



# (12) 发明专利申请

(10) 申请公布号 CN 119731328 A

(43) 申请公布日 2025. 03. 28

(21) 申请号 202380049353.8

(22) 申请日 2023.05.12

(30) 优先权数据

63/341,665 2022.05.13 US

63/478,472 2023.01.04 US

(85) PCT国际申请进入国家阶段日

2024.12.23

(86) PCT国际申请的申请数据

PCT/US2023/066969 2023.05.12

(87) PCT国际申请的公布数据

W02023/220742 EN 2023.11.16

(71) 申请人 海鞘治疗公司

地址 美国

(72) 发明人 R·克伦巴赫 S·杜利 A·多伊

K·伯克哈特 J·格雷 彭凌涛

D·吴 A·诺码 K·戈西克

S·斯洛莫维奇 A·克莱门斯

R·贝尔

(74) 专利代理机构 北京市联德律师事务所

11361

专利代理师 黄大正 张来光

(51) Int.Cl.

C12N 15/86 (2006.01)

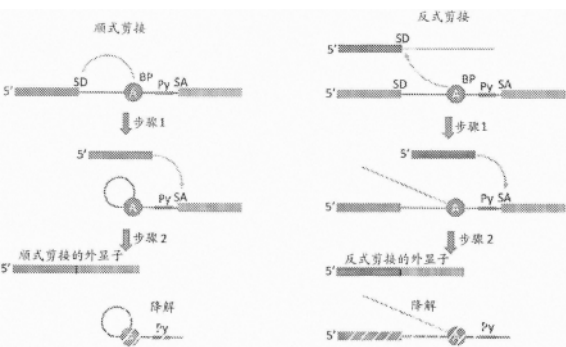
A61K 48/00 (2006.01)

(54) 发明名称

ABCA4反式剪接分子

(57) 摘要

本文提供了能够纠正ABCA4基因突变的核酸反式剪接分子(例如,mRNA前体反式剪接分子(RTM);RNA外显子编辑分子)。这类分子可用于治疗如ABCA4相关的视网膜营养不良(例如,斯特格氏病或视锥-视杆细胞营养不良)等疾病。本文还描述了使用本文所述的核酸反式剪接分子纠正ABCA4中的突变,从而治疗与ABCA4中的突变相关的病症的方法,以及本文所述的核酸反式剪接分子治疗与ABCA4中的突变相关的病症以及制备供治疗与ABCA4中的突变相关的病症的药物的用途。



1. 一种核酸反式剪接分子,所述核酸反式剪接分子包含:
  - (a) 编码结构域序列(CDS),所述编码结构域序列包含ABCA4外显子;
  - (b) 接头结构域,所述接头结构域包含SEQ ID NO:27或与SEQ ID NO:27具有至少85%同一性的序列;以及
  - (c) 结合结构域,所述结合结构域退火到内源ABCA4 mRNA前体内的结合位点,其中所述CDS、所述接头结构域和所述结合结构域在5'至3'方向上操作性连接。
2. 一种核酸反式剪接分子,所述核酸反式剪接分子包含:
  - (a) CDS,所述CDS包含SEQ ID NO:53-55中的任一者的变体、基本上由其组成或由其组成,其中所述SEQ ID NO:53-55中的任一者的变体在表3中列出的至少一个隐蔽剪接位点中包含至少一个核苷酸变异,其中所述至少一个核苷酸变异减少包含至少一个核苷酸变异的所述至少一个隐蔽剪接位点的每一者处隐蔽剪接位点的使用;以及
  - (b) 接头结构域;以及
  - (c) 结合结构域,所述结合结构域退火到内源ABCA4 mRNA前体内的结合位点,其中所述CDS、所述接头结构域和所述结合结构域在5'至3'方向上操作性连接。
3. 如权利要求2所述的核酸反式剪接分子,其中所述SEQ ID NO:55的变体包含SEQ ID NO:56-59中的任一者。
4. 如权利要求3所述的核酸反式剪接分子,其中所述SEQ ID NO:55的变体包含SEQ ID NO:56。
5. 如前述权利要求中任一项所述的核酸反式剪接分子,其中所述接头结构域包含与SEQ ID NO:27具有至少85%同一性、至少86%同一性、至少87%同一性、至少88%同一性或至少89%同一性、90%同一性、91%同一性、至少92%同一性、至少93%同一性、至少94%同一性、至少95%同一性、至少96%同一性、至少97%同一性、至少98%同一性或至少99%同一性的序列。
6. 如前述权利要求中任一项所述的核酸反式剪接分子,其中所述接头结构域包含SEQ ID NO:27、基本上由其组成或由其组成。
7. 如前述权利要求中任一项所述的核酸反式剪接分子,其中所述结合结构域退火到所述内源ABCA4 mRNA前体中的内含子。
8. 如权利要求7所述的核酸反式剪接分子,其中所述内源ABCA4 mRNA前体中的所述内含子是内含子22。
9. 如前述权利要求中任一项所述的核酸反式剪接分子,其中所述结合结构域包含SEQ ID NO:18或SEQ ID NO:20。
10. 如前述权利要求中任一项所述的核酸反式剪接分子,其中所述结合结构域基本上由SEQ ID NO:18或SEQ ID NO:20组成或由其组成。
11. 如前述权利要求中任一项所述的核酸反式剪接分子,其中所述结合结构域基本上由SEQ ID NO:20组成或由其组成。
12. 如权利要求1或5-11中任一项所述的核酸反式剪接分子,其中所述CDS包含ABCA4外显子1-22或其变体。
13. 如权利要求12所述的核酸反式剪接分子,其中所述CDS包含SEQ ID NO:53-59中的任一者、基本上由其组成或由其组成。

14. 一种核酸反式剪接分子, 所述核酸反式剪接分子包含:

(a) 接头结构域, 所述接头结构域包含SEQ ID NO:27或与SEQ ID NO:27具有至少91%同一性、至少92%同一性、至少93%同一性、至少94%同一性、至少95%同一性、至少96%同一性、至少97%同一性、至少98%同一性或至少99%同一性的序列、基本上由其组成或由其组成;

(b) 结合结构域, 所述结合结构域包含SEQ NO:18或ID SEQ ID NO:20、基本上由其组成或由其组成; 以及

(c) CDS, 所述CDS包含ABCA4外显子;

其中所述CDS、所述接头结构域和所述结合结构域在5'至3'方向上操作性连接。

15. 如权利要求14所述的核酸反式剪接分子, 其中所述CDS包含SEQ ID NO:53-59中的任一者、基本上由其组成或由其组成。

16. 一种核酸反式剪接分子, 所述核酸反式剪接分子包含:

(a) CDS, 所述CDS包含SEQ ID NO:53-59中的任一者、基本上由其组成或由其组成;

(b) 接头结构域, 所述接头结构域包含SEQ ID NO:27或与SEQ ID NO:27具有至少91%同一性、至少92%同一性、至少93%同一性、至少94%同一性、至少95%同一性、至少96%同一性、至少97%同一性、至少98%同一性或至少99%同一性的序列、基本上由其组成或由其组成; 以及

(c) 结合结构域, 所述结合结构域包含SEQ ID NO:20、基本上由其组成或由其组成, 其中所述CDS、所述接头结构域和所述结合结构域在5'至3'方向上操作性连接。

17. 如前述权利要求中任一项所述的核酸反式剪接分子, 所述核酸反式剪接分子还包含增效子, 所述增效子包括包含GGTAAGT或SEQ ID NO:61的至少一个序列。

18. 如权利要求17所述的核酸反式剪接分子, 其中所述增效子在3'可操作地连接到所述接头结构域。

19. 如权利要求17或18中任一项所述的核酸反式剪接分子, 其中所述增效子包含三个串联排列并可操作地连接的序列, 其中所述三个序列中的每一个包含GGTAAGT或SEQ ID NO:61。

20. 如权利要求17-19中任一项所述的核酸反式剪接分子, 其中所述增效子包含三个串联排列的序列, 其中所述三个序列包含SEQ ID NO:62。

21. 如前述权利要求中任一项所述的核酸反式剪接分子, 所述核酸反式剪接分子还包含翻译增效子, 其中所述翻译增效子包含SEQ ID NO:63或与SEQ ID NO:63具有至少90%同一性的序列, 其中SEQ ID NO:63或所述与SEQ ID NO:63具有至少90%同一性的序列在3'可操作地连接到所述增效子。

22. 如权利要求21所述的核酸反式剪接分子, 其中所述翻译增效子与SEQ ID NO:63具有至少91%同一性, 与SEQ ID NO:63具有至少92%同一性, 与SEQ ID NO:63具有至少93%同一性, 与SEQ ID NO:63具有至少94%同一性, 与SEQ ID NO:63具有至少95%同一性, 与SEQ ID NO:63具有至少96%同一性, 与SEQ ID NO:63具有至少97%同一性, 与SEQ ID NO:63具有至少98%同一性, 或与SEQ ID NO:63具有至少99%同一性。

23. 如权利要求21或22中任一项所述的核酸反式剪接分子, 其中所述翻译增效子包含SEQ ID NO:63、基本上由其组成或由其组成。

24. 如前述权利要求中任一项所述的核酸反式剪接分子,所述核酸反式剪接分子还包括包含GTAAGT、GTAAGG、GTAAGC或GTAAC中任一者的剪接结构域。

25. 如前述权利要求中任一项所述的核酸反式剪接分子,所述核酸反式剪接分子还包括包含GTAAGT、基本上由其组成或由其组成的剪接结构域。

26. 如前述权利要求中任一项所述的核酸反式剪接分子,所述核酸反式剪接分子还包含5'非翻译区。

27. 如权利要求26所述的核酸反式剪接分子,其中所述5'非翻译区包含SEQ ID NO:13或SEQ ID NO:64、基本上由其组成或由其组成。

28. 如前述权利要求中任一项所述的核酸反式剪接分子,所述核酸反式剪接分子还包含3'转录终止子结构域。

29. 如权利要求28所述的核酸反式剪接分子,其中所述3'转录终止子结构域编码形成三螺旋结构的核糖核酸,所述三螺旋结构保护所述反式剪接分子的3'端。

30. 如权利要求28或权利要求29所述的核酸反式剪接分子,其中所述3'转录终止子结构域包含野生型MALAT1序列(SEQ ID NO:33)或与SEQ ID NO:33具有至少80%同一性的核酸序列。

31. 如权利要求30所述的核酸反式剪接分子,其中所述3'转录终止子结构域包含如下核酸序列,所述核酸序列与SEQ ID NO:33具有至少80%同一性;与SEQ ID NO:33具有至少85%同一性;与SEQ ID NO:33具有至少90%同一性;与SEQ ID NO:33具有至少91%同一性;与SEQ ID NO:33具有至少92%同一性;与SEQ ID NO:33具有至少93%同一性;与SEQ ID NO:33具有至少94%同一性;与SEQ ID NO:33具有至少95%同一性;与SEQ ID NO:33具有至少96%同一性;与SEQ ID NO:33具有至少97%同一性;与SEQ ID NO:33具有至少98%同一性;与SEQ ID NO:33具有至少99%同一性;或与SEQ ID NO:33100%同一。

32. 如权利要求28或权利要求29所述的核酸反式剪接分子,其中所述3'转录终止子结构域包含SEQ ID NO:33或SEQ ID NO:66、基本上由其组成或由其组成。

33. 如前述权利要求中任一项所述的核酸反式剪接分子,所述核酸反式剪接分子还包含至少一个在3'操作性连接到所述剪接结构域的终止密码子。

34. 如权利要求33所述的核酸反式剪接分子,所述核酸反式剪接分子在5'至3'方向上包含操作性连接的以下各者:

(a) 5'非翻译区;

(b) CDS,所述CDS基本上由SEQ ID NO:53-59中的任一者组成或由其组成;

(c) 剪接结构域;

(d) 至少一个终止密码子;

(e) 接头结构域,所述接头结构域基本上由SEQ ID NO:27组成或由其组成;

(f) 至少一个增效子;

(g) 翻译增效子;

(h) 结合结构域,所述结合结构域基本上由SEQ ID NO:18或SEQ ID NO:20组成或由其组成;以及

(i) 3'转录终止子结构域。

35. 如权利要求34所述的核酸反式剪接分子,其中

- (a) 所述5'非翻译区包含SEQ ID NO:13或SEQ ID NO:64、基本上由其组成或由其组成；
- (b) 所述剪接结构域包含GTAAGT、基本上由其组成或由其组成；
- (c) 所述至少一个终止密码子包含两个终止密码子、基本上由其组成或由其组成，其中所述两个终止密码子中的第一个终止密码子的序列与所述剪接结构域的序列部分重叠；
- (d) 所述至少一个增效子包含三个串联排列的序列、基本上由其组成或由其组成，其中所述三个序列中的每一个包含GGTAAGT或SEQ ID NO:61；
- (e) 所述翻译增效子包含SEQ ID NO:63、基本上由其组成或由其组成；并且
- (f) 所述3'转录终止子结构域包含SEQ ID NO:66、基本上由其组成或由其组成。

36. 一种核酸反式剪接分子，所述核酸反式剪接分子包含：

- (a) 5'非翻译区；
- (b) CDS，所述CDS包含SEQ ID NO:53-59中的任一者、基本上由其组成或由其组成；
- (c) 剪接结构域；
- (d) 任选地，至少一个终止密码子；
- (e) 接头结构域，所述接头结构域包含SEQ ID NO:27、基本上由其组成或由其组成；
- (f) 至少一个增效子；
- (g) 翻译增效子；
- (h) 结合结构域，所述结合结构域包含SEQ ID NO:18或SEQ ID NO:20、基本上由其组成或由其组成；以及
- (i) 3'转录终止子结构域，

其中所述5'非翻译区；所述CDS；所述剪接结构域；所述至少一个终止密码子在存在时；所述接头结构域；所述至少一个增效子；所述翻译增效子；所述结合结构域；以及所述3'转录终止子结构域在5'至3'方向上操作性连接。

37. 如权利要求36所述的核酸反式剪接分子，其中

- (a) 所述5'非翻译区包含SEQ ID NO:13或SEQ ID NO:64、基本上由其组成或由其组成；
- (b) 所述剪接结构域包含GTAAGT、基本上由其组成或由其组成；
- (c) 所述至少一个终止密码子存在且包含两个终止密码子、基本上由其组成或由其组成，其中所述两个终止密码子中的第一个终止密码子的序列与所述剪接结构域的序列部分重叠；
- (d) 所述至少一个增效子包含三个串联排列的序列、基本上由其组成或由其组成，其中所述三个序列中的每一个包含GGTAAGT或SEQ ID NO:61；
- (e) 所述翻译增效子包含SEQ ID NO:63、基本上由其组成或由其组成；并且
- (f) 所述3'转录终止子结构域包含SEQ ID NO:66、基本上由其组成或由其组成。

38. 一种核酸反式剪接分子，所述核酸反式剪接分子在5'至3'方向上包含操作性连接的以下各者：

- (a) 5'非翻译区，所述5'非翻译区包含SEQ ID NO:13或SEQ ID NO:64、基本上由其组成或由其组成；
- (b) CDS，所述CDS包含SEQ ID NO:53-59中的任一者、基本上由其组成或由其组成；
- (c) 剪接结构域，所述剪接结构域包含GTAAGT、基本上由其组成或由其组成；
- (d) 两个终止密码子，其中所述两个终止密码子中的第一个终止密码子的序列与所述

剪接结构域的序列部分重叠并且其中所述剪接结构域的序列与所述两个终止密码子组合包含SEQ ID NO:68、基本上由其组成或由其组成；

(e) 接头结构域,所述接头结构域包含SEQ ID NO:27、基本上由其组成或由其组成；

(f) 增效子,所述增效子包含SEQ ID NO:62、基本上由其组成或由其组成；

(g) 翻译增效子,所述翻译增效子包含SEQ ID NO:63、基本上由其组成或由其组成；

(h) 结合结构域,所述结合结构域包含SEQ ID NO:18或SEQ ID NO:20、基本上由其组成或由其组成；以及

(i) 3'转录终止子结构域,所述3'转录终止子结构域包含SEQ ID NO:66、基本上由其组成或由其组成。

39. 如权利要求37或权利要求38所述的核酸反式剪接分子,其中所述核酸反式剪接分子包含SEQ ID NO:69、SEQ ID NO:78或SEQ ID NO:90、基本上由其组成或由其组成。

40. 如权利要求1-20中任一项所述的核酸反式剪接分子,其中所述核酸反式剪接分子包含SEQ ID NO:70或SEQ ID NO:71、基本上由其组成或由其组成。

41. 如权利要求1-12中任一项所述的核酸反式剪接分子,其中所述核酸反式剪接分子包含与SEQ ID NO:40-44、46或48-50中的任一者具有至少80%同一性的核酸序列。

42. 一种核糖核酸反式剪接分子,所述核糖核酸反式剪接分子从如权利要求1-41中任一项所述的核酸反式剪接分子转录。

43. 如权利要求1-41中任一项所述的核酸反式剪接分子,其中所述内源ABCA4 mRNA前体包含至少一个与ABCA4相关的视网膜营养不良相关的突变。

44. 如权利要求43所述的核酸反式剪接分子,其中所述ABCA4相关的视网膜营养不良是隐性遗传的。

45. 如权利要求43或44所述的核酸反式剪接分子,其中所述ABCA4相关的视网膜营养不良与跨越所述ABCA4基因的外显子1-22的等位基因致病突变和与视锥-视杆细胞营养不良(例如,视锥-视杆细胞营养不良-3)、斯特格氏黄斑营养不良(例如,斯特格氏病-1)、眼底黄色斑点症、视网膜色素变性-19或年龄相关性黄斑变性-2一致的表型相关。

46. 如权利要求1-45中任一项所述的核酸反式剪接分子,其中包含所述至少一个突变的所述内源ABCA4 mRNA前体在感光细胞或视网膜色素上皮细胞中表达。

47. 一种载体,所述载体包含如权利要求1-46中任一项所述的核酸反式剪接分子。

48. 如权利要求47所述的载体,其中所述载体包含在5'操作性连接到所述CDS的5'调节结构域。

49. 如权利要求48所述的载体,其中所述5'调节结构域操作性连接到5'非翻译区。

50. 如权利要求48-49中任一项所述的载体,其中所述5'调节结构域包含组成型启动子或组织特异性启动子。

51. 如权利要求50所述的载体,其中所述组成型启动子是CMV启动子。

52. 一种前病毒质粒,所述前病毒质粒包含如权利要求1-46中任一项所述的核酸反式剪接分子。

53. 一种腺相关病毒(AAV),所述腺相关病毒包含如权利要求1-46中任一项所述的核酸反式剪接分子,其中所述AAV任选地包含5'调节结构域,所述5'调节结构域在5'操作性连接到所述核酸反式剪接分子。

54. 如权利要求53所述的AAV,其中所述AAV包含在5'操作性连接到所述CDS的5'调节结构域。

55. 如权利要求53或54中任一项所述的AAV,其中所述5'调节结构域操作性连接到5'非翻译区。

56. 如权利要求53-55中任一项所述的AAV,其中所述5'调节结构域包含组成型启动子。

57. 如权利要求56所述的AAV,其中所述组成型启动子是CMV启动子。

58. 如权利要求53-57中任一项所述的AAV,其中所述AAV优先靶向感光细胞和/或视网膜色素上皮细胞。

59. 如权利要求53-58中任一项所述的AAV,其中所述AAV是AAV8、AAV5或AAV2。

60. 一种组合物,所述组合物包含如权利要求1-46中任一项所述的核酸反式剪接分子、如权利要求47-51中任一项所述的载体、如权利要求52所述的前病毒质粒或如权利要求53-59中任一项所述的AAV。

61. 如权利要求60所述的组合物,所述组合物包含药学上可接受的赋形剂。

62. 一种在靶细胞中表达生物活性ABCA4以恢复所述靶细胞中ABCA4蛋白的功能水平的方法,所述方法包括将所述靶细胞用如权利要求1-46中任一项所述的核酸反式剪接分子、如权利要求47-51中任一项所述的载体、如权利要求52所述的前病毒质粒、如权利要求53-59中任一项所述的AAV或如权利要求60或61所述的组合物转导。

63. 如权利要求62所述的方法,其中所述靶细胞中的至少5%、至少10%或至少15%的靶ABCA4 mRNA被替换。

64. 如权利要求63所述的方法,其中所述靶细胞中的至少20%的靶ABCA4 mRNA被替换。

65. 如权利要求64所述的方法,其中所述靶细胞中的至少35%的靶ABCA4 mRNA被替换。

66. 如权利要求62-65中任一项所述的方法,其中通过表达具有生物功能的全长ABCA4蛋白,使所述靶细胞中ABCA4的功能水平恢复。

67. 一种减少受试者中ABCA4相关的脂褐素和/或A2E的积累的方法,所述方法包括将所述受试者中的靶视网膜细胞用如权利要求1-46中任一项所述的核酸反式剪接分子、如权利要求47-51中任一项所述的载体、如权利要求52所述的前病毒质粒、如权利要求53-59中任一项所述的AAV或如权利要求60或61所述的组合物转染或转导。

68. 一种纠正受试者的靶细胞中ABCA4外显子序列中的至少一个突变的方法,所述方法包括向所述受试者施用如权利要求1-46中任一项所述的核酸反式剪接分子、如权利要求47-51中任一项所述的载体、如权利要求52所述的前病毒质粒、如权利要求53-59中任一项所述的AAV或如权利要求60或61所述的组合物。

69. 一种治疗有需要的受试者的ABCA4相关的视网膜营养不良的方法,所述方法包括以治疗有效量向所述受试者施用如权利要求1-46中任一项所述的核酸反式剪接分子、如权利要求47-51中任一项所述的载体、如权利要求52所述的前病毒质粒、如权利要求53-59中任一项所述的AAV或如权利要求60或61所述的组合物。

70. 如权利要求67-69中任一项所述的方法,所述方法包括视网膜下向所述受试者施用如权利要求1-46中任一项所述的核酸反式剪接分子、如权利要求47-51中任一项所述的载体、如权利要求52所述的前病毒质粒、如权利要求53-59中任一项所述的AAV或如权利要求60或61所述的组合物。

71. 如权利要求67-70中任一项所述的方法,其中所述受试者是非人灵长类动物或人。

72. 如权利要求67-71中任一项所述的方法,其中所述受试者已被诊断为患有ABCA4相关的视网膜营养不良。

73. 如权利要求1-46中任一项所述的核酸反式剪接分子、如权利要求47-51中任一项所述的载体、如权利要求52所述的前病毒质粒、如权利要求53-59中任一项所述的AAV或如权利要求60或61所述的组合物,其用于治疗有需要的受试者的ABCA4相关的视网膜营养不良。

74. 如权利要求1-46中任一项所述的核酸反式剪接分子、如权利要求47-51中任一项所述的载体、如权利要求52所述的前病毒质粒、如权利要求53-59中任一项所述的AAV或如权利要求60或61所述的组合物,其用于制备供治疗有需要的受试者的ABCA4相关的视网膜营养不良的药物。

75. 如权利要求67-72中任一项所述的方法或权利要求73或74中任一项的用途,其中所述ABCA4相关的视网膜营养不良与跨越所述ABCA4基因的外显子1-22的等位基因致病突变和与视锥-视杆细胞营养不良、斯特格氏黄斑营养不良或眼底黄色斑点症一致的表型相关。

76. 一种核酸反式剪接分子,所述核酸反式剪接分子包含:

(a) 接头结构域,所述接头结构域包含SEQ ID NO:27或与SEQ ID NO:27具有至少90%同一性的序列、基本上由其组成或由其组成;以及

(b) 任选地,增效子,所述增效子包含SEQ ID NO:62、基本上由其组成或由其组成;以及

(c) 任选地,翻译增效子,所述翻译增效子包含SEQ ID NO:63、基本上由其组成或由其组成,其中所述接头结构域;所述增效子在存在时;以及所述翻译增效子在存在时,在5'至3'方向上操作性连接。

77. 如权利要求76所述的核酸反式剪接分子,所述核酸反式剪接分子还包含结合结构域,所述结合结构域退火到内源ABCA4mRNA前体内的结合位点,

其中所述接头结构域;所述增效子在存在时;所述翻译增效子在存在时;以及所述结合结构域,在5'至3'方向上操作性连接。

78. 如权利要求76或权利要求77所述的核酸反式剪接分子,所述核酸反式剪接分子还包括包含ABCA4外显子的CDS,

其中所述CDS;所述接头结构域;所述增效子在存在时;所述翻译增效子在存在时;以及所述结合结构域在存在时,在5'至3'方向上操作性连接。

79. 一种5'核酸反式剪接分子,所述5'核酸反式剪接分子包含:

(a) 增效子,所述增效子包含GGTAAGT或SEQ ID NO:61或与GGTAAGT或SEQ ID NO:61具有至少90%同一性、至少91%同一性、至少92%同一性、至少93%同一性、至少94%同一性、至少95%同一性、至少96%同一性、至少97%同一性、至少98%同一性或至少99%同一性的序列中的至少一者、基本上由其组成或由其组成;以及

(b) 任选地,接头结构域,所述接头结构域包含SEQ ID NO:27或与SEQ ID NO:27具有至少90%同一性的序列、基本上由其组成或由其组成;

(c) 任选地,翻译增效子,所述翻译增效子包含SEQ ID NO:63、基本上由其组成或由其组成,

其中所述接头结构域在存在时;所述增效子;以及所述翻译增效子在存在时,在5'至3'方向上操作性连接。

80. 一种5'核酸反式剪接分子,所述5'核酸反式剪接分子包含:

(a) 增效子,所述增效子包含SEQ ID NO:62或与SEQ ID NO:62具有至少90%同一性的序列、基本上由其组成或由其组成;以及

(b) 任选地,接头结构域,所述接头结构域包含SEQ ID NO:27或与SEQ ID NO:27具有至少90%同一性的序列、基本上由其组成或由其组成;

(c) 任选地,翻译增效子,所述翻译增效子包含SEQ ID NO:63、基本上由其组成或由其组成,

其中所述接头结构域在存在时;所述增效子;以及所述翻译增效子在存在时,在5'至3'方向上操作性连接。

81. 如权利要求79或权利要求80所述的核酸反式剪接分子,所述核酸反式剪接分子还包含结合结构域,所述结合结构域退火到内源ABCA4 mRNA前体内的结合位点,

其中所述接头结构域在存在时;所述增效子;所述翻译增效子在存在时;以及所述结合结构域,在5'至3'方向上操作性连接。

82. 如权利要求79-81中任一项所述的核酸反式剪接分子,所述核酸反式剪接分子还包括包含ABCA4外显子的CDS,

其中所述CDS;所述接头结构域在存在时;所述增效子;所述翻译增效子在存在时;以及所述结合结构域在存在时,在5'至3'方向上操作性连接。

83. 一种核酸反式剪接分子,所述核酸反式剪接分子包含:

(a) 翻译增效子,所述翻译增效子包含SEQ ID NO:63或与SEQ ID NO:63具有至少90%同一性的序列、基本上由其组成或由其组成;

(b) 任选地,增效子,所述增效子包含SEQ ID NO:62、基本上由其组成或由其组成;以及

(c) 任选地,接头结构域,所述接头结构域包含SEQ ID NO:27或与SEQ ID NO:27具有至少90%同一性的序列、基本上由其组成或由其组成,其中所述接头结构域在存在时;所述增效子在存在时;以及所述翻译增效子,在5'至3'方向上操作性连接。

84. 如权利要求83所述的核酸反式剪接分子,所述核酸反式剪接分子还包含结合结构域,所述结合结构域退火到内源ABCA4mRNA前体内的结合位点,

其中所述接头结构域在存在时;所述增效子在存在时;所述翻译增效子;以及所述结合结构域,在5'至3'方向上操作性连接。

85. 如权利要求83或权利要求84所述的核酸反式剪接分子,所述核酸反式剪接分子还包括包含ABCA4外显子的CDS,

其中所述CDS;所述接头结构域在存在时;所述增效子在存在时;所述翻译增效子;以及所述结合结构域在存在时,在5'至3'方向上操作性连接。

86. 一种核酸反式剪接分子,所述核酸反式剪接分子包含:

包含接头结构域和增效子的盒,其中所述盒包含SEQ ID NO:72、基本上由其组成或由其组成。

87. 如权利要求86所述的核酸反式剪接分子,所述核酸反式剪接分子还包含结合结构域,所述结合结构域退火到内源ABCA4mRNA前体内的结合位点,

其中所述盒和所述结合结构域在5'至3'方向上操作性连接。

88. 如权利要求86或权利要求87所述的核酸反式剪接分子,所述核酸反式剪接分子还

包括包含ABCA4外显子的CDS,

其中所述CDS;所述盒;以及所述结合结构域在存在时,在5'至3'方向上操作性连接。

89.一种核酸反式剪接分子,所述核酸反式剪接分子包含:

包含接头结构域、增效子和翻译增效子的盒,其中所述盒包含SEQ ID NO:73、基本上由其组成或由其组成。

90.如权利要求89所述的核酸反式剪接分子,所述核酸反式剪接分子还包含结合结构域,所述结合结构域退火到内源ABCA4mRNA前体内的结合位点,

其中所述盒和所述结合结构域在5'至3'方向上操作性连接。

91.如权利要求89或权利要求90所述的核酸反式剪接分子,所述核酸反式剪接分子还包括包含ABCA4外显子的CDS,

其中所述CDS;所述盒;以及所述结合结构域在存在时,在5'至3'方向上操作性连接。

92.一种核酸反式剪接分子,所述核酸反式剪接分子包含:

(a) 3'转录终止子结构域,所述3'转录终止子结构域包含SEQ ID NO:66或与SEQ ID NO:66具有至少95%同一性的序列、基本上由其组成或由其组成。

93.如权利要求92所述的核酸反式剪接分子,其中所述3'转录终止子结构域包含如下核酸序列,所述核酸序列与SEQ ID NO:66具有至少96%同一性;与SEQ ID NO:66具有至少97%同一性;与SEQ ID NO:66具有至少98%同一性;与SEQ ID NO:66具有至少99%同一性;或与SEQ ID NO:66100%同一。

94.如权利要求92或权利要求93所述的核酸反式剪接分子,其中所述3'转录终止子结构域包含SEQ ID NO:66、基本上由其组成或由其组成。

## ABCA4反式剪接分子

[0001] 交叉参考

[0002] 本申请要求2022年5月13日提交的美国临时专利申请No.63/341,665和2023年1月4日提交的美国临时专利申请No.63/478,472的权益,所述临时专利申请以引用的方式整体并入本文中。

[0003] 序列表

[0004] 本申请含有序列表,所述序列表已经以XML格式以电子方式提交并且在此以引用的方式整体并入。该XML拷贝创建于2023年5月10日,名为61313-703\_601\_SL.xml,大小为1,315,456字节。

### 技术领域

[0005] 本文描述了涉及用于纠正ABCA4 (ATP结合盒亚家族A成员4) 中的突变的反式剪接分子的组合物、方法和用途。

### 背景技术

[0006] ABCA4的隐性遗传突变是导致渐进性失明,包括斯特格氏病 (Stargardt Disease) 和视锥-视杆细胞营养不良的原因。ABCA4位于感光细胞外节视盘膜的边缘区域,起到将N-亚视黄基-磷脂酰乙醇胺 (N-Ret-PE) 从管腔转运到视盘膜细胞质小叶的转运蛋白(或“翻转酶”)的作用。ABCA4蛋白缺陷会导致有毒类视黄酸化合物的积累,最终导致光感受器和视网膜色素上皮细胞死亡。ABCA4的6882bp编码序列太大,无法通过单个AAV载体完整递送。此外,由于在整个基因中发现了数百种致病突变,所以单一的碱基编辑方法无法解决大量患者的问题。因此,本领域需要纠正ABCA4基因中的突变的替代策略。

### 发明内容

[0007] ABCA4相关的视网膜病变,例如斯特格氏病,是影响成人和儿童的严重罕见遗传性视网膜疾病,由ABCA4基因的双等位基因致病突变引起。据估计,全世界每8,000至10,000人中就有1人患有斯特格氏病,约占视网膜营养不良的7%。ABCA4被认定为导致常染色体隐性遗传斯特格氏病的基因,然而其它几种视网膜表型也与ABCA4基因突变有关。事实上,ABCA4相关的视网膜病变被统一认为是由于孟德尔遗传导致的视网膜变性的最常见原因。患有ABCA4相关的视网膜病变的个体通常在生命的前二十年内被诊断出来,并且根据特定突变的严重程度,可能在成年时完全失去中央视野。

[0008] 目前,尚无批准的针对ABCA4相关的视网膜病变的治疗。因此,针对受到功能丧失型ABCA4突变影响的成人和儿童,开发改善疾病的疗法存在着巨大的未满足的医疗需求。

[0009] 通过广泛的研究,本发明人已鉴定出赋予所掺入的核酸反式剪接分子特性的核酸元件,这些特性解决了与例如提高反式剪接效率和由此产生的反式剪接蛋白的水平相关的技术挑战。因此,核酸反式剪接分子(例如,RNA反式剪接分子)中包括所鉴定的元件为过去与实施核酸反式剪接分子相关的问题(包括与反式剪接效率低下有关的问题)提供了多种

解决方案。因此,本发明人设计了核酸反式剪接分子,实现了为与基因突变相关的疾病的治疗干预提供选择的长期寻求的目标,从而在RNA水平上纠正那些基因突变可以为患者带来巨大的益处。许多患有与基因突变相关的疾病的患者几乎没有治疗选择,因此他们的需求基本上得不到满足。例如,患有与ABCA4基因突变相关的渐进性失明的患者代表着一个需求未得到满足的患者群体。

[0010] 为了满足这些未满足的需求,本文描述了一种反式剪接解决方案(也称为外显子编辑),其包括递送治疗性核酸反式剪接构建体,该构建体编码RNA外显子编辑器,该RNA外显子编辑器被设计为反式剪接到包含至少一个突变的内源ABCA4 mRNA前体中,其中反式剪接用来自治疗性RNA外显子编辑器的编码功能性(例如,野生型)氨基酸序列的外显子替换内源ABCA4 mRNA前体中的外显子(其中至少一个外显子包含至少一个突变),从而纠正内源ABCA4 mRNA前体中的至少一个突变。在一些实施方案中, RNA外显子编辑器可以直接递送到细胞。术语“核酸反式剪接分子”、“RNA外显子编辑分子”、“RNA外显子编辑器”、“外显子编辑器”和“RNA反式剪接构建体”在本文中可互换使用。

[0011] 在一些实施方案中,本文描述一种核酸反式剪接分子,所述核酸反式剪接分子包含:(a) 包含ABCA4外显子的序列(CDS); (b) 接头结构域,所述接头结构域包含SEQ ID NO:27或与SEQ ID NO:27具有至少90%同一性的序列;以及(c) 结合结构域,所述结合结构域退火到内源ABCA4 mRNA前体内的结合位点,其中CDS、接头结构域和结合结构域在5'至3'方向上操作性连接。本文还描述了其治疗与ABCA4突变相关的视网膜病变的用途和在制备用于治疗与ABCA4突变相关的视网膜病变的药物中的用途,以及治疗与ABCA4突变相关的视网膜病变的方法。

[0012] 在一些实施方案中,本文描述一种核酸反式剪接分子,所述核酸反式剪接分子包含:(a) CDS,所述CDS包含SEQ ID NO:53-55中的任一者的变体、基本上由其组成或由其组成,其中SEQ ID NO:53-55中的任一者的变体在表3中列出的至少一个隐蔽剪接位点中包含至少一个核苷酸变异,其中所述至少一个核苷酸变异减少包含至少一个核苷酸变异的至少一个隐蔽剪接位点的每一者处隐蔽剪接位点的使用;以及(b) 任选地,接头结构域;以及(c) 任选地,结合结构域,所述结合结构域退火到内源ABCA4 mRNA前体内的结合位点,其中CDS;接头结构域在存在时;以及结合结构域在存在时;在5'至3'方向上操作性连接。在一些实施方案中,SEQ ID NO:55的变体包含SEQ ID NO:56-59中的任一者。在一些实施方案中,SEQ ID NO:55的变体包含SEQ ID NO:56。本文还描述了其治疗与ABCA4突变相关的视网膜病变的用途和在制备用于治疗与ABCA4突变相关的视网膜病变的药物中的用途,以及治疗与ABCA4突变相关的视网膜病变的方法。

[0013] 在一方面,本文提供了一种核酸反式剪接分子(或含有或编码核酸反式剪接分子的载体),其中所述核酸反式剪接分子含有在5'至3'方向上操作性连接的以下各者:(a) 5'调节结构域,所述5'调节结构域包含例如天然5' ABCA4非翻译区;(b) CDS,所述CDS包含5' ABCA4外显子的功能序列;(c) 剪接结构域,所述剪接结构域被配置成介导反式剪接;以及(d) 结合结构域,所述结合结构域被配置成结合内源ABCA4 mRNA前体内的结合位点,其中所述核酸反式剪接分子被配置成将CDS反式剪接到内源ABCA4,从而用5' ABCA4外显子的功能序列替换内源5' ABCA4外显子序列。在一些实施方案中,5'调节结构域还包括组成型启动子,例如CMV/CMV启动子(CMV增强子和启动子)或其变体。在一些实施方案中,天然5' ABCA4

非翻译区包含与SEQ ID NO:12或13具有至少80%序列同一性(例如,与SEQ ID NO:12或13具有至少85%序列同一性、至少90%序列同一性、至少91%序列同一性、至少92%序列同一性、至少93%序列同一性、至少94%序列同一性、至少95%序列同一性、至少96%序列同一性、至少97%序列同一性、至少98%序列同一性、至少99%序列同一性或100%序列同一性)的核酸序列。在一些实施方案中,5'调节结构域包含与SEQ ID NO:14或15具有至少80%序列同一性(例如,与SEQ ID NO:14或15具有至少85%序列同一性、至少90%序列同一性、至少91%序列同一性、至少92%序列同一性、至少93%序列同一性、至少94%序列同一性、至少95%序列同一性、至少96%序列同一性、至少97%序列同一性、至少98%序列同一性、至少99%序列同一性或100%序列同一性)的核酸序列。

[0014] 在一些实施方案中,结合位点在ABCA4内含子22内。在一些实施方案中,结合位点包含SEQ ID NO:16的核苷酸1至510或880至1,350中的任一个或多个(例如,SEQ ID NO:16的核苷酸1至510或880至1,350中的任六个或更多个、任八个或更多个、任10个或更多个、任12个或更多个、任18个或更多个、任20个或更多个、任24个或更多个、任50个或更多个、任100个或更多个、或者任150个或更多个(例如,150个))。在一些实施方案中,结合位点包含SEQ ID NO:16的核苷酸1160至1309中的任一个或多个(例如,SEQ ID NO:16的核苷酸1160至1309中的任六个或更多个、任八个或更多个、任10个或更多个、任12个或更多个、任18个或更多个、任20个或更多个、任24个或更多个、任50个或更多个、任100个或更多个、或者任150个或更多个(例如,150个))。在一些实施方案中,结合于结合位点的结合结构域与SEQ ID NO:16内的结合位点具有最少80%序列同一性(例如,与SEQ ID NO:16内的结合位点具有至少85%序列同一性、至少90%序列同一性、至少91%序列同一性、至少92%序列同一性、至少93%序列同一性、至少94%序列同一性、至少95%序列同一性、至少96%序列同一性、至少97%序列同一性、至少98%序列同一性、至少99%序列同一性或100%序列同一性。在一些实施方案中,结合结构域包含SEQ ID NO:17或18。

[0015] 在一些实施方案中,剪接结构域包含GTAAGT、GUAAGT、GTAAGG、GUAAGG、GTAAGC、GUAAGC、GTAAC或GUAACU。

[0016] 在一些实施方案中,核酸反式剪接分子还包含接头结构域。在一些实施方案中,接头结构域长度超过25个核苷酸(例如,长度在25个与50个核苷酸之间、长度在35个与45个核苷酸之间或长度为约40个核苷酸(例如,40聚体接头))。在一些情况下,接头结构域为或包含与CTGGTGCCCGCGGCCGCGGAACCGTTGGGGCATGTAC(SEQ ID NO:27)具有至少80%同一性(例如,至少85%序列同一性、至少90%序列同一性、至少91%序列同一性、至少92%序列同一性、至少93%序列同一性、至少94%序列同一性、至少95%序列同一性、至少96%序列同一性、至少97%序列同一性、至少98%序列同一性、至少99%序列同一性或100%序列同一性)的核酸序列。在其它情况下,接头结构域为或包含与CCGAATACGACACGTACAAGATCT(SEQ ID NO:29)具有至少80%同一性(例如,至少85%序列同一性、至少90%序列同一性、至少91%序列同一性、至少92%序列同一性、至少93%序列同一性、至少94%序列同一性、至少95%序列同一性、至少96%序列同一性、至少97%序列同一性、至少98%序列同一性、至少99%序列同一性或100%序列同一性)的核酸序列。在其它情况下,接头结构域为或包含与GCTGCTGCTCAGTCTCCTGGGCTGG(SEQ ID NO:31)具有至少80%同一性(例如,至少85%序列同一性、至少90%序列同一性、至少91%序列同一性、至少92%序列同一性、至少93%序列同

一性、至少94%序列同一性、至少95%序列同一性、至少96%序列同一性、至少97%序列同一性、至少98%序列同一性、至少99%序列同一性或100%序列同一性)的核酸序列。

[0017] 在另一方面,提供了一种核酸反式剪接分子,所述核酸反式剪接分子在5'至3'方向上包含操作性连接的以下各者:(a) CDS,所述CDS包含5'ABCA4外显子的功能序列;(b) 剪接结构域,所述剪接结构域被配置成介导反式剪接,其中所述剪接结构域包含GTAAGT、GUAAGT、GTAAGG、GUAAGG、GTAAGC、GUAAGC、GTAAGT或GUAACU;以及(c) 结合结构域,所述结合结构域被配置成结合内源ABCA4内含子22内的结合位点,其中所述核酸反式剪接分子被配置成将CDS反式剪接到内源ABCA4外显子23,从而用5'ABCA4外显子的功能序列替换内源5'ABCA4外显子序列。在一些实施方案中,核酸反式剪接分子包含在5'操作性连接到CDS的5'调节结构域。在一些实施方案中,5'调节结构域包含天然5'ABCA4非翻译区,例如与SEQ ID NO:12或13具有至少80%序列同一性(例如,与SEQ ID NO:12或13具有至少85%序列同一性、至少90%序列同一性、至少91%序列同一性、至少92%序列同一性、至少93%序列同一性、至少94%序列同一性、至少95%序列同一性、至少96%序列同一性、至少97%序列同一性、至少98%序列同一性、至少99%序列同一性或100%序列同一性)的天然5'ABCA4非翻译区。在一些实施方案中,5'调节结构域还包含组成型启动子,例如CMV启动子。

[0018] 在一些实施方案中,核酸反式剪接分子还包含接头结构域。在一些实施方案中,接头结构域长度超过25个核苷酸(例如,长度在25个与50个核苷酸之间、长度在35个与45个核苷酸之间或长度为约40个核苷酸(例如,40聚体接头))。在一些情况下,接头结构域为或包含与CTGGTGGCCGCGGGCCGCGGAACCGGTTGGGGCATGTAC(SEQ ID NO:27)具有至少80%同一性(例如,至少85%序列同一性、至少90%序列同一性、至少91%序列同一性、至少92%序列同一性、至少93%序列同一性、至少94%序列同一性、至少95%序列同一性、至少96%序列同一性、至少97%序列同一性、至少98%序列同一性、至少99%序列同一性或100%序列同一性)的核酸序列。在其它情况下,接头结构域为或包含与CCGAATACGACACGTACAAGATCT(SEQ ID NO:29)具有至少80%同一性(例如,至少85%序列同一性、至少90%序列同一性、至少91%序列同一性、至少92%序列同一性、至少93%序列同一性、至少94%序列同一性、至少95%序列同一性、至少96%序列同一性、至少97%序列同一性、至少98%序列同一性、至少99%序列同一性或100%序列同一性)的核酸序列。在其它情况下,接头结构域为或包含与GCTGCTGCTCAGTCTCTGGGCTGG(SEQ ID NO:31)具有至少80%同一性(例如,至少85%序列同一性、至少90%序列同一性、至少91%序列同一性、至少92%序列同一性、至少93%序列同一性、至少94%序列同一性、至少95%序列同一性、至少96%序列同一性、至少97%序列同一性、至少98%序列同一性、至少99%序列同一性或100%序列同一性)的核酸序列。

[0019] 在一些实施方案中,结合位点包含SEQ ID NO:16的核苷酸1至510或880至1,350中的任一个或多个(例如,SEQ ID NO:16的核苷酸1至510或880至1,350中的任六个或更多个、任八个或更多个、任10个或更多个、任12个或更多个、任18个或更多个、任20个或更多个、任24个或更多个、任50个或更多个、任100个或更多个、或者任150个或更多个(例如,150个))。在一些实施方案中,结合位点包含SEQ ID NO:16的核苷酸1160至1309中的任一个或多个(例如,SEQ ID NO:16的核苷酸1160至1309中的任六个或更多个、任八个或更多个、任10个或更多个、任12个或更多个、任18个或更多个、任20个或更多个、任24个或更多个、任50个或更多个、任100个或更多个、或者任150个或更多个(例如,150个))。在一些实施方案中,结合

于结合位点的结合结构域与SEQ ID NO:17或18具有最少80%序列同一性(例如,与SEQ ID NO:17或18具有至少85%序列同一性、至少90%序列同一性、至少91%序列同一性、至少92%序列同一性、至少93%序列同一性、至少94%序列同一性、至少95%序列同一性、至少96%序列同一性、至少97%序列同一性、至少98%序列同一性、至少99%序列同一性或100%序列同一性)。在一些实施方案中,结合结构域包含SEQ ID NO:17或18。

[0020] 在另一方面,提供了一种核酸反式剪接分子,所述核酸反式剪接分子在5'至3'方向上包含操作性连接的以下各者:(a)CDS,所述CD S包含5'ABCA4外显子的功能序列;(b)剪接结构域,所述剪接结构域被配置成介导反式剪接;(c)长度在25个与50个核苷酸之间的接头结构域;以及(d)结合结构域,所述结合结构域被配置成结合内源ABCA4内含子22内的结合位点,其中所述核酸反式剪接分子被配置成将CDS反式剪接到内源ABCA4外显子23,从而用5'ABCA4外显子的功能序列替换内源5'ABCA4外显子序列。在一些实施方案中,接头结构域长度在35个与45个核苷酸之间或长度为约40个核苷酸(例如,40聚体接头)。在一些情况下,接头结构域为或包含与CTGG TGCCGCGGGCCGCGGAACCGTTGGGGGCATGTAC (SEQ ID NO:27)具有至少80%同一性(例如,至少85%序列同一性、至少90%序列同一性、至少91%序列同一性、至少92%序列同一性、至少93%序列同一性、至少94%序列同一性、至少95%序列同一性、至少96%序列同一性、至少97%序列同一性、至少98%序列同一性、至少99%序列同一性或100%序列同一性)的核酸序列。在其它情况下,接头结构域为或包含与CCGAATACGACACGTACAAGATCT (SEQ ID NO:29)具有至少80%同一性(例如,至少85%序列同一性、至少90%序列同一性、至少91%序列同一性、至少92%序列同一性、至少93%序列同一性、至少94%序列同一性、至少95%序列同一性、至少96%序列同一性、至少97%序列同一性、至少98%序列同一性、至少99%序列同一性或100%序列同一性)的核酸序列。在其它情况下,接头结构域为或包含与GCTGCTGCTCAGTCTCCTGGGCTGG (SEQ ID NO:31)具有至少80%同一性(例如,至少85%序列同一性、至少90%序列同一性、至少91%序列同一性、至少92%序列同一性、至少93%序列同一性、至少94%序列同一性、至少95%序列同一性、至少96%序列同一性、至少97%序列同一性、至少98%序列同一性、至少99%序列同一性或100%序列同一性)的核酸序列。

[0021] 在一些实施方案中,核酸反式剪接分子包含在5'操作性连接到CDS的5'调节结构域。在一些实施方案中,5'调节结构域包含天然5'ABCA4非翻译区,例如与SEQ ID NO:12或13具有至少80%序列同一性(例如,与SEQ ID NO:12或13具有至少85%序列同一性、至少90%序列同一性、至少91%序列同一性、至少92%序列同一性、至少93%序列同一性、至少94%序列同一性、至少95%序列同一性、至少96%序列同一性、至少97%序列同一性、至少98%序列同一性、至少99%序列同一性或100%序列同一性)的天然5'ABCA4非翻译区。在一些实施方案中,5'调节结构域还包含组成型启动子,例如CMV/CMV启动子。

[0022] 在一些实施方案中,剪接结构域包含GTAAGT、GUAAGT、GTAAGG、GUAAGG、GTAAGC、GUAAGC、GTAAGT或GUAACU。

[0023] 在一些实施方案中,结合位点包含SEQ ID NO:16的核苷酸1至510或880至1,350中的任一个或多个(例如,SEQ ID NO:16的核苷酸1至510或880至1,350中的任六个或更多个、任八个或更多个、任10个或更多个、任12个或更多个、任18个或更多个、任20个或更多个、任24个或更多个、任50个或更多个、任100个或更多个、或者任150个或更多个(例如,150个))。

在一些实施方案中,结合位点包含SEQ ID NO:16的核苷酸1160至1309中的任一个或多个(例如,SEQ ID NO:16的核苷酸1160至1309中的任六个或更多个、任八个或更多个、任10个或更多个、任12个或更多个、任18个或更多个、任20个或更多个、任24个或更多个、任50个或更多个、任100个或更多个、或者任150个或更多个(例如,150个))。在一些实施方案中,结合于结合位点的结合结构域与SEQ ID NO:17或18具有最少80%序列同一性(例如,与SEQ ID NO:17或18具有至少85%序列同一性、至少90%序列同一性、至少91%序列同一性、至少92%序列同一性、至少93%序列同一性、至少94%序列同一性、至少95%序列同一性、至少96%序列同一性、至少97%序列同一性、至少98%序列同一性、至少99%序列同一性或100%序列同一性)。在一些实施方案中,结合结构域包含SEQ ID NO:17或18。

[0024] 在任一前述方面的一些实施方案中,核酸反式剪接分子包含3'转录终止子结构域。在一些实施方案中,3'转录终止子结构域形成将反式剪接分子的3'端有效封端的三螺旋结构。在一些实施方案中,3'转录终止子结构域包含野生型MALAT1序列。在一些实施方案中,3'转录终止子结构域包含与SEQ ID NO:32或33具有至少80%序列同一性(例如,与SEQ ID NO:32或33具有至少85%序列同一性、至少90%序列同一性、至少91%序列同一性、至少92%序列同一性、至少93%序列同一性、至少94%序列同一性、至少95%序列同一性、至少96%序列同一性、至少97%序列同一性、至少98%序列同一性、至少99%序列同一性或100%序列同一性)的核酸序列。

[0025] 在任一前述方面的一些实施方案中,CDS包含ABCA4外显子1-22或其变体(例如,其密码子优化变体)。

[0026] 在任一前述方面的一些实施方案中,核酸反式剪接分子包含与SEQ ID NO:40-50中的任一者具有至少80%序列同一性(例如,与SEQ ID NO:40-50中的任一者具有至少85%序列同一性、至少90%序列同一性、至少91%序列同一性、至少92%序列同一性、至少93%序列同一性、至少94%序列同一性、至少95%序列同一性、至少96%序列同一性、至少97%序列同一性、至少98%序列同一性、至少99%序列同一性或100%序列同一性)的核酸序列。

[0027] 在任一前述方面的一些实施方案中,内源ABCA4 mRNA前体具有至少一个与ABCA4相关的视网膜营养不良相关的突变(例如,隐性遗传的ABCA4相关的视网膜营养不良,例如斯特格氏病或视锥-视杆细胞营养不良)。在一些实施方案中,突变在感光细胞和/或视网膜色素上皮细胞中表达。

[0028] 在另一方面,描述了一种核酸反式剪接分子,所述核酸反式剪接分子在5'至3'方向上包含操作性连接的以下各者:(a) CDS,所述CDS包含5'ABCA4外显子的功能序列;(b) 剪接结构域,所述剪接结构域被配置成介导反式剪接;以及(c) 结合结构域,所述结合结构域与SEQ ID NO:17-24中的任一者具有至少80%序列同一性(例如,至少85%序列同一性、至少90%序列同一性、至少91%序列同一性、至少92%序列同一性、至少93%序列同一性、至少94%序列同一性、至少95%序列同一性、至少96%序列同一性、至少97%序列同一性、至少98%序列同一性、至少99%序列同一性或100%序列同一性),其中所述核酸反式剪接分子被配置成将CDS反式剪接到内源ABCA4外显子23,从而用5'ABCA4外显子的功能序列替换内源5'ABCA4外显子序列。

[0029] 在另一方面,本文提供了一种包含SEQ ID NO:18的分离的DNA分子或其载体。

[0030] 在另一方面,本文提供了一种包含SEQ ID NO:17的分离的RNA分子或其载体。

[0031] 在另一方面,本文提供了一种分离的核酸(例如,DNA或RNA),其编码任一前述实施方案或任一前述方面的核酸反式剪接分子,例如编码任一前述实施方案或任一前述方面的核酸反式剪接分子的载体。

[0032] 在另一方面,本文提供了一种前病毒质粒,所述前病毒质粒包含任一前述实施方案或任一前述方面的核酸反式剪接分子。

[0033] 在另一方面,本文提供了一种腺相关病毒(AAV),所述腺相关病毒包含编码任一前述实施方案或任一前述方面的核酸反式剪接分子的序列。在一些实施方案中,AAV优先靶向感光细胞和/或视网膜色素上皮细胞。在一些实施方案中,AAV是AAV8、AAV5或AAV2。

[0034] 在另一方面,本文提供了一种组合物,所述组合物包含任一前述实施方案或任一前述方面的核酸反式剪接分子、载体、前病毒质粒或AAV。在一些实施方案中,组合物包括药学上可接受的赋形剂。

[0035] 在另一方面,本文提供了一种在靶细胞中(例如,体外或体内)表达功能性ABCA4的方法。在一些实施方案中,方法包括将靶细胞用任一前述方面的核酸反式剪接分子、载体、AAV或组合物转导。在一些实施方案中,方法替换了靶细胞中20%或更多的靶ABCA4 mRNA(例如,在培养的人体细胞中,例如通过反式剪接的ABCA4 RNA或由反式剪接的RNA翻译产生的相应蛋白质测量)。在一些实施方案中,方法替换了靶细胞中38%或更多的靶ABCA4 mRNA(例如,在培养的人细胞中,例如通过反式剪接的ABCA4RNA或由反式剪接的RNA翻译产生的相应蛋白质测量)。在一些实施方案中,功能性ABCA4(由反式剪接的RNA翻译产生的相应蛋白质)是全长ABCA4蛋白。

[0036] 在另一方面,本文提供了一种减少受试者(例如哺乳动物,例如灵长类动物,例如人)中ABCA4视网膜病变相关的脂褐素和/或A2E积累的方法。在一些实施方案中,所述方法包括将受试者中的靶视网膜细胞用任一前述实施方案或任一前述方面的核酸反式剪接分子、载体、AAV或组合物转导。在一些实施方案中,受试者是非人灵长类动物或人(例如,已被诊断为患有ABCA4相关的视网膜营养不良的人)。在一些实施方案中,通过眼部施用(例如,通过视网膜下施用)向受试者施用核酸反式剪接分子、载体或AAV。

[0037] 在另一方面,本文提供了一种纠正受试者(例如哺乳动物,例如灵长类动物,例如人)的靶细胞中5'ABCA4外显子序列的至少一个突变的方法。在一些实施方案中,所述方法包括向受试者施用任一前述实施方案或任一前述方面的核酸反式剪接分子、载体、组合物或AAV。在一些实施方案中,受试者是非人灵长类动物或人(例如,已被诊断为患有ABCA4相关的视网膜营养不良的人)。在一些实施方案中,通过眼部施用(例如,通过视网膜下施用)向受试者施用核酸反式剪接分子、载体或AAV。

[0038] 在另一方面,本文提供了一种治疗受试者(例如哺乳动物,例如灵长类动物,例如人)的ABCA4相关的视网膜营养不良的方法。在一些实施方案中,所述方法包括以治疗有效量向受试者施用任一前述实施方案或任一前述方面的核酸反式剪接分子、载体、AAV或组合物。在一些实施方案中,ABCA4相关的视网膜营养不良与5'ABCA4外显子序列中的突变(例如,ABCA4外显子1-22中的任一者或多者中的突变)相关。在一些实施方案中,ABCA4相关的视网膜营养不良是斯特格氏病或视锥-视杆细胞营养不良。

## 附图说明

[0039] 图1是显示具有各种剪接结构域(SS)、接头和结合结构域(BD)排列的ABCA4反式剪接分子的反式剪接效率的图。

[0040] 图2是显示具有各种剪接结构域序列的ABCA4反式剪接分子的反式剪接效率的图。

[0041] 图3A和3B显示了AAV转导的各种ABCA4反式剪接分子在培养的人细胞中的反式剪接效率。图3A显示了RT-qPCR的结果。图3B和3C显示了蛋白质印迹的结果。图3D是定量图3C的结果的图。

[0042] 图4显示了AAV8反式剪接分子制剂的蛋白质印迹,显示了VP1、VP2和VP3条带。

[0043] 图5A-5C显示了在视网膜下注射后,GFP反式剪接分子在非人灵长类动物(NHP)体内的GFP蛋白表达。图5A是显示GFP蛋白的实时图像。图5B是显示第28天阴性对照中的视网膜切片(光感受器和脉络膜被标记)的抗GFP免疫组织化学(IHC)染色的显微照片。图5C是显示第28天经过GFP反式剪接分子处理的动物的视网膜切片(光感受器和脉络膜被标记)的抗GFP IHC染色的显微照片。

[0044] 图6是基线、第14天和第28天的视网膜下注射后的一组cSLO、OCT图像。水泡在“注射后”图上标记。

[0045] 图7是第28天用媒介物、外显子编辑器443和外显子编辑器274处理的NHP眼睛的一组彩色眼底图像(前视图、后视图、上视图和下视图)。

[0046] 图8A是显示反式剪接分子的RT-qPCR测试方法的示意图。带有5'标签的反式剪接分子(外显子编辑器RNA)反式剪接到AGM mRNA前体中,并且通过其所得到的独特序列对编辑的(例如,反式剪接的)全长mRNA进行定量。图8B显示了外显子编辑器443、274、497和505的定量RT-qPCR结果。如所示,外显子编辑器443、274、497和505中的每一个都以统计上显著的频率与体内AGM视网膜中的内源AGM mRNA前体进行反式剪接,该频率通过定量每10ng RNA中编辑的拷贝数来确定。

[0047] 图9是显示因外显子编辑器443、274、497和505反式剪接到AGM视网膜中的AGMmRNA前体中而产生的全长嵌合(人类/AGM)ABCA4蛋白的蛋白质印迹。在视网膜的膜提取物中检测到了蛋白质标签。

[0048] 图10A和10B是显示人和AGMABCA4蛋白的独特肽的检测的质谱结果。图10A显示了阴性对照(媒介物处理)视网膜组织的结果。顶行显示在人蛋白质尖峰应出现的位置缺少尖峰的对照样品;第二行显示加入了标准品的对照样品,以显示人蛋白质尖峰会出现的位置;第三行显示含有代表AGM蛋白的尖峰的对照样品;并且底行显示加入了标准品的样品,其含有AGM蛋白尖峰。图10B显示了经过外显子编辑器处理的视网膜组织的结果,显示检测到人和AGM肽。顶行显示对照样品,其表现出人肽尖峰,这反映了人/AGM嵌合蛋白的存在(与图10A中所示的阴性对照形成对比);第二行显示加入了标准品的对照样品;第三行显示含有代表AGM蛋白的尖峰的对照样品;并且底行显示加入了标准品的样品,其含有AGM蛋白尖峰。

[0049] 图11A-11F是显示用视网膜下施用的AAV8外显子编辑器或媒介物处理后AGM眼部健康状况的条形图。数据显示为媒介物的平均值和外显子编辑器443、274、497和505的平均值±标准差。图11A显示了基线、第14天和第28天的视网膜厚度。图11B显示了基线、第7天、第14天、第21天和第28天通过眼压计测量的眼内压。图11C-11F显示了眼部炎症指标的结果。图11C显示了基线、第7天、第14天、第21天和第28天的总临床评分;图11D显示了基线、第

7天、第14天、第21天和第28天的玻璃体细胞评分;图11E显示了基线、第7天、第14天、第21天和第28天的房水细胞评分;并且图11F显示了基线、第7天、第14天、第21天和第28天的炎症角膜颗粒。

[0050] 图12A和12B显示了用视网膜下施用的AAV8外显子编辑器或媒介物处理后AGM的一般健康指标的结果。图12A显示了基线、第7天、第14天、第21天和第28天的体重测量值。图12B显示了基线和第28天的速率(心跳/分钟(bpm))。

[0051] 图13描绘了示例性RNA外显子编辑器的元件,该编辑器被设计用于将外显子编辑器的人ABCA4基因的外显子1-22反式剪接到内源ABCA4 mRNA前体转录本的外显子23-50。示例性外显子编辑器被设计用于治疗携带跨越ABCA4外显子1-22的突变的受试者。数字标签如下:1:基于AAV2的反向末端重复序列(ITR),用于ABCA4RNA外显子编辑器的AAV包装;2:巨细胞病毒(CMV)启动子(CMV即刻早期增强子和CMV即刻早期启动子)驱动ABCA4 RNA外显子编辑器在感光细胞中的表达;3:来自人ABCA4 mRNA的天然5'非翻译区(UTR)保持天然5'调节;4:ABCA4外显子1-21,其包含同义密码子变化,通过反式剪接替换相应的内源编码序列。同义密码子变化可用于例如去除隐蔽的剪接受体位点以减少潜在的替代剪接反应;5:ABCA4外显子22,其包含使得能够进行基于序列的定量的同义密码子变化;6:剪接供体(本文中也称为剪接结构域)反式剪接到内源ABCA4内含子22剪接受体位点,引起内源ABCA4mRNA前体上游序列被包括外显子22在内的ABCA4 RNA外显子编辑器编码结构域序列(CDS)替换;7:终止密码子(双终止)减少/阻止在外显子22下游(例如接头、3×U1、ARE、BD结构域区域)氨基酸从ABCA4外显子编辑器翻译;8:接头为剪接结构域位点提供结构灵活性和可及性;9:三个串联U1结合位点(3×U1),预测它们会与U1 snRNP缔合,并且本文证实它们会增强ABCA4 RNA外显子编辑器介导的活性;10:富AU元件(ARE),本文证实它们会降低进入细胞质的任何未剪接的ABCA4 RNA外显子编辑器的稳定性;11:结合结构域(BD)通过序列互补性介导ABCA4 RNA外显子编辑器靶向内源ABCA4 mRNA前体内含子22;12:MALAT1终止子促进有效的反式剪接;13:基于AAV2的ITR,用于外显子编辑器盒的AAV包装。与所描绘的示例性RNA外显子编辑器相对应的示例性序列呈现于SEQ ID NO:90(ABCA4-01)中。SEQ ID NO:90对应于SEQ ID NO:82的核苷酸4284-9073。SEQ ID NO:90包括SEQ ID NO:78(从5'UTR跨越到包含抗mut1的MALAT1-Masc RNA的末端)和SEQ ID NO:69(从5'UTR内的ATG跨越到包含抗mut1的MALAT1-Masc RNA的末端)。SEQ ID NO:82编码SEQ ID NO:90。

[0052] 图14A和14B描绘了RNA反式剪接和顺式剪接反应。(A)示出了RNA反式剪接反应示意图,其中针对ABCA4 RNA外显子编辑器呈现了通过反式剪接进行外显子编辑的一般机制。示例性ABCA4外显子编辑器RNA在靶向的ABCA4 mRNA前体上方描绘。与内含子22互补的结合结构域(BD)将外显子编辑器与mRNA前体缔合,并使剪接体介导的反式剪接反应在外显子编辑器的剪接供体(SD)位点(或剪接结构域)与天然内含子22剪接受体(SA)位点之间发生,从而用外显子编辑器提供的编码序列(CDS)替换靶向的mRNA前体的上游区域。mRNA前体分子逐渐经历常见的顺式剪接过程,去除任何残留的下游内含子,产生成熟的mRNA,然后可以进行翻译,形成功能性(生物活性)ABCA4蛋白。内源mRNA前体图中的红色“x”突变描绘了单个mRNA前体转录本上ABCA4基因中存在的代表性个别等位基因突变。SD和SA站点以红色箭头表示。通过反式剪接形成的外显子-外显子接合处以红色圈出。(B)描绘了RNA顺式剪接和反式剪接机制。RNA顺式剪接和反式剪接均依赖内源剪接体介导剪接,并从位于多聚嘧啶区

(Py)上游的分支点(BP)与SD位点的第一个核苷酸之间形成共价键开始。在顺式剪接过程中会形成套索,而在反式剪接过程中会形成Y型分支。接下来,剪接体介导上游外显子与下游外显子的连接,而套索和Y型分支分子则迅速降解。

[0053] 图15呈现了个别RNA外显子编辑器活性的RT-qPCR分析示意图。从用RNA外显子编辑器处理的细胞或组织中提取RNA后, RNA进行逆转录并进行qPCR或ddPCR。在示意图中,使用紫色正向引物(标记为1)和绿色反向引物(标记为2)测量RNA外显子编辑器RNA拷贝数。可以使用蓝色正向(标记为3)和反向(标记为4)引物对专门定量天然ABCA4 RNA。由于ABCA4 RNA外显子编辑器的末端外显子22中发生同义密码子变化,所以蓝色正向引物(标记为3)与该位置的互补性较低,并被紫色正向引物(标记为1)替换,以便定量外显子编辑的ONT RNA产物。紫色正向(标记为1)和反向(标记为5)引物对用于一起定量RNA外显子编辑器、ONT和OFTRNA分子。

[0054] 图16描绘了示例性ABCA4外显子编辑器的筛选和鉴定。应用基于NGS文库的方法(顶部)和基于单个外显子编辑器生物发光的方法(底部)鉴定出BD117为ABCA4基因内含子22中指导外显子编辑器的示例性位置。

[0055] 图17A和17B显示了40聚体接头的性能明显优于LinkD接头。(A)呈现了一个直方图,描绘每个所示构建体或未转染的阴性对照的ABCA4 RNA替换百分比。与含有前体接头(LinkD)的外显子编辑器相比,包含40聚体接头的外显子编辑器表现出ABCA4 RNA替换百分比显著增加且具有统计学意义。(B)描绘了蛋白质印迹,显示40聚体接头相对于LinkD接头可促进更高的反式剪接活性,这反映在靶(ONT)反式剪接RNA翻译而来的蛋白质水平显著更高。在使用经过工程化以过表达突变型ABCA4 mRNA的HEK293T 17+06细胞进行的转染实验中,与含有LinkD接头的外显子编辑器相比,含有40聚体接头的外显子编辑器具有显著更高的反式剪接ABCA4蛋白(ONT)水平和显著更低的非剪接蛋白(NSP)水平。利用对ABCA4的C端具有特异性的抗ABCA4抗体(上图)、对并入反式剪接蛋白和NSP的N端中的V5表位标签具有特异性的抗V5表位标签抗体(中图)、以及作为蛋白质上样对照的对钠/钾ATP酶( $\text{Na}^+/\text{K}^+$ ATP酶)具有特异性的抗体对蛋白质印迹进行探测。

[0056] 图18漫画示出了全长蛋白质和NSP。C端Ab仅免疫沉淀全长ABCA4蛋白。

[0057] 图19A和19B示出了在示例性ABCA4 5' RTM的背景下每个所示RTM元件单独和组合的活性。(A)呈现了一个直方图,描绘每个所示构建体的ABCA4 RNA替换百分比。每个构建体的替换百分比范围约为20-30%。有趣的是,相对于单独的40聚体,40聚体+3x重复U1结合位点( $3 \times \text{UBS}$ )的组合显示出ABCA4 RNA替换百分比增加的趋势。(B)描绘了蛋白质印迹,显示所示构建体40聚体、40聚体+ $3 \times \text{UBS}$ 、40聚体+富AU元件(ARE)以及40聚体+ $3 \times \text{UBS}$ +ARE表达反式剪接的ABCA4蛋白(ONT),并且这些RTM元件组合在一起表现出相对于单独的40聚体来说ONT水平增加的趋势。蛋白质印迹还显示,相对于单独的40聚体,40聚体+ $3 \times \text{UBS}$ 、40聚体+ARE和40聚体+ $3 \times \text{UBS}$ +ARE的每个组合都对非剪接蛋白(NSP)的减少表现出组合效应。相对于单独的40聚体,40聚体+ $3 \times \text{UBS}$ +ARE对NSP减少的组合效应尤为明显。在最小NSP背景下(含有终止密码子,紧跟外显子编辑器中切片供体序列之后),两种独立降低NSP水平的策略的组合——包括 $3 \times \text{UBS}$ 和富AU元件——对降低NSP水平显示出协同效应。与仅40聚体相比,这种组合将NSP水平降低了约70%,降至无终止的外显子编辑器中看到的水平。示例性构建体包含SEQ ID NO:84(40聚体泳道);SEQ ID NO:85(40聚体+3XUBS泳道);SEQ ID NO:86(40聚体

+ARE);SEQ ID NO:87(40聚体+3XUBS+ARE)。

[0058] 图20A和20B显示,MALAT1终止子(三螺旋+mascRNA)显著改善示例性ABCA45'外显子编辑器中的ONT反式剪接,并且抗mut1MALAT1终止子在ONT蛋白和NSP表达方面表现出与MALAT1终止子相似的活性。(A)呈现了一个直方图,描绘在示例性ABCA45'外显子编辑器的背景下,与前体终止子相比,MALAT1终止子赋予的相对反式剪接。相对于前体终止子,MALAT1终止子使反式剪接增加了23倍。(B)描绘了蛋白质印迹,显示对于三种不同的ABCA45'外显子编辑器,MALAT1和抗mut1 MALAT1在反式剪接的ABCA4蛋白(ONT)和非剪接蛋白(NSP)的表达水平方面表现出相似的活性。利用对ABCA4的N端具有特异性的抗ABCA4抗体(上图)、对ABCA4的C端具有特异性的抗ABCA4抗体(中图)、以及作为蛋白质上样对照的对钠/钾ATP酶( $\text{Na}^+/\text{K}^+$ ATP酶)具有特异性的抗体对蛋白质印迹进行探测。

[0059] 图21A和21B显示了转导到被工程化成过表达突变型ABCA4mRNA的HEK293T 17+06细胞中的示例性AAV2-ABCA4 RNA外显子编辑器构建体的剂量反应,其中ABCA4 RNA外显子编辑器描绘于图13中(从5'UTR跨越到抗mut1 MALAT1终止子的末端;SEQ ID NO:78)并且包含40聚体接头+3XUBS+ARE的盒(SEQ ID NO:73)。将AAV2-ABCA4外显子编辑器以生物重复的形式转导到HEK293T17+06细胞中,一系列感染复数(MOI)如下:0、1k、10k(1e4)、100k(1e5)、1M(1e6),然后(A)进行RNA提取,用于ddRT-PCR RNA替换%,或(B)进行膜蛋白提取和利用ABCA4特异性抗体探测的蛋白质印迹法。利用对ABCA4的N端具有特异性的抗ABCA4抗体(上图)进行探测的蛋白质印迹表明,在MOI 1e4时可检测到ONT蛋白和NSP,并增加直到MOI1e6。利用对ABCA4的N端具有特异性的抗ABCA4抗体(上图;20秒暴露)和作为蛋白质上样对照的对钠/钾ATP酶具有特异性的抗体( $\text{Na}^+/\text{K}^+$ ATP酶;5秒暴露)对蛋白质印迹进行探测。这些结果证明了在ABCA4 RNA%和蛋白质替换水平方面增加MOI的预期剂量反应效应。

[0060] 图22显示了体内ABCA4 RNA外显子编辑。通过RT-qPCR测量AGM视网膜中的外显子编辑。通过RT-qPCR分析处理区域的神经视网膜穿孔。(A)RNA替换%显示为总ABCA4(天然的和编辑的)的百分比。(B)每10ng RNA的外显子编辑器拷贝数相对于RNF20归一化。该体内研究中使用的示例性AAV8 RNA外显子编辑器构建体(AAV8-REEC)包括包含SEQ ID NO:77的RNA外显子编辑器。显示1个月(4周)和3个月(12周)时间点的结果。

[0061] 图23显示了编码ABCA4-01(SEQ ID NO:90)的AAV载体质粒(SEQ ID NO:82)的示意图。

[0062] 图24描绘了从外显子编辑的ABCA4 RNA翻译的蛋白质的功能活性。这些结果提供了证实通过ABCA4 RNA外显子编辑拯救的ABCA4蛋白的功能活性的证据。未处理(阴性对照)或用包含SEQ ID NO:78的外显子编辑器的V5表位标记型式(V5标记的REEC)转染后,通过从ABCA4 KO细胞中免疫沉淀富集的ABCA4蛋白的ATP酶活性,其中显示了V5表位标记型式(SEQ ID NO:81)的活性。显示了没有ATR(基线)或添加40 $\mu\text{M}$  ATR(+ATR)的ABCA4 ATP酶活性测量值。数据是进行技术双重或三重测量的单一生物学重复。

[0063] 图25描绘了OFT-seq捕获的RNA种类。GSP靶向外显子编辑器的外显子22(Ex22)中的密码子优化(CO)序列。红色虚线区域由Illumina测序平台测序。紫色区域示出了外显子编辑器提供的ABCA4CDS。出于说明的目的,提供了三个假定的OFT接合(A、B、C)和两个假定的自剪接外显子编辑器序列。

[0064] 图26呈现了描绘在HEK17+06细胞中针对包含所示元件的示例性ABCA4构建体所鉴

定的隐蔽剪接位点以及其使用频率的结果。鉴定相似的一组隐蔽剪接位点并在体内在NHP视网膜中表征。通过实验确定包含前体序列的ABCA4构建体在已鉴定的隐蔽剪接位点中引入同义突变之前 (SEQ ID NO:46) (上图) 以及在已鉴定的隐蔽剪接位点中引入同义突变之后 (下图) 的隐蔽剪接位点频率, 所述引入产生剪接位点抗性序列 (SEQ ID NO:79), 这由剪接频率的显著降低所证明。

[0065] 图27A和27B示出了在体外检查的靶mRNA前体的示例性内含子 (内含子A) 的5'RTM背景下每个所示RTM元件单独和组合的活性。(A) 呈现了一个直方图, 描绘每个所示构建体的靶RNA替换百分比。有趣的是, 相对于单独的40聚体, 40聚体+3×UBS的组合、40聚体+ARE的组合或40聚体+3×UBS+ARE的组合显示出靶RNA替换百分比降低的趋势。(B) 描绘了蛋白质印迹, 显示所示构建体40聚体、40聚体+3×UBS、40聚体+ARE以及40聚体+3×UBS+ARE表达反式剪接的在靶蛋白 (ONT), 并且相对于单独的40聚体, 每个组合都对NSP的减少表现出组合效应。更具体地说, 与单独的40聚体相比, 40聚体+3×UBS的组合使NSP水平降低了约75%; 40聚体+ARE的组合使NSP水平降低了约40%; 并且40聚体+3×UBS+ARE的组合使NSP水平降低了约88%。

[0066] 图28A和28B示出了在体外检查的靶mRNA前体的示例性内含子 (内含子B) 的5'RTM背景下每个所示RTM元件单独和组合的活性。(A) 呈现了一个直方图, 描绘每个所示构建体的靶RNA替换百分比。有趣的是, 相对于单独的40聚体, 40聚体+3×UBS的组合或40聚体+ARE的组合显示出靶RNA替换百分比增加的趋势。(B) 描绘了蛋白质印迹, 显示所示构建体40聚体、40聚体+3×UBS、40聚体+ARE以及40聚体+3×UBS+ARE表达反式剪接的靶蛋白 (ONT), 并且相对于单独的40聚体, 40聚体+3×UBS、40聚体+ARE和40聚体+3×UBS+ARE每个组合都对NSP的减少表现出组合效应。相对于单独的40聚体, 40聚体+3×UBS+ARE对NSP减少的组合效应尤为明显。更具体地说, 与单独的40聚体相比, 40聚体+3×UBS的组合使NSP水平降低了约38%; 40聚体+ARE的组合使NSP水平降低了约33%; 并且40聚体+3×UBS+ARE的组合使NSP水平降低了约66%。

[0067] 图29描绘了在体外检查的靶mRNA前体的示例性内含子 (内含子C) 的5'RTM背景下每个所示RTM元件单独和组合的活性。简单地说, 将HEK293T细胞用含有以下的RTM转染: 非结合结构域 (NBD)+40聚体; BD34 (特异性结合内含子C)+40聚体; BD34+40聚体+3×UBS; 或无RTM (转染对照)。转染后48小时收获细胞并通过RT-qPCR测定反式剪接效率。BD34+40聚体的替换百分比约为1%, 而对于包含BD34+40聚体+3×UBS的RTM, 替换百分比增加到约3%。因此, 在所示的示例性RTM靶向内含子C中, 3×UBS元件的替换百分比增加了约3倍。

[0068] 图30A-30C呈现了证明在使用示例性AAV8 RNA外显子编辑器构建体处理六 (6) 个月后, 在NHP中在体内实现了强大的ABCA4 RNA和蛋白质替换的结果。(A) 病毒基因组拷贝数 (VGC) (x轴) 和外显子编辑器RNA拷贝数 (y轴); (B) ABCA4 RNA替换%; 以及 (C) 用ABCA4-01 (3.5E11vg/眼) 处理6个月后, n=3只食蟹猴动物中的蛋白质替换情况。在A和B中, 使用基于液滴数字聚合酶链反应 (ddPCR) 的测定法评估DNA和RNA水平。在C中, 使用合格的IA MS测定法对总蛋白质百分比 (编辑的ABCA4相对于编辑的和NHP天然蛋白质的总和) 进行定量。该体内研究中使用的示例性AAV8 RNA外显子编辑器构建体包括包含SEQ ID NO:90的RNA外显子编辑器。(B) 和 (C) 中显示了四 (4) 个NHP的结果, 并指定为1-4:1 (用媒介物处理); 2-4 (用3.5E11vg/眼处理)。

[0069] 图31呈现了证明示例性AAV8RNA外显子编辑器构建体在离体人视网膜中编辑ABCA4 RNA的结果。该离体研究中使用的示例性AAV8 RNA外显子编辑器构建体包括包含SEQ ID NO:90的RNA外显子编辑器。该图描绘了用媒介物或包含SEQ ID NO:90的示例性AAV8 RNA外显子编辑器构建体处理21天后来自3个独立供体的人视网膜外植体中的ABCA4 RNA替换百分比。

### 具体实施方式

[0070] 本文描述的组合物和方法涉及用于治疗由ABCA4基因突变引起的疾病或病症的反式剪接分子(例如,mRNA前体反式剪接分子)。本文描述的组合物和方法采用mRNA前体反式剪接分子进行基因疗法(例如体内基因疗法,例如通过腺相关病毒递送)以治疗由ABCA4中的至少一个突变引起的疾病,例如斯特格氏病或视锥-视杆细胞营养不良。

[0071] ABCA4大小约为6.8kb,含有50个外显子,并且编码250kDa ATP结合盒转运蛋白,该转运蛋白在感光细胞和RPE细胞中表达。ABCA4是一种跨膜蛋白,其特异性地定位于视杆和视锥感光细胞外节盘的边缘和切口处。在视网膜中,ABCA4将N-亚视黄基-磷脂酰乙醇胺从管腔转运到感光细胞膜的细胞质小叶,该过程与化学异构化相结合,有助于去除潜在的毒性类视黄醇化合物。相反,ABCA4功能受损会导致双类视黄醇(如二视黄醛结合物A2E)在视网膜色素上皮(RPE)中积累。这种碎片也称为脂褐素,认为其会扰乱RPE代谢,并最终导致细胞死亡。当前的基因型-表型模型将残留的ABCA4活性程度与视网膜疾病表型的严重程度相关联。由于ABCA4相关的视网膜病变通常以常染色体隐性方式遗传,因此光感受器中ABCA4水平为50%的个体预计具有正常的视网膜功能。Abca4敲除小鼠的数据表明,10-25%范围的ABCA4水平纠正足以实现相关的治疗作用,包括减少视网膜中毒性脂质的积累。

[0072] 值得注意的是,本发明人已证明在携带模拟疾病的过早终止密码子突变的体外工程化人ABCA4蛋白敲除(KO)细胞系中完全拯救ABCA4蛋白水平,并在NHP视网膜和人视网膜外植体中实现了体内治疗相关水平的RNA外显子编辑。更具体地说,本发明人已观察到引起>20%ABCA4 RNA外显子编辑效率(RNA替换%)的AAV-ABCA4外显子编辑器处理足以在ABCA4蛋白KO细胞系中实现体外ABCA4蛋白水平100%的拯救。参见例如图3C。亲本ABCA4敲入(KI)细胞系含有敲入内源ABCA4基因座上游的强CAG启动子,其用作比较ABCA4蛋白拯救效率的参考。如本文所述,通过将模拟疾病的过早终止密码子突变引入ABCA4基因中,自亲本ABCA4 KI细胞系获得体外工程化人ABCA4蛋白KO细胞系。因此,体外工程化人ABCA4蛋白敲除细胞系可用作预测本文描述的ABCA4 RNA外显子编辑器(核酸反式剪接分子)的治疗效果的体外模型系统。

[0073] 此外,本发明人已显示,在NHP眼中,单次视网膜下施用示例性的基于AAV8的ABCA4外显子编辑分子可在3个月(12周)的时间点引起体内高达37%的RNA替换。参见例如图22。如本文所用,37%替换意味着从处理的细胞或组织中定量的总ABCA4 mRNA的37%以外显子编辑形式存在。本研究使用的示例性的基于AAV8的ABCA4外显子编辑分子包含SEQ ID NO:77,其共享SEQ ID NO:90中存在的相同调节元件和双终止密码子。本研究还证明了RNA反式剪接引起全长人/NHPABCA4蛋白表达的翻译,并且表达水平趋势与RNA替换%一致。另外的NHP研究使用了多种其它的基于AAV8的ABCA4外显子编辑分子,在处理后的不同时间点进行检查,证明了RNA反式剪接引起持久的RNA外显子编辑和人/NHPABCA4蛋白表达。参见例

