G)改变为腺嘌呤(A)。在一些情况下,影响靶多核苷酸编辑或突变的基因编辑部分将鸟嘌呤(G)改变为尿嘧啶(U)。在一些情况下,影响靶多核苷酸编辑或突变的基因编辑部分将鸟嘌呤(G)改变为肌苷(I)。在一些实施方式中,影响靶多核苷酸序列编辑或突变的基因编辑部分(例如,蛋白质,或其功能结构域或功能片段)是将胞嘧啶(C)改变为不同核苷酸的基因编辑部分。在一些情况下,影响靶多核苷酸编辑或突变的基因编辑部分将胞嘧啶(C)改变为鸟嘌呤(G)。在一些情况下,影响靶多核苷酸编辑或突变的基因编辑部分将胞嘧啶(C)改变为胸腺嘧啶(T)。在一些情况下,影响靶多核苷酸编辑或突变的基因编辑部分将胞嘧啶(C)改变为腺嘌呤(A)。在一些情况下,影响靶多核苷酸编辑或突变的基因编辑部分将胞嘧啶(C)改变为尿嘧啶(U)。在一些情况下,影响靶多核苷酸编辑或突变的基因编辑部分将胞嘧啶(C)改变为肌苷(I)。在一些实施方式中,影响靶多核苷酸序列编辑或突变的基因编辑部分(例如,蛋白质,或其功能结构域或功能片段)是将胸腺嘧啶(T)改变为不同核苷酸的基因编辑部分。在一些情况下,影响靶多核苷酸编辑或突变的基因编辑部分将胸腺嘧啶(T)改变为胞嘧啶(C)。在一些情况下,影响靶多核苷酸编辑或突变的基因编辑部分将胸腺嘧啶(T)改变为鸟嘌呤(G)。在一些情况下,影响靶多核苷酸编辑或突变的基因编辑部分将胸腺嘧啶(T)改变为腺嘌呤(A)。在一些情况下,影响靶多核苷酸编辑或突变的基因编辑部分将胸腺嘧啶(T)改变为尿嘧啶(U)。在一些情况下,影响靶多核苷酸编辑或突变的基因编辑部分将胸腺嘧啶(T)改变为肌苷(I)。在一些实施方式中,影响靶多核苷酸序列编辑或突变的基因编辑部分(例如,蛋白质,或其功能结构域或功能片段)是将腺嘌呤(A)改变为不同核苷酸的基因编辑部分。在一些情况下,影响靶多核苷酸编辑或突变的基因编辑部分将腺嘌呤(A)改变为胞嘧啶(C)。在一些情况下,影响靶多核苷酸编辑或突变的基因编辑部分将腺嘌呤(A)改变为胸腺嘧啶(T)。在一些情况下,影响靶多核苷酸编辑或突变的基因编辑部分将腺嘌呤(A)改变为鸟嘌呤(G)。在一些情况下,影响靶多核苷酸编辑或突变的基因编辑部分将腺嘌呤(A)改变为尿嘧啶(U)。在一些情况下,影响靶多核苷酸编辑或突变的基因编辑部分将腺嘌呤(A)改变为肌苷(I)。在一些实施方式中,影响靶多核苷酸序列编辑或突变的基因编辑部分(例如,蛋白质,或其功能结构域或功能片段)是将尿嘧啶(U)改变为不同核苷酸的基因编辑部分。在一些情况下,影响靶多核苷酸编辑或突变的基因编辑部分将尿嘧啶(U)改变为胞嘧啶(C)。在一些情况下,影响靶多核苷酸编辑或突变的基因编辑部分将尿

嘧啶(U)改变为胸腺嘧啶(T)。在一些情况下,影响靶多核苷酸编辑或突变的基因编辑部分将尿嘧啶(U)改变为腺嘌呤(A)。在一些情况下,影响靶多核苷酸编辑或突变的基因编辑部分将尿嘧啶(U)改变为鸟嘌呤(G)。在一些情况下,影响靶多核苷酸编辑或突变的基因编辑部分将尿嘧啶(U)改变为肌苷(I)。在一些实施方式中,影响靶多核苷酸序列编辑或突变的基因编辑部分(例如,蛋白质,或其功能结构域或功能片段)是将肌苷(I)改变为不同核苷酸的基因编辑部分。在一些情况下,影响靶多核苷酸编辑或突变的基因编辑部分将肌苷(I)改变为胞嘧啶(C)。在一些情况下,影响靶多核苷酸编辑或突变的基因编辑部分将肌苷(I)改变为胸腺嘧啶(T)。在一些情况下,影响靶多核苷酸编辑或突变的基因编辑部分将肌苷(I)改变为腺嘌呤(A)。在一些情况下,影响靶多核苷酸编辑或突变的基因编辑部分将肌苷(I)改变为尿嘧啶(U)。在一些情况下,影响靶多核苷酸编辑或突变的基因编辑部分将肌苷(I)改变为鸟嘌呤(G)。

[0158] 在一些实施方式中,本文所述的工程化核酸酶(例如,本文所述的核酸酶缺陷或核酸酶无效的工程化核酸酶)可以偶联(例如,共价或非共价)至将一个或多个点突变引入靶多核苷酸的伴侣(例如,基因编辑部分)。在一些实施方式中,本文所述的工程化核酸酶(例如,本文所述的核酸酶缺陷或核酸酶无效的工程化核酸酶)可以偶联(例如,共价或非共价)至将一个或多个缺失(例如,一个或多个核苷酸的缺失)引入靶多核苷酸的伴侣(例如,基因编辑部分)。在一些实施方式中,本文所述的工程化核酸酶(例如,本文所述的核酸酶缺陷或核酸酶无效的工程化核酸酶)可以偶联(例如,共价或非共价)至将一个或多个插入(例如,一个或多个核苷酸的插入)引入靶多核苷酸的伴侣(例如,基因编辑部分)。在一些实施方式中,本文所述的工程化核酸酶(例如,本文所述的核酸酶缺陷或核酸酶无效的工程化核酸酶)可以偶联(例如,共价或非共价)至将一个或多个倒置(例如,两个或更多个核苷酸的倒置)引入靶多核苷酸的伴侣(例如,基因编辑部分)。在一些实施方式中,本文所述的工程化核酸酶(例如,本文所述的核酸酶缺陷或核酸酶无效的工程化核酸酶)可以偶联(例如,共价或非共价)至将一个或多个易位(例如,一个或多个核苷酸的易位)引入靶多核苷酸的伴侣(例如,基因编辑部分)。在一些实施方式中,本文所述的工程化核酸酶(例如,本文所述的核酸酶缺陷或核酸酶无效的工程化核酸酶)可以偶联(例如,共价或非共价)至将一个或多个转座引入靶多核苷酸的伴侣(例如,基因编辑部分)。

[0159] 在一些实施方式中,基因编辑部分可以是碱基编辑蛋白或碱基编辑酶。在一些实施方式中,碱基编辑蛋白或碱基编辑酶是脱氨酶。在一些情况下,脱氨酶是胞苷脱氨酶。在一些情况下,胞苷脱氨酶催化胞嘧啶(C)向尿嘧啶(U)的反应,尿嘧啶(U)具有胸腺嘧啶的碱基配对特性。在一些实施方式中,例如当多核苷酸是双链(例如,双链DNA)时,尿苷碱基随后可以被胸苷碱基置换(例如,通过细胞修复机制)以产生C·G到T·A的转变。在一些实施方式中,脱氨酶是腺嘌呤脱氨酶。在一些情况下,腺嘌呤脱氨酶催化腺苷(A)向肌苷(I)的反应。适用于本文的脱氨酶的非限制性实例包括但不限于APOBEC 1脱氨酶、APOBEC2脱氨酶、APOBEC3脱氨酶、APOBEC3A脱氨酶、APOBEC3B脱氨酶、APOBEC3C脱氨酶、APOBEC3D脱氨酶、APOBEC3E脱氨酶、APOBEC3F脱氨酶、APOBEC3G脱氨酶、APOBEC3H脱氨酶、APOBEC4脱氨酶、活化诱导的胞苷脱氨酶(AID)、腺苷脱氨酶1(ADAR1)、腺苷脱氨酶2(ADAR2)、腺苷脱氨酶3(ADAR3)或TadA。

[0160] 在一些实施方式中,本文所述的工程化核酸酶可用于先导编辑(prime editing)。

例如,本文所述的工程化核酸酶可以偶联至逆转录酶(例如,工程化M-MLV逆转录酶)和先导编辑RNA(pegRNA)。在这种情况下,工程化核酸酶可以包含切口酶活性。在一些实施方式中,先导编辑可用于介导靶向插入、缺失或碱基到碱基的转换。

[0161] 在一些实施方式中,本文所述的工程化核酸酶可用于基因写入。

[0162] 在一些实施方式中,本文所述的工程化核酸酶可以偶联至使靶底物甲基化的伴侣(例如,蛋白质,或其功能结构域或功能片段)。在一些情况下,使靶底物甲基化的伴侣(例如,蛋白质,或其功能结构域或功能片段)是甲基转移酶。在一些情况下,甲基转移酶是DNA甲基转移酶、组蛋白甲基转移酶或RNA甲基转移酶。在一些情况下,DNA甲基转移酶是DNMT1或DNMT3。

[0163] 在一些实施方式中,本文所述的工程化核酸酶可以偶联至具有去甲基化酶活性(例如可以从核酸、蛋白质或其他分子中去除甲基基团)的伴侣(例如蛋白质,或其功能结构域或功能片段)。在一些情况下,具有去甲基化酶活性的伴侣(例如蛋白质,或其功能结构域或功能片段)是组蛋白赖氨酸去甲基化酶,例如但不限于KDM1、KDM2、KDM3、KDM4、KDM5和KDM6。

[0164] 在一些实施方式中,本文所述的工程化核酸酶可以偶联至具有歧化酶活性的伴侣(例如,蛋白质,或其功能结构域或功能片段)。在一些情况下,具有歧化酶活性的伴侣(例如,蛋白质,或其功能结构域或功能片段)是超氧化物歧化酶、甲醛歧化酶或亚氯酸盐歧化酶。

[0165] 在一些实施方式中,本文所述的工程化核酸酶可以偶联至具有烷基化活性的伴侣(例如,蛋白质,或其功能结构域或功能片段)。在一些情况下,具有烷基化活性的伴侣(例如,蛋白质,或其功能结构域或功能片段)是异戊烯基转移酶、萜烯环化酶或萜烯合酶。

[0166] 在一些实施方式中,本文所述的工程化核酸酶可以偶联至具有脱嘌呤活性的伴侣(例如,蛋白质,或其功能结构域或功能片段)。在一些情况下,具有脱嘌呤活性的伴侣(例如,蛋白质,或其功能结构域或功能片段)是DNA糖基化酶。

[0167] 在一些实施方式中,本文所述的工程化核酸酶可以偶联至具有氧化活性的伴侣(例如,蛋白质,或其功能结构域或功能片段)。在一些情况下,具有氧化活性的伴侣(例如,蛋白质,或其功能结构域或功能片段)是过氧化物酶或氧化酶。

[0168] 在一些实施方式中,本文所述的工程化核酸酶可以偶联至具有嘧啶二聚体形成活性的伴侣(例如,蛋白质,或其功能结构域或功能片段)。

[0169] 在一些实施方式中,本文所述的工程化核酸酶可以偶联至具有整合酶活性的伴侣(例如,蛋白质,或其功能结构域或功能片段)。在一些情况下,具有整合酶活性的伴侣(例如,蛋白质,或其功能结构域或功能片段)是逆转录病毒整合酶或HIV整合酶。

[0170] 在一些实施方式中,本文所述的工程化核酸酶可以偶联至具有转座酶活性的伴侣(例如,蛋白质,或其功能结构域或功能片段)。在一些情况下,具有转座酶活性的伴侣(例如,蛋白质,或其功能结构域或功能片段)是ty1、Mariner转座酶、Tn3、转座酶(Tnp)Tn5或Tn7转座子。

[0171] 在一些实施方式中,本文所述的工程化核酸酶可以偶联至具有重组酶活性的伴侣(例如,蛋白质,或其功能结构域或功能片段)。在一些情况下,具有重组酶活性的伴侣(例如,蛋白质,或其功能结构域或功能片段)是酪氨酸重组酶、Rad51重组酶、RecA重组酶或

Dmc1重组酶。

[0172] 在一些实施方式中,本文所述的工程化核酸酶可以偶联至具有聚合酶活性的伴侣(例如,蛋白质,或其功能结构域或功能片段)。在一些情况下,具有聚合酶活性的伴侣(例如,蛋白质,或其功能结构域或功能片段)是DNA聚合酶、RNA聚合酶、逆转录酶或RdRp复制酶。

[0173] 在一些实施方式中,本文所述的工程化核酸酶可以偶联至具有连接酶活性的伴侣(例如,蛋白质,或其功能结构域或功能片段)。在一些情况下,具有连接酶活性的伴侣(例如,蛋白质,或其功能结构域或功能片段)是泛素连接酶、谷氨酸-半胱氨酸连接酶、氨酰基tRNA合成酶、琥珀酰辅酶A合成酶、乙酰CoA合成酶、丙酮酸羧化酶、乙酰CoA羧化酶、丙酰CoA羧化酶、甲基巴豆酰CoA羧化酶、DNA连接酶、镁螯合酶、钴螯合酶或DNA合成酶。

[0174] 在一些实施方式中,本文所述的工程化核酸酶可以偶联至具有解旋酶活性的伴侣(例如,蛋白质,或其功能结构域或功能片段)。在一些情况下,具有解旋酶活性的伴侣(例如,蛋白质,或其功能结构域或功能片段)是DNA解旋酶、RNA解旋酶、染色质域(chromodomain)解旋酶、或DEAD box/DEAD/DEAH box解旋酶。

[0175] 在一些实施方式中,本文所述的工程化核酸酶可以偶联至具有光解酶活性的伴侣(例如,蛋白质,或其功能结构域或功能片段)。在一些情况下,具有光解酶活性的伴侣(例如,蛋白质,或其功能结构域或功能片段)是光复活酶、DNA光解酶、DNA光复活酶、DNA环丁烷二嘧啶光解酶、DNA光解酶、脱氧核糖核酸光解酶、脱氧核糖二嘧啶光解酶、光解酶、PRE、PhrB光解酶、脱氧核糖核酸环丁烷二嘧啶光解酶、phr A光解酶、二嘧啶光解酶(光敏感)或脱氧核糖核酸嘧啶二聚体裂解酶(光敏感)。

[0176] 在一些实施方式中,本文所述的工程化核酸酶可以偶联至具有糖基化酶活性的伴侣(例如,蛋白质,或其功能结构域或功能片段)。在一些情况下,具有糖基化酶活性的伴侣(例如,蛋白质,或其功能结构域或功能片段)是N-甲基嘌呤DNA糖基化酶、UNG、hOGG1、hNTH1、hNEIL1、hMYH、hSMUG1、TDG、MBD4、Mag1、Ung1、Ogg1、Ntg1、AlkE、Ntg2、hNEIL2、hNEIL3、AlkC、AlkD、MutY、Nei、Nth、Fpg或UDG。

[0177] 在一些实施方式中,本文所述的工程化核酸酶可以偶联至具有乙酰转移酶活性的伴侣(例如,蛋白质,或其功能结构域或功能片段)。在一些情况下,具有乙酰转移酶活性的伴侣(例如,蛋白质,或其功能结构域或功能片段)是CBP组蛋白乙酰转移酶、胆碱乙酰转移酶、氯霉素乙酰转移酶、血清素N-乙酰转移酶、NatA乙酰转移酶或NatB乙酰转移酶。

[0178] 在一些实施方式中,本文所述的工程化核酸酶可以偶联至具有脱乙酰酶活性的伴侣(例如,蛋白质,或其功能结构域或功能片段)。在一些情况下,具有脱乙酰酶活性的伴侣(例如,蛋白质,或其功能结构域或功能片段)是HDAC-1、HDAC-2、HDAC-3、HDAC-4、HDAC-5、HDAC-6、HDAC-7或HDAC-8。

[0179] 在一些实施方式中,本文所述的工程化核酸酶可以偶联至具有激酶活性的伴侣(例如,蛋白质,或其功能结构域或功能片段)。在一些情况下,具有激酶活性的伴侣(例如,蛋白质,或其功能结构域或功能片段)是Ca2+/钙调蛋白依赖性蛋白激酶、细胞周期蛋白依赖性激酶、核苷二磷酸激酶、磷脂酰肌醇磷酸激酶、胸苷激酶、胸苷酸激酶或壁相关激酶。

[0180] 在一些实施方式中,本文所述的工程化核酸酶可以偶联至具有磷酸酶活性的伴侣(例如,蛋白质,或其功能结构域或功能片段)。在一些情况下,具有磷酸酶活性的伴侣(例

如,蛋白质,或其功能结构域或功能片段)是酸性磷酸酶、碱性磷酸酶、核酸内切酶/核酸外切酶/磷酸酶家族、激酶、phosphatome、磷酸转移酶、蛋白磷酸酶或蛋白磷酸酶2。

[0181] 在一些实施方式中,本文所述的工程化核酸酶可以偶联至具有泛素连接酶活性的伴侣(例如,蛋白质,或其功能结构域或功能片段)。在一些情况下,具有泛素连接酶活性的伴侣(例如,蛋白质,或其功能结构域或功能片段)是E3A、mdm2、后期促进复合体(APC)、UBR5(EDD1)、SOCS/BC-box/eloBC/CUL5/RING、LNXp80、CBX4、CBLL1、HACE1、HECTD1、HECTD2、HECTD3、HECTD4、HECW1、HECW2、HERC1、HERC2、HERC3、HERC4、HERC5、HERC6、HUWE1、ITCH、NEDD4、NEDD4L、PPIL2、PRPF19、PIAS1、PIAS2、PIAS3、PIAS4、RANBP2、RNF4、RBX1、SMURF1、SMURF2、STUB1、TOPORS、TRIP12、UBE3A、UBE3B、UBE3C、UBE3D、UBE4A、UBE4B、UBOX5、UBR5、VHL、WWP1、WWP2、Parkin或MKRN1。

[0182] 在一些实施方式中,本文所述的工程化核酸酶可以偶联至具有去泛素化活性的伴侣(例如,蛋白质,或其功能结构域或功能片段)。在一些情况下,具有去泛素化活性的伴侣(例如,蛋白质,或其功能结构域或功能片段)是去泛素化肽酶、去泛素化异肽酶、去泛素化酶、泛素蛋白酶、泛素水解酶或泛素异肽酶。

[0183] 在一些实施方式中,本文所述的工程化核酸酶可以偶联至具有腺苷酰化活性的伴侣(例如,蛋白质,或其功能结构域或功能片段)。在一些情况下,具有腺苷酰化活性的伴侣(例如,蛋白质,或其功能结构域或功能片段)是羧酸还原酶。

[0184] 在一些实施方式中,本文所述的工程化核酸酶可以偶联至具有脱腺苷酸化活性的伴侣(例如,蛋白质,或其功能结构域或功能片段)。在一些情况下,具有脱腺苷酸化活性的伴侣(例如,蛋白质,或其功能结构域或功能片段)是5'-脱腺苷酸酶、CNOT6脱腺苷酸酶、CNOT6L脱腺苷酸酶或CCR4-NOT脱腺苷酸酶。

[0185] 在一些实施方式中,本文所述的工程化核酸酶可以偶联至具有SUMO化活性的伴侣(例如,蛋白质,或其功能结构域或功能片段)。在一些情况下,具有SUMO化活性的伴侣(例如,蛋白质,或其功能结构域或功能片段)是小泛素相关修饰物(SUMO-1)、SUMO-2或SUMO-3。

[0186] 在一些实施方式中,本文所述的工程化核酸酶可以偶联至具有脱SUMO化活性的伴侣(例如,蛋白质,或其功能结构域或功能片段)。在一些情况下,具有脱SUMO化活性的伴侣(例如,蛋白质,或其功能结构域或功能片段)是SENP1、SENP2、SENP3或SENP5。

[0187] 在一些实施方式中,本文所述的工程化核酸酶可以偶联至具有核糖基化活性的伴侣(例如,蛋白质,或其功能结构域或功能片段)。在一些情况下,具有核糖基化活性的伴侣(例如,蛋白质,或其功能结构域或功能片段)是单(ADP-核糖基)转移酶、聚(ADP-核糖)聚合酶或组蛋白核糖基化酶。

[0188] 在一些实施方式中,本文所述的工程化核酸酶可以偶联至具有脱核糖基化活性的伴侣(例如,蛋白质,或其功能结构域或功能片段)。在一些情况下,具有脱核糖基化活性的伴侣(例如,蛋白质,或其功能结构域或功能片段)是组蛋白赖氨酸脱核糖基化酶或ADP-核糖脱核糖基化酶。

[0189] 在一些实施方式中,本文所述的工程化核酸酶可以偶联至具有肉豆蔻酰化活性的伴侣(例如,蛋白质,或其功能结构域或功能片段)。在一些情况下,具有肉豆蔻酰化活性的伴侣(例如,蛋白质,或其功能结构域或功能片段)是N-肉豆蔻酰基转移酶(NMT)1、N-肉豆蔻酰基转移酶(NMT)2或甘氨酰肽N-十四烷酰基转移酶。

[0190] 在一些实施方式中,本文所述的工程化核酸酶可以偶联至具有重塑活性的伴侣(例如,蛋白质,或其功能结构域或功能片段)。在一些情况下,具有重塑活性的伴侣(例如,蛋白质,或其功能结构域或功能片段)是组蛋白乙酰转移酶(HAT)、脱乙酰酶或甲基转移酶。

[0191] 在一些实施方式中,本文所述的工程化核酸酶可以偶联至具有蛋白酶活性的伴侣(例如,蛋白质,或其功能结构域或功能片段)。在一些情况下,具有蛋白酶活性的伴侣(例如,蛋白质,或其功能结构域或功能片段)是胰蛋白酶、糜蛋白酶、弹性蛋白酶、木瓜蛋白酶、菠萝蛋白酶、丝氨酸蛋白酶、半胱氨酸蛋白酶、苏氨酸蛋白酶、天冬氨酸蛋白酶、谷氨酸蛋白酶、金属蛋白酶或天冬酰胺肽裂解酶。

[0192] 在一些实施方式中,本文所述的工程化核酸酶可以偶联至具有氧化还原酶活性的伴侣(例如,蛋白质,或其功能结构域或功能片段)。在一些情况下,具有氧化还原酶活性的伴侣(例如,蛋白质,或其功能结构域或功能片段)是供体脱氢酶、过氧化物酶、还原酶、脱氢酶、氧化酶、加氧酶、羟化酶、荧光素酶、DMSO还原酶、葡萄糖氧化酶、L-古洛糖酸内酯氧化酶、硫胺素氧化酶、黄嘌呤氧化酶、乙醛脱氢酶、丙酮酸脱氢酶、酮戊二酸脱氢酶、单胺氧化酶、胆绿素还原酶、二氢叶酸还原酶、亚甲基四氢叶酸还原酶、肌氨酸氧化酶或二氢苯并菲啶氧化酶。

[0193] 在一些实施方式中,本文所述的工程化核酸酶可以偶联至具有转移酶活性的伴侣(例如,蛋白质,或其功能结构域或功能片段)。在一些情况下,具有转移酶活性的伴侣(例如,蛋白质,或其功能结构域或功能片段)是辅酶A转移酶、酰基转移酶、肽基转移酶、N-乙酰转移酶或丙酮酸脱氢酶。

[0194] 在一些实施方式中,本文所述的工程化核酸酶可以偶联至具有水解酶活性的伴侣(例如,蛋白质,或其功能结构域或功能片段)。在一些情况下,具有水解酶活性的伴侣(例如,蛋白质,或其功能结构域或功能片段)是酯酶、蛋白酶、糖苷酶或脂肪酶。

[0195] 在一些实施方式中,本文所述的工程化核酸酶可以偶联至具有裂解酶活性的伴侣(例如,蛋白质,或其功能结构域或功能片段)。在一些情况下,具有裂解酶活性的伴侣(例如,蛋白质,或其功能结构域或功能片段)是苯丙氨酸氨裂解酶、柠檬酸裂解酶、异柠檬酸裂解酶、羟基腈(hydroxynitrile)、果胶酸裂解酶、精氨琥珀酸裂解酶、丙酮酸甲酸裂解酶、藻酸裂解酶或果胶裂解酶。

[0196] 在一些实施方式中,本文所述的工程化核酸酶可以偶联至具有异构酶活性的伴侣(例如,蛋白质,或其功能结构域或功能片段)。在一些情况下,具有异构酶活性的伴侣(例如,蛋白质,或其功能结构域或功能片段)是核糖磷酸异构酶、双磷酸甘油酸变位酶或光异构酶。

[0197] 在一些实施方式中,本文所述的工程化核酸酶可以偶联至具有合酶活性的伴侣(例如,蛋白质,或其功能结构域或功能片段)。在一些情况下,具有合酶活性的伴侣(例如,蛋白质,或其功能结构域或功能片段)是ATP合酶、柠檬酸合酶、色氨酸合酶、假尿苷合酶或脂肪酸合酶。

[0198] 在一些实施方式中,本文所述的工程化核酸酶可以偶联至具有脱肉豆蔻酰化活性的伴侣(例如,蛋白质,或其功能结构域或功能片段)。在一些情况下,具有脱肉豆蔻酰化活性的伴侣(例如,蛋白质,或其功能结构域或功能片段)是T3SS效应蛋白。

[0199] 在一些实施方式中,本文所述的工程化核酸酶可以偶联至具有转座活性的伴侣

(例如,蛋白质,或其功能结构域或功能片段)。在一些情况下,具有转座活性的伴侣(例如,蛋白质,或其功能结构域或功能片段)是转座酶Tn5或睡美人转座酶。

[0200] 本文公开的工程化核酸酶可以以任何形式提供。例如,工程化核酸酶可以以蛋白质(例如单独的工程化核酸酶或与引导核酸复合作为核糖核蛋白)的形式提供。工程化核酸酶可以以复合体的形式提供,例如与引导核酸和/或本公开的一种或多种异源基因效应因子复合。工程化核酸酶可以以编码至少工程化核酸酶的核酸(例如RNA(例如,信使RNA(mRNA))或DNA)的形式提供。编码至少工程化核酸酶的核酸可以进行密码子优化以便在特定细胞或生物体中有效翻译成蛋白质(例如,人类密码子优化)。

[0201] 编码至少本文公开的工程化核酸酶、其片段或衍生物的核酸可以稳定地整合到细胞的基因组中。编码至少工程化核酸酶的核酸可以可操作地连接到启动子,例如,在细胞中组成型或诱导型活性的启动子。编码至少工程化核酸酶的核酸可以可操作地连接到表达构建体中的启动子。表达构建体可以包括能够指导感兴趣的基因或其他核酸序列(例如,至少工程化核酸酶)的表达并且可以将这种感兴趣核酸序列转移到靶细胞的任何核酸构建体。

[0202] 在一些实施方式中,本文公开的工程化核酸酶可以与单个引导RNA(sgRNA)相关联以激活或阻抑靶基因(例如,靶内源基因)的转录,例如,与本文所公开的异源基因效应因子组合。可以将sgRNA引入表达本文提供的工程化核酸酶或其变体的细胞中。在一些情况下,此类细胞可以包含靶向相同靶基因(例如,靶内源基因)或靶基因调控序列的一个或多个不同的sgRNA。在另一些情况下,sgRNA靶向细胞中的不同核酸(例如,不同靶基因、不同靶基因调控序列,或者相同靶基因或靶基因调控序列内的不同序列)。

[0203] 酶失活(例如,核酸酶缺陷)可以指这样的核酸酶,其可以以序列特异性方式结合多核苷酸中的核酸序列,但可能不会切割靶多核苷酸或将以显著降低的频率切割它。酶失活引导部分可以包含酶失活结构域(例如,核酸酶结构域)。酶失活可以指无活性。酶失活可以指基本无活性。酶失活可以指基本上无活性。酶无活性可以指与相当的野生型活性(例如,核酸切割活性、野生型Cas活性)相比,活性不超过1%、不超过2%、不超过3%、不超过4%、不超过5%、不超过6%、不超过7%、不超过8%、不超过9%或不超过10%活性。

[0204] 在一些实施方式中,本文公开的工程化核酸酶的靶核酸可以是dsDNA。在这样的实施方式中,dsDNA靶向特异性至少部分地由两个参数决定:靶向靶dsDNA中的原间隔序列(protospacer)(靶dsDNA中对应于非互补DNA链上的gRNA间隔子的序列)的gRNA间隔子和位于非互补DNA链上原间隔序列的紧接5'(上游)的短序列原间隔序列临近基序(PAM)。在一些实施方式中,PAM是5'-TTTG-3'、5'-TTTA-3'或5'-TTTR-3'。在一些实施方式中,PAM是5'-TTTG-3'。在一些实施方式中,PAM是5'-TTTA-3'。在一些实施方式中,PAM是5'-TTTR-3'。

[0205] 在一些实施方式中,本文公开的工程化核酸酶的靶核酸可以是RNA。在这样的实施方式中,RNA靶向特异性至少部分由靶向靶RNA中的原间隔序列样序列(靶RNA中与gRNA间隔子互补的序列)的gRNA间隔子决定,并且与位于原间隔序列样序列的紧接5'(上游)的序列无关。在一些实施方式中,工程化核酸酶还可以能够靶向dsDNA分子,其中gRNA间隔子被选择为靶向具有选自5'-TTTG-3'、5'-TTTA-3'和5'-TTTR-3'的PAM的靶dsDNA分子中的原间隔序列。在另一些实施方式中,工程化核酸酶不能靶向dsDNA分子,其中gRNA间隔子被选择为使得gRNA间隔子靶向的dsDNA分子中的任何原间隔序列都不具有选自5'-TTTG-3'、5'-TTTA-3'和5'-TTTR-3'的PAM。

[0206] 在一些实施方式中,包含工程化核酸酶的异源多肽(例如,和/或包含异源多肽的复合体)可以调控靶基因(例如,靶内源基因)的表达和/或活性。在一些实施方式中,异源多肽和/或其复合体可以编辑核酸(例如,基因和/或基因产物)的序列。工程化核酸酶的核酸酶活性变体可以通过在靶多核苷酸中产生双链断裂或单链断裂来编辑核酸序列。

[0207] 在一些实施方式中,包含工程化核酸酶的异源多肽(例如,和/或包含异源多肽的复合体)可以在靶多核苷酸(例如DNA)中产生双链断裂。DNA中的双链断裂可导致DNA断裂修复,从而允许引入基因修饰(例如,核酸编辑)。在一些实施方式中,核酸酶诱导位点特异性单链DNA断裂或缺口,从而导致HDR。

[0208] DNA中的双链断裂可导致DNA断裂修复,从而允许引入基因修饰(例如,核酸编辑)。DNA断裂修复可以通过非同源末端连接(NHEJ)或同源介导修复(homology-directed repair,HDR)发生。在HDR中,可提供供体DNA修复模板或模板多核苷酸,其包含靶DNA位点侧翼的同源臂。

[0209] 在一些实施方式中,包含工程化核酸酶的异源多肽(例如,和/或包含异源多肽的复合体)不在靶多核苷酸(例如DNA)中产生双链断裂。异源多肽或包含异源多肽的复合体(例如,包含工程化核酸酶的核酸酶缺陷变体和引导RNA的复合体)在没有核酸断裂的情况下的结合可能足以调控靶基因(例如,内源靶基因)的表达(例如,增强或抑制)。

靶基因

[0210] 本公开提供了用于调节一个或多个靶基因的表达的组合物、方法和系统。靶基因可以是一个或多个内源靶基因,例如(i)致病等位基因,例如突变等位基因,和/或(ii)非致病等位基因,例如野生型等位基因。例如,本文公开了一种或多种复合体,其包含引导部分和一个或多个异源多肽,所述异源多肽包含本文公开的工程化多肽(例如包含工程化核酸酶),其可以调节(例如增加或减少)靶基因(例如在细胞中)的活性或表达水平。包含引导部分(例如小引导RNA)和工程化核酸酶的此类复合体可以通过切割靶基因来实现靶基因表达的调节。替代地或另外,此类复合体包含可操作地偶联至(例如,融合至)工程化核酸酶的基因调节因子,使得所述复合体可以在不切割靶基因的情况下实现靶基因表达的调节。在一些情况下,基因调节因子可以实现靶基因的表达增加。在一些情况下,基因调节因子可以实现靶基因的表达减少。在一些情况下,基因调节因子(例如,本文所述的基因编辑部分)可以实现靶基因的编辑,例如,以修正靶基因中的不良突变,使得突变基因的表达减少,并且经修正的基因的表达增加。

[0211] 在一些实施方式中,靶基因或其调控序列对于细胞是内源的,例如,存在于细胞的基因组中,或者对于受试者是内源的,例如,存在于受试者的基因组中。在一些实施方式中,靶基因或其调控序列不是工程化报告系统的一部分。

[0212] 在一些实施方式中,靶基因对于宿主受试者是外源的,例如,病原体靶基因或作为治疗干预(例如基因疗法和/或细胞疗法)的结果而表达的外源基因。在一些实施方式中,靶基因是外源报告基因。在一些实施方式中,靶基因是外源合成基因。

[0213] 在一些实施方式中,本文公开的系统和方法可以调节(例如,增加或减少)靶基因的表达(例如,在将包含异源多肽的复合体引入细胞或细胞群后)或其持续时间。在一些实施方式中,表达水平是RNA表达水平,可以通过例如RNAseq、qPCR、微阵列、基因阵列、FISH等来测量。在一些实施方式中,表达水平是蛋白质表达水平,可以通过例如蛋白质印迹、

ELISA、多重免疫测定、质谱法、NMR、蛋白质组学、流式细胞术、质谱流式细胞术等来测量。

[0214] 在一些实施方式中,本文公开的系统和方法可以将靶基因的表达(例如,在将包含异源多肽的复合体引入细胞或细胞群后)或其持续时间调节(例如,增加或减少)至少约10%、至少约20%、至少约30%、至少约40%、至少约50%、至少约60%、至少约70%、至少约80%、至少约90%、至少约2倍、至少约3倍、至少约4倍、至少约5倍、至少约6倍、至少约7倍、至少约8倍、至少约9倍、至少约10倍、至少约11倍、至少约12倍、至少约13倍、至少约14倍、至少约15倍、至少约20倍、至少约30倍、至少约40倍、至少约50倍、至少约60倍、至少约70倍、至少约80倍、至少约90倍、至少约100倍、至少约150倍、至少约200倍、至少约250倍、至少约300倍、至少约350倍、至少约400倍、至少约500倍、至少约600倍、至少约700倍、至少约800倍、至少约900倍、至少约1000倍、至少约1500倍、至少约2000倍、或至少约3000倍。

[0215] 在一些实施方式中,本文公开的系统和方法可以将靶基因的表达(例如,在将包含异源多肽的复合体引入细胞或细胞群后)或其持续时间调节(例如,增加或减少)至多约50%、至多约60%、至多约70%、至多约80%、至多约90%、至多约2倍、至多约3倍、至多约4倍、至多约5倍、至多约6倍、至多约7倍、至多约8倍、至多约9倍、至多约10倍、至多约11倍、至多约12倍、至多约13倍、至多约14倍、至多约15倍、至多约20倍、至多约30倍、至多约40倍、至多约50倍、至多约60倍、至多约70倍、至多约80倍、至多约90倍、至多约100倍、至多约150倍、至多约200倍、至多约250倍、至多约300倍、至多约350倍、至多约400倍、至多约500倍、至多约600倍、至多约700倍、至多约800倍、至多约900倍、至多约1000倍、至多约1500倍、至多约2000倍、至多约3000倍、至多约5000倍、或至多约10000倍。

[0216] 在一些实施方式中,本文公开的系统和方法可以将靶基因的表达(例如,在将包含异源多肽的复合体引入细胞或细胞群后)或其持续时间调节(例如,增加或减少)约10%、约20%、约30%、约40%、约50%、约60%、约70%、约80%、约90%、约2倍、约3倍、约4倍、约5倍、约6倍、约7倍、约8倍、约9倍、约10倍、约11倍、约12倍、约13倍、约14倍、约15倍、约20倍、约30倍、约40倍、约50倍、约60倍、约70倍、约80倍、约90倍、约100倍、约150倍、约200倍、约250倍、约300倍、约350倍、约400倍、约500倍、约600倍、约700倍、约800倍、约900倍、约1000倍、约1500倍、约2000倍、约3000倍、约5000倍、或约10000倍。

[0217] 在一些实施方式中,本文公开的系统和方法可以将靶基因的表达(例如,将包含异源多肽的复合体引入细胞或细胞群后)或其持续时间从低于检测限调节(例如,增加或减少)到可检测水平。

[0218] 在一些实施方式中,表达或其持续时间的变化程度是相对于将本公开的系统(例如,包含异源多肽的复合体)引入细胞或细胞群之前。在一些实施方式中,表达或其持续时间的变化程度是相对于未用本公开的系统处理的相应对照细胞或细胞群。在一些实施方式中,表达或其持续时间的变化程度是相对于用本公开系统的替代品处理的相应对照细胞或细胞群。

[0219] 在一些实施方式中,表达或其持续时间的变化程度是相对于对照核酸酶。对照核酸酶可以包括天然存在的核酸酶(例如,(i)SEQ ID NO:1的多肽序列或(ii)选自表2的任一种Cas蛋白的多肽序列)或其任何修饰(例如,具有降低的核酸酶活性并可操作地偶联至基因调节因子的其变体)。例如,对照核酸酶可以是与相同基因调节因子偶联(例如,融合)的dCasMINI,其中相同基因调节因子与本文所公开的工程化核酸酶偶联。

[0220] 在一些实施方式中,本文公开的系统和方法可以调节(例如,增加或减少)靶基因的活性水平(例如,在将包含含有本文公开的工程化核酸酶的异源多肽的复合体引入细胞或细胞群中后)或其持续时间。活性水平可以通过针对所讨论的靶基因的合适的功能测定来确定,这取决于靶基因的功能特征。例如,作为有丝分裂原的靶基因的活性水平可以通过测量细胞增殖来确定;诱导凋亡的靶基因的活性水平可以通过膜联蛋白V测定或其他合适的细胞死亡测定来测量;抗炎细胞因子的活性水平可以通过LPS诱导的细胞因子释放测定来测量。

[0221] 在一些情况下,本公开的系统和方法可以引起靶基因(例如,靶内源基因)的表达和/或活性水平的变化,这种变化持续时间比使用替代组合物和方法(例如,通过RNAi抑制,例如,使用siRNA)所能实现的更长。在一些实施方式中,与瞬时调节相比,基因表达的持续调节是有利的。

[0222] 在一些实施方式中,本文公开的系统和方法可以调节(例如,增加或减少)靶基因的表达和/或活性水平至少约1小时、至少约2小时、至少约3小时、至少约4小时、至少约5小时、至少约6小时、至少约7小时、至少约8小时、至少约9小时、至少约10小时、至少约12小时、至少约14小时、至少约18小时、至少约20小时、至少约1天、至少约2天、至少约3天、至少约4天、至少约5天、至少约6天、至少约7天、至少约8天、至少约9天、至少约10天、至少约14天、至少约21天、至少约28天、至少约5周、至少约6周、至少约7周、至少约8周、至少约9周、至少约10周、至少约12周、至少约14周、至少约18周、至少约20周、至少约26周、或至少约5个月、至少约6个月、至少约9个月、至少约12个月或更久。

[0223] 在一些实施方式中,本文公开的系统和方法可以调节(例如,增加或减少)靶基因(例如,靶内源基因)的表达和/或活性水平至高于某一阈值,持续至少或多至约1小时、至少或多至约2小时、至少或多至约3小时、至少或多至约4小时、至少或多至约5小时、至少或多至约6小时、至少或多至约7小时、至少或多至约8小时、至少或多至约9小时、至少或多至约10小时、至少或多至约12小时、至少或多至约14小时、至少或多至约18小时、至少或多至约20小时、至少或多至约1天、至少或多至约2天、至少或多至约3天、至少或多至约4天、至少或多至约5天、至少或多至约6天、至少或多至约7天、至少或多至约8天、至少或多至约9天、至少或多至约10天、至少或多至约14天、至少或多至约21天、至少或多至约28天、至少或多至约5周、至少或多至约6周、至少或多至约7周、至少或多至约8周、至少或多至约9周、至少或多至约10周、至少或多至约12周、至少或多至约14周、至少或多至约18周、至少或多至约20周、至少或多至约26周、或至少或多至约5个月、至少或多至约6个月、至少或多至约9个月、或至少或多至约12个月。

[0224] 在一些实施方式中,本文公开的系统和方法可以调节(例如,增加或减少)靶基因(例如,靶内源基因)的表达和/或活性水平至高于某一阈值,持续约1小时、约2小时、约3小时、约4小时、约5小时、约6小时、约7小时、约8小时、约9小时、约10小时、约12小时、约14小时、约18小时、约20小时、约1天、约2天、约3天、约4天、约5天、约6天、约7天、约8天、约9天、约10天、约14天、约21天、约28天、约5周、约6周、约7周、约8周、约9周、约10周、约12周、约14周、约18周、约20周、约26周、约5个月、约6个月、约9个月或约12个月。

[0225] 在一些实施方式中,与对照多肽(例如,可操作地偶联至相同基因调节因子的对照核酸酶)对靶基因的调节相比,本文公开的工程化多肽(例如,可操作地偶联至基因调节因

子的工程化核酸酶)可能能够增强调节靶基因或可以实现对靶基因的增强的调节。对照核酸酶可以是天然存在的核酸酶(例如,(i)SEQ ID NO:1的多肽序列或(ii)选自表2中的任一种Cas蛋白的多肽序列)或其任何修饰(例如,本文公开的dCasMINI)。

[0226] 在一些情况下,靶基因的增强的调节可以通过靶基因表达水平的变化大于对照多肽导致的变化来表征。在一些实例中,这种变化可以是靶基因的提高的表达水平。本文公开的工程化多肽导致的靶基因的提高的表达水平可以比对照多肽导致的高至少或多至约1%、至少或多至约2%、至少或多至约3%、至少或多至约4%、至少或多至约5%、至少或多至约10%、至少或多至约15%、至少或多至约20%、至少或多至约30%、至少或多至约40%、至少或多至约50%、至少或多至约60%、至少或多至约70%、至少或多至约80%、至少或多至约90%、至少或多至约100%、至少或多至约110%、至少或多至约120%、至少或多至约150%、至少或多至约200%、至少或多至约300%、至少或多至约400%、或至少或多至约500%。本文公开的工程化多肽导致的靶基因的提高的表达水平可以比对照多肽导致的高至少或多至约0.1倍、至少或多至约0.2倍、至少或多至约0.5倍、至少或多至约1倍、至少或多至约1.5倍、至少或多至约2倍、至少或多至约3倍、至少或多至约4倍、至少或多至约5倍、至少或多至约6倍、至少或多至约7倍、至少或多至约8倍、至少或多至约9倍、至少或多至约10倍、至少或多至约15倍、至少或多至约20倍、至少或多至约30倍、至少或多至约40倍、或至少或多至约100倍。在一些实例中,这种变化可以是靶基因的减少的(或降低的)表达水平。本文公开的工程化多肽导致的靶基因的减少的表达水平可以比对照多肽导致的低至少或多至约1%、至少或多至约2%、至少或多至约3%、至少或多至约4%、至少或多至约5%、至少或多至约10%、至少或多至约15%、至少或多至约20%、至少或多至约30%、至少或多至约40%、至少或多至约50%、至少或多至约60%、至少或多至约70%、至少或多至约80%、至少或多至约90%、至少或多至约100%、至少或多至约110%、至少或多至约120%、至少或多至约150%、至少或多至约200%、至少或多至约300%、至少或多至约400%、或至少或多至约500%。本文公开的工程化多肽导致的靶基因的减少的(或降低的)表达水平可以比对照多肽导致的低至少或多至约0.1倍、至少或多至约0.2倍、至少或多至约0.5倍、至少或多至约1倍、至少或多至约1.5倍、至少或多至约2倍、至少或多至约3倍、至少或多至约4倍、至少或多至约5倍、至少或多至约6倍、至少或多至约7倍、至少或多至约8倍、至少或多至约9倍、至少或多至约10倍、至少或多至约15倍、至少或多至约20倍、至少或多至约30倍、至少或多至约40倍、或至少或多至约100倍。

[0227] 在一些情况下,靶基因的增强的调节可以通过靶基因表达水平的延长的变化(例如,如本文所公开的,高于某阈值的增加的表达或减少的表达)比对照多肽导致的更长来表征。本文公开的工程化多肽导致的靶基因表达水平的延长的变化可以比对照多肽导致的长至少或多至约1%、至少或多至约2%、至少或多至约3%、至少或多至约4%、至少或多至约5%、至少或多至约10%、至少或多至约15%、至少或多至约20%、至少或多至约30%、至少或多至约40%、至少或多至约50%、至少或多至约60%、至少或多至约70%、至少或多至约80%、至少或多至约90%、至少或多至约100%、至少或多至约110%、至少或多至约120%、至少或多至约150%、至少或多至约200%、至少或多至约300%、至少或多至约400%、或至少或多至约500%。本文公开的工程化多肽导致的靶基因表达水平的延长的变化可以比对照多肽导致的长至少或多至约0.1倍、至少或多至约0.2倍、至少或多至约0.5倍、至少或多

至约1倍、至少或多至约1.5倍、至少或多至约2倍、至少或多至约3倍、至少或多至约4倍、至少或多至约5倍、至少或多至约6倍、至少或多至约7倍、至少或多至约8倍、至少或多至约9倍、至少或多至约10倍、至少或多至约15倍、至少或多至约20倍、至少或多至约30倍、至少或多至约40倍、或至少或多至约100倍。例如,阈值水平可以相对于(i)增强的调节之前靶基因的表达水平,(ii)对于激活靶基因或其部分来说靶基因表达水平的最大增加(例如,至少50%、至少60%、至少70%、至少80%或至少90%),或(iii)对于阻抑靶基因或其部分来说靶基因表达水平的最大降低(例如,至少50%、至少60%、至少70%、至少80%或至少90%)。在一些实施方式中,靶基因(例如,内源靶基因)可以是致病等位基因,例如野生型等位基因的突变变体。疾病可以是遗传疾病,例如遗传性紊乱。遗传疾病的非限制性实例可以包括杜氏肌营养不良症(DMD)、血友病、囊性纤维化、亨廷顿舞蹈症、家族性高胆固醇血症(LDL受体缺陷)、肝母细胞瘤、威尔逊氏病、先天性肝卟啉症、遗传性肝代谢障碍、莱施-尼汉综合征(Lesch Nyhan syndrome)、镰状细胞性贫血、地中海贫血、着色性干皮病、范科尼贫血(Fanconi's anemia)、视网膜色素变性、毛细血管扩张性共济失调、布卢姆综合征(Bloom's syndrome)、视网膜母细胞瘤和泰-萨克斯病(Tay-Sachs disease)。在一些情况下,靶基因可以是编码蛋白质的基因。在一些情况下,靶基因可以是基因调控序列(例如,启动子、增强子、阻抑因子、沉默子、绝缘子、顺式调控元件、反式调控元件、表观遗传修饰(例如,DNA甲基化)位点等),其可以影响编码本文提供的感兴趣的蛋白质的基因的表达。例如,靶基因调控序列可以物理地位于编码靶基因产物的转录单元或开放阅读框之外。

[0228] 在一些实施方式中,靶基因调控序列不包含对于受试者或宿主细胞而言外源的核苷酸序列。在一些实施方式中,靶基因调控序列不包含工程化或人工产生或引入的核苷酸序列。

[0229] 在一些实施方式中,靶基因(例如,靶内源基因)是在疾病或病症中过表达或表达不足的基因。在一些实施方式中,靶基因是在可遗传的遗传疾病中过表达或表达不足的基因。

[0230] 在一些实施方式中,靶基因(例如靶内源基因)是在癌症中过表达或表达不足的基因,例如急性白血病、星形细胞瘤、胆道癌(胆管癌)、骨癌、乳腺癌、脑干神经胶质瘤、支气管肺泡细胞肺癌、肾上腺癌、肛门区癌、膀胱癌、内分泌系统癌、食道癌、头颈癌、肾癌、甲状旁腺癌、阴茎癌、胸膜/腹膜癌、唾液腺癌、小肠癌、甲状腺癌、输尿管癌、尿道癌、宫颈癌、子宫内膜癌、输卵管癌、肾盂癌、阴道癌、外阴癌、宫颈癌、慢性白血病、结肠癌、结直肠癌、皮肤黑色素瘤、室管膜瘤、表皮样肿瘤、尤文氏肉瘤、胃癌、胶质母细胞瘤、多形性胶质母细胞瘤、胶质瘤、血液系统恶性肿瘤、肝细胞(肝)癌、肝细胞瘤、霍奇金病、眼内黑色素瘤、卡波西肉瘤、肺癌、淋巴瘤、髓母细胞瘤、黑色素瘤、脑膜瘤、间皮瘤、多发性骨髓瘤、肌肉癌、中枢神经系统(CNS)赘生物、神经元癌、小细胞肺癌、非小细胞肺癌、骨肉瘤、卵巢癌、胰腺癌、儿童恶性肿瘤、垂体腺瘤、前列腺癌、直肠癌、肾细胞癌、软组织肉瘤、神经鞘瘤、皮肤癌、脊柱肿瘤、鳞状细胞癌、胃部癌、滑膜肉瘤、睾丸癌、子宫癌、或肿瘤及其转移,包括上述任何癌症的难治性版本,或其组合。

[0231] 如本文所公开的,靶基因或编码感兴趣的蛋白质的基因的非限制性实例包括在表1中。

引导核酸分子

[0232] 在一些方面,本公开提供了引导核酸分子(例如,工程化的引导核酸分子),所述引导核酸分子被配置为与Cas蛋白形成复合体。Cas蛋白可以是天然存在的蛋白质。Cas蛋白可以是本文提供的工程化的核酸酶变体。引导核酸分子可以包含表现出与可操作地偶联至靶基因(例如,在细胞中)的靶多核苷酸序列的特异性结合的间隔序列。靶多核苷酸序列可以是靶基因的一部分。替代地,靶多核苷酸序列可以在靶基因的上游(例如,靶基因的启动子序列的一部分或邻近靶基因的启动子序列)或下游(例如,靶基因的终止序列的一部分或邻近靶基因的终止序列)。引导核酸分子可以包含用于与Cas蛋白形成复合体的支架序列。间隔序列和支架序列可以是单个多核苷酸序列(例如单个引导核酸分子,如sgRNA)的一部分。替代地,间隔序列和支架序列可以是单独的分子,其杂交以与Cas蛋白形成复合体。

[0233] 不希望受理论的束缚,本文公开的引导核酸分子可以可操作地偶联至(例如可以与其形成功能复合体)一种或多种Cas蛋白,包括但不限于Un1Cas12f1、选自表2的Cas蛋白或整个本公开中提供的任何工程化核酸酶变体(例如SEQ ID NO:12的多肽)。

[0234] 在一些实施方式中,本文公开的支架序列与SEQ ID NO:500的多核苷酸序列不相同。与SEQ ID NO:500的多核苷酸序列相比(例如,当比对时),支架序列可以包含至少一个缺失。不希望受理论的束缚,支架序列的至少一个缺失可以通过对SEQ ID NO:500的对照支架序列进行缺失景观研究(例如,迭代和/或全面缺失)来确定。与SEQ ID NO:500的多核苷酸序列相比(例如,当比对时),支架序列可以包含至少一个突变。不希望受理论的束缚,支架序列的至少一个突变可以通过对SEQ ID NO:500的对照支架序列进行突变景观研究(例如,迭代和/或全面突变)来确定。本文公开的至少一个缺失可以是核苷酸的去除。替代地,至少一个缺失可以是用不同的核苷酸替换核苷酸(例如,突变)。

[0235] 在一些实施方式中,当与SEQ ID NO:500的对照多核苷酸序列比对(或比较)时,支架序列可以包含一个或多个核苷酸缺失。一个或多个核苷酸缺失可以包括单个缺失。一个或多个核苷酸缺失可以包括多个核苷酸缺失,例如至少或多至约2个缺失、至少或多至约3个缺失、至少或多至约4个缺失、至少或多至约5个缺失、至少或多至约6个缺失、至少或多至约7个缺失、至少或多至约8个缺失、至少或多至约9个缺失、至少或多至约10个缺失、至少或多至约11个缺失、至少或多至约12个缺失、至少或多至约13个缺失、至少或多至约14个缺失、至少或多至约15个缺失、至少或多至约16个缺失、至少或多至约17个缺失、至少或多至约18个缺失、至少或多至约19个缺失、至少或多至约20个缺失、至少或多至约22个缺失、至少或多至约24个缺失、至少或多至约25个缺失、至少或多至约26个缺失、至少或多至约28个缺失、至少或多至约30个缺失、至少或多至约32个缺失、至少或多至约34个缺失、至少或多至约35个缺失、至少或多至约36个缺失、至少或多至约38个缺失、至少或多至约40个缺失、至少或多至约42个缺失、至少或多至约44个缺失、至少或多至约45个缺失、至少或多至约46个缺失、至少或多至约48个缺失、至少或多至约50个缺失、至少或多至约52个缺失、至少或多至约54个缺失、至少或多至约55个缺失、至少或多至约56个缺失、至少或多至约58个缺失、至少或多至约60个缺失、至少或多至约70个缺失、或至少或多至约80个缺失。当与SEQ ID NO:500的多核苷酸序列比对时,多个核苷酸缺失可以彼此相邻(例如,连续)。支架序列可以包含单个连续缺失。支架序列可以包含多个连续缺失,其中当与SEQ ID NO:500的多核苷酸序列比对时,一个连续缺失与另一个连续缺失不直接相邻。

[0236] 在一些实施方式中,当与SEQ ID NO:500的对照多核苷酸序列比对时,本文公开的

支架序列可以在来自SEQ ID NO:500的核苷酸1-10、核苷酸11-20、核苷酸21-30、核苷酸31-40、核苷酸41-50、核苷酸51-60、核苷酸61-70、核苷酸71-80、核苷酸81-90、核苷酸91-100、核苷酸101-110、核苷酸111-120、核苷酸121-130、核苷酸131-140、核苷酸141-150和/或核苷酸151-159的一个或多个成员中包含一个或多个核苷酸缺失。

[0237] 在一些实施方式中,当与SEQ ID NO:500的对照多核苷酸序列比对时,本文公开的支架序列可以包含SEQ ID NO:500的多核苷酸序列的核苷酸1-25,例如SEQ ID NO:500的多核苷酸序列的核苷酸1-23、核苷酸3-23、核苷酸5-23、核苷酸7-23、核苷酸9-23、核苷酸11-23、核苷酸13-23、核苷酸15-23、核苷酸17-23、核苷酸19-23和/或核苷酸21-23中的一个或多个核苷酸缺失。当与SEQ ID NO:500的对照多核苷酸序列比对时,本文公开的支架序列可以包含SEQ ID NO:500的多核苷酸序列的核苷酸1-23、核苷酸1-21、核苷酸1-19、核苷酸1-17、核苷酸1-15、核苷酸1-13、核苷酸1-11、核苷酸1-9、核苷酸1-7、核苷酸1-5和/或核苷酸1-3中的一个或多个核苷酸缺失。当与SEQ ID NO:500的对照多核苷酸序列比对时,本文公开的支架序列可以包含SEQ ID NO:500的多核苷酸序列的核苷酸1-5、核苷酸6-10、核苷酸11-15、核苷酸16-20和/或核苷酸21-23中的一个或多个核苷酸缺失。当与SEQ ID NO:500的对照多核苷酸序列比对时,本文公开的支架序列可以包含SEQ ID NO:500的多核苷酸序列的核苷酸1、核苷酸2、核苷酸3、核苷酸4、核苷酸5、核苷酸6、核苷酸7、核苷酸8、核苷酸9、核苷酸10、核苷酸11、核苷酸12、核苷酸13、核苷酸14、核苷酸15、核苷酸16、核苷酸17、核苷酸18、核苷酸19、核苷酸20、核苷酸21、核苷酸22、核苷酸23、核苷酸24和/或核苷酸25中的一个或多个核苷酸缺失。

[0238] 在一些实施方式中,当与SEQ ID NO:500的对照多核苷酸序列比对时,本文公开的支架序列可以包含SEQ ID NO:500的多核苷酸序列的核苷酸35-65,例如SEQ ID NO:500的多核苷酸序列的核苷酸35-61、核苷酸37-61、核苷酸39-61、核苷酸41-61、核苷酸43-61、核苷酸45-61、核苷酸47-61、核苷酸49-61、核苷酸51-61、核苷酸53-61、核苷酸55-61、核苷酸57-61和/或核苷酸59-61中的一个或多个核苷酸缺失。当与SEQ ID NO:500的对照多核苷酸序列比对时,本文公开的支架序列可以包含SEQ ID NO:500的多核苷酸序列的核苷酸35-61、核苷酸35-59、核苷酸35-57、核苷酸35-55、核苷酸35-53、核苷酸35-51、核苷酸35-49、核苷酸35-47、核苷酸35-45、核苷酸35-43、核苷酸35-41、核苷酸35-39和/或核苷酸35-37中的一个或多个核苷酸缺失。当与SEQ ID NO:500的对照多核苷酸序列比对时,本文公开的支架序列可以包含SEQ ID NO:500的多核苷酸序列的核苷酸35、核苷酸36、核苷酸37、核苷酸38、核苷酸39、核苷酸40、核苷酸41、核苷酸42、核苷酸43、核苷酸44、核苷酸45、核苷酸46、核苷酸47、核苷酸48、核苷酸49、核苷酸50、核苷酸51、核苷酸52、核苷酸53、核苷酸54、核苷酸55、核苷酸56、核苷酸57、核苷酸58、核苷酸59、核苷酸60、核苷酸61、核苷酸62、核苷酸63、核苷酸64和/或核苷酸65中的一个或多个核苷酸缺失。

[0239] 在一些实施方式中,当与SEQ ID NO:500的对照多核苷酸序列比对时,本文公开的支架序列可以包含SEQ ID NO:500的多核苷酸序列的核苷酸135-150,例如SEQ ID NO:500的多核苷酸序列的核苷酸136-149、核苷酸137-149、核苷酸139-149、核苷酸141-149、核苷酸143-149、核苷酸145-149和/或核苷酸147-149中的一个或多个核苷酸缺失。当与SEQ ID NO:500的对照多核苷酸序列比对时,本文公开的支架序列可以包含SEQ ID NO:500的多核苷酸序列的核苷酸136-149、核苷酸136-147、核苷酸136-145、核苷酸136-143、核苷酸136-

141、核苷酸136-139和/或核苷酸136-137中的一个或多个核苷酸缺失。当与SEQ ID NO:500的对照多核苷酸序列比对时,本文公开的支架序列可以包含SEQ ID NO:500的多核苷酸序列的核苷酸135、核苷酸136、核苷酸137、核苷酸138、核苷酸139、核苷酸140、核苷酸141、核苷酸142、核苷酸143、核苷酸144、核苷酸145、核苷酸146、核苷酸147、核苷酸148、核苷酸149和/或核苷酸150中的一个或多个核苷酸缺失。

[0240] 在一些实施方式中,当与SEQ ID NO:500的对照多核苷酸序列比对时,本文公开的支架序列可以包含SEQ ID NO:500的多核苷酸序列的核苷酸136-151中的一个或多个核苷酸缺失。在一些情况下,本文公开的支架序列可以包含SEQ ID NO:500的多核苷酸序列的核苷酸T136、T137、C138、A139、T140、T141、T142、G143、A144、A145、T146、G147、A148、A149、G150和/或G151中的一个或多个核苷酸缺失。在一些情况下,本文公开的支架序列可以包含SEQ ID NO:500的多核苷酸序列的核苷酸T136、T137、C138、A139、T140、T141、T142、A144、A145、T146、A148和/或A149中的一个或多个核苷酸缺失(例如至少1个、至少2个、至少3个、至少4个、至少5个、至少6个、至少7个、至少8个、至少9个、至少10个、至少11个或全部12个)。在一些情况下,当与SEQ ID NO:500的多核苷酸序列比对时,本文公开的支架序列可以包含选自于由G143、G147、G150和G151所组成的组中的至少或多至约1个核苷酸、至少或多至约2个核苷酸、至少或多至约3个核苷酸、或全部4个核苷酸。

[0241] 在一些实施方式中,本文公开的支架序列与SEQ ID NO:549和SEQ ID NO:550的组合的多核苷酸序列不相同。例如,SEQ ID NO:549和SEQ ID NO:550的多核苷酸可以分别偶联至间隔序列的5'端和3'端,以用作对照sgRNA分子,以比较本文提供的任何支架序列的活性。在一些实施方式中,本文公开的支架序列与SEQ ID NO:551和SEQ ID NO:552的组合的多核苷酸序列不相同。例如,SEQ ID NO:551和SEQ ID NO:552的多核苷酸可以分别偶联至间隔序列的5'端和3'端,以用作对照sgRNA分子,以比较本文提供的任何支架序列的活性。

[0242] 在一些实施方式中,支架序列(例如,支架序列的连续多核苷酸序列)可表征为表现出与选自表6B、表7B和表8B的成员的多核苷酸序列的至少或高至约60%、至少或高至约65%、至少或高至约70%、至少或高至约71%、至少或高至约72%、至少或高至约73%、至少或高至约74%、至少或高至约75%、至少或高至约76%、至少或高至约77%、至少或高至约78%、至少或高至约79%、至少或高至约80%、至少或高至约81%、至少或高至约82%、至少或高至约83%、至少或高至约84%、至少或高至约85%、至少或高至约86%、至少或高至约87%、至少或高至约88%、至少或高至约89%、至少或高至约90%、至少或高至约91%、至少或高至约92%、至少或高至约93%、至少或高至约94%、至少或高至约95%、至少或高至约96%、至少或高至约97%、至少或高至约98%、至少或高至约99%或基本上100%序列同一性(或互补性)。

[0243] 在一些情况下,所述成员可以选自于由SEQ ID NO:503-152、519、524、528和553所组成的组。在一些情况下,所述成员可以选自于由SEQ ID NO:555、557、558、568、569、578和580所组成的组。在一些情况下,所述成员可以选自于由SEQ ID NO:555、557、568、569、576、577、578、580、593、519和528所组成的组。

[0244] 在一些情况下,支架序列的长度可以为至少或多至约80个核苷酸、至少或多至约85个核苷酸、至少或多至约90个核苷酸、至少或多至约91个核苷酸、至少或多至约92个核苷酸、至少或多至约93个核苷酸、至少或多至约94个核苷酸、至少或多至约95个核苷酸、至少

或多至约96个核苷酸、至少或多至约97个核苷酸、至少或多至约98个核苷酸、至少或多至约99个核苷酸、至少或多至约100个核苷酸、至少或多至约101个核苷酸、至少或多至约102个核苷酸、至少或多至约103个核苷酸、至少或多至约104个核苷酸、至少或多至约105个核苷酸、至少或多至约106个核苷酸、至少或多至约107个核苷酸、至少或多至约108个核苷酸、至少或多至约109个核苷酸、至少或多至约110个核苷酸、至少或多至约112个核苷酸、至少或多至约114个核苷酸、至少或多至约115个核苷酸、至少或多至约116个核苷酸、至少或多至约118个核苷酸、至少或多至约120个核苷酸、至少或多至约122个核苷酸、至少或多至约124个核苷酸、至少或多至约125个核苷酸、至少或多至约126个核苷酸、至少或多至约128个核苷酸、至少或多至约130个核苷酸、至少或多至约135个核苷酸、至少或多至约140个核苷酸、至少或多至约145个核苷酸、至少或多至约150个核苷酸、至少或多至约155个核苷酸、或至少或多至约160个核苷酸。

[0245] 在一些实施方式中,支架序列可以包含连续多核苷酸序列,所述连续多核苷酸序列表现出与SEQ ID NO:597或SEQ ID NO:598的多核苷酸序列的至少或高至约60%、至少或高至约65%、至少或高至约70%、至少或高至约71%、至少或高至约72%、至少或高至约73%、至少或高至约74%、至少或高至约75%、至少或高至约76%、至少或高至约77%、至少或高至约78%、至少或高至约79%、至少或高至约80%、至少或高至约81%、至少或高至约82%、至少或高至约83%、至少或高至约84%、至少或高至约85%、至少或高至约86%、至少或高至约87%、至少或高至约88%、至少或高至约89%、至少或高至约90%、至少或高至约91%、至少或高至约92%、至少或高至约93%、至少或高至约94%、至少或高至约95%、至少或高至约96%、至少或高至约97%、至少或高至约98%、至少或高至约99、或基本上100%序列同一性(或互补性)。支架序列的连续多核苷酸序列的长度可以为至少或多至约15个核苷酸、至少或多至约16个核苷酸、至少或多至约17个核苷酸、至少或多至约18个核苷酸、至少或多至约19个核苷酸、至少或多至约20个核苷酸、至少或多至约21个核苷酸、至少或多至约22个核苷酸、至少或多至约23个核苷酸、至少或多至约24个核苷酸、至少或多至约25个核苷酸、至少或多至约26个核苷酸、至少或多至约27个核苷酸、至少或多至约28个核苷酸、至少或多至约29个核苷酸、至少或多至约30个核苷酸、至少或多至约31个核苷酸、至少或多至约32个核苷酸、至少或多至约33个核苷酸、至少或多至约34个核苷酸、至少或多至约35个核苷酸、至少或多至约36个核苷酸、至少或多至约37个核苷酸、至少或多至约38个核苷酸、至少或多至约39个核苷酸、或至少或多至约40个核苷酸。连续多核苷酸序列可配置于支架序列的N端或C端。连续多核苷酸序列可配置于支架序列的N端50%、N端45%、N端40%、N端35%、N端30%、N端25%、N端20%、N端15%或N端10%。替代地,连续多核苷酸序列可配置于支架序列的C端50%、C端45%、C端40%、C端35%、C端30%、C端25%、C端20%、C端15%或C端10%。连续多核苷酸序列可配置于支架序列之间。

[0246] 在一些实施方式中,引导核酸分子的间隔序列可以具有至少或多至约12个核苷酸、至少或多至约13个核苷酸、至少或多至约14个核苷酸、至少或多至约15个核苷酸、至少或多至约16个核苷酸、至少或多至约17个核苷酸、至少或多至约18个核苷酸、至少或多至约19个核苷酸、至少或多至约20个核苷酸、至少或多至约21个核苷酸、或至少或多至约22个核苷酸的长度。

[0247] 在一些实施方式中,引导核酸分子可以具有至少或多至约80个核苷酸、至少或多

至约85个核苷酸、至少或多至约90个核苷酸、至少或多至约95个核苷酸、至少或多至约96个核苷酸、至少或多至约97个核苷酸、至少或多至约98个核苷酸、至少或多至约99个核苷酸、至少或多至约100个核苷酸、至少或多至约101个核苷酸、至少或多至约102个核苷酸、至少或多至约103个核苷酸、至少或多至约104个核苷酸、至少或多至约105个核苷酸、至少或多至约106个核苷酸、至少或多至约107个核苷酸、至少或多至约108个核苷酸、至少或多至约109个核苷酸、至少或多至约110个核苷酸、至少或多至约111个核苷酸、至少或多至约112个核苷酸、至少或多至约113个核苷酸、至少或多至约114个核苷酸、至少或多至约115个核苷酸、至少或多至约116个核苷酸、至少或多至约117个核苷酸、至少或多至约118个核苷酸、至少或多至约119个核苷酸、至少或多至约120个核苷酸、至少或多至约121个核苷酸、至少或多至约122个核苷酸、至少或多至约123个核苷酸、至少或多至约124个核苷酸、至少或多至约125个核苷酸、至少或多至约130个核苷酸、至少或多至约135个核苷酸、至少或多至约140个核苷酸、至少或多至约145个核苷酸、至少或多至约150个核苷酸、至少或多至约155个核苷酸、或至少或多至约160个核苷酸的长度。

[0248] 在一些实例中,支架序列的连续多核苷酸序列可以为至少约20个、至少约25个或至少约30个核苷酸长,并且此类连续多核苷酸序列可以表现出与选自表6B、表7B和表8B的成员(例如来自SEQ ID NO:555、557、568、569、576、577、578、580、593、519和528的一个或多个成员)的(i)N端30个核苷酸序列或(ii)C端30个核苷酸序列的多核苷酸序列的至少或高至约60%、至少或高至约65%、至少或高至约70%、至少或高至约71%、至少或高至约72%、至少或高至约73%、至少或高至约74%、至少或高至约75%、至少或高至约76%、至少或高至约77%、至少或高至约78%、至少或高至约79%、至少或高至约80%、至少或高至约81%、至少或高至约82%、至少或高至约83%、至少或高至约84%、至少或高至约85%、至少或高至约86%、至少或高至约87%、至少或高至约88%、至少或高至约89%、至少或高至约90%、至少或高至约91%、至少或高至约92%、至少或高至约93%、至少或高至约94%、至少或高至约95%、至少或高至约96%、至少或高至约97%、至少或高至约98%、至少或高至约99、或基本上100%序列同一性(或互补性)。

异源多核苷酸

[0249] 在一些实施方式中,本公开的系统(例如,包括保留其核酸酶活性的至少一部分的工程化核酸酶的变体)可以靶向靶基因以编辑所述靶基因。在一些情况下,包含(i)含有本文公开的工程化核酸酶的异源多肽和(ii)引导核酸(例如,sgRNA)的复合体可以识别、结合靶基因并在靶基因中产生切口(一条链)或断裂(两条链),例如在靶基因内的复合体的靶序列处或附近。在一些情况下,切口或断裂可以通过非同源末端连接(NHEJ)修复。在一些情况下,切口或断裂可以通过同源介导修复(HDR)或通过同源重组(HR)用多核苷酸修饰模板(例如,供体模板,如供体DNA模板)修复。在一些实例中,可将编码感兴趣的基因的异源多核苷酸修饰模板提供给细胞,使得感兴趣的基因可被插入靶基因中,例如用于基因替代疗法。

[0250] 在一些实施方式中,本公开的系统和组合物将异源多核苷酸(例如,编码感兴趣基因,如选自表1的一个或多个基因)引入细胞,而不通过本公开的工程化核酸酶的作用进入细胞的基因组。在一些情况下,编码感兴趣的基因的此类异源多核苷酸可以通过其他方式进入细胞的基因组,例如通过腺相关病毒载体(例如,AAV2或AAV8)。替代地,编码感兴趣的基因的此类异源多核苷酸可以被引入细胞的细胞内部分并保持非染色体(例如,作为非染

色体质粒)。

[0251] 因此,该系统和组合物可以包括如上所述的靶基因的非致病野生型或变体。替代地或另外,该系统和组合物可以包括编码(或包含)本文所公开的靶基因(例如,内源靶基因)的至少非致病野生型或变体的异源多核苷酸序列。

组合物

[0252] 在一些方面,本公开提供了组合物,所述组合物包含如上所述的系统的至少一部分,例如如本文所公开的(i)包含工程化核酸酶的异源多肽或编码所述异源多肽的异源多核苷酸和/或(ii)引导核酸或编码所述引导核酸的异源多核苷酸,其用于本文所公开的任何方法。主题组合物可用于体外、离体或体内修饰细胞。主题组合物可用于治疗或增强受试者的状况,如本文所公开的。

[0253] 本文公开的组合物可以包含活性成分(例如,包含工程化核酸酶的异源多肽、引导核酸等)和任选的另外的成分(例如,赋形剂)。如果需要和/或期望,可以将组合物分割、成型和/或包装成期望的单剂量或多剂量单位或者单植入或多植入单位。

[0254] 在一些实施方式中,该组合物可以包含编码本文公开的活性成分的一种或多种异源多核苷酸。当活性成分中存在不同成员时,每个成员可以由不同的异源多核苷酸编码。替代地,两种或更多种(例如,所有)成分可以由单个异源多核苷酸编码。在一些情况下,单个异源多核苷酸可编码(i)包含工程化核酸酶(例如,dCas-转录效应因子融合蛋白,如dCas-KRAB、dCas-DNMT、dCas-ADA)的异源多肽和(ii)用于靶向靶基因的特定区域或序列的一个或多个引导核酸(例如,至少1个、至少2个、至少3个、至少4个、至少5个或更多个引导核酸)。

[0255] 一个或多个异源多核苷酸还可以包含一个或多个启动子(或一个或多个转录控制元件,如本文中可互换使用的)。由一个或多个异源多核苷酸编码的不同活性成分可处于相同启动子或不同启动子的控制之下。本文公开的启动子可在真核细胞、哺乳动物细胞、非人类哺乳动物细胞或人类细胞中具有活性。启动子可以是诱导型或组成型活性启动子。替代地或另外,启动子可以是组织特异性的或细胞特异性的。合适的真核启动子(即在真核细胞中有功能的启动子)的非限制性实例可以包括来自巨细胞病毒(CMV)立即早期、单纯疱疹病毒(HSV)胸苷激酶、早期和晚期SV40、逆转录病毒的长末端重复序列(LTR)、人类延伸因子-1启动子(EF1)、包含与鸡β肌动蛋白启动子(CAG)融合的巨细胞病毒(CMV)增强子的混合构建体、鼠干细胞病毒启动子(MSCV)、磷酸甘油酸激酶-1基因座启动子(PGK)和小鼠金属硫蛋白-I的那些。启动子可以是真菌启动子。启动子可以是植物启动子。可以找到植物启动子的数据库(例如PlantProm)。表达载体还可以包含用于翻译起始的核糖体结合位点和转录终止子。表达载体还可以包括用于扩增表达的适当序列。在一些情况下,本文公开的启动子可以是本文提供的任何组织特异性的启动子,或本文提供的任何细胞类型特异性的启动子。

[0256] 一个或多个异源多核苷酸(例如,单个异源多核苷酸)中的异源多核苷酸的尺寸可以为至少或多至约2.5千碱基、至少或多至约2.6千碱基、至少或多至约2.7千碱基、至少或多至约2.8千碱基、至少或多至约2.9千碱基、至少或多至约3.0千碱基、至少或多至约3.1千碱基、至少或多至约3.2千碱基、至少或多至约3.3千碱基、至少或多至约3.4千碱基、至少或多至约3.5千碱基、至少或多至约3.6千碱基、至少或多至约3.7千碱基、至少或多至约3.8千碱基、至少或多至约3.9千碱基、至少或多至约4.0千碱基、至少或多至约4.1千碱基、至少或多至约4.2千碱基、至少或多至约4.3千碱基、至少或多至约4.4千碱基、至少或多至约4.5千

碱基、至少或多至约4.6千碱基、至少或多至约4.7千碱基、至少或多至约4.8千碱基、至少或多至约4.9千碱基、至少或多至约5.0千碱基、至少或多至约5.5千碱基、至少或多至约6.0千碱基、至少或多至约6.5千碱基、至少或多至约7.0千碱基、至少或多至约7.5千碱基、至少或多至约8.0千碱基、至少或多至约9.0千碱基、或至少或多至约10千碱基。在一些情况下,一个或多个异源多核苷酸(例如,单个异源多核苷酸)中的异源多核苷酸的尺寸可以为约3千碱基至约5千碱基、约3千碱基至约4.8千碱基、约3千碱基至约4.6千碱基、约3千碱基至约4.4千碱基、约3千碱基至约4.2千碱基、约3千碱基至约4.0千碱基、约3千碱基至约3.5千碱基、约3.5千碱基至约5千碱基、约3.5千碱基至约4.8千碱基、约3.5千碱基至约4.6千碱基、约3.5千碱基至约4.4千碱基、约3.5千碱基至约4.2千碱基、约3.5千碱基至约4千碱基、约4千碱基至约5千碱基、约4千碱基至约4.9千碱基、约4千碱基至约4.8千碱基、约4千碱基至约4.7千碱基、约4千碱基至约4.6千碱基、约4千碱基至约4.5千碱基、约4千碱基至约4.4千碱基、约4千碱基至约4.3千碱基、约4千碱基至约4.2千碱基、或约4千碱基至约4.1千碱基。

[0257] 载体(或表达盒)可以编码至少(i)Cas蛋白和(ii)包含间隔序列和支架序列的引导核酸分子,如本文所提供的。载体可以包含编码Cas蛋白的第一多核苷酸序列、编码支架序列的第二多核苷酸序列、和/或编码支架序列的第三多核苷酸序列。组合的第一多核苷酸序列的长度与第二多核苷酸序列的长度的和可以为至少或多至约1400个核苷酸、至少或多至约1420个核苷酸、至少或多至约1440个核苷酸、至少或多至约1450个核苷酸、至少或多至约1460个核苷酸、至少或多至约1480个核苷酸、至少或多至约1500个核苷酸、至少或多至约1520个核苷酸、至少或多至约1540个核苷酸、至少或多至约1550个核苷酸、至少或多至约1560个核苷酸、至少或多至约1580个核苷酸、至少或多至约1600个核苷酸、至少或多至约1620个核苷酸、至少或多至约1640个核苷酸、至少或多至约1650个核苷酸、至少或多至约1660个核苷酸、至少或多至约1680个核苷酸、至少或多至约1700个核苷酸、至少或多至约1720个核苷酸、至少或多至约1740个核苷酸、或至少或多至约1750个核苷酸。在一些实施方式中,组合的第一多核苷酸序列的长度与第二多核苷酸序列的长度的和可以少于1746个核苷酸、少于1737个核苷酸、或少于1720个核苷酸。

[0258] 在一些实施方式中,第一多核苷酸序列的长度可以为至少或多至约1400个核苷酸、至少或多至约1420个核苷酸、至少或多至约1440个核苷酸、至少或多至约1450个核苷酸、至少或多至约1460个核苷酸、至少或多至约1480个核苷酸、至少或多至约1500个核苷酸、至少或多至约1520个核苷酸、至少或多至约1540个核苷酸、至少或多至约1550个核苷酸、至少或多至约1560个核苷酸、至少或多至约1580个核苷酸、至少或多至约1600个核苷酸、至少或多至约1620个核苷酸、至少或多至约1640个核苷酸、至少或多至约1650个核苷酸、至少或多至约1660个核苷酸、至少或多至约1680个核苷酸、或至少或多至约1700个核苷酸。

[0259] 在一些实施方式中,第二多核苷酸序列的长度可以为至少或多至约80个核苷酸、至少或多至约85个核苷酸、至少或多至约90个核苷酸、至少或多至约91个核苷酸、至少或多至约92个核苷酸、至少或多至约93个核苷酸、至少或多至约94个核苷酸、至少或多至约95个核苷酸、至少或多至约96个核苷酸、至少或多至约97个核苷酸、至少或多至约98个核苷酸、至少或多至约99个核苷酸、至少或多至约100个核苷酸、至少或多至约101个核苷酸、至少或多至约102个核苷酸、至少或多至约103个核苷酸、至少或多至约104个核苷酸、至少或多至

约105个核苷酸、至少或多至约106个核苷酸、至少或多至约107个核苷酸、至少或多至约108个核苷酸、至少或多至约109个核苷酸、至少或多至约110个核苷酸、至少或多至约112个核苷酸、至少或多至约114个核苷酸、至少或多至约115个核苷酸、至少或多至约116个核苷酸、至少或多至约118个核苷酸、至少或多至约120个核苷酸、至少或多至约122个核苷酸、至少或多至约124个核苷酸、至少或多至约125个核苷酸、至少或多至约126个核苷酸、至少或多至约128个核苷酸、至少或多至约130个核苷酸、至少或多至约135个核苷酸、至少或多至约140个核苷酸、至少或多至约145个核苷酸、至少或多至约150个核苷酸、至少或多至约155个核苷酸、或至少或多至约160个核苷酸。

[0260] 在一些实施方式中,组合的第一多核苷酸序列的长度与第二多核苷酸序列的长度的和可以足够小,使得编码至少Cas蛋白和引导核酸分子的载体可以(i)是小/紧凑的和/或(ii)具有足够的空间用于额外的货物(例如,可操作地偶联至Cas蛋白的基因调节因子,或本文提供的异源多核苷酸)。即使载体具有小/紧凑尺寸,载体编码的包含Cas蛋白和引导核酸分子的复合体也可以具有功能性。在一些情况下,载体编码的复合体可以具有功能活性,以结合靶多核苷酸序列并编辑(例如,切割、缺失核苷酸、添加核苷酸、编辑碱基等)靶多核苷酸序列的至少一部分。在一些情况下,载体编码的复合体可以具有功能活性,以实现细胞中靶基因的调节的表达水平。因此,(A1)复合体导致的靶基因的调节的表达水平(替代地或另外,其活性水平)可以与(A2)包含Cas蛋白和对照引导核酸分子的对照复合体导致的相当或优于后者。

[0261] 在一些情况下,(A1)复合体导致的靶基因的调节的表达水平可以与(A2)相当,使得(A1)与(A2)相差不超过(A2)的50%、不超过45%、不超过40%、不超过35%、不超过30%、不超过25%、不超过20%、不超过15%、不超过10%、不超过8%、不超过6%、不超过5%、不超过4%、不超过3%、不超过2%或不超过1%。

[0262] 在一些情况下,靶基因的表达水平可以由本文公开的工程化核酸酶变体和/或引导核酸分子激活,并且(A1)复合体导致的靶基因的调节的表达水平可优于(A2),使得(A1)比(A2)大(A2)的至少或多至约1%、至少或多至约2%、至少或多至约5%、至少或多至约10%、至少或多至约15%、至少或多至约20%、至少或多至约30%、至少或多至约40%、至少或多至约50%、至少或多至约60%、至少或多至约70%、至少或多至约80%、至少或多至约90%、至少或多至约100%、至少或多至约150%、至少或多至约200%、至少或多至约250%、至少或多至约300%、至少或多至约350%、至少或多至约400%、至少或多至约450%、或至少或多至约500%,或者使得与(A2)相比,(A1)比(A2)大至少或多至约0.1倍、至少或多至约0.2倍、至少或多至约0.3倍、至少或多至约0.4倍、至少或多至约0.5倍、至少或多至约1倍、至少或多至约1.5倍、至少或多至约2倍、至少或多至约3倍、至少或多至约4倍、至少或多至约5倍、至少或多至约6倍、至少或多至约7倍、至少或多至约8倍、至少或多至约9倍、或至少或多至约10倍。

[0263] 在一些情况下,靶基因的表达水平可以通过本文公开的工程化核酸酶变体和/或引导核酸分子而降低(例如,阻抑),并且(A1)复合体导致的靶基因的调节的表达水平可优于(A2),使得(A1)比(A2)小(A2)的至少或多至约1%、至少或多至约2%、至少或多至约5%、至少或多至约10%、至少或多至约15%、至少或多至约20%、至少或多至约30%、至少或多至约40%、至少或多至约50%、至少或多至约60%、至少或多至约70%、至少或多至约80%、

