



1. 一种在无RNA介导的情况下采用Cas蛋白从DNA溶液中分离DNA的方法,其中所述方法包括步骤a) DNA溶液与Cas蛋白体系混合;c) 从Cas蛋白体系与DNA结合的复合物中分离得到DNA,其中所述的Cas蛋白体系为:

- 1) 偶联了平板状、膜状、柱状的固体支持物的Cas蛋白;或
- 2) 偶联了珠形的固体支持物的Cas蛋白。

2. 权利要求1所述的方法,包括步骤a) DNA溶液与Cas蛋白体系混合;b) 富集步骤a) 中的Cas蛋白体系与DNA结合的复合物;c) 从步骤b) 富集的复合物中分离得到DNA,其中所述的Cas蛋白体系为偶联了珠形的固体支持物的Cas蛋白。

3. 权利要求2所述的方法,其中所述的Cas蛋白体系为Cas蛋白与至少一种连接序列形成的Cas融合蛋白。

4. 权利要求3所述的方法,其中在步骤a) 中加入偶联了亲和分子的固体支持物,孵育后,再通过步骤b) 富集。

5. 权利要求3所述的方法,其中所述步骤b) 包括离心富集或磁性富集。

6. 权利要求1-5任一项所述的方法,其中所述的DNA是双链DNA。

7. 权利要求1所述的方法,其中所述的Cas蛋白体系为Cas蛋白通过化学共价键偶联到固体支持物而形成的。

8. 权利要求3所述的方法,其中所述的Cas蛋白体系为Cas融合蛋白与固体支持物上的亲和分子结合而形成的。

9. 权利要求1、4、5、7或8所述的方法,其中所述的固体支持物是能与氨基酸残基形成化学键的固体物质。

10. 权利要求7所述的方法,其中所述的固体支持物选自凝胶材料、磁性材料、纤维素材料、硅胶材料、玻璃材料和人造高分子聚合物制得的固体物质中的至少一种。

11. 权利要求1所述的方法,其中所述的珠形的固体支持物为磁珠、凝胶珠和/或硅胶珠。

12. 权利要求3所述的方法,其中所述的连接序列和其亲和分子选自下组配体和受体组合中的一种:酶与底物、抗原与抗体、生物素与亲和素。

13. 权利要求12所述的方法,其中所述的酶与底物选自谷胱甘肽转移酶与谷胱甘肽。

14. 权利要求12所述的方法,其中所述的抗原-抗体选自下组中的一种:A蛋白或其保留了结合Fc区域功能性的片段与免疫球蛋白或其Fc或Fab片段、G蛋白或其保留了结合Fc区域功能性的片段与免疫球蛋白或其Fc或Fab片段、组氨酸标签与抗组氨酸标签、多组氨酸标签与抗多组氨酸标签和FLAG标签与抗FLAG标签。

15. 权利要求12所述的方法,其中所述的生物素与亲和素选自下组中的一种:生物素与亲和素或链酶亲和素、连接了生物素的生物素受体蛋白与亲和素或链酶亲和素、Strep-tag与亲和素或链酶亲和素。

16. 权利要求1或2所述的方法,其中所述的步骤a) 的反应时间为5分钟到6小时。

17. 权利要求16所述的方法,其中所述的步骤a) 的反应时间为15分钟到1小时。

18. 权利要求1或2所述的方法,其中所述的步骤a) 的反应温度为20-37°C。

19. 权利要求18所述的方法,其中所述的步骤a) 的反应温度为30-37°C。

20. 权利要求4所述的方法,其中所述的步骤b), 在溶液中加入偶联了亲和分子的固体

支持物的反应时间为30分钟至过夜。

21. 权利要求20所述的方法,其中所述的步骤b),在溶液中加入偶联了亲和分子的固体支持物的反应时间为1-2小时至过夜。

22. 权利要求20所述的方法,其中所述的步骤b)的反应温度为4℃到37℃。

23. 权利要求22所述的方法,其中所述的步骤b)的反应温度为20-30℃。

24. 权利要求2、4或5所述的方法,其中所述b)的富集为离心富集或磁性富集。

25. 权利要求1、2、4或5所述的方法,其中所述步骤c)通过洗脱分离DNA。

26. 权利要求1所述的方法,其中所述的Cas是具有DNA结合功能的Cas蛋白。

27. 权利要求26所述的方法,其中所述的Cas蛋白是天然的Cas蛋白、突变后失去了gRNA结合能力的Cas蛋白,或者突变后核酸酶失活的Cas蛋白。

28. 权利要求26所述的方法,其中Cas蛋白选自Cas1、Cas1B、Cas2、Cas3、Cas4、Cas5、Cas6、Cas7、Cas8、Cas9、Cas10、Cas12、Cas13和Cas14。

29. 权利要求28所述的方法,其中所述的Cas蛋白是Cas9蛋白。

30. 权利要求29所述的方法,其中所述的Cas9蛋白是dCas9蛋白。

## 一种Cas蛋白体系分离DNA的方法

### 技术领域

[0001] 本发明涉及一种分离DNA的方法,尤其涉及一种采用Cas蛋白从溶液中分离DNA的方法。

### 背景技术

[0002] 通常情况下,绝大多数脱氧核糖核酸(DNA)位于细胞中,但不管是在微生物、植物还是动物的细胞外,也常常存在着DNA的片段,这类核酸被统称为游离DNA。Mandel和Metasis在1948年首次报道了在外周血中发现游离DNA(Mandel P.C R Hebd Seances Acad Sci(Paris),1948,142(3):241-3)。经过半个世纪的研究,科学家发现不论是健康人群还是患病人群,都有小部分DNA游离于细胞外。这类DNA通常存在于血液、血浆、血清和其他体液中。在各个样品中发现的细胞外DNA由于被核酸酶所保护(例如它们以蛋白脂质复合物的形式被分泌出来,常常与蛋白质结合,或被包含在囊泡中),因此并不容易降解。在患有疾病,如恶性肿瘤和传染病的人群中,这类细胞外DNA的含量常常比正常人要高很多,因此在疾病筛查、诊断、预后、病程进展监测、甚至是治疗靶标的识别等方面都有很大的应用前景。此外,母体血液中常伴随有相当含量的胎儿DNA,这类DNA可以用于多种检测,例如性别鉴定,染色体异常的测试和评估,同时还能监测妊娠相关的并发症。也正是由于其较强的临床诊断相关性,游离DNA在无创诊断和预后方面有着巨大的应用潜力,可用于例如无创产前基因检测、肿瘤检测、移植医学等许多其他疾病。

[0003] 目前,许多从各种生物液体中提取并分析单链或双链的游离DNA的方法已经被报道,其中包括阴离子交换法(WO2016198571 A1),二氧化硅珠提取法(WO2015120445A1),曲拉通/加热/苯酚法(Xue et al.Clinica Chimica Acta 2009.404:100-104)等。这些方法通常会使用分解液剂将游离DNA从蛋白质的包裹中释放出来,然后通过物理吸附或析出的方法将游离DNA分离出来。然而,使用这些方法提取DNA的产率非常低,很难从大样本中提取少量DNA。例如,目前市售的用于游离核酸提取的QIAamp®血清/血浆核酸纯化试剂盒(Circulating Nucleic Acid kit,QIAGEN(凯杰)公司),对游离DNA的提取效率只有不到50%。

[0004] 成簇规律间隔短回文重复序列(CRISPR)和CRISPR结合蛋白(Cas蛋白)组成了一套强大的核酸酶系统,能够对真核细胞中几乎所有与原型间隔子邻近基序(proto-spacer-adjacent motif,PAM)相邻的基因组序列进行切割(Cong et al.Science 2013.339:819-823)。到目前为止,所有和CRISPR/Cas系统相关的应用除了需要Cas蛋白之外,还包含了RNA组分。所述RNA组分是由一条CRISPR RNA(crRNA)和一条反式激活crRNA(tracrRNA)组成的双链指导RNA结构,指导Cas蛋白切割与crRNA序列互补的DNA位点(Jinek et al.Science 2012.337:816-821)。为了进一步简化CRISPR/Cas9系统,研究人员利用工程学方法将双链指导RNA的两个组分(crRNA和tracrRNA)的部分连接成嵌合型单链指导RNA(sgRNA)。这体现了CRISPR/Cas9系统强大的基因识别和基因编辑能力,但是,Cas蛋白本身的性质并没有被深入的研究。到目前为止,还没有研究人员发现单独使用作为核酸酶的Cas蛋白及其变体在

识别并结合DNA上的潜力。

[0005] 目前,市场上存在对高效提取游离DNA方法的强烈需求,以便从各种生物液体的大样本中高效提取游离DNA,用于后续分析检测。

[0006] 本发明的简述

[0007] 为了提高对溶液中DNA的提取效率,本发明提供了如下技术方案:

[0008] (1)一种采用Cas蛋白从溶液中分离DNA的方法。

[0009] (2)根据技术方案(1)所述的方法,包括步骤a) DNA溶液与Cas蛋白体系混合;c)从Cas蛋白体系与DNA结合的复合物中分离得到DNA

[0010] (3)根据技术方案(1)所述的方法,包括步骤a)在DNA的溶液中加入Cas蛋白体系;b)富集步骤a)中的Cas蛋白体系与DNA结合的复合物;c)从步骤b)富集的复合物中分离得到DNA。

[0011] (4)根据技术方案(2)所述的方法,其中所述的Cas蛋白体系为偶联了平板状、膜状或柱状的固体支持物的Cas蛋白。

[0012] (5)根据技术方案(3)所述的方法,其中所述的Cas蛋白体系为偶联了珠形固体支持物的Cas蛋白。

[0013] (6)根据技术方案(3)所述的方法,其中所述的Cas蛋白体系为Cas蛋白与至少一种连接序列形成的Cas融合蛋白。

[0014] (7)根据技术方案(6)所述的方法,其中步骤b)在溶液中加入偶联了亲和分子的固体支持物,再通过离心富集或磁性富集。

[0015] (8)根据技术方案(6)所述的方法,其中在步骤a)中加入偶联了亲和分子的固体支持物,孵育后,再通过步骤b)离心富集或磁性富集。

[0016] (9)根据技术方案(1)-(8)任意一项所述的方法,其中所述的DNA是双链DNA。

[0017] (10)根据技术方案(4)或(5)所述的方法,其中所述的Cas蛋白体系为Cas蛋白通过化学共价键偶联到固体支持物。

[0018] (11)根据技术方案(6)所述的方法,其中所述的Cas蛋白体系为Cas融合蛋白与固体支持物上的亲和分子结合而形成的。

[0019] (12)根据技术方案(4)-(11)任一项所述的方法,其中所述的固体支持物是能与氨基酸残基形成化学键的固体物质。

[0020] (13)根据技术方案(10)所述的方法,其中所述的固体支持物选自至少一种凝胶材料、磁性材料、纤维素材料、硅胶材料、玻璃材料和人造高分子聚合物制得的固体物质。

[0021] (14)根据技术方案(5)-(13)任一项所述的方法,其中所述的固体支持物为磁珠、凝胶珠和/或硅胶珠。

[0022] (15)根据技术方案(6)-(14)任一项所述的方法,其中所述的连接序列和亲和分子选自下组配体和受体组合中的一种,酶与底物、抗原与抗体、生物素与亲和素。

[0023] (16)根据技术方案(15)所述的方法,其中所述的酶与底物选自谷胱甘肽转移酶与谷胱甘肽。

[0024] (17)根据技术方案(15)所述的方法,其中所述的抗原-抗体选自下组中的一种:A蛋白或其保留了结合Fc区域功能性的片段与免疫球蛋白或其Fc或Fab片段、G蛋白或其保留了结合Fc区域功能性的片段与免疫球蛋白或其Fc或Fab片段、组氨酸标签与抗组氨酸标签、

多组氨酸标签与抗多组氨酸标签和FLAG标签与抗FLAG标签。

[0025] (18) 根据技术方案(15)所述的方法,其中所述的生物素与亲和素选自下组中的一种:生物素与亲和素或链酶亲和素、连接了生物素的生物素受体蛋白与亲和素或链酶亲和素、Strep-tag与亲和素或链酶亲和素。

[0026] (19) 根据前述技术方案(2) - (18)任意一项所述的方法,其中所述的步骤a)的反应时间为5分钟到6小时,优选反应时间为15分钟到1小时。

[0027] (20) 根据前述技术方案(2) - (18)任意一项所述的方法,其中所述的步骤a)的反应温度为20-37℃,优选反应温度为30-37℃。

[0028] (21) 根据前述技术方案(7)、(12) - (20)任意一项所述的方法,其中所述的步骤b),在溶液中加入偶联了亲和分子的固体支持物的反应时间为30分钟至过夜,优选反应时间为1-2小时。

