



## (12)发明专利申请

(10)申请公布号 CN 109414450 A

(43)申请公布日 2019.03.01

(21)申请号 201780036135.5

(22)申请日 2017.04.21

(30)优先权数据

62/326,700 2016.04.22 US

(85)PCT国际申请进入国家阶段日

2018.12.11

(86)PCT国际申请的申请数据

PCT/US2017/028981 2017.04.21

(87)PCT国际申请的公布数据

W02017/185054 EN 2017.10.26

(71)申请人 因特利亚治疗公司

地址 美国马萨诸塞州

(72)发明人 S·M·伯恩斯 B·A·穆雷

S·B·赫西

(74)专利代理机构 北京市中咨律师事务所

11247

代理人 凌立 黄革生

(51)Int.Cl.

A61K 31/7105(2006.01)

A61K 38/46(2006.01)

A61K 47/00(2006.01)

A61K 48/00(2006.01)

权利要求书5页 说明书101页 附图2页

(54)发明名称

用于治疗与转录因子4中三核苷酸重复相关的疾病的组合物和方法

(57)摘要

本申请涉及用于切除如见于患有角膜上皮营养不良(FECD)、PSC和精神分裂症的个体中的包含在TCF4的内含子3中的三核苷酸重复(TNR)的组合物和方法。本申请还公开包含靶向胶原蛋白VIII  $\alpha$  2亚基的向导序列的组合物,用于治疗其中可促成FECD的突变。

1. 组合物,其包含至少一个含有指导核酸酶至选自SEQ ID NO:1-1084的靶序列的向导序列的向导RNA。

2. 组合物,其包含至少一个含有与选自SEQ ID NO:1089-1278的序列至少99%、98%、97%、96%、95%、94%、93%、92%、91%或90%同一的向导序列的向导RNA。

3. 组合物,其包含至少一个含有与选自SEQ ID NO:1089-1278的序列同一的向导序列的向导RNA。

4. 权利要求1的组合物,其中向导RNA靶向转录因子4 (TCF4) 基因中三核苷酸重复 (TNR) 处或附近的序列,并指导核酸酶至选自SEQ ID NO:1-190的靶序列。

5. 权利要求4的组合物,其包含至少一个含有与选自SEQ ID NO:1089-1278的序列至少99%、98%、97%、96%、95%、94%、93%、92%、91%或90%同一的向导序列的向导RNA。

6. 组合物,其包含选自以下向导RNA配对的两个向导RNA:

a. 指导核酸酶至SEQ ID NO:83的第一向导RNA和指导核酸酶至SEQ ID NO:109的第二向导RNA;

b. 指导核酸酶至SEQ ID NO:85的第一向导RNA和指导核酸酶至SEQ ID NO:109的第二向导RNA;

c. 指导核酸酶至SEQ ID NO:86的第一向导RNA和指导核酸酶至SEQ ID NO:112的第二向导RNA;

d. 指导核酸酶至SEQ ID NO:85的第一向导RNA和指导核酸酶至SEQ ID NO:112的第二向导RNA;

e. 指导核酸酶至SEQ ID NO:86的第一向导RNA和指导核酸酶至SEQ ID NO:109的第二向导RNA;

f. 指导核酸酶至SEQ ID NO:85的第一向导RNA和指导核酸酶至SEQ ID NO:107的第二向导RNA;

g. 指导核酸酶至SEQ ID NO:83的第一向导RNA和指导核酸酶至SEQ ID NO:125的第二向导RNA;

h. 指导核酸酶至SEQ ID NO:86的第一向导RNA和指导核酸酶至SEQ ID NO:125的第二向导RNA;

i. 指导核酸酶至SEQ ID NO:86的第一向导RNA和指导核酸酶至SEQ ID NO:107的第二向导RNA;

j. 指导核酸酶至SEQ ID NO:64的第一向导RNA和指导核酸酶至SEQ ID NO:106的第二向导RNA;

k. 指导核酸酶至SEQ ID NO:85的第一向导RNA和指导核酸酶至SEQ ID NO:114的第二向导RNA;

l. 指导核酸酶至SEQ ID NO:86的第一向导RNA和指导核酸酶至SEQ ID NO:114的第二向导RNA;

m. 指导核酸酶至SEQ ID NO:83的第一向导RNA和指导核酸酶至SEQ ID NO:114的第二向导RNA;

n. 指导核酸酶至SEQ ID NO:53的第一向导RNA和指导核酸酶至SEQ ID NO:114的第二向导RNA;

o. 指导核酸酶至SEQ ID NO:83的第一向导RNA和指导核酸酶至SEQ ID NO:112的第二向导RNA;和

p. 指导核酸酶至SEQ ID NO:74的第一向导RNA和指导核酸酶至SEQ ID NO:114的第二向导RNA。

7. 权利要求6的组合物,其中指导核酸酶至SEQ ID NO:83的第一向导RNA包含SEQ ID NO:1177,指导核酸酶至SEQ ID NO:109的第二向导RNA包含SEQ ID NO:1197。

8. 权利要求6的组合物,其中指导核酸酶至SEQ ID NO:85的第一向导RNA包含SEQ ID NO:1173,指导核酸酶至SEQ ID NO:109的第二向导RNA包含SEQ ID NO:1197。

9. 权利要求6的组合物,其中指导核酸酶至SEQ ID NO:86的第一向导RNA包含SEQ ID NO:1174,指导核酸酶至SEQ ID NO:112的第二向导RNA包含SEQ ID NO:1200。

10. 权利要求6的组合物,其中指导核酸酶至SEQ ID NO:85的第一向导RNA包含SEQ ID NO:1173,指导核酸酶至SEQ ID NO:112的第二向导RNA包含SEQ ID NO:1200。

11. 权利要求6的组合物,其中指导核酸酶至SEQ ID NO:86的第一向导RNA包含SEQ ID NO:1174,指导核酸酶至SEQ ID NO:109的第二向导RNA包含SEQ ID NO:1197。

12. 权利要求6的组合物,其中指导核酸酶至SEQ ID NO:85的第一向导RNA包含SEQ ID NO:1173,指导核酸酶至SEQ ID NO:107的第二向导RNA包含SEQ ID NO:1195。

13. 权利要求6的组合物,其中指导核酸酶至SEQ ID NO:83的第一向导RNA包含SEQ ID NO:1171,指导核酸酶至SEQ ID NO:125的第二向导RNA包含SEQ ID NO:1213。

14. 权利要求6的组合物,其中指导核酸酶至SEQ ID NO:86的第一向导RNA包含SEQ ID NO:1174,指导核酸酶至SEQ ID NO:125的第二向导RNA包含SEQ ID NO:1213。

15. 权利要求6的组合物,其中指导核酸酶至SEQ ID NO:86的第一向导RNA包含SEQ ID NO:1174,指导核酸酶至SEQ ID NO:107的第二向导RNA包含SEQ ID NO:1195。

16. 权利要求6的组合物,其中指导核酸酶至SEQ ID NO:64的第一向导RNA包含SEQ ID NO:1152,指导核酸酶至SEQ ID NO:106的第二向导RNA包含SEQ ID NO:1194。

17. 权利要求6的组合物,其中指导核酸酶至SEQ ID NO:85的第一向导RNA包含SEQ ID NO:1173,指导核酸酶至SEQ ID NO:114的第二向导RNA包含SEQ ID NO:1202。

18. 权利要求6的组合物,其中指导核酸酶至SEQ ID NO:86的第一向导RNA包含SEQ ID NO:1174,指导核酸酶至SEQ ID NO:114的第二向导RNA包含SEQ ID NO:1202。

19. 权利要求6的组合物,其中指导核酸酶至SEQ ID NO:83的第一向导RNA包含SEQ ID NO:1171,指导核酸酶至SEQ ID NO:114的第二向导RNA包含SEQ ID NO:1202。

20. 权利要求6的组合物,其中指导核酸酶至SEQ ID NO:53的第一向导RNA包含SEQ ID NO:1141,指导核酸酶至SEQ ID NO:114的第二向导RNA包含SEQ ID NO:1202。

21. 权利要求6的组合物,其中指导核酸酶至SEQ ID NO:83的第一向导RNA包含SEQ ID NO:1171,指导核酸酶至SEQ ID NO:112的第二向导RNA包含SEQ ID NO:1200。

22. 权利要求6的组合物,其中指导核酸酶至SEQ ID NO:74的第一向导RNA包含SEQ ID NO:1162,指导核酸酶至SEQ ID NO:114的第二向导RNA包含SEQ ID NO:1202。

23. 权利要求1的组合物,其中向导RNA靶向VIII型胶原 $\alpha 2$ 亚基基因(Co18A2),并指导核酸酶至选自SEQ ID NO:191-1063的靶序列。

24. 权利要求23的组合物,其包含至少一个向导RNA,该向导RNA包含与选自SEQ ID NO:

191-1063的靶序列的前20个核苷酸互补或同一的序列至少99%、98%、97%、96%、95%、94%、93%、92%、91%或90%同一的向导序列,其中用尿嘧啶替换SEQ ID NO:191-1063的前20个核苷酸中的胸腺嘧啶。

25. 权利要求1的组合物,其中向导RNA靶向Co18A2基因产物中的Gln455Lys突变,并指导核酸酶至选自SEQ ID NO:1064-1069的靶序列。

26. 权利要求25的组合物,其包含至少一个向导RNA,该向导RNA包含与选自SEQ ID NO:1064-1069的靶序列的前20个核苷酸互补或同一的序列至少99%、98%、97%、96%、95%、94%、93%、92%、91%或90%同一的向导序列,其中用尿嘧啶替换SEQ ID NO:1064-1069的前20个核苷酸中的胸腺嘧啶。

27. 权利要求1的组合物,其中向导RNA靶向Co18A2基因产物中的Gln455Val突变,并指导核酸酶至选自SEQ ID NO:1070-1075的靶序列。

28. 权利要求27的组合物,其包含至少一个向导RNA,该向导RNA包含与选自SEQ ID NO:1070-1075的靶序列的前20个核苷酸互补或同一的序列至少99%、98%、97%、96%、95%、94%、93%、92%、91%或90%同一的向导序列,其中用尿嘧啶替换SEQ ID NO:1070-1075的前20个核苷酸中的胸腺嘧啶。

29. 权利要求1的组合物,其中向导RNA靶向Co18A2基因产物中的Leu450Trp突变,并指导核酸酶至选自SEQ ID NO:1076-1084的靶序列。

30. 权利要求29的组合物,其包含至少一个向导RNA,该向导RNA包含与选自SEQ ID NO:1076-1084的靶序列的前20个核苷酸互补或同一的序列至少99%、98%、97%、96%、95%、94%、93%、92%、91%或90%同一的向导序列,其中用尿嘧啶替换SEQ ID NO:1076-1084的前20个核苷酸中的胸腺嘧啶。

31. 权利要求1-30中任一项的组合物,其中向导RNA是双向导。

32. 权利要求1-30中任一项的组合物,其中向导RNA是单向导。

33. 权利要求1-32中任一项的组合物,其中至少一个向导RNA包含crRNA、trRNA或crRNA和trRNA。

34. 权利要求1-33中任一项的组合物,其中至少一个向导序列编码在载体上。

35. 权利要求34的组合物,其中载体包含第一向导序列和第二向导序列。

36. 权利要求1-33中任一项的组合物,其中第一向导序列和第二向导序列编码在不同载体上。

37. 权利要求34或35的组合物,其中第一向导序列和第二向导序列由相同的启动子和/或调节序列控制。

38. 权利要求1-37中任一项的组合物,其中向导序列与靶基因正链中的靶序列互补。

39. 权利要求1-37中任一项的组合物,其中向导序列与靶基因负链中的靶序列互补。

40. 权利要求1-39中任一项的组合物,其中第一向导序列和第二向导序列与靶基因相反链中的第一靶序列和第二靶序列互补。

41. 权利要求1-39中任一项的组合物,其中向导RNA经化学修饰。

42. 权利要求1-41中任一项的组合物,其进一步包含核酸酶。

43. 权利要求42的组合物,其中核酸酶是Cas蛋白质。

44. 权利要求43的组合物,其中Cas蛋白质来自I型、II型或III型CRISPR/Cas系统。

45. 权利要求43的组合物,其中Cas蛋白质是Cas9。
46. 权利要求43的组合物,其中Cas蛋白质是Cpf1。
47. 权利要求42的组合物,其中核酸酶是切口酶。
48. 权利要求42的组合物,其中核酸酶经修饰。
49. 权利要求48的组合物,其中修饰核酸酶包含核定位信号(NLS)。
50. 药物制剂,其包含权利要求1至49中任一项的组合物和可药用载体。
51. 在人个体中切除转录因子4 (TCF4) 基因中的至少部分三核苷酸重复 (TNR) 的方法,其包括施用权利要求1-49中任一项的组合物或权利要求50的药物制剂。
52. 权利要求51的方法,其中使用两个向导RNA,其中第一向导RNA指导核酸酶至TNR的5' 序列,第二向导RNA指导核酸酶至TNR的3' 序列。
53. 权利要求51的方法,其中人个体患有角膜上皮营养不良 (FECD)。
54. 权利要求53的方法,其中个体具有FECD家族史。
55. 权利要求51-54中任一项的方法,其中个体由于施用而具有视敏度的改善、稳定化或下降减慢。
56. 权利要求51-54中任一项的方法,其中个体由于施用而具有通过角膜厚度测量法测量的改善、稳定化或变化减慢。
57. 权利要求51-54中任一项的方法,其中个体由于施用而具有基于角膜内皮显微术的改善、稳定化或变化减慢。
58. 权利要求51-54中任一项的方法,其中个体由于施用而延迟需要角膜移植的时间。
59. 权利要求51-58中任一项的方法,其中TNR等于或大于约40个三核苷酸重复。
60. 权利要求51-59中任一项的方法,其中切除整个TNR。
61. 权利要求51-60中任一项的方法,其中组合物或药物制剂经病毒载体施用。
62. 权利要求51-60中任一项的方法,其中组合物或药物制剂经脂质纳米颗粒施用。
63. 权利要求51-62中任一项的方法,其进一步包括滴眼液或眼膏的共同施用。
64. 权利要求51-63中任一项的方法,其进一步包括软性隐形眼镜的使用。
65. 权利要求51的方法,其中人个体患有精神分裂症。
66. 权利要求51的方法,其中人个体患有原发性硬化性胆管炎 (PSC)。
67. 在人个体中减少COL8A2基因的突变等位基因 (如Gln455Lys、Gln455Val或Leu450Trp) 的表达或改变核苷酸序列来矫正所述突变等位基因的方法,其包括施用权利要求1-50中任一项的组合物或权利要求51的药物制剂。
68. 权利要求67的方法,其中人个体患有角膜上皮营养不良 (FECD) 或多形性角膜后层营养不良 (PPCD)。
69. 权利要求68的方法,其中个体具有FECD家族史。
70. 权利要求67-69中任一项的方法,其中个体由于施用而具有视敏度的改善、稳定化或下降减慢。
71. 权利要求67-70中任一项的方法,其中个体由于施用而具有通过角膜厚度测量法测量的改善、稳定化或变化减慢。
72. 权利要求67-71中任一项的方法,其中个体由于施用而具有基于角膜内皮显微术的改善、稳定化或变化减慢。

73. 权利要求67-72中任一项的方法,其中个体由于施用而延迟需要角膜移植的时间。

74. 权利要求67-73中任一项的方法,其中导致表达Gln455Lys、Gln455Val或Leu450Trp基因产物的突变分别是c.1364C>A、c.1363-1364CA>GT或c.1349T>G。

75. 权利要求67-74中任一项的方法,其中组合物或药物制剂经病毒载体施用。

76. 权利要求67-74中任一项的方法,其中组合物或药物制剂经脂质纳米颗粒施用。

77. 权利要求67-76中任一项的方法,其进一步包括滴眼液或眼膏的共同施用。

78. 权利要求67-77中任一项的方法,其进一步包括软性隐形眼镜的使用。

79. 权利要求1至50中任一项的组合物或权利要求51的药物制剂的用途,用于制备用于治疗TCF4基因中具有TNR扩增、或COL8A2基因中具有导致基因产物具有Gln455Lys、Gln455Val或Leu450Trp突变的突变的人个体的药物。

## 用于治疗与转录因子4中三核苷酸重复相关的疾病的组合物 和方法

### 技术领域

[0001] 本申请涉及用于治疗与转录因子4 (TNF4) 基因中的三核苷酸重复相关的疾病的组合物和方法, 该疾病包括角膜上皮营养不良、多形性角膜后层营养不良 (PPCD)、原发性硬化性胆管炎 (PSC) 和精神分裂症。

[0002] 角膜上皮营养不良 (FECD) (也称为Fuchs营养不良) 是影响角膜内部内皮细胞单层的变性疾病。角膜内皮的作用是通过维持内皮屏障和执行泵功能来确保角膜透明。在FECD中, 角膜内皮中存在灶性增生 (称为水滴 (guttae)) 和异常胶原的累积。认为散布在角膜内皮和基质细胞间的水滴的存在是疾病的临床标志。晚期FECD表征为大范围水滴、内皮细胞丧失和间质性水肿。

[0003] FECD可以导致视力丧失, 晚期FECD仅可用角膜移植治疗。据估计, 美国约5%的中年白人患有FECD。此外, 据估计, FECD每年导致超过14,000例角膜移植。与角膜移植相关的风险包括急性排斥、慢性排斥、移植物黏附宿主床失败、感染和宿主眼睛损伤。大多数移植给受体带来不到20/20的视力, 需要长达六个月的恢复期, 需要患者在手术后使用免疫抑制剂滴眼液两年或更长时间。长期使用免疫抑制剂滴眼液可提高白内障或青光眼的风险。

[0004] 已报道了遗传因素在FECD中的作用, 包括转录因子4 (TCF4) 基因中的单核苷酸多态性和三核苷酸重复 (TNR) 扩增。TCF4基因第三内含子中的TNR占对疾病的可遗传易感性的大部分, 高于50个重复的重复长度与FECD的临床诊断相关 (Wieben等, PLOS One, 7:11, e49083 (2012))。最近的研究表明, 此TNR扩增导致受影响的TCF4RNA的聚集和关键RNA剪接因子的螯合 (sequestration) (Mootha等, Invest Ophthalmol VisSci. 55 (1):33-42 (2014); Mootha等, Invest Ophthalmol Vis Sci. 56 (3):2003-11 (2015); Vasanth等, Invest Ophthalmol Vis Sci. 56 (8):4531-6 (2015); Soliman等, JAMA Ophthalmol. 133 (12):1386-91 (2015))。这种螯合可以导致基因表达的整体改变, 包括细胞功能的深远改变, 这最终导致细胞死亡 (Du等, JBiological Chem. 290:10, 5979-5990 (2015))。TCF4突变还与原发性硬化性胆管炎 (PSC) 和精神分裂症相关, 参见Ellinghas等, HEPATOLOGY, 58:3, 1074-1083 (2013) 和Forrest等, Trends in Molecular Medicine 20:6 (2014)。

[0005] 在其他重复扩增疾病中, 已提出了RNA毒性。在RNA毒性的情况下, 扩增微卫星DNA序列可见于多种基因的非编码区, 重复元件转录为毒性功能获得RNA或毒性蛋白质种类 (参见Mohan等, Brain Res. 1584, 3-14 (2014))。最近, RNA毒性也在FECD患者中得到显示 (参见Du 2015)。此外, 已提出TCF4 TNR转录物主要累积在角膜内皮中, 从而产生FECD的病理特征。虽然RNA毒性的作用有助于描绘FECD中的潜在疾病机制, 但治疗仍限于角膜移植。

[0006] 其他形式的早发型FECD与COL8A2中的突变相关 (参见Vedana等, Clinical Ophthalmology 10321-330 (2016))。正常情况下, 胶原VIII或COL8 (包含COL8A1和COL8A2) 有规律地分布在角膜的后弹性膜中。但是, 来自COL8A2中有突变的个体的角膜具有以不协调的方式无规律镶嵌沉积的不同量的COL8A1和COL8A2。COL8A2中的三个突变 (Gln455Lys、Gln455Val和Leu450Trp) 导致突变胶原VIII肽的包内累积, 并可引起早发型FECD, 以及相关

障碍多形性角膜后层营养不良 (PPCD)。PPCD表征为角膜后弹性膜和内皮层中的变化。最常与COL8A2中的突变相关的PPCD形式是PPCD2。

[0007] 需要直接调节TCF4中的 (CTG)<sub>n</sub>TNR和COL8A2中的点突变的手段来治疗导致FECD、PPCD、PSC和精神分裂症的遗传突变。最近研究的基因编辑/破坏技术基于细菌CRISPR (规律成簇间隔短回文重复) 系统。CRISPR基因编辑依赖于可在DNA中诱导位点特异性断裂的单核苷酸酶, 如“CRISPR结合蛋白9” (Cas9) 和Cpf1所包含的单核苷酸酶。Cas内切核苷酸酶由小RNA分子 (称为trRNA和crRNA) 连同邻近靶基因的原间隔区 (protospacer) 邻近基序 (PAM) 引导至特异性DNA序列。trRNA和crRNA一起形成向导RNA, 也称为gRNA。trRNA和crRNA可以组合为单向导RNA (sgRNA) 以便于Cas蛋白质的靶向, 或者可以作为双向导 (dgRNA) 系统同时使用但不组合。与trRNA和crRNA组合的Cas内切核苷酸酶称为Cas核糖核蛋白 (RNP) 复合物。

[0008] 发明概述

[0009] 我们在本文中描述CRISPR组合物及其使用方法, 在一些实施方案中, 其设计用于切除TCF4内包含TNR扩增的一些或全部区域。在一些实施方案中, 这些TNR扩增见于患有FECD的个体中。这样做防止了与扩增相关的毒性。TCF4内TNR的减少或消除将减少TNR的下游影响, 如RNA毒性, 并改善病程。因此, 与TCF4内含子3TNR侧翼靶序列互补的向导RNA和核苷酸酶 (或Cas RNP) 的其他修饰可以是治疗在TCF4中显示TNR的FECD的遗传形式以及PSC和精神分裂症中的TNR的手段。此外, 公开用于设计与核苷酸酶一起敲除或编辑在胶原VIII $\alpha$ 亚基中显示突变的FECD和PPCD形式中的COL8A2的向导RNA的向导序列。

[0010] 根据该描述, 在一些实施方案中, 描述指导CRISPR/Cas内切核苷酸酶至TCF4基因中TNR扩增的5' 和3' 区域的向导RNA的组合物。该组合物用于从TCF4基因切除TNR扩增, 以及治疗FECD、PPCD、PSC和精神分裂症。在其他实施方案中, 描述靶向至COL8A2基因的区域向导RNA的组合物, 包括靶向至与FECD相关的突变等位基因的向导RNA。取决于向导RNA的设计和用于治疗的方法, 这些向导RNA将与CRISPR核苷酸酶一起用于切除TNR, 产生插入/缺失, 或通过经双链断裂进行的同源重组 (HR) 或同源性指导修复 (HDR) 矫正基因。

[0011] 在一个实施方案中, 本发明包含含有至少一个向导RNA的组合物, 该向导RNA包含选自SEQ ID NO:1-1084的指导核苷酸酶至靶序列的向导序列。在一些实施方案中, 本发明包含含有至少一个向导RNA的组合物, 该向导RNA包含与选自SEQ ID NO:1089-1278的序列至少99%、98%、97%、96%、95%、94%、93%、92%、91%或90%同一的向导序列。

[0012] 在一些实施方案中, 提供包含至少一个向导RNA的组合物, 该向导RNA包含与选自SEQ ID NO:1089-1278的序列同一的向导序列。

[0013] 在一些实施方案中, 该向导RNA靶向转录因子4 (TCF4) 基因中的三核苷酸重复 (TNR), 并指导核苷酸酶至选自SEQ ID NO:1-190的靶序列。在一些实施方案中, 本发明包含至少一个向导RNA, 该向导RNA包含与选自SEQ ID NO:1089-1278的序列至少99%、98%、97%、96%、95%、94%、93%、92%、91%或90%同一的向导序列。

[0014] 提供包含选自以下向导RNA配对的两个向导RNA的组合物:

[0015] a. 指导核苷酸酶至SEQ ID NO:83的第一向导RNA和指导核苷酸酶至SEQ ID NO:109的第二向导RNA;

[0016] b. 指导核苷酸酶至SEQ ID NO:85的第一向导RNA和指导核苷酸酶至SEQ ID NO:109的第二向导RNA;

- [0017] c. 指导核酸酶至SEQ ID NO:86的第一向导RNA和指导核酸酶至SEQ ID NO:112的第二向导RNA;
- [0018] d. 指导核酸酶至SEQ ID NO:85的第一向导RNA和指导核酸酶至SEQ ID NO:112的第二向导RNA;
- [0019] e. 指导核酸酶至SEQ ID NO:86的第一向导RNA和指导核酸酶至SEQ ID NO:109的第二向导RNA;
- [0020] f. 指导核酸酶至SEQ ID NO:85的第一向导RNA和指导核酸酶至SEQ ID NO:107的第二向导RNA;
- [0021] g. 指导核酸酶至SEQ ID NO:83的第一向导RNA和指导核酸酶至SEQ ID NO:125的第二向导RNA;
- [0022] h. 指导核酸酶至SEQ ID NO:86的第一向导RNA和指导核酸酶至SEQ ID NO:125的第二向导RNA;
- [0023] i. 指导核酸酶至SEQ ID NO:86的第一向导RNA和指导核酸酶至SEQ ID NO:107的第二向导RNA;
- [0024] j. 指导核酸酶至SEQ ID NO:64的第一向导RNA和指导核酸酶至SEQ ID NO:106的第二向导RNA;
- [0025] k. 指导核酸酶至SEQ ID NO:85的第一向导RNA和指导核酸酶至SEQ ID NO:114的第二向导RNA;
- [0026] l. 指导核酸酶至SEQ ID NO:86的第一向导RNA和指导核酸酶至SEQ ID NO:114的第二向导RNA;
- [0027] m. 指导核酸酶至SEQ ID NO:83的第一向导RNA和指导核酸酶至SEQ ID NO:114的第二向导RNA;
- [0028] n. 指导核酸酶至SEQ ID NO:53的第一向导RNA和指导核酸酶至SEQ ID NO:114的第二向导RNA;
- [0029] o. 指导核酸酶至SEQ ID NO:83的第一向导RNA和指导核酸酶至SEQ ID NO:112的第二向导RNA; 和
- [0030] p. 指导核酸酶至SEQ ID NO:74的第一向导RNA和指导核酸酶至SEQ ID NO:114的第二向导RNA。
- [0031] 在一些包含两个gRNA的实施方案中, 指导核酸酶至SEQ ID NO:83的第一向导RNA包含SEQ ID NO:1177, 指导核酸酶至SEQ ID NO:109的第二向导RNA包含SEQ ID NO:1197。
- [0032] 在一些包含两个gRNA的实施方案中, 指导核酸酶至SEQ ID NO:85的第一向导RNA包含SEQ ID NO:1173, 指导核酸酶至SEQ ID NO:109的第二向导RNA包含SEQ ID NO:1197。
- [0033] 在一些包含两个gRNA的实施方案中, 指导核酸酶至SEQ ID NO:86的第一向导RNA包含SEQ ID NO:1174, 指导核酸酶至SEQ ID NO:112的第二向导RNA包含SEQ ID NO:1200。
- [0034] 在一些包含两个gRNA的实施方案中, 指导核酸酶至SEQ ID NO:85的第一向导RNA包含SEQ ID NO:1173, 指导核酸酶至SEQ ID NO:112的第二向导RNA包含SEQ ID NO:1200。
- [0035] 在一些包含两个gRNA的实施方案中, 指导核酸酶至SEQ ID NO:86的第一向导RNA包含SEQ ID NO:1174, 指导核酸酶至SEQ ID NO:109的第二向导RNA包含SEQ ID NO:1197。
- [0036] 在一些包含两个gRNA的实施方案中, 指导核酸酶至SEQ ID NO:85的第一向导RNA

包含SEQ ID NO:1173,指导核酸酶至SEQ ID NO:107的第二向导RNA包含SEQ ID NO:1195。

[0037] 在一些包含两个gRNA的实施方案中,指导核酸酶至SEQ ID NO:83的第一向导RNA包含SEQ ID NO:1171,指导核酸酶至SEQ ID NO:125的第二向导RNA包含SEQ ID NO:1213。

[0038] 在一些包含两个gRNA的实施方案中,指导核酸酶至SEQ ID NO:86的第一向导RNA包含SEQ ID NO:1174,指导核酸酶至SEQ ID NO:125的第二向导RNA包含SEQ ID NO:1213。

[0039] 在一些包含两个gRNA的实施方案中,指导核酸酶至SEQ ID NO:86的第一向导RNA包含SEQ ID NO:1174,指导核酸酶至SEQ ID NO:107的第二向导RNA包含SEQ ID NO:1195。

[0040] 在一些包含两个gRNA的实施方案中,指导核酸酶至SEQ ID NO:64的第一向导RNA包含SEQ ID NO:1152,指导核酸酶至SEQ ID NO:106的第二向导RNA包含SEQ ID NO:1194。

[0041] 在一些包含两个gRNA的实施方案中,指导核酸酶至SEQ ID NO:85的第一向导RNA包含SEQ ID NO:1173,指导核酸酶至SEQ ID NO:114的第二向导RNA包含SEQ ID NO:1202。

[0042] 在一些包含两个gRNA的实施方案中,指导核酸酶至SEQ ID NO:86的第一向导RNA包含SEQ ID NO:1174,指导核酸酶至SEQ ID NO:114的第二向导RNA包含SEQ ID NO:1202。

[0043] 在一些包含两个gRNA的实施方案中,指导核酸酶至SEQ ID NO:83的第一向导RNA包含SEQ ID NO:1171,指导核酸酶至SEQ ID NO:114的第二向导RNA包含SEQ ID NO:1202。

[0044] 在一些包含两个gRNA的实施方案中,指导核酸酶至SEQ ID NO:53的第一向导RNA包含SEQ ID NO:1141,指导核酸酶至SEQ ID NO:114的第二向导RNA包含SEQ ID NO:1202。

[0045] 在一些包含两个gRNA的实施方案中,指导核酸酶至SEQ ID NO:83的第一向导RNA包含SEQ ID NO:1171,指导核酸酶至SEQ ID NO:112的第二向导RNA包含SEQ ID NO:1200。

[0046] 在一些包含两个gRNA的实施方案中,指导核酸酶至SEQ ID NO:74的第一向导RNA包含SEQ ID NO:1162,指导核酸酶至SEQ ID NO:114的第二向导RNA包含SEQ ID NO:1202。

[0047] 在一些实施方案中,该向导RNA靶向VIII型胶原 $\alpha 2$ 亚基基因(Co18A2),并指导核酸酶至选自SEQ ID NO:191-1063的靶序列。在一些实施方案中,本发明包含至少一个向导RNA,该向导RNA包含与选自SEQ ID NO:191-1063的靶序列(例如缺乏PAM的靶序列)的前20个核苷酸互补或同一的序列至少99%、98%、97%、96%、95%、94%、93%、92%、91%或90%同一的向导序列,其中用尿嘧啶替换SEQ ID NO:191-1063的前20个核苷酸中的胸腺嘧啶。

[0048] 在一些实施方案中,该向导RNA靶向Co18A2基因产物中的Gln455Lys突变,并指导核酸酶至选自SEQ ID NO:1064-1069的靶序列。在一些实施方案中,本发明包含至少一个向导RNA,该向导RNA包含与选自SEQ ID NO:1064-1069的靶序列(例如缺乏PAM的靶序列)的前20个核苷酸互补或同一的序列至少99%、98%、97%、96%、95%、94%、93%、92%、91%或90%同一的向导序列,其中用尿嘧啶替换SEQ ID NO:1064-1069的前20个核苷酸中的胸腺嘧啶。

[0049] 在一些实施方案中,该向导RNA靶向Co18A2基因产物中的Gln455Val突变,并指导核酸酶至选自SEQ ID NO:1070-1075的靶序列。在一些实施方案中,本发明包含至少一个向导RNA,该向导RNA包含与选自SEQ ID NO:1070-1075的靶序列(例如缺乏PAM的靶序列)的前20个核苷酸互补或同一的序列至少99%、98%、97%、96%、95%、94%、93%、92%、91%或90%同一的向导序列,其中用尿嘧啶替换SEQ ID NO:1070-1075的前20个核苷酸中的胸腺嘧啶。

[0050] 在一些实施方案中,该向导RNA靶向Co18A2基因产物中的Leu450Trp突变,并指导核酸酶至选自SEQ ID NO:1076-1084的靶序列。在一些实施方案中,本发明包含至少一个向导RNA,该向导RNA包含与选自SEQ ID NO:1076-1084的靶序列(例如缺乏PAM的靶序列)的前20个核苷酸互补或同一的序列至少99%、98%、97%、96%、95%、94%、93%、92%、91%或90%同一的向导序列,其中用尿嘧啶替换SEQ ID NO:1076-1084的前20个核苷酸中的胸腺嘧啶。

[0051] 在一些实施方案中,该向导RNA是双向导。在一些实施方案中,该向导RNA是单向导。在一些实施方案中,至少一个向导RNA包含crRNA、trRNA或crRNA和trRNA。

[0052] 在一些实施方案中,至少一个向导序列编码在载体上。在一些实施方案中,第一向导序列和第二向导序列编码在同一载体上。在一些实施方案中,第一向导序列和第二向导序列编码在不同载体上。在一些实施方案中,第一向导序列和第二向导序列由同一启动子和/或调节序列控制。

[0053] 在一些实施方案中,该向导序列与靶基因正链中的靶序列互补。在一些实施方案中,该向导序列与靶基因负链中的靶序列互补。在一些实施方案中,第一向导序列和第二向导序列与靶基因(即基因组DNA中的目的区域,如TCF4中的TNR)相反链中的第一靶序列和第二靶序列互补。

[0054] 在一些实施方案中,该向导RNA经化学修饰。在一些实施方案中,本发明进一步包含核酸酶。在一些实施方案中,该核酸酶是Cas蛋白质或其他切割双链或单链DNA的核酸酶。在一些实施方案中,该Cas蛋白质来自I型、II型或III型CRISPR/Cas系统。在一些实施方案中,该Cas蛋白质是Cas9或Cpf1。在一些实施方案中,该核酸酶是切口酶。在一些实施方案中,该核酸酶经修饰。在一些实施方案中,该修饰核酸酶包含核定位信号(NLS)。

[0055] 在一些实施方案中,本发明包含向导RNA和可药用载体的药物制剂。在一些实施方案中,该药物制剂包含一个或多个向导RNA和编码Cas蛋白质的mRNA。在一些实施方案中,该药物制剂包含一个或多个向导RNA和Cas蛋白质。

[0056] 在一些实施方案中,本发明包含在人个体中切除转录因子4(TCF4)基因中的至少部分三核苷酸重复(TNR)的方法。在一些实施方案中,使用两个向导RNA,其中第一向导RNA与TNR的5'序列互补,第二向导RNA与TNR的3'序列互补。在使用两个向导序列时,切除所靶向的基因组DNA区域之间的DNA序列。

[0057] 在一些实施方案中,该TNR等于或大于约40个三核苷酸重复。在一些实施方案中,该TNR等于或大于约50、45、40、35、30、25、20、15、10或5个三核苷酸重复。在一些实施方案中,该TNR等于或大于约40、50、60、70、80、90、100、110、120、130、140、150、160、170、180、190或200个三核苷酸重复。

[0058] 在一些实施方案中,该组合物或药物制剂包含至少两个切除至少部分TNR的向导。在一些实施方案中,切除整个TNR。

[0059] 在一些实施方案中,该组合物或药物制剂经病毒载体施用。在一些实施方案中,该组合物或药物制剂经脂质纳米颗粒施用。本领域技术人员已知的任何脂质纳米颗粒都适用于可选地与编码Cas蛋白质的mRNA一起递送本文提供的一个或多个向导RNA。在一些实施方案中,利用2017年3月30日提交的PCT/US2017/024973中所述的脂质纳米颗粒。在一些实施方案中,该脂质纳米颗粒包含本文提供的一个或多个向导RNA和编码Cas蛋白质的mRNA。

在一些实施方案中,该脂质纳米颗粒包含本文提供的一个或多个向导RNA而不含编码Cas蛋白质的mRNA。

[0060] 在一些实施方案中,本发明进一步包含滴眼液或眼膏的共同施用。在一些实施方案中,本发明进一步包含软性隐形眼镜的使用。

[0061] 在一些实施方案中,该人个体患有精神分裂症。

[0062] 在一些实施方案中,该人个体患有原发性硬化性胆管炎 (PSC)。

[0063] 在一些实施方案中,本发明包含在人个体中减少COL8A2基因的突变等位基因(如Gln455Lys、Gln455Val或Leu450Trp)的表达或改变核苷酸序列来矫正该突变等位基因的方法。

[0064] 在一些实施方案中,该人个体患有角膜上皮营养不良 (FECD) 或多形性角膜后层营养不良 (PPCD)。在一些实施方案中,该人个体患有FECD。在一些实施方案中,该个体具有FECD家族史。

[0065] 在一些实施方案中,该个体由于施用而具有视敏度的改善、稳定化或下降减慢。在一些实施方案中,该个体由于施用而具有通过角膜厚度测量法测量的改善、稳定化或变化减慢。在一些实施方案中,该个体由于施用而具有基于角膜内皮显微术 (specular microscopy) 的改善、稳定化或变化减慢。在一些实施方案中,该个体由于施用而延迟需要角膜移植的时间。

[0066] 在一些实施方案中,本发明包含组合物或药物在制备用于治疗TCF4基因中具有TNR扩增或COL8A2基因中具有导致基因产物中Gln455Lys、Gln455Val或Leu450Trp突变的突变的人个体的药物中的用途。

[0067] 其他目的和优势将部分在以下发明详述中给出,部分将从发明详述显而易见,或者可以通过实施来了解。该目的和优势可以借助所附权利要求书中具体指出的要素和组合来实现和获得。

[0068] 应理解,前面的概述和以下发明详述仅是示例性和解释性的,并非限制权利要求。

[0069] 并入本说明书并构成本说明书的一部分的附图举例说明一个(若干)实施方案,与描述一起用于解释本文所述的原理。

[0070] 附图简述

[0071] 图1提供用一对gRNA切除TCF4内含子3中的TNR扩增区的示意图,一个gRNA具有靶向TNR 5'的内含子3区域的向导序列,另一个gRNA具有靶向TNR 3'的内含子3区域的向导序列。虽然该附图显示切除发生在TNR的精确边界,但实际上切除可以更大或更小,且包括内含子的上游和/或下游区域。