至少或多至约90％、至少或多至约100％、至少或多至约150％、至少或多至约200％、至少或多至约250％、至少或多至约300％、至少或多至约350％、至少或多至约400％、至少或多至约450％、或至少或多至约500％，或者使得与(A2)相比，(A1)比(A2)小至少或多至约0.1倍、至少或多至约0.2倍、至少或多至约0.3倍、至少或多至约0.4倍、至少或多至约0.5倍、至少或多至约1倍、至少或多至约1.5倍、至少或多至约2倍、至少或多至约3倍、至少或多至约4倍、至少或多至约5倍、至少或多至约6倍、至少或多至约7倍、至少或多至约8倍、至少或多至约9倍、或至少或多至约10倍。

[0264] 在一些情况下，对照引导核酸分子可以比本文公开的载体编码的引导核酸分子长。对照引导核酸分子的对照支架序列可以比载体编码的引导核酸分子的支架序列长至少或多至约1个核苷酸、至少或多至约2个核苷酸、至少或多至约5个核苷酸、至少或多至约10个核苷酸、至少或多至约15个核苷酸、至少或多至约20个核苷酸、至少或多至约25个核苷酸、至少或多至约30个核苷酸、至少或多至约35个核苷酸、至少或多至约40个核苷酸、至少或多至约45个核苷酸、至少或多至约50个核苷酸、至少或多至约55个核苷酸、至少或多至约60个核苷酸、至少或多至约65个核苷酸、至少或多至约70个核苷酸、至少或多至约75个核苷酸、或至少或多至约80个核苷酸。例如，对照引导核酸分子可以包含SEQ ID NO:10的多肽序列。

[0265] 将本文提供的一种或多种异源多核苷酸递送至细胞的方法可以涉及病毒递送方法或非病毒递送方法。因此，一种或多种异源多核苷酸可以是一种或多种病毒载体(例如，一种或多种AAV载体)。替代地，一种或多种异源多核苷酸可以是与非病毒递送部分(例如阳离子脂质和/或脂质颗粒(例如，脂质纳米颗粒(LNP)))复合或被非病毒递送部分包裹的非病毒载体。

[0266] 核酸的非病毒递送方法可以包括脂质转染、核转染、微注射、基因枪、病毒颗粒、脂质体、免疫脂质体、聚阳离子或脂质:核酸缀合物、裸DNA、人工病毒体和药剂增强的DNA摄取。可以使用适合于多核苷酸的有效受体识别脂质转染的阳离子的和中性的脂质。可以递送至细胞(例如，体外或离体施用)或靶组织(例如，体内施用)。

[0267] 基于RNA或DNA病毒的系统可用于靶向体内特定细胞并将病毒有效载荷运送到细胞的核。病毒载体可直接施用(体内)，或病毒载体可用于体外处理细胞，并且可任选施用(离体)经修饰的细胞。基于病毒的系统可以包括用于基因转移的逆转录病毒、慢病毒、腺病毒、腺相关病毒和单纯疱疹病毒载体。宿主基因组中的整合可利用逆转录病毒、慢病毒和腺相关病毒基因转移方法进行，这可导致插入的转基因的长期表达。在许多不同细胞类型和靶组织中可观察到高转导效率。

[0268] 逆转录病毒的趋向性可以通过引入外来包膜蛋白来改变，从而扩大靶细胞的潜在靶群体。慢病毒载体是可转导或感染非分裂细胞并产生高病毒滴度的逆转录病毒载体。逆转录病毒基因转移系统的选择可取决于靶组织。逆转录病毒载体可以包括顺式作用长末端重复序列，具有包装高至6-10kb外来序列的能力。最小的顺式作用LTR可能足以复制和包装载体，其可用于将治疗基因整合到靶细胞中以提供永久的转基因表达。逆转录病毒载体可以包括基于鼠白血病病毒(MuLV)、长臂猿白血病病毒(GaLV)、猿猴免疫缺陷病毒(SIV)、人类免疫缺陷病毒(HIV)及其组合的逆转录病毒载体。

[0269] 可以使用基于腺病毒的系统。基于腺病毒的系统可以导致转基因的瞬时表达。基

于腺病毒的载体在细胞中可以具有高转导效率,并且可能不需要细胞分裂。使用基于腺病毒的载体可以获得高滴度和表达水平。腺相关病毒("AAV")载体可用于用靶核酸转导细胞,例如,在体外产生核酸和肽,以及用于体内和离体基因疗法程序。

[0270] 包装细胞可用于形成能够感染宿主细胞的病毒颗粒。此类细胞可以包括293细胞(例如,用于包装腺病毒)和Psi2细胞或PA317细胞(例如,用于包装逆转录病毒)。可以通过产生将核酸载体包装到病毒颗粒中的细胞系来产生病毒载体。载体可以包含包装和随后整合到宿主中所需的最小病毒序列。载体可以包含由要表达的多核苷酸的表达盒替换的其他病毒序列。缺失的病毒功能可以由包装细胞系反式提供。例如,AAV载体可以包含来自AAV基因组的ITR序列,这些序列是包装和整合到宿主基因组中所必需的。病毒DNA可以包装在细胞系中,该细胞系可以包含编码其他AAV基因(即rep和cap)的辅助质粒,但缺少ITR序列。细胞系还可以用作为辅助的腺病毒感染。辅助病毒可以促进AAV载体的复制和辅助质粒中AAV基因的表达。可以通过例如热处理减少腺病毒污染,腺病毒对热处理比AAV更敏感。

[0271] 宿主细胞可以用本文所述的一种或多种载体瞬时或非瞬时地转染。细胞可以以它在受试者中自然发生的方式被转染。细胞可以从受试者中取出或衍生并被转染。细胞可以衍生自从受试者取出的细胞,例如细胞系。在一些实施方式中,将用本文所述的一种或多种载体转染的细胞用于建立包含一种或多种载体衍生序列的新细胞系。在一些实施方式中,将用本公开的组合物瞬时转染(例如通过一种或多种载体瞬时转染或用RNA转染)并通过包含本文公开的工程化核酸酶的异源多肽的活性进行修饰的细胞用于建立包含含有修饰但缺乏任何其他外源序列的细胞的新细胞系。

[0272] 任何与宿主细胞相容的合适载体都可以与本公开的方法一起使用。用于真核宿主细胞的载体的非限制性实例包括pXT1、pSG5(Stratagene™)、pSVK3、pBPV、pMSG和pSVLSV40(Pharmacia™)。

[0273] 在一些实施方式中,本文公开的组合物的另外的成分可以包括赋形剂。赋形剂的非限制性实例可以包括溶剂、分散介质、稀释剂或其他液体媒介物、分散或悬浮助剂、表面活性剂、等渗剂、增稠剂或乳化剂、防腐剂、类脂质(lipidoids)、脂质体、脂质纳米颗粒、聚合物、脂质复合体(lipoplexes)、核壳纳米颗粒、肽、蛋白质、透明质酸酶、纳米颗粒模拟物、惰性稀释剂、缓冲剂、润滑剂、油及其组合。在一些实例中,本文公开的组合物可以包含一种或多种赋形剂,处于一定量的每种赋形剂共同增加(i)异源多肽或编码其的异源基因和/或(ii)细胞或经修饰的细胞的稳定性。

[0274] 在一些方面,本公开提供了试剂盒,所述试剂盒包含此类组合物和说明书,所述说明书指导(i)使细胞与所述组合物接触(例如,体外、离体或体内),或(ii)将包含本文公开的任一种组合物的细胞施用给受试者。受试者可能患有或可能被怀疑患有疾病,例如遗传性疾病。

[0275] 在一些实施方式中,本文公开的任何组合物都可以通过口服、腹膜内、静脉内、动脉内、经皮、肌内、脂质体、通过导管或支架局部递送、皮下、脂肪内(intraadiposally)或鞘内施用给受试者。

[0276] 本文公开的组合物(例如,药物组合物)可适合施用于人类。此外,此类组合物可适合施用于任何其他动物,例如,非人类动物,例如,非人类哺乳动物。对适合施用于人类的药物组合物进行修饰以使该组合物适合施用于各种动物是众所周知的,并且普通熟练的兽医

药理学家可以仅通过普通的(如果有的话)实验来设计和/或执行此类修饰。考虑施用所述药物组合物的受试者包括但不限于人类和/或其他灵长类动物;哺乳动物,包括商业相关哺乳动物,例如牛、猪、马、绵羊、猫、狗、小鼠和/或大鼠;和/或禽类,包括商业相关禽类,例如家禽,鸡、鸭、鹅和/或火鸡。

细胞

[0277] 在一些实施方式中,本文提供的细胞可以称为靶细胞。在一些实施方式中,本文提供的系统、组合物和方法可以用于修饰靶细胞(例如,修饰靶细胞的靶基因(例如表1中的一个或多个基因)的表达谱)。靶细胞可以包括多种细胞类型。靶细胞可以是体外的。靶细胞可以是体内的。靶细胞可以是离体的。靶细胞可以是分离的细胞。靶细胞可以是生物体内的细胞。靶细胞可以是生物体。靶细胞可以是细胞培养物中的细胞。靶细胞可以是细胞集合中的一个。靶细胞可以是哺乳动物细胞或源自哺乳动物细胞。靶细胞可以是啮齿动物细胞或源自啮齿动物细胞。靶细胞可以是人细胞或源自人细胞。靶细胞可以是原核细胞或源自原核细胞。靶细胞可以是细菌细胞或可以源自细菌细胞。靶细胞可以是古细菌细胞或源自古细菌细胞。靶细胞可以是真核细胞或源自真核细胞。靶细胞可以是多能干细胞。靶细胞可以是植物细胞或源自植物细胞。靶细胞可以是动物细胞或源自动物细胞。靶细胞可以是无脊椎动物细胞或源自无脊椎动物细胞。靶细胞可以是脊椎动物细胞或源自脊椎动物细胞。靶细胞可以是微生物细胞或源自微生物细胞。靶细胞可以是真菌细胞或源自真菌细胞。靶细胞可以来自特定器官或组织。

[0278] 靶细胞可以是干细胞或祖细胞。靶细胞可以包括干细胞(例如,成体干细胞、胚胎干细胞、诱导多能干细胞(iPS)细胞)和祖细胞(例如,心脏祖细胞、神经祖细胞等)。靶细胞可以包括哺乳动物干细胞和祖细胞,包括啮齿动物干细胞、啮齿动物祖细胞、人干细胞、人祖细胞等。克隆细胞可以包括细胞的后代。靶细胞可以包含靶核酸。靶细胞可以存在于活体生物体中。靶细胞可以是基因改造细胞。靶细胞可以是宿主细胞。

[0279] 靶细胞可以是原代细胞。例如,原代细胞的培养物可以传代0次、1次、2次、4次、5次、10次、15次或更多次。细胞可以是单细胞生物体。细胞可以在培养物中生长。

[0280] 靶细胞可以是患病细胞。患病细胞可以具有改变的代谢、基因表达和/或形态特征。患病细胞可以是癌细胞、糖尿病细胞和凋亡细胞。患病细胞可以是来自患病受试者的细胞。示例性疾病可以包括血液紊乱、癌症、代谢紊乱、眼部紊乱、器官紊乱、肌肉骨骼紊乱、心脏病等。

[0281] 如果靶细胞是原代细胞,则可以通过任何方法从个体中收获它们。例如,可以通过单采血液成分法(apheresis)、白细胞去除法(leukocytapheresis)、密度梯度分离等收获白细胞。可以通过活检收获来自如皮肤、肌肉、骨髓、脾、肝、胰腺、肺、肠、胃等组织的细胞。

[0282] 可以作为靶细胞的细胞的非限制性实例包括但不限于淋巴细胞,例如B细胞、T细胞(细胞毒性T细胞、自然杀伤T细胞、调节性T细胞、T辅助细胞)、自然杀伤细胞、细胞因子诱导的杀伤(CIK)细胞;髓系细胞,例如粒细胞(嗜碱性粒细胞、嗜酸性粒细胞、嗜中性粒细胞/多叶核嗜中性粒细胞(Hypersegmented neutrophil))、单核细胞/巨噬细胞、红细胞(网织红细胞)、肥大细胞、血栓细胞/巨核细胞、树突状细胞;来自内分泌系统的细胞,包括甲状腺(甲状腺上皮细胞、滤泡旁细胞)、甲状旁腺(甲状旁腺主细胞、嗜酸性细胞)、肾上腺(嗜铬细胞)、松果体(松果体细胞)细胞;神经系统细胞,包括胶质细胞(星形胶质细胞、小胶质细

胞)、大细胞神经分泌细胞、星状细胞、Boettcher细胞、垂体(促性腺激素细胞(Gonadotrope)、促皮质激素细胞(Corticotrope)、促甲状腺激素细胞(Thyrotrope)、促生长激素细胞(Somatotrope)、促乳激素细胞(Lactotroph));呼吸系统细胞,包括肺泡上皮细胞(Pneumocyte)(I型肺泡上皮细胞、II型肺泡上皮细胞)、克拉拉细胞(Clara cell)、杯状细胞(Goblet cell)、尘细胞(Dust cell);循环系统细胞,包括心肌细胞(Myocardiocyte)、周细胞;消化系统细胞,包括胃(胃主细胞、壁细胞)、杯状细胞、潘氏细胞(Paneth cell)、G细胞、D细胞、ECL细胞、I细胞、K细胞、S细胞;肠内分泌细胞,包括肠嗜铬细胞(enterochromaffm cell)、APUD细胞、肝(肝细胞、库普弗细胞)、软骨/骨/肌肉;骨细胞,包括成骨细胞、骨细胞、破骨细胞、牙齿(成牙骨质细胞、成釉质细胞);软骨组织细胞,包括成软骨细胞、软骨细胞;皮肤细胞,包括毛细胞(Trichocyte)、角质形成细胞、黑色素细胞(痣细胞);肌肉细胞,包括肌细胞;泌尿系统细胞,包括足状突细胞(Podocyte)、近肾小球细胞、肾小球内系膜细胞/肾小球外系膜细胞、肾近端小管刷状缘细胞(Kidney proximal tubule brush border cell)、致密斑细胞;生殖系统细胞,包括精子、塞托利细胞(Sertoli cell)、间质细胞、卵细胞;及其他细胞,包括脂肪细胞、成纤维细胞、肌腱细胞、表皮角质形成细胞(分化表皮细胞)、表皮基底细胞(干细胞)、指甲和趾甲角质形成细胞、甲床基底细胞(干细胞)、髓质毛干细胞(Medullary hair shaft cell)、皮质毛干细胞(Cortical hair shaft cell)、表皮毛干细胞(Cuticular hair shaft cell)、表皮毛根鞘细胞(Cuticular hair root sheath cell)、赫胥黎层毛根鞘细胞(Hair root sheath cell of Huxley's layer)、亨利层毛根鞘细胞(Hair root sheath cell of Henle's layer)、外部毛根鞘细胞、毛基质细胞(干细胞)、湿复层屏障上皮细胞(Wet stratified barrier epithelial cells),角膜、舌、口腔、食道、肛管、远端尿道和阴道的复层鳞状上皮的表面上皮细胞,角膜、舌、口腔、食道、肛管、远端尿道和阴道上皮的基底细胞(干细胞),泌尿上皮细胞(膀胱和尿道内衬)、外分泌分泌上皮细胞、唾液腺粘液细胞(富含多糖的分泌物)、唾液腺浆液细胞(富含糖蛋白酶的分泌物)、舌内Von Ebner腺细胞(冲洗味蕾)、乳腺细胞(乳汁分泌)、泪腺细胞(泪水分泌)、耳内耵聍腺细胞(分泌耵聍)、小汗腺暗细胞(分泌糖蛋白)、小汗腺透明细胞(小分子分泌)。顶泌汗腺细胞(有气味的分泌,对性激素敏感)、眼睑中的Moll腺细胞(特化汗腺)、皮脂腺细胞(富含脂质的皮脂分泌)、鼻中的Bowman腺细胞(冲洗嗅觉上皮)、十二指肠中的Brunner腺细胞(酶和碱性粘液)、精囊细胞(分泌精液成分,包括游动精子的果糖)、前列腺细胞(分泌精液成分)、尿道球腺细胞(粘液分泌)、巴氏腺细胞(阴道润滑剂分泌)、Littre腺细胞(粘液分泌)、子宫内膜细胞(碳水化合物分泌)、呼吸道和消化道分离的杯状细胞(粘液分泌)、胃壁粘液细胞(粘液分泌)、胃腺酶原细胞(胃蛋白酶原分泌)、胃腺泌酸细胞(盐酸分泌)、胰腺腺泡细胞(碳酸氢盐和消化酶分泌)、小肠潘氏细胞(溶菌酶分泌)、肺Ⅱ型肺泡细胞(表面活性物质分泌)、肺克拉拉细胞、激素分泌细胞、垂体前叶细胞、促生长激素细胞、促乳激素细胞、促甲状腺细胞、促性腺激素细胞、促皮质激素细胞、垂体中叶细胞、大细胞神经分泌细胞、肠道及呼吸道细胞、甲状腺细胞、甲状腺上皮细胞、滤泡旁细胞、甲状旁腺细胞、甲状旁腺主细胞、嗜酸性细胞、肾上腺细胞、嗜铬细胞、睾丸Ley dig细胞、卵巢卵泡内膜细胞、破裂的卵巢卵泡黄体细胞、颗粒黄体细胞、膜黄体细胞、肾小球旁细胞(肾素分泌)、肾脏的致密斑细胞、代谢和储存细胞、屏障功能细胞(肺、肠、外分泌腺和泌尿生殖道)、肾、I型肺泡细胞(肺气腔内衬)、胰管细胞(泡心细胞)、无横纹管细胞(汗腺、唾液腺、乳

腺等)、管道细胞(精囊、前列腺等)、封闭的内部体腔内衬上皮细胞、具有推进功能的纤毛细胞、细胞外基质分泌细胞、收缩细胞;骨骼肌细胞、干细胞、心肌细胞、血液和免疫系统细胞、红血细胞(红细胞)、巨核细胞(血小板前体)、单核细胞、结缔组织巨噬细胞(各种类型)、表皮朗格汉斯细胞、破骨细胞(骨骼中)、树突状细胞(淋巴组织中)、小胶质细胞(中枢神经系统中)、嗜中性粒细胞、嗜酸性粒细胞、嗜碱性粒细胞、肥大细胞、辅助性T细胞、抑制性T细胞、细胞毒性T细胞、自然杀伤T细胞、B细胞、自然杀伤细胞、网织红细胞、血液和免疫系统的干细胞和定向祖细胞(各种类型)、多能干细胞、全能干细胞、诱导多能干细胞、成体干细胞、感觉传感细胞、自主神经元细胞、感觉器官和周围神经元支持细胞、中枢神经系统神经元和胶质细胞、晶状体细胞、色素细胞、黑素细胞、视网膜色素上皮细胞、生殖细胞、卵原细胞/卵母细胞、精子细胞、精母细胞、精原细胞(精母细胞的干细胞)、精子、护理细胞、卵泡细胞、塞托利细胞(睾丸中)、胸腺上皮细胞、间质细胞和间质肾细胞。

[0283] 细胞(或靶细胞)可被工程化以包含(或表现出)本文公开的任一系统或组合物,或者可以通过本文公开的任一方法在体外或离体进行处理,然后施用于受试者,例如,以治疗受试者的病症。例如,可将任何受试者修饰细胞产品施用于受试者,以治疗受试者的身体组织病症。在一些情况下,细胞可驻留在受试者体内,并且可将任何其系统或组合物施用于受试者,以通过系统/组合物接触细胞(例如,用系统/组合物对细胞进行工程化)。

实施例

[0284] 实施例1:工程化核酸酶

[0285] 由SEQ ID NO:1的多肽序列编码的Cas蛋白具有529个氨基酸残基的尺寸。当在载体(例如病毒载体,如AAV载体)中递送编码Cas蛋白的基因时,减小此类Cas蛋白的尺寸可在载体(例如,具有载货尺寸或长度限制的病毒载体)内提供更多载货空间。载体内增加的载货空间可用于递送(例如,编码)至少一种另外的组分(例如,一种或多种异源基因效应因子、一种或多种引导核酸分子、一种或多种用于治疗基因递送的cDNA等),以实现期望的结果(例如,治疗效果)。替代地或另外,当在递送媒介物(例如,脂质纳米颗粒、病毒衣壳等)中递送本文公开的Cas蛋白的重组版本时,减小Cas蛋白的尺寸可提供更多载货空间以容纳例如至少一种另外的组分,以实现期望的结果。不希望受理论的束缚,与不同的对照核酸酶相比,使用本文公开的工程化核酸酶以及至少一种另外的组分可以增强其活性(例如,其靶向基因结合、切割、编辑和/或调节)。

[0286] 在一些实施方式中,在整个本公开的实施例中,可以评估(例如,在体外)本公开的一种或多种工程化核酸酶以评估一种或多种工程化核酸酶中的每一种在结合、切割和/或编辑靶多核苷酸序列中的活性,例如,调控由靶多核苷酸序列编码或可操作地偶联至靶多核苷酸序列的多肽(例如,蛋白质)的靶多核苷酸序列的表达和/或活性水平。在一些实例中,可以在细胞中(例如,在体外)测试包含工程化核酸酶的异源多肽和包含针对靶多核苷酸的引导核酸(例如,sgRNA)的异源多核苷酸,以评估靶多核苷酸处的基因敲除效率。在一些实例中,可以在细胞中(例如,在体外)测试包含工程化核酸酶的核酸酶缺陷变体(其偶联(例如,融合)至对于工程化核酸酶而言异源的基因效应因子(例如,基因激活因子、基因阻抑因子))的异源多肽以及包含针对靶多核苷酸的引导核酸(例如,sgRNA)的异源多核苷酸,以评估调控偶联至(或包含)靶多核苷酸的基因的表达和/或活性水平的能力。

[0287] 在一些实施方式中,在整个本公开的实施例中,可以通过跨越缺失区域的相反朝

向的引物利用全质粒扩增来生成多个工程化核酸酶候选物的文库。

实施例2:基于结构比较的工程化核酸酶

[0289] A.方法

[0290] 在一些实施方式中,CasMini或失活CasMini(dCasMini)(例如SEQ ID NO:1的多肽序列的突变变体)的尺寸(例如氨基酸残基的数量)可以进一步减小,同时保持或增强其活性(例如整体表观遗传基因调控活性)。例如,小于SEQ ID NO:1的Cas蛋白的其他Cas蛋白(例如,天然存在的Cas12f蛋白,如Un2Cas12f1(SEQ ID NO:2)或AsCas12f(SEQ ID NO:3))可作为参考点,以确定SEQ ID NO:1的Cas蛋白的可以被修饰(例如,缺失)的至少一个氨基酸残基和/或至少一个三级结构,例如使其活性受到的损害最小或不受影响。

[0291] 在一些实施方式中,可以通过产生至少一个缺失来对由SEQ ID NO:1编码的天然Un1Cas12f1核酸酶进行工程化,以产生本公开的工程化核酸酶,例如由于实施例1中描述的一个或多个原因。

[0292] 在一些情况下,可以在天然Un1Cas12f1核酸酶的一个或多个在结构上不与AsCas12f(SEQ ID NO:3)对齐的区域中存在工程化核酸酶的氨基酸序列的至少一个缺失。与AsCas12f相比时,Un1Cas12f1包含另外的氨基酸残基1-71结构域,并且可以去除该结构域的至少一部分和/或该结构域附近的一个或多个氨基酸残基,例如,使得工程化核酸酶活性(例如,工程化核酸酶与引导核酸分子(例如,crRNA)的相互作用)的降低最小或基本上不降低。如图1所示,Un1Cas12f1包含可能在AsCas12f中不保守的结构域110(例如,包含另外的氨基酸残基1-71结构域的至少一部分)。另一方面,Un1Cas12f1包含可能在AsCas12f中保守的结构域120。

[0293] 在一些情况下,可以在天然Un1Cas12f1核酸酶的一个或多个在结构上不与Un2Cas12f1(SEQ ID NO:2)对齐的区域中存在工程化核酸酶的氨基酸序列的至少一个缺失。与Un2Cas12f1相比时,Un1Cas12f1包含另外的氨基酸残基41-71结构域,并且可以去除该结构域的至少一部分和/或该结构域附近的一个或多个氨基酸残基,例如,使得工程化核酸酶活性(例如,工程化核酸酶与引导核酸分子(例如,crRNA)的相互作用)的降低最小或基本上不降低。如图2所示,Un1Cas12f1包含可能在Un2Cas12f1中不保守的结构域210(例如,包含另外的氨基酸残基41-71结构域的至少一部分)。另一方面,Un1Cas12f1包含可能在Un2Cas12f1中保守的结构域220。

[0294] B.示例性文库设计

[0295] 图3A示意性地示出了如SEQ ID NO:1编码的Un1Cas12f1的不同区域以及可以至少部分缺失以产生一种或多种本文公开的工程化核酸酶的示例性结构域(310、320和330)。例如,结构域310可以涉及N端缺失(例如,与SEQ ID NO:1比对时氨基酸残基2-76),可以通过从蛋白质的N端一次增量去除3个氨基酸来产生25个变体来评估。在另一实例中,结构域320可以涉及N端部分缺失(例如,与SEQ ID NO:1比对时的氨基酸残基41-71),可以通过从这个区域的中间每次增量去除额外的2个氨基酸来产生16个变体来评估(例如,del55-56、del54-57、del54-58、del53-59、del52-60、del51-61、del50-62、del49-63、del48-64、del47-65、del46-66、del45-67、del44-68、del43-69、del42-70、del41-71等)。在不同的实例中,结构域330可以涉及C端缺失(例如,与SEQ ID NO:1比对时的最后75个氨基酸残基),可以通过从C端一次增量去除3个氨基酸来产生25个变体来评估。

[0296] 根据本公开产生的具有降低的核酸酶活性的工程化核酸酶的实例可以包括如本文提供的SEQ ID NO:4-9。

[0297] SEQ ID NO:4(包含N端缺失,例如,与SEQ ID NO:1比对时的氨基酸残基2-21(AKNTITKTLKLRIVRPYNSA))

```
  1 MEVEKIVADE KNNREKIALE KNKDKVKEAC SKHLKVAAYC TTQVERNACL
 51 FCKARKLDDK FYQKLRGQFP DAVFWQEISE IFRQLQKQAA EIYNQSLIEL
101 YYEIFIKGKG IANASSVEHY LSRVCYRRAA ELFKNAAIAS GLRSKIKSNF
151 RLKELKNMKS GLPTTKSDNF PIPLVKQKGG QYTGFEISNH NSDFIIKIPF
201 GRWQVKKEID KYRPWEKFDF EQVQKSPKPI SLLLSTQRRK RNKGWSKDEG
251 TEAEIKKVMN GDYQTSYIEV KRGSKICEKS AWMLNLSIDV PKIDKGVDPS
301 IIGGIAVGVR SPLVCAINNA FSRYSISDND LFHFNKKMFA RRRILLKKNR
351 HKRAGHGAKN KLKPITILTE KSERFRKKLI ERWACEIADF FIKNKVGTVQ
401 MENLESMKRK EDSYFNIRLR GFWPYAEMQN KIEFKLKQYG IEIRKVAPNN
451 TSKTCSKCGH LNNYFNFEYR KKNKFPHFKC EKCNFKENAA YNAALNISNP
501 KLKSTKERP
```

[0298] SEQ ID NO:5(包含N端缺失,例如与SEQ ID NO:1比对时的氨基酸残基2-31(AKNTITKTLKLRIVRPYNSAEVEKIVADEK))

```
  1 MNNREKIALE KNKDKVKEAC SKHLKVAAYC TTQVERNACL FCKARKLDDK
 51 FYQKLRGQFP DAVFWQEISE IFRQLQKQAA EIYNQSLIEL YYEIFIKGKG
101 IANASSVEHY LSRVCYRRAA ELFKNAAIAS GLRSKIKSNF RLKELKNMKS
151 GLPTTKSDNF PIPLVKQKGG QYTGFEISNH NSDFIIKIPF GRWQVKKEID
201 KYRPWEKFDF EQVQKSPKPI SLLLSTQRRK RNKGWSKDEG TEAEIKKVMN
251 GDYQTSYIEV KRGSKICEKS AWMLNLSIDV PKIDKGVDPS IIGGIAVGVR
301 SPLVCAINNA FSRYSISDND LFHFNKKMFA RRRILLKKNR HKRAGHGAKN
351 KLKPITILTE KSERFRKKLI ERWACEIADF FIKNKVGTVQ MENLESMKRK
401 EDSYFNIRLR GFWPYAEMQN KIEFKLKQYG IEIRKVAPNN TSKTCSKCGH
451 LNNYFNFEYR KKNKFPHFKC EKCNFKENAA YNAALNISNP KLKSTKERP
```

[0299] SEQ ID NO:6(包含C端缺失,例如与SEQ ID NO:1比对时的氨基酸残基510-529(DYNAALNISNPKLKSTKEEP))

```
  1 MAKNTITKTL KLRIVRPYNS AEVEKIVADE KNNREKIALE KNKDKVKEAC
 51 SKHLKVAAYC TTQVERNACL FCKARKLDDK FYQKLRGQFP DAVFWQEISE
101 IFRQLQKQAA EIYNQSLIEL YYEIFIKGKG IANASSVEHY LSRVCYRRAA
151 ELFKNAAIAS GLRSKIKSNF RLKELKNMKS GLPTTKSDNF PIPLVKQKGG
201 QYTGFEISNH NSDFIIKIPF GRWQVKKEID KYRPWEKFDF EQVQKSPKPI
251 SLLLSTQRRK RNKGWSKDEG TEAEIKKVMN GDYQTSYIEV KRGSKICEKS
301 AWMLNLSIDV PKIDKGVDPS IIGGIAVGVR SPLVCAINNA FSRYSISDND
351 LFHFNKKMFA RRRILLKKNR HKRAGHGAKN KLKPITILTE KSERFRKKLI
401 ERWACEIADF FIKNKVGTVQ MENLESMKRK EDSYFNIRLR GFWPYAEMQN
451 KIEFKLKQYG IEIRKVAPNN TSKTCSKCGH LNNYFNFEYR KKNKFPHFKC
501 EKCNFKENA
```

[0300] SEQ ID NO:7(包含C端缺失,例如与SEQ ID NO:1比对时的氨基酸残基500-529(CEKCNFKENADYNAALNISNPKLKSTKEEP))

```
  1 MAKNTITKTL KLRIVRPYNS AEVEKIVADE KNNREKIALE KNKDKVKEAC
 51 SKHLKVAAYC TTQVERNACL FCKARKLDDK FYQKLRGQFP DAVFWQEISE
101 IFRQLQKQAA EIYNQSLIEL YYEIFIKGKG IANASSVEHY LSRVCYRRAA
151 ELFKNAAIAS GLRSKIKSNF RLKELKNMKS GLPTTKSDNF PIPLVKQKGG
201 QYTGFEISNH NSDFIIKIPF GRWQVKKEID KYRPWEKFDF EQVQKSPKPI
251 SLLLSTQRRK RNKGWSKDEG TEAEIKKVMN GDYQTSYIEV KRGSKICEKS
301 AWMLNLSIDV PKIDKGVDPS IIGGIAVGVR SPLVCAINNA FSRYSISDND
351 LFHFNKKMFA RRRILLKKNR HKRAGHGAKN KLKPITILTE KSERFRKKLI
401 ERWACEIADF FIKNKVGTVQ MENLESMKRK EDSYFNIRLR GFWPYAEMQN
451 KIEFKLKQYG IEIRKVAPNN TSKTCSKCGH LNNYFNFEYR KKNKFPHFK
```

[0301] SEQ ID NO:8(包含部分N端缺失,例如与SEQ ID NO:1比对时的氨基酸残基47-66(KEACSKHLKVAAYCTTQVER))

```
  1 MAKNTITKTL KLRIVRPYNS AEVEKIVADE KNNREKIALE KNKDKVNACL
 51 FCKARKLDDK FYQKLRGQFP DAVFWQEISE IFRQLQKQAA EIYNQSLIEL
101 YYEIFIKGKG IANASSVEHY LSRVCYRRAA ELFKNAAIAS GLRSKIKSNF
151 RLKELKNMKS GLPTTKSDNF PIPLVKQKGG QYTGFEISNH NSDFIIKIPF
201 GRWQVKKEID KYRPWEKFDF EQVQKSPKPI SLLLSTQRRK RNKGWSKDEG
251 TEAEIKKVMN GDYQTSYIEV KRGSKICEKS AWMLNLSIDV PKIDKGVDPS
301 IIGGIAVGVR SPLVCAINNA FSRYSISDND LFHFNKKMFA RRRILLKKNR
351 HKRAGHGAKN KLKPITILTE KSERFRKKLI ERWACEIADF FIKNKVGTVQ
401 MENLESMKRK EDSYFNIRLR GFWPYAEMQN KIEFKLKQYG IEIRKVAPNN
451 TSKTCSKCGH LNNYFNFEYR KKNKFPHFKC EKCNFKENAA YNAALNISNP
501 KLKSTKERP
```

[0302] SEQ ID NO:9(包含部分N端缺失,例如与SEQ ID NO:1比对时的氨基酸残基41-71(KNKDKVKEACSKHLKVAAYCTTQVERNACLF))

```
  1 MAKNTITKTL KLRIVRPYNS AEVEKIVADE KNNREKIALE CKARKLDDKF
 51 YQKLRGQFPD AVFWQEISEI FRQLQKQAAE IYNQSLIELY YEIFIKGKGI
101 ANASSVEHYL SRVCYRRAAE LFKNAAIASG LRSKIKSNFR LKELKNMKSG
151 LPTTKSDNFP IPLVKQKGGQ YTGFEISNHN SDFIIKIPFG RWQVKKEIDK
201 YRPWEKFDFE QVQKSPKPIS LLLSTQRRKR NKGWSKDEGT EAEIKKVMNG
251 DYQTSYIEVK RGSKICEKSA WMLNLSIDVP KIDKGVDPSI IGGIAVGVRS
301 PLVCAINNAF SRYSISDNDL FHFNKKMFAR RRILLKKNRH KRAGHGAKNK
351 LKPITILTEK SERFRKKLIE RWACEIADFF IKNKVGTVQM ENLESMKRKE
401 DSYFNIRLRG FWPYAEMQNK IEFKLKQYGI EIRKVAPNNT SKTCSKCGHL
451 NNYFNFEYRK KNKFPHFKCE KCNFKENAAY NAALNISNPK LKSTKERP
```

[0303] 实施例3:基于缺失景观的工程化核酸酶

[0304] 可以通过产生至少一个缺失来对由SEQ ID NO:1编码的天然Un1Cas12f1核酸酶进行工程化,以产生本公开的工程化核酸酶,例如由于实施例1中描述的一个或多个原因。如本文所公开的,工程化核酸酶的氨基酸序列的至少一个缺失可以基于Cas蛋白(或其核酸酶缺陷变体)的缺失(或截短)景观来确定。

[0305] 例如,如图4所示,dCas9的截短景观的映射表明对显著缺失(例如dCas9的氨基酸残基167-316的缺失)的耐受性。因此,可以利用类似的缺失景观来确定天然Un1Cas12f1(或其核酸酶缺陷变体)的一个或多个氨基酸残基(例如,多个连续的氨基酸残基),以产生Un1Cas12f1蛋白的缺失突变体,而不显著影响活性(例如,不影响活性)。在一些情况下,可以确定约20个氨基酸残基的缺失(例如,来自一个连续区域,或来自两个非连续区域)以产生Un1Cas12f1的突变变体,其与野生型Un1Cas12f1相比仍表现出相当(或增强)的活性。在一些情况下,可以评估整个天然Un1Cas12f1蛋白的小缺失窗口(例如,约3至约5个氨基酸的窗口),以用于缺失景观。

[0306] 图3B示意性地说明了用于产生Un1Cas12f1的工程化核酸酶变体的缺失景观方法。可以扫描贯穿Un1Cas12f1的至少一部分(例如,整个Un1Cas12f1,除了二聚化结构域和/或PAM结构域)中的多个氨基酸残基(例如,多个连续氨基酸残基,例如5个连续氨基酸残基)的缺失块510,以生成例如多至95种变体用于个体转染筛选。

[0307] 实施例4:基于缺失景观的工程化核酸酶

[0308] 除了从由SEQ ID NO:1编码的天然Un1Cas12f1核酸酶中缺失一个或多个氨基酸残基以产生工程化核酸酶之外,与天然Un1Cas12f1核酸酶相比,可以使一个或多个氨基酸残基突变,以进一步修饰本公开的工程化核酸酶,例如由于实施例1中描述的一个或多个原因。如本文所公开的,工程化核酸酶的氨基酸序列的至少一个缺失可以基于位点饱和诱变(SSM)来确定。例如,编码天然Un1Cas12f1蛋白(或其核酸酶缺陷变体)的多核苷酸的单个密码子或一组密码子可以在增强活性(例如,为了表观遗传调控的改善,如增强的激活/阻抑、不同的PAM识别等)、稳定性、表达、与相应的引导核酸分子结合等的位置上被一个或多个可能的氨基酸置换。

[0309] 实施例5:工程化核酸酶的评估

[0310] 本文公开的工程化核酸酶的核酸酶缺陷变体(例如,根据实施例1-实施例4产生)可以偶联(例如,融合)至异源基因效应因子(例如,VP16、VP64、p65、Rta、VPR等),并且可以

将编码此类融合蛋白和/或引导RNA的一种或多种异源多肽转染到细胞中以评估靶基因的表达/活性水平的激活。例如,可以使用带有合成报告子的工程化HEK93T细胞,其中可以测量荧光激活作为读出。当鉴定出多个工程化核酸酶时,可以重复筛选方法(例如,通过使用工程化HEK93T细胞)以(i)确认先前的筛选结果和/或(ii)鉴定最高命中。

[0311] 根据第一轮工程化的结果,可以组合多个氨基酸残基的缺失(例如,当与SEQ ID NO:1的多肽序列比对时)以产生一种或多种额外的工程化核酸酶。替代地或另外,可以通过评估第一轮工程化中的一个或多个先导周围的一个或多个颗粒缺失来产生一种或多种额外的工程化核酸酶。

[0312] A.第一轮工程化

[0313] dCasMINI(SEQ ID NO:10)与在细菌中具有报告的核酸酶活性的一种或多种天然存在的Cas12f蛋白直系同源物(例如,Un1Cas12f1、Un2Cas12f1、AsCas12f和表2中提供的其他直系同源物)的多序列比对,例如,鉴定在产生本文公开的工程化核酸酶变体中的一种或多种潜在有益的突变。还对Cas12f直系同源物进行了同源性建模,以鉴定结构保守性(例如,并非完全基于基于氨基酸的保守性),从而生成工程化核酸酶变体。结合这两种方法(例如,序列比对和结构保守性分析),氨基酸序列的延伸以及单个残基突变的组合被确定为产生所述工程化核酸酶变体的有希望的候选者。基于所鉴定的信息,通过序列交换设计嵌合蛋白变体,并且通过一个或多个残基突变设计突变变体。在一些情况下,序列交换变体还导致整体减小的蛋白质尺寸。

[0314] 在第一轮工程化中,生成、测试和筛选工程化核酸酶变体(例如,截短变体、嵌合蛋白变体和/或突变变体)以鉴定先导命中。例如,设计了多种具有降低的核酸酶活性的嵌合工程化核酸酶变体(如表3B中提供)并与基因调节因子(例如基因激活因子或基因阻抑因子)组合(例如融合)进行测试,以测试它们与对照dCasMINI(SEQ ID NO:10)相比在基因调节中的功效。

[0315] 通过将每种工程化核酸酶变体与基因激活因子融合来评估基因激活。将具有降低的核酸酶活性的每种工程化核酸酶变体单独与基因激活因子(例如VPR)在框内克隆到dCas质粒中以用于转录激活。用相同的sgRNA质粒和个体的dCas变体质粒一式三份或一式四份重复在96孔板格式中转染HEK293T细胞。转染后几天之后(例如,2天之后、3天之后或4天之后),通过活细胞的细胞表面抗体染色然后进行流式细胞术来量化CD2蛋白水平,同时使用ELISA测量细胞培养上清液中的分泌的IFNγ(IFNγ)蛋白水平。将每种工程化核酸酶变体的基因激活水平与作为对照的与相同基因激活因子融合的dCasMINI的活性进行比较。

[0316] 通过将每种工程化核酸酶变体与基因阻抑因子融合来评估基因阻抑。将具有降低的核酸酶活性的每种工程化核酸酶变体单独与基因阻抑因子在框内克隆到dCas质粒中,用于转录抑制。在96孔板格式中,用相同的sgRNA质粒和个体的dCas变体质粒一式三份或一式四份重复转染HEK293T GFP报告细胞(例如ESR221)。转染后几天之后(例如5至7天),通过流式细胞术测量GFP表达的抑制。将每种工程化核酸酶变体的基因阻抑水平与作为对照的与相同基因激活因子融合的dCasMINI的活性进行比较。

[0317] 如图5所示,在嵌合工程化核酸酶变体cA1至cA9中,嵌合工程化核酸酶变体cA2导致最大的IFNγ表达激活(例如,大于dCasMINI对照的激活)以及CD2激活(如表3A中总结的)。设计了相对于Un1Cas12f1具有一个或多个突变的另外的工程化核酸酶变体,而与

dCasMINI相比,工程化核酸酶变体的整体尺寸没有显著改变。参见表4B中此类工程化核酸酶变体的氨基酸序列。如图5所示,当与基因激活因子融合时,多种工程化核酸酶(例如,mD2、mD4)在激活靶基因(例如,IFNγ或CD2)方面表现出与对照dCasMINI相比相当或更高的功效(如表4A中总结的)。如图6所示,多种工程化核酸酶中的一些(例如,mD2、mD4)在阻抑靶基因(例如,eGFP)方面表现出与对照dCasMINI相比相当或更高的功效(如表4A中总结的)。还鉴定了另一种工程化核酸酶变体(t1),其与SEQ ID NO:1相比包含C端区域上的缺失,但不牺牲基因调节活性。如图15所示,当与基因激活因子融合时,工程化核酸酶变体t1在激活靶基因(例如,IFNγ)方面表现出与对照dCasMINI相比相当的功效。

[0318] 设计了嵌合工程化核酸酶变体cA2,其在锌结合基序处具有序列缺失。不希望受理论的束缚,在特定锌结合基序处的此类缺失导致较小的核酸酶变体,当与一种或多种基因调节因子可操作地连接(例如,融合)时,其维持或甚至改善了基因调节活性(例如,表观遗传修饰活性)。

[0319] 表5A总结了嵌入在每个工程化核酸酶变体中的相对于Un1Cas12f1(SEQ ID NO:1)的氨基酸序列的每组修饰,与dCasMINI(SEQ ID NO:10)相比,其产生增强的基因调节活性。

[0320] 不希望受理论的束缚,可以对来自表3A、表4A、表5A和表5B中的任何工程化核酸酶变体进行一个或多个额外的修饰,以增强期望的活性(例如,与引导核酸分子形成复合体、靶序列核酸酶活性、增加靶基因的表达、降低靶基因的表达等)。

[0321] B.第二轮工程化

[0322] 在第二轮工程化中,选择工程化核酸酶变体“cA2”(SEQ ID NO:12)作为支架,通过将表5A中鉴定和列出的突变和/或截短的一个或多个组合移植到cA2支架上来生成新的工程化核酸酶变体文库。表5B显示了新文库中的工程化核酸酶变体列表,表明了已移植到每个工程化核酸酶变体上的修饰的组合。表5C显示了表5B中列出的每个工程化核酸酶变体的相应氨基酸序列。如图7所示,在与基因激活因子(例如VPR)组合的情况下测试其激活靶基因(例如CD2、IFNγ和CXCR4)的能力时,多种工程化核酸酶变体表现出与对照dCasMINI相当或增强的基因激活,一些变体的表现优于起始支架cA2(例如cA2.6、cA2.39、cA2.69、cA2.29、cA2.10、cA2.4、cA2.21、cA2.13、cA2.3、cA2.16、cA2.23、cA2.8、cA2.31、cA2.30、cA2.11、cA2.5、cA2.41、cA2.49、cA2.26、cA2.14、cA2.20、cA2.1、cA2.24、cA2.58、cA2.61、cA2.38、cA2.88、cA2.2、cA2.51、cA2.34、cA2.25、cA2.85、cA2.54、cA2.15、cA2.75、cA2.32、cA2.90、cA2.89或cA2.46)。相比之下,一些新工程化的核酸酶变体与对照dCasMINI相比没有表现出增强的活性(例如,cA2.55、cA2.84)。如图16所示,在与基因阻抑调节因子(例如ZNF10-KRAB-hDNMT3L)组合的情况下测试其抑制靶基因(例如CXCR4)的能力时,多种工程化核酸酶变体表现出与对照dCasMINI相当或增强的基因抑制,其中一些工程化核酸酶变体的表现与起始支架cA2相当(例如cA2.69、cA.2.29、cA2.4、cA2.2、cA2.34或cA2.7)。相比之下,一些新工程化核酸酶与对照dCasMINI相比没有表现出增强的基因抑制活性(例如cA2.26、cA2.63)。

[0323] 总而言之,本文将两个序列缺失区域鉴定为紧凑型Cas变体。还鉴定了额外的点突变,这些点突变在与基因调节因子和引导核酸分子组合使用时有助于改善基因调节活性。结合这些序列修饰,cA2蛋白(SEQ ID NO:12)及其变体(具有额外的点突变和序列缺失,参见表5C)在多个内源基因座上的转录激活和抑制活性方面均优于对照dCasMINI。

[0324] 不希望受理论的束缚,本文中鉴定的用于增强工程化核酸酶变体的基因调节功效的任何氨基酸缺失、结构缺失或氨基酸修饰(例如,突变/置换)都可以“移植”回任何其他天然或非天然存在的Cas蛋白(例如,选自表2的天然存在的Cas蛋白或其失活的核酸酶变体)中,以产生一种或多种额外的工程化核酸酶变体,其可以比对照Cas蛋白(例如,本文公开的dCasMINI)产生更高的基因调节功效。

[0325] 实施例6:工程化核酸酶的评估

[0326] 可在细胞中测试保持了核酸酶活性的至少一部分(例如,与SEQ ID NO:1相比)的本文公开(例如,根据实施例1-实施例4产生)的工程化核酸酶的工程化核酸酶变体的核酸酶活性。可将包含工程化核酸酶变体和/或引导RNA的多肽转染到细胞中,以评估包含工程化核酸酶和引导RNA的复合体在靶多核苷酸序列中产生断裂的能力(例如,在包含适当PAM的靶多核苷酸序列中或附近产生双链断裂)。可以通过体外酶促测定(例如,用纯化的酶或细胞裂解物)或在体内(例如,在大肠杆菌中或真核细胞中)观察此类核酸酶活性。

[0327] 实施例7:引导核酸支架工程化

[0328] Cas蛋白可以包含天然存在的Cas蛋白(例如,(i)SEQ ID NO:1的多肽序列或(ii)选自表2的任一种Cas蛋白的多肽序列)或其任何修饰(例如,包含本文公开的工程化核酸酶变体的工程化多肽)。Cas蛋白可以与引导核酸分子(例如,sgRNA)形成复合体(例如,核糖核蛋白(RNP)复合体),该复合体可结合靶多核苷酸序列以调节靶基因。引导核酸分子可以包含(i)支架序列,以至少与Cas蛋白形成复合体;和(ii)间隔序列,该间隔序列表现出与靶多核苷酸序列的至少部分序列互补性。不希望受理论的束缚,产生更小和更紧凑的引导核酸支架序列(例如,与dCasMINI的相比)可以增强复合体的整体生物活性(例如,增强的靶基因调节、减少的脱靶效应等)。

[0329] 在整个本实施例中使用的对照引导核酸表示为“SQ”(参见图8中的示意图和表6B中的其支架多核苷酸序列)。对照引导核酸SQ长179个核苷酸(nt),包括20nt间隔序列。

[0330] 不希望受理论的束缚,可以对来自引导核酸序列(例如,SQ)的以下区域中的一个或多个进行修饰(例如,突变或缺失),同时保持所得RNP复合体的生物活性或改善此类生物活性。如图9所示,此类区域可以包括支架序列内结构失向(disoriented)和/或溶剂暴露的环(例如,茎环)。对于SQ引导核酸序列,此类区域的非限制性实例可以包括区域1(例如,1-23nt)、区域2(例如,35-61nt)和区域3(例如,138-143nt),每个nt位置相对于SQ的多核苷酸序列。

[0331] A.截短的引导核酸分子

[0332] 生成引导RNA支架变体,以包含上述讨论的三个区域中的逐步缺失的组合。对于区域1,进行逐步碱基对修剪。对于区域2,除了去除失向茎环外,我们还安装了RNA中常见的稳定发夹结构(例如,cUUCGg)。对于区域3及其相邻序列,进行了一个或多个缺失(例如,136-149nt的缺失),并且进行了其他序列的额外逐步截断,以进一步减小区域3附近的支架尺寸。

[0333] 通过测试的gRNA支架变体和与基因阻抑因子融合的工程化核酸酶变体来评估基因阻抑。将每个gRNA支架变体单独克隆到具有靶向感兴趣的基因的相同间隔序列的sgRNA质粒中。在96孔板格式中,将HEK293T GFP报告细胞ESR221(在每个孔中)用相同的dCas质粒(例如,编码基因抑制调节因子,例如与基因阻抑因子(例如KRAB)融合的dCas(例如本文公

开的dCasMINI))和个体的sgRNA变体质粒一式三份或一式四份重复转染。转染后五至七天后,通过流式细胞术测量GFP表达的抑制。将每个gRNA支架变体的GFP抑制水平与SQ对照gRNA的活性进行比较。

[0334] 通过测试gRNA支架变体和与基因激活因子融合的工程化核酸酶变体来评估基因激活。将每个gRNA支架变体单独克隆到具有靶向感兴趣的基因的相同间隔序列的sgRNA质粒中。在96孔板格式中,将HEK293T细胞用相同的dCas质粒(例如,编码基因激活调节因子,例如与基因激活因子(例如VPR)融合的dCas(例如本文公开的dCasMINI))和个体的sgRNA变体质粒一式三份或一式四份转染。转染后二至四天,测量靶基因激活(例如,通过活细胞的细胞表面抗体染色然后用流式细胞术量化CD2蛋白水平,而分泌的IFN γ 蛋白水平使用ELISA对细胞培养上清液进行测量)。将每个gRNA支架变体的基因激活水平与SQ对照gRNA的活性进行比较。

[0335] B.第一轮

[0336] 在第一轮中,用靶向合成GFP启动子区域的相同间隔序列和用相同基因抑制调节因子测试表6B中提供的gRNA支架变体(例如,SEQ ID NO:501-554和600)和对照gRNA支架序列SQ(SEQ ID NO:500)。转染后第5天,通过流式细胞术量化GFP表达水平并将其相对于阴性对照(具有非靶向间隔序列的SQ支架)进行归一化。如图10所示并总结在表6A中,小于SQ阳性对照的多种gRNA支架变体表现出(i)与SQ阳性对照相当的基因阻抑或(ii)比SQ阳性对照更大的基因阻抑(例如,SEQ ID NO:503-152、519、524、528、553等)。

[0337] C.第二轮

[0338] 在第二轮中,用靶向合成GFP启动子区域的相同间隔序列和用相同基因抑制调节因子测试了表7B中提供的gRNA支架变体(例如,SEQ ID NO:555-588)和对照gRNA支架序列SQ(SEQ ID NO:500)。转染后第5天,通过流式细胞术对eGFP表达水平进行量化并将其相对于阴性对照(具有非靶向间隔序列的SQ支架)进行归一化。如图11所示并总结在表7A中,小于SQ阳性对照的多种gRNA支架变体表现出(i)与SQ阳性对照相当的基因阻抑或(ii)比SQ阳性对照更强的基因阻抑(例如,SEQ ID NO:555、557、558、568、569、578、580等)。

[0339] D.第三轮

[0340] 在第三轮中,用靶向CD2基因座上游(例如靶向内源DNA序列)的相同间隔序列和用相同基因激活调节因子测试了来自第一轮和第二轮的先导gRNA支架变体(例如,在表6B、表7B和表8B中提供)和对照gRNA支架序列SQ(SEQ ID NO:500)。转染后第2天,通过抗体染色和流式细胞术量化CD2表达水平,并将其相对于阴性对照(具有非靶向间隔序列的SQ支架)进行归一化。如图12所示并总结在表8A中,小于SQ阳性对照的多种gRNA支架变体表现出(i)与SQ阳性对照相当的基因激活或(ii)比SQ阳性对照更高的基因激活(例如,SEQ ID NO:555、557、568、569、576、577、578、580、593、519、528等)。例如,gRNA支架变体如SEQ ID NO:555(gRNA支架变体“45.1”)、SEQ ID NO:557(gRNA支架变体“2-6”)、SEQ ID NO:556(gRNA支架变体“2-17”)和SEQ ID NO:578(gRNA支架变体“2-27”)导致比SQ阳性对照更高的CD2激活(例如,与SQ阳性对照相比,基因激活程度增加超过0.5倍、超过1倍或超过1.5倍)。图13中示意性地示出了包含gRNA支架变体45.1和2-27以及2-17的单个引导RNA分子。

[0341] 本文公开的gRNA支架变体(例如45.1、2-17或2-27)可以表征为包含其N端序列的至少一部分,例如来自表9的SEQ ID NO:597的多核苷酸序列。

[0342] 本文公开的gRNA支架变体(例如2-6)可以表征为包含其N端序列的至少一部分,例如来自表9的SEQ ID NO:598的多核苷酸序列。

[0343] E.包含gRNA支架变体和工程化核酸酶变体的RNP复合体

[0344] 本文公开的gRNA支架变体(例如45.1或SEQ ID NO:555)可用于生成sgRNA,其可以与本文公开的任何工程化核酸酶变体(例如,SEQ ID NO:12)组合用于(i)减小编码sgRNA和工程化核酸酶变体的载体的尺寸,如图14所示,和/或(ii)进一步改善靶基因调节(例如,更大程度或调节、调节的持续时间更长等)。

表格

表1.靶基因(例如,编码感兴趣的蛋白质)的实例列表

感兴趣的靶基因/蛋白的列表
A1BG, CCND3, FAM163A, KCNK10, NRG4, REM1, TECTB, A1CF, CCNDBP1, FAM163B, KCNK12, NRGN, REM2, TEDDM1, A2M, CCNE1, FAM166A, KCNK13, NRIP1, REN, TEF, A2ML1, CCNE2, FAM166B, KCNK15, NRIP2, RENBP, TEFM, A3GALT2, CCNF, FAM167A, KCNK16, NRIP3, REP15, TEK, A4GALT, CCNG1, FAM167B, KCNK17, NRK, REPIN1, TEKT1, A4GNT, CCNG2, FAM168A, KCNK18, NRL, REPS1, TEKT2, AAAS, CCNH, FAM168B, KCNK2, NRM, REPS2, TEKT3, AACS, CCNI, FAM169A, KCNK3, NRN1, RER1, TEKT4, AADAC, CCNI2, FAM169B, KCNK4, NRN1L, RERE, TEKT5, AADACL2, CCNJ, FAM170A, KCNK5, NRP1, RERG, TELO2, AADACL3, CCNJL, FAM170B, KCNK6, NRP2, RERGL, TEN1, AADACL4, CCNK, FAM171A1, KCNK7, NRROS, RESP18, TENC1, AADAT, CCNL1, FAM171A2, KCNK9, NRSN1, REST, TENM1, AAED1, CCNL2, FAM171B, KCNMA1, NRSN2, RET, TENM2, AAGAB, CCNO, FAM172A, KCNMB1, NRTN, RETN, TENM3, AAK1, CCNT1, FAM173A, KCNMB2, NRXN1, RETNLB, TENM4, AAMDC, CCNT2, FAM173B, KCNMB3, NRXN2, RETSAT, TEP1, AAMP, CCNY, FAM174A, KCNMB4, NRXN3, REV1, TEPP, AANAT, CCNYL1, FAM174B, KCNN1, NSA2, REV3L, TERF1, AAR2, CCP110, FAM175A, KCNN2, NSD1, REXO1, TERF2, AARD, CCPG1, FAM175B, KCNN3, NSDHL, REXO2, TERF2IP, AARS, CCR1, FAM177A1, KCNN4, NSF, REXO4, TERT, AARS2, CCR10, FAM177B, KCNQ1, NSFL1C, RFC1, TES, AARSD1, CCR2, FAM178A, KCNQ2, NSG1, RFC2, TESC, AASDH, CCR3, FAM178B, KCNQ3, NSL1, RFC3, TESK1, AASDHPPT, CCR4, FAM179A, KCNQ4, NSMAF, RFC4, TESK2, AASS, CCR5, FAM179B, KCNQ5, NSMCE1, RFC5, TESPA1, AATF, CCR6, FAM180A, KCNRG, NSMCE2, RFESD, TET1, AATK, CCR7, FAM180B, KCNS1, NSMCE4A, RFFL, TET2, ABAT, CCR8, FAM181A, KCNS2, NSMF, RFK, TET3, ABCA1, CCR9, FAM181B, KCNS3, NSRP1, RFNG, TEX10, ABCA10, CCRL2, FAM183A, KCNT1, NSUN2, RFPL1, TEX101, ABCA12, CCRN4L, FAM184A, KCNT2, NSUN3, RFPL2, TEX11, ABCA13, CCS, FAM184B,

KCNU1, NSUN4, RFPL3, TEX12, ABCA2, CCSAP, FAM185A, KCNV1, NSUN5, RFPL4A, TEX13A, ABCA3, CCSER1, FAM186A, KCNV2, NSUN6, RFPL4AL1, TEX13B, ABCA4, CCSER2, FAM186B, KCP, NSUN7, RFPL4B, TEX14, ABCA5, CCT2, FAM187B, KCTD1, NT5C, RFT1, TEX15, ABCA6, CCT3, FAM188A, KCTD10, NT5C1A, RFTN1, TEX19, ABCA7, CCT4, FAM188B, KCTD11, NT5C1B, RFTN2, TEX2, ABCA8, CCT5, FAM189A1, KCTD12, NT5C1B-RDH14, RFWD2, TEX22, ABCA9, CCT6A, FAM189A2, KCTD13, NT5C2, RFWD3, TEX26, ABCB1, CCT6B, FAM189B, KCTD14, NT5C3A, RFX1, TEX261, ABCB10, CCT7, FAM192A, KCTD15, NT5C3B, RFX2, TEX264, ABCB11, CCT8, FAM193A, KCTD16, NT5DC1, RFX3, TEX28, ABCB4, CCT8L2, FAM193B, KCTD17, NT5DC2, RFX4, TEX29, ABCB5, CCZ1, FAM194A, KCTD18, NT5DC3, RFX5, TEX30, ABCB6, CCZ1B, FAM194B, KCTD19, NT5E, RFX6, TEX33, ABCB7, CD101, FAM195A, KCTD2, NT5M, RFX7, TEX35, ABCB8, CD109, FAM195B, KCTD20, NTAN1, RFX8, TEX36, ABCB9, CD14, FAM196A, KCTD21, NTF3, RFXANK, TEX37, ABCC1, CD151, FAM196B, KCTD3, NTF4, RFXAP, TEX38, ABCC10, CD160, FAM198A, KCTD4, NTHL1, RGAG1, TEX40, ABCC11, CD163, FAM198B, KCTD5, NTM, RGAG4, TEX9, ABCC12, CD163L1, FAM199X, KCTD6, NTMT1, RGCC, TF, ABCC2, CD164, FAM19A1, KCTD7, NTN1, RGL1, TFAM, ABCC3, CD164L2, FAM19A2, KCTD8, NTN3, RGL2, TFAP2A, ABCC4, CD177, FAM19A3, KCTD9, NTN4, RGL3, TFAP2B, ABCC5, CD180, FAM19A4, KDELC1, NTN5, RGL4, TFAP2C, ABCC6, CD19, FAM19A5, KDELC2, NTNG1, RGMA, TFAP2D, ABCC8, CD1A, FAM200A, KDELR1, NTNG2, RGMB, TFAP2E, ABCC9, CD1B, FAM203A, KDELR2, NTPCR, RGN, TFAP4, ABCD1, CD1C, FAM203B, KDELR3, NTRK1, RGP1, TFB1M, ABCD2, CD1D, FAM204A, KDM1A, NTRK2, RGPD1, TFB2M, ABCD3, CD1E, FAM205A, KDM1B, NTRK3, RGPD2, TFCP2, ABCD4, CD2, FAM206A, KDM2A, NTS, RGPD3, TFCP2L1, ABCE1, CD200, FAM207A, KDM2B, NTSR1, RGPD4, TFDP1, ABCF1, CD200R1, FAM208A, KDM3A, NTSR2, RGPD5, TFDP2, ABCF2, CD200R1L, FAM208B, KDM3B, NUAK1, RGPD6, TFDP3, ABCF3, CD207, FAM209A, KDM4A, NUAK2, RGPD8, TFE3, ABCG1, CD209, FAM209B, KDM4B, NUB1, RGR, TFEB, ABCG2, CD22, FAM20A, KDM4C, NUBP1, RGS1, TFEC, ABCG4, CD226, FAM20B, KDM4D, NUBP2, RGS10, TFF1, ABCG5, CD24, FAM20C, KDM4E, NUBPL, RGS11, TFF2, ABCG8, CD244, FAM210A, KDM5A, NUCB1, RGS12, TFF3, ABHD1, CD247, FAM210B, KDM5B, NUCB2, RGS13, TFG, ABHD10, CD248, FAM211A, KDM5C, NUCKS1, RGS14, TFIP11, ABHD11, CD27, FAM211B, KDM5D, NUDC, RGS16, TFPI, ABHD12, CD274, FAM212A, KDM6A, NUDCD1, RGS17, TFPI2, ABHD12B, CD276, FAM212B, KDM6B, NUDCD2, RGS18, TFPT, ABHD13, CD28, FAM213A, KDM8, NUDCD3, RGS19, TFR2, ABHD14A, CD2AP, FAM213B, KDR, NUDT1, RGS2, TFRC, ABHD14B, CD2BP2, FAM214A, KDSR, NUDT10, RGS20, TG, ABHD15, CD300A, FAM214B, KEAP1, NUDT11, RGS21, TGDS, ABHD16A, CD300C, FAM216A, KEL, NUDT12, RGS22, TGFA, ABHD16B, CD300E, FAM216B, KERA, NUDT13, RGS3, TGFB1, ABHD17A, CD300LB, FAM217A, KHDC1, NUDT14, RGS4, TGFB1I1, ABHD17B, CD300LD, FAM217B, KHDC1L, NUDT15, RGS5, TGFB2, ABHD17C, CD300LF, FAM218A, KHDC3L, NUDT16, RGS6, TGFB3,

ABHD2, CD300LG, FAM219A, KHDRBS1, NUDT16L1, RGS7, TGFBI, ABHD3, CD302, FAM219B, KHDRBS2, NUDT17, RGS7BP, TGFBR1, ABHD4, CD320, FAM21A, KHDRBS3, NUDT18, RGS8, TGFBR2, ABHD5, CD33, FAM21B, KHK, NUDT19, RGS9, TGFBR3, ABHD6, CD34, FAM21C, KHNYN, NUDT2, RGS9BP, TGFBR3L, ABHD8, CD36, FAM220A, KHSRP, NUDT21, RGSL1, TGFBRAP1, ABI1, CD37, FAM221A, KIAA0020, NUDT22, RHAG, TGIF1, ABI2, CD38, FAM221B, KIAA0040, NUDT3, RHBDD1, TGIF2, ABI3, CD3D, FAM222A, KIAA0100, NUDT4, RHBDD2, TGIF2-C20orf24, ABI3BP, CD3E, FAM222B, KIAA0101, NUDT5, RHBDD3, TGIF2LX, ABL1, CD3EAP, FAM227A, KIAA0141, NUDT6, RHBDF1, TGIF2LY, ABL2, CD3G, FAM227B, KIAA0195, NUDT7, RHBDF2, TGM1, ABLIM1, CD4, FAM228A, KIAA0196, NUDT8, RHBDL1, TGM2, ABLIM2, CD40, FAM228B, KIAA0226, NUDT9, RHBDL2, TGM3, ABLIM3, CD40LG, FAM229A, KIAA0226L, NUF2, RHBDL3, TGM4, ABO, CD44, FAM229B, KIAA0232, NUFIP1, RHBG, TGM5, ABR, CD46, FAM230A, KIAA0247, NUFIP2, RHCE, TGM6, ABRA, CD47, FAM24A, KIAA0319, NUGGC, RHCG, TGM7, ABRACL, CD48, FAM24B, KIAA0319L, NUMA1, RHD, TGOLN2, ABT1, CD5, FAM25A, KIAA0355, NUMB, RHEB, TGS1, ABTB1, CD52, FAM25C, KIAA0368, NUMBL, RHEBL1, TH, ABTB2, CD53, FAM25G, KIAA0391, NUP107, RHNO1, THADA, ACAA1, CD55, FAM26D, KIAA0408, NUP133, RHO, THAP1, ACAA2, CD58, FAM26E, KIAA0430, NUP153, RHOA, THAP10, ACACA, CD59, FAM26F, KIAA0513, NUP155, RHOB, THAP11, ACACB, CD5L, FAM32A, KIAA0556, NUP160, RHOBTB1, THAP2, ACAD10, CD6, FAM35A, KIAA0586, NUP188, RHOBTB2, THAP3, ACAD11, CD63, FAM3A, KIAA0753, NUP205, RHOBTB3, THAP4, ACAD8, CD68, FAM3B, KIAA0754, NUP210, RHOC, THAP5, ACAD9, CD69, FAM3C, KIAA0825, NUP210L, RHOD, THAP6, ACADL, CD7, FAM3D, KIAA0895, NUP214, RHOF, THAP7, ACADM, CD70, FAM43A, KIAA0895L, NUP35, RHOG, THAP8, ACADS, CD72, FAM43B, KIAA0907, NUP37, RHOH, THAP9, ACADSB, CD74, FAM45A, KIAA0922, NUP43, RHOJ, THBD, ACADVL, CD79A, FAM46A, KIAA0930, NUP50, RHOQ, THBS1, ACAN, CD79B, FAM46B, KIAA0947, NUP54, RHOT1, THBS2, ACAP1, CD80, FAM46C, KIAA1009, NUP62, RHOT2, THBS3, ACAP2, CD81, FAM46D, KIAA1024, NUP62CL, RHOU, THBS4, ACAP3, CD82, FAM47A, KIAA1024L, NUP85, RHOV, THEG, ACAT1, CD83, FAM47B, KIAA1033, NUP88, RHOXF1, THEG5, ACAT2, CD84, FAM47C, KIAA1045, NUP93, RHOXF2, THEGL, ACBD3, CD86, FAM47E, KIAA1107, NUP98, RHOXF2B, THEM4, ACBD4, CD8A, FAM47E-STBD1, KIAA1109, NUPL1, RHPN1, THEM5, ACBD5, CD8B, FAM49A, KIAA1143, NUPL2, RHPN2, THEM6, ACBD6, CD9, FAM49B, KIAA1147, NUPR1, RIBC1, THEMIS, ACBD7, CD93, FAM50A, KIAA1161, NUPR1L, RIBC2, THEMIS2, ACCS, CD96, FAM50B, KIAA1191, NUS1, RIC3, THG1L, ACCSL, CD97, FAM53A, KIAA1199, NUSAP1, RIC8A, THNSL1, ACD, CD99, FAM53B, KIAA1210, NUTF2, RIC8B, THNSL2, ACE, CD99L2, FAM53C, KIAA1211, NUTM1, RICTOR, THOC1, ACE2, CDA, FAM57A, KIAA1211L, NUTM2A, RIF1, THOC2, ACER1, CDADC1, FAM57B, KIAA1217, NUTM2B, RIIAD1, THOC3, ACER2, CDAN1, FAM58A, KIAA1239, NUTM2F, RILP, THOC5, ACER3, CDC123, FAM60A, KIAA1244, NUTM2G,

RILPL1, THOC6, ACHE, CDC14A, FAM63A, KIAA1257, NVL, RILPL2, THOC7, ACIN1, CDC14B, FAM63B, KIAA1279, NWD1, RIMBP2, THOP1, ACKR1, CDC16, FAM64A, KIAA1324, NXF1, RIMBP3, THPO, ACKR2, CDC20, FAM65A, KIAA1324L, NXF2, RIMBP3B, THRA, ACKR3, CDC20B, FAM65B, KIAA1328, NXF2B, RIMBP3C, THRAP3, ACKR4, CDC23, FAM65C, KIAA1377, NXF3, RIMKLA, THRB, ACLY, CDC25A, FAM69A, KIAA1407, NXF5, RIMKLB, THRSP, ACMSD, CDC25B, FAM69B, KIAA1429, NXN, RIMS1, THSD1, ACN9, CDC25C, FAM69C, KIAA1430, NXNL1, RIMS2, THSD4, ACO1, CDC26, FAM71A, KIAA1432, NXNL2, RIMS3, THSD7A, ACO2, CDC27, FAM71B, KIAA1456, NXPE1, RIMS4, THSD7B, ACOT1, CDC34, FAM71C, KIAA1462, NXPE2, RIN1, THTPA, ACOT11, CDC37, FAM71D, KIAA1467, NXPE3, RIN2, THUMPD1, ACOT12, CDC37L1, FAM71E1, KIAA1468, NXPE4, RIN3, THUMPD2, ACOT13, CDC40, FAM71E2, KIAA1522, NXPH1, RING1, THUMPD3, ACOT2, CDC42, FAM71F1, KIAA1524, NXPH2, RINL, THY1, ACOT4, CDC42BPA, FAM71F2, KIAA1549, NXPH3, RINT1, THYN1, ACOT6, CDC42BPB, FAM72A, KIAA1549L, NXPH4, RIOK1, TIA1, ACOT7, CDC42BPG, FAM72B, KIAA1551, NXT1, RIOK2, TIAF1, ACOT8, CDC42EP1, FAM72D, KIAA1586, NXT2, RIOK3, TIAL1, ACOT9, CDC42EP2, FAM73A, KIAA1598, NYAP1, RIPK1, TIAM1, ACOX1, CDC42EP3, FAM73B, KIAA1614, NYAP2, RIPK2, TIAM2, ACOX2, CDC42EP4, FAM76A, KIAA1644, NYNRIN, RIPK3, TICAM1, ACOX3, CDC42EP5, FAM76B, KIAA1671, NYX, RIPK4, TICAM2, ACOXL, CDC42SE1, FAM78A, KIAA1683, OAF, RIPPLY1, TICRR, ACP1, CDC42SE2, FAM78B, KIAA1715, OARD1, RIPPLY2, TIE1, ACP2, CDC45, FAM81A, KIAA1731, OAS1, RIPPLY3, TIFA, ACP5, CDC5L, FAM81B, KIAA1737, OAS2, RIT1, TIFAB, ACP6, CDC6, FAM83A, KIAA1751, OAS3, RIT2, TIGD2, ACPL2, CDC7, FAM83B, KIAA1755, OASL, RLBP1, TIGD3, ACPP, CDC73, FAM83C, KIAA1804, OAT, RLF, TIGD4, ACPT, CDCA2, FAM83D, KIAA1841, OAZ1, RLIM, TIGD5, ACR, CDCA3, FAM83E, KIAA1919, OAZ2, RLN1, TIGD6, ACRBP, CDCA4, FAM83F, KIAA1958, OAZ3, RLN2, TIGD7, ACRC, CDCA5, FAM83G, KIAA1984, OBFC1, RLN3, TIGIT, ACRV1, CDCA7, FAM83H, KIAA2013, OBP2A, RLTPR, TIMD4, ACSBG1, CDCA7L, FAM84A, KIAA2018, OBP2B, RMDN1, TIMELESS, ACSBG2, CDCA8, FAM84B, KIAA2022, OBSCN, RMDN2, TIMM10, ACSF2, CDCP1, FAM86A, KIAA2026, OBSL1, RMDN3, TIMM10B, ACSF3, CDCP2, FAM86B1, KIDINS220, OC90, RMI1, TIMM13, ACSL1, CDH1, FAM86B2, KIF11, OCA2, RMI2, TIMM17A, ACSL3, CDH10, FAM86C1, KIF12, OCEL1, RMND1, TIMM17B, ACSL4, CDH11, FAM86KP, KIF13A, OCIAD1, RMND5A, TIMM21, ACSL5, CDH12, FAM89A, KIF13B, OCIAD2, RMND5B, TIMM22, ACSL6, CDH13, FAM89B, KIF14, OCLM, RNASE1, TIMM23, ACSM1, CDH15, FAM8A1, KIF15, OCLN, RNASE10, TIMM23B, ACSM2A, CDH16, FAM90A1, KIF16B, OCM, RNASE11, TIMM44, ACSM2B, CDH17, FAM91A1, KIF17, OCM2, RNASE12, TIMM50, ACSM3, CDH18, FAM92A1, KIF18A, OCRL, RNASE13, TIMM8A, ACSM4, CDH19, FAM92B, KIF18B, OCSTAMP, RNASE2, TIMM8B, ACSM5, CDH2, FAM96A, KIF19, ODAM, RNASE3, TIMM9, ACSS1, CDH20, FAM96B, KIF1A, ODC1, RNASE4, TIMMDC1, ACSS2, CDH22, FAM98A, KIF1B, ODF1, RNASE6, TIMP1, ACSS3, CDH23, FAM98B, KIF1C,

ODF2, RNASE7, TIMP2, ACTA1, CDH24, FAM98C, KIF20A, ODF2L, RNASE8, TIMP3, ACTA2, CDH26, FAM9A, KIF20B, ODF3, RNASE9, TIMP4, ACTB, CDH3, FAM9B, KIF21A, ODF3B, RNASEH1, TINAG, ACTBL2, CDH4, FAM9C, KIF21B, ODF3L1, RNASEH2A, TINAGL1, ACTC1, CDH5, FAN1, KIF22, ODF3L2, RNASEH2B, TINF2, ACTG1, CDH6, FANCA, KIF23, ODF4, RNASEH2C, TIPARP, ACTG2, CDH7, FANCB, KIF24, OFCC1, RNASEK, TIPIN, ACTL10, CDH8, FANCC, KIF25, OFD1, RNASEL, TIPRL, ACTL6A, CDH9, FANCD2, KIF26A, OGDH, RNASET2, TIRAP, ACTL6B, CDHR1, FANCD2OS, KIF26B, OGDHL, RND1, TJAP1, ACTL7A, CDHR2, FANCE, KIF27, OGFOD1, RND2, TJP1, ACTL7B, CDHR3, FANCF, KIF28P, OGFOD2, RND3, TJP2, ACTL8, CDHR4, FANCG, KIF2A, OGFOD3, RNF10, TJP3, ACTL9, CDHR5, FANCI, KIF2B, OGFR, RNF103, TK1, ACTN1, CDIP1, FANCL, KIF2C, OGFRL1, RNF103-CHMP3, TK2, ACTN2, CDIPT, FANCM, KIF3A, OGG1, RNF11, TKT, ACTN3, CDK1, FANK1, KIF3B, OGN, RNF111, TKTL1, ACTN4, CDK10, FAP, KIF3C, OGT, RNF112, TKTL2, ACTR10, CDK11A, FAR1, KIF4A, OIP5, RNF113A, TLCD1, ACTR1A, CDK11B, FAR2, KIF4B, OIT3, RNF113B, TLCD2, ACTR1B, CDK12, FARP1, KIF5A, OLA1, RNF114, TLDC1, ACTR2, CDK13, FARP2, KIF5B, OLAH, RNF115, TLDC2, ACTR3, CDK14, FARS2, KIF5C, OLFM1, RNF121, TLE1, ACTR3B, CDK15, FARSA, KIF6, OLFM2, RNF122, TLE2, ACTR3C, CDK16, FARSB, KIF7, OLFM3, RNF123, TLE3, ACTR5, CDK17, FAS, KIF9, OLFM4, RNF125, TLE4, ACTR6, CDK18, FASLG, KIFAP3, OLFML1, RNF126, TLE6, ACTR8, CDK19, FASN, KIFC1, OLFML2A, RNF128, TLK1, ACTRT1, CDK2, FASTK, KIFC2, OLFML2B, RNF13, TLK2, ACTRT2, CDK20, FASTKD1, KIFC3, OLFML3, RNF130, TLL1, ACTRT3, CDK2AP1, FASTKD2, KIN, OLIG1, RNF133, TLL2, ACVR1, CDK2AP2, FASTKD3, KIR2DL1, OLIG2, RNF135, TLN1, ACVR1B, CDK3, FASTKD5, KIR2DL3, OLIG3, RNF138, TLN2, ACVR1C, CDK4, FAT1, KIR2DL4, OLR1, RNF139, TLR1, ACVR2A, CDK5, FAT2, KIR2DS4, OMA1, RNF14, TLR10, ACVR2B, CDK5R1, FAT3, KIR3DL1, OMD, RNF141, TLR2, ACVRL1, CDK5R2, FAT4, KIR3DL2, OMG, RNF144A, TLR3, ACY1, CDK5RAP1, FATE1, KIR3DL3, OMP, RNF144B, TLR4, ACY3, CDK5RAP2, FAU, KIRREL, ONECUT1, RNF145, TLR5, ACYP1, CDK5RAP3, FAXC, KIRREL2, ONECUT2, RNF146, TLR6, ACYP2, CDK6, FAXDC2, KIRREL3, ONECUT3, RNF148, TLR7, ADA, CDK7, FBF1, KISS1, OOEP, RNF149, TLR8, ADAD1, CDK8, FBL, KISS1R, OOSP2, RNF150, TLR9, ADAD2, CDK9, FBLIM1, KIT, OPA1, RNF151, TLX1, ADAL, CDKAL1, FBLN1, KITLG, OPA3, RNF152, TLX1NB, ADAM10, CDKL1, FBLN2, KL, OPALIN, RNF157, TLX2, ADAM11, CDKL2, FBLN5, KLB, OPCML, RNF165, TLX3, ADAM12, CDKL3, FBLN7, KLC1, OPHN1, RNF166, TM2D1, ADAM15, CDKL4, FBN1, KLC2, OPLAH, RNF167, TM2D2, ADAM17, CDKL5, FBN2, KLC3, OPN1LW, RNF168, TM2D3, ADAM18, CDKN1A, FBN3, KLC4, OPN1MW, RNF169, TM4SF1, ADAM19, CDKN1B, FBP1, KLF1, OPN1MW2, RNF17, TM4SF18, ADAM2, CDKN1C, FBP2, KLF10, OPN1SW, RNF170, TM4SF19, ADAM20, CDKN2A, FBRS, KLF11, OPN3, RNF175, TM4SF20, ADAM21, CDKN2AIP, FBRSL1, KLF12, OPN4, RNF180, TM4SF4, ADAM22, CDKN2AIPNL, FBXL12, KLF13, OPN5, RNF181, TM4SF5, ADAM23, CDKN2B, FBXL13, KLF14, OPRD1, RNF182, TM6SF1, ADAM28, CDKN2C, FBXL14,

KLF15, OPRK1, RNF183, TM6SF2, ADAM29, CDKN2D, FBXL15, KLF16, OPRL1, RNF185, TM7SF2, ADAM30, CDKN3, FBXL16, KLF17, OPRM1, RNF186, TM7SF3, ADAM32, CDNF, FBXL17, KLF2, OPTC, RNF187, TM9SF1, ADAM33, CDO1, FBXL18, KLF3, OPTN, RNF19A, TM9SF2, ADAM7, CDON, FBXL19, KLF4, OR10A2, RNF19B, TM9SF3, ADAM8, CDPF1, FBXL2, KLF5, OR10A3, RNF2, TM9SF4, ADAM9, CDR1, FBXL20, KLF6, OR10A4, RNF20, TMA16, ADAMDEC1, CDR2, FBXL21, KLF7, OR10A5, RNF207, TMA7, ADAMTS1, CDR2L, FBXL22, KLF8, OR10A6, RNF208, TMBIM1, ADAMTS10, CDRT1, FBXL3, KLF9, OR10A7, RNF212, TMBIM4, ADAMTS12, CDRT15, FBXL4, KLHDC1, OR10AG1, RNF213, TMBIM6, ADAMTS13, CDRT15L2, FBXL5, KLHDC10, OR10C1, RNF214, TMC1, ADAMTS14, CDRT4, FBXL6, KLHDC2, OR10G2, RNF215, TMC2, ADAMTS15, CDS1, FBXL7, KLHDC3, OR10G3, RNF216, TMC3, ADAMTS16, CDS2, FBXL8, KLHDC4, OR10G4, RNF217, TMC4, ADAMTS17, CDSN, FBXO10, KLHDC7A, OR10G7, RNF219, TMC5, ADAMTS18, CDT1, FBXO11, KLHDC7B, OR10G8, RNF220, TMC6, ADAMTS19, CDV3, FBXO15, KLHDC8A, OR10G9, RNF222, TMC7, ADAMTS2, CDX1, FBXO16, KLHDC8B, OR10H1, RNF223, TMC8, ADAMTS20, CDX2, FBXO17, KLHDC9, OR10H2, RNF224, TMCC1, ADAMTS3, CDX4, FBXO18, KLHL1, OR10H3, RNF24, TMCC2, ADAMTS4, CDY1, FBXO2, KLHL10, OR10H4, RNF25, TMCC3, ADAMTS5, CDY1B, FBXO21, KLHL11, OR10H5, RNF26, TMCO1, ADAMTS6, CDY2A, FBXO22, KLHL12, OR10J1, RNF31, TMCO2, ADAMTS7, CDY2B, FBXO24, KLHL13, OR10J3, RNF32, TMCO3, ADAMTS8, CDYL, FBXO25, KLHL14, OR10J5, RNF34, TMCO4, ADAMTS9, CDYL2, FBXO27, KLHL15, OR10K1, RNF38, TMCO5A, ADAMTSL1, CEACAM1, FBXO28, KLHL17, OR10K2, RNF39, TMCO6, ADAMTSL2, CEACAM16, FBXO3, KLHL18, OR10P1, RNF4, TMED1, ADAMTSL3, CEACAM18, FBXO30, KLHL2, OR10Q1, RNF40, TMED10, ADAMTSL4, CEACAM19, FBXO31, KLHL20, OR10R2, RNF41, TMED2, ADAMTSL5, CEACAM20, FBXO32, KLHL21, OR10S1, RNF43, TMED3, ADAP1, CEACAM21, FBXO33, KLHL22, OR10T2, RNF44, TMED4, ADAP2, CEACAM3, FBXO34, KLHL23, OR10V1, RNF5, TMED5, ADAR, CEACAM4, FBXO36, KLHL24, OR10W1, RNF6, TMED6, ADARB1, CEACAM5, FBXO38, KLHL25, OR10X1, RNF7, TMED7, ADARB2, CEACAM6, FBXO39, KLHL26, OR10Z1, RNF8, TMED7-TICAM2, ADAT1, CEACAM7, FBXO4, KLHL28, OR11A1, RNFT1, TMED8, ADAT2, CEACAM8, FBXO40, KLHL29, OR11G2, RNFT2, TMED9, ADAT3, CEBPA, FBXO41, KLHL3, OR11H1, RNGTT, TMEFF1, ADC, CEBPB, FBXO42, KLHL30, OR11H12, RNH1, TMEFF2, ADCK1, CEBPD, FBXO43, KLHL31, OR11H2, RNLS, TMEM100, ADCK2, CEBPE, FBXO44, KLHL32, OR11H4, RNMT, TMEM101, ADCK3, CEBPG, FBXO45, KLHL33, OR11H6, RNMTL1, TMEM102, ADCK4, CEBPZ, FBXO46, KLHL34, OR11L1, RNPC3, TMEM104, ADCK5, CECR1, FBXO47, KLHL35, OR12D2, RNPEP, TMEM105, ADCY1, CECR2, FBXO48, KLHL36, OR12D3, RNPEPL1, TMEM106A, ADCY10, CECR5, FBXO5, KLHL38, OR13A1, RNPS1, TMEM106B, ADCY2, CECR6, FBXO6, KLHL4, OR13C2, ROBO1, TMEM106C, ADCY3, CEL, FBXO7, KLHL40, OR13C3, ROBO2, TMEM107, ADCY4, CELA1, FBXO8, KLHL41, OR13C4, ROBO3, TMEM108, ADCY5, CELA2A, FBXO9, KLHL42, OR13C5,