[0029] (22) 根据技术方案(21)所述的方法,其中所述的步骤b)的反应温度为4℃到37℃,优选反应温度为20-30℃。

[0030] (23) 根据前述技术方案(3) - (22)任意一项所述的方法,其中所述b)的富集为离心富集或磁性富集。

[0031] (24) 根据前述技术方案(2) - (23)任意一项所述的方法,其中所述步骤c)通过洗脱分离DNA。

[0032] (25) 根据前述技术方案(2) - (24)任意一项所述的方法,其中所述的Cas是具有DNA结合功能的Cas蛋白。

[0033] (26) 根据技术方案(25)所述的方法,其中所述的Cas蛋白是天然的Cas蛋白、突变后失去了gRNA结合能力的Cas蛋白,或者突变后核酸酶失活的Cas蛋白。

[0034] (27) 根据技术方案(25)或(26)所述的方法,其中所述的Cas蛋白是Cas9蛋白。

[0035] (28) 根据技术方案(27)所述的方法,其中所述的Cas9蛋白是dCas9蛋白。

[0036] (29) 一种从溶液中分离DNA的试剂盒,包含组分a) Cas蛋白体系,和组分b) DNA洗脱液。

[0037] (30) 根据技术方案(28)所述的试剂盒,其中组分a) Cas蛋白体系是偶联了固体支持物的Cas蛋白溶液体系,或是含有Cas融合蛋白和偶联了亲和分子的固体支持物的溶液体系。

[0038] (31) 根据技术方案(29)所述的试剂盒,其中组分a) Cas蛋白体系由两个独立的体系组成,i) Cas蛋白与至少一种连接序列形成的Cas融合蛋白,和ii) 偶联了亲和分子的固体支持物。

[0039] (32) 根据技术方案(29)所述的试剂盒,其中组分b) DNA洗脱液可以是以下液体:0.5-10mg/ml蛋白酶K溶液、1-5M浓度的NaCl或者KCl盐溶液、具有0-0.3M Tris和0-0.5M NaCl的pH 11-12.5的碱洗脱液。

[0040] (33) 根据技术方案(30)或(31)所述的试剂盒,其中的固体支持物为平板状、膜状或柱状的固体支持物。

[0041] (34) 根据技术方案(33)所述的试剂盒,其中的固体支持物选自凝胶材料、磁性材料、纤维素材料、硅胶材料、玻璃材料和人造高分子聚合物制得的固体物质中的至少一种。

[0042] (35) 根据技术方案(34)所述的试剂盒,其中的固体支持物为磁珠、凝胶珠和/或硅

胶珠。

[0043] (36) 根据技术方案 (31) - (35) 所述的试剂盒,其中所述的连接序列和亲和分子选自下组配体和受体组合中的一种:酶与底物、抗原与抗体、生物素与亲和素。

[0044] (37) 根据技术方案 (36) 所述的试剂盒,其中所述的与底物选自谷胱甘肽转移酶与谷胱甘肽。

[0045] (38) 根据技术方案 (37) 所述的试剂盒,其中所述的抗原-抗体选自下组中的一种:A蛋白或其保留了结合Fc区域功能性的片段与免疫球蛋白或其Fc或Fab片段、G蛋白或其保留了结合Fc区域功能性的片段与免疫球蛋白或其Fc或Fab片段、组氨酸标签与抗组氨酸标签、多组氨酸标签与抗多组氨酸标签和FLAG标签与抗FLAG标签。

[0046] (39) 根据技术方案 (36) 所述的试剂盒,其中所述的生物素与亲和素选自下组中的一种:生物素与亲和素或链酶亲和素、连接了生物素的生物素受体蛋白与亲和素或链酶亲和素、Strep-tag与亲和素或链酶亲和素。

[0047] (40) 根据技术方案 (29) - (39) 所述的试剂盒,其中所述的Cas是具有DNA结合功能的Cas蛋白。

[0048] (41) 根据技术方案 (40) 所述的试剂盒,其中所述的Cas蛋白是天然的Cas蛋白、突变后失去了gRNA结合能力的Cas蛋白,或者突变后核酸酶失活的Cas蛋白。

[0049] (42) 根据技术方案 (41) 所述的试剂盒,其中所述的Cas蛋白是Cas9蛋白。

[0050] (43) 根据技术方案 (42) 所述的试剂盒,其中所述的Cas9蛋白是dCas9蛋白(突变后核酸酶失活的Cas9蛋白)。

[0051] 在同等含量的DNA样本条件下,本发明中的提取方法省去了繁琐的样本处理过程,减少了样本的使用量要求,同时大幅提高了提取游离DNA的效率。

## 附图说明

[0052] 图1:包含his-S.P.dCas9-FLAG基因片段的表达载体pSMART-his-S.P.dCas9-FLAG

[0053] 图2:486bp的游离DNA(SEQ ID NO.8)与dCas蛋白的结合电泳条带。

[0054] 泳道1是游离DNA自身的条带

[0055] 泳道2是dCas9和游离DNA(SEQ ID NO.8)结合之后的情况

[0056] 本发明的详述

[0057] Cas蛋白

[0058] 成簇规律间隔短回文重复序列(CRISPR)和CRISPR结合蛋白(Cas蛋白)组成了一套强大的核酸酶系统。其中CAS蛋白可分为四个不同的功能模块:目标识别模块(间隔获取);表达模块(crRNA加工和靶标结合);干扰模块(靶标切割);和辅助模块(监管和其他CRISPR相关功能)。近年来,核心Cas蛋白(Cas1-Cas10)已经积累了大量的结构和功能信息,这使得它们可以被分类到这些模块中。CRISPR相关内切酶Cas蛋白可以通过单链指导RNA(sgRNA)靶向特定的基因组序列。以化脓链球菌(*Streptococcus pyogenes*)Cas9蛋白(spCas9)蛋白为例,该蛋白包含有目标识别模块和核酸酶模块(干扰模块)组成的双叶结构,其在界面处的带正电的凹槽中容纳sgRNA-DNA异源双链体。而目标识别区对于结合sgRNA和DNA是必需的,核酸酶区包含HNH和RuvC核酸酶结构域,它们分别适合于切割靶DNA的互补链和非互补链。核酸酶区还含有负责与原型间隔子相邻基序(PAM)相互作用的羧基末端结构域。但是,

到目前为止,还没有任何研究组发表关于Cas蛋白直接结合DNA的结构区域,或者Cas-DNA复合物的晶体结构。

[0059] 因此在本发明中,在没有RNA介导的情况下,Cas蛋白本身就是一种核酸结合系统。通常情况下,Cas蛋白在没有RNA介导时便能够识别并结合几乎所有的DNA序列,同时并不对所结合的DNA序列进行剪切。

[0060] 本发明所述的Cas蛋白是具有DNA结合功能的Cas蛋白,例如天然的Cas蛋白,突变的Cas蛋白,突变后失去了gRNA结合能力的Cas蛋白,突变后核酸酶失活的Cas蛋白(dead Cas,dCas)。

[0061] Cas蛋白的非限制性实例包括:Cas1、Cas1B、Cas2、Cas3、Cas4、Cas5、Cas6、Cas7、Cas8、Cas9(也称为Csn1或Csx12)、Cas10、Cas12、Cas13、Cas14、Csy1、Csy2、Csy3、Cse1、Cse2、Csc1、Csc2、Csa5、Csn2、Csm2、Csm3、Csm4、Csm5、Csm6、Cmr1、Cmr3、Cmr4、Cmr5、Cmr6、Csb1、Csb2、Csb3、Csx17、Csx14、Csx10、Csx16、CsaX、Csx3、Csx1、Csx15、Csf1、Csf2、Csf3、Csf4其在不同物种中的同源蛋白、核酸内切酶失活的突变蛋白、或其修饰形式。

[0062] 在一个具体实施方式中,本发明的Cas蛋白是Cas9蛋白,更优选突变后核酸酶失活的Cas9蛋白(dead Cas9,dCas9)

[0063] Cas9,也称为Csn1或Csx12,是既参与crRNA生物合成又参与摧毁入侵DNA的巨型蛋白质。已经在不同的细菌物种如嗜热链球菌(*S. thermophiles*)、无害利斯特氏菌(*Listeria innocua*) (Gasiunas,Barrangou et al.2012;Jinek,Chylinski et al.2012)和化脓性链球菌(*S.Pyogenes*) (Deltcheva,Chylinski et al.2011)中描述了Cas9。本发明的Cas9包括但不限于以下类型:化脓链球菌(*Streptococcus pyogenes*)Cas9蛋白,其氨基酸序列见SwissProt数据库登录号Q99ZW2;脑膜炎奈瑟氏菌(*Neisseria meningitides*)Cas9蛋白,其氨基酸序列见UniProt数据库编号A1IQ68;嗜热链球菌(*Streptococcus thermophilus*)Cas9蛋白,其氨基酸序列见UniProt数据库编号Q03LF7;金黄色葡萄球菌(*Staphylococcus aureus*)Cas9蛋白,其氨基酸序列见UniProt数据库编号J7RUA5。

[0064] 作用于DNA双链的Cas蛋白一般具有两个内切酶活性位点,如果有且只有一个位点发生突变或缺失造成该位点酶活失效,则可以得到切割DNA单链的Nickase,例如Cas9-nickase,因其高保真率以及切割DNA单链的特性而被广泛应用。

[0065] 另外通过突变筛选出Cas蛋白变体存在一个或多个点突变的情况被证明可以提高其特异性,降低基因组编辑的脱靶率,或使兼容更多样的PAM序列,例如eSpCas9,SpCas9-HF,HeFSpCas9和HIFI-SpCas9,xCas9(Christopher A.V et al.Nature Medicine 2018 (24),pp1216-1224;Johnny H.H et al.Nature 2018 556(7699),pp57-63)等。

[0066] 鉴于Cas蛋白包含多个功能模块,其蛋白序列较长,为了利于包装转运以及功能的控制,其蛋白的序列也会通过蛋白质工程研究被截短从而形成Cas蛋白系统,例如Split-SpCas9系统包含截短蛋白形成的复合体(Jana M et al.Plant Biotechnology Journal 2017 15,pp.917-926)。

[0067] 在一些实施例中,该Cas蛋白是含有上述蛋白或其突变体的融合蛋白或蛋白复合物,包括但不限于在Cas上连接其他氨基酸序列的变体。在以上Cas蛋白或其变体基础上融合其他的功能性蛋白,可以增加Cas蛋白功能的特异性和有效性,也可以对基因组产生切割之外的效应。例如在Cas或dCas上融合FokI,这样可以增加Cas9蛋白对基因组切割的特异

性,因为FokI只有二聚化才会有切割活性,这样就要求有一对识别区域才能完成切割进而减少了脱靶率。又例如在截去部分功能结构域但保存其DNA结合能力的Cas蛋白上连接FokI (Ma et al. ACS Synth. Biol., 2018, 7(4), pp 978-985)。又例如在Cas-nickase或dCas上融合碱基的修饰酶(如脱氨酶,胞嘧啶脱氨酶,腺嘌呤脱氨酶),这样可以高效的对基因组靶点区域进行定向的碱基修改(Komor et al. Sci. Adv. 2017. 3: eaao4774)。又例如对dCas9融合一些可以调节基因表达的蛋白结构域,可以有效对靶点基因进行表达调节,例如融合VP64, VPR等转录激活因子的dCas9可以在gRNA的引导下结合到靶向基因附近而激活表达;相反的,如果融合了转录抑制因子(如SRDX)的dCas9会对靶基因产生下调作用。dSpCas9-Tet1和dSpCas9-Dnmt3a可以用于修改表观遗传状态,调节内源基因启动子的甲基化状态来调节蛋白表达。

#### [0068] 具有DNA结合功能的Cas蛋白

[0069] 具有DNA结合功能的Cas蛋白可以是天然的Cas蛋白,或者在Cas蛋白的DNA结合功能区以外做的突变,或者在DNA结合功能区进行突变后获得的Cas蛋白。

[0070] 检测Cas蛋白是否具有DNA结合功能,可以采用凝胶电泳法对比结合蛋白的DNA和未结合蛋白的DNA的胶图条带位置来判别(检测方法参见实施例1)。

#### [0071] 核酸酶失活的突变Cas蛋白(dCas蛋白)

[0072] 核酸酶失活的突变Cas蛋白(dCas蛋白)是Cas蛋白经突变而获得的变体,其核酸内切酶活性失活或基本失活,使Cas蛋白失去或者基本失去核酸内切酶活性,进而无法切割靶序列。前述列举的Cas蛋白非限制性实例都可通过核酸内切酶失活突变改造为dCas蛋白,所述突变包括一个或多个氨基酸残基的插入、缺失或取代等。

