



(12) 发明专利申请

(10) 申请公布号 CN 117178959 A

(43) 申请公布日 2023. 12. 08

(21) 申请号 202311154585.8

(22) 申请日 2020.04.02

(30) 优先权数据

62/829,321 2019.04.04 US

(62) 分案原申请数据

202080038651.3 2020.04.02

(71) 申请人 瑞泽恩制药公司

地址 美国纽约州塔里敦

(72) 发明人 唐亚俊 丹·夏洛索恩

林登·米特璠尔 洛里·莫顿

达莉亚·扎莫洛奇科夫

妮可·阿莱桑德里-哈贝尔

林恩·麦克唐纳

(74) 专利代理机构 广州三环专利商标代理有限公司 44202

专利代理师 郝传鑫

(51) Int.Cl.

A01K 67/027 (2006.01)

C12Q 1/6851 (2018.01)

C12Q 1/6876 (2018.01)

权利要求书1页 说明书75页

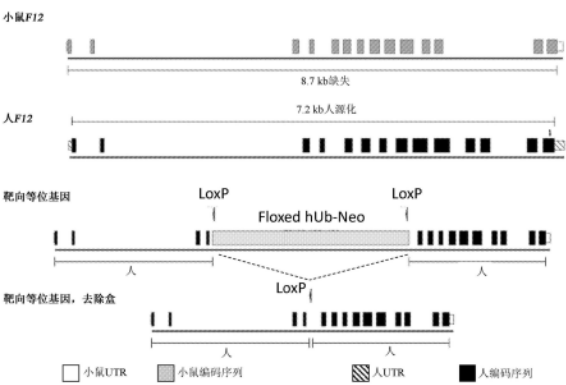
序列表(电子公布) 附图16页

(54) 发明名称

包括人源化凝血因子12基因座的非人动物

(57) 摘要

提供了包括人源化凝血因子XII (F12) 基因座的非人动物基因组、非人动物细胞和非人动物以及制备和使用此类非人动物基因组、非人动物细胞和非人动物的方法。包括人源化F12基因座的非人动物细胞或非人动物表达人凝血因子XII蛋白或嵌合凝血因子XII蛋白,所述嵌合凝血因子XII蛋白的片段来自人凝血因子XII。提供了用于使用包括人源化F12基因座的此类非人动物来评估如被设计成靶向人F12的核酸酶试剂等靶向人凝血因子XII的试剂的体内功效的方法。还提供了在脑中局部产生的F12的短同种型以及使用所述短同种型的方法。



1. 一种评估短F12 mRNA同种型在生物样品中的表达的方法,所述方法包括:

(a) 使用针对F12基因的外显子9-14中的一个或多个外显子的引物和/或探针检测和量化RNA转录物;以及

(b) 使用针对所述F12基因的外显子1-6中的一个或多个外显子的引物和/或探针检测和量化RNA转录物,

其中所述短同种型由针对外显子9-14的引物和/或探针检测,但不由针对外显子1-6的引物和/或探针检测。

2. 根据权利要求1所述的方法,其中所述生物样品来自脑。

3. 根据权利要求1所述的方法,其中所述生物样品包括神经元。

4. 根据权利要求1所述的方法,其中所述生物样品来自人。

5. 根据权利要求1所述的方法,其中所述生物样品来自具有人源化F12基因座的非人动物。

6. 根据权利要求1所述的方法,其中所述生物样品来自包括人神经元。

7. 根据权利要求1所述的方法,其中所述生物样品包括来自具有人源化F12基因座的非人动物的神经元。

8. 根据权利要求1-7中任一项所述的方法,其中步骤(a)包括使用针对F12基因的外显子11-14中的一个或多个外显子的引物和/或探针检测和量化RNA转录物,其中所述短同种型由针对外显子11-14的引物和/或探针检测,但不由针对外显子1-6的引物和/或探针检测。

9. 根据权利要求8所述的方法,其中所述生物样品包括人神经元。

10. 根据权利要求8所述的方法,其中所述生物样品包括来自具有人源化F12基因座的非人动物的神经元。

11. 根据权利要求1-7、9和10中任一项所述的方法,其中步骤(a)和(b)中所述检测和量化包括定量PCR。

12. 根据权利要求1-7、9和10中任一项所述的方法,进一步包括评估所述短F12 mRNA同种型的活性。

13. 根据权利要求12所述的方法,其中评估所述短F12 mRNA同种型的活性包括测量肝细胞生长因子(HGF)-酪氨酸-蛋白激酶Met(Met)信号传导活性。

14. 根据权利要求12所述的方法,其中评估所述短F12 mRNA同种型的活性包括评估前激肽释放酶对激肽释放酶的活化作用、评估FXI的活化作用、评估pro-HGF对HGF的活化作用、或评估经典补体途径的活化作用。

15. 根据权利要求14所述的方法,其中所述生物样品包括人神经元。

16. 根据权利要求14所述的方法,其中所述生物样品包括来自具有人源化F12基因座的非人动物的神经元。

包括人源化凝血因子12基因座的非人动物

[0001] 分案申请说明

[0002] 本申请是申请日为2020年04月02日,申请号为202080038651.3,发明名称为“包括人源化凝血因子12基因座的非人动物”的发明专利申请的分案申请。

[0003] 相关申请的交叉引用

[0004] 本申请要求于2019年4月4日提交的美国申请第62/829,321号的权益,所述美国申请出于所有目的通过引用整体并入本文。

[0005] 通过EFSWEB以文本文件形式提交的序列表

[0006] 写入文件10554CN02.xml中的序列表为172千字节,创建于2023年9月5日(实际序列内容创建于2020年3月22日),并且特此通过引用并入。

背景技术

[0007] 凝血因子连同血小板是血管损伤期间的止血过程中相关的血液组分。众所周知,当调节(即,产生和/或活动)中出现不平衡时,这些组分可能是血栓形成的驱动因素。血栓性疾病被认为主要由外在途径(通过组织因子)的异常活化引起,但最近对F12缺陷型小鼠和靶向活化凝血因子XII(因子XIIa)的分子进行的临床前研究表明,内在凝血途径(也称为接触途径)也有所涉及。因子XIIa是接触途径的核心,并且可以通过因子XI的切割来驱动凝血,或者可以通过血浆前激肽释放酶的切割来驱动炎症。因此,对因子XII活性的抑制可能导致与接触途径活化相关的血栓性凝血和炎症减少。

[0008] 然而,仍然需要合适的非人动物提供靶向人凝血因子XII的试剂的真实人靶标或真实人靶标的近似靶标,由此能够在活体动物中测试此类药剂的功效和作用模式以及药代动力学和药效学研究。

发明内容

[0009] 提供了包括人源化凝血因子XII(F12)基因座的非人动物以及制备和使用此类非人动物的方法。还提供了包括人源化凝血因子XII(F12)基因座的非人动物基因组或细胞以及制备和使用此类非人动物基因组或细胞的方法。还提供了人源化非人动物F12基因、核酸酶试剂和/或用于使非人动物F12基因人源化的靶向载体以及制备和使用此类人源化F12基因的方法。

[0010] 一方面,提供了包括人源化凝血因子XII(F12)基因座的非人动物基因组、非人动物细胞或非人动物。此类非人动物基因组、非人动物细胞或非人动物可以在其基因组中包括人源化内源F12基因座,在所述基因座中,内源F12基因座的区段已被缺失并替换为对应的人F12序列。

[0011] 在一些此类非人动物基因组、非人动物细胞或非人动物中,人源化内源F12基因座对包括人凝血因子XII重链的蛋白质进行编码。在一些此类非人动物基因组、非人动物细胞或非人动物中,人源化内源F12基因座对包括人凝血因子XII轻链的蛋白质进行编码。在一些此类非人动物基因组、非人动物细胞或非人动物中,人源化内源F12基因座对包括人凝血

因子XII信号肽的蛋白质进行编码。

[0012] 在一些此类非人动物基因组、非人动物细胞或非人动物中,包括编码序列和非编码序列两者的内源F12基因座的区域已被缺失并替换为包括编码序列和非编码序列两者的对应的人F12序列。在一些此类非人动物基因组、非人动物细胞或非人动物中,人源化内源F12基因座包括内源F12启动子,其中人F12序列可操作地连接到内源F12启动子。在一些此类非人动物基因组、非人动物细胞或非人动物中,内源F12基因座的至少一个内含子和至少一个外显子已被缺失并替换为对应的人F12序列。

[0013] 在一些此类非人动物基因组、非人动物细胞或非人动物中,内源F12基因座的整个F12编码序列已被缺失并替换为对应的人F12序列。任选地,内源F12基因座从起始密码子到终止密码子的区域已被缺失并替换为对应的人F12序列。

[0014] 在一些此类非人动物基因组、非人动物细胞或非人动物中,内源F123'非翻译区尚未被缺失并且尚未替换为对应的人F12序列。在一些此类非人动物基因组、非人动物细胞或非人动物中,内源F125'非翻译区尚未被缺失并且尚未替换为对应的人F12序列。

[0015] 在一些此类非人动物基因组、非人动物细胞或非人动物中,内源F12基因座从起始密码子到终止密码子的区域已被缺失并替换为对应的人F12序列,并且内源F12启动子尚未被缺失并且尚未替换为对应的人F12序列。

[0016] 在一些此类非人动物基因组、非人动物细胞或非人动物中,人源化内源F12基因座处的人F12序列包括与SEQ ID NO:31或32中所示的序列至少90%、95%、96%、97%、98%、99%或100%相同的序列。在一些此类非人动物基因组、非人动物细胞或非人动物中,人源化内源F12基因座对蛋白质进行编码,所述蛋白质包括与SEQ ID NO:5或46中所示的序列至少90%、95%、96%、97%、98%、99%或100%相同的序列。在一些此类非人动物基因组、非人动物细胞或非人动物中,人源化内源F12基因座包括编码序列,所述编码序列包括与SEQ ID NO:13或48中所示的序列至少90%、95%、96%、97%、98%、99%或100%相同的序列。在一些此类非人动物基因组、非人动物细胞或非人动物中,人源化内源F12基因座包括与SEQ ID NO:17或18中所示的序列至少90%、95%、96%、97%、98%、99%或100%相同的序列。

[0017] 在一些此类非人动物基因组、非人动物细胞或非人动物中,人源化内源F12基因座不包括选择盒或报告基因。在一些此类非人动物基因组、非人动物细胞或非人动物中,人源化内源F12基因座包括位于对应的人F12序列的内含子内的选择盒或报告基因。

[0018] 在一些此类非人动物基因组、非人动物细胞或非人动物中,非人动物对于人源化内源F12基因座是纯合的。

[0019] 在一些此类非人动物基因组、非人动物细胞或非人动物中,非人动物包括其种系中的人源化内源F12基因座。

[0020] 在一些此类非人动物基因组、非人动物细胞或非人动物中,非人动物是哺乳动物。任选地,非人动物是大鼠或小鼠。任选地,非人动物是小鼠。

[0021] 在一些此类非人动物基因组、非人动物细胞或非人动物中,非人动物在脑中表达短人凝血因子XII同种型。任选地,短人凝血因子XII同种型由F12 mRNA进行编码,所述F12 mRNA由针对人F12的外显子11-14的探针检测,但不由针对人F12的外显子1-6的探针检测。任选地,短人凝血因子XII同种型是FXII₂₉₇₋₅₉₆。任选地,短人凝血因子XII同种型在脑的神经元中表达。任选地,短人凝血因子XII同种型可以由血浆激肽释放酶活化,并将pro-HGF转

化为活性HGF。

[0022] 在一些此类非人动物基因组、非人动物细胞或非人动物中,非人动物具有与对应的野生型非人动物中的活化部分凝血活酶时间(aPTT)无显著差异的aPTT,和/或其中非人动物具有与对应的野生型非人动物中的前凝血酶时间(PT)无显著差异的PT。

[0023] 在一些此类非人动物基因组、非人动物细胞或非人动物中,非人具有至少约5 μ g/mL的人凝血因子XII血浆水平。

[0024] 在一些此类非人动物基因组、非人动物细胞或非人动物中,内源F12基因座从起始密码子到终止密码子的区域已被缺失并替换为对应的人F12序列,内源F12启动子尚未被缺失并且尚未替换为对应的人F12序列,并且非人动物在脑中表达短人凝血因子XII同种型,任选地其中:(i)短人凝血因子XII同种型由F12 mRNA进行编码,所述F12 mRNA由针对人F12的外显子11-14的探针检测,但不由针对人F12的外显子1-6的探针检测;和/或(ii)短人凝血因子XII同种型是FXII₂₉₇₋₅₉₆;和/或(iii)短人凝血因子XII同种型在脑的神经元中表达;和/或(iv)短人凝血因子XII同种型可以由血浆激肽释放酶活化,并将pro-HGF转化为活性HGF。

[0025] 另一方面,提供了一种用于产生人源化内源F12基因座的靶向载体,在所述基因座中,内源F12基因座的区段已被缺失并替换为对应的人F12序列,其中所述靶向载体包括插入核酸,所述插入核酸包括侧接有靶向所述内源F12基因座处的5'靶序列的5'同源臂和靶向所述内源F12基因座处的3'靶序列的3'同源臂的对应的人F12序列。

[0026] 另一方面,提供了一种人源化非人动物F12基因,其中内源F12基因座的区段已被缺失并替换为对应的人F12序列。

[0027] 另一方面,提供了一种在体内评估靶向人凝血因子XII的试剂的活性的方法。一些此类方法包括:(a)向上述非人动物中的任何非人动物施用靶向人凝血因子XII的试剂;以及(b)评估靶向人凝血因子XII的试剂在非人动物中的活性。

[0028] 在一些此类方法中,施用包括腺相关病毒(AAV)介导的递送、脂质纳米颗粒(LNP)介导的递送、流体动力学递送(HDD)或注射。

[0029] 在一些此类方法中,步骤(b)包括从非人动物中分离出血浆并评估靶向人凝血因子XII的试剂在血浆中的活性。在一些此类方法中,步骤(b)包括评估靶向人凝血因子XII的试剂在非人动物的肝脏或脑中的活性。在一些此类方法中,步骤(b)包括评估凝血或凝血酶生成。在一些此类方法中,步骤(b)包括评估HGF-Met信号传导。在一些此类方法中,步骤(b)包括测量由人源化内源F12基因座编码的F12信使RNA的表达。在一些此类方法中,步骤(b)包括测量由人源化内源凝血因子XII基因座编码的凝血因子XII蛋白的表达。在一些此类方法中,靶向人凝血因子XII的试剂是基因组编辑药剂,并且步骤(b)包括评估人源化内源F12基因座的修饰。在一些此类方法中,步骤(b)包括测量人源化内源F12基因座内插入或缺失的频率。

[0030] 在一些此类方法中,靶向人凝血因子XII的试剂包括被设计成靶向人F12基因区域的核酸酶试剂。任选地,核酸酶试剂包括Cas蛋白和被设计成靶向人F12基因中的向导RNA靶序列的向导RNA。任选地,Cas蛋白是Cas9蛋白。

[0031] 在一些此类方法中,靶向人凝血因子XII的试剂包括外源供体核酸,其中所述外源供体核酸被设计成靶向人F12基因,并且任选地其中所述外源供体核酸通过AAV进行递送。

在一些此类方法中,靶向人凝血因子XII的试剂是RNAi试剂或反义寡核苷酸。在一些此类方法中,靶向人凝血因子XII的试剂是抗原结合蛋白。在一些此类方法中,靶向人凝血因子XII的试剂是小分子。

[0032] 另一方面,提供了一种在体内优化靶向人凝血因子XII的试剂的活性的方法。一些此类方法包括: (I) 在第一非人动物中首次执行以上方法中的任何方法,所述第一非人动物在其基因组中包括人源化内源F12基因座; (II) 改变变量并在第二非人动物中用改变的变量第二次执行步骤(I)的方法,所述第二非人动物在其基因组中包括人源化内源F12基因座;以及(III)将步骤(I)中的靶向人凝血因子XII的试剂的活性与步骤(II)中的靶向人凝血因子XII的试剂的活性进行比较,并选择产生更高活性的方法。

[0033] 在一些此类方法中,步骤(II)中改变的变量是将靶向人凝血因子XII的试剂引入到非人动物中的递送方法。在一些此类方法中,步骤(II)中改变的变量是将靶向人凝血因子XII的试剂引入到非人动物中的施用途径。在一些此类方法中,步骤(II)中改变的变量是引入到非人动物中的靶向人凝血因子XII的试剂的浓度或量。在一些此类方法中,步骤(II)中改变的变量是引入到非人动物中的靶向人凝血因子XII的试剂的形式。在一些此类方法中,步骤(II)中改变的变量是引入到非人动物中的靶向人凝血因子XII的试剂。

[0034] 另一方面,提供了制备以上非人动物中的任何非人动物的方法。一些此类方法包括: (a) 向非人动物胚胎干(ES)细胞中引入: (i) 靶向内源F12基因座中的靶序列的核酸酶试剂;以及(ii)包括核酸插入物的靶向载体,所述核酸插入物包括侧接有对应于内源F12基因座中的5'靶序列的5'同源臂和对应于内源F12基因座中的3'靶序列的3'同源臂的人F12序列,其中靶向载体与内源F12基因座重组,以产生在其基因组中包括包含人F12序列的人源化内源F12基因座的基因修饰的非人ES细胞; (b) 将基因修饰的非人ES细胞引入到非人动物宿主胚胎中;以及(c)在代孕母体中孕育非人动物宿主胚胎,其中代孕母体产生在其基因组中包括包含人F12序列的人源化内源F12基因座的F0后代基因修饰的非人动物。任选地,靶向载体是长度为至少10kb的大靶向载体,或者在所述大靶向载体中,5'和3'同源臂的总长度为至少10kb。

[0035] 一些此类方法包括: (a) 向非人动物单细胞阶段胚胎中引入: (i) 靶向内源F12基因座中的靶序列的核酸酶试剂;以及(ii)包括核酸插入物的靶向载体,所述核酸插入物包括侧接有对应于内源F12基因座中的5'靶序列的5'同源臂和对应于内源F12基因座中的3'靶序列的3'同源臂的人F12序列,其中靶向载体与内源F12基因座重组,以产生在其基因组中包括包含人F12序列的人源化内源F12基因座的基因修饰的非人单细胞阶段胚胎; (b) 在代孕母体中孕育基因修饰的非人动物单细胞阶段胚胎,以产生在其基因组中包括包含人F12序列的人源化内源F12基因座的基因修饰的F0代非人动物。

[0036] 在一些此类方法中,核酸酶试剂包括Cas蛋白和向导RNA。任选地,Cas蛋白是Cas9蛋白。在一些此类方法中,步骤(a)进一步包括引入靶向内源F12基因座内的第二靶序列的第二向导RNA。

[0037] 一些此类方法包括: (a) 向非人动物宿主胚胎中引入在其基因组中包括人源化内源F12基因座的基因修饰的非人动物胚胎干(ES)细胞,在所述基因座中,内源F12基因座的区段已被缺失并替换为对应的人F12序列;以及(b)在代孕母体中孕育非人动物宿主胚胎,其中所述代孕母体产生包括人源化内源F12基因座的F0后代基因修饰的非人动物。

[0038] 一些此类方法包括：(a) 向非人动物胚胎干 (ES) 细胞中引入包括核酸插入物的靶向载体，所述核酸插入物包括侧接有对应于内源F12基因座中的5'靶序列的5'同源臂和对应于内源F12基因座中的3'靶序列的3'同源臂的人F12序列，其中靶向载体与内源F12基因座重组，以产生在其基因组中包括包含人F12序列的人源化内源F12基因座的基因修饰的非人ES细胞；(b) 将基因修饰的非人ES细胞引入到非人动物宿主胚胎中；以及(c) 在代孕母体中孕育非人动物宿主胚胎，其中代孕母体产生在其基因组中包括包含人F12序列的人源化内源F12基因座的F0后代基因修饰的非人动物。任选地，靶向载体是长度为至少10kb的大靶向载体，或者在所述大靶向载体中，5'和3'同源臂的总长度为至少10kb。

[0039] 一些此类方法包括：(a) 向非人动物单细胞阶段胚胎中引入包括核酸插入物的靶向载体，所述核酸插入物包括侧接有对应于内源F12基因座中的5'靶序列的5'同源臂和对应于内源F12基因座中的3'靶序列的3'同源臂的人F12序列，其中靶向载体与内源F12基因座重组，以产生在其基因组中包括包含人F12序列的人源化内源F12基因座的基因修饰的非人单细胞阶段胚胎；(b) 在代孕母体中孕育基因修饰的非人动物单细胞阶段胚胎，以产生在其基因组中包括包含人F12序列的人源化内源F12基因座的基因修饰的F0代非人动物。

[0040] 在一些此类方法中，步骤(a)进一步包括引入核酸酶试剂或对核酸酶试剂进行编码的核酸，其中核酸酶试剂靶向内源F12基因座中的靶序列。任选地，核酸酶试剂包括Cas蛋白和向导RNA。任选地，Cas蛋白是Cas9蛋白。任选地，步骤(a)进一步包括引入靶向内源F12基因座内的第二靶序列的第二向导RNA。

[0041] 在一些此类方法中，非人动物是小鼠或大鼠。任选地，非人动物是小鼠。

[0042] 另一方面，提供了评估短F12mRNA同种型在生物样品中的表达的方法。一些此类方法包括：(a) 使用针对F12基因的外显子9-14中的一个或多个外显子的引物和/或探针检测和量化RNA转录物；以及(b) 使用针对F12基因的外显子1-6中的一个或多个外显子的引物和/或探针检测和量化RNA转录物，其中短同种型由针对外显子9-14的引物和/或探针检测，但不由针对外显子1-6的引物和/或探针检测。

附图说明

[0043] 图1(未按比例)示出了用于小鼠凝血因子XII (F12) 基因座的人源化的靶向方案的示意图。图的上半部分示出了内源性小鼠F12基因座和内源性人F12基因座，并且图的下半部分示出了具有或不具有自缺失选择盒的人源化基因座。

[0044] 图2(未按比例)示出了用于筛选小鼠F12基因座的人源化的TAQMAN[®]测定的示意图。等位基因获得 (GOA) 测定包含7374hU和7374hD。等位基因丢失 (LOA) 测定包含7374mTU和7374mTD。

[0045] 图3示出了人和小鼠凝血因子XII蛋白(分别为hF12和mF12)的比对。信号肽加框，并注明重链和轻链区域。

[0046] 图4示出了来自人源化F12小鼠和野生型 (VG WT) 小鼠的血浆样品中的人凝血因子XII水平。汇集的正常人血浆用作阳性对照，并且人FXII-ID (FXII-免疫耗竭的) 血浆用作阴性对照。斯图登氏t测试 (Student's t-test)；***对于被测人血浆，相对于正常人血浆， $p < 0.001$ ，或对于小鼠血浆，相对于野生型， $p < 0.001$ 。

[0047] 图5示出了来自野生型小鼠 (FXII^{+/+})、F12敲除小鼠 (FXII^{-/-}) 和人源化F12

(FXII^{hum/hum})小鼠的血浆样品的蛋白质印迹分析。测量了小鼠凝血因子XII和人凝血因子XII的表达。白蛋白用作上样对照。

[0048] 图6示出了野生型和人源化F12小鼠的活化部分凝血活酶时间(aPTT)结果。

[0049] 图7A是示出qPCR探针相对于FXII蛋白结构域的定位的示意图。每个探针下方的线表示F12 mRNA扩增的区域。

[0050] 图7B示出了qPCR分析的结果,证明了所有F12 mRNA区域在肝脏中表达,但只有外显子7-14在全脑、海马体和颞叶中表达。F12表达被相对于GAPDH归一化,并绘制为肝脏水平百分比。

[0051] 图7C示出了来自GTEx门户网站的RNAseq数据,其示出了在人肝脏中从外显子1到外显子14的完整F12 mRNA覆盖,但在人海马体和皮层中仅覆盖外显子7到14。F12外显子从右到左排序,因为基因定位在负链上。数据于2018年12月从GTEx门户网站获得。

[0052] 图8A是示出人和小鼠F12基因和转录物的UCSC基因组浏览器分析的示意图。人F12具有跨越外显子8-12的由1364个碱基(CpG二核苷酸计数=141;68.9% GC含量)组成的大CpG岛,所述岛在小鼠F12中是缺乏的,其中CpG岛具有低得多的CpG计数19,并且由231个碱基组成。GenBank示出了从外显子9开始的5个人F12转录物,这在小鼠F12中是不存在的。在这5个转录物中,有两个得到了实验证据的支持,而三个是合成的。从外显子9开始的F12转录物的登录号为:BC012390、KJ901422、KR710723、KR710724和BT007350。数据从UCSC基因组浏览器(<http://genome.ucsc.edu>)获得。2013年12月(GRCh38/hg38)汇编用于人类数据,并且2011年12月(GRCm38/mm10)汇编用于小鼠数据。

[0053] 图8B示出了在脑代码中发现的F12 mRNA的三个开放阅读框(ORF),其中ORF 1和2对类似于 β FXIIa(N335-S596)的蛋白质进行编码。

[0054] 图9A-9D示出FXII₂₉₇₋₅₉₆在由激肽释放酶活化后获得催化活性并且可以将pro-HGF转化为HGF。

[0055] 在图9A中,FXII₂₉₇₋₅₉₆在存在或不存在激肽释放酶(15nM)的情况下温育,并且在用FXIIa显色底物S-2302测量FXII₂₉₇₋₅₉₆的活性之前用抑肽酶抑制激肽释放酶活性。来自用激肽释放酶预温育的FXII₂₉₇₋₅₉₆的信号不是由于残留的激肽释放酶活性而产生的,因为在省略FXII₂₉₇₋₅₉₆的样品中没有观察到活性。示出平均值 $\pm$ SEM。

[0056] 在图9B中,FXII(1.25 μ M)用组织蛋白酶K(Cath K;100nM)或媒剂温育,并且反应产物通过非还原SDS-PAGE进行分析。约40kDa的FXII切割产物可能是由于L296与M297之间的切割而产生的,从而产生FXII₂₉₇₋₅₉₆。组织蛋白酶K切割的FXII(375nM FXII,30nM组织蛋白酶K)的活性通过显色底物测定进行分析。在不存在激肽释放酶活化的情况下观察到的组织蛋白酶K切割的FXII的活性可以忽略不计,这指示单独的组织蛋白酶K切割不会产生活性FXII。用激肽释放酶(15nM)温育的组织蛋白酶切割的FXII的活性比用激肽释放酶温育的全长FXII高得多。单独的FXII、激肽释放酶和组织蛋白酶K的活性可以忽略不计。示出平均值 $\pm$ SEM。

[0057] 在图9C中,25 μ g/mL pro-HGF由HGFA和 β FXIIa切割,而不是由tPA或凝血酶(均为50nM)切割。

[0058] 在图9D中,FXII₂₉₇₋₅₉₆(225nM)和FXII(50nM)用激肽释放酶(15nM)活化,并且在用Pro-HGF(25 μ g/mL)温育之前用抑肽酶抑制激肽释放酶活性。Pro-HGF由激肽释放酶活化的

FXII₂₉₇₋₅₉₆和激肽释放酶活化的FXII(箭头)切割,而不是由单独的经灭活的激肽释放酶切割。tPA和HGFA(225nM)作为阳性和阴性对照包含在内。

[0059] 图10A-10B示出了来自人和小鼠组织的RNAseq数据。图10A示出了于2018年12月从GTEX门户网站获得的选定人体组织中的F12表达。图10B示出了从C57BL/6小鼠组织的RNAseq分析获得的选定小鼠组织中的F12表达。图上的竖线指示每百万映射读段的每千碱基转录物的片段(FPKM) = 1。

[0060] 图11示出FXII₂₉₇₋₅₉₆(225nM)由激肽释放酶(15nM)活化,并且在用前激肽释放酶(200nM)温育之前用抑肽酶抑制激肽释放酶活性。前激肽释放酶由激肽释放酶活化的FXII₂₉₇₋₅₉₆切割,而不是由单独的经灭活的激肽释放酶切割。由于用于活化FXII₂₉₇₋₅₉₆的激肽释放酶的水平低,因此在激肽释放酶对照泳道中未观察到激肽释放酶重链。

[0061] 定义

[0062] 本文可互换使用的术语“蛋白质”、“多肽”、和“肽”包含任何长度的聚合形式的氨基酸,包含编码氨基酸和非编码氨基酸以及经化学或生物化学修饰或衍生的氨基酸。这些术语还包含已经修饰的聚合物,如具有经修饰的肽主链的多肽。术语“结构域”是指具有特定功能或结构的蛋白质或多肽的任何部分。

[0063] 本文可互换使用的术语“核酸”和“多核苷酸”包含任何长度的聚合形式的核苷酸,包含核糖核苷酸、脱氧核糖核苷酸或其类似物或经修饰的版本。所述核苷酸包含单链、双链和多链DNA或RNA、基因组DNA、cDNA、DNA-RNA杂交体和包括嘌呤碱基、嘧啶碱基或其它天然的、经化学修饰的、经生物化学修饰的、非天然的或衍生的核苷酸碱基的聚合物。

[0064] 术语“基因组整合的”是指已被引入到细胞中使得核苷酸序列整合到细胞的基因组中的核酸。可以使用任何方案用于将核酸稳定地掺入到细胞的基因组中。

[0065] 术语“靶向载体”是指可以通过同源重组、非同源末端接合介导的连结或任何其它重组方式引入到细胞基因组中的靶位置的重组核酸。

[0066] 术语“病毒载体”是指包含至少一种病毒来源元素并包含足以或允许包装成病毒载体颗粒的元素的重组核酸。载体和/或颗粒可以用于在体外、离体或在体内将DNA、RNA或其它核酸转移到细胞中的目的。许多形式的病毒载体是已知的。

[0067] 关于细胞、组织、蛋白质和核酸的术语“分离的”包含相对于其它细菌、病毒、细胞或通常可能原位存在的其它组分而言相对纯化的细胞、组织、蛋白质和核酸,直至并包含细胞、组织、蛋白质和核酸的基本上纯的调配物。术语“分离的”还包含不具有天然存在的对应物、已经被化学合成并且因此基本上未被其它细胞、组织、蛋白质和核酸污染或者已经从其天然伴随的大多数其它组分(例如,细胞组分)(例如,其它细胞蛋白、多核苷酸或细胞组分)中分离或纯化的细胞、组织、蛋白质和核酸。

[0068] 术语“野生型”包含具有如在正常(与突变、患病、改变等相比)状态或情况下发现的结构和/或活性的实体。野生型基因和多肽通常以多种不同形式(例如,等位基因)存在。

[0069] 术语“内源序列”是指天然存在于细胞或非人动物内的核酸序列。例如,非人动物的内源性F12序列是指天然存在于非人动物的F12基因座处的天然F12序列。

[0070] “外源”分子或序列包含通常不以所述形式或位置(例如,基因组基因座)存在于细胞中的分子或序列。正常存在包含关于细胞的特定发育阶段和环境条件的存在。例如,外源分子或序列可以包含细胞内对应的内源序列的突变版本,如内源序列的人源化版本,或者

可以包含与细胞内的内源序列相对应但形式不同(即,不在染色体内)的序列。相比之下,“内源”分子或序列包含在特定环境条件下在特定发育阶段在特定细胞中通常以所述形式和位置存在的分子或序列。

[0071] 当在核酸或蛋白质的上下文中使用时,术语“异源”指示核酸或蛋白质包括至少两个在同一分子中不天然存在的区段。例如,当关于核酸片段或蛋白质片段使用时,术语“异源”指示核酸或蛋白质包括在自然界中未发现彼此之间有相同关系(例如,连接在一起)的两个或更多个子序列。作为一个实例,核酸载体的“异源”区域是在自然界中未发现与另一分子相关联的另一个核酸分子内或与其相连的核酸区段。例如,核酸载体的异源区域可以包含侧接有在自然界中未发现与编码序列相关联的序列的编码序列。同样地,蛋白质的“异源”区域是在自然界中未发现与其它肽分子相关联的另一个肽分子(例如,融合蛋白或带有标签的蛋白质)内或与其相连的氨基酸区段。类似地,核酸或蛋白质可以包括异源标记或异源分泌或定位序列。

[0072] “密码子优化”利用密码子的简并性,如通过指定氨基酸的三碱基对密码子组合的多样性所展示的,并且通常包含通过用宿主细胞的基因中更频繁或最频繁使用的密码子替换自然序列的至少一个密码子同时保持自然氨基酸序列来修饰核酸序列以在特定宿主细胞中增强表达的过程。例如,可以修饰对Cas9蛋白进行编码的核酸以取代与天然存在的核酸序列相比在包含细菌细胞、酵母细胞、人细胞、非人细胞、哺乳动物细胞、啮齿动物细胞、小鼠细胞、大鼠细胞、仓鼠细胞或任何其它宿主细胞的给定原核或真核细胞中具有更高使用频率的密码子。密码子使用表例如在“密码子使用数据库”处很容易获得。这些表可以通过多种方式进行修改。参见Nakamura等人(2000)《核酸研究(Nucleic Acids Research)》28:292,所述文献出于所有目的通过引用整体并入本文。还可获得用于在特定宿主中表达的特定序列的密码子优化的计算机算法(参见例如《基因伪造(Gene Forge)》)。

[0073] 术语“基因座”是指基因(或重要序列)、DNA序列、多肽编码序列的具体位置,或生物体基因组的染色体上的位置。例如,“凝血因子XII基因座”或“F12基因座”可以指F12基因的特定位置、F12 DNA序列、凝血因子XII编码序列或F12在已经被识别为这种序列所在的生物体基因组的染色体上的位置。“F12基因座”可以包括F12基因的调节元件,包含例如增强子、启动子、5'和/或3'非翻译区(UTR)或其组合。

[0074] 术语“基因”是指染色体中的DNA序列,所述染色体如果天然存在可以含有至少一个编码区和至少一个非编码区。染色体中编码产物(例如但不限于RNA产物和/或多肽产物)的DNA序列可以包含被非编码内含子中断的编码区和在5'和3'端两者上邻近编码区定位使得基因对应于全长mRNA的序列(包含5'和3'非翻译序列)。另外,其它非编码序列,包含调节序列(例如但不限于启动子、增强子和转录因子结合位点)、聚腺苷酸化信号、内部核糖体进入位点、沉默子、绝缘序列和基质附着区可以存在于基因中。这些序列可以接近基因的编码区(例如但不限于在10kb内)或位于远处位点,并且这些序列会影响基因的转录和翻译水平或速率。

[0075] 术语“等位基因”是指基因的变体形式。一些基因具有多种不同的形式,其位于染色体上的相同位置或遗传基因座。二倍体生物在每个基因座处有两个等位基因。每对等位基因表示特异性遗传基因座的基因型。如果在特定基因座处有两个相同的等位基因,则基因型被描述为纯合的,如果两个等位基因不同,则基因型被描述为杂合的。

[0076] “启动子”是DNA的调节区域,其通常包括能够指导RNA聚合酶II在特定多核苷酸序列的适当转录起始位点处引发RNA合成的TATA盒。启动子可以另外包括影响转录起始速率的其它区域。本文所公开的启动子序列调节可操作地连接的多核苷酸的转录。启动子可以在本文所公开的细胞类型(例如,真核细胞、非人哺乳动物细胞、人细胞、啮齿动物细胞、多能细胞、单细胞阶段胚胎、分化细胞或其组合)中的一种或多种细胞类型中具有活性。启动子可以是例如组成型活性启动子、条件型启动子、诱导型启动子、时间受限启动子(例如,发育调节启动子)或空间受限启动子(例如,细胞特异性或组织特异性启动子)。启动子的实例可以例如在W0 2013/176772中找到,所述文献出于所有目的通过引用整体并入本文。

[0077] 诱导型启动子的实例包含例如化学调节启动子和物理调节启动子。化学调节启动子包含例如醇调节启动子(例如,醇脱氢酶(alcA)基因启动子)、四环素调节启动子(例如,四环素反应性启动子、四环素操纵子序列(tetO)、tet-On启动子或tet-Off启动子)、类固醇调节启动子(例如,大鼠糖皮质激素受体、雌激素受体启动子或蜕皮激素受体启动子)或金属调节启动子(例如,金属蛋白启动子)。物理调节启动子包含例如温度调节启动子(例如,热休克启动子)和光调节启动子(例如,光诱导型启动子或光抑制型启动子)。

[0078] 组织特异性启动子可以是例如神经元特异性启动子、神经胶质特异性启动子、肌肉细胞特异性启动子、心脏细胞特异性启动子、肾细胞特异性启动子、骨细胞特异性启动子、内皮细胞特异性启动子或免疫细胞特异性启动子(例如,B细胞启动子或T细胞启动子)。

[0079] 发育调节启动子包含例如仅在胚胎发育阶段或仅在成体细胞中具有活性的启动子。

[0080] “可操作的连接”或“可操作地连接”包含将两种或多种组分(例如,启动子和另一种序列元件)并置使得两种组分正常发挥功能,并使得至少一种组分能够介导施加在至少一种其它组分上的功能。例如,如果启动子响应于存在或不存在一种或多种转录调节因子而控制编码序列的转录水平,则可以将所述启动子可操作地连接到编码序列。可操作的连接可以包含此类彼此邻接或以反式作用的序列(例如,调节序列可以在一定距离处起作用以控制编码序列的转录)。

[0081] 核酸的“互补性”意指一条核酸链中的核苷酸序列由于其核碱基的取向而与相对核酸链上的另一序列形成氢键。DNA中的互补碱基通常是A与T和C与G。在RNA中,所述互补碱基通常是C与G和U与A。互补可以是完全互补,也可以是基本互补/充分互补。两个核酸之间的完全互补意指两个核酸可以形成双链体,其中双链体中的每个碱基通过沃森-克里克(Watson-Crick)配对与互补碱基键合。“基本”或“充分”互补意指一条链中的序列与相对链中的序列不完整和/或不完全互补,但两条链上的碱基之间发生充分键合以在设定杂交条件(例如,盐浓度和温度)下形成稳定的杂交复合物。此类条件可以通过使用序列和标准数学计算来预测杂交链的 T_m (熔融温度),或者通过使用常规方法通过 T_m 经验确定来预测。 T_m 包含两条核酸链之间形成的杂交复合物群体50%变性(即,双链核酸分子群体半解离成单链)时的温度。在低于 T_m 的温度下,有利于杂交复合物的形成,而在高于 T_m 的温度下,有利于杂交复合物中的链的熔融或分离。可以通过使用例如 $T_m = 81.5 + 0.41 (\%G+C)$ 来估计在1MNaCl水溶液中具有已知G+C含量的核酸的 T_m ,尽管其它已知的 T_m 计算考虑了核酸结构特性。

[0082] 尽管碱基之间的失配是可能的,但是杂交要求所述两个核酸含有互补序列。两个

核酸之间杂交的适当条件取决于核酸的长度和互补程度,这些变量是众所周知的。两个核苷酸序列之间的互补程度越大,具有这些序列的核酸杂交体的熔融温度(T_m)的值就越大。对于具有短互补性段(例如,互补性超过35个或更少、30个或更少、25个或更少、22个或更少、20个或更少或18个或更少的核苷酸)的核酸之间的杂交,失配的位置变得很重要(参见 Sambrook等人,同上,11.7-11.8)。通常,可杂交核酸的长度为至少约10个核苷酸。可杂交核酸的说明性最小长度包含至少约15个核苷酸、至少约20个核苷酸、至少约22个核苷酸、至少约25个核苷酸和至少约30个核苷酸。此外,可以根据如互补区域的长度和互补程度等因素根据需要调整温度和洗涤溶液盐浓度。

[0083] 多核苷酸序列不需要与其靶核酸的序列100%互补才能特异性杂交。此外,多核苷酸可以在一个或多个区段上杂交,使得中间或邻近区段不参与杂交事件(例如,环结构或发夹结构)。多核苷酸(例如,gRNA)可以与其所靶向的靶核酸序列内的靶区域具有至少70%、至少80%、至少90%、至少95%、至少99%或100%的序列互补性。例如,20个核苷酸中有18个核苷酸与靶区域互补并因此特异性杂交的gRNA将表示90%互补。在此实例中,剩余的非互补核苷酸可以与互补核苷酸聚集或穿插在一起,并且不必彼此邻接或与互补核苷酸邻接。

[0084] 核酸内特定核酸序列段之间的互补百分比可以通过以下常规地确定:使用BLAST程序(基本局部比对搜索工具)和PowerBLAST程序(Altschul等人(1990)《分子生物学杂志(J.Mol.Biol.)》215:403-410;Zhang和Madden(1997)《基因组研究(Genome Res.)》7:649-656),所述文献中的每个文献出于所有目的通过引用整体并入本文,或使用Gap程序(威斯康星州麦迪逊大学研究园(University Research Park, Madison Wis.),遗传学计算机组,Unix第8版,威斯康星序列分析包),所述程序使用默认设置,使用史密斯-沃特曼(Smith and Waterman)算法(1981)《应用数学进展(Adv.Appl.Math.)》2:482-489,所述文献出于所有目的通过引用整体并入本文。

[0085] 本文所提供的方法和组合物采用多种不同的组分。贯穿说明书的一些组分可以具有活性变体和片段。此类组分包含例如Cas蛋白、CRISPR RNA、tracrRNA和向导RNA。这些组分中的每种组分的生物活性在本文别处进行描述。术语“功能性”是指蛋白质或核酸(或其片段或变体)表现出生物活性或功能的先天能力。此类生物活性或功能可以包含例如Cas蛋白与向导RNA和靶DNA序列结合的能力。与原始分子相比,功能性片段或变体的生物学功能可以相同或实际上可以改变(例如,关于其特异性或选择性或功效)但保留分子的基本生物学功能。

[0086] 术语“变体”是指与群体中最普遍的序列不同的核苷酸序列(例如,相差一个核苷酸)或与群体中最普遍的序列不同的蛋白质序列(例如,相差一个氨基酸)。

[0087] 当提及蛋白质时,术语“片段”意指比全长蛋白质更短或具有更少氨基酸的蛋白质。当提及核酸时,术语“片段”意指比全长核酸更短或具有更少核苷酸的核酸。当提及蛋白质片段时,片段可以是例如N端片段(即,去除蛋白质的C末端的一部分)、C端片段(即,去除蛋白质的N末端的一部分)或内部片段(即,去除蛋白质的N末端和C末端中的每个末端的一部分)。当提及核酸片段时,片段可以是例如5'片段(即,去除核酸的3'端的一部分)、3'片段(即,去除核酸的5'端的一部分)或内部片段(即,去除核酸的5'端和3'端中的每个端的一部分)。

[0088] 在两个多核苷酸或多肽序列的上下文中,“序列同一性”或“同一性”是指当在指定的比较窗口上针对最大对应性进行比对时两个序列中相同的残基。当提及蛋白质的序列同一性的百分比时,不相同的残基位置通常因保守性氨基酸取代而不同,其中氨基酸残基被具有相似化学性质(例如,电荷或疏水性)的其它氨基酸残基取代,并且因此不改变分子的功能性质。当序列的保守性取代不同时,可以将百分比序列同一性向上调整以校正取代的保守性质。因此类保守性取代而不同的序列被视为具有“序列相似性”或“相似性”。用于进行这种调整的方法是众所周知的。通常,这涉及将保守性取代计为部分错配而不是完全错配,从而增加百分比序列同一性。因此,例如,当相同氨基酸的所得评分为1,非保守性取代的所得评分为零时,保守性取代的所得评分介于零与1之间。例如,通过在项目PC/GENE(加利福尼亚州山景城的Intelligenetics公司(Intelligenetics, Mountain View, California))中的实施方式计算保守性取代的评分。

[0089] “序列同一性的百分比”包含指通过在比较窗口上比较两个最佳比对序列确定的值(完全匹配残基的最大数量),其中在比较窗口中的多核苷酸序列部分与参考序列(不包括添加物或缺失部分)相比可以包括添加物或缺失部分(即缺口),以实现两个序列的最佳比对。通过确定在两个序列中出现相同核酸碱基或氨基酸残基的位置数计算百分比来得到匹配位置数,用匹配位置数除以比较窗口中的位置总数,并将结果乘以100以得到序列同一性的百分比。除非另有说明(例如较短的序列包含连接的异源序列),否则所述比较窗口为两个所比较序列中较短序列的全长。

