



(12)发明专利申请

(10)申请公布号 CN 107208086 A

(43)申请公布日 2017.09.26

(21)申请号 201580069250.3

(22)申请日 2015.10.16

(30)优先权数据

62/065,602 2014.10.17 US

(85)PCT国际申请进入国家阶段日

2017.06.16

(86)PCT国际申请的申请数据

PCT/US2015/056048 2015.10.16

(87)PCT国际申请的公布数据

W02016/061523 EN 2016.04.21

(71)申请人 霍华德休斯医学研究所

地址 美国马里兰州

(72)发明人 罗伯特·辛格 邓伍兰

蒂莫西·莱昂尼特

(74)专利代理机构 北京路浩知识产权代理有限公司 11002

代理人 王朋飞 张晶

(51)Int.Cl.

G12N 15/01(2006.01)

权利要求书5页 说明书29页
序列表36页 附图11页

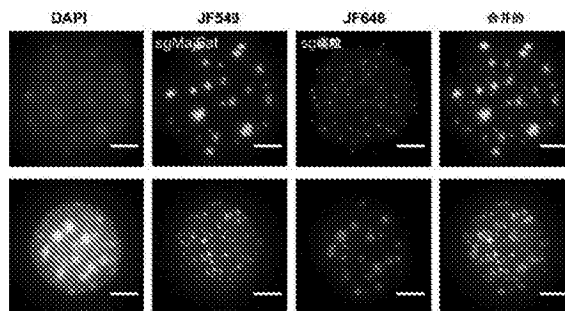
(54)发明名称

基因组探针

(57)摘要

本发明公开了标记的探针及其使用方法。标记的探针包含与对靶核酸序列包括基因组DNA序列具有特异性的gRNA缀合的Cas多肽。该探针和方法可用于标记核酸序列而无需总体DNA变性。本发明公开的主题满足部分或全部上述已确定的需求,在研究了本文中提供的信息之后,这对本领域普通技术人员而言将会是显而易见的。

着丝粒和近着丝粒区的一步双色CASFISH



1. 探针,其包含:Cas多肽,与所述Cas多肽复合的、对靶核酸序列具有特异性的gRNA,以及与所述Cas多肽、所述gRNA或两者结合的一个或多个标记物。

2. 根据权利要求1所述的探针,其中所述Cas多肽选自Cas1、Cas1B、Cas2、Cas3、Cas4、Cas5、Cas6、Cas7、Cas8、Cas9、Cas10、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和Cpf1多肽。

3. 根据权利要求1或2所述的探针,其中所述Cas多肽是核酸酶缺陷的或几乎没有核酸酶活性。

4. 根据权利要求1所述的探针,其中所述Cas多肽包括Cas9多肽。

5. 根据权利要求1所述的探针,其中所述Cas多肽包括dmCas9多肽。

6. 根据权利要求5所述的探针,其中所述dmCas9多肽由包含SEQ ID NO.1序列的核苷酸编码,或所述dmCas9多肽包含SEQ ID NO.2序列。

7. 根据权利要求1或2所述的探针,其还包含核定位信号(NLS)。

8. 根据权利要求7所述的探针,其中所述Cas多肽由包含SEQ ID NO.3序列的核苷酸编码,或所述Cas多肽包含SEQ ID NO.4序列。

9. 根据权利要求1所述的探针,其还包含第二Cas多肽。

10. 根据权利要求9所述的探针,其中融合蛋白包含所述Cas多肽和所述第二Cas多肽。

11. 根据权利要求1所述的探针,其中所述Cas多肽为Cas多肽的裂解形式。

12. 根据权利要求1所述的探针,其中标签与所述Cas多肽结合。

13. 根据权利要求12所述的探针,其中所述标签有助于所述探针的检测。

14. 根据权利要求12所述的探针,其中所述标签促进所述一个或多个标记物与所述Cas多肽的连接。

15. 根据权利要求12所述的探针,其中所述Cas多肽被表达为具有标签的融合蛋白。

16. 根据权利要求12-15中任一项所述的探针,其中所述标签选自HaloTag、多组氨酸标签和醛标签。

17. 根据权利要求12-15中任一项所述的探针,其中所述标签为HaloTag。

18. 根据权利要求12所述的探针,其中HaloTag和多组氨酸标签结合至所述Cas多肽。

19. 根据权利要求1所述的探针,其中所述Cas多肽被表达为具有HaloTag和多组氨酸标签的融合蛋白。

20. 根据权利要求16、18和19中任一项所述的探针,其中所述多组氨酸标签为六组氨酸标签。

21. 根据权利要求1所述的探针,其中所述Cas多肽被表达为具有HaloTag的融合蛋白。

22. 根据权利要求21所述的探针,其中所述一个或多个标记物经由所述HaloTag与所述Cas多肽结合。

23. 根据权利要求22所述的探针,其中所述gRNA不包括所述一个或多个标记物中的任何一种。

24. 根据权利要求1所述的探针,其中所述gRNA包含单个RNA分子。

25. 根据权利要求1所述的探针,其中所述gRNA包含多个RNA分子。

26. 根据权利要求1所述的探针,其中所述gRNA靶向所述靶核酸序列并与其杂交。

27. 根据权利要求1所述的探针,其中所述靶核酸序列选自DNA序列、RNA序列或另一种核酸类似物的序列。

28. 根据权利要求1所述的探针,其中所述靶核酸序列为其中一个或多个残基被修饰的核酸分子。

29. 根据权利要求1所述的探针,其中所述靶核酸序列为未变性的天然DNA。

30. 根据权利要求1所述的探针,其中所述靶核酸序列在尚未变性的染色体DNA内。

31. 根据权利要求1所述的探针,其中所述靶核酸序列在含有遗传物质的样品中。

32. 根据权利要求31所述的探针,其中所述样品为细胞。

33. 根据权利要求32所述的探针,其中所述细胞为活细胞。

34. 根据权利要求32所述的探针,其中所述细胞为固定细胞。

35. 根据权利要求32所述的探针,其中所述细胞为植物细胞或动物细胞。

36. 根据权利要求23所述的探针,其中所述细胞为人细胞。

37. 根据权利要求31所述的探针,其中所述样品包含病毒、细菌或真菌。

38. 根据权利要求31所述的探针,其中所述样品为选自以下的生物样品:用于遗传测试的染色体涂片、培养物、产前材料、用于体外受精的样品、拭子、空气过滤器、水冷却塔、食品、饮料、头发、粪便、尿液、唾液、血液、淋巴、痰液、子宫颈涂片、精子、活检组织切片、尸体解剖组织切片及其组合。

39. 根据权利要求1所述的探针,其中所述Cas多肽与所述一个或多个标记物结合。

40. 根据权利要求1所述的探针,其中所述gRNA与所述一个或多个标记物结合。

41. 根据权利要求39或40所述的探针,其中所述一个或多个标记物选自荧光标记物和荧光团、染料、量子点、金颗粒、放射性标记物、磁性颗粒、酶、催化剂和分光光度标记物。

42. 根据权利要求39或40所述的探针,其中所述一个或多个标记物选自Alexa 488、DY547、Cy5、JF549和JF646。

43. 根据权利要求42所述的探针,其包含多个标记物,其中每个标记物具有不同的发光颜色。

44. 根据权利要求1所述的探针,其中所述一个或多个标记物包含用于随后的二次标记的标签。

45. 根据权利要求44所述的探针,其中所述标签包含抗体靶标。

46. 根据权利要求1所述的探针,其中所述探针与靶核酸序列结合。

47. 根据权利要求1所述的探针,其还包含对一个或多个其他靶核酸序列具有特异性的一个或多个其他gRNA,其中每个gRNA与所述Cas多肽复合。

48. 根据权利要求47所述的探针,其中多个标记物与Cas多肽、gRNA或两者结合。

49. 根据权利要求48所述的探针,其中多个标记物与所述Cas多肽结合。

50. 根据权利要求49所述的探针,其中每个标记物具有不同的发光颜色。

51. 根据权利要求50所述的探针,其中所述标记物选自Alexa 488、DY547、Cy5、JF549和JF646。

52. 根据权利要求50所述的探针,其中每个标记物经由HaloTag与Cas多肽结合。

53. 根据权利要求51所述的探针,其中所述探针与每个gRNA的靶核酸序列结合。

54. 根据权利要求1所述的探针,其还包含含有一个或多个其他gRNA的一个或多个其他

探针,其中每个gRNA与Cas多肽复合,每个gRNA对不同的靶核酸序列具有特异性,并且每个探针包括具有不同发光颜色的标记物。

55. 根据权利要求54所述的探针,其中所述标记物选自Alexa 488、DY547、Cy5、JF549和JF646。

56. 根据权利要求54所述的探针,其中每个标记物经由HaloTag与Cas多肽结合。

57. 根据权利要求51或55所述的探针,其中所述Cas多肽包含Cas9多肽。

58. 根据权利要求51或55所述的探针,其中所述Cas多肽包含dmCas9多肽。

59. 根据权利要求58所述的探针,其中所述dmCas9多肽由包含SEQ ID NO.1序列的核苷酸编码或所述dmCas9多肽包含SEQ ID.2序列。

60. 根据权利要求57所述的探针,其还包含核定位信号(NLS)。

61. 根据权利要求60所述的探针,其中所述Cas多肽由包含SEQ ID NO.3序列的核苷酸编码或所述Cas多肽包含SEQ ID.4序列。

62. 根据权利要求1所述的探针,其中在体外组装探针。

63. 一系列探针,其中每个探针包含与Cas多肽复合的gRNA,每个gRNA对不同的靶核酸序列具有特异性,并且每个探针包括具有不同发光颜色的标记物,其中每个探针在体外组装。

64. 试剂盒,其包含一个或多个权利要求1-63中任一项所述的探针。

65. 试剂盒,其包含一个或多个权利要求1所述的探针。

66. 试剂盒,其包含:Cas多肽;以及能够结合至Cas多肽的一种或多种标记物,用于与所述Cas多肽复合的、对靶核酸序列具有特异性的gRNA,或两者。

67. 根据权利要求66所述的试剂盒,其中所述标记物包含用于随后的二次标记的标签。

68. 根据权利要求67所述的试剂盒,其中所述标签包含抗体靶标。

69. 根据权利要求68所述的试剂盒,其还包含针对抗体靶标的抗体。

70. 根据权利要求65-69中任一项所述的试剂盒,其还包含用于制备含有靶核酸序列的样品的试剂。

71. 根据权利要求70所述的试剂盒,其还包含用于固定所述样品的试剂。

72. 根据权利要求65-69中任一项所述的试剂盒,其还包含对一个或多个其他靶核酸序列具有特异性的一个或多个其他gRNA,其中每个gRNA均可与所述Cas多肽复合,使得能够组装包含与单个Cas复合的多个gRNA的Cas/gRNA复合物,或者能够组装包含单个gRNA和单个Cas的多个Cas/gRNA复合物。

73. 根据权利要求70所述的试剂盒,其还包含多个标记物,其中每个标记物具有不同的发光颜色。

74. 根据权利要求73所述的试剂盒,其中所述标记物选自Alexa 488、DY547、Cy5、JF549和JF646。

75. 根据权利要求74所述的试剂盒,其中所述Cas多肽包含Cas9多肽。

76. 根据权利要求75所述的试剂盒,其中所述Cas多肽包括dmCas9多肽。

77. 根据权利要求76所述的试剂盒,其中所述dmCas9多肽由包含SEQ ID NO.1序列的核苷酸编码或所述dmCas9多肽包含SEQ ID.2序列。

78. 根据权利要求75所述的试剂盒,其还包含核定位信号(NLS)。

79. 根据权利要求76所述的试剂盒,其中所述Cas多肽由包含SEQ ID NO.3序列的核苷酸编码或所述Cas多肽包含SEQ ID NO.4序列。

80. 检测靶核酸序列的方法,其包括:

使含有遗传物质的样品与权利要求1所述的探针接触;以及

检测所述探针是否结合所述靶核酸序列,其中探针与核酸的结合表明样品中存在靶核酸序列。

81. 根据权利要求80所述的方法,其还包括对所述样品成像以确定所述靶核酸序列的位置和/或相对拷贝数。

82. 根据权利要求80所述的方法,其中所述探针还包含对一个或多个其他靶核酸序列具有特异性的一个或多个其他gRNA,其中每个gRNA与所述Cas多肽复合;检测所述探针是否结合每一个靶核酸序列,其中探针与核酸的结合表明样品中存在靶核酸序列。

83. 根据权利要求80所述的方法,其还包括一个或多个权利要求1所述的其他探针,每个探针包含与Cas多肽复合的gRNA,每个gRNA对不同的靶核酸序列具有特异性,并且每个探针包含具有不同发光颜色的标记物。

84. 根据权利要求83所述的方法,其还包括同时检测所述探针是否与所述多个靶核酸序列中的每一个结合。

85. 根据权利要求83所述的方法,其还包括对样本进行成像以确定多个靶核酸序列中每一个的位置和/或相对拷贝数。

86. 根据权利要求80-85中任一项所述的方法,其中所述标记物选自Alexa 488、DY547、Cy5、JF549和JF646。

87. 根据权利要求83所述的方法,其中每个标记物经由HaloTag与每个Cas多肽结合。

88. 根据权利要求80-85中任一项所述的方法,其中所述样品为选自以下的生物样品:用于遗传测试的染色体涂片、培养物、产前材料、用于体外受精的样品、拭子、空气过滤器、水冷却塔、食品、饮料、头发、粪便、尿液、唾液、血液、淋巴、痰液、子宫颈涂片、精子、活检组织切片、尸体解剖组织切片及其组合。

89. 根据权利要求80-85中任一项所述的方法,其中所述样品包括细胞,并且其中所述接触步骤通过选自微珠负载、显微注射、纳米颗粒或脂质介导的转导及其组合的方式进行。

90. 根据权利要求80-85中任一项所述的方法,其中所述靶核酸序列与疾病或病症相关,所述样品来自个体,并且其中所述方法还包括如果在样品中检测到靶核酸序列,则确定个体罹患所述疾病或病症的可能性增加。

91. 根据权利要求90所述的方法,其还包括固定所述样品。

92. 根据权利要求80-85中任一项所述的方法,其中所述靶核酸序列为未变性的天然DNA。

93. 根据权利要求80-85中任一项所述的方法,其中所述靶核酸序列在尚未变性的染色体DNA内。

94. 根据权利要求80-85中任一项所述的方法,其中所述靶核酸序列在含有遗传物质的样品中。

95. 根据权利要求80-85中任一项所述的方法,其中所述样品为细胞。

96. 根据权利要求95所述的方法,其中所述细胞为活细胞。

97. 根据权利要求95所述的方法,其中所述细胞为固定细胞。
98. 根据权利要求95所述的方法,其中所述细胞为植物细胞或动物细胞。
99. 根据权利要求95所述的方法,其中所述细胞为人细胞。
100. 权利要求1所述的探针用于检测靶核酸序列的应用。
101. 体外组装探针的方法,其包括:
- 选择能够靶向目标核酸序列的gRNA,其任选地为标记的gRNA;
- 提供任选地标记的Cas多肽,其中gRNA和Cas多肽中的至少一个被标记;
- 混合并孵育所述gRNA和所述Cas多肽。

基因组探针

[0001] 相关申请

[0002] 本申请要求2014年10月17日提交的第62/065,602号美国临时申请的优先权,其全部公开内容通过引用并入本文。

技术领域

[0003] 本文公开的主题涉及基因组探针。具体地,本文公开的主题涉及包含Cas多肽的探针,该CAS多肽与靶核酸序列包括基因组DNA序列具有特异性的gRNA复合。

背景技术

[0004] 针对核酸序列的探针用于各种目的。这类探针可用于鉴定在DNA或RNA样品上能否发现特定序列,并鉴定特定序列的拷贝数。特定序列可以是与罹患某些疾病或病症的可能性相关的序列,与某些疾病或病症的阶段相关的序列,来自感染病毒、细菌和其他感染微生物的序列,编码某些基因的序列,重复序列等。这类探针可用于检测多个序列的相对位置,例如在各种癌症中频繁检测到的基因易位。因此,探测细胞的DNA可以提供可用于预测、诊断、医学和物种识别目的的有价值的信息。

[0005] 衍生自化脓链球菌 (*Streptococcus pyogenes*) 的II型成簇地规律间隔的短回文重复序列 (CRISPR) -CRISPR相关半胱天冬酶9 (Cas9) 系统已经成为靶向基因组编辑的革命性工具,其核酸酶缺陷型衍生物 (dCas9) 也分别通过与转录调控结构域或荧光蛋白融合用于控制活细胞中基因表达和基因座的可视化。CRISPR系统提供的固有多路复用功能在高通量测定中的应用前景十分广阔。通过编程gRNA序列,Cas9蛋白可以被引导到感兴趣的任何目标DNA序列。然而,因为在一个细胞中缺乏对数十个到数百个gRNA的有效转导,因为活细胞中Cas9和gRNA的组装依赖随机碰撞,并且因为一种类型的Cas蛋白只能用一种颜色标记来成像一组靶DNA序列,所以将基因编码的Cas9蛋白和gRNA用于在活细胞中成像基因组DNA仍然受到极大限制。此外,Cas9和gRNA的基因编码表达需要将遗传物质递送至培养的活细胞,这通常不适用于原代细胞和组织。

[0006] 荧光原位杂交 (FISH) 目前用于鉴定固定细胞和组织中的特异性DNA或RNA序列。FISH通常使用甲酰胺化学和热处理来进行DNA变性。FISH包括样品变性、杂交和杂交后洗涤的步骤,这需要花费数小时至数天。(Lawrence JB,Villnave CA,Singer RH.Cell.1988Jan 15;52 (1) :51-61.PMID:2830981;Lawrence JB,Singer RH.Nucleic Acids Res.1985Mar 11;13 (5) :1777-99.PMID:3889842;Cremer T.Landegent J.Brueckner A,Scholl HP,Schardin M,Hager HD,Devilee P.Pearson P.van der Ploeg M. (1986) ,"Detection of chromosome aberrations in the human interphase nucleus by visualization of specific target DNAs with radioactive and non-radioactive in situ hybridization techniques:diagnosis of trisomy 18with probe L.84,"Hum.Genet 74:346-352;Pinkel D,Straume T.Gray JW. (1986) "Cytogenetic analysis using quantitative,high-sensitivity,fluorescence hybridization,"Proc Natl Acad Sci

USA 83:2934-2938)。

[0007] 尽管持续改进,DNA FISH需要诸如加热和甲酰胺的苛刻处理来使dsDNA变性以允许探针杂交,从而产生影响生物结构和基因组组织完整性的风险。在细胞中,基因组DNA在三维上被高度折叠和构造。染色质在细胞核中的空间构造和不同染色质区域的相对位置与正常发育和疾病中的基因调控密切相关。虽然使用处于其天然状态的DNA研究间期基因组构造可能是有益的,但染色质构造的研究不可能使用使天然染色体变性的FISH方法实现。

[0008] DNA FISH也在检测分辨率方面受到BAC探针的限制,并且由于寡核苷酸探针的高成本而受限制。目前的FISH方案需要对样品进行变性并对所有单独的DNA探针进行荧光标记,这增加了当前FISH成像方法的复杂性和成本。DNA FISH的冗长程序也阻碍了其在研究和临床中的广泛应用。

[0009] 因此,仍然需要用于成像核酸序列(包括DNA)且不需要使样品变性的组合物和方法,其包括用于成像天然状态的基因组DNA的组合物和方法。还需要成像探针和可以快速操作并且不需要对样品进行长时间和苛刻处理的方法。

发明内容

[0010] 本文公开的主题满足一些或所有上述需求,在研究了本文中提供的信息后,这对本领域普通技术人员而言将是显而易见的。

[0011] 本发明内容描述了本文公开的主题的一些实施方案,并且在许多情况下列出了这些实施方案的变化和变换。本发明内容仅仅是众多不同实施方案的示例。对给定实施方案的一个或多个代表性特征的提及同样是示例性的。这样的实施方案通常可以具有或不具有所提及的特征;同样,这些特征可以应用于本文公开的主题的其他实施方案,而无论其是否在本发明内容中列出。为了避免过度重复,此发明内容未列出或显示所有这些特征的可能的组合。

[0012] 本文提供了用于检测固定细胞和组织的未干扰的细胞核内的核酸序列的技术。检测目标可以是DNA、RNA和其他核酸类似物或经修饰的核酸。在一个示例性实施方案中,将来自细菌CRISPR(成簇地规律间隔的短回文重复序列)系统的蛋白质与作为探针的RNA序列组合以在完整基因组中找到目标基因。这种方法保持遗传因子的空间关系,这对于了解基因表达和检测基因易位是重要的;此外,该方法非常快速(15分钟)、方便,可直接用于组织以诊断疾病。

[0013] 本文公开的主题包括探针。探针包括Cas(CRISPR相关半胱天冬酶)多肽或Cas多肽的片段和/或变体,与Cas多肽缀合的、对预定核酸序列具有特异性的引导RNA(gRNA);以及与Cas多肽、gRNA或两者结合的标记物。

[0014] 在一些实施方案中,所述Cas多肽与所述标记物结合或融合。在某些情况下,标记物可以是荧光蛋白、荧光标记物、染料、量子点、金颗粒、放射性标记物、化学发光酶,或修饰底物以使其通过显微镜可见的酶,或用于显微镜或其他检测方法的其他分析物。

[0015] 一些优选实施方案中的Cas多肽包括Cas9多肽或Cas9多肽的片段或变体。在一些实施方案中,Cas9多肽包括核酸酶缺陷型dmCas9多肽、片段和/或变体。当Cas9多肽是dmCas9多肽时,在一些实施方案中,dmCas9多肽由SEQ ID NO.1的核苷酸序列编码,或者dmCas9多肽具有SEQ ID NO.2序列。在一些实施方案中,Cas多肽包括CRISPR I、II、III系统

中的其他CRISPR相关蛋白,如本领域普通技术人员所知晓的。

[0016] 在一些情况下,探针不具有核酸酶活性。在其他情况下,探针可以具有很少的核酸酶活性、降低的核酸酶活性或完全的野生型核酸酶活性。

[0017] 在一些实施方案中,多肽包括核定位信号(NLS)。NLS可以帮助多肽被细胞核结合。包含NLS的多肽的一个实例由SEQ ID NO.3的核苷酸序列编码,或该多肽具有SEQ ID NO.4序列。

[0018] 在一些实施方案中,所述gRNA与所述标记物结合或融合。标记物在某些情况下可以是荧光标记物、染料、量子点、金颗粒、放射性标记物、可被直接或间接标记的RNA适体、酶或用于显微镜或其他检测方法的其他分析物。

[0019] 在一些实施方案中,探针包含含有恒定主链序列和对预定核酸(其为DNA或RNA)具有特异性的可变序列的gRNA。gRNA通常包含靶区段,其中靶区段与预定核酸序列具有完全或部分互补性。在一些实施方案中,gRNA的靶区段长度为约15个至约25个核苷酸。

[0020] gRNA的主链序列含有用于与Cas蛋白相互作用的发夹结构。在一些情况下,可以修饰gRNA主链的长度和核酸组成以具有更长或更短、更动态或更稳定的发夹结构。在其他情况下,可以修饰gRNA以获得其他特定性质。

[0021] 在一些情况下,gRNA可以被两个或更多个RNA分子取代,该RNA分子能够形成相互作用,并且保持与Cas蛋白结合以及与核酸底物杂交的功能。

[0022] 在一些实施方案中,对靶核酸的检测是由Cas蛋白或gRNA上的标记物实现的。在某些情况下,依赖于DNA上的Cas/gRNA复合物的结合或核酸酶活性的次级反应可以产生检测信号。

[0023] 本文还公开了用于成像核酸的方法,其包括提供包含核酸的样品的第一步骤;将样品与本文公开的任何实施方案的探针接触的第二步骤,以及随后的检测探针是否结合核酸的步骤。探针与核酸的结合表明在核酸样品中存在预定的核酸序列。

[0024] 在一些情况下,样品选自用于遗传测试的染色体涂片(spread)、培养物、产前材料、用于体外受精的样品、拭子、空气过滤器、水冷却塔、食品、饮料、头发、粪便、尿液、唾液、血液、淋巴、痰液、子宫颈涂片、精子、活检组织切片、尸体解剖组织切片及其组合。任何遗传物质的来源均可预期用于本文公开的主题。人类、动物、植物、细菌、真菌和其他微生物是遗传物质来源的一些实例。可以不对样品进行处理,或使用非交联固定剂(如甲醇、乙醇、丙酮、乙酸或其他物质)或交联固定剂(如甲醛、多聚甲醛、福尔马林或其他物质)对样品进行处理。

[0025] 在样品包括活细胞的实施方案中,接触步骤可以通过选自微珠负载(bead loading)、显微注射、纳米颗粒或脂质介导的转导及其组合的方式进行。

[0026] 本方法可以包括使样品与两个或更多个探针接触,每个探针具有不同的标记物,并且每个探针对于不同的预定核酸序列具有选择性。这些方法使得能够检测多个核酸序列。

[0027] 在一些实施方案中,预定的核酸序列可以与疾病或病症相关,或与罹患疾病或病症的可能性相对增加相关,并且本文公开的方法可以包括如果在来自个体的核酸样品中检测到预定的核酸序列,则诊断或预测个体患有所述疾病或病症。

[0028] 本文也公开了试剂盒。试剂盒可以包括探针,该探针包含Cas多肽或其片段和/或

变体;能够与Cas多肽缀合的、对预定核酸序列具有特异性的gRNA;以及能够结合该Cas多肽或gRNA或两者的标记物。在一些实施方案中,标记物与Cas多肽结合。在一些实施方案中,Cas9多肽包括dmCas9多肽,或其片段和/或变体。试剂盒可以包括两个或更多个探针,每个探针具有不同的标记物,并且每个探针对不同的预定核酸序列具有选择性。可以根据研究者的需要调整gRNA和探针,所述需要包括例如所需的成像、预定的核酸序列以及将在其上使用探针的样品的类型(例如组织样品、固定样品、未固定样品)。

