



(12)发明专利

(10)授权公告号 CN 105874071 B

(45)授权公告日 2020.04.14

(21)申请号 201480067248.8

E·J·瑞巴 T·韦克斯勒 张雷

(22)申请日 2014.12.09

(74)专利代理机构 上海专利商标事务有限公司 31100

(65)同一申请的已公布的文献号

代理人 余颖 陶家蓉

申请公布号 CN 105874071 A

(51)Int.Cl.

C12N 15/64(2006.01)

C12N 15/66(2006.01)

C12N 5/00(2006.01)

C12N 5/02(2006.01)

(43)申请公布日 2016.08.17

(30)优先权数据

61/913,838 2013.12.09 US

61/943,884 2014.02.24 US

(85)PCT国际申请进入国家阶段日

2016.06.08

(86)PCT国际申请的申请数据

PCT/US2014/069352 2014.12.09

(87)PCT国际申请的公布数据

W02015/089077 EN 2015.06.18

(73)专利权人 桑格摩生物科学股份有限公司

地址 美国加利福尼亚州

(72)发明人 J·C·米勒 D·帕斯乔恩

(56)对比文件

US 2013/0177960 A1, 2013.07.11,

Nathwani AC等. Adenovirus-associated virus vector-mediated gene transfer in hemophilia B.《The New England Journal of Medicine》.2011, 第365卷(第25期), 2357-2365.

审查员 刘树柏

权利要求书2页 说明书38页

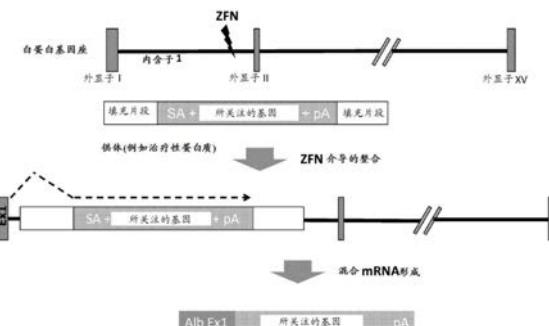
序列表18页 附图21页

(54)发明名称

基因组工程化的方法和组合物

(57)摘要

本文公开用于插入编码在疾病或病症诸如溶酶体贮积病中异常表达的蛋白质的转基因序列的方法和组合物。



1. 一种锌指蛋白核酸酶 (ZFN) 对, 其包含左 ZFN 和右 ZFN, 每个 ZFN 包含野生型或工程化裂解结构域或裂解半结构域和称为并且排序为 F1 至 F5 或 F1 至 F6 的五个或六个锌指结构域, 其中所述的锌指结构域每一个包含一个锌指识别螺旋区, 所述识别螺旋区如下所示:

(i) SBS#47171 的 F1 至 F5 的 锌指识别螺旋区, 其依次分别由 SEQ ID NO:23, 18, 24, 20 和 21 组成;

(ii) SBS#47898 的 F1 至 F6 的 锌指识别螺旋区, 其依次分别由 SEQ ID NO:14, 28, 25, 15, 16 和 12 组成;

(iii) SBS#47079 或 SBS#47169 的 F1 至 F6 的 锌指识别螺旋区, 其依次分别由 SEQ ID NO:14, 9, 10, 15, 16 和 12 组成;

(iv) SBS#47931 的 F1 至 F6 的 锌指识别螺旋区, 其依次分别由 SEQ ID NO:14, 9, 25, 15, 16 和 12 组成;

(v) SBS#47863 的 F1 至 F5 的 锌指识别螺旋区, 其依次分别由 SEQ ID NO:2, 26, 24, 20, 和 27 组成; 或

(vi) SBS#47192 的 F1 至 F5 的 锌指识别螺旋区, 其依次分别由 SEQ ID NO:2, 26, 24, 20 和 21 组成,

其中左 ZFN 包括选自 (i)、(v) 和 (vi) 的 锌指识别螺旋区, 右 ZFN 包括选自 (ii)、(iii) 和 (iv) 的 锌指识别螺旋区。

2. 如权利要求 1 所述的 锌指蛋白核酸酶 对, 其中左 ZFN 包括 (i) 的 锌指识别螺旋区, 右 ZFN 包含 (ii) 的 锌指识别螺旋区。

3. 一种或多种多核苷酸, 其编码权利要求 1 或 权利要求 2 所述的 锌指核酸酶 对。

4. 一种或多种表达载体, 其包含如 权利要求 3 所述的一种或多种多核苷酸。

5. 如 权利要求 4 所述的一种或多种表达载体, 其中所述载体是 AAV 载体。

6. 如 权利要求 5 所述的表达载体, 其中所述 AAV 载体是 AAV2/8 载体。

7. 一种药物组合物, 其包含如 权利要求 4 至 6 中任一项所述的表达载体。

8. 一种裂解细胞中的白蛋白基因的体外方法, 所述方法包括:

将一种或多种如 权利要求 4 至 6 中任一项所述的表达载体在使得所述一种或多种蛋白  
质得以表达并且所述白蛋白基因得以裂解的条件下引入所述细胞中。

9. 如 权利要求 8 所述的方法, 其中供体序列整合至所述裂解白蛋白基因中。

10. 如 权利要求 9 所述的方法, 其中使用 AAV 载体引入所述供体序列。

11. 如 权利要求 10 所述的方法, 其中所述 AAV 载体是 AAV2/8 载体。

12. 如 权利要求 9 至 11 中任一项所述的方法, 其中所述供体序列编码在溶酶体贮积病  
(LSD) 中缺少或有缺陷的蛋白质。

13. 如 权利要求 8 至 12 中任一项所述的方法, 其中所述细胞是肝细胞。

14. 如 权利要求 4 至 6 中任一项所述的表达载体或如 权利要求 7 所述的药物组合物的用  
途, 所述表达载体或药物组合物介导编码在患有溶酶体贮积病的受试者中缺少的功能性蛋  
白质的转基因靶向整合至内源性白蛋白基因中, 所述用途是用于制备治疗溶酶体贮积病的  
药物。

15. 如 权利要求 14 所述的用途, 其中静脉内施用所述表达载体或药物组合物和转基因。

16. 一种用于切割白蛋白基因的试剂盒, 其包括如 权利要求 4 至 6 中任一项所述的表达

载体。

## 基因组工程化的方法和组合物

- [0001] 相关申请的交叉引用
- [0002] 本申请要求2013年12月9日提交的美国临时申请号61/913,838和2014年2月24日提交的美国临时申请号61/943,884的权益，其公开内容全部以引用方式并入本文。
- [0003] 在联邦科研资助下进行的本发明权利的声明
- [0004] 不适用。

### 技术领域

[0005] 本公开是涉及基因修饰和血友病治疗领域。

[0006] 背景

[0007] 基因疗法的尤其有吸引力的应用涉及治疗由于分泌基因产物不足所导致的或可通过分泌治疗性蛋白质来治疗的病症。这些病症潜在地可经由将治疗性转基因递送至适度数量的细胞来处理，只要每个受体细胞表达高水平的治疗性基因产物。在此类情况下，免除将基因递送至大量细胞的需要可实现成功研发基因疗法供治疗用其他手段难治疗的适应症。这些应用需要持久、安全和很高水平的转基因表达。因此，研发展现这些性质的安全港将在基因治疗领域提供实质性效用。

[0008] 血友病诸如甲型血友病和乙型血友病，是凝血系统的遗传病症，其特征是出血至关节和软组织中，和过量出血至经历外伤或经历手术的任何位点。甲型血友病在临幊上与乙型血友病难分辨，但是因子VIII (FVIII或F8) 在甲型血友病中缺乏或不存在，而因子IX (FIX或F.IX) 在患有乙型血友病的患者中缺乏或不存在。F8基因编码血浆糖蛋白，所述血浆糖蛋白以其非活性形式与冯威勒布兰德氏因子相关联地进行循环。在表面损伤后，内在凝血级联启动并且因子VIII从复合物中释放并且变得活化。活化形式与因子IX一起操作来激活因子X变成活化Xa，最后导致纤维蛋白原变化至纤维蛋白和凝结诱导。参见，Levinson等人 (1990) *Genomics* 7(1) :1-11. 40-50% 的甲型血友病患者患有涉及F8内含子22的染色体倒位 (也称为IVS22)。倒位的原因是F8基因的内含子22中的9.6kb序列与位于F8基因远侧约300kb处的两个密切相关反向序列中的一个之间的染色体内重组事件，导致相对于外显子23至26的外显子1至22倒位。参见，*Textbook of Hemophilia*, Lee等人 (编) 2005, Blackwell Publishing。其他甲型血友病患者在F8中具有缺陷，包括活性位点突变，和无义和错义突变。就其本身来说，因子IX (F.IX) 编码涉及凝血系统的丝氨酸蛋白水解酶中的一个，并且已经证明恢复野生型因子IX蛋白质的正常循环水平的甚至3% 可预防自发出血。额外血友病与其他凝血因子的异常表达相关联。举例来说，因子VII缺陷是在300,000至500,000人中的约1个人中发生的常染色体隐性性状并且与患者体内的不适当因子VII水平相关联。类似地，因子X缺陷也是在每500,000至1百万人中的1个人中发生的常染色体隐性性状，并且由FX基因的遗传变体导致。因子X缺陷可在患者群体中具有不同程度的严重性。

[0009] 乙型血友病的当前治疗依赖于长期、重复静脉内输注纯化重组因子IX并且受许多缺点影响。这包括需要重复静脉内输注，与抑制剂形成相关联，并且是预防性而非治愈性的。

[0010] 已经描述患有甲型或乙型血友病的患者的基因疗法,其涉及引入质粒和编码功能F8或F.IX蛋白质的其他载体(例如,AAV)。参见例如美国专利号6,936,243;7,238,346和6,200,560;Shi等人(2007)J Thromb Haemost.(2):352-61;Lee等人(2004)Pharm.Res.7:1229-1232;Graham等人(2008)Genet Vaccines Ther.3:6-9;Manno等人(2003)Blood 101(8):2963-72;Manno等人(2006)Nature Medicine 12(3):342-7;Nathwani等人(2011)Molecular Therapy 19(5):876-85;Nathwani等人(2011);N Engl J Med.365(25):2357-65。然而,在这些方案中,形成抑制性抗-因子VIII或IX(抗-F8或抗-F.IX)抗体和针对递送媒介物的抗体仍然是血友病的基于F8和F.IX替换的疗法的主要并发症。参见例如Scott&Lozier(2012)Br J Haematol.156(3):295-302。

[0011] 溶酶体贮积疾病(LSD)是一组罕见代谢单基因疾病,其特征是缺乏通常参与分解废弃脂质、糖蛋白和粘多糖的功能性个别溶酶体蛋白质。这些疾病的特征是这些化合物在细胞中积聚,因为由于特定酶失效,细胞不能对这些化合物进行处理以便再循环。最常见实例是戈谢氏(葡萄糖脑苷脂酶缺陷-基因名称:GBA)、法布里氏( $\alpha$ 半乳糖苷酶缺陷-GLA)、亨特氏(艾杜糖酸酯-2-硫酸酯酶缺陷-IDS)、赫尔勒氏( $\alpha$ -L艾杜糖苷酸酶缺陷-IDUA),和尼曼匹克氏(鞘磷脂磷酸二酯酶1缺陷-SMPD1)疾病。当组合在一起时,LSD具有在7000个新生儿的群体中约1个病例的发病率。这些疾病对于患病的人具有破坏性效应。其通常首先在婴儿中诊断,这些婴儿可具有特有面部和身体生长模式并且可患有中度至重度智力迟钝。治疗选项包括酶替换疗法(ERT),其中缺少的酶通常以大剂量经由静脉内注射来给予患者。这类治疗只治疗症状并且不能治愈,因而患者必须在其余生接受这些蛋白质的重复给药,并且潜在地可显现针对所注射蛋白质的中和抗体。经常这些蛋白质具有短血清半衰期,并且因此患者还必须承受蛋白质的频繁输注。举例来说,接受Cerezyme®产品(伊米昔酶)的戈谢氏疾病患者必须每周输注三次。酶的产生和纯化也是成问题的,并且因此治疗成本极高(>\$100,000/年/患者)。

[0012] 已经描述基因组DNA的靶向裂解的各种方法和组合物。这些靶向裂解事件可用于例如诱导靶向诱变、诱导细胞DNA序列的靶向缺失,和促进预定染色体基因座处的靶向重组。参见,例如,美国专利号8,623,618;8,034,598;8,586,526;6,534,261;6,599,692;6,503,717;6,689,558;7,067,317;7,262,054;7,888,121;7,972,854;7,914,796;7,951,925;8,110,379;8,409,861;美国专利公布20030232410;20050208489;20050026157;20060063231;20080159996;201000218264;20120017290;20110265198;20130137104;20130122591;20130177983和20130177960和美国申请号14/278,903,其公开内容出于所有目的以引用方式全部并入。这些方法经常涉及使用工程化裂解系统来诱导双链断裂(DSB)或靶DNA序列中的切口以使得通过产生误差的过程诸如非同源末端连接(NHEJ)来修复断裂或使用修复模板的修复(同源性引导修复或HDR)可产生基因的剔除或插入所关注的序列(靶向整合)。此技术也可用于经由使用供体寡核苷酸来在基因组序列中引入位点特异性变化,包括引入基因组区域的特异性缺失,或特异性点突变或局部改变(也称为基因校正)。裂解可经由使用特异性核酸酶诸如工程化锌指核酸酶(ZFN)、转录活化剂样效应核酸酶(TALEN)来发生,或使用具有工程化crRNA/tracr RNA('单一引导RNA')的CRISPR/Cas系统来引导特异性裂解。此外,开发基于Argonaute系统(例如,来自嗜热栖热菌,被称为'TtAgo',参见Swarts等人(2014)Nature 507(7491):258-261)的靶向核酸酶,所述系统还

可具有用于基因组编辑和基因疗法的潜力。

[0013] 如与典型整合方法相比,此核酸酶介导的靶向转基因插入方法提供改进转基因表达、增加安全和表达持久性的希望,因为它允许精确转基因定位以将基因沉默或活化相邻致癌基因的风险减少到最低限度。

[0014] 转基因可靶向整合至其同源基因座中,例如,将野生型转基因插入内源性基因座以校正突变基因。或者,转基因可插入非同源基因座例如“安全港”基因座中。已经描述一些安全港基因座,包括CCR5、HPRT、AAVS1、Rosa和白蛋白。参见,例如,美国专利号7,951,925和8,110,379;美国公布号20080159996;201000218264;20120017290;20110265198;20130137104;20130122591;20130177983和20130177960和美国申请号14/278,903。举例来说,美国专利公布号20110027235涉及将功能性蛋白质靶向整合至分离干细胞中并且美国公布号20120128635描述治疗乙型血友病的方法。另外,美国公布号2014-0017212和2014-0112896描述治疗溶酶体贮积疾病的方法。也参见Li等人(2011)Nature 475 (7355) :217-221和Anguela等人(2013)Blood 122:3283-3287。

[0015] 然而,仍然需要向患有一种或多种蛋白质缺少、缺陷和/或异常表达的疾病或病症的受试者提供治疗性蛋白质的额外组合物和方法。

[0016] 概述

[0017] 本文公开可用于在体内白蛋白启动子(例如,内源性或外源性白蛋白启动子)的控制下表达转基因的方法和组合物。在一些方面,转基因可编码所关注的治疗性蛋白质。转基因可编码蛋白质以使得本发明方法可用于产生缺陷或缺少的蛋白质(例如,“蛋白质替换”)。在一些情况下,蛋白质可涉及治疗溶酶体贮积病。可表达其他治疗性蛋白质,包括如大疱性表皮松解、糖尿病、癌症、凝血病症或AAT缺陷肺气肿不同的病状的蛋白质治疗剂。在其他方面,转基因可包含序列(例如,工程化序列)以使得表达的蛋白质具有给予其新的和所需特征的特性(增加的半衰期、改变的血浆清除特性等)。工程化序列也可包括从白蛋白序列得到的氨基酸。在一些方面,转基因编码治疗性蛋白质、治疗性激素、血浆蛋白质、抗体等。在一些方面,转基因可编码涉及血液病症诸如凝血病症的蛋白质。在一些方面,转基因编码结构性核酸(shRNA、RNAi、miRNA等)。

[0018] 在一方面,在本文中公开用于靶向整合编码功能性凝血因子蛋白质(例如,因子VII、因子VIII、因子IX和/或因子X)的序列的方法和组合物。功能因子VIII(“F8”)和/或因子IX(“F.IX”或“FIX”)蛋白质的表达可导致例如治疗和/或预防甲型血友病(F8)和/或乙型血友病(F.IX),而功能因子VII或因子X的表达可治疗或预防与因子VII和/或因子X缺陷相关联的血友病。

[0019] 在另一方面,本文公开用于靶向整合编码在患有溶酶体贮积病的受试者中缺少的功能蛋白质的序列的方法和组合物。核酸酶,例如工程化大范围核酸酶、锌指核酸酶(ZFN)、TALE-核酸酶(TALEN,包括TALE效应结构域与来自限制性内切酶和/或大范围核酸酶的核酸酶结构域的融合(诸如兆TALE和紧凑TALEN))、Ttago系统和/或CRISPR/Cas核酸酶系统用于在基因所插入的细胞中使‘安全港’基因座(例如CCR5、AAVS1、HPRT、Rosa或白蛋白)处的DNA裂解。供体转基因的靶向插入可经由同源性引导修复(HDR)或非同源性修复机制(例如,NHEJ供体捕获)。核酸酶可诱导靶DNA中的双链(DSB)或单链断裂(切口)。在一些实施方案中,两个切口酶用于通过引入两个切口来产生DSB。在一些情况下,切口酶是ZFN,而在其他

情况下,切口酶是TALEN或CRISPR/Cas切口酶。

[0020] 在一方面,本文描述的是非天然存在的锌指蛋白(ZFP),该锌指蛋白结合至基因组中所关注的区域(例如白蛋白基因)中的靶位点,其中ZFP包含一个或多个工程化的锌指结合结构域。在一个实施方案中,ZFP是裂解所关注靶基因组区域的锌指核酸酶(ZFN),其中ZFN包含一个或多个工程化的锌指结合结构域和核酸酶裂解结构域或裂解半结构域。裂解结构域和裂解半结构域可以例如从各种限制性内切核酸酶和/或归巢内切核酸酶中获得。在一个实施方案中,裂解半结构域从IIS型限制性内切核酸酶(例如,Fok I)中得到。在某些实施方案中,锌指结构域识别白蛋白基因中的靶位点,例如具有如在表5的单行中所示来排序的识别螺旋结构域的锌指蛋白。

[0021] 在另一方面,本文描述转录活化剂样效应(TALE)蛋白质,其结合至基因组中的所关注的区域(例如,白蛋白基因)中的靶位点,其中TALE包含一个或多个工程化TALE结合结构域。在一个实施方案中,TALE是裂解所关注靶基因组区域的核酸酶(TALEN),其中TALEN包含一个或多个工程化的TALE DNA结合结构域和核酸酶裂解结构域或裂解半结构域。裂解结构域和裂解半结构域可以例如从各种限制性内切核酸酶和/或归巢内切核酸酶(大范围核酸酶)中获得。在一个实施方案中,裂解半结构域从IIS型限制性内切核酸酶(例如,Fok I)中得到。在其他实施方案中,裂解结构域从大范围核酸酶得到,所述大范围核酸酶结构域还可展现DNA结合功能。

[0022] 在另一方面,本文描述结合至基因组中的所关注的区域(例如,白蛋白基因)中的靶位点的CRISPR/Cas系统,其中CRISPR/Cas系统包括一种或多种工程化单一引导RNA或功能等效物,以及Cas9核酸酶。

[0023] 如本文描述的核酸酶(例如,ZFN、CRISPR/Cas系统、Ttago和/或TALEN)可结合至且/或裂解在基因内或在基因附近的编码或非编码区域中的所关注的区域,例如像,前导序列、尾随序列或内含子,或在编码区域上游或下游的非转录区域。在某些实施方案中,核酸酶(例如,ZFN)结合至且/或裂解白蛋白基因。

[0024] 在另一方面,本文描述编码一种或多种核酸酶(例如,本文描述的ZFN、CRISPR/Cas系统、Ttago和/或TALEN)的多核苷酸。多核苷酸可以是例如mRNA。在一些方面,mRNA可化学修饰(参见例如Kormann等人(2011)Nature Biotechnology 29 (2):154-157)。

[0025] 在另一方面,本文描述的是ZFN、CRISPR/Cas系统、Ttago和/或TALEN表达载体,其包含有效地连接至启动子的编码如本文描述的一种或多种核酸酶(例如,ZFN、CRISPR/Cas系统、Ttago和/或TALEN)的多核苷酸。在一个实施方案中,表达载体是病毒载体。在一方面,病毒载体展现组织特异性趋向性。

[0026] 在另一方面,本文描述宿主细胞,其包含一种或多种核酸酶(例如,ZFN、CRISPR/Cas系统、Ttago和/或TALEN)表达载体。

[0027] 在另一方面,提供包含如本文描述的表达载体的药物组合物。在一些实施方案中,药物组合物可包含一种以上表达载体。在一些实施方案中,药物组合物包含包含第一多核苷酸的第一表达载体,和包含第二多核苷酸的第二表达载体。在一些实施方案中,第一多核苷酸和第二多核苷酸是不同的。在一些实施方案中,第一多核苷酸和第二多核苷酸是大致上相同的。药物组合物可进一步包含供体序列(例如,编码疾病或病症诸如LSD或血友病中缺少或缺陷的蛋白质的转基因)。在一些实施方案中,供体序列与表达载体相关联。

[0028] 在一些实施方案中,提供包含锌指蛋白和野生型或工程化裂解结构域或裂解半结构域的融合蛋白。

[0029] 在一些实施方案中,提供药物组合物,其包含:(i)编码锌指核酸酶的第一多核苷酸(例如,质粒、mRNA、Ad载体、AAV载体等),锌指核酸酶包含FokI裂解结构域和包含以F1至F5或F1至F6排序的5或6个锌指结构域的锌指蛋白,其中每个锌指结构域包含识别螺旋区域并且其中锌指蛋白的识别螺旋区域在表1、2或5的单行中示出;(ii)编码锌指核酸酶的第二多核苷酸(例如,质粒、mRNA、Ad载体、AAV载体等),所述锌指核酸酶包含FokI裂解结构域和包含以F1至F5或F1至F6排序的5或6个锌指结构域的锌指蛋白,其中每个锌指结构域包含识别螺旋区域并且其中锌指蛋白的识别螺旋区域在表1、2或5的单行中示出;和(iii)第三多核苷酸(例如,质粒、mRNA、Ad载体、AAV载体等)载体,其包含编码在疾病或病症(例如,LSD或血友病)中缺少或缺陷的蛋白质的供体。两个ZFN的ZFP可相同或不同。类似地,两个ZFN的裂解结构域可相同或不同(例如,可为形成专性异源二聚体的突变体)。在一些实施方案中,(i)、(ii)和(iii)以约1:1:1、约1:1:2、约1:1:3、约1:1:4、约1:1:5、约1:1:6、约1:1:7、约1:1:8、约1:1:9、约1:1:10、约1:1:11、约1:1:12、约1:1:13、约1:1:14、约1:1:15、约1:1:16、约1:1:17、约1:1:18、约1:1:19或约1:1:20比率提供。

[0030] 在一方面,本发明方法和组合物包括遗传修饰细胞,所述细胞包括表达在血友病中异常表达的功能形式蛋白质(因子VII、F8、F.IX和/或因子X蛋白质)的转基因,其中转基因整合至细胞基因组的内源性安全港基因(例如,白蛋白基因)中。在另一方面,本发明方法和组合物包括遗传修饰细胞,所述细胞包括在患有溶酶体贮积病的受试者中表达缺少或异常表达的功能形式蛋白质的转基因。在某些实施方案中,转基因使用至少一种核酸酶以位点特异性(靶向)方式来整合。在某些实施方案中,核酸酶(例如,ZFN、TALEN、Ttago和/或CRISPR/Cas系统)对于安全港基因(例如CCR5、HPRT、AAVS1、Rosa或白蛋白具有特异性。参见例如美国专利号7,951,925和8,110,379;美国公布号20080159996;201000218264;20120017290;20110265198;20130137104;20130122591;20130177983和20130177960和美国申请号14/278,903)。在一些实施方案中,安全港是白蛋白基因。

[0031] 在另一方面,本文描述在体外和/或体内遗传修饰细胞以产生治疗性蛋白质(例如,在疾病或病症诸如血友病(因子VII、F8、F.IX和/或因子X)或溶酶体贮积病(IDS、IDUA等)中缺少的蛋白质的方法,所述方法包括使用一种或多种核酸酶(例如,ZFN、TALEN、CRISPR/Cas)来裂解细胞中的内源性安全港基因以使得编码治疗性蛋白质的转基因整合至安全港基因座中并且在细胞中表达。在某些实施方案中,安全港基因是aCCR5、HPRT、AAVS1、Rosa或白蛋白基因。在另一个方面,本文描述在体外和/或体内遗传修饰细胞以产生在溶酶体贮积病中缺少的蛋白质的方法。这些疾病的最常见实例是葡萄糖脑苷脂酶缺陷(基因名称:GBA),其与戈谢氏疾病相关联; $\alpha$ 半乳糖苷酶缺陷(基因名称:GLA),其与法布里氏病相关联;艾杜糖酸酯-2-硫酸酯酶缺陷(基因名称:IDS),其与亨特氏疾病相关联; $\alpha$ -L艾杜糖苷酸酶缺陷(基因名称:IDUA),其与赫尔勒氏疾病相关联;和鞘磷脂磷酸二酯酶1缺陷(基因名称:SMPD1),其与尼曼匹克氏疾病相关联。在某些实施方案中,细胞是哺乳动物细胞。在某些实施方案中,所述细胞为灵长类动物细胞。在某些实施方案中,细胞是人细胞。在一组实施方案中,提供裂解细胞(例如,肝细胞)中的白蛋白基因的方法,包括将本文公开的一种或多种表达载体在使得表达一种或多种蛋白质并且裂解白蛋白基因的条件下引入细胞中。白蛋白

基因可例如通过将供体序列整合至所裂解的白蛋白基因中来修饰。在某些实施方案中，所述方法包括遗传修饰细胞以产生凝血因子或在溶酶体贮积病中缺少的蛋白质，所述方法包括向细胞施用表5示出的锌指核酸酶 (ZFN) (或编码这些ZFN的多核苷酸) 和供体。ZFN和供体可以任何组合存在于相同或不同载体上，例如在各自携带一种组分的3个独立载体 (例如，AAV载体) 上；在携带两种组分的一个载体和携带第3种组分的独立载体上；或在携带所有3种组分的一个载体上。

[0032] 在其他方面，本发明包括将供体核酸递送至靶细胞。供体可在编码核酸酶的核酸之前、之后递送或与其一起递送。供体核酸可包含待整合至细胞的基因组，例如，内源性基因座中的外源性序列(转基因)。在一些实施方案中，供体可包含由具有靶向裂解位点的同源区域侧接的全长基因或其片段。在一些实施方案中，供体缺少同源区域并且经由与同源性无关的机制(即NHEJ)来整合至靶基因座中。供体可包含任何核酸序列，例如核酸在用作核酸酶诱导双链断裂的同源介导修复的底物时导致在内源性染色体基因座产生供体指定缺失或，替代地(或除此以外)导致建立内源性基因座的新等位基因形式(例如，清除转录因子结合位点的点突变)。在一些方面，供体核酸是寡核苷酸，其中整合导致基因校正事件，或靶向缺失。在一些方面，供体包含治疗性蛋白质，例如凝血因子。

