



(12) 发明专利申请

(10) 申请公布号 CN 118542949 A

(43) 申请公布日 2024. 08. 27

(21) 申请号 202410368500.4	A61P 29/00 (2006.01)
(22) 申请日 2014.09.05	A61P 37/02 (2006.01)
(30) 优先权数据	A61P 31/00 (2006.01)
61/874,746 2013.09.06 US	A61P 25/00 (2006.01)
14/462,163 2014.08.18 US	A61P 9/10 (2006.01)
14/462,189 2014.08.18 US	A61P 7/02 (2006.01)
(62) 分案原申请数据	A61P 7/00 (2006.01)
201480060598.1 2014.09.05	A61P 35/00 (2006.01)
(71) 申请人 哈佛大学的校长及成员们	A61P 11/00 (2006.01)
地址 美国马萨诸塞州	A61P 1/00 (2006.01)
(72) 发明人 D·R·刘 J·A·祖里斯	A61P 1/04 (2006.01)
D·B·汤普森	A61P 19/02 (2006.01)
(74) 专利代理机构 北京市柳沈律师事务所	A61P 19/10 (2006.01)
11105	A61P 13/12 (2006.01)
专利代理师 涂滔	A61P 25/24 (2006.01)
(51) Int. Cl.	A61P 25/18 (2006.01)
A61K 47/64 (2017.01)	A61P 17/00 (2006.01)
A61K 38/46 (2006.01)	G12N 9/22 (2006.01)

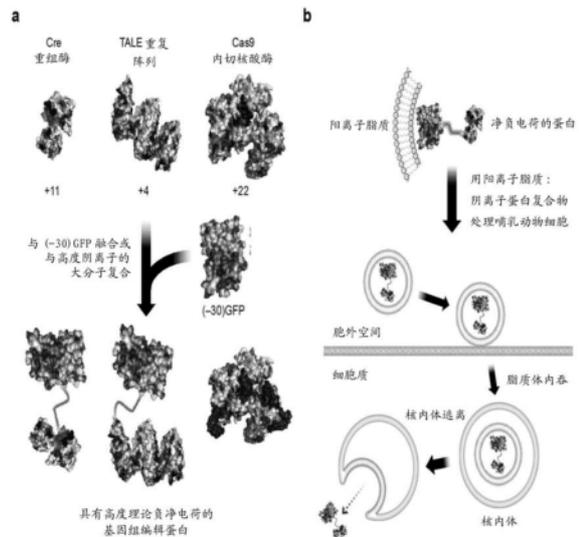
权利要求书1页 说明书108页
序列表(电子公布) 附图24页

(54) 发明名称

用于功能性核酸酶的投递系统

(57) 摘要

提供了用于超电荷蛋白介导的功能性效应器蛋白对细胞的体内、离体、或体外投递的组合物、方法、策略、试剂盒和系统。还提供了使用阳离子脂质和阳离子聚合物来投递功能性效应器蛋白的组合物、方法、策略、试剂盒和系统。功能性效应器蛋白包括但不限于转录调控物(例如阻遏物或激活物)、重组酶、核酸酶(例如可RNA编程的核酸酶,如Cas9蛋白;TALE核酸酶,和锌指核酸酶)、脱氨酶,和其他基因修饰/编辑酶。功能性效应器蛋白包括TALE效应器蛋白,例如TALE转录激活物和阻遏物,以及TALE核酸酶。用于将功能性效应器蛋白投递到细胞中的组合物、方法、策略和系统可用于治疗和研究目的,包括但不限于与疾病有关的基因的靶向性操作,与疾病有关的基因的表达水平的调控,和细胞命运的编程。



1. 一种包含与功能性效应器蛋白联合的超电荷蛋白的组合物,其中所述超电荷蛋白具有大于其相应未经修饰蛋白的总体正电荷并且为充足的量,并且配制为用于穿透入细胞中。
2. 权利要求1的组合物,其中所述功能性效应器蛋白是核酸酶。
3. 权利要求2的组合物,其中所述核酸酶是可RNA编程的核酸酶。
4. 权利要求3的组合物,其中所述核酸酶是TALE核酸酶、Cas9核酸酶或锌指核酸酶。
5. 权利要求2-4中任一项的组合物,其中所述核酸酶特异性结合并切割核酸序列。
6. 权利要求5的组合物,其中所述核酸序列包含在与疾病或病症有关的基因或等位基因中。
7. 权利要求5或6的组合物,其中所述组合物还包含核酸分子,所述核酸分子包含与由所述核酸酶结合并切割的核酸序列相同或同源的序列。
8. 权利要求中5-7任一项的组合物,其中所述核酸序列是包含在基因中的序列,所述基因是治疗靶物。
9. 权利要求8的组合物,其中所述基因是控制细胞命运的基因。
10. 权利要求9的组合物,其中所述基因是诱导或抑制朝向褐色脂肪细胞细胞命运的细胞程序化的基因。

用于功能性核酸酶的投递系统

[0001] 本申请是基于申请日为2014年9月5日,优先权日为2013年9月6日,申请号为201480060598.1,发明名称为“用于功能性核酸酶的投递系统”的专利申请的分案申请。

[0002] 相关申请

[0003] 本申请要求35U.S.C.§365(c)下2014年8月18日提交的美国申请U.S.S.N.14/462,189和2014年8月18日提交的美国申请U.S.S.N.14/462,163的优先权,其还要求35U.S.C.§119(e)下2013年9月6日提交的美国临时专利申请U.S.S.N.61/874,746的优先权,其各自通过提述并入本文。

[0004] 发明背景

[0005] 将大分子投递到哺乳动物细胞中是用于细胞操作的一种有吸引力的办法,因为其将允许调控基因表达和修饰基因组,这相应地将开启研究和实现对目前视为通过小分子“不可给药”的分子的靶向性治疗的新途径。具体而言,靶向与疾病有关的基因或等位基因的重组核酸酶具有作为治疗剂的极大潜力。目前大分子投递的方法包括核酸分子的病毒投递,核酸或蛋白质的受体介导的投递,和具有细胞穿透肽如TAT、Arg9或Penetratin用于投递蛋白质的蛋白融合物的使用。这些投递系统的每一种提供了对特定应用的益处;然而,在大多数情况下,仍存在关于功效、细胞毒性和制备容易性的问题。能够将大分子(例如功能性效应器蛋白)有效投递到多种细胞系而没有显著细胞毒性或其他不良副作用的容易制备的试剂仍是重要的关注点所在。

[0006] 大多数蛋白质不自发进入哺乳动物细胞中,因此其用作研究工具和其作为治疗剂的潜力受到天然限制。最近开发了用于将蛋白质投递到哺乳动物细胞中的技术来针对胞内靶物。这些技术包括使用基于脂质的试剂(Zelphati等,J.Biol.Chem.276,35103-35110,2001),纳米颗粒(Hasadsri等,J.Biol.Chem.,2009),vault核糖核蛋白颗粒(Lai等,ACS Nano 3,691-699,2009);对受体配体的遗传或化学融合(Gabel等,J.Cell Biol.103,1817-1827,1986;Rizk等,Proc.Natl.Acad.Sci.U.S.A.106,11011-11015,2009);和对细胞穿透肽的融合(Wadia等,Curr.Protein Pept.Sci.4,97-104,2003;Zhou等,Cell Stem Cell 4,381-384,2009)。可能用于蛋白质投递的最常见的方法是对蛋白转导域(PTD)的遗传融合,所述转导域包含HIV-1转录反式激活物(transactivator of transcription)(Tat)肽和聚精氨酸肽。这些阳离子PTD促进与带负电荷的细胞表面结构的关联和后续对外源蛋白的胞吞作用。Tat和聚精氨酸两者均已被用于将多种大分子体外和体内投递到细胞中(Wadia等,Curr.Protein Pept.Sci.4,97-104,2003;Zhou等,Cell Stem Cell 4,381-384,2009;Myou等,J.Immunol.169,2670-2676,2002;Bae等,Clin.Exp.Immunol.157,128-138,2009;Schwarze等,Science 285,1569-1572,1999)。尽管有这些进展,但使用外源蛋白仍难以影响胞内靶物,且甚至适度成功也可能需要毒性浓度的相应转导剂,原因是将蛋白质投递到细胞中的效率较低(Zhou等,Cell Stem Cell 4,381-384,2009;Wang等,Nat.Biotechnol.26,901-908,2008)。因此,仍然需要用于使功能性效应器蛋白进入细胞中以靶向胞内生物分子的更好的投递系统。

[0007] 发明概述

[0008] 本公开提供用于将功能性效应器蛋白如例如结合核酸的位点特异性蛋白质投递到细胞中的新的系统、组合物、制备物、试剂盒和相关方法,其使用超电荷(supercharged)的蛋白质(例如带正电荷的超电荷蛋白)、阳离子聚合物或阳离子脂质。在一些实施方案中,所述核酸酶是TALE核酸酶、可RNA编程的核酸酶或工程化的可RNA编程的基因组编辑酶(如Cas9及其变体或融合物)、或锌指核酸酶。在一些实施方案中,转录因子是TALE转录激活物或阻遏物。在一些实施方案中,效应器蛋白是重组酶。如本文中以更多细节描述的,使功能性效应器蛋白(例如核酸酶、转录激活物/阻遏物、Cas9蛋白包括其变体和融合物,等等)与带正电荷的超电荷蛋白融合或关联允许将该蛋白质投递到细胞内部,例如来影响基因表达或基因组修饰。还发现使功能性效应器蛋白与带负电荷的超电荷蛋白融合或关联允许该蛋白与阳离子脂质或阳离子聚合物关联,这使得该蛋白有力投递到细胞内部。此外,天然带负电荷的功能性效应器蛋白(例如VP64转录激活物、阴离子3xFLAG肽标签、及其融合物)或与固有地带负电荷的核酸(例如导引RNA;“gRNA”)关联的功能性效应器蛋白(例如Cas9蛋白、及其变体和融合物)可与阳离子脂质或阳离子聚合物关联以投递至细胞(例如在没有超电荷蛋白的情况下)。

[0009] 虽然已证明效应器蛋白的投递对于胞外靶物有效,但由于大多数蛋白不能自发进入哺乳动物细胞,其针对胞内靶物的用途相比之下开发较少。最常见通过以下来实现外源蛋白对胞内靶物的访问:经由化学转染、电穿孔或病毒投递来投递其编码DNA序列。然而,将外源DNA引入细胞带来了永久重组到基因组中、可能破坏内源基因、和长期暴露于编码的物质的可能性。对于一些研究或治疗应用(包括寻求实现基因组DNA的一次性、永久修饰的基因组编辑应用),不可复制的蛋白质剂的功能性投递可以提供改进的安全性或更广的适用性。此外,虽然使用阳离子化合物如脂质和聚合物来投递蛋白质仍然是技术上困难且在许多情况中诱导细胞毒性,但令人惊讶地发现,使用本文中提供的组合物和方法,可以以无毒性或最小毒性将某些功能性效应器蛋白(例如Cas9蛋白及其变体和融合物、重组酶、转录激活物/阻遏物等等)投递至细胞,在某些情况中介导基因组修饰时具有效率中的显著改进和降低的脱靶(off-target)效应。例如,如实施例7中描述的,将Cas9:gRNA复合物与阳离子脂质一起投递是高度有效的(从单次处理,培养的人细胞中有多达80%修饰)且相比于质粒转染还诱导更高的基因组修饰特异性,通常在人细胞中产生高>10倍的在靶:脱靶DNA修饰比。

[0010] 因此,在一个方面,使用超电荷蛋白来将功能性效应器蛋白投递到细胞中,例如核酸酶、转录激活物/阻遏物、Cas9蛋白(包括其融合物和变体)等。在一些实施方案中,所述超电荷蛋白已工程化为相比于相应的未经修饰的蛋白展现其总体表面电荷的增加。在其他实施方案中,所述超电荷蛋白已工程化为相比于相应的未经修饰的蛋白展现其总体表面电荷的降低。在其他实施方案中,用于本公开背景中的超电荷蛋白是天然存在的超电荷蛋白。所述超电荷蛋白可以与要投递的蛋白经由共价或非共价相互作用关联。不希望受任何具体理论束缚的,与gRNA关联的Cas9蛋白、变体或融合蛋白与带正电荷的超电荷蛋白的关联受净负电荷促进。在某些实施方案中,与超电荷蛋白关联的功能性效应器蛋白进一步与阳离子聚合物或阳离子脂质关联以形成适用于投递到细胞中的组合物。适宜工程化的或天然存在的超电荷蛋白的例子记载于2007年6月1日提交的国际PCT专利申请PCT/US07/70254,在2007年12月13日公开为W0 2007/143574;2009年4月28日提交的国际PCT申请PCT/US09/041984,在2009年11月5日公开为W0 2009/134808;和2010年4月28日提交的国际PCT申请

PCT/US10/001250,在2010年11月11日公开为WO 2010/129023;其各自完整内容通过提述并入本文。本文中描述了用于将核酸酶投递到细胞的超电荷蛋白的别的例子。适宜功能性效应器蛋白的另外的例子,例如核酸酶和可RNA编程的效应器蛋白如Cas9蛋白,记载于2013年8月22日提交的美国临时专利申请U.S.S.N.61/868,846,题目为“Engineered Transcription Activator-Like Effector (TALE) Domains and Uses Thereof”,2013年9月6日提交的美国临时专利申请U.S.S.N.61/874,609,题目为“Cas9 Variants and Uses Thereof”,2014年9月6日提交的美国临时专利申请U.S.S.N.61/874,682,题目为“Switchable Cas9 Nucleases and Uses Thereof”,2014年6月20日提交的美国非临时申请U.S.S.N.14/320,519,题目为“Engineered Transcription Activator-Like Effector (TALE) Domains and Uses Thereof”,2014年6月30日提交的美国非临时申请U.S.S.N.14/320,498,题目为“Cas9-FokI Fusion Proteins And Uses Thereof”,2014年6月30日提交的美国非临时申请U.S.S.N.14/320,467,题目为“Cas9-重组酶Fusion Proteins And Uses Thereof”,2014年6月8日提交的美国非临时申请U.S.S.N.14/326,329,题目为“Switchable gRNAs Comprising Aptamers”,2014年7月8日提交的美国非临时申请U.S.S.N.14/326,340,题目为“mRNA-Sensing Switchable gRNAs”,2014年7月8日提交的美国非临时申请U.S.S.N.14/326,361,题目为“Extended DNA-Sensing gRNAs”,2014年7月8日提交的美国非临时申请U.S.S.N.14/325,815,题目为“Fusions Of Cas9 Domains And Nucleic Acid-Editing Domains”,2014年7月8日提交的美国非临时申请U.S.S.N.14/326,109,题目为“Methods For Nucleic Acid Editing”,2014年7月8日提交的美国非临时申请U.S.S.N.14/326,140,题目为“Methods For Correcting PI3K Point Mutations”,2014年7月9日提交的美国非临时申请U.S.S.N.14/326,269,题目为“Methods For Correcting Presenilin Point Mutations”,2014年7月8日提交的美国非临时申请U.S.S.N.14/326,290,题目为“Methods For Correcting α -Antitrypsin Point Mutations”,2014年7月8日提交的美国非临时申请U.S.S.N.14/326,318,题目为“Methods For Correcting Von Willebrand Factor Point Mutations”,2014年7月8日提交的美国非临时申请U.S.S.N.14/326,303,题目为“Methods For Correcting Caspase-9 Point Mutations”,和美国临时申请U.S.S.N.62/030,943,题目为“Cas9 Proteins Including Ligand-Dependent Inteins”,其各自完整内容通过提述并入本文。

[0011] 在一些实施方案中,工程化或天然存在的超电荷蛋白是带正电荷的。在其他实施方案中,例如涉及使用阳离子脂质和/或阳离子聚合物投递特定效应器蛋白的那些实施方案中,所述超电荷蛋白是带负电荷的。在某些实施方案中,超正电荷或超级带负电荷的蛋白与效应器蛋白非共价关联。或者,超正电荷或超级带负电荷的蛋白可以共价结合于效应器蛋白。在一些实施方案中,效应器蛋白融合于超电荷蛋白。在某些实施方案中,所得融合蛋白包含接头,例如可切割的接头,其在超电荷蛋白和效应器蛋白之间。

[0012] 本公开的一些方面提供了组合物,其包含与功能性效应器蛋白(例如核酸酶、转录激活物/阻遏物、重组酶、Cas9蛋白包含其变体和融合物,等等)关联的超电荷蛋白。在一些实施方案中,所述组合物进一步包含阳离子脂质。在一些实施方案中,所述组合物进一步包含阳离子聚合物。在一些实施方案中,所述组合物进一步包含缓冲剂或赋形剂。在一些实施方案中,所述超电荷蛋白具有大于其相应未经修饰蛋白的总体正电荷,其量足以用于配制

且配制为用于投递至和穿透入细胞。在其他实施方案中,例如涉及使用阳离子脂质和/或阳离子聚合物投递特定效应器蛋白的那些实施方案中,所述超电荷蛋白具有大于其相应未经修饰蛋白的总体负电荷。在一些实施方案中,所述功能性效应器蛋白是位点特异性酶,例如核酸酶、Cas9蛋白、重组酶等。在一些实施方案中,Cas9蛋白是野生型Cas9蛋白,Cas9切口酶,或包含核酸酶失活(dCas9)蛋白。在一些实施方案中,Cas9蛋白是包含dCas9的融合蛋白。在一些实施方案中,所述融合蛋白包含转录激活物(例如VP64)、转录阻遏物(例如KRAB, SID)、核酸酶域(例如FokI)、重组酶域(例如Hin, Gin或Tn3)、脱氨酶(例如胞苷脱氨酶或腺苷脱氨酶)或表观遗传修饰物(epigenetic modifier)域(例如TET1)。在牵涉核酸酶的一些实施方案中,所述核酸酶是TALE核酸酶、Cas9核酸酶、Cas9切口酶或锌指核酸酶。在一些实施方案中,所述核酸酶特异性结合和切割核酸序列。在一些实施方案中,靶向的核酸序列是作为治疗靶物的基因的序列,例如期望在疾病的治疗中失活的基因。在一些实施方案中,靶向的核酸序列是PRDM16、PPAR γ 、VEGF-A、Oct-4、PI3K、早老蛋白(presenilin)、 α -抗胰蛋白酶、血管假性血友病因子(von willebrand factor)、或胱天蛋白酶-9基因序列。

[0013] 在一些实施方案中,所述功能性效应器蛋白是转录因子。在一些实施方案中,所述功能性效应器蛋白是TALE转录激活物或阻遏物。在一些实施方案中,转录因子、转录激活物或转录阻遏物特异性结合并活化或抑制基因。在一些实施方案中,所述基因是治疗靶物。在一些实施方案中,所述功能性效应器蛋白是TALE效应器。在一些实施方案中,所述超电荷蛋白共价结合于功能性效应器蛋白,由此形成融合蛋白。在一些实施方案中,所述超电荷蛋白与功能性效应器蛋白经由接头关联。在一些实施方案中,接头是可切割的接头。在一些实施方案中,所述接头是UV可切割的接头或可通过溶酶体酶切割的接头。在一些实施方案中,所述超电荷蛋白与功能性效应器蛋白非共价关联,由此形成复合物。在一些实施方案中,超电荷蛋白具有总体净正电荷。在其他实施方案中,超电荷蛋白具有总体净负电荷,且一种或多种蛋白与阳离子脂质关联。在其他实施方案中,超电荷蛋白具有总体净负电荷,且一种或多种蛋白与阳离子聚合物关联。在一些实施方案中,总体净正电荷为约+5至约+40,或总体净负电荷为约-5至约-50。在一些实施方案中,超电荷蛋白在生理pH比其相应未经修饰的蛋白带更多正电荷或带更多负电荷的。在一些实施方案中,相应未经修饰的蛋白是天然存在的蛋白。在一些实施方案中,所述超电荷蛋白在生理pH比其相应未经修饰的蛋白带多至少+5的正电荷或带多至少-5的负电荷。在一些实施方案中,所述超电荷蛋白是荧光蛋白。在一些实施方案中,所述超电荷蛋白是绿色荧光蛋白(GFP)。在一些实施方案中,超电荷蛋白是超级带正电荷的GFP。在一些实施方案中,所述超电荷蛋白是超级带正电荷的GFP(+36GFP),其包含以下序列的至少20个连续氨基酸残基:GGASKGERLFRGKVPILVELKGDVNGHKFSVRGKGGDA TRGKLTCLKFICTTGKLPVWPVTLVTTLYGVQCFSRYPKHMKRHDFFKSAMPKGYVQERTISFKKDGKYKTRAEVK FEGRTL VNRIK LKGRDFKEKGNILGHKLRYNFN SHKVYITADKRKNGIKAKFKIRHNVKDG SVQLADHYQQNTPIGRGPVLLPRNHYLSTRSKLSKDPKEKRDHMLLEFVTAAGIKHGRDERYK (SEQ ID NO:1)。

[0014] 在一些实施方案中,所述超电荷蛋白包含SEQ ID NO:1中列出的氨基酸序列。在一些实施方案中,超电荷蛋白由SEQ ID NO:1中列出的氨基酸序列组成。在一些实施方案中,组合物是药物组合物。在一些实施方案中,组合物包含药学可接受的赋形剂。在一些实施方案中,组合物被配置为用于施用给受试者且包含足以有效投递给受试者的至少一个细胞的量的超电荷蛋白和功能性效应器蛋白。在一些实施方案中,组合物包含超电荷蛋白和功能

性效应器蛋白,其量足以在施用给受试者后有效诱导可测量的治疗效果。

[0015] 本公开的一些方面提供组合物,其包含与gRNA和阳离子脂质关联的Cas9蛋白。令人惊讶地发现,当Cas9蛋白与gRNA关联时,复合物可由阳离子脂质包封并有效投递至细胞。这可在有或无超电荷蛋白的情况下完成。在一些实施方案中,所述组合物包含与超级负电荷蛋白(例如超级带负电荷的GFP)和阳离子脂质关联的Cas9蛋白,其也提供对细胞的成功投递。在一些实施方案中,所述组合物在投递至细胞群体时展现出较低的毒性,例如其中至少60%,至少65%,至少70%,至少75%,至少80%,至少85%,至少90%,至少95%或至少99%的细胞在施用组合物后是存活的。在一些实施方案中,所述Cas9蛋白是野生型Cas9蛋白、Cas9切口酶、或包含核酸酶失活(dCas9)蛋白。在一些实施方案中,Cas9蛋白是包含dCas9的融合蛋白。在一些实施方案中,所述融合蛋白包含转录激活物(例如VP64)、转录阻遏物(例如KRAB,SID)、核酸酶域(例如FokI)、重组酶域(例如Hin,Gin或Tn3)、脱氨酶(例如胞苷脱氨酶或腺苷脱氨酶)或表观遗传修饰物域(例如TET1)。

[0016] 本公开的其他方面提供组合物,其包含与gRNA和阳离子聚合物关联的Cas9蛋白。与具有阳离子脂质的情况相同,当Cas9蛋白与gRNA关联时,该复合物可与阳离子聚合物关联并有效投递至细胞。这可在有或无超电荷蛋白的情况下完成。在一些实施方案中,所述组合物包含与超级负电荷蛋白(例如超级带负电荷的GFP)和阳离子聚合物关联的Cas9蛋白,其也提供对细胞的成功投递。在一些实施方案中,所述组合物在投递至细胞群体时展现出较低的毒性,例如其中至少60%,至少65%,至少70%,至少75%,至少80%,至少85%,至少90%,至少95%或至少99%的细胞在施用组合物后是存活的。在一些实施方案中,所述Cas9蛋白是野生型Cas9蛋白、Cas9切口酶、或包含核酸酶失活(dCas9)蛋白。在一些实施方案中,Cas9蛋白是包含dCas9的融合蛋白。在一些实施方案中,所述融合蛋白包含转录激活物(例如VP64)、转录阻遏物(例如KRAB,SID)、核酸酶域(例如FokI)、重组酶域(例如Hin,Gin或Tn3)、脱氨酶(例如胞苷脱氨酶或腺苷脱氨酶)或表观遗传修饰物域(例如LSD1,TET1)。

[0017] 本公开的一些方面提供用于将本文提供的组合物施用给受试者的方法。在一些实施方案中,所述方法包括将本文中描述的组合物施用给受试者。在一些实施方案中,所述受试者易感于或患有或展现出疾病、病症或状况的一种或多种症状。在一些实施方案中,将所述组合物以足以使至少一种迹象或症状作为施用结果改善的量和在适用于使至少一种迹象或症状作为施用结果改善的条件下施用给受试者。在一些实施方案中,施用的步骤在足以使功能性效应器蛋白穿透受试者的细胞的条件下实施。在一些实施方案中,所述疾病、病症或状况与mRNA、蛋白或其组合的异常升高的水平有关。在一些实施方案中,所述组合物包含特异性结合和切割基因组序列的核酸酶,所述序列例如正常或病原性等位基因;与对疾病的易感性或疾病的发作或进展有关的基因;编码病原性RNA或蛋白质的基因;或编码在患病细胞或组织中以异常高水平表达的RNA或蛋白质的基因。在一些实施方案中,施用的步骤包括选自下组的施用路径:口服、静脉内、肌内、动脉内、皮下、心室内、局部、吸入和粘膜投递。

[0018] 本公开的一些方面提供用于将功能性效应器蛋白引入细胞的方法。在一些实施方案中,所述方法包括使细胞与如本文描述的包含超电荷蛋白和功能性效应器蛋白的组合物在适用于所述功能性效应器蛋白进入细胞的条件下接触,由此将功能性效应器蛋白引入细胞中。在一些实施方案中,所述方法包括使细胞与包含Cas9蛋白和阳离子脂质和/或阳离子

聚合物的组合物在适用于使Cas9蛋白进入细胞的条件下接触,由此将Cas9蛋白引入细胞。在一些实施方案中,Cas9蛋白进入细胞的细胞核,例如通过将核定位信号(NLS)包含在蛋白中将Cas9蛋白导向细胞核。在一些实施方案中,所述方法进一步包括确认功能性效应器蛋白(例如包含Cas9)已穿透入细胞。在一些实施方案中,细胞在受试者中,且所述接触是体内完成的。在一些实施方案中,受试者诊断为患有或有风险形成与某基因的异常表达水平有关的疾病,且其中功能性效应器蛋白(例如包含Cas9)调控所述基因的表达水平。在一些实施方案中,所述方法还包括检测基因表达水平的变化或检测受试者中的治疗应答。在一些实施方案中,所述细胞是体细胞。在一些实施方案中,所述细胞与所述组合物或药物组合物在足以诱导细胞的程序化到达期望的细胞命运的量、时间段和条件下接触。在一些实施方案中,所述方法还包括在细胞替换治疗办法中使用经编程的细胞。在一些实施方案中,所述细胞是携带与疾病有关的基因组等位基因的细胞且功能性效应器蛋白特异性靶向该等位基因。在一些实施方案中,细胞离体接触和在由功能性效应器蛋白成功靶向于不期望的等位基因后重施用给受试者。

[0019] 本公开的一些方面提供包含如本文中描述的组合物,例如包含与功能性效应器蛋白关联的超电荷蛋白的组合物试剂盒。在一些实施方案中,该试剂盒包含Cas9蛋白和超电荷蛋白。在一些实施方案中,所述试剂盒包含Cas9蛋白和阳离子脂质。在一些实施方案中,该试剂盒包含Cas9蛋白和阳离子聚合物。在一些实施方案中,该试剂盒还包含用于使用试剂盒中包含的组分的用法说明。

[0020] 本发明的这些和其他方面和实施方案以及多种优点和效用从本发明的附图和详细描述将变得更明显。

[0021] 本发明包括以下实施方案:

[0022] 1.一种包含与功能性效应器蛋白联合的超电荷蛋白的组合物,其中所述超电荷蛋白具有大于其相应未经修饰蛋白的总体正电荷并且为充足的量,并且配制为用于穿透入细胞中。

[0023] 2.实施方案1的组合物,其中所述功能性效应器蛋白是核酸酶。

[0024] 3.实施方案2的组合物,其中所述核酸酶是可RNA编程的核酸酶。

[0025] 4.实施方案3的组合物,其中所述核酸酶是TALE核酸酶、Cas9核酸酶或锌指核酸酶。

[0026] 5.实施方案2-4中任一项的组合物,其中所述核酸酶特异性结合并切割核酸序列。

[0027] 6.实施方案5的组合物,其中所述核酸序列包含在与疾病或病症有关的基因或等位基因中。

[0028] 7.实施方案5或6的组合物,其中所述组合物还包含核酸分子,所述核酸分子包含与由所述核酸酶结合并切割的核酸序列相同或同源的序列。

[0029] 8.实施方案中5-7任一项的组合物,其中所述核酸序列是包含在基因中的序列,所述基因是治疗靶物。

[0030] 9.实施方案8的组合物,其中所述基因是控制细胞命运的基因。

[0031] 10.实施方案9的组合物,其中所述基因是诱导或抑制朝向褐色脂肪细胞细胞命运的细胞程序化的基因。

[0032] 11.实施方案9或10的组合物,其中所述基因是PRDM16或PPAR γ 。

- [0033] 12. 实施方案1的组合物,其中所述功能性效应器蛋白是转录因子。
- [0034] 13. 实施方案1的组合物,其中所述功能性效应器蛋白是TALE转录激活物或阻遏物。
- [0035] 14. 实施方案12或13的组合物,其中所述转录因子、转录激活物或转录阻遏物特异性结合并激活或抑制基因。
- [0036] 15. 实施方案14的组合物,其中所述基因是治疗靶物。
- [0037] 16. 实施方案1的组合物,其中所述功能性效应器蛋白是TALE效应器。
- [0038] 17. 实施方案1-16中任一项的组合物,其中所述超电荷蛋白经由共价键与所述功能性效应器蛋白联合,从而形成融合蛋白。
- [0039] 18. 实施方案1-17中任一项的组合物,其中所述超电荷蛋白经由接头与功能性效应器蛋白联合。
- [0040] 19. 实施方案18的组合物,其中所述接头是可切割接头。
- [0041] 20. 实施方案19的组合物,其中所述接头是UV可切割接头或由溶酶体酶切割的接头。
- [0042] 21. 实施方案1-16中任一项的组合物,其中所述超电荷蛋白经由非共价相互作用与功能性效应器蛋白联合,从而形成复合物。
- [0043] 22. 实施方案21的组合物,其中所述超电荷蛋白具有总体净正电荷。
- [0044] 23. 实施方案21的组合物,其中所述总体净正电荷为约+5。
- [0045] 24. 实施方案21的组合物,其中所述总体净正电荷为约+10。
- [0046] 25. 实施方案21的组合物,其中所述总体净正电荷为约+15。
- [0047] 26. 实施方案21的组合物,其中所述总体净正电荷为约+20。
- [0048] 27. 实施方案21的组合物,其中所述总体净正电荷为约+25。
- [0049] 28. 实施方案21的组合物,其中所述总体净正电荷为约+30。
- [0050] 29. 实施方案21的组合物,其中所述总体净正电荷为约+35。
- [0051] 30. 实施方案21的组合物,其中所述总体净正电荷为约+40。
- [0052] 31. 实施方案1-30中任一项的组合物,其中所述超电荷蛋白在生理学pH比其相应未经修饰的蛋白带更多的正电荷。
- [0053] 32. 实施方案31的组合物,其中所述相应未经修饰的蛋白是天然存在的蛋白。
- [0054] 33. 实施方案1-32中任一项的组合物,其中所述超电荷蛋白在生理学pH比其相应未经修饰的蛋白至少多带+5个正电荷。
- [0055] 34. 实施方案1-32中任一项的组合物,其中所述超电荷蛋白在生理学pH比其相应未经修饰的蛋白至少多带+10个正电荷。
- [0056] 35. 实施方案1-32中任一项的组合物,其中所述超电荷蛋白在生理学pH比其相应未经修饰的蛋白至少多带+15个正电荷。
- [0057] 36. 实施方案1-32中任一项的组合物,其中所述超电荷蛋白在生理学pH比其相应未经修饰的蛋白至少多带+5个正电荷。
- [0058] 37. 实施方案1-36中任一项的组合物,其中所述超电荷蛋白是荧光蛋白。
- [0059] 38. 实施方案1-37中任一项的组合物,其中所述超电荷蛋白是绿色荧光蛋白(GFP)。

- [0060] 39. 实施方案1-38中任一项的组合物,其中所述超电荷蛋白是超带正电荷的GFP。
- [0061] 40. 实施方案中任一项的组合物,1-39,其中所述超电荷蛋白是超带正电荷的GFP(+36GFP),其包含以下序列的至少20个连续氨基酸残基:GGASKGERLFRGKVPILVELKGDVNGHKFSVRGKGGDATRGKLTCLKFICTTGKLPVPWPTLVTTLYGVQCFSRYPKHKRHDFFKSAMPKGYVQERTISFKKDGKYKTRAEVVKFEGRTLVRNRIKLGKGRDFKEKGNILGHKLRYNFNHSHKVVYITADKRKNGIKAKFKIRHNVDGQSVQLADHYQQNTPIGRGPVLLPRNHYLSTRSKLSKDPKEKRDHMLLEFVTAAGIKHGRDERYK (SEQ ID NO:1)。
- [0062] 41. 实施方案40的组合物,其中所述超电荷蛋白包含SEQ ID NO:1中列出的氨基酸序列或由SEQ ID NO:1中列出的氨基酸序列组成。
- [0063] 42. 实施方案1-41中任一项的组合物,其中所述组合物是药物组合物。
- [0064] 43. 实施方案42的药物组合物,其中所述组合物包含药学可接受的赋形剂。
- [0065] 44. 实施方案42或43的药物组合物,其中所述组合物配制为用于施用给受试者,且以有效投递给所述受试者的至少一个细胞的量包含所述超电荷蛋白和所述功能性效应器蛋白。
- [0066] 45. 实施方案42-44中任一项的药物组合物,其中所述组合物以在施用给受试者后有效诱导可测量的治疗效果的量包含所述超电荷蛋白和所述功能性效应器蛋白。
- [0067] 46. 一种方法,包括:
- [0068] 提供易感疾病、病症或状况、患有疾病、病症或状况或展现出疾病、病症或状况的一种或多种症状的受试者;和
- [0069] 向所述受试者施用实施方案1-41中任一项的组合物或实施方案42-45中任一项的药物组合物,从而使得至少一种症状得到改善。
- [0070] 47. 实施方案46的方法,其中所述施用的步骤在足以使得所述功能性效应器蛋白穿透所述受试者的细胞的条件下实施。
- [0071] 48. 实施方案46或47的方法,其中所述疾病、病症或状况与mRNA、蛋白或其组合的异常升高水平有关。
- [0072] 49. 实施方案46-48中任一项的方法,其中所述组合物包含特异性结合并切割编码致病性RNA或蛋白的基因,或编码在患病细胞或组织中以异常高水平表达的RNA或蛋白的基因的核酸酶。
- [0073] 50. 实施方案46-49中任一项的方法,其中所述施用的步骤包括选自下组的施用途径:口服、静脉内、肌内、动脉内、皮下、心室内、局部、吸入和粘膜投递。
- [0074] 51. 一种将功能性效应器蛋白导入细胞中的方法,该方法包括使所述细胞与实施方案1-41中任一项的组合物或实施方案42-45中任一项的药物组合物在适合所述功能性效应器蛋白进入所述细胞的条件下接触,由此将所述功能性效应器蛋白导入所述细胞中。
- [0075] 52. 实施方案51的方法,其进一步包括确认所述功能性效应器蛋白已穿透所述细胞。
- [0076] 53. 实施方案51或52的方法,其中所述细胞包含在受试者中,并且所述接触在体内。
- [0077] 54. 实施方案53的方法,其中所述受试者诊断为患有或有风险形成与基因的异常表达水平有关的病症,且其中所述功能性效应器蛋白调控所述基因的表达水平。
- [0078] 55. 实施方案54的方法,其中所述方法进一步包括检测所述基因表达水平的变化

或检测所述受试者中的治疗应答。

[0079] 56. 实施方案51-55中任一项的方法,其中所述细胞是体细胞。

[0080] 57. 实施方案51-56中任一项的方法,其中在一定条件下使所述细胞与一定量的所述组合物或所述药物组合物接触一定时间,所述条件、量和时间足以将细胞的程序化诱导到期望的细胞命运。

[0081] 58. 实施方案57的方法,其进一步包括在细胞替换治疗办法中使用所述程序化的细胞。

[0082] 59. 实施方案51-58中任一项的方法,其中所述细胞是携带与疾病有关的基因组等位基因的细胞且所述功能性效应器蛋白是特异性靶向所述等位基因的核酸酶。

[0083] 60. 实施方案59的方法,其中所述细胞离体接触,且在所述核酸酶成功靶向不期望的等位基因后重新施用给所述受试者。

[0084] 61. 试剂盒,其包含实施方案1-41中任一项的组合物,或实施方案42-45中任一项的药物组合物。

[0085] 62. 用于实施实施方案46-60中任一项的方法的试剂盒。

[0086] 63. 包含Cas9蛋白和阳离子脂质的组合物,其中所述Cas9蛋白与gRNA联合,且所述组合物能将所述Cas9蛋白投递到细胞的内部。

[0087] 64. 实施方案63的组合物,其中所述组合物在施用给细胞群体时展现出低毒性。

[0088] 65. 实施方案63或64的组合物,其中至少60%,至少65%,至少70%,至少75%,至少80%,至少85%,至少90%,至少95%或至少99%的细胞在施用所述组合物之后是活的。

[0089] 66. 实施方案63-65中任一项的组合物,其中所述阳离子脂质选自下组:
Lipofectamine® 2000, **Lipofectamine® 3000**, **Lipofectamine® RNAiMAX** 和 **Lipofectamine® LTX**。

[0090] 67. 实施方案63-66中任一项的组合物,其中所述Cas9蛋白包含野生型Cas9蛋白或其变体、Cas9切口酶、或核酸酶失活的Cas9 (dCas9) 或其变体。

[0091] 68. 实施方案67的组合物,其中所述dCas9融合于转录激活物或转录阻遏物。

[0092] 69. 实施方案68的组合物,其中所述转录激活物选自下组:VP16,VP64和p65;或其中所述转录阻遏物是KRAB蛋白或SID蛋白。

[0093] 70. 实施方案67的组合物,其中所述dCas9融合于核酸酶域。

[0094] 71. 实施方案70的组合物,其中所述核酸酶域包含FokI核酸酶域。

[0095] 72. 实施方案67的组合物,其中所述dCas9融合于重组酶催化域。

[0096] 73. 实施方案72的组合物,其中所述重组酶催化域包括Hin重组酶催化域、Gin重组酶催化域或Tn3重组酶催化域。

[0097] 74. 实施方案67的组合物,其中所述dCas9融合于脱氨酶。

[0098] 75. 实施方案74的组合物,其中所述脱氨酶包括选自下组的胞苷脱氨酶:APOBEC1、AID和ACF1/ASE脱氨酶;或其中所述脱氨酶包括腺苷脱氨酶,其任选地包含ADAT家族脱氨酶。

[0099] 76. 实施方案67的组合物,其中所述dCas9融合于表观遗传修饰物。

[0100] 77. 实施方案76的组合物,其中所述表观遗传修饰物选自下组:组蛋白脱甲基酶、

组蛋白甲基转移酶、羟化酶、组蛋白脱乙酰基酶和组蛋白乙酰转移酶。

[0101] 78. 实施方案77的组合物,其中所述表观遗传修饰物包含LSD1组蛋白脱甲基酶或TET1羟化酶。

[0102] 79. 实施方案67的组合物,其中所述Cas9蛋白与具有净负电荷的超电荷蛋白联合,其中所述联合具有净负电荷,任选地其中所述联合是共价联合。

[0103] 80. 实施方案79的组合物,其中所述超电荷蛋白是荧光蛋白,任选地其中所述超电荷蛋白是超带负电荷的GFP。

[0104] 81. 一种组合物,其包含效应器蛋白和阳离子脂质、阳离子聚合物、或阳离子脂质和阳离子聚合物两者,其中所述效应器蛋白与具有净负电荷的超电荷蛋白联合,所述联合具有净负电荷,且所述组合物能将所述效应器蛋白投递至细胞内部。

[0105] 82. 实施方案81的组合物,其中所述组合物在施用给细胞群体时展现出低毒性。

[0106] 83. 实施方案81或82的组合物,其中至少60%,至少65%,至少70%,至少75%,至少80%,至少85%,至少90%,至少95%或至少99%的细胞在施用所述组合物之后是活的。

[0107] 84. 实施方案81-83中任一项的组合物,其中所述阳离子脂质选自下组:
Lipofectamine® 2000, Lipofectamine® 3000, Lipofectamine® RNAiMAX 和 Lipofectamine® LTX.

[0108] 85. 实施方案81-84中任一项的组合物,其中所述超电荷蛋白是荧光蛋白,任选地其中所述超电荷蛋白是超带负电荷的GFP。

[0109] 86. 实施方案81-85中任一项的组合物,其中所述效应器蛋白是重组酶。

[0110] 87. 实施方案86的组合物,其中所述重组酶是Cre重组酶。

[0111] 88. 实施方案81-87中任一项的组合物,其中所述效应器蛋白包含TALE蛋白。

[0112] 89. 实施方案88的组合物,其中所述TALE蛋白包含VP64转录激活物。

[0113] 90. 实施方案81-89中任一项的组合物,其中所述组合物是药物组合物。

[0114] 91. 方法,包括:

[0115] 对此需要的受试者施用实施方案90的药物组合物。

[0116] 92. 一种将Cas9蛋白导入细胞中的方法,所述方法包括使所述细胞与实施方案81-90中任一项的组合物在适合所述Cas9蛋白进入所述细胞的条件下接触,从而将所述Cas9蛋白导入所述细胞。

[0117] 定义

[0118] 如本文和权利要求中使用的,单数形式“一个”、“一种”和“该”包括单数和复数指代物,除非上下文清楚地另外指示。因此,例如对“一个/种药剂”的指代包括单个药剂和多种药剂。

[0119] 如本文中在两种或更多种模块(例如蛋白质或蛋白质域)的上下文中使用的,术语“与…关联”指模块与另一模块实体直接或经由一个或多个另外的充当连接剂的模块关联或连接以形成以下结构的事实,该结构足够稳定,从而所述模块在使用该结构的条件,例如在生理学条件下保持实体关联。超电荷蛋白可以经由非共价相互作用(例如静电相互作用)与功能性效应器蛋白(例如核酸酶、转录激活物/阻遏物、重组酶、Cas9蛋白包含其变体和融合物,等等)关联。在某些实施方案中,超电荷蛋白可以经由静电相互作用与功能性效应器蛋白关联以形成复合物。在一些实施方案中,充足数量的更弱的相互作用能提供使得模块

在多种不同条件下保持实体关联的足够稳定性。在某些实施方案中,超电荷蛋白经由共价键(例如酰胺键)与功能性效应器蛋白关联。在一些实施方案中,功能性效应器蛋白通过肽键与超电荷蛋白直接关联,或经由接头间接关联。

[0120] 术语“Cas9”或“Cas9核酸酶”指包含Cas9蛋白的RNA导引的核酸酶,或其片段(例如包含Cas9的活性或失活的DNA切割域或部分失活的DNA切割域(例如Cas9“切口酶”),和/或Cas9的gRNA结合域的蛋白质)。在一些实施方案中,术语“Cas9”指包含Cas9或其片段的融合蛋白。

[0121] 在一些实施方案中,Cas9指来自以下的Cas9:溃疡棒状杆菌(*Corynebacterium ulcerans*) (NCBI Refs:NC_015683.1,NC_017317.1);白喉棒状杆菌(*Corynebacterium diphtheria*) (NCBI Refs:NC_016782.1,NC_016786.1);Spiroplasma syrphidicola (NCBI Ref:NC_021284.1);中间普雷沃菌(*Prevotella intermedia*) (NCBI Ref:NC_017861.1);台湾螺原体(*Spiroplasma taiwanense*) (NCBI Ref:NC_021846.1);海豚链球菌(*Streptococcus iniae*) (NCBI Ref:NC_021314.1);Belliiella baltica (NCBI Ref:NC_018010.1);扭曲冷弯曲菌(*Psychroflexus torquis*) (NCBI Ref:NC_018721.1);嗜热链球菌(*Streptococcus thermophilus*) (NCBI Ref:YP_820832.1);无害利斯特菌(*Listeria innocua*) (NCBI Ref:NP_472073.1);空肠弯曲菌(*Campylobacter jejuni*) (NCBI Ref:YP_002344900.1);或脑膜炎奈瑟氏菌(*Neisseria meningitidis*) (NCBI Ref:YP_002342100.1)。

[0122] 术语“阳离子脂质”指在生理pH具有阳离子或正电荷的脂质。阳离子脂质可采取多种形式,包括但不限于脂质体或微团(micelle)。可用于本公开的某些方面的阳离子脂质是本领域公知的,且一般包含极性和非极性域两者,结合聚阴离子如核酸分子或超负电荷蛋白,且一般已知促进核酸投递到细胞中。可用的阳离子脂质的例子包括聚乙烯亚胺,聚酰胺-胺(polyamidoamine) (PAMAM) starburst树状分子,Lipofectin(DOTMA和DOPE的组合),Lipofectase,LIPOFECTAMINE[®](例如LIPOFECTAMINE[®] 2000,LIPOFECTAMINE[®] 3000,LIPOFECTAMINE[®] RNAiMAX,LIPOFECTAMINE[®] LTX),SAINT-RED(Synvolux Therapeutics,Groningen Netherlands),DOPE,Cytofectin(Gilead Sciences,Foster City,Calif.)和Eufectins(JBL,San Luis Obispo,Calif.)。例示性阳离子脂质体可从以下制备:N-[1-(2,3-二油氧基(dioleoloxo))-丙基]-N,N,N-三甲基铵氯(DOTMA),N-[1-(2,3-二油氧基)-丙基]-N,N,N-三甲基铵硫酸甲酯(DOTAP),3β-[N-(N',N'-二甲基氨基乙烷)氨基甲酰]胆固醇(DC-Chol),2,3,-二油氧基-N-[2(精胺羧基氨基(sperminecarboxamido))乙基]-N,N-二甲基-1-丙基三氟乙酸铵(DOSPA),1,2-双十四烷基氧丙基-3-二甲基-羟乙基溴化铵;和二甲基双十八烷基溴化铵(DDAB)。阳离子脂质已在本领域中用于投递核酸分子至细胞(参见例如美国专利Nos.5,855,910;5,851,548;5,830,430;5,780,053;5,767,099;8,569,256;8,691,750;8,748,667;8,758,810;8,759,104;8,771,728;Lewis等1996.Proc.Natl.Acad.Sci.USA 93:3176;Hope等1998.Molecular Membrane Biology15:1)。另外,其他脂质组合物在本领域中也是已知的,且包括例如美国专利No.4,235,871;美国专利Nos.4,501,728;4,837,028;4,737,323中教导的那些。

[0123] 如本文中使用的,术语“阳离子聚合物”指具有净正电荷的聚合物。阳离子聚合物是本领域中公知的,且包括以下中描述的那些:Samal等,Cationic polymers and their

therapeutic potential. Chem Soc Rev. 2012 Nov 7; 41(21): 7147-94; 公开的美国专利申请 U.S. 2014/0141487A1, U.S. 2014/0141094A1, U.S. 2014/0044793A1, U.S. 2014/0018404A1, U.S. 2014/0005269A1 和 U.S. 2013/0344117A1; 和美国专利 Nos. 8,709,466; 8,728,526; 8,759,103; 和 8,790,664; 其各自完整内容通过提述并入本文。例示性阳离子聚合物包括但不限于, 聚丙烯胺 (PAH); 聚乙烯亚胺 (PEI); 聚(L-赖氨酸) (PLL); 聚(L-精氨酸) (PLA); 聚乙烯胺同型或共聚物; 聚(乙烯基苯甲基-三-C₁-C₄-烷基铵盐); 脂肪族或脂肪族二卤化物和脂肪族N,N,N',N'-四-C₁-C₄-烷基-烯二胺的聚合物; 聚(乙烯基吡啶(vinylpyridin)) 或聚(乙烯基吡啶盐); 聚(N,N-二烯丙基-N,N-二-C₁-C₄-烷基-卤化铵); 季铵化二-C₁-C₄-烷基-氨基乙基丙烯酸酯或甲基丙烯酸酯的同型或共聚物; 聚QUAD™; 聚氨基酰胺; 等等。

[0124] 术语“脱氨酶”指催化脱氨基反应的酶。在一些实施方案中, 所述脱氨酶是胞苷脱氨酶, 其催化胞苷或脱氧胞苷分别水解脱氨基为尿嘧啶或脱氧尿嘧啶。

[0125] 如本文中使用的, 术语“有效量”指足以引发期望的生物学应答的生物活性试剂/药剂的量。例如, 在一些实施方案中, 功能性效应器蛋白(例如核酸酶、转录激活物/阻遏物、重组酶、Cas9蛋白包含其变体和融合物, 等等)的有效量可以指足以诱导可检测的效果(例如靶位点的切割, 靶位点的修饰, 基因表达的调控, 等等)的蛋白质的量。这类效果可以在适宜的测定法中检测, 例如在无细胞测定法中或在靶细胞、组织或受试生物体中。如熟练技术人员将林辉的, 药剂(agent)例如功能性效应器蛋白的有效量可以随着多种因素而变化, 如例如期望的生物学应答, 要靶向的具体等位基因, 靶向的基因组、靶位点、细胞或组织, 以及使用的超电荷蛋白。

[0126] 术语“效应器蛋白”指当引入细胞时调控细胞的生物学功能的蛋白, 例如修饰细胞中的核酸分子(如切割、脱氨基、重组等等), 或调控基因在细胞中的表达或表达水平(例如增加或降低)。

[0127] 如本文中使用的, 术语“工程化的”指由人设计、产生、制备、合成和/或制造的蛋白质分子、复合物、物质或实体。因此, 工程化的产物是不在自然界中存在的产物。在一些实施方案中, 工程化的蛋白或组合物, 例如与功能性效应器蛋白如核酸酶、Cas9蛋白(包括其变体和融合物)关联的工程化的超电荷蛋白是以下超电荷蛋白, 其已设计为满足特定要求或具有特定的期望特征, 例如具有指定的净电荷, 特异性结合和/或切割或修饰感兴趣的靶序列, 具有特定的最小或最大切割或酶活性, 和/或具有特定稳定性。

[0128] 如本文中使用的, 术语“表观遗传修饰物(epigenetic modifier)”指一种蛋白质或其催化域, 其具有导致DNA例如染色体DNA的表观遗传修饰的酶活性。表观遗传修饰包括但不限于DNA甲基化和脱甲基化; 组蛋白修饰, 包括甲基化和脱甲基化(例如单、双和三甲基化), 组蛋白乙酰化和脱乙酰化, 以及组蛋白泛素化(ubiquitylation), 磷酸化和类泛素化。

[0129] 术语“功能蛋白”指处于以下形式的蛋白, 其表现出用来表征其的特性和/或活性。

[0130] 术语“融合蛋白”指包含多个异源蛋白质、蛋白质域或肽的蛋白质, 例如超电荷蛋白和功能性效应器蛋白, 其经由肽连接彼此关联, 从而形成单个氨基酸序列。在某些实施方案中, 融合蛋白由基因编码。

[0131] 术语“基因”具有如其在本领域中所理解的含义。本领域普通技术人员将领会, 术语“基因”可以包括基因调控序列(例如启动子、增强子等等)和/或内含子序列。还将领会基因的定义包括对不编码蛋白质而是编码功能性RNA分子(如RNAi试剂、核酶、tRNA等等)的核

酸的指代。为了清楚目的,应注意,如本申请中使用的,术语“基因”一般指编码蛋白质的核酸的一部分;该术语可任选地涵盖调控序列,如本领域普通技术人员从上下文将是清楚的。该定义不意图排除术语“基因”对非蛋白编码表达单元的应用,而是为了澄清在大多数情况下,如本文件中使用的术语指编码蛋白的核酸。

[0132] 术语“分离的”指一种分子、复合物、物质或实体,其已(1)从当初始产生时(在自然界中或在实验背景中)其所关联的至少一些组分分离,和/或(2)由人产生、制备、合成和/或制备。分离的物质和/或实体可与至少约10%,约20%,约30%,约40%,约50%,约60%,约70%,约80%,约90%或更多的其所初始关联的其他组分分离。在一些实施方案中,分离的试剂/物质是超过约80%,约85%,约90%,约91%,约92%,约93%,约94%,约95%,约96%,约97%,约98%,约99%,或超过约99%纯的。如本文中使用的,如果某物质基本上没有其他组分,则该物质是“纯”的。

[0133] 如本文中使用的,术语“接头”指连接两个分子或模块例如超电荷蛋白个核酸酶的化学基团或分子。典型地,接头位于以下之间或由以下侧接:两个基团、分子或其他实体,且经由共价键连接每个,从而连接两者。在一些实施方案中,接头包含一个氨基酸或多个氨基酸(例如肽或蛋白质)。在一些实施方案中,接头是有机分子、基团、聚合物或化学模块。在一些实施方案中,接头是可切割的接头,例如该接头包含在暴露于切割活性如UV光或水解酶(如溶酶体蛋白酶)时可切割的键。在一些实施方案中,接头是具有至少1,至少2,至少3,至少4,至少5,至少6,至少7,至少8,至少9,至少10,至少15,至少20,至少25,至少30,至少40,至少50或更多个氨基酸的任意区段(stretch)氨基酸。在一些实施方案中,肽接头包含三肽Gly-Gly-Ser的重复,例如包含序列(GGS)_n,其中n代表至少1,2,3,4,5,6,7,8,9,10或更多个重复。在一些实施方案中,接头包含序列(GGS)₆(SEQ ID NO:2)。在一些实施方案中,肽接头是16残基的“XTEN”接头或其变体(参见例如Schellenberger等Arecombinant polypeptide extends the in vivo half-life of peptides and proteins in a tunable manner.Nat.Biotechnol.27,1186-1190(2009))。在一些实施方案中,XTEN接头包含序列SGSETPGTSESATPES(SEQ ID NO:3),SGSETPGTSESA(SEQ ID NO:4)或SGSETPGTSESATPEGGSGGS(SEQ ID NO:5)。在一些实施方案中,肽接头是选自以下的一种或多种:VPFLLEPDNINGKTC(SEQ ID NO:6),GSAGSAAGSGEF(SEQ ID NO:7),SIVAQLSRPDPA(SEQ ID NO:8),MKIIEQLPSA(SEQ ID NO:9),VRHKLKRVGS(SEQ ID NO:10),GHGTGSTGSGSS(SEQ ID NO:11),MSRPDPA(SEQ ID NO:12);或GGSM(SEQ ID NO:13)。

[0134] 术语“核酸酶”,如本文中使用的,指能切割连接核酸分子中的核苷酸残基的磷酸二酯键的试剂,例如蛋白质或小分子。在一些实施方案中,核酸酶是蛋白质,例如能结合核酸分子和切割连接核酸分子中的核苷酸残基的磷酸二酯键的酶。核酸酶可以是内切核酸酶,其切割多核苷酸链中的磷酸二酯键,或者是外切核酸酶,其切割在多核苷酸链末端处的磷酸二酯键。在一些实施方案中,核酸酶是位点特异性核酸酶,其结合和/或切割特定核苷酸序列内的特定磷酸二酯键,该核苷酸序列在本文中亦称为“识别序列”、“核酸酶靶位点”或“靶位点”。在一些实施方案中,核酸酶识别单链的靶位点,而在其他实施方案中,核酸酶识别双链的靶位点,例如双链的DNA靶位点。许多天然存在的核酸酶例如许多天然存在的DNA限制核酸酶的靶位点是本领域技术人员公知的。在许多情况中,DNA核酸酶如EcoRI、HindIII或BamHI识别长度为4至10个碱基对的回文双链DNA靶位点,并在该靶位点内的特定

位置处切割两条DNA链的每条。一些内切核酸酶对称地切割双链核酸靶位点,即在相同位置切割两条链,从而使得末端包含碱基配对的核苷酸,在本文中也称为平末端。其他内切核酸酶不对称地切割双链核酸靶位点,即在不同位置切割每条链,从而使得末端包含不配对的核苷酸。双链DNA分子的末端处不配对的核苷酸也称为“突出”,例如“5'-突出”或“3'-突出”,根据不配对的核苷酸形成的是相应DNA链的5'还是3'端。以不配对核苷酸为末端的双链DNA分子端也称为粘端,因此它们能“黏着于”包含互补的不配对核苷酸的其他双链DNA分子端。核酸酶蛋白通常包含介导蛋白与核酸底物的相互作用的“结合域”和催化核酸骨架内的磷酸二酯键的切割的“切割域”。在一些实施方案中,核酸酶蛋白能以单体形式结合和切割核酸分子,而在其他实施方案中,核酸酶蛋白必须二聚化或多聚化以切割靶核酸分子。天然存在的核酸酶的结合域和切割域,以及能组合以创建结合特定靶位点的核酸酶的模式结合域和切割域,是本领域技术人员公知的。例如,转录激活物元件可用作结合域以特异性结合期望的靶位点,并融合或缀合于切割域,例如FokI的切割域,以创造切割期望靶位点的工程化的核酸酶。

[0135] 术语“核酸”和“核酸分子”,如本文中可交换使用的,指包含核苷、核苷酸或核苷酸聚合物的化合物。通常地,聚合核酸,例如包含3个或更多个核苷酸的核酸分子是线性分子,其中相邻的核苷酸经由磷酸二酯键彼此连接。在一些实施方案中,“核酸”指单个核酸残基(例如核苷酸和/或核苷)。在一些实施方案中,“核酸”指包含3个或更多个个体核苷酸残基的寡核苷酸链。如本文中使用的,术语“寡核苷酸”和“多核苷酸”可以交换使用以指核苷酸的聚合物(例如至少3个核苷酸的串)。在一些实施方案中,“核酸”涵盖RNA以及单链和/或双链DNA。核酸可以是天然存在的,例如在基因组、转录本、mRNA、tRNA、rRNA、siRNA、snRNA、质粒、粘粒、染色体、染色单体、或其他天然存在的核酸分子的背景中。在另一方面,核酸分子可以是非天然存在的分子,例如重组的DNA或RNA,人工染色体,工程化的基因组,或其片段,或合成的DNA、RNA、DNA/RNA杂合体,或包含非天然存在的核苷酸或核苷。例如,术语“核酸”、“DNA”、“RNA”和/或类似术语包括核酸类似物,即具有磷酸二酯骨架以外的形式的类似物。可从天然来源纯化核酸,使用重组表达系统产生,并任选地纯化,化学合成等等。在适宜时,例如在化学合成的分子的情况中,核酸可以包含核苷类似物如具有经化学修饰的碱基或糖和骨架修饰的类似物。除非另外说明,核酸序列以5'至3'方向呈现。在一些实施方案中,核酸是或包含天然的核苷(例如腺苷、胸苷、鸟苷、胞苷、尿苷、脱氧腺苷、脱氧胸苷、脱氧鸟苷、和脱氧胞苷);核苷类似物(例如2-氨基腺苷、2-硫代胸苷、肌苷、吡咯并嘧啶、3-甲基腺苷、5-甲基胞苷、2-氨基腺苷、C5-溴尿苷、C5-氟尿苷、C5-碘尿苷、C5-丙炔-尿苷、C5-丙炔-胞苷、C5-甲基胞苷、2-氨基腺苷、7-脱氮腺苷(deazaadenosine)、7-脱氮尿苷、8-氧腺苷、8-氧鸟苷、0(6)-甲基鸟嘌呤和2-硫代胞苷);经化学修饰的碱基;经生物修饰的碱基(例如甲基化碱基);插入的碱基;经修饰的糖(例如2'-氟核糖、核糖、2'-脱氧核糖、阿拉伯糖和己糖);和/或经修饰的磷酸根基团(例如硫代磷酸酯和5'-N-亚磷酰胺连接)。

[0136] 术语“药物组合物”,如本文中使用的,指在疾病或病症治疗的背景中可施用给受试者的组合物。在一些实施方案中,药物组合物包含活性成分,例如与功能性效应器蛋白如核酸酶关联的超电荷蛋白或编码超电荷蛋白和功能性效应器蛋白(例如以融合蛋白的形式)的核酸,和药学可接受的赋形剂。

[0137] 如本文中使用的,术语“生理学pH”指在正常、非病理细胞或受试者中可见的pH值。

在一些实施方案中,生理学pH为pH 5-8之间。在一些实施方案中,生理学pH是pH 7-7.5,例如pH 7.0,pH 7.1,pH 7.2,pH 7.3,pH 7.4或pH 7.5。在一些实施方案中,生理学pH是pH 6.5-7.5。在一些实施方案中,生理学pH是pH 5,pH 5.5,pH 6,pH 6.5,pH 7,pH 7.5或pH 8。

[0138] 术语“预防”或“防止”指对有风险形成疾病、病症或状况的受试者(例如相比于对照受试者或对照受试者组有升高的风险,或相比于年龄匹配和/或性别匹配的受试者的平均风险有升高的风险)的防护性治疗,其导致受试者形成所述疾病、病症或状况的可能性的降低(如相比于无预防时的可能性),和/或抑制已经建立的病症的进一步进展。

[0139] 术语“增殖性疾病”,如本文中使用的,指其中由于细胞或细胞群体展现出异常升高的增殖率而细胞或组织内稳态被扰乱的任何疾病。增殖性疾病包括增殖过多疾病,如赘生前增生状况和赘生性(neoplastic)疾病。赘生性疾病特征在于细胞的异常增殖且包括良性和恶性赘生物形成。恶性赘生物形成也称为癌症。

[0140] 术语“蛋白质”在本文中可与术语“肽”和“多肽”交换使用,且指通过肽(酰胺)键连接在一起的氨基酸残基的聚合物。该术语指任意大小、结构或功能的蛋白质、肽或多肽。通常地,蛋白质、肽或多肽长度将至少为3个氨基酸。蛋白质、肽或多肽可以指个体蛋白质或蛋白质的集合。可以修饰蛋白质、肽或多肽中的一个或多个氨基酸,例如通过添加化学实体如糖基团、羟基基团、磷酸基团、法呢基(farnesyl)基团、异法呢基基团、脂肪酸基团、用于缀合的接头,功能化,或其他修饰等等。蛋白质、肽或多肽还可以是单分子或可以是多分子复合物。蛋白质、肽或多肽可以仅是天然存在的蛋白质或肽的片段。蛋白质、肽或多肽可以是天然存在的、重组的或合成的,或其任意组合。蛋白质可以包含不同的域,例如TALE效应器蛋白可以包含核酸结合域和效应器域,例如核酸切割域或转录激活物或阻遏物域。在一些实施方案中,蛋白质包含蛋白质性部分,例如构成核酸结合域的氨基酸序列,和有机化合物,例如可充当核酸切割剂的化合物。

[0141] 术语“可RNA编程的核酸酶”和“RNA导引的核酸酶”在本文中可交换使用且指与一种或多种不是切割靶物的RNA形成复合物(例如结合或关联)的核酸酶。在一些实施方案中,可RNA编程的核酸酶在与RNA复合时可以称为核酸酶:RNA复合物。可RNA编程的核酸酶包括Cas9。通常,结合的一种或多种RNA称为导引RNA(gRNA)。以单RNA分子存在的gRNA可称为单导引RNA(sgRNA),不过“gRNA”可交换使用以指作为单分子或作为两个或更多分子的复合物存在的导引RNA。单RNA分子存在的gRNA通常包含两个域:(1)与靶核酸共享同源性的域(例如,指导Cas9复合物对靶物的结合);和(2)结合Cas9蛋白的域。gRNA包含与靶位点互补的核苷酸序列,其介导核酸酶/RNA复合物对所述靶位点的结合并提供核酸酶:RNA复合物的序列特异性。

[0142] 如本文中使用的,术语“重组酶”指一种位点特异性酶,其介导重组酶识别序列之间的DNA的重组,导致重组酶识别序列之间的DNA片段的切除、整合、倒置或交换(例如易位)。重组酶可分类为两个不同的家族:丝氨酸重组酶(例如解离酶和转化酶)和酪氨酸重组酶(例如整合酶)。丝氨酸重组酶的例子包括但不限于,Hin,Gin,Tn3, β -six,CinH,ParA, γ δ ,Bxb1, ϕ C31,TP901,TG1, ϕ BT1,R4, ϕ RV1, ϕ FC1,MR11,A118,U153和gp29。酪氨酸重组酶的例子包括但不限于,Cre,FLP,R,Lambda,HK101,HK022和pSAM2。丝氨酸和酪氨酸重组酶名称来自重组酶用于攻击DNA且在链交换期间变为共价连接于DNA的保守的亲核氨基酸残基。重组酶具有众多应用,包括创建基因敲除/敲入和基因疗法应用。参见,例如Brown等,

“Serine recombinases as tools for genome engineering.”Methods.2011;53(4):372-9;Hirano等,“Site-specific recombinases as tools for heterologous gene integration.”Appl.Microbiol.Biotechnol.2011;92(2):227-39;Chavez和Calos,“Therapeutic applications of the Φ C31 integrase system.”Curr.Gene Ther.2011;11(5):375-81;Turan和Bode,“Site-specific recombinases:from tag-and-target-to tag-and-exchange-based genomic modifications.”FASEB J.2011;25(12):4088-107;Venken和Bellen,“Genome-wide manipulations of *Drosophila melanogaster* with transposons,Flp recombinase,and Φ C31 integrase.”Methods Mol.Biol.2012;859:203-28;Murphy,“Phage recombinases and their applications.”Adv.Virus Res.2012;83:367-414;Zhang等,“Conditional gene manipulation:Cre-ating a new biological era.”J.Zhejiang Univ.Sci.B.2012;13(7):511-24;Karpenshif和Bernstein,“From yeast to mammals:recent advances in genetic control of homologous recombination.”DNA Repair(Amst).2012;1;11(10):781-8;其各自全部内容通过提述完整并入本文。本文中提供的重组酶不意为仅指可在本发明的实施方案中使用的重组酶的例子。本发明的方法和组合物可通过挖掘数据库中的新直系同源重组酶或设计具有限定的DNA特异性的合成重组酶来扩展(参见,例如Groth等,“Phage integrases:biology and applications.”J.Mol.Biol.2004;335,667-678;Gordley等,“Synthesis of programmable integrases.”Proc.Natl.Acad.Sci.U S A.2009;106,5053-5058;其各自全部内容通过提述完整并入本文)。可用于本文中描述的方法和组合物的重组酶的其他例子是本领域技术人员已知的,且预期发现或生成的任何新的重组酶都能用于本发明的不同实施方案中。在一些实施方案中,重组酶(或其催化域)融合于Cas9蛋白(例如dCas9)。

[0143] 术语“重组”,在核酸修饰(例如基因组修饰)的背景中,用于指通过重组酶蛋白的作用修饰两种或更多种核酸分子,或单个核酸分子的两个或更多个区域的过程。重组可导致例如一个或多个核酸分子中或之间的核酸序列的插入、倒置、切除或易位。

[0144] 如本文中使用的,术语“受试者”,指个体生物。在一些实施方案中,受试者是处于任何发育阶段的任意性别的人。在一些实施方案中,受试者是非人哺乳动物。在一些实施方案中,受试者是非人灵长类。在一些实施方案中,受试者是啮齿动物。在一些实施方案中,受试者是实验室动物,例如小鼠、大鼠、沙鼠、豚鼠、鱼、蛙或蝇。在一些实施方案中,受试者是农场动物,例如绵羊、山羊、猪或牛。在一些实施方案中,受试者是伴侣动物,例如猫或犬。在一些实施方案中,受试者是脊椎动物、两栖动物、爬行动物、鱼、昆虫、蝇或线虫。在一些实施方案中,受试者是经遗传工程化的,例如经遗传工程化的非人受试者。

[0145] 术语“超电荷”指导致蛋白质总体净电荷的增加或降低的蛋白质的任意修饰。修饰包括但不限于,氨基酸序列的改变或带电荷模块的添加(例如羧酸基团,磷酸根基团,硫酸根基团,氨基基团)。超电荷还指试剂/药剂与带电荷的蛋白质(天然存在或经修饰的)联合以形成相对于单独的试剂/药剂具有增加或降低的电荷的复合物。

[0146] 术语“靶位点”,如本文中在结合核酸分子的功能性效应器蛋白(如核酸酶和转录激活物或阻遏物)的背景中使用的,指核酸分子内的序列,其被效应器蛋白结合和作用,例如被核酸酶切割或分别被转录激活物或阻遏物转录激活或抑制。靶位点可以是单链或双链的。在RNA导引的(例如可RNA编程的)核酸酶(例如Cas9)的背景中,靶位点通常包含与可RNA