[0072] 图2提供示意图,显示进行切除测试的93对gRNA的预测切除片段大小。数字对应于所测试的向导的各靶序列的SEQ ID NO。gRNA对按切除百分比排序(列表中最上面的一对具有最高的切除率)。“0”标记TNR区的中心。

[0073] 序列描述

[0074]

**表 1 提供本文中引用的某些序列的列表。表 1: 序列描述**

序列	描述	SEQ ID NO:
表 2 中提供的序列	TCF4 内含子 3 中 TNR 5'的靶序列	1-93
表 2 中提供的序列	TCF4 内含子 3 中 TNR 3'的靶序列	94-190
表 3 中提供的序列	野生型 COL8A2 的靶序列	191-1063
表 4 中提供的序列	COL8A2 Gln455Lys 突变的靶序列	1064-1069
表 5 中提供的序列	COL8A2 Gln455Val 突变的靶序列	1070-1075
表 6 中提供的序列	COL8A2 Leu450Trp 突变的靶序列	1076-1084
GTTTGTGTGA TTTTGCTAAA ATGCATCACC AACAGCGAAT GGCTGCCTTA GGGACGGACA AAGAGCTGAG TGATTTACTG GATTTACAGTG CGgtaagaaa gaacggtgga aactaacaac agctgtgaaa aaaacaaaac aaaaacccaa acacttcagc tagaaaccag taggaatcta aaggacagta ataattttta attggctgaa tccttggtaa atatgaaggt ctttttgaca agtttttaac tataattttg tgggtgtgat gaagattcag gctttttttt ttttttgagt tttattactg gccttcaatt ccctaccacc tgattacccc aaataatgga atctcacc agtggaaagc aaaaatagac acccctaaaa ctaaaccacc cctaaaactt ggccatgtct gaacactgag actactaata ctttgcacac tactcttctg tttatttatt gtttttgaa atggaaaata gaaaatagga gaccagttg tctctttaa gttttaagct aatgatgctt tggattggta ggacctgtc cttacatctt acctcctagt tacatctttt cctaggattc ttaaactag tatggatatg ctgagcatac attctttaga accttttggg ctgttttggf aaatctcgta gtcgtaggat cagcacaag cggaacttga cacacttgtg gagttttacg gctgtacttg gtccttctcc atcccttgc ttccttttcc taaaccaagt cccagacatg tcaggagaat gaattcattt ttaatgccag atgagtttgg tgtaagatgc atttgtaaag caaaaataaa agaatccaca aaacacacaa ataaaatcca aacggccttc caagtggggc tctttcatgc <b>tgctgctgct</b> <b>gctgctgctg</b> <b>ctgctgctgc</b> <b>tgctgctgct</b> <b>gctgctgctg</b> <b>ctgctgctgc</b> <b>tgctgctgct</b> gctcctctcc ctcctcctcc ttctcctcct cctcctcctc ttctagacct tcttttgag aatggcctt cggaagtttt gccaggaaac gtagccctag gcaggcagct ttgcagcccc ctttctgctt gttgcacttt ctccattcgt tcctttgctt tttgcaggct ctgactcagg gaaggtgtgc attatccact agatacgtcg aagaagaggg aaaccaatta gggtcgaaat aaatgctgga gagagagggg gtgaaagaga gaggagagt gagagagaga gagagtcttg cttcaaattg ctctcctggt agagacgaaa tgagaattta gtgcagggtg cacttttatt tttatttggg ttcacatatg acaggcaaat cctatacgag atggaaatgg acattgccac gtttatggcc aaggttttca atataaaaca aaacaacttt ttctctctcc ttggtgaaac tagtgttttt ctagagaggc tgctgcctc caacctgaat cttgataaca ttatggggac tgtgtttggt ccaaatgtag cagtagtact gcttggccat ctaataaacc tgaggaaaaa gaaagaacag agtgataatg ggggctgggg tgggatctgt aatgttgtt ctcttttagt tttaagttgg atggtgatgt attttactaa ataaaccctt agcataaact ctaagctggt tggtaacagt atgaaagatc tttgaggagc tctgaaggca caagtgtctt cttttcaact gtaataatttc tttgtttctt ttagATGTTT TCACCTCTG TGAGCAGTGG GAAAAATGGA CCAACTTCTT TGGCAAGTGG ACATTTTACT GGCTCAA	含侧翼外显子的 TCF4 内含子 3 序列, 反向链(GRCh37/hg19)。虽然通常称为内含子 3, 但存该基因的许多选择性剪接同工型, 使得此内含子可以不处在每个转录物的第三和第四外显子之间。 黑体字表示 ctg 重复(TNR)。此区域大小可变。 大写字母表示相邻的 5'和 3'外显子的序列。	1085
mN*mN*mN*NNNNNNNNNNNNNNNNNGUUUUAG AmGmCmUmAmGmAmAmUmAmGmCAAGUUAAA AUAAGGCUAGUCCGUUAUCAmAmCmUmUmGmAm AmAmAmAmGmUmGmGmCmAmCmCmGmAmGmUmC mGmGmUmGmCmU*mU*mU*mU	sgRNA 修饰序列 “N”可以是任意天然或非天然核苷酸。 * = PS 连接; 'm' = 2'-O-Me 核苷酸	1086

[0075]

NNNNNNNNNNNNNNNNNNNNNGUUUUAGAGCUAUGCUGUUUUUG	crRNA 序列 “N”可以是任意天然或非天然核苷酸。	1087
AACAGCAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUG AAAAAGUGGCACCGAGUCGGUGCUUUUUUU	trRNA 序列	1088

[0076] 发明详述

[0077] 定义

[0078] 本文所用的术语“治疗”涵盖用于个体中疾病的治疗剂的任何施用或应用,包括抑制疾病、阻断其发展、减轻疾病的一种或多种症状、治愈疾病、或防止疾病的一种或多种症状的复发。例如,FECD的治疗可以包括减轻FECD的症状,以及减少TCF4基因中TNR的数目,导致FECD症状改善、减慢疾病进展或治愈疾病/防止疾病症状复发。

[0079] 本文所用的“FECD”指角膜上皮营养不良,也称为Fuchs营养不良。FECD还可包括无症状但有遗传障碍(如与FECD发生率提高关联的TCF4内含子3中的TNR扩增)的个体。FECD还可包括无症状但有已知的FECD家族史和TCF4内含子3中的TNR扩增的个体。

[0080] 本文所用的“TNR”指三核苷酸重复。“微卫星重复”指包含一组2个至9个碱基对的多个重复的DNA短序列。术语微卫星重复涵盖TNR。“TNR扩增”指高于正常的三核苷酸重复数。对于TCF4的内含子3,例如,TNR扩增可以表征为约50或更多个TNR。与疾病相关的TNR扩增的范围通常在50和1000之间,但已鉴定了一些具有>1000个重复的患者。通常不认为在TCF4的内含子3中具有<50个TNR的患者通过TNR扩增机制处于提高的疾病风险,但他们仍可以从减少TNR数目受益。

[0081] 由TNR引起和/或表征为存在TNR的疾病可以称为“三核苷酸重复障碍”、“三核苷酸重复扩增障碍”、“三联体重复扩增障碍”或“密码子重复障碍”。

[0082] “向导RNA”和“gRNA”在本文中可互换使用。gRNA包含或由CRISPR RNA(crRNA)和trRNA(也称为tracrRNA)。crRNA和trRNA可以结合在一个RNA分子上(单向导RNA(sgRNA)),或者可以解离在分开的RNA分子上(双向导RNA(dgRNA))。

[0083] 本申请中所用的“向导序列”指crRNA或trRNA内约20个碱基对的序列,其与靶序列互补并发挥功能来指导RNA至靶序列,进行核酸酶切割。也可以用略短或略长的序列作为向导,例如长度为15、16、17、18、19、21、22、23、24或25个碱基对。在一些实施方案中,向导序列的长度对应于例如本文所述靶序列的长度。

[0084] 本文所用的“靶序列”指向导RNA指导核酸酶切割的核酸序列。靶序列在个体的基因组DNA内。在一些实施方案中,Cas蛋白质可由向导RNA指导至靶序列,其中向导RNA与靶序列杂交,核酸酶切割靶序列。靶序列包括基因组DNA的正链和负链二者(即给出的序列和该序列的反向互补序列),因为Cas蛋白质的核酸底物是双链核酸。因此,在说向导序列“与靶序列互补”时,理解为该向导序列可以指导向导RNA(例如在RNP中)与本文中提供的靶序列的反向互补序列结合。因此,在一些实施方案中,在向导序列结合靶序列的反向互补序列时,除在向导序列中用U取代T外,向导序列与靶序列(例如不包含PAM的靶序列)的前20个核苷酸相同。

[0085] 本文所用的“PAM”或“原间隔区邻近基序”指必须与靶序列相邻的序列。所需PAM取决于具体CRISPR系统而变。在源自酿脓链球菌(*Streptococcus pyogenes*)的CRISPR/Cas系

统中,为了最佳切割,靶DNA必须正好在5'-NGG PAM(其中“N”是任意核碱基,后面跟着两个鸟嘌呤核碱基)之前,而其他Cas9直向同源物具有不同的要求。虽然酿脓链球菌Cas9也可以识别5'-NAG PAM,但它在这些PAM位点处的切割效率似乎较低。表2的靶序列包含PAM。

[0086] 在一些实施方案中,向导RNA和Cas蛋白质可以形成“核糖核蛋白”(RNP)。在一些实施方案中,向导RNA引导诸如Cas9的核酸酶至靶序列,向导RNA与靶序列杂交,核酸酶切割靶序列。

[0087] 本文所用的“插入/缺失(indel)”指包含在核酸中双链断裂(DSB)位点处插入或缺失的许多核苷酸的插入/缺失突变。

[0088] 本文所用的“切除片段”指在将两个或多个向导RNA与Cas mRNA或蛋白质一起使用时可以发生的连续多个核苷酸的缺失。

[0089] 组合物

[0090] 描述用于治疗FECD的组合物。在一些方面,该组合物包含向导RNA,该向导RNA指导核酸酶至TCF4基因中的TNR,从而切割TNR,从而治疗在TCF4基因中具有TNR的疾病,包括FECD、PPCD、PSC和精神分裂症。在一些实施方案中,该组合物包含两个向导RNA,该向导RNA指导核酸酶至TCF4内含子3中的第一和第二位置,其中核酸酶在第一和第二位置处切割TCF4内含子3,切除第一和第二切割之间的核酸片段,从而切除一些或全部包含在TCF4内含子3内的TNR,治疗在TCF4基因中具有TNR的疾病,包括FECD、PPCD、PSC和精神分裂症。在其他方面,该组合物包含向导RNA,该向导RNA经DNA中的靶序列指导核酸酶至COL8A2基因,从而介导以切割序列为目的的NHEJ,导致引入插入/缺失或介导HR或HDR,其中可通过使用模板来矫正DNA中的突变,并治疗FECD或PPCD。下文描述组合物的实施方案。

[0091] 向导RNA

[0092] 在一些实施方案中,本发明的组合物包含含有指导核酸酶如Cas9至靶DNA序列的一个或多个向导序列的向导RNA(gRNA)。gRNA包含crRNA和trRNA。在本文所述的每个组合物和方法实施方案中,crRNA和trRNA可以结合在一个RNA上(sgRNA),或者可以解离在分开的RNA上(dsRNA)。

[0093] 在本文所述的每个组合物和方法实施方案中,向导RNA可以包含两个RNA分子作为“双向导RNA”或“dgRNA”。dgRNA包含含有crRNA的第一RNA分子和含有trRNA的第二RNA分子。第一和第二RNA分子共价连接,但可以通过crRNA和trRNA上的旗杆区之间的碱基配对形成RNA双链体。

[0094] 在本文所述的每个组合物和方法实施方案中,向导RNA可以包含单个RNA分子作为“单向导RNA”或“sgRNA”。sgRNA包含与trRNA共价连接的crRNA。在一些实施方案中,crRNA和trRNA通过接头共价连接。在一些实施方案中,sgRNA通过crRNA和trRNA上的旗杆区之间的碱基配对形成茎-环结构。

[0095] 在一些实施方案中,trRNA可以包含来自天然存在的CRISPR/Cas系统的全部或部分野生型trRNA序列。在一些实施方案中,trRNA可以包含截短或修饰的野生型trRNA。trRNA的长度取决于所使用的CRISPR/Cas系统。在一些实施方案中,trRNA包含5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、25、30、40、50、60、70、80、90、100或超过100个核苷酸,或由5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、25、30、40、50、60、70、80、90、100或超过100个核苷酸组成。在某些实施方案中,trRNA长度为至少26个核苷酸。在其他实施方案中,

trRNA长度为至少40个核苷酸。在一些实施方案中, trRNA可以包含某些二级结构, 例如一个或多个发夹或茎-环结构、或一个或多个凸起结构。

[0096] 在一些实施方案中, gRNA经化学修饰。将含有一个或多个修饰核苷或核苷酸的向导gRNA称为“修饰”gRNA或“化学修饰”gRNA, 以描述替换或附加至规范A、G、C和U残基使用的一个或多个非天然和/或天然存在的成分或构型的存在。在一些实施方案中, 用非规范核苷或核苷酸(本文中称为“修饰的”)合成修饰gRNA。修饰核苷和核苷酸可以包含以下一种或多种: (i) 改变(例如替换)一个或两个非连接磷酸氧和/或主链磷酸二酯键中的一个或多个连接磷酸氧(示例性主链修饰); (ii) 改变(例如替换)核糖的组分, 例如核糖上的2' 羟基(示例性糖修饰); (iii) 用“脱磷酸”接头广泛取代磷酸部分(示例性主链修饰); (iv) 修饰或替换天然存在的核碱基, 包括用非规范核碱基替换(示例性碱基修饰); (v) 替换或修饰磷酸核糖主链(示例性主链修饰); (vi) 修饰寡核苷酸的3' 端或5' 端, 例如去除、修饰或替换末端磷酸基团或缀合部分、帽或接头(如3' 或5' 帽修饰可以包含糖和/或主链修饰); 和(vii) 修饰或替换糖(示例性糖修饰)。

[0097] 可以组合上文所列修饰, 以提供包含可以具有两个、三个、四个或多个修饰的核苷和核苷酸(统称“残基”)的修饰gRNA。例如, 修饰残基可以具有修饰糖和修饰核碱基。在一些实施方案中, 修饰gRNA的每个碱基, 例如, 所有碱基都具有修饰磷酸基团, 如硫代磷酸基团。在某些实施方案中, 用硫代磷酸基团替换gRNA分子的全部或基本全部磷酸基团。在一些实施方案中, 修饰gRNA在RNA的5' 端或附近包含至少一个修饰残基。在一些实施方案中, 修饰gRNA在RNA的3' 端或附近包含至少一个修饰残基。

[0098] 在一些实施方案中, gRNA包含一个、两个、三个或多个修饰残基。在一些实施方案中, 修饰gRNA中至少5% (例如至少约5%、至少约10%、至少约15%、至少约20%、至少约25%、至少约30%、至少约35%、至少约40%、至少约45%、至少约50%、至少约55%、至少约60%、至少约65%、至少约70%、至少约75%、至少约80%、至少约85%、至少约90%、至少约95%或约100%) 的位置是修饰核苷或核苷酸。

[0099] 非修饰核酸可倾向于由例如细胞核酸酶降解。例如, 核酸酶可以水解核酸磷酸二酯键。因此, 一方面, 本文所述gRNA可以包含一个或多个修饰核苷或核苷酸, 例如以引入对核酸酶的稳定性。在一些实施方案中, 在引入体内和离体细胞群体时, 本文所述修饰gRNA分子可以显示减少的先天免疫反应。术语“先天免疫反应”包括对外源核酸(包括单链核酸)的细胞反应, 其涉及细胞因子表达和释放(尤其是干扰素)及细胞死亡的诱导。

[0100] 在主链修饰的一些实施方案中, 可以通过用不同取代基替换一个或多个氧来修饰修饰残基的磷酸基团。此外, 修饰残基(例如存在于修饰核酸中的修饰残基)可以包括用本文所述修饰磷酸基团广泛替换非修饰磷酸部分。在一些实施方案中, 磷酸主链的主链修饰可以包括产生不带电荷接头或具有不对称电荷分布的带电荷接头的改变。

[0101] 修饰磷酸基团的实例包括硫代磷酸、硒代磷酸(phosphoroselenates)、borano磷酸、borano磷酸酯、氢膦酸酯、磷酰胺、烷基或芳基磷酸和磷酸三酯。非修饰磷酸基团中的磷原子是非手性的。但是, 用以上原子或原子团之一替换非桥连氧之一可以使磷原子成为手性的。产生立体构型的磷原子可具有“R”构型(本文中称为Rp)或“S”构型(本文中称为Sp)。还可以通过用氮(桥连磷酰胺)、硫(桥连硫代磷酸)和碳(桥连亚甲磷酸)替换桥连氧(即使磷酸与核苷连接的氧)来修饰主链。替换可以发生在任一连接氧处或两个连接氧处。

[0102] 在某些主链修饰中,可以用含有非磷的连接子替换磷酸基团。在一些实施方案中,可以用中性部分替换带电荷的磷酸基团。可以替换磷酸基团的部分的实例可以非限制性地包括例如甲基磷酸、羟氨基、硅氧烷、碳酸、羧甲基、氨基甲酸、酰胺、硫醚、环氧乙烷接头、磺酸、磺酰胺、thioformacetal、formacetal、脞、亚甲基亚氨基、亚甲基甲基亚氨基、亚甲基亚胍基、亚甲基二甲基亚胍基和亚甲基氧甲基亚氨基。

[0103] 也可以构建可以模拟核酸的支架,其中用核酸酶抗性核苷或核苷酸替代物替换磷酸接头和核糖。这类修饰可以包括主链和糖修饰。在一些实施方案中,该核碱基可束缚于替代主链。实例可非限制性地包括吗啉代、环丁基、吡咯烷和肽核酸(PNA)核苷替代物。

[0104] 修饰核苷和修饰核苷酸可以包含一个或多个对糖基团的修饰,即在糖修饰处。例如,可以修饰2' 羟基(OH),例如用许多不同的“氧”或“脱氧”取代基替换。在一些实施方案中,对2' 羟基的修饰可以增强核酸的稳定性,因为羟基可以不再去质子化形成2' -醇盐离子。

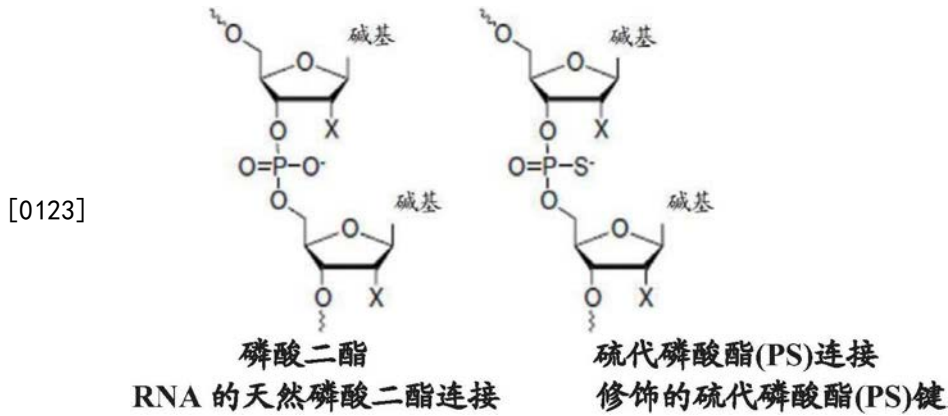
[0105] 2' 羟基修饰的实例可以包括:烷氧基或芳氧基(OR,其中“R”可以是例如烷基、环烷基、芳基、芳烷基、杂芳基或糖);聚乙二醇(PEG)  $O(CH_2CH_2O)_nCH_2CH_2OR$ ,其中R可以是例如H或可选地取代的烷基,n可以是2至20(例如从0至4、从0至8、从0至10、从0至16、从1至4、从1至8、从1至10、从1至16、从1至20、从2至4、从2至8、从2至10、从2至16、从2至20、从4至8、从4至10、从4至16和从4至20)的整数。在一些实施方案中,2' 羟基修饰可以是2'-O-Me。在一些实施方案中,2' 羟基修饰可以是2'-氟修饰,其用氟替换2' 羟基。在一些实施方案中,2' 羟基修饰可以包括“锁定”核酸(LNA),其中可以使2' 羟基例如通过 $C_{1-6}$ 亚烷基或 $C_{1-6}$ 杂亚烷基桥与同一核糖的4' 碳连接,其中示例性桥可以包括亚甲基、亚丙基、醚或氨基桥;O-氨基(其中氨基可以是例如 $NH_2$ 、烷基氨基、二烷基氨基、杂环、芳基氨基、二芳基氨基、杂芳基氨基、或二杂芳基氨基、乙二胺或多聚氨基)和氨基烷氧基 $O(CH_2)_n$ -氨基(其中氨基可以是例如 $NH_2$ 、烷基氨基、二烷基氨基、杂环、芳基氨基、二芳基氨基、杂芳基氨基、或二杂芳基氨基、乙二胺或多聚氨基)。在一些实施方案中,2' 羟基修饰可以包括“非锁定”核酸(UNA),其中核糖环缺乏 $C2'-C3'$ 键。在一些实施方案中,2' 羟基修饰可以包括甲氧基乙基(MOE)  $(OCH_2CH_2OCH_3)$ ,例如PEG衍生物)。

[0106] “脱氧”2' 修饰可以包括:氢(即脱氧核糖,例如在部分dsRNA的突出部分);卤素(例如溴、氯、氟或碘);氨基(其中氨基可以是例如例如 $-NH_2$ 、烷基氨基、二烷基氨基、杂环、芳基氨基、二芳基氨基、杂芳基氨基、或二杂芳基氨基或氨基酸); $NH(CH_2CH_2NH)_nCH_2CH_2$ -氨基(其中氨基可以是例如上文所述)、 $-NHC(O)R$ (其中R可以是例如烷基、环烷基、芳基、芳烷基、杂芳基或糖)、氰基;巯基;烷基-硫-烷基;硫代烷氧基;及烷基、环烷基、芳基、烯基和炔基,其可以可选地用例如本文所述氨基取代。

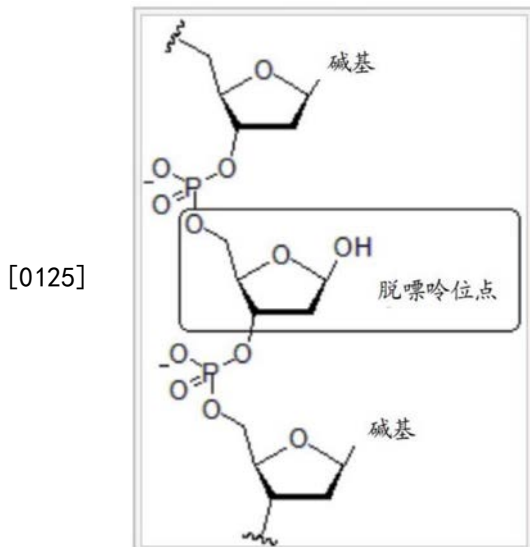
[0107] 糖修饰可以包含糖基,其还可以包含一个或多个与核糖中相应碳的构型相比具有相反立体化学构型的碳。因此,修饰核酸可以包含含有例如阿拉伯糖作为糖的核苷酸。修饰核酸还可以包含脱碱基糖。这些脱碱基糖还可以进一步在一个或多个组成糖原子处修饰。修饰核酸还可以包含一个或多个处于L形式的糖,例如L-核昔。

[0108] 可以掺入修饰核酸的本文所述修饰核苷和核苷酸可以包含修饰碱基,也称为核碱基。核碱基的实例包括但不限于腺嘌呤(A)、鸟嘌呤(G)、胞嘧啶(C)和尿嘧啶(U)。可以修饰或完全取代这些核碱基,以提供可以掺入修饰核酸的修饰残基。核苷酸的核碱基可以独立

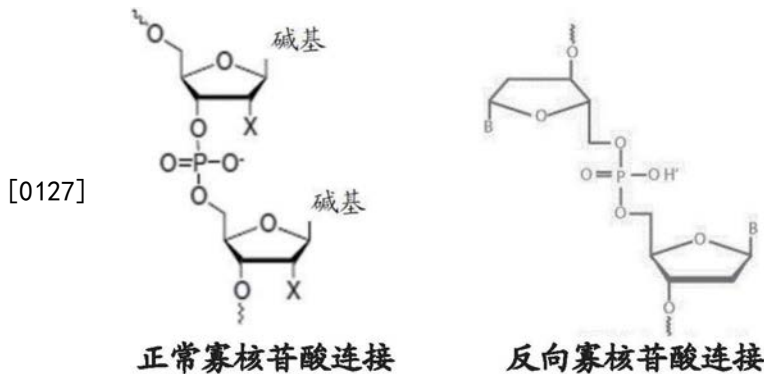




[0124] 脱碱基核苷酸指缺乏含氮碱基的那些核苷酸。下图显示含有缺乏碱基的脱嘌呤位点的寡核苷酸(也称为脱嘌呤)位点的寡核苷酸:



[0126] 反向碱基指具有与正常5'至3'连接相反的连接(即5'至5'连接或3'至3'连接)的那些。例如:



[0128] 脱碱基核苷酸可以用反向连接附着。例如,脱碱基核苷酸可以经5'至5'连接附着于末端5'核苷酸,或者脱碱基核苷酸可以经3'至3'连接附着于末端3'核苷酸。末端5'或3'核苷酸处的反向脱碱基核苷酸也可以称为反向脱碱基末端帽。

[0129] 在一些实施方案中,修饰向导RNA的5'端前3、4或5个核苷酸中的一个或多个和3'端最后3、4或5个核苷酸中的一个或多个。在一些实施方案中,该修饰是2'-O-Me、2'-F、反向脱碱基核苷酸、PS键、或本领域公知的提高稳定性和/或性能的其他核苷酸修饰。

[0130] 在一些实施方案中,5'端的前4个核苷酸和3'端的最后4个核苷酸以硫代磷酸酯(PS)键连接。

[0131] 在一些实施方案中,5'端的前3个核苷酸和3'端的最后3个核苷酸包含2'-O-甲基(2'-O-Me)修饰核苷酸。在一些实施方案中,5'端的前3个核苷酸和3'端的最后3个核苷酸包含2'-氟(2'-F)修饰核苷酸。在一些实施方案中,5'端的前3个核苷酸和3'端的最后3个核苷酸包含反向脱碱基核苷酸。

[0132] 在一些实施方案中,该向导RNA包含修饰sgRNA。在一些实施方案中,该sgRNA包含SEQ ID NO:1086中所示的修饰模式,其中N是任意天然或非天然核苷酸,其中N的总体包含本文所述的指导核酸酶至TC4靶序列的向导序列。用于TCF4的向导RNA。

[0133] 在一些实施方案中,该组合物包含至少一个向导RNA(gRNA),该向导RNA包含与SEQ ID NO:1-190的核酸中的任一个互补的向导序列,或由与SEQ ID NO:1-190的核酸中的任一个互补的向导序列组成。在一些实施方案中,该组合物包含至少一个向导RNA(gRNA),该向导RNA包含指导核酸酶至SEQ ID NO:1-190的核酸中任一个的向导序列,或由指导核酸酶至SEQ ID NO:1-190的核酸中任一个的向导序列组成。在一方面,该组合物包含至少一个gRNA,该gRNA包含与靶序列互补的向导序列,或由与靶序列互补的向导序列组成,该靶序列与SEQ ID NO:1-190的核酸中任一个至少99%、98%、97%、96%、95%、94%、93%、92%、91%或90%同一。在一方面,该组合物包含至少一个gRNA,该gRNA包含指导核酸酶至靶序列的向导序列,或由指导核酸酶至靶序列的向导序列组成,该靶序列与SEQ ID NO:1-190的核酸中任一个至少99%、98%、97%、96%、95%、94%、93%、92%、91%或90%同一。

[0134] 在一些方面,该组合物包含至少一个gRNA,该gRNA包含或由以下序列组成:与SEQ ID NO:1089-1278的核酸中任一个至少99%、98%、97%、96%、95%、94%、93%、92%、91%或90%同一的向导序列。在一些方面,该组合物包含至少一个gRNA,该gRNA包含或由以下序列组成:与SEQ ID NO:1089-1278的核酸中任一个同一的向导序列。

[0135] 在其他实施方案中,该组合物包含至少两个gRNA,该gRNA包含或由以下序列组成:至少两个与靶序列中的任一个互补的向导序列,该靶序列选自SEQ ID NO:1-190的核酸中的任意两个或多个。在一些实施方案中,该组合物包含至少两个gRNA,该gRNA包含或由以下序列组成:至少两个与靶序列中的任一个互补的向导序列,该靶序列选自与SEQ ID NO:1-190的核酸中的任一个至少99%、98%、97%、96%、95%、94%、93%、92%、91%或90%同一的核酸中的任意两个或多个。

[0136] 在一些实施方案中,将靶向TCF4 TNR 5'序列的gRNA与靶向TCF4 TNR 3'序列的gRNA一起用于切除TCF4 TNR的目的。在一些实施方案中,将与SEQ ID NO:1-93的靶序列互补的向导序列和与SEQ ID NO:94-190的靶序列互补的向导序列一起使用。

[0137] 在一些实施方案中,靶向TCF4 TNR 5'序列的gRNA与靶向TCF4 TNR 3'序列的gRNA的一起使用切除了具有延长(extended) TNR序列的患者中的TCF4内含子3中的TNR的全序列。例如,在一些实施方案中,靶向TNR扩增5'和3'序列的gRNA的组合切除具有至少40、至少50、至少60、至少70、至少80、至少90、至少100、至少150、至少200、至少250、至少300、至少400、至少500、至少600、至少700、至少800、至少900、或至少1000或更多个重复的TNR。在一些实施方案中,此方法用于切除数目高于40的TNR扩增。在一些实施方案中,靶向TCF4 TNR 5'序列的gRNA与靶向TNR重复内的gRNA的一起使用、或靶向TNR重复内的gRNA与靶向

TCF4 TNR 3' 序列的gRNA的一起使用切除了具有延长TNR序列的患者中的TCF4内含子中的部分延长TNR,从而缩短TNR的长度。在一些实施方案中,一个向导RNA靶向TCF4 TNR的5' 序列,另一个向导RNA靶向TCF4 TNR的3' 序列,从而切除全部TNR。

[0138] 两个或多个靶向TCF4的向导RNA的组合

[0139] 在某些实施方案中,该组合物包含超过一个gRNA。每个gRNA可以包含不同的向导序列,使得所结合的核酸酶切割超过一个靶序列。在一些实施方案中,该gRNA可以具有相同或不同的特性,如在RNP复合体内的活性或稳定性。在一些涉及载体的实施方案中,在使用超过一个gRNA时,各gRNA可以编码在同一载体上或不同载体上。用于驱动该超过一个gRNA表达的启动子可以相同或不同。在某些涉及脂质纳米颗粒的实施方案中,该两个或多个gRNA可以配制在同一脂质纳米颗粒中或分开的脂质纳米颗粒中。

[0140] 在一些实施方案中,各gRNA的向导序列与TCF4基因同一条链中的靶序列互补。在一些实施方案中,各gRNA的向导序列与TCF4基因正链中的靶序列互补。在一些方面,各gRNA的向导序列与TCF4基因负链中的靶序列互补。在一些实施方案中,各gRNA的向导序列与TCF4基因的相反链中的靶序列互补。

[0141] 在一些方面,该组合物包含至少两个gRNA,其中该至少两个gRNA包含将核酸酶靶向至两个不同位置的向导序列。在一些实施方案中,该两个gRNA可以在TCF4基因的TNR侧翼(即该两个gRNA在TNR的任一侧;换言之,一个gRNA在TNR的5',另一个gRNA在TNR的3')。在一些实施方案中,一个gRNA在TCF4基因的TNR内,另一个gRNA在TCF4基因的TNR外(即在TNR侧翼)。在一些实施方案中,该两个gRNA将核酸酶靶向至相隔约3000、2500、2000、1500、1000、500、400、300、200、150、100、50或30个核苷酸的靶序列。在一些实施方案中,该核酸酶切割各位置,切除包含TCF4内含子3的TNR扩增区的DNA片段。

[0142] 在一些实施方案中,仅使用一个gRNA。在一些实施方案中,使用靶向TCF4 TNR 5' 序列的gRNA。在一些实施方案中,该向导序列与SEQ ID NO:1-93的靶序列互补。在一些实施方案中,使用靶向TCF4 TNR 3' 序列的gRNA。在一些实施方案中,使用与SEQ ID NO:94-190的靶序列互补的向导。在一些实施方案中,使用靶向TCF4中的TNR重复扩增内的序列的gRNA。在一些实施方案中,单向导的使用导致在NHEJ期间形成插入/缺失,其减少或消除TNR序列。在一些实施方案中,单向导的使用导致在NHEJ期间形成插入/缺失,其减少或消除部分TNR序列。

[0143] 用于COL8A2的向导RNA

[0144] 在一些实施方案中,该组合物包含至少一个向导RNA (gRNA),该向导RNA包含或由以下序列组成:与SEQ ID NO:191-1084的核酸中的任一个互补的向导序列。在一方面,该组合物包含至少一个gRNA,该gRNA包含与靶序列互补的向导序列,或由与靶序列互补的向导序列组成,该靶序列与SEQ ID NO:191-1084的核酸中的任一个至少99%、98%、97%、96%、95%、94%、93%、92%、91%或90%同一。

[0145] 在其他实施方案中,该组合物包含至少两个gRNA,该gRNA包含或由以下序列组成:至少两个与SEQ ID NO:191-1084的核酸中的任意两个或多个互补的向导序列。在一些实施方案中,该组合物包含至少两个gRNA,该gRNA包含或由以下序列组成:至少两个与SEQ ID NO:191-1084的核酸序列至少99%、98%、97%、96%、95%、94%、93%、92%、91%或90%同一的任意两个或多个核酸互补的向导序列。

[0146] 在一些实施方案中,使用靶向无已知突变的野生型COL8A2中的序列的gRNA。在一些实施方案中,使用与SEQ ID NO:191-1063的靶序列互补的向导序列。

[0147] 在一些实施方案中,使用靶向对应于COL8A2中已知产生Gln455Lys突变的突变的序列的gRNA。在一些实施方案中,使用与SEQ ID NO:1064-1069的靶序列互补的向导序列,例如以选择性编辑由c.1364C>A核苷酸改变引起的Gln455Lys突变。

[0148] 在一些实施方案中,使用靶向对应于COL8A2中已知产生Gln455Val突变的突变的序列的gRNA。在一些实施方案中,使用与SEQ ID NO:1070-1075的靶序列互补的向导序列,例如以选择性编辑由c.1363-1364CA>GT核苷酸改变引起的Gln455Val突变。

[0149] 在一些实施方案中,使用靶向对应于COL8A2中已知产生Leu450Trp突变的突变的序列的gRNA。在一些实施方案中,使用与SEQ ID NO:1076-1084的靶序列互补的向导序列,例如以选择性编辑由c.1349T>G核苷酸改变引起的Leu450Trp突变。

[0150] 靶序列

[0151] 在一些实施方案中,该向导RNA将核酸酶靶向至COL8A2基因。在一些方面,该crRNA包含与TCF4基因中TNR侧翼的靶序列互补和杂交的向导序列。在一些实施方案中,利用两个gRNA。在这类实施方案中,该两个gRNA可以在TCF4基因的TNR侧翼(即该两个gRNA在TNR的任一侧)。在一些实施方案中,一个gRNA在TCF4基因的TNR内,另一个gRNA在TCF4基因的TNR外(即侧翼)。在一些实施方案中,该crRNA进一步包含与trRNA的部分互补和杂交的旗杆区。在一些实施方案中,crRNA可以与从细菌CRISPR基因座转录的天然存在的crRNA的结构平行,其中向导序列作为CRISPR/Cas9系统的“间隔区”,旗杆对应于CRISPR基因座上间隔区侧翼的一部分重复序列。

[0152] TCF4的靶序列

[0153] 本发明的组合物可以指导至并切割TCF4基因中TNR内或侧翼的靶序列。例如,所提供的核酸酶可以识别并切割TNR靶序列。在一些实施方案中,向导RNA可以指导Cas蛋白质至TCF4基因中TNR侧翼的靶序列,其中向导RNA的向导序列与靶序列或其反向互补序列杂交并指导Cas蛋白质切割靶序列。在一些实施方案中,向导RNA可以指导Cas蛋白质至TCF4基因中TNR内的靶序列。在一些实施方案中,超过一个向导RNA可以指导Cas蛋白质至TCF4基因中TNR侧翼的两个靶序列。在一些实施方案中,超过一个向导RNA可以指导Cas蛋白质至两个靶序列,其中一个靶序列在TCF4基因中TNR的侧翼,另一个靶序列在TCF4基因中的TNR内。

[0154] 在一些实施方案中,根据TCF4基因中TNR附近的靶序列确定该一个或多个向导RNA的选择。例如,在一些实施方案中,该一个或多个向导RNA包含与TCF4基因中TNR侧翼的靶序列互补的向导。在一些实施方案中,该一个或多个向导RNA的crRNA序列与选自SEQ ID NO:1-190的靶序列互补和杂交。

[0155] 在一些实施方案中,该靶序列可以与向导RNA的向导序列互补。在一些实施方案中,向导RNA的向导序列和它对应的靶序列之间的互补程度或同一性可以是约50%、55%、60%、65%、70%、75%、80%、85%、90%、95%、96%、97%、98%、99%或100%。在一些实施方案中,靶序列和gRNA的向导序列可以100%互补或同一。在其他实施方案中,靶序列和gRNA的向导序列可以包含至少一个错配。例如,靶序列和gRNA的向导序列可以包含1、2、3、4、5、6、7、8、9或10个错配,其中向导序列的总长度为约20。在一些实施方案中,靶序列和gRNA的向导序列可以包含1-6个错配,其中向导序列为约20核酸。在一些实施方案中,靶序

列和gRNA的向导序列可以包含1或2个错配,其中向导序列为约20核酸。

[0156] 靶序列的长度可取决于所使用的核酸酶系统。例如,CRISPR/Cas系统的靶序列可以包含5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、35、40、45、50或超过50个核苷酸。在一些实施方案中,靶序列可以包含18-24个核苷酸。在一些实施方案中,靶序列可以包含19-21个核苷酸。在一些实施方案中,靶序列可以包含20个核苷酸。在使用切口酶时,靶序列可以在DNA分子的相反链上包含一对由一对切口酶识别的靶序列。