[0090] 除非另有说明,否则序列同一性/相似性值包含使用以下参数使用第10版GAP获得的值:使用GAP权重50和长度权重3以及nwsgapdna.cmp评分矩阵的核苷酸序列的同一性百分比和相似性百分比;使用GAP权重8和长度权重2以及BLOSUM62评分矩阵的氨基酸序列的同一性百分比和相似性百分比;或其任何等效程序。“等效程序”包含当与第10版GAP生成的对应比对进行比较时针对所讨论的任何两个序列产生具有相同核苷酸或氨基酸残基匹配和相同百分比序列同一性的比对的任何序列比较程序。

[0091] 术语“保守性氨基酸取代”是指用具有相似大小、电荷或极性的不同氨基酸取代序列中正常存在的氨基酸。保守性取代的实例包含用如异亮氨酸、缬氨酸或亮氨酸等非极性(疏水性)残基取代另一种非极性残基。同样地,保守性取代的实例包含用一种极性(亲水性)残基取代另一种极性残基,如精氨酸与赖氨酸之间的极性残基、谷氨酰胺与天冬酰胺之间的极性残基或甘氨酸与丝氨酸之间的极性残基。另外,用如赖氨酸、精氨酸或组氨酸等碱性残基取代另一种碱性残基或者用一种如天冬氨酸或谷氨酸等酸性残基取代另一种酸性残基是保守性取代的另外的实例。非保守性取代的实例包含用如异亮氨酸、缬氨酸、亮氨酸、丙氨酸或甲硫氨酸等非极性(疏水性)氨基酸残基取代如半胱氨酸、谷氨酰胺、谷氨酸或赖氨酸等极性(亲水性)残基和/或用极性残基取代非极性残基。典型的氨基酸分类总结在下表1中。

[0092] 表1.氨基酸分类。

[0093]	丙氨酸	Ala	A	非极性	中性	1.8
	精氨酸	Arg	R	极性	阳性	-4.5
	天冬酰胺	Asn	N	极性	中性	-3.5
	天冬氨酸	Asp	D	极性	阴性	-3.5
	半胱氨酸	Cys	C	非极性	中性	2.5
	谷氨酸	Glu	E	极性	阴性	-3.5
	谷氨酰胺	Gln	Q	极性	中性	-3.5
	甘氨酸	Gly	G	非极性	中性	-0.4
	组氨酸	His	H	极性	阳性	-3.2
	异亮氨酸	Ile	I	非极性	中性	4.5
	亮氨酸	Leu	L	非极性	中性	3.8
	赖氨酸	Lys	K	极性	阳性	-3.9
	甲硫氨酸	Met	M	非极性	中性	1.9
	苯丙氨酸	Phe	F	非极性	中性	2.8
	脯氨酸	Pro	P	非极性	中性	-1.6
	丝氨酸	Ser	S	极性	中性	-0.8
	苏氨酸	Thr	T	极性	中性	-0.7
	色氨酸	Trp	W	非极性	中性	-0.9
	酪氨酸	Tyr	Y	极性	中性	-1.3
	缬氨酸	Val	V	非极性	中性	4.2

[0094] “同源”序列(例如,核酸序列)包含与已知参考序列相同或基本上类似的序列,使得其例如与已知参考序列至少50%、至少55%、至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或100%相同。同源序列可以包含例如直系同源序列和旁系同源序列。例如,同源基因通常通过物种形成事件(直系同源基因)或遗传复制事件(旁系同源基因)从共同的祖先DNA序列衍生而来。“直系同源”基因包含不同物种中通过物种形成从共同的祖先基因进化而来的基因。直系同源物通常在进化过程中保留相同的功能。“旁系同源”基因包含通过基因组内的复制相关的基因。旁系同源物可以在进化过程中进化出新的功能。

[0095] 术语“体外(in vitro)”包含人工环境以及在人工环境(例如,试管或分离的细胞或细胞系)内发生的过程或反应。术语“体内”包含自然环境(例如,细胞、生物体或身体)以及在自然环境中发生的过程或反应。术语“离体”包含已从个体体内取出的细胞以及在此类细胞内发生的过程或反应。

[0096] 术语“报告基因”是指具有对基因产物(通常是酶)进行编码的序列的核酸,当包括与异源启动子和/或增强子元件可操作地连接的报告基因序列的构建体被引入到含有(或可以制成含有)启动子和/或增强子元件活化所必需的因子的细胞中时,所述序列可容易且可定量地测定。报告基因的实例包含但不限于对 β -半乳糖苷酶(lacZ)进行编码的基因、细菌氯霉素乙酰转移酶(cat)基因、萤火虫荧光素酶基因、对 β -葡萄糖醛酸酶(GUS)进行编码的基因和对荧光蛋白进行编码的基因。“报告蛋白”是指由报告基因编码的蛋白质。

[0097] 如本文所使用的,术语“荧光报告蛋白”意指可基于荧光检测的报告蛋白,其中荧光可以直接来自报告蛋白、报告蛋白在荧光底物上的活性,或者具有与荧光标记的化合物结合的亲和力的蛋白。荧光蛋白的实例包含绿色荧光蛋白(例如,GFP、GFP-2、tagGFP、turboGFP、eGFP、祖母绿、Azami绿、单体Azami绿、CopGFP、AceGFP和ZsGreen1)、黄色荧光蛋白(例如,YFP、eYFP、柠檬黄、Venus、YPet、PhiYFP和ZsYellow1)、蓝色荧光蛋白(例如,BFP、

eBFP、eBFP2、石青、mKalamal、GFPuv、天蓝色和T-天蓝色)、青色荧光蛋白(例如,CFP、eCFP、蔚蓝色、CyPet、AmCyan1和Midoriishi-青)、红色荧光蛋白(例如,RFP、mKate、mKate2、mPlum、DsRed单体、mCherry、mRFP1、DsRed-表达、DsRed2、DsRed-单体、HcRed-Tandem、HcRed1、AsRed2、eqFP611、mRaspberry、mStrawberry和Jred)、橙色荧光蛋白(例如,mOrange、mK0、Kusabira-橙、单体Kusabira-橙、mTangerine和tdTomato),以及可以通过流式细胞术方法检测到其在细胞中的存在的任何其它合适的荧光蛋白。

[0098] 响应于双链断裂(DSB)的修复主要通过两个保守的DNA修复途径发生:同源重组(HR)和非同源末端接合(NHEJ)。参见Kasperek和Humphrey(2011)《细胞与发育生物研讨会(Semin.Cell Dev.Biol.)》22(8):886-897,所述文献出于所有目的通过引用整体并入本文。同样地,由外源供体核酸介导的靶核酸的修复可以包含两种多核苷酸之间的任何遗传信息交换过程。

[0099] 术语“重组”包含所述两个多核苷酸之间基因信息交换的任何过程,并且可以通过任何机制发生。重组可以通过同源定向修复(HDR)或同源重组(HR)发生。HDR或HR包含可能要求核苷酸序列同源性的核酸修复形式,使用“供体”分子作为模板来修复“靶”分子(即,经历双链断裂的分子),并引导遗传信息从供体到靶标的转移。不希望受任何特定理论的束缚,这种转移可以涉及在断裂的靶与供体之间形成的异源双链DNA的错配校正和/或合成依赖性链退火(synthesis-dependent strand annealing),其中供体用于重新合成将成为靶的一部分的遗传信息和/或相关过程。在一些情况下,供体多核苷酸、供体多核苷酸的一部分、供体多核苷酸的拷贝或供体多核苷酸的拷贝的一部分整合到靶DNA中。参见Wang等人(2013)《细胞(Cell)》153:910-918;Mandalos等人(2012)《公共科学图书馆·综合(PLoS ONE)》7:e45768:1-9;以及Wang等人(2013)《自然生物技术(Nat.Biotechnol.)》31:530-532,所述文献中的每个文献出于所有目的通过引用整体并入本文。

[0100] 非同源末端接合(NHEJ)包含在不需要同源模板的情况下通过将断裂端直接彼此连接或与外源序列连接来修复核酸中的双链断裂。通过NHEJ连接非连续序列通常会导致双链断裂位点附近的缺失、插入或易位。例如,NHEJ还可以通过断裂端与外源供体核酸端的直接连接(即,基于NHEJ的捕获)导致外源供体核酸的靶向整合。当同源定向修复(HDR)途径不易使用时(例如,在非分裂细胞、原代细胞和基于同源性的DNA修复执行较差的细胞中),此类NHEJ介导的靶向整合可以优选用于外源供体核酸的插入)。另外,与同源定向修复相反,不需要关于侧接切割位点的较大序列同一性区域的知识,这在尝试靶向插入到具有基因组序列知识有限的基因组的生物体中时可能是有益的。整合可以通过连接外源供体核酸与经切割的基因组序列之间的平端来进行,或者通过使用外源供体核酸连接粘性端(即,具有5'或3'突出端)来进行,所述外源供体核酸侧接有突出端,所述突出端与由经切割的基因组序列中的核酸酶试剂产生的那些物质相容。参见例如US 2011/020722、WO 2014/033644、WO 2014/089290和Maresca等人(2013)《基因组研究(GenomeRes.)》23(3):539-546,所述文献中的每个文献出于所有目的通过引用整体并入本文。如果平端被连接,则可能需要靶和/或供体切除来产生片段连接所需的微同源区域,这可能在靶序列中产生不期望的改变。

[0101] 术语“抗原结合蛋白”包含与抗原结合的任何蛋白质。抗原结合蛋白的实例包含抗体、抗体的抗原结合片段、多特异性抗体(例如,双特异性抗体)、scFV、双-scFV、双抗体、三抗体、四抗体、V-NAR、VHH、VL、F(ab)、F(ab)₂、DVD(双可变结构域抗原结合蛋白)、SVD(单可

变结构域抗原结合蛋白)、双特异性T细胞衔接蛋白(BiTE)或戴维斯体(Davisbody)(美国专利第8,586,713号,所述文献出于所有目的通过引用整体并入本文)。

[0102] “包括”或“包含”一个或多个所列举的要素的组合物或方法可以包含其它未具体列举的要素。例如,“包括”或“包含”蛋白质的组合物可以单独含有蛋白质或与其它成分组合的蛋白质。过渡短语“基本上由……组成”意指权利要求的范围应被解释为涵盖权利要求中所列举的特定要素以及不会实质上影响所要求保护的发明的基本和新颖特性的那些要素。因此,当在本发明的权利要求中使用术语“基本上由……组成”不旨在被解释为等同于“包括”。

[0103] “任选的”或“任选地”意指随后描述的事件或情形可以发生也可以不发生,并且描述包含事件或情形发生的情况以及事件或情形不发生的情况。

[0104] 数值范围的指定包含所述范围内或定义所述范围的所有整数以及由所述范围内的整数定义的所有子范围。

[0105] 除非上下文另有说明,否则术语“约”涵盖所述值 ± 5 的值。

[0106] 术语“和/或”是指并且涵盖关联的所列项中的一个或多个所列项的任何和所有可能组合以及在以替代性方案(“或”)解释时组合的缺少。

[0107] 术语“或”是指特定列表中的任何一个成员,并且还包含所述列表成员的任何组合。

[0108] 除非上下文另外明确指明,否则本文中的单数形式“一个(a)”、“一种(an)”和“所述(the)”包含复数个提及物。例如,术语“一种蛋白质”或“至少一种蛋白质”可以包含多种蛋白质,包含其混合物。

[0109] 统计学上显著意指 $p \leq 0.05$ 。

具体实施方式

[0110] I. 概述

[0111] 本文公开了包括人源化凝血因子XII(F12)基因座的非人动物基因组、非人动物细胞和非人动物以及使用此类非人动物细胞和非人动物的方法。本文还公开了人源化非人动物F12基因,其包括使非人动物F12基因和核酸酶试剂以及用于使非人动物F12基因人源化的靶向载体人源化的靶向基因修饰。包括人源化F12基因座的非人动物细胞或非人动物表达人凝血因子XII蛋白或包括人凝血因子XII蛋白的一个或多个片段的嵌合凝血因子XII蛋白。此类非人动物细胞和非人动物可以用于在体外、离体或在体内评估靶向人凝血因子XII的药剂(例如,CRISPR/Cas9基因组编辑药剂或抗原结合蛋白)的递送或功效,并且可以用于在体外、离体或在体内优化此类药剂的递送和功效的方法中。

[0112] 在本文所公开的非人动物细胞和非人动物中的一些中,人F12基因组DNA中的大部分或全部被插入到对应的直系同源非人动物F12基因座中。在本文所公开的非人动物细胞和非人动物中的一些中,非人动物F12基因组DNA中的大部分或全部被一对一替换为对应的直系同源人F12基因组DNA。与具有cDNA插入的非人动物相比,当保持内含子-外显子结构和剪接机制时,表达水平应该更高,因为保守的调节元件更有可能保持完整,并且经过RNA加工的剪接转录物比cDNA更稳定。相比之下,将人F12cDNA插入到非人动物F12基因座将消除保守的调节元件。用对应的直系同源人基因组序列替换非人动物基因组序列或者在对应的

直系同源非人F12基因座中插入人F12基因组序列更有可能导致从内源F12基因座忠实表达转基因。例如,人F12基因组序列可能具有驱动较短同种型在神经元中的表达的不同内部调节元件(与小鼠F12基因组序列相比),所述较短同种型是在F12人源化小鼠中重现的表型。在随机基因组基因座而不是内源非人动物F12基因座处转基因插入人凝血因子XII编码序列的转基因非人动物也不会准确地反映F12表达的内源调节。通过用对应的直系同源人基因组DNA一对一替换非人动物基因组DNA中的大部分或全部或者在对应的直系同源非人F12基因座中插入人F12基因组序列而产生的人源化F12等位基因将提供真实人靶标或靶向人凝血因子XII的试剂的真实人靶标的近似靶标(例如,被设计成靶向人F12的CRISPR/Cas9试剂,或被设计成靶向人凝血因子XII的抗体或小分子),由此实现对此类药剂在活体动物中的功效和作用方式的测试以及在人源化蛋白质和人源化基因是存在的凝血因子XII和F12的唯一版本的情况下的药代动力学和药效学研究。

[0113] II. 包括人源化凝血因子XII (F12) 基因座的非人动物

[0114] 本文所公开的非人动物基因组、非人动物细胞和非人动物包括人源化凝血因子XII (F12) 基因座。包括人源化F12基因座的细胞或非人动物表达人凝血因子XII蛋白或天然凝血因子XII蛋白的一个或多个片段已替换为来自人凝血因子XII的对应片段的部分人源化的嵌合凝血因子XII蛋白。

[0115] A. 凝血因子XII (F12)

[0116] 本文所述的细胞和非人动物包括人源化凝血因子XII (F12) 基因座。凝血因子XII (也被称为因子XII、FXII、哈格曼 (Hageman) 因子、 β -因子XIIa第1部分、 β -因子XIIa第2部分、凝血因子XIIa重链或凝血因子XIIa轻链) 由F12基因 (也被称为FXII、HAE3、HAEX或HAF) 编码。凝血因子XII由肝脏中的肝细胞合成并分泌到循环中,在所述循环中其启动接触活化系统。尽管通常被认为仅限于循环,但在阿尔茨海默病和多发性硬化症患者的脑中已发现了因子XII蛋白,并且在人脑和脑脊髓液中观察到了接触系统活化。凝血因子XII是一种启动接触活化系统的循环丝氨酸蛋白酶。在带负电荷的表面上或通过血浆激肽释放酶活化因子XII会产生活化的因子XII (因子XIIa)。因子XIIa启动内在凝血级联反应,从而导致凝血酶生成和凝块形成。因子XIIa还将血浆前激肽释放酶转化为激肽释放酶,这活化了激肽释放酶-激肽途径,从而导致血管活性和促炎肽缓激肽的释放。除了这些经过充分研究的功能外,因子XIIa还被证明在体外活化肝细胞生长因子 (HGF) 和补体系统。因子XII作为单链酶原循环,所述单链酶原在活化期间在R353处切割后变成具有酶活性的双链分子 (α FXIIa)。 α FXIIa被进一步加工成 β FXIIa,其由含有催化结构域的轻链和仅一小部分重链组成。

[0117] 人F12在5号染色体上映射到5q35.3 (NCBI RefSeq基因ID:2161;汇编GRCh38.p12 (GCF_000001405.38);位置NC_000005.10 (177402138..177409576,补体))。据报道,基因具有14个外显子。野生型人凝血因子XII蛋白已被分配UniProt登录号P00748。一种异构体的序列P00748-1 (与NCBI登录号NP_000496.2相同) 在SEQ ID NO:46中示出。对此同种型进行编码的mRNA (cDNA) 被分配NCBI登录号NM_000505.3,并且在SEQ ID NO:34中示出。此同种型的示例性编码序列 (CDS) 的CCDS ID被分配为CCDS34302.1,并且在SEQ ID NO:48中示出。位置619是外显子6中的C的另一个示例性CDS (rs17876030;c.619G>C;p.Ala207Pro) 在SEQ ID NO:13中示出并对SEQ ID NO:5中所示的蛋白质进行编码。SEQ ID NO:5中所示的全长人凝血因子XII蛋白具有615个氨基酸,包含信号肽 (氨基酸1-19)、重链 (氨基酸20-372) 和轻链

(氨基酸373-615)。类似地,SEQ ID NO:46中所示的全长人凝血因子XII蛋白具有615个氨基酸,包含信号肽(氨基酸1-19)、重链(氨基酸20-372)和轻链(氨基酸373-615)。这些结构域之间的描述在UniProt中指定。对人凝血因子XII或F12的参考包含规范(野生型)形式以及所有等位基因形式和同种型。任何其它形式的人凝血因子XII具有针对与野生型形式(SEQ ID NO:5或46)的最大比对进行编码的氨基酸,所比对的氨基酸被指定为相同的编号。

[0118] 小鼠F12在13号染色体上映射到13;13B1 (NCBI RefSeq基因ID:58992;汇编GRCm38.p4 (GCF_000001635.24);位置NC_000079.6 (55417958..55426804,补体))。据报道,基因具有14个外显子。野生型小鼠凝血因子XII蛋白已被分配UniProt登录号Q80YC5。小鼠凝血因子XII的序列(与NCBI登录号NP_067464.2相同)在SEQ ID NO:1中示出。对规范同种型(SEQ ID NO:1)进行编码的示例性mRNA (cDNA) 被分配NCBI登录号NM_021489.3,并且在SEQ ID NO:33中示出。示例性编码序列(CDS) (CCDS ID:CCDS36675.1)在SEQ ID NO:9中示出。SEQ ID NO:1中所示的规范全长小鼠凝血因子XII蛋白具有597个氨基酸,包含信号肽(氨基酸1-19)、重链(氨基酸20-354)和轻链(氨基酸355-597)。这些结构域之间的描述在UniProt中指定。对小鼠凝血因子XII或F12的参考包含规范(野生型)形式以及所有等位基因形式和同种型。任何其它形式的小鼠凝血因子XII具有针对与野生型形式的最大比对进行编码的氨基酸,所比对的氨基酸被指定为相同的编号。

[0119] 许多其它非人动物的凝血因子XII序列也是已知的。这些非人动物包含例如大鼠(UniProt登录号:D3ZTE0;NCBI RefSeq基因ID:306761)、牛(UniProt登录号:P98140;NCBI RefSeq基因ID:280789)、豚鼠(UniProt登录号:Q04962)和猪(UniProt登录号:097507;NCBI RefSeq基因ID:397474)。

[0120] B. 人源化F12基因座

[0121] 人源化F12基因座是内源F12基因座的片段已被缺失并替换为直系同源人F12序列的F12基因座。人源化F12基因座可以是整个F12基因被替换为对应的直系同源人F12序列的F12基因座,或者其可以是仅F12基因的一部分被替换为对应的直系同源人F12序列(即,人源化)的F12基因座。可替代地,人源化F12基因座可以是直系同源人F12基因座的一部分被插入的F12基因座,或者其可以是F12基因的一部分被缺失并且直系同源人F12基因座的一部分被插入的F12基因座。被插入的直系同源人F12基因座的部分可以例如包括比从内源F12基因座中缺失的更多的人F12基因座。例如,内源F12基因座处的整个F12编码序列可以被缺失并替换为对应的人F12序列。对应于内源F12序列的特定区段的人F12序列是指当人F12与内源F12任选地比对时(最大数量为完全匹配的残基)人F12与内源F12序列的特定区段比对的区域。对应的直系同源人序列可以包括例如互补DNA (cDNA) 或基因组DNA。任选地,对应的直系同源人F12序列基于非人动物中的密码子使用被修饰为密码子优化的。经替换的(即,人源化)区域可以包含如外显子等编码区、如内含子等非编码区、非翻译区或调节区(例如,启动子、增强子或转录抑制子结合元件)或其任何组合。作为一个实例,对应于人F12基因的1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、11个、12个、13个或所有14个外显子的外显子可以是人源化的。例如,对应于人F12基因的所有外显子(即,外显子1-14)的外显子可以是人源化的。同样地,对应于人F12基因的1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、11个、12个或所有13个内含子的内含子可以是人源化的或者可以保持内源性。例如,对应于人F12基因的所有内含子(即,内含子1-13)的内含子可以是人源化的。包含调节序列

的侧接非翻译区也可以是人源化的或者保持内源性。例如,5'非翻译区(UTR)、3'UTR或5'UTR和3'UTR两者可以是人源化的,或者5'UTR、3'UTR或5'UTR和3'UTR两者可以保持内源性。在具体实例中,5'UTR和3'UTR两者保持内源性。根据直系同源序列替换的程度,如启动子等调节序列可以是内源的或由替换的人直系同源序列提供。例如,人源化F12基因座可以包含内源非人动物F12启动子(即,人F12序列可以可操作地连接到内源非人动物启动子)。

[0122] 对信号肽、重链或轻链进行编码的区域中的一个或多个或所有区域可以是人源化的,或者此类区域中的一个或多个区域可以保持内源性。小鼠凝血因子XII信号肽、重链和轻链的示例性编码序列分别在SEQ ID NO:10-12中示出。人凝血因子XII信号肽、重链和轻链的示例性编码序列分别SEQ ID NO:14-16中示出。人凝血因子XII重链的另一个示例性编码序列在SEQ ID NO:49中示出。

[0123] 例如,对信号肽进行编码的F12基因座中的全部或部分区域可以是人源化的,和/或对重链进行编码的F12基因座中的全部或部分区域可以是人源化的,和/或对轻链进行编码的F12基因座中的全部或部分区域可以是人源化的。可替代地或另外,对信号肽进行编码的F12基因座中的全部或部分区域可以保持内源性,和/或对重链进行编码的F12基因座中的全部或部分区域可以保持内源性,和/或对轻链进行编码的F12基因座中的全部或部分区域可以保持内源性。在一个实例中,对信号肽、重链和轻链进行编码的F12基因座的全部或部分区域是人源化的。任选地,F12基因座的人源化区域的CDS包括以下、基本上由以下组成或由以下组成:与SEQ ID NO:13(或其简并变体)至少约85%、至少约90%、至少约95%、至少约96%、至少约97%、至少约98%、至少约99%或约100%相同的序列。任选地,F12基因座的人源化区域的CDS包括以下、基本上由以下组成或由以下组成:与SEQ ID NO:48(或其简并变体)至少约85%、至少约90%、至少约95%、至少约96%、至少约97%、至少约98%、至少约99%或约100%相同的序列。任选地,F12基因座的人源化区域包括与SEQ ID NO:31和/或32至少约85%、至少约90%、至少约95%、至少约96%、至少约97%、至少约98%、至少约99%或约100%相同的序列。任选地,人源化F12基因座对蛋白质进行编码,所述蛋白质包括以下、基本上由以下组成或由以下组成:与SEQ ID NO:5或46至少约85%、至少约90%、至少约95%、至少约96%、至少约97%、至少约98%、至少约99%或约100%相同的序列。任选地,人源化F12基因座包括以下、基本上由以下组成或由以下组成:与SEQ ID NO:17或18至少约85%、至少约90%、至少约95%、至少约96%、至少约97%、至少约98%、至少约99%或约100%相同的序列。在每种情况下,人源化凝血因子XII蛋白可以保留天然凝血因子XII蛋白和/或人凝血因子XII蛋白的活性。

[0124] 由人源化F12基因座编码的凝血因子XII蛋白可以包括来自人凝血因子XII蛋白的一个或多个结构域和/或来自内源(即天然)凝血因子XII蛋白的一个或多个结构域。小鼠凝血因子XII信号肽、重链和轻链的示例性氨基酸序列分别在SEQ ID NO:2-4中示出。人凝血因子XII信号肽、重链和轻链的示例性氨基酸序列分别在SEQ ID NO:6-8中示出。人凝血因子XII重链的另一个示例性氨基酸序列在SEQ ID NO:47中示出。

[0125] 凝血因子XII蛋白可以包括人凝血因子XII信号肽、人凝血因子XII重链和人凝血因子XII轻链中的一个或多个或全部。可替代地或另外,凝血因子XII蛋白可以包括来自内源(即天然)非人动物凝血因子XII蛋白的一个或多个结构域。例如,凝血因子XII蛋白可以包括来自内源(即天然)非人动物凝血因子XII蛋白的信号肽和/或来自内源(即天然)非人

动物凝血因子XII蛋白的重链和/或来自内源(即天然)非人动物凝血因子XII蛋白的轻链。作为一个实例,凝血因子XII蛋白可以包括人信号肽、重链和轻链。

[0126] 嵌合凝血因子XII蛋白中来自人凝血因子XII蛋白的结构域可以由完全人源化的序列编码(即,对所述结构域进行编码的整个序列被替换为直系同源人F12序列)或者可以由部分人源化的序列编码(即,对所述结构域进行编码的序列中的一些序列被替换为直系同源人F12序列,而对所述结构域进行编码的其余内源(即天然)序列对与直系同源人F12序列相同的氨基酸进行编码,使得经编码的结构域与人凝血因子XII蛋白中的所述结构域相同)。同样地,嵌合蛋白中来自内源凝血因子XII蛋白的结构域可以由完全内源序列编码(即,对所述结构域进行编码的整个序列是内源F12序列)或者可以由部分人源化序列编码(即,对所述结构域进行编码的序列中的一些序列被替换为直系同源人F12序列,但直系同源人F12序列对与经替换的内源F12序列相同的氨基酸进行编码,使得经编码的结构域与内源凝血因子XII蛋白中的所述结构域相同)。

[0127] 作为一个实例,由人源化F12基因座编码的凝血因子XII蛋白可以包括人凝血因子XII信号肽。任选地,人凝血因子XII信号肽包括以下、基本上由以下组成或由以下组成:与SEQ ID NO:6至少85%、90%、95%、96%、97%、98%、99%或100%相同的序列。任选地,人凝血因子XII信号肽的编码序列包括以下、基本上由以下组成或由以下组成:与SEQ ID NO:14至少85%、90%、95%、96%、97%、98%、99%或100%相同的序列。作为另一个实例,由人源化F12基因座编码的凝血因子XII蛋白可以包括人凝血因子XII重链。任选地,人凝血因子XII重链包括以下、基本上由以下组成或由以下组成:与SEQ ID NO:7或47至少约85%、至少约90%、至少约95%、至少约96%、至少约97%、至少约98%、至少约99%或约100%相同的序列。任选地,人凝血因子XII重链的编码序列包括以下、基本上由以下组成或由以下组成:与SEQ ID NO:15或49至少约85%、至少约90%、至少约95%、至少约96%、至少约97%、至少约98%、至少约99%或约100%相同的序列。作为另一个实例,由人源化F12基因座编码的凝血因子XII蛋白可以包括人轻链。任选地,人轻链包括以下、基本上由以下组成或由以下组成:与SEQ ID NO:8至少约85%、至少约90%、至少约95%、至少约96%、至少约97%、至少约98%、至少约99%或约100%相同的序列。任选地,人轻链的编码序列包括以下、基本上由以下组成或由以下组成:与SEQ ID NO:16至少约85%、至少约90%、至少约95%、至少约96%、至少约97%、至少约98%、至少约99%或约100%相同的序列。例如,由人源化F12基因座编码的凝血因子XII蛋白可以包括以下、基本上由以下组成或由以下组成:与SEQ ID NO:5或46至少约85%、至少约90%、至少约95%、至少约96%、至少约97%、至少约98%、至少约99%或约100%相同的序列。任选地,由人源化F12基因座编码的F12 CDS可以包括以下、基本上由以下组成或由以下组成:与SEQ ID NO:13或48(或其简并变体)至少约85%、至少约90%、至少约95%、至少约96%、至少约97%、至少约98%、至少约99%或约100%相同的序列。在每种情况下,人源化凝血因子XII蛋白可以保留天然凝血因子XII蛋白和/或人凝血因子XII蛋白的活性。

[0128] 任选地,人源化F12基因座可以包括其它元件。此类元件的实例可以包含选择盒、报告基因、重组酶识别位点或其它元件。在一个实例中,另外的元件(例如,选择盒)定位在人源化F12基因座处插入的人序列的内含子处。在另一个实例中,另外的元件(例如,选择盒)定位在人源化F12基因座处插入的人序列的3'处。可替代地,人源化F12基因座可能缺少

其它元件(例如,可能缺少选择标志物或选择盒)。合适的报告基因和报告蛋白的实例在本文别处公开。合适的选择标志物的实例包含新霉素磷酸转移酶(neo^r)、潮霉素B磷酸转移酶(hyg^r)、嘌呤霉素-N-乙酰转移酶(puro^r)、杀稻瘟素S脱氨酶(bsr^r)、黄嘌呤/鸟嘌呤磷酸核糖转移酶(gpt)和单纯疱疹病毒胸苷激酶(HSV-k)。重组酶的实例包含Cre、Flp和Dre重组酶。Cre重组酶基因的一个实例是Crei,其中两个对Cre重组酶进行编码的外显子被内含子隔开,以防止其在原核细胞中表达。此类重组酶可以进一步包括用于促进定位到核(例如,NLS-Crei)的核定位信号。重组酶识别位点包含由位点特异性重组酶识别并且可以用作重组事件的底物的核苷酸序列。重组酶识别位点的实例包含FRT、FRT11、FRT71、attp、att、rox和lox位点,如loxP、lox511、lox2272、lox66、lox71、loxM2和lox5171。

[0129] 如报告基因或选择盒等其它元件可以是侧接有重组酶识别位点的自缺失盒。参见例如US 8,697,851和US 2013/0312129,所述文献中的每个文献出于所有目的通过引用整体并入本文。作为实例,自缺失盒可以包括可操作地连接到小鼠Prm1启动子的Crei基因(包括对Cre重组酶进行编码的由内含子分开的两个外显子)和可操作地连接到人泛素启动子的新霉素抗性基因。通过采用Prm1启动子,可以在F0动物的雄性生殖细胞中特异性地缺失自缺失盒。对选择标志物进行编码的多核苷酸可以可操作地连接到在被靶向的细胞中具有活性的启动子。启动子的实例在本文别处描述。作为另一个具体实例,自缺失选择盒可以包括可操作地连接到一个或多个启动子(例如,人泛素和EM7启动子两者)的潮霉素抗性基因编码序列,接着是多聚腺苷酸化信号,接着是可操作地连接到一个或多个启动子(例如,mPrm1启动子)的Crei编码序列,接着是另一个聚腺苷酸化信号,其中整个盒侧接有loxP位点。

[0130] 人源化F12基因座也可以是条件等位基因。例如,条件等位基因可以是多功能等位基因,如US 2011/0104799中所述,所述文献出于所有目的通过引用整体并入本文。例如,条件等位基因可以包括:(a)相对于基因转录在有义取向上的启动序列;(b)在有义或反义取向上的药物选择盒(DSC);(c)在反义取向上的所关注核苷酸序列(NSI);以及(d)在相反取向上的条件反转模块(COIN,其利用外显子分裂内含子和可逆基因捕获样模块)。参见例如US 2011/0104799。条件等位基因可以进一步包括可重组单元,所述可重组单元在暴露于第一重组酶后重组以形成条件等位基因,所述条件等位基因(i)缺少启动序列和DSC;并且(ii)含有有义取向上的NSI和反义取向上的COIN。参见例如US 2011/0104799。

[0131] 一种示例性人源化F12基因座(例如,人源化小鼠F12基因座)是从起始密码子到终止密码子的区域被替换为对应的人序列的基因座。参见图1以及SEQ ID NOS:17和18。在一个具体实例中,人源化内源F12基因座处的人F12序列包括与SEQ ID NO:31或32中所示的序列至少约85%、至少约90%、至少约95%、至少约96%、至少约97%、至少约98%、至少约99%或约100%列相同的序列。在另一个具体实例中,人源化内源F12基因座对蛋白质进行编码,所述蛋白质包括与SEQ ID NO:5或46中所示的序列至少约85%、至少约90%、至少约95%、至少约96%、至少约97%、至少约98%、至少约99%或约100%相同的序列。在另一个具体实例中,人源化内源F12基因座包括编码序列,所述编码序列包括与SEQ ID NO:13和/或48中所示的序列至少约85%、至少约90%、至少约95%、至少约96%、至少约97%、至少约98%、至少约99%或约100%相同的序列。在另一个具体实例中,人源化内源F12基因座包括与SEQ ID NO:17或18中所示的序列至少约85%、至少约90%、至少约95%、至少约96%、至

少约97%、至少约98%、至少约99%或约100%相同的序列。

[0132] C.包括人源化凝血因子XII (F12) 基因座的非人动物基因组、非人动物细胞和非人动物

[0133] 提供了如本文别处所述的包括人源化凝血因子XII (F12) 基因座的非人动物基因组、非人动物细胞和非人动物。基因组、细胞或非人动物可以是雄性或雌性。基因组、细胞或非人动物对于人源化F12基因座可以是杂合的也可以是纯合的。二倍体生物在每个基因座处有两个等位基因。每对等位基因表示特异性遗传基因座的基因型。如果在特定基因座处有两个相同的等位基因,则基因型被描述为纯合的,如果两个等位基因不同,则基因型被描述为杂合的。包括人源化F12基因座的非人动物可以在其种系中包括人源化内源F12基因座。

[0134] 本文所提供的非人动物基因组或细胞可以是例如包括F12基因座或与人F12基因座同源或直系同源的基因组基因座的任何非人动物基因组或细胞。基因组可以来自真核细胞或者细胞可以是真核细胞,所述真核细胞包含例如真菌细胞(例如,酵母)、植物细胞、动物细胞、哺乳动物细胞、非人哺乳动物细胞和人细胞。术语“动物”包含动物界的任何成员,包含例如哺乳动物、鱼类、爬行动物、两栖动物、鸟类和蠕虫。哺乳动物细胞可以是例如非人哺乳动物细胞、啮齿动物细胞、大鼠细胞、小鼠细胞或仓鼠细胞。其它非人哺乳动物包含例如非人灵长类动物、猴子、猿、猩猩、猫、狗、兔、马、公牛、鹿、野牛、家畜(例如,牛种,如奶牛、公牛等;绵羊种,如绵羊、山羊等;以及猪种,如猪和野猪)。鸟类包含例如鸡、火鸡、鸵鸟、鹅、鸭等。家养动物和农业动物也包含在内。术语“非人”不包含人。

[0135] 细胞也可以呈任何类型的未分化或分化状态。例如,细胞可以是全能细胞、多能细胞(例如,人多能细胞或非人多能细胞,如小鼠胚胎干(ES)细胞或大鼠ES细胞)或非多能细胞。全能细胞包含可以产生任何细胞类型的未分化细胞,并且多能细胞包含具有发育成多于一种分化细胞类型的能力的未分化细胞。此类多能和/或全能细胞可以是例如ES细胞或ES样细胞,如诱导多能干(iPS)细胞。ES细胞包含在引入到胚胎中时能够对发育胚胎的任何组织做出贡献的胚胎衍生的全能或多能细胞。ES细胞可以源自囊胚的内细胞团,并且能够分化成三种脊椎动物胚层(内胚层、外胚层和中胚层)中的任何层的细胞。

[0136] 本文所提供的细胞还可以是生殖细胞(例如,精子或卵母细胞)。细胞可以是有丝分裂感受态细胞或有丝分裂非活性细胞、减数分裂感受态细胞或减数分裂非活性细胞。类似地,细胞还可以是初生体细胞或不是初生体细胞的细胞。体细胞包含任何不是配子、生殖细胞、配子母细胞或未分化干细胞的细胞。例如,细胞可以是肝细胞(liver cell),如成肝细胞或肝细胞(hepatocyte)。可替代地,细胞可以是神经元细胞,如锥体神经元细胞或皮层神经元细胞。

[0137] 本文所提供的合适的细胞还包含原代细胞。原代细胞包含直接从生物体、器官或组织中分离的细胞或细胞培养物。原代细胞包含既不转化也不永生的细胞。所述原代细胞包含从生物体、器官或组织中获得的所有细胞,所述细胞先前未在组织培养物中进行传代,或者先前已经在组织培养物中进行传代但不能无限期地在组织培养中进行传代。此类细胞可以通过常规技术分离并且包含例如肝细胞或神经元细胞。

[0138] 本文所提供的其它合适的细胞包含永生化细胞。永生化细胞包含来自多细胞生物体的细胞,其通常不会无限增殖但由于突变或改变而逃避正常细胞衰老并且替代地可以继

续进行分裂。此类突变或改变可以天然存在或有意诱导。永生化细胞系的具体实例是HepG2人肝癌细胞系。多种类型的永生化细胞是众所周知的。永生化或原代细胞包含通常用于培养或表达重组基因或蛋白质的细胞。

[0139] 本文所提供的细胞还包含单细胞阶段胚胎(即,受精的卵母细胞或受精卵)。此类单细胞阶段胚胎可以来自任何遗传背景(例如,对于小鼠,BALB/c、C57BL/6、129或其组合)、可以是新鲜的或冷冻的并且可以衍生自自然配种或体外受精。

[0140] 本文所提供的细胞可以是正常的、健康的细胞,或者可以是患病或携带突变体的细胞。

[0141] 包括如本文所述的人源化F12基因座的非人动物可以通过本文别处所述的方法制备。术语“动物”包含动物界的任何成员,包含例如哺乳动物、鱼类、爬行动物、两栖动物、鸟类和蠕虫。在具体实例中,非人动物是非人哺乳动物。非人哺乳动物包含例如非人灵长类动物、猴子、猿、猩猩、猫、狗、马、公牛、鹿、野牛、绵羊、兔、啮齿动物(例如,小鼠、大鼠、仓鼠和豚鼠)和家畜(例如,牛种,如奶牛和公牛;绵羊种,如绵羊和山羊;以及猪种,如猪和野猪)。鸟类包含例如鸡、火鸡、鸵鸟、鹅和鸭。家养动物和农业动物也包含在内。术语“非人动物”不包含人。优选的非人动物包含例如啮齿动物,如小鼠和大鼠。

[0142] 非人动物可以来自任何遗传背景。例如,合适的小鼠可以来自129品系、C57BL/6品系、129和C57BL/6的混合、BALB/c品系或Swiss Webster品系。129品系的实例包含129P1、129P2、129P3、129X1、129S1(例如,129S1/SV,129S1/Sv1m)、129S2、129S4、129S5、129S9/SvEvH、129S6(129/SvEvTac)、129S7、129S8、129T1和129T2。参见例如Festing等人(1999)《哺乳动物基因组(Mamm.Genome)》10(8):836,所述文献出于所有目的通过引用整体并入本文。C57BL品系的实例包含C57BL/A、C57BL/An、C57BL/GrFa、C57BL/Ka1_wN、C57BL/6、C57BL/6J、C57BL/6ByJ、C57BL/6NJ、C57BL/10、C57BL/10ScSn、C57BL/10Cr和C57BL/01a。合适的小鼠还可以来自上述129品系和上述C57BL/6品系的混合(例如,50%129和50%C57BL/6)。同样地,合适的小鼠可以来自上述129品系的混合或上述BL/6品系的混合(例如,129S6(129/SvEvTac)品系)。

[0143] 类似地,大鼠可以来自任何大鼠品系,包含例如ACI大鼠品系、黑刺鼠(DA)大鼠品系、威斯塔(Wistar)大鼠品系、LEA大鼠品系、斯泼累格多雷(Sprague Dawley,SD)大鼠品系或费舍尔(Fischer)大鼠品系,如费舍尔F344或费舍尔F6。大鼠还可以从源自上述两种或更多种品系的混合品系中获得。例如,合适的大鼠可以来自DA品系或ACI品系。ACI大鼠品系的特征在于具有腹部和足部呈白色的黑刺鼠以及RT1^{av1}单倍型。此类品系可从多种来源获得,包含哈兰实验室(Harlan Laboratories)。黑刺鼠(DA)大鼠品系的特征在于具有刺鼠皮毛和RT1^{av1}单倍型。此类大鼠可从多种来源获得,包含查尔斯河和哈兰实验室(Charles River and Harlan Laboratories)。一些合适的大鼠可以来自近交大鼠品系。参见例如US 2014/0235933,所述文献出于所有目的通过引用整体并入本文。

[0144] 包括如本文所述的人源化F12基因座的非人动物可以在脑中如在脑的神经元中表达短FXII同种型。例如,短FXII同种型可以由F12 mRNA编码,所述F12 mRNA由针对外显子11-14的探针检测,但不由针对外显子1-6的探针检测。在具体实例中,短同种型可以是FXII₂₉₇₋₅₉₆,一种含有FXII的富脯氨酸和催化结构域的蛋白质(M297-S596;SEQ ID NO:74,由SEQ ID NO:75编码)。例如,短FXII同种型可以由血浆激肽释放酶活化,并且可以例如将

pro-HGF转化为活性HGF。例如,短FXII同种型也可以在由激肽释放酶切割后被活化。例如,活化的短FXII同种型可以将前激肽释放酶相互活化为激肽释放酶。例如,短FXII同种型能够启动激肽释放酶-激肽途径。例如,短FXII同种型活化可以导致缓激肽的产生。

[0145] 包括如本文所述的人源化F12基因座的非人动物可以具有与对应的野生型非人动物中的活化部分凝血活酶时间(aPTT)凝血时间无显著差异的aPTT凝血时间。同样地,包括如本文所述的人源化F12基因座的非人动物可以具有与对应的野生型非人动物中的前凝血酶时间(PT)值无显著差异的PT值。

[0146] 包括如本文所述的人源化F12基因座的非人动物在血浆中的人凝血因子XII的水平可以在约1到约30 $\mu\text{g/mL}$ 之间、约1到约25 $\mu\text{g/mL}$ 之间、约1到约20 $\mu\text{g/mL}$ 之间、约2到约30 $\mu\text{g/mL}$ 之间、约2到约25 $\mu\text{g/mL}$ 之间、约2到约20 $\mu\text{g/mL}$ 之间、约3到约30 $\mu\text{g/mL}$ 之间、约3到约25 $\mu\text{g/mL}$ 之间、约3到约20 $\mu\text{g/mL}$ 之间、约4到约30 $\mu\text{g/mL}$ 之间、约4到约25 $\mu\text{g/mL}$ 之间、约4到约20 $\mu\text{g/mL}$ 之间、约5到约30 $\mu\text{g/mL}$ 之间、约5到约25 $\mu\text{g/mL}$ 之间或约5到约20 $\mu\text{g/mL}$ 之间。例如,非人动物中的人凝血因子XII可以在约5到约18 $\mu\text{g/mL}$ 之间。可替代地,包括如本文所述的人源化F12基因座的非人动物在血浆中的人凝血因子XII的水平可以为至少约1 $\mu\text{g/mL}$ 、至少约2 $\mu\text{g/mL}$ 、至少约3 $\mu\text{g/mL}$ 、至少约4 $\mu\text{g/mL}$ 、至少约5 $\mu\text{g/mL}$ 、至少约6 $\mu\text{g/mL}$ 、至少约7 $\mu\text{g/mL}$ 、至少约8 $\mu\text{g/mL}$ 、至少约9 $\mu\text{g/mL}$ 或至少约10 $\mu\text{g/mL}$ 。