编程的核酸酶的sgRNA互补的核苷酸序列,和在临近gRNA互补性序列的3'端处的原间隔区临近基序(PAM)。对于RNA导引的核酸酶Cas9(或具有gRNA结合活性的变体或融合物),所述靶位点在一些实施方案中可以是20个碱基对加上3碱基对PAM(例如NNN,其中N代表任意核苷酸)。通常地,PAM的第一个核苷酸可以是任意核苷酸,而两个下游核苷酸根据具体的RNA导引的核酸酶指定。RNA导引的核酸酶如Cas9的例示性靶位点是本领域技术人员已知的,且包括但不限于,NNG,NGN,NAG和NGG,其中N代表任意核苷酸。另外,来自不同物种(例如嗜热链球菌(*S. thermophilus*)而非酿脓链球菌(*S. pyogenes*))的Cas9核酸酶识别包含序列NGGNG的PAM。已知另外的PAM序列,包括但不限于NNAGAAW和NAAR(参见,例如Esvelt和Wang, *Molecular Systems Biology*, 9:641 (2013),其完整内容通过提述并入本文)。例如,RNA导引的核酸酶如例如Cas9的靶位点可以包含结构[NZ]-[PAM],其中每个N独立地为任意核苷酸,且Z是1至50的整数。在一些实施方案中,Z为至少2,至少3,至少4,至少5,至少6,至少7,至少8,至少9,至少10,至少11,至少12,至少13,至少14,至少15,至少16,至少17,至少18,至少19,至少20,至少25,至少30,至少35,至少40,至少45或至少50。在一些实施方案中,Z为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。在一些实施方案中,Z为20。在一些实施方案中,“靶位点”还可以指被核酸酶结合但不被其切割的核酸分子内的序列。例如,本文中描述的某些实施方案提供包含无活性(或失活)Cas9 DNA切割域的蛋白。这类蛋白(例如当还包含Cas9RNA结合域时)能结合gRNA指定的靶位点,然而由于DNA切割位点是失活的,因此靶位点不被具体蛋白切割。然而,如本文中描述的这类蛋白质通常与另一蛋白(例如核酸酶、转录激活物、重组酶、脱氨酶,等等)或分子缀合、融合或结合,所述另一蛋白或分子介导核酸分子的修饰。在一些实施方案中,实际切割的序列将依赖于介导核酸分子切割的蛋白(例如核酸酶)或分子,且例如在一些情况中,将与失活的一种或多种Cas9蛋白结合的近端或所相距距离。在二聚化的核酸酶,例如包含失活Cas9(或Cas9 RNA结合域)和DNA切割域(例如FokI切割域或活性Cas9切割域)的蛋白二聚体的背景中,靶位点通常包含左半位点(由一个蛋白结合)、右半位点(由第二蛋白结合),和其中进行切割的两个各半位点之间的间隔序列。在一些实施方案中,切割左半位点或右半位点(而非间隔序列)。该结构([左半位点]-[间隔序列]-[右半位点])在本文中称为LSR结构。在一些实施方案中,左半位点和/或右半位点对应于RNA导引的靶位点(例如Cas9靶位点)。在一些实施方案中,两个半位点之一或两者比Cas9靶向的典型区域更短或更长,例如短于或长于20个核苷酸。在一些实施方案中,左和右半位点包含不同的核酸序列。在一些实施方案中,靶位点是包含三(3)个RNA导引的核酸酶靶位点序列的序列,例如3个对应于Cas9靶位点序列的序列,其中第一和第二,以及第二和第三Cas9靶位点序列由间隔序列分开。在一些实施方案中,间隔序列长至少5,至少6,至少7,至少8,至少9,至少10,至少11,至少12,至少13,至少14,至少15,至少16,至少17,至少18,至少19,至少20,至少25,至少30,至少35,至少40,至少45,至少50,至少60,至少70,至少80,至少90,至少100,至少125,至少150,至少175,至少200或至少250bp。

[0147] 术语“转录激活物”和“转录阻遏物”指结合靶核酸序列且导致与靶核酸序列有关的基因产物的表达水平分别增加或降低的物质/试剂,如蛋白质(例如转录因子或其片段)。例如,如果靶核酸序列位于基因的调控区内,则转录激活物导致由该基因编码的基因产物的表达水平增加(相反,转录阻遏物导致由基因编码的基因产物的表达水平降低)。所述基

因产物可以是来自基因转录的RNA(例如mRNA)或从自基因转录的mRNA翻译的多肽。通常mRNA水平中的增加或降低导致从其翻译的多肽水平的增加或降低。可以使用测量mRNA或蛋白的标准技术来测定表达水平。

[0148] 术语“转录激活物样效应子”(TALE),如本文中使用的,指包含DNA结合域的效应器蛋白,该DNA结合域含有高度保守的33-34氨基酸序列,该序列包含高度可变的两氨基酸基序(重复可变二残基,Repeat Variable Di-residue,RVD)。RVD基序决定对核酸序列的结合特异性,且可以依照本领域技术人员公知的方法工程化以特异性结合期望的DNA序列(参见,例如Miller,Jeffrey;等(February 2011).“A TALE nuclease architecture for efficient genome editing”.*Nature Biotechnology* 29(2):143-8;Zhang,Feng;等(February 2011).“Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription”.*Nature Biotechnology* 29(2):149-53;Geißler,R.;Scholze,H.;Hahn,S.;Streubel,J.;Bonas,U.;Behrens,S.E.;Boch,J.(2011),Shiu,Shin-Han.编“Transcriptional Activators of Human Genes with Programmable DNA-Specificity”.*PLoS ONE* 6(5):e19509;Boch,Jens(February 2011).“TALEs of genome targeting”.*Nature Biotechnology* 29(2):135-6;Boch,Jens;等(December 2009).“Breaking the Code of DNA Binding Specificity of TAL-Type III Effectors”.*Science* 326(5959):1509-12;和Moscou,Matthew J.;Adam J.Bogdanove(December 2009).“A Simple Cipher Governs DNA Recognition by TAL Effectors”.*Science* 326(5959):1501;其各自完整内容通过提述并入本文)。氨基酸序列和DNA识别之间的这种简单关系允许通过选择含有适宜RVD的重复区段的组合来工程化特定的DNA结合域。TALE效应器蛋白包括但不限于,TALE核酸酶(TALEN)和TALE转录激活物和阻遏物。

[0149] 术语“转录激活物样元件核酸酶”(TALEN),如本文中使用的,指包含针对DNA切割域例如FokI域的转录激活物样效应子DNA结合域的人工核酸酶。已报道了一些用于生成工程化的TALE构建体的模式装配方案(Zhang,Feng;等(February 2011).“Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription”.*Nature Biotechnology* 29(2):149-53;Geißler,R.;Scholze,H.;Hahn,S.;Streubel,J.;Bonas,U.;Behrens,S.E.;Boch,J.(2011),Shiu,Shin-Han.编“Transcriptional Activators of Human Genes with Programmable DNA-Specificity”.*PLoS ONE* 6(5):e19509;Cermak,T.;Doyle,E.L.;Christian,M.;Wang,L.;Zhang,Y.;Schmidt,C.;Baller,J.A.;Somia,N.V.等(2011).“Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting”.*Nucleic Acids Research*;Morbiter,R.;Elsaesser,J.;Hausner,J.;Lahaye,T.(2011).“Assembly of custom TALE-type DNA binding domains by modular cloning”.*Nucleic Acids Research*;Li,T.;Huang,S.;Zhao,X.;Wright,D.A.;Carpenter,S.;Spalding,M.H.;Weeks,D.P.;Yang,B.(2011).“Modularly assembled designer TAL effector nucleases for targeted gene knockout and gene replacement in eukaryotes”.*Nucleic Acids Research*;Weber,E.;Gruetzner,R.;Werner,S.;Engler,C.;Marillonnet,S.(2011).Bendahmane,Mohammed.编“Assembly of Designer TAL Effectors by Golden Gate Cloning”.*PLoS ONE* 6(5):e19722;其各自内容通过提述并入

本文)。

[0150] 术语“转录阻遏物”指结合靶核酸序列且导致与该靶核酸序列有关的基因产物的表达水平降低的转录因子,例如蛋白质。例如,如果靶核酸序列位于基因的调控区内,则转录阻遏物导致由该基因编码的基因产物表达水平的降低。所述基因产物可以是来自基因转录的RNA(例如mRNA)或从自基因转录的mRNA翻译的多肽。通常mRNA水平中的降低导致从其翻译的多肽水平的降低。可以使用测量mRNA或蛋白的标准技术来测定表达水平。

[0151] 术语“锌指核酸酶”,如本文中使用的,指包含缀合于结合域的核酸切割域的核酸酶,所述结合域包含锌指阵列。在一些实施方案中,切割域是II型限制内切核酸酶FokI的切割域。锌指核酸酶可以设计为靶向给定核酸分子中用于切割的实际上任意期望的序列,且设计锌指结合域结合复杂基因组背景中的独特位点的可能性允许对活细胞中单个基因组位点的靶向性切割,例如以实现具有治疗价值的靶向性基因组改变。将双链断裂靶向至期望的基因组基因座可用于将移码突变(frame-shift mutation)引入基因的编码序列中,这是由非同源DNA修复途径的易错性质所致。可通过本领域技术人员公知的方法生成锌指核酸酶以靶向感兴趣的位点。例如,可以设计具有期望特异性的锌指结合域,其通过组合具有已知特异性的个体锌指基序。结合DNA的锌指蛋白Zif268的结构已显示了本领域中的很多工作,且已描述了对于64种可能的碱基对三联体的每种,获得锌指然后混合和匹配这些模式锌指来设计具有任何期望的序列特异性的蛋白质的思路(Pavletich NP,Pabo CO(May 1991).“Zinc finger-DNA recognition:crystal structure of a Zif268-DNA complex at 2.1Å”.Science 252(5007):809-17,其完整内容通过提述并入本文)。在一些实施方案中,组合各自识别3碱基对DNA序列的分别的锌指以生成识别长度范围从9碱基对至18碱基对的靶位点的3、4、5或6锌指阵列。在一些实施方案中,涵盖更长的阵列。在其他实施方案中,组合识别6-8个核苷酸的2锌指模块以生成4-、6-或8-锌指阵列。在一些实施方案中,利用细菌或噬菌体展示来开发识别期望核酸序列,例如长度为3-30bp的期望核酸酶靶位点的锌指域。在一些实施方案中,锌指核酸酶包含融合或者经由接头(例如多肽接头)彼此缀合的锌指结合域和切割域。接头的长度决定切割离锌指域结合的核酸序列的距离。如果使用较短的接头,那么切割域将切割更接近结合的核酸序列的核酸,而更长的接头将导致切割与结合的核酸序列之间的更大的距离。在一些实施方案中,锌指核酸酶的切割域必须二聚化以切割结合的核酸。在一些这类实施方案中,二聚体是两个单体的异二聚体,其各自包含不同的锌指结合域。例如在一些实施方案中,二聚体可以包含包括缀合于FokI切割域的锌指域A的一个单体,和包含缀合于FokI切割域的锌指域B的一个单体。在该非限制性例子中,锌指域A在靶位点的一侧结合核酸序列,锌指域B在靶位点的另一侧结合核酸序列,且二聚化FokI域切割锌指域结合位点之间的核酸。

[0152] 术语“锌指”,如本文中使用的,指小核酸结合蛋白结构基序,其特征在于折叠和稳定该折叠的一个或多个锌离子的配位。锌指涵盖很多种不同的蛋白结构(参见,例如Klug A,Rhodes D(1987).“Zinc fingers:a novel protein fold for nucleic acid recognition”.Cold Spring Harb.Symp.Quant.Biol.52:473-82,其完整内容通过提述并入本文)。锌指可以设计为结合核苷酸的特定序列,且包含一系列锌指的融合物的锌指阵列可以设计为结合实际上任何期望的靶序列。这类锌指阵列可以形成蛋白质例如核酸酶的结合域,例如如果缀合于核酸切割域时。不同类型的锌指基序是本领域技术人员已知的,包括

但不限于Cys₂His₂、Gag指节(knuckle)、Treble clef、锌带(Zinc ribbon)、Zn₂/Cys₆和TAZ2域样基序(参见,例如Krishna SS,Majumdar I,Grishin NV(January2003).“Structural classification of Zinc fingers:survey and summary”.Nucleic Acids Res.31(2):532-50)。单个锌指基序通常结合核酸分子的3或4个核苷酸。因此,包含2个锌指基序的锌指域可以结合6-8个核苷酸,包含3个锌指基序的锌指域可以结合9-12个核苷酸,包含4个锌指基序的锌指域可以结合12-16个核苷酸,如此类推。可以采用任何适宜的蛋白质工程化技术来改变锌指的DNA结合特异性和/或设计新的锌指融合物来结合实际上任何期望的长度从3-30个核苷酸的靶序列(参见,例如Pabo CO,Peisach E,Grant RA(2001).“Design and selection of novel cys2His2 Zinc finger proteins”.Annual Review of Biochemistry 70:313-340;Jamieson AC,Miller JC,Pabo CO(2003).“Drug discovery with engineered zinc-finger proteins”.Nature Reviews Drug Discovery 2(5):361-368;and Liu Q,Segal DJ,Ghiara JB,Barbas CF(May1997).“Design of polydactyl zinc-finger proteins for unique addressing within complex genomes”.Proc.Natl.Acad.Sci.U.S.A.94(11);其各自完整内容通过提述并入本文)。工程化的锌指阵列和切割核酸的蛋白质域之间的融合物可用于生成“锌指核酸酶”。锌指核酸酶通常包含结合核酸分子内的特定靶位点的锌指域,和切割在由结合域结合的靶位点内或附近处的核酸分子的核酸切割域。典型的工程化的锌指核酸酶包含结合域,其具有3至6个个体锌指基序和长度范围从9个碱基对至18个碱基对的结合靶位点。更长的靶位点在期望结合和切割在给定基因组中是独特的靶位点的情况下特别有吸引力。

[0153] 术语“治疗/处理”指目的为如本文中描述的疾病或病症或其一种或多种症状的逆转、减轻、延迟其发作、或抑制其进展的临床干预。如本文中使用的,术语“治疗/处理”指目的为如本文中描述的疾病或病症或其一种或多种症状的逆转、减轻、延迟其发作、或抑制其进展的临床干预。在一些实施方案中,治疗可以在一种或多种症状已形成和/或已诊断出疾病后施用。在其他实施方案中,治疗可以在不存在症状的情况下施用,例如以预防或延迟症状的发作或抑制疾病的发作或进展。例如,治疗可以在症状发作之前对易感个体施用(例如根据症状史和/或根据遗传或其他易感性因素)。治疗还可以在症状已消退后持续,例如以预防或延迟其复发。

[0154] 附图简述

[0155] 图1.大分子投递到哺乳动物细胞内的示意图。

[0156] 图2.编程的脂肪细胞命运:从白色脂肪组织(WAT)转换为棕色脂肪组织(BAT)。

[0157] 图3.使用超电荷的投递平台投递编程为靶向PPAR γ 或PRDM16的TALE激活物。

[0158] 图4.包含+36GFP融合物、18.5聚体TALE域和VP64激活域的融合蛋白的示意图。

[0159] 图5.+36GFP-TALE激活物融合蛋白的表达和纯化。

[0160] 图6.在投递+36GFP PPAR γ 和PRDM16 TALE激活物融合蛋白后测试脂肪细胞调控基因的激活。

[0161] 图7.+36GFP TALE激活物融合蛋白在不同浓度的投递功效。

[0162] 图8.两种不同的+36GFP-PRDM16 TALE融合蛋白在NIH 3T3细胞中的投递功效的比较。

[0163] 图9.PPAR γ -TALE激活物的融合物投递后PPAR γ 的基因表达和与多种对照的比

较。

[0164] 图10.RDM16-TALE激活物的融合物投递后PRDM16的基因表达和与多种对照的比较。

[0165] 图11.在存在血清时观察到适中的TALE活性。

[0166] 图12.验证PPAR γ 的病毒投递继之以用脂肪生成混合物 (cocktail) 的7天处理。

[0167] 图13.用于将成纤维细胞编程为WAT和BAT的测定法的示意图。

[0168] 图14.用+36GFP TALE激活物融合蛋白处理后观察到的脂肪细胞形成。

[0169] 图15.多种处理7天后用脂质TOX红染色显示在病毒投递之后以及在超电荷的PPAR γ TALE激活物融合蛋白投递之后脂肪细胞的形成。

[0170] 图16.多种处理7天后用脂质TOX红染色显示在病毒投递之后以及在超电荷的PPAR γ TALE激活物融合蛋白投递之后脂肪细胞的形成。

[0171] 图17.在病毒投递之后以及在超电荷的PPAR γ TALE激活物融合蛋白投递之后WAT生物标志物基因的表达。

[0172] 图18.投递超电荷的PRDM16 TALE激活物融合蛋白以体内诱导棕色脂肪的脂肪细胞。在PPAR γ 和PRDM16的病毒投递之后以及在超电荷的TALE激活物蛋白的融合物投递之后观察到有力的脂肪细胞形成。

[0173] 图19.通过PPAR γ 和PRDM16的表达比较棕色脂肪标志物的TALE/TALE、病毒/TALE和病毒/病毒诱导的表达。

[0174] 图20.RT-qPCR评估与通过脂质TOX染色观察到的脂肪细胞分化一致。

[0175] 图21.将功能性TALE激活物融合蛋白作为与+36GFP的复合物投递改进了投递后TALE激活物的活性。

[0176] 图22.在以融合物(+36GFP PRDM16 TALE-3)或与+36GFP的复合物(+36GFP+PRDM16 TALE-3)进行TALE激活物的融合物投递后的PRDM16基因表达。复合物的投递倾向于增加TALE激活物活性。

[0177] 图23.Aurein肽与+36GFP融合对TALE激活物融合物投递(以融合物或与+36GFP的复合物)后PRDM16基因表达的影响。

[0178] 图24.以融合物(+36GFP PRDM16 TALE-3)或与Lipofectamine LTX的复合物(+36GFP+PRDM16 TALE-3)进行TALE激活物融合物投递后的PRDM16基因表达。

[0179] 图25.将超电荷的融合蛋白或与Cas9的复合物投递到哺乳动物细胞中。(GGG)9-T-ALAL-PKKKRKV对应于SEQ ID NO:251。

[0180] 图26.野生型Cas9蛋白和具有+36GFP和Aurein-GGS9的Cas9融合蛋白的纯化。

[0181] 图27A-B.显示了用于将蛋白质投递到哺乳动物细胞中的策略,其通过与聚阴离子大分子融合或非共价复合和用阳离子脂质包封。图27(A)显示重组酶,转录激活物样效应器(TALE)蛋白和Cas9内切核酸酶结合核酸且天然是阳离子(理论净电荷显示为黑色)且未被阳离子脂质有效包封。然而,这些蛋白质可通过融合于超级带负电荷的蛋白如(-30)GFP,或通过与聚阴离子核酸复合使其成为高度阴离子性的。图27(B)显示代表通常用于转染DNA和RNA包封所得高度阴离子性蛋白或蛋白:核酸复合物,介导其投递到哺乳细胞中的阳离子脂质。

[0182] 图28A-F.Cre重组酶对培养的人细胞的投递。在图28(A)中,显示了高度阳离子(+

36) GFP或高度阴离子 (-30) GFP对Cre重组酶的融合。使用在Cre介导的重组时表达DsRed的HeLa报道细胞系来评估Cre投递功效; (GGs) 9对应于SEQ ID NO:252和His6对应于SEQ ID NO:253。在图28(B)中,将HeLa dsRed细胞用10nM(-30) GFP-Cre和阳离子脂质RNAiMAX处理。在含有10%胎牛血清(FBS)的培养基中温育48小时后使细胞可见。图28(C)显示在10% FBS培养基或在无血清培养基中投递(+36) GFP-Cre,和在全血清培养基中有或无阳离子脂质RNAiMAX的情况下投递(-30) GFP-Cre。图28(D)呈现了48小时后在含有10% FBS的275 μ L培养基中,阳离子脂质剂量对于功能性(-30) GFP-Cre投递功效的影响。图28(E)是几种商品化的阳离子脂质和聚合物对于(-30) dGFP-Cre的功能投递功效的比较。图28(F)显示融合于Cre的多种阴离子肽或蛋白质序列的RNAiMAX介导的投递。误差条反映来自在不同日实施的3个生物重复的标准偏差。

[0183] 图29A-B. TALE转录激活物投递到培养的人细胞中。图29(A)显示18.5个重复的TALE激活物的设计,其C末端融合于VP64激活域而N末端融合于(-30) GFP和NLS; (GGs) 9对应于SEQ ID NO:251和His6对应于SEQ ID NO:252。融合物的总体净理论电荷为-43。图29(B)显示了对NTF3转录的激活,其通过编码靶向NTF3基因中位点的TALE-VP64激活物的质粒的传统转染,或通过相应NTF3靶向性(-30) GFP-TALE-VP64蛋白的RNAiMAX阳离子脂质介导的投递。通过qRT-PCR测量基因表达水平并相对于GAPDH表达水平标准化。误差条反映来自在不同日实施的3个生物重复的标准偏差。

[0184] 图30A-E. Cas9:sgRNA、Cas9 D10A切口酶、和dCas9-VP64转录激活物投递到培养的人细胞中。图30(A)显示将与EGFP靶向性sgRNA或VEGF靶向性sgRNA复合的Cas9蛋白变体以阳离子脂质介导的方式投递到U2OS EGFP报道细胞中。将结果与标准转染Cas9和sgRNA表达质粒的结果比较。图30(B)显示采用T7内切核酸酶I(T7EI)测定法测量来自以下处理的EGFP修饰的结果:无处理(第1道),仅用EGFP靶向性sgRNA(第2道),仅Cas9蛋白(第3道),Cas9蛋白+VEGF靶向性sgRNA+RNAiMAX(第4道),转染表达Cas9和EGFP靶向性sgRNA的质粒(第5道),或Cas9蛋白+EGFP靶向性sgRNA+RNAiMAX(第6道)。通过密度测定计算的Indel效率显示在凝胶图像下方。图30(C)呈现了使用与4种sgRNA和RNAiMAX复合的Cas9的单投递,在EGFP和3种内源基因处基因组修饰的T7EI测定法的结果。通过密度测定计算的Indel效率显示在凝胶图像下方。图30(D)显示Cas9 D10A切口酶和sgRNA对的投递,其通过质粒转染或RNAiMAX介导的蛋白质:RNA复合物投递。预期破坏EGFP的sgRNAs GFP g1+GFP g5或GFP g3+GFP g7导致基因破坏,而GFP g5+GFP g7靶向同一链,因此预期是非功能性的。图30(E)显示靶向NTF3的催化性死亡的(dCas9)-VP64转录激活物的投递,其通过质粒转染或RNAiMAX介导的蛋白质投递。VEGF g3和VEGF g5 sgRNA两者的投递充当NTF3基因失活的阴性对照。误差条反映来自在不同日实施的6个生物重复的标准偏差。

[0185] 图31A-D. 显示了通过质粒转染或阳离子脂质介导的蛋白质:sgRNA投递,在培养的人细胞中Cas9介导的内源性基因切割的DNA序列特异性。在图31(A)中,对于内源性CLTA,EMX和VEGF基因的在靶(on-target)修饰实施T7EI测定法。在图31(B-D)中,显示了对于质粒转染或阳离子脂质介导的蛋白质:sgRNA投递,来自Cas9:sgRNA的在靶:脱靶(off-target)DNA修饰比率。调整每次处理的条件以产生~10%在靶切割,从而实现对在靶基因修饰效率可比的条件下两种投递方法之间的DNA切割特异性的比较。P值列于表2。对每份在靶和脱靶样品测序一次,其中每份在靶样品有>10,000个分析的序列,和每份脱靶样品平均有>

111,000个分析的序列(表2)。

[0186] 图32A-D.显示了Cre重组酶和Cas9:sgRNA复合物对小鼠内耳中毛细细胞的体内投递。在图32(A)中,将P0 floxP-tdTomato小鼠(n=4)的中阶(*scala media*) (蜗管)注射有0.3 μ L的50% RNAiMAX中23 μ M (-30) GFP-Cre或仅RNAiMAX(对照)注射。5天后,使用免疫组织学使得指示Cre介导的重组的tdTomato表达可视化。红色=tdTomato;绿色=Myo7a;白色=Sox2;蓝色=DAPI。黄色括弧指示外部毛细胞(OHC)层。图32(B)显示,在(-30)GFP-Cre投递后10天,存在tdTomato阳性外部毛细胞的完整的表达*espin* (*Esp*)的硬纤毛(箭头),类似于对照耳蜗中的硬纤毛。红色=tdTomato;绿色=*Esp*;白色=Sox2;蓝色=DAPI。图32(C)等同于图32(A),只不过使用Lipofectamine 2000而不是RNAiMAX。(n=4)。上部分和下部分的图是小鼠耳蜗分别在低和高放大率的图,详细显示了投递的效率以及对耳蜗体系和毛细胞丧失的影响。图32(D)显示当将P2 *Atoh1*-GFP小鼠(n=3)的中阶(蜗管)注射有0.3 μ L的50% RNAiMAX或Lipofectamine 2000中33 μ M Cas9,33 μ M sgRNA的结果。当10天后看时,Cas9介导的基因破坏导致GFP表达的丧失。上部分图仅显示GFP信号,而下部分图包括另外的免疫组织学标志。下部分图中的黄框突出了具有丧失的GFP表达的毛细胞。红色=tdTomato;绿色=Myo7a;白色/浅蓝色=Sox2;蓝色=DAPI。以白色显示的所有比例尺均为10 μ m。

[0187] 图33A-C.阳离子脂质介导的Cre重组酶的投递的优化。图33(A)显示优化(-30)GFP-Cre在BSR-TdTomato细胞中的投递,该细胞是用于处理Cre重组效率的第二种报道细胞系。图33B显示RNAiMAX剂量对HeLa dsRed报道细胞中(-30)GFP-Cre重组效率的影响和毒性,如通过FACS测量的。通过前向散射(forward-scatter)和侧向散射设门来分选HeLa细胞以鉴定保留正常形态的活细胞。图33(C)例示了融合于Cre重组酶的蛋白质的静电荷与阳离子脂质介导的功能性Cre投递效率之间的关系。将列于25nM的与域融合的Cre重组酶与1.5 μ L RNAiMAX组合并与HeLa dsRed报道细胞温育。2天后,通过FACS测量重组效率。误差条反映来自在不同日实施的3个生物重复的标准偏差。

[0188] 图34A-D.通过阳离子脂质介导的投递相对于超级带正电荷的阳离子蛋白投递的蛋白质摄取。图34(A)量化了来自细胞的GFP荧光,所述细胞在用PBS+heparin(20U/mL)清洗以除去未结合蛋白后用(-30)GFP-Cre和RNAiMAX或(+36)GFP-Cre处理。图34(B)显示相对于来自与(+36)GFP融合的Cre重组酶投递效率,(-30)GFP-Cre+1.5 μ L RNAiMAX的功能性Cre重组酶投递效率。图34(C)提供了通过(-30)GFP融合物+1.5 μ M RNAiMAX处理相对于(+36)GFP融合物的mCherry摄取的比较,其通过测量在处理和用PBS+肝素清洗细胞后48小时时总细胞群体的均值mCherry荧光。图34(D)显示在存在或缺失RNAiMAX的情况下,(-30)GFP-Cre或(+36)GFP-Cre的总细胞GFP荧光。图35.设计为靶向NTF3基因的TALE激活物的投递优化。将HEK293T细胞通过转染NTF3 TALE质粒处理或通过脂质体投递NTF3 TALE蛋白处理。在经验确定的两种处理的最佳温育时间后,收获细胞并通过qRT-PCR分析NTF3的mRNA水平。选择最佳蛋白质(25-50nM)和脂质剂量(1.5 μ L RNAiMAX)来比较图29B中的两种投递技术。误差条反映来自在不同日实施的6个生物重复的标准偏差。

[0189] 图36A-D.确定EGFP报道基因的基因破坏频率,其通过投递Cas9:sgRNA和通过流式细胞术分析。图36(A)提供通过由Cas9双链短链诱导的NHEJ在U2OS细胞中的EGFP破坏的示意图。图36(B)显示与(-30)dGFP-Cas9复合的EGFP靶向性sgRNA或脱靶sgRNA的投递,其使用RNAiMAX连同质粒转染阳性对照(橙色)。图36(C)提供EGFP荧光的破坏不是细胞毒性的结果

的确认,其通过用T0-PRO-3存活/死亡染料(Life Technologies, Carlsbad CA)处理样品并通过流式细胞术分析所得细胞。图36(D)显示对T0-PRO-3染色的测试,其通过添加细胞渗透性但非完全膜裂解性的去污剂(0.5% Tween)。

[0190] 图37A-D. Cas9:sgRNA功能性投递的优化。在图37(A)中,两种测试构建体的阳离子脂质介导的投递效率显示,在低蛋白和sgRNA浓度,与天然Cas9相比,更多的阴离子(-30) dGFP-NLS-Cas9促进更有效的投递。图37(B)显示(-30) dGFP-NLS-Cas9的投递优化作为蛋白和sgRNA浓度的函数。图37(C)显示没有任何融合物或标签的Cas9蛋白的投递作为蛋白和sgRNA浓度的函数。图37(D)提供对于(-30) dGFP-NLS-Cas9和天然Cas9的NAiMAX介导的投递,最佳的sgRNA对蛋白比率。误差条反映来自在不同日实施的3个生物重复的标准偏差。

[0191] 图38A-C. NLS和/或(-30) dGFP对功能性Cas9投递的影响作为sgRNA和Cas9两者浓度的函数。在3个固定的sgRNA浓度测量EGFP基因破坏:10nM(图38(A))、25nM(图38(B))和50nM(图38(C)),连同图中显示的不同蛋白浓度。使用0.8 μ L RNAiMAX进行投递并在48小时后通过FACS测定EGFP荧光信号的丧失。

[0192] 图39A-C. RNAiMAX和Lipofectamine 2000对Cas9:sgRNA投递效率和细胞毒性的影响。在图39(A)中,显示了在用RNAiMAX或Lipofectamine 2000处理16小时的U2OS EGFP报道细胞中,在不同Cas9蛋白浓度和恒定剂量的100nM EGFP sgRNA时的EGFP基因破坏。在16小时后,除去培养基并向细胞添加新鲜培养基直至蛋白投递处理后48-72小时的测定终点。通过FACS使用T0-PRO-3存活/死亡染料测定活细胞群体。图39(B)显示Cas9:sgRNA投递至U2OS细胞的毒性概况(profile)作为Lipofectamine 2000剂量的函数。图39(C)提供细胞的毒性概况作为RNAiMAX剂量的函数。误差条反映来自在不同日实施的3个生物重复的标准偏差。

[0193] 图40. 在不同的蛋白和sgRNA浓度靶向NTF3基因的dCas9-VP64投递的优化。将HEK293T细胞用dCas9-VP64激活物和NTF3靶向性sgRNA g2或所有6种NTF3靶向性sgRNA的混合物处理16小时,和0.8 μ L RNAiMAX在48孔板形式中(275 μ L终体积)。通过qRT-PCR测定NTF3 mRNA水平并相对于GAPDH的那些标准化。误差条反映来自在不同日实施的6个生物重复的标准偏差。

[0194] 图41A-C. 绘出了通过高通量测序测量的7个人类基因的Indel频率,所述人类基因通过伪处理、通过转染Cas9质粒和sgRNA线性DNA PCR产物、或通过阳离子脂质介导的蛋白质:sgRNA投递处理。伪处理涉及EGFP靶向性sgRNA而非3种人类基因靶向性sgRNA之一的阳离子脂质介导的蛋白质:sgRNA投递。图41(A)显示CLTA基因的在靶和脱靶indel频率。图41(B)提供EMX基因的在靶和脱靶indel频率。图41(C)显示VEGF基因的在靶和脱靶indel频率。对每份在靶和脱靶样品测序一次,其中每份在靶样品有>10,000个分析的序列,和每份脱靶样品平均有>111,000个分析的序列(表2)。

[0195] 图42A-C. Cas9内切核酸酶对小鼠胚胎干细胞的投递。图42(A)显示用Cas9蛋白和RNAiMAX但无sgRNA(对照)处理的浮球(floating sphere)保留强GFP荧光(右),而用Cas9:sgRNA和RNAiMAX处理的那些展现降低的GFP荧光(左)。比例尺为100 μ m。图42(B)显示细胞粘附后的对照祖细胞,且实质上所有对照祖细胞是GFP阳性的(右小图)。Cas9:sgRNA处理导致GFP表达的显著降低(左小图)且许多祖细胞在细胞粘附后显示完全的GFP敲低(箭头)。比例尺为20 μ m。图42(C)显示在成像后收获的干细胞上的T7E1测定确认了GFP报道子的切割。类似的基因靶物修饰效率从阳离子脂质介导的Cas9:sgRNA投递和Cas9和EGFP sgRNA质粒的

转染观察到。

[0196] 图43A-B. 在体内内源性基因座处由阳离子脂质介导的Cas9核酸酶和sgRNA的蛋白投递诱导的基因组修饰。图43(A) 显示在小鼠毛细胞中在Cas9和EGFP sgRNA的阳离子脂质介导的投递之后, 在EGFP在靶基因座处基因组DNA序列的代表性例子。对于显示的每个例子, 未经修饰的基因组位点是第一序列, 之后是含有缺失的丰度最高的8个序列和含有插入的3个序列。每个序列之前的数字指示序列计数。sgRNA靶位点为粗体并绿色划线。插入和缺失红色显示。PAM位点蓝色显示。图43(B) 显示对于EMX在靶位点在小鼠毛细胞中与图42(A) 中相同的分析。图43(A) 中显示的序列从上到下对应于SEQ ID NOs: 223-236; 且图43(B) 中显示的序列从上到下对应于SEQ ID NOs: 237-250。

[0197] 本发明特定实施方案的详细描述

[0198] 本发明提供用于投递功能性效应器蛋白至细胞的复合物、组合物、制备物、试剂盒、系统和相关方法, 所述功能性效应器蛋白为例如核酸酶、重组酶和Cas9蛋白(包括其变体和融合物, 例如Cas9切口酶和Cas9融合至脱氨酶、基因编辑酶、转录阻遏物和激活物、表观遗传修饰物等等), 其通过将功能性效应器蛋白与超电荷蛋白、阳离子聚合物、和/或阳离子脂质的一种或多种联合。典型地, 将功能性效应器蛋白投递至细胞内部, 例如以引起细胞中的生物学作用, 如切割基因组靶序列或调控靶基因的表达。在一些实施方案中, 所述生物学作用对该细胞所在的受试者施加治疗益处。用于投递功能性效应器蛋白的复合物、组合物、制备物、系统、试剂盒和相关方法可用于将效应器蛋白引入细胞, 例如在操纵细胞用于研究或治疗目的的背景中。本文中提供的用于投递功能性效应器蛋白的组合物、制备物、系统、试剂盒和相关方法展现出改进的功效和降低的细胞毒性, 以及相比于当前技术的便于制备。使用本文中提供的组合物、制备物、系统、试剂盒和相关方法来投递位点特异性蛋白质, 如TALEN或Cas9蛋白(或其变体或融合物) 允许在体外或体内靶向性操纵/修饰宿主细胞的基因组, 同时避免使用更侵入性的投递方法, 如编码位点特异性蛋白的载体的病毒投递。

[0199] 在一些实施方案中, 本发明技术使用超电荷蛋白来将功能性效应器蛋白投递到细胞中。在某些实施方案中, 所述超电荷蛋白是工程化的蛋白。在一些实施方案中, 所述超电荷蛋白是天然存在的超电荷蛋白。本发明的一些方面基于以下认知, 即超电荷蛋白被细胞内吞; 可与超电荷蛋白关联的功能性效应器蛋白能被细胞连同超电荷蛋白一起有效摄取; 以及这类功能性效应器蛋白在细胞摄取后保留其生物学功能, 例如它们能切割或修饰基因组靶位点或调控靶基因的转录。

[0200] 在一些实施方案中, 本文中提供的包含与功能性效应器蛋白(例如核酸酶、转录激活物/阻遏物、重组酶、Cas9蛋白包含其变体和融合物, 等等) 关联的超电荷蛋白的组合物可用作治疗剂、诊断剂或研究工具。在一些实施方案中, 功能性效应器蛋白, 如核酸酶或转录因子, 可以是治疗活性的, 例如其靶向与疾病或病症有关的基因。在一些实施方案中, 将本文提供的包含超电荷蛋白和功能性效应器蛋白(如核酸酶或转录因子) 的组合物用于调控细胞中基因的表达或调控细胞中的生物途径(例如信号传导途径、代谢途径)。在一些实施方案中, 使细胞与本文中描述的创造性组合物接触以将功能性效应器蛋白引入细胞中。在一些实施方案中, 将创造性的组合物施用给有此需要的受试者以将功能性效应器蛋白引入受试者的细胞内, 例如与疾病或病症有关的细胞内。依照本公开的一些方面用于投递功能性效应器蛋白的适宜细胞和细胞类型包括但不限于, 人细胞, 哺乳动物细胞, T细胞, 神经

元,干细胞,祖细胞,血细胞,成纤维细胞,上皮细胞,赘生性细胞和肿瘤细胞。

[0201] 超电荷蛋白

[0202] 用于本发明的超电荷蛋白可通过将蛋白表面上的非保守氨基酸改变为更极性或带电荷的氨基酸残基产生。在某些实施方案中,蛋白表面上的非保守氨基酸突变为在生理学pH (pH ~ 7.4) 带正电荷的氨基酸。要修饰的氨基酸残基可以是疏水、亲水、带电荷的或其组合。超电荷蛋白还可以通过将带电荷模块附接于蛋白质以使蛋白质带超电荷来产生。超电荷蛋白经常对于聚集呈抗性,具有增加的重折叠能力,抵抗不适当折叠,具有改进的溶解度,且一般在宽范围的条件下更稳定,包括变性条件如热或存在去污剂。

[0203] 适用于依照本发明的方面的超电荷蛋白是本领域中已知的且包括但不限于,以下中公开的那些超电荷蛋白:2007年6月1日提交的国际PCT专利申请PCT/US07/70254,在2007年12月13日公开为WO 2007/143574;2009年4月28日提交的国际PCT申请PCT/US09/041984,在2009年11月5日公开为WO 2009/134808;和2010年4月28日提交的国际PCT申请PCT/US10/001250,在2010年11月11日公开为WO 2010/129023;其各自完整内容通过提述并入本文。在一些实施方案中,所述超电荷蛋白是工程化的超电荷蛋白。在一些实施方案中,所述超电荷蛋白是天然存在的超电荷蛋白,例如在以下公开的天然超电荷蛋白:2010年4月28日提交的国际PCT申请PCT/US10/001250,在2010年11月11日公开为WO 2010/129023;其各自通过提述并入本文。在一些实施方案中,所述超电荷蛋白,工程化的或天然存在的,展现出大于0.8的电荷:分子量比,例如 ≥ 0.85 , ≥ 0.9 , ≥ 0.95 , ≥ 1 , ≥ 1.1 , ≥ 1.2 , ≥ 1.3 , ≥ 1.4 , ≥ 1.5 , ≥ 1.6 , ≥ 1.7 , ≥ 1.8 , ≥ 1.9 , ≥ 2 , ≥ 2.5 , ≥ 3 , ≥ 4 , ≥ 5 , ≥ 6 , ≥ 7 , ≥ 8 或 ≥ 10 。

[0204] 采用的超电荷蛋白可以源自任意物种的植物、动物和/或微生物。在某些实施方案中,所述超电荷蛋白是哺乳动物蛋白。在某些实施方案中,所述超电荷蛋白是人蛋白。在某些实施方案中,所述蛋白源自研究中通常使用的生物体。例如,要修饰的蛋白可能来自灵长类(例如猿、猴)、啮齿类(例如家兔、仓鼠、沙鼠)、猪、犬、猫、鱼(例如Danio rerio)、线虫(例如秀丽线虫(C.elegans))、酵母(例如酿酒酵母(Saccharomyces cerevisiae))或细菌(例如大肠杆菌(E.coli))。在某些实施方案中,所述蛋白是非免疫原性的。在某些实施方案中,所述蛋白是非抗原性的。在某些实施方案中,所述蛋白不具有内在的生物学活性或已经修饰为不具生物学活性。在某些实施方案中,蛋白的选择基于其靶向能力。在某些实施方案中,所述蛋白是绿色荧光蛋白。在一些实施方案中,所述超电荷蛋白是带超电荷的谷胱甘肽S-转移酶(GST)。在一些实施方案中,所述超电荷蛋白是带超电荷的链霉亲合素。

[0205] 在一些实施方案中,使用已修饰为增加总体净电荷,或者增加蛋白表面上带电荷残基的总数目的超电荷蛋白。在某些实施方案中,超电荷蛋白的理论净电荷相比于未经修饰的蛋白增加了至少+1,至少+2,至少+3,至少+4,至少+5,至少+10,至少+15,至少+20,至少+25,至少+30,至少+35或至少+40。在某些实施方案中,超电荷蛋白的理论净电荷在生理学pH(即~7.4)为至少+1,至少+2,至少+3,至少+4,至少+5,至少+10,至少+15,至少+20,至少+25,至少+30,至少+35或至少+40。

[0206] 在其他实施方案中,例如涉及使用阳离子脂质和/或阳离子聚合物的那些实施方案中,使用已经过修饰以降低总体净电荷或降低蛋白表面带电荷残基的总数目的超电荷蛋白。在某些实施方案中,超电荷蛋白的理论净电荷相比于未经修饰的蛋白降低了(“负(minus)”或“阴(negative)”由“-”表示)至少-1,至少-2,至少-3,至少-4,至少-5,至少-10,

至少-15,至少-20,至少-25,至少-30,至少-35,至少-40,至少-45或至少-50。在某些实施方案中,超电荷蛋白的理论净电荷为至少-1,至少-2,至少-3,至少-4,至少-5,至少-10,至少-15,至少-20,至少-25,至少-30,至少-35,至少-40,至少-45或至少-50。

[0207] 尽管本文中描述了一些例示性超电荷蛋白以例示本发明的技术,但本公开在此方面不受限制。本领域技术人员基于本公开将能够确定另外的适宜超电荷蛋白用于将功能性效应器蛋白投递至细胞。可以修饰一些天然存在的蛋白质来生成适宜的超电荷蛋白。这类蛋白中的期望修饰可使用本领域中已知的任何技术完成。用于将这类变化引入蛋白序列的重组DNA技术是本领域公知的。在某些实施方案中,所述修饰可通过对编码蛋白质的多核苷酸的定点诱变完成。用于引入突变的其他技术论述于Molecular Cloning: A Laboratory Manual, 2nd Ed., Sambrook, Fritsch和Maniatis编(Cold Spring Harbor Laboratory Press:1989); 论文Methods in Enzymology (Academic Press, Inc., N.Y.); Ausubel等, Current Protocols in Molecular Biology (John Wiley & Sons, Inc., New York, 1999); 其各自通过提述并入本文。

[0208] 超电荷蛋白可进一步修饰。包含超电荷蛋白的蛋白可使用本领域技术人员已知的技术修饰。例如,可以化学或生物学修饰超电荷蛋白。可以从初始序列添加、删除或改变一个或多个氨基酸。例如,可以将聚组氨酸标签或其他标签添加至超电荷蛋白以辅助蛋白质的纯化。可将其他肽或蛋白质添加至超电荷蛋白上以改变该蛋白的生物学、生物化学和/或生物物理学特性。例如,可将核内体裂解性(endosomlytic)肽添加至超电荷蛋白的初始序列,或将靶向性肽添加至超电荷蛋白的初始序列。超电荷蛋白的其他修饰包括但不限于,翻译后修饰(例如糖基化、磷酸化、酰化、脂质化、法呢基化(farnesylation)、乙酰化、蛋白酶解等等)。在某些实施方案中,修饰所述超电荷蛋白以降低其免疫原性。在某些实施方案中,修饰所述超电荷蛋白以增强其投递功能性效应器蛋白(例如核酸酶、转录激活物/阻遏物、重组酶、Cas9蛋白包含其变体和融合物,等等)至细胞的能力。在某些实施方案中,所述超电荷蛋白缀合于聚合物。例如,可以通过将蛋白缀合于聚乙二醇(PEG)聚合物使蛋白PEG化。可使用其他方法来产生超电荷蛋白而不修饰蛋白质序列。例如,可将改变静电荷的模块附接于蛋白(例如通过化学或酶反应)以提供表面电荷来实现带超电荷。在某些实施方案中,使用记载于Shaw等,Protein Science 17:1446,2008的修饰蛋白质的方法来使当前公开的本发明技术中使用的蛋白质带超电荷。

[0209] 适用于本发明公开的技术的几种不同超电荷蛋白的变体的设计和创造记载于2007年6月1日提交的国际PCT专利申请PCT/US07/70254,在2007年12月13日公开为W0 2007/143574;2009年4月28日提交的国际PCT申请PCT/US09/041984,在2009年11月5日公开为W0 2009/134808;和2010年4月28日提交的国际PCT申请PCT/US10/001250,在2010年11月11日公开为W0 2010/129023;其各自完整内容通过提述并入本文。其中描述的一些公开的超电荷蛋白已显示为更稳定且保留其生物学功能,例如在荧光蛋白的情况中保留其荧光。例如,来自维多利亚发光水母(Aequorea victoria)的绿色荧光蛋白(GFP)记载于GenBank登录号P42212,通过提述并入本文。该野生型GFP的氨基酸序列如下:

[0210] MSKGEELFTGVVPIVELDGDVNGHKFSVSGEGEGDATYGKLT LKFICTTGKLPVPWPTLVTTF SYGVQ
C

[0211] FSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGH

- [0212] KLEYNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALS
- [0213] KDPNEKRDHMLLEFVTAAGITHGMDLYK (SEQ ID NO:14)
- [0214] 野生型GFP具有的理论净电荷为-7。已报道了具有理论净电荷-29, -30, -25, +15, +25, +36, +48和+49的变体, 例如在2010年4月28日提交的国际PCT申请PCT/US10/001250中, 在2010年11月11日公开为W0 2010/129023, 其全部内容通过提述并入本文。即使将该+36GFP加热至95°C后, 100%的变体蛋白是可溶的且该蛋白保留其荧光的 $\geq 70\%$ 。
- [0215] 本公开的一些方面基于以下发现, +36GFP有效将功能性效应器蛋白(例如核酸酶、转录激活物/阻遏物、重组酶、Cas9蛋白包含其变体和融合物, 等等)投递至靶细胞, 以及如此投递的效应器蛋白保留其生物学功能。因此, 认为具有静电荷为至少+15, 至少+25, 至少+30, 至少+35或至少+40的GFP或其他蛋白对于将功能性效应器蛋白引入细胞特别有用。
- [0216] 在一些实施方案中, 特别有用的超电荷蛋白是允许类似于+36GFP的电荷分布或表面电荷密度的蛋白质。进一步地, 在一些实施方案中, 特别有用的超电荷蛋白是展现出不容易被超电荷扰乱的稳定折叠结构的蛋白, 由此允许该超电荷蛋白较好地折叠。在一些实施方案中, 特别有用的超电荷蛋白是与本文或以下申请中描述的超电荷蛋白共享结构特征的蛋白: 2007年6月1日提交的国际PCT专利申请PCT/US07/70254, 在2007年12月13日公开为W0 2007/143574; 2009年4月28日提交的国际PCT申请PCT/US09/041984, 在2009年11月5日公开为W0 2009/134808; 和2010年4月28日提交的国际PCT申请PCT/US10/001250, 在2010年11月11日公开为W0 2010/129023; 其各自完整内容通过提述并入本文; 例如球状结构或 β 桶结构。可以通过本文中提供的方法和算法和本领域技术人员已知的其他方法将蛋白质折叠、蛋白质折叠结构稳定性和通过将特定氨基酸取代为不同电荷的氨基酸对蛋白质折叠的扰乱、电荷分布和表面电荷密度用计算机建模。因此, 对于本领域技术人员来说, 论述的超电荷蛋白是否将较好折叠在常规实验范围以内是明显的。如此, 本领域技术人员将能从给定的氨基酸序列鉴定出给定的超电荷蛋白是否将可用于依照本文描述的技术的功能性效应器蛋白的细胞投递。
- [0217] 一些例示性的GFP的适宜变体包括但不限于:
- [0218] +15GFP:
- [0219] MGHHHHHHGGASKGERLFTGVVPILEVELDGDVNGHKFSVRGEGEGDATRGKLTCLKFICTTGKLPVWP
- [0220] TLVTTLTYGVCFSRYPKHKMRHDFFKSAMPEGYVQERTISFKKDGTYKTRAEVKFEGRTLNVRIELKG
- [0221] RDFKEKGNILGHKLEYNFNSHNVYITADKRKNGIKANFKIRHNVKDGSVQLADHYQQNTPIGRGPVLL
PRNHYLSTRSALS KDPKEKRDHMLLEFVTAAGITHGMDLYK (SEQ ID NO:15)
- [0222] +25GFP:
- [0223] MGHHHHHHGGASKGERLFTGVVPILEVELDGDVNGHKFSVRGKKGKGDATRGKLTCLKFICTTGKLPVWP
- [0224] TLVTTLTYGVCFSRYPKHKMRHDFFKSAMPKGYVQERTISFKKDGTYKTRAEVKFEGRTLNVRIKLKG
- [0225] RDFKEKGNILGHKLRYNFNSHNVYITADKRKNGIKANFKIRHNVKDGSVQLADHYQQNTPIGRGPVLL
PRNHYLSTRSALS KDPKEKRDHMLLEFVTAAGITHGMDLYK (SEQ ID NO:16)
- [0226] +36GFP:
- [0227] MGHHHHHHGGASKGERLFRGKVPILVELKGDVNGHKFSVRGKKGKGDATRGKLTCLKFICTTGKLPVWP
- [0228] TLVTTLTYGVCFSRYPKHKMRHDFFKSAMPKGYVQERTISFKKDGTYKTRAEVKFEGRTLNVRIKLKG
- [0229] RDFKEKGNILGHKLRYNFNSHKVYITADKRKNGIKAKFKIRHNVKDGSVQLADHYQQNTPIGRGPVLL

PRNHYLSTRSKLSKDPKEKRDHMLLEFVTAAGIKHGRDERYK (SEQ ID NO:17)

[0230] +42GFP:

[0231] MGHHHHHHGGRSKGKRLFRGKVPILVELKGDVNGHKFSVRGKKGDATRGKLTCLKFICTTGKLPVPWP

[0232] TLVTTLTYGVCFSRYPKHMKRHDFFKSAMPKGYVQERTISFKKDGKYKTRAEVKFEGRTLNVRIKLG

[0233] RDFKEKGNILGHKLRYNFNNSHKVYITADKRKNGIKAKFKIRHNVKDGSVQLADHYQQNTPIGRGPVLL

PRKHYLSTRSKLSKDPKEKRDHMLLEFVTAAGIKHGRKERYK (SEQ ID NO:18)

[0234] +48GFP:

[0235] MGHHHHHHGGRSKGKRLFRGKVPILVCLKGDVNGHKFSVRGKKGDATRGKLTCLKFICTTGKLPVPWP

[0236] TLVTTLTYGVCFSRYPKHMKRHDFFKSAMPKGYVQERTISFKKDGKYKTRAEVKFKGRTLNVRIKLG

[0237] RDFKEKGNILGHKLRYNFNNSHKVYITADKRKNGIKAKFKIRHNVKDGSVQLAKHYQQNTPIGRGPVLL

PRKHYLSTRSKLSKDPKEKRDHMLLEFVTAAGIKHGRKERYK (SEQ ID NO:19)

[0238] +49GFP:

[0239] MGHHHHHHGGRSKGKRLFRGKVPILVCLKGDVNGHKFSVRGKKGDATRGKLTCLKFICTTGKLPVPWP

[0240] TLVTTLTYGVCFSRYPKHMKRHDFFKSAMPKGYVQERTISFKKDGKYKTRAEVKFKGRTLNVRIKLG

[0241] RDFKEKGNILGHKLRYNFNNSHKVYITADKRKNGIKAKFKIRHNVKDGSVQLAKHYQQNTPIGRGPVLL

PRKHYLSTRSKLSKDPKEKRDHMLKEFVTAAGIKHGRKERYK (SEQ ID NO:20)

[0242] (-) 30GFP:

[0243] MGHHHHHHGASKGEELFDGVVPILEVELDGDVNGHEFSVRGEGEGDATEGELTLCLKFICTTGELPVPWPPT

[0244] LVTTLTYGVQCFSYDPDMDQHDFFKSAMPEGYVQERTISFKDDGTYKTRAEVKFEGDTLVNRIELKGI

[0245] DFKEDGNILGHKLEYNFNNSHDVYITADKQENGIKAEFEIRHNVEDGSVQLADHYQQNTPIGDGPVLLP

DDHYLSTESALSKDPNEDRDHMLLEFVTAAGIDHGMDLYK (SEQ ID NO:21)

[0246] 对于熟练技术人员而言以下将是明显的,上述序列包含N末端His6标签,和没有这类标签或具有不同标签的序列也是适宜的。

[0247] 为了在投递至细胞后促进功能性效应器蛋白(例如核酸酶、转录激活物/阻遏物、重组酶、Cas9蛋白包含其变体和融合物,等等)的生物学功能,可能期望在细胞摄取后增强功能性效应器蛋白的核内体逃离。可将超电荷蛋白或功能性效应器蛋白与已知增强核内体降解或核内体的裂解的蛋白、肽或其他实体融合或联合。在某些实施方案中,所述肽是血凝素2(HA2)肽,已知其增加核内体降解。在某些具体的实施方案中,HA2肽融合于带超电荷的GFP(例如+36GFP)。在某些具体的实施方案中,融合蛋白具有序列:

[0248] +36GFP-HA2

[0249] MGHHHHHHGASKGERLFRGKVPILVELKGDVNGHKFSVRGKKGDATRGKLTCLKFICTTGKLPVPWP

[0250] TLVTTLTYGVCFSRYPKHMKRHDFFKSAMPKGYVQERTISFKKDGKYKTRAEVKFEGRTLNVRIKLG

[0251] RDFKEKGNILGHKLRYNFNNSHKVYITADKRKNGIKAKFKIRHNVKDGSVQLADHYQQNTPIGRGPVLLP

[0252] RNHYLSTRSKLSKDPKEKRDHMLLEFVTAAGIKHGRDERYKGSAGSAGSGEFLFGAIAGFIENGWEGMIDG (SEQ ID NO:22)

[0253] 在某些实施方案中,核内体裂解性肽是蜂毒肽(melittin)(GIGAVLKVLTTGLPALISWIKRKRQQ,SEQ ID NO:23)(Meyer等,JACS130(11):3272-3273,2008;其通过提述并入本文)。在某些实施方案中,通过1、2、3、4或5个氨基酸取代、缺失和/或添加修饰蜂毒肽。在某些实施方案中,蜂毒肽具有序列:CIGAVLKVLTTGLPALISWIKRKRQQ (SEQ ID

NO:24)。在某些具体的实施方案中,蜂毒肽融合于带超电荷的GFP(例如+36GFP)。

[0254] 在某些实施方案中,核内体裂解性肽是penetratin肽(RQIKIWFQNRRLKWKK-酰胺, SEQ ID NO:25),牛PrP(1-30)肽(MVKSKIGSWILVLFVAMWSDVGLCKKRPKP-酰胺, SEQ ID NO:26),MPG Δ^{NLS} 肽(由于K->S取代,其缺少功能性核定位序列)(GALFLGWLGAAGSTMGAPKSKRKV, SEQ ID NO:27),TP-10肽(AGYLLGKINLKALAALAKKIL-酰胺, SEQ ID NO:28),和/或EB1肽(LIRLWSHLIHIWFQNRRLKWKKK-酰胺, SEQ ID NO:29)(Lundberg等,2007,FASEB J.21:2664;通过提述并入本文)。在某些实施方案中,通过1、2、3、4或5个氨基酸取代、缺失和/或添加修饰penetratin,PrP(1-30),MPG,TP-10,和/或EB1肽。在某些具体的实施方案中,PrP(1-30)、MPG、TP-10和/或EB1肽融合于带超电荷的GFP(例如+36GFP)。在一些实施方案中,Aurein肽融合于所述超电荷蛋白。

[0255] 可将其他肽或蛋白质融合于超电荷蛋白或包含超电荷蛋白和功能性效应器蛋白(例如核酸酶、转录激活物/阻遏物、重组酶、Cas9蛋白包含其变体和融合物,等等)的融合蛋白。例如,靶向性肽可融合于超电荷蛋白以将功能性效应器蛋白选择性投递至特定细胞类型。还可以使用增强功能性效应器蛋白的细胞摄取的肽或蛋白。在某些实施方案中,融合于超电荷蛋白的肽是肽激素。在某些实施方案中,融合于超电荷蛋白的肽是肽配体。

[0256] 本文中详细描述的例子性超电荷蛋白不意图限制本公开,且本领域技术人员将领会其他超电荷蛋白可用于功能性效应器蛋白(例如核酸酶、转录激活物/阻遏物、重组酶、Cas9蛋白包含其变体和融合物,等等)的细胞投递,所述功能性效应器蛋白包括但不限于其他GFP类型的荧光蛋白。在某些实施方案中,所述超电荷蛋白是蓝色荧光蛋白的超电荷型式。在某些实施方案中,所述超电荷蛋白是青色荧光蛋白的超电荷型式。在某些实施方案中,所述超电荷蛋白是黄色荧光蛋白的超电荷型式。例子性的适宜荧光蛋白包括但不限于,增强的绿色荧光蛋白(EGFP),AcGFP,TurboGFP,Emerald,Azami Green,ZsGreen,EBFP,Sapphire,T-Sapphire,ECFP,mCFP,Cerulean,CyPet,AmCyan1,Midori-Ishi Cyan,mTFP1(Teal),增强的黄色荧光蛋白(EYFP),Topaz,Venus,mCitrine,YPet,PhiYFP,ZsYellow1,mBanana,Kusabira Orange,mOrange,dTomato,dTomato-Tandem,DsRed,DsRed2,DsRed-Express(T1),DsRed-Monomer,mTangerine,mStrawberry,AsRed2,mRFP1,JRed,mCherry,HcRed1,mRaspberry,HcRed1,HcRed-Tandem,mPlum和AQ143。

[0257] 仍然有其他蛋白可以超电荷和使用在例如如本文中公开的功能性效应器蛋白(例如核酸酶、转录激活物/阻遏物、重组酶、Cas9蛋白包含其变体和融合物,等等)的投递中,其包括组蛋白组分或组蛋白样蛋白、高速泳动族蛋白(high-mobility-group proteins, HMG)、酶(例如淀粉酶、果胶酶、水解酶、蛋白酶、葡萄糖异构酶、脂肪酶、肌醇六磷酸酶、阿糖脑苷酶(alglucerase)、伊米苷酶(imiglucerase)、半乳糖苷酶(agalsidase) beta, α -1-艾杜糖苷酸酶(iduronidase)、酸性 α -葡萄糖苷酶和艾杜糖醛酸-2-硫酸酯酶、N-乙酰半乳糖胺-4-硫酸酯酶。

[0258] 还可以使用带电荷的聚合物而非蛋白来投递功能性效应器蛋白。另外,如本文中更多细节描述的,还可以使用阳离子脂质和脂质样材料以及阳离子聚合物来投递功能性效应器蛋白。适宜的阳离子脂质、脂质样材料和阳离子聚合物在本文中公开且另外的适宜脂质和脂质样材料是本领域技术人员已知的(参见,例如记载于Akinc等,Nature Biotechnology 26,561-569(2008)的那些,其完整内容通过提述并入本文)。

[0259] 使用超电荷蛋白投递功能性效应器蛋白

[0260] 本发明提供用于体内、离体或体外向细胞投递功能性效应器蛋白(例如核酸酶、转录激活物/阻遏物、重组酶、Cas9蛋白包含其变体和融合物,等等)的系统和方法。这类系统和方法通常牵涉将功能性效应器蛋白与超电荷蛋白联合以形成复合物或融合蛋白,并将该复合物或融合蛋白投递给细胞。在一些实施方案中,要通过超电荷蛋白投递的功能性效应器蛋白具有治疗活性。在一些实施方案中,将所述复合物或融合蛋白投递给细胞牵涉将包含与功能性效应器蛋白联合的超电荷蛋白的复合物或融合蛋白施用给有此需要的受试者。

[0261] 在一些实施方案中,功能性效应器蛋白(例如核酸酶、转录激活物/阻遏物、重组酶、Cas9蛋白包含其变体和融合物,等等)本身不能进入细胞,但当与超电荷蛋白联合(例如经由共价键或非共价相互作用)时能进入细胞。在一些实施方案中,提供包含共价结合于超电荷蛋白的功能性效应器蛋白的组合物。在一些实施方案中,所述组合物包含经由肽键,例如经由直接融合或经由肽接头融合于超电荷蛋白的功能性效应器蛋白。在一些实施方案中,所述组合物包含功能性效应器蛋白,其通过非共价相互作用结合于超电荷蛋白。在一些实施方案中,利用超电荷蛋白以允许功能性效应器蛋白进入细胞。在一些实施方案中,在投递至细胞后将投递给细胞的与超电荷蛋白联合的功能性效应器蛋白与超电荷蛋白分开,例如通过在特定细胞微环境例如核内体中的细胞蛋白酶(例如核内体蛋白酶)切割接头肽或通过功能性效应器蛋白从超电荷蛋白解离。在一些实施方案中,通过本公开的系统或方法投递给细胞的功能性效应器蛋白具有治疗活性。

[0262] 在一些实施方案中,通过本文中提供的系统、组合物或方法体内、离体或体外地将功能性效应器蛋白(例如核酸酶、转录激活物/阻遏物、重组酶、Cas9蛋白包含其变体和融合物,等等)投递给细胞。在一些实施方案中,功能性效应器蛋白是能在靶细胞内实施生物学功能的蛋白质,例如能结合其底物和催化靶细胞内的酶反应的酶,例如能结合和切割靶细胞内的核酸分子的核酸酶,或能与靶细胞的基因组相互作用和活化或抑制细胞中靶基因的转录的转录因子。

[0263] 在一些实施方案中,生成功能性效应器蛋白和超电荷蛋白的融合物的方法包括生成表达核酸构建体,其含有功能蛋白和超电荷蛋白的编码序列,以及任选地,符合阅读框的肽接头;在原核或真核细胞中培养来表达这类重组融合蛋白,融合蛋白的提取和纯化。在一些实施方案中,以表达载体的形式生成核酸构建体,例如适用于在细菌宿主中增殖和适用于在原核或真核细胞中表达的载体。

[0264] 在一些实施方案中,通过以下生成适用于融合蛋白表达的载体:将编码要投递的功能性效应器蛋白的核苷酸序列克隆到克隆载体中,所述载体包含在真核和/或原核启动子控制下的编码超电荷蛋白的核苷酸序列,通过产生两个编码序列均彼此符合阅读框地克隆办法来进行。在一些实施方案中,所述克隆载体包含在编码超电荷蛋白的核苷酸序列和限制性位点之间的编码肽接头的核苷酸序列,所述限制性位点可用于插入编码与该接头和超电荷蛋白符合阅读框的蛋白质的核苷酸序列。在一些实施方案中,该克隆载体还包含另外的序列,其增强融合蛋白在原核或真核细胞中的表达或促进所表达的融合蛋白从这类细胞的纯化,例如所述序列是稳定化编码融合蛋白的转录本的序列,如聚A信号、可剪接内含子、编码符合阅读框的肽或蛋白质域标签的序列(例如Arg标签、钙调蛋白结合肽标签、纤维素结合域标签、DsbA标签、c-myc标签、谷胱甘肽S-转移酶标签、FLAG标签、HAT标签、His标

签、麦芽糖结合性蛋白标签、NusA标签、S标签、SBP标签、Strep标签或硫氧还蛋白标签),或选择标志或报道盒,其允许鉴定携带且表达所述表达构建体的细胞和/或量化这类细胞中的表达水平。用于克隆和表达融合蛋白的方法是本领域技术人员公知的,参见例如,Sambrook等,Molecular Cloning:ALaboratory Manual,Volume 1-3,CSHL Press(1989); Gellissen等,Production of recombinant proteins,Wiley-VCH,2005。

[0265] 在一些实施方案中,所述功能性效应器蛋白与超电荷的GFP例如+36GFP或-30GFP联合以投递至靶细胞。先前已证明了核内体破坏在通过超电荷蛋白投递大分子中的益处(Wadia等,Nat.Med.10,310-315,2004)且在一些实施方案中,实施另外的步骤来实现增强的核内体逃离,如本文提供的或本领域已知的。当与有效的核内体释放偶联时,高度有效的蛋白质内化具有最小化外源蛋白剂的必要剂量的潜力,从而增强其作为研究工具和导致治疗进展的潜力。

[0266] 在一些实施方案中,在分离和/或纯化后将包含与超电荷蛋白联合的功能性效应器蛋白的组合物施用给靶细胞。蛋白质纯化方法和技术是本领域技术人员公知的,且包括例如,亲和层析或免疫沉淀。适用于分离和/或纯化特定功能性效应器蛋白、超电荷蛋白、和/或融合蛋白的方法将取决于相应蛋白质的性质。例如,带His标签的融合蛋白可容易地经由Ni或Co离子层析分离和纯化,而带有其他肽或域的融合蛋白或未带标签的融合蛋白可通过其他确立的方法来纯化。

[0267] 适用于通过本文提供的系统或方法体内、离体或体外投递至靶细胞的功能性效应器蛋白对于本领域技术人员将是明显的,且包括例如DNA结合蛋白,如转录因子和核酸酶,以及Cas9蛋白(包括其变体和融合物)。

[0268] 在一些实施方案中,使用本文提供的方法、组合物或系统来投递功能性效应器蛋白至细胞。治疗性蛋白的例子包括但不限于,靶向与疾病或病症有关的基因组等位基因的核酸酶和Cas9蛋白(包括其变体和融合物),和激活有益基因或抑制致病基因的转录因子。

[0269] 在一些实施方案中,Cas9融合于超电荷蛋白以投递至细胞。在一些实施方案中,所述超电荷蛋白是带正电荷的。在一些实施方案中,融合于Cas9的所述超电荷蛋白是(+36)GFP。在一些实施方案中,Cas9和(+36)GFP的融合物包含SEQ ID NO:30的氨基酸序列(例如有或无核定位信号(NLS)和有或无6xHis标签),或包含与SEQ ID NO:30的氨基酸序列至少80%,至少85%,至少90%,至少95%,至少98%或至少99%相同的氨基酸序列(例如有或无核定位信号(NLS)和有或无6xHis标签)。在一些实施方案中,融合于Cas9的所述超电荷蛋白是(-30)GFP。在一些实施方案中,Cas9和(-30)GFP的融合物包含SEQ ID NO:31的氨基酸序列(例如有或无核定位信号(NLS)和有或无6xHis标签),或包含与SEQ ID NO:31的氨基酸序列至少80%,至少85%,至少90%,至少95%,至少98%或至少99%相同的氨基酸序列(例如有或无核定位信号(NLS)和有或无6xHis标签)。

[0270] (+36) dGFP-NLS-Cas9-6xHis **(Y67S)**:

MGASKGERLFRGKVPILVELKGDVNGHKFSVRGKGGKGDATRGLTLKFICTTGKLPV
PWPTLVTTLTSGVQCFSRYPKHMKRHDFFKSAMPKGYVQERTISFKKDGKYKTRAEV
KFEGRTL VNRIKLKGRDFKEKGNILGHKLRYNFNESHKYYITADKRKNGIKAKFKIRHN
VKDGSVQLADHYQQNTPIGRGPVLLPRNHYLSTRSKLSKDPKEKRDHMLLEFVTA
GIKHGRDERYKTGGSGGSGGSGGSGGSGGSGGSGGSGGSGGTALALPKKKRKVMDKKYSI
GLDIGTNSVGWAVITDEYKVPSSKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLK
RTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLVESFLVEEDKKHERHPIFGNIVD
EVAYHEKYPTIYHLRKKLVDSTDKADLRILIYALAHMIKFRGHFLIEGDLNPDNSDVD
KLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLI
ALSLGLTPNFKSNFDLAEDAQLSKDQYDDDLNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGY
AGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGE
LHAILRRQEDFYFPLKDNREKIEKILTFRIPIYVGPLARGNSRFAWMTRKSEETITPWNF
EEVVDKGASQAQSFIERMTNFDKNLPNEKVLPHSLLYEYFTVYNELTKVKYVTEGMR
KPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGT
YHDLLKIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKYAHLFDDKVMKQLKR
RRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKA
QVSGQGDSLHEHIANLAGSPAIKKILQTVKVVDELVKVMGRHKPENIVIAMARENQT
TQKGQKNSRERMKRIEELGKELGSQLKEHPVENTQLQNEKLYLYLQNGRDMYVDQ
ELDINRLSDYDVDHIVPQSFLKDDSIDNKVLRSDKNRKGKSDNVPSEEVVKKMKNY
WRQLLNAKLITQRKFDNLTAKAERGGLSELDKAGFIKQRLVETRQITKHVAQILDSRMNT
KYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVRINNYHHAHDAYLNAVVGALIK
KYPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIR
KRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNVKKTVEVQTGGFSKESILPKRNS
DKLIARKKDWDPKKGFFDSPTVAYSVLVVAKEKGGKSKLKSVELLGITIMERS
SSFEKNPIDFLEAKGYKEVKKDLIKLPKYSLELENGRKRMLASAGELQKGNELALPSKY
VNFLYLASHYEKLGKSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVLADANL
DKVLSAYNKHDKPIREQAENIHLFTLNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQ
SITGLYETRIDLSQLGGDHHHHHHH (SEQ ID NO:30)