如图9和10B。因此,本发明人通过剪接体介导的反式剪接实现了治疗相关的RNA外显子编辑,并因此在NHP中实现了由单个AAV载体介导的全长ABCA4蛋白表达。这些研究表明,与ABCA4-01 (SEQ ID NO:90) 共有的选定调节元件和40聚体接头和BD赋予NHP中长达3个月时间点的持续表达和外显子编辑活性。参见例如图22A和22B。

[0074] 本文呈现的另外证据表明,作为由包含SEQ ID NO:82的AAV载体质粒编码的一种示例性的AAV8包装的RNA外显子编辑器,ABCA4-01 (SEQ ID NO:90) 可以介导内源RNA前体的反式剪接,其程度足以至少部分恢复体外和体内细胞中的ABCA4生物活性。ABCA4-01 (SEQ ID NO:90) 描绘于图13中,并跨越RNA外显子编辑器从5' ITR到3' ITR的长度。在一些实施方案中,SEQ ID NO:78[SEQ ID NO:90中从5' UTR跨越到MALAT1-MascRNA (抗mut1) 末端的部分;参见图13]可用作RNA外显子编辑器。在一些实施方案中,图13中描绘的从外显子1-22CDS跨越到包含抗mut1的MALAT1-MascRNA的末端的SEQ ID NO:78的一部分(例如,SEQ ID NO:69) 可以与ABCA4 RNA外显子编辑器中的不同5' UTR组合,并合理预期成功实现ABCA4生物活性的可比恢复。

[0075] 除上述内容外,图30A-30C还呈现了证明在用示例性AAV8 RNA外显子编辑器构建体(ABCA4-01;SEQ ID NO:90) 处理六(6) 个月后,在NHP中在体内实现了强大的ABCA4 RNA和蛋白质替换的结果。简单地说,本发明人用单次视网膜下剂量的ABCA4-01 (3.5E11 vg/眼) 处理野生型食蟹猴。然后评估AAV载体基因组DNA拷贝(VGC) 的生物分布、外显子编辑器RNA表达以及6个月时间点编辑的ABCA4 RNA和蛋白质的百分比。外显子编辑器RNA表达与VGC有很好的相关性(图30A),并产生强大的RNA替换%(将天然ABCA4RNA拷贝转化为编辑的拷贝),该RNA替换%计算为神经视网膜样品中存在的总(编辑+天然)ABCA4 RNA群体中编辑的ABCA4 RNA的百分比(图30B)。如图30B所示,对于该剂量的ABCA4-01, RNA替换百分比范围约为40%-60%。这使所得的人-NHP嵌合ABCA4蛋白达到治疗水平,如通过合格的IA-MS测定法所测量(图30C)。更具体地说,发现由这些动物中的ABCA4 RNA编辑产生的人NHP嵌合ABCA4蛋白占测试样品中总(人-NHP+NHP)ABCA4蛋白的20-40%。此外,本发明人还证明,用通过单次视网膜下注射递送的较高剂量的ABCA4-01处理的野生型食蟹猴中RNA替换高达66%,并且人/NHP嵌合ABCA4蛋白高达45%。这些结果表明,ABCA4-01在NHP中实现了治疗相关水平的编辑ABCA4蛋白表达,这些水平超过了先前在Abca4 KO小鼠模型中显示赋予治疗益处的拯救ABCA4蛋白的水平。

[0076] 本发明人还通过用ABCA4-01(一种包含SEQ ID NO:90的示例性AAV8 RNA外显子编辑器构建体) 处理人供体视网膜外植体来评估人光感受器中的ABCA4外显子编辑。如本文所详细描述,本发明人用ABCA4-01 (3.9E11vg/视网膜穿孔) 处理来自多个供体的外植体,孵育细胞21天,然后提取RNA。对提取的RNA的分析表明大约20-30%的内源外显子1-22被ABCA4-01编码的相同对应的外显子替换的RNA替换(图31)。

[0077] 本发明人还证明,在用编码ABCA4-01的RNA外显子编辑器的N端V5表位标记型式(SEQ ID NO:81) 的过表达质粒转染的ABCA4蛋白KO细胞中拯救的ABCA4蛋白表现出ATP酶活性。如本文所述,在ABCA4蛋白免疫沉淀之后,在存在或不存在40μM全反式视黄醛(ATR) 的情况下评估由ABCA4-01的RNA外显子编辑器的V5表位标记型式(SEQ ID NO:81) 介导的反式剪接后拯救的ABCA4蛋白的活性。结果显示,在ABCA4蛋白KO细胞中拯救的ABCA4蛋白表现出ATP酶活性,这反映在加入40μM全反式视黄醛后ATP酶活性基础水平的增加。参见图24。这些

结果证明,由ABCA4-01 (SEQ ID NO:90) 处理产生的外显子编辑的ABCA4蛋白表现出从光感受器中清除ATR和11-顺式视黄醛所需的功能性ATP酶活性。这些结果描绘了在表达示例性ABCA4 RNA外显子编辑器后ABCA4生物活性的恢复。

[0078] 因此,本文中呈现的结果证明了本文中描述的ABCA4 RNA外显子编辑器(例如,包含SEQ ID NO:90、SEQ ID NO:78或SEQ ID NO:69中任一个的RNA外显子编辑器)以及包含其的组合物用于治疗ABCA4相关的视网膜病变和制备用于治疗ABCA4相关的视网膜病变的药物,以及在治疗患有ABCA4相关的视网膜病变的受试者的方法中的治疗潜力。可使用本文所述的ABCA4 RNA外显子编辑器(例如,包含SEQ ID NO:90、SEQ ID NO:78或SEQ ID NO:69中任一个的RNA外显子编辑器)和包含其的组合物治疗的ABCA4相关的视网膜病变包括斯特格氏病-1 (STGD1;眼底黄色斑点症、视网膜色素变性-19 (RP19);视锥-视杆细胞营养不良-3 (CORD3);以及年龄相关性黄斑变性-2 (ARMD2)。在一些实施方案中,可使用本文所述的ABCA4RNA外显子编辑器(例如,包含SEQ ID NO:90、SEQ ID NO:78或SEQ ID NO:69中任一个的RNA外显子编辑器)和包含其的组合物治疗的ABCA4相关的视网膜病变包括具有跨越ABCA4基因中外显子1-22的两个等位基因致病突变的遗传性视网膜疾病,其呈现与视锥-视杆细胞营养不良、斯特格氏黄斑营养不良或眼底黄色斑点症一致的表型。根据现有数据,大约60%的已知ABCA4突变可使用本文所述的ABCA4 RNA外显子编辑器(例如,由包含SEQ ID NO:90、SEQ ID NO:78或SEQ ID NO:69中任一个的核酸反式剪接分子编码的RNA外显子编辑器)来纠正,并且70-85%的经诊断患有ABCA4相关的视网膜病变的患者将受益于使用本文所述的ABCA4RNA外显子编辑器(例如,包含SEQ ID NO:90、SEQ ID NO:78或SEQ ID NO:69中任一个的RNA外显子编辑器)的治疗。

[0079] 除上述内容外,本领域技术人员认识到,在ABCA4基因中经常会鉴定新的突变,并且这些突变不一定归因于特定的表型。技术人员还认识到CORD3→STGD1→RP19似乎代表了基于突变严重程度的表型谱。因此,RNA外显子编辑器和编码其的构建体以及包含它们的组合物可用于有利地治疗与ABCA4基因外显子1-22中鉴定的任何和所有突变相关的疾患/疾病,同时理解,随后归属的表型可被表征为有关ABCA4视网膜病变的知识的发展。

[0080] 在一些实施方案中,受试者/患者被鉴定为在ABCA4基因的外显子1-22中具有至少一个突变。在一些实施方案中,对受试者/患者进行基因分型,以鉴定在ABCA4基因的外显子1-22中具有至少一个突变的受试者/患者。在一些实施方案中,获得已经进行基因分型以鉴定在ABCA4基因的外显子1-22中具有至少一个突变的受试者/患者的受试者/患者。在任何模型下,受试者/患者都是基于ABCA4基因的外显子1-22中具有至少一个突变而选择的,无论选择过程所基于的基因分型相对于使用本文所述的RNA外显子编辑器对这类受试者/患者的建议治疗的时间选择如何。这类受试者/患者将受益于使用本文所述的ABCA4 RNA外显子编辑器治疗ABCA4相关的视网膜病变、包含本文所述的ABCA4RNA外显子编辑器的药物治疗ABCA4相关的视网膜病变、以及需要施用本文所述的ABCA4RNA外显子编辑器或包含其的组合物治疗ABCA4相关的视网膜病变的方法。

[0081] 斯特格氏病-1 (STGD1) 是由1p22染色体上的ABCA4基因 (OMIM#601691) 中的纯合或复合杂合突变引起的。据报道,ABCA4基因中有超过900种不同的突变,包括错义、剪接、截断和移码变化,且显示与视网膜变性有关。许多这些突变会改变ABCA4蛋白中的单个氨基酸,导致感光细胞中的ABCA4活性缺乏或降低。ABCA4活性降低导致维生素A衍生物和脂褐素在

视网膜细胞中积累,对视网膜色素上皮和光感受器有毒性。斯特格氏黄斑变性是一种遗传性眼部病症,它影响视网膜,更具体地说,影响负责清晰中心视力的称为黄斑的区域,中心视力对于包括阅读、驾驶和面部识别等多种视觉任务至关重要。除了中心视力丧失之外,患有斯特格氏黄斑变性的患者还会出现夜视受损,并且一些受影响的个体还会出现色觉受损。在病程后期,周边视野也可能出现中度至广泛的限制。斯特格氏黄斑变性的体征和症状通常出现在儿童期至成年早期,患者会随着时间的推移表现出渐进性双侧视力丧失,通常会达到20/200或更差。

[0082] 眼底黄色斑点症 (FFM) 是斯特格氏病的等位基因亚型,与ABCA4基因和外周蛋白2 (PRPH2) 基因突变有关。眼底黄色斑点症是斑点状眼底病的一种形式,由因脂褐素沉积而出现许多白黄色斑点,广泛分布于眼底而得名。FFM的特征是视网膜色素上皮网络萎缩和脉络膜血管萎缩。中心视力丧失、色觉丧失、畏光、旁中心暗点和暗适应缓慢是FFM的典型特征。FFM比其它斯特格氏病亚型发病晚,并且进展过程较慢。如果视力丧失在生命的前20年内开始,患者通常会被诊断为患有斯特格氏病,而如果疾病在生命的后期出现并且进展过程较慢,患者通常会被诊断为患有FFM。

[0083] 视网膜色素变性-19 (RP19) 可能是由1p22染色体上ABCA4基因的纯合或复合杂合突变引起的。视网膜色素变性是一种遗传性眼部病症,其特征是周边视力逐渐丧失和夜视受损。许多基因与视网膜色素变性有关。视网膜色素变性-19的特征是夜盲症、周边视力丧失、渐进性视网膜变性、隧道视野、渐进性视力丧失、夜间或弱光下视力下降、晚期中心视力丧失以及视网膜色素上皮 (RPE) 斑驳。临床上,眼底表现为广泛的外层视网膜变性、骨针色素沉积、血管衰减、蜡状视盘苍白和严重的萎缩性改变以及黄斑脉络膜瘢痕化。

[0084] 视锥-视杆细胞营养不良-3 (CORD3) 是由1p22染色体上ABCA4基因的纯合或复合杂合突变引起的。ABCA4是与视锥视杆细胞营养不良的发病机理相关的四大致病基因之一,并且ABCA4突变与30%至60%的常染色体隐性视锥视杆细胞营养不良有关。CORD3是一种常染色体隐性、临床异质性视网膜病症,通常特征在于症状性快速进展性视力丧失以及最初表现为畏光和色彩辨别能力丧失,随后出现夜盲症和渐进性视野丧失。视锥细胞变性在生命早期出现,累及视网膜中心,接着几年后视杆细胞发生变性。视锥视杆细胞营养不良的诊断基于临床病史、眼底检查和视网膜电图。视锥细胞-视杆细胞营养不良通常比其它ABCA4视网膜病变出现得更早,进展也快得多,并且在发病数年后才会出现大面积黄斑萎缩。目前尚无可以阻止疾病发展或恢复视力的疗法,并且视力预后较差。管理旨在减缓变性过程,治疗并发症,并帮助患者做好应对失明所带来的社会和心理影响的挑战的准备。

[0085] 1p22染色体上的ABCA4基因变异导致了对年龄相关性黄斑变性-2 (ARMD2) 的易感性。ARMD2是一种复杂的病症,其特征是玻璃膜疣(黄色沉积物堆积)积聚在RPE内和下以及黄斑RPE的渐进性萎缩。这些变化导致感光功能丧失和视力障碍。年龄相关性黄斑变性是发达国家老年人视力丧失的主要原因。视力丧失通常在人六七十岁时变得明显,并且随着时间的推移变得更糟。年龄相关性黄斑变性主要影响中心视力,中心视力对于例如阅读、驾驶和面部识别等细节任务至关重要。已描述了两种主要类型的年龄相关性黄斑变性:干性黄斑变性和湿性黄斑变性。干性黄斑变性更为常见,占有年龄相关性黄斑变性病例的85%至90%。其特征是玻璃膜疣累积在视网膜下、黄斑萎缩以及视力逐渐缓慢丧失。这种疾病通常会影响双眼。湿性年龄相关性黄斑变性与突然发生的严重视力丧失有关,病情可能迅速

恶化。这种形式疾患的特征是由于VEGF的过度表达导致黄斑下方生长异常、脆弱的血管。这些血管会漏出血液和流体,从而损害黄斑,从而使中心视力模糊和扭曲。湿性黄斑变性可以使用多种疗法治疗。干性AMD只有一种可用的治疗方法,即C3抑制剂,仅适用于出现地图状萎缩的患者。目前,对于仅出现玻璃膜疣沉积而没有萎缩或流体渗漏的早期形式的ARMD,尚无疗法。

#### [0086] I. 定义

[0087] 如本文所用,“反式剪接”意指通过剪接体介导的机制,将含有一个或多个外显子(例如外源外显子或属于反式剪接分子的CDS的一部分的外显子)的第一RNA分子与第二RNA分子(例如mRNA前体分子,例如内源mRNA前体分子)连接,并用第一RNA分子的一部分替换第二RNA分子的一部分。图14B说明了RNA反式剪接反应的一般机制。

[0088] “核酸反式剪接分子”或“反式剪接分子”具有三个主要元件:(a)结合结构域,其通过将反式剪接分子拴系到其靶基因(例如,mRNA前体)而赋予特异性;(b)剪接结构域(例如,具有3'或5'剪接位点的剪接结构域);以及(c)CDS,其配置成反式剪接到靶核酸上,其可以替换靶核酸中的一个或多个外显子(例如,一个或多个突变的外显子)。“mRNA前体反式剪接分子”或“RTM”是指靶向mRNA前体的核酸反式剪接分子。术语“核酸反式剪接分子”和“反式剪接分子”是指(1)编码RNA的DNA,其中RNA转录本是与靶mRNA前体物理结合的效应分子;和(2)RNA转录本本身。为了清楚起见,本文使用术语“编码序列”(例如,反式剪接分子编码序列)来指定受试者编码效应物(例如,编码序列是DNA而效应物是RNA)。在一些实施方案中,反式剪接分子编码序列可以包括cDNA,例如作为功能性外显子(例如,功能性ABCA4外显子)的一部分,用于替换突变的ABCA4外显子。

[0089] 如本文所用,术语“外显子编辑器”可用于指反式剪接分子或包含其的载体(例如,包含编码作为反式剪接分子的RNA转录本的DNA的AAV载体)。

[0090] 本文中使用的“反式剪接效率”是指检测到的所需反式剪接RNA产物(即,包括由RNA反式剪接反应产生的可操作地连接到内源靶mRNA前体的反式剪接分子的功能性外显子的嵌合RNA分子)的表达水平与为了反式剪接分子而引入的DNA或RNA(或参考分子)的量的比率。在某些情况下,在使用RNA-seq从细胞或组织中分离的RNA中检测反式剪接RNA产物的表达水平。

[0091] 如本文所用,“RNA替换%”是指已成功进行反式剪接(TS)的总靶mRNA群体的部分,并通过以下公式计算:在靶(ONT)TS% =  $100 \times (\text{ONT拷贝数} / (\text{ONT拷贝数} + \text{天然拷贝数}))$ 。

[0092] 如本文所用,“相对反式剪接效率”是指测试反式剪接效率与参考反式剪接效率的比率,其中测试反式剪接效率为反式剪接分子(例如,本文所述的核酸反式剪接分子)的反式剪接效率,而参考反式剪接效率为参考分子(例如,具有与核酸反式剪接分子相同的元件,但结合结构域被替换为杂乱的结合结构域(例如,包含SEQ ID NO:1或由其组成的结合结构域,例如包含SEQ ID NO:2或由其组成的结合结构域)的参考分子)的反式剪接效率。反式剪接分子的相对反式剪接效率可以表示为在类似条件下测试的测试反式剪接RNA效率与参考反式剪接效率的比率(又称增加倍数)。

[0093] 如本文所用,术语“可操作地连接”或“操作性连接”是指元件的排列,其中所描述的组件被配置成可执行其通常的功能。当一个核酸与另一个核酸序列具有功能关系时,该核酸就与另一个核酸序列“可操作地连接”。元件不需要相邻即可可操作地连接。因此,例

如,间插序列可存在于可操作地连接地序列之间(例如,结合结构域和编码序列可由间插序列分隔,而结合结构域仍被认为“可操作地连接”到编码序列)。

[0094] 如本文所用,术语“CDS”是指编码蛋白质(例如,正在纠正突变的靶蛋白质)的一部分的核酸序列(例如,RNA序列、DNA序列或RNA与DNA的组合)。因此,CDS可包括一个或多个功能性外显子(例如,功能性外显子序列)。在一些情况下,CDS的一个或多个功能性外显子不被内含子分隔(例如,在内源mRNA前体中),而是彼此相邻(例如,作为cDNA)。在一些情况下,CDS可以包括一个或多个内含子(例如,天然内含子)或位于外显子之间或与之相邻(例如,上游或下游)的非翻译区(UTR,例如天然UTR)。

[0095] 如本文所用,“天然5'ABCA4非翻译区”或“天然5'ABCA4 UTR”是指长度大于20个核苷酸的序列,其与位于ATG起始密码子5'的天然ABCA4基因(例如,人ABCA4基因)的区域具有至少90%的序列同一性。天然5'ABCA4非翻译区的一个实例由SEQ ID NO:13的DNA序列或SEQ ID NO:12的RNA序列给出。本文还涵盖天然5'ABCA4非翻译区的变体,例如SEQ ID NO:64的DNA序列或SEQ ID NO:65的RNA序列。

[0096] 如本文所用,“5'ABCA4外显子的功能序列”是指包含ABCA4外显子1-22中的一个或多个的核酸序列,其编码ABCA4蛋白的功能(生物活性)部分。当在结合位点的3'反式剪接到内源ABCA4外显子时,5'ABCA4外显子的功能序列可表达功能性ABCA4蛋白(例如,未突变的ABCA4蛋白)。在一些情况下,5'ABCA4外显子的功能序列包括与反式剪接分子正在反式剪接的外显子相邻的外显子序列(例如,结合ABCA4内含子22并与内源ABCA4外显子23反式剪接的反式剪接分子可以包括包含外显子22、21、20、19等的5'ABCA4外显子的功能序列)。

[0097] 如本文所用,术语“功能性”在蛋白质的上下文中使用时是指生物活性蛋白质。术语“功能性”也可用于指支持正常细胞功能所必需的蛋白质活性量。对于ABCA4,术语“功能性”可用于指恢复ABCA4活性水平以支持例如感光细胞和/或视网膜上皮细胞背景下的正常细胞功能所必需的ABCA4蛋白活性量。这类水平足以降低或防止双类视黄醇化合物在感光细胞和/或视网膜上皮细胞中积累到毒性水平。更具体地说,有缺陷的(无功能)ABCA4蛋白会导致11-顺式和全反式视黄醛在光感受器中积聚,以及脂褐素在视网膜色素上皮中积累。在治疗与降低或可忽略的ABCA4活性相关的疾患或使用包含本文所述的核酸反式剪接分子的治疗剂的背景下,功能性是指恢复足以消除与ABCA4活性水平降低相关的疾患(例如ABCA4相关的视网膜营养不良)的一种或多种症状的ABCA4蛋白的量。ABCA4相关的视网膜营养不良包括例如斯特格氏黄斑营养不良(斯特格氏病)、眼底黄色斑点症和ABCA4相关的视锥-视杆细胞营养不良。在一些实施方案中,这类方法或用途引起ABCA4蛋白活性(功能性生物活性ABCA4蛋白活性)增加。在一些实施方案中,相对于ABCA4以正常野生型水平(例如表达未突变的ABCA4的细胞(例如,光感受器和/或视网膜上皮细胞)中存在的水平)存在的细胞,ABCA4蛋白活性的这种增加将ABCA4活性水平恢复到ABCA4活性的至少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%(例如,96%、97%、98%、99%或100%)。

[0098] 如本文所用,“纠正隐蔽剪接位点”、“减少隐蔽剪接位点”或“抗隐蔽剪接位点”是指核酸反式剪接分子或其部分(例如,其中的编码结构域序列)已被修饰以改变其中的个别核苷酸,从而降低在核酸反式剪接分子背景下鉴别的隐蔽剪接位点处发生的剪接频率。在

一些实施方案中,修饰不会导致由此编码的氨基酸序列发生任何变化。在一些实施方案中,修饰导致由此编码的氨基酸序列发生变化。在一些实施方案中,核酸反式剪接分子内的抗隐蔽剪接位点核酸序列是编码结构域序列(CDS)。在一些实施方案中,抗隐蔽剪接位点ABCA4 CDS包含ABCA4基因的外显子1-22、基本上由其组成或由其组成,其中已在核酸反式剪接分子的背景下鉴定隐蔽剪接位点,并且其中至少一个隐蔽剪接位点已被修饰以降低至少一个位点处的剪接频率,同时不改变由此编码的氨基酸。表3中呈现了ABCA4外显子1-22中鉴定的示例性隐蔽剪接位点以及这些位点的剪接频率。表中带有rtm\_pos标题的列中所示的核苷酸位置是指第3列或第4列中所示的最后一个核苷酸相对于ATG密码子(翻译起始位点)的位置,该密码子中的A被指定为位置1。示例性隐蔽剪接抗性ABCA4外显子1-22序列包含SEQ ID NO:56-59中的任一者、基本由其组成或由其组成。在一些实施方案中,示例性隐蔽剪接抗性ABCA4外显子1-22序列包含SEQ ID NO:56、基本上由其组成或由其组成。

[0099] 如本文所用,“剪接结构域”是指具有被剪接体识别并介导反式剪接的基序的核酸序列。剪接结构域包括剪接位点(例如单个剪接位点,即有且仅有一个剪接位点),该剪接位点可以是3'剪接位点或5'剪接位点。剪接结构域可能包括其它调节元件。例如,在一些实施方案中,剪接结构域包括剪接增强子(例如,外显子剪接增强子(ESE)和内含子剪接增强子(ISE))。在一些实施方案中,剪接结构域包含GUAAGT或GTAAGT。在一些实施方案中,剪接位点主要由GUAAGT或GTAAGT组成。在一些实施方案中,剪接位点主要由GUAAGT或GTAAGT组成。

[0100] 如本文所用,反式剪接分子的“结合结构域”是通过杂交(即与结合位点完全或部分互补)在结合位点与靶基因结合的多核苷酸序列。

[0101] 如本文所用,术语“结合位点”是指靶mRNA(例如,内源基因如ABCA4的mRNA)中与核酸反式剪接分子的结合结构域结合的内源多核苷酸序列。结合位点从结合结构域结合的最5'核苷酸延伸到结合结构域结合的最3'核苷酸。在一些实施方案中,结合位点的长度与结合结构域的长度相同。在其它实施方案中,结合位点比结合结构域长或短1-10个残基(即,结合位点或结合结构域的一些残基未杂交)。在涉及具有至少两个非重叠序列且与结合位点具有至少80%互补性的结合结构域的实施方案中,结合位点可以显著短于结合结构域。

[0102] 如本文所用,“互补性”及其语法变体是指给定序列通过氢键与参考序列配对的核苷酸碱基的百分比。在没有给定互补百分比的情况下,术语“互补(complement)”和“互补(complementary)”是指100%互补。

[0103] 如本文所用,如果给定序列(例如结合结构域序列)的每个核苷酸碱基通过氢键与参考序列(例如内源mRNA前体结合位点)配对,从而杂交形成双链序列(例如,通过沃森-克里克碱基配对(Watson-Crick base-pairing),例如每个A与T或U配对,并且每个C与G配对),则给定序列与参考序列“100%互补”,或具有“100%互补性”。例如,与结合位点反义取向的结合结构域与结合位点互补。RNA配对包括G与U配对(摆动碱基配对);因此,与结合位点进行G-U配对的RNA结合结构域可以与结合位点100%互补。因此,可以修饰与结合位点完全反向互补的结合结构域(即,结合结构域的A与结合位点的U配对),以将任意一个或多个A替换为G,而不会显著影响结合。

[0104] 如本文所用,如果给定序列(例如结合结构域序列)的X%核苷酸碱基通过氢键与参考序列(例如内源mRNA前体结合位点)配对,例如杂交形成双链序列(例如,通过沃森-克里克碱基配对,例如A与T或U配对,并且C与G配对),则给定序列与参考序列“至少X%互补”,

或具有“X%互补性”。例如,如果长度为150个碱基的结合结构域序列的150个残基中至少有135个残基通过沃森-克里克碱基配对通过氢键与长度为150个碱基的结合位点配对,剩下15个或更少的错配核苷酸,则该结合结构域序列与该结合位点至少90%互补。

[0105] 如本文所用,结合结构域与内含子之间的“结合”是指结合结构域与靶内含子之间以足以通过将反式剪接分子与靶标(例如,mRNA前体)缔合来介导反式剪接的程度形成氢键(例如,双螺旋形成或沃森-克里克配对)。在一些实施方案中,结合结构域与靶内含子之间的氢键位于互补且彼此反义取向(例如,彼此杂交)的核苷酸碱基之间。

[0106] 结合结构域的序列可以与靶序列(例如,mRNA前体)的靶序列至少50%、55%、60%、65%、70%、75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或99.5%互补。在一些实施方案中,结合结构域的序列可以与mRNA前体转录本的靶序列100%互补。

[0107] 在一些实施方案中,结合结构域的序列相对于mRNA前体转录本的靶序列可以具有4个或更少的错配。在一些实施方案中,结合结构域的序列相对于mRNA前体转录本的靶序列可以具有3个或更少的错配。在一些实施方案中,结合结构域的序列相对于mRNA前体转录本的靶序列可以具有2个或更少的错配。在一些实施方案中,结合结构域的序列相对于mRNA前体转录本的靶序列可以具有1个或更少的错配。在一些实施方案中,结合结构域的序列相对于mRNA前体转录本的靶序列可以没有错配。

[0108] 结合结构域可以特异性地与mRNA前体转录本的靶序列杂交。例如,结合结构域可以与mRNA前体转录本的靶序列具有91%、92%、93%、94%、95%、96%、97%、98%、99%、99.5%或100%的序列互补性。在一些实施方案中,杂交可以在生理杂交条件下(在体外或体内细胞中)进行。在一些实施方案中,杂交可以在高严格杂交条件下进行。

[0109] 在一些实施方案中,结合结构域与mRNA前体的靶序列“特异性杂交”或对其具有“特异性”。这类杂交通常发生在显著高于37℃、更特别地至少50℃、并且通常在60℃至约90℃之间的T<sub>m</sub>下。这类杂交更具体地对应于严格杂交条件。在给定的离子强度和pH值下,T<sub>m</sub>是50%的靶序列与互补寡核苷酸杂交的温度。

[0110] 如本文所用,“人工内含子”是指将结合结构域(直接或间接)连接到CDS的非编码核酸序列。人工内含子包括剪接结构域并且还可包括一个或多个间隔序列和/或其它调节元件。

[0111] 如本文所用,术语“突变”可用于指导致有缺陷的蛋白质产物(例如,无功能性蛋白质产物、非生物活性蛋白质、功能降低的蛋白质产物、具有异常功能的蛋白质产物和/或产生的量少于正常量或多于正常量的蛋白质产物)的任何异常核酸序列。突变包括碱基对突变(例如,单核苷酸多态性)、错义突变、移码突变、缺失、插入和剪接突变。在一些实施方案中,突变是指序列的一个或多个部分与相应的野生型核酸序列或其功能变体不同的核酸序列。在一些实施方案中,突变是指编码具有与相应的野生型蛋白质或其功能变体不同的氨基酸序列的蛋白质的核酸序列。“突变的外显子”(例如,突变的ABCA4外显子)是指含有突变的外显子或反映不同区域突变的外显子序列,例如由内含子突变引起的隐蔽外显子。

[0112] 除非另有说明,否则术语“ABCA4”(ATP结合盒亚家族A成员4)是指来自任何脊椎动物来源的任何天然ABCA4(也称为ARMD2、CORD3、ABCR、FFM、视网膜特异性磷脂转运ATP酶ABCA4、斯特格氏病蛋白、RIMABC转运蛋白、RIMProteinv、STGD1、RP19或STGD),包括哺乳动

物,例如灵长类动物(例如,人、非洲绿猴和食蟹猴)和啮齿动物(例如,小鼠和大鼠),以及其功能等效或改良的变体(例如,天然或合成变体)、突变体、突变蛋白、类似物、亚基、受体复合物、同型、剪接变体和片段。基于已知的ABCA4信号传导,可以确定功能等效和改良的变体。ABCA4涵盖全长未加工ABCA4,以及由细胞内天然加工产生的任何形式的ABCA4。示例性人ABCA4序列以美国国家生物技术信息中心(NCBI)参考序列NG\_009073.1提供。在一些情况下,ABCA4由与SEQ ID NO:11具有至少95%序列同一性(例如,与SEQ ID NO:11、其功能部分和/或其密码子优化变体具有至少96%、至少97%、至少98%、至少99%或100%序列同一性)的治疗基因编码。

[0113] 如本文所用,“变体”是指在至少一个核酸残基上与参考多核苷酸序列(例如天然存在的多核苷酸序列)不同的多核苷酸,或在至少一个氨基酸残基上与参考多肽序列(例如天然存在的多肽序列或例如本文所述的任何rAAV序列)不同的多肽(例如AAV衣壳序列)。在本上下文中,至少一个残基的差异可以包括例如一个核酸残基被另一个核酸取代、缺失或插入,或者一个氨基酸残基被另一个氨基酸取代。变体可以是如本文所定义的多核苷酸的同源物、异构体或转录本变体,其中同源物、同种型或转录本变体分别通过如本文所定义的同源性或同源性程度表征。

[0114] 在一些情况下,多核苷酸或多肽的变体包括至少一个核酸取代(例如,1-100个核酸或氨基酸取代、1-50个核酸或氨基酸取代、1-20个核酸或氨基酸取代、1-10个核酸或氨基酸取代,例如1个核酸或氨基酸取代、2个核酸或氨基酸取代、3个核酸或氨基酸取代、4个核酸或氨基酸取代、5个核酸或氨基酸取代、6个核酸或氨基酸取代、7个核酸或氨基酸取代、8个核酸或氨基酸取代、9个核酸或氨基酸取代或10个核酸或氨基酸取代)。使表达的多肽具有来自同一类别的交换氨基酸的核酸取代在本文中称为保守取代。具体来说,这些是具有脂肪族侧链、带正电或负电的侧链、侧链中具有芳香族基团的氨基酸,或者侧链可以形成氢桥,例如具有羟基官能团的侧链的氨基酸。通过保守构成,例如可以将具有极性侧链的氨基酸替换为具有相应极性侧链的另一种氨基酸,或者例如可以将特征在于疏水侧链的氨基酸替换为具有相应疏水侧链的另一种氨基酸(例如,丝氨酸(苏氨酸)被苏氨酸(丝氨酸)取代,或亮氨酸(异亮氨酸)被异亮氨酸(亮氨酸)取代)。

[0115] 在一些情况下,术语变体还涵盖插入、缺失和/或非保守取代,例如在不会导致蛋白质三维结构发生实质性改变的那些位置处的插入、缺失和/或非保守性取代。本领域技术人员可以容易地确定通过插入或缺失对三维结构进行的修改,例如使用CD光谱(圆二色性光谱)。

[0116] 术语“同源”是指两个核酸序列之间的同一性程度。序列的同源性是通过在要比较的序列长度上比较在标准条件下比对的两个序列来确定的。本文要比较的序列在两个序列的最佳比对中可能会有添加或缺失(例如空位等)。在一些实施方案中,序列同源性是通过使用例如ClustalW算法(Nucleic Acid Res.,1994,22(22):46734680)创建比对来计算的。也可以使用常见的序列分析软件,例如Vector NTI、GENETYX、BLAST或公共数据库提供的分析工具。

[0117] 如本文所用,术语“AAV”或“AAV血清型”是指数十种天然存在和可用的腺相关病毒以及人工AAV。在从人或非人灵长类动物(NHP)中分离或工程化并且已充分表征的AAV中,人AAV2是第一个被开发为基因转移载体的AAV;它已广泛用于在不同靶组织和动物模型中进

行有效的基因转移实验。

[0118] 如本文所用,涉及AAV,术语变体意指源自已知AAV序列的任何AAV序列,包括在氨基酸或核酸序列上共享至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少97%、至少99%或更大序列同一性的序列。在另一实施方案中,AAV衣壳包括可能包括与任何已描述或已知的AAV衣壳序列高达约10%的变异的变体。也就是说,AAV衣壳与本文提供的和/或本领域已知的AAV衣壳共享约90%同一性至约99.9%同一性、约95%至约99%同一性或约97%至约98%同一性。在一个实施方案中,AAV衣壳与AAV衣壳具有至少95%同一性。在确定AAV衣壳的同一性百分比时,可以对任何可变蛋白(例如vp1、vp2或vp3)进行比较。

[0119] 可以使用本领域技术人员可利用的技术从AAV中容易地分离ITR或其它AAV组件或进行工程化。这类AAV可以从学术、商业或公共来源(例如,美国典型培养物保藏中心(Manassas, Va.))分离、工程化或获得。或者,AAV序列可通过合成或其它合适方式参考已公开的序列(例如在文献中或数据库(例如GenBank、PubMed等)中可获得的序列)进行工程化。AAV病毒可以通过传统的分子生物学技术进行工程化,从而可以优化这些颗粒,以实现核酸序列的细胞特异性递送、最小化免疫原性、调整稳定性和颗粒寿命、有效降解、准确递送到细胞核等。

[0120] 如本文所用,术语“受试者”、“个体”或“患者”包括需要这些治疗或预防方法的任何哺乳动物,包括灵长类动物,例如人。需要这类治疗或预防的其它哺乳动物包括非人灵长类动物(NHP;例如食蟹猴和非洲绿猴)、狗、猫或其它驯养动物、马、牲畜、实验动物(例如,小鼠、大鼠或兔子)等。个体可以是雄性或雌性。在一个实施方案中,个体患有由ABCA4基因突变引起的疾病或病症(例如,ABCA4相关的视网膜营养不良)。在另一实施方案中,个体有出现由ABCA4基因突变引起的疾病或病症的风险。在另一实施方案中,个体已表现出由ABCA4基因突变引起的疾病或病症,例如ABCA4相关的视网膜营养不良的临床体征。个体可以是治疗或预防性疗法可能有益的任何年龄。例如,在一些实施方案中,个体是0-5岁、5-10岁、10-20岁、20-30岁、30-50岁、50-70岁或超过70岁。在另一实施方案中,个体的年龄为12个月或更大、18个月或更大、2岁或更大、3岁或更大、4岁或更大、5岁或更大、6岁或更大、7岁或更大、8岁或更大、9岁或更大、或者10岁或更大。在另一实施方案中,个体具有活视网膜细胞(例如,足以维持部分视觉功能或完全视觉功能的活视网膜细胞)。

[0121] 如本文所用,术语“与突变相关的病症”或“与病症相关的突变”是指疾病与突变之间的相关性。在一些实施方案中,已知或怀疑与突变相关的病症完全或部分、直接或间接地由该突变引起。例如,具有突变的个体可能有出现该病症的风险,并且该风险可能还取决于其它因素,例如其它(例如,独立的)突变(例如,在相同或不同的基因中)或环境因素。

[0122] 如本文所用,术语“治疗”或其语法衍生词定义为减少疾病的进展、减轻疾病症状的严重程度、延缓疾病症状的进展、消除疾病症状或延迟疾病的发作。在一些实施方案中,术语“治疗”用于指治疗剂(例如本文描述的RNA外显子编辑器)的持续或持久效果。例如,图22和31中呈现了证明本文所述的RNA外显子编辑器的持久治疗效果的证据。在一些实施方案中,本文提供的治疗效果持续至少1个、2个、3个、4个、5个、6个月或更长时间。在一些实施方案中,核酸反式剪接分子在一生中仅施用一次。

[0123] 如本文所用,术语“预防”病症或其语法衍生词被定义为降低疾病发作的风险,例

如,对有出现与突变相关的病症风险的个体进行预防性治疗。根据本领域已知或本文描述的任何合适方法,通过鉴定与病症相关的突变,可将个体表征为“有出现病症的风险”。在一些实施方案中,有出现病症风险的个体具有一个或多个与该病症相关的ABCA4突变。此外或替代地,如果个体有家族病症史,则该个体可以被描述为“有出现该病症的风险”。

[0124] 如本文所用,术语“ABCA4相关的视网膜病变”包含与ABCA4突变相关的疾病/疾患,包含例如斯特格氏病-1 (STGD1);眼底黄色斑点症、视网膜色素变性-19 (RP19);视锥-视杆细胞营养不良-3 (CORD3);以及年龄相关性黄斑变性-2 (ARMD2)。

[0125] 可以通过直接向个体施用反式剪接分子或RNA外显子编辑器(例如,在载体,例如AAV载体或AAV颗粒内)来治疗或预防个体的病症。或者,可以向个体施用含有反式剪接分子的宿主细胞。

[0126] 如本文所述的方法中所用,术语“施用”或其语法衍生词是指将反式剪接分子或RNA外显子编辑器(例如,在载体,例如AAV载体或AAV颗粒内)或其组合物,或离体处理的细胞,递送到有需要的个体,例如具有ABCA4突变或缺陷的个体。例如,在一个以眼部细胞(例如,光感受器)为目标的实施方案中,该方法涉及通过视网膜下注射向个体递送反式剪接分子或RNA外显子编辑器(例如,在载体,例如AAV载体或AAV颗粒内)或其组合物。在另一实施方案中,可进行玻璃体内注射或经眼睑静脉注射来施用。在另一实施方案中,组合物是全身施用的(例如静脉内)。鉴于本公开内容,本领域技术人员还可以选择其它施用方法。

[0127] 如本文所用,“调节ABCA4表达”是指降低内源突变(无功能)ABCA4的表达和/或增加反式剪接的ABCA4的表达。调节ABCA4的表达可用于指例如降低内源(例如突变的)ABCA4的表达和/或增加反式剪接的ABCA4的表达(例如,相对于内源突变的转录本或蛋白质产物,具有反式剪接分子介导的纠正突变位点的ABCA4转录本或蛋白质产物)。通过反式剪接替换构成突变位点的内源ABCA4外显子后,就会表达功能性的ABCA4蛋白。

[0128] 如本文所用,“密码子优化”是指修饰核酸序列以改变个别核酸,但不会导致编码的氨基酸发生任何改变。以这种方式修饰的序列在本文中称为“密码子优化的”。可以对本说明书中描述的任何序列进行该过程以增强表达或稳定性。密码子优化可以以例如美国专利No.7,561,972、7,561,973和7,888,112中描述的方式进行,每个专利都以引用的方式整体并入本文中。可以根据已知方法将翻译起始位点周围的序列转化为共有科扎克序列(consensus Kozak sequence)。参见例如Kozak等人,1987.Nucleic Acids Res.15(20):8125-8148,其以引用的方式整体并入本文中。

[0129] 术语“药学上可接受的”是指对于哺乳动物(例如人)施用是安全的。在一些实施方案中,药学上可接受的组合物由联邦监管机构或州政府批准或者在美国药典或其它通常公认的药典中列出用于动物并且更尤其用于人。

[0130] 术语“载剂”是指与治疗分子(例如,本发明的反式剪接分子或包括反式剪接分子的载体或细胞)一起施用的稀释剂、佐剂、赋形剂或媒介物。合适的药物载剂的实例描述于“Remington's Pharmaceutical Sciences,”Mack Publishing Co.,Easton,PA.,第18版。

[0131] 术语“一”和“一个”意指“一个或多个”。例如,“一个基因”被理解为代表一个或多个这样的基因。因此,术语“一”和“一个”、“一个或多个(或一个)”和“至少一个(或一个)”在本文中可互换使用。

[0132] 除非另有说明,否则单链核酸(例如,mRNA前体转录本、寡核苷酸等)序列的左端为

5'端,并且单链或双链核酸序列的左侧方向称为5'方向。类似地,核酸序列(单链或双链)的右端或方向是3'端或方向。一般来说,核酸中相对于参考点在5'的区域或序列称为“上游”,并且核酸中相对于参考点在3'的区域或序列称为“下游”。一般来说,mRNA的5'方向或端是启动或起始密码子所在的位置,而3'端或方向是终止密码子所在的位置。

[0133] 除非另有说明,否则如本文所用,术语“约”是指相对于参考值在 $\pm 10\%$ 变化范围内的值。

#### [0134] II.反式剪接分子

[0135] 外显子编辑器的设计和发现采用了一种组合方法,其包括基于下一代测序(NGS)的外显子编辑器文库筛选,然后进行计算分析,以及转染个别外显子编辑器,然后进行RT-qPCR分析以评估RNA替换效率和进行蛋白质印迹分析以评估蛋白质拯救。前一种方法利用合理的设计和从过去的筛选中获得的累积知识以及多并行通量的功能,而后一种方法用于测试新的假设并验证基于NGS的方法得出的结论。

[0136] 本文提供了核酸反式剪接分子,其可用于通过用功能性ABCA4外显子(例如,相对于结合位点在5'的ABCA4外显子,例如ABCA4的外显子1-22)替换至少一个突变的ABCA4外显子来纠正ABCA4中的突变。在一些实施方案中,核酸反式剪接分子是RNA前体反式剪接分子(RTM)。反式剪接分子的设计允许用核酸序列,例如具有功能性(例如,正常)序列而无突变的外显子替换mRNA前体外显子的缺陷或突变部分。功能序列可以是野生型、天然存在的序列或进行一些其它修饰(例如密码子优化)的纠正序列。

[0137] 反式剪接分子包含结合结构域、剪接结构域和CDS。在一些实施方案中,核酸反式剪接分子具有5'调节结构域,所述5'调节结构域具有天然5'ABCA4非翻译区(例如,与SEQ ID NOS:12-15中的任一者具有至少80%序列同一性的序列)。在一些实施方案中,核酸反式剪接分子具有GUAAGT或GTAAGT的剪接位点。在一些实施方案中,核酸反式剪接分子具有长度超过25个核苷酸的接头结构域。在一些实施方案中,核酸反式剪接分子具有接头结构域,所述接头结构域包含SEQ ID NO:27或与SEQ ID NO:27具有至少91%同一性、至少92%同一性、至少93%同一性、至少94%同一性、至少95%同一性、至少96%同一性、至少97%同一性、至少98%同一性或至少99%同一性的序列、基本上由其组成或由其组成。

[0138] 在一些实施方案中,本文所述的反式剪接分子包括在5'到3'方向上操作性连接的以下各者:5'非翻译区、编码结构域序列(例如CDS,例如编码ABCA4外显子的功能序列,例如相对于结合位点在5'的ABCA4外显子的功能序列的序列)、剪接结构域(例如剪接位点)、接头结构域、结合结构域、3'下游区域和终止子结构域。

[0139] 在一些实施方案中,本文所述的核酸反式剪接分子被配置成通过结合于靶内含子22并介导具有5'ABCA4外显子功能序列的CDS与靶内含子3'的内源性ABCA4外显子的反式剪接来纠正位于ABCA4基因5'区域(例如,内含子225'的区域)中的至少一个突变(例如两个不同的突变,其中每个突变存在于受试者的ABCA4基因的不同等位基因上)。从而这种反式剪接通过替换有缺陷的外显子并去除靶mRNA前体的有缺陷部分来修复个体靶细胞中的有缺陷的ABCA4基因,产生能够在细胞中转录功能性ABCA4蛋白的功能性ABCA4 mRNA。

[0140] 在一些实施方案中,反式剪接分子的长度可达5,000个核苷酸碱基,例如,长度可达4,900个核苷酸碱基,长度可达4,800个核苷酸碱基,或长度可达4,700个核苷酸碱基(例如,长度为3,000至4,000个核苷酸碱基,长度为3,100至3,800个核苷酸碱基,长度为3,200

至3,700个核苷酸碱基,或长度为3,300至3,500个核苷酸碱基,例如,长度为3,000至3,100个核苷酸碱基,长度为3,100至3,200个核苷酸碱基,长度为3,200至3,300个核苷酸碱基,长度为3,300至3,400个核苷酸碱基,长度为3,400至3,500个核苷酸碱基,长度为3,500至3,600个核苷酸碱基,长度为3,600至3,700个核苷酸碱基,长度为3,700至3,800个核苷酸碱基,长度为3,800至3,900个核苷酸碱基,长度为3,900至4,000个核苷酸碱基,长度为4,000至4,100个核苷酸碱基,长度为4,100至4,200个核苷酸碱基,长度为4,200至4,300个核苷酸碱基,长度为4,300至4,400个核苷酸碱基,长度为4,400至4,500个核苷酸碱基,长度为4,500至4,600个核苷酸碱基长度,长度为4,600至4,700个核苷酸碱基,长度为4,700至4,800个核苷酸碱基,长度为4,800至4,900个核苷酸碱基或长度为4,900至5,000个核苷酸碱基)。

[0141] 由于ABCA4基因尺寸大且受到AAV递送的尺寸限制,被配置用于在AAV载体内包装的单个反式剪接分子(例如,作为反式剪接分子编码序列)可能无法跨越可能与病症相关的ABCA4基因中的所有突变,因此可能无法纠正整个ABCA4基因长度上的突变。因此,本文所述的反式剪接分子可作为下述方法的一部分进行调整,以纠正跨越ABCA4基因整个长度的多个突变。

[0142] ABCA4

[0143] 本文所述的反式剪接分子所靶向的ABCA4基因可以含有一个或多个与ABCA4相关的视网膜营养不良疾病相关(例如,导致疾病或与疾病相关),例如斯特格氏病或视锥-视杆细胞营养不良的突变。示例性人ABCA4序列由美国国家生物技术信息中心(NCBI)参考序列NG\_009073提供。除了已公开的序列外,还包括所有后来获得的纠正或在人或其它哺乳动物群体中存在的天然存在的保守和非致病变异序列。还包括另外的保守核苷酸替换或导致密码子优化的核苷酸替换。数据库登录号提供的序列也可用于搜索相同或另一哺乳动物生物体中的同源序列。

[0144] 预计ABCA4核酸序列和所得到的表达的蛋白质可以容忍核酸层面的某些微小修饰,包括例如对沉默的核苷酸碱基(如偏好密码子)的修饰。在其它实施方案中,设想改变氨基酸的核苷酸碱基修饰,例如,以改善所得肽/蛋白质的表达(例如,密码子优化)。在一些实施方案中,设想由遗传密码的自然简并性引起的等位基因变异的修饰。

[0145] ABCA4基因的修饰还包括类似物,或编码的氨基酸序列的修饰型式。通常,这类类似物与特定鉴定的蛋白质的差异仅在于一到四个密码子变化。保守替换是发生在侧链和化学性质相关的氨基酸家族内的替换。

[0146] 功能性ABCA4基因的核酸序列可以源自任何天然表达功能性ABCA4或其同源物的哺乳动物。在其它实施方案中,对ABCA4基因序列进行某些修饰以增强在靶细胞中的表达。这些修饰包括密码子优化。

[0147] 例如,在国际专利公布No.WO 2017/087900中列出了特别是ABCA4外显子的突变,该专利公布以引用的方式并入本文中。

[0148] 如上所述,ABCA4视网膜病变是由ABCA4基因中的致病变异体引起的,其以常染色体隐性方式遗传。这类致病变异体(突变)包括错义变异体、无义变异体、剪接变异体、结构变异体和深内含子变异体。错义变异体占与ABCA4视网膜病变相关的变异体大多数。在患有ABCA4视网膜病变的个体中,个体的每个等位基因上的ABCA4都会发生突变。包含反式剪接分子的组合物可以纠正两个等位基因上的突变,无论突变在ABCA4基因内的位置如何。例

如,对于第一个等位基因上具有突变的ABCA4外显子2和第二个等位基因上具有突变的ABCA4外显子9的个体,可以使用编码外显子1-22的单个5'反式剪接分子(5'RNA编辑分子)来替换突变的ABCA4外显子2和突变的ABCA4外显子9。参见例如图14A。因此,在一些实施方案中,其中两个或更多个突变位于可被跨越两个或更多个突变的相同反式剪接分子替换的ABCA4基因的一部分上,具有含有功能性ABCA4外显子的编码区的单个反式剪接分子可替换含有突变的一个或多个外显子。

[0149] 或者,在一些实施方案中,个体具有ABCA4突变,其涉及ABCA4基因内远处外显子的突变。例如,其中个体在第一个等位基因上具有突变的ABCA4外显子2并且在第二个等位基因上具有突变的ABCA4外显子49,可以结合使用两种RNA编辑分子(例如,在组合物中)来替换每个突变的外显子。这种方法将涉及使用5'反式剪接分子来替换突变的ABCA4外显子2,并使用3'反式剪接分子来替换突变的ABCA4外显子49。在这类实施方案中,两种反式剪接分子可作为同一AAV载体的一部分共同递送或在单独的AAV载体中递送(例如,在两种反式剪接分子都超过AAV的包装限制的情况下)。

[0150] 编码结构域序列

[0151] 在一些实施方案中,5'反式剪接分子的CDS包括在靶ABCA4内含子(例如,ABCA4内含子22)5'的所有ABCA4外显子(例如,功能性ABCA4外显子)。例如,在5'反式剪接分子靶向ABCA4内含子22的实施方案中,CDS可包括功能性ABCA4外显子1-22。在这类实施方案中,CDS的长度可为2,000至4,500bp。在一些情况下,两个等位基因突变都发生在靶基因的5'部分,并选择5'反式剪接分子来纠正两个突变。在一个实施方案中,结合结构域与内含子22结合,并且CDS包括功能性ABCA4外显子1-22。

[0152] 在一些实施方案中,CDS(例如,编码RTM的转基因的)包括ABCA4外显子的cDNA(例如,功能性ABCA4外显子),用于替换突变的ABCA4外显子。例如,CDS内的一个或多个功能性ABCA4外显子可以是cDNA序列。在一些实施方案中,整个CDS是cDNA序列。另外或可替代地,CDS的全部或部分,或其一个或多个功能性ABCA4外显子,可以是天然存在的序列(例如,与内源ABCA4外显子具有90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%序列同一性的序列)。

[0153] 在一些实施方案中,CDS的全部或部分,或其一个或多个功能性ABCA4外显子,是密码子优化序列,其中核酸序列已被修饰,例如以增强表达或稳定性,而不会导致编码的氨基酸发生改变。密码子优化可以以例如美国专利No.7,561,972、7,561,973和7,888,112中描述的方式进行,每个专利都以引用的方式整体并入本文中。对于如本文所述通过重组AAV递送,在一个实施方案中,CDS可以是长度可达4,000个核苷酸碱基的核酸序列(例如,长度为3,000至4,000个核苷酸碱基,长度为3,100至3,800个核苷酸碱基,长度为3,200至3,700个核苷酸碱基或长度为3,300至3,500个核苷酸碱基,例如长度为3,000至3,100个核苷酸碱基,长度为3,100至3,200个核苷酸碱基,长度为3,200至3,300个核苷酸碱基,长度为3,300至3,400个核苷酸碱基,长度为3,400至3,500个核苷酸碱基,长度为3,500至3,600个核苷酸碱基,长度为3,600至3,700个核苷酸碱基,长度为3,700至3,800个核苷酸碱基,长度为3,800至3,900个核苷酸碱基或长度为3,900至4,000个核苷酸碱基,例如长度为约3,108个核苷酸碱基,长度为约3,285个核苷酸碱基,长度为约3,375个核苷酸碱基,长度为约3,503个核苷酸碱基,长度为约3,630个核苷酸碱基,长度为约3,540个核苷酸碱基,长度为约3,363

个核苷酸碱基,长度为约3,273个核苷酸碱基,长度为约3,145个核苷酸碱基或长度约3,018个核苷酸碱基)。

[0154] 在一些实施方案中,CDS的全部或部分,或其一个或多个功能性ABCA4外显子,被修饰以使其中鉴定的隐蔽剪接位点突变。本发明人通过大量实验和计算分析结果发现,在本文所述的ABCA4外显子编辑器的一些实施方案的背景下使用隐蔽剪接位点。简单地说,本发明人进行了大量实验和分析,包括生成脱靶 (OFT) 文库和其生物信息学分析,并且意外地发现在本文所述的反式剪接分子的背景下包含外显子1-22的ABCA4编码序列 (CDS) 内发生隐蔽剪接的大量位置。这一发现揭示了一个以前未被重视的问题,即RNA外显子编辑器(在本文中也称为RNA反式剪接分子或核酸反式剪接分子)可以在隐蔽剪接位点进行剪接,这可能导致具有潜在下游效应的不需要的剪接变体种类。为了提高本文所述的反式剪接分子的反式剪接效率和保真度,从而提供对隐蔽剪接位点利用问题的解决方案,通过引入同义突变来修饰多个隐蔽剪接位点的序列以损害和/或使隐蔽剪接位点失效。在一些实施方案中,根据通过OFT文库的生物信息学分析检测到的隐蔽剪接位点被利用的次数来鉴定被修饰的隐蔽剪接位点。在一些实施方案中,修饰的隐蔽剪接位点解释了多个隐蔽剪接位点事件。这种方法利用了遗传密码的简并性,允许引入所需的修饰以减少隐蔽剪接位点的使用,而不会导致编码氨基酸发生改变。参见例如表6和图26。

[0155] 使用下一代测序 (NGS) 方法,本发明人能够鉴定隐蔽剪接位点并定量包含外显子1-22的ABCA4 CDS内每个不同位置的隐蔽剪接使用频率,从而建立了数据驱动的、优先减少其中隐蔽剪接使用的方案。参见表3和图26。如表3所示,所鉴定的隐蔽剪接位点按使用频率顺序排列,使用频率由不包括破坏隐蔽剪接位点的同义突变的序列中每百万映射/合并读数的平均计数 (CPM) 决定。因此,结果和其生物信息学分析建立了ABCA4外显子1-22中的隐蔽剪接位点的优先列表,这些位点可以被修饰以生成在反式剪接分子背景下具有改良的特性和功能性的ABCA4 CDS。

[0156] 根据本文中呈现的结果,本发明人建立了使用频率阈值,当高于该阈值时,可以通过引入同义突变有利地对减少隐蔽剪接位点进行分析和设计。参见表6。在一些实施方案中,基于表3中所呈现的结果的阈值等于或大于10CPM。

[0157] 还实施了NGS方法来评估引入同义突变后的隐蔽剪接使用频率。参见表3和图26。通过这种方法确认隐蔽剪接位点的减少将序列修饰(同义突变)的结构变化与由此赋予的功能联系起来,即生成抗隐蔽剪接位点ABCA4 CDS。

[0158] 在一些实施方案中,RNA外显子编辑器分子(例如ABCA4外显子编辑器)在与外显子编辑器分子内的脱靶剪接相关的一个或多个位点处具有减少隐蔽剪接位点的核苷酸改变。任何或所有改变的核苷酸位置可以位于表3中鉴定的位置。在一些实施方案中,ABCA4外显子编辑器在以下一个或多个位置处具有核苷酸变化(根据例如SEQ ID NO:56编号),其中下面指示的核苷酸位置是指相对于ATG密码子的G(翻译起始位点)减少的隐蔽剪接位点核苷酸序列的5'和3'端,或其任何组合。在下面列出的序列中,所做的核苷酸改变以下划线标出。另参见表3:

[0159] 跨越核苷酸位置3066-3086的序列:

[0160] ACATGCTGTTCTATGCCAG (SEQ ID NO:91)

[0161] ACATGCTGTTCTATGCACAA (SEQ ID NO:92)

- [0162] 跨越核苷酸位置2911-2931的序列:
- [0163] CACCTTGTCATCCTCACAG (SEQ ID NO:93)
- [0164] CACCTTGTCATCCTCACTG (SEQ ID NO:94)
- [0165] 跨越核苷酸位置2556-2576的序列:
- [0166] TCGCTTGGTACCTTGATCAG (SEQ ID NO:95)
- [0167] TCGCTTGGTACCTTGATCAA (SEQ ID NO:96)
- [0168] 跨越核苷酸位置3009-3029的序列(剪接位点已纠正):
- [0169] GCCTTGGCATGTGTCCACAG (SEQ ID NO:97)
- [0170] CCCTTGGCATGTGTCCACAA (SEQ ID NO:98)
- [0171] 跨越核苷酸位置107-127的序列:
- [0172] CTGGTCTTGATCTGGTTAAG (SEQ ID NO:99)
- [0173] CTGGTCTTGATCTGGTTACG (SEQ ID NO:100)
- [0174] 跨越核苷酸位置2328-2348的序列:
- [0175] TCCTGTGCTTCGCCTGGCAG (SEQ ID NO:101)
- [0176] TCCTGTGCTTCGCCTGGCAA (SEQ ID NO:102)
- [0177] 跨越核苷酸位置1815-1835的序列:
- [0178] GCGGGTTTGCCTATCTGCAG (SEQ ID NO:103)
- [0179] GCGGGTTTGCATATCTGCAA (SEQ ID NO:104)
- [0180] 跨越核苷酸位置1198-1218的序列:
- [0181] CACTCCTGATTCACCTGCAG (SEQ ID NO:105)
- [0182] CACTCCTGATTCACCTGCTG (SEQ ID NO:106)
- [0183] 跨越核苷酸位置3063-3083的序列[下面指示的核苷酸变化是在位置3086(接近位置3083)的背景下进行的改变而引起的,并且没有纠正3083位置的剪接位点、多聚嘧啶区和/或分支点]:
- [0184] AGCACATGCTGTTCTATGGC (SEQ ID NO:107)
- [0185] AGCACATGCTGTTCTATGCA (SEQ ID NO:108)
- [0186] 跨越核苷酸位置2090-2110的序列:
- [0187] TGTACCTGGTTCCTGCACAG (SEQ ID NO:109)
- [0188] TGTACCTGGTTCCTGGACTC (SEQ ID NO:110)
- [0189] 下面列出的最后一个核苷酸对应于位置2608-2628(分支点已纠正):
- [0190] TTGGTACTTTCTTCTACAAG (SEQ ID NO:111)
- [0191] TTGGTACTTTCTTCTGCAAG (SEQ ID NO:112)
- [0192] 跨越核苷酸位置3083-3103的序列[下面指示的核苷酸变化是在位置3086(接近位置3103)的背景下进行的改变而引起的,并且没有纠正3103位置的剪接位点、多聚嘧啶区和/或分支点]:
- [0193] CAGCTGAAAGGAAAGTCCCA (SEQ ID NO:113)
- [0194] CAAGTAAAGGAAAGTCCCA (SEQ ID NO:114)
- [0195] 跨越核苷酸位置3073-3093的序列[下面指示的核苷酸变化是在位置3086(接近位置3093)的背景下进行的改变而引起的,并且没有纠正3093位置的剪接位点、多聚嘧啶区

和/或分支点]:

[0196] GTTCTATGCCCAGCTGAAAG (SEQ ID NO:115)

[0197] GTTCTATGCACAAGCTGAAAG (SEQ ID NO:116)

[0198] 跨越核苷酸位置681-701的序列(剪接位点已纠正):

[0199] CCCTGTGCTCCCTCTCCCAG (SEQ ID NO:117)

[0200] CCCTGTGCTCCCTCTCCCAA (SEQ ID NO:118)

[0201] 跨越核苷酸位置508-528的序列:

[0202] CATCGGCCTGTCTGACTCAG (SEQ ID NO:119)

[0203] CATCGGCCTGTCTGACTCTG (SEQ ID NO:120)

[0204] 跨越核苷酸位置1431-1451的序列(分支点已纠正):

[0205] TCCTAAACTTCCTCTACAAG (SEQ ID NO:121)

[0206] TCCTGAAGCTTCCTCTACAAG (SEQ ID NO:122)

[0207] 跨越核苷酸位置1995-2015的序列(剪接位点已纠正):

[0208] CTGTCTCCATGACTGTGAAG (SEQ ID NO:123)

[0209] CTGTCTCCATGACTGTAAAA (SEQ ID NO:124)

[0210] 跨越核苷酸位置3159-3179的序列(可以用非同义突变进行纠正):

[0211] AGCGGAATGAAGAGGCTCAG (SEQ ID NO:125)

[0212] AGCGGAATGAAGAGGCTCAG (SEQ ID NO:125)

[0213] 跨越核苷酸位置2234-2254的序列(剪接位点已纠正):

[0214] ATGCTGTGCTTTCTGCTCAG (SEQ ID NO:126)

[0215] ATGCTGTGCTTTCTGCTCTC (SEQ ID NO:127)

[0216] 跨越核苷酸位置1104-1124的序列(剪接位点、分支点已纠正):

[0217] TTTGTAATGCATTGATCCAG (SEQ ID NO:128)

[0218] TTTGTAATGCCTTGATACAA (SEQ ID NO:129)

[0219] 跨越核苷酸位置1887-1907的序列(剪接位点已纠正):

[0220] TTGGAATCTACCICAGCAG (SEQ ID NO:130)

[0221] TTGGAATCTACCTCCAACAA (SEQ ID NO:131)

[0222] 然而,熟练的从业者会意识到可以进行其它核苷酸改变以破坏所鉴定的隐蔽剪接位点,并且能够根据本文呈现的结果设想和设计这类改变。在一些实施方案中,ABCA4外显子编辑器在上述位置具有一个或多个核苷酸变化。另参见表3。在一些实施方案中,ABCA4外显子编辑器具有与SEQ ID NO:56-59所示的任一个序列具有至少80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或99.9%序列同一性或具有100%序列同一性的CDS序列。在一些实施方案中,ABCA4外显子编辑器在一个或多个上述位置(根据例如SEQ ID NO:56编号)具有减少隐蔽剪接位点的核苷酸变化并且具有与SEQ ID NO:56-58所示的任一个序列具有至少80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或99.9%序列同一性或具有100%序列同一性的CDS序列。在一些实施方案中,ABCA4外显子编辑器在一个或多个上述位置(根据例如SEQ ID NO:56编号)具有减少隐蔽剪接位点的核苷酸变化并且与SEQ ID NO:56所示的序列具有至少80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或99.9%序列同一性或具有100%序列同一性的

CDS序列。在一些实施方案中,ABCA4外显子编辑器在表3中所阐述的1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、25、30、35、40或45个位置中的至少一者处具有减少隐蔽剪接位点的核苷酸变化。

[0223] 在一些实施方案中,ABCA4外显子编辑器在表3中所阐述的剪接位点使用频率为至少10个计数/百万(又称为CPM)的1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个位置中的至少一者处具有减少隐蔽剪接位点的核苷酸变化,例外是表3中列出的不包括剪接位点的那些位点(例如,位置3083、3103或3093)。在一些实施方案中,如以下实施例7中所述的测定法所测量,ABCA4外显子编辑器分子在使用频率为至少10 cpm的1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个或所有隐蔽剪接位点[例外是表3中列出的不包括剪接位点的那些位点(例如,位置3083、3103或3093)]处具有减少隐蔽剪接位点的核苷酸变化。

[0224] 在一些实施方案中,ABCA4外显子编辑器在表3中所阐述的剪接位点使用频率为至少5计数/百万(又称为CPM)的1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个位置中的至少一者处具有减少隐蔽剪接位点的核苷酸变化,例外是表3中列出的不包括剪接位点的那些位点(例如,位置3083、3103或3093)。在一些实施方案中,如以下实施例7中所述的测定法所测量,ABCA4外显子编辑器分子在使用频率为至少5cpm的1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个或所有隐蔽剪接位点处[例外是表3中列出的不包括剪接位点的那些位点(例如,位置3083、3103或3093)]具有减少隐蔽剪接位点的核苷酸变化。

[0225] 在一些实施方案中,ABCA4外显子编辑器在表3中所阐述的剪接位点使用频率为至少3计数/百万(又称为CPM)的1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个位置中的至少一者处具有减少隐蔽剪接位点的核苷酸变化,例外是表3中列出的不包括剪接位点的那些位点(例如,位置3083、3103或3093)。在一些实施方案中,如以下实施例7中所述的测定法所测量,ABCA4外显子编辑器分子在使用频率为至少3cpm的1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个或所有隐蔽剪接位点处[例外是表3中列出的不包括剪接位点的那些位点(例如,位置3083、3103或3093)]具有减少隐蔽剪接位点的核苷酸变化。

[0226] 在一些实施方案中,ABCA4外显子编辑器在表3中所阐述的剪接位点使用频率为至少2计数/百万(又称为CPM)的1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个位置中的至少一者处具有减少隐蔽剪接位点的核苷酸变化,例外是表3中列出的不包括剪接位点的那些位点(例如,位置3083、3103或3093)。在一些实施方案中,如以下实施例7中所述的测定法所测量,ABCA4外显子编辑器分子在使用频率为至少2cpm的1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个或所有隐蔽剪接位点处[例外是表3中列出的不包括剪接位点的那些位点(例如,位置3083、3103或3093)]具有减少隐蔽剪接位点的核苷酸变化。

[0227] 在一些实施方案中,ABCA4外显子编辑器在表3中所阐述的剪接位点使用频率为至少1计数/百万(又称为CPM)的1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个位置中的至少一者处具有减少隐蔽剪接位点的核苷酸变化,例外是表3中列出的不包括剪接位点的那些位点(例如,位置3083、3103或3093)。在一些实施方案中,如以下实施例7中

所述的测定法所测量,ABCA4外显子编辑器分子在使用频率为至少1cpm的1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个或所有隐蔽剪接位点处[例外是表3中列出的不包括剪接位点的那些位点(例如,位置3083、3103或3093)]具有减少隐蔽剪接位点的核苷酸变化。

[0228] 除了ABCA4外显子编辑器之外,本文所述的实施方案还包括对其它基因靶标具有特异性的具有减少隐蔽剪接位点的核苷酸变化的外显子编辑器。减少隐蔽剪接位点的核苷酸变化可以包括消除或降低隐蔽剪接位点在剪接反应中使用的能力的变化。例如,在RNA外显子编辑器的背景下鉴定的隐蔽剪接位点通常包含剪接位点、多聚嘧啶区和分支点。在一些实施方案中,可以将一个或多个核苷酸变化引入到在RNA外显子编辑器的背景下鉴定的隐蔽剪接位点的剪接位点、多聚嘧啶区或分支点中的至少一个或它们的任何组合中。在一些实施方案中,确定核苷酸变化以尽量减少对其编码的蛋白质的潜在影响。本领域技术人员将理解,如果为降低隐蔽剪接位点使用频率而进行的核苷酸改变也改变了反式剪接RNA编码的氨基酸,则保守性氨基酸改变将优于非保守性氨基酸改变。此外,这类技术人员可以容易地分析蛋白质序列和结构,着眼于其中的功能结构域和重要序列,以评估是否可以合理地预期这种变化会改变由反式剪接蛋白质编码的蛋白质的功能。本领域技术人员还可以使用本领域已知的测定法测试包含这种氨基酸变化的蛋白质,以确定生物活性是否发生改变。在一些实施方案中,在所鉴定的隐蔽剪接位点内,超过一个核苷酸发生变化。在一些情况下,根据计算机预测和/或实验结果凭经验来确定应该改变多少个核苷酸。在一些实施方案中,可以将一个或多个(本文也称为至少一个)同义突变引入到在RNA外显子编辑器的背景下鉴定的隐蔽剪接位点的剪接位点、多聚嘧啶区或分支点中的至少一个或它们的任何组合中。同义突变不会改变反式剪接RNA编码的蛋白质的氨基酸序列。在一些实施方案中,可以将超过一个同义突变引入到在RNA外显子编辑器的背景下鉴定的隐蔽剪接位点的剪接位点、多聚嘧啶区或分支点中的至少一个或它们的任何组合中。

[0229] 除上述内容外,本发明人对实验结果和序列信息进行了大致如下的分析。通过搜索和替换剪接供体位点的某些元件,进行变化以去除实验中发现的隐蔽剪接位点。对于引入核苷酸变化,剪接供体位点末端的AG位点(以及更强的CAG位点)优先。如果未找到AG位点或无法在不引入非同义突变的情况下改变该位点,则扫描剪接位点上游42-44个碱基对的序列以查找分支点(与YNAH匹配的序列)。然后分析所鉴定的任何这类分支点序列并考虑用于引入核苷酸突变以减少实验中鉴定的隐蔽剪接位点处的隐蔽剪接位点的使用。此外,还扫描序列是否存在多聚嘧啶区(紧靠末端AG上游的多个Y)。典型地,这种多聚嘧啶区包含位于剪接位点上游10个碱基对内的至少5个嘧啶。一旦鉴定出,然后就分析这类多聚嘧啶区并考虑用于引入核苷酸突变以减少实验中鉴定的隐蔽剪接位点处的隐蔽剪接位点的使用。

[0230] 在一些实施方案中,被改变以减少脱靶剪接的隐蔽剪接位点或脱靶剪接位点是已经根据经验被鉴定为脱靶剪接位点的位点。例如,可以使用下面实施例6和7中描述的技术来鉴定这类位点。在一些实施方案中,所有使用频率高于预定阈值的隐蔽剪接位点都会被减少隐蔽剪接位点的核苷酸变化改变。

[0231] 在一些实施方案中,被改变以减少脱靶剪接的隐蔽剪接位点或脱靶剪接位点是已经预测是脱靶剪接位点的位点。可以基于序列分析做出这类预测,以鉴定其中推定的隐蔽剪接位点的典型剪接位点、多聚嘧啶区和/或分支点。在人类基因组中,隐蔽剪接位点的数

量比真正的剪接位点多一个数量级,并且可能具有与真正的剪接位点相同的剪接识别序列,但通常受到迄今为止尚不了解的分子机制的抑制。隐蔽5'剪接位点通常具有共有的NNN/GUNNN或NNN/GCNNN,其中N是任意核苷酸,并且“/”是外显子-内含子边界。隐蔽3'剪接位点通常具有共有的NAG/N。隐蔽剪接位点的激活受到周围核苷酸的正面影响,这些核苷酸可能使隐蔽剪接位点看起来更像真实剪接位点的典型共有序列,即分别为MAG/GURAGU和YAG/G,其中M为C或A,R为G或A,并且Y为C或U。

[0232] 在一些实施方案中,减少隐蔽剪接位点的核苷酸变化引起与典型剪接位点共有序列匹配的核苷酸序列不再与典型序列匹配。在一些实施方案中,减少隐蔽剪接位点的核苷酸变化消除潜在的剪接位点核苷酸。在一些实施方案中,减少隐蔽剪接位点的核苷酸变化消除潜在的多聚嘧啶区核苷酸。在一些实施方案中,减少隐蔽剪接位点的核苷酸变化消除潜在的分支点核苷酸。在一些实施方案中,减少隐蔽剪接位点的核苷酸变化是同义核苷酸变化。在一些实施方案中,减少隐蔽剪接位点的核苷酸变化引起外显子编辑器编码的氨基酸发生变化。在一些实施方案中,氨基酸变化是保守氨基酸取代。

[0233] 如例如表3和实施例6和7所述,本发明人通过实验观察了本文所述的RNA外显子编辑器中的隐蔽剪接位点的使用。对定位/鉴定的隐蔽剪接位点的进一步分析可鉴定在实验中鉴定的隐蔽剪接位点内的推定的剪接位点、多聚嘧啶区和分支点,这反过来又提供了关于应修饰哪些序列以减少隐蔽剪接位点接合的指导。这些结果广泛适用于一般的RNA外显子编辑器。

[0234] 在一些实施方案中,本文描述了一种RNA分子,其在5'到3'方向上包含:(a) cDNA编码结构域序列;(b) 剪接供体序列;以及(c) 结合结构域序列,所述结合结构域序列被配置成结合于内源RNA分子的内含子;其中编码结构域序列相对于内源RNA分子序列包含核苷酸突变,其中所述核苷酸突变破坏编码结构域序列内的隐蔽剪接位点。在一些实施方案中,核苷酸突变是同义核苷酸突变。在一些实施方案中,隐蔽剪接位点是通过实验鉴定的。在一些实施方案中,隐蔽剪接位点是根据计算机分析预测的。

[0235] 本文还涵盖一种修饰细胞中的RNA分子的方法,所述方法包括向细胞提供外源RNA分子,所述外源RNA分子在5'到3'方向上包含:(a) cDNA编码结构域序列,所述cDNA编码结构域序列包含与细胞中的内源靶RNA分子不同的核苷酸突变;(b) 剪接供体序列,所述剪接供体序列被配置成剪接到内源靶RNA分子的剪接受体序列;以及(c) 结合结构域序列,所述结合结构域序列被配置成结合于内源靶RNA分子的内含子;其中核苷酸突变破坏外源RNA分子的编码结构域序列内的隐蔽剪接位点。在一些实施方案中,核苷酸突变是同义核苷酸突变。在一些实施方案中,隐蔽剪接位点是通过实验鉴定的。在一些实施方案中,隐蔽剪接位点是根据计算机分析预测的。

[0236] 本文还涵盖一种增加RNA外显子编辑器的反式剪接效率的方法,所述方法包括将突变引入到RNA外显子编辑器的编码结构域序列中,其中所述突变破坏RNA外显子编辑器的编码结构域序列中的隐蔽剪接位点。在一些实施方案中,核苷酸突变是同义核苷酸突变。参见例如表6。在一些实施方案中,隐蔽剪接位点是通过实验鉴定的。在一些实施方案中,隐蔽剪接位点是根据计算机分析预测的。

[0237] 结合结构域

[0238] 本文所述的ABCA4反式剪接分子特征为结合结构域(BD),所述结合结构域被配置

成结合/退火靶ABCA4内含子和/或外显子。在一些情况下,靶ABCA4内含子是ABCA4内含子22。在一个实施方案中,结合结构域是与靶ABCA4 mRNA前体内含子(例如,靶ABCA4内含子)的序列至少80%互补(例如,至少85%互补、至少90%互补、至少91%互补、至少92%互补、至少93%互补、至少94%互补、至少95%互补、至少96%互补、至少97%互补、至少98%互补、至少99%互补或100%互补)的核酸序列,其可以抑制内源靶顺式剪接,同时增强反式剪接分子与靶ABCA4 mRNA前体之间的反式剪接(例如,通过创建具有内源ABCA4 mRNA的一部分和编码野生型ABCA4氨基酸序列的具有一个或多个功能性ABCA4外显子的编码结构域序列的嵌合分子)。在涉及反式剪接分子编码序列(例如,编码反式剪接分子的载体)的一个实施方案中,结合结构域编码序列编码与靶ABCA4mRNA前体内含子的序列至少80%互补(例如,至少85%互补、至少86%互补、至少87%互补、至少88%互补、至少89%互补、至少90%互补、至少91%互补、至少92%互补、至少93%互补、至少94%互补、至少95%互补、至少96%互补、至少97%互补、至少98%互补、至少99%互补或100%互补)的核酸序列。

[0239] 在一些情况下,本发明提供了在内含子22处结合ABCA4的反式剪接分子(或其载体),例如,其中核酸反式剪接分子被配置为将CDS反式剪接到内源ABCA4外显子23。具体地说,本文所述的反式剪接分子包括其中结合结构域与具有SEQ ID NO:16的核苷酸1至510或880至1,350中的任一个或多个(例如六个或更多个、八个或更多个、十个或更多个、或者十二个或更多个)核苷酸的结合位点结合的反式剪接分子。

[0240] 在一些情况下,结合结构域包括ABCA4内含子22的核苷酸1至510或880至1,350内的任六个或更多个连续核苷酸(例如,ABCA4内含子22的核苷酸1至510或880至1,350内的任八个或更多个连续核苷酸、ABCA4内含子22的核苷酸1至510或880至1,350内的任十个或更多个连续核苷酸、ABCA4内含子22的核苷酸1至510或880至1,350内的任12个或更多个连续核苷酸、ABCA4内含子22的核苷酸1至510或880至1,350内的任20个或更多个连续核苷酸、ABCA4内含子22的核苷酸1至510或880至1,350内的任30个或更多个连续核苷酸、ABCA4内含子22的核苷酸1至510或880至1,350内的任40或更多个连续核苷酸、ABCA4内含子22的核苷酸1至510或880至1,350内的任50或更多个连续核苷酸、ABCA4内含子22的核苷酸1至510或880至1,350内的任100或更多个连续核苷酸、ABCA4内含子22的核苷酸1至510或880至1,350内的任150或更多个连续核苷酸、ABCA4内含子22的核苷酸1至510或880至1,350内的任200或更多个连续核苷酸或ABCA4内含子22的核苷酸1至510或880至1,350内的任250个或更多个连续核苷酸)。

[0241] 在一些情况下,结合位点包括ABCA4内含子22的核苷酸880至1,350内的任六个或更多个连续核苷酸(例如,ABCA4内含子22的核苷酸880至1,350内的任八个或更多个连续核苷酸、ABCA4内含子22的核苷酸880至1,350内的任十个或更多个连续核苷酸、ABCA4内含子22的核苷酸880至1,350内的任12或更多个连续核苷酸、ABCA4内含子22的核苷酸880至1,350内的任20或更多个连续核苷酸、ABCA4内含子22的核苷酸880至1,350内的任30或更多个连续核苷酸、ABCA4内含子22的核苷酸880至1,350内的任40或更多个连续核苷酸、ABCA4内含子22的核苷酸880至1,350内的任50或更多个连续核苷酸、ABCA4内含子22的核苷酸880至1,350内的任100或更多个连续核苷酸、ABCA4内含子22的核苷酸880至1,350内的任150个或更多个连续核苷酸、ABCA4内含子22的核苷酸1至510或880至1,350内的任200或更多个连续核苷酸或ABCA4内含子22的核苷酸1至510或880至1,350内的任250个或更多个连续核苷酸)。

[0242] 在一些情况下,结合结构域具有与结合位点具有至少80%互补性的至少两个不重

叠的序列。

[0243] 在一些情况下,结合结构域包括与UUGUUGGUAAGGGUGUACAGCAAUUUCCUGGCUAUUAUAGUUGCAAAAAUCAGGAGUUAACUAAAAAAAAAACCCAAGGGAACUAAUUCAGCAGCAAA UUCCAGCAUUAUUGGACAAUAAUACCAACAUAUUCUAGCUCCUACAUAAC (SEQ ID NO:17) 具有至少80%同一性(例如,至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%、至少99%或100%同一性;例如,90%、91%、92%、93%、94%、95%、96%、97%、98%或99%同一性)的核酸序列。

[0244] 在一些实施方案中,结合结构域是与SEQ ID NO:18TTGTTGGTAAAGGGTGTACAGCAATTTCCTGGCTATATAGTTGCAAAAATCAGGAGTTAACTAAAAAAAAAACCCAAGGGAATAATTCAGCAGCAAATCCAGCATATTGGGACAATAATAACCAACATTTTCATAGCTTCCTACATAC (SEQ ID NO:18) 具有至少80%同一性(例如,至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%、至少99%或100%同一性;例如,90%、91%、92%、93%、94%、95%、96%、97%、98%或99%同一性)的DNA序列。

[0245] 在一些情况下,结合结构域包括与SEQ ID NO:19具有至少80%同一性(至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%、至少99%或100%同一性;例如,90%、91%、92%、93%、94%、95%、96%、97%、98%或99%同一性)的核酸序列。

[0246] 在一些实施方案中,结合结构域是与SEQ ID NO:20具有至少80%同一性(例如,至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%、至少99%或100%同一性;例如,90%、91%、92%、93%、94%、95%、96%、97%、98%或99%同一性)的DNA序列。

[0247] 在一些情况下,结合结构域包括与SEQ ID NO:21具有至少80%同一性(至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%、至少99%或100%同一性;例如,90%、91%、92%、93%、94%、95%、96%、97%、98%或99%同一性)的核酸序列。

[0248] 在一些实施方案中,结合结构域是与SEQ ID NO:22具有至少80%同一性(例如,至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%、至少99%或100%同一性;例如,90%、91%、92%、93%、94%、95%、96%、97%、98%或99%同一性)的DNA序列。

[0249] 在一些情况下,结合结构域包括与SEQ ID NO:23具有至少80%同一性(例如,至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%、至少99%或100%同一性;例如,90%、91%、92%、93%、94%、95%、96%、97%、98%

或99%同一性)的核酸序列。

[0250] 在一些实施方案中,结合结构域是与SEQ ID NO:24具有至少80%同一性(例如,至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%、至少99%或100%同一性;例如,90%、91%、92%、93%、94%、95%、96%、97%、98%或99%同一性)的DNA序列。

[0251] 结合结构域可以在3'可操作地连接到剪接结构域(例如,直接连接到剪接结构域或具有连接剪接结构域的3'端和结合结构域的5'端的间插序列)。

[0252] 如本文所详述,ABCA4外显子编辑器设计的第一阶段始于筛选和选择与靶向的mRNA前体内含子互补的高效BD序列。为了实现高CDS替换潜力并提供在AAV包装容量(约4.7Kb)的限制范围内可用于纠正高百分比ABCA4相关的视网膜病变患者群体中的突变的外显子编辑器,选择ABCA4基因的内含子22进行BD筛选。本发明人应用基于NGS的BD文库方法和基于生物发光的方法,用BD序列不同的外显子编辑器变体沿内含子22序列扫描。将两种方法的结果比对并成功鉴定出内含子内适合进行RNA外显子编辑的区域,从而在一些实施方案中选择BD117。参见图16。