[0073] 例如,某些Cas9突变可以使Cas9蛋白失去或者基本失去核酸内切酶活性,进而无法切割靶序列。对于某一物种的Cas9,例如spCas9而言,降低或消除核酸内切酶活性的示例性突变包括以下位置中的一个或多个突变:D10, G12, G17, E762, H840, N854, N863, H982, H983, A984, D986, 或A987。文献证明导向RNA(gRNA)介导的核酸内切酶失活突变Cas9(称为dCas9)可导致大肠杆菌特异性内源基因、以及人细胞中EGFP报告基因的表达抑制(Qi et al. Cell 2013. 152: 1173-1183)。这项研究证明了使用gRNA介导的dCas9技术可准确的识别并结合相应的基因组。

[0074] dCas蛋白可以是非天然存在于自然界中,并且是由蛋白质工程或通过随机诱变获得的变体,其核酸内切酶活性失活或基本失活。例如,可以通过突变,即化脓性链球菌Cas9核酸内切酶的氨基酸序列中至少一个残基的缺失或插入或取代获得相应dCas9蛋白。

[0075] 在一些实施例中,dCas9蛋白可以为不同物种的Cas9经核酸内切酶活性失活或基本失活突变而获得的变体,包括但不限于,化脓链球菌(*Streptococcus pyogenes*)Cas9蛋白、脑膜炎奈瑟氏菌(*Neisseria meningitidis*)Cas9蛋白、嗜热链球菌(*Streptococcus thermophilus*)Cas9蛋白、金黄色葡萄球菌(*Staphylococcus aureus*)Cas9蛋白、肺炎链球菌Cas9蛋白等。

#### [0076] 连接序列和亲和分子

[0077] 本发明所述的连接序列是指直接或间接连接于Cas蛋白的C端或者N端的分子,其最终与Cas蛋白形成融合蛋白。

[0078] 本发明所述的亲和分子是偶联在固相物质上的,能与前述Cas蛋白上的连接序列

形成特异性结合的分子。

[0079] 本发明的连接序列和亲和分子互为受体和配体关系。任何蛋白质连接序列都可以应用在本发明的实践中,只要它可与亲和分子特异性结合。连接序列和亲和分子选自下列配体受体组合中的一种:酶与底物、抗原-抗体、生物素与亲和素。其中所述的酶与底物选自下组中的一种:谷胱甘肽转移酶与谷胱甘肽;所述的抗原-抗体选自下组中的一种:A蛋白或其保留了结合Fc区域功能性的片段与免疫球蛋白或其Fc或Fab片段、G蛋白或其保留了结合Fc区域功能性的片段与免疫球蛋白或其Fc或Fab片段、组氨酸标签与抗组氨酸标签、多组氨酸标签与抗多组氨酸标签、FLAG标签(氨基酸序列:ATLAAAAL)与抗FLAG标签;所述的生物素与亲和素选自下组中的一种:生物素(biotin)与亲和素(Avidin,UniProt数据库编号P02701)或链酶亲和素(Streptavidin,UniProt数据库编号P22629)、连接了生物素的生物素受体蛋白(如AviTag,氨基酸序列:GLNDIFEAQKIEWHE)与亲和素或链酶亲和素、Strep-tag(氨基酸序列:WSHPQFEK)与亲和素或链酶亲和素。

[0080] 在本发明中,连接序列和亲和分子通过亲和作用特异性结合,并且可以互换。比如,如果连接序列是受体,则亲和分子为与该受体特异结合的配体;连接序列也可是所述配体,那么,亲和分子为与该配体特异结合的受体。又比如,如果连接序列为链酶亲和素,则亲和分子为生物素;连接序列也可以为生物素,那么亲和分子为链酶亲和素。

[0081] 固相支持物

[0082] 本发明所述的固相物质是能与氨基酸残基形成化学键的固体物质,因而可以与亲和分子或者Cas蛋白通过化学键偶联。

[0083] 合适的固相支持物从材料区分可以是凝胶材料、磁性材料、纤维素材料、硅胶材料、玻璃材料,或人造高分子聚合物制得的固体物质。凝胶材料可以是水凝胶、有机凝胶、干凝胶、或纳米复合水凝胶;磁性材料可以由高分子材料包裹的金属氧化物(如铁钴镍的氧化物);纤维素材料可以是纤维素、醋酸纤维素、硝酸纤维素;硅胶材料可以是二氧化硅支持物或者有机硅胶支持物;人造高分子聚合物可以是、尼龙、聚酯、聚醚砜、聚烯烃、聚偏1,1-二氟乙烯和它们的组合。固相支持物从形式上区分可以是珠形、化学膜类柱、玻璃平板、6孔/24孔/48孔/96孔/384孔板,PCR反应管等。

[0084] 结合了DNA的Cas蛋白与固体支持物,例如化学膜类柱、玻璃平板、6孔/24孔/48孔/96孔/384孔板、PCR反应管结合后,通过洗脱液将DNA洗脱下来。

[0085] 结合了DNA的Cas蛋白与珠形的固体支持物结合后,经过富集,例如离心富集或磁性富集后,通过洗脱液将DNA洗脱下来。

[0086] 本发明所述的固体支持物优选是珠形的,例如磁珠、凝胶珠和/或硅胶珠。可以使用的凝胶珠非限制性实例包括:Anti-6X Histag®antibody(琼脂糖凝胶)(Abcam),Anti-6xHisAlphaLISA®Acceptor beads(PerkinElmer),Dynabeads™ His-Tag Isolation and Pulldown(ThermoFisher),Anti-His-tag mAb-Magnetic Beads(MBL Life science),His Tag Antibody Plate(GenScript)。

[0087] 偶联

[0088] 可以采用常规的方法将亲和分子或Cas蛋白偶联到固体支持物上。将A蛋白和G蛋白连接到固体支持物上,可以参见,例如Hermanson et al.1992,Immobilized Affinity Ligand Techniques,Academic Press。典型地,固体支持物是用反应官能团(“活化基团”)

如环氧化物(表氯醇)、氰(溴化氰CNBr),N,N-二琥珀酰亚胺基碳酸酯(DSC)、醛或活化的羧酸(如N-羟基丁二酰亚胺(NHS)酯、羰基二咪唑(CDI)活化酯)来活化的。这些活化基团能直接连接到固体支持物上,如CNBr,或者它们也可以是“接头”或间隔分子的一部分,典型地是碳、氧和氮原子的直链,如在接头丁二醇二缩水甘油醚(一种常用的环氧化物偶联剂)中发现的碳和氧的十元链。在偶联条件下再将活化的固体支持物与Cas蛋白质或亲和分子相平衡。偶联反应完成,彻底洗涤介质。

#### [0089] Cas融合蛋白

[0090] 本发明的Cas融合蛋白是将连接序列添加到Cas蛋白的N端或C端或其蛋白序列中间任意部分而形成的融合蛋白。本发明的融合Cas蛋白可以通过基因工程的方法,使Cas蛋白和连接序列一起表达。

#### [0091] Cas蛋白体系

[0092] 本发明所述的Cas蛋白体系是一种包含Cas蛋白的复合物,用于结合溶液中的DNA。

[0093] 在一个具体的实施方式中,Cas蛋白体系是Cas蛋白通过化学共价键偶联到固体支持物上而形成的复合物。

[0094] 在一个具体的实施方式中,Cas蛋白体系是Cas蛋白与一个或多个连接序列形成的Cas融合蛋白。

[0095] 在另一个具体的实施方式中,Cas蛋白体系是Cas融合蛋白与固相物质上偶联的亲合分子结合后形成的复合物。

#### [0096] Cas融合蛋白和偶联亲和分子的固相物质的结合

[0097] Cas蛋白与连接序列一同组成Cas融合蛋白。连接序列可直接或间接位于Cas蛋白的N端或C端。以固相物质为凝胶珠为例,在一种优选的实施方式中,使用偶联亲和分子的凝胶珠结合Cas融合蛋白时,对于每10 $\mu$ l的凝胶珠,在加入凝胶珠之前,反应液中Cas融合蛋白的初始浓度为0.001-20 $\mu$ g/ $\mu$ l,优选为0.01-10 $\mu$ g/ $\mu$ l的Cas9融合蛋白。

[0098] Cas融合蛋白与固相物质的结合在4 $^{\circ}$ C到37 $^{\circ}$ C进行,优选4 $^{\circ}$ C或室温(20-25 $^{\circ}$ C)。采用的结合反应液可以为去离子水,1倍磷酸盐缓冲生理盐水(Phosphate buffered saline: PBS),Tris缓冲液(50mM Tris-HCl,150mM NaCl,pH7.4),或者其他盐离子浓度低于150mM、不含油去垢剂和变性剂的pH为约中性的缓冲液。反应时间为30分钟到过夜,优选过夜。

[0099] Cas蛋白也可以不经连接蛋白而直接与偶联亲和分子的固相物质结合。

#### [0100] Cas蛋白和游离DNA的结合

[0101] 本发明的Cas蛋白可以结合任何游离状态的DNA。该DNA可以是双链或者单链DNA。与Cas蛋白结合的游离DNA可通过固相物质结合Cas蛋白的方式将其分离出液相。

[0102] 与本发明的Cas蛋白结合的DNA可以为任何游离DNA。例如,以基因组为模板经PCR得到的任意序列DNA片段、序列为人工设计的DNA片段、细胞裂解后游离的DNA、细胞裂解后得到的基因组、人体或动物体液中游离的DNA。在一些实施方式中,DNA是以基因组为模板经PCR得到的DNA片段、序列为人工设计的DNA片段,或人体或动物体液中游离的DNA。

[0103] 作为优选的实施方式,Cas蛋白或Cas融合蛋白结合游离DNA时,DNA结合反应液中蛋白浓度为0.001-20 $\mu$ g/ $\mu$ l。优选0.01-10 $\mu$ g/ $\mu$ l的Cas蛋白或其变体或Cas融合蛋白。在一种优选的实施方式中,反应在室温(20-25 $^{\circ}$ C)到37 $^{\circ}$ C进行。采用的DNA结合反应液可以为去离子水,1倍磷酸盐缓冲生理盐水(Phosphate buffered saline:PBS),Tris缓冲液(50mM

Tris-HCl, 150mM NaCl, pH7.4), 或者其他盐离子浓度低于150mM、不含油去垢剂和变性剂的pH为中性的缓冲液。优选为50mM KCl, 10mM EDTA, 30mM Tris-HCl, 0.2% Triton X-100, 12% 甘油(glycerol)。反应时间为5分钟到6小时, 在一种实施方式中, 优选时间为5分钟到2小时。在另一种实施方式中, 优选时间为15分钟到1小时。

[0104] 将游离DNA从溶液中分离出来的方法

[0105] 本发明的Cas蛋白分离游离DNA的方法能有三种使用方式:

[0106] 方法一, Cas蛋白或Cas融合蛋白先在DNA结合反应液中与游离DNA结合, 之后, 再在蛋白结合反应液中通过偶联亲和分子的固相物结合Cas蛋白或Cas融合蛋白。首先, 在前面所述DNA结合反应条件下将Cas蛋白或Cas融合蛋白与游离DNA在DNA结合反应液孵育结合, 优选的时间为5分钟到6小时, 在另一种实施方式中, 优选时间为15分钟到1小时。反应温度为室温(20-25℃)到37度; DNA结合反应结束后, 直接在溶液中加入偶联亲和分子的固相物质, 如前面所述蛋白质结合反应条件下, 使Cas蛋白或Cas融合蛋白与偶联亲和分子的固相物质结合, 优选反应时间为30分钟到过夜, 反应温度为4℃到37℃; 最后通过离心分离固相物质, 弃上清; 在清洗固相物质数次之后, 将Cas蛋白或其变体或Cas融合蛋白以及结合的DNA从固相物质上洗脱下来, 得到洗脱溶液, 可以使用的洗脱液为含0.5-10mg/ml蛋白酶K溶液、1-5M浓度的NaCl或者KCl盐溶液、碱洗脱液(0-0.3M Tris, 0-0.5M NaCl, pH 11-12.5)、酸洗脱液(0-0.3M甘氨酸HCl, pH 2.5-3.5)、竞争洗脱液(DYKDDDK或者FLAG氨基酸的TBS溶液, 氨基酸浓度可以是100到500μg/ml), 或PAGE胶样本溶液(0.01M Tris-HCl, 10% Glycerol, 0.016% 溴酚蓝)。

[0107] 方法二, 偶联亲和分子的固相物质首先结合Cas蛋白或其变体或Cas融合蛋白, 再于DNA结合反应液中孵育结合游离DNA。使用的反应条件和溶液、洗脱条件和方法一一致。

[0108] 方法三, 将偶联亲和分子的固相物质、Cas9蛋白或其变体或Cas融合蛋白同时加入DNA结合反应液中, 共同孵育后分离固相物质。所述方法中的溶液可以是上述Cas蛋白或其变体或Cas融合蛋白结合游离DNA使用的溶液, 优选反应时间为30分钟到6小时, 反应温度为室温(20-25℃)到37度。洗脱条件和方法一一致。