[0271] (-30) dGFP-NLS-Cas9-6xHis **(Y67S)**:

MGASKGEELFDGVPILVELDGDVNGHEFSVRGEGEGDATEGELTLKFICTTGELPVPWPTLV
TTLTSGVQCFSYDPDHMDQHDFFKSAMPEGYVQERTISFKDDGTYKTRAEVKFEGDTLVNRIE
LKGIDFKEDGNILGHKLEYNFNSHDVYITADKQENGIKAEFEIRHNVEDGSVQLADHYQQNTPI
GDGPVLLPDDHYLSTESALSKDPNEDRDHMLLEFVTAAGIDHGMDELKTGGSGGSGGSGG
SGGSGGSGGSGGSGGSGGTALALPKKKRKVMDKKYSIGLDIGTNSVGWAVITDEYKVPSSKFKVL
GNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFH
RLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRILIYALAHMIK
FRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQ
LPGEKKNGLFGNLIASLGLTPNFKSNFDLAEDAQLSKDQYDDDLNLLAQIGDQYADLFL
AAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSK
NGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELH
AILRRQEDFYFPLKDNREKIEKILTFRIPIYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKG
ASAQSFIERMTNFDKNLPNEKVLPHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKA
VDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIKDKDFLDNEEN
EDILEDIVLTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSG
KTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKILQTV
KVVDDELVKVMGRHKPENIVIAMARENQTTQKGQKNSRERMKRIEELGKELGSQLKEHPVENT
QLQNEKLYLYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLRSDKNRKGK
SDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTAKAERGGLSELDKAGFIKQRLVETRQITK

[0272] HVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKREINNYHHHAHDAYLNAV
VVGITALIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLANG
EIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIKKTEVQTGGFSKESILPKRNSDKLI
ARKKDWDPKKYGGFDSPTVAYSVLVAKVEKGKSKLKSVKELLGITIMERSSEKFNPIDFLE
AKGYKEVKKDLIKLPKYSLENGRKRMLASAGELQKGNELALPSKYVNFYLYASHYEKLLK
GSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVLADANLDKVL SAYNKHRDKPIREQAENIHH
LFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSI TGLYETRIDLSQLGGDHHHHHHH
(SEQ ID NO:31)

[0273] 功能性效应器蛋白和阳离子脂质的组合物

[0274] 本公开的某些方面涉及使用阳离子脂质来投递效应器蛋白(例如核酸酶、转录激活物/阻遏物、重组酶、Cas9蛋白包含其变体和融合物,等等),例如与投递“裸”蛋白制备物相对。令人惊讶地,发现已工程化用于投递核酸如DNA和RNA的现有脂质体投递试剂有效体外或体内投递某些效应器蛋白(例如Cas9蛋白包含其变体和融合物),如本文中描述的。核酸投递极大地受益于在过去二十年的脂质体试剂的进展。阳离子脂质体制剂使得DNA和RNA转染变为基础研究的常规技术且已用于临床试验。脂质体的脂质双层保护包封的核酸免于降解且可防止可结合核酸的裸制备物的抗体的特异性中和。重要地是,在核内体熟化期间脂质体与核内体膜的融合能实现高度有效的阳离子脂质投递货物的核内体逃离。其他非阳离子的然而可逆离子化的脂质纳米颗粒制备物实现核酸的有效包封和投递,同时避免非特异性静电相互作用和相应的扣留。然而,蛋白质是化学上多样的,因此不同于高度阴离子的核酸,脂质体制剂对于有效投递蛋白质不那么成功。例如,虽然蛋白质可以由再水合的脂质非特异性包封并体外投递,包封的功效依赖于蛋白质浓度且通常低效,如此没有广泛应用。本公开的方面涉及以下认知,阴离子蛋白或蛋白复合物(包含与核酸联合的那些蛋白)可以能利用与阳离子脂质体试剂用于核酸投递相同的静电驱动的包封。尽管很少的蛋白质天然拥有在核酸的磷酸骨架中见到的负电荷的密度,但如本文描述的翻译融合于或非共价联合于阴离子载体如带超负电荷的蛋白或核酸使得所得效应器蛋白或蛋白复合物足够阴离子化的,从而驱动这类蛋白货物被阳离子脂质体试剂有效包封。

[0275] 在一些实施方案中,与工程化的带超负电荷的GFP联合或融合能驱动通过常用于转染核酸的阳离子脂质将蛋白质有效包封和投递到培养的哺乳动物细胞中。此办法甚至在低纳米摩尔蛋白浓度和存在血清时也有效,从而比使用与阳离子肽或蛋白质的融合物的蛋白投递方法高1,000倍有效的投递功能性蛋白。如实施例中显示的,投递功效在一些实施方案中依赖于例如融合标签的理论净电荷,和流行的天然阴离子肽标签例如3xFLAG和VP64可类似地实现脂质体蛋白投递。

[0276] 实施例进一步显示与聚阴离子导引RNA(gRNA)联合的Cas9核酸酶蛋白可以通过这些常用的阳离子脂质体制剂以功能性形式有效投递到哺乳动物细胞中,因为,不希望受任何具体理论束缚地,认为gRNA充当另外阳离子的Cas9蛋白和阳离子脂质之间的聚阴离子介导体。Cas9:gRNA复合物的投递不仅高度有效(例如从单次处理高达80%修饰),而且还导致与质粒转染相比显著更高的基因组修饰特异性,其通常导致在在靶:脱靶修饰比上高>10倍,推定是由于投递的Cas9:gRNA活性的瞬时性质。在一些实施方案中,Cas9:gRNA复合物的投递导致至少2倍,3倍,4倍,5倍,6倍,7倍,8倍,9倍,10倍,11倍,12倍,13倍,14倍,15倍,20倍或25倍或更高的在靶:脱靶修饰比。实施例还显示该蛋白投递办法可以是体内有效的,例如通过将功能性Cre重组酶和功能性Cas9:gRNA复合物投递至小鼠内耳的毛细胞。

[0277] 因此,本公开的一些方面提供包含Cas9蛋白(例如如本文描述的;见例如下文Cas9效应器蛋白)和阳离子脂质的组合物,其能将Cas9蛋白投递至细胞的内部。在一些实施方案中,Cas9蛋白与gRNA联合,后者例如提供阴离子电荷给复合物,从而允许Cas9:gRNA复合物被阳离子脂质包封。在一些实施方案中,Cas9蛋白不需要与gRNA联合以被阳离子脂质有效包封,而是与超负电荷蛋白联合,如本文描述的。在其中Cas9蛋白与超负电荷蛋白联合的一些实施方案中,Cas9蛋白还与gRNA联合。在一些实施方案中,Cas9蛋白是野生型Cas9蛋白,野生型Cas9蛋白的片段,或野生型Cas9蛋白的变体。在一些实施方案中,Cas9蛋白包含dCas9域(例如如本文描述的)。在一些实施方案中,Cas9蛋白是包含dCas9域(例如如本文描述的)的融合蛋白。在一些实施方案中,Cas9蛋白是Cas9切口酶。

[0278] 在其他实施方案中,提供包含效应器蛋白(例如除Cas9蛋白以外的)和阳离子脂质的组合物,其能将效应器蛋白投递至细胞内部(例如至细胞的核)。效应器蛋白是天然带负电荷的,经修饰以具有总体净负电荷的,或与超负电荷蛋白关联,如本文描述的。在一些实施方案中,效应器蛋白是本文描述的任意效应器蛋白。在一些实施方案中,效应器蛋白是重组酶,例如本文描述的任意重组酶。在一些实施方案中,重组酶是Cre重组酶。在一些实施方案中,所述Cre重组酶包含SEQ ID NO:32的氨基酸序列(例如有或无6xHis标签)。在一些实施方案中,Cre重组酶包含与SEQ ID NO:32的氨基酸序列至少80%,至少85%,至少90%,至少95%,至少98%或至少99%相同的氨基酸序列(例如有或无6xHis标签)。在一些实施方案中,Cre重组酶融合于超电荷蛋白(例如+36GFP或-30GFP)。在一些实施方案中,融合于超电荷蛋白的Cre重组酶包含SEQ ID NO:33(例如有或无6xHis标签)或SEQ ID NO:34(例如有或无6xHis标签)的氨基酸序列,或包含与SEQ ID NO:33或SEQ ID NO:34的氨基酸序列至少80%,至少85%,至少90%,至少95%,至少98%或至少99%相同的氨基酸序列(例如有或无6xHis标签)。在一些实施方案中,效应器蛋白是TALE蛋白(例如如本文描述的包含实施例中提供的那些)。在一些实施方案中,TALE蛋白包含一个或多个VP64转录激活物域(例如SEQ ID NO:35)。在一些实施方案中,具有VP64转录激活物域的TALE蛋白还包含选自下组的氨基酸序列:SEQ ID NO:36-39(例如有或无6xHis标签)。在一些实施方案中,具有VP64转录激活物域的TALE蛋白包含与选自下组的氨基酸序列至少80%,至少85%,至少90%,至少95%,至少98%或至少99%相同的氨基酸序列(例如有或无6xHis标签):SEQ ID NO:36-39。在一些实施方案中,TALE效应器蛋白包含(-30)GFP域(例如SEQ ID NO:21或SEQ ID NO:40),TALE域的N末端区(例如SEQ ID NO:41),可变重复域(例如18.5聚体重复域,如Maeder等,“Robust, synergistic regulation of human gene expression using TALE activators.”*Nat. Methods*.2013;10,243-245中提供的),C末端TALE域(例如SEQ ID NO:42),VP64激活域(例如SEQ ID NO:35),和任选地在任何域之间的一个或多个接头(例如GGS(9),SEQ ID NO:252)以及任选地序列标签(例如6xHis,SEQ ID NO:253)。

[0279] 尽管已知货物如DNA和RNA的脂质体投递诱导靶向的细胞中的毒性,但发现本文描述的本发明组合物令人惊讶地在体外和体内均无毒性或低毒性地投递其货物。例如,在一些实施方案中,本文描述的包含Cas9蛋白或其他效应器蛋白的组合物在施用给(例如体外或体内)细胞群体时展现出低毒性。在一些实施方案中,群体中至少60%,至少65%,至少70%,至少75%,至少80%,至少85%,至少90%,至少95%或至少99%的细胞在施用本发明组合物后是活的,所述组合物包含Cas9蛋白或其他效应器蛋白和阳离子脂质。用于评估当

施用给细胞群体时组合物毒性的方法是本领域公知的且包括实施例中描述的那些。

[0280] Cre-6xHis (6xHis标签划线) :

[0281] MASNLLTVHQNLPALPVDATSDEVKRNLMDFRDRQAFSEHTWKMLLSVCRSWAAWCKLNNRKWFP
PAEPEDVRDYLLYLQARGLA VKTIQQHLGQLNMLHRRSGLPRPSDSNAVSLVMRRIRKENVDAGERAKQALAFERTDFD
QVRSLMENS DRCQDIRNLAFLGIA YNTLLRIA E IARIRVKDISRTD GGRMLIHIGRTKTLVSTAGVEKALS LGVTK
LVERWISVSGVADDPNNYLCRVRKNGVAAPSATSQLSTRALEGIFEATHRLIYGAKDDSGQRYLAWSGHSARVGA
ARDMARAGVSIPEIMQAGGWTNVNIVMNYIRNLDSETGAMVRLLEDGDGGSHHHHHH (SEQ ID NO:32)

[0282] (+36)GFP-Cre-6xHis (+36GFP双下划线;6xHis标签划线) :

MGASKGERLFRGKVPILVELKGDVNGHKFSVRGKKGKDATRGLTKLTKFICTTGKLPVPWPTL
VTTLTYGVQCFSRYPKHMKRHDFFKSAMPKGYVQERTISFKKDGKYKTRAEVKFEGRTL VNR
IKLKGRDFKEKGNILGHKLRYNFNESHK VYITADKRKNGIKAKFKIRHNVKDGSVQLADHYQQ
NTPIGRGPVLLPRNHYLSTRSKLSKDPKEKRDHMLLEFVTAAGIKHGRDERYKTGGSGGSGG
SGGSGGSGGSGGSGGSGGTASNLLTVHQNLPALPVDATSDEVKRNLMDFRDRQAFSEHTW
[0283] KMLLSVCRSWAAWCKLNNRKWFP
PAEPEDVRDYLLYLQARGLA VKTIQQHLGQLNMLHRRS
GLPRPSDSNAVSLVMRRIRKENVDAGERAKQALAFERTDFDQVRSLMENS DRCQDIRNLAFLG
IA YNTLLRIA E IARIRVKDISRTD GGRMLIHIGRTKTLVSTAGVEKALS LGVTKLVERWISVSGV
ADDPNNYLCRVRKNGVAAPSATSQLSTRALEGIFEATHRLIYGAKDDSGQRYLAWSGHSAR
VGAARDMARAGVSIPEIMQAGGWTNVNIVMNYIRNLDSETGAMVRLLEDGDGGSHHHHHH
(SEQ ID NO:33)

[0284] (-30)GFP-Cre-6xHis (-30GFP双下划线;6xHis标签划线) :

MGASKGEELFDGVVPILVELDGDVNGHEFSVRGEGEGDATEGELTKLTKFICTTGELPVPWPTLV
TTLTYGVQCFSDYPDHMDQHDFFKSAMPEGYVQERTISFKDDGTYKTRAEVKFEGDTLVNRIE
LKGIDFKEDGNILGHKLEYNFNESHVYITADKQENGIKAEFEIRHNVEDGSVQLADHYQQNTPI
GDGPVLLPDDHYLSTESALSKDPNEDRDHMLLEFVTAAGIDHGMDEL YKTGGSGGSGGSGG
SGGSGGSGGSGGSGGSGGTASNLLTVHQNLPALPVDATSDEVKRNLMDFRDRQAFSEHTWKML
[0285] LSVCRSWAAWCKLNNRKWFP
PAEPEDVRDYLLYLQARGLA VKTIQQHLGQLNMLHRRSGLPR
PSDSNAVSLVMRRIRKENVDAGERAKQALAFERTDFDQVRSLMENS DRCQDIRNLAFLGIA YN
TLLRIA E IARIRVKDISRTD GGRMLIHIGRTKTLVSTAGVEKALS LGVTKLVERWISVSGVADDP
NNYLCRVRKNGVAAPSATSQLSTRALEGIFEATHRLIYGAKDDSGQRYLAWSGHSARVGA
RDMARAGVSIPEIMQAGGWTNVNIVMNYIRNLDSETGAMVRLLEDGDGGSHHHHHH (SEQ
ID NO:34)

[0286] (+36)GFP-PPAR γ -TALE-2 (+36GFP双下划线;6xHis标签划线) :

MGASKGERLFRGKVPILVELKGDVNGHKFSVRGKKGKDATRGLTKLTKFICTTGKLPVPWPTL
VTTLTYGVQCFSRYPKHMKRHDFFKSAMPKGYVQERTISFKKDGKYKTRAEVKFEGRTL VNR
IKLKGRDFKEKGNILGHKLRYNFNESHK VYITADKRKNGIKAKFKIRHNVKDGSVQLADHYQQ
NTPIGRGPVLLPRNHYLSTRSKLSKDPKEKRDHMLLEFVTAAGIKHGRDERYKTGGSGGSGG
SGGSGGSGGSGGSGGSGGTAPKKKRKVGHIRGVPMVDLRTLGYSSQQQEKIKPKVRSTVAQH
HEALVGHGFTHAHIVALSQHPAALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEAL
LTVAGELRGPPQLDTGQLLKIAKRGVTA VEA VHA WRNALTGAPLNLTPDQVVAIASNIGG
KQALETVQRLLPVLCQDHGLTPEQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIA
SNGGGKQALETVQRLLPVLCQAHGLTPAQVVAIANNNGGKQALETVQRLLPVLCQDHGLTPD
[0287] QVVAIASNIGGKQALETVQRLLPVLCQDHGLTPEQVVAIASNIGGKQALETVQRLLPVLCQAH
GLTPDQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPAQVVAIASHDGGKQALETVQRLLP
VLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPEQVVAIASNIGGKQALETV
QRLLPVLCQAHGLTPDQVVAIANNNGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNIGGK
QALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPEQVVAIAS
NIGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPAQ
VVAIANNNGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDH
GLTPEQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNIGGRPALESIVAQLSRP
DPALAALTNDHLVALACLGGPALDAVKKGLPHAPALIKRTNRRIPERTSHRVADHAQVVRV

[0288] LGFFQCHSHPAQAFDDAMTQFGMSGGSGRADALDDFDLDMLGSDALDDFDLDMLGSDALD
DFDLDMLGSDALDDFDLDMLHHHHHHH (SEQ ID NO:36)

[0289] (+36)GFP-PRDM16_TALE-3(+36GFP双下划线;6xHis标签划线):

MGASKGERLFRGKVPILVELKGDVNGHKFSVRGKGGKGDATRGKLTTKFICTTGKLPVPWPTL
VTTLTYGVQCFSRYPKHMKRHDFFKSAMPKGYVQERTISFKKDGKYKTRAEVKFEGRTLVNR
IKLKGRDFKEKGNILGHKLRYNFNSHKVYITADKRKNGIKAKFKIRHNVKDGSVQLADHYQQ
NTPIGRGPVLLPRNHYLSTRSKLSKDPKEKRDHMLLEFVTAAGIKHGRDERYKTGGSGGSGG
SGGSGGSGGSGGSGGSGGTAPKKKRVGIHRGVPMVDLRTLGYSSQQQEKIKPKVRSTVAQH
HEALVGHGFTHAHIVALSQHPAALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEAL
LTVAGELRGPPLQLDTGQLLKIARGGVTAVEAVHAWRNALTGAPLNLTPDQVVAIASNGGG
KQALETVQRLLPVLCQDHGLTPEQVVAIANNNGGKQALETVQRLLPVLCQAHGLTPDQVVAI
ANNNGGKQALETVQRLLPVLCQAHGLTPAQVVAIASHDGGKQALETVQRLLPVLCQDHGLT
DQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPEQVVAIASHDGGKQALETVQRLLPVLCQ
[0290] AHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPAQVVAIANNNGGKQALETVQRLL
LPVLCQDHGLTPDQVVAIANNNGGKQALETVQRLLPVLCQDHGLTPEQVVAIASHDGGKQAL
ETVQRLLPVLCQAHGLTPDQVVAIANNNGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNIG
GKQALETVQRLLPVLCQDHGLTPDQVVAIANNNGGKQALETVQRLLPVLCQDHGLTPEQVVA
IANNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIANNNGGKQALETVQRLLPVLCQAHGLT
PAQVVAIANNNGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNGGGKQALETVQRLLPVLC
QDHGLTPEQVVAIANNNGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNGGGRPALESIVAQ
LSRPDPAALTNDHLVALACLGGRPALDAVKKGLPHAPALIKRTNRRIPERTSHRVADHAQV
VRVLGFFQCHSHPAQAFDDAMTQFGMSGGSGRADALDDFDLDMLGSDALDDFDLDMLGSD
ALDDFDLDMLGSDALDDFDLDMLHHHHHH (SEQ ID NO:37)

[0291] (-30)GFP-PPAR γ -TALE-2 (-30GFP双下划线;6xHis标签划线):

MGASKGEELFDGVPVILVELDGDVNGHEFSVRGEGEGDATEGELTLKFICTTGELPVPWPTLV
TTTLTYGVQCFSYDPDHMDQHDFFKSAMPEGYVQERTISFKDDGTYKTRAEVKFEGDTLVNRIE
LKGIDFKEDGNILGHKLEYNFNSHDVYITADKQENGIKAEFEIRHNVEDGSVQLADHYQQNTPI
GDGPVLLPDDHYLSTESALSKDPNEDRDHMLLEFVTAAGIDHGMDELYKAPKKKRVGIHR
GVPMVDLRTLGYSSQQQEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSQHPAALGTVAVKY
QDMIAALPEATHEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDTGQLLKIARGGVTA
EAVHAWRNALTGAPLNLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPEQVVAIASNI
GGKQALETVQRLLPVLCQAHGLTPDQVVAIASNGGGKQALETVQRLLPVLCQAHGLTPAQVVA
AIANNNGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNGGGKQALETVQRLLPVLCQDHGL
[0292] TPEQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIASHDGGKQALETVQRLLPVLC
QAHGLTPAQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQR
LLPVLCQDHGLTPEQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIANNNGGKQAL
ETVQRLLPVLCQAHGLTPAQVVAIASNGGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHD
GGKQALETVQRLLPVLCQDHGLTPEQVVAIASNGGGKQALETVQRLLPVLCQAHGLTPDQVVA
IASNGGGKQALETVQRLLPVLCQAHGLTPAQVVAIANNNGGKQALETVQRLLPVLCQDHGL
TPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPEQVVAIASHDGGKQALETVQRLLPVLC
QAHGLTPEQVVAIASNIGGRPALESIVAQLSRPDPAALTNDHLVALACLGGRPALDAVKKG
LPHAPALIKRTNRRIPERTSHRVADHAQVVRVLGFFQCHSHPAQAFDDAMTQFGMSGGSGR
ADALDDFDLDMLGSDALDDFDLDMLGSDALDDFDLDMLGSDALDDFDLDMLHHHHHH
(SEQ ID NO:38)

[0293] (-30)GFP-PRDM16_TALE-3 (-30GFP双下划线;6xHis标签划线):

MGASKGEELFDGVPVILVELDGDVNGHEFSVRGEGEGDATEGELTLKFICTTGELPVPWPTLV
TTTLTYGVQCFSYDPDHMDQHDFFKSAMPEGYVQERTISFKDDGTYKTRAEVKFEGDTLVNRIE
LKGIDFKEDGNILGHKLEYNFNSHDVYITADKQENGIKAEFEIRHNVEDGSVQLADHYQQNTPI
GDGPVLLPDDHYLSTESALSKDPNEDRDHMLLEFVTAAGIDHGMDELYKAPKKKRVGIHR
GVPMVDLRTLGYSSQQQEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSQHPAALGTVAVKY
QDMIAALPEATHEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDTGQLLKIARGGVTA
EAVHAWRNALTGAPLNLTPDQVVAIASNGGGKQALETVQRLLPVLCQDHGLTPEQVVAIANN
[0294] NGGKQALETVQRLLPVLCQAHGLTPDQVVAIANNNGGKQALETVQRLLPVLCQAHGLTPAQV

[0307] 因此,本公开的一些方面提供包含Cas9蛋白(例如如本文描述的;见例如下文Cas9效应器蛋白)和阳离子聚合物的组合物,其能将Cas9蛋白投递至细胞的内部。在一些实施方案中,Cas9蛋白与gRNA联合,后者例如提供阴离子电荷给复合物,从而允许Cas9:gRNA复合物被阳离子聚合物有效包封和/或与其联合。在一些实施方案中,Cas9蛋白不需要与gRNA联合以被阳离子聚合物有效包封和/或与其联合,而是与超负电荷蛋白联合,如本文描述的。在其中Cas9蛋白与超负电荷蛋白联合的一些实施方案中,Cas9蛋白还与gRNA联合。在一些实施方案中,Cas9蛋白是野生型Cas9蛋白,野生型Cas9蛋白的片段,或野生型Cas9蛋白的变体。在一些实施方案中,Cas9蛋白包含dCas9域(例如如本文描述的)。在一些实施方案中,Cas9蛋白是包含dCas9域(例如如本文描述的)的融合蛋白。在一些实施方案中,Cas9蛋白是Cas9切口酶。

[0308] 在其他实施方案中,提供包含效应器蛋白(例如除Cas9蛋白以外的)和阳离子聚合物的组合物,其能将效应器蛋白投递至细胞内部(例如至细胞的核)。效应器蛋白是天然带负电荷的,经修饰以具有总体净负电荷的,或与超负电荷蛋白关联,如本文描述的。在一些实施方案中,效应器蛋白是本文描述的任意效应器蛋白。在一些实施方案中,效应器蛋白是重组酶,例如本文描述的任意重组酶。在一些实施方案中,重组酶是Cre重组酶。在一些实施方案中,所述Cre重组酶包含SEQ ID NO:32的氨基酸序列(例如有或无6xHis标签)。在一些实施方案中,Cre重组酶包含与SEQ ID NO:32的氨基酸序列至少80%,至少85%,至少90%,至少95%,至少98%或至少99%相同的氨基酸序列(例如有或无6xHis标签)。在一些实施方案中,Cre重组酶融合于超电荷蛋白(例如+36GFP或-30GFP)。在一些实施方案中,融合于超电荷蛋白的Cre重组酶包含SEQ ID NO:33(例如有或无6xHis标签)或SEQ ID NO:34(例如有或无6xHis标签)的氨基酸序列,或包含与SEQ ID NO:33或SEQ ID NO:34的氨基酸序列至少80%,至少85%,至少90%,至少95%,至少98%或至少99%相同的氨基酸序列(例如有或无6xHis标签)。在一些实施方案中,效应器蛋白是TALE蛋白(例如如本文描述的包含实施例中提供的那些)。在一些实施方案中,TALE蛋白包含一个或多个VP64转录激活物域(例如SEQ ID NO:35)。在一些实施方案中,具有VP64转录激活物域的TALE蛋白还包含选自下组的氨基酸序列:SEQ ID NO:36-39(例如有或无6xHis标签)。在一些实施方案中,具有VP64转录激活物域的TALE蛋白包含与选自下组的氨基酸序列至少80%,至少85%,至少90%,至少95%,至少98%或至少99%相同的氨基酸序列(例如有或无6xHis标签):SEQ ID NO:36-39。在一些实施方案中,TALE效应器蛋白包含(-30)GFP域(例如SEQ ID NO:21或SEQ ID NO:40),TALE域的N末端区(例如SEQ ID NO:41),可变重复域(例如18.5聚体重复域,如Maeder等,“Robust, synergistic regulation of human gene expression using TALE activators.”*Nat. Methods*.2013;10,243-245中提供的),C末端TALE域(例如SEQ ID NO:42),VP64激活域(例如SEQ ID NO:35),和任选地在任何域之间的一个或多个接头(例如GGG(9),SEQ ID NO:252)以及任选地序列标签(例如6xHis,SEQ ID NO:253)。

[0309] 在一些实施方案中,本文描述的包含Cas9蛋白或其他效应器蛋白的组合物在施用给(例如体外或体内)细胞群体时展现出低毒性。在一些实施方案中,群体中至少60%,至少65%,至少70%,至少75%,至少80%,至少85%,至少90%,至少95%或至少99%的细胞在施用本发明组合物后是活的,所述组合物包含Cas9蛋白或其他效应器蛋白和阳离子聚合物。用于评估当施用给细胞群体时组合物毒性的方法是本领域公知的且包括实施例中描述

的那些。

[0310] Cas9效应器蛋白

[0311] 在一些实施方案中,通过本文提供的系统或方法将包含可RNA编程的蛋白(或其片段或变体)的效应器蛋白投递给靶细胞。在一些实施方案中,通过本文提供的系统或方法将RNA导引的或可RNA编程的核酸酶投递给靶细胞。在一些实施方案中,所述可RNA编程的蛋白是Cas9核酸酶、Cas9变体、或Cas9蛋白的融合物,其通过本文提供的系统或方法投递给靶细胞。

[0312] 在一些实施方案中,所述可RNA编程的核酸酶是(CRISPR-associated system, CRISPR关联系统) Cas9内切核酸酶,例如来自酿脓链球菌的Cas9 (Csn1) (参见,例如“Complete genome sequence of an M1 strain of *Streptococcus pyogenes*.” Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K., Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P., Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L. expand/collapse author list McLaughlin R.E., Proc.Natl.Acad.Sci.U.S.A.98:4658-4663(2001); “CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III.” Deltcheva E., Chylinski K., Sharma C.M., Gonzales K., Chao Y., Pirzada Z.A., Eckert M.R., Vogel J., Charpentier E., Nature 471:602-607(2011); 和“A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity.” Jinek M., Chylinski K., Fonfara I., Hauer M., Doudna J.A., Charpentier E. Science 337:816-821(2012), 其各自完整内容通过提述并入本文。由于可RNA编程的核酸酶(例如Cas9)利用RNA:DNA杂交来确定靶DNA切割位点,因此这些蛋白质在原理上能切割由导引RNA指定的任意序列。使用可RNA编程的核酸酶如Cas9用于位点特异性切割(例如以修饰基因组)的方法是本领域中公知的(参见例如Cong, L.等 Multiplex genome engineering using CRISPR/Cas systems. Science 339, 819-823(2013); Mali, P.等 RNA-guided human genome engineering via Cas9. Science 339, 823-826(2013); Hwang, W.Y.等 Efficient genome editing in zebrafish using a CRISPR-Cas system. Nature biotechnology 31, 227-229(2013); Jinek, M.等 RNA-programmed genome editing in human cells. eLife 2, e00471(2013); Dicarlo, J.E.等 Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. Nucleic acids research (2013); Jiang, W.等 RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Nature biotechnology 31, 233-239(2013); 其各自完整内容通过提述并入本文)。

[0313] Cas9核酸酶有时还称为casn1核酸酶或CRISPR(成簇的规则间隔的短回文重复(clustered regularly interspaced short palindromic repeat))有关的核酸酶。CRISPR是提供对移动性遗传元件(例如病毒、转座元件和接合质粒)的保护的适应性免疫系统。CRISPR簇含有间隔物、与前期移动元件互补的序列和靶物侵入性核酸。CRISPR簇转录和加工成CRISPR RNA (crRNA)。在II型CRISPR系统中,对前crRNA的正确加工需要反式编码的小RNA(tracrRNA)、内源核糖核酸酶3(rnc)和Cas9蛋白。tracrRNA充当核糖核酸酶3辅助的对前crRNA加工的导引。随后,Cas9/crRNA/tracrRNA以内切核苷酸方式切割与间隔物互补的线性或环状dsDNA靶物。不与crRNA互补的靶链首先被内切核苷酸方式切割,然后以外切核苷酸方式在3'-5'修剪。在自然界中,DNA结合和切割通常需要蛋白质和两种RNA种类。然

而,可以工程化单导引RNA(“sgRNA”或简称为“gNRA”)以将crRNA和tracrRNA两者的示象(aspect)均纳入到单RNA种类中。参见例如,Jinek M.,Chylinski K.,Fonfara I.,Hauer M.,Doudna J.A.,Charpentier E.Science 337:816-821(2012),其完整内容通过提述据此并入本文。Cas9识别CRISPR重复序列中的短基序(PAM或原间隔区临近基序)来帮助区分自身和非自身。Cas9核酸酶序列和结构是本领域技术人员公知的(参见,例如“Complete genome sequence of an M1 strain of *Streptococcus pyogenes*.”Ferretti J.J.,McShan W.M.,Ajdic D.J.,Savic D.J.,Savic G.,Lyon K.,Primeaux C.,Sezate S.,Suvorov A.N.,Kenton S.,Lai H.S.,Lin S.P.,Qian Y.,Jia H.G.,Najar F.Z.,Ren Q.,Zhu H.,Song L.expand/collapse author list McLaughlin R.E.,Proc.Natl.Acad.Sci.U.S.A.98:4658-4663(2001);“CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III.”Deltcheva E.,Chylinski K.,Sharma C.M.,Gonzales K.,Chao Y.,Pirzada Z.A.,Eckert M.R.,Vogel J.,Charpentier E.,Nature 471:602-607(2011);和“Aprogrammable dual-RNA导引的DNA endonuclease in adaptive bacterial immunity.”Jinek M.,Chylinski K.,Fonfara I.,Hauer M.,Doudna J.A.,Charpentier E.Science 337:816-821(2012),其各自完整内容通过提述并入本文)。

[0314] 已在多种物种中描述了Cas9直系同源物,包括但不限于酿脓链球菌(*S. pyogenes*)和嗜热链球菌(*S. thermophilus*)。基于本公开,其他适宜的Cas9核酸酶和序列对于本领域技术人员将是明显的,而且这类Cas9核酸酶和序列包括来自Chylinski,Rhun和Charpentier,“The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems”(2013)RNA Biology 10:5,726-737(其完整内容通过提述并入本文)中公开的生物和基因座的Cas9序列。在一些实施方案中,包含Cas9或其片段的蛋白质称为“Cas9变体”。Cas9变体与Cas9或其片段共享同源性。例如,Cas9变体与野生型Cas9至少约70%相同,至少约80%相同,至少约90%相同,至少约95%相同,至少约98%相同,至少约99%相同,至少约99.5%相同,或至少约99.9%相同。在一些实施方案中,Cas9变体包含Cas9的片段(例如gRNA结合域或DNA切割域、N末端域或C末端域,等等),从而该片段与野生型Cas9的相应片段至少约70%相同,至少约80%相同,至少约90%相同,至少约95%相同,至少约98%相同,至少约99%相同,至少约99.5%相同,或至少约99.9%相同。在一些实施方案中,野生型Cas9对应于来自酿脓链球菌(*Streptococcus pyogenes*)的Cas9(NCBI参考序列:NC_017053.1,SEQ ID NO:43(核苷酸);SEQ ID NO:44(氨基酸))。在一些实施方案中,Cas9蛋白具有无活性的(例如失活的)DNA切割域。核酸酶失活的Cas9蛋白可交换地称为“dCas9”蛋白(对于核酸酶“死亡的”Cas9)。在一些实施方案中,dCas9对应于或包含,部分或全部地,如下文SEQ ID NO:45中所列的氨基酸。在一些实施方案中,提供dCas9的变体(例如SEQ ID NO:45的变体)。例如,在一些实施方案中,提供具有D10A和H840A以外的突变的变体,其导致核酸酶失活的Cas9(dCas9)举例而言,这类突变包括在D10和H840处的其他氨基酸取代,或Cas9核酸酶域内的其他取代(例如HNH核酸酶子域和/或RuvC1子域中的取代)。在一些实施方案中,提供dCas9的变体或同源物(例如SEQ ID NO:45的变体),其与SEQ ID NO:45至少约70%相同,至少约80%相同,至少约90%相同,至少约95%相同,至少约98%相同,至少约99%相同,至少约99.5%相同或至少约99.9%相同。在一些实施方案中,提供dCas9的变体(例如SEQ ID NO:45的变体),其具有比SEQ ID NO:45短或长约5个氨基酸、约10个氨基酸、约15个氨基酸、

约20个氨基酸、约25个氨基酸、约30个氨基酸、约40个氨基酸、约50个氨基酸、约75个氨基酸、约100个氨基酸或更多的氨基酸序列。在一些实施方案中,提供Cas9“切口酶”,其包含使Cas9中单个核酸酶域失活的突变。这类切口酶诱导靶核酸中的单链断裂,如与双链断裂相对的。

[0315] Cas9

[0316] ATGGATAAGAAATACTCAATAGGCTTAGATATCGGCACAAATAGCGTCGGATGGGCGGTGATCACT
[0317] GATGATTATAAGGTTCCGTCTAAAAAGTTCAAGTTCTGGGAAATACAGACCGCCACAGTATCAAA
[0318] AAAAATCTTATAGGGGCTCTTTTATTTGGCAGTGGAGAGACAGCGGAAGCGACTCGTCTCAAACGG
[0319] ACAGCTCGTAGAAGGTATACACGTCGGAAGAATCGTATTTGTTATCTACAGGAGATTTTTTCAAATG
[0320] AGATGGCGAAAGTAGATGATAGTTTCTTTTCATCGACTTGAAGAGTCTTTTTTGGTGGAGAAGACAA
[0321] GAAGCATGAACGTCATCCTATTTTTGGAAATATAGTAGATGAAGTTGCTTATCATGAGAAAATATCCA
[0322] ACTATCTATCATCTGCGAAAAAATTGGCAGATTCTACTGATAAAGCGGATTTGCGCTTAATCTATTT
[0323] GGCCTTAGCGCATATGATTAAGTTTTCGTGGTCATTTTTTTGATTGAGGGAGATTTAAATCCTGATAATA
[0324] GTGATGTGGACAACTATTTATCCAGTTGGTACAAATCTACAATCAATTATTTGAAGAAAAACCTAT
[0325] TAACGCAAGTAGAGTAGATGCTAAAGCGATTCTTTCTGCACGATTGAGTAAATCAAGACGATTAGA
[0326] AAATCTCATTGCTCAGCTCCCCGGTGAGAAGAGAAATGGCTTGTTTGGGAATCTCATTGCTTTGTCAT
[0327] TGGGATTGACCCCTAATTTTTAAATCAAATTTTGATTTGGCAGAAGATGCTAAATTACAGCTTTCAAA
[0328] AGATACTTACGATGATGATTTAGATAATTTATTGGCGCAAATTGGAGATCAATATGCTGATTTGTTTT
[0329] TGGCAGCTAAGAATTTATCAGATGCTATTTTACTTTTACAGATATCCTAAGAGTAAATAGTGAAAATAAC
[0330] TAAGGCTCCCCTATCAGCTTCAATGATTAAGCGCTACGATGAACATCATCAAGACTTGACTCTTTTA
[0331] AAAGCTTTAGTTCGACAACAACCTCCAGAAAAGTATAAAAGAAATCTTTTTTGATCAATCAAAAAACG
[0332] GATATGCAGGTTATATTGATGGGGGAGCTAGCCAAGAAGAATTTTATAAATTTATCAAACCAATTTT
[0333] AGAAAAAATGGATGGTACTGAGGAATTATTGGTAAAATAAATCGTGAAGATTTGCTGCGCAAGCA
[0334] ACGGACCTTTGACAACGGCTCTATTCCCATCAAATTCACCTGGGTGAGCTGCATGCTATTTTGAGA
[0335] AGACAAGAAGACTTTTATCCATTTTTAAAAGACAATCGTGAGAAGATTGAAAAAATCTTGACTTTTC
[0336] GAATTCCTTATTATGTTGGTCCATTGGCGCGTGGCAATAGTCGTTTTGCATGGATGACTCGGAAGTCT
[0337] GAAGAAACAATTACCCCATGGAATTTTGAAGAAGTTGTCGATAAAGGTGCTTCAGCTCAATCATTTA
[0338] TTGAACGCATGACAACTTTGATAAAAAATCTTCCAAATGAAAAAGTACTACAAAAACATAGTTTGCT
[0339] TTATGAGTATTTTACGGTTTATAACGAATTGACAAAGGTCAAATATGTTACTGAGGGAATGCGAAAA
[0340] CCAGCATTTCTTTTCCAGGTGAACAGAAGAAAGCCATTGTTGATTTACTCTTCAAAAACAAATCGAAAA
[0341] TAACCGTTAAGCAATTAAGAAGATTATTTCAAAAAATAGAATGTTTTGATAGTGTGAAATTTTC
[0342] AGGAGTTGAAGATAGATTTAATGCTTCATTAGGCGCTACCATGATTTGCTAAAAATTTAATAAGAT
[0343] AAAGATTTTTTGGATAATGAAGAAAATGAAGATATCTTAGAGGATATTGTTTTAACATTGACCTTAT
[0344] TTGAAGATAGGGGGATGATTGAGGAAAGACTTAAAACATATGCTCACCTCTTTGATGATAAGGTGAT
[0345] GAAACAGCTTAAACGTCGCCGTTATACTGGTTGGGGACGTTTGTCTCGAAAAATTGATTAATGGTATT
[0346] AGGGATAAGCAATCTGGCAAAAACAATATTAGATTTTTTGAATCAGATGGTTTTGCCAATCGCAATT
[0347] TTATGCAGCTGATCCATGATGATAGTTTGACATTTAAAGAAGATATTCAAAAAGCACAGGTGTCTGG
[0348] ACAAGGCCATAGTTTACATGAACAGATTGCTAACTTAGCTGGCAGTCCTGCTATTAAAAAAGGTATT
[0349] TTACAGACTGTAAAAATTGTTGATGAACTGGTCAAAGTAATGGGGCATAAGCCAGAAAAATATCGTT

[0350] ATTGAAATGGCACGTGAAAATCAGACAACTCAAAAGGGCCAGAAAAATTCGCGAGAGCGTATGAA
 [0351] ACGAATCGAAGAAGGTATCAAAGAATTAGGAAGTCAGATTCTTAAAGAGCATCCTGTTGAAAATAC
 [0352] TCAATTGCAAAATGAAAAGCTCTATCTCTATTATCTACAAAATGGAAGAGACATGTATGTGGACCAA
 [0353] GAATTAGATATTAATCGTTTAAAGTGATTATGATGTCGATCACATTGTTCCACAAAAGTTTCATTAAGA
 [0354] CGATTCAATAGACAATAAGGTACTAACGCGTTCTGATAAAAAATCGTGGTAAATCGGATAACGTTCCA
 [0355] AGTGAAGAAGTAGTCAAAAAGATGAAAACTATTGGAGACAACTTCTAAACGCCAAGTTAATCACT
 [0356] CAACGTAAGTTTGATAATTTAACGAAAGCTGAACGTGGAGGTTTGAGTGAACCTTGATAAAGCTGGTT
 [0357] TTATCAAACGCCAATTGGTTGAAACTCGCCAAATCACTAAGCATGTGGCACAAAATTTGGATAGTCG
 [0358] CATGAATACTAAATACGATGAAAATGATAAATTTATTCGAGAGGTTAAAGTGATTACCTTAAATCT
 [0359] AAATTAGTTTCTGACTTCCGAAAAGATTTCCAATTCTATAAAGTACGTGAGATTAACAATTACCATC
 [0360] ATGCCCATGATGCGTATCTAAATGCCGTCGTTGGAAGTCTTTGATTAAGAAAATATCCAAAATCTGA
 [0361] ATCGGAGTTTGTCTATGGTGATTATAAAGTTTATGATGTTTCGTA AAAATGATTGCTAAGTCTGAGCAA
 [0362] GAAATAGGCAAAGCAACCGCAAATATTTCTTTTACTCTAATATCATGAACTTCTTCAAAAACAGAAA
 [0363] TTACACTTGCAAATGGAGAGATTCGCAAACGCCCTCTAATCGAAACTAATGGGGAAAATGGAGAAA
 [0364] TTGTCTGGGATAAAGGGCGAGATTTTGCCACAGTGCGCAAAGTATTGTCCATGCCCAAGTCAATAT
 [0365] TGTCAAGAAAACAGAAGTACAGACAGGCGGATTCTCCAAGGAGTCAATTTTACCAAAAAGAAAATTC
 [0366] GGACAAGCTTATTGCTCGTAAAAAAGACTGGGATCCAAAAAATATGGTGGTTTTGATAGTCCAAC
 [0367] GGTAGCTTATTCAGTCTAGTGGTTGCTAAGGTGGAAAAAGGGAAAATCGAAGAAGTTAAAATCCGT
 [0368] TAAAGAGTTACTAGGGATCACAATTATGGAAAGAAGTTCCTTTGAAAAAATCCGATTGACTTTTTA
 [0369] GAAGCTAAAGGATATAAGGAAGTTAAAAAAGACTTAATCATTAAACTACCTAAATATAGTCTTTTTG
 [0370] AGTTAGAAAACGGTTCGTA AACGGATGCTGGCTAGTGCCGGAGAATTACAAAAAGGAAATGAGCTGG
 [0371] CTCTGCCAAGCAAATATGTGAATTTTTTATATTTAGCTAGTCATTATGAAAAGTTGAAGGGTAGTCC
 [0372] AGAAGATAACGAACAAAAACAATTGTTTGTGGAGCAGCATAAGCATTATTTAGATGAGATTATTGA
 [0373] GCAAATCAGTGAATTTTCTAAGCGTGTTATTTTAGCAGATGCCAATTTAGATAAAGTTCTTAGTGCAT
 [0374] ATAACAAACATAGAGACAAACCAATACGTGAACAAGCAGAAAATATTATTCATTTATTTACGTTGAC
 [0375] GAATCTTGGAGCTCCCGCTGCTTTTAAATATTTTGATACAACAATTGATCGTAAACGATATACGTCTA
 [0376] CAAAAGAAGTTTTAGATGCCACTCTTATCCATCAATCCATCACTGGTCTTTATGAAACACGCATTGAT
 TTGAGTCAGCTAGGAGGTGACTGA (SEQ ID NO:43)

[0377] MDKKYSIGLAIGTNSVGWAVITDEYKVPSSKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARR
 RYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKK
 LVDSTDKADRLIYLALAHMIKFRGHFLIEGDLNPDNSVDKLFQLVQTYNQLFEENPINASGVDAKAIL
 SARLSKSRRLLENLIAQLPGEKKNLFGNLIASLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIG
 DQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQ
 SKNGYAGYIDGASQEEFYKFIKPILEKMDGTEELLVKNREDLLRKQRTFDNGSIPHQIHLGELHAILRR
 QEDFYPLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWFEEVVDKKGASAQSFIERMT
 NFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLK
 EDYFKKIECFDSVEISGVEDRFNASLGTYHLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKT
 YAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMLIHDDSLTFKED
 IQKAQVSGQGDSLHEHIANLAGSPAIKKGIQTQVQVDELVKVMGRHKPENIVIEMARENQTTQKGQKN
 SRERMKRIEIKELGSQILKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYDVDHIVPQSF

[0378] LKDDSIDNKVLRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKA
GFIKRQLVETROITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKREINNYHHA
HDAYLNAVVGTAALIKKYPKLESEFVYGDYKVDYVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLA
NGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKK
 DWDPKKGFFDSPTVAYSVLVVAKVEKGKSKLKSVEKLLGITIMERSSEFKNPIDFLEAKGYKEVKKD
 LIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLGSPEDNEQKQLFVEQHK
 HYLDEIIEQISEFSKRVILADANLDKVL SAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKR
 YTSTKEVL DATLIHQ SITGLYETRIDLSQLGGD (SEQ ID NO:44)

[0379] (单下划线:HNH域;双下划线:RuvC域)

[0380] dCas9 (D10A和H840A) :

MDKKYSIGLAIGTNSVGVAVITDEYKVPKSKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARR
 RYTRRKNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKK
 LVDSTDKADRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAIL
 SARLSKSRLENLIAQLPGEKKNLFGNLIALSGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIG
 DQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQ
 SKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRR
 QEDFYFPLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMT
 NFDKNLPNEKVLPHKSLLEYEFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLK
 EDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKT
 YAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDFANRNFMLIHDDSLTFKED
 IQKAQVSGQDLSLHEHIANLAGSPAIKKGILOTVKVVDLVKVMGRHKPENIVEMARENQTTQKQKN
 SRERMKRIEIEGKELGSQILKEHPVENTQLQNEKLYLYLQNGRDMYVDOELDINRLSDYDVAIVPOSF
LKDDSIDNKVLRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKA
GFIKRQLVETROITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKREINNYHHA
HDAYLNAVVGTAALIKKYPKLESEFVYGDYKVDYVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLA
NGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKK
 DWDPKKGFFDSPTVAYSVLVVAKVEKGKSKLKSVEKLLGITIMERSSEFKNPIDFLEAKGYKEVKKD
 LIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLGSPEDNEQKQLFVEQHK
 HYLDEIIEQISEFSKRVILADANLDKVL SAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKR
 YTSTKEVL DATLIHQ SITGLYETRIDLSQLGGD (SEQ ID NO:45)

[0382] (单下划线:HNH域;双下划线:RuvC域)

[0383] Cas9切口酶 (D10A) :

[0384] MDKKYSIGLAIGTNSVGVAVITDEYKVPKSKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
 R

[0385] RYTRRKNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLR
 KK

[0386] LVDSTDKADRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKA
 IL

[0387] SARLSKSRLENLIAQLPGEKKNLFGNLIALSGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQ
 IG

[0388] DQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFF
 DQ

[0389] SKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAIL
 RR

[0390] QEDFYFPLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIER
 MT

[0391] NFDKNLPNEKVLPHKSLLEYEFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLK

[0392] EDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEE
 RLKT

[0393] YAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDFANRNFMLIHDDSLTFKED
 [0394] IQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVEMARENQTTQKGGQKN
 [0395] SRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYDVDHIVPQS
 F
 [0396] LKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLELSDKA
 [0397] GFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKREINNYHH
 A
 [0398] HDAYLNAVVGITALIKKYPKLESEFVYGDYKVYDVRKMKIAKSEQEIGKATAKYFFYSNIMNFFKTEITLA
 [0399] NGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNI VKKTEVQTGGFSKESILPKRNSDKLIAR
 KK
 [0400] DWDPKKGFFSPTVAYSVLVVAKEVKGKSKKLKSVKELLGITIMERSSEKPNIDFLEAKGYKEVKKD
 [0401] LIIKLPKYSLENGRKRMLASAGELQKGNELALPSKYVNFYLYLASHYEKLGSPEDNEQKQLFVEQH
 K
 [0402] HYLDEIIEQISEFSKRVI LADANLDKVL SAYNKHRDKPIREQAENI IHLFTLTNLGAPAAFKYFDTTI
 DRKRYTSTKEVLDTLIHQSI TGLYETRIDLSQLGGD (SEQ ID NO:46)

[0403] 在一些实施方案中,提供包含Cas9蛋白的融合蛋白用于本文描述的任意组合物和方法中。在一些实施方案中,所述融合蛋白包含dCas9蛋白(例如如本文描述的)。在一些实施方案中,所述融合蛋白包含dCas9和一个或多个域(例如酶域)之间的接头(例如如本文描述的)。在一些实施方案中,所述融合蛋白包含dCas9和转录激活物域、转录阻遏物域、重组酶域、基因编辑域(例如脱氨酶域)、或表观遗传修饰物域。

[0404] 在一些实施方案中,本文中提供的例示性融合蛋白的一般体系结构包含以下结构:

[0405] [NH₂] - [酶域] - [dCas9] - [COOH] 或

[0406] [NH₂] - [dCas9] - [酶域] - [COOH];

[0407] 其中NH₂是融合蛋白的N末端,COOH是融合蛋白的C末端,且所述酶域包含核酸酶域(例如FokI)、重组酶催化域(例如Hin, Gin或Tn3重组酶干)、核酸编辑域(例如脱氨酶域)、转录激活物域(例如VP64、p65)、转录阻遏物域(例如KRAB、SID)、或表观遗传修饰物(例如LSD1组蛋白去甲基化酶、TET1羟甲基化酶)。

[0408] 可以存在另外的特征,例如某些域之间的一个或多个接头序列。其他可能存在的例示性特征是定位序列,如核定位序列(NLS;例如MAPKKKRVGIHRGVP (SEQ ID NO:47));细胞质定位序列;输出序列,如核输出序列;或其他定位序列,以及可用于增溶、纯化或检测融合蛋白的序列标签。本文中提供了适宜的信号序列和蛋白标签序列且是本领域已知的,包括但不限于,生物素羧化酶载体蛋白(BCCP)标签、myc标签、钙调蛋白标签、FLAG标签(例如3xFLAG标签:MDYKDHDGDYKDHDIDYKDDDDK (SEQ ID NO:48))、血凝素(HA)标签、多组氨酸标签(也称为组氨酸标签或His标签)、麦芽糖结合蛋白(MBP)标签、nus标签、谷胱甘肽-S-转移酶(GST)标签、绿色荧光蛋白(GFP)标签、硫氧还蛋白标签、S标签、Softag(例如Softag 1、Softag 3)、strep标签、生物素连接酶标签、FlAsH标签、V5标签和SBP标签。其他适宜序列对于本领域技术人员将是明显的。

[0409] 在一些实施方案中,所述酶域包含核酸酶或其催化域。例如,在一些实施方案中,

例示性的具有核酸酶域的配体依赖性dCas9融合蛋白的一般体系结构包含以下结构：

[0410] $[\text{NH}_2] - [\text{NLS}] - [\text{dCas9}] - [\text{核酸酶}] - [\text{COOH}]$,

[0411] $[\text{NH}_2] - [\text{NLS}] - [\text{核酸酶}] - [\text{dCas9}] - [\text{COOH}]$,

[0412] $[\text{NH}_2] - [\text{dCas9}] - [\text{核酸酶}] - [\text{COOH}]$, 或

[0413] $[\text{NH}_2] - [\text{核酸酶}] - [\text{dCas9}] - [\text{COOH}]$;

[0414] 其中NLS是核定位信号, NH_2 是融合蛋白的N末端, 且COOH是融合蛋白的C末端。在一些实施方案中, 接头插入在dCas9和核酸酶域之间。在一些实施方案中, 接头插入在NLS与核酸酶和/或dCas9域之间。在一些实施方案中, NLS位于核酸酶和/或dCas9域的C末端。在一些实施方案中, NLS位于核酸酶和dCas9域之间。还可以存在另外的特征, 如序列标签。在一些方面, 所述核酸酶域是需要二聚化以切割靶核酸(例如DNA)的核酸酶(例如两个核酸酶的单体在一起)。在一些实施方案中, 核酸酶域是FokIDNA切割域的单体。FokIDNA切割域是已知的, 且在一些方面对应于FokI (NCBI登录号J04623)的氨基酸388-583。在一些实施方案中, FokIDNA切割域对应于FokI的氨基酸300-583, 320-583, 340-583或360-583。亦参见Wah等, “Structure of FokI has implications for DNA cleavage” Proc. Natl. Acad. Sci. USA. 1998; 1; 95 (18) : 10564-9; Li等, “TAL nucleases (TALNs) : hybrid proteins composed of TAL effectors and FokIDNA-cleavage domain” Nucleic Acids Res. 2011; 39 (1) : 359-72; Kim等, “Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain” Proc. Natl. Acad. Sci. USA. 1996; 93: 1156-1160; 其各自完整内容通过提述并入本文)。在一些实施方案中, 所述FokIDNA切割域对应于或包含, 部分或全部地, 如SEQ ID NO: 49中所列的氨基酸序列。在一些实施方案中, 所述FokIDNA切割域是FokI的变体(例如SEQ ID NO: 49的变体), 如本文描述的。使用dCas9-核酸酶融合蛋白的其他例示性组合物和方法可见于2014年6月30日提交的美国专利申请U.S.S.N 14/320,498; 名称为“Cas9-FokI fusion Proteins and Uses Thereof”, 其完整内容通过提述并入本文。

[0415] FokI核酸酶域:

[0416] GSQLVKSELEEKKSELRHKLKYPHEYIELIEIARNSTQDRILEMKVMEFFMKVYGYRGKHLGGSRKPDGAIYTVGSPIDYGVIVDTKAYSGGYNLPIGQADEMQRVYVEENQTRNKHINPNEWKVPSSVTEFKFLVSGHFKGNYKAQLTRLNHNITCNGAVLSVEELLIGGEMIKAGTLTLEEVRKFNNGEINF (SEQ ID NO: 49)

[0417] fCas9(例如dCas9-NLS-GGS3接头-FokI):

[0418] ATGGATAAAAAGTATTCTATTGGTTTAGCTATCGGCACTAATTCCGTTGGATGGGCTGTCATAACCGA TGAATACAAAGTACCTTCAAAGAAATTTAAGGTGTTGGGGAACACAGACCGTCATTCGATTA AAAAGAATCTTATC GGTGCCCTCCTATTTCGATAGTGGCGAAACGGCAGAGGCGACTCGCCTGAAACGAACCGCTCGGAGAAGGTATACAC GTCGCAAGAACCGAATATGTTACTTACAAGAAATTTTTAGCAATGAGATGGCCAAAGTTGACGATTCTTTCTTTCA CCGTTTGAAGAGTCCTTCTTGTGCAAGAGGACAAGAAACATGAACGGCACCCCATCTTTGGAAACATAGTAGAT GAGGTGGCATATCATGAAAAGTACCCAACGATTTATCACCTCAGAAAAAGCTAGTTGACTCAACTGATAAAGCGG ACCTGAGGTTAATCTACTTGGCTCTTGGCCATATGATAAAGTTCCGTGGGCACTTTCTCATTGAGGGTGATCTAAA TCCGACAACCTCGGATGTCGACAACTGTTTCATCCAGTTAGTACAAACCTATAATCAGTTGTTTGAAGAGAACCCT ATAAATGCAAGTGGCGTGGATGCGAAGGCTATTCTTAGCGCCCGCTCTCTAAATCCCGACGGCTAGAAAACCTGA TCGCACAATTACCCGGAGAGAAGAAAAATGGGTTGTTTCGGTAACCTTATAGCGCTCTCACTAGGCCTGACACCAA

TTTAAAGTCGAACTTCGACTTAGCTGAAGATGCCAAATTGCAGCTTAGTAAGGACACGTACGATGACGATCTCGAC
AATCTACTGGCACAATTTGGAGATCAGTATGCGGACTTATTTTTGGCTGCCAAAAACCTTAGCGATGCAATCCTCC
TATCTGACATACTGAGAGTTAATACTGAGATTACCAAGGCGCCGTTATCCGCTTCAATGATCAAAAGGTACGATGA
ACATCACCAAGACTTGACACTTCTCAAGGCCCTAGTCCGTGAGCAACTGCCTGAGAAAATAAAGGAAATATTCTTT
GATCAGTCGAAAAACGGGTACGCAGGTTATATTGACGGCGGAGCGAGTCAAGAGGAATCTACAAGTTTATCAAAC
CCATATTAGAGAAGATGGATGGGACGGAAGAGTTGCTTGTA AAACTCAATCGCGAAGATCTACTGCGAAAGCAGCG
GACTTTTCGACAACGGTAGCATTCCACATCAAATCCACTTAGGCGAATTGCATGCTATACTTAGAAGGCAGGAGGAT
TTTTATCCGTTCTCAAAGACAATCGTGAAAAGATTGAGAAAATCCTAACCTTTCGCATACCTTACTATGTGGGAC
CCCTGGCCCCGAGGAACTCTCGTTTCGCATGGATGACAAGAAAGTCCGAAGAAACGATTACTCCATGGAATTTTGA
GGAAGTTGTCGATAAAGGTGCGTCAGCTCAATCGTTCATCGAGAGGATGACCAACTTTGACAAGAATTTACCGAAC
GAAAAAGTATTGCCTAAGCACAGTTTACTTTACGAGTATTTTACAGTGTACAATGAACTCACGAAAGTTAAGTATG
TCACTGAGGGCATGCGTAAACCCGCCTTTCTAAGCGGAGAACAGAAGAAAGCAATAGTAGATCTGTTATTCAAGAC
CAACCGCAAAGTGACAGTTAAGCAATTGAAAGAGGACTACTTTAAGAAAATGAAATGCTTCGATTCTGTGAGATC
TCCGGGGTAGAAGATCGATTTAATGCGTCACTTGGTACGTATCATGACCTCCTAAAAGATAATTAAGATAAGGACT
TCCTGGATAACGAAGAGAATGAAGATATCTTAGAAGATATAGTGTGACTCTTACCCTCTTTGAAGATCGGGAAAT
GATTGAGGAAAGACTAAAAACATACGCTCACCTGTTGACGATAAGGTTATGAAACAGTTAAAGAGGCGTCGCTAT
ACGGGCTGGGGACGATTGTCGCGGAACTTATCAACGGGATAAGAGACAAGCAAAGTGGTAAAACCTATTCTCGATT
TTCTAAAGAGCGACGGCTTCGCCAATAGGAACTTTATGCAGCTGATCCATGATGACTCTTTAACCTTCAAAGAGGA
TATAAAAAGGCACAGGTTTCCGGACAAGGGGACTCATTGCACGAACATATTGCGAATCTTGCTGGTTCCGCAGCC
ATCAAAAAGGCATACTCCAGACAGTCAAAGTAGTGGATGAGCTAGTTAAGGTCATGGGACGTCACAAACCGGAAA
ACATTGTAATCGAGATGGCACGCGAAAATCAAACGACTCAGAAGGGGCAAAAAACAGTCGAGAGCGGATGAAGAG
AATAGAAGAGGGTATTAAAGAACTGGGCAGCCAGATCTTAAAGGAGCATCCTGTGGAAAATACCCAATTGCAGAAC
GAGAACTTTACCTCTATTACCTACAAAATGGAAGGGACATGTATGTTGATCAGGAACTGGACATAAACCGTTTAT
CTGATTACGACGTCGATGCCATTGTACCCCAATCCTTTTTGAAGGACGATTCAATCGACAATAAAGTGCTTACACG
CTCGGATAAGAACCGAGGGAAAAGTGACAATGTTCCAAGCGAGGAAAGTCGTAAAAGAAAATGAAGAACTATTGGCGG
CAGCTCCTAAATGCGAACTGATAACGCAAAGAAAGTTCGATAACTTAACTAAAGCTGAGAGGGGTGGCTTGTCTG
AACTTGACAAGGCCGGATTTATTAACGTCAGCTCGTGGAAACCCGCCAAATCACAAAGCATGTTGCACAGATACT
AGATTCCCGAATGAATACGAAATACGACGAGAACGATAAGCTGATTCGGGAAGTCAAAGTAATCACTTTAAAGTCA
AAATTGGTGTCGACTTCAGAAAGGATTTTCAATTCTATAAAGTTAGGGAGATAAATAACTACCACCATGCGCACG
ACGCTTATCTTAATGCCGTCGTAGGGACCGCACTCATTAAAGAAATACCCGAAGCTAGAAAAGTGAGTTTGTGTATGG
TGATTACAAAGTTTATGACGTCCGTAAGATGATCGCGAAAAGCGAACAGGAGATAGGCAAGGCTACAGCCAAATAC
TTCTTTTATTCTAACATTATGAATTTCTTTAAGACGAAAATCACTCTGGCAAACGGAGAGATACGCAAACGACCTT
TAATTGAAACCAATGGGGAGACAGGTGAAATCGTATGGGATAAGGGCCGGGACTTCGCGACGGTGAGAAAAGTTTT
GTCCATGCCCAAGTCAACATAGTAAAGAAAACCTGAGGTGCAGACCCGAGGGTTTTCAAAGGAATCGATTCTTCCA
AAAAGGAATAGTGATAAGCTCATCGCTCGTAAAAGGACTGGGACCCGAAAAAGTACGGTGGCTTCGATAGCCCTA
CAGTTGCCTATTCTGTCTAGTAGTGGCAAAAGTTGAGAAGGGAAAATCCAAGAACTGAAGTCAGTCAAAGAATT
ATTGGGATAACGATTATGGAGCGCTCGTCTTTTGAAGAACCACATCGACTTCCTTGAGGCGAAAGGTTACAAG
GAAGTAAAAAGGATCTCATAATTAACCTACCAAAGTATAGTCTGTTTGAAGTTAGAAAATGGCCGAAAACGGATGT
TGGCTAGCGCCGAGAGCTTCAAAGGGGAACGAACTCGCACTACCGTCTAAATACGTGAATTTCTGTATTTAGC

GTCCCATTACGAGAAGTTGAAAGGTTACCTGAAGATAACGAACAGAAGCAACTTTTTGTTGAGCAGCACAAACAT
TATCTCGACGAAATCATAGAGCAAATTCGGAATTCAGTAAGAGAGTCATCCTAGCTGATGCCAATCTGGACAAAG
TATTAAGCGCATAACAAGCACAGGGATAAACCCATACGTGAGCAGGCGGAAAATATTATCCATTTGTTTACTCT
TACCAACCTCGGCGCTCCAGCCGCATTCAAGTATTTTGACACAACGATAGATCGCAAACGATACACTTCTACCAAG
GAGGTGCTAGACGCGACTGATTCACCAATCCATCACGGGATTATATGAAACTCGGATAGATTTGTCACAGCTTG
GGGGTGACGGATCCCCAAGAAGAAGAGGAAAGTCTCGAGCGACTACAAAGACCATGACGGTGATTATAAAGATCA
TGACATCGATTACAAGGATGACGATGACAAGGCTGCAGGATCAGGTGGAAGTGGCGGCAGCGGAGTTCTGGATCC
CAACTAGTCAAAAAGTGAAGTGGAGGAGAAGAAATCTGAACTTCGTCATAAAATGAAATATGTGCCTCATGAATATA
TTGAATTAATTGAAATGCGCAGAAATCCACTCAGGATAGAATTCCTGAAATGAAGGTAATGGAATTTTTTATGAA
AGTTTATGGATATAGAGGTAAACATTTGGGTGGATCAAGGAAACCGGACGGAGCAATTTATACTGTCCGATCTCCT
ATTGATTACGGTGTGATCGTGGATACTAAAGCTTATAGCGGAGGTTATAATCTGCCAATTGGCCAAGCAGATGAAA
TGCAACGATATGTCGAAGAAAATCAAACACGAAACAAACATATCAACCCTAATGAATGGTGGAAAGTCTATCCATC
TTCTGTAACGGAATTTAAGTTTTTATTTGTGAGTGGTCACTTTAAAGGAAACTACAAAGCTCAGCTTACACGATTA
AATCATATCACTAATTGTAATGGAGCTGTTCTTAGTGTAGAAGAGCTTTTAATTGGTGGAGAAATGATTAAGCCG
GCACATTAACCTTAGAGGAAGTCAGACGGAAATTTAATAACGGCGAGATAAACTTT (SEQ ID NO:50)

[0419] fCas9 (例如NLS-dCas9-GGS3接头-FokI) :

[0420] ATGGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGACGATGACAA
GATGGCCCCAAGAAGAAGAGGAAGGTGGGCATTCACCGCGGGGTACCTATGGATAAAAAGTATTCTATTGGTTTA
GCTATCGGCACTAATTCGGTTGGATGGGCTGTCATAACCGATGAATACAAAGTACCTTCAAAGAAATTTAAGGTGT
TGGGGAACACAGACCGTCATTCGATTA AAAAGAATCTTATCGGTGCCCTCCTATTCGATAGTGGCGAAACGGCAGA
GGCGACTCGCCTGAAACGAACCGCTCGGAGAAGGTATACACGTGCGAAGAACCGAATATGTTACTTACAAGAAATT
TTTAGCAATGAGATGGCCAAAGTTGACGATTCTTTCTTTCACCGTTTGAAGAGTCCTTCCTTGTCGAAGAGGACA
AGAAACATGAACGGCACCCCATCTTTGAAAACATAGTAGATGAGGTGGCATATCATGAAAAGTACCCAACGATTTA
TCACCTCAGAAAAAGCTAGTTGACTCAACTGATAAAGCGGACCTGAGGTTAATCTACTTGGCTCTTGCCCATATG
ATAAAGTTCCGTGGGCACTTTCTCATTGAGGGTGATCTAAATCCGGACAACCTCGGATGTCGACAAACTGTTTCATCC
AGTTAGTACAAACCTATAATCAGTTGTTTGAAGAGAACCCTATAAATGCAAGTGGCGTGGATGCGAAGGCTATTCT
TAGCGCCCGCTCTCTAAATCCCGACGGCTAGAAAACCTGATCGCACAATTACCCGGAGAGAAGAAAAATGGGTTG
TTCGGTAACCTTATAGCGCTCTCACTAGGCCTGACACCAAATTTAAGTCGAACTTCGACTTAGCTGAAGATGCCA
AATTGCAGCTTAGTAAGGACACGTACGATGACGATCTCGACAATCTACTGGCACAATTTGGAGATCAGTATGCGGA
CTTATTTTTGGCTGCCAAAAACCTTAGCGATGCAATCCTCCTATCTGACATACTGAGAGTTAATACTGAGATTACC
AAGGCGCCGTTATCCGCTTCAATGATCAAAGGTACGATGAACATCACCAAGACTTGACACTTCTCAAGGCCCTAG
TCCGTGACCAACTGCCTGAGAAAATATAAGGAAATATTCTTTGATCAGTCGAAAAACGGGTACGCAGGTTATATTGA
CGGCGGAGCGAGTCAAGAGGAATTCTACAAGTTTATCAAACCCATATTAGAGAAGATGGATGGGACGGAAGAGTTG
CTTGTA AAAACTCAATCGCGAAGATCTACTGCGAAAGCAGCGGACTTTGACAACGGTAGCATTCCACATCAAATCC
ACTTAGGCGAATTGCATGCTATACTTAGAAGGCAGGAGGATTTTTATCCGTTCTCAAAGACAATCGTGA AAAGAT
TGAGAAAATCCTAACCTTTCGCATACCTTACTATGTGGGACCCCTGGCCCCGAGGAACTCTCGGTTTCGCATGGATG
ACAAGAAAGTCCGAAGAAACGATTACTCCATGGAATTTTGGAGGAAAGTTGTCGATAAAGGTGCGTCAGCTCAATCGT
TCATCGAGAGGATGACCAACTTTGACAAGAATTTACCGAACGAAAAAGTATTGCCTAAGCACAGTTTACTTTACGA
GATTTTACAGTGTACAATGAACTCACGAAAGTTAAGTATGTCACTGAGGGCATGCGTAAACCCGCTTTCTAAGC