[0033] 在一些实施方案中，编码DNA结合蛋白的多核苷酸是mRNA。在一些方面，mRNA可经化学修饰(参见例如Kormann等人(2011) Nature Biotechnology 29 (2) :154-157)。在其他方面，mRNA可包含ARCA帽(参见美国专利7,074,596和8,153,773)。在其他实施方案中，mRNA可包含未修饰和修饰核苷酸的混合物(参见美国专利公布2012-0195936)。

[0034] 在另一方面，本文提供用于在哺乳动物，或灵长类动物，诸如人灵长类动物，诸如患有LSD和/或血友病的人患者中提供缺少或缺陷的一种或多种功能性蛋白质的方法，例如用于通过供应在受试者中缺少或缺陷的蛋白质来治疗疾病。在另一方面，本文提供用于提供在哺乳动物，或灵长类动物，诸如人灵长类动物，诸如患有乙型血友病的人患者中缺少或缺陷的功能性蛋白质(例如，F.IX)的方法，例如用于治疗乙型血友病。在另一方面，本文提供用于向哺乳动物，或灵长类动物，诸如人灵长类动物，诸如人患者提供功能性蛋白质(例如因子VII)以治疗与因子VII缺陷相关联的血友病的方法。在另一方面，本文提供用于提供功能性蛋白质(例如因子X)以治疗与因子X缺陷相关联的血友病的方法。在某些实施方案中，所述方法包括使用核酸酶以将编码功能性因子VII、F8、F.IX和/或因子X蛋白质的序列整合至有需要的受试者的细胞中。在其他实施方案中，所述方法包括使用核酸酶来整合编码在溶酶体贮积病中缺少或缺陷的功能性蛋白质的序列。在其他实施方案中，所述方法包括将遗传修饰细胞(表达在患有血友病的受试者中异常表达的功能形式蛋白质)施用至受试者。因此，分离的细胞可引入受试者中(离体细胞疗法)或细胞可在它是受试者的一部分(在体内)时加以修饰。还提供使用本文描述的供体和/或核酸酶用于治疗血友病(例如，用因子VIII供体治疗甲型血友病，用因子IX供体治疗乙型血友病，用因子VII治疗因子VII缺陷，用因子X治疗因子X缺陷，用GBA供体治疗戈谢氏病，用GLA供体治疗法布里氏病，用IDS供体治疗亨特氏病，用IDUA供体治疗赫尔勒氏病，和/或用SMPD1供体治疗尼曼匹克氏病)，例如，在制备用于治疗疾病的药物中。在某些实施方案中，F8蛋白质包括B-结构域缺失。在某些实施方案中，F8-和/或F.IX-编码序列使用病毒载体、非病毒载体(例如，质粒)和/或其组合来递送。

[0035] 在所描述的任何组合物和方法中,核酸酶和/或转基因可在AAV载体上携带,包括但不限于AAV1、AAV3、AAV4、AAV5、AAV6、AAV8、AAV9和AAVrh10或假型AAV诸如AAV2/8、AAV8.2、AAV2/5和AAV2/6等。在某些实施方案中,核酸酶和转基因供体使用相同AAV载体类型来递送。在其他实施方案中,核酸酶和转基因供体使用不同AAV载体类型来递送。核酸酶和转基因可使用一个或多个载体来递送,例如,一个载体携带转基因和核酸酶;两个载体,其中一个携带核酸酶(例如,例如具有2A肽的ZFN对的左和右ZFN)并且一个携带转基因;或三个载体,其中一个载体携带核酸酶对的一个核酸酶(例如,左ZFN),单独载体携带核酸酶对中的另一个核酸酶(例如,右ZFN)并且第三个单独载体携带转基因。参见图2。在使用两个或更多个载体的实施方案中,载体可以相同浓度或以不同比率来使用,例如,转基因供体载体可用比核酸酶载体高2倍、3倍、4倍、5倍或更高的浓度来施用。在某些实施方案中,核酸酶和/或转基因供体经由静脉内(例如,门静脉内)施用来递送至完整动物的肝脏。

[0036] 在本文描述的任何组合物和方法中,由转基因编码的蛋白质可包括F8蛋白质,例如B-域缺失因子VIII(BDD-F8)。在其他实施方案中,由转基因编码的蛋白质包括F.IX蛋白质。在其他实施方案中,由转基因编码的蛋白质包括因子VII蛋白质。在其他实施方案中,由转基因编码的蛋白质包括因子X蛋白质。在一些实施方案中,由转基因编码的蛋白质包括葡糖脑苷脂酶。在其他实施方案中,由转基因编码的蛋白质包括 $\alpha$ 半乳糖苷酶。在其他实施方案中,由转基因编码的蛋白质包括艾杜糖酸酯-2-硫酸酯酶。在一些实施方案中,由转基因编码的蛋白质包括鞘磷脂磷酸二酯酶。在本文描述的任何组合物或方法中,转基因还包括转录调控因子,而在其他组合物或方法中,它不包括转录调控因子并且转录通过内源性调控因子来调控。在另一方面,本发明方法包括用于治疗性治疗有需要的受试者的组合物。在一些实施方案中,所述组合物包含工程化干细胞,其包括安全港特异性核酸酶,和编码因子VII、F8、F.IX、因子X、GBA、GLA、IDS、IDUA和/或SMPD1蛋白质或其功能片段和/或截断形式的转基因供体。在其他实施方案中,所述组合物包含工程化干细胞,其已经修饰并且表达编码因子VII、F8、F.IX、因子X、GBA、GLA、IDS、IDUA和/或SMPD1蛋白质或其功能片段和/或截断形式的转基因供体。

[0037] 在本文描述的任何组合物或方法中,细胞可为真核细胞。合适细胞的非限制实例包括真核细胞或细胞系诸如分泌细胞(例如,肝脏细胞、粘膜细胞、唾液腺细胞、垂体细胞等)、血细胞(红细胞)、先驱红细胞、肝细胞、肌肉细胞、干细胞(例如胚胎干细胞、诱导多能干细胞、肝干细胞、造血干细胞(例如CD34+))或内皮细胞(例如血管、肾小球和管状内皮细胞)。因此,靶细胞可为灵长类动物细胞,例如人细胞,或靶细胞可为哺乳动物细胞(包括兽医动物),例如尤其非人灵长类动物和以下目的哺乳动物:啮齿目(小鼠、大鼠、仓鼠)、兔形目(兔)、食肉目(猫、犬)和偶蹄目(奶牛、猪、绵羊、山羊、马)。在一些方面,靶细胞包括组织(例如肝脏)。在一些方面,靶细胞是干细胞(例如,胚胎干细胞、诱导多能干细胞、肝干细胞等)或通过本文描述的任何方法的动物胚胎,然后植入胚胎以使得活动物出生。然后将动物培育至性成熟并且允许产生后代,其中至少一些后代包括基因组修饰。细胞也可包括胚胎细胞,例如小鼠、大鼠、兔或其他哺乳动物细胞胚胎。细胞可来自任何生物体,例如人、非人灵长类动物、小鼠、大鼠、兔、猫、犬或其他哺乳动物细胞。细胞可分离或可为生物体(例如,受试者)的一部分。

[0038] 在本文所述的任何方法或组合物中,含有工程化基因座(例如,白蛋白基因座)的

细胞可为可适用于治疗性用途的干细胞。可用于本发明方法和组合物的具体干细胞类型包括胚胎干细胞(ESC)、诱导多能干细胞(iPSC)和肝或肝脏干细胞。iPSC可从患者样品和正常对照得到，其中来源于患者的iPSC可突变至所关注的基因的正常基因序列，或正常细胞可改变成所关注的基因的已知疾病等位基因。类似地，肝干细胞可从患者分离。然后，将这些细胞工程化以表达所关注的转基因、扩增、然后重新引入患者体内。

[0039] 在本文所述的任何方法和组合物中，转基因可整合至内源性安全港基因中以使得一些、全部或无内源性基因表达，例如具有整合转基因的融合蛋白。在一些实施方案中，内源性安全港基因是白蛋白基因并且内源性序列是白蛋白序列。内源性可存在于外源性蛋白质的氨基(N)-末端部分和/或外源性蛋白质的羧基(C)-末端部分上。白蛋白序列可包括全长野生型或突变型白蛋白序列或，替代地，可包括部分白蛋白氨基酸序列。在某些实施方案中，白蛋白序列(全长或部分)用于增加与其融合且/或充当载体的转基因所表达的多肽的血清半衰期。在其他实施方案中，转基因包含白蛋白序列并且靶向插入基因组内的另一个安全港。此外，转基因可包括驱动其表达的外源性启动子(例如，组成型或可诱导启动子)或其表达可由内源性控制序列(例如，内源性白蛋白启动子)驱动。在一些实施方案中，供体包括额外修饰，包括但不限于密码子优化、添加糖基化位点等。

[0040] 在本文描述的任何组合物或方法中，裂解可经由使用特异性核酸酶诸如工程化锌指核酸酶(ZFN)、转录活化剂样效应核酸酶(TALEN)来发生，或使用具有工程化crRNA/tracr RNA(‘单一引导RNA’)的Ttago或CRISPR/Cas系统来引导特异性裂解。在一些实施方案中，使用两个切口酶通过引入两个切口来产生DSB。在一些情况下，切口酶是ZFN，而在其他情况下，切口酶是TALEN或CRISPR/Cas系统。靶向整合可经由同源性引导修复机制(HDR)且/或经由非同源性修复机制(例如，NHEJ供体捕获)来发生。如本文描述的核酸酶可结合至且/或裂解在基因内或在基因附近的编码或非编码区域中的所关注的区域，例如像，前导序列、尾随序列或内含子，或在编码区域上游或下游的非转录区域内。在某些实施方案中，核酸酶裂解在结合位点(例如，表5所示结合位点)处或附近的靶序列。裂解可导致例如经由插入、缺失或其组合来修饰基因。在某些实施方案中，修饰在核酸酶结合和/或裂解位点处或附近，例如，裂解位点上游或下游1-300个(或在其间的任何值)碱基对内，更优选地结合和/或裂解位点任一侧的1-100个碱基对(或在其间的任何值)内，甚至更优选地结合和/或裂解位点任一侧的1至50个碱基对(或在其间的任何值)内。

[0041] 所描述的方法和组合物可用于治疗或预防有需要的受试者的血友病。在一些实施方案中，组合物包含载体并且用于靶向肝脏细胞。在其他实施方案中，组合物包含工程化干细胞并且作为骨髓移植植物来给予患者。在一些情况下，患者在移植之前部分或完全免疫清除。在其他情况下，在内源性基因的核酸酶介导的修饰(例如，将转基因靶向整合至白蛋白基因座中)之前、期间和/或之后，患者用一种或多种免疫抑制剂治疗。此外，本文描述的任何方法可进一步包括额外步骤，包括部分肝切除术或用增强转导且/或诱导肝细胞经历细胞周期的辅助剂治疗。辅助剂的实例包括 $\gamma$ 辐照、UV辐照、氯化核苷酸诸如胸昔、顺铂、依托泊苷、羟基脲、阿非迪霉素、泼尼松龙、四氯化碳和/或腺病毒。

[0042] 本文所述方法可以在体外、离体或体内实施。在某些实施方案中，将组合物引入活的完整哺乳动物中。在递送时，哺乳动物可在任何发育阶段，例如，胚胎、胎儿、新生儿、婴儿、幼年或成年。另外，靶向细胞可为健康或患病细胞。在某些实施方案中，将组合物中的一

种或多种静脉内递送(例如,经由门静脉内到达肝脏,例如尾静脉注射)、动脉内、腹膜内、肌肉内、进入肝实质(例如,经由注射)、进入肝动脉(例如,经由注射),和/或经由胆道系统(例如,经由注射)。

[0043] 在一个具体方面,本文所述的方法和组合物包括治疗性组合物,其包含(i)编码FVIII的供体转基因(ii)核酸酶(例如,ZFN、TALEN、Ttago或CRISPR/Cas系统),其靶向相应地除了FVIII以外的内源性基因的基因座,例如,靶向哺乳动物,或灵长类动物或人,诸如血友病患者的内源性白蛋白基因。在某些实施方案中,治疗性组合物包含在分开的独立载体诸如分开AAV载体中的不同量的FVIII供体转基因和白蛋白基因特异性核酸酶,其可一起施用(例如,混合成单一溶液或同时施用)或替代地可个别地施用(例如,在单独溶液中以相应施用之间的实质性延迟,例如,10分钟或更长、30分钟或更长、1小时或更长、2小时或更长、3小时或更长或更长时间来施用)。在某些实施方案中,施用治疗性组合物以将FVIII供体转基因整合至非FVIII基因座中并且随后表达所整合的FVIII以在哺乳动物,或灵长类动物或人,或血友病患者的血浆中获得治疗水平的FVIII。在某些实施方案中,治疗水平的FVIII可包括,例如,大于2%、大于4%、大于5%、大于6%、大于8%、大于10%、大于12%、大于15%、大于20%、大于25%或更大临床可接受正常血浆浓度的FVIII。替代地或另外地,治疗水平的FVIII可包括在将FVIII供体转基因和白蛋白基因核酸酶施用至此个体之前在个别哺乳动物、或灵长类动物或人、或血友病患者中测量的例如大于2%、大于4%、大于5%、大于6%、大于8%、大于10%、大于12%、大于15%、大于20%、大于25%、大于30%、大于35%或更大血浆浓度的功能性FVIII。

[0044] 在一个具体方面,本文所述的方法和组合物包括治疗性组合物,其包含(i)编码在溶酶体贮积蛋白质中缺陷的蛋白质的供体转基因(ii)核酸酶(例如,ZFN、TALEN、Ttago或CRISPR/Cas系统),其靶向相应地除了在溶酶体贮积病中缺陷的蛋白质的基因以外的内源性基因的基因座,例如,靶向哺乳动物,或灵长类动物或人,诸如患有溶酶体贮积病的受试者的内源性白蛋白基因。在某些实施方案中,治疗性组合物包含在分开的独立载体诸如分开AAV载体中的不同量的选自GBA、GLA、IDS、IDUA和/或SMPD1的供体转基因和白蛋白基因特异性核酸酶,其可一起施用(例如,混合成单一溶液或同时施用)或替代地可个别地施用(例如,在单独溶液中以相应施用之间的实质性延迟,例如,10分钟或更长、30分钟或更长、1小时或更长、2小时或更长、3小时或更长或更长时间来施用)。在某些实施方案中,施用治疗性组合物以使GBA、GLA、IDS、IDUA和/或SMPD1供体转基因整合至相应地除了编码GBA、GLA、IDS、IDUA和/或SMPD1以外的基因座,并且随后表达所整合的GBA、GLA、IDS、IDUA和/或SMPD1以在哺乳动物,或灵长类动物或人,或血友病患者的血浆中获得治疗水平的GBA、GLA、IDS、IDUA和/或SMPD1。在某些实施方案中,治疗水平的GBA、GLA、IDS、IDUA和/或SMPD1可包括例如大于2%、大于4%、大于5%、大于6%、大于8%、大于10%、大于12%、大于15%、大于20%、大于25%或更大临床可接受正常血浆浓度的GBA、GLA、IDS、IDUA和/或SMPD1。替代地或另外地,治疗水平的GBA、GLA、IDS、IDUA和/或SMPD1可包括在将GBA、GLA、IDS、IDUA和/或SMPD1供体转基因和白蛋白基因核酸酶施用至此个体之前在个别哺乳动物、或灵长类动物或人、或患者中测量的例如大于2%、大于4%、大于5%、大于6%、大于8%、大于10%、大于12%、大于15%、大于20%、大于25%、大于30%、大于35%或更大血浆浓度的功能性GBA、GLA、IDS、IDUA和/或SMPD1。

[0045] 在另一个具体方面,本文所述的方法和组合物包括治疗性组合物,其包含(i)编码FIX的供体转基因(ii)核酸酶((例如,ZFN、TALEN、Ttago或CRISPR/Cas系统),其靶向相应地除了FIX以外的内源性基因的基因座,例如,靶向哺乳动物,或灵长类动物或人,诸如血友病患者的内源性白蛋白基因。在某些实施方案中,治疗性组合物包含在分开的独立载体诸如分离AAV载体中的不同量的FIX供体转基因和白蛋白基因特异性核酸酶,其可一起施用(例如,混合成单一溶液或同时施用)或替代地可个别地施用(例如,在单独溶液中以相应施用之间的实质性延迟,例如,10分钟或更长、30分钟或更长、1小时或更长、2小时或更长、3小时或更长或更长时间来施用)。在某些实施方案中,施用治疗性组合物以将FIX供体转基因整合至非FIX基因座中并且随后表达所整合的FIX以在哺乳动物,或灵长类动物或人,或血友病患者的血浆中获得治疗水平的FIX。在某些实施方案中,治疗水平的FIX可包括,例如,大于2%、大于4%、大于5%、大于6%、大于8%、大于10%、大于12%、大于15%、大于20%、大于25%或更大临床可接受正常血浆浓度的FIX。替代地或另外地,治疗水平的FIX可包括在将FIX供体转基因和白蛋白基因核酸酶施用至此个体之前在个别哺乳动物、或灵长类动物或人、或血友病患者中测量的例如大于2%、大于4%、大于5%、大于6%、大于8%、大于10%、大于12%、大于15%、大于20%、大于25%、大于30%、大于35%或更大血浆浓度的功能性FIX。

[0046] 为了将组合物靶向至具体类型细胞,例如,血小板、成纤维细胞、肝细胞等,施用的组合物中的一种或多种可与特异性结合至细胞表面受体的归巢剂相关联。举例来说,载体可偶联至配体(例如,半乳糖),某些肝系统细胞具有针对所述配体的受体。偶联可为共价的,例如,交联剂诸如戊二醛,或非共价的,例如,抗生素素化配体结合至生物素化载体。另一种形式的共价偶联通过将用于制备载体原料的AAV辅助质粒工程化来提供以使得所编码的外壳蛋白中的一种或多种是原生AAV外壳蛋白和肽或蛋白质配体的混合物,从而使配体暴露于病毒颗粒的表面上。

[0047] 还提供试剂盒,其包括本发明的组合物(例如,遗传修饰细胞、ZFP、CRISPR/Cas系统和/或TALEN和任选地转基因供体)。试剂盒可以包括编码核酸酶的核酸(例如包含在适合表达载体中RNA分子或核酸酶编码基因)、供体分子、适合的宿主细胞系、用于实施本发明方法的说明书等。

[0048] 这些和其他方面鉴于总体公开内容易于为本领域技术人员显而易知。

[0049] 附图简述

[0050] 图1是描绘锌指核酸酶介导的将编码蛋白质(例如,治疗性蛋白质)的转基因插入内源性白蛋白基因座处的示意图。“SA”是指剪接受体部点;“pA”是指多聚腺苷酸化信号;并且“Alb Ex1”是指内源性白蛋白基因座的外显子1。

[0051] 图2A至2C是描绘F8转基因供体的示例性核酸酶和供体设计的示意图。图2A示出示例性供体设计并且图2B示出另一个示例性设计(“优化供体”),其包括密码子优化、使用不同多聚腺苷酸化信号和/或添加假定糖基化基序(“V3肽”)。图2A和2B描绘的供体缺少启动子/增强子区域并且大约为4.4至4.7kb大小,适合封装至AAV中。图2C是描绘用于安全港裂解以利于F8和/或F9转基因的靶向整合的一对ZFN中的每个ZFN的分开载体的设计的示意图。

[0052] 图3A至3D通过展示使用AAV2/8载体来施用ZFN和供体之后的HA/CD4-/-小鼠的血

浆中的hFVIII活性来描绘锌指核酸酶介导的将编码人F8蛋白质的(hFVIII)序列的靶向整合至内源性白蛋白基因座中的结果。图3A是描绘AAV2/8转基因供体的示意图。图3B和3C是描绘在注射ZFN载体和转基因供体(“ZFN+供体”)或空载体(没有ZFN序列)和转基因供体(“模拟+供体”)的HA/CD4<sup>-/-</sup>小鼠中作为正常水平的百分比的hFVIII血浆水平的图。示出施用后2周(图3B)或施用后2和8周(图3C)的结果。示出载体和施用量。图3D是展示施用编码左和右ZFN的单一载体(“2A融合”)或如图2C所示各自编码一对ZFN中的一个ZFN的分开载体(“个别ZFN”)的小鼠中的白蛋白基因修饰水平的图。所使用的载体是AAV2/8并且剂量沿着水平轴(病毒基因组(“VG”)/小鼠)作图。

[0053] 图4是描绘注射白蛋白靶向ZFN和图2B示出的优化(V3)供体构建体的HA/CD4<sup>-/-</sup>小鼠中的作为正常水平的百分比的hFVIII的血浆水平的图。使用AAV2/8-ZFN( $5 \times 10^{10}$  vg的每个ZFN)+AAV2/8供体( $1 \times 10^{11}$  vg/小鼠)。

[0054] 图5A和5B描绘F.IX转基因的核酸酶介导整合至白蛋白基因座中。图5A是所使用的供体转基因F.IX构建体的示意图。供体包含与人F9基因座同源的供体臂(“人臂左”和“人臂右”),因而预期其在此实验中不促进HDR。因此,供体的插入依赖于经由末端捕获的NHEJ。图5B示出将靶向内源性小鼠白蛋白基因座(“mAlb ZFN”)或人因子IX基因座(“hF9ZFN”)的ZFN和hF9供体转基因(“供体”)施用至野生型小鼠之后的循环hFIX水平。AAV载体和施用量如下:对于mAlb ZFN和F.IX供体来说,AAV2/8-ZFN是 $1 \times 10^{11}$  vg/小鼠并且AAV2/8-供体是 $5 \times 10^{11}$  vg/小鼠,并且对于hF9ZFN和F.IX供体来说,AAV2/8-ZFN是 $1 \times 10^{11}$  vg/小鼠并且AAV2/8-供体是 $5 \times 10^{11}$  vg/小鼠。注意hF9ZFN不裂解内源性小鼠F9基因座。

[0055] 图6是描绘将白蛋白靶向ZFN和hF.IX供体施用至小鼠之后的hF.IX水平的图。AAV2/8-ZFN以指示剂量施用并且AAV2/8-供体是ZFN的5x剂量。基因组编辑在三个数量级下与AAV剂量成比例。

[0056] 图7A和7B是描绘用ZFN和hF.IX供体处理的HB小鼠中的凝血时间的图。图7A示出指示动物中的血浆hF.IX水平并且图7B示出活化部分凝血活酶时间(aPTT)。AAV剂量在底部示出。

[0057] 图8A至8D是展示在只接受白蛋白靶向ZFN的非人灵长类动物中施用ZFN后第65天的ELISPOT数据的图。“LN”是指淋巴结。如示出,没有针对AAV8衣壳或ZFN的免疫响应。

[0058] 图9是示出原代人肝细胞中的白蛋白基因座的核酸酶介导靶向的图。人原代肝细胞在体外用AAV2/6hF9供体(MOI  $9 \times 10^5$  vg/细胞)转导并且24小时后用500ng的hALB ZFN mRNA转导。底部图:通过MiSeq分析测量的插入和缺失%。第5天(每个组的最左侧棒)、第7天(每个组的中间棒)和第9天(每个组的最右侧棒)获得的上清液针对hFIX蛋白质水平通过ELISA来分析。误差棒=s.e.m.数据代表至少2个独立实验。

[0059] 图10A至10D示出mAlb ZFN在体外和体内的表征。图10A示出用指示量的ZFN或GFP mRNA转染的Hepa 1-6细胞中的通过插入和缺失测量的ZFN活性。将基因组DNA分离并且将靶序列PCR扩增供Illumina MiSeq测序。百分比指示与裂解和NHEJ修复一致的含有插入和/或缺失的读取。图10B示出在IV注射 $5 \times 10^{11}$  vg AAV8-hF9-供体和 $1 \times 10^{11}$  vg AAV8-mALB-ZFN之后,治疗小鼠中的hF.IX水平保持稳定超过一年。图10C示出治疗之后的血浆ALT值未偏离正常范围(阴影面积)。图10D描绘使用定量PCR来确定“混合F9-mAlb”相比于野生型白蛋白mRNA的相对丰度。将小鼠注射1:1比率的指示剂量的ZFN:供体。注射后2周,总RNA从小鼠的

肝脏分离。作为阴性对照,荧光素酶(模拟物)+供体在各自 $5\times10^{11}$ vg的较高剂量下给予。2-加尾曼-惠特尼测试用于比较2个组。n=6-8小鼠/组。误差棒=s.e.m.\*\*P<0.01相比于模拟物。

[0060] 图11是展示核酸酶介导插入F.IX转基因的非人灵长类动物(恒河猴)研究的研究设计的示意图。“ALB ZFN”是指白蛋白靶向ZFN,如美国公布号20130177983所描述。“FokI-eHi/fi”是指形成如在ZFN中所使用的专性异源二聚体的工程化FokI裂解结构域。参见,例如,美国专利号7,914,796,8,034,598;美国公布号20110201055和20120142062。AAV载体和血清型如所示以高剂量(单一ZFN:1.5e11vg/kg;供体:1.5e14vg/kg)或低剂量(单一ZFN:5e12vg/kg;供体:5e13vg/kg)来使用。

[0061] 图12A和12B示出分别用于靶向小鼠白蛋白基因座的内含子1的示例性ZFN SBS#30724和SBS#30725的氨基酸序列。图12A是在研究中所使用的ZFN<sub>左</sub>(SEQ ID NO:115)氨基酸序列并且图12B是ZFN<sub>右</sub>(SEQ ID NO:116)氨基酸序列。3xFLAG标签以斜体字来标注。SV40大T抗原核定位序列通过下划线来标注。FokI结构域以粗字体来标注。识别螺旋区域加双下划线和斜体字来示出。

[0062] 图13A至13D示出在用白蛋白特异性ZFN和供体修饰之后的细胞上清液中的在溶酶体贮积疾病中缺陷的蛋白质的表达和酶活性的存在。图13A示出经由蛋白质印迹分析(图13B)的α-L艾杜糖昔酸酶(IDUA)蛋白质的酶表达和细胞上清液中的IDUA酶活性的存在。培养物在转染和低或高剂量的ZFN和IDUA供体之后第3和6天取样。图13C和13D示出测量在ZFN和IDS供体转染之后的艾杜糖酸酯-2-硫酸酯酶缺陷(IDS)蛋白质的存在和酶活性的一组类似数据。

### [0063] 详述

[0064] 本文公开用于修饰细胞以产生一种或多种蛋白质的组合物和方法,所述蛋白质的表达或基因序列在修饰之前是异常的并且与疾病或病症,例如,血友病或溶酶体贮积病(LSD)相关联。细胞通过将编码一种或多种功能蛋白质的转基因靶向插入细胞的安全港基因(例如,白蛋白)来修饰。在一些实施方案中,转基因插入内源性白蛋白基因中。转基因可编码涉及血友病的任何蛋白质或肽,例如因子VII、F8、F.IX、因子X、GBA、GLA、IDS、IDUA、SMPD1和/或其功能片段。还公开使用如本文描述的细胞和/或通过如本文描述来修饰细胞(离体或体内)来治疗一种或多种蛋白质缺少或缺陷的病症(例如,血友病或溶酶体贮积病)的方法。进一步描述用于修饰细胞的包含编码核酸酶的核酸和供体分子的组合物,和在体内或离体修饰细胞的方法。另外,描述包含通过本发明方法和组合物修饰的细胞的组合物。