[0147] III. 使用包括人源化凝血因子XII (F12) 基因座的非人动物在体内或离体评估靶向人凝血因子XII的试剂的功效的方法

[0148] 提供了用于使用包括如本文别处所述的人源化F12基因座的非人动物在体内或离体评估或优化靶向人凝血因子XII的试剂(例如,治疗性分子或复合物)的递送或功效的各种方法。由于非人动物包括人源化F12基因座,因此非人动物将更准确地反映靶向人凝血因子XII的试剂的活性。此类非人动物对于测试被设计成靶向人F12基因的基因组编辑试剂特别有用,因为本文所公开的非人动物包括人源化内源F12基因座而不是人F12序列在随机基因组基因座处的转基因插入,并且人源化内源F12基因座可以包括来自编码区和非编码区两者的直系同源人基因组F12序列而不是人工cDNA序列。

[0149] A. 在体内或离体测试靶向人凝血因子XII的试剂的活性

[0150] 提供了用于使用包括如本文别处所述的人源化F12基因座的非人动物在体内评估靶向人凝血因子XII的试剂的递送或功效的各种方法。此类方法可以包括:(a) 向非人动物中引入靶向人凝血因子XII的试剂(即,将靶向人凝血因子XII的试剂施用于非人动物);以及(b) 评估靶向人凝血因子XII的试剂的活性。

[0151] 靶向人凝血因子XII的试剂可以是靶向人F12基因座(人F12基因)、人F12 mRNA或人凝血因子XII蛋白的任何生物或化学试剂。靶向人凝血因子XII的试剂的实例在本文别处公开,并且包含例如基因组编辑试剂或抗原结合蛋白。例如,靶向人凝血因子XII的试剂可以是F12靶向核酸(例如,CRISPR/Cas9向导RNA、短发夹RNA(shRNA)或小干扰RNA(siRNA))或对F12靶向蛋白(例如,Cas蛋白,如Cas9、ZFN或TALEN)进行编码的核酸。可替代地,人F12靶向试剂可以是凝血因子XII靶向抗体或抗原结合蛋白,或靶向人凝血因子XII的任何其它大分子或小分子。在一个实例中,靶向人凝血因子XII的试剂是基因组编辑试剂,如核酸酶试剂和/或外源供体核酸(例如,靶向载体)。

[0152] 此类靶向人凝血因子XII的试剂可以通过本文别处更详细地公开的任何递送方法

(例如, AAV、LNP、HDD或注射)和通过任何施用途来施用。递送治疗复合物和分子的方式以及施用途在本文别处更详细地公开。在特定方法中, 试剂通过AAV介导的递送进行递送。例如, AAV8可以用于靶向肝脏。在其它特定方法中, 试剂通过LNP介导的递送进行递送。在其它特定方法中, 试剂通过流体动力学递送(HDD)进行递送。剂量可以是任何合适的剂量。例如, 在试剂(例如, Cas9 mRNA和gRNA)通过LNP介导的递送进行递送的一些方法中, 剂量可以在约0.01与约10mg/kg之间、约0.01与约5mg/kg之间、约0.01与约4mg/kg之间、约0.01与约3mg/kg之间、约0.01与约2mg/kg之间、约0.01与约1mg/kg、约0.1与约10mg/kg之间、约0.1与约6mg/kg之间、约0.1与约5mg/kg之间、约0.1与约4mg/kg之间、约0.1与约3mg/kg之间、约0.1与约2mg/kg之间、约0.1与约1mg/kg之间、约0.3与约10mg/kg之间、约0.3与约6mg/kg之间、约0.3与约5mg/kg之间、约0.3与约4mg/kg之间、约0.3与约3mg/kg之间、约0.3与约2mg/kg之间、约0.3与约1mg/kg之间、约0.1mg/kg、约0.3mg/kg、约1mg/kg、约2mg/kg或约3mg/kg。在具体实例中, 剂量在约0.1与约6mg/kg之间、约0.1与约3mg/kg之间或约0.1与约2mg/kg之间。

[0153] 用于评估靶向人凝血因子XII的试剂的活性的方法是众所周知的并且在本文别处提供。活性评估可以在如本文别处公开的任何细胞类型、任何组织类型或任何器官类型中进行。在一些方法中, 活性评估在血浆中进行。在一些方法中, 活性评估在肝细胞中进行。如果凝血因子XII靶向试剂是基因组编辑试剂(例如, 核酸酶试剂), 则此类方法可以包括评估人源化F12基因座的修饰。作为一个实例, 评估可以包括测量人源化F12基因座处的非同源末端接合(NHEJ)活性。这可以包括例如测量人源化F12基因座内插入或缺失的频率。例如, 评估可以包括对从非人动物中分离出的一个或多个细胞中的人源化F12基因座进行测序(例如, 下一代测序)。评估可以包括从非人动物中分离出靶器官(例如, 肝脏)或组织并评估靶器官或组织中的人源化F12基因座的修饰。评估还可以包括评估靶器官或组织内两种或更多种不同细胞类型中的人源化F12基因座的修饰。类似地, 评估可以包括从非人动物中分离出非靶器官或组织(例如, 两个或多个非靶器官或组织)并评估非靶器官或组织中的人源化F12基因座的修饰。

[0154] 此类方法还可以包括测量由人源化F12基因座产生的mRNA的表达水平, 或测量由人源化F12基因座编码的蛋白质的表达水平。例如, 可以测量特定细胞、组织或器官类型(例如, 肝脏)中的蛋白质水平, 或者可以测量血浆或血清中的分泌水平。用于评估从人源化F12基因座表达的F12 mRNA或蛋白质的表达的方法在本文别处提供并且是众所周知的。

[0155] 作为一个具体实例, 如果靶向人凝血因子XII的试剂是基因组编辑试剂(例如, 核酸酶试剂), 则可以(例如, 在肝细胞中)评估人源化F12基因座处的编辑百分比(例如, 在PCR反应中从裂解细胞池中读取的序列总数上观察到的插入或缺失总数)。

[0156] 以上提供的用于在体内评估活性的各种方法还可以用于评估如本文别处所述的靶向人凝血因子XII的试剂的活性。

[0157] 在一些方法中, 靶向人凝血因子XII的试剂是靶向人F12基因的核酸酶试剂, 如CRISPR/Cas核酸酶试剂。此类方法可以包括, 例如: (a) 向非人动物中引入被设计成切割人F12基因的核酸酶试剂(例如, 如Cas9等Cas蛋白和被设计成靶向人F12基因中的向导RNA靶序列的向导RNA); 以及(b) 评估人源化F12基因座的修饰。

[0158] 例如, 在CRISPR/Cas核酸酶的情况下, 当向导RNA与Cas蛋白形成复合物并将Cas蛋

白引导到人源化F12基因座,并且Cas/向导RNA复合物切割向导RNA靶序列,从而触发细胞的修复(例如,如果不存在供体序列,则通过非同源末端接合(NHEJ)触发)时,人源化F12基因座的修饰将被诱导。

[0159] 任选地,可以引入两个或更多个向导RNA,各自被设计成靶向人F12基因内不同的向导RNA靶序列。例如,两个向导RNA可以设计成切除两个向导RNA靶序列之间的基因组序列。当第一向导RNA与Cas蛋白形成复合物并将Cas蛋白引导到人源化F12基因座,第二向导RNA与Cas蛋白形成复合物并将Cas蛋白引导到人源化F12基因座,第一Cas/向导RNA复合物切割第一向导RNA靶序列,并且第二Cas/向导RNA复合物切割第二向导RNA靶序列,从而导致插入序列的切除时,人源化F12基因座的修饰将被诱导。

[0160] 另外或可替代地,能够与人F12基因重组并对其进行修饰的外源供体核酸(例如,靶向载体)也被引入到非人动物中。任选地,核酸酶试剂或Cas蛋白可以栓系到如本文别处所述的外源供体核酸。例如,当向导RNA与Cas蛋白形成复合物并将Cas蛋白引导到人源化F12基因座,Cas/向导RNA复合物切割向导RNA靶序列,并且人源化F12基因座与外源供体核酸重组以修饰人源化F12基因座时,人源化F12基因座的修饰将被诱导。外源供体核酸可以例如通过同源定向修复(HDR)或通过NHEJ介导的插入与人源化F12基因座重组。可以使用任何类型的外源供体核酸,其实例在本文别处提供。

[0161] B.在体内或离体优化靶向人凝血因子XII的试剂的递送或功效的方法

[0162] 提供了用于优化靶向人凝血因子XII的试剂到细胞或非人动物的递送或在体内优化靶向人凝血因子XII的试剂的活性或功效的各种方法。此类方法可以包括,例如:(a)在包括人源化F12基因座的第一非人动物或第一细胞中首次执行测试如上所述的靶向人凝血因子XII的试剂的功效的方法;(b)改变变量并用改变的变量在包括人源化F12基因座的第二非人动物(即,相同物种的)或第二细胞中第二次执行所述方法;以及(c)将步骤(a)中的靶向人凝血因子XII的试剂的活性与步骤(b)中的靶向人凝血因子XII的试剂的活性进行比较,并选择产生更高活性的方法。

[0163] 测量靶向人凝血因子XII的试剂的递送、功效或活性的方法在本文别处公开。例如,此类方法可以包括测量人源化F12基因座的修饰。取决于非人动物或细胞内的期望效果,人源化F12基因座的更有效的修饰可以意指不同的事情。例如,人源化F12基因座的更有效的修饰可以意指更高水平的修饰、更高的精确度、更高的一致性或更高的特异性中的一个或多个或全部。人源化F12基因座的更高水平的修饰(即,更高的功效)是指在特定靶细胞类型内、在特定靶组织内或在特定靶器官(例如,肝脏)内靶向更高百分比的细胞。更高的精确度是指人源化F12基因座的更精确的修饰(例如,更高百分比的靶向细胞具有相同的修饰或具有期望的修饰而没有额外的非预期插入和缺失(例如,NHEJ插入缺失))。更高的一致性是指在多于一种类型的细胞、组织或器官被靶向的情况下,不同类型的靶向细胞、组织或器官之间的人源化F12基因座的更一致的修饰(例如,肝脏内更多数量的细胞类型的修饰)。如果靶向特定器官,则更高的一致性也可以指在器官(例如,肝脏)内的所有位置中更一致的修饰。更高的特异性可以指对于所靶向的基因组基因座或基因座的更高的特异性、对于所靶向的细胞类型的更高的特异性、对于所靶向的组织类型的更高的特异性或者对于所靶向的器官的更高的特异性。例如,增加的基因组基因座特异性是指脱靶基因组基因座的修饰较少(例如,代替或除了靶基因组基因座的修饰之外,在非预期的脱靶基因组基因座处具有

修饰的靶向细胞的百分比较低)。同样地,增加的细胞类型、组织或器官类型特异性是指在靶向特定的细胞类型、组织类型或器官类型的情况下(例如,当靶向特定器官(例如,肝脏)时,器官或组织中不是预期靶标的细胞的修饰较少),脱靶细胞类型、组织类型或器官类型的修饰较少。

[0164] 可替代地,此类方法可以包括测量F12 mRNA或凝血因子XII蛋白的表达。在一个实例中,更有效的靶向人凝血因子XII的药剂导致F12 mRNA或凝血因子XII蛋白表达的更大降低。可替代地,此类方法可以包括测量凝血因子XII活性。在一个实例中,更有效的靶向人凝血因子XII的药剂导致凝血因子XII活性的更大降低。

[0165] 改变的变量可以是任何参数。作为一个实例,改变的变量可以是靶向人凝血因子XII的试剂引入到细胞或非人动物中的包装或递送方法(例如,递送媒介)。如LNP、HDD和AAV等递送方法或递送媒介的实例在本文别处公开。例如,改变的变量可以是AAV血清型。类似地,施用可以包括LNP介导的递送,并且改变的变量可以是LNP调配物。作为另一个实例,改变的变量可以是用于将靶向人凝血因子XII的试剂引入到细胞或非人动物中的施用途径。如静脉内、玻璃体内、实质内和鼻内滴注等施用途径的实例在本文别处公开。

[0166] 作为另一个实例,改变的变量可以是所引入的靶向人凝血因子XII的试剂的浓度或量。作为另一个实例,改变的变量可以是所引入的一种靶向人凝血因子XII的试剂(例如,向导RNA、Cas蛋白、外源供体核酸、RNAi 药剂或ASO)的浓度或量相对于所引入的另一种靶向人凝血因子XII的试剂(例如,向导RNA、Cas蛋白、外源供体核酸、RNAi 药剂或ASO)的浓度或量。

[0167] 作为另一个实例,改变的变量可以是引入靶向人凝血因子XII的试剂的时间相对于评估试剂活性或功效的时间。作为另一个实例,改变的变量可以是引入靶向人凝血因子XII的试剂的次数或频率。作为另一个实例,改变的变量可以是所引入的一种靶向人凝血因子XII的试剂(例如,向导RNA、Cas蛋白、外源供体核酸、RNAi 药剂或ASO)的引入时间相对于所引入的另一种靶向人凝血因子XII的试剂(例如,向导RNA、Cas蛋白、外源供体核酸、RNAi 药剂或ASO)的引入时间。

[0168] 作为另一个实例,改变的变量可以是引入靶向人凝血因子XII的试剂的形式。例如,向导RNA可以以DNA的形式或以RNA的形式引入。Cas蛋白(例如,Cas9)可以以DNA的形式、以RNA的形式或以蛋白质的形式(例如,与向导RNA复合)引入。外源供体核酸可以是DNA、RNA、单链的、双链的、线状的、环状的,等等。类似地,为了稳定性,组分中的每个组分可以包括各种修饰组合,以减少脱靶效应、促进递送,等等。同样地,例如,为了稳定性,RNAi 药剂和ASO可以包括各种修饰组合,以减少脱靶效应、促进递送,等等。

[0169] 作为另一个实例,改变的变量可以是所引入的靶向人凝血因子XII的试剂。例如,如果靶向人凝血因子XII的试剂包括向导RNA,则改变的变量可以是引入具有不同序列的不同向导RNA(例如,靶向不同的向导RNA靶序列)。类似地,如果靶向人凝血因子XII的试剂包括RNAi 药剂或ASO,则改变的变量可以是引入具有不同序列的不同RNAi 药剂或ASO。同样地,如果靶向人凝血因子XII的试剂包括Cas蛋白,则改变的变量可以是引入不同的Cas蛋白(例如,引入具有不同序列的不同Cas蛋白,或具有不同序列(例如,密码子优化的)但对相同Cas蛋白氨基酸序列进行编码的核酸)。同样地,如果靶向人凝血因子XII的试剂包括外源供体核酸,则改变的变量可以是引入具有不同序列的不同外源供体核酸(例如,不同的插入核酸

或不同的同源臂(例如,更长或更短的同源臂或靶向人F12基因的不同区域的同源臂))。

[0170] 在具体实例中,靶向人凝血因子XII的试剂包括Cas蛋白和被设计成靶向人F12基因中的向导RNA靶序列的向导RNA。在此类方法中,改变的变量可以是向导RNA序列和/或向导RNA靶序列。在一些此类方法中,Cas蛋白和向导RNA可以各自以RNA的形式施用,并且改变的变量可以是Cas mRNA与向导RNA的比率(例如,在LNP调配物中)。在一些此类方法中,改变的变量可以是向导RNA修饰(例如,将具有修饰的向导RNA与没有修饰的向导RNA进行比较)。

[0171] C. 靶向人凝血因子XII的试剂

[0172] 靶向人凝血因子XII的试剂可以是靶向人F12基因、人F12 mRNA或人凝血因子XII蛋白的任何试剂。例如,其可以是基因组编辑试剂,如切割人F12基因内的靶序列和/或与人F12基因重组的外源供体序列的核酸酶试剂,其可以是靶向人F12mRNA的反义寡核苷酸,其可以是靶向人凝血因子XII蛋白表位的抗原结合蛋白,或者其可以是靶向人凝血因子XII的小分子。本文所公开的方法中的靶向人凝血因子XII的试剂可以是已知的靶向人凝血因子XII的试剂,可以是假定的靶向人凝血因子XII的试剂(例如,被设计成靶向人凝血因子XII的候选试剂),或者可以是针对人凝血因子XII靶向活性进行筛选的试剂。

[0173] (1) 靶向人F12基因的核酸酶试剂

[0174] 靶向人凝血因子XII的试剂可以是基因组编辑试剂,如切割人F12基因内的靶序列的核酸酶试剂。核酸酶靶序列包含由核酸酶试剂诱导切口或双链断裂的DNA序列。核酸酶试剂的靶序列可以是细胞内源的(或自然的),或者靶序列可以是细胞外源的。细胞外源性靶序列并非天然存在于细胞的基因组中。靶序列还可以对于期望定位在靶基因座处的所关注的多核苷酸是外源的。在一些情况下,靶序列在宿主细胞的基因组中仅存在一次。

[0175] 靶序列的长度可以变化,并且包含例如对于锌指核酸酶(ZFN)对约30-36bp(即,对于每个ZFN约15-18bp)、对于转录激活子样效应物核酸酶(TALEN)约36bp或对于CRISPR/Cas9向导RNA约20bp的靶序列。

[0176] 可以在本文所公开的方法和组合物中使用在期望靶序列处诱导切口或双链断裂的任何核酸酶试剂。可以采用天然存在的或天然的核酸酶试剂,只要所述核酸酶试剂在期望的靶序列中诱导切口或双链断裂即可。可替代地,可以采用经修饰的或工程化的核酸酶试剂。“工程化核酸酶试剂”包含从其自然形式工程化(修饰或衍生)以特异性识别和诱导期望靶序列中的切口或双链断裂的核酸酶。因此,工程化核酸酶试剂可以源自自然的、天然存在的核酸酶试剂,或者可以人工产生或合成。例如,工程化核酸酶可以在靶序列中诱导切口或双链断裂,其中靶序列不是可以被自然(非工程化或非修饰的)核酸酶试剂识别的序列。核酸酶试剂的修饰可以仅仅是蛋白质切割剂中的一个氨基酸或核酸切割剂中的一个核苷酸。在靶序列或其它DNA中产生切口或双链断裂在本文中可以被称为“切割(cutting)”或“切割(cleaving)”靶序列或其它DNA。

[0177] 还提供了示例性靶序列的活性变体和片段。此类活性变体可以与给定靶序列具有至少65%、70%、75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或更多的序列同一性,其中活性变体保留生物活性并因此能够以序列特异性方式被核酸酶试剂识别和切割。由核酸酶试剂测量靶序列双链断裂的测定是众所周知的。参见例如Frendwey等人,(2010)《酶学方法(Methods in Enzymology)》476:295-307,所述文献出于所有目的整体并入本文。

[0178] 核酸酶试剂的靶序列可以定位在F12基因座中或附近的任何位置。靶序列可以定位在F12基因的编码区内,或者在影响基因表达的调节区内。核酸酶试剂的靶序列可以定位在内含子、外显子、启动子、增强子、调节区或任何非蛋白质编码区中。

[0179] 一种类型的核酸酶试剂是转录激活子样效应物核酸酶(TALEN)。TAL效应物核酸酶是一类序列特异性核酸酶,其可以用于在原核或真核生物基因组中的具体靶序列处使双链断裂。TAL效应子核酸酶是通过将天然的或工程化转录激活子样(TAL)效应子或其功能部分融合到核酸内切酶例如FokI的催化结构域而产生的。独特的模块化TAL效应物DNA结合结构域允许具有潜在地任何给定DNA识别特异性的蛋白质的设计。因此,TAL效应物核酸酶的DNA结合结构域可以被工程化为识别具体的DNA靶位点,并且因此用于在期望靶序列处使双链断裂。参见W0 2010/079430;Morbitzer等人(2010)《美国国家科学院院刊(Proc.Natl.Acad.Sci.U.S.A.)》107(50):21617-21622;Scholze和Boch(2010)《毒力(Virulence)》1:428-432;Christian等人《遗传学(Genetics)》(2010)186:757-761;Li等人(2010)《核酸研究》(2010)39(1):359-372;以及Miller等人(2011)《自然生物技术》29:143-148,所述文献中的每个文献通过引用整体并入本文。

[0180] 合适的TAL核酸酶的实例和用于制备合适的TAL核酸酶的方法公开于例如US 2011/0239315 A1、US 2011/0269234 A1、US 2011/0145940 A1、US 2003/0232410 A1、US 2005/0208489 A1、US 2005/0026157 A1、US 2005/0064474 A1、US 2006/0188987 A1和US 2006/0063231 A1中,所述文献中的每个文献通过引用整体并入本文。在各个实施例中,TAL效应核酸酶被工程化,其在例如所关注基因座或所关注基因组基因座中的靶核酸序列中或附近切割,其中靶核酸序列在将被靶向载体修饰的序列处或附近。适合与本文所提供的各种方法和组合物一起使用的TAL核酸酶包含那些专门设计成在将由如本文所述的靶向载体修饰的靶核酸序列处或附近结合的核酸酶。

[0181] 在一些TALEN中,TALEN的每个单体包括通过两个高变残基识别单个碱基对的33-35个TAL重复序列。在一些TALEN中,核酸酶试剂是包括可操作地连接到如FokI核酸内切酶等独立核酸酶的基于TAL重复序列的DNA结合结构域的嵌合蛋白。例如,核酸酶试剂可以包括第一基于TAL重复序列的DNA结合结构域和第二基于TAL重复序列的DNA结合结构域,其中第一和第二基于TAL重复序列的DNA结合结构域中的每一个可操作地连接到FokI核酸酶,其中第一和第二基于TAL重复序列的DNA结合结构域识别由不同长度(12-20bp)的间隔序列隔开的靶DNA序列的每条链中的两个连续靶DNA序列,并且其中FokI核酸酶亚基发生二聚化以产生在靶序列处使双链断裂的活性核酸酶。

[0182] 在本文所公开的各种方法和组合物中采用的核酸酶试剂可以进一步包括锌指核酸酶(ZFN)。在一些ZFN中,ZFN的每个单体包括3个或更多个基于锌指的DNA结合结构域,其中每个基于锌指的DNA结合结构域与3bp亚位点结合。在其它ZFN中,ZFN是包括可操作地连接到如FokI核酸内切酶等独立核酸酶的基于锌指的DNA结合结构域的嵌合蛋白。例如,核酸酶试剂可以包括第一ZFN和第二ZFN,其中第一ZFN和第二ZFN中的每一个可操作地连接到FokI核酸酶亚基,其中第一和第二ZFN识别由约5-7bp间隔区隔开的靶DNA序列的每条链中的两个连续靶DNA序列,并且其中FokI核酸酶亚基发生二聚化以产生使双链断裂的活性核酸酶。参见例如US20060246567;US20080182332;US20020081614;US20030021776;W0/2002/057308A2;US20130123484;US20100291048;W0/2011/017293A2;以及Gaj等人(2013)《生物

技术趋势(Trends in Biotechnology)》31(7):397-405,所述文献中的每个文献通过引用整体并入本文。

[0183] 另一种类型的核酸酶试剂是工程化大范围核酸酶。大范围核酸酶基于保守序列基序分类为四个家族,这些家族是LAGLIDADG、GIY-YIG、H-N-H和His-Cys盒家族。这些基序参与金属离子的配位和磷酸二酯键的水解。大范围核酸酶因其长的靶序列以及在其DNA底物中耐受一些序列多态性而著称。大范围核酸酶结构域、结构和功能是已知的,参见例如Guhan和Muniyappa (2003)《生物化学与分子生物学评论(Crit Rev Biochem Mol Biol)》38:199-248;Lucas等人,(2001)《核酸研究》29:960-9;Jurica和Stoddard,(1999)《细胞分子生命科学(Cell Mol Life Sci)》55:1304-26;Stoddard,(2006)《生物物理学季评(Q Rev Biophys)》38:49-95;以及Moure等人,(2002),《自然结构生物学(Nat Struct Biol)》9:764。在一些实例中,使用天然存在的变体和/或工程化衍生的大范围核酸酶。用于修饰动力学、辅因子相互作用、表达、最佳条件和/或靶序列特异性和筛选活性的方法是已知的。参见例如Epinat等人,(2003)《核酸研究》31:2952-62;Chevalier等人,(2002)《分子细胞(Mol. Cell)》10:895-905;Gimble等人(2003)《分子生物学(Mol Biol)》334:993-1008;Seligman等人,(2002)《核酸研究》30:3870-9;Sussman等人,(2004)《分子生物学杂志》342:31-41;Rosen等人,(2006)《核酸研究》34:4791-800;Chames等人,(2005)《核酸研究》33:e178;Seligman等人,(2006)《核酸研究》34:e149;Gruen等人,(2002)《核酸研究》30:e29;Chen和Zhao,(2005)《核酸研究》33:e154;W02005105989;W02003078619;W02006097854;W02006097853;W02006097784;以及W02004031346,所述文献中的每个文献通过引用整体并入本文。

[0184] 可以使用任何大范围核酸酶,包含例如I-SceI、I-SceII、I-SceIII、I-SceIV、I-SceV、I-SceVI、I-SceVII、I-CeuI、I-CeuAIIP、I-CreI、I-CrepsbIP、I-CrepsbIIP、I-CrepsbIIIP、I-CrepsbIVP、I-TliI、I-PpoI、PI-PspI、F-SceI、F-SceII、F-SuvI、F-TevI、F-TevII、I-AmaI、I-AniI、I-ChuI、I-CmoI、I-CpaI、I-CpaII、I-CsmI、I-CvuI、I-CvuAIP、I-DdiI、I-DdiII、I-DirI、I-DmoI、I-HmuI、I-HmuII、I-HsNIP、I-LlaI、I-MsoI、I-NaaI、I-NanI、I-NciIP、I-NgrIP、I-NitI、I-NjaI、I-Nsp236IP、I-PakI、I-PboIP、I-PcuIP、I-PcuAI、I-PcuVI、I-PgrIP、I-PobIP、I-PorI、I-PorIIP、I-PbpIP、I-SpBetaIP、I-ScaI、I-SexIP、I-SneIP、I-SpomI、I-SpomCP、I-SpomIP、I-SpomIIP、I-SquIP、I-Ssp6803I、I-SthPhiJP、I-SthPhiST3P、I-SthPhiSTe3bP、I-TdeIP、I-TevI、I-TevII、I-TevIII、I-UarAP、I-UarHGPAIP、I-UarHGPA13P、I-VinIP、I-ZbiIP、PI-MtuI、PI-MtuHIP PI-MtuHIIP、PI-PfuI、PI-PfuII、PI-PkoI、PI-PkoII、PI-Rma43812IP、PI-SpBetaIP、PI-SceI、PI-TfuI、PI-TfuII、PI-ThyI、PI-TliI、PI-TliII或其任何活性变体或片段。

[0185] 大范围核酸酶可以识别例如12个到40个碱基对的双链DNA序列。在一些情况下,大范围核酸酶识别基因组中一个完美匹配的靶序列。

[0186] 一些大范围核酸酶是归巢核酸酶。归巢核酸酶的一种类型是归巢核酸酶的LAGLIDADG家族,包含例如I-SceI、I-CreI和I-Dm1。

[0187] 核酸酶试剂可以进一步包括如以下更详细地描述的CRISPR/Cas系统。

[0188] 还提供了核酸酶试剂的活性变体和片段(即,工程化核酸酶试剂)。此类活性变体可以包括与天然核酸酶试剂具有至少65%、70%、75%、80%、85%、90%、91%、92%、93%、

94%、95%、96%、97%、98%、99%或更多的序列同一性,其中活性变体保留在期望的靶序列处切割的能力并因此保留切口或双链断裂诱导活性。例如,本文所描述的核酸酶试剂中的任何核酸酶试剂都可以从天然核酸内切酶序列进行修饰,并被设计成在未被天然核酸酶试剂识别的靶序列处识别和诱导切口或双链断裂。因此,一些工程化核酸酶具有特异性以在不同于对应的天然核酸酶试剂靶序列的靶序列处诱导切口或双链断裂。对切口或双链断裂诱导活性的测定是已知的,并且通常测量内切核酸酶对含有靶序列的DNA底物的总体活性和特异性。

[0189] 核酸酶试剂可以通过任何已知方式引入到细胞或非人动物中。对核酸酶试剂进行编码的多肽可以被直接引入到细胞或非人动物中。可替代地,对核酸酶试剂进行编码的多核苷酸可以被引入到细胞或非人动物中。当引入对核酸酶试剂进行编码的多核苷酸时,核酸酶试剂可以在细胞内瞬时地、有条件地或组成性地表达。对核酸酶试剂进行编码的多核苷酸可以包含在表达盒中并且可操作地连接到条件型启动子、诱导型启动子、组成型启动子或组织特异性启动子。启动子的实例在本文别处进一步详细讨论。可替代地,核酸酶试剂可以作为对核酸酶试剂进行编码的mRNA被引入到细胞中。

[0190] 对核酸酶试剂进行编码的多核苷酸可以稳定地整合在细胞的基因组中,并且可操作地连接到在细胞中具有活性的启动子。可替代地,对核酸酶试剂进行编码的多核苷酸可以在靶向载体中。

[0191] 当通过引入对核酸酶试剂进行编码的多核苷酸向细胞提供核酸酶试剂时,可以修饰这种对核酸酶试剂进行编码的多核苷酸以取代与对核酸酶试剂进行编码的天然存在的多核苷酸序列相比在所关注的细胞中具有更高使用频率的密码子。例如,与天然存在的多核苷酸序列相比,可以修饰对核酸酶试剂进行编码的多核苷酸以取代在给定的所关注真核细胞中具有更高使用频率的密码子,包含人细胞、非人细胞、哺乳动物细胞、啮齿动物细胞、小鼠细胞、大鼠细胞或任何其它所关注的宿主细胞。

[0192] (2) 靶向人F12基因的CRISPR/Cas系统

[0193] 特定类型的靶向人凝血因子XII的试剂可以是靶向人F12基因的成簇规律间隔短回文重复序列(CRISPR)/CRISPR相关(Cas)系统。CRISPR/Cas系统包含转录物和其它参与Cas基因表达或指导其活动的元件。CRISPR/Cas系统可以是例如I型、II型、III型系统或V型系统(例如,V-A亚型或V-B亚型)。在本文所公开的组合物和方法中使用的CRISPR/Cas系统可以是非天然存在的。“非天然存在的”系统包含指示人类参与的任何事物,如系统的一个或多个组分从其天然存在的状态改变或突变,至少基本上不含至少一种与其在自然界中天然相关的其它组分,或者与至少一种与其不天然相关的其它成分相关。例如,一些CRISPR/Cas系统采用包括天然不会同时存在的gRNA和Cas蛋白的非天然存在的CRISPR复合物,采用不会天然存在的Cas蛋白,或采用不会天然存在的gRNA。

[0194] Cas蛋白以及对Cas蛋白进行编码的多核苷酸。Cas蛋白通常包括至少一个可以与向导RNA(gRNA)相互作用的RNA识别或结合结构域。Cas蛋白还可以包括核酸酶结构域(例如,DNase结构域或RNase结构域)、DNA结合结构域、解旋酶结构域、蛋白质-蛋白质相互作用结构域、二聚化结构域和其它结构域。一些此类结构域(例如,DNase结构域)可以来自天然Cas蛋白。可以添加其它此类结构域以制作经修饰的Cas蛋白。核酸酶结构域对核酸切割具有催化活性,所述核酸切割包含核酸分子的共价键的断裂。切割可以产生平端或交错端,并

且切割可以是单链的或双链的。例如,野生型Cas9蛋白通常会产生平切割产物。可替代地,野生型Cpf1蛋白(例如,FnCpf1)可以产生具有5个核苷酸5'突出端的切割产物,其中切割发生在来自非靶向链上的PAM序列的第18个碱基对之后和靶向链上的第23个碱基之后。Cas蛋白可以具有完全切割活性以在靶基因组基因座处产生双链断裂(例如,具有平端的双链断裂),或者其可以是在靶基因组基因座处产生单链断裂的切口酶。

[0195] Cas蛋白的实例包含Cas1、Cas1B、Cas2、Cas3、Cas4、Cas5、Cas5e (CasD)、Cas6、Cas6e、Cas6f、Cas7、Cas8a1、Cas8a2、Cas8b、Cas8c、Cas9 (Csn1或Csx12)、Cas10、Cas10d、CasF、CasG、CasH、Csy1、Csy2、Csy3、Cse1 (CasA)、Cse2 (CasB)、Cse3 (CasE)、Cse4 (CasC)、Csc1、Csc2、Csa5、Csn2、Csm2、Csm3、Csm4、Csm5、Csm6、Cmr1、Cmr3、Cmr4、Cmr5、Cmr6、Csb1、Csb2、Csb3、Csx17、Csx14、Csx10、Csx16、CsaX、Csx3、Csx1、Csx15、Csf1、Csf2、Csf3、Csf4和Cu1966,以及其同系物或修饰版本。

[0196] 示例性Cas蛋白是Cas9蛋白或源自Cas9蛋白的蛋白质。Cas9蛋白来自II型CRISPR/Cas系统,并且通常共享具有保守结构的四个关键基序。基序1、2和4是RuvC样基序,并且基序3是HNH基序。示例性Cas9蛋白来自酿脓链球菌(*Streptococcus pyogenes*)、嗜热链球菌(*Streptococcus thermophilus*)、链球菌属(*Streptococcus* sp.)、金黄色葡萄球菌(*Staphylococcus aureus*)、达松维尔拟诺卡氏菌(*Nocardiopsis dassonvillei*)、始旋链霉菌(*Streptomyces pristinaespiralis*)、绿色产色链霉菌(*Streptomyces viridochromogenes*)、绿色产色链霉菌、玫瑰链孢囊菌(*Streptosporangium roseum*)、玫瑰链孢囊菌、酸热脂环酸芽孢杆菌(*Alicyclobacillus acidocaldarius*)、假蕈状芽孢杆菌(*Bacillus pseudomycoides*)、亚硒酸杆菌(*Bacillus selenitireducens*)、唐松草微小杆菌(*Exiguobacterium sibiricum*)、德氏乳杆菌(*Lactobacillus delbrueckii*)、唾液乳杆菌(*Lactobacillus salivarius*)、海洋微颤菌(*Microscilla marina*)、伯克霍尔德氏菌(*Burkholderiales bacterium*)、食茶极地单胞菌(*Polaromonas naphthalenivorans*)、极地单胞菌属(*Polaromonas* sp.)、瓦氏鳄球藻(*Crocospaera watsonii*)、蓝杆藻属(*Cyanothece* sp.)、铜绿微囊藻(*Microcystis aeruginosa*)、聚球藻属(*Synechococcus* sp.)、阿拉伯糖醋盐杆菌(*Acetohalobium arabaticum*)、丹氏制氮菌(*Ammonifex degensii*)、热解纤维素菌(*Caldicelulosiruptor beccarii*)、金矿菌(*Candidatus Desulforudis*)、肉毒杆菌(*Clostridium botulinum*)、艰难梭菌(*Clostridium difficile*)、大芬戈尔德菌(*Finegoldia magna*)、嗜热盐碱厌氧菌(*Natranaerobius thermophilus*)、嗜热丙酸厌氧肠状菌(*Pelotomaculum thermopropionicum*)、喜温嗜酸硫杆菌(*Acidithiobacillus caldus*)、嗜酸氧化亚铁硫杆菌(*acidithiobacillus ferrooxidans*)、酒色闪杆菌(*Allochromatium vinosum*)、海杆菌属(*Marinobacter* sp.)、嗜盐硝化球菌(*Nitrosococcus halophilus*)、瓦氏亚硝化球菌(*Nitrosococcus watsoni*)、游海假交替单胞菌(*Pseudoalteromonas haloplanktis*)、消旋纤线杆菌(*Ktedonobacter racemifer*)、调查甲烷盐菌(*Methanohalobium evestigatum*)、多变鱼腥藻(*Anabaena variabilis*)、泡沫节球藻(*Nodularia spumigena*)、念珠藻属(*Nostoc* sp.)、极大节旋藻(*Arthrospira maxima*)、钝项节旋藻(*Arthrospira platensis*)、节旋藻属(*Arthrospira* sp.)、鞘丝藻属(*Lyngbya* sp.)、原型微鞘藻(*Microcoleus chthonoplastes*)、颤藻属(*Oscillatoria* sp.)、运动石袍菌(*Petrogloa mobilis*)、非洲栖

热腔菌 (*Thermosipho africanus*)、深海单细胞蓝藻 (*Acaryochloris marina*)、脑膜炎奈瑟氏菌 (*Neisseria meningitidis*) 或空肠弯曲杆菌 (*Campylobacter jejuni*)。Cas9 家族成员的另外的实例在 WO 2014/131833 中描述, 所述文献出于所有目的通过引用整体并入本文。来自酿脓链球菌的 Cas9 (SpCas9) (所分配的 SwissProt 登录号为 Q99ZW2) 是示例性 Cas9 蛋白。来自金黄色葡萄球菌的 Cas9 (SaCas9) (所分配的 UniProt 登录号为 J7RUA5) 是另一种示例性 Cas9 蛋白。来自空肠弯曲杆菌的 Cas9 (CjCas9) (所分配的 UniProt 登录号为 Q0P897) 是另一种示例性 Cas9 蛋白。参见例如 Kim 等人 (2017) 《自然通讯 (Nat. Comm.)》8:14500, 所述文献出于所有目的通过引用整体并入本文。SaCas9 小于 SpCas9, 并且 CjCas9 小于 SaCas9 和 SpCas9 两者。来自脑膜炎奈瑟氏菌的 Cas9 (Nme2Cas9) 是另一种示例性 Cas9 蛋白。参见例如 Edraki 等人 (2019) 《分子细胞》73(4):714-726, 所述文献出于所有目的通过引用整体并入本文。来自嗜热链球菌的 Cas9 蛋白 (例如, 由 CRISPR1 基因座编码的嗜热链球菌 LMD-9Cas9 (St1Cas9) 或来自 CRISPR3 基因座的嗜热链球菌 Cas9 (St3Cas9)) 是其它示例性 Cas9 蛋白。来自新凶手弗朗西斯菌 (*Francisella novicida*) (FnCas9) 的 Cas9 或识别置换性 PAM (E1369R/E1449H/R1556A 取代) 的 RHA 新凶手弗朗西斯菌 Cas9 变体是其它示例性 Cas9 蛋白。这些和其它示例性 Cas9 蛋白例如在 Cebrian-Serrano 和 Davies (2017) 《哺乳动物基因组 (Mamm. Genome)》28(7):247-261 中综述, 所述文献出于所有目的通过引用整体并入本文。示例性 Cas9 蛋白序列可以包括以下、基本上由以下组成或由以下组成: SEQ ID NO: 35。对 Cas9 蛋白进行编码的示例性 DNA 可以包括以下、基本上由以下组成或由以下组成: SEQ ID NO: 36。

[0197] Cas 蛋白的另一个实例是 Cpf1 (来自普雷沃氏菌 (*Prevotella*) 和弗朗西斯氏菌 (*Francisella*) 1 的 CRISPR) 蛋白。Cpf1 是含有与 Cas9 对应结构域同源的 RuvC 样核酸酶结构域以及 Cas9 表征性富精氨酸簇的对应物的大蛋白质 (约 1300 个氨基酸)。然而, Cpf1 缺乏存在于 Cas9 蛋白中的 HNH 核酸酶结构域, 并且 RuvC 样结构域在 Cpf1 序列中是连续的, 与 Cas9 相比, 其含有包含 HNH 结构域的长插入物。参见例如 Zetsche 等人 (2015) 《细胞》163(3):759-771, 所述文献出于所有目的通过引用整体并入本文。示例性 Cpf1 蛋白来自土拉弗朗西斯菌 (*Francisella tularensis*) 1、土拉弗朗西斯菌新凶手亚种 (*Francisella tularensis* subsp. *novicida*)、易北河普雷沃氏菌 (*Prevotella albensis*)、毛螺菌科细菌 (*Lachnospiraceae bacterium*) MC20171、解蛋白丁酸弧菌 (*Butyrivibrio proteoclasticus*)、异域菌门细菌 (*Peregrinibacteria bacterium*) GW2011_GWA2_33_10、俭菌超门细菌 (*Parcubacteria bacterium*) GW2011_GWC2_44_17、史密斯氏菌属 (*Smithella* sp.) SCADC、氨基酸球菌属 (*Acidaminococcus* sp.) BV3L6、毛螺菌科细菌 (*Lachnospiraceae bacterium*) MA2020、候选白蚁甲烷支原体 (*Candidatus Methanoplasma termitum*)、挑剔真杆菌 (*Eubacterium eligens*)、牛眼莫拉氏菌 (*Moraxella bovoculi*) 237、稻田氏钩端螺旋体 (*Leptospira inadai*)、毛螺菌科细菌 (*Lachnospiraceae bacterium*) ND2006、狗口腔卟啉单胞菌 (*Porphyromonas crevioricanis*) 3、解糖脲普雷沃氏菌 (*Prevotella disiens*) 和猕猴卟啉单胞菌 (*Porphyromonas macacae*)。来自新凶手弗朗西斯菌 U112 的 Cpf1 (FnCpf1; 所分配的 UniProt 登录号为 A0Q7Q2) 是示例性 Cpf1 蛋白。

[0198] Cas 蛋白可以是野生型蛋白 (即, 自然界中存在的蛋白)、经修饰的 Cas 蛋白 (即, Cas 蛋白变体) 或者野生型或经修饰的 Cas 蛋白的片段。关于野生型或经修饰的 Cas 蛋白的催化

活性,Cas蛋白也可以是活性变体或片段。关于催化活性,活性变体或片段可以与野生型或经修饰的Cas蛋白或其部分具有至少80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或更多的序列同一性,其中活性变体保留在期望切割位点切割的能力并且因此保留切口诱导活性或双链断裂诱导活性。对切口诱导活性或双链断裂诱导活性的测定是已知的,并且通常测量Cas蛋白在含有切割位点的DNA底物上的整体活性和特异性。

[0199] 可以对Cas蛋白进行修饰以增加或减少核酸结合亲和力、核酸结合特异性和酶活性中的一种或多种。还可以对Cas蛋白进行修饰以改变蛋白质的任何其它活性或特性,如稳定性。例如,可以修饰、缺失或灭活Cas蛋白的一个或多个核酸酶结构域,或者可以截短Cas蛋白以去除对蛋白质功能不是必需的结构域或优化(例如,增强或减少)Cas蛋白的活性或性质。

[0200] 经修饰的Cas蛋白的一个实例是经修饰的SpCas9-HF1蛋白,其是具有设计成减少非特异性DNA接触的改变(N497A/R661A/Q695A/Q926A)的酿脓链球菌Cas9的高保真变体。参见例如Kleinstiver等人,(2016)《自然(Nature)》529(7587):490-495,所述文献出于所有目的整体并入本文。经修饰的Cas蛋白的另一个实例是设计成减少脱靶效应的经修饰的eSpCas9变体(K848A/K1003A/R1060A)。参见例如Slaymaker等人,(2016)《科学(Science)》351(6268):84-88,所述文献出于所有目的整体并入本文。其它SpCas9变体包含K855A和K810A/K1003A/R1060A。这些和其它经修饰的Cas蛋白例如在Cebrian-Serrano和Davies(2017)《哺乳动物基因组》28(7):247-261中综述,所述文献出于所有目的通过引用整体并入本文。经修饰的Cas9蛋白的另一个实例是xCas9,其是可以识别更大范围的PAM序列的SpCas9变体。参见例如Hu等人(2018)《自然》556:57-63,所述文献出于所有目的通过引用整体并入本文。

[0201] Cas蛋白可以包括至少一个核酸酶结构域,如DNase结构域。例如,野生型Cpf1蛋白通常包括可能是二聚体构型的切割靶DNA的两条链的RuvC样结构域。Cas蛋白还可以包括至少两个核酸酶结构域,如DNase结构域。例如,野生型Cas9蛋白通常包括RuvC样核酸酶结构域和HNH样核酸酶结构域。RuvC和HNH结构域可以各自切割不同的双链DNA链,以在DNA中形成双链断裂。参见例如Jinek等人(2012)《科学(Science)》337(6096):816-821,所述文献出于所有目的通过引用整体并入本文。