## 实施例

[0109] 实施例仅为举例说明, 不旨在对本发明造成任何方式上的限制。

[0110] 实施例1: Cas蛋白的DNA结合活性

[0111] Cas蛋白的DNA结合活性可以采用凝胶电泳法, 对比结合Cas蛋白的DNA和未结合Cas蛋白的DNA的胶图条带位置来判别。采用如下反应条件:

[0112] 反应缓冲液(BA): 50mM三羟甲基氨基甲烷盐酸盐(Tris-HCl, pH 7.4); 150mM氯化钠(NaCl)

[0113] 反应体系(20μl):

[0114] dCas蛋白: 2μg

[0115] DNA: 0.25μg

[0116] 无核酸酶双蒸水: 定容到20μl

[0117] 反应过程:

[0118] 上述反应体系加入200μl PCR反应管中混合均匀后, 室温或37℃反应15分钟到1小

时。反应结束后,使用非变性胶,在BioRad电泳槽中,以80V电压,400mA电流运行80分钟。取出非变性胶后用DuRed(上海翊圣生物科技有限公司,10202ES76)将核酸染色,继而使用凝胶成像仪拍摄取得图片(见图2)。结合Cas蛋白的DNA的电泳条带明显滞后于未结合Cas蛋白的DNA的电泳条带。

[0119] 实施例2:游离DNA的制备

[0120] 本发明中的游离DNA可以是任意序列的DNA。在本实施例中,我们合成、构建了不同长度的11种DNA序列,如表1所示。

[0121] 表1:游离DNA序列(5' -3')

	SEQ ID NO:	长度 (bp)
[0122]	1	100
	2	100
	3	100
	4	200
	5	200
	6	200
[0123]	7	957
	8	486
	9	166
	10	609
	11	2655

[0124] SEQ ID NO.1-6号DNA序列通过正反向引物(表2)直接PCR而合成,并未使用任何模板。使用这六个DNA的主要目的是想要检测原型间隔子相邻基序(PAM)区域是否会对Cas蛋白与DNA的结合产生影响。其中SEQ ID NO.1-3为100bp,4-6为200bp,其中1和4号序列上没有Cas9相应的PAM区域(NGG),2和5号序列上各有一个PAM区域,而3和6号序列上各有3个PAM区域。详细结果会在实施例4中描述讨论。

[0125] SEQ ID NO.7-11号DNA序列则是通过分别设计相对应的正反向引物,以293T细胞(上海生科院,GNHu17)基因组作为模板PCR反应制成。

[0126] 表2:不同游离DNA序列所对应的上下游引物,以及各自PCR反应的退火温度。

	上游引物	下游引物	退火温度(°C)
SEQ ID NO:1	SEQ ID NO:12	SEQ ID NO:13	55
SEQ ID NO:2	SEQ ID NO:14	SEQ ID NO:15	55
SEQ ID NO:3	SEQ ID NO:16	SEQ ID NO:17	55
SEQ ID NO:4	SEQ ID NO:18	SEQ ID NO:19	58
SEQ ID NO:5	SEQ ID NO:20	SEQ ID NO:21	58
SEQ ID NO:6	SEQ ID NO:22	SEQ ID NO:23	58
SEQ ID NO:7	SEQ ID NO:24	SEQ ID NO:25	68
SEQ ID NO:8	SEQ ID NO:26	SEQ ID NO:27	60
SEQ ID NO:9	SEQ ID NO:28	SEQ ID NO:29	60
SEQ ID NO:10	SEQ ID NO:30	SEQ ID NO:31	59
SEQ ID NO:11	SEQ ID NO:32	SEQ ID NO:33	66

[0128] SEQ ID NO.1-6号DNA的PCR反应使用上述的2条上下游引物与Q5热启动超保真2X Master Mix试剂盒(New England Biolabs,M0494S)进行,每条引物的终浓度为0.5 $\mu$ M,加入10 $\mu$ l Q5 2X Master Mix,再加水至总体积为50 $\mu$ l,然后根据厂家的指南进行PCR。SEQ ID NO.7-11号DNA序列在上述反应体系的基础上,额外加入1 $\mu$ l细胞基因组提取液,最后定容至50 $\mu$ l。

[0129] 实施例3:构建并纯化偶联了FLAG的dCas9蛋白

[0130] 1蛋白表达

[0131] 1.1表达载体构建:pSMART-his-S.P.dCas9-FLAG。以下为载体环状示意图,his-S.P.dCas9-FLAG基因片段(SEQ ID No:34)通过EcoRI和BamHI限制性核酸内切酶整合到pSMART载体上(见图1);如果使用其他连接序列,则将载体中的FLAG(FLAG序列:GATTACAAGGATGACGATGACAAG)替换为相应的序列,如AVI-Tag,其他表达纯化蛋白的步骤相同:

[0132] 1.2表达菌株选择:Rosetta2(NOVAGEN,17400)

[0133] 1.3培养条件:摇瓶培养

[0134] 仪器耗材:摇床,LB培养基(生工,A507002)

[0135] 步骤:

[0136] 从4度冰箱里拿出平板(Rosetta/BL21)活化,37度过夜;转接:500ml的LB接了菌液4ml,37度恒温培养4小时,OD600=0.8,拿出放在室温,降摇床温度至18度,加入500 $\mu$ l的IPTG(异丙基- $\beta$ -D-硫代吡喃半乳糖苷,天根RT108-01),过夜诱导,诱导时间16小时。

[0137] 2蛋白纯化

[0138] 2.1菌体收获及裂解

[0139] 仪器耗材:离心机,匀浆机,高速离心机,0.45 $\mu$ m滤膜(默克密理博,SLHV033),裂解液

[0140] 步骤:菌体经5000rpm,30分钟离心收集,加入裂解液进行重悬,经过匀浆机进行破碎。高速离心分离上清和沉淀。上清经过0.45 $\mu$ m的滤膜后准备进入下一步层析纯化。

[0141] 2.2层析柱纯化流程:

[0142] 仪器:AKTA-purify,层析柱,层析溶液

[0143] 2.2.1组蛋白亲和层析

[0144] 步骤:20ml组蛋白亲和层析柱直接用购买的介质(GE生命科学,17-5318-02)装柱,用0.5%的溶液B(20mM Tris-HCl,PH=8.0;250mM NaCl;1M咪唑)进行平衡。根据AKTA仪器说明操作得到纯化后蛋白溶液。

[0145] 2.2.2阳离子交换层析

[0146] 步骤:使用20ml阳离子交换层析柱(GE生命科学,17-0407-01),用0.5%的溶液B(20mM HEPES-KOH PH=7.5;1M KCL)对其进行平衡。根据AKTA仪器说明操作得到纯化后蛋白溶液。

[0147] 3浓缩换液

[0148] 仪器耗材:30kD浓缩管(默克密理博,UFC903096),低温高速离心机

[0149] 步骤:将蛋白溶液加入浓缩管中,4 $^{\circ}$ C离心,5000rpm,40分钟,取出浓缩管,清除穿透液,加入15ml蛋白保存液(20mM Hepes,PH=7.5;150mM KCl;1%蔗糖;30%甘油;1mM二硫

苏糖醇 (DTT)), 4℃ 离心40分钟;重复上述步骤3次,得到最终his-S.P.dCas9-FLAG蛋白溶液,分装后放入-80℃ 保存。

[0150] 实施例4:游离DNA和Cas蛋白结合

[0151] 通常情况下,CRISPR/Cas系统会基于PAM区域(序列:NGG,N可以是A/T/C/G)以及RNA的指导对特定的DNA序列进行识别和切割。为了验证Cas蛋白及其变体与核酸的直接结合是否也依赖于PAM区域,我们制备了SEQ ID NO.1-6号DNA序列。其中1-3为100bp,4-6为200bp,其中1和4号序列上没有Cas9相应的PAM区域(NGG),2和5号序列上各有一个PAM区域,而3和6号序列上各有3个PAM区域。反应所用的dCas9蛋白与实施5、6和7相同,即为实施例3中生产纯化的FLAG-dCas9蛋白。反应条件如下:

[0152] 反应缓冲液(BA):50mM三羟甲基氨基甲烷盐酸盐(Tris-HCl,pH 7.4);150mM氯化钠(NaCl)

[0153] 反应体系:

[0154] 总体积:20μl

[0155] dCas9蛋白:2μg

[0156] DNA:0.25μg

[0157] 无核酸酶双蒸水:定容到20μl

[0158] 反应过程及结果:

[0159] 上述反应体系加入200μl PCR反应管中混合均匀后,室温或37℃ 反应15分钟到1小时。反应结束后,使用非变性胶,在BioRad电泳槽中,以80V电压,400mA电流运行80分钟。取出非变性胶后用DuRed(上海翊圣生物科技有限公司,10202ES76)将核酸染色,继而使用凝胶成像仪拍摄取得图片(图2)。图2展示了500bp左右的游离DNA(SEQ ID NO.7)与dCas蛋白的结合电泳条带。第1泳道是游离DNA自身的条带,属于阴性对照;而2号泳道描述的是dCas9和游离DNA(SEQ ID NO.7)结合之后的情况。在阴性对照中,游离DNA的条带清晰完整,电泳成功,而第2条泳道由于游离DNA和dCas9相结合,导致DNA无法正常进行电泳,几乎所有DNA都留在了泳道井中,故而在相对应处没有明显条带。为了计算dCas9和DNA的结合比例(E1),我们使用以下公式进行计算:

$$[0160] \quad E1 = \frac{1 \text{ 号泳道条带亮度} - 2 \text{ 号泳道相对应位置条带亮度}}{1 \text{ 号泳道条带亮度}}$$

[0161] 该公式将1号泳道条带亮度作为总DNA量的基准,2号泳道相对应位置条带亮度为未与dCas9结合的DNA量,于是分子计算后便是与dCas9蛋白结合的DNA的量,再除以总DNA量得出dCas9和DNA的结合效率,计算得到dCas9蛋白和DNA结合效率为97.4%。使用相同的实验和计算方法,表3总结了不同游离DNA长度所对应的dCas9-DNA结合比例,结果显示dCas9能无差别高效的结合上述11种DNA,展现出良好的从溶液中分离核酸的潜质。

[0162] 表3:各个DNA对应的dCas9-DNA结合比例

	SEQ ID NO.										
	1	2	3	4	5	6	7	8	9	10	11
[0163] DCAS9-DNA 结合比例 (E1)	97.1%	96.9%	96.1%	96.9%	97.0%	96.8%	97.1%	97.4%	98.4%	97.2%	96.4%

[0164] 实施例5: 偶联了FLAG的dCas9蛋白在溶液中结合游离DNA后, 与固相珠结合并将其分离出来

[0165] 反应体系:

[0166] 总体积: 100 $\mu$ l

[0167] dCas9蛋白: 10 $\mu$ g

[0168] DNA (SEQ ID NO. 7、8、9、11, 代表不同游离DNA的长度): 每种DNA各0.25 $\mu$ g

[0169] 无核酸酶双蒸水: 定容到100 $\mu$ l

[0170] 上述反应体系加入1.5ml PCR反应管中混合均匀后, 室温或37 $^{\circ}$ C反应15分钟到1小时。

[0171] 取15 $\mu$ l anti-FLAG亲和凝胶珠 (Bimake, B23101) 加入1.5毫升离心管中, 加入150 $\mu$ l 实施例4中的缓冲液 (BA), 混合均匀后以5000RPM的转速离心30秒, 弃上清, 重复上述步骤一次。弃上清后将上述100 $\mu$ l反应体系加入清洗好的亲和凝胶珠中, 室温中反转混匀, 反应1小时。

[0172] 反应结束后以5000RPM的转速离心30秒, 弃上清, 加入150 $\mu$ l 实施例4中的缓冲液 (BA), 混合均匀后以5000RPM的转速离心30秒, 弃上清, 重复上述步骤一次。弃上清后在凝胶珠中加入30 $\mu$ l蛋白酶K (天根生化, RT403), 并加入20 $\mu$ l BA。混合均匀后55 $^{\circ}$ C水浴加热30分钟, 以5000RPM的转速离心30秒, 取上清得到分离出来的游离DNA。经电泳跑胶并与阳性对照 (0.25 $\mu$ g DNA溶于30 $\mu$ l蛋白酶K加20 $\mu$ l BA溶液中) 比较之后, 根据以下公式计算出DNA分离效率E2, 并总结在表4中:

$$[0173] \quad E2 = \frac{\text{DNA 抓取泳道亮度}}{\text{阳性对照泳道亮度}}$$

[0174] 表4: 各个DNA对应的dCas9-DNA分离效率

	SEQ ID NO.			
	7	8	9	11
[0175] DNA 分离效率 (E2)	79.5%	77.2%	77.6%	70.9%

[0176] 实施例6: 偶联了FLAG的dCas9蛋白与固相珠结合后, 在溶液中结合游离DNA并将其分离出来

[0177] 第一步: 取15 $\mu$ l Anti-FLAG亲和凝胶珠 (GenScript, L00439-1) 加入1.5毫升离心管中, 加入150 $\mu$ l 实施例4中的缓冲液 (BA), 混合均匀后以5000RPM的转速离心30秒, 弃上清, 重复上述步骤一次。弃上清后在凝胶珠中加入500 $\mu$ l BA, 将凝胶珠混合均匀。然后在混合液中加入1 $\mu$ l浓度为10 $\mu$ g/ $\mu$ l dCas9蛋白溶液, 室温中反转混匀, 反应1小时。