[0065] 本文描述的基因组修饰的细胞通常经由核酸酶介导(ZFN、TALEN和/或CRISPR/Cas)靶向整合来修饰以插入编码治疗性蛋白质(例如,因子VII、因子VIII(F8)、因子IX、因子X、葡糖脑苷脂酶、α半乳糖昔酶、艾杜糖酸酯-2-硫酸酯酶(IDS)、α-L艾杜糖昔酸酶(IDUA)和/或鞘磷脂磷酸二酯酶1)的序列,其中与疾病相关的其基因发生改变或处于异常状态下的蛋白质进入受试者的一种或多种细胞的基因组(在体内或离体),以使得细胞在体内产生蛋白质。在某些实施方案中,所述方法进一步包括诱导受试者的细胞,尤其肝脏细胞增殖(进入细胞周期),例如通过部分肝切除术和/或通过施用诱导肝细胞经历细胞周期的一种或多种化合物。受试者包括但不限于人,非人灵长类动物,兽医动物诸如猫、犬、兔、大鼠、小鼠、豚鼠、奶牛、猪、马、山羊等。

[0066] 总则

[0067] 除非另外指示,否则这些方法的实施以及本文公开的组合物的制备和用途将使用分子生物学、生物化学、染色质结构和分析、计算化学、细胞培养、重组DNA以及本领域技术内的相关领域中的传统技术。这些技术在文献中得以充分说明。参见例如Sambrook等人MOLECULAR CLONING:A LABORATORY MANUAL,第二版,Cold Spring Harbor Laboratory Press,1989和2001第三版;Ausubel等人,CURRENT PROTOCOLS IN MOLECULAR BIOLOGY,John Wiley&Sons,New York,1987和定期更新;丛书METHODS IN ENZYMOLOGY,Academic Press,San Diego;Wolffe,CHROMATIN STRUCTURE AND FUNCTION,第三版,Academic Press,San Diego,1998;METHODS IN ENZYMOLOGY,Vol.304,“Chromatin”(P.M.Wasserman和A.P.Wolffe编),Academic Press,San Diego,1999;和METHODS IN MOLECULAR BIOLOGY,Vol.119,“Chromatin Protocols”(P.B.Becker,ed.)Humana Press,Totowa,1999。

[0068] 定义

[0069] 术语“核酸”、“多核苷酸”以及“寡核苷酸”可交换使用,并且指的是呈直链或环状构象,以及呈单链或双链形式的脱氧核糖核苷酸或核糖核苷酸聚合物。出于本公开的目的,这些术语不应被解释为相对于聚合物的长度为限制性的。所述术语可涵盖天然核苷酸的已知类似物,以及在碱基、糖和/或磷酸部分(例如,硫代磷酸主链)中被修饰的核苷酸。通常,具体核苷酸的类似物具有相同的碱基成对特异性;即A的类似物将会与T碱基成对。

[0070] 术语“多肽”、“肽”以及“蛋白质”可交换使用,指的是氨基酸残基的聚合物。该术语也适用于其中一种或多种氨基酸是对应的天然存在氨基酸的化学类似物或修饰衍生物的氨基酸聚合物。

[0071] “结合”指的是大分子之间(例如,蛋白质与核酸之间)序列特异、非共价相互作用。不是所有的结合相互作用的组分都需要序列特异(例如,与DNA主链中磷酸盐残基接触),只要相互作用总体上是序列特异即可。这种相互作用通常特征在于 $10^{-6}M^{-1}$ 或更低的分解常数( $K_d$ )。“亲和力”指的是结合的强度:增加的结合亲和力与较低 $K_d$ 有关。

[0072] “结合蛋白”是能够非共价结合另一分子的蛋白质。结合蛋白可以结合至例如DNA分子(DNA结合蛋白)、RNA分子(RNA结合蛋白)和/或蛋白质分子(蛋白质结合蛋白)。就蛋白质结合蛋白来说,它可以结合至其自身(以形成同源二聚体、同源三聚体等)和/或它可以结合至不同蛋白质的一个或多个分子。结合蛋白可以具有多于一种类型的结合活性。举例来说,锌指蛋白具有DNA结合、RNA结合以及蛋白质结合活性。

[0073] “锌指DNA结合蛋白”(或结合结构域)是通过一个或多个锌指以序列特异的方式结合DNA的蛋白质,或较大蛋白质内的一个结构域,所述一个或多个锌指是结合结构域内的氨基酸序列区,该序列区的结构通过锌离子配位作用来稳定。术语锌指DNA结合蛋白经常缩写为锌指蛋白或ZFP。

[0074] “TALE DNA结合结构域”或“TALE”是包含一个或多个TALE重复结构域/单元的多肽。重复结构域涉及TALE结合至其同源靶DNA序列。单一“重复单元”(也称为“重复序列”)通常是33-35氨基酸长度并且展现至少一些与天然存在TALE蛋白质内的其他TALE重复序列的序列同源性。参见例如美国专利号8,586,526。

[0075] 锌指和TALE结合结构域可“工程化”以结合至预定核苷酸序列,例如经由天然存在锌指或TALE蛋白质的识别螺旋区域的工程化(改变一个或多个氨基酸)。因此,工程化DNA结

合蛋白(锌指或TALE)是非天然存在的蛋白质。用于将DNA结合蛋白工程化的方法的非限制性实例是设计和选择。经设计的DNA结合蛋白是自然界中不存在的蛋白质,该蛋白质的设计/组成主要是合理标准的结果。用于设计的合理标准包括取代规则以及用于处理数据库信息的计算机算法的应用,该数据库存储了现有ZFP和/或TALE设计和结合数据的信息。参见例如美国专利6,140,081;6,453,242;6,534,261;和8,586,526也参见WO 98/53058;WO 98/53059;WO 98/53060;WO 02/016536和WO 03/016496。

[0076] “选择的”锌指蛋白或TALE是没有在自然界中发现的蛋白质,该蛋白质的制造主要是经验过程的结果,如噬菌体展示、相互捕获或杂种选择。参见例如US 5,789,538;US 5,925,523;US 6,007,988;US 6,013,453;US 6,200,759;WO 95/19431;WO 96/06166;WO 98/53057;WO 98/54311;WO 00/27878;WO 01/60970 WO 01/88197和WO 02/099084。

[0077] “TtAgo”是被认为涉及基因沉默的原核Argonaute蛋白质。TtAgo来源于细菌嗜热栖热菌。参见例如Swarts等人,同前,G.Sheng等人(2013) Proc.Natl.Acad.Sci.U.S.A.111,652)。“TtAgo系统”是所需要的所有组分,包括,例如,用于由TtAgo酶裂解的引导DNA。

[0078] “重组”是指两个多核苷酸之间交换遗传信息的过程,包括但不限于通过非同源末端连接(NHEJ)和同源重组的供体捕获。对本公开内容来说,“同源重组(HR)”指的是这种交换的特定形式,该交换发生在例如经由同源导向修复机理所进行的细胞中双链断裂修复期间。这个过程需要核苷酸序列同源性,使用“供体”分子以成为“靶”分子(即,经历双链断裂的分子)的修复模板,并且这个过程被不同地称为“无交叉基因转变”或“短序列基因转变”,因为它导致基因信息从供体向靶传递。不希望被任何具体理论束缚,这种传递可以涉及在断裂的靶和供体之间形成的异源双链体DNA的错配校正,和/或“取决于合成的链退火”,在该退火中供体用来重新合成将成为靶一部分的基因信息,和/或相关的过程。这种特定的HR经常导致靶分子序列的变化,这样使得供体多核苷酸的部分或所有序列并入靶多核苷酸中。

[0079] 在本公开的方法中,如本文所描述的一种或多种靶向核酸酶可在预先确定的位点中,在靶序列(例如,细胞染色质)中产生双链断裂,并且与断裂区中的核苷酸序列具有同源性的“供体”多核苷酸可以引入细胞中。已经展示了存在双链断裂可促进供体序列的整合。供体序列可以物理整合或,可替代地,供体多核苷酸被用作用于经由同源组合来修复断裂的模板,这将导致引入供体中的全部或部分核酸序列至细胞染色质中。因此,可以改变细胞染色质中的第一序列,并且在某些实施方案中,可以转变成存在于供体多核苷酸中的序列。因此,使用术语“替换”或“更换”可以理解为表示用另一核苷酸序列来替换一个核苷酸序列,(即,在信息意义上替换序列),并且不必然需要用另一多核苷酸来物理或化学地替换一个多核苷酸。

[0080] 在本文描述的任何方法中,额外的锌指蛋白或TALEN对可以用于细胞内额外靶位点的额外双链裂解。

[0081] 在用于靶向重组和/或替换和/或改变细胞染色质中所关注区域序列方法的实施方案中,染色体序列通过用外源“供体”核苷酸序列同源重组来改变。如果存在与断裂区同源的序列,那么可通过细胞染色质中的双链断裂的存在来刺激这种同源重组。

[0082] 在本文描述的任何方法中,第一核苷酸序列(“供体序列”)可以包含与所关注区域中基因组序列同源但不同的序列,因此将刺激同源重组以在所关注区域插入不相同序列。

因此,在某些实施方案中,与所关注区域序列同源的供体序列部分展现了与所替换的基因组序列约80%至99%之间(或任何在其间的整数)的序列同一性。在其他实施方案中,供体和基因组序列之间的同源性高于99%,例如如果具有超过100个邻接碱基对的供体与基因组序列之间只有1个核苷酸不同。在某些情况下,供体序列的非同源部分可以包含不存在于所关注区域中的序列,这样使得新的序列引入所关注区域中。在这些情况下,非同源序列通常侧接具有50至1,000个碱基对(或在其间的任何整数值)或任何大于1,000个碱基对数目的序列,这些序列与所关注区域中的序列同源或相同。在其他实施方案中,供体序列与第一序列是非同源的,并且供体序列通过非同源重组机理插入基因组中。

[0083] 本文描述的任何方法可以通过靶向整合中断所关注基因表达的供体序列来用于细胞中一种或多种靶序列的部分或完全失活。也提供了具有部分或完全失活基因的细胞系。

[0084] 此外,如本文所描述的靶向整合方法也可以用来整合一种或多种外源序列。外源核酸序列可以包括,例如一种或多种基因或cDNA分子,或任何类型编码或非编码的序列,以及一种或多种控制成分(例如,启动子)。另外,外源核酸序列可能产生一种或多种RNA分子(例如,小发夹结构的RNA(shRNA)、抑制性RNA(RNAi)、微型RNA(miRNA)等)。

[0085] “裂解”指的是DNA分子共价主链的断裂。可通过各种各样的方法来开始裂解,所述方法包括但不限于磷酸二酯键的酶水解或化学水解。单链裂解和双链裂解均是可能的,并且双链裂解可由于两个相异单链裂解事件而发生。DNA裂解可导致平端或交错端产生。在某些实施方案中,融合多肽用于靶向双链DNA裂解。

[0086] “裂解半结构域”是与第二多肽(相同的或不同的)偶联的多肽序列,从而形成具有裂解活性(优选地双链裂解活性)的复合物。术语“第一和第二裂解半结构域”;“+和-裂解半结构域”以及“右和左裂解半结构域”可交换使用,指的是二聚化的裂解半结构域对。

[0087] “工程化裂解半结构域”是已经修饰以便与另一裂解半结构域(例如,另一工程化裂解半结构域)形成专性异源二聚体的裂解半结构域。还参见美国专利号7,914,796;8,034,598;和8,623,618;以引用方式全部并入本文。

[0088] 术语“序列”指的是任何长度的核苷酸序列,该核苷酸序列可以是DNA或RNA;可以是直链的、环状的或支链的,以及可以是单链或双链的。术语“供体序列”指的是插入基因组的核苷酸序列。供体序列可以是任何长度,例如2与10,000个核苷酸之间的长度(或任何在其间或其上的整数值),优选地约100与1,000个核苷酸之间的长度(或任何在其间的整数值),更优选地约200与500个核苷酸之间的长度。

[0089] “染色质”是包括细胞基因组的核蛋白结构。细胞染色质包括核酸(主要为DNA)以及蛋白质,该蛋白质包括组蛋白和非组蛋白的染色体蛋白。大多数真核细胞染色质以核小体的形式存在,其中核小体核心包括近似150个与八聚物相关的DNA碱基对,该八聚物包括组蛋白H2A、H2B、H3以及H4的每两个;并且接头DNA(取决于有机体具有可变长度)在核小体核心之间伸展。组蛋白H1分子一般与接头DNA相关。就本公开内容来说,术语“染色质”意指包含所有类型的细胞核蛋白,不论是原核的还是真核的。细胞染色质包括染色体染色质和游离基因染色质。

[0090] “染色体”是包括细胞所有或部分基因组的染色质复合物。细胞基因组的特征通常在于它的核型,该核型是包括细胞基因组的所有染色体集合。细胞的基因组可以包括一种

或多种染色体。

[0091] “游离基因”是复制的核酸、核蛋白复合物或其他包含核酸的结构，该核酸不是细胞染色体核型的一部分。游离基因的实例包括质粒和某些病毒基因组。

[0092] “靶位点”或“靶序列”是定义结合分子所结合(只要存在足以结合的条件)核酸的一部分的核酸序列。

[0093] “外源”分子是通常不存在于细胞中，但可以通过一种或多种基因的、生物化学的或其他方法来引入细胞中的分子。“通常存在于细胞中”是相对于细胞的具体发育阶段和环境条件而测定。因此，例如，仅在肌肉的胚胎发育期间存在的分子相对于成年肌肉细胞是外源分子。类似地，通过热激诱导的分子相对于非热激的细胞是外源分子。外源分子可以包括，例如功能失常的内源分子的功能形式或功能正常的内源分子的功能失常形式。

[0094] 其中，外源分子可以是小分子，如通过组合化学过程而产生的小分子，或大分子如蛋白质、核酸、碳水化合物、脂质、糖蛋白、脂蛋白、多糖、任何上述分子的修饰衍生物或任何包含上述分子的一种或多种的复合物。核酸包括DNA和RNA，可以是单链或双链的；可以是直链、支链或环状的；并且可以是任何长度。核酸包括能够形成双链体以及三链体的那些核酸。参见例如美国专利号5,176,996和5,422,251。蛋白质包括，但不限于DNA结合蛋白、转录因子、染色质重塑因子、甲基化的DNA结合蛋白、聚合酶、甲基化酶、脱甲基酶、乙酰基转移酶、脱乙酰酶、激酶、磷酸酶、整合酶、重组酶、连接酶、拓扑异构酶、促旋酶以及解旋酶。

[0095] 外源分子可以是与内源分子相同类型的分子，例如外源蛋白质或核酸。举例来说，外源核酸可以包括感染的病毒基因组、引入细胞中的质粒或游离基因、或通常不存在于细胞中的染色体。用于将外源分子引入细胞的方法是本领域技术人员已知的，并且包括但不限于脂质介导的传递(即，脂质体，包括中性和阳离子脂质)、电穿孔、直接注入、细胞融合、粒子轰击、磷酸钙共沉淀、DEAE-葡聚糖介导的传递以及病毒载体介导的传递。外源性分子也可为与内源性分子相同类型的分子，但是来源于与产生细胞的物种不同的物种。举例来说，人核酸序列可引入最初来源于小鼠或仓鼠的细胞系。

[0096] 相反，“内源”分子是在具体环境条件下，在具体发育阶段通常存在于具体细胞中的分子。举例来说，内源核酸可以包括染色体、线粒体、叶绿体或其他细胞器的基因组、或天然存在的游离核酸。额外的内源分子可以包括蛋白质，例如转录因子和酶。

[0097] “融合”分子是其中两个或更多个亚单位分子得以连接优选地共价连接的分子。亚单位分子可以是相同化学类型的分子，或可以是不同化学类型的分子。第一种类型融合分子的实例包括但不限于融合蛋白(例如，ZFP或TALE DNA结合结构域与一个或多个活化结构域之间融合)以及融合核酸(例如，编码上文中描述的融合蛋白的核酸)。第二种类型融合分子的实例包括但不限于形成三链体核酸与多肽之间融合以及小沟结合剂与核酸之间融合。

[0098] 细胞中融合蛋白的表达可以由融合蛋白向细胞递送或通过编码融合蛋白的多核苷酸向细胞递送而引起，其中转录多核苷酸，并且翻译转录物以产生融合蛋白。分子间剪接、多肽裂解以及多肽连接反应也可以在细胞中的蛋白质表达中涉及到。用于向细胞递送多核苷酸和多肽的方法在本公开内容别处提出。

[0099] “基因”对本公开内容来说包括编码基因产物(参见下文)的DNA区，以及调控基因产物产生的所有DNA区，无论这些调控序列是否与编码和/或转录序列邻近。因此，基因包括，但不一定限于启动子序列、终止子、翻译调控序列如核糖体结合位点和内核糖体进入位

点、增强子、沉默子、绝缘子、边界元件、复制起点、基质结合位点以及基因座控制区。

[0100] “基因表达”指的是将包含在基因中的信息转变成基因产物。基因产物可以是基因的直接转录产物(例如,mRNA,tRNA,rRNA、反义RNA、核酶、结构RNA或任何其他类型的RNA)或通过mRNA翻译产生的蛋白质。基因产物也包括通过如加帽、聚腺苷酸化、甲基化以及编辑的方法来修饰的RNA,以及通过例如甲基化、乙酰化、磷酸化、遍在蛋白化、ADP-核糖基化、肉豆蔻基化以及糖基化来修饰的蛋白质。

[0101] 基因表达“调节”指的是基因活性变化。表达调节可以包括但不限于基因活化和基因阻抑。基因编辑(例如,裂解、变化、失活、随机突变)可以用来调节表达。基因失活指的是与不包括相较于如本文所描述的ZFP的细胞,基因表达的任何减少。因此,基因失活可能是部分的或完全的。

[0102] “所关注区”是任何细胞染色质区,如,例如基因或基因内部或基因附近的非编码序列,其中结合外源分子是可取的。结合可以是为了靶向DNA裂解和/或靶向重组的目的。所关注区可以存在于例如染色体、游离基因、细胞器基因组(例如,线粒体、叶绿体)或感染病毒基因组中。所关注区可以在基因编码区内,在转录的非编码区内,如例如前导序列、尾随序列或内含子,或在编码区上游或下游内的非转录区内。所关注区可以小至单个核苷酸对,或高达2,000个核苷酸对长度,或任何整数值的核苷酸对。

[0103] “真核”细胞包括但不限于真菌细胞(诸如酵母)、植物细胞、动物细胞、哺乳动物细胞和人细胞(例如,T-细胞)。

[0104] 术语“有效连接”或“有效地连接”(或“可操作连接”)关于并列两种或更多种组分(如序列成分)可交换使用,其中组分经布置,这样既使得组分功能正常又使得组分的至少一种可以介导对于其他组分的至少一种所施加的功能。作为说明,如果响应于一种或多种转录调控因子存在或缺失,转录调控序列控制编码序列转录水平,那么转录调控序列如启动子可有效连接至编码序列。转录调控序列一般以顺式与编码序列有效连接,但无需直接邻近编码序列。举例来说,增强子是有效连接至编码序列的转录调控序列,虽然它们不是邻接的。

[0105] 相对于融合多肽,术语“有效地连接”可以指的是如下事实:每个组分在与其他组分的连接状态下进行与其在未这样连接的情况下所进行的功能相同的功能。举例来说,相对于ZFP DNA结合结构域融合至活化结构域的融合多肽,如果在融合多肽中,ZFP DNA结合结构域能够结合它的靶位点和/或它的结合位点,同时活化结构域能够上调基因表达,那么ZFP DNA结合结构域和活化结构域是有效连接的。相对于ZFP DNA结合结构域融合至裂解结构域的融合多肽,如果在融合多肽中,ZFP DNA结合结构域能够结合它的靶位点和/或它的结合位点,同时裂解结构域能够在靶位点附近裂解DNA,那么ZFP DNA结合结构域和裂解结构域是有效连接的。

[0106] 蛋白质、多肽或核酸的“功能片段”是序列不同于全长蛋白质、多肽或核酸,但仍保留与全长蛋白质、多肽或核酸相同功能的蛋白质、多肽或核酸。功能片段可以拥有比对应天然分子更多、更少或相同数量的残基,和/或可以包含一种或多种氨基酸或核酸取代物。用于测定核酸功能(例如,编码功能,可杂交另一核酸的能力)的方法在本领域是众所周知的。类似地,确定蛋白质功能的方法是熟知的。举例来说,多肽的DNA结合功能可以例如通过过滤结合、电泳迁移位移或免疫沉淀实验来测定。DNA裂解可以通过凝胶电泳来分析。参见

Ausubel等人,同上文。蛋白质可与另一蛋白质相互作用的能力可以通过例如共免疫沉淀、双杂交实验或基因和生物化学两者的互补作用来测定。参见例如Fields等人(1989)Nature340:245-246;美国专利号5,585,245和PCT WO 98/44350。

[0107] “载体”能够将基因序列传递至靶细胞。典型地,“载体构建体”、“表达载体”和“基因转移载体”是指能够引导所关注基因的表达并且可将基因序列转移至靶细胞的任何核酸构建体。因此,此术语包括克隆,和表达媒介物,以及整合载体。

[0108] “安全港”基因座是基因组内的基因座,其中基因可插入而对于宿主细胞没有任何有害效应。最有利的是所插入基因序列的表达不受来自相邻基因的任何通读表达干扰的安全港基因座。由核酸酶靶向的安全港基因座的非限制实例包括CCR5、CCR5、HPRT、AAVS1、Rosa和白蛋白。参见例如美国专利号7,951,925和8,110,379;美国公布号20080159996;201000218264;20120017290;20110265198;20130137104;20130122591;20130177983和20130177960和美国临时申请号61/823,689)

[0109] 核酸酶

[0110] 本文描述组合物,尤其适用于整合编码在患有疾病或病症的受试者中缺少、缺陷或异常表达的功能性蛋白质的序列的核酸酶(例如,在患有LSD的受试者中缺少或缺陷的蛋白质和/或来自患有甲型或乙型血友病的受试者或在所述受试者体内的细胞的基因组中的凝血因子(例如F8和/或F.IX)蛋白质)。在某些实施方案中,核酸酶是天然存在的。在其他实施方案中,核酸酶非天然存在的,即,在DNA-结合结构域和/或裂解结构域中工程化。举例来说,可改变天然存在核酸酶的DNA-结合结构域以结合至选定靶位点(例如,已经工程化以结合至不同于同源结合位点的大范围核酸酶)。在其他实施方案中,核酸酶包含异源DNA-结合和裂解结构域(例如,锌指核酸酶;TAL-效应结构域DNA结合蛋白;具有异源裂解结构域的大范围核酸酶DNA-结合结构域)和/或使用工程化单一引导RNA的CRISPR/Cas系统)。

[0111] A.DNA结合结构域

[0112] 任何DNA-结合结构域可用于本文公开的组合物和方法中,包括但不限于锌指DNA-结合结构域、TALE DNA结合结构域、CRISPR/Cas核酸酶的DNA-结合部分,或来自大范围核酸酶的DNA-结合结构域。

[0113] 在某些实施方案中,核酸酶是天然存在或工程化(非天然存在的)大范围核酸酶(归巢内切核酸酶)。示例性归巢核酸内切酶包括I-SceI、I-CeuI、PI-PspI、PI-Sce、I-SceIV、I-CsmI、I-PanI、I-SceII、I-PpoI、I-SceIII、I-CreI、I-TevI、I-TevII和I-TevIII。它们的识别序列是已知的。还参见美国专利号5,420,032;美国专利号6,833,252;Bel fort等人(1997)Nucleic AcidsRes.25:3379-3388;Dujon等人(1989)Gene 82:115-118;Perler等人(1994)Nucleic Acids Res.22,1125-1127;Jasin(1996)Trends Genet.12:224-228;Gimble等人(1996)J.Mol.Biol.263:163-180;Argast等人(1998)J.Mol.Biol.280:345-353和the New England Biolabs catalogue。工程化大范围核酸酶描述于例如美国专利公布号20070117128中。归巢核酸内切酶和大范围核酸酶的DNA-结合结构域可在核酸酶的总体情况下改变(即,使得核酸酶包括同源裂解结构域)或可融合至异源裂解结构域。来自大范围核酸酶的DNA-结合结构域还可展现核酸酶活性(例如,cTALEN)。

[0114] 在其他实施方案中,DNA-结合结构域包括天然存在或工程化(非天然存在的)TAL效应DNA结合结构域。参见例如美国专利号8,586,526,以引用的方式整体并入本文。黄单孢

菌属的植物病原菌已知在重要农作物中导致许多疾病。黄单孢菌属的致病性取决于保守类型III分泌(T3S)系统,其将超过25种不同效应蛋白质注入植物细胞中。在这些注入蛋白质之中的是转录活化剂样(TAL)效应物,其模拟植物转录活化剂并且操纵植物转录组(参见Kay等人(2007)Science318:648-651)。这些蛋白质含有DNA结合结构域和转录活化结构域。最好表征的TAL-效应物中的一种是来自野油菜黄单胞菌辣椒斑点病致病变型的AvrBs3(参见Bonas等人(1989)Mol Gen Genet 218:127-136和W02010079430)。TAL-效应物含有串联重复序列的集中结构域,每个重复序列含有大约34个氨基酸,所述重复序列对于这些蛋白质的DNA结合特异性是关键性的。另外,其含有核定位序列和酸性转录活化结构域(关于综述,参见Schornack S,等人(2006)J Plant Physiol 163 (3):256-272)。另外,在植物病原细菌青枯菌中,已经发现称做brg11和hpx17的两种基因,所述基因与青枯菌生物变型1菌株GMI1000和生物变型4菌株RS1000中的黄单孢菌的AvrBs3家族同源(参见Heuer等人(2007)Appl and Envir Micro 73 (13):4379-4384)。这些基因在核苷酸序列上彼此98.9%相同但是不同之处在于在hpx17的重复结构域中缺失1,575个碱基对。然而,两种基因产物与黄单孢菌属的AvrBs3家族蛋白质具有小于40%序列同一性。参见例如美国专利号8,586,526,以引用的方式整体并入本文。

[0115] 这些TAL效应物的特异性取决于在串联重复序列中发现的序列。重复序列包含大约102个碱基对并且重复序列通常彼此91-100%同源(Bonas等人,同前)。重复序列的多态性通常位于位置12和13并且在位置12和13处的高变双残基(重复可变双残基或RVD区域)的同一性与TAL-效应物的靶序列中的邻接核苷酸的同一性之间似乎存在一一对应的关系(参见Moscou和Bogdanove,(2009)Science 326:1501和Boch等人(2009)Science 326:1509-1512)。实验上,已经确定这些TAL-效应物的DNA识别的天然代码以使得位置12和13处的HD序列(重复可变双残基或RVD)导致a结合至胞嘧啶(C),NG结合至T,NI结合至A、C、G或T,NN结合至A或G,和ING结合至T。这些DNA结合重复序列已经以新的组合和多个重复序列来组装成蛋白质,制得人工转录因子,所述因子能够与新的序列相互作用并且活化植物细胞中的非内源性报道基因的表达(Boch等人,同前)。工程化TAL蛋白质已经连接至FokI裂解半结构域以产生TAL效应结构域核酸酶融合(TALEN),包括具有非典型RVD的TALEN。参见,例如美国专利号8,586,526。

