



## (12) 发明专利申请

(10) 申请公布号 CN 114381470 A

(43) 申请公布日 2022. 04. 22

(21) 申请号 202210036661.4

A61P 29/00 (2006.01)

(22) 申请日 2016.08.22

A61P 35/00 (2006.01)

(30) 优先权数据

A61P 37/04 (2006.01)

PCT/EP2015/001755 2015.08.28 EP

(62) 分案原申请数据

201680049973.1 2016.08.22

(71) 申请人 库瑞瓦格股份公司

地址 德国图宾根

(72) 发明人 托马斯·施拉克

斯缔芬尼·格朗德

(74) 专利代理机构 中科专利商标代理有限责任

公司 11021

代理人 柴云峰 张莹

(51) Int. Cl.

C12N 15/67 (2006.01)

权利要求书11页 说明书144页

序列表34页 附图11页

(54) 发明名称

人工核酸分子

(57) 摘要

本发明涉及人工核酸分子,其包含至少一个可读框和至少一个3'-非翻译区元件(3'-UTR元件)和/或至少一个5'-非翻译区元件(5'-UTR元件),其中所述人工核酸分子的特征在于高翻译效率。所述翻译效率至少部分归因于所述5'-UTR元件或所述3'-UTR元件,或者归因于所述5'-UTR元件和所述3'-UTR元件二者。本发明还涉及所述人工核酸分子在基因治疗和/或基因疫苗接种中的用途。此外,提供了新型的3'-UTR元件和5'-UTR元件。

1. 一种人工核酸分子,其包含

- a. 至少一个可读框 (ORF); 和
- b. 至少一个3' -非翻译区元件 (3' -UTR元件) 和/或至少一个5' -非翻译区元件 (5' -UTR元件),

其中所述人工核酸分子的特征为高翻译效率。

2. 权利要求1的人工核酸分子,其中所述高翻译效率由所述至少一个3' -非翻译区元件 (3' -UTR元件) 和/或所述至少一个5' -非翻译区元件 (5' -UTR元件) 提供。

3. 权利要求1或权利要求2的人工核酸分子,其中所述人工核酸分子的翻译效率与参比核酸分子的翻译效率进行比较,

其中所述参比核酸分子包含至少一个可读框 (ORF), 所述可读框与所述人工核酸分子的至少一个ORF相同; 并且, 其中所述参比核酸分子不包含所述人工核酸分子的至少一个3' -非翻译区元件 (3' -UTR元件) 和/或所述人工核酸分子的至少一个5' -非翻译区元件 (5' -UTR元件),

其中所述人工核酸分子的翻译效率与所述参比核酸分子的翻译效率通过包括下述步骤的方法进行比较:

(i) 用所述人工核酸分子转染哺乳动物细胞, 并且在转染后特定时间点 (例如, 24或48小时) 测量由所述人工核酸分子的ORF编码的蛋白的表达量,

(ii) 用所述参比核酸分子转染哺乳动物细胞, 并且在转染后相同时间点测量所述参比核酸分子的ORF编码的蛋白的表达量,

(iii) 计算从所述人工核酸分子表达的蛋白量与从所述参比核酸分子表达的蛋白量的比率,

其中在 (iii) 中计算的比率 $\geq 1$ , 优选 $>1$ 。

4. 根据权利要求3所述的人工核酸分子, 其中所述ORF编码能被定量的蛋白, 优选报道蛋白。

5. 权利要求3或4的人工核酸分子, 其中所述参比核酸分子包含在所述人工核酸分子中不存在的3' -UTR元件,

或

其中所述参比核酸分子包含在所述人工核酸分子中不存在的5' -UTR元件。

6. 权利要求5的人工核酸分子, 其中所述参比核酸分子包含源自下述的3' -UTR元件: 白蛋白基因的3' -UTR, 优选对应于SEQ ID NO:207所示的序列的3' -UTR,

或

其中所述参比核酸分子包含源自下述的5' -UTR元件: TOP基因的5' -UTR, 优选缺少5' 端寡嘧啶束的人大核糖体蛋白32的5' -UTR, 优选对应于SEQ ID NO:208所示的序列的5' -UTR。

7. 权利要求3或4的人工核酸分子, 其中所述参比核酸分子不包含任意3' -非翻译区元件 (3' -UTR元件),

或

其中所述参比核酸分子不包含任意5' -非翻译区元件 (5' -UTR元件)。

8. 权利要求6的人工核酸分子, 其中所述方法进一步的特征在于:

(a) 在转染步骤中, 所述细胞是选自HDF、L929、HEP2G和HeLa细胞的组的哺乳动物细胞;

(b) 测量的时间点是转染后24小时;和

(c) 所述参比核酸分子是SEQ ID NO:205的mRNA (图1A)。

9. 根据权利要求1至4中任一项所述的人工核酸分子,其中所述可读框源自与所述至少一个3'-UTR元件和/或所述至少一个5'-UTR元件来源的基因不同的基因。

10. 根据权利要求1至9中任一项所述的人工核酸分子,其包含至少一个3'-UTR元件和至少一个5'-UTR元件。

11. 据权利要求10所述的人工核酸分子,其中所述至少一个可读框、所述至少一个3'-UTR元件和所述至少一个5'-UTR元件中的每一个彼此异源。

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不包含下述的3'-UTR和/或5'-UTR:核糖体蛋白S6,RPL36AL,rps16或核糖体蛋白L9,并且其中所述人工核酸分子的可读框不编码GFP蛋白。

13. 根据权利要求12所述的人工核酸分子,其中所述人工核酸分子的可读框不编码报道蛋白。

14. 根据权利要求1至13中任一项所述的人工核酸分子,其中所述至少一个3'-UTR元件和/或所述至少一个5'-UTR元件源自稳定mRNA。

15. 根据权利要求1至14中任一项所述的人工核酸分子,其中所述至少一个3'-UTR元件和/或所述至少一个5'-UTR元件包含下述核酸序列或由下述核酸序列组成:所述核酸序列源自真核生物蛋白编码基因的3'-UTR和/或5'-UTR,优选源自脊椎动物蛋白编码基因的3'-UTR和/或5'-UTR,更优选源自哺乳动物蛋白编码基因的3'-UTR和/或5'-UTR,甚至更优选源自灵长类蛋白编码基因的3'-UTR和/或5'-UTR,特别源自人或鼠蛋白编码基因的3'-UTR和/或5'-UTR。

16. 根据权利要求1至15中任一项所述的人工核酸分子,其中,与从缺少3'-UTR和/或至少一个5'-UTR的参比核酸分子的蛋白产生相比,所述至少一个3'-UTR元件和/或所述至少一个5'-UTR元件分别使所述人工核酸分子的翻译效率增加至少1.2倍,优选至少1.5倍,更优选至少2倍,甚至更优选至少2.5倍,和/或,其中,与从缺少3'-UTR和/或至少一个5'-UTR的参比核酸分子的蛋白产生相比,所述至少一个3'-UTR元件和/或所述至少一个5'-UTR元件为所述人工核酸分子提供至少1.5倍、优选至少2倍、更优选至少2.5倍的高翻译效率。

17. 根据权利要求1至16中任一项所述的人工核酸分子,其中所述至少一个3'-UTR元件和/或所述至少一个5'-UTR元件包含下述核酸序列或由下述核酸序列组成:所述核酸序列源自选自由下述组成的组的基因的转录本的3'-UTR和/或5'-UTR:ZNF460,TGM2,IL7R,BGN,TK1,RAB3B,CBX6,FZD2,COL8A1,NDUFS7,PHGDH,PLK2,TSP0,PTGS1,FBX032,NID2,ATP5D,EXOSC4,NOL9,UBB4B,VPS18,ORMDL2,FSCN1,TMEM33,TUBA4A,EMP3,TMEM201,CRIP2,BRAT1,SERPINH1,CD9,DPYSL2,CDK9,TFRC,PSMB3 5'-UTR,FASN,PSMB6,PRSS56,KPNA6,SFT2D2,PARD6B,LPP,SPARC,SCAND1,VASN,SLC26A1,LCLAT1,FBXL18,SLC35F6,RAB3D,MAP1B,VMA21,CYBA,SEZ6L2,PCOLCE,VTN,ALDH16A1,RAVER1,KPNA6,SERINC5,JUP,CPN2,CRIP2,EPT1,PNP0,SSSCA1,POLR2L,LIN7C,UQCR10,PYCRL,AMN,MAP1S,NDUFS7,PHGDH,TSP0,ATP5D,EXOSC4,TUBB4B,TUBA4A,EMP3,CRIP2,BRAT1,CD9,CDK9,PSMB3,PSMB6,PRSS56,SCAND1,AMN,CYBA,PCOLCE,MAP1S,VTN,ALDH16A1 (全部优选是人的) 和Dpys12,Ccnd1,Acox2,Cbx6,Ubc,Ldlr,Nudt22,Pcyox11,Ankrd1,Tmem37,Tsps14,Slc7a3,Cst6,Aacs,Nosip,Itga7,Ccnd2,

Ebp, Sf3b5, Fasn, Hmgcs1, Osr1, Lmnb1, Vma21, Kif20a, Cdca8, Slc7a1, Ubqln2, Prps2, Shmt2, Aurkb, Fignl1, Cad, Anln, Slfn9, Ncaph, Pole, Uhrf1, Gja1, Fam64a, Kif2c, Tspan10, Scand1, Gpr84, Fads3, Cers6, Cxcr4, Gprc5c, Fen1, Cspg4, Mrpl34, Comtd1, Armc6, Emr4, Atp5d, 1110001J03Rik, Csf2ra, Aarsd1, Kif22, Cth, Tpgs1, Ccl17, Alkbh7, Ms4a8a, Acox2, Ubc, Slpi, Pcyox11, Igf2bp1, Tmem37, Slc7a3, Cst6, Ebp, Sf3b5, Plk1, Cdca8, Kif22, Cad, Cth, Pole, Kif2c, Scand1, Gpr84, Tpgs1, Ccl17, Alkbh7, Ms4a8a, Mrpl34, Comtd1, Armc6, Atp5d, 1110001J03Rik, Nudt22, Aarsd1 (全部优选是小鼠的)。

18. 根据权利要求17所述的人工核酸分子, 其中所述至少一个5' -UTR元件包含源自选自自由下述组成的组的基因的转录本的5' -UTR的核酸序列: ZNF460-5' -UTR, TGM2-5' -UTR, IL7R-5' -UTR, BGN-5' -UTR, TK1-5' -UTR, RAB3B-5' -UTR, CBX6-5' -UTR, FZD2-5' -UTR, COL8A1-5' -UTR, NDUFS7-5' -UTR, PHGDH-5' -UTR, PLK2-5' -UTR, TSP0-5' -UTR, PTGS1-5' -UTR, FBX032-5' -UTR, NID2-5' -UTR, ATP5D-5' -UTR, EXOSC4-5' -UTR, NOL9-5' -UTR, UBB4B-5' -UTR, VPS18-5' -UTR, ORMDL2-5' -UTR, FSCN1-5' -UTR, TMEM33-5' -UTR, TUBA4A-5' -UTR, EMP3-5' -UTR, TMEM201-5' -UTR, CRIP2-5' -UTR, BRAT1-5' -UTR, SERPINH1-5' -UTR, CD9-5' -UTR, DPYSL2-5' -UTR, CDK9-5' -UTR, TFRC-5' -UTR, PSMB3-5' -UTR, FASN-5' -UTR, PSMB6-5' -UTR, PRSS56-5' -UTR, KPNA6-5' -UTR, SFT2D2-5' -UTR, PARD6B-5' -UTR, LPP-5' -UTR, SPARC-5' -UTR, SCAND1-5' -UTR, VASN-5' -UTR, SLC26A1-5' -UTR, LCLAT1-5' -UTR, FBXL18-5' -UTR, SLC35F6-5' -UTR, RAB3D-5' -UTR, MAP1B-5' -UTR, VMA21-5' -UTR, CYBA-5' -UTR, SEZ6L2-5' -UTR, PCOLCE-5' -UTR, VTN-5' -UTR, ALDH16A1-5' -UTR, RAVR1-5' -UTR, KPNA6-5' -UTR, SERINC5-5' -UTR, JUP-5' -UTR, CPN2-5' -UTR, CRIP2-5' -UTR, EPT1-5' -UTR, PNPO-5' -UTR, SSSCA1-5' -UTR, POLR2L-5' -UTR, LIN7C-5' -UTR, UQCR10-5' -UTR, PYCRL-5' -UTR, AMN-5' -UTR, MAP1S-5' -UTR, (全部优选是人的) 和Dpys12-5' -UTR, Ccnd1-5' -UTR, Acox2-5' -UTR, Cbx6-5' -UTR, Ubc-5' -UTR, Ldlr-5' -UTR, Nudt22-5' -UTR, Pcyox11-5' -UTR, Ankrd1-5' -UTR, Tmem37-5' -UTR, Tspyl4-5' -UTR, Slc7a3-5' -UTR, Cst6-5' -UTR, Aacs-5' -UTR, Nosip-5' -UTR, Itga7-5' -UTR, Ccnd2-5' -UTR, Ebp-5' -UTR, Sf3b5-5' -UTR, Fasn-5' -UTR, Hmgcs1-5' -UTR, Osr1-5' -UTR, Lmnb1-5' -UTR, Vma21-5' -UTR, Kif20a-5' -UTR, Cdca8-5' -UTR, Slc7a1-5' -UTR, Ubqln2-5' -UTR, Prps2-5' -UTR, Shmt2-5' -UTR, Aurkb-5' -UTR, Fignl1-5' -UTR, Cad-5' -UTR, Anln-5' -UTR, Slfn9-5' -UTR, Ncaph-5' -UTR, Pole-5' -UTR, Uhrf1-5' -UTR, Gja1-5' -UTR, Fam64a-5' -UTR, Kif2c-5' -UTR, Tspan10-5' -UTR, Scand1-5' -UTR, Gpr84-5' -UTR, Fads3-5' -UTR, Cers6-5' -UTR, Cxcr4-5' -UTR, Gprc5c-5' -UTR, Fen1-5' -UTR, Cspg4-5' -UTR, Mrpl34-5' -UTR, Comtd1-5' -UTR, Armc6-5' -UTR, Emr4-5' -UTR, Atp5d-5' -UTR, 1110001J03Rik-5' -UTR, Csf2ra-5' -UTR, Aarsd1-5' -UTR, Kif22-5' -UTR, Cth-5' -UTR, Tpgs1-5' -UTR, Ccl17-5' -UTR, Alkbh7-5' -UTR, Ms4a8a-5' -UTR (全部优选是小鼠的)。

19. 根据权利要求17所述的人工核酸分子, 其中所述至少一个3' -UTR元件包含源自选自自由下述组成的组的基因的转录本的3' -UTR的核酸序列: NDUFS7-3' -UTR, PHGDH-3' -UTR, TSP0-3' -UTR, ATP5D-3' -UTR, EXOSC4-3' -UTR, TUBB4B-3' -UTR, TUBA4A-3' -UTR, EMP3-3' -UTR, CRIP2-3' -UTR, BRAT1-3' -UTR, CD9-3' -UTR, CDK9-3' -UTR, PSMB3-3' -UTR, PSMB6-3' -UTR, PRSS56-3' -UTR, SCAND1-3' -UTR, AMN-3' -UTR, CYBA-3' -UTR, PCOLCE-3' -UTR, MAP1S-3' -UTR, VTN-3' -UTR, ALDH16A1-3' -UTR (全部优选是人的) 和Acox2-3' -UTR, Ubc-3' -UTR,



Slpi-3'-UTR, Pcyox11-3'-UTR, Igf2bp1-3'-UTR, Tmem37-3'-UTR, Slc7a3-3'-UTR, Cst6-3'-UTR, Ebp-3'-UTR, Sf3b5-3'-UTR, Plk1-3'-UTR, Cdca8-3'-UTR, Kif22-3'-UTR, Cad-3'-UTR, Cth-3'-UTR, Pole-3'-UTR, Kif2c-3'-UTR, Scand1-3'-UTR, Gpr84-3'-UTR, Tpgs1-3'-UTR, Cc117-3'-UTR, Alkbh7-3'-UTR, Ms4a8a-3'-UTR, Mrpl34-3'-UTR, Comtd1-3'-UTR, Armc6-3'-UTR, Atp5d-3'-UTR, 1110001J03Rik-3'-UTR, Nudt22-3'-UTR, Aarsd1-3'-UTR, (全部优选是小鼠的)。