[0178] 第二步: 反应结束后以5000RPM的转速离心30秒, 弃上清, 加入150 $\mu$ l 实施例4中的缓冲液 (BA), 混合均匀后以5000RPM的转速离心30秒, 弃上清, 重复上述步骤一次。将清洗后的凝胶珠分别加入4管1ml含有不同长度游离DNA溶液中 (游离DNA使用了SEQ ID NO. 7、8、9、11, 分别代表不同游离DNA的长度957、486、166和2655bp), 室温或37 $^{\circ}$ C反转混匀反应15分钟到1小时。

[0179] 第三步:反应结束后以5000RPM的转速离心30秒,弃上清,加入150 $\mu$ l实施例4中的缓冲液(BA),混合均匀后以5000RPM的转速离心30秒,弃上清,重复上述步骤一次。弃上清后在凝胶珠中加入30 $\mu$ l蛋白酶K(天根生化,RT403),并加入20 $\mu$ l BA。混合均匀后55 $^{\circ}$ C水浴加热30分钟,以5000RPM的转速离心30秒,取上清得到分离出来的游离DNA。

[0180] 为了与凯杰现有的试剂盒进行比较,第二步中的1ml游离DNA溶液,使用QIAamp<sup>®</sup> Circulating Nucleic Acid kit(QIAGEN(凯杰)公司),根据使用说明书的步骤提取溶液中的游离DNA。

[0181] 阳性对照:相同量的DNA(游离DNA使用了SEQ ID NO.7、8、9、11,分别代表不同游离DNA的长度957、486、166和2655bp)溶解于30 $\mu$ l蛋白酶K(天根生化,RT403)中,并加入20 $\mu$ l BA。

[0182] 将本发明抓取的DNA、凯杰试剂盒抓取的DNA和阳性对照一起使用Agilent 2100 High Sensitivity DNA Analysis Kit(Cat.5067-4626)进行DNA的定量分析,所得出的抓取DNA量和抓取百分比结果总结在表4中。提取效率为表4中本发明或者凯杰试剂盒抓取DNA的量和阳性对照DNA的量的比值(即为抓取DNA百分比)。本发明游离DNA提取效率接近80%,而QIAamp<sup>®</sup> Circulating Nucleic Acid kit只有不到50%,因此本发明将游离DNA的提取效率提高了60%左右。

[0183] 表4:各长度游离DNA用本发明和QIAamp<sup>®</sup> Circulating Nucleic Acid kit的提取效率对比

	DNA 大小 [BP]	本发明抓取 DNA 量[pg/ $\mu$ l]	Qiagen Kit 抓 取 DNA 量 [pg/ $\mu$ l]	阳性对照 DNA 量[pg/ $\mu$ l]	本发明提 取效率	Qiagen Kit 提 取效率
[0184]	166	1,025.24	631.79	1,305.28	78.55%	48.40%
	486	1,012.37	549.34	1,261.83	80.23%	38.22%
	957	818.39	308.745	1077.58	75.95%	28.65%
	2655	711.04	436.435	1179.08	60.30%	28.65%

[0185] 实施例7:带有His-Tag的Cpf1(Cas12a)蛋白与固相珠结合后,在溶液中结合游离DNA并将其分离出来

[0186] 第一步:带有His-Tag的Cpf1(Cas12a)蛋白(IDT,Alt-R<sup>®</sup>A.s.Cas12a(Cpf1)V3,Cat.No.1081068)与His-Tag分离磁珠(Dynabeads<sup>™</sup> His-Tag Isolation and Pulldown,Cat.No.10103D)连接。取300 $\mu$ l Dynabeads放入1.5ml离心管中,加1ml pH=7.4的PBS,混合均匀,将离心管置于磁性分离器(海狸公司,Cat.No.60201)对应的样品孔中,静置两分钟,使磁珠被吸附聚集于管壁,溶液恢复澄清,小心的吸出上清,重复上述步骤两遍。40 $\mu$ l Cpf1蛋白与Dynabeads混匀,加1ml pH=7.4的PBS,室温放混匀仪上转动20分钟至2小时,将离心管置于磁性分离器对应的样品孔中,静置两分钟,弃上清,加入500 $\mu$ l实施例4中的缓冲液(BA),混合均匀后将离心管置于磁性分离器对应的样品孔中,静置两分钟,弃上清,重复上述步骤一次。将清洗后的Dynabeads取10到100 $\mu$ l,加入1ml含有SEQ ID NO.9号DNA的溶液中,室温或37 $^{\circ}$ C反转混匀反应15分钟到6小时。

[0187] 第二步:反应结束后将装有反应溶液和Dynabeads的离心管置于磁性分离器对应的样品孔中,静置两分钟,弃上清,加入150 $\mu$ l实施例4中的缓冲液(BA),将离心管置于磁性分离器对应的样品孔中,静置两分钟,弃上清,重复上述步骤一次。弃上清后在Dynabeads中

加入30 $\mu$ l蛋白酶K(天根生化,RT403),并加入20 $\mu$ l BA。混合均匀后55 $^{\circ}$ C水浴加热30分钟,将离心管置于磁性分离器对应的样品孔中,静置两分钟,取上清得到分离出来的游离DNA。

[0188] 为了与凯杰现有试剂盒进行对比,第三步中的1ml游离DNA溶液,使用QIAamp<sup>®</sup>Circulating Nucleic Acid kit(QIAGEN(凯杰)公司),根据使用说明书的步骤提取血浆中的游离DNA。

[0189] 阳性对照:相同量的DNA(游离DNA使用了SEQ ID NO.9,代表游离DNA的长度166bp)溶解于30 $\mu$ l蛋白酶K(天根生化,RT403)中,并加入20 $\mu$ l BA。

[0190] 将本发明抓取的DNA、凯杰试剂盒抓取的DNA和阳性对照一起使用Agilent 2100 High Sensitivity DNA Analysis Kit(Cat.5067-4626)进行DNA的定量分析,所得出的抓取DNA量和抓取百分比结果总结在表5中。提取效率为表5中本发明或者凯杰试剂盒抓取DNA的量和阳性对照DNA的量的比值(即为抓取DNA百分比)。本发明游离DNA提取效率接近80%,而QIAamp<sup>®</sup>Circulating Nucleic Acid kit只有51%,因此本发明将在血浆中提取游离DNA的提取效率提高了54.7%。

[0191] 表5:160bp游离DNA用本发明和QIAamp<sup>®</sup>Circulating Nucleic Acid kit的提取效率对比

[0192]	DNA 大小 [BP]	本发明抓取 DNA 量[pg/ $\mu$ l]	QIAGEN KIT 抓 取 DNA 量[pg/ $\mu$ l]	阳性对照 DNA 量[pg/ $\mu$ l]	本发明提取 效率	QIAGEN KIT 提取效率
	166	1,018.62	658.34	1,293.72	78.74%	50.89%

[0193] 实施例8:偶联了AVI-Tag的dCas9蛋白与固相珠结合后,在血浆中结合外源游离DNA并将其分离出来

[0194] 第一步:his-S.P.dCas9-AVI蛋白的纯化,参考实施例3,将FLAG序列替换为AVI-Tag序列(DNA序列:ggcctgaacgatatttttgaagcgcagaaaattgaatggcatgaa)。

[0195] 第二步:使用AViDITY公司的BirA-500kit(Cat.BirA500),根据试剂盒的说明书将AVI-Tag-dCas9蛋白生物素化。在1.5ml的离心管中加入相应的反应物进行反应,反应体系如下:

[0196] 10 $\times$ BiomixA 8 $\mu$ l

[0197] 10 $\times$ BiomixB 8 $\mu$ l

[0198] AVI-Tag-dCas9蛋白 50 $\mu$ l

[0199] BirA酶 0.8 $\mu$ l

[0200] 超纯水 补充总体积至80 $\mu$ l

[0201] 之后将反应液放室温反应1小时或过夜。

[0202] 第三步:生物素化的蛋白与Streptavidin凝胶珠连接。取400 $\mu$ l GenScript Streptavidin凝胶珠(Cat.No.L00353)放1.5ml离心管中,加1ml pH=7.4的PBS,室温5000g离心30s,取出后小心的吸出上清,重复上述步骤两遍。80 $\mu$ l生物素化的蛋白与Streptavidin凝胶珠混匀,加1ml pH=7.4的PBS,室温放混匀仪上转动20分钟至2小时,室温5000g离心30秒,弃上清,加入500 $\mu$ l实施例4中的缓冲液(BA),混合均匀后以5000RPM的转速离心30秒,弃上清,重复上述步骤一次。将清洗后的凝胶珠取10到100 $\mu$ l,加入1ml含有SEQ ID NO.9号DNA的血浆中,室温或37 $^{\circ}$ C反转混匀反应15分钟到6小时。

[0203] 第四步:反应结束后以5000RPM的转速离心30秒,弃上清,加入150 $\mu$ l实施例4中的

缓冲液(BA),混合均匀后以5000RPM的转速离心30秒,弃上清,重复上述步骤一次。弃上清后在凝胶珠中加入30 $\mu$ l蛋白酶K(天根生化,RT403),并加入20 $\mu$ l BA。混合均匀后55 $^{\circ}$ C水浴加热30分钟,以5000RPM的转速离心30秒,取上清得到分离出来的游离DNA。

[0204] 为了与凯杰现有试剂盒进行对比,第三步中的1ml游离DNA溶液,使用QIAamp<sup>®</sup>Circulating Nucleic Acid kit(QIAGEN(凯杰)公司),根据使用说明书的步骤提取血浆中的游离DNA。

[0205] 阳性对照:相同量的DNA(游离DNA使用了SEQ ID NO.9,代表游离DNA的长度166bp)溶解于30 $\mu$ l蛋白酶K(天根生化,RT403)中,并加入20 $\mu$ l BA。

[0206] 将本发明抓取的DNA、凯杰试剂盒抓取的DNA和阳性对照一起使用Agilent 2100 High Sensitivity DNA Analysis Kit(Cat.5067-4626)进行DNA的定量分析,所得出的抓取DNA量和抓取百分比结果总结在表6中。提取效率为表6中本发明或者凯杰试剂盒抓取DNA的量和阳性对照DNA的量的比值(即为抓取DNA百分比)。本发明游离DNA提取效率接近78%,而QIAamp<sup>®</sup>Circulating Nucleic Acid kit只有51%,因此本发明将在血浆中提取游离DNA的提取效率提高了53%。

[0207] 表6:160bp游离DNA用本发明和QIAamp<sup>®</sup>Circulating Nucleic Acid kit的提取效率对比

	DNA 大小 [BP]	本发明抓取 DNA 量[pg/ $\mu$ l]	QIAGEN KIT 抓取 DNA 量[pg/ $\mu$ l]	阳性对照 DNA 量[pg/ $\mu$ l]	本发明提取 效率	QIAGEN KIT 提 取效率
[0208]	166	1,007.05	674.35	1,305.45	77.14%	51.66%

[0209] 实施例9:带有His-Tag的Cas9蛋白与固相珠结合后,在血浆中结合外源游离DNA并将其分离出来

[0210] 第一步:带有His-Tag的Cas9蛋白(IDT,Alt-R<sup>®</sup>S.p.Cas9 Nuclease V3,Cat.No.1081058)与His-Tag分离磁珠(Dynabeads<sup>™</sup> His-Tag Isolation and Pulldown,Cat.No.10103D)连接。取300 $\mu$ l Dynabeads放入1.5ml离心管中,加1ml pH=7.4的PBS,混合均匀,将离心管置于磁性分离器(海狸公司,Cat.No.60201)对应的样品孔中,静置两分钟,使磁珠被吸附聚集于管壁,溶液恢复澄清,小心的吸出上清,重复上述步骤两遍。40 $\mu$ lCas9蛋白与Dynabeads混匀,加1ml pH=7.4的PBS,室温放混匀仪上转动20分钟至2小时,将离心管置于磁性分离器对应的样品孔中,静置两分钟,弃上清,加入500 $\mu$ l实施例4中的缓冲液(BA),混合均匀后将离心管置于磁性分离器对应的样品孔中,静置两分钟,弃上清,重复上述步骤一次。将清洗后的Dynabeads取10到100 $\mu$ l,加入1ml含有SEQ ID NO.9号DNA的血浆中,室温或37 $^{\circ}$ C反转混匀反应15分钟到6小时。

