



## (12)发明专利申请

(10)申请公布号 CN 110914433 A

(43)申请公布日 2020.03.24

(21)申请号 201880020941.8

(74)专利代理机构 北京纪凯知识产权代理有限公司 11245

(22)申请日 2018.03.23

代理人 王永伟

(30)优先权数据

PCT/EP2017/057110 2017.03.24 EP

PCT/EP2017/076775 2017.10.19 EP

(51)Int.Cl.

C12N 15/67(2006.01)

(85)PCT国际申请进入国家阶段日

2019.09.24

(86)PCT国际申请的申请数据

PCT/EP2018/057552 2018.03.23

(87)PCT国际申请的公布数据

W02018/172556 EN 2018.09.27

(71)申请人 库尔维科公司

地址 德国蒂宾根

(72)发明人 F·切维斯图内森 M·波尼什

T·施莱克

权利要求书12页 说明书136页 附图10页

(54)发明名称

编码CRISPR相关蛋白质的核酸及其用途

(57)摘要

本发明涉及生物药物领域,并且具体地涉及治疗性核酸领域。本发明提供编码CRISPR相关蛋白质的人工核酸,具体地RNA。还提供包括其的(药物)组合物和多部分试剂盒。此外,本发明涉及用于药物和具体地用于治疗 and/或预防适于用CRISPR相关蛋白质治疗的疾病的人工核酸、(药物)组合物、或多部分试剂盒。

1. 人工核酸分子, 包括
  - a. 至少一个编码区, 编码至少一种CRISPR相关蛋白质;
  - b. 至少一个5' 非翻译区 (5' UTR) 元件, 衍生自选自SLC7A3、ATP5A1、RPL32、HSD17B4、NOSIP、ASAHI、RPL31、TUBB4B、UBQLN2、MP68和NDUFA4的基因的5' UTR; 和
  - c. 至少一个3' 非翻译区 (3' UTR) 元件, 衍生自选自GNAS、CASP1、PSMB3、ALB、COX6B1、NDUFA1和RPS9的基因的3' UTR。
2. 根据权利要求1所述的人工核酸分子, 其中各所述基因包括天然存在的DNA序列及其同源物、变体、片段、和相应RNA序列。
3. 根据权利要求1或2所述的人工核酸分子, 包括
  - a. 衍生自SLC7A3基因的5' UTR或其相应RNA序列、同源物、片段或变体的至少一个5' UTR元件、和衍生自GNAS基因的3' UTR或其相应RNA序列、同源物、片段或变体的至少一个3' UTR元件; 或
  - b. 衍生自HSD17B4基因的5' UTR或其相应RNA序列、同源物、片段或变体的至少一个5' UTR元件、和衍生自GNAS基因的3' UTR或其相应RNA序列、同源物、片段或变体的至少一个3' UTR元件; 或
  - c. 衍生自ATP5A1基因的5' UTR或其相应RNA序列、同源物、片段或变体的至少一个5' UTR元件、和衍生自CASP1基因的3' UTR或其相应RNA序列、同源物、片段或变体的至少一个3' UTR元件; 或
  - d. 衍生自NDUFA4基因的5' UTR或其相应RNA序列、同源物、片段或变体的至少一个5' UTR元件、和衍生自PSMB3基因的3' UTR或其相应RNA序列、同源物、片段或变体的至少一个3' UTR元件; 或
  - e. 衍生自HSD17B4基因的5' UTR或其相应RNA序列、同源物、片段或变体的至少一个5' UTR元件、和衍生自PSMB3基因的3' UTR或其相应RNA序列、同源物、片段或变体的至少一个3' UTR元件; 或
  - f. 衍生自RPL32基因的5' UTR或其相应RNA序列、同源物、片段或变体的至少一个5' UTR元件、和衍生自ALB基因的3' UTR或其相应RNA序列、同源物、片段或变体的至少一个3' UTR元件; 或
  - g. 衍生自HSD17B4基因的5' UTR或其相应RNA序列、同源物、片段或变体的至少一个5' UTR元件、和衍生自CASP1基因的3' UTR或其相应RNA序列、同源物、片段或变体的至少一个3' UTR元件; 或
  - h. 衍生自SLC7A3基因的5' UTR或其相应RNA序列、同源物、片段或变体的至少一个5' UTR元件、和衍生自CASP1基因的3' UTR或其相应RNA序列、同源物、片段或变体的至少一个3' UTR元件; 或
  - i. 衍生自SLC7A3基因的5' UTR或其相应RNA序列、同源物、片段或变体的至少一个5' UTR元件、和衍生自PSMB3基因的3' UTR或其相应RNA序列、同源物、片段或变体的至少一个3' UTR元件; 或
  - j. 衍生自NOSIP基因的5' UTR或其相应RNA序列、同源物、片段或变体的至少一个5' UTR元件、和衍生自PSMB3基因的3' UTR或其相应RNA序列、同源物、片段或变体的至少一个3' UTR元件; 或

k. 衍生自NDUFA4基因的5' UTR或其相应RNA序列、同源物、片段或变体的至少一个5' UTR元件、和衍生自RPS9基因的3' UTR或其相应RNA序列、同源物、片段或变体的至少一个3' UTR元件;或

1. 衍生自HSD17B4基因的5' UTR或其相应RNA序列、同源物、片段或变体的至少一个5' UTR元件、和衍生自RPS9基因的3' UTR或其相应RNA序列、同源物、片段或变体的至少一个3' UTR元件;或

m. 衍生自ATP5A1基因的5' UTR或其同源物、片段或变体的至少一个5' UTR元件、和衍生自GNAS基因的3' UTR或其同源物、片段或变体的至少一个3' UTR元件;或

n. 衍生自NDUFA4基因的5' UTR或其同源物、片段或变体的至少一个5' UTR元件、和衍生自COX6B1基因的3' UTR或其同源物、片段或变体的至少一个3' UTR元件;或

n. 衍生自NDUFA4基因的5' UTR或其同源物、片段或变体的至少一个5' UTR元件、和衍生自GNAS基因的3' UTR或其同源物、片段或变体的至少一个3' UTR元件;或

o. 衍生自NDUFA4基因的5' UTR或其同源物、片段或变体的至少一个5' UTR元件、和衍生自NDUFA1基因的3' UTR或其同源物、片段或变体的至少一个3' UTR元件;或

p. 衍生自NOSIP基因的5' UTR或其同源物、片段或变体的至少一个5' UTR元件、和衍生自NDUFA1基因的3' UTR或其同源物、片段或变体的至少一个3' UTR元件;或

q. 衍生自RPL31基因的5' UTR或其同源物、片段或变体的至少一个5' UTR元件、和衍生自GNAS基因的3' UTR或其同源物、片段或变体的至少一个3' UTR元件;或

r. 衍生自TUBB4B基因的5' UTR或其同源物、片段或变体的至少一个5' UTR元件、和衍生自RPS9基因的3' UTR或其同源物、片段或变体的至少一个3' UTR元件;或

s. 衍生自UBQLN2基因的5' UTR或其同源物、片段或变体的至少一个5' UTR元件、和衍生自RPS9基因的3' UTR或其同源物、片段或变体的至少一个3' UTR元件;

t. 衍生自MP68基因的5' UTR或其同源物、片段或变体的至少一个5' UTR元件、和衍生自GNAS基因的3' UTR或其同源物、片段或变体的至少一个3' UTR元件;或

u. 衍生自MP68基因的5' UTR或其同源物、片段或变体的至少一个5' UTR元件、和衍生自NDUFA1基因的3' UTR或其同源物、片段或变体的至少一个3' UTR元件。

4. 根据权利要求3所述的人工核酸分子,包括根据d、e、g或1的UTR元件。

5. 根据权利要求1至4中任一项所述的人工核酸分子,其中

-所述衍生自HSD17B4基因的5' UTR元件包括下列或由下列组成:根据SEQ ID NO:1的DNA序列或按递增优选顺序与根据SEQ ID NO:1的核酸序列具有至少50%、60%、70%、80%、90%、95%、96%、97%、98%、或99%序列同一性的DNA序列、或其片段或变体;或根据SEQ ID NO:2的RNA序列、或按递增优选顺序与根据SEQ ID NO:2的核酸序列具有至少50%、60%、70%、80%、90%、95%、96%、97%、98%、或99%序列同一性的RNA序列、或其片段或变体;

-所述衍生自RPL32基因的5' UTR元件包括下列或由下列组成:根据SEQ ID NO:21的DNA序列或按递增优选顺序与根据SEQ ID NO:21的核酸序列具有至少50%、60%、70%、80%、90%、95%、96%、97%、98%、或99%序列同一性的DNA序列、或其片段或变体;或根据SEQ ID NO:22的RNA序列、或按递增优选顺序与根据SEQ ID NO:22的核酸序列具有至少50%、60%、70%、80%、90%、95%、96%、97%、98%、或99%序列同一性的RNA序列、或其片段或

变体;

-所述衍生自NDUFA4基因的5' UTR元件包括下列或由下列组成:根据SEQ ID NO:9的DNA序列、或按递增优选顺序与根据SEQ ID NO:9的核酸序列具有至少50%、60%、70%、80%、90%、95%、96%、97%、98%、或99%序列同一性的DNA序列、或其片段或变体;或根据SEQ ID NO:10的RNA序列、或按递增优选顺序与根据SEQ ID NO:10的核酸序列具有至少50%、60%、70%、80%、90%、95%、96%、97%、98%、或99%序列同一性的RNA序列、或其片段或变体;

-所述衍生自SLC7A3基因的5' UTR元件包括下列或由下列组成:根据SEQ ID NO:15的DNA序列、或按递增优选顺序与根据SEQ ID NO:15的核酸序列具有至少50%、60%、70%、80%、90%、95%、96%、97%、98%、或99%序列同一性的DNA序列、或其片段或变体;或根据SEQ ID NO:16的RNA序列、或按递增优选顺序与根据SEQ ID NO:16的核酸序列具有至少50%、60%、70%、80%、90%、95%、96%、97%、98%、或99%序列同一性的RNA序列、或其片段或变体;

-所述衍生自NOSIP基因的5' UTR元件包括下列或由下列组成:根据SEQ ID NO:11的DNA序列、或按递增优选顺序与根据SEQ ID NO:11的核酸序列具有至少50%、60%、70%、80%、90%、95%、96%、97%、98%、或99%序列同一性的DNA序列、或其片段或变体;或根据SEQ ID NO:12的RNA序列、或按递增优选顺序与根据SEQ ID NO:12的核酸序列具有至少50%、60%、70%、80%、90%、95%、96%、97%、98%、或99%序列同一性的RNA序列、或其片段或变体;

-所述衍生自ATP5A1基因的5' UTR元件包括下列或由下列组成:根据SEQ ID NO:5的DNA序列、或按递增优选顺序与根据SEQ ID NO:5的核酸序列具有至少50%、60%、70%、80%、90%、95%、96%、97%、98%、或99%序列同一性的DNA序列、或其片段或变体;或根据SEQ ID NO:6的RNA序列、或按递增优选顺序与根据SEQ ID NO:6的核酸序列具有至少50%、60%、70%、80%、90%、95%、96%、97%、98%、或99%序列同一性的RNA序列、或其片段或变体;

所述衍生自ASAHI1基因的5' UTR元件包括下列或由下列组成:根据SEQ ID NO:3的DNA序列、或按递增优选顺序与根据SEQ ID NO:3的核酸序列具有至少50%、60%、70%、80%、90%、95%、96%、97%、98%、或99%序列同一性的DNA序列、或其片段或变体;或根据SEQ ID NO:4的RNA序列、或按递增优选顺序与根据SEQ ID NO:4的核酸序列具有至少50%、60%、70%、80%、90%、95%、96%、97%、98%、或99%序列同一性的RNA序列、或其片段或变体;

所述衍生自Mp68基因的5' UTR元件包括下列或由下列组成:根据SEQ ID NO:7的DNA序列、或按递增优选顺序与根据SEQ ID NO:7的核酸序列具有至少50%、60%、70%、80%、90%、95%、96%、97%、98%、或99%序列同一性的DNA序列、或其片段或变体;或根据SEQ ID NO:8的RNA序列、或按递增优选顺序与根据SEQ ID NO:8的核酸序列具有至少50%、60%、70%、80%、90%、95%、96%、97%、98%、或99%序列同一性的RNA序列、或其片段或变体;

所述衍生自Rp131基因的5' UTR元件包括下列或由下列组成:根据SEQ ID NO:13的DNA序列、或按递增优选顺序与根据SEQ ID NO:13的核酸序列具有至少50%、60%、70%、80%、



90%、95%、96%、97%、98%、或99%序列同一性的DNA序列、或其片段或变体；或根据SEQ ID NO:14的RNA序列、或按递增优选顺序与根据SEQ ID NO:14的核酸序列具有至少50%、60%、70%、80%、90%、95%、96%、97%、98%、或99%序列同一性的RNA序列、或其片段或变体；

所述衍生自TUBB4B基因的5' UTR元件包括下列或由下列组成：根据SEQ ID NO:17的DNA序列、或按递增优选顺序与根据SEQ ID NO:17的核酸序列具有至少50%、60%、70%、80%、90%、95%、96%、97%、98%、或99%序列同一性的DNA序列、或其片段或变体；或根据SEQ ID NO:18的RNA序列、或按递增优选顺序与根据SEQ ID NO:18的核酸序列具有至少50%、60%、70%、80%、90%、95%、96%、97%、98%、或99%序列同一性的RNA序列、或其片段或变体；

所述衍生自Ubqln2基因的5' UTR元件包括下列或由下列组成：根据SEQ ID NO:19的DNA序列、或按递增优选顺序与根据SEQ ID NO:19的核酸序列具有至少50%、60%、70%、80%、90%、95%、96%、97%、98%、或99%序列同一性的DNA序列、或其片段或变体；或根据SEQ ID NO:20的RNA序列、或按递增优选顺序与根据SEQ ID NO:20的核酸序列具有至少50%、60%、70%、80%、90%、95%、96%、97%、98%、或99%序列同一性的RNA序列、或其片段或变体；

-所述衍生自GNAS基因的3' UTR元件包括下列或由下列组成：根据SEQ ID NO:29的DNA序列、或按递增优选顺序与根据SEQ ID NO:29的核酸序列具有至少50%、60%、70%、80%、90%、95%、96%、97%、98%、或99%序列同一性的DNA序列、或其片段或变体；根据SEQ ID NO:30的RNA序列、或按递增优选顺序与根据SEQ ID NO:30的核酸序列具有至少50%、60%、70%、80%、90%、95%、96%、97%、98%、或99%序列同一性的RNA序列、或其片段或变体；

-所述衍生自CASP1基因的3' UTR元件包括下列或由下列组成：根据SEQ ID NO:25的DNA序列、或按递增优选顺序与根据SEQ ID NO:25的核酸序列具有至少50%、60%、70%、80%、90%、95%、96%、97%、98%、或99%序列同一性的DNA序列、或其片段或变体；或根据SEQ ID NO:26的RNA序列、或按递增优选顺序与根据SEQ ID NO:26的核酸序列具有至少50%、60%、70%、80%、90%、95%、96%、97%、98%、或99%序列同一性的RNA序列、或其片段或变体；

-所述衍生自PSMB3基因的3' UTR元件包括下列或由下列组成：根据SEQ ID NO:23的DNA序列、或按递增优选顺序与根据SEQ ID NO:23的核酸序列具有至少50%、60%、70%、80%、90%、95%、96%、97%、98%、或99%序列同一性的DNA序列、或其片段或变体；或根据SEQ ID NO:24的RNA序列、或按递增优选顺序与根据SEQ ID NO:24的核酸序列具有至少50%、60%、70%、80%、90%、95%、96%、97%、98%、或99%序列同一性的RNA序列、或其片段或变体；

-所述衍生自ALB基因的3' UTR元件包括下列或由下列组成：根据SEQ ID NO:35的DNA序列、或按递增优选顺序与根据SEQ ID NO:35的核酸序列具有至少50%、60%、70%、80%、90%、95%、96%、97%、98%、或99%序列同一性的DNA序列、或其片段或变体；或根据SEQ ID NO:36的RNA序列、或按递增优选顺序与根据SEQ ID NO:36的核酸序列具有至少50%、60%、70%、80%、90%、95%、96%、97%、98%、或99%序列同一性的RNA序列、或其片段或变体；

-所述衍生自RPS9基因的3' UTR元件包括下列或由下列组成:根据SEQ ID NO:33的DNA序列、或按递增优选顺序与根据SEQ ID NO:33的核酸序列具有至少50%、60%、70%、80%、90%、95%、96%、97%、98%、或99%序列同一性的DNA序列、或其片段或变体;或根据SEQ ID NO:34的RNA序列、或按递增优选顺序与根据SEQ ID NO:34的核酸序列具有至少50%、60%、70%、80%、90%、95%、96%、97%、98%、或99%序列同一性的RNA序列、或其片段或变体;

-所述衍生自COX6B1基因的3' UTR元件包括下列或由下列组成:根据SEQ ID NO:27的DNA序列、或按递增优选顺序与根据SEQ ID NO:27的核酸序列具有至少50%、60%、70%、80%、90%、95%、96%、97%、98%、或99%序列同一性的DNA序列、或其片段或变体;或根据SEQ ID NO:28的RNA序列、或按递增优选顺序与根据SEQ ID NO:28的核酸序列具有至少50%、60%、70%、80%、90%、95%、96%、97%、98%、或99%序列同一性的RNA序列、或其片段或变体;或

-所述衍生自Ndufa1基因的3' UTR元件包括下列或由下列组成:根据SEQ ID NO:31的DNA序列、或按递增优选顺序与根据SEQ ID NO:31的核酸序列具有至少50%、60%、70%、80%、90%、95%、96%、97%、98%、或99%序列同一性的DNA序列、或其片段或变体;或根据SEQ ID NO:32的RNA序列、或按递增优选顺序与根据SEQ ID NO:32的核酸序列具有至少50%、60%、70%、80%、90%、95%、96%、97%、98%、或99%序列同一性的RNA序列、或其片段或变体。

6. 根据权利要求1至5中任一项所述的人工核酸分子,其中所述CRISPR相关蛋白质包括CRISPR相关野生型蛋白质、其同源物、变体、片段和衍生物。

7. 根据权利要求1至6中任一项所述的人工核酸分子,其中所述CRISPR相关蛋白质选自Cas9、Cpf1 (Cas12)、C2c1、C2c3、Cas13、CasX或CasY。

8. 根据权利要求1至7中任一项所述的人工核酸分子,包括编码CRISPR相关蛋白质的核酸序列的所述人工核酸包括下列或由下列组成:根据SEQ ID NO:428-441;10999-11001;442-1345中任一者的氨基酸序列、或按递增优选顺序与根据SEQ ID NO:428-441;10999-11001;442-1345中任一者的氨基酸序列具有至少50%、60%、70%、80%、90%、95%、96%、97%、98%、或99%序列同一性的氨基酸序列、或这些序列中任意者的变体或片段。

9. 根据权利要求8所述的人工核酸分子,其中所述CRISPR相关蛋白质衍生物包括至少一个其他效应结构域,所述其他效应结构域任选地选自KRAB、CSD、WRPW、VP64、p65AD和Mxi。

10. 根据权利要求1至9中任一项所述的人工核酸分子,其中所述人工核酸进一步包括编码核定位信号(NLS)的至少一种核酸序列,所述核定位信号(NLS)任选地选自包括下列或由下列组成的NLS:根据SEQ ID NO:426;427;10575;381;382;384;11957;11958-11964的氨基酸序列、或按递增优选顺序与根据SEQ ID NO:426;427;10575;381;382;384;11957;11958-11964的氨基酸序列具有至少50%、60%、70%、80%、90%、95%、96%、97%、98%、或99%序列同一性的氨基酸序列;和包括下列或由下列组成的NLS:根据SEQ ID NO:426;427;10575;381;382;384;11957;11958-11964的氨基酸序列、或按递增优选顺序与根据SEQ ID NO:426;427;10575;381;382;384;11957;11958-11964的氨基酸序列具有至少50%、60%、70%、80%、90%、95%、96%、97%、98%、或99%序列同一性的氨基酸序列;或具有按递增优选顺序与根据SEQ ID NO:12021-14274的氨基酸序列具有至少50%、60%、70%、

80%、90%、95%、96%、97%、98%、或99%序列同一性的氨基酸序列的NLS。

11. 根据权利要求1至10中任一项所述的人工核酸分子,其中所述人工核酸进一步包括编码蛋白质或肽标签的至少一种核酸序列。

12. 根据权利要求1至11中任一项所述的人工核酸分子,其中所述人工核酸分子的所述至少一个编码区包括下列或由下列组成:根据SEQ ID NO:411;2540-2553;11117-11119;11355-11357;2554-3457;1380-1393;3700-3713;4860-4873;6020-6033;7180-7193;8340-8353;11237-11239;11473-11475;11591-11593;11709-11711;11827-11829;11945-11947;1394-2297;3714-4617;4874-5777;6034-6937;7194-8097;8354-9257;412;3474-3887;2314-2327;4634-4647;5794-5807;6954-6967;8114-8127;413-425;3490-3503;3506-3519;3522-3535;3538-3551;3554-3567;3570-3583;3586-3599;3602-3615;3618-3631;3634-3647;3650-3663;3666-3679;3682-3695;9514-9527;9626-9639;9738-9751;9850-9863;9962-9975;10074-10087;10186-10199;10298-10311;2330-2343;2346-2359;2362-2375;2378-2391;2394-2407;2410-2423;2426-2439;2442-2455;2458-2471;2474-2487;2490-2503;2506-2519;2522-2535;9498-9511;9610-9623;9722-9735;9834-9847;9946-9959;10058-10071;10170-10183-10282-10295;4650-4663;4666-4679;4682-4695;4698-4711;4714-4727;4730-4743;4746-4759;4762-4775;4778-4791;4794-4807;4810-4823;4826-4839;4842-4855;9530-9543;9642-9655;9754-9767;9866-9879;9978-9991;10090-10103;10202-10215;10314-10327;5810-5823;5826-5839;5842-5855;5858-5871;5874-5887;5890-5903;5906-5919;5922-5935;5938-5951;5954-5967;5970-5983;5986-5999;6002-6015;9546-9559;9658-9671;9770-9783;9882-9895;9994-10007;10106-10119;10218-10231;10330-10343;6970-6983;6986-6999;7002-7015;7018-7031;7034-7047;7050-7063;7066-7079;7082-7095;7098-7111;7114-7127;7130-7143;7146-7159;7162-7175;9562-9575;9674-9687;9786-9799;9898-9911;10010-10023;10122-10135;10234-10247;10346-10359;8130-8143;8146-8159;8162-8175;8178-8191;8194-8207;8210-8223;8226-8239;8242-8255;8258-8271;8274-8287;8290-8302;8306-8319;8322-8335;9578-9591;9690-9703;9802-9815;9914-9927;10026-10039;10138-10151;10250-10263;10362-10375;9290-9303;9306-9319;9322-9335;9338-9351;9354-9367;9370-9383;9386-9399;9402-9415;9418-9431;9434-9447;9450-9463;9466-9479;9482-9495;9594-9607;9706-9719;9818-9831;9930-9943;10042-10055;10154-10167;10266-10279;10378-10391中任一者的核酸序列;或按递增优选顺序与所述核酸序列中任一者具有至少50%、60%、70%、80%、90%、95%、96%、97%、98%、或99%序列同一性的核酸序列。

13. 根据权利要求1至12中任一项所述的人工核酸分子,其中所述人工核酸分子包括根据SEQ ID NO:10552;3458-3459;3460-3473;2298-2299;4618-4619;5778-5779;6938-6939;8098-8099;9258-9259;2300-2313;4620-4633;5780-5793;6940-6953;8100-8113;9260-9273;3488-3489;10396;2328-2329;10395;4648-4649;10397;5808-5809;10398;6968-6969;10399;8128-8129;10400;9274-9287;3504-3505;3520-3521;3536-3537;3552-3553;3568-3669;3584-3585;3600-3601;3616-3617;3632-3633;3648-3649;3664-3665;3680-3681;3696-3697;9528-9529;9640-9641;9752-9753;9864-9865;9976-9977;10088-10089;10200-10201;10312-10313;10403;10410;10417;10424;10431;10438;10445;

10452;10459;10466;10473;10480;10487;10494;10501;10508;10515;10522;10529;  
10536;10543;2344-2345;2360-2361;2376-2377;2392-2393;2408-2409;2424-2425;2440-  
2441;2456-2457;2472-2473;2489-2490;2504-2505;2520-2521;2536-2537;9512-9513;  
9624-9625;9736-9737;9848-9849;9960-9961;10072-10073;10184-10185;10296-10297;  
10402;10409;10416;10423;10430;10437;10444;10451;10458;10465;10472;10479;  
10486;10493;10500;10507;10514;10521;10528;10535;10542;4664-4665;4680-4681;  
4696-4697;4712-4713;4728-4729;4744-4745;4760-4761;4776-4777;4792-4793;4808-  
4809;4824-4825;4840-4841;4856-4857;9544-9545;9656-9657;9768-9769;9880-9881;  
9992-9993;10104-10105;10216-10217;10328-10329;10404;10411;10418;10425;10432;  
10439;10446;10453;10460;10467;10474;10481;10488;10495;10502;10509;10516;  
10523;10530;10537;10544;5824-5825;5840-5841;5856-5857;5872-5873;5888-5889;  
5904-5905;5920-5921;5936-5937;5952-5953;5968-5969;5984-5985;6000-6001;6016-  
6017;9560-9561;9672-9673;9784-9785;9896-9897;10008-10009;10120-10121;10232-  
10233;10344-10345;10405;10412;10419;10426;10433;10440;10447;10454;10461;  
10468;10475;10482;10489;10496;10503;10510;10517;10524;10531;10538;10545;7033;  
7048-7049;7064-7065;7080-7081;7096-7097;7112-7113;7128-7129;7144-7145;7160-  
7161;7176-7177;9576-9577;9688-9689;9800-9801;9912-9913;10024-10025;10136-  
10137;10248-10249;10360-10361;10406;10413;10420;10427;10434;10441;10448;  
10455;10462;10469;10476;10483;10490;10497;10504;10511;10518;10525;10532;  
10539;10546;8144-8145;8160-8160;8176-8177;8192-8193;8208-8209;8224-8225;8240-  
8241;8256-8257;8272-8273;8288-8289;8304-8305;8320-8321;8336-8337;9592-9593;  
9704-9705;9816-9817;9928-9929;10040-10041;10152-10153;10264-10265;10376-  
10377;10407;10414;10421;10428;10435;10442;10449;10456;10463;10470;10477;  
10484;10491;10498;10505;10512;10519;10526;10533;10540;10547;9288-9289;10401;  
10553;10582-10583;10579-10580;10585-10586;10588-10589;10591-10592;10594-10595;  
10597-10598;10554-10574;10601;10602;10615;10616;10629;10630;10643;10644;  
10657;10658;10671;10672;10685;10686;10699;10700;10713;10714;10727;10728;  
10741;10742;10755;10756;10769;10770;10783;10784;10797;10798;10811;10812;  
10825;10826;10839;10840;10853;10854;10867;10868;10881;10882;10603;10604;  
10617;10618;10631;10632;10645;10646;10659;10660;10673;10674;10687;10688;  
10701;10702;10715;10716;10729;10730;10743;10744;10757;10758;10771;10772;  
10785;10786;10799;10800;10813;10814;10827;10828;10841;10842;10855;10856;  
10869;10870;10883;10884;10605;10606;10619;10620;10633;10634;10647;10648;  
10661;10662;10675;10676;10689;10690;10703;10704;10717;10718;10731;10732;  
10745;10746;10759;10760;10773;10774;10787;10788;10801;10802;10815;10816;  
10829;10830;10843;10844;10857;10858;10871;10872;10885;10886;10607;10608;  
10621;10622;10635;10636;10649;10650;10663;10664;10677;10678;10691;10692;  
10705;10706;10719;10720;10733;10734;10747;10748;10761;10762;10775;10776;  
10789;10790;10803;10804;10817;10818;10831;10832;10845;10846;10859;10860;

10873;10874;10887;10888;10609;10610;10623;10624;10637;10638;10651;10652;10665;10666;10679;10680;10693;10694;10707;10708;10721;10722;10735;10736;10749;10750;10763;10764;10777;10778;10791;10792;10805;10806;10819;10820;10833;10834;10847;10848;10861;10862;10875;10876;10889;10890;10611;10612;10625;10626;10639;10640;10653;10654;10667;10668;10681;10682;10695;10696;10709;10710;10723;10724;10737;10738;10751;10752;10765;10766;10779;10780;10793;10794;10807;10808;10821;10822;10835;10836;10849;10850;10863;10864;10877;10878;10891;10892;9304-9305;9320-9321;9336-9337;9352-9353;9368-9369;9384-9385;9400-9401;9416-9417;9432-9433;9448-9449;9464-9465;9480-9481;9496-9497;9608-9609;9720-9721;9832-9833;9944-9945;10056-10057;10168-10169;10280-10281;10392-10393;10408;10415;10422;10429;10436;10443;10450;10457;10464;10471;10478;10485;10492;10499;10506;10513;10520;10527;10534;10541;10548中任一者的核酸序列、或按递增优选顺序与所述核酸序列中任一者具有至少50%、60%、70%、80%、90%、95%、96%、97%、98%、或99%序列同一性的核酸序列。

14. 根据权利要求1至13中任一项所述的人工核酸分子,其中所述人工核酸分子包括下列或由下列组成:根据SEQ ID NO:11011-11042;11249-11280;11131-11162;11367-11398;11485-11516;11603-11634;11721-11752;11839-11870;11044-11116;11282-11354;11164-11236;11400-11472;11518-11590;11636-11708;11754-11826;11872-11944;11011-11042;11249-11280;11044-11116;11282-11354;11131-11162;11367-11398;11485-11516;11603-11634;11721-11752;11839-11870;11164-11236;11400-11472;11518-11590;11636-11708;11754-11826;11872-11944;11120-11122;11240;11241;11358;11359;11476;11477;11594;11595;11712;11713;11830;11831;11948;11949;11123-11130;11360-11366;11242-11248;11478-11484;11596-11602;11714-11720;11832-11838;11950-11956中任一者的核酸序列、或按递增优选顺序与所述核酸序列中任一者具有至少50%、60%、70%、80%、90%、95%、96%、97%、98%、或99%序列同一性的核酸序列。

15. 根据权利要求1至14中任一项所述的人工核酸分子,其中所述人工核酸分子是RNA。

16. 根据权利要求15所述的RNA,其中所述RNA是单顺反子、双顺反子、或多顺反子的。

17. 根据权利要求14或15所述的RNA,其中所述RNA是mRNA、病毒RNA或复制子RNA。

18. 根据权利要求1至17中任一项所述的人工核酸,优选地RNA,其中所述人工核酸是修饰核酸,优选地稳定化的核酸。

19. 根据权利要求1至18中任一项所述的人工核酸,优选地RNA,其中

-所述人工核酸的所述至少一个编码区的G/C含量与相应野生型人工核酸的相应编码序列的G/C含量相比增加,和/或其中

-所述人工核酸的所述至少一个编码区的C含量与相应野生型人工核酸的相应编码序列的C含量相比增加,和/或其中

-所述人工核酸的所述至少一个编码区的密码子适于密码子使用,其中在所述人工核酸的所述至少一种编码序列中密码子适应指数(CAI)优选增加或最大化,

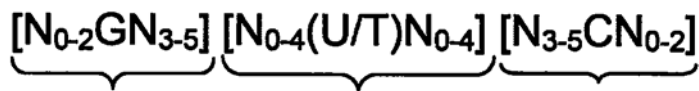
-其中所述人工核酸编码的氨基酸序列优选地与相应野生型人工核酸编码的氨基酸序列相比不被修饰。

20. 根据权利要求1至19中任一项所述的人工核酸, 优选地RNA, 其包括5' -帽结构, 优选地m7GpppN或帽1。

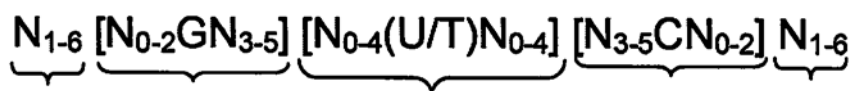
21. 根据权利要求1至20中任一项所述的人工核酸, 优选地RNA, 其包括至少一种组蛋白茎-环。

22. 根据权利要求21所述的人工核酸, 优选地RNA, 其中所述至少一种组蛋白茎-环包括根据下列式(I)或(II)的核酸序列:

式(I) (茎-环序列, 无茎边界元件):



式(II) (茎-环序列, 有茎边界元件):



茎 1

茎 1

环

茎 2

茎 2

边界元件

边界元件

其中:

茎1或茎2边界元件 $N_{1-6}$ 是1至6个, 优选2至6个, 更优选2至5个, 甚至更优选3至5个, 最优选4至5个或5个N的连续序列, 其中各N彼此独立地选自以下核苷酸: 选自A、U、T、G和C、或其核苷酸类似物;

茎1 $[N_{0-2}GN_{3-5}]$ 与元件茎2反向互补或部分反向互补, 并且是5至7个核苷酸的连续序列;

其中 $N_{0-2}$ 是0至2个, 优选0至1个, 更优选1个N的连续序列, 其中各N彼此独立地选自以下核苷酸: 选自A、U、T、G和C或其核苷酸类似物;

其中 $N_{3-5}$ 是3至5个, 优选4至5个, 更优选4个N的连续序列, 其中各N彼此独立地选自以下核苷酸: 选自A、U、T、G和C或其核苷酸类似物, 并且

其中G是鸟苷或其类似物, 并且可任选地被胞苷或其类似物替代, 条件是茎2中其互补核苷酸胞苷被鸟苷替代;

环序列 $[N_{0-4}(U/T)N_{0-4}]$ 位于元件茎1和茎2之间, 并且是3至5个核苷酸, 更优选4个核苷酸的连续序列;

其中各 $N_{0-4}$ 彼此独立地是0至4个, 优选1至3个, 更优选1至2个N的连续序列, 其中各N彼此独立地选自以下核苷酸: 选自A、U、T、G和C或其核苷酸类似物; 并且

其中U/T表示尿苷、或任选地胸苷;

茎2 $[N_{3-5}CN_{0-2}]$ 与元件茎1反向互补或部分反向互补, 并且是5至7个核苷酸的连续序列;

其中 $N_{3-5}$ 是3至5个, 优选4至5个, 更优选4个N的连续序列, 其中各N彼此独立地选自以下核苷酸: 选自A、U、T、G和C或其核苷酸类似物;

其中 $N_{0-2}$ 是0至2个, 优选0至1个, 更优选1个N的连续序列, 其中各N彼此独立地选自以下核苷酸: 选自A、U、T、G和C或其核苷酸类似物; 并且

其中C是胞苷或其类似物, 并且可任选地被鸟苷或其类似物替代, 条件是茎1中其互补



29. 根据权利要求28所述的组合物,其中所述人工核酸分子,优选地RNA,与一种或多种阳离子或聚阳离子化合物,优选地与阳离子或聚阳离子聚合物、阳离子或聚阳离子肽或蛋白质,例如鱼精蛋白、阳离子或聚阳离子多糖和/或阳离子或聚阳离子脂质复合。

30. 根据权利要求29所述的组合物,其中所述人工核酸分子,优选地RNA,与所述一种或多种阳离子或聚阳离子肽或蛋白质的N/P比在约0.1至10的范围内,包括约0.3至4、约0.5至2、约0.7至2和约0.7至1.5的范围。

31. 根据权利要求28至30中任一项所述的组合物,其中所述人工核酸分子,优选地RNA,与一种或多种脂质复合,从而形成脂质体、脂质纳米颗粒和/或脂质复合物。

32. 根据权利要求28至31中任一项所述的组合物,进一步包括至少一种向导RNA (gRNA) 或编码其的核酸,所述gRNA能够将CRISPR相关蛋白质靶向至目的靶DNA序列、或其可操作地连接的调控元件。

33. 试剂盒,优选地多部分试剂盒,其包括根据权利要求1至27中任一项所述的人工核酸分子,优选地RNA,或根据权利要求28至32中任一项所述的组合物,和任选地液体媒介和/或任选地具有关于所述人工核酸分子或所述组合物的给予和剂量的信息的技术说明书。

34. 根据权利要求33所述的试剂盒,其中所述试剂盒包含林格-乳酸盐溶液作为一部分。

35. 根据权利要求33或34所述的试剂盒,进一步包括向导RNA (gRNA) 或编码其的核酸,所述gRNA能够将CRISPR相关蛋白质靶向至目的靶DNA序列、或其可操作地连接的调控元件。

36. 根据权利要求1至27中任一项所述的人工核酸分子,优选地RNA;根据权利要求28至32中任一项所述的组合物;或根据权利要求33至35所述的试剂盒,其用作药物。

37. 根据权利要求1至27中任一项所述的人工核酸分子,优选地RNA;根据权利要求28至32中任一项所述的组合物;或根据权利要求33至35所述的试剂盒,其用于基因疗法。

38. 根据权利要求1至27中任一项所述的人工核酸分子,优选地RNA;根据权利要求28至32中任一项所述的组合物;或根据权利要求33至35所述的试剂盒,其用于调控目的基因的表达的方法,所述方法包括向有此需要的患者给予 (a) 所述人工核酸分子,优选地RNA,所述组合物或所述试剂盒;和 (b) 向导RNA (gRNA) 或编码其的核酸,所述sgRNA能够将CRISPR相关蛋白质靶向至目的基因、或其可操作地连接的调控元件。

39. 根据权利要求1至27中任一项所述的人工核酸分子,优选地RNA;根据权利要求28至32中任一项所述的组合物;或根据权利要求33至35所述的试剂盒,其用作适于通过表达所述至少一种编码序列编码的CRISPR相关蛋白质进行治疗的疾病、障碍或状况的药物或用于适于通过表达所述至少一种编码序列编码的CRISPR相关蛋白质进行治疗的疾病、障碍或状况的基因疗法。

40. 根据权利要求1至27中任一项所述的人工核酸分子,优选地RNA;根据权利要求28至32中任一项所述的组合物;或根据权利要求33至35所述的试剂盒,其用作通过敲入、敲除或操纵目的基因、或通过调控目的基因的表达可处理的疾病、障碍或状况的药物或用于通过敲入、敲除或操纵目的基因、或通过调控目的基因的表达可处理的疾病、障碍或状况的基因疗法。

41. 根据权利要求40所述的用途的人工核酸分子,优选地RNA,组合物或试剂盒,其中所述疾病、障碍或状况选自遗传疾病、癌症、自身免疫疾病、炎性疾病和感染性疾病。



42. 根据权利要求1至27中任一项所述的人工核酸分子, 优选地RNA; 根据权利要求28至32中任一项所述的组合物; 或根据权利要求33至35所述的试剂盒用于增加编码的所述CRISPR相关蛋白质的表达的用途, 任选地在基因疗法中。

43. 根据权利要求1至27中任一项所述的人工核酸分子, 优选地RNA; 根据权利要求28至32中任一项所述的组合物; 或根据权利要求33至35所述的试剂盒用于调控编码的所述CRISPR相关蛋白质所靶向的目的基因的表达的用途。

44. 用于调控目的基因的表达的方法, 包括以下步骤:

- a) 提供根据权利要求1至27中任一项所述的人工核酸分子, 优选地RNA;
- b) 提供向导RNA (gRNA) 或编码其的核酸, 所述gRNA能够将CRISPR相关蛋白质靶向至目的靶DNA序列、或其可操作地连接的调控元件,
- c) 使细胞、组织或生物体与所述人工核酸分子, 优选地RNA, 和所述gRNA或编码其的核酸在适于调控所述目的基因的表达效力的条件下接触。

45. 治疗或预防障碍的方法, 其中所述方法包括向有此需要的对象给予有效量的根据权利要求1至27中任一项所述的人工核酸分子, 优选地RNA, 根据权利要求28至32中任一项所述的组合物, 或根据权利要求33至35所述的试剂盒; 和向导RNA (gRNA) 或编码其的核酸, 所述gRNA能够将CRISPR相关蛋白质靶向至目的靶DNA序列、或其可操作地连接的调控元件。

46. 根据权利要求45所述的方法, 其中所述障碍是适于通过表达编码的所述CRISPR相关蛋白质进行治疗, 优选地适于通过调控所述CRISPR相关蛋白质所靶向的目的基因的表达进行治疗的疾病、障碍或状况。

47. 根据权利要求45或46所述的方法, 其中所述障碍是通过敲入、敲除或通过突变目的基因、或通过改变目的基因的表达可处理的疾病、障碍或状况。

48. 增加包括编码CRISPR相关蛋白质的编码区的人工核酸分子, 优选地RNA, 的表达效力的方法, 所述方法包括

(a) 使所述编码区与至少一个5' UTR元件缔合, 所述5' UTR元件衍生自选自ATP5A1、RPL32、HSD17B4、SLC7A3、NOSIP、或NDUFA4的基因的5' UTR、或其相应RNA序列、同源物、片段或变体;

(b) 使所述编码区与至少一个3' UTR元件缔合, 所述3' UTR元件衍生自选自GNAS、CASP1、PSMB3、ALB、或RPS9的基因的3' UTR、或其相应RNA序列、同源物、片段或变体; 和

(c) 获得根据权利要求1至47中任一项所述的人工核酸分子, 优选地RNA。

## 编码CRISPR相关蛋白质的核酸及其用途

[0001] 本发明涉及编码CRISPR相关蛋白质的人工核酸——具体地RNA——和包括其的(药物)组合和多部分试剂盒(kit-of-parts)。所述人工核酸——具体地RNA——(药物)组合和试剂盒设想用于药物等,例如用于基因疗法,和具体地用于治疗 and/或预防适于用CRISPR相关蛋白质(例如通过基因编辑、敲入、敲除或调控目的靶基因表达)治疗的疾病。

[0002] CRISPR(成簇规律间隔的短回文重复序列)-Cas系统通过使用反义RNA以序列特异性方式识别和切割外源DNA,赋予细菌和古细菌针对入侵DNA元件(例如病毒,质粒)的适应性免疫保护。在最新的分类中,各种CRISPR-Cas系统根据其效应物的构型被分为两类:1类CRISPR系统利用几种Cas(CRISPR相关)蛋白和CRISPR-RNA(crRNA)作为向导RNA(gRNA)形成效应复合物,而2类CRISPR系统利用大的单组分Cas蛋白与crRNA一起介导对外源DNA元件的干扰。已经鉴定了多个1类CRISPR-Cas系统——其包括I型和III型系统,并且对其进行了详细功能表征。迄今已鉴定和实验表征的大多数2类CRISPR-Cas系统利用Cas9家族的同源RNA向导型核酸内切酶作为效应子——其充当多结构域核酸内切酶,连同crRNA和反式激活crRNA(tracrRNA),或者连同合成的单向导RNA(sgRNA),以切割侵入靶DNA的两条链(Sander and Joung, Nat Biotechnol. 2014年4月; 32(4): 347-355; Boettcher and McManus Mol Cell. 2015年5月21日; 58(4): 575-585)。

[0003] 天然CRISPR/Cas9 II型系统基本上分三个步骤发挥作用。在暴露于外源DNA后,短外源DNA序列(前间隔序列(protospacer))被并入细菌基因组中CRISPR基因座中的短回文序列之间。前间隔序列附近的小段保守核苷酸(前间隔序列相邻基序(PAM))被用于获得前间隔序列(获得或适应阶段)。随后,宿主原核生物转录和加工CRISPR基因座以产生成熟CRISPR RNA(crRNA),该成熟CRISPR RNA(crRNA)含有CRISPR重复元件和整合的外源DNA间隔遗传区段——相应于此前非自身的DNA元件,连同反式激活CRISPR RNA(tracrRNA)(表达或成熟步骤)。最后,crRNA和Cas9与tracrRNA结合产生crRNA:tracrRNA:Cas9复合物,其与入侵DNA中的互补序列结合。Cas9核酸内切酶然后将DNA双链断裂(DSB)引入靶DNA(干扰阶段)(Sander and Joung, Nat Biotechnol. 2014年4月; 32(4): 347-355)。

[0004] 哺乳动物细胞通过非同源末端连接法(NHEJ)或同源定向修复(HDR)而响应DSB。NHEJ可引入小段核苷酸碱基的随机插入或删除,导致基因突变和功能丧失效应。在HDR中,具有与DNA双链断裂两侧侧翼序列具有同源性的区域的DNA区段的引入将导致宿主细胞机构进行的修复(Sander and Joung, Nat Biotechnol. 2014年4月; 32(4): 347-355)。

[0005] 最近在几个细菌基因组中鉴定了第二个推定的2类CRISPR系统,暂指定为V型。推定的V型CRISPR-Cas系统包含大的~1300个氨基酸的蛋白质,称为Cpf1或Cas12(来自普氏菌(Prevotella)、弗朗西斯氏菌1(Francisella 1)、氨基酸球菌(Acidaminococcus)属BV3L6(AsCpf1)和毛螺菌科(Lachnospiraceae)细菌ND2006(LbCpf1)的CRISPR)。Cpf1只需一个短crRNA识别和结合其靶DNA序列,代替Cas9的~100-nt向导RNA(crRNA和tracrRNA)。即,Cpf1通常显示单个42nt,其具有在其3'端的与靶DNA序列的前间隔序列互补的23nt,前间隔序列的TTTN PAM 5',并产生作为与spCas9的钝端相比DSB 5'突出端(overhangs)。

[0006] Cpf1有效地切割靶DNA,其通过短富T的前间隔序列相邻基序(PAM)进行,与Cas9系

统的靶DNA后富G PAM形成对比。第三,Cpf1引入交错的DNA双链断裂,具有4nt或5nt的5'突出端(Zetsche et al.Cell.201510月22日;163(3):759-771)。Cpf1在人细胞中的靶标效率是与spCas9相当的,并且Cpf1显示没有或减少的脱靶切割。

[0007] 自从CRISPR/Cas系统在哺乳动物基因组中应用以来,该技术迅速发展:不呈现核酸内切酶活性的无催化活性或“死亡”的Cas9(dCas9)可以被适当的gRNA招募,以靶向目的DNA序列。这种Cas蛋白及其变体和衍生物被特别关注作为多用途、序列特异性和非诱变基因调控工具。例如,适当的gRNA可用于靶向dCas9衍生物,其中转录抑制或激活结构域靶向基因,导致转录抑制(称为CRISPR干扰,CRISPRi)或激活(称为CRISPR激活,CRISPRa)。

[0008] 通过这些连续的创新,CRISPR-Cas系统已经广泛适用于基因组工程。CRISPR-Cas系统是多用途的且容易定制,因为针对目的靶基因的gRNA可以容易制备,而Cas蛋白不需要任何修饰。通过引入数种gRNA(“多元复合”)可以容易靶向多个基因座。

[0009] CRISPR/Cas系统已经被成功地用作细菌和真核生物中基因组编辑和转录激活/抑制的牢靠的、多用途的和精确的工具,并且已经引发了用于研究和治疗目的的有前景的新方法的发展。然而,尽管其具有众多优点,CRISPR/Cas系统的应用通常受到Cas蛋白表达不良的阻碍。

[0010] 本发明的一个目的是满足这些需要并提供改进的治疗方法,用于治疗癌症、感染性疾病和本文限定的其他疾病和状况。本发明蕴涵的目的通过所请求保护的主题来解决。

[0011] 尽管下文详细描述了本发明,但应理解本发明不限于本文所述的具体方法、方案和试剂,因为这些可以变化。还应理解,本文使用的术语不旨在限制本发明的范围,本发明的范围仅受所附权利要求的限制。除非另外限定,本文使用的所有技术和科学术语均与本领域普通技术人员的通常理解具有相同的含义。

[0012] 在下文中,将描述本发明的要素。这些要素随具体实施方式一起列出,然而,应该理解,其可以以任何方式和任何数量组合以产生另外的实施方式。不应将各种描述的实例和优选实施方式解释为将本发明仅限于明确描述的实施方式。这种描述应被理解为支持和包含将明确描述的实施方式与任何数量的公开和/或优选要素组合的实施方式。此外,除非上下文另有说明,本申请中的所有所述要素的任何排列和组合应该被认为已被本申请的描述公开。

[0013] 贯穿说明书和所附权利要求书,除非上下文另有要求,术语“包括”和诸如“含有”和“包含”的变型将被理解为暗示包括所述的成员、整数或步骤,但不排除任何其他未提及的成员、整数或步骤。术语“由.....组成”是术语“包括”的特定实施方式,其中排除任何其他未提及的成员、整数或步骤。在本发明的环境中,术语“包括”包括术语“由.....组成”。因此,术语“包括”包括“包含”以及“由.....组成”,例如,“包括”X的组合物可以仅由X组成,或者可以包括另外的一些,例如X+Y。

[0014] 除非在本文中另有说明或明显与上下文相矛盾,在描述本发明的环境(特别是在权利要求的环境中)中使用的术语“一个”和“一种”和“所述”以及类似的用词应被解释为同时涵盖单数和复数。本文中对数值范围的描述仅旨在充当分别提及落入该范围内的每个单独值的简写法。除非本文另有说明,每个单独的值被并入说明书中,如同其在本文中被单独描述一样。说明书中的任何语言都不应被解释为暗示任何未主张的要素对于本发明的实践必不可少。

[0015] “基本上”一词不排除“完全”，例如，“基本上不含”Y的组合物可以完全不含Y。必要时，可以从本发明的限定中省略“基本上”一词。

[0016] 与数值x相关的术语“约”表示 $x \pm 10\%$ 。

[0017] 在本发明中，如果没有另外说明，替代方式和实施方式的不同特征可以彼此组合。

[0018] 为清楚和可读起见，提供了以下定义。这些定义提及的任何技术特征可以在本发明的各实施方式上读到。另外的定义和解释可在这些实施方式的环境中被具体提供。

[0019] 定义

[0020] 人工核酸分子：人工核酸分子可一般被理解为非天然存在的核酸分子，例如DNA或RNA。换句话说，人工核酸分子可被理解为非天然核酸分子。这种核酸分子可以因其个体序列（非天然存在）和/或因非天然存在的其他修饰，例如核苷酸的结构修饰，而是非天然的。人工核酸分子可以是DNA分子、RNA分子或包括DNA和RNA部分的杂种分子。一般，人工核酸分子可通过基因工程方法被设计和/或生成以相应于期望的人工核苷酸序列（异源序列）。在此环境中，人工序列通常是可以非天然存在的序列，即其与野生型序列存在至少一种核苷酸的区别。术语“野生型”可被理解为自然界中存在的序列。进一步，术语“人工核酸分子”不限于意为“单一一个分子”，而一般被理解为包括相同的分子的集合。因此，其可涉及等份中包含的多个相同的分子。

[0021] DNA：DNA是脱氧核糖核酸的常规缩写。其是一种核酸分子，即由核苷酸组成的聚合物。这些核苷酸通常是脱氧-腺苷-一磷酸、脱氧-胸苷-一磷酸、脱氧-鸟苷-一磷酸和脱氧-胞苷-一磷酸单体，该单体本身由糖部分（脱氧核糖）、碱基部分和磷酸部分组成，并且通过特征性骨架结构聚合。该骨架结构一般通过第一相邻单体的核苷酸糖部分（即脱氧核糖）和第二个相邻单体的磷酸部分之间的磷酸二酯键形成。单体的具体相继，即连接至糖/磷酸骨架的碱基的顺序称为DNA序列。DNA可以是单链或双链的。在双链形式中，第一链的核苷酸一般与第二链的核苷酸杂交，例如通过A/T碱基配对和G/C碱基配对。

[0022] 异源序列：如果两个序列不是可衍生自相同基因的，则其一般被理解是“异源的”。即，尽管异源序列可以是可衍生自相同生物体的，但其不天然地（在自然界中）存在于相同的核酸分子中，如在相同的mRNA中。

[0023] 克隆位点：克隆位点一般被理解为适于插入核酸序列（例如包含开放阅读框的核酸序列）的核酸分子区段。插入可以通过本领域技术人员已知的任何分子生物学方法进行，例如通过限制和连接。克隆位点一般包含一种或多种限制酶识别位点（限制位点）。这一种或多种限制性位点可以被在这些位点切割DNA的限制酶识别。包含多于一个限制位点的克隆位点也可称为多克隆位点（MCS）或多接头。

[0024] 核酸分子：核酸分子是包含核酸组分（优选由核酸组分组成）的分子。术语核酸分子优选指代DNA或RNA分子。其优选与术语“多核苷酸”同义使用。优选地，核酸分子是包含核苷酸单体或由核苷酸单体组成的聚合物，该核苷酸单体通过糖/磷酸骨架的磷酸二酯键彼此共价连接。术语“核酸分子”还包括修饰的核酸分子，如碱基修饰、糖修饰或骨架修饰等的DNA或RNA分子。

[0025] 开放阅读框：在本发明的环境中，开放阅读框（ORF）一般可以是几个核苷酸三联体的序列，其可以翻译成肽或蛋白质。开放阅读框优选含有在其5'末端的起始密码子——即通常编码氨基酸甲硫氨酸（ATG）的三个后续核苷酸的组合，和后续区域——其通常呈现3核

核苷酸的倍数长度。ORF优选通过终止密码子(例如,TAA、TAG、TGA)终止。一般,这是开放阅读框架的唯一终止密码子。因此,在本发明环境中的开放阅读框优选是这样的核苷酸序列:由可被3整除的数量的核苷酸组成,起始于起始密码子(例如ATG),并且优选终止于终止密码子(例如,TAA、TGA或TAG)。开放阅读框可被分离,或者其可以被并入较长的核酸序列中,例如载体或mRNA中。开放阅读框也可称为“(蛋白质)编码序列”,或优选地“编码序列”。

[0026] 肽:肽或多肽一般是通过肽键连接的氨基酸单体的聚合物。其一般含有少于50个单体单元。然而,术语肽不是不主张具有多于50个单体单元分子。长肽也称为多肽,一般具有50至600个单体单元。

[0027] 蛋白质:蛋白质一般包含一种或多种肽或多肽。蛋白质一般折叠成三维形式,这可能是蛋白质发挥其生物学功能所需的。

[0028] 限制位点:限制位点,也称为限制酶识别位点,是由限制酶识别的核苷酸序列。限制位点一般是短的,优选回文的,核苷酸序列,例如包含4至8个核苷酸的序列。限制位点优选被限制酶特异性识别。限制酶一般在限制位点处切割包含此位点的核苷酸序列。在双链核苷酸序列如双链DNA序列中,限制酶一般切割核苷酸序列的两条链。

[0029] RNA,mRNA:RNA是核糖核酸的常规缩写。其是一种核酸分子,即由核苷酸组成的聚合物。这些核苷酸通常是腺苷-一磷酸、尿苷-一磷酸、鸟苷-一磷酸和胞苷-一磷酸单体,该单体沿着所谓的骨架相互连接。该骨架通过第一相邻单体的糖(即核糖)与第二相邻单体的磷酸酯部分之间的磷酸二酯键形成。单体的具体序列称为RNA序列。通常,RNA可以是可通过DNA序列的转录获得的,例如在细胞内。在真核细胞中,转录一般在细胞核或线粒体内进行。在体内,DNA的转录通常导致所谓的成熟前RNA,其必须被加工成所谓的信使RNA,通常缩写为mRNA。成熟前RNA的加工——例如在真核生物中——包括多种不同的转录后修饰,如剪接、5'-加帽、多腺苷酸化、从细胞核或线粒体输出等。这些过程的总和也称为RNA的成熟。成熟的信使RNA通常提供可以被翻译成具体肽或蛋白质的氨基酸序列的核苷酸序列。一般,成熟的mRNA包含5'-帽、5'-UTR、开放阅读框、3'-UTR和聚(A)序列。除了信使RNA之外,还存在几种非编码类型的RNA,其可能涉及转录和/或翻译的调控。

[0030] 核酸分子的序列:核酸分子的序列一般被理解为具体的和个体的顺序,即其核苷酸的相继。蛋白质或肽的序列一般被理解为顺序,即其氨基酸的相继。

[0031] 序列同一性:如果两个或更多个序列呈现相同的核苷酸或氨基酸长度和顺序,则其是相同的。同一性百分比一般描述两个序列的相同程度,即其一般描述核苷酸在其序列位置相应于参考序列的相同核苷酸的百分比。为了确定同一性程度(“%同一性”),一般认为待比较序列呈现相同的长度,即待比较序列中最长序列的长度。这意味着由8个核苷酸组成的第一序列与由包含第一序列的10个核苷酸组成的第二序列80%相同。换句话说,在本发明的环境中,序列的同一性优选地涉及序列中的、在具有相同长度的两个或更多个序列中具有相同的位置的核苷酸或氨基酸的百分比。具体地,两个氨基酸序列或两个核酸序列的“%同一性”可以通过如下确定:以最优比较为目的比对序列(例如,可以在任一序列中引入空位以与另一序列最优比对)和比较相应位置的氨基酸或核苷酸。空位通常被视为不相同的位置,而无论其在比对中的实际位置。“最优比对”一般是导致最高同一性百分比的两个序列的比对。同一性百分比通过所比较的序列中相同核苷酸的数量来确定(即,%同一性=相同位置数/总位置数×100)。可以利用本领域技术人员已知的数学算法完成两个序列

之间同一性百分比的确定。

[0032] 稳定化的核酸分子: 稳定化的核酸分子是这样的核酸分子 (优选DNA或RNA分子): 经修饰使得其与未经修饰的核酸分子相比对于崩解或降解 (例如通过环境因子或酶促消化, 如通过外切核酸酶或内切核酸酶降解) 更稳定。优选地, 在本发明的环境中, 稳定化的核酸分子在细胞中稳定化, 如原核或真核细胞, 优选哺乳动物细胞, 如人细胞。稳定化效果也可以在细胞外施加, 例如在缓冲溶液等中, 例如在包含稳定化核酸分子的药物组合物的制备方法中。

[0033] 转染: 术语“转染”是指将核酸分子 (如DNA或RNA (例如mRNA) 分子) 引入细胞, 优选引入真核细胞。在本发明的环境中, 术语“转染”包括本领域技术人员已知的用于将核酸分子引入细胞, 优选引入真核细胞, 如引入哺乳动物细胞的任何方法。这种方法包括, 例如, 电穿孔, 脂质转染, 例如, 基于阳离子脂质和/或脂质体、磷酸钙沉淀、基于纳米颗粒的转染、基于病毒的转染、或基于阳离子聚合物如DEAE-葡聚糖或聚乙烯亚胺等的转染。优选地, 所述引入是非病毒的。

[0034] 载体: 术语“载体”是指一种核酸分子, 优选人工核酸分子。本发明上下文中的载体适于包含或含有期望的核酸序列, 如包括开放阅读框的核酸序列。这种载体可以是储存载体、表达载体、克隆载体、转移载体等。储存载体是实现核酸分子例如mRNA分子方便储存的载体。因此, 该载体可包含相应于例如期望的mRNA序列或其部分的序列, 如相应于mRNA的编码序列和3'-UTR的序列。表达载体可用于产生表达产物, 如RNA, 例如mRNA, 或肽、多肽或蛋白质。例如, 表达载体可以包含转录该载体的序列段所需的序列, 如启动子序列, 例如RNA聚合酶启动子序列。克隆载体一般是含有克隆位点的载体, 该克隆位点可用于将核酸序列并入载体中。克隆载体可以是例如质粒载体或噬菌体载体。转移载体可以是适合于将核酸分子转移到细胞或生物体的载体, 例如病毒载体。本发明上下文中的载体可以是例如RNA载体或DNA载体。优选地, 载体是DNA分子。优选地, 本申请意义上的载体包含克隆位点、选择标记如抗生素抗性因子、和适于载体增殖的序列如复制起点。

[0035] 媒介: 载体一般被理解为适合于储存、运输和/或给予化合物 (如药物活性化合物) 的材料。例如, 其可以是适于储存、运输和/或给予药物活性化合物的生理可接受的液体。

[0036] 本发明部分基于如下令人惊讶的发现: 具体的3'和/或5'UTR元件可介导编码序列 (特别是编码CRISPR相关 (Cas) 蛋白如Cas9或Cpf1的那些) 的表达增加。本发明人具体地发现3'和5'UTR元件的某些组合对于提供期望的表达模式和表达蛋白量是特别有利的。具体地, 对于许多应用可能需要短时间 (约24小时, “脉冲表达”) 进行高Cas蛋白表达, 例如为了最小化基因组DNA的暴露以减少脱靶效应 (即对任何一种或多种靶标、基因或细胞转录物的任何非预期效果)。在编码CRISPR相关蛋白质的人工核酸中这种3'和5'UTR元件的协同作用在体外或体内需要瞬时表达大量这种蛋白质时特别有利。因此, 这种人工核酸适于各种如下治疗应用等: 适于通过引入突变、基因敲除或敲入、或调控目的基因的表达来进行治疗。

[0037] 因此, 在第一方面, 本发明因此涉及人工核酸分子, 其包含a. 至少一个编码区, 编码至少一种CRISPR相关蛋白质; b. 至少一个5'非翻译区 (5'UTR) 元件, 衍生自选自ATP5A1、RPL32、HSD17B4、SLC7A3、NOSIP和NDUFA4的基因的5'UTR; 和c. 至少一个3'非翻译区 (3'UTR) 元件, 衍生自选自GNAS、CASP1、PSMB3、ALB和RPS9的基因的3'UTR。

[0038] 术语“UTR”是指如本文限定的人工核酸的编码序列侧翼的“非翻译区”。在本文中,

“UTR元件”包含下列核酸序列或由其组成：衍生自具体基因的（天然存在的，野生型）UTR的核酸序列，优选如本文示例。

[0039] 当提及“衍生自”具体UTR的UTR元件时，所说的是与所述UTR（“母体UTR”）或所述UTR的同源物、变体或片段的序列相应的核酸序列。该术语包括相应于所述UTR的完整（全长）野生型序列、或其同源物、变体或片段的序列，包括全长同源物和变体，以及所述全长野生型序列的片段、同源物和变体，以及所述片段的变体。术语“相应于或与……相应”意为衍生自“母体UTR”的核酸序列可以是RNA序列（例如，等于用于限定所述母体UTR序列的RNA序列）、或相应于这种RNA序列的DNA序列（有义和反义链，成熟和未成熟）。

[0040] 当提及衍生自基因的UTR的“UTR元件”或“其同源物、片段或变体”时，表述“或其同源物、片段或变体”可以指该基因、或UTR、或两者。

[0041] 在基因（或其衍生或由所述基因包含的核酸序列，如UTR）的上下文中的术语“同源物”是指通过从共同的祖先DNA序列下降（descent）而涉及第二基因（或这种核酸序列）的基因（或其衍生或由所述基因包含的核酸序列）。术语“同源物”包括通过物种形成事件而分离的基因（“直系同源物”）和通过遗传复制事件而分离的基因（“旁系同源物”）。

[0042] 在基因的核酸序列的上下文中，术语“变体”是指核酸序列变体，即包含与参考（或“母体”）核酸或基因的参考（或“母体”）核酸序列有至少一种核酸的区别的核酸序列的核酸序列或基因。因此，变体核酸或基因可优选与其对应的参考序列相比在其核酸序列中包含至少一种突变、取代、插入或缺失（删除，deletion）。优选地，如本文所用的术语“变体”包括核酸序列或基因的天然存在变体和工程变体。因此，如本文所定义的“变体”可以衍生自，分离自，涉及，基于参考核酸序列或与参考核酸序列同源。“变体”可以优选与对应的天然存在的（野生型）核酸序列或基因、或其同源物、片段或衍生物的核酸序列具有至少5%、10%、20%、30%、40%、50%、60%、70%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、或99%的序列同一性，优选至少70%，更优选至少80%，甚至更优选至少85%，甚至更优选至少90%，最优选至少95%或甚至97%。

[0043] 在核酸序列或基因的上下文中，术语“片段”是指全长参考（或“母体”）核酸序列或基因的连续子序列。换句话说，“片段”一般可以是全长核酸序列或基因的较短部分。因此，片段一般由与全长核酸序列或基因内的相应段相同的序列组成。该术语包括天然存在片段和工程片段。在本发明的环境中，优选的序列片段由连续一段核酸组成，所述连续一段核酸相应于该片段的来源核酸或基因中的连续一段实体，其表示该片段的来源总（全长）核酸序列或基因的至少20%，优选至少30%，更优选至少40%，更优选至少50%，甚至更优选至少60%，甚至更优选至少70%，并且最优选至少80%。关于这种片段所述的序列同一性优选涉及整个核酸序列或基因。优选地，“片段”可包含具有与其来源参考核酸序列或基因具有至少5%、10%、20%、30%、40%、50%、60%、70%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、或99%，优选至少70%，更优选至少80%，甚至更优选至少85%，甚至更优选至少90%，并且最优选至少95%或甚至97%的序列同一性的核酸序列。

[0044] 在本发明的环境中使用的UTR元件优选是功能性的，即能够引发与其来源的天然存在的（野生型）UTR相同的期望的生物学效应，即特别是控制（即，调控，优选地增强）可操作地连接的编码序列的表达。本文所用的术语“可操作地连接”是指“与编码序列处于功能

关系”。本文限定的UTR元件优选与本发明的人工核酸的编码序列可操作地连接,即处于功能关系,优选以允许其控制(即调控,优选地增强)所述编码序列的表达的方式。本文使用的术语“表达”总体上包括蛋白质生物合成的所有步骤,尤其包括转录、mRNA加工和翻译。本文所述的UTR元件,特别是在所述组合中,尤其被设想增强编码本文所述的CRISPR相关蛋白质的编码序列的转录。

[0045] 因此,本发明的人工核酸有利地包含5'UTR元件和3'UTR元件,各元件均衍生自选自本文所述那些的基因。适当的5'UTR元件选自衍生自选自ATP5A1、RPL32、HSD17B4、SLC7A3、NOSIP和NDUFA4的基因的5'-UTR的5'-UTR元件,优选如本文所定义。适当的3'UTR元件选自衍生自选自GNAS、CASP1、PSMB3、ALB和RPS9的基因的3'UTR的3'UTR元件,优选如本文所定义。

[0046] 一般,本发明的人工核酸分子的5'-或3'-UTR元件与至少一种编码序列是异源的。

[0047] 优选地,本文指出的UTR(充当本发明的人工核酸的UTR元件的“母体UTR”)包括天然存在的(野生型)UTR,以及其同源物、片段、变体和相应RNA序列。

[0048] 换句话说,人工核酸可以优选包含a.至少一个编码区,编码至少一种CRISPR相关蛋白质;b.至少一个5'非翻译区(5'UTR)元件,衍生自选自ATP5A1、RPL32、HSD17B4、SLC7A3、NOSIP和NDUFA4的基因的5'UTR、或任一所述5'UTR的同源物,片段、变体或相应RNA序列;和c.至少一个3'非翻译区(3'UTR)元件,衍生自选自GNAS、CASP1、PSMB3、ALB和RPS9的基因的3'UTR,或任一所述3'UTR的同源物、片段、变体或相应RNA序列。

[0049] 5'UTR和3'UTR优选可操作地连接至本发明的人工核酸的编码序列。

[0050] UTR

[0051] 5'UTR

[0052] 本文所述的人工核酸包括衍生自如本文所述的基因的5'UTR的至少一个5'-UTR元件或其同源物、变体或片段。

[0053] 术语“5'-UTR”是指核酸分子的一部分,其位于开放阅读框的5'(即“上游”)并且不翻译成蛋白质。在本发明的环境中,5'-UTR始于转录起始位点并终于开放阅读框的起始密码子前一个核苷酸。5'-UTR可包含用于控制基因表达的元件,也称为“调控元件”。这种调控元件可以是例如核糖体结合位点。5'-UTR可以被转录后修饰,例如通过添加5'-帽。因此,5'-UTR可以优选地相应于位于5'-帽和起始密码子之间的核酸(具体地,成熟mRNA)序列,并且更具体地相应于从5'-帽的3'位置的核苷酸(优选5'-帽的3'紧邻位置的核苷酸)至蛋白质编码序列起始密码子的5'位置(转录起始位点)(优选蛋白质编码序列的起始密码子的5'紧邻位置(转录起始位点)的核苷酸)的核苷酸延伸的序列。成熟mRNA的5'-帽的3'紧邻位置的核苷酸一般相应于转录起始位点。5'UTR一般具有少于500、400、300、250或少于200个核苷酸的长度。在一些实施方式中,其长度可以在至少10、20、30或40,优选上至100或150个核苷酸的范围内。

[0054] 优选地,所述至少一个5'UTR元件包含如下核酸序列或由如下核酸序列组成:衍生自脊索动物基因,优选脊椎动物基因,更优选哺乳动物基因,最优选人基因的5'UTR,或衍生自脊索动物基因,优选脊椎动物基因,更优选哺乳动物基因,最优选人基因的3'UTR的变体。

[0055] 包括扩展名“.1”或“var”的UTR名称与没有所述扩展名的UTR相同。

[0056] TOP基因衍生的5'UTR元件



[0057] 一些本文指定的5'UTR元件可以衍生自TOP基因的5'UTR或其同源物、变体或片段。

[0058] 因此, TOP基因一般特征在于存在5'端寡嘧啶束(TOP), 并且进一步地一般在于生长相关的翻译调控。然而, 具有组织特异性翻译调控的TOP基因也是已知的。含有5'TOP的mRNA通常称为TOP mRNA。因此, 提供这种信使RNA的基因被称为TOP基因。例如, 在编码肽延伸因子和核糖体蛋白的基因和mRNA中发现了TOP序列。

[0059] 5'端寡嘧啶束(“5'TOP”或“TOP”)一般是位于核酸分子的5'端区域的一段嘧啶核苷酸, 如某些mRNA分子的5'端区域或功能实体的5'端区域, 例如某些基因的转录区域。

[0060] TOP基因的5'UTR相应于衍生自TOP基因的成熟mRNA的5'UTR的序列, 其优选从位于5'-CAP的3'的核苷酸延伸至位于起始密码子的5'的核苷酸。TOP序列一般始于胞苷, 胞苷通常相应于转录起始位点, 然后是一段通常约3至30个嘧啶核苷酸。嘧啶段以及因此5'TOP终止于位于TOP下游第一个嘌呤核苷酸的5'的一个核苷酸。

[0061] TOP基因的5'UTR一般不包含任何起始密码子, 优选不包含上游AUG(uAUG)或上游开放阅读框(uORF)。其中, 上游AUG和上游开放阅读框一般被理解为AUG和存在于应翻译的开放阅读框的起始密码子(AUG)的5'的开放阅读框。TOP基因的5'UTR总体上相当短。TOP基因的5'UTR的长度可以在20个核苷酸上至500个核苷酸之间变化, 并且一般小于约200个核苷酸, 优选小于约150个核苷酸, 更优选小于约100个核苷酸。例如, TOP可包括3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30或甚至更多个核苷酸。

[0062] 在本发明的环境中, “TOP基序”是相应于如上定义的5'TOP的核酸序列。因此, 本发明上下文中的TOP基序优选是长度为3-30个核苷酸的一段嘧啶核苷酸。优选地, TOP-基序由至少3个, 优选至少4个, 更优选至少6个, 更优选至少7个, 最优选至少8个嘧啶核苷酸组成, 其中嘧啶核苷酸段优选在其5'端始于胞嘧啶核苷酸。在TOP基因和TOP mRNA中, “TOP-基序”优选在其5'端始于转录起始位点, 且并终于所述基因或mRNA中的第一个嘌呤残基的5'的一个核苷酸。本发明意义上的“TOP基序”优选位于表示5'UTR的序列的5'端, 或位于编码5'UTR的序列的5'端。因此, 优选地, 一段3个或更多个嘧啶核苷酸称为本发明意义上的“TOP基序”——如果该段位于对应序列如人工核酸分子、人工核酸分子的5'UTR元件、或衍生自本文所述的TOP基因的5'UTR的核酸序列的5'端。换句话说, 不位于5'UTR或5'UTR元件的5'端而位于5'UTR或5'UTR元件内的任何位置的一段3个或更多个嘧啶核苷酸优选不称为“TOP基序”。

[0063] 在特别优选的实施方式中, 本文示例的衍生自TOP基因的5'UTR的5'UTR元件不包含TOP-基序或5'TOP, 如上限定。因此, 衍生自TOP基因的5'UTR的5'UTR元件的核酸序列可以在其3'端终于位于其来源基因或mRNA的起始密码子(例如A(U/T)G)上游的第1、2、3、4、5、6、7、8、9或10位的核苷酸。因此, 5'UTR元件不包含蛋白质编码序列的任何部分。因此, 优选地, 人工核酸的唯一氨基酸编码部分由编码CRISPR相关蛋白质(和任选地如本文所述的其他氨基酸序列)的编码序列提供。