[0202] 核酸酶结构域中的一个或多个或所有核酸酶结构域可以缺失或突变,使得其不再具有功能或具有降低的核酸酶活性。例如,如果Cas9蛋白中的核酸酶结构域之一缺失或突变,则所得Cas9蛋白可以被称为切口酶,并且可以在双链靶DNA内产生单链断裂但不会产生双链断裂(即,其可以切割互补链或非互补链,但不能切割两者)。如果两个核酸酶结构域都缺失或突变,则所得Cas蛋白(例如,Cas9)切割双链DNA(例如,核酸酶无效或核酸酶失活的Cas蛋白,或催化死亡的Cas蛋白(dCas))的两条链的能力将降低。将Cas9转化为切口酶的突变的实例是来自酿脓链球菌的Cas9的RuvC结构域中的D10A(在Cas9的位置10处天冬氨酸转化为丙氨酸)突变。同样地,来自酿脓链球菌的Cas9的HNH结构域中的H939A(在氨基酸位置839处组氨酸转化为丙氨酸)、H840A(在氨基酸位置840处组氨酸转化为丙氨酸)或N863A(在氨基酸位置863处天冬酰胺转化为丙氨酸)可以将Cas9转化为切口酶。将Cas9转化为切口酶的突变的其它实例包含嗜热链球菌对Cas9的对应突变。参见例如Sapranaukas等人,(2011)《核酸研究》39(21):9275-9282以及WO 2013/141680,所述文献中的每个文献出于所

有目的整体并入本文。可以使用如定点诱变、PCR介导的诱变或总基因合成等方法产生此类突变。产生切口酶的其他突变的实例可以例如在WO 2013/176772和WO 2013/142578中找到,所述文献中的每个文献出于所有目的通过引用整体并入本文。如果Cas蛋白中的全部核酸酶结构域都进行缺失或突变(例如,Cas9蛋白中的两个核酸酶结构域都进行缺失或突变),则所得Cas蛋白(例如,Cas9)切割双链DNA的两条链的能力将降低(例如,核酸酶无效或核酸酶失活Cas蛋白)。一个具体实例是D10A/H840A酿脓链球菌Cas9双突变体或者当与酿脓链球菌Cas9最佳比对时来自另一物种的Cas9中的对应双突变体。另一个具体实例是D10A/N863A酿脓链球菌Cas9双突变体或者当与酿脓链球菌Cas9最佳比对时来自另一物种的Cas9中的对应双突变体。

[0203] xCas9催化结构域中的失活突变的实例与以上针对SpCas9所述的相同。金黄色葡萄球菌Cas9蛋白的催化结构域中的失活突变的实例也是已知的。例如,金黄色葡萄球菌Cas9酶(SaCas9)可以包括位置N580处的取代(例如,N580A取代)和位置D10处的取代(例如,D10A取代),用于产生核酸酶失活的Cas蛋白。参见例如WO 2016/106236,所述文献出于所有目的通过引用整体并入本文。Nme2Cas9的催化结构域中的失活突变的实例也是已知的(例如,D16A和H588A的组合)。St1Cas9催化结构域中的失活突变的实例也是已知的(例如,D9A、D598A、H599A和N622A的组合)。St3Cas9催化结构域中的失活突变的实例也是已知的(例如,D10A和N870A的组合)。CjCas9催化结构域中的失活突变的实例也是已知的(例如,D8A和H559A的组合)。FnCas9和RHAFnCas9催化结构域中的失活突变的实例也是已知的(例如,N995A)。

[0204] Cpf1蛋白的催化结构域中失活突变的实例也是已知的。参考来自新凶手弗朗西丝菌U112(FnCpf1)、氨基酸球菌属BV3L6(AsCpf1)、毛螺菌科细菌ND2006(LbCpf1)和牛眼莫拉氏菌237(MbCpf1 Cpf1)的Cpf1蛋白,此类突变可以包含AsCpf1的位置908、993或1263处或Cpf1直系同源物中的对应位置处或者LbCpf1的位置832、925、947或1180处或Cpf1直系同源物中的对应位置处的突变。此类突变可以包含例如AsCpf1的突变D908A、E993A和D1263A或Cpf1直系同源物中的对应突变或者LbCpf1的突变D832A、E925A、D947A和D1180A或Cpf1直系同源物中的对应突变中的一个或多个突变。参见例如US 2016/0208243,所述文献出于所有目的通过引用整体并入本文。

[0205] Cas蛋白(例如,核酸酶活性Cas蛋白或核酸酶失活Cas蛋白)也可以作为融合蛋白可操作地连接到异源多肽。例如,Cas蛋白可以与切割结构域或表观基因修饰结构域融合。参见WO 2014/089290,所述文献出于所有目的通过引用整体并入本文。Cas蛋白也可以与异源多肽融合,从而提高或降低稳定性。融合结构域或异源多肽可以定位在N端、C端或Cas蛋白内部。

[0206] 作为一个实例,Cas蛋白可以与提供亚细胞定位的一种或多种异源多肽融合。此类异源多肽可以包含例如一种或多种核定位信号(NLS),如用于靶向细胞核的单分SV40 NLS和/或双分 α -输入蛋白NLS、用于靶向线粒体的线粒体定位信号、ER保留信号等等。参见例如Lange等人(2007)《生物化学杂志(J.Biol.Chem.)》282(8):5101-5105,所述文献出于所有目的通过引用整体并入本文。此类亚细胞定位信号可以定位在N端、C端或Cas蛋白内的任何位置。NLS可以包括碱性氨基酸段,并且可以是单组分序列或双组份序列。任选地,Cas蛋白可以包括两个或更多个NLS,包含N端处的NLS(例如, α -输入蛋白NLS或单组分NLS)和C端处

的NLS(例如,SV40 NLS或双组分NLS)。Cas蛋白还可以包括N端处的两个或更多个NLS和/或C端处的两个或更多个NLS。

[0207] Cas蛋白还可以可操作地连接到细胞穿透结构域或蛋白质转导结构域。例如,细胞穿透结构域可以源自HIV-1TAT蛋白、来自人乙型肝炎病毒的TLM细胞穿透基序、MPG、Pep-1、VP22、来自单纯性疱疹病毒的细胞穿透肽或聚精氨酸肽序列。参见例如W0 2014/089290和W0 2013/176772,所述文献中的每个文献出于所有目的通过引用整体并入本文。细胞穿透结构域可以定位在N端、C端或Cas蛋白内的任何位置。

[0208] Cas蛋白还可以可操作地连接到异源多肽,以便于追踪或纯化如荧光蛋白、纯化标签或表位标签。荧光蛋白的实例包含绿色荧光蛋白(例如,GFP、GFP-2、tagGFP、turboGFP、eGFP、祖母绿、Azami绿、单体Azami绿、CopGFP、AceGFP、ZsGreen1)、黄色荧光蛋白(例如,YFP、eYFP、柠檬黄、Venus、YPet、PhiYFP、ZsYellow1)、蓝色荧光蛋白(例如,eBFP、eBFP2、石青、mKalamal、GFPuv、天蓝色、T-天蓝色)、青色荧光蛋白(例如,eCFP、蔚蓝色、CyPet、AmCyan1、Midoriishi-青)、红色荧光蛋白(例如,mKate、mKate2、mPlum、DsRed单体、mCherry、mRFP1、DsRed-表达、DsRed2、DsRed-单体、HcRed-Tandem、HcRed1、AsRed2、eqFP611、mRaspberry、mStrawberry、Jred)、橙色荧光蛋白(例如,mOrange、mKO、Kusabira-橙、单体Kusabira-橙、mTangerine、tdTomato),以及任何其它合适的荧光蛋白。标签的实例包含谷胱甘肽-S-转移酶(GST)、几丁质结合蛋白(CBP)、麦芽糖结合蛋白、硫氧还蛋白(TRX)、聚(NANP)、串联亲和纯化(TAP)标签、myc、AcV5、AU1、AU5、E、ECS、E2、FLAG、血凝素(HA)、nus、Softag 1、Softag 3、Strep、SBP、Glu-Glu、HSV、KT3、S、S1、T7、V5、VSV-G、组氨酸(His)、生物素羧基载体蛋白(BCCP)和钙调蛋白。

[0209] Cas蛋白还可以拴系到外源供体核酸或标记的核酸。这种拴系(即,物理连接)可以通过共价相互作用或非共价相互作用来实现,并且拴系可以是直接的(例如,通过直接融合或化学缀合,这可以通过蛋白质上的半胱氨酸或赖氨酸残基的修饰或内含子修饰来实现),或者可以通过如链霉亲和素或适配子等一个或多个中间接头或衔接子分子来实现。参见例如Pierce等人(2005)《药物化学短评(Mini Rev.Med.Chem.)》5(1):41-55;Duckworth等人(2007)《应用化学-英文国际版(Angew.Chem.Int.Ed.Engl.)》46(46):8819-8822;Schaeffer和Dixon(2009)《澳大利亚化学杂志(Australian J.Chem.)》62(10):1328-1332;Goodman等人(2009)《生物化学(Chembiochem.)》10(9):1551-1557;以及Khatwani等人(2012)《生物有机化学与医药化学(Bioorg.Med.Chem.)》20(14):4532-4539,所述文献中的每个文献出于所有目的通过引用整体并入本文。用于合成蛋白质-核酸缀合物的非共价策略包含生物素-链霉亲和素和镍-组氨酸方法。可以通过使用多种化学反应连接适当功能化的核酸和蛋白质来合成共价蛋白质-核酸缀合物。这些化学反应中的一些化学反应涉及将寡核苷酸直接连接到蛋白质表面上的氨基酸残基(例如,赖氨酸胺或半胱氨酸硫醇),而其它更复杂的方案需要蛋白质的翻译后修饰或者催化或反应蛋白结构域的参与。用于蛋白质与核酸共价连接的方法可以包含例如寡核苷酸与蛋白质赖氨酸或半胱氨酸残基的化学交联、表达的蛋白质连接、化学酶法和光适体的使用。外源供体核酸或标记的核酸可以拴系到Cas蛋白内的C端、N端或内部区域。在一个实例中,外源供体核酸或标记的核酸拴系到Cas蛋白的C端或N端。同样地,Cas蛋白可以拴系到外源供体核酸或标记的核酸的5'端、3'端或内部区域。也就是说,外源供体核酸或标记的核酸可以以任何取向和极性拴系。例如,Cas蛋白

可以栓系到外源供体核酸或标记的核酸的5'端或3'端。

[0210] Cas蛋白可以以任何形式提供。例如,Cas蛋白可以以蛋白质的形式提供,如与gRNA复合的Cas蛋白。可替代地,Cas蛋白可以以编码Cas蛋白的核酸形式提供,如RNA(例如,信使RNA(mRNA))或DNA。任选地,对Cas蛋白进行编码的核酸可以是密码子优化的,以在特定细胞或生物体中高效翻译成蛋白质。例如,可以修饰对Cas蛋白进行编码的核酸以取代与天然存在的多核苷酸序列相比在细菌细胞、酵母细胞、人细胞、非人细胞、哺乳动物细胞、啮齿动物细胞、小鼠细胞、大鼠细胞或任何其它所关注的宿主细胞中具有更高使用频率的密码子。当对Cas蛋白进行编码的核酸被引入到细胞中时,Cas蛋白可以在细胞中瞬时地、有条件地或组成性地表达。

[0211] 可以修饰作为mRNA提供的Cas蛋白以提高稳定性和/或免疫原性性质。可以对mRNA内的一种或多种核苷进行修饰。对mRNA核碱基进行化学修饰的实例包含假尿苷、1-甲基-假尿苷和5-甲基-胞苷。例如,可以使用含有N1-甲基假尿苷的封端的和聚腺苷酸化Cas mRNA。同样地,可以通过使用同义密码子耗尽尿苷来修饰Cas mRNA。

[0212] 对Cas蛋白进行编码的核酸可以稳定地整合在细胞的基因组中,并可操作地连接到在细胞中具有活性的启动子。可替代地,对Cas蛋白进行编码的核酸可以可操作地连接到表达构建体中的启动子。表达构建体包含能够指导基因或其它所关注核酸序列(例如,Cas基因)的表达并且可以将此类所关注核酸序列转移到靶细胞的任何核酸构建体。例如,对Cas蛋白进行编码的核酸可以在包括核酸插入物的靶向载体和/或包括对gRNA进行编码的DNA的载体中。可替代地,其可以在与包括核酸插入物的靶向载体分离和/或与包括对gRNA进行编码的DNA的载体分离的载体或质粒中。可以在表达构建体中使用的启动子包含例如在真核细胞、人细胞、非人细胞、哺乳动物细胞、非人哺乳动物细胞、啮齿动物细胞、小鼠细胞、大鼠细胞、仓鼠细胞、兔细胞、多能细胞、胚胎干(ES)细胞或受精卵中的一个或多个中具有活性的启动子。此类启动子可以是例如条件型启动子、诱导型启动子、组成型启动子或组织特异性启动子。任选地,启动子可以是在一个方向上驱动Cas蛋白的表达并且在另一个方向上驱动向导RNA的表达的双向启动子。此类双向启动子可以由以下组成:(1)完整的、常规的、单向Pol III启动子,其含有3个外部控制元件:远侧序列元件(DSE)、近侧序列元件(PSE)和TATA框;(2)第二基本的Pol III启动子,其包含PSE和TATA盒和以相反取向与DSE的5'端融合的TATA盒。例如,在H1启动子中,DSE邻近PSE和TATA框,并且可以通过产生杂合启动子使启动子双向化,其中通过源自U6启动子的附加PSE和TATA盒来控制反向转录。参见例如US 2016/0074535,所述文献出于所有目的通过引用整体并入本文。使用双向启动子同时表达对Cas蛋白和向导RNA进行编码的基因允许生成紧凑表达盒以促进递送。

[0213] 向导RNA“向导RNA”或“gRNA”是与Cas蛋白(例如,Cas9蛋白)结合并将Cas蛋白靶向靶DNA内的具体位置的RNA分子。向导RNA可以包括两个段:“DNA靶向区段”和“蛋白质结合段”。“段”包含分子的一部分或区域,如RNA中的核苷酸的连续段。一些gRNA,如用于Cas9的gRNA,可以包括两个单独的RNA分子:“激活子-RNA”(例如,tracrRNA)和“靶向子-RNA”(例如,CRISPR RNA或crRNA)。其它gRNA是单个RNA分子(单个RNA多核苷酸),其也可以被称为“单分子gRNA”、“单向导RNA”或“sgRNA”。参见例如WO 2013/176772、WO 2014/065596、WO 2014/089290、WO 2014/093622、WO 2014/099750、WO 2013/142578和WO 2014/131833,所述文献中的每个文献出于所有目的通过引用整体并入本文。例如,对于Cas9,单向导RNA可以

包括(例如,通过接头)融合到tracrRNA的crRNA。例如,对于Cpf1,仅需要一个crRNA来实现与靶序列的结合和/或切割。术语“向导RNA”和“gRNA”包含双分子(即,模块化)gRNA和单分子gRNA两者。

[0214] 示例性双分子gRNA包括crRNA样(“CRISPR RNA”或“靶向子-RNA”或“crRNA”或“crRNA重复序列”)分子和对应的tracrRNA样(“反式作用CRISPR RNA”或“激活子RNA”或“tracrRNA”)分子。crRNA包括gRNA的DNA靶向区段(单链)和形成gRNA的蛋白质结合段的dsRNA双链体的一半的核苷酸段(即,crRNA尾部)两者。定位在DNA靶向区段下游(3')的crRNA尾的实例包括以下、基本上由以下组成或由以下组成:GUUUUAGAGCUAUGCU (SEQ ID NO:37)。本文所公开的DNA靶向区段中的任何区段可以与SEQ ID NO:37的5'端连接以形成crRNA。

[0215] 对应的tracrRNA(激活子-RNA)包括形成gRNA的蛋白质结合区段的dsRNA双链体另一半的核苷酸段。crRNA的核苷酸段与tracrRNA的核苷酸段互补并且杂交,以形成gRNA的蛋白质结合结构域的dsRNA双链体。如此,可以说每个crRNA具有对应的tracrRNA。tracrRNA序列的实例包括以下、基本上由以下组成或由以下组成:AGCAUAGCAAGUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGA GUCGGUGCUUU (SEQ ID NO:38)。tracrRNA序列的其它实例包括以下、基本上由以下组成或由以下组成:AAACAGCAUAGCAAGUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCA CCGAGUCGGUGCUUUU (SEQ ID NO:80)或GUUGGAACCAUUCAAAACAGCAUAGCAAGUAAAAUAAGGCUAGUCCGUUAUCAAC UUGAAAAAGUGGCACCGAGUCGGUGC (SEQ ID NO:81)。

[0216] 在需要crRNA和tracrRNA两者的系统中,crRNA和对应的tracrRNA杂交以形成gRNA。在需要仅crRNA的系统中,crRNA可以是gRNA。crRNA另外提供与靶DNA的互补链杂交的单链DNA靶向区段。如果用于细胞内修饰,则给定crRNA或tracrRNA分子的确切序列可以被设计成对将使用RNA分子的物种具有特异性。参见例如Mali等人(2013)《科学》339(6121):823-826;Jinek等人(2012)《科学》337(6096):816-821;Hwang等人(2013)《自然生物技术》31(3):227-229;Jiang等人(2013)《自然生物技术》31(3):233-239;以及Cong等人(2013)《科学》339(6121):819-823,所述文献中的每个文献出于所有目的通过引用整体并入本文。

[0217] 给定gRNA的DNA靶向区段(crRNA)包括与靶DNA的互补链上的序列互补的核苷酸序列,如以下更详细描述。gRNA的DNA靶向区段通过杂交(即,碱基配对)以序列特异性方式与靶DNA相互作用。如此,DNA靶向区段的核苷酸序列可能会有所不同,并确定将会与gRNA和靶DNA相互作用的靶DNA内的位置。可以修饰主题gRNA的DNA靶向区段以与靶DNA内的任何期望序列杂交。天然存在的crRNA根据CRISPR/Cas系统和生物体的不同而有所不同,但通常含有21到72个核苷酸长度的靶向区段,侧接有长度为21至46个核苷酸的两个直接重复序列(DR)(参见例如WO 2014/131833,所述文献出于所有目的通过引用整体并入本文)。在酿脓链球菌的情况下,DR的长度为36个核苷酸并且靶向区段的长度为30个核苷酸。定位在3'处的DR与对应的tracrRNA互补并杂交,进而与Cas蛋白结合。

[0218] DNA靶向区段的长度可以例如为至少约12个、15个、17个、18个、19个、20个、25个、30个、35个或40个核苷酸。此类DNA靶向区段的长度可以为例如约12个到约100个、约12个到约80个、约12个到约50个、约12个到约40个、约12个到约30个、约12个到约25个或约12个到约20个核苷酸。例如,DNA靶向区段可以为约15个到约25个核苷酸(例如,约17个到约20个核苷酸或约17个、18个、19个或20个核苷酸)。参见例如US 2016/0024523,所述文献出于所有

目的通过引用整体并入本文。对于来自酿脓链球菌的Cas9,典型的DNA靶向区段的长度在16个与20个核苷酸之间或者17个与20个核苷酸之间。对于来自金黄色葡萄球菌的Cas9,典型的DNA靶向区段的长度在21个与23个核苷酸之间。对于Cpf1,典型的DNA靶向区段的长度为至少16个核苷酸或长度为至少18个核苷酸。

[0219] TracrRNA可以呈任何形式(例如,全长tracrRNA或活化部分tracrRNA)并具有不同长度。其可以包含初级转录物或经处理的形式。例如, tracrRNA(作为单向导RNA的一部分,或作为双分子gRNA的一部分的单独分子)可以包括以下、基本上由以下组成或由以下组成:野生型tracrRNA序列的全部或一部分(例如,野生型tracrRNA序列的约或大于约20个、26个、32个、45个、48个、54个、63个、67个、85个或更多个核苷酸)。来自酿脓链球菌的野生型tracrRNA序列的实例包含171-核苷酸、89-核苷酸、75-核苷酸和65-核苷酸版本。参见例如Deltcheva等人(2011)《自然》471(7340):602-607;WO 2014/093661,所述文献中的每个文献出于所有目的通过引用整体并入本文。单向导RNA(sgRNA)内的tracrRNA的实例包含在+48、+54、+67和+85版本的sgRNA中发现的tracrRNA区段,其中“+n”指示野生型tracrRNA的至多+n个核苷酸包含在sgRNA中。参见US 8,697,359,所述文献出于所有目的通过引用整体并入本文。

[0220] 向导RNA的DNA靶向区段与靶DNA的互补链之间的互补性百分比可以为至少60%(例如,至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少97%、至少98%、至少99%或100%)。DNA靶向区段与靶DNA的互补链之间的互补性百分比在约20个连续核苷酸上可以为至少60%。作为实例,DNA靶向区段与靶DNA的互补链之间的互补性百分比在靶DNA的互补链的5'端处的14个连续核苷酸上可以为100%,并且在剩余部分上低至0%。在这种情况下,可以认为DNA靶向区段的长度为14个核苷酸。作为另一个实例,DNA靶向区段与靶DNA的互补链之间的互补性百分比在靶DNA的互补链的5'端处的七个连续核苷酸上可以为100%,并且在剩余部分上低至0%。在这种情况下,可以认为DNA靶向区段的长度为7个核苷酸。在一些向导RNA中,DNA靶向区段内的至少17个核苷酸与靶DNA的互补链互补。例如,DNA靶向区段的长度可以为20个核苷酸,并且可以包括与靶DNA的互补链的1个、2个或3个错配。在一个实例中,错配不与对应于前间区序列邻近基序(PAM)序列的互补链的区域(即,PAM序列的反向互补序列)邻近(例如,错配位于向导RNA的DNA靶向区段的5'端,或者错配与对应于PAM序列的互补链的区域相距至少2个、3个、4个、5个、6个、7个、8个、9个、10个、11个、12个、13个、14个、15个、16个、17个、18个或19个碱基对)。

[0221] gRNA的蛋白质结合区段可以包括两个彼此互补的核苷酸段。蛋白质结合区段的互补核苷酸杂交以形成双链RNA双链体(dsRNA)。主题gRNA的蛋白质结合区段与Cas蛋白相互作用,并且gRNA通过DNA靶向区段将结合的Cas蛋白引导到靶DNA内的具体核苷酸序列。

[0222] 单向导RNA可以包括与支架序列(即,向导RNA的蛋白质结合序列或Cas结合序列)连接的DNA靶向区段。例如,此类向导RNA可以具有连接到3'支架序列的5'DNA靶向区段。示例性支架序列包括以下、基本上由以下组成或由以下组成:GUUUUAGAGCUAGAAUAGCAAGUUA AAAUAAGGCUAGUCCGUUAUCAACUUGAAA AAGUGGCACCGAGUCGGUGCU(第1版;SEQ ID NO:39);GUU GGAACCAUUC AAAACAGCAUAGCAAGUUA AAAU AAGGCUAGUCCGUUAUCAAC UUGAAAAAGUGGCACCGAGUCGGUGC(第2版;SEQ ID NO:40);GUUUUAGAGCUAGAAUAGCAAGUUA AAA UAAGGCUAGUCCGUUAUCAACUUGAAA AAGUGGCACCGAGUCGGUGC(第3版;SEQ ID NO:41);以及GUU

U A A G A G C U A U G C U G G A A A C A G C A U A G C A A G U U U A A A U A A G G C U A G U C C G U U A U C A A C U U G A A A A A G U G G C A C C G A G U C G G U C (第4版; SEQ ID NO: 42)。其它示例性支架序列包括以下、基本上由以下组成或由以下组成: GUUUUAGAGCUAGAAAUAGCAAGUUAUUAAAGGCUAGUCCGUUAU AUCAACUUGAAA AAGUGGCACCGAGUCGGUGCUUUUUU (版本5; SEQ ID NO: 82); GUUUUAGAGCUAGAA AUAGCAAGUUAUUAAAGGCUAGUCCGUUAU AUCAACUUGAAA AAGUGGCACCGAGUCGGUGCUUUU (第6版; SEQ ID NO: 83); 或 GUUUUAGAGCUAUGCUGGAAACAGCAUAGCAAGUUAUUAAAGGCUAGUCCGUUAU CAACUUG AAAAAGUGGCACCGAGUCGGUGCUUUUUU (第7版; SEQ ID NO: 84)。靶向任何向导RNA靶序列的向导RNA可以包含例如与向导RNA的3'端上的示例性向导RNA支架序列中的任何序列融合的向导RNA的5'端上的DNA靶向区段。也就是说, 本文所公开的DNA靶向区段中的任何区段可以与SEQ ID NO: 39-42中的任一项的5'端连接以形成单个向导RNA (嵌合向导RNA)。如本文其它地方所公开的向导RNA版本1、2、3和4是指分别与支架版本1、2、3和4连接的DNA靶向区段 (即, 向导序列或向导)。同样地, 本文所公开的DNA靶向区段中的任何区段可以与SEQ ID NO: 82-84中的任一项的5'端连接以形成单个向导RNA (嵌合向导RNA)。如本文其它地方所公开的向导RNA版本5、6和7是指分别与支架版本5、6和7连接的DNA靶向区段 (即, 向导序列或向导)。

[0223] 向导RNA可以包含提供另外的期望特征的修饰或序列 (例如, 经修饰或经调控的稳定性; 亚细胞靶向; 用荧光标记追踪; 蛋白质或蛋白质复合物的结合位点; 等等)。此类修饰的实例包含例如5'帽 (例如, 7-甲基鸟苷酸帽 (m7G)); 3'聚腺苷酸化尾部 (即, 3' poly (A) 尾部); 核糖开关序列 (例如, 以允许调控稳定性和/或调控蛋白质和/或蛋白质复合物的可及性); 稳定性控制序列; 形成dsRNA双链体 (即, 发夹) 的序列; 将RNA靶向亚细胞位置 (例如, 细胞核、线粒体、叶绿体等等) 的修饰或序列; 提供用于追踪的修饰或序列 (例如, 与荧光分子直接缀合、与促进荧光检测的部分缀合、允许荧光检测的序列等等); 为蛋白质 (例如, 作用于DNA的蛋白质, 包含DNA甲基转移酶、DNA去甲基化酶、组蛋白乙酰转移酶、组蛋白去乙酰化酶等等) 提供结合位点的修饰或序列; 和其组合。修饰的其它实例包含工程化茎环双链体结构、工程化凸起区、茎环双链体结构的工程化发夹3'或其任何组合。参见例如US 2015/0376586, 所述文献出于所有目的整体并入本文。凸起可以由crRNA样区和最小tracrRNA样区组成的双链体内的核苷酸的未配对区。凸起在双链体的一侧可以包括未配对的5'-XXX-3', 其中X是任何嘌呤, 并且Y可以包括可以与相对链上的核苷酸形成摇摆对的核苷酸; 并且在双链体的另一侧包括未配对核苷酸区。

[0224] 未经修饰的核酸可能易于降解。外源核酸也可以诱导先天免疫应答。修饰可以有助于引入稳定性并降低免疫原性。向导RNA可以包括经修饰的核苷和经修饰的核苷酸, 包含例如以下中的一种或多种: (1) 磷酸二酯骨架键中的非连接磷酸氧中的一个或两个和/或连接磷酸氧中的一个或多个的改变或置换; (2) 核糖成分的改变或置换, 如核糖上的2'羟基的改变或置换; (3) 用“脱磷”接头大规模置换磷酸部分; (4) 天然存在的核碱基的修饰或置换; (5) 核糖磷酸骨架的置换或修饰; (6) 寡核苷酸的3'端或5'端的修饰 (例如, 末端磷酸基的去除、修饰或置换或部分的缀合); 以及 (7) 糖的修饰。其它可能的向导RNA修饰包含尿嘧啶或聚尿嘧啶束 (poly-uracil tracts) 的修饰或置换。参见例如WO 2015/048577和US 2016/0237455中, 所述文献中的每个文献出于所有目的整体并入本文。可以对如Cas mRNA等Cas编码核酸进行类似的修饰。例如, 可以通过使用同义密码子耗尽尿苷来修饰Cas mRNA。

[0225] 作为一个实例,向导RNA的5'或3'端处的核苷酸可以包含硫代磷酸酯键(例如,碱基可以具有经修饰的磷酸酯基团,即硫代磷酸酯基团)。例如,向导RNA可以在向导RNA的5'或3'端处的2个、3个或4个末端核苷酸之间包含硫代磷酸酯键。作为另一个实例,向导RNA的5'和/或3'端的核苷酸可以具有2'-O-甲基修饰。例如,向导RNA可以在向导RNA的5'和/或3'端(例如,5'端)的2个、3个或4个末端核苷酸处包含2'-O-甲基修饰。参见例如WO 2017/173054 A1和Finn等人(2018)《细胞报告(Cell Rep.)》22(9):2227-2235,所述文献中的每个文献出于所有目的通过引用整体并入本文。在一个具体实例中,向导RNA包括前三个5'和3'端RNA残基处的2'-O-甲基类似物和3'硫代磷酸酯核苷酸间键。在另一个具体实例中,向导RNA被修饰,使得所有不与Cas9蛋白相互作用的2'OH基团被2'-O-甲基类似物替换,并且与Cas9蛋白的相互作用最小的向导RNA的尾部区域用5'和3'硫代磷酸酯核苷酸间键修饰。参见例如Yin等人(2017)《自然生物技术》35(12):1179-1187,所述文献出于所有目的通过引用整体并入本文。经修饰的向导RNA的其它实例例如在WO 2018/107028 A1中提供,所述文献出于所有目的通过引用整体并入本文。例如,此类化学修饰可以为向导RNA提供更高的稳定性和免受核酸外切酶影响的保护性,使其在细胞内的存留时间比未经修饰的向导RNA更长。例如,此类化学修饰还可以防止先天的细胞内免疫应答,所述先天的细胞内免疫应答可以主动降解RNA或触发导致细胞死亡的免疫级联反应。

[0226] 向导RNA可以以任何形式提供。例如,gRNA可以以RNA的形式提供,作为两个分子(单独的crRNA和tracrRNA)或作为一个分子(sgRNA),并且任选地以与Cas蛋白的复合物的形式提供。gRNA还可以以对gRNA进行编码的DNA的形式提供。对gRNA进行编码的DNA可以对单个RNA分子(sgRNA)或单独的RNA分子(例如,单独的crRNA和tracrRNA)进行编码。在后一种情况下,对gRNA进行编码的DNA可以作为一个DNA分子或作为分别对crRNA和tracrRNA进行编码的单独的DNA分子提供。

[0227] 当gRNA以DNA的形式提供时,gRNA可以在细胞中瞬时地、有条件地或组成性地表达。对gRNA进行编码的DNA可以稳定地整合到细胞的基因组中,并且可操作地连接到在细胞中具有活性的启动子。可替代地,对gRNA进行编码的DNA可以可操作地连接到表达构建体中的启动子。例如,对gRNA进行编码的DNA可以在包括异源核酸的载体中,如对Cas蛋白进行编码的核酸。可替代地,其可以在与包括对Cas蛋白进行编码的核酸的载体分离的载体或质粒中。可以在此类表达构建体中使用的启动子包含例如在真核细胞、人细胞、非人细胞、哺乳动物细胞、非人哺乳动物细胞、啮齿动物细胞、小鼠细胞、大鼠细胞、仓鼠细胞、兔细胞、多能细胞、胚胎干(ES)细胞、成体干细胞、发育受限的祖细胞、诱导多能干(iPS)细胞或单细胞阶段胚胎中的一个或多个中具有活性的启动子。此类启动子可以是例如条件型启动子、诱导型启动子、组成型启动子或组织特异性启动子。此类启动子还可以是例如双向启动子。合适的启动子的具体实例包含RNA聚合酶III启动子,如人U6启动子、大鼠U6聚合酶III启动子或小鼠U6聚合酶III启动子。

[0228] 可替代地,可以通过各种其它方法制备gRNA。例如,可以使用例如T7 RNA聚合酶通过体外转录来制备gRNA(参见例如WO 2014/089290和WO 2014/065596,所述文献中的每个文献出于所有目的通过引用整体并入本文)。向导RNA还可以是通过化学合成制备的合成产生的分子。例如,向导RNA可以化学合成为在前三个5'和3'端RNA残基处包含2'-O-甲基类似物和3'硫代磷酸酯核苷酸间键。

[0229] 向导RNA (或对向导RNA进行编码的核酸) 可以在包括一种或多种向导RNA (例如, 1种、2种、3种、4种或更多种向导RNA) 和增加向导RNA稳定性 (例如, 延长在给定储存条件 (例如, -20°C 、 4°C 或环境温度) 下降解产物保持在阈值以下的时间, 如低于起始核酸或蛋白质重量的0.5%; 或增加体内稳定性) 的载体的组合物中。此类载体的非限制性实例包含聚乳酸 (PLA) 微球体、聚 (D, L-乳酸-乙醇酸共聚物) (PLGA) 微球体、脂质体、胶束、反胶束、脂质螺旋体和脂质微管。此类组合物可以进一步包括Cas蛋白, 如Cas9蛋白或对Cas蛋白进行编码的核酸。

[0230] 向导RNA靶序列向导RNA的靶DNA包含存在于DNA中的核酸序列, 所述核酸序列将与gRNA的DNA靶向区段结合, 前提是存在足够的结合条件。合适的DNA/RNA结合条件包含通常存在于细胞中的生理条件。其它合适的DNA/RNA结合条件 (例如, 无细胞系统中的条件) 在本领域是已知的 (参见例如《分子克隆: 实验室手册 (Molecular Cloning: A Laboratory Manual)》, 第3版 (Sambrook等人, 港湾实验室出版社 (Harbor Laboratory Press 2001)), 所述文献出于所有目的通过引用整体并入本文)。与gRNA互补并杂交的靶DNA链可以被称为“互补链”, 并且与“互补链”互补 (并且因此不与Cas蛋白或gRNA互补) 的靶DNA链可以被称为“非互补链”或“模板链”。

[0231] 靶DNA包含与向导RNA杂交的互补链上的序列和非互补链上的对应序列两者 (例如, 与前间区序列邻近基序 (PAM) 邻近)。如本文所使用的术语“向导RNA靶序列”具体指对应于与向导RNA在互补链上杂交的序列 (即, 反向互补序列) 的非互补链上的序列。也就是说, 向导RNA靶序列是指与PAM邻近 (例如, 在Cas9的情况下在PAM的上游或5') 的非互补链上的序列。向导RNA靶序列等同于向导RNA的DNA靶向区段, 但具有胸腺嘧啶而不是尿嘧啶。作为一个实例, SpCas9酶的向导RNA靶序列可以指非互补链上的5'-NGG-3' PAM上游的序列。向导RNA被设计成与靶DNA的互补链互补, 其中向导RNA的DNA靶向区段与靶DNA的互补链之间的杂交促进了CRISPR复合物的形成。如果有足够的互补性来引起杂交并促进CRISPR复合物的形成, 则不一定需要完全互补。如果向导RNA在本文中被称作靶向向导RNA靶序列, 则意味着向导RNA与靶DNA的互补链序列杂交, 所述互补链序列是非互补链上的向导RNA靶序列的反向互补序列。

[0232] 靶DNA或向导RNA靶序列可以包括任何多核苷酸, 并且可以定位在例如细胞的细胞核或细胞质中或者细胞的细胞器如线粒体或叶绿体内。靶DNA或向导RNA靶序列可以是细胞内源或外源的任何核酸序列。向导RNA靶序列可以是编码基因产物 (例如, 蛋白质) 的序列或非编码序列 (例如, 调节序列) 或者可以包含两者。

[0233] 由Cas蛋白对靶DNA的位点特异性结合和切割可以发生在由以下两者确定的位置处: (i) 向导RNA与靶DNA的互补链之间的碱基配对互补性和(ii) 短基序, 其被称为原间隔子相邻基序 (PAM), 位于靶DNA的非互补链中。PAM可以侧接向导RNA靶序列。任选地, 向导RNA靶序列可以侧接在PAM的3'端上 (例如, 对于Cas9)。可替代地, 向导RNA靶序列可以侧接在PAM的5'端上 (例如, 对于Cpf1)。例如, Cas蛋白的切割位点可以在PAM序列上游或下游 (例如, 在向导RNA靶序列内) 约1到约10或约2到约5个碱基对 (例如, 3个碱基对) 处。在SpCas9的情况下, PAM序列 (即, 非互补链上的PAM序列) 可以是5'-N₁GG-3', 其中N₁是任何DNA核苷酸, 并且其中PAM紧邻靶DNA的非互补链上的向导RNA靶序列的3'。如此, 对应于互补链上的PAM的序列 (即, 反向互补序列) 将是5'-CCN₂-3', 其中N₂是任何DNA核苷酸, 并且是向导RNA的DNA靶向

区段在靶DNA的互补链上杂交的序列的紧接着的5'。在一些这种情况下, N_1 和 N_2 可以互补, 并且 N_1 - N_2 碱基对可以是任何碱基对(例如, $N_1=C$ 和 $N_2=G$; $N_1=G$ 和 $N_2=C$; $N_1=A$ 和 $N_2=T$; 或 $N_1=T$ 和 $N_2=A$)。在来自金黄色葡萄球菌的Cas9的情况下, PAM可以是NNGRRT或NNGRR, 其中N可以是A、G、C或T, 并且R可以是G或A。在来自空肠弯曲菌的Cas9的情况下, PAM可以是例如NNNACAC或NNNRYAC, 其中N可以是A、G、C或T, 并且R可以是G或A。在一些情况下(例如, 对于FnCpf1), PAM序列可以位于5'端上游并且具有序列5'-TTN-3'。

[0234] 向导RNA靶序列的实例是紧接在由SpCas9蛋白识别的NGG基序之前的20个核苷酸的DNA序列。例如, 向导RNA靶序列加PAM的两个实例是 $GN_{19}NGG$ (SEQ ID NO:43)或 $N_{20}NGG$ (SEQ ID NO:44)。参见例如WO 2014/165825, 所述文献出于所有目的通过引用整体并入本文。5'端处的鸟嘌呤可以促进细胞中RNA聚合酶的转录。向导RNA靶序列加PAM的其它实例可以包含5'端处的两个鸟嘌呤核苷酸(例如, $GGN_{20}NGG$; SEQ ID NO:45)以促进体外T7聚合酶的高效转录。参见例如WO 2014/065596, 所述文献出于所有目的通过引用整体并入本文。其它向导RNA靶序列加PAM可以具有长度为4-22个核苷酸的SEQ ID NO:43-45, 包含5'G或GG和3'GG或NGG。又其它向导RNA靶序列PAM可以具有长度为14与20个核苷酸之间的SEQ ID NO:43-45。

[0235] 与靶DNA杂交的CRISPR复合物的形成可以导致靶DNA的一条或两条链在对应于向导RNA靶序列(即, 靶DNA的非互补链上的向导RNA靶序列和与向导RNA杂交的互补链上的反向互补序列)的区域内或附近发生切割。例如, 切割位点可以在向导RNA靶序列内(例如, 在相对于PAM序列的限定位置处)。“切割位点”包含Cas蛋白产生单链断裂或双链断裂的靶DNA的位置。切割位点可以仅在一条链上(例如, 当使用切口酶时)或者在双链DNA的两条链上。切割位点可以在两条链上的相同位置处(产生平端; 例如, Cas9)或可以在每条链上的不同位点处(产生交错末端(即, 突出端); 例如, Cpf1)。例如, 可以通过使用两种Cas蛋白产生交错端, 所述两种蛋白的每一种在不同链上的不同切割位点处产生单链断裂, 由此产生双链断裂。例如, 第一切口酶可以在双链DNA(dsDNA)的第一链上产生单链断裂, 并且第二切口酶可以在dsDNA的第二链上产生单链断裂, 使得产生突出序列。在一些情况下, 第一链上的切口酶的向导RNA靶序列或切割位点与第二链上的切口酶的向导RNA靶序列或切割位点相隔至少2个、3个、4个、5个、6个、7个、8个、9个、10个、15个、20个、25个、30个、40个、50个、75个、100个、250个、500个或1,000个碱基对。

[0236] (3) 靶向人F12基因的外源供体核酸

[0237] 本文所公开的方法和组合物可以利用外源供体核酸在用核酸酶试剂切割人源化F12基因座后或独立于用核酸酶试剂切割人源化F12基因座来修饰人源化F12基因座。在使用核酸酶试剂的此类方法中, 核酸酶试剂蛋白质切割人源化F12基因座以产生单链断裂(切口)或双链断裂, 并且外源供体核酸通过非同源末端接合(NHEJ)介导的连接或通过同源定向修复事件重组人源化F12基因座。任选地, 用外源供体核酸进行修复去除或破坏了核酸酶靶序列, 使得已经靶向的等位基因不能被核酸酶试剂重新靶向。

[0238] 外源供体核酸可以靶向人F12基因中的任何序列。一些外源供体核酸包括同源臂。其它外源供体核酸不包括同源臂。外源供体核酸能够通过同源定向修复插入到人源化F12基因座中, 和/或其能够通过非同源末端接合插入到人源化F12基因座中。

[0239] 外源供体核酸可以包括脱氧核糖核酸(DNA)或核糖核酸(RNA), 其可以是单链或双链的, 并且其可以呈线性或环状形式。例如, 外源供体核酸可以是单链寡脱氧核苷酸

(ssODN)。参见例如Yoshimi等人(2016)《自然通讯》7:10431,所述文献出于所有目的通过引用整体并入本文。外源供体核酸可以是裸核酸或者可以通过如AAV等病毒递送。在具体实例中,外源供体核酸可以通过AAV递送并且能够通过非同源末端接合插入到人源化F12基因座中(例如,外源供体核酸可以是不包括同源臂的核酸)。

[0240] 示例性外源供体核酸的长度在约50个核苷酸到约5kb之间、约50个核苷酸到约3kb之间或约50个到约1,000个核苷酸之间。其它示例性外源供体核酸的长度在约40个到约200个核苷酸之间。例如,外源供体核酸的长度可以在约50-60、60-70、70-80、80-90、90-100、100-110、110-120、120-130、130-140、140-150、150-160、160-170、170-180、180-190或190-200个核苷酸之间。可替代地,外源供体核酸的长度可以在约50-100、100-200、200-300、300-400、400-500、500-600、600-700、700-800、800-900或900-1000个核苷酸之间。可替代地,外源供体核酸的长度可以在约1-1.5、1.5-2、2-2.5、2.5-3、3-3.5、3.5-4、4-4.5或4.5-5kb之间。可替代地,外源供体核酸的长度可以例如不超过5kb、4.5kb、4kb、3.5kb、3kb、2.5kb、2kb、1.5kb、1kb、900个核苷酸、800个核苷酸、700个核苷酸、600个核苷酸、500个核苷酸、400个核苷酸、300个核苷酸、200个核苷酸、100个核苷酸或50个核苷酸。外源供体核酸(例如,靶向载体)也可以更长。

[0241] 在一个实例中,外源供体核酸是长度在约80个核苷酸与约200个核苷酸之间的ssODN。在另一个实例中,外源供体核酸是长度在约80个核苷酸与约3kb之间的ssODN。此类ssODN可以具有例如各自长度在约40个核苷酸与约60个核苷酸之间的同源臂。此类ssODN还可以具有例如各自长度在约30个核苷酸与100个核苷酸之间的同源臂。同源臂可以是对称的(例如,长度为各自40个核苷酸或各自60个核苷酸),或者其可以是不对称的(例如,一个同源臂的长度为36个核苷酸,并且一个同源臂的长度为91个核苷酸)。

[0242] 外源供体核酸可以包含提供另外的期望特征的修饰或序列(例如,经修饰的或经调节的稳定性;用荧光标记跟踪或检测;蛋白质或蛋白质复合物的结合位点;等等)。外源供体核酸可以包括一种或多种荧光标记、纯化标签、表位标签或其组合。例如,外源供体核酸可以包括一个或多个荧光标记(例如,荧光蛋白或其它荧光团或染料),如至少1个、至少2个、至少3个、至少4个或至少5个荧光标记。示例性荧光标记包含荧光团,如荧光素(例如,6-羧基荧光素(6-FAM))、德克萨斯红(Texas Red)、HEX、Cy3、Cy5、Cy5.5、太平洋蓝、5-(和-6)-羧基四甲基罗丹明(TAMRA)和Cy7。多种荧光染料可商购获得,用于标记寡核苷酸(例如,来自整合DNA技术公司(Integrated DNA Technologies))。此类荧光标记(例如,内部荧光标记)可以用于例如检测已经直接整合到经切割的靶核酸中的外源供体核酸,所述经切割的靶核酸具有与外源供体核酸的端相容的突出端。标签或标记可以位于外源供体核酸的5'端、3'端或内部。例如,外源供体核酸可以在5'端与来自整合DNA技术公司(5'IRDYE®700)的IR700荧光团缀合。