[0253] 剪接结构域

[0254] 剪接结构域可以包括剪接位点、分支点和/或多聚嘧啶区(PPT)以介导反式剪接。在一些实施方案中,剪接结构域具有单个剪接位点,这表明由于缺乏相应的剪接位点,该剪接位点被设计用于优先进行反式剪接,而不是顺式剪接。

[0255] 本领域技术人员可以根据已知的方法和原理来选择可变剪接结构域。在一个实施方案中,5'剪接位点共有序列是核酸序列AG/GURAGU(其中/表示剪接位点)。在另一实施方案中,可采用与剪接位点近端的外显子和内含子相对应的内源剪接位点来维持任何剪接调节信号。

[0256] 在一个实施方案中,合适的具有间隔子的5'剪接位点为:5'-GTAAGA GAG CTC GTT GCG ATA TTA T-3'(SEQ ID NO:25)。在一个实施方案中,合适的5'剪接位点为AGGT。在一些实施方案中,剪接位点(例如,5'剪接位点)为或包含GTAAGT或GUAAGT。在一些实施方案中,剪接位点(例如,5'剪接位点)为或包含GTAAGG、GUAAGG、GTAAGC、GUAAGC、GTAAGT、GUAACU、CAAAGT或CAAAGU。

[0257] 剪接结构域可以在5'可操作地连接到终止子结构域(例如,直接连接到终止子结构域,或者具有连接剪接结构域的3'端和终止子结构域的5'端的间插序列,例如接头结构域和/或3'下游序列)。

[0258] 5'非翻译区

[0259] 在一些情况下,核酸反式剪接分子包括5'非翻译区。在一些实施方案中,5'非翻译区包含天然5'ABCA4非翻译区、基本上由其组成或由其组成。在一些实施方案中,5'非翻译区包含与SEQ ID NO:12-13中的任一者具有至少80%序列同一性(例如,至少85%序列同一性、至少86%序列同一性、至少87%序列同一性、至少88%序列同一性、至少89%序列同一性、至少90%序列同一性、至少91%序列同一性、至少92%序列同一性、至少93%序列同一性、至少94%序列同一性、至少95%序列同一性、至少96%序列同一性、至少97%序列同一性、至少98%序列同一性、至少99%序列同一性或100%序列同一性)的序列。

[0260] 在一些实施方案中,5'非翻译区可以在5'可操作地连接到CDS(例如,直接连接到CDS或具有连接5'非翻译区的3'端和CDS的5'端的间插序列)。

[0261] 5'调节结构域

[0262] 在一些情况下,核酸反式剪接分子操作性地连接到在5'操作性地连接到CDS的5'调节结构域(例如,直接连接到CDS,或通过中间结构域,例如非翻译区连接)。5'调节结构域可包括启动子(例如组成型启动子,例如CMV启动子或EF1- $\alpha$ 启动子)。在一些情况下,5'调节结构域包括操作性地连接到天然5'ABCA4非翻译区的启动子(例如组成型启动子,例如CMV启动子)。在一些实施方案中,操作性地连接到天然5'ABCA4非翻译区的5'调节结构域包含与SEQ ID No:14-15中的任一者具有至少80%序列同一性(例如,至少85%序列同一性、至少86%序列同一性、至少87%序列同一性、至少88%序列同一性、至少89%序列同一性、至少90%序列同一性、至少91%序列同一性、至少92%序列同一性、至少93%序列同一性、至少94%序列同一性、至少95%序列同一性、至少96%序列同一性、至少97%序列同一性、至少98%序列同一性、至少99%序列同一性或100%序列同一性)的序列。

[0263] 在一些实施方案中,5'调节结构域可以在5'可操作地连接到CDS(例如,直接连接到CDS或具有连接5'调节结构域的3'端和CDS的5'端的间插序列)。

[0264] 接头结构域

[0265] 如本文所讨论的,提高反式剪接效率仍然是将核酸反式剪接分子用作治疗剂的一个重要目标。为了设计出具有能够满足其在使用要求的反式剪接效率的核酸反式剪接分子,本发明人在剪接结构域(SD)与结合结构域(BD)之间测试了数千个接头序列,并鉴定出一个示例性接头,称为40聚体,其性能相对于前体接头(LinkD)具有统计学上显著程度的改善。参见例如图17A和17B。

[0266] 除上述内容外,核酸反式剪接分子可在分子的一个或多个位置处包括接头结构域。在一些实施方案中,接头结构域在3'操作性地连接到剪接结构域或剪接位点(例如,直接连接到剪接结构域或剪接位点)。接头结构域可以是任何合适的尺寸。在一些实施方案中,接头结构域长度超过25个核苷酸(例如,长度在25个与50个核苷酸之间、长度在35个与45个核苷酸之间或长度为约40个核苷酸(例如,40聚体接头))。在一些情况下,接头结构域包含与CUGGUGCCCCGCGGGCCGCGGAACCGGUUGGGGGCAUGUAC(SEQ ID NO:26)或CTGGTGCCCCGCGGGCCGCGGAACCGGTTGGGGGCATGTAC(SEQ ID NO:27)具有至少80%同一性(例如,至少85%序列同一性、至少86%序列同一性、至少87%序列同一性、至少88%序列同一性、至少89%序列同一性、至少90%序列同一性、至少91%序列同一性、至少92%序列同一性、至少93%序列同一性、至少94%序列同一性、至少95%序列同一性、至少96%序列同一性、至少97%序列同一性、至少98%序列同一性、至少99%序列同一性或100%序列同一性)的核酸序列、基本上由其组成或由其组成。

[0267] 如上所述,从文库筛选中选择出包含SEQ ID NO:26(RNA)或SEQ ID NO:27(DNA)的40聚体接头,其中对数千种不同的元件在反式剪接分子背景下的功能进行评估。例如,图1比较了不同元件组合背景下的三种不同接头,由此可以明显看出40聚体接头的优越性。在每种背景下,40聚体接头(表示为与最浅灰度色调相对应的接头;在每个ss的背景下从左到右表示的第三个接头)的存在与最高的相对反式剪接效率相关。接头序列通常包括在反式剪接分子中,其中它位于剪接供体与结合结构域之间,为每个元件提供灵活性和可及性。如

本文所述,在不同组合中,针对在反式剪接分子背景下赋予的活性来评估不同元件的功能贡献。在该筛选中,40聚体接头作为主要候选元件出现,其在反式剪接分子的背景下赋予改良的反式剪接活性,如相对于LinkD,ABCA4RNA的在靶(ONT)替换%在统计学上显著增加(图17A)并且反式剪接的ABCA4蛋白(ONT)水平显著增加并且非剪接蛋白(NSP)水平显著降低(图17B)所证明。

[0268] 在其它情况下,接头结构域包含与CCGAUACGACACGUAC AAGAUCU(SEQ ID NO:28)或CCGAATACGACACGTACAAGAT CT(SEQ ID NO:29)具有至少80%同一性(例如,至少85%序列同一性、至少86%序列同一性、至少87%序列同一性、至少88%序列同一性、至少89%序列同一性、至少90%序列同一性、至少91%序列同一性、至少92%序列同一性、至少93%序列同一性、至少94%序列同一性、至少95%序列同一性、至少96%序列同一性、至少97%序列同一性、至少98%序列同一性、至少99%序列同一性或100%序列同一性)的核酸序列、基本上由其组成或由其组成。在其它情况下,接头结构域包含与GCUGCUGCUCAGUCUCCUGGGCUGG(SEQ ID NO:30)或GCTGCTGCTCAGTCTCCTGGGCTGG(SEQ ID NO:31)具有至少80%同一性(例如,至少85%序列同一性、至少86%序列同一性、至少87%序列同一性、至少88%序列同一性、至少89%序列同一性、至少90%序列同一性、至少91%序列同一性、至少92%序列同一性、至少93%序列同一性、至少94%序列同一性、至少95%序列同一性、至少96%序列同一性、至少97%序列同一性、至少98%序列同一性、至少99%序列同一性或100%序列同一性)的核酸序列、基本上由其组成或由其组成。

[0269] RNA外显子编辑器活性的增效子

[0270] 本发明人使用多方面的实验方法来设计具有有利于用于治疗干预的特性的核酸反式剪接分子。因此,本发明人试图鉴定可以例如通过增加在靶(ONT)反式剪接、增加ONT ABCA4反式剪接蛋白水平、或降低非剪接蛋白(NSP)水平、或它们的任何组合而增强RNA外显子编辑器活性的另外的元件。因此,测试多种元件以确定相对于没有测试元件的RNA外显子编辑器所观察到的活性,测试元件的存在是否增强RNA外显子编辑器的活性。通过大量的反复试验程序,其中在个别反式剪接分子的背景下评估不同元件和其不同组合的评分,本发明人鉴定出一个序列元件,即U1结合位点(UBS),相对于缺乏UBS的RNA外显子编辑器,其使包含UBS作为附加元件的RNA外显子编辑器的ONT反式剪接增加和NSP水平降低。参见例如图13(3×U1)和图19A和19B(3×UBS)。在一些实施方案中,3×UBS(又称为3×U1)包含SEQ ID NO:62。在一些实施方案中,3×UBS(又称为3×U1)包含至少一个包含GGTAAGT或GGTAAGTAGAGTG(SEQ ID NO:61)或GGTGAGTAGAGTG(SEQ ID NO:83)的序列。在其一具体实施方案中,3×UBS位于接头结构域的3'。在其一具体实施方案中,3×UBS位于接头结构域的3'和结合结构域的5'。在其一具体实施方案中,3×UBS包含至少一个包含GGTAAGT的序列,其中至少一个包含GGTAAGT的序列中的每一者还包含侧翼序列。在其实施方案中,3×UBS包含GGTAAGTAGAGTG(SEQ ID NO:61)或GGTGAGTAGAGTG(SEQ ID NO:83)中的至少一者。在一些实施方案中,3×UBS包含以下中的至少一者:CGTGGTAAGTAGAGTG(SEQ ID NO:74;UBS1);ATG GGTAAGT AGAGTG(SEQ ID NO:75;UBS2);或TGG GGTAAGT AGAGTG(SEQ ID NO:76;UBS3),或它们的任何组合。

[0271] 因此,在一些实施方案中,包含至少一个UBS的RNA外显子编辑器包含至少一个包含GGTAAGT或GGTAAGTAGAGTG(SEQ ID NO:61)的序列。在其一具体实施方案中,3×UBS位于

RNA外显子编辑器中的接头结构域的3'。在其一具体实施方案中,3×UBS位于RNA外显子编辑器中的接头结构域的3'和结合结构域的5'。在一些实施方案中,包含至少一个UBS的RNA外显子编辑器包含以下中的至少一者:CGTGGTAAGTAGAGTG (UBS1;SEQ ID NO:74);ATG GGTAAGT AGAGTG (UBS2;SEQ ID NO:75);或TGG GGTAAGT AGAGTG (UBS3;SEQ ID NO:76);或它们的任何组合。在一些实施方案中,包含至少一个UBS的RNA外显子编辑器包含至少一个CGT GGTAAGT AGAGTG (UBS1;SEQ ID NO:74)。在一些实施方案中,包含至少一个UBS的RNA外显子编辑器包含CGT GGTAAGT AGAGTG (SEQ ID NO:74;UBS1)、ATG GGTAAGT AGAGTG (SEQ ID NO:75;UBS2)和TGGGGTAAGT AGAGTG (SEQ ID NO:76;UBS3)。

[0272] 根据本文中呈现的结果,在反式剪接分子背景下存在的至少一个UBS在本文中称为增效子。本文所述的一示例性增效子包含以下中的至少一者:CGTGGTAAGT AGAGTG (UBS1;SEQ ID NO:74)、ATG GGTAAGTAGAGTG (UBS2;SEQ ID NO:75)或TGG GGTAAGT AGAGTG (UBS3;SEQ ID NO:76),或它们的任何组合。在一些实施方案中,本文所述的一示例性增效子包含至少一个UBS,所述UBS包含至少一个CGT GGTAAGT AGAGTG (UBS1;SEQ ID NO:74)。在一些实施方案中,本文所述的一示例性增效子包含UBS,所述UBS包含CGT GGTAAGT AGAGTG (UBS1;SEQ ID NO:74)、ATG GGTAAGT AGAGTG (UBS2;SEQ ID NO:75)和TGG GGTAAGT AGAGTG (UBS3;SEQ ID NO:76)。

[0273] UBS (作为附加元件包括在RNA外显子编辑器中,并且位于其中相对于接头的3'和相对于BD的5';参见图13)增强RNA外显子编辑器活性的能力令人惊讶,至少因为可以合理地预测,至少一个可能充当替代5'剪接位点的附加元件的存在可能会通过潜在地与RNA外显子编辑器的剪接结构域(参见例如图13;SD;现有的真正的5'剪接位点)竞争和/或干扰其而减少ONT反式剪接,这种竞争/干扰会减少ONT反式剪接。因此,观察到的由RNA外显子编辑器中增效子(UBS中的至少一者,例如UBS1、UBS2或UBS3,或它们的任何组合)的存在赋予的ONT反式剪接的增加是出乎意料且令人惊讶的。增效子还使NSP统计上显著下降,这一观察结果进一步强调了该元件在RNA外显子编辑器背景下令人惊讶的功能特性。

[0274] 本文所述的增效子的功能特性也已在包含不同结合结构域的RNA外显子编辑器的背景下得到证实,所述不同结合结构域结合/靶向除ABCA4 mRNA前体之外的不同靶mRNA前体中的不同内含子。来自示例性RNA外显子编辑器的结果表明,包含3×UBS的增效子在靶向非ABCA4 mRNA前体中的内含子A和内含子B的RNA外显子编辑器背景下赋予令人惊讶的有益特性。更具体地说,相对于仅包含40聚体 (SEQ ID NO:27) 的RNA外显子编辑器,包含40聚体+3×UBS (SEQ ID NO:62) 组合的靶向非ABCA4 mRNA前体的RNA外显子编辑器表现出显著降低的NSP水平。参见例如图27B和28B。在降低NSP的同时,相对于单独的40聚体,40聚体+3×UBS (SEQ ID NO:62) 的组合维持相似水平的ONT反式剪接蛋白。通过比较,其中3×UBS显示可以减少NSP的每个RNA外显子编辑器[靶ABCA4mRNA前体的内含子22 (图19B);内含子A (图27B);和内含子B (图28B)]包括相同的40聚体+3×UBS序列盒,这支持以下结论:无论RNA外显子编辑器的靶内含子和/或靶mRNA前体如何,3×UBS序列 (SEQ ID NO:62) 的存在赋予的功能都是一致的。

[0275] 核酸反式剪接分子中3×UBS序列 (SEQ ID NO:62) 的存在所赋予的功能性在RNA反式剪接%水平上也很明显。图19A (靶向ABCA4mRNA前体的内含子22)、图28A (内含子B) 和图29 (内含子C) 证明3×UBS序列 (SEQ ID NO:62) 的存在可使RNA反式剪接%增加,这反映在

RNA替换%增加的趋势上。事实上,在靶向内含子C的示例性反式剪接分子中,3×UBS序列(SEQ ID NO:62)的存在使替换%水平增加了3倍。这些结果表明,无论RNA外显子编辑器的靶内含子和/或靶mRNA前体如何,3×UBS序列(SEQ ID NO:62)促进替换%增加的趋势的这种附加的功能特性都是一致的。值得注意的是,包含靶向内含子C的结合结构域(BD34)的RNA外显子编辑器包括一个替代的5'剪接位点,该位点不同于包含靶向ABCA4 mRNA前体的内含子22、内含子A和内含子B的结合结构域的每个RNA外显子编辑器中使用的5'剪接位点。这些结果表明,3×UBS增效子可以在除结合结构域和mRNA前体靶向特异性之外的元件方面有所不同的RNA外显子编辑器的背景下赋予功能性。

[0276] 因此,本发明人已呈现了证据证明,如本文所述的增效子(例如,SEQ ID NO:62)的增效子以不依赖于结合结构域和靶标的方式发挥作用,因此,可在一般RNA外显子编辑器的背景下使用,预期增效子的存在将赋予有利的特性,包括降低的NSP水平和可能增加的反式剪接活性,如RNA替换%所反映的。

[0277] RNA外显子编辑器活性的翻译增效子

[0278] 对能够通过增加在靶(ONT)反式剪接、增加ONTABCA4反式剪接蛋白水平或降低非剪接蛋白(NSP)水平或其任何组合来增强RNA外显子编辑器活性的额外元件的搜索也使本发明人发现富AU元件(ARE)在本文所述的RNA外显子编辑器的背景下表现出令人惊讶的功能特性。ARE的惊人特性是通过大量的反复试验程序发现的,其中在个别反式剪接分子的背景下评估不同元件和其不同组合的评分。本发明人证明,ARE的存在使得包含ARE作为附加元件(40聚体+ARE)的RNA外显子编辑器的NSP水平相对于缺乏ARE(例如,仅包含40聚体)的RNA外显子编辑器降低。在一些实验中,ARE的存在还使得包含ARE作为附加元件(40聚体+ARE)的RNA外显子编辑器中ONT反式剪接ABCA4蛋白相对于缺乏ARE(例如,仅包含40聚体)的RNA外显子编辑器增加。参见例如图13(ARE)和图19B(40聚体+ARE)。由于ARE使NSP水平降低和ONT反式剪接ABCA4蛋白水平增加,因此在本文中将它们称为“翻译增效子”。

[0279] 在一些实施方案中,一示例性ARE是粒细胞-巨噬细胞集落刺激因子(GM-CSF)ARE,其包括包含ATTTATATATTTATATTTTAAAA TATTTATTTATTTATTTATTTA(SEQ ID NO:63)的序列或与SEQ ID NO:63具有至少90%同一性的序列。在一些实施方案中,翻译增效子与SEQ ID NO:63具有至少91%同一性,与SEQ ID NO:63具有至少92%同一性,与SEQ ID NO:63具有至少93%同一性,与SEQ ID NO:63具有至少94%同一性,与SEQ ID NO:63具有至少95%同一性,与SEQ ID NO:63具有至少96%同一性,与SEQ ID NO:63具有至少97%同一性,与SEQ ID NO:63具有至少98%同一性,或与SEQ ID NO:63具有至少99%同一性。在一些实施方案中,SEQ ID NO:63或与SEQ ID NO:63具有至少90%同一性的序列在3'可操作地连接到增效子(例如,SEQ ID NO:62)。核心序列通常是AUUUA(RNA;编码其的DNA载体中的ATTTA),并且该核心AUUUA元件的重复通常是功能所必需的。ARE元件通常约等于或大于50个碱基。因此,在一些实施方案中,ARE包含约等于或大于50个碱基的核心AUUUA元件的重复。

[0280] ARE(作为附加元件包括在RNA外显子编辑器中;参见图13)增强RNA外显子编辑器活性的能力令人惊讶,至少因为它增加了ONT反式剪接ABCA4蛋白水平,同时降低了NSP水平。因此,ARE介导的对编码蛋白质水平的影响通过优先增加治疗性蛋白质水平(ONT反式剪接ABCA4蛋白水平)和降低NSP水平来促进RNA外显子编辑器的治疗潜力。

[0281] 还值得注意的是,示例性增效子与示例性翻译增效子(40聚体+3×UBS+ARE;SEQ

ID NO:73) 组合的盒使得NSP水平相对于仅包含40聚体、包含40聚体+3×UBS或包含40聚体+ARE的RNA外显子编辑器的NSP水平更加显著的降低。因此,该元件组合(40聚体+3×UBS+ARE;SEQ ID NO:73)在如本文所述的反式剪接分子的反式剪接分子背景下表现出特别有利的特性。

[0282] 本文所述的翻译增效子的功能特性也已在包含不同结合结构域的RNA外显子编辑器的背景下得到证实,所述不同结合结构域结合/靶向除ABCA4 mRNA前体之外的不同靶mRNA前体中的不同内含子。来自示例性RNA外显子编辑器的结果表明,包含ARE(SEQ ID NO:63)的翻译增效子在靶向非ABCA4 mRNA前体中的内含子A和内含子B的RNA外显子编辑器背景下赋予令人惊讶的有益特性。更具体地说,相对于仅包含40聚体(SEQ ID NO:27)的RNA外显子编辑器,包含40聚体+ARE(SEQ ID NO:63)组合的靶向非ABCA4mRNA前体的RNA外显子编辑器表现出显著降低的NSP水平。参见例如图27B和28B。在降低NSP的同时,相对于单独的40聚体,40聚体+ARE(SEQ ID NO:63)的组合维持相似水平的ONT反式剪接蛋白。通过比较,其中ARE显示可以减少NSP的每个RNA外显子编辑器[靶ABCA4 mRNA前体的内含子22(图19B);内含子A(图27B);和内含子B(图28B)]包括相同的40聚体+ARE序列(SEQ ID NO:86的序列3444-3529),这支持以下结论:无论RNA外显子编辑器的靶内含子和/或靶mRNA前体如何,ARE序列(SEQ ID NO:63)的存在赋予的功能都是一致的。使用40聚体+3×UBS+ARE(SEQ ID NO:73)盒时,NSP水平的降低更为明显,表明这些元件具有组合效应。这些结果表明,ARE翻译增效子可以在结合结构域和mRNA前体靶标特异性方面有所不同的RNA外显子编辑器的背景下赋予功能性,并且可以通过与其它元件组合的方式起作用以促进RNA外显子编辑器的功效。

[0283] 因此,本发明人已呈现了证据证明,如本文所述的增效子(例如,SEQ ID NO:63)的增效子以不依赖于结合结构域和靶标的方式发挥作用,因此,可在一般RNA外显子编辑器的背景下使用,预期翻译增效子的存在将赋予有利的特性,包括降低的NSP水平。

[0284] 另外的组件或修饰

[0285] 在一些情况下,反式剪接分子包括3'转录终止子结构域。在一些实施方案中,这类3'转录终止子结构域形成将反式剪接分子的3'端有效封端的三螺旋结构。在一些情况下,3'转录终止子结构域来自人长非编码RNAMALAT1(例如,野生型MALAT1)。在一些实施方案中,3'转录终止子结构域包括tRNA样结构域。可用作本发明ABCA4反式剪接分子的一部分的3'转录终止子结构域在国际专利公布No.WO2020/214973中进行了描述,该专利公布以引用的方式整体并入本文中。例如,在一些实施方案中,可操作地连接到结合结构域的3'端的RTM区域包括终止子结构域,所述终止子结构域包含野生型MALAT1+mascRNA结构域,例如SEQ ID NO:32或SEQ ID NO:33,基本上由其组成或由其组成。在一些实施方案中,可操作地连接到结合结构域的3'端的RNA外显子编辑器的区域包括终止子结构域,所述终止子结构域包含突变的MALAT1+mascRNA(抗Mut1 mascRNA)结构域,例如SEQ ID NO:66(DNA)或SEQ ID NO:67(RNA),基本上由其组成或由其组成。

[0286] 在一些情况下,反式剪接分子包括位于结合结构域下游(即,在3'操作性连接)和/或在终止子结构域上游(即,在5'操作性连接)的3'下游(DS)区域。在一些情况下,3'DS区域是终止子结构域的一部分。3'DS区域可将结合结构域直接连接到终止子区域。在一些实施方案中,3'DS区域长度为6-10个核苷酸(例如,长度为8个核苷酸)。在一些实施方案中,3'DS

区域包含以下任一者:3′DS-MALAT1侧翼8聚体RNA,AGGGUCAU;3′DS-MALAT1侧翼8聚体DNA,AGGGTCAT;3′DS-随机8聚体RNA,CGAGCCUC;3′DS-随机8聚体DNA,CGAGCCTC;3′DS-MALAT1侧翼前6个8聚体RNA,AGGGU;以及3′DS-MALAT1侧翼前6个8聚体DNA,AGGGT。在一些实施方案中,3′DS区域由以下中的任一者组成:3′DS-MALAT1侧翼8聚体RNA,AGGGUCAU;3′DS-MALAT1侧翼8聚体DNA,AGGGTCAT;3′DS-随机8聚体RNA,CGAGCCUC;3′DS-随机8聚体DNA,CGAGCCTC;3′DS-MALAT1侧翼前6个8聚体RNA,AGGGU;以及3′DS-MALAT1侧翼前6个8聚体DNA,AGGGT。

[0287] 在一些实施方案中,反式剪接分子与靶mRNA前体的结合由互补性百分比(即基于核酸的碱基配对特性)、三螺旋形成或蛋白质-核酸相互作用(如本文引用的文献所述)介导。在一个实施方案中,核酸反式剪接分子包括DNA、RNA或DNA/RNA杂交分子,其中DNA或RNA是单链或双链的。本文还包括可与前述RNA或DNA之一杂交的RNA或DNA,优选在严格条件下杂交,例如在60℃下在2.5x SSC缓冲液中杂交,并在37℃下在较低缓冲液浓度(例如0.5xSSC缓冲液)中洗涤几次。这些核酸可以编码表现出脂质磷酸酶活性和/或与质膜缔合的蛋白质。当在体外合成反式剪接分子时,可以在碱基部分、糖部分或磷酸骨架上对这类反式剪接分子进行修饰,例如以提高分子的稳定性、与靶mRNA的杂交、转运到细胞中、在细胞中对酶促裂解的稳定性等。例如,修饰反式剪接分子以降低总电荷可以增强分子的细胞摄取。此外,还可以进行修饰以降低对核酸酶或化学降解的敏感性。核酸分子可以通过与另一分子,例如肽、杂交触发的交联剂、转运剂、杂交触发的裂解剂等缀合的方式合成。

[0288] 可以引入各种其它众所周知的对核酸分子的修饰作为增加细胞内稳定性和半衰期的方法(另见上文关于寡核苷酸的内容)。可能的修饰是本领域已知的。可以对合成的反式剪接分子的结构进行的修饰包括骨架修饰。

[0289] 示例性ABCA4 RNA外显子编辑器盒图谱及概览

[0290] 一般来说,ABCA4 RNA外显子编辑器盒(ABCA4 REEC)包含驱动RNA外显子编辑器表达的调节组件、替换相应天然序列和靶天然序列内可能存在的任何突变的5′UTR和CDS、与天然剪接受体(SA)相互作用的剪接供体(SD)、可为这种相互作用的发生提供灵活性和可及性的接头、增强反式剪接(TS)的调节元件、靶向mRNA前体中正确位置的BD和促进TS和稳定性的终止子序列。图13和下表1中指示并简要描述了这些特征,其中呈现了示例性ABCA4 REEC:

[0291] 表1:

[0292]

|   |                                                                                                                           |
|---|---------------------------------------------------------------------------------------------------------------------------|
| 1 | 基于 AAV2 的反向末端重复序列(ITR), 对于 AAV 包装 <i>ABCA4</i> 外显子编辑器盒来说是不可或缺的。                                                           |
| 2 | 巨细胞病毒(CMV)启动子驱动 <i>ABCA4</i> 外显子编辑器的表达。                                                                                   |
| 3 | 来自人 <i>ABCA4</i> mRNA 的天然 5'非翻译区(UTR)维持天然 5'调节。                                                                           |
| 4 | <i>ABCA4</i> 外显子 1-21 包含同义密码子变化, 旨在通过 RNA 外显子编辑替换相应的内源编码序列。同义密码子变化去除隐蔽剪接受体位点, 以减少潜在的可变剪接反应。                               |
| 5 | <i>ABCA4</i> 外显子 22 包含同义密码子变化, 可实现基于序列的定量。                                                                                |
| 6 | 剪接供体(SD)反式剪接到内源 <i>ABCA4</i> 内含子 22 剪接受体位点, 使得内源 <i>ABCA4</i> mRNA 前体上游序列被包括外显子 22 在内的 <i>ABCA4</i> 外显子编辑器编码结构域序列(CDS)替换。 |
| 7 | 双终止密码子阻止从外显子 22 下游的 <i>ABCA4</i> 外显子编辑器(例如接头、3 × U1、富 AU 元件[ARE]、BD 结构域区域)翻译氨基酸。                                          |
| 8 | 接头为剪接供体位点提供了结构灵活性和可及性。                                                                                                    |

[0293]

|    |                                                                        |
|----|------------------------------------------------------------------------|
| 9  | 三个串联 U1 结合位点(3 × U1), 假定与 U1 snRNP 相关; 已证实可提高 RNA 外显子编辑效率。             |
| 10 | ARE, 假定会降低进入细胞质的任何未剪接的 <i>ABCA4</i> 外显子编辑器的稳定性。                        |
| 11 | 结合结构域(BD)通过序列互补性介导 <i>ABCA4</i> 外显子编辑器靶向内源 <i>ABCA4</i> mRNA 前体内含子 22。 |
| 12 | MALAT1 终止子(例如, MALAT1-MascRNA 抗 mut1)促进有效的 RNA 外显子编辑。                  |
| 13 | 基于 AAV2 的 ITR, 对于 AAV 包装外显子编辑器盒来说是不可或缺的。                               |

[0294] *ABCA4* REEC特征

[0295] 下面概述了关于图13中所呈现的示例性*ABCA4* REEC实施方案的特征的细节和背景。

[0296] 反向末端重复 (ITR) 序列: 5' 和 3' ITR 序列位于示例性 RNA 外显子编辑器序列的两侧, 并且对于腺相关病毒 (AAV) 包装转基因来说是不可或缺的。一示例性 *ABCA4* REEC 含有源自 AAV2 的“翻转方向”130 核苷酸 ITR 序列, 其常用于基因疗法中。

[0297] CMV/CMV启动子:该调节序列驱动示例性外显子编辑器转录本的表达。它包含巨细胞病毒(CMV)即刻早期增强子和CMV即刻早期增强子,已证明可在非人灵长类动物(NHP)光感受器中驱动高表达,并且本文显示可在NHP视网膜和人视网膜中驱动高表达。参见例如图8、9、31和32。

[0298] 5'UTR:示例性外显子编辑器RNA分子的5'端以天然ABCA45'UTR序列开头,该序列保留了由TS产生的最终ONT mRNA分子中的天然RNA序列。已通过五个点突变对UTR序列进行修饰,这些点突变消除可疑的或数据驱动的隐蔽剪接受体位点,以减少外显子编辑器转录本之间发生顺式剪接或TS的可能性。

[0299] 编码序列(CDS):一示例性ABCA4REEC编码ABCA4mRNA序列的外显子1至22,以在TS后替换相应天然序列(有关外显子22的另外的细节如下所呈现)。已通过超过40个同义突变对外显子1至21进行修饰,这些点突变消除可疑的或数据驱动的隐蔽剪接受体位点,以减少外显子编辑器转录本之间发生顺式剪接或TS的可能性。

[0300] 末端外显子22:示例性ABCA4 REEC中编码的CDS包括外显子22,称为末端外显子,因为在外显子编辑器序列中它前面紧挨着剪接供体(SD)位点,后面通过TS连接到天然外显子23,形成正确的外显子22-23接合。为了便于鉴定和定量由TS形成的ONT分子,已通过29个同义点突变对外显子22进行修饰。通过促进RT-qPCR分析(定量)的选择性引物/探针设计,以及基于下一代测序(NGS)的分析(鉴定和定量),而不引起蛋白质序列的氨基酸变化,这些突变提供了区分反式剪接转录本与内源转录本的机会。

[0301] 剪接供体位点(剪接结构域;SD):示例性ABCA4 REEC包含共有U1剪接供体位点(SD),它位于末端外显子22的直接下游。在剪接体介导的TS反应期间,该SD与内源内含子22剪接受体(SA)位点相互作用,代替了内源内含子22剪接供体位点与SA位点之间原本会发生的天然顺式剪接反应。该TS事件的结果是精确形成的外显子22-23接合,其中上游区域源自外显子编辑器,下游区域包含天然ABCA4mRNA前体的剩余部分,ABCA4mRNA前体继续进行其天然顺式剪接事件,以形成成熟的mRNA分子。

[0302] 终止密码子:示例性ABCA4 REEC包括双终止密码子,它紧接在SD位点之后。事实上,GTAAGT SD位点的最后一个核苷酸充当TAGTGA双终止密码子的第一个核苷酸。这一故意设置的终止密码子的一个目的是,在剪接前的外显子编辑器转录本离开细胞核、进入细胞质并进行翻译从而产生截短的非剪接蛋白质(NSP)的情况下,消除在C端添加外来氨基酸。NSP的C端的任何这类外来氨基酸都可能带来潜在的免疫原性风险,而包括终止密码子则可以避免这种风险。终止密码子还可促进富AU元件(ARE)介导的衰变,本发明人假定这将有助于降解逃逸到细胞质中的剪接前的外显子编辑器RNA,如下文所论述。

[0303] 40聚体接头:位于SD位点与下游元件之间的接头序列可以为RNA结构提供灵活性和对SD的可及性,这对于有效的TS是有利的。在对数千个接头候选物进行全面筛选后,40聚体接头作为一种高效序列脱颖而出。参见例如图1。包含40聚体接头的外显子编辑器相关区域的预测二级结构表明40聚体接头的存在会导致开放且可接近的SD。

[0304] 3×U1位点:一示例性ABCA4REEC在40聚体接头的下游包含一个增效子,它包含三个串联U1位点。在剪接体介导的剪接过程中,U1位点与U1 snRNP结合。在该外显子编辑器的主要SD位点下游包括三个串联U1位点会引起RNATS增加,以及观察到的NSP水平降低。本发明人假定这些结果可能是由于与剪接体的缔合增加以及在发生TS的细胞核中的保留所致。

[0305] 富AU元件 (ARE) : 一示例性ABCA4 REEC包含源自粒细胞-巨噬细胞集落刺激因子 (GM-CSF) 基因的mRNA的富AU元件 (ARE)。该ARE以及上游终止密码子的存在旨在模仿3'UTR样序列。本发明人假定,在剪接前的外显子编辑器RNA进入细胞质的情况下,ARE RNA结合蛋白可能会识别ARE序列并引起分子快速降解,从而降低NSP的水平。

[0306] 结合结构域 (BD) : 示例性ABCA4 REEC的编码的BD直接将外显子编辑器RNA与天然ABCA4前mRNA关联。其150个核苷酸 (nt) 序列与mRNA前体的内含子22 (chr1:94041452-94041601) 内相同长度的区域互补。BD的内含子位置是通过使用外显子编辑器阵列扫描内含子22的迭代过程来选择的,这些外显子编辑器在不同内含子位置的BD序列方面有所不同。该扫描过程需要采用基于大规模并行库NGS的方法评估数千个靶向内含子22的BD,并在个别外显子编辑器层面使用生物发光或qPCR作为读数进行评估。参见例如图16。

[0307] MALAT1三螺旋终止子: 示例性ABCA4 REEC还包括源自MALAT1长非编码RNA (lncRNA) 的终止子序列。MALAT1 lncRNA的这一部分含有三螺旋结构,该结构可以赋予外显子编辑器RNA稳定性和核保留。MALAT1终止子使得TS效率远远超过应用替代终止子 (例如自裂解核酶) 时所达到的效率。参见例如图20A。三螺旋后面是核糖核酸酶P裂解位点和一段包含MALAT1相关小细胞质RNA (mascRNA) 的短序列,如下所论述。三螺旋下游的核糖核酸酶P裂解决定了外显子编辑器转录本的3'端。

[0308] MascRNA: 示例性ABCA4 REEC中的mascRNA序列两侧是上游核糖核酸酶P裂解位点和下游核糖核酸酶Z裂解位点,认为一旦经过处理,它就会离开细胞核,而外显子编辑器转录本则会保留下来并参与TS。尽管在核糖核酸酶P裂解后mascRNA不再存在于外显子编辑器转录本中,但包括其天然存在时的MALAT1三螺旋下游序列可能有助于实现高效的TS结果。在细胞质中,mascRNA被怀疑通过与多tRNA合成酶复合物组分谷氨酰胺基-tRNA合成酶 (QARS) 相互作用和上调来促进整体蛋白质翻译和细胞增殖。为了降低这种潜在风险,示例性ABCA4 REEC中编码的mascRNA序列已进行突变,以在其反密码子茎环内包括一个点突变,称为抗Mut1,从而消除了其与QARS的相互作用以及随之而来的整体蛋白质合成率的可能增加。第二个点突变通过与抗Mut1突变进行碱基配对,帮助维持mascRNA的天然三叶草结构。值得注意的是,将抗Mut1引入mascRNA序列不会损害TS效率。参见例如图20B。

[0309] 如本文所述,示例性ABCA4 REEC编码示例性ABCA4 5'外显子编辑器RNA分子,该分子在由内源剪接体介导的5'反式剪接反应中发挥作用。如图14A所示,在外显子编辑器RNA转录后,外显子编辑器内与天然ABCA4 mRNA前体的内含子22互补的RNA序列 (结合结构域; BD) 将外显子编辑器RNA定位到mRNA前体。接下来,外显子编辑器的剪接供体位点 (SD) 和mRNA前体的内含子22剪接受体位点 (SA) 之间发生反式剪接 (TS) 反应。该反应导致外显子1至外显子22的天然内源mRNA前体序列被外显子编辑器RNA提供的相应编码序列 (CDS) 替换,从而纠正天然内源mRNA前体该区域内可能存在的任何突变。由此产生的在靶 (ONT) 分子经过翻译形成具有生物活性的ABCA4蛋白。参见例如图14A。

[0310] 除上述内容外,不受理论的束缚,本文所述的RNA外显子编辑器被认为是通过剪接体介导的反式剪接来操作的,反式剪接通过类似于顺式剪接的机制进行。转录后,外显子编辑RNA在细胞核内扩散,U1 snRNP沉积在其未配对的SD上。此外,在细胞核中,内源 (天然) mRNA前体靶标从其基因座转录,并且其内含子通过侧翼SD与剪接受体 (SA) 位点之间的顺式剪接而去除发生在每个内含子的SA出现后的数秒内。反式剪接过程由外显子编辑器结合结

构域(BD)与其位于内源ABCA4 mRNA前体内含子22内的互补序列的碱基配对启动,该过程必须在通过顺式剪接去除内含子之前发生。此时,剪接体可以进行以下:(1)使用内源SD和SA位点形成套索进行顺式剪接,或(2)使用外显子编辑器的SD与内源SA配对形成“Y分支”进行反式剪接。在后一种情况下,外显子编辑器的SD的第一个核苷酸(鸟嘌呤)与内含子的分支点形成共价键,该分支点通常位于SA上游15-50bp处。该键的形成导致外显子编辑器的下游部分与终止于外显子22的CDS部分分离,而外显子22仍然与剪接体相互作用。接下来,剪接体介导外显子编辑器的外显子22与内源mRNA前体的外显子23的连接,同时内含子被分离,从而形成正确的外显子22-23接合。最后,由外显子编辑器RNA的下游部分和分离的内含子序列组成的Y分支RNA会快速降解,类似于顺式剪接过程中形成的套索RNA的命运。参见例如图14B。

[0311] ABCA4外显子编辑器可以包括或排除上表1中列出的一个或多个组件。可以将各个功能组件有利地包括在ABCA4外显子编辑器中,无需包括表1中列出的所有组件。例如,虽然元件3、7、8、9、10和12为ABCA4外显子编辑器提供了优势,但是可以在没有这些元件中的一个或多个的情况下制成有效的ABCA4外显子编辑器。

[0312] 细胞系测定

[0313] 在一些情况下,在培养的细胞系中测试本文所述的反式剪接分子。为了筛选、选择和改善RNA外显子编辑器的功能,可以获得或工程化培养的细胞系以使靶向的ABCA4 mRNA前体以足够的水平表达。由于ABCA4在常见、强大且易于转染/转导的转化细胞系中表达极低,因此设计了一种基于HEK293T的细胞系,称为B6,它可表达高ABCA4RNA和蛋白质水平。这是通过CRISPR介导的同源性定向修复(HDR)将组成型CAGGS启动子敲入(KI)到内源ABCA4启动子与外显子1之间实现的。基因组DNA的PCR显示CAG启动子插入了B6细胞中至少一个等位基因中。然后使用四种不同的qPCR测定法确认B6细胞中的ABCA4 RNA表达。这与不表达ABCA4的亲本HEK293T细胞形成对比。通过蛋白质印迹法确认B6细胞中的蛋白质翻译。参见图3C。

[0314] 为了表征ABCA4 RNA外显子编辑器有效纠正靶天然mRNA前体的能力,评估所得的反式剪接在靶(ONT)RNA进行翻译并产生全长ABCA4蛋白的能力。为此,基于B6细胞系设计了一种敲除(KO)细胞系,称为17+06。17+06细胞继续产生高水平的ABCA4 RNA,但与B6细胞系截然不同,17+06细胞不产生ABCA4蛋白。通过使用CRISPR和靶向外显子3和4的引导RNA(gRNA)将突变引入ABCA4基因编码序列,实现了对17+06细胞的工程化。B6细胞经过两轮连续转染以实现高突变效率并破坏ABCA4基因序列,首先通过应用靶向外显子3的gRNA,然后通过应用靶向外显子3的gRNA与靶向外显子4的gRNA的混合物。通过对由17+06基因组DNA以及cDNA产生的PCR扩增子进行桑格测序(Sanger sequencing)来确认突变,其中cDNA可确认突变发生在生成亲本细胞系B6的过程中已成功进行CAG插入的等位基因内。RT-qPCR证实了17+06细胞系中ABCA4RNA表达的持续存在,而蛋白质印迹分析表明ABCA4蛋白翻译被消除。参见图3C。

[0315] RNA外显子编辑器筛选平台

[0316] 如图13所述,5'RNA外显子编辑器包含几个功能序列元件,例如用于靶向mRNA前体的结合结构域(BD)和能够接近剪接供体(SD)位点的接头。当为了给定的基因靶标对RNA外显子编辑器进行工程化时,将针对有助于提高反式剪接(TS)效率的能力,测试这些元件每

一者的多种序列选项。这种测试可以通过以下来完成：(A) 克隆和转染个别RNA外显子编辑器变体并通过RT-qPCR/ddPCR和蛋白质印迹分析效率；或(B) 克隆和汇集RNA外显子编辑器在基于高通量(HT)文库的方法中，该方法依赖于下一代测序(NGS)和计算分析来评估效率。ABCA4REEC是一种示例性ABCA4RNA外显子编辑器构建体，它是通过结合这些方法开发的。下面介绍了这两种方法。

[0317] 个别格式的RNA外显子编辑器筛选：

[0318] 该方法可用于在启动基于文库的多重筛选之前测试RNA外显子编辑器序列中的少量可变元件，以验证在多重筛选中鉴定的RNA外显子编辑器的性能，和/或提高主要候选物的性能。TS效率的评估发生在RNA和蛋白质水平。

[0319] 在RNA水平上，通过从细胞中分离总RNA，然后进行逆转录和实时定量PCR(RT-qPCR)测量例如以下靶标的RNA拷贝数来评估TS活性：RNF20(用于归一化的管家基因)；天然(ABCA4)mRNA；外显子编辑器RNA；在靶的外显子编辑的RNA(ONT)，它是阳性TS的产物；ONT+外显子编辑器+OFT(脱靶)——捕获所有这三个靶标的单一测定。OFT表示RNA外显子编辑器可能反式剪接的不正确的RNA分子。

[0320] ONTTS效率，也称为替换百分比，表示已成功TS的总ABCA4mRNA群体的部分，并通过以下等式计算： $ONT\ TS\ \% = 100 \times (ONT\text{拷贝数} / (ONT\text{拷贝数} + \text{天然拷贝数}))$ 。

[0321] RNA外显子编辑器TS效率是已正确反式剪接到ABCA4 RNA的RNA外显子编辑器转录本群体的部分，并通过以下等式计算： $\text{外显子编辑器TS}\ \% = 100 \times (ONT\text{拷贝数} / (ONT\text{拷贝数} + \text{外显子编辑器拷贝数} + \text{OFT拷贝数}))$ 。

[0322] 在蛋白质水平上，通过对从细胞或组织样品中提取的蛋白质进行蛋白质印迹分析来测量TS活性，在ABCA4的情况下，通过富集膜蛋白的方案进行。Na<sup>+</sup>/K<sup>+</sup>ATP酶是一种膜蛋白，用作上样对照。ONT蛋白的水平通过ABCA4特异性抗体(Ab)来测量。对于那些包括标签(例如，N端的V5标签)的构建体，可以使用对该标签具有特异性的Ab(例如，V5特异性Ab)来探测蛋白质印迹，从而评估ONT蛋白水平。

[0323] 文库格式的RNA外显子编辑器筛选：

[0324] 除上述内容外，RNA外显子编辑器可以以依赖于DNA寡核苷酸文库的合成的高通量文库格式进行筛选，所述格式可能侧重于外显子编辑器的具体特征，例如BD或接头。

[0325] 全长外显子编辑的ABCA4 mRNA序列的确认：

[0326] 为了确认由外显子编辑产生的ABCA4成熟mRNA具有预测由精确的反式剪接反应产生的序列，包括正确的外显子22-外显子23接合，本发明人从用编码ABCA4-01外显子编辑器的质粒瞬时转染的ABCA4 KO细胞中分离编辑的ABCA4 RT-PCR产物并进行测序。桑格测序证实了多个随机选择的克隆中成熟mRNA的序列正确。除了确认正确的外显子22-外显子23接合之外，这些测序结果还证实了全长外显子编辑的ABCA4成熟mRNA分子在其整个长度上具有预期的序列，包括外显子编辑器衍生的跨越外显子1至22的CDS和跨越外显子23至50的内源序列。这些结果提供了对通过ABCA4-01处理介导的ABCA4外显子编辑会产生预期的正确的ABCA4成熟mRNA序列的额外信心。

[0327] 外显子编辑的ABCA4蛋白的ATP酶功能活性评估：

[0328] ABCA4蛋白是ATP结合盒(ABC)转运蛋白超家族的成员，主要位于光感受器外节盘膜的边缘区域。ABC转运蛋白利用ATP水解的能量跨细胞膜单向转移多种底物，从离子到脂

质和肽。一旦进入眼睛的光转换成电信号(光转导),ABCA4蛋白就会变得活跃。光转导过程中会形成具有潜在毒性的维生素A衍生物,而ABCA4蛋白有助于从光感受器外节中去除这些物质中的至少两种(全反式视黄醛和11-顺式视黄醛)。这种光感受器清除机制帮助保护光感受器和视网膜色素上皮细胞免受这些分子引起的积聚和二次损伤。如上所述,隐性遗传的功能丧失型ABCA4突变会导致这些维生素A衍生物的积累,从而引起脂褐素的积聚,包括A2E(在脂褐素中发现的双类视黄醇吡啶盐N-亚视黄基-N-视黄基乙醇胺荧光团),导致视网膜细胞毒性,最终导致视力逐渐丧失。

[0329] ATP酶活性(包括由于底物结合而引起的活性增加)对ABCA4蛋白的功能至关重要。因此,评估从通过外显子编辑产生的ABCA4mRNA翻译而来的ABCA4蛋白的ATP酶活性可用作ABCA4生物活性的总体指标。先前的报告显示,40 $\mu$ M全反式视黄醛(ATR)与纯化的野生型(WT)ABCA4蛋白一起孵育会引起ATP酶活性增加1.8至2.5倍。考虑到这些信息,设计了实验来评估外显子编辑的ABCA4蛋白的生物活性。

[0330] 利用已建立的方法进行ABCA4蛋白免疫沉淀后,在存在或不存在40 $\mu$ M全反式视黄醛(ATR)的情况下,评估在用编码ABCA4-01的RNA外显子编辑器的N端V5表位标记型式(SEQ ID NO:81)的过表达质粒转染的ABCA4K0细胞中拯救的ABCA4蛋白的ATP酶活性。初始数据证明ATP酶活性的基础水平在添加40 $\mu$ M全反式视黄醛后有所增加。参见图24。在缺乏ABCA4的未转染样品中检测到最小的背景ATP酶活性。这些结果提供的证据表明,由ABCA4-01(SEQ ID NO:90)处理产生的外显子编辑的ABCA4蛋白将表现出从光感受器外节转运全反式视黄醛和过量的11-顺式视黄醛所需的功能性ATP酶活性。在一些实施方案中,SEQ ID NO:90的片段(例如,SEQ ID NO:69)可与不同的5'UTR组合,获得同等的结果。这些结果描绘了在表达示例性ABCA4 RNA外显子编辑器后ABCA4生物活性的恢复。

[0331] 支持使用ABCA4外显子编辑器治疗ABCA4相关的视网膜病变患者的体内NHP研究摘要:

[0332] 实施例4呈现的证据表明,基于AAV8的V5表位标记的ABCA4外显子编辑器在单次视网膜下施用1E11vg/注射1个月和在非洲绿猴(AGV)中产生治疗相关的外显子编辑活性。值得注意的是,本研究中测试的包含(SEQ ID NO:42)的早期先导载体用作进一步优化和最终选择ABCA4-01(SEQ ID NO:90)中存在的元件的基础,其包括包含SEQ ID NO:78的RNA外显子编辑器。参见例如图4-12;表1和表2。

[0333] 实施例8呈现的证据表明,ABCA4-01(SEQ ID NO:90)中所选择的调节元件在AGM中赋予长达3个月时间点的持续表达和外显子编辑活性。本研究中测试的AAV8构建体包含一种RNA外显子编辑器,该编辑器包含SEQ ID NO:77(其中前104个核苷酸包含天然ABCA45'UTR,后跟用于起始翻译的ATG起始密码子),其与包含(SEQ ID NO:42)的早期先导载体非常相似,但不同之处在于它缺少V5表位标签,在剪接结构域后含有双终止密码子,并且与ABCA4-01(SEQ ID NO:90)相比,在体外产生更高水平的截短ABCA4非剪接蛋白(NSP)。值得注意的是,本研究中测试的AAV8构建体包含与包含(SEQ ID NO:42)的早期先导载体非常相似的外显子编辑器(SEQ ID NO:77),其在所有动物中单次视网膜下注射4.3E10 vg/眼后1个月和3个月均耐受性良好,这进一步支持使用相关构建体ABCA4-01(SEQ ID NO:90)的临床方法。参见图22。包含SEQ ID NO:90(其又包含SEQ ID NO:78)的ABCA4-01 RNA外显子编辑器与本研究中使用的包含SEQ ID NO:77的RNA外显子编辑器不同之处在于SEQ ID NO:77

不包括增效子(例如,SEQ ID NO:62)或翻译增效子(例如,SEQ ID NO:63)。然而,SEQ ID NO:77和SEQ ID NO:78每一者均具有相同的调节元件和双终止密码子。

[0334] 实施例10呈现了证据表明ABCA4-01(SEQ ID NO:90)在NHP中赋予长达6个月时间点的持续表达和外显子编辑活性。更具体地说,图30A-30C呈现了证明在使用示例性AAV8 RNA外显子编辑器构建体(ABCA4-01;SEQ ID NO:90)处理六(6)个月,在NHP在体内实现了强大的ABCA4RNA和蛋白质替换的结果。简单地说,用单剂量视网膜下ABCA4-01处理野生型食蟹猴(呈现示例剂量为3.5E11vg/眼的结果)。对于该剂量的ABCA4-01,RNA替换百分比范围约为40%-60%。这使所得的人-NHP嵌合ABCA4蛋白达到治疗水平,如通过合格的IA-MS测定法所测量(图30C)。更具体地说,发现由这些动物中的ABCA4 RNA编辑产生的人NHP嵌合ABCA4蛋白占测试样品中总(人-NHP+NHP)ABCA4蛋白的20-40%。用通过单次视网膜下注射递送的较高剂量的ABCA4-01(1E12 vg/眼)处理的野生型食蟹猴中RNA替换高达66%,并且人/NHP嵌合ABCA4蛋白高达45%。这些结果表明,ABCA4-01在NHP中实现了治疗相关水平的编辑ABCA4蛋白表达,这些水平超过了先前在Abca4K0小鼠模型中显示赋予治疗益处的拯救ABCA4蛋白的水平。

[0335] 实施例12呈现了证据表明ABCA4-01(包含SEQ ID NO:90的示例性AAV8 RNA外显子编辑器构建体)在人光感受器背景下发挥作用。简单地说,用ABCA4-01(3.9E11vg)处理来自多个人类供体的供体来源的视网膜外植体,以评估ABCA4-01在人光感受器中的活性。ABCA4-01处理引起大约20-30%的内源外显子1-22被ABCA4-01编码的相同对应的外显子替换的RNA替换(图31)。值得注意的是,同一批ABCA4-01也用于实施例10中提到的调查性NHP研究,进一步证明了ABCA4-01结果从NHP到人类受试者的可转化性。

### [0336] III. 载体

[0337] 可以使用各种技术,例如使用重组腺相关病毒(AAV)载体或其它载体形式,例如非病毒载体,将反式剪接分子递送到个体的靶细胞。因此,本文提供了包含/编码反式剪接分子的载体(例如包含/编码反式剪接分子的病毒或非病毒载体,例如包含/编码反式剪接分子的DNA载体)。任何合适的核酸载体均可与本发明的组合物和方法结合使用,以设计和装配反式剪接分子和重组AAV的组件。在一个实施方案中,载体是携带反式剪接分子的重组AAV,其由启动子驱动,在个体的选定细胞中表达反式剪接分子。重组载体的装配方法是本领域已知的。参见例如Ausubel等人,Current Protocols in Molecular Biology,John Wiley&Sons,New York,1989;Kay,M.A.等人,Nat.Medic,2001,7(1):33-40;以及Walther W.和Stein U.,Drugs 2000,60(2):249-71。

[0338] 在本文所述的某些实施方案中,反式剪接分子借助于AAV载体递送到需要治疗的选定细胞,例如感光细胞。可利用多种天然存在的AAV血清型。AAV衣壳存在多种天然变体,允许鉴定和使用具有特别适合眼部或耳蜗细胞的特性的AAV。人工AAV载体可以通过常规分子生物学技术进行工程化,从而可以优化这些颗粒,以便将反式剪接分子核酸序列特异性地递送到细胞,最大限度地降低免疫原性,调节稳定性和颗粒寿命,实现高效降解,准确递送到细胞核等。例如,这种人工衣壳可以通过任何合适的技术,使用选定的AAV序列(例如,vp1衣壳蛋白的片段)与可以从不同的选定AAV、相同AAV的非连续部分、非AAV病毒源或非病毒源获得的异源序列组合来生成。人工AAV可以是但不限于假型AAV、嵌合AAV衣壳、重组AAV衣壳或“人源化”AAV衣壳。其中一个AAV的衣壳被异源衣壳蛋白替换的假型载体可用于递送

本文所述的反式剪接分子。

[0339] 可通过用含有包含/编码所需反式剪接分子的序列的重组工程化AAV或人工AAV递送,在选定细胞中实现本文所述的反式剪接分子的表达。使用AAV是外源DNA递送的一种常见方式,因为它相对无毒,提供有效的基因转移,并且针对特定目的容易优化。在从人或非人灵长类动物中分离出的已充分表征的AAV血清型中,人血清型2已广泛用于在不同靶组织和动物模型中进行有效的基因转移实验。

[0340] 在一些实施方案中,AAV是AAV1或其变体(例如,US20030138772的SEQ ID NO:6或64或US20150159173的SEQ ID NO:11或27)、AAV2或其变体(例如,US20030138772的SEQ ID NO:7或70、US20150159173的SEQ ID NO:7或23或US20150159173的SEQ ID NO:7)、AAV2G9或其变体、AAV3或其变体(例如,US20030138772的SEQ ID NO:8或71)、AAV3a或其变体、AAV3b或其变体(例如,美国专利No.6,156,303的SEQ ID NO:1和10)、AAV3-3或其变体(例如,WO2005033321的SEQ ID NO:200和217)、AAV4或其变体(例如,US20030138772的SEQ ID NO:63)、AAV4-4或其变体(例如,WO2005033321的SEQ ID NO:201或218)、AAV5或其变体(例如,US20030138772的SEQ ID NO:114)、AAV6或其变体(例如,US20030138772的SEQ ID NO:65)、AAV6.1或其变体(例如,US20150159173的SEQ ID NO:29)、AAV6.2或其变体、AAV6.1.2或其变体、AAV7或其变体(例如,US20030138772的SEQ ID NO:1-3)、AAV7.2或其变体、AAV8或其变体(例如,US20030138772的SEQ ID NO:4和95或AAV8(b)(与野生型AAV8相比,在氨基酸位置587-595上具有Pro-Glu-Arg-Thr-Ala-Met-Ser-Leu-Pro(SEQ ID NO:132)的氨基酸序列,如美国专利No.9,567,376(其以引用的方式整体并入本文中)中所述))、AAV9或其变体(例如,US20030138772的SEQ ID NO:5和100)、AAV9.9或其变体、AAV9.11或其变体、AAV9.13或其变体、AAV9.16或其变体、AAV9.24或其变体、AAV9.45或其变体、AAV9.47或其变体、AAV9.61或其变体、AAV9.68或其变体、AAV9.84或其变体(参见例如N.Pulicherla等人,Molecular Therapy 19(6):1070-1078(2011),以引用的方式整体并入本文中)、AAV10或其变体(例如,US20030138772的SEQ ID NO:117)、AAV11或其变体(例如,US20030138772的SEQ ID NO:118)、AAV12或其变体(例如,US20030138772的SEQ ID NO:119)、AAV16.3或其变体、AAV24.1或其变体、AAV27.3或其变体、AAV42.12或其变体、AAV42-1b或其变体、AAV42-2或其变体、AAV42-3a或其变体、AAV42-3b或其变体、AAV42-4或其变体、AAV42-5a或其变体、AAV42-5b或其变体、AAV42-6b或其变体、AAV42-8或其变体、AAV42-10或其变体、AAV42-11或其变体、AAV42-12或其变体、AAV42-13或其变体、AAV42-15或其变体、AAV42-aa或其变体、AAV43-1或其变体、AAV43-12或其变体、AAV43-20或其变体、AAV43-21或其变体、AAV43-23或其变体、AAV43-25或其变体、AAV43-5或其变体、AAV44.1或其变体、AAV44.2或其变体、AAV44.5或其变体、AAV223.1或其变体、AAV223.2或其变体、AAV223.4或其变体、AAV223.5或其变体、AAV223.6或其变体、AAV223.7或其变体、AAV1-7/rh.48或其变体、AAV1-8/rh.49或其变体、AAV2-15/rh.62或其变体、AAV2-3/rh.61或其变体、AAV2-4/rh.50或其变体、AAV2-5/rh.51或其变体、AAV3.1/hu.6或其变体、AAV3.1/hu.9或其变体、AAV3-9/rh.52或其变体、AAV3-11/rh.53或其变体、AAV4-8/rh.64或其变体、AAV4-9/rh.54或其变体(例如,WO2005033321的SEQ ID NO:116)、AAV4-19/rh.55或其变体(例如,WO2005033321的SEQ ID NO:117)、AAV5-3/rh.57或其变体、AAV5-22/rh.58或其变体、AAV7.3/hu.7或其变体、AAV16.8/hu.10或其变体、AAV16.12/hu.11或其变体、AAV29.3/bb.1或其变体、AAV29.5/

bb.2或其变体、AAV106.1/hu.37或其变体、AAV114.3/hu.40或其变体、AAV127.2/hu.41或其变体、AAV127.5/hu.42或其变体、AAV128.3/hu.44或其变体、AAV130.4/hu.48或其变体、AAV145.1/hu.53或其变体、AAV145.5/hu.54或其变体、AAV145.6/hu.55或其变体、AAV161.10/hu.60或其变体、AAV161.6/hu.61或其变体、AAV33.12/hu.17或其变体、AAV33.4/hu.15或其变体、AAV33.8/hu.16或其变体、AAV52/hu.19或其变体、AAV52.1/hu.20或其变体、AAV58.2/hu.25或其变体、AAVA3.3或其变体、AAVA3.4或其变体、AAVA3.5或其变体、AAVA3.7或其变体、AAVC1或其变体、AAVC2或其变体、AAVC5或其变体、AA V-DJ或其变体(例如,US20140359799的SEQ ID NO:2或3)、AA V-DJ8或其变体、AAVF3或其变体、AAVF5或其变体、AAVH2或其变体、AAVH6或其变体、AAVLK03或其变体、AAVH-1/hu.1或其变体、AAVH-5/hu.3或其变体、AAVLG-10/rh.40或其变体、AAVLG-4/rh.38或其变体、AAVLG-9/hu.39或其变体、AAVN721-8/rh.43或其变体、AAVCh.5或其变体(例如,US20150159173的SEQ ID N 0 46)、AAVCh.5R1或其变体、AAVcy.2或其变体、AAVcy.3或其变体、AAVcy.4或其变体、AAVcy.5或其变体(例如,US20150159173的SEQ ID NO:8和24)、AAVCy.SR1或其变体、AAVCy.5R2或其变体、AAVCy.5R3或其变体、AAVCy.5R4或其变体、AAVcy.6或其变体、AAVhu.1或其变体(例如,W02005033321的SEQ ID NO:144)、AAVhu.2或其变体(例如,W02005033321的SEQ ID NO:143)、AAVhu.3或其变体(例如,W02005033321的SEQ ID NO:145)、AAVhu.4或其变体(例如,W02005033321的SEQ ID NO:141)、AAVhu.5或其变体、AAVhu.6或其变体(例如,W02005033321的SEQ ID NO:84)、AAVhu.7或其变体(例如,W02005033321的SEQ ID NO:150)、AAVhu.9或其变体(例如,W02005033321的SEQ ID NO:155)、AAVhu.10或其变体(例如,W02005033321的SEQ ID NO:156)、AAVhu.11或其变体(例如,W02005033321的SEQ ID NO:153)、AAVhu.13或其变体(US20150159173的SEQ ID NO:16和32)、AAVhu.15或其变体(例如,W02005033321的SEQ ID NO:147)、AAVhu.16或其变体(例如,W02005033321的SEQ ID NO:148)、AAVhu.17或其变体(例如,W02005033321的SEQ ID NO:83)、AAVhu.18或其变体(例如,W02005033321的SEQ ID NO:149)、AAVhu.19或其变体(例如,W02005033321的SEQ ID NO:133)、AAVhu.20或其变体(例如,W02005033321的SEQ ID NO:134)、AAVhu.21或其变体(例如,W02005033321的SEQ ID NO:135)、AAVhu.22或其变体(例如,W02005033321的SEQ ID NO:138)、AAVhu.23.2或其变体(例如,W02005033321的SEQ ID NO:137)、AAVhu.24或其变体(例如,W02005033321的SEQ ID NO:136)、AAVhu.25或其变体(例如,W02005033321的SEQ ID NO:146)、AAVhu.26或其变体(例如,US20150159173的SEQ ID NO:17和33)、AAVhu.27或其变体(例如,W02005033321的SEQ ID NO:140)、AAVhu.28或其变体(例如,US20150159173的SEQ ID NO:42)、AAVhu.29或其变体(例如,W02005033321的SEQ ID NO:132)、AAVhu.29R或其变体、AAVhu.31或其变体(例如,W02005033321的SEQ ID NO:121)、AAVhu.32或其变体(W02005033321的SEQ ID NO:122)、AAVhu.34或其变体(例如,W02005033321的SEQ ID NO:125)、AAVhu.35或其变体(例如,W02005033321的SEQ ID NO:164)、AAVhu.37或其变体(例如,US20150159173的SEQ ID NO:18和34)、AAVhu.39或其变体(例如,W02005033321的SEQ ID NO:102)、AAVhu.40或其变体(例如,W02005033321的SEQ ID NO:87)、AAVhu.41或其变体(例如,W02005033321的SEQ ID NO:91)、AAVhu.42或其变体(例如,W02005033321的SEQ ID NO:85)、AAVhu.43或其变体(例如,W02005033321的SEQ ID NO:160)、AAVhu.44或其变体(例如,US20150159173的SEQ ID NO:45)、AAVhu.44R1或其变体、AAVhu.44R2或其变体、

AAVhu.44R3或其变体、AAVhu.45或其变体(例如,W02005033321的SEQ ID NO:127)、AAVhu.46或其变体(例如,W02005033321的SEQ ID NO:159)、AAVhu.47或其变体(例如,W02005033321的SEQ ID NO:128)、AAVhu.48或其变体(例如,US20150159173的SEQ ID NO:38)、AAVhu.48R1或其变体、AAVhu.48R2或其变体、AAVhu.48R3或其变体、AAVhu.49或其变体(例如,W02005033321的SEQ ID NO:189)、AAVhu.51或其变体(例如,W02005033321的SEQ ID NO:190)、AAVhu.52或其变体(例如,W02005033321的SEQ ID NO:191)、AAVhu.53或其变体(例如,US20150159173的SEQ ID NO:19和35)、AAVhu.54或其变体(例如,W02005033321的SEQ ID NO:188)、AAVhu.55或其变体(例如,W02005033321的SEQ ID NO:187)、AAVhu.56或其变体(例如,W02005033321的SEQ ID NO:192)、AAVhu.57或其变体(例如,W02005033321的SEQ ID NO:193)、AAVhu.58或其变体(例如,W02005033321的SEQ ID NO:194)、AAVhu.60或其变体(例如,W02005033321的SEQ ID NO:184)、AAVhu.61或其变体(例如,W02005033321的SEQ ID NO:185)、AAVhu.63或其变体(例如,W02005033321的SEQ ID NO:195)、AAVhu.64或其变体(例如,W02005033321的SEQ ID NO:196)、AAVhu.66或其变体(例如,W02005033321的SEQ ID NO:197)、AAVhu.67或其变体(例如,W02005033321的SEQ ID NO:198)、AAVhu.14/9或其变体、AAVhu.t 19或其变体、AAVrh.2或其变体(例如,US20150159173的SEQ ID NO:39)、AAVrh.2R或其变体、AAVrh.8或其变体(例如,US20150159173的SEQ ID NO:41)、AAVrh.8R或其变体、AAVrh.10或其变体(例如,US20150159173的SEQ ID NO:9和25)、AAVrh.12或其变体、AAVrh.13或其变体(例如,US20150159173的SEQ ID NO:10和26)、AAVrh.13R或其变体、AAVrh.14或其变体、AAVrh.17或其变体、AAVrh.18或其变体、AAVrh.19或其变体、AAVrh.20或其变体(例如,US20150159173的SEQ ID NO:1)、AAVrh.21或其变体、AAVrh.22或其变体、AAVrh.23或其变体、AAVrh.24或其变体、AAVrh.25或其变体、AAVrh.31或其变体、AAVrh.32或其变体、AAVrh.33或其变体、AAVrh.34或其变体、AAVrh.35或其变体、AAVrh.36或其变体、AAVrh.37或其变体(例如,US20150159173的SEQ ID NO:40)、AAVrh.37R2或其变体、AAVrh.38或其变体(例如,W02005033321的SEQ ID NO:86)、AAVrh.39或其变体(例如,US20150159173的SEQ ID NO:3、20或36)、AAVrh.40或其变体(例如,W02005033321的SEQ ID NO:92)、AAVrh.43或其变体(例如,US20150159173的SEQ ID NO:21和37)、AAVrh.46或其变体(例如,US20150159173的SEQ ID NO:4和22)、AAVrh.48或其变体(例如,US20150159173的SEQ ID NO:44)、AAVrh.48.1或其变体(例如,US20150159173的SEQ ID NO:44)、AAVrh.48.1.2或其变体、AAVrh.48.2或其变体、AAVrh.49或其变体(例如,W02005033321的SEQ ID NO:103)、AAVrh.50或其变体(例如,W02005033321的SEQ ID NO:108)、AAVrh.51或其变体(例如,W02005033321的SEQ ID NO:104)、AAVrh.52或其变体(例如,W02005033321的SEQ ID NO:96)、AAVrh.53或其变体(例如,W02005033321的SEQ ID NO:97)、AAVrh.54或其变体(例如,US20150159173的SEQ ID NO:49)、AAVrh.56或其变体(例如,W02005033321的SEQ ID NO:152)、AAVrh.57或其变体(例如,W02005033321的SEQ ID NO:105)、AAVrh.58或其变体(例如,US20150159173的SEQ ID NO:48)、AAVrh.61或其变体(例如,W02005033321的SEQ ID NO:107)、AAVrh.62或其变体(例如,W02005033321的SEQ ID NO:114)、AAVrh.64或其变体(例如,US20150159173的SEQ ID NO:43)、AAVrh.64R1或其变体、AAVrh.64R2或其变体、AAVrh.67或其变体(例如,US20150159173的SEQ ID NO:47)、AAVrh.73或其变体(例如,US20150159173的SEQ ID NO:5)或AAVrh.74或其变体(例如,

US2015015917的SEQ ID NO:6)。变体的非限制性实例包括US20030138772的SEQ ID NO:9、27-45、47-62、66-69、73-81、84-94、96、97、99和101-113(其内容以引用的方式整体并入本文中)和W02005033321的SEQ ID NO:1、2、4-82、89、90、93-95、98、100、101、109-113、118-120、124、126、131、139、142、151、154、158、161、162、165-183、202、204-212、215、219和224-236(其内容以引用的方式整体并入本文中)。在一个实施方案中,AAV血清型是U.S.2021/0189430(其内容以引用的方式整体并入本文中)中描述的那些AAV血清型中的任何一种。AAV的氨基酸序列可以在AAV衣壳蛋白中与硫酸肝素蛋白聚糖相互作用的一个或多个位置处或者在对应于氨基酸484、487、527、532、585或588的一个或多个位置处包括一个或多个氨基酸取代,编号基于AAV2的VP1编号。

[0341] 除非另有说明,否则本文所述的AAV ITR和其它选定的AAV组件可以从任何AAV血清型中容易地选择,所述血清型包括但不限于AAV1、AAV2、AAV3、AAV4、AAV5、AAV6、AAV7、AAV8、AAV9或其它已知和未知的AAV血清型。在一个实施方案中,ITR来自AAV2。可以使用本领域技术人员可利用的技术从AAV血清型中容易地分离这些ITR或其它AAV组件。这类AAV可以从学术、商业或公共来源(例如,美国典型培养物保藏中心(Manassas, Va.))分离或获得。或者,AAV序列可通过合成或其它合适方式参考已公开的序列(例如在文献中或数据库(例如GenBank、PubMed等)中可获得的序列)获得。

[0342] 适合装配成载体的AAV片段包括cap蛋白(包括vp1、vp2、vp3和高变区)、rep蛋白(包括rep 78、rep 68、rep 52和rep 40)以及编码这些蛋白质的序列。这些片段可容易地用于各种载体系统和宿主细胞。这些片段可以单独使用,也可以与其它AAV血清型序列或片段组合使用,或者与来自其它AAV或非AAV病毒序列的元件组合使用。如本文所用,人工AAV血清型包括但不限于具有非天然存在的衣壳蛋白的AAV。这种人工衣壳可以通过任何合适的技术,使用选定的AAV序列(例如,vp1衣壳蛋白的片段)与可以从不同的选定AAV血清型、相同AAV血清型的非连续部分、非AAV病毒源或非病毒源获得的异源序列组合来生成。人工AAV血清型可以是但不限于假型AAV、嵌合AAV衣壳、重组AAV衣壳或“人源化”AAV衣壳。其中一个AAV的衣壳与具有不同衣壳蛋白的AAV的ITR一起使用的假型载体如本文所述是有用的。在一个实施方案中,AAV是AAV2/5(即,具有AAV2 ITR和AAV5衣壳的AAV)。在另一实施方案中,AAV是AAV2/8(即,具有AAV2 ITR和AAV8衣壳的AAV)。在一个实施方案中,AAV包括AAV8衣壳。这类AAV8衣壳包括在NCBI参考序列:YP\_077180.1下发现的氨基酸序列。在另一实施方案中,AAV8衣壳包括由GenBank登录号:AF513852.1的nt 2121至4337编码的衣壳。

[0343] 在一个实施方案中,在本文所述的组合物和方法中有用的载体至少含有编码选定的AAV血清型衣壳(例如AAV2衣壳)或其片段的序列。在另一实施方案中,有用的载体至少含有编码选定的AAV血清型rep蛋白(例如AAV2 rep蛋白)或其片段的序列。任选地,这类载体可以同时包含AAV cap和rep蛋白。在提供AAV rep和cap的载体中,AAV rep和AAV cap序列可以都来自一种血清型,例如AAV2来源。

[0344] 或者,可使用其中rep序列来自与提供cap序列的AAV血清型不同的AAV血清型的载体。在一个实施方案中,rep和cap序列由不同的来源(例如,不同的载体,或宿主细胞和载体)表达。在另一实施方案中,这些rep序列与不同AAV血清型的cap序列同框融合,形成嵌合AAV载体,例如美国专利No.7,282,199(以引用的方式并入本文中)中描述的那些。

[0345] 通过培养宿主细胞来生成合适的重组AAV(rAAV),该宿主细胞含有编码如本文定

义的AAV血清型衣壳蛋白或其片段的核酸序列;功能性rep基因;由例如AAV ITR和反式剪接分子核酸序列构成的小基因;以及足够的辅助功能以允许将小基因包装到AAV衣壳蛋白中。在宿主细胞中培养以将AAV小基因包装到AAV衣壳中所需的组件可以呈反式提供给宿主细胞。或者,任何一个或多个所需组件(例如,小基因、rep序列、cap序列和/或辅助功能)可由稳定的宿主细胞提供,该宿主细胞已使用本领域技术人员已知的方法被工程化为含有一个或多个所需组件。

[0346] 在一个实施方案中,AAV包括启动子(或启动子的功能片段)。可以从大量能够在所需靶细胞中表达所选转基因的组成型或诱导型启动子中选择出将rAAV中采用的启动子。参见例如国际专利公布No.WO2014/012482(以引用的方式并入本文中)中鉴定的启动子列表。在一个实施方案中,启动子是细胞特异性的。术语“细胞特异性”意指为被选择用于重组载体的特定启动子可以指导所选转基因在特定细胞类型中的表达。在一个实施方案中,启动子特异性地在感光细胞中表达转基因。在另一实施方案中,启动子特异性地在视杆细胞和/或视锥细胞中表达。在另一实施方案中,启动子特异性地在视网膜色素上皮(RPE)细胞中表达转基因。在另一实施方案中,启动子特异性地在神经节细胞中表达转基因。在另一实施方案中,启动子特异性地在苗勒细胞(Mueller cell)中表达转基因。在另一实施方案中,启动子特异性地在双极细胞中表达转基因。在另一实施方案中,启动子特异性地在水平细胞中表达转基因。在另一实施方案中,启动子特异性地在无长突细胞中表达转基因。在另一实施方案中,转基因在上述任意细胞中表达。

[0347] 在另一实施方案中,启动子是要表达的靶基因的天然启动子。有用的启动子包括但不限于:杆状视蛋白启动子、红绿视蛋白启动子、蓝视蛋白启动子、cGMP-磷酸二酯酶启动子、小鼠视蛋白启动子、视紫红质启动子、视锥细胞转导的 $\alpha$ 亚基、 $\beta$ 磷酸二酯酶(PDE)启动子、视网膜色素变性启动子、NXNL2/NXNL 1启动子、RPE65启动子、视网膜缓慢变性/周围蛋白2(Rds/perph2)启动子和VMD2启动子。

[0348] 小基因或rAAV中含有的其它常规调节序列也在如WO2014/124282等文献以及本文引用并以引用的方式并入的其它文献中公开。本领域技术人员可以在不脱离本文所述的范围下在这些和其它表达控制序列中做出选择。

[0349] 所选择的遗传元件可以通过任何合适的方法,包括本文所述的方法来递送。用于构建本文所述的任何实施方案的方法是核酸操纵领域的技术人员已知的,并且包括基因工程、重组工程和合成技术。参见例如Sambrook等人,Molecular Cloning.A Laboratory Manual,Cold Spring Harbor Press,Cold Spring Harbor,NY。类似地,生成rAAV病毒体的方法是众所周知的,并且合适方法的选择并不限制本文所述的方法和构建体。参见例如K.Fisher等人,J.Virology,199370:520-532和美国专利5,478,745,每一者以引用的方式并入本文中。

[0350] 在另一实施方案中,反式剪接分子包括在前病毒质粒中,例如国际专利公布No.WO 2012/158757(以引用的方式并入本文中)中所公开的那些。这样的前病毒质粒含有模块化重组AAV基因组,所述AAV基因组包含操作性缔合的以下各者:野生型5'AAV2ITR序列,两侧是独特的限制性位点,允许很容易地去除或替换所述ITR;启动子,所述启动子包含位于巨细胞病毒(CMV)-鸡 $\beta$ 肌动蛋白序列上游的49个核酸巨细胞病毒序列;或光感受器特异性启动子/增强子,所述启动子两侧是独特的限制性位点,允许很容易地去除或替换整个启动子

序列,并且上游序列两侧是独特的限制性位点,允许很容易地从启动子序列中仅去除或替换上游CMV或增强子序列。本文所述的反式剪接分子可插入多克隆多聚接头的位点中,其中反式剪接分子可操作地连接到启动子并受启动子的调节控制。两侧是独特的限制性位点,允许很容易地去除或替换所述多聚腺苷酸序列的牛生长激素多聚腺苷酸化序列;以及两侧是独特的限制性位点,允许很容易地去除或替换3' ITR的野生型3' AAV2 ITR序列;也是这种质粒的一部分。质粒骨架包含在细菌细胞中复制所必需的元件,例如卡那霉素(kanamycin)抗性基因,并且其本身两侧是转录终止子/绝缘子序列。

[0351] 在一个实施方案中,前病毒质粒包含:(a)模块化重组AAV基因组,所述AAV基因组包含操作性缔合的以下各者:(i)野生型5' AAV2 ITR序列,两侧是独特的限制性位点,允许很容易地去除或替换所述ITR;(ii)启动子,所述启动子包含(A)CMV-鸡 $\beta$ 肌动蛋白序列上游的49个核苷酸CMV序列;(B)光感受器特异性启动子/增强子;或(C)神经元细胞特异性启动子/增强子。启动子两侧是独特的限制性位点,允许很容易地去除或替换整个启动子序列,并且上游序列两侧是独特的限制性位点,允许很容易地从启动子序列中仅去除或替换上游CMV或增强子序列。该前病毒质粒的一部分还包括多克隆多接头序列,该多接头序列允许插入包括本文所述任何序列的反式剪接分子序列,其中反式剪接分子可操作地连接到启动子并受启动子的调节控制;牛生长激素多聚腺苷酸化序列,两侧是独特的限制性位点,允许很容易地去除或替换所述多聚腺苷酸序列;以及野生型3' AAV2 ITR序列,两侧是独特的限制性位点,允许很容易地去除或替换3' ITR。前病毒质粒还含有质粒骨架,该质粒骨架包含在细菌细胞中复制所必需的元件,还包含卡那霉素抗性基因,所述质粒骨架两侧是转录终止子/绝缘子序列。本文所述的前病毒质粒还可以在质粒骨架中含有非编码的 $\lambda$ 噬菌体5.1kb填充序列,以增加骨架长度并防止无功能性AAV基因组的反向包装。

[0352] 在又一方面,修饰前病毒质粒的启动子以减小启动子的尺寸,从而允许更大的反式剪接分子序列插入到rAAV中。在一个实施方案中,CMV/CBA混杂启动子(其通常包括总计约1,000个碱基对的非编码外显子和内含子)被130个碱基对的嵌合内含子替换,如国际专利公布No. WO 2017/087900(以引用的方式整体并入本文中)中所述。

[0353] 然后以目前常规的包装方法采用这些前病毒质粒来生成表达前病毒质粒携带的反式剪接分子转基因的重组病毒。本领域技术人员很容易选择合适的生产细胞系。例如,合适的宿主细胞可以从任何生物体中选择,包括原核(例如,细菌)细胞,和真核细胞,包括昆虫细胞、酵母细胞和哺乳动物细胞。简单地说,将前病毒质粒转染到选定的包装细胞中,它可能在那里暂时存在。或者,小基因或基因表达盒与其侧翼ITR以染色体形式或附加体形式稳定地整合到宿主细胞的基因组中。合适的转染技术是已知的并且可以容易地用来将重组AAV基因组递送到宿主细胞。通常,前病毒质粒在表达cap和/或rep蛋白的宿主细胞中培养。在宿主细胞中,由反式剪接分子与侧翼AAV ITR组成的小基因被拯救并包装到衣壳蛋白或包膜蛋白中,形成感染性病毒颗粒。因此,通过在足够的病毒序列存在的情况下培养携带前病毒质粒的包装细胞以允许将基因表达盒病毒基因组包装到感染性AAV包膜或衣壳中来产生重组AAV感染性颗粒。

[0354] 或者,可使用非AAV载体(例如非病毒载体)来递送反式剪接分子。可以使用本领域已知或本文描述的任何合适的非病毒载体技术。适合于递送反式剪接分子的这类非病毒载体包括脂质体(例如阳离子脂质体、单层脂质体或多层脂质体)、纳米颗粒(例如聚合物纳米

颗粒、脂质纳米颗粒 (LNP)、聚乙二醇化纳米颗粒 (例如聚乙二醇化LNP)、肽纳米颗粒、金属纳米颗粒等)、树枝状聚合物 (例如阳离子树枝状聚合物, 例如聚丙烯亚胺树枝状聚合物)、外泌体 (例如免疫惰性和/或靶向外泌体, 例如使用Alvarez-Erviti, 等人, 2011, Nat. Biotechnol. 29:341中描述的技术制成) 和微泡。在一些情况下, 可使用细胞穿透肽 (CPP) 来递送本文所述的反式剪接分子, CPP可转移靶细胞的质膜并促进反式剪接分子递送到靶细胞内部。

#### [0355] IV. 药物组合物和药盒

[0356] 本文提供了药物组合物, 所述药物组合物包括核酸反式剪接分子、前病毒质粒或包含本文所述的任意ABCA4核酸反式剪接分子的rAAV。在一些实施方案中, 药物组合物包括本文所述的任意5'反式剪接分子。在一些实施方案中, 药物组合物包括5'反式剪接分子和3'反式剪接分子, 例如; 其中5'反式剪接分子和3'反式剪接分子一起含有ABCA4外显子的功能序列并结合同一个靶ABCA4内含子 (例如内含子22)。

[0357] 可以制备这类药物组合物以使其纯净无污染并适合于体内施用。可以通过常规方法评估本文所述的药物组合物的污染情况, 然后配制成打算用于合适施用途径的药物组合物。还可以使用合适的载剂以类似方式配制含有反式剪接分子 (例如裸DNA) 的其它组合物。这种制剂涉及使用药理学上和/或生理学上可接受的媒介物或载剂, 特别是指定施用于靶细胞 (例如, 眼部细胞)。在一个实施方案中, 适合于施用于靶细胞的载剂包括缓冲盐水、等渗氯化钠溶液或其它缓冲液 (例如HEPES) 以将pH维持在适当的生理水平, 以及任选的其它药物制剂、药剂、稳定剂、缓冲剂、载剂、佐剂、稀释剂等。