ROBO4, TMEM109, ADCY6, CELA2B, FBXW10, KLHL5, OR13C8, ROCK1, TMEM11, ADCY7, CELA3A, FBXW11, KLHL6, OR13C9, ROCK2, TMEM110, ADCY8, CELA3B, FBXW12, KLHL7, OR13D1, ROGDI, TMEM110-MUSTN1, ADCY9, CELF1, FBXW2, KLHL8, OR13F1, ROM1, TMEM114, ADCYAP1, CELF2, FBXW4, KLHL9, OR13G1, ROMO1, TMEM115, ADCYAP1R1, CELF3, FBXW5, KLK1, OR13H1, ROPN1, TMEM116, ADD1, CELF4, FBXW7, KLK10, OR13J1, ROPN1B, TMEM117, ADD2, CELF5, FBXW8, KLK11, OR14A16, ROPN1L, TMEM119, ADD3, CELF6, FBXW9, KLK12, OR14C36, ROR1, TMEM120A, ADGB, CELSR1, FCAMR, KLK13, OR14J1, ROR2, TMEM120B, ADH1A, CELSR2, FCAR, KLK14, OR1A1, RORA, TMEM121, ADH1B, CELSR3, FCER1A, KLK15, OR1A2, RORB, TMEM123, ADH1C, CEMP1, FCER1G, KLK2, OR1B1, RORC, TMEM125, ADH4, CEND1, FCER2, KLK3, OR1C1, ROS1, TMEM126A, ADH5, CENPA, FCF1, KLK4, OR1D2, RP1, TMEM126B, ADH6, CENPB, FCGBP, KLK5, OR1D5, RP1L1, TMEM127, ADH7, CENPBD1, FCGR1A, KLK6, OR1E1, RP2, TMEM128, ADHFE1, CENPC, FCGR1B, KLK7, OR1E2, RP9, TMEM129, ADI1, CENPE, FCGR2A, KLK8, OR1F1, RPA1, TMEM130, ADIG, CENPF, FCGR2B, KLK9, OR1G1, RPA2, TMEM131, ADIPOQ, CENPH, FCGR2C, KLKB1, OR1I1, RPA3, TMEM132A, ADIPOR1, CENPI, FCGR3A, KLLN, OR1J1, RPA4, TMEM132B, ADIPOR2, CENPJ, FCGR3B, KLRB1, OR1J4, RPAIN, TMEM132C, ADIRF, CENPK, FCGRT, KLRC1, OR1K1, RPAP1, TMEM132D, ADK, CENPL, FCHO1, KLRC2, OR1L1, RPAP2, TMEM132E, ADM, CENPM, FCHO2, KLRC3, OR1L3, RPAP3, TMEM133, ADM2, CENPN, FCHSD1, KLRC4, OR1L4, RPE, TMEM134, ADM5, CENPO, FCHSD2, KLRC4-KLRK1, OR1L6, RPE65, TMEM135, ADNP, CENPP, FCN1, KLRD1, OR1L8, RPF1, TMEM136, ADNP2, CENPQ, FCN2, KLRF1, OR1M1, RPF2, TMEM138, ADO, CENPT, FCN3, KLRF2, OR1N1, RPGR, TMEM139, ADORA1, CENPU, FCRL1, KLRG1, OR1N2, RPGRIP1, TMEM140, ADORA2A, CENPV, FCRL2, KLRG2, OR1Q1, RPGRIP1L, TMEM141, ADORA2B, CENPW, FCRL3, KLRK1, OR1S1, RPH3A, TMEM143, ADORA3, CEP104, FCRL4, KMO, OR1S2, RPH3AL, TMEM144, ADPGK, CEP112, FCRL5, KMT2A, OR2A12, RPIA, TMEM145, ADPRH, CEP120, FCRL6, KMT2B, OR2A14, RPL10, TMEM147, ADPRHL1, CEP128, FCRLA, KMT2C, OR2A2, RPL10A, TMEM14A, ADPRHL2, CEP135, FCRLB, KMT2D, OR2A25, RPL10L, TMEM14B, ADPRM, CEP152, FDCSP, KMT2E, OR2A4, RPL11, TMEM14C, ADRA1A, CEP164, FDFT1, KNCN, OR2A5, RPL12, TMEM14E, ADRA1B, CEP170, FDPS, KNDC1, OR2A7, RPL13, TMEM150A, ADRA1D, CEP170B, FDX1, KNG1, OR2AE1, RPL13A, TMEM150B, ADRA2A, CEP19, FDX1L, KNOP1, OR2AG1, RPL14, TMEM150C, ADRA2B, CEP192, FDXACB1, KNSTRN, OR2AG2, RPL15, TMEM151A, ADRA2C, CEP250, FDXR, KNTC1, OR2AK2, RPL17, TMEM151B, ADRB1, CEP290, FECH, KPNA1, OR2AP1, RPL17-C18orf32, TMEM154, ADRB2, CEP350, FEM1A, KPNA2, OR2AT4, RPL18, TMEM155, ADRB3, CEP41, FEM1B, KPNA3, OR2B11, RPL18A, TMEM156, ADRBK1, CEP44, FEM1C, KPNA4, OR2B2, RPL19, TMEM158, ADRBK2, CEP55, FEN1, KPNA5, OR2B3, RPL21, TMEM159, ADRM1, CEP57, FER, KPNA6, OR2B6, RPL22, TMEM160, ADSL, CEP57L1, FER1L5, KPNA7, OR2C1, RPL22L1, TMEM161A, ADSS, CEP63, FER1L6, KPNB1, OR2C3, RPL23, TMEM161B, ADSSL1, CEP68, FERD3L, KPRP,

OR2D2, RPL23A, TMEM163, ADTRP, CEP70, FERMT1, KPTN, OR2D3, RPL24, TMEM164, AEBP1, CEP72, FERMT2, KRAS, OR2F1, RPL26, TMEM165, AEBP2, CEP76, FERMT3, KRBA1, OR2F2, RPL26L1, TMEM167A, AEN, CEP78, FES, KRBA2, OR2G2, RPL27, TMEM167B, AES, CEP85, FETUB, KRBOX1, OR2G3, RPL27A, TMEM168, AFAP1, CEP85L, FEV, KRBOX4, OR2G6, RPL28, TMEM169, AFAP1L1, CEP89, FEZ1, KRCC1, OR2H1, RPL29, TMEM17, AFAP1L2, CEP95, FEZ2, KREMEN1, OR2J2, RPL3, TMEM170A, AFF1, CEP97, FEZF1, KREMEN2, OR2J3, RPL30, TMEM170B, AFF2, CEPT1, FEZF2, KRI1, OR2K2, RPL31, TMEM171, AFF3, CER1, FFAR1, KRIT1, OR2L13, RPL32, TMEM173, AFF4, CERCAM, FFAR2, KRR1, OR2L2, RPL34, TMEM174, AFG3L2, CERK, FFAR3, KRT1, OR2L3, RPL35, TMEM175, AFM, CERKL, FFAR4, KRT10, OR2L5, RPL35A, TMEM176A, AFMID, CERS1, FGA, KRT12, OR2L8, RPL36, TMEM176B, AFP, CERS2, FGB, KRT13, OR2M2, RPL36A, TMEM177, AFTPH, CERS3, FGD1, KRT14, OR2M3, RPL36A-HNRNPH2, TMEM178A, AGA, CERS4, FGD2, KRT15, OR2M4, RPL36AL, TMEM178B, AGAP1, CERS5, FGD3, KRT16, OR2M5, RPL37, TMEM179, AGAP10, CERS6, FGD4, KRT17, OR2M7, RPL37A, TMEM179B, AGAP11, CES1, FGD5, KRT18, OR2S2, RPL38, TMEM18, AGAP2, CES2, FGD6, KRT19, OR2T1, RPL39, TMEM180, AGAP3, CES3, FGF1, KRT2, OR2T10, RPL39L, TMEM181, AGAP4, CES4A, FGF10, KRT20, OR2T11, RPL3L, TMEM182, AGAP5, CES5A, FGF11, KRT222, OR2T12, RPL4, TMEM183A, AGAP6, CETN1, FGF12, KRT23, OR2T2, RPL41, TMEM183B, AGAP7, CETN2, FGF13, KRT24, OR2T27, RPL5, TMEM184A, AGAP8, CETN3, FGF14, KRT25, OR2T29, RPL6, TMEM184B, AGAP9, CETP, FGF16, KRT26, OR2T3, RPL7, TMEM184C, AGBL1, CFB, FGF17, KRT27, OR2T33, RPL7A, TMEM185A, AGBL2, CFC1, FGF18, KRT28, OR2T34, RPL7L1, TMEM185B, AGBL3, CFC1B, FGF19, KRT3, OR2T35, RPL8, TMEM186, AGBL4, CFD, FGF2, KRT31, OR2T4, RPL9, TMEM187, AGBL5, CFDP1, FGF20, KRT32, OR2T5, RPLP0, TMEM189, AGER, CFH, FGF21, KRT33A, OR2T6, RPLP1, TMEM189-UBE2V1, AGFG1, CFHR1, FGF22, KRT33B, OR2T8, RPLP2, TMEM19, AGFG2, CFHR2, FGF23, KRT34, OR2V1, RPN1, TMEM190, AGGF1, CFHR3, FGF3, KRT35, OR2V2, RPN2, TMEM191B, AGK, CFHR4, FGF4, KRT36, OR2W1, RPP14, TMEM191C, AGL, CFHR5, FGF5, KRT37, OR2W3, RPP21, TMEM192, AGMAT, CFI, FGF6, KRT38, OR2W5, RPP25, TMEM194A, AGMO, CFL1, FGF7, KRT39, OR2Y1, RPP25L, TMEM194B, AGO1, CFL2, FGF8, KRT4, OR2Z1, RPP30, TMEM196, AGO2, CFLAR, FGF9, KRT40, OR3A1, RPP38, TMEM198, AGO3, CFP, FGFBP1, KRT5, OR3A2, RPP40, TMEM199, AGO4, CFTR, FGFBP2, KRT6A, OR3A3, RPRD1A, TMEM2, AGPAT1, CGA, FGFBP3, KRT6B, OR4A15, RPRD1B, TMEM200A, AGPAT2, CGB, FGFR1, KRT6C, OR4A16, RPRD2, TMEM200B, AGPAT3, CGB1, FGFR1OP, KRT7, OR4A47, RPRM, TMEM201, AGPAT4, CGB2, FGFR1OP2, KRT71, OR4A5, RPRML, TMEM202, AGPAT5, CGB5, FGFR2, KRT72, OR4B1, RPS10, TMEM203, AGPAT6, CGB7, FGFR3, KRT73, OR4C11, RPS10-NUDT3, TMEM204, AGPAT9, CGB8, FGFR4, KRT74, OR4C12, RPS11, TMEM205, AGPS, CGGBP1, FGFRL1, KRT75, OR4C13, RPS12, TMEM206, AGR2, CGN, FGG, KRT76, OR4C15, RPS13, TMEM207, AGR3, CGNL1, FGGY, KRT77, OR4C16, RPS14, TMEM208, AGRN, CGREF1, FGL1, KRT78,

OR4C3, RPS15, TMEM209, AGRP, CGRRF1, FGL2, KRT79, OR4C45, RPS15A, TMEM210, AGT, CH25H, FGR, KRT8, OR4C46, RPS16, TMEM211, AGTPBP1, CHAC1, FH, KRT80, OR4C6, RPS17, TMEM212, AGTR1, CHAC2, FHAD1, KRT81, OR4D1, RPS17L, TMEM213, AGTR2, CHAD, FHDC1, KRT82, OR4D10, RPS18, TMEM214, AGTRAP, CHADL, FHIT, KRT83, OR4D11, RPS19, TMEM215, AGXT, CHAF1A, FHL1, KRT84, OR4D2, RPS19BP1, TMEM216, AGXT2, CHAF1B, FHL2, KRT85, OR4D5, RPS2, TMEM217, AHCTF1, CHAMP1, FHL3, KRT86, OR4D6, RPS20, TMEM218, AHCY, CHAT, FHL5, KRT9, OR4D9, RPS21, TMEM219, AHCYL1, CHCHD1, FHOD1, KRTAP10-1, OR4E2, RPS23, TMEM220, AHCYL2, CHCHD10, FHOD3, KRTAP10-10, OR4F15, RPS24, TMEM221, AHDC1, CHCHD2, FIBCD1, KRTAP10-11, OR4F16, RPS25, TMEM222, AHI1, CHCHD3, FIBIN, KRTAP10-12, OR4F21, RPS26, TMEM223, AHNAK, CHCHD4, FIBP, KRTAP10-2, OR4F29, RPS27, TMEM225, AHNAK2, CHCHD5, FICD, KRTAP10-3, OR4F3, RPS27A, TMEM229A, AHR, CHCHD6, FIG4, KRTAP10-4, OR4F4, RPS27L, TMEM229B, AHRR, CHCHD7, FIGF, KRTAP10-5, OR4F5, RPS28, TMEM230, AHSA1, CHD1, FIGLA, KRTAP10-6, OR4F6, RPS29, TMEM231, AHSA2, CHD1L, FIGN, KRTAP10-7, OR4K1, RPS3, TMEM232, AHSG, CHD2, FIGNL1, KRTAP10-8, OR4K13, RPS3A, TMEM233, AHSP, CHD3, FIGNL2, KRTAP10-9, OR4K14, RPS4X, TMEM234, AICDA, CHD4, FILIP1, KRTAP1-1, OR4K15, RPS4Y1, TMEM235, AIDA, CHD5, FILIP1L, KRTAP11-1, OR4K17, RPS4Y2, TMEM236, AIF1, CHD6, FIP1L1, KRTAP12-1, OR4K2, RPS5, TMEM237, AIF1L, CHD7, FIS1, KRTAP12-2, OR4K5, RPS6, TMEM238, AIFM1, CHD8, FITM1, KRTAP12-3, OR4L1, RPS6KA1, TMEM239, AIFM2, CHD9, FITM2, KRTAP12-4, OR4M1, RPS6KA2, TMEM240, AIFM3, CHDC2, FIZ1, KRTAP1-3, OR4M2, RPS6KA3, TMEM241, AIG1, CHDH, FJX1, KRTAP13-1, OR4N2, RPS6KA4, TMEM242, AIM1, CHEK1, FKBP10, KRTAP13-2, OR4N4, RPS6KA5, TMEM243, AIM1L, CHEK2, FKBP11, KRTAP13-3, OR4N5, RPS6KA6, TMEM244, AIM2, CHERP, FKBP14, KRTAP13-4, OR4P4, RPS6KB1, TMEM245, AIMP1, CHFR, FKBP15, KRTAP1-4, OR4Q3, RPS6KB2, TMEM246, AIMP2, CHGA, FKBP1A, KRTAP1-5, OR4S1, RPS6KC1, TMEM247, AIP, CHGB, FKBP1B, KRTAP15-1, OR4S2, RPS6KL1, TMEM248, AIPL1, CHI3L1, FKBP2, KRTAP16-1, OR4X1, RPS7, TMEM249, AIRE, CHI3L2, FKBP3, KRTAP17-1, OR4X2, RPS8, TMEM25, AJAP1, CHIA, FKBP4, KRTAP19-1, OR51A2, RPS9, TMEM251, AJUBA, CHIC1, FKBP5, KRTAP19-2, OR51A4, RPSA, TMEM252, AK1, CHIC2, FKBP6, KRTAP19-3, OR51A7, RPTN, TMEM253, AK2, CHID1, FKBP7, KRTAP19-4, OR51B2, RPTOR, TMEM254, AK3, CHIT1, FKBP8, KRTAP19-5, OR51B4, RPUSD1, TMEM255A, AK4, CHKA, FKBP9, KRTAP19-6, OR51B5, RPUSD2, TMEM255B, AK5, CHKB, FKBPL, KRTAP19-7, OR51B6, RPUSD3, TMEM256, AK6, CHL1, FKRP, KRTAP19-8, OR51D1, RPUSD4, TMEM257, AK7, CHM, FKTN, KRTAP20-1, OR51E1, RQCD1, TMEM258, AK8, CHML, FLAD1, KRTAP20-2, OR51E2, RRAD, TMEM259, AK9, CHMP1A, FLCN, KRTAP20-3, OR51F1, RRAGA, TMEM26, AKAP1, CHMP1B, FLG, KRTAP2-1, OR51F2, RRAGB, TMEM260, AKAP10, CHMP2A, FLG2, KRTAP21-1, OR51G1, RRAGC, TMEM261, AKAP11, CHMP2B, FLI1, KRTAP21-2, OR51G2, RRAGD, TMEM27, AKAP12, CHMP3, FLII, KRTAP21-3, OR51I1, RRAS, TMEM30A, AKAP13,

CHMP4A, FLJ22184, KRTAP2-2, OR51I2, RRAS2, TMEM30B, AKAP14, CHMP4B, FLJ25363, KRTAP22-1, OR51L1, RRBP1, TMEM31, AKAP17A, CHMP4C, FLJ44313, KRTAP22-2, OR51M1, RREB1, TMEM33, AKAP2, CHMP5, FLJ44635, KRTAP2-3, OR51Q1, RRH, TMEM35, AKAP3, CHMP6, FLJ45513, KRTAP23-1, OR51S1, RRM1, TMEM37, AKAP4, CHMP7, FLNA, KRTAP2-4, OR51T1, RRM2, TMEM38A, AKAP5, CHN1, FLNB, KRTAP24-1, OR51V1, RRM2B, TMEM38B, AKAP6, CHN2, FLNC, KRTAP25-1, OR52A1, RRN3, TMEM39A, AKAP7, CHODL, FLOT1, KRTAP26-1, OR52A5, RRNAD1, TMEM39B, AKAP8, CHORDC1, FLOT2, KRTAP27-1, OR52B2, RRP1, TMEM40, AKAP8L, CHP1, FLRT1, KRTAP29-1, OR52B4, RRP12, TMEM41A, AKAP9, CHP2, FLRT2, KRTAP3-1, OR52B6, RRP15, TMEM41B, AKIP1, CHPF, FLRT3, KRTAP3-2, OR52D1, RRP1B, TMEM42, AKIRIN1, CHPF2, FLT1, KRTAP3-3, OR52E2, RRP36, TMEM43, AKIRIN2, CHPT1, FLT3, KRTAP4-1, OR52E4, RRP7A, TMEM44, AKNA, CHRAC1, FLT3LG, KRTAP4-11, OR52E6, RRP8, TMEM45A, AKNAD1, CHRD, FLT4, KRTAP4-12, OR52E8, RRP9, TMEM45B, AKR1A1, CHRDL1, FLVCR1, KRTAP4-2, OR52H1, RRS1, TMEM47, AKR1B1, CHRDL2, FLVCR2, KRTAP4-3, OR52I1, RS1, TMEM5, AKR1B10, CHRFAM7A, FLYWCH1, KRTAP4-4, OR52I2, RSAD1, TMEM50A, AKR1B15, CHRM1, FLYWCH2, KRTAP4-5, OR52J3, RSAD2, TMEM50B, AKR1C1, CHRM2, FMN1, KRTAP4-6, OR52K1, RSBN1, TMEM51, AKR1C2, CHRM3, FMN2, KRTAP4-7, OR52K2, RSBN1L, TMEM52, AKR1C3, CHRM4, FMNL1, KRTAP4-8, OR52L1, RSC1A1, TMEM52B, AKR1C4, CHRM5, FMNL2, KRTAP4-9, OR52M1, RSF1, TMEM53, AKR1D1, CHRNA1, FMNL3, KRTAP5-1, OR52N1, RSG1, TMEM54, AKR1E2, CHRNA10, FMO1, KRTAP5-10, OR52N4, RSL1D1, TMEM55A, AKR7A2, CHRNA2, FMO2, KRTAP5-11, OR52N5, RSL24D1, TMEM55B, AKR7A3, CHRNA3, FMO3, KRTAP5-2, OR52R1, RSPH1, TMEM56, AKT1, CHRNA4, FMO4, KRTAP5-3, OR52W1, RSPH10B, TMEM56-RWDD3, AKT1S1, CHRNA5, FMO5, KRTAP5-4, OR56A1, RSPH10B2, TMEM57, AKT2, CHRNA6, FMOD, KRTAP5-5, OR56A3, RSPH3, TMEM59, AKT3, CHRNA7, FMR1, KRTAP5- 6, OR56A4, RSPH4A, TMEM59L, AKTIP, CHRNA9, FMR1NB, KRTAP5-7, OR56A5, RSPH6A, TMEM60, ALAD, CHRNB1, FN1, KRTAP5-8, OR56B1, RSPH9, TMEM61, ALAS1, CHRNB2, FN3K, KRTAP5-9, OR56B4, RSPO1, TMEM62, ALAS2, CHRNB3, FN3KRP, KRTAP6-1, OR5A1, RSPO2, TMEM63A, ALB, CHRNB4, FNBP1, KRTAP6-2, OR5A2, RSPO3, TMEM63B, ALCAM, CHRND, FNBP1L, KRTAP6-3, OR5AC2, RSPO4, TMEM63C, ALDH16A1, CHRNE, FNBP4, KRTAP7-1, OR5AK2, RSPRY1, TMEM64, ALDH18A1, CHRNG, FNDC1, KRTAP8-1, OR5AN1, RSRC1, TMEM65, ALDH1A1, CHST1, FNDC3A, KRTAP9-1, OR5AP2, RSRC2, TMEM66, ALDH1A2, CHST10, FNDC3B, KRTAP9-2, OR5AR1, RSU1, TMEM67, ALDH1A3, CHST11, FNDC4, KRTAP9-3, OR5AS1, RTBDN, TMEM68, ALDH1B1, CHST12, FNDC5, KRTAP9-4, OR5AU1, RTCA, TMEM69, ALDH1L1, CHST13, FNDC7, KRTAP9-6, OR5B12, RTCB, TMEM70, ALDH1L2, CHST14, FNDC8, KRTAP9-7, OR5B17, RTDR1, TMEM71, ALDH2, CHST15, FNDC9, KRTAP9- 8, OR5B2, RTEL1, TMEM72, ALDH3A1, CHST2, FNIP1, KRTAP9-9, OR5B21, RTF1, TMEM74, ALDH3A2, CHST3, FNIP2, KRTCAP2, OR5B3, RTFDC1, TMEM74B, ALDH3B1, CHST4, FNTA, KRTCAP3, OR5C1, RTKN,

TMEM79, ALDH3B2, CHST5, FNTB, KRTDAP, OR5D13, RTKN2, TMEM80, ALDH4A1, CHST6, FOCAD, KSR1, OR5D14, RTL1, TMEM81, ALDH5A1, CHST7, FOLH1, KSR2, OR5D16, RTN1, TMEM82, ALDH6A1, CHST8, FOLH1B, KTI12, OR5D18, RTN2, TMEM86A, ALDH7A1, CHST9, FOLR1, KTN1, OR5F1, RTN3, TMEM86B, ALDH8A1, CHSY1, FOLR2, KXD1, OR5H1, RTN4, TMEM87A, ALDH9A1, CHSY3, FOLR3, KY, OR5H14, RTN4IP1, TMEM87B, ALDOA, CHTF18, FOLR4, KYNU, OR5H15, RTN4R, TMEM88, ALDOB, CHTF8, FOPNL, L1CAM, OR5H2, RTN4RL1, TMEM88B, ALDOC, CHTOP, FOS, L1TD1, OR5H6, RTN4RL2, TMEM89, ALG1, CHUK, FOSB, L2HGDH, OR5I1, RTP1, TMEM8A, ALG10, CHURC1, FOSL1, L3HYPDH, OR5J2, RTP2, TMEM8B, ALG10B, CHURC1-FNTB, FOSL2, L3MBTL1, OR5K1, RTP3, TMEM8C, ALG11, CIAO1, FOXA1, L3MBTL2, OR5K2, RTP4, TMEM9, ALG12, CIAPIN1, FOXA2, L3MBTL3, OR5K3, RTTN, TMEM91, ALG13, CIB1, FOXA3, L3MBTL4, OR5K4, RUFY1, TMEM92, ALG14, CIB2, FOXB1, LACC1, OR5L1, RUFY2, TMEM95, ALG1L, CIB3, FOXB2, LACE1, OR5L2, RUFY3, TMEM97, ALG1L2, CIB4, FOXC1, LACRT, OR5M1, RUFY4, TMEM98, ALG2, CIC, FOXC2, LACTB, OR5M10, RUNDC1, TMEM99, ALG3, CIDEA, FOXD1, LACTB2, OR5M11, RUNDC3A, TMEM9B, ALG5, CIDEB, FOXD2, LACTBL1, OR5M3, RUNDC3B, TMF1, ALG6, CIDEC, FOXD3, LAD1, OR5M8, RUNX1, TMIE, ALG8, CIITA, FOXD4, LAG3, OR5M9, RUNX1T1, TMIGD1, ALG9, CILP, FOXD4L1, LAGE3, OR5P2, RUNX2, TMIGD2, ALK, CILP2, FOXD4L2, LAIR1, OR5P3, RUNX3, TMLHE, ALKBH1, CINP, FOXD4L3, LAIR2, OR5R1, RUSC1, TMOD1, ALKBH2, CIR1, FOXD4L4, LALBA, OR5T1, RUSC1-AS1, TMOD2, ALKBH3, CIRBP, FOXD4L5, LAMA1, OR5T2, RUSC2, TMOD3, ALKBH4, CIRH1A, FOXD4L6, LAMA2, OR5T3, RUVBL1, TMOD4, ALKBH5, CISD1, FOXE1, LAMA3, OR5V1, RUVBL2, TMPO, ALKBH6, CISD2, FOXE3, LAMA4, OR5W2, RWDD1, TMPPE, ALKBH7, CISD3, FOXF1, LAMA5, OR6A2, RWDD2A, TMPRSS11A, ALKBH8, CISH, FOXF2, LAMB1, OR6B1, RWDD2B, TMPRSS11B, ALLC, CIT, FOXG1, LAMB2, OR6B3, RWDD3, TMPRSS11D, ALMS1, CITED1, FOXH1, LAMB3, OR6C1, RWDD4, TMPRSS11E, ALOX12, CITED2, FOXI1, LAMB4, OR6C2, RXFP1, TMPRSS11F, ALOX12B, CITED4, FOXI2, LAMC1, OR6C3, RXFP2, TMPRSS12, ALOX15, CIZ1, FOXI3, LAMC2, OR6C4, RXFP3, TMPRSS13, ALOX15B, CKAP2, FOXJ1, LAMC3, OR6C6, RXFP4, TMPRSS15, ALOX5, CKAP2L, FOXJ2, LAMP1, OR6C65, RXRA, TMPRSS2, ALOX5AP, CKAP4, FOXJ3, LAMP2, OR6C68, RXRB, TMPRSS3, ALOXE3, CKAP5, FOXK1, LAMP3, OR6C70, RXRG, TMPRSS4, ALPI, CKB, FOXK2, LAMP5, OR6C74, RYBP, TMPRSS5, ALPK1, CKLF, FOXL1, LAMTOR1, OR6C75, RYK, TMPRSS6, ALPK2, CKLF-CMTM1, FOXL2, LAMTOR2, OR6C76, RYR1, TMPRSS7, ALPK3, CKM, FOXM1, LAMTOR3, OR6F1, RYR2, TMPRSS9, ALPL, CKMT1A, FOXN1, LAMTOR4, OR6K2, RYR3, TMSB10, ALPP, CKMT1B, FOXN2, LAMTOR5, OR6K3, S100A1, TMSB15A, ALPPL2, CKMT2, FOXN3, LANCL1, OR6K6, S100A10, TMSB15B, ALS2, CKS1B, FOXN4, LANCL2, OR6M1, S100A11, TMSB4X, ALS2CL, CKS2, FOXO1, LANCL3, OR6N1, S100A12, TMSB4Y, ALS2CR11, CLASP1, FOXO3, LAP3, OR6N2, S100A13, TMTC1, ALS2CR12, CLASP2, FOXO4, LAPTM4A, OR6P1, S100A14, TMTC2, ALX1, CLASRP, FOXO6,

LAPTM4B, OR6Q1, S100A16, TMTC3, ALX3, CLC, FOXP1, LAPTM5, OR6S1, S100A2, TMTC4, ALX4, CLCA1, FOXP2, LARGE, OR6T1, S100A3, TMUB1, ALYREF, CLCA2, FOXP3, LARP1, OR6V1, S100A4, TMUB2, AMACR, CLCA4, FOXP4, LARP1B, OR6X1, S100A5, TMX1, AMBN, CLCC1, FOXQ1, LARP4, OR6Y1, S100A6, TMX2, AMBP, CLCF1, FOXR1, LARP4B, OR7A10, S100A7, TMX3, AMBRA1, CLCN1, FOXR2, LARP6, OR7A17, S100A7A, TMX4, AMD1, CLCN2, FOXRED1, LARP7, OR7A5, S100A7L2, TNC, AMDHD1, CLCN3, FOXRED2, LARS, OR7C1, S100A8, TNF, AMDHD2, CLCN4, FOXS1, LARS2, OR7C2, S100A9, TNFAIP1, AMELX, CLCN5, FPGS, LAS1L, OR7D2, S100B, TNFAIP2, AMELY, CLCN6, FPGT, LASP1, OR7D4, S100G, TNFAIP3, AMER1, CLCN7, FPGT-TNNI3K, LAT, OR7E24, S100P, TNFAIP6, AMER2, CLCNKA, FPR1, LAT2, OR7G1, S100PBP, TNFAIP8, AMER3, CLCNKB, FPR2, LATS1, OR7G2, S100Z, TNFAIP8L1, AMFR, CLDN1, FPR3, LATS2, OR7G3, S1PR1, TNFAIP8L2, AMH, CLDN10, FRA10AC1, LAX1, OR8A1, S1PR2, TNFAIP8L2-SCNM1, AMHR2, CLDN11, FRAS1, LAYN, OR8B12, S1PR3, TNFAIP8L3, AMICA1, CLDN12, FRAT1, LBH, OR8B2, S1PR4, TNFRSF10A, AMIGO1, CLDN14, FRAT2, LBP, OR8B3, S1PR5, TNFRSF10B, AMIGO2, CLDN15, FREM1, LBR, OR8B4, SAA1, TNFRSF10C, AMIGO3, CLDN16, FREM2, LBX1, OR8B8, SAA2, TNFRSF10D, AMMECR1, CLDN17, FREM3, LBX2, OR8D1, SAA2-SAA4, TNFRSF11A, AMMECR1L, CLDN18, FRG1, LCA5, OR8D2, SAA4, TNFRSF11B, AMN, CLDN19, FRG2, LCA5L, OR8D4, SAAL1, TNFRSF12A, AMN1, CLDN2, FRG2B, LCAT, OR8G1, SAC3D1, TNFRSF13B, AMOT, CLDN20, FRG2C, LCE1A, OR8G2, SACM1L, TNFRSF13C, AMOTL1, CLDN22, FRK, LCE1B, OR8G5, SACS, TNFRSF14, AMOTL2, CLDN23, FRMD1, LCE1C, OR8H1, SAE1, TNFRSF17, AMPD1, CLDN24, FRMD3, LCE1D, OR8H2, SAFB, TNFRSF18, AMPD2, CLDN25, FRMD4A, LCE1E, OR8H3, SAFB2, TNFRSF19, AMPD3, CLDN3, FRMD4B, LCE1F, OR8I2, SAG, TNFRSF1A, AMPH, CLDN4, FRMD5, LCE2A, OR8J1, SAGE1, TNFRSF1B, AMT, CLDN5, FRMD6, LCE2B, OR8J3, SALL1, TNFRSF21, AMTN, CLDN6, FRMD7, LCE2C, OR8K1, SALL2, TNFRSF25, AMY1A, CLDN7, FRMD8, LCE2D, OR8K3, SALL3, TNFRSF4, AMY1B, CLDN8, FRMPD1, LCE3A, OR8K5, SALL4, TNFRSF6B, AMY1C, CLDN9, FRMPD2, LCE3B, OR8S1, SAMD1, TNFRSF8, AMY2A, CLDND1, FRMPD3, LCE3C, OR8U1, SAMD10, TNFRSF9, AMY2B, CLDND2, FRMPD4, LCE3D, OR8U8, SAMD11, TNFSF10, AMZ1, CLEC10A, FRRS1, LCE3E, OR9A2, SAMD12, TNFSF11, AMZ2, CLEC11A, FRRS1L, LCE4A, OR9A4, SAMD13, TNFSF12, ANAPC1, CLEC12A, FRS2, LCE5A, OR9G1, SAMD14, TNFSF12-TNFSF13, ANAPC10, CLEC12B, FRS3, LCE6A, OR9G4, SAMD15, TNFSF13, ANAPC11, CLEC14A, FRY, LCK, OR9G9, SAMD3, TNFSF13B, ANAPC13, CLEC16A, FRYL, LCLAT1, OR9I1, SAMD4A, TNFSF14, ANAPC15, CLEC17A, FRZB, LCMT1, OR9K2, SAMD4B, TNFSF15, ANAPC16, CLEC18A, FSBP, LCMT2, OR9Q1, SAMD5, TNFSF18, ANAPC2, CLEC18B, FSCB, LCN1, OR9Q2, SAMD7, TNFSF4, ANAPC4, CLEC18C, FSCN1, LCN10, ORAI1, SAMD8, TNFSF8, ANAPC5, CLEC19A, FSCN2, LCN12, ORAI2, SAMD9, TNFSF9, ANAPC7, CLEC1A, FSCN3, LCN15, ORAI3, SAMD9L, TNIK, ANG, CLEC1B, FSD1, LCN2, ORAOV1, SAMHD1, TNIP1, ANGEL1, CLEC2A, FSD1L, LCN6, ORC1, SAMM50, TNIP2,

ANGEL2, CLEC2B, FSD2, LCN8, ORC2, SAMSN1, TNIP3, ANGPT1, CLEC2D, FSHB, LCN9, ORC3, SAP130, TNK1, ANGPT2, CLEC2L, FSHR, LCNL1, ORC4, SAP18, TNK2, ANGPT4, CLEC3A, FSIP1, LCOR, ORC5, SAP25, TNKS, ANGPTL1, CLEC3B, FSIP2, LCORL, ORC6, SAP30, TNKS1BP1, ANGPTL2, CLEC4A, FST, LCP1, ORM1, SAP30BP, TNKS2, ANGPTL3, CLEC4C, FSTL1, LCP2, ORM2, SAP30L, TNMD, ANGPTL4, CLEC4D, FSTL3, LCT, ORMDL1, SAPCD1, TNN, ANGPTL5, CLEC4E, FSTL4, LCTL, ORMDL2, SAPCD2, TNNC1, ANGPTL6, CLEC4F, FSTL5, LDB1, ORMDL3, SAR1A, TNNC2, ANGPTL7, CLEC4G, FTCD, LDB2, OS9, SAR1B, TNNI1, ANHX, CLEC4M, FTH1, LDB3, OSBP, SARDH, TNNI2, ANK1, CLEC5A, FTH1P18, LDHA, OSBP2, SARM1, TNNI3, ANK2, CLEC6A, FTHL17, LDHAL6A, OSBPL10, SARNP, TNNI3K, ANK3, CLEC7A, FTL, LDHAL6B, OSBPL11, SARS, TNNT1, ANKAR, CLEC9A, FTMT, LDHB, OSBPL1A, SARS2, TNNT2, ANKDD1A, CLECL1, FTO, LDHC, OSBPL2, SART1, TNNT3, ANKDD1B, CLGN, FTSJ1, LDHD, OSBPL3, SART3, TNP1, ANKEF1, CLHC1, FTSJ2, LDLR, OSBPL5, SASH1, TNP2, ANKFN1, CLIC1, FTSJ3, LDLRAD1, OSBPL6, SASH3, TNPO1, ANKFY1, CLIC2, FUBP1, LDLRAD2, OSBPL7, SASS6, TNPO2, ANKH, CLIC3, FUBP3, LDLRAD3, OSBPL8, SAT1, TNPO3, ANKHD1, CLIC4, FUCA1, LDLRAD4, OSBPL9, SAT2, TNR, ANKHD1-EIF4EBP3, CLIC5, FUCA2, LDLRAP1, OSCAR, SATB1, TNRC18, ANKIB1, CLIC6, FUK, LDOC1, OSCP1, SATB2, TNRC6A, ANKK1, CLINT1, FUNDC1, LDOC1L, OSER1, SATL1, TNRC6B, ANKLE1, CLIP1, FUNDC2, LEAP2, OSGEP, SAV1, TNRC6C, ANKLE2, CLIP2, FUOM, LECT1, OSGEPL1, SAYSD1, TNS1, ANKMY1, CLIP3, FURIN, LECT2, OSGIN1, SBDS, TNS3, ANKMY2, CLIP4, FUS, LEF1, OSGIN2, SBF1, TNS4, ANKRA2, CLK1, FUT1, LEFTY1, OSM, SBF2, TNXB, ANKRD1, CLK2, FUT10, LEFTY2, OSMR, SBK1, TOB1, ANKRD10, CLK3, FUT11, LEKR1, OSR1, SBK2, TOB2, ANKRD11, CLK4, FUT2, LELP1, OSR2, SBK3, TOE1, ANKRD12, CLLU1, FUT3, LEMD1, OST4, SBNO1, TOLLIP, ANKRD13A, CLLU1OS, FUT4, LEMD2, OSTC, SBNO2, TOM1, ANKRD13B, CLMN, FUT5, LEMD3, OSTF1, SBSN, TOM1L1, ANKRD13C, CLMP, FUT6, LENEP, OSTM1, SBSPON, TOM1L2, ANKRD13D, CLN3, FUT7, LENG1, OSTN, SC5D, TOMM20, ANKRD16, CLN5, FUT8, LENG8, OTC, SCAF1, TOMM20L, ANKRD17, CLN6, FUT9, LENG9, OTOA, SCAF11, TOMM22, ANKRD18A, CLN8, FUZ, LEO1, OTOF, SCAF4, TOMM34, ANKRD18B, CLNK, FXN, LEP, OTOG, SCAF8, TOMM40, ANKRD2, CLNS1A, FXR1, LEPR, OTOGL, SCAI, TOMM40L, ANKRD20A1, CLOCK, FXR2, LEPRE1, OTOL1, SCAMP1, TOMM5, ANKRD20A2, CLP1, FXYD1, LEPREL1, OTOP1, SCAMP2, TOMM6, ANKRD20A3, CLPB, FXYD2, LEPREL2, OTOP2, SCAMP3, TOMM7, ANKRD20A4, CLPP, FXYD3, LEPREL4, OTOP3, SCAMP4, TOMM70A, ANKRD22, CLPS, FXYD4, LEPROT, OTOR, SCAMP5, TONSL, ANKRD23, CLPSL1, FXYD5, LEPROTL1, OTOS, SCAND1, TOP1, ANKRD24, CLPSL2, FXYD6, LETM1, OTP, SCAND3, TOP1MT, ANKRD26, CLPTM1, FXYD6-FXYD2, LETM2, OTUB1, SCAP, TOP2A, ANKRD27, CLPTM1L, FXYD7, LETMD1, OTUB2, SCAPER, TOP2B, ANKRD28, CLPX, FYB, LEUTX, OTUD1, SCARA3, TOP3A, ANKRD29, CLRN1, FYCO1, LFNG, OTUD3, SCARA5, TOP3B, ANKRD30A, CLRN2, FYN, LGALS1, OTUD4, SCARB1, TOPAZ1,

ANKRD30B, CLRN3, FYTTD1, LGALS12, OTUD5, SCARB2, TOPBP1, ANKRD31, CLSPN, FZD1, LGALS13, OTUD6A, SCARF1, TOPORS, ANKRD32, CLSTN1, FZD10, LGALS14, OTUD6B, SCARF2, TOR1A, ANKRD33, CLSTN2, FZD2, LGALS16, OTUD7A, SCCPDH, TOR1AIP1, ANKRD33B, CLSTN3, FZD3, LGALS2, OTUD7B, SCD, TOR1AIP2, ANKRD34A, CLTA, FZD4, LGALS3, OTX1, SCD5, TOR1B, ANKRD34B, CLTB, FZD5, LGALS3BP, OTX2, SCEL, TOR2A, ANKRD34C, CLTC, FZD6, LGALS4, OVCA2, SCFD1, TOR3A, ANKRD35, CLTCL1, FZD7, LGALS7, OVCH1, SCFD2, TOR4A, ANKRD36, CLU, FZD8, LGALS7B, OVCH2, SCG2, TOX, ANKRD36B, CLUAP1, FZD9, LGALS8, OVGP1, SCG3, TOX2, ANKRD36C, CLUH, FZR1, LGALS9, OVOL1, SCG5, TOX3, ANKRD37, CLUL1, G0S2, LGALS9B, OVOL2, SCGB1A1, TOX4, ANKRD39, CLVS1, G2E3, LGALS9C, OVOL3, SCGB1C1, TP53, ANKRD40, CLVS2, G3BP1, LGALSL, OVOS, SCGB1D1, TP53AIP1, ANKRD42, CLYBL, G3BP2, LGI1, OVOS2, SCGB1D2, TP53BP1, ANKRD44, CMA1, G6PC, LGI2, OXA1L, SCGB1D4, TP53BP2, ANKRD45, CMAS, G6PC2, LGI3, OXCT1, SCGB2A1, TP53I11, ANKRD46, CMBL, G6PC3, LGI4, OXCT2, SCGB2A2, TP53I13, ANKRD49, CMC1, G6PD, LGMN, OXER1, SCGB2B2, TP53I3, ANKRD50, CMC2, GAA, LGR4, OXGR1, SCGB3A1, TP53INP1, ANKRD52, CMC4, GAB1, LGR5, OXLD1, SCGB3A2, TP53INP2, ANKRD53, CMIP, GAB2, LGR6, OXNAD1, SCGN, TP53RK, ANKRD54, CMKLR1, GAB3, LGSN, OXR1, SCHIP1, TP53TG3, ANKRD55, CMPK1, GAB4, LHB, OXSM, SCIMP, TP53TG3B, ANKRD6, CMPK2, GABARAP, LHCGR, OXSR1, SCIN, TP53TG3C, ANKRD60, CMSS1, GABARAPL1, LHFP, OXT, SCLT1, TP53TG3D, ANKRD61, CMTM1, GABARAPL2, LHFPL1, OXTR, SCLY, TP53TG5, ANKRD62, CMTM2, GABBR1, LHFPL2, P2RX1, SCMH1, TP63, ANKRD63, CMTM3, GABBR2, LHFPL3, P2RX2, SCML1, TP73, ANKRD65, CMTM4, GABPA, LHFPL4, P2RX3, SCML2, TPBG, ANKRD66, CMTM5, GABPB1, LHFPL5, P2RX4, SCML4, TPBGL, ANKRD7, CMTM6, GABPB2, LHPP, P2RX5, SCN10A, TPCN1, ANKRD9, CMTM7, GABRA1, LHX1, P2RX6, SCN11A, TPCN2, ANKS1A, CMTM8, GABRA2, LHX2, P2RX7, SCN1A, TPD52, ANKS1B, CMTR1, GABRA3, LHX3, P2RY1, SCN1B, TPD52L1, ANKS3, CMTR2, GABRA4, LHX4, P2RY10, SCN2A, TPD52L2, ANKS4B, CMYA5, GABRA5, LHX5, P2RY11, SCN2B, TPD52L3, ANKS6, CNBD1, GABRA6, LHX6, P2RY12, SCN3A, TPGS1, ANKUB1, CNBD2, GABRB1, LHX8, P2RY13, SCN3B, TPGS2, ANKZF1, CNBP, GABRB2, LHX9, P2RY14, SCN4A, TPH1, ANLN, CNDP1, GABRB3, LIAS, P2RY2, SCN4B, TPH2, ANO1, CNDP2, GABRD, LIF, P2RY4, SCN5A, TPI1, ANO10, CNEP1R1, GABRE, LIFR, P2RY6, SCN7A, TPK1, ANO2, CNFN, GABRG1, LIG1, P2RY8, SCN8A, TPM1, ANO3, CNGA1, GABRG2, LIG3, P4HA1, SCN9A, TPM2, ANO4, CNGA2, GABRG3, LIG4, P4HA2, SCNM1, TPM3, ANO5, CNGA3, GABRP, LILRA1, P4HA3, SCNN1A, TPM4, ANO6, CNGA4, GABRQ, LILRA2, P4HB, SCNN1B, TPMT, ANO7, CNGB1, GABRR1, LILRA3, P4HTM, SCNN1D, TPO, ANO8, CNGB3, GABRR2, LILRA4, PA2G4, SCNN1G, TPP1, ANO9, CNIH1, GABRR3, LILRA5, PAAF1, SCO1, TPP2, ANP32A, CNIH2, GAD1, LILRA6, PABPC1, SCO2, TPPP, ANP32B, CNIH3, GAD2, LILRB1, PABPC1L, SCOC, TPPP2, ANP32C, CNIH4, GADD45A, LILRB2, PABPC1L2A, SCP2, TPPP3, ANP32D, CNKSR1,

GADD45B, LILRB3, PABPC1L2B, SCP2D1, TPR, ANP32E, CNKSR2, GADD45G, LILRB4, PABPC3, SCPEP1, TPRA1, ANPEP, CNKSR3, GADD45GIP1, LILRB5, PABPC4, SCRG1, TPRG1, ANTXR1, CNN1, GADL1, LIM2, PABPC4L, SCRIB, TPRG1L, ANTXR2, CNN2, GAGE1, LIMA1, PABPC5, SCRN1, TPRKB, ANTXRL, CNN3, GAGE10, LIMCH1, PABPN1, SCRN2, TPRN, ANXA1, CNNM1, GAGE12B, LIMD1, PABPN1L, SCRN3, TPRX1, ANXA10, CNNM2, GAGE12C, LIMD2, PACRG, SCRT1, TPSAB1, ANXA11, CNNM3, GAGE12D, LIME1, PACRGL, SCRT2, TPSB2, ANXA13, CNNM4, GAGE12E, LIMK1, PACS1, SCT, TPSD1, ANXA2, CNOT1, GAGE12F, LIMK2, PACS2, SCTR, TPSG1, ANXA2R, CNOT10, GAGE12G, LIMS1, PACSIN1, SCUBE1, TPST1, ANXA3, CNOT11, GAGE12H, LIMS2, PACSIN2, SCUBE2, TPST2, ANXA4, CNOT2, GAGE12I, LIMS3, PACSIN3, SCUBE3, TPT1, ANXA5, CNOT3, GAGE12J, LIN28A, PADI1, SCXA, TPTE, ANXA6, CNOT4, GAGE13, LIN28B, PADI2, SCXB, TPTE2, ANXA7, CNOT6, GAGE2A, LIN37, PADI3, SCYL1, TPX2, ANXA8, CNOT6L, GAGE2B, LIN52, PADI4, SCYL2, TRA2A, ANXA8L1, CNOT7, GAGE2C, LIN54, PADI6, SCYL3, TRA2B, ANXA8L2, CNOT8, GAGE2D, LIN7A, PAEP, SDAD1, TRABD, ANXA9, CNP, GAGE2E, LIN7B, PAF1, SDC1, TRABD2A, AOAH, CNPPD1, GAGE4, LIN7C, PAFAH1B1, SDC2, TRABD2B, AOC1, CNPY1, GAGE5, LIN9, PAFAH1B2, SDC3, TRADD, AOC2, CNPY2, GAGE6, LINC00452, PAFAH1B3, SDC4, TRAF1, AOC3, CNPY3, GAGE7, LINC00984, PAFAH2, SDCBP, TRAF2, AOX1, CNPY4, GAGE8, LINGO1, PAG1, SDCBP2, TRAF3, AP1AR, CNR1, GAK, LINGO2, PAGE1, SDCCAG3, TRAF3IP1, AP1B1, CNR2, GAL, LINGO3, PAGE2, SDCCAG8, TRAF3IP2, AP1G1, CNRIP1, GAL3ST1, LINGO4, PAGE2B, SDE2, TRAF3IP3, AP1G2, CNST, GAL3ST2, LINS, PAGE4, SDF2, TRAF4, AP1M1, CNTD1, GAL3ST3, LIPA, PAGE5, SDF2L1, TRAF5, AP1M2, CNTD2, GAL3ST4, LIPC, PAGR1, SDF4, TRAF6, AP1S1, CNTF, GALC, LIPE, PAH, SDHA, TRAF7, AP1S2, CNTFR, GALE, LIPF, PAICS, SDHAF1, TRAFD1, AP1S3, CNTLN, GALK1, LIPG, PAIP1, SDHAF2, TRAIP, AP2A1, CNTN1, GALK2, LIPH, PAIP2, SDHB, TRAK1, AP2A2, CNTN2, GALM, LIPI, PAIP2B, SDHC, TRAK2, AP2B1, CNTN3, GALNS, LIPJ, PAK1, SDHD, TRAM1, AP2M1, CNTN4, GALNT1, LIPK, PAK1IP1, SDK1, TRAM1L1, AP2S1, CNTN5, GALNT10, LIPM, PAK2, SDK2, TRAM2, AP3B1, CNTN6, GALNT11, LIPN, PAK3, SDPR, TRANK1, AP3B2, CNTNAP1, GALNT12, LIPT1, PAK4, SDR16C5, TRAP1, AP3D1, CNTNAP2, GALNT13, LIPT2, PAK6, SDR39U1, TRAPPC1, AP3M1, CNTNAP3, GALNT14, LITAF, PAK7, SDR42E1, TRAPPC10, AP3M2, CNTNAP3B, GALNT15, LIX1, PALB2, SDR9C7, TRAPPC11, AP3S1, CNTNAP4, GALNT16, LIX1L, PALD1, SDS, TRAPPC12, AP3S2, CNTNAP5, GALNT18, LLGL1, PALLD, SDSL, TRAPPC13, AP4B1, CNTRL, GALNT2, LLGL2, PALM, SEBOX, TRAPPC2, AP4E1, CNTROB, GALNT3, LLPH, PALM2, SEC11A, TRAPPC2L, AP4M1, COA1, GALNT4, LMAN1, PALM2-AKAP2, SEC11C, TRAPPC3, AP4S1, COA3, GALNT5, LMAN1L, PALM3, SEC13, TRAPPC3L, AP5B1, COA4, GALNT6, LMAN2, PALMD, SEC14L1, TRAPPC4, AP5M1, COA5, GALNT7, LMAN2L, PAM, SEC14L2, TRAPPC5, AP5S1, COA6, GALNT8, LMBR1, PAM16, SEC14L3, TRAPPC6A, AP5Z1, COASY, GALNT9, LMBR1L, PAMR1, SEC14L4, TRAPPC6B, APAF1, COBL, GALNTL5, LMBRD1, PAN2, SEC14L5, TRAPPC8,

APBA1, COBLL1, GALNTL6, LMBRD2, PAN3, SEC14L6, TRAPPC9, APBA2, COCH, GALP, LMCD1, PANK1, SEC16A, TRAT1, APBA3, COG1, GALR1, LMF1, PANK2, SEC16B, TRDMT1, APBB1, COG2, GALR2, LMF2, PANK3, SEC22A, TRDN, APBB1IP, COG3, GALR3, LMLN, PANK4, SEC22B, TREH, APBB2, COG4, GALT, LMNA, PANX1, SEC22C, TREM1, APBB3, COG5, GAMT, LMNB1, PANX2, SEC23A, TREM2, APC, COG6, GAN, LMNB2, PANX3, SEC23B, TREML1, APC2, COG7, GANAB, LMO1, PAOX, SEC23IP, TREML2, APCDD1, COG8, GANC, LMO2, PAPD4, SEC24A, TREML4, APCDD1L, COIL, GAP43, LMO3, PAPD5, SEC24B, TRERF1, APCS, COL10A1, GAPDH, LMO4, PAPD7, SEC24C, TREX1, APEH, COL11A1, GAPDHS, LMO7, PAPL, SEC24D, TREX2, APEX1, COL11A2, GAPT, LMOD1, PAPLN, SEC31A, TRH, APEX2, COL12A1, GAPVD1, LMOD2, PAPOLA, SEC31B, TRHDE, APH1A, COL13A1, GAR1, LMOD3, PAPOLB, SEC61A1, TRHR, APH1B, COL14A1, GAREM, LMTK2, PAPOLG, SEC61A2, TRIAP1, API5, COL15A1, GAREML, LMTK3, PAPPA, SEC61B, TRIB1, APIP, COL16A1, GARNL3, LMX1A, PAPPA2, SEC61G, TRIB2, APITD1, COL17A1, GARS, LMX1B, PAPSS1, SEC62, TRIB3, APITD1-CORT, COL18A1, GART, LNP1, PAPSS2, SEC63, TRIL, APLF, COL19A1, GAS1, LNPEP, PAQR3, SECISBP2, TRIM10, APLN, COL1A1, GAS2, LNX1, PAQR4, SECISBP2L, TRIM11, APLNR, COL1A2, GAS2L1, LNX2, PAQR5, SECTM1, TRIM13, APLP1, COL20A1, GAS2L2, LOH12CR1, PAQR6, SEH1L, TRIM14, APLP2, COL21A1, GAS2L3, LONP1, PAQR7, SEL1L, TRIM15, APMAP, COL22A1, GAS6, LONP2, PAQR8, SEL1L2, TRIM16, APOA1, COL23A1, GAS7, LONRF1, PAQR9, SEL1L3, TRIM16L, APOA1BP, COL24A1, GAS8, LONRF2, PARD3, SELE, TRIM17, APOA2, COL25A1, GAST, LONRF3, PARD3B, SELENBP1, TRIM2, APOA4, COL26A1, GATA1, LOR, PARD6A, SELK, TRIM21, APOA5, COL27A1, GATA2, LOX, PARD6B, SELL, TRIM22, APOB, COL28A1, GATA3, LOXHD1, PARD6G, SELM, TRIM23, APOBEC1, COL2A1, GATA4, LOXL1, PARG, SELO, TRIM24, APOBEC2, COL3A1, GATA5, LOXL2, PARK2, SELP, TRIM25, APOBEC3A, COL4A1, GATA6, LOXL3, PARK7, SELPLG, TRIM26, APOBEC3B, COL4A2, GATAD1, LOXL4, PARL, SELRC1, TRIM27, APOBEC3C, COL4A3, GATAD2A, LPA, PARM1, SELT, TRIM28, APOBEC3D, COL4A3BP, GATAD2B, LPAR1, PARN, SELV, TRIM29, APOBEC3F, COL4A4, GATC, LPAR2, PARP1, SEMA3A, TRIM3, APOBEC3G, COL4A5, GATM, LPAR3, PARP10, SEMA3B, TRIM31, APOBEC3H, COL4A6, GATS, LPAR4, PARP11, SEMA3C, TRIM32, APOBEC4, COL5A1, GATSL1, LPAR5, PARP12, SEMA3D, TRIM33, APOBR, COL5A2, GATSL2, LPAR6, PARP14, SEMA3E, TRIM34, APOC1, COL5A3, GATSL3, LPCAT1, PARP15, SEMA3F, TRIM35, APOC2, COL6A1, GBA, LPCAT2, PARP16, SEMA3G, TRIM36, APOC3, COL6A2, GBA2, LPCAT3, PARP2, SEMA4A, TRIM37, APOC4, COL6A3, GBA3, LPCAT4, PARP3, SEMA4B, TRIM38, APOD, COL6A5, GBAS, LPGAT1, PARP4, SEMA4C, TRIM39, APOE, COL6A6, GBE1, LPHN1, PARP6, SEMA4D, TRIM39-RPP21, APOF, COL7A1, GBF1, LPHN2, PARP8, SEMA4F, TRIM4, APOH, COL8A1, GBGT1, LPHN3, PARP9, SEMA4G, TRIM40, APOL1, COL8A2, GBP1, LPIN1, PARPBP, SEMA5A, TRIM41, APOL2, COL9A1, GBP2, LPIN2, PARS2, SEMA5B, TRIM42, APOL3, COL9A2, GBP3, LPIN3, PARVA, SEMA6A,

TRIM43, APOL4, COL9A3, GBP4, LPL, PARVB, SEMA6B, TRIM43B, APOL5, COLCA2, GBP5, LPO, PARVG, SEMA6C, TRIM44, APOL6, COLEC10, GBP6, LPP, PASD1, SEMA6D, TRIM45, APOLD1, COLEC11, GBP7, LPPR1, PASK, SEMA7A, TRIM46, APOM, COLEC12, GBX1, LPPR2, PATE1, SEMG1, TRIM47, APOO, COLGALT1, GBX2, LPPR3, PATE2, SEMG2, TRIM48, APOOL, COLGALT2, GC, LPPR4, PATE3, SENP1, TRIM49, APOPT1, COLQ, GCA, LPPR5, PATE4, SENP2, TRIM49B, APP, COMMD1, GCAT, LPXN, PATL1, SENP3, TRIM49C, APPBP2, COMMD10, GCC1, LRAT, PATL2, SENP5, TRIM49D1, APPL1, COMMD2, GCC2, LRBA, PATZ1, SENP6, TRIM49D2P, APPL2, COMMD3, GCDH, LRCH1, PAWR, SENP7, TRIM5, APRT, COMMD3-BMI1, GCFC2, LRCH2, PAX1, SENP8, TRIM50, APTX, COMMD4, GCG, LRCH3, PAX2, SEPHS1, TRIM51, AQP1, COMMD5, GCGR, LRCH4, PAX3, SEPHS2, TRIM52, AQP10, COMMD6, GCH1, LRCOL1, PAX4, SEPN1, TRIM54, AQP11, COMMD7, GCHFR, LRFN1, PAX5, SEPP1, TRIM55, AQP12A, COMMD8, GCK, LRFN2, PAX6, SEPSECS, TRIM56, AQP12B, COMMD9, GCKR, LRFN3, PAX7, SEPT1, TRIM58, AQP2, COMP, GCLC, LRFN4, PAX8, SEPT10, TRIM59, AQP3, COMT, GCLM, LRFN5, PAX9, SEPT11, TRIM6, AQP4, COMTD1, GCM1, LRG1, PAXBP1, SEPT12, TRIM60, AQP5, COPA, GCM2, LRGUK, PAXIP1, SEPT14, TRIM61, AQP6, COPB1, GCN1L1, LRIF1, PBDC1, SEPT15, TRIM62, AQP7, COPB2, GCNT1, LRIG1, PBK, SEPT2, TRIM63, AQP8, COPE, GCNT2, LRIG2, PBLD, SEPT3, TRIM64, AQP9, COPG1, GCNT3, LRIG3, PBOV1, SEPT4, TRIM64B, AQPEP, COPG2, GCNT4, LRIT1, PBRM1, SEPT5, TRIM64C, AQR, COPRS, GCNT7, LRIT2, PBX1, SEPT6, TRIM65, AR, COPS2, GCOM1, LRIT3, PBX2, SEPT7, TRIM66, ARAF, COPS3, GCSAM, LRMP, PBX3, SEPT8, TRIM67, ARAP1, COPS4, GCSAML, LRP1, PBX4, SEPT9, TRIM68, ARAP2, COPS5, GCSH, LRP10, PBXIP1, SEPW1, TRIM69, ARAP3, COPS6, GDA, LRP11, PC, SERAC1, TRIM6-TRIM34, ARC, COPS7A, GDAP1, LRP12, PCBD1, SERBP1, TRIM7, ARCN1, COPS7B, GDAP1L1, LRP1B, PCBD2, SERF1A, TRIM71, AREG, COPS8, GDAP2, LRP2, PCBP1, SERF1B, TRIM72, AREGB, COPZ1, GDE1, LRP2BP, PCBP2, SERF2, TRIM73, AREL1, COPZ2, GDF1, LRP3, PCBP3, SERGEF, TRIM74, ARF1, COQ10A, GDF10, LRP4, PCBP4, SERHL2, TRIM77, ARF3, COQ10B, GDF11, LRP5, PCCA, SERINC1, TRIM8, ARF4, COQ2, GDF15, LRP5L, PCCB, SERINC2, TRIM9, ARF5, COQ3, GDF2, LRP6, PCDH1, SERINC3, TRIML1, ARF6, COQ4, GDF3, LRP8, PCDH10, SERINC4, TRIML2, ARFGAP1, COQ5, GDF5, LRPAP1, PCDH11X, SERINC5, TRIO, ARFGAP2, COQ6, GDF6, LRPPRC, PCDH11Y, SERP1, TRIOBP, ARFGAP3, COQ7, GDF7, LRR1, PCDH12, SERP2, TRIP10, ARFGEF1, COQ9, GDF9, LRRC1, PCDH15, SERPINA1, TRIP11, ARFGEF2, CORIN, GDI1, LRRC10, PCDH17, SERPINA10, TRIP12, ARFIP1, CORO1A, GDI2, LRRC10B, PCDH18, SERPINA11, TRIP13, ARFIP2, CORO1B, GDNF, LRRC14, PCDH19, SERPINA12, TRIP4, ARFRP1, CORO1C, GDNF-AS1, LRRC14B, PCDH20, SERPINA3, TRIP6, ARG1, CORO2A, GDPD1, LRRC15, PCDH7, SERPINA4, TRIQK, ARG2, CORO2B, GDPD2, LRRC16A, PCDH8, SERPINA5, TRIT1, ARGFX, CORO6, GDPD3, LRRC16B, PCDH9, SERPINA6, TRMT1, ARGLU1, CORO7, GDPD4, LRRC17, PCDHA1, SERPINA7, TRMT10A, ARHGAP1, CORO7-PAM16, GDPD5, LRRC18, PCDHA10, SERPINA9, TRMT10B, ARHGAP10,

CORT, GDPGP1, LRRC19, PCDHA11, SERPINB1, TRMT10C, ARHGAP11A, COTL1, GEM, LRRC2, PCDHA12, SERPINB10, TRMT11, ARHGAP11B, COX10, GEMIN2, LRRC20, PCDHA13, SERPINB11, TRMT112, ARHGAP12, COX11, GEMIN4, LRRC23, PCDHA2, SERPINB12, TRMT12, ARHGAP15, COX14, GEMIN5, LRRC24, PCDHA3, SERPINB13, TRMT13, ARHGAP17, COX15, GEMIN6, LRRC25, PCDHA4, SERPINB2, TRMT1L, ARHGAP18, COX16, GEMIN7, LRRC26, PCDHA5, SERPINB3, TRMT2A, ARHGAP19, COX17, GEMIN8, LRRC27, PCDHA6, SERPINB4, TRMT2B, ARHGAP20, COX18, GEN1, LRRC28, PCDHA7, SERPINB5, TRMT44, ARHGAP21, COX19, GET4, LRRC29, PCDHA8, SERPINB6, TRMT5, ARHGAP22, COX20, GFAP, LRRC3, PCDHA9, SERPINB7, TRMT6, ARHGAP23, COX4I1, GFER, LRRC30, PCDHAC1, SERPINB8, TRMT61A, ARHGAP24, COX4I2, GFI1, LRRC31, PCDHAC2, SERPINB9, TRMT61B, ARHGAP25, COX5A, GFI1B, LRRC32, PCDHB1, SERPINC1, TRMU, ARHGAP26, COX5B, GFM1, LRRC34, PCDHB10, SERPIND1, TRNAU1AP, ARHGAP27, COX6A1, GFM2, LRRC36, PCDHB11, SERPINE1, TRNP1, ARHGAP28, COX6A2, GFOD1, LRRC37A2, PCDHB12, SERPINE2, TRNT1, ARHGAP29, COX6B1, GFOD2, LRRC37A3, PCDHB13, SERPINE3, TRO, ARHGAP30, COX6B2, GFPT1, LRRC37B, PCDHB14, SERPINF1, TROAP, ARHGAP31, COX6C, GFPT2, LRRC38, PCDHB15, SERPINF2, TROVE2, ARHGAP32, COX7A1, GFRA1, LRRC39, PCDHB16, SERPING1, TRPA1, ARHGAP33, COX7A2, GFRA2, LRRC3B, PCDHB2, SERPINH1, TRPC1, ARHGAP35, COX7A2L, GFRA3, LRRC3C, PCDHB3, SERPINI1, TRPC3, ARHGAP36, COX7B, GFRA4, LRRC4, PCDHB4, SERPINI2, TRPC4, ARHGAP39, COX7B2, GFRAL, LRRC40, PCDHB5, SERTAD1, TRPC4AP, ARHGAP4, COX7C, GGA1, LRRC41, PCDHB6, SERTAD2, TRPC5, ARHGAP40, COX8A, GGA2, LRRC42, PCDHB7, SERTAD3, TRPC5OS, ARHGAP42, COX8C, GGA3, LRRC43, PCDHB8, SERTAD4, TRPC6, ARHGAP44, CP, GGACT, LRRC45, PCDHB9, SERTM1, TRPC7, ARHGAP5, CPA1, GGCT, LRRC46, PCDHGA1, SESN1, TRPM1, ARHGAP6, CPA2, GGCX, LRRC47, PCDHGA10, SESN2, TRPM2, ARHGAP8, CPA3, GGH, LRRC48, PCDHGA11, SESN3, TRPM3, ARHGAP9, CPA4, GGN, LRRC49, PCDHGA12, SESTD1, TRPM4, ARHGDIA, CPA5, GGNBP2, LRRC4B, PCDHGA2, SET, TRPM5, ARHGDIB, CPA6, GGPS1, LRRC4C, PCDHGA3, SETBP1, TRPM6, ARHGDIG, CPAMD8, GGT1, LRRC52, PCDHGA4, SETD1A, TRPM7, ARHGEF1, CPB1, GGT2, LRRC55, PCDHGA5, SETD1B, TRPM8, ARHGEF10, CPB2, GGT5, LRRC56, PCDHGA6, SETD2, TRPS1, ARHGEF10L, CPD, GGT6, LRRC57, PCDHGA7, SETD3, TRPT1, ARHGEF11, CPE, GGT7, LRRC58, PCDHGA8, SETD4, TRPV1, ARHGEF12, CPEB1, GGTLC1, LRRC59, PCDHGA9, SETD5, TRPV2, ARHGEF15, CPEB2, GGTLC2, LRRC6, PCDHGB1, SETD6, TRPV3, ARHGEF16, CPEB3, GH1, LRRC61, PCDHGB2, SETD7, TRPV4, ARHGEF17, CPEB4, GH2, LRRC63, PCDHGB3, SETD8, TRPV5, ARHGEF18, CPED1, GHDC, LRRC66, PCDHGB4, SETD9, TRPV6, ARHGEF19, CPLX1, GHITM, LRRC69, PCDHGB5, SETDB1, TRRAP, ARHGEF2, CPLX2, GHR, LRRC7, PCDHGB6, SETDB2, TRUB1, ARHGEF25, CPLX3, GHRH, LRRC70, PCDHGB7, SETMAR, TRUB2, ARHGEF26, CPLX4, GHRHR, LRRC71, PCDHGC3, SETSIP, TSACC, ARHGEF28, CPM, GHRL, LRRC72, PCDHGC4, SETX, TSC1, ARHGEF3, CPN1, GHSR,

LRRC73, PCDHGC5, SEZ6, TSC2, ARHGEF33, CPN2, GID4, LRRC8A, PCDP1, SEZ6L, TSC22D1, ARHGEF35, CPNE1, GID8, LRRC8B, PCED1A, SEZ6L2, TSC22D2, ARHGEF37, CPNE2, GIF, LRRC8C, PCED1B, SF1, TSC22D3, ARHGEF38, CPNE3, GIGYF1, LRRC8D, PCF11, SF3A1, TSC22D4, ARHGEF39, CPNE4, GIGYF2, LRRC8E, PCGF1, SF3A2, TSEN15, ARHGEF4, CPNE5, GIMAP1, LRRCC1, PCGF2, SF3A3, TSEN2, ARHGEF40, CPNE6, GIMAP1-GIMAP5, LRRD1, PCGF3, SF3B1, TSEN34, ARHGEF5, CPNE7, GIMAP2, LRRFIP1, PCGF5, SF3B14, TSEN54, ARHGEF6, CPNE8, GIMAP4, LRRFIP2, PCGF6, SF3B2, TSFM, ARHGEF7, CPNE9, GIMAP5, LRRIQ1, PCID2, SF3B3, TSG101, ARHGEF9, CPO, GIMAP6, LRRIQ3, PCIF1, SF3B4, TSGA10, ARID1A, CPOX, GIMAP7, LRRIQ4, PCK1, SF3B5, TSGA10IP, ARID1B, CPPED1, GIMAP8, LRRK1, PCK2, SFI1, TSGA13, ARID2, CPQ, GIMD1, LRRK2, PCLO, SFMBT1, TSHB, ARID3A, CPS1, GIN1, LRRN1, PCM1, SFMBT2, TSHR, ARID3B, CPSF1, GINM1, LRRN2, PCMT1, SFN, TSHZ1, ARID3C, CPSF2, GINS1, LRRN3, PCMTD1, SFPQ, TSHZ2, ARID4A, CPSF3, GINS2, LRRN4, PCMTD2, SFR1, TSHZ3, ARID4B, CPSF3L, GINS3, LRRN4CL, PCNA, SFRP1, TSKS, ARID5A, CPSF4, GINS4, LRRTM1, PCNP, SFRP2, TSKU, ARID5B, CPSF4L, GIP, LRRTM2, PCNT, SFRP4, TSLP, ARIH1, CPSF6, GIPC1, LRRTM3, PCNX, SFRP5, TSN, ARIH2, CPSF7, GIPC2, LRRTM4, PCNXL2, SFSWAP, TSNARE1, ARIH2OS, CPT1A, GIPC3, LRSAM1, PCNXL3, SFT2D1, TSNAX, ARL1, CPT1B, GIPR, LRTM1, PCNXL4, SFT2D2, TSNAXIP1, ARL10, CPT1C, GIT1, LRTM2, PCOLCE, SFT2D3, TSPAN1, ARL11, CPT2, GIT2, LRTOMT, PCOLCE2, SFTA2, TSPAN10, ARL13A, CPVL, GJA1, LRWD1, PCP2, SFTA3, TSPAN11, ARL13B, CPXCR1, GJA10, LSAMP, PCP4, SFTPA1, TSPAN12, ARL14, CPXM1, GJA3, LSG1, PCP4L1, SFTPA2, TSPAN13, ARL14EP, CPXM2, GJA4, LSM1, PCSK1, SFTPB, TSPAN14, ARL14EPL, CPZ, GJA5, LSM10, PCSK1N, SFTPC, TSPAN15, ARL15, CR1, GJA8, LSM11, PCSK2, SFTPD, TSPAN16, ARL16, CR1L, GJA9, LSM12, PCSK4, SFXN1, TSPAN17, ARL17A, CR2, GJB1, LSM14A, PCSK5, SFXN2, TSPAN18, ARL17B, CRABP1, GJB2, LSM14B, PCSK6, SFXN3, TSPAN19, ARL2, CRABP2, GJB3, LSM2, PCSK7, SFXN4, TSPAN2, ARL2BP, CRADD, GJB4, LSM3, PCSK9, SFXN5, TSPAN3, ARL3, CRAMP1L, GJB5, LSM4, PCTP, SGCA, TSPAN31, ARL4A, CRAT, GJB6, LSM5, PCYOX1, SGCB, TSPAN32, ARL4C, CRB1, GJB7, LSM6, PCYOX1L, SGCD, TSPAN33, ARL4D, CRB2, GJC1, LSM7, PCYT1A, SGCE, TSPAN4, ARL5A, CRB3, GJC2, LSMD1, PCYT1B, SGCG, TSPAN5, ARL5B, CRBN, GJC3, LSMEM1, PCYT2, SGCZ, TSPAN6, ARL5C, CRCP, GJD2, LSMEM2, PDAP1, SGIP1, TSPAN7, ARL6, CRCT1, GJD3, LSP1, PDC, SGK1, TSPAN8, ARL6IP1, CREB1, GJD4, LSR, PDCD1, SGK2, TSPAN9, ARL6IP4, CREB3, GK, LSS, PDCD10, SGK223, TSPEAR, ARL6IP5, CREB3L1, GK2, LST1, PDCD11, SGK3, TSPO, ARL6IP6, CREB3L2, GK5, LTA, PDCD1LG2, SGK494, TSPO2, ARL8A, CREB3L3, GKAP1, LTA4H, PDCD2, SGMS1, TSPY1, ARL8B, CREB3L4, GKN1, LTB, PDCD2L, SGMS2, TSPY10, ARL9, CREB5, GKN2, LTB4R, PDCD4, SGOL1, TSPY2, ARMC1, CREBBP, GLA, LTB4R2, PDCD5, SGOL2, TSPY3, ARMC10, CREBL2, GLB1, LTBP1, PDCD6, SGPL1, TSPY4, ARMC12, CREBRF, GLB1L, LTBP2, PDCD6IP, SGPP1, TSPY8, ARMC2, CREBZF, GLB1L2, LTBP3, PDCD7, SGPP2, TSPYL1,

ARMC3, CREG1, GLB1L3, LTBP4, PDCL, SGSH, TSPYL2, ARMC4, CREG2, GLCCI1, LTBR, PDCL2, SGSM1, TSPYL4, ARMC5, CRELD1, GLCE, LTC4S, PDCL3, SGSM2, TSPYL5, ARMC6, CRELD2, GLDC, LTF, PDDC1, SGSM3, TSPYL6, ARMC7, CREM, GLDN, LTK, PDE10A, SGTA, TSR1, ARMC8, CRH, GLE1, LTN1, PDE11A, SGTB, TSR2, ARMC9, CRHBP, GLG1, LTV1, PDE12, SH2B1, TSR3, ARMCX1, CRHR1, GLI1, LUC7L, PDE1A, SH2B2, TSSC1, ARMCX2, CRHR2, GLI2, LUC7L2, PDE1B, SH2B3, TSSC4, ARMCX3, CRIM1, GLI3, LUC7L3, PDE1C, SH2D1A, TSSK1B, ARMCX4, CRIP1, GLI4, LUM, PDE2A, SH2D1B, TSSK2, ARMCX5, CRIP2, GLIPR1, LURAP1, PDE3A, SH2D2A, TSSK3, ARMCX5-GPRASP2, CRIP3, GLIPR1L1, LURAP1L, PDE3B, SH2D3A, TSSK4, ARMCX6, CRIPAK, GLIPR1L2, LUZP1, PDE4A, SH2D3C, TSSK6, ARMS2, CRIPT, GLIPR2, LUZP2, PDE4B, SH2D4A, TST, ARNT, CRISP1, GLIS1, LUZP4, PDE4C, SH2D4B, TSTA3, ARNT2, CRISP2, GLIS2, LUZP6, PDE4D, SH2D5, TSTD1, ARNTL, CRISP3, GLIS3, LXN, PDE4DIP, SH2D6, TSTD2, ARNTL2, CRISPLD1, GLMN, LY6D, PDE5A, SH2D7, TSTD3, ARPC1A, CRISPLD2, GLO1, LY6E, PDE6A, SH3BGR, TTBK1, ARPC1B, CRK, GLOD4, LY6G5B, PDE6B, SH3BGRL, TTBK2, ARPC2, CRKL, GLOD5, LY6G5C, PDE6C, SH3BGRL2, TTC1, ARPC3, CRLF1, GLP1R, LY6G6C, PDE6D, SH3BGRL3, TTC12, ARPC4, CRLF2, GLP2R, LY6G6D, PDE6G, SH3BP1, TTC13, ARPC4-TTLL3, CRLF3, GLRA1, LY6G6F, PDE6H, SH3BP2, TTC14, ARPC5, CRLS1, GLRA2, LY6H, PDE7A, SH3BP4, TTC16, ARPC5L, CRMP1, GLRA3, LY6K, PDE7B, SH3BP5, TTC17, ARPP19, CRNKL1, GLRA4, LY75, PDE8A, SH3BP5L, TTC18, ARPP21, CRNN, GLRB, LY75-CD302, PDE8B, SH3D19, TTC19, ARR3, CROCC, GLRX, LY86, PDE9A, SH3D21, TTC21A, ARRB1, CROT, GLRX2, LY9, PDF, SH3GL1, TTC21B, ARRB2, CRP, GLRX3, LY96, PDGFA, SH3GL2, TTC22, ARRDC1, CRTAC1, GLRX5, LYAR, PDGFB, SH3GL3, TTC23, ARRDC2, CRTAM, GLS, LYG1, PDGFC, SH3GLB1, TTC23L, ARRDC3, CRTAP, GLS2, LYG2, PDGFD, SH3GLB2, TTC24, ARRDC4, CRTC1, GLT1D1, LYL1, PDGFRA, SH3KBP1, TTC25, ARRDC5, CRTC2, GLT6D1, LYN, PDGFRB, SH3PXD2A, TTC26, ARSA, CRTC3, GLT8D1, LYNX1, PDGFRL, SH3PXD2B, TTC27, ARSB, CRX, GLT8D2, LYPD1, PDHA1, SH3RF1, TTC28, ARSD, CRY1, GLTP, LYPD2, PDHA2, SH3RF2, TTC29, ARSE, CRY2, GLTPD1, LYPD3, PDHB, SH3RF3, TTC3, ARSF, CRYAA, GLTPD2, LYPD4, PDHX, SH3TC1, TTC30A, ARSG, CRYAB, GLTSCR1, LYPD5, PDIA2, SH3TC2, TTC30B, ARSH, CRYBA1, GLTSCR1L, LYPD6, PDIA3, SH3YL1, TTC31, ARSI, CRYBA2, GLTSCR2, LYPD6B, PDIA4, SHANK1, TTC32, ARSJ, CRYBA4, GLUD1, LYPD8, PDIA5, SHANK2, TTC33, ARSK, CRYBB1, GLUD2, LYPLA1, PDIA6, SHANK3, TTC34, ART1, CRYBB2, GLUL, LYPLA2, PDIK1L, SHARPIN, TTC36, ART3, CRYBB3, GLYAT, LYPLAL1, PDILT, SHB, TTC37, ART4, CRYBG3, GLYATL1, LYRM1, PDK1, SHBG, TTC38, ART5, CRYGA, GLYATL2, LYRM2, PDK2, SHC1, TTC39A, ARTN, CRYGB, GLYATL3, LYRM4, PDK3, SHC2, TTC39B, ARV1, CRYGC, GLYCTK, LYRM5, PDK4, SHC3, TTC39C, ARVCF, CRYGD, GLYR1, LYRM7, PDLIM1, SHC4, TTC4, ARX, CRYGN, GM2A, LYRM9, PDLIM2, SHCBP1, TTC40, AS3MT, CRYGS, GMCL1, LYSMD1, PDLIM3, SHCBP1L, TTC5, ASAH1, CRYL1, GMDS, LYSMD2, PDLIM4, SHD, TTC6, ASAH2, CRYM, GMEB1, LYSMD3, PDLIM5,

SHE, TTC7A, ASAH2B, CRYZ, GMEB2, LYSMD4, PDLIM7, SHF, TTC7B, ASAP1, CRYZL1, GMFB, LYST, PDP1, SHFM1, TTC8, ASAP2, CS, GMFG, LYVE1, PDP2, SHH, TTC9, ASAP3, CSAD, GMIP, LYZ, PDPK1, SHISA2, TTC9B, ASB1, CSAG1, GML, LYZL1, PDPN, SHISA3, TTC9C, ASB10, CSAG2, GMNC, LYZL2, PDPR, SHISA4, TTF1, ASB11, CSAG3, GMNN, LYZL4, PDRG1, SHISA5, TTF2, ASB12, CSDC2, GMPPA, LYZL6, PDS5A, SHISA6, TTI1, ASB13, CSDE1, GMPPB, LZIC, PDS5B, SHISA7, TTI2, ASB14, CSE1L, GMPR, LZTFL1, PDSS1, SHISA8, TTK, ASB15, CSF1, GMPR2, LZTR1, PDSS2, SHISA9, TTL, ASB16, CSF1R, GMPS, LZTS1, PDX1, SHKBP1, TTLL1, ASB17, CSF2, GNA11, LZTS2, PDXDC1, SHMT1, TTLL10, ASB18, CSF2RA, GNA12, LZTS3, PDXK, SHMT2, TTLL11, ASB2, CSF2RB, GNA13, M1AP, PDXP, SHOC2, TTLL12, ASB3, CSF3, GNA14, M6PR, PDYN, SHOX, TTLL13, ASB4, CSF3R, GNA15, MAATS1, PDZD11, SHOX2, TTLL2, ASB5, CSGALNACT1, GNAI1, MAB21L1, PDZD2, SHPK, TTLL3, ASB6, CSGALNACT2, GNAI2, MAB21L2, PDZD3, SHPRH, TTLL4, ASB7, CSH1, GNAI3, MAB21L3, PDZD4, SHQ1, TTLL5, ASB8, CSH2, GNAL, MACC1, PDZD7, SHROOM1, TTLL6, ASB9, CSHL1, GNAO1, MACF1, PDZD8, SHROOM2, TTLL7, ASCC1, CSK, GNAQ, MACROD1, PDZD9, SHROOM3, TTLL8, ASCC2, CSMD1, GNAS, MACROD2, PDZK1, SHROOM4, TTLL9, ASCC3, CSMD2, GNAT1, MAD1L1, PDZK1IP1, SI, TTN, ASCL1, CSMD3, GNAT2, MAD2L1, PDZRN3, SIAE, TTPA, ASCL2, CSN1S1, GNAT3, MAD2L1BP, PDZRN4, SIAH1, TTPAL, ASCL3, CSN2, GNAZ, MAD2L2, PEA15, SIAH2, TTR, ASCL4, CSN3, GNB1, MADCAM1, PEAK1, SIAH3, TTYH1, ASCL5, CSNK1A1, GNB1L, MADD, PEAR1, SIDT1, TTYH2, ASF1A, CSNK1A1L, GNB2, MAEA, PEBP1, SIDT2, TTYH3, ASF1B, CSNK1D, GNB2L1, MAEL, PEBP4, SIGIRR, TUB, ASGR1, CSNK1E, GNB3, MAF, PECAM1, SIGLEC1, TUBA1A, ASGR2, CSNK1G1, GNB4, MAF1, PECR, SIGLEC10, TUBA1B, ASH1L, CSNK1G2, GNB5, MAFA, PEF1, SIGLEC11, TUBA1C, ASH2L, CSNK1G3, GNE, MAFB, PEG10, SIGLEC12, TUBA3C, ASIC1, CSNK2A1, GNG10, MAFF, PEG3, SIGLEC14, TUBA3D, ASIC2, CSNK2A2, GNG11, MAFG, PELI1, SIGLEC15, TUBA3E, ASIC3, CSNK2A3, GNG12, MAFK, PELI2, SIGLEC5, TUBA4A, ASIC4, CSNK2B, GNG13, MAG, PELI3, SIGLEC6, TUBA8, ASIC5, CSPG4, GNG2, MAGEA1, PELO, SIGLEC7, TUBAL3, ASIP, CSPG5, GNG3, MAGEA10, PELP1, SIGLEC8, TUBB, ASL, CSPP1, GNG4, MAGEA10-MAGEA5, PEMT, SIGLEC9, TUBB1, ASMT, CSRNP1, GNG5, MAGEA11, PENK, SIGLECL1, TUBB2A, ASMTL, CSRNP2, GNG7, MAGEA12, PEPD, SIGMAR1, TUBB2B, ASNA1, CSRNP3, GNG8, MAGEA3, PER1, SIK1, TUBB3, ASNS, CSRP1, GNGT1, MAGEA4, PER2, SIK2, TUBB4A, ASNSD1, CSRP2, GNGT2, MAGEA5, PER3, SIK3, TUBB4B, ASPA, CSRP2BP, GNL1, MAGEA6, PERM1, SIKE1, TUBB6, ASPDH, CSRP3, GNL2, MAGEA8, PERP, SIL1, TUBB8, ASPG, CST1, GNL3, MAGEB1, PES1, SIM1, TUBD1, ASPH, CST11, GNL3L, MAGEB10, PET100, SIM2, TUBE1, ASPHD1, CST2, GNLY, MAGEB16, PET112, SIMC1, TUBG1, ASPHD2, CST3, GNMT, MAGEB17, PET117, SIN3A, TUBG2, ASPM, CST4, GNPAT, MAGEB18, PEX1, SIN3B, TUBGCP2, ASPN, CST5, GNPDA1, MAGEB2, PEX10, SIPA1, TUBGCP3, ASPRV1, CST6, GNPDA2, MAGEB3, PEX11A, SIPA1L1, TUBGCP4, ASPSCR1, CST7, GNPNAT1, MAGEB4, PEX11B, SIPA1L2, TUBGCP5,