[0116] 在一些实施方案中,TALEN包含内切核酸酶(例如,FokI)裂解结构域或裂解半结构域。在其他实施方案中,TALE-核酸酶是兆TAL。这些兆TAL核酸酶是包含TALE DNA结合结构域和大范围核酸酶裂解结构域的融合蛋白。大范围核酸酶裂解结构域作为单体来起作用并且不需要二聚化以获得活性。(参见Boisse1等人(2013)Nucl Acid Res:1-13,doi:10.1093/nar/gkt1224)。

[0117] 在更进一步实施方案中,核酸酶包括紧凑TALEN。这些核酸酶是将TALE DNA结合结构域连接至TevI核酸酶结构域的单链融合蛋白。融合蛋白可充当由TALE区域定位的切口酶,或可产生双链断裂,取决于TALE DNA结合结构域相对于TevI核酸酶结构域安置的位置(参见Beurdeley等人(2013)Nat Comm:1-8DOI:10.1038/ncomms2782)。另外,核酸酶结构域还可展现DNA-结合功能。任何TALEN可与具有一个或多个兆TALE的额外TALEN(例如,一种或多种TALEN(cTALEN或FokI-TALEN)组合使用。

[0118] 在某些实施方案中,DNA结合结构域包括锌指蛋白。优选地,锌指蛋白是非天然存

在的,因为它经过工程化以结合至选定靶位点。参见例如Beerli等人(2002)Nature Biotechnol.20:135-141;Pabo等人(2001)Ann.Rev.Biochem.70:313-340;Isalan等人(2001)Nature Biotechnol.19:656-660;Segal等人(2001)Curr.Opin.Biotechnol.12:632-637;Choo等人(2000)Curr.Opin.Struct.Biol.10:411-416;美国专利号6,453,242;6,534,261;6,599,692;6,503,717;6,689,558;7,030,215;6,794,136;7,067,317;7,262,054;7,070,934;7,361,635;7,253,273;和美国专利公布号2005/0064474;2007/0218528;2005/0267061,以引用方式全部并入本文。

[0119] 与天然存在的锌指蛋白相比,工程化的锌指结合结构域可具有新的结合特异性。工程化方法包括但不限于合理设计和各种类型的选择。合理设计包括,例如使用包括三联体(或四联体)核苷酸序列和单独锌指氨基酸序列的数据库,在其中每个三联体或四联体核苷酸序列与锌指的一个或多个氨基酸序列相关,该锌指结合具体的三联体或四联体序列。参见,例如共有的美国专利6,453,242和6,534,261,该专利以引用的方式全部并入本文。

[0120] 包括噬菌体展示和双杂交系统的示例性选择方法公开于美国专利5,789,538;5,925,523;6,007,988;6,013,453;6,410,248;6,140,466;6,200,759以及6,242,568;以及W0 98/37186;W0 98/53057;W0 00/27878;W0 01/88197以及GB 2,338,237中。另外,已经在例如共有的W0 02/077227中描述了用于锌指结合结构域的结合特异性增强。

[0121] 另外,如这些和其他参考文献中所公开,锌指结构域和/或多指锌指蛋白可使用任何合适接头序列,包括例如5个或更多个氨基酸长度的接头来连接在一起。关于6个或更多个氨基酸长度的示例性接头序列,还参见美国专利号6,479,626;6,903,185;和7,153,949。本文描述的蛋白质可以包括单独蛋白质锌指之间适合接头的任何组合。

[0122] 靶位点的选择;ZFP和设计并构建融合蛋白(和编码所述蛋白质的多核苷酸)的方法为本领域技术人员已知并且详细描述于美国专利号6,140,081;5,789,538;6,453,242;6,534,261;5,925,523;6,007,988;6,013,453;6,200,759;W0 95/19431;W0 96/06166;W0 98/53057;W0 98/54311;W0 00/27878;W0 01/60970;W0 01/88197;W0 02/099084;W0 98/53058;W0 98/53059;W0 98/53060;W0 02/016536和W0 03/016496中。

[0123] 另外,如这些和其他参考文献中所公开,锌指结构域和/或多指锌指蛋白可使用任何合适接头序列,包括例如5个或更多个氨基酸长度的接头来连接在一起。关于6个或更多个氨基酸长度的示例性接头序列,还参见美国专利号6,479,626;6,903,185;和7,153,949。本文描述的蛋白质可以包括单独蛋白质锌指之间适合接头的任何组合。

#### [0124] B. 裂解结构域

[0125] 任何合适裂解结构域可有效地连接至DNA-结合结构域以形成核酸酶,诸如锌指核酸酶、TALEN或CRISPR/Cas核酸酶系统。参见例如美国专利号7,951,925;8,110,379和8,586,526;美国公布号20080159996;201000218264;20120017290;20110265198;20130137104;20130122591;20130177983和20130177960和美国临时申请号61/823,689。

[0126] 如以上提及,裂解结构域可与DNA-结合结构域异源,例如锌指DNA-结合结构域和裂解结构域来自核酸酶或TALEN DNA-结合结构域和裂解结构域,或大范围核酸酶DNA-结合结构域和裂解结构域来自不同核酸酶。异源裂解结构域可从任何内切核酸酶或外切核酸酶获得。可以从其中得到裂解结构域的示例性的内切核酸酶包括但不限于限制性内切核酸酶和归巢内切核酸酶。参见例如2002-2003Catalogue,New England Biolabs,Beverly,MA;和

Belfort等人(1997)Nucleic Acids Res.25:3379-3388。裂解DNA的其他酶是已知的(例如S1核酸酶、绿豆核酸酶、胰腺DNA酶I、微球菌核酸酶、酵母H0核酸内切酶;还参见Linn等人(编)Nucleases,Cold Spring Harbor Laboratory Press,1993)。这些酶(或其功能片段)的一种或多种可以用作裂解结构域和裂解半结构域的来源。

[0127] 类似地,裂解半结构域可以从任何核酸酶或其部分中得到,如上所阐述,该裂解半结构域或其部分需要用于裂解活性的二聚作用。通常地,如果融合蛋白包含裂解半结构域,那么裂解需要两个融合蛋白。可替代地,可以使用单个包含两个裂解半结构域的蛋白质。两个裂解半结构域可以从相同内切核酸酶(或其功能片段)中得到,或每个裂解半结构域可以从不同内切核酸酶(或其功能片段)中得到。另外,两个融合蛋白的靶位点是优选地相对于彼此布置的,这样使得两个融合蛋白与它们各自的靶位点的结合将裂解半结构域放置于使得裂解半结构域例如通过二聚作用来形成功能裂解结构域的彼此空间定位中。因此,在某些实施方案中,靶位点的近侧通过5至8个核苷酸或通过15至18个核苷酸来分开。然而任何数目的核苷酸或核苷酸对可以介于两个靶位点之间(例如,从2至50个核苷酸对或更多)。通常地,裂解的位点位于靶位点之间。

[0128] 限制核酸内切酶(限制酶)存在于许多物种中且能够以序列特异性方式结合DNA(在识别位点处),且在结合位点处或附近裂解DNA。某些限制酶(例如IIS型)在远离识别位点的位点处裂解DNA且具有可分开的结合结构域和裂解结构域。例如,IIS型酶FokI在距其于一条链上的识别位点9个核苷酸处,并且在距其于另一条链上的识别位点13个核苷酸处催化DNA的双链裂解。参见例如美国专利5,356,802;5,436,150和5,487,994;以及Li等人(1992)Proc.Natl.Acad.Sci.USA89:4275-4279;Li等人(1993)Proc.Natl.Acad.Sci.USA 90:2764-2768;Kim等人(1994a)Proc.Natl.Acad.Sci.USA 91:883-887;Kim等人(1994b)J.Biol.Chem.269:31,978-31,982。因此,在一个实施方案中,融合蛋白包含来自至少一种IIS型限制性酶的裂解结构域(或裂解半结构域)以及一个或多个锌指结合结构域,该锌指结合结构域可能是或可能不是工程化的。

[0129] 裂解域与结合结构域分离的示例性IIS型限制酶是Fok I。这个具体的酶以二聚体的形式起作用.Bitinaite等人(1998)Proc.Natl.Acad.Sci.USA95:10,570-10,575。因此,对本公开内容来说,认为用于公开融合蛋白中的Fok I酶部分是裂解半结构域。因此,对于使用锌指Fok I融合体来靶向双链裂解和/或靶向取代细胞序列,可以使用两个融合蛋白(每个该融合蛋白包含Fok I裂解半结构域)来重新构成催化活性的裂解结构域。可替代地,也可以使用包含锌指结合结构域和两个Fok I裂解半结构域的单个多肽分子。在本公开内容别处提供了使用锌指Fok I融合体来靶向裂解和靶向序列变化的参数。

[0130] 裂解结构域或裂解半结构域可以是蛋白质的任何部分,该蛋白质保留裂解活性或保留多聚(例如,二聚)以形成功能裂解结构域的能力。

[0131] 示例性的IIS型限制性酶描述于国际公布WO 07/014275中,该公布以引用的方式全部并入本文。额外的限制性酶也包含独立的结合和裂解结构域,并且这些酶是本公开内容涵盖的。参见例如Roberts等人(2003)Nucleic Acids Res.31:418-420。

[0132] 在某些实施方案中,裂解结构域包括最大限度地减少或防止同源二聚化的一种或多种工程化裂解半结构域(也称为二聚化结构域突变体),如例如美国专利公布号20050064474;20060188987;20070305346和20080131962所描述,其公开内容以引用方式全

部并入本文中。在Fok I的位置446、447、479、483、484、486、487、490、491、496、498、499、500、531、534、537以及538处的氨基酸残基全部是用于影响Fok I裂解半结构域二聚的靶。

[0133] 形成专性异源二聚体的Fok I的示例性工程化裂解半结构域包括如下对,在该对中第一种裂解半结构域包括在Fok I的位置490和538处的氨基酸残基的突变,并且第二种裂解半结构域包括在氨基酸残基486和499处的突变。

[0134] 因此,在一个实施方案中,490处的突变用Lys (K) 替换Glu (E);在538处的突变用Lys (K) 替换Iso (I);486处的突变用Glu (E) 替换Gln (Q);并且位置499处的突变用Lys (K) 替换Iso (I)。更确切地说,本文描述的工程化裂解半结构域通过在一个裂解半结构域中位置490 (E→K) 和538 (I→K) 突变以产生指定的“E490K:I538K”工程化裂解半结构域,以及通过在另一裂解半结构域中位置486 (Q→E) 和499 (I→L) 突变以产生指定的“Q486E:I499L”工程化裂解半结构域来制备。本文描述的工程化裂解半结构域是专性异源二聚突变体,在该突变体中异常裂解被减至最低或消除。参见例如美国专利号7,914,796和8,034,598,其公开内容出于所有目的以引用方式全部并入。在某些实施方案中,工程化裂解半结构域包含在位置486、499以及496(相对于野生型Fok I编号)处的突变,例如在位置486处用Glu (E) 残基替换野生型Gln (Q) 残基、在位置499处用Leu (L) 残基替换野生型Iso (I) 残基以及在位置496处用Asp (D) 或Glu (E) 残基(也分别称为“ELD”和“ELE”结构域)替换野生型Asn (N) 残基的突变。在其他实施方案中,工程化裂解半结构域包含在位置490、538以及537(相对于野生型Fok I编号)处的突变,例如在位置490处用Lys (K) 残基替换野生型Glu (E) 残基、在位置538处用Lys (K) 残基替换野生型Iso (I) 残基以及在位置537处用Lys (K) 残基或Arg (R) 残基(也分别称为“KKK”和“KKR”结构域)替换野生型His (H) 残基的突变。在其他实施方案中,工程化裂解半结构域包含在位置490和537(相对于野生型Fok I编号)处的突变,例如在位置490处用Lys (K) 残基替换野生型Glu (E) 残基以及在位置537处用Lys (K) 或Arg (R) 残基(也分别称为“KIK”和“KIR”结构域)替换野生型His (H) 残基的突变。(参见美国专利第8,623,618号)。在其他实施方案中,工程化裂解半结构域包含“Sharkey”和/或“Sharkey’”突变(参见Guo等人(2010)J.Mol.Biol.400 (1):96–107)。

[0135] 本文描述的工程化裂解半结构域可以使用任何适合的方法来制备,例如通过如美国专利号7,888,121;7,914,796;8,034,598和8,623,618中所描述的野生型裂解半结构域(Fok I)位点导向突变发生来制备。

[0136] 或者,核酸酶可在体内在使用所谓“分裂酶”技术核酸靶位点处组装(参见,例如美国专利公布号20090068164)。这类分裂酶的组分可表达于分开表达构建体中,或可在开放解读码组中连接,其中个别组分例如通过自裂解2A肽或IRES序列来分开。组分可为个别锌指结合结构域或大范围核酸酶核酸结合结构域的结构域。

[0137] 核酸酶可在使用之前针对活性来筛选,例如在基于酵母的染色体系统中,如W02009/042163和20090068164所描述。核酸酶表达构建体可使用在本领域中已知的方法容易地设计。参见例如美国专利公布20030232410;20050208489;20050026157;20050064474;20060188987;20060063231;和国际公布WO 07/014275。核酸酶的表达可在组成型启动子或可诱导启动子的控制下,例如在蜜三糖和/或半乳糖存在下活化(去抑制)并且在葡萄糖存在下被抑制的半乳糖激酶启动子。

[0138] 在某些实施方案中,核酸酶包括CRISPR/Cas系统。编码系统的RNA组分的CRISPR

(集群规则间隔短旋转对称重复序列)基因座,和编码蛋白质的cas(CRISPR相关联)基因座(Jansen等人2002.Mol.Microbiol.43:1565-1575;Makarova等人2002.Nucleic Acids Res.30:482-496;Makarova等人2006.Biol.Direct 1:7;Haft等人2005.PLoS Comput.Biol.1:e60)构成CRISPR/Cas核酸酶系统的基因序列。微生物宿主中的CRISPR基因座含有CRISPR-相关联(Cas)基因的组合以及能够设计CRISPR-介导核酸裂解的特异性的非编码RNA成分。

[0139] II型CRISPR是最好表征的系统之一并且在四个相继步骤中进行靶向DNA双链断裂。首先,两个非编码RNA,前crRNA阵列和tracrRNA,从CRISPR基因座转录。其次,tracrRNA杂交至前-crRNA的重复序列区域并且介导将前-crRNA处理至含有个别间隔序列的成熟crRNA。第三,经由crRNA上的间隔序列与靶DNA上的原型间隔序列相邻基序(PAM)附近的原型间隔序列之间的Watson-Crick碱基配对,成熟crRNA:tracrRNA复合物将Cas9引导至靶DNA,这是靶识别的附加要求。最后,Cas9介导靶DNA的裂解以产生原型间隔序列内的双链断裂。CRISPR/Cas系统的活动包括三个步骤:(i)将外来DNA序列插入CRISPR阵列以阻止将来攻击,这是称为‘适应’的过程,(ii)表达相关蛋白质,以及表达并处理阵列,随后(iii)RNA-介导干扰外来核酸。因此,在细菌细胞中,多个所谓‘Cas’蛋白质涉及CRISPR/Cas系统的天然功能并且在诸如插入外来DNA等的功能中发挥作用。

[0140] 在某些实施方案中,Cas蛋白质可为天然存在Cas蛋白质的“功能衍生物”。原生序列多肽的“功能衍生物”是具有与原生序列多肽相同定性生物特性的化合物。“功能衍生物”包括但不限于原生序列的片段和原生序列多肽和其片段的衍生物,只要其具有与相应原生序列多肽相同的生物活性。在本文中涵盖的生物活性是功能衍生物将DNA底物水解成片段的能力。术语“衍生物”涵盖多肽的氨基酸序列变体、其共价修饰和融合。Cas多肽或其片段的合适衍生物包括但不限于Cas蛋白质或其片段的突变体、融合、共价修饰。Cas蛋白质,包括Cas蛋白质或其片段,以及Cas蛋白质或其片段的衍生物,可从细胞获得或化学合成或通过这两种程序的组合获得。细胞可为天然产生Cas蛋白质的细胞,或天然产生Cas蛋白质并且遗传工程化以在较高表达水平下产生内源性Cas蛋白质或从外源引入核酸产生Cas蛋白质的细胞,所述核酸编码与内源性Cas相同或不同的Cas。在一些情况下,细胞不天然产生Cas蛋白质并且遗传工程化以产生Cas蛋白质。

[0141] 靶向安全港和其他基因的示例性CRISPR/Cas核酸酶系统公开于例如美国临时申请号61/823,689中。

[0142] 因此,核酸酶包含DNA-结合结构域,其特异性结合至需要插入供体(转基因)的任何基因中的靶位点。

[0143] 靶位点

[0144] 如以上详细描述,DNA-结合结构域可工程化以结合至任何选定序列,例如在安全港基因座诸如白蛋白中。工程化的DNA结合结构域与天然存在的DNA结合结构域相比,可以具有新的结合特异性。工程化方法包括但不限于合理设计和各种类型的选择。合理设计包括,例如使用包括三联体(或四联体)核苷酸序列和单独锌指氨基酸序列的数据库,在其中每个三联体或四联体核苷酸序列与锌指的一个或多个氨基酸序列相关,该锌指结合具体的三联体或四联体序列。参见,例如共有的美国专利6,453,242和6,534,261,该专利以引用的方式全部并入本文。也可进行TAL-效应结构域的合理设计。参见,例如美国专利号8,586,

526。

[0145] 可适用于DNA结合结构域的包括噬菌体展示和双杂交系统的示例性选择方法公开于美国专利5,789,538;5,925,523;6,007,988;6,013,453;6,410,248;6,140,466;6,200,759以及6,242,568;以及WO 98/37186;WO 98/53057;WO 00/27878;WO 01/88197以及GB 2,338,237中。另外,已经在例如共有的WO 02/077227中描述了用于锌指结合结构域的结合特异性增强。

[0146] 靶位点的选择;核酸酶和设计并构建融合蛋白(和编码所述蛋白质的多核苷酸)的方法为本领域技术人员已知并且详细描述于美国专利申请公布号20050064474和20060188987中,所述公布以引用方式全部并入本文中。

[0147] 另外,如这些和其他参考文献中所公开,DNA结合结构域(例如多指锌指蛋白)可使用任何合适接头序列,包括例如5个或更多个氨基酸长度的接头来连接在一起。关于6个或更多个氨基酸长度的示例性接头序列,参见例如美国专利号6,479,626;6,903,185;和7,153,949。本文描述的蛋白质可以包括蛋白质的单独DNA结合结构域之间适合接头的任何组合。还参见美国专利号8,586,526。

[0148] 为了经由靶向插入编码功能F8和/或F.IX蛋白质的序列来治疗血友病,受试者的基因组中的任何所需插入位点用核酸酶裂解,所述核酸酶刺激携带F8-和/或F.IX-编码序列的供体多核苷酸的靶向插入。核酸酶的DNA-结合结构域可靶向基因组中的任何所需位点。在某些实施方案中,核酸酶的DNA-结合结构域靶向内源性安全港基因座,例如内源性白蛋白基因座。

#### [0149] 供体序列

[0150] 任何供体可经由如本文描述的核酸酶介导靶向整合来插入。在某些实施方案中,供体包含编码治疗性蛋白质,例如在患有疾病或病症的受试者中缺少、缺陷或异常表达的蛋白质的多核苷酸(转基因)。这类病症的非限制实例包括大疱性表皮松解、糖尿病、癌症、凝血病症或AAT缺陷肺气肿、凝血病症和/或溶酶体贮积疾病。

[0151] 为了治疗血友病,供体序列(还称为“外源性序列”或“供体”或“转基因”)包含编码功能凝血因子蛋白质的序列或其一部分,以便在整合供体之后产生编码并表达功能凝血因子蛋白质的序列。凝血因子蛋白质转基因的非限制实例包括因子VIII和/或因子IX,包括这些蛋白质的功能片段。在某些实施方案中,F8蛋白质的B-结构域缺失。参见例如Chuah等人(2003)101(5):1734-1743。在其他实施方案中,转基因包含编码功能F.IX蛋白质的序列,或其一部分,以便在整合供体之后产生编码并表达功能F.IX蛋白质的序列。类似地,为了治疗LSD,供体序列编码在患有LSD的受试者中缺少的一种或多种蛋白质。这类蛋白质的非限制实例包括葡糖脑苷脂酶(GBA),其在戈谢氏病中有缺陷; $\alpha$ -半乳糖苷酶(GLA),其在法布里氏病中有缺陷;艾杜糖酸酯-2-硫酸酯酶缺陷(IDS),其在亨特氏病中有缺陷; $\alpha$ -L艾杜糖苷酸酶(IDUA),其在赫尔勒氏病中有缺陷;鞘磷脂磷酸二酯酶1(SMPD1),其在尼曼匹克氏病中有缺陷。

[0152] 供体分子可插入内源性基因中以使得所有、一些或没有内源性基因得以表达。举例来说,包含如本文描述的功能凝血因子蛋白质(例如,F8和/或F.IX)序列的转基因可插入内源性白蛋白基因座以使得一些或没有内源性白蛋白与转基因一起表达。

[0153] 供体(转基因)序列可在表达融合蛋白(例如,核酸酶)之前、同时或之后引入细胞。

供体多核苷酸可含有与基因组序列的足够同源性(连续或不连续区域)以支持在它与和它带有同源性的基因组序列之间的同源重组(或同源介导修复)或,替代地,供体序列可经由非HDR机制(例如,NHEJ供体捕获)来整合,在此情况下供体多核苷酸(例如,载体)不需要含有与细胞染色质中的所关注区域同源的序列。参见例如美国专利号7,888,121和7,972,843和美国专利公布号20110281361;20100047805和20110207221。

[0154] 供体多核苷酸可为DNA或RNA、单链、双链或部分单链和部分双链并且可以线性或环状(例如,小环)形式引入细胞。参见例如美国专利公布号20100047805、20110281361、20110207221。如果以线性形式引入,那么供体序列的末端可通过为本领域技术人员已知的方法来保护(例如,避免核酸外切降解)。举例来说,一个或多个双脱氧核苷酸残基添加至线性分子的3'末端且/或自身互补寡核苷酸连接至一个或两个末端。参见例如Chang等人(1987) Proc.Natl.Acad.Sci.USA84:4959-4963;Nehls等人(1996) Science 272:886-889。用于保护外源多核苷酸免受降解的另外方法包括但不限于添加末端氨基和使用修饰的核苷酸间键联,例如像硫代磷酸酯、氨基磷酸酯和O-甲基核糖或脱氧核糖残基。可将多核苷酸引入到细胞中作为载体分子的一部分,所述载体分子具有额外序列例如像复制起点、启动子和编码抗生素耐药性的基因。此外,供体多核苷酸可以裸核酸形式,以与试剂诸如脂质体或泊洛沙姆复合的核酸形式引入,或可通过病毒(例如,腺病毒、AAV、疱疹病毒、逆转录病毒、慢病毒)来递送。

[0155] 供体总体上插入以使得其表达通过整合位点处的内源性启动子(例如,在供体整合至患者的白蛋白基因座中时的内源性白蛋白启动子)来驱动。因此,转基因通常缺少驱动其表达的控制元件(例如,启动子和/或增强子)(例如,还被称为“无启动子构建体”)。但是,显然供体可包含在整合后驱动功能性蛋白质表达的启动子和/或增强子,例如组成型启动子或可诱导或组织特异性(例如,肝脏或血小板特异性)启动子。

[0156] 供体序列可特异性整合至任何选定靶位点,从而消除传统基因疗法中的与随机整合相关联的问题。

[0157] 当白蛋白序列(内源性或转基因的一部分)与转基因一起表达时,白蛋白序列可为全长序列(野生型或突变型)或部分序列。优选地白蛋白序列是功能性的。这些全长或部分白蛋白序列的功能的非限制实例包括增加由转基因(例如,治疗性基因)表达的多肽的血清半衰期且/或充当载体。

[0158] 此外,虽然并非为表达所必需的,外源性序列还可包括转录或翻译调控序列,例如,启动子、增强子、绝缘子、内部核糖体进入位点、编码2A肽和/或多聚腺苷酸化信号的序列。

[0159] 任何供体序列可包括以下修饰中的一个或多个:密码子优化(例如,人密码子)和/或添加一个或多个糖基化位点。参见例如McIntosh等人(2013) Blood (17):3335-44。外源性序列还可包含允许靶向递送治疗性蛋白质的肽序列。举例来说,编码人p97多肽和/或其片段的核酸序列可连接至供体外源性序列以使得融合蛋白具有穿过血脑屏障的潜力(参见例如美国临时专利申请号20130183368和Karkan等人(2008) PLOS One.DOI:10.1371/journal.pone.0002469)或其他肽可用于将转基因供体编码蛋白质靶向至细胞内细胞器诸如线粒体(例如Jacotot等人(2006) Biochim Biophys Acta Bioenerg 1757:1312-1323)。

[0160] 递送

[0161] 核酸酶、编码这些核酸酶的多核苷酸、供体多核苷酸和包含本文所述蛋白质和/或多核苷酸的组合物可在体内或离体以任何合适手段递送。

[0162] 如本文描述的递送核酸酶的方法描述于例如美国专利号8,586,526;6,453,242;6,503,717;6,534,261;6,599,692;6,607,882;6,689,558;6,824,978;6,933,113;6,979,539;7,013,219;和7,163,824中,其公开内容以引用的方式全部并入本文。

[0163] 如本文描述的核酸酶和/或供体构建体还可使用含有编码锌指蛋白、TALEN蛋白质和/或CRISPR/Cas系统中的一个或多个的序列的载体来递送。可以使用任何载体系统,包括但不限于质粒载体、逆转录病毒载体、慢病毒载体、腺病毒载体、痘病毒载体;疱疹病毒载体以及腺相关病毒载体等。还参见美国专利号6,534,261;6,607,882;6,824,978;6,933,113;6,979,539;7,013,219;和7,163,824,所述专利以引用的方式全部并入本文。此外,显然任何这些载体可包含所需要的序列中的一个或多个。因此,在将一种或多种核酸酶和供体构建体引入细胞中时,可在相同载体或不同载体上携带核酸酶和/或供体多核苷酸。当使用多个载体时,每个载体可以包含编码一个或多个核酸酶和/或供体构建体的序列。在某些实施方案中,一个载体用于携带转基因和核酸酶。在其他实施方案中,使用两个载体(相同或不同载体类型),其中一个载体携带核酸酶(例如,例如具有2A肽的ZFN对的左和右ZFN)并且一个载体携带转基因。在更进一步实施方案中,使用三个载体,其中第一个载体携带核酸酶对的一个核酸酶(例如,左ZFN),第二个载体携带核酸酶对中的另一个核酸酶(例如,右ZFN)并且第三个载体携带转基因。参见图2。