[0211] 第二步:反应结束后将含有血浆和Dynabeads的离心管置于磁性分离器对应的样品孔中,静置两分钟,弃上清,加入150 $\mu$ l实施例4中的缓冲液(BA),将离心管置于磁性分离器对应的样品孔中,静置两分钟,弃上清,重复上述步骤一次。弃上清后在Dynabeads中加入30 $\mu$ l蛋白酶K(天根生化,RT403),并加入20 $\mu$ l BA。混合均匀后55 $^{\circ}$ C水浴加热30分钟,将离心管置于磁性分离器对应的样品孔中,静置两分钟,取上清得到分离出来的游离DNA。

[0212] 为了与凯杰现有试剂盒进行对比,第三步中的1ml游离DNA溶液,使用QIAamp<sup>®</sup>Circulating Nucleic Acid kit(QIAGEN(凯杰)公司),根据使用说明书的步骤提取血浆中的游离DNA。

[0213] 阳性对照:相同量的DNA(游离DNA使用了SEQ ID NO.9,代表游离DNA的长度166bp)溶解于30 $\mu$ l蛋白酶K(天根生化,RT403)中,并加入20 $\mu$ l BA。

[0214] 将本发明抓取的DNA、凯杰试剂盒抓取的DNA和阳性对照一起使用Agilent 2100 High Sensitivity DNA Analysis Kit(Cat.5067-4626)进行DNA的定量分析,所得出的抓取DNA量和抓取百分比结果总结在表7中。提取效率为表7中本发明或者凯杰试剂盒抓取DNA的量和阳性对照DNA的量的比值(即为抓取DNA百分比)。本发明游离DNA提取效率超过85%,而QIAamp®Circulating Nucleic Acid kit只有55%,因此本发明将在血浆中提取游离DNA的提取效率提高了54.5%。

[0215] 表7:160bp游离DNA用本发明和QIAamp®Circulating Nucleic Acid kit的提取效率对比

	DNA 大小 [BP]	本发明抓取 DNA 量[pg/ $\mu$ l]	QIAGEN KIT 抓 取 DNA 量[pg/ $\mu$ l]	阳性对照 DNA 量[pg/ $\mu$ l]	本发明提取 效率	QIAGEN KIT 提取效率
[0216]	170	1,160.66	754.54	1,363.28	85.14%	55.35%

[0217] 实施例10:偶联了AVI-Tag的dCas9蛋白与固相珠结合后,在血浆中结合内源游离DNA并将其分离出来

[0218] 第一步与第二步与实施例8中的第一步和第二步相同,得到生物素化的dCas9蛋白。

[0219] 第三步:生物素化的蛋白与Streptavidin凝胶珠连接。取400 $\mu$ l GenScript Streptavidin凝胶珠(Cat.No.L00353)放1.5ml离心管中,加1ml pH=7.4的PBS,室温5000g离心30s,取出后小心的吸出上清,重复上述步骤两遍。80 $\mu$ l生物素化的蛋白与Streptavidin凝胶珠混匀,加1ml pH=7.4的PBS,室温放混匀仪上转动20分钟至2小时,室温5000g离心30秒,弃上清,加入500 $\mu$ l实施例4中的缓冲液(BA),混合均匀后以5000RPM的转速离心30秒,弃上清,重复上述步骤一次。将清洗后的凝胶珠取10到100 $\mu$ l,加入1ml血浆中,室温或37 $^{\circ}$ C反转混匀反应15分钟到6小时。

[0220] 第四步:反应结束后以5000RPM的转速离心30秒,弃上清,加入150 $\mu$ l实施例4中的缓冲液(BA),混合均匀后以5000RPM的转速离心30秒,弃上清,重复上述步骤一次。弃上清后在凝胶珠中加入30 $\mu$ l蛋白酶K(天根生化,RT403),并加入20 $\mu$ l BA。混合均匀后55 $^{\circ}$ C水浴加热30分钟,以5000RPM的转速离心30秒,取上清得到分离出来的游离DNA。

[0221] 为了与凯杰现有试剂盒进行对比,第三步中的1ml游离DNA溶液,使用QIAamp®Circulating Nucleic Acid kit(QIAGEN(凯杰)公司),根据使用说明书的步骤提取血浆中的游离DNA。

[0222] 将本发明抓取的DNA和凯杰试剂盒抓取的DNA一起使用Agilent 2100 High Sensitivity DNA Analysis Kit(Cat.5067-4626)进行DNA的定量分析,所得出的抓取DNA量结果总结在表8中。本发明游离DNA提取量接近28.49pg/ $\mu$ l,而QIAamp®Circulating Nucleic Acid kit只有24.68pg/ $\mu$ l,因此本发明将在血浆中提取内源游离DNA的提取效率提高了15.4%。

[0223] 表8:血浆游离DNA用本发明和QIAamp®Circulating Nucleic Acid kit的提取量对比

	DNA 大小 [BP]	本发明抓取 DNA 量[pg/ $\mu$ l]	QIAGEN KIT 抓 取 DNA 量[pg/ $\mu$ l]	本发明提取效 率提高
[0224]	177	28.49	24.68	15.4%

[0001]		SEQUENCE LISTING	
[0002]	<110>	苏州克睿基因生物科技有限公司	
[0003]	<120>	一种Cas蛋白体系分离DNA的方法	
[0004]	<130>	AJFS2017BJ692	
[0005]	<160>	34	
[0006]	<170>	PatentIn version 3.5	
[0007]	<210>	1	
[0008]	<211>	100	
[0009]	<212>	DNA	
[0010]	<213>	synthetic sequence	
[0011]	<400>	1	
[0012]		ctgcgtaaac gtcgctgtgc tagctctttt gcagtcacag cttttcgtgc atgagtacgt	60
[0013]		atthtgaaac tcaagatcgc attcatgcgt cttcacgtga	100
[0014]	<210>	2	
[0015]	<211>	100	
[0016]	<212>	DNA	
[0017]	<213>	synthetic sequence	
[0018]	<400>	2	
[0019]		ctgcgtaaac gtcgctgtgc tagctctttt gcaggcacag cttttcgtgc atgagtacgt	60
[0020]		atthtgaaac tcaagatcgc attcatgcgt cttcacgtga	100
[0021]	<210>	3	
[0022]	<211>	100	
[0023]	<212>	DNA	
[0024]	<213>	synthetic sequence	
[0025]	<400>	3	
[0026]		ctgcgtaaac gtcgcggtgc tagctctttt gcaggcacag cttttcgtgc atgagtatgg	60
[0027]		atthtgaaac tcaagatcgc attcatgcgt cttcacgtga	100
[0028]	<210>	4	
[0029]	<211>	200	
[0030]	<212>	DNA	
[0031]	<213>	synthetic sequence	
[0032]	<400>	4	
[0033]		ctgcttcgtg catgagtacg tattctcttt acgtgactgc gtaagtaaac gtcgctgtgc	60
[0034]		tagtgcagtc acagttgaaa ctcaagatcg cattcaagct ctttactgct tacgtgactg	120
[0035]		cgttgcagtc acagcttagt acgtattgcg tcttcagtgc atgttgaaac tcaagatcgc	180
[0036]		attcatgcgt cttcacgtga	200
[0037]	<210>	5	
[0038]	<211>	200	
[0039]	<212>	DNA	
[0040]	<213>	synthetic sequence	
[0041]	<400>	5	

[0042]	ctgcttcgtg catgagtacg tattctcttt acgtgactgc gtaagtaaac gtcgctgtgc	60
[0043]	tagtgcagtc acagttgaaa ctcaagatcg cattcaaggt ctttactgct tacgtgactg	120
[0044]	cgttgcagtc acagcttagt acgtattgcg tcttcagtc atgttgaaac tcaagatcgc	180
[0045]	attcatgcgt cttcacgtga	200
[0046]	<210>	6
[0047]	<211>	200
[0048]	<212>	DNA
[0049]	<213>	synthetic sequence
[0050]	<400>	6
[0051]	ctgcttcgtg catgcggacg tattctcttt acgtgactgc gtaagtaaac gtcgctgtgc	60
[0052]	tagtgcagtc acagttgaaa ctcaagatcg cattcaaggt cgttactgct tacgtgactg	120
[0053]	cgttgcagtc acagcttagt acgtattgcg tcttcagtc atgttgaaac tcaaggtcgc	180
[0054]	attcatgcgt cttcacgtga	200
[0055]	<210>	7
[0056]	<211>	957
[0057]	<212>	DNA
[0058]	<213>	Homo sapiens
[0059]	<400>	7
[0060]	gccctcatga tttttaaaa cacagcatcc tcaacctga ggcggagtc ttcataacaa	60
[0061]	agatactatc agttccaaa ctcaagatc aggtgactcc gactcctcct ttatccaatg	120
[0062]	tgctcctcat ggccactgtt gcttgggct ctctgtcatg gggaatcccc agatgcacc	180
[0063]	aggaggggcc ctctcccact gcatctgtca cttcacagcc ctgcgtaaac gtcctgtgc	240
[0064]	taggtctttt gcaggcacag ctttctctcc atgagtacgt attttgaaac tcaagatcgc	300
[0065]	attcatgcgt cttcacctgg aaggggtcca tgtgcccctc cttctggcca ccatgcgaag	360
[0066]	ccacactgac gtgcctctcc ctccctccag gaagcctacg tgatggccag cgtggacaac	420
[0067]	ccccactgtg gccgcctgct gggcatctgc ctacactcca ccgtgcagct catcacgcag	480
[0068]	ctcatgccct tcggctgcct cctggactat gtccgggaac acaaagacaa tattggctcc	540
[0069]	cagtacctgc tcaactggtg tgtgcagatc gcaaaggtaa tcaggggaag gagatacggg	600
[0070]	gaggggagat aaggagccag gatcctcaca tgcggtctgc gctcctggga tagcaagagt	660
[0071]	ttgccatggg gatatgtgtg tgcgtgcatg cagcacacac acattccttt attttgatt	720
[0072]	caatcaagtt gatctcttg tgcacaaac agtgacctgc ccatctgcat gtggaaactc	780
[0073]	tcatcaatca gctaccttg aagaattttc tctttattga gtgctcagtg tggctgatg	840
[0074]	tctctgttct ttttctctg gaattcttg tgaatactgt ggtgattgt agtggagaag	900
[0075]	gaatattgct tccccattc aggacttgat aacaaggtaa gcaagccagg ccaaggc	957
[0076]	<210>	8
[0077]	<211>	486
[0078]	<212>	DNA
[0079]	<213>	Homo sapiens
[0080]	<400>	8
[0081]	ctctcccact gcatctgtca cttcacagcc ctgcgtaaac gtcctgtgc taggtctttt	60
[0082]	gcaggcacag ctttctctcc atgagtacgt attttgaaac tcaagatcgc attcatgcgt	120
[0083]	cttcacctgg aaggggtcca tgtgcccctc cttctggcca ccatgcgaag ccacactgac	180

[0084]	gtgcctctcc ctccctccag gaagcctacg tgatggccag cgtggacaac ccccacgtgt	240
[0085]	gccgcctgct gggcatctgc ctcacctcca ccgtgcagct catcacgcag ctcacgtcct	300
[0086]	tcggetgect cctggactat gtccgggaac acaaagacaa tattggctcc cagtacctgc	360
[0087]	tcaactggtg tgtgcagatc gcaaaggtaa tcagggaagg gagatacggg gaggggagat	420
[0088]	aaggagccag gatcctcaca tgcggtctgc gctcctggga tagcaagagt ttgccatggg	480
[0089]	gatatg	486
[0090]	<210> 9	
[0091]	<211> 166	
[0092]	<212> DNA	
[0093]	<213> Homo sapiens	
[0094]	<400> 9	
[0095]	ctccaggaag cctacgtgat ggccagcgtg gacaaccccc acgtgtgccg cctgctgggc	60
[0096]	atctgcctca cctccaccgt gcagctcacc acgcagctca tgcccttcgg ctgcctcctg	120
[0097]	gactatgtcc gggaacacaa agacaatatt ggctcccagt acctgc	166
[0098]	<210> 10	
[0099]	<211> 609	
[0100]	<212> DNA	
[0101]	<213> Homo sapiens	
[0102]	<400> 10	
[0103]	tctccacaag gaggcattga aaggctgtag ttgttcacct gcccaagaac taggaggtct	60
[0104]	gggggtgggag agtcagcctg ctctggatgc tgaagaatg tctgtttttc ctttttagaaa	120
[0105]	gttctctgta tgtcaagctg gtcgagaaaa gctttgaaac aggtaagaca ggggtctagc	180
[0106]	ctgggtttgc acaggattgc ggaagtgatg aaccgcgaat aaccctgcct ggatgagga	240
[0107]	gtgggaagaa attagtagat gtgggaatga atgatgagga atggaacag cggttcaaga	300
[0108]	cctgccaga gctgggtggg gtctctcctg aatcctctc accatctctg actttccatt	360
[0109]	ctaagcactt tgaggatgag tttctagctt caatagacca aggactctct cctaggcctc	420
[0110]	tgtattcctt tcaacagctc cactgtcaag agagccagag agagcttctg ggtggcccag	480
[0111]	ctgtgaaatt tctgagctcc ttagggatag ccctaaacga accagatcat cctgaggaca	540
[0112]	gccaaagagt tttgccttct ttcaagacaa gcaacagtac tcacatagge tgtgggcaat	600
[0113]	ggtcctgtc	609
[0114]	<210> 11	
[0115]	<211> 2655	
[0116]	<212> DNA	
[0117]	<213> Homo sapiens	
[0118]	<400> 11	
[0119]	cgttccagaa gcggcagaag cttacctccg aggggtgccg caagctcctg ctagacacct	60
[0120]	tgtgagtgcg gcgggcccgg ggcgcggga gccgttgcca cgcggacccc ctcggcagga	120
[0121]	gtcgggctcg cagcccgcgt gcaggccttg ggcgccttcc agctctgagc cttccacgtt	180
[0122]	acggagccga cgcggtctcc ccgtagcac tggggttatt ccttaattct aagggcggcg	240
[0123]	gggcggggcg gtgcttctag ggctgtgtgg cctgggaatg gaaagggggc ttcgtccgga	300
[0124]	gcttgagggc gcgagcagcc gtcttgggac acaaagagtc tttcccggcc ataagcatcg	360
[0125]	tgctcgggga cgttgtctct tgctggcgcg caaagggtgt gggggccttt cttagtga	420