ASRGL1, CST8, GNPTAB, MAGEB5, PEX11G, SIPA1L3, TUBGCP6, ASS1, CST9, GNPTG, MAGEB6, PEX12, SIRPA, TUFM, ASTE1, CST9L, GNRH1, MAGEC1, PEX13, SIRPB1, TUFT1, ASTL, CSTA, GNRH2, MAGEC2, PEX14, SIRPB2, TULP1, ASTN1, CSTB, GNRHR, MAGEC3, PEX16, SIRPD, TULP2, ASTN2, CSTF1, GNS, MAGED1, PEX19, SIRPG, TULP3, ASUN, CSTF2, GOLGA1, MAGED2, PEX2, SIRT1, TULP4, ASXL1, CSTF2T, GOLGA2, MAGEE1, PEX26, SIRT2, TUSC1, ASXL2, CSTF3, GOLGA3, MAGEE2, PEX3, SIRT3, TUSC2, ASXL3, CSTL1, GOLGA4, MAGEF1, PEX5, SIRT4, TUSC3, ASZ1, CT45A1, GOLGA5, MAGEH1, PEX5L, SIRT5, TUSC5, ATAD1, CT45A2, GOLGA6A, MAGEL2, PEX6, SIRT6, TUT1, ATAD2, CT45A3, GOLGA6B, MAGI1, PEX7, SIRT7, TVP23A, ATAD2B, CT45A4, GOLGA6C, MAGI2, PF4, SIT1, TVP23B, ATAD3A, CT45A5, GOLGA6D, MAGI3, PF4V1, SIVA1, TVP23C, ATAD3B, CT45A6, GOLGA6L1, MAGIX, PFAS, SIX1, TVP23C-CDRT4, ATAD3C, CT47A1, GOLGA6L10, MAGOH, PFDN1, SIX2, TWF1, ATAD5, CT47A10, GOLGA6L2, MAGOHB, PFDN2, SIX3, TWF2, ATAT1, CT47A11, GOLGA6L3, MAGT1, PFDN4, SIX4, TWIST1, ATCAY, CT47A12, GOLGA6L4, MAK, PFDN5, SIX5, TWIST2, ATE1, CT47A2, GOLGA6L6, MAK16, PFDN6, SIX6, TWISTNB, ATF1, CT47A3, GOLGA6L9, MAL, PFKFB1, SKA1, TWSG1, ATF2, CT47A4, GOLGA7, MAL2, PFKFB2, SKA2, TXK, ATF3, CT47A5, GOLGA7B, MALL, PFKFB3, SKA3, TXLNA, ATF4, CT47A6, GOLGA8A, MALSU1, PFKFB4, SKAP1, TXLNB, ATF5, CT47A7, GOLGA8B, MALT1, PFKL, SKAP2, TXLNG, ATF6, CT47A8, GOLGA8H, MAMDC2, PFKM, SKI, TXN, ATF6B, CT47A9, GOLGA8J, MAMDC4, PFKP, SKIDA1, TXN2, ATF7, CT47B1, GOLGA8K, MAML1, PFN1, SKIL, TXNDC11, ATF7IP, CT62, GOLGA8M, MAML2, PFN2, SKIV2L, TXNDC12, ATF7IP2, CTAG1A, GOLGA8O, MAML3, PFN3, SKIV2L2, TXNDC15, ATG10, CTAG1B, GOLGA8R, MAMLD1, PFN4, SKOR1, TXNDC16, ATG12, CTAG2, GOLGB1, MAMSTR, PGA3, SKOR2, TXNDC17, ATG13, CTAGE1, GOLIM4, MAN1A1, PGA4, SKP1, TXNDC2, ATG14, CTAGE15, GOLM1, MAN1A2, PGA5, SKP2, TXNDC5, ATG16L1, CTAGE4, GOLPH3, MAN1B1, PGAM1, SLA, TXNDC8, ATG16L2, CTAGE5, GOLPH3L, MAN1C1, PGAM2, SLA2, TXNDC9, ATG2A, CTAGE6, GOLT1A, MAN2A1, PGAM4, SLAIN1, TXNIP, ATG2B, CTAGE8, GOLT1B, MAN2A2, PGAM5, SLAIN2, TXNL1, ATG3, CTAGE9, GON4L, MAN2B1, PGAP1, SLAMF1, TXNL4A, ATG4A, CTBP1, GOPC, MAN2B2, PGAP2, SLAMF6, TXNL4B, ATG4B, CTBP2, GORAB, MAN2C1, PGAP3, SLAMF7, TXNRD1, ATG4C, CTBS, GORASP1, MANBA, PGBD1, SLAMF8, TXNRD2, ATG4D, CTC1, GORASP2, MANBAL, PGBD2, SLAMF9, TXNRD3, ATG5, CTCF, GOSR1, MANEA, PGBD3, SLBP, TXNRD3NB, ATG7, CTCFL, GOSR2, MANEAL, PGBD4, SLC10A1, TYK2, ATG9A, CTDNEP1, GOT1, MANF, PGBD5, SLC10A2, TYMP, ATG9B, CTDP1, GOT1L1, MANSC1, PGC, SLC10A3, TYMS, ATHL1, CTDSP1, GOT2, MANSC4, PGD, SLC10A4, TYR, ATIC, CTDSP2, GP1BA, MAOA, PGF, SLC10A5, TYRO3, ATL1, CTDSPL, GP1BB, MAOB, PGGT1B, SLC10A6, TYROBP, ATL2, CTDSPL2, GP2, MAP10, PGK1, SLC10A7, TYRP1, ATL3, CTF1, GP5, MAP1A, PGK2, SLC11A1, TYSND1, ATM, CTGF, GP6, MAP1B, PGLS, SLC11A2, TYW1, ATMIN, CTH, GP9, MAP1LC3A, PGLYRP1, SLC12A1, TYW1B, ATN1, CTHRC1, GPA33, MAP1LC3B, PGLYRP2, SLC12A2, TYW3, ATOH1, CTIF,

GPAA1, MAP1LC3B2, PGLYRP3, SLC12A3, TYW5, ATOH7, CTLA4, GPALPP1, MAP1LC3C, PGLYRP4, SLC12A4, U2AF1, ATOH8, CTNNA1, GPAM, MAP1S, PGM1, SLC12A5, U2AF1L4, ATOX1, CTNNA2, GPANK1, MAP2, PGM2, SLC12A6, U2AF2, ATP10A, CTNNA3, GPAT2, MAP2K1, PGM2L1, SLC12A7, U2SURP, ATP10B, CTNNAL1, GPATCH1, MAP2K2, PGM3, SLC12A8, UACA, ATP10D, CTNNB1, GPATCH11, MAP2K3, PGM5, SLC12A9, UAP1, ATP11A, CTNNBIP1, GPATCH2, MAP2K4, PGP, SLC13A1, UAP1L1, ATP11B, CTNNBL1, GPATCH2L, MAP2K5, PGPEP1, SLC13A2, UBA1, ATP11C, CTNND1, GPATCH3, MAP2K6, PGPEP1L, SLC13A3, UBA2, ATP12A, CTNND2, GPATCH4, MAP2K7, PGR, SLC13A4, UBA3, ATP13A1, CTNS, GPATCH8, MAP3K1, PGRMC1, SLC13A5, UBA5, ATP13A2, CTPS1, GPBAR1, MAP3K10, PGRMC2, SLC14A1, UBA52, ATP13A3, CTPS2, GPBP1, MAP3K11, PGS1, SLC14A2, UBA6, ATP13A4, CTR9, GPBP1L1, MAP3K12, PHACTR1, SLC15A1, UBA7, ATP13A5, CTRB1, GPC1, MAP3K13, PHACTR2, SLC15A2, UBAC1, ATP1A1, CTRB2, GPC2, MAP3K14, PHACTR3, SLC15A3, UBAC2, ATP1A2, CTRC, GPC3, MAP3K15, PHACTR4, SLC15A4, UBALD1, ATP1A3, CTRL, GPC4, MAP3K19, PHAX, SLC15A5, UBALD2, ATP1A4, CTSA, GPC5, MAP3K2, PHB, SLC16A1, UBAP1, ATP1B1, CTSB, GPC6, MAP3K3, PHB2, SLC16A10, UBAP1L, ATP1B2, CTSC, GPCPD1, MAP3K4, PHC1, SLC16A11, UBAP2, ATP1B3, CTSD, GPD1, MAP3K5, PHC2, SLC16A12, UBAP2L, ATP1B4, CTSE, GPD1L, MAP3K6, PHC3, SLC16A13, UBASH3A, ATP2A1, CTSF, GPD2, MAP3K7, PHEX, SLC16A14, UBASH3B, ATP2A2, CTSG, GPER1, MAP3K7CL, PHF1, SLC16A2, UBB, ATP2A3, CTSH, GPHA2, MAP3K8, PHF10, SLC16A3, UBC, ATP2B1, CTSK, GPHB5, MAP3K9, PHF11, SLC16A4, UBD, ATP2B2, CTSL, GPHN, MAP4, PHF12, SLC16A5, UBE2A, ATP2B3, CTSO, GPI, MAP4K1, PHF13, SLC16A6, UBE2B, ATP2B4, CTSS, GPIHBP1, MAP4K2, PHF14, SLC16A7, UBE2C, ATP2C1, CTSV, GPKOW, MAP4K3, PHF19, SLC16A8, UBE2D1, ATP2C2, CTSW, GPLD1, MAP4K4, PHF2, SLC16A9, UBE2D2, ATP4A, CTSZ, GPM6A, MAP4K5, PHF20, SLC17A1, UBE2D3, ATP4B, CTTN, GPM6B, MAP6, PHF20L1, SLC17A2, UBE2D4, ATP5A1, CTTNBP2, GPN1, MAP6D1, PHF21A, SLC17A3, UBE2E1, ATP5B, CTTNBP2NL, GPN2, MAP7, PHF21B, SLC17A4, UBE2E2, ATP5C1, CTU1, GPN3, MAP7D1, PHF23, SLC17A5, UBE2E2-AS1, ATP5D, CTU2, GPNMB, MAP7D2, PHF3, SLC17A6, UBE2E3, ATP5E, CTXN1, GPR1, MAP7D3, PHF5A, SLC17A7, UBE2F, ATP5F1, CTXN2, GPR101, MAP9, PHF6, SLC17A8, UBE2G1, ATP5G1, CTXN3, GPR107, MAPK1, PHF7, SLC17A9, UBE2G2, ATP5G2, CUBN, GPR108, MAPK10, PHF8, SLC18A1, UBE2H, ATP5G3, CUEDC1, GPR110, MAPK11, PHGDH, SLC18A2, UBE2I, ATP5H, CUEDC2, GPR111, MAPK12, PHGR1, SLC18A3, UBE2J1, ATP5I, CUL1, GPR112, MAPK13, PHIP, SLC18B1, UBE2J2, ATP5J, CUL2, GPR113, MAPK14, PHKA1, SLC19A1, UBE2K, ATP5J2, CUL3, GPR114, MAPK15, PHKA2, SLC19A2, UBE2L3, ATP5J2-PTCD1, CUL4A, GPR115, MAPK1IP1L, PHKB, SLC19A3, UBE2L6, ATP5L, CUL4B, GPR116, MAPK3, PHKG1, SLC1A1, UBE2M, ATP5L2, CUL5, GPR119, MAPK4, PHKG2, SLC1A2, UBE2N, ATP5O, CUL7, GPR12, MAPK6, PHLDA1, SLC1A3, UBE2NL, ATP5S, CUL9, GPR123, MAPK7, PHLDA2, SLC1A4, UBE2O, ATP5SL, CUTA, GPR124, MAPK8, PHLDA3, SLC1A5, UBE2Q1,

ATP6AP1, CUTC, GPR125, MAPK8IP1, PHLDB1, SLC1A6, UBE2Q2, ATP6AP1L, CUX1, GPR126, MAPK8IP2, PHLDB2, SLC1A7, UBE2QL1, ATP6AP2, CUX2, GPR128, MAPK8IP3, PHLDB3, SLC20A1, UBE2R2, ATP6V0A1, CUZD1, GPR132, MAPK9, PHLPP1, SLC20A2, UBE2S, ATP6V0A2, CWC15, GPR133, MAPKAP1, PHLPP2, SLC22A1, UBE2T, ATP6V0A4, CWC22, GPR135, MAPKAPK2, PHOSPHO1, SLC22A10, UBE2U, ATP6V0B, CWC25, GPR137, MAPKAPK3, PHOSPHO2, SLC22A11, UBE2V1, ATP6V0C, CWC27, GPR137B, MAPKAPK5, PHOSPHO2-KLHL23, SLC22A12, UBE2V2, ATP6V0D1, CWF19L1, GPR137C, MAPKBP1, PHOX2A, SLC22A13, UBE2W, ATP6V0D2, CWF19L2, GPR139, MAPRE1, PHOX2B, SLC22A14, UBE2Z, ATP6V0E1, CWH43, GPR142, MAPRE2, PHPT1, SLC22A15, UBE3A, ATP6V0E2, CX3CL1, GPR143, MAPRE3, PHRF1, SLC22A16, UBE3B, ATP6V1A, CX3CR1, GPR148, MAPT, PHTF1, SLC22A17, UBE3C, ATP6V1B1, CXADR, GPR149, 1-Mar, PHTF2, SLC22A18, UBE3D, ATP6V1B2, CXCL1, GPR15, 10-Mar, PHYH, SLC22A18AS, UBE4A, ATP6V1C1, CXCL10, GPR150, 11- Mar, PHYHD1, SLC22A2, UBE4B, ATP6V1C2, CXCL11, GPR151, 2-Mar, PHYHIP, SLC22A20, UBFD1, ATP6V1D, CXCL12, GPR152, 3-Mar, PHYHIPL, SLC22A23, UBIAD1, ATP6V1E1, CXCL13, GPR153, 4- Mar, PHYKPL, SLC22A24, UBL3, ATP6V1E2, CXCL14, GPR155, 5-Mar, PI15, SLC22A25, UBL4A, ATP6V1F, CXCL16, GPR156, 6-Mar, PI16, SLC22A3, UBL4B, ATP6V1G1, CXCL17, GPR157, 7-Mar, PI3, SLC22A31, UBL5, ATP6V1G2, CXCL2, GPR158, 8-Mar, PI4K2A, SLC22A4, UBL7, ATP6V1G3, CXCL3, GPR160, 9-Mar, PI4K2B, SLC22A5, UBLCP1, ATP6V1H, CXCL5, GPR161, MARCKS, PI4KA, SLC22A6, UBN1, ATP7A, CXCL6, GPR162, MARCKSL1, PI4KB, SLC22A7, UBN2, ATP7B, CXCL9, GPR17, MARCO, PIANP, SLC22A8, UBOX5, ATP8A1, CXCR1, GPR171, MARK1, PIAS1, SLC22A9, UBP1, ATP8A2, CXCR2, GPR173, MARK2, PIAS2, SLC23A1, UBQLN1, ATP8B1, CXCR3, GPR174, MARK3, PIAS3, SLC23A2, UBQLN2, ATP8B2, CXCR4, GPR176, MARK4, PIAS4, SLC23A3, UBQLN3, ATP8B3, CXCR5, GPR179, MARS, PIBF1, SLC24A1, UBQLN4, ATP8B4, CXCR6, GPR18, MARS2, PICALM, SLC24A2, UBQLNL, ATP9A, CXorf21, GPR180, MARVELD1, PICK1, SLC24A3, UBR1, ATP9B, CXorf22, GPR182, MARVELD2, PID1, SLC24A4, UBR2, ATPAF1, CXorf23, GPR183, MARVELD3, PIDD, SLC24A5, UBR3, ATPAF2, CXorf27, GPR19, MAS1, PIEZO1, SLC25A1, UBR4, ATPIF1, CXorf30, GPR20, MAS1L, PIEZO2, SLC25A10, UBR5, ATR, CXorf36, GPR21, MASP1, PIF1, SLC25A11, UBR7, ATRAID, CXorf38, GPR22, MASP2, PIFO, SLC25A12, UBTD1, ATRIP, CXorf40A, GPR25, MAST1, PIGA, SLC25A13, UBTD2, ATRN, CXorf40B, GPR26, MAST2, PIGB, SLC25A14, UBTF, ATRNL1, CXorf48, GPR27, MAST3, PIGC, SLC25A15, UBTFL1, ATRX, CXorf49, GPR3, MAST4, PIGF, SLC25A16, UBXN1, ATXN1, CXorf49B, GPR31, MASTL, PIGG, SLC25A17, UBXN10, ATXN10, CXorf51A, GPR32, MAT1A, PIGH, SLC25A18, UBXN11, ATXN1L, CXorf51B, GPR33, MAT2A, PIGK, SLC25A19, UBXN2A, ATXN2, CXorf56, GPR34, MAT2B, PIGL, SLC25A2, UBXN2B, ATXN2L, CXorf57, GPR35, MATK, PIGM, SLC25A20, UBXN4, ATXN3, CXorf58, GPR37, MATN1, PIGN, SLC25A21, UBXN6, ATXN3L, CXorf61, GPR37L1, MATN2, PIGO, SLC25A22, UBXN7, ATXN7, CXorf64, GPR39, MATN3, PIGP, SLC25A23, UBXN8, ATXN7L1, CXorf65, GPR4, MATN4,

PIGQ, SLC25A24, UCHL1, ATXN7L2, CXorf66, GPR45, MATR3, PIGR, SLC25A25, UCHL3, ATXN7L3, CXXC1, GPR50, MAU2, PIGS, SLC25A26, UCHL5, ATXN7L3B, CXXC11, GPR52, MAVS, PIGT, SLC25A27, UCK1, AUH, CXXC4, GPR55, MAX, PIGU, SLC25A28, UCK2, AUNIP, CXXC5, GPR56, MAZ, PIGV, SLC25A29, UCKL1, AUP1, CYB561, GPR6, MB, PIGW, SLC25A3, UCMA, AURKA, CYB561A3, GPR61, MB21D1, PIGX, SLC25A30, UCN, AURKAIP1, CYB561D1, GPR62, MB21D2, PIGY, SLC25A31, UCN2, AURKB, CYB561D2, GPR63, MBD1, PIGZ, SLC25A32, UCN3, AURKC, CYB5A, GPR64, MBD2, PIH1D1, SLC25A33, UCP1, AUTS2, CYB5B, GPR65, MBD3, PIH1D2, SLC25A34, UCP2, AVEN, CYB5D1, GPR68, MBD3L1, PIH1D3, SLC25A35, UCP3, AVIL, CYB5D2, GPR75, MBD4, PIK3AP1, SLC25A36, UEVLD, AVL9, CYB5R1, GPR75-ASB3, MBD5, PIK3C2A, SLC25A37, UFC1, AVP, CYB5R2, GPR78, MBD6, PIK3C2B, SLC25A38, UFD1L, AVPI1, CYB5R3, GPR82, MBIP, PIK3C2G, SLC25A39, UFL1, AVPR1A, CYB5R4, GPR83, MBL2, PIK3C3, SLC25A4, UFM1, AVPR1B, CYB5RL, GPR84, MBLAC1, PIK3CA, SLC25A40, UFSP1, AVPR2, CYBA, GPR85, MBLAC2, PIK3CB, SLC25A41, UFSP2, AWAT1, CYBB, GPR87, MBNL1, PIK3CD, SLC25A42, UGCG, AWAT2, CYBRD1, GPR88, MBNL2, PIK3CG, SLC25A43, UGDH, AXDND1, CYC1, GPR89A, MBNL3, PIK3IP1, SLC25A44, UGGT1, AXIN1, CYCS, GPR89B, MBOAT1, PIK3R1, SLC25A45, UGGT2, AXIN2, CYFIP1, GPR89C, MBOAT2, PIK3R2, SLC25A46, UGP2, AXL, CYFIP2, GPR97, MBOAT4, PIK3R3, SLC25A47, UGT1A1, AZGP1, CYGB, GPR98, MBOAT7, PIK3R4, SLC25A48, UGT1A10, AZI1, CYHR1, GPRASP1, MBP, PIK3R5, SLC25A5, UGT1A3, AZI2, CYLC1, GPRASP2, MBTD1, PIK3R6, SLC25A51, UGT1A4, AZIN1, CYLC2, GPRC5A, MBTPS1, PIKFYVE, SLC25A52, UGT1A5, AZU1, CYLD, GPRC5B, MBTPS2, PILRA, SLC25A53, UGT1A6, B2M, CYorf17, GPRC5C, MC1R, PILRB, SLC25A6, UGT1A7, B3GALNT1, CYP11A1, GPRC5D, MC2R, PIM1, SLC26A1, UGT1A8, B3GALNT2, CYP11B1, GPRC6A, MC3R, PIM2, SLC26A10, UGT1A9, B3GALT1, CYP11B2, GPRIN1, MC4R, PIM3, SLC26A11, UGT2A1, B3GALT2, CYP17A1, GPRIN2, MC5R, PIN1, SLC26A2, UGT2A2, B3GALT4, CYP19A1, GPRIN3, MCAM, PIN4, SLC26A3, UGT2A3, B3GALT5, CYP1A1, GPS1, MCAT, PINK1, SLC26A4, UGT2B10, B3GALT6, CYP1A2, GPS2, MCC, PINLYP, SLC26A5, UGT2B11, B3GALTL, CYP1B1, GPSM1, MCCC1, PINX1, SLC26A6, UGT2B15, B3GAT1, CYP20A1, GPSM2, MCCC2, PIP, SLC26A7, UGT2B17, B3GAT2, CYP21A2, GPSM3, MCCD1, PIP4K2A, SLC26A8, UGT2B28, B3GAT3, CYP24A1, GPT, MCEE, PIP4K2B, SLC26A9, UGT2B4, B3GNT1, CYP26A1, GPT2, MCF2, PIP4K2C, SLC27A1, UGT2B7, B3GNT2, CYP26B1, GPX1, MCF2L, PIP5K1A, SLC27A2, UGT3A1, B3GNT3, CYP26C1, GPX2, MCF2L2, PIP5K1B, SLC27A3, UGT3A2, B3GNT4, CYP27A1, GPX3, MCFD2, PIP5K1C, SLC27A4, UGT8, B3GNT5, CYP27B1, GPX4, MCHR1, PIP5KL1, SLC27A5, UHMK1, B3GNT6, CYP27C1, GPX5, MCHR2, PIPOX, SLC27A6, UHRF1, B3GNT7, CYP2A13, GPX6, MCIDAS, PIR, SLC28A1, UHRF1BP1, B3GNT8, CYP2A6, GPX7, MCL1, PIRT, SLC28A2, UHRF1BP1L, B3GNT9, CYP2A7, GPX8, MCM10, PISD, SLC28A3, UHRF2, B3GNTL1, CYP2B6, GRAMD1A, MCM2, PITHD1, SLC29A1, UIMC1, B4GALNT1, CYP2C18, GRAMD1B, MCM3, PITPNA, SLC29A2, ULBP1, B4GALNT2, CYP2C19,

GRAMD1C, MCM3AP, PITPNB, SLC29A3, ULBP2, B4GALNT3, CYP2C8, GRAMD2, MCM4, PITPNC1, SLC29A4, ULBP3, B4GALNT4, CYP2C9, GRAMD3, MCM5, PITPNM1, SLC2A1, ULK1, B4GALT1, CYP2D6, GRAMD4, MCM6, PITPNM2, SLC2A10, ULK2, B4GALT2, CYP2E1, GRAP, MCM7, PITPNM3, SLC2A11, ULK3, B4GALT3, CYP2F1, GRAP2, MCM8, PITRM1, SLC2A12, ULK4, B4GALT4, CYP2J2, GRAPL, MCM9, PITX1, SLC2A13, UMOD, B4GALT5, CYP2R1, GRASP, MCMBP, PITX2, SLC2A14, UMODL1, B4GALT6, CYP2S1, GRB10, MCMDC2, PITX3, SLC2A2, UMPS, B4GALT7, CYP2U1, GRB14, MCOLN1, PIWIL1, SLC2A3, UNC119, B9D1, CYP2W1, GRB2, MCOLN2, PIWIL2, SLC2A4, UNC119B, B9D2, CYP39A1, GRB7, MCOLN3, PIWIL3, SLC2A4RG, UNC13A, BAALC, CYP3A4, GREB1, MCPH1, PIWIL4, SLC2A5, UNC13B, BAAT, CYP3A43, GREM1, MCRS1, PJA1, SLC2A6, UNC13C, BABAM1, CYP3A5, GREM2, MCTP1, PJA2, SLC2A7, UNC13D, BACE1, CYP3A7, GRHL1, MCTP2, PKD1, SLC2A8, UNC45A, BACE2, CYP3A7- CYP3AP1, GRHL2, MCTS1, PKD1L1, SLC2A9, UNC45B, BACH1, CYP46A1, GRHL3, MCU, PKD1L2, SLC30A1, UNC50, BACH2, CYP4A11, GRHPR, MCUR1, PKD1L3, SLC30A10, UNC5A, BAD, CYP4A22, GRIA1, MDC1, PKD2, SLC30A2, UNC5B, BAG1, CYP4B1, GRIA2, MDFI, PKD2L1, SLC30A3, UNC5C, BAG2, CYP4F11, GRIA3, MDFIC, PKD2L2, SLC30A4, UNC5CL, BAG3, CYP4F12, GRIA4, MDGA1, PKDCC, SLC30A5, UNC5D, BAG4, CYP4F2, GRID1, MDGA2, PKDREJ, SLC30A6, UNC79, BAG5, CYP4F22, GRID2, MDH1, PKHD1, SLC30A7, UNC80, BAG6, CYP4F3, GRID2IP, MDH1B, PKHD1L1, SLC30A8, UNC93A, BAGE, CYP4F8, GRIFIN, MDH2, PKIA, SLC30A9, UNC93B1, BAGE2, CYP4V2, GRIK1, MDK, PKIB, SLC31A1, UNCX, BAGE3, CYP4X1, GRIK2, MDM1, PKIG, SLC31A2, UNG, BAHCC1, CYP4Z1, GRIK3, MDM2, PKLR, SLC32A1, UNK, BAHD1, CYP51A1, GRIK4, MDM4, PKM, SLC33A1, UNKL, BAI1, CYP7A1, GRIK5, MDN1, PKMYT1, SLC34A1, UPB1, BAI2, CYP7B1, GRIN1, MDP1, PKN1, SLC34A2, UPF1, BAI3, CYP8B1, GRIN2A, ME1, PKN2, SLC34A3, UPF2, BAIAP2, CYR61, GRIN2B, ME2, PKN3, SLC35A1, UPF3A, BAIAP2L1, CYS1, GRIN2C, ME3, PKNOX1, SLC35A2, UPF3B, BAIAP2L2, CYSLTR1, GRIN2D, MEA1, PKNOX2, SLC35A3, UPK1A, BAIAP3, CYSLTR2, GRIN3A, MEAF6, PKP1, SLC35A4, UPK1B, BAK1, CYSTM1, GRIN3B, MECOM, PKP2, SLC35A5, UPK2, BAMBI, CYTH1, GRINA, MECP2, PKP3, SLC35B1, UPK3A, BANF1, CYTH2, GRIP1, MECR, PKP4, SLC35B2, UPK3B, BANF2, CYTH3, GRIP2, MED1, PLA1A, SLC35B3, UPK3BL, BANK1, CYTH4, GRIPAP1, MED10, PLA2G10, SLC35B4, UPP1, BANP, CYTIP, GRK1, MED11, PLA2G12A, SLC35C1, UPP2, BAP1, CYTL1, GRK4, MED12, PLA2G12B, SLC35C2, UPRT, BARD1, CYYR1, GRK5, MED12L, PLA2G15, SLC35D1, UQCC1, BARHL1, D2HGDH, GRK6, MED13, PLA2G16, SLC35D2, UQCC2, BARHL2, DAAM1, GRK7, MED13L, PLA2G1B, SLC35D3, UQCR10, BARX1, DAAM2, GRM1, MED14, PLA2G2A, SLC35E1, UQCR11, BARX2, DAB1, GRM2, MED15, PLA2G2C, SLC35E2, UQCRB, BASP1, DAB2, GRM3, MED16, PLA2G2D, SLC35E2B, UQCRC1, BATF, DAB2IP, GRM4, MED17, PLA2G2E, SLC35E3, UQCRC2, BATF2, DACH1, GRM5, MED18, PLA2G2F, SLC35E4, UQCRFS1, BATF3, DACH2, GRM6, MED19, PLA2G3, SLC35F1, UQCRH, BAX, DACT1, GRM7, MED20, PLA2G4A,

SLC35F2, UQCRHL, BAZ1A, DACT2, GRM8, MED21, PLA2G4B, SLC35F3, UQCRQ, BAZ1B, DACT3, GRN, MED22, PLA2G4C, SLC35F4, URAD, BAZ2A, DAD1, GRP, MED23, PLA2G4D, SLC35F5, URB1, BAZ2B, DAG1, GRPEL1, MED24, PLA2G4E, SLC35F6, URB2, BBC3, DAGLA, GRPEL2, MED25, PLA2G4F, SLC35G1, URGCP, BBIP1, DAGLB, GRPR, MED26, PLA2G5, SLC35G2, URGCP-MRPS24, BBOX1, DAK, GRSF1, MED27, PLA2G6, SLC35G3, URI1, BBS1, DALRD3, GRTP1, MED28, PLA2G7, SLC35G4, URM1, BBS10, DAND5, GRWD1, MED29, PLA2R1, SLC35G5, UROC1, BBS12, DAO, GRXCR1, MED30, PLAA, SLC35G6, UROD, BBS2, DAOA, GRXCR2, MED31, PLAC1, SLC36A1, UROS, BBS4, DAP, GSAP, MED4, PLAC4, SLC36A2, USB1, BBS5, DAP3, GSC, MED6, PLAC8, SLC36A3, USE1, BBS7, DAPK1, GSC2, MED7, PLAC8L1, SLC36A4, USF1, BBS9, DAPK2, GSDMA, MED8, PLAC9, SLC37A1, USF2, BBX, DAPK3, GSDMB, MED9, PLAG1, SLC37A2, USH1C, BCAM, DAPL1, GSDMC, MEDAG, PLAGL1, SLC37A3, USH1G, BCAN, DAPP1, GSDMD, MEF2A, PLAGL2, SLC37A4, USH2A, BCAP29, DARS, GSE1, MEF2B, PLAT, SLC38A1, USHBP1, BCAP31, DARS2, GSG1, MEF2BNB, PLAU, SLC38A10, USMG5, BCAR1, DAW1, GSG1L, MEF2BNB-MEF2B, PLAUR, SLC38A11, USO1, BCAR3, DAXX, GSG2, MEF2C, PLB1, SLC38A2, USP1, BCAS1, DAZ1, GSK3A, MEF2D, PLBD1, SLC38A3, USP10, BCAS2, DAZ2, GSK3B, MEFV, PLBD2, SLC38A4, USP11, BCAS3, DAZ3, GSKIP, MEGF10, PLCB1, SLC38A5, USP12, BCAS4, DAZ4, GSN, MEGF11, PLCB2, SLC38A6, USP13, BCAT1, DAZAP1, GSPT1, MEGF6, PLCB3, SLC38A7, USP14, BCAT2, DAZAP2, GSPT2, MEGF8, PLCB4, SLC38A8, USP15, BCCIP, DAZL, GSR, MEGF9, PLCD1, SLC38A9, USP16, BCDIN3D, DBF4, GSS, MEI1, PLCD3, SLC39A1, USP17L24, BCHE, DBF4B, GSTA1, MEI4, PLCD4, SLC39A10, USP17L25, BCKDHA, DBH, GSTA2, MEIG1, PLCE1, SLC39A11, USP17L26, BCKDHB, DBI, GSTA3, MEIOB, PLCG1, SLC39A12, USP17L28, BCKDK, DBN1, GSTA4, MEIS1, PLCG2, SLC39A13, USP17L29, BCL10, DBNDD1, GSTA5, MEIS2, PLCH1, SLC39A14, USP18, BCL11A, DBNDD2, GSTCD, MEIS3, PLCH2, SLC39A2, USP19, BCL11B, DBNL, GSTK1, MELK, PLCL1, SLC39A3, USP2, BCL2, DBP, GSTM1, MEMO1, PLCL2, SLC39A4, USP20, BCL2A1, DBR1, GSTM2, MEN1, PLCXD1, SLC39A5, USP21, BCL2L1, DBT, GSTM3, MEOX1, PLCXD2, SLC39A6, USP22, BCL2L10, DBX1, GSTM4, MEOX2, PLCXD3, SLC39A7, USP24, BCL2L11, DBX2, GSTM5, MEP1A, PLCZ1, SLC39A8, USP25, BCL2L12, DCAF10, GSTO1, MEP1B, PLD1, SLC39A9, USP26, BCL2L13, DCAF11, GSTO2, MEPCE, PLD2, SLC3A1, USP27X, BCL2L14, DCAF12, GSTP1, MEPE, PLD3, SLC3A2, USP28, BCL2L15, DCAF12L1, GSTT1, MERTK, PLD4, SLC40A1, USP29, BCL2L2, DCAF12L2, GSTT2, MESDC1, PLD5, SLC41A1, USP3, BCL2L2-PABPN1, DCAF13, GSTT2B, MESDC2, PLD6, SLC41A2, USP30, BCL3, DCAF15, GSTZ1, MESP1, PLEC, SLC41A3, USP31, BCL6, DCAF16, GSX1, MESP2, PLEK, SLC43A1, USP32, BCL6B, DCAF17, GSX2, MEST, PLEK2, SLC43A2, USP33, BCL7A, DCAF4, GTDC1, MET, PLEKHA1, SLC43A3, USP34, BCL7B, DCAF4L1, GTF2A1, METAP1, PLEKHA2, SLC44A1, USP35, BCL7C, DCAF4L2, GTF2A1L, METAP1D, PLEKHA3, SLC44A2, USP36, BCL9, DCAF5, GTF2A2, METAP2, PLEKHA4, SLC44A3, USP37, BCL9L, DCAF6, GTF2B, METRN, PLEKHA5, SLC44A4, USP38, BCLAF1,

DCAF7, GTF2E1, METRNL, PLEKHA6, SLC44A5, USP39, BCMO1, DCAF8, GTF2E2, METTL1, PLEKHA7, SLC45A1, USP4, BCO2, DCAF8L1, GTF2F1, METTL10, PLEKHA8, SLC45A2, USP40, BCOR, DCAKD, GTF2F2, METTL11B, PLEKHB1, SLC45A3, USP41, BCORL1, DCBLD1, GTF2H1, METTL12, PLEKHB2, SLC45A4, USP42, BCR, DCBLD2, GTF2H2, METTL13, PLEKHD1, SLC46A1, USP43, BCS1L, DCC, GTF2H2C, METTL14, PLEKHF1, SLC46A2, USP44, BDH1, DCD, GTF2H3, METTL15, PLEKHF2, SLC46A3, USP45, BDH2, DCDC1, GTF2H4, METTL16, PLEKHG1, SLC47A1, USP46, BDKRB1, DCDC2, GTF2H5, METTL17, PLEKHG2, SLC47A2, USP47, BDKRB2, DCDC2B, GTF2I, METTL18, PLEKHG3, SLC48A1, USP48, BDNF, DCDC2C, GTF2IRD1, METTL20, PLEKHG4, SLC4A1, USP49, BDP1, DCDC5, GTF2IRD2, METTL21A, PLEKHG4B, SLC4A10, USP5, BEAN1, DCHS1, GTF2IRD2B, METTL21B, PLEKHG5, SLC4A11, USP50, BECN1, DCHS2, GTF3A, METTL21C, PLEKHG6, SLC4A1AP, USP51, BECN1P1, DCK, GTF3C1, METTL22, PLEKHG7, SLC4A2, USP53, BEGAIN, DCLK1, GTF3C2, METTL23, PLEKHH1, SLC4A3, USP54, BEND2, DCLK2, GTF3C3, METTL24, PLEKHH2, SLC4A4, USP6, BEND3, DCLK3, GTF3C4, METTL25, PLEKHH3, SLC4A5, USP6NL, BEND4, DCLRE1A, GTF3C5, METTL2A, PLEKHJ1, SLC4A7, USP7, BEND5, DCLRE1B, GTF3C6, METTL2B, PLEKHM1, SLC4A8, USP8, BEND6, DCLRE1C, GTPBP1, METTL3, PLEKHM2, SLC4A9, USP9X, BEND7, DCN, GTPBP10, METTL4, PLEKHM3, SLC50A1, USP9Y, BEST1, DCP1A, GTPBP2, METTL5, PLEKHN1, SLC51A, USPL1, BEST2, DCP1B, GTPBP3, METTL6, PLEKHO1, SLC51B, UST, BEST3, DCP2, GTPBP4, METTL7A, PLEKHO2, SLC52A1, UTF1, BEST4, DCPS, GTPBP6, METTL7B, PLEKHS1, SLC52A2, UTP11L, BET1, DCST1, GTPBP8, METTL8, PLET1, SLC52A3, UTP14A, BET1L, DCST2, GTSCR1, METTL9, PLG, SLC5A1, UTP14C, BEX1, DCSTAMP, GTSE1, MEX3A, PLGLB1, SLC5A10, UTP15, BEX2, DCT, GTSF1, MEX3B, PLGLB2, SLC5A11, UTP18, BEX4, DCTD, GTSF1L, MEX3C, PLGRKT, SLC5A12, UTP20, BEX5, DCTN1, GUCA1A, MEX3D, PLIN1, SLC5A2, UTP23, BFAR, DCTN2, GUCA1B, MFAP1, PLIN2, SLC5A3, UTP3, BFSP1, DCTN3, GUCA1C, MFAP2, PLIN3, SLC5A4, UTP6, BFSP2, DCTN4, GUCA2A, MFAP3, PLIN4, SLC5A5, UTRN, BGLAP, DCTN5, GUCA2B, MFAP3L, PLIN5, SLC5A6, UTS2, BGN, DCTN6, GUCD1, MFAP4, PLK1, SLC5A7, UTS2B, BHLHA15, DCTPP1, GUCY1A2, MFAP5, PLK1S1, SLC5A8, UTY, BHLHA9, DCUN1D1, GUCY1A3, MFF, PLK2, SLC5A9, UVRAG, BHLHB9, DCUN1D2, GUCY1B3, MFGE8, PLK3, SLC6A1, UVSSA, BHLHE22, DCUN1D3, GUCY2C, MFHAS1, PLK4, SLC6A11, UXS1, BHLHE23, DCUN1D4, GUCY2D, MFI2, PLK5, SLC6A12, UXT, BHLHE40, DCUN1D5, GUCY2F, MFN1, PLLP, SLC6A13, VAC14, BHLHE41, DCX, GUF1, MFN2, PLN, SLC6A14, VAMP1, BHMT, DCXR, GUK1, MFNG, PLOD1, SLC6A15, VAMP2, BHMT2, DDA1, GULP1, MFRP, PLOD2, SLC6A16, VAMP3, BICC1, DDAH1, GUSB, MFSD1, PLOD3, SLC6A17, VAMP4, BICD1, DDAH2, GXYLT1, MFSD10, PLP1, SLC6A18, VAMP5, BICD2, DDB1, GXYLT2, MFSD11, PLP2, SLC6A19, VAMP7, BID, DDB2, GYG1, MFSD12, PLRG1, SLC6A2, VAMP8, BIK, DDC, GYG2, MFSD2A, PLS1, SLC6A20, VANGL1, BIN1, DDHD1, GYLTL1B, MFSD2B, PLS3, SLC6A3, VANGL2, BIN2, DDHD2, GYPA, MFSD3, PLSCR1, SLC6A4, VAPA,

BIN3, DDI1, GYPB, MFSD4, PLSCR2, SLC6A5, VAPB, BIRC2, DDI2, GYPC, MFSD5, PLSCR3, SLC6A6, VARS, BIRC3, DDIT3, GYPE, MFSD6, PLSCR4, SLC6A7, VARS2, BIRC5, DDIT4, GYS1, MFSD6L, PLSCR5, SLC6A8, VASH1, BIRC6, DDIT4L, GYS2, MFSD7, PLTP, SLC6A9, VASH2, BIRC7, DDN, GZF1, MFSD8, PLVAP, SLC7A1, VASN, BIRC8, DDO, GZMA, MFSD9, PLXDC1, SLC7A10, VASP, BIVM, DDOST, GZMB, MGA, PLXDC2, SLC7A11, VAT1, BIVM-ERCC5, DDR1, GZMH, MGAM, PLXNA1, SLC7A13, VAT1L, BLCAP, DDR2, GZMK, MGARP, PLXNA2, SLC7A14, VAV1, BLID, DDRGK1, GZMM, MGAT1, PLXNA3, SLC7A2, VAV2, BLK, DDT, H1F0, MGAT2, PLXNA4, SLC7A3, VAV3, BLM, DDTL, H1FNT, MGAT3, PLXNB1, SLC7A4, VAX1, BLMH, DDX1, H1FOO, MGAT4A, PLXNB2, SLC7A5, VAX2, BLNK, DDX10, H1FX, MGAT4B, PLXNB3, SLC7A6, VBP1, BLOC1S1, DDX11, H2AFB1, MGAT4C, PLXNC1, SLC7A6OS, VCAM1, BLOC1S2, DDX17, H2AFB2, MGAT5, PLXND1, SLC7A7, VCAN, BLOC1S3, DDX18, H2AFB3, MGAT5B, PM20D1, SLC7A8, VCL, BLOC1S4, DDX19A, H2AFJ, MGEA5, PM20D2, SLC7A9, VCP, BLOC1S5, DDX19B, H2AFV, MGLL, PMAIP1, SLC8A1, VCPIP1, BLOC1S6, DDX20, H2AFX, MGME1, PMCH, SLC8A2, VCPKMT, BLVRA, DDX21, H2AFY, MGMT, PMEL, SLC8A3, VCX, BLVRB, DDX23, H2AFY2, MGP, PMEPA1, SLC8B1, VCX2, BLZF1, DDX24, H2AFZ, MGRN1, PMF1, SLC9A1, VCX3A, BMF, DDX25, H2BFM, MGST1, PMF1-BGLAP, SLC9A2, VCX3B, BMI1, DDX26B, H2BFWT, MGST2, PMFBP1, SLC9A3, VCY, BMP1, DDX27, H3F3A, MGST3, PML, SLC9A3R1, VCY1B, BMP10, DDX28, H3F3B, MIA, PMM1, SLC9A3R2, VDAC1, BMP15, DDX31, H3F3C, MIA2, PMM2, SLC9A4, VDAC2, BMP2, DDX39A, H6PD, MIA3, PMP2, SLC9A5, VDAC3, BMP2K, DDX39B, HAAO, MIB1, PMP22, SLC9A6, VDR, BMP3, DDX3X, HABP2, MIB2, PMPCA, SLC9A7, VEGFA, BMP4, DDX3Y, HABP4, MICA, PMPCB, SLC9A8, VEGFB, BMP5, DDX4, HACE1, MICAL1, PMS1, SLC9A9, VEGFC, BMP6, DDX41, HACL1, MICAL2, PMS2, SLC9B1, VENTX, BMP7, DDX42, HADH, MICAL3, PMVK, SLC9B2, VEPH1, BMP8A, DDX43, HADHA, MICALCL, PNCK, SLC9C1, VEZF1, BMP8B, DDX46, HADHB, MICALL1, PNISR, SLC9C2, VEZT, BMPER, DDX47, HAGH, MICALL2, PNKD, SLCO1A2, VGF, BMPR1A, DDX49, HAGHL, MICB, PNKP, SLCO1B1, VGLL1, BMPR1B, DDX5, HAL, MICU1, PNLDC1, SLCO1B3, VGLL2, BMPR2, DDX50, HAMP, MICU2, PNLIP, SLCO1B7, VGLL3, BMS1, DDX51, HAND1, MICU3, PNLIPRP1, SLCO1C1, VGLL4, BMX, DDX52, HAND2, MID1, PNLIPRP2, SLCO2A1, VHL, BNC1, DDX53, HAO1, MID1IP1, PNLIPRP3, SLCO2B1, VHLL, BNC2, DDX54, HAO2, MID2, PNMA2, SLCO3A1, VIL1, BNIP1, DDX55, HAP1, MIDN, PNMA3, SLCO4A1, VILL, BNIP2, DDX56, HAPLN1, MIEF1, PNMA5, SLCO4C1, VIM, BNIP3, DDX58, HAPLN2, MIEF2, PNMA6A, SLCO5A1, VIMP, BNIP3L, DDX59, HAPLN3, MIEN1, PNMA6C, SLCO6A1, VIP, BNIPL, DDX6, HAPLN4, MIER1, PNMAL1, SLFN11, VIPAS39, BOC, DDX60, HARBI1, MIER2, PNMAL2, SLFN12, VIPR1, BOD1, DDX60L, HARS, MIER3, PNMT, SLFN12L, VIPR2, BOD1L1, DEAF1, HARS2, MIF, PNN, SLFN13, VIT, BOD1L2, 1-Dec, HAS1, MIF4GD, PNO1, SLFN14, VKORC1, BOK, DECR1, HAS2, MIIP, PNOC, SLFN5, VKORC1L1, BOLA1, DECR2, HAS3, MILR1, PNP, SLFNL1, VLDLR, BOLA2, DEDD, HAT1, MINA, PNPLA1, SLIRP, VMA21,

BOLA2B, DEDD2, HAUS1, MINK1, PNPLA2, SLIT1, VMAC, BOLA3, DEF6, HAUS2, MINOS1, PNPLA3, SLIT2, VMO1, BOLL, DEF8, HAUS3, MINOS1-NBL1, PNPLA4, SLIT3, VMP1, BOP1, DEFA1, HAUS4, MINPP1, PNPLA5, SLITRK1, VN1R1, BORA, DEFA1B, HAUS5, MIOS, PNPLA6, SLITRK2, VN1R2, BPGM, DEFA3, HAUS6, MIOX, PNPLA7, SLITRK3, VN1R4, BPHL, DEFA4, HAUS7, MIP, PNPLA8, SLITRK4, VN1R5, BPI, DEFA5, HAUS8, MIPEP, PNPO, SLITRK5, VNN1, BPIFA1, DEFA6, HAVCR1, MIPOL1, PNPT1, SLITRK6, VNN2, BPIFA2, DEFB1, HAVCR2, MIR205HG, PNRC1, SLK, VOPP1, BPIFA3, DEFB103A, HAX1, MIS12, PNRC2, SLMAP, VPRBP, BPIFB1, DEFB103B, HBA1, MIS18A, POC1A, SLMO1, VPREB1, BPIFB2, DEFB104A, HBA2, MIS18BP1, POC1B, SLMO2, VPREB3, BPIFB3, DEFB104B, HBB, MISP, POC1B-GALNT4, SLN, VPS11, BPIFB4, DEFB105A, HBD, MITD1, POC5, SLPI, VPS13A, BPIFB6, DEFB105B, HBE1, MITF, PODN, SLTM, VPS13B, BPIFC, DEFB106A, HBEGF, MIXL1, PODNL1, SLU7, VPS13C, BPNT1, DEFB106B, HBG1, MKI67, PODXL, SLURP1, VPS13D, BPTF, DEFB107A, HBG2, MKKS, PODXL2, SLX1A, VPS16, BPY2, DEFB107B, HBM, MKL1, POF1B, SLX1B, VPS18, BPY2B, DEFB108B, HBP1, MKL2, POFUT1, SLX4, VPS25, BPY2C, DEFB110, HBQ1, MKLN1, POFUT2, SLX4IP, VPS26A, BRAF, DEFB112, HBS1L, MKNK1, POGK, SMAD1, VPS26B, BRAP, DEFB113, HBZ, MKNK2, POGLUT1, SMAD2, VPS28, BRAT1, DEFB114, HCAR1, MKRN1, POGZ, SMAD3, VPS29, BRCA1, DEFB115, HCAR2, MKRN2, POLA1, SMAD4, VPS33A, BRCA2, DEFB116, HCAR3, MKRN3, POLA2, SMAD5, VPS33B, BRCC3, DEFB118, HCCS, MKS1, POLB, SMAD6, VPS35, BRD1, DEFB119, HCFC1, MKX, POLD1, SMAD7, VPS36, BRD2, DEFB121, HCFC1R1, MLANA, POLD2, SMAD9, VPS37A, BRD3, DEFB123, HCFC2, MLC1, POLD3, SMAGP, VPS37B, BRD4, DEFB124, HCK, MLEC, POLD4, SMAP1, VPS37C, BRD7, DEFB125, HCLS1, MLF1, POLDIP2, SMAP2, VPS37D, BRD8, DEFB126, HCN1, MLF2, POLDIP3, SMARCA1, VPS39, BRD9, DEFB127, HCN2, MLH1, POLE, SMARCA2, VPS41, BRDT, DEFB128, HCN3, MLH3, POLE2, SMARCA4, VPS45, BRE, DEFB129, HCN4, MLIP, POLE3, SMARCA5, VPS4A, BRF1, DEFB130, HCRT, MLKL, POLE4, SMARCAD1, VPS4B, BRF2, DEFB131, HCRTR1, MLLT1, POLG, SMARCAL1, VPS51, BRI3, DEFB132, HCRTR2, MLLT10, POLG2, SMARCB1, VPS52, BRI3BP, DEFB133, HCST, MLLT11, POLH, SMARCC1, VPS53, BRICD5, DEFB134, HDAC1, MLLT3, POLI, SMARCC2, VPS54, BRINP1, DEFB135, HDAC10, MLLT4, POLK, SMARCD1, VPS72, BRINP2, DEFB136, HDAC11, MLLT6, POLL, SMARCD2, VPS8, BRINP3, DEFB4A, HDAC2, MLN, POLM, SMARCD3, VPS9D1, BRIP1, DEFB4B, HDAC3, MLNR, POLN, SMARCE1, VRK1, BRIX1, DEGS1, HDAC4, MLPH, POLQ, SMC1A, VRK2, BRK1, DEGS2, HDAC5, MLST8, POLR1A, SMC1B, VRK3, BRMS1, DEK, HDAC6, MLX, POLR1B, SMC2, VRTN, BRMS1L, DENND1A, HDAC7, MLXIP, POLR1C, SMC3, VSIG1, BROX, DENND1B, HDAC8, MLXIPL, POLR1D, SMC4, VSIG10, BRPF1, DENND1C, HDAC9, MLYCD, POLR1E, SMC5, VSIG10L, BRPF3, DENND2A, HDC, MMAA, POLR2A, SMC6, VSIG2, BRS3, DENND2C, HDDC2, MMAB, POLR2B, SMCHD1, VSIG4, BRSK1, DENND2D, HDDC3, MMACHC, POLR2C, SMCO2, VSIG8, BRSK2, DENND3, HDGF, MMADHC, POLR2D, SMCO3, VSNL1, BRWD1, DENND4A, HDGFL1, MMD, POLR2E,

SMCO4, VSTM1, BRWD3, DENND4B, HDGFRP2, MMD2, POLR2F, SMCP, VSTM2A, BSCL2, DENND4C, HDGFRP3, MME, POLR2G, SMCR8, VSTM2B, BSDC1, DENND5A, HDHD1, MMEL1, POLR2H, SMCR9, VSTM2L, BSG, DENND5B, HDHD2, MMGT1, POLR2I, SMDT1, VSTM4, BSN, DENND6A, HDHD3, MMP1, POLR2J, SMEK1, VSTM5, BSND, DENND6B, HDLBP, MMP10, POLR2J2, SMEK2, VSX1, BSPH1, DENR, HDX, MMP11, POLR2J3, SMG1, VSX2, BSPRY, DEPDC1, HEATR1, MMP12, POLR2K, SMG5, VTA1, BST1, DEPDC1B, HEATR2, MMP13, POLR2L, SMG6, VTCN1, BST2, DEPDC4, HEATR3, MMP14, POLR2M, SMG7, VTI1A, BSX, DEPDC5, HEATR4, MMP15, POLR3A, SMG8, VTI1B, BTAF1, DEPDC7, HEATR5A, MMP16, POLR3B, SMG9, VTN, BTBD1, DEPTOR, HEATR5B, MMP17, POLR3C, SMIM1, VWA1, BTBD10, DERA, HEATR6, MMP19, POLR3D, SMIM10, VWA2, BTBD11, DERL1, HEBP1, MMP2, POLR3E, SMIM11, VWA3A, BTBD16, DERL2, HEBP2, MMP20, POLR3F, SMIM12, VWA3B, BTBD17, DERL3, HECA, MMP21, POLR3G, SMIM13, VWA5A, BTBD18, DES, HECTD1, MMP23B, POLR3GL, SMIM14, VWA5B1, BTBD19, DESI1, HECTD2, MMP24, POLR3H, SMIM15, VWA5B2, BTBD2, DESI2, HECTD3, MMP25, POLR3K, SMIM17, VWA7, BTBD3, DET1, HECTD4, MMP26, POLRMT, SMIM18, VWA8, BTBD6, DEXI, HECW1, MMP27, POM121, SMIM19, VWA9, BTBD7, DFFA, HECW2, MMP28, POM121C, SMIM2, VWC2, BTBD8, DFFB, HEG1, MMP3, POM121L12, SMIM20, VWC2L, BTBD9, DFNA5, HELB, MMP7, POM121L2, SMIM21, VWCE, BTC, DFNB31, HELLS, MMP8, POMC, SMIM22, VWDE, BTD, DFNB59, HELQ, MMP9, POMGNT1, SMIM3, VWF, BTF3, DGAT1, HELT, MMRN1, POMGNT2, SMIM4, WAC, BTF3L4, DGAT2, HELZ, MMRN2, POMK, SMIM5, WAPAL, BTG1, DGAT2L6, HELZ2, MMS19, POMP, SMIM6, WARS, BTG2, DGCR14, HEMGN, MMS22L, POMT1, SMIM7, WARS2, BTG3, DGCR2, HEMK1, MN1, POMT2, SMIM8, WAS, BTG4, DGCR6, HENMT1, MNAT1, POMZP3, SMIM9, WASF1, BTK, DGCR6L, HEPACAM, MND1, PON1, SMKR1, WASF2, BTLA, DGCR8, HEPACAM2, MNDA, PON2, SMLR1, WASF3, BTN1A1, DGKA, HEPH, MNS1, PON3, SMN1, WASH1, BTN2A1, DGKB, HEPHL1, MNT, POP1, SMN2, WASL, BTN2A2, DGKD, HEPN1, MNX1, POP4, SMNDC1, WBP1, BTN3A1, DGKE, HERC1, MOAP1, POP5, SMO, WBP11, BTN3A2, DGKG, HERC2, MOB1A, POP7, SMOC1, WBP1L, BTN3A3, DGKH, HERC3, MOB1B, POPDC2, SMOC2, WBP2, BTNL10, DGKI, HERC4, MOB2, POPDC3, SMOX, WBP2NL, BTNL2, DGKK, HERC5, MOB3A, POR, SMPD1, WBP4, BTNL3, DGKQ, HERC6, MOB3B, PORCN, SMPD2, WBP5, BTNL8, DGKZ, HERPUD1, MOB3C, POSTN, SMPD3, WBSCR16, BTNL9, DGUOK, HERPUD2, MOB4, POT1, SMPD4, WBSCR17, BTRC, DHCR24, HES1, MOBP, POTEA, SMPD5, WBSCR22, BUB1, DHCR7, HES2, MOCOS, POTEB, SMPDL3A, WBSCR27, BUB1B, DHDDS, HES3, MOCS1, POTEB2, SMPDL3B, WBSCR28, BUB3, DHDH, HES4, MOCS2, POTEC, SMPX, WDFY1, BUD13, DHFR, HES5, MOCS3, POTED, SMR3A, WDFY2, BUD31, DHFRL1, HES6, MOG, POTEE, SMR3B, WDFY3, BVES, DHH, HES7, MOGAT1, POTEF, SMS, WDFY4, BYSL, DHODH, HESX1, MOGAT2, POTEG, SMTN, WDHD1, BZRAP1, DHPS, HEXA, MOGAT3, POTEH, SMTNL1, WDPCP, BZW1, DHRS1, HEXB, MOGS, POTEI, SMTNL2, WDR1, BZW2, DHRS11, HEXDC, MOK,

POTEJ, SMU1, WDR11, C10orf10, DHRS12, HEXIM1, MON1A, POTEM, SMUG1, WDR12, C10orf105, DHRS13, HEXIM2, MON1B, POU1F1, SMURF1, WDR13, C10orf107, DHRS2, HEY1, MON2, POU2AF1, SMURF2, WDR16, C10orf11, DHRS3, HEY2, MORC1, POU2F1, SMYD1, WDR17, C10orf111, DHRS4, HEYL, MORC2, POU2F2, SMYD2, WDR18, C10orf112, DHRS4L1, HFE, MORC3, POU2F3, SMYD3, WDR19, C10orf113, DHRS4L2, HFE2, MORC4, POU3F1, SMYD4, WDR20, C10orf118, DHRS7, HFM1, MORF4L1, POU3F2, SMYD5, WDR24, C10orf12, DHRS7B, HGC6.3, MORF4L2, POU3F3, SNAI1, WDR25, C10orf120, DHRS7C, HGD, MORN1, POU3F4, SNAI2, WDR26, C10orf126, DHRS9, HGF, MORN2, POU4F1, SNAI3, WDR27, C10orf128, DHRSX, HGFAC, MORN3, POU4F2, SNAP23, WDR3, C10orf129, DHTKD1, HGS, MORN4, POU4F3, SNAP25, WDR31, C10orf131, DHX15, HGSNAT, MORN5, POU5F1, SNAP29, WDR33, C10orf137, DHX16, HHAT, MOS, POU5F1B, SNAP47, WDR34, C10orf2, DHX29, HHATL, MOSPD1, POU5F2, SNAP91, WDR35, C10orf25, DHX30, HHEX, MOSPD2, POU6F1, SNAPC1, WDR36, C10orf32, DHX32, HHIP, MOSPD3, POU6F2, SNAPC2, WDR37, C10orf35, DHX33, HHIPL1, MOV10, PP2D1, SNAPC3, WDR38, C10orf53, DHX34, HHIPL2, MOV10L1, PPA1, SNAPC4, WDR4, C10orf54, DHX35, HHLA1, MOXD1, PPA2, SNAPC5, WDR41, C10orf55, DHX36, HHLA2, MPC1, PPAN, SNAPIN, WDR43, C10orf62, DHX37, HHLA3, MPC1L, PPAN-P2RY11, SNCA, WDR44, C10orf67, DHX38, HIAT1, MPC2, PPAP2A, SNCAIP, WDR45, C10orf68, DHX40, HIATL1, MPDU1, PPAP2B, SNCB, WDR45B, C10orf71, DHX57, HIBADH, MPDZ, PPAP2C, SNCG, WDR46, C10orf76, DHX58, HIBCH, MPEG1, PPAPDC1A, SND1, WDR47, C10orf82, DHX8, HIC1, MPG, PPAPDC1B, SNED1, WDR48, C10orf88, DHX9, HIC2, MPHOSPH10, PPAPDC2, SNF8, WDR49, C10orf90, DIABLO, HID1, MPHOSPH6, PPAPDC3, SNIP1, WDR5, C10orf91, DIAPH1, HIF1A, MPHOSPH8, PPARA, SNN, WDR52, C10orf95, DIAPH2, HIF1AN, MPHOSPH9, PPARD, SNPH, WDR53, C10orf99, DIAPH3, HIF3A, MPI, PPARG, SNRK, WDR54, C11orf1, DICER1, HIGD1A, MPL, PPARGC1A, SNRNP200, WDR55, C11orf16, DIDO1, HIGD1B, MPLKIP, PPARGC1B, SNRNP25, WDR59, C11orf21, DIEXF, HIGD1C, MPND, PPAT, SNRNP27, WDR5B, C11orf24, DIMT1, HIGD2A, MPO, PPBP, SNRNP35, WDR6, C11orf30, DIO1, HILPDA, MPP1, PPCDC, SNRNP40, WDR60, C11orf31, DIO2, HINFP, MPP2, PPCS, SNRNP48, WDR61, C11orf35, DIO3, HINT1, MPP3, PPDPF, SNRNP70, WDR62, C11orf40, DIP2A, HINT2, MPP4, PPEF1, SNRPA, WDR63, C11orf42, DIP2B, HINT3, MPP5, PPEF2, SNRPA1, WDR64, C11orf44, DIP2C, HIP1, MPP6, PPFIA1, SNRPB, WDR65, C11orf45, DIRAS1, HIP1R, MPP7, PPFIA2, SNRPB2, WDR66, C11orf48, DIRAS2, HIPK1, MPPE1, PPFIA3, SNRPC, WDR7, C11orf49, DIRAS3, HIPK2, MPPED1, PPFIA4, SNRPD1, WDR70, C11orf52, DIRC1, HIPK3, MPPED2, PPFIBP1, SNRPD2, WDR72, C11orf53, DIRC2, HIPK4, MPRIP, PPFIBP2, SNRPD3, WDR73, C11orf54, DIS3, HIRA, MPST, PPHLN1, SNRPE, WDR74, C11orf57, DIS3L, HIRIP3, MPV17, PPIA, SNRPF, WDR75, C11orf58, DIS3L2, HIST1H1A, MPV17L, PPIAL4A, SNRPG, WDR76, C11orf63, DISC1, HIST1H1B, MPV17L2, PPIAL4B, SNRPN, WDR77, C11orf65, DISP1, HIST1H1C, MPZ, PPIAL4C, SNTA1, WDR78, C11orf68, DISP2, HIST1H1D, MPZL1, PPIAL4D, SNTB1, WDR81, C11orf70, DIXDC1,

HIST1H1E, MPZL2, PPIAL4E, SNTB2, WDR82, C11orf71, DKC1, HIST1H1T, MPZL3, PPIAL4F, SNTG1, WDR83, C11orf73, DKK1, HIST1H2AA, MR1, PPIAL4G, SNTG2, WDR83OS, C11orf74, DKK2, HIST1H2AB, MRAP, PPIB, SNTN, WDR86, C11orf80, DKK3, HIST1H2AC, MRAP2, PPIC, SNUPN, WDR87, C11orf82, DKK4, HIST1H2AD, MRAS, PPID, SNURF, WDR88, C11orf83, DKKL1, HIST1H2AE, MRC2, PPIE, SNW1, WDR89, C11orf84, DLAT, HIST1H2AG, MRE11A, PPIF, SNX1, WDR90, C11orf85, DLC1, HIST1H2AH, MREG, PPIG, SNX10, WDR91, C11orf86, DLD, HIST1H2AI, MRFAP1, PPIH, SNX11, WDR92, C11orf87, DLEC1, HIST1H2AJ, MRFAP1L1, PPIL1, SNX12, WDR93, C11orf88, DLEU7, HIST1H2AK, MRGBP, PPIL2, SNX13, WDR96, C11orf91, DLG1, HIST1H2AL, MRGPRD, PPIL3, SNX14, WDSUB1, C11orf94, DLG2, HIST1H2AM, MRGPRE, PPIL4, SNX15, WDTC1, C11orf95, DLG3, HIST1H2BA, MRGPRF, PPIL6, SNX16, WDYHV1, C11orf96, DLG4, HIST1H2BB, MRGPRG, PPIP5K1, SNX17, WEE1, C12orf10, DLG5, HIST1H2BC, MRGPRX1, PPIP5K2, SNX18, WEE2, C12orf23, DLGAP1, HIST1H2BD, MRGPRX2, PPL, SNX19, WFDC1, C12orf29, DLGAP2, HIST1H2BE, MRGPRX3, PPM1A, SNX2, WFDC10A, C12orf39, DLGAP4, HIST1H2BF, MRGPRX4, PPM1B, SNX20, WFDC10B, C12orf4, DLGAP5, HIST1H2BG, MRI1, PPM1D, SNX21, WFDC11, C12orf40, DLK1, HIST1H2BH, MRM1, PPM1E, SNX22, WFDC12, C12orf42, DLK2, HIST1H2BI, MRO, PPM1F, SNX24, WFDC13, C12orf43, DLL1, HIST1H2BJ, MROH1, PPM1G, SNX25, WFDC2, C12orf44, DLL3, HIST1H2BK, MROH2A, PPM1H, SNX27, WFDC3, C12orf45, DLL4, HIST1H2BL, MROH2B, PPM1J, SNX29, WFDC5, C12orf49, DLST, HIST1H2BM, MROH5, PPM1K, SNX3, WFDC6, C12orf5, DLX1, HIST1H2BN, MROH6, PPM1L, SNX30, WFDC8, C12orf50, DLX2, HIST1H2BO, MROH7, PPM1M, SNX31, WFDC9, C12orf52, DLX3, HIST1H3A, MROH8, PPM1N, SNX32, WFIKKN1, C12orf54, DLX4, HIST1H3B, MROH9, PPME1, SNX33, WFIKKN2, C12orf55, DLX5, HIST1H3C, MRP63, PPOX, SNX4, WFS1, C12orf56, DLX6, HIST1H3D, MRPL1, PPP1CA, SNX5, WHAMM, C12orf57, DMAP1, HIST1H3E, MRPL10, PPP1CB, SNX6, WHSC1, C12orf60, DMBT1, HIST1H3F, MRPL11, PPP1CC, SNX7, WHSC1L1, C12orf61, DMBX1, HIST1H3G, MRPL12, PPP1R10, SNX8, WIBG, C12orf65, DMC1, HIST1H3H, MRPL13, PPP1R11, SNX9, WIF1, C12orf66, DMD, HIST1H3I, MRPL14, PPP1R12A, SOAT1, WIPF1, C12orf68, DMGDH, HIST1H3J, MRPL15, PPP1R12B, SOAT2, WIPF2, C12orf71, DMKN, HIST1H4A, MRPL16, PPP1R12C, SOBP, WIPF3, C12orf73, DMP1, HIST1H4B, MRPL17, PPP1R13B, SOCS1, WIPI1, C12orf74, DMPK, HIST1H4C, MRPL18, PPP1R13L, SOCS2, WIPI2, C12orf75, DMRT1, HIST1H4D, MRPL19, PPP1R14A, SOCS3, WISP1, C12orf76, DMRT2, HIST1H4E, MRPL2, PPP1R14B, SOCS4, WISP2, C12orf77, DMRT3, HIST1H4F, MRPL20, PPP1R14C, SOCS5, WISP3, C12orf79, DMRTA1, HIST1H4G, MRPL21, PPP1R14D, SOCS6, WIZ, C12orf80, DMRTA2, HIST1H4H, MRPL22, PPP1R15A, SOCS7, WLS, C13orf35, DMRTB1, HIST1H4I, MRPL23, PPP1R15B, SOD1, WNK1, C13orf45, DMRTC1, HIST1H4J, MRPL24, PPP1R16A, SOD2, WNK2, C14orf1, DMRTC1B, HIST1H4K, MRPL27, PPP1R16B, SOD3, WNK3, C14orf105, DMRTC2, HIST1H4L, MRPL28, PPP1R17, SOGA1, WNK4, C14orf119, DMTF1, HIST2H2AA3, MRPL3, PPP1R18, SOGA2, WNT1, C14orf132, DMTN, HIST2H2AA4, MRPL30,

PPP1R1A, SOGA3, WNT10A, C14orf142, DMWD, HIST2H2AB, MRPL32, PPP1R1B, SOHLH1, WNT10B, C14orf159, DMXL1, HIST2H2AC, MRPL33, PPP1R1C, SOHLH2, WNT11, C14orf164, DMXL2, HIST2H2BE, MRPL34, PPP1R2, SON, WNT16, C14orf166, DNA2, HIST2H2BF, MRPL35, PPP1R21, SORBS1, WNT2, C14orf166B, DNAAF1, HIST2H3A, MRPL36, PPP1R26, SORBS2, WNT2B, C14orf169, DNAAF2, HIST2H3C, MRPL37, PPP1R27, SORBS3, WNT3, C14orf177, DNAAF3, HIST2H3D, MRPL38, PPP1R32, SORCS1, WNT3A, C14orf178, DNAH1, HIST2H4A, MRPL39, PPP1R35, SORCS2, WNT4, C14orf180, DNAH10, HIST2H4B, MRPL4, PPP1R36, SORCS3, WNT5A, C14orf182, DNAH11, HIST3H2A, MRPL40, PPP1R37, SORD, WNT5B, C14orf183, DNAH12, HIST3H2BB, MRPL41, PPP1R3A, SORL1, WNT6, C14orf2, DNAH14, HIST3H3, MRPL42, PPP1R3B, SORT1, WNT7A, C14orf28, DNAH17, HIST4H4, MRPL43, PPP1R3C, SOS1, WNT7B, C14orf37, DNAH2, HIVEP1, MRPL44, PPP1R3D, SOS2, WNT8A, C14orf39, DNAH3, HIVEP2, MRPL45, PPP1R3E, SOST, WNT8B, C14orf79, DNAH5, HIVEP3, MRPL46, PPP1R3F, SOSTDC1, WNT9A, C14orf80, DNAH6, HJURP, MRPL47, PPP1R3G, SOWAHA, WNT9B, C14orf93, DNAH7, HK1, MRPL48, PPP1R42, SOWAHB, WRAP53, C15orf26, DNAH8, HK2, MRPL49, PPP1R7, SOWAHC, WRAP73, C15orf27, DNAH9, HK3, MRPL50, PPP1R8, SOWAHD, WRB, C15orf32, DNAI1, HKDC1, MRPL51, PPP1R9A, SOX1, WRN, C15orf38, DNAI2, HKR1, MRPL52, PPP1R9B, SOX10, WRNIP1, C15orf38-AP3S2, DNAJA1, HLA-A, MRPL53, PPP2CA, SOX11, WSB1, C15orf39, DNAJA2, HLA-B, MRPL54, PPP2CB, SOX12, WSB2, C15orf40, DNAJA3, HLA-C, MRPL55, PPP2R1A, SOX13, WSCD1, C15orf41, DNAJA4, HLA-DMA, MRPL9, PPP2R1B, SOX14, WSCD2, C15orf43, DNAJB1, HLA-DMB, MRPS10, PPP2R2A, SOX15, WT1, C15orf48, DNAJB11, HLA-DOA, MRPS11, PPP2R2B, SOX17, WTAP, C15orf52, DNAJB12, HLA-DOB, MRPS12, PPP2R2C, SOX18, WTH3DI, C15orf53, DNAJB13, HLA-DPA1, MRPS14, PPP2R2D, SOX2, WTIP, C15orf54, DNAJB14, HLA-DPB1, MRPS15, PPP2R3A, SOX21, WWC1, C15orf56, DNAJB2, HLA- DQA1, MRPS16, PPP2R3B, SOX3, WWC2, C15orf57, DNAJB3, HLA-DQA2, MRPS17, PPP2R3C, SOX30, WWC3, C15orf59, DNAJB4, HLA-DQB1, MRPS18A, PPP2R4, SOX4, WWOX, C15orf60, DNAJB5, HLA-DQB2, MRPS18B, PPP2R5A, SOX5, WWP1, C15orf61, DNAJB6, HLA-DRA, MRPS18C, PPP2R5B, SOX6, WWP2, C15orf62, DNAJB7, HLA-DRB1, MRPS2, PPP2R5C, SOX7, WWTR1, C15orf65, DNAJB8, HLA- DRB5, MRPS21, PPP2R5D, SOX8, XAB2, C16orf11, DNAJB9, HLA-E, MRPS22, PPP2R5E, SOX9, XAF1, C16orf13, DNAJC1, HLA-F, MRPS23, PPP3CA, SP1, XAGE1D, C16orf3, DNAJC10, HLA-G, MRPS24, PPP3CB, SP100, XAGE2, C16orf45, DNAJC11, HLCS, MRPS25, PPP3CC, SP110, XAGE3, C16orf46, DNAJC12, HLF, MRPS26, PPP3R1, SP140, XAGE5, C16orf47, DNAJC13, HLTF, MRPS27, PPP3R2, SP140L, XBP1, C16orf52, DNAJC14, HLX, MRPS28, PPP4C, SP2, XCL1, C16orf54, DNAJC15, HM13, MRPS30, PPP4R1, SP3, XCL2, C16orf58, DNAJC16, HMBOX1, MRPS31, PPP4R2, SP4, XCR1, C16orf59, DNAJC17, HMBS, MRPS33, PPP4R4, SP5, XDH, C16orf62, DNAJC18, HMCES, MRPS34, PPP5C, SP6, XG, C16orf70, DNAJC19, HMCN1, MRPS35, PPP5D1, SP7, XIAP, C16orf71, DNAJC2, HMCN2, MRPS36, PPP6C, SP8, XIRP1, C16orf72, DNAJC21, HMG20A, MRPS5, PPP6R1, SP9, XIRP2,

C16orf74, DNAJC22, HMG20B, MRPS6, PPP6R2, SPA17, XK, C16orf78, DNAJC24, HMGA1, MRPS7, PPP6R3, SPACA1, XKR3, C16orf80, DNAJC25, HMGA2, MRPS9, PPRC1, SPACA3, XKR4, C16orf82, DNAJC25-GNG10, HMGB1, MRRF, PPT1, SPACA4, XKR5, C16orf86, DNAJC27, HMGB2, MRS2, PPT2, SPACA5, XKR6, C16orf87, DNAJC28, HMGB3, MRTO4, PPTC7, SPACA5B, XKR7, C16orf89, DNAJC3, HMGB4, MRVI1, PPWD1, SPACA7, XKR8, C16orf90, DNAJC30, HMGCL, MS4A1, PPY, SPAG1, XKR9, C16orf91, DNAJC4, HMGCLL1, MS4A10, PQBP1, SPAG11A, XKRX, C16orf92, DNAJC5, HMGCR, MS4A12, PQLC1, SPAG11B, XKRY, C16orf93, DNAJC5B, HMGCS1, MS4A13, PQLC2, SPAG16, XKRY2, C16orf95, DNAJC5G, HMGCS2, MS4A14, PQLC3, SPAG17, XPA, C16orf96, DNAJC6, HMGN1, MS4A15, PRAC1, SPAG4, XPC, C16orf97, DNAJC7, HMGN2, MS4A2, PRAC2, SPAG5, XPNPEP1, C17orf100, DNAJC8, HMGN3, MS4A3, PRADC1, SPAG6, XPNPEP2, C17orf102, DNAJC9, HMGN4, MS4A4A, PRAF2, SPAG7, XPNPEP3, C17orf103, DNAL1, HMGN5, MS4A4E, PRAM1, SPAG8, XPO1, C17orf104, DNAL4, HMGXB3, MS4A5, PRAME, SPAG9, XPO4, C17orf105, DNALI1, HMGXB4, MS4A6A, PRAMEF1, SPAM1, XPO5, C17orf107, DNASE1, HMHA1, MS4A6E, PRAMEF10, SPANXA1, XPO6, C17orf112, DNASE1L1, HMHB1, MS4A7, PRAMEF11, SPANXA2, XPO7, C17orf47, DNASE1L2, HMMR, MS4A8, PRAMEF12, SPANXB1, XPOT, C17orf49, DNASE1L3, HMOX1, MSANTD1, PRAMEF13, SPANXB2, XPR1, C17orf50, DNASE2, HMOX2, MSANTD2, PRAMEF14, SPANXC, XRCC1, C17orf51, DNASE2B, HMP19, MSANTD3, PRAMEF15, SPANXD, XRCC2, C17orf53, DND1, HMSD, MSANTD3-TMEFF1, PRAMEF16, SPANXN1, XRCC3, C17orf58, DNER, HMX1, MSANTD4, PRAMEF17, SPANXN2, XRCC4, C17orf59, DNHD1, HMX2, MSC, PRAMEF18, SPANXN3, XRCC5, C17orf62, DNLZ, HMX3, MSGN1, PRAMEF19, SPANXN4, XRCC6, C17orf64, DNM1, HN1, MSH2, PRAMEF2, SPANXN5, XRCC6BP1, C17orf66, DNM1L, HN1L, MSH3, PRAMEF20, SPARC, XRN1, C17orf67, DNM2, HNF1A, MSH4, PRAMEF21, SPARCL1, XRN2, C17orf70, DNM3, HNF1B, MSH5, PRAMEF22, SPAST, XRRA1, C17orf72, DNMBP, HNF4A, MSH6, PRAMEF23, SPATA1, XXYLT1, C17orf74, DNMT1, HNF4G, MSI1, PRAMEF25, SPATA12, XYLB, C17orf75, DNMT3A, HNMT, MSI2, PRAMEF3, SPATA13, XYLT1, C17orf77, DNMT3B, HNRNPA0, MSL1, PRAMEF4, SPATA16, XYLT2, C17orf78, DNMT3L, HNRNPA1, MSL2, PRAMEF5, SPATA17, YAE1D1, C17orf80, DNPEP, HNRNPA1L2, MSL3, PRAMEF6, SPATA18, YAF2, C17orf82, DNPH1, HNRNPA2B1, MSLN, PRAMEF7, SPATA19, YAP1, C17orf85, DNTT, HNRNPA3, MSMB, PRAMEF8, SPATA2, YARS, C17orf89, DNTTIP1, HNRNPAB, MSMO1, PRAMEF9, SPATA20, YARS2, C17orf96, DNTTIP2, HNRNPC, MSMP, PRAP1, SPATA21, YBEY, C17orf97, DOC2A, HNRNPCL1, MSN, PRB1, SPATA22, YBX1, C17orf98, DOC2B, HNRNPCP5, MSR1, PRB2, SPATA24, YBX2, C17orf99, DOCK1, HNRNPD, MSRA, PRB3, SPATA25, YBX3, C18orf21, DOCK10, HNRNPDL, MSRB1, PRB4, SPATA2L, YDJC, C18orf25, DOCK11, HNRNPF, MSRB2, PRC1, SPATA3, YEATS2, C18orf32, DOCK2, HNRNPH1, MSRB3, PRCC, SPATA31A1, YEATS4, C18orf42, DOCK3, HNRNPH2, MSS51, PRCD, SPATA31A2, YES1, C18orf54, DOCK4, HNRNPH3, MST1, PRCP, SPATA31A3, YIF1A, C18orf56, DOCK5, HNRNPK,