[0157] 用于COL8A2的靶序列

[0158] 本发明的组合物可以指导至COL8A2基因中的靶序列。例如,所提供的核酸酶可以识别并切割COL8A2靶序列。在一些实施方案中,向导RNA可以指导Cas蛋白质至COL8A2的靶序列,其中向导RNA的向导序列与靶序列杂交,Cas蛋白质切割靶序列。

[0159] 在一些实施方案中,根据COL8A2基因中的靶序列确定该一个或多个向导RNA的选择。在一些实施方案中,该一个或多个向导RNA的crRNA序列与选自SEQ ID NO:191-1084的靶序列互补和杂交。

[0160] 在一些实施方案中,根据不具有已知导致胶原VIII $\alpha$ 亚基(COL8A2)的异常功能的突变的野生型COL8A2基因中的靶序列确定该一个或多个向导RNA的选择。在一些实施方案中,该一个或多个向导RNA的crRNA序列与选自SEQ ID NO:191-1063的靶序列互补和杂交。

[0161] 在一些实施方案中,根据对应于COL8A2蛋白质中的Gln455Lys突变(由c.1364C>A核苷酸改变引起)的COL8A2基因中的靶序列确定该一个或多个向导RNA的选择。在一些实施方案中,该一个或多个向导RNA的crRNA序列与选自SEQ ID NO:1064-1069的靶序列互补和杂交。

[0162] 在一些实施方案中,根据对应于COL8A2蛋白质中的Gln455Val突变(由c.1363-1364CA>GT核苷酸改变引起)的COL8A2基因中的靶序列确定该一个或多个向导RNA的选择。在一些实施方案中,该一个或多个向导RNA的crRNA序列与选自SEQ ID NO:1070-1075的靶序列互补和杂交。

[0163] 在一些实施方案中,根据对应于COL8A2蛋白质中的Leu450Trp突变(由c.1349T>G核苷酸改变引起)的COL8A2基因中的靶序列确定该一个或多个向导RNA的选择。在一些实施方案中,该一个或多个向导RNA的crRNA序列与选自SEQ ID NO:1076-1084的靶序列互补和杂交。

[0164] 在一些实施方案中,该靶序列可以与向导RNA的向导序列互补。在一些实施方案中,向导RNA的向导序列和它对应的靶序列之间的互补程度或同一性可以是约50%、55%、60%、65%、70%、75%、80%、85%、90%、95%、96%、97%、98%、99%或100%。在一些实施方案中,靶序列和gRNA的向导序列可以100%互补或同一。在其他实施方案中,靶序列和gRNA的向导序列可以包含至少一个错配。例如,靶序列和gRNA的向导序列可以包含1、2、3、4、5、6、7、8、9或10个错配,其中向导序列的总长度为约20。在一些实施方案中,靶序列和gRNA的向导序列可以包含1-6个错配,其中向导序列为约20核酸。在一些实施方案中,靶序列和gRNA的向导序列可以包含1或2个错配,其中向导序列为约20核酸。

[0165] 靶序列的长度可取决于所使用的核酸酶系统。例如,CRISPR/Cas系统的靶序列可以包含5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、

30、35、40、45、50或超过50个核苷酸。在一些实施方案中,靶序列可以包含18-24个核苷酸。在一些实施方案中,靶序列可以包含19-21个核苷酸。在一些实施方案中,靶序列可以包含20个核苷酸。靶序列可以包含PAM。在使用切口酶时,靶序列可以在DNA分子的相反链上包含一对由一对切口酶识别的靶序列。

#### [0166] 载体

[0167] 在本发明的某些实施方案中,该组合物包含编码本文所述的任意向导RNA的DNA载体。在一些实施方案中,除向导RNA序列外,该载体进一步包含不编码向导RNA的核酸。不编码向导RNA的核酸包括但不限于启动子、增强子、调节序列和编码核酸酶如Cas9的核酸。在一些实施方案中,该载体包含编码crRNA、trRNA或crRNA和trRNA的核苷酸序列。在一些实施方案中,该编码crRNA、trRNA或crRNA和trRNA的核苷酸序列包含或由以下序列组成:侧翼为来自天然存在的CRISPR/Cas系统的全部或部分重复序列的向导序列。包含crRNA、trRNA或crRNA和trRNA的核酸,或由crRNA、trRNA或crRNA和trRNA的核酸组成的核酸可以进一步包含载体序列,其中该载体序列包含或由以下序列组成:不与该crRNA、trRNA或crRNA和trRNA天然见于一起的核酸。

[0168] 在一些实施方案中,该crRNA和trRNA由一个载体内的非相邻核酸编码。在一些实施方案中,该crRNA和trRNA可以由相邻核酸编码。在一些实施方案中,该crRNA和trRNA由单个核酸的相反链编码。在其他实施方案中,该crRNA和trRNA由单个核酸的同一条链编码。在一些实施方案中,该载体编码一个或多个sgRNA。在其他实施方案中,该载体编码两个或多个sgRNA。

#### [0169] 核酸酶

[0170] 在一些实施方案中,除该至少一个gRNA外,该组合物进一步包含核酸酶。在一些实施方案中,该gRNA与核酸酶一起称为核糖核蛋白复合物(RNP)。在一些实施方案中,该核酸酶是Cas蛋白质。在一些实施方案中,该gRNA与Cas蛋白质一起称为Cas RNP。在一些实施方案中,该Cas包含I型、II型或III型成分。在一些实施方案中,该Cas蛋白质来自I型CRISPR/Cas系统。在一些实施方案中,该Cas蛋白质来自II型CRISPR/Cas系统。在一些实施方案中,该Cas蛋白质来自III型CRISPR/Cas系统。在一些实施方案中,该Cas蛋白质是Cas9。在一些实施方案中,该Cas蛋白质是Cpf1。在一些实施方案中,该Cas蛋白质是来自II型CRISPR/Cas系统的Cas9蛋白质。在一些实施方案中,该gRNA与Cas9一起称为Cas9RNP。

[0171] 在涵盖Cas核酸酶的实施方案中,该Cas核酸酶可以来自IIA型、IIB型或IIC型系统。Cas核酸酶或其他RNP成分可以源自的非限制性示例性物种包括:酿脓链球菌、嗜热链球菌(*Streptococcus thermophilus*)、链球菌属物种(*Streptococcus sp.*)、金黄色葡萄球菌(*Staphylococcus aureus*)、无害利斯特氏菌(*Listeria innocua*)、戈氏乳杆菌(*Lactobacillus gasseri*)、新凶手巴斯德氏菌(*Francisella novicida*)、产琥珀酸沃林氏菌(*Wolinella succinogenes*)、*Sutterella wadsworthensis*、*Gammaproteobacterium*、脑膜炎奈氏球菌(*Neisseria meningitidis*)、空肠弯曲杆菌(*Campylobacter jejuni*)、多杀巴斯德氏菌(*Pasteurella multocida*)、产琥珀酸丝状杆菌(*Fibrobacter succinogene*)、深红红螺菌(*Rhodospirillum rubrum*)、达松维尔拟诺卡氏菌(*Nocardiosis 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、脑膜炎奈氏球菌 (*Alicyclobacillus acidocaldarius*)、假真菌样芽孢杆菌 (*Bacillus pseudomycooides*)、*Bacillus selenitireducens*、*Exiguobacterium sibiricum*、德氏乳杆菌 (*Lactobacillus delbrueckii*)、唾液乳杆菌 (*Lactobacillus salivarius*)、布氏乳杆菌 (*Lactobacillus buchneri*)、齿垢密螺旋体 (*Treponema denticola*)、海洋微颤菌 (*Microscilla marina*)、*Burkholderiales bacterium*、*Polaromonas naphthalenivorans*、极地单胞菌属物种 (*Polaromonas sp.*)、*Crocospaera watsonii*、蓝丝菌属物种 (*Cyanothece sp.*)、铜绿微囊蓝细菌 (*Microcystis aeruginosa*)、聚球蓝细菌属物种 (*Synechococcus sp.*)、阿拉伯糖醋盐杆菌 (*Acetohalobium arabaticum*)、*Ammonifex degensii*、*Caldicelulosiruptor becscii*、*Candidatus Desulforudis*、肉毒梭状芽孢杆菌 (*Clostridium botulinum*)、艰难梭菌 (*Clostridium difficile*)、*Finegoldia magna*、*Natranaerobius thermophilus*、*Pelotomaculum thermopropionium*、*Acidithiobacillus caldus*、*Acidithiobacillus ferrooxidans*、*Allochromatium vinosum*、海杆菌属物种 (*Marinobacter sp.*)、嗜盐亚硝化球菌 (*Nitrosococcus halophilus*)、*Nitrosococcus watsoni*、*Pseudoalteromonas haloplanktis*、*Ktedonobacter racemifer*、*Methanohalobium evestigatum*、多变鱼腥蓝细菌 (*Anabaena variabilis*)、产泡沫节球蓝细菌 (*Nodularia spumigena*)、念珠蓝细菌属物种 (*Nostoc sp.*)、最大节螺菌 (*Arthrospira maxima*)、盘状节螺蓝细菌 (*Arthrospira platensis*)、节螺菌属物种 (*Arthrospira sp.*)、鞘丝蓝细菌属物种 (*Lyngbya sp.*)、土质体微鞘蓝细菌 (*Microcoleus chthonoplastes*)、颤蓝细菌属物种 (*Oscillatoria sp.*)、*Petrotoga mobilis*、非洲栖热腔菌 (*Thermosiphon africanus*)、*Streptococcus pasteurianus*、灰烬奈瑟氏菌 (*Neisseria cinerea*)、红嘴鸥弯曲杆菌 (*Campylobacter lari*)、*Parvibaculum lavamentivorans*、白喉棒杆菌 (*Corynebacterium diphtheria*)、氨基酸球菌属物种 (*Acidaminococcus sp.*)、毛螺菌科 (*Lachnospiraceae*) 细菌ND2006和 *A Caryochloris marina*。在一些实施方案中,该Cas核酸酶是来自酿脓链球菌的Cas9蛋白质。在一些实施方案中,该Cas核酸酶是来自嗜热链球菌的Cas9蛋白质。在一些实施方案中,该Cas核酸酶是来自脑膜炎奈氏球菌的Cas9蛋白质。一些实施方案中,该Cas核酸酶是来自金黄色葡萄球菌的Cas9蛋白质。在一些实施方案中,该Cas核酸酶是来自新凶手巴斯德氏菌的Cpf1蛋白质。在一些实施方案中,该Cas核酸酶是来自氨基酸球菌属物种的Cpf1蛋白质。在一些实施方案中,该Cas核酸酶是来自Lachnospiraceae细菌ND2006的Cpf1蛋白质。

[0172] 野生型Cas9具有两个核酸酶结构域:RuvC和HNH。RuvC结构域切割非靶DNA链,HNH结构域切割靶DNA链。在一些实施方案中,该Cas9蛋白质包含超过一个RuvC结构域和/或超过一个HNH结构域。在一些实施方案中,该Cas9蛋白质是野生型Cas9。在每个组合物和方法实施方案中,该Cas在靶DNA中诱导双链断裂。

[0173] 具有一个失活的催化结构域 (RuvC或HNH) 的Cas9修饰形式称为“切口酶”。切口酶仅切割靶DNA上的一条链,从而产生单链断裂。单链断裂也可以称为“切口”。在一些实施方案中,该组合物和方法包含切口酶。在一些实施方案中,该组合物和方法包含在靶DNA中诱导切口而不是取链断裂的切口酶Cas9。

[0174] 在一些实施方案中,Cas蛋白质可以修饰为仅包含一个功能性核酸酶结构域。例如,可以这样修饰Cas蛋白质,使得核酸酶结构域之一突变或完全或部分缺失,以降低其核

酸切割活性。在一些实施方案中,使用具有活性降低的RuvC结构域的切口酶Cas。在一些实施方案中,使用具有失活的RuvC结构域的切口酶Cas。在一些实施方案中,使用具有活性降低的HNH结构域的切口酶Cas。在一些实施方案中,使用具有失活的HNH结构域的切口酶Cas。

[0175] 在一些实施方案中,取代Cas蛋白质核酸酶结构域内的保守氨基酸以降低或改变核酸酶活性。在一些实施方案中,Cas蛋白质可以在RuvC或RuvC样核酸酶结构域中包含氨基酸取代。RuvC或RuvC样核酸酶结构域中的示例性氨基酸取代包括D10A(基于酿脓链球菌Cas9蛋白质)。在一些实施方案中,Cas蛋白质可以在HNH或HNH样核酸酶结构域中包含氨基酸取代。HNH或HNH样核酸酶结构域中的示例性氨基酸取代包括E762A、H840A、N863A、H983A和D986A(基于酿脓链球菌Cas9蛋白质)。

[0176] 在一些实施方案中,该组合物包含切口酶和一对向导RNA。在一些实施方案中,该对向导RNA分别与靶序列的有义和反义链互补。在此实施方案中,该向导RNA指导切口酶至靶序列,并通过在靶序列的相反链上产生切口(即双切口)来引入DSB。在一些实施方案中,双切口的使用可以改善特异性,降低脱靶效应。在一些实施方案中,将切口酶Cas与两个分开的靶向DNA的相反链的向导RNA一起使用,以在靶DNA中产生双切口。在一些实施方案中,将切口酶Cas与选择为密切靠近的两个分开的向导RNA一起使用,以在靶DNA中产生双切口。

[0177] 在一些实施方案中,使用嵌合Cas蛋白质,其中用不同蛋白质的部分替换蛋白质的一个结构域或区域。在一些实施方案中,用来自不同核酸酶如Fok1的结构域替换Cas核酸酶结构域。在一些实施方案中,Cas蛋白质可以是修饰核酸酶。

[0178] 在一些实施方案中,使用Cas9-脱氨酶融合,其中Cas9不能切割双链DNA(dCas9)。术语“脱氨酶”指催化脱氨反应的酶。在一些实施方案中,该脱氨酶是胞嘧啶脱氨酶,其将胞嘧啶(C)转化为尿嘧啶(U),然后尿嘧啶由细胞转化为胸腺嘧啶(T)。在一些实施方案中,该脱氨酶是鸟嘌呤脱氨酶,其将鸟嘌呤(G)转化为黄嘌呤,然后黄嘌呤由细胞转化为腺嘌呤(A)。在一些实施方案中,该脱氨酶是APOBEC 1家族脱氨酶、激活诱导胞嘧啶脱氨酶(AID)和腺苷脱氨酶,如ADAT家族脱氨酶或作用于RNA的腺苷脱氨酶(ADAR),其将腺嘌呤(A)转化为次黄嘌呤,然后次黄嘌呤由细胞转化为鸟嘌呤(G)。

[0179] 在其他实施方案中,该Cas蛋白质可以来自I型CRISPR/Cas系统。在一些实施方案中,该Cas蛋白质可以是I型CRISPR/Cas系统的级联复合物的成分。在一些实施方案中,该Cas蛋白质可以是Cas3蛋白质。在一些实施方案中,该Cas蛋白质可以来自III型CRISPR/Cas系统。在一些实施方案中,该Cas蛋白质可以具有RNA切割活性。

[0180] PAM

[0181] 在一些实施方案中,靶序列可以与PAM相邻。在一些实施方案中,PAM可以与靶序列3'端相邻或在靶序列3'端的1、2、3或4个核苷酸之内。PAM的长度和序列可取决于所使用的Cas蛋白质。例如,PAM可以选自具体Cas9蛋白质或Cas9直向同源物的共有或具体PAM序列,包括Ran等,Nature,520:186-191(2015)的图1中公开的那些,在此引入作为参考。在一些实施方案中,PAM可以包含长度2、3、4、5、6、7、8、9或10个核苷酸。非限制性示例性PAM序列包括NGG、NAG、NGA、NGAG、NGCG、NNGRRT、TTN、NGGNG、NG、NAAAAN、NNAAAAW、NNNNACA、GNNNCNNA和NNNNGATT(其中N定义为任意核苷酸,W定义为A或T,R定义为A或G)。在一些实施方案中,PAM序列可以是NGG。在一些实施方案中,PAM序列可以是NGGNG。在一些实施方案中,PAM序列可以是NNAAAAW。

**[0182] 切除TNR的方法**

**[0183]** TCF4中的TNR与FECD风险提高相关。此外,TCF4中的突变与精神分裂症和PSC相关。向导RNA与Cas蛋白质(或编码Cas蛋白质的核酸)一起递送可以用作这些障碍的治疗,例如通过从TCF4基因切除TNR(或其部分)。因此,本文提供的某些实施方案涉及从TCF4切除TNR的方法。在一些实施方案中,该方法包括递送本文提供CRISPR/Cas组合物中的任一种至细胞,该组合物包含指导核酸酶至本文表2中提供的靶序列的一个或多个gRNA。在一些实施方案中,该方法包括将两个gRNA与Cas蛋白质(或编码Cas蛋白质的核酸)一起递送至细胞,其中第一gRNA包含靶向TNR的5'区域且选自SEQ ID NO:1089-1181的向导序列,第二gRNA包含靶向TNR的3'区域且选自SEQ ID NO:1182-1278的向导序列。在一些实施方案中,该细胞是人细胞,例如人角膜内皮细胞。在一些实施方案中,该方法产生这样的细胞群体,其中群体的某部分从TCF4基因切除了TNR。在一些实施方案中,群体内至少5%、至少10%、至少15%、至少20%、至少25%、至少30%、至少35%、至少40%、至少45%、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%或至少95%或更多的细胞从TCF4基因切除了TNR。用于测量细胞群体内的切除百分比的方法是已知的,且包括本文中提供的那些,例如下一代测序(NGS)法,例如其中切除百分比定义为包含TNR缺失的测序读出数除以与靶区域重叠的读出总数。

**[0184]** CRISPR/Cas系统的使用可以在DNA中产生双链断裂,或者在使用切口酶时在DNA中产生单链断裂。

**[0185]** NHEJ是通过断裂末端的重新连接来修复DNA中的双链断裂(DSB)的过程,其可以产生插入/缺失突变形式的错误。NHEJ因此可以是敲除具体基因产物或降低具体基因产物的水平的手段,因为发生在编码外显子内的插入/缺失可以产生移码突变和提前终止密码子。

**[0186]** HR和HDR是可用于在外源引入的修复模板存在下在靶基因座处产生精确、确定的修饰的主要旁路DNA修复途径。这可以用来矫正单碱基改变、缺失、插入、倒位和其他突变。在一些情况下,使用这样的修复模板,该模板在DNA内引入防止被用于起始修复过程的CRISPR核酸酶识别的沉默(即同义)核苷酸改变,从而防止在所矫正的基因内形成插入/缺失。

**[0187]** 在一些实施方案中,该模板可以用于HR,例如以修饰诸如TCF4和/或COL8A2的靶基因。在一些实施方案中,该HR可以导致模板序列或部分模板序列整合入靶核酸分子。在一些实施方案中,可以提供单模板。在其他实施方案中,可以提供两个或多个模板,使得可以在两个或多个靶位点发生HR。例如,可以提供不同模板来修复细胞中的单个基因或细胞中的两个不同基因。在一些实施方案中,为细胞提供多个拷贝的至少一个模板。在一些实施方案中,该不同模板可以按独立的拷贝数或独立的量提供。

**[0188]** 在其他实施方案中,该模板可以用于HDR,例如以修饰诸如TCF4和/或COL8A2的靶基因。HDR涉及核酸中切割位点处的DNA链侵入。在一些实施方案中,该HDR可以导致在所编辑的靶核酸分子中包含模板序列。在一些实施方案中,可以提供单模板。在其他实施方案中,HDR可以在两个或多个位点处使用两个或多个具有不同序列的模板。例如,可以提供不同模板来修复细胞中的单个基因或细胞中的两个不同基因。在一些实施方案中,为细胞提供多个拷贝的至少一个模板。在一些实施方案中,该不同模板可以按独立的拷贝数或独立的量提供。

[0189] 还在其他实施方案中,该模板可以用于NHEJ介导的基因编辑,例如以修饰诸如TCF4和/或COL8A2的靶基因。在一些实施方案中,该模板序列与切割位点附近的核酸序列无相似性。在一些实施方案中,掺入该模板或部分模板序列。在一些实施方案中,可以提供单模板。在其他实施方案中,NHEJ可以在两个或多个位点处插入两个或多个具有不同序列的模板。例如,可以提供不同模板来在细胞中插入单个模板或在细胞中插入两个不同模板。在一些实施方案中,该不同模板可以按独立的拷贝数提供。在一些实施方案中,该模板包含侧翼反向末端重复(ITR)序列。

[0190] 模板可以具有任何适宜的长度。在一些实施方案中,模板可以包含长度10、15、20、25、50、75、100、150、200、500、1000、1500、2000、2500、3000、3500、4000、4500、5000、5500、6000或更多个核苷酸。模板可以是单链核酸。模板可以是双链或部分双链核酸。在某些实施方案中,单链模板长度为20、30、40、50、75、100、125、150、175或200个核苷酸。在一些实施方案中,模板可以包含与靶核酸分子的含有靶序列的部分互补的核苷酸序列(即“同源臂”)。在一些实施方案中,模板可以包含与靶核酸位子上位于切割位点上游或下游的序列互补的同源臂。在一些实施方案中,模板可以包含分别与位于切割位点上游和下游的序列互补的第一同源臂和第二同源臂(也称为第一和第二核苷酸序列)。在模板包含两个同源臂时,每个臂可以是相同长度或不同长度,同源臂之间的序列可以与同源臂之间的靶序列基本相似或相同,或者它可以完全不相关。在一些实施方案中,模板上的第一核苷酸序列和切割位点上游序列之间以及模板上的第二核苷酸序列和切割位点下游序列之间的互补程度可允许模板和靶核酸分子之间的同源重组,如高保真同源重组。在一些实施方案中,互补程度可以是约50%、55%、60%、65%、70%、75%、80%、85%、90%、95%、97%、98%、99%或100%。在一些实施方案中,互补程度可以是约95%、97%、98%、99%或100%。在一些实施方案中,互补程度可以是至少98%、99%或100%。在一些实施方案中,互补程度可以是100%。

[0191] 在一些实施方案中,模板包含含有侧翼反向末端重复(ITR)序列的ssDNA或dsDNA。在一些实施方案中,模板作为质粒、小环、纳米环或PCR产物提供。

[0192] 切除片段

[0193] 切除片段的产生意指利用CRISPR技术的力量来通过使用与这些靶序列互补的两个向导RNA精确去除两个靶序列之间的小的DNA区域。在一些实施方案中,该两个向导RNA将核酸酶靶向至相隔约3000、2500、2000、1500、1000、500、400、300、200、150、100、50或30个核苷酸的序列,导致切除靶序列之间的DNA片段。

[0194] 用CRISPR/Cas组合物治疗FECD

[0195] 可以对个体施用本文所述的任意组合物来在具有导致FECD风险提高的遗传突变的个体中治疗FECD。

[0196] 可以对个体施用本文所述的任意组合物来在TCF4内含子3中具有TNR扩增的个体中的治疗FECD。涵盖包括施用本文所述的任意组合物的治疗FECD的方法。在一些方面,该组合物按治疗有效量施用。在一些实施方案中,涵盖切除TCF4的TNR、突变TCF4的TNR、减少TCF4的TNR的拷贝数、改善TCF4的TNR和/或消除TCF4的TNR的方法,其包括施用本文所述的一种或多种组合物。在一些实施方案中,提供在个体中切割每个细胞的一个或两个TCF4拷贝的TNR、减少每个细胞的一个或两个TCF4拷贝的TNR的拷贝数、改善每个细胞的一个或两个TCF4拷贝的TNR和/或消除每个细胞的一个或两个TCF4拷贝的TNR的方法,其包括施用本

文所述的一种或多种组合物。在一些实施方案中,该细胞是角膜内皮细胞。

[0197] 在一些方面,涵盖降低、抑制或改善TCF4的RNA毒性的方法,其包括施用一种或多种本文所述的组合物。在一些实施方案中,涵盖抑制RNA毒性的方法,其包括施用一种或多种本文所述的组合物,其中治疗后TCF4的毒性RNA产物的水平不回到施用前的水平,恢复对角膜内皮细胞的正常功能,并防止细胞死亡。

[0198] 在一些实施方案中,治疗可以是使用递送入眼前房的包含一个或多个适当的向导的载体和/或脂质纳米颗粒。在一些实施方案中,治疗可以是使用递送入眼后房的包含一个或多个适当的向导的载体和/或脂质纳米颗粒。在一些实施方案中,治疗可以是使用递送入角膜本身的包含一个或多个适当的向导的载体和/或脂质纳米颗粒。在一些实施方案中,治疗可以是使用递送入角膜基质的包含一个或多个适当的向导的载体和/或脂质纳米颗粒。在一些实施方案中,治疗可以是使用递送入角膜缘的包含一个或多个适当的向导的载体和/或脂质纳米颗粒。在一些实施方案中,治疗可以是使用局部递送至角膜上皮表面的包含一个或多个适当的向导的载体和/或脂质纳米颗粒。在本段落的任意前述实施方案以及本文所述的其他实施方案中,治疗进一步包括例如用脂质纳米颗粒递送Cas蛋白质(例如Cas9),或用载体和/或脂质纳米颗粒递送编码Cas蛋白质的核酸。在一些实施方案(例如使用脂质纳米颗粒的那些)中,编码Cas蛋白质的核酸是mRNA。在一些实施方案中,通过与用于递送该一个或多个适当的向导的载体和/或脂质纳米颗粒相同的载体和/或脂质纳米颗粒来递送Cas蛋白质或编码Cas蛋白质的核酸。在一些实施方案中,通过与用于递送该一个或多个适当的向导的载体和/或脂质纳米颗粒不同的载体和/或脂质纳米颗粒来递送Cas蛋白质或编码Cas蛋白质的核酸。

[0199] 在一些实施方案中,本发明的CRISPR组合物的单次使用可足以矫正与疾病相关的潜在遗传缺陷或突变。在其他实施方案中,超过一次施用CRISPR治疗剂可以是有益的,以通过累积效应最大化编辑所有靶细胞和所有等位基因。

[0200] 涵盖本文所述组合物在制备用于治疗FECD的药物中的用途。在一些实施方案中,在启动本发明的组合物治疗之前针对TCF4中的TNR筛查FECD、可能为FECD和/或提示FECD家族史的患者。在一些实施方案中,如果TCF4内含子3中存在50或更多个TNR,则在患者中启动治疗。

[0201] COL8A2中的突变与FECD和PPCD风险提高相关。可以对个体施用本文所述的任意组合物来在COL8A2中具有导致基因产物具有氨基酸突变的突变的个体中治疗FECD。在一些实施方案中,这些氨基酸突变是Gln455Lys、Gln455Val或Leu450Trp。

[0202] 涵盖治疗FECD的方法,其包括施用本文所述的任意组合物。在一些方面,该组合物按治疗有效量施用。在一些实施方案中,涵盖切割、突变、改善和/或消除COL8A2中的突变的方法,其包括施用本文所述的一种或多种组合物。在一些实施方案中,CRISPR/Cas组合物的使用与NHEJ的过程一起进行,导致插入/缺失的产生和COL8A2等位基因的丢失。在一些实施方案中,为了矫正导致胶原VIII $\alpha$ 2亚基中氨基酸突变的核酸突变的目的,CRISPR/Cas组合物的使用与使用外源HR/HDR模板或使用正常内源等位基因作为HR/HDR模板一起进行。在一些实施方案中,所矫正的COL8A2基因中的突变是由c.1364C>A核苷酸改变引起的Gln455Lys突变。在一些实施方案中,所矫正的COL8A2基因中的突变是由c.1363-1364CA>GT核苷酸改变引起的Gln455Val突变。在一些实施方案中,所矫正的COL8A2基因中的突变是

由c.1349T>G核苷酸改变引起的Leu450Trp突变。在一些实施方案中,模板与Cas RNP一起使用导致核酸序列的矫正,使得突变不再存在。在一些实施方案中,该细胞是角膜内皮细胞。

[0203] 在一些方面,涵盖减少、抑制或改善由突变COL8A2形成的异常胶原的方法,其包括施用一种或多种本文所述的组合物。在一些实施方案中,涵盖抑制异常胶原VIII $\alpha$ 亚基(COL8A2)产生的方法,其包括施用一种或多种本文所述的组合物,其中异常COL8A2的水平在治疗后不回到施用前的水平。在一些实施方案中,涵盖用HR或HDR矫正遗传突变使得仅产生正常胶原的方法,其包括施用一种或多种本文所述的组合物。突变形式的胶原的减少或矫正应阻止见于FECD患者角膜中的异常胶原沉积。

[0204] 涵盖本文所述组合物在制备用于治疗FECD的药物中的用途。在一些实施方案中,在启动本发明的组合物治疗之前针对COL8A2中的突变筛查FECD、可能为FECD和/或提示FECD家族史的患者。在一些实施方案中,在启动本发明的组合物治疗之前针对COL8A2中的突变筛查PPCD、可能为PPCD和/或提示PPCD家族史的患者。在一些实施方案中,如果存在突变(如c.1364C>A核苷酸改变引起的Gln455Lys突变、c.1363-1364CA>GT核苷酸改变引起的Gln455Val突变、或c.1349T>G核苷酸改变引起的Leu450Trp突变),则在患者中启动治疗。

[0205] 在一些实施方案中,本发明的CRISPR组合物的单次使用可足以矫正与疾病相关的潜在遗传缺陷或突变。在其他实施方案中,超过一次施用CRISPR治疗剂可以是有益的,以通过累积效应最大化编辑所有靶细胞和所有等位基因。在一些实施方案中,本发明的组合物的治疗功效在递送后1年、2年、3年、4年、5年或10天可见。

[0206] 可以用许多不同类型的评估来确定对FECD的治疗功效,参见Eghrari和Gottsch, *Expert Rev Ophthalmol.* 5(2):147-159 (2010)。在一些实施方案中,组合物的治疗功效基于随时间推移通过裂隙灯显微镜检查进行的评估。在一些实施方案中,组合物的治疗功效基于通过角膜厚度测量法测量角膜厚度随时间的变化来定量测量疾病进展。在一些实施方案中,组合物的治疗功效基于角膜厚度测量中角膜随时间推移的改善、稳定化或减慢改变。

[0207] 在一些实施方案中,组合物的治疗功效基于随时间推移评估视敏度。在一些实施方案中,组合物的治疗功效基于视敏度随时间推移的改善、稳定化或下降减慢。

[0208] 在一些实施方案中,组合物的治疗功效基于角膜内皮显微术(specular microscopy)。在一些实施方案中,用此角膜内皮显微术来记录水滴的存在。在一些实施方案中,组合物的治疗功效基于新水滴形成的减少。在一些实施方案中,组合物的治疗功效基于已有水滴的存在的减少。

[0209] 在一些实施方案中,组合物的治疗功效基于患者保持可接受的视敏度,避免了进行角膜移植的需要。在一些实施方案中,组合物的治疗功效基于需要进行角膜移植的时间的延迟。此角膜移植可以是全角膜移植或角膜内层移植。

[0210] 除与FECD相关外,TCF4基因中的遗传变体与另外两种病症相关:原发性硬化性胆管炎(PSC)和精神分裂症(参见Forrest MP等, *Trends Mol Med.* 2014 Jun; 20(6):322-31)。尚不清楚TCF4基因中的非编码变体如何提高PSC和精神分裂症的风险。一种可能性是,这些变体作为内含子3内同一TNR区中的共遗传扩增的标志,其与FECD中RNA介导的毒性关联。虽然此假设仍未得到证明,但与PSC和精神分裂症相关的变体在物理上和单体型上靠近内含

子3中含有TNR的区域,提示这些邻近区域中的变体的共遗传。此外,与PSC和精神分裂症相关的风险变体与TCF4基因的表达变化不相关,提示涉及另一机制,如在内含子3中具有TNR扩增的患者中看到的RNA毒性。

[0211] 联合治疗

[0212] 在一些实施方案中,本发明的组合物作用于治疗FECD、PPCD、PSC和/或精神分裂症的单一活性剂。

[0213] 在一些实施方案中,本发明的组合物与其他用于FECD、PPCD、PSC和/或精神分裂症的治疗联合。在一些实施方案中,该联合治疗是软性隐形眼镜。在一些实施方案中,这些软性隐形眼镜弄平眼睛表面的微观肿胀。在一些实施方案中,本发明的组合物与从角膜抽出流体的滴眼液或眼膏联合使用。在一些实施方案中,这些滴眼液或眼膏是Muro 128<sup>®</sup> 5% (氯化钠高渗眼用溶液,5%,Bausch and Lomb)、Muro 128 5%软膏(氯化钠高渗眼用软膏,5%)(Bausch and Lomb)、或其他盐溶液或眼泪替代物。

[0214] 在一些实施方案中,将糖皮质激素或皮质类固醇与本发明的组合物一起用于减少对治疗剂的免疫反应。

[0215] 联合治疗可以通过治疗的单个成分的同时、顺次或分开给药的方式来达到。在顺次或分开施用,施用第二成分的延迟不应使得丧失联合的有益作用。

[0216] CRISPR/Cas组合物的递送

[0217] 在一些实施方案中,本文所述CRISPR/Cas组合物可以通过包含一个或多个适当的向导的载体和/或脂质纳米颗粒施用。

[0218] 病毒载体

[0219] CRISPR/Cas组合物可以通过载体系统递送。在一些实施方案中,CRISPR/Cas组合物可以提供在一个或多个载体上。在一些实施方案中,该载体可以是DNA载体。在其他实施方案中,该载体可以是RNA载体。在一些实施方案中,该载体可以是环状载体。在其他实施方案中,该载体可以是线性载体。在一些实施方案中,该载体可以包封在脂质纳米颗粒、脂质体、非脂质纳米颗粒或病毒壳体中。非限制性示例性载体包括质粒、噬菌粒、黏粒、人工染色体、微型染色体、转座子、病毒载体和表达载体。

[0220] 在一些实施方案中,该载体可以是病毒载体。在一些实施方案中,该病毒载体可以从其野生型对应物遗传修饰。例如,该病毒载体可以包含一个或多个核苷酸的插入、缺失或取代,以便于克隆或使得改变该载体的一种或多种特性。这类特性可以包括包装容量、转导效率、免疫原性、基因组整合、复制、转录和翻译。在一些实施方案中,可以缺失部分病毒基因组,使得病毒能够包装具有更大尺寸的外源序列。在一些实施方案中,该病毒载体可以具有增强的转导效率。在一些实施方案中,可以减少病毒在宿主中诱导的免疫反应。在一些实施方案中,可以突变促进病毒序列整合入宿主基因组的病毒基因(例如整合酶),使得病毒成为非整合型。在一些实施方案中,该病毒载体可以是复制缺陷型。在一些实施方案中,该病毒载体可以包含外源转录或翻译控制序列来驱动载体上的编码序列表达。在一些实施方案中,该病毒可以是辅助病毒(helper-dependent)依赖型。例如,该病毒可能需要一种或多种辅助病毒来提供将载体扩增并包装入病毒颗粒所需的病毒成分(如病毒蛋白质)。在这种情况下,可将一个或多个辅助成分(包括一个或多个编码病毒成分的载体)连同本文所述的载体系统引入宿主细胞。在其他实施方案中,该病毒可以是非辅助病毒依赖型。例如,该病毒

可以在无任何辅助病毒的情况下扩增和包装载体。在一些实施方案中,本文所述的载体系统也可以编码病毒扩增和包装所需的病毒成分。

[0221] 非限制性实例性病毒载体包括腺相关病毒(AAV)载体、慢病毒载体、腺病毒载体、辅助病毒依赖型腺病毒载体(HDAd)、单纯疱疹病毒(HSV-1)载体、噬菌体T4、杆状病毒载体和反转录病毒载体。在一些实施方案中,该病毒载体可以是AAV载体。在一些实施方案中,该AAV载体具有血清型2、3、5、7、8、9或rh.10。在其他实施方案中,该病毒载体可以是慢病毒载体。在一些实施方案中,该慢病毒载体可以是非整合型。

[0222] 在一些实施方案中,该病毒载体可以是腺病毒载体。在一些实施方案中,该腺病毒可以是高克隆容量或“无病毒基因(gutless)”腺病毒,其中从病毒缺失了远离5'和3'反向末端重复(ITR)和包装信号(I)的所有病毒编码区,以提高其包装容量。还在其他实施方案中,该病毒载体可以是HSV-1载体。在一些实施方案中,基于HSV-1的载体是辅助病毒依赖型,在其他实施方案中,它是非辅助病毒依赖型。例如,仅保留包装序列的扩增子载体需要具有用于包装的结构成分的辅助病毒,而去除了非必需病毒功能的30kb缺失HSV-1载体无需辅助病毒。在其他实施方案中,该病毒载体可以是噬菌体T4。在一些实施方案中,在清空病毒头部时,该噬菌体T4可以能够包装任何线性或环状DNA或RNA分子。在其他实施方案中,该病毒载体可以是杆状病毒载体。还在其他实施方案中,该病毒载体可以是反转录病毒载体。在使用具有较小克隆容量的AAV或慢病毒载体的实施方案中,可以有必要用超过一个载体来递送本文公开的载体系统的所有成分。例如,一个AAV载体可以包含编码Cas蛋白质的序列,而第二AAV载体可以包含一个或多个向导序列。但是,在一些实施方案中,单个AAV载体可以包含编码Cas蛋白质和一个或多个向导序列的序列。在一些涉及用单个AAV来递送本文所述的CRISPR/Cas成分的实施方案中,使用小Cas9直向同源物。在一些实施方案中,该小Cas9直向同源物源自脑膜炎奈氏球菌、空肠弯曲杆菌或金黄色葡萄球菌。

[0223] 在一些实施方案中,该载体可以能够驱动一个或多个编码序列在细胞中表达。在一些实施方案中,该细胞可以是原核细胞,例如细菌细胞。在一些实施方案中,该细胞可以是真核细胞,如酵母、植物、昆虫或哺乳动物细胞。在一些实施方案中,该真核细胞可以是哺乳动物细胞。在一些实施方案中,该真核细胞可以是啮齿动物细胞。在一些实施方案中,该真核细胞可以是人细胞。适于在不同类型细胞中驱动表达的启动子为本领域已知。在一些实施方案中,该启动子可以是野生型启动子。在其他实施方案中,该启动子可以针对更有效率或更有效果的表达进行修饰。还在其他实施方案中,该启动子可以是截短的,但保持其功能。例如,该启动子可以具有适于将载体包装入病毒的正常大小或减小的大小。