GGAGAACAGAAGAAAGCAATAGTAGATCTGTTATTCAAGACCAACCGCAAAGTGACAGTTAAGCAATTGAAAGAGG
ACTACTTTAAGAAAATTGAATGCTTCGATTCTGTGAGATCTCCGGGGTAGAAGATCGATTTAATGCGTCACCTGG
TACGTATCATGACCTCCTAAAGATAATTAAGATAAGGACTTCCTGGATAACGAAGAGAATGAAGATATCTTAGAA
GATATAGTGTGACTCTTACCCTCTTTGAAGATCGGGAAATGATTGAGGAAAGACTAAAAACATACGCTCACCTGT
TCGACGATAAGGTTATGAAACAGTTAAAGAGGCGTCGCTATACGGGCTGGGGACGATTGTGCGGAAACTTATCAA
CGGGATAAGAGACAAGCAAAGTGGTAAAACCTATTCTCGATTTTCTAAAGAGCGACGGCTTCGCCAATAGGAACTTT
ATGCAGCTGATCCATGATGACTCTTTAACCTTCAAAGAGGATATACAAAAGGCACAGTTTTCCGGACAAGGGGACT
CATTGCACGAACATATTGCGAATCTTGCTGGTTCGCCAGCCATCAAAAAGGGCATACTCCAGACAGTCAAAGTAGT
GGATGAGCTAGTTAAGGTCATGGGACGTCACAAACCGGAAAACATTGTAATCGAGATGGCACGCGAAAATCAAACG
ACTCAGAAGGGGCAAAAAACAGTCGAGAGCGGATGAAGAGAATAGAAGAGGGTATTAAGAAGTGGGCAGCCAGA
TCTTAAAGGAGCATCCTGTGGAAAATACCCAATTGCAGAACGAGAAAACCTTACCTCTATTACCTACAAAATGGAAG
GGACATGTATGTTGATCAGGAACTGGACATAAACCGTTTATCTGATTACGACGTCGATGCCATTGTACCCCAATCC
TTTTTGAAGGACGATTCAATCGACAATAAAGTGCTTACACGCTCGGATAAGAACCGAGGGAAAAGTGACAATGTTC
CAAGCGAGGAAGTCGTAAAGAAAATGAAGAACTATTGGCGGCAGCTCCTAAATGCGAAACTGATAACGCAAAGAAA
GTTTCGATAACTTAACTAAAGCTGAGAGGGGTGGCTTGTCTGAACTTGACAAGGCCGGATTTATTAACGTCAGCTC
GTGGAACCCGCCAAATCACAAAGCATGTTGCACAGATACTAGATTCCCGAATGAATACGAAATACGACGAGAACG
ATAAGCTGATTCGGGAAGTCAAAGTAATCACTTTAAAGTCAAATTTGGTGTGCGACTTCAGAAAGGATTTTCAATT
CTATAAAGTTAGGGAGATAAATACTACCACCATGCGCACGACGCTTATCTTAATGCCGTCGTAGGGACCGCACTC
ATTAAGAAATACCCGAAGCTAGAAAGTGAGTTTGTGTATGGTGATTACAAAGTTTATGACGTCGGTAAGATGATCG
CGAAAAGCGAACAGGAGATAGGCAAGGCTACAGCCAAATACTTCTTTTATTCTAACATTATGAATTTCTTTAAGAC
GGAAATCACTCTGGCAAACGGAGAGATACGCAAACGACCTTAAATTGAAACCAATGGGGAGACAGGTGAAATCGTA
TGGGATAAGGGCCGGGACTTCGCGACGGTGAGAAAAGTTTTGTCCATGCCCAAGTCAACATAGTAAAGAAAACCTG
AGGTGCAGACCCGAGGGTTTTCAAAGGAATCGATTCTTCCAAAAGGAATAGTGATAAGCTCATCGCTCGTAAAAA
GGACTGGGACCCGAAAAAGTACGGTGGCTTCGATAGCCCTACAGTTGCCTATTCTGTCCTAGTAGTGCCAAAAGTT
GAGAAGGGAAAATCCAAGAACTGAAGTCAGTCAAAGAATTATTGGGGATAACGATTATGGAGCGCTCGTCTTTTG
AAAAGAACCCCATCGACTTCCTTGAGGCGAAAGGTTACAAGGAAGTAAAAAAGGATCTCATAATTAACCTACCAA
GTATAGTCTGTTTGTAGTTAGAAAATGGCCGAAAACGGATGTTGGCTAGCGCCGGAGAGCTTCAAAGGGGAACGAA
CTCGCACTACCGTCTAAATACGTGAATTTCTGTATTTAGCGTCCCATTACGAGAAGTTGAAAGGTTACCTGAAG
ATAACGAACAGAAGCAACTTTTTGTTGAGCAGCACAAACATTATCTCGACGAAATCATAGAGCAAATTTGGAATT
CAGTAAGAGAGTCATCCTAGCTGATGCCAATCTGGACAAAGTATTAAGCGCATAACAAGCACAGGGATAAACC
ATACGTGAGCAGGCGGAAAAATATTATCCATTTGTTTACTCTTACCAACCTCGGCGCTCCAGCCGATTCAAGTATT
TTGACACAACGATAGATCGCAAACGATACACTTCTACCAAGGAGGTGCTAGACGCGACTGATTACCAATCCAT
CACGGGATTATATGAAACTCGGATAGATTTGTCACAGCTTGGGGGTGACTCAGGTGGAAGTGGCGGCAGCGGAGGT
TCTGGATCCCAACTAGTCAAAGTGAAGTGGAGGAGAAGAAATCTGAACTTCGTCATAAATTGAAATATGTGCCTC
ATGAATATATTGAATTAATTGAAATTGCCAGAAATCCACTCAGGATAGAATTCTTGAATGAAGGTAATGGAATT
TTTTATGAAAGTTTATGGATATAGAGGTAACATTTGGGTGGATCAAGGAAAACCGACGGAGCAATTTATACTGTC
GGATCTCCTATTGATTACGGTGTGATCGTGGATACTAAAGCTTATAGCGGAGGTTATAATCTGCCAATTGGCCAAG
CAGATGAAATGCAACGATATGTCGAAGAAAATCAAACACGAAACAAACATATCAACCCTAATGAATGGTGGAAAGT
CTATCCATCTTCTGTAACGGAATTTAAGTTTTTATTTGTGAGTGGTCACTTTAAAGGAACTACAAAGCTCAGCTT

ACACGATTAATCATATCACTAATTGTAATGGAGCTGTTCTTAGTGTAGAAGAGCTTTAATTGGTGGAGAAATGA
TTAAAGCCGGCACATTAACCTTAGAGGAAGTCAGACGGAAATTTAATAACGGCGAGATAAACTTT (SEQ ID NO:
51)

[0421] fCas9 (例如FokI-GGS3接头-dCas9-NLS) :

[0422] ATGGGATCCCAACTAGTCAAAAGTGAAGTGGAGGAGAAGAAAATCTGAACTTCGTCATAAATTGAAATA
TGTGCCTCATGAATATATTGAATTAATTGAAATTGCCAGAAATCCACTCAGGATAGAATTCTTGAATGAAGGTA
ATGGAATTTTTTATGAAAGTTTATGGATATAGAGGTAAACATTTGGGTGGATCAAGGAAACCGGACGGAGCAATTT
ATACTGTCGGATCTCCTATTGATTACGGTGTGATCGTGGATACTAAAGCTTATAGCGGAGGTTATAATCTGCCAAT
TGGCCAAGCAGATGAAATGCAACGATATGTCGAAGAAAATCAAACACGAAAACAAACATATCAACCCTAATGAATGG
TGGAAAGTCTATCCATCTTCTGTAACGGAATTTAAGTTTTTATTTGTGAGTGGTCACTTTAAAGGAACTACAAAG
CTCAGCTTACACGATTAATCATATCACTAATTGTAATGGAGCTGTTCTTAGTGTAGAAGAGCTTTAATTGGTGG
AGAAATGATTAAGCCGGCACATTAACCTTAGAGGAAGTCAGACGGAAATTTAATAACGGCGAGATAAACTTTGGC
GGTAGTGGGGGATCTGGGGGAAGTATGGATAAAAAGTATTCTATTGGTTTAGCTATCGGCACTAATCCGTTGGAT
GGGCTGTCATAACCGATGAATACAAAGTACCTTCAAAGAAAATTTAAGGTGTTGGGGAACACAGACCGTCATTCCAT
TAAAAAGAATCTTATCGGTGCCCTCTATTCGATAGTGGCGAAACGGCAGAGGCGACTCGCCTGAAACGAACCGCT
CGGAGAAGGTATACACGTCGCAAGAACCGAATATGTTACTTACAAGAAAATTTTAGCAATGAGATGGCCAAAGTTG
ACGATTCTTTCTTTCACCGTTTGAAGAGTCCTTCTTGTGCAAGAGGACAAGAAACATGAACGGCACCCCATCTT
TGAAACATAGTAGATGAGGTGGCATATCATGAAAAGTACCCAACGATTTATCACCTCAGAAAAAGCTAGTTGAC
TCAACTGATAAAGCGGACCTGAGGTTAATCTACTTGGCTCTTGCCCATATGATAAAGTCCGTGGGCACTTTCTCA
TTGAGGGTATCTAAATCCGGACAACCTCGGATGTCGACAAACTGTTTCATCCAGTTAGTACAAACCTATAATCAGTT
GTTTGAAGAGAACCCTATAAATGCAAGTGGCGTGGATGCGAAGGCTATTCTTAGCGCCCGCCTCTCTAAATCCCGA
CGGCTAGAAAACCTGATCGCACAATTACCCGGAGAGAAGAAAATGGGTTGTTTCGGTAACCTTATAGCGCTCTCAC
TAGGCCTGACACCAAATTTAAGTCAACTTCGACTTAGCTGAAGATGCCAAATTGCAGCTTAGTAAGGACACGTA
CGATGACGATCTCGACAATCTACTGGCACAATTGGAGATCAGTATGCGGACTTATTTTTGGCTGCCAAAAACCTT
AGCGATGCAATCCTCCTATCTGACATACTGAGAGTTAATACTGAGATTACCAAGGCGCCGTTATCCGCTTCAATGA
TCAAAAGGTACGATGAACATCACCAGACTTGACACTTCTCAAGGCCCTAGTCCGTCAGCAACTGCCTGAGAAATA
TAAGGAAATATTCTTTGATCAGTCGAAAAACGGGTACGCAGGTTATATTGACGGCGGAGCGAGTCAAGAGGAATTC
TACAAGTTTATCAAACCATATTAGAGAAGATGGATGGGACGGAAGAGTTGCTTGTAAAACCTCAATCGCGAAGATC
TACTGCGAAAGCAGCGGACTTTCGACAACGGTAGCATTCCACATCAAATCCACTTAGGCGAATTGCATGCTATACT
TAGAAGGCAGGAGGATTTTTATCCGTTCCCTCAAAGACAATCGTGAAAAGATTGAGAAAATCCTAACCTTTTCGATA
CCTTACTATGTGGGACCCCTGGCCCGAGGAACTCTCGTTTCGCATGGATGACAAGAAAGTCCGAAGAAACGATTA
CTCCATGGAATTTTGAAGAAAGTTGTCGATAAAGGTGCGTCAGCTCAATCGTTCATCGAGAGGATGACCAACTTTGA
CAAGAATTTACCGAACGAAAAAGTATTGCCTAAGCACAGTTTACTTTACGAGTATTTACAGTGTACAATGAACTC
ACGAAAGTTAAGTATGTCAGTGGGATGCGTAAACCCGCTTTCTAAGCGGAGAACAGAAGAAAGCAATAGTAG
ATCTGTTATTCAAGACCAACCGCAAAGTGACAGTTAAGCAATTGAAAGAGGACTACTTTAAGAAAATTGAATGCTT
CGATTCTGTGAGATCTCCGGGGTAGAAGATCGATTTAATGCGTCACTTGGTACGTATCATGACCTCCTAAAGATA
ATTAAGATAAGGACTTCTGGATAACGAAGAGAATGAAGATATCTTAGAAGATATAGTGTGACTCTTACCCTCT
TTGAAGATCGGGAAATGATTGAGGAAAGACTAAAAACATACGCTCACCTGTTGACGATAAGGTTATGAAACAGTT
AAAGAGGCGTCGCTATACGGGCTGGGACGATTGTCGCGGAAACTTATCAACGGGATAAGAGACAAGCAAAGTGGT

AAAAC TATTCTCGATTTTCTAAAGAGCGACGGCTTCGCCAATAGGAACTTTATGCAGCTGATCCATGATGACTCTT
TAACCTTCAAAGAGGATATACAAAAGGCACAGGTTTCCGGACAAGGGGACTCATTGCACGAACATATTGCGAATCT
TGCTGGTTCCGACCCATCAAAAAGGGCATACTCCAGACAGTCAAAGTAGTGGATGAGCTAGTTAAGGTCATGGGA
CGTCACAAAACCGGAAAAACATTGTAATCGAGATGGCACGCGAAAAATCAAACGACTCAGAAGGGGCAAAAAACAGTC
GAGAGCGGATGAAGAGAATAGAAGAGGTATTAAGAAGTGGGCAGCCAGATCTTAAAGGAGCATCTGTGAAAA
TACCCAATTGCAGAACGAGAACTTTACCTCTATTACCTACAAAATGGAAGGGACATGTATGTTGATCAGGAACTG
GACATAAACCGTTTATCTGATTACGACGTCGATGCCATTGTACCCCAATCCTTTTTGAAGGACGATTCAATCGACA
ATAAAGTGCTTACACGCTCGGATAAGAACCGAGGGAAAAGTGACAATGTTCCAAGCGAGGAAGTCGTAAAGAAAAT
GAAGA ACTATTGGCGCAGCTCCTAAATGCGAACTGATAACGCAAAGAAAAGTTCGATAACTTAACTAAAGCTGAG
AGGGGTGGCTTGTCTGAACTTGACAAGGCCGATTTATTAACGTCAGCTCGTGGAAACCCGCCAAATCACAAAGC
ATGTTGCACAGATACTAGATTCCCGAATGAATACGAAATACGACGAGAACGATAAGCTGATTCCGGGAAGTCAAAGT
AATCACTTTAAAGTCAAAAATTGGTGTCCGACTTCAGAAAGGATTTTCAATTCTATAAAGTTAGGGAGATAAATAAC
TACCACCATGCGCAGCAGCTTATCTTAATGCCGTCGTAGGGACCGCACTCATTAAGAAATACCCGAAGCTAGAAA
GTGAGTTTGTGTATGGTGATTACAAAGTTTATGACGTCCGTAAGATGATCGCGAAAAGCGAACAGGAGATAGGCAA
GGCTACAGCCAAATACTTCTTTATTCTAACATTATGAATTTCTTTAAGACGAAAATCACTCTGGCAAACGGAGAG
ATACGCAAACGACCTTAAATTGAAACCAATGGGGAGACAGGTGAAATCGTATGGGATAAGGGCCGGGACTTCGCGA
CGGTGAGAAAAGTTTTGTCCATGCCCAAGTCAACATAGTAAAGAAAACACTGAGGTGCAGACCGGAGGGTTTTCAA
GGAATCGATTCTTCAAAAAGGAATAGTGATAAGCTCATCGCTCGTAAAAAGGACTGGGACCCGAAAAAGTACGGT
GGCTTCGATAGCCCTACAGTTGCCTATTCTGTCTAGTAGTGCCAAAAGTTGAGAAGGGAAAATCCAAGAACTGA
AGTCAGTCAAAGAATTATTGGGGATAACGATTATGGAGCGCTCGTCTTTTGAAGAACCCCATCGACTTCCTTGA
GGCGAAAAGGTTACAAGGAAGTAAAAAGGATCTCATAATTAACCTACCAAAGTATAGTCTGTTGAGTTAGAAAAT
GGCCGAAAACGGATGTTGGCTAGCGCCGAGAGCTTCAAAGGGGAACGAACTCGCACTACCGTCTAAATACGTGA
ATTTCTGTATTTAGCGTCCCATTACGAGAAGTTGAAAGGTTACCTGAAGATAACGAACAGAAGCAACTTTTTGT
TGAGCAGCACAAAACATTATCTCGACGAAATCATAGAGCAAATTTCCGAATTCAGTAAGAGAGTCATCCTAGCTGAT
GCCAATCTGGACAAAAGTATTAAGCGCATAACAAGCACAGGGATAAAACCCATACGTGAGCAGGCGGAAAATATTA
TCCATTTGTTTACTCTTACCAACCTCGGCGCTCCAGCCGATTCAAGTATTTTGACACAACGATAGATCGCAAACG
ATACACTTCTACCAAGGAGGTGCTAGACGCGACACTGATTCACCAATCCATCACGGGATTATATGAACTCGGATA
GATTTGTACAGCTTGGGGGTGACGGATCCCCAAGAAGAAGAGGAAAAGTCTCGAGCGACTACAAAGACCATGACG
GTGATTATAAAGATCATGACATCGATTACAAGGATGACGATGACAAGGCTGCAGGA (SEQ ID NO: 52)

[0423] fCas9 (例如NLS-FokI-GGS3接头-dCas9) :

[0424] ATGGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGACGATGACAA
GATGGCCCCAAGAAGAAGAGGAAGGTGGGCATTACCAGCGGGGTACCTGGAGGTTCTATGGGATCCCAACTAGTC
AAAAGTGAACCTGGAGGAGAAGAAATCTGAACTTCGTCATAAATTGAAATATGTGCCTCATGAATATATTGAATTA
TTGAAATTGCCAGAAATCCACTCAGGATAGAATTCTTGAAATGAAGGTAATGGAATTTTTTATGAAAGTTTATGG
ATATAGAGGTAAACATTTGGGTGGATCAAGGAAACCGACGGAGCAATTTATACTGTCGGATCTCCTATTGATTAC
GGTGTGATCGTGGATACTAAAGCTTATAGCGGAGTTATAATCTGCCAATTGGCCAAGCAGATGAAATGCAACGAT
ATGTGCAAGAAAATCAAACACGAAACAAACATATCAACCCTAATGAATGGTGGAAAGTCTATCCATCTTCTGTAAC
GGAATTTAAGTTTTTATTTGTGAGTGGTCACTTTAAAGGAACTACAAAGCTCAGCTTACACGATTAATCATATC
ACTAATTGTAATGGAGCTGTTCTTAGTGTAGAAGAGCTTTAATTGGTGGAGAAATGATTAAAGCCGGCACATTA

CCTTAGAGGAAGTCAGACGGAAATTTAATAACGGCGAGATAAACTTTGGCGGTAGTGGGGGATCTGGGGGAAGTAT
GGATAAAAAGTATTCTATTGGTTTAGCTATCGGCTAATTCGGTTGGATGGGCTGCATAACCGATGAATACAAA
GTACCTTCAAAGAAATTTAAGGTGTTGGGGAACACAGACCGTCATTTCGATTAAGAAAGTCTTATCGGTGCCCTCC
TATTCGATAGTGGCGAAACGGCAGAGGCGACTCGCCTGAAACGAACCGCTCGGAGAAGGTATACACGTCGCAAGAA
CCGAATATGTTACTTACAAGAAATTTTTAGCAATGAGATGGCCAAAAGTTGACGATTCTTTCTTTACCCTTTGGAA
GAGTCCTTCTTGTGCGAAGAGGACAAGAAACATGAACGGCACCCCATCTTTGGAAACATAGTAGATGAGGTGGCAT
ATCATGAAAAGTACCAACGATTTATCACCTCAGAAAAAGCTAGTTGACTCAACTGATAAAGCGGACCTGAGGTT
AATCTACTTGGCTCTTGCCCATATGATAAAGTTCCGTGGGCACTTTCTCATTGAGGGTGATCTAAATCCGGACAAC
TCGGATGTCGACAAAAGTTCATCCAGTTAGTACAAACCTATAATCAGTTGTTTGAAGAGAACCCTATAAATGCAA
GTGGCGTGGATGCGAAGGCTATTCTTAGCGCCCGCTCTCTAAATCCCGACGGCTAGAAAACCTGATCGCACAATT
ACCCGGAGAGAAGAAAAATGGGTTGTTCCGTAACCTTATAGCGCTCTCACTAGGCCCTGACACCAAATTTAAGTCG
AACTTCGACTTAGCTGAAGATGCCAAATTGCAGCTTAGTAAGGACACGTACGATGACGATCTCGACAATCTACTGG
CACAAATTGGAGATCAGTATGCGGACTTATTTTTGGCTGCCAAAAACCTTAGCGATGCAATCCTCCTATCTGACAT
ACTGAGAGTTAATACTGAGATTACCAAGGCGCCGTTATCCGCTTCAATGATCAAAAAGGTACGATGAACATCACCAA
GACTTGACACTTCTCAAGGCCCTAGTCCGTCAGCAACTGCCTGAGAAATATAAGGAAATATTCTTTGATCAGTCGA
AAAACGGGTACGCAGGTTATATTGACGGCGGAGCGAGTCAAGAGGAATTCTACAAGTTTATCAAACCCATATTAGA
GAAGATGGATGGGACGGAAGAGTTGCTTGTA AAACTCAATCGCGAAGATCTACTGCGAAAAGCAGCGGACTTTTCGAC
AACGGTAGCATTCCACATCAAATCCACTTAGGCGAATTGCATGCTATACTTAGAAGGCAGGAGGATTTTTATCCGT
TCCTCAAAGACAATCGTGA AAAAGATTGAGAAAATCCTAACCTTTTCGCATACCTTACTATGTGGGACCCCTGGCCCG
AGGGAACCTCTCGGTTTCGCATGGATGACAAGAAAGTCCGAAGAAACGATTACTCCATGGAATTTGAGGAAGTTGTC
GATAAAGGTGCGTCAGCTCAATCGTTCATCGAGAGGATGACCAACTTTGACAAGAAATTTACCGAACGAAAAAGTAT
TGCCTAAGCACAGTTTACTTTACGAGTATTTACAGTGTACAATGAACTCACGAAAGTTAAGTATGTCACTGAGGG
CATGCGTAAACCCGCTTTCTAAGCGGAGAACAGAAGAAAGCAATAGTAGATCTGTTATTCAAGACCAACCGCAAA
GTGACAGTTAAGCAATTGAAAGAGGACTACTTTAAGAAAATTGAATGCTTCGATTCTGTCGAGATCTCCGGGGTAG
AAGATCGATTTAATGCGTCACTTGGTACGTATCATGACCTCCTAAAGATAATTAAGATAAGGACTTCCTGGATAA
CGAAGAGAATGAAGATATCTTAGAAGATATAGTGTGACTCTTACCCTCTTTGAAGATCGGGAAATGATTGAGGAA
AGACTAAAAACATACGCTCACCTGTTTCGACGATAAGGTTATGAAACAGTTAAAGAGGCGTCGCTATACGGGCTGGG
GACGATTGTCGCGGAAAAGTATCAACGGGATAAGAGACAAGCAAAGTGGTAAAAGTATTTCTCGATTTTCTAAAGAG
CGACGGCTTCGCCAATAGGAACTTTATGCAGCTGATCCATGATGACTCTTTAACCTTCAAAGAGGATATACAAAAG
GCACAGGTTTCCGGACAAGGGGACTCATTGCACGAACATATTGCGAATCTTGCTGGTTCCGACCCATCAAAAAGG
GCATACTCCAGACAGTCAAAGTAGTGGATGAGCTAGTTAAGGTCATGGGACGTCACAAACCGGAAAACATTGTAAT
CGAGATGGCAGCGGAAAATCAAACGACTCAGAAGGGGCAAAAAACAGTCGAGAGCGGATGAAGAGAATAGAAGAG
GGTATTAAGAACTGGGCAGCCAGATCTTAAAGGAGCATCCTGTGGAAAATACCCAATTGCAGAACGAGAACTTT
ACCTCTATTACCTACAAAATGGAAGGGACATGTATGTTGATCAGGAACTGGACATAAACCGTTTATCTGATTACGA
CGTCGATGCCATTGTACCCCAATCCTTTTTGAAGGACGATTCAATCGACAATAAAGTGCTTACACGCTCGGATAAG
AACCGAGGGAAAAGTGACAATGTTCCAAGCGAGGAAGTCGTAAGAAAATGAAGAACTATTGGCGGCAGCTCCTAA
ATGCGAACTGATAACGCAAAGAAAGTTCGATAACTTAACTAAAGCTGAGAGGGGTGGCTTGTCTGAACTTGACAA
GGCCGGATTTATTAACGTCAGCTCGTGAAACCCGCCAAATCACAAAGCATGTTGCACAGATACTAGATTCCCGA
ATGAATACGAAATACGACGAGAACGATAAGCTGATTCGGGAAGTCAAAGTAATCACTTTAAAGTCAAATTTGGTGT

CGGACTTCAGAAAGGATTTTCAATTCTATAAAGTTAGGGAGATAAATAACTACCACCATGCGCAGCAGCCTTATCT
TAATGCCGTCGTAGGGACCGCACTCATTAAGAAATACCCGAAGCTAGAAAAGTGAGTTTGTGTATGGTGATTACAAA
GTTTATGACGTCCGTAAGATGATCGCGAAAAGCGAACAGGAGATAGGCAAGGCTACAGCCAAATACTTCTTTTATT
CTAACATTATGAATTTCTTTAAGACGGAAATCACTCTGGCAAACGGAGAGATACGCAAACGACCTTTAATTGAAAC
CAATGGGGAGACAGGTGAAATCGTATGGGATAAGGGCCGGGACTTCGCGACGGTGAGAAAAGTTTTGTCCATGCC
CAAGTCAACATAGTAAAGAAAAGTGAAGTGCAGACCGGAGGGTTTTCAAAGGAATCGATTCTTCCAAAAGGAATA
GTGATAAGCTCATCGCTCGTAAAAAGGACTGGGACCCGAAAAGTACGGTGGCTTCGATAGCCCTACAGTTGCCTA
TTCTGTCTAGTAGTGGCAAAAGTTGAGAAGGGAAAATCCAAGAAAAGTGAAGTCAGTCAAAGAATTATTGGGGATA
ACGATTATGGAGCGCTCGTCTTTTAAAAGAACCCCATCGACTTCCTTGAGGCGAAAAGGTTACAAGGAAGTAAAA
AGGATCTCATAATTAAGTACCAAAGTATAGTCTGTTTGTAGTTAGAAAATGGCCGAAAACGGATGTTGGCTAGCGC
CGGAGAGCTTCAAAGGGGAACGAACCTCGACTACCGTCTAAATACGTGAATTTCTGTATTTAGCGTCCCATTAC
GAGAAGTTGAAAGGTTACCTGAAGATAACGAACAGAAGCAACTTTTTGTGAGCAGCACAAACATTATCTCGACG
AAATCATAGAGCAAATTTTCGGAATTCAGTAAGAGAGTCATCTAGCTGATGCCAATCTGGACAAAGTATTAAGCGC
ATACAACAAGCACAGGGATAAACCATACGTGAGCAGGCGGAAAATATTATCCATTTGTTTACTCTTACCAACCTC
GGCGCTCCAGCCGATTCAAGTATTTGACACAACGATAGATCGCAAACGATACACTTCTACCAAGGAGGTGCTAG
ACGCGACACTGATTCACCAATCCATCACGGGATTATATGAAACTCGGATAGATTTGTACAGCTTGGGGGTGAC
(SEQ ID NO:53)

[0425] fCas9:

[0426] ATGGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGACGATGACAA
GATGGCCCCAAGAAGAAGAGGAAGGTGGGCATTCACCGCGGGGTACCTGGAGGTTCTGGATCCCAACTAGTCAAA
AGTGAAGTGGAGGAGAAGAAATCTGAAGTTCGTCATAAATTGAAATATGTGCCTCATGAATATATTGAATTAATTG
AAATTGCCAGAAATTCCTCAGGATAGAATTCTTGAATGAAGGTAATGGAATTTTTTATGAAAGTTTATGGATA
TAGAGGTAACATTTGGGTGGATCAAGGAAACCGGACGGAGCAATTTATACTGTCCGATCTCCTATTGATTACGGT
GTGATCGTGATACTAAAGCTTATAGCGGAGGTTATAATCTGCCAATTGGCCAAGCAGATGAAATGCAACGATATG
TCGAAGAAAATCAAACACGAAACAAACATATCAACCCTAATGAATGGTGGAAAGTCTATCCATCTTCTGTAACGGA
ATTTAAGTTTTTATTTGTGAGTGGTCACTTTAAAGGAAACTACAAAGCTCAGCTTACACGATTAATCATATCACT
AATTGTAATGGAGCTGTTCTTAGTGTAGAAGAGCTTTTAATTGGTGGAGAAATGATTAAAGCCGGCACATTAACCT
TAGAGGAAGTCAGACGGAAATTTAATAACGGCAGATAAACTTTAGCGGCAGCGAGACTCCCGGGACCTCAGAGTC
CGCCACACCCGAAAGTGATAAAAAGTATTCTATTGGTTTAGCTATCGGCACTAATCCGTTGGATGGGCTGTCATA
ACCGATGAATACAAAGTACCTTCAAAGAAATTTAAGGTGTTGGGGAACACAGACCGTCATTCGATTAAGAAGATC
TTATCGGTGCCCTCCTATTCGATAGTGGCGAAACGGCAGAGGGGACTCGCCTGAAACGAACCGCTCGGAGAAGGTA
TACACGTCGCAAGAACCGAATATGTTACTTACAAGAAATTTTTAGCAATGAGATGGCCAAAGTTGACGATTCTTTC
TTTACCGTTTTGGAAGAGTCCCTTCTTGTGCAAGAGGACAAGAAACATGAACGGCACCCCATCTTTGAAACATAG
TAGATGAGGTGGCATATCATGAAAAGTACCAACGATTTATCACCTCAGAAAAAAGCTAGTTGACTCAACTGATAA
AGCGGACCTGAGGTTAATCTACTTGGCTCTTGCCATATGATAAAGTTCCGTGGGCACTTTCTCATTGAGGGTGAT
CTAAATCCGGACAACCTCGGATGTCGACAAACTGTTTCATCCAGTTAGTACAAACCTATAATCAGTTGTTTGAAGAGA
ACCCTATAAATGCAAGTGGCGTGGATGCGAAGGCTATTCTTAGCGCCCGCCTCTCTAAATCCCGACGGCTAGAAAA
CCTGATCGCACAAATACCCGGAGAGAAGAAAATGGGTTGTTGCGTAACCTTATAGCGCTCTCACTAGGCCTGACA
CCAAATTTTAAGTCGAACTTCGACTTAGCTGAAGATGCCAATTGCAGCTTAGTAAGGACACGTACGATGACGATC

TCGACAATCTACTGGCACAAATTGGAGATCAGTATGCGGACTTATTTTTGGCTGCCAAAAACCTTAGCGATGCAAT
CCTCCTATCTGACATACTGAGAGTTAATACTGAGATTACCAAGGCGCCGTTATCCGCTTCAATGATCAAAAGGTAC
GATGAACATCACCAAGACTTGACACTTCTCAAGGCCCTAGTCCGTCAGCAACTGCCTGAGAAATATAAGGAAATAT
TCTTTGATCAGTCGAAAAACGGGTACGCAGGTTATATTGACGGCGGAGCGAGTCAAGAGGAATTCTACAAGTTTAT
CAAACCCATATTAGAGAAGATGGATGGGACGGAAGAGTTGCTTGAAAACTCAATCGCGAAGATCTACTGCGAAAG
CAGCGGACTTTGACAACGGTAGCATTCCACATCAAATCCACTTAGGCGAATTGCATGCTATACTTAGAAGGCAGG
AGGATTTTTATCCGTTCCCTCAAAGACAATCGTGAAAAGATTGAGAAAACTCTAACCTTTCGCATACCTTACTATGT
GGGACCCCTGGCCGAGGGAACCTCGGTTTCGCATGGATGACAAGAAAAGTCCGAAGAAACGATTACTCCATGGAAT
TTTGAGGAAGTTGTCGATAAAGGTGCGTCAGCTCAATCGTTCATCGAGAGGATGACCAACTTTGACAAGAATTTAC
CGAACGAAAAAGTATTGCCTAAGCACAGTTTACTTTACGAGTATTTACAGTGTACAATGAACTCACGAAAGTTAA
GTATGTCACTGAGGGCATGCGTAAACCCGCCTTTCTAAGCGGAGAACAGAAGAAAAGCAATAGTAGATCTGTTATTC
AAGACCAACCGCAAAGTGACAGTTAAGCAATTGAAAGAGGACTACTTTAAGAAAAATTGAATGCTTCGATTCTGTCTG
AGATCTCCGGGGTAGAAGATCGATTTAATGCGTCACTTGGTACGTATCATGACCTCCTAAAGATAATTAAGATAA
GGACTTCCTGGATAACGAAGAGAATGAAGATATCTTAGAAGATATAGTGTGACTCTTACCCTCTTTGAAGATCGG
GAAATGATTGAGGAAAGACTAAAAACATACGCTCACCTGTTTCGACGATAAGGTTATGAAACAGTTAAAGAGGCGTC
GCTATACGGGCTGGGGACGATTGTCGCGGAACTTATCAACGGGATAAGAGACAAGCAAAGTGGTAAAACTATTCT
CGATTTTCTAAAGAGCGACGGCTTCGCCAATAGGAACTTTATGCAGCTGATCCATGATGACTCTTTAACCTTCAAA
GAGGATATACAAAAGGCACAGGTTCCGGACAAGGGGACTCATTGCACGAACATATTGCGAATCTTGCTGGTTTCGC
CAGCCATCAAAAAGGCATACTCCAGACAGTCAAAGTAGTGGATGAGCTAGTTAAGGTCATGGGACGTCACAAACC
GGAAAACATTGTAATCGAGATGGCACGCGAAAATCAAACGACTCAGAAGGGGCAAAAAACAGTCGAGAGCGGATG
AAGAGAATAGAAGAGGGTATTAAGAAGTGGGACCCAGATCTTAAAGGAGCATCCTGTGAAAAATACCCAATTGC
AGAACGAGAACTTTACCTCTATTACCTACAAAATGGAAGGGACATGTATGTTGATCAGGAACTGGACATAAACCG
TTTATCTGATTACGACGTCGATGCCATTGTACCCCAATCCTTTTTGAAGGACGATTCAATCGACAATAAAGTGCTT
ACACGCTCGGATAAGAACCGAGGGAAAAGTGACAATGTTCCAAGCGAGGAAGTCGTAAAGAAAATGAAGAACTATT
GGCGGCAGCTCCTAAATGCGAACTGATAACGCAAAGAAAGTTGATAACTTAACTAAAGCTGAGAGGGGTGGCTT
GTCTGAACTTGACAAGGCCGATTTATTAACGTCAGCTCGTGGAAAACCCGCCAAATCACAAAGCATGTTGCACAG
ATACTAGATTCCCGAATGAATACGAAATACGACGAGAACGATAAGCTGATTCCGGGAAGTCAAAGTAATCACTTTAA
AGTCAAAATTGGTGTGCGGACTTCAGAAAGGATTTTCAATTCTATAAAGTTAGGGAGATAAATACTACCACCATGC
GCACGACGCTTATCTTAATGCCGTCGTAGGGACCGCACTCATTAAAGAAATACCCGAAGCTAGAAAGTGAGTTTGTG
TATGGTGATTACAAAAGTTTATGACGTCGTAAGATGATCGCGAAAAGCGAACAGGAGATAGGCAAGGCTACAGCCA
AATACTTCTTTTATTCTAACATTATGAATTTCTTTAAGACGGAAATCACTCTGGCAAACGGAGAGATACGCAAACG
ACCTTTAATTGAAACCAATGGGGAGACAGGTGAAATCGTATGGGATAAGGGCCGGGACTTCGCGACGGTGAGAAAA
GTTTTGTCCATGCCCAAGTCAACATAGTAAAGAAAAGTGGAGTGCAGACCGGAGGGTTTTCAAAGGAATCGATTC
TTCCAAAAAGGAATAGTGATAAGCTCATCGCTCGTAAAAAGGACTGGGACCCGAAAAAGTACGGTGGCTTCGATAG
CCCTACAGTTGCCTATTCTGTCCTAGTAGTGGCAAAAGTTGAGAAGGGAAAAATCCAAGAACTGAAGTCAGTCAAA
GAATTATTGGGGATAACGATTATGGAGCGCTCGTCTTTTGAAGAAGACCCATCGACTTCCTTGAGGCGAAAGGTT
ACAAGGAAGTAAAAAAGGATCTCATAATTAAGTACCAAAGTATAGTCTGTTGAGTTAGAAAATGGCCGAAAACG
GATGTTGGCTAGCGCCGAGAGCTTCAAAGGGGAACGAACTCGCACTACCGTCTAAATACGTGAATTTCTGTAT
TTAGCGTCCCATTACGAGAAGTTGAAAGGTTACCTGAAGATAACGAACAGAAGCAACTTTTTGTTGAGCAGCACA

AACATTATCTCGACGAAATCATAGAGCAAATTTTCGGAATTCAGTAAGAGAGTCATCCTAGCTGATGCCAATCTGGA
CAAAGTATTAAGCGCATACAACAAGCACAGGGATAAACCCATACGTGAGCAGGCGGAAAATATTATCCATTTGTTT
ACTCTTACCAACCTCGGCGCTCCAGCCGATTCAAGTATTTTACACAACGATAGATCGCAAACGATACACTTCTA
CCAAGGAGGTGCTAGACGCGACTGATTCACCAATCCATCACGGGATTATATGAAACTCGGATAGATTTGTCACA
GCTTGGGGGTGAC (SEQ ID NO:54)

[0427] 在一些实施方案中,所述酶域包含重组酶或其催化域。例如,在一些实施方案中,例示性的具有重组酶域的配体依赖性dCas9融合蛋白的一般体系结构包含以下结构:

[0428] [NH₂] - [NLS] - [dCas9] - [重组酶] - [COOH],

[0429] [NH₂] - [NLS] - [重组酶] - [dCas9] - [COOH],

[0430] [NH₂] - [dCas9] - [重组酶] - [COOH], 或

[0431] [NH₂] - [重组酶] - [dCas9] - [COOH];

[0432] 其中NLS是核定位信号,NH₂是融合蛋白的N末端,且COOH是融合蛋白的C末端。在一些实施方案中,接头插入在dCas9和重组酶域之间。在一些实施方案中,接头插入在NLS与重组酶和/或dCas9域之间。在一些实施方案中,NLS位于重组酶和/或dCas9域的C末端。在一些实施方案中,NLS位于重组酶和dCas9域之间。还可以存在另外的特征,如序列标签。“重组酶的催化域”意指融合蛋白包含包括重组酶的氨基酸序列(例如衍生于重组酶)的域,从而使该域在与靶核酸接触(单独地或与包含别的重组酶催化域的其他因子组合,所述别的重组酶催化域可能或可能不形成融合蛋白的部分)时足以诱导重组)。在一些实施方案中,重组酶的催化域不包含重组酶的DNA结合域。在一些实施方案中,重组酶的催化域包含部分或全部的重组酶,例如催化域可以包含重组酶域和DNA结合域,或其部分,或者催化域可以包含重组酶域和DNA结合域,其突变或截短以消除DNA结合活性。重组酶和重组酶的催化域是本领域技术人员已知的,且包括例如本文描述的那些。在一些实施方案中,所述催化域源自任意重组酶。在一些实施方案中,重组酶催化域是aTn3解离酶、Hin重组酶或Gin重组酶的催化域。在一些实施方案中,催化域包含Tn3解离酶(例如Stark Tn3重组酶),其由包含部分或全部的如下文提供的SEQ ID NO:55的核苷酸序列编码。在一些实施方案中,Tn3催化域由SEQ ID NO:55的变体编码。在一些实施方案中,Tn3催化域由编码对应于SEQ ID NO:56的多肽的多核苷酸(或其变体)编码。在一些实施方案中,所述催化域包含Hin重组酶,其由包含部分或全部的如下文提供的SEQ ID NO:57的核苷酸序列编码。在一些实施方案中,Hin催化域由SEQ ID NO:57的变体编码。在一些实施方案中,Hin催化域由编码对应于SEQ ID NO:58的多肽的多核苷酸(或其变体)编码。在一些实施方案中,所述催化域包含Gin重组酶(例如Gin beta重组酶),其由包含部分或全部的如下文提供的SEQ ID NO:59的核苷酸序列编码。在一些实施方案中,Gin催化域由SEQ ID NO:59的变体编码。在一些实施方案中,Gin催化域由编码对应于SEQ ID NO:60的多肽的多核苷酸(或其变体)编码。使用dCas9-重组酶融合蛋白的其他例示性组合物和方法可见于2014年6月30日提交的美国专利申请U.S.S.N 14/320,467;名称为“Cas9Variants and Uses Thereof”;其完整内容通过提述并入本文。

[0433] Stark Tn3重组酶(核苷酸:SEQ ID NO:55;氨基酸:SEQ ID NO:56):

[0434] ATGGCCCTGTTTGGCTACGCACGCGTGTCTACCAGTCAACAGTCACTCGATTGCAAGTGAGGGCTCT
TAAAGATGCCGGAGTGAAGGCAAACAGAATTTTTACTGATAAGGCCAGCGGAAGCAGCACAGACAGAGAGGGGCTG
GATCTCCTGAGAATGAAGGTAAAGGAGGGTGATGTGATCTTGGTCAAAAAATTGGATCGACTGGGGAGAGACACAG

CTGATATGCTTCAGCTTATTAAGAGTTTGACGCTCAGGGTGTTCGGTGAGGTTTATCGATGACGGCATCTCAAC
CGACTCCTACATTGGTCTTATGTTTGTGACAATTTTGTCCGCTGTGGCTCAGGCTGAGCGGAGAAGGATTCTCGAA
AGGACGAATGAGGGACGGCAAGCAGCTAAGTTGAAAGGTATCAAATTTGGCAGACGAAGG (SEQ ID NO:55)

[0435] MALFGYARVSTSQQSLDLQVRALKDAGVKANRIFTDKASGSSTDREGLDLLRMKVKEGDVILVKKLDR
LGRDTADMLQLIKEFDAQGVAVRFIDGISTDSYIGLMFVTILSAVAQAERRRILERTNEGRQAAKLGKIKFGRRR
(SEQ ID NO:56)

[0436] Hin重组酶(核苷酸:SEQ ID NO:57;氨基酸:SEQ ID NO:58):

[0437] ATGGCAACCATTGGCTACATAAGGGTGTCTACCATCGACCAAAATATCGACCTGCAGCGCAACGCTCT
GACATCCGCCAACTGCGATCGGATCTTCGAGGATAGGATCAGTGGAAGATCGCCAACCGGCCCGGTCTGAAGCGG
GCTCTGAAGTACGTGAATAAGGGCGATACTCTGGTTGTGTGGAAGTTGGATCGCTTGGGTAGATCAGTGAAGAATC
TCGTAGCCCTGATAAGCGAGCTGCACGAGAGGGGTGCACATTTCCATTCTCTGACCGATTCCATCGATACGTCTAG
CGCCATGGGCCGATTCTTCTTTACGTCATGTCCGCCCTCGCTGAAATGGAGCGCGAACTTATTGTTGAACGGACT
TTGGCTGGACTGGCAGCGGCTAGAGCACAGGGCCGACTTGA (SEQ ID NO:57)

[0438] MATIGYIRVSTIDQNIDLQRNALTSANCDRIFEDRISGKIANRPLKRALKYVNKGDTLVWKLDRLG
RSVKNLVALISELHERGAHFHSLTDSIDTSSAMGRFFFYVMSALAEMERELIVERTLAGLAAARAQRLG (SEQ ID
NO:58)

[0439] Gin beta重组酶(核苷酸:SEQ ID NO:59;氨基酸:SEQ ID NO:60):

[0440] ATGCTCATTGGCTATGTAAGGGTCAGCACCAATGACCAAAACACAGACTTGCAACGCAATGCTTTGGT
TTGCGCCGGATGTGAACAGATATTTGAAGATAAACTGAGCGGCACTCGGACAGACAGACCTGGGCTTAAGAGAGCA
CTGAAAAGACTGCAGAAGGGGGACACCCTGGTCGTCTGGAACTGGATCGCCTCGGACGCAGCATGAAACATCTGA
TTAGCCTGGTTGGTGAGCTTAGGGAGAGGAATCAACTTCAGAAGCCTGACCGACTCCATCGACACCAGTAGCCC
CATGGGACGATTCTTCTTCTATGTGATGGGAGCACTTGCTGAGATGGAAAGAGAGCTTATTATCGAAAGAACTATG
GCTGGTATCGCTGCTGCCCGGAACAAAGGCAGACGGTTCCGGCAGACCCGCGAAGAGCGGC (SEQ ID NO:59)

[0441] MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLKRALKRLQKGDTLVWKLDRLLGR
SMKHLISLVGELRERGINFRSLTDSIDTSSP MGRFFFYVMGALAEMERELI IERTMAGIAAARNKGRFRPPKSG
(SEQ ID NO:60)

[0442] 在一些实施方案中,所述酶域包含脱氨酶或其催化域。例如,在一些实施方案中,具有脱氨酶或域的例示性dCas9融合蛋白的一般体系结构包含以下结构:

[0443] [NH₂] - [NLS] - [Cas9] - [脱氨酶] - [COOH],

[0444] [NH₂] - [NLS] - [脱氨酶] - [Cas9] - [COOH],

[0445] [NH₂] - [Cas9] - [脱氨酶] - [COOH], 或

[0446] [NH₂] - [脱氨酶] - [Cas9] - [COOH];

[0447] 其中NLS是核定位信号,NH₂是融合蛋白的N末端,且COOH是融合蛋白的C末端。在一些实施方案中,接头插入在dCas9和脱氨酶域之间。在一些实施方案中,接头插入在NLS与脱氨酶和/或dCas9域之间。在一些实施方案中,NLS位于脱氨酶和/或dCas9域的C末端。在一些实施方案中,NLS位于脱氨酶和dCas9域之间。还可以存在另外的特征,如序列标签。一个例示性的核酸编辑酶和域的适宜类型是胞嘧啶脱氨酶,例如,胞嘧啶脱氨酶的载脂蛋白BmRNA编辑复合物(APOBEC)家族的,其包含活化诱导的胞苷脱氨酶(AID)和载脂蛋白B编辑复合物3(APOBEC3)酶。适用于本发明的另一个例示性的核酸编辑酶和域的适宜类型包括腺苷脱氨

酶。例如,ADAT家族腺苷脱氨酶可以融合于dCas9域。下文提供了一些例示性的适宜核酸编辑酶和域,例如脱氨酶和脱氨酶域,其可依照本公开的方面融合于dCas9域。将理解,在一些实施方案中,可以使用相应序列的活性域,例如没有定位信号(核定位信号,没有核输出信号,细胞质定位信号)的域。使用dCas9-核酸酶融合蛋白的其他例示性组合和方法可见于2014年7月8日提交的美国专利申请U.S.S.N14/325,815;名称为“Fusions of Cas9 Domains and Nucleic Acid-Editing Domains”;其完整内容通过提述并入本文

[0448] 人AID:

MDSLLMNRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNKGCHVELLFLRYISDW
DLDPGRCYRVTWFTSWSPCYDCARHVADFLRGPNLSLRIFTARLYFCEDRKAEPGLRRLHRAGVQIAI

[0449] MTFKDYFYCWNTFVENHERTFKAWEGLHENSVRLSRQLRRILLPLYEVDDLDRDAFRTLGL (SEQ ID NO:61)

[0450] (划线:核定位信号;双划线:核输出信号)

[0451] 小鼠AID:

MDSLLMKQKFLYHFKNVRWAKGRHETYLCYVVKRRDSATSCSLDFGHLRNKSGCHVELLFLRYISDW
DLDPGRCYRVTWFTSWSPCYDCARHVAEFLRWPNLSLRIFTARLYFCEDRKAEPGLRRLHRAGVQIGI

[0452] MTFKDYFYCWNTFVENRERTFKAWGLHENSVRLTRQLRRILLPLYEVDDLDRDAFRMLGF (SEQ ID NO:62)

[0453] (划线:核定位信号;双划线:核输出信号)

[0454] 犬AID:

MDSLLMKQKFLYHFKNVRWAKGRHETYLCYVVKRRDSATSFSLDFGHLRNKSGCHVELLFLRYISDW
DLDPGRCYRVTWFTSWSPCYDCARHVADFLRGYPNLSLRIFAARLYFCEDRKAEPGLRRLHRAGVQIAI

[0455] MTFKDYFYCWNTFVENREKTFKAWEGLHENSVRLSRQLRRILLPLYEVDDLDRDAFRTLGL (SEQ ID NO:63)

[0456] (划线:核定位信号;双划线:核输出信号)

[0457] 牛AID:

MDSLLKKQRQFLYQFKNVRWAKGRHETYLCYVVKRRDSPTSFSLDFGHLRNKAGCHVELLFLRYISDW
DLDPGRCYRVTWFTSWSPCYDCARHVADFLRGYPNLSLRIFTARLYFCDKERKAEPGLRRLHRAGVQI

[0458] AIMTFKDYFYCWNTFVENHERTFKAWEGLHENSVRLSRQLRRILLPLYEVDDLDRDAFRTLGL (SEQ ID NO:64)

[0459] (划线:核定位信号;双划线:核输出信号)

[0460] 小鼠APOBEC-3:

MGPFCLGCSHRKCYSPIRNLSQETFKFHFKNLGYAKGRKDTFLCYEVTRKDCDSPVSLHHGVFKNKDNI
HAEICFLYWFHDKVLKVLSPREEFKITWYMSWSPCFECAEQIVRFLATHHNLSLDIFSSRLYNVQDPETQ
QNLRLVQEGAQVAAMDLYEFKCKWKKFVDNNGRRFRPWKRLLTNFRYQDSKLQEILRPCYIPVSSSS
STLSNICLTKGLPETRFCVEGRRMDPLSEEFYSQFYNQRVKHLCCYHMKPYLCYQLEQFNGQAPLKG
CLLSEKQKQHAEIFLFDKIRSMELSQVTITCYLTWSPCPNCAWQLAAFKRDRPDLILHIYTSRLYFHWKRP
FQKGLCSLWQSGILVDVMDLPQFTDCWTNFWNPKRPFWPWKGLEIISRRTQRRRLRIKESWGLQDLVND
FGNLQLGPPMS (SEQ ID NO:65)

[0462] (划线:核酸编辑域)

[0463] 大鼠APOBEC-3:

MGPFCLGCSHRKCYSPIRNLSQETFKFHFKNLRYAIDRKDTFLCYEVTRKDCDSPVSLHHGVFKNKDNIH
AEICFLYWFHDKVLKVLSPREEFKITWYMSWSPCFECAEQVLRFLATHHNLSLDIFSSRLYNIRDPENQON
LCRLVQEGAQVAAMDLYEFKCKWKKFVDNNGRRFRPWKLLTNFRYQDSKLQEILRPCYIPVSSSSST
LSNICLTKGLPETRFCVERRRVHLLSEEFYSQFYNQRVKHLCCYHGVKPYLCYQLEQFNGQAPLKGCLL
SEKQKQHAEIFLFDKIRSMELSQVIITCYLTWSPCPNCAWQLAAFKRDRPDLILHIYTSRLYFHWKRP
FQKGLCSLWQSGILVDVMDLPQFTDCWTNFWNPKRPFWPWKGLEIISRRTQRRRLRIKESWGLQDLVND
FGNLQLGPPMS (SEQ ID NO:66)

[0465] (划线:核酸编辑域)

[0466] 恒河猴 (Rhesus macaque) APOBEC-3G:

MVEPMDPRTFVSNFNRPILSGLNTVWLCCEVTKDKPSGPPDLAKIFQGKVYSKAKYHPEMRFLRWFH
KWRQLHHDQEYKVTWYVSWSPCTRCANSVATFLAKDPKVTLTIFVARLYYFWKPDYQQALRILCQKRG
GPHATMKIMNYNEFQDCWNKFVDGRGKPFKPRNLPKHHTLLQATLGELLRHLMDPGTFTSNFNKPKW
[0467] VSGQHETLYCYKVERLHNDTWVPLNQHRGFLRNQAPNIHGFPKGRHAELCFLDLIPFWKLDGQQYRV
CTFSWSPCFSCAQEMAKFISNNEHVSLCIFAARIYDDQGRYQEGLRALHRDGAKIAMMNYSEFEYCWDT
FVDRQGRPFQPWDGLDEHSQALSGLRLRAI (SEQ ID NO:67)

[0468] (粗斜体:核酸编辑域;划线:细胞质定位信号)

[0469] 黑猩猩 (Chimpanzee) APOBEC-3G:

MKPHFRNPVERMYQDTFSDNFYNRPILSHRNTVWLCYEVKTKGSPRPDLAKIFRGQVYSKLYHPEMR
FFHWFSKWRKLHRDQEYEVTWYISWSPCTKCTRVAFLAEDPKVTLTIFVARLYYFWDYQALRSL
CQKRDGPRATMKIMNYDEFQHCWSKFVYSQRELFEPWNNLPKYIILLHIMLGEILRHSMDPPTFTSNFN
[0470] NELVWRGRHETLYCYEVERLHNDTWVLLNQRRGFLCNQAPHKHGFLEGRHAELCFLDVIPFWKLDLHQ
DYRVTCFTSWSPCFSCAQEMAKFISNNKHVSLCIFAARIYDDQGRQCQGLRRLAKAGAKISIMTYSEFKH
CWDTFVDHQGCPFPWDGLEEHSQALSGLRLRAILQNQGN (SEQ ID NO:68)

[0471] (划线:核酸编辑域;双划线:细胞质定位信号)

[0472] 绿猴 (Green monkey) APOBEC-3G:

MNPQIRNMVEQMEPDIFVYYFNNRPILSGRNTVWLCYEVKTKDPSGPPDLANIFQGKLYPEAKDHPEMK
FLHWFRKWRQLHRDQEYEVTWYISWSPCTKCTRDMATFLAEDPKVTLTIFVARLYYFWDYQALRSL
LCQERGGPHATMKIMNYNEFQHCWNEFVDGQGKPFKPRKNLPKHHTLLHATLGELLRHVMMDPGTFTSNFN
[0473] FNNKPVWVSGQRETYLCYKVERSHNDTWVLLNQHRGFLRNQAPDRHGFPKGRHAELCFLDLIPFWKLD
QQYRVTCFTSWSPCFSCAQKMAKFISSNNKHVSLCIFAARIYDDQGRQCQGLRRLHRDGAKIAVMNYSEF
EYCWDTFVDRQGRPFQPWDGLDEHSQALSGLRLRAI (SEQ ID NO:69)

[0474] (划线:核酸编辑域;双划线:细胞质定位信号)

[0475] 人 APOBEC-3G:

MKPHFRNTVERMYRDTFSYNFYNRPILSRRNTVWLCYEVKTKGSPRPDLAKIFRGQVYSELKYHPEMR
FFHWFSKWRKLHRDQEYEVTWYISWSPCTKCTRDMATFLAEDPKVTLTIFVARLYYFWDYQALRSL
LCQKRDGPRATMKIMNYDEFQHCWSKFVYSQRELFEPWNNLPKYIILLHIMLGEILRHSMDPPTFTSNFN
[0476] NEPVWRGRHETLYCYEVERMHNDTWVLLNQRRGFLCNQAPHKHGFLEGRHAELCFLDVIPFWKLDLD
QDYRVTCFTSWSPCFSCAQEMAKFISNNKHVSLCIFTARIYDDQGRQCQGLRRLAEAGAKISIMTYSEFKH
CWDTFVDHQGCPFPWDGLDEHSQDLSGLRLRAILQNQEN (SEQ ID NO:70)

[0477] (划线:核酸编辑域;双划线:细胞质定位信号)

[0478] 人 APOBEC-3F:

MKPHFRNTVERMYRDTFSYNFYNRPILSRRNTVWLCYEVKTKGSPRPDLAKIFRGQVYSQPEHHAEMC
FLSWFCGNQLPAYKCFQITWFVSWTPCPDCVAKLAFLAEPNVTLTISAARLYYWERDYRRALCRSL
QAGARVKIMDDEFAFCWENFVYSEGQPFMPWYKFDNYAFLHRTLKEILRNPMMEAMYPHIFYHFNK
[0479] LRKAYGRNESWLCFTMEVVKHHSPVSWKRGVFRNQVDPETHCHAERCFLSWFCDDILSPNTNYEVTWY
TSWSPCECAGEVAEFLARHSNVNLTIFTARLYYFWDYQGLRSLSQEGASVEIMGYKDFKYCWENF
VYNDDEPFKPKWGLKYNFLFLDSKLQEILE (SEQ ID NO:71)

[0480] (划线:核酸编辑域)

[0481] 人 APOBEC-3B:

MNPQIRNPMERMYRDTFYDNFENEPILYGRSYTWLCYEVKIKRGRSNLLWDTGVFRGQVYFKPQYHAE
MCFLSWFCGNQLPAYKCFQITWFVSWTPCPDCVAKLAFLSEHPNVTLTISAARLYYWERDYRRALCR
LSQAGARVTIMDYEEFAFCWENFVYNEGQFMPWYKFDENY AFLHRTLKEILRYLMDPDTFTSNFNND
[0482] PLVLRRTYLYCYEVERLDNGTWVLMQHMGLCNEAKNLLCGFYGRHAELRFLDLVPSLQDPAQIY
RVTWFISWSPCFSWGCAGEVRAFLQENTHVRIRIFAARIYDYDPLYKEALQMLRDAGAQVSIMTYDEFE
YCWDTFVYRQGCPFPWDGLEEHSQALSGLRLRAILQNQGN (SEQ ID NO:72)

[0483] (划线:核酸编辑域)

[0484] 人 APOBEC-3C:

- [0485] MNPQIRNPMKAMYPGTFYFQFKNLWEANDRNETWLCFTVEGIKRRSVVSWKTGVFRNQVDSETHCHA
ERCFLSWFCDDILSPNTKYQVTWYTSWSPCPDCAGEVAEFLARHSNVNLTIFTARLYYFYQYPCYQEG
 LSQEGVAVEIMDYEDFKYCWENFVYNDNEPFKPKWGLKTNFRLLKRRLRESLQ (SEQ ID
 NO:73)
- [0486] (划线:核酸编辑域)
- [0487] 人APOBEC-3A:
- [0488] MEASPASGPRHLMDPHIFTSNFNNGIGRHKTYLCYEVEERLDNGTSVKMDQHRGFLHNQAKNLLCGFYG
RHAELRFLDLVPSLQLDPAQIYRVTWFIWSPCFSWGCAGEVRAFLQENTHVRLRIFAARIYDYDPLYKEALQMLR
 DAGAQVSIMTYDEFKHCWDTFVDHQGCPFQPWDGLDEHSQALSGRRLRAILQNQGN (SEQ ID NO:74)
- [0489] (划线:核酸编辑域)
- [0490] 人APOBEC-3H:
- [0491] MALLTAETFRLQFNKRRLRRPYPRKALLCYQLTPQNGSTPTRGYFENKKKCHAEICFINEIKSMGL
DETQCYQVTCYLTWSPCSSCAWELVDFIKAHDLNLGIFASRLYYHWCKPQQKGLRLLCGSQVPVEVMGFPKFADC
 WENFVDHEKPLSFNPYKMLEELDKNSRAIKRRLERIKIPGVRAQGRYMDILCDAEV (SEQ ID NO:75)
- [0492] (划线:核酸编辑域)
- [0493] 人APOBEC-3D:
- [0494] MNPQIRNPMERMYRDTFYDNFENEPILYGRSYTWLCYEVKIKRGRSNLLWDTGVFRGPVLPKRQSNHR
QEVYFRFENHAEMCFLSWFCGNRLPANRRFQITWVFSWNPCLPCVVKVTKFLAEHPNVTLTISAARLYYRDRDWR
WVLLRLHKAGARVKIMDYEDFAWCWENFVCNEGQPFMPWYKFDDNYASLHRTLKEILRNPMEAMYPHIFYFHFKNL
LKACGRNESWLCTMEVTKHSAVFRKRGVFRNQVDPETHCHAERCFLSWFCDDILSPNTNYEVTWYTSWSPCPEC
AGEVAEFLARHSNVNLTIFTARLCYFWDTDYQEGLCSLSQEGASVKIMGYKDFVSCWKNFVYSDDPEFPKPKGLQT
 NFRLKRLREILQ (SEQ ID NO:76)
- [0495] (划线:核酸编辑域)
- [0496] 人APOBEC-1:
- [0497] MTSEKGPSTGDPTLRRRIEPWEFDVFYDPRELKREACLLYEIKWGMSRKIWRSSGKNTTNHVEVNFIK
KFTSERDFHPSMCSITWFLSWSPCWECSQAIREFLSRHPGVTLVIYVARLFWHMDQQNRQGLRDLVNSGVTIQIM
RASEYYHCWRNFVNYPPGDEAHWPQYPLWMLYALELHCIILSLPPCLKISRRWQNHLTFFRLHLQNCHYQTIPP
 HILLATGLIHPSVAWR (SEQ ID NO:77)
- [0498] 小鼠APOBEC-1:
- [0499] MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELKRETCCLYEINWGRHSVWRHTSQNTSNHVEVNFLE
KFTTERYFRPNTRCSITWFLSWSPCGECSRAITEFLSRHPVTLFIYIARLYHHTDQRNRQGLRDLISSGVTIQIM
TEQEYCYCWRNFVNYPPSNEAYWPRYPHLWVKLYVLELYCIIILGLPPCLKILRRKQPQLTFFTITLQTCHYQRIPP
 HLLWATGLK (SEQ ID NO:78)
- [0500] 大鼠APOBEC-1:
- [0501] MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELKRETCCLYEINWGRHSIWRHTSQNTNKHVEVNFIE
KFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIM
TEQESGYCWRNFVNYSPSNEAHWPYPHLWVRLYVLELYCIIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPP
 HILWATGLK (SEQ ID NO:79)
- [0502] 人ADAT-2:
- [0503] MEAKAAPKPAASGACSVSAEETEKWMEEAMHMAKEALENTEVPVGCLMVYNNEVVGKGRNEVNQTKNA

TRHAEMVAIDQVLDWCRQSGKSPSEVFEHTVLVYTVTEPCIMCAAALRLMKIPLVVYGCQNERFGGCGSVLNIASAD
LPNTGRPFQCIPGYRAEEAVEMLKTFYKQENPNAPKSKVRKKECQKS (SEQ ID NO:80)

[0504] 小鼠ADAT-2:

[0505] MEEKVESTTTPDGPCVVSQETEKWMEEAMRMAKEALENIEVPVGCMLVYNNEVVGKGRNEVNQTKNA
TRHAEMVAIDQVLDWCHQHGGQSPSTVFEHTVLVYTVTEPCIMCAAALRLMKIPLVVYGCQNERFGGCGSVLNIASAD
LPNTGRPFQCIPGYRAEEAVELLKTFYKQENPNAPKSKVRKKECQKS (SEQ ID NO:81)

[0506] 小鼠ADAT-1:

[0507] MWTADEIAQLCYAHYNVRLPKQKGPEPNREWTLLAAVVKIQASANQACDIPEKEVQVTKEVSMGTGT
KCIGQSKMRESGDILNDSHAEIIARRSFQRYLLHQLHLAAVLKEDSIFVPGTQRGLWRLRPDLFSVFFSSHTPCGD
ASIIIPMLEFEEQPCCPVIRSWANNSPVQETENLEDSKDKRNCEDPASPVAKKMRLGTPARSLSNCAHHGTQESGP
VKPDVSSDLTKEEPDAANGIASGSFRVVDVYRTGAKCVPGETGDLREPGAAYHQVGLLRVKPGRGDRTCMSMCS
KMARWNVLGCQGALLMHFLEKPIYLSAVVIGKCPYSQEAMRRALTGRCEETLVLPRGFGVQELEIQQSGLLFEQSR
CAVHRKRGDSPGRLVPCGAAISWSAVPQQPLDVTANGFPQGTTKKEIGSPRARSRISKVELFRSFQKLLSSIADE
QPDSIRVTKKLDITYQEYKDAASAYQEAWGALRRIQPFASWIRNPPDYHQFK (SEQ ID NO:82) (划线:核酸编
辑域)

[0508] 人ADAT-1:

[0509] MWTADEIAQLCYEHYGI RLPKKGKPEPNHEWTLLAAVVKIQSPADKACDTPDKPVQVTKEVSMGTGT
KCIGQSKMRKNGDILNDSHAEVIARRSFQRYLLHQLQLAATLKEDSIFVPGTQKGVWKLRRDLIFVFFSSHTPCGD
ASIIIPMLEFEDQPCCPVFRNWAHNSSEASSNLEAPGNERKCEDPDSVTKMRLEPGTAAREVTNGAAHHSFGK
QKSGPISPGIHSCDLTVEGLATVTRIAPGSAKVIDVYRTGAKCVPGEAGDSGKPGA AFHQVGLLRVKPGRGDRTRS
MSCSDKMARWNVLGCQGALLMHLLEPIYLSAVVIGKCPYSQEAMQRALIGRCQNVSALPKGFGVQELKILQSDLL
FEQSRSAVQAKRADSPGRLVPCGAAISWSAVPEQPLDVTANGFPQGTTKKTIGSLQARSQISKVELFRSFQKLLSR
IARDKWPHSLRVQKLDITYQEYKEAASSYQEAWSTLRKQVFGSWIRNPPDYHQFK (SEQ ID NO:83) (划线:核
酸编辑域)

[0510] 在一些实施方案中,所述酶域包含一个或多个转录激活物。例如,在一些实施方案中,具有转录激活物域的例示性dCas9融合蛋白的一般体系结构包含以下结构:

[0511] $[\text{NH}_2] - [\text{NLS}] - [\text{Cas9}] - [(\text{转录激活物})_n] - [\text{COOH}]$,

[0512] $[\text{NH}_2] - [\text{NLS}] - [(\text{转录激活物})_n] - [\text{Cas9}] - [\text{COOH}]$,

[0513] $[\text{NH}_2] - [\text{Cas9}] - [(\text{转录激活物})_n] - [\text{COOH}]$, 或

[0514] $[\text{NH}_2] - [(\text{转录激活物})_n] - [\text{Cas9}] - [\text{COOH}]$;

[0515] 其中NLS是核定位信号, NH_2 是融合蛋白的N末端,且COOH是融合蛋白的C末端。在一些实施方案中,融合蛋白包含一个或多个转录激活物重复,例如其中 $n=1-10$ (例如 n 是1,2,3,4,5,6,7,8,9或10)。在一些实施方案中, $n=1-20$ 。在一些实施方案中,接头插入在dCas9和转录激活物域之间。在一些实施方案中,接头插入在NLS与转录激活物和/或dCas9域之间。在一些实施方案中,NLS位于转录激活物和/或dCas9域的C末端。在一些实施方案中,NLS位于转录激活物域和dCas9域之间。还可以存在另外的特征,如序列标签。在一些实施方案中,转录激活物选自下组:VP64, (SEQ ID NO:84或SEQ ID NO:35),VP16 (SEQ ID NO:85)和p65 (SEQ ID NO:86)。在一些实施方案中,dCas9-VP64融合蛋白包含SEQ ID NO:87的氨基酸序列(例如有或无6xHis标签),或包含与SEQ ID NO:87的氨基酸序列至少80%,至少85%,

至少90%，至少95%，至少98%或至少99%相同的氨基酸序列(例如有或无6xHis标签)。

[0516] VP64GSGRADALDDFDLMLGSDALDDFDLMLGSDALDDFDLMLGSDALDDFDLMLIN (SEQ ID NO:84)

[0517] VP16

[0518] APPTDVSLGDELHLDGEDVAMAHADALDDFDLMLGSDGSPGPGFTPHDSAPYGALDMADFEFEQMFT DALGIDEYGGFPGIRR (SEQ ID NO:85)

[0519] p65:

[0520] PSGQISNQALALAPSSAPVLAQTMVPSSAMVPLAQQPAPAPVLTGPPQSLSAPVPKSTQAGEGTLSE ALLHLQFDAEDL GALLGNSTDPGVFTDLASVDNSEFQQLLNQGVSMHSTAEPMLMEYPEAITRLVTGSQRPPDP APTPLGTSGLPNGLSGDEDFSSIADMDFSALLSQISSSQ (SEQ ID NO:86)

[0521] dCas9-VP64-6xHis:

[0522] MDKKYSIGLAIGTNSVGWAVITDEYKVPKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTA RRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERHP IFGNIVDEVAYHEKYPTIYHLRKKLVD STDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSR RLENLIAQLPGEKKNLFGNLIALSLGLTPNFKSNFDLAEDAQLSKDQYDDDLNLLAQIGDQYADLFLAAKNL SDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF YKFIKPILEKMDGTEELLVKNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRI PYYVGPLARGNSRFAMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPHSLLYEYFTVYNEL TKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKI IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSG KTILDFLKSDFANRNFMLIHDDSLTFKEDIQKAQVSGQGSLSHEHIANLAGSPAIKKGILQTVKVVDELVKVMG RHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGKELGSQILKEHPVENTQLQNEKLYLYLQNGRDMYVDQEL DINRLSDYDVDAIVPQSFLKDDSIDNKVLRSDKNRKGSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAE RGGSELKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVI TLKSKLVSDFRKDFQFYKVVREINN YHHAHDAYLNAVVG TALIKKYPKLESEFVYGDYKVDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGE IRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNI VKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYG GFDSPTVAYSVLVVAKEVGKSKLKSVELLGITIMERSSEFEKNPIDFLEAKGYKEVKKDLI IKLPKYSLFELEN GRKRMLASAGELQKGNELALPSKYVNFYLASHYEKLGKSPEDNEQKQLFVEQHKHYLDEIEEQISEFSKRVLAD ANLDKVL SAYNKHRDKPIREQAENI IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSI TGLYETRI DLSQLGGDGSPKKRKVSSDYKDHDGDYKDHDIDYKDDDDKAAGGGGSGRADALDDFDLMLGSDALDDFDLMLG SDALDDFDLMLGSDALDDFDLMLHHHHH (SEQ ID NO:87)

[0523] 在一些实施方案中,所述酶域包含一个或多个转录阻遏物。例如,在一些实施方案中,具有转录阻遏物的例示性dCas9融合蛋白的一般体系结构包含以下结构:

[0524] $[\text{NH}_2] - [\text{NLS}] - [\text{Cas9}] - [(\text{转录阻遏物})_n] - [\text{COOH}]$,

[0525] $[\text{NH}_2] - [\text{NLS}] - [(\text{转录阻遏物})_n] - [\text{Cas9}] - [\text{COOH}]$,

[0526] $[\text{NH}_2] - [\text{Cas9}] - [(\text{转录阻遏物})_n] - [\text{COOH}]$, 或

[0527] $[\text{NH}_2] - [(\text{转录阻遏物})_n] - [\text{Cas9}] - [\text{COOH}]$;

[0528] 其中NLS是核定位信号, NH_2 是融合蛋白的N末端,且COOH是融合蛋白的C末端。在一些实施方案中,融合蛋白包含一个或多个转录阻遏物重复,例如其中 $n=1-10$ (例如 n 是1,2,

3,4,5,6,7,8,9或10)。在一些实施方案中,n=1-20。在一些实施方案中,接头插入在dCas9和转录阻遏物域之间。在一些实施方案中,接头插入在NLS与转录阻遏物和/或dCas9域之间。在一些实施方案中,NLS位于转录阻遏物和/或dCas9域的C末端。在一些实施方案中,NLS位于转录阻遏物域和dCas9域之间。还可以存在另外的特征,如序列标签。在一些实施方案中,the转录阻遏物选自下组:Kox1的KRAB (Krüppel 关联框 (associated box)) 域、SID (mSin3相互作用域)、HP1 α 的CS (Chromo Shadow) 域、或Hes1的WRPW域。这些和其他阻遏物域是本领域已知的,且在一些实施方案中对应于Urrutia,KRAB-containing zinc-finger repressor proteins.Genome Biol.2003;4(10):231;Gilbert等CRISPR mediated modular RNA-guided regulation of transcription in eukaryotes.Cell.2013;154,442-451;Konermann等,Optical control of mammalian endogenous transcription and epigenetic states.Nature.2013;500,472-476;和公开的美国专利申请U.S.S.N.14/105,017(其公开为U.S.2014/0186958A1)中描述的那些,其完整内容通过提述并入本文。在一些实施方案中,所述转录阻遏物域包含KRAB域的一个或多个重复(例如2,3,4,5,6,7,8,9或10个重复)。在一些实施方案中,KRAB域包含选自下组的氨基酸序列:SEQ ID NOs:88-91。在一些实施方案中,所述转录阻遏物域包含SID蛋白的一个或多个重复。在一些实施方案中,所述SID蛋白包含如SEQ ID NO:80中列出的氨基酸序列。在一些实施方案中,所述阻遏物域包含SID蛋白的2,3,4,5,6,7,8,9或10个重复(例如SEQ ID NO:92)。在一些实施方案中,所述阻遏物域包含SID的4个重复(例如SID4x;SEQ ID NO:93)。

[0529] KRAB(人;GenBank:AAD20972.1)

[0530] MNMFKEAVTFKDVAVAFTEELGLLGPAQRKLYRDVMVENFRNLLSVGHPPFKQDVSPIERNEQLWIM TTATRRQGNLDTLPVKALLLYDLAQT (SEQ ID NO:88)

[0531] KRAB蛋白域,部分(人;GenBank:CAB52478.1):

[0532] EQVSFKDVCVDFTQEEWYLLDPAQKILYRDVILENYSNLVSVGYCITKPEVIFKIEQGEEPWILEKGF PSQCHP (SEQ ID NO:89)

[0533] KRAB A域,部分(人;GenBank:AAB03530.1):

[0534] EAVTFKDVAVVFTEELGLLDPAQRKLYRDVMLENFRNLLSV (SEQ ID NO:90)

[0535] KRAB(小鼠;含C2H2型域的蛋白;GenBank:CAM27971.1):

[0536] MDLVTYDDVHVNFQDEWALLDPSQKSLYKGVMLETYKNLTAIGYIWEHTIEDHFQTSRSHGSNKKTH (SEQ ID NO:91)

[0537] SID阻遏物域:

[0538] GSGMNIQMLLEAADYLERREREAHEGYASMLP (SEQ ID NO:92)

[0539] SID4x阻遏物域:

[0540] GSGMNIQMLLEAADYLERREREAHEGYASMLPGSGMNIQMLLEAADYLERREREAHEGYASMLPGSGM NIQMLLEAADYLERREREAHEGYASMLPGSGMNIQMLLEAADYLERREREAHEGYASMLPSR (SEQ ID NO:93)

[0541] 在一些实施方案中,所述酶域包含表观遗传修饰物或其催化域。例如,在一些实施方案中,具有表观遗传修饰物或域的例示性dCas9融合蛋白的一般体系结构包含以下结构:

[0542] [NH₂]-[NLS]-[Cas9]-[表观遗传修饰物]-[COOH],

[0543] [NH₂]-[NLS]-[表观遗传修饰物]-[Cas9]-[COOH],

[0544] [NH₂]-[Cas9]-[表观遗传修饰物]-[COOH],或

[0545] [NH₂]-[表观遗传修饰物]-[Cas9]-[COOH];

[0546] 其中NLS是核定位信号,NH₂是融合蛋白的N末端,且COOH是融合蛋白的C末端。在一些实施方案中,接头插入在dCas9和表观遗传修饰物域之间。在一些实施方案中,接头插入在NLS与表观遗传修饰物和/或dCas9域之间。在一些实施方案中,NLS位于表观遗传修饰物和/或dCas9域的C末端。在一些实施方案中,NLS位于表观遗传修饰物域和dCas9域之间。还可以存在另外的特征,如序列标签。表观遗传修饰物是本领域公知的,且通常催化DNA甲基化(和脱甲基化)或组蛋白修饰(例如组蛋白甲基化/脱甲基化,乙酰化/去乙酰化,泛素化,磷酸化,类泛素化(sumoylation)等等)。一种或多种表观遗传修饰的存在可以影响一种或多种基因的转录活性,例如将基因从“打开”状态调为“关闭”状态,反之亦然。表观遗传修饰物包括但不限于,组蛋白脱甲基酶、组蛋白甲基转移酶、羟化酶、组蛋白脱乙酰基酶和组蛋白乙酰转移酶。例示性的表观遗传修饰蛋白可见于Koner mann等,Optical control of mammalian endogenous transcription and epigenetic states.Nature.2013;500,472-476;Mendenhall等,Locus-specific editing of histone modifications at endogenous enhancers.Nat.Biotechnol.2013;31,1133-1136;和Maeder等,Targeted DNA demethylation and activation of endogenous genes using programmable TALE-TET1 fusion proteins.Nat.Biotechnol.2013;31,1137-1142;其各自完整内容通过提述并入本文。在一些实施方案中,所述表观遗传修饰物域是LSD1(赖氨酸(K)特异性脱甲基酶1A)组蛋白脱甲基酶,其在一些实施方案中全部或部分地包含如SEQ ID NO:94或SEQ ID NO:95中所列的氨基酸序列。在一些实施方案中,所述表观遗传修饰物域是TET1羟化酶催化域,其在一些实施方案中,包含如SEQ ID NO:96中所列的氨基酸序列。在一些实施方案中,所述表观遗传修饰物是组蛋白去乙酰基酶(HDAC)效应子域。在一些实施方案中,所述HDAC效应子域全部或部分地包含对应于如Koner mann等,Optical control of mammalian endogenous transcription and epigenetic states.Nature.2013;500,472-476;SEQ ID NOs:97-108的附录表2中提供的任意HDAC效应器蛋白的氨基酸序列。在一些实施方案中,所述表观遗传修饰物是组蛋白甲基转移酶(HMT)效应子域。在一些实施方案中,所述HMT效应子域全部或部分地包含对应于如Koner mann等,Optical control of mammalian endogenous transcription and epigenetic states.Nature.2013;500,472-476;SEQ ID NOs:97-108的附录表3中提供的任意HDAC效应器蛋白的氨基酸序列。

[0547] LSD1,同等型a(人):

[0548] MLSGKKAATAAAAAAAAAATGTEAGPGTAGGSENGSEVAAQPAGLSGPAEAVGPGAVGERTPRKKEPPRA
SPPGGLAEPGSGAGPQAGPTVVPGSATPMETGIAETPEGRRTSRRKRAKVEYREMDLANLSEDEYYSEEERNAK
AEKEKKLPPPPPQAPPEEENESEPEEPSGQAGGLQDDSSGGYGDGQASGVEGAFAFQSRLPHDRMSTQEAACFPDII
SGPQQTQKVFLFIRNRTLQLWLDNPKIQLTFEATLQQLEAPYNSDTVLVHRVHSYLERHGLINFGIYKRIKPLPTK
KTGKVIIGSGVSGLAAARQLQSFQMDVTLLIARDRVGGRVATFRKGNVYADLGAMVVTGLGPNMAVVSQVNM
LAKIKQKCPLYEANGQADTVKVPKEKDEMVEQEFNRLLEATSYLSHQDFNVLNKPVSLGQALEVVIQLQEKHV
KDEQIEHWKKIVKTQEELKELLNKMNKELKELHQQYKEASEVKPPRDITAEFLVKS KHRDLTALCKEYDELAET
QGKLEEKLQLEANPPSDVYLSSRDRQILDWHFANLEFANATPLSTLSLKHWDQDDDFEFTGSHLTVRNGYSCVPV
ALAEGLDIKLNTAVRQVRYTASGCEVIAVNTRSTSQTFIYKCAVLCCTPLGVLKQQPPAVQFVPLPEWKTS
SAVQRMGFGNLNKVVLCFDRVFWDPVNLFGHVGSTTASRGELFLFWNLYKAPILLALVAGEAAGIMENISDDVIVGRCL

AILKGFSSAVPQPKETVVSRRADPWARGSYSYVAAGSSGNDYDLMAQPIITPGPSIPGAPQPIPRLFFAGEHTI
RNYPATVHGALLSGLREAGRIADQFLGAMYTLPRQATPGVPAQQSPSM (SEQ ID NO:94)

[0549] LSD1, 同等型b (人):

[0550] MLSGKKAATAAATGTEAGPGTAGGSENGSEVAAQPAGLSGPAEVGPGAVGERTPRKKEPPRA
SPPGGLAEPGSAGPQAGPTVVPGSATPMETGIAETPEGRRTSRRKRAKVEYREMDLANLSEDEYYSEEERNAK
AEKEKKLPPPPQAPPEEENESEPEEPSGVEGAAFQSRLPHDRMTSQEAACFPDIISGPQQTQKVFLFIRNRTLQL
WLDNPKIQLTFEATLQQLEAPYNSDTVLVHRVHSYLERHGLINFGIYKRIKPLPTKKTGKVIIGSGVSGLAAARQ
LQSFQMDVTLLEARDRVGGRVATFRKGNVADLGAMVVTGLGGNPMVAVVSKQVNMELAKIKQKCLYEANGQAVPK
EKDEMVEQEFNRLLEATSLSHQLDFNVLNNKPVSLGQALEVVIQLQEKHKVDEQIEHWKKIVKTQEELKELLNKM
VNLKEKIKELHQQYKEASEVKPPRDIATAEFLVSKHRDLTALCKEYDELAETQGKLEEKLELEANPPSDVYLSSR
DRQILDWHFANLEFANATPLSTLSLKHWDQDDDFEFTGSHLTVRNGYSCVPVALAEGLDIKLNTAVRQVRYTASGC
EVIAVNTRSTSQTFIYKCAVLCTPLGLVKKQPPAVQFVPLPEWKTSAVQRMGFGNLKVVLCFDRVFWDPVSN
LFGHVSTTASRGELFLFWNLKAPILLALVAGEAAGIMENISDDVIVGRCLAILKGFSSAVPQPKETVVSRR
ADPWARGSYSYVAAGSSGNDYDLMAQPIITPGPSIPGAPQPIPRLFFAGEHTIRNYPATVHGALLSGLREAGRIADQ
FLGAMYTLPRQATPGVPAQQSPSM (SEQ ID NO:95)

[0551] TET1 催化域:

[0552] SIVAQLSRPDPALAAALNDHLVALACLGGPALDAVKKGLPHAPALIKRTNRRIPERTSHRVADHAQV
VRVLGFFQCHSHPAQAFDDAMTQFGMSGGSLPTCSCLDRVIQKDKGPYYTHLGAGPSVAAVREIMENRYGQKGN
IRIEIVVYTGKEGKSSHGCPKIAKWVLRSSDEEKVLCVLRQRTGHHCTAVMVVIMVWDGIPLMADRLYTEL
NLKSYNGHPTDRRCTLNENRTCTCQIDPETCGASFSFGCSWSMYFNGCKFGRSPSPRRFRIDPSSPLHEKNLEDN
LQSLATRLAPIYKQYAPVAYQNQVEYENVARECRLGSKEGRPFSGVTACLDFCAHPRDIHNMNGSTVVCTLTRE
DNRS LGVIPQDEQLHVLPLYKLSDTDEFSGKEGMEAKIKSGAIEVLAPRRKKRTCFQPVPRSGKKRAAMTEVLA
HKIRAVEKKPIPRIKRKNSTTTNNSKPSSLPTLGSNTETVQPEVKSETEPHFILKSSDNTKTYSLMPSAPHPVKE
ASPGFSWSPKTASATPAPLKN DATASCGFSERSSTPHCTMPSGRLSGANAAAADGPGISQLGEVAPLPTLSAPVME
PLINSEPTGVTEPLTPHQPNHQPSFLTSPQDLASSPMEEDEQHSEADEPPSDEPLSDDPLSPAEEKLPHIDEYWS
DSEHIFLDANIGGVAIAPA HGSVLI ECARRELHATTPVEHPNRNHPTRL SLVIFYQHKNLNKPKQHGFELNKIKFEAK
EAKNKKMKASEQKDQAANEGPEQSSEVNELNQP SHKALTLTHDNVTVSPYALTHVAGPYNHV (SEQ ID NO:
96)

[0553] HDAC 效应子域:

[0554] HDAC8 (光滑爪蟾 (*X. laevis*)):

[0555] ASSPKKKRVEASMSRVVVKVASMEEAAFH TDAYLQHLHKVSEEGDND DPETLEYGLGYDCPITEG
IYDYAAA VGGATLTAAEQ LIEGKTRIAVNWPGGWHHAKKDEASGFCYLND AVLGILKLREKFDRVLYVDMDLHHGD
GVEDAFSFTSKVMTVSLHKFSPGFFPGTGDVSDIGLGKGRYYSINVPLQDGIQDDKYYQICEGLKEVFTTFNPEA
VVLQLGADTIAGDPMCSFNMTPEGIGKCLKYVLQWLPTLILGGGGYHLPNTARCWYLTALIVGRTLSSEIPDHE
FFTEYGPDYVLEITPSCR PDRNDTQKVQEILQSIKGNLKR VVEF (SEQ ID NO:97)

[0556] RPD3 (酿酒酵母):

[0557] ASSPKKKRVEASRRVAYFYDADVGNAYGAGHPMKPHRIRMAHSLIMNYGLYKKMEIYRAK PATKQE
MCQFHTDEYIDFLSRVTPDNLEMFKRESVKNVGD DCPVFDGLYEYCSISGGGSMEGAARLN RGKCDVAVNYAGGL
HHAKKSEASGFCYLNDIVLGI IELLYHPRVLYIDIDVHHGDGVVEAFYTTDRVMTCSFH KYGEFFPGTGELRDIG

VGAGKNYAVNVPLRDGIDDATYRSVFEPVIKKIMEWYQPSAVVLQCGGDSLSGDRLGCFNLSMEGHANCVNYVKSFGIPMMVVGGGGYTMRNVARTWCFETGLLNNVLDKDLPEYF (SEQ ID NO:98)

[0558] MesoLo4 (*M. loti*):

[0559] ASSPKKKRKVEASMLQIVHHPDYDAGFATNHRFPMSKYPLLMEALRARGLASPDALNTTEPAPASWLKLAHAADYVDQVIVSCSVPEKIEREIGFPVGPVSLRAQLATGGTILAARLALRHGIACNTAGGSHHARRAQGAGFC TFNDVAVASLVLLDEGAAQNILVVDLDVHQGDGTADILSDEPGVFTFSMHGERNYPVRKIASDLDIAPDGTGDAA YLRRLATILPELSARARWDIVFYNAGVDVHAEDRLGRLALSNGGLRARDEMVIHGFALGIPVCGVIGGGYSTDVP ALASRHAILFEVASTYAEF (SEQ ID NO:99)

[0560] HDAC11 (人):

[0561] ASSPKKKRKVEASMLHTTQLYQHVPETRWPIVYSPRYNITFMGLEKLHPFDAGKWGKVINFLKEEKLLSDSMLVEAREASEEDLLVVHTRRYLNELKWSFAVATITEIPPVIFLPNFLVQRKVLRLPLRTQTGGTIMAGKLAVER GWAINVGGGFHHCSSDRGGGFCAYADITLAIKFLFERVEGISRATIIDLDAHQNGHERDFMDDKRVYIMDVYNRH IYPGDRFAKQAIRRKVELEWGTEDDEYLDKVERNIKKSLQEHLDPVVVYNAGTDILEGDRLGGLSISPAGIVKRDE LVFRMVRGRRVPILMVTSGGYQKRTARIADSILNLFGLGLIGPESPSVSAQNSDTPLPPAVPEF (SEQ ID NO: 100)

[0562] HDT1 (拟南芥 (*A. thaliana*)):

[0563] ASSPKKKRKVEASMEFWGIEVKSQKPVTVTPEEGILIHVSQASLGECKNKKGEFVPLHVKVGNNQNLVLTSTENIPQLFCDLVFDKEFELSHTWGKGSVYFVGKTPNIEPQGYSEEEEEEEEEVPAGNAKAVAKPKAKPAE VKPAVDDEEDESDDSDGMEDESDGEDSEEEPTPKKPASSKKRANETTPKAPVSAKAKAVAVTPQKTDEKKKGGKA ANQSEF (SEQ ID NO:101)

[0564] SIRT3 (人):

[0565] ASSPKKKRKVEASMGAGISTPSGIPDFRSPGSLYSNLQQYDLPYPEAIFELPFFFHNPKPFFTLAK ELYPGNYKPNVTHYFLRLLHDKGLLLRLYTQNIIDGLERVSGIPASKLVEAHGTFASATCTVCQRPFGEDIRADVM ADRVPRCPVCTGVVVKPDIVFFGEPLPQRFLHVVDFPMADLLLILGTSLEVEPFASLTEAVRSSVPRLLINRDLVG PLAWHPRS RDVAQLGDVVHGVESLVELLWGTEEMRDLVQRETGKLDGPDKEF (SEQ ID NO:102)

[0566] HST2 (酿酒酵母):

[0567] ASSPKKKRKVEASTEMSVRKIAAHMKSNNPAKVIFMVGAGISTSCGIPDFRSPGTGLYHNLARLKLPHY PEAVFDVDFQSDPLPFYTLAKELYPGNFRPSKFHYLLKLFQDKDVLKRVYTQNIIDTLERQAGVKDDLIIEAHGSF AHCHCIGCGKVYPQVFKSKLAEHPKDFVKCDVCGELVKPAIVFFGEDLPDSFSETWLNSEWLREKITTSGKHP QQPLVIVVGTSLAVYPFASLPEEIPRKVKRVLNLETVGDVFKANKRPTDLIVHQVSDEFQAEQLVEELGWQEDFEKI LTAQGGMGEF (SEQ ID NO:103)

[0568] CobB (大肠杆菌 (K12)):

[0569] ASSPKKKRKVEASMEKPRVLTGAGISAESGIRTFRAADGLWEEHRVEDVATPEGFDRDPELVQAFY NARRRQLQQPEIQPNAAHLALAKLQDALGDRFLLVTQNIIDNLHERAGNTNVIHMHGELLKVRCSQSGQVLDWTGDV TPEDKCHCCQFPAPLRPHVWFGEPLGMDEIYMALSMADIFIAIGTSGHVYPAAGFVHEAKLHGAHTVELNLEPS QVGNEFAEKYYGPASQVPEFVEKLLKGLKAGSIAEF (SEQ ID NO:104)

[0570] HST2 (白色念珠菌 (*C. albicans*)):

[0571] ASSPKKKRKVEASMPSLDDILKPVAEAVKNGKKVTFNAGISTGAGIPDFRSPDTGLYANLAKLNLP FAEAVFDIDFFKEDPKPFYTLAEELYPGNFAPTKFHFIKLLQDQGS�KRVYTQNIIDTLERLAGVEDKYIVEAHGS

FASNHCVDCHKEMTTETLKYMKDKKIPSCQHCEGYVKPDIVFFGEGLPVKFFDLWEDDCEDVEVAIVAGTSLTVF
PFASLPGEVNNKCLRVLNKEKVGTFKHEPRKSDIIALHDCDIVAERLCTLLGLDDKLNEVYEKEKIKYKAETKE
IKMHEIEDKLKEEAHLKEDKHTTKVDKKEKQNDANDKELEQLIDKAKAEF (SEQ ID NO:105)

[0572] SIRT5(人):

[0573] ASSPKKKRKVEASSSSMADFRKFFAKAKHIVIISGAGVSAESGVPTFRGAGGYWRKWQAQDLATPLAF
AHNPSRVWEFYHYRREVMSKEPNAGHRAIAECETRLGKQGRVVVITQNIDELHRKAGTKNLEIHGSLFKTRCT
SCGVVAENYKSPICPALSGKGAPEPGTQDASIPVEKLPRCEEAGCGLLRPHVWVWFGENLDPAILLEVDRELAHCD
LCLVVGTSVVYPAAMFAPQVAARGVPVAEFNTETTPATNRFHFHQGPCGTTLPEALACHENETVSEF (SEQ ID
NO:106)

[0574] Sir2A(恶性疟原虫(*P.falciparum*)):

[0575] ASSPKKKRKVEASMGNLMISFLKQDTSITLLELAKIIKKCKHVVALTGSSTSAESNIPSRGSSNSI
WSKYDPRIYGTIWFWKYPEKIWEVIRDISSDYEIEINNGHVALSTLESGLYKSVVTQNVLDGLHEASGNTKVISL
HGNVFEAVCCTCNKIVKLNKIMLQKTSHFHQLPPECPGGIFKPNILFGEVVSSDLLKEAEEEEIAKCDLLLIVIG
TSSTVSTATNLCHFACKKKKKIVEINISKTYITNKMSDYHVCAKFSELTQVANILKGSSEKNKKIMEF (SEQ ID
NO:107)

[0576] SIRT6(人):

[0577] ASSPKKKRKVEASMSVNYAAGLSPYADKGGKGLPEIFDPPEELERKVVWELARLVWQSSSVVFTGAGI
STASGIPDFRPHGVWVMEERGLAPKFDTFESARPTQTHMALVQLERVGLLRFLVSQNVLDGLHVRSGFPRDKLAE
LHGNMFVEECAKCKTQYVRDVTVMGLKATGRLCTVAKARGLRACRGELRDTILDWEDSLPDRDLALADEASRNA
DLSITLGTSLQIRPSGNLPLATKRRGGRLVIVNLQPTKHDRHADLRHGYVDEVMTRLMKHLGLEIPAWDGPVLE
RALPPLEF (SEQ ID NO:108)

[0578] HMT效应子域:

[0579] NUE(*C.trachomatis*):

[0580] ASSPKKKRKVEASMTTNSTQDTLYLSLHGGIDSAIPYPVRRVEQLLQFSFLPELQFQNAAVKQRIQRL
CYREEKRLAVSSLAKWLGQLHKQRLRAPKNPPVAICWINSYVGYGVFARESIPAWSYIGEYTGILRRRQALWLDEN
DYCFRYPVPRYSFRYFTIDSGMQGNVTRFINHSDNPNEAIGAFENGIFHIIIRA IKDILPGEELCYHYGPLYWKH
RKKREEFVPQEEEF (SEQ ID NO:109)

[0581] vSET(*P.bursaria chlorella virus*):

[0582] ASSPKKKRKVEASMFNDRVIVKKSPLGGYGVFARKSFEKGEELVEECLCIVRHNDWGTALDYLF SRK
NMSAMALGFGAIFNHSKDPNARHELTAGLKRMRIFTIKPIAIGEEITISYGDDYWLSRPRLTQNEF (SEQ ID NO:
110)

[0583] SUV39H1(人):

[0584] ASSPKKKRKVEASNLKCVRIKQFHKDLERELRRHRSKTPRHLDPSLANYLQKAKQRRALRRWEQ
ELNAKRSHLGRITVENEVDLDGPPRAFVYINEYRVGEGITLNQVAVGCECQDCLWAPTGGCCPGASLHKFAYNDQG
QVRLRAGLPYECNSRCRCGYDCPNRVVQKGI RYDLCIFRTDDGRGWGVRTLEKIRKNSFVMEYVGEIITSEEAE
RGQIYDRQGATYLFDLDYVEDVYTVDAAYYGNISHFVNHSCDPNLQVYNVFIDNLDERLPRIAFFATRTIRAGEEL
TFDYNMQVDPVDMESTRMSNFGLAGLPGSPKKRVRIECKCGTESCRKYLFEF (SEQ ID NO:111)

[0585] DIM5(*N.crassa*):

[0586] ASSPKKKRKVEASMEKAFRPHFFNHGKPDANPKEKKNCHWCQIRSFATHAQLPISIVNREDDAFLNPN

FRFIDHSIIIGKNVPVADQSFVRVGCSCASDEECMYSTCQCLDEMAPDSDEEADPYTRKKRFAYYSQGAKKGLLRDRV
LQSQEPIYECHQGCACSKDCPNRVVERGRTPVPLQIFRTKDRGWGKCPVNIKRGQFVDRYLGEIITSEADRRAE
STIARRKDVYLFALDKFSDPDSLDPPLLAGQPLEVDGEYMSGPTRFINHSCDPNMAIFARVGDHADKHIHDLALFAI
KDIPKGTTELTFDYVNGLTGLESDAHPSKISEMTKCLCGTAKCRGYLWEF (SEQ ID NO:112)

[0587] KYP (拟南芥):

[0588] ASSPKKKRKVEASDISGGLEFKGIPATNRVDDSPVSPTSFTYIKSLIIEPNVIIPKSSTGCNCRGSC
TDSKKCACAKLNGGNFPYVDLNDGRLIESRDVVFECGPHCGGPKCVNRTSQKRLRFNLEVFSAKKGWAVRSWEY
IPAGSPVCEYIGVVRRTADVDTISDNEYIFEIDCQQTMQGLGGRQRRLRDVAVPMNNGVSQSSEDENAPEFCIDAG
STGNFARFINHSCEPNLFVQCVLSSHQDIRLARVVLFAADNISPMQELTYDYGALDSVHEF (SEQ ID NO:113)

[0589] SUV4 (拟南芥):

[0590] ASSPKKKRKVEASQSAYLHVSLARISDEDCCANCKGNCLSAFDPCTCARETSGEYAYTKEGLLKEKFL
DTCLKMKKEPDSFPKVVYCKDCPLERDHDKGTYGKCDGHLIRKFIKECWKRCGCDMQCGNRVVQGRIRCQLQVYFTQ
EGKGWGLRTLQDLPKGTFICEYIGEILTNTELYDRNVRSSSERHTYPVTLADADWGSEKDLKDEEALCLDATICGNV
ARFINHRCEDANMIDIPIEIEIETPDRHYHYHIAFFTLRDVKAMDELTDYIMDFNDKSHPVKAFRCCCGSESCRDRKI
KGSQKSIERRKIVSAKKQGSKEVSKKRKEF (SEQ ID NO:114)

[0591] Set4 (秀丽线虫):

[0592] ASSPKKKRKVEASMLHEQIANISVTFNDIPRSDHSMTPTLFCYDDFATTLVDSVLNFTTHKMSKK
RRYLYQDEYRTARTVMKTFREQRDWTNAIYGLLTLRSVSHFLSKLPPNKLFEFRDHI VRFLNMFILDSGYTIQECK
RYSQEGHQGAKLVSTGVWSRGDKIERLSGVVCLLSEDEDSILAQEGSDFSVMYSTRKRCSTLWLGPGAYINHDCR
PTCEFVSHGSTAHIRVLRDMVPGDEITCFYGSEFFGPNNDCECCTCEKNMNGAFSYLRGNENAEP I ISEKKTKEYE
LRSRSEF (SEQ ID NO:115)

[0593] Set1 (秀丽线虫):

[0594] ASSPKKKRKVEASMKVAAKKLATSRMRKDRAAAASPSSDIENSENPSSLASHSSSSGRMTPSKNTRSR
KGVSVKDVSNHKITEFFQVRRSNRKTQSKISDEAKHALRDTVLKGTNERLLEVYKDVVKGGRIRTKVNFEEKGFV
EYRGVMMYSEAKVIEEQSYNDEEIGSYMYFEHNNKWCIDATKESPWKGRLINHSVLRPNLKTQVVEIDGSHHL
ILVARRQIAQGEELLYDYGDRSAETIAKNPWLVNTEF (SEQ ID NO:116)

[0595] SETD8 (人)

[0596] ASSPKKKRKVEASSCDSTNAIAKQALKKPIKKGQAPRKAQGGKQQNRKLTDFYPVRRSSRKSKAEL
QSEERKRIDELIESGKEEGMKIDLIDGKGRGVIATKQFSRGDFVVEYHGDLEIETDAKKREALYAQDPSTGCYMY
FQYLSKTYCVDATRETNRLGRLINHSKCGNCQTKLHDIDGVPHLILIASRDI AAGEELLYDYGDRSKASIEAFPWL
KHEF (SEQ ID NO:117)

[0597] TgSET8 (T. gondii):

[0598] ASSPKKKRKVEASASRRTGEFLRDAQAPSRWLKRSKTGQDDGAFCLWTLAGAGDDAAGGERGRDREG
AADKAKQREERRQKELEERFEEMKVEFEKAQRMIARRAALTGEIYSDGKGSKKPRVPSLPENDDALIEI I IDPE
QGILKWPLSVMSIRQRTVIYQECLRRDLTACIHLTKVPGKGRAVFAADTILKDDFVVEYKGE LCSEREAREQRY
NRSKVPMGSFMFYFKNGSRMMAIDATDEKQDFGPRLINHSRRNPNTPRAITLGDNFSEPRLIFVARRNIEKGEE
LLVDYGERDPDVIKEHPWLNSEF (SEQ ID NO:118)

[0599] 本领域技术人员将理解使用本发明公开的技术可将任意例示性的Cas9蛋白,包括例示性的Cas9核酸酶、变体及其融合物,例如本文描述的,投递至细胞,且本公开在此方面

不受限。

[0600] 核酸酶效应器蛋白

[0601] TALE核酸酶或TALEN是包含与DNA切割域(例如FokI域)联合的转录激活物样效应子DNA结合域的人工核酸酶。已报道了一些用于生成工程化的TALE构建体的模式装配设计方案(Zhang,Feng;等(February2011).”Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription”.Nature Biotechnology 29 (2):149-53;Geißler,R.;Scholze,H.;Hahn,S.;Streubel,J.;Bonas,U.;Behrens,S.E.;Boch,J.(2011),Shiu,Shin-Han.编“Transcriptional Activators of人Genes with Programmable DNA-Specificity”.PLoS ONE 6(5):e19509;Cermak,T.;Doyle,E.L.;Christian,M.;Wang,L.;Zhang,Y.;Schmidt,C.;Baller,J.A.;Somia,N.V.等(2011).”Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting”.Nucleic Acids Research;Morbitzer,R.;Elsaesser,J.;Hausner,J.;Lahaye,T.(2011).”Assembly of custom TALE-type DNA binding domains by modular cloning”.Nucleic Acids Research;Li,T.;Huang,S.;Zhao,X.;Wright,D.A.;Carpenter,S.;Spalding,M.H.;Weeks,D.P.;Yang,B.(2011).”Modularly assembled designer TAL effector nucleases for targeted gene knockout and gene replacement in eukaryotes”.Nucleic Acids Research.;Weber,E.;Gruetzner,R.;Werner,S.;Engler,C.;Marillonnet,S.(2011).Bendahmane,Mohammed.编“Assembly of Designer TAL Effectors by Golden Gate Cloning”.PLoS ONE 6(5):e19722;其各自完整内容通过提述并入本文)。本领域技术人员将理解TALE核酸酶可经工程化以高特异性靶向实际上任何基因组序列,且这类工程化的核酸酶可用于本发明技术的实施方案中以操纵细胞的基因组,例如通过经由本文中公开的方法或策略在适合TALEN结合并切割细胞基因组内的其靶序列的情况下投递相应的TALEN。在一些实施方案中,投递的TALEN靶向与疾病或病症有关的基因或等位基因。在一些实施方案中,对受试者投递TALEN赋予对受试者的治疗益处。

[0602] 锌指核酸酶是一类人工核酸酶,其包含DNA切割域和锌指DNA结合域。在一些实施方案中,所述DNA切割域是限制性内切核酸酶,例如FokI的非特异性DNA切割域。在一些实施方案中,所述DNA切割域是仅在与同一类型的第二DNA切割域二聚化时切割双链DNA的域。在一些实施方案中,所述DNA切割域经由接头,例如肽接头融合于锌指域的C末端。在一些实施方案中,所述锌指域包含约3至约6个锌指且特异性识别和结合长度约9-20个核苷酸的靶序列。在一些实施方案中,通过本发明提供的系统或方法将多个锌指核酸酶分子投递给靶细胞,其中一个锌指核酸酶分子的锌指域在第二锌指核酸酶分子的靶序列的紧密近端结合靶序列。在一些实施方案中,在彼此的紧密近端结合靶序列的锌指核酸酶分子的锌指域是不同的。在一些实施方案中,通过本文提供的系统或方法投递至细胞的锌指核酸酶分子在另一锌指核酸酶分子的靶序列的紧密近端结合靶核酸序列,从而分子的DNA切割域二聚化并在两个靶序列之间的位点处切割DNA分子。

[0603] 在一些实施方案中,经由本文公开的策略或方法通过投递至细胞的核酸酶来编辑靶细胞的基因组,例如通过TALEN、或锌指核酸酶、或多种这类核酸酶或这类核酸酶的组合。在一些实施方案中,通过核酸酶将单链或双链断裂引入靶细胞基因组内的特定定位点,从而

导致靶向的基因组序列的破坏。在一些实施方案中,靶向的基因组序列是基因编码区内的核酸序列。在一些实施方案中,核酸酶引入的链断裂导致靶基因内的突变,其损害编码的基因产物的表达。在一些实施方案中,核酸与核酸酶一起共投递至细胞。在一些实施方案中,所述核酸包含与邻近核酸酶靶位点的序列相同或同源的序列。在一些这类实施方案中,通过细胞DNA修复体系修复由核酸酶引起的链断裂以将全部或部分的共投递的核酸在断裂位点处引入细胞DNA,从而导致共投递的核酸或其部分的靶向性插入。在一些实施方案中,插入导致致病性等位基因的破坏或修复。在一些实施方案中,通过适宜的测定法检测插入,例如DNA测序测定、southern印迹测定、或针对由共投递的核酸编码的报道基因的测定法,例如荧光蛋白或对抗生素的抗性。在一些实施方案中,所述核酸通过与超电荷蛋白联合来共投递。在一些实施方案中,所述超电荷蛋白还与功能性效应器蛋白例如核酸酶联合。在一些实施方案中,核酸酶投递至靶细胞导致基因功能的临床或治疗上有益的破坏。

[0604] 在一些实施方案中,从受试者获得细胞并通过本文提供的系统或方法离体将核酸酶或其他效应器蛋白投递至细胞。在一些实施方案中,从经处理的细胞选择其中已引起期望的核酸酶介导的基因组编辑事件的那些细胞。在一些实施方案中,将携带期望的基因组突变或改变的经处理细胞返回至获得其的受试者。

[0605] 用于靶向特定序列的核酸酶(例如TALE或锌指核酸酶)的工程化、生成和分离,和在特定靶序列处编辑细胞基因组的方法是本领域公知的(参见,例如Mani等,Biochemical and Biophysical Research Communications 335:447-457,2005;Perez等,Nature Biotechnology 26:808-16,2008;Kim等,Genome Research,19:1279-88,2009;Urnov等,Nature 435:646-51,2005;Carroll等,Gene Therapy 15:1463-68,2005;Lombardo等,Nature Biotechnology 25:1298-306,2007;Kandavelou等,Biochemical and Biophysical Research Communications 388:56-61,2009;和Hockemeyer等,Nature Biotechnology 27(9):851-59,2009,以及在每种核酸酶的相应部分中记载的参考文献)。熟练技术人员基于本文中提供的指导将能够确定适用于本公开上下文的方法。

[0606] TALE效应器蛋白

[0607] 在一些实施方案中,通过本文提供的系统或方法将包含TALE域的效应器蛋白投递给靶细胞。在一些实施方案中,通过由本发明的方面提供的系统或方法将TALE效应器,例如工程化的TALE转录因子(包含TALE DNA结合域和异源转录激活物或阻遏物域)投递给细胞。在一些实施方案中,所述TALE效应器,例如转录因子,以足以激活或抑制细胞中转录因子对靶基因的转录的量投递给细胞。在一些实施方案中,以足以引起靶细胞内的表型变化,例如细胞功能的变化或发育潜力的变化的量和一段时间段来投递转录因子。例示性TALE转录因子在本文描述,且熟练技术人员将能基于本文提供的指导和本领域中这类TALE转录因子的知识来鉴定另外的适宜TALE转录因子。

[0608] 在一些实施方案中,使靶细胞例如体细胞与TALE转录因子或这类因子的组合接触,所述因子与本文中提供的超电荷蛋白联合。在一些实施方案中,所述靶细胞是原代体细胞且体外或离体与联合于超电荷蛋白的TALE转录因子接触。在一些实施方案中,所述TALE转录因子与正电荷的超电荷蛋白联合,例如如本文描述的。在一些实施方案中,所述TALE转录因子与带负电荷的超电荷蛋白联合,例如如本文描述的。在一些实施方案中,所述TALE转录因子与阳离子脂质和/或阳离子聚合物联合,例如如本文描述的。在一些实施方案中,所

述TALE转录因子与带负电荷的超电荷蛋白和阳离子脂质和/或阳离子聚合物联合,例如如本文描述的。

[0609] 在一些实施方案中,使靶细胞与如本文中提供的联合于超电荷蛋白(和任选地阳离子脂质和/或阳离子聚合物)的TALE转录因子接触或反复接触,并检测到细胞表型或基因表达中的期望变化。在一些实施方案中,使靶细胞与如本文中提供的联合于超电荷蛋白(和任选地阳离子脂质和/或阳离子聚合物)的TALE转录因子反复接触,直至检测到期望的细胞表型的形成。用于检测细胞表型和基因表达的方法是本领域技术人员公知的,且包括例如,形态学分析,和通过确立的方法如免疫组化、荧光活化的细胞分选(FACS)或荧光显微术来检测标志物基因表达。在一些实施方案中,使靶细胞与如本文中提供的联合于超电荷蛋白的TALE转录因子接触至少3小时,至少6小时,至少12小时,至少1天,至少2天,至少3天,至少4天,至少5天,至少6天,至少7天,至少10-12天,至少12-15天,至少15-20天,至少20-25天,至少25-30天,至少30-40天,至少40-50天,至少50-60天,至少60-70或至少70-100天的时间段。

[0610] 在一些实施方案中,使靶细胞与如本文中提供的联合于超电荷蛋白(和任选地阳离子脂质和/或阳离子聚合物)的TALE转录因子接触,其量和时间段足以将细胞程序化导向至不同的细胞状态。如对于本领域技术人员明显的,程序化或再程序化细胞的所需量将取决于多种因素,例如细胞类型和处理时程安排。一般地,通过本文提供的系统或方法投递TALE转录因子至靶向的体细胞将在低于可观察到显著毒性的浓度的浓度。临界浓度将依赖于例如具体的TALE转录因子,其所联合的超电荷蛋白,联合类型,和所处理的细胞类型。

[0611] 本领域技术人员可通过常规实验手段确立用于投递至特定细胞类型的联合于超电荷蛋白(和任选地阳离子脂质和/或阳离子聚合物)的功能性效应器蛋白的有用浓度。在一些实施方案中,使靶细胞与浓度为约1pM至约1 μ M的联合于超电荷蛋白(和任选地阳离子脂质和/或阳离子聚合物)的功能性效应器蛋白体外或离体接触。在一些实施方案中,使靶细胞与浓度为以下的联合于超电荷蛋白的功能性效应器蛋白体外或离体接触:约1pM、约2.5pM、约5pM、约7.5pM、约10pM、约20pM、约25pM、约30pM、约40pM、约50pM、约60pM、约70pM、约75pM、约80pM、约90pM、约100pM、约200pM、约250pM、约300pM、约400pM、约500pM、约600pM、约700pM、约750pM、约800pM、约900pM、约1nM、约2nM、约3nM、约4nM、约5nM、约6nM、约7nM、约8nM、约9nM、约10nM、约20nM、约25nM、约30nM、约40nM、约50nM、约60nm、约70nM、约75nM、约80nM、约90nM、约100nM、约200nM、约250nM、约300nM、约400nM、约500nM、约600nM、约700nM、约750nM、约800nM、约900nM,或约1 μ M。靶细胞暴露于功能性效应器蛋白,和如果必要的话,在缺少功能性效应器蛋白情况下的施用后温育的有用时间,以及可用于实现期望的生物学效果(例如基因转录变化,通过投递的核酸酶切割靶位点等等)的施用/温育循环数,或期望的细胞表型也可以由本领域技术人员通过常规实验手段确定。

[0612] 在一些实施方案中,用于通过本文提供的系统或方法投递功能性效应器蛋白的靶细胞是通过从受试者的活组织检查获得的原代细胞。在一些实施方案中,该受试者诊断为患有疾病。在一些实施方案中,所述疾病是变性疾病(degenerative disease),其特征为特定细胞类型例如神经细胞的降低的功能。在一些实施方案中,依照本文公开的策略或方法用功能性效应器蛋白处理的细胞,或这类细胞的后代,用在细胞替换治疗办法中。在一些实施方案中,在自体同源细胞替换治疗办法中对获得体细胞的受试者施用经处理的细胞。

[0613] 在一些实施方案中,将能使细胞从一个分化状态转化为另一个的功能性效应器蛋白,例如TALE转录因子,通过本文提供的系统或方法体外或体内投递给靶细胞。引起转分化的转录因子是本领域中已知的(参见,例如Zhou等,Nature 455:627-33,2008)。在一些实施方案中,将调控PPAR γ 或PRDM16的表达的TALE转录因子通过如本发明提供的系统或方法投递至成纤维细胞。本领域已知表达这些转录因子在将成纤维细胞程序化为褐色脂肪或白色脂肪细胞状态中是关键步骤。在一些实施方案中,从自需要褐色脂肪细胞的受试者获得的成纤维细胞生成程序化的褐色脂肪细胞,并施用给该受试者,例如用在涉及该受试者的细胞替换治疗办法中。

[0614] 复合物的形成

[0615] 本发明提供包含与要投递的一种或多种功能性效应器蛋白联合的超电荷蛋白的复合物。在一些实施方案中,超电荷蛋白经由非共价相互作用与要投递的一种或多种功能性效应器蛋白联合。在一些实施方案中,超电荷蛋白经由静电相互作用与一种或多种功能性效应器蛋白联合。在某些实施方案中,超电荷蛋白具有总体净正电荷,而要投递的功能性效应器蛋白具有总体净负电荷。在一些实施方案中,复合物还包含阳离子脂质和/或阳离子聚合物。例如,在一些实施方案中,复合物的超电荷蛋白是超级带负电荷的,从而允许与阳离子脂质和/或聚合物联合。

[0616] 在某些实施方案中,超电荷蛋白经由共价键与要投递的一种或多种功能性效应器蛋白联合。例如,超电荷蛋白可以融合于要投递的功能性效应器蛋白。共价附接可以是直接或间接的(例如经由接头)。在一些实施方案中,共价附接经由一个或多个接头介导。在一些实施方案中,接头是可切割接头。在某些实施方案中,所述可切割的接头包含酰胺、酯或二硫键。例如,所述接头可以是可被细胞酶切割的氨基酸序列。在某些实施方案中,所述酶是蛋白酶。在其他实施方案中,所述酶是酯酶。在一些实施方案中,所述酶是在特定细胞类型中比在其他细胞类型中更高度表达的酶。例如,所述酶可以是在肿瘤细胞中比在非肿瘤细胞中更高度表达的酶。例示性接头和切割那些接头的酶在下文呈现。

[0617] 可切割接头

[0618]

接头序列	靶向接头的酶
X ¹ -AGVF-X (SEQ ID NO: 256)	溶酶体硫醇蛋白酶(参见,例如Duncan等,1982, Biosci. Rep., 2:1041-46; 通过提述并入本文)
X-GFLG-X (SEQ ID NO: 257)	溶酶体半胱氨酸蛋白酶(参见,例如Vasey等, Clin. Canc. Res., 1999, 5:83-94; 通过提述并入本文)
X-FK-X	组织蛋白酶(Cathepsin) B - 遍在的,在许多实体瘤中过表达,如乳腺癌(参见,例如Dubowchik等,2002, Bioconjugate Chem., 13:855-69; 通过提述并入本文)
X-A*L-X	组织蛋白酶B - 遍在的,在许多实体瘤中过表达,如乳腺癌(参见,例如Trouet等,1982, Proc. Natl. Acad. Sci., USA, 79:626-29; 通过提述并入本文)
X-A*LA*L-X (SEQ ID NO: 258)	组织蛋白酶B - 遍在的,在许多实体瘤中过表达(参

	见, 例如 Schmid 等, 2007, Bioconjugate Chem, 18:702-16; 通过提述并入本文)
[0619]	X-AL*AL*A-X (SEQ ID NO: 259)
	组织蛋白酶D -遍在的(参见, 例如Czerwinski等, 1998, Proc. Natl. Acad. Sci., USA, 95:11520-25; 通过提述并入本文)

[0620] ¹X表示超电荷蛋白或要投递的功能性效应器蛋白

[0621] *指观察到的切割位点

[0622] 为了仅给出一个具体的例子是, +36GFP可以与要投递的功能性效应器蛋白通过可切割接头如ALAL (SEQ ID NO:254) 联合, 以产生+36GFP-(GGG)₄-ALAL-(GGG)₄-[功能性效应器蛋白X] (SEQ ID NO:255)。

[0623] 在某些实施方案中, 使要投递的功能性效应器蛋白与超电荷蛋白接触以形成复合物。在一些实施方案中, 复合物的形成在pH 7或约pH 7进行。在一些实施方案中, 复合物的形成在约pH 5、约pH 6、约pH 7、约pH 8或约pH 9进行。复合物的形成通常在不对超电荷蛋白和/或功能性效应器蛋白的功能有不利影响的pH进行。在一些实施方案中, 复合物的形成在室温进行。在一些实施方案中, 复合物的形成在37°C或其附近进行。在一些实施方案中, 复合物的形成在低于4°C, 在约4°C, 在约10°C, 在约15°C, 在约20°C, 在约25°C, 在约30°C, 在约35°C, 在约37°C, 在约40°C或高于40°C进行。复合物的形成通常在不对超电荷蛋白和/或功能性效应器蛋白的功能有不利影响的温度进行。在一些实施方案中, 复合物的形成在无血清培养基中进行。在一些实施方案中, 复合物的形成在存在CO₂ (例如约1%、约2%、约3%、约4%、约5%、约6%或更多) 的情况下进行。

[0624] 在一些实施方案中, 使用约100nM的功能性效应器蛋白的浓度实施复合物的形成。在一些实施方案中, 使用浓度为约25nM、约50nM、约75nM、约90nM、约100nM、约110nM、约125nM、约150nM、约175nM或约200nM的功能性效应器蛋白的浓度实施复合物的形成。在一些实施方案中, 使用约40nM的超电荷蛋白浓度实施复合物的形成。在一些实施方案中, 使用浓度为约10nM、约20nM、约30nM、约40nM、约50nM、约60nM、约70nM、约80nM、约90nM或约100nM的超电荷蛋白的浓度实施复合物的形成。

[0625] 在一些实施方案中, 在过量功能性效应器蛋白的条件下实施复合物的形成。在一些实施方案中, 以功能性效应器蛋白:超电荷蛋白为约20:1、约10:1、约9:1、约8:1、约7:1、约6:1、约5:1、约4:1、约3:1、约2:1或约1:1的比率实施复合物的形成。在一些实施方案中, 以功能性效应器蛋白:超电荷蛋白的比率为约3:1实施复合物的形成。在一些实施方案中, 超电荷蛋白:功能性效应器蛋白为约20:1、约10:1、约9:1、约8:1、约7:1、约6:1、约5:1、约4:1、约3:1、约2:1或约1:1实施复合物的形成。

[0626] 在一些实施方案中, 通过将超电荷蛋白与功能性效应器蛋白混合, 并搅动混合物 (例如通过颠倒) 实施复合物的形成。在一些实施方案中, 通过将超电荷蛋白与功能性效应器蛋白混合并允许混合物静置实施复合物的形成。在一些实施方案中, 在存在药学可接受载体或赋形剂的情况下实施复合物的形成。在一些实施方案中, 将所述复合物进一步与药学可接受载体或赋形剂组合。例示性赋形剂或载体包括水、溶剂、脂质、蛋白、肽、核内体裂解剂 (例如氯奎、茈丁酸)、小分子、糖类、缓冲剂、天然聚合物、合成聚合物 (例如PLGA、聚氨酯、聚酯、聚己酸内酯、聚磷腈 (polyphosphazenes))、药学试剂等。

[0627] 在一些实施方案中,包含超电荷蛋白和功能性效应器蛋白的复合物可能比单独的超电荷蛋白或单独的功能性效应器蛋白在凝胶电泳测定法中更缓慢地迁移。

[0628] 应用

[0629] 本发明提供包含与要投递至细胞的功能性效应器蛋白(例如核酸酶、转录激活物/阻遏物、重组酶、Cas9蛋白包含其变体和融合物,等等)联合的超电荷蛋白(天然存在或经工程化的)的组合物,以及使用这类组合物的方法和这类组合物的用途。在某些实施方案中,提供包含Cas9蛋白(例如其中Cas9蛋白与gRNA联合)和阳离子脂质的组合物。在某些实施方案中,提供包含Cas9蛋白(例如其中Cas9蛋白与gRNA联合)和阳离子聚合物的组合物。本发明的组合物可用于治疗或预防任何可能受益于例如药剂投递至细胞的疾病。本发明组合物还可以用于转染或处理细胞用于研究目的。

[0630] 在一些实施方案中,依照本发明的组合物可用于研究目的,例如以在研究背景中将功能性效应器蛋白有效投递至细胞。在一些实施方案中,依照本发明的组合物可用于治疗目的。在一些实施方案中,依照本发明的组合物可用于治疗任意的多种疾病、病症和/或状况,包括但不限于以下一种或多种:自身免疫性病症(例如糖尿病、狼疮、多发性硬化、银屑病、类风湿性关节炎);炎性病症(例如关节炎、盆腔炎(pelvic inflammatory disease));感染性疾病(例如病毒感染(例如HIV,HCV,RSV)、细菌感染、真菌感染、败血症);神经学病症(例如阿尔兹海默病(Alzheimer's disease)、亨廷顿氏病(Huntington's disease);孤独症;迪谢内肌营养不良(Duchenne muscular dystrophy));新血管病症(例如动脉粥样硬化、血胆固醇过多、血栓形成、凝血病症、血管生成性病症如黄斑变性);增殖性病症(例如癌症、良性赘生物);呼吸病症(例如慢性阻塞性肺病);消化病症(例如炎性肠病、溃疡);肌骨骼病症(例如纤维肌痛(fibromyalgia)、关节炎);内分泌、代谢和营养病症(例如糖尿病、骨质疏松);泌尿病症(例如肾病);心理病症(例如抑郁、精神分裂症);皮肤病症(例如伤口、湿疹);血液和淋巴病症(例如贫血、血友病);等等。

[0631] 本发明的组合物可用于临床背景。例如,超电荷蛋白可以与可用于治疗应用的功能性效应器蛋白联合。这类功能性效应器蛋白可以是,例如核酸酶或转录激活物。也可以使用包含Cas9蛋白和阳离子脂质的其他组合物用于治疗应用。

[0632] 在一些实施方案中,与超电荷蛋白联合的超电荷蛋白或功能性效应器蛋白包含可检测标记物。这些分子可用于检测、成像、疾病分期、诊断或患者选择。适宜的标签包括荧光、化学发光、酶标记物,比色、磷光、基于密度的标记物,例如基于电子密度的标记物,和一般的对比剂(contrast agent)和/或放射性标记物。

[0633] 药物组合物

[0634] 本发明提供包含与要投递的至少一种功能性效应器蛋白(且在一些实施方案中被阳离子脂质包封)联合的超电荷蛋白的组合物。提供包含Cas9蛋白和阳离子脂质的其他组合物。如此,本发明提供药物组合物,其包含与功能性效应器蛋白联合的一种或多种超电荷蛋白;和/或与阳离子脂质和/或阳离子聚合物联合的一种或多种功能性效应器蛋白,以及一种或多种药学可接受的赋形剂。药物组合物可任选地包含一种或多种另外的治疗活性物质。依照一些实施方案,提供将包含与要投递的功能性效应器蛋白联合的一种或多种超电荷蛋白的药物组合物施用给有此需要的受试者的方法。在一些实施方案中,组合物施用给人。就本公开目的而言,短语“活性成分”一般指与功能性效应器蛋白联合的Cas9蛋白和/或

超电荷蛋白,或如本文描述的要投递的功能性效应器蛋白。

[0635] 尽管本文中提供的药物组合物的描述主要针对适用于施用给人的药物组合物,熟练技术人员将理解这类组合物一般适用于施用给所有种类的动物。能较好地理解对适用于施用给人的药物组合物更改以使得该组合物适用于施用给各种动物,且普通技能的兽医药理学家能仅使用普通的实验手段(若有的话)设计和/或实施这类更改。施用所述药物组合物的受试者包括但不限于,人和/或其他灵长类;哺乳动物,包括商业相关的哺乳动物如牛、猪、马、绵羊、猫、犬、小鼠和/或大鼠;和/或禽类,包括商业相关的禽类如鸡、鸭、鹅和/或火鸡。

[0636] 可以通过药理学领域中已知或后来开发的任何方法制备本文中描述的药物组合物的配制剂。一般地,此类制备方法包括下述步骤,即使活性成分与赋形剂和/或一种或多种其它辅助成分联合,然后若必要和/或想要的话,运输和/或包装产品成期望的单剂量或多剂量形式。

[0637] 依照本发明的药物组合物可以以单单位剂量和/或多个单单位剂量大批制备、包装和/或销售。如本文中使用的,“单位剂量”是包含预定量的活性成分的药物组合物的离散量。活性成分的量一般等于将施用给受试者的活性成分剂量和/或这类剂量的方便分数,例如,这类剂量的一半或三分之一。

[0638] 依照本发明的药物组合物中活性成分、药学可接受赋形剂和/或任何另外的成分的相对量将随着治疗的受试者的身份、块头和/或状况以及进一步随着要施用组合物的路径而变化。例如,所述组合物可以包含0.1%至100%(w/w)的活性成分。

[0639] 另外,药物组合物可以包含药学可接受赋形剂,如本文中使用的,其包括任何和所有溶剂、分散介质、稀释剂或其它液体媒介物、分散或悬浮助剂、表面活性剂、等张剂、增稠剂或乳化剂、防腐剂、固体粘合剂、润滑剂等,如适合于期望的特定剂量形式的。Remington's The Science and Practice of Pharmacy, 21st Edition, A.R. Gennaro (Lippincott, Williams & Wilkins, Baltimore, MD, 2006; 其通过提及完整并入本文)公开了配制药物组合物中使用的各种赋形剂及其已知的制备技术。除非任何常规赋形剂介质与物质或其衍生物不相容,诸如通过产生任何不想要的生物效应或以其它方式与药物组合物的任何其它组分以有害的方式相互作用,涵盖其用途在本公开内容的范围内。

[0640] 在一些实施方案中,药学可接受的赋形剂是至少95%,至少96%,至少97%,至少98%,至少99%或100%纯的。在一些实施方案中,批准赋形剂用于人和用于兽医用途。在一些实施方案中,赋形剂被美国食品和药物管理局批准。在一些实施方案中,赋形剂是药物级的。在一些实施方案中,赋形剂满足美国药典(USP)、欧洲药典(EP)、英国药典和/或国际药典的标准。

[0641] 用于制造药物组合物的药学可接受的赋形剂包括但不限于,惰性稀释剂、分散剂和/或粒化剂、表面活性剂和/或乳化剂、崩解剂、结合剂、防腐剂、缓冲剂、润滑剂和/或油。这类赋形剂可以任选地包含在药学配制剂中。赋形剂如可可油和栓剂蜡(suppository waxe)、着色剂、包被剂、甜化剂、调味剂和/或芳香剂可根据配制者的判断而存在于组合物中。

[0642] 例示性稀释剂包括但不限于,碳酸钙、碳酸钠、磷酸钙、磷酸二钙、硫酸钙、磷酸氢钙、乳糖磷酸钠、蔗糖、纤维素、微晶纤维素、高岭土、甘露醇、山梨糖醇、肌醇、氯化钠、干淀

粉、玉米淀粉、粉化糖、等等和/或其组合。

[0643] 例示性粒化剂和/或分散剂包括但不限于,马铃薯淀粉、玉米淀粉、木薯淀粉、羧基乙酸淀粉钠、黏土、褐藻酸、瓜尔豆胶(guar gum)、柑橘浆、琼脂、膨润土、纤维素和木材产品、天然海绵、阳离子交换树脂、碳酸钙、硅酸钠、碳酸钠、交联的聚(乙烯吡咯烷酮)(交聚维酮, crospovidone)、羧甲基淀粉钠(羧基乙酸淀粉钠)、羧甲基纤维素、交联的羧甲基纤维素钠(交联甲羧纤维素, croscarmellose)、甲基纤维素、预胶凝淀粉(淀粉1500)、微晶淀粉、水不溶性淀粉、羧甲基纤维素钙、硅酸镁铝(Veegum)、十二醇硫酸钠(sodium lauryl sulfate)、季铵化合物,等等和/或其组合。

[0644] 例示性表面活性剂和/或乳化剂包括但不限于,天然乳化剂(例如,阿拉伯树胶、琼脂、褐藻酸、藻酸钠、黄芪胶、chondrux、胆固醇、黄原胶、果胶、明胶、蛋黄、酪蛋白、羊毛脂、胆固醇、蜡和卵磷脂)、胶体状黏土(例如,膨润土[硅酸铝]和Veegum[®][硅酸镁铝])、长链氨基酸衍生物、高分子量醇(例如硬脂酰醇、十六烷醇、油醇、三乙酸甘油酯单硬脂酸酯(triacetin monostearate)、乙二醇二硬脂酸酯(ethylene glycol distearate)、甘油基单硬脂酸酯、和丙二醇单硬脂酸酯、聚乙烯醇)、卡波姆(carbomer)(例如羧基聚亚甲基、聚丙烯酸、丙烯酸聚合物和羧基乙烯基聚合物)、角叉胶、纤维素衍生物(例如羧甲基纤维素钠、粉状纤维素、羟甲基纤维素、羟丙基纤维素、羟丙基甲基纤维素、甲基纤维素)、山梨聚糖脂肪酸酯(例如聚氧乙烯山梨聚糖单月桂酸酯(Tween[®]20)、聚氧乙烯山梨聚糖(Tween[®]60)、聚氧乙烯山梨聚糖单油酸酯[Tween[®]80]、山梨聚糖单棕榈酸酯[Span[®]40]、山梨聚糖单硬脂酸酯[Span[®]60]、山梨聚糖三硬脂酸酯[Span[®]65]、甘油基单油酸酯、山梨聚糖单油酸酯[Span[®]80])、聚氧乙烯酯(例如聚氧乙烯单硬脂酸酯[Myrij[®]45]、聚氧乙烯氢化蓖麻油、聚乙氧基化蓖麻油、聚氧亚甲基硬脂酸酯和Solutol[®])、蔗糖脂肪酸酯、聚乙二醇脂肪酸酯(例如Cremophor[®])、聚氧乙烯醚(例如聚氧乙烯月桂基醚(Brij[®]30))、聚(乙烯-吡咯烷酮)、二乙二醇单月桂酸酯、三乙醇胺油酸酯、油酸钠、油酸钾、油酸乙酯、油酸、月桂酸乙酯、十二烷基硫酸钠、Pluronic[®]F 68、Poloxamer[®]188、西曲溴铵(cetrimonium bromide)、氯化十六烷基吡啶(cetylpyridinium chloride)、苯扎氯铵(benzalkonium chloride)、多库酯钠(docusate sodium)等等和/或其组合。

[0645] 例示性结合剂包括但不限于,淀粉(例如玉米淀粉和淀粉糊);明胶;糖(例如蔗糖、葡萄糖、右旋糖、糊精、糖蜜、乳糖、乳糖醇、甘露醇);天然和合成的胶质(例如阿拉伯树胶、藻酸钠、爱尔兰藓提取物(extract of Irish moss)、Panwar胶、印度胶(ghatti gum)、isapol壳的胶浆、羧甲基纤维素、甲基纤维素、乙基纤维素、羟乙基纤维素、羟丙基纤维素、羟丙基甲基纤维素、微晶纤维素、醋酸纤维素、聚乙烯吡咯烷酮、硅酸镁铝(Veegum[®])、和落叶松阿拉伯半乳糖);藻酸盐;聚环氧乙烷;聚乙二醇;无机钙盐;硅酸;聚甲基丙烯酸酯;蜡;水;醇;等等;及其组合。

[0646] 例示性防腐剂可以包括但不限于,抗氧化剂、螯合剂、抗微生物防腐剂、抗真菌防腐剂、醇防腐剂、酸性防腐剂和/或其他防腐剂。例示性抗氧化剂包括但不限于, alpha生育酚、抗坏血酸、维生素c棕榈酸酯、丁羟茴醚、丁羟甲苯、二羟基丙硫醇、偏亚硫酸氢钾、丙酸、

没食子酸丙酯、抗坏血酸钠、重亚硫酸钠、偏亚硫酸氢钠和/或亚硫酸钠。例示性螯合剂包括乙二胺四乙酸(EDTA)、柠檬酸一水合物、依地酸二钠、依地酸二钾、依地酸、延胡索酸、苹果酸、磷酸、依地酸钠、酒石酸和/或依地酸三钠。例示性抗微生物防腐剂包括但不限于,苯扎氯铵、苜蓿氯铵、苯甲醇、溴硝丙二醇(bronopol)、溴化十六烷基三甲铵、西吡氯铵、氯己定、三氯叔丁醇、氯化甲酚、盐酸辛可芬、甲酚、乙醇、甘油、海克替啶、咪脲、酚、苯氧基乙醇、苯乙基醇、硝酸苯汞、丙二醇和/或硫柳汞(thimerosal)。例示性抗真菌防腐剂包括但不限于,尼泊金丁酯(butyl paraben)、尼泊金甲酯、尼泊金乙酯、尼泊金丙酯、苯甲酸、羟基苯甲酸、苯甲酸钾、山梨酸钾、苯甲酸钠、丙酸钠和/或山梨酸。例示性醇防腐剂包括但不限于,乙醇、聚乙二醇、苯酚、酚化合物、双酚、氯丁醇、羟基苯甲酸和/或苯乙基醇。例示性酸性防腐剂包括但不限于,维生素A、维生素C、维生素E、beta-胡萝卜素、柠檬酸、乙酸、脱氢乙酸、抗坏血酸、山梨酸和/或植酸(phytic acid)。其他防腐剂包括但不限于,生育酚、醋酸生育酚、甲磺酸去铁胺(deteroxime mesylate)、溴化十六烷基三甲铵、丁羟茴醚(BHA)、丁羟甲苯(BHT)、乙二胺、十二烷基硫酸钠(SLS)、月桂基乙醚硫酸钠(SLES)、重亚硫酸钠、偏亚硫酸氢钠、亚硫酸钾、偏亚硫酸氢钾、Glydant Plus[®]、Phenonip[®]、尼泊金甲酯、

Germall[®]115、Germaben[®]II、Neolone[™]、Kathon[™]和/或Euxyl[®]。

[0647] 例示性缓冲剂包括但不限于,柠檬酸盐缓冲溶液、乙酸盐缓冲溶液、磷酸盐缓冲溶液、氯化铵、碳酸钙、氯化钙、柠檬酸钙、葡乳醛酸钙、葡庚糖酸钙(calcium gluceptate)、葡糖酸钙、D-葡糖酸、甘油磷酸钙、乳酸钙、丙酸、乙酰丙酸钙、戊酸、二碱式磷酸钙、磷酸、三碱式磷酸钙、磷酸氢氧化钙、氯化钾、葡糖酸钾、钾混合物、二碱式磷酸钾、单碱式磷酸钾、磷酸钾混合物、乙酸钠、碳酸氢钠、氯化钠、柠檬酸钠、乳酸钠、二碱式磷酸钠、单碱式磷酸钠、磷酸钠混合物、氨丁三醇、氢氧化镁、氢氧化铝、褐藻酸、无致热原水、等张盐水、林格氏溶液、乙醇等等和/或其组合。

[0648] 例示性润滑剂包括但不限于,硬脂酸镁、硬脂酸钙、硬脂酸、二氧化硅、滑石、麦芽、glyceryl behenate、氢化植物油、聚乙二醇、苯甲酸钠、乙酸钠、氯化钠、亮氨酸、十二烷基硫酸镁、十二烷基硫酸钠等等及其组合。

[0649] 例示性油包括但不限于,杏仁油、杏核油、鳄梨油、巴巴苏仁油、佛手柑油、黑醋栗籽油、琉璃苣油、刺楸油、甘菊油、芸苔油、蒿子油、蜡棕榈油、蓖麻油、肉桂油、可可脂、椰子油、鱼肝油、咖啡油、玉米油、棉籽油、鹌鹑油、桉树油、月见草油、鱼油、亚麻籽油、香叶油、葫芦油、葡萄籽油、榛子油、牛膝草油、肉豆蔻酸异丙酯、荷荷巴油、夏威夷果油、熏衣草精油(lavandin)、熏衣草油、柠檬油、山苍子(litsea cubeba)油、澳洲胡桃油、锦葵油、芒果籽油、白芒花籽油、貂油、肉豆蔻油、橄榄油、橙油、深海鱼油(orange roughy)、棕榈油、棕榈仁油、桃仁油、花生油、罂粟子油、南瓜子油、油菜籽油、米糠油、迷迭香油、红花油、檀香油、山茶花油、香油、沙棘油、芝麻油、牛油树脂、硅油、大豆油、向日葵油、茶树油、薊油、椿油、香根草油、胡桃油、和麦芽油。例示性的油包括但不限于硬脂酸丁酯、辛酸甘油三酯、癸酸甘油三酯、环甲硅油、癸二酸二乙酯、二甲硅油360、肉豆蔻酸异丙酯、矿物油、辛基十二烷醇、油醇、硅油、和/或其组合。