MST1L, PRDM1, SPATA31A4, YIF1B, C18orf63, DOCK6, HNRNPL, MST1R, PRDM10, SPATA31A5, YIPF1, C18orf8, DOCK7, HNRNPLL, MST4, PRDM11, SPATA31A6, YIPF2, C19orf10, DOCK8, HNRNPM, MSTN, PRDM12, SPATA31A7, YIPF3, C19orf12, DOCK9, HNRNPR, MSTO1, PRDM13, SPATA31C1, YIPF4, C19orf18, DOHH, HNRNPU, MSX1, PRDM14, SPATA31C2, YIPF5, C19orf24, DOK1, HNRNPUL1, MSX2, PRDM15, SPATA31D1, YIPF6, C19orf25, DOK2, HNRNPUL2, MT1A, PRDM16, SPATA31D3, YIPF7, C19orf26, DOK3, HOGA1, MT1B, PRDM2, SPATA31D4, YJEFN3, C19orf33, DOK4, HOMER1, MT1E, PRDM4, SPATA31E1, YKT6, C19orf35, DOK5, HOMER2, MT1F, PRDM5, SPATA32, YLPM1, C19orf38, DOK6, HOMER3, MT1G, PRDM6, SPATA33, YME1L1, C19orf40, DOK7, HOMEZ, MT1H, PRDM7, SPATA4, YOD1, C19orf43, DOLK, HOOK1, MT1HL1, PRDM8, SPATA5, YPEL1, C19orf44, DOLPP1, HOOK2, MT1M, PRDM9, SPATA5L1, YPEL2, C19orf45, DONSON, HOOK3, MT1X, PRDX1, SPATA6, YPEL3, C19orf47, DOPEY1, HOPX, MT2A, PRDX2, SPATA6L, YPEL4, C19orf48, DOPEY2, HORMAD1, MT3, PRDX3, SPATA7, YPEL5, C19orf52, DOT1L, HORMAD2, MT4, PRDX4, SPATA8, YRDC, C19orf53, DPAGT1, HOXA1, MTA1, PRDX5, SPATA9, YTHDC1, C19orf54, DPCD, HOXA10, MTA2, PRDX6, SPATC1, YTHDC2, C19orf55, DPCR1, HOXA11, MTA3, PREB, SPATC1L, YTHDF1, C19orf57, DPEP1, HOXA13, MTAP, PRELID1, SPATS1, YTHDF2, C19orf59, DPEP2, HOXA2, MTBP, PRELID2, SPATS2, YTHDF3, C19orf60, DPEP3, HOXA3, MTCH1, PRELP, SPATS2L, YWHAB, C19orf66, DPF1, HOXA4, MTCH2, PREP, SPC24, YWHAE, C19orf67, DPF2, HOXA5, MTCP1, PREPL, SPC25, YWHAG, C19orf68, DPF3, HOXA6, MTDH, PREX1, SPCS1, YWHAH, C19orf69, DPH1, HOXA7, MTERF, PREX2, SPCS2, YWHAQ, C19orf70, DPH2, HOXA9, MTERFD1, PRF1, SPCS3, YWHAZ, C19orf71, DPH3, HOXB1, MTERFD2, PRG2, SPDEF, YY1, C19orf73, DPH3P1, HOXB13, MTERFD3, PRG3, SPDL1, YY1AP1, C19orf77, DPH5, HOXB2, MTF1, PRG4, SPDYA, YY2, C19orf80, DPH6, HOXB3, MTF2, PRH1, SPDYC, ZACN, C19orf81, DPH7, HOXB4, MTFMT, PRH2, SPDYE1, ZADH2, C19orf82, DPM1, HOXB5, MTFP1, PRICKLE1, SPDYE2, ZAK, C1D, DPM2, HOXB6, MTFR1, PRICKLE2, SPDYE2B, ZAN, C1GALT1, DPM3, HOXB7, MTFR1L, PRICKLE3, SPDYE3, ZAP70, C1GALT1C1, DPP10, HOXB8, MTFR2, PRICKLE4, SPDYE4, ZAR1, C1orf100, DPP3, HOXB9, MTG1, PRIM1, SPDYE5, ZAR1L, C1orf101, DPP4, HOXC10, MTG2, PRIM2, SPDYE6, ZBBX, C1orf105, DPP6, HOXC11, MTHFD1, PRIMA1, SPECC1, ZBED1, C1orf106, DPP7, HOXC12, MTHFD1L, PRIMPOL, SPECC1L, ZBED2, C1orf109, DPP8, HOXC13, MTHFD2, PRKAA1, SPEF1, ZBED3, C1orf110, DPP9, HOXC4, MTHFD2L, PRKAA2, SPEF2, ZBED4, C1orf111, DPPA2, HOXC5, MTHFR, PRKAB1, SPEG, ZBED5, C1orf112, DPPA3, HOXC6, MTHFS, PRKAB2, SPEM1, ZBED6, C1orf115, DPPA4, HOXC8, MTHFSD, PRKACA, SPEN, ZBED6CL, C1orf116, DPPA5, HOXC9, MTIF2, PRKACB, SPERT, ZBP1, C1orf122, DPRX, HOXD1, MTIF3, PRKACG, SPESP1, ZBTB1, C1orf123, DPT, HOXD10, MTL5, PRKAG1, SPG11, ZBTB10, C1orf127, DPY19L1, HOXD11, MTM1, PRKAG2, SPG20, ZBTB11, C1orf131, DPY19L2, HOXD12, MTMR1, PRKAG3, SPG21, ZBTB12, C1orf137, DPY19L3, HOXD13, MTMR10, PRKAR1A, SPG7, ZBTB14, C1orf141, DPY19L4, HOXD3, MTMR11, PRKAR1B,

SPHAR, ZBTB16, C1orf146, DPY30, HOXD4, MTMR12, PRKAR2A, SPHK1, ZBTB17, C1orf158, DPYD, HOXD8, MTMR14, PRKAR2B, SPHK2, ZBTB18, C1orf159, DPYS, HOXD9, MTMR2, PRKCA, SPHKAP, ZBTB2, C1orf162, DPYSL2, HP, MTMR3, PRKCB, SPI1, ZBTB20, C1orf167, DPYSL3, HP1BP3, MTMR4, PRKCD, SPIB, ZBTB21, C1orf168, DPYSL4, HPCA, MTMR6, PRKCDBP, SPIC, ZBTB22, C1orf172, DPYSL5, HPCAL1, MTMR7, PRKCE, SPICE1, ZBTB24, C1orf173, DQX1, HPCAL4, MTMR8, PRKCG, SPIDR, ZBTB25, C1orf174, DR1, HPD, MTMR9, PRKCH, SPIN1, ZBTB26, C1orf177, DRAM1, HPDL, MTNR1A, PRKCI, SPIN2A, ZBTB3, C1orf185, DRAM2, HPGD, MTNR1B, PRKCQ, SPIN2B, ZBTB32, C1orf186, DRAP1, HPGDS, MTO1, PRKCSH, SPIN3, ZBTB33, C1orf189, DRAXIN, HPN, MTOR, PRKCZ, SPIN4, ZBTB34, C1orf192, DRC1, HPR, MTPAP, PRKD1, SPINK1, ZBTB37, C1orf194, DRD1, HPRT1, MTPN, PRKD2, SPINK13, ZBTB38, C1orf195, DRD2, HPS1, MTR, PRKD3, SPINK14, ZBTB39, C1orf198, DRD3, HPS3, MTRF1, PRKDC, SPINK2, ZBTB4, C1orf204, DRD4, HPS4, MTRF1L, PRKG1, SPINK4, ZBTB40, C1orf21, DRD5, HPS5, MTRNR2L10, PRKG2, SPINK5, ZBTB41, C1orf210, DRG1, HPS6, MTRNR2L2, PRKRA, SPINK6, ZBTB42, C1orf216, DRG2, HPSE, MTRNR2L3, PRKRIP1, SPINK7, ZBTB43, C1orf226, DRGX, HPSE2, MTRNR2L4, PRKRIR, SPINK8, ZBTB44, C1orf227, DROSHA, HPX, MTRNR2L5, PRKX, SPINK9, ZBTB45, C1orf228, DRP2, HR, MTRNR2L6, PRL, SPINT1, ZBTB46, C1orf229, DSC1, HRAS, MTRNR2L7, PRLH, SPINT2, ZBTB47, C1orf233, DSC2, HRASLS, MTRNR2L9, PRLHR, SPINT3, ZBTB48, C1orf234, DSC3, HRASLS2, MTRR, PRLR, SPINT4, ZBTB49, C1orf27, DSCAM, HRASLS5, MTSS1, PRM1, SPIRE1, ZBTB5, C1orf35, DSCAML1, HRC, MTSS1L, PRM2, SPIRE2, ZBTB6, C1orf43, DSCC1, HRCT1, MTTP, PRM3, SPN, ZBTB7A, C1orf50, DSCR3, HRG, MTURN, PRMT1, SPNS1, ZBTB7B, C1orf51, DSCR4, HRH1, MTUS1, PRMT10, SPNS2, ZBTB7C, C1orf52, DSE, HRH2, MTUS2, PRMT2, SPNS3, ZBTB8A, C1orf53, DSEL, HRH3, MTX1, PRMT3, SPO11, ZBTB8B, C1orf54, DSG1, HRH4, MTX2, PRMT5, SPOCD1, ZBTB8OS, C1orf56, DSG2, HRK, MTX3, PRMT6, SPOCK1, ZBTB9, C1orf61, DSG3, HRNR, MUC1, PRMT7, SPOCK2, ZC2HC1A, C1orf63, DSG4, HRSP12, MUC12, PRMT8, SPOCK3, ZC2HC1B, C1orf64, DSN1, HS1BP3, MUC13, PRND, SPON1, ZC2HC1C, C1orf65, DSP, HS2ST1, MUC15, PRNP, SPON2, ZC3H10, C1orf68, DSPP, HS3ST1, MUC16, PROB1, SPOP, ZC3H11A, C1orf74, DST, HS3ST2, MUC17, PROC, SPOPL, ZC3H12A, C1orf85, DSTN, HS3ST3A1, MUC19, PROCA1, SPP1, ZC3H12B, C1orf86, DSTNP4, HS3ST3B1, MUC2, PROCR, SPP2, ZC3H12C, C1orf87, DSTYK, HS3ST4, MUC20, PRODH, SPPL2A, ZC3H12D, C1orf94, DTD1, HS3ST5, MUC21, PRODH2, SPPL2B, ZC3H13, C1orf95, DTD2, HS3ST6, MUC22, PROK1, SPPL2C, ZC3H14, C1QA, DTHD1, HS6ST1, MUC3A, PROK2, SPPL3, ZC3H15, C1QB, DTL, HS6ST2, MUC4, PROKR1, SPR, ZC3H18, C1QBP, DTNA, HS6ST3, MUC5AC, PROKR2, SPRED1, ZC3H3, C1QC, DTNB, HSBP1, MUC5B, PROL1, SPRED2, ZC3H4, C1QL1, DTNBP1, HSBP1L1, MUC6, PROM1, SPRED3, ZC3H6, C1QL2, DTWD1, HSCB, MUC7, PROM2, SPRN, ZC3H7A, C1QL3, DTWD2, HSD11B1, MUC8, PROP1, SPRR1A, ZC3H7B, C1QL4, DTX1, HSD11B1L, MUCL1, PROS1, SPRR1B, ZC3H8, C1QTNF1, DTX2, HSD11B2, MUL1, PROSC, SPRR2A, ZC3HAV1, C1QTNF2, DTX3, HSD17B1,

MUM1, PROSER1, SPRR2B, ZC3HAV1L, C1QTNF3, DTX3L, HSD17B10, MUM1L1, PROSER2, SPRR2D, ZC3HC1, C1QTNF4, DTX4, HSD17B11, MURC, PROX1, SPRR2E, ZC4H2, C1QTNF5, DTYMK, HSD17B12, MUS81, PROX2, SPRR2F, ZCCHC10, C1QTNF6, DUOX1, HSD17B13, MUSK, PROZ, SPRR2G, ZCCHC11, C1QTNF7, DUOX2, HSD17B14, MUSTN1, PRPF18, SPRR3, ZCCHC12, C1QTNF8, DUOXA1, HSD17B2, MUT, PRPF19, SPRR4, ZCCHC13, C1QTNF9, DUOXA2, HSD17B3, MUTYH, PRPF3, SPRTN, ZCCHC14, C1QTNF9B, DUPD1, HSD17B4, MVB12A, PRPF31, SPRY1, ZCCHC16, C1QTNF9B-AS1, DUS1L, HSD17B6, MVB12B, PRPF38A, SPRY2, ZCCHC17, C1R, DUS2, HSD17B7, MVD, PRPF38B, SPRY3, ZCCHC18, C1RL, DUS3L, HSD17B8, MVK, PRPF39, SPRY4, ZCCHC2, C1S, DUS4L, HSD3B1, MVP, PRPF4, SPRYD3, ZCCHC24, C2, DUSP1, HSD3B2, MX1, PRPF40A, SPRYD4, ZCCHC3, C20orf112, DUSP10, HSD3B7, MX2, PRPF40B, SPRYD7, ZCCHC4, C20orf141, DUSP11, HSDL1, MXD1, PRPF4B, SPSB1, ZCCHC5, C20orf144, DUSP12, HSDL2, MXD3, PRPF6, SPSB2, ZCCHC6, C20orf166, DUSP13, HSF1, MXD4, PRPF8, SPSB3, ZCCHC7, C20orf173, DUSP14, HSF2, MXI1, PRPH, SPSB4, ZCCHC8, C20orf194, DUSP15, HSF2BP, MXRA5, PRPH2, SPTA1, ZCCHC9, C20orf195, DUSP16, HSF4, MXRA7, PRPS1, SPTAN1, ZCRB1, C20orf196, DUSP18, HSF5, MXRA8, PRPS1L1, SPTB, ZCWPW1, C20orf197, DUSP19, HSFX1, MYADM, PRPS2, SPTBN1, ZCWPW2, C20orf201, DUSP2, HSFX2, MYADML2, PRPSAP1, SPTBN2, ZDBF2, C20orf202, DUSP21, HSFY1, MYB, PRPSAP2, SPTBN4, ZDHHC1, C20orf203, DUSP22, HSFY2, MYBBP1A, PRR11, SPTBN5, ZDHHC11, C20orf24, DUSP23, HSH2D, MYBL1, PRR12, SPTLC1, ZDHHC11B, C20orf26, DUSP26, HSP90AA1, MYBL2, PRR13, SPTLC2, ZDHHC12, C20orf27, DUSP27, HSP90AB1, MYBPC1, PRR14, SPTLC3, ZDHHC13, C20orf62, DUSP28, HSP90B1, MYBPC2, PRR14L, SPTSSA, ZDHHC14, C20orf78, DUSP3, HSPA12A, MYBPC3, PRR15, SPTSSB, ZDHHC15, C20orf85, DUSP4, HSPA12B, MYBPH, PRR15L, SPTY2D1, ZDHHC16, C20orf96, DUSP5, HSPA13, MYBPHL, PRR16, SPZ1, ZDHHC17, C21orf140, DUSP6, HSPA14, MYC, PRR18, SQLE, ZDHHC18, C21orf2, DUSP7, HSPA1A, MYCBP, PRR19, SQRDL, ZDHHC19, C21orf33, DUSP8, HSPA1B, MYCBP2, PRR20A, SQSTM1, ZDHHC2, C21orf58, DUSP9, HSPA1L, MYCBPAP, PRR20B, SRA1, ZDHHC20, C21orf59, DUT, HSPA2, MYCL, PRR20C, SRBD1, ZDHHC21, C21orf62, DUXA, HSPA4, MYCN, PRR20D, SRC, ZDHHC22, C21orf91, DVL1, HSPA4L, MYCT1, PRR20E, SRCAP, ZDHHC23, C22orf15, DVL2, HSPA5, MYD88, PRR21, SRCIN1, ZDHHC24, C22orf23, DVL3, HSPA6, MYEF2, PRR22, SRCRB4D, ZDHHC3, C22orf24, DXO, HSPA8, MYEOV, PRR23A, SRD5A1, ZDHHC4, C22orf26, DYDC1, HSPA9, MYEOV2, PRR23B, SRD5A2, ZDHHC5, C22orf29, DYDC2, HSPB1, MYF5, PRR23C, SRD5A3, ZDHHC6, C22orf31, DYM, HSPB11, MYF6, PRR24, SREBF1, ZDHHC7, C22orf39, DYNAP, HSPB2, MYH1, PRR25, SREBF2, ZDHHC8, C22orf42, DYNC1H1, HSPB3, MYH10, PRR3, SREK1, ZDHHC9, C22orf43, DYNC1I1, HSPB6, MYH11, PRR30, SREK1IP1, ZEB1, C22orf46, DYNC1I2, HSPB7, MYH13, PRR4, SRF, ZEB2, C2CD2, DYNC1LI1, HSPB8, MYH14, PRR5, SRFBP1, ZER1, C2CD2L, DYNC1LI2, HSPB9, MYH15, PRR5-ARHGAP8, SRGAP1, ZFAND1, C2CD3, DYNC2H1, HSPBAP1, MYH2, PRR5L, SRGAP2, ZFAND2A, C2CD4A,

DYNC2LI1, HSPBP1, MYH3, PRR7, SRGAP2B, ZFAND2B, C2CD4B, DYNLL1, HSPD1, MYH4, PRR9, SRGAP2C, ZFAND3, C2CD4C, DYNLL2, HSPE1, MYH6, PRRC1, SRGAP3, ZFAND4, C2CD4D, DYNLRB1, HSPE1-MOB4, MYH7, PRRC2A, SRGN, ZFAND5, C2CD5, DYNLRB2, HSPG2, MYH7B, PRRC2B, SRI, ZFAND6, C2orf15, DYNLT1, HSPH1, MYH8, PRRC2C, SRL, ZFAT, C2orf16, DYNLT3, HTATIP2, MYH9, PRRG1, SRM, ZFC3H1, C2orf27A, DYRK1A, HTATSF1, MYL1, PRRG2, SRMS, ZFHX2, C2orf27B, DYRK1B, HTN1, MYL10, PRRG3, SRP14, ZFHX3, C2orf40, DYRK2, HTN3, MYL12A, PRRG4, SRP19, ZFHX4, C2orf42, DYRK3, HTR1A, MYL12B, PRRT1, SRP54, ZFP1, C2orf43, DYRK4, HTR1B, MYL2, PRRT2, SRP68, ZFP14, C2orf44, DYSF, HTR1D, MYL3, PRRT3, SRP72, ZFP2, C2orf47, DYTN, HTR1E, MYL4, PRRT4, SRP9, ZFP28, C2orf48, DYX1C1, HTR1F, MYL5, PRRX1, SRPK1, ZFP3, C2orf49, DZANK1, HTR2A, MYL6, PRRX2, SRPK2, ZFP30, C2orf50, DZIP1, HTR2B, MYL6B, PRSS1, SRPK3, ZFP36, C2orf53, DZIP1L, HTR2C, MYL7, PRSS12, SRPR, ZFP36L1, C2orf54, DZIP3, HTR3A, MYL9, PRSS16, SRPRB, ZFP36L2, C2orf57, E2F1, HTR3B, MYLIP, PRSS2, SRPX, ZFP37, C2orf61, E2F2, HTR3C, MYLK, PRSS21, SRPX2, ZFP41, C2orf62, E2F3, HTR3D, MYLK2, PRSS22, SRR, ZFP42, C2orf66, E2F4, HTR3E, MYLK3, PRSS23, SRRD, ZFP57, C2orf68, E2F5, HTR4, MYLK4, PRSS27, SRRM1, ZFP62, C2orf69, E2F6, HTR5A, MYLPF, PRSS3, SRRM2, ZFP64, C2orf70, E2F7, HTR6, MYNN, PRSS33, SRRM3, ZFP69, C2orf71, E2F8, HTR7, MYO10, PRSS35, SRRM4, ZFP69B, C2orf72, E4F1, HTRA1, MYO15A, PRSS36, SRRM5, ZFP82, C2orf73, EAF1, HTRA2, MYO16, PRSS37, SRRT, ZFP90, C2orf74, EAF2, HTRA3, MYO18A, PRSS38, SRSF1, ZFP91, C2orf76, EAPP, HTRA4, MYO18B, PRSS41, SRSF10, ZFP92, C2orf78, EARS2, HTT, MYO19, PRSS42, SRSF11, ZFPL1, C2orf80, EBAG9, HUNK, MYO1A, PRSS44, SRSF12, ZFPM1, C2orf81, EBF1, HUS1, MYO1B, PRSS45, SRSF2, ZFPM2, C2orf82, EBF2, HUS1B, MYO1C, PRSS46, SRSF3, ZFR, C2orf83, EBF3, HUWE1, MYO1D, PRSS48, SRSF4, ZFR2, C2orf88, EBF4, HVCN1, MYO1E, PRSS50, SRSF5, ZFX, C2orf91, EBI3, HYAL1, MYO1F, PRSS53, SRSF6, ZFY, C3, EBLN1, HYAL2, MYO1G, PRSS54, SRSF7, ZFYVE1, C3AR1, EBLN2, HYAL3, MYO1H, PRSS55, SRSF8, ZFYVE16, C3orf14, EBNA1BP2, HYAL4, MYO3A, PRSS56, SRSF9, ZFYVE19, C3orf17, EBP, HYDIN, MYO3B, PRSS57, SRXN1, ZFYVE20, C3orf18, EBPL, HYI, MYO5A, PRSS58, SRY, ZFYVE21, C3orf20, ECD, HYKK, MYO5B, PRSS8, SS18, ZFYVE26, C3orf22, ECE1, HYLS1, MYO5C, PRTFDC1, SS18L1, ZFYVE27, C3orf27, ECE2, HYOU1, MYO6, PRTG, SS18L2, ZFYVE28, C3orf30, ECEL1, HYPK, MYO7A, PRTN3, SSB, ZFYVE9, C3orf33, ECH1, IAH1, MYO7B, PRUNE, SSBP1, ZG16, C3orf35, ECHDC1, IAPP, MYO9A, PRUNE2, SSBP2, ZG16B, C3orf36, ECHDC2, IARS, MYO9B, PRX, SSBP3, ZGLP1, C3orf38, ECHDC3, IARS2, MYOC, PRY, SSBP4, ZGPAT, C3orf43, ECHS1, IBA57, MYOCD, PRY2, SSC5D, ZHX1, C3orf52, ECI1, IBSP, MYOD1, PSAP, SSFA2, ZHX1-C8ORF76, C3orf55, ECI2, IBTK, MYOF, PSAPL1, SSH1, ZHX2, C3orf56, ECM1, ICA1, MYOG, PSAT1, SSH2, ZHX3, C3orf58, ECM2, ICA1L, MYOM1, PSCA, SSH3, ZIC1, C3orf62, ECSCR, ICAM1, MYOM2, PSD, SSMEM1, ZIC2, C3orf67, ECSIT, ICAM2, MYOM3, PSD2, SSNA1, ZIC3, C3orf70, ECT2, ICAM3, MYOT, PSD3, SSPN, ZIC4, C3orf72, ECT2L, ICAM4, MYOZ1, PSD4, SSPO, ZIC5, C3orf79, EDA,

ICAM5, MYOZ2, PSEN1, SSR1, ZIK1, C3orf80, EDA2R, ICK, MYOZ3, PSEN2, SSR2, ZIM2, C3orf83, EDAR, ICMT, MYPN, PSENEN, SSR3, ZIM3, C3orf84, EDARADD, ICOS, MYPOP, PSG1, SSR4, ZKSCAN1, C4A, EDC3, ICOSLG, MYRF, PSG11, SSRP1, ZKSCAN2, C4B, EDC4, ICT1, MYRFL, PSG2, SSSCA1, ZKSCAN3, C4B_2, EDDM3A, ID1, MYRIP, PSG3, SST, ZKSCAN4, C4BPA, EDDM3B, ID2, MYSM1, PSG4, SSTR1, ZKSCAN5, C4BPB, EDEM1, ID3, MYT1, PSG5, SSTR2, ZKSCAN7, C4orf17, EDEM2, ID4, MYT1L, PSG6, SSTR3, ZKSCAN8, C4orf19, EDEM3, IDE, MYZAP, PSG7, SSTR4, ZMAT1, C4orf21, EDF1, IDH1, MZB1, PSG8, SSTR5, ZMAT2, C4orf22, EDIL3, IDH2, MZF1, PSG9, SSU72, ZMAT3, C4orf26, EDN1, IDH3A, MZT1, PSIP1, SSUH2, ZMAT4, C4orf27, EDN2, IDH3B, MZT2A, PSKH1, SSX1, ZMAT5, C4orf29, EDN3, IDH3G, MZT2B, PSKH2, SSX2, ZMIZ1, C4orf3, EDNRA, IDI1, N4BP1, PSMA1, SSX2B, ZMIZ2, C4orf32, EDNRB, IDI2, N4BP2, PSMA2, SSX2IP, ZMPSTE24, C4orf33, EEA1, IDNK, N4BP2L1, PSMA3, SSX3, ZMYM1, C4orf36, EED, IDO1, N4BP2L2, PSMA4, SSX4, ZMYM2, C4orf40, EEF1A1, IDO2, N4BP3, PSMA5, SSX4B, ZMYM3, C4orf45, EEF1A2, IDS, N6AMT1, PSMA6, SSX5, ZMYM4, C4orf46, EEF1B2, IDUA, N6AMT2, PSMA7, SSX7, ZMYM5, C4orf47, EEF1D, IER2, NAA10, PSMA8, ST13, ZMYM6, C4orf48, EEF1E1, IER3, NAA11, PSMB1, ST14, ZMYM6NB, C4orf50, EEF1G, IER3IP1, NAA15, PSMB10, ST18, ZMYND10, C4orf51, EEF2, IER5, NAA16, PSMB11, ST20, ZMYND11, C4orf6, EEF2K, IER5L, NAA20, PSMB2, ST20-MTHFS, ZMYND12, C5, EEFSEC, IFFO1, NAA25, PSMB3, ST3GAL1, ZMYND15, C5AR1, EEPD1, IFFO2, NAA30, PSMB4, ST3GAL2, ZMYND19, C5AR2, EFCAB1, IFI16, NAA35, PSMB5, ST3GAL3, ZMYND8, C5orf15, EFCAB11, IFI27, NAA38, PSMB6, ST3GAL4, ZNF10, C5orf20, EFCAB12, IFI27L1, NAA40, PSMB7, ST3GAL5, ZNF100, C5orf22, EFCAB13, IFI27L2, NAA50, PSMB8, ST3GAL6, ZNF101, C5orf24, EFCAB14, IFI30, NAA60, PSMB9, ST5, ZNF106, C5orf28, EFCAB2, IFI35, NAAA, PSMC1, ST6GAL1, ZNF107, C5orf30, EFCAB3, IFI44, NAALAD2, PSMC2, ST6GAL2, ZNF112, C5orf34, EFCAB4A, IFI44L, NAALADL1, PSMC3, ST6GALNAC1, ZNF114, C5orf38, EFCAB4B, IFI6, NAALADL2, PSMC3IP, ST6GALNAC2, ZNF117, C5orf42, EFCAB5, IFIH1, NAB1, PSMC4, ST6GALNAC3, ZNF12, C5orf45, EFCAB6, IFIT1, NAB2, PSMC5, ST6GALNAC4, ZNF121, C5orf46, EFCAB7, IFIT1B, NABP1, PSMC6, ST6GALNAC5, ZNF124, C5orf47, EFCAB8, IFIT2, NABP2, PSMD1, ST6GALNAC6, ZNF131, C5orf48, EFCAB9, IFIT3, NACA, PSMD10, ST7, ZNF132, C5orf49, EFCC1, IFIT5, NACA2, PSMD11, ST7L, ZNF133, C5orf50, EFEMP1, IFITM1, NACAD, PSMD12, ST8SIA1, ZNF134, C5orf51, EFEMP2, IFITM10, NACC1, PSMD13, ST8SIA2, ZNF135, C5orf52, EFHB, IFITM2, NACC2, PSMD14, ST8SIA3, ZNF136, C5orf55, EFHC1, IFITM3, NADK, PSMD2, ST8SIA4, ZNF138, C5orf58, EFHC2, IFITM5, NADK2, PSMD3, ST8SIA5, ZNF14, C5orf60, EFHD1, IFLTD1, NADSYN1, PSMD4, ST8SIA6, ZNF140, C5orf63, EFHD2, IFNA1, NAE1, PSMD5, STAB1, ZNF141, C5orf64, EFNA1, IFNA10, NAF1, PSMD6, STAB2, ZNF142, C6, EFNA2, IFNA13, NAGA, PSMD7, STAC, ZNF143, C6orf1, EFNA3, IFNA14, NAGK, PSMD8, STAC2, ZNF146, C6orf10, EFNA4, IFNA16, NAGLU, PSMD9, STAC3, ZNF148, C6orf106, EFNA5, IFNA17, NAGPA, PSME1, STAG1, ZNF154, C6orf118, EFNB1, IFNA2, NAGS, PSME2, STAG2,

ZNF155, C6orf132, EFNB2, IFNA21, NAIF1, PSME3, STAG3, ZNF157, C6orf136, EFNB3, IFNA4, NAIP, PSME4, STAM, ZNF16, C6orf141, EFR3A, IFNA5, NALCN, PSMF1, STAM2, ZNF160, C6orf15, EFR3B, IFNA6, NAMPT, PSMG1, STAMBP, ZNF165, C6orf163, EFS, IFNA7, NANOG, PSMG2, STAMBPL1, ZNF169, C6orf165, EFTUD1, IFNA8, NANOGNB, PSMG3, STAP1, ZNF17, C6orf183, EFTUD2, IFNAR1, NANOS1, PSMG4, STAP2, ZNF174, C6orf195, EGF, IFNAR2, NANOS2, PSORS1C1, STAR, ZNF175, C6orf201, EGFL6, IFNB1, NANOS3, PSORS1C2, STARD10, ZNF177, C6orf203, EGFL7, IFNE, NANP, PSPC1, STARD13, ZNF18, C6orf211, EGFL8, IFNG, NANS, PSPH, STARD3, ZNF180, C6orf222, EGFLAM, IFNGR1, NAP1L1, PSPN, STARD3NL, ZNF181, C6orf223, EGFR, IFNGR2, NAP1L2, PSRC1, STARD4, ZNF182, C6orf226, EGLN1, IFNK, NAP1L3, PSTK, STARD5, ZNF184, C6orf25, EGLN2, IFNL1, NAP1L4, PSTPIP1, STARD6, ZNF185, C6orf47, EGLN3, IFNL2, NAP1L5, PSTPIP2, STARD7, ZNF189, C6orf48, EGR1, IFNL3, NAPA, PTAFR, STARD8, ZNF19, C6orf52, EGR2, IFNL4, NAPB, PTAR1, STARD9, ZNF195, C6orf57, EGR3, IFNLR1, NAPEPLD, PTBP1, STAT1, ZNF197, C6orf58, EGR4, IFNW1, NAPG, PTBP2, STAT2, ZNF2, C6orf62, EHBP1, IFRD1, NAPRT1, PTBP3, STAT3, ZNF20, C6orf7, EHBP1L1, IFRD2, NAPSA, PTCD1, STAT4, ZNF200, C6orf89, EHD1, IFT122, NARF, PTCD2, STAT5A, ZNF202, C6orf99, EHD2, IFT140, NARFL, PTCD3, STAT5B, ZNF205, C7, EHD3, IFT172, NARG2, PTCH1, STAT6, ZNF207, C7orf10, EHD4, IFT20, NARR, PTCH2, STATH, ZNF208, C7orf25, EHF, IFT27, NARS, PTCHD1, STAU1, ZNF211, C7orf26, EHHADH, IFT43, NARS2, PTCHD2, STAU2, ZNF212, C7orf31, EHMT1, IFT46, NASP, PTCHD3, STBD1, ZNF213, C7orf33, EHMT2, IFT52, NAT1, PTCHD4, STC1, ZNF214, C7orf34, EI24, IFT57, NAT10, PTCRA, STC2, ZNF215, C7orf43, EID1, IFT74, NAT14, PTDSS1, STEAP1, ZNF217, C7orf49, EID2, IFT80, NAT16, PTDSS2, STEAP1B, ZNF219, C7orf50, EID2B, IFT81, NAT2, PTEN, STEAP2, ZNF22, C7orf55, EID3, IFT88, NAT6, PTER, STEAP3, ZNF221, C7orf55-LUC7L2, EIF1, IGBP1, NAT8, PTF1A, STEAP4, ZNF222, C7orf57, EIF1AD, IGDCC3, NAT8B, PTGDR, STH, ZNF223, C7orf60, EIF1AX, IGDCC4, NAT8L, PTGDR2, STIL, ZNF224, C7orf61, EIF1AY, IGF1, NAT9, PTGDS, STIM1, ZNF225, C7orf62, EIF1B, IGF1R, NAV1, PTGER1, STIM2, ZNF226, C7orf63, EIF2A, IGF2, NAV2, PTGER2, STIP1, ZNF227, C7orf65, EIF2AK1, IGF2BP1, NAV3, PTGER3, STK10, ZNF229, C7orf66, EIF2AK2, IGF2BP2, NBAS, PTGER4, STK11, ZNF23, C7orf69, EIF2AK3, IGF2BP3, NBEA, PTGES, STK11IP, ZNF230, C7orf71, EIF2AK4, IGF2R, NBEAL1, PTGES2, STK16, ZNF232, C7orf72, EIF2B1, IGFALS, NBEAL2, PTGES3, STK17A, ZNF233, C7orf73, EIF2B2, IGFBP1, NBL1, PTGES3L, STK17B, ZNF234, C7orf76, EIF2B3, IGFBP2, NBN, PTGES3L-AARSD1, STK19, ZNF235, C8A, EIF2B4, IGFBP3, NBPF14, PTGFR, STK24, ZNF236, C8B, EIF2B5, IGFBP4, NBPF3, PTGFRN, STK25, ZNF239, C8G, EIF2D, IGFBP5, NBPF4, PTGIR, STK3, ZNF24, C8orf22, EIF2S1, IGFBP6, NBPF6, PTGIS, STK31, ZNF248, C8orf31, EIF2S2, IGFBP7, NBPF7, PTGR1, STK32A, ZNF25, C8orf33, EIF2S3, IGFBPL1, NBR1, PTGR2, STK32B, ZNF250, C8orf34, EIF3A, IGFL1, NCALD, PTGS1, STK32C, ZNF251, C8orf37, EIF3B, IGFL2, NCAM1, PTGS2, STK33, ZNF253, C8orf4, EIF3C, IGFL3, NCAM2, PTH, STK35, ZNF254, C8orf44, EIF3CL, IGFL4, NCAN, PTH1R, STK36, ZNF256,

C8orf44-SGK3, EIF3D, IGFLR1, NCAPD2, PTH2, STK38, ZNF257, C8orf46, EIF3E, IGFN1, NCAPD3, PTH2R, STK38L, ZNF259, C8orf47, EIF3F, IGHMBP2, NCAPG, PTHLH, STK39, ZNF26, C8orf48, EIF3G, IGIP, NCAPG2, PTK2, STK4, ZNF260, C8orf58, EIF3H, IGJ, NCAPH, PTK2B, STK40, ZNF263, C8orf59, EIF3I, IGLL1, NCAPH2, PTK6, STMN1, ZNF264, C8orf74, EIF3J, IGLL5, NCBP1, PTK7, STMN2, ZNF266, C8orf76, EIF3K, IGLON5, NCBP2, PTMA, STMN3, ZNF267, C8orf82, EIF3L, IGSF1, NCCRP1, PTMS, STMN4, ZNF268, C8orf86, EIF3M, IGSF10, NCDN, PTN, STMND1, ZNF273, C8orf87, EIF4A1, IGSF11, NCEH1, PTOV1, STOM, ZNF274, C9, EIF4A2, IGSF21, NCF1, PTP4A1, STOML1, ZNF275, C9orf106, EIF4A3, IGSF22, NCF2, PTP4A2, STOML2, ZNF276, C9orf114, EIF4B, IGSF23, NCF4, PTP4A3, STOML3, ZNF277, C9orf116, EIF4E, IGSF3, NCK1, PTPDC1, STON1, ZNF28, C9orf117, EIF4E1B, IGSF5, NCK2, PTPLA, STON1-GTF2A1L, ZNF280A, C9orf129, EIF4E2, IGSF6, NCKAP1, PTPLAD1, STON2, ZNF280B, C9orf131, EIF4E3, IGSF8, NCKAP1L, PTPLAD2, STOX1, ZNF280C, C9orf135, EIF4EBP1, IGSF9, NCKAP5, PTPLB, STOX2, ZNF280D, C9orf139, EIF4EBP2, IGSF9B, NCKAP5L, PTPMT1, STPG1, ZNF281, C9orf142, EIF4EBP3, IHH, NCKIPSD, PTPN1, STPG2, ZNF282, C9orf152, EIF4ENIF1, IK, NCL, PTPN11, STRA13, ZNF283, C9orf153, EIF4G1, IKBIP, NCLN, PTPN12, STRA6, ZNF284, C9orf156, EIF4G2, IKBKAP, NCMAP, PTPN13, STRA8, ZNF285, C9orf16, EIF4G3, IKBKB, NCOA1, PTPN14, STRADA, ZNF286A, C9orf163, EIF4H, IKBKE, NCOA2, PTPN18, STRADB, ZNF286B, C9orf169, EIF5, IKBKG, NCOA3, PTPN2, STRAP, ZNF287, C9orf170, EIF5A, IKZF1, NCOA4, PTPN20A, STRBP, ZNF292, C9orf171, EIF5A2, IKZF2, NCOA5, PTPN20B, STRC, ZNF296, C9orf173, EIF5AL1, IKZF3, NCOA6, PTPN21, STRIP1, ZNF3, C9orf24, EIF5B, IKZF4, NCOA7, PTPN22, STRIP2, ZNF30, C9orf3, EIF6, IKZF5, NCOR1, PTPN23, STRN, ZNF300, C9orf37, ELAC1, INTERLEUKIN, IL10, NCOR2, PTPN3, STRN3, ZNF302, C9orf40, ELAC2, IL10RA, NCR1, PTPN4, STRN4, ZNF304, C9orf41, ELANE, IL10RB, NCR2, PTPN5, STS, ZNF311, C9orf43, ELAVL1, IL11, NCR3, PTPN6, STT3A, ZNF316, C9orf47, ELAVL2, IL11RA, NCR3LG1, PTPN7, STT3B, ZNF317, C9orf50, ELAVL3, IL12A, NCS1, PTPN9, STUB1, ZNF318, C9orf57, ELAVL4, IL12B, NCSTN, PTPRA, STX10, ZNF319, C9orf62, ELF1, IL12RB1, NDC1, PTPRB, STX11, ZNF32, C9orf64, ELF2, IL12RB2, NDC80, PTPRC, STX12, ZNF320, C9orf66, ELF3, IL13, NDE1, PTPRCAP, STX16, ZNF322, C9orf69, ELF4, IL13RA1, NDEL1, PTPRD, STX17, ZNF324, C9orf72, ELF5, IL13RA2, NDFIP1, PTPRE, STX18, ZNF324B, C9orf78, ELFN1, IL15, NDFIP2, PTPRF, STX19, ZNF326, C9orf84, ELFN2, IL15RA, NDN, PTPRG, STX1A, ZNF329, C9orf85, ELK1, IL16, NDNF, PTPRH, STX1B, ZNF330, C9orf89, ELK3, IL17A, NDNL2, PTPRJ, STX2, ZNF331, C9orf9, ELK4, IL17B, NDOR1, PTPRK, STX3, ZNF333, C9orf91, ELL, IL17C, NDP, PTPRM, STX4, ZNF334, C9orf92, ELL2, IL17D, NDRG1, PTPRN, STX5, ZNF335, C9orf96, ELL3, IL17F, NDRG2, PTPRN2, STX6, ZNF337, CA1, ELMO1, IL17RA, NDRG3, PTPRO, STX7, ZNF33A, CA10, ELMO2, IL17RB, NDRG4, PTPRQ, STX8, ZNF33B, CA11, ELMO3, IL17RC, NDST1, PTPRR, STXBP1, ZNF34, CA12, ELMOD1, IL17RD, NDST2, PTPRS, STXBP2, ZNF341, CA13, ELMOD2, IL17RE, NDST3, PTPRT, STXBP3, ZNF343, CA14, ELMOD3, IL17REL, NDST4, PTPRU, STXBP4, ZNF345, CA2,

ELMSAN1, IL18, NDUFA1, PTPRZ1, STXBP5, ZNF346, CA3, ELN, IL18BP, NDUFA10, PTRF, STXBP5L, ZNF347, CA4, ELOF1, IL18R1, NDUFA11, PTRH1, STXBP6, ZNF35, CA5A, ELOVL1, IL18RAP, NDUFA12, PTRH2, STYK1, ZNF350, CA5B, ELOVL2, IL19, NDUFA13, PTRHD1, STYX, ZNF354A, CA6, ELOVL3, IL1A, NDUFA2, PTS, STYXL1, ZNF354B, CA7, ELOVL4, IL1B, NDUFA3, PTTG1, SUB1, ZNF354C, CA8, ELOVL5, IL1F10, NDUFA4, PTTG1IP, SUCLA2, ZNF358, CA9, ELOVL6, IL1R1, NDUFA4L2, PTTG2, SUCLG1, ZNF362, CAAP1, ELOVL7, IL1R2, NDUFA5, PTX3, SUCLG2, ZNF365, CAB39, ELP2, IL1RAP, NDUFA6, PTX4, SUCNR1, ZNF366, CAB39L, ELP3, IL1RAPL1, NDUFA7, PUF60, SUCO, ZNF367, CABIN1, ELP4, IL1RAPL2, NDUFA8, PUM1, SUDS3, ZNF37A, CABLES1, ELP5, IL1RL1, NDUFA9, PUM2, SUFU, ZNF382, CABLES2, ELP6, IL1RL2, NDUFAB1, PURA, SUGP1, ZNF383, CABP1, ELSPBP1, IL1RN, NDUFAF1, PURB, SUGP2, ZNF384, CABP2, ELTD1, IL2, NDUFAF2, PURG, SUGT1, ZNF385A, CABP4, EMB, IL20, NDUFAF3, PUS1, SULF1, ZNF385B, CABP5, EMC1, IL20RA, NDUFAF4, PUS10, SULF2, ZNF385C, CABP7, EMC10, IL20RB, NDUFAF5, PUS3, SULT1A1, ZNF385D, CABS1, EMC2, IL21, NDUFAF6, PUS7, SULT1A2, ZNF391, CABYR, EMC3, IL21R, NDUFAF7, PUS7L, SULT1A3, ZNF394, CACFD1, EMC4, IL22, NDUFB1, PUSL1, SULT1A4, ZNF395, CACHD1, EMC6, IL22RA1, NDUFB10, PVALB, SULT1B1, ZNF396, CACNA1A, EMC7, IL22RA2, NDUFB11, PVR, SULT1C2, ZNF397, CACNA1B, EMC8, IL23A, NDUFB2, PVRIG, SULT1C3, ZNF398, CACNA1C, EMC9, IL23R, NDUFB3, PVRL1, SULT1C4, ZNF404, CACNA1D, EMCN, IL24, NDUFB4, PVRL2, SULT1E1, ZNF407, CACNA1E, EMD, IL25, NDUFB5, PVRL3, SULT2A1, ZNF408, CACNA1F, EME1, IL26, NDUFB6, PVRL4, SULT2B1, ZNF41, CACNA1G, EME2, IL27, NDUFB7, PWP1, SULT4A1, ZNF410, CACNA1H, EMG1, IL27RA, NDUFB8, PWP2, SULT6B1, ZNF414, CACNA1I, EMID1, IL2RA, NDUFB9, PWWP2A, SUMF1, ZNF415, CACNA1S, EMILIN1, IL2RB, NDUFC1, PWWP2B, SUMF2, ZNF416, CACNA2D1, EMILIN2, IL2RG, NDUFC2, PXDC1, SUMO1, ZNF417, CACNA2D2, EMILIN3, IL3, NDUFC2-KCTD14, PXDN, SUMO2, ZNF418, CACNA2D3, EML1, IL31, NDUFS1, PXDNL, SUMO3, ZNF419, CACNA2D4, EML2, IL31RA, NDUFS2, PXK, SUMO4, ZNF420, CACNB1, EML3, IL32, NDUFS3, PXMP2, SUN1, ZNF423, CACNB2, EML4, IL33, NDUFS4, PXMP4, SUN2, ZNF425, CACNB3, EML5, IL34, NDUFS5, PXN, SUN3, ZNF426, CACNB4, EML6, IL36A, NDUFS6, PXT1, SUN5, ZNF428, CACNG1, EMP1, IL36B, NDUFS7, PYCARD, SUOX, ZNF429, CACNG2, EMP2, IL36G, NDUFS8, PYCR1, SUPT16H, ZNF43, CACNG3, EMP3, IL36RN, NDUFV1, PYCR2, SUPT20H, ZNF430, CACNG4, EMR1, IL37, NDUFV2, PYCRL, SUPT3H, ZNF431, CACNG5, EMR2, IL3RA, NDUFV3, PYDC1, SUPT4H1, ZNF432, CACNG6, EMR3, IL4, NEB, PYDC2, SUPT5H, ZNF433, CACNG7, EMX1, IL4I1, NEBL, PYGB, SUPT6H, ZNF436, CACNG8, EMX2, IL4R, NECAB1, PYGL, SUPT7L, ZNF438, CACTIN, EN1, IL5, NECAB2, PYGM, SUPV3L1, ZNF439, CACUL1, EN2, IL5RA, NECAB3, PYGO1, SURF1, ZNF44, CACYBP, ENAH, IL6, NECAP1, PYGO2, SURF2, ZNF440, CAD, ENAM, IL6R, NECAP2, PYHIN1, SURF4, ZNF441, CADM1, ENC1, IL6ST, NEDD1, PYROXD1, SURF6, ZNF442, CADM2, ENDOD1, IL7, NEDD4, PYROXD2, SUSD1, ZNF443, CADM3, ENDOG, IL7R,

NEDD4L, PYURF, SUSD2, ZNF444, CADM4, ENDOU, IL8, NEDD8, PYY, SUSD3, ZNF445, CADPS, ENDOV, IL9, NEDD8-MDP1, PZP, SUSD4, ZNF446, CADPS2, ENG, IL9R, NEDD9, QARS, SUSD5, ZNF449, CAGE1, ENGASE, ILDR1, NEFH, QDPR, SUV39H1, ZNF45, CALB1, ENHO, ILDR2, NEFL, QKI, SUV39H2, ZNF451, CALB2, ENKD1, ILF2, NEFM, QPCT, SUV420H1, ZNF454, CALCA, ENKUR, ILF3, NEGR1, QPCTL, SUV420H2, ZNF460, CALCB, ENO1, ILK, NEIL1, QPRT, SUZ12, ZNF461, CALCOCO1, ENO2, ILKAP, NEIL2, QRFP, SV2A, ZNF462, CALCOCO2, ENO3, ILVBL, NEIL3, QRFPR, SV2B, ZNF467, CALCR, ENO4, IMMP1L, NEK1, QRICH1, SV2C, ZNF468, CALCRL, ENOPH1, IMMP2L, NEK10, QRICH2, SVEP1, ZNF469, CALD1, ENOSF1, IMMT, NEK11, QRSL1, SVIL, ZNF470, CALHM1, ENOX1, IMP3, NEK2, QSER1, SVIP, ZNF471, CALHM2, ENOX2, IMP4, NEK3, QSOX1, SVOP, ZNF473, CALHM3, ENPEP, IMPA1, NEK4, QSOX2, SVOPL, ZNF474, CALM1, ENPP1, IMPA2, NEK5, QTRT1, SWAP70, ZNF479, CALM2, ENPP2, IMPACT, NEK6, QTRTD1, SWI5, ZNF48, CALM3, ENPP3, IMPAD1, NEK7, R3HCC1, SWSAP1, ZNF480, CALML3, ENPP4, IMPDH1, NEK8, R3HCC1L, SWT1, ZNF483, CALML4, ENPP5, IMPDH2, NEK9, R3HDM1, SYAP1, ZNF484, CALML5, ENPP6, IMPG1, NELFA, R3HDM2, SYBU, ZNF485, CALML6, ENPP7, IMPG2, NELFB, R3HDM4, SYCE1, ZNF486, CALN1, ENSA, INA, NELFCD, R3HDML, SYCE1L, ZNF488, CALR, ENTHD1, INADL, NELFE, RAB10, SYCE2, ZNF490, CALR3, ENTHD2, INCA1, NELL1, RAB11A, SYCE3, ZNF491, CALU, ENTPD1, INCENP, NELL2, RAB11B, SYCN, ZNF492, CALY, ENTPD2, INF2, NEMF, RAB11FIP1, SYCP1, ZNF493, CAMK1, ENTPD3, ING1, NENF, RAB11FIP2, SYCP2, ZNF496, CAMK1D, ENTPD4, ING2, NEO1, RAB11FIP3, SYCP2L, ZNF497, CAMK1G, ENTPD5, ING3, NES, RAB11FIP4, SYCP3, ZNF500, CAMK2A, ENTPD6, ING4, NET1, RAB11FIP5, SYDE1, ZNF501, CAMK2B, ENTPD7, ING5, NETO1, RAB12, SYDE2, ZNF502, CAMK2D, ENTPD8, INHA, NETO2, RAB13, SYF2, ZNF503, CAMK2G, ENY2, INHBA, NEU1, RAB14, SYK, ZNF506, CAMK2N1, EOGT, INHBB, NEU2, RAB15, SYMPK, ZNF507, CAMK2N2, EOMES, INHBC, NEU3, RAB17, SYN1, ZNF510, CAMK4, EP300, INHBE, NEU4, RAB18, SYN2, ZNF511, CAMKK1, EP400, INIP, NEURL1, RAB19, SYN3, ZNF512, CAMKK2, EPAS1, INMT, NEURL1B, RAB1A, SYNC, ZNF512B, CAMKMT, EPB41, INO80, NEURL2, RAB1B, SYNCRIP, ZNF513, CAMKV, EPB41L1, INO80B, NEURL3, RAB20, SYNDIG1, ZNF514, CAMLG, EPB41L2, INO80C, NEURL4, RAB21, SYNDIG1L, ZNF516, CAMP, EPB41L3, INO80D, NEUROD1, RAB22A, SYNE1, ZNF517, CAMSAP1, EPB41L4A, INO80E, NEUROD2, RAB23, SYNE2, ZNF518A, CAMSAP2, EPB41L4B, INPP1, NEUROD4, RAB24, SYNE3, ZNF518B, CAMSAP3, EPB41L5, INPP4A, NEUROD6, RAB25, SYNE4, ZNF519, CAMTA1, EPB42, INPP4B, NEUROG1, RAB26, SYNGAP1, ZNF521, CAMTA2, EPC1, INPP5A, NEUROG2, RAB27A, SYNGR1, ZNF524, CAND1, EPC2, INPP5B, NEUROG3, RAB27B, SYNGR2, ZNF526, CAND2, EPCAM, INPP5D, NEXN, RAB28, SYNGR3, ZNF527, CANT1, EPDR1, INPP5E, NF1, RAB2A, SYNGR4, ZNF528, CANX, EPG5, INPP5F, NF2, RAB2B, SYNJ1, ZNF529, CAP1, EPGN, INPP5J, NFAM1, RAB30, SYNJ2, ZNF530, CAP2, EPHA1, INPP5K, NFASC, RAB31, SYNJ2BP, ZNF532, CAPG, EPHA10, INPPL1, NFAT5, RAB32, SYNJ2BP-COX16, ZNF534, CAPN1, EPHA2,

INS, NFATC1, RAB33A, SYNM, ZNF536, CAPN10, EPHA3, INSC, NFATC2, RAB33B, SYNPO, ZNF540, CAPN11, EPHA4, INSIG1, NFATC2IP, RAB34, SYNPO2, ZNF541, CAPN12, EPHA5, INSIG2, NFATC3, RAB35, SYNPO2L, ZNF543, CAPN13, EPHA6, INS-IGF2, NFATC4, RAB36, SYNPR, ZNF544, CAPN14, EPHA7, INSL3, NFE2, RAB37, SYNRG, ZNF546, CAPN15, EPHA8, INSL4, NFE2L1, RAB38, SYP, ZNF547, CAPN2, EPHB1, INSL5, NFE2L2, RAB39A, SYPL1, ZNF548, CAPN3, EPHB2, INSL6, NFE2L3, RAB39B, SYPL2, ZNF549, CAPN5, EPHB3, INSM1, NFIA, RAB3A, SYS1, ZNF550, CAPN6, EPHB4, INSM2, NFIB, RAB3B, SYT1, ZNF551, CAPN7, EPHB6, INSR, NFIC, RAB3C, SYT10, ZNF552, CAPN8, EPHX1, INSRR, NFIL3, RAB3D, SYT11, ZNF554, CAPN9, EPHX2, INTS1, NFIX, RAB3GAP1, SYT12, ZNF555, CAPNS1, EPHX3, INTS10, NFKB1, RAB3GAP2, SYT13, ZNF556, CAPNS2, EPHX4, INTS12, NFKB2, RAB3IL1, SYT14, ZNF557, CAPRIN1, EPM2A, INTS2, NFKBIA, RAB3IP, SYT15, ZNF558, CAPRIN2, EPM2AIP1, INTS3, NFKBIB, RAB40A, SYT16, ZNF559, CAPS, EPN1, INTS4, NFKBID, RAB40AL, SYT17, ZNF559-ZNF177, CAPS2, EPN2, INTS5, NFKBIE, RAB40B, SYT2, ZNF560, CAPSL, EPN3, INTS6, NFKBIL1, RAB40C, SYT3, ZNF561, CAPZA1, EPO, INTS7, NFKBIZ, RAB41, SYT4, ZNF562, CAPZA2, EPOR, INTS8, NFRKB, RAB42, SYT5, ZNF563, CAPZA3, EPPIN, INTS9, NFS1, RAB43, SYT6, ZNF564, CAPZB, EPPIN-WFDC6, INTU, NFU1, RAB44, SYT7, ZNF565, CARD10, EPPK1, INVS, NFX1, RAB4A, SYT8, ZNF566, CARD11, EPRS, IP6K1, NFXL1, RAB4B, SYT9, ZNF567, CARD14, EPS15, IP6K2, NFYA, RAB5A, SYTL1, ZNF568, CARD16, EPS15L1, IP6K3, NFYB, RAB5B, SYTL2, ZNF569, CARD17, EPS8, IPCEF1, NFYC, RAB5C, SYTL3, ZNF57, CARD18, EPS8L1, IPMK, NGB, RAB6A, SYTL4, ZNF570, CARD6, EPS8L2, IPO11, NGDN, RAB6B, SYTL5, ZNF571, CARD8, EPS8L3, IPO13, NGEF, RAB6C, SYVN1, ZNF572, CARD9, EPSTI1, IPO4, NGF, RAB7A, SZT2, ZNF573, CARF, EPT1, IPO5, NGFR, RAB7L1, T, ZNF574, CARHSP1, EPX, IPO7, NGFRAP1, RAB8A, TAAR1, ZNF575, CARKD, EPYC, IPO8, NGLY1, RAB8B, TAAR2, ZNF576, CARM1, EQTN, IPO9, NGRN, RAB9A, TAAR5, ZNF577, CARNS1, ERAL1, IPP, NHEJ1, RAB9B, TAAR6, ZNF578, CARS, ERAP1, IPPK, NHLH1, RABAC1, TAAR8, ZNF579, CARS2, ERAP2, IQCA1, NHLH2, RABEP1, TAAR9, ZNF580, CARTPT, ERAS, IQCB1, NHLRC1, RABEP2, TAB1, ZNF581, CASC1, ERBB2, IQCC, NHLRC2, RABEPK, TAB2, ZNF582, CASC10, ERBB2IP, IQCD, NHLRC3, RABGAP1, TAB3, ZNF583, CASC3, ERBB3, IQCE, NHLRC4, RABGAP1L, TAC1, ZNF584, CASC4, ERBB4, IQCF1, NHP2, RABGEF1, TAC3, ZNF585A, CASC5, ERC1, IQCF2, NHP2L1, RABGGTA, TAC4, ZNF585B, CASD1, ERC2, IQCF3, NHS, RABGGTB, TACC1, ZNF586, CASK, ERCC1, IQCF5, NHSL1, RABIF, TACC2, ZNF587, CASKIN1, ERCC2, IQCF6, NHSL2, RABL2A, TACC3, ZNF587B, CASKIN2, ERCC3, IQCG, NICN1, RABL2B, TACO1, ZNF589, CASP1, ERCC4, IQCH, NID1, RABL3, TACR1, ZNF592, CASP10, ERCC5, IQCJ, NID2, RABL5, TACR2, ZNF593, CASP12, ERCC6, IQCJ-SCHIP1, NIF3L1, RABL6, TACR3, ZNF594, CASP14, ERCC6L, IQCK, NIFK, RAC1, TACSTD2, ZNF595, CASP16, ERCC6L2, IQGAP1, NIM1, RAC2, TADA1, ZNF596, CASP2, ERCC6-PGBD3, IQGAP2, NIN, RAC3, TADA2A, ZNF597, CASP3, ERCC8, IQGAP3, NINJ1, RACGAP1, TADA2B, ZNF598, CASP4, EREG, IQSEC1, NINJ2, RAD1, TADA3, ZNF599,

CASP5, ERF, IQSEC2, NINL, RAD17, TAF1, ZNF600, CASP6, ERG, IQSEC3, NIP7, RAD18, TAF10, ZNF605, CASP7, ERGIC1, IQUB, NIPA1, RAD21, TAF11, ZNF606, CASP8, ERGIC2, IRAK1, NIPA2, RAD21L1, TAF12, ZNF607, CASP8AP2, ERGIC3, IRAK1BP1, NIPAL1, RAD23A, TAF13, ZNF608, CASP9, ERH, IRAK2, NIPAL2, RAD23B, TAF15, ZNF609, CASQ1, ERI1, IRAK3, NIPAL3, RAD50, TAF1A, ZNF610, CASQ2, ERI2, IRAK4, NIPAL4, RAD51, TAF1B, ZNF611, CASR, ERI3, IREB2, NIPBL, RAD51AP1, TAF1C, ZNF613, CASS4, ERICH1, IRF1, NIPSNAP1, RAD51AP2, TAF1D, ZNF614, CAST, ERICH2, IRF2, NIPSNAP3A, RAD51B, TAF1L, ZNF615, CASZ1, ERLEC1, IRF2BP1, NIPSNAP3B, RAD51C, TAF2, ZNF616, CAT, ERLIN1, IRF2BP2, NISCH, RAD51D, TAF3, ZNF618, CATSPER1, ERLIN2, IRF2BPL, NIT1, RAD52, TAF4, ZNF619, CATSPER2, ERMAP, IRF3, NIT2, RAD54B, TAF4B, ZNF620, CATSPER3, ERMARD, IRF4, NKAIN1, RAD54L, TAF5, ZNF621, CATSPER4, ERMN, IRF5, NKAIN2, RAD54L2, TAF5L, ZNF622, CATSPERB, ERMP1, IRF6, NKAIN3, RAD9A, TAF6, ZNF623, CATSPERD, ERN1, IRF7, NKAIN4, RAD9B, TAF6L, ZNF624, CATSPERG, ERN2, IRF8, NKAP, RADIL, TAF7, ZNF625, CAV1, ERO1L, IRF9, NKAPL, RAE1, TAF7L, ZNF626, CAV2, ERO1LB, IRG1, NKD1, RAET1E, TAF8, ZNF627, CAV3, ERP27, IRGC, NKD2, RAET1G, TAF9, ZNF628, CBFA2T2, ERP29, IRGM, NKG7, RAET1L, TAF9B, ZNF629, CBFA2T3, ERP44, IRGQ, NKIRAS1, RAF1, TAGAP, ZNF630, CBFB, ERRFI1, IRS1, NKIRAS2, RAG1, TAGLN, ZNF638, CBL, ERVFRD-1, IRS2, NKPD1, RAG2, TAGLN2, ZNF639, CBLB, ERVMER34-1, IRS4, NKRF, RAI1, TAGLN3, ZNF641, CBLC, ERVV-1, IRX1, NKTR, RAI14, TAL1, ZNF644, CBLL1, ERVV-2, IRX2, NKX1-2, RAI2, TAL2, ZNF645, CBLN1, ESAM, IRX3, NKX2-1, RALA, TALDO1, ZNF646, CBLN2, ESCO1, IRX4, NKX2-2, RALB, TAMM41, ZNF648, CBLN3, ESCO2, IRX5, NKX2-3, RALBP1, TANC1, ZNF649, CBLN4, ESD, IRX6, NKX2-4, RALGAPA1, TANC2, ZNF652, CBR1, ESF1, ISCA1, NKX2-5, RALGAPA2, TANGO2, ZNF653, CBR3, ESM1, ISCA2, NKX2-6, RALGAPB, TANGO6, ZNF654, CBR4, ESPL1, ISCU, NKX2-8, RALGDS, TANK, ZNF655, CBS, ESPN, ISG15, NKX3- 1, RALGPS1, TAOK1, ZNF658, CBWD1, ESPNL, ISG20, NKX3-2, RALGPS2, TAOK2, ZNF660, CBWD2, ESR1, ISG20L2, NKX6-1, RALY, TAOK3, ZNF662, CBWD3, ESR2, ISL1, NKX6-2, RALYL, TAP1, ZNF664, CBWD5, ESRP1, ISL2, NKX6-3, RAMP1, TAP2, ZNF664-FAM101A, CBWD6, ESRP2, ISLR, NLE1, RAMP2, TAPBP, ZNF665, CBWD7, ESRRA, ISLR2, NLGN1, RAMP3, TAPBPL, ZNF667, CBX1, ESRRB, ISM1, NLGN2, RAN, TAPT1, ZNF668, CBX2, ESRRG, ISM2, NLGN3, RANBP1, TARBP1, ZNF669, CBX3, ESX1, ISOC1, NLGN4X, RANBP10, TARBP2, ZNF670, CBX4, ESYT1, ISOC2, NLGN4Y, RANBP17, TARDBP, ZNF671, CBX5, ESYT2, ISPD, NLK, RANBP2, TARM1, ZNF672, CBX6, ESYT3, IST1, NLN, RANBP3, TARP, ZNF674, CBX7, ETAA1, ISX, NLRC3, RANBP3L, TARS, ZNF675, CBX8, ETF1, ISY1, NLRC4, RANBP6, TARS2, ZNF676, CBY1, ETFA, ISY1-RAB43, NLRC5, RANBP9, TARSL2, ZNF677, CBY3, ETFB, ISYNA1, NLRP1, RANGAP1, TAS1R1, ZNF678, CC2D1A, ETFDH, ITCH, NLRP10, RANGRF, TAS1R2, ZNF679, CC2D1B, ETHE1, ITFG1, NLRP11, RAP1A, TAS1R3, ZNF680, CC2D2A, ETNK1, ITFG2, NLRP12, RAP1B, TAS2R1, ZNF681, CC2D2B, ETNK2, ITFG3, NLRP13, RAP1GAP, TAS2R10, ZNF682, CCAR1,

ETNPPL, ITGA1, NLRP14, RAP1GAP2, TAS2R13, ZNF683, CCAR2, ETS1, ITGA10, NLRP2, RAP1GDS1, TAS2R14, ZNF684, CCBE1, ETS2, ITGA11, NLRP3, RAP2A, TAS2R16, ZNF687, CCBL1, ETV1, ITGA2, NLRP4, RAP2B, TAS2R19, ZNF688, CCBL2, ETV2, ITGA2B, NLRP5, RAP2C, TAS2R20, ZNF689, CCDC101, ETV3, ITGA3, NLRP6, RAPGEF1, TAS2R3, ZNF69, CCDC102A, ETV3L, ITGA4, NLRP7, RAPGEF2, TAS2R30, ZNF691, CCDC102B, ETV4, ITGA5, NLRP8, RAPGEF3, TAS2R31, ZNF692, CCDC103, ETV5, ITGA6, NLRP9, RAPGEF4, TAS2R38, ZNF695, CCDC104, ETV6, ITGA7, NLRX1, RAPGEF5, TAS2R39, ZNF696, CCDC105, ETV7, ITGA8, NMB, RAPGEF6, TAS2R4, ZNF697, CCDC106, EVA1A, ITGA9, NMBR, RAPGEFL1, TAS2R40, ZNF699, CCDC107, EVA1B, ITGAD, NMD3, RAPH1, TAS2R41, ZNF7, CCDC108, EVA1C, ITGAE, NME1, RAPSN, TAS2R42, ZNF70, CCDC109B, EVC, ITGAL, NME1-NME2, RARA, TAS2R43, ZNF700, CCDC11, EVC2, ITGAM, NME2, RARB, TAS2R46, ZNF701, CCDC110, EVI2A, ITGAV, NME3, RARG, TAS2R5, ZNF703, CCDC112, EVI2B, ITGAX, NME4, RARRES1, TAS2R50, ZNF704, CCDC113, EVI5, ITGB1, NME5, RARRES2, TAS2R60, ZNF705A, CCDC114, EVI5L, ITGB1BP1, NME6, RARRES3, TAS2R7, ZNF705B, CCDC115, EVL, ITGB1BP2, NME7, RARS, TAS2R8, ZNF705D, CCDC116, EVPL, ITGB2, NME8, RARS2, TAS2R9, ZNF705E, CCDC117, EVPLL, ITGB3, NME9, RASA1, TASP1, ZNF705G, CCDC12, EVX1, ITGB3BP, NMI, RASA2, TAT, ZNF706, CCDC120, EVX2, ITGB4, NMNAT1, RASA3, TATDN1, ZNF707, CCDC121, EWSR1, ITGB5, NMNAT2, RASA4, TATDN2, ZNF708, CCDC122, EXD1, ITGB6, NMNAT3, RASA4B, TATDN3, ZNF709, CCDC124, EXD2, ITGB7, NMRAL1, RASAL1, TAX1BP1, ZNF71, CCDC125, EXD3, ITGB8, NMRK1, RASAL2, TAX1BP3, ZNF710, CCDC126, EXO1, ITGBL1, NMRK2, RASAL3, TAZ, ZNF711, CCDC127, EXO5, ITIH1, NMS, RASD1, TBATA, ZNF713, CCDC129, EXOC1, ITIH2, NMT1, RASD2, TBC1D1, ZNF714, CCDC13, EXOC2, ITIH3, NMT2, RASEF, TBC1D10A, ZNF716, CCDC130, EXOC3, ITIH4, NMU, RASGEF1A, TBC1D10B, ZNF717, CCDC132, EXOC3L1, ITIH5, NMUR1, RASGEF1B, TBC1D10C, ZNF718, CCDC134, EXOC3L2, ITIH6, NMUR2, RASGEF1C, TBC1D12, ZNF720, CCDC135, EXOC3L4, ITK, NNAT, RASGRF1, TBC1D13, ZNF721, CCDC136, EXOC4, ITLN1, NNMT, RASGRF2, TBC1D14, ZNF726, CCDC137, EXOC5, ITLN2, NNT, RASGRP1, TBC1D15, ZNF727, CCDC138, EXOC6, ITM2A, NOA1, RASGRP2, TBC1D16, ZNF728, CCDC14, EXOC6B, ITM2B, NOB1, RASGRP3, TBC1D17, ZNF729, CCDC140, EXOC7, ITM2C, NOBOX, RASGRP4, TBC1D19, ZNF730, CCDC141, EXOC8, ITPA, NOC2L, RASIP1, TBC1D2, ZNF732, CCDC142, EXOG, ITPK1, NOC3L, RASL10A, TBC1D20, ZNF735, CCDC144A, EXOSC1, ITPKA, NOC4L, RASL10B, TBC1D21, ZNF736, CCDC144NL, EXOSC10, ITPKB, NOD1, RASL11A, TBC1D22A, ZNF737, CCDC146, EXOSC2, ITPKC, NOD2, RASL11B, TBC1D22B, ZNF74, CCDC147, EXOSC3, ITPR1, NODAL, RASL12, TBC1D23, ZNF740, CCDC148, EXOSC4, ITPR2, NOG, RASSF1, TBC1D24, ZNF746, CCDC149, EXOSC5, ITPR3, NOL10, RASSF10, TBC1D25, ZNF747, CCDC15, EXOSC6, ITPRIP, NOL11, RASSF2, TBC1D26, ZNF749, CCDC150, EXOSC7, ITPRIPL1, NOL12, RASSF3, TBC1D27, ZNF750, CCDC151, EXOSC8, ITPRIPL2, NOL3, RASSF4, TBC1D28, ZNF75A, CCDC152, EXOSC9, ITSN1, NOL4, RASSF5, TBC1D29, ZNF75D,

CCDC153, EXPH5, ITSN2, NOL6, RASSF6, TBC1D2B, ZNF76, CCDC154, EXT1, IVD, NOL7, RASSF7, TBC1D3, ZNF761, CCDC155, EXT2, IVL, NOL8, RASSF8, TBC1D30, ZNF763, CCDC157, EXTL1, IVNS1ABP, NOL9, RASSF9, TBC1D31, ZNF764, CCDC158, EXTL2, IWS1, NOLC1, RAVER1, TBC1D32, ZNF765, CCDC159, EXTL3, IYD, NOM1, RAVER2, TBC1D3B, ZNF766, CCDC160, EYA1, IZUMO1, NOMO1, RAX, TBC1D3C, ZNF768, CCDC166, EYA2, IZUMO2, NOMO2, RAX2, TBC1D3F, ZNF77, CCDC167, EYA3, IZUMO3, NOMO3, RB1, TBC1D3G, ZNF770, CCDC168, EYA4, IZUMO4, NONO, RB1CC1, TBC1D3H, ZNF771, CCDC169, EYS, JADE1, NOP10, RBAK, TBC1D4, ZNF772, CCDC169-SOHLH2, EZH1, JADE2, NOP14, RBAK-RBAKDN, TBC1D5, ZNF773, CCDC17, EZH2, JADE3, NOP16, RBBP4, TBC1D7, ZNF774, CCDC170, EZR, JAG1, NOP2, RBBP5, TBC1D8, ZNF775, CCDC171, F10, JAG2, NOP56, RBBP6, TBC1D8B, ZNF776, CCDC172, F11, JAGN1, NOP58, RBBP7, TBC1D9, ZNF777, CCDC173, F11R, JAK1, NOP9, RBBP8, TBC1D9B, ZNF778, CCDC174, F12, JAK2, NOS1, RBBP8NL, TBCA, ZNF780A, CCDC175, F13A1, JAK3, NOS1AP, RBBP9, TBCB, ZNF780B, CCDC176, F13B, JAKMIP1, NOS2, RBCK1, TBCC, ZNF781, CCDC177, F2, JAKMIP2, NOS3, RBFA, TBCCD1, ZNF782, CCDC178, F2R, JAKMIP3, NOSIP, RBFOX1, TBCD, ZNF783, CCDC179, F2RL1, JAM2, NOSTRIN, RBFOX2, TBCE, ZNF784, CCDC18, F2RL2, JAM3, NOTCH1, RBFOX3, TBCEL, ZNF785, CCDC180, F2RL3, JARID2, NOTCH2, RBKS, TBCK, ZNF786, CCDC181, F3, JAZF1, NOTCH2NL, RBL1, TBK1, ZNF787, CCDC19, F5, JDP2, NOTCH3, RBL2, TBKBP1, ZNF789, CCDC22, F7, JHDM1D, NOTCH4, RBM10, TBL1X, ZNF79, CCDC23, F8, JKAMP, NOTO, RBM11, TBL1XR1, ZNF790, CCDC24, F8A1, JMJD1C, NOTUM, RBM12, TBL1Y, ZNF791, CCDC25, F8A2, JMJD4, NOV, RBM12B, TBL2, ZNF792, CCDC27, F8A3, JMJD6, NOVA1, RBM14, TBL3, ZNF793, CCDC28A, F9, JMJD7, NOVA2, RBM14-RBM4, TBP, ZNF799, CCDC28B, FA2H, JMJD7-PLA2G4B, NOX1, RBM15, TBPL1, ZNF8, CCDC3, FAAH, JMJD8, NOX3, RBM15B, TBPL2, ZNF80, CCDC30, FAAH2, JMY, NOX4, RBM17, TBR1, ZNF800, CCDC33, FABP1, JOSD1, NOX5, RBM18, TBRG1, ZNF804A, CCDC34, FABP12, JOSD2, NOXA1, RBM19, TBRG4, ZNF804B, CCDC36, FABP2, JPH1, NOXO1, RBM20, TBX1, ZNF805, CCDC37, FABP3, JPH2, NOXRED1, RBM22, TBX10, ZNF806, CCDC38, FABP4, JPH3, NPAP1, RBM23, TBX15, ZNF808, CCDC39, FABP5, JPH4, NPAS1, RBM24, TBX18, ZNF81, CCDC40, FABP6, JRK, NPAS2, RBM25, TBX19, ZNF812, CCDC41, FABP7, JRKL, NPAS3, RBM26, TBX2, ZNF813, CCDC42, FABP9, JSRP1, NPAS4, RBM27, TBX20, ZNF814, CCDC42B, FADD, JTB, NPAT, RBM28, TBX21, ZNF816, CCDC43, FADS1, JUN, NPB, RBM3, TBX22, ZNF816-ZNF321P, CCDC47, FADS2, JUNB, NPBWR1, RBM33, TBX3, ZNF821, CCDC50, FADS3, JUND, NPBWR2, RBM34, TBX4, ZNF823, CCDC51, FADS6, JUP, NPC1, RBM38, TBX5, ZNF827, CCDC53, FAF1, KAAG1, NPC1L1, RBM39, TBX6, ZNF829, CCDC54, FAF2, KAL1, NPC2, RBM4, TBXA2R, ZNF83, CCDC57, FAH, KALRN, NPDC1, RBM41, TBXAS1, ZNF830, CCDC58, FAHD1, KANK1, NPEPL1, RBM42, TC2N, ZNF831, CCDC59, FAHD2A, KANK2, NPEPPS, RBM43, TCAIM, ZNF835, CCDC6, FAHD2B, KANK3, NPFF, RBM46, TCAP, ZNF836, CCDC60, FAIM, KANK4, NPFFR1, RBM47, TCEA1, ZNF837, CCDC61, FAIM2,

KANSL1, NPFFR2, RBM48, TCEA2, ZNF839, CCDC62, FAIM3, KANSL1L, NPHP1, RBM4B, TCEA3, ZNF84, CCDC63, FAM101A, KANSL2, NPHP3, RBM5, TCEAL1, ZNF841, CCDC64, FAM101B, KANSL3, NPHP4, RBM6, TCEAL2, ZNF843, CCDC64B, FAM102A, KARS, NPHS1, RBM7, TCEAL3, ZNF844, CCDC65, FAM102B, KAT2A, NPHS2, RBM8A, TCEAL4, ZNF845, CCDC66, FAM103A1, KAT2B, NPIPA1, RBMS1, TCEAL5, ZNF846, CCDC67, FAM104A, KAT5, NPIPA2, RBMS2, TCEAL6, ZNF85, CCDC68, FAM104B, KAT6A, NPIPA3, RBMS3, TCEAL7, ZNF850, CCDC69, FAM105A, KAT6B, NPIPA5, RBMX, TCEAL8, ZNF852, CCDC7, FAM105B, KAT7, NPIPA7, RBMX2, TCEANC, ZNF853, CCDC70, FAM107A, KAT8, NPIPA8, RBMXL1, TCEANC2, ZNF860, CCDC71, FAM107B, KATNA1, NPIPB11, RBMXL2, TCEB1, ZNF862, CCDC71L, FAM109A, KATNAL1, NPIPB15, RBMXL3, TCEB2, ZNF865, CCDC73, FAM109B, KATNAL2, NPIPB3, RBMY1A1, TCEB3, ZNF878, CCDC74A, FAM110A, KATNB1, NPIPB4, RBMY1B, TCEB3B, ZNF879, CCDC74B, FAM110B, KATNBL1, NPIPB5, RBMY1D, TCEB3C, ZNF880, CCDC77, FAM110C, KAZALD1, NPIPB6, RBMY1E, TCEB3CL, ZNF883, CCDC78, FAM110D, KAZN, NPIPB8, RBMY1F, TCEB3CL2, ZNF888, CCDC79, FAM111A, KBTBD11, NPIPB9, RBMY1J, TCERG1, ZNF891, CCDC8, FAM111B, KBTBD12, NPL, RBP1, TCERG1L, ZNF90, CCDC80, FAM114A1, KBTBD13, NPLOC4, RBP2, TCF12, ZNF91, CCDC81, FAM114A2, KBTBD2, NPM1, RBP3, TCF15, ZNF92, CCDC82, FAM115A, KBTBD3, NPM2, RBP4, TCF19, ZNF93, CCDC83, FAM115C, KBTBD4, NPM3, RBP5, TCF20, ZNF98, CCDC84, FAM117A, KBTBD6, NPNT, RBP7, TCF21, ZNF99, CCDC85A, FAM117B, KBTBD7, NPPA, RBPJ, TCF23, ZNFX1, CCDC85B, FAM118A, KBTBD8, NPPB, RBPJL, TCF24, ZNHIT1, CCDC85C, FAM118B, KCMF1, NPPC, RBPMS, TCF25, ZNHIT2, CCDC86, FAM120A, KCNA1, NPR1, RBPMS2, TCF3, ZNHIT3, CCDC87, FAM120AOS, KCNA10, NPR2, RBX1, TCF4, ZNHIT6, CCDC88A, FAM120B, KCNA2, NPR3, RC3H1, TCF7, ZNRD1, CCDC88B, FAM120C, KCNA3, NPRL2, RC3H2, TCF7L1, ZNRF1, CCDC88C, FAM122A, KCNA4, NPRL3, RCAN1, TCF7L2, ZNRF2, CCDC89, FAM122B, KCNA5, NPS, RCAN2, TCFL5, ZNRF3, CCDC9, FAM122C, KCNA6, NPSR1, RCAN3, TCHH, ZNRF4, CCDC90B, FAM124A, KCNA7, NPTN, RCBTB1, TCHHL1, ZP1, CCDC91, FAM124B, KCNAB1, NPTX1, RCBTB2, TCHP, ZP2, CCDC92, FAM126A, KCNAB2, NPTX2, RCC1, TCIRG1, ZP3, CCDC93, FAM126B, KCNAB3, NPTXR, RCC2, TCL1A, ZP4, CCDC94, FAM127A, KCNB1, NPVF, RCCD1, TCL1B, ZPBP, CCDC96, FAM127B, KCNB2, NPW, RCE1, TCN1, ZPBP2, CCDC97, FAM127C, KCNC1, NPY, RCHY1, TCN2, ZPLD1, CCER1, FAM129A, KCNC2, NPY1R, RCL1, TCOF1, ZRANB1, CCHCR1, FAM129B, KCNC3, NPY2R, RCN1, TCP1, ZRANB2, CCIN, FAM129C, KCNC4, NPY4R, RCN2, TCP10, ZRANB3, CCK, FAM131A, KCND1, NPY5R, RCN3, TCP10L, ZRSR2, CCKAR, FAM131B, KCND2, NQO1, RCOR1, TCP10L2, ZSCAN1, CCKBR, FAM131C, KCND3, NQO2, RCOR2, TCP11, ZSCAN10, CCL1, FAM132A, KCNE1, NR0B1, RCOR3, TCP11L1, ZSCAN12, CCL11, FAM132B, KCNE1L, NR0B2, RCSD1, TCP11L2, ZSCAN16, CCL13, FAM133A, KCNE2, NR1D1, RCVRN, TCP11X1, ZSCAN18, CCL14, FAM133B, KCNE3, NR1D2, RD3, TCP11X2, ZSCAN2, CCL15, FAM134A, KCNE4, NR1H2, RD3L, TCTA, ZSCAN20, CCL16,

FAM134B, KCNF1, NR1H3, RDH10, TCTE1, ZSCAN21, CCL17, FAM134C, KCNG1, NR1H4, RDH11, TCTE3, ZSCAN22, CCL18, FAM135A, KCNG2, NR1I2, RDH12, TCTEX1D1, ZSCAN23, CCL19, FAM135B, KCNG3, NR1I3, RDH13, TCTEX1D2, ZSCAN25, CCL2, FAM136A, KCNG4, NR2C1, RDH14, TCTEX1D4, ZSCAN26, CCL20, FAM13A, KCNH1, NR2C2, RDH16, TCTN1, ZSCAN29, CCL21, FAM13B, KCNH2, NR2C2AP, RDH5, TCTN2, ZSCAN30, CCL22, FAM13C, KCNH3, NR2E1, RDH8, TCTN3, ZSCAN31, CCL23, FAM149A, KCNH4, NR2E3, RDM1, TDG, ZSCAN32, CCL24, FAM149B1, KCNH5, NR2F1, RDX, TDGF1, ZSCAN4, CCL25, FAM150A, KCNH6, NR2F2, REC8, TDO2, ZSCAN5A, CCL26, FAM150B, KCNH7, NR2F6, RECK, TDP1, ZSCAN5B, CCL27, FAM151A, KCNH8, NR3C1, RECQL, TDP2, ZSCAN9, CCL28, FAM151B, KCNIP1, NR3C2, RECQL4, TDRD1, ZSWIM1, CCL3, FAM153A, KCNIP2, NR4A1, RECQL5, TDRD10, ZSWIM2, CCL3L1, FAM153B, KCNIP3, NR4A2, REEP1, TDRD12, ZSWIM3, CCL3L3, FAM154A, KCNIP4, NR4A3, REEP2, TDRD3, ZSWIM4, CCL4, FAM154B, KCNJ1, NR5A1, REEP3, TDRD5, ZSWIM5, CCL4L1, FAM155A, KCNJ10, NR5A2, REEP4, TDRD6, ZSWIM6, CCL4L2, FAM155B, KCNJ11, NR6A1, REEP5, TDRD7, ZSWIM7, CCL5, FAM156A, KCNJ12, NRAP, REEP6, TDRD9, ZSWIM8, CCL7, FAM156B, KCNJ13, NRARP, REG1A, TDRKH, ZUFSP, CCL8, FAM157A, KCNJ14, NRAS, REG1B, TDRP, ZW10, CCM2, FAM157B, KCNJ15, NRBF2, REG3A, TEAD1, ZWILCH, CCM2L, FAM159A, KCNJ16, NRBP1, REG3G, TEAD2, ZWINT, CCNA1, FAM159B, KCNJ18, NRBP2, REG4, TEAD3, ZXDA, CCNA2, FAM160A1, KCNJ2, NRCAM, REL, TEAD4, ZXDB, CCNB1, FAM160A2, KCNJ3, NRD1, RELA, TEC, ZXDC, CCNB1IP1, FAM160B1, KCNJ4, NRDE2, RELB, TECPR1, ZYG11A, CCNB2, FAM160B2, KCNJ5, NREP, RELL1, TECPR2, ZYG11B, CCNB3, FAM161A, KCNJ6, NRF1, RELL2, TECR, ZYX, CCNC, FAM161B, KCNJ8, NRG1, RELN, TECRL, ZZEF1, CCND1, FAM162A, KCNJ9, NRG2, RELT, TECTA, ZZZ3, CCND2, FAM162B, KCNK1, NRG, RHODOPSIN, RdCVF, RdCVFL, GIRK, DUX4, 和 DBET (或 DBET lncRNA).