[0243] 外源供体核酸还可以包括核酸插入物,所述核酸插入物包含要在人源化F12基因座处整合的DNA区段。核酸插入物在人源化F12基因座处的整合会导致所关注核酸序列到人源化F12基因座的添加,所关注核酸序列在人源化F12基因座处的缺失,或所关注核酸序列在人源化F12基因座处的替换(即,缺失和插入)。一些外源供体核酸被设计成在人源化F12基因座处插入核酸插入物而在人源化F12基因座处没有任何对应的缺失。其它外源供体核酸被设计成在人源化F12基因座处缺失所关注核酸序列而无需核酸插入物的任何对应的插

入。又其它外源供体核酸被设计成在人源化F12基因座处缺失所关注的核酸序列并用核酸插入物将其替换。

[0244] 在人源化F12基因座处被缺失和/或替换的核酸插入物或对应的核酸可以具有各种长度。在人源化F12基因座处被缺失和/或替换的示例性核酸插入物或对应的核酸的长度在约1个核苷酸到约5kb之间或在约1个核苷酸到约1,000个核苷酸之间。例如,在人源化F12基因座处被缺失和/或替换的核酸插入物或对应的核酸的长度在约1-10、10-20、20-30、30-40、40-50、50-60、60-70、70-80、80-90、90-100、100-110、110-120、120-130、130-140、140-150、150-160、160-170、170-180、180-190或190-120个核苷酸之间。同样地,在人源化F12基因座处被缺失和/或替换的核酸插入物或对应的核酸的长度在1-100、100-200、200-300、300-400、400-500、500-600、600-700、700-800、800-900或900-1000个核苷酸之间。同样地,在人源化F12基因座处被缺失和/或替换的核酸插入物或对应的核酸的长度在约1-1.5、1.5-2、2-2.5、2.5-3、3-3.5、3.5-4、4-4.5或4.5-5kb之间或更长。

[0245] 核酸插入物可以包括与旨在用于替换的序列中的所有或部分序列同源或直系同源的序列。例如,核酸插入物可以包括与在人源化F12基因座处旨在用于替换的序列相比包括一个或多个点突变(例如,1个、2个、3个、4个、5个或更多个)的序列。任选地,此类点突变可能导致经编码的多肽中的保守氨基酸取代(例如,用谷氨酸[Glu,E]取代天冬氨酸[Asp,D])。

[0246] 一些外源供体核酸可以对不由野生型内源F12基因座编码或表达的外源蛋白进行编码(例如,可以包括对外源蛋白进行编码的插入核酸)。

[0247] 用于非同源末端接合介导的插入的供体核酸一些外源供体核酸能够通过非同源末端接合插入到人源化F12基因座中。在一些情况下,此类外源供体核酸不包括同源臂。例如,此类外源供体核酸可以在用核酸酶试剂切割后插入到平端双链断裂中。在具体实例中,外源供体核酸可以通过AAV递送并且能够通过非同源末端接合插入到人源化F12基因座中(例如,外源供体核酸可以是不包括同源臂的核酸)。

[0248] 其它外源供体核酸可以在5'端和/或3'端处具有与由核酸酶介导的切割在人源化F12基因座处产生的一个或多个突出端互补的短单链区域。这些突出端也可以被称为5'和3'同源臂。例如,一些外源供体核酸在5'端和/或3'端处具有与由核酸酶介导的切割在人源化F12基因座的5'和/或3'靶序列处产生的一个或多个突出端互补的短单链区域。一些此类外源供体核酸仅在5'端或仅在3'端处具有互补区。例如,一些此类外源供体核酸仅在与人源化F12基因座的5'靶序列处产生的突出端互补的5'端处或者仅在与人源化F12基因座的3'靶序列处产生的突出端互补的3'端处具有互补区。其它此类外源供体核酸在5'和3'端两者处都有互补区域。例如,其它此类外源供体核酸在5'和3'端两者处都有由核酸酶介导的切割在人源化F12基因座处产生的例如分别与第一突出端和第二突出端互补的互补区。例如,如果外源供体核酸是双链的,则单链互补区可以从供体核酸的顶部链的5'端和供体核酸的底部链的5'端延伸,从而在每端产生5'突出端。可替代地,单链互补区可以从供体核酸的顶部链的3'端和模板的底部链的3'端延伸,从而产生3'突出端。

[0249] 互补区可以具有足以促进外源供体核酸与靶核酸之间的连接的任何长度。示例性互补区的长度在约1个到约5个核苷酸之间、在约1个到约25个核苷酸之间或在约5个到约150个核苷酸之间。例如,互补区的长度可以为至少约1个、2个、3个、4个、5个、6个、7个、8个、

9个、10个、11个、12个、13个、14个、15个、16个、17个、18个、19个、20个、21个、22个、23个、24个或25个核苷酸。可替代地,互补区的长度可以为约5-10、10-20、20-30、30-40、40-50、50-60、60-70、70-80、80-90、90-100、100-110、110-120、120-130、130-140或140-150个核苷酸或更长。

[0250] 此类互补区可以与由两对切口酶产生的突出端互补。通过使用切割相反的DNA链以产生第一双链断裂的第一切口酶和第二切口酶以及切割相反的DNA链以产生第二双链断裂的第三切口酶和第四切口酶,可以产生具有交错端的两个双链断裂。例如,Cas蛋白可以用于切割与第一、第二、第三和第四向导RNA对应的第一、第二、第三和第四向导RNA靶序列。第一和第二向导RNA靶序列可以被定位成产生第一切割位点,使得第一和第二切口酶在第一和第二DNA链上产生的切口产生双链断裂(即,第一切割位点包括第一和第二向导RNA靶序列内的切口)。同样地,第三和第四向导RNA靶序列可以被定位成产生第二切割位点,使得第三和第四切口酶在第一和第二DNA链上产生的切口产生双链断裂(即,第二切割位点包括第三和第四向导RNA靶序列内的切口)。优选地,第一和第二向导RNA靶序列和/或第三和第四向导RNA靶序列内的切口可以是产生突出端的偏移切口。偏移窗口可以为例如至少约5bp、10bp、20bp、30bp、40bp、50bp、60bp、70bp、80bp、90bp、100bp或更多。参见Ran等人(2013)《细胞》154:1380-1389;Mali等人(2013)《自然生物技术》31:833-838;以及Shen等人(2014)《自然方法(Nat.Methods)》11:399-404,所述文献中的每个文献出于所有目的通过引用整体并入本文。在这种情况下,双链外源供体核酸可以被设计成具有单链互补区,所述单链互补区与由第一和第二向导RNA靶序列内的切口以及第三和第四向导RNA靶序列内的切口产生的突出端互补。然后此类外源供体核酸可以通过非同源末端接合介导的连接插入。

[0251] 通过同源导向修复插入的供体核酸。一些外源供体核酸包括同源臂。如果外源供体核酸还包括核酸插入物,则同源臂可以侧接核酸插入物。为了便于参考,同源臂在本文中被称作5'和3'(即,上游和下游)同源臂。此术语涉及同源臂与外源供体核酸内的核酸插入物的相对位置。5'和3'同源臂对应于人源化F12基因座内的区域,所述区域在本文中分别被称为“5'靶序列”和“3'靶序列”。

[0252] 当同源臂和靶序列彼此共享足够水平的序列同一性时,则这两个区域彼此“对应(correspond或corresponding)”以充当同源重组反应的底物。术语“同源性”包含与对应序列相同或共享序列同一性的DNA序列。给定靶序列与外源供体核酸中存在的对应同源臂之间的序列同一性可以是允许发生同源重组的任何程度的序列同一性。例如,外源供体核酸(或其片段)的同源臂与靶序列(或其片段)所共享的序列同一性的量可以是至少50%、55%、60%、65%、70%、75%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%序列同一性,使得序列经历同源重组。此外,同源臂与对应的靶序列之间对应的同源区域可以具有足以促进同源重组的任何长度。示例性同源臂的长度在约25个核苷酸到约2.5kb之间、约25个核苷酸到约1.5kb之间或约25个到约500个核苷酸之间。例如,给定同源臂(或同源臂中的每个同源臂)和/或对应的靶序列可以包括长度在约25-30、30-40、40-50、50-60、60-70、70-80、80-90、90-100、100-150、150-200、200-250、250-300、300-350、350-400、400-450或450-500个核苷酸之间的对应同源区域,使得同源臂具有足以与靶核酸内对应的靶序列经历同源重组的同

源性。可替代地,给定同源臂(或每个同源臂)和/或对应的靶序列可以包括长度为以下的对应同源区域:约0.5kb到约1kb、约1kb到约1.5kb、约1.5kb到约2kb或约2kb到约2.5kb。例如,同源臂各自的长度可以为约750个核苷酸。同源臂可以是对称的(每个长度的大小大约相同),或者可以是不对称的(一个比另一个长)。

[0253] 当核酸酶试剂与外源供体核酸组合使用时,5'和3'靶序列优选地定位在足够接近核酸酶切割位点的位置(例如,在足够接近核酸酶靶序列的位置内),以在核酸酶切割位点处的单链断裂(切口)或双链断裂后促进靶序列与同源臂之间同源重组事件的发生。术语“核酸酶切割位点”包含由核酸酶试剂(例如,与向导RNA复合的Cas9蛋白)在其中产生切口或双链断裂的DNA序列。靶基因座内对应于外源供体核酸的5'和3'同源臂的靶序列“定位在足够接近”核酸酶切割位点的位置,如果这样的距离是为了在核酸酶切割位点处的单链断裂或双链断裂后促进5'和3'靶序列与同源臂之间同源重组事件的发生。因此,对应于外源供体核酸的5'和/或3'同源臂的靶序列可以例如在给定核酸酶切割位点的至少1个核苷酸内,或在给定核酸酶切割位点的至少10个核苷酸到约1,000个核苷酸内。作为实例,核酸酶切割位点可以紧邻靶序列的至少一个或两个靶序列。

[0254] 对应于外源供体核酸的同源臂和核酸酶切割位点的靶序列的空间关系可以变化。例如,靶序列可以定位到核酸酶切割位点的5',靶序列可以定位到核酸酶切割位点的3',或者靶序列可以侧接核酸酶切割位点。

[0255] (4) 其它靶向人凝血因子XII的试剂

[0256] 还可以使用本文所公开的非人动物对任何其它已知或假定的靶向人凝血因子XII的试剂的活性进行评估。类似地,可以使用本文所公开的非人动物来筛选任何其它分子的人凝血因子XII靶向活性。

[0257] 作为一个实例,靶向人凝血因子XII的试剂可以是靶向人凝血因子XII蛋白的表位的抗原结合蛋白。术语“抗原结合蛋白”包含与抗原结合的任何蛋白质。抗原结合蛋白的实例包含抗体、抗体的抗原结合片段、多特异性抗体(例如,双特异性抗体)、scFV、双-scFV、双抗体、三抗体、四抗体、V-NAR、VHH、VL、F(ab)、F(ab)₂、DVD(双可变结构域抗原结合蛋白)、SVD(单可变结构域抗原结合蛋白)、双特异性T细胞衔接蛋白(BiTE)或戴维斯体(Davisbody)(美国专利第8,586,713号,所述文献出于所有目的通过引用整体并入本文)。其它靶向人凝血因子XII的试剂包含靶向人凝血因子XII蛋白的小分子。

[0258] 其它靶向人凝血因子XII的试剂可以包含RNAi试剂。“RNAi试剂”是包括能够以序列特异性方式促进如信使RNA(mRNA)等靶RNA的翻译的降级或抑制的小双链RNA或RNA样(例如,化学修饰的RNA)寡核苷酸分子的组合物。RNAi试剂中的寡核苷酸是经连接的核苷的聚合物,所述经连接的核苷中的每个核苷可以是独立地修饰的或未经修饰的。RNAi试剂通过RNA干扰机制起作用(即,通过与哺乳动物细胞的RNA干扰途径机制(RNA诱导的沉默复合物或RISC)相互作用来诱导RNA干扰)。虽然如本文所使用的术语RNAi试剂被认为主要通过RNA干扰机制起作用,但所公开的RNAi试剂不受任何特定途径或作用机制的约束或限制。本文所公开的RNAi试剂包括有义链和反义链,并且包含但不限于短干扰RNA(siRNA)、双链RNA(dsRNA)、微小RNA(miRNA)、短发夹RNA(shRNA)和dicer底物。本文所述的RNAi试剂的反义链至少部分地与靶RNA中的序列(即,连续或有序的核碱基或核苷酸,使用标准命名法用连续字母描述)互补。

[0259] 其它靶向人凝血因子XII的试剂可以包含反义寡核苷酸(ASO)。单链ASO和RNA干扰(RNAi)共享寡核苷酸通过沃森-克里克碱基配对与靶RNA结合的基本原理。不希望受理论束缚,在RNAi期间,小RNA双链体(RNAi药剂)与RNA诱导的沉默复合物(RISC)相关联,一条链(随从链)丢失,并且剩下的链(引导链)与RISC协同结合互补RNA。然后,RISC的催化组分Argonaute 2(Ago2)切割靶RNA。引导链总是与互补的有义链或蛋白质(RISC)相关联。相比之下,ASO必须存活下来并作为单链发挥作用。ASO与靶RNA结合并阻断核糖体或如剪接因子等其它因子与RNA结合或者募集如核酸酶等蛋白质。根据期望的作用机制为ASO选择不同的修饰和靶区域。Gapmer是一种在侧接DNA的中央8-10个碱基间隙的每个端上含有2-5个化学修饰的核苷酸(例如LNA或2'-MOE)的ASO寡核苷酸。在与靶RNA结合之后,DNA-RNA杂交体充当RNase H的底物。

[0260] D. 将靶向人凝血因子XII的试剂施用于非人动物或细胞

[0261] 本文所公开的方法可以包括向非人动物或细胞中引入各种分子(例如,靶向人凝血因子XII的试剂,如治疗分子或复合物),包含核酸、蛋白质、核酸-蛋白质复合物、蛋白质复合物或小分子。“引入”包含以进入细胞内部或非人动物内的细胞内部的方式向细胞或非人动物呈递分子(例如,核酸或蛋白质)。引入可以通过任何方式完成,并且可以以任何组合同时或顺序地将组分中的两种或更多种组分(例如,组分中的两种组分,或组分中的全部组分)引入到细胞或非人动物中。例如,可以在引入向导RNA之前将Cas蛋白引入到细胞或非人动物中,或者可以在引入向导RNA之后将其引入。作为另一个实例,可以在引入Cas蛋白和向导RNA前引入外源供体核酸,或者可以在引入Cas蛋白和向导RNA后引入外源供体核酸(例如,外源供体核酸可以在引入Cas蛋白和向导RNA之前或之后约1小时、2小时、3小时、4小时、8小时、12小时、24小时、36小时、48小时或72小时施用)。参见例如US 2015/0240263和US 2015/0110762,所述文献中的每个文献出于所有目的通过引用整体并入本文。另外,可以通过相同的递送方法或不同的递送方法将组分中的两种或更多种组分引入到细胞或非人动物中。类似地,可以通过相同的施用途径或不同的施用途径将组分中的两种或更多种组分引入到非人动物中。

[0262] 在一些方法中,CRISPR/Cas系统的组分被引入到非人动物或细胞中。可以将向导RNA以RNA的形式(例如,体外转录的RNA)或对向导RNA进行编码的DNA的形式引入到非人动物或细胞中。当以DNA的形式引入时,对向导RNA进行编码的DNA可以可操作地连接到在非人动物的细胞中具有活性的启动子。例如,向导RNA可以通过AAV递送,并在U6启动子下在体内表达。此类DNA可以在一种或多种表达构建体中。例如,此类表达构建体可以是单个核酸分子的组分。可替代地,其可以在两个或多个核酸分子之间以任何组合分离(即,对一种或多种CRISPR RNA进行编码的DNA和对一种或多种tracrRNA进行编码的DNA可以是单独的核酸分子的组分)。

[0263] 同样地,Cas蛋白可以以任何形式提供。例如,Cas蛋白可以以蛋白质的形式提供,如与gRNA复合的Cas蛋白。可替代地,Cas蛋白可以以编码Cas蛋白的核酸形式提供,如RNA(例如,信使RNA(mRNA))或DNA。任选地,对Cas蛋白进行编码的核酸可以是密码子优化的,以在特定细胞或生物体中高效翻译成蛋白质。例如,可以修饰对Cas蛋白进行编码的核酸以取代与天然存在的多核苷酸序列相比在哺乳动物细胞、啮齿动物细胞、小鼠细胞、大鼠细胞或任何其它所关注宿主细胞中具有更高使用频率的密码子。当对Cas蛋白进行编码的核酸被

引入到非人动物中时,Cas蛋白可以在非人动物的细胞中瞬时地、有条件地或组成性地表达。

[0264] 对Cas蛋白或向导RNA进行编码的核酸可以可操作地连接到表达构建体中的启动子。表达构建体包含能够指导基因或其它所关注核酸序列(例如,Cas基因)的表达并且可以将此类所关注核酸序列转移到靶细胞的任何核酸构建体。例如,对Cas蛋白进行编码的核酸可以在包括对一种或多种gRNA进行编码的DNA的载体中。可替代地,其可以在与包括对一种或多种gRNA进行编码的DNA的载体分离的载体或质粒中。可以在表达构建体中使用的合适的启动子包含例如在真核细胞、人细胞、非人细胞、哺乳动物细胞、非人哺乳动物细胞、啮齿动物细胞、小鼠细胞、大鼠细胞、仓鼠细胞、兔细胞、多能细胞、胚胎干(ES)细胞、成体干细胞、发育受限的祖细胞、诱导多能干(iPS)细胞或单细胞阶段胚胎中的一个或多个中具有活性的启动子。此类启动子可以是例如条件型启动子、诱导型启动子、组成型启动子或组织特异性启动子。任选地,启动子可以是在一个方向上驱动Cas蛋白的表达并且在另一个方向上驱动向导RNA的表达的双向启动子。此类双向启动子可以由以下组成:(1)完整的、常规的、单向Pol III启动子,其含有3个外部控制元件:远侧序列元件(DSE)、近侧序列元件(PSE)和TATA框;(2)第二基本的Pol III启动子,其包含PSE和TATA盒和以相反取向与DSE的5'端融合的TATA盒。例如,在H1启动子中,DSE邻近PSE和TATA框,并且可以通过产生杂合启动子使启动子双向化,其中通过源自U6启动子的附加PSE和TATA盒来控制反向转录。参见例如US 2016/0074535,所述文献出于所有目的通过引用整体并入本文。使用双向启动子同时表达对Cas蛋白和向导RNA进行编码的基因允许生成紧凑表达盒以促进递送。

[0265] 引入到非人动物或细胞中的分子(例如,Cas蛋白或向导RNA或RNAi药剂或ASO)可以在包括增加所引入分子的稳定性(例如,延长在给定储存条件(例如,-20℃、4℃或环境温度)下降解产物保持在阈值以下的时间,如低于起始核酸或蛋白质重量的0.5%;或增加体内稳定性)的载体的组合物中提供。此类载体的非限制性实例包含聚乳酸(PLA)微球体、聚(D,L-乳酸-乙醇酸共聚物)(PLGA)微球体、脂质体、胶束、反胶束、脂质螺旋体和脂质微管。

[0266] 本文提供了允许将分子(例如,核酸或蛋白质)引入到细胞或非人动物中的多种方法和组合物。用于将分子引入到各种细胞类型中的方法是已知的,并且包含例如稳定转染方法、瞬时转染方法和病毒介导的方法。

[0267] 转染方案以及用于将分子引入到细胞中的方案可能会有所不同。非限制性转染方法包含使用以下的基于化学的转染方法:脂质体;纳米颗粒;磷酸钙(Graham等人(1973)《病毒学(Virology)》52(2):456-67;Bacchetti等人(1977)《美国国家科学院院刊》74(4):1590-4;以及Kriegler,M(1991)《转移和表达:实验室手册(Transfer and Expression:A Laboratory Manual.)》纽约W.H.弗里曼出版社(W.H.Freeman and Company)第96-97页);树枝状聚合物;或阳离子聚合物,如DEAE葡聚糖或聚乙烯亚胺。非化学方法包含电穿孔、声穿孔和光转染。基于颗粒的转染包含使用基因枪或磁体辅助的转染(Bertram(2006)《当今药物生物技术(Current Pharmaceutical Biotechnology)》7,277-28)。病毒方法也可以用于转染。

[0268] 将核酸或蛋白质引入到细胞中也可以通过电穿孔、胞质内注射、病毒感染、腺病毒、腺相关病毒、慢病毒、逆转录病毒、转染、脂质介导的转染或核转染来介导。核转染是使得核酸底物不仅能够递送到细胞质而且能够通过核膜进入到细胞核中的改进的电穿孔技

术。另外,在本文所公开的方法中使用核转染通常需要比常规电穿孔少得多的细胞(例如,与常规电穿孔需要700万细胞相比,仅需要约200万细胞)。在一个实例中,使用LONZA® NUCLEOFECTOR™系统执行核转染。

[0269] 将分子(例如,核酸或蛋白质)引入到细胞(例如,受精卵)中还可以通过显微注射来完成。在受精卵(即,单细胞阶段胚胎)中,显微注射可以注入到母本和/或父本原核或细胞质中。如果显微注射仅注入到一个原核,则优选父本原核,因为其尺寸较大。优选将mRNA显微注射到细胞质中(例如,将mRNA直接递送到翻译装置),而优选将Cas蛋白或对Cas蛋白或RNA进行编码的多核苷酸显微注射到细胞核/原核中。可替代地,可以通过注射到细胞核/原核和细胞质两者中来进行显微注射:可以首先将针引入到细胞核/原核中并注射第一量,并且当从单细胞阶段胚胎中取出针时,可以将第二量注射到细胞质中。如果Cas蛋白被注射到细胞质中,则Cas蛋白优选地包括用于确保递送到细胞核/原核的核定位信号。用于进行显微注射的方法是众所周知的。参见例如Nagy等人(Nagy A、Gertsenstein M、Vintersten K、Behringer R.,2003,《操纵小鼠胚胎(Manipulating the Mouse Embryo.)》纽约冷泉港:冷泉港实验室出版社);还参见Meyer等人(2010)《美国国家科学院院刊》107:15022-15026和Meyer等人(2012)《美国国家科学院院刊》109:9354-9359。

[0270] 用于将分子(核酸或蛋白质)引入到细胞或非人动物中的其它方法可以包含例如载体递送、颗粒介导的递送、外泌体介导的递送、脂质纳米颗粒介导的递送、细胞穿透肽介导的递送或可植入装置介导的递送。作为具体实例,可以将核酸或蛋白质以如聚乳酸(PLA)微球体、聚(D,L-乳酸-乙醇酸共聚物)(PLGA)微球体、脂质体、胶束、反胶束、脂质螺旋体或脂质微管等载体引入到细胞或非人动物中。向非人动物递送的一些具体实例包含流体动力学递送、病毒介导的递送(例如,腺相关病毒(AAV)介导的递送)和脂质纳米颗粒介导的递送。

[0271] 将核酸和蛋白质引入到细胞或非人动物中可以通过流体动力递送(HDD)来完成。流体动力递送已成为用于细胞内DNA体内递送的方法。对于到实质细胞的基因递送,只需要通过选定的血管注射必需的DNA序列,从而消除与当前病毒和合成载体相关的安全问题。当注入到血流中时,DNA能够到达血液可及的不同组织中的细胞。流体动力递送采用将大量溶液快速注射到循环中不可压缩的血液中所产生的力来解决阻止大的和不可透过膜的化合物进入实质细胞的内皮和细胞膜的物理屏障问题。除了递送DNA外,此方法还可用于RNA、蛋白质和其它小化合物在体内的高效细胞内递送。参见例如Bonamassa等人(2011)《药学研究(Pharm.Res.)》28(4):694-701,所述文献出于所有目的通过引用整体并入本文。

[0272] 核酸的引入也可以通过病毒介导的递送来完成,如AAV介导的递送或慢病毒介导的递送。其它示例性病毒/病毒性载体包含逆转录病毒、腺病毒、牛痘病毒、痘病毒和单纯疱疹病毒。病毒可以感染分裂细胞、非分裂细胞或分裂细胞和非分裂细胞两者。病毒可以整合到宿主基因组中,或者可替代地不整合到宿主基因组中。此类病毒还可以被工程化为具有降低的免疫力。病毒可能具有复制能力,也可能具有复制缺陷(例如,在另外轮次的病毒粒子复制和/或包装所必需的一个或多个基因中存在缺陷)。病毒可以引起瞬时表达、长期表达(例如,至少1周、2周、1个月、2个月或3个月)或永久表达(例如,Cas9和/或gRNA)。示例性病毒滴度(例如,AAV滴度)包含 10^{12} 、 10^{13} 、 10^{14} 、 10^{15} 和 10^{16} 个载体基因组/mL。

[0273] ssDNA AAV基因组由两个开放阅读框Rep和Cap组成,其侧接有允许合成互补DNA链

的两个反向末端重复序列。当构建AAV转移质粒时,转基因放置在两个ITR之间,并且Rep和Cap可以反式提供。除了Rep和Cap之外,AAV还可能含有腺病毒基因的辅助质粒。这些基因(E4、E2a和VA)介导AAV复制。例如,转移质粒、Rep/Cap和辅助质粒可以转染到含有腺病毒基因E1+的HEK293细胞中,以产生感染性AAV颗粒。可替代地,将Rep、Cap和腺病毒辅助基因可以组合成单个质粒。类似的包装细胞和方法可以用于其它病毒,如逆转录病毒。

[0274] 已识别出多种AAV血清型。这些血清型在其感染的细胞类型(即,其趋向性)方面不同,允许优先转导特定细胞类型。CNS组织的血清型包含AAV1、AAV2、AAV4、AAV5、AAV8和AAV9。心脏组织的血清型包含AAV1、AAV8和AAV9。肾组织的血清型包含AAV2。肺组织的血清型包含AAV4、AAV5、AAV6和AAV9。胰腺组织的血清型包含AAV8。感光细胞的血清型包含AAV2、AAV5和AAV8。视网膜色素上皮组织的血清型包含AAV1、AAV2、AAV4、AAV5和AAV8。骨骼肌组织的血清型包含AAV1、AAV6、AAV7、AAV8和AAV9。肝组织的血清型包含AAV7、AAV8和AAV9,并且特别是AAV8。

[0275] 趋向性可以通过假型进一步细化,所述假型即混合来自不同病毒血清型的衣壳和基因组。例如,AAV2/5指示包装在来自血清型5的衣壳中的含有血清型2基因组的病毒。使用假型病毒可以提高转导效率以及改变趋向性。源自不同血清型的杂交衣壳也可以用于改变病毒趋向性。例如,AAV-DJ含有来自八种血清型的杂交衣壳,并在广泛的体内细胞类型中显示出高感染性。AAV-DJ8是显示AAV-DJ性质的另一个实例,但具有增强的脑摄取。AAV血清型还可以通过突变进行修饰。AAV2突变修饰的实例包含Y444F、Y500F、Y730F和S662V。AAV3突变修饰的实例包含Y705F、Y731F和T492V。AAV6突变修饰的实例包含S663V和T492V。其它假型的/经修饰的AAV变体包含AAV2/1、AAV2/6、AAV2/7、AAV2/8、AAV2/9、AAV2.5、AAV8.2和AAV/SASTG。

[0276] 为了加速转基因表达,可以使用自身互补型AAV(scAAV)变体。由于AAV依赖于细胞的DNA复制机制来合成AAV单链DNA基因组的互补链,因此转基因表达可能会延迟。为了解决这种延迟问题,可以使用含有能够在感染后自发退火的互补序列的scAAV,从而消除对宿主细胞DNA合成的需要。然而,也可以使用单链AAV(ssAAV)载体。

[0277] 为了提高包装能力,可以将较长的转基因在两个AAV转移质粒之间拆分,第一个具有3'剪接供体并且第二个具有5'剪接受体。在细胞共感染后,这些病毒形成多联体,拼接在一起,并且全长转基因可以被表达。虽然这允许更长的转基因表达,但表达效率较低。用于增加容量的类似方法利用同源重组。例如,转基因可以在两个转移质粒之间分开但是具大量的序列重叠,使得共表达诱导全长转基因的同源重组和表达。

[0278] 核酸和蛋白质的引入也可以通过脂质纳米颗粒(LNP)介导的递送来完成。例如,LNP介导的递送可以用于递送Cas mRNA和向导RNA的组合或Cas蛋白和向导RNA的组合。通过此类方法递送导致瞬时Cas表达,并且生物可降解脂质提高清除率、提高耐受性并降低免疫原性。脂质调配物可以保护生物分子免于降解,同时改善其细胞摄取。脂质纳米颗粒是通过分子间力彼此物理相关的多个脂质分子的颗粒。这些颗粒包含微球体(包含单层和多层囊泡,例如,脂质体)、乳液中的分散相、胶束或悬浮液中的内相。此类脂质纳米颗粒可以用于封装一个或多个核酸或蛋白质以供递送。含有阳离子脂质的调配物可用于递送如核酸等聚阴离子。其它可以包含在内的脂质是中性脂质(即,不带电荷或两性离子脂质)、阴离子脂质、增强转染的辅助脂质和增加纳米颗粒可以在体内存在的时间长度的隐形脂质。合适

的阳离子脂质、中性脂质、阴离子脂质、辅助脂质和隐形脂质的实例可以在WO 2016/010840 A1中找到,所述文献出于所有目的通过引用整体并入本文。示例性脂质纳米颗粒可以包括阳离子脂质和一种或多种其它组分。在一个实例中,其它组分可以包括如胆固醇等辅助脂质。在另一个实例中,其它组分可以包括如胆固醇等辅助脂质和如DSPC等中性脂质。在另一个实例中,其它组分可以包括如胆固醇等辅助脂质、如DSPC等任选的中性脂质以及如S010、S024、S027、S031或S033等隐形脂质。

[0279] LNP可以含有以下中的一种或多种或全部:(i)用于封装和用于内体逃逸的脂质;(ii)用于稳定的中性脂质;(iii)用于稳定的辅助脂质;(iv)隐形脂质。参见例如Finn等人(2018)《细胞报告》22:1-9和WO 2017/173054 A1,所述文献中的每个文献出于所有目的通过引用整体并入本文。在某些LNP中,负荷物可以包含向导RNA或对向导RNA进行编码的核酸。在某些LNP中,负荷物可以包含对如Cas9等Cas核酸酶进行编码的mRNA以及向导RNA或对向导RNA进行编码的核酸。

[0280] 用于包封和内体逃逸的脂质可以是阳离子脂质。脂质还可以是生物可降解脂质,如生物可降解可电离脂质。合适的脂质的一个实例是脂质A或LP01,即(9Z,12Z)-3-((4,4-双(辛氧基)丁酰基)氧基)-2-((((3-(二乙氨基)丙氧基)羰基)氧基)甲基)丙基十八-9,12-二烯酸酯,也被称为3-((4,4-双(辛氧基)丁酰基)氧基)-2-((((3-(二乙氨基)丙氧基)羰基)氧基)甲基)丙基(9Z,12Z)-十八-9,12-二烯酸酯。参见例如Finn等人(2018)《细胞报告》22:1-9和WO 2017/173054 A1,所述文献中的每个文献出于所有目的通过引用整体并入本文。合适的脂质的另一个实例是脂质B,即((5-((二甲氨基)甲基)-1,3-亚苯基)双(氧基))双(辛烷-8,1-二基)双(癸酸酯),也被称为((5-((二甲氨基)甲基)-1,3-亚苯基)双(氧基))双(辛烷-8,1-二基)双(癸酸酯)。合适的脂质的另一个实例是脂质C,即2-((4-(((3-(二甲氨基)丙氧基)羰基)氧基)十六酰基)氧基)丙烷-1,3-二基(9Z,9'Z,12Z,12'Z)-双(十八-9,12-二烯酸酯)。合适的脂质的另一个实例是脂质D,即3-(((3-(二甲氨基)丙氧基)羰基)氧基)-13-(辛酰氧基)十三烷基3-辛基十一烷酸酯。其它合适的脂质包含三十七-6,9,28,31-四烯-19-基4-(二甲氨基)丁酸酯(也被称为Dlin-MC3-DMA(MC3))。

[0281] 适用于本文所述的LNP的一些此类脂质在体内是生物可降解的。例如,包括此类脂质的LNP包含在8小时、10小时、12小时、24小时或48小时或3天、4天、5天、6天、7天或10天内从血浆中清除脂质的至少75%的那些。作为另一个实例,LNP的至少50%在8小时、10小时、12小时、24小时或48小时或3天、4天、5天、6天、7天或10天内从血浆中清除。

[0282] 根据其所在的介质的pH值,此类脂质可以是可电离的。例如,在微酸性介质中,脂质可以被质子化并且因此带有正电荷。相反,在弱碱性介质中,例如在pH大约为7.35的血液中,脂质可能不会被质子化并且因此不带电荷。在一些实施例,脂质可以在至少约9、9.5或10的pH下质子化。这种脂质带电荷的能力与其内在pKa有关。例如,脂质可以独立地具有范围为约5.8到约6.2的pKa。

[0283] 中性脂质的作用是稳定和改善LNP的处理。合适的中性脂质的实例包含各种中性、不带电荷或两性离子脂质。适用于本公开的中性磷脂的实例包含但不限于5-十七烷基苯-1,3-二醇(间苯二酚)、二棕榈酰磷脂酰胆碱(DPPC)、二硬脂酰磷脂酰胆碱(DSPC)、磷酸胆碱(DOPC)、二肉豆蔻酰磷脂酰胆碱(DMPC)、磷脂酰胆碱(PLPC)、1,2-二硬脂酰-sn-甘油-3-磷酸胆碱(DAPC)、磷脂酰乙醇胺(PE)、卵磷脂酰胆碱(EPC)、二月桂酰磷脂酰胆碱(DLPC)、二肉

豆蔻酰磷脂酰胆碱 (DMPC)、1-肉豆蔻酰-2-棕榈酰磷脂酰胆碱 (MPPC)、1-棕榈酰-2-肉豆蔻酰磷脂酰胆碱 (PMPC)、1-棕榈酰-2-硬脂酰磷脂酰胆碱 (PSPC)、1,2-二花生酰-sn-甘油-3-磷酸胆碱 (DBPC)、1-硬脂酰-2-棕榈酰磷脂酰胆碱 (SPPC)、1,2-二二十碳烯酰-sn-甘油-3-磷酸胆碱 (DEPC)、棕榈酰油酰磷脂酰胆碱 (POPC)、溶血磷脂酰胆碱、二油酰磷脂酰乙醇胺 (DOPE)、二亚油酰磷脂酰胆碱二硬脂酰磷脂酰乙醇胺 (DSPE)、二肉豆蔻酰磷脂酰乙醇胺 (DMPE)、二棕榈酰磷脂酰乙醇胺 (DPPE)、棕榈酰油酰磷脂酰乙醇胺 (POPE)、溶血磷脂酰乙醇胺和其组合。例如,中性磷脂可以选自由二硬脂酰磷脂酰胆碱 (DSPC) 和二肉豆蔻酰磷脂酰乙醇胺 (DMPE) 组成的组。

[0284] 辅助脂质包含增强转染的脂质。辅助脂质增强转染的机制可以包含增强颗粒稳定性。在某些情况下,辅助脂质可以增强膜融合性。辅助脂质包含类固醇、甾醇和烷基间苯二酚。合适的辅助脂质的实例包含胆固醇、5-十七烷基间苯二酚和胆固醇半琥珀酸酯。在一个实例中,辅助脂质可以是胆固醇或胆固醇半琥珀酸酯。

[0285] 隐形脂质包含改变纳米颗粒可以在体内存在的时间长度的脂质。隐形脂质可以通过例如减少颗粒聚集和控制粒度来帮助调配过程。隐形脂质可以调节LNP的药代动力学性质。合适的隐形脂质包含具有连接到脂质部分的亲水性头部基团的脂质。

[0286] 隐形脂质的亲水性头部基团可以包括例如选自基于PEG (有时称为聚(环氧乙烷))、聚(噁唑啉)、聚(乙烯醇)、聚(甘油)、聚(N-乙基吡咯烷酮)、聚氨基酸和聚N-(2-羟丙基)甲基丙烯酰胺的聚合物的聚合物部分。术语PEG意指任何聚乙二醇或其它聚亚烷基醚聚合物。在某些LNP调配物中,PEG是PEG-2K,也被称为PEG 2000,其平均分子量为约2,000道尔顿。参见例如WO 2017/173054 A1,所述文献出于所有目的通过引用整体并入本文。

[0287] 隐形脂质的脂质部分可以衍生自例如二酰基甘油或二烷基甘酰胺,其包含包括二烷基甘油或二烷基甘酰胺基团的那些,所述二烷基甘油或二烷基甘酰胺基团具有独立地包括约C4到约C40个饱和或不饱和碳原子的烷基链长度,其中链可以包括一个或多个官能团,例如酰胺或酯。二酰基甘油或二烷基甘酰胺基团可以进一步包括一个或多个经取代的烷基。

[0288] 作为一个实例,隐形脂质可以选自PEG-二月桂酸甘油酯、PEG-二肉豆蔻酰甘油 (PEG-DMG)、PEG-二棕榈酰甘油、PEG-二硬脂酰甘油 (PEG-DSPE)、PEG-二月桂甘酰胺、PEG-二肉豆蔻甘酰胺、PEG-二棕榈酰甘酰胺和PEG-二硬脂酰甘酰胺、PEG-胆固醇(1-[8'-(胆甾-5-en-3[β]-氧基)甲酰胺基-3',6'-二氧杂辛基]氨甲酰基-[ω]-甲基-聚(乙二醇)、PEG-DMB(3,4-二十四烷基苄基-[ω]-甲基-聚(乙二醇)醚)、1,2-二肉豆蔻酰-sn-甘油-3-磷酸乙醇胺-N-[甲氧基(聚乙二醇)-2000] (PEG2k-DMG)、1,2-二硬脂酰-sn-甘油-3-磷酸乙醇胺-N-[甲氧基(聚乙二醇)-2000] (PEG2k-DSPE)、1,2-二硬脂酰-sn-甘油、甲氧基聚乙二醇(PEG2k-DSG)、聚(乙二醇)-2000-二甲基丙烯酸酯(PEG2k-DMA)和1,2-二硬脂氧基丙基-3-胺-N-[甲氧基(聚乙二醇)-2000] (PEG2k-DSA)。在一个特定实例中,隐形脂质可以是PEG2k-DMG。

[0289] LNP可以包括调配物中相应摩尔比的组分脂质。CCD脂质的mol-%可以为例如约30mol-%到约60mol-%、约35mol-%到约55mol-%、约40mol-%到约50mol-%、约42mol-%到约47mol-%或约45%。辅助脂质的mol-%可以为例如约30mol-%到约60mol-%、约35mol-%到约55mol-%、约40mol-%到约50mol-%、约41mol-%到约46mol-%或约

44mol-%。中性脂质的mol-%可以为例如约1mol-%到约20mol-%、约5mol-%到约15mol-%、约7mol-%到约12mol-%或约9mol-%。隐形脂质的mol-%可以为例如约1mol-%到约10mol-%、约1mol-%到约5mol-%、约1mol-%到约3mol-%、约2mol-%或约1mol-%。

[0290] LNP在生物可降解脂质(N)的带正电荷的胺基团与待封装的核酸的带负电荷的磷酸基团(P)之间可以具有不同的比率。这可以由等式N/P在数学上表示。例如,N/P比率可以为约0.5到约100、约1到约50、约1到约25、约1到约10、约1到约7、约3到约5、约4到约5、约4、约4.5或约5。N/P比率也可以是约4到约7或约4.5到约6。在具体实例中,N/P比率可以为4.5或者可以为6。

[0291] 在一些LNP中,负荷物可以包括Cas mRNA和gRNA。Cas mRNA和gRNA的比率可以不同。例如,LNP调配物可以包含范围为约25:1到约1:25、约10:1到约1:10、约5:1到约1:5或约1:1的Cas mRNA与gRNA核酸的比率。可替代地,LNP调配物可以包含约1:1到约1:5或约10:1的Cas mRNA与gRNA核酸的比率。可替代地,LNP调配物可以包含约1:10、25:1、10:1、5:1、3:1、1:1、1:3、1:5、1:10或1:25的Cas mRNA与gRNA核酸的比率。可替代地,LNP调配物可以包含约1:1到约1:2的Cas mRNA与gRNA核酸的比率。在具体实例中,Cas mRNA与gRNA的比率可以为约1:1或约1:2。

[0292] 在一些LNP中,负荷物可以包括外源供体核酸和gRNA。外源供体核酸和gRNA的比率可以不同。例如,LNP调配物可以包含范围为约25:1到约1:25、约10:1到约1:10、约5:1到约1:5或约1:1的外源供体核酸与gRNA核酸的比率。可替代地,LNP调配物可以包含约1:1到约1:5、约5:1到约1:1、约10:1或约1:10的外源供体核酸与gRNA核酸的比率。可替代地,LNP调配物可以包含约1:10、25:1、10:1、5:1、3:1、1:1、1:3、1:5、1:10或1:25的外源供体核酸与gRNA核酸的比率。

[0293] 合适的LNP的具体实例的氮磷(N/P)比为4.5,并且含有摩尔比为45:44:9:2的生物可降解阳离子脂质、胆固醇、DSPC和PEG2k-DMG。生物可降解阳离子脂质可以是(9Z,12Z)-3-((4,4-双(辛氧基)丁酰基)氧基)-2-(((3-(二乙氨基)丙氧基)羰基)氧基)甲基)丙基十八-9,12-二烯酸酯,也被称为3-((4,4-双(辛氧基)丁酰基)氧基)-2-(((3-(二乙氨基)丙氧基)羰基)氧基)甲基)丙基(9Z,12Z)-十八-9,12-二烯酸酯。参见例如Finn等人(2018)《细胞报告》22:1-9,所述文献出于所有目的通过引用整体并入本文。Cas9 mRNA与向导RNA的重量比可以为1:1。合适的LNP的另一个具体实例含有摩尔比为50:38.5:10:1.5的Dlin-MC3-DMA(MC3)、胆固醇、DSPC和PEG-DMG。

[0294] 合适的LNP的另一个具体实例的氮磷(N/P)比为6,并且含有摩尔比为50:38:9:3的生物可降解阳离子脂质、胆固醇、DSPC和PEG2k-DMG。生物可降解阳离子脂质可以是(9Z,12Z)-3-((4,4-双(辛氧基)丁酰基)氧基)-2-(((3-(二乙氨基)丙氧基)羰基)氧基)甲基)丙基十八-9,12-二烯酸酯,也被称为3-((4,4-双(辛氧基)丁酰基)氧基)-2-(((3-(二乙氨基)丙氧基)羰基)氧基)甲基)丙基(9Z,12Z)-十八-9,12-二烯酸酯。Cas9 mRNA与向导RNA的重量比可以为1:2。

[0295] 可以选择降低免疫原性的递送方式。例如,Cas蛋白和gRNA可以通过不同的模式(例如,双模式递送)进行递送。这些不同的模式可以赋予受试者递送的分子(例如,Cas或核酸编码、gRNA或核酸编码或外源供体核酸/修复模板)不同的药效学或药代动力学性质。例如,不同的模式会导致不同的组织分布、不同的半衰期或不同的时间分布。一些递送模式

(例如,通过自主复制或基因组整合来递送在细胞中持续存在的核酸载体)导致分子的表达和存在更持久,而其它递送模式是瞬时的且不太持久(例如,RNA或蛋白质的递送)。以更瞬时的方式递送Cas蛋白例如作为mRNA或蛋白质可以确保Cas/gRNA复合物仅在短时间内存在和具有活性,并且可以降低由MHC分子在细胞表面上展示的来自细菌来源的Cas酶的肽引起的免疫原性。此类瞬时递送还可以减少脱靶修饰的可能性。