20. 根据权利要求17-19中任一项所述的人工核酸分子, 其中所述至少一个5'-UTR元件包含源自选自由下述组成的组的人基因的转录本的5'-UTR的核酸序列: ZNF460-5'-UTR, TGM2-5'-UTR, IL7R-5'-UTR, COL8A1-5'-UTR, NDUFS7-5'-UTR, PLK2-5'-UTR, FBX032-5'-UTR, ATP5D-5'-UTR, TUBB4B-5'-UTR, ORMDL2-5'-UTR, FSCN1-5'-UTR, CD9-5'-UTR, PYSL2-5'-UTR, PSMB3-5'-UTR, PSMB6-5'-UTR, KPNA6-5'-UTR, SFT2D2-5'-UTR, LCLAT1-5'-UTR, FBXL18-5'-UTR, SLC35F6-5'-UTR, VMA21-5'-UTR, SEZ6L2-5'-UTR, PCOLCE-5'-UTR, VTN-5'-UTR, ALDH16A1-5'-UTR, KPNA6-5'-UTR, JUP-5'-UTR, CPN2-5'-UTR, PNPO-5'-UTR, SSSCA1-5'-UTR, POLR2L-5'-UTR, LIN7C-5'-UTR, UQCR10-5'-UTR, PYCRL-5'-UTR, AMN-5'-UTR, MAP1S-5'-UTR; 或源自选自由下述组成的组的小鼠基因的转录本的5'-UTR的核酸序列: Dpysl2-5'-UTR, Acox2-5'-UTR, Ubc-5'-UTR, Nudt22-5'-UTR, Pcyox11-5'-UTR, Ankrd1-5'-UTR, Tspyl4-5'-UTR, Slc7a3-5'-UTR, Aacs-5'-UTR, Nosip-5'-UTR, Itga7-5'-UTR, Ccnd2-5'-UTR, Ebp-5'-UTR, Sf3b5-5'-UTR, Fasn-5'-UTR, Hmgcs1-5'-UTR, Osr1-5'-UTR, Lmnb1-5'-UTR, Vma21-5'-UTR, Kif20a-5'-UTR, Cdca8-5'-UTR, Slc7a1-5'-UTR, Ubqln2-5'-UTR, Prps2-5'-UTR, Shmt2-5'-UTR, Fignl1-5'-UTR, Cad-5'-UTR, Anln-5'-UTR, Slfn9-5'-UTR, Ncaph-5'-UTR, Pole-5'-UTR, Uhrf1-5'-UTR, Gja1-5'-UTR, Fam64a-5'-UTR, Tspan10-5'-UTR, Scand1-5'-UTR, Gpr84-5'-UTR, Cers6-5'-UTR, Cxcr4-5'-UTR, Gprc5c-5'-UTR, Fen1-5'-UTR, Cspg4-5'-UTR, Mrpl34-5'-UTR, Comtd1-5'-UTR, Armc6-5'-UTR, Emr4-5'-UTR, Atp5d-5'-UTR, Csf2ra-5'-UTR, Aarsd1-5'-UTR, Cth-5'-UTR, Tpgs1-5'-UTR, Cc117-5'-UTR, Alkbh7-5'-UTR, Ms4a8a-5'-UTR。

21. 根据权利要求17-19中任一项所述的人工核酸分子, 其中所述至少一个3'-UTR元件包含源自选自由下述组成的组的人基因的转录本的3'-UTR的核酸序列: NDUFS7-3'-UTR, PHGDH-3'-UTR, TSP0-3'-UTR, ATP5D-3'-UTR, EXOSC4-3'-UTR, TUBB4B-3'-UTR, TUBA4A-3'-UTR, EMP3-3'-UTR, CRIP2-3'-UTR, BRAT1-3'-UTR, PSMB3-3'-UTR, PSMB6-3'-UTR, SCAND1-3'-UTR, AMN-3'-UTR, CYBA-3'-UTR, PCOLCE-3'-UTR, MAP1S-3'-UTR, VTN-3'-UTR, ALDH16A1-3'-UTR; 或源自选自由下述组成的组的小鼠基因的转录本的3'-UTR的核酸序列: Acox2-3'-UTR, Ubc-3'-UTR, Slpi-3'-UTR, Igf2bp1-3'-UTR, Tmem37-3'-UTR, Slc7a3-3'-UTR, Cst6-3'-UTR, Ebp-3'-UTR, Sf3b5-3'-UTR, Cdca8-3'-UTR, Kif22-3'-UTR, Cad-3'-UTR, Pole-3'-UTR, Kif2c-3'-UTR, Scand1-3'-UTR, Gpr84-3'-UTR, Tpgs1-3'-UTR, Cc117-3'-UTR, Alkbh7-3'-UTR, Ms4a8a-3'-UTR, Mrpl34-3'-UTR, Comtd1-3'-UTR, Armc6-3'-UTR, Atp5d-3'-UTR, 1110001J03Rik-3'-UTR, Nudt22-3'-UTR。

22. 根据权利要求1-21中任一项所述的人工核酸分子, 其中所述至少一个3'-UTR元件包含下述核酸序列或由下述核酸序列组成: 所述核酸序列与选自SEQ ID NOs: 152-204组成的组的核酸序列具有至少约50%, 优选至少约60%, 优选至少约70%, 更优选至少约

80%，更优选至少约90%，甚至更优选至少约95%，甚至更优选至少约99%的同一性，或者，其中所述至少一个3'-UTR元件包含下述核酸序列的片段或由下述核酸序列的片段组成：所述核酸序列的片段与选自SEQ ID NOs:152-204组成的组的核酸序列的片段具有至少约40%，优选至少约50%，优选至少约60%，优选至少约70%，更优选至少约80%，更优选至少约90%，甚至更优选至少约95%，甚至更优选至少约99%的同一性。

23. 根据权利要求1-22中任一项所述的人工核酸分子，其中所述至少一个5'-UTR元件包含下述核酸序列或由下述核酸序列组成：所述核酸序列与选自SEQ ID NOs:1-151组成的组的核酸序列具有至少约50%，优选至少约60%，优选至少约70%，更优选至少约80%，更优选至少约90%，甚至更优选至少约95%，甚至更优选至少约99%的同一性，或者，其中所述至少一个5'-UTR元件包含下述核酸序列的片段或由下述核酸序列的片段组成：所述核酸序列的片段与选自SEQ ID NOs:1-151组成的组的核酸序列具有至少约40%，优选至少约50%，优选至少约60%，优选至少约70%，更优选至少约80%，更优选至少约90%，甚至更优选至少约95%，甚至更优选至少约99%的同一性。

24. 根据权利要求22或23所述的人工核酸分子，其中所述片段呈现3-约500个核苷酸、优选5-约150个核苷酸、更优选10-100个核苷酸、甚至更优选15-90个核苷酸、最优选20-70个核苷酸的长度。

25. 根据权利要求1-24中任一项所述的人工核酸分子，其中所述至少一个3'-UTR元件和/或所述至少一个5'-UTR元件呈现3-约500个核苷酸、优选5-约150个核苷酸、更优选10-100个核苷酸、甚至更优选15-90个核苷酸、最优选20-70个核苷酸的长度。

26. 根据权利要求1-25中任一项所述的人工核酸分子，其还包含：

c. 聚腺苷酸序列和/或聚腺苷酸化信号。

27. 根据权利要求26所述的人工核酸分子，其中所述聚腺苷酸序列或所述聚腺苷酸化信号位于所述3'-UTR元件的3'。

28. 根据权利要求26或27所述的人工核酸分子，其中所述聚腺苷酸化信号包含共有序列NN(U/T)ANA，其中N=A或U，优选为AA(U/T)AAA或A(U/T)(U/T)AAA。

29. 根据权利要求26-28中任一项所述的人工核酸分子，其中所述聚腺苷酸化信号，优选所述共有序列NNUANA，位于所述3'-UTR元件的3'-端下游少于约50个核酸。

30. 根据权利要求26-28中任一项所述的人工核酸分子，其中所述聚腺苷酸序列具有约20-约300个腺嘌呤核苷酸、优选约40-约200个腺嘌呤核苷酸、更优选约50-约100个腺嘌呤核苷酸、甚至更优选约60-约70个腺嘌呤核苷酸的长度。

31. 根据权利要求1-30中任一项所述的人工核酸分子，其还包含5'-帽结构、聚胞苷酸序列、组蛋白茎环和/或IRES-基序。

32. 根据权利要求31所述的人工核酸分子，其中所述组蛋白茎环包含SEQ ID NO:209所述的序列。

33. 根据权利要求1-32中任一项所述的人工核酸分子，其中所述核酸包含启动子。

34. 根据权利要求1-33中任一项所述的人工核酸分子，其中所述核酸包含5'-TOP UTR。

35. 根据权利要求1-34中任一项所述的人工核酸分子，其中所述核酸包含3'-UTR，所述3'-UTR包含源自白蛋白基因3'-UTR的核酸序列或由源自白蛋白基因3'-UTR的核酸序列组成。

36. 根据权利要求1-35中任一项所述的人工核酸分子,其中所述人工核酸分子,优选所述可读框,是至少部分G/C修饰的,优选其中所述可读框的G/C含量与野生型可读框相比是增加的。

37. 根据权利要求1-36中任一项所述的人工核酸分子,其中所述可读框包含密码子优化的区域,优选地,其中所述可读框是密码子优化的。

38. 根据权利要求1-37中任一项所述的人工核酸分子,其是RNA,优选是mRNA分子。

39. 包含根据权利要求1-38中任一项所述的人工核酸分子的载体。

40. 根据权利要求39所述的载体,其是DNA载体。

41. 根据权利要求39或40所述的载体,其是质粒载体或病毒载体,优选是质粒载体。

42. 根据权利要求39-41中任一项所述的载体,其是环形分子。

43. 根据权利要求42所述的载体,其中编码链的所述聚腺苷酸序列、所述聚胞苷酸序列、所述组蛋白茎环或所述3' -UTR元件在5' → 3' 方向上后接用于使环形载体分子线性化的限制性位点。

44. 包含根据权利要求1-38中任一项所述的人工核酸分子或权利要求39-43中任一项的载体的细胞。

45. 根据权利要求44的细胞,其是哺乳动物细胞。

46. 根据权利要求44或45所述的细胞,其是哺乳动物受试者的细胞,优选是哺乳动物受试者分离的细胞,优选是人受试者分离的细胞。

47. 药物组合物,其包含根据权利要求1-38中任一项所述的人工核酸分子、根据权利要求39-43中任一项所述的载体或根据权利要求44-46中任一项所述的细胞。

48. 根据权利要求47所述的药物组合物,其还包含一种或多种药用载体、稀释剂和/或赋形剂和/或一种或多种佐剂。

49. 根据权利要求1-38中任一项所述的人工核酸分子、根据权利要求39-43中任一项所述的载体、根据权利要求44-46中任一项所述的细胞或根据权利要求47或48所述的药物组合物,其用作药物。

50. 根据权利要求1-38中任一项所述的人工核酸分子、根据权利要求39-43中任一项所述的载体、根据权利要求44-46中任一项所述的细胞或根据权利要求47或48所述的药物组合物,其用作疫苗或用于基因治疗。

51. 一种用于治疗或预防病症的方法,所述方法包括向有此需要的受试者施用根据权利要求1-38中任一项所述的人工核酸分子、根据权利要求39-43中任一项所述的载体、根据权利要求44-46中任一项所述的细胞或根据权利要求47或48所述的药物组合物。

52. 一种治疗或预防病症的方法,所述方法包括用根据权利要求1-38中任一项所述的人工核酸分子或权利要求39-43中任一项所述的载体转染细胞。

53. 根据权利要求52所述的方法,其中转染细胞在体外/离体进行,并且将转染的细胞施用给有此需要的受试者,优选是人患者。

54. 根据权利要求53所述的方法,其中要在体外转染的细胞是受试者分离的细胞,优选是人患者分离的细胞。

55. 根据权利要求51-55中任一项所述的方法,其是疫苗接种方法或基因治疗方法。

56. 一种用于增加人工核酸分子、优选mRNA分子或载体的翻译效率的方法,所述方法包

括连接可读框与3'-UTR元件和/或5'-UTR元件的步骤,其中所述3'-UTR元件和/或所述5'-UTR元件增加从所得的人工核酸分子的翻译效率,从而得到根据权利要求1-38中任一项所述的人工核酸分子,优选mRNA分子,或根据权利要求39-43中任一项所述的载体。

57. 根据权利要求52所述的用于增加人工核酸分子、优选mRNA分子或载体的翻译效率的方法,其中所述3'-UTR元件和/或所述5'-UTR元件包含下述核酸序列或由下述核酸序列组成:所述核酸序列源自选自由下述组成的组的基因的转录本的3'-UTR和/或5'-UTR: ZNF460, TGM2, IL7R, BGN, TK1, RAB3B, CBX6, FZD2, COL8A1, NDUFS7, PHGDH, PLK2, TSP0, PTGS1, FBX032, NID2, ATP5D, EXOSC4, NOL9, UBB4B, VPS18, ORMDL2, FSCN1, TMEM33, TUBA4A, EMP3, TMEM201, CRIP2, BRAT1, SERPINH1, CD9, DPYSL2, CDK9, TFRC, PSMB3 5'-UTR, FASN, PSMB6, PRSS56, KPNA6, SFT2D2, PARD6B, LPP, SPARC, SCAND1, VASN, SLC26A1, LCLAT1, FBXL18, SLC35F6, RAB3D, MAP1B, VMA21, CYBA, SEZ6L2, PCOLCE, VTN, ALDH16A1, RAVER1, KPNA6, SERINC5, JUP, CPN2, CRIP2, EPT1, PNP0, SSSCA1, POLR2L, LIN7C, UQCR10, PYCRL, AMN, MAP1S, NDUFS7, PHGDH, TSP0, ATP5D, EXOSC4, TUBB4B, TUBA4A, EMP3, CRIP2, BRAT1, CD9, CDK9, PSMB3, PSMB6, PRSS56, SCAND1, AMN, CYBA, PCOLCE, MAP1S, VTN, ALDH16A1 (全部优选是人的) 和 Dpysl2, Ccnd1, Acox2, Cbx6, Ubc, Ldlr, Nudt22, Pcyox11, Ankrd1, Tmem37, Tspyl4, Slc7a3, Cst6, Aacs, Nosip, Itga7, Ccnd2, Ebp, Sf3b5, Fasn, Hmgcs1, Osrl, Lmnb1, Vma21, Kif20a, Cdca8, Slc7a1, Ubqln2, Prps2, Shmt2, Aurkb, Fignl1, Cad, Anln, Slfn9, Ncaph, Pole, Uhrf1, Gja1, Fam64a, Kif2c, Tspan10, Scand1, Gpr84, Fads3, Cers6, Cxcr4, Gprc5c, Fen1, Cspg4, Mrpl34, Comtd1, Armc6, Emr4, Atp5d, 1110001J03Rik, Csf2ra, Aarsd1, Kif22, Cth, Tpgs1, Cc117, Alkbh7, Ms4a8a, Acox2, Ubc, Slpi, Pcyox11, Igf2bp1, Tmem37, Slc7a3, Cst6, Ebp, Sf3b5, Plk1, Cdca8, Kif22, Cad, Cth, Pole, Kif2c, Scand1, Gpr84, Tpgs1, Cc117, Alkbh7, Ms4a8a, Mrpl34, Comtd1, Armc6, Atp5d, 1110001J03Rik, Nudt22, Aarsd1 (全部优选是小鼠的)。

58. 3'-UTR元件和/或5'-UTR元件用于增加人工核酸分子、优选mRNA分子或载体的翻译效率的用途,其中所述3'-UTR元件和/或所述5'-UTR元件包含下述核酸序列或由下述核酸序列组成:所述核酸序列源自选自由下述组成的组的基因的转录本的3'-UTR和/或5'-UTR: ZNF460, TGM2, IL7R, BGN, TK1, RAB3B, CBX6, FZD2, COL8A1, NDUFS7, PHGDH, PLK2, TSP0, PTGS1, FBX032, NID2, ATP5D, EXOSC4, NOL9, UBB4B, VPS18, ORMDL2, FSCN1, TMEM33, TUBA4A, EMP3, TMEM201, CRIP2, BRAT1, SERPINH1, CD9, DPYSL2, CDK9, TFRC, PSMB3 5'-UTR, FASN, PSMB6, PRSS56, KPNA6, SFT2D2, PARD6B, LPP, SPARC, SCAND1, VASN, SLC26A1, LCLAT1, FBXL18, SLC35F6, RAB3D, MAP1B, VMA21, CYBA, SEZ6L2, PCOLCE, VTN, ALDH16A1, RAVER1, KPNA6, SERINC5, JUP, CPN2, CRIP2, EPT1, PNP0, SSSCA1, POLR2L, LIN7C, UQCR10, PYCRL, AMN, MAP1S, NDUFS7, PHGDH, TSP0, ATP5D, EXOSC4, TUBB4B, TUBA4A, EMP3, CRIP2, BRAT1, CD9, CDK9, PSMB3, PSMB6, PRSS56, SCAND1, AMN, CYBA, PCOLCE, MAP1S, VTN, ALDH16A1 (全部优选是人的) 和 Dpysl2, Ccnd1, Acox2, Cbx6, Ubc, Ldlr, Nudt22, Pcyox11, Ankrd1, Tmem37, Tspyl4, Slc7a3, Cst6, Aacs, Nosip, Itga7, Ccnd2, Ebp, Sf3b5, Fasn, Hmgcs1, Osrl, Lmnb1, Vma21, Kif20a, Cdca8, Slc7a1, Ubqln2, Prps2, Shmt2, Aurkb, Fignl1, Cad, Anln, Slfn9, Ncaph, Pole, Uhrf1, Gja1, Fam64a, Kif2c, Tspan10, Scand1, Gpr84, Fads3, Cers6, Cxcr4, Gprc5c, Fen1, Cspg4, Mrpl34, Comtd1, Armc6, Emr4, Atp5d, 1110001J03Rik, Csf2ra, Aarsd1, Kif22, Cth, Tpgs1,