附图说明

[0029] 图1包括针对近着丝粒区、端粒区和着丝粒的CASFISH示意图,以及细胞中基因组DNA中体外组装的dCas9/sgRNA荧光标记的图像。(A) CASFISH策略示意图。(B,上)在鼠染色体上的所示DNA元件的相对位置。(B,下)所示sgRNA序列。sgMajSat: TTTCTTGCCATATTCCACGTCCTACAGTGG (SEQ ID NO.5);sg端粒: TTAGGGTTAGGGTTAGGGTTAGGGTTAGGG (SEQ ID NO.6);sgMinSat: ACTCATCTAATATGTTCTACAGTGTGG (SEQ ID NO.7)。(C)使用在体外组装的所示荧光dCas9/sgMajSat探针的针对MEF中近着丝粒区的CASFISH。用JF646标记dCas9,用DY547标记sgMajSat。用指定的滤光片设置拍摄荧光图像,并进行伪着色以达到可视化目的。(D,左)使用所示荧光dCas9/sgTelomere复合物的针对MEF中的端粒的CASFISH。(D,右)每个细胞中检测到的端粒数的直方图;n=134个细胞。(E)针对着丝粒中的小卫星元件的MEF细胞中的CASFISH。显示了z-堆栈的最大投影。(比例尺,5 μ M。)

[0030] 图2包括使用dCas9 sgRNA复合物的图像和示意图,包括活细胞成像(A)表达dCas9-Halo和sgMajSat的MEF中的近着丝粒区的活细胞成像。(B)sgMajSat-2序列和针对mEF中sgMajSat-2的CASFISH。(C)与sgMajSatCASFISH共染色的近着丝粒区中的抗HP1 α 免疫荧光染色。(D)使用指定试剂的CASFISH对照显示近着丝粒区的检测需要由dCas9蛋白和全长sgRNA形成的复合物。显示了z-堆栈的最大投影。(比例尺,5 μ M。)

[0031] 图3包括使用Cy5标记的dCas9蛋白的CASFISH的示意图和图像。(A)DY547标记的全长sgMajSat检测近着丝粒区。(B)未标记的sgRNA的所示截短形式或不存在的sgRNA未能检测到近着丝粒区。显示了z-堆栈的最大投影。(比例尺,5 μ M。)

[0032] 图4包括基于MEF细胞系的MBS结合位点的CASFISH的图像和数据(A)对含有重复MS2结合位点(MBS)阵列的U2OS细胞系中的MBS进行的CASFISH。如代表性图中所示,所有检查的细胞均被标记。z-堆栈的最大投影由比例尺表示。(B)在本研究中使用的、G-显带中期的MEF细胞系上进行的代表性细胞遗传学分析。(C)10个MEF细胞的复合中期染色体组型。MEF细胞系在细胞中具有可变的染色体组成,平均为94.5。

[0033] 图5包括CASFISH的示意图和测定以及显示由dCas9、sgRNA和DNA形成的三元复合物的稳定性的直方图:(A)在1%琼脂糖凝胶上对使用所示荧光或未标记的试剂的反应进行的EMSA测定。用两个指明的滤光器设置扫描凝胶,绿色信道检测JF549和DY547,红色信道检测JF646和Cy5。加号的颜色表示所示的分子被荧光标记(绿色或红色)或未标记(黑色)。+,10nM;++,40nM;+++,400nM。泳道1-4显示了凝胶上单独的dCas9蛋白(泳道1和2)、单独的sgRNA(泳道3)或单独的DNA底物(泳道4)的迁移。如图所示,由dCas9和sgRNA形成的二元复合物(泳道5)以及由dCas9、sgRNA和DNA形成的三元复合物(泳道6和8)在凝胶上的不同位置

迁移。dCas9和DNA (泳道7) 不形成复合物,因为在凝胶上没有检测到新的条带。正如所预期的,在复合物形成前添加10×未标记的靶DNA (泳道10) 导致结合到三元复合物中的Cy5-DNA减小10倍。(B,左) 在溶液中组装具有Cy5标记的sgMajSat靶DNA的所示荧光三元复合物,并在存在或不存在30×未标记的靶DNA作为竞争剂的情况下,进一步在37°C下孵育指定的分钟。(B,右) 将对应于三元复合物的条带的Cy5荧光强度绘制为直方图。由于缺乏竞争,复合物的稳定性显而易见。误差条代表三个独立实验的标准偏差。

[0034] 图6包括双色CASFIISH示意图、测定、直方图和成像 (A) 评估三元复合物的稳定性的竞争EMSA。(A,左) 将指明的荧光三元复合物在溶液中组装,并在37°C下与指定倍数的量的、作为竞争剂的、未标记的sgMajSat靶DNA一起进一步孵育15分钟。(A,右) 将三元复合物的条带的荧光强度绘制为直方图。由于缺乏竞争,复合物的稳定性显而易见。对照泳道(ctrl) 表示由Cy5 DNA和30倍的未标记的靶DNA的混合物与dCas9/sgRNA二元复合物制备的反应产物。如预期的那样,只有一小部分Cy5-DNA被结合到dCas9三元复合物中。(B) 评估二元复合物稳定性的竞争性EMSA。(B,左) 将指定的荧光二元复合物在室温下组装10分钟,并与指定倍数的量的未标记的sgMajSat在室温下进一步孵育10分钟。随后将未标记的靶DNA加入到溶液中形成三元复合物,接着进行凝胶电泳。(B,右) 将三元复合物的条带的DY547强度绘制为直方图。dCas9和DY547-sgRNA的二元复合物的稳定性从缺乏竞争中可以看出。对照泳道(ctrl) 表示由DY547-sgMajSat和100倍未标记的sgMajSat的混合物与dCas9蛋白质形成的二元复合物的反应产物。如预期的那样,只有一小部分DY547-sgMajSat结合到dCas9三元复合物。(C) 双色CASFIISH方法的示意图。(D) 使用指定的JF549标记的dCas9蛋白或JF646标记的dCas9蛋白和未标记的sgRNA,针对MEF中近着丝粒区和端粒进行的顺序CASFIISH荧光成像。顺序CASFIISH的结果与一步法的那些结果相似。显示z-堆栈的最大投影。(比例尺,5μM。) 误差条代表三个独立实验的标准偏差。

[0035] 图7是使用指定的荧光dCas9蛋白和未标记的sgRNA通过一步法、双色CASFIISH产生的近着丝粒区和端粒的荧光成像。显示z-堆栈的最大投影。(比例尺,5μM。)

[0036] 图8包括在人细胞中的内源性编码基因的CASFIISH成像。(A) 针对MUC4基因的内含子3 (sgMUC4-I1) 或外显子2 (sgMUC4-E2) 内的两个重复DNA序列的顺序CASFIISH的荧光成像。(B) 如在A中,通过针对sgMUC4-I1 (绿色) 和sgMUC4-E2 (红色) 的顺序CASFIISH定量每个细胞的MUC4基因座的数目和染色的基因座总数。 $n=80$ 个细胞。两个探针标记的基因座都是黄色的。(C) 针对MUC4和MUC1基因的顺序CASFIISH的荧光成像。(D) 针对MUC4基因的重复内含子3 (sgMUC4-I1) 和非重复性内含子1 (sgMUC4-瓦样覆盖) 的顺序CASFIISH。白色箭头表示标记的MUC4基因。如图所示,dCas9被JF549或JF646标记。显示z-堆栈的最大投影。(比例尺,5μM。)

[0037] 图9包括与CASFIISH灵敏度相关的研究 (A) MUC4CASFIISH的代表性视野的荧光成像。所有细胞均被有效标记。显示z-堆栈的最大投影。(B) 复制细胞中MUC4标记的一个实例。测量每个斑点的荧光强度,并减去背景强度。显示z-堆栈的最大投影。(比例尺,5μM。)

[0038] 图10为组织切片上的CASFIISH成像和快速CASFIISH方案。(A) 显示针对主要卫星和端粒的小鼠冷冻脑切片的CASFIISH。(B,左) 在HeLa细胞中MUC4基因的十五分钟CASFIISH和标准CASFIISH。(B,右) 标记的基因座的强度曲线,如左图虚线所示。dCas9被JF646标记。显示z-堆栈的最大投影。(比例尺,5μM。)

[0039] 图11描绘了CASFIISH多色和多路复用 (A) dCas9-Halo用于多色和多路复用CASFIISH

的优点。dCas9-Halo可以用各种着色染料标记,并与各种靶向不同DNA序列的sgRNA复合。(B)使用单色或组合颜色编码以通过CASFISH对多个靶标成像的概念。如图所示,四种不同颜色的组合可以检测15种独特的靶标。

具体实施方式

[0040] 本文公开的主题的一个或多个实施方案的细节在本文中阐述。在对本文件中提供的信息进行研究之后,对本文中描述的实施方案的修改以及其他实施方案对于本领域普通技术人员是显而易见的。本文中提供的信息,特别是所描述的示例性实施方案的具体细节主要是为了清楚理解而提供的,不能将其理解为不必要的限制。在存在冲突的情况下,本文的具体描述(包括定义)将进行约束。

[0041] 本文公开的主题包括用于检测核酸序列的探针、试剂盒和方法。本文公开的主题的实施方案可以在需要或不需要样品的变性和/或固定的情况下操作,并且在一些实施方案中,探针可以检测基因组DNA序列。在这方面,由于本申请公开的探针的一些实施方案不需要对样品进行变性,因此利用本发明探针的成像方法可以排除通常与其他已知的成像技术相关的、耗时且破坏样品的变性步骤。这反过来可以提高本发明的探针成像核酸序列的快速性,这对于时间敏感的方法(如某些诊断、治疗或筛选方法)是特别有利的。

[0042] 本文提供了体外合成然后体外组装的Cas蛋白和gRNA分子用作可编程且序列依赖性的探针以使基因组DNA原位可见的用途。所公开的体外合成和体外组装方法使含有任何给定的用于检测的修饰物、任何给定的gRNA和任何给定数量的gRNA的Cas9蛋白能够高效而有效地组装。所有组装的Cas/gRNA复合物可以一次施用于其靶DNA。我们和其他人的CRISPR系统的体外研究表明Cas9/sgRNA对其靶DNA具有很强的和稳定的亲和力。本文还公开了一种用多种颜色成像多种靶DNA元件的方法,该方法通过将用不同颜色标记的多个探针一次施加至细胞或通过施加多轮探针反应且每两轮反应间进行洗涤步骤而实现。这些方法允许对细胞和组织中的基因组序列进行多色的、广泛的标记。

[0043] 本文提供了Cas/gRNA二元复合物作为高度特异性且高效的酶探针用于标记没有总体变性的核酸(例如DNA)的用途,所述总体变性是在DNA FISH方案中通过热或化学处理产生的。本文还公开了一种利用CRISPR/Cas系统进行核酸FISH研究的方法,其使得能够在细胞和组织中对基因组序列进行多色的、广泛的标记。

[0044] 本文公开的主题包括探针,该探针包含Cas多肽、与Cas多肽复合的gRNA。gRNA与目标靶核酸序列具有特异性。本文公开的探针的某些实施方案中所选的gRNA和Cas多肽在施用于样品之前组装在一起。在某些情况下,Cas多肽和gRNA可以同时或依序施用于样品。探针还包括与Cas多肽、gRNA或两者结合的一个或多个标记物。

[0045] 本文公开的探针的实施方案可以在体外组装。实际上,已经发现本文公开的复合探针在体外组装后是稳定的,允许将针对不同靶核酸序列的多个探针同时施用于样品,而不会显著地分解和/或重新组装复合物,而复合物的分解和/或重新组装将导致意想不到的标记物与靶序列的组合。

[0046] 术语“多肽”、“蛋白质”和“肽”在本文中可互换使用,其是指蛋白质氨基酸或氨基酸类似物的聚合物,而不管其大小或功能如何。虽然“蛋白质”通常用于指相对较大的多肽,而“肽”通常用于指小的多肽,但是本领域中这些术语的使用是重叠和变化的。除非另有说

明,本文所用的术语“多肽”是指肽、多肽和蛋白质。当指代基因产物时,术语“蛋白质”、“多肽”和“肽”在本文中可互换使用。因此,示例性多肽包括前述的基因产物、天然存在的蛋白质、同系物、直系同源物、旁系同源物、片段和其他等同物、变体和类似物。此外,术语“融合多肽”在本文中使用时通常是指由两个或多个不同多肽形成的多肽。

[0047] 当术语“多肽片段”或“片段”的使用与参照多肽有关时,其是指与参照多肽本身相比氨基酸残基缺失的多肽,但其中剩余的氨基酸序列通常与参照多肽中相应的位置相同。这种缺失可以发生在参照多肽的氨基末端、参照多肽的羧基末端、参照多肽的中间区域,或其组合。另外或替代地,一些实施方案包括替代的组合缺失。片段也可以是“功能片段”,在这种情况下,片段保留本文所述的参照多肽的一些或全部活性。例如,特定Cas多肽变体的功能性片段保留了一些或全部Cas样活性。在一些实施方案中,相对于野生型,所述Cas多肽片段和/或变体被从其C末端和/或N末端去除了多达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、47、48、49或50个残基。在一些实施方案中,可以除去其他残基。在一些实施方案中,可以去除Cas9蛋白的确定或不确定结构域的部分或全部残基,例如N末端结构域、HNH核酸酶结构域、RuvC核酸酶结构域、PAM结合结构域、C末端结构域,或者该残基可以被其他残基取代或插入其他残基。在一些实施方案中,所述Cas多肽片段和/或变体还包括对核酸酶活性有用的残基的缺失,例如HNH和RuvC核酸酶结构域中的残基的缺失。在一些实施方案中,可以修饰PAM结合结构域以改变Cas9可以识别的PAM基序。

[0048] 术语“变体”是指有一个或多个氨基酸与参照多肽不同的氨基酸序列,例如有一个或多个氨基酸取代。在一些情况下,变体包括氨基酸插入或提供使变体与参照多肽区分的多肽的标记。这些变体可以调节多肽的性质,例如调节其活性或稳定性。在一些实施方案中,例如,插入可以包括核定位信号(NLS);小表位标签,其包括但不限于Flag、HA、Myc、生物素和组氨酸;以及多肽,其包括但不限于Halo、SNAP、CLIP、GFP、YFP、RFP、mCherry和HRP。在一些实施方案中,例如,插入可以在N末端,或C末端,或内部或其组合。

[0049] 包括多肽变体和/或多肽片段在内的多肽突变体可以影响多肽的各种活性,包括但不限于多肽的酶活性、DNA识别活性、核酸酶活性和/或靶向活性。一些实施方案利用包含Cas9多肽在内的探针来介导DNA-FISH过程,在本文中可以称为“CASFISH”。

[0050] 本领域普通技术人员已知的Cas多肽实例包括但不限于Cas1、Cas1B、Cas2、Cas3、Cas4、Cas5、Cas6、Cas7、Cas8、Cas9、Cas10、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、Cpf1和不同微生物物种中的直向同源多肽,以及其片段和/或变体。Cas多肽在不同物种中具有许多直向同源基因。例如,Cas9包括但不限于化脓性链球菌Cas9(spCas9)、嗜热链球菌Cas9(St1Cas9)、金黄色葡萄球菌Cas9(SaCas9)和脑膜炎奈瑟氏球菌Cas9(Nm Cas9)。在本文公开的主题的具体实施方案中,Cas多肽包括化脓性链球菌Cas9(spCas9)及其变体和/或片段。

[0051] CRISPR/Cas系统将是本领域技术人员已知的,并且可以用在本文公开的主题的探针中,在研究本文后,这对本领域技术人员而言将会是显而易见的。例如,在一些实施方案中,I型和II型和/或II型系统可适用于本发明。

[0052] 在一些实施方案中,Cas多肽是野生型Cas多肽的变体或片段。在一些实施方案中,

Cas多肽可以被修饰以进一步包括能够调节多肽的活性和/或稳定性的插入物。

[0053] 突变体和/或变体多肽的另一个实例是Cas多肽的裂解形式。例如,裂解的Cas多肽可以包括表达为单独的多肽的核酸酶裂片和 α -螺旋裂片。突变体和/或变体多肽的另一个实例是可被工程化以识别替代PAM序列的Cas多肽。(参考:Engineered CRISPR-Cas9 nucleases with altered PAM specificities.Nature.2015 Jul 23;523(7561):481-5.doi:10.1038/nature14592.Epub 2015 Jun 22)在其他实施方案中,两种或更多种Cas多肽或其变体可以用于例如融合以增强靶向效率和/或特异性。在某些情况下,将Cas蛋白融合至裂解荧光蛋白可以提高特异性,因为只有当两个相邻的Cas/gRNA同时与DNA模板结合时,裂解的荧光蛋白才会显示荧光。一个Cas/gRNA与一半的荧光蛋白在非靶位点的结合将是暗的,从而降低噪音。在某些情况下,可以感测到与DNA的结合并触发荧光或其他活性的工程化的Cas多肽和/或gRNA可增强特异性。例如,传感器可以是Förster共振能量传递(FRET)传感器。

[0054] 在一些实施方案中,提供的Cas多肽几乎不具有核酸酶活性,其可破坏双链DNA序列。在其他实施方案中,蛋白质可以是具有完全活性的野生型蛋白质。所提供的多肽几乎不具有核酸酶活性的实施方案提供主要用作DNA序列探针的探针,该探针不切割或不改变正在成像的DNA的序列。几乎不具有核酸酶活性的多肽的实施方案可以包括例如dmCas9,其是具有相对地几乎无核酸酶活性的Cas9突变体多肽。例如,SEQ ID NO.1提供了Halo标记的dmCas9多肽的核酸序列。同样地,SEQ ID NO.2提供了SEQ ID NO.1的蛋白质序列。

[0055] 在一些实施方案中,突变体多肽可以表现出改进的靶向活性。例如,在一些实施方案中,多肽被赋予核定位信号(NLS),该核定位信号可以帮助多肽被细胞核结合。如本文所述,包括NLS的多肽的实施方案可以并入细胞核中,并且在一些情况下可以允许检测未固定和/或天然样品中的预定基因组DNA序列。例如,SEQ ID NO.3提供了被Halo标记并包括NLS的示例性Cas9多肽的核酸序列。同样地,SEQ ID NO.4提供了SEQ ID NO.3的蛋白质序列。

[0056] 本文公开的探针的实施方案还可以包括标签,提供该标签以赋予探针各种期望的特征,例如,使探针易于制备、使探针易于检测、使标记物易于与探针连接和在研究本文后对本领域普通技术人员而言将是显而易见的其他期望特征。

[0057] 在一些实施方案中,例如由于探针的荧光,通过直接观察探针上的标记物实现靶核酸的检测。在其他实施方案中,从识别探针的活性的二级反应中检测靶核酸,包括探针与靶的结合和/或探针的全部或部分核酸酶活性。例如,针对探针的抗体可以允许检测。也就是说,二次反应可以识别探针的活性,从而允许检测。因此,在一些实施方案中,标记物本身不可检测直到其在杂交后经历一个或多个反应。

[0058] 在这方面,标签可以用于随后的二次标记。探针不需要直接标记,而是可以包含标签,例如含有抗体靶标的标签。这样的标签可以用于随后的二次标记,例如,利用针对标签的抗体靶标的抗体,并且二次标记反应可以识别探针的活性,例如探针结合靶标的活性和/或探针的完全或部分核酸酶活性。因此,在一些实施方案中,探针的一个或多个标记物可以是标签。

[0059] 示例性标签还可以包括但不限于多组氨酸标签、醛标签和HaloTag™ (Promega Corporation, 麦迪逊, 威斯康星州, 美国)。其他标签SNAP、CLIP和众多荧光蛋白(如GFP、YFP、RFP、TagRFP、串联的Tomato和mCherry)以及酶HRP。这些蛋白质中的一些可能是可光激

活的或可光开关的。

[0060] 在一些实施方案中,标签与Cas多肽结合。在一些实施方案中,Cas多肽被表达为具有标签的融合蛋白。例如,在一些实施方案中,Cas多肽被表达为具有HaloTag和/或多组氨酸标签的融合蛋白。在其他实施方案中,所述Cas多肽与特定标签(例如HaloTag、多组氨酸标签和/或醛标签)结合。

[0061] 某些标签用于使用一个或多个标记物容易地标记Cas多肽。在这方面,尽管本申请公开的探针的实施方案包括标记物缀合的sgRNA探针,但是它们可能是昂贵的。然而,利用使用标签的Cas多肽标记可以是高效的且成本效益好的。HaloTag是一个实例,其允许使用不同标记物(例如,荧光标记物)对Cas多肽容易地进行标记,这使得能够使用不同标记的探针进行多路复用。本领域技术人员将认识到,HaloTag有各种替代,例如SNAP或CLIP。

[0062] 如本文所述,本申请公开的探针的实施方案可包括与Cas多肽、gRNA或两者结合的一个或多个标记物。可以通过允许检测样品内的探针的任何方式对探针进行标记。因此,标记是允许直接或间接检测样品内探针的任何化合物或元素,包括随后的二次标记。

[0063] 可以在本申请公开的探针中使用的标记物的实例包括但不限于荧光标记物和荧光团(例如花青染料(Cy染料)、Alexas Fluor染料系列、Quasar染料)、染料、量子点、金颗粒、放射性标记物、磁性颗粒、酶、催化剂、分光光度标记物以及用于显微镜或其他检测方法(如MRI,CAT或PET)的其他分析物。其他的非限制性实例包括Alexa 488、DY547、Cy5、JF549和JF646。可以认识到,与蛋白质和核酸缀合的染料可以与本发明结合使用。

[0064] 在多肽被标记的一些实施方案中,探针可以表现出相对于其他DNA探针增加的灵敏度。不受理论的约束,认为增加的灵敏度至少部分归因于以下事实:相对于由核酸组成的其他已知探针,多肽可以接受更高浓度的标记物。

[0065] 在一些实施方案中,所述Cas多肽与所述一个或多个标记物结合。在一些实施方案中,gRNA与一个或多个标记物结合。在一些实施方案中,Cas多肽和gRNA都与一个或多个标记物结合。在一些实施方案中,多个标记物中的每一个具有不同的发光颜色。如本领域普通技术人员在研究本文后将认识到的,存在与对特定核酸序列具有特异性的探针相关联的多个标记物时,具有放大与该目标序列的检测相关联的信号的能力。以这种方式,可以有助于低拷贝核酸序列的检测。

[0066] 本文公开的主题也有利地允许多色和/或多路复用探针,参考图11A和图11B可以方便地理解其实例。

[0067] 在一些实施方案中,探针可以包括Cas多肽和对一个或多个其他靶核酸序列具有特异性的多个gRNA,其中每个gRNA与Cas多肽复合。在一些实施方案中,探针还可以包括多个结合到Cas多肽的标记物、gRNA或两者。在一些实施方案中,多个标记物与Cas多肽结合。这样的多个标记物可以例如经由标签(如HaloTag)与Cas多肽结合。在一些实施方案中,每个标签具有不同的发光颜色。在一些实施方案中,标记物选自Alexa 488、DY547、Cy5、JF549和JF646。

[0068] 在一些实施方案中,存在一系列探针,每个探针包含与Cas多肽复合的gRNA,其中每个gRNA对不同的靶核酸序列具有特异性,并且每个探针包括具有不同发光颜色的标记物。在一些实施方案中,标记物与Cas多肽、gRNA或两者结合。在一些实施方案中,标记物与每个Cas多肽结合。这样的标记物可以例如经由标签(如HaloTag)结合Cas多肽。在一些实施

方案中,每个标记物具有不同的发光颜色。在一些实施方案中,标记物选自Alexa 488、DY547、Cy5、JF549和JF646。参考图11B,通过提供一个限制性示例,具有四种不同发光颜色的标记物面板可以允许检测15个独特的目标物。

[0069] 如本文所述,本申请公开的探针还包括对目标核酸序列具有特异性的gRNA,所述gRNA与Cas多肽复合。在本发明的一些实施方案中,探针可包括摩尔比为1:1至1:4的Cas多肽与gRNA。可以使用大范围的摩尔比来增强Cas/gRNA复合物的形成并优化成像结果。当Cas多肽被标记时,等量或更多的gRNA,比例为1:1、1:2、1:3、...、1:20或更大比例,一起反应以增强复合物的形成。当gRNA被标记时,使用等量或更多的Cas多肽。当两者都被标记时,使用1:1的比例。一旦随后进行大量的洗涤步骤,比例的变化可能仍然给出良好的结果。

[0070] 术语“gRNA”在本文中用于指被选择用于特异性靶向特定目标核酸序列(“靶核酸序列”或“预定核酸序列”)的RNA序列。Cas多肽与gRNA之间的复合可以包括通过共价键、氢键和/或其他非共价键将gRNA和多肽结合,并且在CRISPR/Cas系统的领域中得到很好的理解。因此,示例性的gRNA可以包含靶向特定的目标核酸序列的部分和用于结合Cas多肽的部分,并且这些部分可以彼此相互排斥,也可以不彼此相互排斥。此外,其他部分可以任选地包含在gRNA中。

[0071] 在一些实施方案中,本探针包括与引导RNA(gRNA)缀合的多肽,其中所述gRNA靶向预定的核酸序列。在一些实施方案中,gRNA具有靶向部分。在一些实施方案中,所述gRNA靶向靶核酸序列并与其杂交。在一些情况下,gRNA可以包含单个RNA分子,并且在其他实施方案中,所述gRNA可以由两个或更多个RNA分子组装。两个或更多个RNA分子单独地也许能够或也许不能与多肽缀合,并且能够或不能靶向预定的核酸序列。然而,在组装时,gRNA将进行靶向预定的核酸序列。

[0072] 在本探针的实施方案中使用的gRNA分子可以在序列长度上变化。gRNA的与靶核酸序列互补的部分,即靶向部分,也可以在长度上变化,并且可以对其进行选择使得gRNA靶向靶核酸序列(或“特定目标核酸序列”或“预定核酸序列”)。在一些实施方案中,gRNA的该部分靶向预定核酸序列,该预定核酸序列或整个gRNA序列的长度为约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、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100或更多个核酸。在一些优选的实施方案中,gRNA的靶向部分的长度为约15-25个核酸。在一些实施方案中,gRNA具有与除靶核酸之外的多肽结合的部分。在一些实施方案中,该部分长度为约80个至约100个核酸,更优选长度为约90个核酸。

[0073] 用于设计gRNA的方法和工具是本领域普通技术人员已知的。例如,可以在以下站点获得多个公众可用的工具:crispr.mit.edu/、<https://chopchop.rc.fas.harvard.edu/>、www.e-crisp.org/、crispr.cos.uni-heidelberg.de/。

[0074] 术语“靶”、“靶向”等指化合物选择性接近和/或选择性结合靶化合物的特征。例如,靶向预定DNA序列的gRNA接近和/或选择性地结合该预定DNA序列。在一些实施方案中,当gRNA可与预定的DNA序列杂交时,gRNA分子被称为靶向预定的DNA序列,其中术语“杂交”是指一个核酸序列与另一个核酸序列以序列特异性的方式结合(例如,非共价结合)。因此,