[0064] 下面详细描述根据本发明设想的具体5'UTR元件。

[0065] HSD17B4衍生的5'UTR元件

[0066] 根据本发明的人工核酸可包含5'UTR元件, 其包含下列或由其组成: 衍生自编码17-β-羟基类固醇脱氢酶4的基因的5'UTR的核酸序列、或其同源物、变体或片段, 优选缺少

5' TOP基序。

[0067] 这种5' UTR元件优选包含下列或由其组成：衍生自17-β-羟基类固醇脱氢酶4（“HSD17B4”，也称为过氧化物酶体多功能酶2型）基因的5' UTR的核酸序列，优选衍生自脊椎动物17-β-羟基类固醇脱氢酶4 (HSD17B4) 基因，更优选衍生自哺乳动物17-β-羟基类固醇脱氢酶4 (HSD17B4) 基因，最优选衍生自人17-β-羟基类固醇脱氢酶4 (HSD17B4) 基因，或任何所述5' UTR的同源物、变体或片段，其中优选地5' UTR元件不包括所述基因的5' TOP。

[0068] 因此，根据本发明的人工核酸可包括衍生自HSD17B4基因的5' UTR元件，其中所述5' UTR元件包括下列或由其组成：根据SEQ ID NO:1的DNA序列或其同源物、变体或片段，具体地按递增优选顺序与根据SEQ ID NO:1的核酸序列具有至少5%、10%、20%、30%、40%、50%、60%、70%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、或99%，优选至少70%，更优选至少80%，甚至更优选至少85%，甚至更优选至少90%，并且最优选至少95%或甚至97%序列同一性的DNA序列，或其中所述5' UTR元件包括下列或由其组成：根据SEQ ID NO:2的RNA序列、或其同源物、变体或片段，具体地按递增优选顺序与根据SEQ ID NO:2的核酸序列具有至少5%、10%、20%、30%、40%、50%、60%、70%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、或99%，优选至少70%，更优选至少80%，甚至更优选至少85%，甚至更优选至少90%，并且最优选至少95%或甚至97%序列同一性的RNA序列。

[0069] RPL32衍生的5' -UTR元件

[0070] 根据本发明的人工核酸可包括5' UTR元件，该5' UTR元件包括如下核酸序列或由其组成：所述核酸序列衍生自编码核糖体大蛋白 (RPL) 的基因的5' UTR；或其同源物、变体或片段，其中所述5' UTR元件优选缺少5' TOP (末端寡嘧啶束) 基序。

[0071] 这种5' UTR元件优选包含下列或由其组成：衍生自核糖体大蛋白32（“RPL32”）基因的5' UTR的核酸序列，优选地衍生自脊椎动物核糖体大蛋白32 (L32) 基因，更优选衍生自哺乳动物核糖体大蛋白32 (L32) 基因，最优选衍生自人核糖体大蛋白32 (L32) 基因，或任何所述5' UTR的同源物、变体或片段，其中5' UTR元件优选不包括所述基因的5' TOP。术语“RPL32”还包括其变体和片段，其在本文中也称为“RPL32var”或“32L4”。

[0072] 因此，根据本发明的人工核酸可包括衍生自RPL32基因的5' UTR元件，其中所述5' UTR元件包括下列或由其组成：根据SEQ ID NO:21的DNA序列或其同源物、变体或片段，具体地按递增优选顺序与根据SEQ ID NO:21的核酸序列具有至少5%、10%、20%、30%、40%、50%、60%、70%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、或99%，优选至少70%，更优选至少80%，甚至更优选至少85%，甚至更优选至少90%，并且最优选至少95%或甚至97%序列同一性的DNA序列，或其中所述5' UTR元件包括下列或由其组成：根据SEQ ID NO:22的RNA序列、或其同源物、变体或片段，具体地按递增优选顺序与根据SEQ ID NO:22的核酸序列具有至少5%、10%、20%、30%、40%、50%、60%、70%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、或99%，优选至少70%，更优选至少80%，甚至更优选至少85%，甚至更优选至少90%，并且最优选至少95%或甚至97%序列同一性的RNA序列。

[0073] NDUFA4衍生的5' -UTR元件

[0074] 根据本发明的人工核酸可包括5' UTR元件，其包括如下核酸序列或由其组成：所述

核酸序列衍生自编码细胞色素c氧化酶亚基 (NDUFA4) 的基因的5' UTR;或其同源物、片段或变体。

[0075] 这种5' UTR元件优选包含下列或由其组成:衍生自细胞色素c氧化酶亚基 (“NDUFA4”或“Ndufa4.1”) 基因的5' UTR的核酸序列,优选衍生自脊椎动物细胞色素c氧化酶亚基 (NDUFA4) 基因,更优选衍生自哺乳动物细胞色素c氧化酶亚基 (NDUFA4) 基因,最优选衍生自人细胞色素c氧化酶亚基 (NDUFA4) 基因,或其同源物、变体或片段。

[0076] 因此,根据本发明的人工核酸可包括衍生自NDUFA4基因的5' UTR元件,其中所述5' UTR元件包括下列或由下列组成:根据SEQ ID NO:9的DNA序列或其同源物、变体或片段,具体地按递增优选顺序与根据SEQ ID NO:9的核酸序列具有至少5%、10%、20%、30%、40%、50%、60%、70%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、或99%,优选至少70%,更优选至少80%,甚至更优选至少85%,甚至更优选至少90%,并且最优选至少95%或甚至97%序列同一性的DNA序列,或其中所述5' UTR元件包括下列或由下列组成:根据SEQ ID NO:10的RNA序列、或其同源物、变体或片段,具体地按递增优选顺序与根据SEQ ID NO:10的核酸序列具有至少5%、10%、20%、30%、40%、50%、60%、70%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、或99%,优选至少70%,更优选至少80%,甚至更优选至少85%,甚至更优选至少90%,并且最优选至少95%或甚至97%序列同一性RNA序列。

[0077] SLC7A3衍生的5' -UTR元件

[0078] 根据本发明的人工核酸可包括5' UTR元件,其包括如下核酸序列或由其组成:所述核酸序列衍生自基因编码溶质载体家族7成员3 (SLC7A3) 的5' UTR;或其同源物、片段或变体。

[0079] 这种5' UTR元件优选包含下列或由其组成:衍生自溶质载体家族7成员3 (“SLC7A3”或“Slc7a3.1”) 基因的5' UTR的核酸序列,优选衍生自脊椎动物溶质载体家族7成员3 (SLC7A3) 基因,更优选衍生自哺乳动物溶质载体家族7成员3 (SLC7A3) 基因,最优选衍生自人溶质载体家族7成员3 (SLC7A3) 基因;或其同源物、变体或片段。

[0080] 因此,根据本发明的人工核酸可包括衍生自SLC7A3基因的5' UTR元件,其中所述5' UTR元件包括下列或由下列组成:根据SEQ ID NO:15的DNA序列或其同源物、变体或片段,具体地按递增优选顺序与根据SEQ ID NO:15的核酸序列具有至少5%、10%、20%、30%、40%、50%、60%、70%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、或99%,优选至少70%,更优选至少80%,甚至更优选至少85%,甚至更优选至少90%,并且最优选至少95%或甚至97%序列同一性的DNA序列,或其中所述5' UTR元件包括下列或由下列组成:根据SEQ ID NO:16的RNA序列、或其同源物、变体或片段,具体地按递增优选顺序与根据SEQ ID NO:16的核酸序列具有至少5%、10%、20%、30%、40%、50%、60%、70%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、或99%,优选至少70%,更优选至少80%,甚至更优选至少85%,甚至更优选至少90%,并且最优选至少95%或甚至97%序列同一性的RNA序列。

[0081] NOSIP衍生的5' UTR元件

[0082] 根据本发明的人工核酸可包括5' UTR元件,其包括下列或由下列组成:衍生自编码一氧化氮合酶相互作用蛋白的基因的5' UTR的核酸序列;或其同源物、变体或片段。

[0083] 这种5' UTR元件优选包含下列或由其组成:衍生自一氧化氮合酶相互作用蛋白(“NOSIP”或“Nosip.1”)基因的5' UTR的核酸序列,优选衍生自脊椎动物一氧化氮合酶相互作用蛋白(NOSIP)基因,更优选衍生自哺乳动物一氧化氮合酶相互作用蛋白(NOSIP)基因,最优选衍生自人一氧化氮合酶相互作用蛋白(NOSIP)基因;或其同源物、变体或片段。

[0084] 因此,根据本发明的人工核酸可包括衍生自NOSIP基因的5' UTR元件,其中所述5' UTR元件包括下列或由其组成:根据SEQ ID NO:11的DNA序列或其同源物、变体或片段,具体地按递增优选顺序与根据SEQ ID NO:11的核酸序列具有至少5%、10%、20%、30%、40%、50%、60%、70%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、或99%,优选至少70%,更优选至少80%,甚至更优选至少85%,甚至更优选至少90%,并且最优选至少95%或甚至97%序列同一性的DNA序列,或其中所述5' UTR元件包括下列或由其组成:根据SEQ ID NO:12的RNA序列、或其同源物、变体或片段,具体地按递增优选顺序与根据SEQ ID NO:12的核酸序列具有至少5%、10%、20%、30%、40%、50%、60%、70%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、或99%,优选至少70%,更优选至少80%,甚至更优选至少85%,甚至更优选至少90%,并且最优选至少95%或甚至97%序列同一性的RNA序列。

[0085] ATP5A1衍生的5' -UTR元件

[0086] 根据本发明的人工核酸可包括5' UTR元件,其包括下列或由其组成:衍生自编码线粒体ATP合酶亚基 $\alpha$  (ATP5A1) 的基因的5' UTR的核酸序列、或其同源物、变体或片段,其中所述5' UTR元件优选缺少5' TOP基序。

[0087] 这种5' UTR元件优选包含下列或由其组成:衍生自线粒体ATP合酶亚基 $\alpha$  (“ATP5A1”) 基因的5' UTR的核酸序列,优选衍生自脊椎动物线粒体ATP合酶亚基 $\alpha$  (ATP5A1) 基因,更优选衍生自哺乳动物线粒体ATP合酶亚基 $\alpha$  (ATP5A1) 基因,最优选衍生自人线粒体ATP合酶亚基 $\alpha$  (ATP5A1) 基因,或其同源物、变体或片段,其中5' UTR元件优选不包括所述基因的5' TOP。

[0088] 因此,根据本发明的人工核酸可包括衍生自ATP5A1基因的5' UTR元件,其中所述5' UTR元件包括下列或由其组成:根据SEQ ID NO:5的DNA序列或其同源物、变体或片段,具体地按递增优选顺序与根据SEQ ID NO:5的核酸序列具有至少5%、10%、20%、30%、40%、50%、60%、70%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、或99%,优选至少70%,更优选至少80%,甚至更优选至少85%,甚至更优选至少90%,并且最优选至少95%或甚至97%序列同一性的DNA序列,或其中所述5' UTR元件包括下列或由其组成:根据SEQ ID NO:6的RNA序列、或其同源物、变体或片段,具体地按递增优选顺序与根据SEQ ID NO:6的核酸序列具有至少5%、10%、20%、30%、40%、50%、60%、70%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、或99%,优选至少70%,更优选至少80%,甚至更优选至少85%,甚至更优选至少90%,并且最优选至少95%或甚至97%序列同一性的RNA序列。

[0089] ASAH1衍生的5' -UTR元件

[0090] 根据本发明的人工核酸可包括5' UTR元件,其包括下列或由其组成:衍生自编码线粒体ATP合酶亚基 $\alpha$  (ASAH1) 的基因的5' UTR的核酸序列、或其同源物、变体或片段,其中所述5' UTR元件优选缺少5' TOP基序。

[0091] 这种5' UTR元件优选包含下列或由其组成:衍生自线粒体ATP合酶亚基 $\alpha$  (“ASAH1”)基因的5' UTR的核酸序列,优选衍生自脊椎动物线粒体ATP合酶亚基 $\alpha$  (ASAH1) 基因,更优选衍生自哺乳动物线粒体ATP合酶亚基 $\alpha$  (ASAH1) 基因,最优选衍生自人线粒体ATP合酶亚基 $\alpha$  (ASAH1) 基因;或其同源物、变体或片段,其中5' UTR元件优选不包括所述基因的5' TOP。

[0092] 因此,根据本发明的人工核酸可包括衍生自ASAH1基因的5' UTR元件,其中所述5' UTR元件包括下列或由其组成:根据SEQ ID NO:3的DNA序列或其同源物、变体或片段,具体地按递增优选顺序与根据SEQ ID NO:3的核酸序列具有至少5%、10%、20%、30%、40%、50%、60%、70%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、或99%,优选至少70%,更优选至少80%,甚至更优选至少85%,甚至更优选至少90%,并且最优选至少95%或甚至97%序列同一性的DNA序列,或其中所述5' UTR元件包括下列或由其组成:根据SEQ ID NO:4的RNA序列、或其同源物、变体或片段,具体地按递增优选顺序与根据SEQ ID NO:4的核酸序列具有至少5%、10%、20%、30%、40%、50%、60%、70%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、或99%,优选至少70%,更优选至少80%,甚至更优选至少85%,甚至更优选至少90%,并且最优选至少95%或甚至97%序列同一性的RNA序列。

[0093] MP68衍生的5' -UTR元件

[0094] 根据本发明的人工核酸可包括5' UTR元件,其包括下列或由其组成:衍生自编码线粒体ATP合酶亚基 $\alpha$  (MP68) 的基因的5' UTR的核酸序列、或其同源物、变体或片段,其中所述5' UTR元件优选缺少5' TOP基序。

[0095] 这种5' UTR元件优选包含下列或由其组成:衍生自线粒体ATP合酶亚基 $\alpha$  (“MP68”或“Mp68”) 基因的5' UTR的核酸序列,优选衍生自脊椎动物线粒体ATP合酶亚基 $\alpha$  (MP68) 基因,更优选衍生自哺乳动物线粒体ATP合酶亚基 $\alpha$  (MP68) 基因,最优选衍生自人线粒体ATP合酶亚基 $\alpha$  (MP68) 基因;或其同源物、变体或片段,其中5' UTR元件优选不包括所述基因的5' TOP。

[0096] 因此,根据本发明的人工核酸可包括衍生自Mp68基因的5' UTR元件,其中所述5' UTR元件包括下列或由其组成:根据SEQ ID NO:7的DNA序列或其同源物、变体或片段,具体地按递增优选顺序与根据SEQ ID NO:7的核酸序列具有至少5%、10%、20%、30%、40%、50%、60%、70%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、或99%,优选至少70%,更优选至少80%,甚至更优选至少85%,甚至更优选至少90%,并且最优选至少95%或甚至97%序列同一性的DNA序列,或其中所述5' UTR元件包括下列或由其组成:根据SEQ ID NO:8的RNA序列、或其同源物、变体或片段,具体地按递增优选顺序与根据SEQ ID NO:8的核酸序列具有至少5%、10%、20%、30%、40%、50%、60%、70%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、或99%,优选至少70%,更优选至少80%,甚至更优选至少85%,甚至更优选至少90%,并且最优选至少95%或甚至97%序列同一性的RNA序列。

[0097] RPL31衍生的5' -UTR元件

[0098] 根据本发明的人工核酸可包括5' UTR元件,其包括下列或由其组成:衍生自编码线粒体ATP合酶亚基 $\alpha$  (RPL31) 的基因的5' UTR的核酸序列、或其同源物、变体或片段,其中所述5' UTR元件优选缺少5' TOP基序。

[0099] 这种5' UTR元件优选包含下列或由其组成:衍生自线粒体ATP合酶亚基 $\alpha$  (“RPL31”

或“Rp131.1”)基因的5' UTR的核酸序列,优选衍生自脊椎动物线粒体ATP合酶亚基 $\alpha$  (RPL31)基因,更优选衍生自哺乳动物线粒体ATP合酶亚基 $\alpha$  (RPL31)基因,最优选衍生自人线粒体ATP合酶亚基 $\alpha$  (RPL31)基因;或其同源物、变体或片段,其中5' UTR元件优选不包括所述基因的5' TOP。

[0100] 因此,根据本发明的人工核酸可包括衍生自RPL31基因的5' UTR元件,其中所述5' UTR元件包括下列或由下列组成:根据SEQ ID NO:13的DNA序列或其同源物、变体或片段,具体地按递增优选顺序与根据SEQ ID NO:13的核酸序列具有至少5%、10%、20%、30%、40%、50%、60%、70%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、或99%,优选至少70%,更优选至少80%,甚至更优选至少85%,甚至更优选至少90%,并且最优选至少95%或甚至97%序列同一性的DNA序列,或其中所述5' UTR元件包括下列或由下列组成:根据SEQ ID NO:14的RNA序列、或其同源物、变体或片段,具体地按递增优选顺序与根据SEQ ID NO:14的核酸序列具有至少5%、10%、20%、30%、40%、50%、60%、70%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、或99%,优选至少70%,更优选至少80%,甚至更优选至少85%,甚至更优选至少90%,并且最优选至少95%或甚至97%序列同一性的RNA序列。

[0101] TUBB4B衍生的5' -UTR元件

[0102] 根据本发明的人工核酸可包括5' UTR元件,其包括下列或由下列组成:衍生自编码线粒体ATP合酶亚基 $\alpha$  (TUBB4B)的基因的5' UTR的核酸序列、或其同源物、变体或片段,其中所述5' UTR元件优选缺少5' TOP基序。

[0103] 这种5' UTR元件优选包含下列或由其组成:衍生自线粒体ATP合酶亚基 $\alpha$  (“TUBB4B”或“TUBB4B.1”)基因的5' UTR的核酸序列,优选衍生自脊椎动物线粒体ATP合酶亚基 $\alpha$  (TUBB4B)基因,更优选衍生自哺乳动物线粒体ATP合酶亚基 $\alpha$  (TUBB4B)基因,最优选衍生自人线粒体ATP合酶亚基 $\alpha$  (TUBB4B)基因;或其同源物、变体或片段,其中5' UTR元件优选不包括所述基因的5' TOP。

[0104] 因此,根据本发明的人工核酸可包括衍生自TUBB4B基因的5' UTR元件,其中所述5' UTR元件包括下列或由下列组成:根据SEQ ID NO:17的DNA序列或其同源物、变体或片段,具体地按递增优选顺序与根据SEQ ID NO:17的核酸序列具有至少5%、10%、20%、30%、40%、50%、60%、70%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、或99%,优选至少70%,更优选至少80%,甚至更优选至少85%,甚至更优选至少90%,并且最优选至少95%或甚至97%序列同一性的DNA序列,或其中所述5' UTR元件包括下列或由下列组成:根据SEQ ID NO:18的RNA序列、或其同源物、变体或片段,具体地按递增优选顺序与根据SEQ ID NO:18的核酸序列具有至少5%、10%、20%、30%、40%、50%、60%、70%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、或99%,优选至少70%,更优选至少80%,甚至更优选至少85%,甚至更优选至少90%,并且最优选至少95%或甚至97%序列同一性的RNA序列。

[0105] UBQLN2衍生的5' -UTR元件

[0106] 根据本发明的人工核酸可包括5' UTR元件,其包括下列或由下列组成:衍生自编码线粒体ATP合酶亚基 $\alpha$  (UBQLN2)的基因的5' UTR的核酸序列、或其同源物、变体或片段,其中所述5' UTR元件优选缺少5' TOP基序。

[0107] 这种5' UTR元件优选包含下列或由其组成:衍生自线粒体ATP合酶亚基 $\alpha$  (“UBQLN2”或“Ubqln2.1”)基因的5' UTR的核酸序列,优选衍生自脊椎动物线粒体ATP合酶亚基 $\alpha$  (UBQLN2)基因,更优选衍生自哺乳动物线粒体ATP合酶亚基 $\alpha$  (UBQLN2)基因,最优选衍生自人线粒体ATP合酶亚基 $\alpha$  (UBQLN2)基因;或其同源物、变体或片段,其中5' UTR元件优选不包括所述基因的5' TOP。

[0108] 因此,根据本发明的人工核酸可包括衍生自UBQLN2基因的5' UTR元件,其中所述5' UTR元件包括下列或由其组成:根据SEQ ID NO:19的DNA序列或其同源物、变体或片段,具体地按递增优选顺序与根据SEQ ID NO:19的核酸序列具有至少5%、10%、20%、30%、40%、50%、60%、70%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、或99%,优选至少70%,更优选至少80%,甚至更优选至少85%,甚至更优选至少90%,并且最优选至少95%或甚至97%序列同一性的DNA序列,或其中所述5' UTR元件包括下列或由其组成:根据SEQ ID NO:20的RNA序列、或其同源物、变体或片段,具体地按递增优选顺序与根据SEQ ID NO:20的核酸序列具有至少5%、10%、20%、30%、40%、50%、60%、70%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、或99%,优选至少70%,更优选至少80%,甚至更优选至少85%,甚至更优选至少90%,并且最优选至少95%或甚至97%序列同一性的RNA序列。

#### [0109] 3' UTR

[0110] 本文所述的人工核酸进一步包括衍生自本文所述基因或所述基因的同源物、变体、片段的3' UTR的至少一个3' -UTR元件。术语“3' -UTR”指代核酸分子中位于开放阅读框的3' (即“下游”)并且不被翻译成蛋白质的部分。在本发明的环境中,3' -UTR相应于位于人工核酸分子(优选地RNA)的蛋白质编码序列终止密码子(优选地紧邻蛋白质编码序列的终止密码子的3')和聚(A)序列之间的序列。

[0111] 优选地,所述至少一个3' UTR元件包括下列或由其组成:衍生自脊索动物基因,优选脊椎动物基因,更优选哺乳动物基因,最优选人基因的3' UTR的核酸序列、或脊索动物基因,优选脊椎动物基因,更优选哺乳动物基因,最优选人基因的3' UTR的变体。

#### [0112] GNAS衍生的3' -UTR元件

[0113] 根据本发明的人工核酸可包括3' UTR元件,其包括下列或由其组成:衍生自编码鸟嘌呤核苷酸结合蛋白G(一种或多种)亚基 $\alpha$ 同种型短(GNAS)的基因的3' UTR的核酸序列、或其同源物、变体或片段。

[0114] 这种3' UTR元件优选包括下列或由其组成:衍生自鸟嘌呤核苷酸结合蛋白G(一种或多种)亚基 $\alpha$ 同种型短 (“GNAS”或“Gnas.1”)基因的3' UTR的核酸序列,优选衍生自脊椎动物鸟嘌呤核苷酸结合蛋白G(一种或多种)亚基 $\alpha$ 同种型短(GNAS)基因,更优选衍生自哺乳动物鸟嘌呤核苷酸结合蛋白G(一种或多种)亚基 $\alpha$ 同种型短(GNAS)基因,最优选衍生自人鸟嘌呤核苷酸结合蛋白G(一种或多种)亚基 $\alpha$ 同种型短(GNAS)基因;或其同源物、变体或片段。

[0115] 因此,根据本发明的人工核酸可包括衍生自GNAS基因的3' UTR元件,其中所述3' UTR元件包括下列或由其组成:根据SEQ ID NO:29的DNA序列或其同源物、变体或片段,具体地按递增优选顺序与根据SEQ ID NO:29的核酸序列具有至少5%、10%、20%、30%、40%、50%、60%、70%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、或99%,优选至少70%,更优选至少80%,甚至更优选至少85%,甚

至更优选至少90%，并且最优选至少95%或甚至97%序列同一性的DNA序列，或其中所述3' UTR元件包括下列或由下列组成：根据SEQ ID NO:30的RNA序列、或其同源物、变体或片段，具体地按递增优选顺序与根据SEQ ID NO:30的核酸序列具有至少5%、10%、20%、30%、40%、50%、60%、70%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、或99%，优选至少70%，更优选至少80%，甚至更优选至少85%，甚至更优选至少90%，并且最优选至少95%或甚至97%序列同一性的RNA序列。

[0116] CASP1衍生的3' -UTR元件

[0117] 根据本发明的人工核酸可包括3' UTR元件，其包括下列或由下列组成：衍生自编码胱天蛋白酶-1 (CASP1) 的基因的3' UTR的核酸序列、或其同源物、变体或片段。

[0118] 这种3' UTR元件优选包括下列或由下列组成：衍生自胱天蛋白酶-1 (“CASP1”或“CASP1.1”) 基因的3' UTR的核酸序列，优选衍生自脊椎动物胱天蛋白酶-1 (CASP1) 基因，更优选衍生自哺乳动物胱天蛋白酶-1 (CASP1) 基因，最优选衍生自人胱天蛋白酶-1 (CASP1) 基因；或其同源物、变体或片段。

[0119] 因此，根据本发明的人工核酸可包括衍生自CASP1基因的3' UTR元件，其中所述3' UTR元件包括下列或由下列组成：根据SEQ ID NO:25的DNA序列或其同源物、变体或片段，具体地按递增优选顺序与根据SEQ ID NO:25的核酸序列具有至少5%、10%、20%、30%、40%、50%、60%、70%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、或99%，优选至少70%，更优选至少80%，甚至更优选至少85%，甚至更优选至少90%，并且最优选至少95%或甚至97%序列同一性的DNA序列，或其中所述3' UTR元件包括下列或由下列组成：根据SEQ ID NO:26的RNA序列、或其同源物、变体或片段，具体地按递增优选顺序与根据SEQ ID NO:26的核酸序列具有至少5%、10%、20%、30%、40%、50%、60%、70%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、或99%，优选至少70%，更优选至少80%，甚至更优选至少85%，甚至更优选至少90%，并且最优选至少95%或甚至97%序列同一性的RNA序列。

[0120] PSMB3衍生的3' -UTR元件

[0121] 根据本发明的人工核酸可包括3' UTR元件，其包括下列或由下列组成：衍生自编码蛋白酶体亚基β3型 (PSMB3) 的基因的3' UTR的核酸序列、或其同源物、变体或片段。

[0122] 这种3' UTR元件优选包括下列或由下列组成：衍生自蛋白酶体亚基β3型 (“PSMB3”或“PSMB3.1”) 基因的3' UTR的核酸序列，优选衍生自脊椎动物蛋白酶体亚基β3型 (PSMB3) 基因，更优选衍生自哺乳动物蛋白酶体亚基β3型 (PSMB3) 基因，最优选衍生自人蛋白酶体亚基β3型 (PSMB3) 基因；或其同源物、变体或片段。

[0123] 因此，根据本发明的人工核酸可包括衍生自PSMB3基因的3' UTR元件，其中所述3' UTR元件包括下列或由下列组成：根据SEQ ID NO:23的DNA序列或其同源物、变体或片段，具体地按递增优选顺序与根据SEQ ID NO:23的核酸序列具有至少5%、10%、20%、30%、40%、50%、60%、70%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、或99%，优选至少70%，更优选至少80%，甚至更优选至少85%，甚至更优选至少90%，并且最优选至少95%或甚至97%序列同一性的DNA序列，或其中所述3' UTR元件包括下列或由下列组成：根据SEQ ID NO:24的RNA序列、或其同源物、变体或片段，具体地按递增优选顺序与根据SEQ ID NO:24的核酸序列具有至少5%、10%、20%、30%、



40%、50%、60%、70%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、或99%，优选至少70%，更优选至少80%，甚至更优选至少85%，甚至更优选至少90%，并且最优选至少95%或甚至97%序列同一性的RNA序列。

[0124] ALB衍生的3' UTR元件

[0125] 根据本发明的人工核酸可包括3' UTR元件，其包括下列或由下列组成：衍生自编码血清白蛋白 (ALB) 的基因的3' UTR的核酸序列、或其同源物、变体或片段。

[0126] 这种3' UTR元件优选包括下列或由下列组成：衍生自血清白蛋白 (“ALB”或“白蛋白7”) 基因的3' UTR的核酸序列，优选衍生自脊椎动物血清白蛋白 (ALB) 基因，更优选衍生自哺乳动物血清白蛋白 (ALB) 基因，最优选衍生自人血清白蛋白 (ALB) 基因；或其同源物、变体或片段。

[0127] 因此，根据本发明的人工核酸可包括衍生自ALB基因的3' UTR元件，其中所述3' UTR元件包括下列或由下列组成：根据SEQ ID NO:35的DNA序列、或其同源物、变体或片段，具体地按递增优选顺序与根据SEQ ID NO:35的核酸序列具有至少5%、10%、20%、30%、40%、50%、60%、70%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、或99%，优选至少70%，更优选至少80%，甚至更优选至少85%，甚至更优选至少90%，并且最优选至少95%或甚至97%序列同一性的DNA序列，或其中所述3' UTR元件包括下列或由下列组成：根据SEQ ID NO:36的RNA序列、或其同源物、变体或片段，具体地按递增优选顺序与根据SEQ ID NO:36的核酸序列具有至少5%、10%、20%、30%、40%、50%、60%、70%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、或99%，优选至少70%，更优选至少80%，甚至更优选至少85%，甚至更优选至少90%，并且最优选至少95%或甚至97%序列同一性的RNA序列。

[0128] RPS9衍生的3' -UTR元件

[0129] 根据本发明的人工核酸可包括3' UTR元件，其包括下列或由下列组成：衍生自40S核糖体蛋白S9 (RPS9) 编码的基因的3' UTR的核酸序列、或其同源物、变体或片段。

[0130] 这种3' UTR元件优选包括下列或由下列组成：衍生自40S核糖体蛋白S9 (“RPS9”或“RPS9.1”) 基因的3' UTR的核酸序列，优选衍生自脊椎动物40S核糖体蛋白S9 (RPS9) 基因，更优选衍生自哺乳动物40S核糖体蛋白S9 (RPS9) 基因，最优选衍生自人40S核糖体蛋白S9 (RPS9) 基因；或其同源物、变体或片段。

[0131] 因此，根据本发明的人工核酸可包括衍生自RPS9基因的3' UTR元件，其中所述3' UTR元件包括下列或由下列组成：根据SEQ ID NO:33的DNA序列或其同源物、变体或片段，具体地按递增优选顺序与根据SEQ ID NO:33的核酸序列具有至少5%、10%、20%、30%、40%、50%、60%、70%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、或99%，优选至少70%，更优选至少80%，甚至更优选至少85%，甚至更优选至少90%，并且最优选至少95%或甚至97%序列同一性的DNA序列，或其中所述5' UTR元件包括下列或由下列组成：根据SEQ ID NO:34的RNA序列、或其同源物、变体或片段，具体地按递增优选顺序与根据SEQ ID NO:34的核酸序列具有至少5%、10%、20%、30%、40%、50%、60%、70%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、或99%，优选至少70%，更优选至少80%，甚至更优选至少85%，甚至更优选至少90%，并且最优选至少95%或甚至97%序列同一性的RNA序列。

[0132] COX6B1衍生的3'-UTR元件

[0133] 根据本发明的人工核酸可包括3' UTR元件,其包括下列或由下列组成:衍生自COX6B1基因的3' UTR的核酸序列、或其同源物、变体或片段。

[0134] 这种3' UTR元件优选包括下列或由下列组成:衍生自COX6B1(或“COX6B1.1”)基因的3' UTR的核酸序列,优选衍生自脊椎动物COX6B1基因,更优选衍生自哺乳动物COX6B1基因,最优选衍生自人COX6B1基因、或其同源物、变体或片段。

[0135] 因此,根据本发明的人工核酸可包括衍生自COX6B1基因的3' UTR元件,其中所述3' UTR元件包括下列或由下列组成:根据SEQ ID NO:27的DNA序列或其同源物、变体或片段,具体地按递增优选顺序与根据SEQ ID NO:27的核酸序列具有至少5%、10%、20%、30%、40%、50%、60%、70%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、或99%,优选至少70%,更优选至少80%,甚至更优选至少85%,甚至更优选至少90%,并且最优选至少95%或甚至97%序列同一性的DNA序列,或其中所述5' UTR元件包括下列或由下列组成:根据SEQ ID NO:28的RNA序列、或其同源物、变体或片段,具体地按递增优选顺序与根据SEQ ID NO:28的核酸序列具有至少5%、10%、20%、30%、40%、50%、60%、70%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、或99%,优选至少70%,更优选至少80%,甚至更优选至少85%,甚至更优选至少90%,并且最优选至少95%或甚至97%序列同一性的RNA序列。

[0136] NDUFA1衍生的3'-UTR元件

[0137] 根据本发明的人工核酸可包括3' UTR元件,其包括下列或由下列组成:衍生自NDUFA1基因的3' UTR的核酸序列、或其同源物、变体或片段。

[0138] 这种3' UTR元件优选包括下列或由下列组成:衍生自NDUFA1(或“Ndufa1.1”)基因的3' UTR的核酸序列,优选衍生自脊椎动物NDUFA1基因,更优选衍生自哺乳动物NDUFA1基因,最优选衍生自人NDUFA1基因、或其同源物、变体或片段。

[0139] 因此,根据本发明的人工核酸可包括衍生自NDUFA1基因的3' UTR元件,其中所述3' UTR元件包括下列或由下列组成:根据SEQ ID NO:31的DNA序列或其同源物、变体或片段,具体地按递增优选顺序与根据SEQ ID NO:31的核酸序列具有至少5%、10%、20%、30%、40%、50%、60%、70%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、或99%,优选至少70%,更优选至少80%,甚至更优选至少85%,甚至更优选至少90%,并且最优选至少95%或甚至97%序列同一性的DNA序列,或其中所述5' UTR元件包括下列或由下列组成:根据SEQ ID NO:32的RNA序列、或其同源物、变体或片段,具体地按递增优选顺序与根据SEQ ID NO:32的核酸序列具有至少5%、10%、20%、30%、40%、50%、60%、70%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、或99%,优选至少70%,更优选至少80%,甚至更优选至少85%,甚至更优选至少90%,并且最优选至少95%或甚至97%序列同一性的RNA序列。

[0140] UTR组合

[0141] 优选地,所述至少一个5' UTR元件和所述至少一个3' UTR元件发生协同作用,以增加可操作地连接至所述UTR的至少一种编码序列的表达。在此设想以任何有用的组合利用所述5'-UTR和3'-UTR。特别有用的5'和3' UTR列举在下表1A中。特别有用的5' UTR和3'-UTR组合列举在下表1B中。本发明的特别优选的实施方式包括选择的CDS(即Cas9、Cpf1、CasX、

CasY、或Cas13)与选自下列的UTR组合的组合:HSD17B4/Gnas.1;Slc7a3.1/Gnas.1;ATP5A1/CASP.1;Ndufa4.1/PSMB3.1;HSD17B4/PSMB3.1;RPL32var/白蛋白7;32L4/白蛋白7;HSD17B4/CASP1.1;Slc7a3.1/CASP1.1;Slc7a3.1/PSMB3.1;Nosip.1/PSMB3.1;Ndufa4.1/RPS9.1;HSD17B4/RPS9.1;ATP5A1/Gnas.1;Ndufa4.1/COX6B1.1;Ndufa4.1/Gnas.1;Ndufa4.1/Ndufa1.1;Nosip.1/Ndufa1.1;Rpl31.1/Gnas.1;TUBB4B.1/RPS9.1;和Ubqln2.1/RPS9.1。

[0142] 表1A

[0143]

描述	序列类型	SEQ ID NO
HSD17B4 5'-UTR	DNA	SEQ ID NO: 1
HSD17B4 5'-UTR	RNA	SEQ ID NO: 2
ASAH1 5'-UTR	DNA	SEQ ID NO: 3
ASAH1 5'-UTR	RNA	SEQ ID NO: 4
ATP5A1 5'-UTR	DNA	SEQ ID NO: 5
ATP5A1 5'-UTR	RNA	SEQ ID NO: 6
Mp68 5'-UTR	DNA	SEQ ID NO: 7
Mp68 5'-UTR	RNA	SEQ ID NO: 8
Ndufa4 5'-UTR	DNA	SEQ ID NO: 9
Ndufa4 5'-UTR	RNA	SEQ ID NO: 10
Nosip 5'-UTR	DNA	SEQ ID NO: 11
Nosip 5'-UTR	RNA	SEQ ID NO: 12
Rpl31 5'-UTR	DNA	SEQ ID NO: 13
Rpl31 5'-UTR	RNA	SEQ ID NO: 14
Slc7a3 5'-UTR	DNA	SEQ ID NO: 15
Slc7a3 5'-UTR	RNA	SEQ ID NO: 16
TUBB4B 5'-UTR	DNA	SEQ ID NO: 17
TUBB4B 5'-UTR	RNA	SEQ ID NO: 18
Ubqln2 5'-UTR	DNA	SEQ ID NO: 19
Ubqln2 5'-UTR	RNA	SEQ ID NO: 20
RPL32 (32L4) 5'-UTR	DNA	SEQ ID NO: 21
RPL32 (32L4) 5'-UTR	RNA	SEQ ID NO: 22
PSMB3 3'-UTR	DNA	SEQ ID NO: 23
PSMB3 3'-UTR	RNA	SEQ ID NO: 24
CASP1 3'-UTR	DNA	SEQ ID NO: 25
CASP1 3'-UTR	RNA	SEQ ID NO: 26
COX6B1 3'-UTR	DNA	SEQ ID NO: 27
COX6B1 3'-UTR	RNA	SEQ ID NO: 28
Gnas 3'-UTR	DNA	SEQ ID NO: 29

[0144]

Gnas3'-UTR	RNA	SEQ ID NO: 30
Ndufa1 3'-UTR	DNA	SEQ ID NO: 31
Ndufa1 3'-UTR	RNA	SEQ ID NO: 32
RPS9 3'-UTR	DNA	SEQ ID NO: 33
RPS9 3'-UTR	RNA	SEQ ID NO: 34
ALB7 3'-UTR	DNA	SEQ ID NO: 35
ALB7 3'-UTR	RNA	SEQ ID NO: 36

[0145] 表1B: 有用的UTR-组合和相应构建体

[0146]

UTR 组合	SEQ ID NO
HSD17B4/Gnas	413; 2330-2345; 3490-3505; 4650-4665; 5810-5825; 6970-6985; 8130-8145; 9290-9305; 10402-10408; 10554; 10599-10612
Slc7a3/Gnas	414; 2346-2361; 3506-3521; 4666-4681; 5826-5841; 6986-7001; 8146-8161; 9306-9321; 10409-10415; 10555; 10613-10626
ATP5A1/CASP	415; 2362-2377; 3522-3537; 4682-4697; 5842-5857; 7002-7017; 8162-8177; 9322-9337; 10416-10422; 10556; 10627-10640
Ndufa4/PSMB3	416; 2378-2393; 3538-3553; 4698-4713; 5858-5873; 7018-7033; 8178-8193; 9338-9353; 10423-10429; 10557; 10641-10654
HSD17B4/PSMB3	417; 2394-2409; 3554-3569; 4714-4729; 5874-5889; 7034-7049; 8194-8209; 9354-9369; 10430-10436; 10558; 10655-10668
RPL32/白蛋白 7	418; 2410-2425; 3570-3585; 4730-4745; 5890-5905; 7050-7065; 8210-8225; 9370-9385; 10437-10443; 10559; 10669-10682
32L4/白蛋白 7 (Gen5, HSL, PolyC)	419; 2426-2441; 3586-3601; 4746-4761; 5906-5921; 7066-7081; 8226-8241; 9386-9401; 10444-10450; 10560; 10683-10696
HSD17B4/CASP1	420; 2442-2457; 3602-3617; 4762-4777; 5922-5937; 7082-7097; 8242-8257; 9402-9417; 10451-10457; 10561; 10697-10710
Slc7a3/CASP1	421; 2458-2473; 3618-3633; 4778-4793; 5938-5953; 7098-7113; 8258-8273; 9418-9433; 10458-10464; 10562; 10711-10724
Slc7a3/PSMB3	422; 2474-2489; 3634-3649; 4794-4809; 5954-5969; 7114-7129; 8274-8289; 9434-9449; 10465-10471; 10563; 10725-10738
Nosip/PSMB3	423; 2490-2505; 3650-3665; 4810-4825; 5970-5985; 7130-7145; 8290-8305; 9459-9450; 10472-10478; 10564; 10739-10752
Ndufa4/RPS9	424; 2506-2521; 3666-3681; 4826-4841; 5986-6001; 7146-7161; 8306-8321; 9466-9481; 10479-10485; 10565; 10753-10766
HSD17B4/RPS9	425; 2522-2537; 3682-3697; 4842-4857; 6002-6017; 7162-7177; 8322-8337; 9482-9497; 10486-10492; 10566; 10767-10780
ATP5A1/Gnas	9498-9609; 10493-10499; 10567; 10781-10794
Ndufa4/COX6B1	9610-9721; 10500-10506; 10568; 10795-10808
Ndufa4/Gnas	9722-9833; 10507-10513; 10569; 10809-10822
Ndufa4/Ndufa1	9834-9945; 10514-10520; 10570; 10823-10836
Nosip/Ndufa1	9946-10057; 10521-10527; 10571; 10837-10850
Rpl31/Gnas	10058-10169; 10528-10534; 10572; 10851-10864
TUBB4B/RPS9	10170-10281; 10535-10541; 10573; 10865-10878
Ubqln2/RPS9	10282-10393; 10542-10548; 10574; 10879-10892
Mp68/Gnas1	14526; 14533; 14540
Mp68/Ndufa1	14527; 14534; 14541

[0147] 在一些实施方式中,本发明的编码CRISPR相关蛋白质的人工核酸包括选自下列的至少一种UTR组合:HSD17B4/Gnas.1;Slc7a3.1/Gnas.1;ATP5A1/CASP.1;Ndufa4.1/PSMB3.1;HSD17B4/PSMB3.1;RPL32var/白蛋白7;32L4/白蛋白7;HSD17B4/CASP1.1;Slc7a3.1/CASP1.1;Slc7a3.1/PSMB3.1;Nosip.1/PSMB3.1;Ndufa4.1/RPS9.1;HSD17B4/RPS9.1;ATP5A1/Gnas.1;Ndufa4.1/COX6B1.1;Ndufa4.1/Gnas.1;Ndufa4.1/Ndufa1.1;Nosip.1/Ndufa1.1;Rpl31.1/Gnas.1;TUBB4B.1/RPS9.1;Ubqln2.1/RPS9.1;MP68/Gnas1.1和MP68/Ndufa1.1。

[0148] 在一些实施方式中,根据本发明的人工核酸包括至少一种UTR组合,其选自PCT/EP2017/076775(其整体通过引用并入本文)中关于编码CRISPR相关蛋白质的人工核酸公开的UTR组合。因此,在一些实施方式中,根据本发明的人工核酸可包括选自下列的至少一种UTR组合:SLC7A3/GNAS;ATP5A1/CASP1;HSD17B4/GNAS;NDUFA4/COX6B1;NOSIP/NDUFA1,

NDUFA4/NDUFA1;ATP5A1/GNAS;MP68/NDUFA1;NDUFA4/RPS9;NDUFA4/GNAS;NDUFA4/PSMB3;TUBB4B/RPS9.1;UQBLN2/RPS9;RPL31/GNAS) 或HSD17B4/PSMB3。

[0149] 在一些实施方式中,根据本发明的人工核酸可因此包含PCT/EP2017/076775中关于编码CRISPR相关蛋白质的人工核酸公开的核酸序列或由其组成。

[0150] 表1中通过参考具体SEQ ID NO限定的各UTR元件可包括其变体或片段,该变体或片段呈现与通过参考其具体SEQ ID NO限定的对应核酸序列至少5%、10%、20%、30%、40%、50%、60%、70%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、或99%,优选至少70%,更优选至少80%,甚至更优选至少85%,甚至更优选至少90%,并且最优选至少95%或甚至97%序列同一性。表1的最后一列明确公开了可与“5' UTR”列(包括SEQ ID NO):和“3' UTR”列(包括SEQ ID NO)显示的具体有利UTR组合——即对于技术人员被公开作为优选本发明实施方式的组合——组合的所有可能的Cas9和Cpf1。特别优选的实施方式相像本发明的Cas9或Cpf1序列带有5' UTR SLC7A3 (SEQ ID NO:15/16)或其衍生序列和带有3' UTR GNAS (SEQ ID NO:29/30)或其衍生序列。

[0151] 为便于参考,表A1描述了特别优选的和有利的CDS和UTR组合。

[0152] 表1中通过参考其具体SEQ ID NO确定的各序列也可被相应DNA序列限定,如本文所述。

[0153] 表1中通过参考其具体SEQ ID NO确定的各序列可被修饰(任选地彼此独立地),如下文所述。

[0154] 根据本发明的优选的人工核酸可包括

[0155] a. 衍生自HSD17B4基因的5' UTR或其同源物、片段或变体的至少一个5' UTR元件和衍生自GNAS基因的3' UTR或其同源物、片段或变体的至少一个3' UTR元件;或

[0156] b. 衍生自SLC7A3基因的5' UTR或其同源物、片段或变体的至少一个5' UTR元件和衍生自GNAS基因的3' UTR或其同源物、片段或变体的至少一个3' UTR元件;或

[0157] c. 衍生自ATP5A1基因的5' UTR或其同源物、片段或变体的至少一个5' UTR元件和衍生自CASP1基因的3' UTR或其同源物、片段或变体的至少一个3' UTR元件;或

[0158] d. 衍生自NDUFA4基因的5' UTR或其同源物、片段或变体的至少一个5' UTR元件和衍生自PSMB3基因的3' UTR或其同源物、片段或变体的至少一个3' UTR元件;或

[0159] e. 衍生自HSD17B4基因的5' UTR或其同源物、片段或变体的至少一个5' UTR元件和衍生自PSSLC7A3MB3基因的3' UTR或其同源物、片段或变体的至少一个3' UTR元件;或

[0160] f. 衍生自RPL32基因的5' UTR或其同源物、片段或变体的至少一个5' UTR元件和衍生自ALB基因的3' UTR或其同源物、片段或变体的至少一个3' UTR元件;或

[0161] g. 衍生自HSD17B4基因的5' UTR或其同源物、片段或变体的至少一个5' UTR元件和衍生自CASP1基因的3' UTR或其同源物、片段或变体的至少一个3' UTR元件;或

[0162] h. 衍生自SLC7A3基因的5' UTR或其同源物、片段或变体的至少一个5' UTR元件和衍生自CASP1基因的3' UTR或其同源物、片段或变体的至少一个3' UTR元件;或

[0163] i. 衍生自SLC7A3基因的5' UTR或其同源物、片段或变体的至少一个5' UTR元件和衍生自PSMB3基因的3' UTR或其同源物、片段或变体的至少一个3' UTR元件;或

[0164] j. 衍生自NOSIP基因的5' UTR或其同源物、片段或变体的至少一个5' UTR元件和衍生自PSMB3基因的3' UTR或其同源物、片段或变体的至少一个3' UTR元件;或

[0165] k. 衍生自NDUFA4基因的5' UTR或其同源物、片段或变体的至少一个5' UTR元件和衍生自RPS9基因的3' UTR或其同源物、片段或变体的至少一个3' UTR元件;或

[0166] l. 衍生自HSD17B4基因的5' UTR或其同源物、片段或变体的至少一个5' UTR元件和衍生自RPS9基因的3' UTR或其同源物、片段或变体的至少一个3' UTR元件;或

[0167] m. 衍生自ATP5A1基因的5' UTR或其同源物、片段或变体的至少一个5' UTR元件和衍生自GNAS基因的3' UTR或其同源物、片段或变体的至少一个3' UTR元件;或

[0168] n. 衍生自NDUFA4基因的5' UTR或其同源物、片段或变体的至少一个5' UTR元件和衍生自COX6B1基因的3' UTR或其同源物、片段或变体的至少一个3' UTR元件;或

[0169] o. 衍生自NDUFA4基因的5' UTR或其同源物、片段或变体的至少一个5' UTR元件和衍生自GNAS基因的3' UTR或其同源物、片段或变体的至少一个3' UTR元件;或

[0170] p. 衍生自NDUFA4基因的5' UTR或其同源物、片段或变体的至少一个5' UTR元件和衍生自NDUFA1基因的3' UTR或其同源物、片段或变体的至少一个3' UTR元件;或

[0171] q. 衍生自NOSIP基因的5' UTR或其同源物、片段或变体的至少一个5' UTR元件和衍生自NDUFA1基因的3' UTR或其同源物、片段或变体的至少一个3' UTR元件;或

[0172] r. 衍生自RPL31基因的5' UTR的至少一个5' UTR或其同源物、片段或变体元件和衍生自GNAS基因的3' UTR或其同源物、片段或变体的至少一个3' UTR元件;或

[0173] s. 衍生自TUBB4B基因的5' UTR或其同源物、片段或变体的至少一个5' UTR元件和衍生自RPS9基因的3' UTR或其同源物、片段或变体的至少一个3' UTR元件;或

[0174] t. 衍生自UBQLN2基因的5' UTR或其同源物、片段或变体的至少一个5' UTR元件和衍生自RPS9基因的3' UTR或其同源物、片段或变体的至少一个3' UTR元件;

[0175] u. 衍生自MP68基因的5' UTR或其同源物、片段或变体的至少一个5' UTR元件和衍生自GNAS基因的3' UTR或其同源物、片段或变体的至少一个3' UTR元件;或

[0176] v. 衍生自MP68基因的5' UTR或其同源物、片段或变体的至少一个5' UTR元件和衍生自NDUFA1基因的3' UTR或其同源物、片段或变体的至少一个3' UTR元件。

[0177] 特别优选的人工核酸可包括根据d、e、g或l的UTR组合。

[0178] 在一些实施方式中,根据本发明的人工核酸可不包括衍生自ALB基因的3' UTR或其同源物、片段或变体的3' UTR元件。

[0179] 编码序列

[0180] CRISPR相关蛋白质

[0181] 根据本发明的人工核酸包括至少一种编码序列编码CRISPR相关蛋白质。

[0182] 术语“CRISPR相关蛋白质”指代作为原核生物用以赋予抵抗外源DNA元件的适应性免疫的CRISPR (成簇规律间隔的短回文重复序列) 系统 (和其同源物、变体、片段或衍生物) 的部分的、RNA向导的核酸内切酶。CRISPR相关蛋白质无限制地包括Cas9、Cpf1 (Cas12)、C2c1、C2c3、C2c2、Cas13、CasX和CasY。如本文所用,术语“CRISPR相关蛋白质”包括野生型蛋白质以及其同源物、变体、片段和衍生物。因此,当提及编码Cas9、Cpf1 (Cas12)、C2c1、C2c3、和C2c2、Cas13、CasX和CasY的人工核酸分子时,所述人工核酸分子可编码对应的野生型蛋白质、或其同源物、变体、片段和衍生物。

[0183] CRISPR相关蛋白质可被任何基因、或其同源物、变体或片段编码。当提及基因时,术语“同源物”或“同源基因”包括“直系同源基因”和“旁系同源基因”。



[0184] CRISPR相关蛋白质(和其同源物、变体、片段或衍生物)优选是功能性的,即呈现期望的生物学性质,和/或发挥期望的生物学功能。所述生物学性质或生物学功能可以是相容的或甚至是与相应参考(或“母体”)蛋白质相比增强的。功能性CRISPR相关蛋白质和其同源物、变体、片段或衍生物优选地保留以序列特异性方式被向导RNA靶向至靶DNA序列的能力。然而,野生型CRISPR相关蛋白质一般发挥的核酸内切酶活性(即将DSB引入靶DNA序列的能力)可以,但不一定,保留在本文所述的CRISPR相关蛋白质的所有“功能性”同源物、变体、片段和衍生物中。

[0185] 具体地,功能性同源物、变体、片段或衍生物优选能够(1)与靶DNA序列特异性地相互作用,(2)与适当的向导RNA缔合,和任选地(3)识别与靶DNA序列并置的前间隔序列相邻基序(PAM)。在此环境下,“与……相互作用”优选意为结合,和任选地(进一步)切割(通过核酸内切酶或切口酶活性),激活或抑制表达,和/或招募效应子。“特异性地”意为CRISPR相关蛋白质与靶DNA序列相互作用比其与非靶DNA序列相互作用更容易。

[0186] 当本文提及具体的CRISPR相关蛋白质(如Cas9,Cpf1)时,对应的蛋白质将被理解为包括其所有翻译后修饰形式。翻译后修饰(PTM)可导致给定蛋白质的共价或非共价修饰。常见的翻译后修饰包括糖基化、磷酸化、泛素化、S-亚硝基化、甲基化、N-乙酰化、脂化、二硫键形成、硫酸化、酰化、脱氨等。不同的PTM可产生例如不同的化学、活性、定位、相互作用或构象。然而,本发明的环境中设想的所有翻译后修饰的CRISPR相关蛋白质优选地保持功能性,如上限定。

[0187] 同源物

[0188] 本文示例的各CRISPR相关蛋白质(如Cas9,Cpf1)优选地还包括同源物。当提及蛋白质,术语“同源物”包括“直系同源物”(或“直系同源蛋白质”)和旁系同源物(或“旁系同源蛋白质”)。在此环境下,“直系同源物”是不同物种中的通过物种形成由相同祖先基因演化的基因所编码的蛋白质。直系同源物通常在演化过程中保留相同功能(一种或多种)。因此,在比较一对直系同源物时,功能可能丧失或获得。然而,在本发明的环境中,直系同源CRISPR相关蛋白质优选地保留其与适当的向导RNA缔合以与靶DNA序列特异性地相互作用的能力(即,是“功能性”的)。“旁系同源物”是基因组中通过基因复制产生的基因。旁系同源物一般演化新功能或可最终变成假基因。在本发明的环境中,旁系同源CRISPR相关蛋白质优选是功能性的,如上限定。

[0189] 变体

[0190] 本文示例的各CRISPR相关蛋白质(如Cas9,Cpf1)优选还包括其变体。

[0191] 本文涉及蛋白质所用的术语“变体”优选指代“序列变体”,即包括与参考(或“母体”)蛋白质的参考(或“母体”)氨基酸序列有至少一种氨基酸残基区别的氨基酸序列的蛋白质。

[0192] 变体蛋白质可因此优选地与其对应的参考序列相比在其氨基酸序列中包括至少一种氨基酸突变、取代、插入或缺失。取代可选自保守或非保守取代。在一些实施方式中,优选本发明人工核酸的至少一种编码序列编码的蛋白质“变体”包含至少一种保守氨基酸取代,其中源自相同类别的氨基酸被交换为另一种。具体地,这些是具有脂族侧链、正电荷或负电荷侧链、侧链或氨基酸中的芳族基团、其侧链可以形成氢桥(例如具有羟基官能团的侧链)的氨基酸。通过保守构成,例如具有极性侧链的氨基酸可以被具有相应极性侧链的另一

种氨基酸取代,或者,例如,具有疏水侧链的氨基酸可以被具有相应疏水侧链的另一种氨基酸取代(例如,丝氨酸(苏氨酸)被苏氨酸(丝氨酸),或亮氨酸(异亮氨酸)被异亮氨酸(亮氨酸)取代)。

[0193] 优选地,如本文所用的术语“变体”包括天然存在的变体,例如,已经进行翻译后蛋白水解加工(这可能涉及去除N-末端甲硫氨酸、信号肽和/或无活性或非功能性蛋白转化为活性蛋白或功能性蛋白)的前蛋白原、前蛋白和CRISPR相关蛋白质和天然存在的突变蛋白。术语“变体”还包括CRISPR相关蛋白质的工程变体,其可以被(序列)修饰以引入或消除某种生物学特性和/或功能。下面详细讨论了工程Cas9变体。术语“转录物变体”或“剪接变体”在蛋白质的环境中是指由信使RNA产生的变体,其最初从相同基因转录,但随后进行可选的(或差异的)剪接,其中基因的特定外显子可以被包括或不包括在最终的加工的信使RNA(mRNA)中。然而,CRISPR相关蛋白质的“转录物变体”优选保留其期望的生物学功能,如上限定。应注意,术语“变体”可本质上以与参考蛋白质相比最小程度的序列同一性(并且优选以及期望的生物学功能/性质)的方式来限定。因此,同源物、片段或某些衍生物(在其氨基酸序列方面也不同于参考蛋白质)也可以归类为“变体”。因此,如本文所限定的“变体”可以衍生自,分离自,相关于,基于参考蛋白质或与参考蛋白质同源,其可以是如本文所限定的CRISPR相关蛋白质参考蛋白质(如Cas9,Cpf1)、或其同源物、片段变体或衍生物。

[0194] 根据本发明的CRISPR相关蛋白质(如Cas9,Cpf1)变体优选与对应的天然存在的(野生型)CRISPR相关蛋白质(如Cas9,Cpf1)或其同源物、片段或衍生物的氨基酸序列具有至少5%、10%、20%、30%、40%、50%、60%、70%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、或99%,优选至少70%,更优选至少80%,甚至更优选至少85%,甚至更优选至少90%,并且最优选至少95%或甚至97%的序列同一性。

[0195] 片段

[0196] 本文示例的各CRISPR相关蛋白质(如Cas9,Cpf1)优选还包括其片段。

[0197] 术语“片段”是指由参考(或“母体”)蛋白质或(多)肽的全长氨基酸序列的连续子序列组成的蛋白质或多肽——就其氨基酸序列而言,与所述参考蛋白质的氨基酸序列相比N端、C端和/或序列内截短。这种截短可分别在氨基酸水平或核酸水平上发生。换句话说,“片段”一般可以是氨基酸序列的全长序列的较短部分。因此,片段一般由与全长氨基酸序列内的相应段相同的序列组成。该术语包括天然存在的片段(如由天然存在的体内蛋白酶活性产生的片段)以及工程片段。

[0198] 如本文所用的术语“片段”可指代肽或多肽包括本文限定的CRISPR相关蛋白质或其同源物、变体或衍生物的氨基酸序列的至少5个连续氨基酸残基、至少10个连续氨基酸残基、至少15个连续氨基酸残基、至少20个连续氨基酸残基、至少25个连续氨基酸残基、至少40个连续氨基酸残基、至少50个连续氨基酸残基、至少60个连续氨基酸残基、至少70个连续氨基酸残基、至少连续80个氨基酸残基、至少连续90个氨基酸残基、至少连续100个氨基酸残基、至少连续125个氨基酸残基、至少150个连续氨基酸残基、至少连续175个氨基酸残基、至少连续200个氨基酸残基、或至少连续250个氨基酸残基的氨基酸序列。

[0199] 在本发明的环境中优选的序列片段由与该片段的来源蛋白质中连续一段实体相应的连续一段氨基酸组成——其代表该片段的来源总(即全长)蛋白质或(多)肽的至少

20%，优选至少30%，更优选至少40%，更优选至少50%，甚至更优选至少60%，甚至更优选至少70%、和最优选至少80%。

[0200] 关于这种片段所述的序列同一性优选说的是参考蛋白质的整个氨基酸序列或编码所述参考蛋白质的整个核酸序列。优选地，CRISPR相关蛋白质(如Cas9或Cpf1)或其同源物、变体或衍生物的“片段”可一般包括与所述CRISPR相关蛋白质(例如Cas9，Cpf1)或所述同源物、变体或衍生物的氨基酸序列具有至少5%、10%、20%、30%、40%、50%、60%、70%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、或99%，优选至少70%，更优选至少80%，甚至更优选至少85%，甚至更优选至少90%，并且最优选至少95%或甚至97%序列同一性的氨基酸序列。

[0201] 衍生物

[0202] 本文示例的各CRISPR相关蛋白质(如Cas9，Cpf1)优选还包括其衍生物。

[0203] 在涉及蛋白质时，术语“衍生物”将被理解为相对于参考(或“母体”)蛋白质而言已被修饰以包括新的或另外的性质或功能的蛋白质。衍生物可被修饰以包括期望的生物学功能(例如通过引入或去除赋予、增强、降低或消除靶标结合亲和力或特异性或酶活性的部分或结构域)、制造性质(例如通过引入赋予溶解度增加或分泌增强或允许纯化的部分)或用于医学用途的药代动力学/药效动力学特性(例如通过引入赋予稳定性增加、生物利用度，吸收；分布和/或清除性降低的部分)。衍生物可以通过引入或去除赋予目的生物学特性或功能的部分或结构域来制备。这种部分或结构域可以利用标准基因工程技术(Sambrook J et al., 2012 (4<sup>th</sup>ed.), Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, 纽约)在翻译后或在核酸序列水平上被引入氨基酸序列(例如在氨基和/或羧基末端残基处)。“衍生物”可以衍生自(并因此任选地包括)天然存在的(野生型)CRISPR相关蛋白质序列或其变体或片段。

[0204] 应当理解，CRISPR相关蛋白质衍生物其氨基酸序列可以与其来源参考蛋白质不同(例如通过引入或去除(多)肽部分和/或蛋白质结构域)，因此可以也符合“变体”。然而，序列变体主要根据其参考氨基酸序列的序列同一性来限定，但衍生物优选特征在于与参考蛋白质相比存在或不存在特定的生物学特性或功能。

[0205] 很多CRISPR相关蛋白质衍生物基于对应天然存在的(野生型)CRISPR相关蛋白质的变体、片段、片段变体或变体片段。例如，根据本发明的CRISPR相关蛋白质衍生物可包括基于包含消除内切核酸酶和/或切口酶活性的突变的、已被进一步工程化以包括赋予新的或另外的生物学性质或功能的效应或衔接结构域的工程蛋白质变体的衍生物。

[0206] 因此，在优选的实施方式中，本发明的人工核酸分子编码本文限定的CRISPR相关蛋白质(例如Cas9，Cpf1)衍生物，其中所述衍生物包括至少一种进一步的效应结构域。

[0207] “效应结构域”将被理解为赋予另外的和/或新的生物学性质或功能的蛋白质部分。在本发明的环境中，效应结构域可以基于其赋予CRISPR相关蛋白质(新的或另外的)生物学功能的能力来选择，优选地不干扰其与合适的向导RNA缔合以与靶DNA序列特异性相互作用的能力。新的或另外的生物学功能可以是，例如，转录抑制(引起CRISPR干扰，CRISPRi)或激活(引起CRISPR激活，CRISPRa)。因此，在本发明的环境中目的效应结构域可以选自转录抑制结构域，包括Krüppel缔合盒(KRAB)结构域、MAX相互作用蛋白1(MXI1)结构域、四个连接的mSin3(SID4X)结构域；或转录激活结构域，包括单纯疱疹病毒VP16激活结构域(VP64

或VP160)、核因子- $\kappa$ B (NF- $\kappa$ B) 反式激活亚基激活结构域 (p65AD)。这种效应结构域(一种或多种)可以与CRISPR相关蛋白质的氨基(N-)或羧基(C-)末端或两者融合。在合适的效应结构域下,转录还可以在表观遗传水平上被调控。组蛋白去甲基化酶LSD1从目标远端增强子中去除组蛋白3 Lys4二甲基化(H3K4me2)标记,导致转录抑制。组蛋白乙酰转移酶p300的催化核心(p300核心)可以在目标近端和远端增强子处乙酰化H3K27 (H3K27ac),这导致转录激活。另外地或可选地,新的或另外的生物学功能可以是招募目的效应结构域。为此,效应结构域可以是“招募结构域”,优选蛋白质-蛋白质相互作用结构域或基序,如WRPW (Trp-Arg-Pro-Trp)基序(Fisher et al. Mol Cell Biol. 1996 Jun; 16(6):2670-7)。另外地或可选地,新的或另外的生物学功能可以是其他目的实体如RNA的招募。为此,效应结构域可以选自蛋白质-RNA相互作用结构域,如冷休克结构域(CSD)。

[0208] 因此,包含效应结构域的CRISPR相关蛋白质衍生物包括(a)与能够与靶DNA序列(或与其可操作地连接的调控元件)相互作用的效应结构域直接融合(任选地通过合适的连接体)的CRISPR相关蛋白质(或其同源物、变体、片段);和(b)与招募其他目的效应子(结构域、蛋白质或核酸)融合(任选地通过合适的连接体)的效应结构域融合的蛋白质(或其同源物、变体、片段),该其他目的效应子进而能够与靶DNA序列(或与其可操作地连接的调控元件)相互作用。

[0209] 效应结构域可以利用标准遗传工程技术(参见Sambrook J et al., 2012 (4<sup>th</sup> ed.), Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, 纽约)与CRISPR相关蛋白质(或其变体或片段)融合,任选地通过合适的(肽)连接体。

[0210] 目的肽连接体是本领域公知的,并且可以分为三种类型:柔性连接体、刚性连接体和可切割连接体。柔性连接体通常在连接的结构域需要一定程度的运动或相互作用时应用,因此在本发明的CRISPR相关蛋白质衍生物的环境下特别受关注。其总体上富含小的非极性(例如Gly)或极性(例如Ser或Thr)氨基酸,以提供良好的柔性和溶解性,并允许连接的蛋白质结构域移动。Ser或Thr的并入可通过与水分子形成氢键来维持连接体在水溶液中的稳定性,并因此减少连接体与蛋白质部分之间的不利相互作用。

[0211] 最常用的柔性连接体具有主要由Gly和Ser残基段(“GS”连接体)组成的序列。最广泛使用的柔性连接体的实例具有(Gly-Gly-Gly-Gly-Ser)<sub>n</sub>的序列。通过调整拷贝数“n”,可以优化该GS连接体的长度以实现蛋白质结构域的适当分离,或维持必要的结构域间相互作用。除了GS连接体外,很多其他柔性连接体也已被设计用于重组融合蛋白。这些柔性连接体也富含小的或极性的氨基酸如Gly和Ser,但可以含有其他氨基酸如Thr和Ala以保持韧性,以及极性氨基酸如Lys和Glu以提高溶解度。

[0212] 几种其他类型的柔性连接体,包括KESGSVSSEQLAQFRSLD、EGKSSGSGSESKST和GSAGSAAGSGEF已被应用于构建融合蛋白。其他柔性连接体包括单独甘氨酸连接体(Gly)<sub>6</sub>或(Gly)<sub>8</sub>。

[0213] 刚性连接体可以在要确保蛋白质结构域分离和其干扰减少时使用。另一方面,可切割连接体可以被引入以在体内释放游离的功能性结构域。陈等人。Chen et al. Adv Drug Deliv Rev. 2013 Oct 15; 65(10):1357-1369综述了最常用的肽连接体及其应用,并其整体通过引用并入本文。

[0214] 除了将期望的效应结构域与CRISPR相关蛋白质融合外,还有几种替代方法用于介导对目的靶序列的期望的生物效应。这些方法本质上利用了能够招募靶DNA序列的目的效应结构域的CRISPR相关蛋白质(或向导RNA)。这些方法可以为多重招募具体目的靶DNA序列的效应结构域(例如启动子或增强子)提供另外的选择和灵活性。

[0215] SunTag激活法利用融合至CRISPR相关蛋白质(例如dCas9)的一系列小肽表位标签来招募多拷贝的单链可变片段(scFV),该单链可变片段(scFV)融合至超折叠GFP(sfGFP;用于改善蛋白质折叠),融合至(a)效应结构域(一种或多种)例如VP64。协同三分(tripartite)激活法(VPR)利用三个效应结构域(例如转录激活因子,VP64、p65和Epstein-Barr病毒R反式激活因子(Rta))的串联融合,以赋予期望的生物学功能。基于适体的招募系统(协同激活介质(SAM))利用CRISPR相关蛋白质(例如dCas9)与具有两个结合位点的向导RNA(例如,四环和第二茎-环上的RNA适体)招募与效应结构域(例如转录激活因子如p65和热休克因子1(HSF1))融合的噬菌体MS2外壳蛋白(MCP)。另外,可以将其他效应结构域(例如VP64)与CRISPR相关蛋白质融合,产生根据本发明的衍生物。

[0216] 上述激活方法可以容易地适应于赋予如本文所述的CRISPR相关蛋白质或其同源物、变体、片段或衍生物以转录抑制功能或其他期望的生物学功能。CRISPR相关蛋白质衍生物和介导CRISPRa和CRISPRi的各种方法在Dominguez et al. Nat Rev Mol Cell Biol. 2016 Jan; 17(1): 5-15中被综述,其全部内容通过引用并入本文。

[0217] 信号肽

[0218] 在一些实施方式中,本发明的人工核酸分子,优选地RNA,包括编码信号肽的至少一种核酸序列。所述核酸序列优选地位于本发明人工核酸分子的编码区(编码CRISPR相关蛋白质)内。因此,本发明的人工核酸分子,优选地RNA,可优选包括编码包括至少一种信号肽的本文限定的CRISPR相关蛋白质或其同源物、变体、片段或衍生物的编码区。

[0219] 信号肽(有时称为信号序列、靶向信号、定位信号、定位序列、转运肽、前导序列或前导肽)一般是短(5-30个氨基酸长)肽,优选地位于被编码的CRISPR相关蛋白质(或其同源物、变体、片段或衍生物)的N端。

[0220] 信号肽优选介导被编码的CRISPR相关蛋白质(或其同源物、变体、片段或衍生物)转运到限定的细胞区室中,例如细胞表面、内质网(ER)或内体-溶酶体区室。因此,信号肽可用于促进被表达的蛋白质从生产细胞系分泌。

[0221] 在本发明的环境中设想的示例性信号肽包括但不限于经典或非经典MHC分子的信号序列(例如MHC I和II分子的信号序列,例如MHC I类分子HLA-A\*0201)、细胞因子或免疫球蛋白的信号序列、免疫球蛋白或抗体的恒定链的信号序列、Lamp1、Tapasin、Erp57、Calretikulin、Calnexin、PLAT、EPO或白蛋白以及其他膜相关蛋白或内质网(ER)或内体-溶酶体区室相关蛋白质的信号序列。最优选地,信号序列衍生自(人)HLA-A2、(人)PLAT、(人)sEPO、(人)ALB、(人)IgE-前导序列、(人)CD5、(人)IL2、(人)CTRB2、(人)IgG-HC、(人)Ig-HC、(人)Ig-LC、GpLuc、(人)Igkappa或任何前述蛋白质的片段或变体,特别是HLA-A2、HsPLAT、sHsEPO、HsALB、HsPLAT(aa1-21)、HsPLAT(aa1-22)、IgE-前导序列、HsCD5(aa1-24)、HsIL2(aa1-20)、HsCTRB2(aa1-18)、IgG-HC(aa1-19)、Ig-HC(aa1-19)、Ig-LC(aa1-19)、GpLuc(1-17)或MmIgkappa。本发明设想前述信号序列或其变体或片段的应用,只要这些变体或片段是功能性的,即能够将CRISPR相关蛋白质靶向细胞(内或外)的目的位置。

[0222] 编码所述信号肽的核酸序列优选通过标准遗传工程技术(参见Sambrook J et al.,2012(4th ed.),Molecular cloning:a laboratory manual.Cold Spring Harbor Laboratory,Cold Spring Harbor,纽约)与本发明人工核酸的编码区中编码CRISPR相关蛋白质(或其同源物、变体、片段或衍生物)的核酸序列融合。所述编码区的表达优选产生包含(优选在其N端、C端或两者)被编码信号肽的CRISPR相关蛋白质。

[0223] 核定位序列(NLS)

[0224] 本发明的人工核酸分子,优选地RNA,可优选地进一步包括编码至少一种核定位序列(NLS)的核酸序列。所述核酸序列优选地位于本发明人工核酸分子的编码区(编码CRISPR相关蛋白质)中。因此,本发明的人工核酸分子,优选地RNA,可优选地包括编码本文限定的CRISPR相关蛋白质或其同源物、变体、片段或衍生物的,包括至少一种核定位序列(NLS)的编码区。

[0225] 核定位信号或序列(NLS)是介导核蛋白质转运到细胞核中的小段氨基酸。由于本发明的人工核酸编码的CRISPR相关蛋白质具体地被设想用于哺乳动物细胞的治疗和研究,其可以被赋予至少一种NLS以使其能够进入细胞核,在此其可以对基因组DNA发挥其作用。NLS优选与核膜中的核孔复合物(NPC)相互作用,从而促进CRISPR相关蛋白质转运到细胞核中。

[0226] 多种NLS序列是本领域已知的,并且其根据本发明的使用(或使用调整)在本领域技术人员的普通技能和知识范围内。最佳表征的转运信号是用于核蛋白输入的经典NLS(cNLS),其由一段(单分)或两段(二分)基本氨基酸组成。一般,单分基序的特征在于碱性残基簇前接螺旋破坏残基。类似地,二分基序由两个被9-12个残基隔开的碱性残基簇组成。单分cNLS的示例是SV40大T抗原NLS(<sup>126</sup>PKKKRRV<sup>132</sup>;SEQ ID NO:381),并且二分cNLS的示例是核质蛋白NLS(<sup>155</sup>KRPAATKKAGQAKKKK<sup>170</sup>;SEQ ID NO:382)。来自单分NLS的N端赖氨酸的连续残基被称为P1、P2等。单分cNLS一般需要在P1位的赖氨酸,然后是P2和P4位的碱性残基以产生K(K/R)X(K/R)的松散共识序列(SEQ ID NO:384;Lange et al.,J Biol Chem.2007 Feb 23;282(8):5101-5105)。