[0224] 在一些实施方案中,该载体可以包含编码本文所述核酸酶的核苷酸序列。在一些实施方案中,该载体编码的核酸酶可以是Cas蛋白质。在一些实施方案中,该载体系统可以包含一个拷贝的编码核酸酶的核苷酸序列。在其他实施方案中,该载体系统可以包含超过一个拷贝的编码核酸酶的核苷酸序列。在一些实施方案中,该编码核酸酶的核苷酸序列可以与至少一个转录或翻译控制序列有效连接。在一些实施方案中,该编码核酸酶的核苷酸序列可以与至少一个启动子有效连接。

[0225] 在一些实施方案中,该启动子可以是组成型启动子、诱导型启动子或组织特异性启动子。在一些实施方案中,该启动子可以是组成型启动子。非限制性示例性组成型启动子包括巨细胞病毒立即早期启动子(CMV)、猿猴病毒(SV40)启动子、腺病毒主要晚期(MLP)启

动子、劳斯肉瘤病毒 (RSV) 启动子、小鼠乳腺肿瘤病毒 (MMTV) 启动子、磷酸甘油酸激酶 (PGK) 启动子、延伸因子- $\alpha$  (EF1a) 启动子、遍在蛋白启动子、肌动蛋白启动子、微管蛋白启动子、免疫球蛋白启动子、其功能性片段、或任意前述的组合。在一些实施方案中,该启动子可以是CMV启动子。在一些实施方案中,该启动子可以是截短的CMV启动子。在其他实施方案中,该启动子可以是EF1a启动子。在一些实施方案中,该启动子可以是诱导型启动子。非限制性示例性诱导型启动子包括可通过热休克、光、化学药品、肽、金属、类固醇、抗生素或醇诱导的那些。在一些实施方案中,该诱导型启动子可以是具有低基础(非诱导)表达水平的启动子,例如Tet-On<sup>®</sup>启动子(Clontech)。

[0226] 在一些实施方案中,该启动子可以是组织特异性启动子,例如在角膜内皮中特异性表达的启动子。

[0227] 该载体可以进一步包含编码本文所述向导RNA的核苷酸序列。在一些实施方案中,该载体包含一个拷贝的向导RNA。在其他实施方案中,该载体包含超过一个拷贝的向导RNA。在使用超过一个向导RNA的实施方案中,该向导RNA可以是不相同的,使得它们靶向不同的靶序列,或者可以是相同的,因为它们靶向同一靶序列。在一些载体包含超过一个向导RNA的实施方案中,各向导RNA可以具有其他不同的特性,如在Cas RNP复合体内的活性或稳定性。在一些实施方案中,该编码向导RNA的核苷酸序列可以与至少一个转录或翻译控制序列(如启动子、3'UTR或5'UTR)有效连接。在一个实施方案中,该启动子可以是tRNA启动子,例如tRNA<sup>Lys3</sup>或tRNA嵌合体。参见Mefferd等,RNA.201521:1683-9;Scherer等,Nucleic Acids Res.200735:2620-2628。在一些实施方案中,该启动子可由RNA聚合酶III(Pol III)识别。Pol III启动子的非限制性实例包括U6和H1启动子。在一些实施方案中,该编码向导RNA的核苷酸序列可以与小鼠或人U6启动子有效连接。在其他实施方案中,该编码向导RNA的核苷酸序列可以与小鼠或人H1启动子有效连接。在使用超过一个向导RNA的实施方案中,用于驱动表达的启动子可以相同或不同。在一些实施方案中,编码向导RNA的crRNA的核苷酸和编码向导RNA的trRNA的核苷酸可以在同一载体上提供。在一些实施方案中,编码crRNA的核苷酸和编码trRNA的核苷酸可以由同一启动子驱动。在一些实施方案中,crRNA和trRNA可以转录为单个转录物。例如,crRNA和trRNA可以从单个转录物加工形成双分子向导RNA。备选地,crRNA和trRNA可以转录为单分子向导RNA。在其他实施方案中,crRNA和trRNA可以由它们在同一载体上的对应启动子驱动。还在其他实施方案中,crRNA和trRNA可以由不同载体编码。

[0228] 在一些实施方案中,编码向导RNA的核苷酸序列可以定位在包含编码Cas蛋白质的核苷酸序列的同一载体上。在一些实施方案中,向导RNA和Cas蛋白质的表达可以由它们自己对应的启动子驱动。在一些实施方案中,向导RNA的表达可以由驱动Cas9蛋白质表达的同一启动子驱动。在一些实施方案中,向导RNA和Cas蛋白质转录物可以包含在单个转录物内。例如,向导RNA可以在Cas蛋白质转录物的非翻译区(UTR)内。在一些实施方案中,向导RNA可以在Cas蛋白质转录物的5'UTR内。在其他实施方案中,向导RNA可以在Cas蛋白质转录物的3'UTR内。在一些实施方案中,通过在其3'UTR内包含向导RNA从而缩短其3'UTR的长度,可以降低Cas蛋白质转录物的胞内半衰期。在其他实施方案中,向导RNA可以在Cas蛋白质转录物的内含子内。在一些实施方案中,可以在向导RNA定位于其内的内含子处添加适宜的剪接位点,使得从转录物正确剪接出向导RNA。在一些实施方案中,在同一载体上密切靠近的Cas蛋白质和向导RNA的表达可便于更有效地形成CRISPR RNP复合体。

[0229] 在一些实施方案中,该组合物包含载体系统,其中该系统包含超过一个载体。在一些实施方案中,该载体系统可以包含一个单载体。在其他实施方案中,该载体系统可以包含两个载体。在其他实施方案中,该载体系统可以包含三个载体。在用不同向导RNA进行多重化时,或在使用多个拷贝的向导RNA时,该载体系统可以包含超过三个载体。

[0230] 在一些实施方案中,该载体系统可以包含诱导型启动子,以仅在将它递送至靶细胞后开始表达。非限制性示例性诱导型启动子包括可通过热休克、光、化学药品、肽、金属、类固醇、抗生素或醇诱导的那些。在一些实施方案中,该诱导型启动子可以是具有低基础(非诱导)表达水平的启动子,例如Tet-On<sup>®</sup>启动子(Clontech)。

[0231] 在其他实施方案中,该载体系统可以包含组织特异性启动子,以仅在将它递送入特异性组织后开始表达。

[0232] 载体可以通过脂质体、纳米颗粒、外来体或微囊递送。载体也可以通过脂质纳米颗粒递送;参见例如PCT/US2017/024973,其于2017年3月30日提交,要求2016年3月30日提交的标题为“用于CRISPR/Cas成分的脂质纳米颗粒制剂”的U.S.S.N.62/315,602的优先权,其内容在此以其整体引入作为参考。

[0233] 在一些实施方案中,载体可以通过直接递送至角膜的溶液来递送。递送可以通过局部应用、注射入角膜本身、注射入前房、注射入后房、注射入角膜缘或其他手段来达到。

[0234] 在一些实施方案中,该载体可以全身递送。

[0235] 脂质纳米颗粒(LNP)

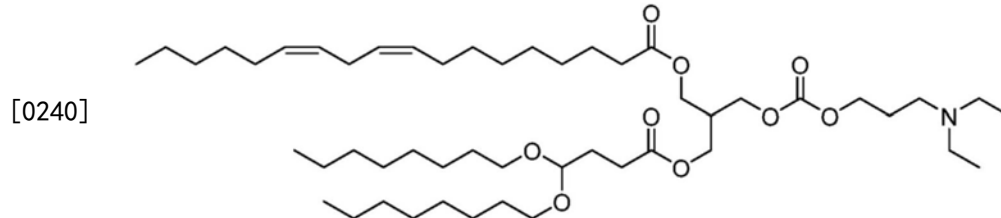
[0236] 在一些实施方案中,通过脂质纳米颗粒施用单独或编码在一个或多个载体上的本文所述向导RNA组合物;参见例如PCT/US2017/024973,其于2017年3月30日提交,要求2016年3月30日提交的标题为“用于CRISPR/Cas成分的脂质纳米颗粒制剂”的U.S.S.N.62/315,602的优先权,其内容在此以其整体引入作为参考。可以利用本领域技术人员已知能够递送核苷酸至个体的任何脂质纳米颗粒来施用本文所述向导RNA,以及编码Cas或Cas脱氨酶融合蛋白或Cas9或Cas9脱氨酶融合蛋白本身的mRNA

[0237] 在一些实施方案中,LNP包含:(i)用于包封和用于内体逃逸的CCD脂质;(ii)用于稳定化的中性脂质;(iii)同样用于稳定化的辅助脂质;和(iv)隐形脂质。LNP载体负荷可以包含以下任意一个或全部:编码Cas核酸酶或Cas-脱氨酶(如Cas9或Cas9-脱氨酶)的mRNA;一个或多个向导RNA或编码一个或多个向导RNA的核酸;及编码Cas9或Cas9-脱氨酶、一个或多个向导RNA、或Cas9/Cas9-脱氨酶和向导RNA二者的一个或多个病毒载体。在一个实施方案中,LNP包含CCD脂质,如脂质A、脂质B、脂质C或脂质D。在一些方面,CCD脂质是脂质A。在一些方面,CCD脂质是脂质B。在一些实施方案中,LNP包含CCD脂质、中性脂质、辅助脂质和隐形脂质。在某些实施方案中,辅助脂质是胆固醇。在某些实施方案中,中性脂质是DSPC。在一些实施方案中,隐形脂质是PEG2k-DMG。在其他实施方案中,LNP包含选自脂质A或脂质B的CCD脂质、胆固醇、DSPC和PEG2k-DMG。

[0238] 在一些实施方案中,适宜的LNP制剂包含CCD脂质,连同辅助脂质、中性脂质和隐形脂质。“脂质纳米颗粒”意指包含多个(例如超过一个)通过分子间力相互物理结合的脂质分子的颗粒。LNP可以是例如微球(包括单层和多层脂质体,例如“脂质体”——层状相脂双层,在一些实施方案中,其基本为球形,在更具体的实施方案中,其可以包含水性核心,例如包含RNA分子的实质部分)、乳剂中的分散相、微团或悬剂中的内相。乳剂、微团和悬剂可以是

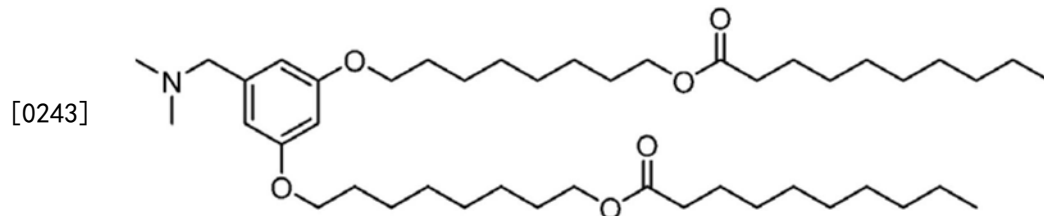
用于局部和/或外部递送的适宜组合物。

[0239] 在一些实施方案中,该CCD脂质是脂质A,其是(9Z,12Z)-3-((4,4-双(辛氧基)丁酰基)氧)-2-(((3-(二乙氨基)丙氧基)羰基)氧)甲基)丙基十八碳-9,12-二烯酸酯,也称为3-((4,4-双(辛氧基)丁酰基)氧)-2-(((3-(二乙氨基)丙氧基)羰基)氧)甲基)丙基(9Z,12Z)-十八碳-9,12-二烯酸酯。脂质A可以描绘为:



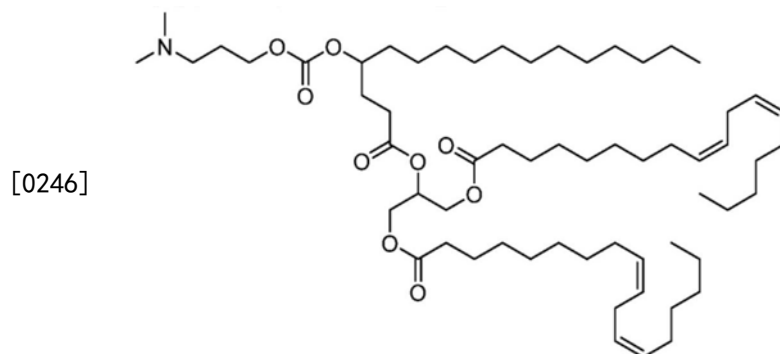
[0241] 脂质A可以按照W02015/095340(例如84-86页)合成,以其整体引入作为参考。

[0242] 在一些实施方案中,该CCD脂质是脂质B,其是((5-((二甲氨基)甲基)-1,3-亚苯基)双(氧))双(辛烷-8,1-亚基)双(癸酸酯),也称为((5-((二甲氨基)甲基)-1,3-亚苯基)双(氧))双(辛烷-8,1-亚基)双(癸酸酯)。脂质B可以描绘为:



[0244] 脂质B可以按照W02014/136086(例如107-09页)合成,以其整体引入作为参考。

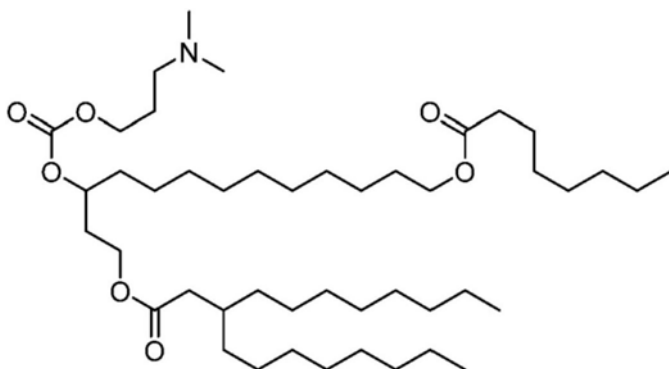
[0245] 在一些实施方案中,该CCD脂质是脂质C,其是2-((4-(((3-(二甲氨基)丙氧基)羰基)氧)十六酰)氧)丙烷-1,3-亚基(9Z,9'Z,12Z,12'Z)-双(十八碳-9,12-二烯酸酯)。脂质C可以描绘为:



[0247] 在一些实施方案中,该CCD脂质是脂质D,其是3-(((3-(二甲氨基)丙氧基)羰基)氧)-13-(辛酰基氧)三癸基3-辛基十一酸酯。

[0248] 脂质D可以描绘为:

[0249]



[0250] 脂质C和脂质D可以按照W02015/095340合成,以其整体引入作为参考。

[0251] 适合用于脂质组合物的“中性脂质”包括例如多种中性、不带电荷或两性离子脂质。适合用于本公开的中性磷脂的实例包括但不限于:5-十七烷基苯-1,3-二酚(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)、二豆蔻酰磷脂酰胆碱(DMPC)、磷脂酰胆碱(PLPC)、1,2-二硬脂酰-锡-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)、二豆蔻酰磷脂酰胆碱(DMPC)、1-豆蔻酰-2-棕榈酰磷脂酰胆碱(MPPC)、1-棕榈酰-2-豆蔻酰磷脂酰胆碱(PMPC)、1-棕榈酰-2-硬脂酰磷脂酰胆碱(PSPC)、1,2-二花生四烯酰(diarachidoyl)-锡-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(dieicosenoyl)-锡-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)、溶血磷脂酰乙醇胺及其组合。在一个实施方案中,中性磷脂可以选自二硬脂酰磷脂酰胆碱(DSPC)和二豆蔻酰磷脂酰乙醇胺(DMPE)。在另一实施方案中,中性磷脂可以是二硬脂酰磷脂酰胆碱(DSPC)。中性脂质发挥功能来稳定和改善LNP的处理。

[0252] “辅助脂质”是增强转染(包含生物活性剂的纳米颗粒的转染)的脂质。辅助脂质增强转染的机制包括增强颗粒稳定性。在某些实施方案中,辅助脂质增强膜fusogenicity。辅助脂质包括类固醇、固醇和烷基间苯二酚。适合用于LNP的辅助脂质包括但不限于胆固醇、5-十七烷基间苯二酚和胆固醇半琥珀酸酯。在一个实施方案中,辅助脂质可以是胆固醇。在一些实施方案中,辅助脂质可以是胆固醇半琥珀酸酯。

[0253] “隐形脂质”是改变纳米颗粒可以在体内(例如在血液中)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颗粒大小来辅助配制过程。本文所用的隐形脂质可以调节LNP的药代动力学特性。适合用于脂质组合物的隐形脂质包括但不限于具有与脂质部分连接的亲水性首基的隐形脂质。适合用于本公开的脂质组合物的隐形脂质和关于这类脂质的生物化学的信息可见于Romberg等,Pharmaceutical Research,第25卷,第1期,2008,55-71页和Hoekstra等,Biochimica et Biophysica Acta 1660(2004)41-52中。其他适宜的PEG脂质公开于例如W0 2006/007712中。

[0254] 在一个实施方案中,隐形脂质的亲水性首基包含选自基于PEG的聚合物(有时称为聚(环氧乙烷))、聚(噁唑啉)、聚(乙二醇)、聚(甘油)、聚(N-乙烯吡咯烷酮)、聚氨基酸和聚[N-(2-羟丙基)甲基丙烯酰胺]。

[0255] 隐形脂质可以包含脂质部分。在一些实施方案中,隐形脂质的脂质部分可以源自

二酰基甘油或二酰基甘油酰胺,包括含有二烷基甘油或二烷基甘油酰胺基团的那些,该二烷基甘油或二烷基甘油酰胺基团具有独立地包含约C4至约C40饱和或不饱和碳原子的烷基链长,其中链可以包含一个或多个官能团,例如酰胺或酯。该二烷基甘油或二烷基甘油酰胺基团可以进一步包含一个或多个取代烷基。

[0256] 除非另有说明,本文所用的术语“PEG”指任何聚乙二醇或其他聚亚烷基醚聚合物。在一些实施方案中,PEG是乙二醇或环氧乙烷的可选地取代的线性或分支聚合物。在一些实施方案中,PEG是非取代的。在一些实施方案中,PEG例如由一个或多个烷基、烷氧基、酰基、羟基或芳基取代。在一些实施方案中,该术语包括PEG共聚物,如PEG-聚氨基甲酸酯或PEG-聚丙烯(参见例如J.Milton Harris, Poly(ethylene glycol) chemistry: biotechnical and biomedical applications (1992)); 在另一实施方案中,该术语不包括PEG共聚物。在一些实施方案中,该PEG具有从约130至约50,000的分子量,在子实施方案中,该PEG具有约150至约30,000的分子量,在子实施方案中,该PEG具有约150至约20,000的分子量,在子实施方案中,该PEG具有约150至约15,000的分子量,在子实施方案中,该PEG具有约150至约10,000的分子量,在子实施方案中,该PEG具有约150至约6,000的分子量,在子实施方案中,该PEG具有约150至约5,000的分子量,在子实施方案中,该PEG具有约150至约4,000的分子量,在子实施方案中,该PEG具有约150至约3,000的分子量,在子实施方案中,该PEG具有约300至约3,000的分子量,在子实施方案中,该PEG具有约1,000至约3,000的分子量,在子实施方案中,该PEG具有约1,500至约2,500的分子量。

[0257] 在某些实施方案中,该PEG(例如与脂质如隐形脂质缀合)是“PEG-2K”,也称为“PEG 2000”,其具有约2,000道尔顿的平均分子量。PEG-2K在本文中表示为以下式(I),其中n为

45,意指数值平均聚合度包含约45个亚单位  $\left[ \text{O}-\text{CH}_2-\text{CH}_2 \right]_n \text{OR}$  (I) 但是,可以使用本领域已

知的其他PEG实施方案,包括例如其中数值平均聚合度包含约23个亚单位( $n=23$ )和/或68个亚单位( $n=68$ )的那些。在一些实施方案中,n可以在从约30至约60的范围内。在一些实施方案中,n可以在从约35至约55的范围内。在一些实施方案中,n可以在从约40至约50的范围内。在一些实施方案中,n可以在从约42至约48的范围内。在一些实施方案中,n可以是45。在一些实施方案中,R可以选自H、取代烷基和非取代烷基。在一些实施方案中,R可以是非取代烷基。在一些实施方案中,R可以是甲基。

[0258] 在本文所述实施方案的任一个中,隐形脂质可以选自PEG-二月桂酰甘油、PEG-二豆蔻酰甘油(PEG-DMG)(来自日本东京NOF的目录#GM-020)、PEG-二棕榈酰甘油、PEG-二硬脂酰甘油(PEG-DSPE)(来自日本东京NOF的目录#DSPE-020CN)、PEG-二月桂酰甘油酰胺、PEG-二豆蔻酰甘油酰胺、PEG-二棕榈酰甘油酰胺、PEG-二硬脂酰甘油酰胺、PEG-胆固醇(1-[8'-(胆甾-5-烯-3[β]-氧)酰胺-3',6'-二噁辛基]氨基甲酰基-[ω]-甲基-聚(乙二醇)、PEG-DMB(3,4-二十四烷氧基苄基-[ω]-甲基-聚(乙二醇)醚)、1,2-二豆蔻酰-锡-甘油-3-磷酸乙醇胺-N-[甲氧基(聚乙二醇)-2000](PEG2k-DMG)(来自Avanti Polar Lipids, Alabaster, Alabama, USA的目录#880150P)、1,2-二硬脂酰-锡-甘油-3-磷酸乙醇胺-N-[甲氧基(聚乙二醇)-2000](PEG2k-DSPE)(来自Avanti Polar Lipids, Alabaster, Alabama, USA的目录#880120C)、1,2-二硬脂酰-锡-甘油-甲氧基聚乙二醇(PEG2k-DSG; GS-020, NOF Tokyo, 日本)、聚(乙二醇)-2000-二甲基丙烯酸(PEG2k-DMA)和1,2-二硬脂酰氧丙基-3-胺-

N-[甲氧基(聚乙二醇)-2000] (PEG2k-DSA)。在一个实施方案中,隐形脂质可以是PEG2k-DMG。在一些实施方案中,隐形脂质可以是PEG2k-DSG。在一个实施方案中,隐形脂质可以是PEG2k-DSPE。在一个实施方案中,隐形脂质可以是PEG2k-DMA。在一个实施方案中,隐形脂质可以是PEG2k-DSA。在一个实施方案中,隐形脂质可以是PEG2k-C11。在一些实施方案中,隐形脂质可以是PEG2k-C14。在一些实施方案中,隐形脂质可以是PEG2k-C16。在一些实施方案中,隐形脂质可以是PEG2k-C18。

[0259] 本公开的实施方案还提供按照成分脂质在制剂中的各摩尔比描述的脂质组合物。在一个实施方案中,CCD脂质的mol-%可以从约30mol-%至约60mol-%。在一个实施方案中,CCD脂质的mol-%可以从约35mol-%至约55mol-%。在一个实施方案中,CCD脂质的mol-%可以从约40mol-%至约50mol-%。在一个实施方案中,CCD脂质的mol-%可以从约42mol-%至约47mol-%。在一个实施方案中,CCD脂质的mol-%可以是45mol-%。在一些实施方案中,LNP批次的CCD脂质mol-%将是目标mol-%的 $\pm 30\%$ 、 $\pm 25\%$ 、 $\pm 20\%$ 、 $\pm 15\%$ 、 $\pm 10\%$ 、 $\pm 5\%$ 或 $\pm 2.5\%$ 。在某些实施方案中,LNP批间变异将小于15%、小于10%或小于5%。

[0260] 在一个实施方案中,辅助脂质的mol-%可以从约30mol-%至约60mol-%。在一个实施方案中,辅助脂质的mol-%可以从约35mol-%至约55mol-%。在一个实施方案中,辅助脂质的mol-%可以从约40mol-%至约50mol-%。在一个实施方案中,辅助脂质的mol-%可以从约41mol-%至约46mol-%。在一个实施方案中,辅助脂质的mol-%可以是约44mol-%。在一些实施方案中,LNP批次的辅助脂质mol-%将是目标mol-%的 $\pm 30\%$ 、 $\pm 25\%$ 、 $\pm 20\%$ 、 $\pm 15\%$ 、 $\pm 10\%$ 、 $\pm 5\%$ 或 $\pm 2.5\%$ 。在某些实施方案中,LNP批间变异将小于15%、小于10%或小于5%。

[0261] 在一个实施方案中,中性脂质的mol-%可以从约1mol-%至约20mol-%。在一个实施方案中,中性脂质的mol-%可以从约5mol-%至约15mol-%。在一个实施方案中,中性脂质的mol-%可以从约7mol-%至约12mol-%。在一个实施方案中,中性脂质的mol-%可以是约9mol-%。在一些实施方案中,LNP批次的中性脂质mol-%将是目标mol-%的 $\pm 30\%$ 、 $\pm 25\%$ 、 $\pm 20\%$ 、 $\pm 15\%$ 、 $\pm 10\%$ 、 $\pm 5\%$ 或 $\pm 2.5\%$ 。在某些实施方案中,LNP批间变异将小于15%、小于10%或小于5%。

[0262] 在一个实施方案中,隐形脂质的mol-%可以从约1mol-%至约10mol-%。在一个实施方案中,隐形脂质的mol-%可以从约1mol-%至约5mol-%。在一个实施方案中,隐形脂质的mol-%可以从约1mol-%至约3mol-%。在一个实施方案中,隐形脂质的mol-%可以是约2mol-%。在一个实施方案中,隐形脂质的mol-%可以是约1mol-%。在一些实施方案中,LNP批次的隐形脂质mol-%将是目标mol-%的 $\pm 30\%$ 、 $\pm 25\%$ 、 $\pm 20\%$ 、 $\pm 15\%$ 、 $\pm 10\%$ 、 $\pm 5\%$ 或 $\pm 2.5\%$ 。在某些实施方案中,LNP批间变异将小于15%、小于10%或小于5%。

#### [0263] 施用位置

[0264] 在一些实施方案中,将该组合物递送入眼前房。在一些实施方案中,将该组合物递送入眼后房。在一些实施方案中,将该组合物递送入角膜本身。在一些实施方案中,将该组合物递送入角膜基质。在一些实施方案中,将该组合物递送入角膜缘。在一些实施方案中,将该组合物都送至角膜上皮表面。在本段落的任意前述实施方案以及本文所述的其他实施方案中,治疗进一步包括例如用脂质纳米颗粒递送Cas蛋白质(例如Cas9),或用载体和/或脂质纳米颗粒递送编码Cas蛋白质的核酸。在一些实施方案(例如使用脂质纳米颗粒的那

些)中,编码Cas蛋白质的核酸是mRNA。在一些实施方案中,通过与用于递送该一个或多个适当的向导的载体和/或脂质纳米颗粒相同的载体和/或脂质纳米颗粒来递送Cas蛋白质或编码Cas蛋白质的核酸。在一些实施方案中,通过与用于递送该一个或多个适当的向导的载体和/或脂质纳米颗粒不同的载体和/或脂质纳米颗粒来递送Cas蛋白质或编码Cas蛋白质的核酸。

[0265] 可对个体施用本文所述的任意组合物来切除TCF4内含子3中的部分或全部TNR扩增。涵盖包括施用本文所述任意组合物的治疗FEC D的方法。在一些方面,该组合物按治疗有效量施用。在一些实施方案中,涵盖切除TCF4的TNR、突变TCF4的TNR、减少TCF4的TNR的拷贝数、改善TCF4的TNR和/或消除TCF4的TNR的方法,其包括施用本文所述的一种或多种组合物。在一些实施方案中,提供在个体中切割每个细胞的一个或两个TCF4拷贝的TNR、突变每个细胞的一个或两个TCF4拷贝的TNR、减少每个细胞的一个或两个TCF4拷贝的TNR的拷贝数、改善每个细胞的一个或两个TCF4拷贝的TNR和/或消除每个细胞的一个或两个TCF4拷贝的TNR的方法,其包括施用本文所述的一种或多种组合物。在一些实施方案中,该细胞是角膜内皮细胞。

[0266] 在一些实施方案中,用两个gRNA来切除TCF4中的全部TNR。在一些实施方案中,TNR 5'的第一向导与TNR 3'的第二向导一起提供,或反之。在考虑两个gRNA时,提供包含以下向导组合中任一个的组合物:

[0267] 组合01:在一些实施方案中,组合物包含两个gRNA,该两个gRNA包含含有SEQ ID NO:1089的第一gRNA和含有选自SEQ ID NO:1182-1278的序列的第二gRNA。

[0268] 组合02:在一些实施方案中,该组合物包含两个gRNA,该两个gRNA包含含有SEQ ID NO:1090的第一gRNA和含有选自SEQ ID NO:1182-1278的序列的第二gRNA。

[0269] 组合03:在一些实施方案中,该组合物包含两个gRNA,该两个gRNA包含含有SEQ ID NO:1091的第一gRNA和含有选自SEQ ID NO:1182-1278的序列的第二gRNA。

[0270] 组合04:在一些实施方案中,该组合物包含两个gRNA,该两个gRNA包含含有SEQ ID NO:1092的第一gRNA和含有选自SEQ ID NO:1182-1278的序列的第二gRNA。

[0271] 组合05:在一些实施方案中,该组合物包含两个gRNA,该两个gRNA包含含有SEQ ID NO:1093的第一gRNA和含有选自SEQ ID NO:1182-1278的序列的第二gRNA。

[0272] 组合06:在一些实施方案中,该组合物包含两个gRNA,该两个gRNA包含含有SEQ ID NO:1094的第一gRNA和含有选自SEQ ID NO:1182-1278的序列的第二gRNA。

[0273] 组合07:在一些实施方案中,该组合物包含两个gRNA,该两个gRNA包含含有SEQ ID NO:1095的第一gRNA和含有选自SEQ ID NO:1182-1278的序列的第二gRNA。

[0274] 组合08:在一些实施方案中,该组合物包含两个gRNA,该两个gRNA包含含有SEQ ID NO:1096的第一gRNA和含有选自SEQ ID NO:1182-1278的序列的第二gRNA。

[0275] 组合09:在一些实施方案中,该组合物包含两个gRNA,该两个gRNA包含含有SEQ ID NO:1097的第一gRNA和含有选自SEQ ID NO:1182-1278的序列的第二gRNA。

[0276] 组合10:在一些实施方案中,该组合物包含两个gRNA,该两个gRNA包含含有SEQ ID NO:1098的第一gRNA和含有选自SEQ ID NO:1182-1278的序列的第二gRNA。

[0277] 组合11:在一些实施方案中,该组合物包含两个gRNA,该两个gRNA包含含有SEQ ID NO:1099的第一gRNA和含有选自SEQ ID NO:1182-1278的序列的第二gRNA。









- [0356] 组合90:在一些实施方案中,该组合物包含两个gRNA,该两个gRNA包含含有SEQ ID NO:1178的第一gRNA和含有选自SEQ ID NO:1182-1278的序列的第二gRNA。
- [0357] 组合91:在一些实施方案中,该组合物包含两个gRNA,该两个gRNA包含含有SEQ ID NO:1179的第一gRNA和含有选自SEQ ID NO:1182-1278的序列的第二gRNA。
- [0358] 组合92:在一些实施方案中,该组合物包含两个gRNA,该两个gRNA包含含有SEQ ID NO:1180的第一gRNA和含有选自SEQ ID NO:1182-1278的序列的第二gRNA。
- [0359] 组合93:在一些实施方案中,该组合物包含两个gRNA,该两个gRNA包含含有SEQ ID NO:1181的第一gRNA和含有选自SEQ ID NO:1182-1278的序列的第二gRNA。
- [0360] 组合94:在一些实施方案中,该组合物包含两个gRNA,该两个gRNA包含含有SEQ ID NO:1182的第一gRNA和含有选自SEQ ID NO:1089-1181的序列的第二gRNA。
- [0361] 组合95:在一些实施方案中,该组合物包含两个gRNA,该两个gRNA包含含有SEQ ID NO:1183的第一gRNA和含有选自SEQ ID NO:1089-1181的序列的第二gRNA。
- [0362] 组合96:在一些实施方案中,该组合物包含两个gRNA,该两个gRNA包含含有SEQ ID NO:1184的第一gRNA和含有选自SEQ ID NO:1089-1181的序列的第二gRNA。
- [0363] 组合97:在一些实施方案中,该组合物包含两个gRNA,该两个gRNA包含含有SEQ ID NO:1185的第一gRNA和含有选自SEQ ID NO:1089-1181的序列的第二gRNA。
- [0364] 组合98:在一些实施方案中,该组合物包含两个gRNA,该两个gRNA包含含有SEQ ID NO:1186的第一gRNA和含有选自SEQ ID NO:1089-1181的序列的第二gRNA。
- [0365] 组合99:在一些实施方案中,该组合物包含两个gRNA,该两个gRNA包含含有SEQ ID NO:1187的第一gRNA和含有选自SEQ ID NO:1089-1181的序列的第二gRNA。
- [0366] 组合100:在一些实施方案中,该组合物包含两个gRNA,该两个gRNA包含含有SEQ ID NO:1188的第一gRNA和含有选自SEQ ID NO:1089-1181的序列的第二gRNA。
- [0367] 组合101:在一些实施方案中,该组合物包含两个gRNA,该两个gRNA包含含有SEQ ID NO:1189的第一gRNA和含有选自SEQ ID NO:1089-1181的序列的第二gRNA。
- [0368] 组合102:在一些实施方案中,该组合物包含两个gRNA,该两个gRNA包含含有SEQ ID NO:1190的第一gRNA和含有选自SEQ ID NO:1089-1181的序列的第二gRNA。
- [0369] 组合103:在一些实施方案中,该组合物包含两个gRNA,该两个gRNA包含含有SEQ ID NO:1191的第一gRNA和含有选自SEQ ID NO:1089-1181的序列的第二gRNA。
- [0370] 组合104:在一些实施方案中,该组合物包含两个gRNA,该两个gRNA包含含有SEQ ID NO:1192的第一gRNA和含有选自SEQ ID NO:1089-1181的序列的第二gRNA。
- [0371] 组合105:在一些实施方案中,该组合物包含两个gRNA,该两个gRNA包含含有SEQ ID NO:1193的第一gRNA和含有选自SEQ ID NO:1089-1181的序列的第二gRNA。
- [0372] 组合106:在一些实施方案中,该组合物包含两个gRNA,该两个gRNA包含含有SEQ ID NO:1194的第一gRNA和含有选自SEQ ID NO:1089-1181的序列的第二gRNA。
- [0373] 组合107:在一些实施方案中,该组合物包含两个gRNA,该两个gRNA包含含有SEQ ID NO:1195的第一gRNA和含有选自SEQ ID NO:1089-1181的序列的第二gRNA。
- [0374] 组合108:在一些实施方案中,该组合物包含两个gRNA,该两个gRNA包含含有SEQ ID NO:1196的第一gRNA和含有选自SEQ ID NO:1089-1181的序列的第二gRNA。
- [0375] 组合109:在一些实施方案中,该组合物包含两个gRNA,该两个gRNA包含含有SEQ

ID NO:1197的第一gRNA和含有选自SEQ ID NO:1089-1181的序列的第二gRNA。

[0376] 组合110:在一些实施方案中,该组合物包含两个gRNA,该两个gRNA包含含有SEQ ID NO:1198的第一gRNA和含有选自SEQ ID NO:1089-1181的序列的第二gRNA。

[0377] 组合111:在一些实施方案中,该组合物包含两个gRNA,该两个gRNA包含含有SEQ ID NO:1199的第一gRNA和含有选自SEQ ID NO:1089-1181的序列的第二gRNA。

[0378] 组合112:在一些实施方案中,该组合物包含两个gRNA,该两个gRNA包含含有SEQ ID NO:1200的第一gRNA和含有选自SEQ ID NO:1089-1181的序列的第二gRNA。

[0379] 组合113:在一些实施方案中,该组合物包含两个gRNA,该两个gRNA包含含有SEQ ID NO:1201的第一gRNA和含有选自SEQ ID NO:1089-1181的序列的第二gRNA。

[0380] 组合114:在一些实施方案中,该组合物包含两个gRNA,该两个gRNA包含含有SEQ ID NO:1202的第一gRNA和含有选自SEQ ID NO:1089-1181的序列的第二gRNA。

[0381] 组合115:在一些实施方案中,该组合物包含两个gRNA,该两个gRNA包含含有SEQ ID NO:1203的第一gRNA和含有选自SEQ ID NO:1089-1181的序列的第二gRNA。

[0382] 组合116:在一些实施方案中,该组合物包含两个gRNA,该两个gRNA包含含有SEQ ID NO:1204的第一gRNA和含有选自SEQ ID NO:1089-1181的序列的第二gRNA。

[0383] 组合117:在一些实施方案中,该组合物包含两个gRNA,该两个gRNA包含含有SEQ ID NO:1205的第一gRNA和含有选自SEQ ID NO:1089-1181的序列的第二gRNA。

[0384] 组合118:在一些实施方案中,该组合物包含两个gRNA,该两个gRNA包含含有SEQ ID NO:1206的第一gRNA和含有选自SEQ ID NO:1089-1181的序列的第二gRNA。

[0385] 组合119:在一些实施方案中,该组合物包含两个gRNA,该两个gRNA包含含有SEQ ID NO:1207的第一gRNA和含有选自SEQ ID NO:1089-1181的序列的第二gRNA。

[0386] 组合120:在一些实施方案中,该组合物包含两个gRNA,该两个gRNA包含含有SEQ ID NO:1208的第一gRNA和含有选自SEQ ID NO:1089-1181的序列的第二gRNA。

[0387] 组合121:在一些实施方案中,该组合物包含两个gRNA,该两个gRNA包含含有SEQ ID NO:1209的第一gRNA和含有选自SEQ ID NO:1089-1181的序列的第二gRNA。

[0388] 组合122:在一些实施方案中,该组合物包含两个gRNA,该两个gRNA包含含有SEQ ID NO:1210的第一gRNA和含有选自SEQ ID NO:1089-1181的序列的第二gRNA。

[0389] 组合123:在一些实施方案中,该组合物包含两个gRNA,该两个gRNA包含含有SEQ ID NO:1211的第一gRNA和含有选自SEQ ID NO:1089-1181的序列的第二gRNA。

[0390] 组合124:在一些实施方案中,该组合物包含两个gRNA,该两个gRNA包含含有SEQ ID NO:1212的第一gRNA和含有选自SEQ ID NO:1089-1181的序列的第二gRNA。

[0391] 组合125:在一些实施方案中,该组合物包含两个gRNA,该两个gRNA包含含有SEQ ID NO:1213的第一gRNA和含有选自SEQ ID NO:1089-1181的序列的第二gRNA。