可杂交序列在本文中可被称为“互补”序列。

[0075] 在一些情况下,核酸(例如,gRNA)的靶向部分与其靶核酸(例如DNA)的100%互补性不是杂交所必需的。在一些情况下,核苷酸可以在一个或多个片段上杂交,使得介于中间的或相邻的片段不参与杂交事件(例如,环结构或发夹结构)。核苷酸的靶向部分可以与可杂交靶序列具有至少50%、至少60%、至少70%、至少80%、至少90%、至少95%、至少99%或100%的序列互补性。

[0076] “互补性”是指核酸与另一种核酸序列通过传统的沃森-克里克(Watson-Crick)模式或其他非传统模式形成氢键的能力。百分比互补性表示核酸分子中可以与第二核酸序列形成氢键(例如,沃森-克里克碱基配对)的残基的百分比(例如,10个中的5、6、7、8、9、10个互补,也即50%、60%、70%、80%、90%和100%互补)。“完全互补”是指核酸序列的所有连续残基将与第二核酸序列中相同数目的连续残基形成氢键。本文所用的“基本互补”是指在8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、30、35、40、45、50或更多个核苷酸的区域内至少50%、55%、60%、65%、70%、75%、80%、85%、90%、95%、97%、98%、99%或100%的互补程度,或指在严格条件下杂交的两个核酸。

[0077] 如本文所用,用于杂交的“严格条件”是指与靶序列互补的核酸主要与靶序列杂交并且基本上不与非靶序列杂交的条件。严格条件通常是依赖于序列的,并且根据多个因素变化。通常,序列越长,序列与其靶序列特异性杂交的温度越高。在下述文献中详细描述了严格条件的非限制性实例:Tijssen(1993),Laboratory Techniques In Biochemistry And Molecular Biology-Hybridization With Nucleic Acid Probes Part 1,Second Charper“Overview of principles of hybridization and the strategy of nucleic acid probe assay”,Elsevier,N.Y.。

[0078] “杂交”是指一个或多个多核苷酸反应以形成通过核苷酸残基的碱基之间的氢键而被稳定化的复合物的反应。氢键可以由沃森-克里克碱基配对、Hoogstein结合或以任何其他序列特异性方式产生。复合物可以包含形成双链结构的两条链、形成多链复合物的三条或更多条链、单条自杂交链或其任何组合。杂交反应可以构成更全面的过程(例如引发PCR或通过酶切割多核苷酸)中的步骤。能够与给定序列杂交的序列被称为给定序列的“补体”。

[0079] 术语“核苷酸”、“多核苷酸”、“核酸”和“核酸序列”在本文中也用于指单链或双链形式的脱氧核糖核苷酸或核糖核苷酸及其聚合物。除非特别限定,否则术语包括含有天然核苷酸的已知类似物的核酸,该类似物具有与参照核酸相似的结合特性并且以类似于天然存在的核苷酸的代谢方式被代谢。除非另有说明,特定的核酸序列还隐含地包括其保守修饰的形式(例如简并密码子取代)和互补序列以及明确指出的序列。因此,术语核苷酸等包括本文所述的gRNA。

[0080] 术语“预定核酸序列”、“靶核酸序列”和“特定目标核酸序列”是指已知的核酸分子的核酸序列,在合成gRNA之前根据本文公开的发明选择该序列。因此,最广义地说,“靶核酸序列”一词仅表示具有特定的已知序列的靶核酸序列。

[0081] 靶核酸序列可以选自DNA序列、RNA序列或另一核酸类似物的序列,例如但不限于肽核酸(PNA)、吗啉代核酸和锁核酸(LNA)、以及乙二醇核酸(glycol nucleic acid,GNA)和苏氨酸核酸(TNA)。在一些实施方案中、靶核酸序列是其中一个或多个核苷酸已被修饰(例

如但不限于甲基化和羟甲基化)的核酸分子的核酸序列。如本文所述,靶核酸序列可以是未变性的天然DNA,靶核酸序列可以在尚未变性的染色体DNA内。

[0082] 靶核酸序列可以包含在含有遗传物质的任何样品中。在这方面,本文公开的探针、试剂盒和方法可用于检测和成像含有遗传物质的样品中的靶核酸序列。

[0083] 样品类型没有特别限制,只要样品包含遗传物质即可。在一些实施方案中,细胞是人细胞。在一些实施方案中,样品包含病毒、细菌或真菌。在一些实施方案中,样品是细胞。细胞可以是活细胞或固定细胞。细胞可以是植物细胞或动物细胞。在一些实施方案中,样品包括整个天然染色体和/或天然染色体的一部分。在一些实施方案中,细胞可以是人细胞。

[0084] 因此,在一些实施方案中,样品包括从个体获得的细胞。来自个体的示例性样品可以包括用于遗传测试的染色体涂片(spread)、培养物、产前材料、用于体外受精的样品、拭子、空气过滤器、水冷却塔、食品、饮料、头发、粪便、尿液、唾液、血液、淋巴、痰液、子宫颈涂片、精子、活检组织切片、尸体解剖组织切片等。样品可以包括通过本领域已知的各种固定方法得到的固定组织样品。化学固定的实例包括甲醛、多聚甲醛、甲酰胺、戊二醛、吐温-20和HCl、乙酸和甲醇以及其他醛和醇。

[0085] 本文公开的主题还包括试剂盒,其包括如本文所公开的探针或探针的组分。示例性试剂盒可以包括Cas多肽;用于与Cas多肽复合的、对靶核酸序列具有特异性的gRNA;以及能够结合Cas多肽、gRNA或两者的标记物。在一些实施方案中,提供试剂盒,其中标记物已经与Cas多肽结合。随后,在执行成像步骤之前,可以混合试剂盒的组分以获得探针或探针的集合。将探针存储为包含单独组分的试剂盒的能力允许与gRNA分离地合成多肽。相对于在多肽和gRNA已经组装的状态下储存的探针或相对其他已知的探针,保持组分彼此分离也可以增加某些具体的探针的保质期。

[0086] 另外,本探针和试剂盒允许探针的各种定制和配置。例如,因为每个探针不是被单独标记的,而是可以预先标记所有多肽,所以可以通过将标记的多肽与不同的gRNA缀合来合成不同的探针。因此,可以通过改变与多肽缀合的gRNA相对容易地使一组标记的多肽适合作为用于各种不同的预定序列的探针。与预先组装的且各自单独标记用于特定的DNA序列的其他已知探针相比,这可以显著节约成本。

[0087] 在一些实施方案中,试剂盒还包括如本文所述的标签。在试剂盒的一些实施方案中,标签是用于随后的二次标记的标签。在一些实施方案中,标签包含抗体靶标,并且试剂盒还包括针对抗体靶标的抗体。

[0088] 在一些实施方案中,试剂盒还包括用于制备含有靶核酸序列的样品的试剂。例如,试剂盒可以包括用于如本文所述或通过本领域已知的任何方法固定样品的试剂。

[0089] 本文公开的主题还涉及使用本申请公开的探针来成像核酸的方法。如本文所述,这种方法允许靶核酸序列和/或含有靶核酸序列的样品的检测和/或可视化,而不使核酸变性,例如,不使双链染色体DNA变性,所述变性可能对细胞、细胞核和染色体的形态有害。在这方面,该方法可以允许细胞中序列的亚核定位,以及使细胞核中染色体的三维结构可视化。

[0090] 在一些实施方案中,所述方法包括将包含核酸的样品与探针接触,然后检测探针是否已结合靶核酸序列。与核酸分子结合的探针的存在表明核酸分子包含探针所靶向的靶核酸序列。本文公开的方法还可以包括成像样品以确定靶核酸序列的位置和/或相对拷贝

数。

[0091] 接触步骤也可以通过使探针与样品直接接触的任何已知方式进行。在一些实施方案中,例如成像活细胞的实施方案,可以通过使用诸如微珠负载、显微注射、纳米颗粒或脂质介导的转导等方法递送探针来使本发明的探针与活细胞接触。因此,可以通过将探针递送至目标样品中的细胞上或细胞中来进行活细胞成像。

[0092] 如本文所述的探针的任何实施方案都可以根据本公开的主题的方法使用,考虑到所需的测定、样品和靶核酸序列,在研究本文后对探针进行特定选择的益处对于普通技术人员而言是显而易见的。在一些实施方案中,探针包括Cas多肽和多个gRNA,每个gRNA对一个或多个其他靶核酸序列具有特异性,其中每个gRNA与Cas多肽复合;并且该方法包括检测探针是否结合每个靶核酸序列,其中探针与核酸的结合表明样品中存在靶核酸序列。在一些实施方案中,存在一系列探针,每个探针包含与Cas多肽复合的gRNA,每个gRNA对不同的靶核酸序列具有特异性,并且每个探针包含具有不同发光颜色的标记物;并且该方法包括同时检测探针是否结合多个靶核酸序列中的每一个。根据需要,该方法可以包括成像样品以确定多个靶核酸序列中每一个的位置和/或相对拷贝数。

[0093] 检测和/或成像步骤可以通过任何已知的方式进行,包括目前用于荧光原位杂交成像的任何已知手段。在一些实施方案中,通过包括磁共振成像(MRI)、X射线计算机断层摄影(CAT)、正电子发射断层摄影(PET)或其组合的过程来执行检测步骤。

[0094] 本发明探针的实施方案具有检测基因组核酸或天然核酸的优异且意想不到的能力。已知的探针通常由核酸序列组成,需要DNA的固定和变性以检测DNA上的某些序列。然而,本发明的实施方案由与gRNA缀合的多肽组成,尽管它们具有相对大的尺寸,但它们能够穿透天然DNA以探测特定的预定基因组DNA序列。本发明探针的实施方案可以穿透细胞核和/或天然DNA。在这方面,包含本发明探针的某些实施方案的多肽可以突变以增强穿透和与天然DNA序列的结合。

[0095] 因为本发明探针的实施方案具有不需要对样品进行固定和/或变性的新特性,因此其还可以加速整个成像过程。例如,可以在约0.1小时、0.5小时、1.0小时、1.5小时、2.0小时、2.5小时或3.0小时内实现本发明的成像方法。当然,可能存在较长的时间帧内难处理的样品;然而,本文公开的主题的一个优点是相对快速的过程。

[0096] 另外,在一些情况下,提供了一种用于成像的方法,其包括提供两个或多个探针,其中每个探针靶向不同的预定DNA序列。可以通过用不同的gRNA形成探针来使探针靶向不同的核酸序列。另外,通过在每种类型的探针上使用不同类型的标记物或通过每种类型的探针上使用发射不同荧光的标记物,可对靶向预定核酸序列的每个探针进行不同的标记。在将样品与两个或多个探针接触后,可以鉴定出任何、一些或全部不同类型的探针是否与核酸序列结合。通过一起施用和成像不同的探针,可以创建包括来自每种探针类型的信号的点源的“光谱条形码”。这使得可以对样品内的两个或多个不同的预定DNA序列进行可视化和表征。该方法在本文中称为双色CASFISH或多色CASFISH。

[0097] 如本文所述,可以提供本发明的成像方法的实施方案用于研究和表征遗传样品中的易位。不需要对核酸样品进行固定和/或变性的一些实施方案可以有利地成像天然染色体,从而使得可以相对容易地鉴定样品内的潜在易位。

[0098] 在其他实施方案中,可以提供本发明的成像方法用于诊断或预测某些疾病和病

症。值得注意的是,由于可以在没有固定和杂交方案的情况下执行某些方法,一些实施方案提供用于对个体进行诊断或预测的相对快速的测定。这种快速性在时间敏感的情况下是特别有利的,例如在手术过程中进行以确定诊断和接下来所需的治疗的诊断方法。

[0099] 在一些实施方案中,靶核酸序列与疾病或病症相关。在这方面,样品是从个体获得的,并且该方法还包括如果在样品中检测到靶核酸序列和/或样品的成像表明个体患有或罹患疾病或病症的可能性增加,则确定个体患有或罹患该疾病或病症的可能性增加。尽管不是必需的,但是在一些实施方案中,该方法可以包括固定样品和/或获得固定的样品,这通常是常见的临床实践。

[0100] 本文所用的术语“诊断”(“diagnosing”和“diagnosis”)是指通过其本领域技术人员可以估计甚至确定个体是否患有给定疾病或病症的方法。与诊断一起,进行“预后”或“预测”可以指在与个体相关的样品中存在预定的DNA序列和/或特定基因的基础上,预测临床结果(在进行或不进行治疗的情况下),选择适当的治疗(或治疗是否有效),或监测当前的治疗并可能改变该治疗。

[0101] 本文所用的“预测”是指方法,通过所述方法本领域技术人员可以预测个体中病症的过程或结果。术语“预后”可以指能够以高达100%准确度预测病情或结果的能力,或者预测给定的过程或结果或多或少可能发生。术语“预后”也可以指某一过程或结果将会发生的可能性增加;也就是说,当与未显示基因突变的那些个体相比时,显示基因突变、反转、增加和/或缺失的个体中更可能发生该过程或结果。在某些实施方案中,预后约为给定预期结果的约5%的几率、约7%的几率、约10%的几率、约12%的几率、约15%的几率、约20%的几率、约30%的几率、约40%的机率、约50%的机率、约60%的几率、约75%的几率、约90%的几率或约95%的几率。如果可以进行准确的预后,可以选择对患者适当的疗法,在某些情况下可以选择不太剧烈的疗法或更有效的疗法。

[0102] 此外,术语“个体”包括人、植物、动物、细菌、病毒、其他微生物和包含遗传物质的任何个体。因此,根据本文公开的主题提供兽医用途,本文公开的主题提供了防止哺乳动物(例如人;以及由于濒危而重要的哺乳动物;具有经济重要性的哺乳动物,如养殖场养殖的供人类消费的动物;和/或对人类具有社会重要性的动物,如作为宠物饲养的动物或动物园中的动物)的氧化损伤的方法。这些动物的实例包括但不限于:诸如猫和狗的食肉动物;猪,包括小猪、阉猪和野猪;反刍动物和/或有蹄类动物,如母牛、公牛、绵羊、长颈鹿、鹿、山羊、野牛和骆驼;以及马。因此,还提供了家畜的治疗,家畜包括但不限于驯养的猪、反刍动物、有蹄类动物、马、家禽等。

[0103] 在这方面,可以通过提供含有对核酸序列(包括DNA序列和/或基因)具有选择性的gRNA的探针来实现诊断或预后,已知该核酸序列引起特定疾病或病症和/或与特定疾病或病症相关,和/或与罹患特定疾病或病症的风险增加或可能性增加相关。对可以通过本发明的方法诊断或预测的疾病和病症没有特别限制,普通技术人员可以认识到许多已知的、可以被探测到以诊断或预测特定疾病或病症的基因。这些疾病和病症可以包括但不限于各种类型的遗传异常、癌症和其他疾病及病症。可能被诊断或预测的具体疾病和病症包括但不限于包括慢性骨髓性白血病和急性早幼粒细胞性白血病在内的急性和慢性白血病、淋巴瘤、多发性骨髓瘤、乳腺癌、肺癌、结肠癌、前列腺癌、肉瘤和间质起源的肿瘤、包括少突神经胶质瘤在内的脑肿瘤、阿尔茨海默病、帕金森病、癫痫、肌萎缩性侧索硬化、多发性硬化、中

风、自闭症、猫叫综合征、1p36缺失综合征、天使人综合征 (Angelman syndrome)、Prader-Willi综合征、软腭-心-面综合征 (Velocardiofacial syndrome)、特纳综合征 (Turner syndrome)、克兰费尔特综合征 (Klinefelter syndrome)、爱德华综合征、唐氏综合征、帕图综合征、以及三染色体8症、三染色体9症和三染色体16症等。其他用途可以用于检测病毒感染和整合以及寄生性微生物 (例如疟原虫引起的疟疾)。本申请公开的方法也可用于农业、植物学研究和育种中的遗传测试。

[0104] 本领域普通技术人员将认识到,可以进一步优化和调整本申请公开的成像方法,其可以包括用于诊断或预后的方法。在某些情况下,本发明的探针和方法可以被修改为与较宽范围的样品类型和应用目的兼容。关于包括固定步骤的方法,这种优化可以包括但不限于改变乙酸和甲醇的比例、处理时间、渗透性甲酰胺和用于固定样品的多聚甲醛。

[0105] 本文公开的主题还包括组装探针的方法,如本文所述,包括在体外组装探针的方法。在一些实施方案中,在体外组装探针的方法包括选择任选地被标记的、能够靶向目标核酸序列的gRNA;提供任选地被标记的Cas多肽,其中gRNA和Cas多肽中的至少一个被标记;混合并孵育gRNA和Cas多肽。

[0106] 本文公开的主题还包括本文公开的探针,其与靶核酸序列结合或复合,和/或其中探针与每个gRNA的靶核酸序列结合。本文公开的主题还包括一系列探针,每个探针与靶核酸序列结合或复合,和/或其中每个探针与每个gRNA的靶核酸序列结合。因此,包含与探针结合的核酸序列的样品包括在本文公开的主题内。

[0107] 本文公开的主题通过以下具体的但非限制性的实施例进一步说明。以下实施例可以包括代表在与本发明相关的开发和实验过程中的不同时间收集的数据的数据汇编。

[0108] 实施例

[0109] 以下实施例使用与各种单引导RNA (sgRNA) 组装的、荧光标记的、核酸酶缺陷型Cas9 (dCas9) 蛋白,以展示对重复DNA元件的近着丝粒区、着丝粒、富G端粒和编码基因位点的快速标记和强健标记。将dCas9与sgRNA瓦片样覆盖的任意靶基因座的阵列进行组装使得非重复基因组序列可视化。如以下实施例所述,dCas9/sgRNA二元复合物是稳定的并且以高亲和力结合其靶DNA,允许顺序地或同时探测多个靶标。本文提供的实施例还示范了使用不同颜色的dCas9/sgRNA复合物的CAS FISH测定,其允许细胞中靶基因座的多色标记。

[0110] 材料和方法

[0111] Sd dCas9构建体和纯化

[0112] 将含有双重核酸酶突变 (D10A和H840A;dCas9) 的酿酒酵母 (*S. pyogenes*) Cas9基因克隆到具有N-末端六组氨酸亲和标签和C-末端Halo标签的PET302/N (Invitrogen) 中。产生具有额外的N-末端醛标记的构建体用于醛特异性Cy5标记 (13)。表达所有dCas9融合蛋白,并通过如本文所述的三步FPLC纯化方案来纯化所有dCas9融合蛋白。

[0113] dCas9 FPLC

[0114] 表达所有dCas9融合蛋白,并通过三步FPLC纯化方案来纯化所有dCas9融合蛋白,该过程按照 (20) 所述进行,具有以下改变。简言之,蛋白质在大肠杆菌BL21 (DE3) (Agilent Technologies) 中表达,用1mM异丙基β-D-1-硫代半乳糖吡喃糖苷 (IPTG) 诱导,然后在LB培养基中在16°C下生长过夜。将细胞在50mM磷酸钠 (pH7.0) 和300mM NaCl中溶解。将澄清的溶解产物施加于1mL His HiSTALON (Clontech) 亲和柱。通过增加咪唑浓度至150mM将结合的

蛋白质在50mM磷酸钠(pH7.0)和300mM NaCl中洗脱,并通过50,000-MWCO离心过滤器(Millipore Amicon)将其交换到缓冲液[50mM Hepes(pH7.5)、100mM KCl、1mM TCEP]中。蛋白质用阳离子交换色谱(HiTrap SP HP;GE Healthcare)进一步纯化,并用含有0.1-1.0M KCl的缓冲液线性梯度洗脱。在使用含有50mM Hepes(pH7.5)、150mM KCl和1mM TCEP的缓冲液的Superdex 200 16/60柱(GE Healthcare)上通过凝胶过滤色谱进一步纯化洗脱的蛋白质。将蛋白质储存在-80℃,另加20%甘油。

[0115] 肉毒杆菌的活细胞成像

[0116] 使用含有dCas9-Halo和sgMajSat的哺乳动物表达质粒转染MEF。在转染后48小时,在细胞培养器中将细胞与300nM JF549-Halo配体在37℃一起孵育15分钟。随后用温热的PBS洗涤细胞两次,用新鲜和温热的培养基一起再孵育30分钟,用温热的PBS洗涤两次,并在成像之前更换成新鲜的培养基。

[0117] 免疫荧光染色

[0118] 将细胞在1%BSA/10%正常山羊血清/0.3M甘氨酸的0.1%PBS-Tween 20溶液中封闭1小时,然后与Alexa 488-缀合的抗HP1 α 抗体(ab185018;Abcam)在4℃和工作稀释度(working dilution)为1:50的条件下一起孵育过夜。细胞用PBS洗涤,并在成像前用DAPI染色。使用白光激发系统(SOLA光引擎;Lumencor)与适当的过滤立方体设置(CFP-A-Basic-NTE;Semrock)一起成像Alexa 488标记。

[0119] dCas9荧光标记

[0120] 为了进行Halo结构域标记,将荧光Halo配体(JF549和JF646)以8:1的比例混合在蛋白质样品中,并在室温下反应30分钟,随后在4℃下孵育过夜。如(13)所述,用Cy5酰肼(GE Healthcare)荧光标记具有醛标签的dCas9-Halo。使用截留分子量(MWCO)为40-K的Zeba旋转脱盐离心柱(Thermo Scientific)除去过量未反应的荧光Halo配体或Cy5酰肼。蛋白质在含有50mM Hepes(pH 7.5)、150mM KCl、1mM三(2-羧乙基)膦(TCEP)和10%(vol/vol)甘油的储存缓冲液中洗脱,在液氮中快速冷冻并在-80℃储存。根据比尔定律由吸收光谱和消光系数计算蛋白质浓度和标记效率。

[0121] sgRNA合成

[0122] 通过夹板连接以下两个片段产生DY547标记的sgMajSat:具有DY547的sgMajSat(GE Dharmacon)的36个合成核苷酸5'和用T7转录和纯化的sgMajSat的77个核苷酸3'。使用具有以下序列的DNA模板,通过T7RNA聚合酶(MEGAshortscript T7Kit;Life Technologies)在体外合成未标记的sgRNA:5'-TAATACGACTCACTATAGGN₁₇₋₂₈GTTTAAAGAGCTATGCTGGAAACAGCATAGCAAGTTTAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACC GAGTCGGTGC-3'(SEQ ID NO:8)。模板含有T7启动子结合序列(下划线的)、sgRNA靶标(GGN₁₇₋₂₈) (粗体的)和如(6)所报道的sgRNA主链。通过gBlock(Integrated DNA Technologies)或PCR反应合成T7模板DNA。通过MEGAclear Transcription Clean-Up Kit(美国生命技术公司)纯化产生的sgRNA,通过10%变性PAGE验证质量。相关序列列于表1-3。

[0123] 表1:通过夹板连接合成DY547标记的sgMajSat

[0124]

5'-sgMajSat	DY547-CCAUAUCCACGUCCUACAGGUUUAAGAGCUAUGCU (SEQ ID NO: 9)
3'-sgMajSat	GGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUC- AACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUUU (SEQ ID NO: 10)
sgMajSat-桥	AAACTTGCTATGCTGTTTCCAGCATAGCTCTTAAACCTGT (SEQ ID NO: 11)

[0125] 表2:用于sgRNA的T7转录的DNA模板的合成:通过gBlock合成的T7转录模板序列

[0126]

gBlock_sgMajSat (SEQ ID NO: 12)	GGTGACACTATAGA AACTCGAGCAGCTGGATCCTAATACGACTCACTATA GGCCATAATCCACGTACAGGTTTAAGAGCTATGCTGGAAACAGCATA GCAAGTTTAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACC GAGTCGGTGC
gBlock_sg端粒 (SEQ ID NO: 13)	GGTGACACTATAGA AACTCGAGCAGCTGGATCCTAATACGACTCACTATA GGGTTAGGGTTAGGGTTAGGGTTAGTTTAAGAGCTATGCTGGAAACA GCATAGCAAGTTTAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGG CACCGAGTCGGTGC
gBlock_sgMinSat (SEQ ID NO: 14)	GGTGACACTATAGA AACTCGAGCAGCTGGATCCTAATACGACTCACTATA ggATCTAATATGTTCTACAGTGGTTTAAGAGCTATGCTGGAAACAGCAT AGCAAGTTTAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCAC CGAGTCGGTGC

[0127]

gBlock_sgMBS (SEQ ID NO: 15)	GGTGACACTATAGA AACTCGAGCAGCTGGATCCTAATACGACTCACTATA GGTCGACTCTAGAAAACATGGTTTAAGAGCTATGCTGGAAACAGCAT AGCAAGTTTAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCAC CGAGTCGGTGC
---------------------------------	--

[0128] T7转录模板的设计 (TAATACGACTCACTATAGGN₁₇₋₂₈GGTTTAAGAGCTATGCTGGAAACAGCATA GCAAGTTTAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC) (SEQ ID NO:8)。小写字母“g”表示不与sgRNA靶标互补的额外的鸟嘌呤。

[0129] 表3:通过PCR合成的T7转录模板

[0130]

表 3	
序列名称	序列
正向引物的设计 (SEQ ID NO: 16)	TAATACGACTCACTATAGGN ₁₇₋₂₈ GTTTAAGAGCTATGCTGG
sgRNA 主链 (SEQ ID NO: 17)	TAATACGACTCACTATAGGGTTTAAGAGCTATGCTGG
sgMUC4-I1 (SEQ ID NO: 18)	TAATACGACTCACTATAgGAGGTGACACCGTGGGCTGGGGGTTTAAGAGCTATGCTGG
sgMUC4-E2 (SEQ ID NO: 19)	TAATACGACTCACTATAgGAAGTGTTCGACAGGAAGAGTTTAAGAGCTATGCTGG
sgMUC1-E2 (SEQ ID NO: 20)	TAATACGACTCACTATAgGAAGGTATGGGTGTGGAAGGTATGTTTAAGAGCTATGCTGG
sgMUC4-覆瓦式正向引物	
sgMUC4_01 (SEQ ID NO: 21)	TAATACGACTCACTATAgGAAGAGTGGAGGCCGTGCGCGGGTTTAAGAGCTATGCTGG
sgMUC4_02 (SEQ ID NO: 22)	TAATACGACTCACTATAgGACCTCAGGTGATCTCCTGCCTGTTTAAGAGCTATGCTGG
sgMUC4_03 (SEQ ID NO: 23)	TAATACGACTCACTATAgGTATATTAGTAGAGACGGTTTAAGAGCTATGCTGG
sgMUC4_04 (SEQ ID NO: 24)	TAATACGACTCACTATAgGAGTAGCTGGAATTACAGGTGGTTTAAGAGCTATGCTGG
sgMUC4_05 (SEQ ID NO: 25)	TAATACGACTCACTATAgGCTCACTGCAACCTCCACCTCCCGTTTAAGAGCTATGCTGG
sgMUC4_06 (SEQ ID NO: 26)	TAATACGACTCACTATAgGACAGAGTCTCGCTCTCTCTCCCGTTTAAGAGCTATGCTGG
sgMUC4_07 (SEQ ID NO: 27)	TAATACGACTCACTATAgGAAGAGGAGAAAAGTGGGGAAGGTTTAAGAGCTATGCTGG