[0227] 因此设想根据优选的实施方式,人工核酸分子进一步包括编码核定位信号(NLS)的至少一种核酸序列。根据本发明的人工核酸分子可因此编码1、2、3、4、5或更多个NLS,其任选地选自本文示例的NLS。所述NLS编码核酸序列优选位于本发明人工核酸的编码区中,并且优选地融合至编码CRISPR相关蛋白质的核酸序列,使得所述编码区的表达产生包括所述至少一种NLS(优选在其N端、C端、或两者)的CRISPR相关蛋白质。换句话说,根据本发明的人工核酸分子可优选地编码包括至少一种NLS(优选在其N端、C端、或两者)的CRISPR相关蛋白质。

[0228] 根据本发明的适当的NLS可包括下列或由下列组成:根据SEQ ID NO:426(MAPKKKRKVGIGHGVPA)的氨基酸序列,在本文中也称为NLS2,其可被根据SEQ ID NO:409;2538,1378;3698;4858;6018;7178;或8338中的任一者的核酸序列、或这些序列中任意者的(功能性)变体或片段,具体地按递增优选顺序与那些序列中的任意者具有至少5%、10%、20%、30%、40%、50%、60%、70%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、或99%,优选至少70%,更优选至少80%,甚至更优选至少85%,甚至更优选至少90%,并且最优选至少95%或甚至97%序列同一性的核酸序列编

码。本发明进一步设想NLS2变体或片段的应用,条件是这些变体和片段优选是功能性的,即能够介导CRISPR相关蛋白质输入细胞核中。这种功能性变体或片段可包括下列或由下列组成:按递增优选顺序与根据SEQ ID NO:426的氨基酸序列具有至少5%、10%、20%、30%、40%、50%、60%、70%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、或99%,优选至少70%,更优选至少80%,甚至更优选至少85%,甚至更优选至少90%,并且最优选至少95%或甚至97%序列同一性的氨基酸序列。

[0229] 根据本发明的其他适当的NLS可包括下列或由下列组成:根据SEQ ID NO:427 (KRPAATKKAGQAKKKK) 的氨基酸序列,在本文中也称为NLS4,其可被根据SEQ ID NO:410; 2539;1379;3699;4859;6019;7179;8339的核酸序列或这些序列中任意者的(功能性)变体或片段,具体地按递增优选顺序与这些序列中的任意者具有至少5%、10%、20%、30%、40%、50%、60%、70%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、或99%,优选至少70%,更优选至少80%,甚至更优选至少85%,甚至更优选至少90%,并且最优选至少95%或甚至97%序列同一性的核酸序列编码。本发明进一步设想NLS4变体或片段的应用,条件是这些变体和片段优选是功能性的,即能够介导CRISPR相关蛋白质输入细胞核中。这种功能性变体或片段可包括下列或由下列组成:按递增优选顺序与根据SEQ ID NO:427的氨基酸序列具有至少5%、10%、20%、30%、40%、50%、60%、70%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、或99%,优选至少70%,更优选至少80%,甚至更优选至少85%,甚至更优选至少90%,并且最优选至少95%或甚至97%序列同一性的氨基酸序列。

[0230] 根据本发明的其他适当的NLS可包括下列或由下列组成:根据SEQ ID NO:427 (KRPAATKKAGQAKKKK) 的氨基酸序列,在本文中也称为NLS4,其可被根据SEQ ID NO:410; 2539;1379;3699;4859;6019;7179;8339的核酸序列、或这些序列中任意者的(功能性)变体或片段,具体地按递增优选顺序与这些序列中的任意者具有至少5%、10%、20%、30%、40%、50%、60%、70%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、或99%,优选至少70%,更优选至少80%,甚至更优选至少85%,甚至更优选至少90%,并且最优选至少95%或甚至97%序列同一性的核酸序列编码。本发明进一步设想NLS4变体或片段的应用,条件是这些变体和片段优选是功能性的,即能够介导CRISPR相关蛋白质输入细胞核中。这种功能性变体或片段可包括下列或由下列组成:按递增优选顺序与根据SEQ ID NO:427的氨基酸序列具有至少5%、10%、20%、30%、40%、50%、60%、70%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、或99%,优选至少70%,更优选至少80%,甚至更优选至少85%,甚至更优选至少90%,并且最优选至少95%或甚至97%序列同一性的氨基酸序列。

[0231] 本文进一步设想使CRISPR相关蛋白质装备两个或更多个NLS,并且这些NLS可例如选自NLS2(以SEQ ID NO:426表征)和NLS4(以SEQ ID NO:427表征)或这些核定位信号中一者或两者的功能性变体或片段。

[0232] 根据本发明的其他适当的NLS可包括下列或由下列组成:根据SEQ ID NO:10575 (KRPAATKKAGQAKKKK) 的氨基酸序列,在本文中也称为NLS3,其可被根据SEQ ID NO:410; 2539;1379;3699;4859;6019;7179;8339,10551;10581,10593;10584;10587;10590;10593; 10596的核酸序列、或这些序列中任意者的(功能性)变体或片段,具体地按递增优选顺序与

这些序列中的任意者具有至少5%、10%、20%、30%、40%、50%、60%、70%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、或99%，优选至少70%，更优选至少80%，甚至更优选至少85%，甚至更优选至少90%，并且最优选至少95%或甚至97%序列同一性的核酸序列编码。本发明进一步设想NLS3变体或片段的应用，条件是这些变体和片段优选是功能性的，即能够介导CRISPR相关蛋白质输入细胞核中。这种功能性变体或片段可包括下列或由下列组成：按递增优选顺序与根据SEQ ID NO:427的氨基酸序列具有至少5%、10%、20%、30%、40%、50%、60%、70%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、或99%，优选至少70%，更优选至少80%，甚至更优选至少85%，甚至更优选至少90%，并且最优选至少95%或甚至97%序列同一性的氨基酸序列。

[0233] 因此，进一步优选的NLS序列可包括下列或由下列组成：根据SEQ ID NO:426;427;10575;381;382;384;11957;11958-11964的氨基酸序列，其可被根据409;2538;410;2539;10551;10581;11973;11974-1198,1378;3698;4858;6018;7178;8338;1379;3699;4859;6019;7179;8339;10593;10584;10587;10590;10593;10596;11965;11981;11989;11997;12005;12013;11966-11972;11982-11988;11990-11996;11998-12004;12006-12012;或12014-12020的核酸序列编码。

[0234] 进一步优选的NLS可包括下列或由下列组成：根据SEQ ID NO:12021-14274的氨基酸序列。

[0235] 如上所述的NLS序列是常用和被接受的NLS的非限制性列举。应理解，任何本文提到的基因编辑酶，例如Cas9或Cpf1，可以与本领域已知的任何其他NLS组合，并且序列中有任何数量的NLS序列。不同NLS序列的组合也被涵盖在本发明的上述公开内容中。

[0236] 本发明的教导内还包括以任何NLS数量和/或组合(即5'/3'NLS)包含本文公开的或本领域已知的NLS的、编码本文公开的基因编辑蛋白的序列或序列表中的序列。

[0237] 蛋白质和肽标签

[0238] 在一些实施方式中，本发明的人工核酸分子，优选地RNA，进一步包括至少一种核酸序列编码蛋白质或肽标签。所述核酸序列优选地位于本发明人工核酸分子的编码区(编码CRISPR相关蛋白质)中。因此，本发明的人工核酸分子，优选地RNA，可优选地包括编码包括至少一种蛋白质或肽标签的本文限定的CRISPR相关蛋白质或其同源物、变体、片段或衍生物的编码区。

[0239] 蛋白质和肽标签是可被引入目的蛋白质中以能够实现纯化、检测、定位或其他目的的氨基酸序列。蛋白质和肽标签可以基于其功能进行分类，无限制地包括亲和标签(如几丁质结合蛋白(CBP)、麦芽糖结合蛋白(MBP)、谷胱甘肽-S-转移酶(GST)、聚(His)标签、Fc标签、Strep标签)、增溶标签(如硫氧还蛋白(TRX)和聚(NANP))、色谱标签(如FLAG标签)、表位标签(V5标签、Myc标签、HA标签和NE标签)、荧光标签(如GFP标签)或其他(Av标签，允许生物素化和后续的分选)。

[0240] 根据本发明的人工核酸分子可因此编码1、2、3、4、5或更多种蛋白质或肽标签，其任选地选自本文示例的蛋白质标签。所述蛋白质或肽标签编码核酸序列优选位于本发明人工核酸分子的编码区中，并且优选融合至编码CRISPR相关蛋白质的核酸序列，使得所述编码区的表达产生包括所述至少一种蛋白质或肽标签的CRISPR相关蛋白质。换句话说，根据



本发明的人工核酸分子可优选地编码包括至少一种蛋白质或肽标签的CRISPR相关蛋白质。引入编码这种蛋白质或肽标签的核酸的手段和方法在本领域技术人员的技能和公知常识范围内。

[0241] 本发明的人工核酸分子,优选地RNA,可编码呈现上述任何特征或特点的CRISPR相关蛋白质(如Cas9,Cpf1)——如合适或必要,该特征或特点彼此任何组合,然而条件是组合的特征或特点不彼此干扰。因此,人工核酸分子,具体地RNA,可编码本文示例的任何CRISPR相关蛋白质、或本文限定的其同源物、变体、片段或衍生物,其可包括一种或多种NLS,和任选地一种或多种信号序列和/或蛋白质或肽标签,条件是编码的CRISPR相关蛋白质(和NLS、信号序列、蛋白质/肽标签)优选地保留其期望的生物学功能或性质,如上限定。

[0242] 本发明的教导还包括具有蛋白质或肽标签而所述标签序列(一种或多种)没有被引入用于纯化、检测、定位或其他目的的情况下的所有序列。换句话说,当标签序列被除去时,序列表中公开的具有肽或蛋白质标签的序列也明确被包括在本发明的教导内。技术人员容易能够从带标签的蛋白质序列除去任何标签序列,即还使用无标签形式的本发明序列。对于聚C和组蛋白茎环序列也是如此——如需,其可以容易地从蛋白质除去或者也可以添加到蛋白质。

[0243] Cas9

[0244] “Cas9”是指可以由化脓性链球菌(*Streptococcus pyogenes*)血清型M1 cas9基因(NCBI参考序列:NC\_002737.2,“SPy1046”;化脓链球菌)编码的、RNA向导的II型CRISPR-Cas DNA核酸内切酶,即spCas9,或其同源物、变体或片段。Cas9可以优选地被向导RNA(gRNA)招募,以利用两个不同的核酸内切酶结构域(HNH和RuvC/RNA酶H样结构域)位点特异性地切割靶DNA序列——DNA双螺旋每条链一个。RuvC和HNH一起产生双链断裂(DSB),并且单独可以产生单链断裂(美国公开专利申请号2014-0068797和Jinek M.,et al.Science.2012 Aug 17;337(6096):816-21)。Cas9优选能够特异性识别(并且优选结合)与靶DNA序列并置的前间隔序列相邻基序(PAM)。PAM一般位于靶DNA的3',并且可包含三核苷酸序列NGG或由三核苷酸序列NGG组成。其一般被位于Cas9的C末端附近的PAM相互作用结构域(PI结构域)识别。

[0245] 大量Cas9蛋白质是本领域已知的,并且在本发明的环境中被设想为CRISPR相关蛋白质。合适的Cas9蛋白列于下表2中。其中,每行相应于相应蛋白质的数据库登录号所标识的Cas9蛋白(第一列,“A”,“登录号”)。表2中的第二列(“B”)表示相应于本文提供的对应氨基酸序列的SEQ ID NO。优选的Cas9蛋白显示在序列表中的SEQ ID NO:428-441;SEQ ID NO:10999-11001;和SEQ ID NO:442-1345。作为本发明优选实施方式的相应优化mRNA序列显示在序列表中的SEQ ID NO:411;2540-2553;11117-11119;11355-11357;2554-3457;1380-1393;3700-3713;4860-4873;6020-6033;7180-7193;8340-8353;11237-11239;11473-11475;11591-11593;11709-11711;11827-11829;11945-11947;1394-2297;3714-4617;4874-5777;6034-6937;7194-8097;和8354-9257。

[0246] 氨基酸序列

[0247] 表2:本发明的Cas9蛋白

[0248]

行	A 列 蛋白质登录号	蛋白质 (Cas9/Cpf1)	B 列 SEQ ID NO
1	Q99ZW2	Cas9_Q99ZW2_prot	428
2	A0Q5Y3	Cas9_A0Q5Y3_prot	429
3	J7RUA5	Cas9_J7RUA5_prot	430
4	G3ECR1	Cas9_G3ECR1_prot	431
5	J3F2B0	Cas9_J3F2B0_prot	432
6	Q03JI6	Cas9_Q03JI6_prot	433
7	C9X1G5	Cas9_C9X1G5_prot	434
8	Q927P4	Cas9_Q927P4_prot	435
9	Q8DTE3	Cas9_Q8DTE3_prot	436
10	Q9CLT2	Cas9_Q9CLT2_prot	437
11	A1IQ68	Cas9_A1IQ68_prot	438
12	Q6NKI3	Cas9_Q6NKI3_prot	439
13	Q0P897	Cas9_Q0P897_prot	440
14	Q03LF7	Cas9_Q03LF7_prot	441
15	T0TDV9	Cas9_T0TDV9_prot	442
16	A0A0D8BYB2	Cas9_A0A0D8BYB2_prot	443
17	A0A0M4TTU2	Cas9_A0A0M4TTU2_prot	444
18	A7H5P1	Cas9_A7H5P1_prot	445
19	A0A0W8KZ82	Cas9_A0A0W8KZ82_prot	446
20	A0A0E1ZMQ3	Cas9_A0A0E1ZMQ3_prot	447

[0249]

21	W8KE67	Cas9_W8KE67_prot	448
22	A0A0B6V308	Cas9_A0A0B6V308_prot	449
23	A0A1E7PM50	Cas9_A0A1E7PM50_prot	450
24	A0A1E7P6J5	Cas9_A0A1E7P6J5_prot	451
25	A0A1D9BML5	Cas9_A0A1D9BML5_prot	452
26	A5KEK9	Cas9_A5KEK9_prot	453
27	D2MWB9	Cas9_D2MWB9_prot	454
28	A0A0H4KTI1	Cas9_A0A0H4KTI1_prot	455
29	A0A0D7V4T2	Cas9_A0A0D7V4T2_prot	456
30	A0A059HXJ1	Cas9_A0A059HXJ1_prot	457
31	A0A1E7NYV5	Cas9_A0A1E7NYV5_prot	458
32	A0A1E7P943	Cas9_A0A1E7P943_prot	459
33	A0A0E2UY67	Cas9_A0A0E2UY67_prot	460
34	A0A1B3X857	Cas9_A0A1B3X857_prot	461
35	A0A0E9LLC5	Cas9_A0A0E9LLC5_prot	462
36	A0A125S8M1	Cas9_A0A125S8M1_prot	463
37	A0A0S8HUI8	Cas9_A0A0S8HUI8_prot	464
38	A0A0A8GXC3	Cas9_A0A0A8GXC3_prot	465
39	A0A0A8GU36	Cas9_A0A0A8GU36_prot	466
40	A0A139BVD9	Cas9_A0A139BVD9_prot	467
41	A3VED0	Cas9_A3VED0_prot	468
42	A0A0A8HTA3	Cas9_A0A0A8HTA3_prot	469
43	A0A125S8L7	Cas9_A0A125S8L7_prot	470
44	T2LKS6	Cas9_T2LKS6_prot	471
45	A0A0A8H849	Cas9_A0A0A8H849_prot	472
46	F5S4M8	Cas9_F5S4M8_prot	473
47	G1UFN3	Cas9_G1UFN3_prot	474
48	B5ZLK9	Cas9_B5ZLK9_prot	475
49	C5ZYI3	Cas9_C5ZYI3_prot	476
50	A0A0G3EK96	Cas9_A0A0G3EK96_prot	477
51	A0A125S8M5	Cas9_A0A125S8M5_prot	478
52	A0A0L6CQ85	Cas9_A0A0L6CQ85_prot	479
53	A0A0L8B0U9	Cas9_A0A0L8B0U9_prot	480
54	A0A178N1Y8	Cas9_A0A178N1Y8_prot	481
55	A0A125S8I0	Cas9_A0A125S8I0_prot	482
56	A0A0A1PPJ7	Cas9_A0A0A1PPJ7_prot	483
57	B8I085	Cas9_B8I085_prot	484
58	A0A1B8J9V3	Cas9_A0A1B8J9V3_prot	485
59	I7GTK8	Cas9_I7GTK8_prot	486
60	D3UFL8	Cas9_D3UFL8_prot	487
61	E1VQA3	Cas9_E1VQA3_prot	488
62	M4V7E7	Cas9_M4V7E7_prot	489
63	F4GDP9	Cas9_F4GDP9_prot	490
64	A0A0Q6WIJ3	Cas9_A0A0Q6WIJ3_prot	491
65	A0A0E9L8G0	Cas9_A0A0E9L8G0_prot	492

[0250]

66	A0A0A1VBC9	Cas9_A0A0A1VBC9_prot	493
67	B1GZM3	Cas9_B1GZM3_prot	494
68	A0A1C0W3U5	Cas9_A0A1C0W3U5_prot	495
69	D5BR51	Cas9_D5BR51_prot	496
70	A0A1A7NZJ6	Cas9_A0A1A7NZJ6_prot	497
71	A0A125S8J2	Cas9_A0A125S8J2_prot	498
72	A0A0A2YBT2	Cas9_A0A0A2YBT2_prot	499
73	A0A099UFG2	Cas9_A0A099UFG2_prot	500
74	A0A0C5JLX1	Cas9_A0A0C5JLX1_prot	501
75	A7HP89	Cas9_A7HP89_prot	502
76	A0A0J6BUV9	Cas9_A0A0J6BUV9_prot	503
77	A0A1C9ZTA2	Cas9_A0A1C9ZTA2_prot	504
78	A0A087N7M8	Cas9_A0A087N7M8_prot	505
79	A0A0Q9CTQ5	Cas9_A0A0Q9CTQ5_prot	506
80	A0A101I188	Cas9_A0A101I188_prot	507
81	V2Q0I9	Cas9_V2Q0I9_prot	508
82	F9ZKQ5	Cas9_F9ZKQ5_prot	509
83	F0Q2T1	Cas9_F0Q2T1_prot	510
84	M4R7E0	Cas9_M4R7E0_prot	511
85	T1DV82	Cas9_T1DV82_prot	512
86	W0Q6X6	Cas9_W0Q6X6_prot	513
87	A0A0E9MLX9	Cas9_A0A0E9MLX9_prot	514
88	A0A0D6MWC5	Cas9_A0A0D6MWC5_prot	515
89	A0A087MCH0	Cas9_A0A087MCH0_prot	516
90	I3TWJ0	Cas9_I3TWJ0_prot	517
91	A0A011P7F8	Cas9_A0A011P7F8_prot	518
92	A0A163RXL7	Cas9_A0A163RXL7_prot	519
93	A9HKP2	Cas9_A9HKP2_prot	520
94	A0A0N1EBR4	Cas9_A0A0N1EBR4_prot	521
95	A0A0A8HLU7	Cas9_A0A0A8HLU7_prot	522
96	E1W6G3	Cas9_E1W6G3_prot	523
97	J4KDT3	Cas9_J4KDT3_prot	524
98	E3CY56	Cas9_E3CY56_prot	525
99	J7RUA5	Cas9_J7RUA5_prot	526
100	A0A151A3A4	Cas9_A0A151A3A4_prot	527
101	A0A1E5TL62	Cas9_A0A1E5TL62_prot	528
102	M4S2X5	Cas9_M4S2X5_prot	529
103	E0F2V7	Cas9_E0F2V7_prot	530
104	A0A0N7KBI5	Cas9_A0A0N7KBI5_prot	531
105	A0A133QCR3	Cas9_A0A133QCR3_prot	532
106	K0G350	Cas9_K0G350_prot	533
107	U5ULJ7	Cas9_U5ULJ7_prot	534
108	F0ET08	Cas9_F0ET08_prot	535
109	A0A0S2F228	Cas9_A0A0S2F228_prot	536
110	A0A060QC50	Cas9_A0A060QC50_prot	537

[0251]

111	C5S1N0	Cas9_C5S1N0_prot	538
112	A0A0K1NCD0	Cas9_A0A0K1NCD0_prot	539
113	A0A099TTS6	Cas9_A0A099TTS6_prot	540
114	A0A0D2S XK1	Cas9_A0A0D2S XK1_prot	541
115	A0A1E4MWW9	Cas9_A0A1E4MWW9_prot	542
116	A0A0M3VQX7	Cas9_A0A0M3VQX7_prot	543
117	A0A0T0PVC7	Cas9_A0A0T0PVC7_prot	544
118	Q7MRD3	Cas9_Q7MRD3_prot	545
119	A0A160JE60	Cas9_A0A160JE60_prot	546
120	J6LE60	Cas9_J6LE60_prot	547
121	A0A0P1D4L3	Cas9_A0A0P1D4L3_prot	548
122	A0A176I8B4	Cas9_A0A176I8B4_prot	549
123	A0A143DGZ8	Cas9_A0A143DGZ8_prot	550
124	G2ZYP2	Cas9_G2ZYP2_prot	551
125	A6VLA7	Cas9_A6VLA7_prot	552
126	A0A151APJ0	Cas9_A0A151APJ0_prot	553
127	V9H606	Cas9_V9H606_prot	554
128	A0A0D6XNZ8	Cas9_A0A0D6XNZ8_prot	555
129	Q13CC2	Cas9_Q13CC2_prot	556
130	A5EIM8	Cas9_A5EIM8_prot	557
131	B1UZL4	Cas9_B1UZL4_prot	558
132	B1BJM3	Cas9_B1BJM3_prot	559
133	Q20XX4	Cas9_Q20XX4_prot	560
134	A0A125S8L3	Cas9_A0A125S8L3_prot	561
135	A0A0B8Z713	Cas9_A0A0B8Z713_prot	562
136	A0A150D6Y2	Cas9_A0A150D6Y2_prot	563
137	A1WH93	Cas9_A1WH93_prot	564
138	R8LDU5	Cas9_R8LDU5_prot	565
139	A0A0F7K1T5	Cas9_A0A0F7K1T5_prot	566
140	R8NC81	Cas9_R8NC81_prot	567
141	A0A0P7L7M3	Cas9_A0A0P7L7M3_prot	568
142	F0PZE9	Cas9_F0PZE9_prot	569
143	C2UN05	Cas9_C2UN05_prot	570
144	T0HC86	Cas9_T0HC86_prot	571
145	R5QL13	Cas9_R5QL13_prot	572
146	A0A0J0YQ19	Cas9_A0A0J0YQ19_prot	573
147	A0A196P6K7	Cas9_A0A196P6K7_prot	574
148	R6QL84	Cas9_R6QL84_prot	575
149	A0A0J5QZM1	Cas9_A0A0J5QZM1_prot	576
150	A0A0P1ETF1	Cas9_A0A0P1ETF1_prot	577
151	A0A125S8J5	Cas9_A0A125S8J5_prot	578
152	A0A1C6WUG4	Cas9_A0A1C6WUG4_prot	579
153	A0A1D3QUT4	Cas9_A0A1D3QUT4_prot	580
154	F2B8K0	Cas9_F2B8K0_prot	581
155	A0A1D3PTA0	Cas9_A0A1D3PTA0_prot	582

[0252]

156	A0A0P7LDT0	Cas9_A0A0P7LDT0_prot	583
157	A0A0R1LQW1	Cas9_A0A0R1LQW1_prot	584
158	A0A159Z911	Cas9_A0A159Z911_prot	585
159	R5Y7W7	Cas9_R5Y7W7_prot	586
160	A8LN05	Cas9_A8LN05_prot	587
161	S0RVL7	Cas9_S0RVL7_prot	588
162	W1K9F9	Cas9_W1K9F9_prot	589
163	A0A1E4DUI9	Cas9_A0A1E4DUI9_prot	590
164	A0A1E4F4V8	Cas9_A0A1E4F4V8_prot	591
165	J8W240	Cas9_J8W240_prot	592
166	C6SFU3	Cas9_C6SFU3_prot	593
167	C5TLV5	Cas9_C5TLV5_prot	594
168	A0A0Y5JFG8	Cas9_A0A0Y5JFG8_prot	595
169	A0A125S8I7	Cas9_A0A125S8I7_prot	596
170	E4ZF34	Cas9_E4ZF34_prot	597
171	A0A0Y6L5Q1	Cas9_A0A0Y6L5Q1_prot	598
172	A0A0T7L299	Cas9_A0A0T7L299_prot	599
173	X5EPV9	Cas9_X5EPV9_prot	600
174	C6SH44	Cas9_C6SH44_prot	601
175	E0NB23	Cas9_E0NB23_prot	602
176	A9M1K5	Cas9_A9M1K5_prot	603
177	D0W2Z9	Cas9_D0W2Z9_prot	604
178	R0TXT9	Cas9_R0TXT9_prot	605
179	C6S593	Cas9_C6S593_prot	606
180	A0A1A6FJT6	Cas9_A0A1A6FJT6_prot	607
181	A0A0D8IYR9	Cas9_A0A0D8IYR9_prot	608
182	A0A0W7TPK7	Cas9_A0A0W7TPK7_prot	609
183	G9RUL1	Cas9_G9RUL1_prot	610
184	A0A0N8K819	Cas9_A0A0N8K819_prot	611
185	A0A0Q7HTH3	Cas9_A0A0Q7HTH3_prot	612
186	A0A0Q0YQ33	Cas9_A0A0Q0YQ33_prot	613
187	E8LGQ1	Cas9_E8LGQ1_prot	614
188	A0A0K1KC97	Cas9_A0A0K1KC97_prot	615
189	A0A150MM34	Cas9_A0A150MM34_prot	616
190	H7F839	Cas9_H7F839_prot	617
191	A0A178TEJ9	Cas9_A0A178TEJ9_prot	618
192	A0A150MP45	Cas9_A0A150MP45_prot	619
193	A0A164FEH7	Cas9_A0A164FEH7_prot	620
194	V6VHM9	Cas9_V6VHM9_prot	621
195	A0A096BCZ5	Cas9_A0A096BCZ5_prot	622
196	A0A0J8GDE4	Cas9_A0A0J8GDE4_prot	623
197	G9QLF2	Cas9_G9QLF2_prot	624
198	D7N2B0	Cas9_D7N2B0_prot	625
199	S5ZZV3	Cas9_S5ZZV3_prot	626
200	A0A0N1BZF2	Cas9_A0A0N1BZF2_prot	627

[0253]

201	A0A0C9MY24	Cas9_A0A0C9MY24_prot	628
202	A0A1C7NZW8	Cas9_A0A1C7NZW8_prot	629
203	H0UDA8	Cas9_H0UDA8_prot	630
204	E3HCA8	Cas9_E3HCA8_prot	631
205	A0A073IJU3	Cas9_A0A073IJU3_prot	632
206	W3RQ02	Cas9_W3RQ02_prot	633
207	A0A0U2W148	Cas9_A0A0U2W148_prot	634
208	G4CMU0	Cas9_G4CMU0_prot	635
209	A0A0H1A177	Cas9_A0A0H1A177_prot	636
210	A0A125S8L4	Cas9_A0A125S8L4_prot	637
211	A0A0T2NHL9	Cas9_A0A0T2NHL9_prot	638
212	A0A1A9FXI0	Cas9_A0A1A9FXI0_prot	639
213	A0A139DPY2	Cas9_A0A139DPY2_prot	640
214	A0A1A7V637	Cas9_A0A1A7V637_prot	641
215	R5KSL2	Cas9_R5KSL2_prot	642
216	R7B4M2	Cas9_R7B4M2_prot	643
217	A0A143X3E0	Cas9_A0A143X3E0_prot	644
218	R6DVD3	Cas9_R6DVD3_prot	645
219	W0A9N2	Cas9_W0A9N2_prot	646
220	R5UJK1	Cas9_R5UJK1_prot	647
221	A5Z395	Cas9_A5Z395_prot	648
222	A0A0X1TKX4	Cas9_A0A0X1TKX4_prot	649
223	R7A6L3	Cas9_R7A6L3_prot	650
224	J2WIFY6	Cas9_J2WIFY6_prot	651
225	W1SA26	Cas9_W1SA26_prot	652
226	A0A099UAI1	Cas9_A0A099UAI1_prot	653
227	U2XW20	Cas9_U2XW20_prot	654
228	R6ACK8	Cas9_R6ACK8_prot	655
229	C4ZA16	Cas9_C4ZA16_prot	656
230	A0A133XDM2	Cas9_A0A133XDM2_prot	657
231	V8C5L2	Cas9_V8C5L2_prot	658
232	E4MSY6	Cas9_E4MSY6_prot	659
233	A0A0A1H768	Cas9_A0A0A1H768_prot	660
234	A0A0Q7WLY8	Cas9_A0A0Q7WLY8_prot	661
235	A0A0R1JQF2	Cas9_A0A0R1JQF2_prot	662
236	A0A142LIG4	Cas9_A0A142LIG4_prot	663
237	R2S872	Cas9_R2S872_prot	664
238	A0A0E2RF34	Cas9_A0A0E2RF34_prot	665
239	A0A0R1IXU4	Cas9_A0A0R1IXU4_prot	666
240	A0A0E2Q4M6	Cas9_A0A0E2Q4M6_prot	667
241	A0A139MDP4	Cas9_A0A139MDP4_prot	668
242	X8HGN9	Cas9_X8HGN9_prot	669
243	A0A1C3YEE6	Cas9_A0A1C3YEE6_prot	670
244	A0A0P6UEB3	Cas9_A0A0P6UEB3_prot	671
245	A0A0M3RT06	Cas9_A0A0M3RT06_prot	672

246	K0ZVL9	Cas9_K0ZVL9_prot	673
247	R0P7Y6	Cas9_R0P7Y6_prot	674
248	S0KIG9	Cas9_S0KIG9_prot	675
249	A0A081Q0Q9	Cas9_A0A081Q0Q9_prot	676
250	F8LWC5	Cas9_F8LWC5_prot	677
251	I0SS54	Cas9_I0SS54_prot	678
252	A0A125S8J6	Cas9_A0A125S8J6_prot	679
253	V8LWT4	Cas9_V8LWT4_prot	680
254	A0A111NJ61	Cas9_A0A111NJ61_prot	681
255	A0A0A0DHL5	Cas9_A0A0A0DHL5_prot	682
256	Q5M542	Cas9_Q5M542_prot	683
257	A0A0Z8LKF1	Cas9_A0A0Z8LKF1_prot	684
258	A0A126UMM8	Cas9_A0A126UMM8_prot	685
259	A0A0N8VNG6	Cas9_A0A0N8VNG6_prot	686
260	A0A081PRN2	Cas9_A0A081PRN2_prot	687
261	T1ZF93	Cas9_T1ZF93_prot	688
262	A0A125S8J9	Cas9_A0A125S8J9_prot	689
263	A0A0H4LAU6	Cas9_A0A0H4LAU6_prot	690
264	A0A0P0N7J4	Cas9_A0A0P0N7J4_prot	691
265	A0A1G0BAB7	Cas9_A0A1G0BAB7_prot	692
266	F8LNX0	Cas9_F8LNX0_prot	693
267	U5P749	Cas9_U5P749_prot	694
268	K6QJ37	Cas9_K6QJ37_prot	695
269	D4KTZ0	Cas9_D4KTZ0_prot	696
270	U1GXL8	Cas9_U1GXL8_prot	697
271	E8KVY4	Cas9_E8KVY4_prot	698
272	A0A173V977	Cas9_A0A173V977_prot	699
273	W3XZF8	Cas9_W3XZF8_prot	700
274	A0A0X8G6A4	Cas9_A0A0X8G6A4_prot	701
275	A0A139RFX7	Cas9_A0A139RFX7_prot	702
276	A0A0R1M5J7	Cas9_A0A0R1M5J7_prot	703
277	A0A0U3F8P4	Cas9_A0A0U3F8P4_prot	704
278	B1SGF4	Cas9_B1SGF4_prot	705
279	A0A0U3EY47	Cas9_A0A0U3EY47_prot	706
280	A0A176T602	Cas9_A0A176T602_prot	707
281	A0A1C3SQ53	Cas9_A0A1C3SQ53_prot	708
282	F5WVI4	Cas9_F5WVI4_prot	709
283	H2A7K0	Cas9_H2A7K0_prot	710
284	A0A091BWC6	Cas9_A0A091BWC6_prot	711
285	F5X275	Cas9_F5X275_prot	712
286	A0A081JGI6	Cas9_A0A081JGI6_prot	713
287	B9M9X8	Cas9_B9M9X8_prot	714
288	A0A1D2U437	Cas9_A0A1D2U437_prot	715
289	E6WZS9	Cas9_E6WZS9_prot	716
290	S0J9K5	Cas9_S0J9K5_prot	717

[0254]



[0255]

291	A0A0R1J9U0	Cas9_A0A0R1J9U0_prot	718
292	X8KGX3	Cas9_X8KGX3_prot	719
293	A0A081R6F9	Cas9_A0A081R6F9_prot	720
294	A0A139QZ91	Cas9_A0A139QZ91_prot	721
295	S1RM25	Cas9_S1RM25_prot	722
296	E0PQK3	Cas9_E0PQK3_prot	723
297	I0QHG7	Cas9_I0QHG7_prot	724
298	A0A0R1XK13	Cas9_A0A0R1XK13_prot	725
299	A0A1E9DYC7	Cas9_A0A1E9DYC7_prot	726
300	A0A139NS17	Cas9_A0A139NS17_prot	727
301	A8AY02	Cas9_A8AY02_prot	728
302	E9DN79	Cas9_E9DN79_prot	729
303	A0A125S8J4	Cas9_A0A125S8J4_prot	730
304	K8Z8F3	Cas9_K8Z8F3_prot	731
305	C7G697	Cas9_C7G697_prot	732
306	A0A0F2E4R3	Cas9_A0A0F2E4R3_prot	733
307	I2NMF2	Cas9_I2NMF2_prot	734
308	A0A173VVZ1	Cas9_A0A173VVZ1_prot	735
309	A0A1F0FMT7	Cas9_A0A1F0FMT7_prot	736
310	K1LQN8	Cas9_K1LQN8_prot	737
311	A0A125S8K1	Cas9_A0A125S8K1_prot	738
312	A0A125S8K3	Cas9_A0A125S8K3_prot	739
313	H8MA21	Cas9_H8MA21_prot	740
314	W0SDH6	Cas9_W0SDH6_prot	741
315	J9E534	Cas9_J9E534_prot	742
316	A0A0V0PNI8	Cas9_A0A0V0PNI8_prot	743
317	A0A171J711	Cas9_A0A171J711_prot	744
318	Q1WVK1	Cas9_Q1WVK1_prot	745
319	C0FXH5	Cas9_C0FXH5_prot	746
320	K0XCK7	Cas9_K0XCK7_prot	747
321	A0A125S8J8	Cas9_A0A125S8J8_prot	748
322	A0A060RE66	Cas9_A0A060RE66_prot	749
323	Q1QGC9	Cas9_Q1QGC9_prot	750
324	D3NT09	Cas9_D3NT09_prot	751
325	A0A0R1MEF5	Cas9_A0A0R1MEF5_prot	752
326	A0A0R2CKA0	Cas9_A0A0R2CKA0_prot	753
327	V4Q7N5	Cas9_V4Q7N5_prot	754
328	Q2RX87	Cas9_Q2RX87_prot	755
329	A0A0R2FKF9	Cas9_A0A0R2FKF9_prot	756
330	R7BDB6	Cas9_R7BDB6_prot	757
331	A0A1C9ZUE2	Cas9_A0A1C9ZUE2_prot	758
332	S2WQ18	Cas9_S2WQ18_prot	759
333	A0A1C9ZTA0	Cas9_A0A1C9ZTA0_prot	760
334	F0RSV0	Cas9_F0RSV0_prot	761
335	A0A0N0IXQ9	Cas9_A0A0N0IXQ9_prot	762

[0256]

336	A0A0R1X611	Cas9_A0A0R1X611_prot	763
337	B2KB46	Cas9_B2KB46_prot	764
338	U2KF13	Cas9_U2KF13_prot	765
339	D5ESN1	Cas9_D5ESN1_prot	766
340	R7HI23	Cas9_R7HI23_prot	767
341	I9J7D5	Cas9_I9J7D5_prot	768
342	A0A0B0BZE8	Cas9_A0A0B0BZE8_prot	769
343	R5W806	Cas9_R5W806_prot	770
344	G6B158	Cas9_G6B158_prot	771
345	U2IU08	Cas9_U2IU08_prot	772
346	A0A096AT21	Cas9_A0A096AT21_prot	773
347	R6ZCR1	Cas9_R6ZCR1_prot	774
348	D1W6R4	Cas9_D1W6R4_prot	775
349	D1VXP4	Cas9_D1VXP4_prot	776
350	U7USL1	Cas9_U7USL1_prot	777
351	E2N8V1	Cas9_E2N8V1_prot	778
352	R6D2P4	Cas9_R6D2P4_prot	779
353	A0A134BD56	Cas9_A0A134BD56_prot	780
354	A0A099BT78	Cas9_A0A099BT78_prot	781
355	R7CVK2	Cas9_R7CVK2_prot	782
356	D2EJF1	Cas9_D2EJF1_prot	783
357	A0A069QG82	Cas9_A0A069QG82_prot	784
358	R5LWG1	Cas9_R5LWG1_prot	785
359	C9LGP5	Cas9_C9LGP5_prot	786
360	Q6KIQ7	Cas9_Q6KIQ7_prot	787
361	A0A180F6C8	Cas9_A0A180F6C8_prot	788
362	C0WRP7	Cas9_C0WRP7_prot	789
363	A0A174NKB5	Cas9_A0A174NKB5_prot	790
364	U2Y346	Cas9_U2Y346_prot	791
365	A0A125S8I5	Cas9_A0A125S8I5_prot	792
366	R7CG17	Cas9_R7CG17_prot	793
367	F3A050	Cas9_F3A050_prot	794
368	D1AUW6	Cas9_D1AUW6_prot	795
369	A0A0X8KN88	Cas9_A0A0X8KN88_prot	796
370	A0A0D5BKQ5	Cas9_A0A0D5BKQ5_prot	797
371	A0A0N1DVV7	Cas9_A0A0N1DVV7_prot	798
372	A0A085Z0I3	Cas9_A0A085Z0I3_prot	799
373	J3TRJ9	Cas9_J3TRJ9_prot	800
374	A0A0F6CLF2	Cas9_A0A0F6CLF2_prot	801
375	A0A199XSD8	Cas9_A0A199XSD8_prot	802
376	A0A0B8YC59	Cas9_A0A0B8YC59_prot	803
377	K2M2X7	Cas9_K2M2X7_prot	804
378	A0A1B9Y472	Cas9_A0A1B9Y472_prot	805
379	A0A0Q4DTQ9	Cas9_A0A0Q4DTQ9_prot	806
380	S4EM46	Cas9_S4EM46_prot	807

381	A0A1D2JYF3	Cas9_A0A1D2JYF3_prot	808
382	A0A0R2FVI8	Cas9_A0A0R2FVI8_prot	809
383	A0A174LFF7	Cas9_A0A174LFF7_prot	810
384	A0A173SPI3	Cas9_A0A173SPI3_prot	811
385	D0DRL9	Cas9_D0DRL9_prot	812
386	A0A175A1Y1	Cas9_A0A175A1Y1_prot	813
387	A0A062XBE5	Cas9_A0A062XBE5_prot	814
388	A0A0K8MIK7	Cas9_A0A0K8MIK7_prot	815
389	A0A0R1TV35	Cas9_A0A0R1TV35_prot	816
390	A0A125S8J7	Cas9_A0A125S8J7_prot	817
391	A0A0R1ZP43	Cas9_A0A0R1ZP43_prot	818
392	W4T7U3	Cas9_W4T7U3_prot	819
393	A0A0J5P9G6	Cas9_A0A0J5P9G6_prot	820
394	A0A0R2CL57	Cas9_A0A0R2CL57_prot	821
395	R6U7U5	Cas9_R6U7U5_prot	822
396	A0A0R2AFH9	Cas9_A0A0R2AFH9_prot	823
397	E7MR72	Cas9_E7MR72_prot	824
398	A0A1C0YPC7	Cas9_A0A1C0YPC7_prot	825
399	A0A179EQS1	Cas9_A0A179EQS1_prot	826
400	W9EE99	Cas9_W9EE99_prot	827
401	A0A0R2BKJ5	Cas9_A0A0R2BKJ5_prot	828
402	E6LI02	Cas9_E6LI02_prot	829
403	V5XLV7	Cas9_V5XLV7_prot	830
404	G2KVM6	Cas9_G2KVM6_prot	831
405	A0A1C5TF27	Cas9_A0A1C5TF27_prot	832
406	H3NFH0	Cas9_H3NFH0_prot	833
407	A0A081BKX9	Cas9_A0A081BKX9_prot	834
408	A0A1C7DHH3	Cas9_A0A1C7DHH3_prot	835
409	R7GMQ9	Cas9_R7GMQ9_prot	836
410	R6QHH1	Cas9_R6QHH1_prot	837
411	A0A174HSW2	Cas9_A0A174HSW2_prot	838
412	A0A0R2HIR8	Cas9_A0A0R2HIR8_prot	839
413	A0A0H3GNI1	Cas9_A0A0H3GNI1_prot	840
414	A0A0F5ZHG0	Cas9_A0A0F5ZHG0_prot	841
415	A0A0H3J2A7	Cas9_A0A0H3J2A7_prot	842
416	A0A166RWM5	Cas9_A0A166RWM5_prot	843
417	A0A1E6FAD7	Cas9_A0A1E6FAD7_prot	844
418	A0A1E8EQS5	Cas9_A0A1E8EQS5_prot	845
419	A0A1E7DWS8	Cas9_A0A1E7DWS8_prot	846
420	A0A1E5ZAU0	Cas9_A0A1E5ZAU0_prot	847
421	A0A1E8EI75	Cas9_A0A1E8EI75_prot	848
422	R3WHR8	Cas9_R3WHR8_prot	849
423	A0A097B8A9	Cas9_A0A097B8A9_prot	850
424	A0A095XEU7	Cas9_A0A095XEU7_prot	851
425	A0A0R1RFJ4	Cas9_A0A0R1RFJ4_prot	852

[0257]

[0258]

426	A0A160NBB3	Cas9_A0A160NBB3_prot	853
427	I6T669	Cas9_I6T669_prot	854
428	H1GG18	Cas9_H1GG18_prot	855
429	A0A017H668	Cas9_A0A017H668_prot	856
430	A0A121IZ21	Cas9_A0A121IZ21_prot	857
431	A0A1B4XLG6	Cas9_A0A1B4XLG6_prot	858
432	U6S081	Cas9_U6S081_prot	859
433	E6GPD8	Cas9_E6GPD8_prot	860
434	H3NQF8	Cas9_H3NQF8_prot	861
435	D4J3S7	Cas9_D4J3S7_prot	862
436	G5JVJ9	Cas9_G5JVJ9_prot	863
437	R9MHT9	Cas9_R9MHT9_prot	864
438	R2SDC4	Cas9_R2SDC4_prot	865
439	H7FYD8	Cas9_H7FYD8_prot	866
440	A0A1D2JQJ5	Cas9_A0A1D2JQJ5_prot	867
441	A0A1D2LU44	Cas9_A0A1D2LU44_prot	868
442	C9BHR2	Cas9_C9BHR2_prot	869
443	L2LBP5	Cas9_L2LBP5_prot	870
444	R6ZAM8	Cas9_R6ZAM8_prot	871
445	A6BJV4	Cas9_A6BJV4_prot	872
446	A0A174GDD3	Cas9_A0A174GDD3_prot	873
447	C9BWE2	Cas9_C9BWE2_prot	874
448	A0A173UVP4	Cas9_A0A173UVP4_prot	875
449	R5BQB0	Cas9_R5BQB0_prot	876
450	D7N6R3	Cas9_D7N6R3_prot	877
451	A0A1C5P2V8	Cas9_A0A1C5P2V8_prot	878
452	B5CL59	Cas9_B5CL59_prot	879
453	A0A0R2JSC5	Cas9_A0A0R2JSC5_prot	880
454	A0A1C6BK34	Cas9_A0A1C6BK34_prot	881
455	R7KBA0	Cas9_R7KBA0_prot	882
456	A0A0R2HM97	Cas9_A0A0R2HM97_prot	883
457	U7PCQ1	Cas9_U7PCQ1_prot	884
458	R5V4T4	Cas9_R5V4T4_prot	885
459	A0A133QT10	Cas9_A0A133QT10_prot	886
460	A0A0E2EP65	Cas9_A0A0E2EP65_prot	887
461	R5MT23	Cas9_R5MT23_prot	888
462	A0A0R2DR00	Cas9_A0A0R2DR00_prot	889
463	R5N3I1	Cas9_R5N3I1_prot	890
464	I0SF74	Cas9_I0SF74_prot	891
465	E6J3R0	Cas9_E6J3R0_prot	892
466	U2YFI6	Cas9_U2YFI6_prot	893
467	A0A0R2DIR3	Cas9_A0A0R2DIR3_prot	894
468	U2U1P0	Cas9_U2U1P0_prot	895
469	A0A134CKK1	Cas9_A0A134CKK1_prot	896
470	A0A0R2N0I6	Cas9_A0A0R2N0I6_prot	897

[0259]

471	A0A125S8I2	Cas9_A0A125S8I2_prot	898
472	A0A0R1ZCI7	Cas9_A0A0R1ZCI7_prot	899
473	A0A0R1SCA3	Cas9_A0A0R1SCA3_prot	900
474	D9PRA6	Cas9_D9PRA6_prot	901
475	A0A0D0ZAW2	Cas9_A0A0D0ZAW2_prot	902
476	F9N0W8	Cas9_F9N0W8_prot	903
477	B0RZQ7	Cas9_B0RZQ7_prot	904
478	R6XMN7	Cas9_R6XMN7_prot	905
479	U2SSY7	Cas9_U2SSY7_prot	906
480	S4NUM0	Cas9_S4NUM0_prot	907
481	A0A072ETA7	Cas9_A0A072ETA7_prot	908
482	R5RU71	Cas9_R5RU71_prot	909
483	A0A174FD97	Cas9_A0A174FD97_prot	910
484	A0A0A8K7X7	Cas9_A0A0A8K7X7_prot	911
485	R5Z6B4	Cas9_R5Z6B4_prot	912
486	S1NSG8	Cas9_S1NSG8_prot	913
487	A0A1D8P523	Cas9_A0A1D8P523_prot	914
488	A0A1C6IPF7	Cas9_A0A1C6IPF7_prot	915
489	A0A0R1MNC7	Cas9_A0A0R1MNC7_prot	916
490	A0A132HQM8	Cas9_A0A132HQM8_prot	917
491	A0A0M9VGT5	Cas9_A0A0M9VGT5_prot	918
492	F9MP31	Cas9_F9MP31_prot	919
493	A0A0R1V7X0	Cas9_A0A0R1V7X0_prot	920
494	A0A0X7BAB3	Cas9_A0A0X7BAB3_prot	921
495	R7K435	Cas9_R7K435_prot	922
496	I3Z8Z5	Cas9_I3Z8Z5_prot	923
497	A0A173YKH0	Cas9_A0A173YKH0_prot	924
498	A0A174PI34	Cas9_A0A174PI34_prot	925
499	A0A1C6E673	Cas9_A0A1C6E673_prot	926
500	R5YGP2	Cas9_R5YGP2_prot	927
501	A0A076P3F6	Cas9_A0A076P3F6_prot	928
502	A0A176Y372	Cas9_A0A176Y372_prot	929
503	D3I574	Cas9_D3I574_prot	930
504	S2D876	Cas9_S2D876_prot	931
505	A0A0R1F6R4	Cas9_A0A0R1F6R4_prot	932
506	J9YH95	Cas9_J9YH95_prot	933
507	R5SXF4	Cas9_R5SXF4_prot	934
508	R6P3Z6	Cas9_R6P3Z6_prot	935
509	A0A0R1IS26	Cas9_A0A0R1IS26_prot	936
510	A0A0H4LAX2	Cas9_A0A0H4LAX2_prot	937
511	A0A0K2LF21	Cas9_A0A0K2LF21_prot	938
512	A0A0R1QGB6	Cas9_A0A0R1QGB6_prot	939
513	A0A143W8R3	Cas9_A0A143W8R3_prot	940
514	M4KKI8	Cas9_M4KKI8_prot	941
515	A0A0R1FUZ5	Cas9_A0A0R1FUZ5_prot	942

[0260]

516	F6ITQ2	Cas9_F6ITQ2_prot	943
517	A0A1E3KQ44	Cas9_A0A1E3KQ44_prot	944
518	A0A173WIE2	Cas9_A0A173WIE2_prot	945
519	G4Q6A5	Cas9_G4Q6A5_prot	946
520	A0A0K1MWW2	Cas9_A0A0K1MWW2_prot	947
521	A0A0H0YP06	Cas9_A0A0H0YP06_prot	948
522	A0A0C9QP69	Cas9_A0A0C9QP69_prot	949
523	A0A0E4H4H8	Cas9_A0A0E4H4H8_prot	950
524	C2CKI6	Cas9_C2CKI6_prot	951
525	A0A0M2FYH7	Cas9_A0A0M2FYH7_prot	952
526	R6TGN6	Cas9_R6TGN6_prot	953
527	I9L4B5	Cas9_I9L4B5_prot	954
528	A0A133KEN0	Cas9_A0A133KEN0_prot	955
529	A0A139NKI7	Cas9_A0A139NKI7_prot	956
530	T5JDL4	Cas9_T5JDL4_prot	957
531	C5F8S2	Cas9_C5F8S2_prot	958
532	S4ZP66	Cas9_S4ZP66_prot	959
533	S2LEI5	Cas9_S2LEI5_prot	960
534	A0A0R1UKG9	Cas9_A0A0R1UKG9_prot	961
535	A0A174P7Q9	Cas9_A0A174P7Q9_prot	962
536	K6R5Z8	Cas9_K6R5Z8_prot	963
537	A0A0R1S2S1	Cas9_A0A0R1S2S1_prot	964
538	A0A0R1MEL8	Cas9_A0A0R1MEL8_prot	965
539	A0A0C9Q7U6	Cas9_A0A0C9Q7U6_prot	966
540	A0A179YJ40	Cas9_A0A179YJ40_prot	967
541	C7TEQ6	Cas9_C7TEQ6_prot	968
542	E0NI75	Cas9_E0NI75_prot	969
543	A0A133ZK65	Cas9_A0A133ZK65_prot	970
544	A0A0R1RRH5	Cas9_A0A0R1RRH5_prot	971
545	E0NJ84	Cas9_E0NJ84_prot	972
546	A0A0R2HZC9	Cas9_A0A0R2HZC9_prot	973
547	A0A180AER3	Cas9_A0A180AER3_prot	974
548	D6GRK4	Cas9_D6GRK4_prot	975
549	A0A1B3WEM9	Cas9_A0A1B3WEM9_prot	976
550	A0A116L128	Cas9_A0A116L128_prot	977
551	A0A127TRM8	Cas9_A0A127TRM8_prot	978
552	A0A0R1W1T1	Cas9_A0A0R1W1T1_prot	979
553	A0A1A5VIM0	Cas9_A0A1A5VIM0_prot	980
554	K6RXS8	Cas9_K6RXS8_prot	981
555	X0QNI0	Cas9_X0QNI0_prot	982
556	R5WWQ0	Cas9_R5WWQ0_prot	983
557	C7XMU0	Cas9_C7XMU0_prot	984
558	D6LEV9	Cas9_D6LEV9_prot	985
559	A0A128ECZ8	Cas9_A0A128ECZ8_prot	986
560	A0A133NAH6	Cas9_A0A133NAH6_prot	987

[0261]

561	A0A0X3Y1U5	Cas9_A0A0X3Y1U5_prot	988
562	A0A116M370	Cas9_A0A116M370_prot	989
563	A0A116KLL2	Cas9_A0A116KLL2_prot	990
564	A0A1B2IXP8	Cas9_A0A1B2IXP8_prot	991
565	A0A0R1LCE0	Cas9_A0A0R1LCE0_prot	992
566	A0A0R1WWN2	Cas9_A0A0R1WWN2_prot	993
567	A0A0C6FZC2	Cas9_A0A0C6FZC2_prot	994
568	A0A127X7N0	Cas9_A0A127X7N0_prot	995
569	A0A1B4Z6K5	Cas9_A0A1B4Z6K5_prot	996
570	R9LW52	Cas9_R9LW52_prot	997
571	A0A0B2XHU2	Cas9_A0A0B2XHU2_prot	998
572	A0A1B2A6P4	Cas9_A0A1B2A6P4_prot	999
573	A0A0P6SHS4	Cas9_A0A0P6SHS4_prot	1000
574	A0A0H3BZZ0	Cas9_A0A0H3BZZ0_prot	1001
575	A0A0R1TGJ3	Cas9_A0A0R1TGJ3_prot	1002
576	Q1JLZ6	Cas9_Q1JLZ6_prot	1003
577	Q48TU5	Cas9_Q48TU5_prot	1004
578	G6CGE4	Cas9_G6CGE4_prot	1005
579	Q1JH43	Cas9_Q1JH43_prot	1006
580	R5C8N0	Cas9_R5C8N0_prot	1007
581	A0A0R1SN52	Cas9_A0A0R1SN52_prot	1008
582	A0A0R2DGS6	Cas9_A0A0R2DGS6_prot	1009
583	A0A0R1SDU2	Cas9_A0A0R1SDU2_prot	1010
584	A0A0D0YUU5	Cas9_A0A0D0YUU5_prot	1011
585	S9AZZ0	Cas9_S9AZZ0_prot	1012
586	A0A0E1XG84	Cas9_A0A0E1XG84_prot	1013
587	R6ET93	Cas9_R6ET93_prot	1014
588	S9BFF6	Cas9_S9BFF6_prot	1015
589	A0A1B3PSQ7	Cas9_A0A1B3PSQ7_prot	1016
590	A0A137PP63	Cas9_A0A137PP63_prot	1017
591	S9KSN8	Cas9_S9KSN8_prot	1018
592	Q8E042	Cas9_Q8E042_prot	1019
593	S8HGI0	Cas9_S8HGI0_prot	1020
594	S8H4C8	Cas9_S8H4C8_prot	1021
595	F0FD37	Cas9_F0FD37_prot	1022
596	J3JPT0	Cas9_J3JPT0_prot	1023
597	F8Y040	Cas9_F8Y040_prot	1024
598	A0A0R2JE56	Cas9_A0A0R2JE56_prot	1025
599	A0A1A9E0X4	Cas9_A0A1A9E0X4_prot	1026
600	A0A0E1EMN2	Cas9_A0A0E1EMN2_prot	1027
601	A0A1C0BC24	Cas9_A0A1C0BC24_prot	1028
602	A0A1E2WAR5	Cas9_A0A1E2WAR5_prot	1029
603	S8FJS0	Cas9_S8FJS0_prot	1030
604	F4FTI2	Cas9_F4FTI2_prot	1031
605	K4Q9P5	Cas9_K4Q9P5_prot	1032

[0262]

606	M4YX12	Cas9_M4YX12_prot	1033
607	F9HIG7	Cas9_F9HIG7_prot	1034
608	F5WVJ4	Cas9_F5WVJ4_prot	1035
609	D6E761	Cas9_D6E761_prot	1036
610	I0Q2W2	Cas9_I0Q2W2_prot	1037
611	C5WH61	Cas9_C5WH61_prot	1038
612	A0A1C2CVQ9	Cas9_A0A1C2CVQ9_prot	1039
613	K8MQ90	Cas9_K8MQ90_prot	1040
614	A0A0R1JG51	Cas9_A0A0R1JG51_prot	1041
615	J9W3C2	Cas9_J9W3C2_prot	1042
616	Q1J6W2	Cas9_Q1J6W2_prot	1043
617	R5GJ26	Cas9_R5GJ26_prot	1044
618	A0A172Q7S3	Cas9_A0A172Q7S3_prot	1045
619	A0A060RIR3	Cas9_A0A060RIR3_prot	1046
620	A0A1C5U497	Cas9_A0A1C5U497_prot	1047
621	S5R5C8	Cas9_S5R5C8_prot	1048
622	A0A0W7V6X6	Cas9_A0A0W7V6X6_prot	1049
623	A0A1C5S579	Cas9_A0A1C5S579_prot	1050
624	A0A125S8J0	Cas9_A0A125S8J0_prot	1051
625	E0PEL3	Cas9_E0PEL3_prot	1052
626	J4K985	Cas9_J4K985_prot	1053
627	U2PI18	Cas9_U2PI18_prot	1054
628	G5KAN2	Cas9_G5KAN2_prot	1055
629	Q7P7J1	Cas9_Q7P7J1_prot	1056
630	A0A0F2D9H7	Cas9_A0A0F2D9H7_prot	1057
631	A0A1D7ZZ65	Cas9_A0A1D7ZZ65_prot	1058
632	E7FPD8	Cas9_E7FPD8_prot	1059
633	A0A176TM67	Cas9_A0A176TM67_prot	1060
634	G6AFY6	Cas9_G6AFY6_prot	1061
635	E9FPR9	Cas9_E9FPR9_prot	1062
636	I7QXF2	Cas9_I7QXF2_prot	1063
637	I0TCL1	Cas9_I0TCL1_prot	1064
638	A0A173R3H4	Cas9_A0A173R3H4_prot	1065
639	A0A178KKP5	Cas9_A0A178KKP5_prot	1066
640	H6PBR9	Cas9_H6PBR9_prot	1067
641	F4AF10	Cas9_F4AF10_prot	1068
642	A0A134C7A8	Cas9_A0A134C7A8_prot	1069
643	A0A0R1SG79	Cas9_A0A0R1SG79_prot	1070
644	A0A0F2DWP8	Cas9_A0A0F2DWP8_prot	1071
645	A0A0R1K630	Cas9_A0A0R1K630_prot	1072
646	A0A135YMA6	Cas9_A0A135YMA6_prot	1073
647	F0I6Z8	Cas9_F0I6Z8_prot	1074
648	E9FJ16	Cas9_E9FJ16_prot	1075
649	C2D302	Cas9_C2D302_prot	1076
650	Q8E5R9	Cas9_Q8E5R9_prot	1077



[0263]

651	E8JP81	Cas9_E8JP81_prot	1078
652	A0A0R1RHH9	Cas9_A0A0R1RHH9_prot	1079
653	A0A0F3FWK9	Cas9_A0A0F3FWK9_prot	1080
654	A0A0R2I8Q5	Cas9_A0A0R2I8Q5_prot	1081
655	A0A150NPH1	Cas9_A0A150NPH1_prot	1082
656	E7S4M3	Cas9_E7S4M3_prot	1083
657	A0A143ASS0	Cas9_A0A143ASS0_prot	1084
658	A0A0R2HDR9	Cas9_A0A0R2HDR9_prot	1085
659	A0A0B2JE32	Cas9_A0A0B2JE32_prot	1086
660	A0A0R2KUQ3	Cas9_A0A0R2KUQ3_prot	1087
661	A0A0W7V0H0	Cas9_A0A0W7V0H0_prot	1088
662	R6TGA0	Cas9_R6TGA0_prot	1089
663	A0A0H5B4T2	Cas9_A0A0H5B4T2_prot	1090
664	U2J559	Cas9_U2J559_prot	1091
665	A0A075SSB9	Cas9_A0A075SSB9_prot	1092
666	A0A096B7Z5	Cas9_A0A096B7Z5_prot	1093
667	L9PS87	Cas9_L9PS87_prot	1094
668	A0A134D9V8	Cas9_A0A134D9V8_prot	1095
669	F7UWL3	Cas9_F7UWL3_prot	1096
670	G7SP82	Cas9_G7SP82_prot	1097
671	A0A0R2E213	Cas9_A0A0R2E213_prot	1098
672	R7I2K1	Cas9_R7I2K1_prot	1099
673	C0WXA2	Cas9_C0WXA2_prot	1100
674	A0A0Z8GCN2	Cas9_A0A0Z8GCN2_prot	1101
675	R5GUN8	Cas9_R5GUN8_prot	1102
676	A0A116RA22	Cas9_A0A116RA22_prot	1103
677	A0A0Z8JWB5	Cas9_A0A0Z8JWB5_prot	1104
678	A0A116KAQ7	Cas9_A0A116KAQ7_prot	1105
679	G0M2G7	Cas9_G0M2G7_prot	1106
680	A0A1C5P5Q5	Cas9_A0A1C5P5Q5_prot	1107
681	A0A0H1TNR9	Cas9_A0A0H1TNR9_prot	1108
682	F2NB82	Cas9_F2NB82_prot	1109
683	J7TMY5	Cas9_J7TMY5_prot	1110
684	A0A125S8I1	Cas9_A0A125S8I1_prot	1111
685	A0A078RYQ2	Cas9_A0A078RYQ2_prot	1112
686	A0A0F3H9Z9	Cas9_A0A0F3H9Z9_prot	1113
687	E5V117	Cas9_E5V117_prot	1114
688	J4TM44	Cas9_J4TM44_prot	1115
689	I7L6U4	Cas9_I7L6U4_prot	1116
690	R5J5B2	Cas9_R5J5B2_prot	1117
691	A0A1B2ULM2	Cas9_A0A1B2ULM2_prot	1118
692	A0A0P6UDU2	Cas9_A0A0P6UDU2_prot	1119
693	V8LSG7	Cas9_V8LSG7_prot	1120
694	D5BC98	Cas9_D5BC98_prot	1121
695	K0MXA7	Cas9_K0MXA7_prot	1122

[0264]

696	G9WGU4	Cas9_G9WGU4_prot	1123
697	A0A1C2D810	Cas9_A0A1C2D810_prot	1124
698	A0A087BJ94	Cas9_A0A087BJ94_prot	1125
699	A0A0R1S4R8	Cas9_A0A0R1S4R8_prot	1126
700	E0QLT3	Cas9_E0QLT3_prot	1127
701	C4VKS7	Cas9_C4VKS7_prot	1128
702	A9DTN2	Cas9_A9DTN2_prot	1129
703	K2PT21	Cas9_K2PT21_prot	1130
704	E7RR33	Cas9_E7RR33_prot	1131
705	A0A0N0CU60	Cas9_A0A0N0CU60_prot	1132
706	A0A0P7LUX3	Cas9_A0A0P7LUX3_prot	1133
707	R7D1C6	Cas9_R7D1C6_prot	1134
708	V8BZU1	Cas9_V8BZU1_prot	1135
709	A0A0F4LG72	Cas9_A0A0F4LG72_prot	1136
710	J4KB57	Cas9_J4KB57_prot	1137
711	W1U735	Cas9_W1U735_prot	1138
712	A0A095ZV18	Cas9_A0A095ZV18_prot	1139
713	R5BUB1	Cas9_R5BUB1_prot	1140
714	F2C4I5	Cas9_F2C4I5_prot	1141
715	E1LI65	Cas9_E1LI65_prot	1142
716	A0A0N0UTU1	Cas9_A0A0N0UTU1_prot	1143
717	C5NZ04	Cas9_C5NZ04_prot	1144
718	A0A081Q742	Cas9_A0A081Q742_prot	1145
719	A0A0F2DF30	Cas9_A0A0F2DF30_prot	1146
720	A0A0F4LMR6	Cas9_A0A0F4LMR6_prot	1147
721	A0A0E2EBU7	Cas9_A0A0E2EBU7_prot	1148
722	A0A0E2EGB1	Cas9_A0A0E2EGB1_prot	1149
723	A0A0C3A2P0	Cas9_A0A0C3A2P0_prot	1150
724	M2CG59	Cas9_M2CG59_prot	1151
725	R5R3T7	Cas9_R5R3T7_prot	1152
726	D6KPM9	Cas9_D6KPM9_prot	1153
727	U2VD49	Cas9_U2VD49_prot	1154
728	D6S374	Cas9_D6S374_prot	1155
729	A0A0R2KGU9	Cas9_A0A0R2KGU9_prot	1156
730	A0A0F6MNV4	Cas9_A0A0F6MNV4_prot	1157
731	A0A0X3ARL2	Cas9_A0A0X3ARL2_prot	1158
732	A0A088RCP8	Cas9_A0A088RCP8_prot	1159
733	S3KPV3	Cas9_S3KPV3_prot	1160
734	M2CIC2	Cas9_M2CIC2_prot	1161
735	M2SLU3	Cas9_M2SLU3_prot	1162
736	A0A0D4CLL6	Cas9_A0A0D4CLL6_prot	1163
737	R6I3U9	Cas9_R6I3U9_prot	1164
738	F5U0T2	Cas9_F5U0T2_prot	1165
739	A0A0F4LIJ0	Cas9_A0A0F4LIJ0_prot	1166
740	A0A0N0CQ86	Cas9_A0A0N0CQ86_prot	1167

[0265]

741	I3C2S4	Cas9_I3C2S4_prot	1168
742	U2QKG2	Cas9_U2QKG2_prot	1169
743	D1YP75	Cas9_D1YP75_prot	1170
744	A0A091BLA4	Cas9_A0A091BLA4_prot	1171
745	A0A100YPE0	Cas9_A0A100YPE0_prot	1172
746	E1LBR5	Cas9_E1LBR5_prot	1173
747	R5BD80	Cas9_R5BD80_prot	1174
748	W3Y2C1	Cas9_W3Y2C1_prot	1175
749	E1QW44	Cas9_E1QW44_prot	1176
750	A0A134A1I6	Cas9_A0A134A1I6_prot	1177
751	A0A0F4M7Y5	Cas9_A0A0F4M7Y5_prot	1178
752	A0A133YSB7	Cas9_A0A133YSB7_prot	1179
753	A0A089Y508	Cas9_A0A089Y508_prot	1180
754	A0A162CL99	Cas9_A0A162CL99_prot	1181
755	A0A133YDF1	Cas9_A0A133YDF1_prot	1182
756	A0A133YY65	Cas9_A0A133YY65_prot	1183
757	A0A0B4S2L0	Cas9_A0A0B4S2L0_prot	1184
758	A0A0R2DLB6	Cas9_A0A0R2DLB6_prot	1185
759	A0A0G3MB19	Cas9_A0A0G3MB19_prot	1186
760	A0A0Q3K6A2	Cas9_A0A0Q3K6A2_prot	1187
761	A0A134AG29	Cas9_A0A134AG29_prot	1188
762	A0A0N1DXX4	Cas9_A0A0N1DXX4_prot	1189
763	A0A1E4DZC0	Cas9_A0A1E4DZC0_prot	1190
764	J9R1Q7	Cas9_J9R1Q7_prot	1191
765	A0A1C4DJV4	Cas9_A0A1C4DJV4_prot	1192
766	A0A077KK20	Cas9_A0A077KK20_prot	1193
767	A0A085ZZC2	Cas9_A0A085ZZC2_prot	1194
768	W1V0U5	Cas9_W1V0U5_prot	1195
769	A0A0K9XVX7	Cas9_A0A0K9XVX7_prot	1196
770	A0A1H5RY71	Cas9_A0A1H5RY71_prot	1197
771	A0A0J7IGI6	Cas9_A0A0J7IGI6_prot	1198
772	A0A086AYB7	Cas9_A0A086AYB7_prot	1199
773	A0A125S8K4	Cas9_A0A125S8K4_prot	1200
774	M3INT0	Cas9_M3INT0_prot	1201
775	R5FLM1	Cas9_R5FLM1_prot	1202
776	U5Q7L9	Cas9_U5Q7L9_prot	1203
777	A0A1E3DW10	Cas9_A0A1E3DW10_prot	1204
778	K0NQV3	Cas9_K0NQV3_prot	1205
779	J2KJ07	Cas9_J2KJ07_prot	1206
780	A0A0U5KB17	Cas9_A0A0U5KB17_prot	1207
781	A0A0D6ZH65	Cas9_A0A0D6ZH65_prot	1208
782	A0A139PB46	Cas9_A0A139PB46_prot	1209
783	A0A139NVJ1	Cas9_A0A139NVJ1_prot	1210
784	A0A139NSX3	Cas9_A0A139NSX3_prot	1211
785	E0Q490	Cas9_E0Q490_prot	1212

786	E3ELL7	Cas9_E3ELL7_prot	1213
787	A0A061CF22	Cas9_A0A061CF22_prot	1214
788	F3UXG6	Cas9_F3UXG6_prot	1215
789	A0A125S8K2	Cas9_A0A125S8K2_prot	1216
790	A0A0B7IR20	Cas9_A0A0B7IR20_prot	1217
791	A0A174J8H3	Cas9_A0A174J8H3_prot	1218
792	D7IW96	Cas9_D7IW96_prot	1219
793	A0A1A9I5Z1	Cas9_A0A1A9I5Z1_prot	1220
794	W0EYD8	Cas9_W0EYD8_prot	1221
795	E2N1F9	Cas9_E2N1F9_prot	1222
796	D7VKD0	Cas9_D7VKD0_prot	1223
797	R7N9S8	Cas9_R7N9S8_prot	1224
798	C7M7G9	Cas9_C7M7G9_prot	1225
799	J0WLS6	Cas9_J0WLS6_prot	1226
800	A0A174L7S6	Cas9_A0A174L7S6_prot	1227
801	S2KA46	Cas9_S2KA46_prot	1228
802	A0A0A2F4C3	Cas9_A0A0A2F4C3_prot	1229
803	U2JCC9	Cas9_U2JCC9_prot	1230
804	A0A136MV65	Cas9_A0A136MV65_prot	1231
805	K1M8Y6	Cas9_K1M8Y6_prot	1232
806	F9YQX1	Cas9_F9YQX1_prot	1233
807	A0A0B7IQ14	Cas9_A0A0B7IQ14_prot	1234
808	A0A0B7IB79	Cas9_A0A0B7IB79_prot	1235
809	A0A0A2EHM8	Cas9_A0A0A2EHM8_prot	1236
810	A0A173V1H2	Cas9_A0A173V1H2_prot	1237
811	R7DKC0	Cas9_R7DKC0_prot	1238
812	U5CHH4	Cas9_U5CHH4_prot	1239
813	L1NKM1	Cas9_L1NKM1_prot	1240
814	A0A127VAB0	Cas9_A0A127VAB0_prot	1241
815	D7JGI6	Cas9_D7JGI6_prot	1242
816	A0A0U3BTM1	Cas9_A0A0U3BTM1_prot	1243
817	A0A0M4G8J7	Cas9_A0A0M4G8J7_prot	1244
818	A0A0A6Y3B0	Cas9_A0A0A6Y3B0_prot	1245
819	A0A015SZB2	Cas9_A0A015SZB2_prot	1246
820	A0A0E2RG29	Cas9_A0A0E2RG29_prot	1247
821	A0A0E2A7Q9	Cas9_A0A0E2A7Q9_prot	1248
822	A0A015Y7X0	Cas9_A0A015Y7X0_prot	1249
823	A0A0E2SQU9	Cas9_A0A0E2SQU9_prot	1250
824	A0A0E2T2C8	Cas9_A0A0E2T2C8_prot	1251
825	A0A017N289	Cas9_A0A017N289_prot	1252
826	A0A015UHU2	Cas9_A0A015UHU2_prot	1253
827	E5C8Y3	Cas9_E5C8Y3_prot	1254
828	C2M5N8	Cas9_C2M5N8_prot	1255
829	J4XAP6	Cas9_J4XAP6_prot	1256
830	A0A1B8ZVU5	Cas9_A0A1B8ZVU5_prot	1257

[0266]

[0267]