Ccl17, Alkbh7, Ms4a8a, Acox2, Ubc, Slpi, Pcyox11, Igf2bp1, Tmem37, Slc7a3, Cst6, Ebp, Sf3b5, Plk1, Cdca8, Kif22, Cad, Cth, Pole, Kif2c, Scand1, Gpr84, Tpgs1, Ccl17, Alkbh7, Ms4a8a, Mrpl34, Comtd1, Armc6, Atp5d, 1110001J03Rik, Nudt22, Aarsd1 (全部优选是小鼠的)。

59. 根据权利要求57所述的方法或根据权利要求58所述的用途, 其中所述至少一个3' -UTR元件和/或所述至少一个5' -UTR元件包含源自选自由下述组成的组的人基因的转录本的3' -UTR和/或5' -UTR的核酸序列: ZNF460, TGM2, IL7R, BGN, TK1, RAB3B, CBX6, FZD2, COL8A1, NDUFS7, PHGDH, PLK2, TSP0, PTGS1, FBX032, NID2, ATP5D, EXOSC4, NOL9, UBB4B, VPS18, ORM DL2, FSCN1, TMEM33, TUBA4A, EMP3, TMEM201, CRIP2, BRAT1, SERPINH1, CD9, DPYSL2, CDK9, TFRC, PSMB3 5' -UTR, FASN, PSMB6, PRSS56, KPNA6, SFT2D2, PARD6B, LPP, SPARC, SCAND1, VASN, SLC26A1, LCLAT1, FBXL18, SLC35F6, RAB3D, MAP1B, VMA21, CYBA, SEZ6L2, PCOLCE, VTN, ALDH16A1, RAVR1, KPNA6, SERINC5, JUP, CPN2, CRIP2, EPT1, PNP0, SSSCA1, POLR2L, LIN7C, UQCR10, PYCRL, AMN, MAP1S, NDUFS7, PHGDH, TSP0, ATP5D, EXOSC4, TUBB4B, TUBA4A, EMP3, CRIP2, BRAT1, CD9, CDK9, PSMB3, PSMB6, PRSS56, SCAND1, AMN, CYBA, PCOLCE, MAP1S, VTN, ALDH16A1。

60. 根据权利要求57所述的方法或根据权利要求58所述的用途, 其中所述至少一个3' -UTR元件和/或所述至少一个5' -UTR元件包含源自选自由下述组成的组的鼠基因的转录本的3' -UTR和/或5' -UTR的核酸序列: Dpysl2, Ccnd1, Acox2, Cbx6, Ubc, Ldlr, Nudt22, Pcyox11, Ankrd1, Tmem37, Tspyl4, Slc7a3, Cst6, Aacs, Nosip, Itga7, Ccnd2, Ebp, Sf3b5, Fasn, Hmgcs1, Osr1, Lmnb1, Vma21, Kif20a, Cdca8, Slc7a1, Ubqln2, Prps2, Shmt2, Aurkb, Fignl1, Cad, Anln, Slfn9, Ncaph, Pole, Uhrf1, Gjal, Fam64a, Kif2c, Tspan10, Scand1, Gpr84, Fads3, Cers6, Cxcr4, Gprc5c, Fen1, Cspg4, Mrpl34, Comtd1, Armc6, Emr4, Atp5d, 1110001J03Rik, Csf2ra, Aarsd1, Kif22, Cth, Tpgs1, Ccl17, Alkbh7, Ms4a8a, Acox2, Ubc, Slpi, Pcyox11, Igf2bp1, Tmem37, Slc7a3, Cst6, Ebp, Sf3b5, Plk1, Cdca8, Kif22, Cad, Cth, Pole, Kif2c, Scand1, Gpr84, Tpgs1, Ccl17, Alkbh7, Ms4a8a, Mrpl34, Comtd1, Armc6, Atp5d, 1110001J03Rik, Nudt22, Aarsd1。

61. 根据权利要求57-59中任一项所述的方法或用途, 其中所述至少一个3' -UTR元件包含源自选自自由下述组成的组的基因的转录本的3' -UTR的核酸序列: NDUFS7-3' -UTR, PHGDH-3' -UTR, TSP0-3' -UTR, ATP5D-3' -UTR, EXOSC4-3' -UTR, TUBB4B-3' -UTR, TUBA4A-3' -UTR, EMP3-3' -UTR, CRIP2-3' -UTR, BRAT1-3' -UTR, CD9-3' -UTR, CDK9-3' -UTR, PSMB3-3' -UTR, PSMB6-3' -UTR, PRSS56-3' -UTR, SCAND1-3' -UTR, AMN-3' -UTR, CYBA-3' -UTR, PCOLCE-3' -UTR, MAP1S-3' -UTR, VTN-3' -UTR, ALDH16A1-3' -UTR (全部优选是人的) 和Acox2-3' -UTR, Ubc-3' -UTR, Slpi-3' -UTR, Pcyox11-3' -UTR, Igf2bp1-3' -UTR, Tmem37-3' -UTR, Slc7a3-3' -UTR, Cst6-3' -UTR, Ebp-3' -UTR, Sf3b5-3' -UTR, Plk1-3' -UTR, Cdca8-3' -UTR, Kif22-3' -UTR, Cad-3' -UTR, Cth-3' -UTR, Pole-3' -UTR, Kif2c-3' -UTR, Scand1-3' -UTR, Gpr84-3' -UTR, Tpgs1-3' -UTR, Ccl17-3' -UTR, Alkbh7-3' -UTR, Ms4a8a-3' -UTR, Mrpl34-3' -UTR, Comtd1-3' -UTR, Armc6-3' -UTR, Atp5d-3' -UTR, 1110001J03Rik-3' -UTR, Nudt22-3' -UTR, Aarsd1-3' -UTR, (全部优选是小鼠的); 优选地, 所述至少一个3' -UTR元件包含下述核酸序列或由下述核酸序列组成: 所述核酸序列源自选自自由下述组成的组的基因的转录本的3' -

UTR:NDUFS7-3'-UTR,PHGDH-3'-UTR,TSP0-3'-UTR,ATP5D-3'-UTR,EXOSC4-3'-UTR,TUBB4B-3'-UTR,TUBA4A-3'-UTR,EMP3-3'-UTR,CRIP2-3'-UTR,BRAT1-3'-UTR,PSMB3-3'-UTR,PSMB6-3'-UTR,SCAND1-3'-UTR,AMN-3'-UTR,CYBA-3'-UTR,PCOLCE-3'-UTR,MAP1S-3'-UTR,VTN-3'-UTR,ALDH16A1-3'-UTR(全部是人的),Acox2-3'-UTR,Ubc-3'-UTR,Slpi-3'-UTR,Igf2bp1-3'-UTR,Tmem37-3'-UTR,Slc7a3-3'-UTR,Cst6-3'-UTR,Ebp-3'-UTR,Sf3b5-3'-UTR,Cdca8-3'-UTR,Kif22-3'-UTR,Cad-3'-UTR,Pole-3'-UTR,Kif2c-3'-UTR,Scand1-3'-UTR,Gpr84-3'-UTR,Tpgs1-3'-UTR,Ccl17-3'-UTR,Alkbh7-3'-UTR,Ms4a8a-3'-UTR,Mrpl34-3'-UTR,Comtd1-3'-UTR,Armc6-3'-UTR,Atp5d-3'-UTR,1110001J03Rik-3'-UTR,Nudt22-3'-UTR(全部是小鼠的)。

62. 根据权利要求57-60中任一项所述的方法或用途,其中所述至少一个5'-UTR元件包含源自选自自由下述组成的组的基因的转录本的5'-UTR的核酸序列:ZNF460-5'-UTR,TGM2-5'-UTR,IL7R-5'-UTR,BGN-5'-UTR,TK1-5'-UTR,RAB3B-5'-UTR,CBX6-5'-UTR,FZD2-5'-UTR,COL8A1-5'-UTR,NDUFS7-5'-UTR,PHGDH-5'-UTR,PLK2-5'-UTR,TSP0-5'-UTR,PTGS1-5'-UTR,FBX032-5'-UTR,NID2-5'-UTR,ATP5D-5'-UTR,EXOSC4-5'-UTR,NOL9-5'-UTR,UBB4B-5'-UTR,VPS18-5'-UTR,ORMDL2-5'-UTR,FSCN1-5'-UTR,TMEM33-5'-UTR,TUBA4A-5'-UTR,EMP3-5'-UTR,TMEM201-5'-UTR,CRIP2-5'-UTR,BRAT1-5'-UTR,SERPINH1-5'-UTR,CD9-5'-UTR,DPYSL2-5'-UTR,CDK9-5'-UTR,TFRC-5'-UTR,PSMB3-5'-UTR,FASN-5'-UTR,PSMB6-5'-UTR,PRSS56-5'-UTR,KPNA6-5'-UTR,SFT2D2-5'-UTR,PARD6B-5'-UTR,LPP-5'-UTR,SPARC-5'-UTR,SCAND1-5'-UTR,VASN-5'-UTR,SLC26A1-5'-UTR,LCLAT1-5'-UTR,FBXL18-5'-UTR,SLC35F6-5'-UTR,RAB3D-5'-UTR,MAP1B-5'-UTR,VMA21-5'-UTR,CYBA-5'-UTR,SEZ6L2-5'-UTR,PCOLCE-5'-UTR,VTN-5'-UTR,ALDH16A1-5'-UTR,RAVER1-5'-UTR,KPNA6-5'-UTR,SERINC5-5'-UTR,JUP-5'-UTR,CPN2-5'-UTR,CRIP2-5'-UTR,EPT1-5'-UTR,PNPO-5'-UTR,SSSCA1-5'-UTR,POLR2L-5'-UTR,LIN7C-5'-UTR,UQCR10-5'-UTR,PYCRL-5'-UTR,AMN-5'-UTR,MAP1S-5'-UTR,(全部优选是人的)和Dpys12-5'-UTR,Ccnd1-5'-UTR,Acox2-5'-UTR,Cbx6-5'-UTR,Ubc-5'-UTR,Ldlr-5'-UTR,Nudt22-5'-UTR,Pcyox11-5'-UTR,Ankrd1-5'-UTR,Tmem37-5'-UTR,Tspyl4-5'-UTR,Slc7a3-5'-UTR,Cst6-5'-UTR,Aacs-5'-UTR,Nosip-5'-UTR,Itga7-5'-UTR,Ccnd2-5'-UTR,Ebp-5'-UTR,Sf3b5-5'-UTR,Fasn-5'-UTR,Hmgcs1-5'-UTR,Osrl-5'-UTR,Lmnb1-5'-UTR,Vma21-5'-UTR,Kif20a-5'-UTR,Cdca8-5'-UTR,Slc7a1-5'-UTR,Ubqln2-5'-UTR,Prps2-5'-UTR,Shmt2-5'-UTR,Aurkb-5'-UTR,Figl1-5'-UTR,Cad-5'-UTR,Anln-5'-UTR,Slfn9-5'-UTR,Ncaph-5'-UTR,Pole-5'-UTR,Uhrf1-5'-UTR,Gja1-5'-UTR,Fam64a-5'-UTR,Kif2c-5'-UTR,Tspan10-5'-UTR,Scand1-5'-UTR,Gpr84-5'-UTR,Fads3-5'-UTR,Cers6-5'-UTR,Cxcr4-5'-UTR,Gprc5c-5'-UTR,Fen1-5'-UTR,Cspg4-5'-UTR,Mrpl34-5'-UTR,Comtd1-5'-UTR,Armc6-5'-UTR,Emr4-5'-UTR,Atp5d-5'-UTR,1110001J03Rik-5'-UTR,Csf2ra-5'-UTR,Aarsd1-5'-UTR,Kif22-5'-UTR,Cth-5'-UTR,Tpgs1-5'-UTR,Ccl17-5'-UTR,Alkbh7-5'-UTR,Ms4a8a-5'-UTR(全部优选是小鼠的);优选地,所述至少一个5'-UTR元件包含下述核酸序列或由下述核酸序列组成:所述核酸序列源自下述各项的转录本的5'-UTR:ZNF460-5'-UTR,TGM2-5'-UTR,IL7R-5'-UTR,COL8A1-5'-UTR,NDUFS7-5'-UTR,PLK2-5'-UTR,FBX032-5'-UTR,ATP5D-5'-UTR,TUBB4B-5'-UTR,ORMDL2-5'-UTR,FSCN1-5'-UTR,CD9-5'-UTR,PYSL2-5'-UTR,PSMB3-5'-UTR,PSMB6-5'-

UTR, KPNA6-5' -UTR, SFT2D2-5' -UTR, LCLAT1-5' -UTR, FBXL18-5' -UTR, SLC35F6-5' -UTR, VMA21-5' -UTR, SEZ6L2-5' -UTR, PCOLCE-5' -UTR, VTN-5' -UTR, ALDH16A1-5' -UTR, KPNA6-5' -UTR, JUP-5' -UTR, CPN2-5' -UTR, PNPO-5' -UTR, SSSCA1-5' -UTR, POLR2L-5' -UTR, LIN7C-5' -UTR, UQCR10-5' -UTR, PYCRL-5' -UTR, AMN-5' -UTR, MAP1S-5' -UTR (全部是人的), Dpys12-5' -UTR, Acox2-5' -UTR, Ubc-5' -UTR, Nudt22-5' -UTR, Pcyox11-5' -UTR, Ankrd1-5' -UTR, Tspyl4-5' -UTR, Slc7a3-5' -UTR, Aacs-5' -UTR, Nosip-5' -UTR, Itga7-5' -UTR, Ccnd2-5' -UTR, Ebp-5' -UTR, Sf3b5-5' -UTR, Fasn-5' -UTR, Hmgcs1-5' -UTR, Osr1-5' -UTR, Lmnb1-5' -UTR, Vma21-5' -UTR, Kif20a-5' -UTR, Cdca8-5' -UTR, Slc7a1-5' -UTR, Ubqln2-5' -UTR, Prps2-5' -UTR, Shmt2-5' -UTR, Fignl1-5' -UTR, Cad-5' -UTR, Anln-5' -UTR, Slfn9-5' -UTR, Ncaph-5' -UTR, Pole-5' -UTR, Uhrf1-5' -UTR, Gja1-5' -UTR, Fam64a-5' -UTR, Tspan10-5' -UTR, Scand1-5' -UTR, Gpr84-5' -UTR, Cers6-5' -UTR, Cxcr4-5' -UTR, Gprc5c-5' -UTR, Fen1-5' -UTR, Cspg4-5' -UTR, Mrpl34-5' -UTR, Comtd1-5' -UTR, Armc6-5' -UTR, Emr4-5' -UTR, Atp5d-5' -UTR, Csf2ra-5' -UTR, Aarsd1-5' -UTR, Cth-5' -UTR, Tpgs1-5' -UTR, Cc117-5' -UTR, Alkbh7-5' -UTR, Ms4a8a-5' -UTR (全部是小鼠的)。

63. 根据权利要求57-61中任一项所述的方法或用途, 其中所述3' -UTR元件包含下述核酸序列或由下述核酸序列组成: 所述核酸序列与选自SEQ ID NOs: 152-204组成的组的序列具有至少约40%, 优选至少约50%, 优选至少约60%, 优选至少约70%, 更优选至少约80%, 更优选至少约90%, 甚至更优选至少约95%, 甚至更优选至少约99%的同一性, 或者, 其中所述3' -UTR元件包含下述核酸序列的片段或由下述核酸序列的片段组成: 所述核酸序列的片段与选自SEQ ID NOs: 152-204组成的组的序列具有至少约40%, 优选至少约50%, 优选至少约60%, 优选至少约70%, 更优选至少约80%, 更优选至少约90%, 甚至更优选至少约95%, 甚至更优选至少约99%的同一性。