[0358] 在一些实施方案中, 载剂是注射液。示例性的生理学上可接受的载剂包括无菌、无热原的水和无菌、无热原的磷酸盐缓冲盐水。美国专利No. 7, 629, 322 (以引用的方式并入本文中) 中提供了多种这类已知载体。在一个实施方案中, 载剂是等渗氯化钠溶液。在另一实施方案中, 载剂为平衡盐溶液。在一个实施方案中, 载剂包括tween。如果要长期保存病毒, 可以在甘油或TWEEN 20存在下冷冻。

[0359] 在其它实施例中, 含有本文所述的反式剪接分子的组合物包括表面活性剂。可以包括有用的表面活性剂, 例如Pluronic F68 (Poloxamer 188, 也称为LUTROL®F68), 因为它们可以防止AAV粘附到惰性表面上, 从而确保所需剂量的递送。举例来说, 一种设计用于治疗本文所述的眼部疾病的示例性组合物包含携带编码本文所述的5'反式剪接分子或3'反式剪接分子的核酸序列的重组腺相关载体和药理学上可接受的载剂, 所述核酸序列受在哺乳动物个体的眼部细胞中表达反式剪接分子的调节序列的控制。载剂为等渗氯化钠溶液, 并包括表面活性剂Pluronic F68。在一个实施方案中, 反式剪接分子是本文中描述的任何反式剪接分子。

[0360] 在又一个示例性实施方案中, 组合物包含rAAV病毒, 所述rAAV病毒包含本文所述的用于ABCA4基因纠正的任何ABCA4反式剪接分子, 所述核酸序列受指导反式剪接分子在感光细胞中的表达的启动子的控制, 其中组合物与载剂和适合于视网膜下注射的另外的组分一起配制。在另一实施方案中, 可以将组合物或用于产生或装配该组合物的组分 (包括载剂、rAAV颗粒、表面活性剂和/或用于生成rAAV的组分) 以及用于制备该组合物的合适的实验室硬件纳入药盒中。这类药盒还可包括将组合物施用于个体的说明书, 例如用于治疗眼部疾病, 例如斯特格氏病或乌谢尔综合征 (Usher Syndrome)。

[0361] 本文另外提供了含有包含5'反式剪接分子的药物组合物的药盒(例如,其中反式剪接分子包装在本文所述的任何AAV载体中)。在一些实施方案中,药盒包括在施用前混合药物组合物的说明书。

[0362] V方法和用途

[0363] 患有ABCA4相关的视网膜病变的患者表现出相似但不相同的临床特征。绝大多数有脂褐素积累和双侧视力丧失。ABCA4相关的斯特格氏病通常显示特征性的黄斑萎缩,在眼底镜检查中常呈现被打青铜色外观。眼底黄色斑点症类似,但可能显示出较少的黄斑萎缩和黄斑中黄白色斑点更大的存在,延伸到中周边区域。患有ABCA4相关的视锥-视杆细胞营养不良的受试者呈现出如视力丧失、色觉受损和中央暗点等视觉症状。

[0364] ABCA4相关的斯特格氏病的临床表现可分为儿童期发病、成年早发型和成年晚发型。这些群体中的每一个可能都有自己的表型变化,这些变化与特定的ABCA4变体相关。

[0365] 儿童期发病:具有双等位基因严重ABCA4变体的儿童呈现出更有攻击性的中心视力丧失,这归因于单独的黄斑功能障碍或与视锥-视杆细胞营养不良组合,通常发生在五至十一岁之间。

[0366] 成年早发型:这些患者通常在二十或三十岁时因中央凹萎缩而呈现中心视力丧失。在呈现时,眼底镜检查通常显示明显的黄斑萎缩,并伴有大量黄斑旁和周边斑点。这些患者通常携带ABCA4变体,对ABCA4功能有中等影响(例如c.6079C>T和c.[2588G>C;5603A>T]),要么是二个等位基因都有影响,要么是与无效/严重ABCA4变体的复合杂合组合。

[0367] 成年晚发型:“晚”发型疾病的定义从年龄 $\geq 35$ 岁到年龄 $\geq 50$ 岁不等。因此,患者可以从40岁起出现症状。患者通常在视网膜筛查时偶然被发现,因为其最佳矫正视力(BCVA)保存得相对较好,认为这可以由萎缩不影响中央凹来解释。

[0368] 大多数ABCA4相关的视网膜病变患者在二十岁左右被诊断出/出现症状,这会对他们的发展、独立性、教育和整体生活质量产生重大影响。此外,一些更严重的形式,例如视锥-视杆细胞营养不良,在儿童早期就开始,最小为5岁。一项针对95名斯特格氏病患者的研究发现,一旦视力(VA)降至20/40以下,视力就会迅速丧失至20/200。此外,19岁时至少有一只眼睛维持20/40或更好的视力的概率为52%,但到39岁时下降到22%。一般来说,较早发病且严重程度较高的视锥-视杆细胞营养不良以及儿童期发病的斯特格氏病通常与ABCA4的双等位基因无效变体有关。这些儿童期/早发病形式呈现有症状的中央暗点和快速进展的萎缩,大多数儿科患者在诊断时就表现出视杆感光细胞受损,这是预后较差的指标。另一个考虑因素是年轻的ABCA4视网膜病变患者可能会出现弱视。弱视是由于早期视觉发育缺陷产生的,会导致视力受损、立体视觉下降和融合受损,应了解,如果在7-9岁的关键时期不予以纠正,则会造成永久性的影响。一些研究表明青少年可能还保留一定的神经可塑性潜力。这使得本发明人预期在关键期结束之前用本文所述的ABCA4外显子编辑器治疗视网膜病变的儿科受试者将具有更大的功效。

[0369] 上述核酸反式剪接分子(例如,核酸反式剪接分子和核酸反式剪接分子编码载体)和组合物可用于在个体的靶细胞(例如,眼细胞、视网膜细胞、光感受器和/或RPE细胞)中表达功能性ABCA4和/或调节ABCA4的表达,例如用于治疗与ABCA4基因突变相关的疾病或病症,例如ABCA4相关的视网膜营养不良,例如斯特格氏病和视锥-视杆细胞营养不良,包括延缓或改善与本文所述的疾病相关的症状的方法中。上述核酸反式剪接分子(例如,核酸反式

剪接分子和核酸反式剪接分子编码载体)和组合物还可用于在个体的靶细胞(例如,眼细胞、视网膜细胞、光感受器和/或RPE细胞)中表达功能性ABCA4和/或调节ABCA4的表达,例如,用于其治疗与ABCA4基因突变相关的疾病或病症,例如ABCA4相关的视网膜营养不良,例如斯特格氏病和视锥-视杆细胞营养不良,包括延迟或改善与本文所述的疾病相关的症状的用途,或其制备供治疗与ABCA4基因突变相关的疾病或病症,例如ABCA4相关的视网膜营养不良,例如斯特格氏病和视锥-视杆细胞营养不良,包括延缓或改善与本文所述的疾病相关的症状的药物的用途。这类方法和用途涉及在将反式剪接分子的CDS剪接到靶ABCA4mRNA前体的条件下使靶ABCA4基因(例如,ABCA4 mRNA前体)与如本文所述的反式剪接分子[例如,5'反式剪接分子,或如本文所述的3'和5'反式剪接分子的混合物,包含其的组合物(例如,药物组合物)或包含其的药物]接触,以用靶向基因的功能性(即健康的)、正常或野生型或纠正的mRNA替换携带一个或多个缺陷或突变的靶向mRNA前体的一部分,从而纠正靶细胞中ABCA4的表达。因此,所述方法和组合物可用于治疗与特定突变和/或基因表达相关的眼部疾病/病变。

[0370] 在一些情况下,本文提供了通过使靶细胞与本文所述的任何核酸反式剪接分子、载体(例如,AAV载体)或组合物接触(例如,转导)来在靶细胞中表达功能性ABCA4的方法。在一个实施方案中,接触涉及将组合物(例如,药物组合物)直接施用于患病个体。在另一实施方案中,接触可以在体外进行,将培养的细胞和处理过的眼部细胞重新植入个体中。在另一实施方案中,所述方法涉及施用携带本文所述的任何5'ABCA4反式剪接分子的rAAV。在又一实施方案中,所述方法涉及施用携带3'反式剪接分子的rAAV与携带5'反式剪接分子的rAAV的混合物。这些方法包括向有需要的个体施用有效浓度的本文所述的任何组合物的组合物。

[0371] 在一些实施方案中,所述方法包括选择一种或多种反式剪接分子来治疗患有与ABCA4突变相关的病症的个体。在一些实施方案中,本文涵盖一种或多种反式剪接分子治疗患有与ABCA4突变相关的病症的个体的用途,或其制备供治疗患有与ABCA4突变相关的病症的个体的药物的用途。这样的方法和用途包括选择一种或多种反式剪接分子来治疗患有与ABCA4突变相关的病症的个体,或使用这样选择的一种或多种反式剪接分子来治疗患有与ABCA4突变相关的病症的个体,或使用这样选择的一种或多种反式剪接分子来制备供治疗患有与ABCA4突变相关的病症的个体的药物。这种选择可以基于个体的基因型。在一些实施方案中,与ABCA4相关的病症可能是常染色体隐性遗传病症。在一些情况下,个体的ABCA4突变是纯合的或复合杂合的。筛选和鉴定ABCA4中的特定突变的方法是本领域已知的。

[0372] 本发明的方法包括根据ABCA4中单个突变(例如个体的一个等位基因的突变)的位置选择单个反式剪接分子。在常染色体隐性突变背景下的一些情况下,仅纠正两个突变中的一个就足以恢复功能性蛋白质活性,例如,其中第二个等位基因在ABCA4基因的相反部分上具有突变,超出了配置为纠正第一个突变的单个AAV递送的反式剪接分子的范围。因此,在一些实施方案中,本发明的方法包括施用单个反式剪接分子来纠正靶基因5'部分上的单个突变,例如,不考虑其它等位基因中突变的位置。

[0373] 此外,本发明提供了涉及选择单个反式剪接分子来纠正ABCA4基因5'部分上的两个或更多个突变的方法,使得能够包装在AAV载体中的单个反式剪接分子能够跨越两个或所有突变,从而纠正两个或所有突变。

[0374] 在其它实施方案中,本文提供了使用两种反式剪接分子(5'反式剪接分子和3'反式剪接分子)纠正ABCA4基因内的多个突变的方法。在一些实施方案中,整个ABCA4基因在两种反式剪接分子结合和反式剪接后被替换,例如,其中5'反式剪接分子和3'反式剪接分子结合相同的靶ABCA4内含子并分别替换靶内含子上游和下游的外显子。

[0375] 在一些实施方案中,作为治疗疾病或病症的方法,将本文提供的组合物施用于个体。在一些实施方案中,个体患有遗传疾病,例如本文所述的ABCA4相关的视网膜病变。在一些实施方案中,个体有患上疾病,例如本文所述的ABCA4相关的视网膜病变的风险。在一些实施方案中,个体患上由蛋白质(例如,ABCA蛋白)量不足或蛋白质(例如,ABCA蛋白)活性不足引起的疾病或病症的风险增加。如果个体患上由蛋白质质量不足或蛋白质活性不足引起的疾病或病症的“风险增加”,则该方法涉及预防性或预防性治疗。例如,由于家族有疾病史,个体患上这种疾病或病症的风险可能会增加。通常,患上这类疾病或病症的风险增加的个体受益于预防性治疗(例如,通过预防或延缓疾病或病症的发作或进展)。

[0376] 携带本文所述的反式剪接分子的重组腺相关病毒的有效浓度范围为每毫升约 $10^8$ 至 $10^{13}$ 个载体基因组(vg/mL)。rAAV感染单位的测量如McLaughlin等人,J.Virol.1988,62:1963中所述。在另一实施方案中,浓度范围在 $10^9$ 至 $10^{13}$ vg/mL之间。在另一实施方案中,有效浓度为约 $1.5 \times 10^{11}$ vg/mL。在另一实施方案中,有效浓度为约 $5 \times 10^{11}$ vg/mL。在一个实施方案中,有效浓度为约 $1.5 \times 10^{10}$ vg/mL。在另一实施方案中,有效浓度为约 $2.8 \times 10^{11}$ vg/mL。在又一实施方案中,有效浓度为约 $1.5 \times 10^{12}$ vg/mL。在另一实施方案中,有效浓度为约 $1.5 \times 10^{13}$ vg/mL。

[0377] 在一些实施方案中,浓度范围在 $4.0 \times 10^{10}$ 与 $1 \times 10^{12}$ vg/眼之间。在一些实施方案中,有效浓度为约 $1.0 \times 10^{11}$ vg/眼。在一些实施方案中,有效浓度为约 $4.3 \times 10^{10}$ vg/眼。

[0378] 希望使用最低有效剂量(递送的总基因组拷贝数)的病毒,以减少不良影响的风险,例如毒性,以及与眼部施用相关的其它问题,例如视网膜发育不良和脱离。携带如本文所述的反式剪接分子的重组腺相关病毒的有效剂量范围为每剂(即每次注射)约 $10^8$ 与 $10^{13}$ 个载体基因组(vg)之间。在一个实施方案中,剂量范围在 $10^9$ 与 $10^{13}$ vg之间。在另一实施方案中,有效剂量为约 $1.5 \times 10^{11}$ vg。在另一实施方案中,有效剂量为约 $5 \times 10^{11}$ vg。在一个实施方案中,有效剂量为约 $1.5 \times 10^{10}$ vg。在另一实施方案中,有效剂量为约 $2.8 \times 10^{11}$ vg。在又一实施方案中,有效剂量为约 $1.5 \times 10^{12}$ vg。在另一实施方案中,有效浓度为约 $1.5 \times 10^{13}$ vg。考虑接受治疗的个体的身体状况,包括个体的年龄、所施用的组合物和特定的病症、靶向的细胞和病症(如果是渐进性的)发展的程度,主治医生还可以选择这些范围内或其它单位的其它剂量。

[0379] 组合物可以以约50 $\mu$ L至约1mL的体积进行递送,包括该范围内的所有数字,取决于待治疗区域的大小、所使用的病毒滴度、施用途径和该方法的期望效果。在一个实施方案中,体积为约50 $\mu$ L。在另一实施方案中,体积为约70 $\mu$ L。在另一实施方案中,体积为约100 $\mu$ L。在另一实施方案中,体积为约125 $\mu$ L。在另一实施方案中,体积为约150 $\mu$ L。在另一实施方案中,体积为约175 $\mu$ L。在又一实施方案中,体积为约200 $\mu$ L。在另一实施方案中,体积为约250 $\mu$ L。在另一实施方案中,体积为约300 $\mu$ L。在另一实施方案中,体积为约350 $\mu$ L。在另一实施方案中,体积为约400 $\mu$ L。在另一实施方案中,体积为约450 $\mu$ L。在另一实施方案中,体积为约500 $\mu$ L。在另一实施方案中,体积为约600 $\mu$ L。在另一实施方案中,体积为约750 $\mu$ L。在另一实

施方案中, 体积为约850 $\mu$ L。在另一实施方案中, 体积为约1,000 $\mu$ L。

[0380] 在一些实施方案中, 包含本文所述的RNA外显子编辑器(例如, 包含SEQ ID NO:90、SEQ ID NO:78或SEQ ID NO:69的RNA外显子编辑器)的组合物可以作为适合注射的悬浮液提供。在一些实施方案中, 本文所述的RNA外显子编辑器(例如, 包含SEQ ID NO:90、SEQ ID NO:78或SEQ ID NO:69的RNA外显子编辑器)可在玻璃体切除术后通过小规格套管以300 $\mu$ L的固定体积进行视网膜下注射来施用。

[0381] 在一个实施方案中, 选择rAAV组合物的体积和浓度, 使得仅影响具有靶细胞的某些解剖区域。在另一实施方案中, rAAV组合物的体积和/或浓度量更大, 以便到达眼睛的更大部分。同样地, 对其它器官的施用剂量也会进行调整。

[0382] 在一些实施方案中, 本文所述的治疗和用途替换了靶细胞中10%或更多的靶ABCA4 mRNA(例如, 靶细胞中11%或更多的靶ABCA4 mRNA、靶细胞中12%或更多的靶ABCA4 mRNA、靶细胞中13%或更多的靶ABCA4 mRNA、靶细胞中14%或更多的靶ABCA4 mRNA、靶细胞中15%或更多的靶ABCA4 mRNA、靶细胞中16%或更多的靶ABCA4 mRNA、靶细胞中17%或更多的靶ABCA4 mRNA、靶细胞中18%或更多的靶ABCA4 mRNA或靶细胞中19%或更多的靶ABCA4 mRNA)。在一些实施方案中, 本文所述的治疗和用途替换了靶细胞中20%或更多的靶ABCA4 mRNA[例如, 靶细胞中21%或更多的靶ABCA4 mRNA、靶细胞中22%或更多的靶ABCA4 mRNA、靶细胞中23%或更多的靶ABCA4 mRNA、靶细胞中24%或更多的靶ABCA4 mRNA、靶细胞中25%或更多的靶ABCA4 mRNA、靶细胞中26%或更多的靶ABCA4 mRNA、靶细胞中27%或更多的靶ABCA4 mRNA、靶细胞中28%或更多的靶ABCA4 mRNA、靶细胞中29%或更多的靶ABCA4 mRNA、靶细胞中30%或更多的靶ABCA4 mRNA、靶细胞中31%或更多的靶ABCA4 mRNA、靶细胞中32%或更多的靶ABCA4 mRNA、靶细胞中33%或更多的靶ABCA4 mRNA、靶细胞中34%或更多的靶ABCA4 mRNA、靶细胞中35%或更多的靶ABCA4 mRNA、靶细胞中36%或更多的靶ABCA4 mRNA、靶细胞中37%或更多的靶ABCA4 mRNA、靶细胞中38%或更多的靶ABCA4 mRNA、靶细胞中39%或更多的靶ABCA4 mRNA、靶细胞中40%或更多的靶ABCA4 mRNA、靶细胞中41%或更多的靶ABCA4 mRNA、靶细胞中42%或更多的靶ABCA4 mRNA、靶细胞中43%或更多的靶ABCA4 mRNA、靶细胞中44%或更多的靶ABCA4 mRNA、靶细胞中45%或更多的靶ABCA4 mRNA、靶细胞中46%或更多的靶ABCA4 mRNA、靶细胞中47%或更多的靶ABCA4 mRNA、靶细胞中48%或更多的靶ABCA4 mRNA、靶细胞中49%或更多的靶ABCA4 mRNA、靶细胞中50%或更多的靶ABCA4 mRNA、靶细胞中51%或更多的靶ABCA4 mRNA、靶细胞中52%或更多的靶ABCA4 mRNA、靶细胞中53%或更多的靶ABCA4 mRNA、靶细胞中54%或更多的靶ABCA4 mRNA、靶细胞中55%或更多的靶ABCA4 mRNA、靶细胞中56%或更多的靶ABCA4 mRNA、靶细胞中57%或更多的靶ABCA4 mRNA、靶细胞中58%或更多的靶ABCA4 mRNA、靶细胞中59%或更多的靶ABCA4 mRNA、靶细胞中60%或更多的靶ABCA4 mRNA、靶细胞中61%或更多的靶ABCA4 mRNA、靶细胞中62%或更多的靶ABCA4 mRNA、靶细胞中63%或更多的靶ABCA4 mRNA、靶细胞中64%或更多的靶ABCA4 mRNA、靶细胞中65%或更多的靶ABCA4 mRNA、靶细胞中66%或更多的靶ABCA4 mRNA、靶细胞中67%或更多的靶ABCA4 mRNA、靶细胞中68%或更多的靶ABCA4 mRNA、靶细胞中69%或更多的靶ABCA4 mRNA、靶细胞中70%或更多的靶ABCA4 mRNA; 靶细胞中75%或更多的靶ABCA4 mRNA; 靶细胞中80%或更多的靶ABCA4 mRNA; 靶细胞中85%或更多的靶ABCA4 mRNA; 靶细胞中90%或更多的靶ABCA4 mRNA; 靶细胞中95%或更多的靶ABCA4 mRNA]。

[0383] 在一些实施方案中,本文所述的治疗和用途替换了靶细胞中5%或10%或更多的靶ABCA4蛋白[例如,靶细胞中11%或更多的靶ABCA4蛋白、靶细胞中12%或更多的靶ABCA4蛋白、靶细胞中13%或更多的靶ABCA4蛋白、靶细胞中14%或更多的靶ABCA4蛋白、靶细胞中15%或更多的靶ABCA4蛋白、靶细胞中16%或更多的靶ABCA4蛋白、靶细胞中17%或更多的靶ABCA4蛋白、靶细胞中18%或更多的靶ABCA4蛋白或靶细胞中19%或更多的靶ABCA4蛋白]。在一些实施方案中,本文所述的治疗和用途替换了靶细胞中20%或更多的靶ABCA4蛋白[例如,靶细胞中21%或更多的靶ABCA4蛋白、靶细胞中22%或更多的靶ABCA4蛋白、靶细胞中23%或更多的靶ABCA4蛋白、靶细胞中24%或更多的靶ABCA4蛋白、靶细胞中25%或更多的靶ABCA4蛋白、靶细胞中26%或更多的靶ABCA4蛋白、靶细胞中27%或更多的靶ABCA4蛋白、靶细胞中28%或更多的靶ABCA4蛋白、靶细胞中29%或更多的靶ABCA4蛋白、靶细胞中30%或更多的靶ABCA4蛋白、靶细胞中31%或更多的靶ABCA4蛋白、靶细胞中32%或更多的靶ABCA4蛋白、靶细胞中33%或更多的靶ABCA4蛋白、靶细胞中34%或更多的靶ABCA4蛋白、靶细胞中35%或更多的靶ABCA4蛋白、靶细胞中36%或更多的靶ABCA4蛋白、靶细胞中37%或更多的靶ABCA4蛋白、靶细胞中38%或更多的靶ABCA4蛋白、靶细胞中39%或更多的靶ABCA4蛋白、靶细胞中40%或更多的靶ABCA4蛋白、靶细胞中41%或更多的靶ABCA4蛋白、靶细胞中42%或更多的靶ABCA4蛋白、靶细胞中43%或更多的靶ABCA4蛋白、靶细胞中44%或更多的靶ABCA4蛋白、靶细胞中45%或更多的靶ABCA4蛋白、靶细胞中46%或更多的靶ABCA4蛋白、靶细胞中47%或更多的靶ABCA4蛋白、靶细胞中48%或更多的靶ABCA4蛋白、靶细胞中49%或更多的靶ABCA4蛋白、靶细胞中50%或更多的靶ABCA4蛋白、靶细胞中51%或更多的靶ABCA4蛋白、靶细胞中52%或更多的靶ABCA4蛋白、靶细胞中53%或更多的靶ABCA4蛋白、靶细胞中54%或更多的靶ABCA4蛋白、靶细胞中55%或更多的靶ABCA4蛋白、靶细胞中56%或更多的靶ABCA4蛋白、靶细胞中57%或更多的靶ABCA4蛋白、靶细胞中58%或更多的靶ABCA4蛋白、靶细胞中59%或更多的靶ABCA4蛋白、靶细胞中60%或更多的靶ABCA4蛋白、靶细胞中61%或更多的靶ABCA4蛋白、靶细胞中62%或更多的靶ABCA4蛋白、靶细胞中63%或更多的靶ABCA4蛋白、靶细胞中64%或更多的靶ABCA4蛋白、靶细胞中65%或更多的靶ABCA4蛋白、靶细胞中66%或更多的靶ABCA4蛋白、靶细胞中67%或更多的靶ABCA4蛋白、靶细胞中68%或更多的靶ABCA4蛋白、靶细胞中69%或更多的靶ABCA4蛋白、靶细胞中70%或更多的靶ABCA4蛋白;靶细胞中75%或更多的靶ABCA4蛋白;靶细胞中80%或更多的靶ABCA4蛋白;靶细胞中85%或更多的靶ABCA4蛋白;靶细胞中90%或更多的靶ABCA4蛋白;靶细胞中95%或更多的靶ABCA4蛋白]。

[0384] 当提到增加功能性ABCA4蛋白水平时,增加可能是临床上显著的。增加可能是相对于未经治疗的受试者中的功能性ABCA4蛋白水平,或者相对于相似受试者群体中的功能性ABCA4蛋白的量。相对于治疗前的受试者或相似受试者群体,功能性ABCA4蛋白的增加可能至少增加10%。相对于治疗前的受试者或相似受试者群体,功能性ABCA4蛋白的增加可能至少增加20%。相对于治疗前的受试者或相似受试者群体,功能性ABCA4蛋白的增加可能至少增加40%。相对于治疗前的受试者或相似受试者群体,功能性ABCA4蛋白的增加可能至少增加50%。相对于治疗前的受试者或相似受试者群体,功能性ABCA4蛋白的增加可能至少增加80%。相对于治疗前的受试者或相似受试者群体,功能性ABCA4蛋白的增加可能至少增加100%。相对于治疗前的受试者或相似受试者群体,功能性ABCA4蛋白的增加可能至少增加

200%。相对于治疗前的受试者或相似受试者群体,功能性ABCA4蛋白的增加可能至少增加500%。

[0385] 在一些情况下,本文所述的方法和用途减少受试者(例如,患有ABCA4相关的视网膜营养不良的受试者)中的ABCA4相关的脂褐素。在其它情况下,本文所述的方法和用途减少受试者(例如,患有ABCA4相关的视网膜营养不良的受试者)中的A2E积累。

[0386] 对于所描述的每种方法和用途,治疗或用途可用于防止进一步损伤的发生,或挽救患有轻度、中度或晚期疾病的组织。如本文所用,术语“拯救”是指防止疾病进展、防止损伤扩散至未受伤的细胞和/或改善受伤细胞的损伤。

[0387] 因此,在一个实施方案中,在疾病发作之前施用组合物。在另一实施方案中,组合物在症状出现之前施用。在另一实施方案中,组合物在症状出现之后施用。在又一实施方案中,当例如与参考组织相比,少于90%的靶细胞发挥功能或剩余时,施用组合物。在又一实施方案中,当例如与参考组织相比,超过10%的靶细胞发挥功能或剩余时,施用组合物。在又一实施方案中,当超过20%的靶细胞发挥功能或剩余时,施用组合物。在又一实施方案中,当超过30%的靶细胞发挥功能或剩余时,施用组合物。在又一实施方案中,当超过40%的靶细胞发挥功能或剩余时,施用组合物。在又一实施方案中,当超过50%的靶细胞发挥功能或剩余时,施用组合物。在又一实施方案中,当超过60%的靶细胞发挥功能或剩余时,施用组合物。在又一实施方案中,当超过70%的靶细胞发挥功能或剩余时,施用组合物。在又一实施方案中,当超过80%的靶细胞发挥功能或剩余时,施用组合物。在又一实施方案中,当超过90%的靶细胞发挥功能或剩余时,施用组合物。在又一实施方案中,当超过95%的靶细胞发挥功能或剩余时,施用组合物。

[0388] 在又一实施方案中,上述任何方法或用途都与另一种疗法或次要疗法组合进行。疗法可以是任何现在已知的或尚未已知的疗法,其有助于预防、阻止或改善这些突变或缺陷或任何与之相关的影响。次要疗法可以在施用上述药物组合物之前、同时或之后施用。在一个实施方案中,次要治疗涉及维持视网膜细胞健康的非特异性方法,例如施用神经营养因子、抗氧化剂和/或抗凋亡剂。非特异性方法是通过注射蛋白质、重组DNA、重组病毒载体、干细胞、胎儿组织或基因修饰细胞实现的。后者可能包括封装的基因修饰细胞。

[0389] 在另一实施方案中,所述方法包括进行功能和成像研究以确定治疗效果。这些研究包括ERG和体内视网膜成像,如美国专利No.8,147,823和国际专利公布No.WO 2014/011210或WO 2014/124282中所述,这些文献以引用的方式并入本文中。此外,还可以进行视野研究、视野检查(例如动态视野检查或全视野静态视野检查)和微视野检查、裂隙灯检查、眼压测量、散瞳眼底镜检查、活动性测试、视力和/或色觉测试。可以在施用如本文所述的核酸反式剪接分子和由其编码的RNA外显子编辑器或包含它们的组合物(例如,药物组合物)之后的时间点测量功效,以确定例如与未治疗的受试者相比,治疗的受试者的ABCA4相关的视网膜病变的症状是否减轻。在一些实施方案中,可以通过例如视力和/或色觉的改善或脂褐素(包括A2E积累)的减少来测量功效。

[0390] 在某些实施方案中,需要进行非侵入性视网膜成像和功能研究,以鉴定疗法所靶向的保留光感受器区域。在这些实施方案中,采用临床诊断测试来确定一次或多次视网膜下注射的精确位置。这些测试可能包括ERG、视野检查、视网膜层的地形绘图和借助于共焦扫描激光检眼镜(cSLO)和光学相干断层扫描(OCT)测量其层的厚度、通过自适应光学(AO)

对视锥密度进行地形绘图、功能性眼科检查等。鉴于成像和功能研究,在一些实施方案中,在同一只眼睛中进行一次或多次注射,以靶向保留光感受器的不同区域。

[0391] 为了在这些方法中使用,每次注射的体积和病毒滴度是单独确定的,并且可以与在同一只眼睛或对侧眼睛中进行的其它注射相同或不同。在另一实施方案中,进行单次较大体积的注射以治疗整个眼睛。剂量、施用和方案可由主治医生根据本公开的教导来确定。

[0392] 以下实施例不限制本文所述的实施方案的范围。本领域技术人员将会意识到,可以在不超出本发明的精神或范围的情况下对以下实施例进行修改。

[0393] 实施例

[0394] 以下实施例提供了生成上述组合物和使用上述组合物的非限制性方法。

[0395] 实施例1.培养的人细胞中的反式剪接效率

[0396] 使用基于NGS的高通量测定法在培养的人细胞中测试ABCA4反式剪接分子,其中测试每个反式剪接分子的各种排列。结果示于图1中。本实施例中的每个ABCA4反式剪接分子包括5'调节结构域(包括CMV启动子和天然5'UTR)和功能性ABCA4 CDS。跨反式剪接分子变化的结构域为接头结构域、结合结构域(BD)、3'DS和终止子。在本实施例中,反式剪接分子作为质粒转染到细胞中,并根据已知的RNA-seq方法通过唯一分子标识符(UMI)进行检测。反式剪接效率读数表明某些结构域在许多或所有排列中显示出改良的反式剪接。例如,此处,相对于SEQ ID NO:29和31的较短接头,SEQ ID NO:27的40聚体接头意外地表现出最高的反式剪接效率。此外,具有SEQ ID NO:18的结合结构域在所有测试的结合结构域中表现出最高的反式剪接效率。还值得注意的是,接头结构域、BD和剪接位点(SS)的某些组合赋予令人惊讶的相对反式剪接效率水平。

[0397] 实施例2.剪接位点评估

[0398] 使用下一代测序对具有总共550个接头序列的ABCA4核酸反式剪接分子文库中的剪接位点进行测试。测试的五个剪接位点分别是CAAAGT、GTAAGT、GTAAGC、GTAAGG和GTAAGT。由此编码的RNA序列在本文中分别称为GUAAGU、GUAAGG、GUAAGC和GUAACU。图2以任意单位显示反式剪接活性(外显子编辑活性)。这些结果表明,在测试的五个剪接位点中,包含GTAAGT的剪接位点赋予最高的反式剪接活性。因此,本文所述的示例性剪接结构域包含GUAAGT,其由GTAAGT编码。

[0399] 实施例3.培养的人细胞中的ABCA4反式剪接

[0400] 将培养的人细胞(HEK-293T)以MOI=100,000用编码各种ABCA4反式剪接分子(RNA外显子编辑器)的AAV转导,并孵育三天,然后收获细胞。AAV-56是编码具有SEQ ID NO:40的ABCA4反式剪接分子的AAV载体。AAV-274是编码具有SEQ ID NO:41的ABCA4反式剪接分子的AAV载体。AAV-443是编码具有SEQ ID NO:42的ABCA4反式剪接分子的AAV载体。AAV-497是编码具有SEQ ID NO:43的ABCA4反式剪接分子的AAV载体。AAV-505是编码具有SEQ ID NO:44的ABCA4反式剪接分子的AAV载体。

[0401] 使用常规方法提取膜蛋白。简单地说,将透化缓冲液和蛋白酶/磷酸酶抑制剂加入细胞中,涡旋,在4℃下孵育10分钟。将透化细胞离心,并将团块再悬浮于含有蛋白酶/磷酸酶抑制剂的溶解缓冲液中。样品在4℃下孵育30分钟,然后离心。将含有膜蛋白的上清液转移并存储于-80℃。

[0402] 通过常规方法提取RNA并通过RT-qPCR进行定量。将在靶ABCA4反式剪接百分比相

对于反式剪接分子拷贝数归一化。图3A显示了RNA结果,显示每个反式剪接分子(外显子编辑)替换的ABCA4mRNA的百分比。图3B显示了蛋白质印迹,使用抗ABCA4 C端抗体,并使用抗NaK/ATP酶抗体作为上样对照。

[0403] AAV-274、AAV-443、AAV-497和AAV-505每一者均表现出相对于AAV-56改良的反式剪接效率(相对于23.8%,分别为37.3%、38.7%、28.5%和28.5%)。与AAV-56相比,AAV-274和AAV-443每一者均表现出明显更高的在靶拷贝数。

[0404] 使用ABCA4敲入细胞作为阳性对照,未处理的HEK293T细胞作为阴性对照,进行类似的测定以比较AAV-443-2(SEQ ID NO:45)与AAV-652(SEQ ID NO:46)之间的反式剪接效率。图3C显示了蛋白质印迹,显示AAV-443-2和AAV-652每一者的重复样品中的蛋白质表达。结果在图3D中以变化倍数给出。值得注意的是,在ABCA4敲除细胞中观察到ABCA4蛋白水平的完全恢复(图3D)。

[0405] 实施例4. 对非人灵长类动物中的ABCA4反式剪接分子的评估

[0406] ABCA4的隐性遗传突变是导致渐进性失明,包括斯特格氏病1和视锥-视杆细胞营养不良3的原因。ABCA4的6882bp编码序列太大,无法通过单个AAV载体完整递送。此外,由于在整个基因中发现了数百种致病突变,所以单一的碱基编辑方法无法解决大量患者的问题。大规模外显子编辑解决方案旨在递送能够反式剪接到内源ABCA4 mRNA前体中的治疗性RNA构建体,从而引入功能性外显子,纠正位于通过反式剪接递送的功能性外显子中的大量突变。事实上,使用单个基于AAV的构建体进行RNA反式剪接可以解决所有患者突变的大约60%。本实施例报告了AAV-ABCA4外显子编辑器在健康非洲绿猴(AGM)非人灵长类动物(NHP)中视网膜下注射后在体内的编辑效率和耐受性。

[0407] 外显子编辑器443(SEQ ID NO:47)、274(SEQ ID NO:48)、497(SEQ ID NO:49)和505(SEQ ID NO:50)包装在AAV8载体中。使用qPCR对共有调节元件进行滴度测定,使用Stunner测量完整衣壳和空衣壳以及测量内毒素水平来表征AAV8载体制剂。结果示于下表2。

[0408] 表2.AAV表征

| [0409] | 外显子编辑器 | 滴度<br>(vg/mL)      | 完整衣壳% | 内毒素(EU/mL) |
|--------|--------|--------------------|-------|------------|
|        | 443    | $1 \times 10^{12}$ | 41.5  | <0.5       |
|        | 274    | $1 \times 10^{12}$ | 46.0  | <0.5       |
|        | 497    | $1 \times 10^{12}$ | 43.0  | <0.5       |
|        | 505    | $1 \times 10^{12}$ | 23.9  | <0.5       |

[0410] 检测每种制剂的VP1、VP2和VP3蛋白,并在SDS凝胶上以 $5 \times 10$ vg/孔目测(图4)。

[0411] 将含有SEQ ID NO:15的5'调节结构域(即含有CMV启动子和天然5'ABCA4非翻译区)的GFP编码AAV载体在非洲绿猴(AGM)非人灵长类动物(NHP)中视网膜下施用。GFP蛋白表达示于图5A-5C中。图5A是显示GFP蛋白的实时图像。图5B是显示第28天阴性对照中的视网膜切片(光感受器和脉络膜被标记)的抗GFP免疫组织化学(IHC)染色的显微照片。图5C是显示第28天经过GFP外显子编辑器处理的动物的视网膜切片(光感受器和脉络膜被标记)的抗GFP IHC染色的显微照片。这些结果表明,5'调节结构域赋予体内光感受器强大的蛋白质表达。

[0412] ABCA4外显子编辑器被配制成AAV8,并在AGM NHP中以每只眼睛两个水泡的形式进行视网膜下施用。每次注射100μl,含1×10<sup>11</sup>vg。图6显示了基线、视网膜下注射后、第14天和第28天的cSLO和OCT图像。注射后观察到上部和下部水泡。到第28天时,观察到水泡区域有轻微的‘阴影’。在第28天,使用彩色眼底成像对媒介物、外显子编辑器443和反式剪接分子274进行临床评分(图7)。媒介物、外显子编辑器443和反式剪接分子274的总临床评分分别为17、0和1。

[0413] 使用定量RT-PCR定量RNA并使用蛋白质印迹定量ABCA4蛋白,评估AGM NHP视网膜中ABCA4外显子编辑器的外显子编辑。为了测量反式剪接的ABCA4 mRNA,使用定量RT-PCR检测人与AGM序列之间的嵌合mRNA接合,如图8A所示。对于每个ABCA4外显子编辑器,反式剪接的mRNA被定量为每10 ng RNA的编辑拷贝数(图8B和表4)。

[0414] 表4. 编辑拷贝数/10ngRNA

[0415]

| 外显子编辑器 | 平均值       | 标准差       |
|--------|-----------|-----------|
| (媒介物)  | 0         | 0         |
| 443    | 464880.29 | 464880.29 |
| 274    | 45380.29  | 45380.29  |
| 497    | 328749.88 | 328749.88 |
| 505    | 30594.16  | 30594.16  |

[0416] 对于每个ABCA4外显子编辑器,通过标签的蛋白质印迹法对视网膜的膜提取物中标记的全长反式剪接的ABCA4蛋白进行定量。图9显示了示例性蛋白质印迹,其显示外显子编辑器443、274、497和505的全长ABCA4表达。

[0417] 为了确认表达的蛋白质是嵌合人/AGM(因此是反式剪接产物),进行质谱实验。通过免疫亲和富集与MRM-MS结合,鉴定出区分完全AGM NHP ABCA4和嵌合的编辑ABCA4的独特肽。通过加入标准品来确认肽的身份。在媒介物处理的视网膜中(图10A),仅观察到NHP肽。在外显子编辑器处理的视网膜中(图10B),检测到了人和NHP肽。该结果证实了全长外显子编辑的mRNA被翻译。

[0418] 在整个研究过程中对眼部健康状况进行监测。总体来说,通过视网膜厚度(图11A)、眼内压(图11B)、总临床评分(图11C)、玻璃体细胞评分(图11D)、房水细胞评分(图11E)和炎性角膜颗粒(图11F)测量,使用AAV8视网膜下递送的外显子编辑器处理后眼部健康得以维持。在整个研究期间,视网膜厚度保持正常(图11A)。眼压计显示,媒介物处理的眼睛在第7天眼内压降低,并在之后的时间点保持稳定(图11B)。外显子编辑器处理的眼睛的眼内压保持在健康范围内。通过裂隙灯生物显微镜和眼底镜检查评估眼部炎症(图11C-11F)。用外显子编辑器处理的个别眼睛显示出一些轻度至中度炎症的迹象,但对于某个编辑器或AAV制剂来说并不一致,这表明这些观察结果是由于个别眼睛的程序造成的。

[0419] 研究中的所有动物都保持良好的健康状况。体重(图12A)和心率(图12B)测量未发现任何指示外显子编辑器给药引起的不良全身影响的临床症状。每天进行笼边观察并监测食物消耗量,未发现任何变化。

[0420] 本文描述的研究构成了NHP视网膜中临床相关效率的第一份反式剪接报告。外显子编辑和替换在非洲绿猴非人灵长类动物中在视网膜下注射后在体内表现出功效和耐受性。所有视网膜下递送的AAV介导的外显子编辑分子均具有良好的耐受性。这些研究强调了

RNA外显子编辑治疗ABCA4相关的视网膜病变的潜力。

[0421] 实施例5. 体外评估示例性ABCA4反式剪接分子

[0422] 转染至ABCA4蛋白KO细胞系后示例性ABCA4 RNA外显子编辑器的舌性

[0423] 如上所述,对基于HEK293T的细胞系(称为B6)进行工程化,其表达高ABCA4RNA和蛋白水平。这是通过CRISPR介导的同源性定向修复(HDR)将组成型CAGGS启动子敲入(KI)到内源ABCA4启动子与外显子1之间实现的。

[0424] 为了创建一个模型系统来表征ABCA4 RNA外显子编辑器纠正靶向的天然mRNA前体、生成反式剪接的在靶(ONT)RNA以及编码全长ABCA4蛋白的能力,基于B6细胞系,工程化一个蛋白质敲除(KO)细胞系(称为17+06)。17+06细胞继续产生高水平的ABCA4RNA,但与B6细胞系截然不同,17+06细胞不产生ABCA4蛋白。通过使用CRISPR和靶向外显子3和4的引导RNA(gRNA)将突变引入ABCA4基因编码序列,实现了对17+06细胞的工程化。RT-qPCR证实了17+06细胞系中ABCA4 RNA表达的持续存在,而蛋白质印迹分析表明ABCA4蛋白质翻被消除。参见图3C。

[0425] 将ABCA4 KO(17+06)细胞用几种ABCA4 RNA外显子编辑器构建体转染,以评估反式剪接效率和所引起的ONT反式剪接的ABCA4蛋白的表达。构建体是相同的,除了指示为区分的元件,每一者反映在图19A和19B中4个泳道的标记上。包含图13中描绘的示例性ABCA4 RNA外显子编辑器的示例性ABCA4 RNA外显子编辑器构建体(5'UTR至MALAT1-MascRNA(抗mut1;SEQ ID NO:78)的末端)在该分析中脱颖而出,成为主要候选者,因为它驱动高水平的反式剪接(ABCA4 RNA替换约20%)和高水平的ONT反式剪接的ABCA4蛋白,同时相对于所测试的其它构建体降低NSP水平。图19A和19B的泳道4显示了示例性ABCA4RNA外显子编辑器的活性。如图19A和19B所示,示例性ABCA4 RNA外显子编辑器包含本文详细描述40聚体接头+3×UBS+ARE元件。

[0426] 除上述内容外,图19A和19B示出了在示例性ABCA4 5'构建体的背景下每个所示RNA外显子编辑器(RTM)元件单独和组合的活性。图19A显示了一个直方图,描绘每个所示构建体的ABCA4 RNA替换百分比。每个构建体的替换百分比范围约为20-30%。有趣的是,相对于单独的40聚体(SEQ ID NO:27),40聚体+3x重复U1结合位点(3×UUBS)的组合(SEQ ID NO:72)显示出ABCA4 RNA替换百分比增加的趋势。图19B呈现了蛋白质印迹,显示包含40聚体(SEQ ID NO:27)、40聚体+3×UBS(SEQ ID NO:72)、40聚体+富AU元件(SEQ ID NO:86)的序列3444-3529和40聚体+3×UBS+ARE(SEQ ID NO:73)的所示构建体表达反式剪接的ABCA4蛋白(ONT),并且这些RTM元件组合在一起表现出相对于单独的40聚体来说ONT水平增加的趋势。蛋白质印迹还显示,相对于单独的40聚体,40聚体+3×UBS、40聚体+ARE和40聚体+3×UBS+ARE的每个组合都对非剪接蛋白质(NSP)的减少表现出组合效应。相对于单独的40聚体(SEQ ID NO:27),40聚体+3×UBS+ARE(SEQ ID NO:73)对NSP减少的组合效应尤为明显。在最小NSP背景下(含有终止密码子,紧跟外显子编辑器中切片供体序列之后),两种独立降低NSP水平的策略的组合——包括3×UBS和富AU元件——对降低NSP水平显示出协同效应。与仅40聚体相比,这种组合将NSP水平降低了约70%,降至无终止的外显子编辑器中看到的水平。

[0427] 值得注意的是,相对于仅包含40聚体、包含40聚体+3×UBS或包含40聚体+ARE的RNA外显子编辑器,40聚体+3×UBS+ARE(SEQ ID NO:73)的盒使NSP水平的降低更为显著。因

此,该元件组合(40聚体+3×UBS+ARE;SEQ ID NO:73)在如本文所述的反式剪接分子的反式剪接分子背景下表现出特别有利的特性。

[0428] 在ABCA4 RNA外显子编辑器的背景下呈现的结果在靶向除ABCA4以外的基因的靶mRNA前体中的不同内含子的其它RNA外显子编辑器的背景下进行了重现。参见例如图27和28,其在下文中有更详细的描述,其中描绘了40聚体+3×UBS(SEQ ID NO:72)、40聚体+ARE(SEQ ID NO:86的序列3444-3529)和40聚体+3×UBS+ARE(SEQ ID NO:73)每一者与单独的40聚体(SEQ ID NO:27)相比降低NSP水平的能力。在不同RNA外显子编辑器背景下40聚体+3×UBS(SEQ ID NO:72)、40聚体+ARE(SEQ ID NO:86的序列3444-3529)和40聚体+3×UBS+ARE(SEQ ID NO:73)的排列和定位与本文所述并在图13中描绘的ABCA4外显子编辑器的排列和定位一致。

[0429] 如图27B所示,通过在靶向内含子A的RNA外显子编辑器中包括3×UBS和ARE,以组合方式减少NSP,图中描绘了结果。靶向内含子A的RNA外显子编辑器靶向不同于ABCA4的mRNA前体中的内含子。简单地说,将HEK293细胞用靶向内含子A的RNA外显子编辑器转染,所述编辑器在其接头区域有所变化以包括所指示的RNA外显子编辑器元件(40聚体、40聚体+3×UBS、40聚体+ARE和40聚体+3×UBS+ARE)。转染后48小时收获细胞,通过RT-qPCR测定反式剪接效率(图27A)或进行蛋白质印迹分析(图28B)。在靶向内含子A的外显子编辑器中,本发明人表明,相对于单独的40聚体,40聚体+3×UBS使NSP水平降低了约75%,相对于单独的40聚体,40聚体+ARE使NSP水平降低了约40%,并且相对于仅40聚体,40聚体+3×UBS+ARE使NSP水平降低了88%(图27B)。

[0430] 如图28B所示,通过在靶向内含子B的RNA外显子编辑器中包括3×UBS和ARE,以组合方式减少NSP,图中描绘了结果。靶向内含子B的RNA外显子编辑器靶向不同于ABCA4和不同于靶向内含子A的RNA外显子编辑器靶向的mRNA前体的mRNA前体中的内含子。简单地说,将HEK293细胞用靶向内含子B的RNA外显子编辑器转染,所述编辑器在其接头区域有所变化以包括所指示的RNA外显子编辑器元件(40聚体、40聚体+3×UBS、40聚体+ARE和40聚体+3×UBS+ARE)。转染后48小时收获细胞,通过RT-qPCR测定反式剪接效率(图28A)或进行蛋白质印迹分析(图28B)。在靶向内含子B的外显子编辑器中,本发明人表明,相对于单独的40聚体,40聚体+3×UBS使NSP水平降低了约38%,相对于单独的40聚体,40聚体+ARE使NSP水平降低了约33%,并且相对于仅40聚体,40聚体+3×UBS+ARE使NSP水平降低了66%(图28B)。

[0431] ABCA4 KO细胞系转导后示例性ABCA4 RNA外显子编辑器的剂量反应评估

[0432] 为了证明对ABCA4 RNA外显子编辑器构建体药理学和作用机制的信心,本发明人将包含图13中所描绘的示例性ABCA4 RNA外显子编辑器的示例性ABCA4 RNA外显子编辑器构建体(ABCA4-01)(SEQ ID NO:90,其包含SEQ ID NO:78)包装到替代AAV2衣壳中,并评估在ABCA4 KO细胞系中的剂量反应。将ABCA4 KO细胞用 $1 \times 10^3$ 至 $1 \times 10^6$  ( $1e3$ - $1e6$ )增加MOI的AAV2-ABCA4 RNA外显子编辑器构建体进行转导,该构建体包含图13中所描绘的示例性ABCA4 RNA外显子编辑器(SEQ ID NO:90,其包含SEQ ID NO:78),然后进行膜蛋白分离并用ABCA4特异性Ab进行蛋白质印迹。包含图13中所描绘的示例性ABCA4 RNA外显子编辑器的AAV2-ABCA4 RNA外显子编辑器构建体(SEQ ID NO:90)的MOI增加引起外显子编辑水平增加,这反映在在靶(ONT)ABCA4蛋白表达和RNA替换%上。参见图21A和21B。

[0433] 实施例6.脱靶(OFT)发现下一代测序(NGS)方法(OFT-seq)

[0434] OFT RNA测序 (OFT-seq) 是本发明人开发的一种全局脱靶发现方法。该方法利用从细胞、NHP神经视网膜样品或人视网膜外植体中提取的RNA,这些RNA要么用质粒转染,要么用编码ABCA4外显子编辑器或相关对照的AAV转导。在外显子22的3'上游的外显子编辑器特定的密码子优化CDS用作靶向引物(基因特异性引物:GSP)区域。由于该方法仅靶向外显子编辑器的上游序列,因此其可用于捕获无偏的全局脱靶非ABCA4RNA序列。因此,该方法允许捕获未剪接的外显子编辑器、自剪接的外显子编辑器以及预期的ONT编辑的ABCA4RNA和OFT(图25)。然后使用Illumina短读NGS平台对这些序列进行测序。输出的NGS读数由内部开发的专有计算生物信息学分析工具进行处理。该分析提供了总体测序质量和靶接合序列以及每个变异序列的定量信息。

[0435] 简单地说,首先用脱氧核糖核酸酶处理RNA样品,以消除残留的基因组DNA、质粒DNA/AAVDNA和可能的污染DNA。然后通过核糖体RNA(rRNA)消耗或PolyA富集来富集RNA。然后将富集的RNA样品分解成碎片。在Illumina接头连接及cDNA制备后,利用靶向GSP的ABCA4密码子优化序列及靶向接头的引物进行PCR。该步骤扩增带有ABCA4密码子优化序列的靶序列-包括外显子编辑器、ONT、OFT和自剪接的外显子编辑器。

[0436] 实施例7.生成抗隐蔽剪接位点ABCA4 CDS

[0437] 本发明人通过体外和体内(NHP)的大量实验和计算分析结果发现,在本文所述的ABCA4外显子编辑器的背景下使用隐蔽剪接位点。本发明人进行了大量实验和分析,包括生成脱靶(OFT)文库和其生物信息学分析,并且意外地发现在本文所述的反式剪接分子的背景下包含外显子1-22的ABCA4编码序列(CDS)内发生隐蔽剪接的大量位置。为了提高本文所述的反式剪接分子的反式剪接效率和保真度,通过在其中引入同义突变来修饰多个隐蔽剪接位点的序列以损害和/或使隐蔽剪接位点失效。这种方法利用了遗传密码的简并性,允许引入所需的修饰以减少隐蔽剪接位点的使用,而不会导致编码氨基酸发生改变。

[0438] 除上述内容外,使用下一代测序(NGS)方法,本发明人鉴定隐蔽剪接位点并定量包含外显子1-22的ABCA4 CDS内每个不同位置的隐蔽剪接使用频率,从而建立了数据驱动的、优先减少其中隐蔽剪接使用的方案。参见表3和图26。如表3所示,所鉴定的隐蔽剪接位点按使用频率顺序排列,使用频率由引入同义突变之前每百万映射合并读数的平均计数(CPM)决定。因此,结果和其生物信息学分析建立了ABCA4外显子1-22中的隐蔽剪接位点的优先列表,这些位点可以被修饰以生成在反式剪接分子背景下具有改良的特性和功能性(例如抗隐蔽剪接位点ABCA4 CDS)的ABCA4 CDS。

[0439] 还实施了NGS方法来评估引入同义突变后的隐蔽剪接使用频率。参见表3和图26。通过这种方法确认隐蔽剪接位点的减少将序列修饰(同义突变)的结构变化与由此赋予的功能联系起来,即生成抗隐蔽剪接位点ABCA4 CDS。图26通过显示在已鉴定的隐蔽剪接位点引入同义突变之前(上图)和在已鉴定的隐蔽剪接位点引入同义突变之后(下图;抗剪接位点序列)的隐蔽剪接位点裂解频率说明了在示例性ABCA4外显子编辑器中通过在已鉴定的隐蔽剪接位点引入同义突变而赋予的功能。图26表明,在隐蔽剪接位点减少之后,观察到剪接频率显著降低。

[0440] 表3:列出在ABCA4编码结构域序列(CDS)中鉴定的隐蔽剪接位点,以及在剪接位点减少之前和之后鉴定的每个剪接位点的剪接频率。SEQ ID NO:46表示引入同义突变之前的示例性序列;SEQ ID NO:79表示引入同义突变之后的示例性序列。rtm\_pos列中指示的核苷

酸位置是指第2列或第3列中指示的最后一个核苷酸相对于ATG密码子(翻译起始位点)的G的位置。缩写“没有”用于表示在所示序列(SEQ ID NO:79)中没有发生核苷酸变化;分支点的缩写为“bp”。

[0441]

| rtm_pos | 剪接位点减少之前的 ABCA4 CDS                              | 剪接位点减少之后的 ABCA4 CDS                                 | 纠正   | 剪接位点减少之前的 ABCA4 CDS_avg_cpm | 剪接位点减少之后的 ABCA4 CDS_avg_cpm | 注释            |
|---------|--------------------------------------------------|-----------------------------------------------------|------|-----------------------------|-----------------------------|---------------|
| 3086    | ACATGCTG<br>TTCTATGCC<br>CAG (SEQ<br>ID NO: 133) | ACATGCT<br>GTTCTAT<br>GCACAA<br>(SEQ ID<br>NO: 92)  | 剪接位点 | 1480.814<br>745             | 1.742685<br>513             | ABCA4<br>ex21 |
| 2931    | CACCTTGT<br>CCATCCTC<br>ACAG (SEQ<br>ID NO: 134) | CACCTTG<br>TCCATCC<br>TCACTG<br>(SEQ ID<br>NO: 94)  | 剪接位点 | 256.3165<br>696             | 0                           | ABCA4<br>ex20 |
| 2576    | TCGCTTGG<br>TACCTTGAT<br>CAG (SEQ<br>ID NO: 135) | TCGCTTG<br>GTACCTT<br>GATCAA<br>(SEQ ID<br>NO: 96)  | 剪接位点 | 142.1908<br>763             | 0.961889<br>9213            | ABCA4<br>ex16 |
| 3029    | GCCTTGGC<br>ATGTGTCC<br>ACAG (SEQ<br>ID NO: 136) | GCCTTGG<br>CATGTGT<br>CCACAG<br>(SEQ ID<br>NO: 97)  | 没有   | 87.63236<br>171             | 183.2327<br>206             | ABCA4<br>ex20 |
| 127     | CTGGTCTT<br>GATCTGGT<br>TAAG (SEQ<br>ID NO: 137) | CTGGTCT<br>TGATCTG<br>GTTACG<br>(SEQ ID<br>NO: 100) | 剪接位点 | 48.27688<br>581             | 0                           | ABCA4<br>ex1  |

[0442]

|      |                                                   |                                                     |                   |                 |                  |               |
|------|---------------------------------------------------|-----------------------------------------------------|-------------------|-----------------|------------------|---------------|
| 2348 | TCCTGTGC<br>TTCGCCTG<br>GCAG (SEQ<br>ID NO: 138)  | TCCTGTG<br>CTTCGCC<br>TGGCAA<br>(SEQ ID<br>NO: 102) | 剪接位<br>点          | 46.36827<br>31  | 0                | ABCA4<br>ex15 |
| 1835 | GCGGGTTT<br>GCCTATCTG<br>CAG (SEQ<br>ID NO: 139)  | GCGGGTT<br>TGCATAT<br>CTGCAA<br>(SEQ ID<br>NO: 104) | 剪接位<br>点, 分<br>支点 | 40.54486<br>027 | 0                | ABCA4<br>ex13 |
| 1218 | CACTCCTG<br>ATTACCT<br>GCAG (SEQ<br>ID NO: 140)   | CACTCCT<br>GATTCAC<br>CTGCTG<br>(SEQ ID<br>NO: 106) | 剪接位<br>点          | 34.90671<br>1   | 0.961889<br>9213 | ABCA4<br>ex9  |
| 3083 | AGCACATG<br>CTGTTCTAT<br>GCC (SEQ ID<br>NO: 141)  | AGCACAT<br>GCTGTTC<br>TATGCA<br>(SEQ ID<br>NO: 108) | 没有                | 29.04542<br>378 | 0                | ABCA4<br>ex21 |
| 2110 | TGTACCTG<br>GTTTCCTGG<br>ACAG (SEQ<br>ID NO: 142) | TGTACCT<br>GGTTCCT<br>GGACTC<br>(SEQ ID<br>NO: 110) | 剪接位<br>点          | 27.63290<br>145 | 0.871342<br>7566 | ABCA4<br>ex14 |
| 2628 | TTGGTACTT<br>TCTTCTACA<br>AG (SEQ ID<br>NO: 143)  | TTGGTAC<br>TTTCTTCT<br>ACAAG<br>(SEQ ID<br>NO: 111) | 没有                | 23.56865<br>795 | 25.48416<br>316  | ABCA4<br>ex17 |
| 3103 | CAGCTGAA<br>AGGAAAGT<br>CCCA (SEQ<br>ID NO: 144)  | CAACTGA<br>AAGGAA<br>AGTCCCA<br>(SEQ ID<br>NO: 114) | 没有                | 20.91693<br>678 | 26.29908<br>764  | ABCA4<br>ex21 |
| 3093 | GTTCTATGC<br>CCAGCTGA<br>AAG (SEQ<br>ID NO: 145)  | GTTCTAT<br>GCACAAC<br>TGAAAG<br>(SEQ ID<br>NO: 116) | 没有                | 20.02438<br>495 | 64.52533<br>239  | ABCA4<br>ex21 |

[0443]

|      |                                                  |                                                     |          |                 |                 |              |
|------|--------------------------------------------------|-----------------------------------------------------|----------|-----------------|-----------------|--------------|
| 701  | CCCTGTGC<br>TCCCTCTCC<br>CAG (SEQ<br>ID NO: 146) | CCCTGTG<br>CTCCCTC<br>TCCCAG<br>(SEQ ID<br>NO: 117) | 没有       | 18.78518<br>611 | 15.44665<br>702 | ABCA4<br>ex6 |
| 528  | CATCGGCC<br>TGTCTGAC<br>TCAG (SEQ<br>ID NO: 147) | CATCGGC<br>CTGTCTG<br>ACTCTG<br>(SEQ ID<br>NO: 120) | 剪接位<br>点 | 15.56367<br>999 | 0               | ABCA4<br>ex5 |
| 1451 | TCCTAAAC<br>TTCCTCTAC<br>AAG (SEQ<br>ID NO: 148) | TCCTAAA<br>CTTCCTC<br>TACAAG<br>(SEQ ID<br>NO: 121) | 没有       | 14.49780<br>466 | 0               |              |
| 2015 | CTGTCTCC<br>ATGACTGT<br>GAAG (SEQ<br>ID NO: 149) | CTGTCTC<br>CATGACT<br>GTGAAG<br>(SEQ ID<br>NO: 123) | 没有       | 12.29235<br>941 | 12.23292<br>748 |              |
| 3179 | AGCGGAAT<br>GAAGAGGC<br>TCAG (SEQ<br>ID NO: 150) | AGCGGAA<br>TGAAGAG<br>GCTCAG<br>(SEQ ID<br>NO: 125) | 没有       | 11.33805<br>306 | 9.166163<br>39  |              |
| 2254 | ATGCTGTG<br>CTTTCTGCT<br>CAG (SEQ<br>ID NO: 151) | ATGCTGT<br>GCTTTCT<br>GCTCAG<br>(SEQ ID<br>NO: 126) | 没有       | 9.194362<br>328 | 44.59728<br>553 |              |
| 1124 | TTTGTAATG<br>CATTGATCC<br>AG (SEQ ID<br>NO: 152) | TTTGTA<br>TGCATTG<br>ATCCAG<br>(SEQ ID<br>NO: 128)  | 没有       | 8.574762<br>912 | 14.84695<br>575 |              |
| 1907 | TTGGAATC<br>TACCTCCA<br>GCAG (SEQ<br>ID NO: 153) | TTGGAAT<br>CTACCTC<br>CAGCAG<br>(SEQ ID<br>NO: 130) | 没有       | 8.178301<br>454 | 24.70336<br>757 |              |

[0444]

|      |                                                  |                                                     |    |                 |                  |  |
|------|--------------------------------------------------|-----------------------------------------------------|----|-----------------|------------------|--|
| 1090 | CCTATCTAT<br>TCTTATGAC<br>AG (SEQ ID<br>NO: 154) | CCTATCTA<br>TTCTTATG<br>ACAG<br>(SEQ ID<br>NO: 671) | 没有 | 6.939102<br>621 | 22.96068<br>206  |  |
| 555  | CCTTCTGAT<br>CAACTCTC<br>AAG (SEQ<br>ID NO: 155) | CCTTCTG<br>ATCAACT<br>CTCAAG<br>(SEQ ID<br>NO: 672) | 没有 | 6.542641<br>163 | 27.31739<br>584  |  |
| 2801 | GGGTATGC<br>GTGAAGAA<br>TCTG (SEQ<br>ID NO: 156) | GGGTGTG<br>CGTGAAG<br>AATCTG<br>(SEQ ID<br>NO: 673) | 没有 | 6.146179<br>706 | 0                |  |
| 3104 | AGCTGAAA<br>GGAAAGTC<br>CCAG (SEQ<br>ID NO: 157) | AACTGAA<br>AGGAAA<br>GTCCCAG<br>(SEQ ID<br>NO: 674) | 没有 | 6.034610<br>726 | 0.871342<br>7566 |  |
| 3147 | AGCCATGT<br>TGGAGGAC<br>ACAG (SEQ<br>ID NO: 158) | AGCCATG<br>TTGGAGG<br>ACACAG<br>(SEQ ID<br>NO: 675) | 没有 | 5.861287<br>227 | 7.332930<br>712  |  |
| 1333 | TGGTACTT<br>CTTTGACA<br>ACAG (SEQ<br>ID NO: 159) | TGGTACT<br>TCTTTGA<br>CAACAG<br>(SEQ ID<br>NO: 676) | 没有 | 5.749718<br>248 | 11.96128<br>599  |  |
| 3116 | AGTCCCAG<br>GAGGAGGC<br>CCAG (SEQ<br>ID NO: 160) | AGTCCCA<br>GGAGGA<br>GGCCCAG<br>(SEQ ID<br>NO: 677) | 没有 | 5.130118<br>831 | 0.961889<br>9213 |  |
| 2082 | TCAGGGTG<br>TCTCCAAT<br>GCAG (SEQ<br>ID NO: 161) | TCAGGGT<br>GTCTCCA<br>ATGCAG<br>(SEQ ID<br>NO: 678) | 没有 | 5.130118<br>831 | 9.075616<br>225  |  |

[0445]

|      |                                                  |                                                     |     |                 |                  |  |
|------|--------------------------------------------------|-----------------------------------------------------|-----|-----------------|------------------|--|
| 3084 | GCACATGC<br>TGTTCTATG<br>CCC (SEQ ID<br>NO: 162) | GCACATG<br>CTGTTCT<br>ATGCAC<br>(SEQ ID<br>NO: 679) | 没有  | 5.130118<br>831 | 0                |  |
| 1904 | CAGTTGGA<br>ATCTACCTC<br>CAG (SEQ<br>ID NO: 163) | CAGTTGG<br>AATCTAC<br>CTCCAG<br>(SEQ ID<br>NO: 680) | 没有  | 4.906980<br>872 | 7.932631<br>974  |  |
| 2874 | CTTCTACG<br>AGAACCAG<br>ATCA (SEQ<br>ID NO: 164) | CTTCTAC<br>GAGAACC<br>AGATCA<br>(SEQ ID<br>NO: 681) | 没有  | 4.510519<br>415 | 5.409150<br>869  |  |
| 3088 | ATGCTGTTC<br>TATGCCCA<br>GCT (SEQ ID<br>NO: 165) | ATGCTGT<br>TCTATGC<br>ACAAC<br>(SEQ ID<br>NO: 682)  | 没有  | 4.287381<br>456 | 1.833232<br>678  |  |
| 3040 | TGTCCACA<br>GCACAACA<br>TCCT (SEQ<br>ID NO: 166) | TGTCCAC<br>AGCACAA<br>CATCCT<br>(SEQ ID<br>NO: 683) | 没有  | 3.890919<br>998 | 0.871342<br>7566 |  |
| 250  | AACAATCC<br>CTGTTTTCA<br>AAG (SEQ<br>ID NO: 167) | AACAATC<br>CCTGTTT<br>TCAAAG<br>(SEQ ID<br>NO: 684) | 没有  | 3.890919<br>998 | 2.614028<br>27   |  |
| 2820 | GGTAAAGA<br>TTTTTGAG<br>CCCT (SEQ<br>ID NO: 168) | GGTGAAG<br>ATTTTG<br>AGCCCT<br>(SEQ ID<br>NO: 685)  | 分支点 | 3.779351<br>019 | 2.795122<br>599  |  |
| 3160 | GACACAGG<br>CCTCCACC<br>ACAA (SEQ<br>ID NO: 169) | GACACAG<br>GCCTCCA<br>CCACAA<br>(SEQ ID<br>NO: 686) | 没有  | 3.606027<br>52  | 2.704575<br>435  |  |

[0446]

|      |                                                  |                                                     |    |                 |                  |  |
|------|--------------------------------------------------|-----------------------------------------------------|----|-----------------|------------------|--|
| 1936 | CCCTGCTT<br>CGTGGACG<br>ATTC (SEQ<br>ID NO: 170) | CCCTGCT<br>TCGTGGA<br>CGATTC<br>(SEQ ID<br>NO: 687) | 没有 | 3.382889<br>561 | 2.795122<br>599  |  |
| 1396 | GTAAAAGA<br>CTTTTGA<br>ATAG (SEQ<br>ID NO: 171)  | GTAAAAG<br>ACTTTT<br>GAATAG<br>(SEQ ID<br>NO: 688)  | 没有 | 3.382889<br>561 | 4.628355<br>277  |  |
| 2703 | AAAGACCG<br>AGCCCCTA<br>ACAG (SEQ<br>ID NO: 172) | AAAGACC<br>GAGCCCC<br>TAACAG<br>(SEQ ID<br>NO: 689) | 没有 | 3.382889<br>561 | 2.885669<br>764  |  |
| 2959 | CCTCCAAC<br>CTCTGGGA<br>CTGT (SEQ<br>ID NO: 173) | CCTCCAA<br>CCTCTGG<br>GACTGT<br>(SEQ ID<br>NO: 690) | 没有 | 3.382889<br>561 | 0.871342<br>7566 |  |
| 783  | TGTGCTTC<br>CCACACTC<br>CTAG (SEQ<br>ID NO: 174) | TGTGCTT<br>CCACAC<br>TCCTAG<br>(SEQ ID<br>NO: 691)  | 没有 | 3.271320<br>582 | 2.795122<br>599  |  |
| 2522 | GCTTCCTG<br>CTGTCCAT<br>GCAG (SEQ<br>ID NO: 175) | GCTTCCT<br>GCTGTCC<br>ATGCAG<br>(SEQ ID<br>NO: 692) | 没有 | 3.048182<br>623 | 21.03690<br>221  |  |
| 3097 | TATGCCCA<br>GCTGAAAG<br>GAAA (SEQ<br>ID NO: 176) | TATGCAC<br>AACTGAA<br>AGGAAA<br>(SEQ ID<br>NO: 693) | 没有 | 3.048182<br>623 | 2.795122<br>599  |  |
| 176  | AATGCCATT<br>TCCCCAAC<br>AAG (SEQ<br>ID NO: 177) | AATGCCA<br>TTTCCCC<br>AACAAG<br>(SEQ ID<br>NO: 694) | 没有 | 3.048182<br>623 | 0.871342<br>7566 |  |

[0447]

|      |                                                  |                                                     |    |                 |                  |  |
|------|--------------------------------------------------|-----------------------------------------------------|----|-----------------|------------------|--|
| 3085 | CACATGCT<br>GTTCTATGC<br>CCA (SEQ ID<br>NO: 178) | CACATGC<br>TGTTCTAT<br>GCACA<br>(SEQ ID<br>NO: 695) | 没有 | 2.986428<br>103 | 0                |  |
| 649  | GAGCGCTT<br>CATCATCTT<br>CAG (SEQ<br>ID NO: 179) | GAGCGCT<br>TCATCAT<br>CTTCAG<br>(SEQ ID<br>NO: 696) | 没有 | 2.986428<br>103 | 0.871342<br>7566 |  |
| 3118 | TCCCAGGA<br>GGAGGCC<br>AGCT (SEQ<br>ID NO: 180)  | TCCCAGG<br>AGGAGGC<br>CCAGCT<br>(SEQ ID<br>NO: 697) | 没有 | 2.763290<br>145 | 0.871342<br>7566 |  |
| 3098 | ATGCCCAG<br>CTGAAAGG<br>AAAG (SEQ<br>ID NO: 181) | ATGCACA<br>ACTGAAA<br>GGAAAG<br>(SEQ ID<br>NO: 698) | 没有 | 2.763290<br>145 | 1.923779<br>843  |  |
| 3128 | AGGCCCAG<br>CTGGAGAT<br>GGAA (SEQ<br>ID NO: 182) | AGGCCCA<br>GCTGGAG<br>ATGGAA<br>(SEQ ID<br>NO: 699) | 没有 | 2.763290<br>145 | 1.833232<br>678  |  |
| 1982 | TCCCTATCT<br>TCATGGTG<br>CTG (SEQ ID<br>NO: 183) | TCCCTAT<br>CTTCATG<br>GTGCTG<br>(SEQ ID<br>NO: 700) | 没有 | 2.651721<br>165 | 0.961889<br>9213 |  |
| 2586 | CCTTGATC<br>AGGTGTTT<br>CCAG (SEQ<br>ID NO: 184) | CCTTGAT<br>CAAGTGT<br>TTCCAG<br>(SEQ ID<br>NO: 701) | 没有 | 2.540152<br>186 | 0.871342<br>7566 |  |
| 2845 | CGGCCAGC<br>TGTGGACC<br>GTCT (SEQ<br>ID NO: 185) | CGGCCAG<br>CTGTGGA<br>CCGTCT<br>(SEQ ID<br>NO: 702) | 没有 | 2.540152<br>186 | 0                |  |

[0448]

|      |                                                   |                                                      |    |                 |                  |  |
|------|---------------------------------------------------|------------------------------------------------------|----|-----------------|------------------|--|
| 1782 | TTGGGATT<br>CTGGTCCC<br>AGAG (SEQ<br>ID NO: 186)  | TTGGGAT<br>TCTGGTC<br>CCAGAG<br>(SEQ ID<br>NO: 703)  | 没有 | 2.366828<br>687 | 0                |  |
| 2901 | CCTGGGCC<br>ACAATGGA<br>GCTG (SEQ<br>ID NO: 187)  | CCTGGGC<br>CACAATG<br>GAGCTG<br>(SEQ ID<br>NO: 704)  | 没有 | 2.255259<br>707 | 3.757012<br>521  |  |
| 310  | AACAACCTC<br>CATCTTGG<br>CAAG (SEQ<br>ID NO: 188) | AACAACCT<br>CCATCTT<br>GGCAAG<br>(SEQ ID<br>NO: 705) | 没有 | 2.255259<br>707 | 8.204273<br>468  |  |
| 2137 | ATCATGTCG<br>ATGAGCAT<br>CTT (SEQ ID<br>NO: 189)  | ATCATGT<br>CGATGAG<br>CATCTT<br>(SEQ ID<br>NO: 706)  | 没有 | 2.255259<br>707 | 0                |  |
| 3067 | CACCTCAC<br>GGTGGCTG<br>AGCA (SEQ<br>ID NO: 190)  | CACCTCA<br>CGGTGGC<br>TGAGCA<br>(SEQ ID<br>NO: 707)  | 没有 | 2.255259<br>707 | 0                |  |
| 3090 | GCTGTTCT<br>ATGCCCAG<br>CTGA (SEQ<br>ID NO: 191)  | GCTGTTC<br>TATGCAC<br>AACTGA<br>(SEQ ID<br>NO: 708)  | 没有 | 2.143690<br>728 | 0.871342<br>7566 |  |
| 3133 | CAGCTGGA<br>GATGGAAG<br>CCAT (SEQ<br>ID NO: 192)  | CAGCTGG<br>AGATGGA<br>AGCCAT<br>(SEQ ID<br>NO: 709)  | 没有 | 2.143690<br>728 | 0.961889<br>9213 |  |
| 3066 | CCACCTCA<br>CGGTGGCT<br>GAGC (SEQ<br>ID NO: 193)  | CCACCTC<br>ACGGTGG<br>CTGAGC<br>(SEQ ID<br>NO: 710)  | 没有 | 2.143690<br>728 | 0                |  |

[0449]

|      |                                                  |                                                     |    |                 |                  |  |
|------|--------------------------------------------------|-----------------------------------------------------|----|-----------------|------------------|--|
| 3152 | TGTTGGAG<br>GACACAGG<br>CCTC (SEQ<br>ID NO: 194) | TGTTGGA<br>GGACACA<br>GGCCTC<br>(SEQ ID<br>NO: 711) | 没有 | 2.032121<br>748 | 0                |  |
| 554  | ACCTTCTG<br>ATCAACTC<br>TCAA (SEQ<br>ID NO: 195) | ACCTTCT<br>GATCAAC<br>TCTCAA<br>(SEQ ID<br>NO: 712) | 没有 | 2.032121<br>748 | 0                |  |
| 2870 | TCACCTTCT<br>ACGAGAAC<br>CAG (SEQ<br>ID NO: 196) | TCACCTT<br>CTACGAG<br>AACCAG<br>(SEQ ID<br>NO: 713) | 没有 | 1.858798<br>25  | 1.833232<br>678  |  |
| 352  | CTCCTCATG<br>AATGCACC<br>AGA (SEQ<br>ID NO: 197) | CTCCTCA<br>TGAATGC<br>ACCAGA<br>(SEQ ID<br>NO: 714) | 没有 | 1.858798<br>25  | 0                |  |
| 1203 | GGGAAAAA<br>TCCTGTAC<br>ACTC (SEQ<br>ID NO: 198) | GGGAAA<br>AATCCTG<br>TACTCTC<br>(SEQ ID<br>NO: 715) | 没有 | 1.858798<br>25  | 0                |  |
| 1966 | ATCCTGAA<br>CCGCTGTT<br>TCCC (SEQ<br>ID NO: 199) | ATCCTGA<br>ACCGCTG<br>TTTCCC<br>(SEQ ID<br>NO: 716) | 没有 | 1.858798<br>25  | 0                |  |
| 3057 | CCTGTTCC<br>ACCACCTC<br>ACGG (SEQ<br>ID NO: 200) | CCTGTTC<br>CACCACC<br>TCACGG<br>(SEQ ID<br>NO: 717) | 没有 | 1.858798<br>25  | 0.871342<br>7566 |  |
| 2996 | ACATTGAA<br>ACCAGCCT<br>GGAT (SEQ<br>ID NO: 201) | ACATTGA<br>AACCAGC<br>CTGGAT<br>(SEQ ID<br>NO: 718) | 没有 | 1.858798<br>25  | 0                |  |

[0450]

|      |                                                  |                                                     |     |                |                  |  |
|------|--------------------------------------------------|-----------------------------------------------------|-----|----------------|------------------|--|
| 2372 | GCATGACC<br>GCTGAGCT<br>GAAG (SEQ<br>ID NO: 202) | GCATGAC<br>CGCTGAG<br>CTGAAG<br>(SEQ ID<br>NO: 719) | 没有  | 1.858798<br>25 | 0                |  |
| 2966 | CCTCTGGG<br>ACTGTGCT<br>CGTT (SEQ<br>ID NO: 203) | CCTCTGG<br>GACTGTG<br>CTCGTT<br>(SEQ ID<br>NO: 720) | 没有  | 1.858798<br>25 | 0                |  |
| 2818 | CTGGTAAA<br>GATTTTGA<br>GCC (SEQ ID<br>NO: 204)  | CTGGTGA<br>AGATTTT<br>TGAGCC<br>(SEQ ID<br>NO: 721) | 分支点 | 1.747229<br>27 | 1.923779<br>843  |  |
| 2811 | GAAGAATC<br>TGGTAAAG<br>ATT (SEQ<br>ID NO: 205)  | GAAGAAT<br>CTGGTGA<br>AGATTT<br>(SEQ ID<br>NO: 722) | 分支点 | 1.747229<br>27 | 0                |  |
| 3080 | CTGAGCAC<br>ATGCTGTC<br>TAT (SEQ ID<br>NO: 206)  | CTGAGCA<br>CATGCTG<br>TTCTAT<br>(SEQ ID<br>NO: 723) | 没有  | 1.747229<br>27 | 0                |  |
| 3081 | TGAGCACA<br>TGCTGTTCT<br>ATG (SEQ ID<br>NO: 207) | TGAGCAC<br>ATGCTGT<br>TCTATG<br>(SEQ ID<br>NO: 724) | 没有  | 1.747229<br>27 | 0.961889<br>9213 |  |
| 2832 | TGAGCCCT<br>GTGGCCGG<br>CCAG (SEQ<br>ID NO: 208) | TGAGCCC<br>TGTGGCC<br>GGCCAG<br>(SEQ ID<br>NO: 725) | 没有  | 1.747229<br>27 | 0                |  |
| 3125 | AGGAGGCC<br>CAGCTGGA<br>GATG (SEQ<br>ID NO: 209) | AGGAGGC<br>CCAGCTG<br>GAGATG<br>(SEQ ID<br>NO: 726) | 没有  | 1.747229<br>27 | 0                |  |

[0451]

|      |                                                  |                                                     |    |                 |                  |  |
|------|--------------------------------------------------|-----------------------------------------------------|----|-----------------|------------------|--|
| 1860 | GGTTGAAC<br>AGGGGATC<br>ACAA (SEQ<br>ID NO: 210) | GGTTGAA<br>CAGGGGA<br>TCACAA<br>(SEQ ID<br>NO: 727) | 没有 | 1.747229<br>27  | 0                |  |
| 3123 | GGAGGAGG<br>CCCAGCTG<br>GAGA (SEQ<br>ID NO: 211) | GGAGGA<br>GGCCCAG<br>CTGGAGA<br>(SEQ ID<br>NO: 728) | 没有 | 1.747229<br>27  | 0                |  |
| 2774 | AGAGAGAA<br>CACCCCGG<br>GTGG (SEQ<br>ID NO: 212) | AGAGAG<br>AACACCC<br>CGGGTGG<br>(SEQ ID<br>NO: 729) | 没有 | 1.635660<br>291 | 7.932631<br>974  |  |
| 3071 | TCACGGTG<br>GCTGAGCA<br>CATG (SEQ<br>ID NO: 213) | TCACGGT<br>GGCTGAG<br>CACATG<br>(SEQ ID<br>NO: 730) | 没有 | 1.635660<br>291 | 0.961889<br>9213 |  |
| 2600 | TTCCAGGA<br>GACTATGG<br>AACC (SEQ<br>ID NO: 214) | TTCCAGG<br>AGACTAT<br>GGAACC<br>(SEQ ID<br>NO: 731) | 没有 | 1.635660<br>291 | 0                |  |
| 3119 | CCCAGGAG<br>GAGGCCCA<br>GCTG (SEQ<br>ID NO: 215) | CCCAGGA<br>GGAGGCC<br>CAGCTG<br>(SEQ ID<br>NO: 732) | 没有 | 1.635660<br>291 | 0                |  |
| 2713 | CCCCTAAC<br>AGAGGAAA<br>CGGA (SEQ<br>ID NO: 216) | CCCCTAA<br>CAGAGGA<br>AACGGA<br>(SEQ ID<br>NO: 733) | 没有 | 1.635660<br>291 | 0                |  |
| 2721 | AGAGGAAA<br>CGGAGGAT<br>CCAG (SEQ<br>ID NO: 217) | AGAGGA<br>AACGGAG<br>GATCCAG<br>(SEQ ID<br>NO: 734) | 没有 | 1.635660<br>291 | 0                |  |

[0452]

|      |                                                  |                                                      |    |                 |                 |  |
|------|--------------------------------------------------|------------------------------------------------------|----|-----------------|-----------------|--|
| 2857 | GACCGTCT<br>GAACATCA<br>CCTT (SEQ<br>ID NO: 218) | GACCGTC<br>TGAACAT<br>CACCTT<br>(SEQ ID<br>NO: 735)  | 没有 | 1.635660<br>291 | 0               |  |
| 3046 | CAGCACAA<br>CATCCTGTT<br>CCA (SEQ ID<br>NO: 219) | CAGCACAA<br>ACATCCT<br>GTTCCA<br>(SEQ ID<br>NO: 736) | 没有 | 1.635660<br>291 | 0               |  |
| 3087 | CATGCTGTT<br>CTATGCCC<br>AGC (SEQ<br>ID NO: 220) | CATGCTG<br>TTCTATG<br>CACAAAC<br>(SEQ ID<br>NO: 737) | 没有 | 1.635660<br>291 | 0               |  |
| 2676 | TTCAACCA<br>GAGAAGAA<br>AGAG (SEQ<br>ID NO: 221) | TTCAACC<br>AGAGAA<br>GAAAGA<br>G (SEQ ID<br>NO: 738) | 没有 | 1.635660<br>291 | 0               |  |
| 2831 | TTGAGCCC<br>TGTGGCCG<br>GCCA (SEQ<br>ID NO: 222) | TTGAGCC<br>CTGTGGC<br>CGGCCA<br>(SEQ ID<br>NO: 739)  | 没有 | 1.524091<br>311 | 3.847559<br>685 |  |
| 148  | AATGCCAA<br>CCCGCTCT<br>ACAG (SEQ<br>ID NO: 223) | AATGCCA<br>ACCCGCT<br>CTACAG<br>(SEQ ID<br>NO: 740)  | 没有 | 1.524091<br>311 | 2.614028<br>27  |  |
| 580  | CCAGAGCA<br>GTTCGCTC<br>ATGG (SEQ<br>ID NO: 224) | CCAGAGC<br>AGTTCGC<br>TCATGG<br>(SEQ ID<br>NO: 741)  | 没有 | 1.524091<br>311 | 0               |  |
| 643  | CTCCTGGA<br>GCGCTTCA<br>TCAT (SEQ<br>ID NO: 225) | CTCCTGG<br>AGCGCTT<br>CATCAT<br>(SEQ ID<br>NO: 742)  | 没有 | 1.524091<br>311 | 0               |  |