[0126]	agaacataaa tcccacggct ttctcttacc cgaatcccct ttttttttgt gtttgcacac	480
[0127]	gtgtttttga gggtaggttg aagggaaagg tagaagcgtt gaaggaagca tgaatagttt	540
[0128]	tgaactcact gggaagcaaa atcccgtcat tgttccaggc actgccaggg agggagcagg	600
[0129]	gggcggaggt tattgatccg ggagagaagc gcgcgcgaag agatgctccg gcctcgcgcg	660
[0130]	gcggcgggga tacctacgta gggaggcttt tatgtgctca gccagctaata gctccctgcc	720
[0131]	gctcgccttc ctggctgctc cgcagtcttt tagatggccc gaggagcctc tctgagtgcc	780
[0132]	gcgatttggg aaagagcttt taactgggct gttaaagggc tgccttcttt tctccaaatg	840
[0133]	cgatctggat tctgctctg gggaacagtt gtgctacaga gagatatttt aacgcctga	900
[0134]	tattatattt cagtgtcaat tttacgaatt aaaaagaaca atcaaaacac tagggggaaa	960
[0135]	aaaccgacaa gaataacctg gcaattgcat aacctgagaa ttacttcttt tcttgaacat	1020
[0136]	gttcctttga gcttcttgcg gaagaaattc tcttaggatg agtcgtcccc gacagaagat	1080
[0137]	acatatttgg tttctgtgt gcttcaactt ttggtttttc atcttggact gcacactgaa	1140
[0138]	tttttaaaaa acgttcatcc atcaaacatt tgagcaccta taggggtgacg gctgcctctg	1200
[0139]	ggcagaagga atacgaagat ccttaggaat gtaaaactca tgggtgtcttc aaggcataca	1260
[0140]	gaaaaaaaga agaattgaaa actcaaatth gagataaatt tttattttaa gacaaactga	1320
[0141]	taaatgggtca agaaaaagtt ttaaaatatt gacataaaaa agctgttttt ctcccactag	1380
[0142]	attgccatga ctttgtactt ataaagtcta ctaaattata ttcaaaaagt gtgtataact	1440
[0143]	gtaccatttt cgttaaaata ttgtgtagaa aaaagtttgg gggaatatat aacaaataat	1500
[0144]	taactgtaga tcctctggg tgtaaagatt acaggaggct ttcactttga gcaactctca	1560
[0145]	atagtgtgaa gtttggagtt tttacaataa gcattatgtg tgcagaaaaa attaaaatac	1620
[0146]	ataaacgtaa agaaacatga aaaaatgatc agtcttaatc atgcaaatga aaactgcagt	1680
[0147]	gagctatttt tcctataatt attgcctgag actgtacagt gaaatgtttt ctcatattt	1740
[0148]	tatcgggtaa ggggtgtcaa tccttttga aagcgatttt aagacattta tcagatctta	1800
[0149]	atgttcattt ctgagtttgg cttatatata agctttaaag ggttaaaatt gtatgtaggc	1860
[0150]	tgggtgaggt ggctcatgcc tgtaatccca gcactttggg aggccgagc gggtagatca	1920
[0151]	cctgaggtgg ggagttgag actgcctgac caacatggag aaacccccgc tctactaaaa	1980
[0152]	atacaaaatt agctgggtgt ggtggtgggc gcctgtaatc ccagctactc gggaggctga	2040
[0153]	ggcaggagaa tcatttgaac cctggagggtg gaggttgcgg tgggctgaaa tgcgctcact	2100
[0154]	gcactccagc ctgggcaaca agaattgaaac tcctctcaaa aaaaaaaaaa ttctatgtaa	2160
[0155]	tgtgtattac aactatgtaa aacgcattgt tatgaagaat actggaaaga aatagagcaa	2220
[0156]	aaatataaga gcctttatta agtttctcaa aaggtctctc tggacctctg gacttgattg	2280
[0157]	agtacacagt cttatttttc atagcagcct gtacctcctt tatcacaatc ataattaatt	2340
[0158]	attatttgag taaattgtta aggtcggtaa gggtagcaat cgtgtctgct gtatcttgtt	2400
[0159]	tagtgtctc tccttcgtgc ctagctcata aaaatattta gtatttgagg cccggagcag	2460
[0160]	tggctcacgc ctgtaatccc agcactttgg gagtccccgt gggtagatca caaggtcagg	2520
[0161]	agttcaagac cagcttgccc aagatgggtga agccctatct ctactaaaaa taaaaatta	2580
[0162]	gcagggcgcg gtggcaggcg cctgtaatcc cagctacttg ggaggctaag gctggagaat	2640
[0163]	tgettcaacc caggt	2655
[0164]	<210> 12	
[0165]	<211> 60	
[0166]	<212> DNA	
[0167]	<213> synthetic sequence	

[0168]	<400> 12	
[0169]	ctgcgtaaac gtcgctgtgc tagctctttt gcagtcacag cttttcgtgc atgagtacgt	60
[0170]	<210> 13	
[0171]	<211> 62	
[0172]	<212> DNA	
[0173]	<213> synthetic sequence	
[0174]	<400> 13	
[0175]	agcttttcgt gcatgagtac gtattttgaa actcaagatc gcattcatgc gtcttcacgt	60
[0176]	ga	62
[0177]	<210> 14	
[0178]	<211> 72	
[0179]	<212> DNA	
[0180]	<213> synthetic sequence	
[0181]	<400> 14	
[0182]	ctgcgtaaac gtcgctgtgc tagctctttt gcaggcacag cttttcgtgc atgagtacgt	60
[0183]	attttgaaac tc	72
[0184]	<210> 15	
[0185]	<211> 55	
[0186]	<212> DNA	
[0187]	<213> synthetic sequence	
[0188]	<400> 15	
[0189]	cgtgcatgag tacgtatttt gaaactcaag atcgcattca tgcgtcttca cgtga	55
[0190]	<210> 16	
[0191]	<211> 63	
[0192]	<212> DNA	
[0193]	<213> synthetic sequence	
[0194]	<400> 16	
[0195]	ctgcgtaaac gtcgcggtgc tagctctttt gcaggcacag cttttcgtgc atgagtatgg	60
[0196]	att	63
[0197]	<210> 17	
[0198]	<211> 61	
[0199]	<212> DNA	
[0200]	<213> synthetic sequence	
[0201]	<400> 17	
[0202]	gcttttcgtg catgagtatg gattttgaaa ctcaagatcg cattcatgcg tcttcacgtg	60
[0203]	a	61
[0204]	<210> 18	
[0205]	<211> 110	
[0206]	<212> DNA	
[0207]	<213> synthetic sequence	
[0208]	<400> 18	
[0209]	ctgcttcgtg catgagtacg tattctcttt acgtgactgc gtaagtaaac gtcgctgtgc	60

[0210]	tagtgcagtc acagttgaaa ctcaagatcg cattcaagct ctttactgct	110
[0211]	<210> 19	
[0212]	<211> 113	
[0213]	<212> DNA	
[0214]	<213> synthetic sequence	
[0215]	<400> 19	
[0216]	tcgcattcaa gctctttact gcttacgtga ctgcggttga gtcacagctt agtacgtatt	60
[0217]	gcgtcttcag tgcattgtga aactcaagat cgcattcatg cgtcttcacg tga	113
[0218]	<210> 20	
[0219]	<211> 114	
[0220]	<212> DNA	
[0221]	<213> synthetic sequence	
[0222]	<400> 20	
[0223]	ctgcttcgtg catgagtacg tattctcttt acgtgactgc gtaagtaaac gtcgctgtgc	60
[0224]	tagtgcagtc acagttgaaa ctcaagatcg cattcaaggt ctttactgct tacg	114
[0225]	<210> 21	
[0226]	<211> 112	
[0227]	<212> DNA	
[0228]	<213> synthetic sequence	
[0229]	<400> 21	
[0230]	cgcattcaag gtctttactg cttacgtgac tgcggttcag tcacagctta gtacgtattg	60
[0231]	cgtcttcagt gcatgttgaa actcaagatc gcattcatgc gtcttcacgt ga	112
[0232]	<210> 22	
[0233]	<211> 110	
[0234]	<212> DNA	
[0235]	<213> synthetic sequence	
[0236]	<400> 22	
[0237]	ctgcttcgtg catgcggacg tattctcttt acgtgactgc gtaagtaaac gtcgctgtgc	60
[0238]	tagtgcagtc acagttgaaa ctcaagatcg cattcaaggt cgttactgct	110
[0239]	<210> 23	
[0240]	<211> 113	
[0241]	<212> DNA	
[0242]	<213> synthetic sequence	
[0243]	<400> 23	
[0244]	tcgcattcaa ggtcgttact gcttacgtga ctgcggttga gtcacagctt agtacgtatt	60
[0245]	gcgtcttcag tgcattgtga aactcaaggt cgcattcatg cgtcttcacg tga	113
[0246]	<210> 24	
[0247]	<211> 50	
[0248]	<212> DNA	
[0249]	<213> synthetic sequence	
[0250]	<400> 24	
[0251]	gccctcatga tattttaaaa cacagcatcc tcaaccttga ggccggaggtc	50

[0252]	<210> 25	
[0253]	<211> 49	
[0254]	<212> DNA	
[0255]	<213> synthetic sequence	
[0256]	<400> 25	
[0257]	ccttgccctg gcttgcttac cttggtatca agtcctgaat gggggaagc	49
[0258]	<210> 26	
[0259]	<211> 20	
[0260]	<212> DNA	
[0261]	<213> synthetic sequence	
[0262]	<400> 26	
[0263]	ctctcccact gcactgtca	20
[0264]	<210> 27	
[0265]	<211> 20	
[0266]	<212> DNA	
[0267]	<213> synthetic sequence	
[0268]	<400> 27	
[0269]	catatcccca tggcaaactc	20
[0270]	<210> 28	
[0271]	<211> 19	
[0272]	<212> DNA	
[0273]	<213> synthetic sequence	
[0274]	<400> 28	
[0275]	cacaactgacg tgcctctcc	19
[0276]	<210> 29	
[0277]	<211> 19	
[0278]	<212> DNA	
[0279]	<213> synthetic sequence	
[0280]	<400> 29	
[0281]	gcaggtactg ggagccaat	19
[0282]	<210> 30	
[0283]	<211> 20	
[0284]	<212> DNA	
[0285]	<213> synthetic sequence	
[0286]	<400> 30	
[0287]	tctccacaag gaggcatgga	20
[0288]	<210> 31	
[0289]	<211> 20	
[0290]	<212> DNA	
[0291]	<213> synthetic sequence	
[0292]	<400> 31	
[0293]	gacaggacca ttgcccacag	20