64. 根据权利要求57-60和62中任一项所述的方法或用途, 其中所述5' -UTR元件包含下述核酸序列或由下述核酸序列组成: 所述核酸序列与选自SEQ ID NOs: 1-151组成的组的序列具有至少约40%, 优选至少约50%, 优选至少约60%, 优选至少约70%, 更优选至少约80%, 更优选至少约90%, 甚至更优选至少约95%, 甚至更优选至少约99%的同一性, 或者, 其中所述5' -UTR元件包含下述核酸序列的片段或由下述核酸序列的片段组成: 所述核酸序列的片段与选自SEQ ID NOs: 1-151组成的组的序列具有至少约40%, 优选至少约50%, 优选至少约60%, 优选至少约70%, 更优选至少约80%, 更优选至少约90%, 甚至更优选至少约95%, 甚至更优选至少约99%的同一性。

65. 根据权利要求63或64所述的方法或用途, 其中所述片段呈现3-约500个核苷酸、优选5-约150个核苷酸、更优选10-100个核苷酸、甚至更优选15-90个、最优选20-70个核苷酸的长度。

66. 根据权利要求57-65中任一项所述的方法或用途, 其中所述3' -UTR元件和/或所述5' -UTR元件呈现3-约500个核苷酸、优选5-约150个核苷酸、更优选10-100个核苷酸、甚至更优选15-90个、最优选20-70个核苷酸的长度。

67. 试剂盒或部件试剂盒, 其包括根据权利要求1-38中任一项所述的人工核酸分子、根据权利要求39-43中任一项所述的载体、根据权利要求44-46中任一项所述的细胞和/或根据权利要求47或48所述的药物组合物。

68. 根据权利要求67所述的试剂盒,其还包括使用说明、用于转染的细胞、佐剂、施用所述药物组合物的装置、药用载体和/或用于溶解或稀释所述人工核酸分子、所述载体、所述细胞或所述药物组合物的药用溶液。

69. 用于产生人工核酸分子的方法,其中合成权利要求1-38中任一项限定的包含至少一个可读框和至少一个3' -UTR元件和/或至少一个5' -UTR元件的人工核酸分子。

70. 根据权利要求69所述的用于产生人工核酸分子的方法,其中使用根据权利要求39-43任一项所述的载体来合成所述人工核酸分子。

71. 能够通过根据权利要求69-70中任一项所述的用于产生人工核酸分子的方法得到的人工核酸分子。



## 人工核酸分子

[0001] 本申请是申请日为2016年8月22日、发明名称为“人工核酸分子”的中国专利申请No.201680049973.1的分案申请。

[0002] 本发明涉及人工核酸分子,其包含可读框、3'-非翻译区元件(3'-UTR元件)和/或5'-非翻译区元件(5'-UTR元件)和任选地聚腺苷酸序列(poly(A)序列)和/或聚腺苷酸化-信号。本发明还涉及包含3'-UTR元件和/或5'-UTR元件的载体,涉及包含所述人工核酸分子或所述载体的细胞,涉及包含所述人工核酸分子或所述载体的药物组合物,以及涉及包含所述人工核酸分子、所述载体和/或所述药物组合物的试剂盒,优选用于基因治疗和/或基因接种疫苗的领域。

[0003] 基因治疗和基因接种疫苗属于最有希望和快速发展的现代医学方法。它们可以为多种疾病的治疗提供高度特异和个性化选择。尤其是,不仅遗传病,而且自身免疫疾病、癌或肿瘤相关疾病以及炎症疾病可以是此种治疗方法的对象。而且,人们设想通过这些方法预防此类疾病的早期发病。

[0004] 基因治疗背后的主要构思理念为适当调节与特定疾病的病理状态相关的受损的基因表达。病理学上改变的基因表达可以导致缺少或过量生产关键基因产物,例如,信号传导因子诸如激素,看家因子(housekeeping factors),代谢酶,结构蛋白等。改变的基因表达可以不仅由于转录和/或翻译的错误调节引起,而且还由于编码特定蛋白的ORF内的突变引起。病理突变可以由例如染色体畸变,或由更特异突变,如点或移码突变引起,它们全部都导致受限的功能,以及可能完全丧失基因产物的功能。然而,如果突变影响编码参与细胞的转录或翻译机制的蛋白的基因,转录或翻译的错误调节也可能发生。此种突变可以导致基因病理上调或下调,所述基因本身是功能性的。编码发挥此种调节功能的基因产物的基因,可以是例如,转录因子、信号受体、信使蛋白等。然而,此种编码调节蛋白的基因的功能缺失,在某些情况下可以由人工引入进一步作用于受损的基因产物下游的其它因子反转。此种基因缺陷还可以通过置换受影响基因本身的基因治疗补偿。

[0005] 基因接种疫苗允许引起对选择的抗原所需的免疫应答,所述选择的抗原诸如细菌表面的特征组分、病毒粒子、肿瘤抗原等。一般而言,疫苗接种是现代医学的关键性成就之一。然而,目前仅对有限的疾病有可用的有效疫苗。因此,不能由疫苗接种预防的感染每年仍然影响数百万人。

[0006] 通常,疫苗可以细分为“第一代”、“第二代”和“第三代”疫苗。“第一代”疫苗典型地是全生物体疫苗。它们基于活的和减毒或被杀灭的病原体,例如病毒、细菌等。活疫苗和减毒疫苗的主要缺点是向危及生命的变体回复的风险。因此,尽管被减毒,但此种病原体本质上仍然可以带有不可预测的风险。杀灭的病原体可能不如所需要的那样有效产生特异免疫应答。为了最小化这些风险,开发了“第二代”疫苗。这些典型地为亚单位疫苗,所述亚单位疫苗由源自病原体的限定的抗原或重组蛋白组分组成。

[0007] 基因疫苗,即用于基因接种的疫苗,通常被理解为“第三代”疫苗。它们典型地由遗传改造的核酸分子构成,所述遗传改造的核酸分子允许在体内表达病原体或肿瘤抗原特征性的肽或蛋白(抗原)片段。在施用于患者,由靶细胞摄取后,基因疫苗表达。施用的核酸的

表达导致产生编码的蛋白。如果这些蛋白被患者的免疫系统识别为外源物,则引发免疫应答。

[0008] 如从上文可以看出的,基因治疗和基因接种疫苗这两种方法基本上都基于向患者施用核酸分子并随后转录和/或翻译编码的遗传信息。备选地,基因接种疫苗或基因治疗还可以包括以下方法:所述方法包括从要治疗的患者分离特定的体细胞,随后离体转染此种细胞,并将处理后的细胞再施用于患者。

[0009] DNA以及RNA可以用作在基因治疗或基因接种疫苗范畴内施用的核酸分子。已知DNA相对稳定并易于操作。然而,使用DNA承担施用的DNA-片段不希望的插入患者基因组中、从而潜在地导致突变事件如受损基因功能的丧失的风险。作为进一步的风险,出现不希望的抗DNA抗体。另一个缺点是在DNA施用后可以获得的编码肽或蛋白的有限表达水平,因为DNA必需进入核,才能被转录,之后得到的mRNA可以被翻译。在其它原因中,施用的DNA的表达水平将依赖于调节DNA转录的特定转录因子的存在。在缺少此种因子的情况下,DNA转录将不产生令人满意量的RNA。从而,获得的翻译的肽或蛋白的水平有限。

[0010] 通过使用RNA替代DNA用于基因治疗或基因接种疫苗,不希望的基因组整合和抗DNA抗体产生的风险被最小化或被避免。然而,RNA被认为是相当不稳定的分子种类,其可以容易地被遍在的RNA酶降解。

[0011] 通常,RNA-降解促进调节RNA半衰期。该效果被认为和证明是精细调整真核基因表达的调节(Friedel等人,2009.Conserved principles of mammalian transcriptional regulation revealed by RNA half-life,Nucleic Acid Research 37(17):1-12)。因此,每个天然存在的mRNA具有依赖于mRNA源头的基因以及其在何种细胞类型中表达的其各自的半衰期。其促进调节该基因的表达水平。不稳定RNA对于及时在不同点实现瞬时基因表达是重要的。然而,耐久的RNA可能与不同蛋白的积累或基因的连续表达相关。在体内,mRNA的半衰期还可以依赖于环境因素,如激素处理,如例如对胰岛素样生长因子I、肌动蛋白和白蛋白mRNA所显示的(Johnson等人,Newly synthesized RNA:Simultaneous measurement in intact cells of transcription rates and RNA stability of insulin-like growth factor I,actin,and albumin in growth hormone-stimulated hepatocytes,Proc.Natl.Acad.Sci.,第88卷,第5287-5291页,1991)。

[0012] 对于基因治疗和基因接种疫苗,通常需要稳定的RNA。一方面,这是由于通常需要由RNA序列编码的产物在体内积累的事实。另一方面,当在其储存过程中制备成合适剂型时,和当施用时,RNA必须保持其结构和功能完整性。因此,对提供用于基因治疗或基因接种疫苗的稳定RNA分子进行尝试,以阻止它们经历早期降解或衰变。

[0013] 据报道,核酸分子的G/C-含量可以影响其稳定性。因此,包含增加量的鸟嘌呤(G)和/或胞嘧啶(C)残基的核酸可以比含有大量腺嘌呤(A)和胸腺嘧啶(T)或尿嘧啶(U)核苷酸的核酸在功能上更稳定。在这方面,W002/098443提供含有由编码区中的序列修饰稳定化的mRNA的药物组合物。此序列修饰利用遗传密码简并性。因此,含有不太有利的核苷酸组合(就RNA稳定性而言不太有利)的密码子可以在不改变编码的氨基酸序列的情况下被替代的密码子置换。该RNA稳定化方法受提供不允许留下所需氨基酸序列间隔的各个单个RNA分子的特定核苷酸序列限制。此外,该方法限于RNA的编码区。

[0014] 作为mRNA稳定化的备选方案,已经发现天然存在的真核mRNA分子含有特有的稳定

化元件。例如,它们可以包含在其5'末端(5'-UTR)和/或在其3'末端(3'-UTR)的所谓非翻译区(UTR)以及其它结构特征,如5'-帽结构或3'-聚腺苷酸(poly(A))尾。5'-UTR和3'-UTR二者都典型地从基因组DNA转录并且因此是成熟前(premature)mRNA元件。在mRNA加工过程中,成熟mRNA的特有结构特征,如5'-帽和3'-聚腺苷酸尾(也称为多聚腺苷酸尾(poly(A) tail)或聚腺苷酸序列(poly(A) sequence))通常被添加于转录的(成熟前)mRNA。

[0015] 3'-聚腺苷酸尾典型地是添加在转录的mRNA的3'末端的一段单调的腺苷核苷酸序列。其可以包含至多约400个腺苷核苷酸。发现此种3'-聚腺苷酸尾的长度对于个体mRNA的稳定性是可能的关键要素。

[0016] 此外,已显示 $\alpha$ -珠蛋白mRNA的3'-UTR可能是公知的 $\alpha$ -珠蛋白mRNA稳定性的重要因素(Rodgers等人,Regulated $\alpha$ -globin mRNA decay is a cytoplasmic event proceeding through 3'-to-5' exosome-dependent decapping, RNA, 8, 第1526-1537页, 2002)。 $\alpha$ -珠蛋白mRNA的3'-UTR明显参与特定核蛋白-复合物( $\alpha$ -复合物)的形成,其存在与mRNA体外稳定性相关(Wang等人, An mRNA stability complex functions with poly(A)-binding protein to stabilize mRNA in vitro, Molecular and Cellular biology, 第19卷, No. 7, 1999年7月, 第4552-4560页)。

[0017] 对于核糖体蛋白mRNA中的UTR已经进一步表明有趣的调节功能:在核糖体蛋白mRNA的5'-UTR控制生长相关的mRNA翻译的同时,该调节的严格性由核糖体蛋白mRNA中的各个3'-UTR赋予(Ledda等人, Effect of the 3'-UTR length on the translational regulation of 5'-terminal oligopyrimidine mRNAs, Gene, 第344卷, 2005, 第213-220页)。该机制促进核糖体蛋白的特异表达,所述核糖体蛋白通常以恒定的方式转录,从而一些核糖体蛋白mRNA如核糖体蛋白S9或核糖体蛋白L32被称为看家基因(Janovick-Guretzky等人, Housekeeping Gene Expression in Bovine Liver is Affected by Physiological State, Feed Intake, and Dietary Treatment, J. Dairy Sci., 第90卷, 2007, 第2246-2252页)。核糖体蛋白的生长相关的表达模式因此主要是由于对翻译水平的调节。

[0018] WO 2014/164253 A1描述了一些具有5'-UTR和/或3'-UTR的具体的核酸分子,但是没有关于此类分子的翻译效率的详细内容。

[0019] 不考虑影响mRNA稳定性的因素,由靶细胞或组织有效翻译施用的核酸分子对于任何使用用于基因治疗或基因接种疫苗的核酸分子的方法都至关重要。从上文引用的实例可以看出,与调节稳定性一起,大多数mRNA的翻译也由结构特征如UTR、5'-帽和3'-聚腺苷酸尾调节。在此种情况下,据报道,聚腺苷酸尾的长度可能也对翻译效率发挥重要作用。然而,稳定化3'-元件也可能对翻译具有减弱效果。

[0020] 本发明的目的是提供可以适于应用于基因治疗和/或基因接种疫苗的核酸分子。尤其是,本发明的目的是提供这样的mRNA种类:其针对早期降解或衰变得稳定,但不显示翻译效率的显著功能损失。本发明还有一个目的是提供以高翻译效率为特征的人工核酸分子,优选mRNA。本发明的一个特别目的是提供mRNA,其中所述翻译(各个编码的蛋白的核糖体产生)效率增强,例如,相对于参比核酸分子(参比mRNA)增强。本发明的另一个目的是提供编码此种优异mRNA种类的核酸分子,其可以适于在基因治疗和/或基因接种疫苗中使用。本发明进一步的目的是提供用于基因治疗和/或基因接种疫苗的药物组合物。归纳起来,本

发明的目的是提供改进的核酸种类,其通过节省成本和直接的途径克服上文讨论的现有技术的缺陷。

[0021] 作为本发明的基础的目的通过所要求的主题得以解决。具体地,本发明人鉴定了提供高翻译效率的UTR元件(5'-UTR元件和3'-UTR元件)。它们均以翻译效率高于先前已知的核酸分子的翻译效率而特征,这是本发明优选的核酸分子的共同方面。还提供了用于确定相对于参比核酸分子的高翻译效率的测定。

[0022] 本发明的进行得到美国政府的支持,DARPA授予的合同号为HR0011-11-3-0001。美国政府拥有本发明的某些权益。

[0023] 为了清楚和易读,提供以下定义。对于这些定义所提到的任何技术特征,可以基于本发明的每个实施方案读懂。可以在这些实施方案范围内具体提供其它定义和解释。

[0024] 适应性免疫应答:适应性免疫应答典型地理解为免疫系统的抗原-特异应答。抗原特异性允许产生适应特异病原体或病原体感染的细胞的应答。发起这些适应的应答的能力通常由“记忆细胞”保持在体内。假如病原体不止一次感染身体,这些特异记忆细胞用于快速清除它。在此种情况下,适应性免疫应答的第一步是激活能够通过抗原呈递细胞诱导抗原特异免疫应答的幼稚(naïve)抗原特异T细胞或不同免疫细胞。这在幼稚T细胞不断通过的淋巴组织和器官中发生。可以用作抗原呈递细胞的三种细胞类型是树突细胞、巨噬细胞和B细胞。这些细胞中的每个在引起免疫应答方面具有不同功能。树突细胞可以通过吞噬作用和大胞饮摄取抗原,并且可以通过与例如外来抗原相接触而被刺激,从而迁移至局部淋巴组织,在那里它们分化为成熟树突细胞。巨噬细胞摄取颗粒抗原如细菌并通过由传染剂或其它合适的刺激物诱导而表达MHC分子。B细胞通过其受体结合和内化可溶蛋白抗原的独特能力对于诱导T细胞也可以是重要的。MHC-分子典型地负责向T-细胞呈递抗原。其中,将抗原呈递到MHC分子上导致T细胞的激活,这诱导其增殖和向武装的效应T细胞的分化。效应T细胞的最重要功能是通过CD8+细胞毒性T细胞杀灭感染的细胞和通过Th1细胞激活巨噬细胞(二者一起构成细胞介导的免疫),和通过Th2和Th1细胞激活B细胞以产生不同种类的抗体,从而驱动体液免疫应答。T细胞通过其T细胞受体识别抗原,其不直接识别和结合抗原,而是识别短肽片段例,如源自病原体的蛋白抗原的片段,例如所谓的表位,其结合其它细胞表面上的MHC分子。

[0025] 适应性免疫系统:适应性免疫系统基本上致力于消除或阻止病原体生长。其典型地通过向脊椎动物免疫系统提供识别和记忆特定病原体(以产生免疫性)并在每次遭遇病原体时产生更强烈的攻击的能力调节适应性免疫应答。因为体细胞超变(加速的体细胞突变过程)和V(D)J重组(抗原受体基因片段的不可逆遗传重组),该系统是高度可适应的。该机制允许小量基因产生大量不同抗原受体,其随后独特地表达在每个个体淋巴细胞上。因为基因重排导致各个细胞DNA的不可逆改变,所以之后此种细胞的所有子代(后代)将遗传编码相同受体特异性的基因,包括作为持久的特异免疫性的关键的记忆B细胞和记忆T细胞。