[0453]

|      |                                                  |                                                     |    |                 |                  |  |
|------|--------------------------------------------------|-----------------------------------------------------|----|-----------------|------------------|--|
| 703  | CTGTGCTC<br>CCTCTCCC<br>AGGG (SEQ<br>ID NO: 226) | CTGTGCT<br>CCCTCTC<br>CCAGGG<br>(SEQ ID<br>NO: 743) | 没有 | 1.524091<br>311 | 0                |  |
| 3070 | CTCACGGT<br>GGCTGAGC<br>ACAT (SEQ<br>ID NO: 227) | CTCACGG<br>TGGCTGA<br>GCACAT<br>(SEQ ID<br>NO: 744) | 没有 | 1.524091<br>311 | 0.961889<br>9213 |  |
| 2783 | ACCCCGGG<br>TGGGTTCC<br>TGGG (SEQ<br>ID NO: 228) | ACCCCGG<br>GTGGGTT<br>CCTGGG<br>(SEQ ID<br>NO: 745) | 没有 | 1.524091<br>311 | 0                |  |
| 3075 | GGTGGCTG<br>AGCACATG<br>CTGT (SEQ<br>ID NO: 229) | GGTGGCT<br>GAGCACA<br>TGCTGT<br>(SEQ ID<br>NO: 746) | 没有 | 1.524091<br>311 | 0                |  |
| 2852 | CTGTGGAC<br>CGTCTGAA<br>CATC (SEQ<br>ID NO: 230) | CTGTGGA<br>CCGTCTG<br>AACATC<br>(SEQ ID<br>NO: 747) | 没有 | 1.524091<br>311 | 0                |  |
| 3107 | TGAAAGGA<br>AAGTCCCA<br>GGAG (SEQ<br>ID NO: 231) | TGAAAGG<br>AAAGTCC<br>CAGGAG<br>(SEQ ID<br>NO: 748) | 没有 | 1.524091<br>311 | 2.795122<br>599  |  |
| 2270 | TCAGCACC<br>TTCTTCTCC<br>AAG (SEQ<br>ID NO: 232) | TCAGCAC<br>CTTCTTC<br>TCCAAG<br>(SEQ ID<br>NO: 749) | 没有 | 1.524091<br>311 | 1.833232<br>678  |  |
| 2988 | GGGAAGGG<br>ACATTGAA<br>ACCA (SEQ<br>ID NO: 233) | GGGAAG<br>GGACATT<br>GAAACCA<br>(SEQ ID<br>NO: 750) | 没有 | 1.239198<br>833 | 0                |  |

[0454]

|      |                                                  |                                                     |    |                 |                  |  |
|------|--------------------------------------------------|-----------------------------------------------------|----|-----------------|------------------|--|
| 3149 | CCATGTTG<br>GAGGACAC<br>AGGC (SEQ<br>ID NO: 234) | CCATGTT<br>GGAGGAC<br>ACAGGC<br>(SEQ ID<br>NO: 751) | 没有 | 1.239198<br>833 | 0                |  |
| 698  | ATGCCCTG<br>TGCTCCCT<br>CTCC (SEQ<br>ID NO: 235) | ATGCCCT<br>GTGCTCC<br>CTCTCC<br>(SEQ ID<br>NO: 752) | 没有 | 1.239198<br>833 | 0                |  |
| 3106 | CTGAAAGG<br>AAAGTCCC<br>AGGA (SEQ<br>ID NO: 236) | CTGAAAG<br>GAAAGTC<br>CCAGGA<br>(SEQ ID<br>NO: 753) | 没有 | 1.239198<br>833 | 0                |  |
| 2691 | AAGAGCCC<br>TGGAAGG<br>ACCG (SEQ<br>ID NO: 237)  | AAGAGCC<br>CTGGAAA<br>AGACCG<br>(SEQ ID<br>NO: 754) | 没有 | 1.239198<br>833 | 0                |  |
| 3135 | GCTGGAGA<br>TGGAAGCC<br>ATGT (SEQ<br>ID NO: 238) | GCTGGAG<br>ATGGAAG<br>CCATGT<br>(SEQ ID<br>NO: 755) | 没有 | 1.239198<br>833 | 0.871342<br>7566 |  |
| 3115 | AAGTCCCA<br>GGAGGAGG<br>CCCA (SEQ<br>ID NO: 239) | AAGTCCC<br>AGGAGG<br>AGGCCCA<br>(SEQ ID<br>NO: 756) | 没有 | 1.239198<br>833 | 1.923779<br>843  |  |
| 2974 | ACTGTGCT<br>CGTTGGGG<br>GAAG (SEQ<br>ID NO: 240) | ACTGTGC<br>TCGTTGG<br>GGGAAG<br>(SEQ ID<br>NO: 757) | 没有 | 1.239198<br>833 | 0                |  |
| 2789 | GGTGGGTT<br>CCTGGGGT<br>ATGC (SEQ<br>ID NO: 241) | GGTGGGT<br>TCCTGGG<br>GTGTGC<br>(SEQ ID<br>NO: 758) | 没有 | 1.239198<br>833 | 0                |  |

[0455]

|      |                                                  |                                                     |    |                 |                  |  |
|------|--------------------------------------------------|-----------------------------------------------------|----|-----------------|------------------|--|
| 2134 | TCCATCATG<br>TCGATGAG<br>CAT (SEQ ID<br>NO: 242) | TCCATCA<br>TGTCGAT<br>GAGCAT<br>(SEQ ID<br>NO: 759) | 没有 | 1.239198<br>833 | 0                |  |
| 2700 | GGAAAAGA<br>CCGAGCCC<br>CTAA (SEQ<br>ID NO: 243) | GGAAAA<br>GACCGAG<br>CCCCTAA<br>(SEQ ID<br>NO: 760) | 没有 | 1.239198<br>833 | 0                |  |
| 566  | ACTCTCAA<br>GTCCGTCC<br>AGAG (SEQ<br>ID NO: 244) | ACTCTCA<br>AGTCCGT<br>CCAGAG<br>(SEQ ID<br>NO: 761) | 没有 | 1.239198<br>833 | 0                |  |
| 3137 | TGGAGATG<br>GAAGCCAT<br>GTTG (SEQ<br>ID NO: 245) | TGGAGAT<br>GGAAGCC<br>ATGTTG<br>(SEQ ID<br>NO: 762) | 没有 | 1.239198<br>833 | 0.961889<br>9213 |  |
| 3096 | CTATGCCC<br>AGCTGAAA<br>GGAA (SEQ<br>ID NO: 246) | CTATGCA<br>CAACTGA<br>AAGGAA<br>(SEQ ID<br>NO: 763) | 没有 | 1.239198<br>833 | 0.871342<br>7566 |  |
| 19   | AGCGGCTT<br>CGTGAGAC<br>AGAT (SEQ<br>ID NO: 247) | TCCGGCT<br>TCGTGAG<br>ACAGAT<br>(SEQ ID<br>NO: 764) | 没有 | 1.239198<br>833 | 0                |  |
| 3092 | TGTTCTATG<br>CCCAGCTG<br>AAA (SEQ<br>ID NO: 248) | TGTTCTAT<br>GCACAAC<br>TGAAA<br>(SEQ ID<br>NO: 765) | 没有 | 1.239198<br>833 | 0                |  |
| 1877 | CAAGGAGC<br>CAGGTGCA<br>GGCG (SEQ<br>ID NO: 249) | CAAGGAG<br>CCAGGTG<br>CAGGCG<br>(SEQ ID<br>NO: 766) | 没有 | 1.239198<br>833 | 0                |  |

[0456]

|      |                                                  |                                                     |    |                 |                  |  |
|------|--------------------------------------------------|-----------------------------------------------------|----|-----------------|------------------|--|
| 857  | TGTCACCA<br>AGAATTCA<br>AGAG (SEQ<br>ID NO: 250) | TGTCACC<br>AAGAATT<br>CAAGAG<br>(SEQ ID<br>NO: 767) | 没有 | 1.127629<br>854 | 0.961889<br>9213 |  |
| 2846 | GGCCAGCT<br>GTGGACCG<br>TCTG (SEQ<br>ID NO: 251) | GGCCAGC<br>TGTGGAC<br>CGTCTG<br>(SEQ ID<br>NO: 768) | 没有 | 1.127629<br>854 | 0                |  |
| 2977 | GTGCTCGT<br>TGGGGGAA<br>GGGA (SEQ<br>ID NO: 252) | GTGCTCG<br>TTGGGGG<br>AAGGGA<br>(SEQ ID<br>NO: 769) | 没有 | 1.127629<br>854 | 0                |  |
| 616  | CTGAAGGA<br>CATCGCCT<br>GCAG (SEQ<br>ID NO: 253) | CTGAAGG<br>ACATCGC<br>CTGCAG<br>(SEQ ID<br>NO: 770) | 没有 | 1.127629<br>854 | 0                |  |
| 2729 | CGGAGGAT<br>CCAGAGCA<br>CCCA (SEQ<br>ID NO: 254) | CGGAGGA<br>TCCAGAG<br>CACCCA<br>(SEQ ID<br>NO: 771) | 没有 | 1.127629<br>854 | 0                |  |
| 2834 | AGCCCTGT<br>GGCCGGCC<br>AGCT (SEQ<br>ID NO: 255) | AGCCCTG<br>TGGCCGG<br>CCAGCT<br>(SEQ ID<br>NO: 772) | 没有 | 1.127629<br>854 | 0                |  |
| 3140 | AGATGGAA<br>GCCATGTT<br>GGAG (SEQ<br>ID NO: 256) | AGATGGA<br>AGCCATG<br>TTGGAG<br>(SEQ ID<br>NO: 773) | 没有 | 1.127629<br>854 | 0                |  |
| 2989 | GGAAGGGA<br>CATTGAAA<br>CCAG (SEQ<br>ID NO: 257) | GGAAGG<br>GACATTG<br>AAACCAG<br>(SEQ ID<br>NO: 774) | 没有 | 1.127629<br>854 | 0                |  |

[0457]

|      |                                                   |                                                      |    |                 |                  |  |
|------|---------------------------------------------------|------------------------------------------------------|----|-----------------|------------------|--|
| 2938 | TCCATCCTC<br>ACAGGCCT<br>GCT (SEQ ID<br>NO: 258)  | TCCATCC<br>TCACTGG<br>CCTGCT<br>(SEQ ID<br>NO: 775)  | 没有 | 1.127629<br>854 | 0                |  |
| 2972 | GGA CTGTG<br>CTCGTTGG<br>GGGA (SEQ<br>ID NO: 259) | GGA CTGT<br>GCTCGTT<br>GGGGGA<br>(SEQ ID<br>NO: 776) | 没有 | 1.127629<br>854 | 0.961889<br>9213 |  |
| 3127 | GAGGCCCA<br>GCTGGAGA<br>TGGA (SEQ<br>ID NO: 260)  | GAGGCC<br>AGCTGGA<br>GATGGA<br>(SEQ ID<br>NO: 777)   | 没有 | 1.127629<br>854 | 0                |  |
| 192  | CAAGGCGA<br>TGCCCTCA<br>GCAG (SEQ<br>ID NO: 261)  | CAAGGCG<br>ATGCCCT<br>CAGCAG<br>(SEQ ID<br>NO: 778)  | 没有 | 1.127629<br>854 | 0.871342<br>7566 |  |
| 2652 | GTATTGGCT<br>TGGCGGTG<br>AAG (SEQ<br>ID NO: 262)  | GTATTGG<br>CTTGGCG<br>GTGAAG<br>(SEQ ID<br>NO: 779)  | 没有 | 1.127629<br>854 | 3.757012<br>521  |  |
| 1469 | AGGGCCCT<br>CGGGAAAG<br>CCAG (SEQ<br>ID NO: 263)  | AGGGCCC<br>TCGGGAA<br>AGCCAG<br>(SEQ ID<br>NO: 780)  | 没有 | 1.127629<br>854 | 0                |  |
| 3162 | CACAGGCC<br>TCCACCAC<br>AAGC (SEQ<br>ID NO: 264)  | CACAGGC<br>CTCCACC<br>ACAAGC<br>(SEQ ID<br>NO: 781)  | 没有 | 1.127629<br>854 | 2.885669<br>764  |  |
| 2705 | AGACCGAG<br>CCCCTAAC<br>AGAG (SEQ<br>ID NO: 265)  | AGACCGA<br>GCCCCTA<br>ACAGAG<br>(SEQ ID<br>NO: 782)  | 没有 | 1.127629<br>854 | 0                |  |

[0458]

|      |                                                  |                                                     |     |                 |                  |  |
|------|--------------------------------------------------|-----------------------------------------------------|-----|-----------------|------------------|--|
| 2665 | GGTGAAGG<br>GTGTTCAA<br>CCAG (SEQ<br>ID NO: 266) | GGTGAAG<br>GGTGTTC<br>AACCAG<br>(SEQ ID<br>NO: 783) | 没有  | 1.127629<br>854 | 0.961889<br>9213 |  |
| 2927 | CCACCACC<br>TTGTCCATC<br>CTC (SEQ ID<br>NO: 267) | CCACCAC<br>CTTGTCC<br>ATCCTC<br>(SEQ ID<br>NO: 784) | 没有  | 1.127629<br>854 | 0                |  |
| 2822 | TAAAGATT<br>TTTGAGCC<br>CTGT (SEQ<br>ID NO: 268) | TGAAGAT<br>TTTTGAG<br>CCCTGT<br>(SEQ ID<br>NO: 785) | 分支点 | 1.127629<br>854 | 0                |  |
| 3124 | GAGGAGGC<br>CCAGCTGG<br>AGAT (SEQ<br>ID NO: 269) | GAGGAG<br>GCCCAGC<br>TGGAGAT<br>(SEQ ID<br>NO: 786) | 没有  | 1.127629<br>854 | 0                |  |
| 3129 | GGCCCAGC<br>TGGAGATG<br>GAAG (SEQ<br>ID NO: 270) | GGCCCAG<br>CTGGAGA<br>TGGAAG<br>(SEQ ID<br>NO: 787) | 没有  | 1.127629<br>854 | 0                |  |
| 2993 | GGGACATT<br>GAAACCAG<br>CCTG (SEQ<br>ID NO: 271) | GGGACAT<br>TGAAACC<br>AGCCTG<br>(SEQ ID<br>NO: 788) | 没有  | 1.016060<br>874 | 0                |  |
| 3101 | CCCAGCTG<br>AAAGGAAA<br>GTCC (SEQ<br>ID NO: 272) | CACAACT<br>GAAAGG<br>AAAGTCC<br>(SEQ ID<br>NO: 789) | 没有  | 1.016060<br>874 | 0                |  |
| 2650 | TCGTATTGG<br>CTTGCGG<br>TGA (SEQ ID<br>NO: 273)  | TCGTATT<br>GGCTTGG<br>CGGTGA<br>(SEQ ID<br>NO: 790) | 没有  | 1.016060<br>874 | 0                |  |

[0459]

|      |                                                  |                                                     |     |                 |                 |  |
|------|--------------------------------------------------|-----------------------------------------------------|-----|-----------------|-----------------|--|
| 2816 | ATCTGGTA<br>AAGATTTT<br>GAG (SEQ<br>ID NO: 274)  | ATCTGGT<br>GAAGATT<br>TTTGAG<br>(SEQ ID<br>NO: 791) | 分支点 | 1.016060<br>874 | 3.757012<br>521 |  |
| 2361 | CTGGCAGG<br>ACCGCATG<br>ACCG (SEQ<br>ID NO: 275) | CTGGCAA<br>GACCGCA<br>TGACCG<br>(SEQ ID<br>NO: 792) | 没有  | 1.016060<br>874 | 0               |  |
| 592  | GCTCATGG<br>AGTCCCGG<br>ACCT (SEQ<br>ID NO: 276) | GCTCATG<br>GAGTCCC<br>GGACCT<br>(SEQ ID<br>NO: 793) | 没有  | 1.016060<br>874 | 0               |  |
| 490  | AAAGATGA<br>AGAAACAC<br>TGAC (SEQ<br>ID NO: 277) | AAAGATG<br>AAGAAAC<br>ACTGAC<br>(SEQ ID<br>NO: 794) | 没有  | 1.016060<br>874 | 0               |  |
| 2260 | TGCTTTCTG<br>CTCAGCAC<br>CTT (SEQ ID<br>NO: 278) | TGCTTTC<br>TGCTCAG<br>CACCTT<br>(SEQ ID<br>NO: 795) | 没有  | 1.016060<br>874 | 0               |  |
| 2104 | ATTTGGTGT<br>ACCTGGTT<br>CCT (SEQ ID<br>NO: 279) | ATTTGGT<br>GTACCTG<br>GTTCTT<br>(SEQ ID<br>NO: 796) | 没有  | 1.016060<br>874 | 0               |  |
| 2285 | CCAAGGCC<br>AGTCTGGC<br>AGCA (SEQ<br>ID NO: 280) | CCAAGGC<br>CAGTCTG<br>GCAGCA<br>(SEQ ID<br>NO: 797) | 没有  | 1.016060<br>874 | 0               |  |
| 2806 | TGCGTGAA<br>GAATCTGG<br>TAAA (SEQ<br>ID NO: 281) | TGCGTGA<br>AGAATCT<br>GGTGAA<br>(SEQ ID<br>NO: 798) | 没有  | 1.016060<br>874 | 1.833232<br>678 |  |

[0460]

|      |                                                  |                                                     |    |                 |                  |  |
|------|--------------------------------------------------|-----------------------------------------------------|----|-----------------|------------------|--|
| 2808 | CGTGAAGA<br>ATCTGGTA<br>AAGA (SEQ<br>ID NO: 282) | CGTGAAG<br>AATCTGG<br>TGAAGA<br>(SEQ ID<br>NO: 799) | 没有 | 1.016060<br>874 | 0.961889<br>9213 |  |
| 2257 | CTGTGCTTT<br>CTGCTCAG<br>CAC (SEQ ID<br>NO: 283) | CTGTGCT<br>TTCTGCT<br>CAGCAC<br>(SEQ ID<br>NO: 800) | 没有 | 1.016060<br>874 | 0.961889<br>9213 |  |
| 2778 | AGAACACC<br>CCGGGTGG<br>GTTC (SEQ<br>ID NO: 284) | AGAACAC<br>CCCGGGT<br>GGGTTC<br>(SEQ ID<br>NO: 801) | 没有 | 1.016060<br>874 | 0.871342<br>7566 |  |
| 3049 | CACAACAT<br>CCTGTTCC<br>ACCA (SEQ<br>ID NO: 285) | CACAACA<br>TCCTGTT<br>CCACCA<br>(SEQ ID<br>NO: 802) | 没有 | 1.016060<br>874 | 1.923779<br>843  |  |
| 2052 | GGAGTTGC<br>GACTGAAG<br>GAGA (SEQ<br>ID NO: 286) | GGAGTTG<br>CGACTGA<br>AGGAGA<br>(SEQ ID<br>NO: 803) | 没有 | 1.016060<br>874 | 0                |  |
| 3121 | CAGGAGGA<br>GGCCCAGC<br>TGGA (SEQ<br>ID NO: 287) | CAGGAGG<br>AGGCCCA<br>GCTGGA<br>(SEQ ID<br>NO: 804) | 没有 | 1.016060<br>874 | 0                |  |
| 767  | TCTTCAAG<br>CTCTTCCGT<br>GTG (SEQ ID<br>NO: 288) | TCTTCAA<br>GCTCTTC<br>CGTGTG<br>(SEQ ID<br>NO: 805) | 没有 | 1.016060<br>874 | 0.871342<br>7566 |  |
| 1861 | GTTGAACA<br>GGGGATCA<br>CAAG (SEQ<br>ID NO: 289) | GTTGAAC<br>AGGGGAT<br>CACAAG<br>(SEQ ID<br>NO: 806) | 没有 | 1.016060<br>874 | 0                |  |

[0461]

|      |                                                  |                                                     |    |                 |                  |  |
|------|--------------------------------------------------|-----------------------------------------------------|----|-----------------|------------------|--|
| 702  | CCTGTGCT<br>CCCTCTCC<br>CAGG (SEQ<br>ID NO: 290) | CCTGTGC<br>TCCCTCT<br>CCCAGG<br>(SEQ ID<br>NO: 807) | 没有 | 1.016060<br>874 | 0                |  |
| 3000 | TGAAACCA<br>GCCTGGAT<br>GCAG (SEQ<br>ID NO: 291) | TGAAACC<br>AGCCTGG<br>ATGCAG<br>(SEQ ID<br>NO: 808) | 没有 | 1.016060<br>874 | 0.871342<br>7566 |  |
| 2098 | GCAGTGAT<br>TTGGTGTA<br>CCTG (SEQ<br>ID NO: 292) | GCAGTGA<br>TTTGGTG<br>TACCTG<br>(SEQ ID<br>NO: 809) | 没有 | 1.016060<br>874 | 1.742685<br>513  |  |
| 2782 | CACCCCGG<br>GTGGGTTC<br>CTGG (SEQ<br>ID NO: 293) | CACCCCG<br>GGTGGGT<br>TCCTGG<br>(SEQ ID<br>NO: 810) | 没有 | 1.016060<br>874 | 0.871342<br>7566 |  |
| 2862 | TCTGAACA<br>TCACCTTCT<br>ACG (SEQ<br>ID NO: 294) | TCTGAAC<br>ATCACCT<br>TCTACG<br>(SEQ ID<br>NO: 811) | 没有 | 1.016060<br>874 | 0                |  |
| 2273 | GCACCTTC<br>TTCTCCAA<br>GGCC (SEQ<br>ID NO: 295) | GCACCTT<br>CTTCTCC<br>AAGGCC<br>(SEQ ID<br>NO: 812) | 没有 | 1.016060<br>874 | 0                |  |
| 569  | CTCAAGTC<br>CGTCCAGA<br>GCAG (SEQ<br>ID NO: 296) | CTCAAGT<br>CCGTCCA<br>GAGCAG<br>(SEQ ID<br>NO: 813) | 没有 | 1.016060<br>874 | 0.961889<br>9213 |  |
| 676  | CGCGGGGC<br>AAAGACGG<br>TGCG (SEQ<br>ID NO: 297) | CGCGGGG<br>CAAAGAC<br>GGTGCG<br>(SEQ ID<br>NO: 814) | 没有 | 1.016060<br>874 | 0                |  |

[0462]

|      |                                                  |                                                     |    |                 |                  |  |
|------|--------------------------------------------------|-----------------------------------------------------|----|-----------------|------------------|--|
| 3141 | GATGGAAG<br>CCATGTTG<br>GAGG (SEQ<br>ID NO: 298) | GATGGAA<br>GCCATGT<br>TGGAGG<br>(SEQ ID<br>NO: 815) | 没有 | 1.016060<br>874 | 0                |  |
| 2627 | CTTGGTAC<br>TTTCTTCTA<br>CAA (SEQ ID<br>NO: 299) | CTTGGTA<br>CTTTCTT<br>CTACAA<br>(SEQ ID<br>NO: 816) | 没有 | 1.016060<br>874 | 1.923779<br>843  |  |
| 2009 | TCTACTCTG<br>TCTCCATG<br>ACT (SEQ ID<br>NO: 300) | TCTACTC<br>TGTCTCC<br>ATGACT<br>(SEQ ID<br>NO: 817) | 没有 | 1.016060<br>874 | 0                |  |
| 1850 | TGCAGGAC<br>ATGGTTGA<br>ACAG (SEQ<br>ID NO: 301) | TGCAAGA<br>CATGGTT<br>GAACAG<br>(SEQ ID<br>NO: 818) | 没有 | 1.016060<br>874 | 0                |  |
| 1887 | GGTGCAGG<br>CGGAGGCT<br>CCAG (SEQ<br>ID NO: 302) | GGTGCAG<br>GCGGAGG<br>CTCCAG<br>(SEQ ID<br>NO: 819) | 没有 | 1.016060<br>874 | 0                |  |
| 2315 | GTGTCATCT<br>ATTTCACCC<br>TC (SEQ ID<br>NO: 303) | GTGTCAT<br>CTATTTC<br>ACCCTG<br>(SEQ ID<br>NO: 820) | 没有 | 1.016060<br>874 | 0                |  |
| 1359 | GATGAACA<br>TGATCAGA<br>GATA (SEQ<br>ID NO: 304) | GATGAAC<br>ATGATCA<br>GAGATA<br>(SEQ ID<br>NO: 821) | 没有 | 1.016060<br>874 | 0                |  |
| 2787 | CGGGTGGG<br>TTCCTGGG<br>GTAT (SEQ<br>ID NO: 305) | CGGGTGG<br>GTTCTG<br>GGGTGT<br>(SEQ ID<br>NO: 822)  | 没有 | 1.016060<br>874 | 0.961889<br>9213 |  |

[0463]

|      |                                                  |                                                      |    |                  |   |  |
|------|--------------------------------------------------|------------------------------------------------------|----|------------------|---|--|
| 2953 | CTGCTGCC<br>TCCAACCT<br>CTGG (SEQ<br>ID NO: 306) | CTGCTGC<br>CTCCAAC<br>CTCTGG<br>(SEQ ID<br>NO: 823)  | 没有 | 1.016060<br>874  | 0 |  |
| 2924 | AAACCACC<br>ACCTTGTC<br>CATC (SEQ<br>ID NO: 307) | AAACCAC<br>CACCTTG<br>TCCATC<br>(SEQ ID<br>NO: 824)  | 没有 | 0.619599<br>4166 | 0 |  |
| 3100 | GCCCAGCT<br>GAAAGGAA<br>AGTC (SEQ<br>ID NO: 308) | GCACAAC<br>TGAAAGG<br>AAAGTC<br>(SEQ ID<br>NO: 825)  | 没有 | 0.619599<br>4166 | 0 |  |
| 2687 | AAGAAAGA<br>GCCCTGGA<br>AAAG (SEQ<br>ID NO: 309) | AAGAAA<br>GAGCCCT<br>GGAAAA<br>G (SEQ ID<br>NO: 826) | 没有 | 0.619599<br>4166 | 0 |  |
| 2446 | GTTCGCTTT<br>GAAGAGCA<br>AGG (SEQ<br>ID NO: 310) | GTTCGCT<br>TTGAAGA<br>GCAAGG<br>(SEQ ID<br>NO: 827)  | 没有 | 0.619599<br>4166 | 0 |  |
| 3113 | GAAAGTCC<br>CAGGAGGA<br>GGCC (SEQ<br>ID NO: 311) | GAAAGTC<br>CCAGGAG<br>GAGGCC<br>(SEQ ID<br>NO: 828)  | 没有 | 0.619599<br>4166 | 0 |  |
| 2661 | TGGCGGTG<br>AAGGGTGT<br>TCAA (SEQ<br>ID NO: 312) | TGGCGGT<br>GAAGGGT<br>GTTCAA<br>(SEQ ID<br>NO: 829)  | 没有 | 0.619599<br>4166 | 0 |  |
| 2681 | CCAGAGAA<br>GAAAGAGC<br>CCTG (SEQ<br>ID NO: 313) | CCAGAGA<br>AGAAAG<br>AGCCCTG<br>(SEQ ID<br>NO: 830)  | 没有 | 0.619599<br>4166 | 0 |  |

[0464]

|      |                                                   |                                                     |    |                  |                  |  |
|------|---------------------------------------------------|-----------------------------------------------------|----|------------------|------------------|--|
| 2786 | CCGGGTGG<br>GTTTCCTGG<br>GGTA (SEQ<br>ID NO: 314) | CCGGGTG<br>GGTTCCT<br>GGGGTG<br>(SEQ ID<br>NO: 831) | 没有 | 0.619599<br>4166 | 0                |  |
| 1880 | GGAGCCAG<br>GTGCAGGC<br>GGAG (SEQ<br>ID NO: 315)  | GGAGCCA<br>GGTGCAG<br>GCGGAG<br>(SEQ ID<br>NO: 832) | 没有 | 0.619599<br>4166 | 0                |  |
| 2632 | TACTTTCTT<br>CTACAAGA<br>GTC (SEQ ID<br>NO: 316)  | TACTTTC<br>TTCTACA<br>AGAGTC<br>(SEQ ID<br>NO: 833) | 没有 | 0.619599<br>4166 | 0                |  |
| 1999 | CTGGCATG<br>GATCTACTC<br>TGT (SEQ ID<br>NO: 317)  | CTGGCAT<br>GGATCTA<br>CTCTGT<br>(SEQ ID<br>NO: 834) | 没有 | 0.619599<br>4166 | 0                |  |
| 3284 | CAAGCGGC<br>GTGGACCC<br>TTAC (SEQ<br>ID NO: 318)  | CAAGCGG<br>CGTGGAC<br>CCTTAC<br>(SEQ ID<br>NO: 835) | 没有 | 0.619599<br>4166 | 0                |  |
| 2543 | TGATGCTC<br>CTTGATGCT<br>GCT (SEQ ID<br>NO: 319)  | TGATGCT<br>CCTTGAT<br>GCTGCT<br>(SEQ ID<br>NO: 836) | 没有 | 0.619599<br>4166 | 0                |  |
| 2792 | GGGTTCCT<br>GGGGTATG<br>CGTG (SEQ<br>ID NO: 320)  | GGGTTC<br>TGGGGTG<br>TGCGTG<br>(SEQ ID<br>NO: 837)  | 没有 | 0.619599<br>4166 | 0                |  |
| 2853 | TGTGGACC<br>GTCTGAAC<br>ATCA (SEQ<br>ID NO: 321)  | TGTGGAC<br>CGTCTGA<br>ACATCA<br>(SEQ ID<br>NO: 838) | 没有 | 0.619599<br>4166 | 0.871342<br>7566 |  |

[0465]

|      |                                                  |                                                     |    |                  |                  |  |
|------|--------------------------------------------------|-----------------------------------------------------|----|------------------|------------------|--|
| 714  | CTCCCAGG<br>GCACCCTA<br>CAGT (SEQ<br>ID NO: 322) | CTCCCAG<br>GGCACCC<br>TACAGT<br>(SEQ ID<br>NO: 839) | 没有 | 0.619599<br>4166 | 0                |  |
| 2158 | CTCCTGAC<br>GATATTCAT<br>CAT (SEQ ID<br>NO: 323) | CTCCTGA<br>CGATATT<br>CATCAT<br>(SEQ ID<br>NO: 840) | 没有 | 0.619599<br>4166 | 0                |  |
| 1158 | TCCTTTAAC<br>CAAAATCG<br>CTT (SEQ ID<br>NO: 324) | TCCTTTA<br>ACCAAAA<br>TCGCTT<br>(SEQ ID<br>NO: 841) | 没有 | 0.619599<br>4166 | 0                |  |
| 2869 | ATCACCTTC<br>TACGAGAA<br>CCA (SEQ ID<br>NO: 325) | ATCACCT<br>TCTACGA<br>GAACCA<br>(SEQ ID<br>NO: 842) | 没有 | 0.619599<br>4166 | 0.871342<br>7566 |  |
| 2772 | CGAGAGAG<br>AACACCCC<br>GGGT (SEQ<br>ID NO: 326) | CGAGAGA<br>GAACACC<br>CCGGGT<br>(SEQ ID<br>NO: 843) | 没有 | 0.619599<br>4166 | 2.614028<br>27   |  |
| 3158 | AGGACACA<br>GGCCTCCA<br>CCAC (SEQ<br>ID NO: 327) | AGGACAC<br>AGGCCTC<br>CACCAC<br>(SEQ ID<br>NO: 844) | 没有 | 0.619599<br>4166 | 0                |  |
| 3230 | GCGTGGCC<br>ATTGCCTTC<br>GTG (SEQ ID<br>NO: 328) | GCGTGGC<br>CATTGCC<br>TTCGTG<br>(SEQ ID<br>NO: 845) | 没有 | 0.619599<br>4166 | 0                |  |
| 2825 | AGATTTTGT<br>AGCCCTGT<br>GGC (SEQ<br>ID NO: 329) | AGATTTT<br>TGAGCCC<br>TGTGGC<br>(SEQ ID<br>NO: 846) | 没有 | 0.619599<br>4166 | 0                |  |

[0466]

|      |                                                  |                                                     |    |                  |   |  |
|------|--------------------------------------------------|-----------------------------------------------------|----|------------------|---|--|
| 3156 | GGAGGACA<br>CAGGCCTC<br>CACC (SEQ<br>ID NO: 330) | GGAGGAC<br>ACAGGCC<br>TCCACC<br>(SEQ ID<br>NO: 847) | 没有 | 0.619599<br>4166 | 0 |  |
| 978  | TGACCTCC<br>TGTGTGGC<br>TACC (SEQ<br>ID NO: 331) | TGACCTC<br>CTGTGTG<br>GCTACC<br>(SEQ ID<br>NO: 848) | 没有 | 0.619599<br>4166 | 0 |  |
| 2663 | GCGGTGAA<br>GGGTGTTC<br>AACC (SEQ<br>ID NO: 332) | GCGGTGA<br>AGGGTGT<br>TCAACC<br>(SEQ ID<br>NO: 849) | 没有 | 0.619599<br>4166 | 0 |  |
| 189  | CAACAAGG<br>CGATGCCC<br>TCAG (SEQ<br>ID NO: 333) | CAACAAG<br>GCGATGC<br>CCTCAG<br>(SEQ ID<br>NO: 850) | 没有 | 0.619599<br>4166 | 0 |  |
| 2201 | GCGACCCA<br>TTCATCCTC<br>TTC (SEQ ID<br>NO: 334) | GCGACCC<br>ATTCTTC<br>CTCTTC<br>(SEQ ID<br>NO: 851) | 没有 | 0.619599<br>4166 | 0 |  |
| 332  | TATATCGAG<br>ATTTTCAA<br>GAA (SEQ<br>ID NO: 335) | TATATCGA<br>GATTTTC<br>AAGAA<br>(SEQ ID<br>NO: 852) | 没有 | 0.619599<br>4166 | 0 |  |
| 1873 | ATCACAAG<br>GAGCCAGG<br>TGCA (SEQ<br>ID NO: 336) | ATCACAA<br>GGAGCCA<br>GGTGCA<br>(SEQ ID<br>NO: 853) | 没有 | 0.619599<br>4166 | 0 |  |
| 1291 | AGGAAGTT<br>GGTCAAAG<br>CCTG (SEQ<br>ID NO: 337) | CGGAAGT<br>TGGTCAA<br>AGCCTG<br>(SEQ ID<br>NO: 854) | 没有 | 0.619599<br>4166 | 0 |  |

[0467]

|      |                                                  |                                                     |    |                  |                  |  |
|------|--------------------------------------------------|-----------------------------------------------------|----|------------------|------------------|--|
| 1413 | TAGGCAGC<br>TTGGTGAA<br>GAAG (SEQ<br>ID NO: 338) | TAGGCAG<br>CTTGGTG<br>AAGAAG<br>(SEQ ID<br>NO: 855) | 没有 | 0.619599<br>4166 | 0                |  |
| 2826 | GATTTTGA<br>GCCCTGTG<br>GCC (SEQ ID<br>NO: 339)  | GATTTT<br>GAGCCCT<br>GTGGCC<br>(SEQ ID<br>NO: 856)  | 没有 | 0.619599<br>4166 | 0.961889<br>9213 |  |
| 661  | ATCTTCAG<br>CCAGAGGC<br>GCGG (SEQ<br>ID NO: 340) | ATCTTCA<br>GCCAGAG<br>GCGCGG<br>(SEQ ID<br>NO: 857) | 没有 | 0.619599<br>4166 | 0                |  |
| 670  | CAGAGGCG<br>CGGGGCAA<br>AGAC (SEQ<br>ID NO: 341) | CAGAGGC<br>GCGGGGC<br>AAAGAC<br>(SEQ ID<br>NO: 858) | 没有 | 0.619599<br>4166 | 0                |  |
| 2550 | CCTTGATG<br>CTGCTGTC<br>TATG (SEQ<br>ID NO: 342) | CCTTGAT<br>GCTGCTG<br>TGTATG<br>(SEQ ID<br>NO: 859) | 没有 | 0.619599<br>4166 | 0                |  |
| 2304 | AGCCTGTA<br>GTGGTGTC<br>ATCT (SEQ<br>ID NO: 343) | AGCCTGT<br>AGTGGTG<br>TCATCT<br>(SEQ ID<br>NO: 860) | 没有 | 0.619599<br>4166 | 0                |  |
| 1633 | CTCTCTCTA<br>CTGGAGGA<br>AAA (SEQ<br>ID NO: 344) | CTCTCTC<br>TACTGGA<br>GGAAAA<br>(SEQ ID<br>NO: 861) | 没有 | 0.619599<br>4166 | 0                |  |
| 1546 | CTCCGCCT<br>GGTCAATC<br>AATA (SEQ<br>ID NO: 345) | CTCCGCC<br>TGGTCAA<br>TCAATA<br>(SEQ ID<br>NO: 862) | 没有 | 0.619599<br>4166 | 0.961889<br>9213 |  |

[0468]

|      |                                                  |                                                     |    |                  |                  |  |
|------|--------------------------------------------------|-----------------------------------------------------|----|------------------|------------------|--|
| 212  | GAATGCTG<br>CCGTGGCT<br>CCAG (SEQ<br>ID NO: 346) | GAATGCT<br>GCCGTGG<br>CTCCAG<br>(SEQ ID<br>NO: 863) | 没有 | 0.619599<br>4166 | 0                |  |
| 2575 | CTCGCTTG<br>GTACCTTG<br>ATCA (SEQ<br>ID NO: 347) | CTCGCTT<br>GGTACCT<br>TGATCA<br>(SEQ ID<br>NO: 864) | 没有 | 0.619599<br>4166 | 0                |  |
| 1498 | ATGGCCAA<br>CTTCGACT<br>GGAG (SEQ<br>ID NO: 348) | ATGGCCA<br>ACTTCGA<br>CTGGAG<br>(SEQ ID<br>NO: 865) | 没有 | 0.619599<br>4166 | 0                |  |
| 1759 | ACCAATAA<br>GATTAAAG<br>ACAG (SEQ<br>ID NO: 349) | ACCAATA<br>AGATTAA<br>AGACAG<br>(SEQ ID<br>NO: 866) | 没有 | 0.619599<br>4166 | 0                |  |
| 2975 | CTGTGCTC<br>GTTGGGGG<br>AAGG (SEQ<br>ID NO: 350) | CTGTGCT<br>CGTTGGG<br>GGAAGG<br>(SEQ ID<br>NO: 867) | 没有 | 0.619599<br>4166 | 0                |  |
| 3012 | GGATGCAG<br>TCCGGCAG<br>AGCC (SEQ<br>ID NO: 351) | GGATGCA<br>GTCCGGC<br>AGAGCC<br>(SEQ ID<br>NO: 868) | 没有 | 0.619599<br>4166 | 0                |  |
| 3142 | ATGGAAGC<br>CATGTTGG<br>AGGA (SEQ<br>ID NO: 352) | ATGGAAG<br>CCATGTT<br>GGAGGA<br>(SEQ ID<br>NO: 869) | 没有 | 0.619599<br>4166 | 0                |  |
| 2690 | AAAGAGCC<br>CTGGAAAA<br>GACC (SEQ<br>ID NO: 353) | AAAGAGC<br>CCTGGAA<br>AAGACC<br>(SEQ ID<br>NO: 870) | 没有 | 0.619599<br>4166 | 0.961889<br>9213 |  |

[0469]

|      |                                                  |                                                     |    |                  |   |  |
|------|--------------------------------------------------|-----------------------------------------------------|----|------------------|---|--|
| 642  | CCTCCTGG<br>AGCGCTTC<br>ATCA (SEQ<br>ID NO: 354) | CCTCCTG<br>GAGCGCT<br>TCATCA<br>(SEQ ID<br>NO: 871) | 没有 | 0.619599<br>4166 | 0 |  |
| 2028 | TGTGAAGA<br>GCATCGTC<br>TTGG (SEQ<br>ID NO: 355) | TGTGAAG<br>AGCATCG<br>TCTTGG<br>(SEQ ID<br>NO: 872) | 没有 | 0.619599<br>4166 | 0 |  |
| 2364 | GCAGGACC<br>GCATGACC<br>GCTG (SEQ<br>ID NO: 356) | GCAAGAC<br>CGCATGA<br>CCGCTG<br>(SEQ ID<br>NO: 873) | 没有 | 0.619599<br>4166 | 0 |  |
| 2716 | CTAACAGA<br>GGAAACGG<br>AGGA (SEQ<br>ID NO: 357) | CTAACAG<br>AGGAAAC<br>GGAGGA<br>(SEQ ID<br>NO: 874) | 没有 | 0.619599<br>4166 | 0 |  |
| 3024 | GCAGAGCC<br>TTGGCATG<br>TGTC (SEQ<br>ID NO: 358) | GCAGAGC<br>CTTGGCA<br>TGTGTC<br>(SEQ ID<br>NO: 875) | 没有 | 0.619599<br>4166 | 0 |  |
| 1820 | TCCGGTAC<br>ATCTGGGG<br>CGGG (SEQ<br>ID NO: 359) | TCCGGTA<br>CATCTGG<br>GGCGGG<br>(SEQ ID<br>NO: 876) | 没有 | 0.619599<br>4166 | 0 |  |
| 3039 | GTGTCCAC<br>AGCACAAC<br>ATCC (SEQ<br>ID NO: 360) | GTGTCCA<br>CAGCACA<br>ACATCC<br>(SEQ ID<br>NO: 877) | 没有 | 0.619599<br>4166 | 0 |  |
| 2645 | AAGAGTCG<br>TATTGGCTT<br>GGC (SEQ<br>ID NO: 361) | AAGAGTC<br>GTATTGG<br>CTTGGC<br>(SEQ ID<br>NO: 878) | 没有 | 0.619599<br>4166 | 0 |  |

[0470]

|      |                                                  |                                                     |    |                  |                  |  |
|------|--------------------------------------------------|-----------------------------------------------------|----|------------------|------------------|--|
| 2696 | CCCTGGAA<br>AAGACCGA<br>GCCC (SEQ<br>ID NO: 362) | CCCTGGA<br>AAAGACC<br>GAGCCC<br>(SEQ ID<br>NO: 879) | 没有 | 0.619599<br>4166 | 0                |  |
| 2759 | ACGACTCC<br>TTCTTCGA<br>GAGA (SEQ<br>ID NO: 363) | ACGACTC<br>CTTCTTC<br>GAGAGA<br>(SEQ ID<br>NO: 880) | 没有 | 0.619599<br>4166 | 0.871342<br>7566 |  |
| 2362 | TGGCAGGA<br>CCGCATGA<br>CCGC (SEQ<br>ID NO: 364) | TGGCAAG<br>ACCGCAT<br>GACCGC<br>(SEQ ID<br>NO: 881) | 没有 | 0.619599<br>4166 | 0                |  |
| 983  | TCCTGTGT<br>GGCTACCC<br>CGAG (SEQ<br>ID NO: 365) | TCCTGTG<br>TGGCTAC<br>CCCGAG<br>(SEQ ID<br>NO: 882) | 没有 | 0.619599<br>4166 | 0                |  |
| 740  | AAGACACT<br>CTGTATGCC<br>AAC (SEQ<br>ID NO: 366) | AAGACAC<br>TCTGTAT<br>GCCAAC<br>(SEQ ID<br>NO: 883) | 没有 | 0.619599<br>4166 | 0                |  |
| 907  | CTGTGGGT<br>GACCAGGC<br>CCCT (SEQ<br>ID NO: 367) | CTGTGGG<br>TGACCAG<br>GCCCCT<br>(SEQ ID<br>NO: 884) | 没有 | 0.619599<br>4166 | 0                |  |
| 750  | GTATGCCA<br>ACGTGGAC<br>TTCT (SEQ<br>ID NO: 368) | GTATGCC<br>AACGTGG<br>ACTTCT<br>(SEQ ID<br>NO: 885) | 没有 | 0.619599<br>4166 | 0                |  |
| 2788 | GGGTGGGT<br>TCCTGGGG<br>TATG (SEQ<br>ID NO: 369) | GGGTGGG<br>TTCCTGG<br>GGTGTG<br>(SEQ ID<br>NO: 886) | 没有 | 0.619599<br>4166 | 0.961889<br>9213 |  |

[0471]

|      |                                                  |                                                     |    |                  |   |  |
|------|--------------------------------------------------|-----------------------------------------------------|----|------------------|---|--|
| 3248 | TGGGCGAC<br>GCCAAGGT<br>TGTG (SEQ<br>ID NO: 370) | TGGGCGA<br>CGCCAAG<br>GTTGTG<br>(SEQ ID<br>NO: 887) | 没有 | 0.619599<br>4166 | 0 |  |
| 2798 | CTGGGGTA<br>TGCGTGAA<br>GAAT (SEQ<br>ID NO: 371) | CTGGGGT<br>GTGCGTG<br>AAGAAT<br>(SEQ ID<br>NO: 888) | 没有 | 0.619599<br>4166 | 0 |  |
| 3143 | TGGAAGCC<br>ATGTTGGA<br>GGAC (SEQ<br>ID NO: 372) | TGGAAGC<br>CATGTTG<br>GAGGAC<br>(SEQ ID<br>NO: 889) | 没有 | 0.619599<br>4166 | 0 |  |
| 2138 | TCATGTCG<br>ATGAGCAT<br>CTTC (SEQ<br>ID NO: 373) | TCATGTC<br>GATGAGC<br>ATCTTC<br>(SEQ ID<br>NO: 890) | 没有 | 0.619599<br>4166 | 0 |  |
| 2940 | CATCCTCA<br>CAGGCCTG<br>CTGC (SEQ<br>ID NO: 374) | CATCCTC<br>ACTGGCC<br>TGCTGC<br>(SEQ ID<br>NO: 891) | 没有 | 0.619599<br>4166 | 0 |  |
| 3300 | TTACAGCA<br>GAAGATCC<br>ATCT (SEQ<br>ID NO: 375) | TTACAGC<br>AGAAGAT<br>CCATCT<br>(SEQ ID<br>NO: 892) | 没有 | 0.619599<br>4166 | 0 |  |
| 2091 | CTCCAATG<br>CAGTGATT<br>TGGT (SEQ<br>ID NO: 376) | CTCCAAT<br>GCAGTGA<br>TTTGGT<br>(SEQ ID<br>NO: 893) | 没有 | 0.619599<br>4166 | 0 |  |
| 591  | CGCTCATG<br>GAGTCCCG<br>GACC (SEQ<br>ID NO: 377) | CGCTCAT<br>GGAGTCC<br>CGGACC<br>(SEQ ID<br>NO: 894) | 没有 | 0.619599<br>4166 | 0 |  |

[0472]

|      |                                                  |                                                      |    |                  |                  |  |
|------|--------------------------------------------------|------------------------------------------------------|----|------------------|------------------|--|
| 2532 | GTCCATGC<br>AGATGATG<br>CTCC (SEQ<br>ID NO: 378) | GTCCATG<br>CAGATGA<br>TGCTCC<br>(SEQ ID<br>NO: 895)  | 没有 | 0.619599<br>4166 | 0                |  |
| 3110 | AAGGAAAG<br>TCCCAGGA<br>GGAG (SEQ<br>ID NO: 379) | AAGGAA<br>AGTCCCA<br>GGAGGA<br>G (SEQ ID<br>NO: 896) | 没有 | 0.619599<br>4166 | 3.575918<br>191  |  |
| 288  | TCCTGGAA<br>TTGTGTCA<br>AACT (SEQ<br>ID NO: 380) | TCCTGGA<br>ATTGTGT<br>CAAAC<br>(SEQ ID<br>NO: 897)   | 没有 | 0.619599<br>4166 | 0                |  |
| 1758 | AACCAATA<br>AGATTAAA<br>GACA (SEQ<br>ID NO: 381) | AACCAAT<br>AAGATTA<br>AAGACA<br>(SEQ ID<br>NO: 898)  | 没有 | 0.619599<br>4166 | 0                |  |
| 2314 | GGTGTCAT<br>CTATTTAC<br>CCT (SEQ ID<br>NO: 382)  | GGTGTC<br>TCTATTT<br>ACCCT<br>(SEQ ID<br>NO: 899)    | 没有 | 0.619599<br>4166 | 0.871342<br>7566 |  |
| 2710 | GAGCCCCT<br>AACAGAGG<br>AAAC (SEQ<br>ID NO: 383) | GAGCCCC<br>TAACAGA<br>GGAAAC<br>(SEQ ID<br>NO: 900)  | 没有 | 0.619599<br>4166 | 0                |  |
| 2192 | TACATTACA<br>GCGACCCA<br>TTC (SEQ ID<br>NO: 384) | TACATTA<br>CAGCGAC<br>CCATTC<br>(SEQ ID<br>NO: 901)  | 没有 | 0.619599<br>4166 | 0                |  |
| 1868 | AGGGGATC<br>ACAAGGAG<br>CCAG (SEQ<br>ID NO: 385) | AGGGGAT<br>CACAAGG<br>AGCCAG<br>(SEQ ID<br>NO: 902)  | 没有 | 0.619599<br>4166 | 0                |  |

[0473]

|      |                                                   |                                                     |    |                  |                  |  |
|------|---------------------------------------------------|-----------------------------------------------------|----|------------------|------------------|--|
| 2809 | GTGAAGAA<br>TCTGGTAA<br>AGAT (SEQ<br>ID NO: 386)  | GTGAAGA<br>ATCTGGT<br>GAAGAT<br>(SEQ ID<br>NO: 903) | 没有 | 0.619599<br>4166 | 0                |  |
| 3091 | CTGTTCTAT<br>GCCCAGCT<br>GAA (SEQ<br>ID NO: 387)  | CTGTTCT<br>ATGCACA<br>ACTGAA<br>(SEQ ID<br>NO: 904) | 没有 | 0.619599<br>4166 | 0.961889<br>9213 |  |
| 753  | TGCCAACG<br>TGGACTION<br>TTCA (SEQ<br>ID NO: 388) | TGCCAAC<br>GTGGACT<br>TCTTCA<br>(SEQ ID<br>NO: 905) | 没有 | 0.619599<br>4166 | 0                |  |
| 980  | ACCTCCTG<br>TGTGGCTA<br>CCCC (SEQ<br>ID NO: 389)  | ACCTCCT<br>GTGTGGC<br>TACCCC<br>(SEQ ID<br>NO: 906) | 没有 | 0.619599<br>4166 | 0                |  |
| 1461 | CCTCTACA<br>AGGGCCCT<br>CGGG (SEQ<br>ID NO: 390)  | CCTCTAC<br>AAGGGCC<br>CTCGGG<br>(SEQ ID<br>NO: 907) | 没有 | 0.619599<br>4166 | 0                |  |
| 2413 | CCGGTGGC<br>ATTTGGATT<br>TGG (SEQ ID<br>NO: 391)  | CCGGTGG<br>CATTTGG<br>ATTTGG<br>(SEQ ID<br>NO: 908) | 没有 | 0.619599<br>4166 | 0                |  |
| 3213 | CATGCAGA<br>GAAACTG<br>AGCG (SEQ<br>ID NO: 392)   | CATGCAG<br>AGAAAC<br>TGAGCG<br>(SEQ ID<br>NO: 909)  | 没有 | 0.619599<br>4166 | 0                |  |
| 3089 | TGCTGTTCT<br>ATGCCAG<br>CTG (SEQ ID<br>NO: 393)   | TGCTGTT<br>CTATGCA<br>CAACTG<br>(SEQ ID<br>NO: 910) | 没有 | 0.619599<br>4166 | 0                |  |

[0474]

|      |                                                  |                                                     |    |                  |                  |  |
|------|--------------------------------------------------|-----------------------------------------------------|----|------------------|------------------|--|
| 2860 | CGTCTGAA<br>CATCACCTT<br>CTA (SEQ ID<br>NO: 394) | CGTCTGA<br>ACATCAC<br>CTTCTA<br>(SEQ ID<br>NO: 911) | 没有 | 0.619599<br>4166 | 0                |  |
| 3068 | ACCTCACG<br>GTGGCTGA<br>GCAC (SEQ<br>ID NO: 395) | ACCTCAC<br>GGTGGCT<br>GAGCAC<br>(SEQ ID<br>NO: 912) | 没有 | 0.619599<br>4166 | 0                |  |
| 2841 | TGGCCGGC<br>CAGCTGTG<br>GACC (SEQ<br>ID NO: 396) | TGGCCGG<br>CCAGCTG<br>TGGACC<br>(SEQ ID<br>NO: 913) | 没有 | 0.619599<br>4166 | 0.871342<br>7566 |  |
| 2965 | ACCTCTGG<br>GACTGTGC<br>TCGT (SEQ<br>ID NO: 397) | ACCTCTG<br>GGACTGT<br>GCTCGT<br>(SEQ ID<br>NO: 914) | 没有 | 0.508030<br>4371 | 0.961889<br>9213 |  |
| 2859 | CCGTCTGA<br>ACATCACC<br>TTCT (SEQ<br>ID NO: 398) | CCGTCTG<br>AACATCA<br>CCTTCT<br>(SEQ ID<br>NO: 915) | 没有 | 0.508030<br>4371 | 0                |  |
| 2283 | CTCCAAGG<br>CCAGTCTG<br>GCAG (SEQ<br>ID NO: 399) | CTCCAAG<br>GCCAGTC<br>TGGCAG<br>(SEQ ID<br>NO: 916) | 没有 | 0.508030<br>4371 | 0.871342<br>7566 |  |
| 3131 | CCCAGCTG<br>GAGATGGA<br>AGCC (SEQ<br>ID NO: 400) | CCCAGCT<br>GGAGATG<br>GAAGCC<br>(SEQ ID<br>NO: 917) | 没有 | 0.508030<br>4371 | 0                |  |
| 2946 | CACAGGCC<br>TGCTGCCT<br>CCAA (SEQ<br>ID NO: 401) | CACTGGC<br>CTGCTGC<br>CTCCAA<br>(SEQ ID<br>NO: 918) | 没有 | 0.508030<br>4371 | 0                |  |

[0475]

|      |                                                  |                                                     |    |                  |                 |  |
|------|--------------------------------------------------|-----------------------------------------------------|----|------------------|-----------------|--|
| 2730 | GGAGGATC<br>CAGAGCAC<br>CCAG (SEQ<br>ID NO: 402) | GGAGGAT<br>CCAGAGC<br>ACCCAG<br>(SEQ ID<br>NO: 919) | 没有 | 0.508030<br>4371 | 0               |  |
| 2971 | GGGACTGT<br>GCTCGTTG<br>GGGG (SEQ<br>ID NO: 403) | GGGACTG<br>TGCTCGT<br>TGGGGG<br>(SEQ ID<br>NO: 920) | 没有 | 0.508030<br>4371 | 0               |  |
| 1797 | CAGAGCTG<br>ATCCCGTG<br>GAAG (SEQ<br>ID NO: 404) | CAGAGCT<br>GATCCCG<br>TGGAAG<br>(SEQ ID<br>NO: 921) | 没有 | 0.508030<br>4371 | 0               |  |
| 2624 | TTCCTTGGT<br>ACTTTCTTC<br>TA (SEQ ID<br>NO: 405) | TTCCTTG<br>GTACTTT<br>CTTCTA<br>(SEQ ID<br>NO: 922) | 没有 | 0.508030<br>4371 | 0               |  |
| 1472 | GCCCTCGG<br>GAAAGCCA<br>GGCT (SEQ<br>ID NO: 406) | GCCCTCG<br>GGAAAGC<br>CAGGCT<br>(SEQ ID<br>NO: 923) | 没有 | 0.508030<br>4371 | 1.923779<br>843 |  |
| 1765 | AAGATTAA<br>AGACAGGT<br>ATTG (SEQ<br>ID NO: 407) | AAGATTA<br>AAGACAG<br>ATATTG<br>(SEQ ID<br>NO: 924) | 没有 | 0.508030<br>4371 | 0               |  |
| 1385 | GGAACCCA<br>ACAGTAAA<br>AGAC (SEQ<br>ID NO: 408) | GGAACCC<br>AACAGTA<br>AAAGAC<br>(SEQ ID<br>NO: 925) | 没有 | 0.508030<br>4371 | 0               |  |
| 1553 | TGGTCAAT<br>CAATACCT<br>GGAG (SEQ<br>ID NO: 409) | TGGTCAA<br>TCAATAC<br>CTGGAG<br>(SEQ ID<br>NO: 926) | 没有 | 0.508030<br>4371 | 0               |  |

[0476]

|      |                                                  |                                                     |    |                  |                  |  |
|------|--------------------------------------------------|-----------------------------------------------------|----|------------------|------------------|--|
| 2824 | AAGATTTT<br>GAGCCCTG<br>TGG (SEQ ID<br>NO: 410)  | AAGATTT<br>TTGAGCC<br>CTGTGG<br>(SEQ ID<br>NO: 927) | 没有 | 0.508030<br>4371 | 0.961889<br>9213 |  |
| 273  | CACCCCAG<br>GAGAATCT<br>CCTG (SEQ<br>ID NO: 411) | CACCCA<br>GGAGAAT<br>CTCCTG<br>(SEQ ID<br>NO: 928)  | 没有 | 0.508030<br>4371 | 0                |  |
| 3060 | GTTCCACC<br>ACCTCACG<br>GTGG (SEQ<br>ID NO: 412) | GTTCCAC<br>CACCTCA<br>CGGTGG<br>(SEQ ID<br>NO: 929) | 没有 | 0.508030<br>4371 | 0                |  |
| 2149 | AGCATCTT<br>CCTCCTGA<br>CGAT (SEQ<br>ID NO: 413) | AGCATCT<br>TCCTCCT<br>GACGAT<br>(SEQ ID<br>NO: 930) | 没有 | 0.508030<br>4371 | 0                |  |
| 2881 | GAGAACCA<br>GATCACCG<br>CATT (SEQ<br>ID NO: 414) | GAGAACC<br>AGATCAC<br>CGCATT<br>(SEQ ID<br>NO: 931) | 没有 | 0.508030<br>4371 | 0                |  |
| 2797 | CCTGGGGT<br>ATGCGTGA<br>AGAA (SEQ<br>ID NO: 415) | CCTGGGG<br>TGTGCGT<br>GAAGAA<br>(SEQ ID<br>NO: 932) | 没有 | 0.508030<br>4371 | 0                |  |
| 2733 | GGATCCAG<br>AGCACCCA<br>GAAG (SEQ<br>ID NO: 416) | GGATCCA<br>GAGCACC<br>CAGAAG<br>(SEQ ID<br>NO: 933) | 没有 | 0.508030<br>4371 | 0                |  |
| 798  | CCTAGACA<br>GCCGTTCT<br>CAAG (SEQ<br>ID NO: 417) | CCTAGAC<br>AGCCGTT<br>CTCAAG<br>(SEQ ID<br>NO: 934) | 没有 | 0.508030<br>4371 | 0                |  |

[0477]

|      |                                                  |                                                     |     |                  |                  |  |
|------|--------------------------------------------------|-----------------------------------------------------|-----|------------------|------------------|--|
| 2849 | CAGCTGTG<br>GACCGTCT<br>GAAC (SEQ<br>ID NO: 418) | CAGCTGT<br>GGACCGT<br>CTGAAC<br>(SEQ ID<br>NO: 935) | 没有  | 0.508030<br>4371 | 0                |  |
| 3117 | GTCCCAGG<br>AGGAGGCC<br>CAGC (SEQ<br>ID NO: 419) | GTCCCAG<br>GAGGAG<br>GCCCAGC<br>(SEQ ID<br>NO: 936) | 没有  | 0.508030<br>4371 | 2.795122<br>599  |  |
| 1780 | TATTGGGAT<br>TCTGGTCC<br>CAG (SEQ<br>ID NO: 420) | TATTGGG<br>ATTCTGG<br>TCCCAG<br>(SEQ ID<br>NO: 937) | 没有  | 0.508030<br>4371 | 0                |  |
| 2817 | TCTGGTAA<br>AGATTTTGG<br>AGC (SEQ<br>ID NO: 421) | TCTGGTG<br>AAGATT<br>TTGAGC<br>(SEQ ID<br>NO: 938)  | 分支点 | 0.508030<br>4371 | 0.871342<br>7566 |  |
| 2894 | CCGCATTC<br>CTGGGCCA<br>CAAT (SEQ<br>ID NO: 422) | CCGCATT<br>CCTGGGC<br>CACAAT<br>(SEQ ID<br>NO: 939) | 没有  | 0.508030<br>4371 | 1.833232<br>678  |  |
| 2807 | GCGTGAAG<br>AATCTGGT<br>AAAG (SEQ<br>ID NO: 423) | GCGTGAA<br>GAATCTG<br>GTGAAG<br>(SEQ ID<br>NO: 940) | 没有  | 0.508030<br>4371 | 0                |  |
| 3008 | GCCTGGAT<br>GCAGTCCG<br>GCAG (SEQ<br>ID NO: 424) | GCCTGGA<br>TGCAGTC<br>CGGCAG<br>(SEQ ID<br>NO: 941) | 没有  | 0.508030<br>4371 | 0                |  |
| 2785 | CCCGGGTG<br>GGTTCCTG<br>GGGT (SEQ<br>ID NO: 425) | CCCGGGT<br>GGGTTC<br>TGGGGT<br>(SEQ ID<br>NO: 942)  | 没有  | 0.508030<br>4371 | 0                |  |

[0478]

|      |                                                  |                                                     |    |                  |                  |  |
|------|--------------------------------------------------|-----------------------------------------------------|----|------------------|------------------|--|
| 608  | ACCTGGCG<br>CTGAAGGA<br>CATC (SEQ<br>ID NO: 426) | ACCTGGC<br>GCTGAAG<br>GACATC<br>(SEQ ID<br>NO: 943) | 没有 | 0.508030<br>4371 | 0                |  |
| 1713 | CCACGTGA<br>AGTATAAG<br>ATCC (SEQ<br>ID NO: 427) | CCACGTG<br>AAGTATA<br>AGATCC<br>(SEQ ID<br>NO: 944) | 没有 | 0.508030<br>4371 | 0                |  |
| 2944 | CTCACAGG<br>CCTGCTGC<br>CTCC (SEQ<br>ID NO: 428) | CTCACTG<br>GCCTGCT<br>GCCTCC<br>(SEQ ID<br>NO: 945) | 没有 | 0.508030<br>4371 | 0                |  |
| 2066 | AGGAGACT<br>TTGAAAAA<br>TCAG (SEQ<br>ID NO: 429) | AGGAGAC<br>ATTGAAA<br>AATCAG<br>(SEQ ID<br>NO: 946) | 没有 | 0.508030<br>4371 | 0.871342<br>7566 |  |
| 2639 | TTCTACAA<br>GAGTCGTA<br>TTGG (SEQ<br>ID NO: 430) | TTCTACA<br>AGAGTCG<br>TATTGG<br>(SEQ ID<br>NO: 947) | 没有 | 0.508030<br>4371 | 0                |  |
| 237  | CTTCTGCA<br>ATGTGAAC<br>AATC (SEQ<br>ID NO: 431) | CTTCTGC<br>AATGTGA<br>ACAATC<br>(SEQ ID<br>NO: 948) | 没有 | 0.508030<br>4371 | 0                |  |
| 1870 | GGGATCAC<br>AAGGAGCC<br>AGGT (SEQ<br>ID NO: 432) | GGGATCA<br>CAAGGAG<br>CCAGGT<br>(SEQ ID<br>NO: 949) | 没有 | 0.508030<br>4371 | 0                |  |
| 3099 | TGCCCAGC<br>TGAAAGGA<br>AAGT (SEQ<br>ID NO: 433) | TGCACAA<br>CTGAAAG<br>GAAAGT<br>(SEQ ID<br>NO: 950) | 没有 | 0.508030<br>4371 | 0                |  |

[0479]

|      |                                                  |                                                     |     |                  |                  |  |
|------|--------------------------------------------------|-----------------------------------------------------|-----|------------------|------------------|--|
| 1841 | TTGCCTATC<br>TGCAGGAC<br>ATG (SEQ ID<br>NO: 434) | TTGCATAT<br>CTGCAAG<br>ACATG<br>(SEQ ID<br>NO: 951) | 分支点 | 0.508030<br>4371 | 0                |  |
| 17   | GCAGCGGC<br>TTCGTGAG<br>ACAG (SEQ<br>ID NO: 435) | GATCCGG<br>CTTCGTG<br>AGACAG<br>(SEQ ID<br>NO: 952) | 没有  | 0.508030<br>4371 | 82.58601<br>767  |  |
| 3151 | ATGTTGGA<br>GGACACAG<br>GCCT (SEQ<br>ID NO: 436) | ATGTTGG<br>AGGACAC<br>AGGCCT<br>(SEQ ID<br>NO: 953) | 没有  | 0.508030<br>4371 | 0                |  |
| 1400 | AAGACTTT<br>TTGAATAG<br>GCAG (SEQ<br>ID NO: 437) | AAGACTT<br>TTTGAAT<br>AGGCAG<br>(SEQ ID<br>NO: 954) | 没有  | 0.508030<br>4371 | 3.575918<br>191  |  |
| 3062 | TCCACCAC<br>CTCACGGT<br>GGCT (SEQ<br>ID NO: 438) | TCCACCA<br>CCTCACG<br>GTGGCT<br>(SEQ ID<br>NO: 955) | 没有  | 0.508030<br>4371 | 0                |  |
| 2769 | CTTCGAGA<br>GAGAACAC<br>CCCG (SEQ<br>ID NO: 439) | CTTCGAG<br>AGAGAAC<br>ACCCCG<br>(SEQ ID<br>NO: 956) | 没有  | 0.508030<br>4371 | 0.871342<br>7566 |  |
| 2032 | AAGAGCAT<br>CGTCTTGG<br>AGAA (SEQ<br>ID NO: 440) | AAGAGCA<br>TCGTCTT<br>GGAGAA<br>(SEQ ID<br>NO: 957) | 没有  | 0.508030<br>4371 | 0                |  |
| 181  | CATTTCCCC<br>AACAAGGC<br>GAT (SEQ ID<br>NO: 441) | CATTTCC<br>CCAACAA<br>GGCGAT<br>(SEQ ID<br>NO: 958) | 没有  | 0.508030<br>4371 | 0                |  |

[0480]

|      |                                                  |                                                     |    |                  |                  |  |
|------|--------------------------------------------------|-----------------------------------------------------|----|------------------|------------------|--|
| 854  | ATATGTCAC<br>CAAGAATT<br>CAA (SEQ ID<br>NO: 442) | ATATGTC<br>ACCAAGA<br>ATTCAA<br>(SEQ ID<br>NO: 959) | 没有 | 0.508030<br>4371 | 0                |  |
| 1899 | GGCTCCAG<br>TTGGAATC<br>TACC (SEQ<br>ID NO: 443) | GGCTCCA<br>GTTGGAA<br>TCTACC<br>(SEQ ID<br>NO: 960) | 没有 | 0.508030<br>4371 | 0                |  |
| 2426 | GATTTGGC<br>ACTGAGTA<br>CCTG (SEQ<br>ID NO: 444) | GATTTGG<br>CACTGAG<br>TACCTG<br>(SEQ ID<br>NO: 961) | 没有 | 0.508030<br>4371 | 0                |  |
| 3077 | TGGCTGAG<br>CACATGCT<br>GTTC (SEQ<br>ID NO: 445) | TGGCTGA<br>GCACATG<br>CTGTTC<br>(SEQ ID<br>NO: 962) | 没有 | 0.508030<br>4371 | 0                |  |
| 2286 | CAAGGCCA<br>GTCTGGCA<br>GCAG (SEQ<br>ID NO: 446) | CAAGGCC<br>AGTCTGG<br>CAGCAG<br>(SEQ ID<br>NO: 963) | 没有 | 0.508030<br>4371 | 0                |  |
| 3308 | GAAGATCC<br>ATCTGGGA<br>TCTG (SEQ<br>ID NO: 447) | GAAGATC<br>CATCTGG<br>GATCTG<br>(SEQ ID<br>NO: 964) | 没有 | 0.508030<br>4371 | 0                |  |
| 927  | CATGCAGA<br>ATGGTGGT<br>CCAG (SEQ<br>ID NO: 448) | CATGCAG<br>AATGGTG<br>GTCCAG<br>(SEQ ID<br>NO: 965) | 没有 | 0.508030<br>4371 | 0.871342<br>7566 |  |
| 3114 | AAAGTCCC<br>AGGAGGAG<br>GCCC (SEQ<br>ID NO: 449) | AAAGTCC<br>CAGGAGG<br>AGGCCC<br>(SEQ ID<br>NO: 966) | 没有 | 0.508030<br>4371 | 0                |  |

[0481]

|      |                                                  |                                                     |    |                  |                  |  |
|------|--------------------------------------------------|-----------------------------------------------------|----|------------------|------------------|--|
| 908  | TGTGGGTG<br>ACCAGGCC<br>CCTC (SEQ<br>ID NO: 450) | TGTGGGT<br>GACCAGG<br>CCCCTC<br>(SEQ ID<br>NO: 967) | 没有 | 0.508030<br>4371 | 0                |  |
| 2127 | CAGCTTCT<br>CCATCATGT<br>CGA (SEQ ID<br>NO: 451) | CTCCTTC<br>TCCATCA<br>TGTCGA<br>(SEQ ID<br>NO: 968) | 没有 | 0.508030<br>4371 | 0                |  |
| 1886 | AGGTGCAG<br>GCGGAGGC<br>TCCA (SEQ<br>ID NO: 452) | AGGTGCA<br>GGCGGAG<br>GCTCCA<br>(SEQ ID<br>NO: 969) | 没有 | 0.508030<br>4371 | 0                |  |
| 2603 | CAGGAGAC<br>TATGGAAC<br>CCCA (SEQ<br>ID NO: 453) | CAGGAGA<br>CTATGGA<br>ACCCCA<br>(SEQ ID<br>NO: 970) | 没有 | 0.508030<br>4371 | 0                |  |
| 3109 | AAAGGAAA<br>GTCCCAGG<br>AGGA (SEQ<br>ID NO: 454) | AAAGGA<br>AAGTCCC<br>AGGAGGA<br>(SEQ ID<br>NO: 971) | 没有 | 0.508030<br>4371 | 0.871342<br>7566 |  |
| 2851 | GCTGTGGA<br>CCGTCTGA<br>ACAT (SEQ<br>ID NO: 455) | GCTGTGG<br>ACCGTCT<br>GAACAT<br>(SEQ ID<br>NO: 972) | 没有 | 0.508030<br>4371 | 0                |  |
| 3108 | GAAAGGAA<br>AGTCCCAG<br>GAGG (SEQ<br>ID NO: 456) | GAAAGG<br>AAAGTCC<br>CAGGAGG<br>(SEQ ID<br>NO: 973) | 没有 | 0.508030<br>4371 | 0                |  |
| 738  | AGAAGACA<br>CTCTGTATG<br>CCA (SEQ ID<br>NO: 457) | AGAAGAC<br>ACTCTGT<br>ATGCCA<br>(SEQ ID<br>NO: 974) | 没有 | 0.508030<br>4371 | 0                |  |

[0482]

|      |                                                  |                                                      |    |                  |   |  |
|------|--------------------------------------------------|------------------------------------------------------|----|------------------|---|--|
| 2614 | GGAACCCC<br>ACTTCCTT<br>GGTA (SEQ<br>ID NO: 458) | GGAACCC<br>CACTTCC<br>TTGGTA<br>(SEQ ID<br>NO: 975)  | 没有 | 0.508030<br>4371 | 0 |  |
| 3095 | TCTATGCCC<br>AGCTGAAA<br>GGA (SEQ<br>ID NO: 459) | TCTATGC<br>ACAACTG<br>AAAGGA<br>(SEQ ID<br>NO: 976)  | 没有 | 0.508030<br>4371 | 0 |  |
| 3016 | GCAGTCCG<br>GCAGAGCC<br>TTGG (SEQ<br>ID NO: 460) | GCAGTCC<br>GGCAGAG<br>CCTTGG<br>(SEQ ID<br>NO: 977)  | 没有 | 0.508030<br>4371 | 0 |  |
| 448  | GAGAGAAT<br>TGCAGGAA<br>GAGG (SEQ<br>ID NO: 461) | GAGAGA<br>ATTGCAG<br>GAAGAG<br>G (SEQ ID<br>NO: 978) | 没有 | 0.508030<br>4371 | 0 |  |
| 1510 | GACTGGAG<br>GGACATATT<br>TAA (SEQ ID<br>NO: 462) | GACTGGA<br>GGGACAT<br>ATTTAA<br>(SEQ ID<br>NO: 979)  | 没有 | 0.508030<br>4371 | 0 |  |
| 2503 | ACGGAAGG<br>GGACGAAT<br>TCAG (SEQ<br>ID NO: 463) | ACGGAAG<br>GGGACGA<br>ATTCAG<br>(SEQ ID<br>NO: 980)  | 没有 | 0.508030<br>4371 | 0 |  |
| 2781 | ACACCCCG<br>GGTGGGTT<br>CCTG (SEQ<br>ID NO: 464) | ACACCCC<br>GGGTGGG<br>TTCCTG<br>(SEQ ID<br>NO: 981)  | 没有 | 0.508030<br>4371 | 0 |  |
| 607  | GACCTGGC<br>GCTGAAGG<br>ACAT (SEQ<br>ID NO: 465) | GACCTGG<br>CGCTGAA<br>GGACAT<br>(SEQ ID<br>NO: 982)  | 没有 | 0.508030<br>4371 | 0 |  |

[0483]

|      |                                                  |                                                     |    |                  |                  |  |
|------|--------------------------------------------------|-----------------------------------------------------|----|------------------|------------------|--|
| 3201 | CCTATCAG<br>GCGGCATG<br>CAGA (SEQ<br>ID NO: 466) | CCTATCA<br>GGCGGCA<br>TGCAGA<br>(SEQ ID<br>NO: 983) | 没有 | 0.508030<br>4371 | 0                |  |
| 1298 | TGGTCAAA<br>GCCTGGGA<br>AGAA (SEQ<br>ID NO: 467) | TGGTCAA<br>AGCCTGG<br>GAAGAA<br>(SEQ ID<br>NO: 984) | 没有 | 0.508030<br>4371 | 0                |  |
| 1243 | AGGATACT<br>GAAGAATG<br>CCAA (SEQ<br>ID NO: 468) | AGGATAC<br>TGAAGAA<br>TGCCAA<br>(SEQ ID<br>NO: 985) | 没有 | 0.508030<br>4371 | 0                |  |
| 1570 | GAGTGCTT<br>GGTCCTGG<br>ATAA (SEQ<br>ID NO: 469) | GAGTGCT<br>TGGTCCT<br>GGATAA<br>(SEQ ID<br>NO: 986) | 没有 | 0.508030<br>4371 | 0                |  |
| 2920 | GGGAAAAC<br>CACCACCT<br>TGTC (SEQ<br>ID NO: 470) | GGGAAA<br>ACCACCA<br>CCTTGTC<br>(SEQ ID<br>NO: 987) | 没有 | 0.508030<br>4371 | 0                |  |
| 2914 | GGAGCTGG<br>GAAAACCA<br>CCAC (SEQ<br>ID NO: 471) | GGAGCTG<br>GGAAAAC<br>CACCAC<br>(SEQ ID<br>NO: 988) | 没有 | 0.508030<br>4371 | 0.871342<br>7566 |  |
| 1819 | TTCCGGTA<br>CATCTGGG<br>GCGG (SEQ<br>ID NO: 472) | TTCCGGT<br>ACATCTG<br>GGGCGG<br>(SEQ ID<br>NO: 989) | 没有 | 0.508030<br>4371 | 0                |  |
| 2601 | TCCAGGAG<br>ACTATGGA<br>ACCC (SEQ<br>ID NO: 473) | TCCAGGA<br>GACTATG<br>GAACCC<br>(SEQ ID<br>NO: 990) | 没有 | 0.508030<br>4371 | 0                |  |

[0484]

|      |                                                  |                                                      |     |                  |                  |  |
|------|--------------------------------------------------|------------------------------------------------------|-----|------------------|------------------|--|
| 3047 | AGCACAAC<br>ATCCTGTT<br>CAC (SEQ ID<br>NO: 474)  | AGCACAA<br>CATCCTG<br>TTCCAC<br>(SEQ ID<br>NO: 991)  | 没有  | 0.508030<br>4371 | 0                |  |
| 2800 | GGGGTATG<br>CGTGAAGA<br>ATCT (SEQ<br>ID NO: 475) | GGGGTGT<br>GCGTGAA<br>GAATCT<br>(SEQ ID<br>NO: 992)  | 没有  | 0.508030<br>4371 | 0                |  |
| 1834 | GGCGGGTT<br>TGCCTATCT<br>GCA (SEQ ID<br>NO: 476) | GGCGGGT<br>TTGCATAT<br>CTGCA<br>(SEQ ID<br>NO: 993)  | 分支点 | 0.508030<br>4371 | 0                |  |
| 309  | TAACAACT<br>CCATCTTG<br>GCAA (SEQ<br>ID NO: 477) | TAACAACT<br>TCCATCT<br>TGGCAA<br>(SEQ ID<br>NO: 994) | 没有  | 0.508030<br>4371 | 0                |  |
| 2823 | AAAGATT<br>TTGAGCCC<br>TGTG (SEQ<br>ID NO: 478)  | GAAGATT<br>TTTGAGC<br>CCTGTG<br>(SEQ ID<br>NO: 995)  | 没有  | 0.508030<br>4371 | 0                |  |
| 3229 | AGCGTGGC<br>CATTGCCTT<br>CGT (SEQ ID<br>NO: 479) | AGCGTGG<br>CCATTGC<br>CTTCGT<br>(SEQ ID<br>NO: 996)  | 没有  | 0.508030<br>4371 | 0                |  |
| 2564 | TCTATGGCT<br>TACTCGCTT<br>GG (SEQ ID<br>NO: 480) | TGTATGG<br>CTTGCTC<br>GCTTGG<br>(SEQ ID<br>NO: 997)  | 分支点 | 0.508030<br>4371 | 0                |  |
| 713  | TCTCCCAG<br>GGCACCCT<br>ACAG (SEQ<br>ID NO: 481) | TCTCCCA<br>GGGCACC<br>CTACAG<br>(SEQ ID<br>NO: 998)  | 没有  | 0.508030<br>4371 | 0.961889<br>9213 |  |

[0485]

|      |                                                  |                                                      |     |                  |   |  |
|------|--------------------------------------------------|------------------------------------------------------|-----|------------------|---|--|
| 1433 | GTATTACTG<br>CTGAAGCC<br>ATC (SEQ ID<br>NO: 482) | GCATTAC<br>TGCGGAG<br>GCCATC<br>(SEQ ID<br>NO: 999)  | 分支点 | 0.508030<br>4371 | 0 |  |
| 2714 | CCCTAACA<br>GAGGAAAC<br>GGAG (SEQ<br>ID NO: 483) | CCCTAAC<br>AGAGGA<br>AACGGAG<br>(SEQ ID<br>NO: 1000) | 没有  | 0.508030<br>4371 | 0 |  |
| 23   | GCTTCGTG<br>AGACAGAT<br>ACAG (SEQ<br>ID NO: 484) | GCTTCGT<br>GAGACAG<br>ATACAG<br>(SEQ ID<br>NO: 1001) | 没有  | 0.508030<br>4371 | 0 |  |
| 1597 | AGCTACAA<br>TGATGAAA<br>CTCA (SEQ<br>ID NO: 485) | AGCTACA<br>ATGATGA<br>AACTCA<br>(SEQ ID<br>NO: 1002) | 没有  | 0.508030<br>4371 | 0 |  |
| 2343 | ACACATCC<br>TGTGCTTC<br>GCCT (SEQ<br>ID NO: 486) | CCACATC<br>CTGTGCT<br>TCGCCT<br>(SEQ ID<br>NO: 1003) | 没有  | 0.508030<br>4371 | 0 |  |
| 2351 | TGTGCTTC<br>GCCTGGCA<br>GGAC (SEQ<br>ID NO: 487) | TGTGCTT<br>CGCCTGG<br>CAAGAC<br>(SEQ ID<br>NO: 1004) | 没有  | 0.508030<br>4371 | 0 |  |
| 2642 | TACAAGAG<br>TCGTATTGG<br>CTT (SEQ ID<br>NO: 488) | TACAAGA<br>GTCGTAT<br>TGGCTT<br>(SEQ ID<br>NO: 1005) | 没有  | 0.508030<br>4371 | 0 |  |
| 2239 | TCCACTGC<br>CACCATCAT<br>GCT (SEQ ID<br>NO: 489) | TCCACTG<br>CCACCAT<br>CATGCT<br>(SEQ ID<br>NO: 1006) | 没有  | 0.508030<br>4371 | 0 |  |