[0164] 供体和/或核酸酶可以任何合适浓度使用。在某些实施方案中,供体和分开核酸酶载体以相同浓度使用。在其他实施方案中,供体和分开核酸酶载体以不同浓度使用,例如,一个载体比另一个载体多2、3、4、5、10或更大倍(例如,供体载体比核酸酶载体更多。当例如AAV载体用于递送时,包含供体和/或核酸酶的病毒载体在 $1 \times 10^8$ 与 $1 \times 10^{13}$ 颗粒/剂量(例如,细胞或动物)之间。

[0165] 常规基于病毒和非病毒的基因转移方法可用于将编码核酸酶和供体构建体的核酸引入细胞(例如,哺乳动物细胞)和靶组织中。非病毒载体递送系统包括DNA质粒、裸核酸和与如脂质体或泊洛沙姆的递送媒介物复合的核酸。病毒载体递送系统包括DNA和RNA病毒,在递送至细胞中之后,其具有游离基因或整合的基因组。关于体内递送工程化DNA-结合蛋白和包含这些结合蛋白的融合蛋白的综述,参见例如,Rebar (2004) Expert Opinion Invest. Drugs 13 (7) :829-839; Rossi等人(2007) Nature Biotech. 25 (12) :1444-1454以及一般基因递送参考文献诸如Anderson, Science 256:808-813 (1992); Nabel&Felgner, TIBTECH 11:211-217 (1993); Mitani&Caskey, TIBTECH 11:162-166 (1993); Dillon, TIBTECH 11:167-175 (1993); Miller, Nature 357:455-460 (1992); Van Brunt, Biotechnology 6 (10) :1149-1154 (1988); Vigne, Restorative Neurology and Neuroscience 8:35-36 (1995); Kremer&Perricaudet, British Medical Bulletin 51 (1) :31-44 (1995); Haddada等人,于Current Topics in Microbiology and Immunology Doerfler and Böhm (编辑) (1995) 中;和Yu等人, Gene Therapy 1:13-26 (1994)。

[0166] 核酸的非病毒递送方法包括电穿孔、脂转染、微注射、基因枪、病毒体、脂质体、免疫脂质体、聚阳离子或脂质:核酸偶联物、裸DNA、人工病毒颗粒以及DNA的试剂增强吸收。使用例如Sonitron2000系统(Rich-Mar)的超声穿孔也可以用于核酸递送。

[0167] 额外示例性的核酸递送系统包括由AmaxaBiosystems (Cologne, Germany)、Maxcyte, Inc. (Rockville, Maryland)、BTX Molecular Delivery Systems (Holliston, MA) 和Copernicus Therapeutics Inc (参见例如U.S.6,008,336) 提供的那些。脂转染描述于例如美国专利号5,049,386;4,946,787;和4,897,355中) 并且脂转染试剂有市售(例如Transfectam<sup>TM</sup>和Lipofectin<sup>TM</sup>)。适用于多核苷酸有效受体识别脂转染的阳离子和中性脂质包括Felgner、WO 91/17424、WO 91/16024的那些。

[0168] 制备脂质:核酸复合物,包括靶向脂质体诸如免疫脂质复合物,为本领域技术人员熟知(参见,例如, Crystal, Science 270:404-410 (1995); Blaese等人 Cancer Gene Ther. 2:291-297 (1995); Behr等人, Bioconjugate Chem. 5:382-389 (1994); Remy等人, Bioconjugate Chem. 5:647-654 (1994); Gao等人, Gene Therapy 2:710-722 (1995); Ahmad等人, Cancer Res. 52:4817-4820 (1992); 美国专利号4,186,183、4,217,344、4,235,871、4,261,975、4,485,054、4,501,728、4,774,085、4,837,028和4,946,787)。

[0169] 额外的递送方法包括使用将待递送的核酸封装至EnGeneIC递送媒介物(EDV)中。这些EDV使用双特异性抗体特异性地向靶组织递送,其中该抗体的一个臂具有对于靶组织的特异性,并且另一个臂具有对于EDV的特异性。抗体将EDV带至靶细胞表面,然后通过胞吞作用将EDV带入细胞中。当在细胞中时,将释放内含物(参见MacDiarmid等人(2009) Nature Biotechnology 27 (7):643)。

[0170] 使用基于RNA或DNA病毒的系统递送核酸利用用于使病毒靶向体内特定细胞以及使病毒有效载荷迁移至核中的高度进展的方法。病毒载体可直接向患者施用(体内)或它们可用于体外处理细胞,且接着向患者施用修饰的细胞(离体)。用于递送核酸酶和/或供体的常规基于病毒的系统包括但不限于逆转录病毒、慢病毒、腺病毒、腺相关、痘苗和单纯性疱疹病毒载体用于基因转移。用逆转录病毒、慢病毒和腺相关病毒基因转移方法使整合在宿主基因组中成为可能,从而常导致插入的转基因的长期表达。另外,已在许多不同细胞类型和靶组织中观察到高转导效率。

[0171] 逆转录病毒的趋向性可通过并入外来包膜蛋白质、扩增靶细胞的潜在靶群体来改变。慢病毒载体是能够转导或感染非分裂细胞并且通常产生高病毒效价的逆转录病毒载体。逆转录病毒基因转移系统的选择取决于靶组织。逆转录病毒载体包含封装能力高达6-10kb外来序列的顺式作用长末端重复序列。最小顺式作用LTR就足以用于载体的复制和封装,其接着用于将治疗基因整合至靶细胞中以提供持久的转基因表达。广泛使用的逆转录病毒载体包括基于鼠白血病病毒(MuLV)、长臂猿白血病病毒(GaLV)、猿免疫缺陷病毒(SIV)、人免疫缺陷病毒(HIV)的载体及其组合(参见,例如Buchscher等人, J.Virol. 66: 2731-2739 (1992); Johann等人, J.Virol. 66:1635-1640 (1992); Sommerfelt等人, Virology 176:58-59 (1990); Wilson等人, J.Virol. 63:2374-2378 (1989); Miller等人, J.Virol. 65:2220-2224 (1991); PCT/US94/05700)。

[0172] 在瞬时表达优选的应用中,可使用基于腺病毒的系统。基于腺病毒的载体能够在许多细胞类型中具有极高转导效率且不需要细胞分裂。用所述载体已获得高效价和表达水平。此载体可在相对简单系统中大量产生。腺病毒相关病毒("AAV")载体还用于向细胞转导靶核酸,例如,体外产生核酸和肽,和体内和离体基因治疗程序(参见,例如, West等人, Virology 160:38-47 (1987); 美国专利号4,797,368; WO 93/24641; Kotin, Human Gene

Therapy 5:793-801 (1994) ;Muzyczka, J.Clin.Invest.94:1351 (1994)。构建重组AAV载体描述于许多公布中,包括美国专利号5,173,414;Tratschin等人,Mol.Cell.Biol.5:3251-3260 (1985);Tratschin等人,Mol.Cell.Biol.4:2072-2081 (1984);Hermonat&Muzyczka,PNAS 81:6466-6470 (1984);和Samulski等人,J.Virol.63:03822-3828 (1989)。

[0173] 至少六种病毒载体方法当前可用于临床试验中的基因转移,其利用涉及通过插入辅助细胞系的基因来补充缺陷载体以产生转导剂的方法。

[0174] pLASN和MFG-S是已被用于临床试验中的逆转录病毒载体的实例(Dunbar等人,Blood 85:3048-305 (1995);Kohn等人,Nat.Med.1:1017-102 (1995);Malech等人,PNAS 94:22 12133-12138 (1997))。PA317/pLASN是用于基因治疗试验中的第一种治疗性载体。(Blaese等人,Science 270:475-480 (1995))。已经观察到MFG-S包装载体的50%或更大的转导效率。(Ellem等人,Immunol Immunother.44 (1):10-20 (1997);Dranoff等人,Hum.Gene Ther.1:111-2 (1997))。

[0175] 重组腺相关病毒载体(rAAV)是基于缺陷性和非病原性细小病毒腺相关2型病毒的有前途的替代性基因递送系统。所有载体都源于仅保留侧接转基因表达盒的AAV 145bp碱基对倒置末端重复序列的质粒。归因于整合至转导的细胞的基因组中的高效基因转移和稳定转基因递送是此载体系统的关键特征。(Wagner等人,Lancet 351:91171702-3 (1998),Kearns等人,Gene Ther.9:748-55 (1996))。其他AAV血清型,包括AAV1、AAV3、AAV4、AAV5、AAV6、AAV8、AAV9和AAVrh10或假型AAV诸如AAV2/8、AAV8.2、AAV2/5和AAV2/6和任何新AAV血清型也可根据本发明来使用。

[0176] 复制缺陷重组腺病毒载体(Ad)可以高效价产生并且容易感染许多不同细胞类型。大多数腺病毒载体被工程化以使转基因替代Ad E1a、E1b和E3基因;随后使复制缺陷性载体在反式提供缺失的基因功能的人293细胞中增殖。Ad载体可在体内转导多种类型的组织,包括非分裂性分化细胞,如见于肝、肾和肌肉组织中的细胞。常规Ad载体具有大携带能力。在临床试验中使用Ad载体的实例涉及通过肌肉内注射来抗肿瘤免疫接种的多核苷酸疗法(Sterman等人,Hum.Gene Ther.7:1083-9 (1998))。在临床试验中使用腺病毒载体来转移基因的额外实例包括Rosenecker等人,Infection 24:1 5-10 (1996);Sterman等人,Hum.Gene Ther.9:7 1083-1089 (1998);Welsh等人,Hum.Gene Ther.2:205-18 (1995);Alvarez等人,Hum.Gene Ther.5:597-613 (1997);Topf等人,Gene Ther.5:507-513 (1998);Sterman等人,Hum.Gene Ther.7:1083-1089 (1998)。

[0177] 封装细胞用于形成能够感染宿主细胞的病毒颗粒。这类细胞包括293细胞,其封装腺病毒,和 $\psi$ 2细胞或PA317细胞,其封装逆转录病毒。用于基因疗法的病毒载体通常由生产细胞系产生,所述细胞系将核酸载体封装至病毒颗粒中。载体通常含有封装并随后整合至宿主中所需要的减少病毒序列(如果可适用),其他病毒序列由编码待表达的蛋白质的表达盒替换。缺失的病毒功能由封装细胞系以反式提供。举例来说,用于基因疗法的AAV载体通常只具有来自AAV基因组的反向末端重复(ITR)序列,所述序列为封装并整合至宿主基因组中所需要。病毒DNA封装于细胞系中,其含有编码其他AAV基因,即rep和cap的辅助质粒,但是缺少ITR序列。细胞系还用作为辅助病毒的腺病毒感染。辅助病毒促进复制AAV载体和从辅助质粒表达AAV基因。由于缺少ITR序列,辅助质粒未大量地封装。以腺病毒的污染可通过例如热处理来减少,与AAV相比,腺病毒对于热处理更敏感。

[0178] 在许多基因治疗应用中,需要基因治疗载体以针对具体组织类型的高度特异性来递送。因此,病毒载体通常被修饰来通过在病毒外表面上以与病毒外壳蛋白的融合蛋白形式表达配体来对给定细胞类型具有特异性。选择配体以对已知存在于目标细胞类型上的受体具有亲和力。举例来说,Han等人,Proc.Natl.Acad.Sci.USA 92:9747-9751(1995),报道Moloney鼠白血病病毒可修饰以表达融合至gp70的人调蛋白,并且重组病毒感染表达人表皮生长因子受体的某些人乳腺癌细胞。此原理可扩展至其他病毒-靶细胞对,其中靶细胞表达受体并且病毒表达包含细胞表面受体的配体的融合蛋白。举例来说,丝状噬菌体可工程化以展示具有实际上任何选定细胞受体的特异性结合亲和力的抗体片段(例如,FAB或Fv)。虽然以上描述主要适用于病毒载体,但是相同原则可施加于非病毒载体。这类载体可工程化以含有有利于由特异性靶细胞吸收的特异性吸收序列。

[0179] 基因疗法载体可通常通过如下所述的全身性施用(例如静脉内、腹膜内、肌肉内、真皮下或颅内输注)或局部施用向个别患者施用来体内递送。或者,载体可离体递送至细胞,如自个别患者外植的细胞(例如淋巴细胞、骨髓吸出物和组织生检体)或万能供血者造血干细胞中,随后通常在选择已并入载体的细胞之后将细胞再植入受试者中。

[0180] 含有核酸酶和/或供体构建体的载体(例如,逆转录病毒、腺病毒、脂质体等)也可直接施用至生物体以便在体内转导细胞。或者,可施用裸DNA。施用是通过通常用于引入分子以最与血液或组织细胞终接触的任何途径,包括但不限于注射、输注、局部施加和电穿孔。施用这类核酸的适合方法是可用的且为本领域技术人员所熟知,且尽管一种以上途径可用于施用特定组合物,但特定途径可常提供比另一途径更直接且更有效的反应。

[0181] 本文描述的适合于引入多核苷酸(例如核酸酶编码和/或供体)的载体包括非整合慢病毒载体(IDLV)。参见例如Ory等人(1996) Proc.Natl.Acad.Sci.USA 93:11382-11388; Dull等人(1998) J.Virol.72:8463-8471; Zuffery等人(1998) J.Virol.72:9873-9880; Follenzi等人(2000) Nature Genetics 25:217-222美国专利公布号2009/054985。

[0182] 药学上可接受的载体部分地通过所施用的具体组合物以及通过用来施用组合物的具体方法来确定。因此,如下所述,存在可用的药物组合物的广泛多种的适合制剂(例如参见Remington's Pharmaceutical Sciences,17th ed.,1989)。

[0183] 体外和体内递送还可经由使用纳米颗粒来实现。当前研究的许多纳米颗粒包含与脂质或聚合物自动组装成纳米结构的治疗性分子。这些颗粒具有将治疗性剂量的核酸递送至靶组织(例如肿瘤细胞、特异性器官等)的潜力。参见例如Rink等人(2013), Curr Opin Oncol:25 (6):p.646-651。

[0184] 显然核酸酶编码序列和供体构建体可使用相同或不同系统来递送。举例来说,核酸酶和供体可由相同载体(例如,AAV)携带。或者,供体多核苷酸可由质粒携带,而一种或多种核酸酶可由不同载体(例如,AAV载体)携带。此外,不同载体可通过相同或不同途径来施用(肌肉内注射、尾静脉注射、其他静脉内注射、腹膜内施用和/或肌肉内注射)。载体可同时或任何相继顺序来递送。

[0185] 因此,本公开包括体内或离体治疗和/或预防蛋白质缺少或缺陷的疾病或病症。例如,甲型血友病可经由核酸酶介导的F8-编码序列的整合来处理。本公开还包括经由F.IX编码序列的核酸酶介导整合来在体内或离体治疗乙型血友病。类似地,本公开包括分别经由因子VII或因子X编码序列的核酸酶介导整合来治疗因子VII缺陷和因子X缺陷相关血友病。

另外,本公开包括经由在LSD中缺少或缺陷的一种或多种蛋白质的核酸酶介导整合来治疗一种或多种LSD。组合物以有效获得血清、肝脏或靶细胞中的治疗性多肽的所需浓度的量来施用至人患者。施用可通过多核苷酸递送至所需靶细胞的任何手段。举例来说,涵盖体内和离体方法。静脉内注射至门静脉是优选施用方法。其他体内施用模式包括例如直接注入肝叶或胆管和在肝脏远端的静脉内注射,包括经由肝动脉,直接注入肝实质,经由肝动脉注射,和/或经由胆道系统的逆行注射。离体施用模式包括切除肝细胞或其他肝脏细胞的体外转导,随后将转导、切除肝细胞输注回到人患者的门脉管、肝实质或胆道系统,参见例如,Grossman等人(1994) *Nature Genetics*, 6:335-341。其他施用模式包括编码因子VII、F8、F.IX、因子X、葡糖脑苷脂酶、 $\alpha$ 半乳糖苷酶、艾杜糖酸酯-2-硫酸酯酶和/或 $\alpha$ -L艾杜糖苷酸酶的转基因的离体核酸酶介导插入患者或同种异体干细胞中的安全港位置。修饰之后,处理细胞然后重新输注至患者体内以治疗疾病或病症(例如,LSD和/或血友病)。

[0186] 待施用的核酸酶和供体(例如,因子VII、F8、F.IX、因子X、GBA、GLA、IDS、IDUA或SMPD1)的有效量将在不同患者之间并且根据所关注的治疗性多肽而变化。因此,有效量最好由施用组合物的医师确定并且适当剂量可由本领域普通技术人员容易地确定。允许足够时间来整合并表达(通常例如4-15天)之后,治疗性多肽的血清或其他组织水平的分析和与施用之前的初始水平比较确定是否施用量太低、在正确范围内或太高。初始和随后施用的合适方案也可变,但是如果有必要,典型地在初始施用之后进行后续施用。随后施用可以可变时间间隔来施用,所述时间间隔在每天一次至每年一次至每隔几年一次范围内。本领域技术人员应认识到可建议适当免疫抑制技术以便通过递送载体的免疫抑制来避免转导的抑制或阻滞,参见例如,Vilquin等人(1995) *Human Gene Ther.* 6:1391-1401。

[0187] 离体和体内施用的制剂包括液体或乳化液体中的悬浮液。活性成分经常与药学上可接受并与活性成分相容的赋形剂混合。合适赋形剂包括例如水、盐水、右旋糖、甘油、乙醇等和其组合。另外,组合物可含有少量辅助物质,诸如,润湿或乳化剂、pH缓冲剂、稳定剂或增强药物组合物的有效性的其他试剂。

#### [0188] 应用

[0189] 本发明方法和组合物可用于需要提供编码一种或多种蛋白质的转基因以使得蛋白质从靶向细胞分泌的任何情况中。因此,此技术适用于患者归因于各种问题(例如,表达水平的问题或所表达的蛋白质功能不足或无功能的问题)而在一些蛋白质中有缺陷的病状。另外,A1AT缺陷病症诸如COPD或肝损伤,或可通过由分泌器官供应外源性蛋白质来缓解的其他病症、病状或疾病可通过本发明方法和组合物来成功地治疗。与代谢疾病诸如糖尿病一样,溶酶体贮积疾病可通过本发明方法和组合物来治疗。

[0190] 在治疗上适用并且通常通过注射或输注来递送的蛋白质也适用于本发明方法和组合物。作为非限制实例,产生C-肽(例如Cebix的Ersatta<sup>TM</sup>)或用于糖尿病疗法中的胰岛素。其他应用包括经由产生胶原蛋白VII来治疗大疱性表皮松解。如本文描述,IGF-1在分泌组织中的表达可用于通过脂蛋白脂肪酶的表达来增加此蛋白质在患有肝硬变和脂蛋白脂肪酶缺陷的患者中的水平。还可出于治疗性益处来分泌抗体,例如,用于治疗癌、自身免疫和其他疾病。治疗性抗体的实例包括针对TNF- $\alpha$ 、EpCAM、CD20、CD19、VEGFR、CD52等的抗体。与凝血相关的其他蛋白质可在分泌组织中产生,包括纤维蛋白原、凝血酶原、组织因子、因子V、因子XI、因子XII(哈格曼因子)、因子XIII(纤维蛋白稳定因子)、冯威勒布兰德氏因子、

激肽释放酶原、高分子量激肽原(菲茨杰拉德因子)、纤连蛋白、抗凝血酶III、肝素辅因子II、蛋白质C、蛋白质S、蛋白质Z、蛋白质Z相关蛋白水解酶抑制剂、纤溶酶原、 $\alpha$ 2-抗纤溶酶、组织纤溶酶原活化剂、尿激酶、纤溶酶原活化物抑制剂-1和纤溶酶原活化物抑制剂-2。

[0191] 本发明方法和组合物也可用于需要供应并表达编码一种或多种非编码或结构核酸(例如shRNA或RNAi)的转基因的任何环境中。这类RNA可形成抑制性结构并且适用于治疗疾病诸如脂质病症(靶向例如ApoB-100、ApoC-III、ANGPTL3、PCSK9)；冠状动脉疾病(靶向例如CRP、Apo(a))；凝血和血液病症(靶向例如F.XI、FVII、抗凝血酶、TMPRSS6)；自身免疫疾病(靶向例如ICAM-1、GCCR、GCGR、PTP-1B、VLA-4)；TTR淀粉样变性；肌肉疾病(靶向例如SMN2、GHR、DMPK)；炎症性疾病(靶向例如PKK)；肥胖症(靶向例如FGFR4)；肝脏疾病(靶向例如DGAT2、ALAS-1、C5、AAT)；癌症(靶向例如丛生蛋白、eIF-4E、Hsp27、AR)；纤维化疾病(靶向例如CTGF)；眼疾病(靶向例如C-raf激酶)；或传染病(靶向例如氨基糖苷类、铁调素、RG-101)。

[0192] 以下实施例涉及本公开的示例性实施方案，其中核酸酶包括锌指核酸酶(ZFN)。应认识到此仅用于例证目的并且可使用其他核酸酶，例如具有工程化DNA-结合结构域的归巢核酸内切酶(大范围核酸酶)和/或天然存在的工程化归巢核酸内切酶(大范围核酸酶)DNA-结合结构域和异源裂解结构域的融合，TALEN和/或包括工程化单一引导RNA的CRISPR/Cas系统。

## 实施例

[0193] 实施例1：F8转基因在体内的靶向整合

[0194] 均通过向尾静脉注射向HA/CD4<sup>-/-</sup>小鼠施用(1)对照AAV2/8载体和编码F8的AAV2/8供体转基因，“模拟物+供体”或(2)编码靶向白蛋白基因座的ZFN对的AAV载体(如美国专利公布20130177983和图10A和10B所描述)和编码F8的AAV供体转基因(“ZFN+供体”)，如美国专利公布号20120128635所描述。所使用的供体如图1和2示意性地展示并且包括大约4.4至4.7kb大小的无启动子B-结构域缺失的因子VIII(BDD-F8)cDNA。剂量以两个给药水平施用至小鼠，“低”给药水平包括AAV8-ZFN( $1\text{e}^{11}$ vg/小鼠)+AAV8-供体( $1\text{e}^{11}$ vg/小鼠)：图3B；并且“高”给药水平包括AAV8-ZFN( $5\text{e}^{11}$ vg/小鼠)+AAV8-供体( $5\text{e}^{11}$ vg/小鼠)：图3C。

[0195] 另外，F8供体通过密码子优化来优化以在哺乳动物细胞中根据标准方案来表达并且添加具有糖基化位点的接头(V3)(参见McIntosh等人，(2013)Blood 121:3335)。在此实验中，向HA/CD4<sup>-/-</sup>小鼠给予AAV8-ZFN( $5\text{e}^{10}$ vg的每一种ZFN)+AAV8-供体( $1\text{e}^{11}$ vg/小鼠)：图4。

[0196] F8的血浆水平使用标准技术来评估。

[0197] 如图3和4示出，因子8供体靶向整合至小鼠白蛋白基因座中产生在使用任一种供体时甲型血友病小鼠中的至多正常水平50%的活性水平。然而，在使用优化F8供体构建体时，只使用20%剂量即可观察到类似F8血浆水平。另外，在使用两个ZFN载体时，每个载体包含此对所需要的两个ZFN中的一个，观察到与当ZFN由2A位点分隔在一个表达载体上一起引入时相比，更高的裂解水平(图3D)。

[0198] 实施例2：F9转基因在体内的靶向整合

[0199] A.人肝细胞

[0200] 我们首先用含有F9供体的AAV6载体转导人原代肝细胞，同时转染编码靶向人血白蛋白的第一内含子内的位点的ZFN对的mRNA。

[0201] 如图9示出,用供体和ZFN处理的肝细胞在培养上清液中展示可测量的人F.IX。

[0202] B. 小鼠

[0203] 我们接下来试图在小鼠体内证明此方法。为了完成此,我们首先工程化如下示出的靶向小鼠白蛋白内含子1中的类似位置(在图12中示出)的ZFN对(在以下表1中示出)并且证实此对在体外鼠肝癌细胞中的活性。

[0204] 表1:小鼠白蛋白特异性核酸酶设计

SBS #, 靶	设计						
	小鼠-白蛋白特异性 ZFN						
	F1	F2	F3	F4	F5	F6	
[0205]	<b>SBS#30724</b> ctGAAGGTgG CAATGGTTcct ctctgct (SEQ ID NO:117)	TSGSLTR (SEQ ID NO:119)	RSDALST (SEQ ID NO:120)	QSATRTK (SEQ ID NO:121)	TSGHLSR (SEQ ID NO:121)	QSGNLAR (SEQ ID NO:2)	N/A
	<b>SBS#30725</b> ttTCCTGTAAC GATCGGgaact ggcatc (SEQ ID NO:118)	RSDHLS A(SEQ ID NO:122)	TKSNRT K(SEQ ID NO:123)	DRSNLSR (SEQ ID NO:5)	WRSSLR A(SEQ ID NO:124)	DSSDRKK Q(SEQ ID NO:125)	N/A

[0206] 如图10示出,此对在鼠肝癌细胞中起作用。

[0207] 另外,向野生型小鼠(3动物/组)经由尾静脉注射来施用编码F.IX的AAV2/8供体转基因(参见,图5A)以及编码小鼠白蛋白靶向ZFN对(“mA1b ZFN”)的AAV载体,如美国公布号20130177983所描述,或人F.IX-靶向ZFN对(“hF9ZFN”),如美国公布号20120128635所描述并且在图12A和12B中示出。载体构建体和剂量如下:白蛋白AAV2/8-ZFN为1x 10<sup>11</sup>vg/小鼠并且AAV8-供体为5x 10<sup>11</sup>vg/小鼠并且F.IX AAV8-ZFN为1x 10<sup>11</sup>vg/小鼠并且AAV8-供体为5x 10<sup>11</sup>vg/小鼠。

[0208] F9的血浆水平使用标准可购得ELISA试剂盒使用可购得抗体来评估。另外,Cel-I测定(Surveyor<sup>TM</sup>, Transgenomics)如美国公布号20120128635和Perez等人,(2008) Nat.Biotechnol.26:808-816和Guschin等人,(2010) Methods Mol Biol.649:247-56)所描述来进行。

[0209] 如图5B示出,稳健循环hFIX水平在白蛋白ZFN和F9供体递送之后获得。人F.IX特异性ZFN不识别内源性小鼠F.IX基因座,并且因此使用此核酸酶对,没有F9供体的明显整合。此外,如图6示出,基因组编辑在三个数量级下与AAV剂量成比例。

[0210] 另外,向血友病小鼠(HB小鼠)施用供体和白蛋白-ZFN,如上所述(AAV8-mA1b-ZFN为1x10<sup>11</sup>vg/小鼠并且AAV8-F9供体为5x10<sup>11</sup>vg/小鼠,还参见Li等人(2011)同前和Anguela等人(2013)同前)并且hF.IX的血浆水平和活化部分凝血活酶时间(aPTT)也通过标准可购得试剂盒(例如来自Rossix和Vitaclot,Vital<sup>®</sup> Diagnostics的Rox因子IX发色试剂盒)来确定。