831	A0A1E5KUU0	Cas9_A0A1E5KUU0_prot	1258
832	A0A1E9S993	Cas9_A0A1E9S993_prot	1259
833	F0P0P2	Cas9_F0P0P2_prot	1260
834	B7B6H7	Cas9_B7B6H7_prot	1261
835	W1R7X2	Cas9_W1R7X2_prot	1262
836	X5KBF9	Cas9_X5KBF9_prot	1263
837	A0A0B7HB18	Cas9_A0A0B7HB18_prot	1264
838	I8UMX3	Cas9_I8UMX3_prot	1265
839	L1PRF6	Cas9_L1PRF6_prot	1266
840	S3CB04	Cas9_S3CB04_prot	1267
841	A0A098LI38	Cas9_A0A098LI38_prot	1268
842	A0A125S8K9	Cas9_A0A125S8K9_prot	1269
843	E6K6M2	Cas9_E6K6M2_prot	1270
844	A0A1E5UGK6	Cas9_A0A1E5UGK6_prot	1271
845	F2IKJ5	Cas9_F2IKJ5_prot	1272
846	A0A150XH78	Cas9_A0A150XH78_prot	1273
847	G8X9H3	Cas9_G8X9H3_prot	1274
848	A0A109Q6P7	Cas9_A0A109Q6P7_prot	1275
849	A0A101CFI9	Cas9_A0A101CFI9_prot	1276
850	A0A0T0M2G2	Cas9_A0A0T0M2G2_prot	1277
851	K1I305	Cas9_K1I305_prot	1278
852	L1P954	Cas9_L1P954_prot	1279
853	J0DFD8	Cas9_J0DFD8_prot	1280
854	H1YII5	Cas9_H1YII5_prot	1281
855	G2Z1C1	Cas9_G2Z1C1_prot	1282
856	A0A1E5TBF5	Cas9_A0A1E5TBF5_prot	1283
857	A0A0K1NMP1	Cas9_A0A0K1NMP1_prot	1284
858	A0A0E3VRY2	Cas9_A0A0E3VRY2_prot	1285
859	A0A133Q212	Cas9_A0A133Q212_prot	1286
860	A0A1E4APC4	Cas9_A0A1E4APC4_prot	1287
861	U2QLH7	Cas9_U2QLH7_prot	1288
862	A0A1D3UU01	Cas9_A0A1D3UU01_prot	1289
863	A0A096CIC5	Cas9_A0A096CIC5_prot	1290
864	I4ZCD3	Cas9_I4ZCD3_prot	1291
865	A0A137SV51	Cas9_A0A137SV51_prot	1292
866	A0A0X8BZ89	Cas9_A0A0X8BZ89_prot	1293
867	A0A096D253	Cas9_A0A096D253_prot	1294
868	A0A134B2X0	Cas9_A0A134B2X0_prot	1295
869	D1W1M7	Cas9_D1W1M7_prot	1296
870	U2LB41	Cas9_U2LB41_prot	1297
871	C9MPM6	Cas9_C9MPM6_prot	1298
872	R7D4J2	Cas9_R7D4J2_prot	1299
873	A0A1C5L2R1	Cas9_A0A1C5L2R1_prot	1300
874	A0A1D3UYE2	Cas9_A0A1D3UYE2_prot	1301
875	R9I6A5	Cas9_R9I6A5_prot	1302

[0268]

876	A0A0M1W3D2	Cas9_A0A0M1W3D2_prot	1303
877	R7NZZ9	Cas9_R7NZZ9_prot	1304
878	A0A0P7AYC1	Cas9_A0A0P7AYC1_prot	1305
879	F3ZS64	Cas9_F3ZS64_prot	1306
880	B6W3J8	Cas9_B6W3J8_prot	1307
881	I9UHX4	Cas9_I9UHX4_prot	1308
882	F9DDR2	Cas9_F9DDR2_prot	1309
883	A0A069SLB0	Cas9_A0A069SLB0_prot	1310
884	K4I9M9	Cas9_K4I9M9_prot	1311
885	F3PY63	Cas9_F3PY63_prot	1312
886	E5WV33	Cas9_E5WV33_prot	1313
887	R5MDQ9	Cas9_R5MDQ9_prot	1314
888	R5K6G6	Cas9_R5K6G6_prot	1315
889	S0FEG1	Cas9_S0FEG1_prot	1316
890	A0A078PYN7	Cas9_A0A078PYN7_prot	1317
891	E5CB73	Cas9_E5CB73_prot	1318
892	U6RJS5	Cas9_U6RJS5_prot	1319
893	A0A1B7ZF33	Cas9_A0A1B7ZF33_prot	1320
894	C9RJP1	Cas9_C9RJP1_prot	1321
895	I8X6S1	Cas9_I8X6S1_prot	1322
896	A0A0D0IUN5	Cas9_A0A0D0IUN5_prot	1323
897	E1Z024	Cas9_E1Z024_prot	1324
898	A0A173UDH4	Cas9_A0A173UDH4_prot	1325
899	A0A0J9G920	Cas9_A0A0J9G920_prot	1326
900	A0A1C2BS21	Cas9_A0A1C2BS21_prot	1327
901	A0A174HS76	Cas9_A0A174HS76_prot	1328
902	R6V444	Cas9_R6V444_prot	1329
903	A0A167Y4I1	Cas9_A0A167Y4I1_prot	1330
904	R7ZSP8	Cas9_R7ZSP8_prot	1331
905	W4PXW0	Cas9_W4PXW0_prot	1332
906	U2DMI6	Cas9_U2DMI6_prot	1333
907	W4PHU4	Cas9_W4PHU4_prot	1334
908	I4A2W8	Cas9_I4A2W8_prot	1335
909	G8XA12	Cas9_G8XA12_prot	1336
910	A0A1B9E9Q0	Cas9_A0A1B9E9Q0_prot	1337
911	R5CLM1	Cas9_R5CLM1_prot	1338
912	A0A180FK19	Cas9_A0A180FK19_prot	1339
913	R6E3D1	Cas9_R6E3D1_prot	1340
914	A0A101CN94	Cas9_A0A101CN94_prot	1341
915	A0A0K8QW18	Cas9_A0A0K8QW18_prot	1342
916	A0A1H6BLP7	Cas9_A0A1H6BLP7_prot	1343
917	R5ZG15	Cas9_R5ZG15_prot	1344
918	I0AP30	Cas9_I0AP30_prot	1345
935	Q99ZW2	NLS2_STRP1 (SF370) cas9_Q99ZW2_NLS4_prot	1362
936	A0Q5Y3	NLS2_cas9_A0Q5Y3_NLS4_prot	1363

[0269]

937	J7RUA5	NLS2_cas9_J7RUA5_NLS4_prot	1364
938	G3ECR1	NLS2_cas9-G3ECR1_NLS4_prot	1365
939	J3F2B0	NLS2_cas9_J3F2B0_NLS4_prot	1366
940	Q03JI6	NLS2_cas9_Q03JI6_NLS4_prot	1367
941	C9X1G5	NLS2_cas9_C9X1G5_NLS4_prot	1368
942	Q927P4	NLS2_cas9_Q927P4_NLS4_prot	1369
943	Q8DTE3	NLS2_cas9_Q8DTE3_NLS4_prot	1370
944	Q9CLT2	NLS2_cas9_Q9CLT2_NLS4_prot	1371
945	A1IQ68	NLS2_cas9_A1IQ68_NLS4_prot	1372
946	Q6NKI3	NLS2_cas9_Q6NKI3_NLS4_prot	1373
947	Q0P897	NLS2_cas9_Q0P897_NLS4_prot	1374
948	Q03LF7	NLS2_cas9_Q03LF7_NLS4_prot	1375

[0270] 在优选的实施方式中,本发明的人工核酸分子因此包括编码序列,该编码序列包括下列或由下列组成:编码表2中对应列提供的数据库登录号所限定的Cas9蛋白的核酸序列、或其同源物、变体、片段或衍生物。具体地,编码的Cas9蛋白可优选地包括下列或由下列组成:表2中对应列显示的氨基酸序列、或其同源物、变体、片段或衍生物。

[0271] 具体地,在优选的实施方式中,本发明的人工核酸分子可因此包括编码序列,该编码序列包括下列或由下列组成:编码Cas9蛋白的核酸序列,该Cas9蛋白包括下列或由下列组成:SEQ ID NO:428-1375中任一者限定的氨基酸序列或其(功能性)同源物、变体、片段或衍生物,具体地按递增优选顺序与这些序列中的任意者具有至少5%、10%、20%、30%、40%、50%、60%、70%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、或99%,优选至少70%,更优选至少80%,甚至更优选至少85%,甚至更优选至少90%,并且最优选至少95%或甚至97%序列同一性的氨基酸序列。

[0272] 具体地,编码的Cas9蛋白可优选地包括至少一种核定位信号(NLS),更优选两种NLS,该NLS选自如上限定的NLS2和NLS4。在优选的实施方式中,本发明的人工核酸分子可因此包括编码序列,该编码序列包括下列或由下列组成:编码具有核定位信号的Cas9蛋白的核酸序列,该Cas9蛋白包括下列或由下列组成:SEQ ID NO:426;427;10575;381;382;384;11957;11958-11964或SEQ ID NO:12021-14274中任一者限定的氨基酸序列、或其(功能性)同源物、变体、片段或衍生物,具体地按递增优选顺序与这些序列中的任意者具有至少5%、10%、20%、30%、40%、50%、60%、70%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、或99%,优选至少70%,更优选至少80%,甚至更优选至少85%,甚至更优选至少90%,并且最优选至少95%或甚至97%序列同一性的氨基酸序列。

[0273] 本文设想的优选功能性Cas9变体包括“死Cas9(dCas9)”和“Cas9切口酶”等。

[0274] 术语“dCas9”指代核酸酶去活的Cas9,也称“催化失活的”,“催化死亡的Cas9”或“死Cas9”。这种核酸酶缺乏全部或部分内切核酸酶活性,因此可用于以RNA向导的方式调控基因(Jinek M et al.Science.2012 Aug 17;337(6096):816-21)。dCas9核酸酶包含使Cas9内切核酸酶活性失活的突变,一般在两个催化残基(RuvC-1结构域中的D10A和HNH结构域中的H840A,编号相对于化脓链球菌Cas9,即spCas9)中。然而,也可以突变其他催化残基以降低该核酸酶结构域中的一者或两者的活性。dCas9优选不能切割dsDNA,但保留其与合适的gRNA缔合和特异性结合靶DNA的能力。氨基酸位置D10A和H840A处有变化的Cas9双突变

体使核酸酶和切口酶活性完全失活。基于“dCas9”的Cas9衍生物可用于将额外的效应结构域穿梭至靶DNA序列,从而引起例如CRISPRa或CRISPRi(如本文其他部分所述)。

[0275] 术语“Cas9切口酶”是指不保留在靶核酸序列中引入双链断裂的能力,但保持在靶位点处结合和引入单链断裂的能力的Cas9变体。这种变体一般在Cas9核酸内切酶结构域(HNH和RuvC)中一者但不是两者中包括突变。因此,Cas9中位置D10A或H840A的氨基酸突变——编号相对于化脓链球菌Cas9(即spCas9)——可导致核酸酶催化活性失活并将Cas9转化为切口酶。

[0276] 进一步的Cas9变体是本领域已知的,并且被设想为根据本发明的变体。美国专利申请号20140273226讨论了化脓链球菌Cas9基因、Cas9蛋白和包括宿主特异性密码子优化的Cas9编码序列的Cas9蛋白变体以及Cas9融合蛋白。美国专利申请号20140315985教导了大量示例性野生型Cas9多肽(例如,美国专利申请号20140273226的SEQ ID NO:1-256、SEQ ID NO:795-1346),其包括来自化脓链球菌的Cas9的序列(美国专利申请号20140273226的SEQ ID NO:8)。还讨论了Cas9蛋白的修饰和变体。这些参考文献的公开内容整体并入本文。

[0277] 在进一步实施方式中,根据本发明的人工核酸编码Cas9蛋白、或其同种型、同源物、变体、片段或衍生物,如PCT/EP2017/076775(其整体通过引用并入本文)的表2所示。例如,本发明人工核酸可因此包括至少一种编码Cas9蛋白的编码序列,该Cas9蛋白包括下列或由下列组成:PCT/EP2017/076775的SEQ ID NO:428-1345或1362-1375中任一者限定的氨基酸序列、或其(功能性)同种型、同源物、变体、片段或衍生物。

[0278] 核酸序列

[0279] 在优选的实施方式中,本发明的人工核酸分子可包括编码序列,该编码序列包括下列或由下列组成:编码本文限定的Cas9蛋白的核酸序列,其中所述核酸序列由下列中的任一者限定:SEQ ID NO:412;3474-3887;2314-2327;4634-4647;5794-5807;6954-6967;8114-8127;413-425;3490-3503;3506-3519;3522-3535;3538-3551;3554-3567;3570-3583;3586-3599;3602-3615;3618-3631;3634-3647;3650-3663;3666-3679;3682-3695;9514-9527;9626-9639;9738-9751;9850-9863;9962-9975,10074-10087;10186-10199;10298-10311;2330-2343;2346-2359;2362-2375;2378-2391;2394-2407;2410-2423;2426-2439;2442-2455;2458-2471;2474-2487;2490-2503;2506-2519;2522-2535;9498-9511;9610-9623;9722-9735;9834-9847;9946-9959;10058-10071;10170-10183-10282-10295;4650-4663;4666-4679;4682-4695;4698-4711;4714-4727;4730-4743;4746-4759;4762-4775;4778-4791;4794-4807;4810-4823;4826-4839;4842-4855;9530-9543;9642-9655;9754-9767;9866-9879;9978-9991;10090-10103;10202-10215;10314-10327;5810-5823;5826-5839;5842-5855;5858-5871;5874-5887;5890-5903;5906-5919;5922-5935;5938-5951;5954-5967;5970-5983,5986-5999;6002-6015;9546-9559;9658-9671;9770-9783;9882-9895;9994-10007;10106-10119;10218-10231;10330-10343;6970-6983;6986-6999;7002-7015;7018-7031;7034-7047;7050-7063;7066-7079;7082-7095;7098-7111;7114-7127;7130-7143;7146-7159;7162-7175;9562-9575;9674-9687;9786-9799;9898-9911;10010-10023;10122-10135;10234-10247;10346-10359;8130-8143;8146-8159;8162-8175;8178-8191;8194-8207;8210-8223;8226-8239;8242-8255;8258-8271;8274-8287;8290-8302;8306-8319;8322-8335;9578-9591;9690-9703;9802-9815;9914-9927;10026-



10039;10138-10151;10250-10263;10362-10375;9290-9303;9306-9319;9322-9335;9338-9351;9354-9367;9370-9383;9386-9399;9402-9415;9418-9431;9434-9447;9450-9463;9466-9479;9482-9495;9594-9607;9706-9719;9818-9831;9930-9943;10042-10055;10154-10167;10266-10279;10378-10391;27;996-1009;2156-2169;3316-3329;4476-4489;5636-5649;6796-6809;7956-7969;1010-1913;2170-3073;3330-4233;4490-5393;5650-6553;6810-7713;7970-8873,或其(功能性)同源物、变体、片段或衍生物,具体地按递增优选顺序与这些序列中的任意者具有至少5%、10%、20%、30%、40%、50%、60%、70%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、或99%,优选至少70%,更优选至少80%,甚至更优选至少85%,甚至更优选至少90%,并且最优选至少95%或甚至97%序列同一性的核酸序列。

[0280] 在优选的实施方式中,本发明人工核酸进一步在其编码区中编码至少一种核定位信号。编码核定位信号(一种或多种)的核酸序列优选地融合至本文限定的编码Cas9蛋白、或其同源物、变体、片段或衍生物的核酸,从而促进所述Cas9蛋白、或其同源物、变体、片段或衍生物运送到细胞核中。在优选的实施方式中,人工核酸因此包含下列或由下列组成:编码融合至少一种核定位信号的Cas9蛋白、或其同源物、变体、片段或衍生物的核酸序列——所述核酸序列优选由下列中的任一者限定:SEQ ID NO:409;2538;410;2539;10551;10581;11973;11974-11980;1378;3698;4858;6018;7178;8338;1379;3699;4859;6019;7179;8339;10593;10584;10587;10590;10593;10596;11965;11981;11989;11997;12005;12013;11966-11972;11982-11988;11990-11996;11998-12004;12006-12012;12014-12020——或编码蛋白SEQ ID NO:12021-14274,或其(功能性)同源物、变体、片段或衍生物的任何核酸序列,具体地按递增优选顺序与这些序列中的任意者具有至少5%、10%、20%、30%、40%、50%、60%、70%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、或99%,优选至少70%,更优选至少80%,甚至更优选至少85%,甚至更优选至少90%,并且最优选至少95%或甚至97%序列同一性的核酸序列。

[0281] 本发明设想CRISPR相关蛋白质编码区与本文限定的UTR的有益组合,以优选地增加所述编码蛋白质的表达。在优选的实施方式中,人工核酸因此包含下列或由下列组成:编码融合至少一种核定位信号的Cas9蛋白或其同源物、变体、片段或衍生物的核酸序列——所述核酸序列优选由下列中的任一者限定:SEQ ID NO:412;3474-3887;2314-2327;4634-4647;5794-5807;6954-6967;8114-8127;413-425;3490-3503;3506-3519;3522-3535;3538-3551;3554-3567;3570-3583;3586-3599;3602-3615;3618-3631;3634-3647;3650-3663;3666-3679;3682-3695;9514-9527;9626-9639;9738-9751;9850-9863;9962-9975;10074-10087;10186-10199;10298-10311;2330-2343;2346-2359;2362-2375;2378-2391;2394-2407;2410-2423;2426-2439;2442-2455;2458-2471;2474-2487;2490-2503;2506-2519;2522-2535;9498-9511;9610-9623;9722-9735;9834-9847;9946-9959;10058-10071;10170-10183-10282-10295;4650-4663;4666-4679;4682-4695;4698-4711;4714-4727;4730-4743;4746-4759;4762-4775;4778-4791;4794-4807;4810-4823;4826-4839;4842-4855;9530-9543;9642-9655;9754-9767;9866-9879;9978-9991;10090-10103;10202-10215;10314-10327;5810-5823;5826-5839;5842-5855;5858-5871;5874-5887;5890-5903;5906-5919;5922-5935;5938-5951;5954-5967;5970-5983;5986-5999;6002-6015;

9546-9559;9658-9671;9770-9783;9882-9895;9994-10007;10106-10119;10218-10231;10330-10343;6970-6983;6986-6999;7002-7015;7018-7031;7034-7047;7050-7063;7066-7079;7082-7095;7098-7111;7114-7127;7130-7143;7146-7159;7162-7175;9562-9575;9674-9687;9786-9799;9898-9911;10010-10023;10122-10135;10234-10247;10346-10359;8130-8143;8146-8159;8162-8175;8178-8191;8194-8207;8210-8223;8226-8239;8242-8255;8258-8271;8274-8287;8290-8302;8306-8319;8322-8335;9578-9591;9690-9703;9802-9815;9914-9927;10026-10039;10138-10151;10250-10263;10362-10375;9290-9303;9306-9319;9322-9335;9338-9351;9354-9367;9370-9383;9386-9399;9402-9415;9418-9431;9434-9447;9450-9463;9466-9479;9482-9495;9594-9607;9706-9719;9818-9831;9930-9943;10042-10055;10154-10167;10266-10279;10378-10391,或其(功能性)同源物、变体、片段或衍生物,具体地按递增优选顺序与这些序列中的任意者具有至少5%、10%、20%、30%、40%、50%、60%、70%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、或99%,优选至少70%,更优选至少80%,甚至更优选至少85%,甚至更优选至少90%,并且最优选至少95%或甚至97%序列同一性的核酸序列。

[0282] 本发明设想CRISPR相关蛋白质编码区与本文限定的UTR的有益组合,以优选地增加所述编码蛋白质的表达。在优选的实施方式中,人工核酸因此包含下列或由下列组成:编码融合至少一种核定位信号的Cas9蛋白或其同源物、变体、片段或衍生物的核酸序列——所述核酸序列优选地由选自下列的任何SEQ ID NO限定:SEQ ID NO:14274、SEQ ID NO:14275、SEQ ID NO:14276、SEQ ID NO:14277、SEQ ID NO:14278、SEQ ID NO:14279、SEQ ID NO:14280、SEQ ID NO:14281、和SEQ ID NO:14282;更优选SEQ ID NO:14281、SEQ ID NO:417 (HSD17B4/PSMB3.1即构建体HSD17B4-NLS2\_STRP1 (SF370)-cas9\_HsOpt-NLS4\_PSMB3.1;Hsopt=Homo sapiens optimization)或SEQ ID NO:414 (Slc7a3.1/Gnas1,即构建体Slc7a3.1-NLS2\_STRP1 (SF370)-cas9\_HsOpt-NLS4\_Gnas.1)。

[0283] 有利地,公开的任何Cas9序列可被选择用于本发明的用途,即上文提到,即本文和/或序列表中公开的任何序列,即Cas9蛋白质序列和编码不同形式的对应Cas9蛋白质序列即WT或优化序列的mRNA。

[0284] 进一步,通过参考PMID 29535594 (其通过引用并入本文),通过Cas9切口酶从亨廷顿基因精确切除CAG束被包括在本发明的教导内。此外,参考PMID 29456189 (其通过引用并入本文),来自脑膜炎奈瑟氏球菌 (*Neisseria meningitidis*) 的天然CRISPR-Cas9系统进行的可编程RNA切割和识别通过被包括在本发明的教导内。通过参考PMID 29499139 (其通过引用并入本文),此外,通过空肠弯曲杆菌 (*Campylobacter jejuni*) Cas9进行的CRISPR RNA依赖性结合和切割内源RNA被包括在本发明的教导内。通过参考PMID 29224783 (其通过引用并入本文),通过CRISPR/Cas9介导的反式-表观遗传调控进行的体内靶基因活化也被包括在本发明的教导内。

[0285] 在进一步的实施方式中,本发明还设想表2A所示的核酸序列。

[0286] 表2A:本发明的进一步优选的优化Cas9序列

[0287]

5' UTR	包括 NLS 的 Cas9	3'UTR	SEQ ID NO
Slc7a3.1	NLS2_STRP1 (SF370) -cas9 (opt1) _NLS4	Gnas.1	14521
Ubqln2.1	NLS2_STRP1 (SF370) -cas9 (opt1) _NLS4	RPS9.1	14522
HSD17B4	NLS2_STRP1 (SF370) -cas9 (opt1) _NLS4	PSBM3	14523
HSD17B4	NLS2_STRP1 (SF370) -cas9 (opt1) _NLS4	Gnas.1	14524
Nosip.1	NLS2_STRP1 (SF370) -cas9 (opt1) _NLS4	Ndufa1.1	14525
Mp68	NLS2_STRP1 (SF370) -cas9 (opt1) _NLS4	Gnas.1	14526
Mp68	NLS2_STRP1 (SF370) -cas9 (opt1) _NLS4	Ndufa1.1	14527
Slc7a3.1	NLS2_STRP1 (SF370) -cas9 (opt2) _NLS4	Gnas.1	14528
Ubqln2.1	NLS2_STRP1 (SF370) -cas9 (opt2) _NLS4	RPS9.1	14529
HSD17B4	NLS2_STRP1 (SF370) -cas9 (opt2)	PSBM3	14530

[0288]

	_NLS4		
HSD17B4	NLS2_STRP1 (SF370) -cas9 (opt2) _NLS4	Gnas.1	14531
Nosip.1	NLS2_STRP1 (SF370) -cas9 (opt2) _NLS4	Ndufa1.1	14532
Mp68	NLS2_STRP1 (SF370) -cas9 (opt2) _NLS4	Gnas.1	14533
Mp68	NLS2_STRP1 (SF370) -cas9 (opt2) _NLS4	Ndufa1.1	14534
Slc7a3.1	NLS2_STRP1 (SF370) -cas9 (opt10) _NLS4	Gnas.1	14535
Ubqln2.1	NLS2_STRP1 (SF370) -cas9 (opt10) _NLS4	RPS9.1	14536
HSD17B4	NLS2_STRP1 (SF370) -cas9 (opt10) _NLS4	PSBM3	14537
HSD17B4	NLS2_STRP1 (SF370) -cas9 (opt10) _NLS4	Gnas.1	14538
Nosip.1	NLS2_STRP1 (SF370) -cas9 (opt10) _NLS4	Ndufa1.1	14539
Mp68	NLS2_STRP1 (SF370) -cas9 (opt10) _NLS4	Gnas.1	14540
Mp68	NLS2_STRP1 (SF370) -cas9 (opt10) _NLS4	Ndufa1.1	14541

[0289] 在进一步的实施方式中,如表2A所示,NLS2\_STRP1 (SF370) -cas9\_HsOpt\_NLS4 (SEQ ID NO:412) 与UTR组合进行组合,即与Slc7a3.1 (SEQ ID NO:15/16) /Gnas.1 (SEQ ID NO:29/30) 组合。

[0290] 表2B:本发明的进一步优选的Cas9序列

[0291]

5' UTR	包括 NLS 的 Cas9 (Hsopt)	3'UTR	SEQ ID NO
Slc7a3.1	NLS2_STRP1 (SF370) -cas9_HsOpt_NLS4	Gnas.1	414
Ubqln2.1	NLS2_STRP1 (SF370) -cas9_HsOpt_NLS4	RPS9.1	14542
HSD17B4 (V2)	NLS2_STRP1 (SF370) -cas9_HsOpt_NLS4	PSBM3	417
HSD17B4 (V2)	NLS2_STRP1 (SF370) -cas9_HsOpt_NLS4	Gnas.1	14543
Nosip.1	NLS2_STRP1 (SF370) -cas9_HsOpt_NLS4	Ndufa1.1	14544
Mp68	NLS2_STRP1 (SF370) -cas9_HsOpt_NLS4	Gnas.1	14545
Mp68	NLS2_STRP1 (SF370) -cas9_HsOpt_NLS4	Ndufa1.1	14546

[0292] 在进一步实施方式中,根据本发明的人工核酸序列可包含下列或由下列组成:根据PCT/EP2017/076775的SEQ ID NO:1380-1393;1394-2297;2314-2327;2330-2343;2346-2359;2362-2375;2378-2391;2394-2407;2410-2423;2426-2439;2442-2455;2458-2471;2474-2487;2490-2503;2506-2519;2522-2535;9498-9511;9610-9623;9722-9735;9834-9847;9946-9959;10058-10071;10170-10183-10282-10295;2540-2553;2554-3457;3474-3887;3490-3503;3506-3519;3522-3535;3538-3551;3554-3567;3570-3583;3586-3599;3602-3615;3618-3631;3634-3647;3650-3663;3666-3679;3682-3695;9514-9527;9626-9639;9738-9751;9850-9863;9962-9975;10074-10087;10186-10199;10298-10311;3700-3713;3714-4617;4634-4647;4650-4663;4666-4679;4682-4695;4698-4711;4714-4727;4730-4743;4746-4759;4762-4775;4778-4791;4794-4807;4810-4823;4826-4839;4842-4855;9530-9543;9642-9655;9754-9767;9866-9879;9978-9991;10090-10103;10202-10215;10314-10327;4860-4873;4874-5777;5794-5807;5810-5823;5826-5839;5842-5855;5858-5871;5874-5887;5890-5903;5906-5919;5922-5935;5938-5951;5954-5967;5970-5983;5986-5999;6002-6015;9546-9559;9658-9671;9770-9783;9882-9895;9994-10007;10106-10119;10218-10231;10330-10343;6020-6033;6034-6937;6954-6967;6970-6983;6986-6999;7002-7015;7018-7031;7034-7047;7050-7063;7066-7079;7082-7095;7098-7111;7114-7127;7130-7143;7146-7159;7162-7175;9562-9575;9674-9687;9786-9799;9898-9911;10010-10023;10122-10135;10234-10247;10346-10359;7180-7193;7194-8097;8114-8127;8130-8143;8146-8159;8162-8175;8178-8191;8194-8207;8210-8223;8226-8239;8242-8255;8258-8271;8274-8287;8290-8302;8306-8319;8322-8335;9578-9591;9690-9703;9802-9815;9914-9927;10026-10039;10138-10151;10250-10263;10362-10375;8340-8353;8354-9257;9274-9287;9290-9303;9306-9319;9322-9335;9338-9351;9354-9367;9370-9383;9386-9399;9402-9415;9418-9431;9434-9447;9450-9463;9466-9479;9482-9495;9594-9607;9706-9719;9818-9831;9930-9943;10042-10055;10154-10167;10266-10279;10378-10391;411;1380-1393;2540-2553;3700-3713;4860-4873;6020-6033;7180-7193;8340-8353;1394-2297;2554-3457;3714-4617;4874-5777;6034-6937;7194-8097;8354-9257;2314-2327;3474-3887;4634-4647;5794-5807;6954-

6967;8114-8127;9274-9287;413-425;2330-2343;2346-2359;2362-2375;2378-2391;2394-2407;2410-2423;2426-2439;2442-2455;2458-2471;2474-2487;2490-2503;2506-2519;2522-2535;9498-9511;9610-9623;9722-9735;9834-9847;9946-9959;10058-10071;10170-10183-10282-10295;3490-3503;3506-3519;3522-3535;3538-3551;3554-3567;3570-3583;3586-3599;3602-3615;3618-3631;3634-3647;3650-3663;3666-3679;3682-3695;9514-9527;9626-9639;9738-9751;9850-9863;9962-9975,10074-10087;10186-10199;10298-10311;4650-4663;4666-4679;4682-4695;4698-4711;4714-4727;4730-4743;4746-4759;4762-4775;4778-4791;4794-4807;4810-4823;4826-4839;4842-4855;9530-9543;9642-9655;9754-9767;9866-9879;9978-9991;10090-10103;10202-10215;10314-10327;5810-5823;5826-5839;5842-5855;5858-5871;5874-5887;5890-5903;5906-5919;5922-5935;5938-5951;5954-5967;5970-5983,5986-5999;6002-6015;9546-9559;9658-9671;9770-9783;9882-9895;9994-10007;10106-10119;10218-10231;10330-10343;6970-6983;6986-6999;7002-7015;7018-7031;7034-7047;7050-7063;7066-7079;7082-7095;7098-7111;7114-7127;7130-7143;7146-7159;7162-7175;9562-9575;9674-9687;9786-9799;9898-9911;10010-10023;10122-10135;10234-10247;10346-10359;8130-8143;8146-8159;8162-8175;8178-8191;8194-8207;8210-8223;8226-8239;8242-8255;8258-8271;8274-8287;8290-8302;8306-8319;8322-8335;9578-9591;9690-9703;9802-9815;9914-9927;10026-10039;10138-10151;10250-10263;10362-10375;9290-9303;9306-9319;9322-9335;9338-9351;9354-9367;9370-9383;9386-9399;9402-9415;9418-9431;9434-9447;9450-9463;9466-9479;9482-9495;9594-9607;9706-9719;9818-9831;9930-9943;10042-10055;10154-10167;10266-10279;10378-10391的核酸序列、或其(功能性)同源物、片段、变体或衍生物。

[0293] Cpf1

[0294] “Cpf1”(“来自普氏菌和弗朗西斯氏菌1的CRISPR”)或“Cas12”是指属于推定的2类V CRISPR-Cas系统(Zetsche et al.,Cell.2015 Oct 22;163(3):759-771)的RNA向导的DNA核酸内切酶、及其同源物、变体、片段和衍生物。Cpf1编码基因包括土拉弗朗西斯菌(*Francisella tularensis*)亚种novacida(菌株U112)cpf1基因(NCBI参考序列:NZ\_CP009633.1,“AW25\_RS03035”)或其同源物、变体或片段。基于序列分析,Cpf1仅包含一个可检测的RuvC核酸内切酶结构域和第二个推定的新核酸酶(NUC)结构域(Zetsche et al.,Cell.2015 Oct 22;163(3):759-771;Gao et al.Cell Res.2016 Aug;26(8):901-13)。

[0295] Cpf1蛋白优选与crRNA缔合,形成Cpf1:crRNA复合物,该复合物优选能够与靶DNA序列特异性地相互作用。Cpf1:crRNA复合物优选地能够有效切割靶DNA——通过位于靶DNA的5'短富T前间隔序列相邻基序(PAM)进行,并且可引入交错的DNA双链断裂——具有4或5-nt 5'突出端(overhang)。

[0296] 氨基酸序列

[0297] 几种Cpf1蛋白质是本领域已知的,并且在本发明的环境中被设想为CRISPR相关蛋白质。合适的Cpf1蛋白列于下表3中。其中,每行相应于通过其数据库登录号标识的Cpf1蛋白(第一列,“A”,“登录号”)。表3中的第二列(“B”)表示相应于本文提供的对应氨基酸序列的SEQ ID NO。优选的Cpf1蛋白显示在序列表中的SEQ ID NO:1346-1347;10576-10577;和

1348-1361。作为本发明优选实施方式的相应优化mRNA序列显示在序列表中的SEQ ID NO: 10552;3458-3459;3460-34732298-2299;4618-4619;5778-5779;6938-6939;8098-8099;9258-9259;2300-2313;4620-4633;5780-5793;6940-6953;8100-8113;和9260-9273。

[0298] 表3:Cpf1蛋白

行	A 列	B 列
	登录号	SEQ ID NO
1	U2UMQ6	1346
2	A0Q7Q2	1347
3	A8WNM2	1348
4	E3LGD2	1349
5	A0A182DWE3	1350
6	A0A0B6KQP9	1351
7	A0A0E1N6W4	1352
8	V6HCU8	1353
9	A0A0E1N9S2	1354
10	A0A1B8PW75	1355
11	A0A1J0L0B6	1356
12	A0A1F3JTA5	1357
13	A0A1G2R4W1	1358
14	A0A1F5S360	1359
15	A0A1F5ENJ2	1360
16	A0A1J4U637	1361
17	U2UMQ6 (NLS2_cpf1_U2UMQ6_NLS4_prot)	1376
18	A0Q7Q2 (NLS2_cpf1_A0Q7Q2_NLS4_prot)	1377

[0301] 在优选的实施方式中,本发明的人工核酸分子因此包括编码序列,该编码序列包括下列或由下列组成:编码表3中对列提供的数据库登录号所限定的Cpf1蛋白的核酸序列、或其同源物、变体、片段或衍生物。具体地,编码Cpf1蛋白可优选地包括下列或由下列组成:表3中对列显示的氨基酸序列、或其同源物、变体、片段或衍生物。

[0302] 具体地,在优选的实施方式中,本发明的人工核酸分子可因此包括编码序列,该编码序列包括下列或由下列组成:编码Cpf1蛋白的核酸序列,该Cpf1蛋白包括下列或由下列组成:SEQ ID NO:1346-1347;10576-10577;或1348-1361中任一者限定的氨基酸序列;或其(功能性)同源物、变体、片段或衍生物,具体地按递增优选顺序与这些序列中的任意者具有至少5%、10%、20%、30%、40%、50%、60%、70%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、或99%,优选至少70%,更优选至少80%,甚至更优选至少85%,甚至更优选至少90%,并且最优选至少95%或甚至97%序列同一性的核酸序列。

[0303] 具体地,编码的Cpf1蛋白可优选地包括至少一种核定位信号(NLS),更优选两个NLS,该NLS选自如上限定的NLS2和NLS4。在优选的实施方式中,本发明的人工核酸分子可因此包括编码序列,该编码序列包括下列或由下列组成:编码具有核定位信号的Cpf1蛋白的

核酸序列,该Cpf1蛋白包括下列或由下列组成:SEQ ID NO:992-993中任一者限定的氨基酸序列,或其(功能性)同源物、变体、片段或衍生物;包括下列或由下列组成:按递增优选顺序与这些序列中的任意者具有至少5%、10%、20%、30%、40%、50%、60%、70%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、或99%,优选至少70%,更优选至少80%,甚至更优选至少85%,甚至更优选至少90%,并且最优选至少95%或甚至97%序列同一性的氨基酸序列。

[0304] 在进一步实施方式中,根据本发明的人工核酸编码Cpf1蛋白、或其同种型、同源物、变体、片段或衍生物,如PCT/EP2017/076775(其整体通过引用并入本文)的表3所示。具体地,本发明的人工核酸(RNA)分子可因此包括编码Cpf1蛋白的编码序列,该Cpf1蛋白包括下列或由下列组成:PCT/EP2017/076775的SEQ ID NO:1346-1361、或1376-1377或10576-10577中任一者限定的氨基酸序列、或其(功能性)同种型、同源物、变体、片段或衍生物。

[0305] 核酸序列

[0306] 在优选的实施方式中,本发明的人工核酸分子可包括编码序列,该编码序列包括下列或由下列组成:编码本文限定的Cpf1蛋白的核酸序列,其中所述核酸序列由下列中的任一者限定:SEQ ID NO:3488-3489;10396;2328-2329;10395;4648-4649;10397;5808-5809;10398;6968-6969;10399;8128-8129;10400;9274-9287;3504-3505;3520-3521;3536-3537;3552-3553;3568-3669;3584-3585;3600-3601;3616-3617;3632-3633;3648-3649;3664-3665;3680-3681;3696-3697;9528-9529;9640-9641;9752-9753;9864-9865;9976-9977;10088-10089;10200-10201;10312-10313;10403;10410;10417;10424;10431;10438;10445;10452;10459;10466;10473;10480;10487;10494;10501;10508;10515;10522;10529;10536;10543;2344-2345;2360-2361;2376-2377;2392-2393;2408-2409;2424-2425;2440-2441;2456-2457;2472-2473;2489-2490;2504-2505;2520-2521;2536-2537;9512-9513;9624-9625;9736-9737;9848-9849;9960-9961;10072-10073;10184-10185;10296-10297;10402;10409;10416;10423;10430;10437;10444;10451;10458;10465;10472;10479;10486;10493;10500;10507;10514;10521;10528;10535;10542;4664-4665;4680-4681;4696-4697;4712-4713;4728-4729;4744-4745;4760-4761;4776-4777;4792-4793;4808-4809;4824-4825;4840-4841;4856-4857;9544-9545;9656-9657;9768-9769;9880-9881;9992-9993;10104-10105;10216-10217;10328-10329;10404;10411;10418;10425;10432;10439;10446;10453;10460;10467;10474;10481;10488;10495;10502;10509;10516;10523;10530;10537;10544;5824-5825;5840-5841;5856-5857;5872-5873;5888-5889;5904-5905;5920-5921;5936-5937;5952-5953;5968-5969;5984-5985;6000-6001;6016-6017;9560-9561;9672-9673;9784-9785;9896-9897;10008-10009;10120-10121;10232-10233;10344-10345;10405;10412;10419;10426;10433;10440;10447;10454;10461;10468;10475;10482;10489;10496;10503;10510;10517;10524;10531;10538;10545;7033;7048-7049;7064-7065;7080-7081;7096-7097;7112-7113;7128-7129;7144-7145;7160-7161;7176-7177;9576-9577;9688-9689;9800-9801;9912-9913;10024-10025;10136-10137;10248-10249;10360-10361;10406;10413;10420;10427;10434;10441;10448;10455;10462;10469;10476;10483;10490;10497;10504;10511;10518;10525;10532;10539;10546;8144-8145;8160-8160;8176-8177;8192-8193;8208-

8209;8224-8225;8240-8241;8256-8257;8272-8273;8288-8289;8304-8305;8320-8321;8336-8337;9592-9593;9704-9705;9816-9817;9928-9929;10040-10041;10152-10153;10264-10265;10376-10377;10407;10414;10421;10428;10435;10442;10449;10456;10463;10470;10477;10484;10491;10498;10505;10512;10519;10526;10533;10540;10547;9288-9289;10401;10553;10582-10583;10579-10580;10585-10586;10588-10589;10591-10592;10594-10595;10597-10598;10554-10574;10601;10602;10615;10616;10629;10630;10643;10644;10657;10658;10671;10672;10685;10686;10699;10700;10713;10714;10727;10728;10741;10742;10755;10756;10769;10770;10783;10784;10797;10798;10811;10812;10825;10826;10839;10840;10853;10854;10867;10868;10881;10882;10603;10604;10617;10618;10631;10632;10645;10646;10659;10660;10673;10674;10687;10688;10701;10702;10715;10716;10729;10730;10743;10744;10757;10758;10771;10772;10785;10786;10799;10800;10813;10814;10827;10828;10841;10842;10855;10856;10869;10870;10883;10884;10605;10606;10619;10620;10633;10634;10647;10648;10661;10662;10675;10676;10689;10690;10703;10704;10717;10718;10731;10732;10745;10746;10759;10760;10773;10774;10787;10788;10801;10802;10815;10816;10829;10830;10843;10844;10857;10858;10871;10872;10885;10886;10607;10608;10621;10622;10635;10636;10649;10650;10663;10664;10677;10678;10691;10692;10705;10706;10719;10720;10733;10734;10747;10748;10761;10762;10775;10776;10789;10790;10803;10804;10817;10818;10831;10832;10845;10846;10859;10860;10873;10874;10887;10888;10609;10610;10623;10624;10637;10638;10651;10652;10665;10666;10679;10680;10693;10694;10707;10708;10721;10722;10735;10736;10749;10750;10763;10764;10777;10778;10791;10792;10805;10806;10819;10820;10833;10834;10847;10848;10861;10862;10875;10876;10889;10890;10611;10612;10625;10626;10639;10640;10653;10654;10667;10668;10681;10682;10695;10696;10709;10710;10723;10724;10737;10738;10751;10752;10765;10766;10779;10780;10793;10794;10807;10808;10821;10822;10835;10836;10849;10850;10863;10864;10877;10878;10891;10892;9304-9305;9320-9321;9336-9337;9352-9353;9368-9369;9384-9385;9400-9401;9416-9417;9432-9433;9448-9449;9464-9465;9480-9481;9496-9497;9608-9609;9720-9721;9832-9833;9944-9945;10056-10057;10168-10169;10280-10281;10392-10393;10408;10415;10422;10429;10436;10443;10450;10457;10464;10471;10478;10485;10492;10499;10506;10513;10520;10527;10534;10541;10548;或其(功能性)同源物、变体、片段或衍生物,具体地按递增优选顺序与这些序列中的任意者具有至少5%、10%、20%、30%、40%、50%、60%、70%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、或99%,优选至少70%,更优选至少80%,甚至更优选至少85%,甚至更优选至少90%,并且最优选至少95%或甚至97%序列同一性的核酸序列。

[0307] 在另一优选的实施方式中,本发明的人工核酸分子可包括编码序列,该编码序列包括下列或由下列组成:编码本文限定的Cpf1蛋白的核酸序列,其中所述核酸序列由下列中的任一者限定:SEQ ID NO:10549(即AsCpf1=32L4\_AsCpf1(Hsopt)-NLS3-3xHA-标签\_白



蛋白7)或SEQ ID NO:10550(即LbCpf1=32L4\_LbCpf1(Hsopt)-NLS3-3xHA-标签\_白蛋白7);或其(功能性)同源物、变体、片段或衍生物,具体地按递增优选顺序与这些序列中的任意者具有至少5%、10%、20%、30%、40%、50%、60%、70%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、或99%,优选至少70%,更优选至少80%,甚至更优选至少85%,甚至更优选至少90%,并且最优选至少95%或甚至97%序列同一性的核酸序列。

[0308] 在优选的实施方式中,本发明的人工核酸进一步在其编码区中编码至少一种核定位信号。编码核定位信号(一种或多种)的核酸序列优选地融合至本文限定的编码Cpf1蛋白或其同源物、变体、片段或衍生物的核酸,从而促进所述Cpf1蛋白或其同源物、变体、片段或衍生物运送到细胞核中。在优选的实施方式中,人工核酸因此包含下列或由下列组成:编码融合至少一种核定位信号的Cpf1蛋白或其同源物,片段、变体或衍生物的核酸序列,所述核酸序列优选由下列中的任一者限定:SEQ ID NO:10551;10581;10593;10584;10587;10590;10593;10596;409;2538;410;2539;10551;10581;11973;11974-11980;1378;3698;4858;6018;7178;8338;1379;3699;4859;6019;7179;8339;10593;10584;10587;10590;10593;10596;11965;11981;11989;11997;12005;12013;11966-11972;11982-11988;11990-11996;11998-12004;12006-12012;12014-12020;或编码SEQ ID NO:12021-14274中的任一者或组合的核酸、或其(功能性)同源物、变体、片段或衍生物,具体地按递增优选顺序与这些序列中的任意者具有至少5%、10%、20%、30%、40%、50%、60%、70%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、或99%,优选至少70%,更优选至少80%,甚至更优选至少85%,甚至更优选至少90%,并且最优选至少95%或甚至97%序列同一性的核酸序列。

[0309] 有利地,根据本发明的人工核酸的编码区中编码CRISPR相关蛋白质如Cpf1或其同源物、变体、片段或衍生物的核酸序列与本文限定的UTR组合,以优选地增加所述编码蛋白质的表达。在优选的实施方式中,人工核酸包含下列或由下列组成:编码进一步融合至少一种核定位信号的Cpf1蛋白或其同源物、变体或片段的核酸序列,所述核酸序列优选由下列中的任一者限定:SEQ ID NO:3488-3489;10396;2328-2329;10395;4648-4649;10397;5808-5809;10398;6968-6969;10399;8128-8129;10400;9274-9287;3504-3505;3520-3521;3536-3537;3552-3553;3568-3669;3584-3585;3600-3601;3616-3617;3632-3633;3648-3649;3664-3665;3680-3681;3696-3697;9528-9529;9640-9641;9752-9753;9864-9865;9976-9977;10088-10089;10200-10201;10312-10313;10403;10410;10417;10424;10431;10438;10445;10452;10459;10466;10473;10480;10487;10494;10501;10508;10515;10522;10529;10536;10543;2344-2345;2360-2361;2376-2377;2392-2393;2408-2409;2424-2425;2440-2441;2456-2457;2472-2473;2489-2490;2504-2505;2520-2521;2536-2537;9512-9513;9624-9625;9736-9737;9848-9849;9960-9961;10072-10073;10184-10185;10296-10297;10402;10409;10416;10423;10430;10437;10444;10451;10458;10465;10472;10479;10486;10493;10500;10507;10514;10521;10528;10535;10542;4664-4665;4680-4681;4696-4697;4712-4713;4728-4729;4744-4745;4760-4761;4776-4777;4792-4793;4808-4809;4824-4825;4840-4841;4856-4857;9544-9545;9656-9657;9768-9769;9880-9881;9992-9993;10104-10105;10216-10217;10328-10329;10404;

10411;10418;10425;10432;10439;10446;10453;10460;10467;10474;10481;10488;  
10495;10502;10509;10516;10523;10530;10537;10544;5824-5825;5840-5841;5856-  
5857;5872-5873;5888-5889;5904-5905;5920-5921;5936-5937;5952-5953;5968-5969;  
5984-5985;6000-6001;6016-6017;9560-9561;9672-9673;9784-9785;9896-9897;10008-  
10009;10120-10121;10232-10233;10344-10345;10405;10412;10419;10426;10433;  
10440;10447;10454;10461;10468;10475;10482;10489;10496;10503;10510;10517;  
10524;10531;10538;10545;7033;7048-7049;7064-7065;7080-7081;7096-7097;7112-  
7113;7128-7129;7144-7145;7160-7161;7176-7177;9576-9577;9688-9689;9800-9801;  
9912-9913;10024-10025;10136-10137;10248-10249;10360-10361;10406;10413;10420;  
10427;10434;10441;10448;10455;10462;10469;10476;10483;10490;10497;10504;  
10511;10518;10525;10532;10539;10546;8144-8145;8160-8160;8176-8177;8192-8193;  
8208-8209;8224-8225;8240-8241;8256-8257;8272-8273;8288-8289;8304-8305;8320-  
8321;8336-8337;9592-9593;9704-9705;9816-9817;9928-9929;10040-10041;10152-  
10153;10264-10265;10376-10377;10407;10414;10421;10428;10435;10442;10449;  
10456;10463;10470;10477;10484;10491;10498;10505;10512;10519;10526;10533;  
10540;10547;9288-9289;10401;10553;10582-10583

[0310] 10579-10580;10585-10586;10588-10589;10591-10592;10594-10595;10597-  
10598;10554-10574;10601;10602;10615;10616;10629;10630;10643;10644;10657;  
10658;10671;10672;10685;10686;10699;10700;10713;10714;10727;10728;10741;  
10742;10755;10756;10769;10770;10783;10784;10797;10798;10811;10812;10825;  
10826;10839;10840;10853;10854;10867;10868;10881;1088210603;10604;10617;10618;  
10631;10632;10645;10646;10659;10660;10673;10674;10687;10688;10701;10702;  
10715;10716;10729;10730;10743;10744;10757;10758;10771;10772;10785;10786;  
10799;10800;10813;10814;10827;10828;10841;10842;10855;10856;10869;10870;  
10883;10884;10605;10606;10619;10620;10633;10634;10647;10648;10661;10662;  
10675;10676;10689;10690;10703;10704;10717;10718;10731;10732;10745;10746;  
10759;10760;10773;10774;10787;10788;10801;10802;10815;10816;10829;10830;  
10843;10844;10857;10858;10871;10872;10885;10886;10607;10608;10621;10622;  
10635;10636;10649;10650;10663;10664;10677;10678;10691;10692;10705;10706;  
10719;10720;10733;10734;10747;10748;10761;10762;10775;10776;10789;10790;  
10803;10804;10817;10818;10831;10832;10845;10846;10859;10860;10873;10874;  
10887;10888;10609;10610;10623;10624;10637;10638;10651;10652;10665;10666;  
10679;10680;10693;10694;10707;10708;10721;10722;10735;10736;10749;10750;  
10763;10764;10777;10778;10791;10792;10805;10806;10819;10820;10833;10834;  
10847;10848;10861;10862;10875;10876;10889;10890;10611;10612;10625;10626;  
10639;10640;10653;10654;10667;10668;10681;10682;10695;10696;10709;10710;  
10723;10724;10737;10738;10751;10752;10765;10766;10779;10780;10793;10794;  
10807;10808;10821;10822;10835;10836;10849;10850;10863;10864;10877;10878;  
10891;10892;9304-9305;9320-9321;9336-9337;9352-9353;9368-9369;9384-9385;9400-

9401;9416-9417;9432-9433;9448-9449;9464-9465;9480-9481;9496-9497;9608-9609;9720-9721;9832-9833;9944-9945;10056-10057;10168-10169;10280-10281;10392-10393;10408;10415;10422;10429;10436;10443;10450;10457;10464;10471;10478;10485;10492;10499;10506;10513;10520;10527;10534;10541;10548;或其(功能性)同源物、变体、片段或衍生物,具体地按递增优选顺序与这些序列中的任一者具有至少5%、10%、20%、30%、40%、50%、60%、70%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、或99%,优选至少70%,更优选至少80%,甚至更优选至少85%,甚至更优选至少90%,并且最优选至少95%或甚至97%序列同一性的核酸序列。

[0311] 有利地,公开的任何Cpf1序列可被选择用于本发明的用途,即上文提到,即本文和/或序列表中公开的任何序列,即Cpf1蛋白序列和编码不同形式的对应Cpf1蛋白序列即WT或优化序列的mRNA。

[0312] 在进一步实施方式中,根据本发明的人工核酸序列可包含下列或由下列组成:根据PCT/EP2017/076775的SEQ ID NO:2298-2299;3458-3459;4618-4619;5778-5779;6938-6939;8098-8099;9258-9259;2300-2313;3460-3473;4620-4633;5780-5793;6940-6953;8100-8113;9260-9273;2328-2329;10395;3488-3489;10396;4648-4649;10397;5808-5809;10398;6968-6969;10399;8128-8129;10400;9288-9289;10401;2344-2345;2360-2361;2376-2377;2392-2393;2408-2409;2424-2425;2440-2441;2456-2457;2472-2473;2489-2490;2504-2505;2520-2521;2536-2537;9512-9513;9624-9625;9736-9737;9848-9849;9960-9961;10072-10073;10184-10185;10296-10297;10402;10409;10416;10423;10430;10437;10444;10451;10458;10465;10472;10479;10486;10493;10500;10507;10514;10521;10528;10535;10542;3504-3505;3520-3521;3536-3537;3552-3553;3568-3669;3584-3585;3600-3601;3616-3617;3632-3633;3648-3649;3664-3665;3680-3681;3696-3697;9528-9529;9640-9641;9752-9753;9864-9865;9976-9977;10088-10089;10200-10201;10312-10313;10403;10410;10417;10424;10431;10438;10445;10452;10459;10466;10473;10480;10487;10494;10501;10508;10515;10522;10529;10536;10543;4664-4665;4680-4681;4696-4697;4712-4713;4728-4729;4744-4745;4760-4761;4776-4777;4792-4793;4808-4809;4824-4825;4840-4841;4856-4857;9544-9545;9656-9657;9768-9769;9880-9881;9992-9993;10104-10105;10216-10217;10328-10329;10404;10411;10418;10425;10432;10439;10446;10453;10460;10467;10474;10481;10488;10495;10502;10509;10516;10523;10530;10537;10544;5824-5825;5840-5841;5856-5857;5872-5873;5888-5889;5904-5905;5920-5921;5936-5937;5952-5953;5968-5969;5984-5985;6000-6001;6016-6017;9560-9561;9672-9673;9784-9785;9896-9897;10008-10009;10120-10121;10232-10233;10344-10345;10405;10412;10419;10426;10433;10440;10447;10454;10461;10468;10475;10482;10489;10496;10503;10510;10517;10524;10531;10538;10545;6984-6985;7000-7001;1016-7017;7032-7033;7048-7049;7064-7065;7080-7081;7096-7097;7112-7113;7128-7129;7144-7145;7160-7161;7176-7177;9576-9577;9688-9689;9800-9801;9912-9913;10024-10025;10136-10137;10248-10249;10360-10361;10406;10413;10420;10427;10434;10441;10448;10455;10462;

10469;10476;10483;10490;10497;10504;10511;10518;10525;10532;10539;10546;8144-8145;8160-8160;8176-8177;8192-8193;8208-8209;8224-8225;8240-8241;8256-8257;8272-8273;8288-8289;8304-8305;8320-8321;8336-8337;9592-9593;9704-9705;9816-9817;9928-9929;10040-10041;10152-10153;10264-10265;10376-10377;10407;10414;10421;10428;10435;10442;10449;10456;10463;10470;10477;10484;10491;10498;10505;10512;10519;10526;10533;10540;10547;9304-9305;9320-9321;9336-9337;9352-9353;9368-9369;9384-9385;9400-9401;9416-9417;9432-9433;9448-9449;9464-9465;9480-9481;9496-9497;9608-9609;9720-9721;9832-9833;9944-9945;10056-10057;10168-10169;10280-10281;10392-10393;10408;10415;10422;10429;10436;10443;10450;10457;10464;10471;10478;10485;10492;10499;10506;10513;10520;10527;10534;10541;10548;10552;10594-10595;10582-10583;10585-10586;10588-10589;10591-10592;10594-10595;10597-10598;10553;10599;10600;10613;10614;10627;10628;10641;10642;10655;10656;10669;10670;10683;10684;10697;10698;10711;10712;10725;10726;10739;10740;10753;10754;10767;10768;10781;10782;10795;10796;10809;10810;10823;10824;10837;10838;10851;10852;10865;10866;10879;10880;;10601;10602;10615;10616;10629;10630;10643;10644;10657;10658;10671;10672;10685;10686;10699;10700;10713;10714;10727;10728;10741;10742;10755;10756;10769;10770;10783;10784;10797;10798;10811;10812;10825;10826;10839;10840;10853;10854;10867;10868;10881;10882;10603;10604;10617;10618;10631;10632;10645;10646;10659;10660;10673;10674;10687;10688;10701;10702;10715;10716;10729;10730;10743;10744;10757;10758;10771;10772;10785;10786;10799;10800;10813;10814;10827;10828;10841;10842;10855;10856;10869;10870;10883;10884;10605;10606;10619;10620;10633;10634;10647;10648;10661;10662;10675;10676;10689;10690;10703;10704;10717;10718;10731;10732;10745;10746;10759;10760;10773;10774;10787;10788;10801;10802;10815;10816;10829;10830;10843;10844;10857;10858;10871;10872;10885;10886;10607;10608;10621;10622;10635;10636;10649;10650;10663;10664;10677;10678;10691;10692;10705;10706;10719;10720;10733;10734;10747;10748;10761;10762;10775;10776;10789;10790;10803;10804;10817;10818;10831;10832;10845;10846;10859;10860;10873;10874;10887;10888;10609;10610;10623;10624;10637;10638;10651;10652;10665;10666;10679;10680;10693;10694;10707;10708;10721;10722;10735;10736;10749;10750;10763;10764;10777;10778;10791;10792;10805;10806;10819;10820;10833;10834;10847;10848;10861;10862;10875;10876;10889;10890;10611;10612;10625;10626;10639;10640;10653;10654;10667;10668;10681;10682;10695;10696;10709;10710;10723;10724;10737;10738;10751;10752;10765;10766;10779;10780;10793;10794;10807;10808;10821;10822;10835;10836;10849;10850;10863;10864;10877;10878;10891;10892;2298-2299;3458-3459;4618-4619;5778-5779;6938-6939;8098-8099;9258-9259;10552;2300-2313;3460-3473;4620-4633;5780-5793;6940-6953;8100-8113;9260-9273;2314-2327;3474-3887;4634-4647;5794-5807;6954-6967;8114-8127;9274-9287;2328-

2329;10395;3488-3489;10396;4648-4649;10397;5808-5809;10398;6968-6969;10399;  
8128-8129;10400;9288-9289;10401;10553;2344-2345;2360-2361;2376-2377;2392-  
2393;2408-2409;2424-2425;2440-2441;2456-2457;2472-2473;2489-2490;2504-2505;  
2520-2521;2536-2537;9512-9513;9624-9625;9736-9737;9848-9849;9960-9961;10072-  
10073;10184-10185;10296-10297;10402;10409;10416;10423;10430;10437;10444;  
10451;10458;10465;10472;10479;10486;10493;10500;10507;10514;10521;10528;  
10535;10542;3504-3505;3520-3521;3536-3537;3552-3553;3568-3669;3584-3585;3600-  
3601;3616-3617;3632-3633;3648-3649;3664-3665;3680-3681;3696-3697;9528-9529;  
9640-9641;9752-9753;9864-9865;9976-9977;10088-10089;10200-10201;10312-10313;  
10403;10410;10417;10424;10431;10438;10445;10452;10459;10466;10473;10480;  
10487;10494;10501;10508;10515;10522;10529;10536;10543;4664-4665;4680-4681;  
4696-4697;4712-4713;4728-4729;4744-4745;4760-4761;4776-4777;4792-4793;4808-  
4809;4824-4825;4840-4841;4856-4857;9544-9545;9656-9657;9768-9769;9880-9881;  
9992-9993;10104-10105;10216-10217;10328-10329;10404;10411;10418;10425;10432;  
10439;10446;10453;10460;10467;10474;10481;10488;10495;10502;10509;10516;  
10523;10530;10537;10544;5824-5825;5840-5841;5856-5857;5872-5873;5888-5889;  
5904-5905;5920-5921;5936-5937;5952-5953;5968-5969;5984-5985;6000-6001;6016-  
6017;9560-9561;9672-9673;9784-9785;9896-9897;10008-10009;10120-10121;10232-  
10233;10344-10345;10405;10412;10419;10426;10433;10440;10447;10454;10461;  
10468;10475;10482;10489;10496;10503;10510;10517;10524;10531;10538;10545;6984-  
6985;7000-7001;1016-7017;7032-7033;7048-7049;7064-7065;7080-7081;7096-7097;  
7112-7113;7128-7129;7144-7145;7160-7161;7176-7177;9576-9577;9688-9689;9800-  
9801;9912-9913;10024-10025;10136-10137;10248-10249;10360-10361;10406;10413;  
10420;10427;10434;10441;10448;10455;10462;10469;10476;10483;10490;10497;  
10504;10511;10518;10525;10532;10539;10546;8144-8145;8160-8160;8176-8177;8192-  
8193;8208-8209;8224-8225;8240-8241;8256-8257;8272-8273;8288-8289;8304-8305;  
8320-8321;8336-8337;9592-9593;9704-9705;9816-9817;9928-9929;10040-10041;  
10152-10153;10264-10265;10376-10377;10407;10414;10421;10428;10435;10442;  
10449;10456;10463;10470;10477;10484;10491;10498;10505;10512;10519;10526;  
10533;10540;10547;9304-9305;9320-9321;9336-9337;9352-9353;9368-9369;9384-  
9385;9400-9401;9416-9417;9432-9433;9448-9449;9464-9465;9480-9481;9496-9497;  
9608-9609;9720-9721;9832-9833;9944-9945;10056-10057;10168-10169;10280-10281;  
10392-10393;10408;10415;10422;10429;10436;10443;10450;10457;10464;10471;  
10478;10485;10492;10499;10506;10513;10520;10527;10534;10541;10548;10554-  
10574;10594-10595;10582-10583;10585-10586;10588-10589;10591-10592;10594-  
10595;10597-10598;10599;10600;10613;10614;10627;10628;10641;10642;10655;  
10656;10669;10670;10683;10684;10697;10698;10711;10712;10725;10726;10739;  
10740;10753;10754;10767;10768;10781;10782;10795;10796;10809;10810;10823;  
10824;10837;10838;10851;10852;10865;10866;10879;10880;10601;10602;10615;

10616;10629;10630;10643;10644;10657;10658;10671;10672;10685;10686;10699;  
10700;10713;10714;10727;10728;10741;10742;10755;10756;10769;10770;10783;  
10784;10797;10798;10811;10812;10825;10826;10839;10840;10853;10854;10867;  
10868;10881;10882;10603;10604;10617;10618;10631;10632;10645;10646;10659;  
10660;10673;10674;10687;10688;10701;10702;10715;10716;10729;10730;10743;  
10744;10757;10758;10771;10772;10785;10786;10799;10800;10813;10814;10827;  
10828;10841;10842;10855;10856;10869;10870;10883;10884;10605;10606;10619;  
10620;10633;10634;10647;10648;10661;10662;10675;10676;10689;10690;10703;  
10704;10717;10718;10731;10732;10745;10746;10759;10760;10773;10774;10787;  
10788;10801;10802;10815;10816;10829;10830;10843;10844;10857;10858;10871;  
10872;10885;10886;10607;10608;10621;10622;10635;10636;10649;10650;10663;  
10664;10677;10678;10691;10692;10705;10706;10719;10720;10733;10734;10747;  
10748;10761;10762;10775;10776;10789;10790;10803;10804;10817;10818;10831;  
10832;10845;10846;10859;10860;10873;10874;10887;10888;10609;10610;10623;  
10624;10637;10638;10651;10652;10665;10666;10679;10680;10693;10694;10707;  
10708;10721;10722;10735;10736;10749;10750;10763;10764;10777;10778;10791;  
10792;10805;10806;10819;10820;10833;10834;10847;10848;10861;10862;10875;  
10876;10889;10890;10611;10612;10625;10626;10639;10640;10653;10654;10667;  
10668;10681;10682;10695;10696;10709;10710;10723;10724;10737;10738;10751;  
10752;10765;10766;10779;10780;10793;10794;10807;10808;10821;10822;10835;  
10836;10849;10850;10863;10864;10877;10878;10891;10892的核酸序列,或其(功能性)  
同源物、片段、变体或衍生物。

[0313] Cas13、CasX、CasY和其他(内切)核酸酶

[0314] 氨基酸序列

[0315] 几种Cas13、CasX和CasY蛋白是本领域已知的并且在本发明的环境中被设想作为CRISPR相关蛋白质。适当的Cas13、CasX和CasY蛋白显示在SEQ ID NO:10893-10925;10926-10998 (Cas13, 即WP15770004、WP18451595、WP21744063、WP21746774、ERK53440、WP31473346、CVRQ01000008、CRZ35554、WP22785443、WP36091002、WP12985477、WP13443710、ETD76934、WP38617242、WP2664492、WP4343973、WP44065294、ADAR2DD、WP47447901、ERI81700、WP34542281、WP13997271、WP41989581、WP47431796、WP14084666、WP60381855、WP14165541、WP63744070、WP65213424、WP45968377、EH006562、WP6261414、EKB06014、WP58700060、WP13446107、WP44218239、WP12458151、ERJ81987、ERJ65637、WP21665475、WP61156637、WP23846767、ERJ87335、WP5873511、WP39445055、WP52912312、WP53444417、WP12458414、WP39417390、EOA10535、WP61156470、WP13816155、WP5874195、WP39437199、WP39419792、WP39431778、WP46201018、WP39442171、WP39426176、WP39418912、WP39434803、WP39428968、WP25000926、EFU31981、WP4343581、WP36884929、BAU18623、AFJ07523、WP14708441、WP36860899、WP61868553、KJJ86756、EGQ18444、EKY00089、WP36929175、WP7412163、WP44072147、WP42518169、WP44074780、WP15024765、WP49354263、WP4919755、WP64970887、WP61710138); 11002; 11003 (CasX, 即OGP07438、OHB99618); 和11004-11010

(CasY, 即OJI08769、OGY82221、OJI06454、APG80656、OJI07455、OJI09436、PIP58309)。

[0316] 有利地, 根据本发明的人工核酸的编码区中编码CRISPR相关蛋白质如Cas13或其同源物、变体、片段或衍生物的核酸序列与本文限定的UTR组合, 以优选地增加所述编码蛋白质的表达。在优选的实施方式中, 人工核酸包含下列或由下列组成: 编码进一步融合至少一种核定位信号的Cas13蛋白或其同源物、变体或片段的核酸序列, 所述核酸序列优选由下列中的任一者限定: SEQ ID NO: 11011-11042; 11249-11280; 11044-11116; 11282-11354; 11131-11162; 11367-11398; 11485-11516; 11603-11634; 11721-11752; 11839-11870; 11164-11236; 11400-11472; 11518-11590; 11636-11708; 11754-11826; 11872-11944, 或其(功能性)同源物、变体、片段或衍生物, 具体地按递增优选顺序与这些序列中的任一者具有至少5%、10%、20%、30%、40%、50%、60%、70%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、或99%, 优选至少70%, 更优选至少80%, 甚至更优选至少85%, 甚至更优选至少90%, 并且最优选至少95%或甚至97%序列同一性的核酸序列。其他Cas13蛋白也被包括在本发明的公开内容中, 即Cas13a、Cas13b、Cas13c和Cas13d(由PMID 29551514和29551272显而易见, 通过引用并入本文)。而且通过引用并入本文Cas13相关出版物PMID 26593719、28976959和29070703。

[0317] 大量Cas13蛋白是本领域已知的并且在本发明的环境中被设想作为CRISPR相关蛋白质。本发明的优选Cas13d序列是Cas13d蛋白(SEQ ID NO: 14294-14321)和其优化mRNA序列, 具有SEQ ID NO: 14322-14349、SEQ ID NO: 14350-14377、SEQ ID NO: 14378-14405、SEQ ID NO: 14406-14433、SEQ ID NO: 14434-14461、SEQ ID NO: 14462-14489、和SEQ ID NO: 14490-14517。

[0318] 有利地, 根据本发明的人工核酸的编码区中编码CRISPR相关蛋白质如CasX或其同源物、变体、片段或衍生物的核酸序列与本文限定的UTR组合, 以优选地增加所述编码蛋白质的表达。在优选的实施方式中, 人工核酸包含下列或由下列组成: 编码进一步融合至少一种核定位信号的CasX蛋白、或其同源物、变体或片段的核酸序列, 所述核酸序列优选由下列中的任一者限定: SEQ ID NO: 11120-11122; 11240; 11241; 11358; 11359; 11476; 11477; 11594; 11595; 11712; 11713; 11830; 11831; 11948; 11949, 或其(功能性)同源物、变体、片段或衍生物, 具体地按递增优选顺序与这些序列中的任一者具有至少5%、10%、20%、30%、40%、50%、60%、70%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、或99%, 优选至少70%, 更优选至少80%, 甚至更优选至少85%, 甚至更优选至少90%, 并且最优选至少95%或甚至97%序列同一性的核酸序列。

[0319] 有利地, 本文或序列表公开的任何Cas13、CasX或CasY序列可被选择用于本发明的用途, 作为本发明环境中的CRISPR相关蛋白质, 即上文提到的, 即本文和/或序列表公开的任何序列, 即Cas13、CasX或CasY蛋白序列和编码不同形式的对应Cas13、CasX或CasY蛋白序列即WT或优化序列的mRNA。

[0320] 进一步, 如PMID 29160308中所述介导基因组DNA中A • T向G • C转化的腺嘌呤碱基编辑因子(ABE)以及出版物PMID 29160308本身通过引用并入本文。根据作者, ABE比当前基于Cas9核酸酶的方法更有效、更干净地引入点突变, 并且具有更少的脱靶基因组修饰, 并且可以在人细胞中安装疾病纠正或疾病抑制突变。

[0321] ARMAN核酸内切酶

[0322] 通过参考PMID 28005056、20421484和17185602(其通过引用并入本文),在本发明的教导还包括使用来自未培养的微生物的新CRISPR-Cas系统,即ARMAN Cas9,即nanoarchaea ARMAN-1(*Candidatus Micrarchaeum acidiphilum* ARMAN-1)和ARMAN-4(*Candidatus Parvarchaeum acidiphilum* ARMAN-4)。

[0323] 有利地,根据本发明的人工核酸的编码区中编码CRISPR相关蛋白质如CasY、或其同源物、变体、片段或衍生物的核酸序列与本文限定的UTR组合,以优选地增加所述编码蛋白质的表达。在优选的实施方式中,人工核酸包含下列或由下列组成:编码进一步融合至少一种核定位信号的CasY蛋白或其同源物、变体或片段的核酸序列,所述核酸序列优选由下列中的任一者限定:SEQ ID NO:11123-11130;11360-11366;11242-11248;11478-11484;11596-11602;11714-11720;11832-11838;11950-11956,或其(功能性)同源物、变体、片段或衍生物,具体地按递增优选顺序与这些序列中的任一者具有至少5%、10%、20%、30%、40%、50%、60%、70%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、或99%,优选至少70%,更优选至少80%,甚至更优选至少85%,甚至更优选至少90%,并且最优选至少95%或甚至97%序列同一性的核酸序列。

[0324] gRNA

[0325] 如本文所讨论的,功能性CRISPR-Cas系统一般需要存在向导RNA(“gRNA”),向导RNA与CRISPR相关蛋白质缔合并将其招募至互补靶DNA序列。向导RNA的结构和特征一般取决于具体CRISPR相关蛋白质的选择。

[0326] 如本文所用,术语“向导RNA”因此涉及能够将CRISPR相关蛋白质靶向目的靶DNA序列的任何RNA分子。向导RNA(gRNA)优选包含i)能够与靶DNA序列特异性杂交的第一互补区域,和ii)与CRISPR相关蛋白质相互作用的第二区域。

[0327] 所述区域——一般位于gRNA的5'端——包含与靶DNA序列互补的短核苷酸序列,并且在本文中也称为“靶向区域”。术语“区域”是指分子的部分/区段,例如RNA中的连续一段核苷酸。靶向区域可以是约17-20,例如约21-23个核苷酸的长度或可以更长或更短(“截短的gRNA”)。其可以优选通过互补碱基对(即配对碱基)之间的氢键与靶DNA序列相互作用。

[0328] “gRNA”可以具有任何长度,条件是其包括“靶向区域”,以及优选地能够以序列特异性方式将CRISPR相关蛋白质招募至靶DNA序列。因此,“gRNA”可以是至少10,至少11,至少12,更优选至少13,至少14,至少15,和最优选至少16或至少17个核苷酸或至少18个核苷酸或至少19个核苷酸或至少20个核苷酸的长度。在一些实施方式中,“gRNA”包括靶向区域,该靶向区域优选是至少21、至少22、至少23、至少24、至少25个核苷酸或更长的长度;和ii)与CRISPR相关蛋白质相互作用的第二区域。

[0329] 如本文所用,术语“gRNA”包括双分子gRNA以及单分子RNA。gRNA可以包含或不包含用于与CRISPR相关蛋白质相互作用的次要结构特征。

[0330] II型CRISPR-Cas9系统天然地使用双分子gRNA。这种双分子gRNA(“tracrRNA/crRNA”)一般包含crRNA(“CRISPR RNA”或“靶向RNA”或“crRNA”或“crRNA重复序列”)和相应的tracrRNA(“反式作用CRISPR RNA”或“激活RNA”或“tracrRNA”)分子。crRNA包含靶向区域(单链)和形成gRNA的Cas9结合区的dsRNA双链体的一半的一段核苷酸(“双链体形成区域”)。相应的tracrRNA包含形成gRNA的Cas9结合区的dsRNA双链体的另一半的一段核苷酸(双链体形成区)。换句话说,crRNA的一段核苷酸与tracrRNA的一段核苷酸互补并杂交,以



形成gRNA的Cas9结合区的dsRNA双链体。因此,可以说每个crRNA具有相应的tracrRNA。crRNA另外提供单链靶向区域。因此,crRNA和tracrRNA(作为相应一对)杂交以形成gRNA。

[0331] crRNA和tracrRNA还可以接合以提供(人工)单分子向导RNA(“单向导RNA”,“sgRNA”)。“sgRNA”一般包含crRNA其3'端通过“环”序列连接至tracrRNA的5'端(参见,例如,美国专利申请号US20140068797)。与crRNA类似,sgRNA包含与靶多核苷酸序列互补的靶向区域,一般与形成碱基对氢键(其形成二级结构)的第二区域相邻,一般为茎结构。“sgRNA”一般为约100个核苷酸的长度,然而,该术语还包括约17-18nt的截短单向导RNA(tru-sgRNA)(参见Fu,Y.et.al.Nat Biotechnol.2014 Mar;32(3):279-84)。该术语还包括去除了可抛弃特征的功能性微型sgRNA,其保留了位于sgRNA中与tracrRNA(非crRNA)相应的部分中的被称为“联系纽带”的必需和保守模块(参见美国专利申请号20140315985和Briner AE et al.Mol Cell.2014 Oct 23;56(2):333-9)。联系纽带位于II型CRISPR-Cas9系统的下部茎的下游紧邻处(即,位于3'方向)。术语“sgRNA”还包括“死RNA”(“dRNA”),其包含11-15个核苷酸的缩短化靶向区域。这种dRNA可用于招募催化活性Cas9内切核酸酶,以靶向DNA序列以改变基因表达而不引起DSB(参见Dahlman JE et al.Nat Biotechnol.2015 Nov;33(11):1159-61)。sgRNA衍生物也被该术语包括在内。这种衍生物一般包括赋予新的或另外的功能的其他部分或实体。具体地,添加到sgRNA四环和/或茎-环结构的MS2适体能够选择性地包含所述MS2结构域的效应蛋白招募到靶DNA(“sgRNA-MS2”)(参见Konermann S et al.Nature.2015 Jan 29;517(7536):583-588)。进一步的修饰在此是可想到的和并被设想。