[0294]	<210>	32	
[0295]	<211>	27	
[0296]	<212>	DNA	
[0297]	<213>	synthetic sequence	
[0298]	<400>	32	
[0299]		cgttccagaa gcggcagaag cttacct	27
[0300]	<210>	33	
[0301]	<211>	30	
[0302]	<212>	DNA	
[0303]	<213>	synthetic sequence	
[0304]	<400>	33	
[0305]		acctgggttg aagcaattct ccagccttag	30
[0306]	<210>	34	
[0307]	<211>	4145	
[0308]	<212>	DNA	
[0309]	<213>	synthetic sequence	
[0310]	<400>	34	
[0311]		catcatcatc atcatcacgg atccatggac aagaagtaca gcatcggcct ggccatcggc	60
[0312]		accaactctg tgggctgggc cgtgatcacc gacgagtaca aggtgcccag caagaaattc	120
[0313]		aaggctgctgg gcaacaccga ccggcacagc atcaagaaga acctgatcgg agccctgctg	180
[0314]		ttcgacagcg gcgaaacagc cgaggccacc cggctgaaga gaaccgccag aagaagatac	240
[0315]		accagacgga agaaccggat ctgctatctg caagagatct tcagcaacga gatggccaag	300
[0316]		gtggacgaca gtttctcca cagactggaa gagtccttcc tgggtggaaga ggataagaag	360
[0317]		cacgagcggc accccatctt cggcaacatc gtggacgagg tggcctacca cgagaagtac	420
[0318]		cccaccatct accacctgag aaagaaactg gtggacagca ccgacaagge cgacctgcgg	480
[0319]		ctgatctatc tggccctggc ccacatgac aagttccggg gccacttctc gatcgagggc	540
[0320]		gacctgaacc ccgacaacag cgacgtggac aagctgttca tccagctggt gcagacctac	600
[0321]		aaccagctgt tcgaggaaaa ccccatcaac gccagcggcg tggacgcaa ggccatcctg	660
[0322]		tctgccagac tgagcaagag cagacggctg gaaaatctga tcgcccaget gcccggcgag	720
[0323]		aagaagaatg gcctgttcgg caacctgatt gccctgagcc tgggcctgac cccaacttc	780
[0324]		aagagcaact tcgacctggc cgaggatgcc aaactgcagc tgagcaagga cacctacgac	840
[0325]		gacgacctgg acaacctgct ggcccagatc ggcgaccagt acgccgacct gtttctggcc	900
[0326]		gccaagaacc tgtccgacgc catcctgctg agcgacatcc tgagagtga caccgagatc	960
[0327]		accaaggccc cctgagcgc ctctatgac aagagatac acgagacca ccaggacctg	1020
[0328]		accctgctga aagctctcgt gcggcagcag ctgcctgaga agtacaaga gattttcttc	1080
[0329]		gaccagagca agaacggcta cgccggctac attgacggcg gagccagcca ggaagagttc	1140
[0330]		tacaagttca tcaagcccat cctggaaaag atggacggca ccgaggaact gctcgtgaag	1200
[0331]		ctgaacagag aggacctgct gcggaagcag cggaccttcg acaacggcag cateccccac	1260
[0332]		cgatccacc tgggagagct gcacgccatt ctgcggcggc aggaagattt ttaccattc	1320
[0333]		ctgaaggaca accgggaaaa gatcgagaag atcctgacct tccgcatccc ctactacgtg	1380
[0334]		ggccctctgg ccaggggaaa cagcagattc gcctggatga ccagaaagag cgaggaaacc	1440
[0335]		atcacccctt ggaacttcga ggaagtgggt gacaaggcgc cttccgcca gagcttcatc	1500

[0336]	gagcggatga ccaacttoga taagaacctg cccaacgaga aggtgctgcc caagcacagc	1560
[0337]	ctgtgtacg agtacttcac cgtgtataac gagctgacca aagtgaaata cgtgaccgag	1620
[0338]	ggaatgagaa agccccctt cctgagcggc gagcagaaaa aggccatcgt ggacctgctg	1680
[0339]	ttcaagacca accgaaagt gaccgtgaag cagctgaaag aggactactt caagaaaatc	1740
[0340]	gagtgttcg actccgtgga aatctccggc gtggaagatc ggttcaacgc ctccctgggc	1800
[0341]	acataccacg atctgtgaa aattatcaag gacaaggact tcctggacaa tgaggaaaac	1860
[0342]	gaggacattc tggaagatat cgtgtgacc ctgacactgt ttgaggacag agagatgatc	1920
[0343]	gaggaacggc tgaaaaccta tgcccactg ttcgacgaca aagtgatgaa gcagctgaag	1980
[0344]	cggcggagat acaccggctg gggcaggctg agccggaagc tgatcaacgg catccgggac	2040
[0345]	aagcagtccg gcaagacaat cctggatttc ctgaagtcg acggcttcgc caacagaaac	2100
[0346]	ttcatgcagc tgatccacga cgacagcctg acctttaaag aggacatcca gaaagcccag	2160
[0347]	gtgtccggcc agggcgatag cctgcacgag cacattgcca atctggccgg cagccccgcc	2220
[0348]	attaagaagg gcatctgca gacagtgaag gtggtggacg agctcgtgaa agtgatgggc	2280
[0349]	cggcacaagc ccgagaacat cgtgatcga atggccagag agaaccagac caccagaag	2340
[0350]	ggacagaaga acagccgca gagaatgaag cggatcgaag agggcatcaa agagctgggc	2400
[0351]	agccagatcc tgaagaaca ccccgaggaa aacaccagc tgcagaacga gaagctgtac	2460
[0352]	ctgtactacc tgcagaatgg gcgggatatg tacgtggacc aggaactgga catcaaccgg	2520
[0353]	ctgtccgact acgatgtgga ccacatcgtg cctcagagct ttctgaagga cgactccatc	2580
[0354]	gacaacaagg tgctgaccag aagcgacaag gcccggggca agagcgacaa cgtgccctcc	2640
[0355]	gaagaggtcg tgaagaagat gaagaactac tggcggcagc tgctgaacgc caagctgatt	2700
[0356]	accagagaa agttcgacaa tctgaccaag gccgagagag gcggcctgag cgaactggat	2760
[0357]	aaggccggct tcatcaagag acagctggtg gaaaccggc agatcacaaa gcacgtggca	2820
[0358]	cagatcctgg actcccgat gaactaag tacgacgaga atgacaagct gatccgggaa	2880
[0359]	gtgaaagtga tcacctgaa gtccaagctg gtgtccgatt tccggaagga tttccagttt	2940
[0360]	tacaaagtgc gcgagatcaa caactaccac cacgcccacg acgcctacct gaacgcctc	3000
[0361]	gtgggaaccg ccctgatcaa aaagtaccct aagctggaaa gcgagttcgt gtacggcgac	3060
[0362]	tacaaggtgt acgacgtcg gaagatgatc gccaaagcgc agcaggaaat cggcaaggct	3120
[0363]	accgcaagt acttcttcta cagcaacatc atgaactttt tcaagaccga gattaccctg	3180
[0364]	gccaacggcg agatccggaa gcggcctctg atcgagacaa acggcgaaac cggggagatc	3240
[0365]	gtgtgggata agggccggga ttttgccacc gtgcggaaag tgctgagcat gcccgaagt	3300
[0366]	aatatcgtga aaaagaccga ggtgcagaca ggcggcttca gcaaagagtc tatcctgcc	3360
[0367]	aagaggaaca gcgataagct gatcgccaga aagaaggact gggaccctaa gaagtacggc	3420
[0368]	ggcttcgaca gccccaccgt ggcctattct gtgctggtgg tggccaaagt ggaaaagggc	3480
[0369]	aagtccaaga aactgaagag tgtgaaagag ctgctgggga tcaccatcat ggaaagaagc	3540
[0370]	agcttcgaga agaatcccat cgactttctg gaagccaagg tgaaccccga caacagctga	3600
[0371]	tcatcaagct gcctaagtac tccctgttcg agctggaaaa cggccggaag agaatgctgg	3660
[0372]	cctctgccgg cgaactgcag aagggaaacg aactggcctt gccctccaaa tatgtgaact	3720
[0373]	tectgtacct ggccagccac tatgagaagc tgaagggctc ccccaggat aatgagcaga	3780
[0374]	aacagctgtt tgtggaacag cacaagcact acctggacga gatcatcgag cagatcagc	3840
[0375]	agttctcaa gagagtgatc ctggccgacg ctaatctgga caaagtgctg tccgcctaca	3900
[0376]	acaagcaccg ggataagccc atcagagagc aggccgagaa tatcatccac ctgtttacc	3960
[0377]	tgaccaatct gggagcccct gccgccttca agtactttga caccaccatc gaccggaaga	4020

---

[0378]	ggtacaccag caccaaagag gtgctggacg ccacctgat ccaccagagc atcaccggcc	4080
[0379]	tgtacgagac acggatcgac ctgtctcagc tgggaggcga cgattacaag gatgacgatg	4140
[0380]	acaag	4145

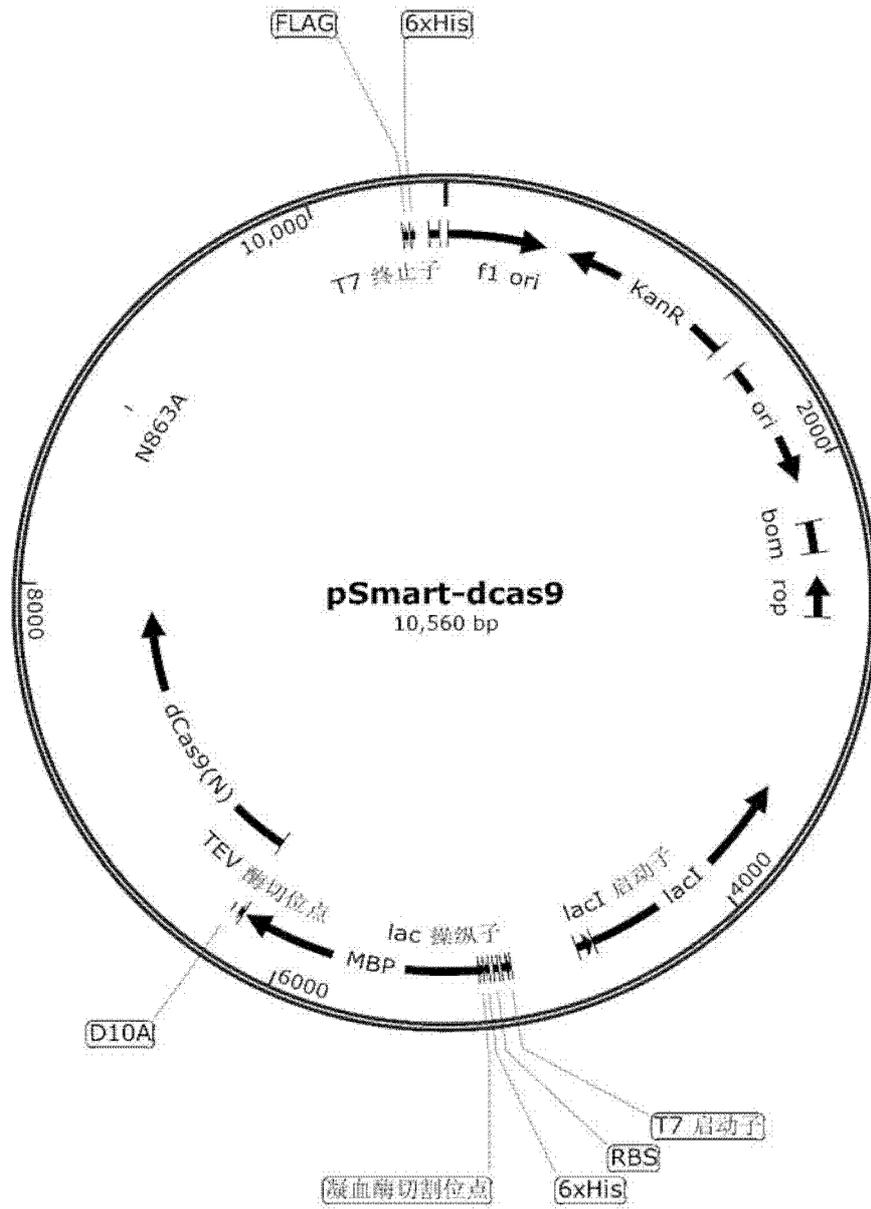


图1

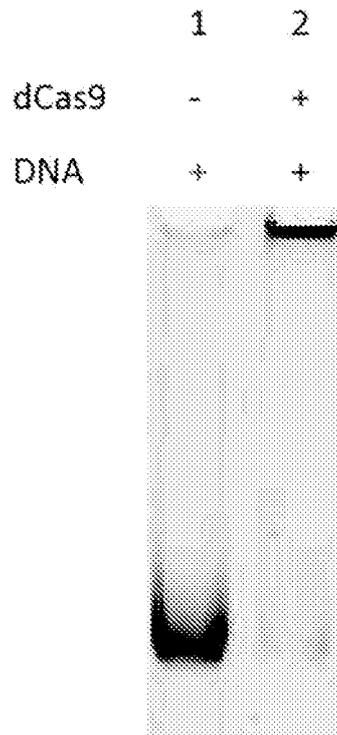


图2