[0026] 佐剂/佐剂组分:佐剂或佐剂组分广义上典型地是可以改良(例如增强)其它试剂(如药物或疫苗)的效果的药剂和/或免疫剂。将在广义上解释它并且指广谱的物质。典型地,这些物质能够增加抗原的免疫原性。例如,佐剂可以被先天免疫系统识别,并且,例如可以引起先天免疫应答。“佐剂”典型地不引起适应性免疫应答。在这个程度上,“佐剂”不算是

抗原。其作用模式不同于由导致适应性免疫应答的抗原引发的效果。

[0027] 抗原:在本发明的背景下,“抗原”典型地指可以被免疫系统,优选被适应性免疫系统识别,并能够例如通过作为适应性免疫应答的一部分的抗体和/或抗原-特异T细胞的形成引发抗原-特异免疫应答的物质。典型地,抗原可以是或可以包含可以被MHC呈递到T-细胞的肽或蛋白。在本发明的意义中,抗原可以是提供的核酸分子(优选本文中限定的mRNA)的翻译产物。在此情况下,肽和蛋白的包含至少一个表位的片段、变体和衍生物也理解为抗原。在本发明的情形中,本文中限定的肿瘤抗原和病原体抗原是特别优选的。

[0028] 换言之,人工核酸分子可以理解为非天然核酸分子。此种核酸分子可以由于其个体序列(其天然不存在)和/或由于其它修饰(例如天然不存在的核苷酸的结构修饰)而是非天然的。人工核酸分子可以是DNA分子、RNA分子或包含DNA和RNA部分的杂合分子。典型地,可以通过遗传工程方法设计和/或产生人工核酸分子以符合需要的人工核苷酸序列(异源序列)。在此种情况下,人工序列通常是可以天然不存在的序列,即其与野生型序列至少有一个核苷酸不同。术语“野生型”可以理解为天然存在的序列。当在本文中描述任意特定的“人工核酸分子”是“基于”任意特定的野生型核酸分子时,则所述人工核酸分子与所述野生型人工核酸分子至少有一个核苷酸不同。进一步地,术语“人工核酸分子”不限于意为“一个单个分子”,而典型地理解为包含全体相同分子。因此,它可以涉及等分部分中含有的多个相同分子。

[0029] 双顺反子RNA,多顺反子RNA:双顺反子或多顺反子RNA典型地是典型地可以具有两个(双顺反子)或更多个(多顺反子)可读框(ORF)的RNA,优选mRNA。在此种情况下,可读框是可翻译为肽或蛋白的密码子序列。

[0030] 载体/聚合载体:本发明情形内的载体典型地可以是促进另一个化合物(货物)的转运和/或复合的化合物。聚合载体典型地是由聚合物形成的载体。载体可以通过共价或非共价相互作用与货物关联。载体可以将核酸(例如RNA或DNA)转运至靶细胞。对于一些实施方案,载体可以是阳离子组分。

[0031] 阳离子组分:术语“阳离子组分”典型地指在pH值典型地为从1至9,优选为或低于9(例如从5至9),为或低于8(例如从5至8),为或低于7(例如从5至7)的pH值,最优选在生理pH(例如从7.3至7.4)带正电荷(阳离子)的带电分子。因此,阳离子组分可以是任何带正电荷的化合物或聚合物,优选在生理条件下,尤其是在体内生理条件下带正电荷的阳离子肽或蛋白。“阳离子肽或蛋白”可以含有至少一个带正电荷的氨基酸,或多于一个带正电荷的氨基酸,例如选自Arg,His,Lys或Orn。因此,“聚阳离子”组分也在给定条件下呈现多于一个正电荷的范围内。

[0032] 5'-帽:5'-帽是实体,典型地是修饰的核苷酸实体,其通常“加帽(cap)”成熟mRNA的5'-末端。5'-帽可以典型地由修饰的核苷酸,尤其是由鸟嘌呤核苷酸衍生物形成。优选地,5'-帽通过5'-5'-三磷酸酯键与5'-末端连接。5'-帽可以是甲基化的,例如m7GpppN,其中N是携带5'-帽的核酸的末端5'核苷酸,典型地是RNA的5'-末端。5'帽结构的进一步实例包括甘油基(反转的脱氧脱碱基残基(部分))、4',5'亚甲基核苷酸,1-(β-D-赤味喃糖基)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,修饰碱基的核苷酸,苏式-戊味喃糖核苷酸,无环3',4'-开环核苷酸,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'-反向核苷酸部分,3'-3'-反向脱碱基部分,3'-2'-反向核

核苷酸部分,3'-2'-反向脱碱基部分,1,4-丁二醇磷酸酯,3'-氨基磷酸酯,己基磷酸酯,氨基己基磷酸酯,3'-磷酸酯,3'硫代磷酸,二硫代磷酸酯,或桥接或非桥接甲基磷酸酯部分。

[0033] 细胞免疫性/细胞免疫应答:细胞免疫性典型地涉及巨噬细胞、天然杀伤细胞(NK)、抗原特异性细胞毒性T-淋巴细胞的激活,和应答抗原时各种细胞因子的释放。更概括地,细胞免疫性不基于抗体,而基于免疫系统细胞的激活。典型地,细胞免疫应答可以例如特征在于激活能够在细胞(例如特定免疫细胞如树突细胞或其它细胞)中诱导凋亡的抗原特异性细胞毒性T-淋巴细胞,将外来抗原的表位展示在其表面。此种细胞可以是病毒感染的或细胞内细菌感染的,或展示肿瘤抗原的癌症细胞。进一步的特征可以是巨噬细胞和天然杀伤细胞的激活,使它们能够破坏病原体并刺激细胞分泌多种影响参与适应性免疫应答和先天免疫应答的其它细胞的功能的细胞因子。

[0034] DNA:DNA是脱氧核糖核酸的常见缩写。其是核酸分子,即由核苷酸组成的聚合物。这些核苷酸通常是脱氧腺苷单磷酸酯,脱氧胸苷单磷酸酯,脱氧鸟苷单磷酸酯和脱氧胞苷单磷酸酯单体,其本身由糖部分(脱氧核糖)、碱基部分和磷酸酯部分构成,并通过特征骨架结构聚合。所述骨架结构典型地通过第一个相邻单体的核苷酸的糖部分,即脱氧核糖和第二个相邻单体的磷酸酯部分之间的磷酸二酯键形成。单体的特定顺序,即与糖/磷酸酯-骨架连接的碱基的顺序,被称为DNA序列。DNA可以是单链或双链的。在双链形式中,第一链的核苷酸典型地与第二链的核苷酸例如通过A/T-碱基配对和G/C碱基配对杂交。

[0035] 表位:表位(也称为“抗原决定簇”)可以在T细胞表位和B细胞表位相区别。本发明范围内的蛋白的T细胞表位或部分可以包含片段,所述片段优选具有约6至约20或甚至更多个氨基酸的长度,例如由I类MHC分子加工和呈递的片段,优选具有约8至约10个氨基酸,例如8,9,或10个,(或甚至11或12个氨基酸)的长度,或由II类MHC分子加工和呈递的片段,优选具有约13或更多个氨基酸,例如13,14,15,16,17,18,19,20或甚至更多个氨基酸的长度,其中这些片段可以选自氨基酸序列的任何部分。这些片段典型地以由肽片段和MHC分子组成的复合物形式被T细胞识别,即所述片段典型地不以其天然形式被识别。B细胞表位典型地是位于如本文限定的(天然)蛋白或肽抗原的外表面的片段,优选具有5至15个氨基酸,更优选具有5至12个氨基酸,甚至更优选具有6至9个氨基酸,其可以即以其天然形式被抗体识别。

[0036] 此外,蛋白或肽的此种表位可以选自任何本文提到的此种蛋白或肽的变体。在此种情况下,抗原决定簇可以由如本文限定的蛋白或肽氨基酸序列不连续的、但在三维结构中汇集在一起的蛋白或肽片段构成的构象表位或不连续表位或由单一多肽链构成的连续或线性表位。

[0037] 序列片段:序列片段可以典型地是例如核酸分子或氨基酸序列的全长序列的较短部分。因此,典型地,片段由与全长序列内相应一段相同的序列组成。本发明范围内优选的序列片段由与所述片段源自的分子中连续段实体对应的连续段的实体(如核苷酸或氨基酸)组成,其代表所述片段源自的总(即全长)分子的至少5%,10%,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%。

[0038] G/C修饰的:G/C-修饰的核酸可以典型地是核酸,优选如本文限定的人工核酸分子,其基于优选包含与野生型序列相比增加的鸟苷和/或胞嘧啶核苷酸数量的修饰的野生

型序列。此种增加的数量可以通过用含有鸟苷或胞嘧啶核苷酸的密码子置换含有腺苷或胸苷核苷酸的密码子产生。如果富集的G/C含量发生在DNA或RNA的编码区,则其使用遗传密码的简并性。因此,所述密码子置换优选不改变编码的氨基酸残基,但仅增加核酸分子的G/C含量。

[0039] 基因治疗:基因治疗可以典型地理解为通过编码肽或蛋白的核酸治疗患者的身体或患者身体的分离的组成部分,例如分离的组织/细胞。其典型地可以包括至少一个以下步骤:a)直接将核酸(优选如本文限定的人工核酸分子)通过任何给药途径施用于患者或体外施用于患者的分离的细胞/组织,其导致体内/离体或体外转染患者的细胞;b)转录和/或翻译引入的核酸分子;和任选地c)如果核酸未直接施用于患者,则将分离、转染的细胞再施用于患者。

[0040] 基因疫苗接种:基因疫苗接种可以典型地理解为通过施用编码抗原或免疫原或其片段的核酸分子进行的疫苗接种。核酸分子可以施用于受试者的身体或施用于受试者的分离的细胞。当转染身体的某些细胞或当转染分离的细胞后,抗原或免疫原可以由那些细胞表达并随后呈递到免疫系统,引起适应性的(即抗原特异性的)免疫应答。因此,基因疫苗接种典型地包括至少一个以下步骤:a)将核酸(优选如本文限定的人工核酸分子)施用给受试者,优选患者,或施用于受试者(优选患者)的分离的细胞,其通常导致体内或体外转染受试者的细胞;b)转录和/或翻译引入的核酸分子;并且任选地c)如果核酸未直接施用于患者,则将分离、转染的细胞再施用于受试者,优选患者。

[0041] 异源序列:如果两条序列不源于相同的基因,则它们典型地被理解为‘异源的’。即,尽管异源序列可以源自相同生物体,但它们天然(在自然中)不存在于相同的核酸分子中,如不存在于相同的mRNA中。

[0042] 体液免疫/体液免疫应答:体液免疫典型地指抗体产生和任选地指伴随抗体产生的附加过程。体液免疫应答可以典型地特征在于,例如Th2激活和细胞因子产生、生发中心形成和同种型转换、亲和力成熟和记忆细胞产生。体液免疫性典型地还可以指抗体的效应功能,其包括病原体 and 毒素中和、经典补体激活、和吞噬作用的调理素促进和病原体清除。

[0043] 免疫原:在本发明范围内,免疫原可以典型地理解为能够刺激免疫应答的化合物。优选地,免疫原是肽、多肽或蛋白。在特别优选的实施方案中,从本发明的意义上说,免疫原是提供的核酸分子(优选如本文限定的人工核酸分子)的翻译产物。典型地,免疫原至少引起适应性免疫应答。

[0044] 免疫刺激组合物:本发明范围内,免疫刺激组合物可以典型地理解为含有至少一个能够诱导免疫应答的组分或能够诱导免疫应答的组分源自的组分的组合物。此种免疫应答可以优选为先天免疫应答或适应性和先天免疫应答的组合。优选地,本发明范围内的免疫刺激组合物含有至少一种人工核酸分子,更优选RNA,例如mRNA分子。免疫刺激组分,如mRNA,可以与合适的载体复合。因此,所述免疫刺激组合物可以包含mRNA/载体-复合物。此外,所述免疫刺激组合物可以包含用于免疫刺激组分(如mRNA)的佐剂和/或合适的载体。

[0045] 免疫应答:免疫应答可以典型地是针对特定抗原的适应性免疫系统的特异反应(所谓的特异或适应性免疫应答)或先天免疫系统的非特异反应(所谓的非特异或先天免疫应答),或其组合。

[0046] 免疫系统:免疫系统可以保护生物体免受感染。如果病原体成功通过生物体的物

理屏障并进入该生物体,则先天免疫系统提供立即、但非特异的应答。如果病原体避开了该先天应答,则脊椎动物拥有第二层保护,即适应性免疫系统。这里,免疫系统在感染过程中改变其应答以改善其对病原体的识别。病原体被消除之后,此种改善的应答随后以免疫记忆的形式保留,并且允许适应性免疫系统每次遭遇该病原体时发起更快速和更强烈的攻击。据此,免疫系统包含先天和适应性免疫系统。这两部分的每个典型地含有所谓体液和细胞组分。

[0047] 免疫刺激RNA:本发明范围内,免疫刺激RNA (isRNA) 可以典型地是能够诱导先天免疫应答的RNA。其通常不具有可读框并且因此不提供肽-抗原或免疫原,但例如通过结合特定类型的Toll样受体 (TLR) 或其它合适的受体引起免疫应答。然而,当然具有可读框并编码肽/蛋白的mRNA也可以诱导先天免疫应答,并且,因此可以是免疫刺激RNA。

[0048] 先天免疫系统:先天免疫系统,也被称为非特异 (non-specific) (或非特异 (unspecific)) 免疫系统,典型地包含以非特异的方式保护宿主免受其它生物体感染的细胞和机制。这意味着先天系统的细胞可以以一般的方式识别和响应于病原体,但不像适应性免疫系统,它不赋予宿主持久或保护性免疫。先天免疫系统可以例如被以下激活:Toll样受体 (TLR) 的配体或其它辅助物质如脂多糖, TNF- $\alpha$ , CD40配体,或细胞因子,单核因子,淋巴因子,白细胞介素或趋化因子, IL-1, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, IL-13, IL-14, IL-15, IL-16, IL-17, IL-18, IL-19, IL-20, IL-21, IL-22, IL-23, IL-24, IL-25, IL-26, IL-27, IL-28, IL-29, IL-30, IL-31, IL-32, IL-33, IFN- $\alpha$ , IFN- $\beta$ , IFN- $\gamma$ , GM-CSF, G-CSF, M-CSF, LT- $\beta$ , TNF- $\alpha$ , 生长因子, 和hGH, 人Toll样受体TLR1、TLR2、TLR3、TLR4、TLR5、TLR6、TLR7、TLR8、TLR9、TLR10的配体, 鼠Toll样受体TLR1、TLR2、TLR3、TLR4、TLR5、TLR6、TLR7、TLR8、TLR9、TLR10、TLR11、TLR12或TLR13的配体, NOD样受体的配体, RIG-I样受体的配体, 免疫刺激核酸, 免疫刺激RNA (isRNA), CpG-DNA, 抗菌剂, 或抗病毒剂。本发明所述的药物组合物可以包含一种以上此种物质。典型地, 先天免疫系统的应答包括:通过产生化学因子招募免疫细胞到感染位点, 所述化学因子包括专用的化学介质, 称为细胞因子; 激活补体级联; 通过专门的白细胞鉴别和去除器官、组织、血液和淋巴中的外来物质; 激活适应性免疫系统; 和/或用作针对传染剂的物理和化学屏障。

[0049] 克隆位点:克隆位点典型地理解为适于插入核酸序列 (例如包含可读框的核酸序列) 的核酸分子片段。插入可以通过任何本领域技术人员已知的分子生物学方法进行, 例如通过限制酶切和连接。克隆位点典型地包含一个以上限制性酶识别位点 (限制性位点)。这些一个以上限制性位点可以被在这些位点切割DNA的限制性酶识别。包含一个以上限制性位点的克隆位点也可以称为多克隆位点 (MCS) 或多聚接头。

[0050] 核酸分子:核酸分子是包含核酸组分的分子, 优选是由核酸组分组成的分子。术语核酸分子优选指DNA或RNA分子。其优选与术语“多核苷酸”同义使用。优选地, 核酸分子是聚合物, 包含通过糖/磷酸酯-骨架的磷酸二酯键彼此共价连接的核苷酸单体或由通过糖/磷酸酯-骨架的磷酸二酯键彼此共价连接的核苷酸单体组成。术语“核酸分子”还包括修饰的核酸分子, 如碱基修饰的、糖修饰的或骨架修饰的等的DNA或RNA分子。