[0211] 如图7示出,F9供体转基因的ZFN介导整合至HB小鼠的白蛋白基因座中导致血浆中的高水平F.IX和持续很久的凝血时间的校正。

[0212] C. 恒河猴

[0213] 为了在较大动物中测试ZFN驱动基因组修饰和转基因插入,进行两个研究。所使用

的ZFN在以下表2中示出。靶序列中的大写字母表示结合核苷酸并且小写字母表示未结合核苷酸。

[0214] 表2:恒河猴白蛋白特异性核酸酶设计

SBS #, 靶	设计						
	F1	F2	F3	F4	F5	F6	
[0215]	<b>SBS#36806 (恒河猴)</b> ttAGGGACA GTTATGAAtt caatcttca (SEQ ID NO:1)	QSGNLA R (SEQ ID NO:2)	LMQNRR Q (SEQ ID NO:3)	LKHHLT D (SEQ ID NO:4)	DRSNLSR (SEQ ID NO:5)	RSDHLTQ (SEQ ID NO:6)	N/A
	<b>SBS#35396 (人/恒河猴)</b> ccTATCCATT GCACTATGC Tttatttaa (SEQ ID NO:7)	QSSDLR (SEQ ID NO:8)	LKWNLRT (SEQ ID NO:9)	DQSNLR A (SEQ ID NO:10)	RPYTLRL (SEQ ID NO:11)	QSSDLR (SEQ ID NO:8)	HRS NLN K (SEQ ID NO:12)
	<b>SBS#37804 (恒河猴)</b> ttAGGGACA GTTATGAAtt caatcttca (SEQ ID NO:1)	QSGNLA R (SEQ ID NO:2)	LMQNRR Q (SEQ ID NO:3)	LAHHLV E (SEQ ID NO:13)	DRSNLSR (SEQ ID NO:5)	RSDHLTQ (SEQ ID NO:6)	N/A

[0216] 表2示出结合其靶位点的所有设计。

[0217] 示例性研究以ZFN对36806和35396(对2)进行如下。将2至4岁龄、具有3至4.6kg体重的恒河猴(预定繁殖)针对rAAV 2/6和2/8中和抗体的存在、白蛋白基因座的基因型和正常血清化学和血液学来预先筛选。将动物群居豢养(相同剂量组的多达3只动物豢养在一起)。载体施用通过在1mL/min的速率下IV输注至外周静脉,历时约10-30分钟范围内的给药持续时间来进行(对于研究1,各自为10mL,对于研究2,各自为29mL)。猴子在整个研究中针对死亡/假死、例行临床观察、笼侧观察和食物消耗(每日)、体重(研究前和每周)、临床病理学包括肝酶水平(ALT和AST)、临床化学和血液学和使用例行方法的凝血来进行评估。进行肝活检并且组织针对组织病理学和rAAV载体的药代动力学来进行检查,并且通过miSEQ(Illumina)针对基因修饰来评估并且通过蛋白质印迹分析针对ZFN表达来评估。抗药物抗体分析在整个研究中进行并且PBMC从全血分离以进行ELISpot分析(参见上述)。在研究结束时,对于组织评估进行肉眼和微观病理学。

[0218] 研究#1:向恒河猴施用AAV2/8载体中的白蛋白靶向ZFN,如美国公布号20130177983所描述。在此项研究中,还使用野生型Fok1裂解结构域的变体,其中序列已经根据标准技术(DNA 2.0)针对哺乳动物表达来优化。剂量组和动物ID在以下表3中示出。

[0219] 表3:NHP研究#1;剂量组

组	描述	剂量	动物 ID
1	阴性对照	无	1001
2	ZFN 对 2, Fok1 WT, 仅 ZFN-	1.5 e+13 每个 ZFN	7001
		1.5 e+13 每个 ZFN	7002
3	ZFN 对 2, 密码子优化的 Fok1 WT, 仅 ZFN	1.5 e+13 每个 ZFN	8001
	ZFN 对 2, 密码子优化的 Fok1 WT, 仅 ZFN	1.5 e+13 每个 ZFN	8002

[0221] 在第65天,对于从动物分离的脾和肠系膜淋巴结组织进行酶联免疫吸附斑点测定(ELISPOT, 参见Markusic等人(2013), EMBO Mol Med 5:1698–1709)并且如图8示出,没有针对动物中的AAV8衣壳或ZFN转基因引起免疫响应。动物7001和7002(图8,图A和C)以及动物8001和8002(图8,图B和C)都对于抗体响应呈阴性。

[0222] 研究#2:在单独研究中,各自两只动物的三个群组对于F9转基因的ZFN介导插入恒河猴的白蛋白基因座中进行评估。参见图11。

[0223] 示例性结果使用如上所述对2来获得,其包含AAV2/8或AAV2/6载体中的如指示的野生型FokI核酸酶裂解结构域(标记为“Fok1WT”)或工程化结构域(标记为“Fok1eHiFi”,参见美国专利号8,623,618)。向接受含有供体的AAV的动物给予AAV2/8载体中的F9供体(具有白蛋白同源性臂)。参见以下表4。在表中,“高”和“低”剂量是指给予AAV的总量。

[0224] 表4:NHP研究#2;剂量组

组	描述	AAV 血清型	剂量	动物 ID
1	阴性对照	-	无	1001
6	ZFN 对 2, Fok1-WT, + F9 供体(1:5 定量 ZFN:供体), 高剂量	AAV2/6	1.5e+13 每个 ZFN, 1.5e+14 供体	6101
			1.5e+13 每个 ZFN, 1.5e+14 供体	6102
7	仅供体	-	1.5e+14	7001

[0226] 只接受ZFN(没有供体)的动物在施用后第14天展示稳健裂解(0.4–4.1%)。

[0227] 对于样品进行蛋白质印迹分析以评估ZFN表达。另外,在已接受ZFN和供体载体的动物的血浆中检测到F.IX蛋白质的表达。在ZFN和供体存在下,血浆中的hFIX水平可检测并且随着时间的推移而增加。

[0228] 总之,这些数据示出因子8或因子9供体靶向插入白蛋白基因座增加活性水平,包括多至甲型血友病小鼠中的正常FVIII的50%。此外,优化供体和ZFN构建体(例如,密码子优化、包含糖基化位点和/或在分开载体上施用ZFN),AAV剂量可减少,同时保持转基因活性。事实上,编码每个白蛋白特异性ZFN的AAV与hF.IX供体的单次静脉内共注射导致可检测DNA裂解和恒河猴血浆中的hF.IX表达。

[0229] 实施例3:人血白蛋白特异性核酸酶的设计、构建和一般表征

[0230] 靶向白蛋白的核酸酶(例如,ZFN、TALEN、CRISPR/Cas)描述于美国专利公布号20130177983和20130177960和美国申请号14/278,903)。对于这些实验,包含ZFP(有效地连接至工程化裂解结构域)的ZFN用于裂解人细胞中的内源性白蛋白基因座。人血白蛋白特异性对在以下表5中示出。表5中的所有核酸酶结合至其靶。

[0231] 表5:人白蛋白特异性核酸酶设计

[0232]	SBS #, 靶	设计					
	人白蛋白特异性 ZFN						
		F1	F2	F3	F4	F5	F6

[0233]	SBS#35396 (人/恒河猴) ccTATCCATTG CACTATGCTtta tttaa(SEQ ID NO:7)	QSSDLR (SEQ ID NO:8)	LKWNLRT(SEQ ID NO:9)	DQSNLR A (SEQ ID NO:10)	RPYTLR L (SEQ ID NO:11)	QSSDLR (SEQ ID NO:8)	HRSNL NK (SEQ ID NO:12)
	SBS#39330 (人) ttTGGGATAGT TATGAAttcaatct tca (SEQ ID NO:103)	QSGNLAR (SEQ ID NO:2)	LKQNLC M (SEQ ID NO:104)	WQSNL QN (SEQ ID NO:105)	TSGNLT R (SEQ ID NO:106)	RQSHLC L (SEQ ID NO:107)	NA
	SBS#43116 (人) ccTATCCATTG CACTATGctttttt aa (SEQ ID NO:7)	LKWNLRT (SEQ ID NO:9)	DQSNLR A (SEQ ID NO:10)	RNFSLT M (SEQ ID NO:15)	QSSTLD T (SEQ ID NO:108)	HRSNLN K (SEQ ID NO:12)	NA
	SBS#47171 (人) ttTGGGATAGT TATGAAttcaatct tca (SEQ ID NO:103)	QSGNLS R (SEQ ID NO:109)	LKQNLC M (SEQ ID NO:104)	WADNL QN (SEQ ID NO:110)	TSGNLT R (SEQ ID NO:106)	RQSHLC L (SEQ ID NO:107)	NA
	SBS#47931 (人) ccTATCCATTG CACTATGCTtta tttaa (SEQ ID NO:7)	TPQLLDR (SEQ ID NO:14)	LKWNLRT (SEQ ID NO:9)	DQSNLN A (SEQ ID NO:111)	RNFSLT M (SEQ ID NO:15)	LRHDL DR (SEQ ID NO:16)	HRSNL NK (SEQ ID NO:12)
	SBS#47863 (人) ttTGGGATAGT TATGAAttcaatct tca (SEQ ID NO:103)	QSGNLAR (SEQ ID NO:2)	LIQYLQS (SEQ ID NO:112)	WADNL QN (SEQ ID NO:110)	TSGNLT R (SEQ ID NO:106)	RQSHLS L (SEQ ID NO:113)	NA
	SBS#47079 (人) ccTATCCATTG CACTATGCTtta tttaa (SEQ ID NO:7)	TPQLLD R (SEQ ID NO:14)	LKWNLRT (SEQ ID NO:9)	DQSNLR A (SEQ ID NO:10)	RNFSLT M (SEQ ID NO:15)	LRHDLD R (SEQ ID NO:16)	HRSNL NK (SEQ ID NO:12)
	SBS#47192 (人) ttTGGGATAGT TATGAAttcaatct tca (SEQ ID NO:103)	QSGNLAR (SEQ ID NO:2)	LIQYLQS (SEQ ID NO:112)	WADNL QN (SEQ ID NO:110)	TSGNLT R (SEQ ID NO:106)	RQSHLC L (SEQ ID NO:107)	NA
	SBS#47898 (人) ccTATCCATTG CACTATGCTtta tttaa (SEQ ID NO:7)	TPQLLDR (SEQ ID NO:14)	LKHNLRT (SEQ ID NO:114)	DQSNL NA (SEQ ID NO:111)	RNFSLT M (SEQ ID NO:15)	LRHDL DR (SEQ ID NO:16)	HRSNL NK (SEQ ID NO:12)
	SBS#47169	TPQLLD	LKWNLRT	DQSNLR	RNFSLT	LRHDL	HRSNL

(人) ccTATCCATTG CACTATGCTtta tttaa (SEQ ID NO:7)	R (SEQ ID NO:14)	T(SEQ ID NO:9)	A(SEQ ID NO:10)	M(SEQ ID NO:15)	DR(SEQ ID NO:16)	NK(SE Q ID NO:12)
SBS#47864 (人) )ttTGGGATAGT TATGAAttcaatct tca(SEQ ID NO:5)	QSGNLAR (SEQ ID NO:2)	LIQYLQS (SEQ ID NO:112)	WQSNL QN (SEQ ID NO:105)	TSGNLT R(SEQ ID NO:106)	RQSHLC L(SEQ ID NO:107)	N/A
SBS#40477 (人) ccTATCCATTG CACTATGCTtta tttaa (SEQ ID NO:7)		LKHNL T(SEQ ID NO:114)	LKHNL T(SEQ ID NO:114)	RPYTLR L(SEQ ID NO:11)	LRPDLE R(SEQ ID NO:126)	HRSNL NK(SE Q ID NO:12)

[0235] 在这些实验中,ZFN以mRNA形式转染至细胞中并且经由通过标准方法的BTX核转染引入。ZFN mRNA的浓度随着实验而改变。NHEJ活性通过根据在本领域中已知的方法进行的MiSeq分析(Illumina)来测量。

[0236] 为了在人原代肝细胞中测试,使用编码此对中的每个ZFN的50ng RNA。结果如下示出(表6)并且证明所有ZFN对具有活性。为了在人HepG2中测试,使用编码此对中的每个ZFN的100ng RNA。结果如下示出(表7)并且证明所有ZFN对具有活性。人K562细胞还使用75ng(表8)重复两次测试并且证明所述对是有活性。

[0237] 表6:人原代肝细胞

左 ZFN						插入 和缺 失%
	47162	47171	47192	47863	47864	
40477	8.9	6.5	6.1	11.7	14.0	
47079	8.2	13.0	12.3	13.7	13.7	
47169	11.2	9.6	11.6	20.3	14.1	
47898	12.1	10.8	11.1	16.8	14.6	
47931	10.1	11.8	11.5	17.2	14.7	

[0239] 表7:人HepG2细胞

左 ZFN						插入 和缺 失%
	47162	47171	47192	47863	47864	
40477	20.3	21.3	23.7	22.2	18.4	
47079	20.8	25.8	22.9	23.5	16.0	
47169	21.0	22.2	22.2	21.6	17.7	
47898	21.5	22.0	13.9	22.4	14.3	
47931	23.1	19.4	23.1	20.9	21.3	

[0242] 表8:K562细胞

ZFN 对	1	2	插入和缺失 平均%
[0243]	47171:47931	73.9	75.7
	47171:47079	77.1	81.7
	47171:47898	75.3	68.6
	47863:47931	76.9	80.9
	47863:47079	86.8	86.7
	47863:47898	91.1	92.7
	47192:47931	69.9	76.1
	47192:47079	80	75.9
	47192:47898	70.1	72.3
			71.2

[0244] 实施例5:整合使用人血白蛋白特异性ZFN的LSD供体。

[0245] 在这些实验中,IDS或IDUA cDNA供体经由AAV2/8颗粒递送,其中cDNA转基因包含侧接切割位点的白蛋白区域的同源性臂。在这些供体构建体中,治疗性基因由与白蛋白基因同源的序列侧接。在转基因的5',供体构建体都含有与鼠白蛋白内含子1同源的序列,而在基因的3',构建体含有与鼠白蛋白内含子1-外显子2边界同源的序列(如美国专利公布2014-0017212所描述)。

[0246] 为了整合IDS或IDUA cDNA转基因并且测定其表达,呈mRNA形式的白蛋白特异性锌指核酸酶转染至人HepG2/C3a细胞中。简单地说,100,000个细胞通过标准方法由病毒递送来转染,zfn:zfn:供体的MOI X1000是100:100:200(“低”或“L”)或300:300:600(“高”或“H”)。表达通过测定细胞上清液中的由转基因编码的蛋白质的酶活性来或通过6天之后对于细胞球团进行蛋白质印迹来分析。图13示出的数据证明供体IDS和IDUA在细胞中的表达。

[0247] 所有本文提到的专利、专利申请以及专利公布特此以引用的方式全部并入。

[0248] 虽然为了清楚理解,比较详细地用说明和实施例的方式提供了公开内容,但本领域技术人员明显可以实施不背离本公开内容精神或范围的各种改变和改进。因此,前面的描述和实施例不应理解为具有限制性。

## 序列表

<110>桑格摩生物科学股份有限公司 (SANGAMO BIOSCIENCES, INC.)

<120> 基因组工程化的方法和组合物

<130> 8325-0115.40

<140> PCT/US2014/069352

<141> 2014-12-09

<150> 61/943,884

<151> 2014-02-24

<150> 61/913,838

<151> 2013-12-09

<160> 41

<170> PatentIn 版体 3.5

[0001]

<210> 1

<211> 28

<212> DNA

<213> 恒河猴

<400> 1

ttagggacag ttatgaattc aatcttca 28

<210> 2

<211> 7

<212> PRT

<213> 人工序列

<220>

<223> 人工序列的描述：合成肽

<400> 2

Gln Ser Gly Asn Leu Ala Arg

1

5

<210> 3

<211> 7

<212> PRT

<213> 人工序列

<220>

<223> 人工序列的描述：合成肽

<400> 3

Leu Met Gln Asn Arg Asn Gln

1 5

<210> 4

<211> 7

<212> PRT

<213> 人工序列

<220>

<223> 人工序列的描述：合成肽

[0002]

<400> 4

Leu Lys His His Leu Thr Asp

1 5

<210> 5

<211> 7

<212> PRT

<213> 人工序列

<220>

<223> 人工序列的描述：合成肽

<400> 5

Asp Arg Ser Asn Leu Ser Arg

1 5

<210> 6

<211> 7

<212> PRT

<213> 人工序列

<220>

<223> 人工序列的描述：合成肽

<400> 6

Arg Ser Asp His Leu Thr Gln

1 5

<210> 7

<211> 28

<212> DNA

<213> 未知

<220>

<223> 未知的描述：

人或恒河猴寡核苷酸

[0003] <400> 7

cctatccatt gcactatgt ttatcaa

28

<210> 8

<211> 7

<212> PRT

<213> 人工序列

<220>

<223> 人工序列的描述：合成肽

<400> 8

Gln Ser Ser Asp Leu Ser Arg

1 5

<210> 9

<211> 7

<212> PRT

<213> 人工序列

<220>

<223> 人工序列的描述：合成肽

<400> 9

Leu Lys Trp Asn Leu Arg Thr

1 5

<210> 10

<211> 7

<212> PRT

<213> 人工序列

<220>

<223> 人工序列的描述：合成肽

<400> 10

Asp Gln Ser Asn Leu Arg Ala

1 5

[0004]

<210> 11

<211> 7

<212> PRT

<213> 人工序列

<220>

<223> 人工序列的描述：合成肽

<400> 11

Arg Pro Tyr Thr Leu Arg Leu

1 5

<210> 12

<211> 7

<212> PRT

<213> 人工序列

<220>

<223> 人工序列的描述：合成肽

<400> 12  
His Arg Ser Asn Leu Asn Lys  
1 5

<210> 13  
<211> 7  
<212> PRT  
<213> 人工序列

<220>  
<223> 人工序列的描述：合成肽

<400> 13  
Leu Ala His His Leu Val Glu  
1 5

[0005] <210> 14  
<211> 7  
<212> PRT  
<213> 人工序列

<220>  
<223> 人工序列的描述：合成肽

<400> 14  
Thr Pro Gln Leu Leu Asp Arg  
1 5

<210> 15  
<211> 7  
<212> PRT  
<213> 人工序列

<220>  
<223> 人工序列的描述：合成肽

<400> 15

Arg Asn Phe Ser Leu Thr Met

1 5

<210> 16

<211> 7

<212> PRT

<213> 人工序列

<220>

<223> 人工序列的描述：合成肽

<400> 16

Leu Arg His Asp Leu Asp Arg

1 5

<210> 17

<211> 28

[0006] <212> DNA

<213> 智人

<400> 17

tttggatag ttatgaattc aatcttca

28

<210> 18

<211> 7

<212> PRT

<213> 人工序列

<220>

<223> 人工序列的描述：合成肽

<400> 18

Leu Lys Gln Asn Leu Cys Met

1 5

<210> 19

<211> 7

<212> PRT

<213> 人工序列

<220>

<223> 人工序列的描述：合成肽

<400> 19

Trp Gln Ser Asn Leu Gln Asn

1 5

<210> 20

<211> 7

<212> PRT

<213> 人工序列

<220>

<223> 人工序列的描述：合成肽

[0007]

<400> 20

Thr Ser Gly Asn Leu Thr Arg

1 5

<210> 21

<211> 7

<212> PRT

<213> 人工序列

<220>

<223> 人工序列的描述：合成肽

<400> 21

Arg Gln Ser His Leu Cys Leu

1 5

<210> 22

<211> 7

<212> PRT

<213> 人工序列

&lt;220&gt;

&lt;223&gt; 人工序列的描述：合成肽

&lt;400&gt; 22

Gln Ser Ser Thr Leu Asp Thr

1 5

&lt;210&gt; 23

&lt;211&gt; 7

&lt;212&gt; PRT

&lt;213&gt; 人工序列

&lt;220&gt;

&lt;223&gt; 人工序列的描述：合成肽

&lt;400&gt; 23

Gln Ser Gly Asn Leu Ser Arg

1 5

[0008]

&lt;210&gt; 24

&lt;211&gt; 7

&lt;212&gt; PRT

&lt;213&gt; 人工序列

&lt;220&gt;

&lt;223&gt; 人工序列的描述：合成肽

&lt;400&gt; 24

Trp Ala Asp Asn Leu Gln Asn

1 5

&lt;210&gt; 25

&lt;211&gt; 7

&lt;212&gt; PRT

&lt;213&gt; 人工序列

&lt;220&gt;

<223> 人工序列的描述：合成肽

<400> 25

Asp Gln Ser Asn Leu Asn Ala

1 5

<210> 26

<211> 7

<212> PRT

<213> 人工序列

<220>

<223> 人工序列的描述：合成肽

<400> 26

Leu Ile Gln Tyr Leu Gln Ser

1 5

[0009]

<210> 27

<211> 7

<212> PRT

<213> 人工序列

<220>

<223> 人工序列的描述：合成肽

<400> 27

Arg Gln Ser His Leu Ser Leu

1 5

<210> 28

<211> 7

<212> PRT

<213> 人工序列

<220>

<223> 人工序列的描述：合成肽

<400> 28

Leu Lys His Asn Leu Leu Thr

1 5

<210> 29

<211> 385

<212> PRT

<213> 人工序列

<220>

<223> 人工序列的描述：合成肽

<400> 29

Met Asp Tyr Lys Asp His Asp Gly Asp Tyr Lys Asp His Asp Ile Asp

1 5 10 15

Tyr Lys Asp Asp Asp Asp Lys Met Ala Pro Lys Lys Lys Arg Lys Val

[0010] 20 25 30

Gly Ile His Gly Val Pro Ala Ala Met Ala Glu Arg Pro Phe Gln Cys

35 40 45

Arg Ile Cys Met Arg Lys Phe Ala Thr Ser Gly Ser Leu Thr Arg His

50 55 60

Thr Lys Ile His Thr Gly Glu Lys Pro Phe Gln Cys Arg Ile Cys Met

65 70 75 80

Arg Asn Phe Ser Arg Ser Asp Ala Leu Ser Thr His Ile Arg Thr His

85 90 95

Thr Gly Glu Lys Pro Phe Ala Cys Asp Ile Cys Gly Arg Lys Phe Ala

100 105 110

Gln Ser Ala Thr Arg Thr Lys His Thr Lys Ile His Thr His Pro Arg

115 120 125

Ala Pro Ile Pro Lys Pro Phe Gln Cys Arg Ile Cys Met Arg Asn Phe

130 135 140

Ser Thr Ser Gly His Leu Ser Arg His Ile Arg Thr His Thr Gly Glu

145 150 155 160

Lys Pro Phe Ala Cys Asp Ile Cys Gly Arg Lys Phe Ala Gln Ser Gly

165 170 175

Asn Leu Ala Arg His Thr Lys Ile His Leu Arg Gly Ser Gln Leu Val

180 185 190

[0011] Lys Ser Glu Leu Glu Glu Lys Ser Glu Leu Arg His Lys Leu Lys

195 200 205

Tyr Val Pro His Glu Tyr Ile Glu Leu Ile Glu Ile Ala Arg Asn Ser

210 215 220

Thr Gln Asp Arg Ile Leu Glu Met Lys Val Met Glu Phe Phe Met Lys

225 230 235 240

Val Tyr Gly Tyr Arg Gly Lys His Leu Gly Gly Ser Arg Lys Pro Asp

245 250 255

Gly Ala Ile Tyr Thr Val Gly Ser Pro Ile Asp Tyr Gly Val Ile Val

260 265 270

Asp Thr Lys Ala Tyr Ser Gly Gly Tyr Asn Leu Pro Ile Gly Gln Ala

275 280 285

Asp Glu Met Glu Arg Tyr Val Glu Glu Asn Gln Thr Arg Asp Lys His

290

295

300

Leu Asn Pro Asn Glu Trp Trp Lys Val Tyr Pro Ser Ser Val Thr Glu

305

310

315

320

Phe Lys Phe Leu Phe Val Ser Gly His Phe Lys Gly Asn Tyr Lys Ala

325

330

335

Gln Leu Thr Arg Leu Asn His Ile Thr Asn Cys Asn Gly Ala Val Leu

340

345

350

Ser Val Glu Glu Leu Leu Ile Gly Gly Glu Met Ile Lys Ala Gly Thr

355

360

365

Leu Thr Leu Glu Glu Val Arg Arg Lys Phe Asn Asn Gly Glu Ile Asn

[0012]

370

375

380

Phe

385

&lt;210&gt; 30

&lt;211&gt; 381

&lt;212&gt; PRT

&lt;213&gt; 人工序列

&lt;220&gt;