[0486]

|      |                                                  |                                                      |     |                  |                  |  |
|------|--------------------------------------------------|------------------------------------------------------|-----|------------------|------------------|--|
| 2223 | GTTCTTGTT<br>GGCTTTCT<br>CCA (SEQ ID<br>NO: 490) | GTTCTTG<br>TTGGCTT<br>TCTCCA<br>(SEQ ID<br>NO: 1007) | 没有  | 0.508030<br>4371 | 0                |  |
| 441  | TCACCCGG<br>AGAGAATT<br>GCAG (SEQ<br>ID NO: 491) | TCACCCG<br>GAGAGA<br>ATTGCAG<br>(SEQ ID<br>NO: 1008) | 没有  | 0.508030<br>4371 | 0.871342<br>7566 |  |
| 1189 | AAGCCTTT<br>GCTGATGG<br>GAAA (SEQ<br>ID NO: 492) | AAGCCTT<br>TGCTGAT<br>GGGAAA<br>(SEQ ID<br>NO: 1009) | 没有  | 0.508030<br>4371 | 0                |  |
| 2555 | ATGCTGCT<br>GTCTATGG<br>CTTA (SEQ<br>ID NO: 493) | ATGCTGC<br>TGTGTAT<br>GGCTTG<br>(SEQ ID<br>NO: 1010) | 分支点 | 0.508030<br>4371 | 0                |  |
| 1636 | TCTCTACTG<br>GAGGAAAA<br>CAT (SEQ ID<br>NO: 494) | TCTCTAC<br>TGGAGGA<br>AAACAT<br>(SEQ ID<br>NO: 1011) | 没有  | 0.508030<br>4371 | 0                |  |
| 229  | CAGGGGAT<br>CTTCTGCA<br>ATGT (SEQ<br>ID NO: 495) | CAGGGGA<br>TCTTCTG<br>CAATGT<br>(SEQ ID<br>NO: 1012) | 没有  | 0.508030<br>4371 | 0.871342<br>7566 |  |
| 705  | GTGCTCCC<br>TCTCCCAG<br>GGCA (SEQ<br>ID NO: 496) | GTGCTCC<br>CTCTCCC<br>AGGGCA<br>(SEQ ID<br>NO: 1013) | 没有  | 0.508030<br>4371 | 0                |  |
| 1747 | GTGGTGGA<br>GAAAACCA<br>ATAA (SEQ<br>ID NO: 497) | GTGGTG<br>GAGAAAAC<br>CAATAA<br>(SEQ ID<br>NO: 1014) | 没有  | 0.508030<br>4371 | 0                |  |

[0487]

|      |                                                  |                                                      |    |                  |                  |  |
|------|--------------------------------------------------|------------------------------------------------------|----|------------------|------------------|--|
| 1318 | GTAGGGCC<br>CCAGATCT<br>GGTA (SEQ<br>ID NO: 498) | GTGGGGC<br>CCCAGAT<br>CTGGTA<br>(SEQ ID<br>NO: 1015) | 没有 | 0.508030<br>4371 | 0.961889<br>9213 |  |
| 3015 | TGCAGTCC<br>GGCAGAGC<br>CTTG (SEQ<br>ID NO: 499) | TGCAGTC<br>CGGCAGA<br>GCCTTG<br>(SEQ ID<br>NO: 1016) | 没有 | 0.508030<br>4371 | 0.961889<br>9213 |  |
| 2103 | GATTTGGT<br>GTACCTGG<br>TTCC (SEQ<br>ID NO: 500) | GATTTGG<br>TGTACCT<br>GGTTCC<br>(SEQ ID<br>NO: 1017) | 没有 | 0.508030<br>4371 | 0                |  |
| 3122 | AGGAGGAG<br>GCCCAGCT<br>GGAG (SEQ<br>ID NO: 501) | AGGAGG<br>AGGCCCA<br>GCTGGAG<br>(SEQ ID<br>NO: 1018) | 没有 | 0.508030<br>4371 | 1.833232<br>678  |  |
| 1555 | GTCAATCA<br>ATACCTGG<br>AGTG (SEQ<br>ID NO: 502) | GTCAATC<br>AATACCT<br>GGAGTG<br>(SEQ ID<br>NO: 1019) | 没有 | 0.508030<br>4371 | 0                |  |
| 1725 | TAAGATCC<br>GAATGGAC<br>ATAG (SEQ<br>ID NO: 503) | TAAGATC<br>CGAATGG<br>ACATAG<br>(SEQ ID<br>NO: 1020) | 没有 | 0.508030<br>4371 | 0                |  |
| 3078 | GGCTGAGC<br>ACATGCTG<br>TTCT (SEQ<br>ID NO: 504) | GGCTGAG<br>CACATGC<br>TGTTCT<br>(SEQ ID<br>NO: 1021) | 没有 | 0.508030<br>4371 | 0                |  |
| 1681 | GACATGTAT<br>CCCTGGAC<br>CAG (SEQ<br>ID NO: 505) | GACATGT<br>ATCCCTG<br>GACCAG<br>(SEQ ID<br>NO: 1022) | 没有 | 0.508030<br>4371 | 1.833232<br>678  |  |

[0488]

|      |                                                  |                                                      |    |                  |                 |  |
|------|--------------------------------------------------|------------------------------------------------------|----|------------------|-----------------|--|
| 2023 | ATGACTGT<br>GAAGAGCA<br>TCGT (SEQ<br>ID NO: 506) | ATGACTG<br>TGAAGAG<br>CATCGT<br>(SEQ ID<br>NO: 1023) | 没有 | 0.508030<br>4371 | 0               |  |
| 3082 | GAGCACAT<br>GCTGTTCT<br>ATGC (SEQ<br>ID NO: 507) | GAGCACA<br>TGCTGTT<br>CTATGC<br>(SEQ ID<br>NO: 1024) | 没有 | 0.508030<br>4371 | 0               |  |
| 103  | CTCGTGTG<br>GCCTTTATC<br>TTT (SEQ ID<br>NO: 508) | CTCGTGT<br>GGCCTTT<br>ATCTTT<br>(SEQ ID<br>NO: 1025) | 没有 | 0.508030<br>4371 | 0               |  |
| 2935 | TTGTCCATC<br>CTCACAGG<br>CCT (SEQ ID<br>NO: 509) | TTGTCCA<br>TCCTCAC<br>TGGCCT<br>(SEQ ID<br>NO: 1026) | 没有 | 0.508030<br>4371 | 0               |  |
| 51   | CTGGAAGA<br>ACTGGACC<br>CTGC (SEQ<br>ID NO: 510) | CTGGAAG<br>AACTGGA<br>CCCTGC<br>(SEQ ID<br>NO: 1027) | 没有 | 0.508030<br>4371 | 0               |  |
| 175  | GAATGCCA<br>TTTCCCCA<br>ACAA (SEQ<br>ID NO: 511) | GAATGCC<br>ATTTCCC<br>CAACAA<br>(SEQ ID<br>NO: 1028) | 没有 | 0.508030<br>4371 | 0               |  |
| 2630 | GGTACTTT<br>CTTCTACA<br>AGAG (SEQ<br>ID NO: 512) | GGTACTT<br>TCTTCTA<br>CAAGAG<br>(SEQ ID<br>NO: 1029) | 没有 | 0.508030<br>4371 | 0               |  |
| 2308 | TGTAGTGG<br>TGTCATCTA<br>TTT (SEQ ID<br>NO: 513) | TGTAGTG<br>GTGTCAT<br>CTATTT<br>(SEQ ID<br>NO: 1030) | 没有 | 0                | 1.742685<br>513 |  |

[0489]

|      |                                                  |                                                       |    |   |                  |  |
|------|--------------------------------------------------|-------------------------------------------------------|----|---|------------------|--|
| 1272 | TGAAGAAC<br>TGGAACAC<br>GTTA (SEQ<br>ID NO: 514) | TGAAGAA<br>CTGGAAC<br>ACGTTC<br>(SEQ ID<br>NO: 1031)  | 没有 | 0 | 0.871342<br>7566 |  |
| 644  | TCCTGGAG<br>CGCTTCATC<br>ATC (SEQ ID<br>NO: 515) | TCCTGGA<br>GCGCTTC<br>ATCATC<br>(SEQ ID<br>NO: 1032)  | 没有 | 0 | 0.871342<br>7566 |  |
| 1162 | TTAACCAA<br>AATCGCTT<br>GGAG (SEQ<br>ID NO: 516) | TTAACCA<br>AAATCGC<br>TTGGAG<br>(SEQ ID<br>NO: 1033)  | 没有 | 0 | 0.871342<br>7566 |  |
| 2917 | GCTGGGAA<br>AACCACCA<br>CCTT (SEQ<br>ID NO: 517) | GCTGGGA<br>AAACCAC<br>CACCTT<br>(SEQ ID<br>NO: 1034)  | 没有 | 0 | 0.871342<br>7566 |  |
| 2889 | GATCACCG<br>CATTCTG<br>GGCC (SEQ<br>ID NO: 518)  | GATCACCC<br>GCATTCC<br>TGGGCC<br>(SEQ ID<br>NO: 1035) | 没有 | 0 | 0.871342<br>7566 |  |
| 3004 | ACCAGCCT<br>GGATGCAG<br>TCCG (SEQ<br>ID NO: 519) | ACCAGCC<br>TGGATGC<br>AGTCCG<br>(SEQ ID<br>NO: 1036)  | 没有 | 0 | 1.742685<br>513  |  |
| 2349 | CCTGTGCT<br>TCGCCTGG<br>CAGG (SEQ<br>ID NO: 520) | CCTGTGC<br>TTCGCCT<br>GGCAAG<br>(SEQ ID<br>NO: 1037)  | 没有 | 0 | 7.242383<br>547  |  |
| 3002 | AAACCAGC<br>CTGGATGC<br>AGTC (SEQ<br>ID NO: 521) | AAACCAG<br>CCTGGAT<br>GCAGTC<br>(SEQ ID<br>NO: 1038)  | 没有 | 0 | 0.871342<br>7566 |  |

[0490]

|      |                                                  |                                                      |     |   |                  |  |
|------|--------------------------------------------------|------------------------------------------------------|-----|---|------------------|--|
| 3076 | GTGGCTGA<br>GCACATGC<br>TGTT (SEQ<br>ID NO: 522) | GTGGCTG<br>AGCACAT<br>GCTGTT<br>(SEQ ID<br>NO: 1039) | 没有  | 0 | 2.704575<br>435  |  |
| 2813 | AGAATCTG<br>GTAAAGAT<br>TTTT (SEQ<br>ID NO: 523) | AGAATCT<br>GGTGAAG<br>ATTTTT<br>(SEQ ID<br>NO: 1040) | 分支点 | 0 | 0.871342<br>7566 |  |
| 2773 | GAGAGAGA<br>ACACCCCG<br>GGTG (SEQ<br>ID NO: 524) | GAGAGA<br>GAACACC<br>CCGGGTG<br>(SEQ ID<br>NO: 1041) | 没有  | 0 | 3.575918<br>191  |  |
| 619  | AAGGACAT<br>CGCCTGCA<br>GCGA (SEQ<br>ID NO: 525) | AAGGACA<br>TCGCCTG<br>CAGCGA<br>(SEQ ID<br>NO: 1042) | 没有  | 0 | 2.614028<br>27   |  |
| 2957 | TGCCTCCA<br>ACCTCTGG<br>GACT (SEQ<br>ID NO: 526) | TGCCTCC<br>AACCTCT<br>GGGACT<br>(SEQ ID<br>NO: 1043) | 没有  | 0 | 1.742685<br>513  |  |
| 2949 | AGGCCTGC<br>TGCCTCCA<br>ACCT (SEQ<br>ID NO: 527) | TGGCCTG<br>CTGCCTC<br>CAACCT<br>(SEQ ID<br>NO: 1044) | 没有  | 0 | 1.833232<br>678  |  |
| 2929 | ACCACCTT<br>GTCCATCCT<br>CAC (SEQ ID<br>NO: 528) | ACCACCT<br>TGTCCAT<br>CCTCAC<br>(SEQ ID<br>NO: 1045) | 没有  | 0 | 4.537808<br>112  |  |
| 2320 | ATCTATTTC<br>ACCCTCTA<br>CCT (SEQ ID<br>NO: 529) | ATCTATT<br>CACCTG<br>TACCT<br>(SEQ ID<br>NO: 1046)   | 没有  | 0 | 0.871342<br>7566 |  |

[0491]

|      |                                                  |                                                      |     |   |                  |  |
|------|--------------------------------------------------|------------------------------------------------------|-----|---|------------------|--|
| 2814 | GAATCTGG<br>TAAAGATT<br>TTTG (SEQ<br>ID NO: 530) | GAATCTG<br>GTGAAGA<br>TTTTTG<br>(SEQ ID<br>NO: 1047) | 分支点 | 0 | 0.871342<br>7566 |  |
| 2297 | TGGCAGCA<br>GCCTGTAG<br>TGGT (SEQ<br>ID NO: 531) | TGGCAGC<br>AGCCTGT<br>AGTGGT<br>(SEQ ID<br>NO: 1048) | 没有  | 0 | 0.871342<br>7566 |  |
| 3073 | ACGGTGGC<br>TGAGCACA<br>TGCT (SEQ<br>ID NO: 532) | ACGGTGG<br>CTGAGCA<br>CATGCT<br>(SEQ ID<br>NO: 1049) | 没有  | 0 | 2.795122<br>599  |  |
| 1785 | GGATTCTG<br>GTCCCAGA<br>GCTG (SEQ<br>ID NO: 533) | GGATTCT<br>GGTCCCA<br>GAGCTG<br>(SEQ ID<br>NO: 1050) | 没有  | 0 | 0.871342<br>7566 |  |
| 3094 | TTCTATGCC<br>CAGCTGAA<br>AGG (SEQ<br>ID NO: 534) | TTCTATG<br>CACAAC<br>GAAAGG<br>(SEQ ID<br>NO: 1051)  | 没有  | 0 | 0.871342<br>7566 |  |
| 173  | ATGAATGC<br>CATTCCCC<br>AAC (SEQ<br>ID NO: 535)  | ATGAATG<br>CCATTTC<br>CCCAAC<br>(SEQ ID<br>NO: 1052) | 没有  | 0 | 0.871342<br>7566 |  |
| 2194 | CATTACAG<br>CGACCCAT<br>TCAT (SEQ<br>ID NO: 536) | CATTACA<br>GCGACCC<br>ATTCAT<br>(SEQ ID<br>NO: 1053) | 没有  | 0 | 1.742685<br>513  |  |
| 2795 | TTCCTGGG<br>GTATGCGT<br>GAAG (SEQ<br>ID NO: 537) | TTCCTGG<br>GGTGTGC<br>GTGAAG<br>(SEQ ID<br>NO: 1054) | 没有  | 0 | 2.614028<br>27   |  |

[0492]

|      |                                                  |                                                      |    |   |                  |  |
|------|--------------------------------------------------|------------------------------------------------------|----|---|------------------|--|
| 13   | GGTGGCAG<br>CGGCTTCG<br>TGAG (SEQ<br>ID NO: 538) | GGTGAT<br>CCGGCTT<br>CGTGAG<br>(SEQ ID<br>NO: 1055)  | 没有 | 0 | 0.871342<br>7566 |  |
| 2242 | ACTGCCAC<br>CATCATGCT<br>GTG (SEQ ID<br>NO: 539) | ACTGCCA<br>CCATCAT<br>GCTGTG<br>(SEQ ID<br>NO: 1056) | 没有 | 0 | 0.871342<br>7566 |  |
| 2863 | CTGAACAT<br>CACCTTCT<br>ACGA (SEQ<br>ID NO: 540) | CTGAACA<br>TCACCTT<br>CTACGA<br>(SEQ ID<br>NO: 1057) | 没有 | 0 | 1.833232<br>678  |  |
| 2435 | CTGAGTAC<br>CTGGTTCG<br>CTTT (SEQ<br>ID NO: 541) | CTGAGTA<br>CCTGGTT<br>CGCTTT<br>(SEQ ID<br>NO: 1058) | 没有 | 0 | 0.871342<br>7566 |  |
| 1225 | GATTCACC<br>TGCAGCAC<br>GAAG (SEQ<br>ID NO: 542) | GATTCAC<br>CTGCTGC<br>ACGAAG<br>(SEQ ID<br>NO: 1059) | 没有 | 0 | 2.614028<br>27   |  |
| 901  | GACTTGCT<br>GTGGGTGA<br>CCAG (SEQ<br>ID NO: 543) | GACTTGC<br>TGTGGGT<br>GACCAG<br>(SEQ ID<br>NO: 1060) | 没有 | 0 | 0.871342<br>7566 |  |
| 1123 | TTTTGTAAT<br>GCATTGATC<br>CA (SEQ ID<br>NO: 544) | TTTTGTA<br>ATGCATT<br>GATCCA<br>(SEQ ID<br>NO: 1061) | 没有 | 0 | 0.871342<br>7566 |  |
| 2130 | CTTCTCCAT<br>CATGTCGAT<br>GA (SEQ ID<br>NO: 545) | CTTCTCC<br>ATCATGT<br>CGATGA<br>(SEQ ID<br>NO: 1062) | 没有 | 0 | 0.871342<br>7566 |  |

[0493]

|      |                                                  |                                                       |     |   |                  |  |
|------|--------------------------------------------------|-------------------------------------------------------|-----|---|------------------|--|
| 3011 | TGGATGCA<br>GTCCGGCA<br>GAGC (SEQ<br>ID NO: 546) | TGGATGC<br>AGTCCGG<br>CAGAGC<br>(SEQ ID<br>NO: 1063)  | 没有  | 0 | 0.871342<br>7566 |  |
| 2081 | ATCAGGGT<br>GTCTCCAA<br>TGCA (SEQ<br>ID NO: 547) | ATCAGGG<br>TGTCTCC<br>AATGCA<br>(SEQ ID<br>NO: 1064)  | 没有  | 0 | 0.871342<br>7566 |  |
| 1092 | TATCTATTC<br>TTATGACA<br>GAA (SEQ<br>ID NO: 548) | TATCTATT<br>CTTATGA<br>CAGAA<br>(SEQ ID<br>NO: 1065)  | 没有  | 0 | 1.742685<br>513  |  |
| 3120 | CCAGGAGG<br>AGGCCAG<br>CTGG (SEQ<br>ID NO: 549)  | CCAGGAG<br>GAGGCC<br>AGCTGG<br>(SEQ ID<br>NO: 1066)   | 没有  | 0 | 1.833232<br>678  |  |
| 3111 | AGGAAAGT<br>CCCAGGAG<br>GAGG (SEQ<br>ID NO: 550) | AGGAAA<br>GTCCCAG<br>GAGGAG<br>G (SEQ ID<br>NO: 1067) | 没有  | 0 | 1.742685<br>513  |  |
| 2839 | TGTGGCCG<br>GCCAGCTG<br>TGGA (SEQ<br>ID NO: 551) | TGTGGCC<br>GGCCAGC<br>TGTGGA<br>(SEQ ID<br>NO: 1068)  | 没有  | 0 | 0.871342<br>7566 |  |
| 2677 | TCAACCAG<br>AGAAGAAA<br>GAGC (SEQ<br>ID NO: 552) | TCAACCA<br>GAGAAG<br>AAAGAGC<br>(SEQ ID<br>NO: 1069)  | 没有  | 0 | 0.871342<br>7566 |  |
| 2821 | GTAAAGAT<br>TTTTGAGC<br>CCTG (SEQ<br>ID NO: 553) | GTGAAGA<br>TTTTTGA<br>GCCCTG<br>(SEQ ID<br>NO: 1070)  | 分支点 | 0 | 0.871342<br>7566 |  |

[0494]

|      |                                                  |                                                      |    |   |                  |  |
|------|--------------------------------------------------|------------------------------------------------------|----|---|------------------|--|
| 2784 | CCCCGGGT<br>GGGTTCCT<br>GGGG (SEQ<br>ID NO: 554) | CCCCGGG<br>TGGGTTC<br>CTGGGG<br>(SEQ ID<br>NO: 1071) | 没有 | 0 | 0.871342<br>7566 |  |
| 1129 | AATGCATT<br>GATCCAGA<br>GCCT (SEQ<br>ID NO: 555) | AATGCAT<br>TGATCCA<br>GAGCCT<br>(SEQ ID<br>NO: 1072) | 没有 | 0 | 0.871342<br>7566 |  |
| 918  | CAGGCCCC<br>TCATGCAG<br>AATG (SEQ<br>ID NO: 556) | CAGGCCC<br>CTCATGC<br>AGAATG<br>(SEQ ID<br>NO: 1073) | 没有 | 0 | 0.871342<br>7566 |  |
| 351  | ACTCCTCAT<br>GAATGCAC<br>CAG (SEQ<br>ID NO: 557) | ACTCCTC<br>ATGAATG<br>CACCAG<br>(SEQ ID<br>NO: 1074) | 没有 | 0 | 0.871342<br>7566 |  |
| 2540 | AGATGATG<br>CTCCTTGAT<br>GCT (SEQ ID<br>NO: 558) | AGATGAT<br>GCTCCTT<br>GATGCT<br>(SEQ ID<br>NO: 1075) | 没有 | 0 | 0.871342<br>7566 |  |
| 3056 | TCCTGTTCC<br>ACCACCTC<br>ACG (SEQ<br>ID NO: 559) | TCCTGTT<br>CCACCAC<br>CTCACG<br>(SEQ ID<br>NO: 1076) | 没有 | 0 | 1.742685<br>513  |  |
| 1465 | TACAAGGG<br>CCCTCGGG<br>AAAG (SEQ<br>ID NO: 560) | TACAAGG<br>GCCCTCG<br>GGAAAG<br>(SEQ ID<br>NO: 1077) | 没有 | 0 | 0.871342<br>7566 |  |
| 2838 | CTGTGGCC<br>GGCCAGCT<br>GTGG (SEQ<br>ID NO: 561) | CTGTGGC<br>CGGCCAG<br>CTGTGG<br>(SEQ ID<br>NO: 1078) | 没有 | 0 | 0.871342<br>7566 |  |

[0495]

|      |                                                  |                                                      |    |   |                  |  |
|------|--------------------------------------------------|------------------------------------------------------|----|---|------------------|--|
| 3065 | ACCACCTC<br>ACGGTGGC<br>TGAG (SEQ<br>ID NO: 562) | ACCACCT<br>CACGGTG<br>GCTGAG<br>(SEQ ID<br>NO: 1079) | 没有 | 0 | 0.871342<br>7566 |  |
| 1150 | GAGTCAAA<br>TCCTTTAAC<br>CAA (SEQ ID<br>NO: 563) | GAGTCAA<br>ATCCTTT<br>AACCAA<br>(SEQ ID<br>NO: 1080) | 没有 | 0 | 0.871342<br>7566 |  |
| 1959 | CATGATCAT<br>CCTGAACC<br>GCT (SEQ ID<br>NO: 564) | CATGATC<br>ATCCTGA<br>ACCGCT<br>(SEQ ID<br>NO: 1081) | 没有 | 0 | 0.871342<br>7566 |  |
| 2656 | TGGCTTGG<br>CGGTGAAG<br>GGTG (SEQ<br>ID NO: 565) | TGGCTTG<br>GCGGTGA<br>AGGGTG<br>(SEQ ID<br>NO: 1082) | 没有 | 0 | 0.871342<br>7566 |  |
| 101  | AACTCGTG<br>TGGCCTTT<br>ATCT (SEQ<br>ID NO: 566) | AACTCGT<br>GTGGCCT<br>TTATCT<br>(SEQ ID<br>NO: 1083) | 没有 | 0 | 0.871342<br>7566 |  |
| 1807 | CCCGTGGA<br>AGATTTC<br>GGTA (SEQ<br>ID NO: 567)  | CCCGTGG<br>AAGATTT<br>CCGGTA<br>(SEQ ID<br>NO: 1084) | 没有 | 0 | 0.871342<br>7566 |  |
| 1270 | TTTGAAGA<br>ACTGGAAC<br>ACGT (SEQ<br>ID NO: 568) | TTTGAAG<br>AACTGGA<br>ACACGT<br>(SEQ ID<br>NO: 1085) | 没有 | 0 | 0.871342<br>7566 |  |
| 2738 | CAGAGCAC<br>CCAGAAGG<br>AATA (SEQ<br>ID NO: 569) | CAGAGCA<br>CCCAGAA<br>GGAATA<br>(SEQ ID<br>NO: 1086) | 没有 | 0 | 0.871342<br>7566 |  |

[0496]

|      |                                                   |                                                       |    |   |                  |  |
|------|---------------------------------------------------|-------------------------------------------------------|----|---|------------------|--|
| 2827 | ATTTTGTGAG<br>CCCTGTGG<br>CCG (SEQ ID<br>NO: 570) | ATTTTGTG<br>AGCCCTG<br>TGGCCG<br>(SEQ ID<br>NO: 1087) | 没有 | 0 | 0.871342<br>7566 |  |
| 2074 | TTGAAAAA<br>TCAGGGTG<br>TCTC (SEQ<br>ID NO: 571)  | TTGAAAA<br>ATCAGGG<br>TGTCTC<br>(SEQ ID<br>NO: 1088)  | 没有 | 0 | 0.871342<br>7566 |  |
| 1414 | AGGCAGCT<br>TGGTGAAG<br>AAGG (SEQ<br>ID NO: 572)  | AGGCAGC<br>TTGGTGA<br>AGAAGG<br>(SEQ ID<br>NO: 1089)  | 没有 | 0 | 0.871342<br>7566 |  |
| 2383 | GAGCTGAA<br>GAAGGCTG<br>TGAG (SEQ<br>ID NO: 573)  | GAGCTGA<br>AGAAGGC<br>TGTGAG<br>(SEQ ID<br>NO: 1090)  | 没有 | 0 | 0.871342<br>7566 |  |
| 2961 | TCCAACCT<br>CTGGGACT<br>GTGC (SEQ<br>ID NO: 574)  | TCCAACC<br>TCTGGGA<br>CTGTGC<br>(SEQ ID<br>NO: 1091)  | 没有 | 0 | 0.871342<br>7566 |  |
| 1355 | CACAGATG<br>AACATGAT<br>CAGA (SEQ<br>ID NO: 575)  | CACAGAT<br>GAACATG<br>ATCAGA<br>(SEQ ID<br>NO: 1092)  | 没有 | 0 | 0.871342<br>7566 |  |
| 987  | GTGTGGCT<br>ACCCCGAG<br>GGAG (SEQ<br>ID NO: 576)  | GTGTGGC<br>TACCCCG<br>AGGGAG<br>(SEQ ID<br>NO: 1093)  | 没有 | 0 | 0.871342<br>7566 |  |
| 2195 | ATTACAGC<br>GACCCATT<br>CATC (SEQ<br>ID NO: 577)  | ATTACAG<br>CGACCCA<br>TTCATC<br>(SEQ ID<br>NO: 1094)  | 没有 | 0 | 0.871342<br>7566 |  |

[0497]

|      |                                                  |                                                       |    |   |                  |  |
|------|--------------------------------------------------|-------------------------------------------------------|----|---|------------------|--|
| 2072 | CTTTGAAA<br>AATCAGGG<br>TGTC (SEQ<br>ID NO: 578) | CATTGAA<br>AAATCAG<br>GGTGTC<br>(SEQ ID<br>NO: 1095)  | 没有 | 0 | 0.871342<br>7566 |  |
| 2899 | TTCCTGGG<br>CCACAATG<br>GAGC (SEQ<br>ID NO: 579) | TTCCTGG<br>GCCACAA<br>TGGAGC<br>(SEQ ID<br>NO: 1096)  | 没有 | 0 | 0.871342<br>7566 |  |
| 2674 | TGTTCAAC<br>CAGAGAAG<br>AAAG (SEQ<br>ID NO: 580) | TGTTCAA<br>CCAGAGA<br>AGAAAG<br>(SEQ ID<br>NO: 1097)  | 没有 | 0 | 0.871342<br>7566 |  |
| 1987 | ATCTTCATG<br>GTGCTGGC<br>ATG (SEQ ID<br>NO: 581) | ATCTTCAT<br>GGTGCTG<br>GCATG<br>(SEQ ID<br>NO: 1098)  | 没有 | 0 | 0.871342<br>7566 |  |
| 498  | AGAAACAC<br>TGACACTA<br>TTTC (SEQ<br>ID NO: 582) | AGAAACA<br>CTGACAC<br>TATTC<br>(SEQ ID<br>NO: 1099)   | 没有 | 0 | 0.871342<br>7566 |  |
| 1955 | CTTTCATGA<br>TCATCCTG<br>AAC (SEQ<br>ID NO: 583) | CTTTCAT<br>GATCATC<br>CTGAAC<br>(SEQ ID<br>NO: 1100)  | 没有 | 0 | 0.871342<br>7566 |  |
| 2777 | GAGAACAC<br>CCCGGGTG<br>GGTT (SEQ<br>ID NO: 584) | GAGAACAC<br>CCCCGGG<br>TGGGTT<br>(SEQ ID<br>NO: 1101) | 没有 | 0 | 0.871342<br>7566 |  |
| 96   | GGTGGAAC<br>TCGTGTGG<br>CCTT (SEQ<br>ID NO: 585) | GGTGGAAC<br>CTCGTGT<br>GGCCTT<br>(SEQ ID<br>NO: 1102) | 没有 | 0 | 0.871342<br>7566 |  |

[0498]

|      |                                                  |                                                      |    |   |                  |  |
|------|--------------------------------------------------|------------------------------------------------------|----|---|------------------|--|
| 1146 | CCTGGAGT<br>CAAATCCT<br>TTAA (SEQ<br>ID NO: 586) | CCTGGAG<br>TCAAATC<br>CTTTAA<br>(SEQ ID<br>NO: 1103) | 没有 | 0 | 0.871342<br>7566 |  |
| 121  | TTATTTCTG<br>GTCTTGATC<br>TG (SEQ ID<br>NO: 587) | TTATTTCT<br>GGTCTTG<br>ATCTG<br>(SEQ ID<br>NO: 1104) | 没有 | 0 | 0.871342<br>7566 |  |
| 2495 | ACAGTCCC<br>ACGGAAGG<br>GGAC (SEQ<br>ID NO: 588) | ACAGTCC<br>CACGGAA<br>GGGGAC<br>(SEQ ID<br>NO: 1105) | 没有 | 0 | 0.871342<br>7566 |  |
| 2203 | GACCCATT<br>CATCCTCTT<br>CCT (SEQ ID<br>NO: 589) | GACCCAT<br>TCATCCT<br>CTTCCT<br>(SEQ ID<br>NO: 1106) | 没有 | 0 | 0.871342<br>7566 |  |
| 3170 | TCCACCAC<br>AAGCGGAA<br>TGAA (SEQ<br>ID NO: 590) | TCCACCA<br>CAAGCGG<br>AATGAA<br>(SEQ ID<br>NO: 1107) | 没有 | 0 | 0.961889<br>9213 |  |
| 3169 | CTCCACCA<br>CAAGCGGA<br>ATGA (SEQ<br>ID NO: 591) | CTCCACC<br>ACAAGCG<br>GAATGA<br>(SEQ ID<br>NO: 1108) | 没有 | 0 | 2.885669<br>764  |  |
| 2086 | GGTGTCTC<br>CAATGCAG<br>TGAT (SEQ<br>ID NO: 592) | GGTGTCT<br>CCAATGC<br>AGTGAT<br>(SEQ ID<br>NO: 1109) | 没有 | 0 | 0.961889<br>9213 |  |
| 240  | CTGCAATG<br>TGAACAAT<br>CCCT (SEQ<br>ID NO: 593) | CTGCAAT<br>GTGAACA<br>ATCCCT<br>(SEQ ID<br>NO: 1110) | 没有 | 0 | 1.923779<br>843  |  |

[0499]

|      |                                                  |                                                       |     |   |                  |  |
|------|--------------------------------------------------|-------------------------------------------------------|-----|---|------------------|--|
| 2653 | TATTGGCTT<br>GGCGGTGA<br>AGG (SEQ<br>ID NO: 594) | TATTGGC<br>TTGGCGG<br>TGAAGG<br>(SEQ ID<br>NO: 1111)  | 没有  | 0 | 0.961889<br>9213 |  |
| 2268 | GCTCAGCA<br>CCTTCTTCT<br>CCA (SEQ ID<br>NO: 595) | GCTCAGC<br>ACCTTCT<br>TCTCCA<br>(SEQ ID<br>NO: 1112)  | 没有  | 0 | 0.961889<br>9213 |  |
| 353  | TCCTCATG<br>AATGCACC<br>AGAG (SEQ<br>ID NO: 596) | TCCTCAT<br>GAATGCA<br>CCAGAG<br>(SEQ ID<br>NO: 1113)  | 没有  | 0 | 1.923779<br>843  |  |
| 2073 | TTTGAAAA<br>ATCAGGGT<br>GTCT (SEQ<br>ID NO: 597) | ATTGAAA<br>AATCAGG<br>GTGTCT<br>(SEQ ID<br>NO: 1114)  | 没有  | 0 | 0.961889<br>9213 |  |
| 2271 | CAGCACCT<br>TCTTCTCCA<br>AGG (SEQ<br>ID NO: 598) | CAGCACCT<br>TTCTTCT<br>CCAAGG<br>(SEQ ID<br>NO: 1115) | 没有  | 0 | 1.923779<br>843  |  |
| 2970 | TGGGACTG<br>TGCTCGTT<br>GGGG (SEQ<br>ID NO: 599) | TGGGACT<br>GTGCTCG<br>TTGGGG<br>(SEQ ID<br>NO: 1116)  | 没有  | 0 | 0.961889<br>9213 |  |
| 2815 | AATCTGGT<br>AAAGATTT<br>TTGA (SEQ<br>ID NO: 600) | AATCTGG<br>TGAAGAT<br>TTTTGA<br>(SEQ ID<br>NO: 1117)  | 分支点 | 0 | 0.961889<br>9213 |  |
| 3132 | CCAGCTGG<br>AGATGGAA<br>GCCA (SEQ<br>ID NO: 601) | CCAGCTG<br>GAGATGG<br>AAGCCA<br>(SEQ ID<br>NO: 1118)  | 没有  | 0 | 0.961889<br>9213 |  |

[0500]

|      |                                                  |                                                      |    |   |                  |  |
|------|--------------------------------------------------|------------------------------------------------------|----|---|------------------|--|
| 2097 | TGCAGTGA<br>TTTGGTGT<br>ACCT (SEQ<br>ID NO: 602) | TGCAGTG<br>ATTTGGT<br>GTACCT<br>(SEQ ID<br>NO: 1119) | 没有 | 0 | 0.961889<br>9213 |  |
| 1894 | GCGGAGGC<br>TCCAGTTG<br>GAAT (SEQ<br>ID NO: 603) | GCGGAGG<br>CTCCAGT<br>TGGAAT<br>(SEQ ID<br>NO: 1120) | 没有 | 0 | 0.961889<br>9213 |  |
| 617  | TGAAGGAC<br>ATCGCCTG<br>CAGC (SEQ<br>ID NO: 604) | TGAAGGA<br>CATCGCC<br>TGCAGC<br>(SEQ ID<br>NO: 1121) | 没有 | 0 | 2.885669<br>764  |  |
| 3027 | GAGCCTTG<br>GCATGTGT<br>CCAC (SEQ<br>ID NO: 605) | GAGCCTT<br>GGCATGT<br>GTCCAC<br>(SEQ ID<br>NO: 1122) | 没有 | 0 | 1.923779<br>843  |  |
| 1977 | CTGTTTCCC<br>TATCTTCAT<br>GG (SEQ ID<br>NO: 606) | CTGTTTC<br>CCTATCTT<br>CATGG<br>(SEQ ID<br>NO: 1123) | 没有 | 0 | 1.923779<br>843  |  |
| 2835 | GCCCTGTG<br>GCCGGCCA<br>GCTG (SEQ<br>ID NO: 607) | GCCCTGT<br>GGCCGGC<br>CAGCTG<br>(SEQ ID<br>NO: 1124) | 没有 | 0 | 0.961889<br>9213 |  |
| 205  | TCAGCAGG<br>AATGCTGC<br>CGTG (SEQ<br>ID NO: 608) | TCAGCAG<br>GAATGCT<br>GCCGTG<br>(SEQ ID<br>NO: 1125) | 没有 | 0 | 0.961889<br>9213 |  |
| 3161 | ACACAGGC<br>CTCCACCA<br>CAAG (SEQ<br>ID NO: 609) | ACACAGG<br>CCTCCAC<br>CACAAG<br>(SEQ ID<br>NO: 1126) | 没有 | 0 | 1.923779<br>843  |  |

[0501]

|      |                                                  |                                                      |    |   |                  |  |
|------|--------------------------------------------------|------------------------------------------------------|----|---|------------------|--|
| 2864 | TGAACATC<br>ACCTTCTA<br>CGAG (SEQ<br>ID NO: 610) | TGAACAT<br>CACCTTC<br>TACGAG<br>(SEQ ID<br>NO: 1127) | 没有 | 0 | 1.923779<br>843  |  |
| 3055 | ATCCTGTTC<br>CACCACCT<br>CAC (SEQ ID<br>NO: 611) | ATCCTGT<br>TCCACCA<br>CCTCAC<br>(SEQ ID<br>NO: 1128) | 没有 | 0 | 0.961889<br>9213 |  |
| 1931 | CCTACCCC<br>TGCTTCGT<br>GGAC (SEQ<br>ID NO: 612) | CCTACCC<br>CTGCTTC<br>GTGGAC<br>(SEQ ID<br>NO: 1129) | 没有 | 0 | 0.961889<br>9213 |  |
| 2872 | ACCTTCTA<br>CGAGAACC<br>AGAT (SEQ<br>ID NO: 613) | ACCTTCT<br>ACGAGAA<br>CCAGAT<br>(SEQ ID<br>NO: 1130) | 没有 | 0 | 1.923779<br>843  |  |
| 2950 | GGCCTGCT<br>GCCTCCAA<br>CCTC (SEQ<br>ID NO: 614) | GGCCTGC<br>TGCCTCC<br>AACCTC<br>(SEQ ID<br>NO: 1131) | 没有 | 0 | 0.961889<br>9213 |  |
| 92   | TTGTGGTG<br>GAACTCGT<br>GTGG (SEQ<br>ID NO: 615) | TTGTGGT<br>GGAATC<br>GTGTGG<br>(SEQ ID<br>NO: 1132)  | 没有 | 0 | 1.923779<br>843  |  |
| 3050 | ACAACATC<br>CTGTTCCA<br>CCAC (SEQ<br>ID NO: 616) | ACAACAT<br>CCTGTTC<br>CACCAC<br>(SEQ ID<br>NO: 1133) | 没有 | 0 | 0.961889<br>9213 |  |
| 1874 | TCACAAGG<br>AGCCAGGT<br>GCAG (SEQ<br>ID NO: 617) | TCACAAG<br>GAGCCAG<br>GTGCAG<br>(SEQ ID<br>NO: 1134) | 没有 | 0 | 0.961889<br>9213 |  |

[0502]

|      |                                                  |                                                      |    |   |                  |  |
|------|--------------------------------------------------|------------------------------------------------------|----|---|------------------|--|
| 2984 | TTGGGGGA<br>AGGGACAT<br>TGAA (SEQ<br>ID NO: 618) | TTGGGGG<br>AAGGGAC<br>ATTGAA<br>(SEQ ID<br>NO: 1135) | 没有 | 0 | 1.923779<br>843  |  |
| 2107 | TGGTGTAC<br>CTGGTTCC<br>TGGA (SEQ<br>ID NO: 619) | TGGTGT<br>CCTGGTT<br>CCTGGA<br>(SEQ ID<br>NO: 1136)  | 没有 | 0 | 0.961889<br>9213 |  |
| 33   | ACAGATAC<br>AGCTTTTG<br>CTCT (SEQ<br>ID NO: 620) | ACAGATA<br>CAGCTTT<br>TGCTCT<br>(SEQ ID<br>NO: 1137) | 没有 | 0 | 1.923779<br>843  |  |
| 2999 | TTGAAACC<br>AGCCTGGA<br>TGCA (SEQ<br>ID NO: 621) | TTGAAAC<br>CAGCCTG<br>GATGCA<br>(SEQ ID<br>NO: 1138) | 没有 | 0 | 0.961889<br>9213 |  |
| 2836 | CCCTGTGG<br>CCGGCCAG<br>CTGT (SEQ<br>ID NO: 622) | CCCTGTG<br>GCCGGCC<br>AGCTGT<br>(SEQ ID<br>NO: 1139) | 没有 | 0 | 0.961889<br>9213 |  |
| 2262 | CTTTCTGCT<br>CAGCACCT<br>TCT (SEQ ID<br>NO: 623) | CTTTCTG<br>CTCAGCA<br>CCTTCT<br>(SEQ ID<br>NO: 1140) | 没有 | 0 | 0.961889<br>9213 |  |
| 2640 | TCTACAAG<br>AGTCGTATT<br>GGC (SEQ<br>ID NO: 624) | TCTACAA<br>GAGTCGT<br>ATTGGC<br>(SEQ ID<br>NO: 1141) | 没有 | 0 | 1.923779<br>843  |  |
| 2717 | TAACAGAG<br>GAAACGGA<br>GGAT (SEQ<br>ID NO: 625) | TAACAGA<br>GGAAACG<br>GAGGAT<br>(SEQ ID<br>NO: 1142) | 没有 | 0 | 0.961889<br>9213 |  |

[0503]

|      |                                                  |                                                      |    |   |                  |  |
|------|--------------------------------------------------|------------------------------------------------------|----|---|------------------|--|
| 3038 | TGTGTCCA<br>CAGCACAA<br>CATC (SEQ<br>ID NO: 626) | TGTGTCC<br>ACAGCAC<br>AACATC<br>(SEQ ID<br>NO: 1143) | 没有 | 0 | 0.961889<br>9213 |  |
| 1579 | GTCCTGGA<br>TAAGTTTG<br>AAAG (SEQ<br>ID NO: 627) | GTCCTGG<br>ATAAGTT<br>TGAAAG<br>(SEQ ID<br>NO: 1144) | 没有 | 0 | 0.961889<br>9213 |  |
| 2307 | CTGTAGTG<br>GTGTCATCT<br>ATT (SEQ ID<br>NO: 628) | CTGTAGT<br>GGTGTCA<br>TCTATT<br>(SEQ ID<br>NO: 1145) | 没有 | 0 | 0.961889<br>9213 |  |
| 3138 | GGAGATGG<br>AAGCCATG<br>TTGG (SEQ<br>ID NO: 629) | GGAGATG<br>GAAGCCA<br>TGTTGG<br>(SEQ ID<br>NO: 1146) | 没有 | 0 | 0.961889<br>9213 |  |
| 2892 | CACCGCAT<br>TCCTGGGC<br>CACA (SEQ<br>ID NO: 630) | CACCGCA<br>TTCCTGG<br>GCCACA<br>(SEQ ID<br>NO: 1147) | 没有 | 0 | 0.961889<br>9213 |  |
| 1558 | AATCAATA<br>CCTGGAGT<br>GCTT (SEQ<br>ID NO: 631) | AATCAAT<br>ACCTGGA<br>GTGCTT<br>(SEQ ID<br>NO: 1148) | 没有 | 0 | 1.923779<br>843  |  |
| 2955 | GCTGCCTC<br>CAACCTCT<br>GGGA (SEQ<br>ID NO: 632) | GCTGCCT<br>CCAACCT<br>CTGGGA<br>(SEQ ID<br>NO: 1149) | 没有 | 0 | 0.961889<br>9213 |  |
| 3146 | AAGCCATG<br>TTGGAGGA<br>CACA (SEQ<br>ID NO: 633) | AAGCCAT<br>GTTGGAG<br>GACACA<br>(SEQ ID<br>NO: 1150) | 没有 | 0 | 1.923779<br>843  |  |

[0504]

|      |                                                  |                                                       |    |   |                  |  |
|------|--------------------------------------------------|-------------------------------------------------------|----|---|------------------|--|
| 2799 | TGGGGTAT<br>GCGTGAAG<br>AATC (SEQ<br>ID NO: 634) | TGGGGTG<br>TGCCTGA<br>AGAATC<br>(SEQ ID<br>NO: 1151)  | 没有 | 0 | 0.961889<br>9213 |  |
| 91   | TTTGTGGT<br>GGAACCTG<br>TGTG (SEQ<br>ID NO: 635) | TTTGTGG<br>TGGAACCT<br>CGTGTG<br>(SEQ ID<br>NO: 1152) | 没有 | 0 | 0.961889<br>9213 |  |
| 1495 | GACATGGC<br>CAACTTCG<br>ACTG (SEQ<br>ID NO: 636) | GACATGG<br>CCAACCTT<br>CGACTG<br>(SEQ ID<br>NO: 1153) | 没有 | 0 | 0.961889<br>9213 |  |
| 1969 | CTGAACCG<br>CTGTTTCCC<br>TAT (SEQ ID<br>NO: 637) | CTGAACC<br>GCTGTTT<br>CCCTAT<br>(SEQ ID<br>NO: 1154)  | 没有 | 0 | 0.961889<br>9213 |  |
| 2340 | GCCACACA<br>TCCTGTGC<br>TTCG (SEQ<br>ID NO: 638) | GCCCCAC<br>ATCCTGT<br>GCTTCG<br>(SEQ ID<br>NO: 1155)  | 没有 | 0 | 1.923779<br>843  |  |
| 578  | GTCCAGAG<br>CAGTTCGC<br>TCAT (SEQ<br>ID NO: 639) | GTCCAGA<br>GCAGTTC<br>GCTCAT<br>(SEQ ID<br>NO: 1156)  | 没有 | 0 | 0.961889<br>9213 |  |
| 2491 | GGGAACAG<br>TCCCACGG<br>AAGG (SEQ<br>ID NO: 640) | GGGAACA<br>GTCCCAC<br>GGAAGG<br>(SEQ ID<br>NO: 1157)  | 没有 | 0 | 0.961889<br>9213 |  |
| 1826 | ACATCTGG<br>GGCGGGTT<br>TGCC (SEQ<br>ID NO: 641) | ACATCTG<br>GGGCGGG<br>TTTGCA<br>(SEQ ID<br>NO: 1158)  | 没有 | 0 | 0.961889<br>9213 |  |

[0505]

|      |                                                  |                                                      |    |   |                  |  |
|------|--------------------------------------------------|------------------------------------------------------|----|---|------------------|--|
| 563  | TCAACTCT<br>CAAGTCCG<br>TCCA (SEQ<br>ID NO: 642) | TCAACTC<br>TCAAGTC<br>CGTCCA<br>(SEQ ID<br>NO: 1159) | 没有 | 0 | 0.961889<br>9213 |  |
| 1535 | CTGATCGC<br>ACCCTCCG<br>CCTG (SEQ<br>ID NO: 643) | CTGATCG<br>CACCTC<br>CGCCTG<br>(SEQ ID<br>NO: 1160)  | 没有 | 0 | 0.961889<br>9213 |  |
| 1444 | GAAGCCAT<br>CCTAAACT<br>TCCT (SEQ<br>ID NO: 644) | GAGGCCA<br>TCCTAAA<br>CTTCCT<br>(SEQ ID<br>NO: 1161) | 没有 | 0 | 0.961889<br>9213 |  |
| 3223 | AAACTGAG<br>CGTGGCCA<br>TTGC (SEQ<br>ID NO: 645) | AAACTGA<br>GCGTGGC<br>CATTGC<br>(SEQ ID<br>NO: 1162) | 没有 | 0 | 0.961889<br>9213 |  |
| 1500 | GGCCAACT<br>TCGACTGG<br>AGGG (SEQ<br>ID NO: 646) | GGCCAAC<br>TTCGACT<br>GGAGGG<br>(SEQ ID<br>NO: 1163) | 没有 | 0 | 1.923779<br>843  |  |
| 3200 | ACCTATCA<br>GGCGGCAT<br>GCAG (SEQ<br>ID NO: 647) | ACCTATC<br>AGGCGGC<br>ATGCAG<br>(SEQ ID<br>NO: 1164) | 没有 | 0 | 0.961889<br>9213 |  |
| 787  | CTTCCCAC<br>ACTCCTAG<br>ACAG (SEQ<br>ID NO: 648) | CTTCCCA<br>CACTCCT<br>AGACAG<br>(SEQ ID<br>NO: 1165) | 没有 | 0 | 0.961889<br>9213 |  |
| 2213 | TCCTCTTC<br>TGTTCTTGT<br>TG (SEQ ID<br>NO: 649)  | TCCTCTT<br>CCTGTTC<br>TTGTTG<br>(SEQ ID<br>NO: 1166) | 没有 | 0 | 0.961889<br>9213 |  |

[0506]

|      |                                                  |                                                      |     |   |                  |  |
|------|--------------------------------------------------|------------------------------------------------------|-----|---|------------------|--|
| 2411 | CTCCGGTG<br>GCATTTGG<br>ATTT (SEQ<br>ID NO: 650) | CTCCGGT<br>GGCATTT<br>GGATTT<br>(SEQ ID<br>NO: 1167) | 没有  | 0 | 0.961889<br>9213 |  |
| 3035 | GCATGTGT<br>CCACAGCA<br>CAAC (SEQ<br>ID NO: 651) | GCATGTG<br>TCCACAG<br>CACAAC<br>(SEQ ID<br>NO: 1168) | 没有  | 0 | 0.961889<br>9213 |  |
| 2812 | AAGAATCT<br>GGTAAAGA<br>TTTT (SEQ<br>ID NO: 652) | AAGAATC<br>TGGTGAA<br>GATTTT<br>(SEQ ID<br>NO: 1169) | 分支点 | 0 | 0.961889<br>9213 |  |
| 626  | TCGCCTGC<br>AGCGAGGC<br>CCTC (SEQ<br>ID NO: 653) | TCGCCTG<br>CAGCGAG<br>GCCCTC<br>(SEQ ID<br>NO: 1170) | 没有  | 0 | 0.961889<br>9213 |  |
| 2005 | TGGATCTA<br>CTCTGTCTC<br>CAT (SEQ ID<br>NO: 654) | TGGATCT<br>ACTCTGT<br>CTCCAT<br>(SEQ ID<br>NO: 1171) | 没有  | 0 | 0.961889<br>9213 |  |
| 65   | CCCTGCGG<br>AAAAGGCA<br>AAAG (SEQ<br>ID NO: 655) | CCCTGCG<br>GAAAAG<br>GCAAAAG<br>(SEQ ID<br>NO: 1172) | 没有  | 0 | 0.961889<br>9213 |  |
| 2928 | CACCACCT<br>TGTCATCC<br>TCA (SEQ ID<br>NO: 656)  | CACCACC<br>TTGTCCA<br>TCCTCA<br>(SEQ ID<br>NO: 1173) | 没有  | 0 | 0.961889<br>9213 |  |
| 2758 | CACGACTC<br>CTTCTTCG<br>AGAG (SEQ<br>ID NO: 657) | CACGACT<br>CCTTCTT<br>CGAGAG<br>(SEQ ID<br>NO: 1174) | 没有  | 0 | 0.961889<br>9213 |  |

[0507]

|      |                                                  |                                                      |    |   |                  |  |
|------|--------------------------------------------------|------------------------------------------------------|----|---|------------------|--|
| 2295 | TCTGGCAG<br>CAGCCTGT<br>AGTG (SEQ<br>ID NO: 658) | TCTGGCA<br>GCAGCCT<br>GTAGTG<br>(SEQ ID<br>NO: 1175) | 没有 | 0 | 0.961889<br>9213 |  |
| 3139 | GAGATGGA<br>AGCCATGT<br>TGGA (SEQ<br>ID NO: 659) | GAGATGG<br>AAGCCAT<br>GTTGGA<br>(SEQ ID<br>NO: 1176) | 没有 | 0 | 1.923779<br>843  |  |
| 169  | CATCATGA<br>ATGCCATTT<br>CCC (SEQ ID<br>NO: 660) | CATCATG<br>AATGCCA<br>TTTCCC<br>(SEQ ID<br>NO: 1177) | 没有 | 0 | 0.961889<br>9213 |  |
| 2964 | AACCTCTG<br>GGACTGTG<br>CTCG (SEQ<br>ID NO: 661) | AACCTCT<br>GGGACTG<br>TGCTCG<br>(SEQ ID<br>NO: 1178) | 没有 | 0 | 0.961889<br>9213 |  |
| 1354 | ACACAGAT<br>GAACATGA<br>TCAG (SEQ<br>ID NO: 662) | ACACAGA<br>TGAACAT<br>GATCAG<br>(SEQ ID<br>NO: 1179) | 没有 | 0 | 0.961889<br>9213 |  |
| 1416 | GCAGCTTG<br>GTGAAGAA<br>GGTA (SEQ<br>ID NO: 663) | GCAGCTT<br>GGTGAAG<br>AAGGCA<br>(SEQ ID<br>NO: 1180) | 没有 | 0 | 0.961889<br>9213 |  |
| 2902 | CTGGGCCA<br>CAATGGAG<br>CTGG (SEQ<br>ID NO: 664) | CTGGGCC<br>ACAATGG<br>AGCTGG<br>(SEQ ID<br>NO: 1181) | 没有 | 0 | 0.961889<br>9213 |  |
| 954  | TACAAAGC<br>TGATGGGC<br>ATCC (SEQ<br>ID NO: 665) | TACAAAG<br>CTGATGG<br>GCATCC<br>(SEQ ID<br>NO: 1182) | 没有 | 0 | 0.961889<br>9213 |  |

[0508]

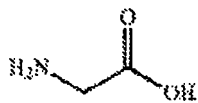
|      |                                                   |                                                      |    |   |                  |  |
|------|---------------------------------------------------|------------------------------------------------------|----|---|------------------|--|
| 2942 | TCCTCACA<br>GGCCTGCT<br>GCCT (SEQ<br>ID NO: 666)  | TCCTCAC<br>TGGCCTG<br>CTGCCT<br>(SEQ ID<br>NO: 1183) | 没有 | 0 | 0.961889<br>9213 |  |
| 94   | GTGGTGGA<br>ACTCGTGT<br>GGCC (SEQ<br>ID NO: 667)  | GTGGTGG<br>AACTCGT<br>GTGGCC<br>(SEQ ID<br>NO: 1184) | 没有 | 0 | 0.961889<br>9213 |  |
| 3064 | CACCACCT<br>CACGGTGG<br>CTGA (SEQ<br>ID NO: 668)  | CACCACC<br>TCACGGT<br>GGCTGA<br>(SEQ ID<br>NO: 1185) | 没有 | 0 | 0.961889<br>9213 |  |
| 2232 | GGCTTTCT<br>CCTACTGCC<br>ACCA (SEQ<br>ID NO: 669) | GGCTTTC<br>TCCACTG<br>CCACCA<br>(SEQ ID<br>NO: 1186) | 没有 | 0 | 0.961889<br>9213 |  |
| 2660 | TTGGCGGT<br>GAAGGGTG<br>TTCA (SEQ<br>ID NO: 670)  | TTGGCGG<br>TGAAGGG<br>TGTTC<br>(SEQ ID<br>NO: 1187)  | 没有 | 0 | 0.961889<br>9213 |  |

[0509] “典型氨基酸”是指下表6所示的人体中天然存在的20种氨基酸。

[0510] 表6:描绘了典型氨基酸和其三字母缩写、一字母缩写、结构和相应的密码子每一者:

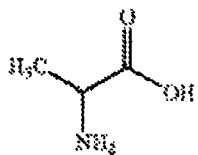
## 非极性脂族残基

甘氨酸 Gly G



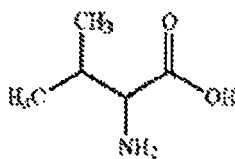
GGU GGC GGA GGG

丙氨酸 Ala A



GCU GCC GCA GCG

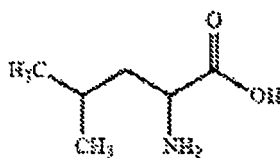
缬氨酸 Val V



GUU GUC GUA GUG

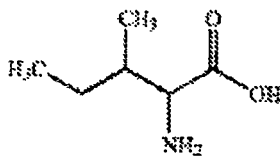
[0511]

亮氨酸 Leu L



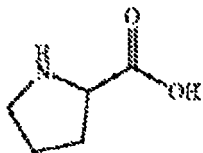
UUA UUG CUU CUC CUA CUG

异亮氨酸 Ile I



AUU AUC AUA

脯氨酸 Pro P

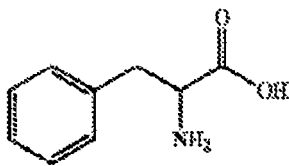


CCU CCC CCA CCG

## 芳香族残基

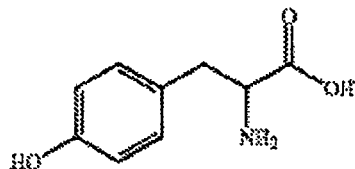
苯丙氨酸 Phe F

UUG/UUC



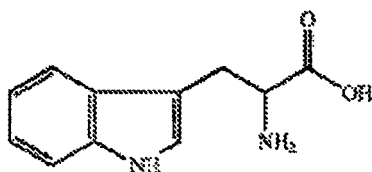
酪氨酸 Tyr Y

UAU/UAC



色氨酸 Trp W

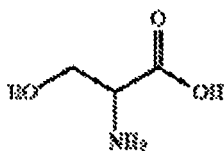
UGG



## 极性不带电残基

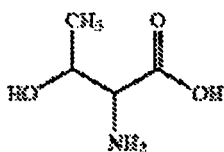
丝氨酸 Ser S

UCU/UCC/UCA/UCG/AGU/AGC



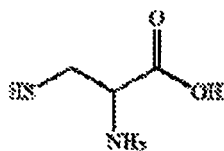
苏氨酸 Thr T

ACU/ACC/ACA/ACG



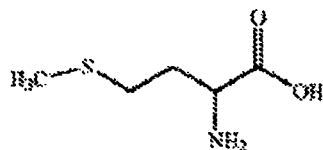
半胱氨酸 Cys C

UGU/UGC



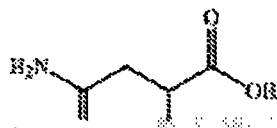
甲硫氨酸 Met M

AUG



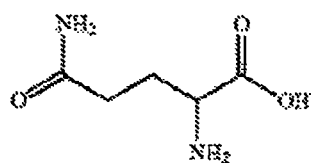
天冬酰胺 Asn N

AAU/AAC



[0512]

谷氨酰胺 Gln Q



CAA CAG

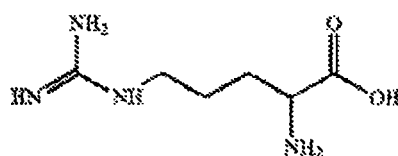
## 带正电残基

赖氨酸 Lys K



AAA AAG

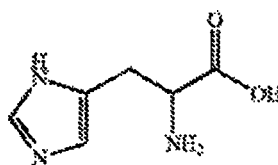
精氨酸 Arg R



CGU CUC CGA CGG AGA AAG

[0513]

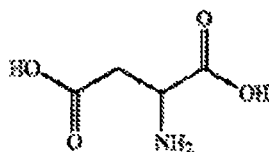
组氨酸 His H



CAU CAC

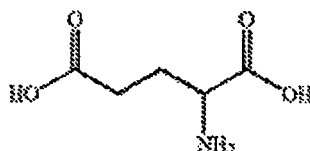
## 带负电残基

天冬氨酸 Asp D



GAU GAC

谷氨酸 Glu E



GAA GAG

[0514] 实施例8. 对体内示例性ABCA4反式剪接分子的评估

[0515] 本研究的主要目的是确认在非洲绿猴 (AGM) 中以  $4.3 \times 10^{10}$  ( $4.3E10$ ) vg/双眼 (OU; 两只眼睛) 的剂量视网膜下施用示例性AAV8-ABCA4RNA外显子编辑器构建体后1个月和3个月时间点的持续外显子编辑器表达和功效。示例性AAV8-ABCA4 RNA外显子编辑器构建体包括SEQ ID NO:77。示例性AAV8-ABCA4 RNA外显子编辑器构建体编码的外显子编辑器包含与图13中所述相同的调节元件(增强子、启动子和5'UTR), 包括阻止NSP翻译超出SD的双终止密码子, 但它不包含 $3 \times U1$ 和ARE特征, 因此与包括增效子( $3 \times U1$ )和翻译增效子(ARE)的构建体相比, 产生明显更高水平的NSP。使用ddPCR测定法, 利用靶向外显子编辑器ABCA4 CDS内的末端外显子22的正向引物和探针, 以及跨越外显子编辑器构建体中的40聚体接头和双终止密码子序列的反向引物来测量AAV8-ABCA4 RNA外显子编辑器构建体滴度。相同的ddPCR测定法已用于计算其它示例性ABCA4RNA外显子编辑器构建体的滴度。

[0516] RT-qPCR用于测量编辑的ABCA4 RNA、天然NHP ABCA4 RNA和示例性AAV8-ABCA4 RNA外显子编辑器构建体的表达, 全部相对于管家基因RNF20归一化。4周(1个月)时ABCA4

RNA替换百分比在14.4%-31.5%范围内,12周(3个月)时在26.3%-37.8%范围内。4周时每10ngRNA的外显子编辑器拷贝数范围为1.3E6至0.6E6,12周时为2.1E6至3.0E6。这些结果表明,本文所述的AAV包装的ABCA4RNA外显子编辑器构建体的转导引起编码的RNA外显子编辑器的持久表达和持久活性,如观察到的替换%所反映。事实上,替换%随着时间的推移不断增加。参见图22A和22B。

[0517] 通过裂隙灯显微镜检查进行的活体眼部评估显示,少数动物出现与视网膜下注射相关的轻度至中度炎症,炎症在第7天达到高峰,并大部分在第28天消退。眼底图像显示眼部变化与视网膜下注射一致。在研究过程中,所有动物的眼内压(IOP)、视网膜厚度和视网膜体积保持稳定。用示例性AAV8-ABCA4 RNA外显子编辑器构建体或媒介物处理的眼睛的共焦扫描激光检眼镜(cSLO)和彩色眼底照相显示出注射区域类似程度的变化。

[0518] 每天两次的一般观察并未发现指示示例性AAV8-ABCA4 RNA外显子编辑器构建体产生不良全身影响的临床体征。研究期间,体重、呼吸速率、心率和体温保持稳定,在发生的变异中没有明显的与AAV8-ABCA4RNA外显子编辑器构建体相关的模式。没有临床病理学[临床化学、全血细胞计数(CBC)、凝血特征]异常可归因于AAV8-ABCA4 RNA外显子编辑器构建体的施用。

[0519] 组织病理学分析证实,所有动物的外侧膝状核、视束、视交叉、脾和肝样品均在正常界限内。根据眼部组织病理学分析,未观察到媒介物或AAV8-ABCA4 RNA外显子编辑器构建体处理的动物中出现免疫浸润或视网膜解剖学异常的迹象。在用AAV8-ABCA4 RNA外显子编辑器构建体处理的两只动物和用媒介物处理的一只动物中,注意到平坦锯齿缘(玻璃体切除部位)的视网膜色素上皮(RPE)轻微至轻度增生和迁移。这些发现与注射程序一致,并且被认为与测试物品(构建体)无关。

[0520] 这些研究表明,与ABCA4-01(参见例如实施例8-10)共有的选定调节元件和40聚体接头和BD赋予AGM中长达3个月时间点的持续表达和外显子编辑活性。参见图22A和22B。

[0521] 实施例9. 调查性毒理学研究:体内示例性ABCA4反式剪接分子ABCA4-01的评估

[0522] 调查性毒理学研究:在食蟹猴中在单次视网膜下注射3.5E11或1E12vg/眼后1个月和6个月完成示例性AAV8-ABCA4 RNA外显子编辑器构建体ABCA4-01(SEQ ID NO:90)(其包括包含SEQ ID NO:78的RNA外显子编辑器)的外显子编辑效率的评估和调查性安全性评估。

[0523] 本研究的主要目的是确认包含SEQ ID NO:78的示例性AAV8-ABCA4 RNA外显子编辑器(ABCA4-01;SEQ ID NO:90)在1个月和6个月时间点的体内安全性和功效。这项研究评估了两种剂量水平;3.5E11和1E12vg/眼,对雄性食蟹猴每只眼(OU)进行单次视网膜下注射。1个月组中每个剂量水平n=2,媒介物组n=2。6个月组每个剂量水平n=3,媒介物组n=2。本研究的所有给药已完成。

[0524] 如本文所述,在神经视网膜样品中测量ABCA4-01载体基因组DNA拷贝(VGC)生物分布、RNA表达、外显子编辑活性(RNA替换%)和编辑的人-NHP嵌合ABCA4蛋白。

[0525] 此外,还对视交叉、视束、视神经的VGC进行了评估。收集以下非眼部组织:脑、背根神经节(腰椎)、心脏、肾、肝、胆囊、外侧膝状核、肺、淋巴结(下颌和肠系膜)、脾和睾丸。最初的计划是仅分析1个月时间点的1E12vg/眼剂量组的所有非眼部组织,以评估下游和未来分析需求的必要。

[0526] 如本文所述,RT-qPCR(或ddPCR)用于测量编辑的ABCA4 RNA、天然NHPABCA4RNA和

ABCA4-01外显子编辑器RNA的表达,全部相对于管家RNF20归一化。除了神经视网膜样品外,仅评估对ABCA4-01 VGC呈阳性的组织的ABCA4-01外显子编辑器RNA表达。

[0527] 每只动物的一只眼睛由经过认证的病理学家进行评估,并生成一份签名的组织病理学报告。收集所有标准全身组织(包括肾上腺、任何肉眼可见的病变、主动脉(升主动脉)、骨髓、脑、食道、胆囊、心脏(左心室壁)、肾、大肠、外侧膝状核、肝、肺、肠系膜淋巴结、视交叉、视束、胰腺、坐骨神经、骨骼肌、皮肤、小肠(空肠)、脊髓、脾、胃、下颌淋巴结、睾丸、胸腺和膀胱),固定并嵌入石蜡块中以支持非临床毒理学评估。意图将所有动物的任何肉眼可见的病变、骨髓涂片(肋骨)、背根神经节、心脏、肾、外侧膝状核、肝、肺、视交叉、视神经、视束和脾带到载玻片上,用H&E染色,并由经过认证的病理学家进行评估。所有其它组织学样品均保存起来,以备将来有需要时进行研究。

[0528] 调查性毒理学研究的初步观察结果:

[0529] 注射3.5E11剂量ABCA4-01后1个月的结果显示体内替换率为54%。这些结果强调了ABCA4-01的治疗功效潜力,至少因为观察到的替换%水平表明,在感光细胞和RPE细胞中产生了治疗有效量的反式剪接ABCA4蛋白,从而使内源ABCA4活性降低或缺失的细胞中恢复了野生型ABCA4功能。

[0530] 迄今为止,两种剂量的ABCA4-01均具有良好的耐受性。这里总结了截至研究第35天的眼科检查初步结果。给药后2-3天、1周、2周、4周或5周监测每只眼睛的炎症迹象。检查房水细胞、玻璃体细胞、结膜状态、虹膜充血、球结膜水肿、分泌物、眼睑、PLR、角膜混浊、角膜沉淀物、角膜血管化、房水闪辉、房水细胞颜色、玻璃体细胞颜色、玻璃体混浊、晶状体和眼底。此外,还测量眼内压。

[0531] 在一些眼睛中和某些时间点看到了因视网膜下程序而导致的轻度至中度的眼部炎症。根据眼科检查结果,适时使用抗炎药治疗的治疗计划。较低阈值(房水细胞评分>3或3且有房水闪辉,无前玻璃体细胞或玻璃体混浊的迹象)触发结膜下注射地塞米松(dexamethasone)(0.1ml,10mg/ml),较高阈值(房水细胞评分大于3或3且有房水闪辉和前玻璃体细胞或玻璃体混浊的迹象)触发玻璃体内注射曲安奈德(triamcinolone)。迄今为止,接受3.5E11vg/眼剂量ABCA4-01的任何动物均无需接受治疗。在一只媒介物处理的动物中,在72小时时施用曲安奈德(仅OD)。在1E12剂量/眼组中,一只动物在第2天需要地塞米松,在第7天需要曲安奈德(仅OS),并在第29天接受地塞米松治疗(仅OD),第二只动物在第8天需要OU曲安奈德治疗,第三只动物在第29天接受曲安西龙(仅OD)。在每种情况下,后续眼部检查均显示炎症减轻。值得注意的是,根据设计,本研究中的动物没有接受任何预防性皮质类固醇或其它免疫抑制方案,以防止掩盖任何不利的免疫反应,尤其是针对ABCA4-01编码的外显子编辑器。上述炎症与视网膜下注射和对较高AAV8剂量的反应一致。因此,这项调查性研究非常适合帮助界定后续研究中的高剂量水平。

[0532] 在整个研究过程中,所有组的眼内压都保持稳定。使用cSLO/OCT对眼睛进行成像显示眼底变化与视网膜下注射一致,并且没有显示归因于ABCA4-01施用的变化(研究正在进行中,对第8天的数据进行评估)。

[0533] 在所有动物中施用ABCA4-01之前对抗AAV8 NAb进行评估。这些滴度范围从低于1:10至 $\geq 1:100$ 。预先存在的NAb滴度数据用于将动物分配到三组中。在4周或5周时,接受媒介物给药的动物的NAb滴度在 $\geq 1:10$ 至 $\geq 1:100$ 范围内,而接受任一剂量ABCA4-01给药的动物

的NAb滴度范围为 $\geq 1:100$ 至 $1 \geq 1:1000$ ,这与针对可能从视网膜下水泡泄漏到全身循环中的少量AAV8衣壳的预期适应性免疫反应一致。

[0534] 笼边监测动物健康状况并测量体重未显示任何不良事件的迹象。

[0535] 实施例10. 确定性GLP毒理学研究:体内示例性ABCA4反式剪接分子的评估

[0536] 在食蟹猴中在单次视网膜下注射 $3.5E11$ 或 $1E12$  vg/眼后1个月和6个月完成示例性AAV8-ABCA4 RNA外显子编辑器构建体ABCA4-01 (SEQ ID NO:90) (其包含SEQ ID NO:78)的外显子编辑效率的评估和调查性安全性评估。

[0537] 确定性GLP毒理学研究:本研究的目的是在食蟹猴中在单次视网膜下施用从 $4E10$ 至 $1E12$ vg/眼范围内不同剂量的ABCA4-01后3个月和6个月,评估包含SEQ ID NO:78的示例性AAV8-ABCA4RNA外显子编辑器ABCA4-01 (SEQ ID NO:90)的视网膜下耐受性。本研究选择的剂量为 $3.5E11$ 或 $1E12$ vg/眼。招募雄性和雌性NHP并且每个视网膜下水泡(OU)注射 $100\mu l$ 。该研究旨在评估每个AAV8-ABCA4 RNA外显子编辑器ABCA4-01 (SEQ ID NO:90)和剂量水平组的每个时间点 $n=3$ 只动物,以及每个时间点 $n=2$ 只媒介物动物。

[0538] 终点包括生命期间的眼部和全身安全性评估以及尸检时收集的样品。在生命期间对一般健康和眼部健康状况的评估如下:详细观察(每周)、体重(测试前、第-1天和之后每周)、食物消耗(每周)、眼科检查(测试前、第1周和第5周以及终末尸检之前;间接检眼镜检查和裂隙灯生物显微镜检查)、眼内压测量(测试前、第1周和第5周以及终末尸检之前)、共焦激光检眼镜检查(测试前、第5周和终末尸检之前)、光学相干断层扫描(测试前、第5周和终末尸检之前)、临床病理学(血液学、凝血、临床化学:测试前两次、第1天给药前、第2天、第8天、第15天、第22天、第29天、第35天、第43天、第50天、第57天以及此后每月直至尸检)。

[0539] 对于每只动物,检查一只眼睛的组织病理学。解剖对侧眼球,并快速冷冻神经视网膜样品和RPE/脉络膜样品。对映射到视网膜处理区域的神经视网膜样品进行ABCA4-01 VGC生物分布、RNA表达和外显子编辑活性(RNA替换%)分析。在神经视网膜样品可用(在确认提取了足够的DNA和RNA以进行所需的分析后)之前,通过合格的IA-LC-MS/MS测定法对编辑的人-NHP嵌合ABCA4蛋白进行评估。

[0540] 此外,在距离上述调查性毒理学研究1个月时仅在 $1E12$ vg/眼剂量组中对ABCA4-01 VGC呈阳性的组织中对VGC进行评估。同样,在对ABCA4-01 VGC呈阳性的组织中测量ABCA4-01 RNA表达。

[0541] 收集所有标准全身组织(包括肾上腺、任何肉眼可见的病变、主动脉、骨髓、脑、食道、胆囊、心脏(左心室壁)、肾、大肠、外侧膝状核、肝、肺、肠系膜淋巴结、视交叉、视束、胰腺、坐骨神经、骨骼肌、皮肤、小肠(空肠)、脊髓、脾、胃、下颌淋巴结、睾丸、胸腺和膀胱),固定并嵌入石蜡块中以支持非临床毒理学评估。

[0542] 所有动物的眼睛均由经过认证的病理学家评估。将在调查性毒理学研究中发现对ABCA4-01 VGC呈阳性的相同非眼部组织类型安装在载玻片上,用H&E染色,然后进行评估。所有其它样品均保存起来,以备将来有需要时进行评估。

[0543] 在上述调查性和确定性GLP研究中,进行了以下评估:ABCA4-01 DNA生物分布和RNA表达、外显子编辑的ABCA4 RNA和蛋白质、临床病理学(化学、血液学、凝血)、生命期间的眼部测量(裂隙灯、光学相干断层扫描(OCT)/cSLO、眼内压)和组织病理学。针对任何免疫反应风险,在调查性和GLP毒理学研究中,均在基线和给药后从血清中测量抗AAV8中和抗体。

此外,在GLP研究中,评估基线、生命期间和终末循环细胞因子和补体,并对已确认ABCA4-01 DNA生物分布的任何组织进行组织病理学评估。该研究的结果表明,ABCA4-01治疗是安全的,且不会引发任何不良事件。

[0544] 除上述内容外,图30A-30C还呈现了证明在用示例性AAV8 RNA外显子编辑器构建体(ABCA4-01;SEQ ID NO:90)处理六(6)个月后,在NHP中在体内实现了强大的ABCA4 RNA和蛋白质替换的结果。图30A显示了病毒基因组拷贝数(VGC)(x轴)和外显子编辑器RNA拷贝数(y轴);图30B显示了ABCA4 RNA替换%;并且图30C显示了在用ABCA4-01处理6个月后食蟹猴(n=3)的ABCA4蛋白质替换情况(显示了3.5E11vg/眼的示例性剂量的结果)。使用基于液滴数字聚合酶链反应(ddPCR)的测定法评估DNA和RNA水平。使用合格的免疫亲和质谱(IA-MS)测定法对总蛋白质百分比(编辑的ABCA4相对于编辑的和NHP天然蛋白质的总和)进行定量。该体内研究中使用的示例性AAV8 RNA外显子编辑器构建体包括包含SEQ ID NO:90的RNA外显子编辑器。(B)和(C)中显示了四(4)个NHP的结果,并指定为1-4:1(用媒介物处理);2-4(用3.5E11vg/眼处理)。

[0545] 简单地说,本发明人用单次视网膜下剂量的ABCA4-01(3.5E11vg/眼)处理野生型食蟹猴。然后评估AAV载体基因组DNA拷贝(VGC)的生物分布、外显子编辑器RNA表达以及6个月时间点编辑的ABCA4RNA和蛋白质的百分比。外显子编辑器RNA表达与VGC有很好的相关性(图30A),并产生强大的RNA替换%(将天然ABCA4RNA拷贝转化为编辑的拷贝),该RNA替换%计算为神经视网膜样品中存在的总(编辑+天然)ABCA4 RNA群体中编辑的ABCA4 RNA的百分比(图30B)。如图30B所示,对于该剂量的ABCA4-01,RNA替换百分比范围约为40%-60%。这使所得的人-NHP嵌合ABCA4蛋白达到治疗水平,如通过合格的IA-MS测定法所测量(图30C)。更具体地说,发现由这些动物中的ABCA4 RNA编辑产生的人NHP嵌合ABCA4蛋白占测试样品中总(人-NHP+NHP)ABCA4蛋白的20-40%。此外,本发明人还证明,用通过单次视网膜下注射递送的较高剂量的ABCA4-01处理的野生型食蟹猴中RNA替换高达66%,并且人/NHP嵌合ABCA4蛋白高达45%。这些结果表明,ABCA4-01在NHP中实现了治疗相关水平的编辑ABCA4蛋白表达,这些水平超过了先前在Abca4 KO小鼠模型中显示赋予治疗益处的拯救ABCA4蛋白的水平。

[0546] 实施例11:包含RNA外显子编辑器SEQ ID NO:78的AAV载体质粒

[0547] 编码ABCA4-01的质粒(SEQ ID NO:82)是一种13,232碱基对(bp) AAV载体质粒,其包含RNA外显子编辑器(SEQ ID NO:78),两侧是AAV2 ITR(SEQ ID NO:90)。质粒骨架含有卡那霉素抗性(KanR)基因、细菌复制起点和两个源自视蛋白1内含子1的非编码区。这些后面序列被描述为‘填充’序列,旨在增加质粒的大小,使其超出AAV的包装限制,从而降低可能的衣壳化质粒DNA骨架组件的水平。图23显示了AAV载体质粒的示意图,表5呈现了有关其的更多细节。

[0548] 表5:包含SEQ ID NO:82的AAV载体质粒内的遗传元件

[0549]

| 特征                         | 质粒中的位置    | 长度<br>(bp) |
|----------------------------|-----------|------------|
| KanR                       | 810-1     | 810        |
| 填充: 人视蛋白 1 内含子 1 (第 2 半部分) | 1185-4127 | 2943       |
| 5' AAV2 ITR                | 4284-4413 | 130        |
| CMV 即刻早期增强子                | 4486-4789 | 304        |
| CMV 即刻早期启动子                | 4790-4993 | 204        |
| ABCA4 5' UTR               | 4994-5099 | 106        |
| ABCA4 编码序列(CDS)外显子 1-22    | 5100-8427 | 3328       |
| 剪接供体(SD)                   | 8428-8433 | 6          |

[0550]

| 特征                         | 质粒中的位置      | 长度<br>(bp) |
|----------------------------|-------------|------------|
| 双终止密码子                     | 8433-8438   | 6          |
| 40 聚体接头                    | 8439-8478   | 40         |
| 3×U1 位点                    | 8479-8549   | 71         |
| GM-CSF ARE                 | 8550-8595   | 46         |
| 结合结构域(BD)                  | 8596-8745   | 150        |
| MALAT1 三螺旋终止子              | 8746-8847   | 102        |
| 抗 Mut1 mascRNA             | 8848-8905   | 58         |
| 3' AAV2 ITR                | 8944-9073   | 130        |
| 填充: 人视蛋白 1 内含子 1 (第 1 半部分) | 9091-12040  | 2950       |
| Ori                        | 12467-13055 | 589        |

[0551] 实施例12.对离体示例性ABCA4反式剪接分子的评估

[0552] 为了评估人光感受器中的ABCA4外显子编辑,本发明人用ABCA4-01(一种包含SEQ ID NO:90的示例性AAV8 RNA外显子编辑器构建体)处理人供体视网膜外植体。本发明人发现,将来自多个供体的外植体与ABCA4-01(3.9E11vg)一起孵育21天,会导致大约20-30%的内源外显子1-22被ABCA4-01编码的相同对应的外显子替换的RNA替换(图31)。本次实验中的供体外植体来自之前未诊断出任何视网膜疾病的个体。值得注意的是,同一批ABCA4-01也用于上述的调查性NHP研究(图30),进一步增强了本发明人对ABCA4-01结果从NHP到人类受试者的可转化性的信心。

[0553] 实施例13.外显子编辑的ABCA4 mRNA序列的确认

[0554] 为了确认由外显子编辑产生的ABCA4成熟mRNA具有预测由精确的反式剪接反应产生的序列,包括正确的外显子22-外外显子23接合,本发明人从用编码ABCA4-01外显子编辑器的包含SEQ ID NO:78的质粒瞬时转染的ABCA4 KO细胞中分离编辑的ABCA4 RT-PCR产物并进行测序。简单地说,通过对从转染的ABCA4 KO细胞中分离的RNA进行RT-PCR,随后进行TOP0克隆和测序来确认全长外显子编辑的mRNA。桑格测序证实了多个随机选择的克隆中成熟mRNA的序列正确。除了确认正确的外显子22-外显子23接合之外,这些测序结果还证实了全长外显子编辑的ABCA4成熟mRNA分子在其整个长度上具有预期的序列,包括外显子编辑器衍生的跨越外显子1至22的CDS和跨越外显子23至50的内源序列。这些结果提供了对通过

ACDN-01处理介导的ABCA4外显子编辑会产生预期的正确的ABCA4成熟mRNA序列的额外信心。

[0555] 编号实施方案I

[0556] 1.一种核酸反式剪接分子,所述核酸反式剪接分子包含:

[0557] (a) 接头结构域,所述接头结构域包含SEQ ID NO:27或与SEQ ID NO:27具有至少90%同一性、至少91%同一性、至少92%同一性、至少93%同一性、至少94%同一性、至少95%同一性、至少96%同一性、至少97%同一性、至少98%同一性或至少99%同一性的序列、基本上由其组成或由其组成;以及

[0558] (b) 结合结构域,所述结合结构域退火到内源ABCA4 mRNA前体内的结合位点;以及

[0559] (c) 任选地,编码结构域序列(CDS),所述编码结构域序列包含ABCA4外显子;

[0560] (d) 任选地,剪接结构域;

[0561] (e) 任选地,增效子,所述增效子包含GGTAAGT或SEQ ID NO:61或与GGTAAGT或SEQ ID NO:61具有至少90%同一性、至少91%同一性、至少92%同一性、至少93%同一性、至少94%同一性、至少95%同一性、至少96%同一性、至少97%同一性、至少98%同一性或至少99%同一性的序列中的至少一者;

[0562] 以及

[0563] (f) 任选地,翻译增效子,所述翻译增效子包含由SEQ ID NO:63组成的序列或与SEQ ID NO:63具有至少90%同一性、至少91%同一性、至少92%同一性、至少93%同一性、至少94%同一性、至少95%同一性、至少96%同一性、至少97%同一性、至少98%同一性或至少99%同一性的序列,

[0564] 其中所述CDS;所述剪接结构域在存在时;所述接头结构域;所述增效子在存在时;所述翻译增效子在存在时;以及所述结合结构域在5'至3'方向上操作性连接。

[0565] 2.实施方案1的核酸反式剪接分子,其中所述结合结构域退火到所述内源ABCA4 mRNA前体中的内含子。

[0566] 3.实施方案2的核酸反式剪接分子,其中所述内源ABCA4mRNA前体中的所述内含子是内含子22。

[0567] 4.实施方案1-3中任一项的核酸反式剪接分子,其中所述结合结构域包含SEQ ID NO:18或SEQ ID NO:20、基本上由其组成或由其组成。

[0568] 5.实施方案1-4中任一项的核酸反式剪接分子,其中所述结合结构域基本上由SEQ ID NO:18或SEQ ID NO:20组成或由其组成。

[0569] 6.实施方案1-5中任一项的核酸反式剪接分子,其中所述结合结构域基本上由SEQ ID NO:20组成或由其组成。

[0570] 7.实施方案1-6中任一项的核酸反式剪接分子,其中所述CDS在存在时包含ABCA4外显子1-22或其变体。

[0571] 8.实施方案1-7中任一项的核酸反式剪接分子,其中所述CDS包含SEQ ID No:53-59中的任一者、基本上由其组成或由其组成。

[0572] 9.实施方案1-8中任一项的核酸反式剪接分子,所述核酸反式剪接分子包含:

[0573] (a) CDS,所述CDS包含SEQ ID NO:53-59中的任一者、基本上由其组成或由其组成;

[0574] (b) 接头结构域,所述接头结构域包含SEQ ID NO:27或与SEQ ID NO:27具有至少

90%同一性、至少91%同一性、至少92%同一性、至少93%同一性、至少94%同一性、至少95%同一性、至少96%同一性、至少97%同一性、至少98%同一性或至少99%同一性的序列、基本上由其组成或由其组成;以及

[0575] (c) 结合结构域,所述结合结构域包含SEQ ID NO:18或SEQ ID NO:20、基本上由其组成或由其组成;以及

[0576] (d) 任选地,剪接结构域,所述剪接结构域包含SEQ ID NO:4、6、8或10中的任一者、基本上由其组成或由其组成;

[0577] (e) 任选地,5'非翻译区;

[0578] (f) 任选地,3'转录终止子结构域;以及

[0579] (g) 任选地,至少一个在3'操作性连接到所述剪接结构域的终止密码子,其中,

[0580] 其中所述5'非翻译区在存在时;所述CDS;所述剪接结构域在存在时;所述至少一个终止密码子在存在时;所述接头结构域;所述增效子在存在时;所述翻译增效子在存在时;所述结合结构域;以及所述3'转录终止子结构域在存在时,在5'至3'方向上操作性连接。

[0581] 10. 实施方案9的核酸反式剪接分子,其中所述增效子包含三个串联排列的序列,其中所述三个序列中的每一个包含GGTAAGT或SEQ ID NO:61或与GGTAAGT或SEQ ID NO:61具有至少90%同一性、至少91%同一性、至少92%同一性、至少93%同一性、至少94%同一性、至少95%同一性、至少96%同一性、至少97%同一性、至少98%同一性或至少99%同一性的序列,或它们的任何组合。

[0582] 11. 实施方案9或10中任一项的核酸反式剪接分子,其中,所述增效子包含SEQ ID NO:62或与SEQ ID NO:62具有至少90%同一性、至少91%同一性、至少92%同一性、至少93%同一性、至少94%同一性、至少95%同一性、至少96%同一性、至少97%同一性、至少98%同一性或至少99%同一性的序列。

[0583] 12. 实施方案9-11中任一项的核酸反式剪接分子,其中所述剪接结构域包含SEQ ID NO:4 (GTAAGT)、基本上由其组成或由其组成。

[0584] 13. 实施方案9-12中任一项的核酸反式剪接分子,其中所述5'非翻译区包含SEQ ID NO:13或SEQ ID NO:64、基本上由其组成或由其组成。

[0585] 14. 实施方案9-13中任一项的核酸反式剪接分子,其中所述3'转录终止子结构域编码形成三螺旋结构的核糖核酸,所述三螺旋结构将所述反式剪接分子的3'端封端。

[0586] 15. 实施方案9-14中任一项的核酸反式剪接分子,其中所述3'转录终止子结构域包含SEQ ID NO:33或SEQ ID NO:66、基本上由其组成或由其组成。

[0587] 16. 实施方案15的核酸反式剪接分子,其中所述3'转录终止子结构域包含如下核酸序列,所述核酸序列与SEQ ID NO:66具有至少80%同一性;与SEQ ID NO:66具有至少85%同一性;与SEQ ID NO:66具有至少90%同一性;与SEQ ID NO:66具有至少91%同一性;与SEQ ID NO:66具有至少92%同一性;与SEQ ID NO:66具有至少93%同一性;与SEQ ID NO:66具有至少94%同一性;与SEQ ID NO:66具有至少95%同一性;与SEQ ID NO:66具有至少96%同一性;与SEQ ID NO:66具有至少97%同一性;与SEQ ID NO:66具有至少98%同一性;与SEQ ID NO:66具有至少99%同一性;或与SEQ ID NO:66 100%同一。

[0588] 17. 实施方案1-16中任一项的核酸反式剪接分子,所述核酸反式剪接分子在5'至

3'方向上包含操作性连接的以下各者:

- [0589] (a) 5'非翻译区;
  - [0590] (b) CDS, 所述CDS包含SEQ ID NO:53-59中的任一者、基本上由其组成或由其组成;
  - [0591] (c) 剪接结构域;
  - [0592] (d) 至少一个终止密码子;
  - [0593] (e) 接头结构域, 所述接头结构域包含SEQ ID NO:27、基本上由其组成或由其组成;
  - [0594] (f) 至少一个增效子;
  - [0595] (g) 翻译增效子;
  - [0596] (h) 结合结构域, 所述结合结构域包含SEQ ID NO:20、基本上由其组成或由其组成; 以及
  - [0597] (i) 3'转录终止子结构域。
- [0598] 18. 实施方案17的核酸反式剪接分子, 其中
- [0599] (a) 所述5'非翻译区包含SEQ ID NO:13或SEQ ID NO:64、基本上由其组成或由其组成;
  - [0600] (b) 所述剪接结构域包含SEQ ID NO:4(GTAAGT)、基本上由其组成或由其组成;
  - [0601] (c) 所述至少一个终止密码子包含两个终止密码子, 其中所述两个终止密码子中的第一个终止密码子的序列与所述剪接结构域的序列部分重叠并且其中所述剪接结构域的序列与所述两个终止密码子组合包含SEQ ID NO:68、基本上由其组成或由其组成;
  - [0602] (d) 所述至少一个增效子包含三个串联排列的序列、基本上由其组成或由其组成, 其中所述三个序列中的每一个包含GGTAAGT或SEQ ID NO:61, 或它们的任何组合;
  - [0603] (e) 所述翻译增效子包含SEQ ID NO:63、基本上由其组成或由其组成; 并且
  - [0604] (f) 所述3'转录终止子结构域包含SEQ ID NO:66、基本上由其组成或由其组成。
- [0605] 19. 实施方案1-18中任一项的核酸反式剪接分子, 其中所述CDS包含SEQ ID NO:56、基本上由其组成或由其组成。
- [0606] 20. 实施方案18或实施方案19的核酸反式剪接分子, 其中所述核酸反式剪接分子包含SEQ ID NO:69、SEQ ID NO:78或SEQ ID NO:90、基本上由其组成或由其组成。
- [0607] 21. 实施方案1-16中任一项的核酸反式剪接分子, 其中所述核酸反式剪接分子包含SEQ ID NO:70或SEQ ID NO:71、基本上由其组成或由其组成。
- [0608] 22. 实施方案1-16中任一项的核酸反式剪接分子, 其中所述核酸反式剪接分子包含与SEQ ID NO:40-44、46或48-50中的任一者具有至少80%同一性的核酸序列。
- [0609] 23. 一种核糖核酸反式剪接分子, 所述核糖核酸反式剪接分子从实施方案1-22中任一项的核酸反式剪接分子转录。
- [0610] 24. 一种核酸反式剪接分子, 所述核酸反式剪接分子包含:
- [0611] (a) 接头结构域, 所述接头结构域包含SEQ ID NO:27或与SEQ ID NO:27具有至少90%同一性的序列、基本上由其组成或由其组成; 以及
  - [0612] (b) 任选地, 增效子, 所述增效子包含GGTAAGT或61中的任一者或它们的任何组合、基本上由其组成或由其组成; 以及
  - [0613] (c) 任选地, 翻译增效子, 所述翻译增效子包含SEQ ID NO:63、基本上由其组成或

由其组成,其中所述接头结构域;所述增效子在存在时;以及所述翻译增效子在存在时,在5'至3'方向上操作性连接。

[0614] 25. 实施方案24的核酸反式剪接分子,其中所述增效子在存在时,包含SEQ ID NO: 62、基本上由其组成或由其组成。

[0615] 26. 实施方案24或25中任一项的核酸反式剪接分子,所述核酸反式剪接分子还包含结合结构域,所述结合结构域退火到内源ABCA4 mRNA前体内的结合位点,

[0616] 其中所述接头结构域;所述增效子在存在时;所述翻译增效子在存在时;以及所述结合结构域,在5'至3'方向上操作性连接。

[0617] 27. 实施方案26的核酸反式剪接分子,所述核酸反式剪接分子还包括包含ABCA4外显子的CDS,

[0618] 其中所述CDS;所述接头结构域;所述增效子在存在时;所述翻译增效子在存在时;以及所述结合结构域在存在时,在5'至3'方向上操作性连接。

[0619] 28. 一种5'核酸反式剪接分子,所述5'核酸反式剪接分子包含:

[0620] (a) 增效子,所述增效子包含GGTAAGT或SEQ ID NO:61或与GGTAAGT或SEQ ID NO: 61具有至少90%同一性、至少91%同一性、至少92%同一性、至少93%同一性、至少94%同一性、至少95%同一性、至少96%同一性、至少97%同一性、至少98%同一性或至少99%同一性的序列中的至少一者、基本上由其组成或由其组成;以及

[0621] (b) 任选地,接头结构域,所述接头结构域包含SEQ ID NO:27或与SEQ ID NO:27具有至少90%同一性的序列、基本上由其组成或由其组成;

[0622] (c) 任选地,翻译增效子,所述翻译增效子包含SEQ ID NO:63、基本上由其组成或由其组成,

[0623] 其中所述接头结构域在存在时;所述增效子;以及所述翻译增效子在存在时,在5'至3'方向上操作性连接。

[0624] 29. 一种5'核酸反式剪接分子,所述5'核酸反式剪接分子包含:

[0625] (a) 增效子,所述增效子包含SEQ ID NO:62或与SEQ ID NO:62具有至少90%同一性的序列、基本上由其组成或由其组成;以及

[0626] (b) 任选地,接头结构域,所述接头结构域包含SEQ ID NO:27或与SEQ ID NO 27具有至少90%同一性的序列、基本上由其组成或由其组成;

[0627] (c) 任选地,翻译增效子,所述翻译增效子包含SEQ ID NO:63、基本上由其组成或由其组成,

[0628] 其中所述接头结构域在存在时;所述增效子;以及所述翻译增效子在存在时,在5'至3'方向上操作性连接。

[0629] 30. 实施方案28或实施方案29的核酸反式剪接分子,所述核酸反式剪接分子还包含结合结构域,所述结合结构域退火到内源ABCA4 mRNA前体内的结合位点,

[0630] 其中所述接头结构域在存在时;所述增效子;所述翻译增效子在存在时;以及所述结合结构域,在5'至3'方向上操作性连接。

[0631] 31. 实施方案28-30中任一项的核酸反式剪接分子,所述核酸反式剪接分子还包括包含ABCA4外显子的CDS,

[0632] 其中所述CDS;所述接头结构域在存在时;所述增效子;所述翻译增效子在存在时;

以及所述结合结构域在存在时,在5'至3'方向上操作性连接。

[0633] 32.一种核酸反式剪接分子,所述核酸反式剪接分子包含:

[0634] (a) 翻译增效子,所述翻译增效子包含SEQ ID NO:63或与SEQ ID NO:63具有至少90%同一性的序列、基本上由其组成或由其组成;

[0635] (b) 任选地,增效子,所述增效子包含GGTAAGT或61中的任一者或它们的任何组合、基本上由其组成或由其组成;以及

[0636] (c) 任选地,接头结构域,所述接头结构域包含SEQ ID NO:27或与SEQ ID NO:27具有至少90%同一性的序列、基本上由其组成或由其组成,其中所述接头结构域在存在时;所述增效子在存在时;以及所述翻译增效子,在5'至3'方向上操作性连接。

[0637] 33.实施方案32的核酸反式剪接分子,其中所述增效子在存在时,包含SEQ ID NO:62、基本上由其组成或由其组成。

[0638] 34.实施方案32或实施方案33的核酸反式剪接分子,所述核酸反式剪接分子还包含结合结构域,所述结合结构域退火到内源ABCA4 mRNA前体内的结合位点,

[0639] 其中所述接头结构域在存在时;所述增效子在存在时;所述翻译增效子;以及所述结合结构域,在5'至3'方向上操作性连接。

[0640] 35.实施方案32-34中任一项的核酸反式剪接分子,所述核酸反式剪接分子还包括包含ABCA4外显子的CDS,

[0641] 其中所述CDS;所述接头结构域在存在时;所述增效子在存在时;所述翻译增效子;以及所述结合结构域在存在时,在5'至3'方向上操作性连接。

[0642] 36.一种核酸反式剪接分子,所述核酸反式剪接分子包含:

[0643] 包含接头结构域和增效子的盒,其中所述盒包含SEQ ID NO:72、基本上由其组成或由其组成。

[0644] 37.一种核酸反式剪接分子,所述核酸反式剪接分子包含:

[0645] 包含接头结构域、增效子和翻译增效子的盒,其中所述盒包含SEQ ID NO:73、基本上由其组成或由其组成。

[0646] 38.一种核酸反式剪接分子,所述核酸反式剪接分子包含:

[0647] (a) 3'转录终止子结构域,所述3'转录终止子结构域包含SEQ ID NO:66或与SEQ ID NO:66具有至少95%同一性的序列、基本上由其组成或由其组成。

[0648] 39.实施方案38的核酸反式剪接分子,其中所述3'转录终止子结构域包含如下核酸序列,所述核酸序列与SEQ ID NO:66具有至少96%同一性;与SEQ ID NO:66具有至少97%同一性;与SEQ ID NO:66具有至少98%同一性;与SEQ ID NO:66具有至少99%同一性;或与SEQ ID NO:66100%同一。

[0649] 40.实施方案38或实施方案39的核酸反式剪接分子,其中所述3'转录终止子结构域包含SEQ ID NO:66、基本上由其组成或由其组成。

[0650] 编号实施方案II

[0651] 1.一种核酸反式剪接分子,所述核酸反式剪接分子包含:

[0652] (a) 编码结构域序列(CDS),所述编码结构域序列包含ABCA4外显子;

[0653] (b) 接头结构域,所述接头结构域包含SEQ ID NO:27或与SEQ ID NO:27具有至少90%同一性的序列;以及

- [0654] (c) 结合结构域,所述结合结构域退火到内源ABCA4 mRNA前体内的结合位点,
- [0655] 其中所述CDS、所述接头结构域和所述结合结构域在5'至3'方向上操作性连接。
- [0656] 2.一种核酸反式剪接分子,所述核酸反式剪接分子包含:
- [0657] (a) CDS,所述CDS包含SEQ ID NO:53-55中的任一者的变体、基本上由其组成或由其组成,其中所述SEQ ID NO:53-55中的任一者的变体在表3中列出的至少一个隐蔽剪接位点中包含至少一个核苷酸变异,其中所述至少一个核苷酸变异减少包含至少一个核苷酸变异的所述至少一个隐蔽剪接位点的每一者处隐蔽剪接位点的使用;以及
- [0658] (b) 任选地,接头结构域;以及
- [0659] (c) 任选地,结合结构域,所述结合结构域退火到内源ABCA4mRNA前体内的结合位点,
- [0660] 其中所述CDS;所述接头结构域在存在时;以及所述结合结构域在存在时;在5'至3'方向上操作性连接。
- [0661] 3.实施方案2的核酸反式剪接分子,其中所述SEQ ID NO:55的变体包含SEQ ID NO:56-59中的任一者。
- [0662] 4.实施方案3的核酸反式剪接分子,其中所述SEQ ID NO:55的变体包含SEQ ID NO:56。
- [0663] 5.前述实施方案中任一项的核酸反式剪接分子,其中所述接头结构域在存在时包含与SEQ ID NO:27具有至少91%同一性、至少92%同一性、至少93%同一性、至少94%同一性、至少95%同一性、至少96%同一性、至少97%同一性、至少98%同一性或至少99%同一性的序列。
- [0664] 6.前述实施方案中任一项的核酸反式剪接分子,其中所述接头结构域在存在时包含SEQ ID NO:27、基本上由其组成或由其组成。
- [0665] 7.前述实施方案中任一项的核酸反式剪接分子,其中所述结合结构域在存在时退火到所述内源ABCA4 mRNA前体中的内含子。
- [0666] 8.实施方案7的核酸反式剪接分子,其中所述内源ABCA4mRNA前体中的所述内含子是内含子22。
- [0667] 9.前述实施方案中任一项的核酸反式剪接分子,其中所述结合结构域在存在时包含SEQ ID NO:18或SEQ ID NO:20。
- [0668] 10.前述实施方案中任一项的核酸反式剪接分子,其中所述结合结构域基本上由SEQ ID NO:18或SEQ ID NO:20组成或由其组成
- [0669] 11.前述实施方案中任一项的核酸反式剪接分子,其中所述结合结构域在存在时基本上由SEQ ID NO:20组成或由其组成。
- [0670] 12.实施方案1或5-11中任一项的核酸反式剪接分子,其中所述CDS包含ABCA4外显子1-22或其变体。
- [0671] 13.实施方案12的核酸反式剪接分子,其中所述CDS包含SEQ ID NO:53-59中的任一者、基本上由其组成或由其组成。
- [0672] 14.一种核酸反式剪接分子,所述核酸反式剪接分子包含:
- [0673] (a) 接头结构域,所述接头结构域包含SEQ ID NO:27或与SEQ ID NO:27具有至少91%同一性、至少92%同一性、至少93%同一性、至少94%同一性、至少95%同一性、至少

96%同一性、至少97%同一性、至少98%同一性或至少99%同一性的序列、基本上由其组成或由其组成；

[0674] (b) 结合结构域,所述结合结构域包含SEQ NO:18或ID SEQ ID NO:20、基本上由其组成或由其组成；以及

[0675] (c) 任选地,CDS,所述CDS包含ABCA4外显子；

[0676] 其中所述CDS在存在时；所述接头结构域；以及所述结合结构域在5'至3'方向上操作性连接。

[0677] 15. 实施方案14的核酸反式剪接分子,其中所述CDS在存在时包含SEQ ID NO:53-59中的任一者、基本上由其组成或由其组成。

[0678] 16. 一种核酸反式剪接分子,所述核酸反式剪接分子包含：

[0679] (a) CDS,所述CDS包含SEQ ID NO:53-59中的任一者、基本上由其组成或由其组成；

[0680] (b) 接头结构域,所述接头结构域包含SEQ ID NO:27或与SEQ ID NO:27具有至少91%同一性、至少92%同一性、至少93%同一性、至少94%同一性、至少95%同一性、至少96%同一性、至少97%同一性、至少98%同一性或至少99%同一性的序列、基本上由其组成或由其组成；以及

[0681] (c) 任选地,结合结构域,所述结合结构域包含SEQ NO:18或SEQ ID NO:20、基本上由其组成或由其组成，

[0682] 其中所述CDS；所述接头结构域；以及所述结合结构域在存在时,在5'至3'方向上操作性连接。

[0683] 17. 前述实施方案中任一项的核酸反式剪接分子,所述核酸反式剪接分子还包含增效子,所述增效子包括包含GGTAAGT或SEQ ID NO:61的至少一个序列。

[0684] 18. 实施方案17的核酸反式剪接分子,其中所述增效子在3'可操作地连接到所述接头结构域。

[0685] 19. 实施方案17或18中任一项的核酸反式剪接分子,其中所述增效子包含三个串联排列的序列,其中所述三个序列中的每一个包含GGTAAGT或SEQ ID NO:61。

[0686] 20. 实施方案17-19中任一项的核酸反式剪接分子,其中所述增效子包含三个串联排列的序列,其中所述三个序列包含SEQ ID NO:62。

[0687] 21. 前述实施方案中任一项的核酸反式剪接分子,所述核酸反式剪接分子还包含翻译增效子,其中所述翻译增效子包含SEQ ID NO:63或与SEQ ID NO:63具有至少90%同一性的序列,其中SEQ ID NO:63或所述与SEQ ID NO:63具有至少90%同一性的序列在3'可操作地连接到所述增效子。

[0688] 22. 实施方案21的核酸反式剪接分子,其中所述翻译增效子与SEQ ID NO:63具有至少91%同一性,与SEQ ID NO:63具有至少92%同一性,与SEQ ID NO:63具有至少93%同一性,与SEQ ID NO:63具有至少94%同一性,与SEQ ID NO:63具有至少95%同一性,与SEQ ID NO:63具有至少96%同一性,与SEQ ID NO:63具有至少97%同一性,与SEQ ID NO:63具有至少98%同一性,或与SEQ ID NO:63具有至少99%同一性。

[0689] 23. 实施方案21或22中任一项的核酸反式剪接分子,其中所述翻译增效子包含SEQ ID NO:63、基本上由其组成或由其组成。

[0690] 24. 前述实施方案中任一项的核酸反式剪接分子,所述核酸反式剪接分子还包括

包含GTAAGT、GTAAGG、GTAAGC或GTAAC中任一者的剪接结构域。

[0691] 25. 前述实施方案中任一项的核酸反式剪接分子, 所述核酸反式剪接分子还包括包含GTAAGT、基本上由其组成或由其组成的剪接结构域。

[0692] 26. 前述实施方案中任一项的核酸反式剪接分子, 所述核酸反式剪接分子还包含5'非翻译区。

[0693] 27. 实施方案26的核酸反式剪接分子, 其中所述5'非翻译区包含SEQ ID NO:13或SEQ ID NO:64、基本上由其组成或由其组成。

[0694] 28. 前述实施方案中任一项的核酸反式剪接分子, 所述核酸反式剪接分子还包含3'转录终止子结构域。

[0695] 29. 实施方案28的核酸反式剪接分子, 其中所述3'转录终止子结构域编码形成三螺旋结构的核糖核酸, 所述三螺旋结构保护所述反式剪接分子的3'端。

[0696] 30. 实施方案28或实施方案29的核酸反式剪接分子, 其中所述3'转录终止子结构域包含野生型MALAT1序列(SEQ ID NO:33)或与SEQ ID NO:33具有至少80%同一性的核酸序列。

[0697] 31. 实施方案30的核酸反式剪接分子, 其中所述3'转录终止子结构域包含如下核酸序列, 所述核酸序列与SEQ ID NO:33具有至少80%同一性; 与SEQ ID NO:33具有至少85%同一性; 与SEQ ID NO:33具有至少90%同一性; 与SEQ ID NO:33具有至少91%同一性; 与SEQ ID NO:33具有至少92%同一性; 与SEQ ID NO:33具有至少93%同一性; 与SEQ ID NO:33具有至少94%同一性; 与SEQ ID NO:33具有至少95%同一性; 与SEQ ID NO:33具有至少96%同一性; 与SEQ ID NO:33具有至少97%同一性; 与SEQ ID NO:33具有至少98%同一性; 与SEQ ID NO:33具有至少99%同一性; 或与SEQ ID NO:33100%同一。

[0698] 32. 实施方案28或实施方案29的核酸反式剪接分子, 其中所述3'转录终止子结构域包含SEQ ID NO:33或SEQ ID NO:66、基本上由其组成或由其组成。

[0699] 33. 前述实施方案中任一项的核酸反式剪接分子, 所述核酸反式剪接分子还包含至少一个在3'操作性连接到所述剪接结构域的终止密码子。

[0700] 34. 实施方案33的核酸反式剪接分子, 所述核酸反式剪接分子在5'至3'方向上包含操作性连接的以下各者:

[0701] (a) 5'非翻译区;

[0702] (b) CDS, 所述CDS基本上由SEQ ID NO:53-59中的任一者组成或由其组成;

[0703] (c) 剪接结构域;

[0704] (d) 至少一个终止密码子;

[0705] (e) 接头结构域, 所述接头结构域基本上由SEQ ID NO:27组成或由其组成;

[0706] (f) 至少一个增效子;

[0707] (g) 翻译增效子;

[0708] (h) 结合结构域, 所述结合结构域基本上由SEQ ID NO:18或SEQ ID NO:20组成或由其组成; 以及

[0709] (i) 3'转录终止子结构域。

[0710] 35. 实施方案34的核酸反式剪接分子, 其中

[0711] (a) 所述5'非翻译区包含SEQ ID NO:13或SEQ ID NO:64、基本上由其组成或由其

组成;

[0712] (b) 所述剪接结构域包含GTAAGT、基本上由其组成或由其组成;

[0713] (c) 所述至少一个终止密码子存在且包含两个终止密码子、基本上由其组成或由其组成,其中所述两个终止密码子中的第一个终止密码子的序列与所述剪接结构域的序列部分重叠;

[0714] (d) 所述至少一个增效子包含三个串联排列的序列、基本上由其组成或由其组成,其中所述三个序列中的每一个包含GGTAAGT或SEQ ID NO:61;

[0715] (e) 所述翻译增效子包含SEQ ID NO:63、基本上由其组成或由其组成;并且;

[0716] (f) 所述3'转录终止子结构域包含SEQ ID NO:66、基本上由其组成或由其组成。

[0717] 36. 一种核酸反式剪接分子,所述核酸反式剪接分子包含:

[0718] (a) 5'非翻译区;

[0719] (b) CDS,所述CDS包含SEQ ID NO:53-59中的任一者、基本上由其组成或由其组成;

[0720] (c) 剪接结构域;

[0721] (d) 任选地,至少一个终止密码子;

[0722] (e) 接头结构域,所述接头结构域包含SEQ ID NO:27、基本上由其组成或由其组成;

[0723] (f) 至少一个增效子;

[0724] (g) 翻译增效子;

[0725] (h) 结合结构域,所述结合结构域包含SEQ ID NO:18或SEQ ID NO:20、基本上由其组成或由其组成;以及

[0726] (i) 3'转录终止子结构域,

[0727] 其中所述5'非翻译区;所述CDS;所述剪接结构域;所述至少一个终止密码子在存在时;所述接头结构域;所述至少一个增效子;所述翻译增效子;所述结合结构域;以及所述3'转录终止子结构域在5'至3'方向上操作性连接。

[0728] 37. 实施方案36的核酸反式剪接分子,其中

[0729] (a) 所述5'非翻译区包含SEQ ID NO:13或SEQ ID NO:64、基本上由其组成或由其组成;

[0730] (b) 所述剪接结构域包含GTAAGT、基本上由其组成或由其组成;

[0731] (c) 所述至少一个终止密码子存在且包含两个终止密码子、基本上由其组成或由其组成,其中所述两个终止密码子中的第一个终止密码子的序列与所述剪接结构域的序列部分重叠;

[0732] (d) 所述至少一个增效子包含三个串联排列的序列、基本上由其组成或由其组成,其中所述三个序列中的每一个包含GGTAAGT或SEQ ID NO:61;

[0733] (e) 所述翻译增效子包含SEQ ID NO:63、基本上由其组成或由其组成;并且

[0734] (f) 所述3'转录终止子结构域包含SEQ ID NO:66、基本上由其组成或由其组成。

[0735] 38. 一种核酸反式剪接分子,所述核酸反式剪接分子在5'至3'方向上包含操作性连接的以下各者:

[0736] (a) 5'非翻译区,所述5'非翻译区包含SEQ ID NO:13或SEQ ID NO:64、基本上由其组成或由其组成;

- [0737] (b) CDS, 所述CDS包含SEQ ID NO:53-59中的任一者、基本上由其组成或由其组成;
- [0738] (c) 剪接结构域, 所述剪接结构域包含GTAAGT、基本上由其组成或由其组成;
- [0739] (d) 两个终止密码子, 其中所述两个终止密码子中的第一个终止密码子的序列与所述剪接结构域的序列部分重叠并且其中所述剪接结构域的序列与所述两个终止密码子组合包含SEQ ID NO:68、基本上由其组成或由其组成;
- [0740] (e) 接头结构域, 所述接头结构域包含SEQ ID NO:27、基本上由其组成或由其组成;
- [0741] (f) 增效子, 所述增效子包含SEQ ID NO:62、基本上由其组成或由其组成;
- [0742] (g) 翻译增效子, 所述翻译增效子包含SEQ ID NO:63、基本上由其组成或由其组成;
- [0743] (h) 结合结构域, 所述结合结构域包含SEQ ID NO:18或SEQ ID NO:20、基本上由其组成或由其组成; 以及
- [0744] (i) 3' 转录终止子结构域, 所述3' 转录终止子结构域包含SEQ ID NO:66、基本上由其组成或由其组成。
- [0745] 39. 实施方案37或实施方案38的核酸反式剪接分子, 其中所述核酸反式剪接分子包含SEQ ID NO:69、SEQ ID NO:78或SEQ ID NO:90、基本上由其组成或由其组成。
- [0746] 40. 实施方案1-20中任一项的核酸反式剪接分子, 其中所述核酸反式剪接分子包含SEQ ID NO:70或SEQ ID NO:71、基本上由其组成或由其组成。
- [0747] 41. 实施方案1-12中任一项的核酸反式剪接分子, 其中所述核酸反式剪接分子包含与SEQ ID NO:40-44、46或48-50中的任一者具有至少80%同一性的核酸序列。
- [0748] 42. 一种核糖核酸反式剪接分子, 所述核糖核酸反式剪接分子从实施方案1-41中任一项的核酸反式剪接分子转录。
- [0749] 43. 实施方案1-41中任一项的核酸反式剪接分子, 其中所述内源ABCA4 mRNA前体包含至少一个与ABCA4相关的视网膜营养不良相关的突变。
- [0750] 44. 实施方案43的核酸反式剪接分子, 其中所述ABCA4相关的视网膜营养不良是隐性遗传的。
- [0751] 45. 实施方案43或44的核酸反式剪接分子, 其中所述ABCA4相关的视网膜营养不良与跨越所述ABCA4基因的外显子1-22的等位基因致病突变和与视锥-视杆细胞营养不良 (例如, 视锥-视杆细胞营养不良-3)、斯特格氏黄斑营养不良 (例如, 斯特格氏病-1)、眼底黄色斑点症、视网膜色素变性-19或年龄相关性黄斑变性-2一致的表型相关。
- [0752] 46. 实施方案1-45中任一项的核酸反式剪接分子, 其中包含所述至少一个突变的所述内源ABCA4mRNA前体在感光细胞或视网膜色素上皮细胞中表达。
- [0753] 47. 一种载体, 所述载体包含实施方案1-46中任一项的核酸反式剪接分子。
- [0754] 48. 实施方案47的载体, 其中所述载体包含在5' 操作性连接到所述CDS的5' 调节结构域。
- [0755] 49. 实施方案48的载体, 其中所述5' 调节结构域操作性连接到5' 非翻译区。
- [0756] 50. 实施方案48-49中任一项的载体, 其中所述5' 调节结构域包含组成型启动子或组织特异性启动子。
- [0757] 51. 实施方案50的载体, 其中所述组成型启动子是CMV启动子。

- [0758] 52.一种前病毒质粒,所述前病毒质粒包含实施方案1-46中任一项的核酸反式剪接分子。
- [0759] 53.一种腺相关病毒(AAV),所述腺相关病毒包含实施方案1-46中任一项的核酸反式剪接分子,其中所述AAV任选地包含5'调节结构域,所述5'调节结构域在5'操作性连接到所述核酸反式剪接分子。
- [0760] 54.实施方案53的AAV,其中所述AAV包含在5'操作性连接到所述CDS的5'调节结构域。
- [0761] 55.实施方案53或54中任一项的AAV,其中所述5'调节结构域操作性连接到5'非翻译区。
- [0762] 56.实施方案53-55中任一项的AAV,其中所述5'调节结构域包含组成型启动子。
- [0763] 57.实施方案56的AAV,其中所述组成型启动子是CMV启动子。
- [0764] 58.实施方案53-57中任一项的AAV,其中所述AAV优先靶向感光细胞和/或视网膜色素上皮细胞。
- [0765] 59.实施方案53-58中任一项的AAV,其中所述AAV是AAV8、AAV5或AAV2。
- [0766] 60.一种组合物,所述组合物包含实施方案1-46中任一项的核酸反式剪接分子、实施方案47-51中任一项的载体、实施方案52的前病毒质粒或实施方案53-59中任一项的AAV。
- [0767] 61.实施方案60的组合物,所述组合物包含药学上可接受的赋形剂。
- [0768] 62.一种在靶细胞中表达生物活性ABCA4以恢复所述靶细胞中ABCA4蛋白的功能水平的方法,所述方法包括将所述靶细胞用实施方案1-46中任一项的核酸反式剪接分子、实施方案47-51中任一项的载体、实施方案52的前病毒质粒、实施方案53-59中任一项的AAV或实施方案60或61的组合物转导。
- [0769] 63.实施方案62的方法,其中所述靶细胞中的至少5%、至少10%或至少15%的靶ABCA4 mRNA被替换。
- [0770] 64.实施方案63的方法,其中所述靶细胞中的至少20%的靶ABCA4 mRNA被替换。
- [0771] 65.实施方案64的方法,其中所述靶细胞中的至少35%的靶ABCA4 mRNA被替换。
- [0772] 66.实施方案62-65中任一项的方法,其中通过表达具有生物功能的全长ABCA4蛋白,使所述靶细胞中ABCA4的功能水平恢复。
- [0773] 67.一种减少受试者中ABCA4相关的脂褐素和/或A2E的积累的方法,所述方法包括将所述受试者中的靶视网膜细胞用实施方案1-46中任一项的核酸反式剪接分子、实施方案47-51中任一项的载体、实施方案52的前病毒质粒、实施方案53-59中任一项的AAV或实施方案60或61的组合物转染或转导。
- [0774] 68.一种纠正受试者的靶细胞中ABCA4外显子序列中的至少一个突变的方法,所述方法包括向所述受试者施用实施方案1-46中任一项的核酸反式剪接分子、实施方案47-51中任一项的载体、实施方案52的前病毒质粒、实施方案53-59中任一项的AAV或实施方案60或61的组合物。
- [0775] 69.一种治疗有需要的受试者的ABCA4相关的视网膜营养不良的方法,所述方法包括以治疗有效量向所述受试者施用实施方案1-46中任一项的核酸反式剪接分子、实施方案47-51中任一项的载体、实施方案52的前病毒质粒、实施方案53-59中任一项的AAV或实施方案60或61的组合物。

[0776] 70. 实施方案67-69中任一项的方法,所述方法包括视网膜下向所述受试者施用实施方案1-46中任一项的核酸反式剪接分子、实施方案47-51中任一项的载体、实施方案52的前病毒质粒、实施方案53-59中任一项的AAV或实施方案60或61的组合物。

[0777] 71. 实施方案67-70中任一项的方法,其中所述受试者是非人灵长类动物或人。

[0778] 72. 实施方案67-71中任一项的方法,其中所述受试者已被诊断为患有ABCA4相关的视网膜营养不良。

[0779] 73. 实施方案1-46中任一项的核酸反式剪接分子、实施方案47-51中任一项的载体、实施方案52的前病毒质粒、实施方案53-59中任一项的AAV或实施方案60或61的组合物,其用于治疗有需要的受试者的ABCA4相关的视网膜营养不良。

[0780] 74. 实施方案1-46中任一项的核酸反式剪接分子、实施方案47-51中任一项的载体、实施方案52的前病毒质粒、实施方案53-59中任一项的AAV或实施方案60或61的组合物,其用于制备供治疗有需要的受试者的ABCA4相关的视网膜营养不良的药物。

[0781] 75. 实施方案67-72中任一项的方法或实施方案73或74中任一项的用途,其中所述ABCA4相关的视网膜营养不良与跨越所述ABCA4基因的外显子1-22的等位基因致病突变和与视锥-视杆细胞营养不良、斯特格氏黄斑营养不良或眼底黄色斑点症一致的表型相关。

[0782] 76. 一种核酸反式剪接分子,所述核酸反式剪接分子包含:

[0783] (a) 接头结构域,所述接头结构域包含SEQ ID NO:27或与SEQ ID NO:27具有至少90%同一性的序列、基本上由其组成或由其组成;以及

[0784] (b) 任选地,增效子,所述增效子包含SEQ ID NO:62、基本上由其组成或由其组成;以及

[0785] (c) 任选地,翻译增效子,所述翻译增效子包含SEQ ID NO:63、基本上由其组成或由其组成,其中所述接头结构域;所述增效子在存在时;以及所述翻译增效子在存在时,在5'至3'方向上操作性连接。

[0786] 77. 实施方案76的核酸反式剪接分子,所述核酸反式剪接分子还包含结合结构域,所述结合结构域退火到内源ABCA4 mRNA前体内的结合位点,

[0787] 其中所述接头结构域;所述增效子在存在时;所述翻译增效子在存在时;以及所述结合结构域,在5'至3'方向上操作性连接。

[0788] 78. 实施方案76或实施方案77的核酸反式剪接分子,所述核酸反式剪接分子还包括包含ABCA4外显子的CDS,

[0789] 其中所述CDS;所述接头结构域;所述增效子在存在时;所述翻译增效子在存在时;以及所述结合结构域在存在时,在5'至3'方向上操作性连接。

[0790] 79. 一种5'核酸反式剪接分子,所述5'核酸反式剪接分子包含:

[0791] (a) 增效子,所述增效子包含GGTAAGT或SEQ ID NO:61或与GGTAAGT或SEQ ID NO:61具有至少90%同一性、至少91%同一性、至少92%同一性、至少93%同一性、至少94%同一性、至少95%同一性、至少96%同一性、至少97%同一性、至少98%同一性或至少99%同一性的序列中的至少一者、基本上由其组成或由其组成;以及

[0792] (b) 任选地,接头结构域,所述接头结构域包含SEQ ID NO:27或与SEQ ID NO:27具有至少90%同一性的序列、基本上由其组成或由其组成;

[0793] (c) 任选地,翻译增效子,所述翻译增效子包含SEQ ID NO:63、基本上由其组成或

由其组成,

[0794] 其中所述接头结构域在存在时;所述增效子;以及所述翻译增效子在存在时,在5'至3'方向上操作性连接。

[0795] 80.一种5'核酸反式剪接分子,所述5'核酸反式剪接分子包含:

[0796] (a) 增效子,所述增效子包含SEQ ID NO:62或与SEQ ID NO:62具有至少90%同一性的序列、基本上由其组成或由其组成;以及

[0797] (b) 任选地,接头结构域,所述接头结构域包含SEQ ID NO:27或与SEQ ID NO 27具有至少90%同一性的序列、基本上由其组成或由其组成;

[0798] (c) 任选地,翻译增效子,所述翻译增效子包含SEQ ID NO:63、基本上由其组成或由其组成,

[0799] 其中所述接头结构域在存在时;所述增效子;以及所述翻译增效子在存在时,在5'至3'方向上操作性连接。

[0800] 81.实施方案79或实施方案80的核酸反式剪接分子,所述核酸反式剪接分子还包含结合结构域,所述结合结构域退火到内源ABCA4 mRNA前体内的结合位点,

[0801] 其中所述接头结构域在存在时;所述增效子;所述翻译增效子在存在时;以及所述结合结构域,在5'至3'方向上操作性连接。

[0802] 82.实施方案79-81中任一项的核酸反式剪接分子,所述核酸反式剪接分子还包括包含ABCA4外显子的CDS,

[0803] 其中所述CDS;所述接头结构域在存在时;所述增效子;所述翻译增效子在存在时;以及所述结合结构域在存在时,在5'至3'方向上操作性连接。

[0804] 83.一种核酸反式剪接分子,所述核酸反式剪接分子包含:

[0805] (a) 翻译增效子,所述翻译增效子包含SEQ ID NO:63或与SEQ ID NO:63具有至少90%同一性的序列、基本上由其组成或由其组成;

[0806] (b) 任选地,增效子,所述增效子包含SEQ ID NO:62、基本上由其组成或由其组成;以及

[0807] (c) 任选地,接头结构域,所述接头结构域包含SEQ ID NO:27或与SEQ ID NO:27具有至少90%同一性的序列、基本上由其组成或由其组成,其中所述接头结构域在存在时;所述增效子在存在时;以及所述翻译增效子,在5'至3'方向上操作性连接。

[0808] 84.实施方案83的核酸反式剪接分子,所述核酸反式剪接分子还包含结合结构域,所述结合结构域退火到内源ABCA4 mRNA前体内的结合位点,

[0809] 其中所述接头结构域在存在时;所述增效子在存在时;所述翻译增效子;以及所述结合结构域,在5'至3'方向上操作性连接。

[0810] 85.实施方案83或实施方案84的核酸反式剪接分子,所述核酸反式剪接分子还包括包含ABCA4外显子的CDS,

[0811] 其中所述CDS;所述接头结构域在存在时;所述增效子在存在时;所述翻译增效子;以及所述结合结构域在存在时,在5'至3'方向上操作性连接。

[0812] 86.一种核酸反式剪接分子,所述核酸反式剪接分子包含:

[0813] 包含接头结构域和增效子的盒,其中所述盒包含SEQ ID NO:72、基本上由其组成或由其组成。

[0814] 87.一种核酸反式剪接分子,所述核酸反式剪接分子包含:

[0815] 包含接头结构域、增效子和翻译增效子的盒,其中所述盒包含SEQ ID NO:73、基本上由其组成或由其组成。

[0816] 88.一种核酸反式剪接分子,所述核酸反式剪接分子包含:

[0817] (a) 3'转录终止子结构域,所述3'转录终止子结构域包含SEQ ID NO:66或与SEQ ID NO:66具有至少95%同一性的序列、基本上由其组成或由其组成。

[0818] 89.实施方案88的核酸反式剪接分子,其中所述3'转录终止子结构域包含如下核酸序列,所述核酸序列与SEQ ID NO:66具有至少96%同一性;与SEQ ID NO:66具有至少97%同一性;与SEQ ID NO:66具有至少98%同一性;与SEQ ID NO:66具有至少99%同一性;或与SEQ ID NO:66100%同一。

[0819] 90.实施方案88或实施方案89的核酸反式剪接分子,其中所述3'转录终止子结构域包含SEQ ID NO:66、基本上由其组成或由其组成。

[0820] 91.权利要求89或权利要求90的核酸反式剪接分子,所述核酸反式剪接分子还包括包含ABCA4外显子的CDS,

[0821] 其中所述CDS;所述盒;以及所述结合结构域在存在时,在5'至3'方向上操作性连接。

[0822] 92.一种核酸反式剪接分子,所述核酸反式剪接分子包含:

[0823] (a) 3'转录终止子结构域,所述3'转录终止子结构域包含SEQ ID NO:66或与SEQ ID NO:66具有至少95%同一性的序列、基本上由其组成或由其组成。

[0824] 93.权利要求92的核酸反式剪接分子,其中所述3'转录终止子结构域包含如下核酸序列,所述核酸序列与SEQ ID NO:66具有至少96%同一性;与SEQ ID NO:66具有至少97%同一性;与SEQ ID NO:66具有至少98%同一性;与SEQ ID NO:66具有至少99%同一性;或与SEQ ID NO:66100%同一。

[0825] 94.权利要求92或权利要求93的核酸反式剪接分子,其中所述3'转录终止子结构域包含SEQ ID NO:66、基本上由其组成或由其组成。

[0826] 编号实施方案III

[0827] 1.一种核酸反式剪接分子,所述核酸反式剪接分子包含:

[0828] (a) 编码结构域序列(CDS),所述编码结构域序列包含相对于野生型ABCA4序列在隐蔽剪接位点处具有至少一个核苷酸变异的ABCA4外显子;以及

[0829] (b) 结合结构域,所述结合结构域与内源ABCA4 mRNA前体内的结合位点互补。

[0830] 2.实施方案1的核酸反式剪接分子,其中所述隐蔽剪接位点是表3中列出的隐蔽剪接位点。

[0831] 3.实施方案1或2的核酸反式剪接分子,其中所述至少一个核苷酸变异是减少或消除所述隐蔽剪接位点处的剪接的同义突变。

[0832] 4.实施方案1至3中任一项的核酸反式剪接分子,其中所述CDS包含SEQ ID NO:92、94、96、98、100、102、104、106、108、110、112、114、116、118、120、122、124、127、129、131、673、674、679、682、685、693、695、698、701、708、721、722、737、758、763、764、765、775、785、789、791、798、799、792、818、820、822、825、831、837、854、859、873、881、886、888、891、903、904、910、918、924、932、938、940、945、946、950、951、952、968、976、992、993、995、997、998、999、

1003、1004、1010、1015、1026、1031、1040、1047、1070、1117或1169中的一者或多者。

[0833] 5. 实施方案1至4中任一项的核酸反式剪接分子,其中所述CDS包含具有所述至少一个核苷酸变异的ABCA4的外显子1-22。

[0834] 6. 实施方案1至5中任一项的核酸反式剪接分子,其中所述CDS包含SEQ ID NO:56-59中的任一者。

[0835] 7. 实施方案1至6中任一项的核酸反式剪接分子,其中所述CDS包含SEQ ID NO:56。

[0836] 8. 实施方案1至7中任一项的核酸反式剪接分子,所述核酸反式剪接分子还包括包含与SEQ ID NO:27具有至少85%同一性的序列的接头结构域,其中所述CDS、所述接头结构域和所述结合结构域在5'至3'方向上操作性连接。

[0837] 9. 实施方案1至8中任一项的核酸反式剪接分子,其中所述结合结构域退火到所述内源ABCA4mRNA前体的内含子22内的序列。

[0838] 10. 实施方案1至9中任一项的核酸反式剪接分子,其中所述结合结构域包含SEQ ID NO:18或20。

[0839] 11. 实施方案1至10中任一项的核酸反式剪接分子,所述核酸反式剪接分子还包含增效子,所述增效子包括包含GGTAAGT或SEQ ID NO:61的至少一个序列,其中所述增效子在3'可操作地连接到所述接头结构域。

[0840] 12. 实施方案11的核酸反式剪接分子,其中所述增效子包含三个串联排列的序列,其中所述三个序列中的每一个包含GGTAAGT或SEQ ID NO:61。

[0841] 13. 实施方案12的核酸反式剪接分子,其中所述增效子包含三个串联排列并可操作地连接的序列,其中所述三个序列一起包含SEQ ID NO:62。

[0842] 14. 实施方案1至13中任一项的核酸反式剪接分子,所述核酸反式剪接分子还包含翻译增效子,其中所述翻译增效子包含与SEQ ID NO:63具有至少85%同一性的序列,在3'可操作地连接到所述增效子。

[0843] 15. 实施方案1至14中任一项的核酸反式剪接分子,所述核酸反式剪接分子还包含剪接结构域,所述剪接结构域包含序列GTAAGT。

[0844] 16. 实施方案1至15中任一项的核酸反式剪接分子,所述核酸反式剪接分子还包含5'非翻译区。

[0845] 17. 实施方案1至16中任一项的核酸反式剪接分子,所述核酸反式剪接分子还包含3'转录终止子结构域,所述3'转录终止子结构域包含与SEQ ID NO:33或SEQ ID NO:66具有至少85%同一性的MALAT1序列。

[0846] 18. 实施方案1至17中任一项的核酸反式剪接分子,所述核酸反式剪接分子在5'至3'方向上包含操作性连接的以下各者:

[0847] (a) 5'非翻译区,所述5'非翻译区包含SEQ ID NO:13或SEQ ID NO:64;

[0848] (b) CDS,所述CDS包含SEQ ID NO:53-59中的任一者;

[0849] (c) 剪接结构域,所述剪接结构域包含序列GTAAGT;

[0850] (d) 两个终止密码子,其中所述两个终止密码子中的第一个终止密码子的序列与所述剪接结构域的序列部分重叠并且其中所述剪接结构域的序列与所述两个终止密码子组合包含SEQ ID NO:68(GTAAGTAGTGA);

[0851] (e) 接头结构域,所述接头结构域包含SEQ ID NO:27;

- [0852] (f) 增效子,所述增效子包含SEQ ID NO:62;
- [0853] (g) 翻译增效子,所述翻译增效子包含SEQ ID NO:63;
- [0854] (h) 结合结构域,所述结合结构域包含SEQ ID NO:18或SEQ ID NO:20;以及
- [0855] (i) 3'转录终止子结构域,所述3'转录终止子结构域包含SEQ ID NO:66。
- [0856] 19. 实施方案1至18中任一项的核酸反式剪接分子,所述核酸反式剪接分子包含SEQ ID NO:69、SEQ ID NO:78或SEQ ID NO:90
- [0857] 20. 一种核糖核酸反式剪接分子,所述核糖核酸反式剪接分子从实施方案19的核酸反式剪接分子转录。
- [0858] 21. 一种载体或腺相关病毒(AAV),所述腺相关病毒包含实施方案1至19中任一项的核酸反式剪接分子。
- [0859] 22. 实施方案21的载体,其中所述载体包含SEQ ID NO:82。
- [0860] 23. 实施方案21或22的AAV,其中所述AAV优先靶向感光细胞和/或视网膜色素上皮细胞。
- [0861] 24. 实施方案23的AAV,其中所述AAV是AAV8、AAV5或AAV2。
- [0862] 25. 一种药物组合物,所述药物组合物包含实施方案21至24中任一项的载体或AAV和药学上可接受的赋形剂。
- [0863] 26. 一种治疗有需要的受试者的ABCA4相关的视网膜营养不良的方法,所述方法包括以治疗有效量向所述受试者施用实施方案25的组合物。
- [0864] 27. 一种在靶细胞中表达生物活性ABCA4的方法,所述方法包括将所述靶细胞用核酸反式剪接分子转染或转导,所述核酸反式剪接分子包含:
- [0865] (a) 编码结构域序列(CDS),所述编码结构域序列包含相对于野生型ABCA4序列在隐蔽剪接位点处具有至少一个核苷酸变异的ABCA4外显子;以及
- [0866] (b) 结合结构域,所述结合结构域与内源ABCA4 mRNA前体内的结合位点互补。
- [0867] 28. 实施方案27的方法,其中所述CDS包含SEQ ID NO:92、94、96、98、100、102、104、106、108、110、112、114、116、118、120、122、124、127、129、131、673、674、679、682、685、693、695、698、701、708、721、722、737、758、763、764、765、775、785、789、791、798、799、792、818、820、822、825、831、837、854、859、873、881、886、888、891、903、904、910、918、924、932、938、940、945、946、950、951、952、968、976、992、993、995、997、998、999、1003、1004、1010、1015、1026、1031、1040、1047、1070、1117或1169中的一者或多者。
- [0868] 29. 一种治疗有需要的受试者的ABCA4相关的视网膜营养不良的方法,所述方法包括向所述受试者施用核酸反式剪接分子,所述核酸反式剪接分子包含:
- [0869] (a) 编码结构域序列(CDS),所述编码结构域序列包含相对于野生型ABCA4序列在隐蔽剪接位点处具有至少一个核苷酸变异的ABCA4外显子;以及
- [0870] (b) 结合结构域,所述结合结构域与内源ABCA4 mRNA前体内的结合位点互补。
- [0871] 30. 实施方案29的方法,其中所述ABCA4相关的视网膜营养不良与ABCA4基因的外显子1-22内的一个或多个等位基因致病突变和与视锥-视杆细胞营养不良(例如,视锥-视杆细胞营养不良-3)、斯特格氏黄斑营养不良(例如,斯特格氏病-1)、眼底黄色斑点症、视网膜色素变性-19或年龄相关性黄斑变性-2一致的表型相关。
- [0872] 其它实施方案

[0873] 本说明书中提及的所有出版物、专利和专利申请以引用的方式并入本文中,就如每个单独的出版物或专利申请具体地和单独地指出以引用的方式并入的程度一样。

[0874] 尽管已经结合具体的实施方案描述本发明,但应理解本发明能够具有另外的修改,并且本申请旨在涵盖大体上符合本发明原理的、以及虽然不属于本发明所公开内容范围但属于本发明所属领域的公知技术或常用的技术手段并可以应用于上文中阐述和权利要求的范围所列出的必要特征的任何变型、用途或者变更。