[0392] 组合126:在一些实施方案中,该组合物包含两个gRNA,该两个gRNA包含含有SEQ ID NO:1214的第一gRNA和含有选自SEQ ID NO:1089-1181的序列的第二gRNA。

[0393] 组合127:在一些实施方案中,该组合物包含两个gRNA,该两个gRNA包含含有SEQ ID NO:1215的第一gRNA和含有选自SEQ ID NO:1089-1181的序列的第二gRNA。

[0394] 组合128:在一些实施方案中,该组合物包含两个gRNA,该两个gRNA包含含有SEQ ID NO:1216的第一gRNA和含有选自SEQ ID NO:1089-1181的序列的第二gRNA。







ID NO:1275的第一gRNA和含有选自SEQ ID NO:1089-1181的序列的第二gRNA。

[0454] 组合188:在一些实施方案中,该组合物包含两个gRNA,该两个gRNA包含含有SEQ ID NO:1276的第一gRNA和含有选自SEQ ID NO:1089-1181的序列的第二gRNA。

[0455] 组合189:在一些实施方案中,该组合物包含两个gRNA,该两个gRNA包含含有SEQ ID NO:1277的第一gRNA和含有选自SEQ ID NO:1089-1181的序列的第二gRNA。

[0456] 组合190:在一些实施方案中,该组合物包含两个gRNA,该两个gRNA包含含有SEQ ID NO:1278的第一gRNA和含有选自SEQ ID NO:1089-1181的序列的第二gRNA。

## 实施例

[0457] 实施例1.用gRNA对从TCF4切除TNR扩增

[0458] 为了从TCF4去除TNR和限制毒性RNA的产生,用特异性靶序列设计了CRISPR向导,来在扩增的任一侧同时切割。这些gRNA设计为与野生型酿脓链球菌Cas9(“Spy Cas9”)一起使用。适于与其他CRISPR核酸酶一起使用的其他gRNA可以相似的方式设计。

[0459] 用含侧翼外显子的TCF4内含子3序列的序列(SEQ ID NO:1085)选择靶序列。此序列基于USCS基因组浏览器,人,2009年2月(GRCh37/hg19)装配。此序列在内含子位置chr18:53252584-53254275内的53253387-53253458范围内包含一组24个CTG重复(TNR)。此内含子中CTG重复的精确范围将根据重复数而变,其中重复数>40与发展疾病的风险提高相关。在hg38构造中,该重复定位在跨chr18:55,585,280-55,587,136的内含子内的chr18:55,586,156-55,586,228。靶序列和对应的向导序列在表2(SEQ ID NO:1-190(靶序列)和SEQ ID NO:1089-1278(向导序列))中列出。此实施例1中使用的crRNA和trRNA的具体形式在表1中分别作为SEQ ID NO:1087和SEQ ID NO:1088提供。5'向导序列(SEQ ID NO:1089-1181)的靶序列定位在Chr18:55,585,285-55,586,153之间,处于TNR位置的上游。3'向导序列的靶序列(SEQ ID NO:94-190)定位在Chr18:55586225-55587203之间,处于TNR位置的下游。表2列出SEQ ID NO:1-190(靶序列)和SEQ ID NO:1089-1278(指导核酸酶至对应的靶序列并结合靶序列的反向互补序列的向导序列)。按照Doench等,NatBiotechnol.2016Feb;34(2):184-191报道的方法,计算机模拟产生各向导序列的切割频率测定(CFD)得分。这些得分(与Doench等如何报告得分相比,已乘以100的因子转化为小数)提供了给定gRNA的脱靶可能性的测量。

[0460]

SEQ ID NO	靶序列 (包括 PAM)	染色体位置	链	方向	与 TNR 起点的距离	SEQ ID No	向导序列	CFD 得分	平均编辑百分比
1	TTGCCAAGTGGAC ATTTACTGG	Chr18:55585285-55585307	-	TCF4 TNR 的 5'	-871	1089	UUGGCAAGUG GACAUUUUAC	422.48	NA
2	TGTCCACTTGCCA AAGAAGTTGG	Chr18:55585294-55585316	+	TCF4 TNR 的 5'	-862	1090	UGUCCACUUGC CAAGAAGU	619.25	NA
3	GGACCAACTTCTT TGGCAAGTGG	Chr18:55585297-55585319	-	TCF4 TNR 的 5'	-859	1091	GGACCAACUUC UUUGGCAAG	402.71	NA
4	GAAAAATGGACCA ACTTCTTTGG	Chr18:55585304-55585326	-	TCF4 TNR 的 5'	-852	1092	GAAAAAUGGA CCAACUUCUU	1569.2 2	NA
5	CCATTTTCCCCT GCTCACAGG	Chr18:55585318-55585340	+	TCF4 TNR 的 5'	-838	1093	CCAUUUUCCCC ACUGCUCAC	809.81	NA
6	CCTGTGAGCAGTG GGAAAAATGG	Chr18:55585318-55585340	-	TCF4 TNR 的 5'	-838	1094	CCUGUGAGCA GUGGGAAAAA	773.35	NA
7	TTTTTCCCCTGCT CACAGGAGG	Chr18:55585321-55585343	+	TCF4 TNR 的 5'	-835	1095	UUUUUCCCACU GCUCACAGG	1673.7 9	NA
8	TTTCCCTCCTGT GAGCAGTGGG	Chr18:55585326-55585348	-	TCF4 TNR 的 5'	-830	1096	UUUCACCCUCCU GUGAGCAGU	1250.2 7	NA
9	TTTTCACCTCCTGT GAGCAGTGG	Chr18:55585327-55585349	-	TCF4 TNR 的 5'	-829	1097	UUUUCACCCUCC UGUGAGCAG	1372.0 8	NA
10	AGATCTTTGAGGA GCTCTGAAGG	Chr18:55585399-55585421	-	TCF4 TNR 的 5'	-757	1098	AGAUCUUUGA GGAGCUCUGA	147.38	27.9
11	AACAGTATGAAAG ATCTTTGAGG	Chr18:55585410-55585432	-	TCF4 TNR 的 5'	-746	1099	AACAGUAUGA AAGAUCUUUG	369.96	32.87
12	AGCAFAAACTCTA AGCTGTTTGG	Chr18:55585434-55585456	-	TCF4 TNR 的 5'	-722	1100	AGCAUAAACU CUAAGCUGUU	37.08	1.83
13	ACAGCTTAGAGTT TATGCTAAGG	Chr18:55585438-55585460	+	TCF4 TNR 的 5'	-718	1101	ACAGCUUAGA GUUUAUGCUA	197.78	7.6
14	CAGCTTAGAGTTT ATGCTAAGGG	Chr18:55585439-55585461	+	TCF4 TNR 的 5'	-717	1102	CAGCUUAGAG UUUAUGCUGA	178.67	1.93
15	TCTTTTAGTTTTAA GTTGGATGG	Chr18:55585483-55585505	-	TCF4 TNR 的 5'	-673	1103	UCUUUUAGUU UUAAAGUUGGA	232.52	10.57

[0461]

SEQ ID NO	靶序列 (包括 PAM)	染色体位置	链	方向	与 TNR 起点的距离	SEQ ID No	向导序列	CFD 得分	平均编辑百分比
16	TTTCTCTTTTAGTT TTAAGTTGG	Chr18:55585487-55585509	-	TCF4 的 TNR 的 5'	-669	1104	UUUCUCUUUU AGUUUUUAAAGU	619.21	2.07
17	GTGATAATGGGG CTGGGGTGGG	Chr18:55585523-55585545	-	TCF4 的 TNR 的 5'	-633	1105	GUGAUAAUGG GGGCUGGGGU	635.78	15.53
18	AGTGATAATGGGG GCTGGGGTGG	Chr18:55585524-55585546	-	TCF4 的 TNR 的 5'	-632	1106	AGUGAUAAUG GGGCUGGGGG	633.13	11.3
19	CAGAGTGATAATG GGGGCTGGGG	Chr18:55585527-55585549	-	TCF4 的 TNR 的 5'	-629	1107	CAGAGUGAUA AUGGGGGCUG	350.31	17.2
20	ACAGAGTGATAAT GGGGCTGGG	Chr18:55585528-55585550	-	TCF4 的 TNR 的 5'	-628	1108	ACAGAGUGAU AAUGGGGGCU	331.09	10.3
21	AACAGAGTGATAA TGGGGCTGG	Chr18:55585529-55585551	-	TCF4 的 TNR 的 5'	-627	1109	AACAGAGUGA UAAUGGGGGC	3776.9 1	12.53
22	AAAGAACAGAGT GATAATGGGG	Chr18:55585533-55585555	-	TCF4 的 TNR 的 5'	-623	1110	AAAGAACAAGA GUGAUAAUGG	372.71	34
23	GAAAGAACAGAG TGATAATGGG	Chr18:55585534-55585556	-	TCF4 的 TNR 的 5'	-622	1111	GAAAGAACAAG AGUGAUAAUG	5837.9 9	17.57
24	AGAAAGAACAGA GTGATAATGG	Chr18:55585535-55585557	-	TCF4 的 TNR 的 5'	-621	1112	AGAAAGAACA GAGUGAUAAU	1439.1 2	17.37
25	AAGAAAGAACAG AGTGATAATGG	Chr18:55585536-55585558	-	TCF4 的 TNR 的 5'	-620	1113	AAGAAAGAAC AGAGUGAUA	418.32	4
26	TCTGTTCTTCTTT TTCCTCAGG	Chr18:55585546-55585568	+	TCF4 的 TNR 的 5'	-610	1114	UCUGUUCUUU CUUUUUCCUC	722.67	4.1
27	TTTTCTCAGGTT CATTAGATGG	Chr18:55585558-55585580	+	TCF4 的 TNR 的 5'	-598	1115	UUUUCCUCAG GUUCAUUAGA	740.15	14.7
28	TTGGCCATCTAATG AACCTGAGG	Chr18:55585562-55585584	-	TCF4 的 TNR 的 5'	-594	1116	UUGGCCAUCU AAUGAACCUG	201.82	28.2
29	AATGTAGCAGTAG TACTGCTTGG	Chr18:55585581-55585583	-	TCF4 的 TNR 的 5'	-575	1117	AAUGUAGCAG UAGUACUGCU	932.03	23
30	AGCAGTACTACTG CTACATTTGG	Chr18:55585584-55585586	+	TCF4 的 TNR 的 5'	-572	1118	AGCAGUACUA CUGCUACAUAU	975.76	4.43

[0462]

SEQ ID NO	靶序列 (包括 PAM)	染色体位置	链	方向	与 TNR 起点的距离	SEQ ID No	向导序列	CFD 得分	平均编辑百分比
31	TGAATCTTGATAA CATTATGGG	Chr18:55585619-55585641	-	TCF4 TNR 的 5'	-537	1119	UGAAUCUUGA UAACAUAUUG	430.8	22.13
32	CTGAATCTTGATA ACATTATGGG	Chr18:55585620-55585642	-	TCF4 TNR 的 5'	-536	1120	CUGAAUCUUG AUAACAUAU	603.7	32.73
33	CCATAATGTTATCA AGATTCAGG	Chr18:55585621-55585643	+	TCF4 TNR 的 5'	-535	1121	CCAUAAUGUU AUCAAGAUUC	473.28	15.53
34	CCTGAATCTTGATA ACATTATGG	Chr18:55585621-55585643	-	TCF4 TNR 的 5'	-535	1122	CCUGAAUCUU GAUAACAUUA	342.57	36.07
35	AATGTTATCAAGAT TCAGGTGG	Chr18:55585625-55585647	+	TCF4 TNR 的 5'	-531	1123	AAUGUUAUCA AGAUUCAGGU	405.03	15.6
36	GTTATCAAGATTC AGGTGGAGG	Chr18:55585628-55585650	+	TCF4 TNR 的 5'	-528	1124	GUUUACAAGA UUCAGGUUGG	355.48	21.3
37	TGTTTTCTAGAG AGGCTGCTGG	Chr18:55585651-55585673	-	TCF4 TNR 的 5'	-505	1125	UGUUUUUCUA GAGAGGCGUC	267.41	3.53
38	AAACTAGTGTTTT TCTAGAGAGG	Chr18:55585658-55585680	-	TCF4 TNR 的 5'	-498	1126	AAACUAGUGU UUUUCUAGAG	609.65	7.43
39	GAAAAACACTAGT TTCACCAAGG	Chr18:55585666-55585688	+	TCF4 TNR 的 5'	-490	1127	GAAAAACACU AGUUUCACCA	1273.03	22.27
40	AACAACTTTTTC TTCTCCTTGG	Chr18:55585683-55585705	-	TCF4 TNR 的 5'	-473	1128	AACAACUUUU UUCUUCUCCU	187.55	3.37
41	TGTTTTATATTGA AAACCTTGG	Chr18:55585706-55585728	+	TCF4 TNR 的 5'	-450	1129	UUGUUUUUAUA UUGAAAAACCU	330.57	5.57
42	GAAAAACCTTGCC ATAAACGTGG	Chr18:55585718-55585740	+	TCF4 TNR 的 5'	-438	1130	GAAAAACCUUG GCCAUAAAACG	242.99	24.23
43	CATTGCCACGTTTA TGGCCAAGG	Chr18:55585723-55585745	-	TCF4 TNR 的 5'	-433	1131	CAUUGCCACGU UUUUGGCCA	374.68	2.3
44	AATGGACATTGCC ACGTTATGG	Chr18:55585729-55585751	-	TCF4 TNR 的 5'	-427	1132	AAUGGACAUU GCCACGUUUA	221.28	19.5
45	TGTCCATTCCATC TCGTATAGG	Chr18:55585744-55585766	+	TCF4 TNR 的 5'	-412	1133	UGUCCAUUUCC AUCUCGUUA	7973.48	12.53

[0463]

SEQ ID NO	靶序列 (包括 PAM)	染色体位置	链	方向	与 TNR 起点的距离	SEQ ID No	向导序列	CFD 得分	平均编辑百分比
46	AATCCTATACGAG ATGGAATGG	Chr18:55585747-55585769	-	TCF4 的 TNR 的 5'	-409	1134	AAUCCUAUAC GAGAUUGGAAA	24066.2	6.87
47	CAGGCAAATCCTA TACGAGATGG	Chr18:55585753-55585775	-	TCF4 的 TNR 的 5'	-403	1135	CAGGCAAUCC UAUACGAGA	1112.86	7.3
48	TATTTGGGTTCCACA TATGACAGG	Chr18:55585772-55585794	-	TCF4 的 TNR 的 5'	-384	1136	UAUUUGGGUU CACAU AUGAC	1223.1	11.3
49	TGGCACITTTTATT TTAATTTGGG	Chr18:55585787-55585809	-	TCF4 的 TNR 的 5'	-369	1137	UGGCACUUUU AUUUUUUUUU	1409	1.37
50	GTGGCACITTTTATT TTTATTTGGG	Chr18:55585788-55585810	-	TCF4 的 TNR 的 5'	-368	1138	GUGGCACUUU UAUUUUUUUU	8296.18	1.17
51	AAATGAGAATTTA GTGCAGGTGG	Chr18:55585807-55585829	-	TCF4 的 TNR 的 5'	-349	1139	AAAUUGAGAAU UUAGUGCAGG	780.66	4.73
52	ACGAAATGAGAAT TTAGTGCAGG	Chr18:55585810-55585832	-	TCF4 的 TNR 的 5'	-346	1140	ACGAAAUUGAG AAUUUAGUGC	372.43	8.9
53	ATTCCTCATTTCTGTC TCTAACACAGG	Chr18:55585820-55585842	+	TCF4 的 TNR 的 5'	-336	1141	AUUCUCAUUU CGUCUCUAAC	182.73	19.17
54	AAATAAATGCTGG AGAGAGAGGG	Chr18:55585898-55585920	-	TCF4 的 TNR 的 5'	-258	1142	AAAUAAAUGC UGGAGAGAGA	283.11	32.93
55	GAAATAAATGCTG GAGAGAGAGG	Chr18:55585899-55585921	-	TCF4 的 TNR 的 5'	-257	1143	GAAAUAAAUG CUGGAGAGAG	516.92	20.5
56	ATTAGGGTCGAAA TAAATGCTGG	Chr18:55585908-55585930	-	TCF4 的 TNR 的 5'	-248	1144	AUUAGGGUCG AAAUAAAUGC	2074.54	31.6
57	GCATTTATTTCTGAC CCTAATTTGG	Chr18:55585911-55585933	+	TCF4 的 TNR 的 5'	-245	1145	GCAUUUAUUU CGACCCUAAU	430.39	12.77
58	AAGAAGAGGGAA ACCAATTAGG	Chr18:55585924-55585946	-	TCF4 的 TNR 的 5'	-232	1146	AAGAAGAGGG AAACCAAUUA	1894.27	47.23
59	GAAGAAGAGGGA AACCAATTAGG	Chr18:55585925-55585947	-	TCF4 的 TNR 的 5'	-231	1147	GAAGAAGAGG GAAACCAAUU	632.04	24
60	ACTAGATACGTCTG AAGAAGAGGG	Chr18:55585937-55585959	-	TCF4 的 TNR 的 5'	-219	1148	ACUAGAUACG UCGAAAGAAGA	554.05	18.97

[0464]

SEQ ID NO	靶序列 (包括 PAM)	染色体位置	链	方向	与 TNR 起点的距离	SEQ ID No	向导序列	CFD 得分	平均编辑百分比
61	CACTAGATACGTC GAAAGAAGAGG	Chr18:55585938-55585960	-	TCF4 TNR 的 5'	-218	1149	CACUAGAUAC GUCGAAGAAG	355.06	11.53
62	CTCTTCTTCGACGT ATCTAGTGG	Chr18:55585939-55585961	+	TCF4 TNR 的 5'	-217	1150	CUCUUCUUCGA CGUAUCUAG	397.65	18.03
63	TGCAGGCTCTGAC TCAGGGAAGG	Chr18:55585972-55585994	-	TCF4 TNR 的 5'	-184	1151	UGCAGGCUCU GACUCAGGGA	611.76	5.97
64	TTTTTGCAGGCTC TGACTCAGGG	Chr18:55585976-55585998	-	TCF4 TNR 的 5'	-180	1152	UUUUUGCAGG CUCUGACUCA	471.42	4.37
65	CTTTTGCAGGCT CTGACTCAGG	Chr18:55585977-55585999	-	TCF4 TNR 的 5'	-179	1153	UUUUUGCAG GCUCUGACUC	588.04	2.13
66	TCAGAGCCTGCAA AAGCAAAGG	Chr18:55585983-55586005	+	TCF4 TNR 的 5'	-173	1154	UCAGAGCCUGC AAAAAGCAA	523.08	13.97
67	TTCGTTCTTTGCT TTTTGCAGG	Chr18:55585989-55586011	-	TCF4 TNR 的 5'	-167	1155	UUCGUUCCUU UGCUUUUUGC	638.97	3.03
68	GCAAAAAGCAA GGAACGAATGG	Chr18:55585992-55586014	+	TCF4 TNR 的 5'	-164	1156	GCAAAAAGCA AAGGAACGAA	287.37	9.73
69	AGAAAGTGCAAC AAGCAGAAAGG	Chr18:55586015-55586037	+	TCF4 TNR 的 5'	-141	1157	AGAAAGUGCA ACAAGCAGAA	563.9	9.17
70	GAAAGTGCAACA AGCAGAAAGG	Chr18:55586016-55586038	+	TCF4 TNR 的 5'	-140	1158	GAAAGUGCAA CAAGCAGAAA	820.22	7.43
71	AAAAGTGCAACA GCAGAAAGGGG	Chr18:55586017-55586039	+	TCF4 TNR 的 5'	-139	1159	AAAAGUGCAAC AAGCAGAAAAG	677.96	30.07
72	AAGTGCAACAAG CAGAAAGGGG	Chr18:55586018-55586040	+	TCF4 TNR 的 5'	-138	1160	AAGUGCAACA AGCAGAAAAGG	423.94	16.47
73	GGTGCAAAAGCTG CTGCCTAGG	Chr18:55586039-55586061	+	TCF4 TNR 的 5'	-117	1161	GGCUGCAAAG CUGCCUGCCU	295.09	1.43
74	GCTGCAAAAGCTGC CTGCCTAGG	Chr18:55586040-55586062	+	TCF4 TNR 的 5'	-116	1162	GCUGCAAAGC UGCCUGCCUA	140464 <sub>9</sub>	37.6
75	CAGAAAACGTAGC CCTAGGCAGG	Chr18:55586052-55586074	-	TCF4 TNR 的 5'	-104	1163	CAGAAAACGU AGCCCUAAGGC	189.68	8.43

[0465]

SEQ ID NO	靶序列 (包括 PAM)	染色体位置	链	方向	与 TNR 起点的距离	SEQ ID No	向导序列	CFD 得分	平均编辑百分比
76	CTGCCTAGGGCTA CGTTTCCTGG	Chr18:55586053-55586075	+	TCF4 的 TNR 的 5'	-103	1164	CUGCCUAGGGC UACGUUUC	139.26	15
77	TTGCCAGGAAACG TAGCCCTAGG	Chr18:55586056-55586078	-	TCF4 的 TNR 的 5'	-100	1165	UUGCCAGGAA ACGUAGCCCU	68.07	31.3
78	TGGCTTTCGGAAG TTTTGCCAGG	Chr18:55586071-55586093	-	TCF4 的 TNR 的 5'	-85	1166	UGGCUUUCGG AAGUUUUGCC	122397	17.97
79	TCTTTTGGAGAAA TGGCTTTCGG	Chr18:55586084-55586106	-	TCF4 的 TNR 的 5'	-72	1167	UCUUUUGGAG AAAUGGCUUU	48.33	18.67
80	AAAGCCATTTC CAAAAGAAGG	Chr18:55586087-55586109	+	TCF4 的 TNR 的 5'	-69	1168	AAAGCCAUUU CUCCAAAAGA	12428.	22.93
81	TAGACCTTCTTTT GAGAAATGG	Chr18:55586091-55586113	-	TCF4 的 TNR 的 5'	-65	1169	UAGACCUUCU UUUGGAGAAA	581837	13
82	TCCAAAAGAAGGT CTAGAAAGG	Chr18:55586098-55586120	+	TCF4 的 TNR 的 5'	-58	1170	UCCAAAAGAA GGUCUAGAAG	146767	21.4
83	TCCTCTCTAGAC CTTCTTTTGG	Chr18:55586099-55586121	-	TCF4 的 TNR 的 5'	-57	1171	UCCUCUUCUAG ACCUUCUUU	5256.5	29.4
84	AAAAGAAGGTCTA GAAGAGGAGG	Chr18:55586101-55586123	+	TCF4 的 TNR 的 5'	-55	1172	AAAAGAAGGU CUAGAAGAGG	103010	23.23
85	AGAAAGTCTAGAA GAGGAGGAGG	Chr18:55586104-55586126	+	TCF4 的 TNR 的 5'	-52	1173	AGAAAGUCUA GAAAGAGGAGG	104079	31.1
86	AGGTCTAGAAGAG GAGGAGGAGG	Chr18:55586107-55586129	+	TCF4 的 TNR 的 5'	-49	1174	AGGUCUAGAA GAGGAGGAGG	2449.4	39.2
87	TCTAGAAGAGGAG GAGGAGGAGG	Chr18:55586110-55586132	+	TCF4 的 TNR 的 5'	-46	1175	UCUAGAAGAG GAGGAGGAGG	1657.4	8.33
88	AGAGGAGGAGGA GGAGGAGAAAGG	Chr18:55586116-55586138	+	TCF4 的 TNR 的 5'	-40	1176	AGAGGAGGAG GAGGAGGAGA	773.69	15.67
89	GGAGGAGGAGGA GGAGAAGGAGG	Chr18:55586119-55586141	+	TCF4 的 TNR 的 5'	-37	1177	GGAGGAGGAG GAGGAGAAGG	420.41	17.23
90	GGAGGAGGAGGA GAAAGGAGGAGG	Chr18:55586122-55586144	+	TCF4 的 TNR 的 5'	-34	1178	GGAGGAGGAG GAGAAGGAGG	394.07	8.03

[0466]

SEQ ID NO	靶序列 (包括 PAM)	染色体位置	链	方向	与 TNR 起点的距离	SEQ ID No	向导序列	CFD 得分	平均编辑百分比
91	GGAGGAGGAGAA GGAGGAGGAGG	Chr18:55586125-55586147	+	TCF4 的 TNR 的 5'	-31	1179	GGAGGAGGAG AAGGAGGAGG	947.52	5.03
92	GGAGGAGAAAGGA GGAGGAGGAGG	Chr18:55586128-55586150	+	TCF4 的 TNR 的 5'	-28	1180	GGAGGAGAAAG GAGGAGGAGG	448.19	5.73
93	GGAGAAGGAGGA GGAGGAGGAGG	Chr18:55586131-55586153	+	TCF4 的 TNR 的 5'	-25	1181	GGAGAAGGAG GAGGAGGAGG	598.33	6
94	CAGCATGAAAGAG CCCCACTTGG	Chr18:55586225-55586247	+	TCF4 的 TNR 的 3'	69	1182	CAGCAUGAAA GAGCCCCACU	6355.32	18.63
95	ATGAAAAGAGCCCC ACTTGAAGG	Chr18:55586229-55586251	+	TCF4 的 TNR 的 3'	73	1183	AUGAAAAGAGC CCCACUUGGA	697.17	26.83
96	AAAGAGCCCCACT TGGAAGGCGG	Chr18:55586232-55586254	+	TCF4 的 TNR 的 3'	76	1184	AAAGAGCCCCA CUUGGAAGG	130.15	22.7
97	GCCCCACTTGGAA GGCGTTTGG	Chr18:55586237-55586259	+	TCF4 的 TNR 的 3'	81	1185	GCCCCACUUGG AAGCGGCUU	203.63	6.7
98	TCCAAAACCGCCTT CCAAAGTGGG	Chr18:55586238-55586260	-	TCF4 的 TNR 的 3'	82	1186	UCCAAAACCGCC UCCCAAGUG	203.16	8.07
99	ATCCAAAACCGCCT TCCAAAGTGGG	Chr18:55586239-55586261	-	TCF4 的 TNR 的 3'	83	1187	AUCCAAAACCGC CUUCCAAAGU	105.14	11.4
100	AATCCAAACCGCC TTCCAAAGTGG	Chr18:55586240-55586262	-	TCF4 的 TNR 的 3'	84	1188	AAUCCAAAACCG CCUCCCAAG	160.67	18.07
101	GATTTAATTGTGT GTTTTGTGG	Chr18:55586259-55586281	+	TCF4 的 TNR 的 3'	103	1189	GAUUUUUUUU GUGUGUUUUU	329.17	0.23
102	CATCTTACACCAA ACTCATCTGG	Chr18:55586308-55586330	+	TCF4 的 TNR 的 3'	152	1190	CAUCUUACACC AAACUCAUC	405.23	12.2
103	TTTTTAATGCCAGA TGAGTTTGG	Chr18:55586317-55586339	-	TCF4 的 TNR 的 3'	161	1191	UUUUUUUUUU CAGAUGAGUU	282.35	8.63
104	ATTCATTCCTCTGA CATGCTGG	Chr18:55586343-55586365	+	TCF4 的 TNR 的 3'	187	1192	AUUCAUUCUCC UGACAUGUC	2000.64	8.23
105	TTCATTCTCCTGAC ATGTCTGGG	Chr18:55586344-55586366	+	TCF4 的 TNR 的 3'	188	1193	UUCAUUCUCCU GACAUGUCU	35953.9	12.3

[0467]

SEQ ID NO	靶序列 (包括 PAM)	染色体位置	链	方向	与 TNR 起点的距离	SEQ ID No	向导序列	CFD 得分	平均编辑百分比
106	CTCCTGACATGTC TGGGACTTGG	Chr18:55586350-55586372	+	TCF4 的 TNR 的 3'	194	1194	CUCUGACAUG UCUGGGACU	683.98	7.03
107	AACCAAGTCCCAG ACATGTCAGG	Chr18:55586352-55586374	-	TCF4 的 TNR 的 3'	196	1195	AACCAAGUCCC AGACAUGUC	5020.06	22.2
108	ACATGCTGGGAC TTGGTTTAGG	Chr18:55586356-55586378	+	TCF4 的 TNR 的 3'	200	1196	ACAUGUCUGG GACUUGGUUU	1201.43	21.03
109	CTGGGACTTGGTT TAGGAAAAGG	Chr18:55586362-55586384	+	TCF4 的 TNR 的 3'	206	1197	CUGGGACUUG GUUUAGGAAA	1784.35	32
110	GGTTTAGGAAAAG GAAAGCAAAGG	Chr18:55586371-55586393	+	TCF4 的 TNR 的 3'	215	1198	GGUUUAGGAA AAGGAAGCAA	1362.04	11.57
111	GTTTAGGAAAAGG AAGCAAAGG	Chr18:55586372-55586394	+	TCF4 的 TNR 的 3'	216	1199	GUUUAGGAAA AGGAAGCAA	4810.53	12.17
112	AGAAAAGGGAAG CAAGGGATGG	Chr18:55586376-55586398	+	TCF4 的 TNR 的 3'	220	1200	AGGAAAAGGA AGCAAAGGGA	814.55	20.47
113	AGGAAAGCAAAGG GATGGAGAAAGG	Chr18:55586382-55586404	+	TCF4 的 TNR 的 3'	226	1201	AGGAAAGCAA GGGAUUGGAGA	878.55	16.2
114	TGGAGTTTTACGG CTGTACTTGG	Chr18:55586406-55586428	-	TCF4 的 TNR 的 3'	250	1202	UGGAGUUUUA CGGCUGUACU	315.87	25.63
115	GACACACTTGTGG AGTTTTACGG	Chr18:55586416-55586438	-	TCF4 的 TNR 的 3'	260	1203	GACACACUUG UGGAGUUUUA	177.25	20.47
116	AGCGGAACCTTGAC ACACTTGTGG	Chr18:55586426-55586448	-	TCF4 的 TNR 的 3'	270	1204	AGCGGAACUU GACACACUUG	135.84	17.3
117	GTCCGTAGGATCAG CACAAAAGCGG	Chr18:55586444-55586466	-	TCF4 的 TNR 的 3'	288	1205	GUCGUAGGAU CAGCACAAAAG	797.01	20.3
118	TTGGTAAATTCGT AGTCGTAGG	Chr18:55586459-55586481	-	TCF4 的 TNR 的 3'	303	1206	UUGGUAAAUU UCGUAGUCGU	200.12	9.3
119	ATTTACCAAAACA GTCCAAAAGG	Chr18:55586473-55586495	+	TCF4 的 TNR 的 3'	317	1207	AUUUACCCAAA ACAGUCCAAA	1602.25	NA
120	TAGAACTTTTGG ACTGTTTTGG	Chr18:55586478-55586500	-	TCF4 的 TNR 的 3'	322	1208	UAGAACCUUU UGGACUGUUU	5716.11	5

[0468]

SEQ ID NO	靶序列 (包括 PAM)	染色体位置	链	方向	与 TNR 起点的距离	SEQ ID No	向导序列	CFD 得分	平均编辑百分比
121	ATACATTCCTTAGA ACCTTTTGG	Chr18:55586488-55586510	-	TCF4 的 TNR 的 3'	332	1209	AUACAUCUU UAGAACCUCUU	345.52	7.5
122	TAGGATTCCTAAA ACTAGTATGG	Chr18:55586522-55586544	-	TCF4 的 TNR 的 3'	366	1210	UAGGAUCUU AAAACUAGUA	1052.1	1.83
123	ATACTAGTTTTAAG AATCCTAGG	Chr18:55586524-55586546	+	TCF4 的 TNR 的 3'	368	1211	AUACUAGUUU UAAAGAAUCCU	1437.3	10.03
124	TCCTAGGAAAAGA TGTAACCTAGG	Chr18:55586540-55586562	+	TCF4 的 TNR 的 3'	384	1212	UCCUAGGAAA AGAUGUAACU	2172.5	20.9
125	TCCTAGTTACATCT TTTCCTAGG	Chr18:55586541-55586563	-	TCF4 的 TNR 的 3'	385	1213	UCCUAGUUAC AUCUUUCCU	1136.6	15.03
126	TAGGAAAAGATGT AACTAGGAGG	Chr18:55586543-55586565	+	TCF4 的 TNR 的 3'	387	1214	UAGGAAAAGA UGUAACUAGG	1044.9	23.3
127	TAACTAGGAGGTA AGATGTAAGG	Chr18:55586555-55586577	+	TCF4 的 TNR 的 3'	399	1215	UAAUCUAGGAG GUAAAGAUGUA	707.33	22.5
128	GGAGGTAAGATGT AAGGAACAGG	Chr18:55586561-55586583	+	TCF4 的 TNR 的 3'	405	1216	GGAGGUAAGA UGUAAGGAAC	473.79	16.03
129	TAATGATGCTTTGG ATTGGTAGG	Chr18:55586585-55586607	-	TCF4 的 TNR 的 3'	429	1217	UAAUUGAUCU UUGGAUUGGU	7.55	19.93
130	AAGCTAATGATGC TTTGGATTGG	Chr18:55586589-55586611	-	TCF4 的 TNR 的 3'	433	1218	AAGCUAAUGA UGC UUUGGAU	48.63	15.27
131	GTTTTAAGCTAAT GATGCTTTGG	Chr18:55586594-55586616	-	TCF4 的 TNR 的 3'	438	1219	GUUUUAAAGCU AAUGAUGCUU	1051.2	3.67
132	TAAAACTTTAAAG AGACAACCTGG	Chr18:55586611-55586633	+	TCF4 的 TNR 的 3'	455	1220	UAAAACUUUA AAGAGACAAC	83.63	12.03
133	AAAACTTTAAAGA GACAACCTGGG	Chr18:55586612-55586634	+	TCF4 的 TNR 的 3'	456	1221	AAAACUUUAA AGAGACAACU	841.09	32.53
134	GGAAATGGAAAAT AGAAAATAGG	Chr18:55586638-55586660	-	TCF4 的 TNR 的 3'	482	1222	GGAAAUGGAA AAUAGAAA AU	22.4	13.73
135	TTAATTAATGTTTT TGGAAAATGG	Chr18:55586663-55586675	-	TCF4 的 TNR 的 3'	497	1223	UUUUUUUUUG UUUUUUGGAAA	2366.7	0.13

[0469]

SEQ ID NO	靶序列 (包括 PAM)	染色体位置	链	方向	与 TNR 起点的距离	SEQ ID No	向导序列	CFD 得分	平均编辑百分比
136	TTCGTTTTATTAT TGTTTTTGG	Chr18:55586659-55586681	-	TCF4 的 TNR 的 3'	503	1224	UUCGUUUUUAU UUUUUUUUUU	1039.9 5	0.07
137	GTAGTCTCAGTGT TCAGACATGG	Chr18:55586702-55586724	+	TCF4 的 TNR 的 3'	546	1225	GUAGUCUCAG UGUUCAGACA	1965.7 9	5.37
138	TTCAGACATGGCC AAGTTTTAGG	Chr18:55586714-55586736	+	TCF4 的 TNR 的 3'	558	1226	UUCAGACAUG GCCAAGUUUU	3320.5	2.33
139	TCAGACATGGCCA AGTTTTAGG	Chr18:55586715-55586737	+	TCF4 的 TNR 的 3'	559	1227	UCAGACAUGG CCAAGUUUUUA	717.05	5.9
140	CAGACATGGCCAA GTTTTAGGG	Chr18:55586716-55586738	+	TCF4 的 TNR 的 3'	560	1228	CAGACAUGGCC AAGUUUUUAG	300.9	6.37
141	ACATGGCCAAAGTT TTAGGGGTGG	Chr18:55586719-55586741	+	TCF4 的 TNR 的 3'	563	1229	ACAUGGCCAA GUUUUAGGGG	301.24	12.73
142	ACTAAACACCC TAAAACTGG	Chr18:55586725-55586747	-	TCF4 的 TNR 的 3'	569	1230	ACUAAAACCACC CCUAAAACU	333.64	1.57
143	TTTAGGGGTGGTT TAGTTTTAGG	Chr18:55586731-55586753	+	TCF4 的 TNR 的 3'	575	1231	UUUAGGGGUG GUUUAGUUUU	171.1	3.2
144	TTAGGGGTGGTTT AGTTTTAGG	Chr18:55586732-55586754	+	TCF4 的 TNR 的 3'	576	1232	UUAGGGGUGG UUUAGUUUUUA	214.26	6.8
145	TAGGGTGGTTTA GTTTTAGGG	Chr18:55586733-55586755	+	TCF4 的 TNR 的 3'	577	1233	UAGGGGUGGU UUAGUUUUAG	147.48	10.37
146	TGCTATTTTTGCT TTCCACTGG	Chr18:55586756-55586778	+	TCF4 的 TNR 的 3'	600	1234	UGUCU AUUUU UGUUUCCAC	995.21	4.33
147	GTCTATTTTTGCTT TCCACTGGG	Chr18:55586757-55586779	+	TCF4 的 TNR 的 3'	601	1235	GUCU AUUUU GCUUCCACU	174.31	1.7
148	TCTATTTTTGCTTT CCACTGGGG	Chr18:55586758-55586780	+	TCF4 的 TNR 的 3'	602	1236	UCU AUUUUUG CUUCCACUG	84.57	5.7
149	ATAATGGAATCTC ACCCAGTGG	Chr18:55586772-55586794	-	TCF4 的 TNR 的 3'	616	1237	AUAAUGGAAU CUCACCCAG	298.73	14.83
150	TGGGTGAGATTC CATTATTGG	Chr18:55586776-55586798	+	TCF4 的 TNR 的 3'	620	1238	UGGGGUGAGA UUCCAUAUU	2434.8 9	4.53

[0470]