[0332] tracrRNA/crRNA或sgRNA作为gRNA的使用不限于Cas9蛋白。任何其他CRISPR相关系统,优选II型CRISPR-Cas系统,可以与这种gRNA结合使用。然而,可能需要其他gRNA来确保其他CRISPR-Cas蛋白的功能,并且这种gRNA也被包括在对应的限定中。例如,V型CRISPR相关蛋白质,如Cpf1,由单和短(42-44nt)crRNA作为gRNA向导,一般包含直接重复序列中的单个茎环。

[0333] gRNA,如tracrRNA/crRNA、sgRNA或crRNA,可以通过任何合适的方式来提供,例如,本文在人工核酸分子的环境中所描述的裸形式或复合形式,例如利用脂质或(聚)阳离子载体,但一般被载体递送。合适的载体(如“定义”标题部分中限定)包括能够优选地广泛表达功能性gRNA(即能够将对应的CRISPR相关蛋白质招募至靶DNA序列)的任何核酸。因此,载体包括质粒和病毒载体,具体地慢病毒载体和腺相关病毒载体(AAV)。

[0334] RNA

[0335] 本发明的人工核酸分子可以优选为RNA。应当理解,术语“RNA”是指核糖核酸分子,其特征在于其核苷酸连接形成所述分子的具体顺序(即,其RNA序列)。因此,本领域技术人员在对应的环境中将容易理解,术语“RNA”可用于指代RNA分子或RNA序列。例如,在本发明的环境中使用的术语“RNA”优选是指RNA分子(所述分子的特征在于通过其具体RNA序列等)。在序列修饰的环境中,术语“RNA”将被理解为涉及修饰的RNA序列,但一般还包括所得的RNA分子(关于其RNA序列被修饰)。在优选的实施方式中,RNA可以是mRNA、病毒RNA或复制子RNA,优选mRNA。

[0336] 单-、双-或多顺反子RNA

[0337] 根据本发明的一些实施方式,人工核酸分子,优选RNA,可以是单-、双-或多顺反子

的,优选如本文定义。双-或多顺反子RNA一般包含两个(双顺反子)或更多个(多顺反子)开放阅读框(ORF)。在此环境中的开放阅读框是可翻译成肽或蛋白质的密码子序列。双-或多顺反子人工核酸分子(优选RNA)中的编码序列优选编码如本文所定义的不同蛋白质。双-或甚至多顺反子人工核酸分子,优选RNA,可以编码例如至少两个、三个、四个、五个、六个或更多个(优选不同的)本文限定的蛋白质(或其同源物、变体、片段或衍生物)。术语“编码两个或更多个蛋白质”可以意为(但不限于)双-或甚至多顺反子人工核酸分子,优选RNA,可以编码例如至少两个、三个、四个、五个、六个或更多个(优选不同的)蛋白质(或其同源物、变体、片段或衍生物)。

[0338] 在一些实施方式中,编码两个或更多个本文限定的CRISPR相关蛋白质或其同源物、变体、片段或衍生物的编码序列可以在双-或甚至多顺反子RNA中被至少一种IRES(内部核糖体进入位点)序列隔开。术语“IRES”(内部核糖体进入位点)是指允许翻译起始的RNA序列。IRES可以充当唯一的核糖体结合位点,但其也可以用于提供双-或甚至多顺反子的人工核酸分子,优选如上限定的RNA,其编码几个蛋白质(或其同源物、变体、片段或衍生物),该蛋白质彼此独立地被核糖体翻译。可根据本发明使用的IRES序列的实例是衍生自小核糖核酸病毒(例如FMDV)、瘟病毒(CFFV)、脊髓灰质炎病毒(PV)、脑炎心肌炎病毒(ECMV)、口蹄疫病毒(FMDV)、丙型肝炎病毒(HCV)、经典猪瘟病毒(CSFV)、小鼠白血病毒(MLV)、猿猴免疫缺陷病毒(SIV)或蟋蟀麻痹病毒(CrPV)的那些。

[0339] 根据进一步的实施方式,本发明的人工核酸分子(优选RNA)的至少一种编码序列可以编码连接或不连接氨基酸连接体序列的至少两个、三个、四个、五个、六个、七个、八个和更多个本文限定的CRISPR相关蛋白质(或其同源物、变体、片段或衍生物),其中所述连接体序列可包括刚性连接体、柔性连接体、可切割接头(例如,自切割肽)或其组合。示例性连接体在标题“衍生物”的部分中被描述。对应的公开内容适用于多个CRISPR相关蛋白质的连接,加以必要的修改。其中,本文限定的CRISPR相关蛋白质可以是相同或不同的或其组合。

[0340] 优选地,人工核酸分子,优选地RNA,包括如下长度:约50至约20000个、或100至约20000个核苷酸,优选约250至约20000个核苷酸,更优选约500至约10000个,甚至更优选约500至约5000个。

[0341] 本发明的人工核酸分子,优选地RNA,可进一步是单链或双链的。当作为双链RNA提供时,人工核酸分子优选包含正义链和相应的反义链。

#### [0342] 核酸修饰

[0343] 本发明的人工核酸分子(优选地RNA)或本文限定的任何其他核酸(例如载体)可以修饰核酸的形式提供。本发明的环境中设想的适当核酸修饰在下文描述。表述“本文限定的任何其他核酸”可以,但一般不,指代gRNA。

[0344] 根据优选的实施方式,如本文限定,本发明的至少一种人工核酸分子(优选地RNA(序列))(或本文限定的任何其他核酸,具体地RNA)被修饰。本文限定的修饰优选导致所述人工核酸分子(优选地RNA)稳定化。更优选,本发明因此提供“稳定化的”人工核酸分子,优选地RNA(或本文限定的任何其他核酸,具体地RNA)。

[0345] 根据优选的实施方式,本发明的人工核酸分子(优选地RNA)(或本文限定的任何其他核酸,具体地RNA)可因此作为“稳定化的”人工核酸分子,优选地RNA,具体地mRNA被提供,即基本上抵抗体内降解(例如通过外切或内切核酸酶)。

[0346] 这种稳定化可以例如通过人工核酸分子(优选地RNA)(或本文限定的任何其他核酸,具体地RNA)的修饰磷酸主链来实现。本发明相关的骨架修饰是这样的修饰:其中所述RNA(或本文限定的任何其他核酸,具体地RNA)中包含的核苷酸骨架的磷酸基被化学修饰。在这方面可以优选使用的核苷酸包含例如硫代磷酸基修饰的磷酸骨架,优选磷酸骨架中包含的至少一种磷酸氧被硫原子取代。稳定化的人工核酸分子(优选RNA)(或本文限定的其他核酸,具体地RNA)可以进一步包括例如:非离子磷酸基类似物,例如烷基和芳基膦酸基,其中带电荷的磷酸氧被烷基或芳基,或磷酸二酯和烷基磷酸三酯代替,其中带电的氧残基以烷基化形式存在。这种骨架修饰一般包括(但不意味着任何限制)下列修饰:甲基膦酸基,氨基磷酸基和硫代磷酸基(例如胞苷-5'-O-(1-硫代磷酸基))。

[0347] 在下文中,描述具体修饰,其优选地能够“稳定化”本发明的人工核酸分子,优选RNA,(或本文限定的任何其他核酸,具体地RNA)。

[0348] 化学修饰

[0349] 本文所用的术语“修饰”可指代化学修饰,包括骨架修饰以及糖修饰或碱基修饰。

[0350] 在此环境下,“修饰”的人工核酸分子,优选地RNA(或本文限定的任何其他核酸,具体地RNA),可包含核苷酸类似物/修饰(修饰的核苷酸或核苷),例如骨架修饰、糖修饰或碱基修饰。

[0351] 本发明相关的骨架修饰是这样的修饰:其中本文所述人工核酸分子(优选地RNA)(或本文限定的任何其他核酸,具体地RNA)中包含的核苷酸骨架的磷酸基被化学修饰。本发明相关的糖修饰是对人工核酸分子(优选地RNA)(或本文限定的任何其他核酸,具体地RNA)的核苷酸的糖的化学修饰。此外,本发明相关的碱基修饰是对人工核酸分子(优选地RNA)(或本文限定的任何其他核酸,具体地RNA)的核苷酸的碱基部分的化学修饰。在此环境下,核苷酸类似物或修饰优选选自适用于转录和/或翻译的核苷酸类似物。

[0352] 糖修饰:

[0353] (化学)修饰的核酸,具体地根据本发明的人工核酸分子,可包括糖修饰,即,其糖部分被修饰的核苷/核苷酸。

[0354] 例如,2'羟基(OH)可被修饰或被多种不同的“氧”或“脱氧”取代基替代。“氧”-2'羟基修饰的实例包括但不限于,烷氧基或芳氧基(-OR,例如,R=H、烷基、环烷基、芳基、芳烷基、杂芳基或糖);聚乙二醇(PEG)、-O(CH<sub>2</sub>CH<sub>2</sub>O)<sub>n</sub>CH<sub>2</sub>CH<sub>2</sub>OR;“锁”核酸(LNA),其中2'羟基例如亚甲基桥连接至相同核糖糖的4'碳;和氨基(-O-氨基,其中氨基,例如NRR,可以是烷基氨基、二烷基氨基、杂环基、芳基氨基、二芳基氨基、杂芳基氨基、或二杂芳基氨基、乙二胺、聚氨基)或氨基烷氧基。

[0355] “脱氧”修饰包括氢、氨基(例如NH<sub>2</sub>;烷基氨基、二烷基氨基、杂环基、芳基氨基、二芳基氨基、杂芳基氨基、二杂芳基氨基、或氨基酸);或氨基可通过连接体附接至糖,其中连接体包括原子C、N和O中的一种或多种。

[0356] 糖基团还可包含与核糖中的相应碳具有相反立体化学构型的一种或多种碳。因此,修饰人工核酸分子(优选地RNA)(或本文限定的任何其他核酸,具体地RNA)可包括包含例如阿拉伯糖作为糖的核苷酸。

[0357] 骨架修饰:

[0358] (化学)修饰的核酸,具体地根据本发明的人工核酸分子,可包括骨架修饰,即,其

磷酸骨架被修饰的核苷/核苷酸。

[0359] 骨架的磷酸基团可以通过用不同的取代基取代一种或多种氧原子来修饰。此外，修饰的核苷和核苷酸可包括用本文所述的修饰磷酸剂完全替代未修饰的磷酸基部分。

[0360] 修饰的磷酸基团的实例包括但不限于硫代磷酸基、磷酸硒酸基、硼酸磷酸基、硼酸磷酸酯、氢膦酸基、氨基磷酸基、烷基或芳基膦酸基和磷酸三酯。二硫代磷酸基的两个非连接氧均被硫替代。磷酸基连接体也可以通过用氮(桥连的氨基磷酸基)、硫(桥连的硫代磷酸基)和碳(桥连的亚甲基-膦酸基)替代连接氧来修饰。

[0361] 碱基修饰：

[0362] (化学)修饰的核酸，具体地根据本发明的人工核酸分子，可包括(核)碱基修饰，即，其核碱基部分被修饰的核苷/核苷酸。

[0363] 在RNA中发现的核碱基的实例包括但不限于腺嘌呤、鸟嘌呤、胞嘧啶和尿嘧啶。例如，本文所述的核苷和核苷酸可在主沟面(major groove face)上被化学修饰。在一些实施方式中，主沟(major groove)化学修饰可包括氨基、硫醇基、烷基或卤基。

[0364] 在一些实施方式中，核苷酸类似物/修饰选自如下碱基修饰，该碱基修饰优选地选自2-氨基-6-氯嘌呤核苷-5'-三磷酸、2-氨基嘌呤-核苷-5'-三磷酸；2-氨基腺苷-5'-三磷酸、2'-氨基-2'-脱氧胞苷-三磷酸、2-硫代胞苷-5'-三磷酸、2-硫代尿苷-5'-三磷酸、2'-氟胸苷-5'-三磷酸、2'-O-甲基-肌苷-5'-三磷酸、4-硫代尿苷-5'-三磷酸、5-氨基烯丙基胞苷-5'-三磷酸、5-氨基烯丙基尿苷-5'-三磷酸、5-溴胞苷-5'-三磷酸、5-溴尿苷-5'-三磷酸、5-溴-2'-脱氧胞苷-5'-三磷酸、5-溴-2'-脱氧尿苷-5'-三磷酸、5-碘胞苷-5'-三磷酸、5-碘-2'-脱氧胞苷-5'-三磷酸、5-碘尿苷-5'-三磷酸、5-碘-2'-脱氧尿苷-5'-三磷酸、5-甲基胞苷-5'-三磷酸、5-甲基尿苷-5'-三磷酸、5-丙炔基-2'-脱氧胞苷-5'-三磷酸、5-丙炔基-2'-脱氧尿苷-5'-三磷酸、6-氮杂胞苷-5'-三磷酸、6-氮杂尿苷-5'-三磷酸、6-氯嘌呤核苷-5'-三磷酸、7-脱氮杂腺苷-5'-三磷酸、7-脱氮杂鸟苷-5'-三磷酸、8-氮杂腺苷-5'-三磷酸、8-叠氮腺苷-5'-三磷酸、苯并咪唑-核苷-5'-三磷酸、N1-甲基腺苷-5'-三磷酸、N1-甲基鸟苷-5'-三磷酸、N6-甲基腺苷-5'-三磷酸、O6-甲基鸟苷-5'-三磷酸、假尿苷-5'-三磷酸、或嘌呤霉素-5'-三磷酸、黄苷-5'-三磷酸。特别优选选自下列碱基修饰核苷酸的核苷酸用于碱基修饰：5-甲基胞苷-5'-三磷酸、7-脱氮杂鸟苷-5'-三磷酸、5-溴胞苷-5'-三磷酸、和假尿苷-5'-三磷酸。

[0365] 在一些实施方式中，修饰核苷包括吡啶-4-酮核糖核苷、5-氮杂-尿苷、2-硫代-5-氮杂-尿苷、2-硫代尿苷、4-硫代-假尿苷、2-硫代-假尿苷、5-羟基尿苷、3-甲基尿苷、5-羧基甲基-尿苷、1-羧基甲基-假尿苷、5-丙炔基-尿苷、1-丙炔基-假尿苷、5-牛磺酸基甲基尿苷、1-牛磺酸基甲基-假尿苷、5-牛磺酸基甲基-2-硫代-尿苷、1-牛磺酸基甲基-4-硫代-尿苷、5-甲基-尿苷、1-甲基-假尿苷、4-硫代-1-甲基-假尿苷、2-硫代-1-甲基-假尿苷、1-甲基-1-脱氮杂-假尿苷、2-硫代-1-甲基-1-脱氮杂-假尿苷、二氢尿苷、二氢假尿苷、2-硫代-二氢尿苷、2-硫代-二氢假尿苷、2-甲氧基尿苷、2-甲氧基-4-硫代-尿苷、4-甲氧基-假尿苷、和4-甲氧基-2-硫代-假尿苷。

[0366] 在一些实施方式中，修饰核苷包括5-氮杂-胞苷、假异胞苷、3-甲基-胞苷、N4-乙酰基胞苷、5-甲酰基胞苷、N4-甲基胞苷、5-羟基甲基胞苷、1-甲基-假异胞苷、吡咯并-胞苷、吡咯并-假异胞苷、2-硫代-胞苷、2-硫代-5-甲基-胞苷、4-硫代-假异胞苷、4-硫代-1-甲基-假

异胞苷、4-硫代-1-甲基-1-脱氮杂-假异胞苷、1-甲基-1-脱氮杂-假异胞苷、泽布拉林(zebularine)、5-氮杂-泽布拉林、5-甲基-泽布拉林、5-氮杂-2-硫代-泽布拉林、2-硫代-泽布拉林、2-甲氧基-胞苷、2-甲氧基-5-甲基-胞苷、4-甲氧基-假异胞苷、和4-甲氧基-1-甲基-假异胞苷。

[0367] 在其他实施方式中,修饰核苷包括2-氨基嘌呤、2,6-二氨基嘌呤、7-脱氮杂-腺嘌呤、7-脱氮杂-8-氮杂-腺嘌呤、7-脱氮杂-2-氨基嘌呤、7-脱氮杂-8-氮杂-2-氨基嘌呤、7-脱氮杂-2,6-二氨基嘌呤、7-脱氮杂-8-氮杂-2,6-二氨基嘌呤、1-甲基腺苷、N6-甲基腺苷、N6-异戊烯基腺苷、N6-(顺式-羟基异戊烯基)腺苷、2-甲基硫-N6-(顺式-羟基异戊烯基)腺苷、N6-甘氨酸基氨甲酰基腺苷、N6-苏氨酸基氨甲酰基腺苷、2-甲基硫-N6-苏氨酸基氨甲酰基腺苷、N6,N6-二甲基腺苷、7-甲基腺嘌呤、2-甲基硫-腺嘌呤、和2-甲氧基-腺嘌呤。

[0368] 在其他实施方式中,修饰核苷包括肌苷、1-甲基-肌苷、丫苷、怀丁苷(wybutosine)、7-脱氮杂-鸟苷、7-脱氮杂-8-氮杂-鸟苷、6-硫代-鸟苷、6-硫代-7-脱氮杂-鸟苷、6-硫代-7-脱氮杂-8-氮杂-鸟苷、7-甲基-鸟苷、6-硫代-7-甲基-鸟苷、7-甲基肌苷、6-甲氧基-鸟苷、1-甲基鸟苷、N2-甲基鸟苷、N2,N2-二甲基鸟苷、8-氧代-鸟苷、7-甲基-8-氧代-鸟苷、1-甲基-6-硫代-鸟苷、N2-甲基-6-硫代-鸟苷、和N2,N2-二甲基-6-硫代-鸟苷。

[0369] 在一些实施方式中,核苷酸可在主沟面上被修饰,并且可包括用甲基或卤基替代尿嘧啶的C-5上的氢。在具体实施方式中,修饰核苷是5'-O-(1-硫代磷酸基)-腺苷、5'-O-(1-硫代磷酸基)-胞苷、5'-O-(1-硫代磷酸基)-鸟苷、5'-O-(1-硫代磷酸基)-尿苷或5'-O-(1-硫代磷酸基)-假尿苷。

[0370] 在一些实施方式中,本发明的修饰RNA(或本文限定的任何其他修饰核酸,具体地RNA)可包括选自下列的核苷修饰:6-氮杂-胞苷、2-硫代-胞苷、 $\alpha$ -硫代-胞苷、假-iso-胞苷、5-氨基烯丙基-尿苷、5-碘-尿苷、N1-甲基-假尿苷、5,6-二氢尿苷、 $\alpha$ -硫代-尿苷、4-硫代-尿苷、6-氮杂-尿苷、5-羟基-尿苷、脱氧-胸苷、5-甲基-尿苷、吡咯并-胞苷、肌苷、 $\alpha$ -硫代-鸟苷、6-甲基-鸟苷、5-甲基-胞苷、8-氧代-鸟苷、7-脱氮杂-鸟苷、N1-甲基-腺苷、2-氨基-6-氯-嘌呤、N6-甲基-2-氨基-嘌呤、假-异-胞苷、6-氯-嘌呤、N6-甲基-腺苷、 $\alpha$ -硫代-腺苷、8-叠氮-腺苷、7-脱氮杂-腺苷。

[0371] 在一些实施方式中,修饰人工核酸分子(优选地RNA)(或本文限定的任何其他核酸,具体地RNA)不包括任何本文所述的化学修饰。然而,这种修饰人工核酸可包括下文所述的脂质修饰或序列修饰。

[0372] 脂质修饰

[0373] 根据进一步实施方式,本发明的人工核酸分子(优选地RNA)(或本文限定的任何其他核酸,具体地RNA)包含至少一种脂质修饰。

[0374] 这种脂质修饰的本发明的人工核酸分子(优选地RNA)(或所述其他核酸,具体地RNA,如本文所述)一般包括(i)本文限定的人工核酸分子,优选地RNA(或所述核酸,具体地RNA),(ii)与所述人工核酸分子(优选地RNA)(或所述其他核酸,具体地RNA)共价连接的至少一种连接体,和(iii)与对应的连接体共价连接的至少一种脂质。

[0375] 可选地,脂质修饰的人工核酸分子(优选地RNA)(或本文限定的其他核酸)包括至少一种人工核酸分子(优选地RNA)(或所述其他核酸,具体地RNA)和与所述人工核酸分子(优选地RNA)(或所述其他核酸,具体地RNA)共价连接(无连接体)的至少一种(双功能性)脂

质。

[0376] 可选地,脂质修饰的人工核酸分子(优选地RNA)(或本文限定的任何其他核酸,具体地RNA)包括(i)人工核酸分子(优选地RNA)(或所述其他核酸,具体地RNA),(ii)与所述人工核酸分子(优选地RNA)(或所述其他核酸,具体地RNA)共价连接的至少一种连接体,和(iii)与对应的连接体共价连接的至少一种脂质,和(iv)与所述人工核酸分子(优选地RNA)(或所述其他核酸,具体地RNA)共价连接(无连接体)的至少一种(双功能性)脂质。

[0377] 在此环境下,特别优选脂质修饰存在于线性人工核酸分子(优选地RNA)(或本文限定的任何其他核酸)的末端。

[0378] 序列修饰

[0379] 根据优选的实施方式,本发明的人工核酸分子,优选地RNA,优选地mRNA,或本文限定的任何其他核酸,是“序列修饰”的,即包括如下文所述的至少一种序列修饰。在不希望束缚于具体理论的情况下,这种序列修饰可增加本发明的人工核酸分子(优选地RNA)的稳定性和/或增强表达。

[0380] G/C含量修饰

[0381] 根据优选的实施方式,本发明的人工核酸分子(优选地RNA,更优选mRNA)(或本文限定的任何其他核酸,具体地RNA)可通过改变其鸟苷/胞嘧啶(G/C)含量,优选地通过改变至少一种编码序列的G/C含量被修饰,因此稳定化。换句话说,本发明的人工核酸分子(优选地RNA)(或本文限定的任何其他核酸,具体地RNA)和优选地其序列可以是G/C修饰。

[0382] “G/C修饰”的核酸(优选地RNA)序列一般指代包括如下核酸(优选地RNA)序列的核酸(优选地RNA):基于修饰的野生型核酸(优选地RNA)序列并且包括与所述野生型核酸(优选地RNA)序列相比改变数量的鸟苷和/或胞嘧啶核苷酸。这种改变数量的G/C核苷酸可通过用包含鸟苷或胞嘧啶核苷酸的“同义”密码子取代包含腺苷或胸苷核苷酸的密码子而生成。因此,密码子取代优选地不改变被编码的氨基酸残基,而仅仅改变核酸(优选地RNA)的G/C含量。

[0383] 在本发明的特别优选的实施方式中,与对应野生型即无修饰核酸的编码序列的G/C含量相比,本发明的人工核酸分子(优选地RNA)(或本文限定的任何其他核酸,具体地RNA)的编码序列的G/C含量被改变,具体地增加。与对应的野生型核酸(优选地RNA)编码的氨基酸序列相比,本发明的人工核酸分子(优选地RNA)(或本文限定的任何其他核酸,具体地RNA)编码的氨基酸序列优选地不被修饰。

[0384] 本发明的人工核酸分子(优选地RNA)(或本文限定的任何其他核酸,具体地RNA)的这种修饰基于以下事实:将被翻译的任何RNA(或其他核酸)区域的序列对于所述RNA(或所述其他核酸)的有效翻译都是重要的。因此,RNA(或所述其他核酸)的组成和各种核苷酸的序列是重要的。具体地,G(鸟苷)/C(胞嘧啶)含量增加的序列比A(腺苷)/U(尿嘧啶)含量增加的序列更稳定。

[0385] 根据本发明,因此,本发明的人工核酸分子(优选地RNA)(或本文限定的任何其他核酸,具体地RNA)的密码子在保持翻译的氨基酸序列的同时与对应的野生型核酸(优选地RNA)(或所述其他核酸)相比是改变的,使得其包括增加量的G/C核苷酸。

[0386] 基于若干密码子编码一种和同种氨基酸(所谓遗传密码简并)的事实,最有利于稳定性的密码子可被确定(所谓替代性密码子使用)。根据本发明的人工核酸分子(优选地

RNA) (或本文限定的任何其他核酸,具体地RNA)所要编码的氨基酸,与其野生型序列相比,修饰其核酸序列存在多种可能性。在氨基酸被仅包含G或C核苷酸的密码子编码的情况下,密码子无修饰是必要的。

[0387] 因此,Pro (CCC或CCG)、Arg (CGC或CGG)、Ala (GCC或GCG) 和Gly (GGC或GGG) 的密码子需要无修饰,因为不存在A或U。相比之下,包含A和/或U核苷酸的密码子可通过编码相同氨基酸但不包含A和/或U的其他密码子的取代而被修饰。这些的实例是:Pro的密码子可从CCU或CCA被修饰成CCC或CCG;Arg的密码子可从CGU或CGA或AGA或AGG被修成CGC或CGG;Ala的密码子可从GCU或GCA被修饰成GCC或GCG;Gly的密码子可从GGU或GGA被修饰成GGC或GGG。在其他情况下,虽然不能从密码子消除A或U核苷酸,但可以通过使用包含较低含量的A和/或U核苷酸的密码子而减少A和U含量。这些的实例是:Phe的密码子可从UUU被修饰成UUC;Leu的密码子可从UUA、UUG、CUU或CUA被修饰成CUC或CUG;Ser的密码子可从UCU或UCA或AGU被修饰成UCC、UCG或AGC;Tyr的密码子可从UAU被修饰成UAC;Cys的密码子可从UGU被修饰成UGC;His的密码子可从CAU被修饰成CAC;Gln的密码子可从CAA被修饰成CAG;Ile的密码子可从AUU或AUA被修饰成AUC;Thr的密码子可从ACU或ACA被修饰成ACC或ACG;Asn的密码子可从AAU被修饰成AAC;Lys的密码子可从AAA被修饰成AAG;Val的密码子可从GUU或GUA被修饰成GUC或GUG;Asp的密码子可从GAU被修饰成GAC;Glu的密码子可从GAA被修饰成GAG;终止密码子UAA可被修饰成UAG或UGA。在Met (AUG) 和Trp (UGG) 的密码子的情况下,另一方面,序列修饰没有可能。上文列举的取代可单独或以所有可能的组合用于增加本发明人工核酸序列(优选地RNA序列)(或本文限定的任何其他核酸序列)的G/C含量——与其具体野生型核酸序列(即原始序列)相比。因此,例如,野生型序列中存在的Thr的所有密码子可被修饰成ACC(或ACG)。然而,优选地,例如,应用上述取代可能性的组合:

[0388] 原始序列(野生型RNA)中所有编码Thr的密码子取代成ACC(或ACG),和

[0389] 所有原始编码Ser的密码子取代成UCC(或UCG或AGC);原始序列中所有编码Ile的密码子的取代成AUC,和

[0390] 所有原始编码Lys的密码子取代成AAG,和

[0391] 所有原始编码Tyr的密码子取代成UAC;原始序列中所有编码Val的密码子取代成GUC(或GUG),和

[0392] 所有原始编码Glu的密码子取代成GAG,和

[0393] 所有原始编码Ala的密码子取代成GCC(或GCG),和

[0394] 所有原始编码Arg的密码子取代成CGC(或CGG);原始序列中所有编码Val的密码子取代成GUC(或GUG),和

[0395] 所有原始编码Glu的密码子取代成GAG,和

[0396] 所有原始编码Ala的密码子取代成GCC(或GCG),和

[0397] 所有原始编码Gly的密码子取代成GGC(或GGG),和

[0398] 所有原始编码Asn的密码子取代成AAC;原始序列中所有编码Val的密码子取代成GUC(或GUG),和

[0399] 所有原始编码Phe的密码子取代成UUC,和

[0400] 所有原始编码Cys的密码子取代成UGC,和

[0401] 所有原始编码Leu的密码子取代成CUG(或CUC),和

[0402] 所有原始编码Gln的密码子取代成CAG,和

[0403] 所有原始编码Pro的密码子取代成CCC(或CCG);等等。

[0404] 优选地,与编码本文限定的至少一种蛋白质的野生型核酸(优选地RNA)(或所述其他核酸,具体地RNA)的编码序列的G/C含量相比,本发明的人工核酸分子(优选地RNA)(或本文限定的任何其他核酸,具体地RNA)的编码序列的G/C含量增加至少7%,更优选至少15%,特别优选至少20%。

[0405] 根据优选的实施方式,编码本文限定的CRISPR相关蛋白质或同源物、变体、片段或衍生物的区域或野生型RNA序列的完整序列中的可取代密码子的至少5%、10%、20%、30%、40%、50%、60%,更优选至少70%,甚至更优选至少80%和最优选至少90%、95%或甚至100%被取代,从而增加所述序列的G/C含量。

[0406] 在此环境下,特别优选使本发明的人工核酸分子(优选地RNA)(或本文限定的任何其他核酸,具体地RNA),优选其至少一种编码序列的G/C含量与野生型序列相比增加至最大(即可取代密码子的100%)。

[0407] 本发明的人工核酸分子(优选地RNA)(或本文限定的任何其他核酸,具体地RNA)的进一步优选的修饰基于如下发现:翻译效率通过tRNA中细胞中的不同存在频率来确定。因此,如果所谓“稀有密码子”中本发明的人工核酸分子(优选地RNA)(或所述其他核酸,具体地RNA)中的存在程度增加,则相应修饰RNA(或所述其他核酸,具体地RNA)序列的翻译程度显著差于编码相对“频繁”的tRNA的密码子存在的情况。

[0408] 在一些优选的实施方式中,在本文限定的修饰人工核酸分子中(优选RNA)(或任何其他核酸),与野生型核酸(优选RNA)的相应区域相比,编码蛋白质的区域被修饰,使得编码细胞中相对稀有tRNA的野生型序列的至少一种密码子交换成编码细胞中相对频繁且与该相对稀有tRNA携带相同氨基酸的tRNA的密码子。

[0409] 由此,本发明的人工核酸分子(优选地RNA)(或本文限定的任何其他核酸,具体地RNA)的序列被修饰以使得频繁存在的tRNA可用的密码子被插入。换句话说,根据本发明,通过这种修饰,编码细胞中相对稀有tRNA的野生型序列的所有密码子都可以在每种情况下交换成编码细胞中相对频繁的tRNA并且在每种情况下与相对稀有tRNA携带相同氨基酸的密码子。哪种tRNA在细胞中相对频繁存在以及那种相比之下相对稀有存在是本领域技术人员已知的;参见例如Akashi, Curr. Opin. Genet. Dev. 2001, 11 (6): 660-666。对于具体氨基酸而言利用最频繁存在的tRNA的密码子,例如利用(人)细胞中最频繁存在的tRNA的Gly密码子是特别优选的。

[0410] 根据本发明的,特别优选将在本发明的修饰人工核酸分子(优选地RNA)(或本文限定的任何其他核酸,具体地RNA)中增加的(特别是最大化的)连续G/C含量与“频繁”的密码子连接,而不改变所述人工核酸分子(优选地RNA)的编码序列所编码的编码氨基酸序列。这种优选的实施方式能够提供特别有效翻译和稳定化的(修饰的)人工核酸分子(优选地RNA)(或本文限定的任何其他核酸)。

[0411] 上述修饰人工核酸分子(优选地RNA)(或本文限定的任何其他核酸)(G/C含量增加;tRNA交换)的确定可以利用WO 02/098443(其公开内容的全部范围被包括在本发明中)中说明的计算机程序进行。使用这种计算机程序,任何期望核酸(具体地RNA)的核苷酸序列可以借助遗传密码或其退化性质修饰,使得结合使用密码子产生最大G/C含量。在细胞中尽



可能频繁地发生tRNA的编码,由修饰的核酸编码的氨基酸序列,特别是RNA,与未修饰的序列相比优选不被修饰。

[0412] 可选地,还可以与原始序列相比仅修饰G/C含量或仅修饰密码子使用。Visual Basic 6.0 (使用的开发环境:Microsoft Visual Studio Enterprise 6.0 with Servicepack 3) 中的源代码还被描述于WO 02/098443。

[0413] A/U含量修饰

[0414] 在本发明的进一步优选的实施方式中,本发明的人工核酸分子(优选地RNA)(或本文限定的任何其他核酸,具体地RNA)的核糖体结合位点环境中的A/U含量与其对应的野生型核酸(优选地RNA)(或所述其他核酸,具体地RNA)的核糖体结合位点环境中的A/U含量相比增加。

[0415] 这种修饰(核糖体结合位点周围A/U含量增加)使核糖体结合至所述人工核酸分子(优选地RNA)(或本文限定的任何其他核酸,具体地RNA)的效率增加。核糖体与核糖体结合位点(Kozak序列)的有效结合进而具有有效翻译人工核酸分子(优选地RNA)(或本文限定的任何其他核酸,具体地RNA)的效果。

[0416] DSE修饰

[0417] 根据本发明的进一步实施方式,本发明的人工核酸分子(优选地RNA)(或本文限定的任何其他核酸,具体地RNA)可关于潜在去稳定序列元件被修饰。具体地,所述人工核酸分子(优选地RNA)(或所述其他核酸,具体地RNA)的编码序列和/或5'和/或3'非翻译区可与对应的野生型核酸(优选地RNA)(或所述其他野生型核酸)相比被修饰,使得其不包含去稳定序列元件,修饰人工核酸分子(优选地RNA)(或所述其他核酸,具体地RNA)的编码氨基酸序列优选与其对应的野生型核酸(优选地RNA)(或所述其他野生型核酸)相比不被修饰。

[0418] 已知例如在真核RNA的序列中,存在去稳定序列元件(DSE),其结合信号蛋白质并调控体内RNA酶促降解。为进一步稳定化修饰人工核酸分子(优选地RNA),任选地在编码本文限定的CRISPR相关蛋白质或本文限定的任何其他核酸的区域中,与野生型核酸(优选地RNA)的相应区域相比,一种或多种这种修饰可因此进行,使得其中不包含或基本上不包含去稳定序列元件。

[0419] 根据本发明的,非翻译区(3'-和/或5'-UTR)中存在的DSE也可通过这种修饰被从人工核酸分子(优选地RNA)(或本文限定的任何其他核酸,具体地RNA)消除。这种去稳定序列是例如富AU序列(AURES),其存在于多种不稳定RNA的3'-UTR部分中(Caput et al., Proc.Natl.Acad.Sci.USA 1986,83:1670-1674)。本发明的人工核酸分子(优选地RNA)(或本文限定的任何其他核酸,具体地RNA)因此优选与对应的野生型核酸(优选地RNA)(或所述对应的其他野生型核酸)相比被修饰,使得所述人工核酸分子(优选地RNA)(或所述其他核酸,具体地RNA)不包含这种去稳定序列。这也适用于被可能的核酸内切酶识别那些序列基序,例如序列GAACAAG——其被包含在编码转铁蛋白受体的基因的3'-UTR区段中(Binder et al., EMBO J.1994,13:1969-1980)。这些序列基序还优选从所述人工核酸分子(优选地RNA)(或本文限定的任何其他核酸,具体地RNA)被去除。

[0420] 适于密码子使用的序列:

[0421] 本发明的人工核酸分子(优选地RNA)(或本文限定的任何其他核酸,具体地RNA)的进一步优选的修饰基于如下发现:编码相同氨基酸的密码子一般以不同的频率存在。根据

进一步优选的实施方式,在修饰人工核酸分子(优选地RNA)(或所述其他核酸,具体地RNA)中,编码序列与对应的野生型核酸(优选地RNA)(或所述其他野生型核酸)的相应区域相比被修饰,使得编码相同氨基酸的密码子的频率相应于该密码子的天然存在频率——根据例如表4所示的人密码子使用。

[0422] 例如,在本发明的人工核酸分子(优选地RNA)(或本文限定的任何其他核酸,具体地RNA)的至少一种编码序列编码的氨基酸序列中存在的氨基酸丙氨酸(Ala)的情况下,野生型编码序列优选地以如下方式调整:密码子“GCC”使用频率为0.40,密码子“GCT”使用频率为0.28,密码子“GCA”使用频率为频率0.22,和密码子“GCG”使用频率为频率0.10,等。(参见表4)。

[0423] 表4:人密码子使用表

[0424]

氨基酸	密码子	分数	/1000	氨基酸	密码子	分数	/1000
Ala	GCG	0.10	7.4	Pro	CCG	0.11	6.9
Ala	GCA	0.22	15.8	Pro	CCA	0.27	16.9
Ala	GCT	0.28	18.5	Pro	CCT	0.29	17.5
Ala	GCC*	0.40	27.7	Pro	CCC*	0.33	19.8
Cys	TGT	0.42	10.6	Gln	CAG*	0.73	34.2
Cys	TGC*	0.58	12.6	Gln	CAA	0.27	12.3
Asp	GAT	0.44	21.8	Arg	AGG	0.22	12.0
Asp	GAC*	0.56	25.1	Arg	AGA*	0.21	12.1
Glu	GAG*	0.59	39.6	Arg	CGG	0.19	11.4
Glu	GAA	0.41	29.0	Arg	CGA	0.10	6.2

[0425]

Phe	TTT	0.43	17.6	Arg	CGT	0.09	4.5
Phe	TTC*	0.57	20.3	Arg	CGC	0.19	10.4
Gly	GGG	0.23	16.5	Ser	AGT	0.14	12.1
Gly	GGA	0.26	16.5	Ser	AGC*	0.25	19.5
Gly	GGT	0.18	10.8	Ser	TCG	0.06	4.4
Gly	GGC*	0.33	22.2	Ser	TCA	0.15	12.2
His	CAT	0.41	10.9	Ser	TCT	0.18	15.2
His	CAC*	0.59	15.1	Ser	TCC	0.23	17.7
Ile	ATA	0.14	7.5	Thr	ACG	0.12	6.1
Ile	ATT	0.35	16.0	Thr	ACA	0.27	15.1
Ile	ATC*	0.52	20.8	Thr	ACT	0.23	13.1
Lys	AAG*	0.60	31.9	Thr	ACC*	0.38	18.9
Lys	AAA	0.40	24.4	Val	GTG*	0.48	28.1
Leu	TTG	0.12	12.9	Val	GTA	0.10	7.1
Leu	TTA	0.06	7.7	Val	GTT	0.17	11.0
Leu	CTG*	0.43	39.6	Val	GTC	0.25	14.5
Leu	CTA	0.07	7.2	Trp	TGG*	1	13.2
Leu	CTT	0.12	13.2	Tyr	TAT	0.42	12.2
Leu	CTC	0.20	19.6	Tyr	TAC*	0.58	15.3
Met	ATG*	1	22.0	Stop	TGA*	0.61	1.6
Asn	AAT	0.44	17.0	Stop	TAG	0.17	0.8
Asn	AAC*	0.56	19.1	Stop	TAA	0.22	1.0

[0426] \*:最频繁的密码子

[0427] 密码子-优化序列:

[0428] 如上所述,根据本发明优选所有编码在细胞中相对稀有的tRNA的野生型序列的密码子交换成编码在细胞相对频繁并且在每种情况下与相对稀有tRNA携带相同氨基酸的tRNA的密码子。

[0429] 因此,特别优选最频繁的密码子用于编码氨基酸(参见表4,最频繁的密码子用星号标记)。这种优化程序增加密码子适应指数(CAI)并且最终最大化CAI。在本发明的环境中,CAI增加或最大化的序列一般被称为“密码子优化”序列和/或CAI增加和/或最大化序列s。根据优选的实施方式,本发明的人工核酸分子(优选地RNA)(或本文限定的任何其他核酸,具体地RNA)包括至少一种编码序列,其中编码序列本文所述被密码子优化。更优选,至少一种编码序列的密码子适应指数(CAI)是至少0.5、至少0.8、至少0.9或至少0.95。最优选,至少一种编码序列的密码子适应指数(CAI)是1。

[0430] 例如,在本发明的人工核酸分子(优选地RNA)(或本文限定的任何其他核酸,具体地RNA)的至少一种编码序列编码的氨基酸序列中存在的氨基酸丙氨酸(Ala)的情况下,野生型编码序列以如下方式调整:最频繁的人密码子“GCC”始终用于所述氨基酸、或用于氨基酸半胱氨酸(Cys);野生型序列以如下方式调整:最频繁的人密码子“TGC”始终用于所述氨基酸等。

[0431] C-优化序列:

[0432] 根据优选的实施方式,本发明的人工核酸分子(优选地RNA)(或本文限定的任何其

他核酸,具体地RNA)通过如下被修饰:改变,优选地增加,所述人工核酸分子(优选地RNA)(或所述其他核酸,具体地RNA)的胞嘧啶(C)含量,具体地在其至少一种编码序列中。

[0433] 在优选的实施方式中,本发明的人工核酸分子(优选地RNA)(或本文限定的任何其他核酸,具体地RNA)的编码序列的C含量与对应野生型(未被修饰)核酸的编码序列的C含量相比改变,优选地增加。本发明的人工核酸分子(优选地RNA)的至少一种编码序列编码的氨基酸序列优选与对应野生型核酸(优选地RNA)(或对应其他野生型核酸)编码的氨基酸序列相比不被修饰。

[0434] 在优选的实施方式中,所述修饰人工核酸分子(优选地RNA)(或本文限定的任何其他核酸,具体地RNA)被修饰使得理论上可能的最大胞嘧啶含量或甚至最大胞嘧啶含量的至少10%、20%、30%、40%、50%、60%、70%或80%、或至少90%实现。

[0435] 在进一步优选的实施方式中,“胞嘧啶含量可优化性”野生型核酸(优选地RNA)序列的密码子的至少10%、20%、30%、40%、50%、60%、70%、80%、90%或甚至100%被胞嘧啶含量高于野生型序列中存在的胞嘧啶含量的密码子替代。

[0436] 在进一步优选的实施方式中,野生型编码序列的一些密码子可另外被修饰使得细胞中相对稀有tRNA的密码子交换成细胞中相对频繁tRNA的密码子,条件是取代的相对频繁tRNA密码子与原始野生型密码子的相对稀有tRNA携带相同的氨基酸。优选地,所有相对稀有tRNA的密码子均被细胞中相对频繁tRNA的密码子替代——除了编码仅由不包含任何胞嘧啶的密码子编码的氨基酸的密码子、或除了谷氨酰胺(Gln),其被两种密码子编码,每种密码子包含相同数量的胞嘧啶。

[0437] 在本发明的进一步优选的实施方式中,修饰人工核酸分子(优选地RNA)(或本文限定的任何其他核酸,具体地RNA)被修饰,使得通过编码细胞中相对频繁的tRNA密码子实现理论上可能的最大胞嘧啶含量或甚至最大胞嘧啶含量的至少80%或至少90%,其中氨基酸序列保持不变。

[0438] 由于天然存在的遗传密码简并,多于一种密码子可编码具体的氨基酸。因此,20种天然存在的氨基酸中的18种被多于一种密码子编码(Tryp和Met除外),例如2种密码子(例如Cys、Asp、Glu),三种密码子(例如Ile),4种密码子(例如Al、Gly、Pro),或6种密码子(例如Leu、Arg、Ser)。然而,不是所有编码相同氨基酸的密码子都在体内条件下以相同频率被利用。根据每一种生物体,建立一般密码子使用配置(usage profile)。

[0439] 本发明的环境中使用的术语‘胞嘧啶含量可优化性密码子’指代呈现胞嘧啶含量低于编码相同氨基酸的其他密码子的密码子。因此,可被编码相同氨基酸的其他密码子替代并在该密码子中呈现更高胞嘧啶数量的任何野生型密码子被认为是胞嘧啶可优化性(C可优化性)。野生型编码序列内任何这种C可优化性野生型密码子被特定C优化密码子的取代增加了其总体C含量,并体现了富C修饰的RNA序列。

[0440] 根据一些优选的实施方式,本发明的人工核酸分子(优选地RNA)(或本文限定的任何其他核酸,具体地RNA)和具体地其至少一种编码序列包括下列或由下列组成:包含所有潜在C可优化性密码子的C优化密码子的C最大化序列。因此,在编码序列全长上理论上可替代的C可优化性密码子的100%或全部优选被C优化密码子替代。

[0441] 在此环境下,胞嘧啶含量可优化性密码子是包含胞嘧啶数量低于编码相同氨基酸的其他密码子的密码子。

[0442] 编码氨基酸Ala的密码子GCG、GCA、GCU中的任一者可交换成编码相同氨基酸的密码子GCC,和/或

[0443] 编码Cys的密码子UGU可交换成编码相同氨基酸的密码子UGC,和/或

[0444] 编码Asp的密码子GAU可交换成编码相同氨基酸的密码子GAC,和/或

[0445] 编码Phe的密码子UUU可交换成编码相同氨基酸的密码子UUC,和/或

[0446] 编码Gly的密码子GGG、GGA、GGU中的任一者可交换成密码子编码相同氨基酸的GGC,和/或

[0447] 编码His的密码子CAU可交换成编码相同氨基酸的密码子CAC,和/或

[0448] 编码Ile的密码子AUA、AUU中的任一者可交换成密码子AUC,和/或

[0449] 编码Leu的密码子UUG、UUA、CUG、CUA、CUU中的任一者可交换成编码相同氨基酸的密码子CUC,和/或

[0450] 编码Asn的密码子AAU可交换成编码相同氨基酸的密码子AAC,和/或

[0451] 编码Pro的密码子CCG、CCA、CCU中的任一者可交换成编码相同氨基酸的密码子CCC,和/或

[0452] 编码Arg的密码子AGG、AGA、CGG、CGA、CGU中的任一者可交换成编码相同氨基酸的密码子CGC,和/或

[0453] 编码Ser的密码子AGU、AGC、UCG、UCA、UCU中的任一者可交换成编码相同氨基酸的密码子UCC,和/或

[0454] 编码Thr的密码子ACG、ACA、ACU中的任一者可交换成编码相同氨基酸的密码子ACC,和/或

[0455] 编码Val的密码子GUG、GUA、GUU中的任一者可交换成编码相同氨基酸的密码子GUC,和/或

[0456] 编码Tyr的密码子UAU可交换成编码相同氨基酸的密码子UAC。

[0457] 在任何上述实例中,每个交换密码子的胞嘧啶数量增加1个。编码序列的所有非C优化密码子(相应于C-可优化性密码子)的交换产生C最大化编码序列。在本发明的环境中,本发明的人工核酸分子(优选地RNA)(或本文限定的任何其他核酸,具体地RNA)的至少一种编码序列中的非C优化密码子的至少70%,优选至少80%,更优选至少90%被C优化密码子替代。

[0458] 可优选,对于一些氨基酸,被C优化密码子替代的C可优化性密码子的百分比小于70%,而对于其他氨基酸,替代密码子的百分比高于70%,以满足C优化总体百分比为编码序列的所有C可优化性野生型密码子的至少70%。

[0459] 优选地,在C优化人工核酸分子(优选地RNA)(或本文限定的任何其他核酸,具体地RNA)中,任何给定氨基酸的C可优化性野生型密码子的至少50%被C优化密码子替代,例如任何修饰富C RNA(或其他核酸,具体地RNA)优选在编码上述氨基酸Ala、Cys、Asp、Phe、Gly、His、Ile、Leu、Asn、Pro、Arg、Ser、Thr、Val和Tyr中任一者的C可优化性野生型密码子位置处包含至少50%C优化密码子,优选至少60%。

[0460] 在此环境中,可不经任何进一步选择过程使用编码不是胞嘧啶含量可优化的但被至少两种密码子编码的氨基酸的密码子。然而,编码细胞(例如人细胞)中相对稀有tRNA的野生型序列的密码子可交换成编码细胞中相对频繁tRNA的密码子,其中二者编码相同的氨

氨酸。因此,编码Glu的相对稀有密码子GAA可交换成编码相同氨基酸的相对频繁密码子GAG,和/或

[0461] 编码Lys的相对稀有密码子AAA可交换成编码相同氨基酸的相对频繁密码子AAG,和/或

[0462] 编码Gln的相对稀有密码子CAA可交换成编码相同氨基酸的相对频繁密码子CAG。

[0463] 在此环境下,各仅被一种密码子编码的氨基酸Met (AUG) 和Trp (UGG) 保持不变。终止密码子不被胞嘧啶含量优化,然而,相对稀有的终止密码子琥珀、赭石 (UAA,UAG) 可交换成相对频繁的终止密码子蛋白石 (UGA)。

[0464] 上文列举的单一取代可单独以及以所有可能的组合使用,以与野生型序列相比优化修饰人工核酸分子(优选地RNA)的胞嘧啶含量。

[0465] 因此,本文限定的至少一种编码序列可与对应野生型核酸(优选地RNA)的编码序列相比以如下方式改变:氨基酸被至少两种或更多种密码子(其中一种包括一个另外的胞嘧啶)编码,这种密码子可交换成包括一个另外的胞嘧啶的C优化密码子,其中氨基酸优选与野生型序列相比不改变。

[0466] 根据特别优选的实施方式,本发明的组合包括人工核酸分子(优选地RNA),其包括(除本文指定的5' UTR和3' UTR外)本文限定的至少一种编码序列,其中(a)所述人工核酸分子(优选地RNA)的至少一种编码序列的G/C含量与相应野生型核酸(优选地RNA)的相应编码序列的G/C含量相比增加,和/或(b)其中所述人工核酸分子(优选地RNA)的至少一种编码序列的C含量与相应野生型核酸(优选地RNA)的相应编码序列的C含量相比增加,和/或(c)其中所述人工核酸分子(优选地RNA)的至少一种编码序列中的密码子适于密码子使用,其中所述人工核酸分子(优选地RNA)的至少一种编码序列中的密码子适应指数(CAI)优选地增加或最大化,并且其中所述人工核酸分子(优选地RNA)编码的氨基酸序列优选与相应野生型核酸(优选地RNA)编码的氨基酸序列相比不被修饰。

[0467] 修饰核酸序列

[0468] 上述序列修饰可总体上适用于本文所述的任何核酸序列,并且具体地设想其适用于包括下列或由下列组成编码序列:编码本文限定的CRISPR相关蛋白质和任选地NLS或其他肽或蛋白质部分、结构域或标签的核酸序列。如适当或必要,修饰(包括化学修饰、脂质修饰和序列修饰)可彼此以任何组合进行组合,条件是组合的修饰不彼此干扰,并且优选地条件是编码的CRISPR相关蛋白质(和NLS、信号序列、蛋白质/肽标签)优选地保留其期望的生物学功能或性质,如上限定。

[0469] 在优选的实施方式中,人工核酸根据本发明包括编码CRISPR相关蛋白质的编码序列,其中所述编码序列已如上所述被修饰。

[0470] 在一些优选的实施方式中,人工核酸根据本发明包括编码Cas9蛋白或其同源物、变体、片段或衍生物的编码序列,其中所述编码序列包括下列或由下列组成:根据SEQ ID NO:412;3474-3887;2314-2327;4634-4647;5794-5807;6954-6967;8114-8127;413-425;3490-3503;3506-3519;3522-3535;3538-3551;3554-3567;3570-3583;3586-3599;3602-3615;3618-3631;3634-3647;3650-3663;3666-3679;3682-3695;9514-9527;9626-9639;9738-9751;9850-9863;9962-9975,10074-10087;10186-10199;10298-10311;2330-2343;2346-2359;2362-2375;2378-2391;2394-2407;2410-2423;2426-2439;2442-2455;2458-

2471;2474-2487;2490-2503;2506-2519;2522-2535;9498-9511;9610-9623;9722-9735;9834-9847;9946-9959;10058-10071;10170-10183-10282-10295;4650-4663;4666-4679;4682-4695;4698-4711;4714-4727;4730-4743;4746-4759;4762-4775;4778-4791;4794-4807;4810-4823;4826-4839;4842-4855;9530-9543;9642-9655;9754-9767;9866-9879;9978-9991;10090-10103;10202-10215;10314-10327;5810-5823;5826-5839;5842-5855;5858-5871;5874-5887;5890-5903;5906-5919;5922-5935;5938-5951;5954-5967;5970-5983;5986-5999;6002-6015;9546-9559;9658-9671;9770-9783;9882-9895;9994-10007;10106-10119;10218-10231;10330-10343;6970-6983;6986-6999;7002-7015;7018-7031;7034-7047;7050-7063;7066-7079;7082-7095;7098-7111;7114-7127;7130-7143;7146-7159;7162-7175;9562-9575;9674-9687;9786-9799;9898-9911;10010-10023;10122-10135;10234-10247;10346-10359;8130-8143;8146-8159;8162-8175;8178-8191;8194-8207;8210-8223;8226-8239;8242-8255;8258-8271;8274-8287;8290-8302;8306-8319;8322-8335;9578-9591;9690-9703;9802-9815;9914-9927;10026-10039;10138-10151;10250-10263;10362-10375;9290-9303;9306-9319;9322-9335;9338-9351;9354-9367;9370-9383;9386-9399;9402-9415;9418-9431;9434-9447;9450-9463;9466-9479;9482-9495;9594-9607;9706-9719;9818-9831;9930-9943;10042-10055;10154-10167;10266-10279;10378-10391的核酸序列或其同源物、变体或片段,具体地按递增优选顺序与这些序列中的任意者具有至少5%、10%、20%、30%、40%、50%、60%、70%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、或99%,优选至少70%,更优选至少80%,甚至更优选至少85%,甚至更优选至少90%,并且最优选至少95%或甚至97%序列同一性的核酸序列。

[0471] 在一些优选的实施方式中,人工核酸根据本发明包括编码Cpf1蛋白或其同源物、变体、片段或衍生物的编码序列,其中所述编码序列包括下列或由下列组成:根据SEQ ID NO:10552;3458-3459;3460-3473;2298-2299;4618-4619;5778-5779;6938-6939;8098-8099;9258-9259;2300-2313;4620-4633;5780-5793;6940-6953;8100-8113;9260-9273;3488-3489;10396;2328-2329;10395;4648-4649;10397;5808-5809;10398;6968-6969;10399;8128-8129;10400;9274-9287;3504-3505;3520-3521;3536-3537;3552-3553;3568-3669;3584-3585;3600-3601;3616-3617;3632-3633;3648-3649;3664-3665;3680-3681;3696-3697;9528-9529;9640-9641;9752-9753;9864-9865;9976-9977;10088-10089;10200-10201;10312-10313;10403;10410;10417;10424;10431;10438;10445;10452;10459;10466;10473;10480;10487;10494;10501;10508;10515;10522;10529;10536;10543

[0472] 2344-2345;2360-2361;2376-2377;2392-2393;2408-2409;2424-2425;2440-2441;2456-2457;2472-2473;2489-2490;2504-2505;2520-2521;2536-2537;9512-9513;9624-9625;9736-9737;9848-9849;9960-9961;10072-10073;10184-10185;10296-10297;10402;10409;10416;10423;10430;10437;10444;10451;10458;10465;10472;10479;10486;10493;10500;10507;10514;10521;10528;10535;10542;4664-4665;4680-4681;4696-4697;4712-4713;4728-4729;4744-4745;4760-4761;4776-4777;4792-4793;4808-4809;4824-4825;4840-4841;4856-4857;9544-9545;9656-9657;9768-9769;9880-9881;9992-9993;10104-10105;10216-10217;10328-10329;10404;10411;10418;10425;10432;

10439;10446;10453;10460;10467;10474;10481;10488;10495;10502;10509;10516;  
10523;10530;10537;10544;5824-5825;5840-5841;5856-5857;5872-5873;5888-5889;  
5904-5905;5920-5921;5936-5937;5952-5953;5968-5969;5984-5985;6000-6001;6016-  
6017;9560-9561;9672-9673;9784-9785;9896-9897;10008-10009;10120-10121;10232-  
10233;10344-10345;10405;10412;10419;10426;10433;10440;10447;10454;10461;  
10468;10475;10482;10489;10496;10503;10510;10517;10524;10531;10538;10545;7033;  
7048-7049;7064-7065;7080-7081;7096-7097;7112-7113;7128-7129;7144-7145;7160-  
7161;7176-7177;9576-9577;9688-9689;9800-9801;9912-9913;10024-10025;10136-  
10137;10248-10249;10360-10361;10406;10413;10420;10427;10434;10441;10448;  
10455;10462;10469;10476;10483;10490;10497;10504;10511;10518;10525;10532;  
10539;10546;8144-8145;8160-8160;8176-8177;8192-8193;8208-8209;8224-8225;8240-  
8241;8256-8257;8272-8273;8288-8289;8304-8305;8320-8321;8336-8337;9592-9593;  
9704-9705;9816-9817;9928-9929;10040-10041;10152-10153;10264-10265;10376-  
10377;10407;10414;10421;10428;10435;10442;10449;10456;10463;10470;10477;  
10484;10491;10498;10505;10512;10519;10526;10533;10540;10547;9288-9289;10401;  
10553;10582-10583;10579-10580;10585-10586;10588-10589;10591-10592;10594-10595;  
10597-10598;10554-10574;10601;10602;10615;10616;10629;10630;10643;10644;  
10657;10658;10671;10672;10685;10686;10699;10700;10713;10714;10727;10728;  
10741;10742;10755;10756;10769;10770;10783;10784;10797;10798;10811;10812;  
10825;10826;10839;10840;10853;10854;10867;10868;10881;10882;10603;10604;  
10617;10618;10631;10632;10645;10646;10659;10660;10673;10674;10687;10688;  
10701;10702;10715;10716;10729;10730;10743;10744;10757;10758;10771;10772;  
10785;10786;10799;10800;10813;10814;10827;10828;10841;10842;10855;10856;  
10869;10870;10883;10884;10605;10606;10619;10620;10633;10634;10647;10648;  
10661;10662;10675;10676;10689;10690;10703;10704;10717;10718;10731;10732;  
10745;10746;10759;10760;10773;10774;10787;10788;10801;10802;10815;10816;  
10829;10830;10843;10844;10857;10858;10871;10872;10885;10886;10607;10608;  
10621;10622;10635;10636;10649;10650;10663;10664;10677;10678;10691;10692;  
10705;10706;10719;10720;10733;10734;10747;10748;10761;10762;10775;10776;  
10789;10790;10803;10804;10817;10818;10831;10832;10845;10846;10859;10860;  
10873;10874;10887;10888;10609;10610;10623;10624;10637;10638;10651;10652;  
10665;10666;10679;10680;10693;10694;10707;10708;10721;10722;10735;10736;  
10749;10750;10763;10764;10777;10778;10791;10792;10805;10806;10819;10820;  
10833;10834;10847;10848;10861;10862;10875;10876;10889;10890;10611;10612;  
10625;10626;10639;10640;10653;10654;10667;10668;10681;10682;10695;10696;  
10709;10710;10723;10724;10737;10738;10751;10752;10765;10766;10779;10780;  
10793;10794;10807;10808;10821;10822;10835;10836;10849;10850;10863;10864;  
10877;10878;10891;10892;9304-9305;9320-9321;9336-9337;9352-9353;9368-9369;  
9384-9385;9400-9401;9416-9417;9432-9433;9448-9449;9464-9465;9480-9481;9496-



9497;9608-9609;9720-9721;9832-9833;9944-9945;10056-10057;10168-10169;10280-10281;10392-10393;10408;10415;10422;10429;10436;10443;10450;10457;10464;10471;10478;10485;10492;10499;10506;10513;10520;10527;10534;10541;10548的核酸序列或其同源物、变体或片段,具体地按递增优选顺序与这些序列中任意者具有至少5%、10%、20%、30%、40%、50%、60%、70%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、或99%,优选至少70%,更优选至少80%,甚至更优选至少85%,甚至更优选至少90%,并且最优选至少95%或甚至97%序列同一性的核酸序列。

[0473] 根据本发明的人工核酸可在相同(单顺反子核酸)或不同(多顺反子核酸)编码区(一种或多种)中编码其他蛋白质或肽。对应的编码核酸序列可进行与上述相同的序列修饰。具体地,本发明人工核酸的编码序列可包括编码一种或多种核定位信号(NLS)的一种或多种序列,该核定位信号优选融合至编码CRISPR相关蛋白质的核酸序列。那些序列也可如上所述被修饰。

[0474] 修饰(或“优化”)编码序列可与本文公开的任何UTR组合。

[0475] 因此,在一些优选的实施方式中,人工核酸根据本发明包括本文限定的至少一个5' UTR元件、本文限定的至少一个3' UTR元件和编码Cas9或Cpf1蛋白或其(功能性)同源物、变体、片段或衍生物的编码序列,其中所述人工核酸分子包括下列或由下列组成:根据SEQ ID NO:413;2330-2345;3490-3505;4650-4665;5810-5825;6970-6985;8130-8145;9290-9305;10402-10408;10554;10599-10612(HSD17B4/Gnas.1);SEQ ID NO:414;2346-2361;3506-3521;4666-4681;5826-5841;6986-7001;8146-8161;9306-9321;10409-10415;10555;10613-10626(Slc7a3.1/Gnas.1);SEQ ID NO:415;2362-2377;3522-3537;4682-4697;5842-5857;7002-7017;8162-8177;9322-9337;10416-10422;10556;10627-10640(ATP5A1/CASP.1);SEQ ID NO:416;2378-2393;3538-3553;4698-4713;5858-5873;7018-7033;8178-8193;9338-9353;10423-10429;10557;10641-10654(Ndufa4.1/PSMB3.1);SEQ ID NO:417;2394-2409;3554-3569;4714-4729;5874-5889;7034-7049;8194-8209;9354-9369;10430-10436;10558;10655-10668(HSD17B4/PSMB3.1);SEQ ID NO:418;2410-2425;3570-3585;4730-4745;5890-5905;7050-7065;8210-8225;9370-9385;10437-10443;10559;10669-10682(RPL32var/白蛋白7);SEQ ID NO:419;2426-2441;3586-3601;4746-4761;5906-5921;7066-7081;8226-8241;9386-9401;10444-10450;10560;10683-10696(32L4/白蛋白7);SEQ ID NO:420;2442-2457;3602-3617;4762-4777;5922-5937;7082-7097;8242-8257;9402-9417;10451-10457;10561;10697-10710(HSD17B4/CASP1.1);SEQ ID NO:421;2458-2473;3618-3633;4778-4793;5938-5953;7098-7113;8258-8273;9418-9433;10458-10464;10562;10711-10724(Slc7a3.1/CASP1.1);SEQ ID NO:422;2474-2489;3634-3649;4794-4809;5954-5969;7114-7129;8274-8289;9434-9449;10465-10471;10563;10725-10738(Slc7a3.1/PSMB3.1);SEQ ID NO:423;2490-2505;3650-3665;4810-4825;5970-5985;7130-7145;8290-8305;9459-9450;10472-10478;10564;10739-10752(Nosip.1/PSMB3.1);SEQ ID NO:424;2506-2521;3666-3681;4826-4841;5986-6001;7146-7161;8306-8321;9466-9481;10479-10485;10565;10753-10766(Ndufa4.1/RPS9.1);SEQ ID NO:425;2522-2537;3682-3697;4842-4857;6002-6017;7162-7177;8322-8337;9482-

9497;10486-10492;10566;10767-10780 (HSD17B4/RPS9.1);SEQ ID NO:9498-9609;10493-10499;10567;10781-10794 (ATP5A1/Gnas.1);SEQ ID NO:9610-9721;10500-10506;10568;10795-10808 (Ndufa4.1/COX6B1.1);SEQ ID NO:9722-9833;10507-10513;10569;10809-10822 (Ndufa4.1/Gnas.1);SEQ ID NO:9834-9945;10514-10520;10570;10823-10836 (Ndufa4.1/Ndufa1.1);SEQ ID NO:9946-10057;10521-10527;10571;10837-10850 (Nosip.1/Ndufa1.1);SEQ ID NO:10058-10169;10528-10534;10572;10851-10864 (Rpl31.1/Gnas.1);SEQ ID NO:10170-10281;10535-10541;10573;10865-10878 (TUBB4B.1/RPS9.1);SEQ ID NO:10282-10393;10542-10548;10574;10879-10892 (Ubqln2.1/RPS9.1)的核酸序列或所述序列中任一者的同源物、变体或片段,具体地按递增优选顺序与这些序列中的任意者具有至少5%、10%、20%、30%、40%、50%、60%、70%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、或99%,优选至少70%,更优选至少80%,甚至更优选至少85%,甚至更优选至少90%,并且最优选至少95%或甚至97%序列同一性的核酸序列。通过序列表的<223>的“其他信息”行(在此公开了对应的SEQ ID NO相像Cpf1变体mRNA产物还是Cas9变体mRNA产物)给出的指导,哪个SEQ ID NO属于那种蛋白质(Cas9或Cpf1)对于技术人员是显而易见的。

#### [0476] 5'帽

[0477] 根据本发明的进一步优选的实施方式,本文限定的修饰人工核酸分子(优选地RNA)(或任何其他核酸,具体地RNA)可通过添加所谓的‘5'帽’结构被修饰,该‘5'帽’优选稳定化所述人工核酸分子(优选地RNA)(或所述其他核酸,具体地RNA),如本文所述。

[0478] 5'-帽是实体,一般是修饰核苷酸实体,其总体上使成熟mRNA的5'端“加帽”。5'-帽可一般由修饰核苷酸形成,具体地由鸟嘌呤核苷酸衍生物形成。优选地,5'-帽通过5'-5'-三磷酸连接而连接至5'端。5'-帽可甲基化,例如m7GpppN,其中N是携带5'-帽的核酸的末端5'核苷酸,一般mRNA的5'端。m7GpppN是5'-帽结构,其天然存在于通过聚合酶II转录的mRNA中并且因此在此环境下优选不被认为是修饰mRNA中包括的修饰。因此,“修饰”人工核酸分子(优选地RNA)(或本文限定的任何其他核酸,具体地RNA)可包括m7GpppN作为5'-帽,但此外所述修饰人工核酸分子(优选地RNA)(或其他核酸)一般包括至少一种本文限定的进一步修饰。

[0479] 5'帽结构的进一步实例包括甘油基、倒置脱氧脱碱基残基(部分)、4',5'亚甲基核苷酸、1-(β-D-赤呋喃糖基)核苷酸、4'-硫代核苷酸、碳环核苷酸、1,5-失水己糖醇核苷酸、L-核苷酸、α-核苷酸、修饰碱基核苷酸、苏-戊呋喃糖基核苷酸、无环3',4'-seco核苷酸、无环3,4-二羟基丁基核苷酸、无环3,5二羟基戊基核苷酸、3'-3'-倒置核苷酸部分、3'-3'-倒置脱碱基部分、3'-2'-倒置核苷酸部分、3'-2'-倒置脱碱基部分、1,4-丁二醇磷酸基、3'-氨基磷酸基、己基磷酸基、氨基己基磷酸基、3'-磷酸基、3'硫代磷酸基、二硫代磷酸基、或桥连或非桥连甲基磷酸基部分。这些修饰5'-帽结构在此环境下被认为是至少一种修饰。

[0480] 特别优选的修饰5'-帽结构是帽1(m7G相邻核苷酸的核糖甲基化)、帽2(m7G下游第2个核苷酸的核糖另外甲基化)、帽3(m7G下游第3个核苷酸的核糖另外甲基化)、帽4(m7G下游第4个核苷酸的核糖甲基化)、ARCA(防逆转帽类似物,修饰ARCA(例如硫代磷酸基修饰的ARCA)、CleanCap(TriLink)和或W02017053297A1(在此通过引用并入)公开的帽结构、肌苷、N1-甲基-鸟苷、2'-氟-鸟苷、7-脱氮杂-鸟苷、8-氧代-鸟苷、2-氨基-鸟苷、LNA-鸟苷、和2-叠

氮-鸟苷。

[0481] 根据优选的实施方式,人工核酸包括在RNA 5'端帽结构(帽-1)相邻第一核苷酸的核糖-2'-0位的2'-0位的甲基。一般,甲基化可通过帽2'-0-甲基转移酶的作用,利用带m7GpppN帽的人工核酸(优选地RNA)作为底物和S-腺苷甲硫氨酸(SAM)作为甲基供体(以甲基化带帽RNA(帽-0),产生帽-1结构)来实现。帽-1结构已被报告增强mRNA翻译效率和因此可有助于提高本发明人工核酸,优选地RNA的表达效力,如本文所述。

[0482] 聚(A)

[0483] 根据进一步优选的实施方式,本发明的人工核酸分子(优选地RNA)(或本文限定的任何其他核酸,具体地RNA)包含聚(A)序列。

[0484] 聚(A)序列,也称聚(A)尾或3'-聚(A)尾,一般被理解为腺苷核苷酸序列,例如,上至约400个腺苷核苷酸,例如约20至约400个,优选约50至约400个,更优选约50至约300个,甚至更优选约50至约250个,最优选约60至约250个腺苷核苷酸。如本文所用,聚(A)序列还可包括约10至200个腺苷核苷酸,优选约10至100个腺苷核苷酸,更优选约40至80个腺苷核苷酸,或甚至更优选约50至70个腺苷核苷酸。聚(A)序列一般位于RNA(具体地mRNA)的3'端。

[0485] 因此,在进一步优选的实施方式中,本发明的人工核酸分子(优选地RNA)(或本文限定的任何其他核酸,具体地RNA)在其3'端包含聚(A)尾,一般约10至200个腺苷核苷酸,优选约10至100个腺苷核苷酸,更优选约40至80个腺苷核苷酸,或甚至更优选约50至70个腺苷核苷酸。

[0486] 优选地,本发明的人工核酸分子(优选地RNA)(或所述其他核酸,具体地RNA)中的聚(A)序列通过RNA体外转录衍生自DNA模板。可选地,聚(A)序列还可通过常用的化学合成方法体外获得,而不一定从DNA祖先转录。

[0487] 此外,聚(A)序列或聚(A)尾可以通过使用市售的聚腺苷酸化试剂盒和本领域已知的相应方案,通过本发明的人工核酸分子(优选RNA)(或所述其他核酸,具体地RNA)的酶促聚腺苷酸化而生成。聚腺苷酸化一般被理解为向核酸分子(如RNA分子,例如成熟前mRNA)添加聚(A)序列。聚腺苷酸化可以由所谓的聚腺苷酸化信号引起。该信号优选位于mRNA的3'端的待聚腺苷酸化的一段核苷酸内。聚腺苷酸化信号一般包含由腺嘌呤和尿嘧啶/胸腺嘧啶核苷酸组成的六聚体,优选六聚体序列AAUAAA。其他序列,优选六聚体序列,也是可想到的。聚腺苷酸化一般在前mRNA(也称为成熟前mRNA)的加工过程中发生。一般,RNA成熟化(从前mRNA到成熟mRNA)包括聚腺苷酸化步骤。

[0488] 因此,本发明的人工核酸分子(优选地RNA)(或本文限定的任何其他核酸,具体地RNA)可包括通过特定蛋白质因子(例如切割和聚腺苷酸化特异性因子(CPSF),切割刺激因子(CstF)、切割因子I和II(CF I和CF II)、聚(A)聚合酶(PAP))将聚腺苷酸化传递给(转录的)RNA的聚腺苷酸化信号。

[0489] 在此环境下,优选共识聚腺苷酸化信号,包括NN(U/T)ANA共识序列。在特别优选的方面,聚腺苷酸化信号包括下列序列中的一种:AA(U/T)AAA或A(U/T)(U/T)AAA(其中尿苷通常存在于RNA中,并且胸苷通常存在于DNA中)。

[0490] 聚(C)

[0491] 根据进一步优选的实施方式,本发明的人工核酸分子(优选地RNA)(或本文限定的任何其他核酸,具体地RNA)包含在3'端的聚(C)尾,一般约10至200个胞嘧啶核苷酸,优选约

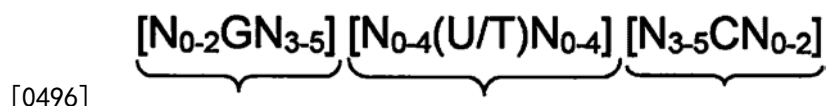
10至100个胞嘧啶核苷酸,更优选约20至70个胞嘧啶核苷酸,或甚至更优选约20至60个或甚至10至40个胞嘧啶核苷酸。