[0051] 可读框:本发明范围内的可读框 (ORF) 可以典型地是一些核苷酸三联体的序列, 其可以翻译为肽或蛋白。可读框优选在其5'末端含有起始密码子, 即通常编码氨基酸甲硫氨酸的三个连续的核苷酸的组合 (ATG), 和通常呈现多个3核苷酸长度的紧接的区域。典型地,



这是可读框仅有的终止密码子。因此,本发明范围内的可读框优选为核苷酸序列,由可以被三整除的很多核苷酸组成,其以起始密码子(例如ATG)起始并且其优选以终止密码子(例如,TAA,TGA,或TAG)终止。可读框可以是分离的或其可以被整合入更长的核酸序列,例如载体或mRNA中。可读框还可以被称为“蛋白编码区”。

[0052] 肽:肽或多肽典型地是通过肽键连接的氨基酸单体的聚合物。其典型地含有少于50个单体单元。尽管如此,术语肽不否认为具有多于50个单体单元分子。长肽也称为多肽,典型地具有50到600个单体单元。

[0053] 药学有效量:本发明范围内的药学有效量典型地理解为例如在病理状况下足以诱导药学效应(如免疫应答)、改变表达的肽或蛋白的病理水平、或替代缺少的基因产物的量。

[0054] 蛋白:蛋白典型地包含一个以上肽或多肽。蛋白典型地折叠为3-维形式,其可以是蛋白发挥其生物学功能所需的。

[0055] 聚腺苷酸(Poly(A))序列:聚腺苷酸序列,也称为聚腺苷酸尾或3'-聚腺苷酸尾,典型地理解为例如多达约400个腺苷核苷酸的腺苷核苷酸序列,例如从约20至约400个,优选从约50至约400个,更优选从约50至约300个,甚至更优选从约50至约250个,最优选从约60至约250个腺苷核苷酸的腺苷核苷酸序列。聚腺苷酸序列典型地位于mRNA的3'末端。本发明范围内,聚腺苷酸序列可以例如通过载体的转录位于mRNA内或位于任何其它核酸分子内,诸如例如,载体内,例如用作产生RNA的模板的载体内,优选mRNA中。

[0056] 聚腺苷酸化:聚腺苷酸化典型地理解为向核酸分子,如RNA分子,例如成熟前mRNA添加聚腺苷酸序列。聚腺苷酸化可以由所谓聚腺苷酸化信号诱导。该信号优选位于要被聚腺苷酸化的核酸分子,如RNA分子的3'末端的一段核苷酸内。聚腺苷酸化信号典型地包含由腺嘌呤和尿嘧啶/胸腺嘧啶核苷酸组成的六聚体,优选为六聚体序列AAUAAA。其它序列,优选六聚体序列,也是可能的。聚腺苷酸化典型地发生在前mRNA(也称为成熟前-mRNA)的加工过程中。典型地,RNA成熟(从前mRNA到成熟mRNA)包含聚腺苷酸化的步骤。

[0057] 限制性位点:限制性位点,也称为‘限制性酶识别位点’,为由限制性酶识别的核苷酸序列。限制性位点典型地是短的,优选回文核苷酸序列,例如包含4至8个核苷酸的序列。限制性位点优选由限制性酶特异识别。限制性酶典型地切割在此位点包含限制性位点的核苷酸序列。在双链核苷酸序列,如双链DNA序列中,限制性酶典型地切割核苷酸序列的两条链。

[0058] RNA,mRNA:RNA是核糖核酸的常见缩写。其是核酸分子,即由核苷酸组成聚合物。这些核苷酸通常是沿所谓骨架彼此连接的腺苷-单磷酸酯、尿苷-单磷酸酯、鸟苷-单磷酸酯和胞苷-单磷酸酯单体。所述骨架由第一相邻单体的糖,即核糖和第二相邻单体的磷酸酯部分之间的磷酸二酯键形成。单体的特定连续被称为RNA-序列。通常RNA是可以通过例如在细胞内DNA-序列的转录获得的。真核细胞中,转录典型地在细胞核或线粒体中进行。通常,DNA的转录通常产生所谓的成熟前RNA,其必须被加工为所谓的信使-RNA,通常简称为mRNA。例如在真核生物体中,成熟前RNA的加工包括多种不同转录后修饰,如剪接、5'-加帽、聚腺苷酸化、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。成熟信使RNA通常提供可以被翻译为特定肽或蛋白氨基酸序列的核苷酸序列。典型地,成熟mRNA包含5'-帽、5'-UTR、可读框、3'-UTR和聚腺苷酸序列。除了信使RNA外,存在一些非编码RNA类型,其可以参与调节转录和/或翻译。

[0059] 核酸分子序列:核酸分子序列典型地理解为特定和独特的顺序,即其核苷酸的连续。蛋白或肽序列典型地理解为该顺序,即其氨基酸的连续。

[0060] 序列同一性:如果它们呈现相同长度和顺序的核苷酸或氨基酸,则两条以上的序列是相同的。同一性百分数典型地描述两条序列相同的程度,即其典型地描述在其序列位置上与参比序列相同的核苷酸对应的核苷酸的百分数。为了确定同一性程度,要比较的序列被认为是呈现相同长度,即要比较的序列的最长序列长度。这意为由8个核苷酸组成的第一序列与包含第一序列的由10个核苷酸组成的第二序列80%相同。换句话说,本发明范围内,序列同一性优选指具有相同长度的两条以上序列中具有相同位置的序列的核苷酸的百分数。通常将缺口认为是不相同的位置,而不考虑其在比对中的真正位置。

[0061] 稳定的核酸分子:稳定的核酸分子是被修饰从而其对例如由环境因素或酶消化,如由核酸外切酶或核酸内切酶降解导致的分解或降解比无修饰的核酸分子更稳定的核酸分子,优选DNA或RNA分子。优选地,本发明范围内,稳定的核酸分子在细胞中稳定,如原核或真核细胞中,优选在哺乳动物细胞,如人细胞中稳定。稳定效应也可以在细胞外发挥,例如在缓冲溶液等中,例如在包含所述稳定化的核酸分子的药物组合物的制造过程中发挥。

[0062] 转染:术语“转染”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。在本发明范围内,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。此种方法包括例如,电穿孔、例如基于阳离子脂质和/或脂质体的脂转染、磷酸钙沉淀、基于纳米粒子的转染、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)的转染等。优选地,所述引入是非病毒性的。

[0063] 翻译效率(或翻译的效率):用于本文时,该术语通常涉及包含可读框(ORF)的核酸分子(例如,mRNA)。翻译效率在实验上是可测量的。翻译效率通常通过测定由ORF翻译的蛋白的量而测量。对于翻译效率的实验测量,ORF优选编码报告蛋白或任意其他可以定量的蛋白。然而,不希望受限于具体的理论,应该理解,高翻译效率通常由特定的UTR元件(特定的5'-UTR元件或特定的3'-UTR元件)提供。由此,在本发明的情形中,术语翻译效率特别用于这样的核酸分子,其中,除了ORF之外,还包含至少一个5'-UTR元件和/或至少一个3'-UTR元件,优选如本文定义。尽管就实验定量翻译效率来说,ORF适当地编码报告蛋白或任意其他可以定量的蛋白,但是,本发明不限于此类目的;因此,本发明的至少一个5'-UTR元件和/或至少一个3'-UTR元件(其提供高翻译效率)可以被包含在含有不编码报告蛋白的ORF的核酸分子中。

[0064] 翻译效率是相对术语。由此,可以确定多种(例如,两种或更多种)核酸分子的翻译效率并进行比较,例如,通过实验定量由ORF编码的蛋白而确定和比较。这可以在标准条件下进行,例如,在本文所述的标准测定中进行。因此,在标准条件下确定的翻译效率是客观特征。优选地,确定并比较两种核酸分子的翻译效率。所述两种核酸分子中的第一种可以称为“受试核酸分子”或“受试构建体”,所述两种核酸分子中的第二种可以称为“参比核酸分子”或“参比构建体”。受试核酸分子可以是本发明所述的人工核酸分子。为了这一目的,参比核酸分子和受试核酸分子共有相同的ORF(相同的核酸序列);并且优选地,受试核酸分子的核酸序列与参比核酸分子的核酸序列相同,不同之处在于测试的UTR元件,即,5'-UTR元件或3'-UTR元件;换言之,优选地,受试核酸分子和参比核酸分子彼此的差异仅在于5'-UTR

元件或3'-UTR元件具有不同的核酸序列;使得5'-UTR元件或3'-UTR元件成为区分所述受试核酸分子和参比核酸分子的唯一的结构特征。

[0065] 在本文所述的测定中,将受试核酸分子和参比核酸分子转染到哺乳动物细胞中,并且确定翻译效率。当受试核酸分子的翻译效率高于参比核酸分子的翻译效率时,则认为所述受试核酸分子的特征是高翻译效率。在这一情形中,认为区分所述受试核酸分子和参比核酸分子的5'-UTR元件或3'-UTR元件(为所述受试核酸分子)提供了高翻译效率。换言之,认为高翻译效率是由存在于受试核酸分子中而不存在于参比核酸分子中的特定的5'-UTR元件或3'-UTR元件“提供的”;并且,认为包含此类5'-UTR或3'-UTR和至少一个ORF的受试核酸分子(或人工核酸分子)的“特征在于”高翻译效率。

[0066] 疫苗:疫苗典型地理解为提供至少一种抗原、优选免疫原的预防或治疗物质。所述抗原或免疫原可以源自任何适于疫苗接种的物质。例如,所述抗原或免疫原可以源自病原体,如源自细菌或病毒粒子等,或源自肿瘤或癌症组织。所述抗原或免疫原刺激身体的适应性免疫系统以提供适应性免疫应答。

[0067] 载体:术语“载体”指核酸分子,优选指人工核酸分子。在本发明范围内,载体适于结合或携带所需核酸序列,如包含可读框的核酸序列。此种载体可以是储存载体、表达载体、克隆载体、转移载体等。储存载体是允许方便储存核酸分子(例如mRNA分子)的载体。因此,所述载体可以包含对应于例如所需mRNA序列或其部分的序列,如对应于mRNA的可读框和3'UTR和/或5'UTR的序列。表达载体可以用于生产表达产物,诸如RNA(例如mRNA)、或肽、多肽或蛋白。例如,表达载体可以包含载体的一段序列的转录所需的序列,如启动子序列,例如RNA聚合酶启动子序列。克隆载体典型地是含有可以用于将核酸序列并入载体的克隆位点的载体。克隆载体可以是,例如质粒载体或噬菌体载体。转移载体可以是适于将核酸分子转移入细胞或生物体的载体,例如,病毒载体。在本发明范围内,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,在本申请的意义中的载体包含克隆位点、选择标记(如抗生素抗性因子)、和适于载体扩增的序列,如复制起点。优选地,本申请范围内的载体是质粒载体。

[0068] 赋形剂(vehicle):赋形剂典型地理解为适于储存、转运和/或施用化合物,如药理学活性化合物的材料。例如,其可以是生理学可接受的适于储存、转运和/或施用药学活性化合物的液体。

[0069] 3'-非翻译区(3'-UTR):通常,术语“3'-UTR”是指人工核酸分子的一部分,其位于可读框的3'(即“下游”),并且其不翻译为蛋白。通常,3'-UTR是mRNA的蛋白编码区(可读框(ORF)或编码序列(CDS))和聚腺苷酸序列之间的一部分mRNA。在本发明的情况中,术语3'-UTR还可以包含这样的元件,其不在模板中编码,RNA由其转录,但在成熟过程中在转录后添加,例如聚腺苷酸序列。mRNA的3'-UTR不翻译为氨基酸序列。3'-UTR序列通常由在基因表达过程中被转录为各自的mRNA的基因编码。基因组序列首先转录为包含任选内含子的成熟前mRNA。成熟前mRNA随后在成熟过程中进一步被加工为成熟mRNA。该成熟过程包含以下步骤:5'加帽、剪接成熟前mRNA以切除任选内含子和3'末端修饰(如成熟前mRNA 3'末端的聚腺苷酸化和任选的核酸内切酶/或核酸外切酶切割等)。在本发明范围内,3'-UTR对应于位于蛋白编码区终止密码子(优选紧接蛋白编码区终止密码子的3'端)与mRNA的聚腺苷酸序列之间的成熟mRNA序列。术语“对应于”意为3'-UTR序列可以是如用于限定3'-UTR序列的mRNA序

列中的RNA序列,或对应于此RNA序列的DNA序列。在本发明范围内,术语“基因的3'-UTR”,为对应于源自该基因的成熟mRNA的3'-UTR的序列,所述成熟mRNA即通过基因转录和成熟前mRNA的成熟获得的mRNA。术语“基因的3'-UTR”包括3'-UTR的DNA序列和RNA序列(正义和反义链二者以及成熟和未成熟的二者)。优选3'UTR具有多于20,30,40或50个核苷酸的长度。

[0070] 5'-非翻译区(5'-UTR):通常,术语“5'-UTR”是指人工核酸分子的一部分,其位于可读框的5'(即“上游”),并且其不翻译为蛋白。5'-UTR通常理解为信使RNA(mRNA)的特定区段,所述信使RNA位于mRNA的可读框的5'。通常,5'-UTR在转录起始位点起始并且在可读框的起始密码子之前的一个核苷酸处终止。优选地,所述5'UTR具有多于20,30,40或50个核苷酸的长度。5'-UTR可以包含用于控制基因表达的元件,也称为调控元件。所述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,例如通过加入5'-帽修饰。mRNA的5'-UTR不翻译为氨基酸序列。5'-UTR序列通常由基因表达过程中转录为各个mRNA的基因编码。基因组序列首先转录为成熟前mRNA,其包含任意的内含子。成熟前mRNA然后在成熟过程中进一步加工为成熟mRNA。该成熟过程包括以下步骤:5'加帽、剪接成熟前mRNA以切除任意的内含子和3'末端修饰(如成熟前mRNA 3'末端的聚腺苷酸化和任意的核酸内切酶/或核酸外切酶切割等)。在本发明范围内,5'-UTR对应于位于起始密码子和例如5'-帽之间的成熟mRNA序列。优选地,5'-UTR对应于从位于5'帽的3'侧的核苷酸,更优选从紧邻5'帽的3'侧核苷酸,向位于蛋白编码区起始密码子5'侧的核苷酸,优选向紧接蛋白编码区起始密码子5'侧的核苷酸延伸的序列。紧接成熟mRNA 5'-帽的3'侧的核苷酸典型地对应于转录起始位点。术语“对应于”意为5'-UTR序列可以是如用于限定5'-UTR序列的mRNA序列中的RNA序列,或与此RNA序列对应的DNA序列。在本发明范围内,术语“基因的5'-UTR”是对应于源自该基因的成熟mRNA的5'-UTR的序列,所述成熟mRNA即通过基因转录和成熟前mRNA的成熟获得的mRNA。术语“基因的5'-UTR”包括5'-UTR的DNA序列和RNA序列(正义和反义链二者以及成熟和未成熟的二者)。

[0071] 5'末端寡嘧啶束(TOP):5'末端寡嘧啶束(TOP)典型地是位于核酸分子的5'末端区域,如某些mRNA分子的5'末端区域或功能实体的5'末端区域,例如某些基因的转录区的一段嘧啶核苷酸。该序列以通常对应于转录起始位点的胞苷起始,并且接着一段通常约3至30个嘧啶核苷酸。例如, TOP可以包含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。所述一段嘧啶序列和由此的5'TOP终止于位于TOP下游的第一个嘌呤核苷酸的5'侧的一个核苷酸。含有5'末端寡嘧啶束的信使RNA常被称为TOP mRNA。因此,提供此种信使RNA的基因被称为TOP基因。例如,已经在编码肽延长因子和核糖体蛋白的基因和mRNA中发现了TOP序列。

[0072] TOP基序:在本发明范围内, TOP基序是对应于如上文限定的5'TOP的核酸序列。因此,本发明范围内的TOP基序优选为具有3-30个核苷酸长度的一段嘧啶核苷酸。优选地,所述TOP基序由至少3个嘧啶核苷酸,优选至少4个嘧啶核苷酸,优选至少5个嘧啶核苷酸,更优选至少6个核苷酸,更优选至少7个核苷酸,最优选至少8个嘧啶核苷酸组成,其中所述一段嘧啶核苷酸优选在其5'末端以胞嘧啶核苷酸起始。在TOP基因和TOP mRNA中,所述TOP-基序优选在其5'末端以转录起始位点起始并终止于所述基因或mRNA中的第一个嘌呤残基的5'侧一个核苷酸。本发明意义中的TOP基序优选位于表示5'-UTR的序列的5'末端或位于编码5'-UTR的序列的5'末端。因此,优选地,从本发明的意义上说,如果该段序列位于各自序列