[0131]

表 3	
序列名称	序列
sgMUC4_08 (SEQ ID NO: 28)	TAATACGACTCACTATAgGAACAGAGGGCCAGAGAGCAGCCGTTTAAGAGCTATGCTGG
sgMUC4_09 (SEQ ID NO: 29)	TAATACGACTCACTATAgGTCTTTCTCTCTGCGAGTAAGCCTGTTTAAGAGCTATGCTGG
sgMUC4_10 (SEQ ID NO: 30)	TAATACGACTCACTATAgGTACACCCTTGTGTACAGAGCTGTTTAAGAGCTATGCTGG
sgMUC4_11 (SEQ ID NO: 31)	TAATACGACTCACTATAgGAAAACCTCATGTAAAGCTGCAGTTTAAGAGCTATGCTGG
sgMUC4_12 (SEQ ID NO: 32)	TAATACGACTCACTATAgGCAAGCAAGGGAAGCGACAAGGGTTTAAGAGCTATGCTGG
sgMUC4_13 (SEQ ID NO: 33)	TAATACGACTCACTATAgGAGGGCGGCCAGGGCGCAGAGTTTAAGAGCTATGCTGG
sgMUC4_14 (SEQ ID NO: 34)	TAATACGACTCACTATAgGCTTTTAAACCCGAGCTCAGTTTAAGAGCTATGCTGG
sgMUC4_15 (SEQ ID NO: 35)	TAATACGACTCACTATAgGTAGCCCCGGCATTGGCCTTGTTTAAGAGCTATGCTGG
sgMUC4_16 (SEQ ID NO: 36)	TAATACGACTCACTATAgGCCTGTGGGAGATGTCCCTCGTTTAAGAGCTATGCTGG
sgMUC4_17 (SEQ ID NO: 37)	TAATACGACTCACTATAgGTCCTGAAGCCAGAGGGACAGCCGTTTAAGAGCTATGCTGG
sgMUC4_18 (SEQ ID NO: 38)	TAATACGACTCACTATAgGAGTCTTTGGGGGAGAGTCTGTTTAAGAGCTATGCTGG
sgMUC4_19 (SEQ ID NO: 39)	TAATACGACTCACTATAgGCTCCTGCCCTGCCTCTCAGCGTTTAAGAGCTATGCTGG
sgMUC4_20 (SEQ ID NO: 40)	TAATACGACTCACTATAgGCATATTTGAGGAGCTTCTGTTTAAGAGCTATGCTGG
sgMUC4_21 (SEQ ID NO: 41)	TAATACGACTCACTATAgGCTGCAAGAGAAGCCATGCGTTTAAGAGCTATGCTGG
sgMUC4_22 (SEQ ID NO: 42)	TAATACGACTCACTATAgGATGTTTCAGGACTAGGCTGAGTTTAAGAGCTATGCTGG
sgMUC4_23 (SEQ ID NO: 43)	TAATACGACTCACTATAgGCTGGAGGGTGGGGAGGTGTAGTTTAAGAGCTATGCTGG
sgMUC4_24 (SEQ ID NO: 44)	TAATACGACTCACTATAgGTGGGATGAGCACTGGAGCGTTTAAGAGCTATGCTGG
sgMUC4_25 (SEQ ID NO: 45)	TAATACGACTCACTATAgGCCCTGCAGATGTGGTTGAGTTTAAGAGCTATGCTGG
sgMUC4_26	TAATACGACTCACTATAgGAGGCTGGGGCTGGGGCGCCGTTTAAGAGCTATGCTGG

[0132]

表 3	
序列名称	序列
(SEQ ID NO: 46)	
sgMUC4_27 (SEQ ID NO: 47)	TAATACGACTCACTATAgGTCTTTGCCGTGAACTGTTTCGTTTAAGAGCTATGCTGG
sgMUC4_28 (SEQ ID NO: 48)	TAATACGACTCACTATAgGACCGGGGCCCTGGGGAGACACGTTTAAGAGCTATGCTGG
sgMUC4_29 (SEQ ID NO: 49)	TAATACGACTCACTATAgGCTGGACACTCAGCTCCATGGTTTAAGAGCTATGCTGG
sgMUC4_30 (SEQ ID NO: 50)	TAATACGACTCACTATAgGAGCGCAGAGGGGCAAGACCTGTTTAAGAGCTATGCTGG
sgMUC4_31 (SEQ ID NO: 51)	TAATACGACTCACTATAgGAGAAGGAGTGAAGGACTGTGTTTAAGAGCTATGCTGG
sgMUC4_32 (SEQ ID NO: 52)	TAATACGACTCACTATAgGCTCCACGACATGCCTAGCTTCTTCGTTTAAGAGCTATGCTGG
sgMUC4_33 (SEQ ID NO: 53)	TAATACGACTCACTATAgGAGCTGGGCCAGGAGAGGAGAGTTTAAGAGCTATGCTGG
sgMUC4_34 (SEQ ID NO: 54)	TAATACGACTCACTATAgGACCGGGCATGACCAGGGCCTGTTTAAGAGCTATGCTGG
sgMUC4_35 (SEQ ID NO: 55)	TAATACGACTCACTATAgGGCAGCCCCACCCCCACAGTTTAAGAGCTATGCTGG
sgMUC4_36 (SEQ ID NO: 56)	TAATACGACTCACTATAgGTTCCTTTTGGCTCCCTGAAGGTTTAAGAGCTATGCTGG
sgMUC4_37 (SEQ ID NO: 57)	TAATACGACTCACTATAgGGTCTGTTTGACACTTGCCTTTAAGAGCTATGCTGG
sgMUC4_38 (SEQ ID NO: 58)	TAATACGACTCACTATAgGCCCCAGCCAGAGGAAAAACACAGTTTAAGAGCTATGCTGG
sgMUC4_39 (SEQ ID NO: 59)	TAATACGACTCACTATAgGTTTCCTTAAGGAACAGCCCGTTTAAGAGCTATGCTGG
sgMUC4_40 (SEQ ID NO: 60)	TAATACGACTCACTATAgGCAGACAGAGGTGGGCTAGACAGTTTAAGAGCTATGCTGG
sgMUC4_41 (SEQ ID NO: 61)	TAATACGACTCACTATAgGCCCCAGGCAGGAATGACTCAGAGTTTAAGAGCTATGCTGG
sgMUC4_42 (SEQ ID NO: 62)	TAATACGACTCACTATAgGACCCAGTTGCCTTTCCTGGTTTAAGAGCTATGCTGG
sgMUC4_43 (SEQ ID NO: 63)	TAATACGACTCACTATAgGACCCAGGGAGGTGACAGGCGTTTAAGAGCTATGCTGG
sgMUC4_44 (SEQ ID NO: 64)	TAATACGACTCACTATAgGCCACAGCGCACTCCACGGGGAAGTTTAAGAGCTATGCTGG

[0133]

表 3	
序列名称	序列
sgMUC4_45 (SEQ ID NO: 65)	TAATACGACTCACTATAgGTCCCAGACTGACAGATAGACCGTTTAAGAGCTATGCTGG
sgMUC4_46 (SEQ ID NO: 66)	TAATACGACTCACTATAgGAGGGGTCTGTGGAGAGTTTGTTTAAGAGCTATGCTGG
sgMUC4_47 (SEQ ID NO: 67)	TAATACGACTCACTATAgGTCCAGCATCAGCGACGCCCTGTTTAAGAGCTATGCTGG
sgMUC4_48 (SEQ ID NO: 68)	TAATACGACTCACTATAgGCCTAAGACTCCAGAGCCAAAGTTTAAGAGCTATGCTGG
sgMUC4_49 (SEQ ID NO: 69)	TAATACGACTCACTATAgGCTACTACGTAGGGTTGTCATGGTTTAAGAGCTATGCTGG
sgMUC4_50 (SEQ ID NO: 70)	TAATACGACTCACTATAgGTAAAAGTAGAAAAGGCATAAAGTTTAAGAGCTATGCTGG
sgMUC4_51 (SEQ ID NO: 71)	TAATACGACTCACTATAgGCACTTTTGGAGGCCAGGCGTTTAAGAGCTATGCTGG
sgMUC4_52 (SEQ ID NO: 72)	TAATACGACTCACTATAgGTGGAGACAGGGTTGGCCAAGCGTTTAAGAGCTATGCTGG
sgMUC4_53 (SEQ ID NO: 73)	TAATACGACTCACTATAgGCTCCCTGCAACCTCTGCCTCCCGTTTAAGAGCTATGCTGG
sgMUC4_54 (SEQ ID NO: 74)	TAATACGACTCACTATAgGCCTCTTTCTCAAACACGTCTTTAGTTTAAGAGCTATGCTGG
sgMUC4_55 (SEQ ID NO: 75)	TAATACGACTCACTATAgGAACCCGGAATGGCACTTGTGTGTTTAAGAGCTATGCTGG
sgMUC4_56 (SEQ ID NO: 76)	TAATACGACTCACTATAgGTGGCTTTTTAGAGGCACGGTTTAAGAGCTATGCTGG
sgMUC4_57 (SEQ ID NO: 77)	TAATACGACTCACTATAgGCTTGGTGTATTCAGAATGGTTTAAGAGCTATGCTGG
sgMUC4_58 (SEQ ID NO: 78)	TAATACGACTCACTATAgGCCTGCCAGGCCAGCCTCTGCCGTTTAAGAGCTATGCTGG
sgMUC4_59 (SEQ ID NO: 79)	TAATACGACTCACTATAgGCTAAGGACAAGAGGCAATGAGGTTTAAGAGCTATGCTGG
sgMUC4_60 (SEQ ID NO: 80)	TAATACGACTCACTATAgGACTCAATTTCTCAGAACATGCTGGTTTAAGAGCTATGCTGG
sgMUC4_61 (SEQ ID NO: 81)	TAATACGACTCACTATAgGACAGAGTTTCTCTCTGTCCCCCGTTTAAGAGCTATGCTGG
sgMUC4_62 (SEQ ID NO: 82)	TAATACGACTCACTATAgGGGGTTTCACCATGTTGGCCGTTTAAGAGCTATGCTGG
sgMUC4_63	TAATACGACTCACTATAgGCTCGCCTCGGCTCCCAAAGTGCCTTAAGAGCTATGCTGG

[0134]

表 3	
序列名称	序列
(SEQ ID NO: 83)	
sgMUC4_64 (SEQ ID NO: 84)	TAATACGACTCACTATAAGGGCATTGTGTTGCACGTGGTTAAGAGCTATGCTGG
sgMUC4_65 (SEQ ID NO: 85)	TAATACGACTCACTATAgGAGTGGAGCTGCGGGCAACCCGTTAAGAGCTATGCTGG
sgMUC4_66 (SEQ ID NO: 86)	TAATACGACTCACTATAgGTAGAGATGCCGCCCGCCCGTTAAGAGCTATGCTGG
sgMUC4_67 (SEQ ID NO: 87)	TAATACGACTCACTATAgGTCCAGTGGCCAGTGGATTTGGTTAAGAGCTATGCTGG
sgMUC4_68 (SEQ ID NO: 88)	TAATACGACTCACTATAgGAGGCAGCTGGGACTAGAACCCGTTAAGAGCTATGCTGG
sgMUC4_69 (SEQ ID NO: 89)	TAATACGACTCACTATAgTCCGGTGGGCTGGGCTGGTTGTTAAGAGCTATGCTGG
sgMUC4_70 (SEQ ID NO: 90)	TAATACGACTCACTATAgGAATGAATGGCTGTCTCAGCAGTTAAGAGCTATGCTGG
sgMUC4_71 (SEQ ID NO: 91)	TAATACGACTCACTATAgGAAACAGACGTGGCCAGTCTCTGTTAAGAGCTATGCTGG
sgMUC4_72 (SEQ ID NO: 92)	TAATACGACTCACTATAgGCTGAGAGCTGCATTTCCAAGTTAAGAGCTATGCTGG
sgMUC4_73 (SEQ ID NO: 93)	TAATACGACTCACTATAgGACAAAGTCAGGAAGGGCCCTGTGGTTAAGAGCTATGCTGG

[0135] PCR模板:

[0136]

GCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTAACTTGCTATGCTGTTTCCAGCATAGCTCTTAAAC (SEQ ID NO:94); 反向引物:GCACCGACTCGGTGCCACTT (SEQ ID NO:95)。小写字母“g”表示不与sgRNA靶标互补的额外的鸟嘌呤。

[0137] 细胞培养和脑部分制备

[0138] 将MEF和HeLa细胞在不含苯酚红 (Gibco) 且补加有10%FBS和1%青霉素/链霉素的DMEM中培养。由Cell Line Genetics公司进行MEF细胞的细胞遗传学分析。由新鲜冷冻的成年小鼠脑制备切片厚度为~20μM的冷冻切片。

[0139] CASFISH方案

[0140] 除非另有说明, 否则标准CASFISH方案如下: 将在35mm MatTek培养皿或组织切片上培养的细胞在-20℃下在预冷的比例为1:1的甲醇和乙酸溶液中固定20分钟。将样品用PBS轻轻振荡洗涤三次 (每次洗涤5分钟), 然后在封闭/反应缓冲液 [20mM HEPES (pH 7.5)、100mM KCl、5mM MgCl₂、新鲜添加的1mM DTT、5% (vol/vol) 甘油、1% BSA和0.1%吐温-20] 中在37℃孵育15分钟。为了组装CASFISH探针, 在封闭/反应缓冲液中将荧光标记的dCas9蛋白 (5-25nM) 分别与标记或未标记的sgRNA以1:1或1:4的摩尔比混合, 在室温下孵育10分钟, 并在下一步之前储存在冰上。将5nM荧光dCas9蛋白用于针对MUC4重复DNA元件的CASFISH。对于所有其他CASFISH实验, 使用25nM dCas9蛋白。将组装的dCas9/sgRNA应用于预先封闭

的细胞并在37℃下孵育5-30分钟。通过去除dCas9/sgRNA溶液并用封闭/反应缓冲液洗涤三次来终止反应。在含有20mM Hepes (pH7.5)、300mM NaCl、3M尿素和1.1% (vol/vol) Nonidet P-40的缓冲液中进一步洗涤sgMUC4-覆瓦式CASFISH样品。将短CASFISH方案修改为以下步骤：将细胞在-20℃下固定5分钟，用PBS漂洗3次，在37℃下用预孵育(5分钟)的dCas9和sgRNA混合物处理5分钟，再用PBS冲洗3次。所有样品在成像前用0.5μg/mL DAPI染色5分钟。

[0141] EMSA测定

[0142] 将dCas9蛋白和sgRNA在室温下一起孵育10分钟以检查该二元复合物，随后与靶DNA在37℃下一起孵育15分钟以检查该三元复合物。对于竞争测定，在加入竞争分子之后进行另外的二元或三元复合反应。将所得反应混合物在含有5mM MgCl₂的1×Tris/硼酸盐/EDTA缓冲液中进行1%琼脂糖凝胶电泳，并用GE Typhoon Trio+成像仪进行成像。

[0143] 显微镜和图像分析

[0144] 将所有CASFISH样品在装有100×油浸物镜(尼康CFI Plan Apo VC 100×Oil, NA 1.4)和EM CCD(Andor iXon Ultra 897)的倒置显微镜(尼康Eclipse Ti)上成像。使用白光激发系统(SOLA光引擎; Lumencor)结合适当的过滤立方体设置(分别为DAPI-1160B-NTE-Zero、LF561/LP-A-NTE或Cy5-3040C-NTE-Zero, Semrock)对DAPI、JF549或JF646标记成像。使用激光激发[分别为561nm(Cobolt Jive)或637nm(Vortran Stradus)]结合多波段二色性(Di01-R405/488/561/635; Semrock)对JF549或JF646标记成像。适用于JF549或JF646的发射滤光片(FF01-593/40或FF01-676/37; Semrock)放置在相机前面。以0.2μm的步长收集Z-堆栈30-40片以成像整个细胞核。使用ImageJ处理图像(21)。

[0145] 实施例1: 荧光标记的dCas9蛋白

[0146] 为了制备荧光标记的dCas9蛋白，我们构建了在N末端含有六组氨酸亲和和标签并在C末端含有Halo标签的dCas9融合蛋白。可以通过与各种有机荧光染料缀合的Halo配体有效地、共价地标记Halo标签以用于不同的成像目的(10)。我们纯化了在大肠杆菌中表达的重组dCas9融合蛋白，并用与Janelia Fluor 646 (JF646)缀合的Halo配体标记该蛋白(11)。(所有荧光染料列于表1。)

[0147] 表1: 荧光染料的光谱性质

[0148]

荧光染料	最大吸收波长, nm	最大发射波长, nm	发光颜色
Alexa 488	495	519	绿色
DY547	557	574	橙色
Cy5	650	670	红色
JF549	549	571	橙色
JF646	646	664	红色

[0149] 在鼠近着丝粒区的高度重复的主要卫星(MajSat)序列上测试本申请公开的策略，从而产生5'-DY547 (Cy3替代) 标记的sgRNA (sgMajSat) (图1B)。将等摩尔量的纯化的JF646-dCas9和DY547-sgMajSat一起孵育以形成dCas9/sgMajSat复合物，然后应用于用甲醇/乙酸溶液固定的小鼠胚胎成纤维细胞(MEF)。在孵育30分钟内，dCas9/sgMajSat探针有效地与近着丝粒区内的靶卫星DNA杂交(图1C)。来自检测sgMajSat的DY547信道的图像、来自检测dCas9蛋白的JF646信道的图像和来自DAPI信道的图像生成细胞核中近着丝粒区的共定位

(colocalized) 图案。如在共表达dCas9-halo和sgMajSat的活MEF细胞中观察到的那样,近着丝粒区具有不同的大小和数量(图2A)。通过靶向主要卫星序列的另一条链的独立sgRNA (sgMajSat-2)的dCas9染色(图2B)以及通过其与异染色质蛋白1 α (HP1 α) 免疫荧光染色的共定位(图2C),进一步验证了荧光dCas9/sgMajSat复合物对近着丝粒区的染色。相比之下,的JF646-dCas9单独孵育或与缺乏目标互补序列的sgMajSat的主链发夹部分一起孵育不能产生任何特异性染色(图2D),这表明dCas9/sgMajSat介导的近着丝粒区的染色是序列特异性的。单独的DY547-sgMajSat没有与目标主要卫星DNA杂交,这表明目标DNA在Cas9介导的FISH (CASFISH) 步骤中没有变性,并且观察到的信号不是简单地由独立于dCas9的RNA-DNA杂交引起的(图2D)。为了展示使用独立的标记方法进行的dCas9介导的荧光染色,我们将基因编码的醛标签设计到dCas9蛋白的N末端,并用Cy5酰肼 (Cy5Hz) 标记醛标签的dCas9 (13)。与DY547-sgMajSat组装的Cy5标记的dCas9蛋白产生与JF646标记的dCas9蛋白相同的近着丝粒区的特异性染色(图3)。总之,这些结果表明,体外组装的荧光dCas9/sgRNA复合物可以有效且特异性地标记细胞中的基因组DNA靶序列。因此,我们将该方法命名为“CRISPR/Cas9介导的FISH”,即CASFISH。

[0150] 实施例2: Halo标签标记

[0151] 荧光染料缀合的sgRNA探针 (~110nt) 的合成是昂贵的,而Halo标签的标记是高效且成本效益好的。因此,本实施例描述了组装荧光dCas9-Halo和未标记的sgRNA的策略以用于CASFISH。具体来说,我们通过Halo配体标记了dCas9-Halo蛋白, Halo配体为荧光染料的选择提供灵活性,并使用T7聚合酶体外合成sgRNA,这是一种结合低成本和可扩展性用于多路复用的方案。如成功的CASFISH成像和体外与靶DNA的特异性结合所证明的(见下文),用两种测试的荧光染料[JF549和JF646 (11)] 标记的dCas9-Halo蛋白保留了它们的活性。通过将标记的dCas9-Halo与T7合成的sgRNA组装,我们成功地染色了各种DNA靶标,包括端粒、小卫星和含有重复的MS2结合序列的基因阵列(图1D、图1E和图4A)。检测到的端粒数与通过细胞遗传学分析MEF细胞中的染色体数量一致(图1D、图4B和图4C)。端粒序列富含G,并且具有比常规基因组DNA更高的解链温度。虽然端粒的常规DNA FISH需要用于杂交的肽核酸探针,但是CASFISH探针可以使用未修饰的sgRNA有效地检测端粒,表明CASFISH在检测富含G的序列方面优于DNA FISH。因此,我们建立了一个成本效益好且方便的途径用于合成针对定制序列的荧光CASFISH探针。

[0152] 实施例3: 用于多个DNA靶标的CASFISH

[0153] 多个染色质结构域或基因的同时成像对于研究染色质相互作用是至关重要的。为了探索使用CASFISH同时检测多个DNA靶标的可能性,我们首先通过电泳迁移率变动分析(EMSA) 研究了CASFISH探针与靶DNA的结合。使用荧光标记的dCas9蛋白、sgMajSat及其靶DNA,我们发现dCas9蛋白与sgMajSat形成二元复合物,并在添加靶DNA时形成三元复合物。这两种复合物在凝胶上迁移形成不同的条带并离开游离dCas9、游离sgMajSat或游离靶DNA(图5A)。如预期的那样,dCas9在没有sgMajSat的情况下未表现出与靶DNA的可检测到的相互作用。为了检测dCas9/sgMajSat与其靶DNA结合的稳定性,我们通过向预先形成的三元复合物JF549-dCas9/sgMajSat/Cy5-DNA中加入过量的未标记的靶DNA进行竞争测定。我们发现,随着孵育中竞争DNA数量的增加,没有将Cy5-DNA从预先形成的三元复合物中置换出来,这可通过即使在竞争DNA过量100倍的情况下三元复合物中的Cy5荧光强度仍然不变来证明

(图6A)。将预组装的三元复合物与30倍的未标记的竞争DNA孵育长达1.5小时也没有将预先形成的三元复合物解离(图5A)。这些结果表明,一旦结合到其DNA底物上,Cas9/sgRNA/DNA复合物异常稳定。这个突出的特征使得可以通过对细胞进行多轮CASFISH测定实现顺序CASFISH策略,所述CASFISH测定使用被不同荧光染料标记并靶向不同DNA序列的探针。

[0154] 实施例5:一步多色成像

[0155] 为了探索在一个步骤中应用多种dCas9/sgRNA物质进行多色成像的可能性,我们通过sgRNA竞争测定法评估了二元复合物的稳定性。我们发现增加量高达100倍的未标记sgRNA竞争剂不能将DY547-sgRNA从预先组装的荧光dCas9/sgRNA中置换出来,正如dCas9复合物的DY547荧光强度未改变所证明的那样(图6B)。相比之下,在形成二元复合物之前添加未标记的sgRNA竞争剂大大减少了DY547-sgRNA与dCas9复合物的结合。这些结果证明了二元复合物的高稳定性和一步多色CASFISH的可行性。因此,我们测试了用于多色CASFISH的两种方法:依序进行多轮CASFISH或一步应用多个探针(图6C)。使用JF549-dCas9/sgMajSat和JF646-dCas9/sgTelomere探针,我们通过顺序CASFISH方法(图6D)或向细胞同时施用两种颜色的探针的一步法(图7),对细胞中的近着丝粒区和端粒进行双色成像。在任一情况下,两个成像信道之间没有可检测的交叉反应性。具有开关荧光标记的探针得到了相同的结果。如预期的那样,使用靶向sgMajSat和sgMajSat-2的两种不同sgRNA的CASFISH显示相同的图案(图6D)。因此,我们证明了CASFISH可用于用多种颜色成像多个序列特异性基因组区域。

[0156] 实施例6: CASFISH灵敏度

[0157] 主要的卫星和端粒序列包含用于sgRNA靶向的成百上千个重复序列。为了评估CASFISH的灵敏度,我们测试了针对具有在数十至数百水平的较低拷贝数的DNA底物的sgRNA。人粘蛋白4(MUC4)和粘蛋白1(MUC1)基因含有已在活性细胞中使用dCas9-EGFP成功成像的重复序列(6)。使用针对MUC4基因(sgMUC4-E2)的外显子2中的~400拷贝的靶标和针对内含子3(sgMUC4-I1)中的~45拷贝的靶标的sgRNA在所有检查的HeLa细胞核中均检测到突出的荧光点,所述的HeLa细胞核使用JF549-或JF646标记的dCas9蛋白(图8A和图9A)。如对分裂间期的HeLa细胞中的MUC4基因座的三个拷贝所预期的那样,在大多数细胞中存在三个荧光点(图8B)。我们还观察到,在相同细胞中,在复制位点的标记(图9B,点3和4)比在姊妹染色单体的标记(图9B,点1和2)更弱。使用这两种sgRNA的双色顺序CASFISH产生了极为接近的斑点,验证了该信号对MUC4基因是特异性的。基于两种颜色标记的基因座的比例,我们估计检测效率为 $\geq 94\%$ 。作为对照,MUC4和MUC1基因(sgMUC4-I1和sgMUC1-E1)的顺序CASFISH显示了这两个基因在细胞核中的不同位置(图8C),揭示了在依序进行的两轮CASFISH探测之间没有可检测的交叉反应性。总之,这些结果表明,CASFISH对于多个给定的基因组DNA序列的同时和多路复用标记是强健和有效的。

[0158] 实施例7:组织中的CASFISH

[0159] 该实施例显示如何通过对组织中存在的基因组序列进行快速且强健的测定来促进诊断程序。为了测试CASFISH是否可以应用于组织切片,我们制备了成年小鼠脑的冷冻切片,并对近着丝粒区和端粒进行CASFISH测定。dCas9/sgRNA探针穿透到15 μ m的脑切片,并在所有检测的细胞中均有效地标记其靶标(图10A)。我们预想在各种情况下通过CASFISH快速检测基因组元件可能是有利的,例如在快速基因诊断和感染剂检测的情况下。为了获得更