SEQ ID NO	靶序列 (包括 PAM)	染色体位置	链	方向	与 TNR 起点的距离	SEQ ID No	向导序列	CFD 得分	平均编辑百分比
151	GGGGTGAGATTCC AATATTGGG	Chr18:55586777-55586799	+	TCF4 的 TNR 的 3'	621	1239	GGGGUGAGAU UCCAUUAAUUU	1205.0 2	4.8
152	GGGTGAGATTCCA TTATTGGGG	Chr18:55586778-55586800	+	TCF4 的 TNR 的 3'	622	1240	GGGUGAGAUU CCAUAUUUUG	2784.1 4	4.63
153	CCATTATTGGGGT AATCAGTGG	Chr18:55586788-55586810	+	TCF4 的 TNR 的 3'	632	1241	CCAUUAAUUUG GGGUAUUCAG	978.57	17.53
154	CCACTGATTACCC CAAATAATGG	Chr18:55586788-55586810	-	TCF4 的 TNR 的 3'	632	1242	CCACUGAUUAC CCCAAAUAA	42.74	12.17
155	CATTATTGGGGTA ATCAGTGG	Chr18:55586789-55586811	+	TCF4 的 TNR 的 3'	633	1243	CAUUAUUUGG GGUAAUCAGU	1266.0 8	19.47
156	ATTGGGGTAATC AGTGGTAGG	Chr18:55586793-55586815	+	TCF4 的 TNR 的 3'	637	1244	AUUUGGGGUA AUCAGUGGGU	251.48	6.2
157	TTTGGGGTAATCA GTGGGTAGG	Chr18:55586794-55586816	+	TCF4 的 TNR 的 3'	638	1245	UUUGGGGUAA UCAGUGGGUA	443.03	8.7
158	ATCAGTGGGTAGG GAATGAAGG	Chr18:55586803-55586825	+	TCF4 的 TNR 的 3'	647	1246	AUCAGUGGGU AGGGAUUUGA	616.38	7.2
159	TTTTTTTGAGTTT TATTACTGG	Chr18:55586826-55586848	-	TCF4 的 TNR 的 3'	670	1247	UUUUUUUUGA GUUUUAAUAC	843.87	1.1
160	TGTGGTGTGATGG AAGATTCAGG	Chr18:55586856-55586878	-	TCF4 的 TNR 的 3'	700	1248	UGUGGUGUGA UGGAAGAUAUC	565.01	6.47
161	ACTATAATTTTGTG GTGTGATGG	Chr18:55586866-55586888	-	TCF4 的 TNR 的 3'	710	1249	ACUAAAUUUU UGUGGUGUGA	4828.9 7	0.5
162	AGTTTTAACTATA ATTTTGTGG	Chr18:55586874-55586896	-	TCF4 的 TNR 的 3'	718	1250	AGUUUUUUAAC UAUAAUUUUG	339.02	1.1
163	AAAGACCTTCATA TTTACCAAGG	Chr18:55586903-55586925	+	TCF4 的 TNR 的 3'	747	1251	AAAGACCUUC AUUUUACCA	142.27	5.87
164	TGAATCCTTGGTA AATATGAAGG	Chr18:55586908-55586930	-	TCF4 的 TNR 的 3'	752	1252	UGAAUCCUUG GUAAAUAUGA	789.33	3.17
165	TTTTTAATGGCTG AATCCTTGG	Chr18:55586920-55586942	-	TCF4 的 TNR 的 3'	764	1253	UUUUUAAUUG GCUGAAUCCU	3433.0 8	8.07

[0471]

SEQ ID NO	靶序列 (包括 PAM)	染色体位置	链	方向	与 TNR 起点的距离	SEQ ID No	向导序列	CFD 得分	平均编辑百分比
166	GGACAGTAATAAT TTTTAATTGG	Chr18:55586932-55586954	-	TCF4 的 TNR 的 3'	776	1254	GGACAGUAAU AAUUUUUUAAU	187.99	0.83
167	ACTGTCCCTTAGAT TCCTACTGG	Chr18:55586948-55586970	+	TCF4 的 TNR 的 3'	792	1255	ACUGUCCUUU AGAUUCCUAC	3697.8 1	8.13
168	AGAAACCAGTAGG AATCTAAAGG	Chr18:55586953-55586975	-	TCF4 的 TNR 的 3'	797	1256	AGAAACCAGU AGGAUUCUAA	1485.3 6	5.8
169	CACTTCAGCTAGA AACCAGTAGG	Chr18:55586963-55586985	-	TCF4 的 TNR 的 3'	807	1257	CACUUCAGCUA GAAACCAGU	1419.4 3	7.7
170	TGGTTTCTAGCTG AAGTGTITGG	Chr18:55586968-55586990	+	TCF4 的 TNR 的 3'	812	1258	UGGUUUCUAG CUGAAGUGUU	1064.1 1	6.83
171	GGTTTCTAGCTGA AGTGTITGG	Chr18:55586969-55586991	+	TCF4 的 TNR 的 3'	813	1259	GGUUUCUAGC UGAAGUGUUU	742.1	8.47
172	AGTGCGGTAAGAA AGAACGGTGG	Chr18:55587028-55587050	-	TCF4 的 TNR 的 3'	872	1260	AGUGCGGUAAA GAAAGAACGG	1308.2	23.43
173	TTCAGTGCGGTAA GAAAGAACGG	Chr18:55587031-55587053	-	TCF4 的 TNR 的 3'	875	1261	UUCAGUGCGG UAAAGAAAAGAA	833.82	23.33
174	TGATTTACTGGATT TCAGTGCGG	Chr18:55587044-55587066	-	TCF4 的 TNR 的 3'	888	1262	UGAUUUACUG GAUUUCAGUG	1281.4 7	NA
175	CAAAGAGCTGAGT GATTTACTGG	Chr18:55587056-55587078	-	TCF4 的 TNR 的 3'	900	1263	CAAAGAGCUG AGUGAUUUAC	1093.0 5	NA
176	CAGCTCTTTGTCC GTCCCTAAGG	Chr18:55587069-55587091	+	TCF4 的 TNR 的 3'	913	1264	CAGCUUUUUG UCCGUCCCUA	2384.9 5	NA
177	GCGAATGGCTGCC TTAGGGACGG	Chr18:55587080-55587102	-	TCF4 的 TNR 的 3'	924	1265	GCGAAUGGCU GCCUUAGGGA	136.05	NA
178	AACAGCGAATGGC TGCCCTTAGGG	Chr18:55587084-55587106	-	TCF4 的 TNR 的 3'	928	1266	AACAGCGAAU GGCUGCCUUA	1946.7 6	NA
179	CAACAGCGAATGG CTGCCTTAGG	Chr18:55587085-55587107	-	TCF4 的 TNR 的 3'	929	1267	CAACAGCGAA UGGCUGCCUU	922.31	NA
180	CTAAGGCAGCCAT TCGCTGTTGG	Chr18:55587086-55587108	+	TCF4 的 TNR 的 3'	930	1268	CUAAGGCAGCC AUUCGCUGU	1288.5 9	NA

[0472]

SEQ ID NO	靶序列 (包括 PAM)	染色体位置	链	方向	与 TNR 起点的距离	SEQ ID No	向导序列	CFD 得分	平均编辑百分比
181	AATGCATCACCAA CAGCGAATGG	Chr18:55587095-55587117	-	TCF4 的 TNR 的 3'	939	1269	AAUGCAUCACC AACACGCGAA	221.14	NA
182	ATCACACAAACCT AGAAACATGG	Chr18:55587126-55587148	+	TCF4 的 TNR 的 3'	970	1270	AUCACACAAAAC CUAGAAACA	1315.96	NA
183	GCGGTTAFTTCCAT GTTTCTAGG	Chr18:55587136-55587158	-	TCF4 的 TNR 的 3'	980	1271	GCGGUUAUUU CCAUGUUUCU	1600	NA
184	GGGACTGGATTT CTGATTGCGG	Chr18:55587155-55587177	-	TCF4 的 TNR 的 3'	999	1272	GGGACUGGAU UUUCUGAUUG	1287.34	NA
185	GAAAATCCAGTCC CAATCCTTGG	Chr18:55587164-55587186	+	TCF4 的 TNR 的 3'	1008	1273	GAAAAUCCAG UCCCAAUCCU	1557.06	NA
186	TTTTCTCCAAGGA TTGGGACTGG	Chr18:55587170-55587192	-	TCF4 的 TNR 的 3'	1014	1274	UUUUCUCCAA GGAUUGGGAC	1644.63	NA
187	TTGTGTTTCTCCA AGGATTGGG	Chr18:55587175-55587197	-	TCF4 的 TNR 的 3'	1019	1275	UUGUGUUUUC UCCAAAGGAU	495.78	NA
188	ATTGTGTTTCTCC AAGGATTGG	Chr18:55587176-55587198	-	TCF4 的 TNR 的 3'	1020	1276	AUUGUGUUUU CUCCAAAGGAU	2305.18	NA
189	ATCCTTGGAGAAA ACACAATCGG	Chr18:55587179-55587201	+	TCF4 的 TNR 的 3'	1023	1277	AUCCUUGGAG AAAACACAAU	527.93	NA
190	ATCCGATTGTGTTT TCTCCAAGG	Chr18:55587181-55587203	-	TCF4 的 TNR 的 3'	1025	1278	AUCCGAUUGU GUUUUCUCCA	125.71	NA

[0473] 以96孔型式筛选具有表2中提供的向导序列的gRNA来确定它们的编辑(例如插入/缺失形成)效率。为此,将组成性表达Spy Cas9的HEK293细胞系(“HEK293\_Cas9”)培养在补

充了10%胎牛血清和500 $\mu$ g/ml G418的DMEM培养基中。转染前20小时,将细胞按10,000细胞/孔的密度接种在96孔板中。按照厂家的流程用Lipofectamine RNAiMAX (ThermoFisher, 目录号13778150) 转染细胞。用包含单个crRNA (25nM)、trRNA (25nM)、Lipofectamine RNAiMAX (0.3 $\mu$ L/孔) 和OptiMem的脂质复合物转染细胞。按照厂家的流程用50 $\mu$ L/孔 BuccalAmp DNA提取液 (Epicentre, 目录号QE09050) 从各孔提取基因组DNA。

[0474] 为了定量测定基因组中靶标位置处的编辑效率,利用深度测序来鉴定通过基因编辑引入的插入和缺失(“插入/缺失”)的存在。在靶位点附近设计PCR引物,扩增目的基因组区域。按照厂家流程(Illumina)进行附加PCR,以为了测序加入必要的化学品。在Illumina MiSeq仪器上测序扩增子。在消除具有低质量得分的那些之后,将读出与人参考基因组比对。将所得到的包含读出的文件定位至参考基因组(BAM文件),其中选择与目的靶区域重叠的读出,计算野生型读出的数目和包含插入、取代或缺失的读出的数目。编辑百分比(例如“编辑效率”或“百分比编辑”)定义为含有插入或缺失的序列读出总数除以包括野生型的序列读出总数。表2中报告所使用的各gRNA的编辑效率数字。

[0475] 完成以上初步评价来鉴定出具有最佳编辑效率的那些之后,筛选gRNA对来确定能够去除含有TNR的DNA的间插节段(如图1中所示)的gRNA对。切除间插节段后,然后细胞将通过非同源末端连接(NHEJ) DNA修复途径修复断裂,该途径甚至在诸如角膜内皮中的那些细胞的非分裂细胞中也高度有效。此过程紧随着两个向导序列之间的DNA片段的切除,其甚至在向导序列相隔>3000个核苷酸时也可以高频率发生。此编辑方法无需其他同源模板DNA,极大地简化了该方法。由于所删除的范围包含在内含子内,预期对TCF4的基因产物无影响,因为内含子不影响最终的mRNA产物或蛋白质产物。

[0476] 去除TNR重复后,TCF4RNA转录物应不在在细胞内聚集,也不整合正常细胞功能所需的剪接因子。内含子3内相关区域的去除不太可能对RNA稳定性或TCF4基因本身的表达有任何有害作用,因为此内含子正常情况下将在最终RNA产物成熟期间通过RNA剪接去除。因此,内含子3内的DNA区域不包含在用于翻译TCF4蛋白质的最终RNA产物内。没有了TNR,TCF4的mRNA和基因产物应正常发挥功能,与具有最低限度TNR扩增的正常等位基因基本一样。此外,由于角膜内皮细胞基本不分裂,一次矫正细胞应导致疾病的永久改善。治疗应阻止疾病特征性的异常胶原沉积(即水滴),且随时间推移可导致现有水滴的吸收。还提出,用此技术治疗已知对FECD易感的个体(如具有疾病家族史并确认具有TCF4内含子3TNR扩增的个体)可以预防疾病的发展。

[0477] 为了证明TNR的切除,形成各具有靶向RNR一侧的gRNA的RNP对。简言之,通过按等摩尔量在95 $^{\circ}$ C水浴中加热crRNA和trRNA 2分钟并使它们在室温下冷却来制备50 $\mu$ M预退火gRNA(例如具有crRNA和trRNA的dgRNA)溶液。将预退火gRNA加至Spy Cas9蛋白质(按50 $\mu$ M浓度),并在室温孵育10分钟,得到具有3.33 $\mu$ M gRNA和1.66 $\mu$ M Cas9蛋白质的最终RNP溶液。将不组成性表达Cas9的HEK293细胞按~50,000细胞/孔在96孔型式中接种在20 $\mu$ L体积的SF电穿孔缓冲液(Lonza)中。向孔中加入5 $\mu$ l各RNP溶液(例如对于所测试的每一对),用Lonza Amaxa仪器电穿孔细胞。电穿孔后,向孔中加入80 $\mu$ L细胞培养基,将细胞转移至96孔平底组织培养板,37 $^{\circ}$ C孵育24小时。然后裂解细胞,并按上文所述提取基因组DNA。

[0478] 为了测定TNR切除效率,按上文针对编辑效率所述进行相似的NGS分析。简言之,进行深度测序来鉴定由TNR侧翼两个位置的基因编辑引起的缺失。在靶位点(TCF4内含子3中

的TNR)附近设计PCR引物,扩增目的基因组区域。按照厂家流程(Illumina)进行附加PCR,以为了测序加入必要的化学品。在Illumina MiSeq仪器上测序所得到的扩增子。过滤读出以消除具有低质量得分的那些,将所得到的读出定位至参考基因组。将与靶区域重叠的读出进一步过滤并局部重比对,以鉴定大的缺失。计算包含跨两个靶区域的缺失的读出数。将切除百分比定义为含有TNR缺失的测序读出数除以与靶区域重叠的读出总数。表7中报告所测试的每一对的切除百分比。

[0479] 如表7和图2中所示,测试了93对gRNA,一些gRNA对达到大于80%切除,一对gRNA尤其达到超过88%切除(例如使用具有指导核酸酶至包含SEQ ID NO:83和SEQ ID NO:109的靶序列的向导序列的gRNA;分别对应于包含SEQ ID NO:1177和SEQ ID NO:1197的向导RNA)。

[0480] 表7

SEQ ID NO (5' 靶序列)	SEQ ID NO (3' 靶序列)	切除百分比
83	109	88.71
85	109	85.56
86	112	81.58
85	125	81.08
86	109	79.99
85	107	78.44
83	125	76.78
86	125	76.67
86	107	71.68
64	106	66.1
85	114	65.86
86	114	61.58
83	114	59.88
53	114	43.8
83	112	27.6
74	114	20.7
85	108	7.35
83	107	6.69
85	115	6.44
58	109	5.69
86	108	5.57
83	96	5.17
74	109	4.46
77	115	4.45
53	96	4.44
83	108	4.4
74	125	4.3
85	94	4.17
86	96	3.53
53	107	3.42

[0481]

[0482]

SEQ ID NO (5' 靶序列)	SEQ ID NO (3' 靶序列)	切除百分比
83	94	3.21
71	115	3.21
77	96	3.12
58	112	3.11
77	109	3.08
85	95	3
53	94	2.9
77	95	2.82
86	115	2.75
85	96	2.65
58	94	2.61
58	115	2.61
71	96	2.56
58	107	2.53
83	95	2.43
58	96	2.36
77	94	2.24
56	94	2.21
77	108	2.17
77	112	2.16
86	94	2.08
77	107	1.9
86	95	1.87
56	96	1.87
54	94	1.72
71	94	1.69
77	114	1.65
71	114	1.64
56	95	1.63
58	95	1.5
53	112	1.32
71	109	1.3
74	112	1.28
54	96	1.17
58	114	1.15
74	108	1.09
53	108	0.79
74	107	0.62
74	94	0.61
71	107	0.56
71	95	0.55
71	112	0.55
74	96	0.47
74	95	0.46

SEQ ID NO (5' 靶序列)	SEQ ID NO (3' 靶序列)	切除百分比
74	115	0.41
54	95	0.37
53	95	0.35
77	125	0.33
54	112	0.09
56	114	0.01
73	101	0.01
54	109	0
54	114	0
[0483] 54	107	0
54	108	0
54	115	0
56	109	0
56	107	0
56	108	0
56	112	0
56	115	0
56	125	0
53	125	0

[0484] 实施例2. 用gRNA治疗COL8A2中的突变

[0485] COL8A2中的三个突变 (Gln455Lys、Gln455Val和Leu450Trp) 与早发型FECD和多形性角膜后层营养不良 (PPDC) 相关, 敲入动物研究已显示与人早发型FECD一致的病理。这些模型与三螺旋结构稳定性改变的突变胶原VIII肽的异常胞内累积相关。因此, 减少具有COL8A2中突变的诊断或家族史的患者中的突变胶原VIII可以改善病程。备选地, 选择性降低含有突变Gln455Lys、Gln455Val或Leu450Trp的COL8A2的水平可以降低突变胶原VIII肽的水平, 并改善病程。另一种方法是矫正DNA中导致胶原VIII $\alpha$ 亚基2 (COL8A2) 中的氨基酸突变的突变, 从而去除异常基因产物。

[0486] 用COL8A2基因转录物变体1的NCBI参考序列NM\_005202.3选择靶序列, 用于发展Cas RNP治疗。此序列不包含已知发生在胶原VIII基因产物氨基酸序列中455和450位的突变, 可以称为“野生型COL8A2序列”。在Chr1:36097532-36100270 (hg38版本) 之间选择表3中所列的靶序列 (SEQ ID NO:191-1063)。可以用与靶序列互补的向导序列来产生用于RNP的gRNA来靶向COL8A2。

[0487]

表 3: 野生型 COL8A2 的靶序列			
SEQ ID No	染色体位置	链	靶序列
191	Chr1:36097532-36097554	+	GGGGAGGAGGCCAGGGCAGCAGG
192	Chr1:36097545-36097567	+	GGGCAGCAGGACCCCCCGCGG
193	Chr1:36097546-36097568	+	GGCAGCAGGACCCCCCGCGGG
194	Chr1:36097554-36097576	+	GACCCCCCGCGGGTTATGTGG
195	Chr1:36097555-36097577	+	ACCCCCCGCGGGTTATGTGGG
196	Chr1:36097556-36097578	+	CCCCCCCCGCGGGTTATGTGGGG
197	Chr1:36097556-36097578	-	CCCCACATAACCCGCGGGGGGG
198	Chr1:36097557-36097579	-	GCCCCACATAACCCGCGGGGGGG
199	Chr1:36097558-36097580	-	TGCCCCACATAACCCGCGGGGGG
200	Chr1:36097559-36097581	-	CTGCCCCACATAACCCGCGGGGG
201	Chr1:36097560-36097582	-	TCTGCCCCACATAACCCGCGGGG
202	Chr1:36097561-36097583	-	CTCTGCCCCACATAACCCGCGGG
203	Chr1:36097562-36097584	-	GCTCTGCCCCACATAACCCGCGG
204	Chr1:36097578-36097600	+	GCAGAGCAAGAATCCTGAAAAGG
205	Chr1:36097581-36097603	+	GAGCAAGAATCCTGAAAAGGAGG
206	Chr1:36097586-36097608	+	AGAATCCTGAAAAGGAGGAGTGG
207	Chr1:36097591-36097613	-	TACATCCACTCCTCCTTTTCAGG
208	Chr1:36097599-36097621	+	GGAGGAGTGGATGTACTCCGTGG
209	Chr1:36097607-36097629	+	GGATGTACTCCGTGGAGTAGAGG
210	Chr1:36097614-36097636	+	CTCCGTGGAGTAGAGGCCGTTGG
211	Chr1:36097616-36097638	-	GGCCAACGGCCTCTACTCCACGG
212	Chr1:36097619-36097641	+	TGGAGTAGAGGCCGTTGGCCTGG

[0488] P98

[0489]

213	Chr1:36097627-36097649	+	AGGCCGTTGGCCTGGTCCGACGG
214	Chr1:36097630-36097652	-	ATGCCGTCGGACCAGGCCAACGG
215	Chr1:36097637-36097659	-	GGTGCAGATGCCGTCGGACCAGG
216	Chr1:36097643-36097665	-	GGTCTGGGTGCAGATGCCGTCGG
217	Chr1:36097646-36097668	+	ACGGCATCTGCACCCAGACCTGG
218	Chr1:36097653-36097675	+	CTGCACCCAGACCTGGTCGTTGG
219	Chr1:36097654-36097676	+	TGCACCCAGACCTGGTCGTTGGG
220	Chr1:36097658-36097680	-	GCGGCCCAACGACCAGGTCTGGG
221	Chr1:36097659-36097681	-	TGCGGCCCAACGACCAGGTCTGG
222	Chr1:36097664-36097686	+	CCTGGTCGTTGGGCCGCAGCTGG
223	Chr1:36097664-36097686	-	CCAGCTGCGGCCCAACGACCAGG
224	Chr1:36097671-36097693	+	GTTGGGCCGCAGCTGGAGCACGG
225	Chr1:36097677-36097699	-	GTGGGGCCGTGCTCCAGCTGCGG
226	Chr1:36097688-36097710	+	GCACGGCCCCACCAGATGCCTGG
227	Chr1:36097694-36097716	+	CCCCACCAGATGCCTGGTCCAGG
228	Chr1:36097694-36097716	-	CCTGGACCAGGCATCTGGTGGGG
229	Chr1:36097695-36097717	-	ACCTGGACCAGGCATCTGGTGGG
230	Chr1:36097696-36097718	-	TACCTGGACCAGGCATCTGGTGG
231	Chr1:36097699-36097721	-	GGCTACCTGGACCAGGCATCTGG
232	Chr1:36097706-36097728	-	CAAGAAGGGCTACCTGGACCAGG
233	Chr1:36097712-36097734	-	TGAGTACAAGAAGGGCTACCTGG
234	Chr1:36097719-36097741	+	GCCCTTCTTGTACTIONCATCGTAGG
235	Chr1:36097720-36097742	-	ACCTACGATGAGTACAAGAAGGG

[0490]

236	Chr1:36097721-36097743	-	TACCTACGATGAGTACAAGAAGG
237	Chr1:36097725-36097747	+	CTTGTACTIONCATCGTAGGTATAGG
238	Chr1:36097728-36097750	+	GTACTCATCGTAGGTATAGGTGG
239	Chr1:36097732-36097754	+	TCATCGTAGGTATAGGTGGCCGG
240	Chr1:36097751-36097773	+	CCGGCACGTTGTTCTTGTACAGG
241	Chr1:36097751-36097773	-	CCTGTACAAGAACAACGTGCCGG
242	Chr1:36097752-36097774	+	CGGCACGTTGTTCTTGTACAGGG
243	Chr1:36097767-36097789	+	GTACAGGGCCACCCACACGTTGG
244	Chr1:36097775-36097797	-	CAAGGGCACCAACGTGTGGGTGG
245	Chr1:36097778-36097800	-	CGTCAAGGGCACCAACGTGTGGG
246	Chr1:36097779-36097801	-	ACGTCAAGGGCACCAACGTGTGG
247	Chr1:36097787-36097809	+	TGGTGCCCTTGACGTGCACATGG
248	Chr1:36097792-36097814	-	GCTTACCATGTGCACGTCAAGGG
249	Chr1:36097793-36097815	-	TGCTTACCATGTGCACGTCAAGG
250	Chr1:36097816-36097838	+	AAGTAGTAGACGCCGCCACAGG
251	Chr1:36097817-36097839	+	AGTAGTAGACGCCGCCACAGGG
252	Chr1:36097821-36097843	+	GTAGACGCCGCCACAGGGCAGG
253	Chr1:36097828-36097850	-	ATCTTACCTGCCCTGTGGGCGG
254	Chr1:36097831-36097853	-	GGCATCTTACCTGCCCTGTGGG
255	Chr1:36097832-36097854	-	TGGCATCTTACCTGCCCTGTGG
256	Chr1:36097836-36097858	+	AGGGCAGGTGAAGATGCCAGTGG
257	Chr1:36097840-36097862	+	CAGGTGAAGATGCCAGTGGCTGG
258	Chr1:36097841-36097863	+	AGGTGAAGATGCCAGTGGCTGGG

[0491]

259	Chr1:36097852-36097874	-	AGCGGCTACAACCCAGCCACTGG
260	Chr1:36097856-36097878	+	TGGCTGGGTTGTAGCCGCTGTGG
261	Chr1:36097870-36097892	-	ACTCTCTACAATGGCCACAGCGG
262	Chr1:36097874-36097896	+	TGTGGCCATTGTAGAGAGTCCGG
263	Chr1:36097879-36097901	-	TTTGACCGGACTCTCTACAATGG
264	Chr1:36097887-36097909	+	GAGAGTCCGGTCAAATTTACGCG
265	Chr1:36097888-36097910	+	AGAGTCCGGTCAAATTTACGGG
266	Chr1:36097893-36097915	-	GCATGCCCGTGAAATTTGACCGG
267	Chr1:36097899-36097921	+	AAATTTACGGGCATGCCCGAGG
268	Chr1:36097902-36097924	+	TTTACGGGCATGCCCGAGGCGG
269	Chr1:36097903-36097925	+	TTTACGGGCATGCCCGAGGCGGG
270	Chr1:36097904-36097926	+	TCACGGGCATGCCCGAGGCGGGG
271	Chr1:36097908-36097930	+	GGGCATGCCCGAGGCGGGGAAGG
272	Chr1:36097909-36097931	+	GGCATGCCCGAGGCGGGGAAGGG
273	Chr1:36097914-36097936	+	GCCCGAGGCGGGGAAGGGCGAGG
274	Chr1:36097915-36097937	-	ACCTCGCCCTTCCCCGCCTCGGG
275	Chr1:36097916-36097938	-	CACCTCGCCCTTCCCCGCCTCGG
276	Chr1:36097932-36097954	+	CGAGGTGAGCACCGCAGTGAAGG
277	Chr1:36097936-36097958	+	GTGAGCACCGCAGTGAAGGCCGG
278	Chr1:36097941-36097963	+	CACCGCAGTGAAGGCCGGTGTGG
279	Chr1:36097943-36097965	-	TGCCACACCGGCCTTCACTGCGG
280	Chr1:36097946-36097968	+	CAGTGAAGGCCGGTGTGGCATGG
281	Chr1:36097947-36097969	+	AGTGAAGGCCGGTGTGGCATGGG

[0492]

282	Chr1:36097955-36097977	-	GCTGTCTGCCCATGCCACACCGG
283	Chr1:36097975-36097997	+	AGCTCGCCCAGCCCAAACACTGTGG
284	Chr1:36097981-36098003	-	GGCAAGCCACAGTTTGGGCTGGG
285	Chr1:36097982-36098004	-	GGGCAAGCCACAGTTTGGGCTGG
286	Chr1:36097986-36098008	-	AGGGGGGCAAGCCACAGTTTGGG
287	Chr1:36097987-36098009	-	AAGGGGGGCAAGCCACAGTTTGG
288	Chr1:36097998-36098020	+	CTTGCCCCCTTGCCCAGCACGG
289	Chr1:36098002-36098024	-	GGTGCCGTGCTGGGCAAGGGGGG
290	Chr1:36098003-36098025	-	GGGTGCCGTGCTGGGCAAGGGGG
291	Chr1:36098004-36098026	-	AGGGTGCCGTGCTGGGCAAGGGG
292	Chr1:36098005-36098027	-	GAGGGTGCCGTGCTGGGCAAGGG
293	Chr1:36098006-36098028	-	GGAGGGTGCCGTGCTGGGCAAGG
294	Chr1:36098011-36098033	-	GGTGTGGAGGGTGCCGTGCTGGG
295	Chr1:36098012-36098034	-	CGGTGTGGAGGGTGCCGTGCTGG
296	Chr1:36098019-36098041	+	GGCACCTCCACACCGCCGTTGG
297	Chr1:36098020-36098042	+	GCACCCTCCACACCGCCGTTGGG
298	Chr1:36098023-36098045	-	CTGCCAACGGCGGTGTGGAGGG
299	Chr1:36098024-36098046	+	CCTCCACACCGCCGTTGGGCAGG
300	Chr1:36098024-36098046	-	CCTGCCAACGGCGGTGTGGAGG
301	Chr1:36098027-36098049	-	GCACCTGCCAACGGCGGTGTGG
302	Chr1:36098032-36098054	-	GGCTTGCACCTGCCAACGGCGG
303	Chr1:36098035-36098057	-	GCAGGCTTGCACCTGCCAACGG
304	Chr1:36098053-36098075	-	TTCGATGAGACTGGCATCGCAGG

[0493]

305	Chr1:36098055-36098077	+	TGCGATGCCAGTCTCATCGAAGG
306	Chr1:36098062-36098084	+	CCAGTCTCATCGAAGGCCCCAGG
307	Chr1:36098062-36098084	-	CCTGGGGCCTTCGATGAGACTGG
308	Chr1:36098063-36098085	+	CAGTCTCATCGAAGGCCCCAGGG
309	Chr1:36098064-36098086	+	AGTCTCATCGAAGGCCCCAGGGG
310	Chr1:36098071-36098093	+	TCGAAGGCCCCAGGGGCACCAGG
311	Chr1:36098072-36098094	+	CGAAGGCCCCAGGGGCACCAGGG
312	Chr1:36098073-36098095	+	GAAGGCCCCAGGGGCACCAGGGG
313	Chr1:36098074-36098096	+	AAGGCCCCAGGGGCACCAGGGGG
314	Chr1:36098078-36098100	-	GGGACCCCTGGTGCCCTGGGG
315	Chr1:36098079-36098101	-	CGGGACCCCTGGTGCCCTGGG
316	Chr1:36098080-36098102	+	CCAGGGGCACCAGGGGGTCCCGG
317	Chr1:36098080-36098102	-	CCGGGACCCCTGGTGCCCTGG
318	Chr1:36098081-36098103	+	CAGGGGCACCAGGGGGTCCCGGG
319	Chr1:36098082-36098104	+	AGGGGCACCAGGGGGTCCCGGGG
320	Chr1:36098083-36098105	+	GGGGCACCAGGGGGTCCCGGGGG
321	Chr1:36098088-36098110	+	ACCAGGGGGTCCCGGGGGCCCGG
322	Chr1:36098089-36098111	+	CCAGGGGGTCCCGGGGGCCCGGG
323	Chr1:36098089-36098111	-	CCCGGGCCCCCGGGACCCCTGG
324	Chr1:36098092-36098114	+	GGGGTCCCGGGGGCCCGGGAGG
325	Chr1:36098098-36098120	+	CCCGGGGGCCCGGGAGGCCCGG
326	Chr1:36098098-36098120	-	CCGGGGCCTCCCGGGCCCCCGGG
327	Chr1:36098099-36098121	-	TCCGGGGCCTCCCGGGCCCCCGG

[0494]

328	Chr1:36098101-36098123	+	GGGGCCCCGGGAGGCCCCGGAGG
329	Chr1:36098102-36098124	+	GGGGCCCCGGGAGGCCCCGGAGGG
330	Chr1:36098106-36098128	-	CGGGCCCTCCGGGGCCTCCCGGG
331	Chr1:36098107-36098129	-	ACGGGCCCTCCGGGGCCTCCCGG
332	Chr1:36098115-36098137	-	CTGGAATCACGGGCCCTCCGGGG
333	Chr1:36098116-36098138	+	CCCGGAGGGCCCGTGATTCCAGG
334	Chr1:36098116-36098138	-	CCTGGAATCACGGGCCCTCCGGG
335	Chr1:36098117-36098139	+	CCGGAGGGCCCGTGATTCCAGGG
336	Chr1:36098117-36098139	-	CCCTGGAATCACGGGCCCTCCGG
337	Chr1:36098118-36098140	+	CGGAGGGCCCGTGATTCCAGGGG
338	Chr1:36098125-36098147	+	CCCGTGATTCCAGGGGAGCCAGG
339	Chr1:36098125-36098147	-	CCTGGCTCCCCTGGAATCACGGG
340	Chr1:36098126-36098148	+	CCGTGATTCCAGGGGAGCCAGGG
341	Chr1:36098126-36098148	-	CCCTGGCTCCCCTGGAATCACGG
342	Chr1:36098134-36098156	+	CCAGGGGAGCCAGGGACCCCTGG
343	Chr1:36098134-36098156	-	CCAGGGGTCCCTGGCTCCCCTGG
344	Chr1:36098135-36098157	+	CAGGGGAGCCAGGGACCCCTGGG
345	Chr1:36098136-36098158	+	AGGGGAGCCAGGGACCCCTGGGG
346	Chr1:36098137-36098159	+	GGGGAGCCAGGGACCCCTGGGGG
347	Chr1:36098143-36098165	-	ACGGGGCCCCCAGGGGTCCCTGG
348	Chr1:36098145-36098167	+	AGGGACCCCTGGGGGCCCCGTGG
349	Chr1:36098146-36098168	+	GGGACCCCTGGGGGCCCCGTGGG
350	Chr1:36098150-36098172	-	TGGGCCACGGGGCCCCCAGGGG

[0495]

351	Chr1:36098151-36098173	-	CTGGGCCCACGGGGCCCCCAGGG
352	Chr1:36098152-36098174	-	GCTGGGCCCACGGGGCCCCCAGG
353	Chr1:36098160-36098182	-	CTGGCACGGCTGGGCCCACGGGG
354	Chr1:36098161-36098183	+	CCCGTGGGCCCAGCCGTGCCAGG
355	Chr1:36098161-36098183	-	CCTGGCACGGCTGGGCCCACGGG
356	Chr1:36098162-36098184	-	ACCTGGCACGGCTGGGCCCACGG
357	Chr1:36098169-36098191	-	CAGGGGAACCTGGCACGGCTGGG
358	Chr1:36098170-36098192	-	GCAGGGGAACCTGGCACGGCTGG
359	Chr1:36098174-36098196	-	GAGAGCAGGGGAACCTGGCACGG
360	Chr1:36098179-36098201	-	GAGGGGAGAGCAGGGGAACCTGG
361	Chr1:36098185-36098207	+	TCCCCTGCTCTCCCCTCTCCAGG
362	Chr1:36098186-36098208	+	CCCCTGCTCTCCCCTCTCCAGGG
363	Chr1:36098186-36098208	-	CCCTGGAGAGGGGAGAGCAGGGG
364	Chr1:36098187-36098209	+	CCCTGCTCTCCCCTCTCCAGGGG
365	Chr1:36098187-36098209	-	CCCCTGGAGAGGGGAGAGCAGGG
366	Chr1:36098188-36098210	+	CCTGCTCTCCCCTCTCCAGGGGG
367	Chr1:36098188-36098210	-	CCCCCTGGAGAGGGGAGAGCAGG
368	Chr1:36098194-36098216	+	CTCCCCTCTCCAGGGGGCCCTGG
369	Chr1:36098196-36098218	-	TGCCAGGGCCCCCTGGAGAGGGG
370	Chr1:36098197-36098219	-	CTGCCAGGGCCCCCTGGAGAGGG
371	Chr1:36098198-36098220	+	CCTCTCCAGGGGGCCCTGGCAGG
372	Chr1:36098198-36098220	-	CCTGCCAGGGCCCCCTGGAGAGG
373	Chr1:36098203-36098225	+	CCAGGGGGCCCTGGCAGGCCTGG

[0496]

374	Chr1:36098203-36098225	-	CCAGGCCTGCCAGGGCCCCCTGG
375	Chr1:36098211-36098233	-	AGGGGGAACCAGGCCTGCCAGGG
376	Chr1:36098212-36098234	-	AAGGGGGAACCAGGCCTGCCAGG
377	Chr1:36098216-36098238	+	GCAGGCCTGGTTCCCCCTTCAGG
378	Chr1:36098221-36098243	+	CCTGGTTCCCCCTTCAGGCCCGG
379	Chr1:36098221-36098243	-	CCGGGCCTGAAGGGGAACCAGG
380	Chr1:36098225-36098247	+	GTTCCCCCTTCAGGCCCGGCAGG
381	Chr1:36098228-36098250	-	AGGCCTGCCGGGCCTGAAGGGGG
382	Chr1:36098229-36098251	-	AAGGCCTGCCGGGCCTGAAGGGG
383	Chr1:36098230-36098252	-	CAAGGCCTGCCGGGCCTGAAGGG
384	Chr1:36098231-36098253	+	CCTTCAGGCCCGGCAGGCCTTGG
385	Chr1:36098231-36098253	-	CCAAGGCCTGCCGGGCCTGAAGG
386	Chr1:36098232-36098254	+	CTTCAGGCCCGGCAGGCCTTGGG
387	Chr1:36098233-36098255	+	TTCAGGCCCGGCAGGCCTTGGGG
388	Chr1:36098239-36098261	-	ATTGGGCCCCAAGGCCTGCCGGG
389	Chr1:36098240-36098262	-	TATTGGGCCCCAAGGCCTGCCGG
390	Chr1:36098242-36098264	+	GGCAGGCCTTGGGGCCCAATAGG
391	Chr1:36098243-36098265	+	GCAGGCCTTGGGGCCCAATAGGG
392	Chr1:36098248-36098270	-	GCTGGCCCTATTGGGCCCCAAGG
393	Chr1:36098251-36098273	+	TGGGGCCCAATAGGGCCAGCTGG
394	Chr1:36098256-36098278	-	AGGGTCCAGCTGGCCCTATTGGG
395	Chr1:36098257-36098279	-	CAGGGTCCAGCTGGCCCTATTGG
396	Chr1:36098258-36098280	+	CAATAGGGCCAGCTGGACCCTGG

[0497]