(如人工核酸分子,人工核酸分子的5'-UTR元件,或源自如本文描述的TOP基因的5'-UTR的核酸序列)的5'末端,则一段3个以上嘧啶核苷酸的序列被称为“TOP基序”。换句话说,不位于5'-UTR或5'-UTR元件的5'末端而是位于5'-UTR或5'-UTR元件内任何位置的一段3个以上嘧啶核苷酸优选不称为“TOP基序”。

[0073] TOP基因:TOP基因典型地以存在5'末端寡嘧啶束为特征。此外,大多数TOP基因以生长相关翻译调节为特征。然而,还已知TOP基因有组织特异翻译调节。如上文限定的,TOP基因的5'-UTR对应于源自TOP基因的成熟mRNA的5'-UTR序列,其优选从位于5'-帽的3'侧的核苷酸向位于起始密码子5'侧的核苷酸延伸。TOP基因的5'-UTR典型地不包含任何起始密码子,优选无上游AUG(uAUG)或上游可读框(uORF)。其中,上游AUG和上游可读框典型地理解为存在于应该被翻译的可读框的起始密码子(AUG)5'侧的AUG和可读框。TOP基因的5'-UTR通常相当短。TOP基因的5'-UTR的长度可以在20个核苷酸到高达500个核苷酸之间变化,并且典型地少于约200个核苷酸,优选少于约150个核苷酸,更优选少于约100个核苷酸。本发明意义中典型的TOP基因的5'-UTR是在根据专利申请W02013/143700的SEQ ID Nos.1-1363的序列中,从位置5的核苷酸到紧接起始密码子(例如ATG)5'侧的核苷酸延伸的核酸序列,W02013/143700的公开内容通过引用结合于本文。在该情况下,特别优选的TOP基因的5'-UTR的片段是缺少5' TOP基序的TOP基因的5'-UTR。术语“TOP基因的5'-UTR”或“5'-TOP UTR”优选指天然存在的TOP基因的5'-UTR。优选的实例为SEQ ID NO:208所示(缺少5'末端寡嘧啶束的人大核糖体蛋白32的5'-UTR);其对应于专利申请W02013/143700的SEQ ID NO.1368。

[0074] 野生型,例如,野生型核酸分子:术语“野生型”可以理解为天然存在的序列。野生型核酸分子通常可以理解为天然存在的核酸分子,例如,DNA或RNA。换言之,人工核酸分子可以理解为天然核酸分子。此类核酸分子可以由于其各自的序列(其是天然存在的)和/或由于其他天然存在的修饰(例如,核苷酸的结构修饰)而是天然的。野生型核酸分子可以是DNA分子、RNA分子或包含DNA和RNA部分的杂合分子。此处,术语“野生型”是指只要是天然存在的任何分子,不需要在公众可访问的序列库(如GenBank)中有反映。国立卫生研究院(National Institute of Health,NIH)提供公众可访问的注解的公众可获得的核苷酸序列的库(“GenBank”,可通过NCBI Entrez检索系统访问:<http://www.ncbi.nlm.nih.gov>), (Nucleic Acids Research,2013;41 (D1):D36-42),包括公众可获得的野生型序列。每个GenBank记录分配特有的不变的标识符,称为登记号,并且出现在GenBank记录的登记线上;并且通过出现在GenBank记录的版本线上的登记号的整数延续追踪序列信息的变化。此外,术语“野生型核酸分子”不限于表示“一种单个分子”,而典型地理解为包含全体相同分子。因此,它可以涉及等分部分中含有的多个相同分子。

[0075] 详述

[0076] 本发明涉及包含至少一个可读框和至少一个3'-非翻译区元件(3'-UTR元件)和/或至少一个5'-非翻译区元件(5'-UTR元件)的人工核酸分子,其中所述人工核酸分子特征在于高翻译效率。所述翻译效率至少部分归因于5'-UTR元件或3'-UTR元件,或归因于5'-UTR元件与3'-UTR元件二者。本发明还涉及此类人工核酸分子在基因治疗和/或基因疫苗接种中的用途。此外,提供了新型的3'-UTR元件和5'-UTR元件。

[0077] 在第一个方面,本发明涉及人工核酸分子,其包含:

[0078] a. 至少一个可读框 (ORF) ; 和

[0079] b. 至少一个3' -非翻译区元件 (3' -UTR元件) 和/或至少一个5' -非翻译区元件 (5' -UTR元件), 其中所述人工核酸分子的特征为高翻译效率。

[0080] 例如, 本发明的人工核酸分子与其他 (例如, 野生型或人工) 核酸分子的不同可能在于至少一个3' -UTR元件 (优选一个3' -UTR元件) 被至少一个不同的3' -UTR元件 (优选一个3' -UTR元件) 替换, 或至少一个5' -UTR元件 (优选一个5' -UTR元件) 被至少一个不同的5' -UTR元件 (优选一个5' -UTR元件) 替换。

[0081] 通常, (在起始核酸分子或参比核酸分子中) 替换一个 (或至少一个) 3' -UTR元件或替换一个 (或至少一个) 5' -UTR元件通常意指所述3' -UTR元件或所述5' -UTR元件的至少一些核苷酸的序列被不同的核酸序列替换。优选地, 当替换5' -UTR元件时, 5' -UTR元件的连续核苷酸序列被不同的连续核苷酸序列 (元件) 替换; 并且当替换3' -UTR元件时, 3' -UTR元件的连续核苷酸序列被不同的连续核苷酸序列 (元件) 替换。初始 (被替换的) 元件的连续核苷酸序列和替换元件的连续核苷酸序列分别可以独立地是任意长度, 如1至大于500个核苷酸, 20至500个核苷酸, 40至400个核苷酸, 60至300个核苷酸, 80至200个核苷酸或100至150个核苷酸。当替换 (5' -或3' -) UTR元件时, 这不需要意指替换所有位于起始密码子5' 侧的核苷酸或所有位于终止密码子3' 侧的核苷酸。在优选的实施方案中, 替换连续的序列, 所述连续的序列是分别位于起始密码子5' 侧的所有核苷酸和位于终止密码子3' 侧的所有核苷酸的子集。其实例显示在图1B (5') 和1C (3') 中。可以准确地替换存在于起始序列或参比序列中的已知的5' -UTR元件或已知的3' -UTR元件。例如, 当 (例如, 参比) 核酸包含对应于SEQ ID NO:208的序列的5' -UTR元件时; 本发明的人工核酸分子可以通过用本发明所述的不同的连续5' -UTR元件 (即, 提供高翻译效率的5' -UTR元件) 准确替换SEQ ID NO:208的完整连续序列而制备。备选地, 可能的是, 已知的UTR元件没有被准确地替代, 例如, 包含另外的核苷酸的连续的核酸序列没有被准确地替代, 即, 除了已知的UTR元件被替代之外, 或者, 例如, 仅替换已知的UTR元件的一部分 (通常是主要部分, 即, 长度的90%以上)。尽管这些可能性已经对5' -UTR元件的情形进行了示例, 但是对于3' -UTR元件也存在相同的可能性。换言之, 所替换的UTR元件不需要被已知UTR元件的准确边界限制。这些可能性通过举例的方式分布进一步在图1B和图1C中示例, 分别与图1A进行比较。类似地, 替换的UTR元件不需要被已知UTR元件的准确边界限制。

[0082] 优选地, 本发明所述的人工核酸分子不包含核糖体蛋白S6、RPL36AL、rps16或核糖体蛋白L9的3' -UTR (元件) 和/或5' -UTR (元件)。更优选地, 本发明所述的人工核酸分子不包含核糖体蛋白S6、RPL36AL、rps16或核糖体蛋白L9的3' -UTR (元件) 和/或5' -UTR (元件), 并且本发明所述的人工核酸分子的可读框不编码GFP蛋白。甚至更优选地, 本发明所述的人工核酸分子不包含核糖体蛋白S6、RPL36AL、rps16或核糖体蛋白L9的3' -UTR (元件) 和/或5' -UTR (元件), 并且本发明所述的人工核酸分子的可读框不编码例如选自由以下各项组成的组的报告蛋白: 球蛋白 (特别是 $\beta$ -球蛋白), 荧光素酶蛋白, GFP蛋白, 葡萄糖醛酸糖苷酶蛋白 (特别是 $\beta$ -葡萄糖醛酸糖苷酶) 或其变体, 例如, 显示与球蛋白, 荧光素酶蛋白, GFP蛋白, 或葡萄糖醛酸糖苷酶蛋白至少70%序列同一性的变体。

[0083] 术语“3' -UTR元件”是指包含以下核酸序列或由以下核酸序列组成的核酸序列: 源自3' -UTR或源自3' -UTR的变体或片段的核酸序列。“3' -UTR元件”优选是指人工核酸序列

(如人工mRNA)的3'-UTR包含的核酸序列。因此,在本发明的意义中,优选地,3'-UTR元件可以由mRNA(优选人工mRNA)的3'-UTR包含,或3'-UTR元件可以由各自的转录模板的3'-UTR包含。优选地,3'-UTR元件是对应于mRNA的3'-UTR、优选人工mRNA(如通过转录基因改造的载体构建体获得的mRNA)的3'-UTR的核酸序列。优选地,本发明的意义中的3'-UTR元件作为3'-UTR行使功能或编码执行3'-UTR的功能的核苷酸序列。

[0084] 因此,术语“5'-UTR元件”是指是指包含以下核酸序列或由以下核酸序列组成的核酸序列:源自5'-UTR或源自5'-UTR的变体或片段的核酸序列。“5'-UTR元件”优选是指人工核酸序列(如人工mRNA)的5'-UTR包含的核酸序列。因此,在本发明的意义中,优选地,5'-UTR元件可以由mRNA(优选人工mRNA)的5'-UTR包含,或5'-UTR元件可以由各自的转录模板的5'-UTR包含。优选地,5'-UTR元件是对应于mRNA的5'-UTR、优选人工mRNA(如通过转录基因改造的载体构建体获得的mRNA)的5'-UTR的核酸序列。优选地,本发明的意义中的5'-UTR元件作为5'-UTR行使功能或编码执行5'-UTR的功能的核苷酸序列。

[0085] 本发明所述的人工核酸分子中的3'-UTR元件和/或5'-UTR元件为所述人工核酸分子提供高翻译效率。因此,本发明所述的人工核酸分子可以尤其包含:

[0086] 一为所述人工核酸分子提供高翻译效率的3'-UTR元件,

[0087] 一为所述人工核酸分子提供高翻译效率的5'-UTR元件,

[0088] 一为所述人工核酸分子提供高翻译效率的3'-UTR元件和为所述人工核酸分子提供高翻译效率的5'-UTR元件。

[0089] 优选地,本发明所述的人工核酸分子包含为所述人工核酸分子提供高翻译效率的3'-UTR元件和/或为所述人工核酸分子提供高翻译效率的5'-UTR元件。

[0090] 优选地,本发明所述的人工核酸分子包含至少一个3'-UTR元件和至少一个5'-UTR元件,即至少一个为所述人工核酸分子提供高翻译效率的3'-UTR元件和延长和至少一个为所述人工核酸分子提供高翻译效率的5'-UTR元件。

[0091] 如下文详细所述,所述至少一个为所述人工核酸分子提供高翻译效率的3'-UTR元件或所述至少一个为所述人工核酸分子提供高翻译效率的5'-UTR元件可以选自天然存在的(优选异源的)3'-UTR元件和5'-UTR元件(同时为天然存在的UTR元件或野生型UTR元件),并且选自人工3'-UTR元件和人工5'-UTR元件(同时为人工UTR元件)。野生型UTR元件可以选自包括在参考文献和公众可获得的数据库(如GenBank(NCBI))中公布的野生型UTR元件与之前没有公布的野生型UTR元件的组。后者可以通过对细胞、优选哺乳动物细胞中存在的mRNA测序进行鉴定,利用该方法,本发明人鉴定了一些之前没有公布的野生型UTR元件,本发明提供了这种类型的UTR元件。术语人工UTR元件没有特别限制,并且是指自然中没有发现的任意核酸序列,即,与野生型UTR元件不同。然而,在优选的实施方案中,用于本发明的人工UTR元件是表现出与野生型UTR元件有特定程度的(如10-99.9%,20-99%,30-98%,40-97%,50-96%,60-95%,70-90%)序列同一性的核酸序列。在优选的实施方案中,除了一个、或两个、或三个、或四个、或五个、或多于五个核苷酸已被相同数量的核苷酸置换(例如,一个核苷酸被一个核苷酸置换)之外,用于本发明的人工UTR与野生型UTR相同。优选地,一个核苷酸置换是各自互补核苷酸的置换。优选的人工UTR元件对应于野生型UTR元件,不同之处在于:(i)野生型5'-UTR元件中的一些或全部的ATG三联体(如果存在)转化为三联体TAG;和/或(ii)通过用互补核苷酸置换在所述特定限制性酶的切割位点内的一个核苷



酸,而消除野生型5' -UTR元件或野生型3' -UTR元件中所选的用于特定限制性酶的切割位点(如果存在),由此去除所述特定限制性酶的切割位点。当(例如,野生型)UTR元件包含用于所述特定限制性酶的切割位点,并且当所述特定限制性酶用于(计划用于)后续克隆步骤时,通常需要后者。由于5' -UTR元件和3' -UTR元件的此种内部切割是不合乎需要的,可以产生消除了所述特定限制性酶的限制性切割位点的人工UTR元件。可以通过本领域技术人员已知的任何适当的方法进行此类置换,例如,使用修饰的引物通过PCR进行。

[0092] 任选地,本发明所述的人工核酸分子包含至少一个3' -UTR元件和至少一个5' -UTR元件,即,至少一个延长和/或增加从所述人工核酸分子的蛋白产生的3' -UTR元件和至少一个延长和/或增加从所述人工核酸分子的蛋白产生的5' -UTR元件。

[0093] 蛋白产生;确定是否延长和/或增加蛋白产生的测定

[0094] “延长和/或增加从所述人工核酸分子的蛋白产生”通常是指与从缺少3' -UTR和/或5' -UTR或包含参比3' -UTR和/或参比5' -UTR(如天然与ORF组合存在的3' -UTR和/或5' -UTR)的各个参比核酸产生的蛋白的量相比,从具有各个3' -UTR元件和/或5' -UTR元件的本发明所述的人工核酸分子产生的蛋白的量。

[0095] 尤其是,与缺少3' -UTR和/或5' -UTR或包含参比3' -UTR和/或5' -UTR(如天然与ORF组合存在的3' -和/或5' -UTR)的各个核酸相比,本发明所述的人工核酸分子的至少一个3' -UTR元件和/或5' -UTR元件延长从本发明所述的人工核酸分子(例如从本发明所述的mRNA)的蛋白产生。

[0096] 尤其是,与缺少3' -和/或5' -UTR或包含参比3' -和/或5' -UTR(如天然与ORF组合存在的3' -和/或5' -UTR)的各个核酸相比,本发明所述的人工核酸分子的至少一个3' -UTR元件和/或5' -UTR元件增加从本发明所述的人工核酸分子(例如从本发明所述的mRNA)的蛋白产生,尤其是蛋白表达和/或总蛋白产生。

[0097] 优选地,与缺少3' -UTR和/或5' -UTR或包含参比3' -UTR和/或参比5' -UTR(如天然与ORF组合存在的3' -UTR和/或5' -UTR)的各个核酸的翻译效率相比,本发明所述的人工核酸分子的所述至少一个3' -UTR元件和/或所述至少一个5' -UTR元件不消极影响核酸的翻译效率。备选地,与其天然情况下各个ORF编码的蛋白的翻译效率相比,翻译效率由3' -UTR和/或5' -UTR增强。

[0098] 本文中使用的术语“各个核酸分子”或“参比核酸分子”意为——除不同3' -UTR和/或5' -UTR之外——参比核酸分子是与包含3' -UTR元件和/或5' -UTR元件的本发明的人工核酸分子相当的,优选相同的。

[0099] 为了评估本发明的人工核酸分子的本文定义的体内或体外的蛋白产生(即体外涉及(“活体”)细胞和/或组织(包括活体受试者的组织);细胞尤其包括细胞系,原代细胞,组织或受试者中的细胞,优选为哺乳动物细胞,例如人细胞和小鼠细胞并且特别优选是人细胞系HeLa,HEPG2和U-937以及小鼠细胞系NIH3T3,JAWSII和L929,此外原代细胞是特别优选的,在尤其优选的实施方案中,人皮肤成纤维细胞(HDF)),在将本发明的人工核酸分子注射/转染入靶细胞/组织后确定编码的蛋白的表达,并且与由参比核酸诱导的蛋白表达相比。用于确定蛋白表达的定量方法是本领域中已知的(例如蛋白质印迹,FACS,ELISA,质谱)。在该情况下特别有用的是确定报告蛋白如荧光素酶、绿色荧光蛋白(GFP)或分泌的碱性磷酸酶(SEAP)的表达。因此,将本发明所述的人工核酸或参比核酸例如通过转染或注射