&lt;223&gt; 人工序列的描述：合成肽

&lt;400&gt; 30

Met Asp Tyr Lys Asp His Asp Gly Asp Tyr Lys Asp His Asp Ile Asp

1

5

10

15

Tyr Lys Asp Asp Asp Asp Lys Met Ala Pro Lys Lys Lys Arg Lys Val

20

25

30

Gly Ile His Gly Val Pro Ala Ala Met Ala Glu Arg Pro Phe Gln Cys

35 40 45

Arg Ile Cys Met Arg Asn Phe Ser Arg Ser Asp His Leu Ser Ala His

50 55 60

Ile Arg Thr His Thr Gly Glu Lys Pro Phe Ala Cys Asp Ile Cys Gly

65 70 75 80

Arg Lys Phe Ala Thr Lys Ser Asn Arg Thr Lys His Thr Lys Ile His

85 90 95

Thr Gly Ser Gln Lys Pro Phe Gln Cys Arg Ile Cys Met Arg Asn Phe

100 105 110

[0013] Ser Asp Arg Ser Asn Leu Ser Arg His Ile Arg Thr His Thr Gly Glu

115 120 125

Lys Pro Phe Ala Cys Asp Ile Cys Gly Arg Lys Phe Ala Trp Arg Ser

130 135 140

Ser Leu Arg Ala His Thr Lys Ile His Thr Gly Glu Lys Pro Phe Gln

145 150 155 160

Cys Arg Ile Cys Met Arg Lys Phe Ala Asp Ser Ser Asp Arg Lys Lys

165 170 175

His Thr Lys Ile His Leu Arg Gly Ser Gln Leu Val Lys Ser Glu Leu

180 185 190

Glu Glu Lys Lys Ser Glu Leu Arg His Lys Leu Lys Tyr Val Pro His

195 200 205

Glu Tyr Ile Glu Leu Ile Glu Ile Ala Arg Asn Ser Thr Gln Asp Arg

210 215 220

Ile Leu Glu Met Lys Val Met Glu Phe Phe Met Lys Val Tyr Gly Tyr

225 230 235 240

Arg Gly Lys His Leu Gly Gly Ser Arg Lys Pro Asp Gly Ala Ile Tyr

245 250 255

Thr Val Gly Ser Pro Ile Asp Tyr Gly Val Ile Val Asp Thr Lys Ala

260 265 270

Tyr Ser Gly Gly Tyr Asn Leu Pro Ile Gly Gln Ala Asp Glu Met Gln

275 280 285

[0014] Arg Tyr Val Lys Glu Asn Gln Thr Arg Asn Lys His Ile Asn Pro Asn

290 295 300

Glu Trp Trp Lys Val Tyr Pro Ser Ser Val Thr Glu Phe Lys Phe Leu

305 310 315 320

Phe Val Ser Gly His Phe Lys Gly Asn Tyr Lys Ala Gln Leu Thr Arg

325 330 335

Leu Asn Arg Lys Thr Asn Cys Asn Gly Ala Val Leu Ser Val Glu Glu

340 345 350

Leu Leu Ile Gly Gly Glu Met Ile Lys Ala Gly Thr Leu Thr Leu Glu

355 360 365

Glu Val Arg Arg Lys Phe Asn Asn Gly Glu Ile Asn Phe

370 375 380

<210> 31

<211> 28

<212> DNA

<213> 小家鼠

<400> 31

ctgaagggtgg caatggttcc tctctgct

28

<210> 32

<211> 28

<212> DNA

<213> 小家鼠

<400> 32

ttcctgtaa cgatcgaa ctggcatc

28

<210> 33

<211> 7

[0015] <212> PRT

<213> 人工序列

<220>

<223> 人工序列的描述：合成肽

<400> 33

Thr Ser Gly Ser Leu Thr Arg

1 5

<210> 34

<211> 7

<212> PRT

<213> 人工序列

<220>

<223> 人工序列的描述：合成肽

<400> 34

Arg Ser Asp Ala Leu Ser Thr

1 5

<210> 35

<211> 7

<212> PRT

<213> 人工序列

<220>

<223> 人工序列的描述：合成肽

<400> 35

Gln Ser Ala Thr Arg Thr Lys

1 5

<210> 36

<211> 7

<212> PRT

[0016] <213> 人工序列

<220>

<223> 人工序列的描述：合成肽

<400> 36

Thr Ser Gly His Leu Ser Arg

1 5

<210> 37

<211> 7

<212> PRT

<213> 人工序列

<220>

<223> 人工序列的描述：合成肽

<400> 37

Arg Ser Asp His Leu Ser Ala

1 5

<210> 38

<211> 7

<212> PRT

<213> 人工序列

<220>

<223> 人工序列的描述：合成肽

<400> 38

Thr Lys Ser Asn Arg Thr Lys

1 5

<210> 39

<211> 7

<212> PRT

<213> 人工序列

[0017] <220>

<223> 人工序列的描述：合成肽

<400> 39

Trp Arg Ser Ser Leu Arg Ala

1 5

<210> 40

<211> 8

<212> PRT

<213> 人工序列

<220>

<223> 人工序列的描述：合成肽

<400> 40

Asp Ser Ser Asp Arg Lys Lys Gln

1 5

<210> 41

<211> 7

<212> PRT

<213> 人工序列

[0018] <220>

<223> 人工序列的描述：合成肽

<400> 41

Leu Arg Pro Asp Leu Glu Arg

1 5

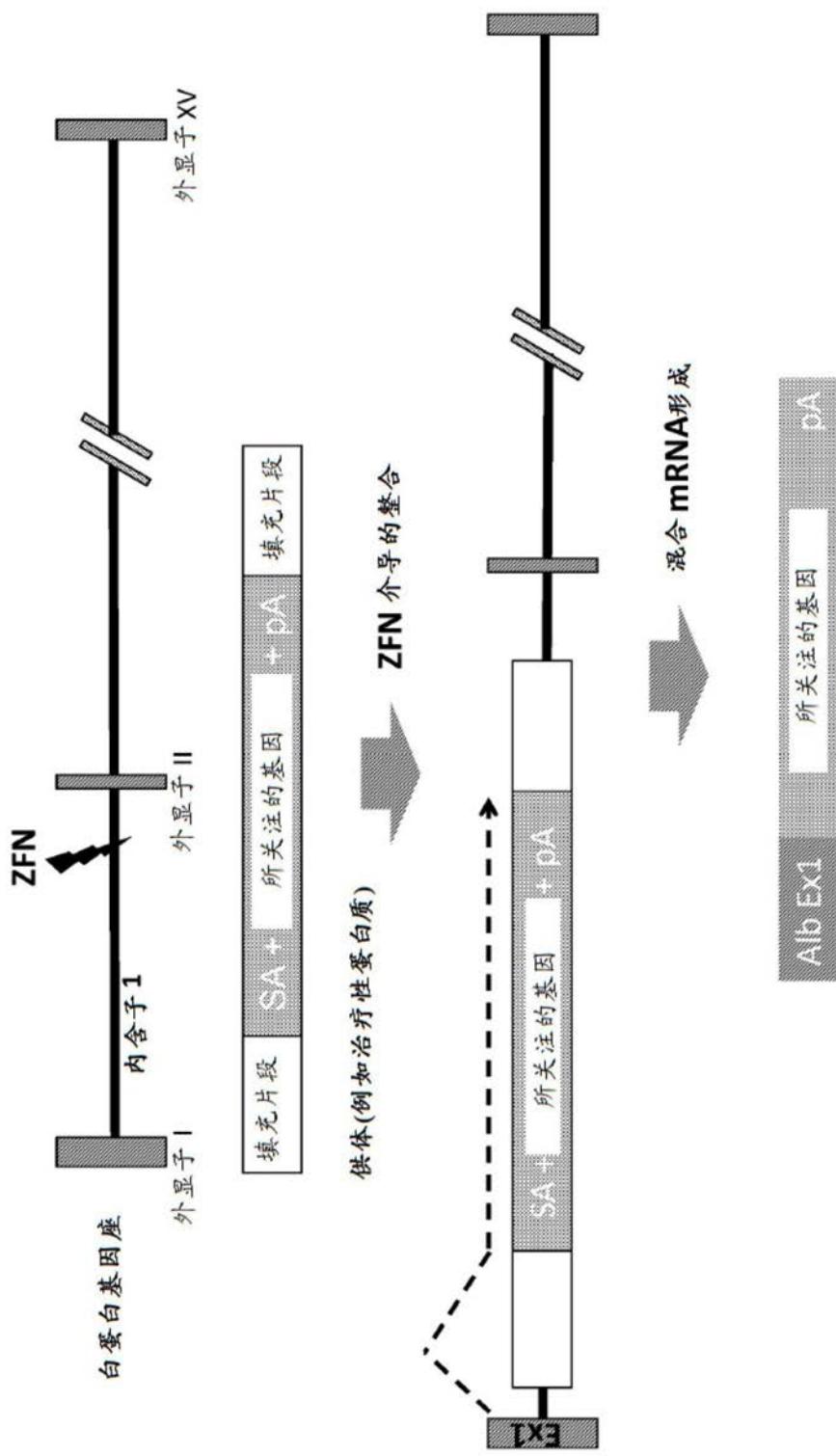
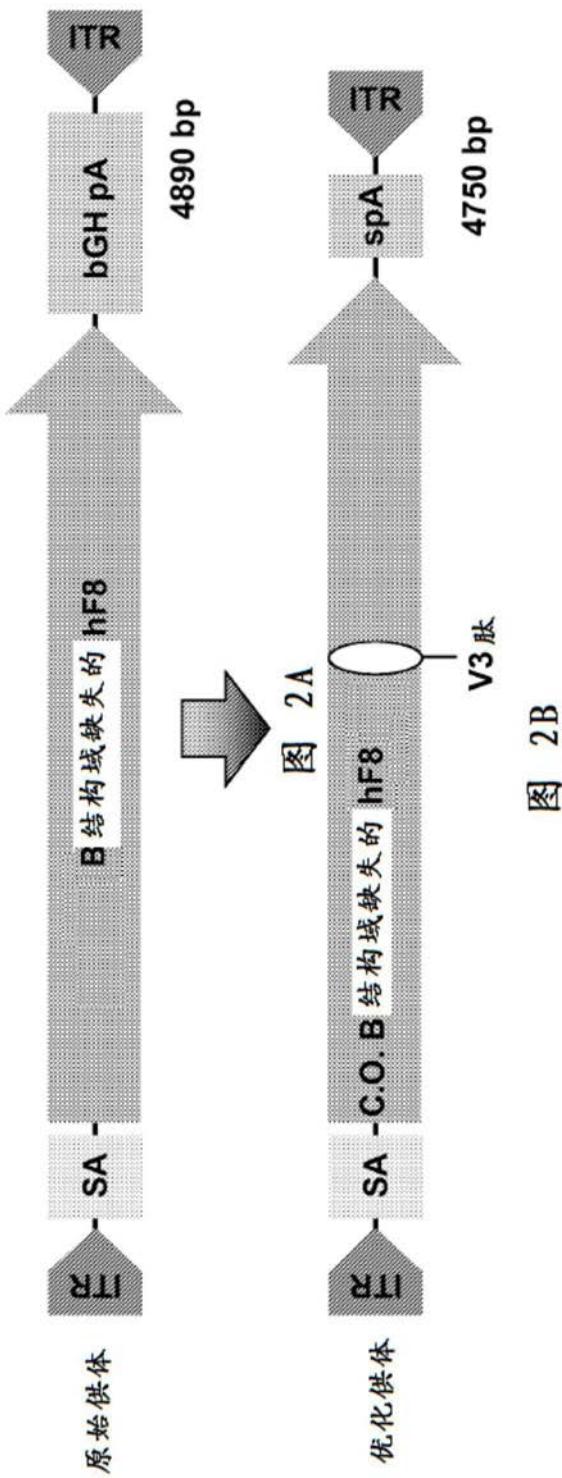


图 1



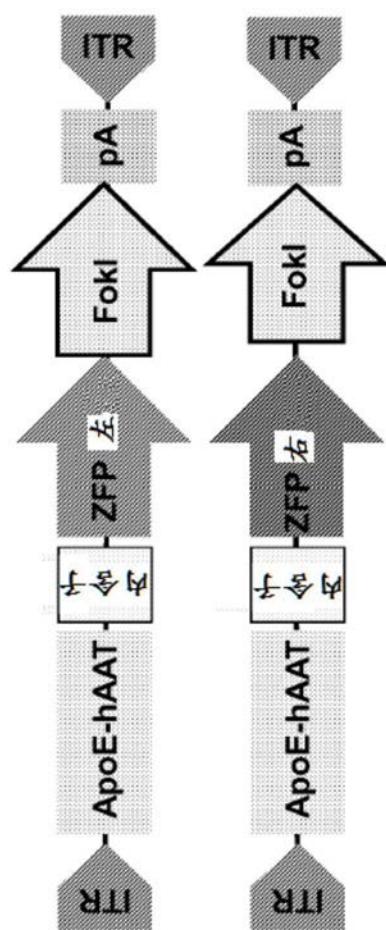


图2C



图3A

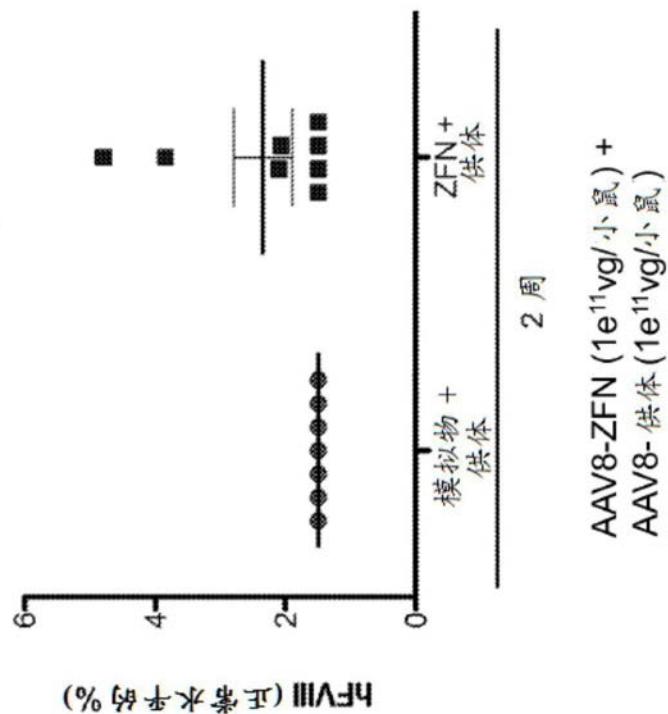


图3B

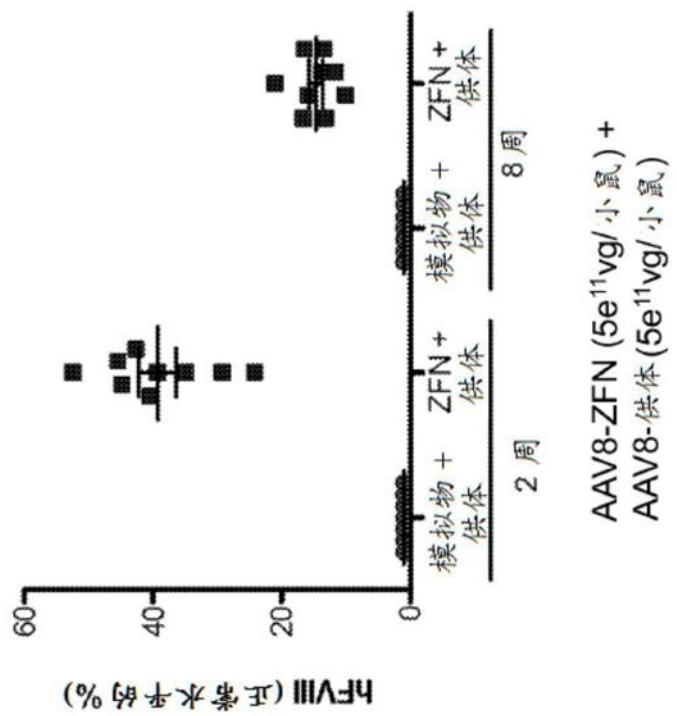


图3C

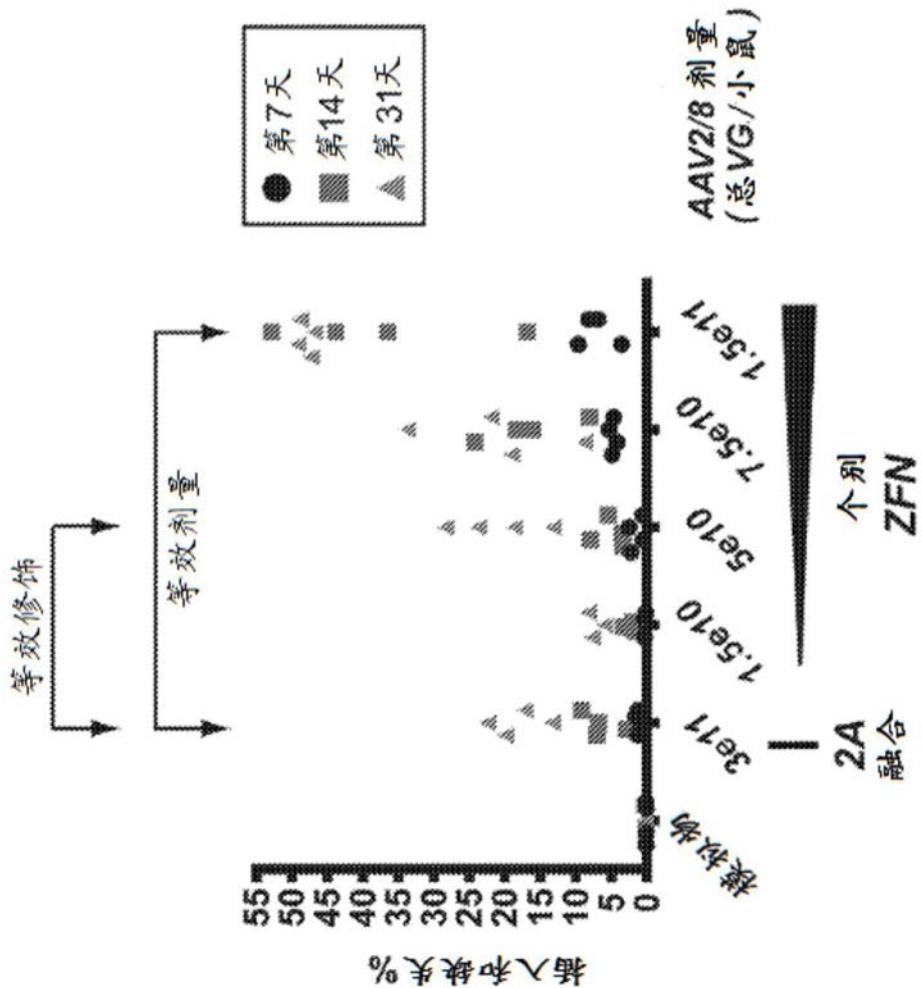


图 3D

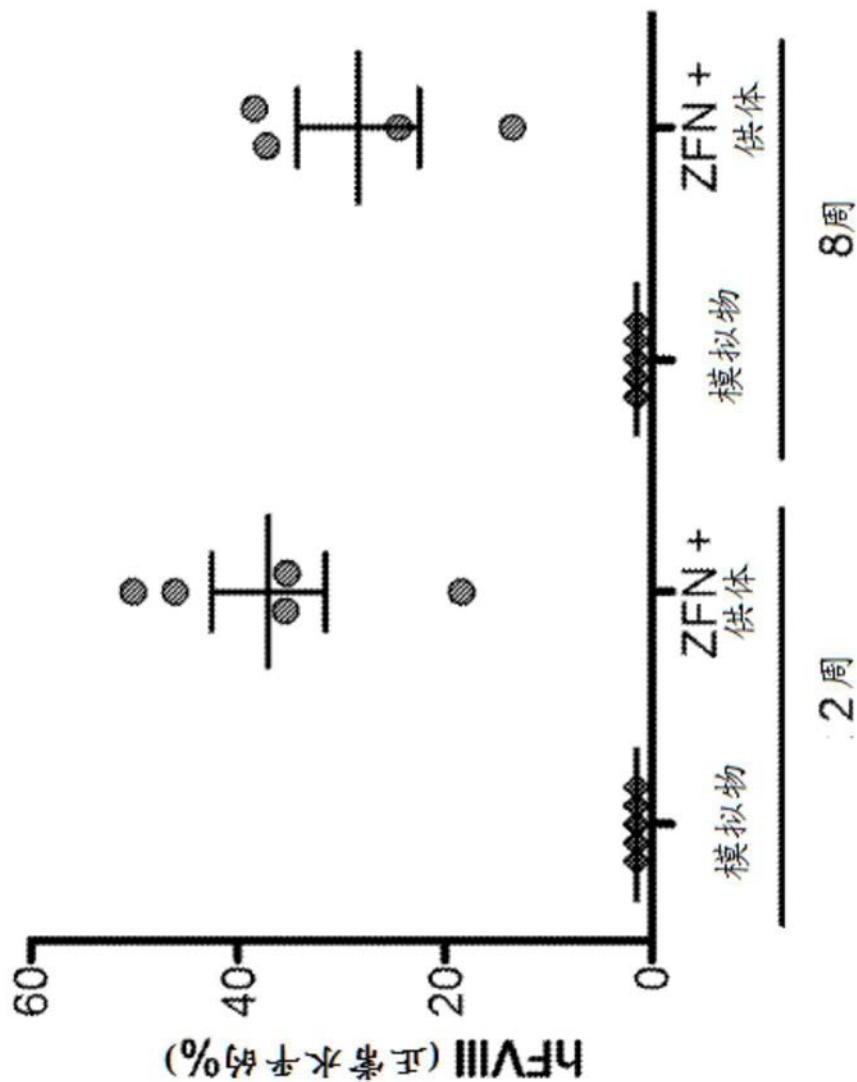


图4



**hF9 供体**

图5A

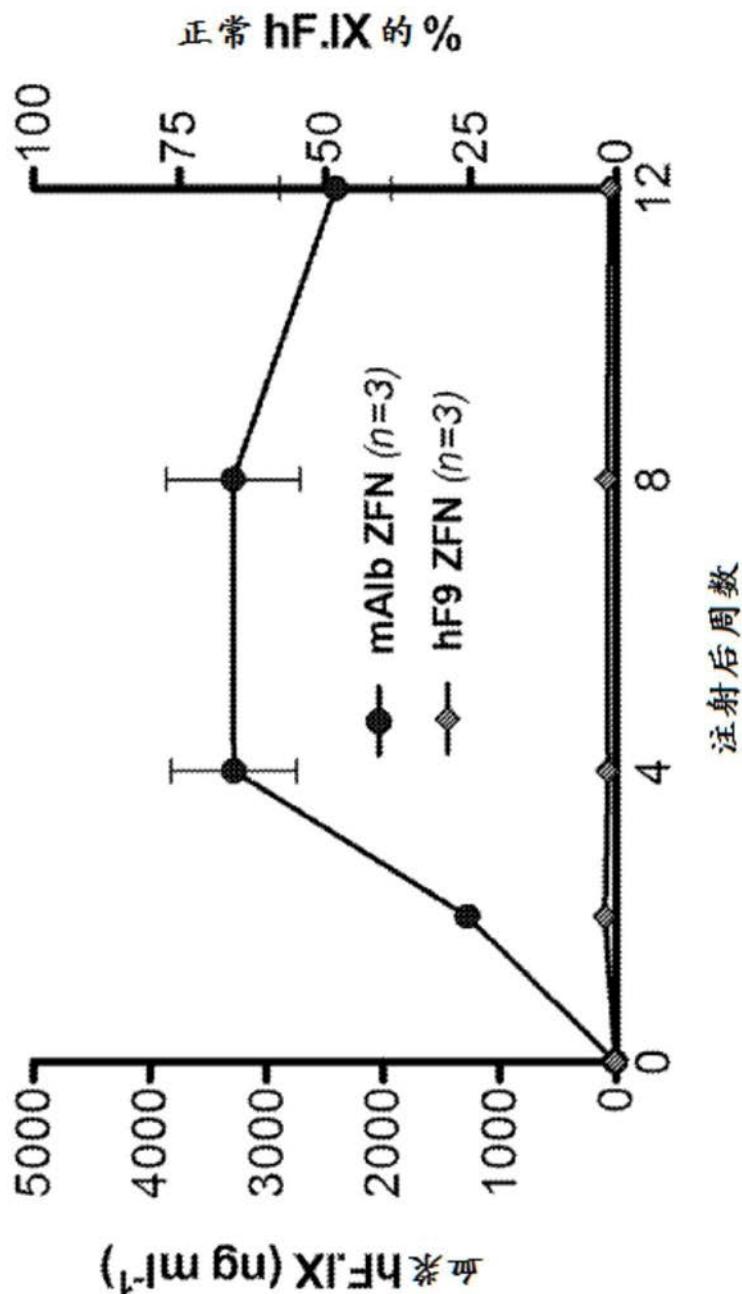


图5B

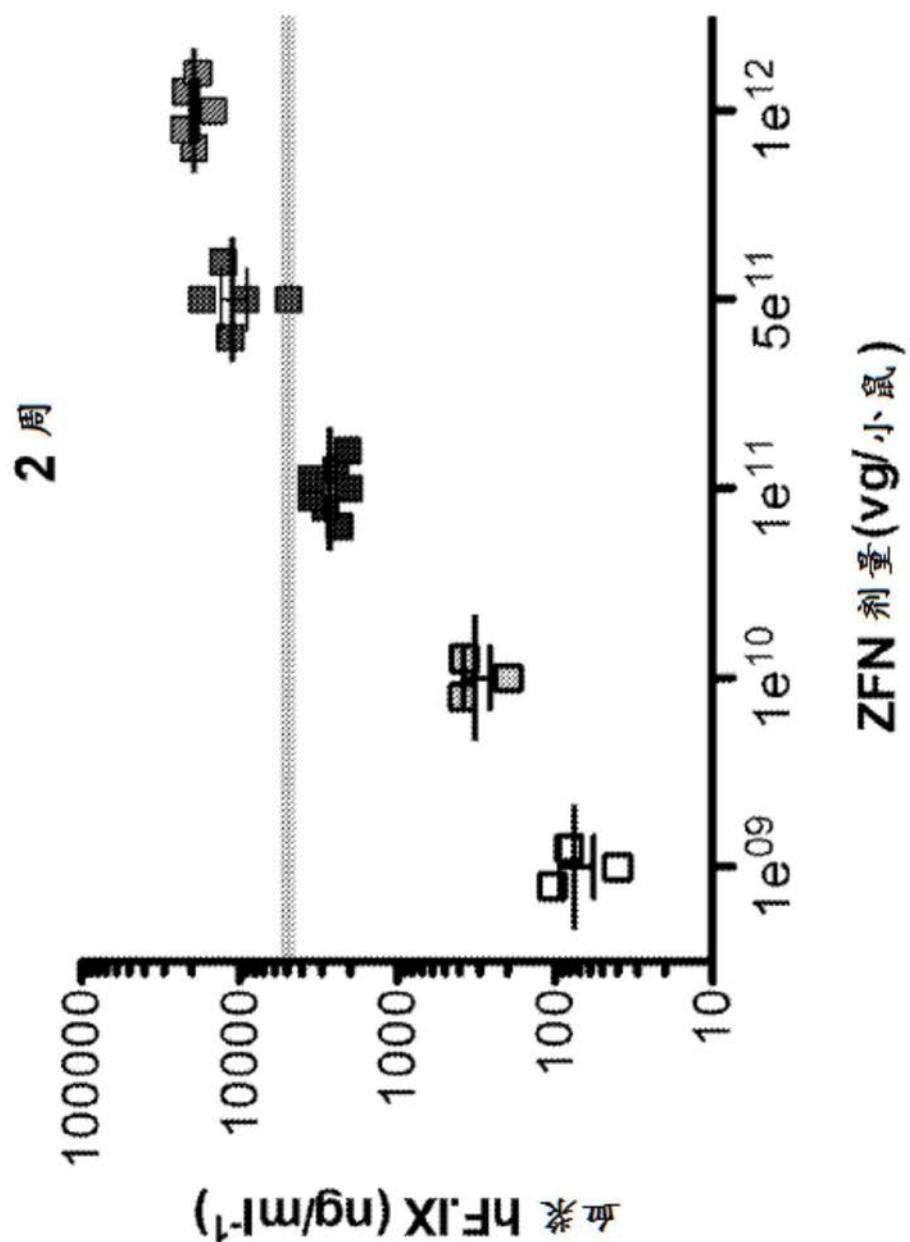
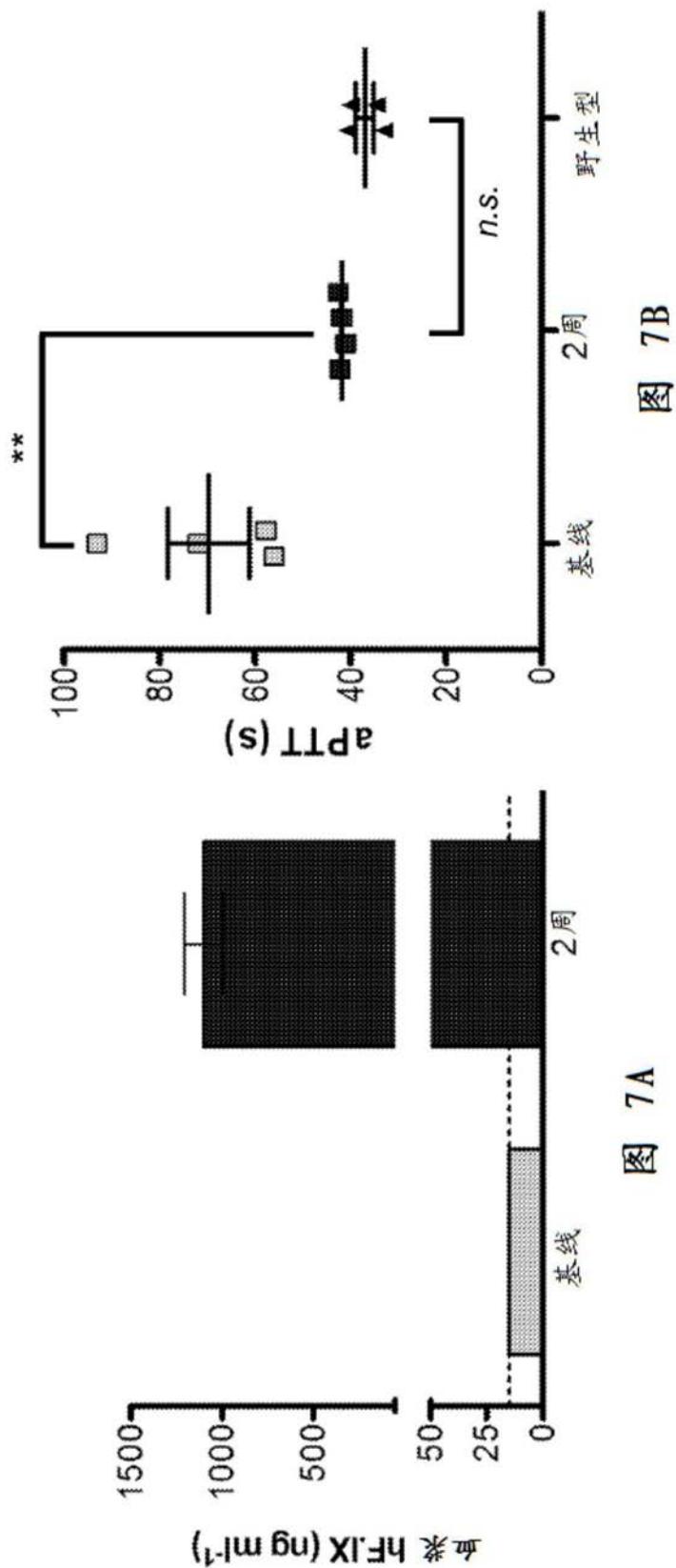