397	Chr1:36098266-36098288	+	CCAGCTGGACCCTGGAGTCCTGG
398	Chr1:36098266-36098288	-	CCAGGACTCCAGGGTCCAGCTGG
399	Chr1:36098267-36098289	+	CAGCTGGACCCTGGAGTCCTGGG
400	Chr1:36098275-36098297	-	TCAGGAATCCCAGGACTCCAGGG
401	Chr1:36098276-36098298	-	CTCAGGAATCCCAGGACTCCAGG
402	Chr1:36098277-36098299	+	CTGGAGTCCTGGGATTCCTGAGG
403	Chr1:36098278-36098300	+	TGGAGTCCTGGGATTCCTGAGGG
404	Chr1:36098284-36098306	-	AGGGGTCCCTCAGGAATCCCAGG
405	Chr1:36098288-36098310	+	GGATTCCTGAGGGACCCCTCAGG
406	Chr1:36098293-36098315	+	CCTGAGGGACCCCTCAGGCCAGG
407	Chr1:36098293-36098315	-	CCTGGCCTGAGGGGTCCCTCAGG
408	Chr1:36098302-36098324	+	CCCCTCAGGCCAGGCTGCCCAGG
409	Chr1:36098302-36098324	-	CCTGGGCAGCCTGGCCTGAGGGG
410	Chr1:36098303-36098325	+	CCCTCAGGCCAGGCTGCCCAGGG
411	Chr1:36098303-36098325	-	CCCTGGGCAGCCTGGCCTGAGGG
412	Chr1:36098304-36098326	-	TCCCTGGGCAGCCTGGCCTGAGG
413	Chr1:36098311-36098333	-	TTGGGGCTCCCTGGGCAGCCTGG
414	Chr1:36098319-36098341	-	AAGGTGACTTGGGGCTCCCTGGG
415	Chr1:36098320-36098342	-	AAAGGTGACTTGGGGCTCCCTGG
416	Chr1:36098328-36098350	-	TGGGGCAGAAAGGTGACTTGGGG
417	Chr1:36098329-36098351	-	CTGGGGCAGAAAGGTGACTTGGG
418	Chr1:36098330-36098352	+	CCAAGTCACCTTTCTGCCCCAGG
419	Chr1:36098330-36098352	-	CCTGGGGCAGAAAGGTGACTTGG

[0498]

420	Chr1:36098331-36098353	+	CAAGTCACCTTTCTGCCCCAGGG
421	Chr1:36098338-36098360	-	GCAGGAGCCCTGGGGCAGAAAGG
422	Chr1:36098346-36098368	-	CAGGGGTGGCAGGAGCCCTGGGG
423	Chr1:36098347-36098369	+	CCCAGGGCTCCTGCCACCCCTGG
424	Chr1:36098347-36098369	-	CCAGGGGTGGCAGGAGCCCTGGG
425	Chr1:36098348-36098370	-	ACCAGGGGTGGCAGGAGCCCTGG
426	Chr1:36098356-36098378	+	CCTGCCACCCCTGGTCCTCCAGG
427	Chr1:36098356-36098378	-	CCTGGAGGACCAGGGGTGGCAGG
428	Chr1:36098357-36098379	+	CTGCCACCCCTGGTCCTCCAGGG
429	Chr1:36098360-36098382	-	TCGCCCTGGAGGACCAGGGGTGG
430	Chr1:36098363-36098385	-	GGGTCGCCCTGGAGGACCAGGGG
431	Chr1:36098364-36098386	-	CGGGTCGCCCTGGAGGACCAGGG
432	Chr1:36098365-36098387	-	ACGGGTGCCCTGGAGGACCAGG
433	Chr1:36098371-36098393	-	GGTTTCACGGGTCGCCCTGGAGG
434	Chr1:36098374-36098396	+	CCAGGGCGACCCGTGAAACCCGG
435	Chr1:36098374-36098396	-	CCGGGTTTCACGGGTCGCCCTGG
436	Chr1:36098383-36098405	-	AAGGGTGAGCCGGGTTTCACGGG
437	Chr1:36098384-36098406	-	CAAGGGTGAGCCGGGTTTCACGG
438	Chr1:36098385-36098407	+	CGTGAAACCCGGCTCACCCCTGG
439	Chr1:36098386-36098408	+	GTGAAACCCGGCTCACCCCTGGG
440	Chr1:36098392-36098414	-	ACTGGGCCCAAGGGTGAGCCGGG
441	Chr1:36098393-36098415	-	AACTGGGCCCAAGGGTGAGCCGG
442	Chr1:36098395-36098417	+	GGCTCACCCCTGGGCCAGTTGG

[0499]

443	Chr1:36098401-36098423	+	CCCTTGGGCCCCAGTTGGTCCAGG
444	Chr1:36098401-36098423	-	CCTGGACCAACTGGGCCCAAGGG
445	Chr1:36098402-36098424	+	CCTTGGGCCCCAGTTGGTCCAGGG
446	Chr1:36098402-36098424	-	CCCTGGACCAACTGGGCCCAAGG
447	Chr1:36098403-36098425	+	CTTGGGCCCCAGTTGGTCCAGGGG
448	Chr1:36098404-36098426	+	TTGGGCCCCAGTTGGTCCAGGGGG
449	Chr1:36098409-36098431	-	ATGGACCCCCTGGACCAACTGGG
450	Chr1:36098410-36098432	-	CATGGACCCCCTGGACCAACTGG
451	Chr1:36098411-36098433	+	CAGTTGGTCCAGGGGGTCCATGG
452	Chr1:36098412-36098434	+	AGTTGGTCCAGGGGGTCCATGGG
453	Chr1:36098419-36098441	+	CCAGGGGGTCCATGGGCCCCAGG
454	Chr1:36098419-36098441	-	CCTGGGGCCCATGGACCCCCTGG
455	Chr1:36098428-36098450	-	AGGGGACTTCCTGGGGCCCATGG
456	Chr1:36098435-36098457	-	AGGTGAGAGGGGACTTCCTGGGG
457	Chr1:36098436-36098458	-	CAGGTGAGAGGGGACTTCCTGGG
458	Chr1:36098437-36098459	+	CCAGGAAGTCCCCTCTCACCTGG
459	Chr1:36098437-36098459	-	CCAGGTGAGAGGGGACTTCCTGG
460	Chr1:36098438-36098460	+	CAGGAAGTCCCCTCTCACCTGGG
461	Chr1:36098446-36098468	+	CCCCTCTCACCTGGGACCCCTGG
462	Chr1:36098446-36098468	-	CCAGGGGTCCCAGGTGAGAGGGG
463	Chr1:36098447-36098469	-	ACCAGGGGTCCCAGGTGAGAGGG
464	Chr1:36098448-36098470	-	AACCAGGGGTCCCAGGTGAGAGG
465	Chr1:36098455-36098477	-	GCTGGGAAACCAGGGGTCCCAGG

[0500]

466	Chr1:36098459-36098481	+	GGACCCCTGGTTTCCCAGCCAGG
467	Chr1:36098462-36098484	-	TGGCCTGGCTGGGAAACCAGGGG
468	Chr1:36098463-36098485	-	GTGGCCTGGCTGGGAAACCAGGG
469	Chr1:36098464-36098486	-	AGTGGCCTGGCTGGGAAACCAGG
470	Chr1:36098467-36098489	+	GGTTTCCCAGCCAGGCCACTAGG
471	Chr1:36098472-36098494	-	AGGGGCCTAGTGGCCTGGCTGGG
472	Chr1:36098473-36098495	-	CAGGGGCCTAGTGGCCTGGCTGG
473	Chr1:36098474-36098496	+	CAGCCAGGCCACTAGGCCCCCTGG
474	Chr1:36098477-36098499	-	TGACCAGGGGCCTAGTGGCCTGG
475	Chr1:36098482-36098504	-	CGAGGTGACCAGGGGCCTAGTGG
476	Chr1:36098490-36098512	-	CTGGCATTGAGGTGACCAGGGG
477	Chr1:36098491-36098513	+	CCCTGGTCACCTCGAATGCCAGG
478	Chr1:36098491-36098513	-	CCTGGCATTGAGGTGACCAGGG
479	Chr1:36098492-36098514	-	GCCTGGCATTGAGGTGACCAGG
480	Chr1:36098500-36098522	+	CCTCGAATGCCAGGCACTCCTGG
481	Chr1:36098500-36098522	-	CCAGGAGTGCCTGGCATTGAGG
482	Chr1:36098501-36098523	+	CTCGAATGCCAGGCACTCCTGGG
483	Chr1:36098502-36098524	+	TCGAATGCCAGGCACTCCTGGGG
484	Chr1:36098503-36098525	+	CGAATGCCAGGCACTCCTGGGGG
485	Chr1:36098509-36098531	-	GGAGGACCCCCAGGAGTGCCTGG
486	Chr1:36098512-36098534	+	GGCACTCCTGGGGGTCCTCCAGG
487	Chr1:36098518-36098540	-	GCAGGGCCTGGAGGACCCCCAGG
488	Chr1:36098527-36098549	-	AAGGGTGAGGCAGGGCCTGGAGG

[0501]

489	Chr1:36098530-36098552	+	CCAGGCCCTGCCTCACCCCTTAGG
490	Chr1:36098530-36098552	-	CCTAAGGGTGAGGCAGGGCCTGG
491	Chr1:36098535-36098557	-	CTGGGCCTAAGGGTGAGGCAGGG
492	Chr1:36098536-36098558	+	CCTGCCTCACCCCTTAGGCCCAGG
493	Chr1:36098536-36098558	-	CCTGGGCCTAAGGGTGAGGCAGG
494	Chr1:36098537-36098559	+	CTGCCTCACCCCTTAGGCCCAGGG
495	Chr1:36098538-36098560	+	TGCCTCACCCCTTAGGCCCAGGGG
496	Chr1:36098539-36098561	+	GCCTCACCCCTTAGGCCCAGGGGG
497	Chr1:36098540-36098562	-	GCCCCCTGGGCCTAAGGGTGAGG
498	Chr1:36098545-36098567	-	CGTGGGCCCCCTGGGCCTAAGGG
499	Chr1:36098546-36098568	-	ACGTGGGCCCCCTGGGCCTAAGG
500	Chr1:36098553-36098575	-	CTGGCAGACGTGGGCCCCCTGGG
501	Chr1:36098554-36098576	+	CCAGGGGGCCCACGTCTGCCAGG
502	Chr1:36098554-36098576	-	CCTGGCAGACGTGGGCCCCCTGG
503	Chr1:36098562-36098584	-	CAGGGCTTCCTGGCAGACGTGGG
504	Chr1:36098563-36098585	-	GCAGGGCTTCCTGGCAGACGTGG
505	Chr1:36098572-36098594	+	CCAGGAAGCCCTGCAGACCCAGG
506	Chr1:36098572-36098594	-	CCTGGGTCTGCAGGGCTTCCTGG
507	Chr1:36098580-36098602	-	CTGGACTTCCTGGGTCTGCAGGG
508	Chr1:36098581-36098603	+	CCTGCAGACCCAGGAAGTCCAGG
509	Chr1:36098581-36098603	-	CCTGGACTTCCTGGGTCTGCAGG
510	Chr1:36098582-36098604	+	CTGCAGACCCAGGAAGTCCAGGG
511	Chr1:36098583-36098605	+	TGCAGACCCAGGAAGTCCAGGGG

[0502]

512	Chr1:36098584-36098606	+	GCAGACCCAGGAAGTCCAGGGGG
513	Chr1:36098589-36098611	-	GGGGTCCCCCTGGACTTCCTGGG
514	Chr1:36098590-36098612	-	GGGGGTCCCCCTGGACTTCCTGG
515	Chr1:36098599-36098621	-	CAGGGTCTTGGGGGTCCCCCTGG
516	Chr1:36098602-36098624	+	GGGGGACCCCCAAGACCCTGTGG
517	Chr1:36098603-36098625	+	GGGGACCCCCAAGACCCTGTGGG
518	Chr1:36098608-36098630	-	CAGGGCCCACAGGGTCTTGGGGG
519	Chr1:36098609-36098631	-	GCAGGGCCCACAGGGTCTTGGGG
520	Chr1:36098610-36098632	-	AGCAGGGCCCACAGGGTCTTGGG
521	Chr1:36098611-36098633	-	GAGCAGGGCCCACAGGGTCTTGG
522	Chr1:36098617-36098639	+	CCCTGTGGGCCCTGCTCCCCTGG
523	Chr1:36098617-36098639	-	CCAGGGGAGCAGGGCCCACAGGG
524	Chr1:36098618-36098640	-	GCCAGGGGAGCAGGGCCCACAGG
525	Chr1:36098626-36098648	-	GATGGGGAGCCAGGGGAGCAGGG
526	Chr1:36098627-36098649	-	GGATGGGGAGCCAGGGGAGCAGG
527	Chr1:36098633-36098655	-	AGGGGAGGATGGGGAGCCAGGGG
528	Chr1:36098634-36098656	-	CAGGGGAGGATGGGGAGCCAGGG
529	Chr1:36098635-36098657	+	CCTGGCTCCCCATCCTCCCCTGG
530	Chr1:36098635-36098657	-	CCAGGGGAGGATGGGGAGCCAGG
531	Chr1:36098642-36098664	-	GGGTGAGCCAGGGGAGGATGGGG
532	Chr1:36098643-36098665	-	GGGGTGTAGCCAGGGGAGGATGGG
533	Chr1:36098644-36098666	-	AGGGGTGAGCCAGGGGAGGATGG
534	Chr1:36098648-36098670	-	GGACAGGGGTGAGCCAGGGGAGG

[0503]

535	Chr1:36098651-36098673	-	GGGGGACAGGGGTGAGCCAGGGG
536	Chr1:36098652-36098674	-	TGGGGGACAGGGGTGAGCCAGGG
537	Chr1:36098653-36098675	-	TTGGGGGACAGGGGTGAGCCAGG
538	Chr1:36098662-36098684	+	CCCCTGTCCCCAAGAGTCCTGG
539	Chr1:36098662-36098684	-	CCAGGACTCTTGGGGGACAGGGG
540	Chr1:36098663-36098685	+	CCCTGTCCCCAAGAGTCCTGGG
541	Chr1:36098663-36098685	-	CCCAGGACTCTTGGGGGACAGGG
542	Chr1:36098664-36098686	-	TCCCAGGACTCTTGGGGGACAGG
543	Chr1:36098669-36098691	-	TGGGGTCCCAGGACTCTTGGGGG
544	Chr1:36098670-36098692	-	CTGGGGTCCCAGGACTCTTGGGG
545	Chr1:36098671-36098693	-	GCTGGGGTCCCAGGACTCTTGGG
546	Chr1:36098672-36098694	-	AGCTGGGGTCCCAGGACTCTTGG
547	Chr1:36098674-36098696	+	AAGAGTCCTGGGACCCCAGCTGG
548	Chr1:36098675-36098697	+	AGAGTCCTGGGACCCCAGCTGGG
549	Chr1:36098680-36098702	-	AGGGGCCAGCTGGGGTCCCAGG
550	Chr1:36098687-36098709	-	GGGGGACAGGGGCCAGCTGGGG
551	Chr1:36098688-36098710	-	AGGGGGACAGGGGCCAGCTGGG
552	Chr1:36098689-36098711	-	AAGGGGGACAGGGGCCAGCTGG
553	Chr1:36098691-36098713	+	AGCTGGGCCCTGTCCCCCTTGG
554	Chr1:36098692-36098714	+	GCTGGGCCCTGTCCCCCTTGGG
555	Chr1:36098693-36098715	+	CTGGGCCCTGTCCCCCTTGGGG
556	Chr1:36098698-36098720	+	CCCCTGTCCCCCTTGGGGCCTGG
557	Chr1:36098698-36098720	-	CCAGGCCCAAGGGGGACAGGGG

[0504]

558	Chr1:36098699-36098721	-	GCCAGGCCCCAAGGGGGACAGGG
559	Chr1:36098700-36098722	-	TGCCAGGCCCCAAGGGGGACAGG
560	Chr1:36098705-36098727	-	AGGACTGCCAGGCCCCAAGGGGG
561	Chr1:36098706-36098728	-	CAGGACTGCCAGGCCCCAAGGGG
562	Chr1:36098707-36098729	+	CCCTTGGGGCCTGGCAGTCCTGG
563	Chr1:36098707-36098729	-	CCAGGACTGCCAGGCCCCAAGGG
564	Chr1:36098708-36098730	-	GCCAGGACTGCCAGGCCCCAAGG
565	Chr1:36098716-36098738	-	TATGGGATGCCAGGACTGCCAGG
566	Chr1:36098724-36098746	+	TCCTGGCATCCCATAGCCAGTGG
567	Chr1:36098725-36098747	+	CCTGGCATCCCATAGCCAGTGGG
568	Chr1:36098725-36098747	-	CCCCTGGCTATGGGATGCCAGG
569	Chr1:36098726-36098748	+	CTGGCATCCCATAGCCAGTGGGG
570	Chr1:36098733-36098755	-	TGATAGGCCCCACTGGCTATGGG
571	Chr1:36098734-36098756	-	CTGATAGGCCCCACTGGCTATGG
572	Chr1:36098740-36098762	+	CCAGTGGGGCCTATCAGCCCAGG
573	Chr1:36098740-36098762	-	CCTGGGCTGATAGGCCCCACTGG
574	Chr1:36098741-36098763	+	CAGTGGGGCCTATCAGCCCAGGG
575	Chr1:36098742-36098764	+	AGTGGGGCCTATCAGCCCAGGGG
576	Chr1:36098743-36098765	+	GTGGGGCCTATCAGCCCAGGGGG
577	Chr1:36098744-36098766	+	TGGGGCCTATCAGCCCAGGGGGG
578	Chr1:36098749-36098771	-	CGGGGCCCCCTGGGCTGATAGG
579	Chr1:36098750-36098772	+	CTATCAGCCCAGGGGGGCCCCGG
580	Chr1:36098751-36098773	+	TATCAGCCCAGGGGGGCCCCGGG

[0505]

581	Chr1:36098757-36098779	-	CAGGGACCCGGGGCCCCCCTGGG
582	Chr1:36098758-36098780	+	CCAGGGGGGGCCCCGGGTCCCTGG
583	Chr1:36098758-36098780	-	CCAGGGACCCGGGGCCCCCCTGG
584	Chr1:36098767-36098789	-	AAAGGGGAGCCAGGGACCCGGGG
585	Chr1:36098768-36098790	-	CAAAGGGGAGCCAGGGACCCGGG
586	Chr1:36098769-36098791	+	CCGGGTCCCTGGCTCCCCTTTGG
587	Chr1:36098769-36098791	-	CCAAAGGGGAGCCAGGGACCCGG
588	Chr1:36098775-36098797	-	CAGGGGCCAAAGGGGAGCCAGGG
589	Chr1:36098776-36098798	-	TCAGGGGCCAAAGGGGAGCCAGG
590	Chr1:36098779-36098801	+	GGCTCCCCTTTGGCCCCTGATGG
591	Chr1:36098780-36098802	+	GCTCCCCTTTGGCCCCTGATGGG
592	Chr1:36098783-36098805	-	GGGCCCATCAGGGGCCAAAGGGG
593	Chr1:36098784-36098806	-	AGGGCCCATCAGGGGCCAAAGGG
594	Chr1:36098785-36098807	-	CAGGGCCCATCAGGGGCCAAAGG
595	Chr1:36098788-36098810	+	TTGGCCCCTGATGGGCCCTGTGG
596	Chr1:36098792-36098814	-	AGGACCACAGGGCCCATCAGGGG
597	Chr1:36098793-36098815	-	CAGGACCACAGGGCCCATCAGGG
598	Chr1:36098794-36098816	+	CCTGATGGGCCCTGTGGTCCTGG
599	Chr1:36098794-36098816	-	CCAGGACCACAGGGCCCATCAGG
600	Chr1:36098803-36098825	-	GCAGGGTTGCCAGGACCACAGGG
601	Chr1:36098804-36098826	-	AGCAGGGTTGCCAGGACCACAGG
602	Chr1:36098812-36098834	+	CCTGGCAACCCTGCTGCCCCTGG
603	Chr1:36098812-36098834	-	CCAGGGGCAGCAGGGTTGCCAGG

[0506]

604	Chr1:36098813-36098835	+	CTGGCAACCCTGCTGCCCCCTGGG
605	Chr1:36098820-36098842	-	TGGGAGTCCCAGGGGCAGCAGGG
606	Chr1:36098821-36098843	-	GTGGGAGTCCCAGGGGCAGCAGG
607	Chr1:36098828-36098850	-	AGACGGTGTGGGAGTCCCAGGGG
608	Chr1:36098829-36098851	-	TAGACGGTGTGGGAGTCCCAGGG
609	Chr1:36098830-36098852	-	GTAGACGGTGTGGGAGTCCCAGG
610	Chr1:36098836-36098858	+	ACTCCCACACCGTCTACTCCAGG
611	Chr1:36098839-36098861	+	CCCACACCGTCTACTCCAGGAGG
612	Chr1:36098839-36098861	-	CCTCCTGGAGTAGACGGTGTGGG
613	Chr1:36098840-36098862	-	ACCTCCTGGAGTAGACGGTGTGG
614	Chr1:36098845-36098867	-	AAAGGACCTCCTGGAGTAGACGG
615	Chr1:36098848-36098870	+	TCTACTCCAGGAGGTCTTTTGG
616	Chr1:36098849-36098871	+	CTACTCCAGGAGGTCTTTTGGG
617	Chr1:36098854-36098876	-	GTGGGCCCAAAGGACCTCCTGG
618	Chr1:36098863-36098885	+	CCTTTTGGGCCACAGCTCCTGG
619	Chr1:36098863-36098885	-	CCAGGAGCTGTGGGCCCAAAGG
620	Chr1:36098872-36098894	-	AGGGGGGAGCCAGGAGCTGTGGG
621	Chr1:36098873-36098895	-	CAGGGGGGAGCCAGGAGCTGTGG
622	Chr1:36098874-36098896	+	CACAGCTCCTGGCTCCCCCTGG
623	Chr1:36098875-36098897	+	ACAGCTCCTGGCTCCCCCTGGG
624	Chr1:36098876-36098898	+	CAGCTCCTGGCTCCCCCTGGGG
625	Chr1:36098881-36098903	+	CCTGGCTCCCCCTGGGGCCTGG
626	Chr1:36098881-36098903	-	CCAGGCCCCAGGGGGGAGCCAGG

[0507]

627	Chr1:36098888-36098910	-	TGGAGTTCCAGGCCCCAGGGGGG
628	Chr1:36098889-36098911	-	CTGGAGTTCCAGGCCCCAGGGGGG
629	Chr1:36098890-36098912	+	CCCCTGGGGCCTGGA ACTCCAGG
630	Chr1:36098890-36098912	-	CCTGGAGTTCCAGGCCCCAGGGG
631	Chr1:36098891-36098913	-	TCCTGGAGTTCCAGGCCCCAGGG
632	Chr1:36098892-36098914	-	CTCCTGGAGTTCCAGGCCCCAGG
633	Chr1:36098893-36098915	+	CTGGGGCCTGGA ACTCCAGGAGG
634	Chr1:36098899-36098921	-	TCTGGGCCTCCTGGAGTTCCAGG
635	Chr1:36098908-36098930	-	AAGGGTGAGTCTGGGCCTCCTGG
636	Chr1:36098916-36098938	-	CAGGAGACAAGGGTGAGTCTGGG
637	Chr1:36098917-36098939	+	CCAGACTCACCTTGTCTCCTGG
638	Chr1:36098917-36098939	-	CCAGGAGACAAGGGTGAGTCTGG
639	Chr1:36098918-36098940	+	CAGACTCACCTTGTCTCCTGGG
640	Chr1:36098919-36098941	+	AGACTCACCTTGTCTCCTGGGG
641	Chr1:36098926-36098948	+	CCCTTGTCTCCTGGGGCCCCAGG
642	Chr1:36098926-36098948	-	CCTGGGGCCCCAGGAGACAAGGG
643	Chr1:36098927-36098949	-	TCCTGGGGCCCCAGGAGACAAGG
644	Chr1:36098935-36098957	-	GATGGGCTTCCTGGGGCCCCAGG
645	Chr1:36098942-36098964	-	TGGTTTGGATGGGCTTCCTGGGG
646	Chr1:36098943-36098965	-	CTGGTTTGGATGGGCTTCCTGGG
647	Chr1:36098944-36098966	+	CCAGGAAGCCCATCCAAACCAGG
648	Chr1:36098944-36098966	-	CCTGGTTTGGATGGGCTTCCTGG
649	Chr1:36098952-36098974	-	TAGGCAAACCTGGTTTGGATGGG

[0508]

650	Chr1:36098953-36098975	-	TTAGGCAAACCTGGTTTGGATGG
651	Chr1:36098957-36098979	-	TGGCTTAGGCAAACCTGGTTTGG
652	Chr1:36098962-36098984	+	CCAGGTTTGCCTAAGCCAGCTGG
653	Chr1:36098962-36098984	-	CCAGCTGGCTTAGGCAAACCTGG
654	Chr1:36098968-36098990	+	TTGCCTAAGCCAGCTGGACCAGG
655	Chr1:36098969-36098991	+	TGCCTAAGCCAGCTGGACCAGGG
656	Chr1:36098971-36098993	-	CTCCCTGGTCCAGCTGGCTTAGG
657	Chr1:36098972-36098994	+	CTAAGCCAGCTGGACCAGGGAGG
658	Chr1:36098976-36098998	+	GCCAGCTGGACCAGGGAGGCCGG
659	Chr1:36098977-36098999	+	CCAGCTGGACCAGGGAGGCCGGG
660	Chr1:36098977-36098999	-	CCCGGCCTCCCTGGTCCAGCTGG
661	Chr1:36098978-36099000	+	CAGCTGGACCAGGGAGGCCGGGG
662	Chr1:36098979-36099001	+	AGCTGGACCAGGGAGGCCGGGGG
663	Chr1:36098980-36099002	+	GCTGGACCAGGGAGGCCGGGGGG
664	Chr1:36098981-36099003	+	CTGGACCAGGGAGGCCGGGGGGG
665	Chr1:36098985-36099007	+	ACCAGGGAGGCCGGGGGGGGCCGG
666	Chr1:36098986-36099008	+	CCAGGGAGGCCGGGGGGGGCCGGG
667	Chr1:36098986-36099008	-	CCCGGCCCCCCCCGGCCTCCCTGG
668	Chr1:36098987-36099009	+	CAGGGAGGCCGGGGGGGGCCGGGG
669	Chr1:36098988-36099010	+	AGGGAGGCCGGGGGGGGCCGGGGG
670	Chr1:36098995-36099017	-	GGGGGTGCCCCCGGCCCCCGG
671	Chr1:36099004-36099026	+	CCGGGGGCACCCCCCTGCCCTGG
672	Chr1:36099004-36099026	-	CCAGGGCAGGGGGGTGCCCCCGG

[0509]

673	Chr1:36099005-36099027	+	CGGGGGCACCCCCCTGCCCTGGG
674	Chr1:36099006-36099028	+	GGGGGCACCCCCCTGCCCTGGGG
675	Chr1:36099013-36099035	+	CCCCCCTGCCCTGGGGCCCCAGG
676	Chr1:36099013-36099035	-	CCTGGGGCCCCAGGGCAGGGGGG
677	Chr1:36099014-36099036	-	GCCTGGGGCCCCAGGGCAGGGGG
678	Chr1:36099015-36099037	-	TGCCTGGGGCCCCAGGGCAGGGG
679	Chr1:36099016-36099038	-	CTGCCTGGGGCCCCAGGGCAGGG
680	Chr1:36099017-36099039	-	GCTGCCTGGGGCCCCAGGGCAGG
681	Chr1:36099021-36099043	+	CCCTGGGGCCCCAGGCAGCCCGG
682	Chr1:36099021-36099043	-	CCGGGCTGCCTGGGGCCCCAGGG
683	Chr1:36099022-36099044	+	CCTGGGGCCCCAGGCAGCCCGGG
684	Chr1:36099022-36099044	-	CCCGGGCTGCCTGGGGCCCCAGG
685	Chr1:36099026-36099048	+	GGGCCCCAGGCAGCCCGGGCTGG
686	Chr1:36099029-36099051	-	GGGCCAGCCCGGGCTGCCTGGGG
687	Chr1:36099030-36099052	-	TGGGCCAGCCCGGGCTGCCTGGG
688	Chr1:36099031-36099053	-	GTGGGCCAGCCCGGGCTGCCTGG
689	Chr1:36099039-36099061	-	ATAATGGAGTGGGCCAGCCCGGG
690	Chr1:36099040-36099062	-	GATAATGGAGTGGGCCAGCCCGG
691	Chr1:36099049-36099071	-	CTCAAGGGGGATAATGGAGTGGG
692	Chr1:36099050-36099072	+	CCACTCCATTATCCCCCTTGAGG
693	Chr1:36099050-36099072	-	CCTCAAGGGGGATAATGGAGTGG
694	Chr1:36099055-36099077	-	CGAGGCCTCAAGGGGGATAATGG
695	Chr1:36099062-36099084	-	AGGTGATCGAGGCCTCAAGGGGG

[0510]

696	Chr1:36099063-36099085	-	CAGGTGATCGAGGCCTCAAGGGG
697	Chr1:36099064-36099086	+	CCCTTGAGGCCTCGATCACCTGG
698	Chr1:36099064-36099086	-	CCAGGTGATCGAGGCCTCAAGGG
699	Chr1:36099065-36099087	+	CCTTGAGGCCTCGATCACCTGGG
700	Chr1:36099065-36099087	-	CCCAGGTGATCGAGGCCTCAAGG
701	Chr1:36099066-36099088	+	CTTGAGGCCTCGATCACCTGGGG
702	Chr1:36099067-36099089	+	TTGAGGCCTCGATCACCTGGGGG
703	Chr1:36099073-36099095	+	CCTCGATCACCTGGGGGCCCAGG
704	Chr1:36099073-36099095	-	CCTGGGGCCCCAGGTGATCGAGG
705	Chr1:36099082-36099104	-	CAGGGGGAGCCTGGGGCCCCCAGG
706	Chr1:36099083-36099105	+	CTGGGGGCCCAGGCTCCCCCTGG
707	Chr1:36099084-36099106	+	TGGGGGCCCAGGCTCCCCCTGGG
708	Chr1:36099085-36099107	+	GGGGGCCCAGGCTCCCCCTGGGG
709	Chr1:36099090-36099112	-	CAGGGCCCCAGGGGGAGCCTGGG
710	Chr1:36099091-36099113	+	CCAGGCTCCCCCTGGGGCCCTGG
711	Chr1:36099091-36099113	-	CCAGGGCCCCAGGGGGAGCCTGG
712	Chr1:36099098-36099120	-	GGGGGAACCAGGGCCCCAGGGGG
713	Chr1:36099099-36099121	-	AGGGGGAACCAGGGCCCCAGGGG
714	Chr1:36099100-36099122	-	CAGGGGGAACCAGGGCCCCAGGG
715	Chr1:36099101-36099123	+	CCTGGGGCCCTGGTTCCCCCTGG
716	Chr1:36099101-36099123	-	CCAGGGGGAACCAGGGCCCCAGG
717	Chr1:36099108-36099130	-	CAGGATTCCAGGGGGAACCAGGG
718	Chr1:36099109-36099131	+	CCTGGTTCCCCCTGGAATCCTGG

[0511]

719	Chr1:36099109-36099131	-	CCAGGATTCCAGGGGAACCAGG
720	Chr1:36099110-36099132	+	CTGGTCCCCCTGGAATCCTGGG
721	Chr1:36099111-36099133	+	TGGTCCCCCTGGAATCCTGGGG
722	Chr1:36099112-36099134	+	GGTCCCCCTGGAATCCTGGGGG
723	Chr1:36099116-36099138	-	AGGGCCCCCAGGATTCCAGGGG
724	Chr1:36099117-36099139	-	CAGGGCCCCCAGGATTCCAGGGG
725	Chr1:36099118-36099140	+	CCCTGGAATCCTGGGGGCCCTGG
726	Chr1:36099118-36099140	-	CCAGGGCCCCCAGGATTCCAGGG
727	Chr1:36099119-36099141	-	GCCAGGGCCCCCAGGATTCCAGG
728	Chr1:36099127-36099149	-	CAAGGGGTGCCAGGGCCCCCAGG
729	Chr1:36099128-36099150	+	CTGGGGGCCCTGGCACCCCTTGG
730	Chr1:36099129-36099151	+	TGGGGGCCCTGGCACCCCTTGGG
731	Chr1:36099135-36099157	-	CAGGTGCCCAAGGGGTGCCAGGG
732	Chr1:36099136-36099158	+	CCTGGCACCCCTTGGGCACCTGG
733	Chr1:36099136-36099158	-	CCAGGTGCCCAAGGGGTGCCAGG
734	Chr1:36099143-36099165	-	TGGAAAACCAGGTGCCCAAGGGG
735	Chr1:36099144-36099166	-	CTGGAAAACCAGGTGCCCAAGGG
736	Chr1:36099145-36099167	+	CCTTGGGCACCTGGTTTTCCAGG
737	Chr1:36099145-36099167	-	CCTGGAAAACCAGGTGCCCAAGG
738	Chr1:36099146-36099168	+	CTTGGGCACCTGGTTTTCCAGGG
739	Chr1:36099154-36099176	-	ATTACTATCCCTGGAAAACCAGG
740	Chr1:36099162-36099184	+	TCCAGGGATAGTAATGCCTGAGG
741	Chr1:36099163-36099185	+	CCAGGGATAGTAATGCCTGAGGG

[0512]

742	Chr1:36099163-36099185	-	CCCTCAGGCATTACTATCCCTGG
743	Chr1:36099164-36099186	+	CAGGGATAGTAATGCCTGAGGGG
744	Chr1:36099169-36099191	+	ATAGTAATGCCTGAGGGGCCCGG
745	Chr1:36099170-36099192	+	TAGTAATGCCTGAGGGGCCCGGG
746	Chr1:36099173-36099195	+	TAATGCCTGAGGGGCCCGGGAGG
747	Chr1:36099178-36099200	+	CCTGAGGGGCCCGGGAGGCCAGG
748	Chr1:36099178-36099200	-	CCTGGCCTCCCGGGCCCCTCAGG
749	Chr1:36099179-36099201	+	CTGAGGGGCCCGGGAGGCCAGGG
750	Chr1:36099180-36099202	+	TGAGGGGCCCGGGAGGCCAGGGG
751	Chr1:36099181-36099203	+	GAGGGGCCCGGGAGGCCAGGGGG
752	Chr1:36099187-36099209	+	CCCGGGAGGCCAGGGGGTCTCTGG
753	Chr1:36099187-36099209	-	CCAGGACCCCCTGGCCTCCCGGG
754	Chr1:36099188-36099210	+	CCGGGAGGCCAGGGGGTCTCTGGG
755	Chr1:36099188-36099210	-	CCCAGGACCCCCTGGCCTCCCGG
756	Chr1:36099189-36099211	+	CGGGAGGCCAGGGGGTCTCTGGGG
757	Chr1:36099190-36099212	+	GGGAGGCCAGGGGGTCTCTGGGGG
758	Chr1:36099196-36099218	-	CGGGGACCCCAGGACCCCCTGG
759	Chr1:36099197-36099219	+	CAGGGGGTCTGGGGGTCCCCGG
760	Chr1:36099200-36099222	+	GGGGTCTCTGGGGGTCCCCGGAGG
761	Chr1:36099205-36099227	-	CAGGGCCTCCGGGGACCCCAGG
762	Chr1:36099206-36099228	+	CTGGGGGTCCCCGGAGGCCCTGG
763	Chr1:36099214-36099236	-	CGAGGGGACCAGGGCCTCCGGGG
764	Chr1:36099215-36099237	-	ACGAGGGGACCAGGGCCTCCGGG

[0513]

765	Chr1:36099216-36099238	-	TACGAGGGGACCAGGGCCTCCGG
766	Chr1:36099223-36099245	+	CCCTGGTCCCCTCGTATTCCCTGG
767	Chr1:36099223-36099245	-	CCAGGAATACGAGGGGACCAGGG
768	Chr1:36099224-36099246	-	GCCAGGAATACGAGGGGACCAGG
769	Chr1:36099230-36099252	-	GGGGGAGCCAGGAATACGAGGGG
770	Chr1:36099231-36099253	-	GGGGGGAGCCAGGAATACGAGGG
771	Chr1:36099232-36099254	-	CGGGGGGAGCCAGGAATACGAGG
772	Chr1:36099241-36099263	+	CCTGGCTCCCCCGAAGCCCCGG
773	Chr1:36099241-36099263	-	CCGGGGCTTCGGGGGAGCCAGG
774	Chr1:36099248-36099270	-	AGGGCAGCCGGGGCTTCGGGGGG
775	Chr1:36099249-36099271	-	CAGGGCAGCCGGGGCTTCGGGGG
776	Chr1:36099250-36099272	+	CCCCGAAGCCCCGGCTGCCCTGG
777	Chr1:36099250-36099272	-	CCAGGGCAGCCGGGGCTTCGGGG
778	Chr1:36099251-36099273	-	ACCAGGGCAGCCGGGGCTTCGGG
779	Chr1:36099252-36099274	-	CACCAGGGCAGCCGGGGCTTCGG
780	Chr1:36099253-36099275	+	CGAAGCCCCGGCTGCCCTGGTGG
781	Chr1:36099258-36099280	-	TCGGGCCACCAGGGCAGCCGGGG
782	Chr1:36099259-36099281	-	GTCGGGCCACCAGGGCAGCCGGG
783	Chr1:36099260-36099282	-	GGTCGGGCCACCAGGGCAGCCGG
784	Chr1:36099267-36099289	-	CTGGCAAGGTCGGGCCACCAGGG
785	Chr1:36099268-36099290	+	CCTGGTGGCCCGACCTTGCCAGG
786	Chr1:36099268-36099290	-	CCTGGCAAGGTCGGGCCACCAGG
787	Chr1:36099269-36099291	+	CTGGTGGCCCGACCTTGCCAGGG

[0514]

788	Chr1:36099276-36099298	-	CAGGGCTCCCTGGCAAGGTCGGG
789	Chr1:36099277-36099299	+	CCGACCTTGCCAGGGAGCCCTGG
790	Chr1:36099277-36099299	-	CCAGGGCTCCCTGGCAAGGTCGG
791	Chr1:36099278-36099300	+	CGACCTTGCCAGGGAGCCCTGGG
792	Chr1:36099279-36099301	+	GACCTTGCCAGGGAGCCCTGGGG
793	Chr1:36099280-36099302	+	ACCTTGCCAGGGAGCCCTGGGGG
794	Chr1:36099281-36099303	-	TCCCCCAGGGCTCCCTGGCAAGG
795	Chr1:36099286-36099308	-	GCTGGTCCCCCAGGGCTCCCTGG
796	Chr1:36099294-36099316	-	TGGGCAAGGCTGGTCCCCCAGGG
797	Chr1:36099295-36099317	-	ATGGGCAAGGCTGGTCCCCCAGG
798	Chr1:36099299-36099321	+	GGGGACCAGCCTTGCCCATCCGG
799	Chr1:36099300-36099322	+	GGGACCAGCCTTGCCCATCCGGG
800	Chr1:36099304-36099326	-	TTCTCCCGGATGGGCAAGGCTGG
801	Chr1:36099308-36099330	-	TGGCTTCTCCCGGATGGGCAAGG
802	Chr1:36099310-36099332	+	TTGCCCATCCGGGAGAAGCCAGG
803	Chr1:36099311-36099333	+	TGCCCATCCGGGAGAAGCCAGGG
804	Chr1:36099312-36099334	+	GCCCATCCGGGAGAAGCCAGGGG
805	Chr1:36099313-36099335	+	CCCATCCGGGAGAAGCCAGGGGG
806	Chr1:36099313-36099335	-	CCCCCTGGCTTCTCCCGGATGGG
807	Chr1:36099314-36099336	-	GCCCCCTGGCTTCTCCCGGATGG
808	Chr1:36099318-36099340	-	CTGGGCCCCCTGGCTTCTCCCGG
809	Chr1:36099322-36099344	+	GAGAAGCCAGGGGGCCAGCAGG
810	Chr1:36099323-36099345	+	AGAAGCCAGGGGGCCAGCAGGG