[0650] 用于口服和胃肠外给药的液体剂量形式包括但不限于,药学可接受的乳剂、微乳剂、溶液、悬浮液、糖浆剂和/或酞剂。除了活性成分,液体剂体剂量形式还可以包含本领域

中常用的惰性稀释剂(诸如例如水或其它溶剂)、增溶剂和乳化剂,诸如乙醇、异丙醇、碳酸乙酯、乙酸乙酯、苯甲醇、苯甲酸苄酯、丙二醇、1,3-丁二醇、二甲基甲酰胺、油(特别是棉籽油、落花生油、玉米油、胚芽油、橄榄油、蓖麻油、和芝麻油)、甘油、四氢糠醇、聚乙二醇和脱水山梨糖醇的脂肪酸酯、及其混合物。在惰性稀释剂外,口服组合物可以包含佐剂诸如湿润剂、乳化和悬浮剂、甜味剂、调味剂、和/或芳香剂。在用于胃肠外给药的一些实施方案中,组合物与增溶剂,诸如Cremophor[®]、醇、油、经修饰的油、二醇、聚山梨酸酯、环糊精、聚合物、和/或其组合。

[0651] 可以使用合适的分散剂、湿润剂和/或悬浮剂根据已知技术配制可注射制剂(例如无菌可注射水性或油脂性悬浮液)。无菌可注射制剂可以是无毒性胃肠外可接受稀释剂或溶剂中的无菌可注射溶液、悬浮液和/或乳剂,例如,作为1,3-丁二醇中的溶液。可以采用的可接受媒介物和溶剂是水、林格氏溶液、U.S.P.和等张氯化钠溶液。另外,常规采用无菌不挥发性油作为溶剂或悬浮介质。出于此目的,可以采用任何温和和不挥发性油,包括合成的单甘油酸酯或甘油二酯。另外,在制备注射液中使用脂肪酸诸如油酸。

[0652] 可以如下灭菌可注射制剂:例如通过截留细菌过滤器过滤,和/或通过掺入无菌固体组合物形式的灭菌剂,所述无菌固体组合物可以在使用前溶解或分散于无菌水或其它无菌可注射介质中。

[0653] 为了延长活性成分的效果,经常期望减缓来自皮下或肌肉注射的活性成分的吸收。这可以通过使用具有较差水溶解度的晶体或无定形材料的液体悬浮液来实现。然后,药物的吸收速率取决于其溶解速率,其继而取决于晶体大小和晶型。或者,胃肠外给药药物形式的延迟吸收通过在油性媒介物中溶解或悬浮药物来实现。通过在可生物降解的聚合物如聚丙交酯-聚乙交酯中形成微胶囊基质来制备可注射的depot形式。根据药物对聚合物的比率以及采用的具体聚合物的性质,可以控制药物释放的速率。其他可生物降解的聚合物的例子包括聚(原酸酯)和聚(酸酐)。Depot可注射配制剂通过将药物囊括在与身体组织相容的脂质体或微乳剂中来制备。

[0654] 用于直肠或阴道给药的组合物通常为栓剂,其可以通过将组合物与合适的非-刺激性赋形剂(例如可可脂、聚乙二醇或栓剂蜡(其在室温为固体但在体温为液体,因此其在直肠或阴道腔中融化并释放活性成分))混合制备。

[0655] 用于口服给药的固体剂量形式包括胶囊、片剂、丸剂、粉剂和颗粒剂。在此类固体剂体剂量形式中,混合活性成分与至少一种惰性、药学可接受赋形剂,诸如柠檬酸钠或磷酸二钙诸如和/或填充剂或增量剂(extender)(诸如淀粉、乳糖、蔗糖、葡萄糖、甘露醇、和硅酸),粘合剂(诸如例如羧甲基纤维素、褐藻酸盐、明胶、聚乙烯吡咯烷酮、蔗糖和阿拉伯树胶),润湿剂(诸如甘油),崩解剂(诸如琼脂、碳酸钙、马铃薯或木薯淀粉、褐藻酸、某些硅酸盐和碳酸钠),溶解延迟剂(诸如石蜡),吸收加速剂(诸如季铵化合物),湿润剂(诸如例如鲸蜡醇和单硬脂酸甘油酯),吸收剂(诸如高岭土和膨润土粘土),和润滑剂(诸如滑石、硬脂酸钙、硬脂酸镁、固体聚乙二醇、十二烷基硫酸钠)、及其混合物。在胶囊、片剂和丸剂的情况下,剂量形式可以包含缓冲剂。

[0656] 可以使用诸如乳糖或乳糖及高分子量聚乙二醇等赋形剂在软和硬的填充明胶胶囊中采用类似形式的固体组合物作为填充剂。可以用包衣和壳,诸如肠溶包衣和药学配制领域中公知的其它包衣制备片剂、锭剂、胶囊、丸剂和颗粒剂的固体剂量形式。任选地,它

们可以包含遮光剂,并且可以属于如下的组合物:它们仅或优先在肠道的某个部分中任选以延迟方式释放活性成分。可以使用的嵌入组合物的例子包括聚合物物质和蜡。可以使用诸如乳糖或乳液糖及高分子量聚乙二醇等赋形剂在软和硬的填充明胶胶囊中采用类似形式的固体组合物作为填充剂。

[0657] 用于局部和/或经皮给药组合物的剂量形式可以包括软膏剂、糊剂、乳膏、洗剂、凝胶、粉剂、溶液、喷雾剂、吸入剂和/或贴片。一般地,在无菌条件下混合活性成分与药学可接受赋形剂和/或任何需要的防腐剂和/或缓冲剂(如果需要的话)。另外,本发明涵盖使用经皮贴片,其经常具有对身体提供化合物的受控递送的额外优点。此类剂量形式可以例如通过在适当的介质中溶解和/或分散化合物来制备。或者/另外,可以通过提供速率控制膜和/或通过聚合物基质和/或凝胶中分散化合物控制速率。

[0658] 适合用于递送本文中描述的皮内药物组合物的装置包括短针装置。皮内组合物可以通过限制针在皮肤内的有效穿透长度的装置及其功能等同物给药。经由液体喷射注射器和/或经由针将液体组合物递送到真皮的喷射注射装置是合适的,所述针穿透角质层,并且产生达到真皮的喷射。弹道粉末/颗粒递送装置是合适的,其使用压缩气体将粉末形式的疫苗加速通过皮肤的外层到真皮。或者/另外,可以在皮内给药的经典mantoux法中使用常规的注射器。

[0659] 适合于局部给药的制剂包括但不限于液体和/或半液体制剂,诸如搽剂、洗剂、水包油和/或油包水乳剂诸如乳膏、软膏剂和/或糊剂、和/或溶液和/或悬浮液。局部可给药制剂可以例如包含约1%-约10% (w/w) 活性成分,尽管活性成分的浓度可以高达活性成分在溶剂中的溶解度限度。用于局部给药的制剂可以进一步包含一种或多种本文中描述的别的成分。

[0660] 药物组合物可以以适合经过口腔用于肺部给药的配制剂制备、包装和/或销售。该配制剂可以包括干颗粒,其包含活性成分且其直径为大约0.5至大约7纳米或大约1至大约6纳米的范围。这类组合物方便地为干粉的形式给药,使用包括干粉储存器(可以向其导入推进剂的气流以分散粉末)的装置,和/或使用自推进溶剂/粉末分散容器,例如装有溶解和/或悬浮在密闭容器的低沸点推进剂中的活性成分的装置。这类粉剂包括颗粒,其中至少98%重量的颗粒的直径大于0.5纳米且至少95%数量的颗粒的直径小于7纳米。或者,至少95%重量的颗粒的直径大于1纳米且至少90%数量的颗粒的直径小于6纳米。干粉组合物可以包括固体细粉稀释剂(例如糖)并且方便地以单位剂量形式提供。

[0661] 低沸点推进剂通常包括液体推进剂,其沸点在大气压低于65°F。通常,所述推进剂可以占所述组合物的50至99.9% (w/w),并且活性成分可以占所述组合物的0.1至20% (w/w)。推进剂还可以包括其它成分,例如液体非-离子和/或固体阴离子表面活性剂和/或固体稀释剂(其粒度与包含活性成分的颗粒为一个量级)。

[0662] 被配制为用于肺部递送的药物组合物还提供溶液和/或悬液的微滴形式的活性成分。这类制剂可以被制备、包装和/或销售为包含活性成分的水性和/或稀醇性溶液和/或悬浮液,任选为无菌的,并且可以方便地使用任何喷雾和/或雾化装置给药。这类制剂还可以包含一种或多种其它成分,包括但不限于,调味剂(例如糖精钠)、挥发油、缓冲剂、表面活性剂和/或防腐剂(例如羟苯甲酯)。通过该给药途径提供的微滴可以具有大约0.1至大约200纳米范围内的平均直径。

[0663] 作为可用于肺部递送的本文所述的制剂可用于鼻内递送药物组合物。适合鼻内给药的另一制剂为包含活性成分的粗粉并且具有大约0.2至500微米的平均颗粒。这类制剂以鼻子呼吸的方式给药,即通过经过鼻道的快速吸入从放在鼻孔附近的粉末容器中给药。

[0664] 适用于鼻腔给药的制剂可以,例如包含大约少至0.1% (w/w) 至多至100% (w/w) 的活性成分,并且可以包含一种或多种本文所述的其它成分。药物组合物可以以适用于含服给药的制剂制备、包装和/或销售。此类制剂可以,例如为使用常规方法制得的片剂和/或锭剂的形式,并且可以包含,例如,0.1至20% (w/w) 的活性成份,包括口腔可溶解和/或可降解组成的平衡,任选地,一种或多中本文所述的活性成分。或者,用于含服给药的制剂可以包括包含活性成分的粉末化形式和/或气化和/或雾化的溶液和/或悬浮液形式。分散时,此类粉末化、气化和/或雾化的制剂可以具有大约0.1至大约200纳米范围内的平均颗粒和/或微滴尺寸,并且还包含一种或多种本文所述的其它成分。

[0665] 药物组合物可以以用于眼部给药的制剂制备、包装和/或销售。此类制剂可以,例如为滴眼剂的形式,包括,例如,0.1/1.0% (w/w) 活性成分在水性或油性液体赋形剂中的溶液和/或悬浮液。此类滴眼剂还可以包括缓冲剂、盐和/或一个或多种本文所述的其它成分。有用的其它可眼部给药的制剂包括以微晶型和/或脂质体制剂包含活性成分的那些。滴耳剂和/或滴眼剂也包含在本发明范围内。

[0666] 在配制和/或制造药物药剂中的通常考虑可见于,例如Remington:The Science and Practice of Pharmacy 21st ed.,Lippincott Williams&Wilkins,2005(通过提述并入本文)。

[0667] 施用

[0668] 本发明提供方法,其包括对有此需要的受试者施用包含与功能性效应器蛋白联合的超电荷蛋白的组合物。在一些实施方案中,提供施用包含其他功能性效应器蛋白(例如Cas9蛋白)和阳离子脂质和/或阳离子聚合物的组合物的方法。这类组合物可使用对于预防、治疗、诊断或成像疾病、病症和/或状况有效的任意量和任意施用路径施用给受试者。需要的确切量将随着受试者的种族、年龄和一般状况,疾病的严重性,具体组合物,其施用模式,其活性模式等在受试者之间变化。依照本发明的组合物通常以剂量单位形式配制以用于施用的便易和剂量的均匀性。然而,将理解本发明的组合物的总日常剂量将由主治医师通过合理的医学判断来决定。任何具体患者的特定治疗有效的、防护有效的或适当的成像剂量水平将依赖于多种因素,包括所治疗的病症和该病症的严重性;采用的特定化合物的活性;采用的特定组合物;患者的年龄、体重、一般健康、性别和饮食;采用的特定化合物的施用时间、施用路径和排泄速率;治疗的持续时间;与采用的特定化合物组合或巧合使用的药物;以及医学领域中公知的类似因素。

[0669] 与要投递的功能性效应器蛋白联合的超电荷蛋白的组合物以及包含例如Cas9蛋白和阳离子脂质的组合物可通过任意路径施用。在一些实施方案中,通过多种路径的一种或多种来施用此类组合物,包括口服、静脉内、肌内、动脉内、髓内、鞘内、皮下、心室内、经皮、皮内、直肠、阴道内、腹膜内、局部(例如通过粉剂、软膏、乳膏、凝胶、洗剂和/或滴剂)、粘膜、鼻、含服、肠、玻璃体、肿瘤内、舌下;通过气管内滴注、支气管滴注和/或吸入;和口腔喷雾、鼻喷雾和/或气雾剂,和/或经由门静脉导管。在一些实施方案中,通过系统性静脉内注射来施用超电荷蛋白或复合物,和/或其药学、防护性、诊断性或成像组合物。在特定的实施

方案中,可静脉内和/或口服施用超电荷蛋白或复合物,和/或其药学、防护性、诊断性或成像组合物。在特定的实施方案中,这类组合物可以以允许功能性效应器蛋白跨越血脑屏障、血管屏障或其他上皮屏障的方式施用。

[0670] 在某些实施方案中,依照本发明的组合物可以以下剂量水平施用,该水平足以将从约0.0001mg/kg至约100mg/kg、从约0.01mg/kg至约50mg/kg、从约0.1mg/kg至约40mg/kg、从约0.5mg/kg至约30mg/kg、从约0.01mg/kg至约10mg/kg、从约0.1mg/kg至约10mg/kg、或从约1mg/kg至约25mg/kg受试者体重的功能性效应器蛋白量每日、每日一次或多次地投递,从而获得期望的治疗、诊断、防护或成像效果。期望的剂量可以每日三次、每日两次、每日一次、每隔一日、每三日、每周、每两周、每三周或每四周投递。在某些实施方案中,可以利用多次施用来投递期望的剂量(例如2、3、4、5、6、7、8、9、10、11、12、13、14或更多次施用)。

[0671] 包含与功能性效应器蛋白联合的超电荷蛋白的组合物可以与一种或多种其他治疗剂、防护剂、诊断剂或成像剂组合施用。“组合”不意图指所述药剂同时施用和/或配制为用于一起投递,尽管这些投递方法在本发明的范围内。组合物可以与一种或多种其他期望的治疗学或医学规程同时、在前或在后施用。一般地,每种药剂将以对该药剂确定的剂量和/或时间安排施用。在一些实施方案中,本发明涵盖投递药学、防护性、诊断性或成像组合物与可以改进其生物利用度、降低和/或改良其代谢、抑制其排泄和/或改良其在体内的分布的药剂的组合。

[0672] 试剂盒

[0673] 本发明提供用于方便和/或有效地实施本发明方法的多种试剂盒。通常试剂盒将包括有效量和/或数量的组分来允许使用者实施对一个或多个受试者的多次处理和/或实施多次试验。在一些实施方案中,试剂盒包含以下一或多种:(i)超电荷蛋白,如本文描述的;(ii)功能性效应器蛋白要投递的;(iii)阳离子脂质和/或阳离子聚合物;和(iv)用于配制包含与超电荷蛋白联合的功能性蛋白的组合物使用说明。在一些实施方案中,试剂盒包含Cas9蛋白和阳离子脂质。在一些实施方案中,试剂盒包含编码超电荷蛋白和/或要投递的功能蛋白的核酸。在一些实施方案中,试剂盒包含克隆载体,其编码超电荷蛋白和克隆位点,所述位点允许符合阅读框地克隆功能性效应器蛋白以生成融合蛋白。在一些实施方案中,试剂盒包含本文提供的组合物,其包含与功能性效应器蛋白联合的超电荷蛋白;注射器、针或给药器具以对受试者施用药物组合物;和用于对受试者施用药物组合物的用法说明。

[0674] 本发明的这些和其他方面将在考虑以下实施例的基础上进一步领会,所述实施例意图例示本发明的某些特定实施方案但不意图限制其范围,该范围由权利要求限定。

实施例

[0675] 实施例1:投递融合于超电荷GFP的TALE激活物

[0676] 用于将成纤维细胞细胞命运重新程序化到褐色或白色脂肪细胞细胞命运的主要目标在于从白色脂肪组织(WAT)到褐色脂肪组织(BAT)的切换,该切换由PRDM16和PPAR γ 的表达控制。将融合于+36GFP的有力的TALE转录激活物工程化,其靶向成纤维细胞中的PPAR γ 和PRDM16基因组序列。融合蛋白使用肝素柱和/或SEC纯化,且凝胶在130kD处显示单个条带。将投递TALE激活物后的表达调控和对细胞表型的影响与在病毒投递PPAR γ cDNA,继之

以用脂肪生成混合物 (cocktail) 处理7天后的调控相比较。观察到在用+36GFP TALEPRDM16 融合物处理后形成脂肪细胞。在投递超电荷的PRDM16 TALE激活物后检测到白色脂肪组织标志物基因的表达。

[0677] 发现针对PPAR γ 的TALE激活物的超电荷蛋白介导的一次性投递诱导白色脂肪基因的表达和将成纤维细胞分化成白色脂肪细胞。PPAR γ 和PRDM16TALE激活物两者的超电荷蛋白介导的投递诱导脂肪细胞的分化,其具有褐色脂肪标志物如PRDM16,cox8b,elov13和cidea的增加的表达以及生热基因表达标志物PGC1a和UCP1的小量增加。

[0678] Aurein肽融合于+36GFP-TALE-激活物融合蛋白的N末端。通过检测经处理细胞的细胞核中的荧光来观察通过肝素柱纯化的Aurein+36GFP TALE的投递。

[0679] 图1显示大分子投递到哺乳动物细胞中的示意图。图2显示从白色脂肪组织 (WAT) 到褐色脂肪组织 (BAT) 的转换的概图。图3显示超电荷的投递平台的示意图,其投递程序化为靶向PPAR γ 或PRDM16的TALE激活物。图4显示融合蛋白的示意图,其包含+36GFP融合、18.5聚体TALE域和VP64激活域。图5显示+36GFP-TALE激活物融合蛋白的表达和纯化。图6显示投递+36GFP PPAR γ 和PRDM16 TALE激活物融合蛋白后用于激活脂肪细胞调控物基因的测试测定法。

[0680] 图7显示在不同浓度+36GFP TALE激活物融合蛋白的投递功效。图8显示两种不同的+36GFP-PRDM16 TALE融合蛋白在NIH 3T3细胞中的投递功效的比较。图9显示在投递PPAR γ -TALE激活物融合物后的PPAR γ 基因表达和与多种对照的比较。图10显示在投递RDM16-TALE激活物融合物后的PRDM16基因表达和与多种对照的比较。图11显示在存在血清情况下观察到的适度TALE活性。

[0681] 图12显示病毒投递PPAR γ ,接着用脂肪生成混合物处理7天的验证。图13显示用于程序化成纤维细胞为WAT和BAT的测定法的示意图。图14显示在用+36GFP TALE激活物融合蛋白处理后观察到的脂肪细胞形成。图15显示在用LipidTOX red染色7天后多种处理,其证明在病毒投递后以及超电荷PPAR γ TALE激活物融合蛋白的投递后形成脂肪细胞。图16显示用脂质TOX red染色在多种处理后7天后的细胞,其证明在病毒投递后以及超电荷PPAR γ TALE激活物融合蛋白的投递后形成脂肪细胞。图17显示在病毒投递后以及超电荷PPAR γ TALE激活物融合蛋白的投递后WAT生物标志物基因的表达。

[0682] 实施例2:体内投递融合于超电荷的GFP的TALE激活物

[0683] 将NIH 3T3细胞培养至70-90%汇合并用在无血清的DMEM中1 μ M或0.5-5 μ M的+36GFP PPAR γ TALE和/或+36GFP PRDM16 TALE融合蛋白处理。选择无血清培养基,因为血清可能降低基于蛋白投递的有效性。将细胞与相应的融合蛋白溶液温育4小时,接着除去培养基并将含有血清的完整DMEM添加回细胞。根据已知方案将对照细胞用编码PPAR γ 或PRDM16的病毒构建体感染以充当WAT和BAT基因的表达的阳性对照(参见,例如Seale等2008Nature 454,961-967,其完整内容通过提述并入本文)。一旦所有细胞达到100%汇合,将含有异丁基甲基黄嘌呤、胰岛素、罗格列酮(rosiglitazone)、地塞米松(dexamethosone)、T3和茚甲新(indomethacin)的脂肪生成混合物添加至细胞并在48小时后用仅含有胰岛素、T3和罗格列酮的混合物形式替换。在混合物的第二轮此替换后48小时,将T3、胰岛素和罗格列酮的另一剂量添加至细胞。次日(即从实验开始到现在一周),用TRIzol收获细胞,提取总RNA,并实施qRT-PCR以测量PPAR γ ,PRDM16和其他褐色脂肪标志物基因如UCP1,PGC1a,Elov13和

Cidea的基因表达水平。

[0684] 图18显示体内投递超电荷的PRDM16 TALE激活物融合蛋白以诱导褐色脂肪脂肪细胞。在PPAR γ 和PRDM16的病毒投递以及还有超电荷的TALE激活物蛋白融合物的投递后观察到有力的脂肪细胞形成。图19显示对细胞的超电荷的(TALE)和病毒投递PPAR γ 和PRDM16的比较。该图显示通过PPAR γ 和PRDM16的表达,TALE/TALE、病毒/TALE和病毒/病毒诱导的褐色脂肪标志物的表达。图20显示与脂肪细胞分化一致的RT-qPCR评估,所述细胞分化也通过脂质TOX染色观察到。

[0685] 实施例3:投递与超电荷的GFP复合的TALE激活物

[0686] 为了改进投递功效,生成其中功能性蛋白与超电荷蛋白非共价联合的蛋白复合物并施用给细胞。图21显示以与+36GFP的复合物的形式投递功能性TALE激活物融合蛋白在投递后改进了TALE激活物活性。图22显示在作为与+36GFP的融合物(+36GFP PRDM16 TALE-3)或复合物(+36GFP+PRDM16TALE-3)的TALE激活物融合物投递后的PRDM16基因表达。观察到复合物的投递倾向于增加TALE激活物活性。

[0687] 实施例4:Aurein融合对投递功效的影响

[0688] 图23显示N末端Aurein肽融合于+36GFP在TALE激活物融合物投递(作为与+36GFP的融合物或复合物)后对PRDM16基因表达的影响。Aurein肽经由GGS(9)(SEQ ID NO:252)接头融合于GFP-TALE构建体的N末端,产生Aurein肽-GGS(9)接头-(+)36GFP蛋白-GGS(9)接头-PRDM16 TALE-3融合蛋白。使用大小排阻层析纯化该蛋白。

[0689] 实施例5:投递与超电荷的GFP或阳离子脂质复合的TALE激活物

[0690] 图24显示在以与+36GFP的融合物(+36GFP PRDM16 TALE-3),与+36GFP的复合物(+36GFP+PRDM16 TALE-3),或与Lipofectamine LTX的复合物(对其观察到基因表达的增加)投递TALE PRDM16激活物蛋白后的PRDM16基因表达。

[0691] 实施例6:投递融合于超电荷的GFP的Cas9

[0692] Cas9到哺乳细胞中的超电荷投递将允许将强大的可RNA编程的核酸酶技术应用到细胞中,却没有先前的投递方法的缺点。为此,使用ALAL接头生成具有+36GFP的Cas9融合物。图25显示具有Cas9的超电荷的融合蛋白的示意图。图26显示纯化野生型Cas9蛋白以及具有+36GFP和Aurein-GGS9的Cas9融合蛋白。将融合蛋白以与上文TALE激活物融合蛋白相同的方式施用给细胞。Cas9一旦投递给细胞,就结合并切割其在细胞基因组中的靶位点。经由合适的测定法,经由southern印迹或测序测定法检测到靶细胞中的核酸酶活性。

[0693] 实施例7:基因组编辑蛋白在体外和体内的有效投递

[0694] 需要蛋白到细胞核或细胞质的有效胞内投递以完全实现包括基因组编辑药剂在内的蛋白治疗物的潜力。目前的蛋白质投递方法经常受限于对血清蛋白的低容忍,较差的核内体逃离,和有限的体内功效。如本实施例证明的,最初开发用于核酸转染的常见阳离子脂质试剂能有力地投递融合于超负电荷蛋白(其含有天然阴离子域或天然结合阴离子核酸)的蛋白。该办法介导Cre重组酶、基于TALE和Cas9的转录激活物,和Cas9:sgRNA核酸酶复合物以低纳米摩尔浓度在含有10%血清的培养基中功能性投递到培养的人细胞中。基于脂质的投递可以比阳离子蛋白投递策略强大>1,000倍。Cas9:sgRNA复合物的投递导致高达80%的基因组修饰效率和相比于标准DNA转染实质性更高的特异性,这可能是由于投递的Cas9:sgRNA复合物的瞬时性质。该办法还介导Cre重组酶和Cas9:sgRNA复合物有效体内投

递到小鼠内耳中,在靶向的毛细胞群体中实现高达90%的Cre介导的充足和20% Cas9介导的基因组修饰。

[0695] 材料和方法

[0696] 构建Cas9,Cre和TALE融合物及sgRNA表达质粒。

[0697] 本文中使用的所有构建体的序列在下文列出或在说明书的别处提供。所有蛋白构建体均从先前报道的用于克隆到pET29a表达质粒的感兴趣的蛋白质的质粒生成。

[0698] 表达和纯化酿脓链球菌Cas9和其他蛋白质。

[0699] 将大肠杆菌BL21 STAR(DE3)感受态细胞(Life Technologies)用pMJ806⁴⁷转化,其编码酿脓链球菌Cas9融合于N末端10xHis标签/麦芽糖结合蛋白。将所得表达菌株接种到含有100 μ g/mL氨苄青霉素的Luria-Bertani(LB)培养液中,在37 $^{\circ}$ C过夜。将细胞以1:100稀释到相同的培养基中并在37 $^{\circ}$ C生长至 $OD_{600} \sim 0.6$ 。将培养物在20 $^{\circ}$ C温育30min,并以0.5mM添加异丙基 β -D-1-硫代半乳糖苷(IPTG)以诱导Cas9表达。在约16小时后,通过在8,000g离心收集细胞并重悬于裂解缓冲液(50mM Tris(羟甲基)-氨基甲烷(Tris)-HCl,pH 8.0,1M NaCl,20%甘油,10mM Tris(2-羧乙基)膦(TCEP))。通过超声裂解细胞(1秒脉冲开,1秒脉冲关,总计达15分钟,6W输出),且通过在20,000g离心30min获得可溶裂解物。

[0700] 将细胞裂解物与His-Pur镍-氮川乙酸(镍-NTA)树脂(Thermo Scientific)在4 $^{\circ}$ C温育30分钟以捕捉带His标签的Cas9。将树脂转移至20-mL柱并用20个柱体积的裂解缓冲液清洗。Cas9在50mM Tris-HCl(pH 8),0.1M KCl,20%甘油,1mM TCEP和300mM咪唑中洗脱,并通过Amicon ultra离心滤器(Millipore,100-kDa分子量截留)浓缩至 ~ 50 mg/mL。除去6xHis标签和麦芽糖结合蛋白,其通过在4 $^{\circ}$ C用TEV蛋白酶处理20小时,并通过第二Ni亲和纯化步骤捕捉。将含有Cas9的洗脱物注入到HiTrap SP HP柱(GE Healthcare)中,该柱在含有50mM Tris-HCl(pH 8),0.1M KCl,20%甘油和10mM TCEP的纯化缓冲液中。将Cas9用含有从0.1M至1M的线性NaCl梯度的纯化缓冲液在5个柱体积中洗脱。将洗脱的含有Cas9的级分浓缩至200 μ M的浓度(如通过二辛可宁酸(Bicinchoninic acid)测定法(BCA)(Pierce Biotechnology)量化的),在液氮中速冻,并以等分试样储藏于-80 $^{\circ}$ C。所有其他蛋白均通过此方法纯化,但是没有TEV切割步骤,且含有(-30)GFP的蛋白通过阴离子交换纯化,其使用相同的纯化方案使用Hi-Trap Q HP阴离子交换柱(GE Healthcare)进行。

[0701] sgRNA的体外转录。

[0702] 依照制造商用法说明使用T7 High Yield RNA合成试剂盒(NEB),将含有T7启动子结合位点之后是20-bp sgRNA靶序列的线性DNA片段体外转录。体外转录的RNA用乙醇沉淀并通过凝胶电泳在Criterion 10%聚丙烯酰胺TBE-Urea凝胶(Bio-Rad)上纯化。在4 $^{\circ}$ C在摇动表面上将切取的凝胶片段过夜提取到420 μ L的300mM NaCl中。将凝胶纯化的sgRNA用乙醇沉淀并再溶于水,并通过UV吸光度最终量化sgRNA浓度,速冻于-80 $^{\circ}$ C。

[0703] 质粒转染。

[0704] 使用Lipofectamine 2000(Life Technologies)按照制造商方案转染质粒DNA。对于TALE激活物质粒,转染300ng的DNA,且对于激活物协同实验,合并60ng的5种质粒的每种并转染。对于Cas9核酸酶投递实验,将表达sgRNA的线性DNA PCR产物用于转染实验中,其靶向CLTA,EMX,VEGF和GFP中的基因组位点(sgRNA GFP g1,GFP g3,GFP g5和GFP g7用于切口酶研究)。生成线性DNA PCR产物,其使用含有U6启动子的质粒作为模板和携带U6启动子上

游序列的正向引物和含有U6下游序列继之以sgRNA序列(对每个靶物独特的20-bp序列加上恒定的sgRNA骨架体系结构序列)的逆向引物。从线性DNA模板表达的sgRNA含有至少两个5'鸟苷以匹配体外转录的sgRNA,这些碱基是T7转录所需的。引物序列和PCR条件在下文列出。对于dCas9激活物实验,将700ng的Cas9或dCas9-VP64质粒DNA与250ng的适宜sgRNA表达质粒共转染。对于激活物协同实验,合并来自6种sgRNA的每种的50ng DNA,并与700ng的dCas9-VP64质粒共转染。

[0705] 在细胞培养物中投递与阳离子脂质复合的转录因子蛋白。

[0706] 对体外和体内两者的基因组编辑蛋白的投递的更深度的描述可见于下文。简言之,将培养的细胞以次日达到~70%汇合所需的细胞密度铺板于48孔形式(250 μ L体积)的具有10% FBS(“完全血清培养基”)和抗生素的Dulbecco氏改良的Eagle's Media加GlutaMAX(Life Technologies,Carlsbad,CA)中。在投递前一小时,将完全血清培养基用不含有抗生素的相同培养基替换。依照制造商用于正常质粒转染的方案(包括温育时间),通过将1nM至1 μ M蛋白(275 μ L终体积中)与0.5-1.5 μ L的商品化的阳离子脂质在25 μ L OPTIMEM培养基(Life Technologies,Carlsbad,CA)中组合来实施Cre和TALE蛋白的投递。对于体外Cas9投递,将转录的sgRNA与Cas9蛋白温育5min,接着与阳离子脂质试剂复合。将25 μ L OPTIMEM培养基中的脂质复合物添加至细胞并在12-16小时后用新鲜培养基替换培养基,除非另外说明。在投递后48小时对细胞测定重组,在投递后4或16小时测定基因活化,和在投递后72小时测定基因修饰。

[0707] 检测基因组修饰的T7内切核酸酶I测定法。

[0708] 将U2OS-EGFP细胞或HEK293T细胞用Cas9表达和sgRNA表达质粒或线性DNA PCR产物转染(如上文描述的)或仅用Cas9蛋白,仅用体外转录的sgRNA或仅用RNAiMAX处理。转染后2天,使用DNAdvance试剂盒(Agencourt)依照制造商用法说明从细胞分离基因组DNA。将200ng的基因组DNA用作PCR反应中的模板以扩增靶向的基因组基因座,其使用如下文制定的侧翼侦测引物对。使用QIAquick PCR纯化试剂盒(Qiagen)纯化PCR产物并使用Quant-iT™ PicoGreen® dsDNA试剂盒(Life Technologies)量化。将250ng纯化的PCR DNA与2 μ L NEBuffer 2(NEB)在总体积19 μ L中组合并变性,然后采用以下热循环重退火:95 $^{\circ}$ C 5分钟,95至85 $^{\circ}$ C以2 $^{\circ}$ C/s;85至20 $^{\circ}$ C以0.2 $^{\circ}$ C/s。将重退火的DNA与1 μ L T7内切核酸酶I(10U/ μ L,NEB)在37 $^{\circ}$ C温育15分钟。将10 μ L 50%甘油添加至T7内切核酸酶反应,并在以200V电泳30分钟的5%TBE 18孔Criterion PAGE凝胶(Bio-Rad)上分析12 μ L,然后用1x SYBR Gold(Life Technologies)染色30min。在AlphaImager HP(Alpha Innotech)上使Cas9诱导的切割条带和未切割的条带可见,并使用ImageJ软件⁵⁴量化。将切割条带的峰值强度除以所有条带(未切割+切割的条带)的总强度来确定切割的分数,其用来如先前描述的⁴⁶估测基因修饰水平。对于每份样品,在不同日一式三份地实施转染和后续修饰测量。

[0709] 干细胞培养和投递。

[0710] 将含有永久GFP基因插入的小鼠胚胎干细胞(ES)系Tau-GFP在具有15% FBS(Gibco),100mM MEM非必需氨基酸(Gibco),0.55mM 2-巯基乙醇和白血病抑制因子(1,000units/ml;Chemicon)的DMEM中培养。5天后,形成展现GFP荧光的浮球。将Cas9:sgRNA和RNAiMAX的复合物添加至含有浮球的培养物达16小时。在Cas9:sgRNA处理后,将细胞在上述培养基中培养3天。将浮球用胰蛋白酶处理5分钟,然后通过70 μ m滤器以收集单细胞。在标记

之前,将细胞在用核纤层蛋白包被的载玻片上在补充有1xN2、1xB27、青霉素-链霉素(100g/mL)和10% FBS的DMEM/F12(1:1)中培养2天。使用抗GFP抗体(#ab13970, Abcam)实施免疫组化以评估GFP表达。为了量化GFP阴性细胞的数目,我们从20X放大率的3个代表性视场中计算了GFP阳性和GFP阴性细胞的总数目,并计算了平均效率。对于每个条件实施了3次独立的实验。

[0711] 将蛋白微注射到小鼠内耳。

[0712] 将P0 floxP-tdTomato小鼠用于(-30)GFP-Cre注射并将P2 Atoh1-GFP小鼠用于Cas9:sgRNA注射。在马萨诸塞州Eye&Ear Infirmary ALCUC委员会批准的方案下使用动物。通过在冰上降低其温度来麻醉小鼠。通过在耳后制造切口以暴露耳泡(otic bulla)来实施耳蜗造口术(Cochleostomy)。使用显微操作器持握的玻璃微量移液器来将复合物投递到蜗管中,其允许到达内耳毛细胞。对于(-30)GFP-Cre的投递,将3 μ L的45 μ M蛋白与3 μ L的RNAiMAX或Lipofectamine 2000混合并在注射前在室温温育30分钟。每个处理组注射4只小鼠。对于Cas9:sgRNA复合物的投递,将1 μ L的200 μ M Cas9蛋白与2 μ L的100 μ M sgRNA混合并在室温温育5分钟,接着与3 μ L的RNAiMAX或Lipofectamine 2000混合并再温育30分钟,然后注射。每个处理组注射3只小鼠。每次注射的总投递体积为0.3 μ L每耳蜗,且释放由显微操作器控制在32nL/秒的速度。

[0713] 免疫组化和定量。

[0714] 注射后5-10天,通过标准方案⁵⁵杀死小鼠并收获耳蜗。对于免疫组化,遵循先前描述的方案⁵⁵来使用针对毛细胞标志物(Myo7a和Esp)和支持细胞(Sox2)的抗体。为了量化Cas9:sgRNA投递后(-30)GFP-Cre或GFP阴性细胞后tdTomato阳性细胞的数目,我们计算了在耳蜗的底部转角(base turn)中注射部位周围跨越200 μ m区域中的外毛细胞的总数目。将(-30)GFP-Cre诱导的重组或Cas9:sgRNA诱导的基因组修饰的效率计算为表达tdTomato或丧失GFP表达的外毛细胞的百分数。

[0715] 对基因组修饰的高通量DNA测序。

[0716] 将HEK293T细胞用Cas9和sgRNA表达质粒或线性DNA PCR产物转染,或用50nM Cas9蛋白、250nM纯化的sgRNA和阳离子脂质处理,如之前对于Cas9蛋白投递到U2OS-EGFP报道细胞所描述的。对于基于质粒的转染实验,将700ng的Cas9表达质粒加250ng的sgRNA质粒或50ng的线性DNAPCR产物(表达用于靶向EMX1, CLTA2或VEGF基因座的任一的sgRNA)与Lipofectamine 2000(Life Technologies)转染,并在2天后分离细胞。对于体内蛋白投递实验,如先前描述的⁵⁵从麻醉的小鼠分离~30mg的小鼠组织,并使用Agencourt DNAAdvance基因组DNA分离试剂盒(Beckman Coulter)提取基因组DNA。对于细胞培养实验,如上文描述的分离基因组DNA。将150ng的基因组DNA用作模板来通过PCR扩增在靶和脱靶基因组位点,其使用下文制定的侧翼HTS引物对。粗PCR产物的相对量通过凝胶电泳定量并将用不同sgRNA对或Cas9核酸酶类型处理的样品分别以等摩尔浓度合并,接着用QIAquick PCR纯化试剂盒(Qiagen)纯化。使用5% TBE 18孔Criterion PAGE凝胶(BioRad)将约150ng的合并的DNA在200V电泳30min,且分离长度为~125bp至~300bp的DNA并通过QIAquick PCR纯化试剂盒(Qiagen)纯化。使用含有测序接头的引物将纯化的DNA通过PCR扩增,纯化,并在MiSeq高通量DNA测序仪(Illumina)上测序,如先前描述的⁴⁷。

[0717] 数据分析

[0718] 如下文列出的使用以Unix Bash编写的脚本来过滤和解析Illumina测序读段。将用于测序实验的样本大小最大化(在实际实验考虑内)以确保检测效果的最大能力。如先前描述的⁵⁶实施对Cas9修饰的基因组位点的统计学分析(表2),其中使用Bonferroni方法进行多比较校正。

[0719] 下面是每个基因组靶位点的上游和下游侧翼序列的列表。

靶位点	下游基因组序列	上游基因组序列
EMX_On	GGCCTGCTTCGTGGCAAT GC (SEQ ID NO:119)	ACCTGGGCCAGGGAGGGA GG (SEQ ID NO:120)
EMX_Off1	CTCACTTAGACTTTCTCTC C (SEQ ID NO:121)	CTCGGAGTCTAGCTCCTGC A (SEQ ID NO:122)
EMX_Off2	TGGCCCCAGTCTCTCTTCT A (SEQ ID NO:123)	CAGCCTCTGAACAGCTCCC G (SEQ ID NO:124)
EMX_Off3	TGACTTGGCCTTTGTAGGA A (SEQ ID NO:125)	GAGGCTACTGAAACATAAG T (SEQ ID NO:126)
EMX_Off4	TGCTACCTGTACATCTGCA C (SEQ ID NO:127)	CATCAATGATTGGGCATTTTC (SEQ ID NO:128)
VEG_On	ACTCCAGTCCCAAATATGT A (SEQ ID NO:129)	ACTAGGGGGCGCTCGGCC AC (SEQ ID NO:130)
VEG_Off1	CTGAGTCAACTGTAAGCA	GGCCAGGTGCAGTGATTCA

[0720]

	TT (SEQ ID NO:131)	T (SEQ ID NO:132)
VEG_Off2	TCGTGTCATCTTGTGTTGTG C (SEQ ID NO:133)	GGCAGAGCCCAGCGGACA CT (SEQ ID NO:134)
VEG_Off3	CAAGGTGAGCCTGGGTCT GT (SEQ ID NO:135)	ATCACTGCCCAAGAAGTGC A (SEQ ID NO:136)
VEG_Off4	TTGTAGGATGTTTAGCAGC A (SEQ ID NO:137)	ACTTGCTCTCTTAGAGAA C (SEQ ID NO:138)
CLT2_On	CTCAAGCAGGCCCGCTG GT (SEQ ID NO:139)	TTTTGGACCAAACCTTTTT G (SEQ ID NO:140)
[0721] CLT2_Off1	TGAGGTTATTTGTCCATTG T (SEQ ID NO:141)	TAAGGGGAGTATTTACACC A (SEQ ID NO:142)
CLT2_Off2	TCAAGAGCAGAAAATGTG AC (SEQ ID NO:143)	CTTGCAGGGACCTTCTGAT T (SEQ ID NO:144)
CLT2_Off3	TGTGTGTAGGACTAAACTC T (SEQ ID NO:145)	GATAGCAGTATGACCTTGG G (SEQ ID NO:146)
EGFP	AGCGTGTCCGGCGAGGGC GA (SEQ ID NO:147)	AGCGTGTCCGGCGAGGGC GA (SEQ ID NO:148)
MusEMX	CAGAATCGGAGGACAAAA TACAAAC (SEQ ID NO:149)	ACGAAGCAGGCCAACGGG GAGGACA (SEQ ID NO:150)

[0722] 用于生成PCR产物以充当sgRNA的T7转录的底物的引物。

[0723] T7_gRNA-Rev在所有情况中使用。使用的DNA模板是如上文说明的EGFP sgRNA质粒。用于dCas9-VP64激活物实验的NTF3和VEGF sgRNA之前有报道 (Maeder等, CRISPR RNA-guided activation of endogenous human genes. Nat. Methods. 2013;10,977-979)。

[0724] T7_EGFP1-Fwd TAA TAC GAC TCA CTA TA GGGCACGGGCAGCTTGCCGG
(SEQ ID NO:151)
T7-GFP g1-Fwd TAA TAC GAC TCA CTA TA GGCCTCGAACTTCACCTCGGCG
GAAAGGACGAAACACC (SEQ ID NO:152)

	T7-GFP g5-Fwd	TAA TAC GAC TCA CTA TA GGCTGAAGGGCATCGACTTCA GAAAGGACGAAACACC (SEQ ID NO:153)
	T7-GFP g3-Fwd	TAA TAC GAC TCA CTA TA GGCAGCTCGATGCGGTTACCA GAAAGGACGAAACACC (SEQ ID NO:154)
	T7-GFP g7-Fwd	TAA TAC GAC TCA CTA TA GGCAAGGAGGACGGCAACATCC GAAAGGACGAAACACC (SEQ ID NO:155)
[0725]	T7-EMX-Fwd	TAA TAC GAC TCA CTA TA GGAGTCCGAGCAGAAGAAGAA GAAAGGACGAAACACC (SEQ ID NO:156)
	T7-VEG-Fwd	TAA TAC GAC TCA CTA TA GGGGTGGGGGAGTTTGCTCC GAAAGGACGAAACACC (SEQ ID NO:157)
	T7-CLT2-Fwd	TAA TAC GAC TCA CTA TA GGCAGATGTAGTGTTCACCA GAAAGGACGAAACACC (SEQ ID NO:158)
	T7_gRNA-Rev	AAAAAAAGCACCGACTCGGTG (SEQ ID NO:159)

[0726] 用于生成线性DNA PCR产物用于转染的引物。

[0727] 在作为模板的含有U6启动子的质粒上使用下文列出的PCR_sgRNA-fwd1, PCR_sgRNA-rev2和适宜的PCR_sgRNA引物的PCR延伸(72°C, 3min)。

PCR_gRNA-fwd1 CTGTACAAAAAGCAGGCTTTA (SEQ ID NO:160)

PCR_gRNA-rev2 AAAAAAGCACCGACTCGGTGCCACTTTTCAAGTTGATAACGG
ACTAGCCTATTTTAACTTGCTATTTCTAGCTCTAAAC (SEQ ID NO:161)

PCR-G-GFP1 GAAAGGACGAAACACC
GGCCTCGAACTTCACCTCGGCGGTTTTAGAGCTAGAAATAGCAA (SEQ ID
NO:162)

PCR-G-GFP3 GAAAGGACGAAACACC
GGCAGCTCGATGCGGTTACCAGTTTTAGAGCTAGAAATAGCAA (SEQ ID
NO:163)

[0728] PCR-G-GFP5 GAAAGGACGAAACACC
GGCTGAAGGGCATCGACTTCAGTTTTAGAGCTAGAAATAGCAA (SEQ ID NO:164)

PCR-G-GFP7 GAAAGGACGAAACACC
GGCAAGGAGGACGGCAACATCCGTTTTAGAGCTAGAAATAGCAA (SEQ ID
NO:165)

PCR-G-CLT2 GAAAGGACGAAACACC
GGCAGATGTAGTGTTCACAGTTTTAGAGCTAGAAATAGCAA (SEQ ID NO:166)

PCR-G-EMX GAAAGGACGAAACACC
GGAGTCCGAGCAGAAGAAGAAGTTTTAGAGCTAGAAATAGCAA (SEQ ID
NO:167)

PCR-G-VEG GAAAGGACGAAACACC
GGGGTGGGGGAGTTTGCTCCGTTTTAGAGCTAGAAATAGCAA (SEQ ID NO:168)

[0729] 用于实施T7内切核酸酶I DNA切割测定法的引物。

[0730] Survey_GFP-fwd TACGGCAAGCTGACCCTGAA (SEQ ID NO:169)

[0731] Survey_GFP-rev GTCCATGCCGAGAGTGATCC (SEQ ID NO:170)

[0732] Survey_CLTA-fwd GCCAGGGCTGTTATCTTGG (SEQ ID NO:171)

[0733] Survey_CLTA-rev ATGCACAGAAGCACAGGTTGA (SEQ ID NO:172)

[0734] Survey_EMX-fwd CTGTGCTCTTCTGCGCT (SEQ ID NO:173)

- [0735] Survey_EMX-rev CTCTCCGAGGAGAAGGCCAA (SEQ ID NO:174)
- [0736] Survey_VEGF-fwd CCACACAGCTTCCCGTTCTC (SEQ ID NO:175)
- [0737] Survey_VEGF-rev GAGAGCCGTTCCCTCTTTGC (SEQ ID NO:176)
- [0738] 用于对人基因组中的在靶和脱靶靶位点的高通量测序的引物.
- HTS_EMX_ON-fwd CACTCTTTCCCTACACGACGCTCTTCCGATCT
wd CCTCCCCATTGGCCTGCTTC (SEQ ID NO:177)
- HTS_EMX_Off1-fwd CACTCTTTCCCTACACGACGCTCTTCCGATCT
wd TCGTCCTGCTCTCACTTAGAC (SEQ ID NO:178)
- HTS_EMX_Off2-fwd CACTCTTTCCCTACACGACGCTCTTCCGATCT
wd TTTTGTGGCTTGGCCCCAGT (SEQ ID NO:179)
- HTS_EMX_Off3-fwd CACTCTTTCCCTACACGACGCTCTTCCGATCT
wd TGCAGTCTCATGACTTGGCCT (SEQ ID NO:180)
- [0739] HTS_EMX_Off4-fwd CACTCTTTCCCTACACGACGCTCTTCCGATCT
wd TTCTGAGGGCTGCTACCTGT (SEQ ID NO:181)
- HTS_VEGF_ON-fwd CACTCTTTCCCTACACGACGCTCTTCCGATCT
wd ACATGAAGCAACTCCAGTCCCA (SEQ ID NO:182)
- HTS_VEGF_Off1-fwd CACTCTTTCCCTACACGACGCTCTTCCGATCT
-fwd AGCAGACCCACTGAGTCAACTG (SEQ ID NO:183)
- HTS_VEGF_Off2-fwd CACTCTTTCCCTACACGACGCTCTTCCGATCT
-fwd CCCGCCACAGTCGTGTCAT (SEQ ID NO:184)
- HTS_VEGF_Off3-fwd CACTCTTTCCCTACACGACGCTCTTCCGATCT
-fwd CGCCCCGGTACAAGGTGA (SEQ ID NO:185)

HTS_VEGF_Off4 CACTCTTTCCCTACACGACGCTCTTCCGATCT
-fwd GTACCGTACATTGTAGGATGTTT (SEQ ID NO:186)

HTS_CLTA2_ON- CACTCTTTCCCTACACGACGCTCTTCCGATCT
fwd CCTCATCTCCCTCAAGCAGGC (SEQ ID NO:187)

HTS_CLTA2_Off CACTCTTTCCCTACACGACGCTCTTCCGATCT
1-fwd ATTCTGCTCTTGAGGTTATTTGT (SEQ ID NO:188)

HTS_CLTA2_Off CACTCTTTCCCTACACGACGCTCTTCCGATCT
2-fwd CACCTCTGCCTCAAGAGCAGAAAA (SEQ ID
NO:189)

HTS_CLTA2_Off CACTCTTTCCCTACACGACGCTCTTCCGATCT
3-fwd TGTGTGTGTGTGTGTGTAGGACT (SEQ ID NO:190)

HTS_EMX_ON-r GGAGTTCAGACGTGTGCTCTTCCGATCT
ev TCATCTGTGCCCTCCCTCC (SEQ ID NO:191)

HTS_EMX_Off-re GGAGTTCAGACGTGTGCTCTTCCGATCT
v CGAGAAGGAGGTGCAGGAG (SEQ ID NO:192)

[0740] HTS_EMX_Off-re GGAGTTCAGACGTGTGCTCTTCCGATCT
v CGGGAGCTGTTCAGAGGCTG (SEQ ID NO:193)

HTS_EMX_Off-re GGAGTTCAGACGTGTGCTCTTCCGATCT
v CTCACCTGGGCGAGAAAGGT (SEQ ID NO:194)

HTS_EMX_Off-re GGAGTTCAGACGTGTGCTCTTCCGATCT
v AAAACTCAAAGAAATGCCCAATCA (SEQ ID
NO:195)

HTS_VEGF_ON-r GGAGTTCAGACGTGTGCTCTTCCGATCT
ev AGACGCTGCTCGCTCCATTC (SEQ ID NO:196)

HTS_VEGF_Off1 GGAGTTCAGACGTGTGCTCTTCCGATCT
-rev ACAGGCATGAATCACTGCACCT (SEQ ID NO:197)

HTS_VEGF_Off2 GGAGTTCAGACGTGTGCTCTTCCGATCT
-rev GCGGCAACTTCAGACAACCGA (SEQ ID NO:198)

HTS_VEGF_Off3 GGAGTTCAGACGTGTGCTCTTCCGATCT
-rev GACCCAGGGGCACCAGTT (SEQ ID NO:199)

HTS_VEGF_Off4 GGAGTTCAGACGTGTGCTCTTCCGATCT

	-rev	CTGCCTTCATTGCTTAAAAGTGGAT (SEQ ID NO:200)
	HTS_CLTA2_ON-	GGAGTTCAGACGTGTGCTCTTCCGATCT
	rev	ACAGTTGAAGGAAGGAAACATGC (SEQ ID NO:201)
	HTS_CLTA2_Off	GGAGTTCAGACGTGTGCTCTTCCGATCT
	1-rev	GCTGCATTTGCCATTTCCA (SEQ ID NO:202)
	HTS_CLTA2_Off	GGAGTTCAGACGTGTGCTCTTCCGATCT
[0741]	2-rev	GTTGGGGGAGGAGGAGCTTAT (SEQ ID NO:203)
	HTS_CLTA2_Off	GGAGTTCAGACGTGTGCTCTTCCGATCT
	3-rev	CTAAGAGCTATAAGGGCAAATGACT (SEQ ID NO:204)
	HTS_EGFP-fwd	CACTCTTCCCTACACGACGCTCTTCCGATCTNNNN ACGTAAACGGCCACAAGTTC (SEQ ID NO:205)
	HTS_EGFP-rev	GGAGTTCAGACGTGTGCTCTTCCGATCT GTCGTCCTTGAAGAAGATGGTG (SEQ ID NO:206)
	HTS_MusEMX_ON-	CACTCTTCCCTACACGACGCTCTTCCGATCT
	fwd	CCAGGTGAAGGTGTGGTTCCAG (SEQ ID NO:207)
	HTS_MusEMX_ON-	GGAGTTCAGACGTGTGCTCTTCCGATCT
	rev	CCCCTAGTCATTGGAGGTGAC (SEQ ID NO:208)

[0742] 结果

[0743] 融合于带超负电荷的蛋白的Cre重组酶的高度有效投递

[0744] 推测使用阳离子脂质将核酸的高度阴离子的静电特性赋予基因组编辑蛋白质可以实现其有效投递到哺乳动物细胞中(图27(A))。对于不天然高度带负电荷的感兴趣的蛋白质,认为与天然或工程化的超带负电荷的蛋白17融合将赋予聚阴离子性质。对于核酸结合蛋白,推测与天然DNA或RNA底物的简单复合可以提供充足的阴离子性质来支持基于阳离子脂质的投递(图27(A))。

[0745] 首先测试工程化的超带负电荷的GFP变体³⁵(-30)GFP是否能介导融合的蛋白货物的包封和投递(图27(B))。将(-30)GFP融合于Cre重组酶和几种商品化的阳离子脂质,并测试其将融合物功能性投递到仅在Cre介导的重组时表达DsRed的HeLa细胞中(图28(A))。在含有10%胎牛血清(FBS)的培养基中投递与1.5 μ L Lipofectamine RNAiMAX(后文成为“RNAiMAX”,Life Technologies,Carlsbad CA)复合的10nM(-30)GFP-Cre导致在经处理细胞中的强DsRed荧光信号。荧光活化的细胞分选(FACS)揭示在处理48小时时,52%的细胞表达DsRed,与Cre重组一致(图28(B))。

[0746] 使用在含有10% FBS的250 μ L培养基中与1.5 μ L RNAiMAX复合的25nM(-30)GFP-Cre,优化导致65%的重组效率(图28(C))。脂质介导的阴离子Cre投递的效力相比于阳离子蛋白介导的投递的效力是显著的。仅需要1nM(-30)GFP-Cre和阳离子脂质来产生15-20%重组的细胞,而需要1 μ M(+36)GFP-Cre来实现这一重组程度,代表了投递效力中的1,000倍差

异(图28(C))。在第二种Cre报道细胞系(BSR TdTomato)中观察到了几乎相同的结果(图33(A))。增加阳离子脂质的量增加了毒性(图33(B))且发现1.5 μ L RNAiMAX每250 μ L样品将重组效率最大化且诱导最小的细胞毒性。在这些条件下,阳离子脂质不增加中性或阳离子Cre重组酶融合物的投递效力(图28(C)和图33(C)),表明需要(-30)GFP-Cre的强负电荷来参与阳离子脂质介导的投递,还观察到增加阳离子脂质的量增加了最大重组所需的蛋白质的浓度,这与其中可投递蛋白质与阳离子脂质的特定化学计量(stoichiometry)复合的模型一致(图28(D))。这些观察统一指示阳离子脂质可能介导聚阴离子蛋白到哺乳动物细胞中的有力投递,即使在存在血清的情况下。

[0747] 为了确定阳离子脂质介导的(-30)GFP-Cre投递相对于阳离子蛋白介导的投递的更高效力来自细胞的更多的总蛋白摄取,或来自更高部分的进入细胞的功能性、非核内体蛋白分子,使用流式细胞术来测量用(+36)GFP-Cre或脂质体(-30)GFP-Cre在其相应的最佳Cre投递条件下处理的细胞的GFP荧光。对细胞荧光和重组效率的比较揭示(-30)GFP-Cre的脂质介导的功能投递比(+36)GFP-Cre的投递为每份量的内吞蛋白效力高9,800倍(图34)。总之,这些结果表明阴离子蛋白的脂质介导投递的不寻常的高效力不是来自于每个细胞中的异常高的蛋白摄取,而是来自于内吞作用后过程,其可能包括核内体逃离到细胞质中和避免溶酶体蛋白降解。

[0748] 为了测试投递聚阴离子蛋白的能力是否依赖于RNAiMAX中的专有组分或者其他阳离子脂质是否能介导类似的有力投递,测试了设计为投递核酸的几种其他转染试剂(图28(E))。尽管RNAiMAX仍然是(-30)GFP-Cre的最有效的功能性投递试剂,但其他阳离子脂质制剂也导致有力投递。Lipofectamine2000和Lipofectamine LTX(Life Technologies, Carlsbad CA),两种基于阳离子脂质制剂的质粒转染试剂²¹,和SAINT-Red(Synvolux Therapeutics, Groningen Netherlands),一种含有合成的含pyridium的阳离子脂质的siRNA投递制剂,均在一定浓度范围内导致较强的功能性(-30)GFP-Cre投递(图28(E))。相比之下,未观察到使用阳离子脂质DOTAP(Roche Diagnostics, Indianapolis IN)或基于肽的核酸投递试剂EZ-PLEX(Ascension Bio, Tampa FL)的强投递(图28(E))。这些观察统一指示几种(并非所有)阳离子脂质能包封和投递带负电荷的蛋白质至人细胞中。

[0749] 推测应可以使用阳离子脂质来投递(-30)GFP以外的聚阴离子蛋白。常用语生物学研究中的工程化的聚阴离子蛋白域包括VP64激活域(-22净理论电荷),其广泛用于与工程化的锌指阵列、TALE重复阵列或用于转录激活的dCas9的融合中,和3x FLAG(-7净理论电荷),一种用于蛋白纯化和可视化的表位标签(图28(F))。观察到VP64和3x FLAG两者增强使用阳离子脂质的Cre重组酶的功能性投递,尽管没有(-30)GFP那么有效,可能是由于其更低的总体负电荷(图33(C))。这些观察证明除(-30)GFP以外的不寻常的带负电荷的蛋白能介导到哺乳动物细胞中的基于阳离子脂质的有效投递。

[0750] TALE激活物蛋白的功能性投递

[0751] 测试了TALE-VP64转录激活物(约+4理论净电荷,依赖于使用的TALE变体)到培养的人细胞中的脂质介导的投递。尽管已在哺乳动物细胞中在缺少血清的情况下使用阳离子肽(如Arg₉³⁶)证明了通过投递的TALEN蛋白进行的适度有效的内源性基因切割,但尚未报道基于TALE的转录因子蛋白的投递,且据我们所知,尚未有在血清中的TALE蛋白的有效投递。靶向神经营养因子(neurotrophin)-3(NTF3)的基因,该蛋白是一种与神经退行性疾病有关

的神经生长因子³⁷。将之前描述的靶向NTF3的TALE-VP64³⁸融合于(-30)GFP(图29(A))并将HEK293T细胞用25nM(-30)GFP-NTF3 TALE1-VP64和RNAiMAX在针对Cre投递优化的条件下处理。在用25nM(-30)GFP-NTF3TALE-VP64和RNAiMAX处理的细胞中,处理后4小时NTF3的基因表达水平比未经处理的细胞、仅用RNAiMAX处理的细胞或用VEGF靶向性TALE转录激活物处理的细胞高3.5倍(图29(B))。在转染编码相同NTF3靶向性TALE-VP64的质粒后48小时观察到NTF3表达的可比的水平(图29(B))。

[0752] 由于已显示靶向同一基因上的不同位点的多个TALE激活物的协同表达增加基因活化³⁸,使用RNAiMAX同时投递融合于(-30)GFP的5种不同的NTF3靶向性TALE激活物。如上文所述制备蛋白质-脂质复合物,其通过以5nM各自添加5种(-30)GFP-NTF3-TALE-VP64蛋白,总共25nM蛋白。在4小时温育后观察到NTF3表达中的6.5倍增加(图29(B)和图35),而所有5种NTF3TALE激活物的质粒共转染继之以48小时温育,导致NTF3表达水平的10倍增加(图29(B))。这些发现证明TALE激活物蛋白可使用阳离子脂质投递以短暂激活人细胞中的基因表达。可编程的转录激活物蛋白的投递可以实现靶基因的一次性激活而同时避免长期基因表达,后者是可编程转录因子的基于DNA投递的一个通常的关注所在。该能力可对当短暂表达时在细胞状态或细胞命运中产生一次性永久变化的蛋白质特别有价值³⁹。

[0753] Cas9:sgRNA蛋白:RNA复合物到人细胞中的高度有效投递

[0754] 鉴于聚阴离子Cre和TALE激活物蛋白变体在全血清培养基中的有力的脂质介导投递,推测CRISPR-Cas9:sgRNA复合物(作为与(-30)GFP的融合物或作为天然聚阴离子Cas9:导引RNA的复合物)也可以使用该办法投递到人细胞中。使用确立的Cas9诱导的基因破坏测定法⁴⁰,靶向人U2OS细胞中基因组EGFP报道基因内的特定定位点(图36(A))。在靶Cas9切割诱导EGFP中的非同源末端连接(NHEJ)和细胞荧光的丧失。为了避免(-30)GFP荧光的干扰,将Y67S突变引入(-30)GFP以消除其荧光,并称此非荧光变体为(-30)dGFP。

[0755] 将U2OS报道细胞用25nM(-30)dGFP-NLS-Cas9和50nM EGFP靶向性sgRNA和RNAiMAX在含有10% FBS的培养基中处理显示在48%的细胞中的EGFP表达的丧失(图30(A))。表达Cas9或sgRNA的质粒的共转染导致37%的细胞中的类似的EGFP丧失(图30(A))。在转染编码仅EGFP sgRNA,仅Cas9的质粒或共转染编码Cas9和sgRNA(设计为靶向VEGF基因座)的质粒后未观察到显著的EGFP破坏(图30(A),图36(B))。确认对EGFP的有力破坏不是细胞毒性的结果(图36(C)-(D))。还观察到在存在10% FBS血清的情况下将细胞用(+36)dGFP-NLS-Cas9和sgRNA处理不导致有效的基因破坏(图30(A)),表明投递Cas9和sgRNA的基于阳离子肽的方法无效,可能是由于带超正电荷的蛋白对gRNA:Cas9复合物形成或核酸酶功能的干扰⁴¹。总之,这些结果建立(-30)dGFP-NLS-Cas9:sgRNA复合物的阳离子脂质介导的投递能导致人细胞中的有效的sgRNA依赖性靶基因破坏。

[0756] 聚阴离子sgRNA对于有效的脂质介导的Cas9投递是必要且充分的

[0757] 由于天然Cas9蛋白(+22净理论电荷)和sgRNA(~103个阴离子磷酸根基团)的复合物应是总体高阴离子性的接下来测试没有融合于聚阴离子蛋白的天然Cas9:sgRNA复合物是否可使用阳离子脂质投递到人细胞中。将U2OS EGFP报道细胞用100nM Cas9,100nM EGFP sgRNA和0.8 μ L RNAiMAX处理导致EGFP报道基因的65%破坏(图30(A))。将细胞用Cas9蛋白和sgRNA但没有RNAiMAX处理导致GFP荧光没有丧失(图30(A))。这些观察表明仅sgRNA,甚至在缺少超带负电荷的融合蛋白的情况下,也能提供介导Cas9的基于阳离子脂质的投递所需

的高度阴离子性质。

[0758] 对来自 (-30) dGFP-NLS-Cas9:sgRNA相对于Cas9:sgRNA的阳离子脂质介导的投递的基因破坏效率的比较揭示低剂量 (-30) dGFP-NLS-Cas9比天然Cas9导致更有效的基因破坏(图37(A)),其在更高的浓度以及每种蛋白的相应最佳蛋白质:sgRNA剂量处不如天然Cas9(图37(B)-37(C))。这些结果进一步建立了sgRNA能供应充足的负电荷来支持复合Cas9蛋白的基于阳离子脂质的投递。

[0759] 还观察到尽管 (-30) dGFP-NLS-Cas9需要比Cas9总体上更少的蛋白用于最佳投递,但对于最大 (-30) dGFP-NLS-Cas9介导的EGFP基因破坏需要比天然Cas9介导的基因破坏更高的sgRNA:蛋白比率(图37(D))。推测需要更多的sgRNA等同物以与 (-30) dGFP-NLS-Cas9复合,因为融合的 (-30) dGFP可能静电干扰Cas9:sgRNA复合。由于对于 (-30) dGFP-NLS-Cas9介导的EGFP基因破坏的理想蛋白剂量比野生型Cas9的低10倍,该结果还表明 (-30) dGFP-Cas9由于其更高的总体负电荷比Cas9:sgRNA更好地被阳离子脂质体包封,但该电荷大小可能干扰Cas9:sgRNA相互作用,从而需要更多的sgRNA每蛋白和潜在地降低投递的Cas9总活性。另外,生成和测试NLS-Cas9和Cas9-NLS蛋白,且观察到尽管NLS在 (-30) dGFP-NLS-Cas9中的存在能至少部分地揭示在非常低的浓度时投递功效中的差异,但Cas9,NLS-Cas9和Cas9-NLS在25nM或更高的浓度均导致比 (-30) dGFP-NLS-Cas9更高的EGFP破坏效率(图38(A)-(C))。

[0760] 还测试了使用阳离子脂质制剂而非RNAiMAX的Cas9:sgRNA投递。使用Lipofectamine 2000的投递比使用RNAiMAX显著地更有效,导致多达80%的Cas9介导的基因破坏(图39(A)),且甚至在1nM蛋白处也维持高效率(60%基因破坏)(图39(A))。然而,由于在细胞培养条件下Lipofectamine 2000比RNAiMAX的稍微更高的毒性(图33(B)-(C)),将RNAiMAX用于所有后续的细胞培养研究。

[0761] 为了验证EGFP破坏来自基因组修饰而不仅是Cas9结合⁴²,使用T7内切核酸酶I(T7EI)测定法⁴³来检测和量化在靶EGFP基因座处Cas9介导的基因组插入/缺失突变的频率(图30(B))。T7EI测定结果显示仅用Cas9和EGFP sgRNA质粒两者,或Cas9蛋白和纯化的EGFP sgRNA处理的那些细胞在靶位点处含有indel。总之,这些发现建立了活性Cas9:sgRNA复合物可使用阳离子脂质,以依赖于由sgRNA提供的负电荷的方式,有力地投递到人细胞中。

[0762] 还将U2OS EGFP报道细胞用Cas9的单个脂质介导的投递来处理,所述Cas9与靶向EGFP,CLTA,EMX和VEGF的四种gRNA的混合物复合。该处理导致所有4种靶物的有效破坏,切割效率分别为58%,28%,16%和40%,如通过T7E1切割测定法测量的。来自50nM Cas9和25nM每种sgRNA(总计100nM sgRNA)的单投递的这些高基因破坏效率证明脂质介导的Cas9:sgRNA投递能支持有效的多重基因组编辑(图30(C))。

[0763] Cas9切口酶和dCas9激活物的功能性投递

[0764] 接着,测试了基于阳离子脂质的蛋白投递是否能扩展到投递其他Cas9衍生的基因组工程化工具如Cas9切口酶⁴⁴和基于Cas9的转录激活物⁴⁵。测量了U2OS EGFP报道细胞中来自Cas9 D10A切口酶的投递的基因破坏效率,其通过共转染切口酶和适宜的成对的EGFP靶向性sgRNA质粒,或使用RNAiMAX与EGFP sgRNA对复合的纯化的蛋白(图30(D))。仅在存在靶向相对链的sgRNA对(sgRNA对g1+g5、和g3+g7)而非靶向相同链的sgRNA对(sgRNA对g5+g7)的情况下(图30(D)),双重Cas9切口酶的质粒和阳离子脂质介导的蛋白:RNA投递两者导致

具有类似效率的EGFP破坏(图30(D)),这与之前的Cas9切口酶切割要求的报道一致⁴⁶。

[0765] 比较了HEK293T细胞中的NTF3转录激活效率,其来自质粒转染或直接的融合于VP64激活域的dCas9的蛋白质:sgRNA复合物投递⁴⁵。通过质粒转染或RNAiMAX介导的蛋白投递来投递dCas9-VP64激活物导致较强的($\geq \sim 10$ 倍)NTF3转录的活化(图30(E)和图40)。对于每种投递方法在最佳测定次数,来自质粒转染的转录激活水平比来自蛋白投递的活化更有力(图30(E)),可能是由于来自质粒的Cas9激活物蛋白和sgRNA两者的持续表达相比于纯化的蛋白和RNA的短暂的单剂量。虽然上述结果指示这类因素不限制通过投递的Cas9核酸酶和切口酶蛋白对不可逆基因组修饰的效力(图40(A)和40(D)),但投递蛋白的低剂量和短暂性质可能更强地限制动态过程如转录激活的效力。但是,这些结果统一指示Cas9切口酶和Cas9转录激活物两者还可通过阳离子脂质介导的蛋白:RNA复合物投递来有效投递。