[0492] 组蛋白茎-环 (HSL)

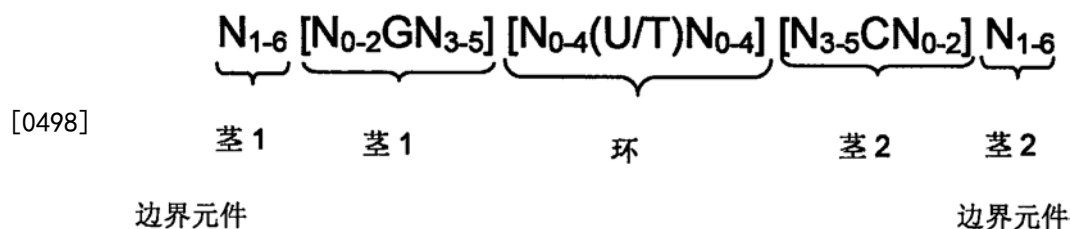
[0493] 在一些优选的实施方式中,本发明的人工核酸分子(优选地RNA)(或本文限定的任何其他核酸,具体地RNA)包括组蛋白茎-环(HSL)序列/结构。这种组蛋白茎-环序列优选选自W0 2012/019780公开的组蛋白茎-环序列,其公开内容提供引用并入本文。

[0494] 适用于本发明的组蛋白茎-环序列优选选自下式(I)或(II)中的至少一种:

[0495] 式(I)(茎-环序列,无茎边界元件):



[0497] 式(II)(茎-环序列,有茎边界元件):



[0499] 其中:

[0500] 茎1或茎2边界元件 $N_{1-6}$ 是1至6个,优选2至6个,更优选2至5个,甚至更优选3至5个,最优选4至5个或5个N的连续序列,其中各N彼此独立地选自以下核苷酸:选自A、U、T、G和C、或其核苷酸类似物;

[0501] 茎1 $[N_{0-2}GN_{3-5}]$ 与元件茎2反向互补或部分反向互补,并且是5至7个核苷酸的连续序列;

[0502] 其中 $N_{0-2}$ 是0至2个,优选0至1个,更优选1个N的连续序列,其中各N彼此独立地选自以下核苷酸:选自A、U、T、G和C或其核苷酸类似物;

[0503] 其中 $N_{3-5}$ 是3至5个,优选4至5个,更优选4个N的连续序列,其中各N彼此独立地选自以下核苷酸:选自A、U、T、G和C或其核苷酸类似物,并且

[0504] 其中G是鸟苷或其类似物,并且可任选地被胞苷或其类似物替代,条件是茎2中其互补核苷酸胞苷被鸟苷替代;

[0505] 环序列 $[N_{0-4}(U/T)N_{0-4}]$ 位于元件茎1和茎2之间,并且是3至5个核苷酸,更优选4个核苷酸的连续序列;

[0506] 其中各 $N_{0-4}$ 彼此独立地是0至4个,优选1至3个,更优选1至2个N的连续序列,其中各N彼此独立地选自以下核苷酸:选自A、U、T、G和C或其核苷酸类似物;并且

[0507] 其中U/T表示尿苷、或任选地胸苷;

[0508] 茎2 $[N_{3-5}CN_{0-2}]$ 与元件茎1反向互补或部分反向互补,并且是5至7个核苷酸的连续序列;

[0509] 其中 $N_{3-5}$ 是3至5个,优选4至5个,更优选4个N的连续序列,其中各N彼此独立地选自以下核苷酸:选自A、U、T、G和C或其核苷酸类似物;

[0510] 其中 $N_{0-2}$ 是0至2个,优选0至1个,更优选1个N的连续序列,其中各N彼此独立地选自以下核苷酸:选自A、U、T、G和C或其核苷酸类似物;并且

[0511] 其中C是胞苷或其类似物,并且可任选地被鸟苷或其类似物替代,条件是茎1中其互补核苷酸鸟苷被胞苷替代;

[0512] 其中

[0513] 茎1和茎2能够彼此碱基配对,形成反向互补序列,其中碱基配对可发生在茎1和茎2之间,例如通过核苷酸A和U/T或G和C的Watson-Crick碱基配对或通过非Watson-Crick碱基配对,例如wobble碱基配对、反向Watson-Crick碱基配对、Hoogsteen碱基配对、反向Hoogsteen碱基配对,或能够彼此碱基配对,形成部分反向互补序列,其中不完全碱基配对可发生在茎1和茎2之间——基础是一个茎中的一种或多种碱基不具有另一个茎的反向互补序列中的互补碱基。

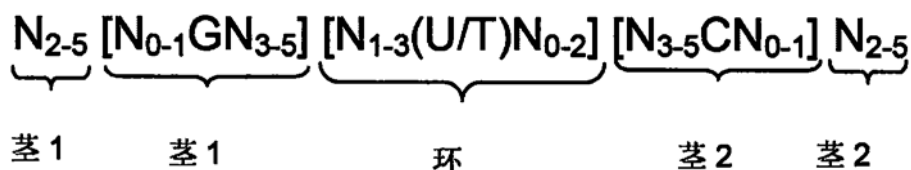
[0514] 根据进一步优选的实施方式,本发明的人工核酸分子(优选地RNA)(或本文限定的任何其他核酸,具体地RNA)可包括根据下列具体式(Ia)或(IIa)中至少一种的至少一种组蛋白茎-环(HSL)序列:

[0515] 式(Ia)(茎-环序列,无茎边界元件):



[0516]

[0517] 式(IIa)(茎-环序列,有茎边界元件):



[0518]

边界元件

边界元件

[0519] 其中:

[0520] N、C、G、T和U如上限定。

[0521] 根据另外特别优选的实施方式,本发明的人工核酸分子(优选地RNA)(或本文限定的任何其他核酸,具体地RNA)可包括根据下列具体式(Ib)或(IIb)中至少一种的至少一种组蛋白茎-环序列:

[0522] 式(Ib)(茎-环序列,无茎边界元件):



[0523]

[0524] 式(IIb)(茎-环序列,有茎边界元件):



[0525]

茎 1

茎 1

环

茎 2

茎 2

边界元件

边界元件

[0526] 其中:

[0527] N、C、G、T和U are如上限定。

[0528] 特别优选的组蛋白茎-环序列是序列CAAAGGCTCTTTTCAGAGCCACCA,或更优选相应的RNA序列CAAAGGCUCUUUCAGAGCCACCA。

[0529] miRNA部分

[0530] 在进一步的实施方式中,非编码部分,即例如miRNA部分,与本发明的序列组合。非编码部分可选自来自以下列举公开的一种或多种的核酸序列:

[0531] 5' -UTR;

[0532] 3' -UTR;

[0533] miRNA部分;

[0534] 帽;

[0535] 聚(C)序列

[0536] 组蛋白茎-环序列

[0537] 聚(A)序列或聚腺苷酸化信号;

[0538] IRES部分

[0539] 发夹部分

[0540] 用于RNA结合蛋白的部分

[0541] 防止3' -5' 降解的部分

[0542] 调控RNA衰减率的部分

[0543] 以上是上位概念。落入这些上位概念的下位部分也被本发明提供。下文提供来自上述列举的部分的详细内容,包括涉及具体实施方式的序列。

[0544] 虽然上述列举提供了单数形式的项目,但相当于可以选择多于一个的对应部分。上述列举未包括的核酸部分可相当地被选择。优选地,至少一种模块或部分来自于上述列举。

[0545] 在一般的实施方式中,选择至少一个5' -UTR部分和/或至少一个3' -UTR部分。优选地,选择至少一个5' -UTR和至少一个3' -UTR。

[0546] miRNA也可被选择作为本发明的部分。可选择本领域已知的任何miRNA部分。这种部分可选自微RNA靶序列、微RNA序列、或微RNA种子。例如,miRNA序列(微RNA靶序列、微RNA序列、或微RNA种子)被描述于W02015085318A2、US2005/0261218、US20170211066、W02017201332、W02017201328、W02017201349、W02017201347、W02017201348、W02017201342、W02017201346、US20160177295、W02014113089、EP2946014、W02016100812、W02013126803和US2005/0059005(所有上述参考文献均通过引用并入本文)。微RNA、微RNA靶区域、和其表达模式的确定和在生物学中的作用已被报告(Bonauer et al., Curr Drug Targets 2010 11:943-949; Anand and Cheresch, Curr Opin Hematol 2011, 18:171-176;

Contreras and Rao Leukemia, 201226:404-413; Barrel, Cell, 2009 136:215-233; Landgraf et al., Cell, 2007129:1401-1414) (所有上述参考文献均通过引用并入本文)。

[0547] 总体上,微RNA(或miRNA)是19-25个核苷酸长的非编码RNA。miRNA结合至核酸分子的3'-UTR。这导致基因表达下调——通过降低核酸分子稳定性或通过抑制翻译。作为本发明的模块,本发明的多核苷酸可包括一种或多种微RNA靶序列、微RNA序列、或微RNA种子。

[0548] 如本文所用,术语“微RNA位点”指代微RNA可结合或以其他方式缔合的多核苷酸序列。“结合”一般通过Watson-Crick杂交而发生;但微RNA与靶序列在微RNA位点处或相邻处的任何其他方式的稳定缔合也被包括在根据本发明的“微RNA位点”的思路内。

[0549] 总体上,微RNA序列包括“种子”区,即,一般在成熟微RNA的第2-8位区域的序列。种子区序列与miRNA靶序列具有完全Watson-Crick互补性。这种微RNA种子可包括成熟微RNA的第2-8位或可选地第2-7位。因此,在一个实施方式中,微RNA种子包括7个核苷酸(例如,成熟微RNA核苷酸2-8),其中相应miRNA靶中的种子互补位点侧翼是腺嘌呤(A)——与微RNA第1位相反。在另一实施方式中,微RNA种子包括6个核苷酸(例如,成熟微RNA的核苷酸2-7),其中相应miRNA靶中的种子互补位点侧翼是腺嘌呤(A)——与微RNA第1位相反。对应的核酸模块被公开于Grimson et al.; Mol Cell. 2007 Jul 6; 27(1):91-105。

[0550] 在本发明中,微RNA靶序列一般被设计以被包括在3'-UTR或以其他方式被包括在开放阅读框的3' (上游)中。在这种情况下,认为miRNA靶序列靶向要降解或减少翻译的分子,条件是所述相应微RNA是可用的。这允许控制递送本发明的核酸分子后任何不期望的脱靶效应。

[0551] 在不期望在肝脏中翻译mRNA但mRNA被运输到肝脏或以其他方式终止于此的情况下,miR-122(肝脏中富含的微RNA)可以抑制本发明核酸的表达\_如果miR-122的一种或多种靶位点存在(例如,被设计)于本发明的多核苷酸的3'-UTR区域中。针对不同微RNA的一种或多种结合位点的引入可被工程化,以进一步影响(例如降低)多核苷酸的寿命、稳定性和蛋白质翻译。

[0552] 相比之下,在确实期望翻译mRNA的情况下,微RNA结合位点可被从其存在的序列工程化掉(即去除),例如以增加在特定组织中的蛋白质表达。例如,可以去除一种或多种miR-122结合位点以提高在肝脏中的蛋白质表达。

[0553] 由此,特定组织中的表达调控可通过引入或去除一个或数个微RNA结合位点而实现。例如,已知微RNA调控mRNA,从而调控蛋白质表达——无限制地在肝脏(miR-122)、心脏(miR-1d、miR-149)、内皮细胞(miR-17-92、miR-126)、脂肪组织(let-7、miR-30c)、肾脏(miR-192、miR-194、miR-204)、骨髓细胞(miR-142-3p、miR-142-5p、miR-16、miR-21、miR-223、miR-24、miR-27)、肌肉(miR-133、miR-206、miR-208)、和肺上皮细胞(let-7、miR-133、miR-126)中。微RNA还可调控复杂的生物学过程,如血管生成(miR-132)(例如Anand and Cheresch, Curr. Opin. Hematol. 2011, 18:171-176)。

[0554] 因此,总体上,根据本发明的模块设计原理,可以去除或引入微RNA结合位点,以便将多核苷酸表达的表达式调整至期望的细胞类型或组织,或相关生物学过程的环境。miRNA序列和结合位点的列举公众可得。本文讨论的文献中公开的任何序列可用于本发明的环境:驱动组织或疾病特异性基因表达的微RNA的实例列于Getner and Naldini, Tissue Antigens. 2012, 80:393-403中。微RNA种子位点并入的实例是将miR-142位点并入表达

UGT1A1的慢病毒载体,这导致在抗原呈递细胞中的表达减少,导致针对病毒表达型UGT1A1的免疫应答不存在,如公开于Schmitt et al.,Gastroenterology 2010;139:999-1007; Gonzalez-Asequinolaza et al.Gastroenterology 2010,139:726-729。因此,在治疗患有完全蛋白质缺乏的患者(UGT1A1 I型、LDLR缺陷患者、CRIM阴性Pompe患者等)的情况下,将一种或多种miR-142种子位点并入mRNA中被认为是重要的。因此,本发明的核酸分子可以被设计以适合这样的目的。

[0555] 可以选择任何具有如下特征的多核苷酸:与任何这样的miRNA序列具有至少80%同一性,至少85%同一性,优选至少90%同一性,更优选至少95%同一性。

[0556] 由于微RNA在不同细胞类型中的表达模式不同,本发明允许特异性地设计用于在特定细胞类型中或在特定生物条件下靶向表达的多核苷酸分子。通过引入组织特异性微RNA结合位点,可以设计多核苷酸,用于在组织中或在生物学条件的环境下的蛋白质表达。

[0557] 在本发明的一个实施方式中,编码本发明CRISPR相关蛋白质(例如Cas9,Cpf1)的本发明人工核酸分子包括选自下列的至少一种miRNA序列:hsa-miR-27a-3p、hsa-miR-99b-5p、bsa-miR-21-5p、hsa-miR-142-5p、hsa-miR-27a-3p、hsa-miR-21-5p、hsa-miR-223-3p、hsa-miR-150-5p、和hsa-miR-142-5p。

[0558] 构建体

[0559] 包括本文限定的至少一种编码序列的本发明人工核酸分子(优选地RNA)包括本文所述的至少一个5' UTR和至少一个3' UTR,和任选地至少一种组蛋白茎-环。

[0560] 本发明的人工核酸分子(优选地RNA)(或本文限定的任何其他核酸,具体地RNA)的3' UTR可进一步包括本文限定的聚(A)和/或聚(C)序列。3' UTR单个元件可在其中沿本发明人工核酸分子(优选地RNA)的序列从5'至3'以任何顺序存在。

[0561] 另外,本文所述的进一步元件也可被包含在内,如本文限定的稳定化序列(例如衍生自珠蛋白基因的UTR)、RES序列等。各元件还可在本发明的人工核酸分子(优选地RNA)中重复至少一次(具体地,在双或多顺反子构建体中),例如两次或更多次。作为实例,单个元件可以下列顺序存在于本发明的人工核酸分子(优选地RNA)中:

[0562] 5'-编码序列-组蛋白茎-环-聚(A)/(C)序列-3';或

[0563] 5'-编码序列-聚(A)/(C)序列-组蛋白茎-环-3';或

[0564] 5'-编码序列-组蛋白茎-环-聚腺苷酸化信号-3';或

[0565] 5'-编码序列-聚腺苷酸化信号-组蛋白茎-环-3';或

[0566] 5'-编码序列-组蛋白茎-环-组蛋白茎-环-聚(A)/(C)序列-3';或

[0567] 5'-编码序列-组蛋白茎-环-组蛋白茎-环-聚腺苷酸化信号-3';或

[0568] 5'-编码序列-稳定化序列-聚(A)/(C)序列-组蛋白茎-环-3';或

[0569] 5'-编码序列-稳定化序列-聚(A)/(C)序列-聚(A)/(C)序列-组蛋白茎-环-3';等。

[0570] 根据进一步实施方式,人工核酸分子(优选地RNA)优选进一步包括下列结构元件中的至少一种:在其3'非翻译区中的组蛋白-茎-环结构,优选组蛋白-茎-环;5'-帽结构;聚A尾;或聚(C)序列。

[0571] 根据一些实施方式,特别优选如果在CRISPR相关蛋白质以外还有其他肽或蛋白质被本文限定的至少一种编码序列编码,则该被编码的肽或蛋白质优选不是组蛋白,不是报告蛋白(例如荧光素酶、GFP和其变体(如eGFP、RFP或BFP),和/或不是标记或选择蛋白——



包括 $\alpha$ -珠蛋白,半乳糖激酶和黄嘌呤:鸟嘌呤磷酸核糖基转移酶(GPT)、次黄嘌呤-鸟嘌呤磷酸核糖基转移酶(HGPRT)、 $\beta$ -半乳糖苷酶、半乳糖激酶、碱性磷酸酶、分泌胚胎碱性磷酸酶(SEAP)或抗性基因(如抗新霉素、嘌呤霉素、潮霉素和博来霉素的抗性基因)。在优选的实施方式中,人工核酸分子(优选地RNA)不编码报告基因或标记基因。在优选的实施方式中,人工核酸分子(优选地RNA)不编码荧光素酶。在其他实施方式中,人工核酸分子(优选地RNA)不编码GFP或其变体。

[0572] 具体地,根据本发明的人工核酸分子,具体地RNA,可优选沿5'至3'方向包括下列元件:

[0573] a) 5'-帽结构,优选地m7GpppN或帽1

[0574] b) 5'-UTR元件,其包括下列或由下列组成:本文限定的衍生自5'-UTR的核酸序列,优选地包括与根据SEQ ID NO:1;3;5;7;9;11;13;15;17;19;或21的核酸序列相应的核酸序列或其同源物、片段或变体;

[0575] c) 本文限定的至少一种编码序列;

[0576] d) 3'-UTR元件,其包括下列或由下列组成:本文限定的衍生自3'-UTR的核酸序列,优选地包括与根据SEQ ID NO:23;25;27;29;31;33或35的核酸序列相应的核酸序列或其同源物、片段或变体,

[0577] e) 任选地聚(A)尾,其优选由10至1000个、10至500个、10至300个、10至200个、10至100个、40至80个、或50至70个腺苷核苷酸组成,

[0578] f) 任选地聚(C)尾,其优选由10至200个、10至100个、20至70个、20至60个、或10至40个胞嘧啶核苷酸组成的,和

[0579] g) 任选地组蛋白茎-环。

[0580] 优选的人工核酸构建体在下文中被详细讨论。

[0581] HSD17B4衍生的5' UTR元件和GNAS衍生的3' UTR元件:

[0582] 在一些优选的实施方式中,根据本发明的人工核酸包括下列或由下列组成:衍生自HSD17B4基因的5' UTR、或其同源物、片段或变体的至少一个5' UTR元件,和衍生自GNAS基因的3' UTR、或其同源物、片段或变体的至少一个3' UTR元件;其中所述人工核酸包括下列或由下列组成:根据SEQ ID NO:413;2330-2345;3490-3505;4650-4665;5810-5825;6970-6985;8130-8145;9290-9305;10402-10408;10554;10599-10612的核酸序列、或其同源物、变体或片段,具体地按递增优选顺序与这些序列中的任意者具有至少5%、10%、20%、30%、40%、50%、60%、70%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、或99%,优选至少70%,更优选至少80%,甚至更优选至少85%,甚至更优选至少90%,并且最优选至少95%或甚至97%序列同一性的核酸序列。

[0583] SLC7A3衍生的5' UTR元件和GNAS衍生的3' UTR元件:

[0584] 在一些优选的实施方式中,人工核酸根据本发明包括下列或由下列组成:衍生自SLC7A3基因的5' UTR、或其同源物、片段或变体的至少一个5' UTR元件,和衍生自GNAS基因的3' UTR、或其同源物、片段或变体的至少一个3' UTR元件,其中所述人工核酸包括下列或由下列组成:根据SEQ ID NO:414;2346-2361;3506-3521;4666-4681;5826-5841;6986-7001;8146-8161;9306-9321;10409-10415;10555;10613-10626的核酸序列、或其同源物、变体或片段,具体地按递增优选顺序与这些序列中的任意者具有至少5%、10%、20%、30%、40%、

50%、60%、70%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、或99%，优选至少70%，更优选至少80%，甚至更优选至少85%，甚至更优选至少90%，并且最优选至少95%或甚至97%序列同一性的核酸序列。

[0585] ATP5A1衍生的5' UTR元件和CASP1衍生的3' UTR元件：

[0586] 在一些优选的实施方式中，人工核酸根据本发明包括下列或由下列组成：衍生自ATP5A1基因的5' UTR、或其同源物、片段或变体的至少一个5' UTR元件，和衍生自CASP1基因的3' UTR、或其同源物、片段或变体的至少一个3' UTR元件，其中所述人工核酸包括下列或由下列组成：根据SEQ ID NO:415;2362-2377;3522-3537;4682-4697;5842-5857;7002-7017;8162-8177;9322-9337;10416-10422;10556;10627-10640的核酸序列、或其同源物、变体或片段，具体地按递增优选顺序与这些序列中的任意者具有至少5%、10%、20%、30%、40%、50%、60%、70%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、或99%，优选至少70%，更优选至少80%，甚至更优选至少85%，甚至更优选至少90%，并且最优选至少95%或甚至97%序列同一性的核酸序列。

[0587] NDUFA4衍生的5' UTR元件和PSMB3衍生的3' UTR元件：

[0588] 在一些优选的实施方式中，人工核酸根据本发明包括下列或由下列组成：衍生自NDUFA4基因的5' UTR、或其同源物、片段或变体的至少一个5' UTR元件，和衍生自PSMB3基因的3' UTR、或其同源物、片段或变体的至少一个3' UTR元件，其中所述人工核酸包括下列或由下列组成：根据SEQ ID NO:416;2378-2393;3538-3553;4698-4713;5858-5873;7018-7033;8178-8193;9338-9353;10423-10429;10557;10641-10654的核酸序列、或其同源物、变体或片段，具体地按递增优选顺序与这些序列中的任意者具有至少5%、10%、20%、30%、40%、50%、60%、70%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、或99%，优选至少70%，更优选至少80%，甚至更优选至少85%，甚至更优选至少90%，并且最优选至少95%或甚至97%序列同一性的核酸序列。

[0589] HSD17B4衍生的5' UTR元件和PSMB3衍生的3' UTR元件：

[0590] 在一些优选的实施方式中，根据本发明的人工核酸包括下列或由下列组成：衍生自HSD17B4基因的5' UTR、或其同源物、片段或变体的至少一个5' UTR元件，和衍生自PSMB3基因的3' UTR、或其同源物、片段或变体的至少一个3' UTR元件，其中所述人工核酸包括下列或由下列组成：根据SEQ ID NO:417;2394-2409;3554-3569;4714-4729;5874-5889;7034-7049;8194-8209;9354-9369;10430-10436;10558;10655-10668的核酸序列、或其同源物、变体或片段，具体地按递增优选顺序与这些序列中的任意者具有至少5%、10%、20%、30%、40%、50%、60%、70%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、或99%，优选至少70%，更优选至少80%，甚至更优选至少85%，甚至更优选至少90%，并且最优选至少95%或甚至97%序列同一性的核酸序列。

[0591] RPL32衍生的5' UTR元件和ALB衍生的3' UTR元件：

[0592] 在一些优选的实施方式中，根据本发明的人工核酸包括下列或由下列组成：衍生自RPL32基因的5' UTR、或其同源物、片段或变体的至少一个5' UTR元件，和衍生自ALB基因的3' UTR、或其同源物、片段或变体的至少一个3' UTR元件，其中所述人工核酸包括下列或由下列组成：根据SEQ ID NO:418;2410-2425;3570-3585;4730-4745;5890-5905;7050-7065;8210-8225;9370-9385;10437-10443;10559;10669-10682的核酸序列或其同源物、变体或

片段,具体地按递增优选顺序与这些序列中的任意者具有至少5%、10%、20%、30%、40%、50%、60%、70%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、或99%,优选至少70%,更优选至少80%,甚至更优选至少85%,甚至更优选至少90%,并且最优选至少95%或甚至97%序列同一性的核酸序列。

[0593] HSD17B4衍生的5' UTR元件和CASP1衍生的3' UTR元件:

[0594] 在一些优选的实施方式中,根据本发明的人工核酸包括下列或由下列组成:衍生自HSD17B4基因的5' UTR、或其同源物、片段或变体的至少一个5' UTR元件,和衍生自CASP1基因的3' UTR、或其同源物、片段或变体的至少一个3' UTR元件,其中所述人工核酸包括下列或由下列组成:根据SEQ ID NO:420;2442-2457;3602-3617;4762-4777;5922-5937;7082-7097;8242-8257;9402-9417;10451-10457;10561;10697-10710的核酸序列、或其同源物、变体或片段,具体地按递增优选顺序与这些序列中的任意者具有至少5%、10%、20%、30%、40%、50%、60%、70%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、或99%,优选至少70%,更优选至少80%,甚至更优选至少85%,甚至更优选至少90%,并且最优选至少95%或甚至97%序列同一性的核酸序列。

[0595] SLC7A3衍生的5' UTR元件和CASP1衍生的3' UTR元件:

[0596] 在一些优选的实施方式中,根据本发明的人工核酸包括下列或由下列组成:衍生自SLC7A3基因的5' UTR、或其同源物、片段或变体的至少一个5' UTR元件,和衍生自CASP1基因的3' UTR、或其同源物、片段或变体的至少一个3' UTR元件,其中所述人工核酸包括下列或由下列组成:根据SEQ ID NO:421;2458-2473;3618-3633;4778-4793;5938-5953;7098-7113;8258-8273;9418-9433;10458-10464;10562;10711-10724的核酸序列、或其同源物、变体或片段,具体地按递增优选顺序与这些序列中的任意者具有至少5%、10%、20%、30%、40%、50%、60%、70%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、或99%,优选至少70%,更优选至少80%,甚至更优选至少85%,甚至更优选至少90%,并且最优选至少95%或甚至97%序列同一性的核酸序列。

[0597] SLC7A3衍生的5' UTR元件和PSMB3衍生的3' UTR元件:

[0598] 在一些优选的实施方式中,根据本发明的人工核酸包括下列或由下列组成:衍生自SLC7A3基因的5' UTR、或其同源物、片段或变体的至少一个5' UTR元件,和衍生自PSMB3基因的3' UTR、或其同源物、片段或变体的至少一个3' UTR元件,其中所述人工核酸包括下列或由下列组成:根据SEQ ID NO:422;2474-2489;3634-3649;4794-4809;5954-5969;7114-7129;8274-8289;9434-9449;10465-10471;10563;10725-10738的核酸序列、或任何所述序列的变体或片段,具体地按递增优选顺序与这些序列中的任意者具有至少5%、10%、20%、30%、40%、50%、60%、70%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、或99%,优选至少70%,更优选至少80%,甚至更优选至少85%,甚至更优选至少90%,并且最优选至少95%或甚至97%序列同一性的核酸序列。

[0599] NOSIP衍生的5' UTR元件和PSMB3衍生的3' UTR元件:

[0600] 在一些优选的实施方式中,根据本发明的人工核酸包括下列或由下列组成:衍生自NOSIP基因的5' UTR、或其同源物、片段或变体的至少一个5' UTR元件,和衍生自PSMB3基因的3' UTR、或其同源物、片段或变体的至少一个3' UTR元件,其中所述人工核酸包括下列或由下列组成:根据SEQ ID NO:423;2490-2505;3650-3665;4810-4825;5970-5985;7130-7145;

8290-8305;9459-9450;10472-10478;10564;10739-10752的核酸序列、或其同源物、变体或片段,具体地按递增优选顺序与这些序列中的任意者具有至少5%、10%、20%、30%、40%、50%、60%、70%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、或99%,优选至少70%,更优选至少80%,甚至更优选至少85%,甚至更优选至少90%,并且最优选至少95%或甚至97%序列同一性的核酸序列。

[0601] NDUFA4衍生的5' UTR元件和RPS9衍生的3' UTR元件:

[0602] 在一些优选的实施方式中,根据本发明的人工核酸包括下列或由下列组成:衍生自NDUFA4基因的5' UTR、或其同源物、片段或变体的至少一个5' UTR元件,和衍生自RPS9基因的3' UTR、或其同源物、片段或变体的至少一个3' UTR元件,其中所述人工核酸包括下列或由下列组成:根据SEQ ID NO:424;2506-2521;3666-3681;4826-4841;5986-6001;7146-7161;8306-8321;9466-9481;10479-10485;10565;10753-10766的核酸序列、或其同源物、变体或片段,具体地按递增优选顺序与这些序列中的任意者具有至少5%、10%、20%、30%、40%、50%、60%、70%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、或99%,优选至少70%,更优选至少80%,甚至更优选至少85%,甚至更优选至少90%,并且最优选至少95%或甚至97%序列同一性的核酸序列。

[0603] HSD17B4衍生的5' UTR元件和RPS9衍生的3' UTR元件:

[0604] 在一些优选的实施方式中,根据本发明的人工核酸包括下列或由下列组成:衍生自HSD17B4基因的5' UTR、或其同源物、片段或变体的至少一个5' UTR元件,和衍生自RPS9基因的3' UTR、或其同源物、片段或变体的至少一个3' UTR元件,其中所述人工核酸包括下列或由下列组成:根据SEQ ID NO:425;2522-2537;3682-3697;4842-4857;6002-6017;7162-7177;8322-8337;9482-9497;10486-10492;10566;10767-10780的核酸序列、或其同源物、变体或片段,具体地按递增优选顺序与这些序列中的任意者具有至少5%、10%、20%、30%、40%、50%、60%、70%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、或99%,优选至少70%,更优选至少80%,甚至更优选至少85%,甚至更优选至少90%,并且最优选至少95%或甚至97%序列同一性的核酸序列。

[0605] 在其他优选的实施方式中,根据本发明的人工核酸包括下列或由下列组成:下文所述的至少一个5' UTR和3' UTR元件(或其同源物、片段或变体),其中人工核酸包括下列或由下列组成:根据具体UTR组合后括号内的SEQ ID NO的核酸序列:

[0606] ATP5A1/Gnas.1 (SEQ ID NO:9498-9609;10493-10499;10567;10781-10794);

[0607] Ndufa4.1/COX6B1.1 (SEQ ID NO:9610-9721;10500-10506;10568;10795-10808);

[0608] Ndufa4.1/Gnas.1 (SEQ ID NO:9722-9833;10507-10513;10569;10809-10822);

[0609] Ndufa4.1/Ndufa1.1 (SEQ ID NO:9834-9945;10514-10520;10570;10823-10836);

[0610] Nosip.1/Ndufa1.1 (SEQ ID NO:9946-10057;10521-10527;10571;10837-10850);

[0611] Rpl31.1/Gnas.1 (SEQ ID NO:10058-10169;10528-10534;10572;10851-10864);

[0612] TUBB4B.1/RPS9.1 (SEQ ID NO:10170-10281;10535-10541;10573;10865-10878);

[0613] Ubqln2.1/RPS9.1 (SEQ ID NO:10282-10393;10542-10548;10574;10879-10892)。

[0614] 在其他优选的实施方式中,本发明构建体的3'端选自A64-C30-HSL-N5;A64-HSL-N5;A64-N5;A64-N5;A64-C5-N5;A64-C5-N5;A64-C10-N5;A64-C10-N5;A64-C15-N5;A40-HSL-A50-N5;HSL-N5;A64-HSL-NR;A64-N5;和A64-C15-N5。

[0615] 复合

[0616] 在优选的实施方式中,本发明的至少一种人工核酸分子(优选地RNA)(或本文所述的任何其他核酸)以复合形式被提供,即与一种或多种(聚)阳离子化合物,优选与(聚)阳离子聚合物、(聚)阳离子肽或蛋白质,例如鱼精蛋白、(聚)阳离子多糖和/或(聚)阳离子脂质复合或缔合。在此环境中,术语“复合”或“缔合”是指所述至少一种人工核酸分子,优选RNA(或所述其他核酸)与一种或多种前述化合物在没有共价结合的情况下基本稳定地组合成较大的复合物或组合物。

[0617] 脂质

[0618] 根据优选的实施方式,本发明的人工核酸分子(优选地RNA)与脂质(具体地阳离子和/或中性脂质)复合或缔合,形成一种或多种脂质体、脂质复合物、脂质纳米颗粒、或纳米脂质体。

[0619] 因此,在一些实施方式中,本发明的人工核酸分子(优选地RNA)(或本文限定的任何其他核酸,具体地RNA)以脂质系制剂形式被提供,具体地包括所述人工核酸分子(优选地RNA)(或所述其他核酸,具体地RNA)的脂质体、脂质复合物、和/或脂质纳米颗粒形式。

[0620] 脂质纳米颗粒

[0621] 根据一些优选实施方式,本发明的人工核酸分子(优选地RNA)(或本文限定的任何其他核酸,具体地RNA)与脂质(具体地阳离子和/或中性脂质)复合或缔合以形成一种或多种脂质纳米颗粒。在一些实施方式中,本发明的纳米颗粒(一种或多种)包括本文所述的编码CRISPR相关蛋白质的至少一种人工核酸分子以及本文所述的至少一种gRNA。

[0622] 优选地,脂质纳米颗粒(LNPs)包括:(a)本发明的至少一种人工核酸分子(优选地RNA)(或本文限定的任何其他核酸),(b)阳离子脂质,(c)聚集减少剂(如聚乙二醇(PEG)脂质或PEG修饰脂质),(d)任选地非阳离子脂质(如中性脂质),和(e)任选地,固醇。

[0623] 在一些实施方式中,除至少一种本发明的人工核酸分子(优选地RNA)(或本文限定的任何其他核酸)外,LNP还包括(i)至少一种阳离子脂质;(ii)中性脂质;(iii)固醇,例如,胆固醇;和(iv)PEG-脂质,其摩尔比为约20-60%阳离子脂质:5-25%中性脂质:25-55%固醇;0.5-15%PEG-脂质。

[0624] 在一些实施方式中,本发明的人工核酸分子(优选地RNA)(或本文限定的任何其他核酸)可被配制在氨基醇类脂质中。可用于本发明的氨基醇类脂质可通过美国专利号8,450,298(其整体通过引用并入本文)描述的方法制备。

[0625] (i) 阳离子脂质

[0626] LNP可包括适于形成脂质纳米颗粒的任何阳离子脂质。优选地,阳离子脂质在大约生理pH下携带净正电荷。

[0627] 阳离子脂质可以是氨基脂质。如本文所用,术语“氨基脂质”意为包括具有一个或两个脂肪酸或脂肪烷基链和可在生理pH下被质子化以形成阳离子脂质的氨基头部基团(包括烷基氨基或二烷基氨基)的那些脂质。

[0628] 阳离子脂质可以是,例如,N,N-二油基(dioleoyl)-N,N-二甲基氯化铵(DODAC)、N,N-二硬脂基-N,N-二甲基溴化铵(DDAB)、1,2-二油酰基三甲基丙烷氯化铵(DOTAP)(也称为N-(2,3-二油酰氧基)丙基)-N,N,N-三甲基氯化铵和1,2-二油基氧基(dioleoyloxy)-3-三甲基氨基丙烷氯化物盐)、N-(1-(2,3-二油基氧基)丙基)-N,N,N-三甲基氯化铵(DOTMA)、N,N-

二甲基-2,3-二油基氧基)丙胺(DODMA)、1,2-二亚油烯基氧基(DiLinoleyloxy)-N,N-二甲基氨基丙烷(DLinDMA)、1,2-二亚麻基氧基(Dilinenyloxy)-N,N-二甲基氨基丙烷(DLenDMA)、1,2-二- $\gamma$ -亚麻基氧基-N,N-二甲基氨基丙烷( $\gamma$ -DLenDMA)、1,2-二亚油烯基氨基甲酰氧基-3-二甲基氨基丙烷(DLin-C-DAP)、1,2-二亚油烯基氧基(dilinoleyloxy)-3-(二甲基氨基)乙酰氧基丙烷(DLin-DAC)、1,2-二亚油烯基氧基-3-吗啉代丙烷(DLin-MA)、1,2-二亚油酰基-3-二甲基氨基丙烷(DLinDAP)、1,2-二亚油烯基硫-3-二甲基氨基丙烷(DLin-S-DMA)、1-亚油酰基-2-亚油烯基氧基-3-二甲基氨基丙烷(DLin-2-DMA)、1,2-二亚油烯基氧基-3-三甲基氨基丙烷氯化物盐(DLin-TMA.Cl)、1,2-二亚油酰基-3-三甲基氨基丙烷氯化物盐(DLin-TAP.Cl)、1,2-二亚油烯基氧基-3-(N-甲基哌嗪基)丙烷(DLin-MPZ)、或3-(N,N-二亚油烯基氨基)-1,2-丙二醇(DLinAP)、3-(N,N-二亚油烯基氨基)-1,2-丙二醇(DOAP)、1,2-二亚油烯基氧基-3-(2-N,N-二甲基氨基)乙氧基丙烷(DLin-EG-DMA)、2,2-二亚油烯基-4-二甲基氨基甲基-[1,3]-二氧戊环(DLin-K-DMA)或其类似物、(3aR,5s,6aS)-N,N-二甲基-2,2-二((9Z,12Z)-十八碳-9,12-二烯基)四氢-3aH-环戊[d][1,3]二氧戊环-5-胺、(6Z,9Z,28Z,31Z)-三十七碳-6,9,28,31-四烯-19-基4-(二甲基氨基)丁酸酯(MC3)、1,1'-(2-(4-(2-((2-(双(2-羟基十二烷基)氨基)乙基)(2-羟基十二烷基)氨基)乙基)哌嗪-1-基)乙基氮烷二基)-二十二烷-2-醇(C12-200)、2,2-二亚油烯基-4-(2-二甲基氨基乙基)-[1,3]-二氧戊环(DLin-K-C2-DMA)、2,2-二亚油烯基-4-二甲基氨基甲基-[1,3]-二氧戊环(DLin-K-DMA)、(6Z,9Z,28Z,31Z)-三十七碳-6,9,28,31-四烯-19-基4-(二甲基氨基)丁酸酯(DLin-M-C3-DMA)、3-((6Z,9Z,28Z,31Z)-三十七碳-6,9,28,31-四烯-19-基氧)-N,N-二甲基丙-1-胺(MC3醚)、4-((6Z,9Z,28Z,31Z)-三十七碳-6,9,28,31-四烯-19-基氧)-N,N-二甲基丁-1-胺(MC4醚)、或前述任意者的任何组合。

[0629] 其他阳离子脂质包括但不限于N,N-二硬脂基-N,N-二甲基溴化铵(DDAB)、3P-(N-(N',N'-二甲基氨基乙烷)-氨基甲酰基)胆固醇(DC-Choi)、N-(1-(2,3-二油基氧基)丙基)-N-2-(精胺甲酰胺基)乙基)-N,N-二甲基三氟乙酸铵(DOSPA)、双十八烷基酰胺基甘氨酸羧基精胺(DOGS)、1,2-二油酰基-sn-3-磷酸乙醇胺(DOPE)、1,2-二油酰基-3-二甲基丙烷铵(DODAP)、N-(1,2-二肉豆蔻氧基丙-3-基)-N,N-二甲基-N-羟基乙基溴化铵(DRIE)和2,2-二亚油烯基-4-二甲基氨基乙基-[1,3]-二氧戊环(XTC)。另外、可以使用阳离子脂质的商业制剂,例如LIPOFECTIN(包括DOTMA和DOPE,可从GIBCO/BRL获得)和LIPOFECTAMINE(包括DOSPA和DOPE,可从GIBCO/BRL获得)。

[0630] 其他适当的阳离子脂质被公开于国际公开号W0 09/086558、W0 09/127060、W0 10/048536、W0 10/054406、W0 10/088537、W0 10/129709、和W0 2011/153493;美国专利公开号2011/0256175、2012/0128760、和2012/0027803;美国专利号8,158,601;和Love et al,PNAS,107(5),1864-69,2010。

[0631] 其他适当的氨基脂质包括具有替代性脂肪酸基团和其他二烷基氨基的那些,包括其中烷基取代基不同(例如,N-乙基-N-甲基氨基、和N-丙基-N-乙基氨基-)的那些。总体上,具有较少饱和酰基链的氨基脂质较容易设定尺寸,特别是当基于过滤灭菌目的复合物必须被设定尺寸到约0.3微米以下时。可使用包含碳链长度在C14至C22范围内的不饱和脂肪酸的氨基脂质。也可使用其他支架材料分隔氨基脂质的氨基和脂肪酸或脂肪烷基部分。

[0632] 在一些实施方式中,氨基或阳离子脂质具有至少一种可质子化或可去质子化的基

团,使得脂质在生理pH(例如pH 7.4)或生理pH以下的pH下带正电荷,并且在第二pH下(优选在生理pH或生理pH以上的pH下)为中性。当然会理解,作为pH的函数的质子添加或去除是平衡过程,并且带电或中性脂质的提及是指主要物种的性质而不要求所有脂质都以带电或中性形式存在。具有多于一个可质子化或可去质子化基团的脂质或两性离子性脂质不被排除在本发明的使用之外。

[0633] 在一些实施方式中,可质子化脂质具有约4至约11范围内的可质子化基团的pKa,例如,约5至约7的pKa。

[0634] LNP可包括两种或更多种阳离子脂质。阳离子脂质可被选择以提供不同的有利性质。例如,在诸如胺pKa、化学稳定性、循环半衰期、组织中的半衰期、组织中的净积累或毒性的性质上不同的阳离子脂质可用于LNP。具体地,阳离子脂质可被选择,使得混合LNP的性质比个体脂质的单LNP的性质更理想。

[0635] 在一些实施方式中,阳离子脂质的存在比例为LNP中存在的脂质总量的约20mol%至约70或75mol%、或约45至约65mol%或约20、25、30、35、40、45、50、55、60、65、或约70mol%。在进一步实施方式中,LNP包括在摩尔基础上约25%至约75%的阳离子脂质,例如,在摩尔基础上(基于脂质纳米颗粒中脂质的100%总摩尔数)约20至约70%、约35至约65%、约45至约65%、约60%、约50%或约40%。在一些实施方式中,阳离子脂质与核酸的比例为约3至约15,如约5至约13、或约7至约11。

[0636] 在一些实施方式中,脂质体可具有1:1与20:1之间的阳离子脂质中氮原子与RNA中磷酸基的摩尔比(N:P比),如国际公开号W02013/006825A1所述,其整体通过引用并入本文。在其他实施方式中,脂质体可具有大于20:1或小于1:1的N:P比。

[0637] 在一些方面,脂质选自98N12-5、C12-200、和cKK-E12。在一个实施方式中,核酸可被配制在氨基醇类脂质中。可用于本发明的氨基醇类脂质可通过美国专利号8,450,298(其整体通过引用并入本文)所述的方法制备。

[0638] 在另一实施方式中,可电离脂质也可以是国际公开号W02015/074085A1、美国申请号61/905,724和15/614,499、W02015074085A1、或美国专利或申请号9,593,077、61/905,724、9,593,077、9,567,296、15/614,499和9,567,296(其整体通过引用并入本文)公开的化合物。

[0639] 可电离脂质也可以是国际公开号W02017/075531A1(其整体通过引用并入本文)的表1、2和3和权利要求1-24公开的化合物。在另一实施方式中,可电离脂质也可以是国际公开号W02015/074085A1(即ATX-001至ATX-032或权利要求1-26提到的化合物)、美国申请号61/905,724和15/614,499或美国专利号9,593,077和9,567,296(其整体通过引用并入本文)公开的化合物。

[0640] (ii) 中性和非阳离子脂质

[0641] 非阳离子脂质可以是中性脂质、阴离子脂质、或两性脂质。中性脂质在存在时可以在生理pH下以不带电或中性两性离子形式存在的多种脂质物种中的任意种。这种脂质包括,例如,二酰基磷脂酰胆碱、二酰基磷脂酰乙醇胺、神经酰胺、鞘磷脂、二氢鞘磷脂、脑磷脂和脑苷脂。用于本文所述颗粒的中性脂质的选择总体上通过考虑例如LNP尺寸和LNP在血流中的稳定性来指导。优选地,中性脂质是具有两个酰基的脂质(例如,二酰基磷脂酰胆碱和二酰基磷脂酰乙醇胺)。

[0642] 在一些实施方式中,中性脂质含有碳链长度在C10至C20范围内的饱和脂肪酸。在其他实施方式中,使用具有碳链长度在C10至C20范围内的单不饱和或二不饱和脂肪酸的中性脂质。另外,可以使用具有饱和脂肪酸链和不饱和脂肪酸链混合物的中性脂质。

[0643] 适当的中性脂质包括但不限于,二硬脂酰基磷脂酰胆碱(DSPC)、二油酰基磷脂酰胆碱(DOPC)、二棕榈酰基磷脂酰胆碱(DPPC)、二油酰基磷脂酰甘油(DOPG)、二棕榈酰基磷脂酰甘油(DPPG)、二油酰基-磷脂酰乙醇胺(DOPE)、棕榈酰基油酰基磷脂酰胆碱(POPC)、棕榈酰基油酰基磷脂酰乙醇胺(POPE)、二油酰基-磷脂酰乙醇胺4-(N-马来酰亚胺基甲基)-环己烷-1-羧酸酯(DOPE-mal)、二棕榈酰基磷脂酰乙醇胺(DPPE)、二肉豆蔻酰基磷酸乙醇胺(DMPE)、二肉豆蔻酰基磷脂酰胆碱(DMPC)、二硬脂酰基-磷脂酰-乙醇胺(DSPE)、SM、16-0-单甲基PE、16-0-二甲基PE、18-1-反式PE、1-硬脂酰基-2-油酰基-磷脂酰乙醇胺(SOPE)、胆固醇、或其混合物。适用于LNP的阴离子脂质包括但不限于,磷脂酰甘油、心磷脂、二酰基磷脂酰丝氨酸、二酰基磷脂酸、N-十二烷酰基磷脂酰乙醇胺、N-琥珀酰基磷脂酰乙醇胺、N-戊二酰基磷脂酰乙醇胺、赖氨酰基磷脂酰甘油、和接合至中性脂质的其他阴离子修饰基团。

[0644] 两性脂质指代任何适当的材料,其中脂质材料的疏水部分朝向疏水相,而亲水部分朝向水相。这种化合物包括但不限于,磷脂、氨基脂质、和鞘脂。代表性磷脂包括鞘磷脂、磷脂酰胆碱、磷脂酰乙醇胺、磷脂酰丝氨酸、磷脂酰肌醇、磷脂酸、棕榈酰基油酰基磷脂酰胆碱、溶血磷脂酰胆碱、溶血磷脂酰乙醇胺、二棕榈酰基磷脂酰胆碱、二油酰基磷脂酰胆碱、二硬脂酰基磷脂酰胆碱、或二亚油酰基磷脂酰胆碱。也可使用其他缺磷化合物,如鞘脂、鞘糖脂脂质家族、二酰基甘油、和 $\beta$ -酰氧基酸。

[0645] 在一些实施方式中,非阳离子脂质的存在比例为LNP中存在的脂质总量的约5mol%至约90mol%、约5mol%至约10mol%、约5、10、15、20、25、30、35、40、45、50、55、60、65、70、75、80、85、或约90mol%。

[0646] 在一些实施方式中,LNP包括在摩尔基础上约0%至约15或45%的中性脂质,例如,约3至约12%、或约5至约10%。例如,LNP可包括在摩尔基础上(基于LNP中脂质100%总摩尔数)约15%、约10%、约7.5%、或约7.1%的中性脂质。

[0647] (iii) 固醇

[0648] 固醇优选是胆固醇。

[0649] 固醇可存在的比例为LNP的约10mol%至约60mol%或约25mol%至约40mol%。在一些实施方式中,固醇的存在比例为LNP中存在的脂质总量的约10、15、20、25、30、35、40、45、50、55、或约60mol%。在其他实施方式中,LNP包括在摩尔基础上约5%至约50%的固醇,例如,在摩尔基础上(基于LNP中脂质100%总摩尔数)约15%至约45%、约20%至约40%、约48%、约40%、约38.5%、约35%、约34.4%、约31.5%或约31%。

[0650] (iv) 聚集减少剂

[0651] 聚集减少剂可以是能够减少聚集的脂质。

[0652] 这种脂质的实例包括但不限于,聚乙二醇(PEG)修饰的脂质、单唾液酸神经节苷脂Gm1、和聚酰胺寡聚物(PAO),如美国专利号6,320,017(其整体通过引用并入)所述那些。在配制过程中防止聚集的具有不带电亲水空间位阻部分的其他化合物,如PEG、Gm1或ATTA,也可偶联至脂质。ATTA-脂质被描述于例如美国专利号6,320,017,并且PEG-脂质缀合物被描述于例如美国专利号5,820,873、5,534,499和5,885,613,其整体均通过引用并入。



[0653] 聚集减少剂可以例如选自聚乙二醇(PEG)-脂质,无限制地包括PEG-二酰基甘油(DAG)、PEG-二烷基甘油、PEG-二烷氧基丙基(DAA)、PEG-磷脂、PEG-神经酰胺(Cer)、或其混合物(如PEG-Cer14或PEG-Cer20)。PEG-DAA缀合物可以是例如PEG-双十二烷氧基丙基(C12)、PEG-双十四烷氧基丙基(C14)、PEG-双十六烷氧基丙基(C16)、或PEG-双十八烷氧基丙基(C18)。其他PEG化脂质包括但不限于,聚乙二醇-二肉豆蔻酰基甘油(C14-PEG或PEG-C14,其中PEG的平均分子量为2000Da)(PEG-DMG);(R)-2,3-双(十八烷氧基)丙基-1-(甲氧基聚(乙二醇)2000)丙基氨基甲酸酯(PEG-DSG);PEG-氨甲酰基-1,2-双十四烷氧基丙基胺,其中PEG的平均分子量为2000Da(PEG-cDMA);N-乙酰基半乳糖胺-((R)-2,3-双(十八烷氧基)丙基-1-(甲氧基聚(乙二醇)2000)丙基氨基甲酸酯))(GalNAc-PEG-DSG);mPEG(mw2000)-二硬脂酰基磷脂酰-乙醇胺(PEG-DSPE);和聚乙二醇-二棕榈酰基甘油(PEG-DPG)。

[0654] 在一些实施方式中,聚集减少剂是PEG-DMG。在其他实施方式中,聚集减少剂是PEG-c-DMA。

[0655] 在另一实施方式中,PEG-脂质还可以是US20150376115A1或W02015199952(其整体通过引用并入本文)公开的化合物。

[0656] LNP组合物

[0657] LNP组合物可受到阳离子脂质组分的选择、阳离子脂质饱和度、PEG化的本质、所有组分的比例和生物物理参数(如其尺寸)等的影响。在Semple等(Semple et al.Nature Biotech.2010 28:172-176;通过引用整体并入本文)的一个实例中,LNP组合物由57.1%阳离子脂质、7.1%二棕榈酰磷脂酰胆碱、34.3%胆固醇和1.4%PEG-c-DMA构成(Basha et al.Mol Ther.2011 19:2186-2200;通过引用整体并入本文)。

[0658] 在一些实施方式中,LNP可包括约35至约45%阳离子脂质、约40%至约50%阳离子脂质、约50%至约60%阳离子脂质、和/或约55%至约65%阳离子脂质。在一些实施方式中,脂质与核酸的比例可在如下范围内:约5:1至约20:1、约10:1至约25:1、约15:1至约30:1、和/或至少30:1。

[0659] PEG修饰的脂质中PEG部分的平均分子量可在约500至约8,000道尔顿(例如、约1,000至约4,000道尔顿)范围内。在一个优选的实施方式中,PEG部分的平均分子量为约2,000道尔顿。

[0660] 聚集减少剂的浓度可在相对于LNP中脂质的100%总摩尔数约0.1至约15mol%范围内。在一些实施方式中,LNP包括小于约3、2、或1摩尔%的PEG或PEG修饰脂质,基于LNP中的脂质总摩尔数。在进一步实施方式中,LNP包括在摩尔基础上约0.1%至约20%的PEG修饰脂质,例如,在摩尔基础上(基于LNP中脂质的100%总摩尔数)约0.5至约10%、约0.5至约5%、约10%、约5%、约3.5%、约1.5%、约0.5%、或约0.3%。

[0661] 具有在摩尔基础上(基于脂质纳米颗粒中的脂质总摩尔数)不同摩尔比的阳离子脂质、非阳离子(或中性)脂质、固醇(例如,胆固醇)、和聚集减少剂(如PEG修饰脂质)的不同LNP显示在下表5中。在优选的实施方式中,本发明的脂质纳米颗粒制剂主要由如下摩尔比的脂质混合物组成:约20-70%阳离子脂质:5-45%中性脂质:20-55%胆固醇,0.5-15%PEG修饰脂质,更优选如下摩尔比:约20-60%阳离子脂质:5-25%中性脂质:25-55%胆固醇:0.5-15%PEG修饰脂质。

[0662] 表5:脂质系制剂

#	脂质的摩尔比 (基于脂质纳米颗粒中脂质的 100%总摩尔数)			
	阳离子脂质	非阳离子(或中性)脂质	固醇	聚集减少剂(例如, PEG-脂质)
1	约 35%至约 65 %	约 3%至约 12%或 15 %	约 15%至约 45 %	约 0.1%至约 10% (优选约 0.5%至约 2%或 3%)
2	约 20%至约 70%	约 5%至约 45%	约 20%至约 55%	约 0.1%至约 10% (优选约 0.5%至约 2%或 3%)
3	约 45%至约 65%	约 5%至约 10%	约 5%至约 45%	约 0.1%至约 3%
4	约 20%至约 60%	约 5%至约 25%	约 25%至约 40%	约 0.1%至约 5% (优选约 0.1%至约 3%)
5	约 40%	约 10%	约 25%至约 55%	约 10%
6	约 35%	约 15%		约 10%
7	约 52%	约 13%		约 5%
8	约 50%	约 10%		约 1.5%

[0664] 在一些实施方式中,LNP以脂质体或脂质复合物存在,如下文进一步详细描述。

[0665] LNP尺寸

[0666] 在一些实施方式中,LNP的中值粒径尺寸为约50nm至约300nm,如约50nm至约250nm,例如约50nm至约200nm。

[0667] 在一些实施方式中,可使用较小的LNP。这种颗粒可包括从0.1 $\mu$ m以下上至100nm的直径,如但不限于,小于0.1 $\mu$ m、小于1.0 $\mu$ m、小于5 $\mu$ m、小于10 $\mu$ m、小于15 $\mu$ m、小于20 $\mu$ m、小于25 $\mu$ m、小于30 $\mu$ m、小于35 $\mu$ m、小于40 $\mu$ m、小于50 $\mu$ m、小于55 $\mu$ m、小于60 $\mu$ m、小于65 $\mu$ m、小于70 $\mu$ m、小于75 $\mu$ m、小于80 $\mu$ m、小于85 $\mu$ m、小于90 $\mu$ m、小于95 $\mu$ m、小于100 $\mu$ m、小于125 $\mu$ m、小于150 $\mu$ m、小于175 $\mu$ m、小于200 $\mu$ m、小于225 $\mu$ m、小于250 $\mu$ m、小于275 $\mu$ m、小于300 $\mu$ m、小于325 $\mu$ m、小于350 $\mu$ m、小于375 $\mu$ m、小于400 $\mu$ m、小于425 $\mu$ m、小于450 $\mu$ m、小于475 $\mu$ m、小于500 $\mu$ m、小于525 $\mu$ m、小于550 $\mu$ m、小于575 $\mu$ m、小于600 $\mu$ m、小于625 $\mu$ m、小于650 $\mu$ m、小于675 $\mu$ m、小于700 $\mu$ m、小于725 $\mu$ m、小于750 $\mu$ m、小于775 $\mu$ m、小于800 $\mu$ m、小于825 $\mu$ m、小于850 $\mu$ m、小于875 $\mu$ m、小于900 $\mu$ m、小于925 $\mu$ m、小于950 $\mu$ m、小于975 $\mu$ m。在另一实施方式中,核酸可利用较小的LNP来递送,该较小的LNP可包括如下直径:约1nm至约100nm、约1nm至约10nm、约1nm至约20nm、约1nm至约30nm、约1nm至约40nm、约1nm至约50nm、约1nm至约60nm、约1nm至约70nm、约1nm至约80nm、约1nm至约90nm、约5nm至约100nm、约5nm至约10nm、约5nm至约20nm、约5nm至约30nm、约5nm至约40nm、约5nm至约50nm、约5nm至约60nm、约5nm至约70nm、约5nm至约80nm、约5nm至约90nm、约10至约50nm、约20至约50nm、约30至约50nm、约40至约50nm、约20至约60nm、约30至约60nm、约40至约60nm、约20至约70nm、约30至约70nm、约40至约70nm、约50至约70nm、约60至约70nm、约20至约80nm、约30至约80nm、约40至约80nm、约50至约80nm、约60至约80nm、约20至约90nm、约30至约90nm、约40至约90nm、约50至约90nm、约60至约90nm和/或约70至约90nm。

[0668] 在一些实施方式中,LNP可具有大于100nm、大于150nm、大于200nm、大于250nm、大于300nm、大于350nm、大于400nm、大于450nm、大于500nm、大于550nm、大于600nm、大于650nm、大于700nm、大于750nm、大于800nm、大于850nm、大于900nm、大于950nm或大于1000nm的直径。

[0669] 在其他实施方式中,LNP具有单峰颗粒尺寸分布(即,其不是双峰或多峰型)。

[0670] 其他组分

[0671] 除上述那些外,LNP可进一步包括一种或多种脂质和/或其他组分。

[0672] 脂质体组合物中可包含其它脂质,用于各种目的,如防止脂质氧化或将配体附接到脂质体表面上。LNP中可存在多种脂质中的任意种,包括两性、中性、阳离子和阴离子脂质。这种脂质可以单独使用或组合使用。

[0673] 可存在于LNP中的其他组分包括双层稳定话组分,如聚酰胺寡聚物(参见,例如,美国专利号6,320,017,其整体通过引用并入)、肽、蛋白质和洗涤剂。

[0674] 脂质体

[0675] 在一些实施方式中,本发明组合物的人工核酸分子(优选地RNA)(或本文限定的任何其他核酸)被配制为脂质体。

[0676] 阳离子脂质系脂质体能够通过静电相互作用与带负电荷的核酸(例如RNA)复合,产生复合物,该复合物具有生物相容性、低毒性和体内临床应用所需的大规模生产的可能性。脂质体可以与质膜融合以便摄取;在处于细胞内后,脂质体就通过内吞途径被加工,然后核酸从内体/载体释放到细胞质中。脂质体由于其优异的生物相容性而长期以来被认为是药物递送媒介,因为脂质体基本上是生物膜的类似物,并且可以由天然和合成磷脂制备(Int J Nanomedicine.2014;9:1833-1843)。

[0677] 脂质体一般由脂质双层组成——该脂质双层可以由阳离子、阴离子或中性(磷酸)脂质和胆固醇构成,该脂质双层包封水性核心。脂质双层和含水空间分别均可包含疏水或亲水化合物。脂质体可具有一种或多种脂质膜。脂质体可以是单层的,称为单层式,或多层的,称为多层式。

[0678] 脂质体特征和体内表现可以通过向脂质体表面添加亲水性聚合物涂层(例如,聚乙二醇(PEG))以赋予空间稳定作用来改变。此外,脂质体可以用于特异性靶向——通过将配体(例如,抗体、肽和碳水化合物)附接到其表面或附接PEG链的末端而(Front Pharmacol.2015年12月1日;6:286)。

[0679] 脂质体一般作为球状媒介存在,并且尺寸可以在20nm至几微米的范围内。

[0680] 脂质体可以具有不同的尺寸,如但不限于多层囊泡(MLV)——其可以直径数百纳米并且可以包含被狭窄水性隔室分隔的一系列同心双层,小的单细胞囊泡(SUV)——其可以直径小于50nm,和大单层囊泡(LUV)——其可以直径在50和500nm之间。脂质体设计可以包括但不限于调理素或配体——以改善脂质体与不健康组织的附接或激活事件如但不限于内吞作用。脂质体可含有低pH或高pH,以改善药物制剂的递送。

[0681] 作为非限制性实例,脂质体如合成膜囊泡可通过美国专利公开号US20130177638、US20130177637、US20130177636、US20130177635、US20130177634、US20130177633、US20130183375、US20130183373和US20130183372(其内容均整体通过引用并入本文)描述的方法、设备和装置来制备。本发明的人工核酸分子(优选地RNA)、(药物)组合物或试剂盒(或本文限定的任何其他核酸,具体地RNA)可被脂质体包封,和/或其可被包含在水性核心中,该水性核心可然后被脂质体包封(参见国际公开号W02012031046、W02012031043、W02012030901和W02012006378和美国专利公开号US20130189351、US20130195969和US20130202684;其内容均整体通过引用并入本文)。

[0682] 在一些实施方式中,本发明的人工核酸分子(优选地RNA)(或本文限定的任何其他核酸)可被配制在脂质体中,如但不限于DiLa2脂质体(Marina Biotech,Bothell,WA)、**SMARTICLES®**(Marina Biotech,Bothell,WA)、中性DOPC(1,2-二油酰基-sn-甘油基-3-磷酸胆碱)系脂质体(例如,siRNA递送,用于卵巢癌(Landen et al.Cancer Biology&Therapy 2006 5(12)1708-1713);其整体通过引用并入本文)和透明质酸(hyaluronan)涂覆的脂质体(Quiet Therapeutics,以色列)。

[0683] 脂质复合物

[0684] 在一些实施方式中,人工核酸分子(优选地RNA)(或本文限定的任何其他核酸)被配制为脂质复合物,即夹在核酸层之间的阳离子脂质双层。

[0685] 阳离子脂质,如DOTAP(1,2-二油酰基-3-三甲基铵-丙烷)和DOTMA(N-[1-(2,3-二油酰氧基)丙基]-N,N,N-三甲基-甲基硫酸铵)可与带负电荷的核酸形成复合物或脂质复合物,以通过静电相互作用形成纳米颗粒,提供高体外转染效率。

[0686] 纳米脂质体

[0687] 在一些实施方式中,人工核酸分子(优选地RNA)(或本文限定的任何其他核酸)被配制为中性脂质系纳米脂质体,如1,2-二油酰基-sn-甘油基-3-磷脂酰胆碱(DOPC)系纳米脂质体(Adv Drug Deliv Rev.2014Feb;66:110-116.)。

[0688] 乳液

[0689] 在一些实施方式中,人工核酸分子(优选地RNA)(或本文限定的任何其他核酸)被配制为乳液。在另一实施方式中,所述人工核酸分子(优选地RNA)被配制在阳离子水包油乳液中,其中乳液颗粒包括油核和阳离子脂质,该阳离子脂质可与核酸(一种或多种)相互作用,将该分子锚定至乳液颗粒(参见国际公开号W02012006380;其整体通过引用并入本文)。在一些实施方式中,所述人工核酸分子(优选地RNA)被配制在油包水乳液中,该油包水乳液包括连续的疏水相,其中分散亲水相。作为非限制性实例,乳液可通过国际公开号W0201087791(其内容整体通过引用并入本文)描述的方法制备。

[0690] (聚)阳离子化合物和载体

[0691] 在优选的实施方式中,本发明的人工核酸分子(优选地RNA)(或本文限定的任何其他核酸)与阳离子或聚阳离子化合物(“(聚)阳离子化合物”)和/或聚合物载体复合或缔合。

[0692] 术语“(聚)阳离子化合物”一般指代在pH值一般为1至9时,优选地pH值为9或9以下(例如5至9),8或8以下(例如5至8),7或7以下(例如5至7),最优选生理pH,例如7.3至7.4时带正电荷(阳离子)的带电荷分子。

[0693] 因此,“(聚)阳离子化合物”可以是任何带正电荷的化合物或聚合物,优选阳离子肽或蛋白质——其在生理条件下,具体地在体内生理条件下带正电荷。“(聚)阳离子肽或蛋白质”可包含至少一种带正电荷的氨基酸、或多于一种带正电荷的氨基酸,其例如选自Arg、His、Lys或Orn。

[0694] (聚)阳离子氨基酸、肽和蛋白质

[0695] 作为对本发明的人工核酸分子(优选RNA)(或本文限定的任何其他核酸)的复合或缔合特别优选的试剂的(聚)阳离子化合物包括鱼精蛋白、核苷、精胺或亚精胺、或其他阳离子肽或蛋白质,如聚L-赖氨酸(PLL)、聚精氨酸、碱性多肽、细胞穿透肽(CPPs),包括HIV结合肽、HIV-1 Tat(HIV)、Tat衍生肽、Penetratin、VP22衍生或类似肽,HSV VP22(单纯疱疹)、

MAP、KALA或蛋白转导结构域(PTD)、PpT620、富脯氨酸肽、富精氨酸肽、富赖氨酸肽、MPG-肽(一种或多种)、Pep-1、L-寡聚体、降钙素肽(一种或多种)、触角足衍生肽(具体地来自触角果蝇(*Drosophila antennapedia*))、pAntp、pIsl、FGF、乳铁蛋白、Transportan、Buforin-2、Bac715-24、SynB、SynB(1)、pVEC、hCT衍生肽、SAP或组蛋白。

[0696] 更优选,本发明的人工核酸分子(优选地RNA)(或本文限定的任何其他核酸)与一种或多种聚阳离子,优选与鱼精蛋白或oligofectamine(下文讨论),最优选与鱼精蛋白复合。在此环境中,鱼精蛋白是特别优选的。

[0697] 另外,优选的(聚)阳离子蛋白质或肽可选自具有下列总式(III)的下列蛋白质或肽:

[0698]  $(\text{Arg})_1; (\text{Lys})_m; (\text{His})_n; (\text{Orn})_o; (\text{Xaa})_x$ , (式(III))

[0699] 其中 $1+m+n+o+x=8-15$ ,并且 $1, m, n$ 或 $o$ 彼此独立地可以是选自0、1、2、3、4、5、6、7、8、9、10、11、12、13、14或15的任何数量,条件是Arg、Lys、His和Orn的总体含量占寡肽所有氨基酸的至少50%;并且Xaa可以是选自除Arg、Lys、His或Orn外的天然(=天然存在的)或非天然氨基酸的任何氨基酸;并且 $x$ 可以是选自0、1、2、3或4的任何数量,条件是Xaa的总体含量不超过寡肽所有氨基酸的50%。在此环境中特别优选的阳离子肽是例如Arg<sub>7</sub>、Arg<sub>8</sub>、Arg<sub>9</sub>、H<sub>3</sub>R<sub>9</sub>、R<sub>9</sub>H<sub>3</sub>、H<sub>3</sub>R<sub>9</sub>H<sub>3</sub>、YSSR<sub>9</sub>SSY、(RKH)<sub>4</sub>、Y(RKH)<sub>2</sub>R、等。在此环境中,WO 2009/030481的公开内容通过引用并入本文。

[0700] (聚)阳离子多糖

[0701] 用于与本发明的人工核酸分子(优选地RNA)(或本文限定的任何其他核酸)复合或缔合的进一步优选的(聚)阳离子化合物包括(聚)阳离子多糖,例如壳聚糖、聚凝胺、阳离子聚合物,例如聚乙烯亚胺(PEI)。

[0702] (聚)阳离子脂质

[0703] 用于与本发明的人工核酸分子(优选地RNA)(或本文限定的任何其他核酸)复合或缔合的进一步优选的(聚)阳离子化合物包括(聚)阳离子脂质,例如DOTMA:[1-(2,3-二油基氧基)丙基]-N,N,N-三甲基氯化铵、DMRIE、二-C14-脘、DOTIM、SAINT、DC-Chol、BGTC、CTAP、DOPC、DODAP、DOPE:二油基磷脂酰乙醇-胺、DOSPA、DODAB、DOIC、DMEPC、DOGS:双十八烷基酰氨基甘氨酸基精胺、DIMRI:二肉豆蔻基-氧丙基二甲基羟基乙基溴化铵、DOTAP:二油酰氧基-3-(三甲基铵)丙烷、DC-6-14:0,0-二十四烷酰基-N-( $\alpha$ -三甲基铵乙酰基)二乙醇胺氯、CLIP1:rac-[2(2,3-双十八烷氧基丙基)(2-羟基乙基)]-二甲基氯化铵、CLIP6:rac-[2(2,3-双十六烷氧基丙基-氧甲氧基)乙基]三甲基铵、CLIP9:rac-[2(2,3-双十六烷氧基丙基-氧琥珀酰氧基)乙基]-三甲基铵、或oligofectamine。

[0704] (聚)阳离子聚合物

[0705] 用于与本发明的人工核酸分子(优选地RNA)(或本文限定的任何其他核酸)复合或缔合的进一步优选的(聚)阳离子化合物包括(聚)阳离子聚合物,例如修饰的聚氨基酸如 $\beta$ -氨基酸-聚合物或反转聚酰胺等、修饰的聚乙烯如PVP(聚(N-乙基-4-乙烯基溴化吡啶))等、修饰的丙烯酸酯如pDMAEMA(聚(二甲基氨基乙基甲基丙烯酸酯))等、修饰的酰氨基胺如pAMAM(聚(酰氨基胺))等、修饰的聚 $\beta$ 氨基酯(PBAE)如二胺端修饰的1,4丁二醇二丙烯酸酯-共-5-氨基-1-戊醇聚合物等、树枝状聚合物如聚丙基胺树枝状聚合物或pAMAM系树枝状聚合物等、聚亚胺(一种或多种)如PEI:聚(乙烯亚胺)、聚(丙烯亚胺)等、聚烯丙基胺、糖骨架

系聚合物如环糊精系聚合物、葡聚糖系聚合物、壳聚糖等、硅烷骨架系聚合物如PMOXA-PDMS共聚物等、或由一种或多种阳离子嵌段(例如选自上述阳离子聚合物)的组合和一种或多种亲水或疏水嵌段(例如聚乙二醇)的组合组成的嵌段聚合物。

[0706] 聚合物载体(运载体, carrier)

[0707] 根据优选的实施方式, 本发明的人工核酸分子(优选地RNA)((或本文限定的任何其他核酸)与聚合物载体复合或缔合。

[0708] 根据本发明使用的“聚合物载体”可以是由二硫化物交联的阳离子组分形成的聚合物载体。二硫化物交联的阳离子组分可彼此相同或不同。聚合物载体还可包含进一步的组分。

[0709] 还特别优选根据本发明使用的聚合物载体包括阳离子肽、蛋白质或聚合物和任选地本文限定的其他组分的混合物——其如本文所述通过二硫键交联。在此环境下, WO 2012/013326的公开内容通过引用并入本文。

[0710] 在此环境下, 通过二硫化物交联构成聚合物载体基础的阳离子组分一般选自适于此目的的任何适当(聚)阳离子肽、蛋白质或聚合物, 具体地能够复合和从而优选地稠合本发明的人工核酸分子(优选地RNA)(或本文限定的任何其他核酸)的任何(聚)阳离子肽、蛋白质或聚合物。(聚)阳离子肽、蛋白质或聚合物优选是线性分子, 然而也可使用分支(聚)阳离子肽、蛋白质或聚合物。

[0711] 可用于复合人工核酸分子(优选地RNA)(或本文限定的任何其他核酸)的聚合物载体的各二硫化物交联(聚)阳离子蛋白质、肽或聚合物包含至少一种-SH部分, 最优选至少一种半胱氨酸残基或呈现-SH部分的在与至少一种进一步(聚)阳离子蛋白质、肽或聚合物作为本文所述聚合物载体的阳离子组分稠合后能够形成二硫化物连接的任何其他化学基团。

[0712] 如上限定, 可用于复合本发明的人工核酸分子(优选地RNA)(或本文限定的任何其他核酸)的聚合物载体可由二硫化物交联的阳离子(或聚阳离子)组分形成。优选地, 包括或另外修饰以包括至少一种-SH部分的聚合物载体的这种(聚)阳离子肽或蛋白质或聚合物选自本文限定的蛋白质、肽和聚合物。

[0713] 在一些实施方式中, 聚合物载体可选自根据通式(IV)的聚合物载体分子:

[0714]  $L-P^1-S-[S-P^2-S]_n-S-P^3-L$  式(IV)

[0715] 其中,

[0716]  $P^1$ 和 $P^3$ 彼此不同或相同并且表示线性或分支亲水聚合物链,  $P^1$ 和 $P^3$ 各呈现至少一种-SH-部分, 在与组分 $P^2$ 或与可选地(AA)、 $(AA)_x$ 或 $[(AA)_x]_z$ (如果这种组分用作 $P^1$ 和 $P^2$ 或 $P^3$ 和 $P^2$ 之间的连接体)和/或其他组分(例如(AA)、 $(AA)_x$ 、 $[(AA)_x]_z$ 或L)稠合后能够形成二硫化物连接, 该线性或分支亲水聚合物链彼此独立地选自聚乙二醇(PEG)、聚-N-(2-羟基丙基)甲基丙烯酰胺、聚-2-(甲基丙烯酰氧基)乙基磷酸胆碱、聚(羟基烷基-L-天冬酰胺)、聚(2-(甲基丙烯酰氧基)乙基磷酸胆碱)、羟基乙基淀粉或聚(羟基烷基-L-谷氨酰胺), 其中亲水聚合物链呈现约1kDa至约100kDa, 优选地约2kDa至约25kDa; 或更优选约2kDa至约10kDa, 例如约5kDa至约25kDa或5kDa至约10kDa的分子量;