[0296] 体内施用可以通过任何合适的途径,包含例如肠胃外、静脉内、口服、皮下、动脉内、颅内、鞘内、腹膜内、局部、鼻内或肌肉内施用。全身施用方式包含例如口服和肠胃外途径。肠胃外途径的实例包含静脉内、动脉内、骨内、肌肉内、皮内、皮下、鼻内和腹膜内途径。具体的实例是静脉输液。鼻滴注和玻璃体内注射是其它具体实例。局部施用方式包含例如鞘内、脑室内、脑实质内(例如,局部脑实质内递送到纹状体(例如,进入尾状核或进入壳核种)、大脑皮层、中央前回、海马体(例如,进入齿状回或CA3区域)、颞叶皮层、杏仁核、额叶皮层、丘脑、小脑、髓质、下丘脑、顶盖、被盖或黑质)、眼内、眶内、结膜下、玻璃体内、视网膜下和经巩膜途径。与全身施用(例如,静脉内)相比,当局部施用(例如,脑实质内或玻璃体内)时,显著更少量的组分(与全身方法相比)可以发挥作用。局部施用方式还可以降低或消除当全身施用治疗有效量的组分时可能发生的潜在毒副作用的发生率。

[0297] 体内施用可以通过任何合适的途径,包含例如肠胃外、静脉内、口服、皮下、动脉内、颅内、鞘内、腹膜内、局部、鼻内或肌肉内施用。具体的实例是静脉输液。可以使用一种或多种生理学和药学上可接受的载体、稀释剂、赋形剂或助剂来调配包括向导RNA和/或Cas蛋白(或对向导RNA和/或Cas蛋白进行编码的核酸)的组合物。调配物可以取决于所选择的施用途径。术语“药学上可接受的”意指载体、稀释剂、赋形剂或助剂与调配物的其它成分相容并且对其接受者基本上无害。

[0298] 施用频率和剂量数可以取决于外源供体核酸、向导RNA或Cas蛋白(或对向导RNA或Cas蛋白进行编码的核酸)的半衰期和施用途径等因素。将核酸或蛋白质引入到细胞或非人动物中可以在一段时间内执行一次或多次。例如,引入可以按以下频率执行:一段时间内至少两次、一段时间内至少三次、一段时间内至少四次、一段时间内至少五次、一段时间内至少六次、一段时间内至少七次、一段时间内至少八次、一段时间内至少九次、一段时间内至少十次、至少十一次、一段时间内至少十二次、一段时间内至少十三次、一段时间内至少十四次、一段时间内至少十五次、一段时间内至少十六次、一段时间内至少十七次、一段时间内至少十八次、一段时间内至少十九次或一段时间内至少二十次。

[0299] E. 在体内或离体测量靶向人凝血因子XII的试剂的递送、活性或功效

[0300] 本文所公开的方法可以进一步包括检测或测量靶向人凝血因子XII的试剂的活性。测量此类试剂的活性可以包括例如测量凝血因子XII的表达或活性或预测的活性。例如,评估可以包括测量凝血或凝血酶生成。可替代地,评估可以包括测量肝细胞生长因子(HGF)-酪氨酸-蛋白激酶Met (Met) 信号传导活性(例如,在神经元中)的潜力,如评估pro-HGF向活性HGF的转化。

[0301] 作为一个实例,如果靶向人凝血因子XII的试剂是基因组编辑试剂(例如,被设计成靶向人F12基因座的CRISPR/Cas),则测量可以包括评估人源化F12基因座的修饰。

[0302] 可以使用各种方法来识别具有靶向基因修饰的细胞。筛选可以包括用于评估亲本染色体的等位基因修饰(MOA)的定量测定。参见例如US 2004/0018626;US 2014/0178879;

US 2016/0145646; WO 2016/081923; 以及Frendewey等人(2010)《酶学方法(Methods Enzymol.)》476:295-307,所述文献中的每个文献出于所有目的通过引用整体并入本文。例如,定量测定可以通过定量PCR进行,如实时PCR(qPCR)。实时PCR可以利用识别靶基因座的第一引物组和识别非靶向参考基因座的第二引物组。引物组可以包括识别扩增序列的荧光探针。合适的定量测定的其它实例包含荧光介导的原位杂交(FISH)、比较基因组杂交、等温DNA扩增、与固定化探针的定量杂交、INVADER[®]探针、TAQMAN[®]分子信标探针或ECLIPSE[™]探针技术(参见例如US 2005/0144655,所述文献出于所有目的通过引用整体并入本文)。

[0303] 下一代测序(NGS)也可以用于筛选。下一代测序也可以被称为“NGS”或“大规模平行测序”或“高通量测序”。除了MOA测定外,NGS还可以用作筛选工具,用于定义靶向基因修饰的确切性质以及其是否在细胞类型或组织类型或器官类型上保持一致。

[0304] 评估对非人动物中的人源化F12基因座的修饰可以在来自任何组织或器官的任何细胞类型中进行。例如,评估可以在来自同一组织或器官的多种细胞类型中进行,或者在来自组织或器官内的多个位置的细胞中进行。这可以提供关于靶向靶组织或器官内的哪些细胞类型或者靶向人凝血因子XII的试剂到达组织或器官的哪些部分的信息。作为另一个实例,评估可以在多种类型的组织或多个器官中进行。在靶向特定组织、器官或细胞类型的方法中,这可以提供关于靶向所述组织或器官的有效性以及其它组织或器官中是否存在脱靶效应的信息。

[0305] 如果试剂被设计成使人源化F12基因座灭活、影响人源化F12基因座的表达、防止人源化F12 mRNA的翻译或清除人源化凝血因子XII蛋白质,则测量可以包括评估人源化F12mRNA或蛋白质表达。这种测量可以在特定细胞类型或器官内进行(例如,肝脏或肝脏内的特定细胞类型或区域,或者脑或大脑内的特定细胞类型或区域,如海马锥体层(锥体神经元)和皮层神经元),或者其可以涉及测量人源化凝血因子XII蛋白的血清或血浆水平。

[0306] 可以使用的测定的一个实例是RNASCOPE[™]和BASESCOPE[™]RNA原位杂交(ISH)测定,这是一种可以在完整的固定组织的情况下定量细胞特异性经编辑的转录物,包含单核苷酸变化的方法。BASESCOPE[™]RNAISH测定可以在基因编辑的表征中补充NGS和qPCR。虽然NGS/qPCR可以提供野生型和编辑序列的定量平均值,但其不提供有关组织内经编辑的细胞的异质性或百分比的信息。BASESCOPE[™]ISH测定可以提供整个组织的景观图并以单细胞分辨率对野生型与经编辑的转录物进行定量,其中可以定量靶组织中含有经编辑的mRNA转录物的细胞的实际数量。BASESCOPE[™]测定使用配对寡核苷酸(“ZZ”)探针在无非特异性背景的情况下放大信号,从而实现单分子RNA检测。然而,BASESCOPE[™]探针设计和信号放大系统利用ZZ探针实现单分子RNA检测,并且可以差异地检测完整固定组织中的单核苷酸编辑和突变。

[0307] 人源化凝血因子XII蛋白的产生和分泌可以通过任何已知的方式进行评估。例如,可以通过使用已知测定法测量非人动物中经编码的mRNA的水平或经编码的蛋白质的水平(例如,测量血浆或血清水平)来评估表达。

[0308] 人源化凝血因子XII蛋白的活性可以通过任何已知的方式进行评估。此类测定可以包含例如评估前激肽释放酶对激肽释放酶的活化作用、评估FXI的活化作用、评估pro-HGF对HGF的活化作用或评估经典补体途径的活化作用。

[0309] IV. 制备包括人源化F12基因座的非人动物的方法

[0310] 提供了用于制备如本文别处所述的包括人源化凝血因子XII (F12) 基因座的非人动物基因组、非人动物细胞和非人动物的各种方法。同样地,提供了用于制备人源化凝血因子XII (F12) 基因或基因座或者制备包括如本文别处公开的人源化凝血因子XII (F12) 基因座的非人动物基因组或非人动物细胞的各种方法。用于产生基因修饰的生物体的任何方便的方法或方案都适用于产生这种基因修饰的非人动物。参见例如Poueymirou等人(2007)《自然生物技术》25(1):91-99;US 7,294,754;US 7,576,259;US 7,659,442;US 8,816,150;US 9,414,575;US 9,730,434;以及US 10,039,269,所述文献中的每个文献出于所有目的通过引用整体并入本文(描述小鼠ES细胞和用于制备基因修饰的小鼠的VELOCIMOUSE®方法)。还参见US 2014/0235933 A1、US 2014/0310828 A1、US 10,385,359和US 10,329,582,所述文献中的每个文献出于所有目的通过引用整体并入本文(描述大鼠ES细胞和用于制备基因修饰的大鼠的方法)。还参见Cho等人(2009)《细胞生物学实验指南(Curr.Protoc.Cell.Biol.)》42:19.11.1-19.11.22(doi:10.1002/0471143030.cb1911s42)以及Gama Sosa等人(2010)《脑结构与功能(Brain Struct.Funct.)》214(2-3):91-109,所述文献中的每个文献出于所有目的通过引用整体并入本文。例如,可以通过在靶标F12基因座上进行基因敲入来生成此类基因修饰的非人动物。

[0311] 例如,产生包括人源化F12基因座的非人动物的方法可以包括:(1)将多能细胞的基因组修饰成包括人源化F12基因座;(2)识别或选择包括人源化F12基因座的基因修饰的多能细胞;(3)将基因修饰的多能细胞引入到非人动物宿主胚胎中;以及(4)将宿主胚胎植入代孕母体中并在其中孕育。例如,产生包括人源化F12基因座的非人动物的方法可以包括:(1)提供包括人源化F12基因座的多能细胞(例如,胚胎干(ES)细胞,如小鼠ES细胞或大鼠ES细胞);(2)将基因修饰的多能细胞引入到非人动物宿主胚胎中;以及(3)在代孕母体中孕育宿主胚胎。

[0312] 作为另一个实例,产生包括人源化F12基因座的非人动物的方法可以包括:(1)将多能细胞的基因组(例如,胚胎干(ES)细胞,如小鼠ES细胞或大鼠ES细胞)修饰成包括人源化F12基因座;(2)识别或选择包括人源化F12基因座的基因修饰的多能细胞;(3)将基因修饰的多能细胞引入到非人动物宿主胚胎中;以及(4)在代孕母体中孕育宿主胚胎。供体细胞可以在任何阶段引入到宿主胚胎中,如胚泡阶段或桑椹胚前阶段(即,4细胞阶段或8细胞阶段)。任选地,包括经修饰的多能细胞(例如,非人ES细胞)的宿主胚胎可以在植入到代孕母体中并在其中孕育以产生F0非人动物之前进行温育,直到胚泡阶段为止。然后,代孕母体可以产生包括人源化F12基因座的F0代非人动物(并且能够通过种系传递基因修饰)。

[0313] 所述方法可以进一步包括识别具有经修饰的靶基因组基因座的细胞或动物。可以使用各种方法来识别具有靶向基因修饰的细胞和动物。

[0314] 可替代地,产生本文别处描述的非人动物的方法可以包括:(1)使用上述用于修饰多能细胞的方法将单细胞阶段胚胎的基因组修饰成包括人源化F12基因座;(2)选择基因修饰的胚胎;(3)在代孕母体中孕育基因修饰的胚胎。产生能够通过种系传递基因修饰的后代。

[0315] 核转移技术还可以用于产生非人哺乳动物。简而言之,用于核转移的方法可以包括以下步骤:(1)去除卵母细胞的核或提供去核卵母细胞;(2)分离或提供要与去核卵母细

胞结合的供体细胞或细胞核；(3) 将细胞或细胞核插入到去核卵母细胞中以形成重组细胞；(4) 将重组细胞植入到动物的子宫中以形成胚胎；以及(5) 允许胚胎发育。在此类方法中，卵母细胞通常从死亡动物身上取得，但卵母细胞也可以从活体动物的输卵管和/或卵巢中分离出来。在去核前，卵母细胞可以在各种众所周知的培养基中成熟。可以以多种众所周知的方式执行卵母细胞的去核。可以通过在融合前在透明带下显微注射供体细胞来将供体细胞或细胞核插入到去核卵母细胞中以形成重组细胞。可以通过在接触/融合平面上施加DC电脉冲(电融合)、通过将细胞暴露于如聚乙二醇等促进融合的化学物质或通过如仙台病毒等灭活病毒的方式来诱导融合。在核供体和受体卵母细胞融合之前、期间和/或之后，可以通过电气方式和/或非电气方式来活化重组细胞。活化方法包含电脉冲、化学诱发的休克、精子渗透、增加卵母细胞中二价阳离子的水平以及减少卵母细胞中细胞蛋白质的磷酸化(如通过激酶抑制剂的方式)。活化的重组细胞或胚胎可以在众所周知的培养基中培养，并且然后转移到动物的子宫中。参见例如US 2008/0092249、WO 1999/005266、US 2004/0177390、WO 2008/017234和US 7,612,250，所述文献中的每个文献出于所有目的通过引用整体并入本文。

[0316] 经修饰的细胞或单细胞阶段胚胎可以例如通过以下方式重组产生：(a) 向细胞中引入包括插入核酸的一种或多种外源供体核酸(例如，靶向载体)，所述插入核酸侧接有例如对应于5'和3'靶位点(例如，侧接内源序列旨在用于缺失并替换为插入核酸的靶位点)的5'和3'同源臂，其中插入核酸包括用于产生人源化F12基因座的人F12序列；以及(b) 识别至少一种在其基因组中包括在内源F12基因座处整合的插入核酸的细胞(即，识别至少一种包括人源化F12基因座的细胞)。同样地，经修饰的非人动物基因组或人源化非人动物F12基因可以例如通过以下方式重组产生：(a) 使基因组或基因与一种或多种外源供体核酸(例如，靶向载体)接触，所述一种或多种外源供体核酸包括对应于5'和3'靶位点(例如，侧接内源序列旨在用于缺失并替换为侧接有5'和3'同源臂的插入核酸(例如，包括用于产生人源化F12基因座的人F12序列)的靶位点)的5'和3'同源臂，其中外源供体核酸被设计用于内源非人动物F12基因座的人源化。

[0317] 可替代地，经修饰的多能细胞或单细胞阶段胚胎可以通过以下方式产生：(a) 向细胞中引入：(i) 核酸酶试剂，其中核酸酶试剂在内源F12基因座内的靶位点处诱导切口或双链断裂；以及(ii) 包括插入核酸的一种或多种外源供体核酸(例如，靶向载体)，所述插入核酸侧接有例如对应于5'和3'靶位点(例如，侧接内源序列旨在用于缺失并替换为插入核酸的靶位点)的5'和3'同源臂，其中插入核酸包括用于产生人源化F12基因座的人F12序列；以及(c) 识别至少一种在其基因组中包括在内源F12基因座处整合的插入核酸的细胞(即，识别至少一种包括人源化F12基因座的细胞)。同样地，经修饰的非人动物基因组或人源化非人动物F12基因可以通过使基因组或基因与以下接触来产生：(i) 核酸酶试剂，其中核酸酶试剂在内源F12基因座或基因内的靶位点处诱导切口或双链断裂；以及(ii) 包括插入核酸(例如，包括用于产生人源化F12基因座的人源化F12序列)的一种或多种外源供体核酸(例如，靶向载体)，所述插入核酸侧接有例如对应于5'和3'靶位点(例如，侧接内源序列旨在用于缺失并替换为插入核酸的靶位点)的5'和3'同源臂，其中外源供体核酸被设计用于内源F12基因座的人源化。可以使用任何诱导切口或双链断裂进入期望识别位点的核酸酶试剂。合适的核酸酶的实例包含转录激活因子样效应物核酸酶(TALEN)、锌指核酸酶(ZFN)、大范围

核酸酶和成簇规律间隔短回文重复序列 (CRISPR) /CRISPR相关 (Cas) 系统 (例如, CRISPR/Cas9系统) 或此类系统 (例如, CRISPR/Cas9) 的组分。参见例如US 2013/0309670和US 2015/0159175, 所述文献中的每个文献出于所有目的通过引用整体并入本文。在一个实例中, 核酸酶包括Cas9蛋白和向导RNA。在另一个实例中, 核酸酶包括Cas9蛋白和两个或更多个、三个或更多个或四个或更多个向导RNA。

[0318] 修饰基因组的步骤可以例如利用外源修复模板或供体核酸 (例如, 靶向载体) 将F12基因座修饰成包括本文所公开的人源化F12基因座。作为一个实例, 靶向载体可以用于在内源F12基因座 (例如, 内源非人动物F12基因座) 处产生人源化F12基因, 其中靶向载体包括核酸插入物, 所述核酸插入物包括要在侧接有靶向内源F12基因座处的5'靶序列的5'同源臂和靶向内源F12基因座处的3'靶序列的3'同源臂的人F12基因座中整合的人F12序列。在F12基因座中整合核酸插入物会导致所关注核酸序列在F12基因座中的添加、所关注核酸序列在F12基因座中的缺失或所关注核酸序列在F12基因座中的替换 (即, 缺失和插入)。同源臂可以侧接包括用于产生人源化F12基因座的人F12序列的插入核酸 (例如, 用于缺失内源F12基因座的区段并替换为直系同源人F12序列)。

[0319] 外源修复模板或供体核酸可以用于非同源末端接合介导的插入或同源重组。外源修复模板或供体核酸可以包括脱氧核糖核酸 (DNA) 或核糖核酸 (RNA), 其可以是单链或双链的, 并且其可以呈线性或环状形式。例如, 修复模板可以是单链寡脱氧核苷酸 (ssODN)。

[0320] 外源供体核酸还可以包括非靶向内源F12基因座处不存在的异源序列。例如, 外源供体核酸可以包括选择盒, 如侧接有重组酶识别位点的选择盒。

[0321] 一些外源供体核酸包括同源臂。如果外源供体核酸还包括核酸插入物, 则同源臂可以侧接核酸插入物。为了便于参考, 同源臂在本文中被称作5'和3' (即, 上游和下游) 同源臂。此术语涉及同源臂与外源供体核酸内的核酸插入物的相对位置。5'和3'同源臂对应于F12基因座内的区域, 所述区域在本文中分别被称为“5'靶序列”和“3'靶序列”。

[0322] 当同源臂和靶序列彼此共享足够水平的序列同一性时, 则这两个区域彼此“对应 (correspond或corresponding)”以充当同源重组反应的底物。术语“同源性”包含与对应序列相同或共享序列同一性的DNA序列。给定靶序列与外源供体核酸中存在的对应同源臂之间的序列同一性可以是允许发生同源重组的任何程度的序列同一性。例如, 外源供体核酸 (或其片段) 的同源臂与靶序列 (或其片段) 所共享的序列同一性的量可以是至少50%、55%、60%、65%、70%、75%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%序列同一性, 使得序列经历同源重组。此外, 同源臂与对应的靶序列之间对应的同源区域可以具有足以促进同源重组的任何长度。在一些靶向载体中, 内源F12基因座中的预期突变包含在侧接有同源臂的插入核酸中。

[0323] 在除单细胞阶段胚胎之外的细胞中, 外源供体核酸可以是包含靶向载体的“大靶向载体”或“LTVEC”, 所述靶向载体包括同源臂, 所述同源臂对应于并来源于比用于在细胞中执行同源重组的其它方法通常使用的核酸序列更大的核酸序列。参见例如US 2004/0018626; WO 2013/163394; US 9,834,786; US 10,301,646; WO 2015/088643; US 9,228,208; US 9,546,384; US 10,208,317; 以及US 2019-0112619, 所述文献中的每个文献出于所有目的通过引用整体并入本文。LTVEC还包含包括核酸插入物的靶向载体, 所述核酸插入物

的核酸序列大于旨在在细胞中进行同源重组的其它方法通常所使用的核酸序列。例如,LTVEC可以对由于其大小限制而无法由传统的基于质粒的靶向载体容纳的大基因座进行修饰。例如,在不存在由核酸酶试剂(例如,Cas蛋白)诱导的切口或双链断裂的情况下,靶向基因座可以是(即,5'和3'同源臂可以对应于)使用常规方法不可靶向或只能错误靶向或仅以显著低效率靶向的细胞的基因座。LTVEC可以是任何长度,并且通常长度为至少10kb。LTVEC中5'同源臂和3'同源臂的总和通常至少为10kb。使用VELOCIGENE®基因工程技术通过细菌同源重组(BHR)反应产生和使用来源于细菌人工染色体(BAC)DNA的大靶向载体(LTVEC)描述于例如US 6,586,251和Valenzuela等人(2003)《自然生物技术》21(6):652-659,所述文献中的每个文献出于所有目的通过引用整体并入本文。通过体外组装方法产生LTVEC描述于例如US 2015/0376628和WO 2015/200334,所述文献中的每个文献出于所有目的通过引用整体并入本文。

[0324] 所述方法可以进一步包括识别具有经修饰的靶基因组基因座的细胞或动物。可以使用各种方法来识别具有靶向基因修饰的细胞和动物。筛选步骤可以包括例如用于评估亲本染色体的等位基因修饰(MOA)的定量测定。参见例如US 2004/0018626;US 2014/0178879;US 2016/0145646;WO 2016/081923;以及Frendewey等人(2010)《酶学方法》476:295-307,所述文献中的每个文献出于所有目的通过引用整体并入本文。例如,定量测定可以通过定量PCR进行,如实时PCR(qPCR)。实时PCR可以利用识别靶基因座的第一引物组和识别非靶向参考基因座的第二引物组。引物组可以包括识别扩增序列的荧光探针。

[0325] 合适的定量测定的其它实例包含荧光介导的原位杂交(FISH)、比较基因组杂交、等温DNA扩增、与固定化探针的定量杂交、INVADER®探针、TAQMAN®分子信标探针或ECLIPSE™探针技术(参见例如US 2005/0144655,所述文献出于所有目的通过引用整体并入本文)。

[0326] 合适的多能细胞的实例是胚胎干(ES)细胞(例如,小鼠ES细胞或大鼠ES细胞)。经修饰的多能细胞可以例如通过以下方式重组产生:(a)向细胞中引入包括插入核酸的一种或多种外源供体核酸(例如,靶向载体),所述插入核酸侧接有例如对应于5'和3'靶位点的5'和3'同源臂,其中插入核酸包括用于产生人源化F12基因座的人F12序列;以及(b)识别至少一种在其基因组中包括在内源F12基因座处整合的插入核酸的细胞(即,识别至少一种包括人源化F12基因座的细胞)。经修饰的多能细胞可以例如通过以下方式重组产生:(a)向细胞中引入包括插入核酸的一种或多种靶向载体,所述插入核酸侧接有对应于5'和3'靶位点的5'和3'同源臂,其中插入核酸包括人源化F12基因座;以及(b)识别至少一种在其基因组中包括在靶基因组基因座处整合的插入核酸的细胞。

[0327] 可替代地,经修饰的多能细胞可以通过以下方式产生:(a)向细胞中引入:(i)核酸酶试剂,其中核酸酶试剂在内源F12基因座内的靶位点处诱导切口或双链断裂;以及(ii)任选地包括插入核酸的一种或多种外源供体核酸(例如,靶向载体),所述插入核酸侧接有例如对应于定位在足够接近核酸酶靶位点的5'和3'靶位点的5'和3'同源臂,其中插入核酸包括用于产生人源化F12基因座的人F12序列;以及(c)识别至少一种在其基因组中包括在内源F12基因座处整合的插入核酸的细胞(即,识别至少一种包括人源化F12基因座的细胞)。可替代地,经修饰的多能细胞可以通过以下方式产生:(a)向细胞中引入:(i)核酸酶试剂,其中核酸酶试剂在靶基因组基因座内的识别位点处诱导切口或双链断裂;以及(ii)包括插

入核酸的一种或多种靶向载体,所述插入核酸侧接有对应于定位在足够接近识别位点的5'和3'靶位点的5'和3'同源臂,其中插入核酸包括人源化F12基因座;以及(c)识别至少一种在靶基因组基因座处包括修饰(例如,插入核酸的整合)的细胞。可以使用任何诱导切口或双链断裂进入期望识别位点的核酸酶试剂。合适的核酸酶的实例包含转录激活子样效应物核酸酶(TALEN)、锌指核酸酶(ZFN)、大范围核酸酶和成簇规律间隔短回文重复序列(CRISPR)/CRISPR相关(Cas)系统(例如,CRISPR/Cas9系统)或此类系统(例如,CRISPR/Cas9)的组分。参见例如US 2013/0309670和US 2015/0159175,所述文献中的每个文献出于所有目的通过引用整体并入本文。

[0328] 供体细胞可以在任何阶段引入到宿主胚胎中,如胚胎阶段或桑椹胚前阶段(即,4细胞阶段或8细胞阶段)。产生能够通过种系传递基因修饰的后代。参见例如美国专利第7,294,754号,所述文献出于所有目的通过引用整体并入本文。

[0329] 可替代地,产生本文别处描述的非人动物的方法可以包括:(1)使用上述用于修饰多能细胞的方法将单细胞阶段胚胎的基因组修饰成包括人源化F12基因座;(2)选择基因修饰的胚胎;以及(3)将基因修饰的胚胎植入到代孕母体中并在其中孕育。可替代地,产生本文别处描述的非人动物的方法可以包括:(1)使用上述用于修饰多能细胞的方法将单细胞阶段胚胎的基因组修饰成包括人源化F12基因座;(2)选择基因修饰的胚胎;(3)在代孕母体中孕育基因修饰的胚胎。产生能够通过种系传递基因修饰的后代。

[0330] 核转移技术还可以用于产生非人哺乳动物。简而言之,用于核转移的方法可以包括以下步骤:(1)去除卵母细胞的核或提供去核卵母细胞;(2)分离或提供要与去核卵母细胞结合的供体细胞或细胞核;(3)将细胞或细胞核插入到去核卵母细胞中以形成重组细胞;(4)将重组细胞植入到动物的子宫中以形成胚胎;以及(5)允许胚胎发育。在此类方法中,卵母细胞通常从死亡动物身上取得,但卵母细胞也可以从活体动物的输卵管和/或卵巢中分离出来。在去核前,卵母细胞可以在各种众所周知的培养基中成熟。可以以多种众所周知的方式执行卵母细胞的去核。可以通过在融合前在透明带下显微注射供体细胞来将供体细胞或细胞核插入到去核卵母细胞中以形成重组细胞。可以通过在接触/融合平面上施加DC电脉冲(电融合)、通过将细胞暴露于如聚乙二醇等促进融合的化学物质或通过如仙台病毒等灭活病毒的方式来诱导融合。在核供体和受体卵母细胞融合之前、期间和/或之后,可以通过电气方式和/或非电气方式来活化重组细胞。活化方法包含电脉冲、化学诱发的休克、精子渗透、增加卵母细胞中二价阳离子的水平以及减少卵母细胞中细胞蛋白质的磷酸化(如通过激酶抑制剂的方式)。活化的重组细胞或胚胎可以在众所周知的培养基中培养,并且然后转移到动物的子宫中。参见例如US 2008/0092249、WO 1999/005266、US 2004/0177390、WO 2008/017234和美国专利第7,612,250号,所述文献中的每个文献出于所有目的通过引用整体并入本文。

[0331] 本文所提供的各种方法允许产生基因修饰的非人F0动物,其中基因修饰的F0动物的细胞包括人源化F12基因座。应认识到,取决于用于生成F0动物的方法,F0动物内具有人源化F12基因座的细胞数量将有所不同。例如,对于小鼠,通过例如**VELOCIMOUSE[®]**方法将供体ES细胞引入到来自对应生物体(例如,8细胞阶段小鼠胚胎)的桑椹胚前阶段胚胎中允许更大百分比的F0小鼠的细胞群包括具有包括靶向基因修饰的所关注核苷酸序列的细胞。例如,非人F0动物的细胞贡献的至少50%、60%、65%、70%、75%、85%、86%、87%、

87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%可以包括具有靶向修饰的细胞群。

[0332] 基因修饰的F0动物的细胞对于人源化F12基因座可以是杂合的,或者对于人源化F12基因座可以是纯合的。

[0333] V. 评估或使用短F12异构体的方法

[0334] 提供了评估或使用在人神经元细胞中表达的短凝血因子XII同种型的各种方法。如实例3中所述,短FXII同种型可以由F12 mRNA编码,所述F12 mRNA由针对外显子11-14的探针检测,但不由针对外显子1-6的探针检测。在具体实例中,短同种型可以是FXII₂₉₇₋₅₉₆ (SEQ ID NO:74,由SEQ ID NO:75编码),一种含有FXII (M297-S596)的富脯氨酸和催化结构域的蛋白质。

[0335] 评估可以包括例如测量生物样品中短同种型的表达或测量短同种型的活性。可以通过使用针对外显子9、10、11、12、13或14的引物和/或探针(例如,用于定量PCR)如针对外显子11-14的探针在mRNA水平上测量短同种型的表达。这可以通过与不会检测到短同种型的针对外显子1、2、3、4、5或6的探针进行比较来完成。可替代地,可以通过使用针对外显子7、8、9、10、11、12、13或14的引物和/或探针(例如,用于定量PCR)如针对外显子11-14的探针在mRNA水平上测量短同种型的表达。这可以通过与不会检测到短同种型的针对外显子1、2、3、4、5或6的探针进行比较来完成。可替代地,可以通过使用针对外显子9、10、11、12、13或14的引物和/或探针(例如,用于定量PCR)如针对外显子11-14的探针在mRNA水平上测量短同种型的表达。这可以通过与不会检测到短同种型的针对外显子1、2、3、4、5、6、7或8的探针进行比较来完成。例如,此类方法可以包括使用如上所述针对F12基因的外显子的引物和/或探针检测和量化RNA转录物。这可以包括例如逆转录PCR (RT-PCR) 或定量PCR (例如,RT-qPCR)。在RT-qPCR中,RNA转录物首先通过反转录将其量化为cDNA,并且随后进行qPCR。与标准PCR一样,DNA通过3个重复步骤进行扩增:变性、退火和伸长。然而,在qPCR中,荧光标记使得能够随着PCR的进展收集数据。

[0336] 评估短同种型的活性可以包括例如测量肝细胞生长因子 (HGF) - 酪氨酸-蛋白激酶 Met (Met) 信号传导活性(例如,在神经元中),如评估pro-HGF向活性HGF的转化。此类测定还可以包含例如评估前激肽释放酶对激肽释放酶的活化作用、评估FXI的活化作用、评估pro-HGF对HGF的活化作用或评估经典补体途径的活化作用。

[0337] 评估(例如,测量表达或活性)可以在任何生物样品中进行,如任何器官、组织类型或细胞类型。例如,测量表达可以在脑中进行,如在神经元(例如,锥体神经元或皮层神经元)中进行。器官、组织或细胞可以来自人或者可以来自非人动物(例如,具有如本文别处详细描述的人源化F12基因座的非人动物)。

[0338] 出于所有目的,上文或下文引用的所有专利申请、网站、其它出版物、登录号等都通过引用整体并入,其程度如同每个单独的项目被单独并且具体地指出通过引用的方式并入。如果序列的不同版本与不同时间的登录号相关联,则意指在本申请的有效提交日期与该登录号相关联的版本。有效提交日期是指实际提交日期或提及登录号的优先权申请的提交日期(在适用情况下)中较早的日期。同样,如果出版物、网站等的不同版本在不同时间发布,除非另有说明,否则指在申请的有效提交日期最近发布的版本。除非另外具体说明,否则本发明的任何特征、步骤、元件、实施例或方面都可以与任何其它特征、步骤、元件、实施

例或方面结合使用。尽管为了清楚和理解起见,已通过图解和实例方式详细地对本发明进行了描述,但显而易见的是,可以在所附权利要求的范围内进行某些改变和修改。

[0339] 序列简要说明

[0340] 使用核苷酸碱基的标准字母缩写和氨基酸的三字母代码示出随附序列表中列出的核苷酸和氨基酸序列。核苷酸序列遵循从序列的5'端开始并且向前(即,在每行中从左到右)到达3'端的标准惯例。每个核苷酸序列仅示出一条链,但任何提及的显示链均应理解为包含互补链。当提供对氨基酸序列进行编码的核苷酸序列时,应理解还提供了对相同氨基酸序列进行编码的其密码子简并变体。氨基酸序列遵循从序列的氨基端开始并且向前(即,在每行中从左到右)到达羧基端的标准惯例。

[0341] 表2. 序列说明。

[0342]

序列号	类型	说明
1	蛋白质	小鼠凝血因子 XII 蛋白 (Q80YC5-1; NP_067464.2)
2	蛋白质	小鼠凝血因子 XII 蛋白信号肽
3	蛋白质	小鼠凝血因子 XII 蛋白重链
4	蛋白质	小鼠凝血因子 XII 蛋白轻链
5	蛋白质	人凝血因子 XII 蛋白
6	蛋白质	人凝血因子 XII 蛋白信号肽
7	蛋白质	人凝血因子 XII 蛋白重链
8	蛋白质	人凝血因子 XII 蛋白轻链
9	DNA	小鼠 <i>F12</i> CDS (CCDS36675.1)
10	DNA	小鼠 <i>F12</i> CDS 信号肽
11	DNA	小鼠 <i>F12</i> CDS 重链
12	DNA	小鼠 <i>F12</i> CDS 轻链
13	DNA	人 <i>F12</i> CDS
14	DNA	人 <i>F12</i> CDS 信号肽
15	DNA	人 <i>F12</i> CDS 重链
16	DNA	人 <i>F12</i> CDS 轻链
17	DNA	MAID7374 (具有 Neo 自缺失盒的 <i>F12</i> 人源化区域)
18	DNA	MAID7375 (不具有 Neo 自缺失盒的 <i>F12</i> 人源化区域)
19	DNA	7374hTU 正向
20	DNA	7374hTU 探针
21	DNA	7374hTU 反向
22	DNA	7374hTD 正向
23	DNA	7374hTD 探针
24	DNA	7374hTD 反向
25	DNA	7374mTU 正向
26	DNA	7374mTU 探针
27	DNA	7374mTU 反向
28	DNA	7374mTD 正向
29	DNA	7374mTD 探针
30	DNA	7374mTD 反向
31	DNA	MAID7374 中的人 <i>F12</i> 序列 (第 I 部分)
32	DNA	MAID7374 中的人 <i>F12</i> 序列 (第 II 部分)
33	DNA	小鼠 <i>F12</i> mRNA (NM_021489.3)
34	DNA	人 <i>F12</i> mRNA (NM_000505.3)
35	蛋白质	Cas9
36	DNA	Cas9
37	RNA	crRNA 尾部

[0343]

序列号	类型	说明
38	RNA	tracrRNA
39	RNA	向导 RNA 支架 v1
40	RNA	向导 RNA 支架 v2
41	RNA	向导 RNA 支架 v3
42	RNA	向导 RNA 支架 v4
43	DNA	向导 RNA 靶序列加 PAM v1
44	DNA	向导 RNA 靶序列加 PAM v2
45	DNA	向导 RNA 靶序列加 PAM v3
46	蛋白质	人凝血因子 XII 蛋白 (P00748-1; NP_000496.2)
47	蛋白质	人凝血因子 XII 蛋白重链 v2
48	DNA	人 <i>F12</i> CDS (CCDS34302.1)
49	DNA	人 <i>F12</i> CDS 重链 v2
50	DNA	F12 外显子 1-2 正向
51	DNA	F12 外显子 3-4 正向
52	DNA	F12 外显子 5-6 正向
53	DNA	F12 外显子 7 正向
54	DNA	F12 外显子 9 正向
55	DNA	F12 外显子 11 正向
56	DNA	F12 外显子 14 正向
57	DNA	GAPDH 正向
58	DNA	F12 外显子 1-2 反向
59	DNA	F12 外显子 3-4 反向
60	DNA	F12 外显子 5-6 反向
61	DNA	F12 外显子 7 反向
62	DNA	F12 外显子 9 反向
63	DNA	F12 外显子 11 反向
64	DNA	F12 外显子 14 反向
65	DNA	GAPDH 反向
66	DNA	F12 外显子 1-2 探针
67	DNA	F12 外显子 3-4 探针
68	DNA	F12 外显子 5-6 探针
69	DNA	F12 外显子 7 探针
70	DNA	F12 外显子 9 探针
71	DNA	F12 外显子 11 探针
72	DNA	F12 外显子 14 探针
73	DNA	GAPDH 探针
74	蛋白质	FXII ₂₉₇₋₅₉₆ 蛋白 (来自实例 3 的 ORF1)
75	DNA	FXII ₂₉₇₋₅₉₆ DNA (来自实例 3 的 ORF1)
76	蛋白质	FXII ₃₅₁₋₅₉₆ 蛋白 (来自实例 3 的 ORF2)
77	DNA	FXII ₃₅₁₋₅₉₆ DNA (来自实例 3 的 ORF2)
78	蛋白质	FXII ₅₂₇₋₅₉₆ 蛋白 (来自实例 3 的 ORF3)
79	DNA	FXII ₅₂₇₋₅₉₆ DNA (来自实例 3 的 ORF3)

[0344]

序列号	类型	说明
80	RNA	tracrRNA v2
81	RNA	tracrRNA v3
82	RNA	向导 RNA 支架 v5
83	RNA	向导 RNA 支架 v6
84	RNA	向导 RNA 支架 v7

[0345] 实例

[0346] 实例1.包括人源化凝血因子XII (F12) 基因座的小鼠的产生

[0347] 产生了大靶向载体 (LTVEC) 以用7.2kb (7,236bp) F12的对应的人序列 (来自基因组构建GRCh38/hg38) 来替换来自小鼠F12基因的8.6kb (8,670bp) 区域,所述大靶向载体包括包含62.4kb小鼠F12基因座 (来自基因组构建GRCm38/mm10) 的5'同源臂和包含107.2kb小鼠F12基因座 (来自基因组构建GRCm38/mm10) 的3'同源臂。关于小鼠和人F12的信息在表3中提供。关于LTVEC的信息在表4中提供。使用VELOCIGENE[®] 基因工程技术通过细菌同源重组 (BHR) 反应产生和使用来源于细菌人工染色体 (BAC) DNA的大靶向载体 (LTVEC) 描述于例如US 6,586,251和Valenzuela等人 (2003) 《自然生物技术》21 (6) :652-659,所述文献中的每个文献出于所有目的通过引用整体并入本文。通过体外组装方法产生LTVEC描述于例如US 2015/0376628和W0 2015/200334,所述文献中的每个文献出于所有目的通过引用整体并入本文。

[0348] 表3.小鼠和人凝血因子XII (F12)。

[0349]		官方符号	NCBI 基因 ID	主要来源	RefSeq mRNA ID	UniProt ID	基因组组装	位置
	小鼠	F12	58992	MGI: 1891012	NM_021489	Q80YC5	GRCm38/mm10	Chr 13
	人	F12	2161	HGNC: 3530	NM_000505	P00748	GRCh38/hg38	Chr 5

[0350] 表4.小鼠F12人源化靶向载体。

[0351]		基因组构建	开始	结束	长度 (bp)
	5'小鼠臂	GRCm38/mm10	Chr13: 55,426,769	Chr13: 55,489,207	62,439
	人插入物	GRCh38/hg38	Chr5: 177,402,292	Chr5: 177,409,527	7,236
	3'小鼠臂	GRCm38/mm10	Chr13: 55,310,888	Chr13: 55,418,098	107,211

[0352] 具体地,从小鼠F12基因座中缺失从ATG起始密码子到终止密码子的区域。插入从ATG起始密码子到终止密码子的人F12的对应区域代替所缺失的小鼠区域。loxP-mPrm1-Crei-pA-hUbl-em7-Neo-pA-loxP盒 (4,766bp) 包含在人内含子4内 (外显子4之后81bp和外显子5之前220bp)。自缺失盒上游的人F12区域包含3,874bp (人ATG到外显子4,以及81bp内含子4),并且自缺失盒下游的人F12区域包含3,362bp (220bp人内含子4,以及外显子5到终止密码子)。这是MAID 7374等位基因。参见图1的下部。盒缺失之后,loxP和克隆位点保留在人F12内含子4中。这是MAID 7375等位基因。

[0353] 小鼠凝血因子XII信号肽、重链和轻链的序列分别在SEQ ID NO:2-4中示出,其中对应的编码序列分别在SEQ ID NO:10-12中示出。人凝血因子XII信号肽、重链和轻链的序列分别在SEQ ID NO:6-8中示出,其中对应的编码序列分别在SEQ ID NO:14-16中示出。预期的经编码的人源化凝血因子XII蛋白与人凝血因子XII蛋白相同。参见图1。小鼠和人凝血因子XII蛋白的比对在图3中提供。小鼠和人F12编码序列分别在SEQ ID NO:9和13中示出。小鼠和人凝血因子XII蛋白序列分别在SEQ ID NO:1和5中示出。预期的人源化ALB编码序列和预期人源化凝血因子XII蛋白的序列分别在SEQ ID NO:13和5中示出。

[0354] 为了产生突变等位基因,上述大靶向载体被引入到F1H4小鼠胚胎干细胞中。F1H4小鼠ES细胞源自通过将雌性C57BL/6NTac小鼠与雄性12956/SvEvTac小鼠杂交产生的杂交

胚胎。参见例如US 2015-0376651和WO 2015/200805,所述文献中的每个文献出于所有目的整体并入本文。选择抗生素后,通过TAQMAN®挑选、扩展和筛选菌落。参见图2。执行等位基因丢失测定以检测内源小鼠等位基因的丢失,并且使用表5中所示的引物和探针执行等位基因获得测定以检测人源化等位基因的获得。

[0355] 表5.筛选测定。

测定	说明	引物/探针	序列
7374hTU	上游人插入	正向	TCGGTGGCAGGCTATGACTTATAG (SEQ ID NO: 19)
		探针 (FAM)	CAGTTCCTGCCTTCTTCTCTCCC (SEQ ID NO: 20)
		反向	GGCTTCCCAAGGTGGAATCTAC (SEQ ID NO: 21)
7374hTD	下游人插入	正向	AAGGGCATGAGTGGGTTTACAAG (SEQ ID NO: 22)
		探针 (Cal)	CGCCTGGAGCAGCTTTGTCCATC (SEQ ID NO: 23)
		反向	ACACAGAGCGCCTTCTTCACA (SEQ ID NO: 24)
7374mTU	上游小鼠 LOA	正向	GTTCTGCCTTCTCTCTCCTA (SEQ ID NO: 25)
		探针 (FAM)	TAGGCTCCACCATGGAAAGACTCCA (SEQ ID NO: 26)
		反向	CCCATCAGGTGCGTCCTTA (SEQ ID NO: 27)
7374mTD	下游小鼠 LOA	正向	TCGCTGCTCCAACTCTAACG (SEQ ID NO: 28)
		探针 (Cal)	ACGCCATTCTCCCTGGGATGCTT (SEQ ID NO: 29)
		反向	ATCGGTGCCTCCCTCCAAGAAG (SEQ ID NO: 30)

[0357] 包含等位基因丢失 (LOA) 和等位基因获得 (GOA) 测定的等位基因修饰 (MOA) 测定描述于例如US 2014/0178879;US 2016/0145646;WO 2016/081923;以及Frendewey等人 (2010)《酶学方法》476:295-307,所述文献中的每个文献出于所有目的通过引用整体并入本文。等位基因丢失 (LOA) 测定颠覆了传统的筛选逻辑,并量化了突变所针对的天然基因座的基因组DNA样品中的拷贝数。在正确靶向的杂合细胞克隆中,LOA测定检测到两个天然等位基因中的一个等位基因 (对于不在X或Y染色体上的基因),另一个等位基因被靶向修饰破坏。可以反向应用相同的原理作为等位基因获得 (GOA) 测定,以量化基因组DNA样品中插入的靶向载体的拷贝数。

[0358] 使用 VELOCIMOUSE®方法从经修饰的ES细胞中产生F0小鼠。具体地,使用 VELOCIMOUSE®方法将通过上述MOA测定选择的小鼠ES细胞克隆注射到8细胞阶段胚胎中。参见例如US 7,576,259;US 7,659,442;US 7,294,754;US 2008/0078000;以及Poueymirou等人 (2007)《自然生物技术》25(1):91-99,所述文献中的每个文献出于所有目的通过引用整体并入本文。在VELOCIMOUSE®方法中,通过激光辅助注射将靶向小鼠胚胎干 (ES) 细胞注射到桑椹胚前阶段胚胎,例如,八细胞阶段胚胎中,这高效地产生完全由ES细胞衍生的F0代小鼠。