图6



*mAlb* ZFN + 供体:AAV8-ZFN (1x10<sup>11</sup> vg/小鼠) + AAV8-供体(5x10<sup>11</sup> vg/小鼠)

图 7A  
图 7B

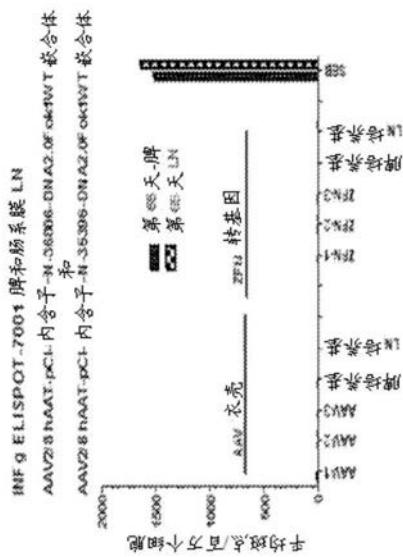


图8A

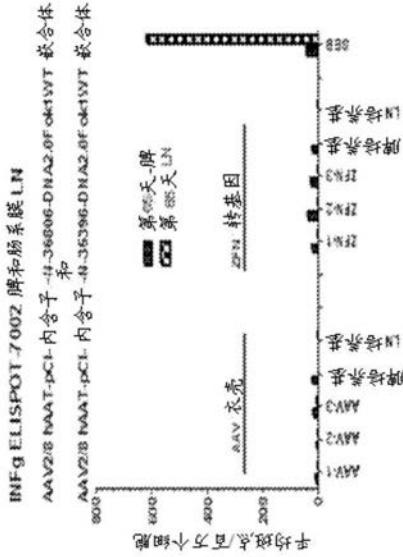


图8B

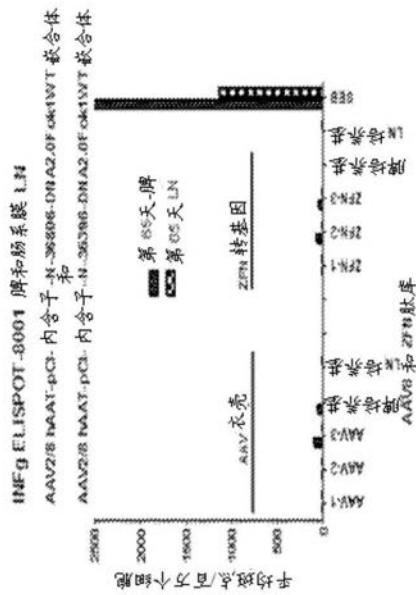


图8C

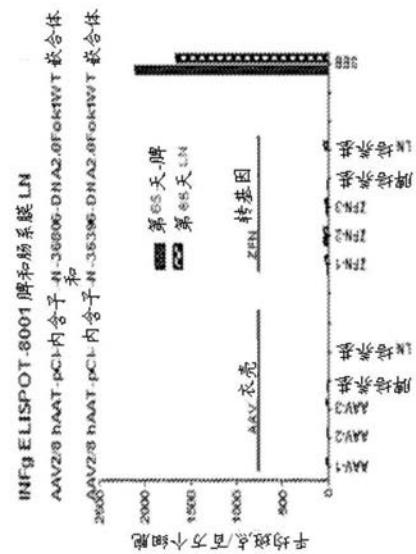


图8D

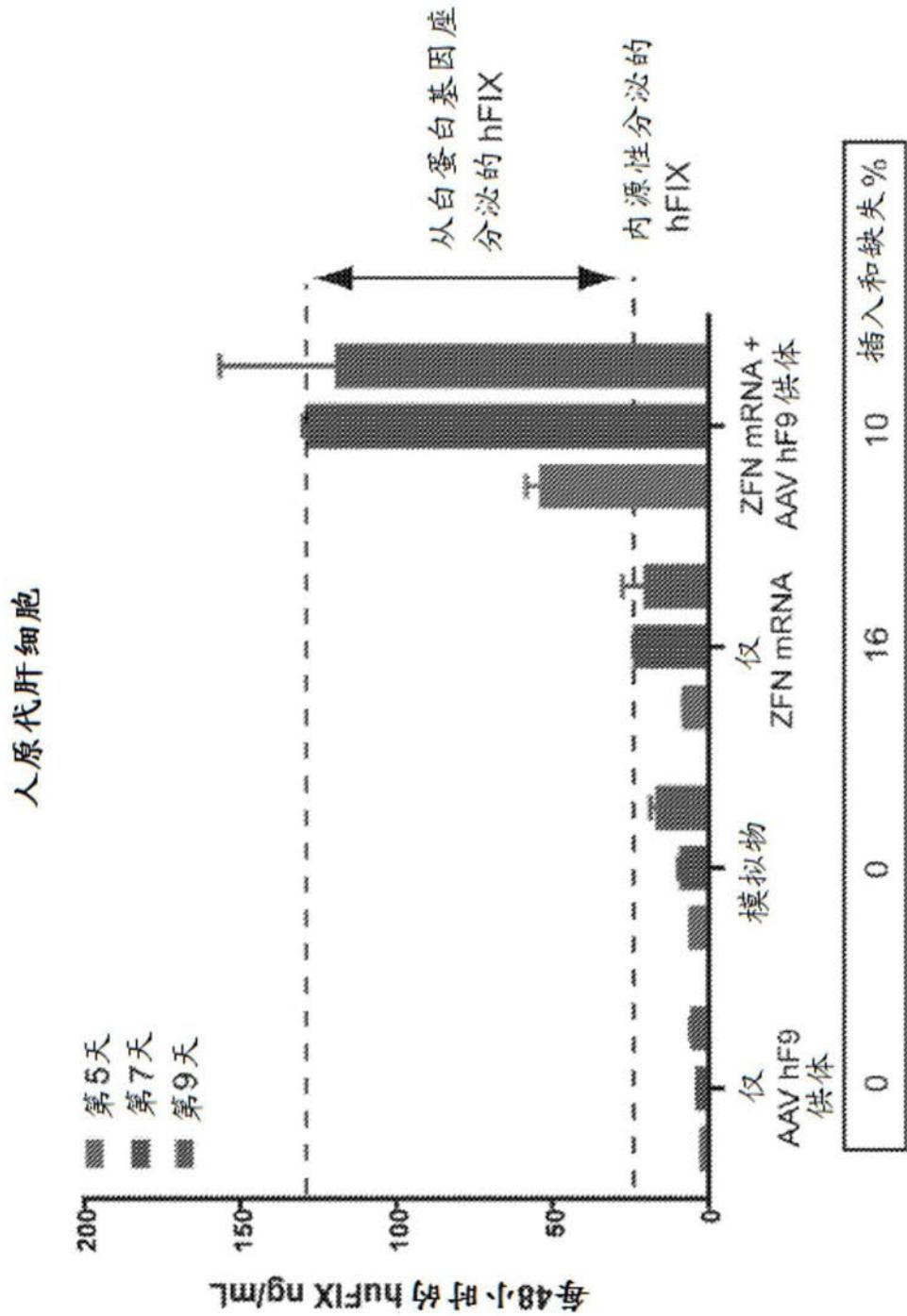


图9

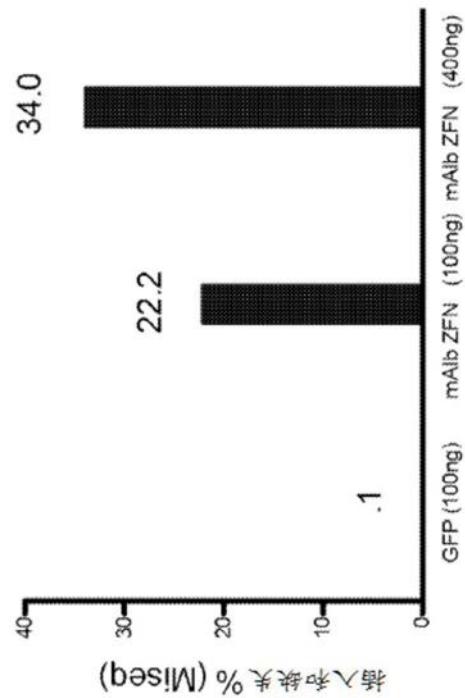


图10A

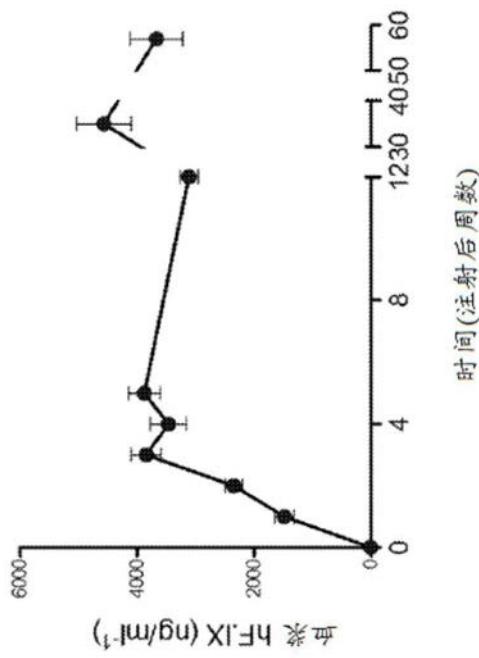


图10B

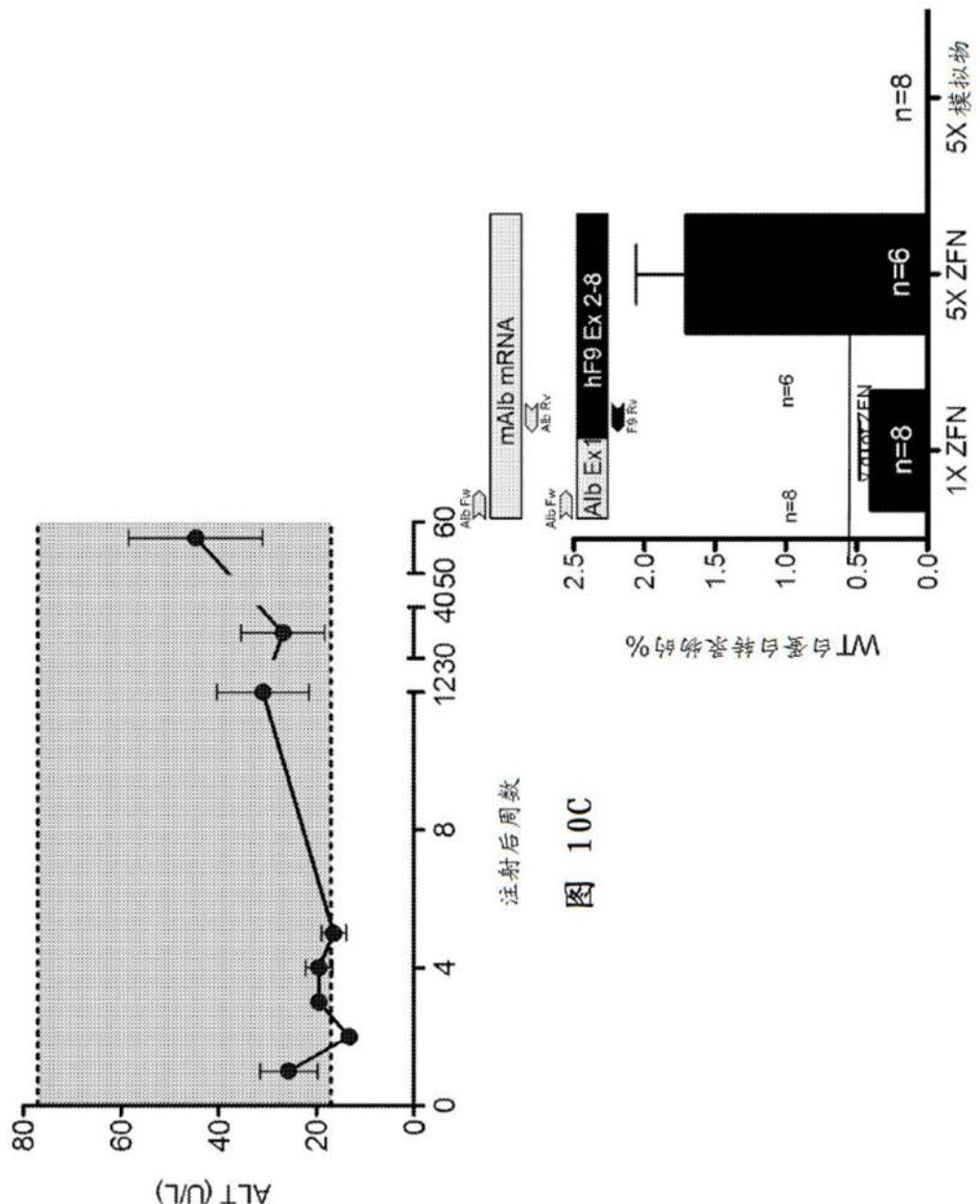


图 10D

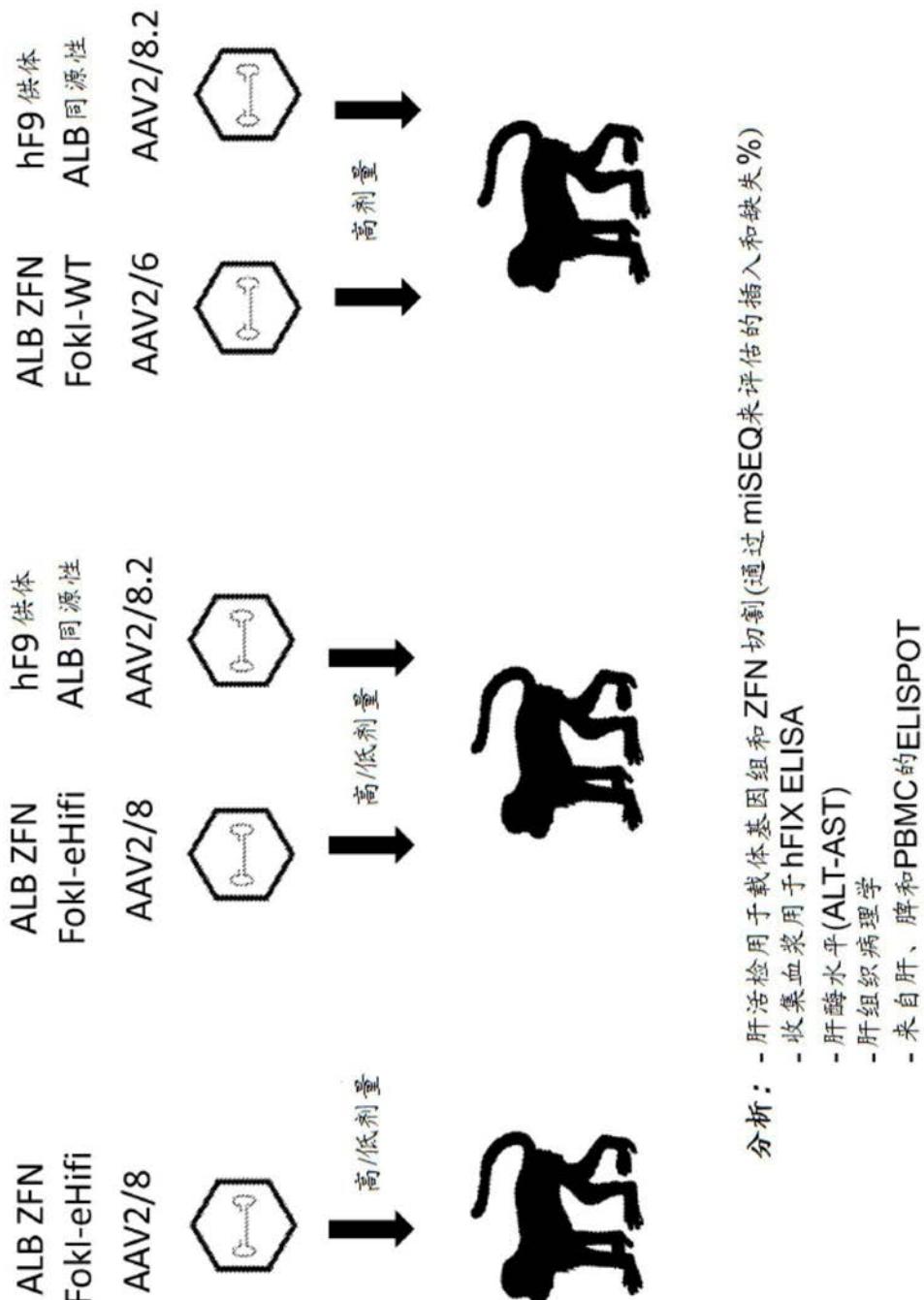


图11

MDYKDHGDYKDHDIDYKDDDKMAPKKKRKVGIHGVPAAMAERPFQCRI  
CMRKFATSGSLTRHTKIHTGEKPFQCRICMRNFSRSDALSTHIRTHTGEK  
PFACDICGRKFAQSATRTKHTKIHTHPRAPIPKPFCRIMRNFSTSGHL  
SRHIRTHTGEKPFACDICGRKFAQSGNLARHTKIHLRGSQLVKSELEEKK  
SELRHKLKYVPHEYIELIEIARNSTQDRILEMKVMEFFMKVYGYRGKHLG  
GSRKPDGAIYTGVSPIDYGIVDTKAYSGGGYNLPIGQADEMERYVEENQT  
RDKHLNPNEWWKVYPSSVTEFKFLFVGHFKGNYKAQLTRLNHTNCNGA  
VLSVEELLIGGEMIKAGTTLLEEVRRKFNNGEINF

图12A

MDYKDHGDYKDHD/DYKDDDKMAPKKKRKVGIHGVPAAMAERPFQCRI  
CMRNFSSRSDHLSAHIRTHTGEKPFAC DICGRKFATKS/NRTKHTKIHTGSQ  
KPFQCRICMRNFSDRSNLSRHIRTHTGEKPFACDICGRKFAWRSSLRAHT  
KIHTGEKPFQCRICMRKFADSSDRKKHTKIHLRGSQLVKSELEEKKSELR  
HKLKYVPHYIELIEIARNSTQDRILEMKVMEFFMKVYGYRGKHLGGSRK  
PDGAIYTVGSPIDYGIVDTKAYSGGYNLPIGQADEMQRYVKENQTRNKH  
INPNEWWWKVYPSSVTEFKFLVSGHFKGNYKAQLTRLNRKTNCNGAVLSV  
EELLIGEMIKAGTLTLEEVRRKFNNGEINF

图12B

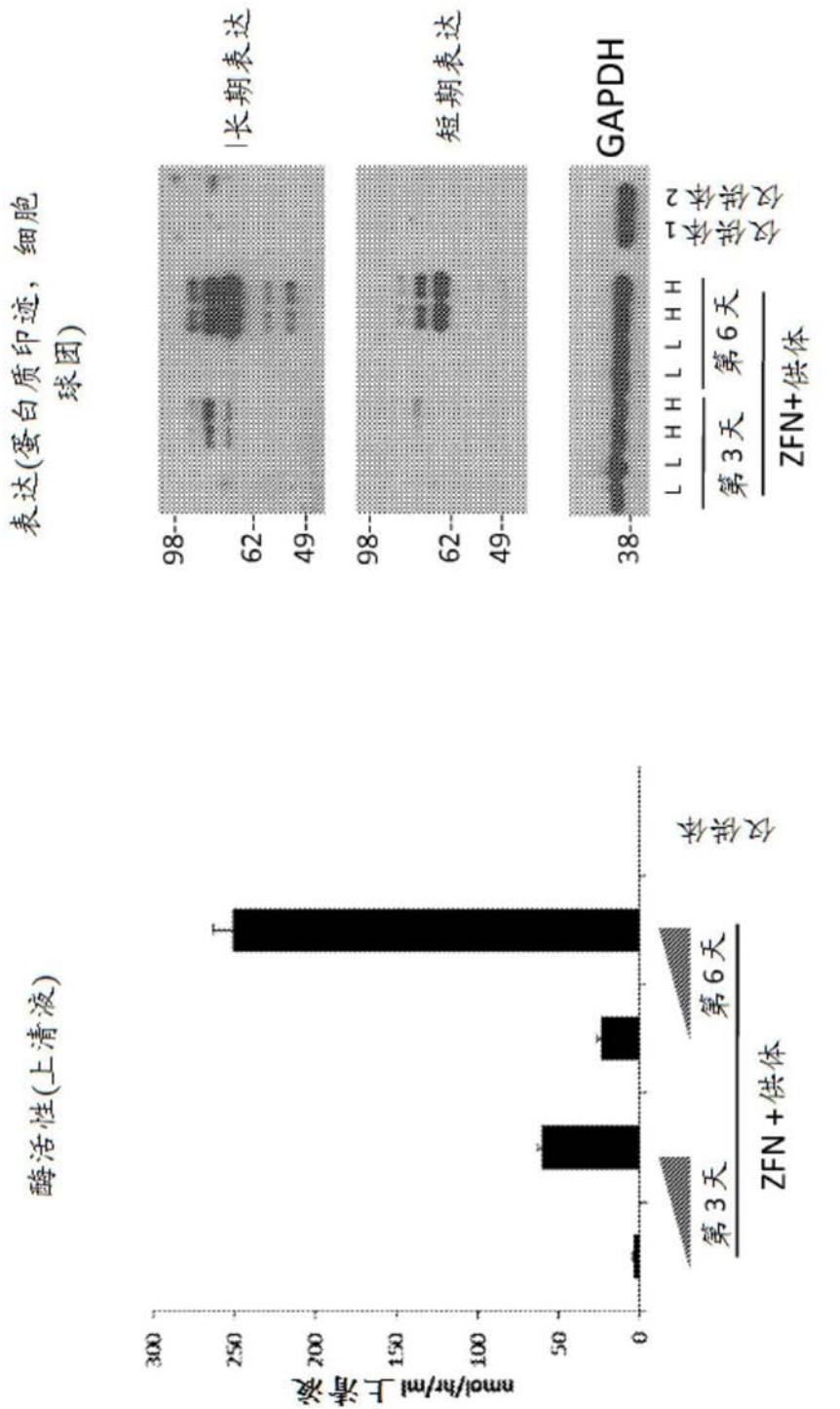


图 13A

IDUA 供体

图 13B

表达(蛋白质印迹, 细胞  
球团)  
酶活性(上清液)

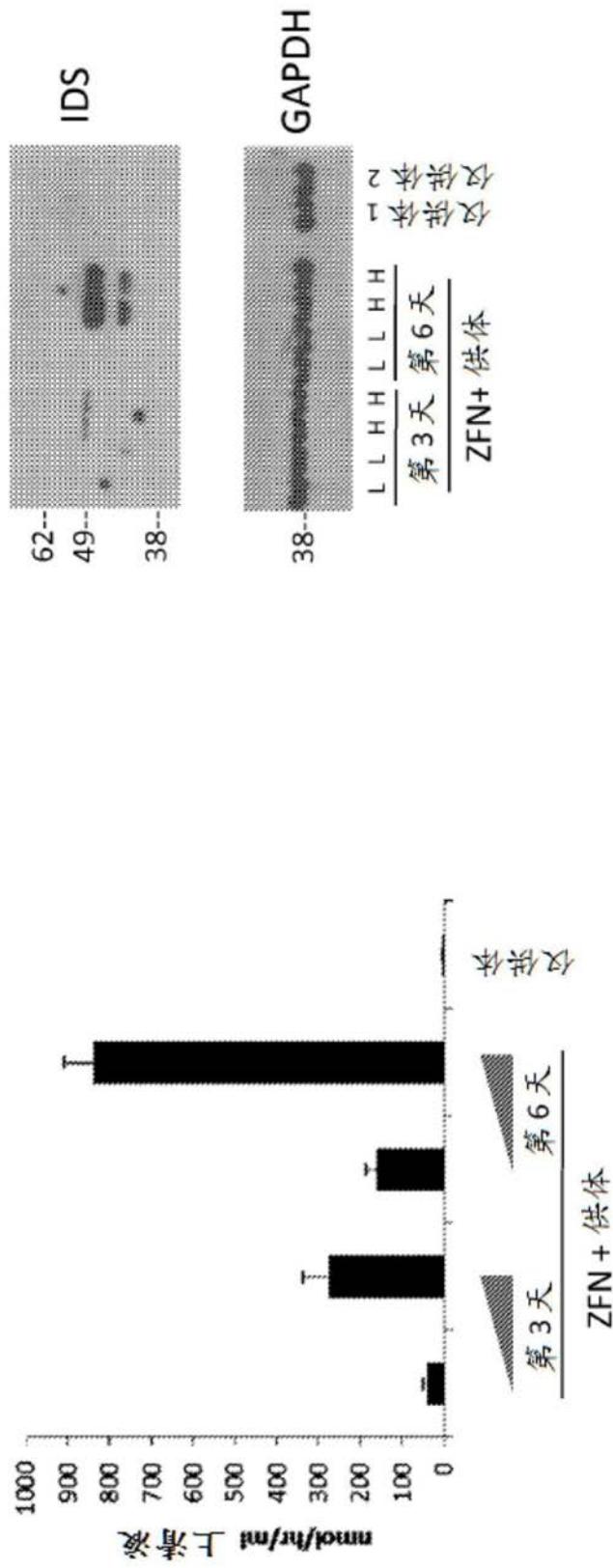


图 13D

IDS 供体

图 13C