[0717]  $P^2$ 是如下(聚)阳离子肽或蛋白质: 例如如上限定用于由二硫化物交联的阳离子组分形成的聚合物载体的, 并且优选具有约3至约100个氨基酸的长度, 更优选具有约3至约50个氨基酸的长度, 甚至更优选具有约3至约25个氨基酸的长度, 例如约3至10个、5至15个、10

至20个或15至25个氨基酸的长度,更优选长度约5至约20个和甚至更优选长度约10至约20个;或

[0718] 是如下(聚)阳离子聚合物:例如如上限定用于由二硫化物交联的阳离子组分形成的聚合物载体的,一般具有约0.5kDa至约30kDa的分子量,包括约1kDa至约20kDa,甚至更优选约1.5kDa至约10kDa的分子量,或具有约0.5kDa至约100kDa的分子量,包括约10kDa至约50kDa,甚至更优选约10kDa至约30kDa的分子量;

[0719] 各 $P^2$ 呈现至少两个-SH-部分,能够在与其他组分 $P^2$ 或组分(一种或多种) $P^1$ 和/或 $P^3$ 或可选地与其他组分(例如(AA)、 $(AA)_x$ 、或 $[(AA)_x]_z$ )稠合后形成二硫化物连接;

[0720] -S-S-是(可逆的)二硫键(为更易理解省略括号),其中S优选地表示形成(可逆的)二硫键的硫或-SH携带部分。(可逆的)二硫键优选地通过组分 $P^1$ 和 $P^2$ 、 $P^2$ 和 $P^2$ 、或 $P^2$ 和 $P^3$ 或任选地本文限定的其他组分(例如L、(AA)、 $(AA)_x$ 、 $[(AA)_x]_z$ 等)的-SH-部分的稠合而形成;-SH-部分可以是这些组分的部分结构或通过下文限定的修饰而被添加;

[0721] L是任选的配体,其可以存在或不存在,并且可以彼此独立地选自RGD、转铁蛋白、叶酸基、信号肽或信号序列、定位信号或序列、核定位信号或序列(NLS)、抗体、细胞穿透肽(例如TAT或KALA)、受体配体(例如细胞因子、激素、生长因子等)、小分子(例如碳水化合物,如甘露糖或半乳糖或合成配体)、小分子激动剂、受体抑制剂或拮抗剂(例如RGD肽模拟类似物)、或本文限定的任何其他蛋白质等;

[0722] n是整数,一般选自约1至50的范围,优选约1、2或3至30的范围,更优选约1、2、3、4、或5至25的范围、或约1、2、3、4、或5至20的范围、或约1、2、3、4、或5至15的范围、或约1、2、3、4、或5至10的范围,包括例如约4至9、4至10、3至20、4至20、5至20、或10至20的范围、或约3至15、4至15、5至15、或10至15的范围、或约6至11或7至10的范围。最优选地,n处于约1、2、3、4、或5至10的范围,更优选约1、2、3、或4至9的范围、约1、2、3、或4至8的范围、或约1、2、或3至7的范围。

[0723] 在此环境下,W0 2011/026641的公开内容通过引用并入本文。各亲水聚合物 $P^1$ 和 $P^3$ 一般呈现至少一种-SH-部分,其中所述至少一种-SH-部分能够在与组分 $P^2$ 或组分(AA)或 $(AA)_x$ (如果如下文限定用作 $P^1$ 和 $P^2$ 或 $P^3$ 和 $P^2$ 之间的连接体)和任选地与其他组分例如L和/或(AA)或 $(AA)_x$ (例如如果包含两个或更多个-SH-部分)反应后形成二硫化物连接。上述通式(IV)中的如下子式“ $P^1-S-S-P^2$ ”和“ $P^2-S-S-P^3$ ”(为更易理解省略括号)——其中S、 $P^1$ 和 $P^3$ 中的任一者均如本文限定——一般表示如下情形:其中上述通式(IV)的亲水聚合物 $P^1$ 和 $P^3$ 的一个-SH-部分与组分 $P^2$ 的一个-SH-部分稠合,其中这些-SH-部分的硫形成式(IV)中如本文限定的二硫键-S-S-。这些-SH-部分一般由各亲水聚合物 $P^1$ 和 $P^3$ 提供,例如通过内部半胱氨酸或携带-SH部分的任何进一步(修饰的)氨基酸或化合物。因此,子式“ $P^1-S-S-P^2$ ”和“ $P^2-S-S-P^3$ ”还可写作“ $P^1-Cys-Cys-P^2$ ”和“ $P^2-Cys-Cys-P^3$ ”——如果-SH-部分由半胱氨酸提供,其中术语Cys-Cys表示通过二硫键而非肽键偶联的两个半胱氨酸。在这种情况下,这些式中的术语“-S-S-”还可写作“-S-Cys”、“-Cys-S”或“-Cys-Cys-”。在此环境下,术语“-Cys-Cys-”不表示肽键,而是两个半胱氨酸通过其-SH-部分连接以形成二硫键。因此,术语“-Cys-Cys-”还可总体上被理解为“-(Cys-S)-(S-Cys)-”,其中在这种具体情况下,S表示半胱氨酸的-SH-部分的硫。同样,术语“-S-Cys”和“-Cys-S”表示含-SH部分和半胱氨酸之间的二硫键,其也可写作“-S-(S-Cys)”和“-(Cys-S)-S”。可选地,亲水聚合物 $P^1$ 和 $P^3$ 可被-SH部分修

饰,优选地通过与携带-SH部分的化合物化学反应,使得各亲水聚合物 $P^1$ 和 $P^3$ 携带至少一种这种-SH部分。携带-SH部分的这种化合物可以是例如(另外的)半胱氨酸或携带-SH部分的任何其他(修饰)氨基酸。这种化合物还可以是包含-SH部分或允许向本文限定的亲水聚合物 $P^1$ 和 $P^3$ 中引入-SH部分的任何非氨基化合物或部分。这种非氨基化合物可通过化学反应或化合物结合而附接至根据本发明的聚合物载体的式(IV)的亲水聚合物 $P^1$ 和 $P^3$ ,例如通过3-硫代丙酸或硫代thioimolane结合,通过酰胺形成(例如羧酸、磺酸、胺等),通过迈克尔加成(例如马来酰亚胺部分、 $\alpha,\beta$ -不饱和羰基等),通过点击化学(例如叠氮化物或炔烃)、烯烃/炔烃变位(methatesis)(例如烯烃或炔烃)、亚胺或腙形成(醛或酮、肼、羟基胺、胺)、复合反应(亲和素、生物素、蛋白G)或允许 $S_n$ 型取代反应的组分(例如卤代烷、硫醇、醇、胺、肼、酰肼、磺酸酯、氧磷盐)或在其他组分附接中可利用的其他化学部分。在此环境中特别优选的PEG衍生物是 $\alpha$ -甲氧基- $\omega$ -巯基聚(乙二醇)。在每种情况下,SH-部分,例如半胱氨酸或任何其他(修饰)氨基酸或化合物的SH-部分,可存在于亲水聚合物 $P^1$ 和 $P^3$ 的末端或内部任何位置。如本文限定,各亲水聚合物 $P^1$ 和 $P^3$ 一般呈现至少一种-SH部分——优选在一端,但也可包含两个或甚至更多个-SH部分,该-SH部分可用于另外附接本文限定的其他组分,优选地进一步功能性肽或蛋白质,例如配体、氨基酸组分(AA)或(AA)<sub>x</sub>、抗体、细胞穿透肽或增强肽(例如TAT、KALA)等。

[0724] 重量比和N/P比

[0725] 在本发明的一些实施方式中,人工核酸分子(优选地RNA)(或所述其他核酸)is与(聚)阳离子化合物或聚合物载体缔合或复合——任选地以选自以下范围的重量比:约6:1(w/w)至约0.25:1(w/w),更优选约5:1(w/w)至约0.5:1(w/w),甚至更优选约4:1(w/w)至约1:1(w/w)、或约3:1(w/w)至约1:1(w/w),和最优选约3:1(w/w)至约2:1(w/w)的核酸:(聚)阳离子化合物和/或聚合物载体的比例;或任选地以以下范围的核酸:(聚)阳离子化合物和/或聚合物载体的氮/磷酸基(N/P)比:约0.1-10的范围,优选约0.3-4或0.3-1的范围,和最优选约0.5-1或0.7-1的范围,甚至最优选约0.3-0.9或0.5-0.9的范围。更优选,至少一种人工核酸分子(优选地RNA):一种或多种聚阳离子的N/P比在约0.1至10的范围内,包括约0.3至4、约0.5至2、约0.7至2和约0.7至1.5的范围。

[0726] 本发明的人工核酸分子(优选地RNA)(或本文限定的任何其他核酸)还可与媒介、转染或复合剂缔合以提高所述人工核酸分子(优选地RNA)的转染效率。

[0727] 在此环境下,特别优选本发明的(药物)组合物包括至少部分与(聚)阳离子化合物和/或聚合物载体(优选阳离子蛋白质或肽)复合的人工核酸分子(优选地RNA)。在此环境下,WO 2010/037539和WO 2012/113513的公开内容通过引用并入本文。“部分”意为仅部分所述人工核酸分子(优选地RNA)与(聚)阳离子化合物和/或聚合物载体复合,而其余所述人工核酸分子(优选地RNA)以非复合形式(“游离”)存在。

[0728] 优选地,复合人工核酸分子(优选地RNA)与游离人工核酸分子(优选地RNA)的摩尔比选自以下摩尔比:约0.001:1至约1:0.001,包括约1:1的比例。更优选地,复合人工核酸分子(优选地RNA)与游离人工核酸分子(优选地RNA)的比例选自约5:1(w/w)至约1:10(w/w)的范围,更优选约4:1(w/w)至约1:8(w/w)的范围,甚至更优选约3:1(w/w)至约1:5(w/w)或1:3(w/w)的范围,和最优选地,复合人工核酸分子(优选地RNA)与游离人工核酸分子(优选地RNA)的比例选自约1:1(w/w)的比例。



[0729] 本发明的复合人工核酸分子(优选地RNA)优选地根据第一步通过如下制备:将人工核酸分子(优选地RNA)与(聚)阳离子化合物和/或与聚合物载体,优选如本文限定,以具体比例复合以形成稳定复合物。在此环境下,高度优选在复合所述人工核酸分子(优选地RNA)后,无游离(聚)阳离子化合物或聚合物载体或仅可忽略不计地少量的游离(聚)阳离子化合物或聚合物载体保持在复合人工核酸分子(优选地RNA)部分中。因此,复合RNA部分中人工核酸分子(优选地RNA)与(聚)阳离子化合物和/或聚合物载体的比例一般在人工核酸分子(优选地RNA)完全复合并且无游离(聚)阳离子化合物或聚合物载体或仅可忽略不计地少量的游离(聚)阳离子化合物或聚合物载体保持在所述部分中的范围内选择。

[0730] 优选地,人工核酸分子(优选地RNA)与(聚)阳离子化合物和/或聚合物载体的比例,优选如本文限定,选自如下的范围:约6:1(w/w)至约0.25:1(w/w),更优选约5:1(w/w)至约0.5:1(w/w),甚至更优选约4:1(w/w)至约1:1(w/w)、或约3:1(w/w)至约1:1(w/w),和最优选约3:1(w/w)至约2:1(w/w)的比例。

[0731] 可选地,人工核酸分子(优选地RNA)与(聚)阳离子化合物和/或聚合物载体的比例还可基于整个复合物的氮/磷酸基比(N/P比)来计算。在本发明的环境中,关于复合物中人工核酸分子(优选地RNA)与(聚)阳离子化合物和/或聚合物载体(优选如本文限定)的比例,N/P比优选在约0.1-10的范围内,优选在约0.3-4的范围,和最优选在约0.5-2或0.7-2的范围内,和最优选在约0.7-1.5、0.5-1或0.7-1的范围内,甚至最优选在约0.3-0.9或0.5-0.9的范围内,优选地条件是复合物中的(聚)阳离子化合物是如上限定的(聚)阳离子蛋白质或肽和/或聚合物载体。

[0732] 在其他实施方式中,人工核酸分子(优选地RNA)可以游离或裸形式被提供和使用,而不与任何其他媒介、转染或复合剂缔合。

[0733] 靶向递送

[0734] 在一些实施方式中,(聚)阳离子化合物、载体、脂质体或LNP可被配制用于靶向递送,以到达不同的器官和/或细胞类型。作为非限制性实例,(聚)阳离子化合物、载体、脂质体或LNP可被配制用于靶向递送至肝脏。用于靶向递送的(聚)阳离子化合物、载体、脂质体或LNP可包括但不限于本文所述的(聚)阳离子化合物、载体、脂质体或LNP。本发明的RNA可编码缀合物,例如共价连接至载体或靶向基团的治疗性蛋白质或其片段或变体。

[0735] 靶向基团可以是蛋白质,例如糖蛋白或肽,例如对共配体具有特异性亲和力的分子,或抗体,例如结合特定细胞类型如上皮细胞、角质形成细胞等的抗体。靶向基团还可包括激素和激素受体。其还可以包括非肽物种,如脂质、凝集素、碳水化合物、维生素、辅因子、多价乳糖、多价半乳糖、N-乙酰基-半乳糖胺、N-乙酰基-葡萄糖胺、多价甘露糖、多价岩藻糖或适体。靶向基团可以是能够靶向特定受体的任何配体。

[0736] 在一些实施方式中,人工核酸分子(优选RNA)和任选地包含其的(药物)组合物或试剂盒适于靶向肝脏(内)。这种人工核酸分子(优选RNA)和任选地包含其的(药物)组合物或试剂盒可特别适用于治疗、预防或减轻代谢疾病,例如由先天性遗传错误引起的代谢疾病,例如鸟氨酸转氨甲酰酶缺乏相关的疾病。

[0737] (药物)组合物

[0738] 在进一步方面,本发明提供组合物,其包括i根据本发明的人工核酸分子(优选地RNA)和至少一种药学上可接受的载体或赋形剂。根据本发明的组合物优选地作为药物组合

物被提供。

[0739] 人工核酸分子(优选地RNA)可以如本文其他部分所述的“复合”或“游离”形式或其混合物作为(药物)组合物的部分被提供。

[0740] 根据本发明的(药物)组合物可进一步包括本文其他部分限定的至少一种gRNA或提供其的载体。

[0741] 根据本发明的(药物)组合物可进一步包括可用于治疗用人工核酸分子(优选地RNA)或包括其的(药物)组合物进行治疗的疾病或状况的至少一种其他活性剂。

[0742] 药学上可接受的赋形剂和载体

[0743] 优选地,根据本发明的(药物)组合物包括至少一种药学上可接受的载体和/或赋形剂。术语“药学上可接受的”指代与一种或多种活性剂(在此:人工核酸分子(优选地RNA))相容并且不干扰和/或显著降低其药物活性的化合物或用剂。药学上可接受的载体优选地具有充分高的纯度和充分低的毒性,以使其适于给予待治疗对象。

[0744] 赋形剂

[0745] 药学上可接受的赋形剂可呈现不同的功能作用,并且无限制地包括稀释剂、填充剂、增量剂、载体、崩解剂、粘合剂、润滑剂、助流剂、涂料、溶剂和共溶剂、缓冲剂、防腐剂、佐剂、抗氧化剂、润湿剂、消泡剂、增稠剂、甜味剂、调味剂和湿润剂。

[0746] 对于液体形式的(药物)组合物,有用的药学上可接受的赋形剂总体上包括溶剂、稀释剂或载体,如(无热原)水、(等渗)盐溶液,如磷酸盐或柠檬酸盐缓冲盐水、固定油、植物油,如例如,花生油、棉籽油、芝麻油、橄榄油、玉米油、乙醇、多元醇(例如,甘油、丙二醇、聚乙二醇等);卵磷脂;表面活性剂;防腐剂如苯甲醇、对羟基苯甲酸酯、氯丁醇、苯酚、抗坏血酸、硫柳汞等;等渗剂,如糖、多元醇如甘露醇、山梨糖醇或氯化钠;单硬脂酸铝或明胶;抗氧化剂,如抗坏血酸或亚硫酸氢钠;螯合剂如乙二胺四乙酸(EDTA);缓冲剂,如乙酸盐、柠檬酸盐或磷酸盐,和调节渗透压的试剂如氯化钠或右旋糖。pH可以用酸或碱调节,如盐酸或氢氧化钠。缓冲剂可以相对于具体参考介质是高渗的、等渗的或低渗的,即缓冲剂可以具有相对于具体参考介质较高、相同或较低的盐含量,其中优选地可以使用不会因渗透或其他浓度效应导致细胞损伤的上述盐的这种浓度。参考介质是例如存在于“体内”方法中的液体,如血液、淋巴液、胞质液或其他体液,或例如在“体外”方法中可用作参考介质的液体,如常见的缓冲剂或液体。这种常见的缓冲剂或液体是技术人员已知的。特别优选林格-乳酸盐溶液作为液体基质。

[0747] 对于(半)固体形式的(药物)组合物,有用的药学上可接受的赋形剂包括粘合剂,如微晶纤维素、黄耆胶或明胶;淀粉或乳糖;糖,如例如乳糖、葡萄糖和蔗糖;淀粉,如玉米淀粉或马铃薯淀粉;纤维素及其衍生物,如例如羧甲基纤维素钠、乙基纤维素、乙酸钠纤维素;崩解剂,如海藻酸等;润滑剂,如硬脂酸镁;助流剂,如硬脂酸、硬脂酸镁;硫酸钙、胶体二氧化硅等;甜味剂,如蔗糖或糖精;和/或调味剂,如薄荷,水杨酸甲酯或橙子调味剂。

[0748] 载体

[0749] 适当的药学上可接受的载体一般基于(药物)组合物的配方来选择。

[0750] 通过注射和具体地通过静脉内注射给予的液体(药物)组合物应无菌且在制造和储存条件下稳定。这种组合物一般配制成不含热原,具有合适的pH,等渗并且保持活性成分(一种或多种)稳定性的肠胃外可接受的水溶液。

[0751] 用于根据本发明的液体(药物)组合物的特别有用的药学上可接受的载体包括水、一般是无热原的水;等渗盐水或缓冲(水)溶液,例如磷酸盐、柠檬酸盐等缓冲溶液。具体地对于本发明(药物)组合物的注射而言,可以使用水或优选缓冲剂,更优选含水性缓冲剂——含有钠盐(优选至少50mM的钠盐)、钙盐(优选至少0.01mM的钙盐)、和任选的钾盐(优选至少3mM的钾盐)。

[0752] 根据优选的实施方式,钠盐、钙盐和任选地钾盐可以如下形式存在:其卤化物形式,例如氯化物、碘化物或溴化物;其氢氧化物、碳酸盐、碳酸氢盐或硫酸盐等形式。在不限于此的情况下,钠盐的实例包括例如NaCl、NaI、NaBr、Na<sub>2</sub>CO<sub>3</sub>、NaHCO<sub>3</sub>、Na<sub>2</sub>SO<sub>4</sub>,任选的钾盐的实例包括例如KCl、KI、KBr、K<sub>2</sub>CO<sub>3</sub>、KHCO<sub>3</sub>、K<sub>2</sub>SO<sub>4</sub>,和钙盐的实例包括例如CaCl<sub>2</sub>、CaI<sub>2</sub>、CaBr<sub>2</sub>、CaCO<sub>3</sub>、CaSO<sub>4</sub>、Ca(OH)<sub>2</sub>。此外,缓冲剂中可以包含上述阳离子的有机阴离子。

[0753] 根据更优选的实施方式,适于如上限定的注射目的的缓冲剂可以含有选自氯化钠(NaCl)、氯化钙(CaCl<sub>2</sub>)和任选地氯化钾(KCl)的盐,其中除氯离子外还可以存在其他阴离子。CaCl<sub>2</sub>也可以用另一种盐如KCl替代。一般,注射缓冲剂中的盐以至少50mM氯化钠(NaCl)、至少3mM氯化钾(KCl)和至少0.01mM氯化钙(CaCl<sub>2</sub>)的浓度存在。注射缓冲剂可以相对于具体参考介质是高渗的、等渗的或低渗的,即缓冲剂可以相对于具有具体参考介质较高、相同或较低的盐含量,其中优选地,可以使用不会由于渗透或其他浓度效应导致细胞损伤的前述盐的这种浓度。参考介质是例如“体内”方法中存在的液体,如血液、淋巴液、胞质液或其他体液,或者例如在“体外”方法中可用作参考介质的液体,如常见的缓冲剂或液体。这种常见的缓冲剂或液体是技术人员已知的。特别优选林格-乳酸盐溶液作为液体基质。

[0754] 制剂

[0755] 总体上,用于表面给予的(药物)组合物可以配制成乳膏、油膏、凝胶、糊剂或粉末。用于口服给予的(药物)组合物可以配制成片剂、胶囊、液体、粉末或持续释放形式。然而,根据优选的实施方式,本发明的(药物)组合物被胃肠外给予,具体地通过皮内或肌肉注射,并因此配制成用于胃肠外给予的液体或冻干形式,如本文其他部分所述。肠胃外制剂一般储存在小瓶、IV袋、安瓿、药筒或预填充注射器中,并且可以作为注射剂、吸入剂或气溶胶给予,优选注射剂。

[0756] 冻干制剂

[0757] 在进一步优选的实施方式中,(药物)组合物以冻干形式提供。优选地,在给予前,使冻干(药物)组合物在合适的缓冲剂中重构——有利地基于含水载体,例如,林格-乳酸盐溶液(其被优选)、林格溶液、磷酸盐缓冲溶液。在一些实施方式中,根据本发明的(药物)组合物含有至少两种、三种、四种、五种、六种或更多种人工核酸分子,优选RNA,其以冻干形式被单独地提供(任选地与至少一种其他添加剂一起),并且其优选地在其使用前单独地在合适的缓冲液(如林格-乳酸盐溶液)中重构,以便允许单独给予每种所述人工核酸分子,优选RNA。

[0758] 液体制剂

[0759] 在进一步优选的实施方式中,(药物)组合物以盐水或脂质系制剂形式提供。脂质系制剂可包括脂质体、脂质复合物、纳米脂质体和脂质纳米颗粒,其在上文在“复合”标记部分中被描述。

[0760] 试剂盒

[0761] 在进一步方面,本发明涉及包括人工核酸分子(优选地RNA)和/或(药物)组合物的试剂盒或多部分试剂盒。试剂盒可进一步包括至少一种gRNA(或提供其的载体)。

[0762] 上述组分可均能在多部分试剂盒中以药物组合物形式被提供。就此而言,上文关于(药物)组合物提供的定义和说明经适当修改同样适用于多部分试剂盒的个体组分。

[0763] 例如,所述至少一种人工核酸分子(优选RNA)和任选地至少一种gRNA或其编码核酸可以彼此独立地以冻干或液体形式提供——任选地与一种或多种药学上可接受的载体、赋形剂或如上文在药物组合物上下文中所述的其他药剂一起。

[0764] 任选地,多部分试剂盒可包含至少一种如本文在药物组合物环境中限定的其他试剂、抗微生物剂、RNA酶抑制剂、增溶剂等。

[0765] 多部分试剂盒可以是两个或更多个部件的试剂盒,并且一般包括在适当容器中的部件。例如,各容器可以如下形式:是小瓶、长瓶、挤压瓶、罐、密封套管、封袋或小袋、管或发泡包装或任何其它适当形式,条件是容器配置成防止组分过早混合。可以单独提供每个不同的组分,或者可以一起提供一些不同的组分(即,在同一容器中)。

[0766] 容器也可以是小瓶、管、罐或封袋、或套管、或发泡包装或长瓶内的隔室或腔室,条件是在药剂师或医生故意混合之前一个隔室的内容物不能与另一个隔室的内容物物理上相关联。

[0767] 多部分试剂盒还可以包含具有关于其任何组分的给予和剂量的信息的技术说明书。

[0768] 医疗用途和治疗

[0769] 本文限定的人工核酸分子(优选RNA)或(药物)组合物或多部分试剂盒可用于人类和兽医医学目的,优选用于人类医学目的。

[0770] 根据另一方面,本发明因此涉及用作药物的人工核酸分子(优选RNA)或(药物)组合物或多部分试剂盒。

[0771] 人工核酸分子(优选RNA)或(药物)组合物或多部分试剂盒尤其可用于治疗和/或预防适于通过表达被编码的CRISPR相关蛋白质而进行治疗(优选适于通过敲入、敲除、操纵或调控目的基因的表达而进行治疗)的疾病。

[0772] 根据另一方面,本发明因此涉及用于适于通过表达被编码的CRISPR相关蛋白质而进行治疗(优选适于通过敲入、敲除、操纵或调控目的基因的表达而进行治疗)的疾病的基因疗法方法和/或适于通过表达被编码的CRISPR相关蛋白质而进行治疗(优选适于通过敲入、敲除、操纵或调控目的基因的表达而进行治疗)的疾病的的治疗的人工核酸分子(优选RNA)或(药物)组合物或多部分试剂盒。这种疾病可选自遗传疾病、癌症、自身免疫疾病、炎症性疾病、感染性疾病、代谢疾病、神经疾病、心血管疾病或其他疾病或状况。

[0773] “基因疗法”优选涉及调控(即恢复、增强、降低或抑制)对象中的基因表达以实现治疗效果。为此,基因疗法一般包括将核酸引入细胞中。该术语总体上指代基于治疗目的的操纵基因组,并且包括使用基因组编辑技术来校正引起疾病的突变、向基因组添加治疗基因、去除有害基因或基因组序列、以及调控基因表达。基因疗法可涉及宿主细胞的体内或离体转化。

[0774] 术语“治疗”(“treatment”或“treating”)疾病包括预防疾病或抗疾病保护(即,导致临床症状不发展);抑制疾病(即阻止或抑制临床症状的发展;和/或缓解疾病(即导致临

床症状消退)。可以理解,“预防”和“抑制”疾病或障碍并不总是可以区分,因为最终的诱导事件(一种或多种)可能是未知的或潜在的。因此,术语“预防”将被理解为构成一种包括“预防”和“抑制”的“治疗”类型。因此,术语“治疗”包括“预防”。

[0775] 本文使用的术语“对象”、“患者”或“个体”总体上包括人和非人动物,并且优选哺乳动物(例如,非人灵长类动物,包括狨猴、小绢猴、蜘蛛猴、猫头鹰猴、黑长尾猴、松鼠猴、和狒狒、猕猴、黑猩猩、猩猩、大猩猩;牛;马;绵羊;猪;鸡;猫;狗;小鼠;大鼠;兔;豚鼠;等等),包括嵌合和转基因动物和疾病模型。在本发明的环境中,术语“对象”优选是指非人灵长类动物或人,最优选人。

[0776] 因此,本发明进一步提供了治疗适于通过表达被编码的CRISPR相关蛋白质而进行治疗(优选适于通过敲入、敲除、操纵或调控目的基因的表达而进行治疗)的疾病,所述方法通过以下进行:向有此需要的对象给予药学有效量的人工核酸分子(优选RNA)或(药物)组合物或多部分试剂盒。这种方法可包括任选的第一步:制备本发明的人工核酸分子(优选RNA)或(药物)组合物或多部分试剂盒;和第二步,包括向有此需要的患者/对象给予(药学和/或治疗有效量的)所述人工核酸分子(优选RNA)或(药物)组合物或多部分试剂盒。

[0777] 给予优选肠胃外完成,例如通过皮下、肌肉或皮内注射,优选通过肌肉或皮内注射,更优选通过皮内注射。优选地,注射通过使用常规针注射或(无针)射流注射,优选通过使用(无针)射流注射进行。

[0778] 本发明还涉及本发明的人工核酸分子(优选RNA)或(药物)组合物或多部分试剂盒的用途,优选用于敲入、敲除、操纵或调控(优选用于诱导或增强)目的基因的表达。

[0779] 给予途径

[0780] 本发明的人工核酸分子(优选地RNA)或(药物)组合物或多部分试剂盒可被例如全身或局部给予。

[0781] 全身给予途径总体上包括例如透皮、口服、肠胃外途径,包括皮下、静脉内、肌肉、动脉内、皮内和腹膜内注射和/或鼻内给予途径。

[0782] 局部给予途径总体上包括例如表面给予途径,但也包括皮内、透皮、皮下或肌肉注射或病灶内、瘤内、颅内、肺内、心内和舌下注射。

[0783] 进一步可想到针对人工核酸分子(优选地RNA)或(药物)组合物或多部分试剂盒的不同组分采用不同的给予途径,例如在所述(药物)组合物或多部分试剂盒包括几种不同核酸(如至少一种编码CRISPR相关蛋白质的人工核酸和至少一种gRNA或提供其的载体)的情况下。

[0784] 根据优选的实施方式,人工核酸分子(优选地RNA)或(药物)组合物或多部分试剂盒通过胃肠外途径被给予,优选通过皮内、皮下、或肌肉途径。优选地,所述人工核酸分子(优选地RNA)或(药物)组合物或多部分试剂盒通过注射被给予,例如皮下、肌肉或皮内注射,其可以是无针和/或针注射。因此,在优选的实施方式中,根据本发明的医疗用途和/或治疗方法涉及通过皮下、肌肉或皮内注射,优选地通过肌肉或皮内注射,更优选通过皮内注射给予所述人工核酸分子(优选地RNA)或(药物)组合物或多部分试剂盒。这种注射可通过利用常规针注射或(无针)射流注射,优选通过利用(无针)射流注射进行。

[0785] 给予方案

[0786] 本发明的(药物)组合物或多部分试剂盒的组分——即,至少一种人工核酸分子

(优选地RNA)和任选地至少一种其他核酸(例如gRNA或提供其的载体)——可向有此需要的对象给予一天几次、每天一次、隔天一次、每周一次、或每月一次;并且可相继或同时被给予。所述组分可通过如上限定的不同给予途径被给予有此需要的对象。

[0787] 根据一些优选的实施方式,本发明的(药物)组合物的组分被同时给予(即在相同时间通过相同或不同的给予途径)。

[0788] 根据其他优选的实施方式,本发明的(药物)组合物或多部分试剂盒的组分被分开给予(即相继在不同时间点和/或通过不同的给予途径)。这种相继给予方案也称为“错时”给予。错时给予可意为人工核酸分子(优选地RNA)例如先于、同步于或晚于gRNA或提供其的载体被给予,或反之。

[0789] 剂量

[0790] 本发明的人工核酸分子(优选地RNA)(药物)组合物或试剂盒优选以安全并且治疗有效量被给予。

[0791] 如本文所用,“安全并且治疗有效量”意为足以在组织、系统、动物或人中引起所图的期望生物学或医学响应的活性剂(一种或多种)量。安全并且治疗有效量优选足以引起待治疗疾病的积极变化,即所治疗疾病的症状减轻、疾病进展减少、或所预防疾病的症状预防。然而,同时,“安全并且治疗有效量”小到足以避免严重的副作用,即允许益处和风险之间的合理关系。

[0792] “安全并且治疗有效量”还将根据待治疗的具体状况以及待治疗患者的年龄、身体状况、体重、性别和饮食、状况的严重程度、治疗的持续时间、伴随治疗的性质、所用具体药学上可接受的载体或赋形剂、治疗方案和类似因素而变化。

[0793] 此外,可以根据人工核酸分子的类型(优选RNA,例如单顺反子、双顺反子或甚至多顺反子RNA)选择人工核酸分子(优选RNA)的“安全并且(治疗)有效量”,因为双顺反子或甚至多顺反子RNA可导致编码的CRISPR相关蛋白质(一种或多种)的表达显着高于使用等量的单顺反子RNA。

[0794] 本发明的人工核酸分子(优选RNA)、(药物)组合物或多部分试剂盒的治疗效力和毒性通过可在细胞培养物或实验动物中的标准药理学程序测定,例如,用于测定LD<sub>50</sub>(群体50%致死剂量)和ED<sub>50</sub>(群体50%治疗有效剂量)。毒性效果和治疗效果之间的剂量比是治疗指数,并且可以表示为LD<sub>50</sub>/ED<sub>50</sub>比。总体上优选呈现大治疗指数的人工核酸(优选RNA)、(药物)组合物或多部分试剂盒。从细胞培养测定和动物研究获得的数据可用于配制一系列用于人类的剂量。这种化合物的剂量优选居于毒性很小或没有毒性的包括ED<sub>50</sub>在内的循环浓度范围内。

[0795] 例如,本文所述的本发明的人工核酸分子(优选RNA)、(药物)组合物或多部分试剂盒的治疗有效剂量可以在如下范围内:每剂量单位约0.001mg至10mg,优选约0.01mg至5mg,更优选约0.1mg至2mg,或每剂量单位约0.01nmol至1mmol,具体地每剂量单位1nmol至1mmol,优选每剂量单位1μmol至1mmol。还设想本发明的人工核酸分子(优选RNA)、(药物)组合物或多部分试剂盒的治疗有效剂量可在如下范围(每kg体重)内:约0.01mg/kg至10g/kg,优选约0.05mg/kg至5g/kg,更优选约0.1mg/kg至2.5g/kg。

[0796] 待给予的本发明的人工核酸分子(优选RNA)、(药物)组合物或多部分试剂盒的安全并且治疗有效量可通过常规实验确定,例如,通过使用动物模型。这种模型包括(不表示

任何限制)兔、绵羊、小鼠、大鼠、狗和非人灵长类动物模型。

[0797] 疾病

[0798] CRISPR技术可用于多种目的,包括基因的功能性敲除或敲入、基因编辑或转录激活或抑制。因此,根据本发明的人工核酸分子(优选RNA)、(药物)组合物或试剂盒可用于治疗多种疾病。其具体地被设想用于适于通过表达由人工核酸分子的至少一种编码序列编码的CRISPR相关蛋白质进行治疗的疾病、障碍或状况的基因疗法。

[0799] 优选地,待治疗的疾病适于通过敲入、敲除或操纵(例如引入或去除突变)目的基因,或通过调控(即改变、诱导、增加、减少、预防或破坏)进行治疗。

[0800] 根据本发明的人工核酸分子(优选RNA)或包含其的(药物)组合物或试剂盒可用于诱导基因敲除(即,使基因无功能性或从基因组去除基因)。因此,可以利用能够将DSB引入基因组DNA的、编码呈现内切核酸酶活性的CRISPR相关蛋白质的人工核酸分子(优选RNA)。所述DSB可诱导非同源末端接合(NHEJ),导致短段核苷酸随机插入或删除,导致密码子阅读框破坏(移码),导致错误转录物和基因表达消除(功能丧失)。这种策略可以例如有效用于敲除介导肿瘤细胞增殖和存活的基因或用于从宿主细胞的基因组去除整合的病毒(基因)。

[0801] 根据本发明的人工核酸分子(优选RNA)或包含其的(药物)组合物或试剂盒可用于诱导基因敲入(即将新基因或修饰基因引入基因组中)。因此,可以利用编码呈现切口酶活性的CRISPR相关蛋白质的人工核酸分子(优选RNA)其因此能够将缺口(即双链基因组DNA的一条链的磷酸二酯键的水解)引入基因组中。此类切口优选诱导同源定向修复(HDR),导致具有同源性区域的DNA区段并入与DNA双链断裂两侧侧翼的序列。利用HDR,可以将任何期望的DNA序列插入基因组DNA,以引起例如功能丧失、功能获得或功能改变(新形态),或研究未知功能状态的变体。为了利用HDR编辑基因组,一般将具有期望序列的DNA修复模板与本发明的人工核酸(和gRNA)一起提供。这种策略可以例如有效用于敲入治疗基因。

[0802] 根据本发明的人工核酸分子(优选RNA)或包含其的(药物)组合物或试剂盒可用于调控基因表达,具体地基因转录。因此,可以使用编码包含合适效应结构域的CRISPR相关蛋白质衍生物的人工核酸分子(优选RNA)。效应结构域可以与靶基因(或其可操作地连接的调控序列)相互作用以调控其表达。这种方法可有效用于与目标靶基因的不期望的(存在或不存在,增加或减少)表达相关的任何疾病。

[0803] 癌症

[0804] 在优选的实施方式中,人工核酸(优选地RNA)、(药物)组合物或试剂盒用于治疗或预防癌症。

[0805] 如本文所用,术语“癌症”是指瘤,其特征不在于细胞的不受控制且通常快速的增殖,该细胞倾向于侵入周围组织和转移至远处身体部位。该术语包括良性和恶性肿瘤。癌症的恶性一般以间变、侵袭和转移为特征;而良性恶性肿瘤一般没有这些特性。该术语包括以肿瘤生长为特征的瘤以及血液和淋巴系统的癌症。

[0806] 在一些实施方式中,根据本发明的人工核酸(优选RNA)、(药物)组合物或试剂盒可以用作药物,具体地用于治疗肿瘤或癌症疾病。在这种环境下,治疗优选涉及肿瘤内施用,尤其通过肿瘤内注射。因此,根据本发明的人工核酸(优选RNA)、(药物)组合物或试剂盒可用于制备用于治疗肿瘤或癌症疾病的药物,所述药物特别适于肿瘤内施用(给予)以治疗肿瘤或癌症。

[0807] 优选地,本文提及的肿瘤和癌症疾病选自优选包括以下的肿瘤或癌症疾病:例如急性成淋巴细胞性白血病、急性髓性白血病、肾上腺皮质癌、AIDS相关癌症、AIDS相关淋巴瘤、肛门癌、阑尾癌、星形细胞瘤、基底细胞癌、胆管癌、膀胱癌、骨癌、骨肉瘤/恶性纤维组织细胞瘤、脑干胶质瘤、脑肿瘤、小脑星形细胞瘤、脑星形细胞瘤/恶性胶质瘤、室管膜瘤、成神经管细胞瘤、幕上原始神经外胚层肿瘤、视觉通路和下丘脑神经胶质瘤、乳腺癌、支气管腺瘤/类癌、伯基特淋巴瘤、儿童类癌肿瘤、胃肠道类癌肿瘤、原发性未知的癌、原发性中枢神经系统淋巴瘤、儿童小脑星形细胞瘤、儿童脑星形细胞瘤/恶性胶质瘤、宫颈癌、儿童癌症、慢性淋巴细胞性白血病、慢性骨髓性白血病、慢性骨髓增生性障碍、结肠癌、皮肤T细胞淋巴瘤、成纤维细胞性小圆形细胞瘤、子宫内膜癌、室管膜瘤、食道癌、尤文氏肿瘤家族中的尤文氏肉瘤、儿童颅外生殖细胞肿瘤、性腺外生殖细胞肿瘤、肝外胆管癌、眼内黑色素瘤、视网膜母细胞瘤、胆囊癌、胃(胃部)癌、胃肠道类癌肿瘤、胃肠道间质肿瘤(GIST)、颅外、性腺外或卵巢生殖细胞肿瘤、妊娠滋养层肿瘤、脑干胶质瘤、儿童脑星形细胞瘤、儿童视觉通路和下丘脑神经胶质瘤、胃类癌、毛细胞白血病、头颈癌、心脏癌、肝细胞(肝)癌、霍奇金淋巴瘤、下咽癌、儿童下丘脑和视觉通路胶质瘤、眼内黑色素瘤、胰岛细胞癌(内分泌胰腺)、卡波西肉瘤、肾癌(肾细胞癌)、喉癌、白血病、急性成淋巴细胞性白血病、急性髓性白血病、慢性淋巴细胞性白血病、慢性骨髓性白血病、毛细胞白血病、唇和口腔癌、脂肪肉瘤、肝癌、非小细胞肺癌、小细胞肺癌、淋巴瘤、AIDS相关淋巴瘤、伯基特淋巴瘤、皮肤T细胞淋巴瘤、霍奇金淋巴瘤、非霍奇金淋巴瘤、原发性中枢神经系统淋巴瘤、**Waldenström**巨球蛋白血症、骨恶性纤维组织细胞瘤/骨肉瘤、儿童成神经管细胞瘤、黑色素瘤、眼内(眼)黑色素瘤、Merkel细胞癌、成人恶性间皮瘤、儿童间皮瘤、原发性隐匿性转移性鳞状颈癌、口腔癌、儿童多发性内分泌瘤形成综合征、多发性骨髓瘤/浆细胞瘤、蕈样真菌病、骨髓增生异常综合征、骨髓增生异常/骨髓增生性疾病、慢性骨髓性白血病、成人急性髓性白血病、儿童急性髓性白血病、多发性骨髓瘤(骨髓癌)、慢性骨髓增生性障碍、鼻腔和鼻窦癌、鼻咽癌、神经母细胞瘤、口腔癌、口咽癌、骨肉瘤/骨恶性纤维组织细胞瘤、卵巢癌、卵巢上皮癌(表面上皮-间质瘤)、卵巢生殖细胞肿瘤、卵巢低恶性潜能肿瘤、胰腺癌、胰岛细胞胰腺癌、鼻窦和鼻腔癌、甲状旁腺癌、阴茎癌、咽癌、嗜铬细胞瘤、松果体星形细胞瘤、松果体生殖细胞瘤、儿童成松果体细胞瘤和幕上原始神经外胚层肿瘤、垂体腺瘤、浆细胞瘤形成/多发性骨髓瘤、胸膜肺母细胞瘤、原发性中枢神经系统淋巴瘤、前列腺癌、直肠癌、肾细胞癌(肾癌)、肾盂和输尿管癌、视网膜母细胞瘤、儿童横纹肌肉瘤、唾液腺癌、尤文氏肿瘤家族肉瘤、卡波西肉瘤、软组织肉瘤、子宫肉瘤、Sézary综合征、皮肤癌(非黑色素瘤)、皮肤癌(黑色素瘤)、Merkel细胞皮肤癌、小肠癌、鳞状细胞癌、原发性隐匿性转移性鳞状颈癌、儿童幕上原始神经外胚层肿瘤、睾丸癌、咽喉癌、儿童胸腺瘤、胸腺瘤和胸腺癌、甲状腺癌、儿童甲状腺癌、肾盂和输尿管的过渡细胞癌、妊娠滋养层肿瘤、尿道癌、子宫内膜子宫癌、子宫肉瘤、阴道癌、儿童视觉通路和下丘脑神经胶质瘤、外阴癌、**Waldenström**巨球蛋白血症和儿童Wilms瘤(肾癌)。

[0808] 适用于肿瘤内给予的特别优选的肿瘤或癌症的实例是前列腺癌、肺癌、乳腺癌、脑癌、头颈癌、甲状腺癌、结肠癌、胃癌、肝癌、胰腺癌、卵巢癌、皮肤癌、膀胱、子宫和子宫颈(癌)。

[0809] 感染性疾病

[0810] 本发明的组合、药物组合物或试剂盒可用于治疗感染性疾病。术语“感染”或“感染



性疾病”涉及正常不存在于体内的微生物如细菌、病毒和寄生虫的侵入和增殖。感染可不引起任何症状并且是亚临床的,或者其可引起症状并且在临床上显著的。感染可以保持局部化,或者其可以通过血液或淋巴系统传播而变成全身性的。在这种环境下,感染性病优选包括病毒、细菌、真菌或原生动物感染性疾病。

[0811] 本发明的人工核酸,优选RNA,被认为特别有效用于去除整合到宿主细胞基因组中的病毒基因组。通过本发明人工核酸(优选RNA)编码的CRISPR相关蛋白质从宿主细胞中根除病毒优选适用于在其生命周期中具有DNA中间体的任何DNA病毒或RNA病毒。因此,具体地设想根据本发明的人工核酸(优选RNA)、(药物)组合物和试剂盒用于治疗人乳头瘤病毒HPV16和HPV18感染、乙型肝炎病毒(HBV)感染、Epstein-Barr病毒(EBV)、HIV-1感染、疱疹病毒感染(包括卡波西氏肉瘤相关疱疹病毒(KSHV、HHV8)感染)和多瘤病毒感染(包括Merkel细胞癌病毒(MCV)、多瘤病毒JC(JCV)和多瘤病毒BK(BKV)感染及相关感染性疾病。

[0812] 联合疗法

[0813] 本发明的人工核酸分子(优选RNA)、(药物)组合物或多部分试剂盒也可以用于联合疗法。可用于治疗或预防本文限定的疾病和障碍的任何其他疗法可与本文公开的用途和方法组合。

[0814] 例如,接受本发明的人工核酸分子(优选RNA)、(药物)组合物或多部分试剂盒的对象可以是癌症(优选如本文限定)或相关状况的接受化学疗法(例如,一线或二线化学疗法)、放射疗法、放化学疗法(化学疗法和放射疗法联合)、酪氨酸激酶抑制剂(如EGFR酪氨酸激酶抑制剂)、抗体治疗和/或抑制性和/或刺激性检查点分子(如CTLA4抑制剂)的患者,或在已接受一种或多种上述治疗后已达到部分响应或病情稳定的患者。或者,接受本发明的人工核酸分子(优选RNA)、(药物)组合物或多部分试剂盒的对象可以是患有感染性疾病(优选如本文限定)的接受抗生素、抗真菌或抗病毒治疗的患者。

[0815] 在另一方面,本发明因此还涉及本发明的人工核酸分子(优选RNA)、(药物)组合物或多部分试剂盒用于支持癌症、感染性疾病或适于通过用所述人工核酸分子、(药物)组合物或试剂盒进行治疗的任何其他疾病的其他疗法的用途。

[0816] 对癌症治疗或预防的“支持”可以是常规癌症治疗方法如手术、放射疗法、化学疗法(例如一线或二线化学疗法)、放化学疗法、酪氨酸激酶抑制剂治疗、抑制性和/或刺激性检查点分子(优选CTLA4抑制剂)治疗、抗体疗法或这些的任何组合与使用本文限定的本发明人工核酸分子(优选RNA)、(药物)组合物或多部分试剂盒的治疗的任何组合。

[0817] 本发明的人工核酸分子(优选RNA)、(药物)组合物或多部分试剂盒的给予可以在给予另一治疗剂或患者接受可用于治疗具体疾病或状况的另一疗法之前、同时和/或之后完成。

[0818] 项目

[0819] 基于以上,本发明可表征在下列项目:

[0820] 1. 人工核酸分子,包括

[0821] a. 至少一个编码区,编码至少一种CRISPR相关蛋白质;

[0822] b. 至少一个5' 非翻译区(5' UTR)元件,衍生自选自ATP5A1、RPL32、HSD17B4、SLC7A3、NOSIP、ASAH1、RPL31、TUBB4B、UBQLN2、MP68和NDUFA4的基因的5' UTR;和

[0823] c. 至少一个3' 非翻译区(3' UTR)元件,衍生自选自GNAS、CASP1、PSMB3、ALB、

COX6B1、NDUFA1和RPS9的基因的3' UTR。

[0824] 2. 根据项目1所述的人工核酸分子,其中各所述基因包括天然存在的DNA序列及其同源物、变体、片段、和相应RNA序列。

[0825] 3. 根据项目1或2所述的人工核酸分子,包括

[0826] a. 衍生自HSD17B4基因的5' UTR或其相应RNA序列、同源物、片段或变体的至少一个5' UTR元件、和衍生自GNAS基因的3' UTR或其相应RNA序列、同源物、片段或变体的至少一个3' UTR元件;或

[0827] b. 衍生自SLC7A3基因的5' UTR或其相应RNA序列、同源物、片段或变体的至少一个5' UTR元件、和衍生自GNAS基因的3' UTR或其相应RNA序列、同源物、片段或变体的至少一个3' UTR元件;或

[0828] c. 衍生自ATP5A1基因的5' UTR或其相应RNA序列、同源物、片段或变体的至少一个5' UTR元件和衍生自CASP1基因的3' UTR或其相应RNA序列、同源物、片段或变体的至少一个3' UTR元件;或

[0829] d. 衍生自NDUFA4基因的5' UTR或其相应RNA序列、同源物、片段的至少一个5' UTR元件或变体和衍生自PSMB3基因的3' UTR或其相应RNA序列、同源物、片段或变体的至少一个3' UTR元件;或

[0830] e. 衍生自HSD17B4基因的5' UTR或其相应RNA序列、同源物、片段或变体的至少一个5' UTR元件和衍生自PSMB3基因的3' UTR或其相应RNA序列、同源物、片段或变体的至少一个3' UTR元件;或

[0831] f. 衍生自RPL32基因的5' UTR或其相应RNA序列、同源物、片段或变体的至少一个5' UTR元件和衍生自ALB基因的3' UTR或其相应RNA序列、同源物、片段或变体的至少一个3' UTR元件;或

[0832] g. 衍生自HSD17B4基因的5' UTR或其相应RNA序列、同源物、片段或变体的至少一个5' UTR元件和衍生自CASP1基因的3' UTR或其相应RNA序列、同源物、片段或变体的至少一个3' UTR元件;或

[0833] h. 衍生自SLC7A3基因的5' UTR或其相应RNA序列、同源物、片段或变体的至少一个5' UTR元件和衍生自CASP1基因的3' UTR或其相应RNA序列、同源物、片段或变体的至少一个3' UTR元件;或

[0834] i. 衍生自SLC7A3基因的5' UTR或其相应RNA序列、同源物、片段或变体的至少一个5' UTR元件和衍生自PSMB3基因的3' UTR或其相应RNA序列、同源物、片段或变体的至少一个3' UTR元件;或

[0835] j. 衍生自NOSIP基因的5' UTR或其相应RNA序列、同源物、片段或变体的至少一个5' UTR元件、和衍生自PSMB3基因的3' UTR或其相应RNA序列、同源物、片段或变体的至少一个3' UTR元件;或

[0836] k. 衍生自NDUFA4基因的5' UTR或其相应RNA序列、同源物、片段或变体的至少一个5' UTR元件、和衍生自RPS9基因的3' UTR或其相应RNA序列、同源物、片段或变体的至少一个3' UTR元件;或

[0837] 1. 衍生自HSD17B4基因的5' UTR或其相应RNA序列、同源物、片段或变体的至少一个5' UTR元件、和衍生自RPS9基因的3' UTR或其相应RNA序列、同源物、片段或变体的至少一个

3' UTR元件;或

[0838] m. 衍生自ATP5A1基因的5' UTR或其同源物、片段或变体的至少一个5' UTR元件、,和衍生自GNAS基因的3' UTR或其同源物、片段或变体的至少一个3' UTR元件;或

[0839] n. 衍生自NDUFA4基因的5' UTR或其同源物、片段或变体的至少一个5' UTR元件、和衍生自COX6B1基因的3' UTR或其同源物、片段或变体的至少一个3' UTR元件;或

[0840] n. 衍生自NDUFA4基因的5' UTR或其同源物、片段或变体的至少一个5' UTR元件、和衍生自GNAS基因的3' UTR或其同源物、片段或变体的至少一个3' UTR元件;或

[0841] o. 衍生自NDUFA4基因的5' UTR或其同源物、片段或变体的至少一个5' UTR元件、和衍生自NDUFA1基因的3' UTR或其同源物、片段或变体的至少一个3' UTR元件;或

[0842] p. 衍生自NOSIP基因的5' UTR或其同源物、片段或变体的至少一个5' UTR元件、和衍生自NDUFA1基因的3' UTR或其同源物、片段或变体的至少一个3' UTR元件;或

[0843] q. 衍生自RPL31基因的5' UTR或其同源物、片段或变体的至少一个5' UTR元件、和衍生自GNAS基因的3' UTR或其同源物、片段或变体的至少一个3' UTR元件;或

[0844] r. 衍生自TUBB4B基因的5' UTR或其同源物、片段或变体的至少一个5' UTR元件、和衍生自RPS9基因的3' UTR或其同源物、片段或变体的至少一个3' UTR元件;或

[0845] s. 衍生自UBQLN2基因的5' UTR或其同源物、片段或变体的至少一个5' UTR元件、和衍生自RPS9基因的3' UTR或其同源物、片段或变体的至少一个3' UTR元件;

[0846] t. 衍生自MP68基因的5' UTR或其同源物、片段或变体的至少一个5' UTR元件、和衍生自GNAS基因的3' UTR或其同源物、片段或变体的至少一个3' UTR元件;或

[0847] u. 衍生自MP68基因的5' UTR或其同源物、片段或变体的至少一个5' UTR元件、和衍生自NDUFA1基因的3' UTR或其同源物、片段或变体的至少一个3' UTR元件。

[0848] 4. 根据项目3所述的人工核酸分子,包括根据d、e、g或l的UTR元件。

[0849] 5. 根据项目1至4中任一项所述的人工核酸分子,其中

[0850] -所述衍生自HSD17B4基因的5' UTR元件包括下列或由下列组成:根据SEQ ID NO:1的DNA序列或按递增优选顺序与根据SEQ ID NO:1的核酸序列具有至少50%、60%、70%、80%、90%、95%、96%、97%、98%、或99%序列同一性的DNA序列、或其片段或变体;或根据SEQ ID NO:2的RNA序列、或按递增优选顺序与根据SEQ ID NO:2的核酸序列具有至少50%、60%、70%、80%、90%、95%、96%、97%、98%、或99%序列同一性的RNA序列、或其片段或变体;

[0851] -所述衍生自RPL32基因的5' UTR元件包括下列或由下列组成:根据SEQ ID NO:21的DNA序列或按递增优选顺序与根据SEQ ID NO:21的核酸序列具有至少50%、60%、70%、80%、90%、95%、96%、97%、98%、或99%序列同一性的DNA序列、或其片段或变体;或根据SEQ ID NO:22的RNA序列、或按递增优选顺序与根据SEQ ID NO:22的核酸序列具有至少50%、60%、70%、80%、90%、95%、96%、97%、98%、或99%序列同一性的RNA序列、或其片段或变体;

[0852] -所述衍生自NDUFA4基因的5' UTR元件包括下列或由下列组成:根据SEQ ID NO:9的DNA序列、或按递增优选顺序与根据SEQ ID NO:9的核酸序列具有至少50%、60%、70%、80%、90%、95%、96%、97%、98%、或99%序列同一性的DNA序列、或其片段或变体;或根据SEQ ID NO:10的RNA序列、或按递增优选顺序与根据SEQ ID NO:10的核酸序列具有至少

50%、60%、70%、80%、90%、95%、96%、97%、98%、或99%序列同一性的RNA序列、或其片段或变体；

[0853] -所述衍生自SLC7A3基因的5' UTR元件包括下列或由下列组成：根据SEQ ID NO:15的DNA序列、或按递增优选顺序与根据SEQ ID NO:15的核酸序列具有至少50%、60%、70%、80%、90%、95%、96%、97%、98%、或99%序列同一性的DNA序列、或其片段或变体；或根据SEQ ID NO:16的RNA序列、或按递增优选顺序与根据SEQ ID NO:16的核酸序列具有至少50%、60%、70%、80%、90%、95%、96%、97%、98%、或99%序列同一性的RNA序列、或其片段或变体；

[0854] -所述衍生自NOSIP基因的5' UTR元件包括下列或由下列组成：根据SEQ ID NO:11的DNA序列、或按递增优选顺序与根据SEQ ID NO:11的核酸序列具有至少50%、60%、70%、80%、90%、95%、96%、97%、98%、或99%序列同一性的DNA序列、或其片段或变体；或根据SEQ ID NO:12的RNA序列、或按递增优选顺序与根据SEQ ID NO:12的核酸序列具有至少50%、60%、70%、80%、90%、95%、96%、97%、98%、或99%序列同一性的RNA序列、或其片段或变体；

[0855] -所述衍生自ATP5A1基因的5' UTR元件包括下列或由下列组成：根据SEQ ID NO:5的DNA序列、或按递增优选顺序与根据SEQ ID NO:5的核酸序列具有至少50%、60%、70%、80%、90%、95%、96%、97%、98%、或99%序列同一性的DNA序列、或其片段或变体；或根据SEQ ID NO:6的RNA序列、或按递增优选顺序与根据SEQ ID NO:6的核酸序列具有至少50%、60%、70%、80%、90%、95%、96%、97%、98%、或99%序列同一性的RNA序列、或其片段或变体；

[0856] 所述衍生自ASAH1基因的5' UTR元件包括下列或由下列组成：根据SEQ ID NO:3的DNA序列、或按递增优选顺序与根据SEQ ID NO:3的核酸序列具有至少50%、60%、70%、80%、90%、95%、96%、97%、98%、或99%序列同一性的DNA序列、或其片段或变体；或根据SEQ ID NO:4的RNA序列、或按递增优选顺序与根据SEQ ID NO:4的核酸序列具有至少50%、60%、70%、80%、90%、95%、96%、97%、98%、或99%序列同一性的RNA序列、或其片段或变体；

[0857] 所述衍生自Mp68基因的5' UTR元件包括下列或由下列组成：根据SEQ ID NO:7的DNA序列、或按递增优选顺序与根据SEQ ID NO:7的核酸序列具有至少50%、60%、70%、80%、90%、95%、96%、97%、98%、或99%序列同一性的DNA序列、或其片段或变体；或根据SEQ ID NO:8的RNA序列、或按递增优选顺序与根据SEQ ID NO:8的核酸序列具有至少50%、60%、70%、80%、90%、95%、96%、97%、98%、或99%序列同一性的RNA序列、或其片段或变体；

[0858] 所述衍生自Rp131基因的5' UTR元件包括下列或由下列组成：根据SEQ ID NO:13的DNA序列、或按递增优选顺序与根据SEQ ID NO:13的核酸序列具有至少50%、60%、70%、80%、90%、95%、96%、97%、98%、或99%序列同一性的DNA序列、或其片段或变体；或根据SEQ ID NO:14的RNA序列、或按递增优选顺序与根据SEQ ID NO:14的核酸序列具有至少50%、60%、70%、80%、90%、95%、96%、97%、98%、或99%序列同一性的RNA序列、或其片段或变体；

[0859] 所述衍生自TUBB4B基因的5' UTR元件包括下列或由下列组成：根据SEQ ID NO:17

的DNA序列、或按递增优选顺序与根据SEQ ID NO:17的核酸序列具有至少50%、60%、70%、80%、90%、95%、96%、97%、98%、或99%序列同一性的DNA序列、或其片段或变体；或根据SEQ ID NO:18的RNA序列、或按递增优选顺序与根据SEQ ID NO:18的核酸序列具有至少50%、60%、70%、80%、90%、95%、96%、97%、98%、或99%序列同一性的RNA序列、或其片段或变体；

[0860] 所述衍生自Ubp1n2基因的5' UTR元件包括下列或由下列组成：根据SEQ ID NO:19的DNA序列、或按递增优选顺序与根据SEQ ID NO:19的核酸序列具有至少50%、60%、70%、80%、90%、95%、96%、97%、98%、或99%序列同一性的DNA序列、或其片段或变体；或根据SEQ ID NO:20的RNA序列、或按递增优选顺序与根据SEQ ID NO:20的核酸序列具有至少50%、60%、70%、80%、90%、95%、96%、97%、98%、或99%序列同一性的RNA序列、或其片段或变体；

[0861] -所述衍生自GNAS基因的3' UTR元件包括下列或由下列组成：根据SEQ ID NO:29的DNA序列、或按递增优选顺序与根据SEQ ID NO:29的核酸序列具有至少50%、60%、70%、80%、90%、95%、96%、97%、98%、或99%序列同一性的DNA序列、或其片段或变体；根据SEQ ID NO:30的RNA序列、或按递增优选顺序与根据SEQ ID NO:30的核酸序列具有至少50%、60%、70%、80%、90%、95%、96%、97%、98%、或99%序列同一性的RNA序列、或其片段或变体；

[0862] -所述衍生自CASP1基因的3' UTR元件包括下列或由下列组成：根据SEQ ID NO:25的DNA序列、或按递增优选顺序与根据SEQ ID NO:25的核酸序列具有至少50%、60%、70%、80%、90%、95%、96%、97%、98%、或99%序列同一性的DNA序列、或其片段或变体；或根据SEQ ID NO:26的RNA序列、或按递增优选顺序与根据SEQ ID NO:26的核酸序列具有至少50%、60%、70%、80%、90%、95%、96%、97%、98%、或99%序列同一性的RNA序列、或其片段或变体；

[0863] -所述衍生自PSMB3基因的3' UTR元件包括下列或由下列组成：根据SEQ ID NO:23的DNA序列、或按递增优选顺序与根据SEQ ID NO:23的核酸序列具有至少50%、60%、70%、80%、90%、95%、96%、97%、98%、或99%序列同一性的DNA序列、或其片段或变体；或根据SEQ ID NO:24的RNA序列、或按递增优选顺序与根据SEQ ID NO:24的核酸序列具有至少50%、60%、70%、80%、90%、95%、96%、97%、98%、或99%序列同一性的RNA序列、或其片段或变体；

[0864] -所述衍生自ALB基因的3' UTR元件包括下列或由下列组成：根据SEQ ID NO:35的DNA序列、或按递增优选顺序与根据SEQ ID NO:35的核酸序列具有至少50%、60%、70%、80%、90%、95%、96%、97%、98%、或99%序列同一性的DNA序列、或其片段或变体；或根据SEQ ID NO:36的RNA序列、或按递增优选顺序与根据SEQ ID NO:36的核酸序列具有至少50%、60%、70%、80%、90%、95%、96%、97%、98%、或99%序列同一性的RNA序列、或其片段或变体；

[0865] -所述衍生自RPS9基因的3' UTR元件包括下列或由下列组成：根据SEQ ID NO:33的DNA序列、或按递增优选顺序与根据SEQ ID NO:33的核酸序列具有至少50%、60%、70%、80%、90%、95%、96%、97%、98%、或99%序列同一性的DNA序列、或其片段或变体；或根据SEQ ID NO:34的RNA序列、或按递增优选顺序与根据SEQ ID NO:34的核酸序列具有至少

50%、60%、70%、80%、90%、95%、96%、97%、98%、或99%序列同一性的RNA序列、或其片段或变体；

[0866] -所述衍生自COX6B1基因的3' UTR元件包括下列或由下列组成：根据SEQ ID NO:27的DNA序列、或按递增优选顺序与根据SEQ ID NO:27的核酸序列具有至少50%、60%、70%、80%、90%、95%、96%、97%、98%、或99%序列同一性的DNA序列、或其片段或变体；或根据SEQ ID NO:28的RNA序列、或按递增优选顺序与根据SEQ ID NO:28的核酸序列具有至少50%、60%、70%、80%、90%、95%、96%、97%、98%、或99%序列同一性的RNA序列、或其片段或变体；或

[0867] -所述衍生自Ndufa1基因的3' UTR元件包括下列或由下列组成：根据SEQ ID NO:31的DNA序列、或按递增优选顺序与根据SEQ ID NO:31的核酸序列具有至少50%、60%、70%、80%、90%、95%、96%、97%、98%、或99%序列同一性的DNA序列、或其片段或变体；或根据SEQ ID NO:32的RNA序列、或按递增优选顺序与根据SEQ ID NO:32的核酸序列具有至少50%、60%、70%、80%、90%、95%、96%、97%、98%、或99%序列同一性的RNA序列、或其片段或变体。

[0868] 6. 根据项目1至5中任一项所述的人工核酸分子，其中CRISPR相关蛋白质包括CRISPR相关野生型蛋白质、其同源物、变体、片段和衍生物。

[0869] 7. 根据项目1至6中任一项所述的人工核酸分子，其中所述CRISPR相关蛋白质选自Cas9、Cpf1 (Cas12)、C2c1、C2c3、Cas13、CasX或CasY。

[0870] 8. 根据项目1至7中任一项所述的人工核酸分子，包括编码CRISPR相关蛋白质的核酸序列的所述人工核酸包括下列或由下列组成：根据SEQ ID NO:428-441;10999-11001;442-1345中任一者的氨基酸序列、或按递增优选顺序与根据SEQ ID NO:428-441;10999-11001;442-1345中任一者的氨基酸序列具有至少50%、60%、70%、80%、90%、95%、96%、97%、98%、或99%序列同一性的氨基酸序列、或这些序列中任意者的变体或片段。

[0871] 9. 根据项目8所述的人工核酸分子，其中所述CRISPR相关蛋白质衍生物包括至少一个其他效应结构域，所述其他效应结构域任选地选自KRAB、CSD、WRPW、VP64、p65AD和Mxi。

[0872] 10. 根据项目1至9中任一项所述的人工核酸分子，其中所述人工核酸进一步包括编码核定位信号(NLS)的至少一种核酸序列，所述核定位信号(NLS)任选地选自包括下列或由下列组成的NLS：根据SEQ ID NO:426;427;10575;381;382;384;11957;11958-11964的氨基酸序列、或按递增优选顺序与根据SEQ ID NO:426;427;10575;381;382;384;11957;11958-11964的氨基酸序列具有至少50%、60%、70%、80%、90%、95%、96%、97%、98%、或99%序列同一性的氨基酸序列；和包括下列或由下列组成的NLS：根据SEQ ID NO:426;427;10575;381;382;384;11957;11958-11964的氨基酸序列、或按递增优选顺序与根据SEQ ID NO:426;427;10575;381;382;384;11957;11958-11964的氨基酸序列具有至少50%、60%、70%、80%、90%、95%、96%、97%、98%、或99%序列同一性的氨基酸序列；或具有按递增优选顺序与根据SEQ ID NO:12021-14274的氨基酸序列具有至少50%、60%、70%、80%、90%、95%、96%、97%、98%、或99%序列同一性的氨基酸序列的NLS。

[0873] 11. 根据项目1至10中任一项所述的人工核酸分子，其中所述人工核酸进一步包括编码蛋白质或肽标签的至少一种核酸序列。

[0874] 12. 根据项目1至11中任一项所述的人工核酸分子，其中所述人工核酸分子的所述

至少一个编码区包括下列或由下列组成:根据SEQ ID NO:411;2540-2553;11117-11119;11355-11357;2554-3457;1380-1393;3700-3713;4860-4873;6020-6033;7180-7193;8340-8353;11237-11239;11473-11475;11591-11593;11709-11711;11827-11829;11945-11947;1394-2297;3714-4617;4874-5777;6034-6937;7194-8097;8354-9257;412;3474-3887

[0875] 2314-2327;4634-4647;5794-5807;6954-6967;8114-8127;413-425;3490-3503;3506-3519;3522-3535;3538-3551;3554-3567;3570-3583;3586-3599;3602-3615;3618-3631;3634-3647;3650-3663;3666-3679;3682-3695;9514-9527;9626-9639;9738-9751;9850-9863;9962-9975;10074-10087;10186-10199;10298-10311;2330-2343;2346-2359;2362-2375;2378-2391;2394-2407;2410-2423;2426-2439;2442-2455;2458-2471;2474-2487;2490-2503;2506-2519;2522-2535;9498-9511;9610-9623;9722-9735;9834-9847;9946-9959;10058-10071;10170-10183-10282-10295;4650-4663;4666-4679;4682-4695;4698-4711;4714-4727;4730-4743;4746-4759;4762-4775;4778-4791;4794-4807;4810-4823;4826-4839;4842-4855;9530-9543;9642-9655;9754-9767;9866-9879;9978-9991;10090-10103;10202-10215;10314-10327;5810-5823;5826-5839;5842-5855;5858-5871;5874-5887;5890-5903;5906-5919;5922-5935;5938-5951;5954-5967;5970-5983;5986-5999;6002-6015;9546-9559;9658-9671;9770-9783;9882-9895;9994-10007;10106-10119;10218-10231;10330-10343;6970-6983;6986-6999;7002-7015;7018-7031;7034-7047;7050-7063;7066-7079;7082-7095;7098-7111;7114-7127;7130-7143;7146-7159;7162-7175;9562-9575;9674-9687;9786-9799;9898-9911;10010-10023;10122-10135;10234-10247;10346-10359;8130-8143;8146-8159;8162-8175;8178-8191;8194-8207;8210-8223;8226-8239;8242-8255;8258-8271;8274-8287;8290-8302;8306-8319;8322-8335;9578-9591;9690-9703;9802-9815;9914-9927;10026-10039;10138-10151;10250-10263;10362-10375;9290-9303;9306-9319;9322-9335;9338-9351;9354-9367;9370-9383;9386-9399;9402-9415;9418-9431;9434-9447;9450-9463;9466-9479;9482-9495;9594-9607;9706-9719;9818-9831;9930-9943;10042-10055;10154-10167;10266-10279;10378-10391中任一者的核酸序列;或按递增优选顺序与所述核酸序列中任一者具有至少50%、60%、70%、80%、90%、95%、96%、97%、98%、或99%序列同一性的核酸序列。

[0876] 13. 根据项目1至12中任一项所述的人工核酸分子,其中所述人工核酸分子包括根据SEQ ID NO:10552;3458-3459;3460-3473;2298-2299;4618-4619;5778-5779;6938-6939;8098-8099;9258-9259;2300-2313;4620-4633;5780-5793;6940-6953;8100-8113;9260-9273;3488-3489;10396;2328-2329;10395;4648-4649;10397;5808-5809;10398;6968-6969;10399;8128-8129;10400;9274-9287;3504-3505;3520-3521;3536-3537;3552-3553;3568-3669;3584-3585;3600-3601;3616-3617;3632-3633;3648-3649;3664-3665;3680-3681;3696-3697;9528-9529;9640-9641;9752-9753;9864-9865;9976-9977;10088-10089;10200-10201;10312-10313;10403;10410;10417;10424;10431;10438;10445;10452;10459;10466;10473;10480;10487;10494;10501;10508;10515;10522;10529;10536;10543;2344-2345;2360-2361;2376-2377;2392-2393;2408-2409;2424-2425;2440-2441;2456-2457;2472-2473;2489-2490;2504-2505;2520-2521;2536-2537;9512-9513;9624-9625;9736-9737;9848-9849;9960-9961;10072-10073;10184-10185;10296-10297;

10402;10409;10416;10423;10430;10437;10444;10451;10458;10465;10472;10479;  
10486;10493;10500;10507;10514;10521;10528;10535;10542;4664-4665;4680-4681;  
4696-4697;4712-4713;4728-4729;4744-4745;4760-4761;4776-4777;4792-4793;4808-  
4809;4824-4825;4840-4841;4856-4857;9544-9545;9656-9657;9768-9769;9880-9881;  
9992-9993;10104-10105;10216-10217;10328-10329;10404;10411;10418;10425;10432;  
10439;10446;10453;10460;10467;10474;10481;10488;10495;10502;10509;10516;  
10523;10530;10537;10544;5824-5825;5840-5841;5856-5857;5872-5873;5888-5889;  
5904-5905;5920-5921;5936-5937;5952-5953;5968-5969;5984-5985;6000-6001;6016-  
6017;9560-9561;9672-9673;9784-9785;9896-9897;10008-10009;10120-10121;10232-  
10233;10344-10345;10405;10412;10419;10426;10433;10440;10447;10454;10461;  
10468;10475;10482;10489;10496;10503;10510;10517;10524;10531;10538;10545;7033;  
7048-7049;7064-7065;7080-7081;7096-7097;7112-7113;7128-7129;7144-7145;7160-  
7161;7176-7177;9576-9577;9688-9689;9800-9801;9912-9913;10024-10025;10136-  
10137;10248-10249;10360-10361;10406;10413;10420;10427;10434;10441;10448;  
10455;10462;10469;10476;10483;10490;10497;10504;10511;10518;10525;10532;  
10539;10546;8144-8145;8160-8160;8176-8177;8192-8193;8208-8209;8224-8225;8240-  
8241;8256-8257;8272-8273;8288-8289;8304-8305;8320-8321;8336-8337;9592-9593;  
9704-9705;9816-9817;9928-9929;10040-10041;10152-10153;10264-10265;10376-  
10377;10407;10414;10421;10428;10435;10442;10449;10456;10463;10470;10477;  
10484;10491;10498;10505;10512;10519;10526;10533;10540;10547;9288-9289;10401;  
10553;10582-10583;10579-10580;10585-10586;10588-10589;10591-10592;10594-10595;  
10597-10598;10554-10574;10601;10602;10615;10616;10629;10630;10643;10644;  
10657;10658;10671;10672;10685;10686;10699;10700;10713;10714;10727;10728;  
10741;10742;10755;10756;10769;10770;10783;10784;10797;10798;10811;10812;  
10825;10826;10839;10840;10853;10854;10867;10868;10881;10882;10603;10604;  
10617;10618;10631;10632;10645;10646;10659;10660;10673;10674;10687;10688;  
10701;10702;10715;10716;10729;10730;10743;10744;10757;10758;10771;10772;  
10785;10786;10799;10800;10813;10814;10827;10828;10841;10842;10855;10856;  
10869;10870;10883;10884;10605;10606;10619;10620;10633;10634;10647;10648;  
10661;10662;10675;10676;10689;10690;10703;10704;10717;10718;10731;10732;  
10745;10746;10759;10760;10773;10774;10787;10788;10801;10802;10815;10816;  
10829;10830;10843;10844;10857;10858;10871;10872;10885;10886;10607;10608;  
10621;10622;10635;10636;10649;10650;10663;10664;10677;10678;10691;10692;  
10705;10706;10719;10720;10733;10734;10747;10748;10761;10762;10775;10776;  
10789;10790;10803;10804;10817;10818;10831;10832;10845;10846;10859;10860;  
10873;10874;10887;10888;10609;10610;10623;10624;10637;10638;10651;10652;  
10665;10666;10679;10680;10693;10694;10707;10708;10721;10722;10735;10736;  
10749;10750;10763;10764;10777;10778;10791;10792;10805;10806;10819;10820;  
10833;10834;10847;10848;10861;10862;10875;10876;10889;10890;10611;10612;