[0359] 实例2.包括人源化凝血因子XII (F12) 基因座的小鼠的验证

[0360] 为了验证人源化F12小鼠,使用人凝血因子XII ELISA试剂盒 (hFXII总抗原ELISA试剂盒,分子创新公司 (Molecular Innovations) 目录#:HFXIIKT-TOT) 测量了来自盒缺失人源化F12小鼠的血浆样品中的人凝血因子XII水平。在正常人血浆中检测到的人凝血因子XII的水平为23.6μg/mL,并且在来自人源化F12小鼠的血浆中为5-18μg/mL。在野生型小鼠血浆中检测到很少的量,这可能是由于与小鼠凝血因子XII的交叉反应很小。参见图4。

[0361] 还对血浆样品进行了蛋白质印迹分析(实例3中提供的方案),以测量野生型小鼠和人源化F12小鼠中小鼠凝血因子XII和人凝血因子XII的表达。血浆的蛋白质印迹分析显示野生型(FXII^{+/+})小鼠血浆中存在小鼠因子XII且缺乏人因子XII,而人源化F12(FXII^{hum/hum})小鼠血浆中存在人因子XII且缺乏小鼠因子XII。在F12敲除小鼠血浆中未检测到人或小鼠因子XII。参见图5。

[0362] 为了进一步验证人源化F12小鼠,执行了两种功能性血浆测定:(1)活化部分凝血活酶时间(aPTT)和(2)前凝血酶时间(PT)。aPTT测试通过测量添加钙和鞣花酸后凝块形成的时间来评估凝血级联的内在和共同途径的所有凝结因子,而PT测试通过测量添加钙和组织因子后凝块形成的时间来评估凝血级联的外在和共同途径的所有凝结因子。

[0363] 在野生型小鼠与人源化F12小鼠之间未观察到aPTT凝血时间的显著差异。参见图6。类似地,野生型小鼠与人源化F12小鼠之间的PT值没有显著差异,其中每个值为9-11秒(数据未示出)。

[0364] 凝血测定:PT-前凝血酶时间

[0365] 此测定用作术前筛查以检测潜在的出血问题。其对外在凝血系统(因子II、V、VII和X;正常人血浆PT为约10-13秒)的缺陷很敏感。所使用的仪器是Start4止血分析仪(思塔高诊断公司(Diagnostica Stago))。所使用的试剂是6mL TriniCLOT PT Excel(思塔高公司T1106)。示例性程序如下:

[0366] 1. 打开分析仪。机器会自动执行“自检查”并预热。测定仅在37°C下执行。

[0367] 2. “测试模式”按键盘上的“1”键,按回车键确认,PT测试按“1”键,并且然后按回车键确认。

[0368] 3. 输入患者ID#或样品#以显示在工作屏幕上。

[0369] 4. 将比色皿条(cuvette-strip)放置在温育区,在37°C下预热。

[0370] 5. 向比色皿条中的每个比色皿分配小球。

[0371] 6. 在每个比色皿中分配50μL血浆样品(对照样品、患者样品或药物治疗样品)(不要产生气泡)。

[0372] 7. 任选地在血浆中添加5μL 2X系列稀释抑制剂(仅在测试药物剂量反应时),并将混合物在37°C下温育5分钟。

[0373] 8. 启动与温育柱对应的计时器(按计时器键)温育180秒。当仪器发出哔哔声时,快速将比色皿条转移到测试柱上。

[0374] 9. 用起始试剂(在37°C下预热的PT试剂)灌注芬兰雷勃移液器(Finnpipette)。

[0375] 10. 在测试柱上,通过按下移液器控制键活化芬兰雷勃移液器(小球开始在比色皿中来回移动),并且然后将预热的100μL PT试剂分配到每个比色皿。

[0376] 11. 当凝块形成时,小球会停止移动并且凝结时间将被记录(在同一比色皿条上的所有四个凝结时间都完成之后打印结果)。

[0377] 经修饰的aPTT:活化部分凝血活酶时间

[0378] 此测定用于检测内在凝血系统的异常,并对因子VIII、IX、X、XI和XII的缺陷(正常人血浆中28-34秒的aPTT)很敏感。所使用的仪器是Start4止血分析仪(思塔高诊断公司,手册参考:0931079A)。所使用的试剂如下:4mL APTT-XL鞣花酸ACTCV(赛默飞世尔公司(Thermo Scientific),目录#:95059-804)、10mL 0.02M CaCl₂(赛默飞世尔公司,目录#:

95059-808)或10mL TriniCLOT aPTT S(思塔高公司T1201)。示例性程序如下:

- [0379] 1. 打开分析仪。机器会自动执行“自检”并预热。测定仅在37℃下执行。
- [0380] 2. “测试模式”按键盘上的“1”键,按回车键确认,APTT测试按“2”键,并且然后按回车键确认。
- [0381] 3. 输入患者ID#或样品#以显示在工作屏幕上。
- [0382] 4. 将比色皿条放置在温育区,在37℃下预热。
- [0383] 5. 向比色皿条中的每个比色皿分配小球。每个比色皿条上连接有4个比色皿。
- [0384] 6. 在每个比色皿中分配50μL血浆样品(对照样品、患者样品或药物治疗样品)(用移液器慢慢地吸取血浆,不要产生气泡),并在37℃下加热1分钟。
- [0385] 7. 任选地在血浆中添加5μL 2X系列稀释抑制剂(仅在测试药物剂量反应时),并将混合物在37℃下温育5分钟。
- [0386] 8. 向每个比色皿添加50μL触发物(即,鞣花酸或高岭土)。
- [0387] 9. 启动与温育柱对应的计时器(按计时器键)温育300秒,并且然后快速将比色皿条转移到测试柱上。
- [0388] 10. 用起始试剂(在37℃下预热的0.02M CaCl_2)灌注芬兰雷勃移液器。
- [0389] 11. 在测试柱上,通过按下移液器控制键活化芬兰雷勃移液器(小球开始在比色皿中来回移动),并且然后将预热的50μL CaCl_2 分配到每个比色皿。
- [0390] 12. 当凝块形成时,小球会停止移动并且凝结时间将被记录。在记录所有四个比色皿的凝结时间之后结果将自动打印出来。

[0391] 实例3. 人与小鼠F12基因组序列的差异可能会驱动人短同种型的表达,所述人短同种型是一种在人源化F12小鼠中重现的表型

[0392] 凝血因子XII (FXII) 由肝脏中的肝细胞合成并分泌到循环中,在所述循环中其启动接触活化系统。尽管通常被认为仅限于循环,但在阿尔茨海默病和多发性硬化症患者的脑中已发现了FXII蛋白,并且在人脑和脑脊髓液中观察到了接触系统活化。然而,FXII蛋白在脑中的来源尚未阐明,并且其在脑中的潜在作用尚不明确。使用原位杂交和RNAseq,示出了较短的FXII同种型由人和FXII人源化小鼠脑中的神经元表达,其中在锥体神经元中观察到的表达最高。这种较短的F12 mRNA转录物含有编码FXII跨越其富脯氨酸和催化结构域的部分的开放阅读框。这种较短的FXII蛋白的重组版本由血浆激肽释放酶活化,并将促肝细胞生长因子(HGF)转化为活性HGF。HGF-Met信号传导在神经元发育和存活中发挥作用,并且其调节异常与神经发育障碍和神经变性有关。在此首次示出了在脑中局部产生F12 mRNA的短同种型,并提出了脑源性FXII可能参与神经元中的HGF-Met信号传导的可能性。

[0393] 凝血因子XII (FXII) 是一种启动接触活化系统的循环丝氨酸蛋白酶。在带负电荷的表面上或通过血浆激肽释放酶活化FXII会产生活化的FXII (FXIIa)。FXIIa启动内在凝血级联反应,从而导致凝血酶生成和凝块形成。FXIIa还将血浆前激肽释放酶转化为激肽释放酶,这活化了激肽释放酶-激肽途径,从而导致血管活性和促炎肽缓激肽的释放。除了这些经过充分研究的功能外,FXIIa还被证明在体外活化肝细胞生长因子(HGF)和补体系统。参见例如Shimomura等人(1995)《欧洲生物化学杂志(Eur. J. Biochem.)》229:257-261; Peek等人(2002)《生物化学杂志》277:47804-47809; 以及Ghebrehiwet等人(1981)《实验医学杂志(J. Exp. Med.)》153:665-676,所述文献中的每个文献出于所有目的通过引用整体并入本

文。FXII作为单链酶原循环,所述单链酶原在活化期间在R353处切割后变成具有酶活性的双链分子(α FXIIa)。 α FXIIa被进一步加工成 β FXIIa,其由含有催化结构域的轻链和仅一小部分重链组成。

[0394] FXII由肝脏中的肝细胞合成并分泌到循环中。尽管通常被认为仅限于循环,但已发现FXII蛋白与 β 阿尔茨海默病(AD)患者的脑和多发性硬化症(MS)患者的脑损伤中的斑块共存。参见例如Yasuhara等人(1994)《大脑研究(Brain Res.)》654:234-240和Gobel等人(2016)《自然通讯》7:11626,所述文献中的每个文献出于所有目的通过引用整体并入本文。由于血脑屏障(BBB)通常会阻止循环蛋白质移动到脑实质中,因此FXII不应进入健康的人脑。因此,在AD和MS患者的脑中观察到的FXII可能来自通过功能失调BBB的循环,或者可能是局部合成的。先前在脑中发现的FXII蛋白的来源是外周的还是局部的还没有得到很好的表征。参见例如Yasuhara等人(1994)《大脑研究》654:234-240和Gobel等人(2016)《自然通讯》7:11626,所述文献中的每个文献出于所有目的通过引用整体并入本文。

[0395] 脑中的FXII可能会触发接触系统的局部活化,这已在人脑和脑脊液中观察到。参见例如Ashby等人(2012)《衰老神经生物学(Neurobiol.Aging)》33:1345-1355和Bergamaschini等人(2001)《衰老与发育机制(Mech.Ageing Dev.)》122:1971-1983,所述文献中的每个文献出于所有目的通过引用整体并入本文。脑FXII还可以活化HGF,后者通过其受体Met或介导突触完整性和可塑性的补体系统在神经元发育和存活中发挥作用。参见例如Kato(2017)《生物医学报告(Biomed.Rep.)》7:495-503;Wright和Harding(2015)《阿尔茨海默病杂志(J.Alzheimers Dis.)》45:985-1000;以及Hammond等人(2018)《细胞与发育生物学年鉴(Annu.Rev.Cell Dev.Biol.)》34:523-544,所述文献中的每个文献出于所有目的通过引用整体并入本文。很可能在病理上发生FXII从外周泄漏,而脑中的局部F12 mRNA合成可能是这些系统的生理活化在脑中的机制。此处示出了潜在地编码 β FXIIa样蛋白的新型F12 mRNA转录物由人脑中的神经元表达,并且这种假定的脑FXII蛋白可以由血浆激肽释放酶活化,并将pro-HGF转化为活性HGF。

[0396] 已发表的研究示出了关于F12在脑中表达的相互矛盾的结果。一项使用跨越外显子7-9的PCR引物的研究检测到了脑F12 mRNA(Yasuhara等人(1994)《大脑研究》654:234-240,所述文献出于所有目的通过引用整体并入本文),而使用外显子3-5内的引物的另一项研究示出几乎没有F12信号(Neth等人(2001)《血栓形成与止血(Thromb.Haemost.)》85:1043-1047,所述文献出于所有目的通过引用整体并入本文)。为了研究这两项研究之间的差异是否可能是由于针对检测选择的外显子造成的,我们针对F12中的七个区域(外显子1-2、外显子3-4、外显子5-6、外显子7、外显子9、外显子11和外显子14;图7A和表6)设计了qPCR探针。我们发现,虽然所有探针都在人肝脏中检测到F12,但只有针对外显子7、9、11和14的探针在人脑中检测到F12(图7B)。

[0397] 表6.qPCR引物和探针。

探针靶标	正向引物	反向引物	探针序列
F12 外显子 1-2	TGGACCAACGGACGG ATG (SEQ ID NO: 50)	CCCAAGGTGGAATCGA AAGTG (SEQ ID NO: 58)	CATGAGGGCTCTGCTGCTCC TGG (SEQ ID NO: 66)
F12 外显子 3-4	TGCCACTTCCCCTTCC AG (SEQ ID NO: 51)	GTAGCACACCAGGGCT GAG (SEQ ID NO: 59)	ACCACCGGCAGCTGTACCAC AAAT (SEQ ID NO: 67)
F12 外显子 5-6	ACTGTCTCTGTCCACA ACACC (SEQ ID NO: 52)	CGGAGAAGCTGAGGCT CAAAG (SEQ ID NO: 60)	TCCTGGAAACCACTGCCA GAAAGAGA (SEQ ID NO: 68)
[0398] F12 外显子 7	ACCAACCCGTGCCTCC AT (SEQ ID NO: 53)	GCAGAAGGCTCCGGTG TAG (SEQ ID NO: 61)	CTGCCTAGAGGTGGAGGGC CACC (SEQ ID NO: 69)
F12 外显子 9	GTGGTGCTTCGTGCTG AAC (SEQ ID NO: 54)	GTTGGGGTCTGGCACTG TG (SEQ ID NO: 62)	CGACCGGCTGAGCTGGGAG TACT (SEQ ID NO: 70)
F12 外显子 11	AGGATCTGACGGTGGT GCTC (SEQ ID NO: 55)	GTAGCTGACGGGCGAG AAG (SEQ ID NO: 63)	CCAGGAACGCCGTAACCAC AGCT (SEQ ID NO: 71)
F12 外显子 14	ACACCGATGTGGCCTA CTAC (SEQ ID NO: 56)	ACTGCGGAATCACCAA GGA (SEQ ID NO: 64)	CTGGATCCGGGAGCACACC GTTT (SEQ ID NO: 72)
GAPDH	CCAGGTGGTCTCCTCT GACT (SEQ ID NO: 57)	GCTTGACAAAGTGGTC GTTGA (SEQ ID NO: 65)	TCAACAGCGACACCCACTCC TC (SEQ ID NO: 73)

[0399] 我们的结果由可通过基因型组织表达 (GTEx) 项目获得的RNAseq数据支持 (G.T. 联盟, 基因型组织表达 (GTEx) 项目, 《自然遗传 (Nat. Genet.)》, 45 (2013) 580-585, 所述文献出于所有目的通过引用整体并入本文)。数据示出肝脏中F12 mRNA的表达高, 而在其它组织中表达要低得多。除了肝脏之外, 最高水平的F12表达中的一些在脑区域中被发现 (图10A)。正如预期的, 对人肝脏中F12 mRNA读段的逐个外显子分析示出覆盖所有14个外显子。然而, 人皮层和海马体中的F12 mRNA读段仅覆盖外显子7与外显子14之间 (图7C)。

[0400] 接下来使用RNA原位杂交来研究F12 mRNA在人海马体中的分布。为此目的, 设计了两种针对F12的RNAscope探针: 一种针对外显子1-6, 一种针对外显子11-14。与我们的qPCR结果和公共GTEx数据一致, F12 mRNA仅由针对人海马体中外显子11-14的探针检测到 (数据未示出)。F12 mRNA信号最丰富的区域是海马锥体层, 其中一些信号在齿状回的颗粒细胞中 (数据未示出)。苏木精染色的细胞核外部的RNAscope信号表明探针检测的不是pre-mRNA, 而是成熟的mRNA。由于难以获得适合RNAscope分析的始终高RNA质量的人脑组织, 我们转向小鼠组织以进一步研究F12在脑中的表达。

[0401] 为了研究F12 mRNA在小鼠脑中的分布, 对野生型小鼠 (FXII^{+/+}) 执行RNAscope分析, 使用不产生因子XII蛋白的F12敲除小鼠 (FXII^{-/-}) 作为对照。正如预期的, 外显子1-6和外显子11-14探针两者均在野生型小鼠肝脏中检测到F12 (数据未示出)。与人脑类似, 只有针对外显子11-14的探针在野生型小鼠脑中检测到F12 mRNA (数据未示出), 而在FXII^{-/-}小鼠的脑或肝脏中未检测到信号, 证实了探针对F12的特异性。来自小鼠脑中F12外显子11-14探针的信号量似乎比在人脑中观察到的要少, 这促使我们比较小鼠与人体组织中肝外F12mRNA的表达水平。与F12在人类中的显著肝外表达相比, 在包含脑的几种组织具有可检测F12 mRNA水平的情况下 (图10A), 未在肝脏外部检测到F12在小鼠中的表达 (图10B)。这些数据支持通过RNAscope在相对于人脑的野生型小鼠脑中检测到的F12 mRNA水平, 并表明通过RNAscope在小鼠脑中发现的F12 mRNA水平不足以通过RNAseq在大块组织中检测到。

[0402] 为了研究人与小鼠脑中短F12 mRNA的差异表达是由于人F12基因内的特定元件还

是由于更普遍的种间差异,接下来分析了仅产生人因子XII的F12人源化小鼠(FXII^{hum/hum})。在FXII^{hum/hum}小鼠肝脏中两个探针都检测到全长人F12 mRNA(数据未示出),但只有针对人F12外显子11-14的探针在FXII^{hum/hum}脑中产生信号。与野生型小鼠相比,FXII^{hum/hum}小鼠脑具有高得多的信号,这表明与小鼠F12相比,人F12可能具有有利于其在脑中的表达的明显调节特征。与在人脑中观察到的模式类似,F12在FXII^{hum/hum}小鼠脑中的表达似乎主要是神经元性的,在海马锥体层和皮层神经元中有强烈的信号(数据未示出)。在主要由神经胶质构成的胼胝体和侧脑室内衬的室管膜细胞中几乎没有观察到信号(数据未示出)。

[0403] 为了进一步研究F12的潜在神经元表达,将RNAscope技术与免疫荧光相结合,使用针对F12外显子11-14的探针和FXII^{hum/hum}小鼠脑的连续切片中的神经元标志物NeuN。观察到连续脑切片中NeuN免疫荧光和F12 RNAscope信号分布的重叠(数据未示出),这表明F12主要在神经元中表达。在海马体的CA2/CA3区域的锥体神经元中检测到F12的最强表达(数据未示出)。

[0404] 由于F12的表达似乎主要是神经元性的,接下来研究了F12 mRNA是否可以通过RNAseq在原代神经元和神经元细胞系中检测到。事实上,F12 mRNA在iPS衍生的人神经元中表达(5.82[1.65-7.38]FPKM),但在小鼠原代皮层神经元中未检测到(数据未示出)。与人脑组织类似,逐个外显子mRNA读段分析示出了人iPS神经元表达较短的F12 mRNA同种型(数据未示出)。人与小鼠之间的脑F12 mRNA表达的这种有趣差异促使我们着眼于物种之间F12基因的差异。

[0405] 我们的结果提出了人F12基因组序列可能具有驱动神经元中较短同种型的表达的不同内部调节元件的可能性。事实上,可通过UCSC浏览器获得的数据(Kent等人(2002)《基因组研究(Genome Res.)》12:996-1006,所述文献出于所有目的通过引用整体并入本文)示出,尽管多个调节元件定位在跨越在人和小鼠F12基因两者中的外显子7-10的区域中,人F12的独特特征是跨越外显子8-12的大CpG岛。CpG岛是基因组中通常与转录起始位点相关的区域,并且表征为升高的G和C碱基组成以及较高的CpG二核苷酸频率。在许多情况下,与已知转录物不相关的位点处存在CpG岛导致新转录物的发现(Deaton等人(2011)《基因与发育(Genes Dev.)》25:1010-1022,所述文献出于所有目的通过引用整体并入本文)。因此,在人而非小鼠F12序列中存在大CpG岛可能是其在人脑中转录增加的原因。

[0406] 接下来检查了在人脑中检测到的F12 mRNA序列是否存在开放阅读框(ORF),以预测将被翻译的假定蛋白质。在人神经元F12序列(图8B)中发现了三个由具有全长F12的框架中的起始密码子和终止密码子的存在定义的ORF。在这三个ORF中,只有ORF 2和3在小鼠中是保守的,其中ORF 1区域中的氨基酸序列比对较差。有趣的是,GENBANK®(Benson等人(2013)《核酸研究》41:D36-42,所述文献出于所有目的通过引用整体并入本文)列出了在对应于ORF 1的外显子9中开始的五个短人F12 mRNA,但是没有识别在外显子10(对应于ORF 2)中开始的人F12 mRNA。没有描述在外显子9或10处开始的小鼠F12 mRNA(图8A)。在人外显子9中发现的ATG在小鼠中并不保守。总之,这些数据支持在外显子9中开始并在人类中编码ORF 1的短F12同种型的更高转录,并确定人和小鼠F12基因中序列和调节元件的差异,这可能有助于解释短F12 mRNA在小鼠脑中的表达。

[0407] 此处,使用qPCR、RNAscope和RNAseq示出F12 mRNA在人和FXII^{hum/hum}小鼠脑的神经元中表达。使用RNAscope,将神经元确定为在脑中表达短F12 mRNA的主要细胞类型,并通过

RNAseq确认其在iPS衍生的谷氨酸能神经元中的表达。有趣的是,通过RNAscope观察到与野生型小鼠脑相比,F12 mRNA在FXII^{hum/hum}小鼠和人类脑中的水平要高得多。RNAseq数据进一步支持了这种种间差异,所述数据示出与人脑相比,F12在小鼠脑中的表达可以忽略不计。对人和小鼠F12基因中的调节元件的分析表明,在外显子9中开始的人F12转录物的表达可能是由于在人而非小鼠F12中发现的跨越外显子8-12的大CpG岛造成的。

[0408] 为了确定短脑FXII同种型的可能的功能作用,评估了脑F12转录物中的三个预测的ORF(图8B)。ORF 1的翻译将产生含有FXII的富脯氨酸和催化结构域的蛋白质(M297-S596;SEQ ID NO:74(蛋白质)和75(DNA))。ORF 2的翻译将产生包括催化结构域和来自重链的三个另外的残基的蛋白质(M351-S596;SEQ ID NO:76(蛋白质)和77(DNA))。ORF 3将产生包含催化结构域中不含催化三联体的部分的蛋白质(M527-S596;SEQ ID NO:78(蛋白质)和79(DNA))。因为由ORF 2产生的蛋白质将具有未配对的半胱氨酸(C467),有可能导致不稳定,并且由于由ORF 3产生的蛋白质中缺乏完整的催化结构域,所以决定聚焦于将由ORF 1产生的蛋白质(FXII₂₉₇₋₅₉₆)。在ORF 1而不是ORF 2和3中开始的GENBANK®中的F12转录物的存在进一步支持了这种选择(图8A)。

[0409] 评估了假定的神经元FXII同种型(FXII₂₉₇₋₅₉₆)是否可以活化由全长FXII触发的途径。由血浆激肽释放酶在R353处切割是全长FXII的活化所需要的,并且导致重链和轻链的分离。测试了缺乏重链中的大部分重链的重组人FXII₂₉₇₋₅₉₆是否需要由血浆激肽释放酶切割以获得酶活性。如图8B所示,当FXII₂₉₇₋₅₉₆在不存在激肽释放酶的情况下用FXIIa底物S-2302温育时没有观察到酶活性,而FXII₂₉₇₋₅₉₆在用激肽释放酶预温育后获得酶活性(图9A)。在添加显色底物之前,所有反应都用激肽释放酶抑制剂抑肽酶进行处理,以确保酶活性不是底物的激肽释放酶介导的裂解的结果。这些结果表明,与全长FXII类似,假定脑FXII同种型在由激肽释放酶切割后被活化。接下来检查活化的FXII₂₉₇₋₅₉₆是否能够将前激肽释放酶相互活化为激肽释放酶。虽然用未活化的FXII₂₉₇₋₅₉₆未观察到前激肽释放酶的裂解,但活化的FXII₂₉₇₋₅₉₆通过激肽释放酶重链的生成将前激肽释放酶转化为激肽释放酶(图11),如所看到的。因此,FXII₂₉₇₋₅₉₆能够启动激肽释放酶-激肽途径,这表明FXII₂₉₇₋₅₉₆活化可以导致缓激肽的产生。

[0410] 最近的研究表明,在富脯氨酸的结构域内切割FXII释放重链的一部分重链,增加了FXII对R353处的激肽释放酶活化的敏感性。参见例如de Maat等人(2019)《血栓形成与止血杂志(J.Thromb.Haemost.)》17:183-194,所述文献出于所有目的通过引用整体并入本文。假设富脯氨酸结构域的重链N端缺失的FXII₂₉₇₋₅₉₆基本上会被预切割,并且可能因此容易受到不会活化全长FXII的激肽释放酶水平的活化。因为本研究中所使用的重组FXII₂₉₇₋₅₉₆是在大肠杆菌中产生的并且缺乏糖基化,所以无法直接将其对活化的敏感性与从人血浆中纯化的全长FXII的进行比较。因此,我们试图从全长FXII酶促地产生类似于FXII₂₉₇₋₅₉₆的蛋白质。使用Prosper软件(Song等人(2012)《公共科学图书馆·综合(PLoS One)》7:e50300,所述文献出于所有目的通过引用整体并入本文)对FXII氨基酸序列中的假定切割位点的分析揭示了L296与M297之间的组织蛋白酶K切割位点,其将产生与可能在脑中表达的同种型相同的截短FXII蛋白(FXII₂₉₇₋₅₉₆)。

[0411] 如图9B所示,用组织蛋白酶K温育全长FXII产生了大小支持L296与M297之间的切割的稳定的FXII片段,所述片段将产生两个约35kDa的片段,由于糖基化,预计所述两个片

段将在约40kDa下运行。接下来比较了全长FXII和通过组织蛋白酶K切割产生的截短FXII的酶活性。正如预期的,全长FXII和组织蛋白酶K切割的FXII本身的活性可以忽略不计,但是当用血浆激肽释放酶温育时,组织蛋白酶K切割的FXII的活性与全长FXII相比显著增加(图9B)。我们的结果与最近的研究一致,示出了组织蛋白酶K与激肽释放酶协同作用以活化FXII(de Maat等人(2019)《血栓形成与止血杂志》17:183-194,所述文献出于所有目的通过引用整体并入本文),并表明与组织蛋白酶-K切割的FXII类似的假定脑FXII₂₉₇₋₅₉₆同种型可能对血浆激肽释放酶和可能地其它蛋白酶的活化具有增加的敏感性。

[0412] 除了活化前激肽释放酶和内在凝血途径外,FXIIa已被证明将pro-HGF转化为活性HGF。参见例如Shimomura等人(1995)《欧洲生物化学杂志》229:257-261,以及Peek等人(2002)《生物化学杂志》277:47804-47809,所述文献中的每个文献出于所有目的通过引用整体并入本文。证实了肝细胞生长因子激活子(HGFA)和 β FXIIa切割pro-HGF的能力(图9C)。另外,我们发现激肽释放酶活化的FXII₂₉₇₋₅₉₆将pro-HGF切割成活性双链形式(图9D),这表明脑FXII可能有助于局部HGF活化。这种活化pro-HGF的能力是特异性的,因为在脑中发现的如凝血酶和组织纤溶酶原激活子等其它丝氨酸蛋白酶并未将pro-HGF转化为HGF(图9C和9D)。

[0413] 越来越多的证据表明,凝血和接触系统蛋白的表达发生在肝脏外部,尤其是在病理条件下。参见例如Ashby等人(2012)《衰老神经生物学》33:1345-1355;Yin等人(2010)《美国病理学杂志(Am. J. Pathol.)》176:1600-1606;以及Takano等人(2003)《大脑研究》978:72-82,所述文献中的每个文献出于所有目的通过引用整体并入本文。肝外FXII表达的作用也正在显现:最近的研究表明,FXII在肺成纤维细胞中表达(Jablonska等人(2010)《生物化学杂志》285:11638-11651,所述文献出于所有目的通过引用整体并入本文),并且中性粒细胞表达的FXII有助于中性粒细胞活化和到炎症部位的迁移(Stavrou等人(2018)《临床研究杂志(J. Clin. Invest.)》128:944-959,所述文献出于所有目的通过引用整体并入本文)。此处,首次使用qPCR、RNAscope和RNAseq示出F12 mRNA在人和FXII^{hum/hum}小鼠脑的神经元中表达。先前的研究产生了关于人脑中的F12表达的相互矛盾的结果。使用跨越外显子7到9的引物,一个组检测到了F12 mRNA(Yasuhara等人(1994)《大脑研究》654:234-240,所述文献出于所有目的通过引用整体并入本文),而使用外显子3-5内的引物和探针执行的qPCR示出几乎没有F12信号(Neth等人(2001)《血栓形成与止血》85:1043-1047,所述文献出于所有目的通过引用整体并入本文)。这些研究使用了不同的方法,并且因此不可直接比较;然而,Yasuhara等人得出的结论是F12 mRNA存在于脑中,而Neth等人得出的结论是F12 mRNA仅存在于肝脏中。我们的结果有助于解决这种差异,因为我们发现对应于外显子7-14而不是外显子1-6的F12 mRNA存在于脑中(图7C),由此支持这两个报告。

[0414] 使用RNAscope,将神经元确定为在脑中表达短F12 mRNA的主要细胞类型,并通过RNAseq确认其在iPS衍生的谷氨酸能神经元中的表达。有趣的是,通过RNAscope观察到与WT小鼠脑相比,F12 mRNA在FXII^{hum/hum}小鼠和人类脑中的水平要高得多。RNAseq数据进一步支持了这种种间差异,所述数据示出与人脑相比,F12在小鼠脑中的表达可以忽略不计。对人和小鼠F12基因中的调节元件的分析表明,在外显子9中开始的人F12转录物的表达可能是由于在人而非小鼠F12中发现的跨越外显子8-12的大CpG岛造成的。人与小鼠神经元之间的这种差异的功能意义尚不明确。

[0415] 最近的工作表明了跨越FXII₂₉₇₋₅₉₆的潜在脑FXII蛋白的功能含义。当FXII首先由如弹性蛋白酶或组织蛋白酶K等其它酶处理时,可以通过亚阈值水平的血浆激肽释放酶实现FXII活化。参见例如de Maat等人(2019)《血栓形成与止血杂志》17:183-194,所述文献出于所有目的通过引用整体并入本文。FXII重链的切割被认为揭示了FXII活化切割位点(R353-V354),从而使其在溶液中更高效地活化。参见例如de Maat等人(2019)《血栓形成与止血杂志》17:183-194,所述文献出于所有目的通过引用整体并入本文。基于计算机分析,假定神经元FXII₂₉₇₋₅₉₆可以通过用组织蛋白酶K切割全长FXII产生。虽然FXII上的确切组织蛋白酶K切割位点仍有待显示,但我们的关于组织蛋白酶K切割的FXII的结果表明,假定神经元FXII同种型可能由低水平的激肽释放酶活化。在这方面,已在不同的脑区域中发现了血浆前激肽释放酶mRNA和蛋白质(Ashby等人(2012)《衰老神经生物学》33:1345-1355和Neth等人(2001)《血栓形成与止血杂志》85:1043-1047,所述文献中的每个文献出于所有目的通过引用整体并入本文),并且其水平相当于(Cerf和Raidoo(2000)《代谢性脑病(Metab. Brain Dis.)》15:315-323,所述文献出于所有目的通过引用整体并入本文)在本研究中使用的水平。此外,血浆前激肽释放酶表达定位于海马神经元(Cerf和Raidoo(2000)《代谢性脑病》15:315-323,所述文献出于所有目的通过引用整体并入本文),所述海马神经元是在其中发现发现最高水平的短F12 mRNA的细胞,这表明神经元FXII的活化可能在体内发生。另外,我们的结果提出了神经元FXII同种型可能对不会活化全长FXII的可能由神经元或其它脑细胞产生的其它蛋白酶的活化具有增加的敏感性的可能性。

[0416] pro-HGF到HGF的活化是其通过Met受体进行信号传导所必需的。虽然已经确定了HGF-Met信号传导在神经元中的作用(Kato(2017)《生物医学报告》7:495-503以及Tyndall和Walikonis(2006)《细胞周期(Cell Cycle)》5:1560-1568,所述文献中的每个文献出于所有目的通过引用整体并入本文),但pro-HGF如何在脑中活化尚未很好的表征。HGFA是循环中的主要HGF激活子。参见例如Itoh等人(2004)《肠胃病学(Gastroenterology)》127:1423-1435,所述文献出于所有目的通过引用整体并入本文。由于HGFA敲除小鼠没有发育异常(Itoh等人(2004)《肠胃病学》127:1423-1435,所述文献出于所有目的通过引用整体并入本文)而HGF敲除小鼠是胚致死的(Uehara等人(1995)《自然》373:702-705,所述文献出于所有目的通过引用整体并入本文),因此其它HGF激活子可能参与组织中的HGF活化。有趣的是,FXII是HGFA的同系物(Ponczek等人(2008)《血栓形成与止血杂志》6:1876-1883,所述文献出于所有目的通过引用整体并入本文)并且已示出为在体外将pro-HGF转化为活性HGF。参见例如Shimomura等人(1995)《欧洲生物化学杂志》229:257-261,以及Peek等人(2002)《生物化学杂志》277:47804-47809,所述文献中的每个文献出于所有目的通过引用整体并入本文。假设假定神经元FXII可以参与脑中pro-HGF向活性HGF的转化,从而有助于神经元HGF-Met信号传导。

[0417] 神经元Met信号传导与多种生理过程有关,包含神经突生长、树突分枝、长时程增强和记忆。参见例如Kato(2017)《生物医学报告》7:495-503;Wright和Harding(2015)《阿尔茨海默病杂志》45:985-1000;以及Tyndall和Walikonis(2006)《细胞周期》5:1560-1568,所述文献中的每个文献出于所有目的通过引用整体并入本文。观察到F12 mRNA在海马和新皮层锥体神经元中的丰富表达,提出了在这些细胞中可能发生假定神经元FXII活化HGF的可能性。有趣的是,由背侧大脑皮层(包含海马和新皮层神经元)产生的神经元中Met的缺失导

致自发交替和活动减退的缺陷 (Thompson和Levitt (2015)《神经发育障碍杂志 (J.Neurodev.Disord.)》7:35,所述文献出于所有目的通过引用整体并入本文),从而支持HGF-Met信号传导在发现表达F12 mRNA的细胞中的重要性。进一步研究短FXII同种型的功能可以阐明其在这类细胞群中的作用。

[0418] 除了源自体外和动物研究的关于神经元HGF-Met信号传导作用的见解之外,越来越多的证据表明HGF-Met信号传导的调节异常会导致人的神经发育障碍和神经变性。参见例如Eagleson等人(2017)《生物精神病学 (Biol.Psychiatry)》81:424-433;Peng等人(2013)《国际神经生物学综述 (Int.Rev.Neurobiol.)》113:135-165;以及Wright和Harding (2015)《阿尔茨海默病杂志》45:985-1000,所述文献中的每个文献出于所有目的通过引用整体并入本文。特别地, Met蛋白的表达在AD患者的海马神经元中降低 (Hamasaki等人(2014)《神经病理学 (Neuropathology)》34:284-290,所述文献出于所有目的通过引用整体并入本文),而HGF蛋白水平在AD脑中增加 (Fenton等人(1998)《大脑研究》779:262-270,所述文献出于所有目的通过引用整体并入本文)。神经元FXII水平在AD或其它疾病状态中是否发生改变以及神经元FXII是否在疾病过程中发挥作用是开放式问题。未来的研究将阐明神经元FXII表达的细节以及神经元FXII同种型在脑中的HGF和接触途径的生理和病理活化中的作用。

[0419] 材料和方法

[0420] qPCR。来自人肝脏、人全脑、海马体、颞叶和皮层的总RNA来自Clontech公司(加利福尼亚州山景城宝生物工程美国公司(Takara Bio USA;Mountain View,CA))。使用SuperScript®VILO™Master Mix (马萨诸塞州沃尔瑟姆英杰生命技术公司(Invitrogen by Life Technologies;Waltham,MA))将mRNA逆转录为cDNA。将cDNA稀释至0.5-5ng/μL,并使用ABI 7900HT序列检测系统(加利福尼亚州福斯特市应用生物系统公司(Applied Biosystems;Foster City,CA))用SensiFAST Hi-ROX MasterMix(马萨诸塞州陶顿比奥立公司(Bioline;Taunton,MA))扩增2.5-25ng cDNA输入。所使用的针对人F12和GAPDH的引物和探针在表5中示出。数据表示为在对照组织中看到的靶基因表达的% ($2^{-(\Delta\Delta Ct)} * 100$)。

[0421] 人脑组织。对于RNAscope分析,使用了两种人脑组织来源。来自哈佛脑组织资源中心(Harvard Brain Tissue Resource Center) (马萨诸塞州贝尔蒙特)的新鲜冷冻海马脑组织被切成10μm,安装在VWR superfrost载玻片上,并储存在-80℃下。预先安装的5μm福尔马林固定的石蜡包埋的人海马体切片来自Abcam公司(马萨诸塞州剑桥)。

[0422] 血浆收集。对于血浆收集,使用柠檬酸钠灌注针从腔静脉抽取血液。将血液在室温下以1500xg离心10分钟,并收集贫血小板血浆,并且储存在-80℃下。对于组织收集,用0.9%盐水经心脏灌注小鼠。取出脑和肝脏,在干冰上冷冻在最佳切割温度化合物中,并储存在-80℃下。将10μm切片安装在Superfrost plus载玻片(VWR;宾夕法尼亚州拉德诺)上,并储存在-80℃下。

[0423] 原位RNA杂交。使用RNAscope 2.5HD系统RED(Advanced Cell Diagnostics;加利福尼亚州纽瓦克高级细胞诊断公司(Advanced Cell Diagnostics;Newark,CA))按照制造商关于新鲜冷冻或福尔马林固定的石蜡包埋的组织执行的方案执行原位RNA杂交。在单重手动RNAscope测定中使用了探针(ACDbio目录号:313901(PPIB(人))、313911(PPIB(小鼠))、310043(DAPB)、493191(F12外显子1-6(人))、473781(F12外显子11-14(人))、493201(F12外

显子1-6(小鼠))和483971(F12外显子11-14(小鼠)),并且使用Aperio载玻片扫描仪(伊利诺伊州布法罗格罗夫徕卡显微系统公司(Leica Microsystems;Buffalo Grove,IL))用40X物镜捕获来自RNAscope探针的比色信号。为了确认组织适合于RNA分析,使用针对普遍表达的基因(亲环蛋白B;PPIB)的探针作为阳性对照,并使用针对细菌基因(4-羟基-四氢吡啶二羧酸还原酶;DapB)的探针作为阴性对照。由于RNAscope探针的高度特异性,设计了对小鼠和人F12具有特异性的单独的RNAscope探针。

[0424] 免疫荧光/RNAscope。在FXII^{hum/hum}小鼠脑的连续切片上执行免疫荧光和RNAscope。如上所述,使用针对F12外显子11-14的探针运行RNAscope,不同之处在于用DAPI对组织进行复染,并在Zeiss AxioScan Z.1载玻片扫描仪(纽约州索恩伍德卡尔蔡司显微镜公司(Carl Zeiss Microscopy;Thornwood,NY))上使用20X物镜获取图像,其中RNAscope RED的Ex/Em为590/617并且DAPI的为358/461。使用Halo软件(新墨西哥州科拉莱斯印第卡实验室(Indica Labs;Corrales,NM))将RNAscope信号假彩色化为白色。对于免疫荧光,将组织用4%多聚甲醛(宾夕法尼亚州哈特菲尔德电子显微镜科技公司(Electron Microscopy Sciences;Hatfield,PA))的磷酸盐缓冲液进行后固定,并且用含有0.25% Triton-X(密苏里州圣路易斯西格玛奥德里奇公司(Sigma Aldrich;St.Louis,MO))的PBS溶液的3%驴血清(缅因州巴尔港杰克逊实验室公司(Jackson Laboratory;Bar Harbor,ME))封闭。然后将载玻片用在封闭缓冲液中以1:500稀释的NeuN抗体(Abcam公司;目录#: Ab177487)在4℃下温育过夜,然后用在3%驴血清中稀释的Alexa-488偶联的次级抗体(赛默飞世尔公司(ThermoFisher))和DAPI(赛默飞世尔公司)在室温下温育1小时。在PBS中洗涤后,使用Aqua-Mount封固剂(赛默飞世尔公司)将载玻片盖上。在AxioScan Z.1载玻片扫描仪上使用20X物镜获取图像,其中Alexa-488的Ex/Em为490/525并且DAPI的为358/461。

[0425] FXIIa活性测定。为了测量血浆激肽释放酶对FXII₂₉₇₋₅₉₆的活化,在37℃下将重组人FXII₂₉₇₋₅₉₆(伊利诺伊州罗斯蒙特蛋白技术公司(Proteintech;Rosemont,IL))用15nM血浆激肽释放酶(密歇根州诺维分子创新(Molecular Innovations;Novi,MI))在含有140mM NaCl pH7.4(HEPES缓冲盐水;HBS)的20mM HEPES中温育18小时。然后使用100KIU/mL抑肽酶(密苏里州圣路易斯西格玛奥德里奇公司)抑制激肽释放酶活性,并使用浓度为0.8mM的显色底物S-2302(H-D-Pro-Phe-Arg-pNA-2HCl;俄亥俄州威彻斯特迪亚制药公司(Diapharma;West Chester,OH))测量FXIIa活性。在37℃下在平底96孔聚苯乙烯MaxiSorp板(赛默飞世尔公司)中执行测定。使用Spectramax i3X酶标仪(宾夕法尼亚州唐宁敦分子装置公司(Molecular Devices;Downingtown,PA))检测酶活性作为405nm处吸光度的变化。

[0426] 组织蛋白酶K切割的FXII通过在HBS中用组织蛋白酶K(100nM;恩佐生命科学公司(Enzo Life Sciences))温育FXII(1.25μM;分子创新公司)而产生。通过添加非还原性样品缓冲液并在100℃下加热5分钟来停止反应。FXII片段通过SDS-PAGE分离并通过考马斯蓝染色可视化。通过将FXII(375nM)用组织蛋白酶K(30nM)或具有血浆激肽释放酶(15nM)的媒剂温育15分钟来分析组织蛋白酶K切割的FXII的蛋白水解活性。然后用抑肽酶抑制激肽释放酶活性,并用S-2302显色底物测量FXIIa活性。

[0427] 为了分析血浆前激肽释放酶切割,首先将FXII₂₉₇₋₅₉₆(300nM)用血浆激肽释放酶(15nM)或媒剂在37℃下温育18小时。然后使用抑肽酶(100KIU/mL;西格玛公司)灭活激肽释放酶活性。在37℃下将从人血浆中纯化的前激肽释放酶(200nM;分子创新公司)用激肽释放

酶活化的或媒剂活化的FXII₂₉₇₋₅₉₆ (225nM) 或FXII (225nM) 在HBS中温育2小时。通过添加还原样品缓冲液使反应停止。如上所述执行SDS-PAGE,并用抗人前激肽释放酶的抗体 (Abcam公司;目录#:ab1006) 分析印迹。

[0428] 蛋白质印迹。在还原条件下通过蛋白质印迹分析来自FXII^{-/-}、FXII^{+/+}和FXII^{hum/hum}小鼠的血浆。使用4-20% Tris-甘氨酸凝胶 (英杰公司) 执行SDS-PAGE,将凝胶转移到0.2μm PVDF膜 (加利福尼亚州赫拉克勒斯伯乐公司 (BioRad;Hercules,CA)) 上,并用抗小鼠因子XII的抗体 (德克萨斯州凯蒂晕克隆公司 (CloudClone Corp; Katy, TX); 目录#:PAA677Mu01)、人因子XII (Abcam公司;目录#:Ab196670) 和白蛋白 (马萨诸塞州沃尔瑟姆赛默飞世尔皮尔斯公司 (Thermo Scientific Pierce;Waltham,MA); 目录#:PA1-29335) 进行分析。