快速的CASFISH方案,我们减少了每个步骤的反应时间,并以sgMUC-E2为例,实现了15分钟方案(5分钟固定并同时进行的5分钟二元复合物组装、5分钟DNA结合和步骤之间的5分钟洗涤)。正如所预期的,快速方案实现了MUC4基因的强健标记,大多数细胞具有三个标记的点(图10A)。标记基因座的荧光强度与标准方案获得的荧光强度相似。这些结果表明,CASFISH可以是一种用于使细胞和原代组织中特定基因组元件可视化的快速而可靠的方法。可以使用类似的方法进行RNA的检测。

[0160] 总结

[0161] CASFISH测定法是快速、成本效益好且方便的(表4)。

[0162] 表4. CASFISH与DNA FISH以及活体Cas9成像的比较

[0163]

	DNA FISH	活体 Cas9 成像	CASFISH
探针	核酸探针	基因编码的 dCas9/sgRNA	体外组装的 dCas9/sgRNA
探针生成的难度	可靠的合成	易变且耗时	可靠的合成
实验持续时间	几小时到几天	立刻	几分钟到1小时
颜色	多色	每个 CRISPR 系统一个颜色	每个 CRISPR 系统多个颜色
高通量多路复用	是	有挑战	是
总体 DNA 变性	是	否	否
固定细胞成像	是	否	是
活细胞成像	否	是	可能
组织成像	是	可能	是

[0164] 首先,CASFISH利用基于CRISPR的机制进行快速DNA杂交。该酶探针比DNA FISH的仅含核酸的探针更有效,所述仅含核酸的探针需要加热和甲酰胺处理以使dsDNA变性。因此,虽然DNA FISH需要数小时或更长时间,CASFISH可以在优化条件下快至15分钟完成。第二,CASFISH的温和条件(室温和37°C)可以更好地保持细胞形态和DNA结构,因此CASFISH可以与单分子超分辨成像相结合作为研究基因组结构的有用工具。第三,CASFISH探针的双组分性质为多路复用提供了巨大的潜力和进一步工程化的空间(图11)。如本研究所证明的,dCas9的Halo标记使荧光标记具有简易性和灵活性以用于各种成像目的。标记的dCas9蛋白质的单一试剂可以以最小的成本与许多定制的未标记的sgRNA组装。因此,CASFISH可用多种单一颜色或组合色码检测多个靶标,有助于研究多个不同基因座的空间关系。CASFISH探针与靶标的快速结合动力学和温和的反应条件将允许在单细胞中进行高度多路复用的DNA分型。此外,CASFISH探针的蛋白质部分为探索其他标记方法和为每分子结合多个荧光染料提供了空间,以实现更高灵敏度和特异性。第四,CASFISH具有很大的应用潜力。CASFISH对于检测以富含G的端粒为例的“困难”DNA FISH序列是有利的。因为前间区序列邻近基序(PAM)附近的sgRNA序列组成对置换敏感,CASFISH可能有利于检测微妙的DNA变异,如SNP。CASFISH策略允许递送(delivery)与数量不限的sgRNA连接的荧光dCas9蛋白,从而允许标记任何给定尺寸的基因座和染色质结构域。原则上,定制CASFISH探针可以通过直接显微注射或蛋白质递送方法递送到活细胞,以便研究天然染色质结构。CASFISH快速并适用于原代组织,从而在快速基因诊断(如检测DNA易位)方面具备优势。CASFISH测定也可以扩展到正交CRISPR系统,使得能够进行广泛的试剂开发,并为组合、复用、多色成像系统的研制提供机会。

[0165] 除非另有定义,本文使用的所有技术和科学术语具有与本文公开主题所属领域的普通技术人员通常理解的含义相同的含义。尽管与本文所述的类似或等同的任何方法、装置和材料可以用于本文公开的主题的实践或测试中,但是现在描述代表性的方法、装置和

材料。

[0166] 根据长期的专利法公约,当在包括权利要求书在内的本申请中使用时,术语一个/一种(“a”、“an”)和所述/该(“the”)指的是“一个或多个”。因此,例如,提及“一个细胞”包括多个这样的细胞,等等。

[0167] 除非另有说明,否则在说明书和权利要求书中使用的表示成分数量、特性如反应条件等的所有数字应被理解为在所有情况下被术语“约”修饰。因此,除非另有说明,否则在本说明书和权利要求书中阐述的数值参数是近似值,其可以根据本文公开的主题寻求获得的期望性质而变化。

[0168] 本文所使用的术语“约”在涉及质量、重量、时间、体积、浓度或百分比的数值或量时,意在包括指定量的变化,该变化在一些实施方案中为 $\pm 20\%$,在一些实施方案中为 $\pm 10\%$,在一些实施方案中为 $\pm 5\%$,在一些实施方案中为 $\pm 1\%$,在一些实施方案中为 $\pm 0.5\%$,在一些实施方案中为 $\pm 0.1\%$,因为这些变化适于实施所公开的方法。

[0169] 本文所使用范围可以表示为“约”一个特定值,和/或表示为“约”另一特定值。还应当理解,本文公开了许多值,并且除该值本身外,每个值在本文中也被公开为“约”该特定值。例如,如果公开了值“10”,则还公开了“约10”。还应当理解,还公开了两个特定单元之间的每个单元。例如,如果公开10和15,则还公开了11、12、13和14。

[0170] 本说明书中提及的所有出版物、专利和专利申请通过引用并入本文,其程度如同每个单独的出版物、专利或专利申请被明确地、单独地被指明通过引用并入,包括下面列表中所述的参考文献:

[0171] 参考文献

[0172] 1.Hübner MR,Eckersley-Maslin MA,Spector DL(2013)Chromatin organization and transcriptional regulation.Curr Opin Genet Dev 23(2):89-95.

[0173] 2.Levsky JM,Singer RH(2003)Fluorescence in situ hybridization:Past, present and future.J Cell Sci 116(Pt 14):2833-2838.

[0174] 3.Beliveau BJ,et al.(2012)Versatile design and synthesis platform for visualizing genomes with Oligopaint FISH probes.Proc Natl Acad Sci USA 109(52):21301-21306.摘要/免费全文期刊.

[0175] 4.Sternberg SH,Doudna JA(2015)Expanding the Biologist's Toolkit with CRISPR-Cas9.Mol Cell 58(4):568-574.

[0176] 5.Mali P,Esvelt KM,Church GM(2013)Cas9as a versatile tool for engineering biology.Nat Methods 10(10):957-963.

[0177] 6.Chen B,et al.(2013)Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system.Cell 155(7):1479-1491.

[0178] 7.Tanenbaum ME,Gilbert LA,Qi LS,Weissman JS,Vale RD(2014)A protein-tagging system for signal amplification in gene expression and fluorescence imaging.Cell 159(3):635-646.

[0179] 8.Ma H,et al.(2015)Multicolor CRISPR labeling of chromosomal loci in human cells.Proc Natl Acad Sci USA 112(10):3002-3007.

[0180] 9.Sternberg SH,Redding S,Jinek M,Greene EC,Doudna JA(2014)DNA

interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature*.

[0181] 10. Encell LP, et al. (2012) Development of a dehalogenase-based protein fusion tag capable of rapid, selective and covalent attachment to customizable ligands. *Curr Chem Genomics* 6:55-71.

[0182] 11. Grimm JB, et al. (2015) A general method to improve fluorophores for live-cell and single-molecule microscopy. *Nat Methods* 12(3):244-250, 3, 250.

[0183] 12. Ziegler-Birling CEL, Miyanari Y, Torres-Padilla M-E (2013) Live visualization of chromatin dynamics with fluorescent TALEs. *Nat Struct Mol Biol* 20:1321-1324.

[0184] 13. Shi X, et al. (2012) Quantitative fluorescence labeling of aldehyde-tagged proteins for single-molecule imaging. *Nat Methods* 9(5):499-503.

[0185] 14. Zijlmans JM, et al. (1997) Telomeres in the mouse have large inter-chromosomal variations in the number of T2AG3 repeats. *Proc Natl Acad Sci USA* 94(14):7423-7428.

[0186] 15. O'Connell MR, et al. (2014) Programmable RNA recognition and cleavage by CRISPR/Cas9. *Nature*.

[0187] 16. Beliveau BJ, et al. (2015) Single-molecule super-resolution imaging of chromosomes and in situ haplotype visualization using Oligopaint FISH probes. *Nat Commun* 6:7147.

[0188] 17. Chen KH, Boettiger AN, Moffitt JR, Wang S, Zhuang X (2015) RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* 348(6233):aaa6090.

[0189] 18. Hsu PD, et al. (2013) DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol* 31(9):827-832.

[0190] 19. Esvelt KM, et al. (2013) Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. *Nat Methods* 10(11):1116-1121.

[0191] 20. Jinek M, et al. (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337(6096):816-821.

[0192] 21. Schneider CA, Rasband WS, Eliceiri KW (2012) NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* 9(7):671-675.

[0193] 22. Janicki SM, et al. (2004) From silencing to gene expression: Real-time analysis in single cells. *Cell* 116(5):683-698.

[0194] 23. Gilbert et al., CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes, *Cell*, 2013 vol. 154(2) pp. 442-51.

[0195] 24. Anton, et al., (2014) "Visualization of specific DNA sequences in living mouse embryonic stem cells with a programmable fluorescent CRISPR/Cas system," *Nucleus* 5(2):163-72.

[0196] 25. Van der Oost, et al., (2014) "Unravelling the structural and mechanistic basis of CRISPR-Cas systems," *Nat Rev Microbiol*, 12(7):479-92.

[0197] 26. Wright, et al., (2015) “Rational design of a split-Cas9 enzyme complex,” Proc Natl Acad Sci USA, 112(10):2984-89.

[0198] 27. 第8,697,359号美国专利, Zhang, “CRISPR-Cas systems and methods for altering expression of gene products.”

[0199] 应当理解,在不脱离本文公开的主题的范围的情况下,可以改变本文公开的主题的各种细节。此外,前述描述仅仅是为了说明的目的,而不是出于限制的目的。

序列表

<110> 霍华德休斯医学研究所

<120> 基因组探针

<130> KHP172110450.1

<150> US62/065,602

<151> 2014-10-17

<160> 95

<170> PatentIn version 3.5

<210> 1

<211> 5031

<212> DNA

<213> 人工序列

<220>

<223> 合成的

<400> 1

```

atgcatcatc atcatcatca cgtggataag aaataactcaa taggcttagc tateggcaca 60
aatagcgtcg gatggggcggg gatcactgat gaatataagg ttccgtctaa aaagttcaag 120
gttctgggaa atacagaccg ccacagtatc aaaaaaaatc ttataggggc tcttttattt 180
gacagtggag agacagcggg agcgactcgt ctcaaacgga cagctcgtag aaggtataca 240
cgtcgggaaga atcgtatttg ttatctacag gagatTTTTT caaatgagat ggcgaaagta 300
gatgatagtt tctttcatcg acttgaagag tcttttttgg tggaagaaga caagaagcat 360
gaacgtcatc ctatttttgg aaatatagta gatgaagttg cttatcatga gaaatatcca 420
actatctatc atctgcgaaa aaaattggta gattctactg ataaagcggg tttgcgctta 480
atctatttgg ccttagcgcg tatgattaag tttcgtggtc attttttgat tgaggagat 540
ttaaactctg ataatagtga tgtggacaaa ctatttatcc agtttgtaca aacctacaat 600
caattatttg aagaaaaccg tattaacgca agtggagtag atgctaaagc gattctttct 660
gcacgattga gtaaatacaag acgattagaa aatctcattg ctacagctccc cggtgagaag 720
aaaaatggct tatttgggaa tctcattgct ttgtcattgg gtttgacccc taattttaaa 780
tcaaattttg atttggcaga agatgctaaa ttacagcttt caaaagatac ttacgatgat 840
gatttagata atttatttgg gcaaatttga gatcaatatg ctgatttgtt tttggcagct 900
aagaatttat cagatgctat tttactttca gatatactaa gagtaaatac tgaataaact 960
aaggtcctcc tatcagcttc aatgattaag cgctacgatg aacatcatca agacttgact 1020
cttttaaaaag ctttagttcg acaacaactt ccagaaaagt ataaagaaat cttttttgat 1080
caatcaaaaa acggatatgc aggttatatt gatggggggg ctagccaaga agaattttat 1140
aaatttatca aaccaatttt agaaaaaatg gatggtactg aggaattatt ggtgaaacta 1200
aatcgtgaag atttgcctgc caagcaacgg acctttgaca acggctctat tcccataca 1260
attcacttgg gtgagctgca tgetattttg agaagacaag aagactttta tccattttta 1320
aaagacaatc gtgagaagat tgaaaaaatc ttgacttttc gaattcctta ttatgttggg 1380

```

ccattggcgc gtggcaatag tcgttttgca tggatgactc ggaagtctga agaacaatt 1440
 accccatgga attttgaaga agttgtcgat aaaggtgctt cagctcaatc atttattgaa 1500
 cgcatgacaa actttgataa aaatcttcca aatgaaaag tactaccaa acatagtttg 1560
 ctttatgagt attttacggt ttataacgaa ttgacaaagg tcaaatatgt tactgaagga 1620
 atgcgaaaac cagcatttct ttcaggtgaa cagaagaaag ccattgttga tttactcttc 1680
 aaaacaaatc gaaaagtaac cgttaagcaa ttaaaagaag attatttcaa aaaaatagaa 1740
 tgttttgata gtgttgaaat ttcaggagtt gaagatagat ttaatgcttc attaggtacc 1800
 taccatgatt tgctaaaaat tattaaagat aaagattttt tggataatga agaaaatgaa 1860
 gatatcttag aggatattgt tttaacattg acctatttg aagataggga gatgattgag 1920
 gaaagactta aaacatatgc tcacctcttt gatgataagg tgatgaaaca gcttaaactg 1980
 cgccgttata ctggttgggg acgtttgtct cgaaaattga ttaatggtat tagggataag 2040
 caatctggca aaacaatatt agattttttg aatcagatg gttttgcca tcgcaatttt 2100
 atgcagctga tccatgatga tagtttgaca tttaaagaag acattcaaaa agcacaagtg 2160
 tctggacaag gcgatagttt acatgaacat attgcaaatt tagctggtag ccctgctatt 2220
 aaaaaaggta ttttacagac tgtaaaagt gttgatgaat tgggtcaaagt aatggggcgg 2280
 cataagccag aaaatatcgt tattgaaatg gcacgtgaaa atcagacaac tcaaaagggc 2340
 cagaaaaatt cgcgagagcg tatgaaacga atcgaagaag gtatcaaaga attaggaagt 2400
 cagattctta aagagcatcc tgttgaaaat actcaattgc aaaatgaaaa gctctatctc 2460
 tattatctcc aaaatggaag agacatgtat gtggaccaag aattagatat taatcgttta 2520
 agtgattatg atgtcgatgc cattgttcca caaagtttcc ttaaagacga ttcaatagac 2580
 aataaggtct taacgcgttc tgataaaaaat cgtggtaaat cggataactg tccaagtgaa 2640
 gaagtagtca aaaagatgaa aaactattgg agacaacttc taaacgcca gttaatcact 2700
 caacgtaagt ttgataatth aacgaaagct gaacgtggag gtttgagtga acttgataaa 2760
 gctggtttta tcaaacgcca attggttgaa actcgccaaa tcaactagca tgtggcacia 2820
 attttgata gtcgcatgaa tactaaatac gatgaaaatg ataaacttat tcgagaggtt 2880
 aaagtgatta ccttaaaatc taaattagtt tctgacttcc gaaaagattt ccaattctat 2940
 aaagtacgtg agattaacaa ttaccatcat gccatgatg cgtatctaaa tgccgtcgtt 3000
 ggaactgctt tgattaagaa atatccaaa cttgaatcgg agtttgtcta tgggtgattat 3060
 aaagtttatg atgttcgtaa aatgattgct aagtctgagc aagaaatagg caaagcaacc 3120
 gcaaaatatt tcttttactc taatatcatg aacttctca aaacagaaat tacacttgca 3180
 aatggagaga ttcgcaaacg ccttctaate gaaactaatg gggaaactgg agaaattgtc 3240
 tgggataaag ggcgagattt tgccacagtg cgcaaagtat tgtccatgcc ccaagtcaat 3300
 attgtcaaga aaacagaagt acagacagge ggattctcca aggagtcaat ttaccaaaa 3360
 agaaattcgg acaagcttat tgetcgtaaa aaagactggg atccaaaaaa atatggtggt 3420
 tttgatagtc caacggtagc ttattcagtc ctagtggttg ctaaggtgga aaaagggaaa 3480
 tcgaagaagt taaaatccgt taaagagtta ctagggatca caattatgga aagaagtcc 3540
 tttgaaaaaa atccgattga ctttttagaa gctaaaggat ataaggaagt taaaaagac 3600
 ttaatcatta aactacctaa atatagtctt tttgagttag aaaacggtcg taaacgatg 3660
 ctggctagtg ccggagaatt acaaaaagga aatgagctgg ctctgccaag caaatatgtg 3720

aattttttat atttagctag tcattatgaa aagttgaagg gtagtccaga agataacgaa 3780
caaaaacaat tgtttggtga gcagcataag cattatntag atgagattat tgagcaaate 3840
agtgaatttt ctaagcgtgt tatttttagca gatgccaatt tagataaagt tcttagtgca 3900
tataacaaac atagagacaa accaatagct gaacaagcag aaaatattat tcatttattt 3960
acgttgacga atcttgagc tcccgtgct tttaaatatt ttgatacaac aattgatcgt 4020
aaacgatata cgtctacaaa agaagtttta gatgccactc ttatccatca atccatcact 4080
ggtctttatg aaacacgcat tgatttgagt cagctaggag gtgacggtgg ctccagatct 4140
gcagaaatcg gtactggcct tccattcgac cccattatg tggaagtcct gggcgagcgc 4200
atgcactacg tcgatgttg tccgcgcgat ggcaccctg tgctgttct gcacggtaac 4260
ccgacctct cctacgtgtg gcgcaacatc atcccgatg ttgcaccgac ccatcgctgc 4320
attgctccag acctgatcgg tatgggcaaa tccgacaaac cagacctggg ttatttcttc 4380
gacgaccacg tccgcttcat ggatgccttc atcgaagecc tgggtctgga agaggctgct 4440
ctggctattc acgactgggg ctccgctctg ggtttccact gggccaagcg caatccagag 4500
cgcgtcaaag gtattgcatt tatggagttc atccgcctc tcccacctg ggacgaatgg 4560
ccagaatttg cccgcgagac cttccaggcc ttccgacca cgcagctcgg ccgcaagctg 4620
atcatcgatc agaacgtttt tatcgagggt acgtgcccga tgggtgtcgt ccgcccgtg 4680
actgaagtcg agatggacca ttaccgcgag ccgttctga atcctgttga ccgcgagcca 4740
ctgtggcgct tccaaaacga gctgccaatc gccggtgagc cagcgaacat cgtcgcgctg 4800
gtcgaagaat acatggactg gctgcaccag tcccctgtcc cgaagctgct gttctggggc 4860
accccaggcg ttctgatccc accggccgaa gccgctcgc tggccaaaag cctgcctaac 4920
tgcaaggctg tggacatcgg cccgggtctg aatctgctgc aagaagacaa cccggacctg 4980
atcggcagcg agatcgcgcg ctggctgtct actctggaga tttccgtta a 5031

<210> 2

<211> 1676

<212> PRT

<213> 人工序列

<220>

<223> 合成的

<400> 2

Met	His	His	His	His	His	His	Val	Asp	Lys	Lys	Tyr	Ser	Ile	Gly	Leu
1				5					10					15	
Ala	Ile	Gly	Thr	Asn	Ser	Val	Gly	Trp	Ala	Val	Ile	Thr	Asp	Glu	Tyr
				20					25					30	
Lys	Val	Pro	Ser	Lys	Lys	Phe	Lys	Val	Leu	Gly	Asn	Thr	Asp	Arg	His
				35					40					45	
Ser	Ile	Lys	Lys	Asn	Leu	Ile	Gly	Ala	Leu	Leu	Phe	Asp	Ser	Gly	Glu
				50					55					60	
Thr	Ala	Glu	Ala	Thr	Arg	Leu	Lys	Arg	Thr	Ala	Arg	Arg	Arg	Tyr	Thr
65					70					75					80

Arg Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu			
	85	90	95
Met Ala Lys Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe			
	100	105	110
Leu Val Glu Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn			
	115	120	125
Ile Val Asp Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His			
	130	135	140
Leu Arg Lys Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu			
145	150	155	160
Ile Tyr Leu Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu			
	165	170	175
Ile Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe			
	180	185	190
Ile Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile			
	195	200	205
Asn Ala Ser Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser			
	210	215	220
Lys Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys			
225	230	235	240
Lys Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr			
	245	250	255
Pro Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln			
	260	265	270
Leu Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln			
	275	280	285
Ile Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser			
	290	295	300
Asp Ala Ile Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr			
305	310	315	320
Lys Ala Pro Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His			
	325	330	335
Gln Asp Leu Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu			
	340	345	350
Lys Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly			
	355	360	365
Tyr Ile Asp Gly Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys			
	370	375	380
Pro Ile Leu Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu			

385	390	395	400
Asn Arg Glu Asp Leu	Leu Arg Lys Gln Arg	Thr Phe Asp Asn Gly Ser	
	405	410	415
Ile Pro His Gln Ile	His Leu Gly Glu Leu	His Ala Ile Leu Arg Arg	
	420	425	430
Gln Glu Asp Phe Tyr	Pro Phe Leu Lys Asp	Asn Arg Glu Lys Ile Glu	
	435	440	445
Lys Ile Leu Thr Phe	Arg Ile Pro Tyr Tyr	Val Gly Pro Leu Ala Arg	
	450	455	460
Gly Asn Ser Arg Phe	Ala Trp Met Thr Arg	Lys Ser Glu Glu Thr Ile	
465	470	475	480
Thr Pro Trp Asn Phe	Glu Glu Val Val Asp	Lys Gly Ala Ser Ala Gln	
	485	490	495
Ser Phe Ile Glu Arg	Met Thr Asn Phe Asp	Lys Asn Leu Pro Asn Glu	
	500	505	510
Lys Val Leu Pro Lys	His Ser Leu Leu Tyr	Glu Tyr Phe Thr Val Tyr	
	515	520	525
Asn Glu Leu Thr Lys	Val Lys Tyr Val Thr	Glu Gly Met Arg Lys Pro	
	530	535	540
Ala Phe Leu Ser Gly	Glu Gln Lys Lys Ala	Ile Val Asp Leu Leu Phe	
545	550	555	560
Lys Thr Asn Arg Lys	Val Thr Val Lys Gln	Leu Lys Glu Asp Tyr Phe	
	565	570	575
Lys Lys Ile Glu Cys	Phe Asp Ser Val Glu	Ile Ser Gly Val Glu Asp	
	580	585	590
Arg Phe Asn Ala Ser	Leu Gly Thr Tyr His	Asp Leu Leu Lys Ile Ile	
	595	600	605
Lys Asp Lys Asp Phe	Leu Asp Asn Glu Glu	Asn Glu Asp Ile Leu Glu	
	610	615	620
Asp Ile Val Leu Thr	Leu Thr Leu Phe Glu	Asp Arg Glu Met Ile Glu	
625	630	635	640
Glu Arg Leu Lys Thr	Tyr Ala His Leu Phe	Asp Asp Lys Val Met Lys	
	645	650	655
Gln Leu Lys Arg Arg	Arg Tyr Thr Gly Trp	Gly Arg Leu Ser Arg Lys	
	660	665	670
Leu Ile Asn Gly Ile	Arg Asp Lys Gln Ser	Gly Lys Thr Ile Leu Asp	
	675	680	685
Phe Leu Lys Ser Asp	Gly Phe Ala Asn Arg	Asn Phe Met Gln Leu Ile	
	690	695	700

His Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val			
705	710	715	720
Ser Gly Gln Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly			
	725	730	735
Ser Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp			
	740	745	750
Glu Leu Val Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile			
	755	760	765
Glu Met Ala Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser			
	770	775	780
Arg Glu Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser			
785	790	795	800
Gln Ile Leu Lys Glu His Pro Val Glu Asn Thr Gln Leu Gln Asn Glu			
	805	810	815
Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val Asp			
	820	825	830
Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp Val Asp Ala Ile			
	835	840	845
Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile Asp Asn Lys Val Leu			
	850	855	860
Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp Asn Val Pro Ser Glu			
865	870	875	880
Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg Gln Leu Leu Asn Ala			
	885	890	895
Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn Leu Thr Lys Ala Glu Arg			
	900	905	910
Gly Gly Leu Ser Glu Leu Asp Lys Ala Gly Phe Ile Lys Arg Gln Leu			
	915	920	925
Val Glu Thr Arg Gln Ile Thr Lys His Val Ala Gln Ile Leu Asp Ser			
	930	935	940
Arg Met Asn Thr Lys Tyr Asp Glu Asn Asp Lys Leu Ile Arg Glu Val			
945	950	955	960
Lys Val Ile Thr Leu Lys Ser Lys Leu Val Ser Asp Phe Arg Lys Asp			
	965	970	975
Phe Gln Phe Tyr Lys Val Arg Glu Ile Asn Asn Tyr His His Ala His			
	980	985	990
Asp Ala Tyr Leu Asn Ala Val Val Gly Thr Ala Leu Ile Lys Lys Tyr			
	995	1000	1005
Pro Lys Leu Glu Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr			