[0766] Cas9:sgRNA投递以比DNA转染更大的特异性修饰基因组

[0767] 功能性Cas9:sgRNA复合物的无DNA投递绕开了与病毒或其他基因投递方法有关的风险,且通过避免修饰靶基因座后基因组编辑剂的不必要的表达从而具有改进基因组修饰的特异性的潜力。为了测试这一所述办法是否能破坏人细胞中的内源性基因,靶向EMX1,CLTA2和VEGF基因中的基因组基因座,原因是所述基因的潜在的生物学相关性和其在之前Cas9脱靶切割活性研究中的使用^{40,46,47}。Cas9:sgRNA复合物到HEK293T细胞中的阳离子脂质介导的投递导致所有3种人基因的有力的切割,其效率与质粒转染方法的那些可比或更大,如通过使用先前对U2OS细胞优化的相同Cas9:sgRNA投递条件的T7EI测定法所揭示的(图31(A))。

[0768] 为了比较对于Cas9质粒相对于蛋白:RNA投递方法的内源性基因修饰特异性,从自HEK293细胞分离的基因组DNA扩增在靶基因座以及几个已知的脱靶位点,所述细胞在产生可比的在靶修饰效率的条件下通过转染Cas9和sgRNA表达质粒,或通过RNAiMAX介导的Cas9:sgRNA复合物投递处理。通过高通量DNA测序分析在3个在靶和11个脱靶位点处的indel频率。对于所有3种靶基因,来自质粒或蛋白质:sgRNA投递的在靶DNA修饰的频率约为10%(图41(A)-(C)),其实现导致非常相似的在靶基因组修饰效率的处理条件下两种技术之间的脱靶修饰的比较。重要地,对于所有11个脱靶位点的脱靶基因组修饰的频率在来自蛋白质:sgRNA投递时比质粒投递低,且因此对于所有测试位点的在靶对脱靶修饰的比率对于蛋白质:sgRNA投递比对于质粒投递高达19倍(图31(B)-(D))。

[0769] 在具有高水平的脱靶修饰的基因座(如4个VEGF脱靶位点,质粒投递对其得到了4至20倍的平均在靶:脱靶修饰比率,而蛋白:RNA投递得到了9至400倍的平均在靶:脱靶修饰比率)以及具有较低水平的脱靶修饰的基因座(如3个EMX脱靶基因座,质粒投递对其得到了低至64倍的平均在靶:脱靶修饰比率,而蛋白:RNA投递得到了500至2,000倍的平均在靶:脱靶修饰比率)处,蛋白质:sgRNA投递的DNA修饰特异性高于质粒投递的。总之,这些结果指示使用阳离子脂质投递Cas9:sgRNA复合物能实现高效率的靶基因修饰,其具有比投递表达Cas9和sgRNA的DNA实质性更高的特异性。

[0770] 将Cas9:sgRNA投递到小鼠胚胎干细胞中

[0771] 有力和短暂的阳离子脂质介导的Cas9:sgRNA投递以实现有效、永久和高度特异性的基因编辑可能在干细胞中特别有用。为了测试该可能性,将表达Tau-EGFP⁴⁸的小鼠胚胎干细胞用Cas9和EGFP靶向性sgRNA处理。在标准的干细胞培养条件下,形成EGFP阳性浮球。将

浮球用与RNAiMAX复合的Cas9:sgRNA处理,或用Cas9和RNAiMAX无sgRNA处理作为对照。处理后3天,观察到与对照样品相比在用Cas9:sgRNA处理的球中的GFP荧光降低(图42(A))。将经处理的球解离,允许细胞附着于用明胶包被的血并分化成祖细胞。使用抗GFP抗体的免疫组化确认了EGFP表达在用Cas9:sgRNA处理的样品的细胞中的敲低,其中许多细胞核具有缺少任何表现的EGFP。相比之下,源自对照球的所有细胞均是EGFP阳性的(图42(B))。对从用Cas9:sgRNA处理的细胞收获的基因组DNA进行T7EI测定,其产生在Tau-EGFP基因座处的indel的清楚证据(图42(C))。从该测定,从阳离子脂质介导的Cas9:sgRNA投递和Cas9和sgRNA DNA的转染两者计算出42%的indel频率。在缺少Cas9:sgRNA或含有Cas9和不相关gRNA的对照样品中未检测到靶物修饰。这些发现证明阳离子脂质介导的Cas9:sgRNA投递能在小鼠胚胎干细胞中实现高度有效的基因破坏。

[0772] Cre重组酶和Cas9:sgRNA的体内阳离子脂质介导的投递

[0773] 功能性基因组编辑蛋白的体内高效投递实现广泛应用,包括校正遗传疾病的非病毒治疗性基因组编辑。为了在活哺乳动物中评估上文描述的蛋白投递方法,选择对小鼠内耳的投递,因为其具有有限的空间、研究较多的内耳细胞类型、以及可以实现将来听力恢复的遗传聋小鼠模型的存在。尝试将两种类型的蛋白体内投递到小鼠内耳中。首先,测试(-30)GFP-Cre蛋白的投递来评估对内耳细胞类型的靶向和功能蛋白投递的效率。其次,评估Cas9:sgRNA复合物对内耳的投递来确定阳离子脂质介导的蛋白:sgRNA复合物投递是否能支持基于CRISPR的体内基因编辑。

[0774] 之前已显示(+36)GFP-Cre可投递到小鼠视网膜¹⁶,虽然该蛋白仅导致重组转化的适度水平,表明低效的体内投递。对于我们的初始内耳投递试验,将(-30)GFP-Cre与RNAiMAX复合,并将复合物注射到出生后第0天(P0)的报道小鼠的耳蜗中,该小鼠具有基因组整合的敲入的(floxed)-STOP tdTomato报道物。如先前在体外Cre报道细胞系中描述的,将Cre功能性投递到内耳细胞中,继之以核内体逃离、核定位和Cre介导的重组导致tdTomato的表达。在注射后,收获耳蜗用于使用内耳细胞标志物进行免疫标记,用于与tdTomato的共定位。将单独的RNAiMAX注射用作对照。在注射(-30)GFP-Cre和RNAiMAX后5天,耳蜗外毛细胞(检测声音的耳部感觉细胞)显示与毛细胞标志物肌球蛋白VIIa(Myo7a)共定位的强烈的tdTomato信号,从而证明了到毛细胞的功能性Cre投递(图32(A)-(B))。在对照耳蜗中未检测到tdTomato表达(图32(A))。tdTomato信号集中在耳蜗底部转角处的注射部位区域中。在耳蜗底处平均 $33 \pm 3\%$ 的外毛细胞是tdTomato阳性的($P < 0.001$;均值 \pm SEM, $n=4$)。

[0775] 为了进一步确定阳离子脂质介导的(-30)GFP-Cre蛋白投递对靶细胞的影响,在注射后10天检查毛细胞静纤毛(stereocilia),一种对于听力必要的精细结构。TdTomato阳性的外毛细胞具有典型的静纤毛结构,如通过espin表达成像的,其类似于对照静纤毛(图32(B))。在对照耳蜗中未检测到tdTomato表达。这些观察表明阳离子脂质介导的(-30)GFP-Cre蛋白投递实现了在耳蜗外毛细胞中的重组,且没有明显地影响毛细胞体系结构。

[0776] 由于靶物体积、蛋白剂量和sgRNA剂量在体内不同于在细胞培养实验中,因此在不同的投递条件下重复上述实验。由于其在体外的高效力,测试了使用Lipofectamine 2000的投递(图39(A))且观察到急剧更高的重组效率:在用(-30)GFP-Cre+Lipofectamine 2000处理的耳蜗中超过90%的外毛细胞是tdTomato阳性的(图32(C))。与对照样品相比,观察到

一些外毛细胞丧失(图32(C)),这与先前观察到的Lipofectamine 2000的高细胞毒性一致,尽管保持总体耳蜗体系结构。

[0777] 为了测试体内Cas9:sgRNA投递的有效性,将Cas9和靶向EGFP的sgRNA与RNAiMAX组合,并将所得复合物注射到出生后第2天(P2)的转基因Atoh1-GFP小鼠耳蜗中,其中所有毛细胞在转录因子Atoh1⁴⁹的毛细胞特异性增强子的调控下表达GFP。使用此模型,Cas9:sgRNA介导的EGFP的破坏导致外毛细胞中EGFP荧光的丧失。注射Cas9:sgRNA和阳离子脂质后10天,在注射部位附近处13%的外毛细胞中观察到GFP的缺失。相比之下,注射有Cas9蛋白和RNAiMAX且无任何sgRNA的对照耳蜗显示出没有EGFP信号的丧失(图32(D))。注射有Cas9:sgRNA RNAiMAX复合物的耳蜗的外毛细胞似乎另外未受影响的,其具有Myo7a的典型表达和健康的细胞核,与最小的毛细胞毒性一致(图32(D))。对从耳蜗组织样品分离的基因组DNA的高通量DNA测序揭示了与经处理样品而非缺少sgRNA的对照样品中的GFP靶基因破坏一致的indel(图43(A))。另外,重复使用靶向EMX基因的sgRNA的内耳体内投递Cas9:sgRNA,且类似地观察到在经处理动物而非对照动物的EMX基因中的indel(图43(B))。

[0778] 由于与Lipofectamine 2000复合的(-30)GFP-Cre产生比与RNAiMAX复合的(-30)GFP-Cre更有效的靶毛细胞群体修饰(图32(A)和32(C)),如上文所述测试其在Cas9:sgRNA投递到Atoh1-GFP耳蜗上的使用。10天后在注射部位附近的20%的外毛细胞中观察到GFP表达的丧失,而在对照耳蜗中所有外毛细胞维持强GFP表达,所述对照耳蜗注射有Cas9和Lipofectamine 2000但无sgRNA(图32(D))。与在Lipofectamine 2000投递(-30)GFP-Cre后观察到的适度的毛细胞丧失(图32(C))形成对比的是,Cas9:sgRNA靶向的外毛细胞展现出无明显的毒性或结构改变(图32(D))。

[0779] 与(-30)GFP-Cre一样,无病毒的、阳离子脂质介导的Cas9:sgRNA投递至小鼠内耳成功地修饰了外毛细胞群体中的特定基因组基因座,从而导致靶基因表达的丧失。几乎一半的所有遗传耳聋类型源自毛细胞丧失或功能障碍⁵⁰,本文中呈现的结果提出了基于Cas9:sgRNA复合物的投递来遗传修饰这些细胞从而实现听力恢复的潜在策略。

[0780] 确定(-30)GFP-Cre的蛋白投递功效

[0781] 为了确定脂质体介导的(-30)GFP-Cre投递比阳离子蛋白投递更高的效力来自细胞的更多的总蛋白摄取或来自细胞摄取的功能性、非核内体蛋白分子的更高分数,使用流式细胞术来测量用(+36)GFP-Cre或脂质体(-30)GFP-Cre在其相应的最佳Cre投递条件下处理的细胞的GFP荧光。细胞荧光报告总的内吞的(-30)GFP-Cre或(+36)GFP-Cre,而不管是核内体或非核内体定位¹。脂质介导的蛋白投递令人惊讶地导致总蛋白摄取中的少量增加(图34(A)),尽管脂质介导的功能性Cre投递具有高效率。虽然(+36)GFP-Cre处理以剂量依赖性方式将细胞GFP荧光增加了高达3个量级(图34(A)),与先前的报道一致^{1,2},但是脂质体(-30)GFP-Cre处理诱导细胞GFP荧光中至多5倍的增加(图34(A))。对细胞荧光和重组效率的比较揭示,(-30)GFP-Cre的脂质介导的功能投递比(+36)GFP-Cre的投递为每份量的内吞蛋白效力高9,800倍(图34(B))。

[0782] 为了测试阴离子(-30)GFP与阳离子脂质的复合是否干扰GFP荧光,由此遮蔽了进入细胞的货物的真实量,将mCherry(其为荧光但不是高度阴离子的)融合于(-30)GFP或(+36)GFP并将两种蛋白融合物投递至HeLa细胞。在用PBS+肝素(20U/mL)洗掉可能附在细胞表面但不进入细胞的蛋白后,在处理4小时和24小时通过FACS分析细胞的mCherry荧光。观

察到融合(-30)GFP的mCherry的脂质介导的投递导致细胞mCherry荧光的仅稍微增加,而在投递(+36)GFP-mCherry后的mCherry荧光一般高 ≥ 100 倍(图34(C)),表明融合于(-30)GFP不导致大量的蛋白货物进入细胞。而且,将脂质添加至(-30)GFP-Cre没有可测量地改变GFP荧光信号(图34(D)),尽管存在阳离子脂质和阴离子(-30)GFP显然相互作用的事实。总之,这些结果表明阴离子蛋白的脂质介导投递的不寻常的高效力不是来自于每个细胞中的异常高的蛋白摄取,而是来自于内吞作用后过程,其可能包括避免溶酶体蛋白降解和核内体逃离到细胞质中。

[0783] 脱靶切割测定法的灵敏性限制

[0784] 用于检测基因组脱靶切割的高通量测序方法的灵敏性受限于输入每个基因组靶位点的PCR扩增中的基因组DNA(gDNA)量。人gDNA的1ng样品代表仅约330个独特的基因组,且如此仅存在每个基因组位点的约330个独特拷贝。在总共150ng的输入gDNA上实施每个基因组靶物的PCR扩增,其分别提供了源自至多50,000个独特gDNA拷贝的扩增子。因此,高通量测序测定法不能检测到以低于1/50,000(0.002%)频率存在的稀有的基因组修饰事件。该限制在表2中陈述。

[0785] 总之,这些发现表明基因组编辑蛋白的阳离子脂质介导的投递能充当用于治疗遗传疾病的强大工具和体内策略。

[0786] 结论

[0787] 体外且尤其是体内的有效胞内蛋白投递一直是生物医学研究和蛋白治疗学中的长久以来的挑战。虽然已对使用阳离子肽和蛋白的投递广泛研究了超过20年,但对血清蛋白的敏感性,抗体的中和,通过胞外和胞内蛋白酶的降解以及内化后较差的核内体逃离限制了使用该办法进行蛋白投递应用的范围。

[0788] 在本发明的实施例中,显示了蛋白投递的一种一般性策略,其利用阴离子蛋白与阳离子脂质体复合。该方法用于向培养的细胞系、干细胞集落和小鼠内耳中的治疗相关的体内部位投递多样化的蛋白类别,包括Cre酪氨酸重组酶、TALE转录激活物和Cas9核酸酶、切口酶、和转录激活物(图27(A))。所描述的办法是高度有效的,其产生与细胞培养中确立的核酸转染方法相当的修饰率,并在活小鼠的内耳毛细胞群体内分别实现高达90%和20%的Cre重组酶和Cas9介导的基因组修饰率(图32(C)-(D))。这些结果还表明可以使用阳离子脂质来有效投递其他核酸结合蛋白(包括诱导细胞命运中的治疗相关变化的转录因子),这是通过将其与核酸复合来进行。

[0789] 基于阳离子脂质的阴离子蛋白投递比有力的阳离子蛋白投递融合伴侣(+36)GFP的性能每分量的内吞蛋白高多达9,800倍,其使用更低数量级的蛋白剂量诱导了对经处理细胞的更有效的修饰(图28(C)34)。对于Cas9核酸酶投递,该办法还导致比传统质粒转染特异性高 >10 倍的基因组修饰(图31(B)-(D)),可能是由于与DNA投递方法相比,每个基因组所被暴露的Cas9活性的短暂窗口,这与先前的报道一致⁵¹。

[0790] 所描述的办法易于执行,仅需要纯化的可投递的蛋白和使用通常的商品化核酸转染试剂(图27(B))。使得给定的蛋白可用于该办法需要简单翻译融合于高度阴离子性的伴侣如(-30)GFP(图27(A)),且使用常见的翻译融合标签(包括VP64激活域和3xFLAG亲和标签)甚至更有效(图28(F)和图33(C))。在某些例子中(如与Cas9蛋白一样),与关联核酸(此例中为sgRNA)的预先复合是足够的(图30(A)),因为部分暴露的结合的核酸可能提供充足

的负电荷来介导与阳离子脂质的复合。

[0791] 其他小组报道了以DNA或mRNA形式体内投递Cas9表达构建体^{52,53}。本发明实施例证明蛋白投递是进行体内基因组编辑的可行办法。

[0792] 表1.

EMX_On	GAGTCCGAGCAGAAGAAGAAGGG (SEQ ID NO:209)_
EMX_Off1	GAGgCCGAGCAGAAGAAgACGG (SEQ ID NO:210)
[0793] EMX_Off2	GAGTCCtAGCAGgAGAAGAAGaG (SEQ ID NO:211)
EMX_Off3	GAGTcTaAGCAGAAGAAGAAGaG (SEQ ID NO:212)
EMX_Off4	GAGTtaGAGCAGAAGAAGAAAGG (SEQ ID NO:213)

VEGF_On	GGGTGGGGGGAGTTTGCTCCTGG (SEQ ID NO:214)
VEGF_Off1	GGaTGGaGGGAGTTTGCTCCTGG (SEQ ID NO:215)
VEGF_Off2	GGGaGGGtGGAGTTTGCTCCTGG (SEQ ID NO:216)
VEGF_Off3	cGGgGGaGGGAGTTTGCTCCTGG (SEQ ID NO:217)
[0794] VEGF_Off4	GGGgaGGGgaAGTTTGCTCCTGG (SEQ ID NO:218)
CLTA_On	GCAGATGTAGTGTTTCCACAGGG (SEQ ID NO:219)
CLTA_Off1	aCAaATGTAGTaTTTCCACAGGG (SEQ ID NO:220)
CLTA_Off2	cCAGATGTAGTaTTcCCACAGGG (SEQ ID NO:221)
CLTA_Off3	ctAGATGaAGTGcTTCCACATGG (SEQ ID NO:222)

[0795] 表1.Cas9:sgRNA(靶向MX,VEGF和CLTA中的位点)的在靶和已知的脱靶底物。显示EMX,VEGF和CLTA的基因组在靶和脱靶位点,来自在靶序列的突变以小写和粗体显示。PAM划线显示。

[0796] 表2

	<u>CLTA 位点</u>	伪处理	质粒转染	蛋白质:sgRNA 投递
	CLTA_On			
	Indel	14	1228	1498
	总计	10000	10000	10000
	经修饰(%)	0.140	12.280	14.980
	P 值		<1.0E-300	<1.0E-300
[0797]	在靶:脱靶特异性	1	1	1
	CLTA_Off1			
	Indel	7	29	14
	总计	41518	205204	125370
	经修饰(%)	0.017	0.014	0.011
	P 值		6.6E-01	4.5E-01

	在靶:脱靶特异性	869	1341
	CLTA_Off2		
	Indel	5	11
	总计	25338	83944
	经修饰(%)	0.020	0.013
	P 值	5.5E-01	5.7E-01
	在靶:脱靶特异性	937	1019
	CLTA_Off3		
	Indel	6	22
	总计	41643	189886
	经修饰(%)	0.014	0.012
	P 值	6.2E-01	5.8E-01
[0798]	在靶:脱靶特异性	1060	1439
	<u>EMX 位点</u>	<u>伪处理</u>	<u>质粒转染</u>
	<u>EMX_On</u>		<u>蛋白质:sgRNA 投递</u>
	Indel	3	930
	总计	10000	10000
	经修饰(%)		0.030
	P 值		1.6E-264
	在靶:脱靶特异性	1	1
	EMX_Off1		
	Indel	0	6
	总计	24623	90935
	经修饰(%)		<0.002
	P 值		3.5E-01
	在靶:脱靶特异性		1409
			1915

	EMX_Off2			
	Indel	16	53	38
	总计	36061	204068	130084
	经修饰(%)		0.044	0.026
	P 值		6.4E-02	1.8E-01
	在靶:脱靶特异性		358	390
	EMX_Off3			
	Indel	20	147	44
	总计	32575	157848	110878
	经修饰(%)		0.061	0.093
	P 值		8.1E-02	1.3E-01
	在靶:脱靶特异性		100	287
	EMX_Off4			
[0799]	Indel	16	141	23
	总计	45548	86586	73451
	经修饰(%)		0.035	0.163
	P 值		2.8E-12	7.4E-01
	在靶:脱靶特异性		57	364
	<u>VEGF 位点</u>	伪处理	质粒转染	蛋白质:sgRNA 投递
	VEGF_On			
	Indel	1	989	785
	总计		10000	10000
	经修饰(%)	0.010	9.890	7.850
	P 值		1.5E-285	5.7E-228
	在靶:脱靶特异性	1	1	1
	VEGF_Off1			
	Indel	4	4240	602

	总计		38625	184554
	经修饰(%)	0.010	2.297	0.394
	P 值		<1.0E-300	3.7E-52
	在靶:脱靶特异性		4	20
	VEGF_Off2			
	Indel	5	727	18
	总计		30301	79164
	经修饰(%)	0.017	0.918	<0.002
	P 值		4.7E-93	1.3E-04
	在靶:脱靶特异性		11	3925
[0800]	VEGF_Off3			
	Indel	2	536	21
	总计		26379	110902
	经修饰(%)	0.008	0.483	0.022
	P 值		2.0E-46	2.0E-01
	在靶:脱靶特异性		20	352
	VEGF_Off4			
	Indel	0	1531	45
	总计		26012	122403

[0801] 表2. EMX、CLTA和VEGF在靶位点和11个已知脱靶位点的Indel频率、P值和在靶:脱靶切割特异性比率。CLTA位点:总计:序列计数的总数目;仅对头10,000个序列分析在靶位点序列。经修饰:Indel数目除以序列总数作为百分比。对没有观察到的Indel的位点计算可能修饰的上限,其通过假定有少于1个indel然后除以总序列计数以得到上限修饰百分比,或取理论检测限(1/49,500;参见上文结果部分),看哪个数值更大。P值:对于伪处理、Cas9质粒转染和脂质体Cas9蛋白:sgRNA投递,使用每个CLTA靶向性处理样品(DNA转染或蛋白质:sgRNA投递)相对于用Cas9蛋白和靶向EGFP的sgRNA处理的对照样品(伪处理)之间的双侧Fisher确切检验来计算P值。在靶:脱靶特异性是对于每个位点,在靶对脱靶基因组修饰频率的比率。EMX位点显示应用于EMX靶位点的CLTA分析的实验和分析方法。VEGF位点显示如应用于VEGF靶位点的CLTA分析的实验和分析方法。从质粒转染和蛋白质:sgRNA投递两者的indel数目减去伪处理对照中的Indel数目来确定Indel的总数目和计算图31正文以及图

41中的在靶:脱靶比率。

[0802] 参考文献

[0803] 1.Putney,S.D.&Burke,P.A.Improving protein therapeutics with sustained-release formulations.Nat.Biotechnol.16,153-157(1998) .

[0804] 2.Mullen,L.等Latent cytokines for targeted therapy of inflammatory disorders.Expert Opin.Drug Deliv.11,101-110(2014) .

[0805] 3.Song,E.等Antibody mediated in vivo delivery of small interfering RNAs via cell-surface receptors.Nat.Biotechnol.23,709-717(2005) .

[0806] 4.Leader,B.,Baca,Q.J.&Golan,D.E.Protein therapeutics:a summary and pharmacological classification.Nat.Rev.Drug Discov.7,21-39(2008) .

[0807] 5.Hartung,S.D.等Correction of Metabolic,Craniofacial,and Neurologic Abnormalities in MPSIMice Treated at Birth with Adeno-associated Virus Vector Transducing the Human α -L-Iduronidase Gene.Mol.Ther.9,866-875(2004) .

[0808] 6.Wang,J.等Neutralizing antibodies to therapeutic enzymes: considerations for testing,prevention and treatment.Nat.Biotechnol.26,901-908 (2008) .

[0809] 7.Urnov,F.D.,Rebar,E.J.,Holmes,M.C.,Zhang,H.S.&Gregory,P.D.Genome editing with engineered zinc finger nucleases.Nat.Rev.Genet.11,636-646(2010) .

[0810] 8.Sander,J.D.&Joung,J.K.CRISPR-Cas systems for editing,regulating and targeting genomes.Nat.Biotechnol.32,347-355(2014) .

[0811] 9.Gaj,T.,Gersbach,C.A.&Barbas,C.F.ZFN,TALEN,and CRISPR/Cas-based methods for genome engineering.Trends Biotechnol.31,397-405(2013) .

[0812] 10.Midoux,P.,Pichon,C.,Yaouanc,J.-J.&Jaffrès,P.-A.Chemical vectors for gene delivery:a current review on polymers,peptides and lipids containing histidine or imidazole as nucleic acids carriers.Br.J.Pharmacol.157,166-178 (2009) .

[0813] 11.Bodles-Brakhop,A.M.,Heller,R.&Draghia-Akli,R.Electroporation for the Delivery of DNA-based Vaccines and Immunotherapeutics:Current Clinical Developments.Mol.Ther.17,585-592(2009) .

[0814] 12.Kay,M.A.,Glorioso,J.C.&Naldini,L.Viral vectors for gene therapy: the art of turning infectious agents into vehicles of therapeutics.Nat.Med.7, 33-40(2001) .

[0815] 13.Zangi,L.等Modified mRNA directs the fate of heart progenitor cells and induces vascular regeneration after myocardial infarction.Nat.Biotechnol.31,898-907(2013) .

[0816] 14.Wadia,J.S.,Stan,R.V.&Dowdy,S.F.Transducible TAT-HA fusogenic peptide enhances escape of TAT-fusion proteins after lipid raft macropinocytosis.Nat.Med.10,310-315(2004) .

[0817] 15.Daniels,D.S.&Schepartz,A.Intrinsically cell-permeable miniature

proteins based on a minimal cationic PPII motif. *J. Am. Chem. Soc.* 129, 14578-14579 (2007).

[0818] 16. Cronican, J. J. 等 Potent delivery of functional proteins into mammalian cells in vitro and in vivo using a supercharged protein. *ACS Chem. Biol.* 5, 747-752 (2010).

[0819] 17. Thompson, D. B., Cronican, J. J. & Liu, D. R. Engineering and identifying supercharged proteins for macromolecule delivery into mammalian cells. *Methods Enzymol.* 503, 293-319 (2012).

[0820] 18. Thompson, D. B., Villaseñor, R., Dorr, B. M., Zerial, M. & Liu, D. R. Cellular uptake mechanisms and endosomal trafficking of supercharged proteins. *Chem. Biol.* 19, 831-843 (2012).

[0821] 19. Heitz, F., Morris, M. C. & Divita, G. Twenty years of cell-penetrating peptides: from molecular mechanisms to therapeutics. *Br. J. Pharmacol.* 157, 195-206 (2009).

[0822] 20. Caron, N. J. 等 Intracellular delivery of a Tat-eGFP fusion protein into muscle cells. *Mol. Ther. J. Am. Soc. Gene Ther.* 3, 310-318 (2001).

[0823] 21. Chesnoy, S. & Huang, L. Structure and function of lipid-DNA complexes for gene delivery. *Annu. Rev. Biophys. Biomol. Struct.* 29, 27-47 (2000).

[0824] 22. Al-Taei, S. 等 Intracellular traffic and fate of protein transduction domains HIV-1TAT peptide and octaarginine. Implications for their utilization as drug delivery vectors. *Bioconjug. Chem.* 17, 90-100 (2006).

[0825] 23. Shete, H. K., Prabhu, R. H. & Patravale, V. B. Endosomal escape: a bottleneck in intracellular delivery. *J. Nanosci. Nanotechnol.* 14, 460-474 (2014).

[0826] 24. Aguilera, T. A., Olson, E. S., Timmers, M. M., Jiang, T. & Tsien, R. Y. Systemic in vivo distribution of activatable cell penetrating peptides is superior to that of cell penetrating peptides. *Integr. Biol. Quant. Biosci. Nano Macro* 1, 371-381 (2009).

[0827] 25. Coelho, T. 等 Safety and efficacy of RNAi therapy for transthyretin amyloidosis. *N. Engl. J. Med.* 369, 819-829 (2013).

[0828] 26. Judge, A. D., Bola, G., Lee, A. C. H. & MacLachlan, I. Design of noninflammatory synthetic siRNA mediating potent gene silencing in vivo. *Mol. Ther. J. Am. Soc. Gene Ther.* 13, 494-505 (2006).

[0829] 27. Basha, G. 等 Influence of cationic lipid composition on gene silencing properties of lipid nanoparticle formulations of siRNA in antigen-presenting cells. *Mol. Ther. J. Am. Soc. Gene Ther.* 19, 2186-2200 (2011).

[0830] 28. Semple, S. C. 等 Rational design of cationic lipids for siRNA delivery. *Nat. Biotechnol.* 28, 172-176 (2010).

[0831] 29. Boeckle, S., Fahrmeir, J., Roedel, W., Ogris, M. & Wagner, E. Melittin analogs with high lytic activity at endosomal pH enhance transfection with purified

targeted PEI polyplexes. *J. Control. Release Off. J. Control. Release Soc.* 112, 240-248 (2006) .

[0832] 30. Allen, T.M. & Cullis, P.R. Liposomal drug delivery systems: from concept to clinical applications. *Adv. Drug Deliv. Rev.* 65, 36-48 (2013) .

[0833] 31. Zelphati, O. 等 Intracellular delivery of proteins with a new lipid-mediated delivery system. *J. Biol. Chem.* 276, 35103-35110 (2001) .

[0834] 32. Adrian, J.E. 等 Targeted SAINT-0-Somes for improved intracellular delivery of siRNA and cytotoxic drugs into endothelial cells. *J. Control. Release Off. J. Control. Release Soc.* 144, 341-349 (2010) .

[0835] 33. Morris, M.C., Depollier, J., Mery, J., Heitz, F. & Divita, G. A peptide carrier for the delivery of biologically active proteins into mammalian cells. *Nat. Biotechnol.* 19, 1173-1176 (2001) .

[0836] 34. Colletier, J.-P., Chaize, B., Winterhalter, M. & Fournier, D. Protein encapsulation in liposomes: efficiency depends on interactions between protein and phospholipid bilayer. *BMC Biotechnol.* 2, 9 (2002) .

[0837] 35. Lawrence, M.S., Phillips, K.J. & Liu, D.R. Supercharging proteins can impart unusual resilience. *J. Am. Chem. Soc.* 129, 10110-10112 (2007) .

[0838] 36. Liu, J., Gaj, T., Patterson, J.T., Sirk, S.J. & Barbas III, C.F. Cell-Penetrating Peptide-Mediated Delivery of TALEN Proteins via Bioconjugation for Genome Engineering. *PLoS ONE* 9, e85755 (2014) .

[0839] 37. Tessarollo, L., Vogel, K.S., Palko, M.E., Reid, S.W. & Parada, L.F. Targeted mutation in the neurotrophin-3 gene results in loss of muscle sensory neurons. *Proc. Natl. Acad. Sci. U.S.A.* 91, 11844-11848 (1994) .

[0840] 38. Maeder, M.L. 等 Robust, synergistic regulation of human gene expression using TALE activators. *Nat. Methods* 10, 243-245 (2013) .

[0841] 39. Jopling, C., Boue, S. & Belmonte, J.C.I. Dedifferentiation, transdifferentiation and reprogramming: three routes to regeneration. *Nat. Rev. Mol. Cell Biol.* 12, 79-89 (2011) .

[0842] 40. Fu, Y., Sander, J.D., Reyon, D., Cascio, V.M. & Joung, J.K. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat. Biotechnol.* 32, 279-284 (2014) .

[0843] 41. McNaughton, B.R., Cronican, J.J., Thompson, D.B. & Liu, D.R. Mammalian cell penetration, siRNA transfection, and DNA transfection by supercharged proteins. *Proc. Natl. Acad. Sci. U.S.A.* 106, 6111-6116 (2009) .

[0844] 42. Qi, L.S. 等 Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. *Cell* 152, 1173-1183 (2013) .

[0845] 43. Guschin, D.Y. 等 A rapid and general assay for monitoring endogenous gene modification. *Methods Mol. Biol. Clifton NJ* 649, 247-256 (2010) .

[0846] 44. Ran, F.A. 等 Double nicking by RNA-guided CRISPR Cas9 for enhanced

genome editing specificity. *Cell* 154,1380-1389(2013).

[0847] 45.Maeder,M.L.等CRISPR RNA-guided activation of endogenous human genes. *Nat.Methods* 10,977-979(2013).

[0848] 46.Guilinger,J.P.,Thompson,D.B.&Liu,D.R.Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. *Nat.Biotechnol.*32,577-582(2014).

[0849] 47.Pattanayak,V.等High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat.Biotechnol.*31, 839-843(2013).

[0850] 48.Li,H.等Differentiation of neurons from neural precursors generated in floating spheres from embryonic stem cells. *BMC Neurosci.*10,122(2009).

[0851] 49.Lumpkin,E.A.等Math1-driven GFP expression in the developing nervous system of transgenic mice. *Gene Expr.Patterns GEP* 3,389-395(2003).

[0852] 50.Van Camp,G.&Smith,R.Hereditary Hearing Loss.at<<http://hereditaryhearingloss.org>>

[0853] 51.Kim,S.,Kim,D.,Cho,S.W.,Kim,J.&Kim,J.-S.Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9ribonucleoproteins. *Genome Res.*24,1012-1019(2014).

[0854] 52.Yin,H.等Genome editing with Cas9 in adult mice corrects a disease mutation and phenotype. *Nat.Biotechnol.*32,551-553(2014).

[0855] 53.Wang,H.等One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell* 153,910-918(2013).

[0856] 54.Schneider,C.A.,Rasband,W.S.&Eliceiri,K.W.NIH Image to ImageJ: 25years of image analysis. *Nat.Methods* 9,671-675(2012).

[0857] 55.Sage,C.等Proliferation of functional hair cells in vivo in the absence of the retinoblastoma protein. *Science* 307,1114-1118(2005).

[0858] 56.Sander,J.D.等In silico abstraction of zinc finger nuclease cleavageprofiles reveals an expanded landscape of off-target sites. *Nucleic Acids Res.*41,e181(2013).

[0859] 等同方案和范围

[0860] 本领域技术人员将认可或能确定,仅使用常规实验法的本文描述的具体实施方案的许多等同体。本发明的范围不意图受限于上述描述,而是如所附权利要求描述的。

[0861] 本领域技术人员将认可或能确定,仅使用常规实验法的根据本文描述的本发明的具体实施方案的许多等同体。本发明的范围不意图受限于上述描述,而是如所附权利要求中列出的。

[0862] 在权利要求中,除非指定为相反含义或从上下文是明显的,冠词如“一个/一种”和“该”可以意指一种或超过一种。除非指定为相反含义或从上下文是明显的,如果一种、超过一种或所有的组成员存在于、用于给定产品或方法与与给定产品或方法相关,则该组的一

个或多个成员之间包括“或”的权利要求或描述被视为满足的。本发明包括以下实施方案，其中组的正好一个成员存在于、用于给定产品或方法与给定产品或方法相关。本发明还包括以下实施方案，其中组的超过一个或所有成员存在于、用于给定产品或方法与给定产品或方法相关。此外，应理解本发明涵盖所有变体、组合和排列，其中来自一个或多个所列权利要求的一个或多个限制、要素、子句、描述术语等被引入另一个权利要求。例如，可将从属于另一权利要求的任何权利要求修改为包括在从属于同一基础权利要求的任何其他权利要求中所见的一个或多个限制。此外，除非另外指示或对于本领域技术人员显而易见将产生冲突或不一致，当权利要求记载组合物时，应理解包括使用该组合物用于本文公开的任意目的的方法，和包括依照本文公开的制备方法或本领域已知的其他方法的制备组合物的方法。

[0863] 当要素呈现为表时，例如马库什组形式时，应理解还公开了要素的每个亚组，且可从该组除去任何一种或多种要素。应理解，一般来说，当本发明或本发明的方面称为包含具体的要素、特征等时，本发明或本发明方面的某些实施方案组成或基本组成为这类要素、特征等。出于简洁目的，那些实施方案并未逐字地具体在本文中列出。还注意术语“包含”意图为开放式的且允许纳入别的要素或步骤。

[0864] 当给出范围时，包括端点。此外，应理解除非另外指示或从上下文或和本领域普通技术人员的理解来看是明显的，以范围表示的值可涵盖在本发明的不同实施方案中在指定范围内的任何具体值或子范围，到该范围的下限的1/10单位，除非上下文清楚地另外指示。

[0865] 另外，应理解可从任意一项或多项权利要求中明确排除落入现有技术中的本发明的任何具体实施方案。由于这类实施方案被认为是本领域普通技术人员已知的，即使未在本文中明确陈述该排除也可将其排除。出于任何原因，不管是否与现有技术存在有关，均可从任意一项或多项权利要求中排除本发明的组合物的任何具体实施方案（例如任何超电荷蛋白；任何核酸；任何生产方法；任何使用方法等）。

[0866] 所有引用来源，例如本文中引用的参考文献、公开出版物、数据库、数据库录入和现有技术均通过提述并入本申请，即使未在引用中明确陈述。在引用来源和本申请有冲突的情况下，以本申请中的说明为准。

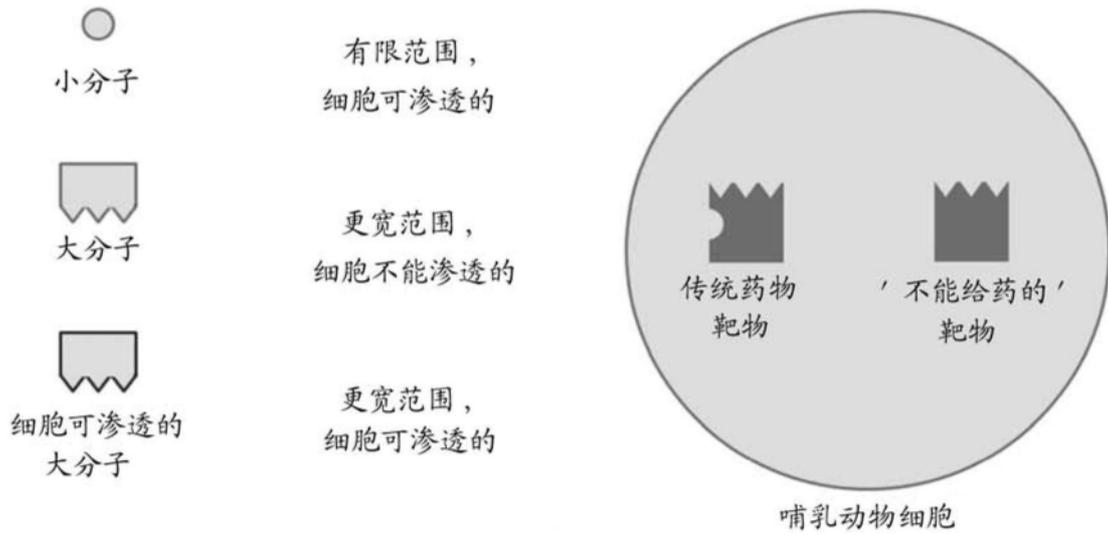


图1

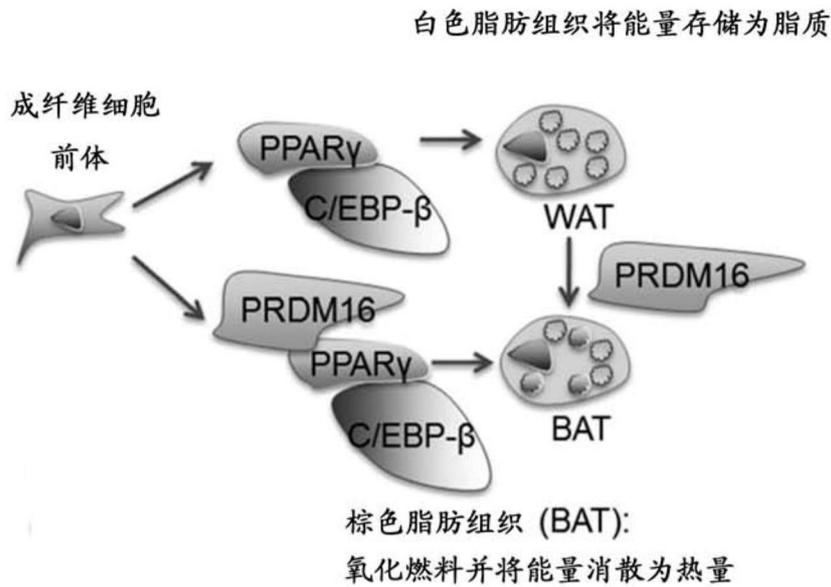


图2

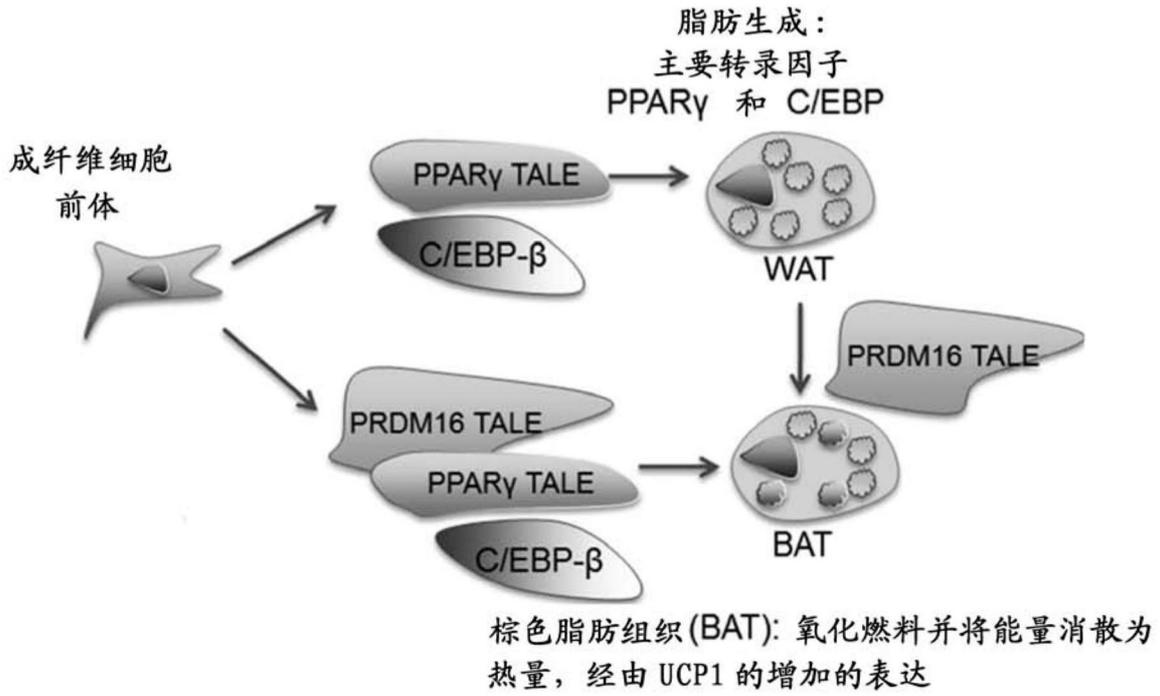
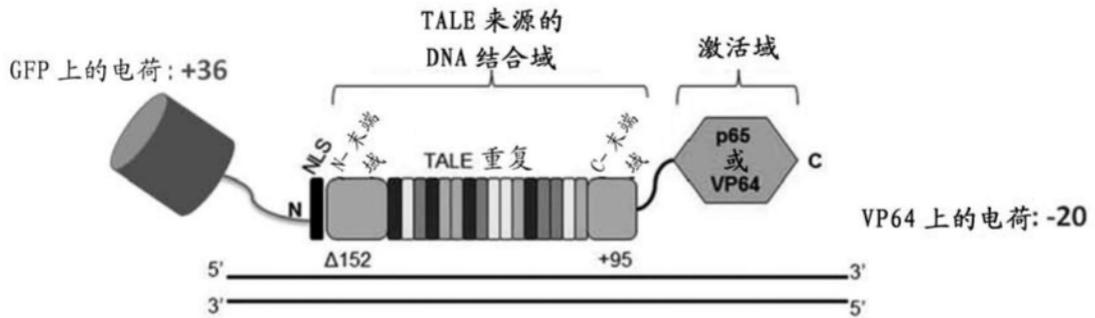