[0429] 为了分析HGF活化,首先将FXII₂₉₇₋₅₉₆ (300nM) 用血浆激肽释放酶 (15nM) 或媒剂在37℃下温育18小时。然后使用抑肽酶 (100KIU/mL; 西格玛公司) 灭活激肽释放酶活性。将重组人pro-HGF (25μg/mL; 明尼苏达州明尼阿波利斯R&D系统公司 (R&D Systems; Minneapolis,MN)) 用激肽释放酶活化的或媒剂活化的FXII₂₉₇₋₅₉₆ (225nM)、重组人肝细胞生长因子激活子 (225nM; 宾夕法尼亚州韦恩义翘神州生物技术公司 (Sino Biological; Wayne,PA))、双链组织纤溶酶原激活子 (225nM; 密歇根州罗切斯特山牛津生物医学研究公司 (Oxford Biomedical Research;Rochester Hills,MI)) 或βFXIIa (50nM; 分子创新公司) 在37℃下在HBS中温育5小时。通过添加还原样品缓冲液使反应停止。如上所述执行SDS-PAGE,并用抗人HGF的抗体 (R&D系统公司; 目录#:AF-294-SP) 分析印迹。

[0430] 细胞培养。根据制造商的说明培养人iPS谷氨酸能神经元 (iCell®谷氨酸神经元; 威斯康星州麦迪逊细胞动力学公司 (Cellular Dynamics;Madison,WI)), 并通过RNAseq验证分化。细胞示出了两种众所周知的神经元标志物微管相关蛋白2 (MAP2) 和囊泡谷氨酸转运蛋白VGLUT2 (SCL17A6) 的高表达以及GFAP (星形胶质细胞标志物) 和NES (神经干细胞和增殖内皮细胞的标志物) 的低表达 (数据未示出)。来自美国模式组织保藏中心 (ATCC; 弗吉尼亚州马纳萨斯) 的SH-SY5Y人神经母细胞瘤细胞和Neuro-2a小鼠神经母细胞瘤细胞在含有L-谷氨酸、青霉素、链霉素和10%胎牛血清 (赛默飞世尔公司) 的EMEM (加利福尼亚州圣安娜欧文科技公司 (Irvine Scientific;Santa Ana,CA)) 中培育。根据制造商的说明培养来自C57BL/6小鼠 (龙沙公司 (Lonza)) 的原代皮层神经元。

[0431] RNAseq。用于RNA提取和测序文库制备的方案类似于先前描述的方案 (Atanasio等人 (2016)《科学报告 (Sci.Rep.)》6:23204,所述文献出于所有目的通过引用整体并入本文)。使用KAPA链mRNA-Seq试剂盒 (KAPA生物系统公司 (KAPA Biosystems)) 由500ng RNA制备链特异性RNA-seq文库。文库通过十二循环PCR扩增。在Illumina HiSeq®2000 (因美纳公司 (Illumina)) 上通过多重单次读取运行 (针对iPS人神经元、小鼠原代神经元和小鼠组织) 或配对读取运行 (针对SH-SY5Y和Neuro-2A细胞系) 执行测序。通过FastQC分析所得FASTQ文件,以确保足够的数据库质量。使用允许两个错配的商业软件ArrayStudio (OmicSoft公司) 将读段映射到人基因组 (hg19) 或小鼠基因组 (mm10)。在我们的样品中,83%-90%的读段唯一地映射。使用OmicSoft ArrayStudio RNASeq分析管道 (Qiagen公司) 从原始测序读段中得出基因表达。对于每个基因,识别唯一地映射到基因的外显子的读段并计数。所得读段计数在基因水平下汇总为原始表达。

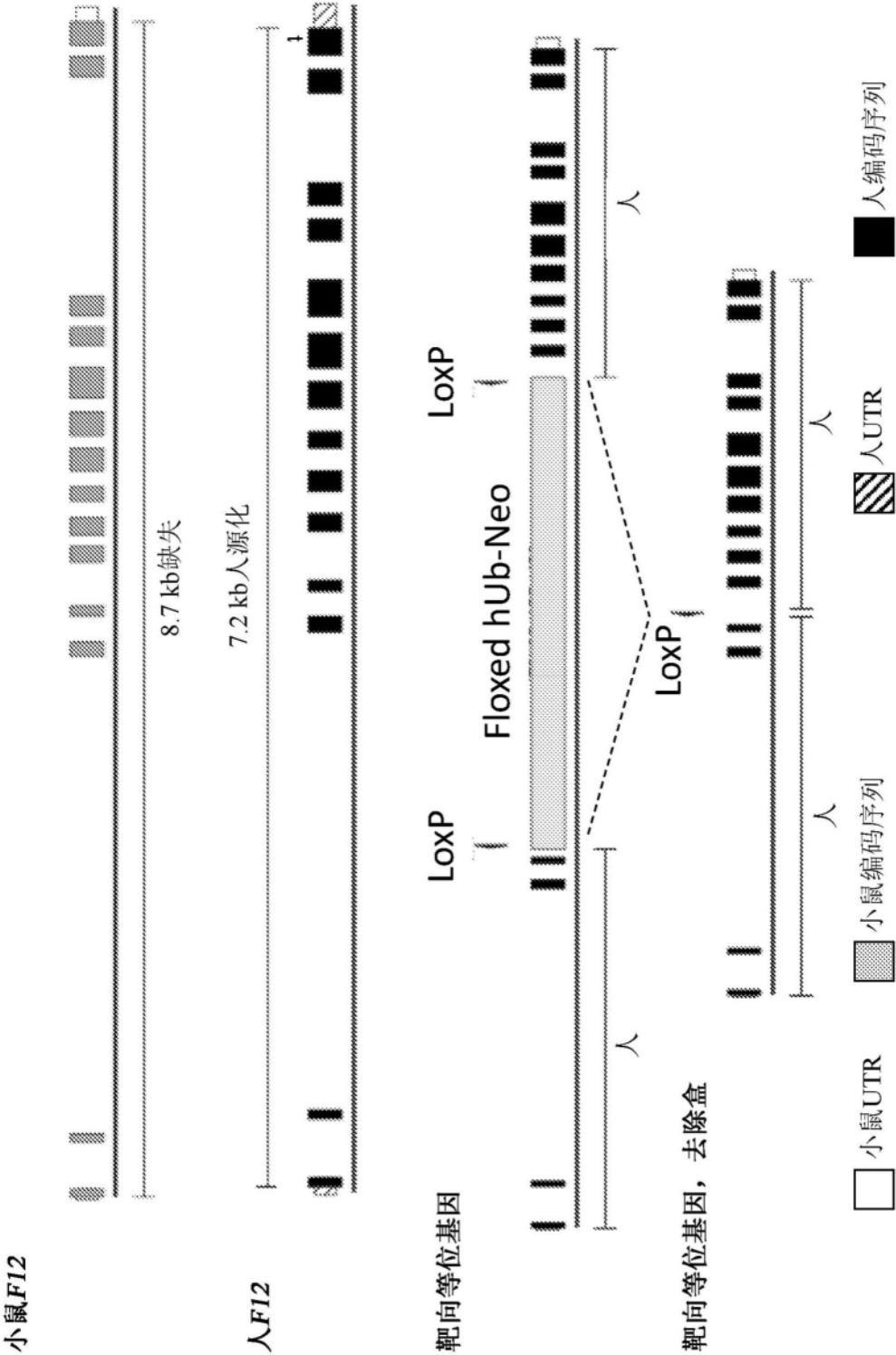


图1

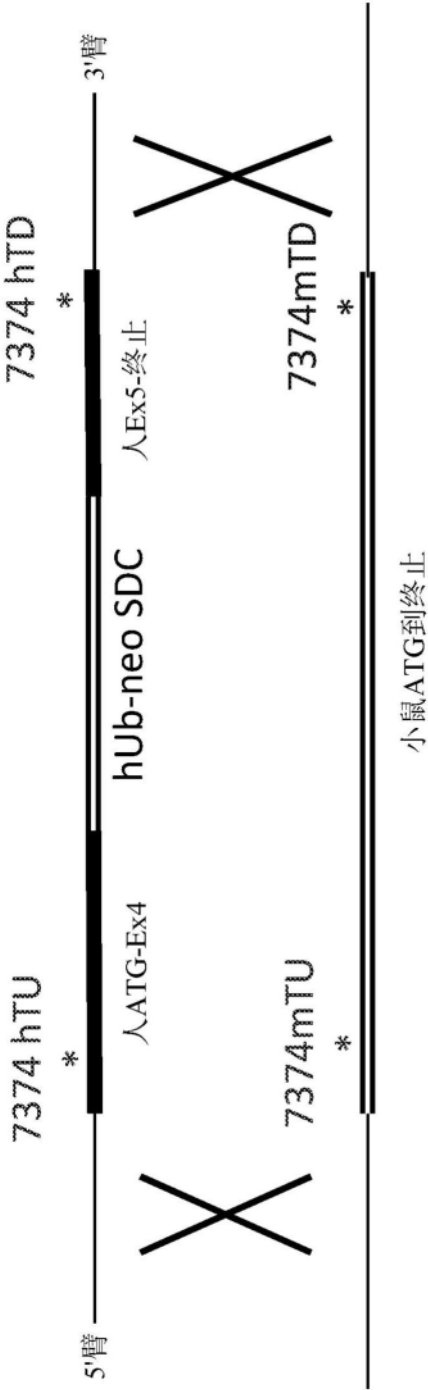


图2

hF12	MRALLLGLLVSLSTLS	IPPEAPKEHKYKAEHTVVLTVTGEPCHFQYHRQL
mF12	MTALLFLGSLMSLDLTLS	APPWKDSKKFKDAPDPTVVLTVDGRLCHFPFQYHRQL
hF12	YHKCTHKGRPGQPWCATTPNFDDQDQRWGYCLEPKKVKDCHSKHSPCQKGGTCVNMP	
mF12	HHKCIHKRRPGSRPWCATTPNFDEDDQWGYCLEPKKVKDCHSKHNPCHKGGTCINTP	
hF12	SGPHCLCPQHLTGNHCQKEKCFEPQLLFFHKNEIWYRTEQAAVARCQCKGPDACHQ	
mF12	NGPHCLCPEHLTGKHCQKEKCFEPQLLKFHENELWFRTGPGGVARCECKGSEAHCK	
hF12	RLASQACRTNPCLHGGRCLVEVEGHRLLCHCPVGTGPFCDVDTKASCYDGRGLSYRGL	
mF12	PVASQAC SINPCNLGGSCLLVEDHPLCRCPTGYTGFCDDLWATCYEGRGLSYRGQ	
hF12	ARTTISGAPCQPWASEATYRNVTAEQARNWGLGGHAFCRNPNDNIRPWCFLNRDRL	
mF12	AGTTQSGAPCQRWTVTEATYRNMTTEKQALSWSGLGHAFCRNPNDNTRPWC FVWSGDRL	
hF12	SWEYCDLAQCQTPTQAAPTTPVSPRLHVPLMPAQAPPKPQPTTRTPPQSQTPGALP	
mF12	SWDYCGLEQCQTPTFAPLVVES	QEEPSQAPSLSHAPNDST
hF12	AKREQPPSLTRNGPLSCGQRLRKSLSSMT	重链 轻链
mF12	DHQTSLSKTNTMCGCQRFKRLSSFMR	VVGGGLVALRGHAHPYIAALYWGHSFCA
hF12	SLIAPCWVLTAAHCLQDRPAPELTVVLGQERRNHSCEPCQTLAVRSYRLHEAFSPV	
mF12	SLIAPCWVLTAAHCLQNRPAPELTVVLGQDRHNQSCWCQTLAVRSYRLHEGFSSI	
hF12	SYQHDLALLRLQEDADGSCALLSPYVQVCLPSGAARPSETTLCQVAGWGHQFEGAE	
mF12	TYQHDLALLRLQESKTNSCAILSPHVQVCLPSGAAPPSETVLCVAGWGHQFEGAE	
hF12	EYASFLQEAQVPFSLERCSAPDVHGSSILPGMLCAGFLEGGTDACQDGGGPLVCE	
mF12	EYSTFLQEAQVPFIALDRCSNSNVHGDAIILPGMLCAGFLEGGTDACQDGGGPLVCE	
hF12	DQAAERRLTLQGIISWGGCGDRNKPVGVTVDVAYYLAWIREHTVS	(SEQ ID NO: 5)
mF12	EGTAEHQTLTRGVISWGGCGDRNKPVGVTVDVANYLAWIQKHAS	(SEQ ID NO: 1)

图3

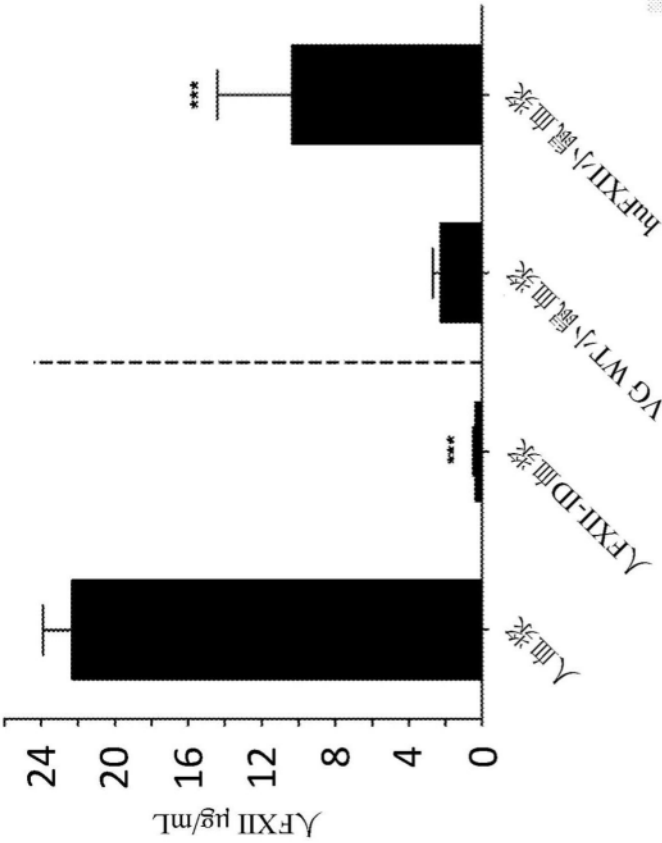


图4

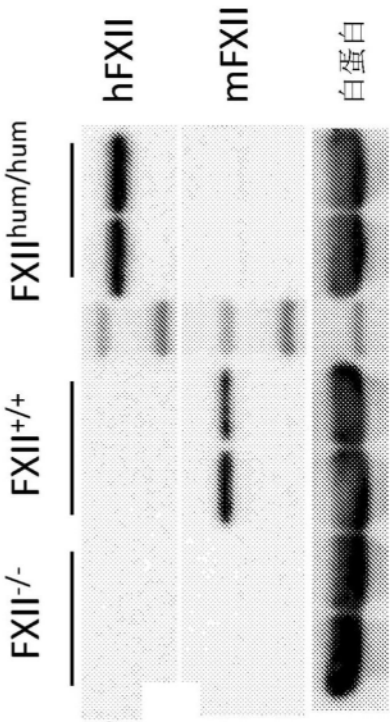


图5

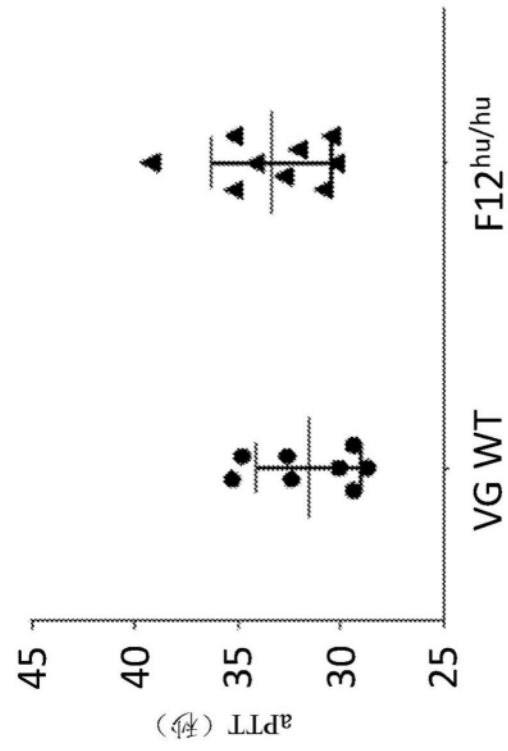


图6

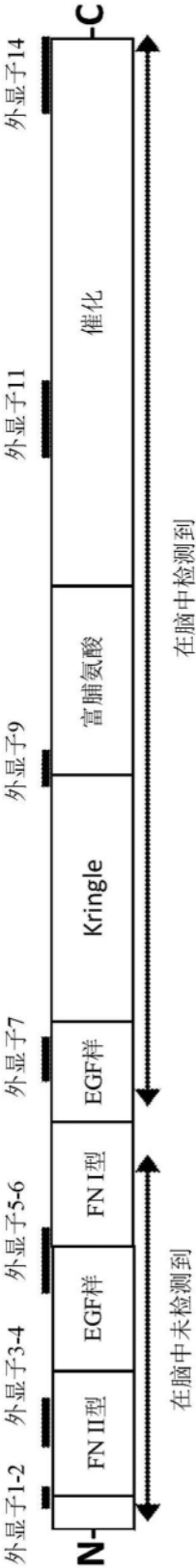


图7A

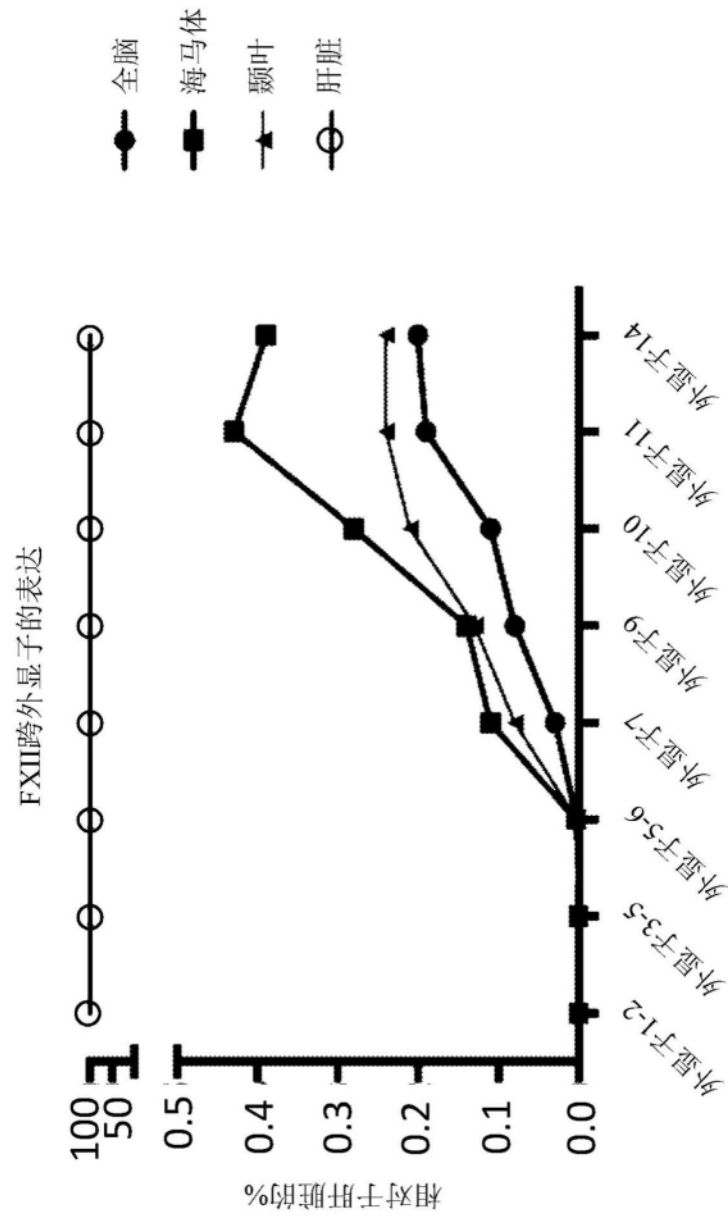


图7B

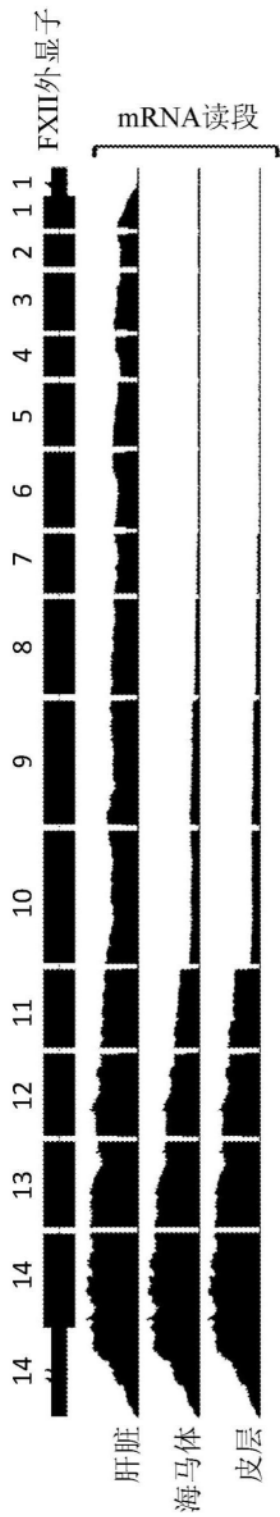


图7C

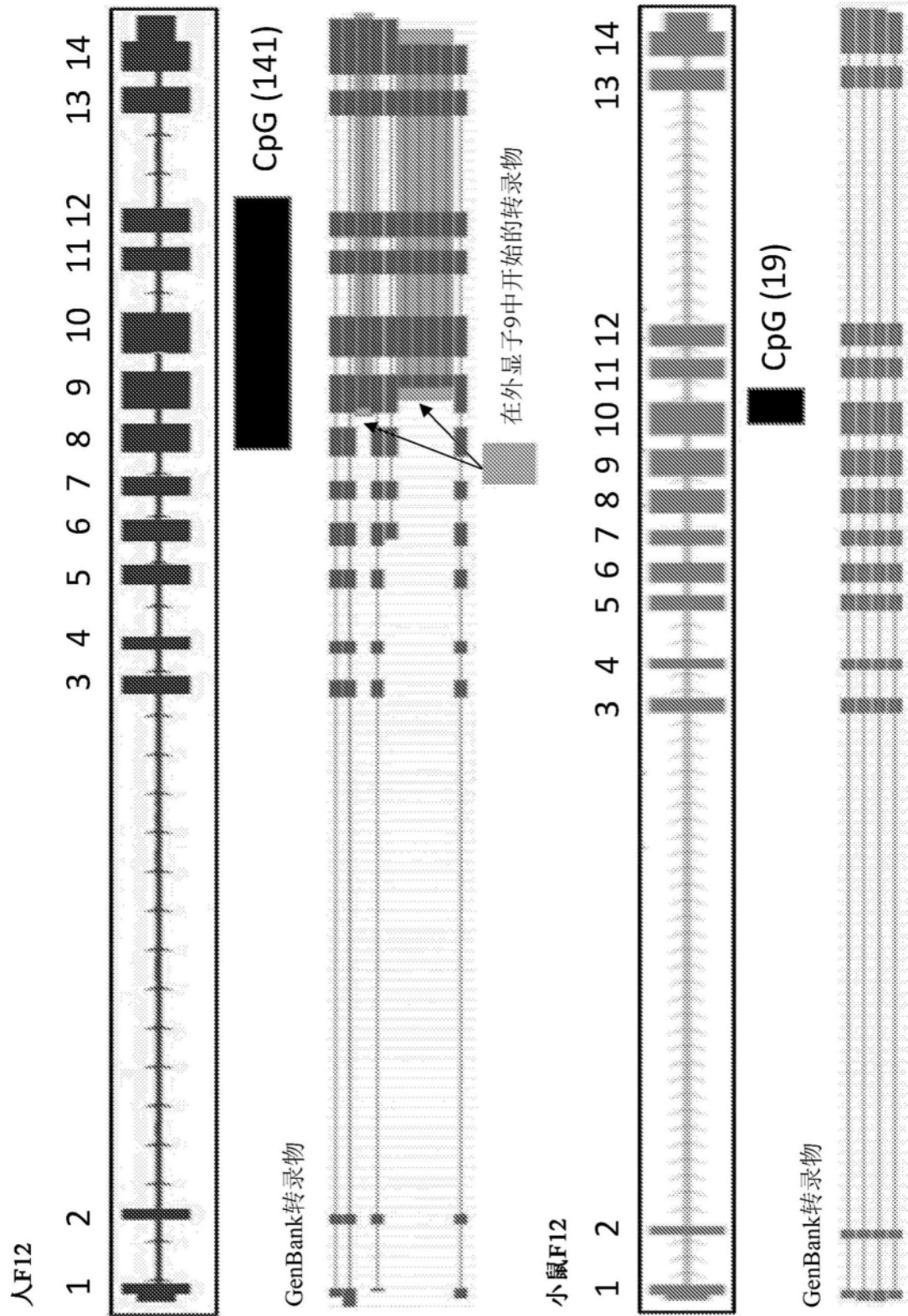


图8A

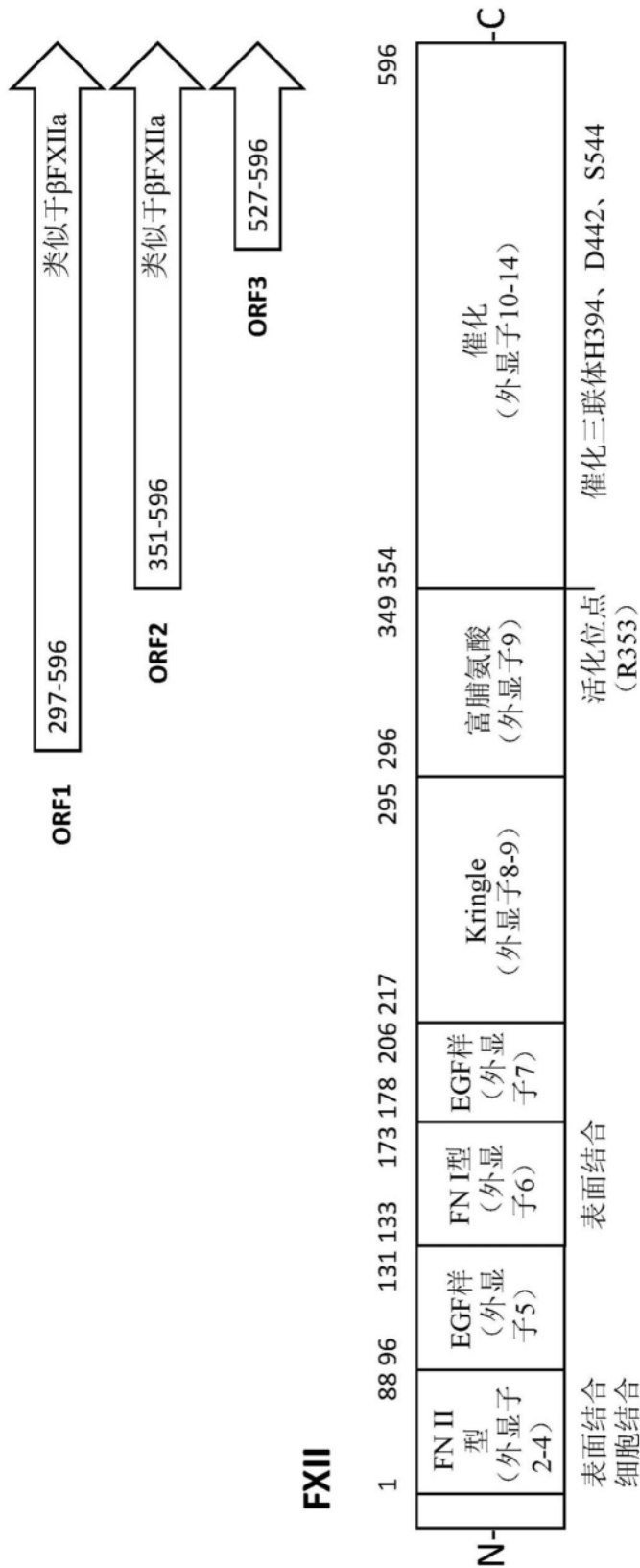


图8B

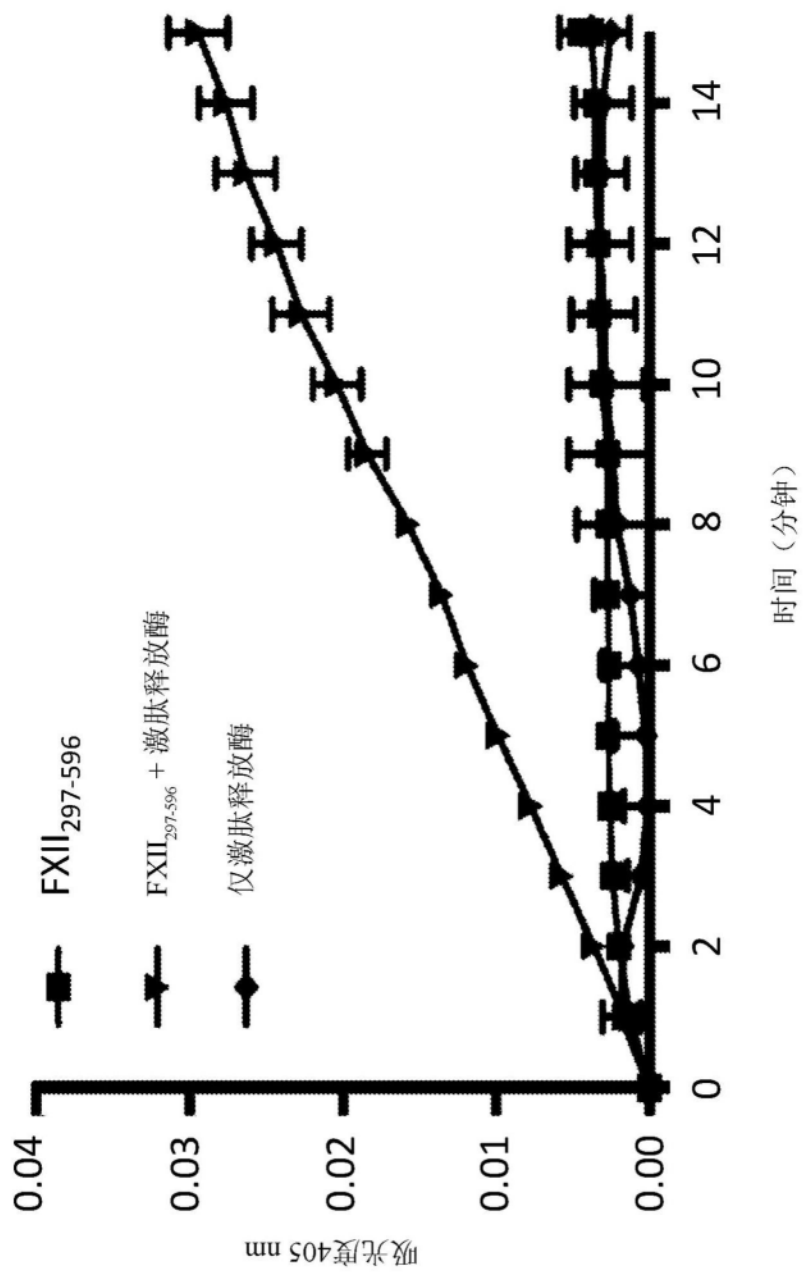


图9A

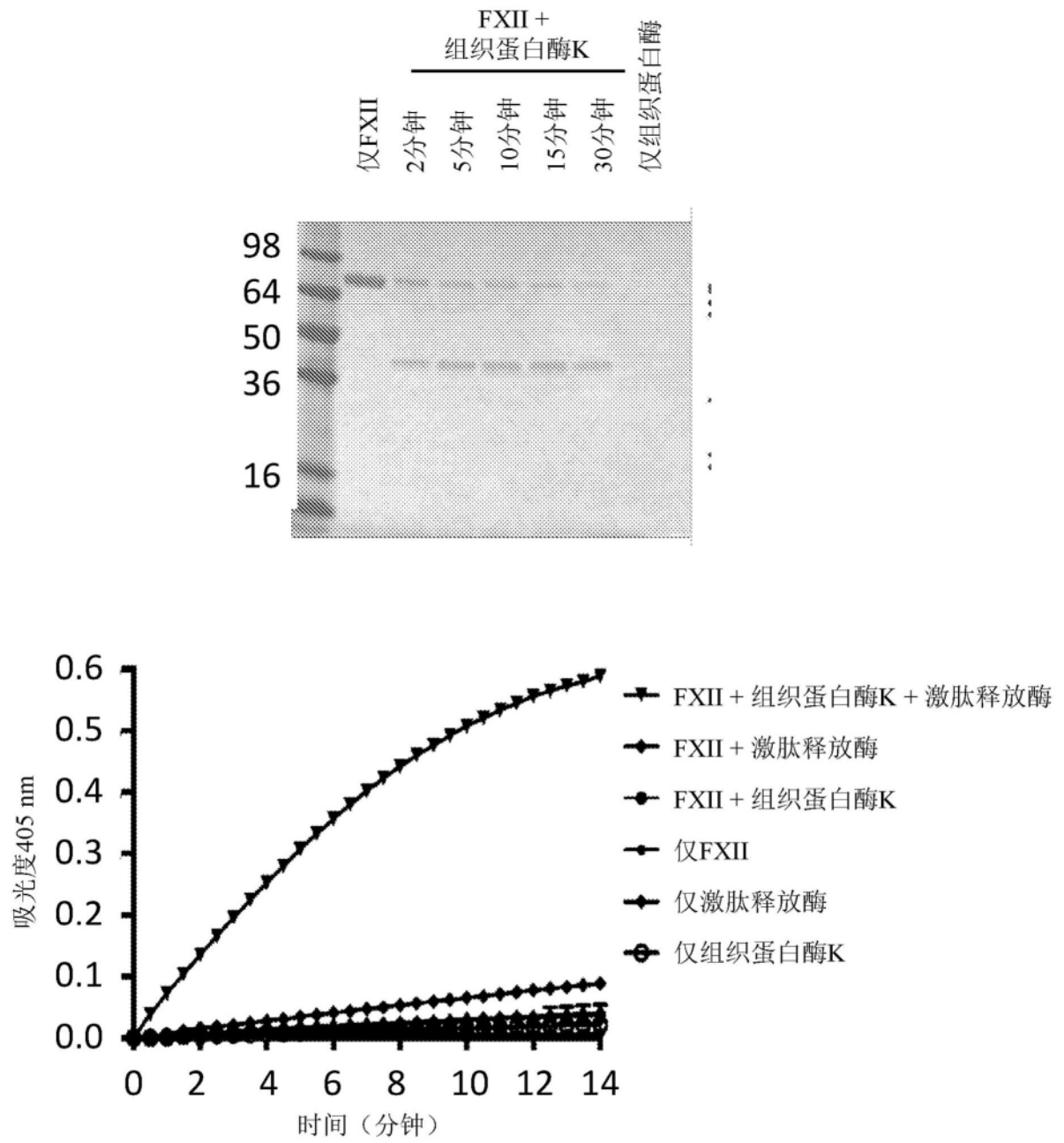


图9B

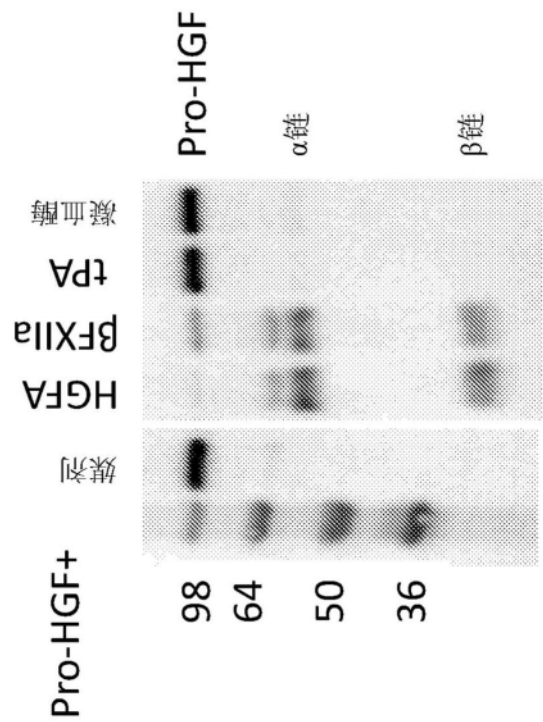


图9C

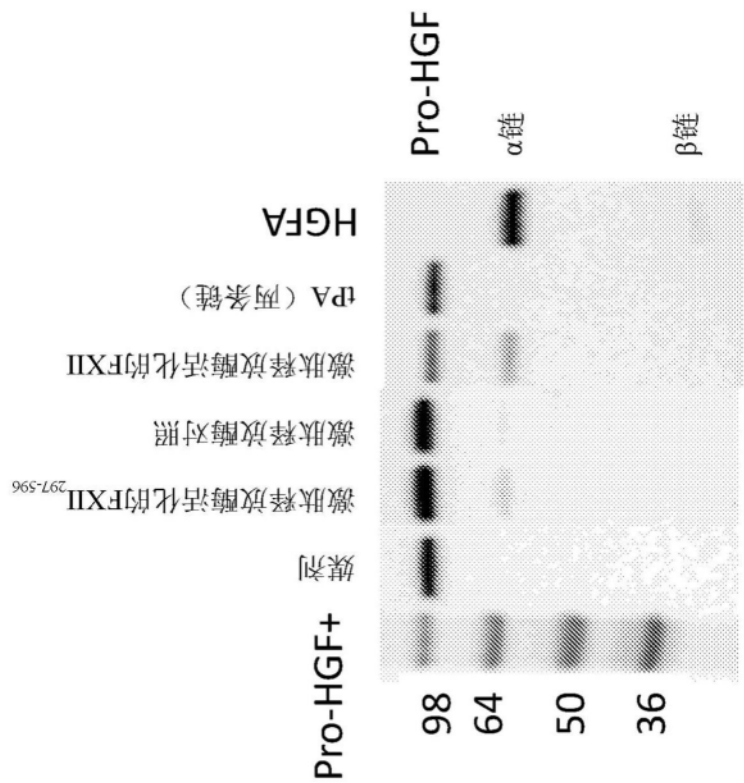


图9D

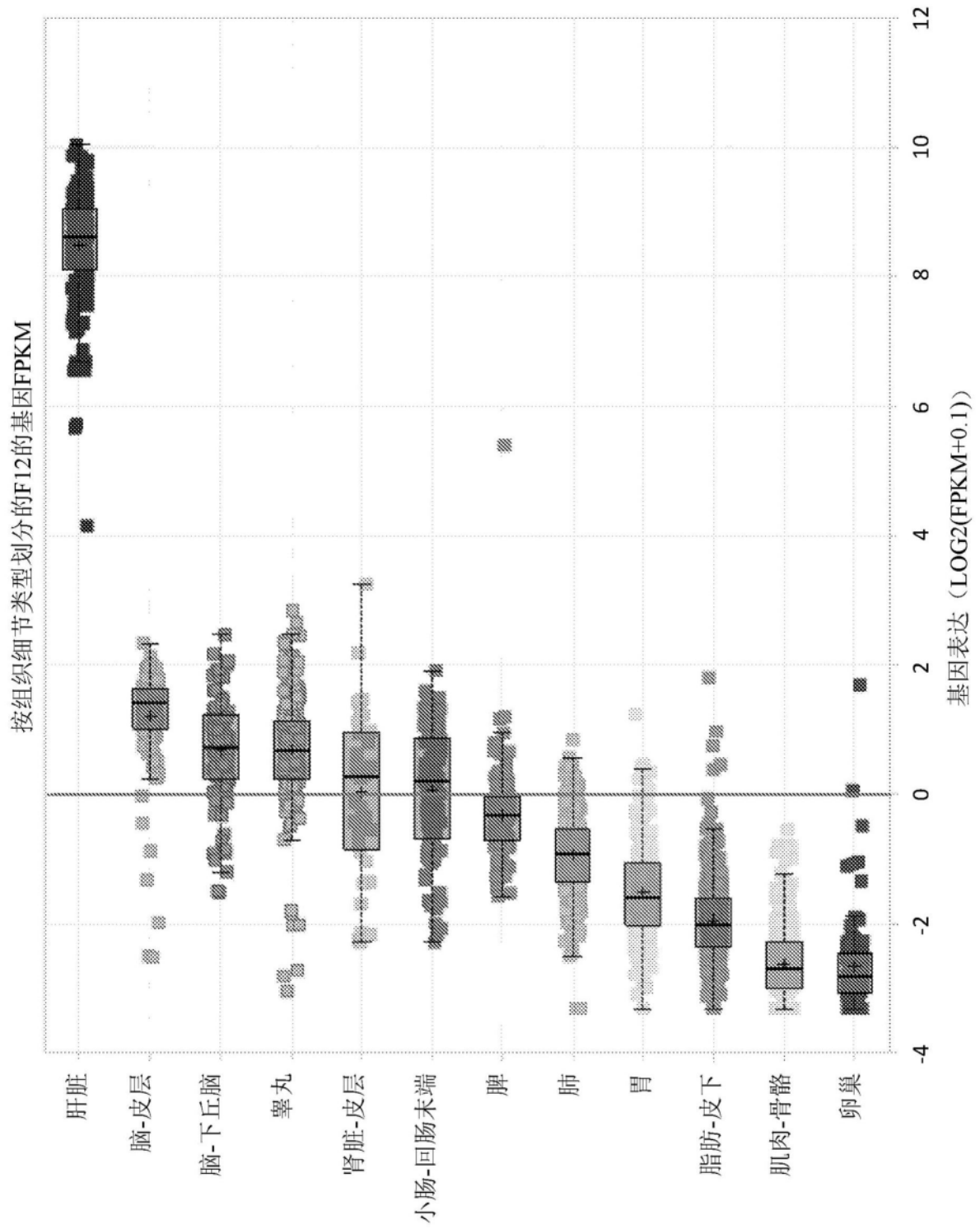


图10A

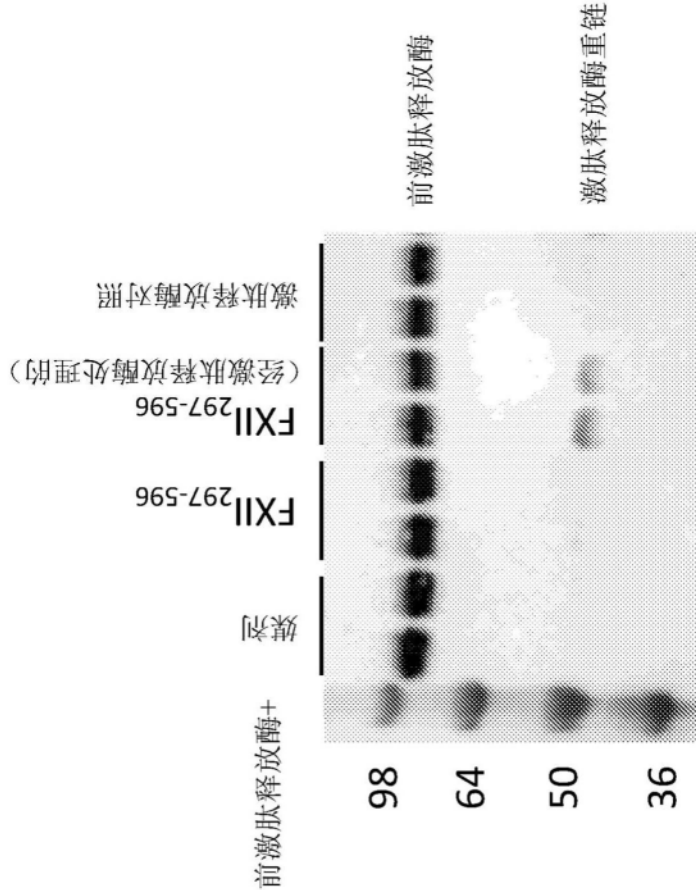


图11