1010	1015	1020
Asp Val Arg Lys Met Ile	Ala Lys Ser Glu Gln	Glu Ile Gly Lys
1025	1030	1035
Ala Thr Ala Lys Tyr Phe	Phe Tyr Ser Asn Ile	Met Asn Phe Phe
1040	1045	1050
Lys Thr Glu Ile Thr Leu	Ala Asn Gly Glu Ile	Arg Lys Arg Pro
1055	1060	1065
Leu Ile Glu Thr Asn Gly	Glu Thr Gly Glu Ile	Val Trp Asp Lys
1070	1075	1080
Gly Arg Asp Phe Ala Thr	Val Arg Lys Val Leu	Ser Met Pro Gln
1085	1090	1095
Val Asn Ile Val Lys Lys	Thr Glu Val Gln Thr	Gly Gly Phe Ser
1100	1105	1110
Lys Glu Ser Ile Leu Pro	Lys Arg Asn Ser Asp	Lys Leu Ile Ala
1115	1120	1125
Arg Lys Lys Asp Trp Asp	Pro Lys Lys Tyr Gly	Gly Phe Asp Ser
1130	1135	1140
Pro Thr Val Ala Tyr Ser	Val Leu Val Val Ala	Lys Val Glu Lys
1145	1150	1155
Gly Lys Ser Lys Lys Leu	Lys Ser Val Lys Glu	Leu Leu Gly Ile
1160	1165	1170
Thr Ile Met Glu Arg Ser	Ser Phe Glu Lys Asn	Pro Ile Asp Phe
1175	1180	1185
Leu Glu Ala Lys Gly Tyr	Lys Glu Val Lys Lys	Asp Leu Ile Ile
1190	1195	1200
Lys Leu Pro Lys Tyr Ser	Leu Phe Glu Leu Glu	Asn Gly Arg Lys
1205	1210	1215
Arg Met Leu Ala Ser Ala	Gly Glu Leu Gln Lys	Gly Asn Glu Leu
1220	1225	1230
Ala Leu Pro Ser Lys Tyr	Val Asn Phe Leu Tyr	Leu Ala Ser His
1235	1240	1245
Tyr Glu Lys Leu Lys Gly	Ser Pro Glu Asp Asn	Glu Gln Lys Gln
1250	1255	1260
Leu Phe Val Glu Gln His	Lys His Tyr Leu Asp	Glu Ile Ile Glu
1265	1270	1275
Gln Ile Ser Glu Phe Ser	Lys Arg Val Ile Leu	Ala Asp Ala Asn
1280	1285	1290
Leu Asp Lys Val Leu Ser	Ala Tyr Asn Lys His	Arg Asp Lys Pro
1295	1300	1305

Ile Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr 1310	1315	1320
Asn Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile 1325	1330	1335
Asp Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr 1340	1345	1350
Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp 1355	1360	1365
Leu Ser Gln Leu Gly Gly Asp Gly Gly Ser Arg Ser Ala Glu Ile 1370	1375	1380
Gly Thr Gly Phe Pro Phe Asp Pro His Tyr Val Glu Val Leu Gly 1385	1390	1395
Glu Arg Met His Tyr Val Asp Val Gly Pro Arg Asp Gly Thr Pro 1400	1405	1410
Val Leu Phe Leu His Gly Asn Pro Thr Ser Ser Tyr Val Trp Arg 1415	1420	1425
Asn Ile Ile Pro His Val Ala Pro Thr His Arg Cys Ile Ala Pro 1430	1435	1440
Asp Leu Ile Gly Met Gly Lys Ser Asp Lys Pro Asp Leu Gly Tyr 1445	1450	1455
Phe Phe Asp Asp His Val Arg Phe Met Asp Ala Phe Ile Glu Ala 1460	1465	1470
Leu Gly Leu Glu Glu Val Val Leu Val Ile His Asp Trp Gly Ser 1475	1480	1485
Ala Leu Gly Phe His Trp Ala Lys Arg Asn Pro Glu Arg Val Lys 1490	1495	1500
Gly Ile Ala Phe Met Glu Phe Ile Arg Pro Ile Pro Thr Trp Asp 1505	1510	1515
Glu Trp Pro Glu Phe Ala Arg Glu Thr Phe Gln Ala Phe Arg Thr 1520	1525	1530
Thr Asp Val Gly Arg Lys Leu Ile Ile Asp Gln Asn Val Phe Ile 1535	1540	1545
Glu Gly Thr Leu Pro Met Gly Val Val Arg Pro Leu Thr Glu Val 1550	1555	1560
Glu Met Asp His Tyr Arg Glu Pro Phe Leu Asn Pro Val Asp Arg 1565	1570	1575
Glu Pro Leu Trp Arg Phe Pro Asn Glu Leu Pro Ile Ala Gly Glu 1580	1585	1590
Pro Ala Asn Ile Val Ala Leu Val Glu Glu Tyr Met Asp Trp Leu		

1595	1600	1605
His Gln Ser Pro Val Pro Lys Leu Leu Phe Trp Gly Thr Pro Gly		
1610	1615	1620
Val Leu Ile Pro Pro Ala Glu Ala Ala Arg Leu Ala Lys Ser Leu		
1625	1630	1635
Pro Asn Cys Lys Ala Val Asp Ile Gly Pro Gly Leu Asn Leu Leu		
1640	1645	1650
Gln Glu Asp Asn Pro Asp Leu Ile Gly Ser Glu Ile Ala Arg Trp		
1655	1660	1665
Leu Ser Thr Leu Glu Ile Ser Gly		
1670	1675	

<210> 3

<211> 5127

<212> DNA

<213> 人工序列

<220>

<223> 合成的

<400> 3

```

atgcatccaa agaagaagcg gaaggtcggg atccacggag tcccagcagc ccatcatcat 60
catcatcacg tggataagaa atactcaata ggcttagcta tcggcacaaa tagcgtcggg 120
tgggcggtga tcaactgatga atataaggtt ccgtctaaaa agttcaaggt tctgggaaat 180
acagaccgcc acagtatcaa aaaaaatctt ataggggctc ttttatttga cagtggagag 240
acagcgggaag cgactcgtct caaacggaca gctcgtagaa ggtatacacg tcggaagaat 300
cgtatttggt atctacagga gattttttca aatgagatgg cgaaagtaga tgatagtttc 360
tttcatcgac ttgaagagtc ttttttgggt gaagaagaca agaagcatga acgtcatcct 420
atttttggaa atatagtaga tgaagttgct tatcatgaga aatatccaac tatctatcat 480
ctgcgaaaaa aattggtaga ttctactgat aaagcggatt tgcgcttaat ctatttggcc 540
ttagcgcata tgattaagtt tcgtgggtcat tttttgattg aggagattt aaatcctgat 600
aatagtgatg tggacaaaact atttatccag ttggtacaaa cctacaatca attatttgaa 660
gaaaacccta ttaacgcaag tggagtagat gctaaagega ttctttctgc acgattgagt 720
aatcaagac gattagaaaa tctcattgct cagctccccg gtgagaagaa aatggctta 780
tttgggaatc tcattgcttt gtcattgggt ttgacccta attttaate aaattttgat 840
ttggcagaag atgctaaatt acagetttca aaagatactt acgatgatga tttagataat 900
ttattggcgc aaattggaga tcaatatget gatttgtttt tggcagctaa gaatttatca 960
gatgetattt tactttcaga tatcetaaga gtaaactg aaataactaa ggctccccta 1020
tcagcttcaa tgattaaacg ctacgatgaa catcatcaag acttgactct tttaaaagct 1080
ttagttcgac aacaacttcc agaaaagtat aaagaaatct tttttgatca atcaaaaaac 1140
ggatatgcag gttatattga tgggggagct agccaagaag aattttataa atttatcaaa 1200
ccaattttag aaaaaatgga tgggtactgag gaattattgg tgaaactaaa tcgtgaagat 1260

```

ttgctgcgca agcaacggac ctttgacaac ggctctatc cccatcaaat tcaactgggt 1320
 gagctgcatg ctattttgag aagacaagaa gacttttate catttttaaa agacaatcgt 1380
 gagaagattg aaaaaatcct gacttttcga attcettatt atgttggctc attggcgcgt 1440
 ggcaatagtc gttttgcatg gatgactcgg aagtctgaag aaacaattac cccatggaat 1500
 tttgaagaag ttgtcgataa aggtgcttca gctcaatcat ttattgaacg catgacaaac 1560
 tttgataaaa atcttccaaa tgaaaaagta ctacccaaac atagtttgct ttatgagtat 1620
 tttacggttt ataacgaatt gacaaaggte aaatatgta ctgaaggaat gcgaaaacca 1680
 gcatttcttt cagggtgaaca gaagaaagcc attgttgatt tactcttcaa acaaatcga 1740
 aaagtaaccg ttaagcaatt aaaagaagat tatttcaaaa aatagaatg ttttgatagt 1800
 gttgaaattt caggagttga agatagattt aatgcttcat taggtacctt ccatgatttg 1860
 ctaaaaaatta ttaaagataa agattttttg gataatgaag aaaatgaaga tatcttagag 1920
 gatattgttt taacattgac cttatttgaa gatagggaga tgattgagga aagacttaaa 1980
 acatatgctc acctctttga tgataaggte atgaaacage ttaaactgct ccgttatact 2040
 ggttggggac gtttgtctcg aaaattgatt aatggtatta gggataagca atctggcaaa 2100
 acaatattag attttttgaa atcagatggg tttgccaate gcaattttat gcagctgac 2160
 catgatgata gtttgacatt taaagaagac attcaaaaag cacaagtgtc tggacaaggc 2220
 gatagtttac atgaacatat tgcaaattha gctggtagcc ctgctattaa aaaaggtatt 2280
 ttacagactg taaaagttgt tgatgaattg gtcaaagtaa tggggcggca taagccagaa 2340
 aatatcgta ttgaaatggc acgtgaaaat cagacaacte aaaaggcca gaaaaattcg 2400
 cgagagcgta tgaaacgaat cgaagaaggt atcaaagaat taggaagtca gattcttaaa 2460
 gagcatcctg ttgaaaatac tcaattgcaa aatgaaaagc tctatcteta ttatctcaa 2520
 aatggaagag acatgtatgt ggaccaagaa ttagatatta atcgtttaag tgattatgat 2580
 gtcgatgcca ttgttcaca aagtttcctt aaagacgatt caatagacaa taaggtctta 2640
 acgcgttctg ataaaaatcg tggtaaactg gataacgttc caagtgaaga agtagtcaaa 2700
 aagatgaaaa actattggag acaacttcta aacgccaagt taatcactca acgtaagttt 2760
 gataattha cgaaagctga acgtggaggt ttgagtgaac ttgataaagc tggttttatc 2820
 aaacgccaat tggttgaaac tcgccaatc actaagcatg tggcacaat tttggatagt 2880
 cgcatgaata ctaaatacga tgaaaatgat aaacttattc gagaggtaaa agtgattacc 2940
 ttaaaatcta aattagtttc tgacttccga aaagatttcc aattctataa agtacgtgag 3000
 attaacaatt accatcatgc ccatgatgag tatctaaatg ccgtcgttg aactgcttg 3060
 attaagaaat atccaaaact tgaatcggag tttgtctatg gtgattataa agtttatgat 3120
 gttcgtaaaa tgattgctaa gtctgagcaa gaaatagcca aagcaaccgc aaaatatttc 3180
 ttttacteta atatcatgaa cttcttcaaa acagaaatta cacttgcaaa tggagagatt 3240
 cgcaaacgcc ctctaactga aactaatggg gaaactggag aaattgtctg ggataaaggg 3300
 cgagattttg ccacagtgcg caaagtattg tccatgcccc aagtcaatat tgtcaagaaa 3360
 acagaagtac agacagcggt attctccaag gagtcaattt taccaaaag aaattcggac 3420
 aagcttattg ctcgtaaaaa agactgggat ccaaaaaaat atggtggttt tgatagtcca 3480
 acggtagctt attcagtcct agtggttgct aaggtggaag aagggaatc gaagaagtta 3540
 aaatccgta aagagtact agggatcaca attatggaag gaagttcctt tgaaaaaaat 3600

ccgattgact ttttagaagc taaaggatat aaggaagtta aaaaagactt aatcattaata 3660
 ctacctaaat atagtctttt tgagttagaa aacggctcga aacggatgct ggctagtgcc 3720
 ggagaattac aaaaaggaaa tgagctggct ctgccaagca aatatgtgaa ttttttatat 3780
 ttagctagtc attatgaaaa gttgaagggt agtccagaag ataacgaaca aaaacaattg 3840
 tttgtggagc agcataagca ttatttagat gagattattg agcaaatacag tgaattttct 3900
 aagcgtgtta ttttagcaga tgccaattta gataaagttc ttagtgcata taacaacat 3960
 agagacaaac caatacgtga acaagcagaa aatattattc atttatttac gttgacgaat 4020
 cttggagctc ccgctgcttt taaatatttt gatacaaca ttgatcgtaa acgatatacg 4080
 tctacaaaag aagttttaga tgccactctt atccatcaat ccatactgg tctttatgaa 4140
 acacgcattg atttgagtca gctaggaggt gacggtggtc ccagatctgc agaaatcgg 4200
 actggctttc cattcgacce ccattatgtg gaagtctctg gcgagcgcac gcactacgtc 4260
 gatgttggtc cgcgcgatgg caccctgtg ctgttctctc acggttaacc gacctctctc 4320
 tacgtgtggc gcaacatcat cccgatggt gcaccgacce atcgtctgat tgctccagac 4380
 ctgatcggta tgggcaaate cgacaaacca gacctgggtt atttcttcca cgaccacgtc 4440
 cgcttcatgg atgccttcat cgaagcctg ggtctggaag aggtctctct ggctattcac 4500
 gactggggct ccgctctggg tttccactgg gccaaagcga atccagagcg cgtcaaagg 4560
 attgcattta tggagtcat ccgcctatc ccgacctggg acgaatggcc agaatttgcc 4620
 cgcgagacct tccaggcctt ccgcaccacc gacgtcggcc gcaagctgat catgatcag 4680
 aacgttttta tcgagggtac gctgccgatg ggtgtctctc gcccgctgac tgaagtcgag 4740
 atggaccatt accgcgagcc gttctgaat cctgttgacc gcgagccact gtggcgcttc 4800
 ccaaacgagc tgccaatcgc cggtagacca gcgaacatcg tcgcgctggt cgaagaatac 4860
 atggactggc tgcaccagtc ccctgtcccg aagctgctgt tctggggcac cccaggcgtt 4920
 ctgatccac cgccgaagc cgctcgcctg gccaaaagcc tgccctaactg caaggctgtg 4980
 gacatcggcc cgggtctgaa tctgctgcaa gaagacaacc cggacctgat cggcagcgag 5040
 atcgcgcgct ggctgtctac tctggagatt tccgtaagc gtctgtctgc tactaagaaa 5100
 gctggtcaag ctaagaaaa gaaataa 5127

<210> 4

<211> 1708

<212> PRT

<213> 人工序列

<220>

<223> 合成的

<400> 4

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

Lys Val Pro Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His
 50 55 60
 Ser Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu
 65 70 75 80
 Thr Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr
 85 90 95
 Arg Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu
 100 105 110
 Met Ala Lys Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe
 115 120 125
 Leu Val Glu Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn
 130 135 140
 Ile Val Asp Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His
 145 150 155 160
 Leu Arg Lys Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu
 165 170 175
 Ile Tyr Leu Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu
 180 185 190
 Ile Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe
 195 200 205
 Ile Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile
 210 215 220
 Asn Ala Ser Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser
 225 230 235 240
 Lys Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys
 245 250 255
 Lys Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr
 260 265 270
 Pro Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln
 275 280 285
 Leu Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln
 290 295 300
 Ile Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser
 305 310 315 320
 Asp Ala Ile Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr
 325 330 335
 Lys Ala Pro Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His
 340 345 350
 Gln Asp Leu Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu

355	360	365
Lys Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly		
370	375	380
Tyr Ile Asp Gly Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys		
385	390	395
Pro Ile Leu Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu		
405	410	415
Asn Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser		
420	425	430
Ile Pro His Gln Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg		
435	440	445
Gln Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu		
450	455	460
Lys Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg		
465	470	475
Gly Asn Ser Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile		
485	490	495
Thr Pro Trp Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln		
500	505	510
Ser Phe Ile Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu		
515	520	525
Lys Val Leu Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr		
530	535	540
Asn Glu Leu Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro		
545	550	555
Ala Phe Leu Ser Gly Glu Gln Lys Lys Ala Ile Val Asp Leu Leu Phe		
565	570	575
Lys Thr Asn Arg Lys Val Thr Val Lys Gln Leu Lys Glu Asp Tyr Phe		
580	585	590
Lys Lys Ile Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp		
595	600	605
Arg Phe Asn Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile		
610	615	620
Lys Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu		
625	630	635
Asp Ile Val Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu		
645	650	655
Glu Arg Leu Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys		
660	665	670

Gln Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys
 675 680 685
 Leu Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp
 690 695 700
 Phe Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile
 705 710 715 720
 His Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val
 725 730 735
 Ser Gly Gln Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly
 740 745 750
 Ser Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp
 755 760 765
 Glu Leu Val Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile
 770 775 780
 Glu Met Ala Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser
 785 790 795 800
 Arg Glu Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser
 805 810 815
 Gln Ile Leu Lys Glu His Pro Val Glu Asn Thr Gln Leu Gln Asn Glu
 820 825 830
 Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val Asp
 835 840 845
 Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp Val Asp Ala Ile
 850 855 860
 Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile Asp Asn Lys Val Leu
 865 870 875 880
 Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp Asn Val Pro Ser Glu
 885 890 895
 Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg Gln Leu Leu Asn Ala
 900 905 910
 Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn Leu Thr Lys Ala Glu Arg
 915 920 925
 Gly Gly Leu Ser Glu Leu Asp Lys Ala Gly Phe Ile Lys Arg Gln Leu
 930 935 940
 Val Glu Thr Arg Gln Ile Thr Lys His Val Ala Gln Ile Leu Asp Ser
 945 950 955 960
 Arg Met Asn Thr Lys Tyr Asp Glu Asn Asp Lys Leu Ile Arg Glu Val
 965 970 975
 Lys Val Ile Thr Leu Lys Ser Lys Leu Val Ser Asp Phe Arg Lys Asp

980	985	990
Phe Gln Phe Tyr Lys Val Arg Glu Ile Asn Asn Tyr His His Ala His		
995	1000	1005
Asp Ala Tyr Leu Asn Ala Val Val Gly Thr Ala Leu Ile Lys Lys		
1010	1015	1020
Tyr Pro Lys Leu Glu Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val		
1025	1030	1035
Tyr Asp Val Arg Lys Met Ile Ala Lys Ser Glu Gln Glu Ile Gly		
1040	1045	1050
Lys Ala Thr Ala Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe		
1055	1060	1065
Phe Lys Thr Glu Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg		
1070	1075	1080
Pro Leu Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp		
1085	1090	1095
Lys Gly Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro		
1100	1105	1110
Gln Val Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe		
1115	1120	1125
Ser Lys Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile		
1130	1135	1140
Ala Arg Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp		
1145	1150	1155
Ser Pro Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu		
1160	1165	1170
Lys Gly Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly		
1175	1180	1185
Ile Thr Ile Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp		
1190	1195	1200
Phe Leu Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile		
1205	1210	1215
Ile Lys Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg		
1220	1225	1230
Lys Arg Met Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu		
1235	1240	1245
Leu Ala Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser		
1250	1255	1260
His Tyr Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys		
1265	1270	1275

Gln Leu Phe Val Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile	1280	1285	1290
Glu Gln Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala	1295	1300	1305
Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys	1310	1315	1320
Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu	1325	1330	1335
Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr	1340	1345	1350
Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala	1355	1360	1365
Thr Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile	1370	1375	1380
Asp Leu Ser Gln Leu Gly Gly Asp Gly Gly Ser Arg Ser Ala Glu	1385	1390	1395
Ile Gly Thr Gly Phe Pro Phe Asp Pro His Tyr Val Glu Val Leu	1400	1405	1410
Gly Glu Arg Met His Tyr Val Asp Val Gly Pro Arg Asp Gly Thr	1415	1420	1425
Pro Val Leu Phe Leu His Gly Asn Pro Thr Ser Ser Tyr Val Trp	1430	1435	1440
Arg Asn Ile Ile Pro His Val Ala Pro Thr His Arg Cys Ile Ala	1445	1450	1455
Pro Asp Leu Ile Gly Met Gly Lys Ser Asp Lys Pro Asp Leu Gly	1460	1465	1470
Tyr Phe Phe Asp Asp His Val Arg Phe Met Asp Ala Phe Ile Glu	1475	1480	1485
Ala Leu Gly Leu Glu Glu Val Val Leu Val Ile His Asp Trp Gly	1490	1495	1500
Ser Ala Leu Gly Phe His Trp Ala Lys Arg Asn Pro Glu Arg Val	1505	1510	1515
Lys Gly Ile Ala Phe Met Glu Phe Ile Arg Pro Ile Pro Thr Trp	1520	1525	1530
Asp Glu Trp Pro Glu Phe Ala Arg Glu Thr Phe Gln Ala Phe Arg	1535	1540	1545
Thr Thr Asp Val Gly Arg Lys Leu Ile Ile Asp Gln Asn Val Phe	1550	1555	1560
Ile Glu Gly Thr Leu Pro Met Gly Val Val Arg Pro Leu Thr Glu			

1565	1570	1575
Val Glu Met Asp His Tyr Arg Glu Pro Phe Leu Asn Pro Val Asp		
1580	1585	1590
Arg Glu Pro Leu Trp Arg Phe Pro Asn Glu Leu Pro Ile Ala Gly		
1595	1600	1605
Glu Pro Ala Asn Ile Val Ala Leu Val Glu Glu Tyr Met Asp Trp		
1610	1615	1620
Leu His Gln Ser Pro Val Pro Lys Leu Leu Phe Trp Gly Thr Pro		
1625	1630	1635
Gly Val Leu Ile Pro Pro Ala Glu Ala Ala Arg Leu Ala Lys Ser		
1640	1645	1650
Leu Pro Asn Cys Lys Ala Val Asp Ile Gly Pro Gly Leu Asn Leu		
1655	1660	1665
Leu Gln Glu Asp Asn Pro Asp Leu Ile Gly Ser Glu Ile Ala Arg		
1670	1675	1680
Trp Leu Ser Thr Leu Glu Ile Ser Gly Lys Arg Pro Ala Ala Thr		
1685	1690	1695
Lys Lys Ala Gly Gln Ala Lys Lys Lys Lys		
1700	1705	

<210> 5

<211> 30

<212> DNA

<213> 人工序列

<220>

<223> 合成的

<400> 5

tttcttgcca tattccacgt cctacagtgg 30

<210> 6

<211> 30

<212> DNA

<213> 人工序列

<220>

<223> 合成的

<400> 6

ttagggtag ggtaggggtt agggtaggg 30

<210> 7

<211> 27

<212> DNA

<213> 人工序列

- <220>
 <223> 合成的
 <400> 7
 actcatctaa tatgttctac agtgtgg 27
 <210> 8
 <211> 106
 <212> DNA
 <213> 人工序列
 <220>
 <223> 合成的
 <220>
 <221> 尚未归类的特征
 <222> (20) .. (20)
 <223> n为a、c、g或t
 <400> 8
 taatacgact cactataggn gtttaagagc tatgctggaa acagcatagc aagtttaaat 60
 aaggctagtc cgttatcaac ttgaaaaagt ggcaccgagt cgggtgc 106
 <210> 9
 <211> 36
 <212> RNA
 <213> 人工序列
 <220>
 <223> 合成的
 <400> 9
 ccauauucca cguccuacag guuuuagagc uaugcu 36
 <210> 10
 <211> 77
 <212> RNA
 <213> 人工序列
 <220>
 <223> 合成的
 <400> 10
 ggaaacagca uagcaaguuu aaauaaggeu aguccguuau caacuugaaa aaguggcacc 60
 gagucggugc uuuuuuuu 77
 <210> 11
 <211> 40
 <212> DNA
 <213> 人工序列
 <220>

<223> 合成的
 <400> 11
 aaacttgcta tgctgtttcc agcatagctc ttaaacctgt 40
 <210> 12
 <211> 154
 <212> DNA
 <213> 人工序列
 <220>
 <223> 合成的
 <400> 12
 ggtgacacta tagaactcga gcagctggat cctaatacga ctactatag gccatattcc 60
 acgtacaggt ttaagageta tgctggaaac agcatagcaa gtttaaataa ggctagtccg 120
 ttatcaactt gaaaaagtgg caccgagtcg gtgc 154
 <210> 13
 <211> 159
 <212> DNA
 <213> 人工序列
 <220>
 <223> 合成的
 <400> 13
 ggtgacacta tagaactcga gcagctggat cctaatacga ctactatag ggtagggtt 60
 agggtaggg ttagtttaag agctatgctg gaaacagcat agcaagtta aataaggcta 120
 gtccgttata aacttgaaaa agtggcaccg agtcgggtgc 159
 <210> 14
 <211> 157
 <212> DNA
 <213> 人工序列
 <220>
 <223> 合成的
 <400> 14
 ggtgacacta tagaactcga gcagctggat cctaatacga ctactatag gatctattat 60
 gttctacagt ggtttaagag ctatgctgga aacagcatag caagtttaa taaggctagt 120
 ccgttatcaa cttgaaaaag tggcaccgag tcggtgc 157
 <210> 15
 <211> 155
 <212> DNA
 <213> 人工序列
 <220>
 <223> 合成的

<400> 15
 ggtgacacta tagaactcga gcagctggat cctaatacga ctactatag gtcgactcta 60
 gaaaacatgg ttaagagct atgctggaaa cagcatagca agtttaaata aggctagtcc 120
 gttatcaact tgaaaaagtg gcaccgagtc ggtgc 155

<210> 16
 <211> 38
 <212> DNA
 <213> 人工序列
 <220>
 <223> 合成的
 <220>
 <221> 尚未归类的特征
 <222> (20) .. (20)
 <223> n为a、c、g或t

<400> 16
 taatacgact cactataggn gtttaagagc tatgctgg 38
 <210> 17
 <211> 37
 <212> DNA
 <213> 人工序列
 <220>
 <223> 合成的
 <400> 17
 taatacgact cactataggg ttaagagct atgctgg 37
 <210> 18
 <211> 58
 <212> DNA
 <213> 人工序列
 <220>
 <223> 合成的
 <400> 18
 taatacgact cactatagga ggtgacaccg tgggctgggg gtttaagagc tatgctgg 58
 <210> 19
 <211> 54
 <212> DNA
 <213> 人工序列
 <220>
 <223> 合成的
 <400> 19

taatacgact cactatagga agtgtcgaca ggaagagttt aagagctatg ctgg 54

<210> 20

<211> 59

<212> DNA

<213> 人工序列

<220>

<223> 合成的

<400> 20

taatacgact cactatagga aggtatgggt gtggaaggta tgtttaagag ctatgctgg 59

<210> 21

<211> 58

<212> DNA

<213> 人工序列

<220>

<223> 合成的

<400> 21

taatacgact cactatagga agagtggagg ccgtgcgcgg gtttaagagc tatgctgg 58

<210> 22

<211> 58

<212> DNA

<213> 人工序列

<220>

<223> 合成的

<400> 22

taatacgact cactatagga cctcaggtaga tctcctgect gtttaagagc tatgctgg 58

<210> 23

<211> 54

<212> DNA

<213> 人工序列

<220>

<223> 合成的

<400> 23

taatacgact cactataggt atatttagta gagacggttt aagagctatg ctgg 54

<210> 24

<211> 57

<212> DNA

<213> 人工序列

<220>

<223> 合成的

<400> 24
taatacgact cactatagga gtagctggaa ttacaggtgg ttttaagagct atgctgg 57
<210> 25
<211> 59
<212> DNA
<213> 人工序列
<220>
<223> 合成的
<400> 25
taatacgact cactatagge tcaactgcaac ctccacctcc cgtttaagag ctatgctgg 59
<210> 26
<211> 59
<212> DNA
<213> 人工序列
<220>
<223> 合成的
<400> 26
taatacgact cactatagga cagagtctcg ctctctctcc cgtttaagag ctatgctgg 59
<210> 27
<211> 58
<212> DNA
<213> 人工序列
<220>
<223> 合成的
<400> 27
taatacgact cactatagga agaggagaaa agtggggaag gtttaagagc tatgctgg 58
<210> 28
<211> 59
<212> DNA
<213> 人工序列
<220>
<223> 合成的
<400> 28
taatacgact cactatagga acagagggcc agagagcagc cgtttaagag ctatgctgg 59
<210> 29
<211> 60
<212> DNA
<213> 人工序列
<220>