图3



总体净电荷 (包含 DNA 结合域): +22

图4

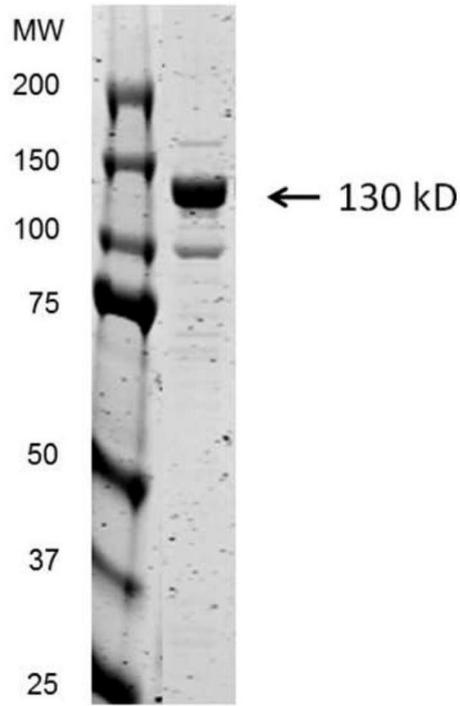


图5

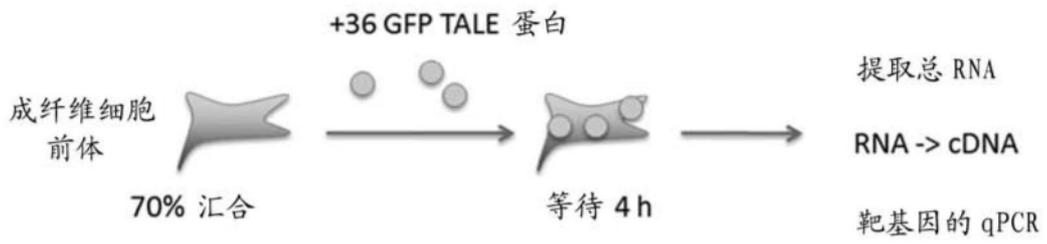


图6

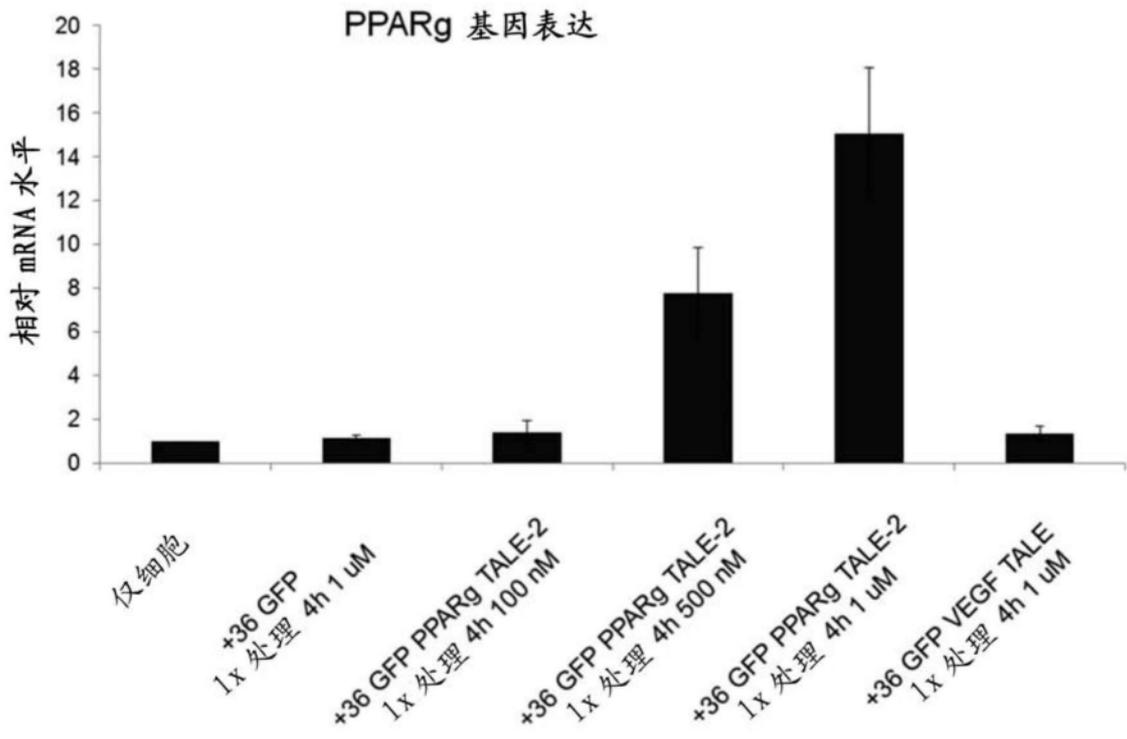


图7

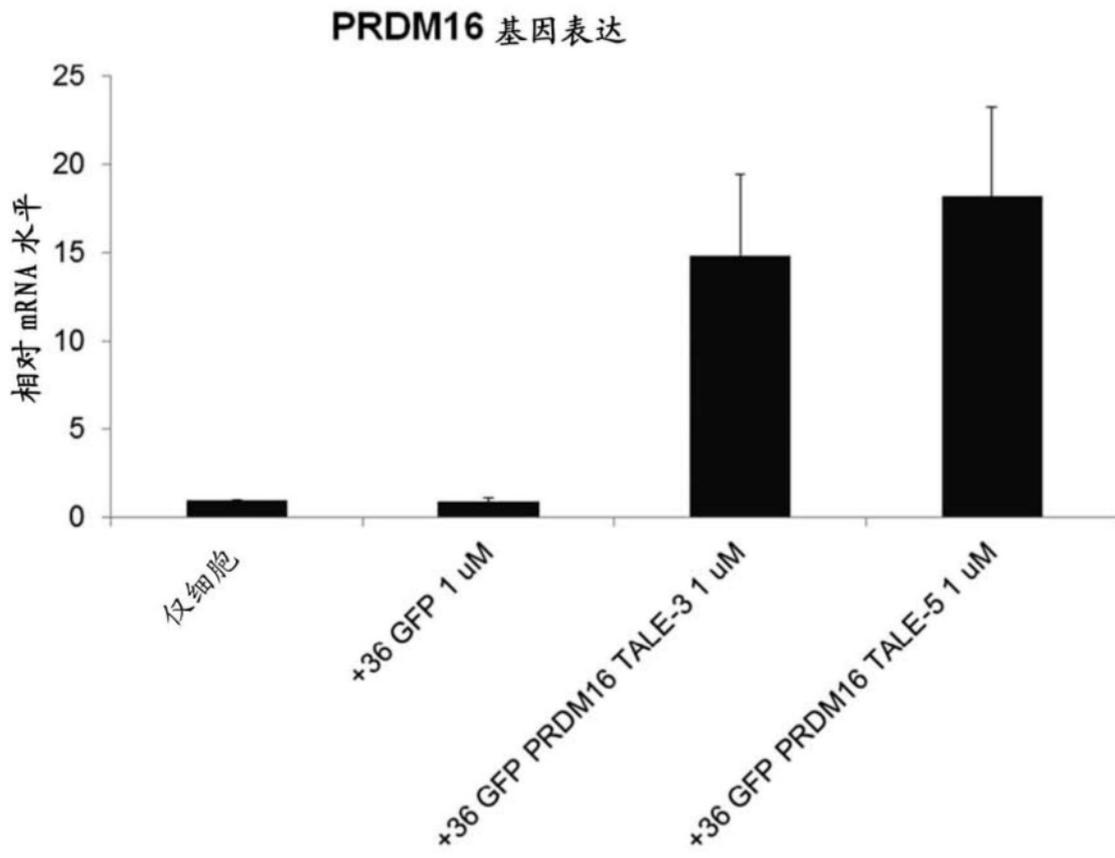


图8

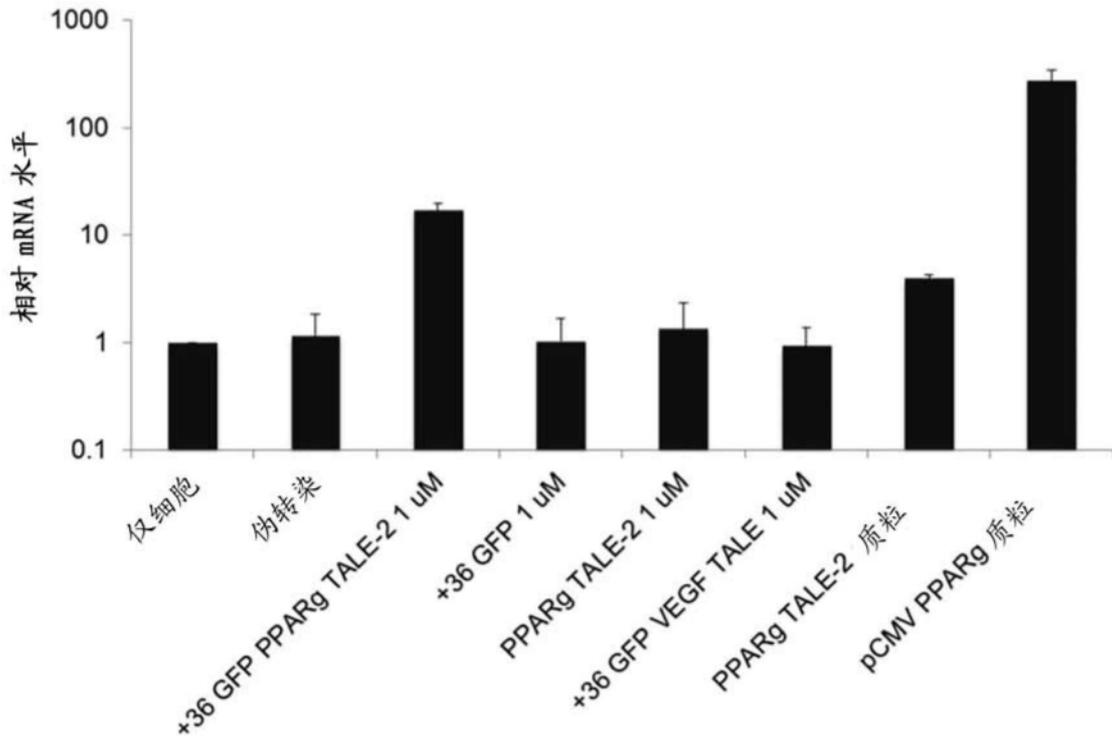


图9

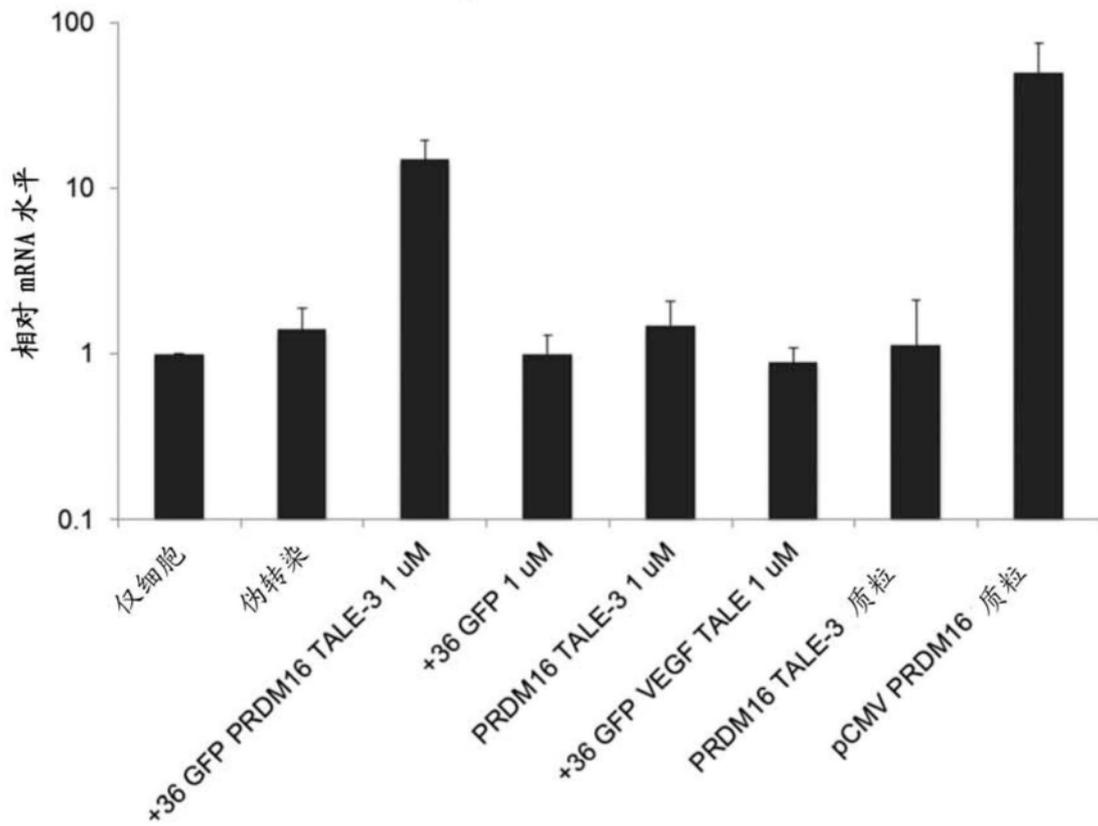


图10

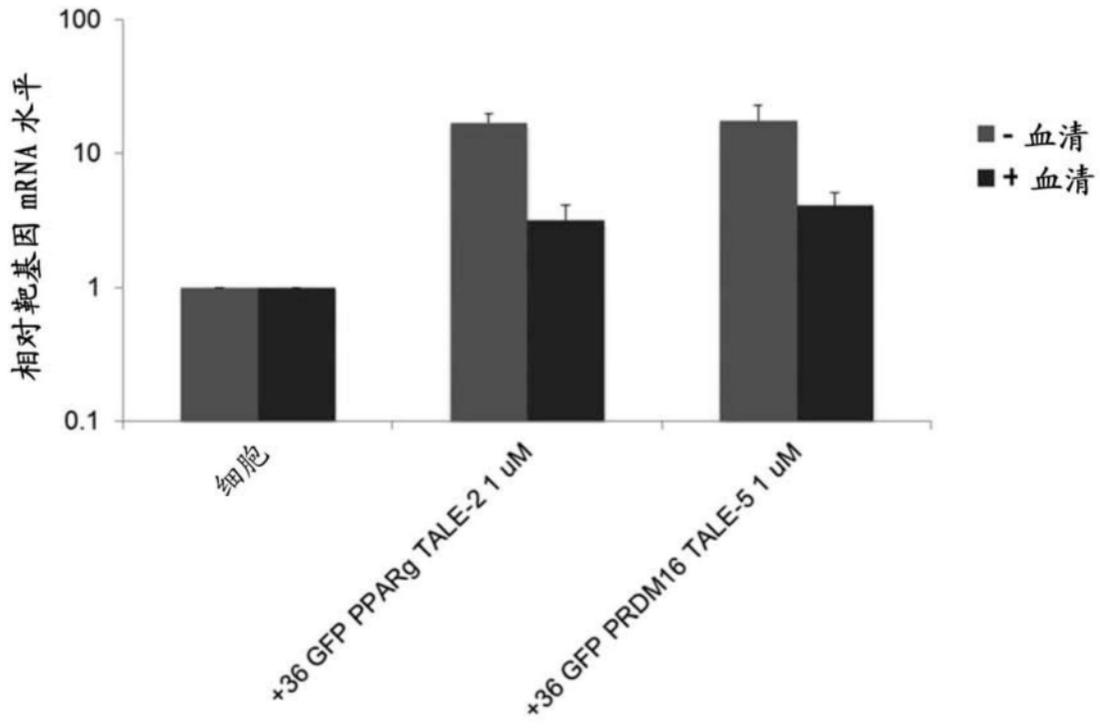


图11

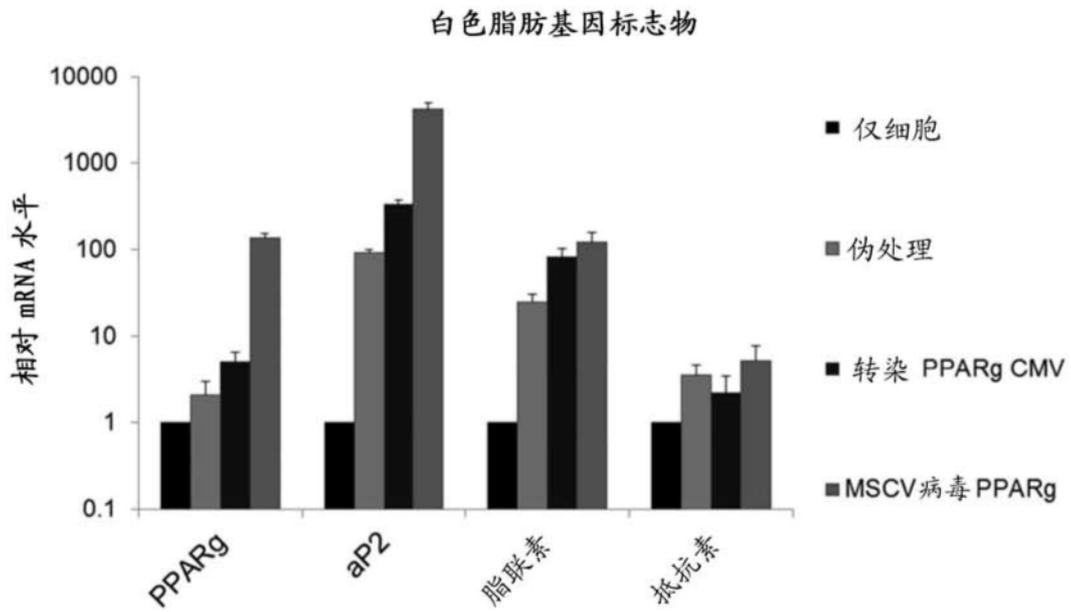


图12

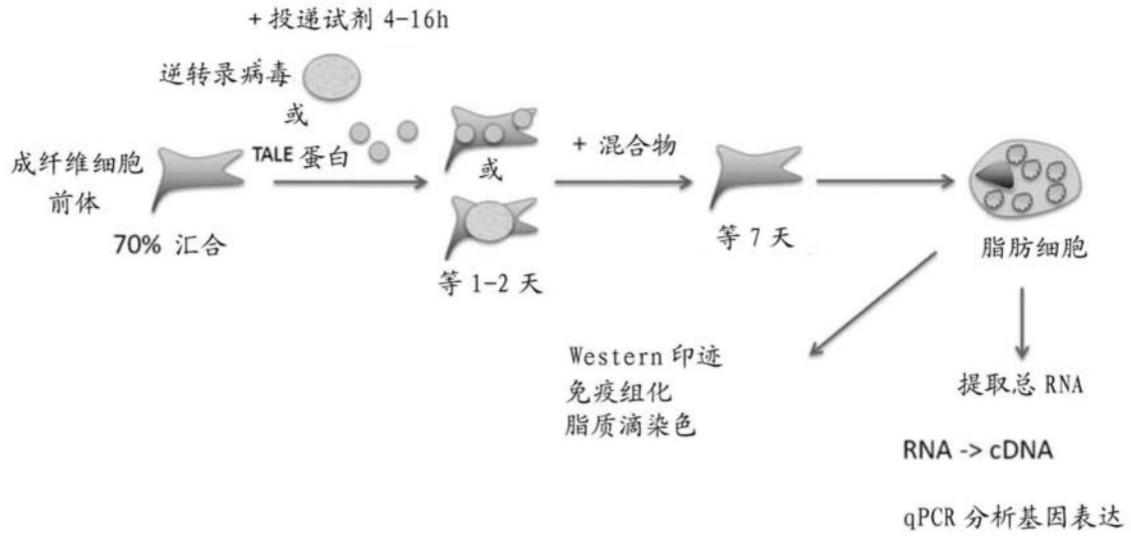


图13

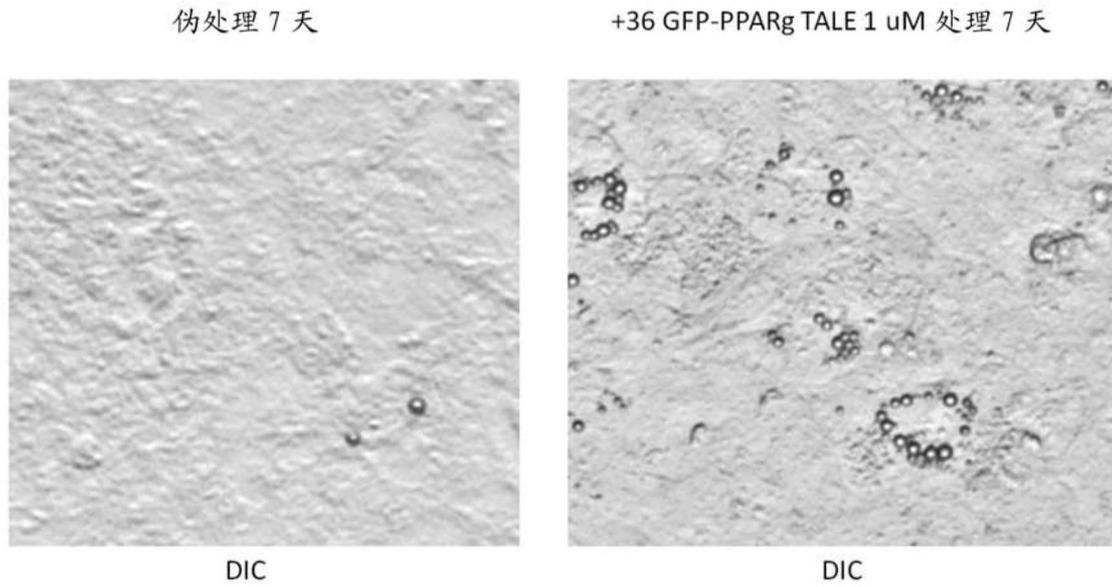


图14

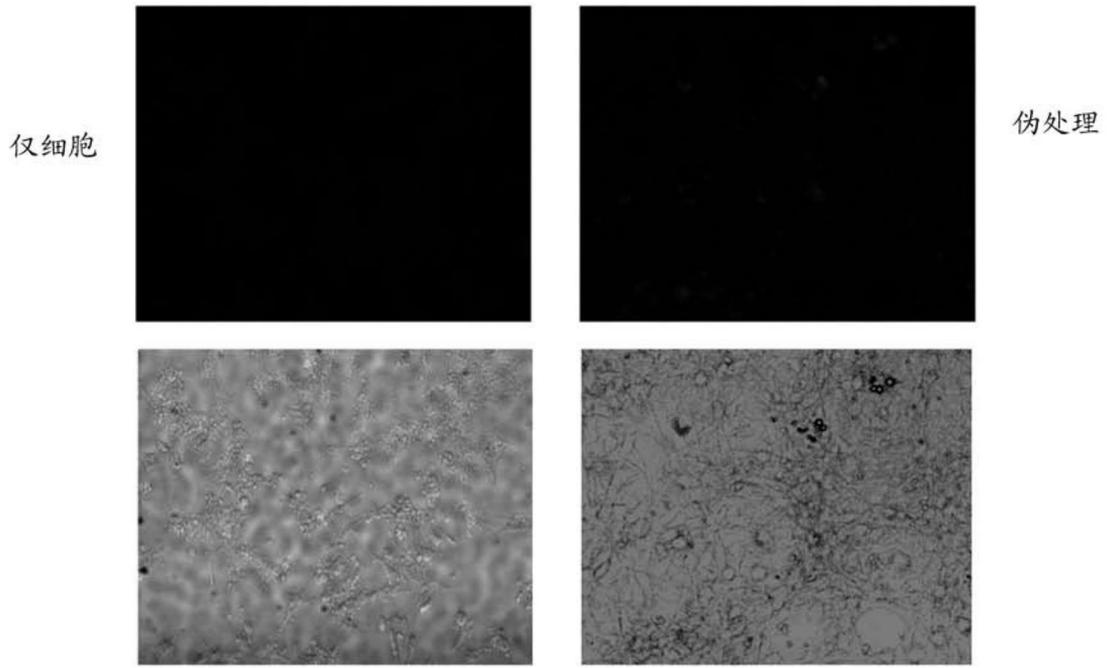


图15

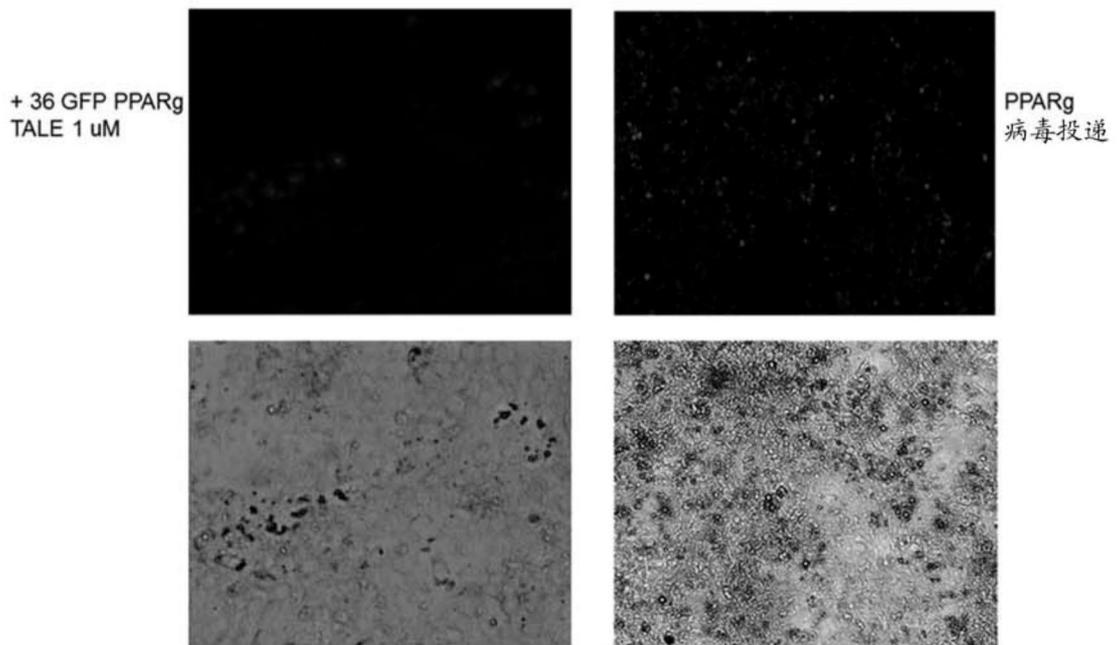


图16

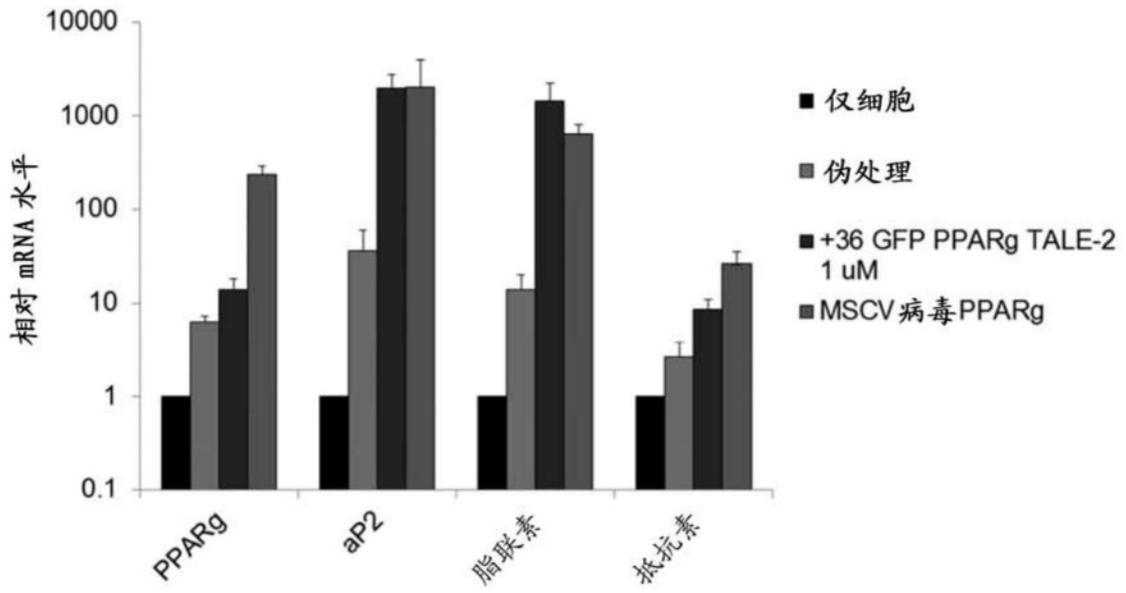


图17

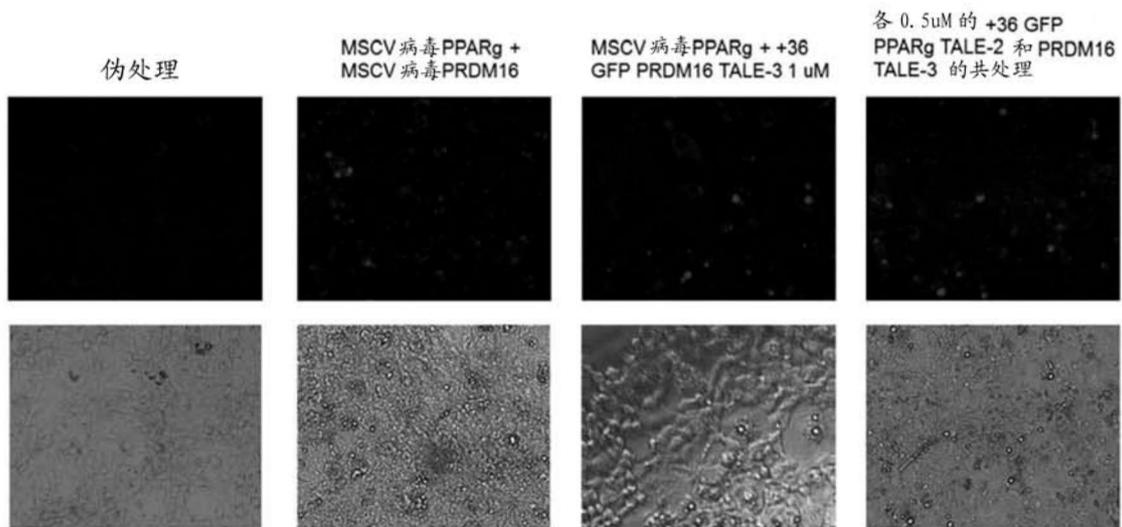


图18

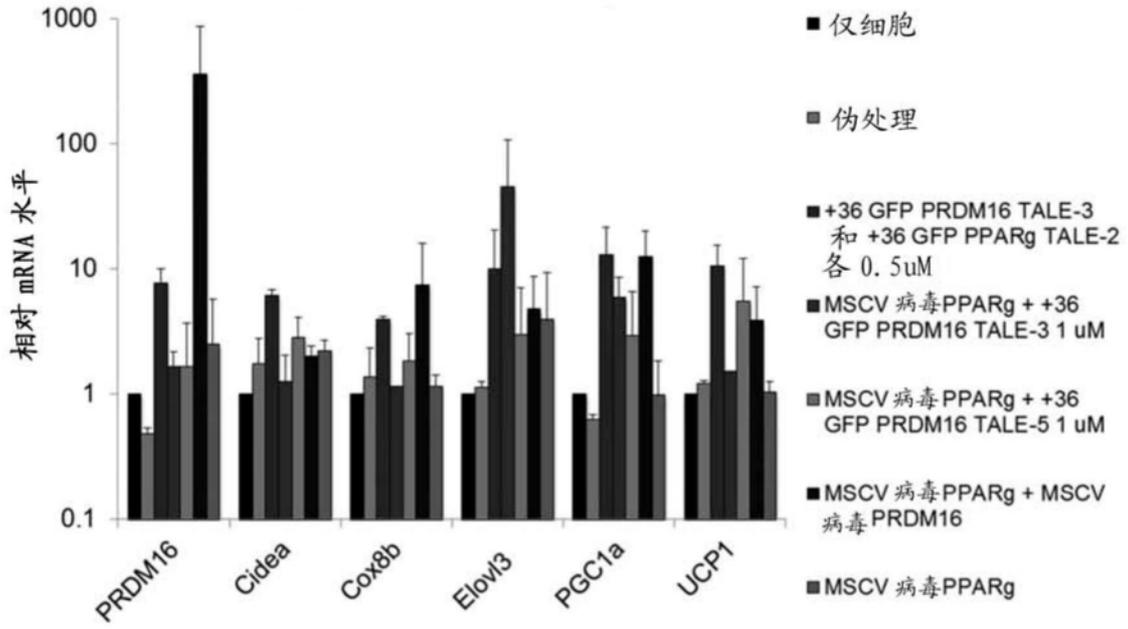


图19

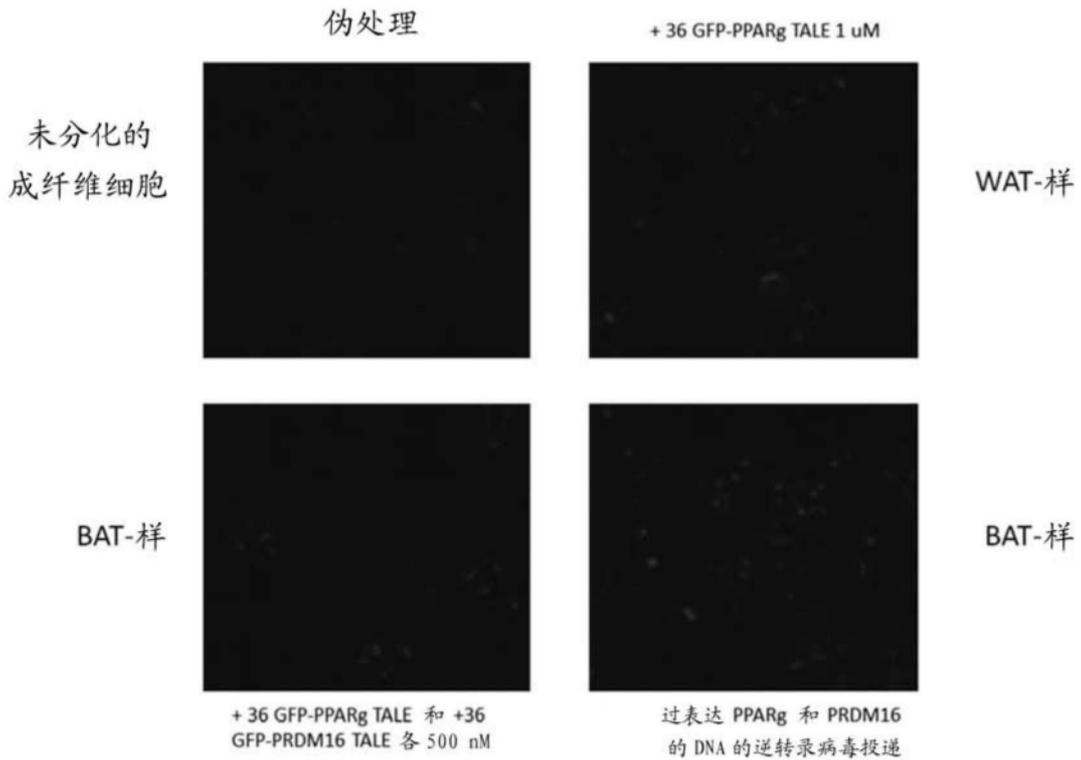


图20

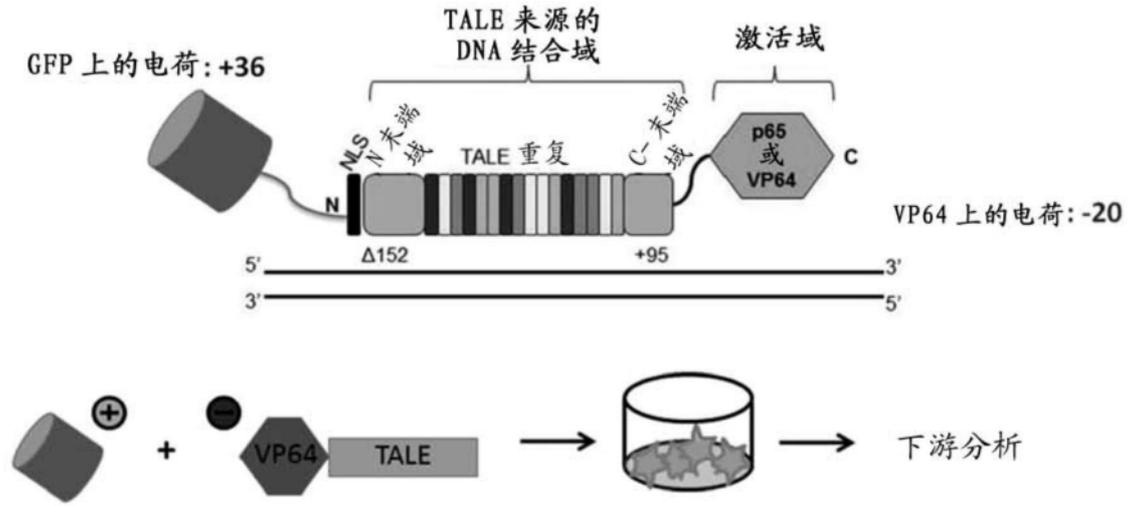


图21

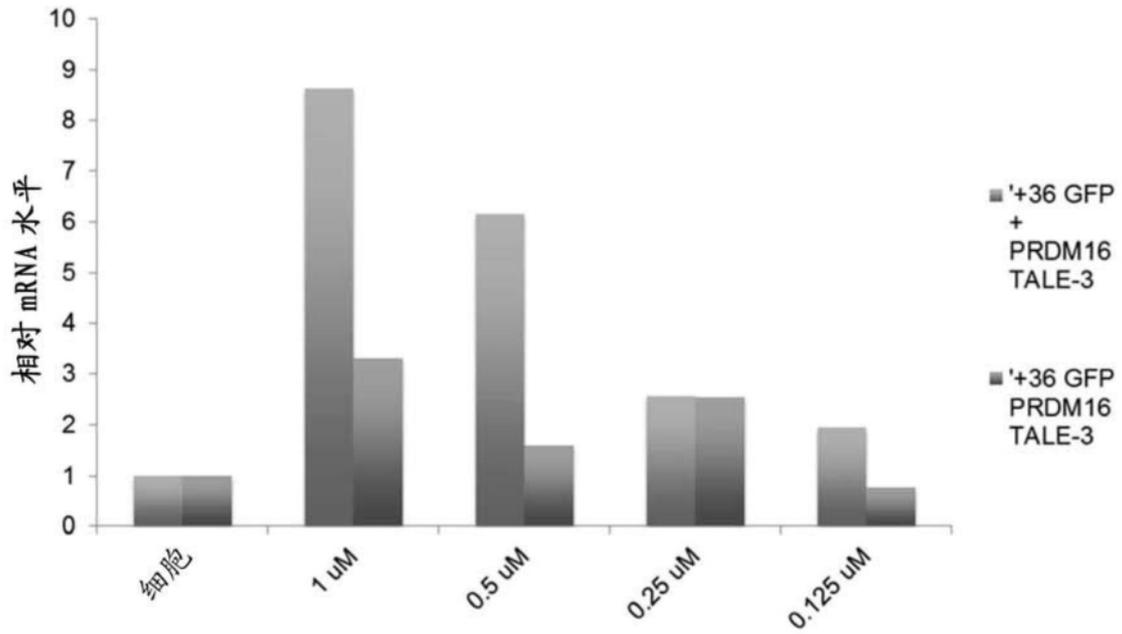


图22

PRDM16 基因表达

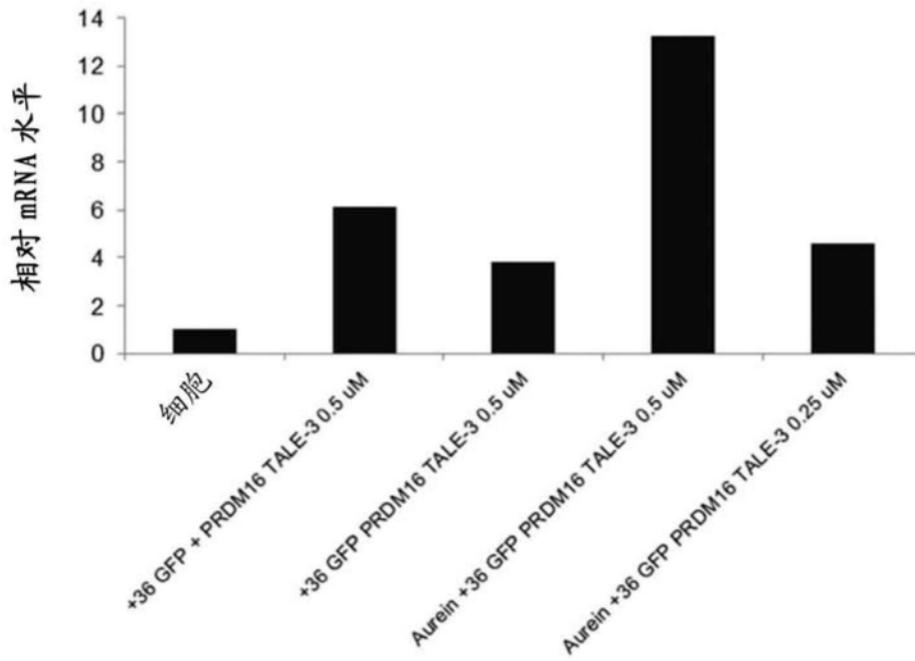


图23

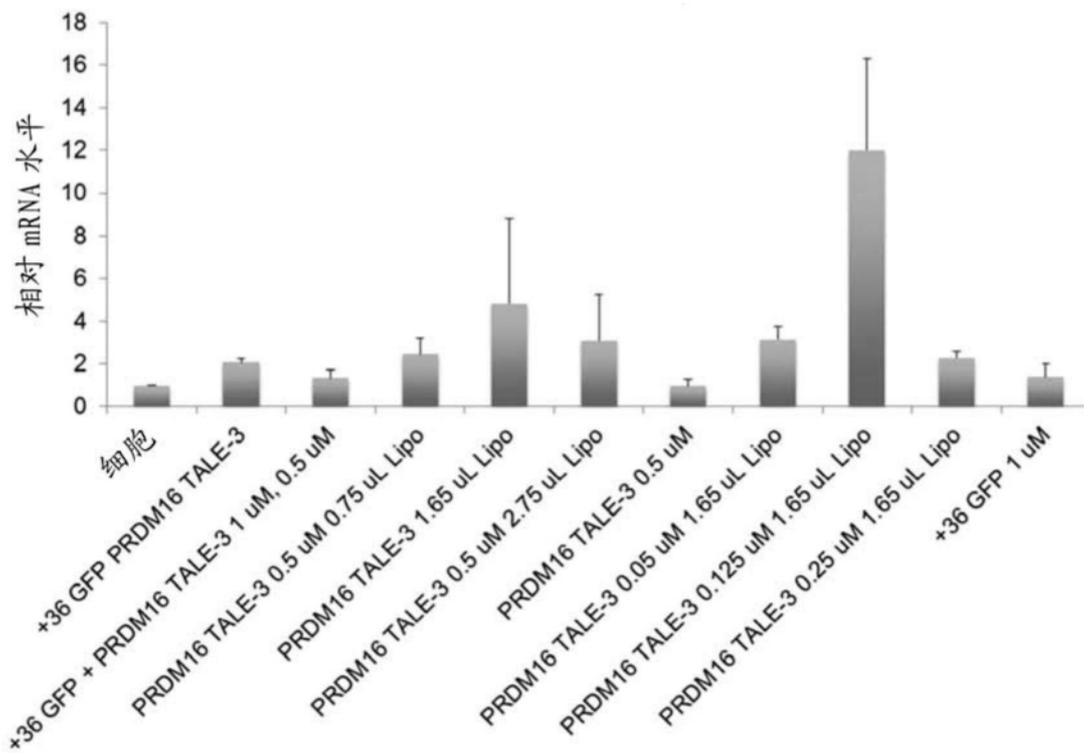


图24

由单个嵌合 RNA 编程的 Cas9

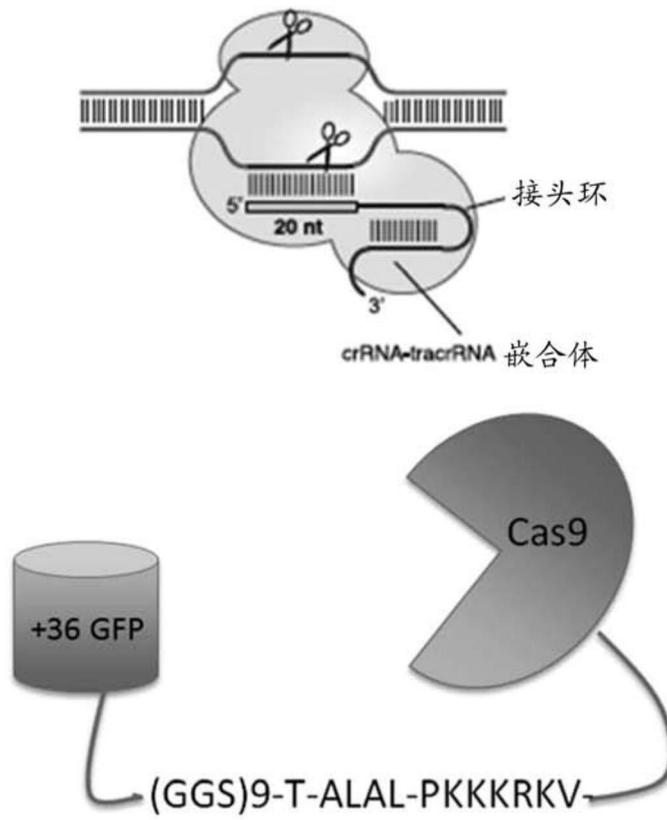


图25

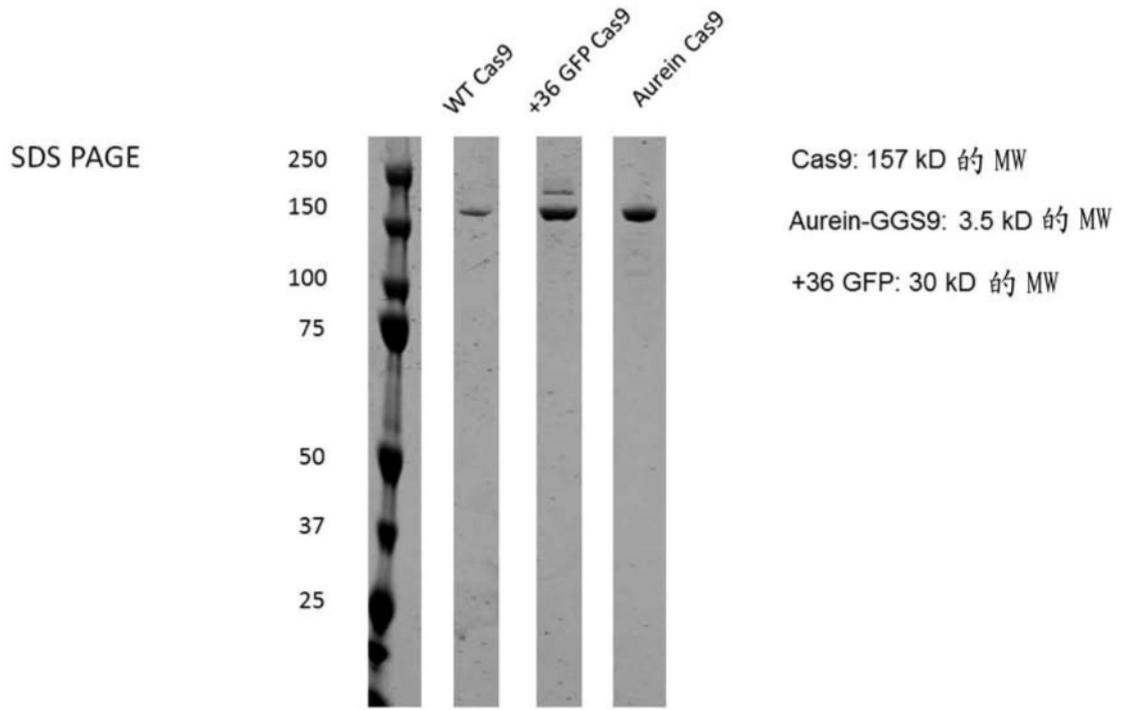


图26

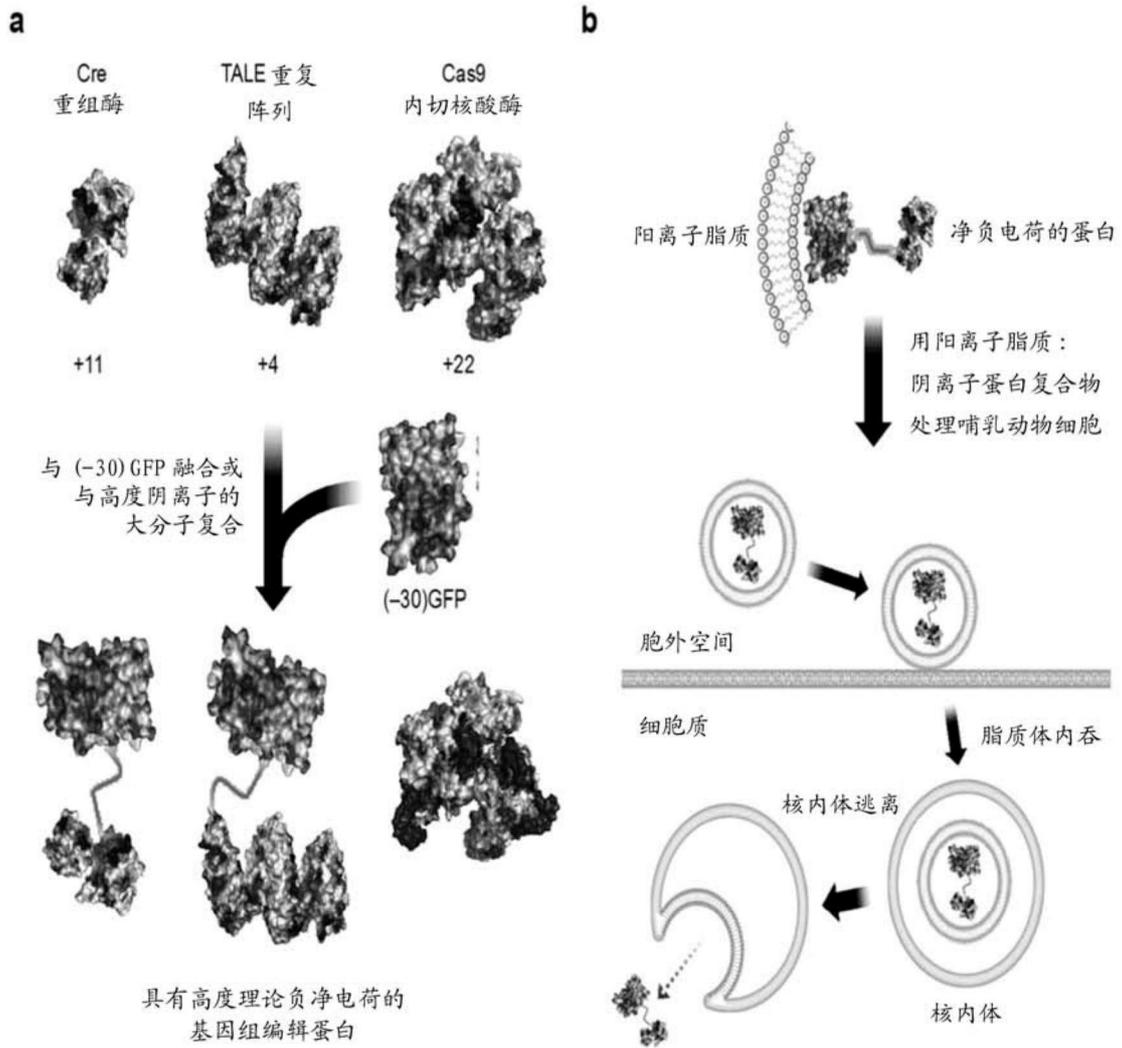


图27

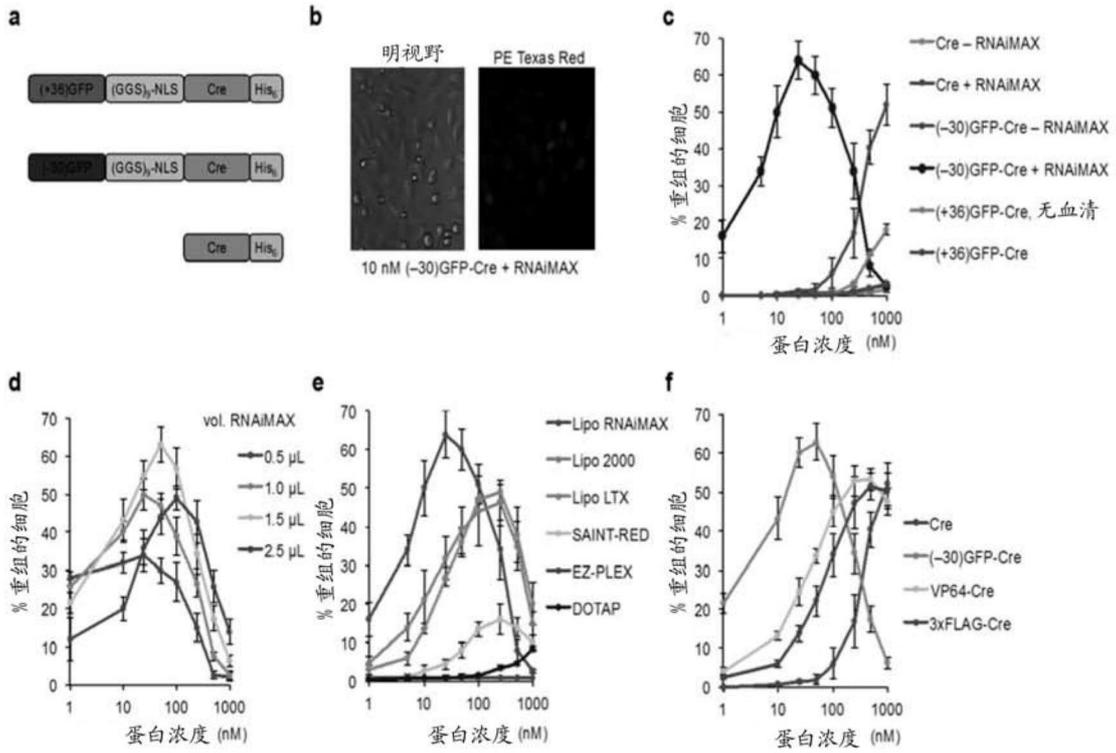


图28

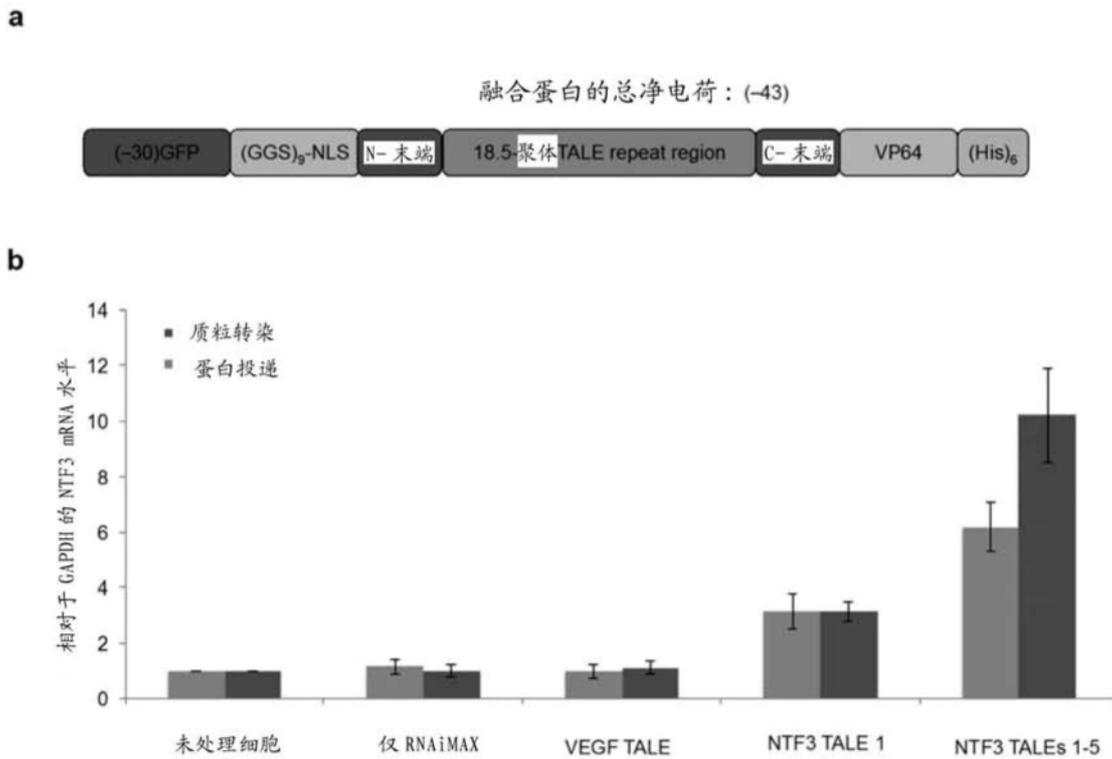


图29

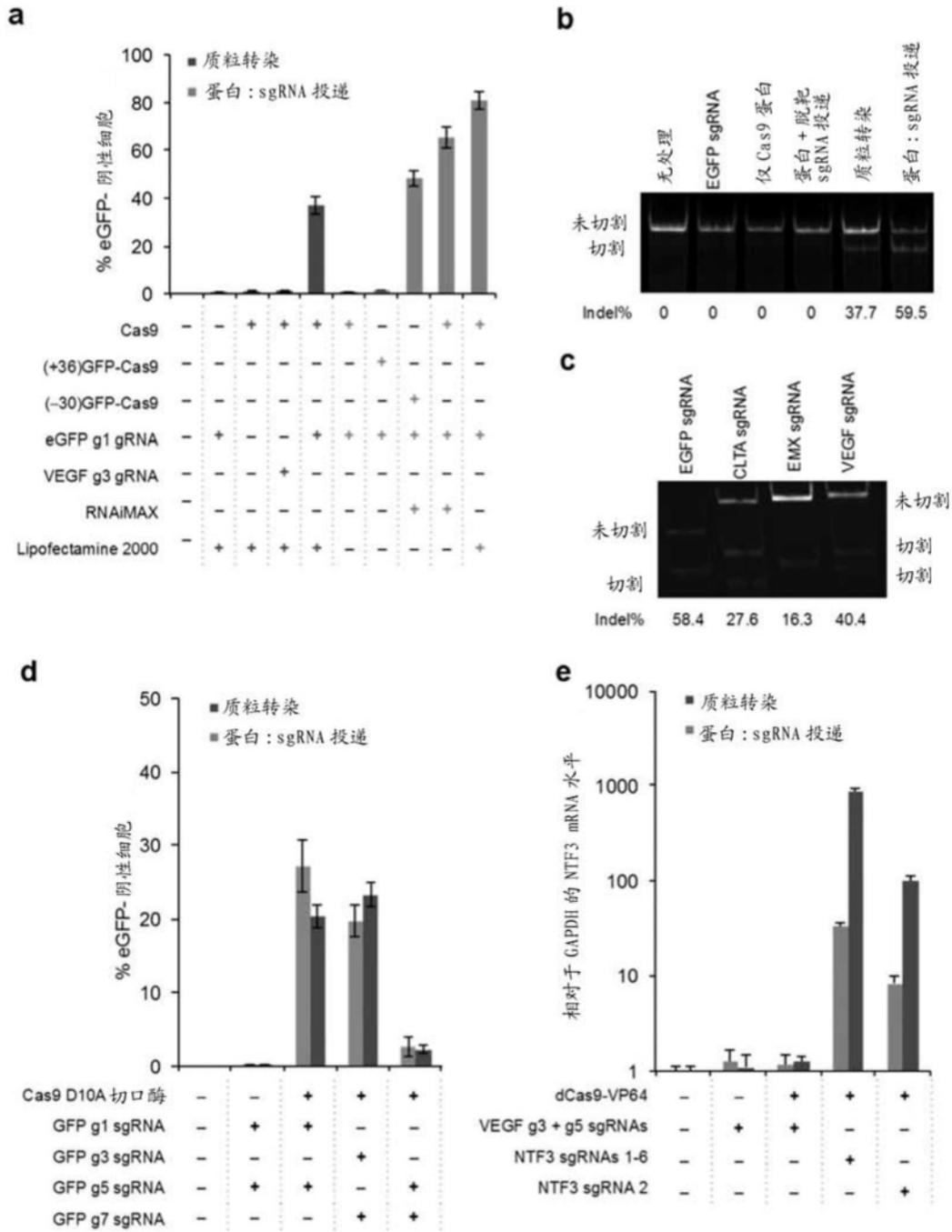


图30

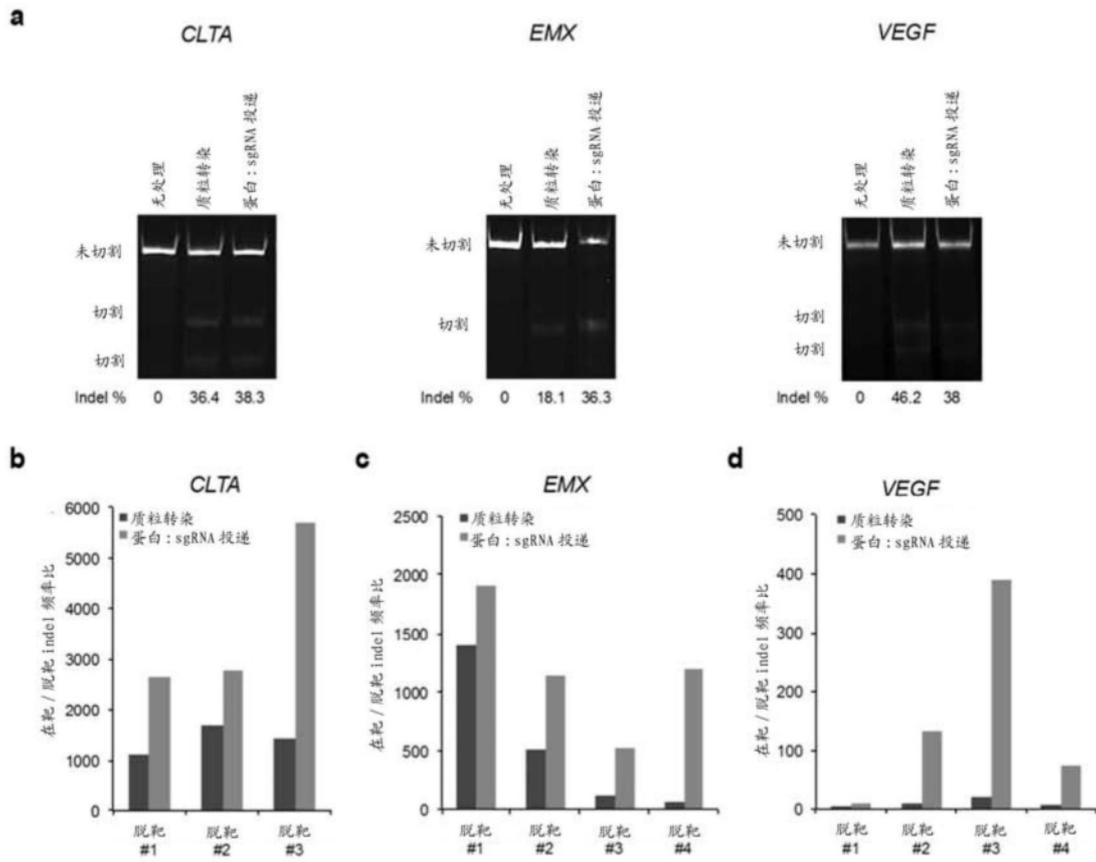


图31

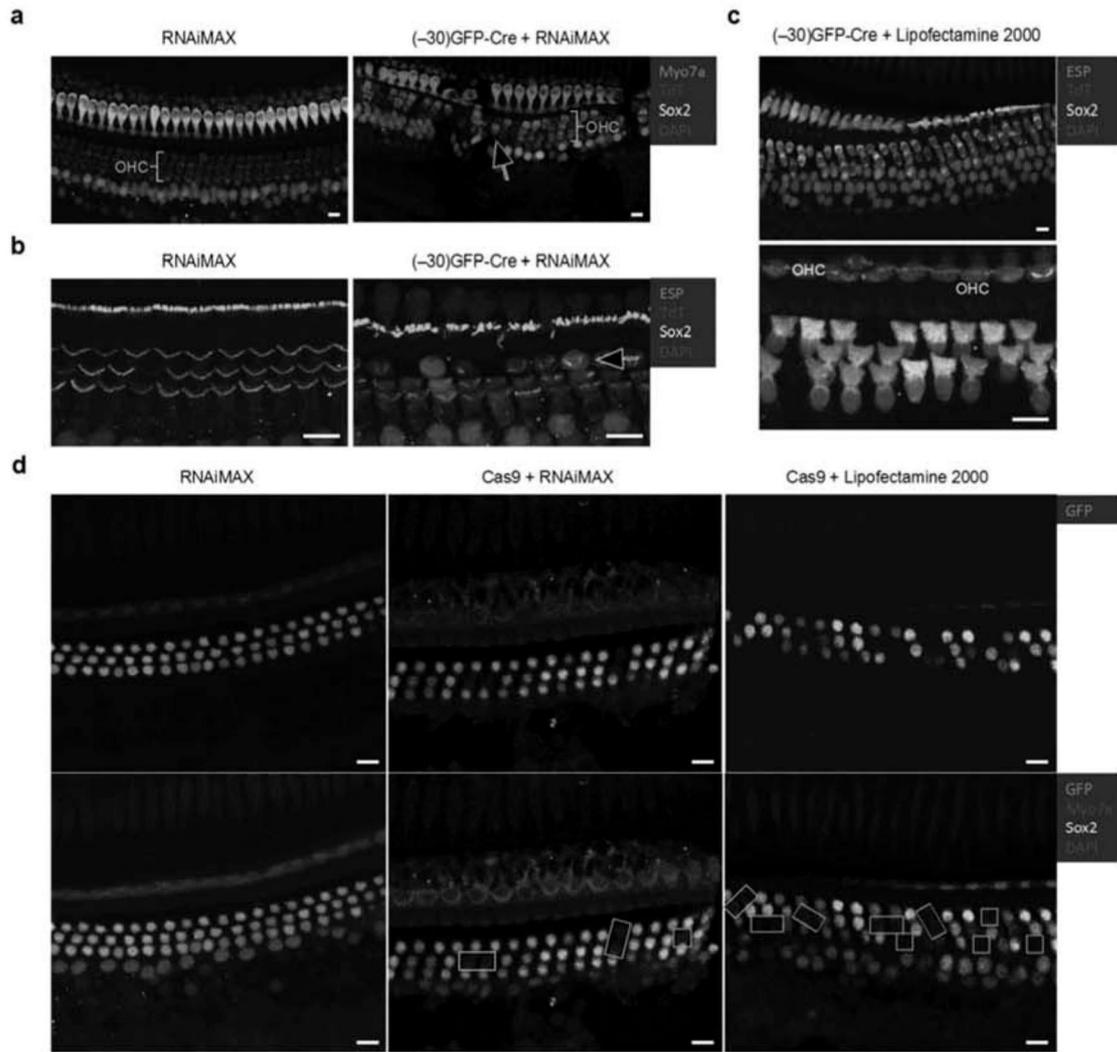


图32

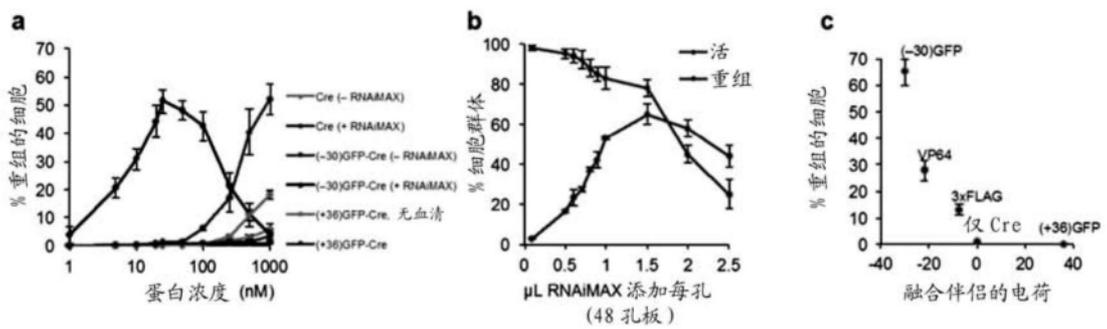


图33

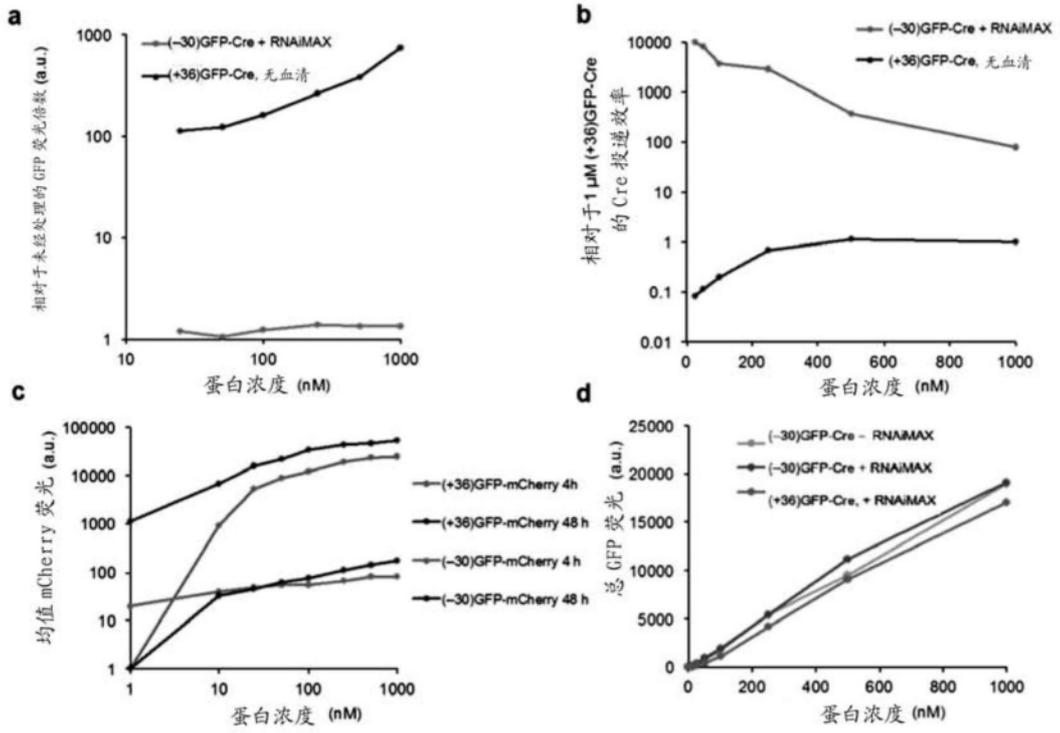


图34

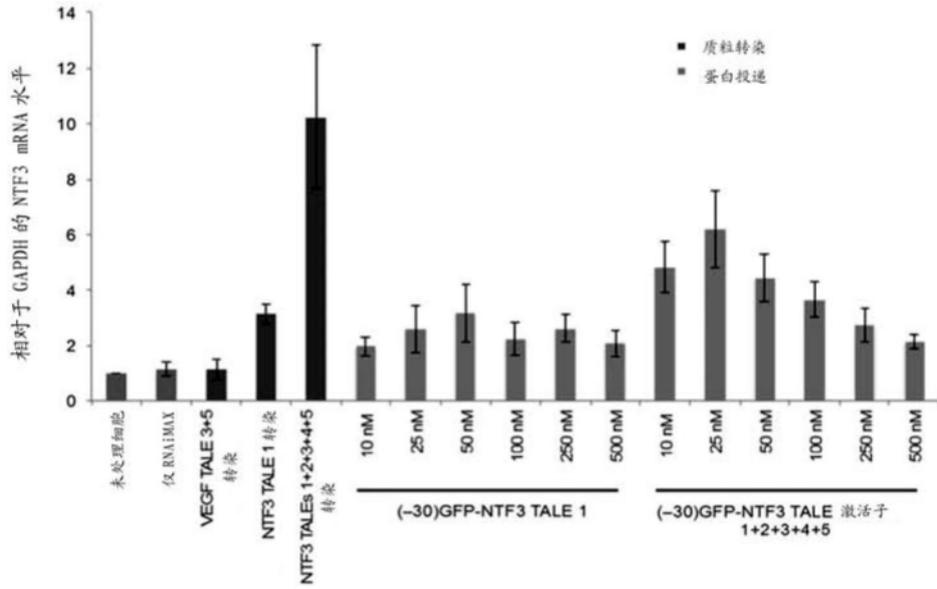


图35

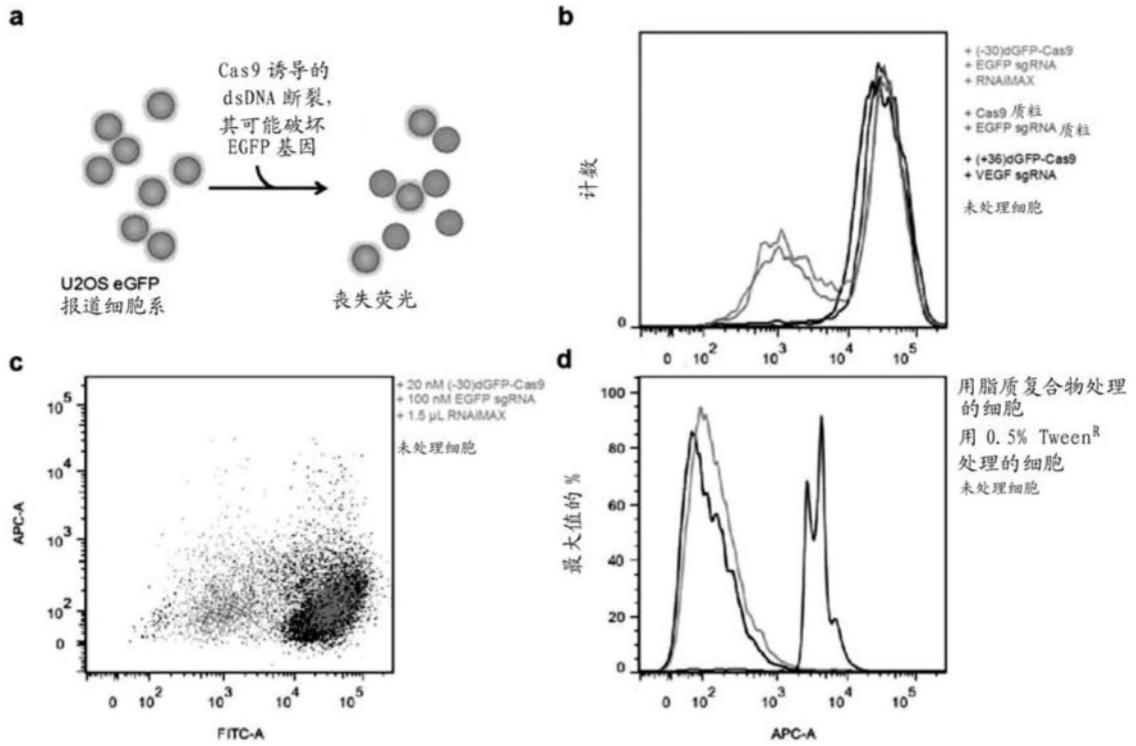


图36

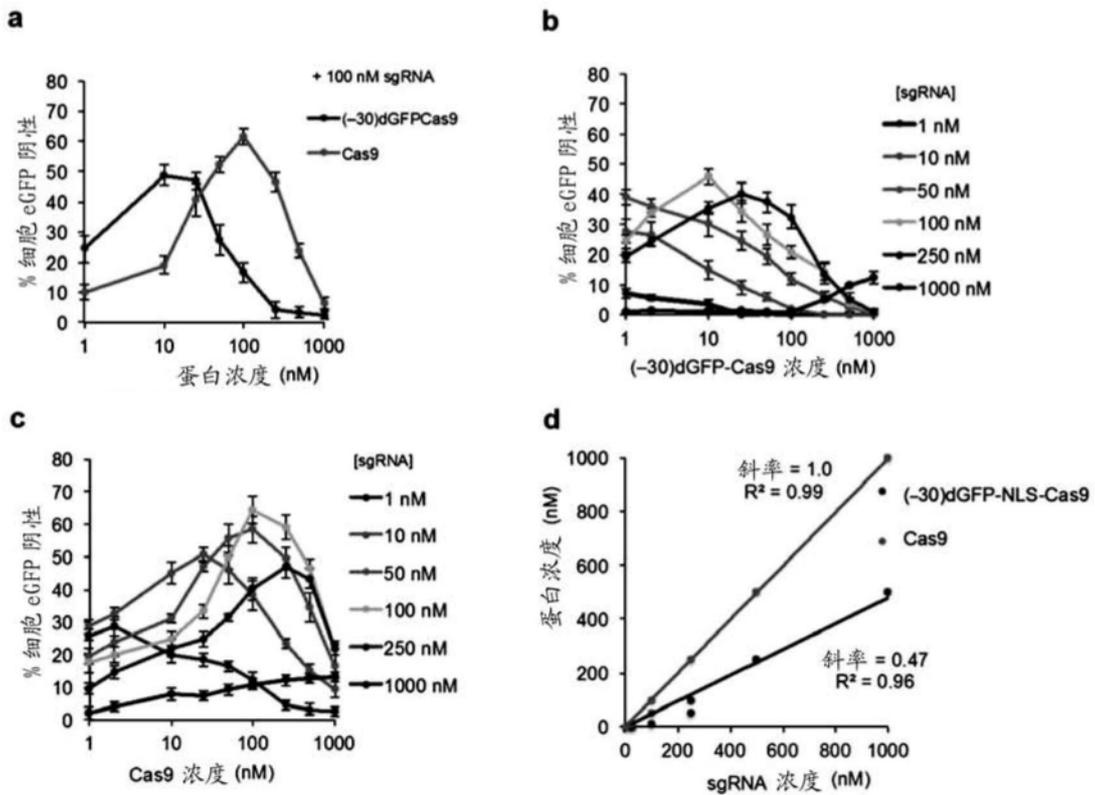


图37

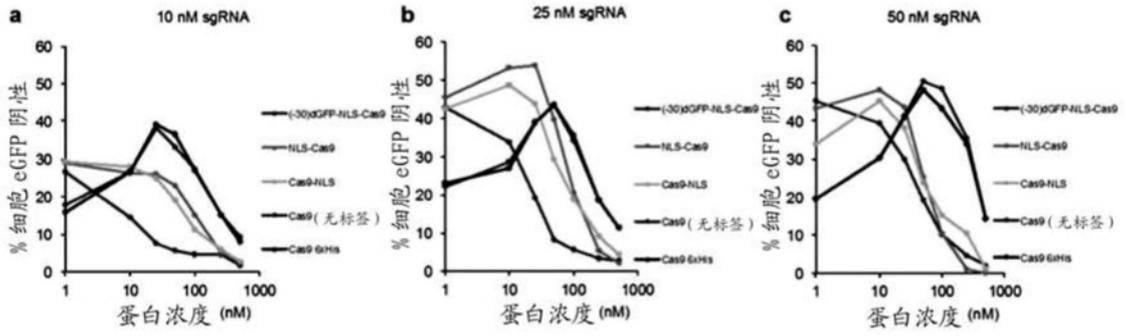


图38

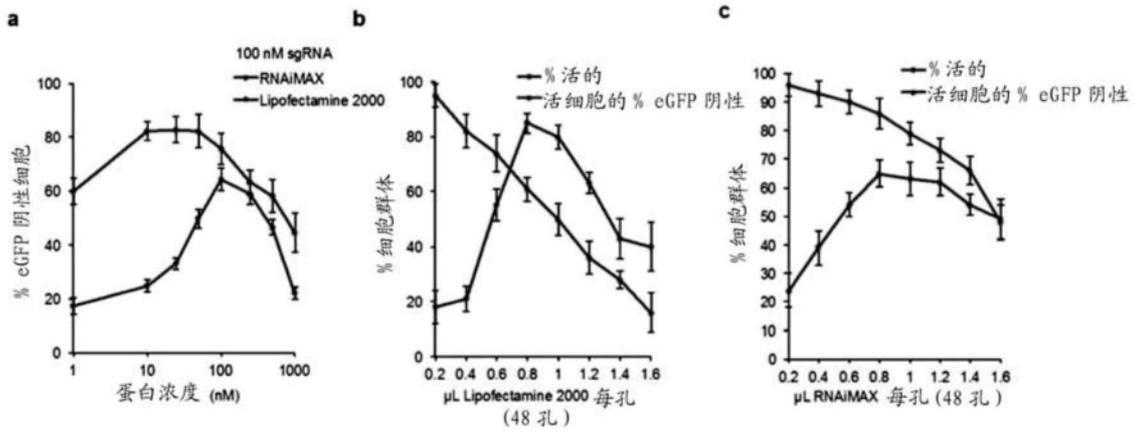


图39

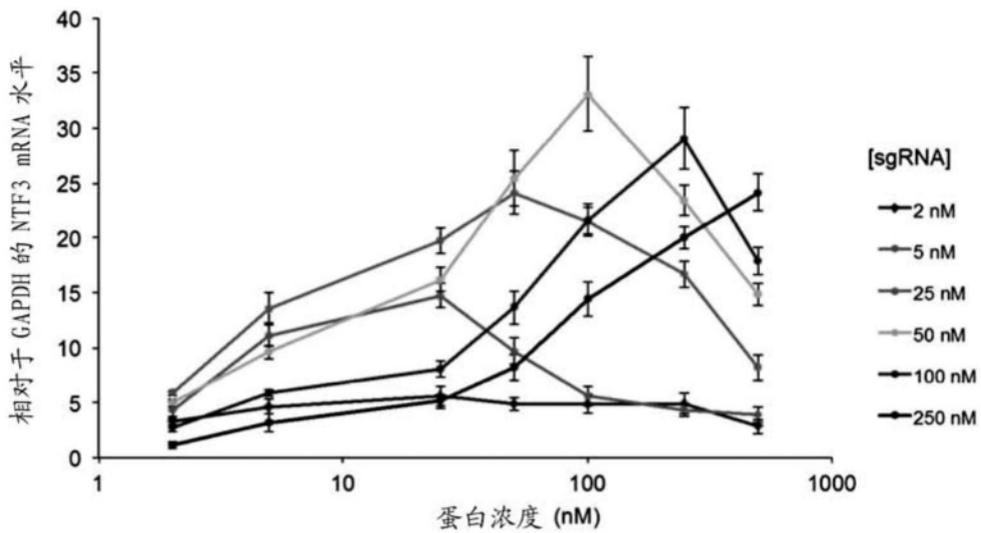


图40

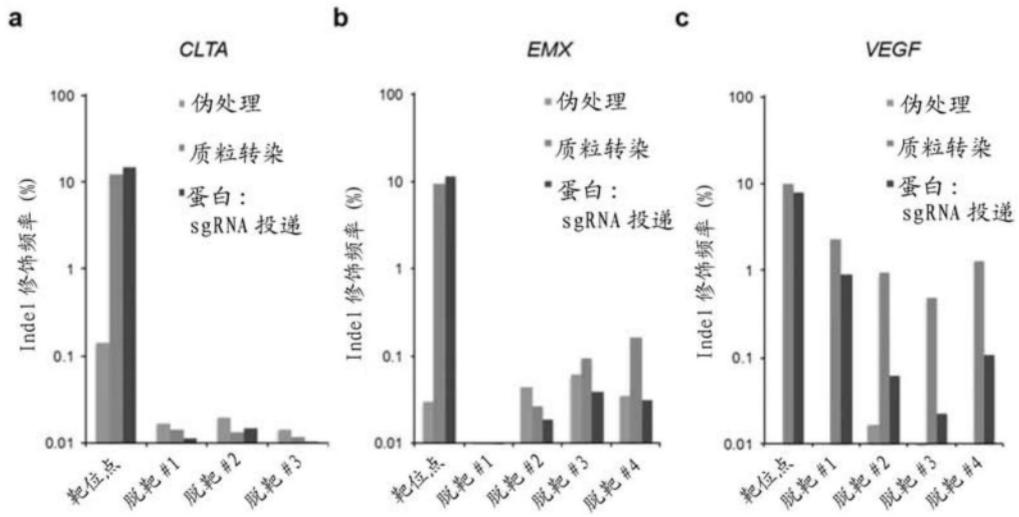


图41

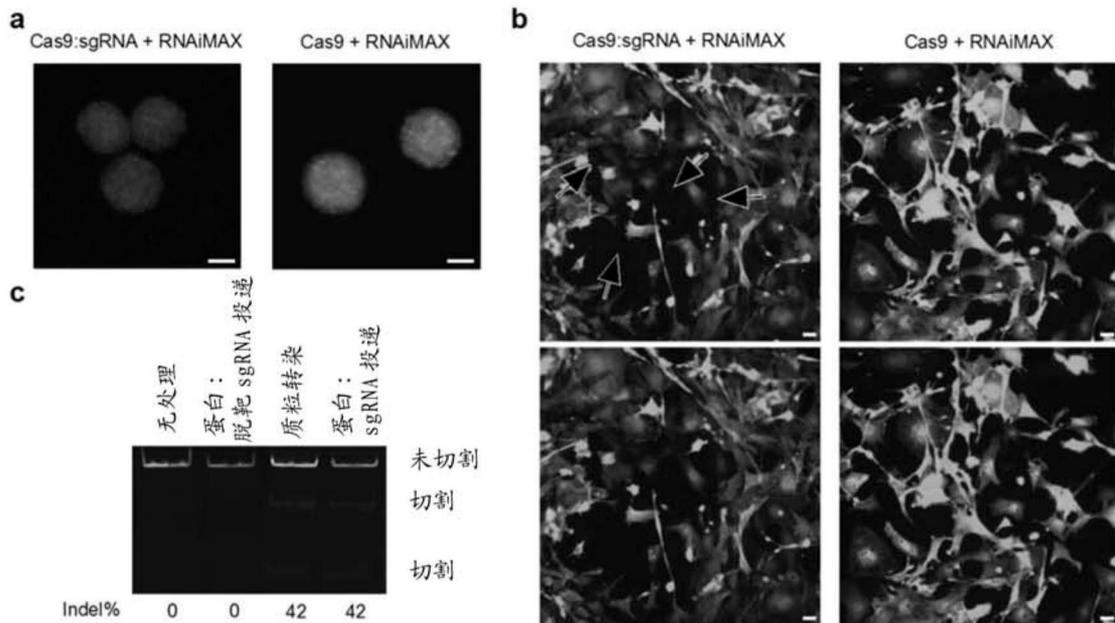


图42

a

野生型 GGGCGATGCCACCTACGGCAAGCTGACCCCTGAAGTTCACTGCAACCACGGGAAAGCTGGCCGTTGCCCTGGCCCACTCGTGACACCCCTGACCTACGGCGTGCAAGTCTCAGCCGCTACCCCGACCCACATG

缺失

12 GGGCGAT-----CGGCGTGCAAGTCTCAGCCGCTACCCCGACCCACATG
9 GGGCGATGCCACCTACGGCAAGCTGACCCCTGAAGTTCACTGCAACCACGGGAAAGCTGGCCGTTGCCCTGGCCCACTCGTGACACCCCTGACCTACGGCGTGCAAGTCTCAGCCGCTACCCCGACCCACATG
8 GGGCGATGCCACCTACGGCAAGCTGACCCCTGAAGTTCACTGCAACCACGGGAAAGCTGGCCGTTGCCCTGGCCCACTCGTGACACCCCTGACCTACGGCGTGCAAGTCTCAGCCGCTACCCCGACCCACATG
7 GGGCGATGCCACCTACGGCAAGCTG-----CCCGTGCCCTGGCCCACTCGTGACACCCCTGACCTACGGCGTGCAAGTCTCAGCCGCTACCCCGACCCACATG
6 GGGCGATGCCACCTACGGCAAGCTGACCCCTGAAGTTCACTGCAACCACGGGAAAGCTGGCCGTTGCCCTGGCCCACTCGTGACACCCCTGACCTACGGCGTGCAAGTCTCAGCCGCTACCCCGACCCACATG
4 GGGCGATGCCACCTACGGCAAGCTGACCCCTGAAGTTCACTGCAACCACGGGAAAGCTGGCCGTTGCCCTGGCCCACTCGTGACACCCCTGACCTACGGCGTGCAAGTCTCAGCCGCTACCCCGACCCACATG
3 GGGCGATGCCACCTACGGCAAGCTGACCCCTGAAGTTCACTGCAACCACGGGAAAGCTGGCCGTTGCCCTGGCCCACTCGTGACACCCCTGACCTACGGCGTGCAAGTCTCAGCCGCTACCCCGACCCACATG
3 GGGCGATGCCACCTACGGCAAGCTGACCCCTGAAGTTCACTGCAACCACGGGAAAGCTGGCCGTTGCCCTGGCCCACTCGTGACACCCCTGACCTACGGCGTGCAAGTCTCAGCCGCTACCCCGACCCACATG
2 GGGCGATGCCACCTACGGCAAGCTGACCCCTGAAGTTCACTGCAACCACGGGAAAGCTGGCCGTTGCCCTGGCCCACTCGTGACACCCCTGACCTACGGCGTGCAAGTCTCAGCCGCTACCCCGACCCACATG
2 GGGCGATGCCACCTACGGCAAGCTGACCCCTGAAGTTCACTGCAACCACGGGAAAGCTGGCCGTTGCCCTGGCCCACTCGTGACACCCCTGACCTACGGCGTGCAAGTCTCAGCCGCTACCCCGACCCACATG

插入

2 GGGCGATGCCACCTACGGCAAGCTGACCCCTGAAGTTCACTGCAACCACGGGAAAGCTGGCCGTTGCCCTGGCCCACTCGTGACACCCCTGACCTACGGCGTGCAAGTCTCAGCCGCTACCCCGACCCACATG
1 GGGCGATGCCACCTACGGCAAGCTGACCCCTGAAGTTCACTGCAACCACGGGAAAGCTGGCCGTTGCCCTGGCCCACTCGTGACACCCCTGACCTACGGCGTGCAAGTCTCAGCCGCTACCCCGACCCACATG
1 GGGCGATGCCACCTACGGCAAGCTGACCCCTGAAGTTCACTGCAACCACGGGAAAGCTGGCCGTTGCCCTGGCCCACTCGTGACACCCCTGACCTACGGCGTGCAAGTCTCAGCCGCTACCCCGACCCACATG

b

野生型 GGCAGAAGCTGGAAGAGGAAGGGCCGAGTCTGAGCAGAAGAAGGGTTCACCATATCAACCGGTGGCGCATCGCC

缺失

58 GGCAGAAGCTGGAAGAGGAAGGGCCGAGTCTGAGCAGAAGAAGGGTTCACCATATCAACCGGTGGCGCATCGCC
8 GGCAGAAGCTGGAAGAGGAAGGGCCGAGTCTGAGCAGAAGAAGGGTTCACCATATCAACCGGTGGCGCATCGCC
7 GGCAGAAGCTGGAAGAGGAAGGGCCGAGTCTGAGCAGAAGAAGGGTTCACCATATCAACCGGTGGCGCATCGCC
6 GGCAGAAGCTGGAAGAGGAAGGGCCGAGTCTGAGCAGAAGAAGGGTTCACCATATCAACCGGTGGCGCATCGCC
6 GGCAGAAGCTGGAAGAGGAAGGGCCGAGTCTGAGCAGAAGAAGGGTTCACCATATCAACCGGTGGCGCATCGCC
5 GGCAGAAGCTGGAAGAGGAAGGGCCGAGTCTGAGCAGAAGAAGGGTTCACCATATCAACCGGTGGCGCATCGCC
4 GGCAGAAGCTGGAAGAGGAAGGGCCGAGTCTGAGCAGAAGAAGGGTTCACCATATCAACCGGTGGCGCATCGCC
3 GGCAGAAGCTGGAAGAGGAAGGGCCGAGTCTGAGCAGAAGAAGGGTTCACCATATCAACCGGTGGCGCATCGCC
3 GGCAGAAGCTGGAAGAGGAAGGGCCGAGTCTGAGCAGAAGAAGGGTTCACCATATCAACCGGTGGCGCATCGCC
3 GGCAGAAGCTGGAAGAGGAAGGGCCGAGTCTGAGCAGAAGAAGGGTTCACCATATCAACCGGTGGCGCATCGCC

插入

4 GGCAGAAGCTGGAAGAGGAAGGGCCGAGTCTGAGCAGAAGAAGGGTTCACCATATCAACCGGTGGCGCATCGCC
1 GGCAGAAGCTGGAAGAGGAAGGGCCGAGTCTGAGCAGAAGAAGGGTTCACCATATCAACCGGTGGCGCATCGCC
1 GGCAGAAGCTGGAAGAGGAAGGGCCGAGTCTGAGCAGAAGAAGGGTTCACCATATCAACCGGTGGCGCATCGCC

图43