表2.天然存在的Cas12f蛋白的列表

Cas 蛋白	SEQ ID NO	氨基酸序列
Un2Cas12f1	2	本文提供的(SEQ ID NO: 2)
AsCas12f	3	本文提供的(SEQ ID NO: 3)

Mi1Cas12f2	13	MNMSKTTISVKLKIIDLSSEKKEFLDNYFNEYAKATTFCQLRIRRLLRN THWLGKKEKSSKKWIFESGICDLCGENKELVNEDRNSGEPAKICKRCY NGRYGNQMIRKLFVSTKKREVQENMDIRRVAKLNNTHYHRIPEEAFD MIKAADTAEKRRKKNVEYDKKRQMEFIEMFNDEKKRAARPKKPNERE TRYVHISKLESPSKGYTLNGIKRKIDGMGKKIERAEKGLSRKKIFGYQG NRIKLDSNWVRFDLAESEITIPSLFKEMKLRITGPTNVHSKSGQIYFAEW FERINKQPNNYCYLIRKTSSNGKYEYYLQYTYEAEVEANKEYAGCLGV DIGCSKLAAAVYYDSKNKKAQKPIEIFTNPIKKIKMRREKLIKLLSRVK VRHRRRKLMQLSKTEPIIDYTCHKTARKIVEMANTAKAFISMENLETGI KQKQQARETKKQKFYRNMFLFRKLSKLIEYKALLKGIKIVYVKPDYTS QTCSSCGADKEKTERPSQAIFRCLNPTCRYYQRDINADFNAAVNIAKK ALNNTEVVTTLL
Mi2Cas12f2	14	MPSETYITKTLSLKLIPSDEEKQALENYFITFQRAVNFAIDRIVDIRSSFR YLNKNEQFPAVCDCCGKKEKIMYVNISNKTFKFKPSRNQKDRYTKDIY TIKPNAHICKTCYSGVAGNMFIRKQMYPNDKEGWKVSRSYNIKVNAP GLTGTEYAMAIRKAISILRSFEKRRRNAERRIIEYEKSKKEYLELIDDVE KGKTNKIVVLEKEGHQRVKRYKHKNWPEKWQGISLNKAKSKVKDIE KRIKKLKEWKHPTLNRPYVELHKNNVRIVGYETVELKLGNKMYTIHF ASISNLRKPFRKQKKKSIEYLKHLLTLALKRNLETYPSIIKRGKNFFLQY PVRVTVKVPKLTKNFKAFGIDRGVNRLAVGCIISKDGKLTNKNIFFFHG KEAWAKENRYKKIRDRLYAMAKKLRGDKTKKIRLYHEIRKKFRHKV KYFRRNYLHNISKQIVEIAKENTPTVIVLEDLRYLRERTYRGKGRSKKA KKTNYKLNTFTYRMLIDMIKYKAEEAGVPVMIIDPRNTSRKCSKCGYV DENNRKQASFKCLKCGYSLNADLNAAVNIAKAFYECPTFRWEEKLHA YVCSEPDK

AuCas12f2	15	MKSFKLKLLPTDEQNVLLNEVFCKWASLCTRMASKGHDKERLAPPDS SGNYFNKTQLNQVNTDVTDHMGALEESASQKERAVEKVKRRLKLISD MLSEPNLRDVSQQKPTTFRPLEWVKEGLLKTKYHTVHYWQKECDKLT KQKERMEKTIEKIKKGKITFKPTKMSLHQNCFSLSFGKGTFSMRPFSDT KRGINLDMLTAPIQPAIGKNDGKSSLRSKEFIARNIENYIIFSIHSQLFGLS RSEELLLNAKKEELVAKRDAMLKKKSDSLSKKIKELEKIVGRKITDSER SEIMSQGGKLSSEKFSEDNSYLKTLKVLAKDIIGREELFRLKKYPIVIRK PLNERKKLKNLKPDEWEYYLQLSYDELEKKEFTPKTIMGIDRGLKHIL AIAIYDPVQNKFVKNMLIPNPILGWKWKLRKIKRSIQHMERRIRAQQN AHVPENQLKKRLKSIENKIDYYYHNVSRQILNLAHDFKSAIVVEDLQN MKQHGRKKSKGLRGLNYALSNFDYGKIMGLVKYKAESENVPLLTVLP AGTSQNCAYCLLYGKEQGNYVRNNVNSKIGKCKLHGEIDADINAARTI AICYHKNINEPKPYGERKTFKRK
PtCas12f1	16	MKYTKVMRYQIIKPLNAEWDELGMVLRDIQKETRAALNKTIQLCWEY QGFSADYKQIHGQYPKPKDVLGYTSMHGYAYDRLKNEFSKIASSNLSQ TIKRAVDKWNSDLKEILRGDRSIPNFRKDCPIDIVKQSTKIQKCNDGYV LSLGLINREYKNELGRKNGVFDVLIKANDKTQQTILERIINGDYTYTAS QIINHKNKWFINLTYQFETKETALDPNNVMGVDLGIVYPVYIAFNNSL HRYHIKGGEIERFRRQVEKRKRELLNQGKYCGDGRKGHGYATRTKSIE SISDKIARFRDTCNHKYSRFIVDMALKHNCGIIQMEDLTGISKESTFLKN WTYYDLQQKIEYKAREAGIQVIKIEPQYTSQRCSKCGYIDKENRQEQA TFKCIECGFKTNADYNAARNIAIPNIDKIIRKTLKMQ
RuCas12f1	17	MTLLVKVVKIHLISEQFDKAGNRIDYEEVNKILWELQKQTREAKNKTV QLLWEWNNFSSDYVKASGIYPKAKDIFGYSSVHGQANKELRTKLALN SSNLSTTTMDVCKNFNTYKKEVWKGKRSVPSYKSDQPLDLHKDSIKLI YENNEFYVRLALLKKAEFAKYGFKDGFRFKMQVKDNSTKTILERCFD EVYKINASKLLYDQKKKKWKLNLSYSFDNKNISELDKEKILGVDVGV NCPLVASVFGDRDRFIIKGGEIEKFRKSVEARRRSMLEQTKYCGDGRIG HGRKKRTEPALNIGDKIARFRDTTNHKYSRALIEYAVKKGCGTIQMEK LTGITSKSDRFLKDWTYYDLQTKIENKAKEVGINVVYIAPKYTSQRCSK CGYIHKDNRPNQAKFRCLECDFESNADYNASQNIGIKNIDKIIEKDLQK QESEVQVNENK

SpCas12f1	18	MGESVKAIKLKILDMFLDPECTKQDDNWRKDLSTMSRFCAEAGNMCL RDLYNYFSMPKEDRISSKDLYNAMYHKTKLLHPELPGKVANQIVNHA KDVWKRNAKLIYRNQISMPTYKITTAPIRLQNNIYKLIKNKNKYIIDVQ LYSKEYSKDSGKGTHRYFLVAVRDSSTRMIFDRIMSKDHIDSSKSYTQ GQLQIKKDHQGKWYCIIPYTFPTHETVLDPDKVMGVDLGVAKAVYW AFNSSYKRGCIDGGEIEHFRKMIRARRVSIQNQIKHSGDARKGHGRKR ALKPIETLSEKEKNFRDTINHRYANRIVEAAIKQGCGTIQIENLEGIADT TGSKFLKNWPYYDLQTKIVNKAKEHGITVVAINPQYTSQRCSMCGYIE KTNRSSQAVFECKQCGYGSRTICINCRHVQVSGDVCEECGGIVKKENV NADYNAAKNISTPYIDQIIMEKCLELGIPYRSITCKECGHIQASGNTCEV CGSTNILKPKKIRKAK
CnCas12f1	19	MITVRKIKLTIMGDKDTRNSQYKWIRDEQYNQYRALNMGMTYLAVN DILYMNESGLEIRTIKDLKDCEKDIDKNKKEIEKLTARLEKEQNKKNSS SEKLDEIKYKISLVENKIEDYKLKIVELNKILEETQKERMDIQKEFKEKY VDDLYQVLDKIPFKHLDNKSLVTQRIKADIKSDKSNGLLKGERSIRNYK RNFPLMTRGRDLKFKYDDNDDIEIKWMEGIKFKVILGNRIKNSLELRHT LHKVIEGKYKICDSSLQFDKNNNLILNLTLDIPIDIVNKKVSGRVVGVD LGLKIPAYCALNDVEYIKKSIGRIDDFLKVRTQMQSRRRRLQIAIQSAK GGKGRVNKLQALERFAEKEKNFAKTYNHFLSSNIVKFAVSNQAEQIN MELLSLKETQNKSILRNWSYYQLQTMIEYKAQREGIKVKYIDPYHTSQ TCSKCGNYEEGQRESQADFICKKCGYKVNADYNAARNIAMSNKYITK KEESKYYKIKESMV

表3A.工程化核酸酶变体例如dCasMINI和嵌合工程化核酸酶变体的列表。星号的数量代表工程化核酸酶变体的基因调节活性的相对程度。

工程化核酸酶变体	尺寸(氨基酸)	IFNγ 的激活	CD2 的激活	嵌合体设计
dCasMINI	529	**	**	SEQ ID NO: 10
cA1	495	-	-	将 dCasMINI 的 K31-L105 置换为 Un2Cas12f1 的 K32-L72
cA2	495	***	*	将 dCasMINI 的 K31-L77 置换为 Un2Cas12f1 的 K32-L44
cA3	496	-	-	将 dCasMINI 的 M1-L77 置换为 Un2Cas12f1 的 M1-L44

cA4	495	*	-	将 dCasMINI 的 P17-L77 置换为 Un2Cas12f1 的 P18-L44
cA5	451	-	-	将 dCasMINI 的 M1-W95 置换为 AsCas12f1 的 M1-W17
cA6	485	-	-	将 dCasMINI 的 M1-D91 置换为 CnCas12f1 的 M1-D47
cA7	462	-	-	将 dCasMINI 的 M1-W95 置换为 SpCas12f1 的 M1-W28
cA8	453	-	-	将 dCasMINI 的 M1-W95 置换为 PtCas12f1 的 M1-W19
cA9	460	-	-	将 dCasMINI 的 M1-E97 置换为 RuCas12f1 的 M1-E28

表3B.来自表3A的嵌合工程化核酸酶变体的氨基酸序列。

工程化核酸酶变体	SEQ ID NO	氨基酸序列
cA1	20	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKERRKQAGGTGELDGGFYKKL EKKHSEMFSFDRLNLLLNQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVE HYLSRVCYRRAAELFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDN FPIPLVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQ VQKSPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRG SKICEKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSIS DNDLFHFNKKMFARRRILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKK LIERWACEIADFFIKNKVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQ NKIEFKLKQYGIEIRKVAPNNTSKTCSKCGHLNNYFNFEYRKKNKFPHFKCE KCNFKENAAYNAALNISNPKLKSTKERP

cA3	21	MEVQKTVMKTLSLRILRPLYSQEIEKEIKEEKERRKQAGGTGELDDKFYQK LRGQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSV EHYLSRVCYRRAAELFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSD NFPIPLVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFE QVQKSPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKR GSKICEKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYS ISDNDLFHFNKKMFARRRILLKKNRHKRAGHGAKNKLKPITILTEKSERFRK KLIERWACEIADFFIKNKVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEM QNKIEFKLKQYGIEIRKVAPNNTSKTCSKCGHLNNYFNFEYRKKNKFPHFK CEKCNFKENAAYNAALNISNPKLKSTKERP
cA4	22	MAKNTITKTLKLRIVRPLYSQEIEKEIKEEKERRKQAGGTGELDDKFYQKLR GQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEH YLSRVCYRRAAELFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFP IPLVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQV QKSPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGS KICEKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSIS DNDLFHFNKKMFARRRILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKK LIERWACEIADFFIKNKVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQ NKIEFKLKQYGIEIRKVAPNNTSKTCSKCGHLNNYFNFEYRKKNKFPHFKCE KCNFKENAAYNAALNISNPKLKSTKERP
cA5	23	MIKVYRYEIVKPLDLDWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIA NASSVEHYLSRVCYRRAAELFKNAAIASGLRSKIKSNFRLKELKNMKSGLP TTKSDNFPIPLVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWE KFDFEQVQKSPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSY IEVKRGSKICEKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNA FSRYSISDNDLFHFNKKMFARRRILLKKNRHKRAGHGAKNKLKPITILTEKS ERFRKKLIERWACEIADFFIKNKVGTVQMENLESMKRKEDSYFNIRLRGFW PYAEMQNKIEFKLKQYGIEIRKVAPNNTSKTCSKCGHLNNYFNFEYRKKNK FPHFKCEKCNFKENAAYNAALNISNPKLKSTKERP
cA6	24	MITVRKIKLTIMGDKDTRNSQYKWIRDEQYNQYRALNMGMTYLAVNDAV FWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCY RRAAELFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQK GGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQKSPKPIS LLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKICEKSA WMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISDNDLFHF NKKMFARRRILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLIERWAC EIADFFIKNKVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQNKIEFKL KQYGIEIRKVAPNNTSKTCSKCGHLNNYFNFEYRKKNKFPHFKCEKCNFKE NAAYNAALNISNPKLKSTKERP

cA7	25	MGESVKAIKLKILDMFLDPECTKQDDNWQEISEIFRQLQKQAAEIYNQSLIE LYYEIFIKGKGIANASSVEHYLSRVCYRRAAELFKNAAIASGLRSKIKSNFRL KELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGFEISNHNSDFIIKIPFGRWQ VKKEIDKYRPWEKFDFEQVQKSPKPISLLLSTQRRKRNKGWSKDEGTEAEI KKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSIDVPKIDKGVDPSIIGGIAV GVRSPLVCAINNAFSRYSISDNDLFHFNKKMFARRRILLKKNRHKRAGHGA KNKLKPITILTEKSERFRKKLIERWACEIADFFIKNKVGTVQMENLESMKRK EDSYFNIRLRGFWPYAEMQNKIEFKLKQYGIEIRKVAPNNTSKTCSKCGHL NNYFNFEYRKKNKFPHFKCEKCNFKENAAYNAALNISNPKLKSTKERP
cA8	26	MKYTKVMRYQIIKPLNAEWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGK GIANASSVEHYLSRVCYRRAAELFKNAAIASGLRSKIKSNFRLKELKNMKS GLPTTKSDNFPIPLVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYR PWEKFDFEQVQKSPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDY QTSYIEVKRGSKICEKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCA INNAFSRYSISDNDLFHFNKKMFARRRILLKKNRHKRAGHGAKNKLKPITIL TEKSERFRKKLIERWACEIADFFIKNKVGTVQMENLESMKRKEDSYFNIRLR GFWPYAEMQNKIEFKLKQYGIEIRKVAPNNTSKTCSKCGHLNNYFNFEYRK KNKFPHFKCEKCNFKENAAYNAALNISNPKLKSTKERP
cA9	27	MTLLVKVVKIHLISEQFDKAGNRIDYEEISEIFRQLQKQAAEIYNQSLIELYY EIFIKGKGIANASSVEHYLSRVCYRRAAELFKNAAIASGLRSKIKSNFRLKEL KNMKSGLPTTKSDNFPIPLVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKK EIDKYRPWEKFDFEQVQKSPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKV MNGDYQTSYIEVKRGSKICEKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVR SPLVCAINNAFSRYSISDNDLFHFNKKMFARRRILLKKNRHKRAGHGAKNK LKPITILTEKSERFRKKLIERWACEIADFFIKNKVGTVQMENLESMKRKEDS YFNIRLRGFWPYAEMQNKIEFKLKQYGIEIRKVAPNNTSKTCSKCGHLNNY FNFEYRKKNKFPHFKCEKCNFKENAAYNAALNISNPKLKSTKERP

表4A.工程化核酸酶变体的列表。星号的数量代表工程化核酸酶变体的基因调节活性的相对程度。

工程化核酸酶变体	尺寸(氨基酸)	IFNγ 的激活	CD2 的激活	eGFP 的阻抑（A型阻抑因子）	eGFP 的阻抑（B型阻抑因子）
dCasMINI	529	**	**	**	**
cA2	495	***	*	****	**
mA8	529	****	**	**	-
mA10	529	-	**	*	-
mA11	529	*	*	**	-
mA12	529	*	*	*	*

mA14	529	*	*	*	*
mB9	529	-	-	*	-
mC7	529	**	*	**	（无数据）
mC16	529	-	*	*	*
mC18	529	***	**	（无数据）	***
mC21	529	-	*	***	*
mD2	529	****	****	****	（无数据）
mD4	529	****	**	***	***
mD5	529	-	*	***	-
mD6	529	-	*	**	-
mD7	529	-	**	***	（无数据）
mD15	529	-	*	*	**

表4B.来自表4A的嵌合工程化核酸酶变体的氨基酸序列

工程化核酸酶变体	SEQ ID NO	氨基酸序列
mA1	28	MAKNTITKTLKLRIVRPYYSQEIEKIVAEEKNRREKIALEKNKDKVKEACSK HLKVAAYCTTQVERNACLFCKARKLDDKFYQKLRGQFPDAVFWQEISEIFR QLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCYRRAAELFKN AAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGFEIS NHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQKSPKPISLLLSTQRRKR NKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSIDVP KIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISDNDLFHFNKKMFARRRI LLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLIERWACEIADFFIKNKV GTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQNKIEFKLKQYGIEIRKVA PNNTSKTCSKCGHLNNYFNFEYRKKNKFPHFKCEKCNFKENAAYNAALNIS NPKLKSTKERP
mA2	29	MAKNTITKTLKLRIVRPYYSAEVEKIVAEEKNNREKIALDKNKDKVKEACS KHLKVAAYCTTQVERNACLFCKARKLDDKFYKKLRGQFPDAVFWQEISEI FRQLQKQAREIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCYRRAAELFK NAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGFE ISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQKSPKPISLLLSTQRR KRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSID VPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISDNDLFHFNKKMFARR RILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLIERWACEIADFFIKNK VGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQNKIEFKLKQYGIEIRKV APNNTSKTCSKCGHLNNYFNFEYRKKNKFPHFKCEKCNFKENAAYNAALN ISNPKLKSTKERP

mA3	30	MAKNTITKTLKLRIVRPYYSAEIEKIVADEKNRREKIALEKNKDKVKEACSK HLKVAAYCTTQVERNACLFCKARKLDDKFYQKLRKQFPDAVFWQEISEIFR QLQKQAREIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCYRRAAELFKN AAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGFEIS NHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQKSPKPISLLLSTQRRKR NKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSIDVP KIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISDNDLFHFNKKMFARRRI LLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLIERWACEIADFFIKNKV GTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQNKIEFKLKQYGIEIRKVA PNNTSKTCSKCGHLNNYFNFEYRKKNKFPHFKCEKCNFKENAAYNAALNIS NPKLKSTKERP
mA4	31	MAKNTITKTLKLRIVRPYNSQEVEKIVAEEKNRREKIALDKNKDKVKEACS KHLKVAAYCTTQVERNACLFCKARKLDDKFYQKLRKQFPDAVFWQEISEI FRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCYRRAAELF KNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGF EISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQKSPKPISLLLSTQR RKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSI DVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISDNDLFHFNKKMFAR RRILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLIERWACEIADFFIKN KVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQNKIEFKLKQYGIEIRK VAPNNTSKTCSKCGHLNNYFNFEYRKKNKFPHFKCEKCNFKENAAYNAAL NISNPKLKSTKERP
mA5	32	MAKNTITKTLKLRIVRPYNSQEVEKIVAEEKNNREKIALDKNKDKVKEACS KHLKVAAYCTTQVERNACLFCKARKLDDKFYKKLRKQFPDAVFWQEISEI FRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCYRRAAELF KNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGF EISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQKSPKPISLLLSTQR RKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSI DVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISDNDLFHFNKKMFAR RRILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLIERWACEIADFFIKN KVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQNKIEFKLKQYGIEIRK VAPNNTSKTCSKCGHLNNYFNFEYRKKNKFPHFKCEKCNFKENAAYNAAL NISNPKLKSTKERP

mA6	33	MAKNTITKTLKLRIVRPYNSQEVEKIVAEEKNNREKIALDKNKDKVKEACS KHLKVAAYCTTQVERNACLFCKARKLDDKFYKKLRGQFPDAVFWQEISEI FRQLQKQAREIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCYRRAAELFK NAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGFE ISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQKSPKPISLLLSTQRR KRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSID VPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISDNDLFHFNKKMFARR RILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLIERWACEIADFFIKNK VGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQNKIEFKLKQYGIEIRKV APNNTSKTCSKCGHLNNYFNFEYRKKNKFPHFKCEKCNFKENAAYNAALN ISNPKLKSTKERP
mA7	34	MAKNTITKTLKLRIVRPYYSAEVEKIVAEEKNNREKIALDKNKDKVKEACS KHLKVAAYCTTQVERNACLFCKARKLDDKFYKKLRKQFPDAVFWQEISEI FRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCYRRAAELF KNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGF EISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQKSPKPISLLLSTQR RKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSI DVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISDNDLFHFNKKMFAR RRILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLIERWACEIADFFIKN KVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQNKIEFKLKQYGIEIRK VAPNNTSKTCSKCGHLNNYFNFEYRKKNKFPHFKCEKCNFKENAAYNAAL NISNPKLKSTKERP
mA8	35	MAKNTITKTLKLRIVRPYNSAEIEKIVADEKNRREKIALDKNKDKVKEACS KHLKVAAYCTTQVERNACLFCKARKLDDKFYKKLRKQFPDAVFWQEISEI FRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCYRRAAELF KNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGF EISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQKSPKPISLLLSTQR RKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSI DVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISDNDLFHFNKKMFAR RRILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLIERWACEIADFFIKN KVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQNKIEFKLKQYGIEIRK VAPNNTSKTCSKCGHLNNYFNFEYRKKNKFPHFKCEKCNFKENAAYNAAL NISNPKLKSTKERP

mA9	36	MAKNTITKTLKLRIVRPYNSAEIEKIVADEKNRREKIALDKNKDKVKEACS KHLKVAAYCTTQVERNACLFCKARKLDDKFYKKLRGQFPDAVFWQEISEI FRQLQKQAREIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCYRRAAELFK NAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGFE ISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQKSPKPISLLLSTQRR KRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSID VPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISDNDLFHFNKKMFARR RILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLIERWACEIADFFIKNK VGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQNKIEFKLKQYGIEIRKV APNNTSKTCSKCGHLNNYFNFEYRKKNKFPHFKCEKCNFKENAAYNAALN ISNPKLKSTKERP
mA10	37	MAKNTITKTLKLRIVRPYNSQEIEKIVAEEKNRREKIALDKNKDKVKEACSK HLKVAAYCTTQVERNACLFCKARKLDDKFYQKLRGQFPDAVFWQEISEIFR QLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCYRRAAELFKN AAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGFEIS NHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQKSPKPISLLLSTQRRKR NKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSIDVP KIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISDNDLFHFNKKMFARRRI LLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLIERWACEIADFFIKNKV GTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQNKIEFKLKQYGIEIRKVA PNNTSKTCSKCGHLNNYFNFEYRKKNKFPHFKCEKCNFKENAAYNAALNIS NPKLKSTKERP
mA11	38	MAKNTITKTLKLRIVRPYNSAEIEKIVAEEKNRREKIALDKNKDKVKEACSK HLKVAAYCTTQVERNACLFCKARKLDDKFYKKLRGQFPDAVFWQEISEIFR QLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCYRRAAELFKN AAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGFEIS NHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQKSPKPISLLLSTQRRKR NKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSIDVP KIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISDNDLFHFNKKMFARRRI LLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLIERWACEIADFFIKNKV GTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQNKIEFKLKQYGIEIRKVA PNNTSKTCSKCGHLNNYFNFEYRKKNKFPHFKCEKCNFKENAAYNAALNIS NPKLKSTKERP

mA12	39	MAKNTITKTLKLRIVRPYNSAEVEKIVAEEKNRREKIALDKNKDKVKEACS KHLKVAAYCTTQVERNACLFCKARKLDDKFYKKLRKQFPDAVFWQEISEI FRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCYRRAAELF KNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGF EISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQKSPKPISLLLSTQR RKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSI DVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISDNDLFHFNKKMFAR RRILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLIERWACEIADFFIKN KVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQNKIEFKLKQYGIEIRK VAPNNTSKTCSKCGHLNNYFNFEYRKKNKFPHFKCEKCNFKENAAYNAAL NISNPKLKSTKERP
mA13	40	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKNRREKIALDKNKDKVKEACS KHLKVAAYCTTQVERNACLFCKARKLDDKFYKKLRKQFPDAVFWQEISEI FRQLQKQAREIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCYRRAAELFK NAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGFE ISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQKSPKPISLLLSTQRR KRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSID VPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISDNDLFHFNKKMFARR RILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLIERWACEIADFFIKNK VGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQNKIEFKLKQYGIEIRKV APNNTSKTCSKCGHLNNYFNFEYRKKNKFPHFKCEKCNFKENAAYNAALN ISNPKLKSTKERP
mA14	41	MAKNTITKTLKLRIVRPYYSAEIEKIVADEKNRREKIALDKNKDKVKEACS KHLKVAAYCTTQVERNACLFCKARKLDDKFYQKLRKQFPDAVFWQEISEI FRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCYRRAAELF KNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGF EISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQKSPKPISLLLSTQR RKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSI DVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISDNDLFHFNKKMFAR RRILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLIERWACEIADFFIKN KVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQNKIEFKLKQYGIEIRK VAPNNTSKTCSKCGHLNNYFNFEYRKKNKFPHFKCEKCNFKENAAYNAAL NISNPKLKSTKERP

mA15	42	MAKNTITKTLKLRIVRPYYSAEIEKIVAEEKNRREKIALEKNKDKVKEACSK HLKVAAYCTTQVERNACLFCKARKLDDKFYQKLRKQFPDAVFWQEISEIFR QLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCYRRAAELFKN AAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGFEIS NHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQKSPKPISLLLSTQRRKR NKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSIDVP KIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISDNDLFHFNKKMFARRRI LLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLIERWACEIADFFIKNKV GTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQNKIEFKLKQYGIEIRKVA PNNTSKTCSKCGHLNNYFNFEYRKKNKFPHFKCEKCNFKENAAYNAALNIS NPKLKSTKERP
mA16	43	MAKNTITKTLKLRIVRPYYSAEIEKIVAEEKNNREKIALDKNKDKVKEACSK HLKVAAYCTTQVERNACLFCKARKLDDKFYQKLRGQFPDAVFWQEISEIFR QLQKQAREIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCYRRAAELFKN AAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGFEIS NHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQKSPKPISLLLSTQRRKR NKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSIDVP KIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISDNDLFHFNKKMFARRRI LLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLIERWACEIADFFIKNKV GTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQNKIEFKLKQYGIEIRKVA PNNTSKTCSKCGHLNNYFNFEYRKKNKFPHFKCEKCNFKENAAYNAALNIS NPKLKSTKERP
mB1	44	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKNNREKIALEKNKDKVKEACS KHLKVAAYCTTQVERNACLFCKARKLDDKFYQKLRGQFPDAVFWQEISEI FRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCYRRAARLF KNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGF KISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFRQVQKSPKPISLLLSTQR RKRNKGWSKDEGTEAEIRKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSI DVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISDNDLFHFNKKMFAR RRILLKKNRHKRAGHGAKNKLKPIEQLTEKSERFRKKLIERWACEIADFFIK NKVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQNKIEFKLKQYGIEIR KVAPNNTSKTCSKCGHLNNYFNFEYRKKNKFPHFKCEKCNFKENAAYNAA LNISNPKLKSTKERP

mB2	45	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKNNREKIALEKNKDKVKEACS KHLKVAAYCTTQVERNACLFCKARKLDDKFYQKLRGQFPDAVFWQEISEI FRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCYRRAAALF KNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGF KISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFRQVQKSPKPISLLLSTQR RKRNKGWSKDEGTEAEIRKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSI DVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISDNDLFHFNKKMFAR RRILLKKNRHKRAGHGAKNKLKPIEQLTEKSERFRKKLIERWACEIADFFIK NKVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQNKIEFKLKQYGIEIR KVAPNNTSKTCSKCGHLNNYFNFEYRKKNKFPHFKCEKCNFKENAAYNAA LNISNPKLKSTKERP
mB3	46	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKNNREKIALEKNKDKVKEACS KHLKVAAYCTTQVERNACLFCKARKLDDKFYQKLRGQFPDAVFWQEISEI FRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCYRRAAGLF KNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGF KISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFRQVQKSPKPISLLLSTQR RKRNKGWSKDEGTEAEIRKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSI DVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISDNDLFHFNKKMFAR RRILLKKNRHKRAGHGAKNKLKPIEQLTEKSERFRKKLIERWACEIADFFIK NKVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQNKIEFKLKQYGIEIR KVAPNNTSKTCSKCGHLNNYFNFEYRKKNKFPHFKCEKCNFKENAAYNAA LNISNPKLKSTKERP
mB4	47	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKNNREKIALEKNKDKVKEACS KHLKVAAYCTTQVERNACLFCKARKLDDKFYQKLRGQFPDAVFWQEISEI FRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCYRRAARLF KNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGF RISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFRQVQKSPKPISLLLSTQR RKRNKGWSKDEGTEAEIRKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSI DVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISDNDLFHFNKKMFAR RRILLKKNRHKRAGHGAKNKLKPIEQLTEKSERFRKKLIERWACEIADFFIK NKVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQNKIEFKLKQYGIEIR KVAPNNTSKTCSKCGHLNNYFNFEYRKKNKFPHFKCEKCNFKENAAYNAA LNISNPKLKSTKERP

mB5	48	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKNNREKIALEKNKDKVKEACS KHLKVAAYCTTQVERNACLFCKARKLDDKFYQKLRGQFPDAVFWQEISEI FRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCYRRAAALF KNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGF RISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFRQVQKSPKPISLLLSTQR RKRNKGWSKDEGTEAEIRKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSI DVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISDNDLFHFNKKMFAR RRILLKKNRHKRAGHGAKNKLKPIEQLTEKSERFRKKLIERWACEIADFFIK NKVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQNKIEFKLKQYGIEIR KVAPNNTSKTCSKCGHLNNYFNFEYRKKNKFPHFKCEKCNFKENAAYNAA LNISNPKLKSTKERP
mB6	49	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKNNREKIALEKNKDKVKEACS KHLKVAAYCTTQVERNACLFCKARKLDDKFYQKLRGQFPDAVFWQEISEI FRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCYRRAAGLF KNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGF RISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFRQVQKSPKPISLLLSTQR RKRNKGWSKDEGTEAEIRKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSI DVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISDNDLFHFNKKMFAR RRILLKKNRHKRAGHGAKNKLKPIEQLTEKSERFRKKLIERWACEIADFFIK NKVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQNKIEFKLKQYGIEIR KVAPNNTSKTCSKCGHLNNYFNFEYRKKNKFPHFKCEKCNFKENAAYNAA LNISNPKLKSTKERP
mB7	50	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKNNREKIALEKNKDKVKEACS KHLKVAAYCTTQVERNACLFCKARKLDDKFYQKLRGQFPDAVFWQEISEI FRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCYRRAARLF KNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGF SISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFRQVQKSPKPISLLLSTQR RKRNKGWSKDEGTEAEIRKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSI DVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISDNDLFHFNKKMFAR RRILLKKNRHKRAGHGAKNKLKPIEQLTEKSERFRKKLIERWACEIADFFIK NKVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQNKIEFKLKQYGIEIR KVAPNNTSKTCSKCGHLNNYFNFEYRKKNKFPHFKCEKCNFKENAAYNAA LNISNPKLKSTKERP

mB8	51	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKNNREKIALEKNKDKVKEACS KHLKVAAYCTTQVERNACLFCKARKLDDKFYQKLRGQFPDAVFWQEISEI FRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCYRRAAALF KNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGF SISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFRQVQKSPKPISLLLSTQR RKRNKGWSKDEGTEAEIRKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSI DVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISDNDLFHFNKKMFAR RRILLKKNRHKRAGHGAKNKLKPIEQLTEKSERFRKKLIERWACEIADFFIK NKVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQNKIEFKLKQYGIEIR KVAPNNTSKTCSKCGHLNNYFNFEYRKKNKFPHFKCEKCNFKENAAYNAA LNISNPKLKSTKERP
mB9	52	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKNNREKIALEKNKDKVKEACS KHLKVAAYCTTQVERNACLFCKARKLDDKFYQKLRGQFPDAVFWQEISEI FRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCYRRAAGLF KNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGF SISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFRQVQKSPKPISLLLSTQR RKRNKGWSKDEGTEAEIRKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSI DVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISDNDLFHFNKKMFAR RRILLKKNRHKRAGHGAKNKLKPIEQLTEKSERFRKKLIERWACEIADFFIK NKVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQNKIEFKLKQYGIEIR KVAPNNTSKTCSKCGHLNNYFNFEYRKKNKFPHFKCEKCNFKENAAYNAA LNISNPKLKSTKERP
mC1	53	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKNNREKIALEKNKDKVKEACS KHLKVAAYCTTQVERNACLFCKARKLDDKFYQKLRGQFPDAVFWQEISEI FRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCYRRAAELF KNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGF EISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQKSPKPISLLLSTQR RKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSI DVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISDNDLENFNKKMFAR RRILLKKNRHKRGGHGAKNKLKPIEQLTEKSERFRKKLIERWACEIADFFIK NKVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQNKIEFKLKQYGIEIR KVAPNNTSKTCSKCGHLNNYFNFEYRKKNKFPHFKCEKCNFKENAAYNAA LNISNPKLKSTKERP

mC2	54	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKNNREKIALEKNKDKVKEACS KHLKVAAYCTTQVERNACLFCKARKLDDKFYQKLRGQFPDAVFWQEISEI FRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCYRRAAELF KNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGF EISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQKSPKPISLLLSTQR RKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSI DVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSIEGGDLFHFNKKMFAR RRILLKKNRHKRAGHGAKNKLKPIEQLTEKSERFRKKLIERWACEIADFFIK NKVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQNKIEFKLKQYGIEIR KVAPNNTSKTCSKCGHLNNYFNFEYRKKNKFPHFKCEKCNFKENAAYNAA LNISNPKLKSTKERP
mC3	55	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKNNREKIALEKNKDKVKEACS KHLKVAAYCTTQVERNACLFCKARKLDDKFYQKLRGQFPDAVFWQEISEI FRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCYRRAAELF KNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGF EISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQKSPKPISLLLSTQR RKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSI DVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSIEGGDLENFNKKMFA RRRILLKKNRHKRGGHGRDKKLKPIEQLTEKSERFRKKLIERWACEIADFFI KNKVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQNKIEFKLKQYGIEI RKVAPNNTSKTCSKCGHLNNYFNFEYRKKNKFPHFKCEKCNFKENAAYNA ALNISNPKLKSTKERP
mC4	56	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKNNREKIALEKNKDKVKEACS KHLKVAAYCTTQVERNACLFCKARKLDDKFYQKLRGQFPDAVFWQEISEI FRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCYRRAAELF KNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGF EISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQKSPKPISLLLSTQR RKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSI DVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSIEGGDLENFNKKMFA RRRILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLIERWACEIADFFIK NKVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQNKIEFKLKQYGIEIR KVAPNNTSKTCSKCGHLNNYFNFEYRKKNKFPHFKCEKCNFKENAAYNAA LNISNPKLKSTKERP

mC5	57	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKNNREKIALEKNKDKVKEACS KHLKVAAYCTTQVERNACLFCKARKLDDKFYQKLRGQFPDAVFWQEISEI FRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCYRRAAELF KNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGF EISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQKSPKPISLLLSTQR RKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSI DVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISDNDLFHFNKKMFAR RRILLKKNRHKRAGHGRDKKLKPIEQLTEKSERFRKKLIERWACEIADFFIK NKVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQNKIEFKLKQYGIEIR KVAPNNTSKTCSKCGHLNNYFNFEYRKKNKFPHFKCEKCNFKENAAYNAA LNISNPKLKSTKERP
mC6	58	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKNNREKIALEKNKDKVKEACS KHLKVAAYCTTQVERNACLFCKARKLDDKFYQKLRGQFPDAVFWQEISEI FRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCYRRAAELF KNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGF EISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQKSPKPISLLLSTQR RKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSI DVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISDNDLENFNKKMFAR RRILLKKNRHKRAGHGRDKKLKPITILTEKSERFRKKLIERWACEIADFFIKN KVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQNKIEFKLKQYGIEIRK VAPNNTSKTCSKCGHLNNYFNFEYRKKNKFPHFKCEKCNFKENAAYNAAL NISNPKLKSTKERP
mC7	59	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKNNREKIALEKNKDKVKEACS KHLKVAAYCTTQVERNACLFCKARKLDDKFYQKLRGQFPDAVFWQEISEI FRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCYRRAAELF KNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGF EISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQKSPKPISLLLSTQR RKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSI DVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISDNDLEHFNKKMFAR RRILLKKNRHKRKGHGAKNKLKPIETLTEKSERFRKKLIERWACEIADFFIK NKVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQNKIEFKLKQYGIEIR KVAPNNTSKTCSKCGHLNNYFNFEYRKKNKFPHFKCEKCNFKENAAYNAA LNISNPKLKSTKERP

mC8	60	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKNNREKIALEKNKDKVKEACS KHLKVAAYCTTQVERNACLFCKARKLDDKFYQKLRGQFPDAVFWQEISEI FRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCYRRAAELF KNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGF EISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQKSPKPISLLLSTQR RKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSI DVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSIDGGDLEHFNKKMFA RRRILLKKNRHKRKGHGAKNKLKPITILTEKSERFRKKLIERWACEIADFFIK NKVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQNKIEFKLKQYGIEIR KVAPNNTSKTCSKCGHLNNYFNFEYRKKNKFPHFKCEKCNFKENAAYNAA LNISNPKLKSTKERP
mC9	61	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKNNREKIALEKNKDKVKEACS KHLKVAAYCTTQVERNACLFCKARKLDDKFYQKLRGQFPDAVFWQEISEI FRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCYRRAAELF KNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGF EISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQKSPKPISLLLSTQR RKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSI DVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSIDGGDLFHFNKKMFA RRRILLKKNRHKRAGHGAKNKLKPIETLTEKSERFRKKLIERWACEIADFFI KNKVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQNKIEFKLKQYGIEI RKVAPNNTSKTCSKCGHLNNYFNFEYRKKNKFPHFKCEKCNFKENAAYNA ALNISNPKLKSTKERP
mC10	62	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKNNREKIALEKNKDKVKEACS KHLKVAAYCTTQVERNACLFCKARKLDDKFYQKLRGQFPDAVFWQEISEI FRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCYRRAAELF KNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGF EISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQKSPKPISLLLSTQR RKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSI DVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSIDGGDLEHFNKKMFA RRRILLKKNRHKRKGHGAKNKLKPIETLTEKSERFRKKLIERWACEIADFFI KNKVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQNKIEFKLKQYGIEI RKVAPNNTSKTCSKCGHLNNYFNFEYRKKNKFPHFKCEKCNFKENAAYNA ALNISNPKLKSTKERP

mC11	63	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKNNREKIALEKNKDKVKEACS KHLKVAAYCTTQVERNACLFCKARKLDDKFYQKLRGQFPDAVFWQEISEI FRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCYRRAAELF KNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGF EISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQKSPKPISLLLSTQR RKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSI DVPKIDKGVDPSIIGGIAVGVRSPLVCAINNSFSRYSIDSNDLFKFNKKMFAR RRILLKKNRHKRKGHGAKNKLKPITELTEKSERFRKKLIERWACEIADFFIK NKVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQNKIEFKLKQYGIEIR KVAPNNTSKTCSKCGHLNNYFNFEYRKKNKFPHFKCEKCNFKENAAYNAA LNISNPKLKSTKERP
mC12	64	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKNNREKIALEKNKDKVKEACS KHLKVAAYCTTQVERNACLFCKARKLDDKFYQKLRGQFPDAVFWQEISEI FRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCYRRAAELF KNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGF EISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQKSPKPISLLLSTQR RKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSI DVPKIDKGVDPSIIGGIAVGVRSPLVCAINNSFSRYSIDSNDLFHFNKKMFAR RRILLKKNRHKRAGHGAAHKLKPITILTEKSERFRKKLIERWACEIADFFIKN KVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQNKIEFKLKQYGIEIRK VAPNNTSKTCSKCGHLNNYFNFEYRKKNKFPHFKCEKCNFKENAAYNAAL NISNPKLKSTKERP
mC13	65	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKNNREKIALEKNKDKVKEACS KHLKVAAYCTTQVERNACLFCKARKLDDKFYQKLRGQFPDAVFWQEISEI FRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCYRRAAELF KNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGF EISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQKSPKPISLLLSTQR RKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSI DVPKIDKGVDPSIIGGIAVGVRSPLVCAINNSFSRYSIDSNDLFKFNKKMFAR RRILLKKNRHKRAGHGAAHKLKPITELTEKSERFRKKLIERWACEIADFFIK NKVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQNKIEFKLKQYGIEIR KVAPNNTSKTCSKCGHLNNYFNFEYRKKNKFPHFKCEKCNFKENAAYNAA LNISNPKLKSTKERP

mC14	66	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKNNREKIALEKNKDKVKEACS KHLKVAAYCTTQVERNACLFCKARKLDDKFYQKLRGQFPDAVFWQEISEI FRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCYRRAAELF KNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGF EISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQKSPKPISLLLSTQR RKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSI DVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSIDSNDLFKFNKKMFAR RRILLKKNRHKRAGHGAAHKLKPITILTEKSERFRKKLIERWACEIADFFIKN KVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQNKIEFKLKQYGIEIRK VAPNNTSKTCSKCGHLNNYFNFEYRKKNKFPHFKCEKCNFKENAAYNAAL NISNPKLKSTKERP
mC15	67	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKNNREKIALEKNKDKVKEACS KHLKVAAYCTTQVERNACLFCKARKLDDKFYQKLRGQFPDAVFWQEISEI FRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCYRRAAELF KNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGF EISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQKSPKPISLLLSTQR RKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSI DVPKIDKGVDPSIIGGIAVGVRSPLVCAINNSFSRYSIDSNDLFKFNKKMFAR RRILLKKNRHKRAGHGAKNKLKPITELTEKSERFRKKLIERWACEIADFFIK NKVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQNKIEFKLKQYGIEIR KVAPNNTSKTCSKCGHLNNYFNFEYRKKNKFPHFKCEKCNFKENAAYNAA LNISNPKLKSTKERP
mC16	68	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKNNREKIALEKNKDKVKEACS KHLKVAAYCTTQVERNACLFCKARKLDDKFYQKLRGQFPDAVFWQEISEI FRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCYRRAAELF KNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGF EISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQKSPKPISLLLSTQR RKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSI DVPKIDKGVDPSIIGGIAVGVRSPLVCAINNSFSRYSISDNDLFKFNKKMFAR RRILLKKNRHKRKGHGAKNKLKPITELTEKSERFRKKLIERWACEIADFFIK NKVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQNKIEFKLKQYGIEIR KVAPNNTSKTCSKCGHLNNYFNFEYRKKNKFPHFKCEKCNFKENAAYNAA LNISNPKLKSTKERP

mC17	69	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKNNREKIALEKNKDKVKEACS KHLKVAAYCTTQVERNACLFCKARKLDDKFYQKLRGQFPDAVFWQEISEI FRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCYRRAAELF KNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGF EISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQKSPKPISLLLSTQR RKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSI DVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSIDSNDLFKFNKKMFAR RRILLKKNRHKRKGHGAKNKLKPITELTEKSERFRKKLIERWACEIADFFIK NKVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQNKIEFKLKQYGIEIR KVAPNNTSKTCSKCGHLNNYFNFEYRKKNKFPHFKCEKCNFKENAAYNAA LNISNPKLKSTKERP
mC18	70	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKNNREKIALEKNKDKVKEACS KHLKVAAYCTTQVERNACLFCKARKLDDKFYQKLRGQFPDAVFWQEISEI FRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCYRRAAELF KNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGF EISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQKSPKPISLLLSTQR RKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSI DVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISDNDLFKFNKKMFAR RRILLKKNRHKRKGHGAAHKLKPITELTEKSERFRKKLIERWACEIADFFIK NKVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQNKIEFKLKQYGIEIR KVAPNNTSKTCSKCGHLNNYFNFEYRKKNKFPHFKCEKCNFKENAAYNAA LNISNPKLKSTKERP
mC19	71	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKNNREKIALEKNKDKVKEACS KHLKVAAYCTTQVERNACLFCKARKLDDKFYQKLRGQFPDAVFWQEISEI FRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCYRRAAELF KNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGF EISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQKSPKPISLLLSTQR RKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSI DVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSIKGGDLERFNKKMFA RRRILLKKNRHKRKGHGAKNKLKPITILTEKSERFRKKLIERWACEIADFFIK NKVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQNKIEFKLKQYGIEIR KVAPNNTSKTCSKCGHLNNYFNFEYRKKNKFPHFKCEKCNFKENAAYNAA LNISNPKLKSTKERP

mC20	72	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKNNREKIALEKNKDKVKEACS KHLKVAAYCTTQVERNACLFCKARKLDDKFYQKLRGQFPDAVFWQEISEI FRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCYRRAAELF KNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGF EISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQKSPKPISLLLSTQR RKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSI DVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSIKGGDLERFNKKMFA RRRILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLIERWACEIADFFIK NKVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQNKIEFKLKQYGIEIR KVAPNNTSKTCSKCGHLNNYFNFEYRKKNKFPHFKCEKCNFKENAAYNAA LNISNPKLKSTKERP
mC21	73	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKNNREKIALEKNKDKVKEACS KHLKVAAYCTTQVERNACLFCKARKLDDKFYQKLRGQFPDAVFWQEISEI FRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCYRRAAELF KNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGF EISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQKSPKPISLLLSTQR RKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSI DVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSIKGGDLEKFNKKMFA RRRILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLIERWACEIADFFIK NKVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQNKIEFKLKQYGIEIR KVAPNNTSKTCSKCGHLNNYFNFEYRKKNKFPHFKCEKCNFKENAAYNAA LNISNPKLKSTKERP
mC22	74	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKNNREKIALEKNKDKVKEACS KHLKVAAYCTTQVERNACLFCKARKLDDKFYQKLRGQFPDAVFWQEISEI FRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCYRRAAELF KNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGF EISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQKSPKPISLLLSTQR RKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSI DVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSIKGGDLFHFNKKMFA RRRILLKKNRHKRAGHGRKKKLKPITILTEKSERFRKKLIERWACEIADFFIK NKVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQNKIEFKLKQYGIEIR KVAPNNTSKTCSKCGHLNNYFNFEYRKKNKFPHFKCEKCNFKENAAYNAA LNISNPKLKSTKERP

mC23	75	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKNNREKIALEKNKDKVKEACS KHLKVAAYCTTQVERNACLFCKARKLDDKFYQKLRGQFPDAVFWQEISEI FRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCYRRAAELF KNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGF EISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQKSPKPISLLLSTQR RKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSI DVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISDNDLEKFNKKMFAR RRILLKKNRHKRAGHGRKKKLKPITILTEKSERFRKKLIERWACEIADFFIKN KVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQNKIEFKLKQYGIEIRK VAPNNTSKTCSKCGHLNNYFNFEYRKKNKFPHFKCEKCNFKENAAYNAAL NISNPKLKSTKERP
mC24	76	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKNNREKIALEKNKDKVKEACS KHLKVAAYCTTQVERNACLFCKARKLDDKFYQKLRGQFPDAVFWQEISEI FRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCYRRAAELF KNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGF EISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQKSPKPISLLLSTQR RKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSI DVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSIKGGDLEKFNKKMFA RRRILLKKNRHKRAGHGRKKKLKPITILTEKSERFRKKLIERWACEIADFFIK NKVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQNKIEFKLKQYGIEIR KVAPNNTSKTCSKCGHLNNYFNFEYRKKNKFPHFKCEKCNFKENAAYNAA LNISNPKLKSTKERP
mD1	77	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKNNREKIALEKNKDKVKEACS KHLKVAAYCTTQVERNACLFCKARKLDDKFYQKLRGQFPDAVFWQEISEI FRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCYRRAAELF KNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGF EISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQKSPKPISLLLSTQR RKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSI DVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISDNDLFHFNKKMFAR RRILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLIERWAKEIADFFIKN KVGTVQMEDLSTMKRKEDSYFNIRLRGFWPYYEMQNKIEFKLKQYGIEIRK VAPNNTSQLCSKCGHLNNYFNFEYRKKNKFPKFKCEKCNFKENAAYNAAR NISTPDIKSTKERP

mD2	78	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKNNREKIALEKNKDKVKEACS KHLKVAAYCTTQVERNACLFCKARKLDDKFYQKLRGQFPDAVFWQEISEI FRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCYRRAAELF KNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGF EISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQKSPKPISLLLSTQR RKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSI DVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISDNDLFHFNKKMFAR RRILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLIERWACEIADFFIKN KVGTVQMEDLESMKRKEDSYFNIRLRGFWPYAEMQNKIEFKLKQYGIEIRK VAPNNTSQLCSKCGHLNNYFNFEYRKKNKFPHFKCEKCNFKENAAYNAAL NISNPDIKSTKERP
mD3	79	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKNNREKIALEKNKDKVKEACS KHLKVAAYCTTQVERNACLFCKARKLDDKFYQKLRGQFPDAVFWQEISEI FRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCYRRAAELF KNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGF EISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQKSPKPISLLLSTQR RKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSI DVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISDNDLFHFNKKMFAR RRILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLIERWAKEIADFFIKN KVGTVQMEDLSTMKRKEDSYFNIRLRGFWPYYEMQNKIEFKLKQYGIEIRK VAPNNTSKTCSKCGHLNNYFNFEYRKKNKFPHFKCEKCNFKENAAYNAAL NISNPKLKSTKERP
mD4	80	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKNNREKIALEKNKDKVKEACS KHLKVAAYCTTQVERNACLFCKARKLDDKFYQKLRGQFPDAVFWQEISEI FRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCYRRAAELF KNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGF EISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQKSPKPISLLLSTQR RKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSI DVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISDNDLFHFNKKMFAR RRILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLIERWACEIADFFIKN KVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQNKIEFKLKQYGIEIRK VAPNNTSQLCSKCGHLNNYFNFEYRKKNKFPKFKCEKCNFKENAAYNAAL NISNPDIKSTKERP

mD5	81	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKNNREKIALEKNKDKVKEACS KHLKVAAYCTTQVERNACLFCKARKLDDKFYQKLRGQFPDAVFWQEISEI FRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCYRRAAELF KNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGF EISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQKSPKPISLLLSTQR RKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSI DVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISDNDLFHFNKKMFAR RRILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLIERWACEIADFFIKN KVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQNKIEFKLKQYGIEIRK VAPNNTSQLCSKCGHLNNYFNFEYRKKNKFPHFKCEKCNFKENAAYNAAL NISTPDIKSTKERP
mD6	82	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKNNREKIALEKNKDKVKEACS KHLKVAAYCTTQVERNACLFCKARKLDDKFYQKLRGQFPDAVFWQEISEI FRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCYRRAAELF KNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGF EISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQKSPKPISLLLSTQR RKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSI DVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISDNDLFHFNKKMFAR RRILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLIERWACEIADFFIKN KVGTVQMEDLSTMKRKEDSYFNIRLRGFWPYYEMQNKIEFKLKQYGIEIRK VAPNNTSKTCSKCGHLNNYFNFEYRKKNKFPKFKCEKCNFKENAAYNAAL NISNPKLKSTKERP
mD7	83	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKNNREKIALEKNKDKVKEACS KHLKVAAYCTTQVERNACLFCKARKLDDKFYQKLRGQFPDAVFWQEISEI FRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCYRRAAELF KNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGF EISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQKSPKPISLLLSTQR RKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSI DVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISDNDLFHFNKKMFAR RRILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLIERWACEIADFFIKN KVGTVQMEDLESMKRKEDSYFNIRLRGFWPYAEMQNKIEFKLKQYGIEIRK VAPNNTSQLCSKCGHLNNYFNFEYRKKNKFPKFKCEKCNFKENAAYNAAL NISTPDIKSTKERP

mD8	84	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKNNREKIALEKNKDKVKEACS KHLKVAAYCTTQVERNACLFCKARKLDDKFYQKLRGQFPDAVFWQEISEI FRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCYRRAAELF KNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGF EISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQKSPKPISLLLSTQR RKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSI DVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISDNDLFHFNKKMFAR RRILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLIERWACEIADFFIKN KVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQNKIEFKLKQYGIEIRK VAPNNTSKTCSKCGHLNNYFNFEYRKKNKFPKFKCEKCNFKENAAYNAAR NISTPDIKSTKERP
mD9	85	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKNNREKIALEKNKDKVKEACS KHLKVAAYCTTQVERNACLFCKARKLDDKFYQKLRGQFPDAVFWQEISEI FRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCYRRAAELF KNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGF EISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQKSPKPISLLLSTQR RKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSI DVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISDNDLFHFNKKMFAR RRILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLIERWSRYIADFFIKN KVGTVQMEDLESMKRKEDSYFNIRLRGFWPYYEMQNKIEFKLKQYGIKIR KVAPNNTSQRCSKCGHLNNYFNFEYRKKNKFPHFKCEKCNFKANAAYNA ARNISNPNIKSTKERP
mD10	86	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKNNREKIALEKNKDKVKEACS KHLKVAAYCTTQVERNACLFCKARKLDDKFYQKLRGQFPDAVFWQEISEI FRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCYRRAAELF KNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGF EISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQKSPKPISLLLSTQR RKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSI DVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISDNDLFHFNKKMFAR RRILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLIERWACYIADFFIK NKVGTVQMEDLESMKRKEDSYFNIRLRGFWPYAEMQNKIEFKLKQYGIEIR KVAPNNTSQRCSKCGHLNNYFNFEYRKKNKFPHFKCEKCNFKENAAYNAA RNISNPNIKSTKERP

mD11	87	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKNNREKIALEKNKDKVKEACS KHLKVAAYCTTQVERNACLFCKARKLDDKFYQKLRGQFPDAVFWQEISEI FRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCYRRAAELF KNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGF EISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQKSPKPISLLLSTQR RKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSI DVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISDNDLFHFNKKMFAR RRILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLIERWACYIADFFIK NKVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQNKIEFKLKQYGIEIR KVAPNNTSQRCSKCGHLNNYFNFEYRKKNKFPHFKCEKCNFKRNAAYNA ARNISNPKLKSTKERP
mD12	88	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKNNREKIALEKNKDKVKEACS KHLKVAAYCTTQVERNACLFCKARKLDDKFYQKLRGQFPDAVFWQEISEI FRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCYRRAAELF KNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGF EISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQKSPKPISLLLSTQR RKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSI DVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISDNDLFHFNKKMFAR RRILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLIERWACYIADFFIK NKVGTVQMEDLESMKRKEDSYFNIRLRGFWPYAEMQNKIEFKLKQYGIEIR KVAPNNTSKTCSKCGHLNNYFNFEYRKKNKFPHFKCEKCNFKENAAYNAA RNISNPNIKSTKERP
mD13	89	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKNNREKIALEKNKDKVKEACS KHLKVAAYCTTQVERNACLFCKARKLDDKFYQKLRGQFPDAVFWQEISEI FRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCYRRAAELF KNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGF EISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQKSPKPISLLLSTQR RKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSI DVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISDNDLFHFNKKMFAR RRILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLIERWARYIADFFIK NKVGTVQMEDLESMKRKEDSYFNIRLRGFWPYYEMQNKIEFKLKQYGIKI RKVAPNNTSKTCSKCGHLNNYFNFEYRKKNKFPHFKCEKCNFKENAAYNA ALNISNPKLKSTKERP

mD14	90	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKNNREKIALEKNKDKVKEACS KHLKVAAYCTTQVERNACLFCKARKLDDKFYQKLRGQFPDAVFWQEISEI FRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCYRRAAELF KNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGF EISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQKSPKPISLLLSTQR RKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSI DVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISDNDLFHFNKKMFAR RRILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLIERWACEIADFFIKN KVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQNKIEFKLKQYGIEIRK VAPNNTSQRCSKCGHLNNYFNFEYRKKNKFPHFKCEKCNFKRNAAYNAAR NISNPNIKSTKERP
mD15	91	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKNNREKIALEKNKDKVKEACS KHLKVAAYCTTQVERNACLFCKARKLDDKFYQKLRGQFPDAVFWQEISEI FRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCYRRAAELF KNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGF EISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQKSPKPISLLLSTQR RKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSI DVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISDNDLFHFNKKMFAR RRILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLIERWACEIADFFIKN KVGTVQMEDLESMKRKEDSYFNIRLRGFWPYAEMQNKIEFKLKQYGIEIRK VAPNNTSQRCSKCGHLNNYFNFEYRKKNKFPHFKCEKCNFKENAAYNAAR NISNPNIKSTKERP
mD16	92	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKNNREKIALEKNKDKVKEACS KHLKVAAYCTTQVERNACLFCKARKLDDKFYQKLRGQFPDAVFWQEISEI FRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCYRRAAELF KNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGF EISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQKSPKPISLLLSTQR RKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSI DVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISDNDLFHFNKKMFAR RRILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLIERWACEIADFFIKN KVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQNKIEFKLKQYGIEIRK VAPNNTSKTCSKCGHLNNYFNFEYRKKNKFPHFKCEKCNFKRNAAYNAAR NISNPNIKSTKERP

mD17	93	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKNNREKIALEKNKDKVKEACS KHLKVAAYCTTQVERNACLFCKARKLDDKFYQKLRGQFPDAVFWQEISEI FRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCYRRAAELF KNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGF EISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQKSPKPISLLLSTQR RKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSI DVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISDNDLFHFNKKMFAR RRILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLIERWANRIADFFIK NKVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQNKIEFKLKQYGIKI RKVAPNNTSQRCSKCGHLNNYFNFEYRKKNKFPHFKCEKCNFKRNAAYNA AKNISNPKLKSTKERP
mD18	94	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKNNREKIALEKNKDKVKEACS KHLKVAAYCTTQVERNACLFCKARKLDDKFYQKLRGQFPDAVFWQEISEI FRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCYRRAAELF KNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGF EISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQKSPKPISLLLSTQR RKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSI DVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISDNDLFHFNKKMFAR RRILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLIERWACEIADFFIKN KVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQNKIEFKLKQYGIEIRK VAPNNTSQRCSKCGHLNNYFNFEYRKKNKFPHFKCEKCNFKRNAAYNAA KNISNPKLKSTKERP
mD19	95	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKNNREKIALEKNKDKVKEACS KHLKVAAYCTTQVERNACLFCKARKLDDKFYQKLRGQFPDAVFWQEISEI FRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCYRRAAELF KNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGF EISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQKSPKPISLLLSTQR RKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSI DVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISDNDLFHFNKKMFAR RRILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLIERWACEIADFFIKN KVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQNKIEFKLKQYGIKIR KVAPNNTSQRCSKCGHLNNYFNFEYRKKNKFPHFKCEKCNFKRNAAYNA AKNISNPKLKSTKERP

mD20	96	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKNNREKIALEKNKDKVKEACS KHLKVAAYCTTQVERNACLFCKARKLDDKFYQKLRGQFPDAVFWQEISEI FRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCYRRAAELF KNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGF EISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQKSPKPISLLLSTQR RKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSI DVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISDNDLFHFNKKMFAR RRILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLIERWANRIADFFIK NKVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQNKIEFKLKQYGIKI RKVAPNNTSQRCSKCGHLNNYFNFEYRKKNKFPHFKCEKCNFKENAAYNA ALNISNPKLKSTKERP
mD21	97	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKNNREKIALEKNKDKVKEACS KHLKVAAYCTTQVERNACLFCKARKLDDKFYQKLRGQFPDAVFWQEISEI FRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCYRRAAELF KNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGF EISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQKSPKPISLLLSTQR RKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSI DVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISDNDLFHFNKKMFAR RRILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLIERWANRIADFFIK NKVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQNKIEFKLKQYGIEIR KVAPNNTSQRCSKCGHLNNYFNFEYRKKNKFPHFKCEKCNFKRNAAYNA ALNISNPKLKSTKERP
mD22	98	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKNNREKIALEKNKDKVKEACS KHLKVAAYCTTQVERNACLFCKARKLDDKFYQKLRGQFPDAVFWQEISEI FRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCYRRAAELF KNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGF EISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQKSPKPISLLLSTQR RKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSI DVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISDNDLFHFNKKMFAR RRILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLIERWSRFIADFFIKN KVGTVQMEDLESMKRKEDSYFNIRLRGFWPYYEMQNKIEFKLKQYGIEIRK VAPNNTSQRCSKCGHLNNYFNFEYRKKNKFPHFKCEKCNFKENAAYNAAR NISNPNIKSTKERP

cB2	99	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKNNREKIALEKNKDKVKEACS KHLKVAAYCTTQVERNACLFCKARKLDDKFYQKLRGQFPDAVFWQEISEI FRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCYRRAAELF KNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGF EISNHNSDFIIDVQLYSKEYSKDSGKGTHRYFLLSTQRRKRNKGWSKDEGT EAEIKKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSIDVPKIDKGVDPSIIG GIAVGVRSPLVCAINNAFSRYSISDNDLFHFNKKMFARRRILLKKNRHKRA GHGAKNKLKPITILTEKSERFRKKLIERWACEIADFFIKNKVGTVQMENLES MKRKEDSYFNIRLRGFWPYAEMQNKIEFKLKQYGIEIRKVAPNNTSKTCSK CGHLNNYFNFEYRKKNKFPHFKCEKCNFKENAAYNAALNISNPKLKSTKE RP
cB3	100	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKNNREKIALEKNKDKVKEACS KHLKVAAYCTTQVERNACLFCKARKLDDKFYQKLRGQFPDAVFWQEISEI FRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCYRRAAELF KNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGF EISNHNSDFIASLSLLSNPAKQEMNVKRKISLLLSTQRRKRNKGWSKDEGTE AEIKKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSIDVPKIDKGVDPSIIGG IAVGVRSPLVCAINNAFSRYSISDNDLFHFNKKMFARRRILLKKNRHKRAG HGAKNKLKPITILTEKSERFRKKLIERWACEIADFFIKNKVGTVQMENLESM KRKEDSYFNIRLRGFWPYAEMQNKIEFKLKQYGIEIRKVAPNNTSKTCSKC GHLNNYFNFEYRKKNKFPHFKCEKCNFKENAAYNAALNISNPKLKSTKERP
cD1	101	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKNNREKIALEKNKDKVKEACS KHLKVAAYCTTQVERNACLFCKARKLDDKFYQKLRGQFPDAVFWQEISEI FRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCYRRAAELF KNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGF EISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQKSPKPISLLLSTQR RKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSI DVPIYERKPNRSIVGGLAVGIRSPLVCAINNSFSRYSVDSNDVFKFSKQVFAF RRRLLSKNSLKRKGHGAAHKLEPITEMTEKNDKFRKKIIERWAKEVTNFFV KNQVGIVQIEDLSTMKDREDHFFNQYLRGFWPYYQMQTLIENKLKEYGIEV KRVQAKYTSQLCSNPNCRYWNNYFNFEYRKVNKFPKFKCEKCNLEISAAY NAARNLSTPDIEKFVAKATKGINLPEK

cD2	102	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKNNREKIALEKNKDKVKEACS KHLKVAAYCTTQVERNACLFCKARKLDDKFYQKLRGQFPDAVFWQEISEI FRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCYRRAAELF KNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGF EISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQKSPKPISLLLSTQR RKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSI DVPTHETVLDPDKVMGVALGVAKAVYWAFNSSYKRGCIDGGEIEHFRKMI RARRVSIQNQIKHSGDARKGHGRKRALKPIETLSEKEKNFRDTINHRYANRI VEAAIKQGCGTIQIENLEGIADTTGSKFLKNWPYYDLQTKIVNKAKEHGITV VAINPQYTSQRCSMCGYIEKTNRSSQAVFECKQCGYGSRTICINCRHVQVSG DVCEECGGIVKKENVNAAYNAAKNISTPYIDQIIMEKCLELGIPYRSITCKEC GHIQASGNTCEVCGSTNILKPKK
cD3	103	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKNNREKIALEKNKDKVKEACS KHLKVAAYCTTQVERNACLFCKARKLDDKFYQKLRGQFPDAVFWQEISEI FRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCYRRAAELF KNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGF EISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQKSPKPISLLLSTQR RKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSI DVPQTRVLDLNKIMGIALGVAVAVYMAFQHTPARYKLEGGEIENFRRQVE SRRISMLRQGKYAGGARGGHGRDKRIKPIEQLRDKIANFRDTTNHRYSRYI VDMAIKEGCGTIQMEDLTNIRDIGSRFLQNWTYYDLQQKIIYKAEEAGIKVI KIDPQYTSQRCSECGNIDSGNRIGQAIFKCRACGYEANAAYNAARNIAIPNI DKIIAESIK
cD4	104	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKNNREKIALEKNKDKVKEACS KHLKVAAYCTTQVERNACLFCKARKLDDKFYQKLRGQFPDAVFWQEISEI FRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCYRRAAELF KNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGF EISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQKSPKPISLLLSTQR RKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSI DVPIDIVNKKVSGRVVGVALGLKIPAYCALNDVEYIKKSIGRIDDFLKVRTQ MQSRRRRLQIAIQSAKGGKGRVNKLQALERFAEKEKNFAKTYNHFLSSNIV KFAVSNQAEQINMELLSLKETQNKSILRNWSYYQLQTMIEYKAQREGIKVK YIDPYHTSQTCSKCGNYEEGQRESQADFICKKCGYKVNAAYNAARNIAMS NKYITKKEESKYYKIKESMV

cD5	105	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKNNREKIALEKNKDKVKEACS KHLKVAAYCTTQVERNACLFCKARKLDDKFYQKLRGQFPDAVFWQEISEI FRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCYRRAAELF KNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGF EISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQKSPKPISLLLSTQR RKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSI DVETKETALDPNNVMGVALGIVYPVYIAFNNSLHRYHIKGGEIERFRRQVE KRKRELLNQGKYCGDGRKGHGYATRTKSIESISDKIARFRDTCNHKYSRFIV DMALKHNCGIIQMEDLTGISKESTFLKNWTYYDLQQKIEYKAREAGIQVIKI EPQYTSQRCSKCGYIDKENRQEQATFKCIECGFKTNAAYNAARNIAIPNIDK IIRKTLKMQ
cD6	106	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKNNREKIALEKNKDKVKEACS KHLKVAAYCTTQVERNACLFCKARKLDDKFYQKLRGQFPDAVFWQEISEI FRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCYRRAAELF KNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGF EISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQKSPKPISLLLSTQR RKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSI DVPKIDKGVNRSIVGGLAVGIRSPLVCAINNSFSRYSVDSNDVFKFSKQVFA FRRRLLSKNSLKRKGHGAAHKLEPITEMTEKNDKFRKKIIERWAKEVTNFF VKNQVGIVQIEDLSTMKDREDHFFNQYLRGFWPYYQMQTLIENKLKEYGIE VKRVQAKYTSQLCSNPNCRYWNNYFNFEYRKVNKFPKFKCEKCNLEISAA YNAARNLSTPDIEKFVAKATKGINLPEK
cD7	107	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKNNREKIALEKNKDKVKEACS KHLKVAAYCTTQVERNACLFCKARKLDDKFYQKLRGQFPDAVFWQEISEI FRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCYRRAAELF KNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGF EISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQKSPKPISLLLSTQR RKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSI DVPKIDKGVDPDKVMGVALGVAKAVYWAFNSSYKRGCIDGGEIEHFRKMI RARRVSIQNQIKHSGDARKGHGRKRALKPIETLSEKEKNFRDTINHRYANRI VEAAIKQGCGTIQIENLEGIADTTGSKFLKNWPYYDLQTKIVNKAKEHGITV VAINPQYTSQRCSMCGYIEKTNRSSQAVFECKQCGYGSRTICINCRHVQVSG DVCEECGGIVKKENVNAAYNAAKNISTPYIDQIIMEKCLELGIPYRSITCKEC GHIQASGNTCEVCGSTNILKPKK

cD8	108	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKNNREKIALEKNKDKVKEACS KHLKVAAYCTTQVERNACLFCKARKLDDKFYQKLRGQFPDAVFWQEISEI FRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCYRRAAELF KNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGF EISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQKSPKPISLLLSTQR RKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSI DVPKIDKGVDLNKIMGIALGVAVAVYMAFQHTPARYKLEGGEIENFRRQV ESRRISMLRQGKYAGGARGGHGRDKRIKPIEQLRDKIANFRDTTNHRYSRYI VDMAIKEGCGTIQMEDLTNIRDIGSRFLQNWTYYDLQQKIIYKAEEAGIKVI KIDPQYTSQRCSECGNIDSGNRIGQAIFKCRACGYEANAAYNAARNIAIPNI DKIIAESIK
cD9	109	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKNNREKIALEKNKDKVKEACS KHLKVAAYCTTQVERNACLFCKARKLDDKFYQKLRGQFPDAVFWQEISEI FRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCYRRAAELF KNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGF EISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQKSPKPISLLLSTQR RKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSI DVPKIDKGVDPNNVMGVALGIVYPVYIAFNNSLHRYHIKGGEIERFRRQVE KRKRELLNQGKYCGDGRKGHGYATRTKSIESISDKIARFRDTCNHKYSRFIV DMALKHNCGIIQMEDLTGISKESTFLKNWTYYDLQQKIEYKAREAGIQVIKI EPQYTSQRCSKCGYIDKENRQEQATFKCIECGFKTNAAYNAARNIAIPNIDK IIRKTLKMQ
cD10	110	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKNNREKIALEKNKDKVKEACS KHLKVAAYCTTQVERNACLFCKARKLDDKFYQKLRGQFPDAVFWQEISEI FRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCYRRAAELF KNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGF EISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQKSPKPISLLLSTQR RKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSI DVPKIDKGVDKEKILGVAVGVNCPLVASVFGDRDRFIIKGGEIEKFRKSVEA RRRSMLEQTKYCGDGRIGHGRKKRTEPALNIGDKIARFRDTTNHKYSRALI EYAVKKGCGTIQMEKLTGITSKSDRFLKDWTYYDLQTKIENKAKEVGINVV YIAPKYTSQRCSKCGYIHKDNRPNQAKFRCLECDFESNAAYNASQNIGIKNI DKIIEKDLQKQESEVQVNENK

t1	111	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKNNREKIALEKNKDKVKEACS KHLKVAAYCTTQVERNACLFCKARKLDDKFYQKLRGQFPDAVFWQEISEI FRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCYRRAAELF KNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGF EISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQKSPKPISLLLSTQR RKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSI DVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISDNDLFHFNKKMFAR RRILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLIERWACEIADFFIKN KVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQNKIEFKLKQYGIEIRK VAPNNTSKTCSKCGHLNNYFNFEYRKKNKFPHFKCEKCNFKENAAYNAAL NI

表5A.工程化核酸酶变体及其相对于SEQ ID NO:1或SEQ ID NO:10的修饰组。

工程化核酸酶变体	**相对于 Un1Cas12f1 (SEQ ID NO: 1)或 dCasMINI (SEQ ID NO: 10)的氨基酸序列的非限制性修饰组，与 dCasMINI (SEQ ID NO: 10)相比具有增强的活性**
cA2	缺失：氨基酸 39-72；突变：N32E、N33R、E35K、K36Q、I37A、A38G、K73G、A74T、R75G、K76E
t1	缺失：氨基酸 518-529
mA8	突变：V23I、N33R、E40D、Q83K 和 G87K
mA10	突变：A21Q、V23I、D29E、N33R 和 E40D
mC16	突变：A340S、H353K、A374K 和 I387E
mD2	突变：N423D、K473Q、T474L、K521D 和 L522I
mD4	突变：K473Q、T474L、H497K、K521D 和 L522I
mD7	突变：N423D、K473Q、T474L、H497K、N519T、K521D 和 L522I
mD15	突变：N423D、K473Q、T474R、L515R、K521N 和 L522I
附加变体 1	突变：E151A
附加变体 2	突变：N423D
附加变体 3	突变：K473Q 和 T474L
附加变体 4	突变：K521D 和 L522I
附加变体 5	突变：K473Q、T474L、K521D 和 L522I

表5B.通过将不同的修饰组合移植到“cA2”起始序列上而产生的工程化核酸酶变体,以及与dCasMINI(SEQ ID NO:10)相比其在增强靶基因活性方面的功效。

工程化核酸酶变体	修饰的组合，来自表 5A	与 dCasMINI（SEQ ID NO: 10）相比增强的活性？	尺寸(氨基酸)
cA2	-	是	495
cA2.1	cA2, E151A, mC16, mD2	是	495
cA2.2	cA2, E151A, mC16, mD4	是	495
cA2.3	cA2, E151A, mC16, mD7	是	495
cA2.4	cA2, E151A, mC16, mD15	是	495
cA2.5	cA2, E151A, mC16	是	495
cA2.6	cA2, E151A, mC16, mD2, t1	是	483
cA2.7	cA2, E151A, mC16, mD4, t1	是	483
cA2.8	cA2, E151A, mC16, mD7, t1	是	483
cA2.9	cA2, E151A, mC16, mD15 ,t1	是	483
cA2.10	cA2, E151A, mC16, t1	是	483
cA2.11	cA2, E151A, mD2	是	495
cA2.12	cA2, E151A, mD4	是	495
cA2.13	cA2, E151A, mD7	是	495
cA2.14	cA2, E151A, mD15	是	495
cA2.15	cA2, E151A	是	495
cA2.16	cA2, E151A, mD2, t1	是	483
cA2.17	cA2, E151A, mD4, t1	是	483
cA2.18	cA2, E151A, mD7, t1	是	483
cA2.19	cA2, E151A, mD15 ,t1	是	483
cA2.20	cA2, E151A, t1	是	483
cA2.21	cA2, mC16, mD2	是	495
cA2.22	cA2, mC16, mD4	是	495
cA2.23	cA2, mC16, mD7	是	495
cA2.24	cA2, mC16, mD15	是	495
cA2.25	cA2, mC16	是	495

cA2.26	cA2, mC16, mD2, t1	是	483
cA2.27	cA2, mC16, mD4, t1	是	483
cA2.28	cA2, mC16, mD7, t1	是	483
cA2.29	cA2, mC16, mD15 ,t1	是	483
cA2.30	cA2, mC16, t1	是	483
cA2.31	cA2, mD2	是	495
cA2.32	cA2, mD4	是	495
cA2.33	cA2, mD7	是	495
cA2.34	cA2, mD15	是	495
cA2.36	cA2, mD2, t1	是	483
cA2.37	cA2, mD4, t1	是	483
cA2.38	cA2, mD7, t1	是	483
cA2.39	cA2, mD15, t1	是	483
cA2.40	cA2, t1	是	483
cA2.41	cA2, mA10, E151A, mC16, mD2	是	495
cA2.42	cA2, mA10, E151A, mC16, mD4	是	495
cA2.43	cA2, mA10, E151A, mC16, mD7	是	495
cA2.44	cA2, mA10, E151A, mC16, mD15	是	495
cA2.45	cA2, mA10, E151A, mC16	是	495
cA2.46	cA2, mA10, E151A, mC16, mD2, t1	是	483
cA2.47	cA2, mA10, E151A, mC16, mD4, t1	是	483
cA2.48	cA2, mA10, E151A, mC16, mD7, t1	是	483
cA2.49	cA2, mA10, E151A, mC16, mD15 ,t1	是	483
cA2.50	cA2, mA10, E151A, mC16, t1	是	483
cA2.51	cA2, mA10, E151A, mD2	是	495
cA2.52	cA2, mA10, E151A, mD4	是	495
cA2.53	cA2, mA10, E151A, mD7	是	495
cA2.54	cA2, mA10, E151A, mD15	是	495
cA2.55	cA2, mA10, E151A	否	495
cA2.56	cA2, mA10, E151A, mD2, t1	是	483
cA2.57	cA2, mA10, E151A, mD4, t1	是	483
cA2.58	cA2, mA10, E151A, mD7, t1	是	483

cA2.59	cA2, mA10, E151A, mD15 ,t1	是	483
cA2.60	cA2, mA10, E151A, t1	是	483
cA2.61	cA2, mA10, mC16, mD2	是	495
cA2.62	cA2, mA10, mC16, mD4	是	495
cA2.63	cA2, mA10, mC16, mD7	是	495
cA2.64	cA2, mA10, mC16, mD15	是	495
cA2.65	cA2, mA10, mC16	是	495
cA2.66	cA2, mA10, mC16, mD2, t1	是	483
cA2.67	cA2, mA10, mC16, mD4, t1	是	483
cA2.68	cA2, mA10, mC16, mD7, t1	是	483
cA2.69	cA2, mA10, mC16, mD15 ,t1	是	483
cA2.70	cA2, mA10, mC16, t1	是	483
cA2.71	cA2, mA10, mD2	是	495
cA2.72	cA2, mA10, mD4	是	495
cA2.73	cA2, mA10, mD7	是	495
cA2.74	cA2, mA10, mD15	是	495
cA2.75	cA2, mA10	是	495
cA2.76	cA2, mA10, mD2, t1	是	483
cA2.77	cA2, mA10, mD4, t1	是	483
cA2.78	cA2, mA10, mD7, t1	是	483
cA2.79	cA2, mA10, mD15 ,t1	是	483
cA2.80	cA2, mA10, t1	是	483
cA2.81	E151A, mC16, mD2	是	529
cA2.82	E151A, mD2	是	529
cA2.83	cA2, mA8, E151A, mD2	是	495
cA2.84	cA2, mA8, E151A, mD4	否	495
cA2.85	cA2, mA8, mD2	是	495
cA2.86	cA2, mA8, mD4	是	495
cA2.87	cA2, N423D	是	495
cA2.88	cA2, K473Q, T474L	是	495
cA2.89	cA2, K521D, L522I	是	495
cA2.90	cA2, K473Q, T474L, K521D, L522I	是	495

表5C.表5B中列出的每种工程化核酸酶变体的氨基酸序列

工程化核酸酶变体	SEQ ID NO	氨基酸序列
cA2.1	112	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKERRKQAGGTGELDDKFYQKL RGQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEH YLSRVCYRRAAALFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFP IPLVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQV QKSPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGS KICEKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNSFSRYSISD NDLFKFNKKMFARRRILLKKNRHKRKGHGAKNKLKPITELTEKSERFRKKL IERWACEIADFFIKNKVGTVQMEDLESMKRKEDSYFNIRLRGFWPYAEMQN KIEFKLKQYGIEIRKVAPNNTSQLCSKCGHLNNYFNFEYRKKNKFPHFKCEK CNFKENAAYNAALNISNPDIKSTKERP
cA2.2	113	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKERRKQAGGTGELDDKFYQKL RGQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEH YLSRVCYRRAAALFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFP IPLVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQV QKSPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGS KICEKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNSFSRYSISD NDLFKFNKKMFARRRILLKKNRHKRKGHGAKNKLKPITELTEKSERFRKKL IERWACEIADFFIKNKVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQN KIEFKLKQYGIEIRKVAPNNTSQLCSKCGHLNNYFNFEYRKKNKFPKFKCEK CNFKENAAYNAALNISNPDIKSTKERP
cA2.3	114	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKERRKQAGGTGELDDKFYQKL RGQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEH YLSRVCYRRAAALFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFP IPLVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQV QKSPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGS KICEKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNSFSRYSISD NDLFKFNKKMFARRRILLKKNRHKRKGHGAKNKLKPITELTEKSERFRKKL IERWACEIADFFIKNKVGTVQMEDLESMKRKEDSYFNIRLRGFWPYAEMQN KIEFKLKQYGIEIRKVAPNNTSQLCSKCGHLNNYFNFEYRKKNKFPKFKCEK CNFKENAAYNAALNISTPDIKSTKERP

cA2.4	115	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKERRKQAGGTGELDDKFYQKL RGQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEH YLSRVCYRRAAALFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFP IPLVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQV QKSPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGS KICEKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNSFSRYSISD NDLFKFNKKMFARRRILLKKNRHKRKGHGAKNKLKPITELTEKSERFRKKL IERWACEIADFFIKNKVGTVQMEDLESMKRKEDSYFNIRLRGFWPYAEMQN KIEFKLKQYGIEIRKVAPNNTSQRCSKCGHLNNYFNFEYRKKNKFPHFKCEK CNFKENAAYNAARNISNPNIKSTKERP
cA2.5	116	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKERRKQAGGTGELDDKFYQKL RGQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEH YLSRVCYRRAAALFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFP IPLVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQV QKSPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGS KICEKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNSFSRYSISD NDLFKFNKKMFARRRILLKKNRHKRKGHGAKNKLKPITELTEKSERFRKKL IERWACEIADFFIKNKVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQN KIEFKLKQYGIEIRKVAPNNTSKTCSKCGHLNNYFNFEYRKKNKFPHFKCEK CNFKENAAYNAALNISNPKLKSTKERP
cA2.6	117	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKERRKQAGGTGELDDKFYQKL RGQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEH YLSRVCYRRAAALFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFP IPLVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQV QKSPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGS KICEKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNSFSRYSISD NDLFKFNKKMFARRRILLKKNRHKRKGHGAKNKLKPITELTEKSERFRKKL IERWACEIADFFIKNKVGTVQMEDLESMKRKEDSYFNIRLRGFWPYAEMQN KIEFKLKQYGIEIRKVAPNNTSQLCSKCGHLNNYFNFEYRKKNKFPHFKCEK CNFKENAAYNAALNI
cA2.7	118	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKERRKQAGGTGELDDKFYQKL RGQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEH YLSRVCYRRAAALFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFP IPLVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQV QKSPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGS KICEKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNSFSRYSISD NDLFKFNKKMFARRRILLKKNRHKRKGHGAKNKLKPITELTEKSERFRKKL

		IERWACEIADFFIKNKVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQN KIEFKLKQYGIEIRKVAPNNTSQLCSKCGHLNNYFNFEYRKKNKFPKFKCEK CNFKENAAYNAALNI
cA2.8	119	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKERRKQAGGTGELDDKFYQKL RGQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEH YLSRVCYRRAAALFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFP IPLVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQV QKSPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGS KICEKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNSFSRYSISD NDLFKFNKKMFARRRILLKKNRHKRKGHGAKNKLKPITELTEKSERFRKKL IERWACEIADFFIKNKVGTVQMEDLESMKRKEDSYFNIRLRGFWPYAEMQN KIEFKLKQYGIEIRKVAPNNTSQLCSKCGHLNNYFNFEYRKKNKFPKFKCEK CNFKENAAYNAALNI
cA2.9	120	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKERRKQAGGTGELDDKFYQKL RGQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEH YLSRVCYRRAAALFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFP IPLVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQV QKSPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGS KICEKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNSFSRYSISD NDLFKFNKKMFARRRILLKKNRHKRKGHGAKNKLKPITELTEKSERFRKKL IERWACEIADFFIKNKVGTVQMEDLESMKRKEDSYFNIRLRGFWPYAEMQN KIEFKLKQYGIEIRKVAPNNTSQRCSKCGHLNNYFNFEYRKKNKFPHFKCEK CNFKENAAYNAARNI
cA2.10	121	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKERRKQAGGTGELDDKFYQKL RGQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEH YLSRVCYRRAAALFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFP IPLVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQV QKSPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGS KICEKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNSFSRYSISD NDLFKFNKKMFARRRILLKKNRHKRKGHGAKNKLKPITELTEKSERFRKKL IERWACEIADFFIKNKVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQN KIEFKLKQYGIEIRKVAPNNTSKTCSKCGHLNNYFNFEYRKKNKFPHFKCEK CNFKENAAYNAALNI

cA2.11	122	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKERRKQAGGTGELDDKFYQKL RGQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEH YLSRVCYRRAAALFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFP IPLVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQV QKSPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGS KICEKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISD NDLFHFNKKMFARRRILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLI ERWACEIADFFIKNKVGTVQMEDLESMKRKEDSYFNIRLRGFWPYAEMQN KIEFKLKQYGIEIRKVAPNNTSQLCSKCGHLNNYFNFEYRKKNKFPHFKCEK CNFKENAAYNAALNISNPDIKSTKERP
cA2.12	123	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKERRKQAGGTGELDDKFYQKL RGQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEH YLSRVCYRRAAALFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFP IPLVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQV QKSPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGS KICEKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISD NDLFHFNKKMFARRRILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLI ERWACEIADFFIKNKVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQN KIEFKLKQYGIEIRKVAPNNTSQLCSKCGHLNNYFNFEYRKKNKFPKFKCEK CNFKENAAYNAALNISNPDIKSTKERP
cA2.13	124	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKERRKQAGGTGELDDKFYQKL RGQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEH YLSRVCYRRAAALFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFP IPLVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQV QKSPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGS KICEKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISD NDLFHFNKKMFARRRILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLI ERWACEIADFFIKNKVGTVQMEDLESMKRKEDSYFNIRLRGFWPYAEMQN KIEFKLKQYGIEIRKVAPNNTSQLCSKCGHLNNYFNFEYRKKNKFPKFKCEK CNFKENAAYNAALNISTPDIKSTKERP
cA2.14	125	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKERRKQAGGTGELDDKFYQKL RGQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEH YLSRVCYRRAAALFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFP IPLVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQV QKSPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGS KICEKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISD NDLFHFNKKMFARRRILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLI

		ERWACEIADFFIKNKVGTVQMEDLESMKRKEDSYFNIRLRGFWPYAEMQN KIEFKLKQYGIEIRKVAPNNTSQRCSKCGHLNNYFNFEYRKKNKFPHFKCEK CNFKENAAYNAARNISNPNIKSTKERP
cA2.15	126	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKERRKQAGGTGELDDKFYQKL RGQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEH YLSRVCYRRAAALFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFP IPLVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQV QKSPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGS KICEKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISD NDLFHFNKKMFARRRILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLI ERWACEIADFFIKNKVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQN KIEFKLKQYGIEIRKVAPNNTSKTCSKCGHLNNYFNFEYRKKNKFPHFKCEK CNFKENAAYNAALNISNPKLKSTKERP
cA2.16	127	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKERRKQAGGTGELDDKFYQKL RGQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEH YLSRVCYRRAAALFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFP IPLVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQV QKSPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGS KICEKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISD NDLFHFNKKMFARRRILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLI ERWACEIADFFIKNKVGTVQMEDLESMKRKEDSYFNIRLRGFWPYAEMQN KIEFKLKQYGIEIRKVAPNNTSQLCSKCGHLNNYFNFEYRKKNKFPHFKCEK CNFKENAAYNAALNI
cA2.17	128	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKERRKQAGGTGELDDKFYQKL RGQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEH YLSRVCYRRAAALFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFP IPLVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQV QKSPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGS KICEKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISD NDLFHFNKKMFARRRILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLI ERWACEIADFFIKNKVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQN KIEFKLKQYGIEIRKVAPNNTSQLCSKCGHLNNYFNFEYRKKNKFPKFKCEK CNFKENAAYNAALNI

cA2.18	129	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKERRKQAGGTGELDDKFYQKL RGQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEH YLSRVCYRRAAALFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFP IPLVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQV QKSPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGS KICEKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISD NDLFHFNKKMFARRRILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLI ERWACEIADFFIKNKVGTVQMEDLESMKRKEDSYFNIRLRGFWPYAEMQN KIEFKLKQYGIEIRKVAPNNTSQLCSKCGHLNNYFNFEYRKKNKFPKFKCEK CNFKENAAYNAALNI
cA2.19	130	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKERRKQAGGTGELDDKFYQKL RGQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEH YLSRVCYRRAAALFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFP IPLVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQV QKSPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGS KICEKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISD NDLFHFNKKMFARRRILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLI ERWACEIADFFIKNKVGTVQMEDLESMKRKEDSYFNIRLRGFWPYAEMQN KIEFKLKQYGIEIRKVAPNNTSQRCSKCGHLNNYFNFEYRKKNKFPHFKCEK CNFKENAAYNAARNI
cA2.20	131	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKERRKQAGGTGELDDKFYQKL RGQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEH YLSRVCYRRAAALFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFP IPLVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQV QKSPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGS KICEKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISD NDLFHFNKKMFARRRILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLI ERWACEIADFFIKNKVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQN KIEFKLKQYGIEIRKVAPNNTSKTCSKCGHLNNYFNFEYRKKNKFPHFKCEK CNFKENAAYNAALNI
cA2.21	132	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKERRKQAGGTGELDDKFYQKL RGQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEH YLSRVCYRRAAELFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFP IPLVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQV QKSPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGS KICEKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNSFSRYSISD NDLFKFNKKMFARRRILLKKNRHKRKGHGAKNKLKPITELTEKSERFRKKL

		IERWACEIADFFIKNKVGTVQMEDLESMKRKEDSYFNIRLRGFWPYAEMQN KIEFKLKQYGIEIRKVAPNNTSQLCSKCGHLNNYFNFEYRKKNKFPHFKCEK CNFKENAAYNAALNISNPDIKSTKERP
cA2.22	133	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKERRKQAGGTGELDDKFYQKL RGQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEH YLSRVCYRRAAELFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFP IPLVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQV QKSPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGS KICEKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNSFSRYSISD NDLFKFNKKMFARRRILLKKNRHKRKGHGAKNKLKPITELTEKSERFRKKL IERWACEIADFFIKNKVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQN KIEFKLKQYGIEIRKVAPNNTSQLCSKCGHLNNYFNFEYRKKNKFPKFKCEK CNFKENAAYNAALNISNPDIKSTKERP
cA2.23	134	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKERRKQAGGTGELDDKFYQKL RGQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEH YLSRVCYRRAAELFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFP IPLVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQV QKSPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGS KICEKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNSFSRYSISD NDLFKFNKKMFARRRILLKKNRHKRKGHGAKNKLKPITELTEKSERFRKKL IERWACEIADFFIKNKVGTVQMEDLESMKRKEDSYFNIRLRGFWPYAEMQN KIEFKLKQYGIEIRKVAPNNTSQLCSKCGHLNNYFNFEYRKKNKFPKFKCEK CNFKENAAYNAALNISTPDIKSTKERP
cA2.24	135	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKERRKQAGGTGELDDKFYQKL RGQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEH YLSRVCYRRAAELFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFP IPLVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQV QKSPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGS KICEKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNSFSRYSISD NDLFKFNKKMFARRRILLKKNRHKRKGHGAKNKLKPITELTEKSERFRKKL IERWACEIADFFIKNKVGTVQMEDLESMKRKEDSYFNIRLRGFWPYAEMQN KIEFKLKQYGIEIRKVAPNNTSQRCSKCGHLNNYFNFEYRKKNKFPHFKCEK CNFKENAAYNAARNISNPNIKSTKERP

cA2.25	136	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKERRKQAGGTGELDDKFYQKL RGQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEH YLSRVCYRRAAELFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFP IPLVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQV QKSPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGS KICEKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNSFSRYSISD NDLFKFNKKMFARRRILLKKNRHKRKGHGAKNKLKPITELTEKSERFRKKL IERWACEIADFFIKNKVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQN KIEFKLKQYGIEIRKVAPNNTSKTCSKCGHLNNYFNFEYRKKNKFPHFKCEK CNFKENAAYNAALNISNPKLKSTKERP
cA2.26	137	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKERRKQAGGTGELDDKFYQKL RGQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEH YLSRVCYRRAAELFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFP IPLVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQV QKSPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGS KICEKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNSFSRYSISD NDLFKFNKKMFARRRILLKKNRHKRKGHGAKNKLKPITELTEKSERFRKKL IERWACEIADFFIKNKVGTVQMEDLESMKRKEDSYFNIRLRGFWPYAEMQN KIEFKLKQYGIEIRKVAPNNTSQLCSKCGHLNNYFNFEYRKKNKFPHFKCEK CNFKENAAYNAALNI
cA2.27	138	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKERRKQAGGTGELDDKFYQKL RGQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEH YLSRVCYRRAAELFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFP IPLVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQV QKSPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGS KICEKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNSFSRYSISD NDLFKFNKKMFARRRILLKKNRHKRKGHGAKNKLKPITELTEKSERFRKKL IERWACEIADFFIKNKVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQN KIEFKLKQYGIEIRKVAPNNTSQLCSKCGHLNNYFNFEYRKKNKFPKFKCEK CNFKENAAYNAALNI
cA2.28	139	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKERRKQAGGTGELDDKFYQKL RGQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEH YLSRVCYRRAAELFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFP IPLVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQV QKSPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGS KICEKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNSFSRYSISD NDLFKFNKKMFARRRILLKKNRHKRKGHGAKNKLKPITELTEKSERFRKKL

		IERWACEIADFFIKNKVGTVQMEDLESMKRKEDSYFNIRLRGFWPYAEMQN KIEFKLKQYGIEIRKVAPNNTSQLCSKCGHLNNYFNFEYRKKNKFPKFKCEK CNFKENAAYNAALNI
cA2.29	140	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKERRKQAGGTGELDDKFYQKL RGQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEH YLSRVCYRRAAELFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFP IPLVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQV QKSPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGS KICEKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNSFSRYSISD NDLFKFNKKMFARRRILLKKNRHKRKGHGAKNKLKPITELTEKSERFRKKL IERWACEIADFFIKNKVGTVQMEDLESMKRKEDSYFNIRLRGFWPYAEMQN KIEFKLKQYGIEIRKVAPNNTSQRCSKCGHLNNYFNFEYRKKNKFPHFKCEK CNFKENAAYNAARNI
cA2.30	141	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKERRKQAGGTGELDDKFYQKL RGQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEH YLSRVCYRRAAELFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFP IPLVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQV QKSPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGS KICEKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNSFSRYSISD NDLFKFNKKMFARRRILLKKNRHKRKGHGAKNKLKPITELTEKSERFRKKL IERWACEIADFFIKNKVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQN KIEFKLKQYGIEIRKVAPNNTSKTCSKCGHLNNYFNFEYRKKNKFPHFKCEK CNFKENAAYNAALNI
cA2.31	142	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKERRKQAGGTGELDDKFYQKL RGQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEH YLSRVCYRRAAELFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFP IPLVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQV QKSPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGS KICEKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISD NDLFHFNKKMFARRRILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLI ERWACEIADFFIKNKVGTVQMEDLESMKRKEDSYFNIRLRGFWPYAEMQN KIEFKLKQYGIEIRKVAPNNTSQLCSKCGHLNNYFNFEYRKKNKFPHFKCEK CNFKENAAYNAALNISNPDIKSTKERP

cA2.32	143	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKERRKQAGGTGELDDKFYQKL RGQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEH YLSRVCYRRAAELFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFP IPLVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQV QKSPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGS KICEKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISD NDLFHFNKKMFARRRILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLI ERWACEIADFFIKNKVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQN KIEFKLKQYGIEIRKVAPNNTSQLCSKCGHLNNYFNFEYRKKNKFPKFKCEK CNFKENAAYNAALNISNPDIKSTKERP
cA2.33	144	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKERRKQAGGTGELDDKFYQKL RGQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEH YLSRVCYRRAAELFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFP IPLVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQV QKSPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGS KICEKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISD NDLFHFNKKMFARRRILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLI ERWACEIADFFIKNKVGTVQMEDLESMKRKEDSYFNIRLRGFWPYAEMQN KIEFKLKQYGIEIRKVAPNNTSQLCSKCGHLNNYFNFEYRKKNKFPKFKCEK CNFKENAAYNAALNISTPDIKSTKERP
cA2.34	145	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKERRKQAGGTGELDDKFYQKL RGQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEH YLSRVCYRRAAELFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFP IPLVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQV QKSPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGS KICEKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISD NDLFHFNKKMFARRRILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLI ERWACEIADFFIKNKVGTVQMEDLESMKRKEDSYFNIRLRGFWPYAEMQN KIEFKLKQYGIEIRKVAPNNTSQRCSKCGHLNNYFNFEYRKKNKFPHFKCEK CNFKENAAYNAARNISNPNIKSTKERP
cA2.35	146	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKERRKQAGGTGELDDKFYQKL RGQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEH YLSRVCYRRAAELFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFP IPLVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQV QKSPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGS KICEKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISD NDLFHFNKKMFARRRILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLI

		ERWACEIADFFIKNKVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQN KIEFKLKQYGIEIRKVAPNNTSKTCSKCGHLNNYFNFEYRKKNKFPHFKCEK CNFKENAAYNAALNISNPKLKSTKERP
cA2.36	147	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKERRKQAGGTGELDDKFYQKL RGQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEH YLSRVCYRRAAELFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFP IPLVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQV QKSPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGS KICEKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISD NDLFHFNKKMFARRRILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLI ERWACEIADFFIKNKVGTVQMEDLESMKRKEDSYFNIRLRGFWPYAEMQN KIEFKLKQYGIEIRKVAPNNTSQLCSKCGHLNNYFNFEYRKKNKFPHFKCEK CNFKENAAYNAALNI
cA2.37	148	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKERRKQAGGTGELDDKFYQKL RGQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEH YLSRVCYRRAAELFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFP IPLVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQV QKSPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGS KICEKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISD NDLFHFNKKMFARRRILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLI ERWACEIADFFIKNKVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQN KIEFKLKQYGIEIRKVAPNNTSQLCSKCGHLNNYFNFEYRKKNKFPKFKCEK CNFKENAAYNAALNI
cA2.38	149	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKERRKQAGGTGELDDKFYQKL RGQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEH YLSRVCYRRAAELFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFP IPLVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQV QKSPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGS KICEKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISD NDLFHFNKKMFARRRILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLI ERWACEIADFFIKNKVGTVQMEDLESMKRKEDSYFNIRLRGFWPYAEMQN KIEFKLKQYGIEIRKVAPNNTSQLCSKCGHLNNYFNFEYRKKNKFPKFKCEK CNFKENAAYNAALNI

cA2.39	150	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKERRKQAGGTGELDDKFYQKL RGQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEH YLSRVCYRRAAELFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFP IPLVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQV QKSPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGS KICEKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISD NDLFHFNKKMFARRRILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLI ERWACEIADFFIKNKVGTVQMEDLESMKRKEDSYFNIRLRGFWPYAEMQN KIEFKLKQYGIEIRKVAPNNTSQRCSKCGHLNNYFNFEYRKKNKFPHFKCEK CNFKENAAYNAARNI
cA2.40	151	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKERRKQAGGTGELDDKFYQKL RGQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEH YLSRVCYRRAAELFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFP IPLVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQV QKSPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGS KICEKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISD NDLFHFNKKMFARRRILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLI ERWACEIADFFIKNKVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQN KIEFKLKQYGIEIRKVAPNNTSKTCSKCGHLNNYFNFEYRKKNKFPHFKCEK CNFKENAAYNAALNI
cA2.41	152	MAKNTITKTLKLRIVRPYNSQEIEKIVAEEKERRKQAGGTGELDDKFYQKLR GQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHY LSRVCYRRAAALFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPI PLVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQ KSPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKI CEKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNSFSRYSISDND LFKFNKKMFARRRILLKKNRHKRKGHGAKNKLKPITELTEKSERFRKKLIER WACEIADFFIKNKVGTVQMEDLESMKRKEDSYFNIRLRGFWPYAEMQNKIE FKLKQYGIEIRKVAPNNTSQLCSKCGHLNNYFNFEYRKKNKFPHFKCEKCN FKENAAYNAALNISNPDIKSTKERP
cA2.42	153	MAKNTITKTLKLRIVRPYNSQEIEKIVAEEKERRKQAGGTGELDDKFYQKLR GQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHY LSRVCYRRAAALFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPI PLVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQ KSPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKI CEKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNSFSRYSISDND LFKFNKKMFARRRILLKKNRHKRKGHGAKNKLKPITELTEKSERFRKKLIER

		WACEIADFFIKNKVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQNKIE FKLKQYGIEIRKVAPNNTSQLCSKCGHLNNYFNFEYRKKNKFPKFKCEKCN FKENAAYNAALNISNPDIKSTKERP
cA2.43	154	MAKNTITKTLKLRIVRPYNSQEIEKIVAEEKERRKQAGGTGELDDKFYQKLR GQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHY LSRVCYRRAAALFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPI PLVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQ KSPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKI CEKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNSFSRYSISDND LFKFNKKMFARRRILLKKNRHKRKGHGAKNKLKPITELTEKSERFRKKLIER WACEIADFFIKNKVGTVQMEDLESMKRKEDSYFNIRLRGFWPYAEMQNKIE FKLKQYGIEIRKVAPNNTSQLCSKCGHLNNYFNFEYRKKNKFPKFKCEKCN FKENAAYNAALNISTPDIKSTKERP
cA2.44	155	MAKNTITKTLKLRIVRPYNSQEIEKIVAEEKERRKQAGGTGELDDKFYQKLR GQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHY LSRVCYRRAAALFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPI PLVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQ KSPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKI CEKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNSFSRYSISDND LFKFNKKMFARRRILLKKNRHKRKGHGAKNKLKPITELTEKSERFRKKLIER WACEIADFFIKNKVGTVQMEDLESMKRKEDSYFNIRLRGFWPYAEMQNKIE FKLKQYGIEIRKVAPNNTSQRCSKCGHLNNYFNFEYRKKNKFPHFKCEKCN FKENAAYNAARNISNPNIKSTKERP
cA2.45	156	MAKNTITKTLKLRIVRPYNSQEIEKIVAEEKERRKQAGGTGELDDKFYQKLR GQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHY LSRVCYRRAAALFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPI PLVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQ KSPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKI CEKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNSFSRYSISDND LFKFNKKMFARRRILLKKNRHKRKGHGAKNKLKPITELTEKSERFRKKLIER WACEIADFFIKNKVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQNKIE FKLKQYGIEIRKVAPNNTSKTCSKCGHLNNYFNFEYRKKNKFPHFKCEKCN FKENAAYNAALNISNPKLKSTKERP

cA2.46	157	MAKNTITKTLKLRIVRPYNSQEIEKIVAEEKERRKQAGGTGELDDKFYQKLR GQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHY LSRVCYRRAAALFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPI PLVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQ KSPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKI CEKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNSFSRYSISDND LFKFNKKMFARRRILLKKNRHKRKGHGAKNKLKPITELTEKSERFRKKLIER WACEIADFFIKNKVGTVQMEDLESMKRKEDSYFNIRLRGFWPYAEMQNKIE FKLKQYGIEIRKVAPNNTSQLCSKCGHLNNYFNFEYRKKNKFPHFKCEKCN FKENAAYNAALNI
cA2.47	158	MAKNTITKTLKLRIVRPYNSQEIEKIVAEEKERRKQAGGTGELDDKFYQKLR GQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHY LSRVCYRRAAALFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPI PLVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQ KSPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKI CEKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNSFSRYSISDND LFKFNKKMFARRRILLKKNRHKRKGHGAKNKLKPITELTEKSERFRKKLIER WACEIADFFIKNKVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQNKIE FKLKQYGIEIRKVAPNNTSQLCSKCGHLNNYFNFEYRKKNKFPKFKCEKCN FKENAAYNAALNI
cA2.48	159	MAKNTITKTLKLRIVRPYNSQEIEKIVAEEKERRKQAGGTGELDDKFYQKLR GQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHY LSRVCYRRAAALFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPI PLVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQ KSPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKI CEKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNSFSRYSISDND LFKFNKKMFARRRILLKKNRHKRKGHGAKNKLKPITELTEKSERFRKKLIER WACEIADFFIKNKVGTVQMEDLESMKRKEDSYFNIRLRGFWPYAEMQNKIE FKLKQYGIEIRKVAPNNTSQLCSKCGHLNNYFNFEYRKKNKFPKFKCEKCN FKENAAYNAALNI
cA2.49	160	MAKNTITKTLKLRIVRPYNSQEIEKIVAEEKERRKQAGGTGELDDKFYQKLR GQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHY LSRVCYRRAAALFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPI PLVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQ KSPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKI CEKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNSFSRYSISDND LFKFNKKMFARRRILLKKNRHKRKGHGAKNKLKPITELTEKSERFRKKLIER

		WACEIADFFIKNKVGTVQMEDLESMKRKEDSYFNIRLRGFWPYAEMQNKIE FKLKQYGIEIRKVAPNNTSQRCSKCGHLNNYFNFEYRKKNKFPHFKCEKCN FKENAAYNAARNI
cA2.50	161	MAKNTITKTLKLRIVRPYNSQEIEKIVAEEKERRKQAGGTGELDDKFYQKLR GQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHY LSRVCYRRAAALFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPI PLVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQ KSPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKI CEKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNSFSRYSISDND LFKFNKKMFARRRILLKKNRHKRKGHGAKNKLKPITELTEKSERFRKKLIER WACEIADFFIKNKVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQNKIE FKLKQYGIEIRKVAPNNTSKTCSKCGHLNNYFNFEYRKKNKFPHFKCEKCN FKENAAYNAALNI
cA2.51	162	MAKNTITKTLKLRIVRPYNSQEIEKIVAEEKERRKQAGGTGELDDKFYQKLR GQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHY LSRVCYRRAAALFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPI PLVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQ KSPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKI CEKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISDN DLFHFNKKMFARRRILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLIE RWACEIADFFIKNKVGTVQMEDLESMKRKEDSYFNIRLRGFWPYAEMQNKI EFKLKQYGIEIRKVAPNNTSQLCSKCGHLNNYFNFEYRKKNKFPHFKCEKC NFKENAAYNAALNISNPDIKSTKERP
cA2.52	163	MAKNTITKTLKLRIVRPYNSQEIEKIVAEEKERRKQAGGTGELDDKFYQKLR GQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHY LSRVCYRRAAALFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPI PLVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQ KSPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKI CEKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISDN DLFHFNKKMFARRRILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLIE RWACEIADFFIKNKVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQNKI EFKLKQYGIEIRKVAPNNTSQLCSKCGHLNNYFNFEYRKKNKFPKFKCEKC NFKENAAYNAALNISNPDIKSTKERP

cA2.53	164	MAKNTITKTLKLRIVRPYNSQEIEKIVAEEKERRKQAGGTGELDDKFYQKLR GQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHY LSRVCYRRAAALFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPI PLVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQ KSPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKI CEKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISDN DLFHFNKKMFARRRILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLIE RWACEIADFFIKNKVGTVQMEDLESMKRKEDSYFNIRLRGFWPYAEMQNKI EFKLKQYGIEIRKVAPNNTSQLCSKCGHLNNYFNFEYRKKNKFPKFKCEKC NFKENAAYNAALNISTPDIKSTKERP
cA2.54	165	MAKNTITKTLKLRIVRPYNSQEIEKIVAEEKERRKQAGGTGELDDKFYQKLR GQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHY LSRVCYRRAAALFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPI PLVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQ KSPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKI CEKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISDN DLFHFNKKMFARRRILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLIE RWACEIADFFIKNKVGTVQMEDLESMKRKEDSYFNIRLRGFWPYAEMQNKI EFKLKQYGIEIRKVAPNNTSQRCSKCGHLNNYFNFEYRKKNKFPHFKCEKC NFKENAAYNAARNISNPNIKSTKERP
cA2.55	166	MAKNTITKTLKLRIVRPYNSQEIEKIVAEEKERRKQAGGTGELDDKFYQKLR GQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHY LSRVCYRRAAALFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPI PLVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQ KSPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKI CEKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISDN DLFHFNKKMFARRRILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLIE RWACEIADFFIKNKVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQNKI EFKLKQYGIEIRKVAPNNTSKTCSKCGHLNNYFNFEYRKKNKFPHFKCEKC NFKENAAYNAALNISNPKLKSTKERP
cA2.56	167	MAKNTITKTLKLRIVRPYNSQEIEKIVAEEKERRKQAGGTGELDDKFYQKLR GQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHY LSRVCYRRAAALFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPI PLVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQ KSPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKI CEKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISDN DLFHFNKKMFARRRILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLIE

		RWACEIADFFIKNKVGTVQMEDLESMKRKEDSYFNIRLRGFWPYAEMQNKI EFKLKQYGIEIRKVAPNNTSQLCSKCGHLNNYFNFEYRKKNKFPHFKCEKC NFKENAAYNAALNI
cA2.57	168	MAKNTITKTLKLRIVRPYNSQEIEKIVAEEKERRKQAGGTGELDDKFYQKLR GQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHY LSRVCYRRAAALFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPI PLVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQ KSPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKI CEKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISDN DLFHFNKKMFARRRILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLIE RWACEIADFFIKNKVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQNKI EFKLKQYGIEIRKVAPNNTSQLCSKCGHLNNYFNFEYRKKNKFPKFKCEKC NFKENAAYNAALNI
cA2.58	169	MAKNTITKTLKLRIVRPYNSQEIEKIVAEEKERRKQAGGTGELDDKFYQKLR GQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHY LSRVCYRRAAALFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPI PLVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQ KSPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKI CEKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISDN DLFHFNKKMFARRRILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLIE RWACEIADFFIKNKVGTVQMEDLESMKRKEDSYFNIRLRGFWPYAEMQNKI EFKLKQYGIEIRKVAPNNTSQLCSKCGHLNNYFNFEYRKKNKFPKFKCEKC NFKENAAYNAALNI
cA2.59	170	MAKNTITKTLKLRIVRPYNSQEIEKIVAEEKERRKQAGGTGELDDKFYQKLR GQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHY LSRVCYRRAAALFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPI PLVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQ KSPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKI CEKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISDN DLFHFNKKMFARRRILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLIE RWACEIADFFIKNKVGTVQMEDLESMKRKEDSYFNIRLRGFWPYAEMQNKI EFKLKQYGIEIRKVAPNNTSQRCSKCGHLNNYFNFEYRKKNKFPHFKCEKC NFKENAAYNAARNI

cA2.60	171	MAKNTITKTLKLRIVRPYNSQEIEKIVAEEKERRKQAGGTGELDDKFYQKLR GQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHY LSRVCYRRAAALFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPI PLVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQ KSPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKI CEKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISDN DLFHFNKKMFARRRILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLIE RWACEIADFFIKNKVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQNKI EFKLKQYGIEIRKVAPNNTSKTCSKCGHLNNYFNFEYRKKNKFPHFKCEKC NFKENAAYNAALNI
cA2.61	172	MAKNTITKTLKLRIVRPYNSQEIEKIVAEEKERRKQAGGTGELDDKFYQKLR GQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHY LSRVCYRRAAELFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIP LVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQK SPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKIC EKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNSFSRYSISDNDL FKFNKKMFARRRILLKKNRHKRKGHGAKNKLKPITELTEKSERFRKKLIER WACEIADFFIKNKVGTVQMEDLESMKRKEDSYFNIRLRGFWPYAEMQNKIE FKLKQYGIEIRKVAPNNTSQLCSKCGHLNNYFNFEYRKKNKFPHFKCEKCN FKENAAYNAALNISNPDIKSTKERP
cA2.62	173	MAKNTITKTLKLRIVRPYNSQEIEKIVAEEKERRKQAGGTGELDDKFYQKLR GQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHY LSRVCYRRAAELFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIP LVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQK SPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKIC EKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNSFSRYSISDNDL FKFNKKMFARRRILLKKNRHKRKGHGAKNKLKPITELTEKSERFRKKLIER WACEIADFFIKNKVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQNKIE FKLKQYGIEIRKVAPNNTSQLCSKCGHLNNYFNFEYRKKNKFPKFKCEKCN FKENAAYNAALNISNPDIKSTKERP
cA2.63	174	MAKNTITKTLKLRIVRPYNSQEIEKIVAEEKERRKQAGGTGELDDKFYQKLR GQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHY LSRVCYRRAAELFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIP LVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQK SPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKIC EKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNSFSRYSISDNDL FKFNKKMFARRRILLKKNRHKRKGHGAKNKLKPITELTEKSERFRKKLIER

		WACEIADFFIKNKVGTVQMEDLESMKRKEDSYFNIRLRGFWPYAEMQNKIE FKLKQYGIEIRKVAPNNTSQLCSKCGHLNNYFNFEYRKKNKFPKFKCEKCN FKENAAYNAALNISTPDIKSTKERP
cA2.64	175	MAKNTITKTLKLRIVRPYNSQEIEKIVAEEKERRKQAGGTGELDDKFYQKLR GQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHY LSRVCYRRAAELFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIP LVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQK SPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKIC EKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNSFSRYSISDNDL FKFNKKMFARRRILLKKNRHKRKGHGAKNKLKPITELTEKSERFRKKLIER WACEIADFFIKNKVGTVQMEDLESMKRKEDSYFNIRLRGFWPYAEMQNKIE FKLKQYGIEIRKVAPNNTSQRCSKCGHLNNYFNFEYRKKNKFPHFKCEKCN FKENAAYNAARNISNPNIKSTKERP
cA2.65	176	MAKNTITKTLKLRIVRPYNSQEIEKIVAEEKERRKQAGGTGELDDKFYQKLR GQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHY LSRVCYRRAAELFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIP LVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQK SPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKIC EKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNSFSRYSISDNDL FKFNKKMFARRRILLKKNRHKRKGHGAKNKLKPITELTEKSERFRKKLIER WACEIADFFIKNKVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQNKIE FKLKQYGIEIRKVAPNNTSKTCSKCGHLNNYFNFEYRKKNKFPHFKCEKCN FKENAAYNAALNISNPKLKSTKERP
cA2.66	177	MAKNTITKTLKLRIVRPYNSQEIEKIVAEEKERRKQAGGTGELDDKFYQKLR GQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHY LSRVCYRRAAELFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIP LVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQK SPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKIC EKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNSFSRYSISDNDL FKFNKKMFARRRILLKKNRHKRKGHGAKNKLKPITELTEKSERFRKKLIER WACEIADFFIKNKVGTVQMEDLESMKRKEDSYFNIRLRGFWPYAEMQNKIE FKLKQYGIEIRKVAPNNTSQLCSKCGHLNNYFNFEYRKKNKFPHFKCEKCN FKENAAYNAALNI

cA2.67	178	MAKNTITKTLKLRIVRPYNSQEIEKIVAEEKERRKQAGGTGELDDKFYQKLR GQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHY LSRVCYRRAAELFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIP LVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQK SPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKIC EKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNSFSRYSISDNDL FKFNKKMFARRRILLKKNRHKRKGHGAKNKLKPITELTEKSERFRKKLIER WACEIADFFIKNKVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQNKIE FKLKQYGIEIRKVAPNNTSQLCSKCGHLNNYFNFEYRKKNKFPKFKCEKCN FKENAAYNAALNI
cA2.68	179	MAKNTITKTLKLRIVRPYNSQEIEKIVAEEKERRKQAGGTGELDDKFYQKLR GQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHY LSRVCYRRAAELFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIP LVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQK SPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKIC EKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNSFSRYSISDNDL FKFNKKMFARRRILLKKNRHKRKGHGAKNKLKPITELTEKSERFRKKLIER WACEIADFFIKNKVGTVQMEDLESMKRKEDSYFNIRLRGFWPYAEMQNKIE FKLKQYGIEIRKVAPNNTSQLCSKCGHLNNYFNFEYRKKNKFPKFKCEKCN FKENAAYNAALNI
cA2.69	180	MAKNTITKTLKLRIVRPYNSQEIEKIVAEEKERRKQAGGTGELDDKFYQKLR GQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHY LSRVCYRRAAELFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIP LVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQK SPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKIC EKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNSFSRYSISDNDL FKFNKKMFARRRILLKKNRHKRKGHGAKNKLKPITELTEKSERFRKKLIER WACEIADFFIKNKVGTVQMEDLESMKRKEDSYFNIRLRGFWPYAEMQNKIE FKLKQYGIEIRKVAPNNTSQRCSKCGHLNNYFNFEYRKKNKFPHFKCEKCN FKENAAYNAARNI
cA2.70	181	MAKNTITKTLKLRIVRPYNSQEIEKIVAEEKERRKQAGGTGELDDKFYQKLR GQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHY LSRVCYRRAAELFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIP LVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQK SPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKIC EKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNSFSRYSISDNDL FKFNKKMFARRRILLKKNRHKRKGHGAKNKLKPITELTEKSERFRKKLIER

		WACEIADFFIKNKVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQNKIE FKLKQYGIEIRKVAPNNTSKTCSKCGHLNNYFNFEYRKKNKFPHFKCEKCN FKENAAYNAALNI
cA2.71	182	MAKNTITKTLKLRIVRPYNSQEIEKIVAEEKERRKQAGGTGELDDKFYQKLR GQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHY LSRVCYRRAAELFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIP LVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQK SPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKIC EKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISDNDL FHFNKKMFARRRILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLIERW ACEIADFFIKNKVGTVQMEDLESMKRKEDSYFNIRLRGFWPYAEMQNKIEF KLKQYGIEIRKVAPNNTSQLCSKCGHLNNYFNFEYRKKNKFPHFKCEKCNF KENAAYNAALNISNPDIKSTKERP
cA2.72	183	MAKNTITKTLKLRIVRPYNSQEIEKIVAEEKERRKQAGGTGELDDKFYQKLR GQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHY LSRVCYRRAAELFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIP LVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQK SPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKIC EKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISDNDL FHFNKKMFARRRILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLIERW ACEIADFFIKNKVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQNKIEF KLKQYGIEIRKVAPNNTSQLCSKCGHLNNYFNFEYRKKNKFPKFKCEKCNF KENAAYNAALNISNPDIKSTKERP
cA2.73	184	MAKNTITKTLKLRIVRPYNSQEIEKIVAEEKERRKQAGGTGELDDKFYQKLR GQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHY LSRVCYRRAAELFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIP LVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQK SPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKIC EKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISDNDL FHFNKKMFARRRILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLIERW ACEIADFFIKNKVGTVQMEDLESMKRKEDSYFNIRLRGFWPYAEMQNKIEF KLKQYGIEIRKVAPNNTSQLCSKCGHLNNYFNFEYRKKNKFPKFKCEKCNF KENAAYNAALNISTPDIKSTKERP

cA2.74	185	MAKNTITKTLKLRIVRPYNSQEIEKIVAEEKERRKQAGGTGELDDKFYQKLR GQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHY LSRVCYRRAAELFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIP LVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQK SPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKIC EKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISDNDL FHFNKKMFARRRILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLIERW ACEIADFFIKNKVGTVQMEDLESMKRKEDSYFNIRLRGFWPYAEMQNKIEF KLKQYGIEIRKVAPNNTSQRCSKCGHLNNYFNFEYRKKNKFPHFKCEKCNF KENAAYNAARNISNPNIKSTKERP
cA2.75	186	MAKNTITKTLKLRIVRPYNSQEIEKIVAEEKERRKQAGGTGELDDKFYQKLR GQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHY LSRVCYRRAAELFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIP LVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQK SPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKIC EKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISDNDL FHFNKKMFARRRILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLIERW ACEIADFFIKNKVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQNKIEF KLKQYGIEIRKVAPNNTSKTCSKCGHLNNYFNFEYRKKNKFPHFKCEKCNF KENAAYNAALNISNPKLKSTKERP
cA2.76	187	MAKNTITKTLKLRIVRPYNSQEIEKIVAEEKERRKQAGGTGELDDKFYQKLR GQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHY LSRVCYRRAAELFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIP LVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQK SPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKIC EKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISDNDL FHFNKKMFARRRILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLIERW ACEIADFFIKNKVGTVQMEDLESMKRKEDSYFNIRLRGFWPYAEMQNKIEF KLKQYGIEIRKVAPNNTSQLCSKCGHLNNYFNFEYRKKNKFPHFKCEKCNF KENAAYNAALNI
cA2.77	188	MAKNTITKTLKLRIVRPYNSQEIEKIVAEEKERRKQAGGTGELDDKFYQKLR GQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHY LSRVCYRRAAELFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIP LVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQK SPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKIC EKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISDNDL FHFNKKMFARRRILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLIERW

		ACEIADFFIKNKVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQNKIEF KLKQYGIEIRKVAPNNTSQLCSKCGHLNNYFNFEYRKKNKFPKFKCEKCNF KENAAYNAALNI
cA2.78	189	MAKNTITKTLKLRIVRPYNSQEIEKIVAEEKERRKQAGGTGELDDKFYQKLR GQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHY LSRVCYRRAAELFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIP LVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQK SPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKIC EKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISDNDL FHFNKKMFARRRILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLIERW ACEIADFFIKNKVGTVQMEDLESMKRKEDSYFNIRLRGFWPYAEMQNKIEF KLKQYGIEIRKVAPNNTSQLCSKCGHLNNYFNFEYRKKNKFPKFKCEKCNF KENAAYNAALNI
cA2.79	190	MAKNTITKTLKLRIVRPYNSQEIEKIVAEEKERRKQAGGTGELDDKFYQKLR GQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHY LSRVCYRRAAELFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIP LVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQK SPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKIC EKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISDNDL FHFNKKMFARRRILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLIERW ACEIADFFIKNKVGTVQMEDLESMKRKEDSYFNIRLRGFWPYAEMQNKIEF KLKQYGIEIRKVAPNNTSQRCSKCGHLNNYFNFEYRKKNKFPHFKCEKCNF KENAAYNAARNI
cA2.80	191	MAKNTITKTLKLRIVRPYNSQEIEKIVAEEKERRKQAGGTGELDDKFYQKLR GQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHY LSRVCYRRAAELFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIP LVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQK SPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKIC EKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISDNDL FHFNKKMFARRRILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLIERW ACEIADFFIKNKVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQNKIEF KLKQYGIEIRKVAPNNTSKTCSKCGHLNNYFNFEYRKKNKFPHFKCEKCNF KENAAYNAALNI

cA2.81	192	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKNNREKIALEKNKDKVKEACS KHLKVAAYCTTQVERNACLFCKARKLDDKFYQKLRGQFPDAVFWQEISEIF RQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCYRRAAALFK NAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGFEI SNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQKSPKPISLLLSTQRRK RNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSIDV PKIDKGVDPSIIGGIAVGVRSPLVCAINNSFSRYSISDNDLFKFNKKMFARRRI LLKKNRHKRKGHGAKNKLKPITELTEKSERFRKKLIERWACEIADFFIKNKV GTVQMEDLESMKRKEDSYFNIRLRGFWPYAEMQNKIEFKLKQYGIEIRKVA PNNTSQLCSKCGHLNNYFNFEYRKKNKFPHFKCEKCNFKENAAYNAALNIS NPDIKSTKERP
cA2.82	193	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKNNREKIALEKNKDKVKEACS KHLKVAAYCTTQVERNACLFCKARKLDDKFYQKLRGQFPDAVFWQEISEIF RQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCYRRAAALFK NAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGFEI SNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQKSPKPISLLLSTQRRK RNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSIDV PKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISDNDLFHFNKKMFARRRI LLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLIERWACEIADFFIKNKV GTVQMEDLESMKRKEDSYFNIRLRGFWPYAEMQNKIEFKLKQYGIEIRKVA PNNTSQLCSKCGHLNNYFNFEYRKKNKFPHFKCEKCNFKENAAYNAALNIS NPDIKSTKERP
cA2.83	194	MAKNTITKTLKLRIVRPYNSAEIEKIVADEKERRKQAGGTGELDDKFYKKL RKQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEH YLSRVCYRRAAALFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFP IPLVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQV QKSPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGS KICEKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISD NDLFHFNKKMFARRRILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLI ERWACEIADFFIKNKVGTVQMEDLESMKRKEDSYFNIRLRGFWPYAEMQN KIEFKLKQYGIEIRKVAPNNTSQLCSKCGHLNNYFNFEYRKKNKFPHFKCEK CNFKENAAYNAALNISNPDIKSTKERP

cA2.84	195	MAKNTITKTLKLRIVRPYNSAEIEKIVADEKERRKQAGGTGELDDKFYKKL RKQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEH YLSRVCYRRAAALFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFP IPLVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQV QKSPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGS KICEKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISD NDLFHFNKKMFARRRILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLI ERWACEIADFFIKNKVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQN KIEFKLKQYGIEIRKVAPNNTSQLCSKCGHLNNYFNFEYRKKNKFPKFKCEK CNFKENAAYNAALNISNPDIKSTKERP
cA2.85	196	MAKNTITKTLKLRIVRPYNSAEIEKIVADEKERRKQAGGTGELDDKFYKKL RKQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEH YLSRVCYRRAAELFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFP IPLVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQV QKSPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGS KICEKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISD NDLFHFNKKMFARRRILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLI ERWACEIADFFIKNKVGTVQMEDLESMKRKEDSYFNIRLRGFWPYAEMQN KIEFKLKQYGIEIRKVAPNNTSQLCSKCGHLNNYFNFEYRKKNKFPHFKCEK CNFKENAAYNAALNISNPDIKSTKERP
cA2.86	197	MAKNTITKTLKLRIVRPYNSAEIEKIVADEKERRKQAGGTGELDDKFYKKL RKQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEH YLSRVCYRRAAELFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFP IPLVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQV QKSPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGS KICEKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISD NDLFHFNKKMFARRRILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLI ERWACEIADFFIKNKVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQN KIEFKLKQYGIEIRKVAPNNTSQLCSKCGHLNNYFNFEYRKKNKFPKFKCEK CNFKENAAYNAALNISNPDIKSTKERP
cA2.87	198	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKERRKQAGGTGELDDKFYQKL RGQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEH YLSRVCYRRAAELFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFP IPLVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQV QKSPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGS KICEKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISD NDLFHFNKKMFARRRILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLI

		ERWACEIADFFIKNKVGTVQMEDLESMKRKEDSYFNIRLRGFWPYAEMQN KIEFKLKQYGIEIRKVAPNNTSKTCSKCGHLNNYFNFEYRKKNKFPHFKCEK CNFKENAAYNAALNISNPKLKSTKERP
cA2.88	199	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKERRKQAGGTGELDDKFYQKL RGQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEH YLSRVCYRRAAELFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFP IPLVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQV QKSPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGS KICEKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISD NDLFHFNKKMFARRRILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLI ERWACEIADFFIKNKVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQN KIEFKLKQYGIEIRKVAPNNTSQLCSKCGHLNNYFNFEYRKKNKFPHFKCEK CNFKENAAYNAALNISNPKLKSTKERP
cA2.89	200	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKERRKQAGGTGELDDKFYQKL RGQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEH YLSRVCYRRAAELFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFP IPLVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQV QKSPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGS KICEKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISD NDLFHFNKKMFARRRILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLI ERWACEIADFFIKNKVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQN KIEFKLKQYGIEIRKVAPNNTSKTCSKCGHLNNYFNFEYRKKNKFPHFKCEK CNFKENAAYNAALNISNPDIKSTKERP
cA2.90	201	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKERRKQAGGTGELDDKFYQKL RGQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEH YLSRVCYRRAAELFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFP IPLVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQV QKSPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGS KICEKSAWMLNLSIDVPKIDKGVDPSIIGGIAVGVRSPLVCAINNAFSRYSISD NDLFHFNKKMFARRRILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLI ERWACEIADFFIKNKVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQN KIEFKLKQYGIEIRKVAPNNTSQLCSKCGHLNNYFNFEYRKKNKFPHFKCEK CNFKENAAYNAALNISNPDIKSTKERP

表6A.来自第一轮的gRNA支架变体的列表。星号的数量代表gRNA支架变体的基因调节活性的相对程度。

引导 NA 支架（第一轮）	不含间隔子的长度 (bp)	GFP 的抑制 (5 d.p.t.)	尺寸减小的情况下与 SQ 相当或改善的活性
SQ (阳性对照)	159	***	
2	139	-	
3	138	*	
4	137	***	是
5	136	***	是
6	151	***	是
7	148	***	是
8	157	***	是
9	153	***	是
10	153	***	是
11	151	***	是
12	153	***	是
13	149	***	是
14	132	*	
15	132	*	
16	131	*	
17	134	*	
18	134	*	
19	136	**	
20	135	***	是
21	133	*	
22	131	*	
23	129	*	
24	126	**	
25	124	***	是
26	122	-	
27	120	*	

28	118	**	
29	116	***	是
30	114	-	
31	161	-	
32	164	-	
33	164	**	
34	169	**	
35	140	*	
36	149	**	
37	142	*	
38	140	-	
39	122	**	
40	120	*	
41	100	**	
42	98	-	
43	110	**	
44	108	*	
45	104	***	是
46	109	**	

表6B.来自表6A的gRNA支架变体和相应多核苷酸序列的列表

引导NA支架	相对于间隔子的位置	SEQ ID NO	引导核酸（NA）支架序列（不含间隔子）
SQ	间隔子的5'	500	GGCTTCACTGATAAAGTGGAGAACCGCTTCACCAAAAGCTGTCCCTTAGGGGATTAGAACTTGAGTGAAGGTGGGCTGCTTGCATCAGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCTCGAAACAAATTCATTTGAATGAAGGAATGCAAC
2	间隔子的5'	501	GAACCGCTTCACCAAAAGCTGTCCCTTAGGGGATTAGAACTTGAGTGAAGGTGGGCTGCTTGCATCAGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCTCGAAACAAATTCATTTGAATGAAGGAATGCAAC
3	间隔子的5'	502	AACCGCTTCACCAAAAGCTGTCCCTTAGGGGATTAGAACTTGAGTGAAGGTGGGCTGCTTGCATCAGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCTCGAAACAAATTCATTTGAATGAAGG

			AATGCAAC
4	间隔子的5'	503	ACCGCTTCACCAAAAGCTGTCCCTTAGGGGATTAGAACTTGAGTGAAGGTGGGCTGCTTGCATCAGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCTCGAAACAAATTCATTTGAATGAAGGAATGCAAC
5	间隔子的5'	504	CCGCTTCACCAAAAGCTGTCCCTTAGGGGATTAGAACTTGAGTGAAGGTGGGCTGCTTGCATCAGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCTCGAAACAAATTCATTTGAATGAAGGAATGCAAC
6	间隔子的5'	505	GGCTTCACTGATAAAGTGGAGAACCGCTTCACCAAAAGCTGTCCCTTAGGGGATTAGAACTTGAGTGAAGGTGGGCTGCTTGCATCAGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCTCGAAACAAGCAATAAGGAATGCAAC
7	间隔子的5'	506	GGCTTCACTGATAAAGTGGAGAACCGCTTCACCAAAAGCTGTCCCTTAGGGGATTAGAACTTGAGTGAAGGTGGGCTGCTTGCATCAGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCTCGAAACAAGAAAGGAATGCAAC
8	间隔子的5'	507	GGCTTCACTGATAAAGTGGAGAACCGCTTCACCAAAAGCTGTCCCTTAGGGGATTAGAACTTGAGTGAAGGTGGGCTGCTTGCATCAGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCTCGAAACAAATCTTCGGATTAAGGAATGCAAC
9	间隔子的5'	508	GGCTTCACTGATAAAGTGGAGAACCGCTTCACCAAAAGCTGTCCCTTAGGGGATTAGAACTTGAGTGAAGGTGGGCTGCTTGCATCAGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCTCGAAACAAATTGCAAAAGGAATGCAAC
10	间隔子的5'	509	GGCTTCACTGATAAAGTGGAGAACCGCTTCACCAAAAGCTGTCCCTTAGGGGATTAGAACTTGAGTGAAGGTGGGCTGCTTGCATCAGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCTCGAAACAAATTCGTTAAGGAATGCAAC
11	间隔子的5'	510	GGCTTCACTGATAAAGTGGAGAACCGCTTCACCAAAAGCTGTCCCTTAGGGGATTAGAACTTGAGTGAAGGTGGGCTGCTTGCATCAGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCTCGAAACAAATGCAAAGGAATGCAAC

12	间隔子的5'	511	GGCTTCACTGATAAAGTGGAGAACCGCTTCACCAAAAGCTGTCCCTTAGGGGATTAGAACTTGAGTGAAGGTGGGCTGCTTGCATCAGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCTCGAAACAAAGCAATTAAGGAATGCAAC
13	间隔子的5'	512	GGCTTCACTGATAAAGTGGAGAACCGCTTCACCAAAAGCTGTCCCTTAGGGGATTAGAACTTGAGTGAAGGTGGGCTGCTTGCATCAGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCTCGAAACAAAGAAAGGAATGCAAC
14	间隔子的5'	513	GGCTTCACTGATAAAGTGGAGAACCGCTTCACTTAGAGTGAAGGTGGGCTGCTTGCATCAGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCTCGAAACAAATTCATTTGAATGAAGGAATGCAAC
15	间隔子的5'	514	GGCTTCACTGATAAAGTGGAGAACCGCTTCACTTCGAGTGAAGGTGGGCTGCTTGCATCAGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCTCGAAACAAATTCATTTGAATGAAGGAATGCAAC
16	间隔子的5'	515	GGCTTCACTGATAAAGTGGAGAACCGCTTCACTTCGGTGAAGGTGGGCTGCTTGCATCAGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCTCGAAACAAATTCATTTGAATGAAGGAATGCAAC
17	间隔子的5'	516	GGCTTCACTGATAAAGTGGAGAACCGCTTCACCTTAGGAGTGAAGGTGGGCTGCTTGCATCAGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCTCGAAACAAATTCATTTGAATGAAGGAATGCAAC
18	间隔子的5'	517	GGCTTCACTGATAAAGTGGAGAACCGCTTCACCTTCGGAGTGAAGGTGGGCTGCTTGCATCAGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCTCGAAACAAATTCATTTGAATGAAGGAATGCAAC
19	间隔子的5'	518	GGCTTCACTGATAAAGTGGAGAACCGCTTCACGCTTCGGCAGTGAAGGTGGGCTGCTTGCATCAGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCTCGAAACAAATTCATTTGAATGAAGGAATGCAAC
20	间隔子的5'	519	ACCGCTTCACCAAAAGCTGTCCTTAGGGATTAGAACTTGAGTGAAGGTGGGCTGCTTGCATCAGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCTCGAAACAAATTCATTTGAATGAAGGAATGCAAC

21	间隔子的5'	520	ACCGCTTCACCAAAAGCTGTCTTAGGATTAGAACTTGAGTGAAGGTGGGCTGCTTGCATCAGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCTCGAAACAAATTCATTTGAATGAAGGAATGCAAC
22	间隔子的5’	521	ACCGCTTCACCAAAAGCTGTTTAGATTAGAACTTGAGTGAAGGTGGGCTGCTTGCATCAGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCTCGAAACAAATTCATTTGAATGAAGGAATGCAAC
23	间隔子的5’	522	ACCGCTTCACCAAAAGCTGTTAGTTAGAACTTGAGTGAAGGTGGGCTGCTTGCATCAGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCTCGAAACAAATTCATTTGAATGAAGGAATGCAAC
24	间隔子的5’	523	ACCGCTTCACCAAAAGCTTTAGAGAACTTGAGTGAAGGTGGGCTGCTTGCATCAGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCTCGAAACAAATTCATTTGAATGAAGGAATGCAAC
25	间隔子的5’	524	ACCGCTTCACCAAAAGCTTCGGCACTTGAGTGAAGGTGGGCTGCTTGCATCAGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCTCGAAACAAATTCATTTGAATGAAGGAATGCAAC
26	间隔子的5’	525	ACCGCTTCACCAAAAGTTCGCACTTGAGTGAAGGTGGGCTGCTTGCATCAGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCTCGAAACAAATTCATTTGAATGAAGGAATGCAAC
27	间隔子的5’	526	ACCGCTTCACCAAAATTCGTCTTGAGTGAAGGTGGGCTGCTTGCATCAGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCTCGAAACAAATTCATTTGAATGAAGGAATGCAAC
28	间隔子的5’	527	ACCGCTTCACCAAGTTCGCTTGAGTGAAGGTGGGCTGCTTGCATCAGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCTCGAAACAAATTCATTTGAATGAAGGAATGCAAC
29	间隔子的5’	528	ACCGCTTCACCAATTCGTTGAGTGAAGGTGGGCTGCTTGCATCAGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCTCGAAACAAATTCATTTGAATGAAGGAATGCAAC
30	间隔子的5’	529	ACCGCTTCACCATTCGTGAGTGAAGGTGGGCTGCTTGCATCAGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCTCGAAACAAATTCATTTGAATGAAGGAATGCAAC
31	间隔子的5’	530	GGCTTCACTGATAAAGTGGAGAACCGCTTCACCAAAAGCTGTCCCTTAGGGGATTAGAACTTGAGTGAAGGTGGGCTGCTTGCATCAGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCTCGAAACAAATTCATTTGAATGAAGGAATGCAAC
	间隔子的	531	TT

	3’		
32	间隔子的5’	532	GGCTTCACTGATAAAGTGGAGAACCGCTTCACCAAAAGCTGTCCCTTAGGGGATTAGAACTTGAGTGAAGGTGGGCTGCTTGCATCAGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCTCGAAACAAATTCATTTGAATGAAGGAATGCAAC
	间隔子的3’	533	TTTTA
33	间隔子的5’	534	GGCTTCACTGATAAAGTGGAGAACCGCTTCACCAAAAGCTGTCCCTTAGGGGATTAGAACTTGAGTGAAGGTGGGCTGCTTGCATCAGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCTCGAAACAAATTCATTTGAATGAAGGAATGCAAC
	间隔子的3’	535	TTTTG
34	间隔子的5’	536	GGCTTCACTGATAAAGTGGAGAACCGCTTCACCAAAAGCTGTCCCTTAGGGGATTAGAACTTGAGTGAAGGTGGGCTGCTTGCATCAGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCTCGAAACAAATTCATTTGAATGAAGGAATGCAAC
	间隔子的3’	537	TTTTATTTTT
35	间隔子的5’	538	GGCTTCACTGATAAAGTGGAGAACCGCTTCACCGAGTGAAGGTGGGCTGCTTGCATCAGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCTCGAAACAAATTCATTTGAATGAAGGAATGCAAC
	间隔子的3’	539	TTTTATTTTT
36	间隔子的5’	540	GGCTTCACTGATAAAGTGGAGAACCGCTTCACCAAAAGCTGTCCCTTAGGGGATTAGAACTTGAGTGAAGGTGGGCTGCTTGCATCAGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCTCGAAACAAAGAAAGGAATGCAAC
37	间隔子的5’	541	GGCTTCACTGATAAAGTGGAGAACCGCTTCACTTAGAGTGAAGGTGGGCTGCTTGCATCAGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCTCGAAACAAATTCATTTGAATGAAGGAATGCAAC
	间隔子的3’	542	TTTTATTTTT
38	间隔子的5’	543	GGCTTCACTGATAAAGTGGAGAACCGCTTCACCGAGTGAAGGTGGGCTGCTTGCATCAGCCTAATGTCGAGAAGTGCTTTCTTCGGA

			AAGTAACCCTCGAAACAAATTCATTTGAATGAAGGAATGCAAC
	间隔子的3’	544	TTTTATTTTT
39	间隔子的5’	545	GGCTTCACTGATAAAGTGGAGAACCGCTTCACTTAGAGTGAAG GTGGGCTGCTTGCATCAGCCTAATGTCGAGAAGTGCTTTCTTCG GAAAGTAACCCTCGAAACAAAGAAAGGAATGCAAC
40	间隔子的5’	546	GGCTTCACTGATAAAGTGGAGAACCGCTTCACCGAGTGAAGGT GGGCTGCTTGCATCAGCCTAATGTCGAGAAGTGCTTTCTTCGGA AAGTAACCCTCGAAACAAAGAAAGGAATGCAAC
41	间隔子的5’	547	ACCGCTTCACTTAGAGTGAAGGTGGGCTGCTTGCATCAGCCTAA TGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCTCGAAACAAAG AAAGGAATGCAAC
42	间隔子的5’	548	ACCGCTTCACCGAGTGAAGGTGGGCTGCTTGCATCAGCCTAAT GTCGAGAAGTGCTTTCTTCGGAAAGTAACCCTCGAAACAAAGA AAGGAATGCAAC
43	间隔子的5’	549	ACCGCTTCACTTAGAGTGAAGGTGGGCTGCTTGCATCAGCCTAA TGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCTCGAAACAAAG AAAGGAATGCAAC
	间隔子的3’	550	TTTTATTTTT
44	间隔子的5’	551	ACCGCTTCACCGAGTGAAGGTGGGCTGCTTGCATCAGCCTAAT GTCGAGAAGTGCTTTCTTCGGAAAGTAACCCTCGAAACAAAGA AAGGAATGCAAC
	间隔子的3’	552	TTTTATTTTT
45	间隔子的5’	553	ACCGCTTCACGCTTCGGCAGTGAAGGTGGGCTGCTTGCATCAG CCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCTCGAAAC AAAGAAAGGAATGCAAC
46	间隔子的5’	554	ACCGCTTCACGCTTCGGCAGTGAAGGTGGGCTGCTTGCATCAG CCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCTCGAAAC AAAGAAAGGAATGCAAC
	间隔子的3’	602	TTTTA

表7A.来自第二轮的gRNA支架变体的列表。星号的数量代表gRNA支架变体的基因调节活性的相对程度。

引导 NA 支架（第二轮）	不含间隔子的长度 (bp)	GFP 的抑制 (5 d.p.t.)	尺寸减小的情况下与 SQ 相当或改善的活性
SQ (阳性对照)	159	***	
45.1	103	***	是
2-5	103	**	
2-6	103	***	是
2-7	103	***	是
2-8	104	**	
2-9	105	*	
2-10	102	**	
2-11	101	-	
2-12	102	*	
2-13	111	-	
2-14	112	*	
2-15	102	**	
2-16	101	**	
2-17	100	***	是
2-18	99	***	是
2-19	98	**	
2-20	97	**	
2-21	96	-	
2-22	95	-	
2-23	94	-	
2-24	93	-	
2-25	97	**	
2-26	99	**	
2-27	98	***	是
2-28	102	*	
2-29	100	***	是
2-30	99	-	
2-31	96	**	

2-32	95	-	
2-33	103	**	
2-34	103	-	
2-35	103	-	
2-36	97	-	
2-37	97	*	

表7B.来自表7A的gRNA支架变体和相应多核苷酸序列的列表。

引导NA支架	相对于间隔子的位置	SEQ ID NO	引导核酸（NA）支架序列（不含间隔子）
45.1	间隔子的5’	555	CCGCTTCACGCTTCGGCAGTGAAGGTGGGCTGCTTGCATCAGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCTCGAAACAAAGAAAGGAATGCAAC
2-5	间隔子的5’	556	CCGCTTCACGCTTAGGCAGTGAAGGTGGGCTGCTTGCATCAGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCTCGAAACAAAGAAAGGAATGCAAC
2-6	间隔子的5’	557	CCGCTTCACTCTTAGGAAGTGAAGGTGGGCTGCTTGCATCAGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCTCGAAACAAAGAAAGGAATGCAAC
2-7	间隔子的5’	558	CCGCTTCACGTTTAGACAGTGAAGGTGGGCTGCTTGCATCAGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCTCGAAACAAAGAAAGGAATGCAAC
2-8	间隔子的5’	559	GCCGCTTCACGCTTCGGCAGTGAAGGTGGGCTGCTTGCATCAGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCTCGAAACAAAGAAAGGAATGCAAC
2-9	间隔子的5’	560	GGCCGCTTCACGCTTCGGCAGTGAAGGTGGGCTGCTTGCATCAGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCTCGAAACAAAGAAAGGAATGCAAC
2-10	间隔子的5’	561	CGCTTCACGCTTCGGCAGTGAAGGTGGGCTGCTTGCATCAGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCTCGAAACAAAGAAAGGAATGCAAC
2-11	间隔子的	562	GCTTCACGCTTCGGCAGTGAAGGTGGGCTGCTTGCATCAG

	5’		CCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCTCG AAACAAAGAAAGGAATGCAAC
2-12	间隔子的 5’	563	GGCTTCACGCTTCGGCAGTGAAGGTGGGCTGCTTGCATCA GCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCTC GAAACAAAGAAAGGAATGCAAC
2-13	间隔子的 5’	564	GCTTCACGCTTCGGCAGTGAAGGTGGGCTGCTTGCATCAG CCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCTCG AAACAAAGAAAGGAATGCAAC
	间隔子的 3’		TTTTATTTTT
2-14	间隔子的 5’	565	GGCTTCACGCTTCGGCAGTGAAGGTGGGCTGCTTGCATCA GCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCTC GAAACAAAGAAAGGAATGCAAC
	间隔子的 3’		TTTTATTTTT
2-15	间隔子的 5’	566	CCGCTTCACGCTTCGGCAGTGAAGGTGGGCTGCTTGCATC AGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCT CGAAACAAAGAAGGAATGCAAC
2-16	间隔子的 5’	567	CCGCTTCACGCTTCGGCAGTGAAGGTGGGCTGCTTGCATC AGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCT CGAAACAAAGAGGAATGCAAC
2-17	间隔子的 5’	568	CCGCTTCACGCTTCGGCAGTGAAGGTGGGCTGCTTGCATC AGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCT CGAAACAAAGGGAATGCAAC
2-18	间隔子的 5’	569	CCGCTTCACGCTTCGGCAGTGAAGGTGGGCTGCTTGCATC AGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCT CGAAACAAAGGAATGCAAC
2-19	间隔子的 5’	570	CCGCTTCACGCTTCGGCAGTGAAGGTGGGCTGCTTGCATC AGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCT CGAAACAAGGAATGCAAC
2-20	间隔子的 5’	571	CCGCTTCACGCTTCGGCAGTGAAGGTGGGCTGCTTGCATC AGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCT CGAAACAGGAATGCAAC

2-21	间隔子的5’	572	CCGCTTCACGCTTCGGCAGTGAAGGTGGGCTGCTTGCATCAGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCTCGAAACGGAATGCAAC
2-22	间隔子的5’	573	CCGCTTCACGCTTCGGCAGTGAAGGTGGGCTGCTTGCATCAGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCTCGAAAGGAATGCAAC
2-23	间隔子的5’	574	CCGCTTCACGCTTCGGCAGTGAAGGTGGGCTGCTTGCATCAGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCTCGAAGGAATGCAAC
2-24	间隔子的5’	575	CCGCTTCACGCTTCGGCAGTGAAGGTGGGCTGCTTGCATCAGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCTCGAGGAATGCAAC
2-25	间隔子的5’	576	CCGCTTCACGCTTCGGCAGTGAAGGTGGGCTGCTTGCATCAGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCTCGAAACAAGAATGCAAC
2-26	间隔子的5’	577	CCGCTTCACGCTTCGGCAGTGAAGGTGGGCTGCTTGCATCAGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCTCGAAACAAAGGAATGCAAC
2-27	间隔子的5’	578	CCGCTTCACGCTTCGGCAGTGAAGGTGGGCTGCTTGCATCAGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCTCGAAACAAAGAATGCAAC
2-28	间隔子的5’	579	CCGCTTCACGCTTCGGCAGTGAAGGTGGGCTGCTTGCATCAGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCTCGAAACAAAGAAAGAATGCAAC
2-29	间隔子的5’	580	CCGCTTCACGCTTCGGCAGTGAAGGTGGGCTGCTTGCATCAGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCTCGAAACAAAGAGAATGCAAC
2-30	间隔子的5’	581	CCGCTTCACGCTTCGGCAGTGAAGGTGGGCTGCTTGCATCAGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCTCGAAACAAAGAAATGCAAC
2-31	间隔子的5’	582	CCGCTTCACGCTTCGGCAGTGAAGGTGGGCTGCTTGCATCAGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCTCGAAACAGAATGCAAC

2-32	间隔子的5’	583	CCGCTTCACGCTTCGGCAGTGAAGGTGGGCTGCTTGCATCAGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCTCGAAACAAATGCAAC
2-33	间隔子的5’	584	CCGCTTCACGCTTCGGCAGTGAAGGTGGGCTGATTGCATCAGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCTCGAAACAAAGAAAGGAATGCAAC
2-34	间隔子的5’	585	CCGCTTCACGCTTCGGCAGTGAAGGTAGGCTGCTTGCATCAGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCTCGAAACAAAGAAAGGAATGCAAC
2-35	间隔子的5’	586	CCGCTTCACGCTTCGGCAGTGAAGGTGGGCTGCTTGCATCAGCCCAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCTCGAAACAAAGAAAGGAATGCAAC
2-36	间隔子的5’	587	GGCTTCACGCTTCGGCAGTGAAGGTAGGCTGCTTGCATCAGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCTCGAAACAAGGAATGCAAC
2-37	间隔子的5’	588	GGCTTCACGCTTCGGCAGTGAAGGTGGGCTGCTTGCATCAGCCCAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCTCGAAACAAGGAATGCAAC

表8A.来自第三轮的gRNA支架变体的列表。星号的数量代表gRNA支架变体的基因调节活性的相对程度。

引导 NA 支架	**不含间隔子的长度(bp)**	**CD2 的激活(2 d.p.t.)**	**尺寸减小的情况下与 SQ 相当或改善的活性**	**SEQ ID NO**
SQ (阳性对照)	159	**		500
45.1	103	****	是	555
2-6	103	****	是	557
2-17	100	***	是	568
2-18	99	***	是	569
2-25	97	**	是	576
2-26	99	**	是	577
2-27	98	***	是	578

2-29	100	**	是	580
2-33	103	-		584
2-31	96	*		582
3-1	103	*		589
3-2	100	*		590
3-3	99	-		591
3-4	98	-		592
3-5	97	**	是	593
3-6	99	*		594
3-7	98	-		595
3-8	96	-		596
6	151	-		505
20	135	**	是	519
29	116	***	是	528

表8B.来自表8A的另外的gRNA支架变体的列表

引导 NA 支架	SEQ ID NO	引导核酸（NA）支架序列（不含间隔子）
3-1	589	CCGCTTCACTCTTAGGAAGTGAAGGTGGGCTGATTGCATCAGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCTCGAAACAAAGAAAGGAATGCAAC
3-2	590	CCGCTTCACGCTTCGGCAGTGAAGGTGGGCTGATTGCATCAGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCTCGAAACAAAGGGAATGCAAC
3-3	591	CCGCTTCACGCTTCGGCAGTGAAGGTGGGCTGATTGCATCAGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCTCGAAACAAAGGAATGCAAC
3-4	592	CCGCTTCACGCTTCGGCAGTGAAGGTGGGCTGATTGCATCAGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCTCGAAACAAGGAATGCAAC
3-5	593	CCGCTTCACGCTTCGGCAGTGAAGGTGGGCTGATTGCATCAGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCTCGAAACAAGAATGCAAC
3-6	594	CCGCTTCACGCTTCGGCAGTGAAGGTGGGCTGATTGCATCAGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCTCGAAACAAAGGAATGCAAC
3-7	595	CCGCTTCACGCTTCGGCAGTGAAGGTGGGCTGATTGCATCAGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCTCGAAACAAAGAATGCAAC
3-8	596	CCGCTTCACGCTTCGGCAGTGAAGGTGGGCTGATTGCATCAGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCTCGAAACAGAATGCAAC

表9.源自gRNA支架变体的片段的列表

引导 NA 支架片段	SEQ ID NO	引导核酸（NA）支架片段序列
45.1、2-27、2-17 和/或 2-18 的片段	597	CCGCTTCACGCTTCGGCAGTGAAGGTGGGC
2-6 的片段	598	CCGCTTCACTCTTAGGAAGTGAAGGTGGGC
2-27 的片段	599	GAAAGTAACCCTCGAAACAAAGAATGCAAC
2-17 的片段	600	AAGTAACCCTCGAAACAAAGGGAATGCAAC
2-18 的片段	601	AAAGTAACCCTCGAAACAAAGGAATGCAAC

实施方式

[0345] 以下非限制性实施方式提供了本发明的说明性实例,但不限制本发明的范围。

[0346] 实施方式1.一种包含工程化核酸酶的工程化多肽,其中所述工程化核酸酶包含与SEQ ID NO:1的多肽序列至少80%相同的氨基酸序列,

其中与SEQ ID NO:1的多肽序列相比,所述氨基酸序列包含来自氨基酸残基2-100的至少一个缺失,

任选地,其中:

(1)与SEQ ID NO:1的多肽序列相比,所述至少一个缺失来自选自于由氨基酸残基30-40、氨基酸残基40-50、氨基酸残基50-60、氨基酸残基60-70和氨基酸残基70-80所组成的组中的一个或多个成员;和/或

(2)与SEQ ID NO:1的多肽序列相比,所述至少一个缺失来自选自于由氨基酸残基30-35、氨基酸残基35-40、氨基酸残基40-45、氨基酸残基45-50、氨基酸残基50-55、氨基酸残基55-60、氨基酸残基60-65、氨基酸残基65-70、氨基酸残基70-75和氨基酸残基75-80所组成的组中的一个或多个成员;和/或

(3)与SEQ ID NO:1的多肽序列相比,所述至少一个缺失包括来自氨基酸残基30-80的多个氨基酸残基;和/或

(4)与SEQ ID NO:1的多肽序列相比,所述至少一个缺失来自氨基酸残基2-80,

任选地,其中:

(a)与SEQ ID NO:1的多肽序列相比,所述至少一个缺失来自氨基酸残基2-60;和/或

(b)与SEQ ID NO:1的多肽序列相比,所述至少一个缺失来自氨基酸残基2-40;和/或

(c)与SEQ ID NO:1的多肽序列相比,所述至少一个缺失来自氨基酸残基2-30;和/或

(5)与SEQ ID NO:1的多肽序列相比,所述至少一个缺失来自氨基酸残基30-100,

任选地,与SEQ ID NO:1的多肽序列相比,所述至少一个缺失来自氨基酸残基30-80,

进一步任选地,其中,与SEQ ID NO:1的多肽序列相比,所述至少一个缺失包含氨基酸残基55-56、氨基酸残基54-57、氨基酸残基54-58、氨基酸残基53-59、氨基酸残基52-60、氨基酸残基51-61、氨基酸残基50-62、氨基酸残基49-63、氨基酸残基48-64、氨基酸残基

47-65、氨基酸残基46-66、氨基酸残基45-67、氨基酸残基44-68、氨基酸残基43-69、氨基酸残基42-70、或氨基酸残基41-71的缺失;和/或

(6)与SEQ ID NO:1的多肽序列相比,所述至少一个缺失来自选自于由氨基酸残基2-10、氨基酸残基11-20、氨基酸残基21-30、氨基酸残基31-40、氨基酸残基41-50、氨基酸残基51-60、氨基酸残基61-70和氨基酸残基71-80所组成的组中的一个或多个成员,

任选地,其中,与SEQ ID NO:1的多肽序列相比,所述至少一个缺失来自选自于由氨基酸残基2-10、氨基酸残基11-20、氨基酸残基21-30、氨基酸残基31-40、氨基酸残基41-50、氨基酸残基51-60、氨基酸残基61-70和氨基酸残基71-80所组成的组中的两个或更多个成员;和/或

(7)与SEQ ID NO:1的氨基酸序列相比,所述工程化核酸酶包含D326或D510处的氨基酸置换,从而降低所述工程化核酸酶的核酸酶活性;和/或

(8)与SEQ ID NO:1的氨基酸序列相比,所述工程化核酸酶包含选自于由D143R、T147R、K330R和E528R所组成的组中的一个或多个氨基酸置换。

[0347] 实施方式2.一种包含工程化核酸酶的工程化多肽,其中,所述工程化核酸酶包含与SEQ ID NO:1的多肽序列至少80%相同的氨基酸序列,

其中,与SEQ ID NO:1的多肽序列相比,所述氨基酸序列包含来自氨基酸残基430-529的至少一个缺失,

任选地,其中:

(1)与SEQ ID NO:1的多肽序列相比,所述至少一个缺失来自氨基酸残基450-529,

任选地,其中:

(a)与SEQ ID NO:1的多肽序列相比,所述至少一个缺失来自氨基酸残基470-529;和/或

(b)与SEQ ID NO:1的多肽序列相比,所述至少一个缺失来自氨基酸残基490-529;和/或

(c)与SEQ ID NO:1的多肽序列相比,所述至少一个缺失来自氨基酸残基500-529;和/或

(2)与SEQ ID NO:1的多肽序列相比,所述至少一个缺失来自选自于由氨基酸残基450-459、氨基酸残基460-469、氨基酸残基470-479、氨基酸残基480-489、氨基酸残基490-499、氨基酸残基500-509、氨基酸残基510-519和氨基酸残基520-529所组成的组中的一个或多个成员,

任选地,其中,与SEQ ID NO:1的多肽序列相比,所述至少一个缺失来自选自于由氨基酸残基450-459、氨基酸残基460-469、氨基酸残基470-479、氨基酸残基480-489、氨基酸残基490-499、氨基酸残基500-509、氨基酸残基510-519、氨基酸残基520-529所组成的组中的两个或更多个成员;和/或

(3)与SEQ ID NO:1的多肽序列相比,所述至少一个缺失来自氨基酸残基500-529;和/或

(4)所述氨基酸序列与SEQ ID NO:1的多肽序列至少85%相同、至少90%相同、至少95%相同、至少96%相同、至少97%相同、至少98%相同或至少99%相同;和/或

(5)所述至少一个缺失包括多个氨基酸缺失,

任选地,其中:

(a)所述多个氨基酸缺失包括至少3个氨基酸缺失、至少4个氨基酸缺失、至少5个氨基酸缺失、至少10个氨基酸缺失、至少15个氨基酸缺失、至少20个氨基酸缺失、至少25个氨基酸缺失、至少30个氨基酸缺失、至少35个氨基酸缺失或至少40个氨基酸缺失;和/或

(b)所述多个氨基酸缺失包括多个非连续氨基酸的缺失;和/或

(6)所述至少一个缺失包括多个连续氨基酸残基的缺失,

任选地,其中,所述多个连续氨基酸残基包括至少3个连续氨基酸残基、至少4个连续氨基酸残基、至少5个连续氨基酸残基、至少10个连续氨基酸残基、至少15个连续氨基酸残基、至少20个连续氨基酸残基、至少25个连续氨基酸残基或至少30个连续氨基酸残基;和/或

(7)与SEQ ID NO:1的多肽序列相比,所述工程化多肽还包含来自氨基酸残基101-429的一个或多个另外的缺失,

任选地,其中,所述一个或多个另外的缺失包括多个另外的缺失;和/或

(8)所述氨基酸序列的长度为至多528个氨基酸、至多527个氨基酸、至多526个氨基酸、至多525个氨基酸、至多524个氨基酸、至多519个氨基酸、至多514个氨基酸、至多509个氨基酸、至多514个氨基酸、至多509个氨基酸、至多504个氨基酸或至多489个氨基酸;和/或

(9)所述工程化核酸酶的长度为至多约600个氨基酸、至多约550个氨基酸、至多约540个氨基酸或至多约530个氨基酸;和/或

(10)与由SEQ ID NO:1编码的蛋白质相比,所述工程化核酸酶表现出降低的核酸酶活性,

任选地,其中,所述工程化核酸酶包含D326和/或D510处的置换,

进一步任选地,其中D326和/或D510被丙氨酸置换;和/或

(11)所述工程化多肽还包含与所述工程化核酸酶偶联的基因调节因子,

任选地,其中:

(a)所述基因调节因子与所述工程化核酸酶融合;和/或

(b)所述基因调节因子是转录激活因子;和/或

(c)所述基因调节因子是转录阻抑因子;和/或

(d)所述基因调节因子是组蛋白修饰剂,进一步任选地,其中所述组蛋白修饰剂是组蛋白甲基化修饰剂;和/或

(e)所述基因调节因子是基因甲基化修饰剂;和/或

(f)所述工程化多肽能够调节细胞中靶基因的表达和/或活性水平,其中,由所述工程化多肽调节的表达和/或活性水平与对照多肽相当,其中,所述对照多肽包含(i)含有SEQ ID NO:10的多肽序列的失活的核酸酶和(ii)基因调节因子;和/或

(12)所述工程化核酸酶包含与SEQ ID NO:11的多肽序列至少约80%相同的氨基酸序列;和/或

(13)所述工程化核酸酶包含与SEQ ID NO:11的多肽序列至少约90%相同的氨基酸序列;和/或

(14)与SEQ ID NO:1的氨基酸序列相比,所述工程化核酸酶包含选自于由D143R、

T147R、K330R和E528R所组成的组中的一个或多个氨基酸置换。

[0348] 实施方式3.一种包含工程化核酸酶的工程化多肽,其中,所述工程化核酸酶包含与SEQ ID NO:12的多肽序列大于92%相同的氨基酸序列,

任选地其中:

(1)所述工程化核酸酶的氨基酸序列与SEQ ID NO:12的多肽序列至少约93%相同;和/或

(2)所述工程化核酸酶的氨基酸序列与SEQ ID NO:12的多肽序列至少约95%、至少约98%或至少约99%相同;和/或

(3)所述工程化核酸酶的氨基酸序列与SEQ ID NO:12的多肽序列基本上相同;和/或

(4)所述氨基酸序列的长度小于或等于约600个氨基酸、小于或等于约550个氨基酸、或小于或等于约500个氨基酸;和/或

(5)与SEQ ID NO:1的氨基酸序列相比,所述工程化核酸酶包含D326或D510处的氨基酸置换,从而降低所述工程化核酸酶的核酸酶活性;和/或

(6)与SEQ ID NO:1的氨基酸序列相比,所述工程化核酸酶包含选自于由D143R、T147R、K330R和E528R所组成的组中的一个或多个氨基酸置换。

[0349] 实施方式4.一种包含工程化核酸酶变体的工程化多肽,其中所述工程化核酸酶变体包含与SEQ ID NO:12的多肽序列至少70%相同的氨基酸序列,

其中,与SEQ ID NO:1的多肽序列相比,所述氨基酸序列包含修饰,其中所述修饰包括选自于由A21Q、V23I、N32E、D29E、N33R、E35K、K36Q、I37A、A38G、E40D、K73G、A74T、R75G、K76E、Q83K、G87K、E151A、A340S、H353K、A374K、I387E、N423D、K473Q、T474L、T474R、H497K、L515R、N519T、K521D、K521N、L522I,以及来自SEQ ID NO:1的氨基酸残基400-529的至少一个缺失所组成的组中的一个或多个成员,

任选地,其中:

(1)所述氨基酸序列与SEQ ID NO:12的多肽序列至少80%、至少85%、至少90%或至少95%相同;和/或

(2)所述氨基酸序列与SEQ ID NO:1的多肽序列至少80%或至少85%相同;和/或

(3)所述氨基酸序列与SEQ ID NO:1的多肽序列至多95%或至多90%相同;和/或

(4)所述修饰包括选自于由A21Q、V23I、N32E、D29E、N33R、E35K、K36Q、I37A、A38G、E40D、K73G、A74T、R75G、K76E、Q83K、G87K、E151A、A340S、H353K、A374K、I387E、N423D、K473Q、T474L、T474R、H497K、L515R、N519T、K521D、K521N、L522I,以及来自SEQ ID NO:1的氨基酸残基400-529的至少一个缺失所组成的组中的至少两个成员、至少三个成员、至少四个成员或至少五个成员;和/或

(5)所述修饰包括来自SEQ ID NO:1的氨基酸残基400-529的至少一个缺失以及选自于由A21Q、V23I、D29E、N33R、E40D、Q83K、G87K、E151A、A340S、H353K、A374K、I387E、N423D、K473Q、T474L、T474R、H497K、L515R、N519T、K521D、K521N和L522I所组成的组中的一个或多个成员;和/或

(6)所述修饰包括来自SEQ ID NO:1的氨基酸残基400-529的至少一个缺失和选自于由N32E、N33R、E35K、K36Q、I37A、A38G、K73G、A74T、R75G、K76E所组成的组中的一个或多个

成员;和/或

(7)所述修饰包括选自于由N32E、N33R、E35K、K36Q、I37A、A38G、K73G、A74T、R75G、K76E所组成的组中的两个或更多个成员;和/或

(8)所述至少一个缺失来自SEQ ID NO:1的氨基酸残基450-529或氨基酸残基500-529;和/或

(9)所述修饰包括选自表5A的修饰组;和/或

(10)所述修饰包括选自表5B的修饰的组合;和/或

(11)所述修饰的组合不是来自表5B的cA2.55或cA2.84;和/或

(12)所述氨基酸序列的长度小于或等于约600个氨基酸、小于或等于约550个氨基酸、或小于或等于约500个氨基酸;和/或

(13)与SEQ ID NO:1的氨基酸序列相比,所述工程化核酸酶变体包含D326或D510处的氨基酸置换,从而降低所述工程化核酸酶变体的核酸酶活性;和/或

(14)与SEQ ID NO:1的氨基酸序列相比,所述工程化核酸酶变体包含选自于由D143R、T147R、K330R和E528R所组成的组中的一个或多个氨基酸置换。

[0350] 实施方式5.一种包含工程化核酸酶变体的工程化多肽,其中所述工程化核酸酶变体是嵌合多肽,所述嵌合多肽包含:

第一多肽序列,所述第一多肽序列包含与第一Cas蛋白共同的至少3个连续氨基酸残基;和

第二多肽序列,所述第二多肽序列包含与第二Cas蛋白共同的至少3个连续氨基酸残基,其中所述第二Cas蛋白不同于所述第一Cas蛋白,

其中所述第一Cas蛋白包含与SEQ ID NO:1的多肽序列至少80%相同的氨基酸序列,

任选地其中:

(1)所述第一Cas蛋白包含与SEQ ID NO:1的多肽序列至少85%、至少90%、至少95%、至少99%或100%相同的氨基酸序列;和/或

(2)所述第二Cas蛋白包含与选自表2的Cas蛋白的多肽序列至少80%、至少85%、至少90%、至少95%、至少99%或100%相同的氨基酸序列;和/或

(3)选自表2的Cas蛋白是Un2Cas12f1;和/或

(4)所述第一多肽包含与所述第一Cas蛋白共同的至少或多至约5个连续氨基酸残基、至少或多至约10个连续氨基酸残基、至少或多至约20个连续氨基酸残基、至少或多至约30个连续氨基酸残基、至少或多至约40个连续氨基酸残基、或至少或多至约50个连续氨基酸残基;和/或

(5)所述第二多肽包含与所述第二Cas蛋白共同的至少或多至约5个连续氨基酸残基、至少或多至约8个连续氨基酸残基、至少或多至约10个连续氨基酸残基、或至少或多至约20个连续氨基酸残基;和/或

(6)所述第一多肽序列的长度大于所述第二多肽序列的长度;和/或

(7)所述第二多肽衍生自所述第二Cas蛋白的N端50%、N端40%或N端20%;和/或

(8)所述第二多肽不包含来自所述第二Cas蛋白的N端的至少前5个氨基酸、至少前10个氨基酸、至少前20个氨基酸、或至少前30个氨基酸;和/或

(9)所述工程化多肽还包含第三多肽序列,所述第三多肽序列包含与所述第一Cas蛋白共同的至少3个连续氨基酸残基,其中所述第一多肽序列和所述第三多肽序列在所述嵌合多肽中不连续;和/或