<223> 合成的
<400> 29
taatacgact cactataggt ctttctctct gcgagtaage ctgtttaaga gctatgctgg 60
<210> 30
<211> 58
<212> DNA
<213> 人工序列
<220>
<223> 合成的
<400> 30
taatacgact cactataggt acaccettgt gtacagagct gtttaagage tatgctgg 58
<210> 31
<211> 57
<212> DNA
<213> 人工序列
<220>
<223> 合成的
<400> 31
taatacgact cactatagga aaactcatgt aaagctgcag ttttaagagct atgctgg 57
<210> 32
<211> 58
<212> DNA
<213> 人工序列
<220>
<223> 合成的
<400> 32
taatacgact cactatagge aagcaaggga agcgacaagg gtttaagage tatgctgg 58
<210> 33
<211> 55
<212> DNA
<213> 人工序列
<220>
<223> 合成的
<400> 33
taatacgact cactatagga ggcggccagg gcgcagagtt taagagctat gctgg 55
<210> 34
<211> 56
<212> DNA
<213> 人工序列

<220>

<223> 合成的

<400> 34

taatacgact cactatagge tttaaacc gagctcaggt ttaagagcta tgctgg 56

<210> 35

<211> 56

<212> DNA

<213> 人工序列

<220>

<223> 合成的

<400> 35

taatacgact cactataggt agccccgca ttggcettgt ttaagagcta tgctgg 56

<210> 36

<211> 57

<212> DNA

<213> 人工序列

<220>

<223> 合成的

<400> 36

taatacgact cactatagge ctgtgggaga tgttcctcg ttaagagct atgctgg 57

<210> 37

<211> 59

<212> DNA

<213> 人工序列

<220>

<223> 合成的

<400> 37

taatacgact cactataggt cctgaagcca gagggacagc cgtttaagag ctatgctgg 59

<210> 38

<211> 56

<212> DNA

<213> 人工序列

<220>

<223> 合成的

<400> 38

taatacgact cactatagga gtctttgggg gagagtctgt ttaagagcta tgctgg 56

<210> 39

<211> 57

<212> DNA

<213> 人工序列
<220>
<223> 合成的
<400> 39
taatacgact cactatagge tcctgcctg cctctcagcg ttaagagct atgctgg 57
<210> 40
<211> 56
<212> DNA
<213> 人工序列
<220>
<223> 合成的
<400> 40
taatacgact cactatagge atatttgagg agettctctgt ttaagagcta tgctgg 56
<210> 41
<211> 55
<212> DNA
<213> 人工序列
<220>
<223> 合成的
<400> 41
taatacgact cactatagge tgcaagagaa gccatgcggt taagagctat gctgg 55
<210> 42
<211> 57
<212> DNA
<213> 人工序列
<220>
<223> 合成的
<400> 42
taatacgact cactatagga tgtttcagga ctaggetgag ttaagagct atgctgg 57
<210> 43
<211> 57
<212> DNA
<213> 人工序列
<220>
<223> 合成的
<400> 43
taatacgact cactatagge tggagggtgg ggaggtgtag ttaagagct atgctgg 57
<210> 44
<211> 55

<212> DNA
<213> 人工序列
<220>
<223> 合成的
<400> 44
taatacgact cactataggt gggatgagca ctggagcgtt taagagctat gctgg 55
<210> 45
<211> 55
<212> DNA
<213> 人工序列
<220>
<223> 合成的
<400> 45
taatacgact cactatagge cctgcagatg tggttgagtt taagagctat gctgg 55
<210> 46
<211> 57
<212> DNA
<213> 人工序列
<220>
<223> 合成的
<400> 46
taatacgact cactatagga ggctggggct tggggcgccg ttaagagct atgctgg 57
<210> 47
<211> 56
<212> DNA
<213> 人工序列
<220>
<223> 合成的
<400> 47
taatacgact cactataggt ctttgccgtg aactgttctg ttaagagcta tgctgg 56
<210> 48
<211> 58
<212> DNA
<213> 人工序列
<220>
<223> 合成的
<400> 48
taatacgact cactatagga ccggggccct ggggagacac gtttaagagc tatgctgg 58
<210> 49

<211> 56

<212> DNA

<213> 人工序列

<220>

<223> 合成的

<400> 49

taatacgact cactatagge tggacactca gctccatggt ttaagagcta tgctgg 56

<210> 50

<211> 57

<212> DNA

<213> 人工序列

<220>

<223> 合成的

<400> 50

taatacgact cactatagga ggcgagaggg gcaagacctg ttaagagct atgctgg 57

<210> 51

<211> 56

<212> DNA

<213> 人工序列

<220>

<223> 合成的

<400> 51

taatacgact cactatagga gaaggagtga aggactgtgt ttaagagcta tgctgg 56

<210> 52

<211> 61

<212> DNA

<213> 人工序列

<220>

<223> 合成的

<400> 52

taatacgact cactatagge tccacgacat gctagcttc ttcgtttaag agctatgctg 60
g 61

<210> 53

<211> 57

<212> DNA

<213> 人工序列

<220>

<223> 合成的

<400> 53

taatacgact cactatagga gctgggccag gagaggagag ttttaagagct atgctgg 57

<210> 54

<211> 57

<212> DNA

<213> 人工序列

<220>

<223> 合成的

<400> 54

taatacgact cactatagga ccgggcatga ccagggctg ttttaagagct atgctgg 57

<210> 55

<211> 55

<212> DNA

<213> 人工序列

<220>

<223> 合成的

<400> 55

taatacgact cactataggg cagccccac cccacagtt taagagctat gctgg 55

<210> 56

<211> 57

<212> DNA

<213> 人工序列

<220>

<223> 合成的

<400> 56

taatacgact cactataggt tccttttggc tcctgaagg ttttaagagct atgctgg 57

<210> 57

<211> 55

<212> DNA

<213> 人工序列

<220>

<223> 合成的

<400> 57

taatacgact cactataggg tetgtttgca cacttgcgtt taagagctat gctgg 55

<210> 58

<211> 59

<212> DNA

<213> 人工序列

<220>

<223> 合成的

- <400> 58
taatacgact cactatagge ccaggccaga ggaaaaacac agtttaagag ctatgctgg 59
- <210> 59
<211> 56
<212> DNA
<213> 人工序列
<220>
<223> 合成的
- <400> 59
taatacgact cactataggt ttcettaagg aacageccgt ttaagagcta tgctgg 56
- <210> 60
<211> 58
<212> DNA
<213> 人工序列
<220>
<223> 合成的
- <400> 60
taatacgact cactatagge agacagaggt gggctagaca gtttaagagc tatgctgg 58
- <210> 61
<211> 59
<212> DNA
<213> 人工序列
<220>
<223> 合成的
- <400> 61
taatacgact cactatagge cccaggcagg aatgactcag agtttaagag ctatgctgg 59
- <210> 62
<211> 56
<212> DNA
<213> 人工序列
<220>
<223> 合成的
- <400> 62
taatacgact cactatagga cccagttgce tttcctggt ttaagagcta tgctgg 56
- <210> 63
<211> 57
<212> DNA
<213> 人工序列
<220>

<223> 合成的
<400> 63
taatacgact cactatagga cccagaggag gtgacaggcg ttaagagct atgctgg 57
<210> 64
<211> 59
<212> DNA
<213> 人工序列
<220>
<223> 合成的
<400> 64
taatacgact cactatagge cacagcgcac tccacgggga agtttaagag ctatgctgg 59
<210> 65
<211> 58
<212> DNA
<213> 人工序列
<220>
<223> 合成的
<400> 65
taatacgact cactataggt cccagactga cagatagacc gtttaagage tatgctgg 58
<210> 66
<211> 56
<212> DNA
<213> 人工序列
<220>
<223> 合成的
<400> 66
taatacgact cactatagga ggggtctgtg gagagtttgt ttaagagcta tgctgg 56
<210> 67
<211> 57
<212> DNA
<213> 人工序列
<220>
<223> 合成的
<400> 67
taatacgact cactataggt ccagcatcag cgacgcctg ttaagagct atgctgg 57
<210> 68
<211> 57
<212> DNA
<213> 人工序列

<220>

<223> 合成的

<400> 68

taatacgact cactatagge ctaagactcc agagccaaag ttaagagct atgctgg 57

<210> 69

<211> 58

<212> DNA

<213> 人工序列

<220>

<223> 合成的

<400> 69

taatacgact cactatagge tactacgtag ggttgcatg gtttaagagc tatgctgg 58

<210> 70

<211> 57

<212> DNA

<213> 人工序列

<220>

<223> 合成的

<400> 70

taatacgact cactataggt aaagtagaaa aggcataaag ttaagagct atgctgg 57

<210> 71

<211> 56

<212> DNA

<213> 人工序列

<220>

<223> 合成的

<400> 71

taatacgact cactatagge acttttggag gccagcgt ttaagagcta tgctgg 56

<210> 72

<211> 58

<212> DNA

<213> 人工序列

<220>

<223> 合成的

<400> 72

taatacgact cactataggt ggagacaggg ttggccaage gtttaagagc tatgctgg 58

<210> 73

<211> 59

<212> DNA

<213> 人工序列
<220>
<223> 合成的
<400> 73
taatacgact cactatagge tccctgcaac ctctgectcc cgtttaagag ctatgctgg 59
<210> 74
<211> 59
<212> DNA
<213> 人工序列
<220>
<223> 合成的
<400> 74
taatacgact cactatagge ctetttetca aacacgtett tagtttaaga gctatgctg 59
<210> 75
<211> 58
<212> DNA
<213> 人工序列
<220>
<223> 合成的
<400> 75
taatacgact cactatagga acccggaatg gcacttgtgt gtttaagage tatgctgg 58
<210> 76
<211> 55
<212> DNA
<213> 人工序列
<220>
<223> 合成的
<400> 76
taatacgact cactataggt ggcttttttag aggcacggtt taagagctat gctgg 55
<210> 77
<211> 55
<212> DNA
<213> 人工序列
<220>
<223> 合成的
<400> 77
taatacgact cactatagge ttggtgtatt cagaatggtt taagagctat gctgg 55
<210> 78
<211> 59

<212> DNA
<213> 人工序列
<220>
<223> 合成的
<400> 78
taatacgact cactatagge actgccagge cagcctctgc cgtttaagag ctatgctgg 59
<210> 79
<211> 58
<212> DNA
<213> 人工序列
<220>
<223> 合成的
<400> 79
taatacgact cactatagge taaggacaag aggcaatgag gtttaagagc tatgctgg 58
<210> 80
<211> 60
<212> DNA
<213> 人工序列
<220>
<223> 合成的
<400> 80
taatacgact cactatagga ctcaatttct cagaacatgc tggtttaaga gctatgctgg 60
<210> 81
<211> 59
<212> DNA
<213> 人工序列
<220>
<223> 合成的
<400> 81
taatacgact cactatagga cagagtttct ctctgteccc cgtttaagag ctatgctgg 59
<210> 82
<211> 55
<212> DNA
<213> 人工序列
<220>
<223> 合成的
<400> 82
taatacgact cactataggg gtttcacat gttggccggt taagagctat gctgg 55
<210> 83

<211> 59

<212> DNA

<213> 人工序列

<220>

<223> 合成的

<400> 83

taatacgact cactataggc tcgcctcggc tcccaaagtg cgtttaagag ctatgctgg 59

<210> 84

<211> 55

<212> DNA

<213> 人工序列

<220>

<223> 合成的

<400> 84

taatacgact cactataggg catttggtt gcacgtggtt taagagctat gctgg 55

<210> 85

<211> 57

<212> DNA

<213> 人工序列

<220>

<223> 合成的

<400> 85

taatacgact cactatagga gtggagctgc gggcaaccg ttaagagct atgctgg 57

<210> 86

<211> 56

<212> DNA

<213> 人工序列

<220>

<223> 合成的

<400> 86

taatacgact cactataggt agagatgccg cccgcccgt ttaagagcta tgctgg 56

<210> 87

<211> 58

<212> DNA

<213> 人工序列

<220>

<223> 合成的

<400> 87

taatacgact cactataggt ccagtgcca gtggattttg gtttaagagc tatgctgg 58

- <210> 88
<211> 58
<212> DNA
<213> 人工序列
<220>
<223> 合成的
<400> 88
taatacgact cactatagga ggcagctggg actagaacc gtttaagagc tatgctgg 58
- <210> 89
<211> 56
<212> DNA
<213> 人工序列
<220>
<223> 合成的
<400> 89
taatacgact cactataggt cggtaggctg ggctggttgt ttaagagcta tgctgg 56
- <210> 90
<211> 57
<212> DNA
<213> 人工序列
<220>
<223> 合成的
<400> 90
taatacgact cactatagga atgaatggct gtctcagcag ttaagagct atgctgg 57
- <210> 91
<211> 59
<212> DNA
<213> 人工序列
<220>
<223> 合成的
<400> 91
taatacgact cactatagga aacagacgtg gccagctctc tgtttaagag ctatgctgg 59
- <210> 92
<211> 56
<212> DNA
<213> 人工序列
<220>
<223> 合成的
<400> 92

taatacgact cactatagge tgagagctgc atttcgaagt ttaagagcta tgctgg 56

<210> 93

<211> 59

<212> DNA

<213> 人工序列

<220>

<223> 合成的

<400> 93

taatacgact cactatagga caagtcagga agggcctgt ggtttaagag ctatgctgg 59

<210> 94

<211> 86

<212> DNA

<213> 人工序列

<220>

<223> 合成的

<400> 94

gcaccgactc ggtgccactt tttcaagttg ataacggact agccttattt aaacttgcta 60

tgctgtttcc agcatagctc ttaaac 86

<210> 95

<211> 20

<212> DNA

<213> 人工序列

<220>

<223> 合成的

<400> 95

gcaccgactc ggtgccactt 20

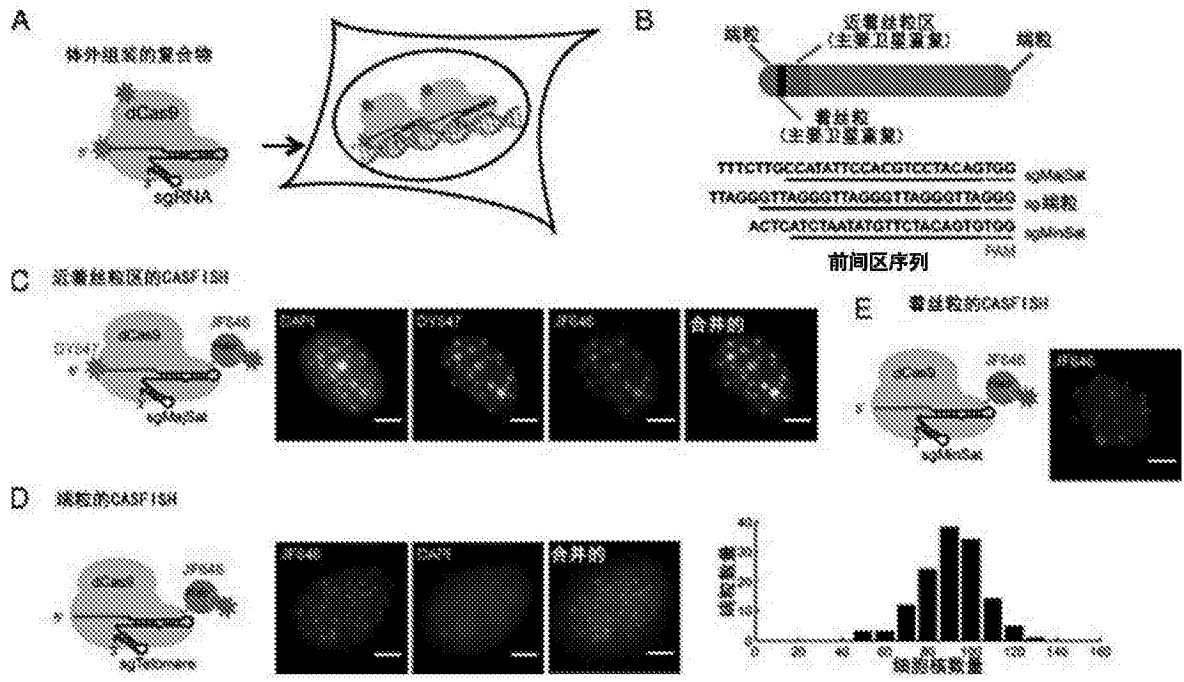


图1A-1E

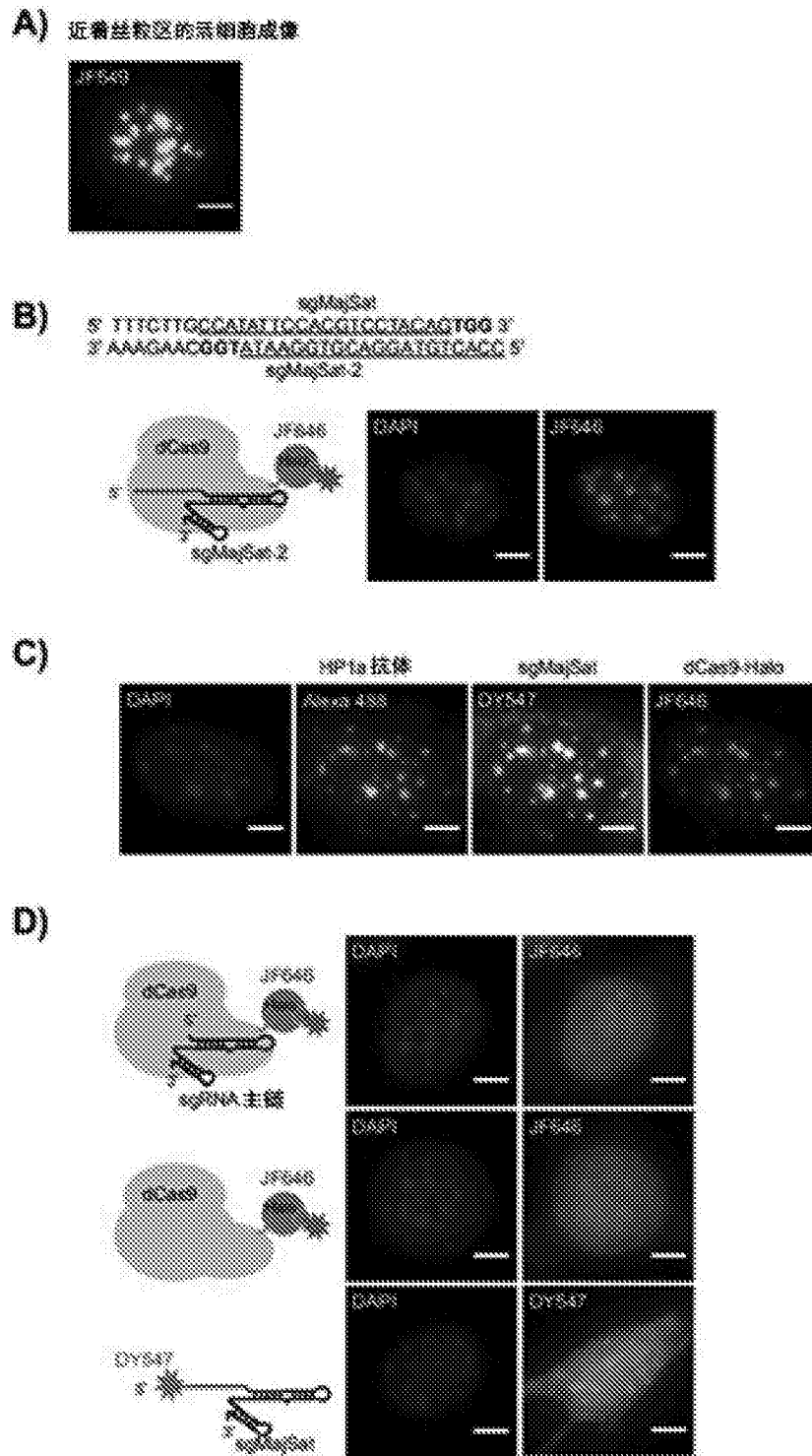


图2A-2D

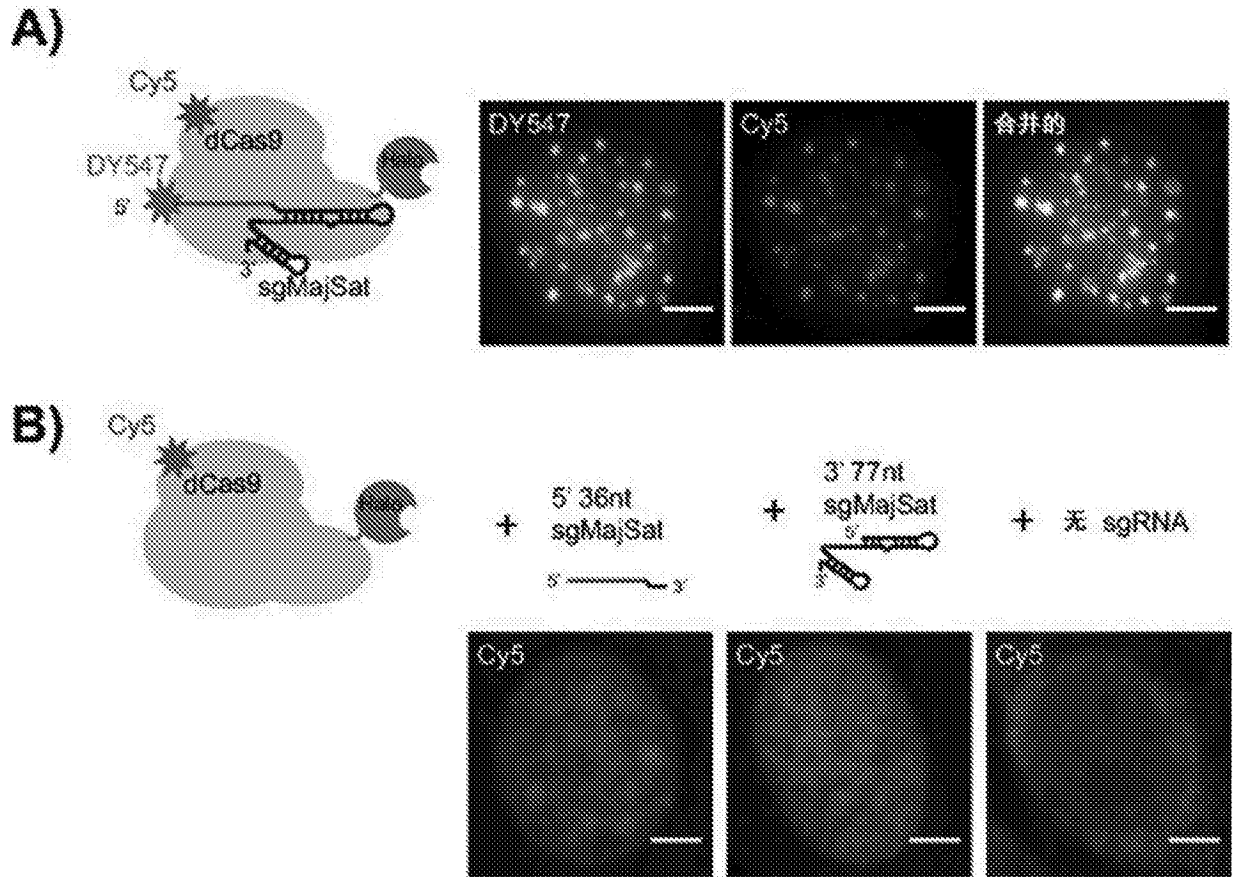
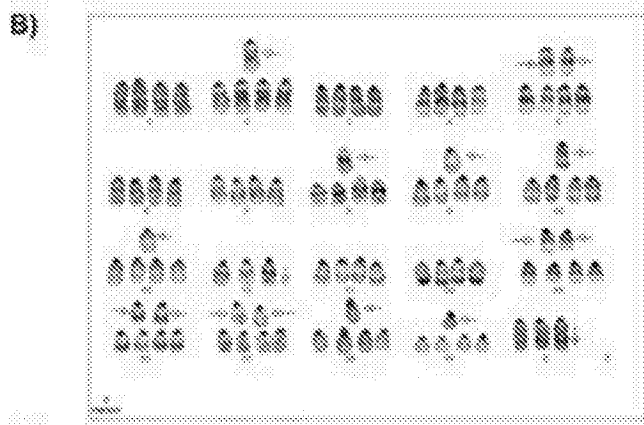
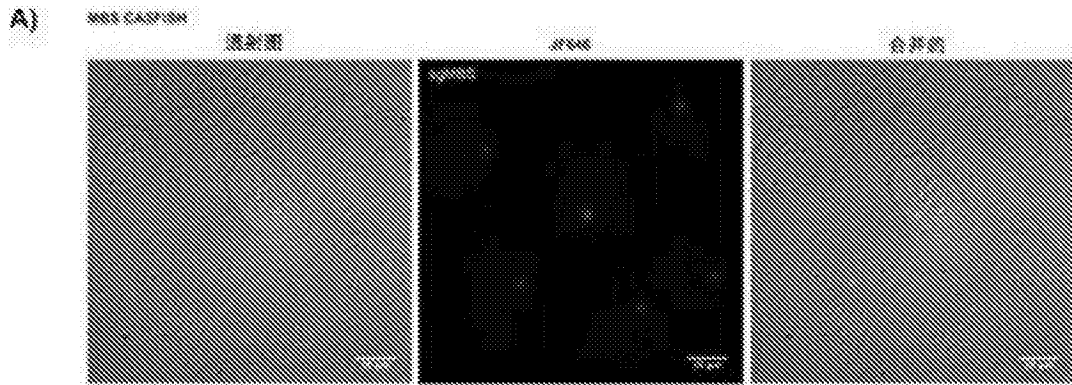