10625;10626;10639;10640;10653;10654;10667;10668;10681;10682;10695;10696;10709;10710;10723;10724;10737;10738;10751;10752;10765;10766;10779;10780;10793;10794;10807;10808;10821;10822;10835;10836;10849;10850;10863;10864;10877;10878;10891;10892;9304-9305;9320-9321;9336-9337;9352-9353;9368-9369;9384-9385;9400-9401;9416-9417;9432-9433;9448-9449;9464-9465;9480-9481;9496-9497;9608-9609;9720-9721;9832-9833;9944-9945;10056-10057;10168-10169;10280-10281;10392-10393;10408;10415;10422;10429;10436;10443;10450;10457;10464;10471;10478;10485;10492;10499;10506;10513;10520;10527;10534;10541;10548中任一者的核酸序列、或按递增优选顺序与所述核酸序列中任一者具有至少50%、60%、70%、80%、90%、95%、96%、97%、98%、或99%序列同一性的核酸序列。

[0877] 14. 根据项目1至13中任一项所述的人工核酸分子,其中所述人工核酸分子包括下列或由下列组成:根据SEQ ID NO:11011-11042;11249-11280;11131-11162;11367-11398;11485-11516;11603-11634;11721-11752;11839-11870;11044-11116;11282-11354;11164-11236;11400-11472;11518-11590;11636-11708;11754-11826;11872-11944;11011-11042;11249-11280;11044-11116;11282-11354;11131-11162;11367-11398;11485-11516;11603-11634;11721-11752;11839-11870;11164-11236;11400-11472;11518-11590;11636-11708;11754-11826;11872-11944;11120-11122;11240;11241;11358;11359;11476;11477;11594;11595;11712;11713;11830;11831;11948;11949;11123-11130;11360-11366;11242-11248;11478-11484;11596-11602;11714-11720;11832-11838;11950-11956中任一者的核酸序列、或按递增优选顺序与所述核酸序列中任一者具有至少50%、60%、70%、80%、90%、95%、96%、97%、98%、或99%序列同一性的核酸序列。

[0878] 15. 根据项目1至14中任一项所述的人工核酸分子,其中所述人工核酸分子是RNA。

[0879] 16. 根据项目15所述的RNA,其中所述RNA是单顺反子、双顺反子、或多顺反子的。

[0880] 17. 根据项目14或15所述的RNA,其中所述RNA是mRNA、病毒RNA或复制子RNA。

[0881] 18. 根据项目1至17中任一者所述的人工核酸,优选地RNA,其中所述人工核酸是修饰核酸,优选地稳定化的核酸。

[0882] 19. 根据项目1至18中任一者的人工核酸,优选地RNA,其中

[0883] -所述人工核酸的所述至少一个编码区的G/C含量与相应野生型人工核酸的相应编码序列的G/C含量相比增加,和/或其中

[0884] -所述人工核酸的所述至少一个编码区的C含量与相应野生型人工核酸的相应编码序列的C含量相比增加,和/或其中

[0885] -所述人工核酸的所述至少一个编码区的密码子适于密码子使用,其中在所述人工核酸的所述至少一种编码序列中密码子适应指数(CAI)优选增加或最大化,

[0886] -其中所述人工核酸编码的氨基酸序列优选地与相应野生型人工核酸编码的氨基酸序列相比不被修饰。

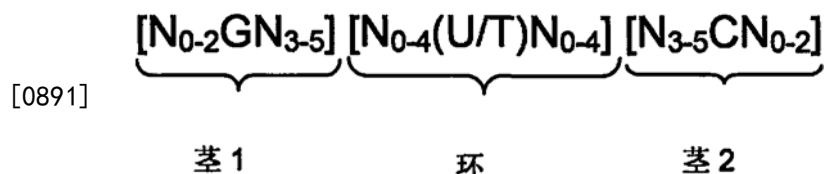
[0887] 20. 根据项目1至19中任一者的人工核酸,优选地RNA,其包括5' -帽结构,优选地m7GpppN或帽1。

[0888] 21. 根据项目1至20中任一者所述的人工核酸,优选地RNA,其包括至少一种组蛋白

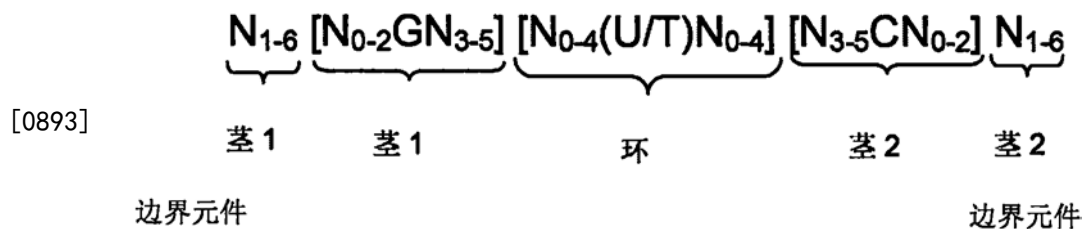
茎-环。

[0889] 22. 根据项目21所述的人工核酸, 优选地RNA, 其中所述至少一种组蛋白茎-环包括根据下列式(I)或(II)的核酸序列:

[0890] 式(I) (苜-环序列,无苜边界元件):



[0892] 式(II) (茎-环序列,有茎边界元件):



[0894] 其中：

[0895] 茎1或茎2边界元件N<sub>1-6</sub>是1至6个,优选2至6个,更优选2至5个,甚至更优选3至5个,最优选4至5个或5个N的连续序列,其中各N彼此独立地选自以下核苷酸:选自A、U、T、G和C、或其核苷酸类似物;

[0896] 茎1 [N<sub>0-2</sub>GN<sub>3-5</sub>] 与元件茎2反向互补或部分反向互补, 并且是5至7个核苷酸的连续序列:

[0897] 其中N<sub>0-2</sub>是0至2个,优选0至1个,更优选1个N的连续序列,其中各N彼此独立地选自以下核苷酸:选自A、U、T、G和C或其核苷酸类似物:

[0898] 其中N<sub>3-5</sub>是3至5个,优选4至5个,更优选4个N的连续序列,其中各N彼此独立地选自以下核苷酸:选自A、U、T、G和C或其核苷酸类似物,并且

[0899] 其中G是鸟苷或其类似物,并且可任选地被胞苷或其类似物替代,条件是茎2中其互补核苷酸胞苷被鸟苷替代;

[0900] 环序列[N<sub>0-4</sub> (U/T) N<sub>0-4</sub>]位于元件茎1和茎2之间,并且是3至5个核苷酸,更优选4个核苷酸的连续序列:

[0901] 其中各N<sub>0-4</sub>彼此独立地是0至4个,优选1至3个,更优选1至2个N的连续序列,其中各N彼此独立地选自以下核苷酸:选自A、U、T、G和C或其核苷酸类似物;并且

[0902] 其中U/T表示尿苷、或任选地胸苷；

[0903] 茎2[N<sub>3-5</sub>CN<sub>0-2</sub>]与元件茎1反向互补或部分反向互补,并且是5至7个核苷酸的连续序列;

[0904] 其中N<sub>3-5</sub>是3至5个,优选4至5个,更优选4个N的连续序列,其中各N彼此独立地选自以下核苷酸:选自A、U、T、G和C或其核苷酸类似物;

[0905] 其中N<sub>0-2</sub>是0至2个,优选0至1个,更优选1个N的连续序列,其中各N彼此独立地选自以下核苷酸:选自A、U、T、G和C或其核苷酸类似物;并且其中C是胞苷或其类似物,并且可任选地被鸟苷或其类似物替代,条件是茎1中其互补核苷酸鸟苷被胞苷替代;

[0906] 其中

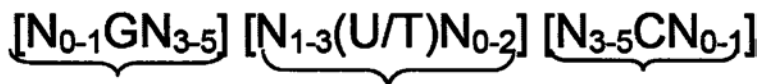
[0907] 茎1和茎2能够彼此碱基配对，

[0908] 形成反向互补序列，其中碱基配对可发生在茎1和茎2之间、或

[0909] 形成部分反向互补序列，其中不完全碱基配对可发生在茎1和茎2之间。

[0910] 24. 根据项目19或20所述的人工核酸，优选地RNA，其中所述至少一种组蛋白茎-环包括根据下列式 (Ia) 或 (IIa) 的核酸序列：

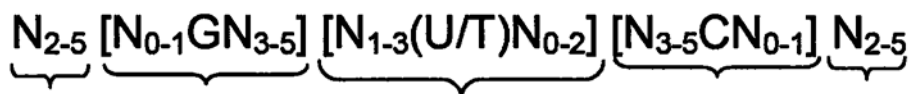
[0911] 式 (Ia) (茎-环序列，无茎边界元件)：



[0912]

茎 1                      环                      茎 2

[0913] 式 (IIa) (茎-环序列，有茎边界元件)：



[0914]

茎 1              茎 1                      环                      茎 2              茎 2

边界元件

边界元件。

[0915] 25. 根据项目1至24中任一者的人工核酸，优选地RNA，任选地包括聚 (A) 序列，优选地包括10至200个、10至100个、40至80个、或50至70个腺苷核苷酸。

[0916] 26. 根据项目1至25中任一者的人工核酸，优选地RNA，任选地包括聚 (C) 序列，优选地包括10至200个、10至100个、20至70个、20至60个、或10至40个胞嘧啶核苷酸。

[0917] 27. 根据项目1至26中任一者的人工核酸，优选地RNA，其包括，优选5' 至3' 方向，下列元件：

[0918] a) 5' -帽结构，优选地m7GpppN或帽1

[0919] b) 5' -UTR元件，其包括下列或由下列组成：项目1至5中任一项限定的衍生自5' -UTR的核酸序列，优选地包括与根据SEQ ID NO:1;3;5;7;9;11;13;15;17;19;或21的核酸序列或其同源物、片段或变体相应的核酸序列，

[0920] c) 项目7至13中任一项限定的至少一种编码序列

[0921] d) 3' -UTR元件，其包括下列或由下列组成：项目1至5中任一项限定的衍生自3' -UTR的核酸序列，优选地包括与根据SEQ ID NO:15;17;19;2129;31;33或35的核酸序列、或其同源物、片段或变体相应的核酸序列，

[0922] e) 任选地聚 (A) 尾，其优选地由10至1000个、10至500个、10至300个、10至200个、10至100个、40至80个、或50至70个腺苷核苷酸组成，

[0923] f) 任选地a聚 (C) 尾，其优选地由10至200个、10至100个、20至70个、20至60个、或10至40个胞嘧啶核苷酸组成，和

[0924] g) 任选地组蛋白茎-环 (HSL)。

[0925] 28. 组合物，包括根据项目1至26中任一项所述的人工核酸分子，优选地RNA，和药学上可接受的载体和/或赋形剂。

[0926] 29. 根据项目28所述的组合物，其中所述人工核酸分子 (优选地RNA) 与一种或多种阳离子或聚阳离子化合物，优选与阳离子或聚阳离子聚合物、阳离子或聚阳离子肽或蛋白

质,例如鱼精蛋白,阳离子或聚阳离子多糖和/或阳离子或聚阳离子脂质复合。

[0927] 30.根据项目29所述的组合物,其中所述人工核酸分子(优选地RNA)与一种或多种阳离子或聚阳离子肽或蛋白质的N/P比在约0.1至10的范围内,包括约0.3至4、约0.5至2、约0.7至2和约0.7至1.5的范围。

[0928] 31.根据项目28至30中任一项所述的组合物,其中所述人工核酸分子(优选地RNA)与一种或多种脂质复合,从而形成脂质体、脂质纳米颗粒和/或脂质复合物。

[0929] 32.根据项目28至31中任一项所述的组合物,进一步包括至少一种向导RNA(gRNA)或编码其的核酸,所述gRNA能够将CRISPR相关蛋白质靶向至目的靶DNA序列、或其可操作地连接的调控元件。

[0930] 33.试剂盒,优选地多部分试剂盒,包括根据项目1至27中任一项所述的人工核酸分子(优选地RNA)或根据项目28至32中任一项所述的组合物,和任选地液体媒介和/或任选地具有关于所述人工核酸分子或所述组合物的给予和剂量的信息的技术说明书。

[0931] 34.根据项目所述的试剂盒33,其中试剂盒包含林格-乳酸盐溶液作为一部分。

[0932] 35.根据项目33或34所述的试剂盒,进一步包括向导RNA(gRNA)或编码其的核酸,所述gRNA能够将CRISPR相关蛋白质靶向至目的靶DNA序列、或其可操作地连接的调控元件。

[0933] 36.根据项目1至27中任一项所述的人工核酸分子(优选地RNA)、根据项目28至32中任一项所述的组合物、或根据项目33至35所述的试剂盒,其用作药物。

[0934] 37.根据项目1至27中任一项所述的人工核酸分子(优选地RNA)、根据项目28至32中任一项所述的组合物、或根据项目33至35所述的试剂盒,其用于基因疗法。

[0935] 38.根据项目1至27中任一项所述的人工核酸分子(优选地RNA)、根据项目28至32中任一项所述的组合物、或根据项目33至35所述的试剂盒,其用于调控目的基因的表达的方法,包括向有此需要的患者给予(a)所述人工核酸分子(优选地RNA)、所述组合物或所述试剂盒,和(b)向导RNA(gRNA)或编码其的核酸,所述sgRNA能够将CRISPR相关蛋白质靶向至目的基因、或其可操作地连接的调控元件。

[0936] 39.根据项目1至27中任一项所述的人工核酸分子(优选地RNA)、根据项目28至32中任一项所述的组合物、或根据项目33至35所述的试剂盒,用作适于通过表达所述至少一种编码序列编码的CRISPR相关蛋白质进行治疗的疾病、障碍或状况的药物或用于适于通过表达所述至少一种编码序列编码的CRISPR相关蛋白质进行治疗的疾病、障碍或状况的基因疗法。

[0937] 40.根据项目1至27中任一项所述的人工核酸分子(优选地RNA)、根据项目28至32中任一项所述的组合物、或根据项目33至35所述的试剂盒,用作通过敲入、敲除或操纵目的基因、或通过调控目的基因的表达可处理的疾病、障碍或状况的药物或用于通过敲入、敲除或操纵目的基因、或通过调控目的基因的表达可处理的疾病、障碍或状况的基因疗法。

[0938] 41.根据项目40所述的用途的人工核酸分子(优选地RNA)、组合物或试剂盒,其中所述疾病、障碍或状况选自遗传性疾病、癌症、自身免疫疾病、炎症性疾病和感染性疾病。

[0939] 42.根据项目1至27中任一项所述的人工核酸分子(优选地RNA)、根据项目28至32中任一项所述的组合物、或根据项目33至35所述的试剂盒用于增加编码的所述CRISPR相关蛋白质的表达的用途,任选地在基因疗法中。

[0940] 43.根据项目1至27中任一项所述的人工核酸分子(优选地RNA)、根据项目28至32

中任一项所述的组合物、或根据项目33至35所述的试剂盒用于调控编码的所述CRISPR相关蛋白质所靶向的目的基因的表达的用途。

[0941] 44. 用于调控目的基因的表达的方法, 包括以下步骤:

[0942] a) 提供根据项目1至27中任一项所述的人工核酸分子(优选地RNA);

[0943] b) 提供向导RNA(gRNA)或编码其的核酸, 所述gRNA能够将CRISPR相关蛋白质靶向至目的靶DNA序列、或其可操作地连接的调控元件,

[0944] c) 使细胞、组织或生物体与所述人工核酸分子(优选地RNA)和所述gRNA或编码其的核酸在适于调控所述目的基因的表达效力的条件下接触。

[0945] 45. 治疗或预防障碍的方法, 其中所述方法包括向有此需要的对象给予有效量的根据项目1至27中任一项所述的人工核酸分子(优选地RNA)、根据项目28至32中任一项所述的组合物、或根据项目33至35所述的试剂盒, 和向导RNA(gRNA)或编码其的核酸, 所述gRNA能够将CRISPR相关蛋白质靶向至目的靶DNA序列、或其可操作地连接的调控元件。

[0946] 46. 根据项目45所述的方法, 其中所述障碍是适于通过表达编码的所述CRISPR相关蛋白质进行治疗, 优选地适于通过调控所述CRISPR相关蛋白质所靶向的目的基因的表达进行治疗的疾病、障碍或状况。

[0947] 47. 根据项目45或46所述的方法, 其中所述障碍是通过敲入、敲除或通过突变目的基因、或通过改变目的基因的表达可处理的疾病、障碍或状况。

[0948] 48. 增加包括编码CRISPR相关蛋白质的编码区的人工核酸分子(优选地RNA)的表达效力的方法, 所述方法包括

[0949] (a) 使所述编码区与至少一个5' UTR元件缔合, 所述5' UTR元件衍生自选自ATP5A1, RPL32, HSD17B4, SLC7A3, NOSIP、或NDUFA4的基因的5' UTR、或其相应RNA序列、同源物、片段或变体;

[0950] (b) 使所述编码区与至少一个3' UTR元件缔合, 所述3' UTR元件衍生自选自GNAS, CASP1, PSMB3, ALB、或RPS9的基因的3' UTR、或其相应RNA序列、同源物、片段或变体; 和

[0951] (c) 获得根据项目1至47中任一项所述的人工核酸分子(优选地RNA)。

[0952] 附图描述

[0953] 图1显示了不同UTR组合对HeLa细胞中Cas9表达水平的影响, 如实施例1(In-Cell Western)中详述。将y轴标准化以显示UTR组合RPL32/ALB7.1的表达水平为100%。

[0954] 图2显示了不同UTR组合对Hek293T细胞中Cas9表达水平的影响, 如实施例2(In-Cell Western)中详述。将y轴标准化以显示UTR组合RPL32/ALB7.1的表达水平为100%。

[0955] 图3显示了不同UTR组合对HepG2细胞中Cas9表达水平的影响, 如实施例1(Western印迹)中详述。将y轴标准化以显示UTR组合RPL32/ALB7.1的表达水平为100%。

[0956] 图4显示了如实施例2(Western印迹)中详述的HeLa细胞中优化的spCas9 mRNA构建体的表达水平。将y轴标准化以显示UTR组合RPL32/ALB7.1的表达水平为100%。

[0957] 图5显示了如实施例2(In-cell Western)中详述的HeLa细胞中优化的spCas9 mRNA构建体的表达水平。将y轴标准化以显示UTR组合RPL32/ALB7.1的表达水平为100%。

[0958] 图6显示了与市售Cas9 mRNA相比, 实施例1(In-cell Western)中详述的HeLa细胞中优化的spCas9 mRNA构建体的表达水平。将y轴标准化以显示商品Cas9 mRNA的表达水平为100%。

[0959] 图7显示了如实施例1(In-cell Western)中详述的HeLa细胞中优化的spCas9 mRNA构建体的表达水平。将y轴标准化以显示UTR组合RPL32/ALB7.1的表达水平为100%

[0960] 图8显示了与市售Cas9 mRNA相比,如实施例2(In-cell Western)中详述的Hek293T细胞中优化的spCas9 mRNA构建体的表达水平。

[0961] 图9显示了与市售Cas9 mRNA相比,如实施例2(Western印迹)中详述的HepG2细胞中优化的spCas9 mRNA构建体的表达水平。将y轴标准化以显示商品Cas9 mRNA的表达水平为100%。

[0962] 图10显示了通过错配核酸酶测定/错配检测测定得到的由实施例3中详述的mRNA构建体表达的spCas9的DNA编辑活性。A:PCR扩增,B:错配检测测定。

[0963] 图11显示了HepG2中优化的Cpf1 mRNA构建体相对于RPL32/ALB7的表达水平(%)。将y-轴标准化以显示RPL32/ALB7 UTR组合Cpf1 mRNA的表达水平为100%。

[0964] 图12显示HeLa中优化的Cpf1 mRNA构建体相对于RPL32/ALB7的表达水平(%)。将y-轴标准化以显示RPL32/ALB7 UTR组合Cpf1 mRNA的表达水平为100%。

## 实施例

[0965] 在下文中,展示了示例本发明的各种实施方式和方面的具体实施例。然而,本发明的范围不限于本文所述的具体实施方式。给出以下制备和实施例以使本领域技术人员能够更清楚地理解和实施本发明。然而,本发明的范围不受示例的实施方式的限制,这些实施方式仅用于示例本发明的单个方面,并且功能等同的方法在本发明的范围内。实际上,根据前文描述、附图和以下实施例,除了本文所述的那些之外,本发明的各种修改也将对于本领域技术人员而言变得显而易见。所有这些修改都落入所附权利要求的范围内。

[0966] 实施例1:利用In-cell-Western和Western印迹分析检测HeLa、Hek293T和HepG2细胞中的Cas9表达

[0967] 将细胞接种于96孔板(Nunc Microplate Black w/Clear Optical Bottom; Thermo Fisher)中的相容完全细胞培养基(200 $\mu$ l)中,HeLa的密度为10,000个细胞/孔;Hek293T和HEPG2的密度为20,000个细胞/孔)。将细胞维持在37 $^{\circ}$ C,5%CO<sub>2</sub>下24小时。转染当天,用50 $\mu$ l无血清Opti-MEM培养基(Thermo Fisher)替换完全培养基。100ng各mRNA,即SEQ ID NO:14274(RPL32/ALB7.1),SEQ ID NO:14275(HSD17B4/CASP1.1),SEQ ID NO:14276(SLC7A3.1/PSMB3.1),SEQ ID NO:14277(SLC7A3.1/CASP1.1),SEQ ID NO:14278(NOSIP.1/PSMB3.1),SEQ ID NO:14279(NDUFA4.1/RPS9.1),SEQ ID NO:14280(NDUFA4.1/PSMB3.1),SEQ ID NO:14281(HSD17B4/PSMB3.1)和SEQ ID NO:14282(HSD17B4/RPS9.1),利用Lipofectamine 3000在50 $\mu$ l Opti-MEM中进行脂质复合,其中mRNA:Lipofectamine 3000的比例为1:1.5。然后将脂质复合的mRNA加入相应的96孔板中。转染后3小时,用100 $\mu$ l完全细胞培养基替换该完全培养基。在进行In-cell-Western之前,将细胞在37 $^{\circ}$ C,5%CO<sub>2</sub>下进一步维持24小时。

[0968] 对于In-cell Western分析(HeLa和Hek293T),将细胞用PBS1X洗涤两次,并用甲醇/丙酮(1:1)溶液固定10分钟。固定后,随后用PBS1X将细胞洗涤3次,每次5分钟。为了避免非特异性结合,将细胞在室温下用补充有0.01%Triton X100的Odyssey封闭缓冲剂(PBS, LI-COR)封闭1小时,然后用一抗(即多克隆兔抗spCas9抗体(1/1000;#632606;Clontech/

Takara)) 培育一个半小时。然后在温和摇动(80rpm)下用PBS1X中的0.1%吐温-20洗涤细胞4次,每次5分钟。

[0969] 随后,将二抗(即**红外染料**<sup>®</sup>800CW山羊抗兔多克隆抗体(1/250;LI-COR))与Odyssey封闭缓冲剂中的Cell-Tag 700染色剂(1/5000;LI-COR)混合,并在室温下在黑暗中培育一小时。在使用**Odyssey**<sup>®</sup>CLx成像系统(LI-COR)扫描之前,如上所述进行洗涤步骤。使用Image Studio<sup>™</sup>Lite软件获得相对定量(800/700)。从测量结果减去由在没有mRNA的情况下脂质转染的孔获得的背景荧光,并将结果与来自市售Cas9编码RNA(TriLink BioTechnologies,LLC,目录号L-6125)的表达进行比较。

[0970] 对于HepG2细胞,利用Western印迹进行分析。将板在PBS1X中洗涤两次,并直接与含有Benzonase内切核酸酶(Millipore)的50μl样品加载缓冲剂1X(Biorad)一起在室温下培育20分钟。然后将板在95℃培育5分钟并离心。

[0971] 将15μl裂解物在10%Mini-Protean TGX凝胶(Biorad)上运行,并转移至硝酸纤维素膜(100V;90分钟)。将膜用PBS中的0.1TritonX100洗涤三次,每次10分钟,并用PBS1X中的10%乳汁(milk)在4℃下饱和过夜。

[0972] PBS1X中的5%乳汁中的多克隆兔抗spCas9抗体(1/1000;#632606;Clontech/Takara)和对照小鼠单克隆抗-β肌动蛋白抗体(1/10000;ab6276;Abcam)在室温下培育1小时。随后将膜在TBS1X中的0.1%吐温-20中洗涤三次。**红外染料**<sup>®</sup>800CW山羊抗兔多克隆和**红外染料**<sup>®</sup>山羊抗小鼠抗体(分别为1/7500和1/10000)用于检测。在使用**Odyssey**<sup>®</sup>CLx成像系统(LI-COR)扫描之前,将如下所述的洗涤步骤重复三次。使用Studio<sup>™</sup>Lite软件(LI-COR)进行条带强度和相对定量。

[0973] 包含根据本发明的UTR组合的Cas9编码mRNA呈现增加的表达(图1、图3、图7),其与市售的Cas9 mRNA相比是高度优越的(图6)。

[0974] 实施例2:使用Western印迹分析检测HeLa细胞中的Cas9表达

[0975] 将Hela细胞接种于12孔板(Nunc;Thermo Fisher)中的完全细胞培养基(RPMI;10%胎牛血清;1%青霉素/链霉素和1%L-谷氨酰胺);Lonza)中,Hela的最终密度为200,000个细胞/孔。将细胞维持在37℃,5%CO<sub>2</sub>下24小时。转染当天,用750μl无血清Opti-MEM培养基(Thermo Fisher)替换该完全培养基。将包含本发明的UTR组合(参见Figure Legends)的1μg Cas9编码mRNA,即SEQ ID NO:14274(RPL32/ALB7.1)、SEQ ID NO:14275(HSD17B4/CASP1.1)、SEQ ID NO:14276(SLC7A3.1/PSMB3.1)、SEQ ID NO:14277(SLC7A3.1/CASP1.1)、SEQ ID NO:14278(NOSIP.1/PSMB3.1)、SEQ ID NO:14279(NDUFA4.1/RPS9.1)、SEQ ID NO:14280(NDUFA4.1/PSMB3.1)、SEQ ID NO:14281(HSD17B4/PSMB3.1)和SEQ ID NO:14282(HSD17B4/RPS9.1),利用Lipofectamine 3000在250μlOpti-MEM中进行脂质复合,其中mRNA:Lipofectamine 3000比为1:1.5。然后将脂质复合的mRNA加入各孔中。转染后3小时,用1ml完全细胞培养基替换该完全培养基。在进行蛋白质提取之前,将细胞在37℃,5%CO<sub>2</sub>下进一步维持24小时。

[0976] 将孔在PBS1X中洗涤两次,并直接与含有Benzonase内切核酸酶(Millipore)的100μl样品加载缓冲剂1X(Biorad)在室温下培育20分钟。刮下细胞,将裂解物转移到Eppendorf管中。然后将样品在95℃下变性5分钟,在冰中冷却5分钟,并在加载前以最大速度离心2分钟。

[0977] 将15 $\mu$ l裂解物在10%Mini-Protean TGX凝胶(Biorad)上运行,并转移至硝酸纤维素膜(100V;90分钟)。将膜用PBS中的0.1 Triton X100洗涤三次,每次10分钟,并用PBS1X中的10%乳汁在4℃下饱和过夜。

[0978] 将多克隆兔抗spCas9抗体(1/1000;#632606;Clontech/Takara)和小鼠单克隆抗- $\beta$ 肌动蛋白抗体(1/10000;ab6276;Abcam)在PBS1X中的5%乳汁中在室温下培育1小时。随后将膜在TBS1X中的0.1%吐温-20中洗涤三次。将**红外染料**<sup>®</sup>800CW山羊抗兔多克隆和**红外染料**<sup>®</sup>山羊抗小鼠抗体(分别为1/7500和1/10000)用于检测。在使用**Odyssey**<sup>®</sup>CLx成像系统(LI-COR)扫描之前,将如下所述的洗涤步骤重复三次。使用Studio<sup>TM</sup>Lite软件(LI-COR)进行条带强度和相对定量。

[0979] 将来自本发明mRNA的Cas9表达与来自市售Cas9编码RNA(TriLink BioTechnologies,LLC,目录号L-6125)的表达进行比较。包含根据本发明的UTR组合的Cas9编码mRNA呈现增加的表达(图2、图4、图5),其与市售的Cas9 mRNA相比是高度优越的(图8、图9)。

[0980] 实施例3:利用错配检测测定法确定体外spCas 9活性

[0981] 将细胞接种于96孔板中的完全细胞培养基(200 $\mu$ l RPMI;10%胎牛血清;1%青霉素/链霉素和1%L-谷氨酰胺;Lonza)中,Hela的密度为50,000个细胞/孔。将细胞维持在37℃,5%CO<sub>2</sub>下24小时。转染当天,利用**TransIT**<sup>®</sup>mRNA转染(Mirrus Bio),将100ng spCas9 mRNA(SEQ ID NO:14274=RPL32/ALB7.1,SEQ ID NO:14281=HSD17B4/PSMB3.1——各mRNA一次无另外的酶促聚腺苷酸化,一次有另外的酶促聚腺苷酸化;作为参考,使用市售Cas9编码RNA TriLink BioTechnologies,LLC,Cat.No.L-6125)与针对人肽基脯氨酰异构酶B的crRNA(PPIB,即亲环蛋白B;Dharmacon;S0-2544646G)(25nM)和tracrRNA(25nM;Dharmacon;U-002000-20)的混合物脂质复合。作为阴性对照,转染混合物中省略spCas9 mRNA或向导RNA或两者。将最终体积10 $\mu$ l加入96孔板中的培养基中。转染后3小时,用100 $\mu$ l完全细胞培养基替换该完全培养基。在进行细胞提取之前,将细胞在37℃,5%CO<sub>2</sub>下进一步维持24小时。

[0982] 将孔用PBS1X洗涤一次,并使用100 $\mu$ l含有1mg/ml蛋白酶K和0.5mg/ml RNA酶A(Thermo Fisher)的Phusion HF缓冲剂1X在56℃下裂解细胞30分钟。在裂解结束时增加95℃5分钟变性步骤。

[0983] 使用针对人PPIB的引物(Edit-R PPIB crRNA控制试剂盒;UK-007060;Dharmacon)和Phusion热启动II高保真DNA聚合酶(Thermo Fisher),将5 $\mu$ l细胞裂解物用于hPPIB片段的PCR扩增。PCR扩增条件如下:

[0984] 1) 变性步骤:98℃,3分钟;

[0985] 2) 降落(Touchdown)PCR循环反应10次(变性,98℃,10秒;降落退火,72℃-1℃/循环,15秒;延伸,72℃,30秒);

[0986] 3) 正常PCR循环反应25次(变性,98℃,10秒;退火,62℃,15秒;延伸,72℃,30秒);

[0987] 4) 最终延伸,72℃,10分钟。

[0988] 然后将PCR样品在95℃加热10分钟并在室温下缓慢冷却超过15分钟。

[0989] 在25分钟的培育期间,在37℃下,错配检测测定利用10 $\mu$ l的PCR反应、NEB缓冲剂2和T7核酸内切酶(New England BioLabs)进行。立即将全部反应体在2%琼脂糖凝胶上运



行。预期的带为505bp(无编辑)或分别330bp和174bp(有编辑)。

[0990] 由包含本发明UTR组合的mRNA表达的spCas9能够编辑靶DNA并因此起作用(图10A、B)

[0991] 实施例4:利用In-cell-Western检测HeLa、Hek293T和HepG2细胞中的Cpf1表达

[0992] 将细胞接种于96孔板(Nunc Microplate Black w/Clear Optical Bottom Thermo Fisher)中的如前使用的相容完全细胞培养基中(200 $\mu$ l),HeLa的密度为10,000个细胞/孔;HepG2和Hek293T的密度为20,000个细胞/孔。将细胞在37 $^{\circ}$ C,5%CO<sub>2</sub>下维持24小时。转染当天,将100ng mRNA,即SEQ ID NO:10549(RPL32/ALB7)、SEQ ID NO:14289(Mp68/Gnas.1)、SEQ ID NO:14290(Ndufa4.1/PSMB3.1)、SEQ ID NO:14291(HSD17B4/Gnas.1)、SEQ ID NO:14292(HSD17B4/PSMB3.1)和SEQ ID NO:14293(Ndufa4.1/Alb7)(用于转染HeLa和Hek293T细胞),和500ng mRNA,即相同的SEQ ID NO(HepG2),利用Lipofectamine Messenger Max在50 $\mu$ l的Opti-MEM中脂质复合,其中mRNA:Lipofectamine Messenger Max的比例为1:1.5。然后将脂质复合的mRNA加入相应的96孔板中。在进行In-cell-Western之前,将细胞在37 $^{\circ}$ C,5%CO<sub>2</sub>下进一步维持24小时。

[0993] 对于In-Cell Western分析(HeLa、HepG2和Hek293T),将细胞用PBS1X洗涤三次,并用多聚甲醛4%固定10分钟。固定后,随后将细胞用PBS1X洗涤3次,每次5分钟,并用PBS中的2%Triton X100透化15分钟。为了避免非特异性结合,将细胞在室温下用Odyssey封闭缓冲剂(PBS,LI-COR)封闭1小时,然后用一抗(即多克隆兔抗HA抗体(1/1000);H6908;Sigma Aldrich))培育一个半小时。然后将细胞在温和摇动(80rpm)下用PBS1X中的0.1%吐温-20洗涤5分钟,4次。

[0994] 随后,将二抗(即**红外染料**<sup>®</sup>800CW山羊抗兔多克隆抗体(1/500;LI-COR))与Odyssey封闭缓冲剂中的Cell-Tag 700染色(1/5000;LI-COR)混合,并在室温下在黑暗中培育一小时。在使用**Odyssey**<sup>®</sup>CLx成像系统(LI-COR)扫描之前,如上所述进行洗涤步骤。使用Image Studio<sup>™</sup>Lite软件获得相对定量(800/700)。从测量结果减去在没有mRNA的情况下的脂质转染的孔获得的背景荧光,并将结果与来自我们的标准mRNA的表达进行比较。

[0995] 包含根据本发明的UTR组合的Cpf1编码mRNA呈现高度增加的表达,其与参考Cpf1 mRNA RPL32/Alb7相比是高度优越的。由图11和图12(HepG2和HeLa细胞)显见,表达相对于RPL32/ALB7以%给出。

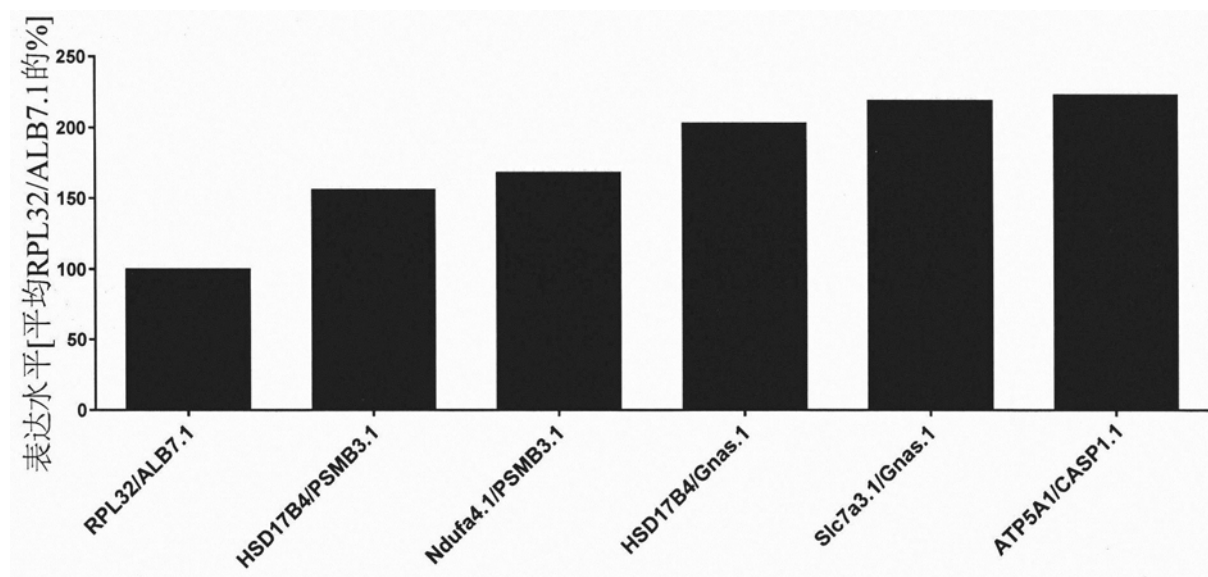


图1

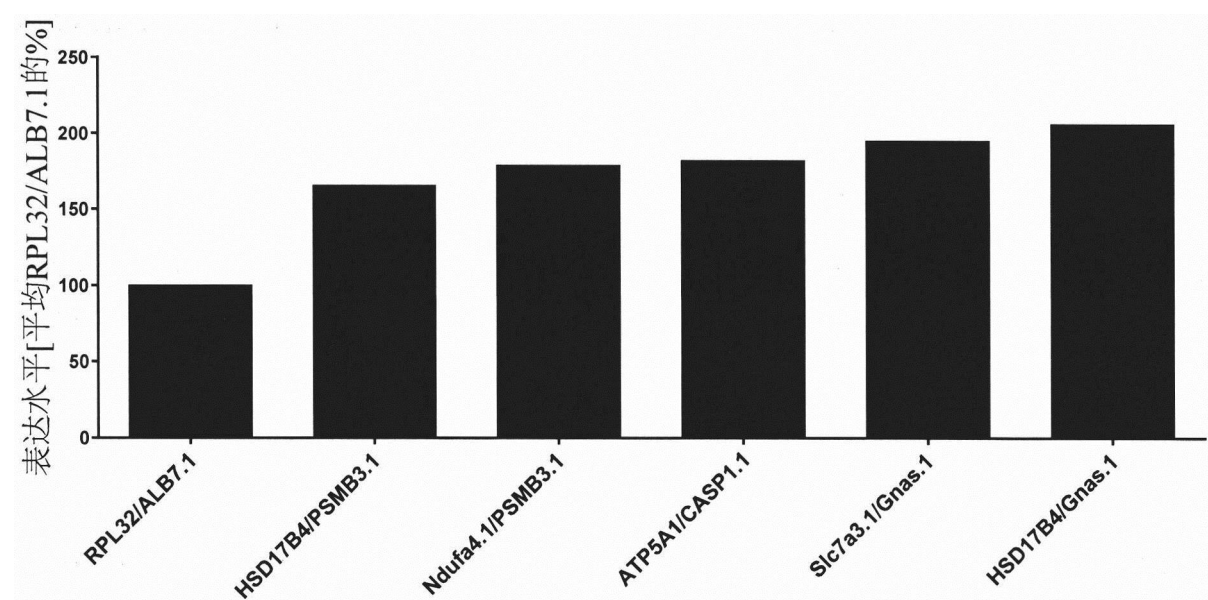


图2

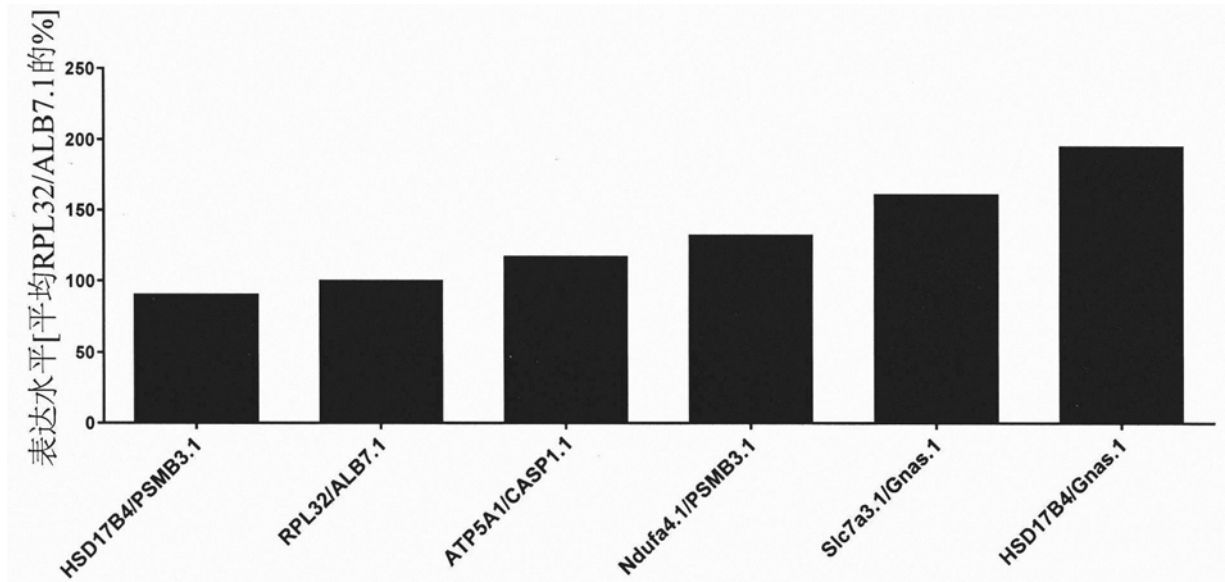


图3

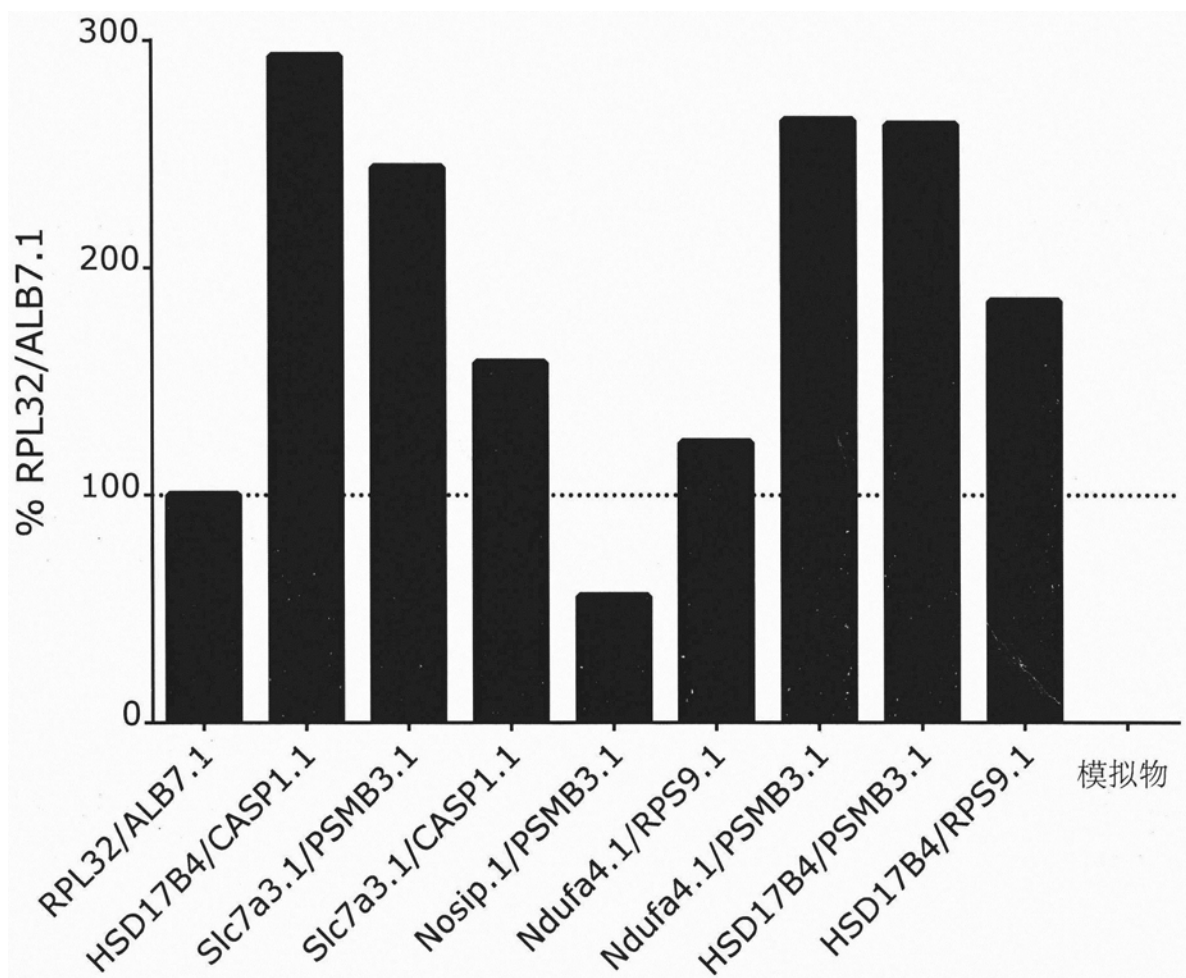


图4

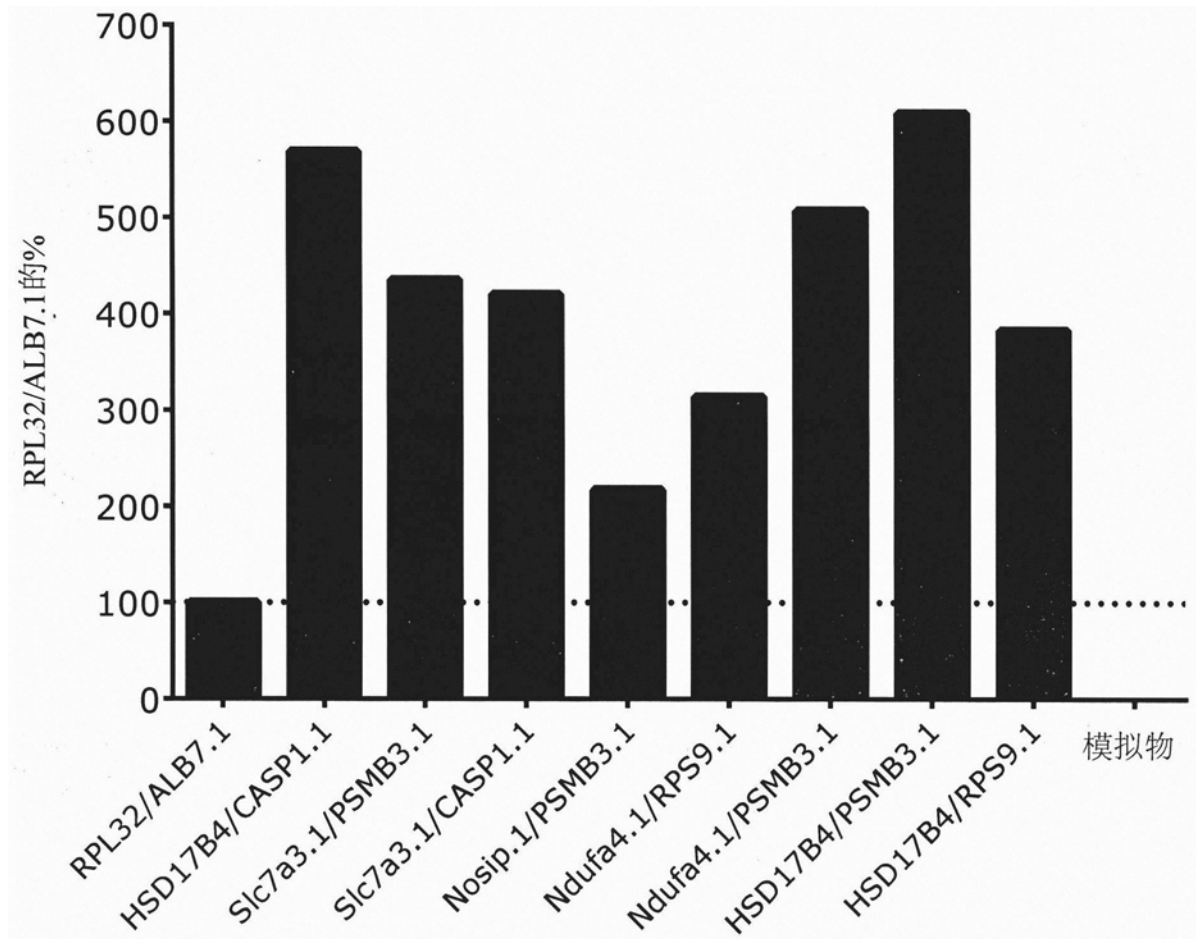


图5

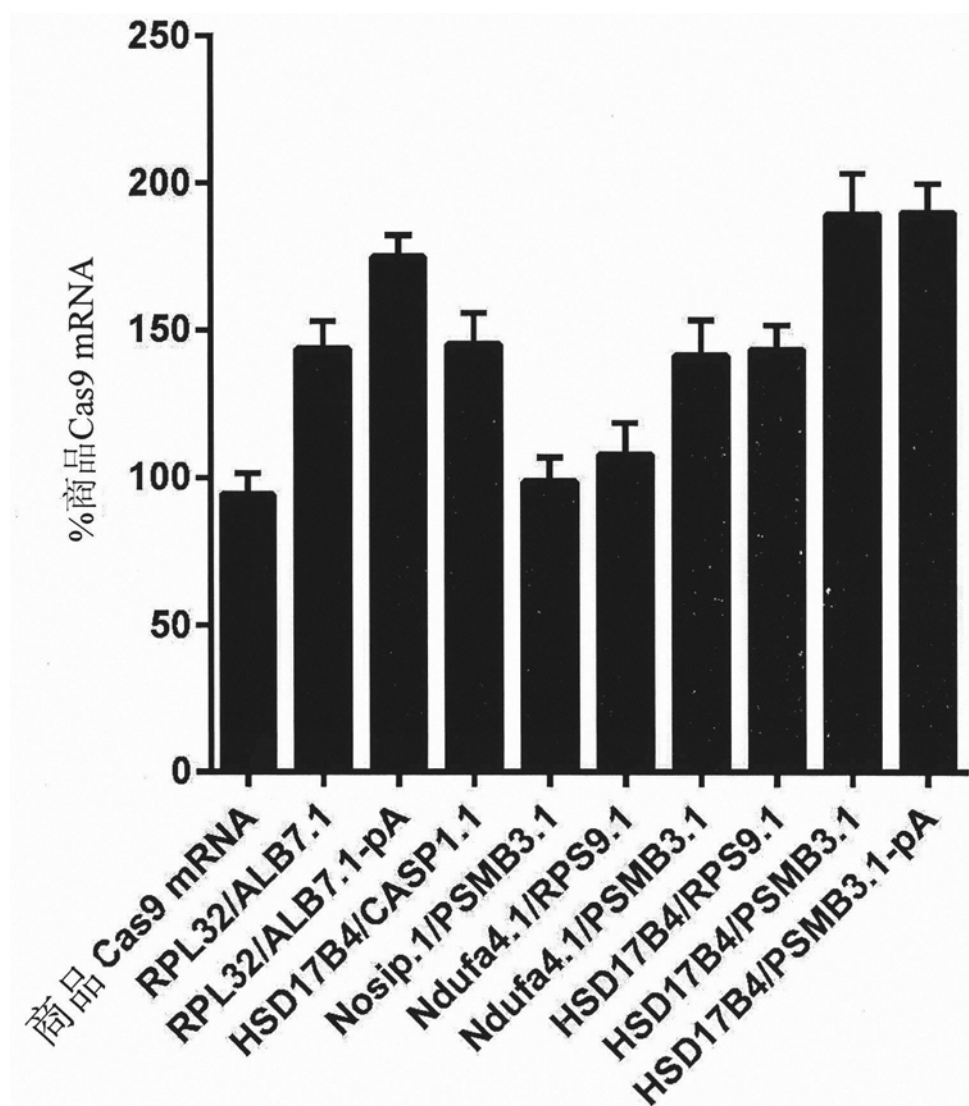


图6

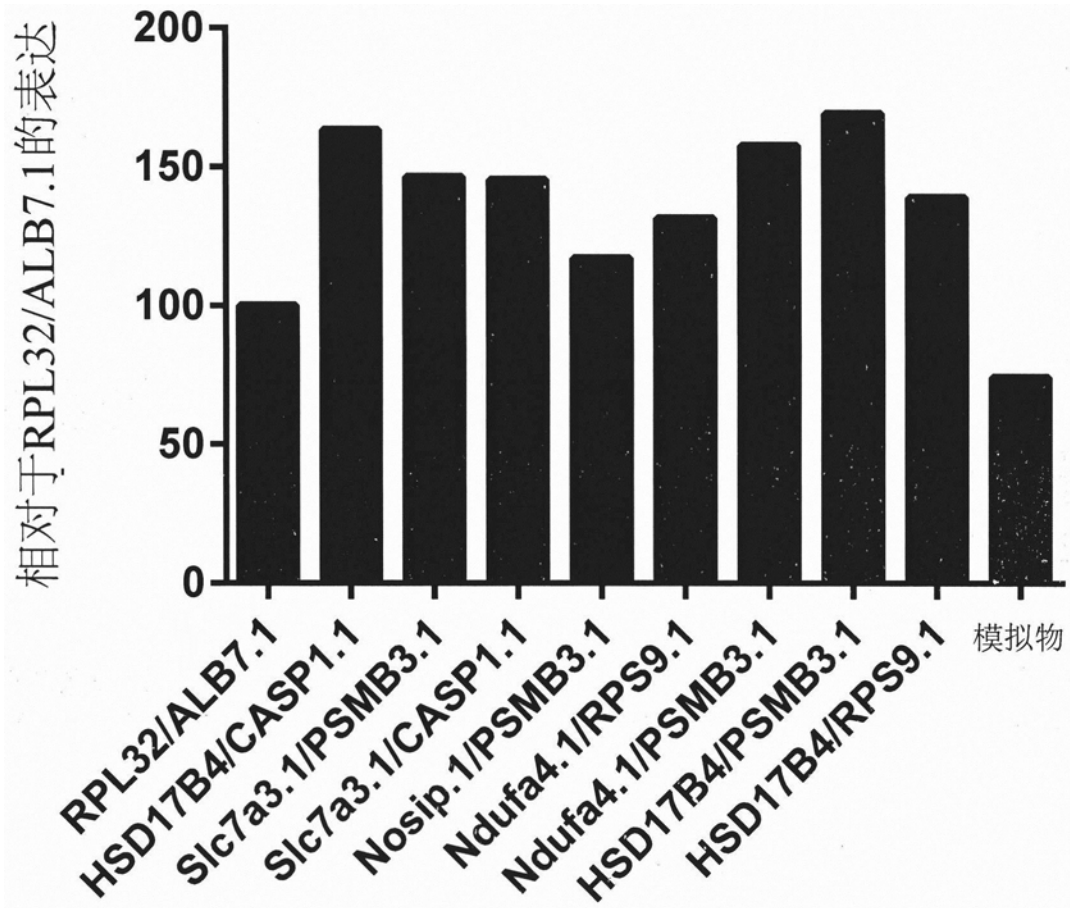


图7

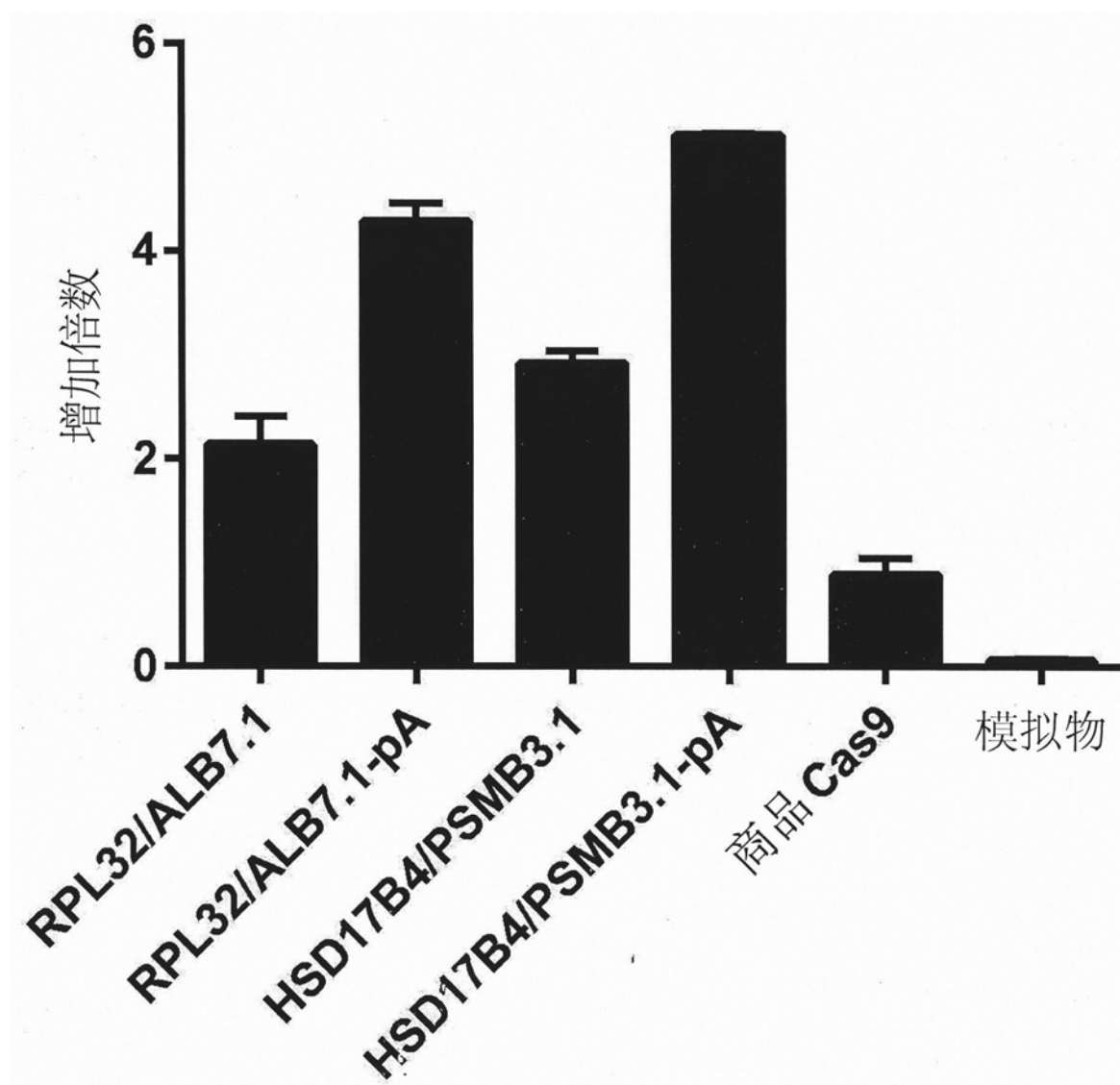


图8

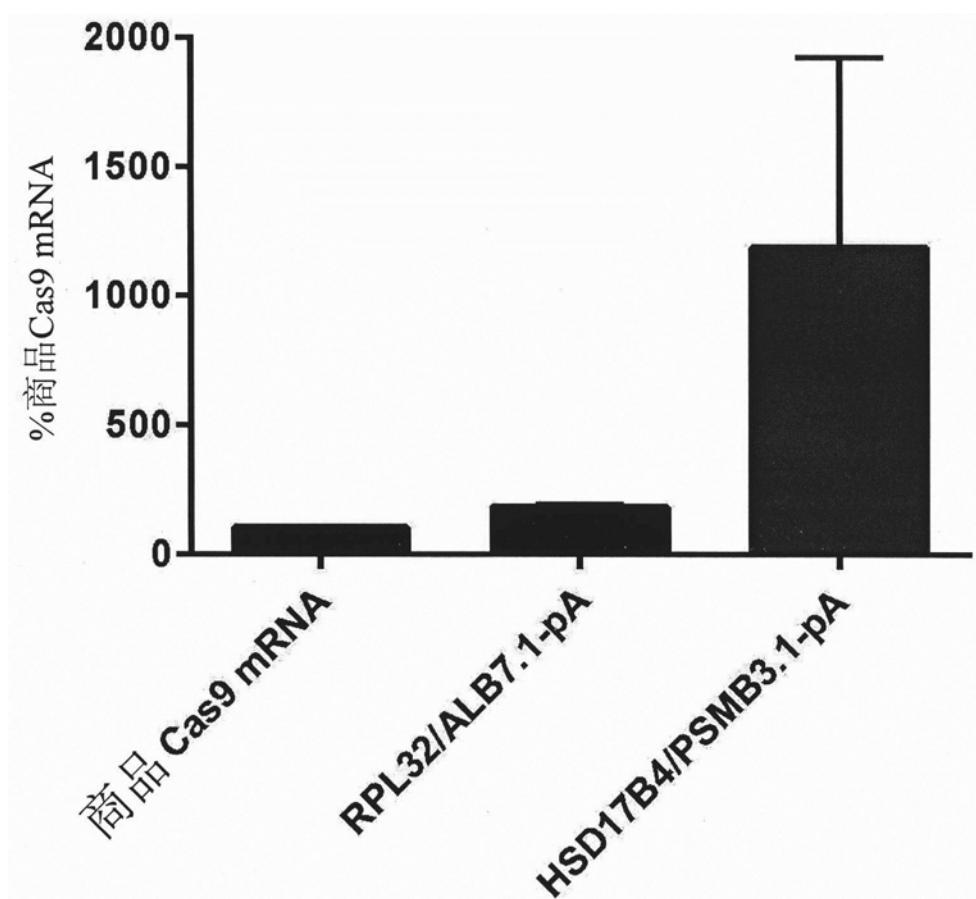


图9



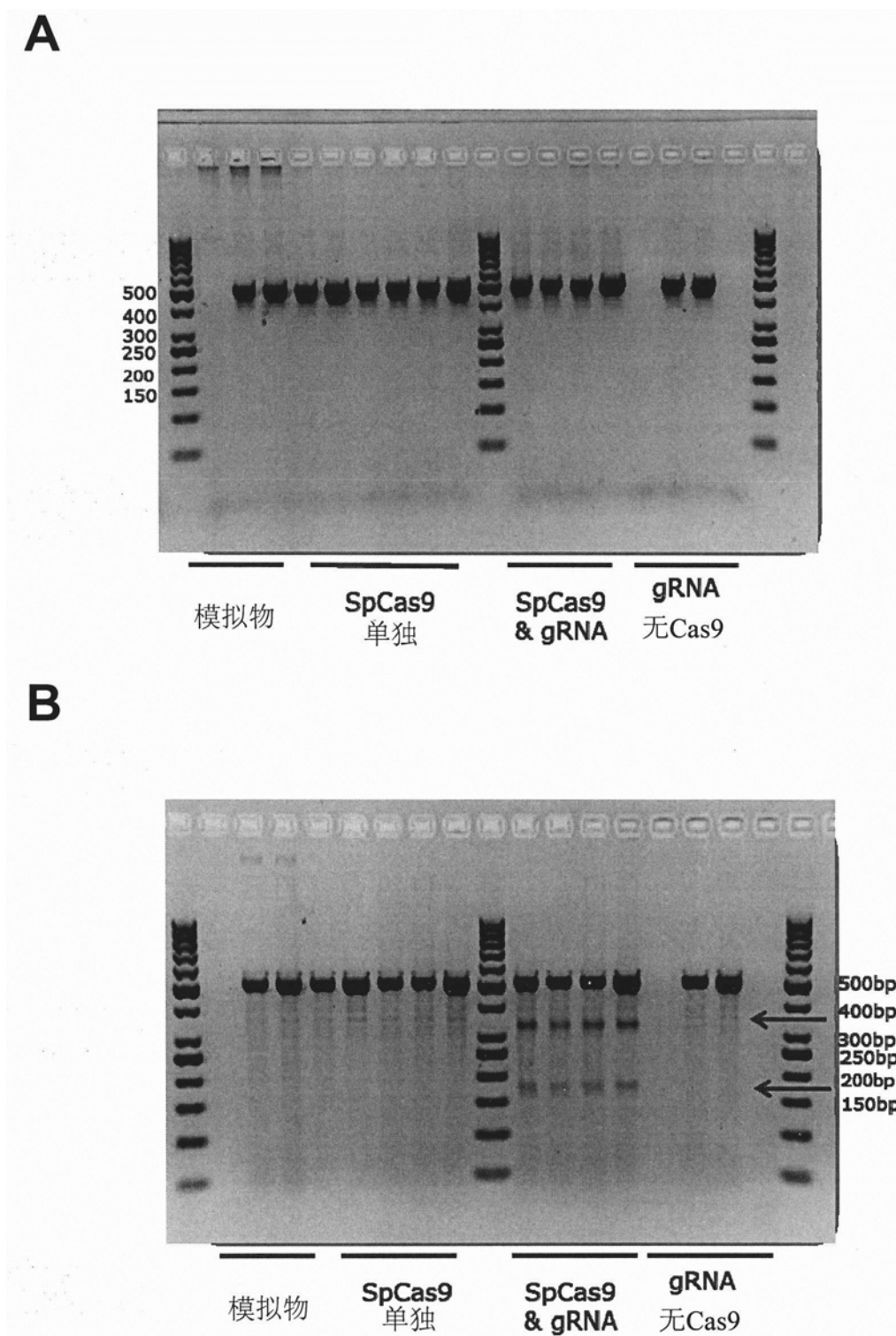


图10

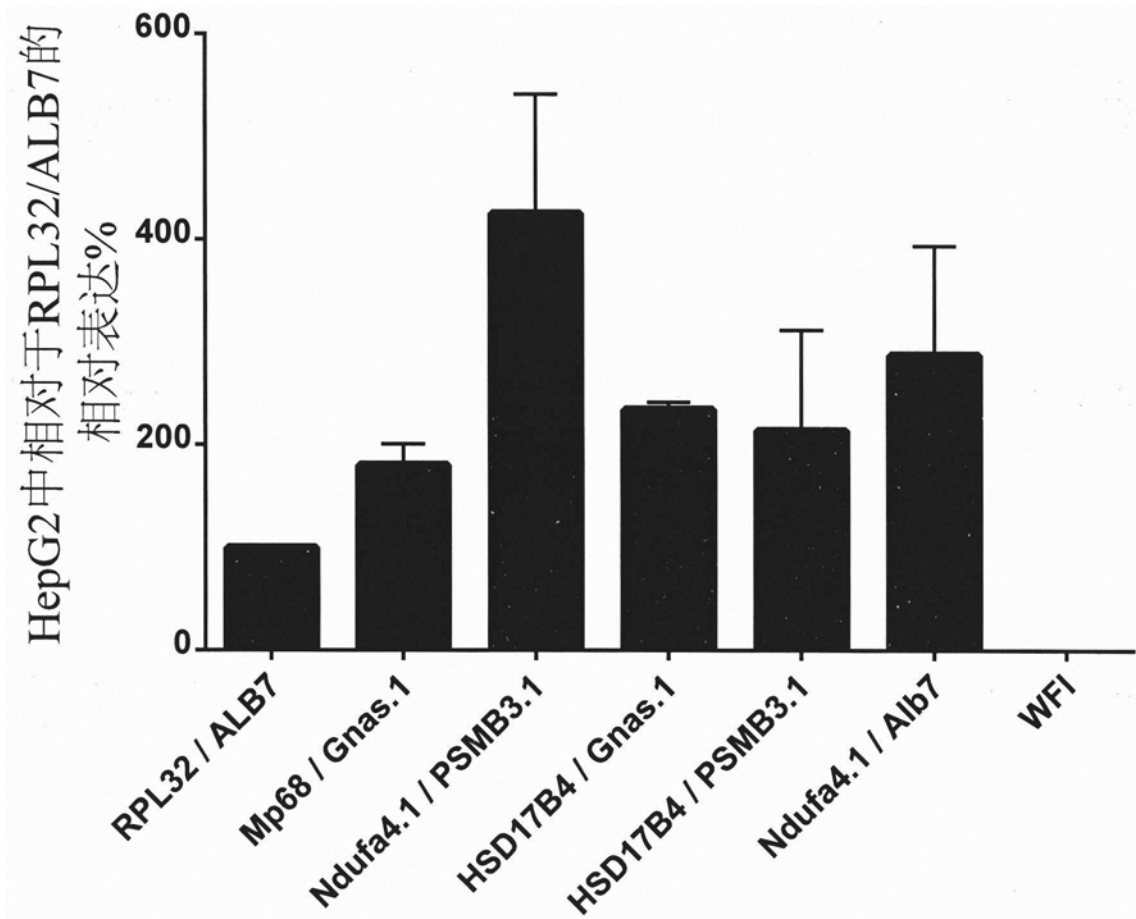


图11

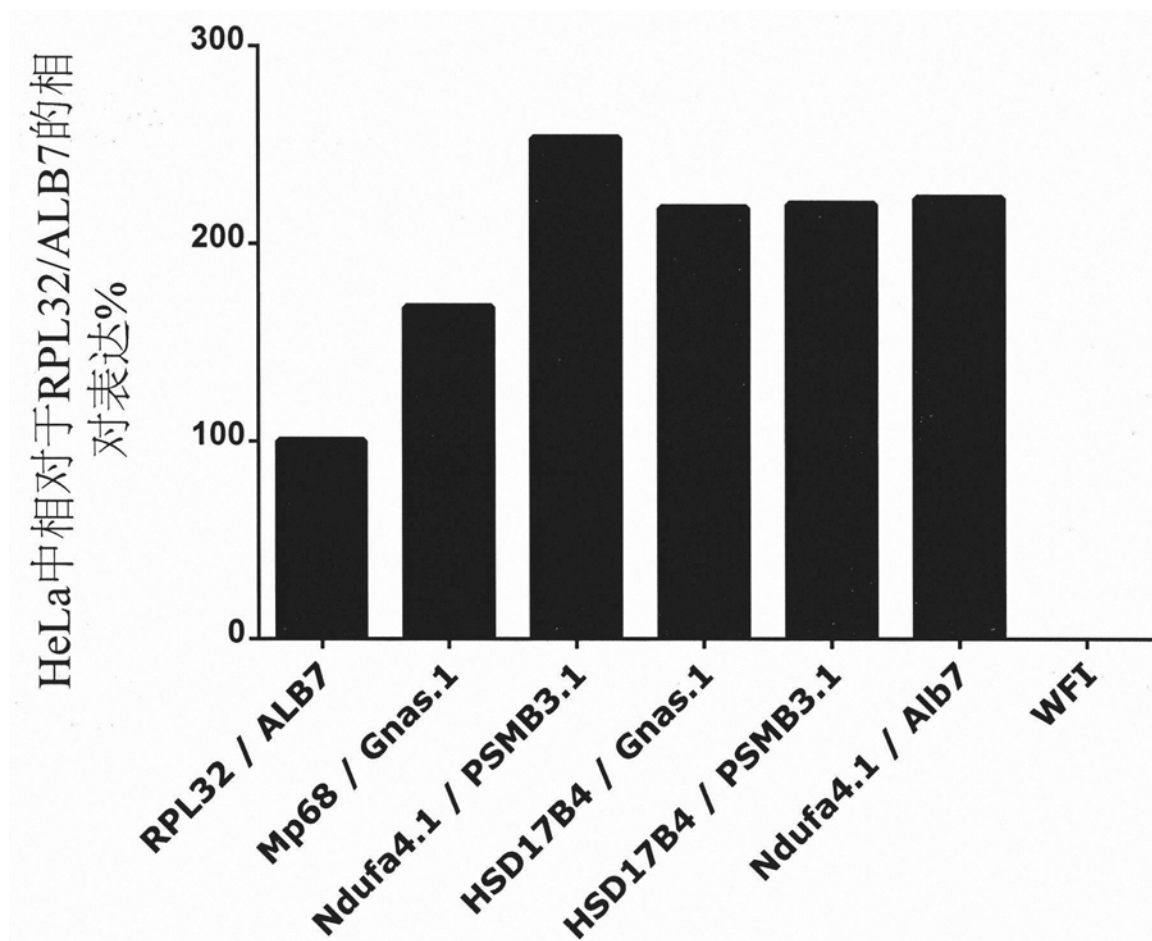


图12