(10)所述嵌合多肽的长度小于或等于约600个氨基酸、小于或等于约550个氨基酸、或小于或等于约500个氨基酸;和/或

(11)所述第一Cas蛋白或所述第二Cas蛋白的天然存在形式的长度小于或等于约600个氨基酸、小于或等于约550个氨基酸、或小于或等于约500个氨基酸;和/或

(12)与SEQ ID NO:1的氨基酸序列相比,所述工程化核酸酶变体包含D326或D510处的氨基酸置换,从而降低所述工程化核酸酶变体的核酸酶活性;和/或

(13)与SEQ ID NO:1的氨基酸序列相比,所述工程化核酸酶变体包含选自于由D143R、T147R、K330R和E528R所组成的组中的一个或多个氨基酸置换。

[0351] 实施方式6.一种包含工程化核酸酶变体的工程化多肽,其中所述工程化核酸酶是嵌合多肽,所述嵌合多肽包含:

第一多肽序列(CP1),所述第一多肽序列包含与第一Cas蛋白的部分共同的至少3个连续氨基酸残基;

第二多肽序列(CP2),所述第二多肽序列包含与不同于所述第一Cas蛋白的第二Cas蛋白的部分共同的至少3个连续氨基酸残基;和

第三多肽序列(CPx),所述第三多肽序列包含与以下共同的至少3个连续氨基酸残基:

(i)所述第一Cas蛋白的另外的部分,其中所述第一Cas蛋白的所述部分和所述另外的部分在所述第一Cas蛋白中彼此不直接相邻;

(ii)所述第二Cas蛋白的另外的部分,其中所述第二Cas蛋白的所述部分和所述另外的部分在所述第二Cas蛋白中彼此不直接相邻;或

(iii)不同于所述第一Cas蛋白和所述第二Cas蛋白的第三Cas蛋白的部分,

其中,所述嵌合多肽的长度小于或等于约1,000个氨基酸,

任选地,其中:

(1)所述嵌合多肽从N端到C端具有如式I所示的结构:

CP1-CP2-CPx(I);和/或

(2)所述第一Cas蛋白、所述第二Cas蛋白或所述第三Cas蛋白不是Cas12a;和/或

(3)所述第一Cas蛋白或所述第三Cas蛋白包含与SEQ ID NO:1的多肽序列至少85%、至少90%、至少95%、至少99%或100%相同的氨基酸序列;和/或

(4)所述第二Cas蛋白包含与来自表2的Cas蛋白的多肽序列至少80%、至少85%、至少90%、至少95%、至少99%或100%相同的氨基酸序列;和/或

(5)选自表2的Cas蛋白是Un2Cas12f1;和/或

(6)所述第三Cas蛋白包含与来自表2的不同Cas蛋白的多肽序列至少80%、至少85%、至少90%、至少95%、至少99%或100%相同的氨基酸序列;和/或

(7)所述CP1多肽包含与所述第一Cas蛋白的所述部分共同的至少或多至约5个连续氨基酸残基、至少或多至约10个连续氨基酸残基、至少或多至约20个连续氨基酸残基、至少或多至约30个连续氨基酸残基、至少或多至约40个连续氨基酸残基、或至少或多至约50

个连续氨基酸残基;和/或

(8)所述CP2多肽包含与所述第二Cas蛋白的所述部分共同的至少或多至约5个连续氨基酸残基、至少或多至约8个连续氨基酸残基、至少或多至约10个连续氨基酸残基、或至少或多至约20个连续氨基酸残基;和/或

(9)所述CPx多肽包含与(i)、(ii)和/或(iii)共同的至少或多至约5个连续氨基酸残基、至少或多至约10个连续氨基酸残基、至少或多至约20个连续氨基酸残基、至少或多至约30个连续氨基酸残基、至少或多至约40个连续氨基酸残基、或至少或多至约50个连续氨基酸残基;和/或

(10)所述嵌合多肽的长度小于或等于约600个氨基酸、小于或等于约550个氨基酸、或小于或等于约500个氨基酸;和/或

(11)所述第一Cas蛋白、所述第二Cas蛋白或所述第三Cas蛋白的天然存在形式的长度小于或等于约600个氨基酸、小于或等于约550个氨基酸、或小于或等于约500个氨基酸;和/或

(12)与SEQ ID NO:1的氨基酸序列相比,所述工程化核酸酶变体包含D326或D510处的氨基酸置换,从而降低所述工程化核酸酶变体的核酸酶活性;和/或

(13)与SEQ ID NO:1的氨基酸序列相比,所述工程化核酸酶变体包含选自于由D143R、T147R、K330R和E528R所组成的组中的一个或多个氨基酸置换。

[0352] 实施方式7.一种包含工程化核酸酶变体的工程化多肽,其中所述工程化核酸酶变体是嵌合多肽,所述嵌合多肽包含:

第一多肽,所述第一多肽包含与第一Cas蛋白共同的至少3个连续氨基酸残基;和

第二多肽,所述第二多肽包含与第二Cas蛋白共同的至少3个连续氨基酸残基,其中所述第二Cas蛋白不同于所述第一Cas蛋白,

其中,所述第二多肽序列的长度小于所述第一多肽序列的长度的约20%,

任选地,其中:

(1)所述第二多肽序列的长度小于所述第一多肽序列的长度的约10%;和/或

(2)所述第二多肽序列的长度大于所述第一多肽序列的长度的约1%;和/或

(3)所述第一Cas蛋白、所述第二Cas蛋白或所述第三Cas蛋白不是Cas12a;和/或

(4)所述第一多肽序列包含第一子结构域和第二子结构域,其(i)各自包含与所述第一Cas蛋白共同的至少3个连续氨基酸残基,(ii)彼此不同,并且(ii)在所述嵌合多肽中不连续;和/或

(5)所述第一Cas蛋白包含与SEQ ID NO:1的多肽序列至少85%、至少90%、至少95%、至少99%或100%相同的氨基酸序列;和/或

(6)所述第二Cas蛋白包含与选自表2的Cas蛋白的多肽序列至少80%、至少85%、至少90%、至少95%、至少99%或100%相同的氨基酸序列;和/或

(7)选自表2的Cas蛋白是Un2Cas12f1;和/或

(8)所述第一多肽包含与所述第一Cas蛋白共同的至少或多至约5个连续氨基酸残基、至少或多至约10个连续氨基酸残基、至少或多至约20个连续氨基酸残基、至少或多至约30个连续氨基酸残基、至少或多至约40个连续氨基酸残基、或至少或多至约50个连续氨基酸残基;和/或

(9)所述第二多肽包含与所述第二Cas蛋白共同的至少或多至约5个连续氨基酸残基、至少或多至约8个连续氨基酸残基、至少或多至约10个连续氨基酸残基、或至少或多至约20个连续氨基酸残基;和/或

(10)所述嵌合多肽的长度小于或等于约600个氨基酸、小于或等于约550个氨基酸、或小于或等于约500个氨基酸;和/或

(11)所述第一Cas蛋白或所述第二Cas蛋白的天然存在形式的长度小于或等于约600个氨基酸、小于或等于约550个氨基酸、或小于或等于约500个氨基酸;和/或

(12)与SEQ ID NO:1的氨基酸序列相比,所述工程化核酸酶变体包含D326或D510处的氨基酸置换,从而降低所述工程化核酸酶变体的核酸酶活性;和/或

(13)与SEQ ID NO:1的氨基酸序列相比,所述工程化核酸酶变体包含选自于由D143R、T147R、K330R和E528R所组成的组中的一个或多个氨基酸置换。

[0353] 实施方式8.一种包含工程化核酸酶变体的工程化多肽,其中所述工程化核酸酶变体:

(i)包含与选自表3B、表4B或表5C的成员的多肽序列至少80%相同的氨基酸序列

(ii)不是SEQ ID NO:1-3、10和13-19中的任一个;并且

(iii)具有小于或等于约800个氨基酸的长度,

任选地其中:

(1)所述氨基酸序列与选自表3B、表4B或表5C的成员的多肽序列至少85%、至少90%或至少95%相同;和/或

(2)所述成员是来自表4B的mA8、mC18、mC21、mD2、mD4、mD5或mD7;和/或

(3)所述成员是来自表5C的cA2.6、cA2.39、cA2.69、cA2.29、cA2.10、cA2.4、cA2.21、cA2.13、cA2.3、cA2.16、cA2.23、cA2.8、cA2.31、cA2.30、cA2.11、cA2.5、cA2.41、cA2.49、cA2.26、cA2.14、cA2.20、cA2.1、cA2.24、cA2.58、cA2.61、cA2.38、cA2.88、cA2.2、cA2.51、cA2.34、cA2.25、cA2.85、cA2.54、cA2.15、cA2.75、cA2.32、cA2.90、cA2.89、orcA2.46;和/或

(4)所述成员不是来自表5C的cA2.55或cA2.84;和/或

(5)所述氨基酸序列的长度小于或等于约600个氨基酸、小于或等于约550个氨基酸、或小于或等于约500个氨基酸;和/或

(6)与SEQ ID NO:1的氨基酸序列相比,所述工程化核酸酶变体包含D326或D510处的氨基酸置换,从而降低所述工程化核酸酶变体的核酸酶活性;和/或

(7)与SEQ ID NO:1的氨基酸序列相比,所述工程化核酸酶变体包含选自于由D143R、T147R、K330R和E528R所组成的组中的一个或多个氨基酸置换。

[0354] 实施方式9.一种包含可操作地偶联至基因调节因子的工程化核酸酶变体的工程化多肽,其中所述工程化核酸酶变体:

(i)包含与选自SEQ ID NO:1-3和13-19的成员的多肽序列至少70%相同的氨基酸序列;并且

(ii)当可操作地偶联至所述基因调节因子时,与包含可操作地偶联至所述基因调节因子的SEQ ID NO:10的对照工程化多肽诱导的相比,诱导细胞中靶基因的增强的调节,

任选地其中:

(1)所述氨基酸序列与SEQ ID NO:1的多肽序列至少70％相同;和/或

(2)所述靶基因的增强的调节的特征在于,所述靶基因的表达水平的更大变化,其比所述对照工程化多肽导致的大至少50％、至少80％或至少100％;和/或

(3)所述靶基因的增强的调节的特征在于,与所述对照工程化多肽导致的相比,所述靶基因的表达水平的更大的增加;和/或

(4)所述靶基因的增强的调节的特征在于,与所述对照工程化多肽导致的相比,所述靶基因的表达水平的更大的降低;和/或

(5)所述靶基因的增强的调节的特征在于,所述靶基因的表达水平的延长的变化,与所述对照工程化多肽导致的相比更长;和/或

(6)所述延长的变化比所述对照工程化多肽导致的长至少20％、至少50％、至少80％或至少100％;和/或

(7)所述氨基酸序列的长度小于或等于约600个氨基酸、小于或等于约550个氨基酸、或小于或等于约500个氨基酸;和/或

(8)所述氨基酸序列的长度小于529个氨基酸;和/或

(9)与由SEQ ID NO:1编码的核酸酶相比,所述工程化核酸酶变体表现出降低的核酸酶活性;和/或

(10)所述氨基酸序列与SEQ ID NO:1的多肽序列至少75％、至少80％或至少85％相同;和/或

(11)所述氨基酸序列与SEQ ID NO:1的多肽序列至多95％相同;和/或

(12)所述氨基酸序列与SEQ ID NO:12的多肽序列至少70％、至少80％、至少85％、至少90％或至少95％相同;和/或

(13)所述细胞是哺乳动物细胞;和/或

(14)所述靶基因是基因组序列;和/或

(15)所述基因调节因子是基因激活因子;和/或

(16)所述基因调节因子是基因阻抑因子;和/或

(17)所述基因调节因子与所述工程化多肽中的工程化核酸酶变体融合;和/或

(18)所述基因调节因子未与所述工程化核酸酶变体融合;和/或

(19)与SEQ ID NO:1的氨基酸序列相比,所述工程化核酸酶变体包含D326或D510处的氨基酸置换,从而降低所述工程化核酸酶变体的核酸酶活性;和/或

(20)与SEQ ID NO:1的氨基酸序列相比,所述工程化核酸酶变体包含选自于由D143R、T147R、K330R和E528R所组成的组中的一个或多个氨基酸置换。

[0355] 实施方式10.一种包含本文提供的任一实施方式的工程化多肽的系统,

任选地,其中:

(1)所述系统还包含能够与所述工程化多肽形成复合体的引导核酸,其中所述复合体表现出与细胞中的靶基因的特异性结合;和/或

(2)本文提供的任一实施方式的引导核酸分子,

任选地,其中所述引导核酸表现出与选自表6B、表7B和表8B的成员的多核苷酸序列的至少80％序列同一性;和/或

(3)所述引导核酸分子的支架序列与选自SEQ ID NO:500、549、550、551和/或552

的成员不相同。

[0356] 实施方式11.编码本文提供的任一实施方式的系统的一种或多种多核苷酸。

[0357] 实施方式12.一种包含本文提供的任一实施方式的系统的细胞。

[0358] 实施方式13.一种控制细胞中的靶基因的方法,所述方法包括使所述细胞与本文提供的任一实施方式的工程化多肽或系统接触,

任选地其中:

(1)所述控制包括所述细胞中所述靶基因中的一个或多个碱基的插入、缺失和/或突变;和/或

(2)所述控制包括调控所述细胞中所述靶基因的表达和/或活性水平;和/或

(3)所述调节包括激活所述靶基因的表达和/或活性水平;和/或

(4)所述调控包括降低所述靶基因的表达和/或活性水平;和/或

(5)与包含可操作地偶联至基因调节因子的SEQ ID NO:10的对照工程化多肽诱导的相比,可操作地偶联至所述基因调节因子的工程核酸酶(或工程核酸酶变体)诱导细胞中靶基因的增强的调节。

[0359] 实施方式14.一种调节细胞中的靶基因的方法,所述方法包括:

使所述细胞与包含可操作地偶联至基因调节因子的工程化核酸酶变体的工程化多肽接触,其中所述工程化核酸酶变体包含与选自SEQ ID NO:1-3和13-19的成员的多肽序列至少70%相同的氨基酸序列,

其中与包含可操作地偶联至所述基因调节因子的SEQ ID NO:10的对照工程化多肽导致的相比,所述接触导致所述细胞中所述靶基因的增强的调节,

任选地,其中:

(1)所述氨基酸序列与SEQ ID NO:1的多肽序列至少70%相同;和/或

(2)所述靶基因的增强的调节的特征在于,所述靶基因的表达水平的更大变化,其比所述对照工程化多肽导致的大至少50%、至少80%或至少100%;和/或

(3)所述靶基因的增强的调节的特征在于,与所述对照工程化多肽导致的相比,所述靶基因的表达水平的更大的增加;和/或

(4)所述靶基因的增强的调节的特征在于,与所述对照工程化多肽导致的相比,所述靶基因的表达水平的更大的降低;和/或

(5)所述靶基因的增强的调节的特征在于,所述靶基因的表达水平的延长的变化,与所述对照工程化多肽导致的相比更长;和/或

(6)所述延长的变化比所述对照工程化多肽导致的长至少20%、至少50%、至少80%或至少100%;和/或

(7)与SEQ ID NO:1的多肽序列相比,所述工程化核酸酶的氨基酸序列包含来自氨基酸残基2-100的至少一个缺失;和/或

(8)所述工程化核酸酶包含与SEQ ID NO:12的多肽序列大于92%相同的氨基酸序列;和/或

(9)所述工程化核酸酶变体包含与SEQ ID NO:12的多肽序列至少70%相同的氨基酸序列,

其中与SEQ ID NO:1的多肽序列相比,所述氨基酸序列包含修饰,其中,所述修饰

包括选自于由A21Q、V23I、N32E、D29E、N33R、E35K、K36Q、I37A、A38G、E40D、K73G、A74T、R75G、K76E、Q83K、G87K、E151A、A340S、H353K、A374K、I387E、N423D、K473Q、T474L、T474R、H497K、L515R、N519T、K521D、K521N、L522I,以及来自SEQ ID NO:1的氨基酸残基400-529的至少一个缺失所组成的组中的一个或多个成员;和/或

(10)所述工程化核酸酶变体是嵌合多肽,所述嵌合多肽包含:

第一多肽序列,所述第一多肽序列包含与第一Cas蛋白共同的至少3个连续氨基酸残基;和

第二多肽序列,所述第二多肽序列包含与第二Cas蛋白共同的至少3个连续氨基酸残基,其中所述第二Cas蛋白不同于所述第一Cas蛋白,

其中所述第一Cas蛋白包含与SEQ ID NO:1的多肽序列至少80%相同的氨基酸序列;和/或

(11)所述工程化核酸酶是嵌合多肽,所述嵌合多肽包含:

第一多肽序列(CP1),所述第一多肽序列包含与第一Cas蛋白的部分共同的至少3个连续氨基酸残基;

第二多肽序列(CP2),所述第二多肽序列包含与不同于所述第一Cas蛋白的第二Cas蛋白的部分共同的至少3个连续氨基酸残基;和

第三多肽序列(CPx),所述第三多肽序列包含与以下共同的至少3个连续氨基酸残基:

(i)所述第一Cas蛋白的另外的部分,其中所述第一Cas蛋白的所述部分和所述另外的部分在所述第一Cas蛋白中彼此不直接相邻;

(ii)所述第二Cas蛋白的另外的部分,其中所述第二Cas蛋白的所述部分和所述另外的部分在所述第二Cas蛋白中彼此不直接相邻;或

(iii)不同于所述第一Cas蛋白和所述第二Cas蛋白的第三Cas蛋白的部分,

其中所述嵌合多肽的长度小于或等于约1,000个氨基酸;和/或

(12)所述工程化核酸酶变体是嵌合多肽,所述嵌合多肽包含:

第一多肽,所述第一多肽包含与第一Cas蛋白共同的至少3个连续氨基酸残基;和

第二多肽,所述第二多肽包含与第二Cas蛋白共同的至少3个连续氨基酸残基,其中所述第二Cas蛋白不同于所述第一Cas蛋白,

其中所述第二多肽序列的长度小于所述第一多肽序列的长度的约20%;和/或

(13)所述工程化核酸酶变体:

(i)包含与选自表3B、表4B或表5C的成员的多肽序列至少80%相同的氨基酸序列;

(ii)不是SEQ ID NO:1-3、10和13-19中的任一个;并且

(iii)具有小于或等于约800个氨基酸的长度;和/或

(14)所述工程化核酸酶变体:

(i)包含与SEQ ID NO:1的多肽序列至少70%相同的氨基酸序列;并且

(ii)当可操作地偶联至所述基因调节因子时,与包含可操作地偶联至所述基因调节因子的SEQ ID NO:10的对照工程化多肽诱导的相比,诱导细胞中靶基因的增强的调节;和/或

(15)所述氨基酸序列的长度小于或等于约600个氨基酸、小于或等于约550个氨基

酸、或小于或等于约500个氨基酸;和/或

(16)所述氨基酸序列的长度小于529个氨基酸;和/或

(17)与由SEQ ID NO:1编码的核酸酶相比,所述工程化核酸酶变体表现出降低的核酸酶活性;和/或

(18)所述氨基酸序列与SEQ ID NO:1的多肽序列至少75%、至少80%或至少85%相同;和/或

(19)所述氨基酸序列与SEQ ID NO:1的多肽序列至多95%相同;和/或

(20)所述氨基酸序列与SEQ ID NO:12的多肽序列至少70%、至少80%、至少85%、至少90%或至少95%相同;和/或

(21)所述细胞是哺乳动物细胞;和/或

(22)所述靶基因是基因组序列;和/或

(23)所述基因调节因子是基因激活因子;和/或

(24)所述基因调节因子是基因阻抑因子;和/或

(25)所述基因调节因子与所述工程化多肽中的工程化核酸酶变体融合;和/或

(26)所述基因调节因子未与所述工程化核酸酶变体融合;和/或

(27)其中所述接触包括用包含所述工程化多肽和对可操作地偶联至所述靶基因的靶多核苷酸序列表现出特异性亲和力的引导核酸分子的复合体转染所述细胞;和/或

(28)所述接触包括用编码所述工程化多肽和对可操作地偶联至所述靶基因的靶多核苷酸序列表现出特异性亲和力的引导核酸分子的载体转染所述细胞;和/或

(29)所述载体是质粒或病毒载体。

[0360] 实施方式15.一种组合物,其包含被配置为与Cas蛋白形成复合体的引导核酸分子,其中所述引导核酸分子包含:

表现出与靶多核苷酸序列的特异性结合的间隔序列;和

用于与所述Cas蛋白形成复合体的支架序列,其中所述支架序列的特征在于:

(i)具有与SEQ ID NO:555的多核苷酸序列具有至少96%序列同一性的连续多核苷酸序列;和/或具有与SEQ ID NO:597的多核苷酸序列具有至少84%序列同一性的连续多核苷酸序列;和/或

(ii)具有与SEQ ID NO:557的多核苷酸序列具有至少97%序列同一性的连续多核苷酸序列;和/或具有与SEQ ID NO:598的多核苷酸序列具有至少88%序列同一性的连续多核苷酸序列;和/或

(iii)具有与SEQ ID NO:578的多核苷酸序列具有至少90%序列同一性的连续多核苷酸序列;和/或具有与SEQ ID NO:597的多核苷酸序列具有至少84%序列同一性的连续多核苷酸序列;和/或具有与SEQ ID NO:599的多核苷酸序列具有至少81%序列同一性的连续多核苷酸序列;和/或

(iv)具有与SEQ ID NO:568的多核苷酸序列具有至少93%序列同一性的连续多核苷酸序列;和/或具有与SEQ ID NO:597的多核苷酸序列具有至少84%序列同一性的连续多核苷酸序列;和/或具有与SEQ ID NO:600的多核苷酸序列具有至少67%序列同一性的连续多核苷酸序列;和/或

(v)具有与SEQ ID NO:569的多核苷酸序列具有至少95%序列同一性的连续多核

苷酸序列;和/或具有与SEQ ID NO:597的多核苷酸序列具有至少84%序列同一性的连续多核苷酸序列;和/或具有与SEQ ID NO:601的多核苷酸序列具有至少71%序列同一性的连续多核苷酸序列,

任选地其中:

(1)支架序列(i-a)具有与SEQ ID NO:555的多核苷酸序列具有至少96%、至少97%、至少98%、至少99%或100%序列同一性的连续多核苷酸序列;或(i-b)具有与SEQ ID NO:597的多核苷酸序列具有至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少89%、至少90%、至少95%、至少98%、至少99%或100%序列同一性的连续多核苷酸序列;和/或

(2)支架序列(ii-a)具有与SEQ ID NO:557的多核苷酸序列具有至少97%、至少98%、至少99%或100%序列同一性的连续多核苷酸序列;或(ii-b)具有与SEQ ID NO:598的多核苷酸序列具有至少88%、至少89%、至少90%、至少95%、至少98%、至少99%或100%序列同一性的连续多核苷酸序列;和/或

(3)支架序列(iii-a)具有与SEQ ID NO:578的多核苷酸序列具有至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%、至少99%或100%序列同一性的连续多核苷酸序列;(iii-b)具有与SEQ ID NO:597的多核苷酸序列具有至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少89%、至少90%、至少95%、至少98%、至少99%或100%序列同一性的连续多核苷酸序列;或(iii-c)具有与SEQ ID NO:599的多核苷酸序列具有至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少89%、至少90%、至少95%、至少98%、至少99%或100%序列同一性的连续多核苷酸序列;和/或

(4)支架序列(iv-a)具有与SEQ ID NO:568的多核苷酸序列具有至少93%、至少94%、至少95%、至少96%、至少97%、至少98%、至少99%或100%序列同一性的连续多核苷酸序列;(iv-b)具有与SEQ ID NO:597的多核苷酸序列具有至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少89%、至少90%、至少95%、至少98%、至少99%或100%序列同一性的连续多核苷酸序列;或(iv-c)具有与SEQ ID NO:600的多核苷酸序列具有至少67%、至少68%、至少69%、至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%序列同一性的连续多核苷酸序列;和/或

(5)支架序列(v-a)具有与SEQ ID NO:569的多核苷酸序列具有至少95%、至少96%、至少97%、至少98%、至少99%或100%序列同一性的连续多核苷酸序列;(v-b)具有与SEQ ID NO:597的多核苷酸序列具有至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少89%、至少90%、至少95%、至少98%、至少99%或100%序列同一性的连续多核苷酸序列;或(v-c)具有与SEQ ID NO:601的多核苷酸序列具有至少71%、至少72%、至少73%、至少74%、至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%序列同一性的连续多核苷酸序列;和/或

(6)所述引导核酸分子的长度为至多约150个、至多约140个、至多约130个或至多约125个核苷酸;和/或

(7)所述支架序列的长度为至多约130个、至多约120个、至多约110个或至多约105个核苷酸;和/或

(8)所述支架序列的长度为至少约95个、至少约99个或至少约100个核苷酸;和/或

(9)所述复合体与细胞中的靶多核苷酸序列的结合导致所述细胞中靶基因的调节的表达水平,其中(A1)所述复合体导致的所述靶基因的调节的表达水平与(A2)包含所述Cas蛋白和含有SEQ ID NO:500的多核苷酸序列的对照引导核酸分子的对照复合体导致的相当或优于后者。

[0361] 实施方式16.一种组合物,其包含被配置为与Cas蛋白形成复合体的引导核酸分子,其中所述引导核酸分子包含:

表现出与可操作地偶联至靶基因的靶多核苷酸序列的特异性结合的间隔序列;和

用于与所述Cas蛋白形成复合体的支架序列,其中所述支架序列表现出与选自表6B、表7B和表8B的成员的多核苷酸序列的至少80%序列同一性,其中所述支架序列与SEQ ID NO:500不同,

任选地,其中所述复合体与细胞中的所述靶多核苷酸序列的结合导致所述细胞中所述靶基因的调节的表达水平,其中(A1)所述复合体导致的所述靶基因的调节的表达水平与(A2)包含所述Cas蛋白和含有SEQ ID NO:500的多核苷酸序列的对照引导核酸分子的对照复合体导致的相当或优于后者,

进一步任选地,其中:

(1)所述支架序列表现出与选自于由SEQ ID NO:503-152、519、524、528和553所组成的组中的成员的多核苷酸序列的至少80%、至少85%、至少90%、至少95%或至少99%序列同一性;和/或

(2)所述支架序列表现出与选自于由SEQ ID NO:555、557、558、568、569、578和580所组成的组中的成员的多核苷酸序列的至少80%、至少85%、至少90%、至少95%或至少99%序列同一性;和/或

(3)所述支架序列表现出与选自于由SEQ ID NO:555、557、568、569、576、577、578、580、593、519和528所组成的组中的成员的多核苷酸序列的至少80%、至少85%、至少90%、至少95%或至少99%序列同一性;和/或

(4)所述支架序列的长度为至多约158个核苷酸。

[0362] 实施方式17.一种组合物,其包含被配置为与Cas蛋白形成复合体的引导核酸分子,其中所述引导核酸分子包含:

表现出与可操作地偶联至靶基因的靶多核苷酸序列的特异性结合的间隔序列;和

用于与所述Cas蛋白形成复合体的支架序列,其中所述支架序列的长度为至多约158个核苷酸,

其中所述复合体与细胞中的靶多核苷酸序列的结合导致所述细胞中所述靶基因的调节的表达水平,其中(A1)所述复合体导致的所述靶基因的调节的表达水平与(A2)包含所述Cas蛋白和含有SEQ ID NO:500的多核苷酸序列的对照引导核酸分子的对照复合体导致的相当或优于后者,

任选地,其中:

(1)所述支架序列表现出与选自表6B、表7B和表8B的成员的多核苷酸序列的至少80%互补性,其中所述支架序列与SEQ ID NO:500不同;和/或

(2)所述支架序列表现出与选自于由SEQ ID NO:503-152、519、524、528和553所组

成的组中的成员的多核苷酸序列的至少80％、至少85％、至少90％、至少95％或至少99％序列同一性;和/或

(3)所述支架序列表现出与选自于由SEQ ID NO:555、557、558、568、569、578和580所组成的组中的成员的多核苷酸序列的至少80％、至少85％、至少90％、至少95％或至少99％序列同一性;和/或

(4)所述支架序列表现出与选自于由SEQ ID NO:555、557、568、569、576、577、578、580、593、519和528所组成的组中的成员的多核苷酸序列的至少80％、至少85％、至少90％、至少95％或至少99％序列同一性。

[0363] 实施方式18.一种组合物,其包含编码Cas蛋白和被配置为与所述Cas蛋白形成复合体的引导核酸分子的载体,其中所述载体包含:

编码所述Cas蛋白的第一多核苷酸序列;和

编码用于与所述Cas蛋白形成复合体的所述引导核酸分子的支架序列的第二多核苷酸序列,

其中所述第一多核苷酸序列的长度与所述第二多核苷酸序列的长度的和为至多约1700个核苷酸,

任选地其中:

(1)所述和为至多约1650个、至多约1620个或至多约1600个核苷酸;和/或

(2)所述第一多核苷酸序列的长度为至多约1550个、至多约1520个或至多约1500个核苷酸;和/或

(3)所述第二多核苷酸序列的长度为至多约135个、至多约130个、至多约125个、至多约120个、至多约115个、至多约110个或至多约105个核苷酸;和/或

(4)所述组合物还包含编码所述引导核酸分子的间隔序列的第三多核苷酸序列,所述间隔序列表现出与所述靶多核苷酸序列的特异性结合;和/或

(5)所述复合体被配置为与可操作地偶联至所述靶基因的靶序列结合,从而实现所述靶基因的调节的表达水平;和/或

(6)所述支架序列表现出与选自表6B、表7B和表8B的成员的多核苷酸序列的至少80％序列同一性,其中所述支架序列与SEQ ID NO:500不同。

[0364] 实施方式19.本文提供的任一实施方式的组合物,其中:

(1)所述支架序列的长度为至多约150个、至多约140个、至多约130个、至多约102个、至多约110个或至多约105个核苷酸;和/或

(2)所述Cas蛋白的长度为至多约535个或至多约530个氨基酸残基;和/或

(3)(A1)与(A2)相当;和/或

(4)(A1)与(A2)相差不超过(A2)的20％、不超过15％、不超过10％、不超过5％、不超过2％或不超过1％;和/或

(5)(A1)优于(A2);和/或

(6)与(A2)相比,(A1)优于(A2)至少5％、至少10％、至少15％、至少20％、至少50％、至少100％、至少150％、至少200％或至少300％;和/或

(7)所述靶基因的调节的表达水平的特征在于所述靶基因的降低的表达水平;和/或

(8)所述靶基因的调节的表达水平的特征在于所述靶基因的提高的表达水平;和/或

(9)所述靶基因的调节的表达水平的特征在于所述靶基因的调节的表达水平的持续时间;和/或

(10)所述复合体导致的所述靶基因的调节的表达水平是在没有所述靶多核苷酸序列的切割或插入缺失(indel)的情况下进行的;和/或

(11)所述Cas蛋白可操作地偶联至基因调节因子;和/或

(12)所述Cas蛋白与所述基因调节因子融合;和/或

(13)所述基因调节因子是基因激活因子;和/或

(14)所述基因调节因子是基因阻抑因子;和/或

(15)所述引导核酸分子还包含被配置为募集所述基因调节因子的适体,以形成所述复合体;和/或

(16)所述组合物还包含本文提供的任一实施方式的工程化多肽。

[0365] 实施方式20.一种包含本文提供的任一实施方式的组合物的系统,

任选地,其中:

(1)所述系统包含所述Cas蛋白;和/或

(2)所述Cas蛋白包含本文提供的任一实施方式的工程化多肽,进一步任选地,其中所述工程化核酸酶变体:

(i)包含与选自表3B、表4B或表5C的成员的多肽序列至少80%相同的氨基酸序列;

(ii)不是SEQ ID NO:1-3、10和13-19中的任一个;并且

(iii)具有小于或等于约800个氨基酸的长度。

[0366] 实施方式21.编码本文提供的任一实施方式的系统的一种或多种多核苷酸。

[0367] 实施方式22.一种包含本文提供的任一实施方式的系统的细胞。

[0368] 实施方式23.一种控制细胞中的靶基因的方法,所述方法包括使所述细胞与本文提供的任一实施方式的组合物或系统接触,

任选地,其中:

(1)所述控制包括所述细胞中所述靶基因中的一个或多个碱基的插入、缺失和/或突变;和/或

(2)所述控制包括调控所述细胞中所述靶基因的表达和/或活性水平;和/或

(3)所述调控包括激活所述靶基因的表达和/或活性水平;和/或

(4)所述调控包括降低所述靶基因的表达和/或活性水平。

[0369] 实施方式24.一种调节细胞中的靶基因的方法,所述方法包括:

使所述细胞与包含引导核酸分子和Cas蛋白的复合体接触,其中所述复合体表现出与可操作地偶联至所述靶基因的靶多核苷酸序列的特异性结合,

其中,所述复合体与所述靶多核苷酸序列的结合导致所述细胞中所述靶基因的调节的表达水平,其中(A1)所述复合体导致的所述靶基因的调节的表达水平与(A2)包含所述Cas蛋白和含有SEQ ID NO:500的多核苷酸序列的对照引导核酸分子的对照复合体导致的相当或优于后者,

任选地,其中:

(1)(A1)与(A2)相当;和/或

(2)(A1)与(A2)相差不超过(A2)的20%、不超过15%、不超过10%、不超过5%、不超过2%或不超过1%;和/或

(3)(A1)优于(A2);和/或

(4)与(A2)相比,(A1)优于(A2)至少5%、至少10%、至少15%、至少20%、至少50%、至少100%、至少150%、至少200%或至少300%;和/或

(5)所述靶基因的调节的表达水平的特征在于所述靶基因的降低的表达水平;和/或

(6)所述靶基因的调节的表达水平的特征在于所述靶基因的提高的表达水平;和/或

(7)所述靶基因的调节的表达水平的特征在于所述靶基因的调节的表达水平的持续时间;和/或

(8)所述引导核酸分子包含:

表现出与所述靶多核苷酸序列的特异性结合的间隔序列;以及用于与所述Cas蛋白形成复合体的支架序列,其中所述支架序列的特征在于:

(i)具有与SEQ ID NO:555的多核苷酸序列具有至少96%序列同一性的连续多核苷酸序列;和/或具有与SEQ ID NO:597的多核苷酸序列具有至少84%序列同一性的连续多核苷酸序列;和/或

(ii)具有与SEQ ID NO:557的多核苷酸序列具有至少97%序列同一性的连续多核苷酸序列;和/或具有与SEQ ID NO:598的多核苷酸序列具有至少88%序列同一性的连续多核苷酸序列;和/或

(iii)具有与SEQ ID NO:578的多核苷酸序列具有至少90%序列同一性的连续多核苷酸序列;和/或具有与SEQ ID NO:597的多核苷酸序列具有至少84%序列同一性的连续多核苷酸序列;和/或具有与SEQ ID NO:599的多核苷酸序列具有至少81%序列同一性的连续多核苷酸序列;和/或

(iv)具有与SEQ ID NO:568的多核苷酸序列具有至少93%序列同一性的连续多核苷酸序列;和/或具有与SEQ ID NO:597的多核苷酸序列具有至少84%序列同一性的连续多核苷酸序列;和/或具有与SEQ ID NO:600的多核苷酸序列具有至少67%序列同一性的连续多核苷酸序列;和/或

(v)具有与SEQ ID NO:569的多核苷酸序列具有至少95%序列同一性的连续多核苷酸序列;和/或具有与SEQ ID NO:597的多核苷酸序列具有至少84%序列同一性的连续多核苷酸序列;和/或具有与SEQ ID NO:601的多核苷酸序列具有至少71%序列同一性的连续多核苷酸序列;和/或

(9)所述引导核酸分子包含:

表现出与所述靶多核苷酸序列的特异性结合的间隔序列;和

用于与所述Cas蛋白形成复合体的支架序列,其中所述支架序列表现出与选自表6B、表7B和表8B的成员的多核苷酸序列的至少80%序列同一性,其中所述支架序列与SEQ ID NO:500不同;和/或

(10)所述引导核酸分子包含:

表现出与所述靶多核苷酸序列的特异性结合的间隔序列;和

用于与所述Cas蛋白形成复合体的支架序列,其中所述支架序列的长度为至多约158个核苷酸;和/或

(11)所述接触包括利用编码所述Cas蛋白和所述引导核酸分子的载体,其中所述载体包含:

编码所述Cas蛋白的第一多核苷酸序列;和

编码用于与所述Cas蛋白形成复合体的所述引导核酸分子的支架序列的第二多核苷酸序列,

其中所述第一多核苷酸序列的长度与第二多核苷酸序列的长度的和为至多约1700个核苷酸;和/或

(12)所述引导核酸分子的支架序列的长度为至多约150个、至多约140个、至多约130个、至多约102个、至多约110个或至多约105个核苷酸;和/或

(13)所述Cas蛋白的长度为至多约535个或至多约530个氨基酸残基;和/或

(14)所述复合体导致的所述靶基因的调节的表达水平是在没有所述靶多核苷酸序列的切割或插入缺失的情况下进行的;和/或

(15)所述Cas蛋白可操作地偶联至基因调节因子;和/或

(16)Cas蛋白与所述基因调节因子融合;和/或

(17)所述基因调节因子是基因激活因子;和/或

(18)所述基因调节因子是基因阻抑因子;和/或

(19)所述引导核酸分子还包含被配置为募集所述基因调节因子的适体,以形成所述复合体;和/或

(20)所述Cas蛋白包含本文提供的任一实施方式的工程化多肽的工程化核酸酶变体。

[0370] 应当理解,本发明的不同方面可以单独、集体或彼此组合地理解。本文描述的本发明的各个方面可以应用于本文公开的任何特定应用。本文在本公开的组合物部分中公开的物质组合物可用于方法部分,包括本文公开的使用和生产方法,反之亦然。

[0371] 虽然本文已展示和描述了本发明的优选实施方式,但本领域技术人员将明白,这些实施方式仅以示例的方式提供。本发明不旨在受说明书中提供的具体实例的限制。虽然本发明已参考上述说明书进行了描述,但本文对实施方式的描述和说明并非旨在以限制意义来解释。本领域技术人员现在将想到许多变化、改变和替换,而不会背离本发明。此外,应理解,本发明的所有方面不限于本文所述的特定描述、配置或相对比例,这些描述、配置或相对比例取决于各种条件和变量。应理解,在实践本发明时,可以采用本文所述的本发明实施方式的各种替代方案。因此,还考虑本发明应涵盖任何此类替代方案、修改、变化或等效物。以下权利要求旨在定义本发明的范围,并且这些权利要求及其等效物范围内的方法和结构因此应被涵盖。

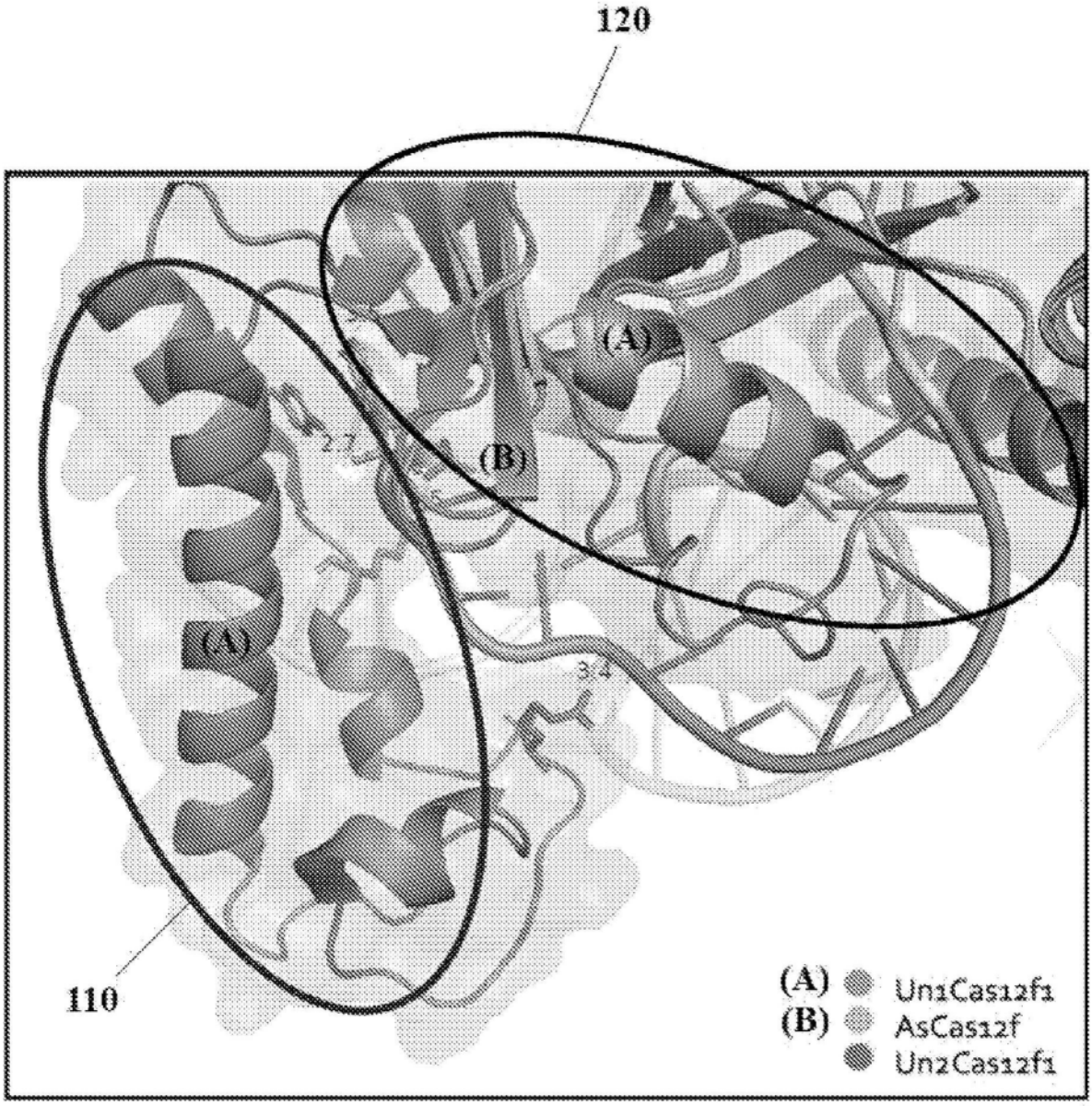


图1

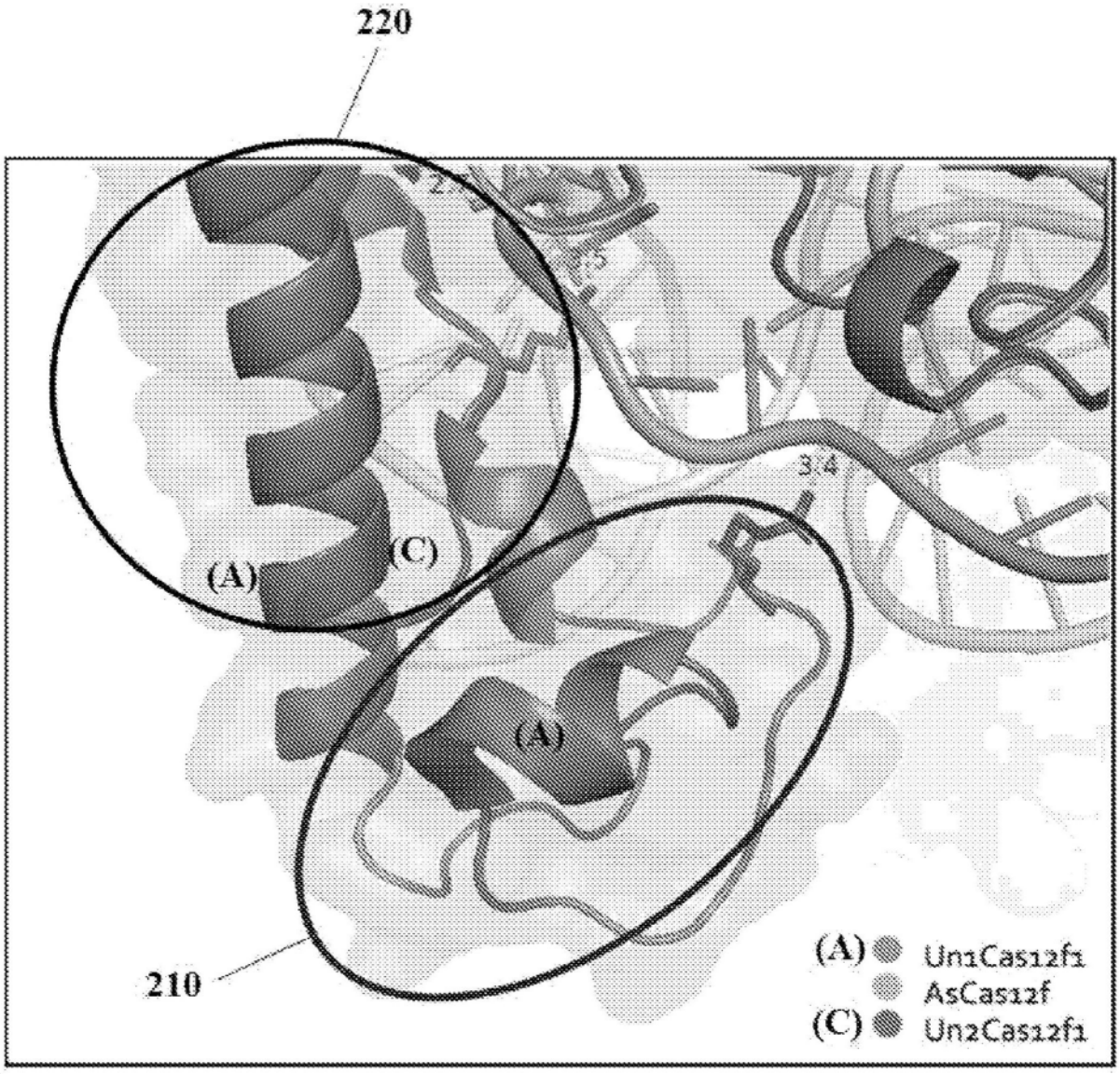

220
(A)
(C)
(A)
210
(A) Un1Cas12f1
AsCas12f
(C) Un2Cas12f1

图2

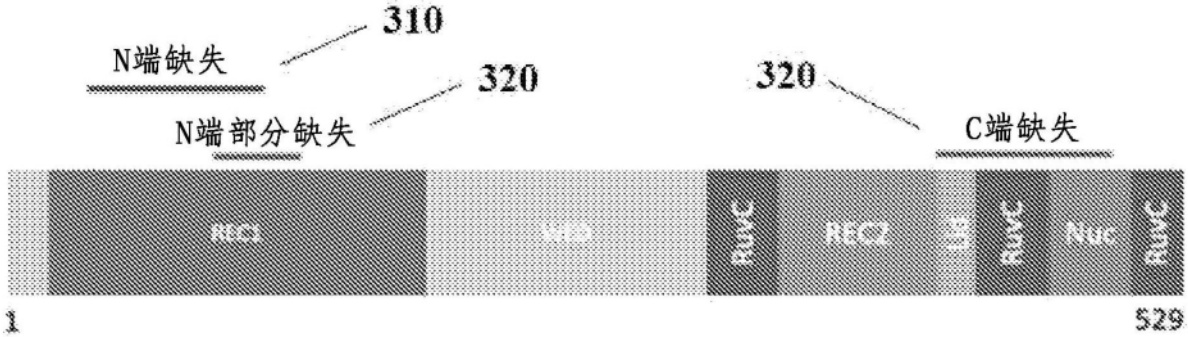

310
N端缺失
320
N端部分缺失
320
C端缺失
REC1
RuvC
REC2
RuvC
Nuc
RuvC
1
529

图3A

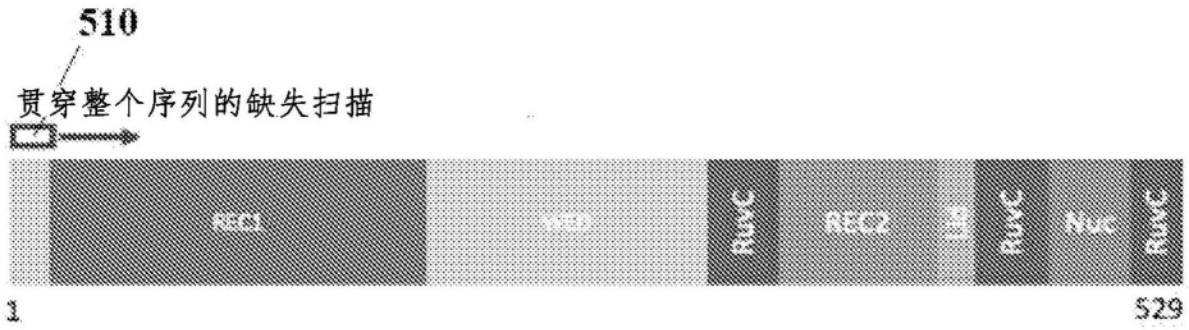


图3B

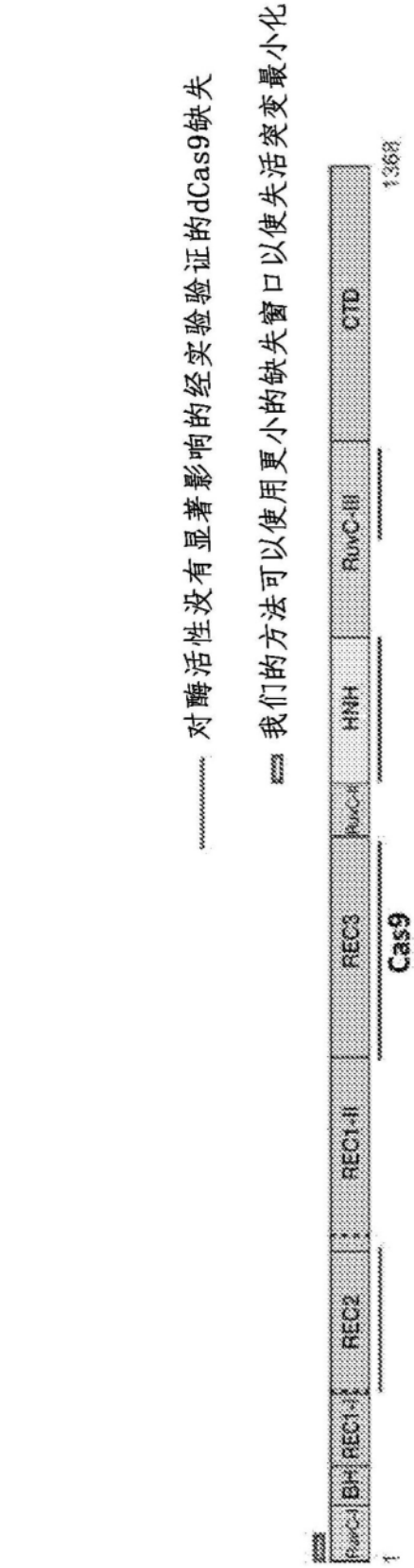

对酶活性没有显著影响的经实验验证的dCas9缺失
我们的方法可以使用更小的缺失窗口以使失活突变最小化
RuvC-I
BH
REC1-I
REC2
REC1-II
REC3
RuvC-II
HNH
RuvC-III
CTD
Cas9
1
1368

图4

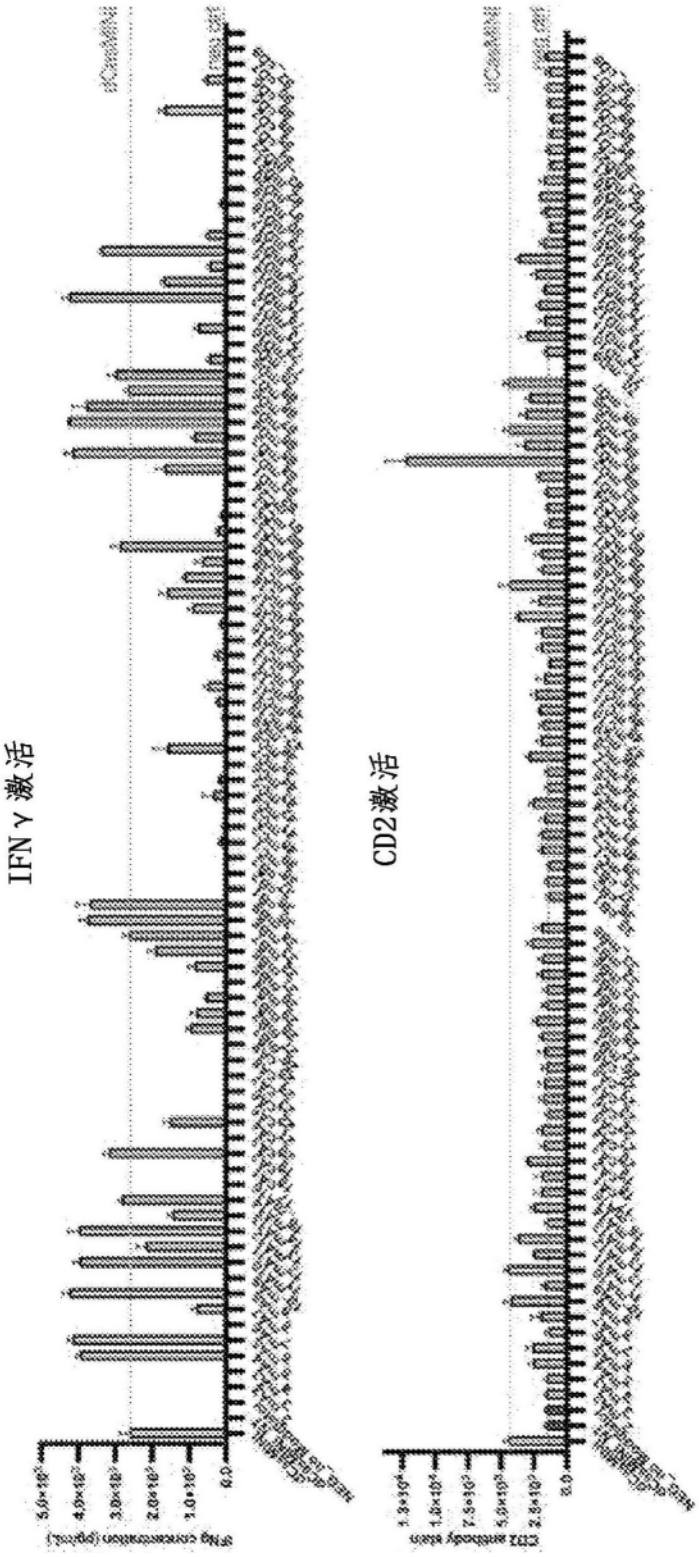


图5

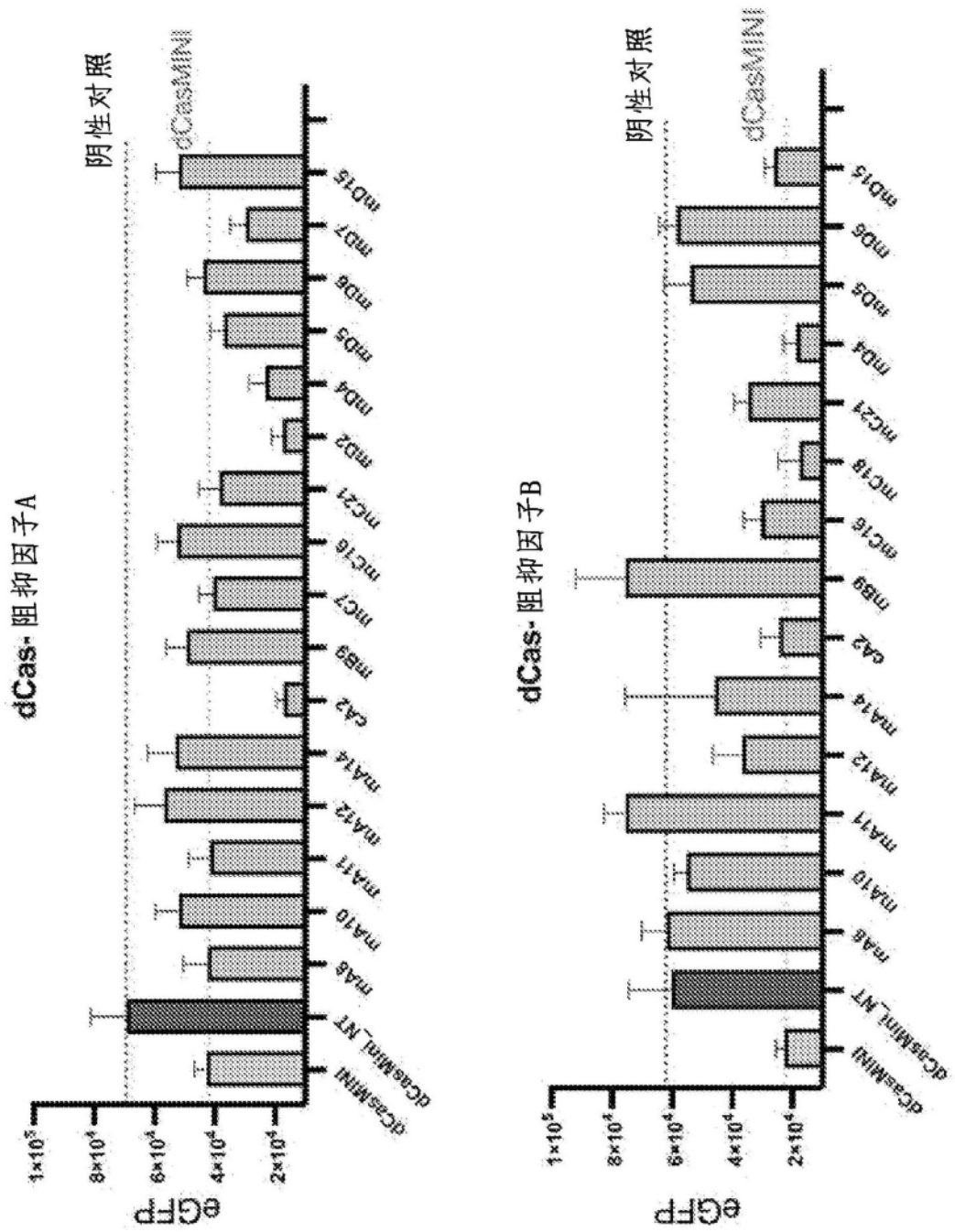

dCas- 阻抑因子A
阴性对照
dCasMINI
mD15
mD7
mD6
mD5
mD4
mD2
mC21
mC16
mC7
mB9
cA2
mA14
mA12
mA11
mA10
mA6
dCasMini_NT
dCasMINI
eGFP
1×10⁵
8×10⁴
6×10⁴
4×10⁴
2×10⁴
dCas- 阻抑因子B
阴性对照
dCasMINI
mD15
mD6
mD5
mD4
mC21
mC19
mC16
mB9
cA2
mA14
mA12
mA11
mA10
mA6
dCasMini_NT
dCasMINI
eGFP
1×10⁵
8×10⁴
6×10⁴
4×10⁴
2×10⁴

图6

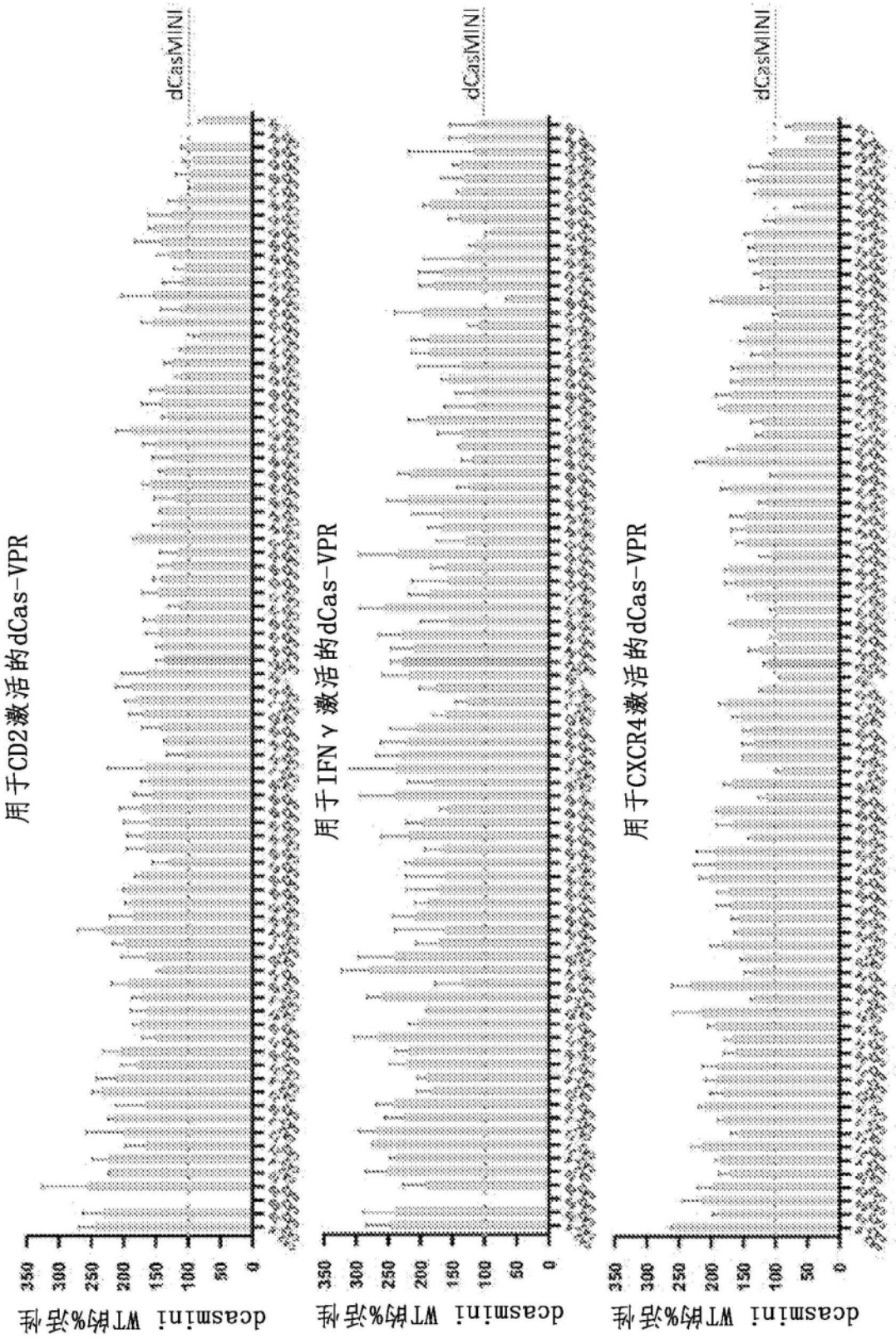

用于CD2激活的dCas-VPR
用于IFNγ激活的dCas-VPR
用于CXCR4激活的dCas-VPR
dCasMINI
dCasMINI
dCasMINI
dcasmini WT的%活性
dcasmini WT的%活性
dcasmini WT的%活性
350
300
250
200
150
100
50
0

图7

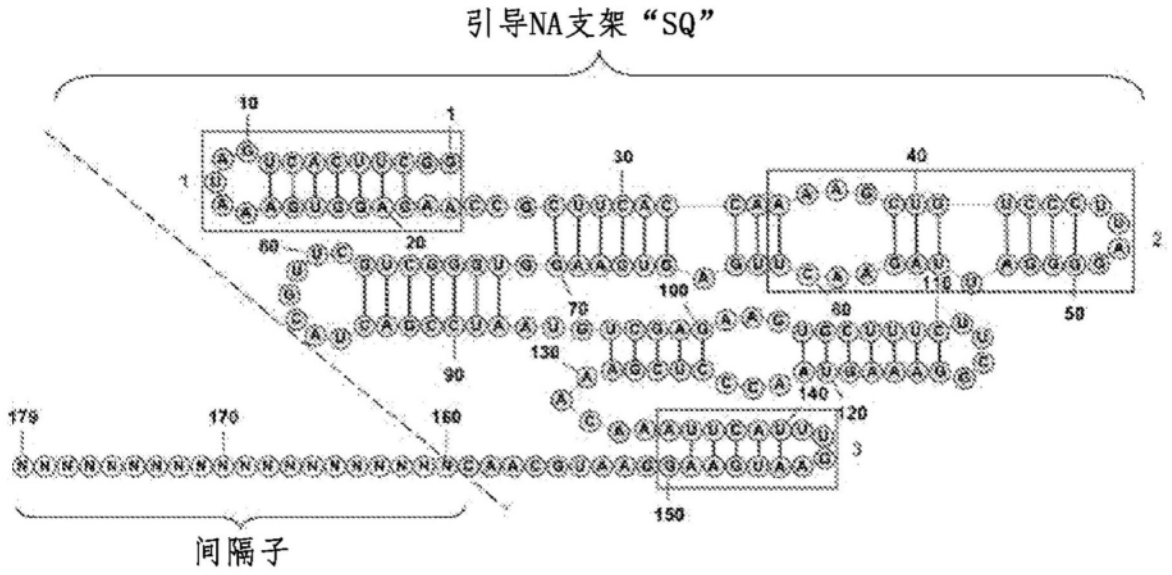

引导NA支架“SQ”
间隔子

图8

截短并且稳定具有共同的mRNA颈环序列的失向的茎2
截短失向的茎1
截短失向的tracr-crRNA界面

图9

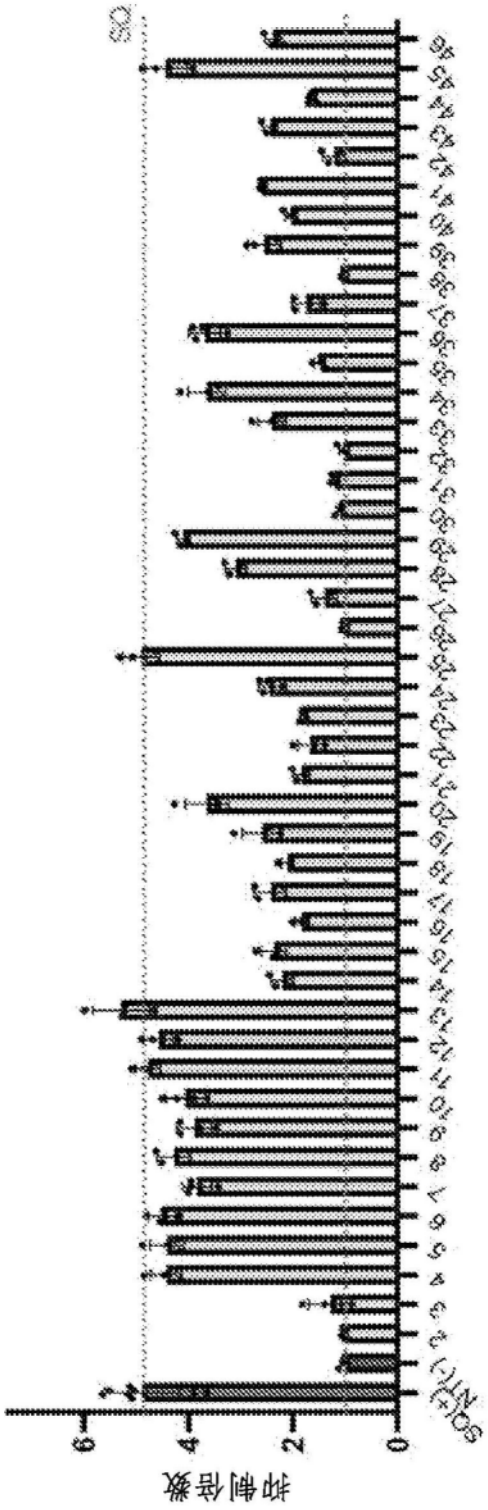

抑制倍数
0
2
4
6
8
SOX(+)
NT(-)
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46

图10

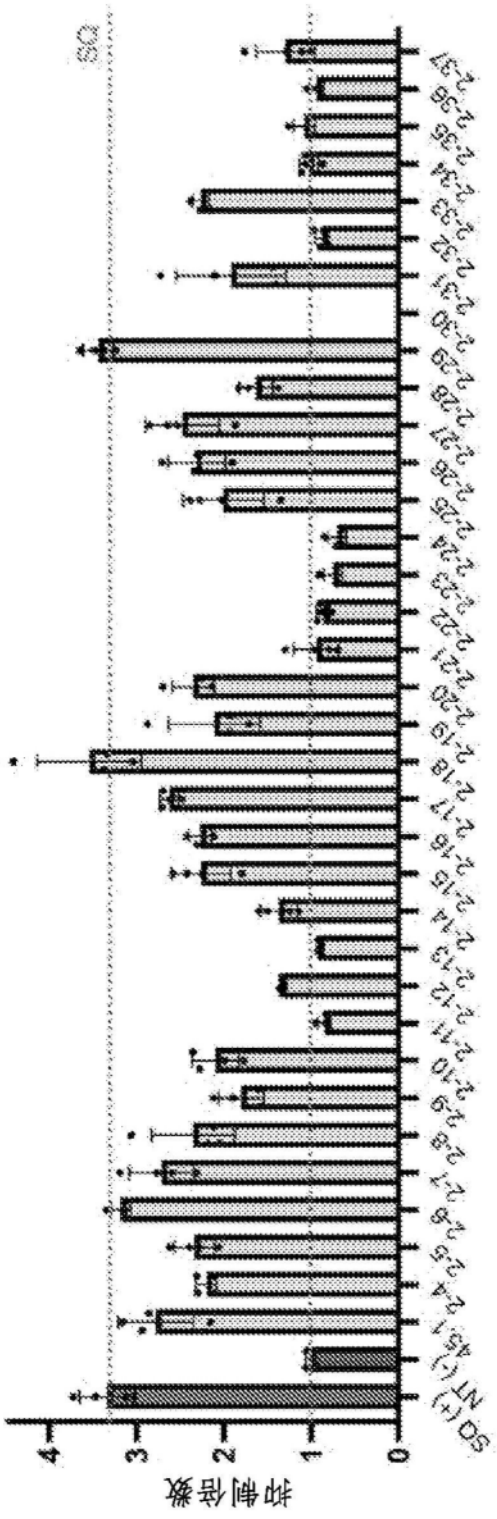

抑制倍数
0
1
2
3
4
SO (+)
NT (-)
2-1
2-4
2-5
2-6
2-7
2-8
2-9
2-10
2-11
2-12
2-13
2-14
2-15
2-16
2-17
2-18
2-19
2-20
2-21
2-22
2-23
2-24
2-25
2-26
2-27
2-28
2-29
2-30
2-31
2-32
2-33
2-34
2-35
2-36
2-37

图11

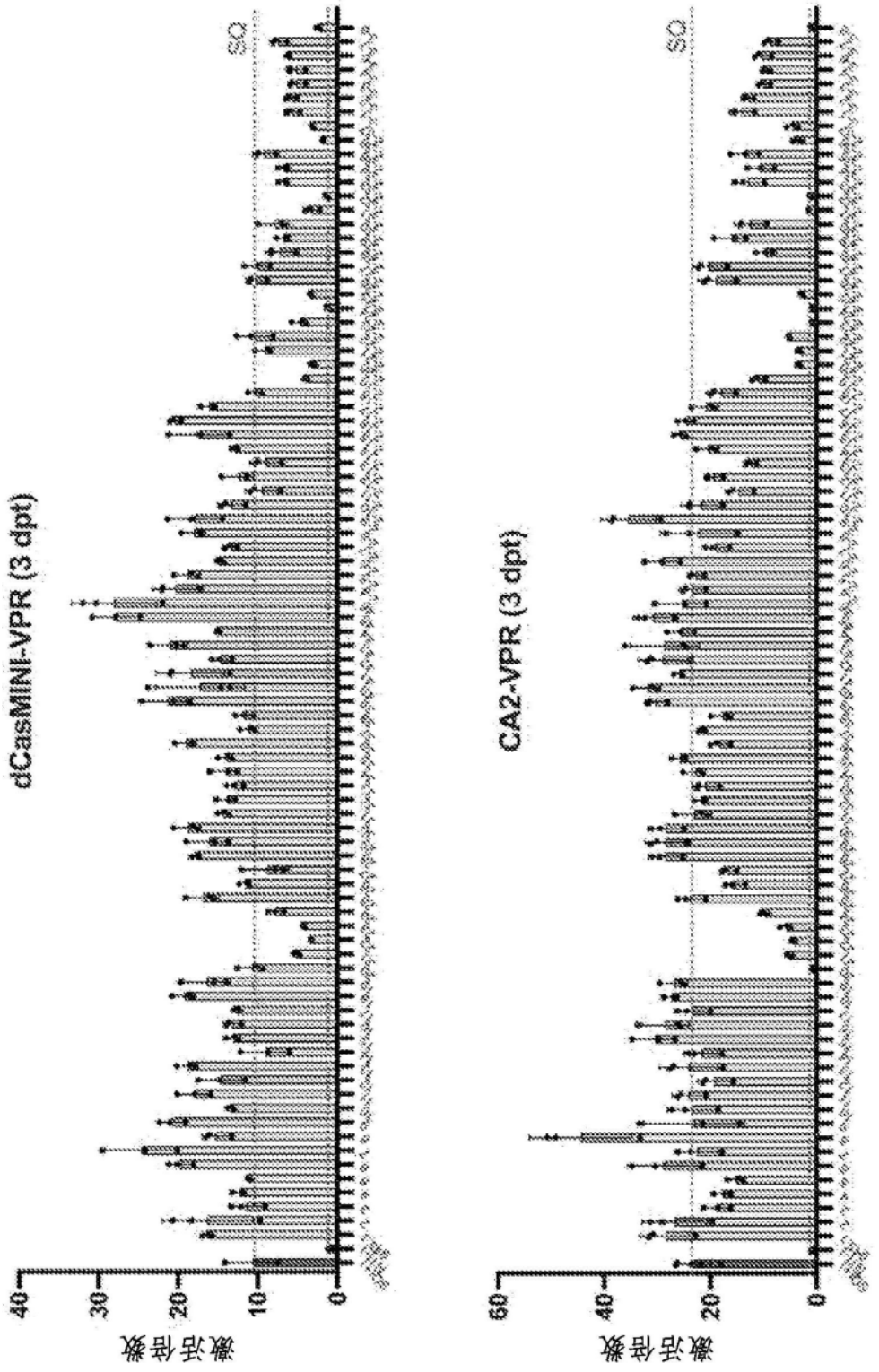

dCasMINI-VPR (3 dpt)
激活倍数
0
10
20
30
40
SO
CA2-VPR (3 dpt)
激活倍数
0
20
40
60
SO

图12

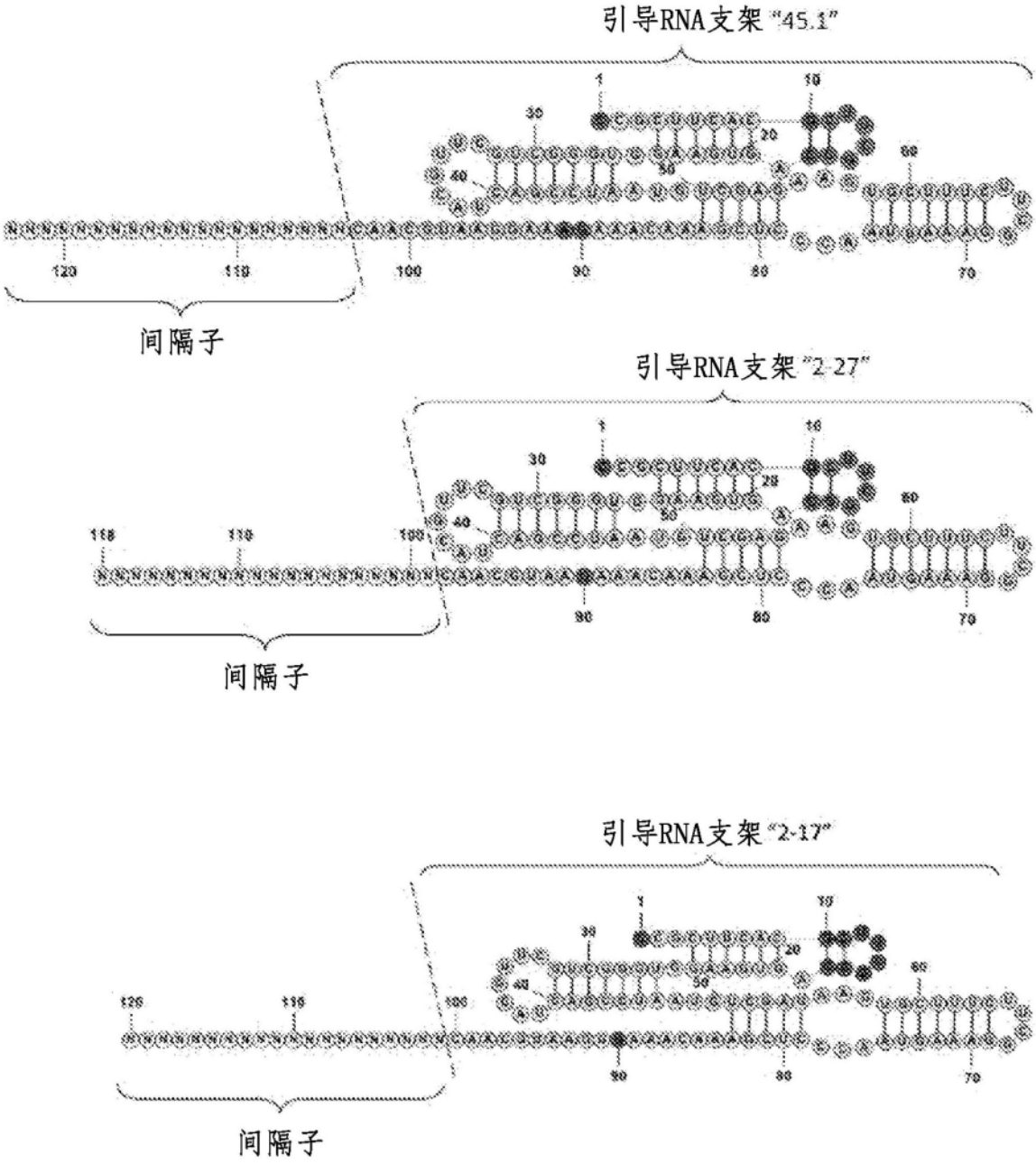


图13

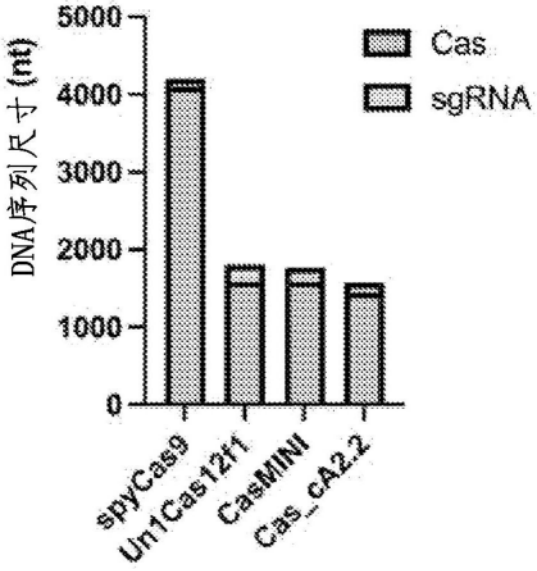

DNA序列尺寸 (nt)
5000
4000
3000
2000
1000
0
Cas
sgRNA
spyCas9
Un1Cas12f1
CasMINI
Cas_cA2.2

图14

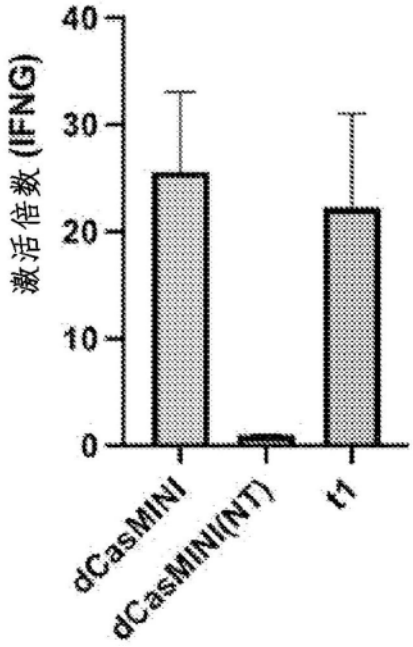

激活倍数 (IFNG)
40
30
20
10
0
dCasMINI
dCasMINI(NT)
t1

图15

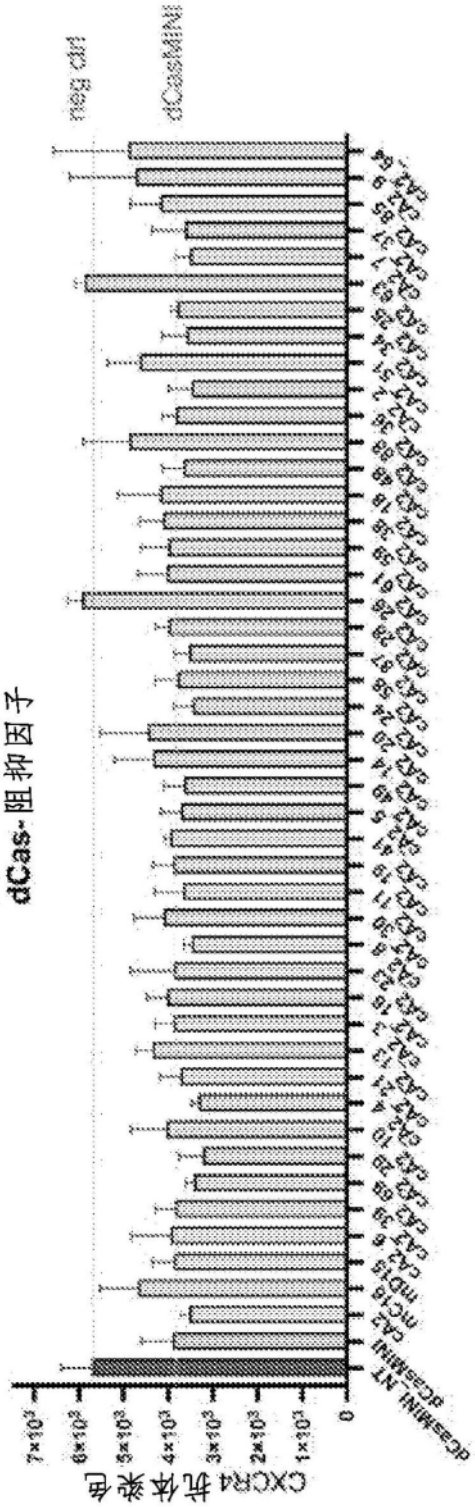

neg ctrl
dCasMINI
dCas-阻抑因子
CXCR4 抗体染色
7×10³
6×10³
5×10³
4×10³
3×10³
2×10³
1×10³
0
dCasMINI_NT
dCasMINI

图16