图3A-3B



C)

中期染色体计数	复合中期染色体组型
85	XXX,X,+2,+5,+6,7,+8,+10,+11,+13,+13,+13,+15,+15,+16,+16,+17,+17,+18,+19
73	XX,X,X,X,X,X,7,7,+10,+12,+16,+17,+19
85	XXX,X,+1,+2,+5,+5,7,+8,+9,+10,+11,+15,+15,+16,+16,+17,+18,+19
122	XXX,X,+1,+1,+1,+2,+2,+4,+4,+5,+5,+8,+8,+10,+10,+10,+10,+11,+11,+12,+13,+13,+13,+13,+15,+15,+15,+15,+16,+16,+16,+16,+17,+17,+17,+17,+18,+18,+18,+18,+19,+19
85	XXX,X,+2,+5,+5,7,+8,+8,+8,+10,+11,+15,+15,+16,+16,+17,+17,+18,+18,+19
84	XXX,X,+2,+2,+4,+5,+5,+7,+10,+11,+12,+13,+15,+16,+16,+17,+19
81	XXX,X,+2,+2,+6,7,+8,+10,+10,+11,+13,+13,+15,+15,+16,+16,+18,+18,+19
84	XXX,X,+2,+2,+5,+5,7,+8,+10,+11,+13,+13,+15,+15,+16,+16,+17,+19
84	XXX,X,+2,+5,+5,+8,+10,+11,+12,+15,+15,+16,+16,+17,+17,+18,+19
82	XXX,X,+2,+5,+5,7,+8,+10,+11,+13,+15,+15,+16,+16,+17,+17,+18
平均值 = 84.5	

图4A-4C

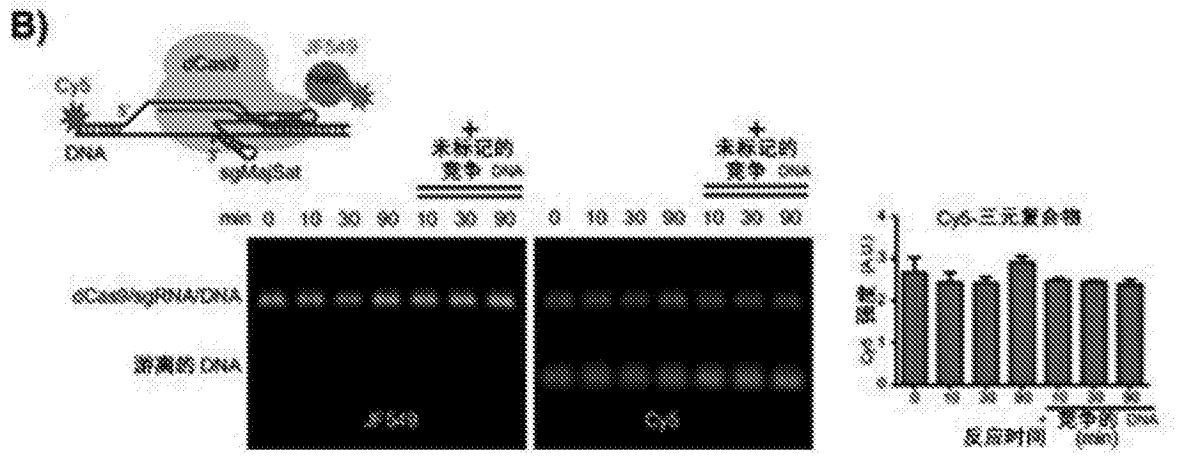
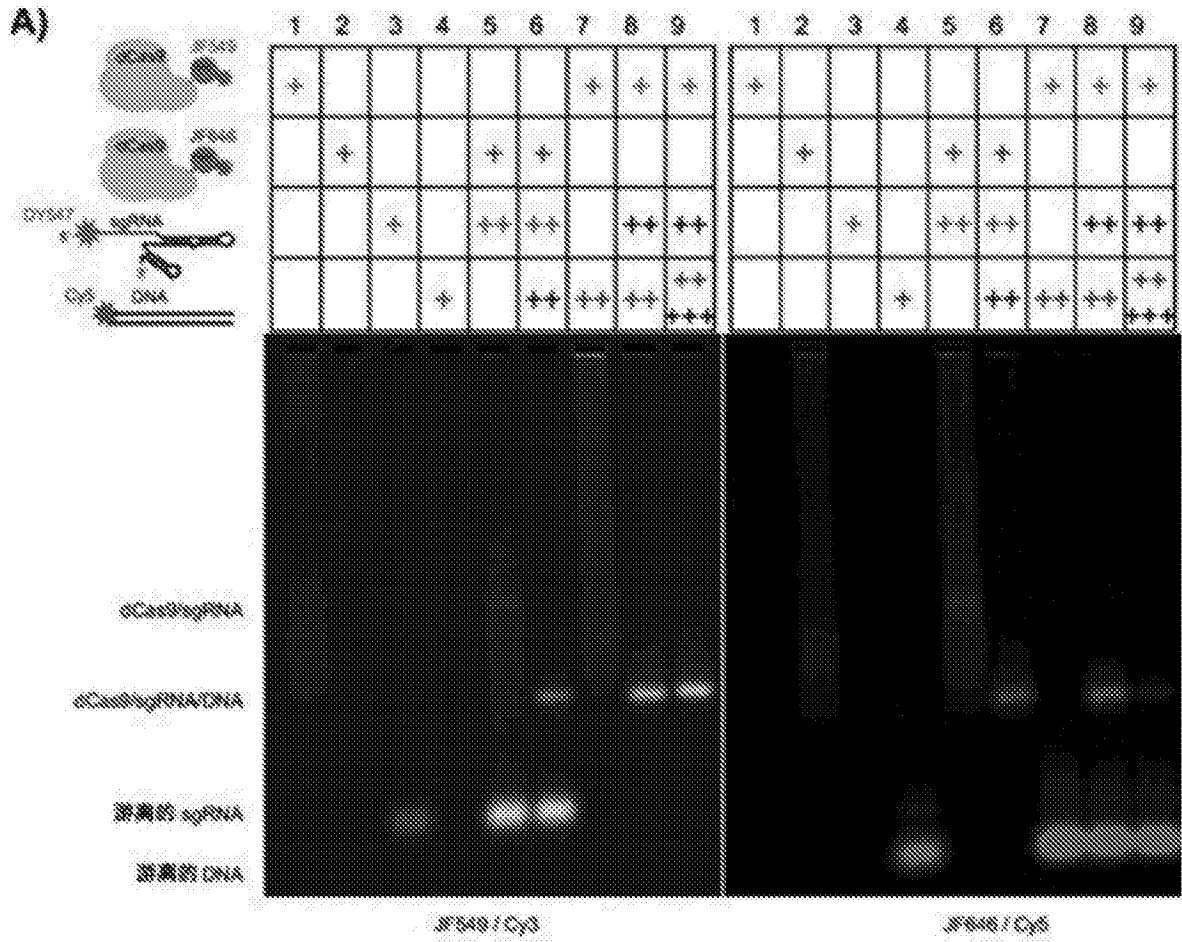


图5A-5B

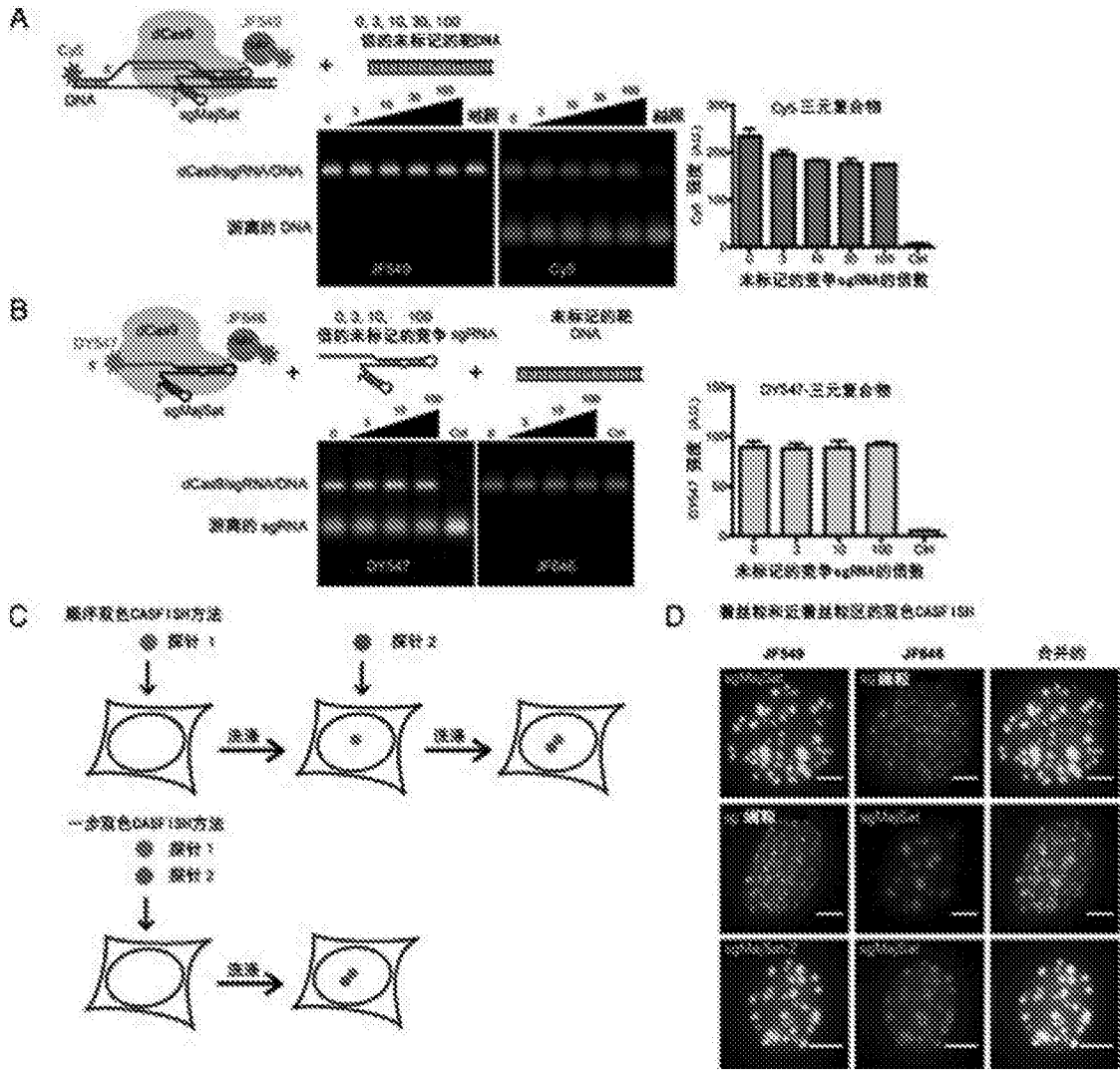


图6A-6D

着丝粒和近着丝粒区的一步双色CASFISH

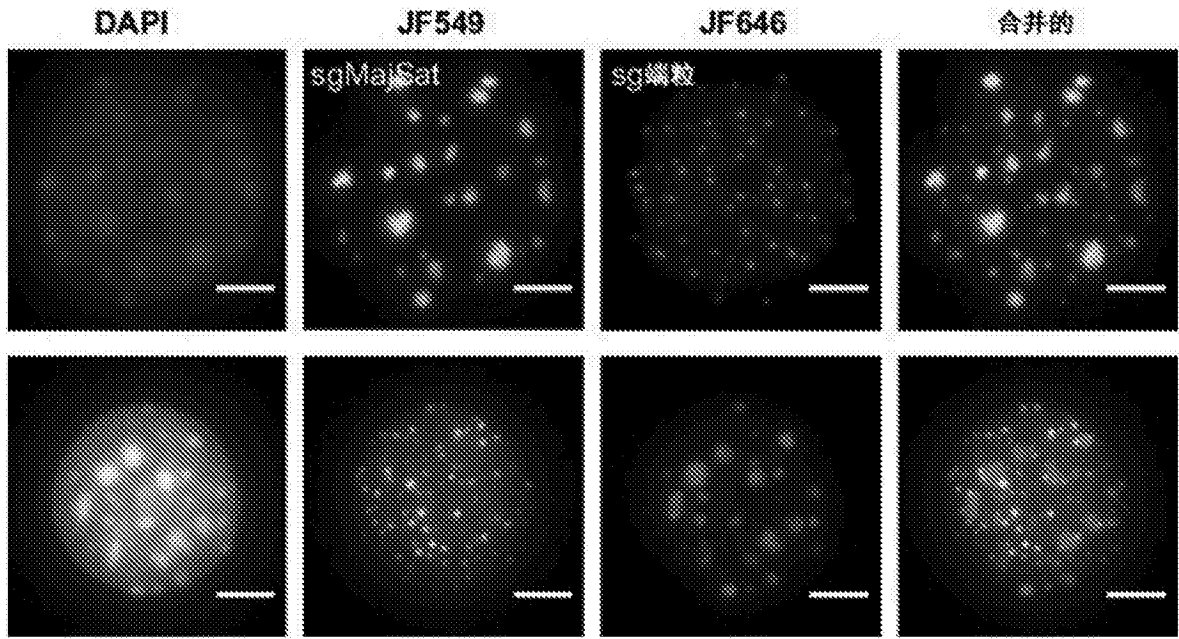


图7

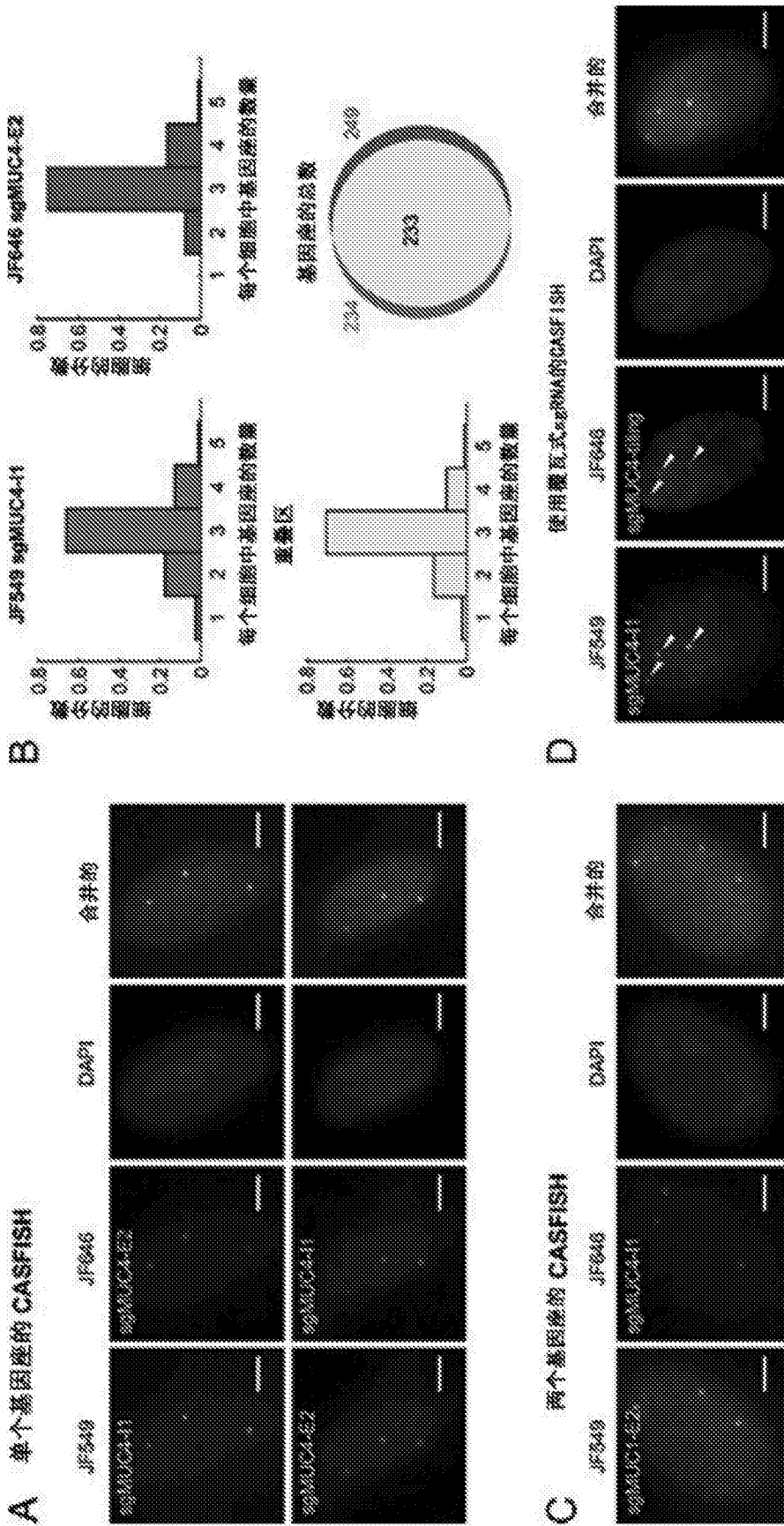
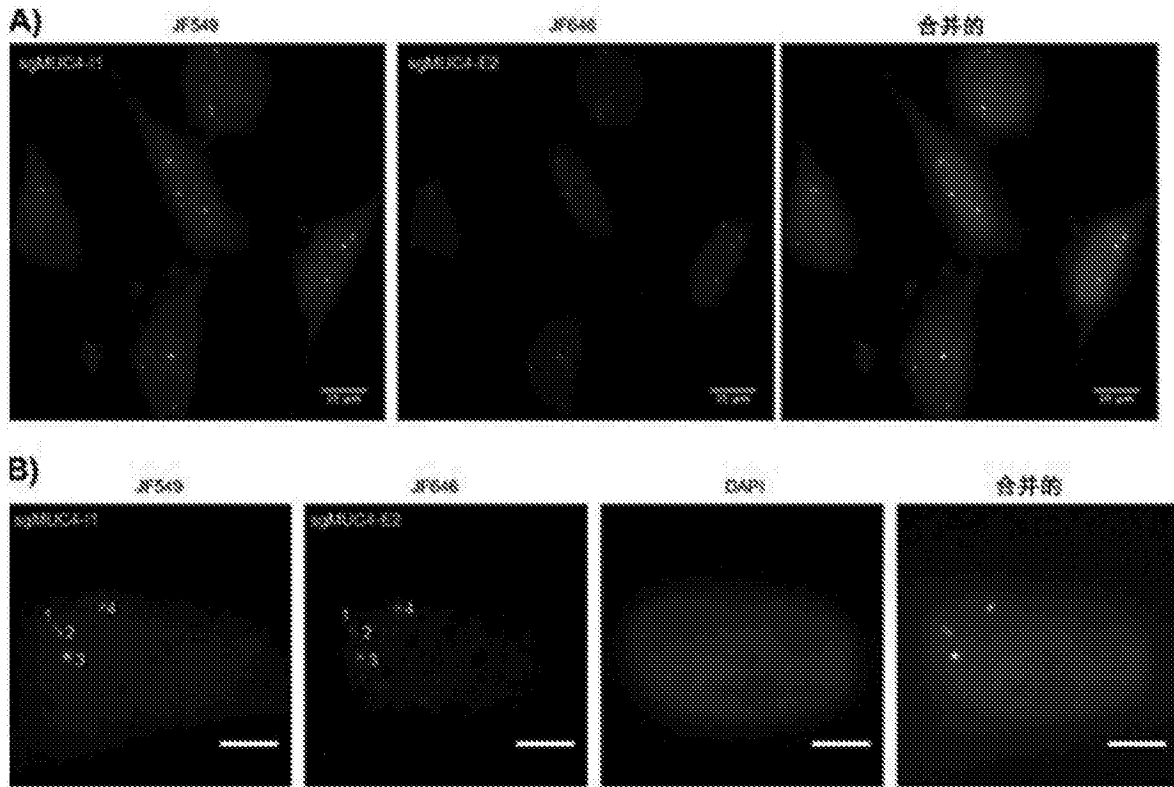


图8A-8D



荧光	通道 #	通道强度
JF548	1	1474
	2	2183
	3	7885
	4	3884
JF646	1	383
	2	231
	3	1515
	4	1428

图9A-9B

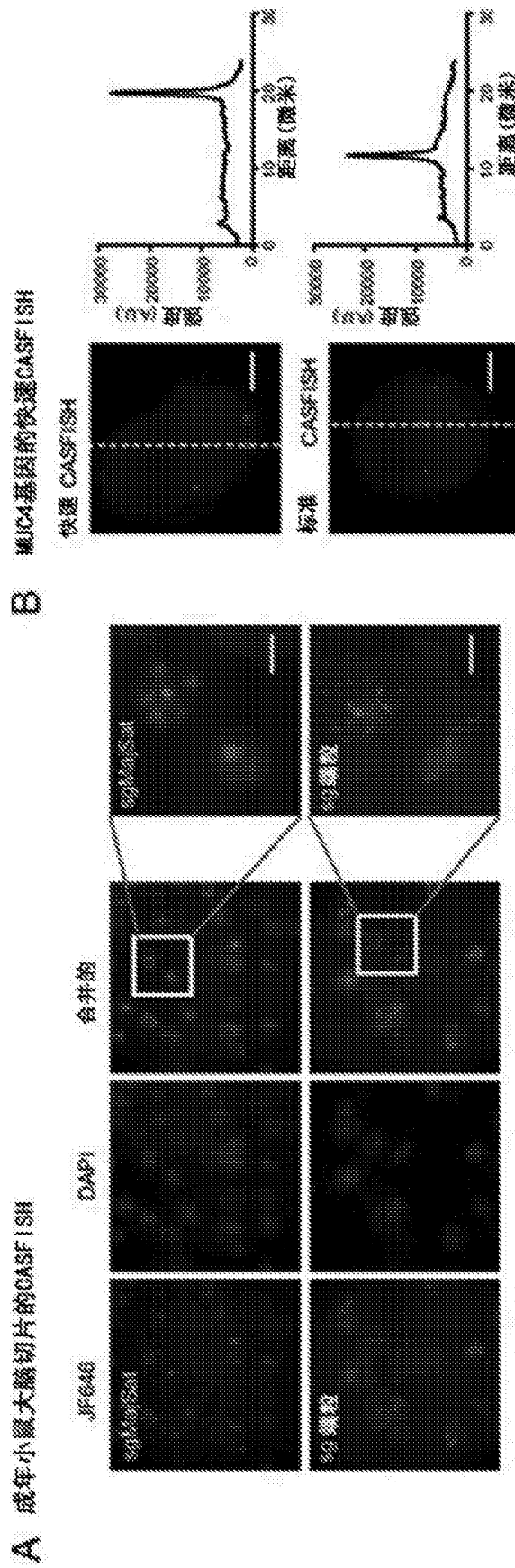


图10A-10B

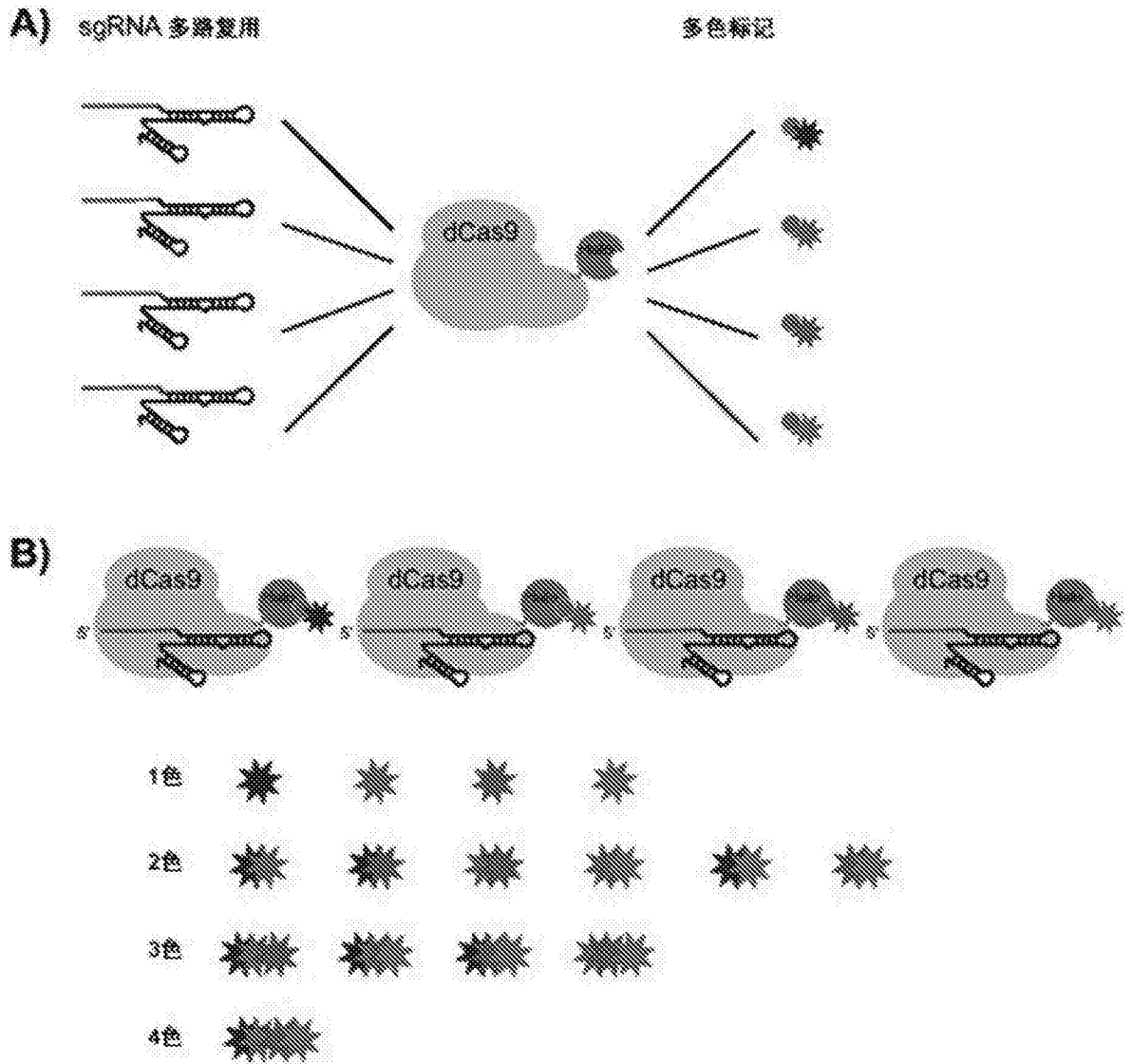


图11A-11B