[0875] 其它实施方案在权利要求书内。

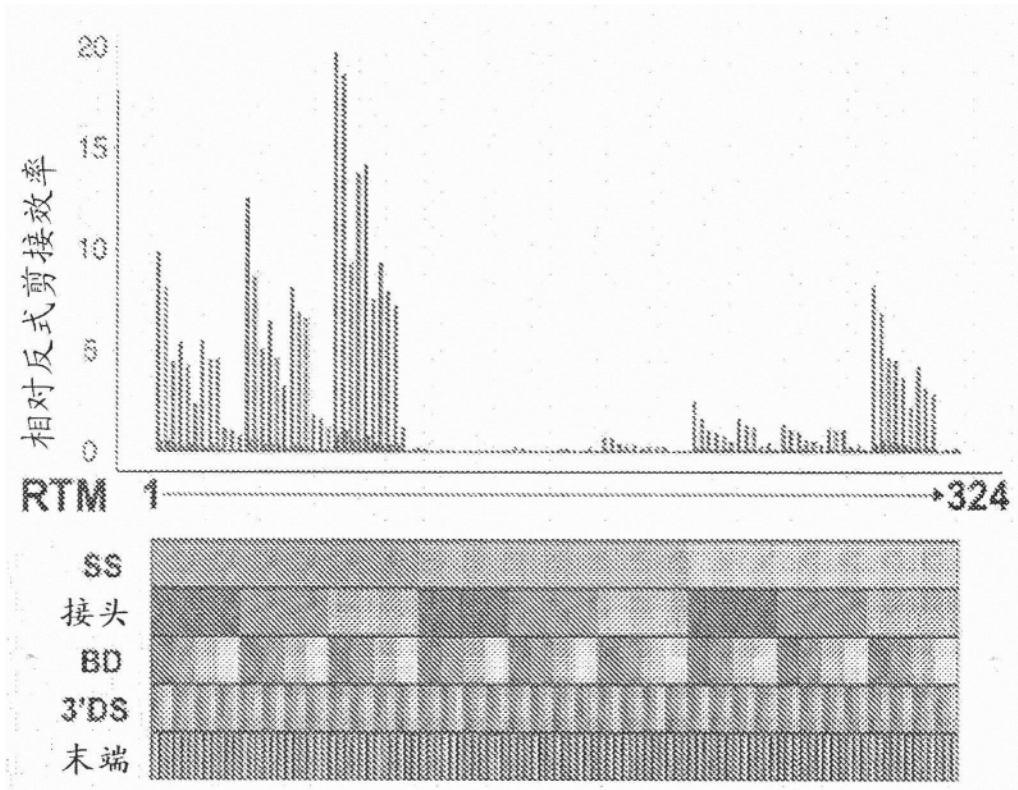


图1

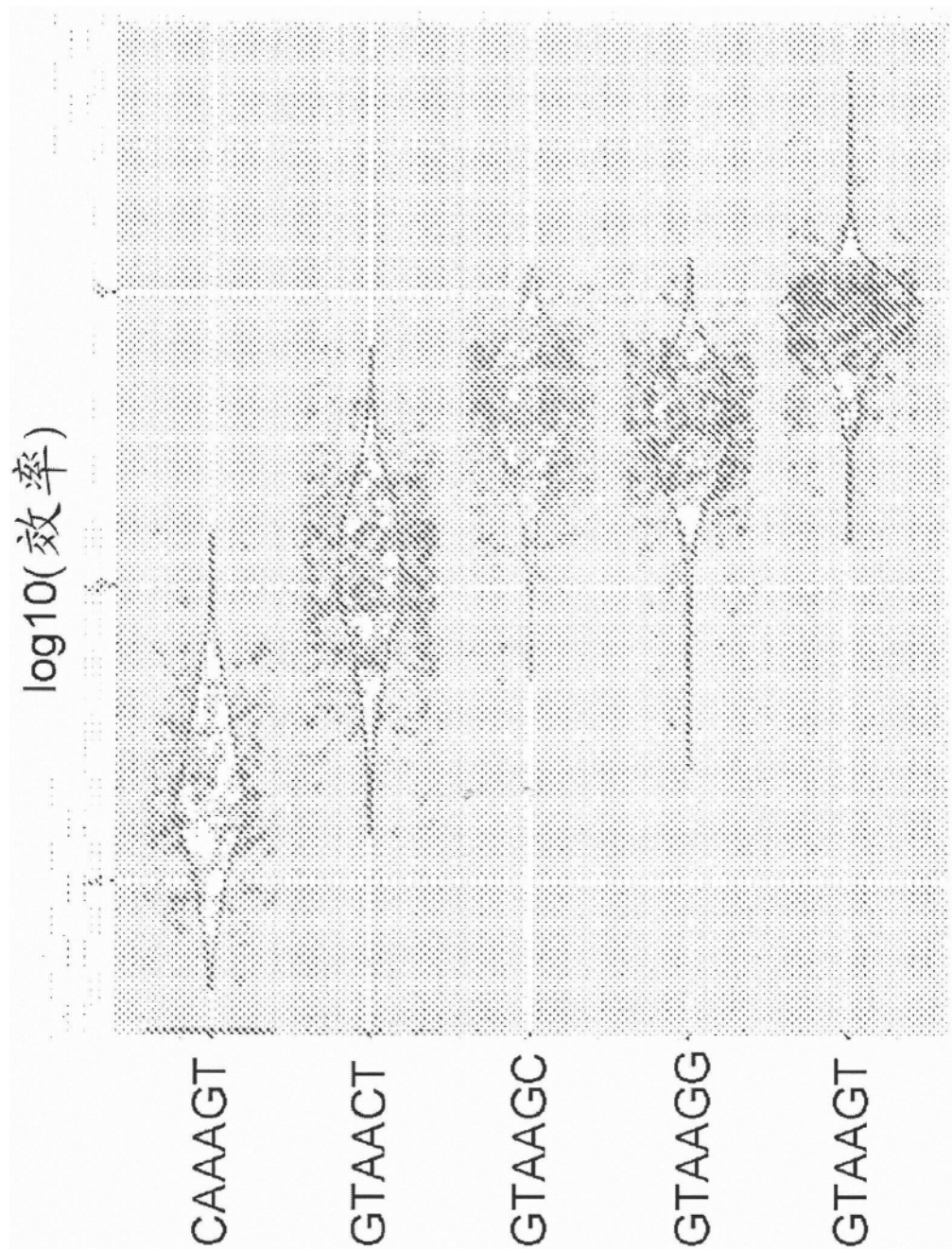


图2

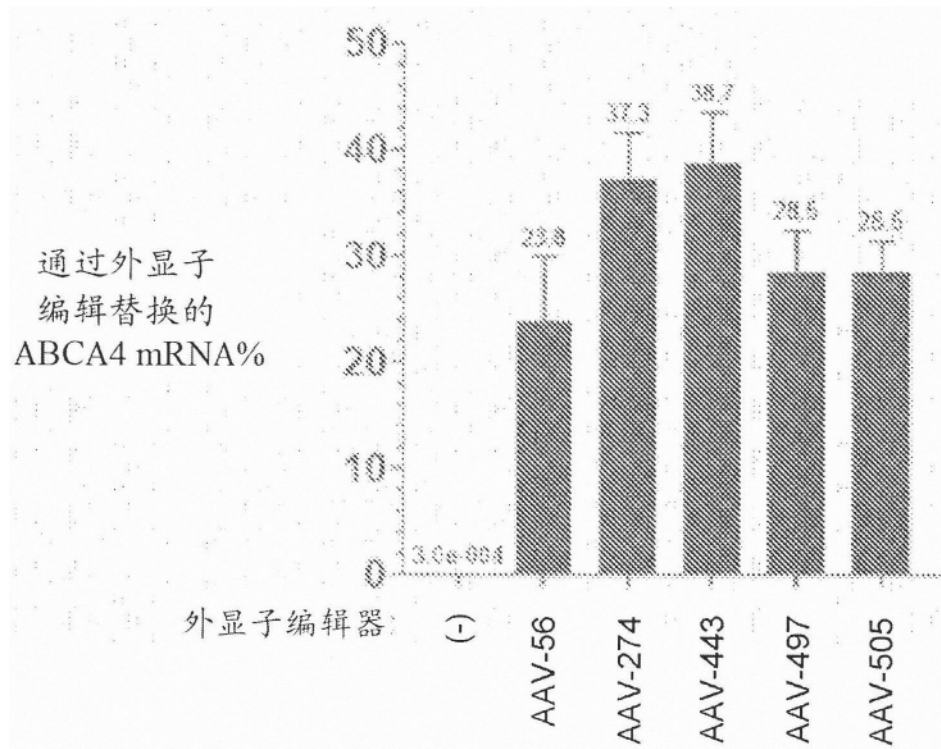


图3A

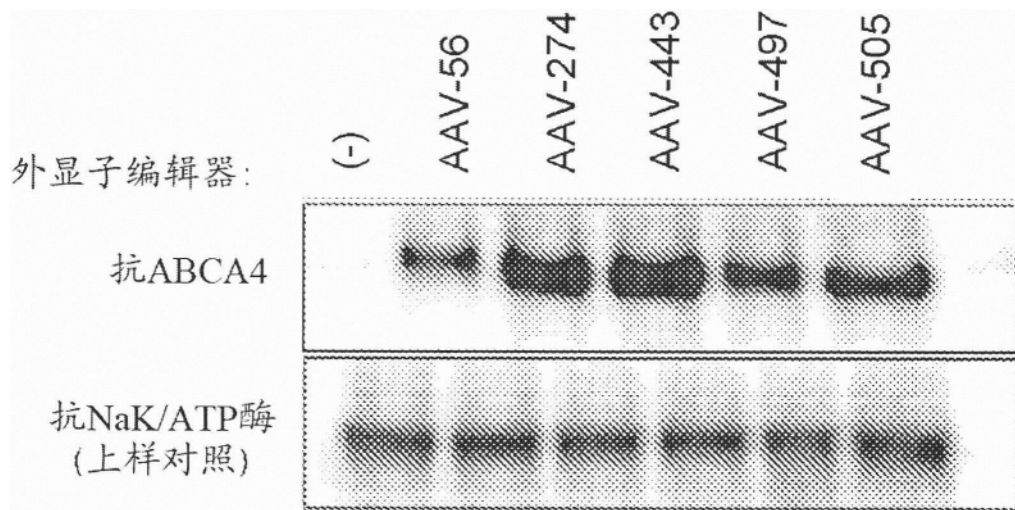


图3B

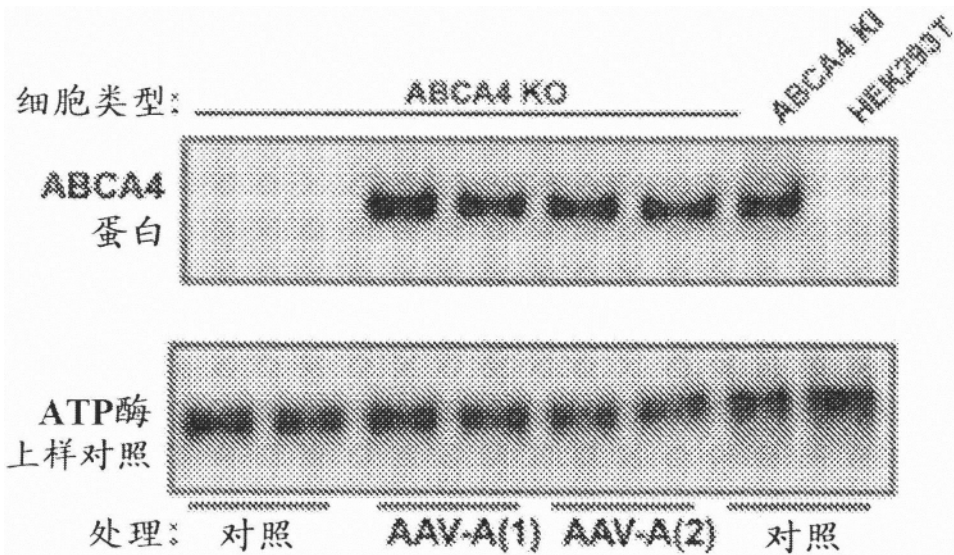


图3C

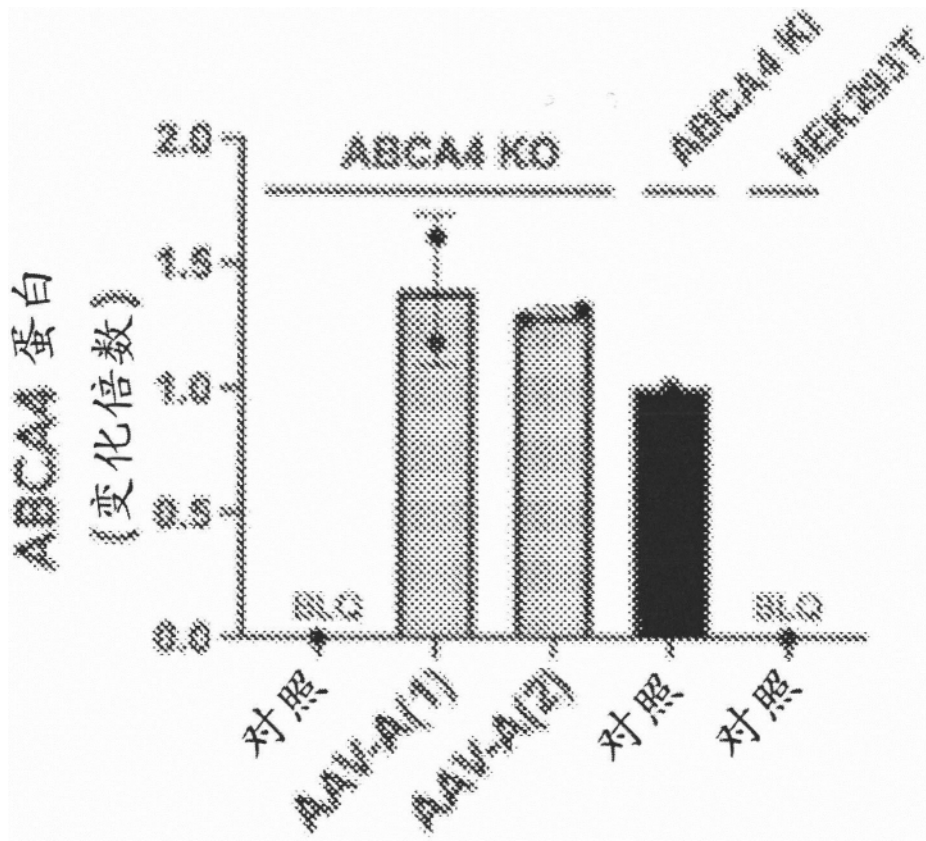


图3D

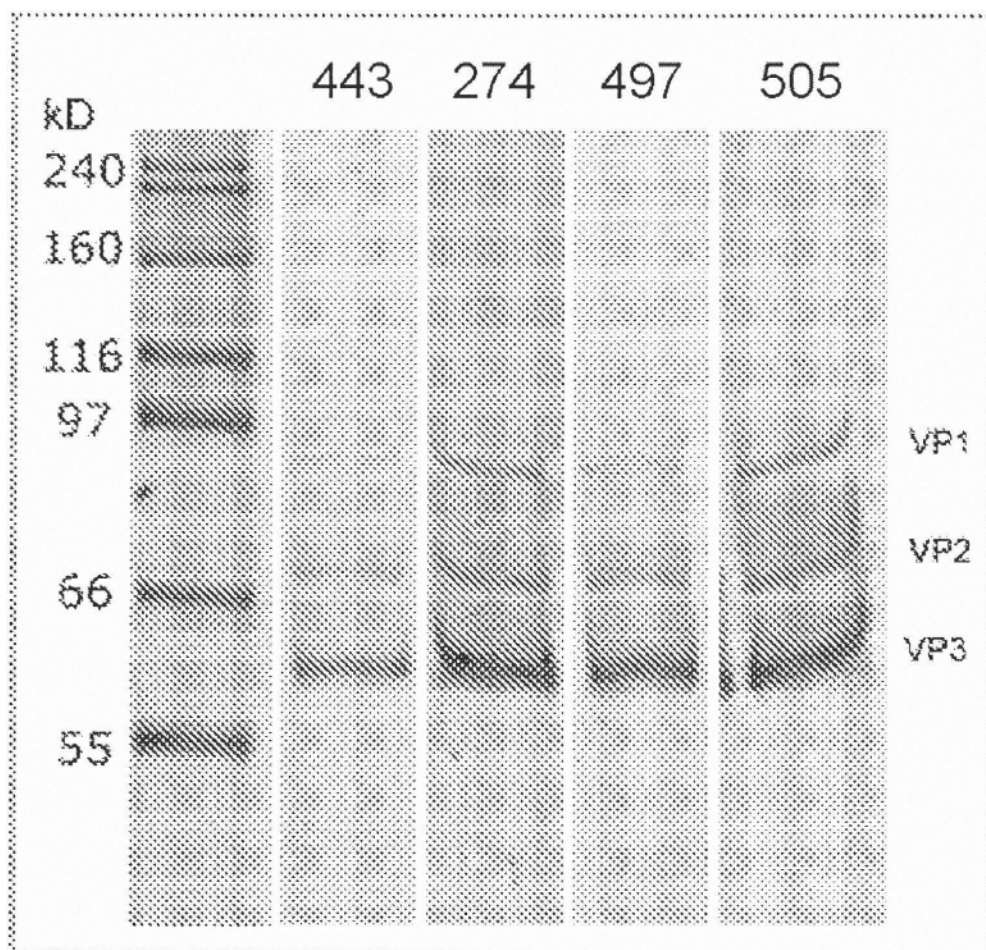


图4

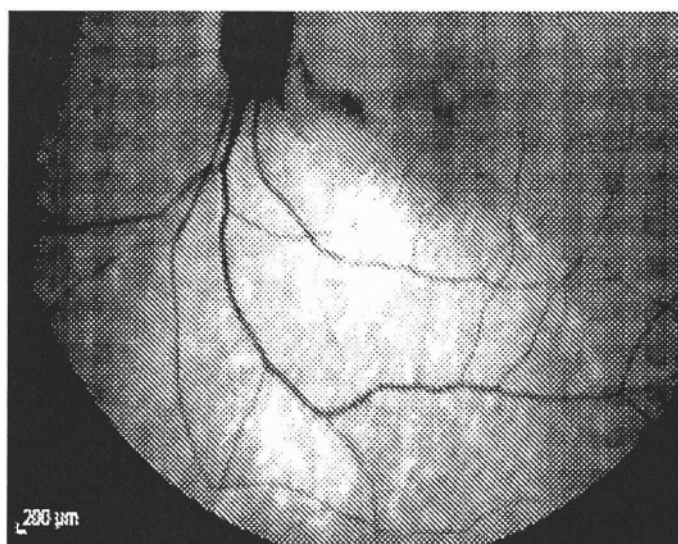


图5A

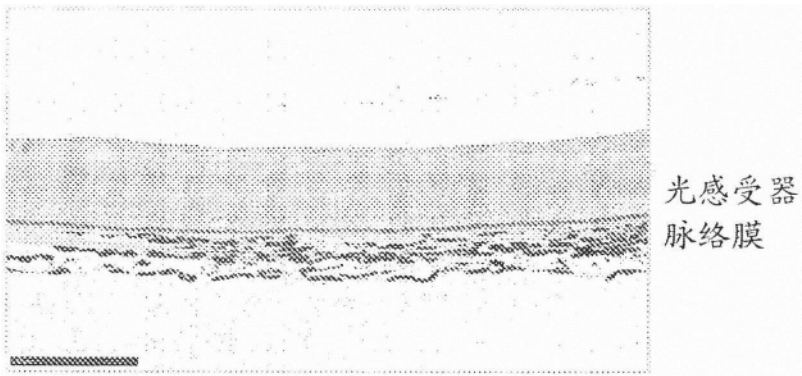


图5B

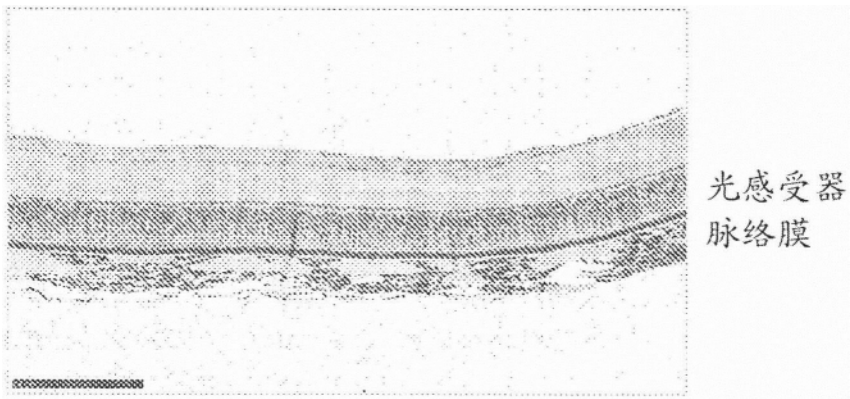


图5C

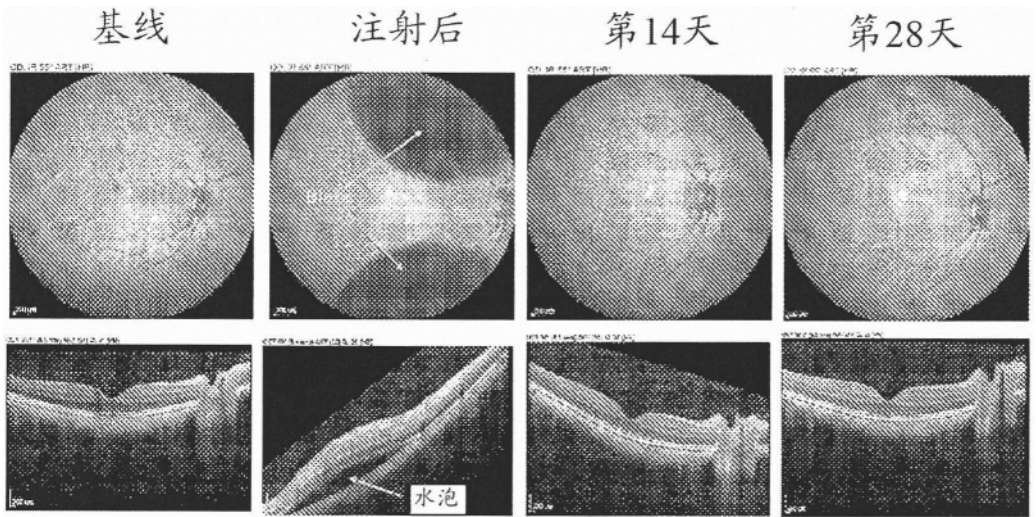


图6

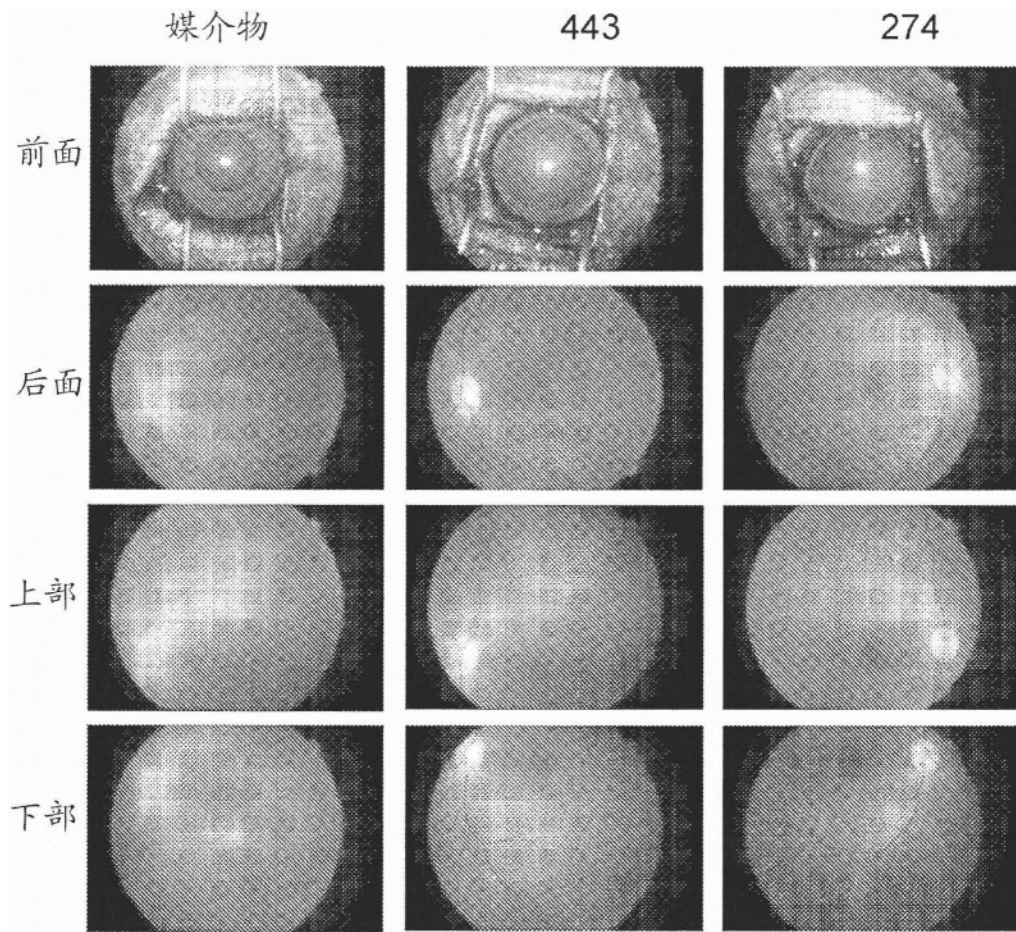


图7

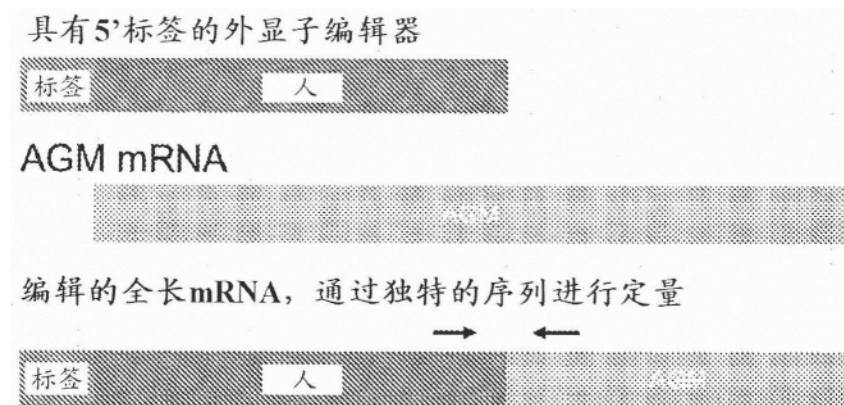


图8A

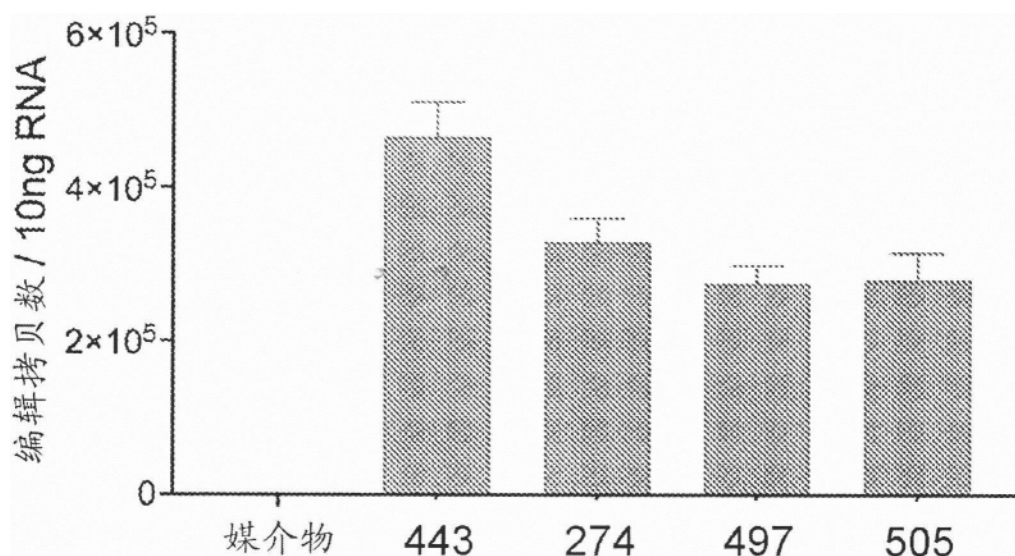


图8B

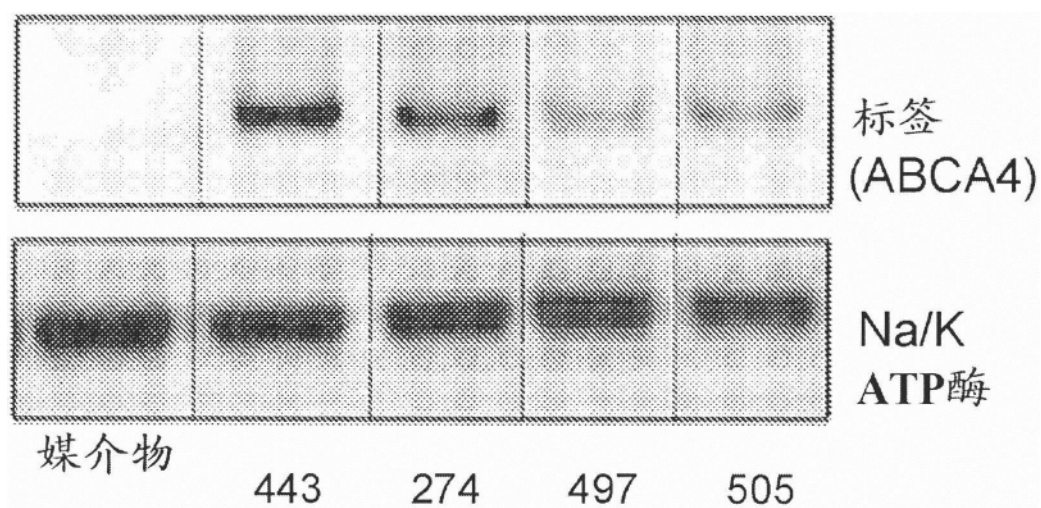


图9

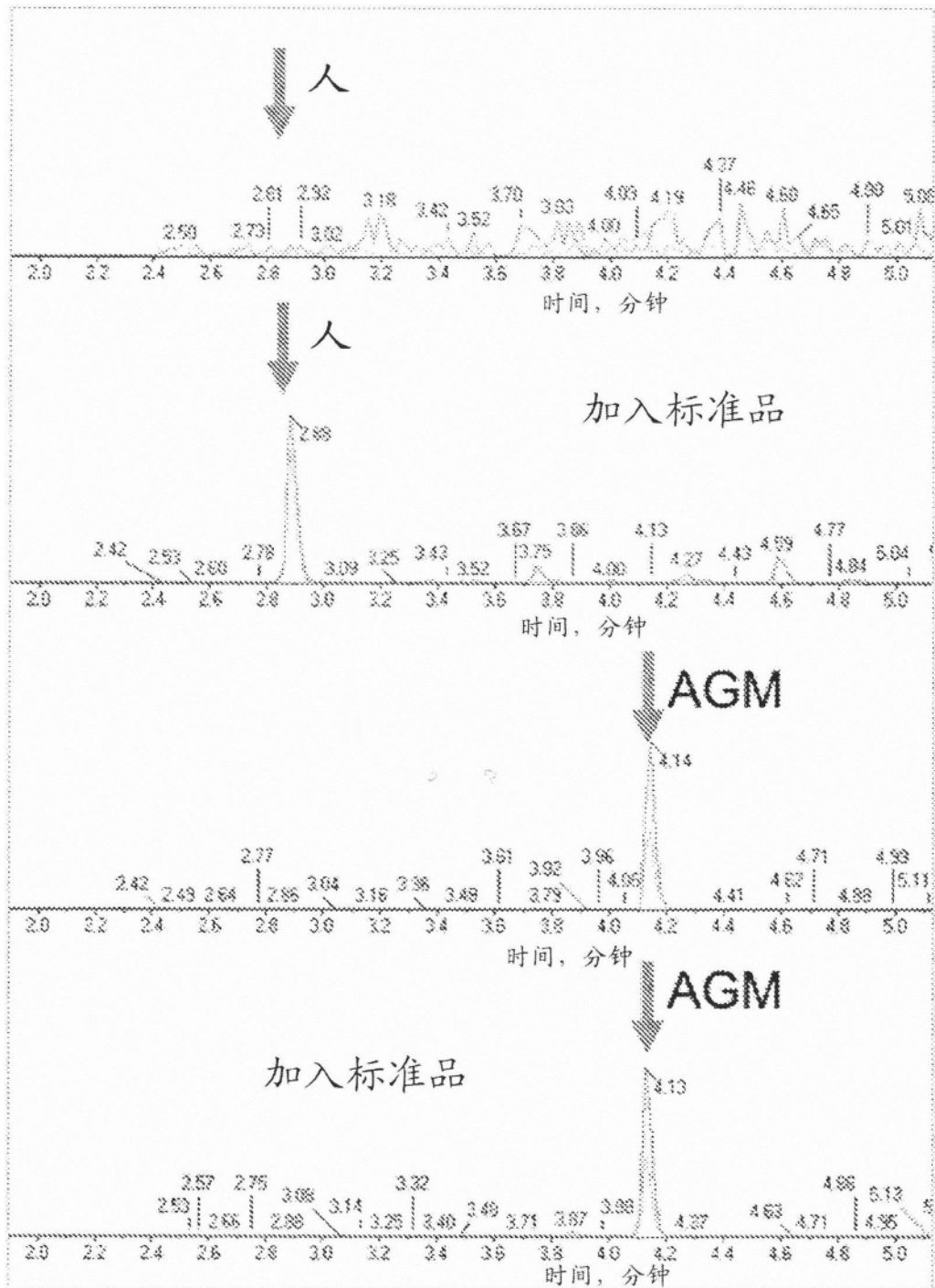


图10A

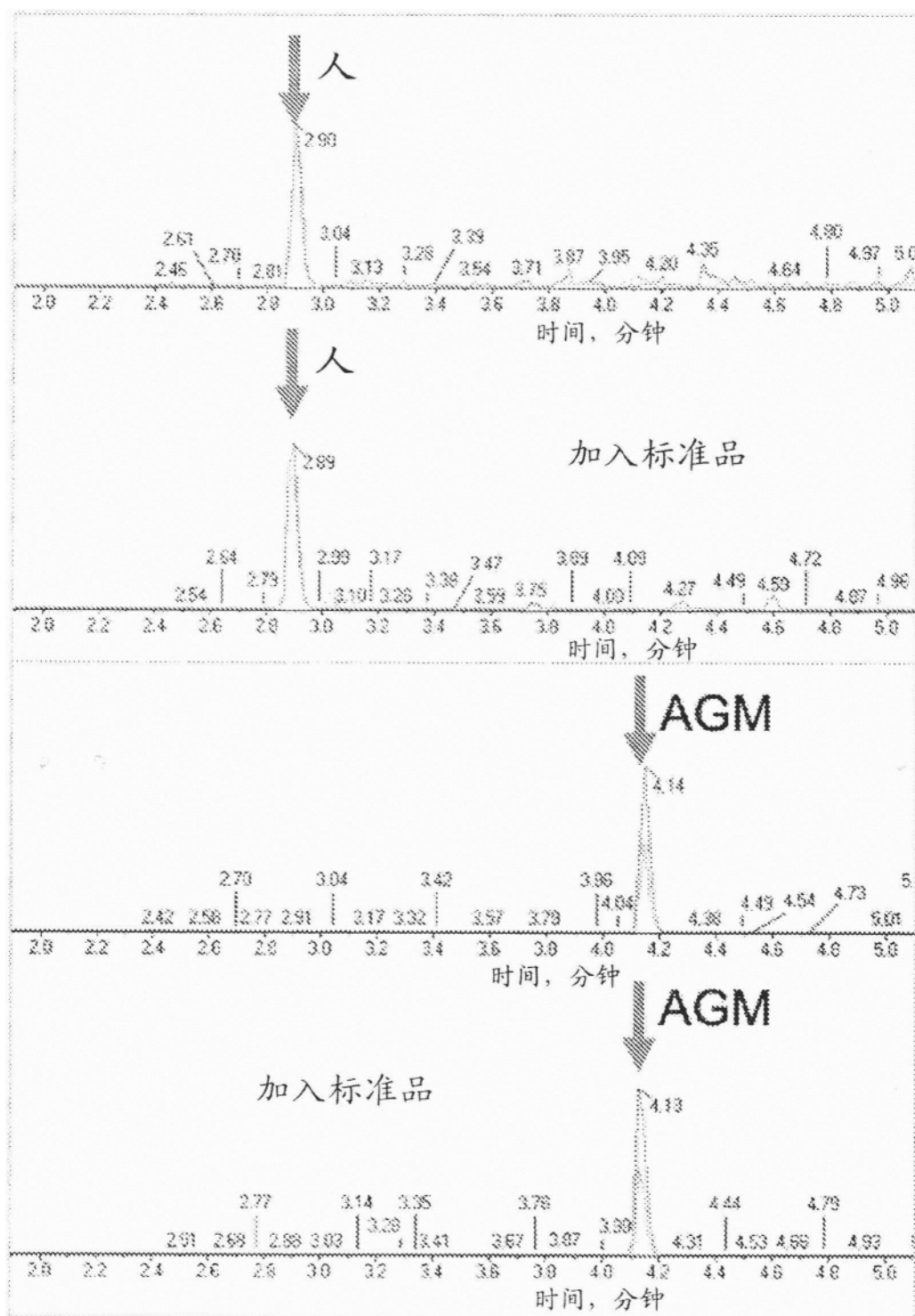


图10B

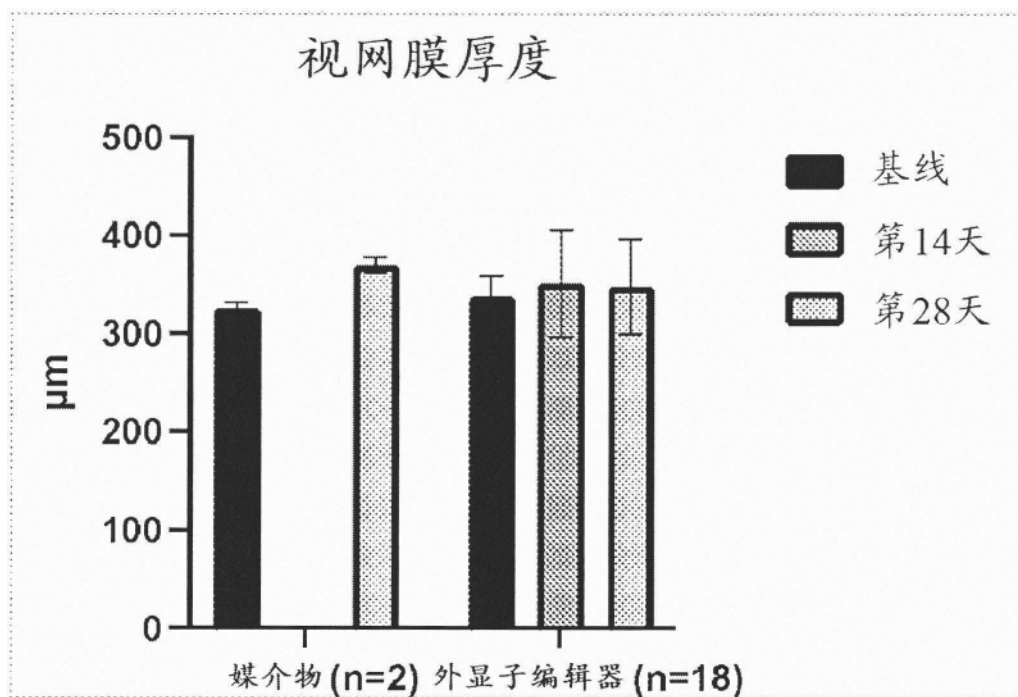


图11A

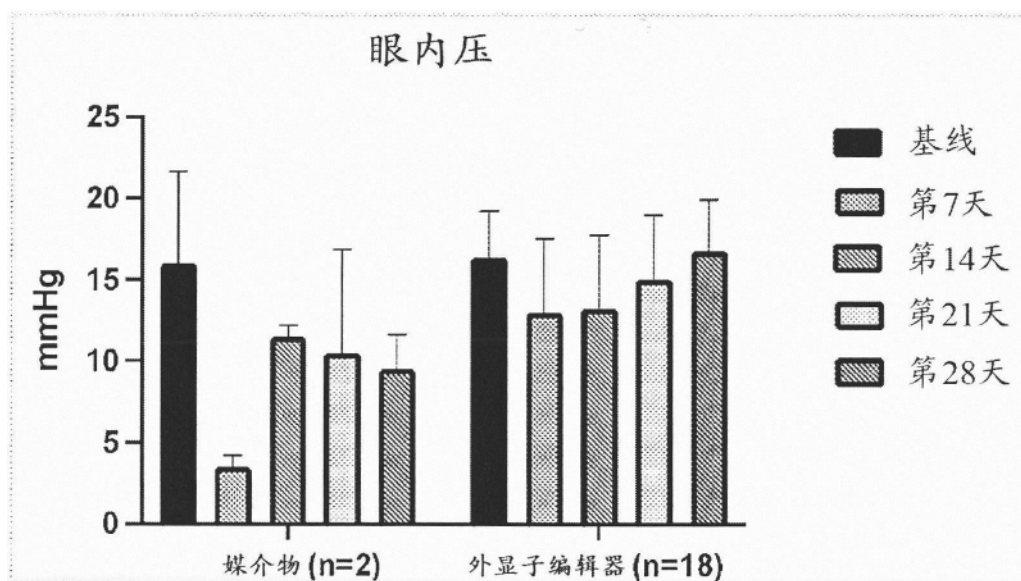


图11B

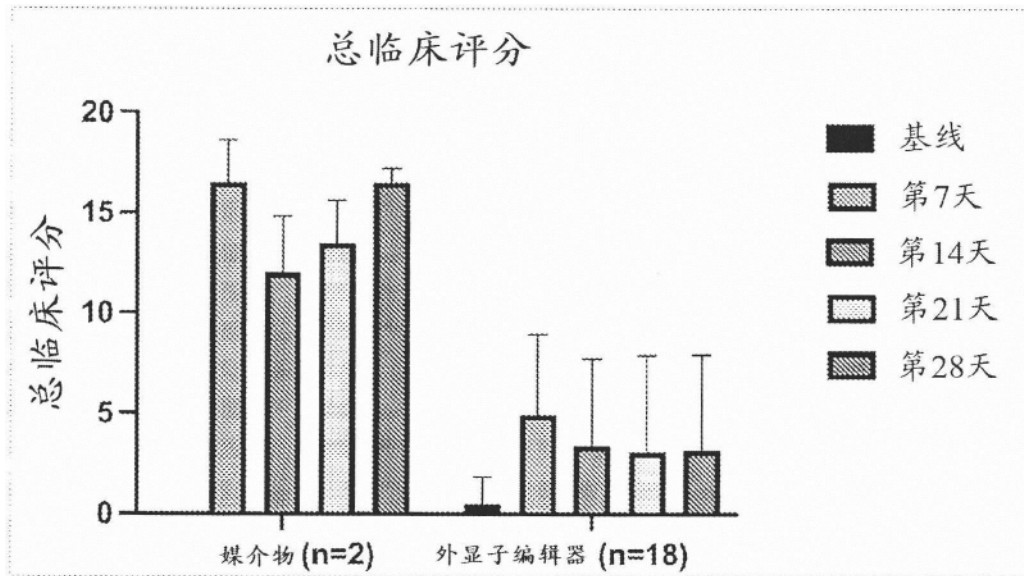


图11C

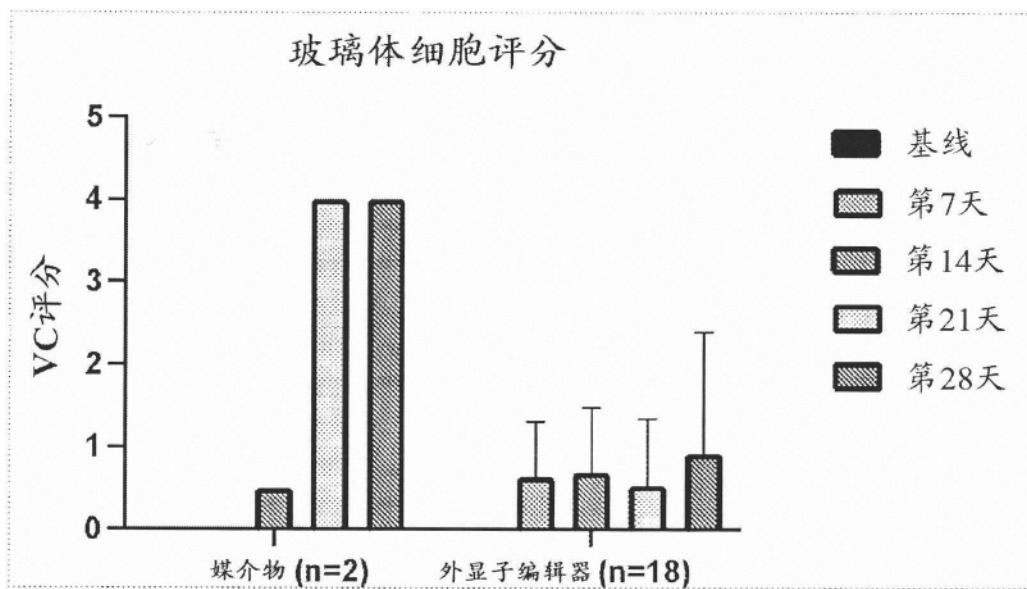


图11D

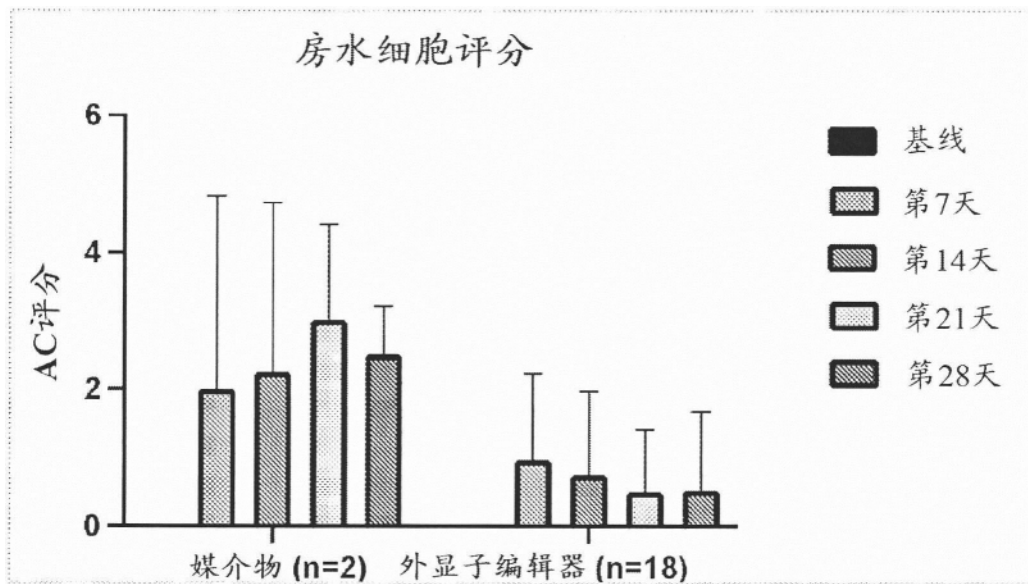


图11E

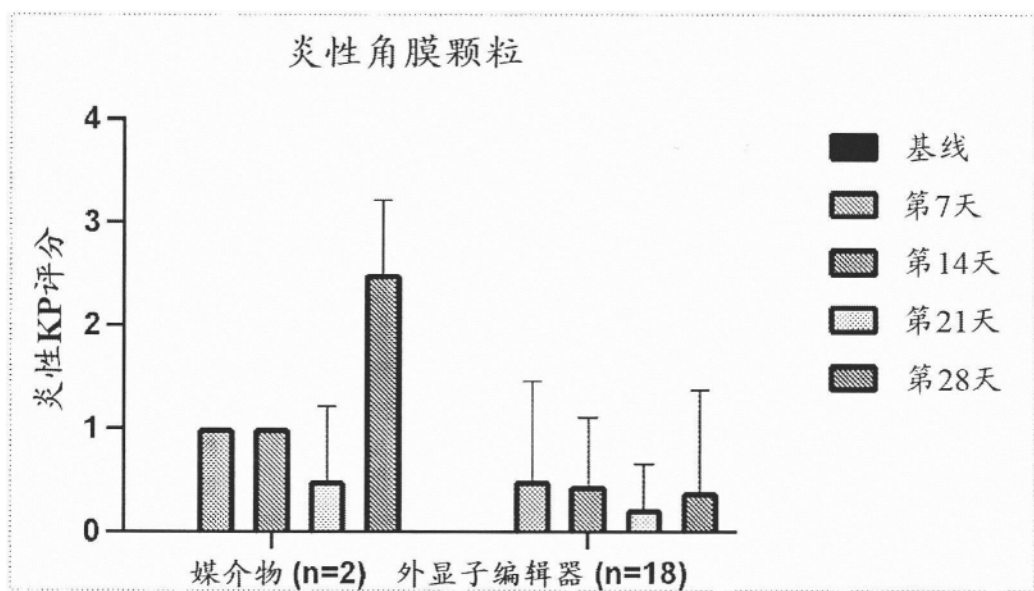


图11F

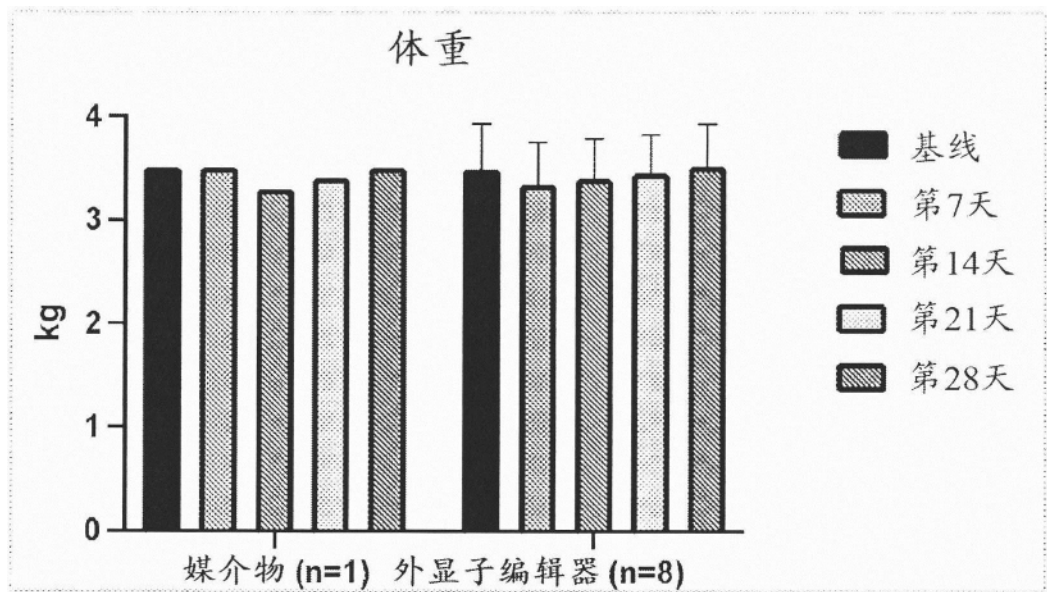


图12A

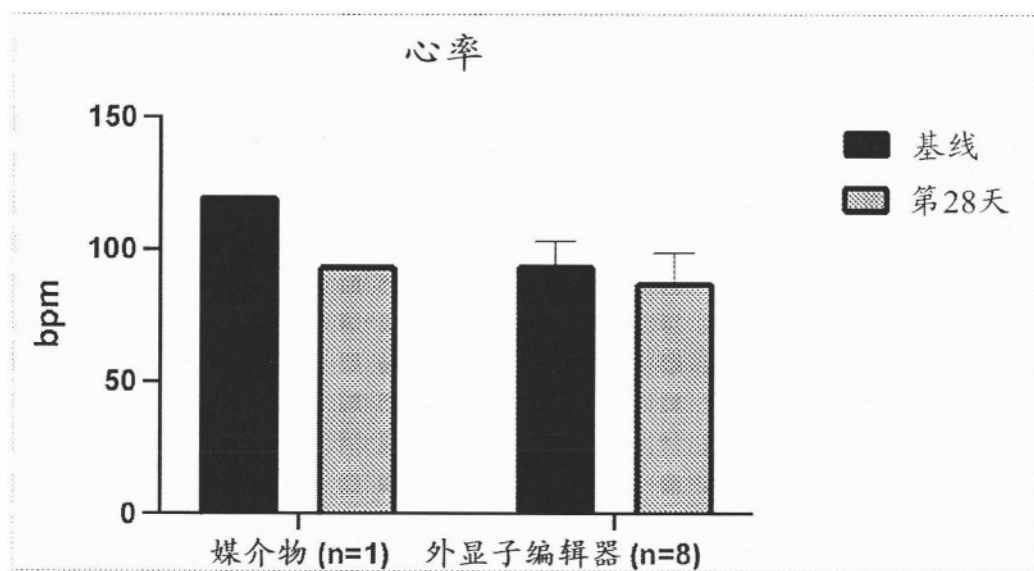


图12B

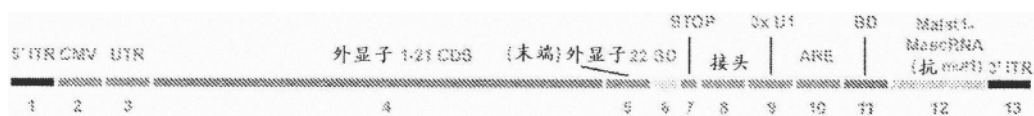


图13

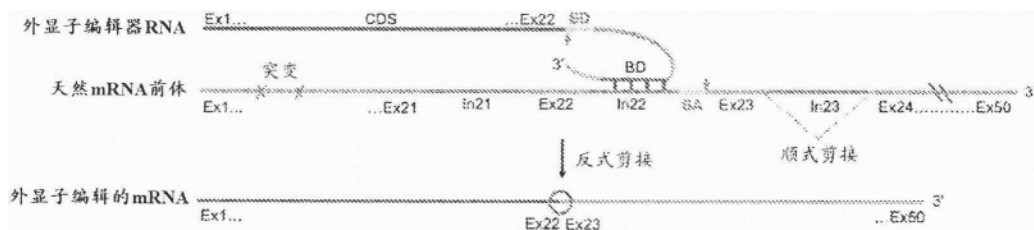


图14A

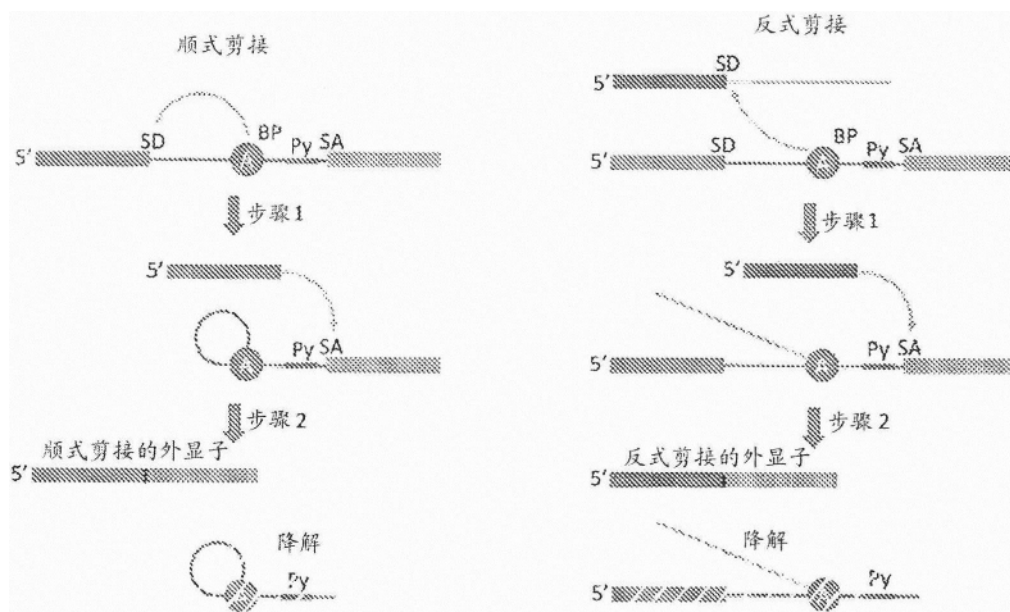


图14B

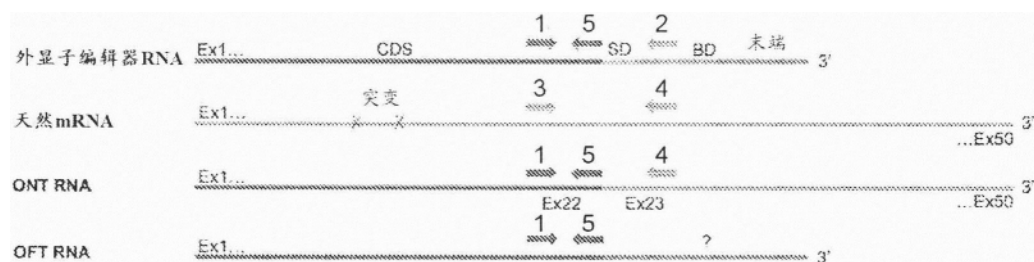


图15

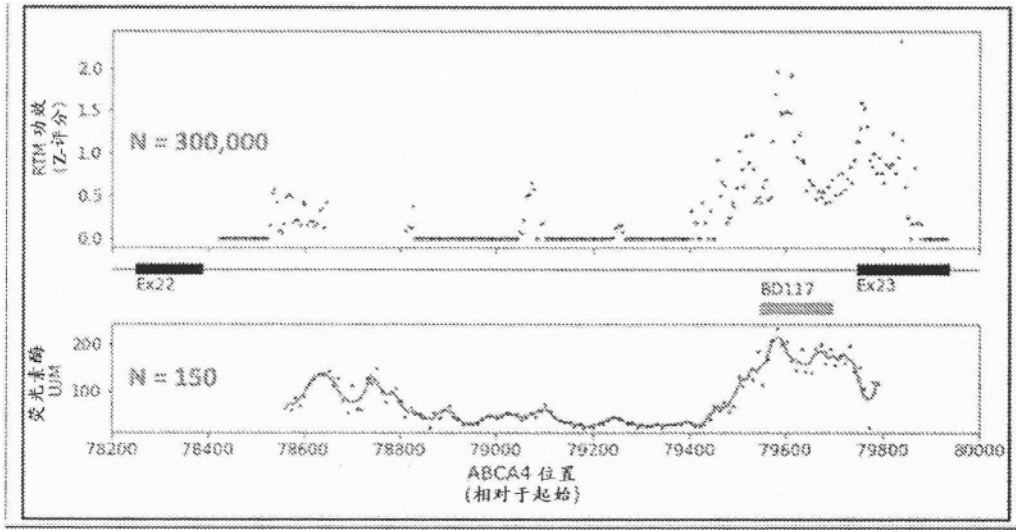


图16

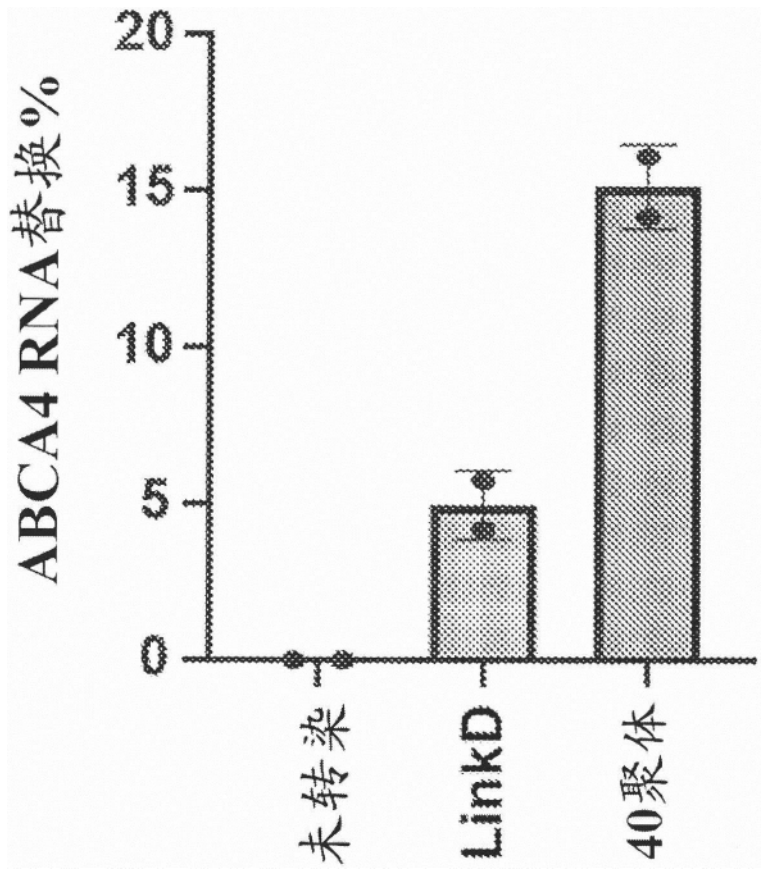


图17A

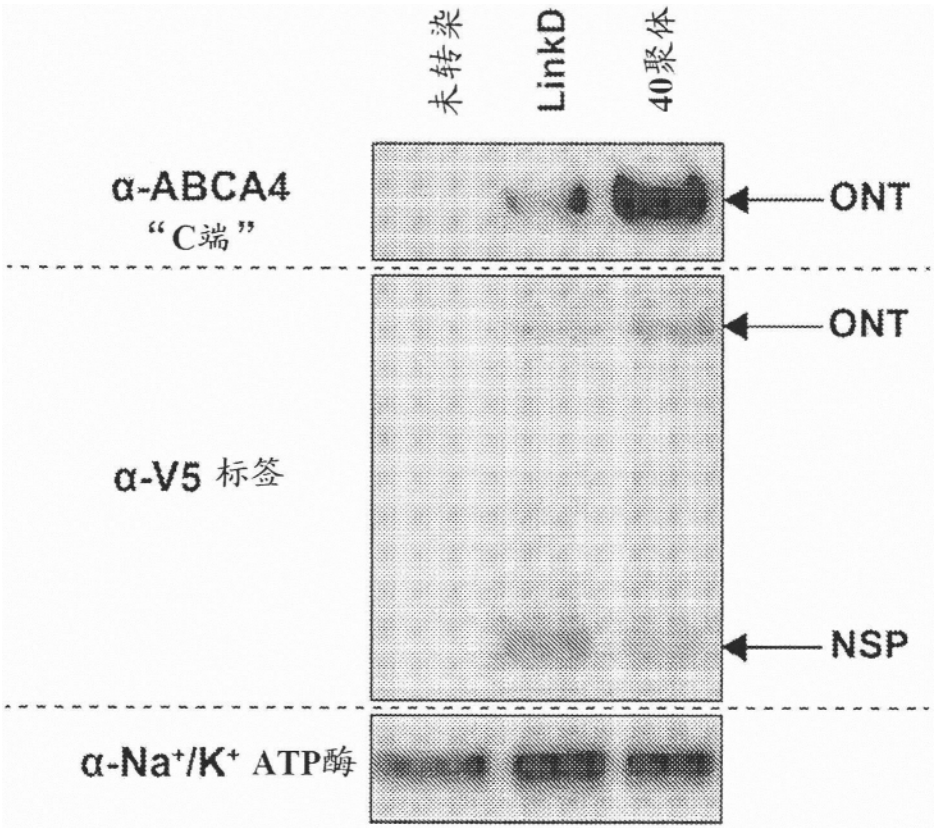


图17B

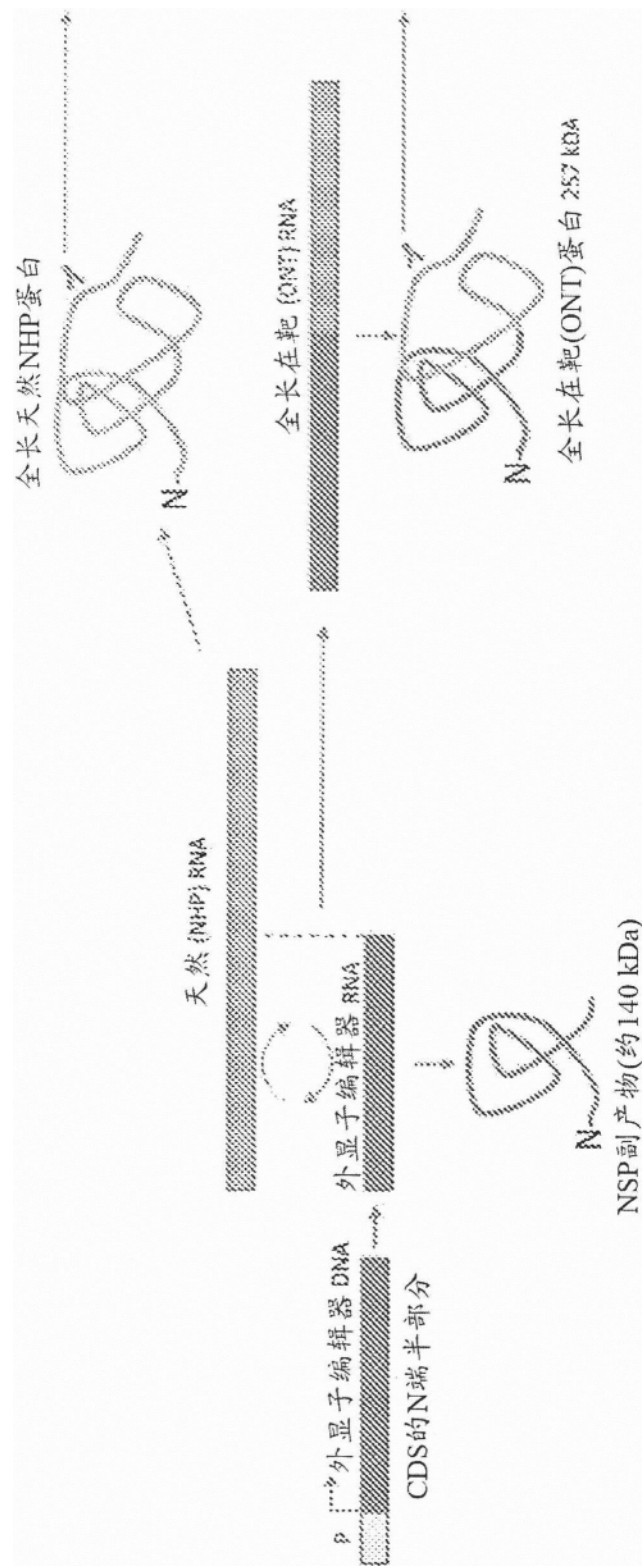


图18

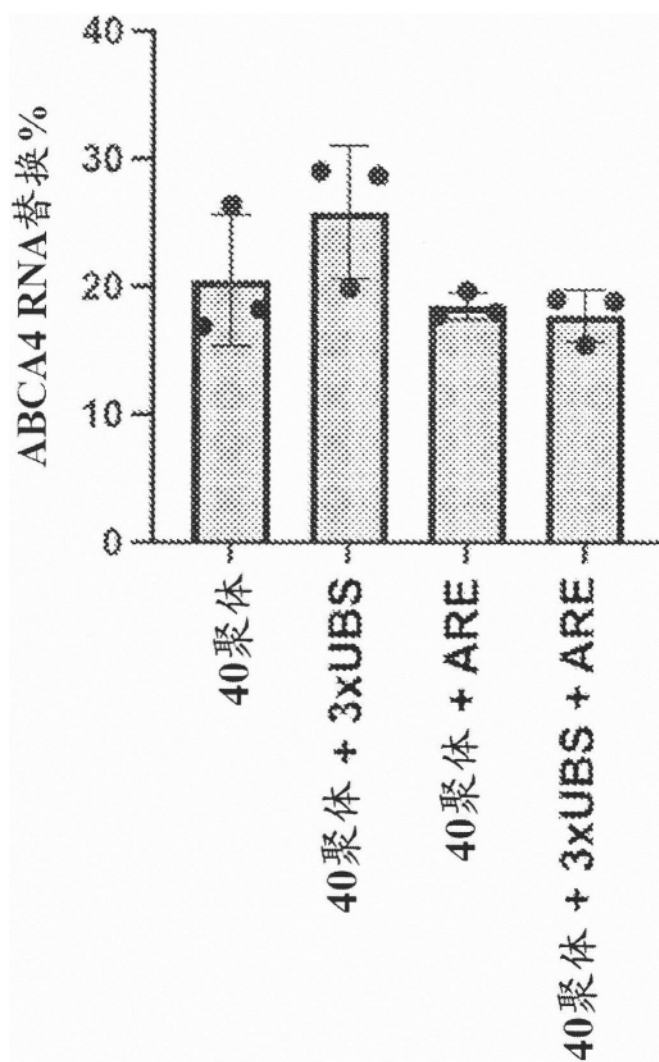


图19A

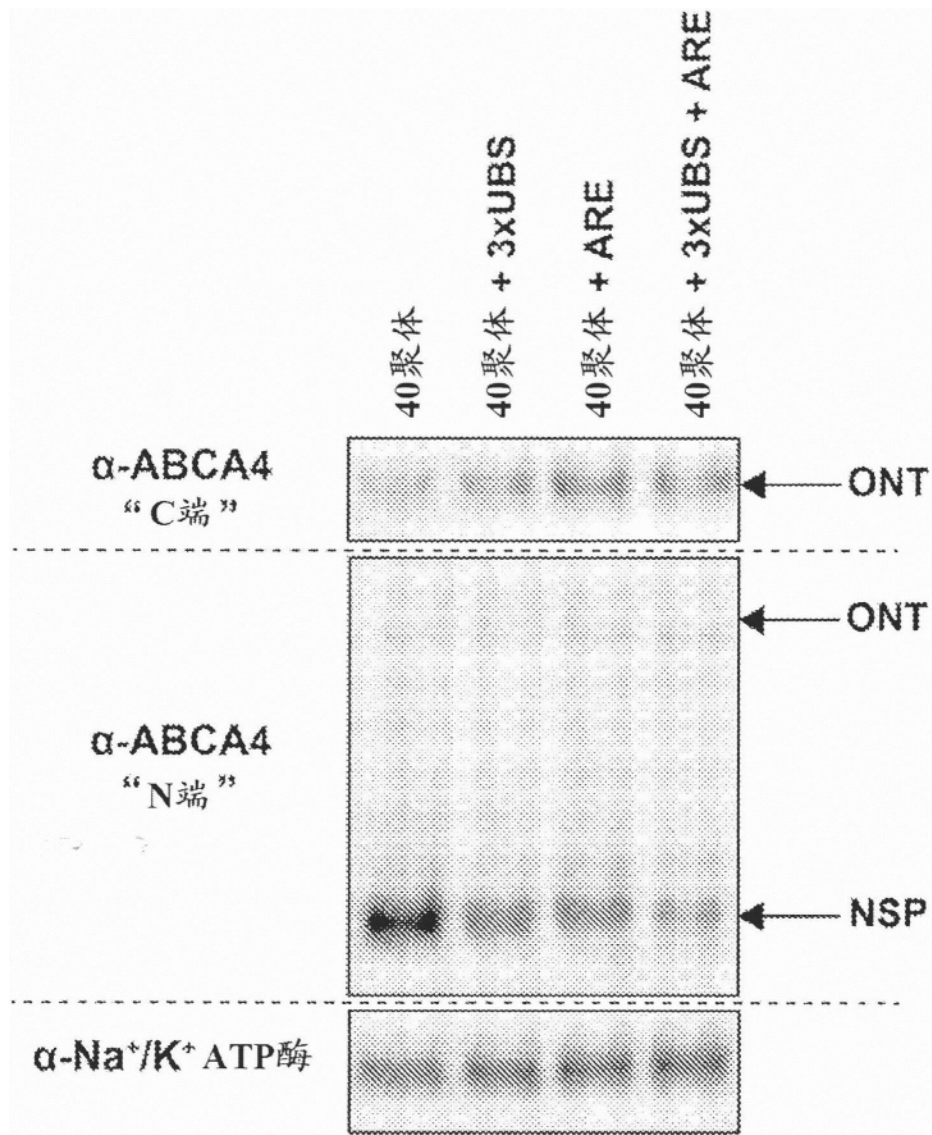


图19B

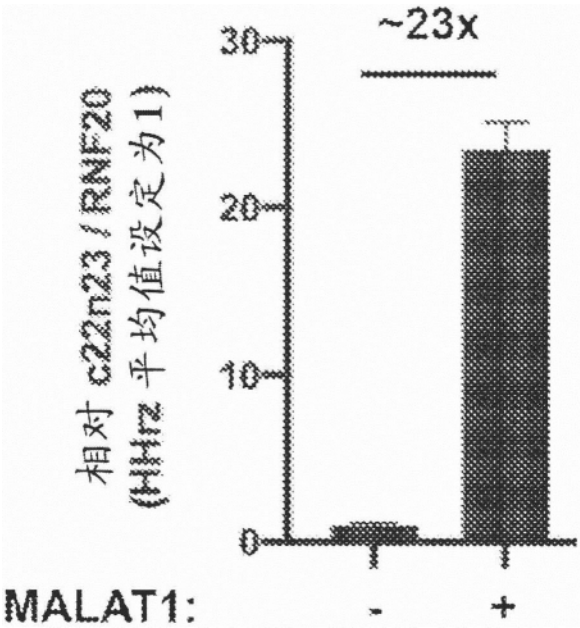


图20A

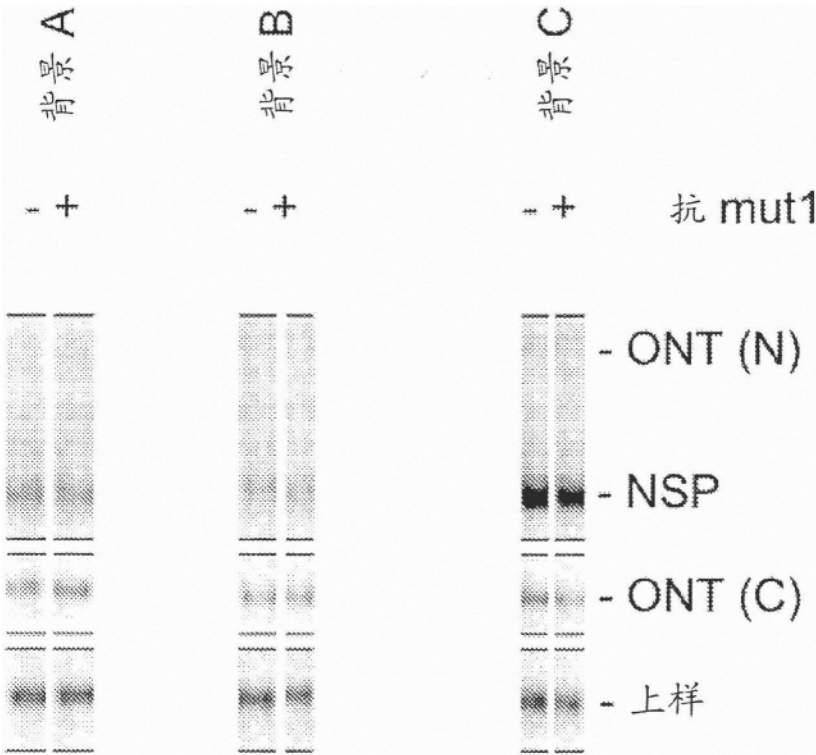


图20B

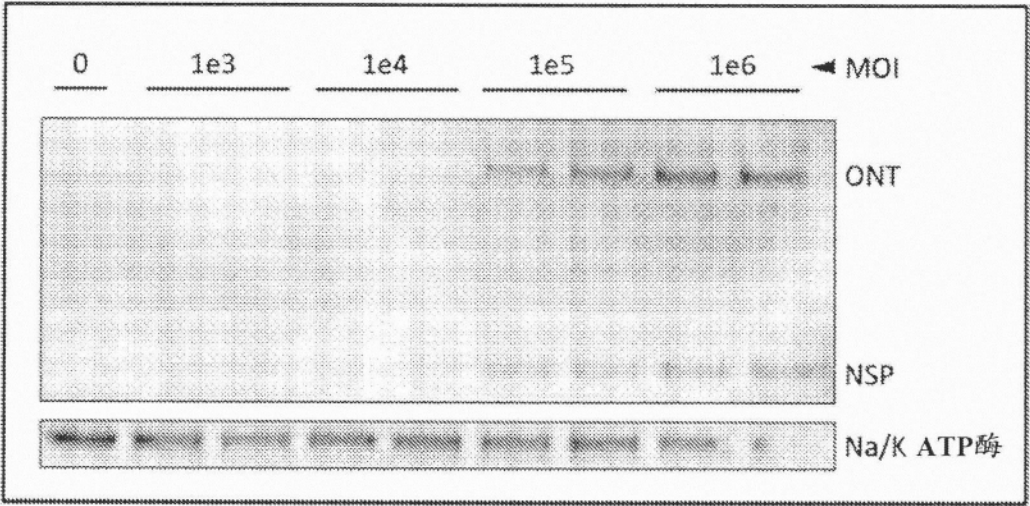


图21A

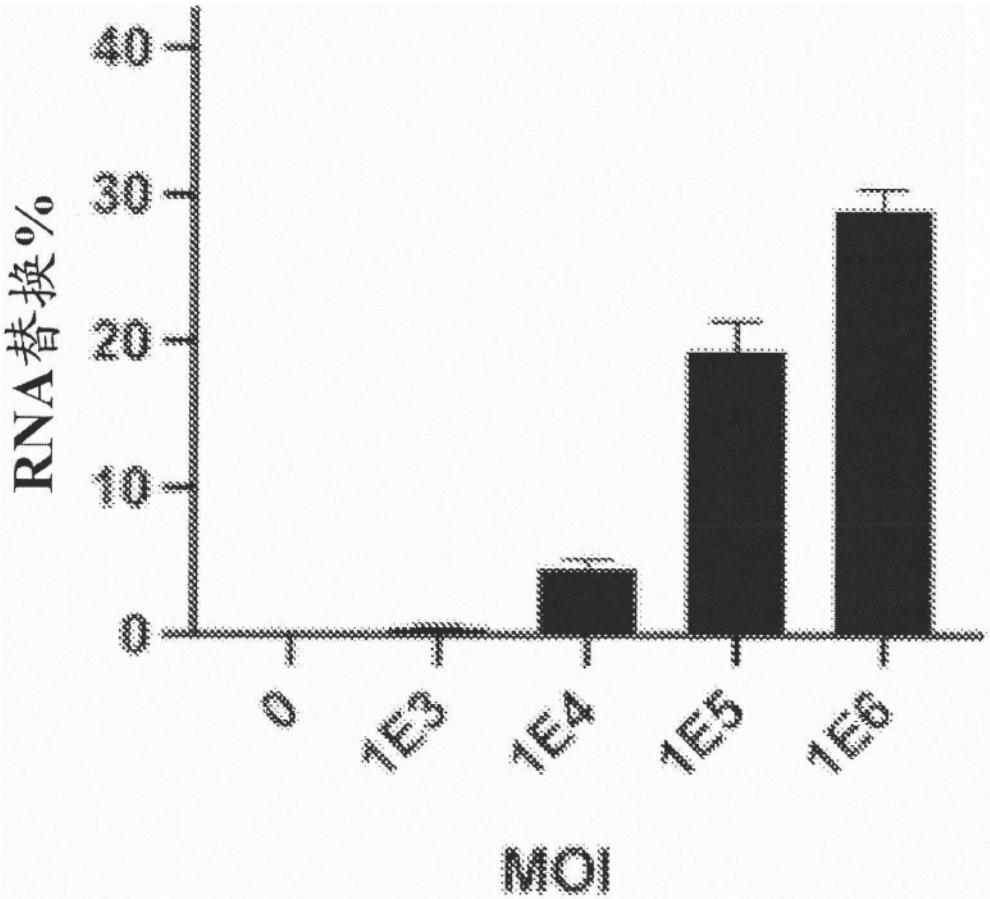


图21B

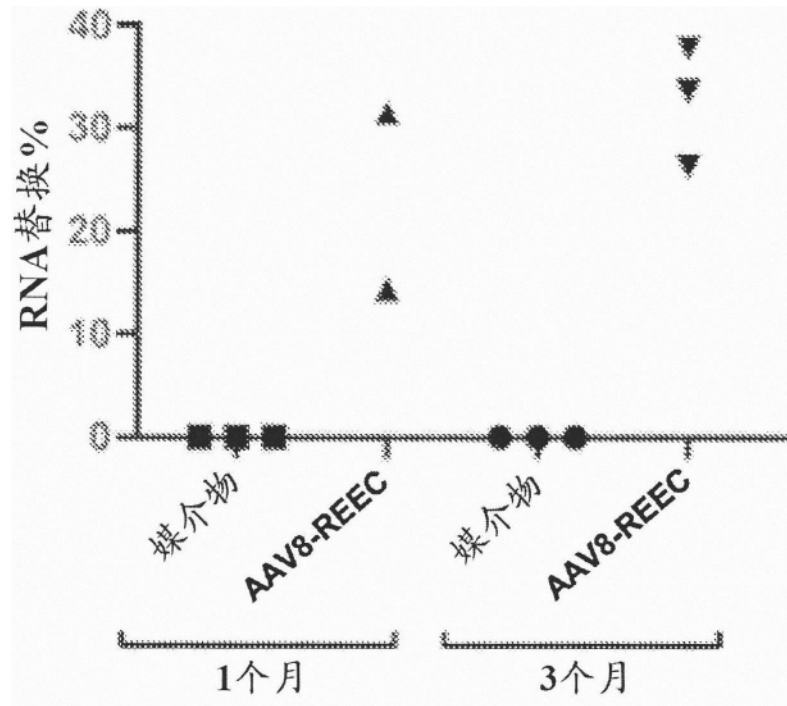


图22A

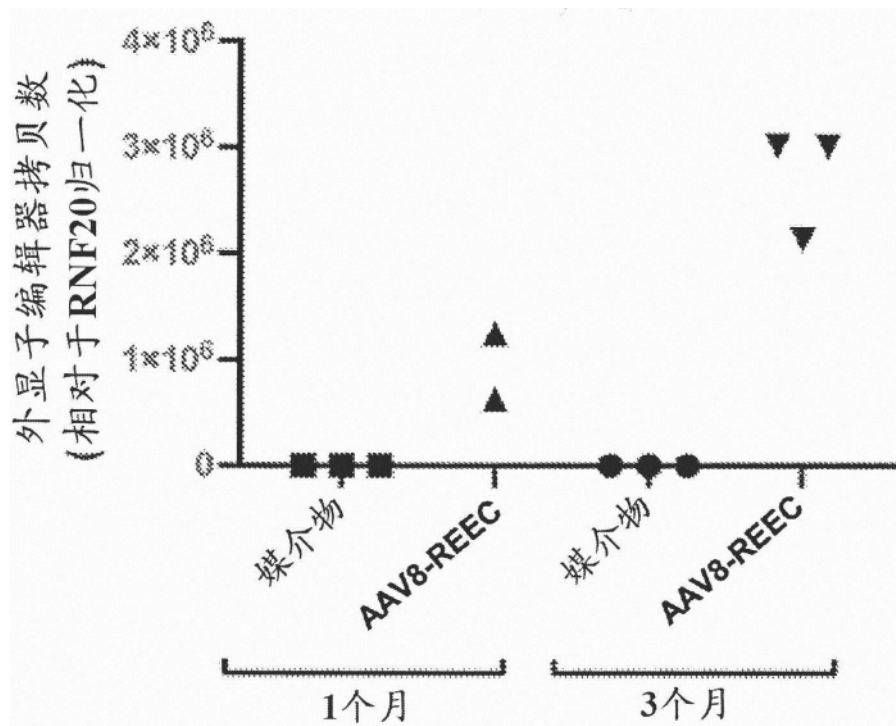


图22B



图23

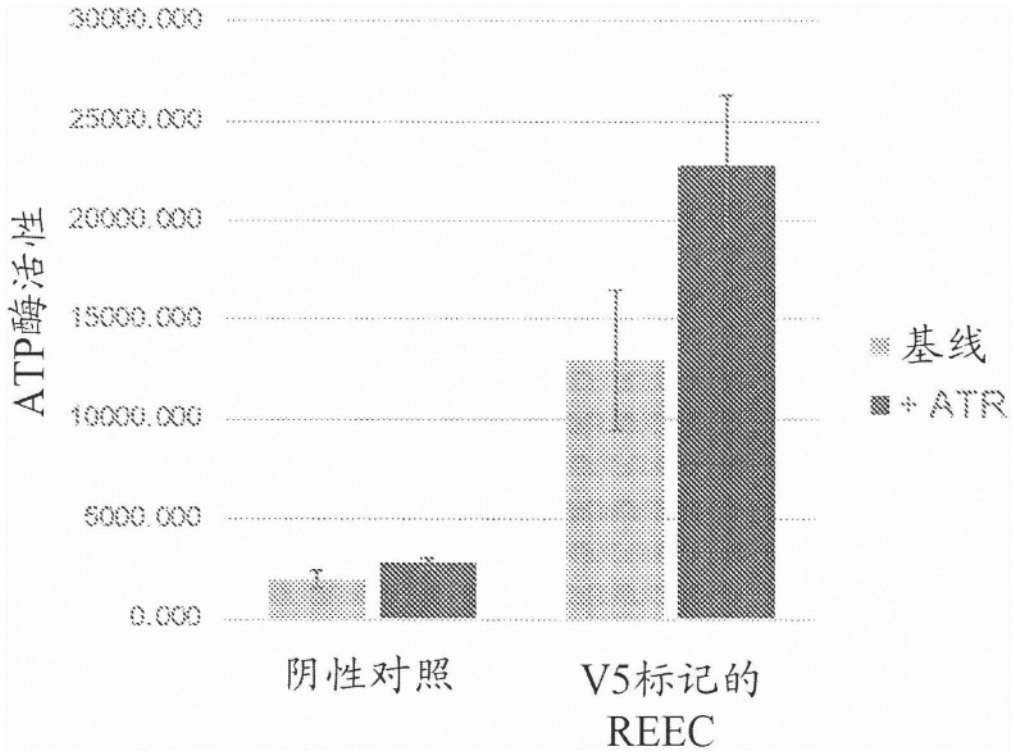


图24

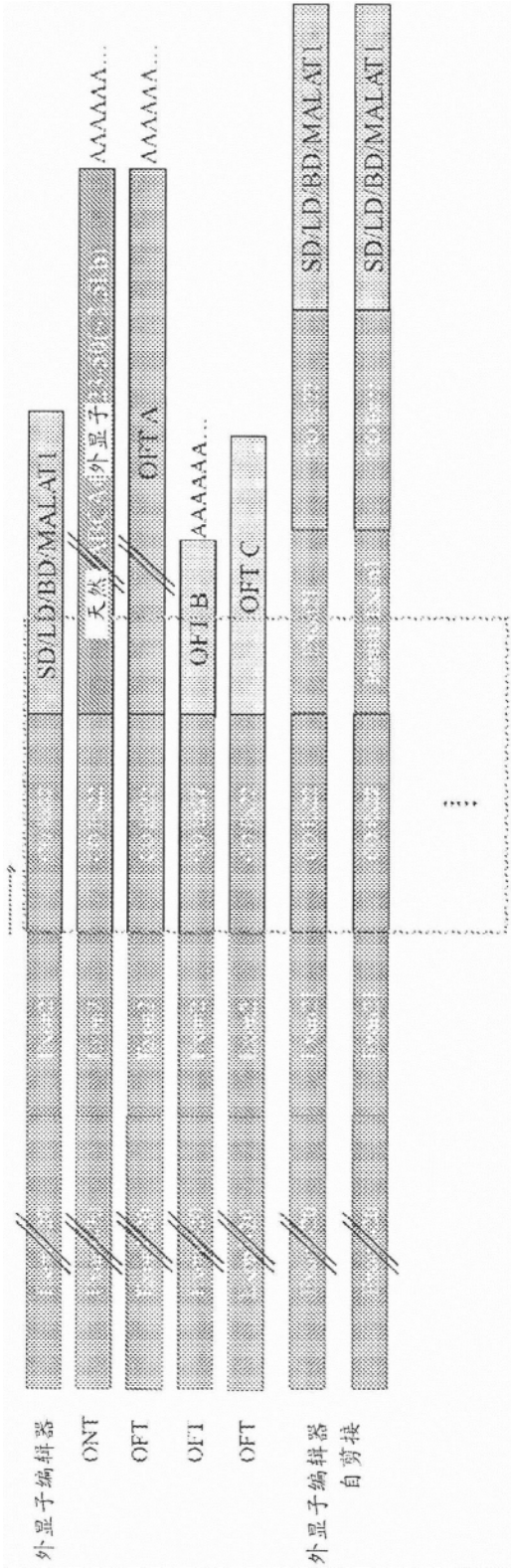


图25

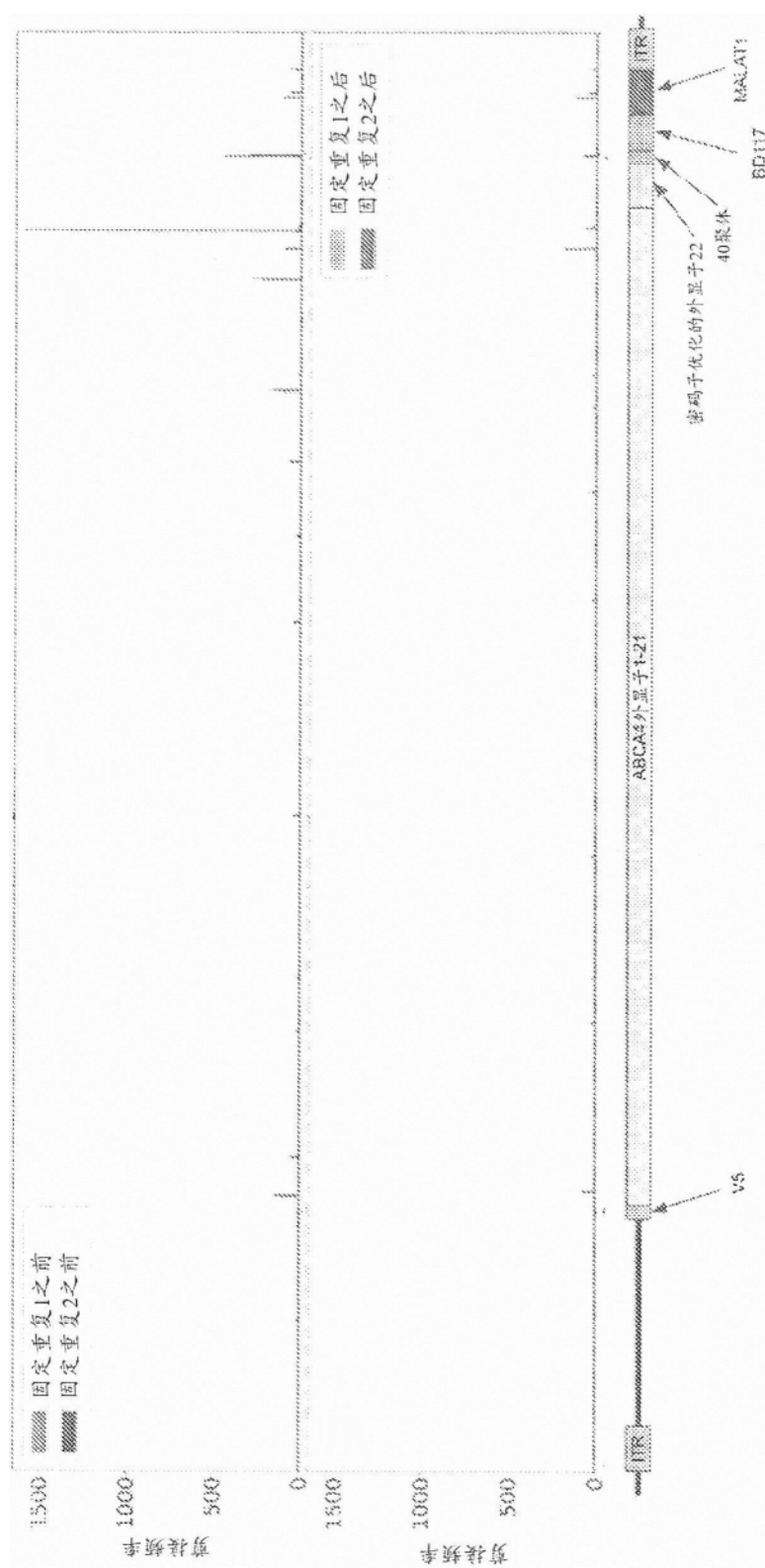


图26

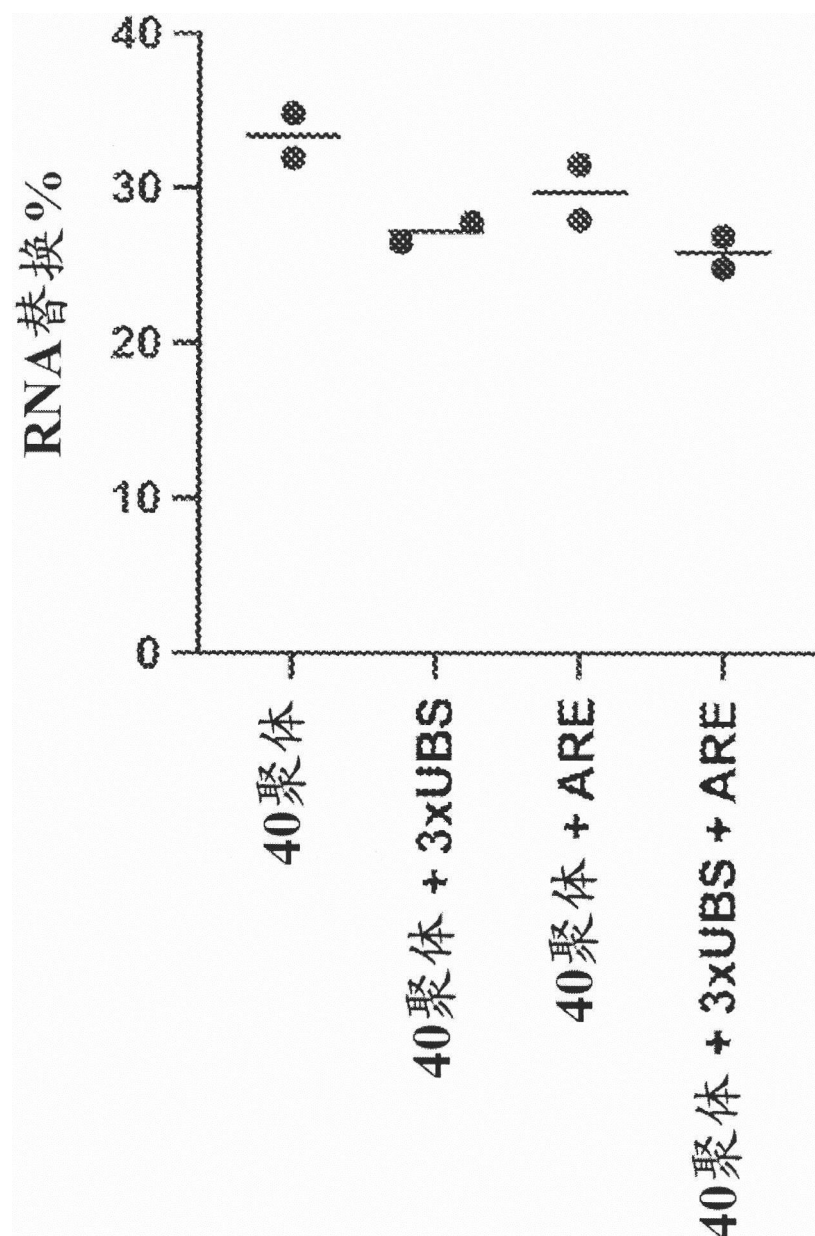


图27A

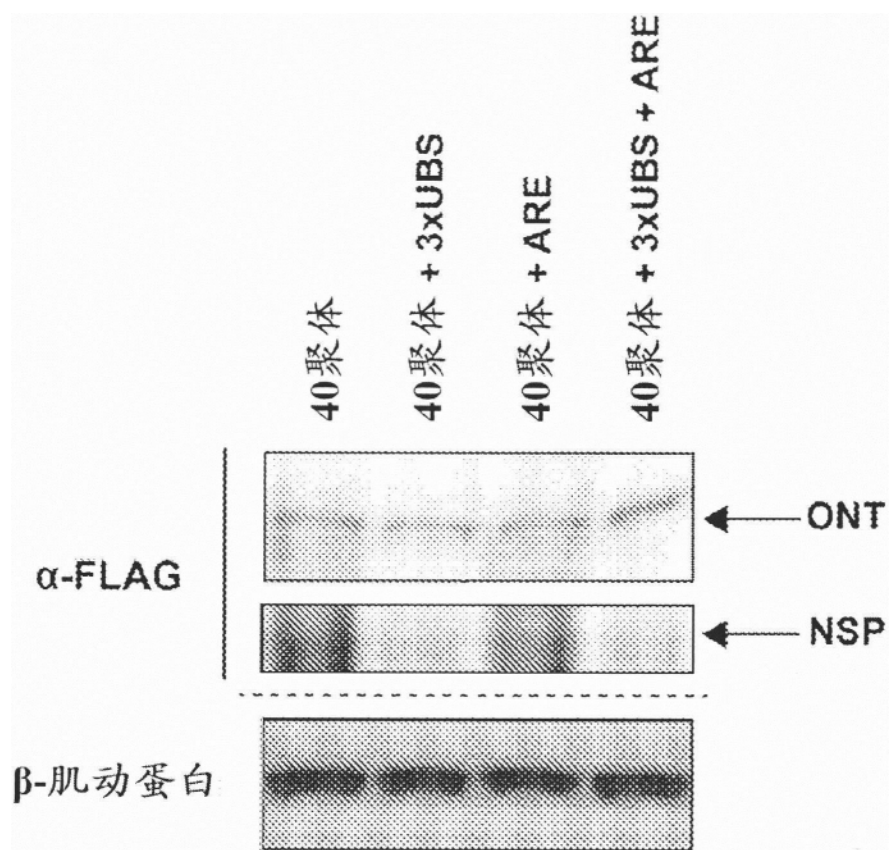


图27B

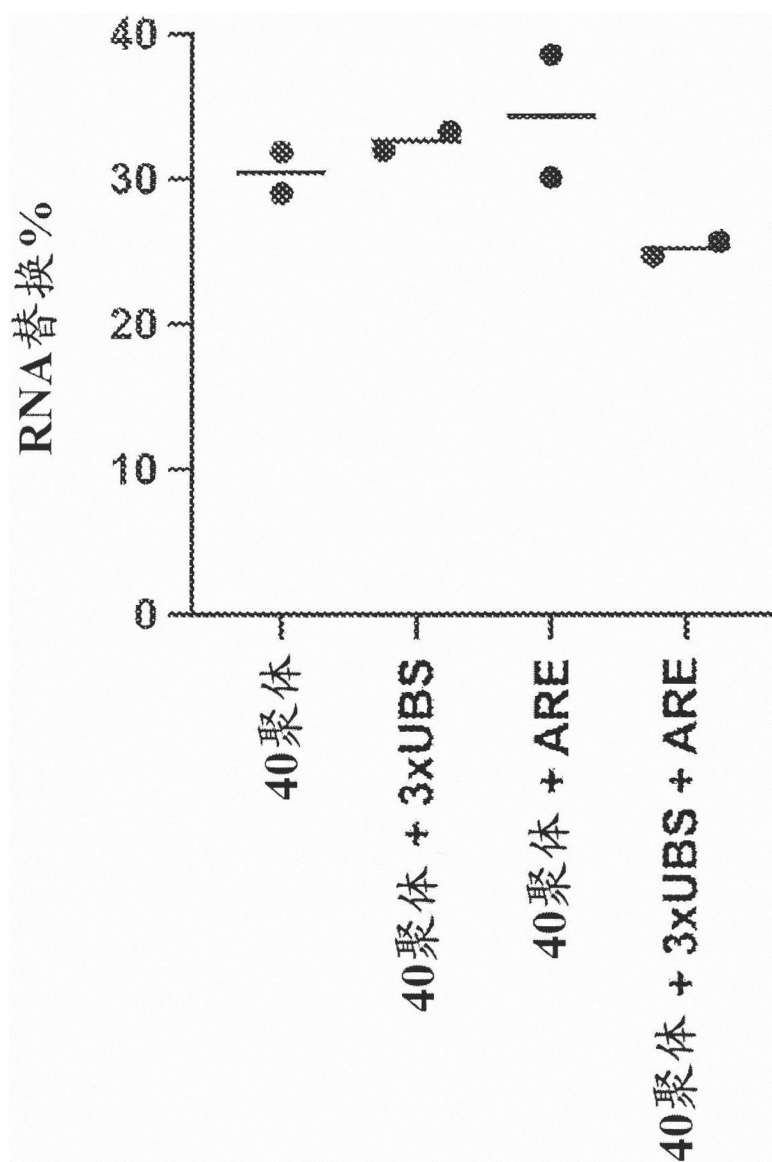


图28A

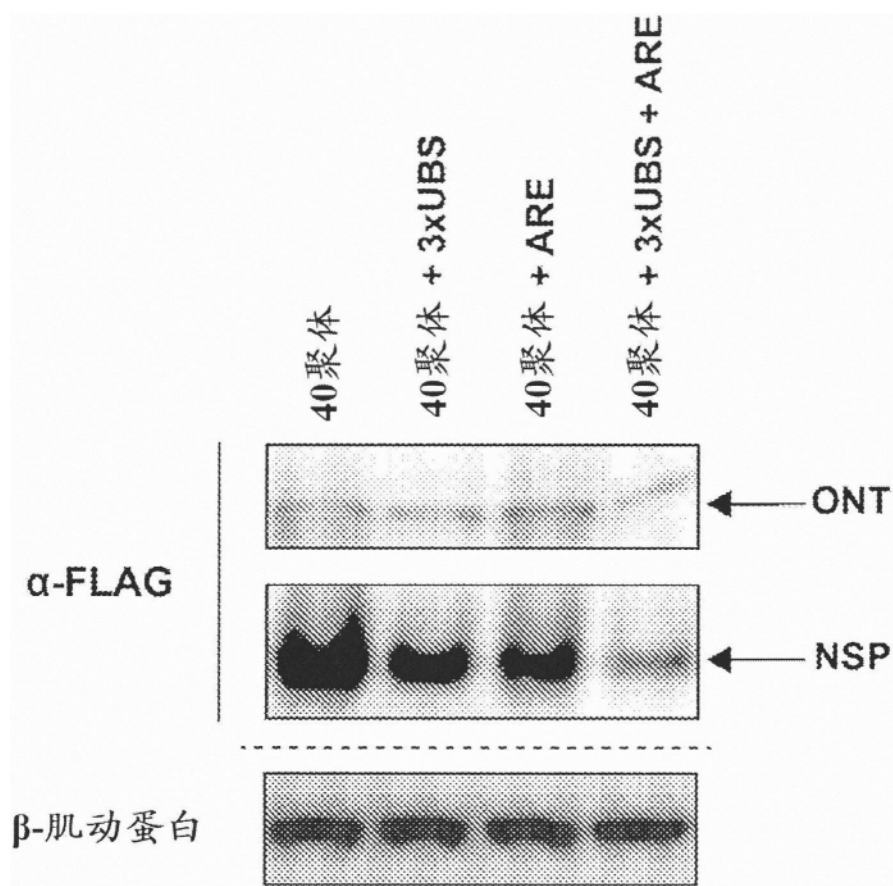


图28B

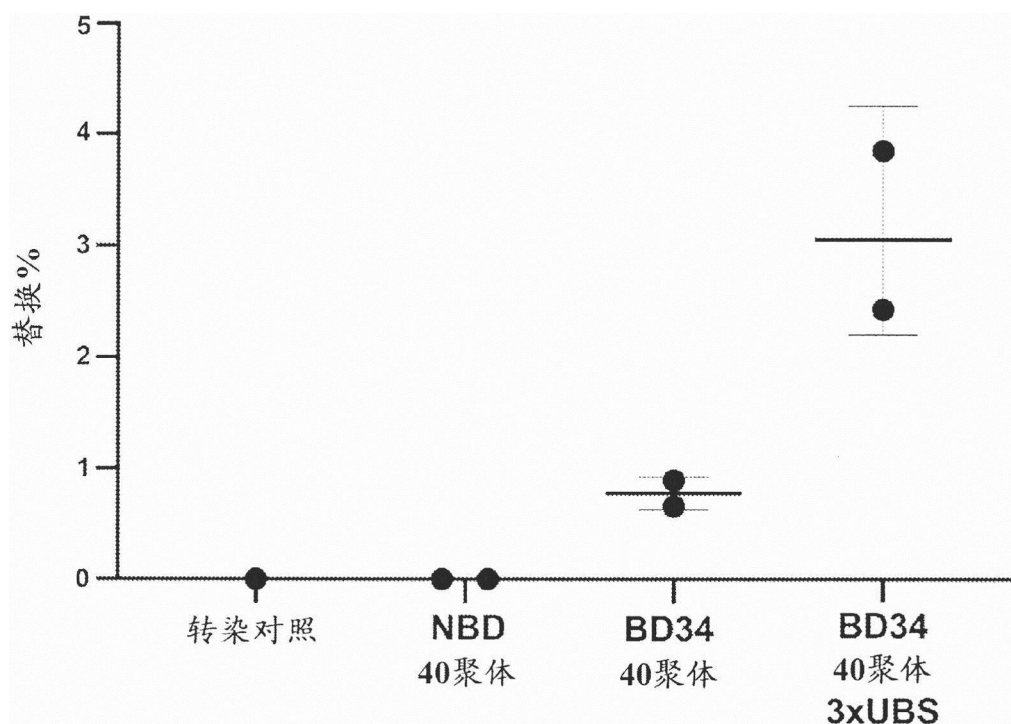


图29

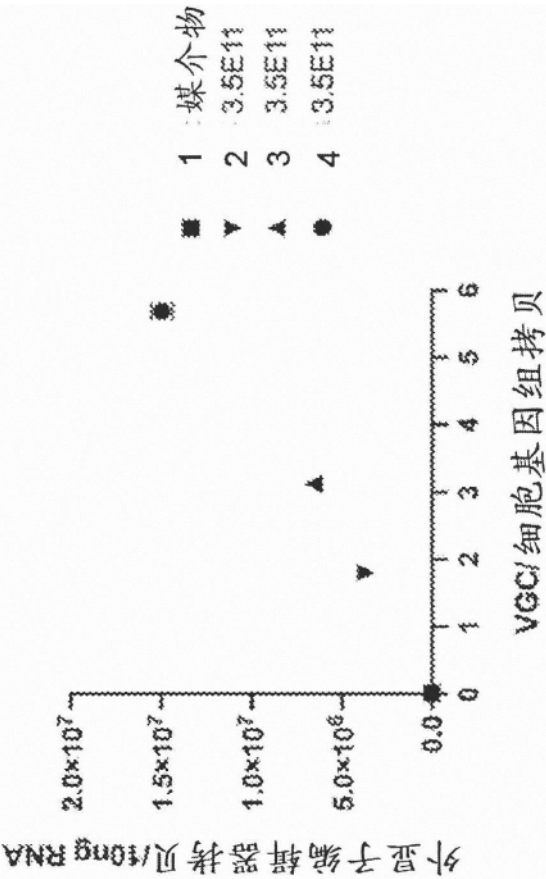


图30A

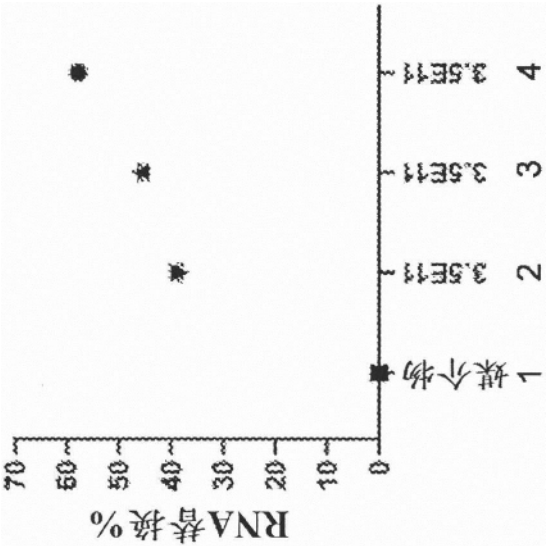


图30B

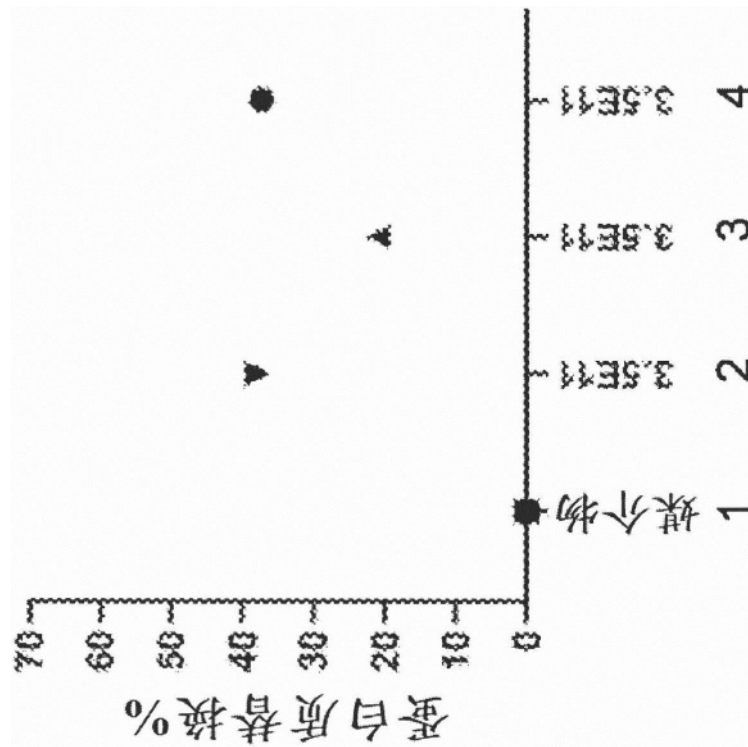


图30C

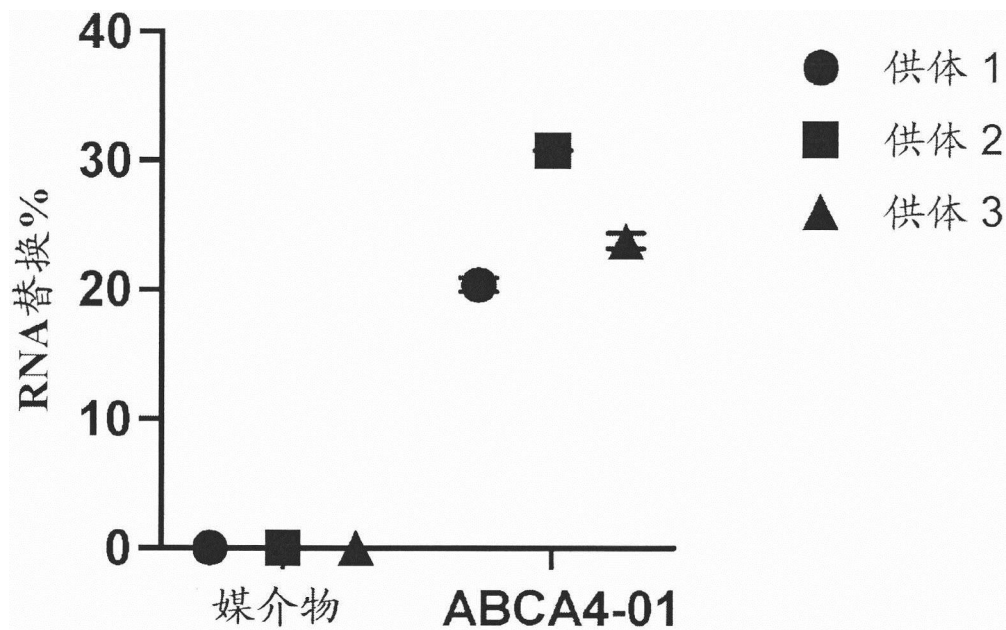


图31