[0515]

811	Chr1:36099328-36099350	+	CCAGGGGGCCAGCAGGGCCAGG
812	Chr1:36099328-36099350	-	CCTGGCCCTGCTGGGCCCCCTGG
813	Chr1:36099336-36099358	-	ATGGGCAGCCTGGCCCTGCTGGG
814	Chr1:36099337-36099359	-	CATGGGCAGCCTGGCCCTGCTGG
815	Chr1:36099338-36099360	+	CAGCAGGGCCAGGCTGCCCATGG
816	Chr1:36099346-36099368	+	CCAGGCTGCCCATGGAGTCCTGG
817	Chr1:36099346-36099368	-	CCAGGACTCCATGGGCAGCCTGG
818	Chr1:36099354-36099376	-	TGGGAAAGCCAGGACTCCATGGG
819	Chr1:36099355-36099377	-	ATGGGAAAGCCAGGACTCCATGG
820	Chr1:36099361-36099383	+	AGTCCTGGCTTCCCATGCCTGG
821	Chr1:36099364-36099386	-	AAACCAGGCATGGGAAAGCCAGG
822	Chr1:36099370-36099392	+	TTCCCATGCCTGGTTTTCTGG
823	Chr1:36099371-36099393	+	TTCCCATGCCTGGTTTTCTGGG
824	Chr1:36099373-36099395	-	TTCCAGGAAAACCAGGCATGGG
825	Chr1:36099374-36099396	-	CTTCCAGGAAAACCAGGCATGG
826	Chr1:36099379-36099401	+	CCTGGTTTTCTGGGAAGCCAGG
827	Chr1:36099379-36099401	-	CCTGGCTTCCCAGGAAAACCAGG
828	Chr1:36099380-36099402	+	CTGGTTTTCTGGGAAGCCAGGG
829	Chr1:36099381-36099403	+	TGGTTTTCTGGGAAGCCAGGGG
830	Chr1:36099382-36099404	+	GGTTTTCTGGGAAGCCAGGGG
831	Chr1:36099383-36099405	+	GTTTTCTGGGAAGCCAGGGGG
832	Chr1:36099388-36099410	+	CCTGGGAAGCCAGGGGGGCCAGG
833	Chr1:36099388-36099410	-	CCTGGCCCCCTGGCTTCCCAGG

[0516]

834	Chr1:36099389-36099411	+	CTGGGAAGCCAGGGGGGCCAGGG
835	Chr1:36099390-36099412	+	TGGGAAGCCAGGGGGGCCAGGGG
836	Chr1:36099391-36099413	+	GGGAAGCCAGGGGGGCCAGGGGG
837	Chr1:36099397-36099419	-	CGGGGTCCCCCTGGCCCCCTGG
838	Chr1:36099400-36099422	+	GGGGGGCCAGGGGGACCCCGAGG
839	Chr1:36099405-36099427	+	GCCAGGGGGACCCCGAGGCCCGG
840	Chr1:36099406-36099428	+	CCAGGGGGACCCCGAGGCCCGGG
841	Chr1:36099406-36099428	-	CCCGGGCCTCGGGGTCCCCCTGG
842	Chr1:36099415-36099437	+	CCCCGAGGCCCGGGCTTCCCAGG
843	Chr1:36099415-36099437	-	CCTGGGAAGCCCGGGCCTCGGGG
844	Chr1:36099416-36099438	+	CCCGAGGCCCGGGCTTCCCAGGG
845	Chr1:36099416-36099438	-	CCCTGGGAAGCCCGGGCCTCGGG
846	Chr1:36099417-36099439	+	CCGAGGCCCGGGCTTCCCAGGGG
847	Chr1:36099417-36099439	-	CCCCTGGGAAGCCCGGGCCTCGG
848	Chr1:36099418-36099440	+	CGAGGCCCGGGCTTCCCAGGGGG
849	Chr1:36099419-36099441	+	GAGGCCCGGGCTTCCCAGGGGGG
850	Chr1:36099423-36099445	+	CCCGGGCTTCCCAGGGGGGCCGG
851	Chr1:36099423-36099445	-	CCGGCCCCCTGGGAAGCCCGGG
852	Chr1:36099424-36099446	+	CCGGGCTTCCCAGGGGGGCCGGG
853	Chr1:36099424-36099446	-	CCCGGCCCCCTGGGAAGCCCGG
854	Chr1:36099432-36099454	-	AGGGAGAGCCCGGCCCCCTGGG
855	Chr1:36099433-36099455	-	AAGGGAGAGCCCGGCCCCCTGG
856	Chr1:36099437-36099459	+	GGGGGCCGGGCTCTCCCTTCAGG

[0517]

857	Chr1:36099442-36099464	-	ATGGACCTGAAGGGAGAGCCCGG
858	Chr1:36099445-36099467	+	GGCTCTCCCTTCAGGTCCATCGG
859	Chr1:36099451-36099473	-	CTGCTGCCGATGGACCTGAAGGG
860	Chr1:36099452-36099474	-	GCTGCTGCCGATGGACCTGAAGG
861	Chr1:36099454-36099476	+	TTCAGGTCCATCGGCAGCAGCGG
862	Chr1:36099460-36099482	+	TCCATCGGCAGCAGCGGTAGAGG
863	Chr1:36099461-36099483	-	GCCTCTACCGCTGCTGCCGATGG
864	Chr1:36099485-36099507	+	TTTCTGAGAAAGAAAGAGAAAGG
865	Chr1:36099486-36099508	+	TTCTGAGAAAGAAAGAGAAAGGG
866	Chr1:36099487-36099509	+	TCTGAGAAAGAAAGAGAAAGGGG
867	Chr1:36099495-36099517	+	AGAAAGAGAAAGGGGCAGTCAGG
868	Chr1:36099496-36099518	+	GAAAGAGAAAGGGGCAGTCAGGG
869	Chr1:36099497-36099519	+	AAAGAGAAAGGGGCAGTCAGGGG
870	Chr1:36099509-36099531	+	GCAGTCAGGGGCCTGAACTGTGG
871	Chr1:36099510-36099532	+	CAGTCAGGGGCCTGAACTGTGGG
872	Chr1:36099511-36099533	+	AGTCAGGGGCCTGAACTGTGGGG
873	Chr1:36099516-36099538	+	GGGGCCTGAACTGTGGGGACAGG
874	Chr1:36099517-36099539	+	GGGCCTGAACTGTGGGGACAGGG
875	Chr1:36099518-36099540	+	GGCCTGAACTGTGGGGACAGGGG
876	Chr1:36099520-36099542	-	GTCCCCTGTCCCCACAGTTCAGG
877	Chr1:36099542-36099564	-	AATGGGGGAATGGGTAGATGGGG
878	Chr1:36099543-36099565	-	GAATGGGGGAATGGGTAGATGGG
879	Chr1:36099544-36099566	-	GGAATGGGGGAATGGGTAGATGG

[0518]

880	Chr1:36099551-36099573	-	TCATACTGGAATGGGGGAATGGG
881	Chr1:36099552-36099574	-	CTCATACTGGAATGGGGGAATGG
882	Chr1:36099553-36099575	+	CATTCCCCATTCCAGTATGAGG
883	Chr1:36099557-36099579	-	TGTACCTCATACTGGAATGGGGG
884	Chr1:36099558-36099580	-	GTGTACCTCATACTGGAATGGGG
885	Chr1:36099559-36099581	-	CGTGTACCTCATACTGGAATGGG
886	Chr1:36099560-36099582	+	CCATTCCAGTATGAGGTACACGG
887	Chr1:36099560-36099582	-	CCGTGTACCTCATACTGGAATGG
888	Chr1:36099561-36099583	+	CATTCCAGTATGAGGTACACGGG
889	Chr1:36099565-36099587	-	CTCTCCCGTGTACCTCATACTGG
890	Chr1:36099566-36099588	+	CAGTATGAGGTACACGGGAGAGG
891	Chr1:36099574-36099596	+	GGTACACGGGAGAGGAAGAATGG
892	Chr1:36099575-36099597	+	GTACACGGGAGAGGAAGAATGGG
893	Chr1:36099576-36099598	+	TACACGGGAGAGGAAGAATGGGG
894	Chr1:36099598-36099620	+	GCTGCCCTTCTGCTCTCATGG
895	Chr1:36099602-36099624	-	TCTTCCATGAGAGCAGGAAGGGG
896	Chr1:36099603-36099625	-	ATCTTCCATGAGAGCAGGAAGGG
897	Chr1:36099604-36099626	-	CATCTTCCATGAGAGCAGGAAGG
898	Chr1:36099605-36099627	+	CTTCTGCTCTCATGGAAGATGG
899	Chr1:36099606-36099628	+	TTCTGCTCTCATGGAAGATGGG
900	Chr1:36099607-36099629	+	TCCTGCTCTCATGGAAGATGGGG
901	Chr1:36099608-36099630	-	ACCCCATCTTCCATGAGAGCAGG
902	Chr1:36099612-36099634	+	CTCTCATGGAAGATGGGGTTTGG

[0519]

903	Chr1:36099613-36099635	+	TCTCATGGAAGATGGGGTTTGGG
904	Chr1:36099614-36099636	+	CTCATGGAAGATGGGGTTTGGGG
905	Chr1:36099615-36099637	+	TCATGGAAGATGGGGTTTGGGGG
906	Chr1:36099618-36099640	+	TGGAAGATGGGGTTTGGGGGTGG
907	Chr1:36099624-36099646	+	ATGGGGTTTGGGGGTGGCCAGG
908	Chr1:36099625-36099647	+	TGGGGTTTGGGGGTGGCCAGGG
909	Chr1:36099626-36099648	+	GGGGTTTGGGGGTGGCCAGGGG
910	Chr1:36099635-36099657	+	GGGTGGCCAGGGGACATCTTGG
911	Chr1:36099636-36099658	+	GGTGGCCAGGGGACATCTTGGG
912	Chr1:36099637-36099659	+	GTGGCCAGGGGACATCTTGGGG
913	Chr1:36099638-36099660	+	TGGCCAGGGGACATCTTGGGGG
914	Chr1:36099641-36099663	-	TTGCCCCAAGATGTCCCCTGGG
915	Chr1:36099642-36099664	-	GTTGCCCCAAGATGTCCCCTGG
916	Chr1:36099645-36099667	+	GGGGACATCTTGGGGGCAACAGG
917	Chr1:36099646-36099668	+	GGGACATCTTGGGGGCAACAGGG
918	Chr1:36099660-36099682	+	GCAACAGGGTGTCTCCTTAAGG
919	Chr1:36099661-36099683	+	CAACAGGGTGTCTCCTTAAGGG
920	Chr1:36099672-36099694	-	GGTGTTAGGAGCCCTTAAGGAGG
921	Chr1:36099675-36099697	-	TTGGGTGTTAGGAGCCCTTAAGG
922	Chr1:36099685-36099707	+	TCCTAACACCCAACCTACCTAGG
923	Chr1:36099686-36099708	-	GCCTAGGTAGGTTGGGTGTTAGG
924	Chr1:36099689-36099711	+	AACACCCAACCTACCTAGGCTGG
925	Chr1:36099690-36099712	+	ACACCCAACCTACCTAGGCTGGG

[0520]

926	Chr1:36099693-36099715	-	AGGCCCAGCCTAGGTAGGTTGGG
927	Chr1:36099694-36099716	-	GAGGCCCAGCCTAGGTAGGTTGG
928	Chr1:36099698-36099720	-	GGAGGAGGCCCAGCCTAGGTAGG
929	Chr1:36099702-36099724	-	TCATGGAGGAGGCCCAGCCTAGG
930	Chr1:36099708-36099730	+	CTGGGCCTCCTCCATGAGCCTGG
931	Chr1:36099713-36099735	-	ATCAGCCAGGCTCATGGAGGAGG
932	Chr1:36099716-36099738	-	AGAATCAGCCAGGCTCATGGAGG
933	Chr1:36099719-36099741	-	GTGAGAATCAGCCAGGCTCATGG
934	Chr1:36099726-36099748	-	ATGAGAGGTGAGAATCAGCCAGG
935	Chr1:36099741-36099763	-	TCAGGTCATGCAGGGATGAGAGG
936	Chr1:36099744-36099766	+	CTCATCCCTGCATGACCTGAAGG
937	Chr1:36099747-36099769	+	ATCCCTGCATGACCTGAAGGTGG
938	Chr1:36099749-36099771	-	CTCCACCTTCAGGTCATGCAGGG
939	Chr1:36099750-36099772	-	ACTCCACCTTCAGGTCATGCAGG
940	Chr1:36099752-36099774	+	TGCATGACCTGAAGGTGGAGTGG
941	Chr1:36099759-36099781	-	CTGGTGGCCACTCCACCTTCAGG
942	Chr1:36099760-36099782	+	CTGAAGGTGGAGTGGCCACCAGG
943	Chr1:36099763-36099785	+	AAGGTGGAGTGGCCACCAGGTGG
944	Chr1:36099775-36099797	-	GGGCTGCTGGTGCCACCTGGTGG
945	Chr1:36099778-36099800	-	GGTGGGCTGCTGGTGCCACCTGG
946	Chr1:36099788-36099810	-	CGGGCTCTAAGGTGGGCTGCTGG
947	Chr1:36099791-36099813	+	GCAGCCCACCTTAGAGCCCGTGG
948	Chr1:36099792-36099814	+	CAGCCCACCTTAGAGCCCGTGGG

[0521]

949	Chr1:36099795-36099817	-	GCTCCCACGGGCTCTAAGGTGGG
950	Chr1:36099796-36099818	-	TGCTCCCACGGGCTCTAAGGTGG
951	Chr1:36099799-36099821	-	CTCTGCTCCCACGGGCTCTAAGG
952	Chr1:36099807-36099829	-	AGGTGGGGCTCTGCTCCCACGGG
953	Chr1:36099808-36099830	-	GAGGTGGGGCTCTGCTCCCACGG
954	Chr1:36099822-36099844	-	AACTGGGAAGTTGGGAGGTGGGG
955	Chr1:36099823-36099845	-	GAACTGGGAAGTTGGGAGGTGGG
956	Chr1:36099824-36099846	-	TGAACTGGGAAGTTGGGAGGTGG
957	Chr1:36099827-36099849	-	AGATGAACTGGGAAGTTGGGAGG
958	Chr1:36099830-36099852	-	GGGAGATGAACTGGGAAGTTGGG
959	Chr1:36099831-36099853	-	GGGGAGATGAACTGGGAAGTTGG
960	Chr1:36099836-36099858	+	TTCCAGTTCATCTCCCCCTTGG
961	Chr1:36099838-36099860	-	TTCCAAGGGGGAGATGAACTGGG
962	Chr1:36099839-36099861	-	CTTCCAAGGGGGAGATGAACTGG
963	Chr1:36099850-36099872	-	GCACAGGTGGTCTTCCAAGGGGG
964	Chr1:36099851-36099873	-	GGCACAGGTGGTCTTCCAAGGGG
965	Chr1:36099852-36099874	-	TGGCACAGGTGGTCTTCCAAGGG
966	Chr1:36099853-36099875	-	CTGGCACAGGTGGTCTTCCAAGG
967	Chr1:36099863-36099885	-	GTGCAGTTAGCTGGCACAGGTGG
968	Chr1:36099866-36099888	-	ACGGTGCAGTTAGCTGGCACAGG
969	Chr1:36099872-36099894	-	CTGGAAACGGTGCAGTTAGCTGG
970	Chr1:36099873-36099895	+	CAGCTAACTGCACCGTTTCCAGG
971	Chr1:36099881-36099903	+	TGCACCGTTTCCAGGCCCTCTGG

[0522]

972	Chr1:36099882-36099904	+	GCACCGTTTCCAGGCCCTCTGGG
973	Chr1:36099883-36099905	+	CACCGTTTCCAGGCCCTCTGGGG
974	Chr1:36099885-36099907	-	TACCCCAGAGGGCCTGGAAACGG
975	Chr1:36099890-36099912	+	TCCAGGCCCTCTGGGGTATTAGG
976	Chr1:36099891-36099913	-	TCCTAATACCCCAGAGGGCCTGG
977	Chr1:36099896-36099918	-	GTTTTTCCTAATACCCCAGAGGG
978	Chr1:36099897-36099919	-	TGTTTTTCCTAATACCCCAGAGG
979	Chr1:36099904-36099926	+	GGTATTAGGAAAAACACTGAAGG
980	Chr1:36099908-36099930	+	TTAGGAAAAACACTGAAGGTAGG
981	Chr1:36099916-36099938	+	AACACTGAAGGTAGGAAAATTGG
982	Chr1:36099919-36099941	+	ACTGAAGGTAGGAAAATTGGTGG
983	Chr1:36099920-36099942	+	CTGAAGGTAGGAAAATTGGTGGG
984	Chr1:36099921-36099943	+	TGAAGGTAGGAAAATTGGTGGGG
985	Chr1:36099928-36099950	+	AGGAAAATTGGTGGGGAATGAGG
986	Chr1:36099936-36099958	+	TGGTGGGGAATGAGGAGCTGTGG
987	Chr1:36099939-36099961	+	TGGGGAATGAGGAGCTGTGGAGG
988	Chr1:36099940-36099962	+	GGGGAATGAGGAGCTGTGGAGGG
989	Chr1:36099949-36099971	+	GGAGCTGTGGAGGGCGCCTGAGG
990	Chr1:36099958-36099980	+	GAGGGCGCCTGAGGATCTGATGG
991	Chr1:36099965-36099987	-	CTGAGAGCCATCAGATCCTCAGG
992	Chr1:36099966-36099988	+	CTGAGGATCTGATGGCTCTCAGG
993	Chr1:36099967-36099989	+	TGAGGATCTGATGGCTCTCAGGG
994	Chr1:36099970-36099992	+	GGATCTGATGGCTCTCAGGGAGG

[0523]

995	Chr1:36099974-36099996	+	CTGATGGCTCTCAGGGAGGCAGG
996	Chr1:36099975-36099997	+	TGATGGCTCTCAGGGAGGCAGGG
997	Chr1:36099976-36099998	+	GATGGCTCTCAGGGAGGCAGGGG
998	Chr1:36099982-36100004	+	TCTCAGGGAGGCAGGGGATTTGG
999	Chr1:36099983-36100005	+	CTCAGGGAGGCAGGGGATTTGGG
1000	Chr1:36099984-36100006	+	TCAGGGAGGCAGGGGATTTGGGG
1001	Chr1:36099985-36100007	+	CAGGGAGGCAGGGGATTTGGGGG
1002	Chr1:36099989-36100011	+	GAGGCAGGGGATTTGGGGGCTGG
1003	Chr1:36099990-36100012	+	AGGCAGGGGATTTGGGGGCTGGG
1004	Chr1:36100002-36100024	+	TGGGGGCTGGGAGCGATTTGAGG
1005	Chr1:36100010-36100032	+	GGGAGCGATTTGAGGCACTGTGG
1006	Chr1:36100011-36100033	+	GGAGCGATTTGAGGCACTGTGGG
1007	Chr1:36100012-36100034	+	GAGCGATTTGAGGCACTGTGGGG
1008	Chr1:36100017-36100039	+	ATTTGAGGCACTGTGGGGTGAGG
1009	Chr1:36100020-36100042	+	TGAGGCACTGTGGGGTGAGGAGG
1010	Chr1:36100032-36100054	+	GGGTGAGGAGGCTCTCACCCAGG
1011	Chr1:36100038-36100060	+	GGAGGCTCTCACCCAGGTACTION
1012	Chr1:36100049-36100071	-	GAGGGCAAAGGCCAGTACCTGGG
1013	Chr1:36100050-36100072	-	TGAGGGCAAAGGCCAGTACCTGG
1014	Chr1:36100053-36100075	+	GGTACTGGCCTTTGCCCTCACGG
1015	Chr1:36100057-36100079	+	CTGGCCTTTGCCCTCACGGAAGG
1016	Chr1:36100058-36100080	+	TGGCCTTTGCCCTCACGGAAGGG
1017	Chr1:36100061-36100083	+	CCTTTGCCCTCACGGAAGGGCGG

[0524]

1018	Chr1:36100061-36100083	-	CCGCCCTTCCGTGAGGGCAAAGG
1019	Chr1:36100067-36100089	-	GTGGGACCGCCCTTCCGTGAGGG
1020	Chr1:36100068-36100090	-	TGTGGGACCGCCCTTCCGTGAGG
1021	Chr1:36100070-36100092	+	TCACGGAAGGGCGGTCCCACAGG
1022	Chr1:36100084-36100106	+	TCCCACAGGTCCTTTCTGCATGG
1023	Chr1:36100085-36100107	+	CCCACAGGTCCTTTCTGCATGGG
1024	Chr1:36100085-36100107	-	CCCATGCAGAAAGGACCTGTGGG
1025	Chr1:36100086-36100108	-	GCCCATGCAGAAAGGACCTGTGG
1026	Chr1:36100089-36100111	+	CAGGTCCTTTCTGCATGGGCTGG
1027	Chr1:36100094-36100116	-	TACATCCAGCCCATGCAGAAAGG
1028	Chr1:36100103-36100125	+	ATGGGCTGGATGTACTTCACTGG
1029	Chr1:36100104-36100126	+	TGGGCTGGATGTACTTCACTGGG
1030	Chr1:36100105-36100127	+	GGGCTGGATGTACTTCACTGGGG
1031	Chr1:36100126-36100148	+	GGCATAGCCCGCCGCCCCACCGG
1032	Chr1:36100133-36100155	-	GGCGGGGCCGGTGGGGCGGCGGG
1033	Chr1:36100134-36100156	-	TGGCGGGGCCGGTGGGGCGGCGG
1034	Chr1:36100137-36100159	-	TGGTGGCGGGGCCGGTGGGGCGG
1035	Chr1:36100140-36100162	-	CTCTGGTGGCGGGGCCGGTGGGG
1036	Chr1:36100141-36100163	+	CCCACCGGCCCCGCCACCAGAGG
1037	Chr1:36100141-36100163	-	CCTCTGGTGGCGGGGCCGGTGGG
1038	Chr1:36100142-36100164	-	TCCTCTGGTGGCGGGGCCGGTGG
1039	Chr1:36100145-36100167	-	GCGTCCTCTGGTGGCGGGGCCGG
1040	Chr1:36100149-36100171	-	GCGGGCGTCCTCTGGTGGCGGGG

[0525]

1041	Chr1:36100150-36100172	-	CGCGGGCGTCCTCTGGTGGCGGG
1042	Chr1:36100151-36100173	+	CCGCCACCAGAGGACGCCCGCGG
1043	Chr1:36100151-36100173	-	CCGCGGGCGTCCTCTGGTGGCGG
1044	Chr1:36100154-36100176	-	GGGCCGCGGGCGTCCTCTGGTGG
1045	Chr1:36100157-36100179	-	TGTGGGCCGCGGGCGTCCTCTGG
1046	Chr1:36100167-36100189	-	GGTGCTGGGGTGTGGGCCGCGGG
1047	Chr1:36100168-36100190	-	TGGTGCTGGGGTGTGGGCCGCGG
1048	Chr1:36100174-36100196	-	TGGTGCTGGTGCTGGGGTGTGGG
1049	Chr1:36100175-36100197	-	CTGGTGCTGGTGCTGGGGTGTGG
1050	Chr1:36100180-36100202	-	TGCTACTGGTGCTGGTGCTGGGG
1051	Chr1:36100181-36100203	-	CTGCTACTGGTGCTGGTGCTGGG
1052	Chr1:36100182-36100204	-	GCTGCTACTGGTGCTGGTGCTGG
1053	Chr1:36100188-36100210	-	GCTGCTGCTGCTACTGGTGCTGG
1054	Chr1:36100194-36100216	-	TTCGCTGCTGCTGCTGCTACTGG
1055	Chr1:36100200-36100222	+	GCAGCAGCAGCAGCGAAGACAGG
1056	Chr1:36100201-36100223	+	CAGCAGCAGCAGCGAAGACAGGG
1057	Chr1:36100202-36100224	+	AGCAGCAGCAGCGAAGACAGGGG
1058	Chr1:36100222-36100244	+	GGGTGTCAGAGTCCCAGCATGG
1059	Chr1:36100231-36100253	+	AGTCCCAGCATGGCGTCCGTGG
1060	Chr1:36100234-36100256	-	CGTCCACGGACGCCATGCTGGGG
1061	Chr1:36100235-36100257	-	ACGTCCACGGACGCCATGCTGGG
1062	Chr1:36100236-36100258	-	CACGTCCACGGACGCCATGCTGG
1063	Chr1:36100248-36100270	-	TCTTCTTTGCAGCACGTCCACGG

[0526] 预期包含与SEQ ID NO:191-1063互补的向导序列或结合SEQ ID NO:191-1063的反向互补序列的gRNA的使用可将核酸酶(例如Cas9或Cas9RNP)靶向至COL8A2的序列。已在

早发型FECD中表征了COL8A2的杂合突变体,靶向具有包含与SEQ ID NO:191-1063的靶序列互补的向导序列的gRNA的Cas RNP可导致通过NHEJ产生插入/缺失。插入/缺失的产生可以减少COL8A2的表达,从而减少产生的毒性胶原-8蛋白质 $\alpha$ -2亚基。毒性COL8A2产物的减少可以改善早发型FECD的病程,因为其他形式的胶原可替换 $\alpha$ -2亚基。某些向导也可以用于切除COL8A2基因的包含已知的疾病相关突变的区域,或改变剪接模式以有利于不包含这类突变的同种型。还可以将用某些向导敲除COL8A2基因与野生型COL8A2替换策略结合使用。例如,去除内源性显性阴性突变体形式的表达后,可通过转基因手段表达野生型COL8A2编码序列。

[0527] 根据突变等位基因核苷酸序列中的差异,还可以鉴定突变等位基因特异性靶序列。

[0528] 表4列出对导致Gln455Lys (由c.1364C>A核苷酸改变引起)的突变特异的靶序列。包含与SEQ ID NO:1064-1069互补的向导序列的gRNA的使用将靶向突变等位基因,而不靶向或低效靶向野生型等位基因。由于具有Gln455Lys突变的个体通常仅有一个等位基因受影响,预期由靶向COL8A2的突变等位基因的Cas RNP介导的NHEJ引起的插入/缺失的选择性产生仅导致此等位基因的丧失,而保留另一野生型COL8A2等位基因。备选地,包含与SEQ ID NO:1064-1069互补的向导序列或结合SEQ ID NO:1064-1069的反向互补序列的向导序列的gRNA也可以与模板一起使用,以介导突变的矫正。

[0529]

**表 4: 具有 Gln455Lys 突变的 COL8A2 的靶序列**

SEQ ID No	靶位置	靶链	靶序列
1064	Chr1:36098302-36098324	+	CCCCTCAGGCCAGGCTTCCCAGG
1065	Chr1:36098302-36098324	-	CCTGGGAAGCCTGGCCTGAGGGG
1066	Chr1:36098303-36098325	+	CCCTCAGGCCAGGTTGCCAGGG
1067	Chr1:36098303-36098325	-	CCCTGGGAAGCCTGGCCTGAGGG
1068	Chr1:36098304-36098326	-	TCCCTGGGAAGCCTGGCCTGAGG
1069	Chr1:36098311-36098333	-	TTGGGGCTCCCTGGGAAGCCTGG

[0530] 表5列出对导致Gln455Val (由c.1363-1364CA>GT核苷酸改变引起)的点突变特异的靶序列。包含指导核酸酶至SEQ ID NO:1070-1075的向导序列的gRNA的使用将靶向突变等位基因,而不靶向或低效靶向野生型等位基因。由于具有Gln455Val突变的个体通常仅有一个等位基因受影响,预期由靶向COL8A2的突变等位基因的核酸酶(例如Cas RNP)介导的NHEJ引起的插入/缺失的选择性产生仅导致此等位基因的丧失,而保留另一野生型COL8A2等位基因。备选地,包含与SEQ ID NO:1070-1075互补的向导序列的gRNA也可以与模板一起使用,以介导突变的矫正。

[0531]

SEQ ID No	靶位置	靶链	靶序列
1070	Chr1:36098302-36098324	+	CCCCTCAGGCCAGGCACCCCAGG
1071	Chr1:36098302-36098324	-	CCTGGGGTGCCTGGCCTGAGGGG
1072	Chr1:36098303-36098325	+	CCCTCAGGCCAGGCACCCCAGGG
1073	Chr1:36098303-36098325	-	CCCTGGGGTGCCTGGCCTGAGGG
1074	Chr1:36098304-36098326	-	TCCCTGGGGTGCCTGGCCTGAGG
1075	Chr1:36098311-36098333	-	TTGGGGCTCCCTGGGGTGCCTGG

[0532] 表6列出对导致Leu450Trp(由c.1349T>G核苷酸改变引起)的点突变特异的靶序列。包含与SEQ ID NO:1076-1084互补的向导序列的gRNA的使用将靶向突变等位基因,而不靶向或低效靶向野生型等位基因。由于具有Leu450Trp突变的个体通常仅有一个等位基因受影响,预期由靶向COL8A2的突变等位基因的Cas RNP介导的NHEJ引起的插入/缺失的选择性产生仅导致此等位基因的丧失,而保留另一野生型COL8A2等位基因。备选地,包含与SEQ ID NO:1076-1084互补的向导序列的gRNA也可以与模板一起使用,以介导突变的矫正。

[0533]

SEQ ID No	靶位置	靶链	靶序列
1076	Chr1:36098311-36098333	-	TGGGGGCTCCCTGGGCAGCCTGG
1077	Chr1:36098319-36098341	-	AAGGTGACTGGGGGCTCCCTGGG
1078	Chr1:36098320-36098342	-	AAAGGTGACTGGGGGCTCCCTGG
1079	Chr1:36098328-36098350	-	TGGGGCAGAAAGGTGACTGGGGG
1080	Chr1:36098329-36098351	-	CTGGGGCAGAAAGGTGACTGGGG
1081	Chr1:36098330-36098352	+	CCCAGTCACCTTTCTGCCCCAGG
1082	Chr1:36098330-36098352	-	CCTGGGGCAGAAAGGTGACTGGG
1083	Chr1:36098331-36098353	+	CCAGTCACCTTTCTGCCCCAGGG
1084	Chr1:36098331-36098353	-	CCCTGGGGCAGAAAGGTGACTGG

[0534] 模板可以与Cas RNP一起使用,以校正导致产生具有Gln455Lys、Gln455Val或Leu450Trp突变的胶原VIII的核苷酸突变。通过这种方式,Cas RNP可以针对突变,启动NHEJ,然后基于外源模板介导突变的校正。靶向Cas RNP以校正导致Gln455Lys产物表达的突变可以使用包含与SEQ ID NO:1064-1069的靶序列互补的向导序列的gRNA以及模板来完成。靶向Cas RNP以校正导致Gln455Val产物表达的突变可以使用包含与SEQ ID NO:1070-1075的靶序列互补的向导序列的gRNA和模板进行。靶向Cas RNP以校正导致Leu450Trp基因产物表达的突变可以使用包含与SEQ ID NO:1076-1084的靶序列互补的向导序列的gRNA和模板进行。以这种方式,可以进行突变等位基因的选择性编辑,以校正由Gln455Lys、Gln455Val或Leu450Trp引起的缺陷型胶原VIII。

[0535] 因此,包含含有与COL8A2的靶序列互补的向导序列的gRNA的Cas RNP的使用可以是治疗FECD或PPCD的新手段。靶序列包括野生型COL8A2的靶序列,以及对可产生COL8A2的突变等位基因并产生含有Gln455Lys、Gln455Val或Leu450Trp突变的基因产物的突变特异的靶序列。表4、5和6中所列的突变特异性靶序列可以用来开发向导RNA,用于具有特异性的Cas(例如在Cas RNP中),用于在突变等位基因中引入其他突变来消除其功能,或备选地,与

模板一起使用以矫正COL8A2中的致病核苷酸突变。

[0536] 等同物

[0537] 认为前文的说明书足以使本领域技术人员能够实施实施方案。前文的描述和实施例详述某些实施方案,并描述发明人考虑的最佳方式。但是,应理解,无论前文在文字上表现得多么详尽,实施方案可以按许多方式实施,应按照所附权利要求书及其任何等同物解释。

[0538] 无论是否明确说明,本文所用的术语“约”指数值,包括例如整数、分数和百分数。术语“约”通常指本领域普通技术人员将认为等同于所列举的值(例如具有相同的功能或结果)的数值范围(例如所列举的范围的 $\pm 5-10\%$ )。在诸如“至少”和“约”的术语放在一系列数值或范围前时,该术语修饰列表中提供的所有值或范围。在一些情况下,术语“约”可包括修约至最近有效数字的数值。

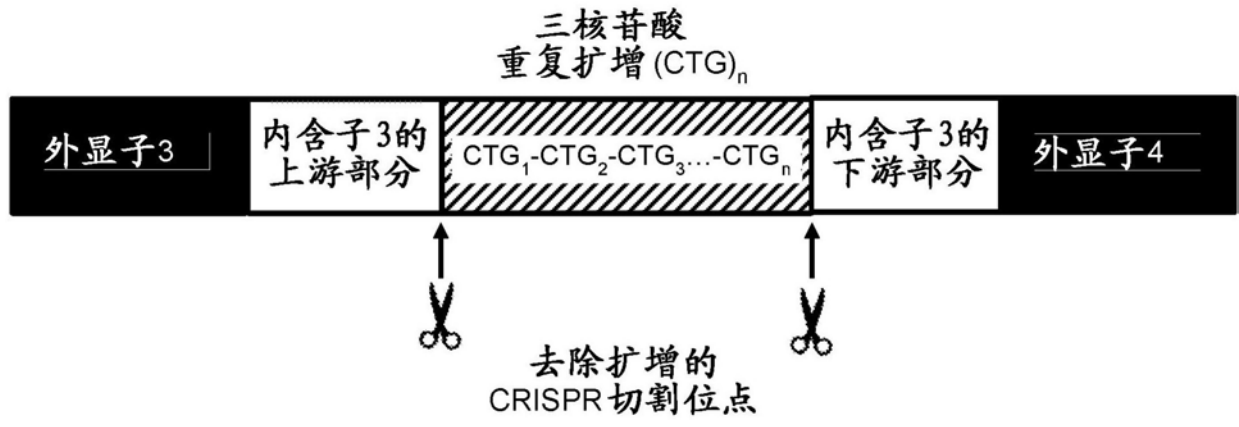


图1

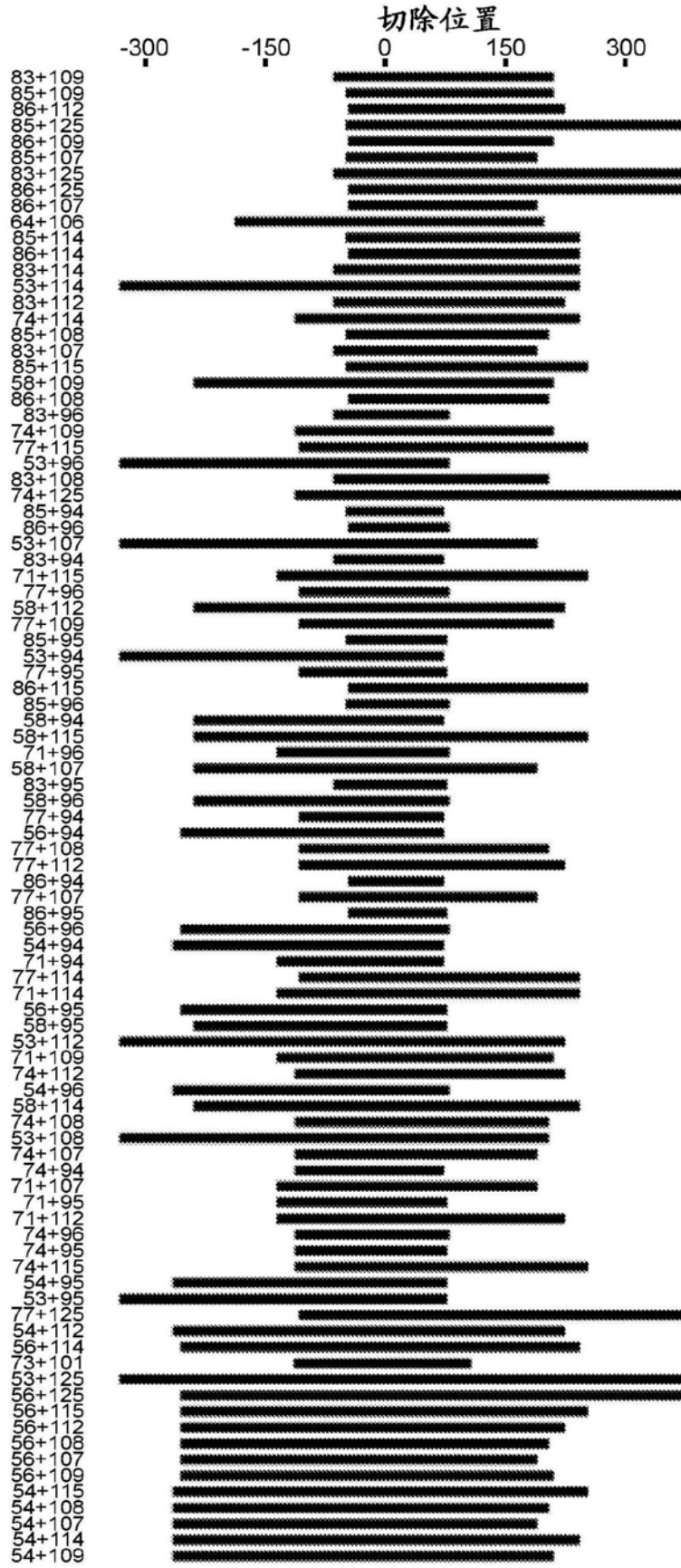


图2