引入靶组织或细胞,优选在哺乳动物表达系统中,如在哺乳动物细胞中,例如HDF、L929、HepG2和/或HeLa细胞中。表达起始后或引入核酸分子后的数小时或数天(例如6,12,24,48或72小时),收集靶细胞样品并通过FACS测量和/或裂解。之后,裂解产物可以用于使用多种方法,例如蛋白质印迹、FACS、ELISA、质谱或通过荧光或发光测量检测表达的蛋白(并且由此确定蛋白表达的效率)。

[0100] 因此,如果将在特定时间点(例如表达起始后或引入核酸分子后6,12,24,48或72小时)本发明所述的人工核酸分子的蛋白表达与从参比核酸分子的蛋白表达相比,两种核酸分子分开引入到靶组织/细胞中,在特定时间点后收集组织/细胞的样品,根据对特定检测方法(例如蛋白质印迹,ELISA,荧光或发光测量等,如本领域中已知的)调整的特定方案制备蛋白裂解产物,并且通过选择的检测方法检测蛋白。作为细胞裂解产物中表达的蛋白量的测量的备选方案——或者,除了在裂解收集的细胞之前测量细胞裂解产物中的蛋白量或平行使用等分部分——蛋白量也可以通过使用FACS分析确定。

[0101] 术语“延长”从人工核酸分子如人工mRNA的“蛋白产生”优选意为,优选在哺乳动物表达系统,如在HDF、L929、HEP2G或HeLa细胞中,与从参比核酸分子如参比mRNA(例如包含参比3'-和/或5'-UTR或缺少3'-和/或5'-UTR)的蛋白产生相比,从人工核酸分子如人工mRNA的蛋白产生延长。因此,与对于从参比核酸分子产生的蛋白可以观察到的相比,从工核酸分子如人工mRNA产生的蛋白可更长期观察到。换句话说,在较晚的时间点,例如转染后48小时或72小时测量的从人工核酸分子如人工mRNA产生的蛋白量,比在相应较晚的时间点从参比核酸分子如参比mRNA产生的蛋白量大。所述“较晚时间点”可以是,例如,超过起始表达后,如转染核酸分子后24小时的任意时间,例如起始表达,即转染后36,48,60,72,96小时。此外,对于相同核酸,可以将较晚时间点产生的蛋白量针对在较早(参考)时间点产生的量进行标准化,例如在较晚时间点的蛋白量可以表示为转染后24小时的蛋白量的百分数。

[0102] 优选地,该延长蛋白产生的效果通过以下步骤确定:(i)测量,例如优选在哺乳动物表达系统中如HDF、L929、HEP2G或HeLa细胞中通过表达编码的报告蛋白如荧光素酶获得的随时间的蛋白量,(ii)确定在“参考”时间点 $t_1$ ,例如后 $t_1=24$ 小时观察到的蛋白量,并且将该蛋白量设为100%,(iii)确定在一个或多个较晚时间点 $t_2$ 、 $t_3$ 等,例如转染后 $t_2=48$ 小时和 $t_3=72$ 小时观察到的蛋白量,并且将在较晚时间点观察到的蛋白的相对量计算为在时间点 $t_1$ 的蛋白量的百分数。例如,在 $t_1$ 以“80”的量表达,在 $t_2$ 以“20”的量表达,并且在 $t_3$ 以“10”的量表达的蛋白,蛋白在 $t_2$ 的相对量将是25%,并且在 $t_3$ 的为12,5%。可以随后将在较晚时间点的这些相对量在步骤(iv)中与对于分别缺少3'-和/或5'-UTR或分别包含参比3'-和/或5'-UTR的核酸分子在相应时间点的相对蛋白量相比。通过比较从本发明所述的人工核酸分子产生的相对蛋白量与从参比核酸分子(即分别缺少3'-和/或5'-UTR或分别包含参比3'-和/或5'-UTR的核酸分子)产生的相对蛋白量,可以确定造成与从参比核酸分子的蛋白产生相比,从本发明所述的人工核酸分子的蛋白产生延长的因素。

[0103] 优选地,与从分别缺少3'-和/或5'-UTR或分别包含参比3'-和/或5'-UTR的参比核酸分子的蛋白产生相比,本发明所述的人工核酸分子的至少一个3'-和/或5'-UTR元件延长从所述人工核酸分子的蛋白产生至少1.2倍,优选至少1.5倍,更优选至少2倍,甚至更优选至少2.5倍。换句话说,对于相同的较晚时间点,与从参比核酸分子(其例如分别缺少3'-和/或5'-UTR或分别包含参比3'-和/或5'-UTR)产生的蛋白的(相对)量相比,在如上文所述的

某较晚时间点从本发明所述的人工核酸分子产生的蛋白的(相对)量以至少1.2,优选至少1.5,更优选至少2,甚至更优选至少2.5的倍数增加。

[0104] 备选地,延长蛋白产生的效应也可以通过以下步骤确定:(i) 测量例如优选在哺乳动物表达系统如在HDF、L929、HEP2G或HeLa细胞中通过表达编码的报告蛋白如荧光素酶获得的随时间的蛋白量,(ii) 确定蛋白量低于例如,在起始表达后1,2,3,4,5,或6小时,例如转染人工核酸分子后1,2,3,4,5,或6小时观察到的蛋白量的时间点,并且(iii) 将蛋白量低于在起始表达后1,2,3,4,5,或6小时观察到的蛋白量的时间点与对于分别缺少3' -和/或5' -UTR或分别包含参比3' -和/或5' -UTR的核酸分子确定的所述时间点相比。

[0105] 例如,与从参比核酸分子(如参比mRNA)的蛋白产生相比,在哺乳动物表达系统中,如在哺乳动物细胞中,例如在HDF、L929、HEP2G或HeLa细胞中,从人工核酸分子如人工mRNA的蛋白产生——以是至少在表达的起始阶段中,如起始表达后,如转染核酸分子后1,2,3,4,5,或6小时观察到的量的量——延长至少约5小时,优选至少约10小时,更优选至少约24小时。因此,与分别缺少3' -和/或5' -UTR或分别包含参比3' -和/或5' -UTR的参比核酸分子相比,本发明所述的人工核酸分子优选允许将以是至少在表达的起始阶段,如表达起始后,如转染后1,2,3,4,5,或6小时观察到的量的量的蛋白产生延长至少约5小时,优选至少约10小时,更优选至少约24小时。

[0106] 在优选的实施方案中,与从分别缺少3' -和/或5' -UTR或分别包含参比3' -和/或5' -UTR的参比核酸分子的蛋白产生相比,从本发明所述的人工核酸分子的蛋白产生期延伸至少1.2倍,优选至少1.5倍,更优选至少2倍,甚至更优选至少2.5倍。

[0107] 优选地,实现蛋白产生的该延长效果,同时例如在48或72小时的时间间隔内从本发明所述的人工核酸分子产生的蛋白总量,至少对应于从分别缺少3' -和/或5' -UTR或分别包含参比3' -和/或5' -UTR(如天然与人工核酸分子的ORF存在的3' -UTR和/或5' -UTR)的参比核酸分子产生的蛋白量。由此,本发明提供允许如上文所述在哺乳动物表达系统中,如哺乳动物细胞中,例如在HDF、L929、HEP2G或HeLa细胞中延长蛋白产生的人工核酸分子,其中从所述人工核酸分子(例如在48或72小时的时间间隔内)产生的蛋白总量,是至少例如在所述时间间隔内从分别缺少3' -和/或5' -UTR或分别包含参比3' -和/或5' -UTR(如天然与人工核酸分子的ORF存在的3' -和/或5' -UTR)的参比核酸分子产生的蛋白的总量。

[0108] 此外,术语“延长的蛋白表达”还包括“稳定化的蛋白表达”,由此“稳定化的蛋白表达”优选意为,当与参比核酸分子(例如,分别包含参比3' -和/或5' -UTR或分别缺少3' -和/或5' -UTR的mRNA)相比时,存在经预定的时期,如经24小时,更优选经48小时,甚至更优选经72小时更均一的从本发明所述的人工核酸分子的蛋白产生。

[0109] 因此,例如在哺乳动物系统中,从本发明所述的包含3' -和/或5' -UTR元件的人工核酸分子,例如从本发明所述的mRNA的蛋白产生水平,优选不降低至对于参比核酸分子,诸如如上文所述的参比mRNA观察到的程度。为了评估从具体核酸分子的蛋白产生降低至的程度,例如,起始表达后24小时,例如将本发明所述的人工核酸分子转染入细胞,如哺乳动物细胞后24小时观察到的蛋白(由各个ORF编码)的量,可以与起始表达后48小时,例如转染后48小时观察到的蛋白量相比。因此,在较晚时间点,例如,起始表达后(例如转染后)48小时观察到的本发明所述的人工核酸分子的ORF编码的蛋白量,如报告蛋白,例如,荧光素酶的量,与在较早时间点,例如,起始表达后(例如转染后)24小时观察到的蛋白量的比率,优选

高于对于分别包含参比3' -和/或5' -UTR或分别缺少3' -和/或5' -UTR的参比核酸分子的相应比率(包括相同时间点)。

[0110] 优选地,在较晚时间点,例如,起始表达后(例如转染后)48小时观察到的本发明所述的人工核酸分子的ORF编码的蛋白量,如报告蛋白,例如,荧光素酶的量,与在较早时间点,例如,起始表达后(例如转染后)24小时观察到的蛋白量的比率,优选为至少0.2,更优选至少约0.3,甚至更优选至少约0.4,甚至更优选至少约0.5,并且特别是优选至少约0.7。对于各个参比核酸分子,例如分别包含参比3' -和/或5' -UTR或分别缺少3' -和/或5' -UTR的mRNA,所述比率可以在,例如约0.05至约0.35之间。

[0111] 由此,本发明提供人工核酸分子,其如上文所述的包含ORF和3' -和/或5' -UTR元件,其中优选在哺乳动物表达系统中,如在哺乳动物细胞中,例如在HDF细胞中或在HeLa细胞中,起始表达后48小时观察到的蛋白量,例如荧光素酶量与起始表达后24小时观察到的蛋白量的比率,优选至少为0.2,更优选至少为约0.3,更优选至少为约0.4,甚至更优选至少为约0.5,甚至更优选至少为约0.6,并且特别是优选至少为约0.7。由此,优选例如在48小时的时间间隔内从所述人工核酸分子产生的蛋白总量,至少对应于例如在所述时间间隔内,从分别缺少3' -和/或5' -UTR或分别包含参比3' -和/或5' -UTR(如天然与人工核酸分子的ORF存在的3' -UTR和/或5' -UTR)的参比核酸分子产生的蛋白总量。

[0112] 优选地,本发明提供如上文所述的包含ORF和3' -UTR元件和/或5' -UTR元件的人工核酸分子,其中优选在哺乳动物表达系统中,如在哺乳动物细胞中,例如在HeLa细胞或HDF细胞中,起始表达后72小时观察到的蛋白量,例如荧光素酶量与起始表达后24小时观察到的蛋白量的比率优选高于约0.05,更优选高于约0.1,更优选高于约0.2,甚至更优选高于约0.3,其中优选例如在72小时的时间间隔内从所述人工核酸分子产生的蛋白的总量,至少是例如在所述时间间隔内,从分别缺少3' -和/或5' -UTR或分别包含参比3' -和/或5' -UTR(如天然与人工核酸分子的ORF存在的3' -和/或5' -UTR)的参比核酸分子产生的蛋白总量。

[0113] 在本发明的情况下“增加的蛋白表达”或“增强的蛋白表达”优选意为与由参比核酸分子引起的表达相比,在起始表达后的一个时间点增加的/增强的蛋白表达或增加的/增强的表达的蛋白的总量。由此,在起始表达后,例如转染本发明所述的人工核酸分子后,例如转染本发明所述的mRNA后某时间点,例如,转染后6,12,24,48或72小时观察到的蛋白水平,优选高于起始表达后,例如转染参比核酸分子,如分别包含参比3' -和/或5' -UTR或分别缺少3' -和/或5' -UTR的参比mRNA后的相同时间点观察到的蛋白水平。在优选的实施方案中,就从分别包含参比3' -和/或5' -UTR或分别缺少3' -和/或5' -UTR的参比核酸表达的蛋白量而言,从人工核酸分子表达的蛋白的最大量(如例如通过蛋白活性或质量确定的)增加。优选例如,转染后在48小时内,更优选24小时内和甚至更优选12小时内达到峰表达水平。

[0114] 优选地,术语“增加的”本发明所述的人工核酸分子的“总蛋白产生”或“增强的”本发明所述的人工核酸分子的“总蛋白产生”是指优选在哺乳动物表达系统中,如在哺乳动物细胞中,例如在HDF、L929、HEP2G或HeLa细胞中,与分别缺少3' -和/或5' -UTR或分别包含参比3' -和/或5' -UTR的参比核酸分子相比,经从人工核酸分子生产蛋白的时间间隔,例如48小时或72小时,增加的/增强的蛋白产生。根据优选的实施方案,当使用本发明所述的人工核酸分子时,表达的蛋白随时间的积累量增加。

[0115] 对于特定时期的蛋白总量可以通过以下步骤确定:(i)在引入人工核酸分子后的

多个时间点(例如起始表达后或引入核酸分子后6,12,24,48和72小时)收集组织或细胞,并且每个时间点的蛋白量可以如上文所述确定。为了计算积累的蛋白量,可以使用确定蛋白的总量的数学方法,例如曲线下面积(AUC)可以根据下式确定:

$$[0116] \quad AUC = \int_a^b f(x) d(x)$$

[0117] 为了计算蛋白总量的曲线下面积,从每个端点(a和b)计算表达曲线的等式的积分。

[0118] 由此,“总蛋白产生”优选是指表示蛋白产生相对时间的曲线下面积(AUC)。

[0119] 优选地,与从分别缺少3'-和/或5'-UTR的参比核酸分子的蛋白产生相比,本发明所述的至少一个3'-或5'-UTR元件增加从所述人工核酸分子的蛋白产生至少1.5倍,优选至少2倍,更优选至少2.5倍。换句话说,与对于相应较晚时间点从参比核酸分子(其例如分别缺少3'-和/或5'-UTR或分别包含参比3'-和/或5'-UTR)产生的蛋白的(相对)量相比,在某时间点,例如起始表达后,例如转染后48小时或72小时,从本发明所述的人工核酸分子产生的蛋白总量以至少1.5,优选至少2,更优选至少2.5的倍数增加。

[0120] 3'-UTR和/或5'-UTR的变体、片段和/或变体片段的mRNA和/或蛋白产生延长效果和效率和/或蛋白产生增加效果和效率以及本发明所述的人工核酸分子的所述至少一个3'-UTR元件和/或所述至少一个5'-UTR元件的mRNA和/或蛋白产生延长效果和效率和/或蛋白产生增加效果和效率可以通过技术人员已知的适于此目的的任意方法确定。

[0121] 例如,可以产生人工mRNA分子,其包含报告蛋白(如荧光素酶)的编码序列/可读框(ORF),和本发明所述的3'-UTR元件,即其延长和/或增加从所述人工mRNA分子的蛋白产生。此外,这样的发明的mRNA分子还可以包含本发明所述的5'-UTR元件,即其延长和/或增加从所述人工mRNA分子的蛋白产生,不包含5'-UTR元件或包含不是本发明所述的5'-UTR元件,例如参比5'-UTR。因此,可以产生人工mRNA分子,其包含报告蛋白(如荧光素酶)的编码序列/可读框(ORF),和本发明所述的5'-UTR元件,即其延长和/或增加从所述人工mRNA分子的蛋白产生。此外,这样的发明的mRNA分子还可以包含本发明所述的3'-UTR元件,即其延长和/或增加从所述人工mRNA分子的蛋白产生,不包含3'-UTR元件或包含不是本发明所述的3'-UTR元件,例如参比3'-UTR。

[0122] 根据本发明,可以例如,通过体外转录各个载体如质粒载体产生mRNA,所述载体例如包含T7启动子和编码各个mRNA序列的序列。产生的mRNA分子可以通过适于转染mRNA的转染方法转染入细胞,例如它们可以脂转染入哺乳动物细胞,如HeLa细胞或HDF细胞,并且可以在转染后的某时间点,例如,转染后6小时,24小时,48小时,和72小时分析样品。可以通过技术人员公知的方法分析所述样品的mRNA量和/或蛋白量。例如,在样品时间点存在于细胞中的报告mRNA的量可以通过定量PCR方法的确定。由各个mRNA编码的报告蛋白的量可以例如,通过蛋白质印迹,ELISA测定,FACS分析或报告子测定如荧光素酶测定确定,这取决于使用的报告蛋白。稳定蛋白表达和/或延长蛋白表达的效果可以,例如,通过确定转染后48小时观察到的蛋白水平和转染后24小时观察到的蛋白水平的比率分析。所述值越接近1,在该时期内的蛋白表达越稳定。所述测量当然还可以在72小时或更长的小时进行,并且可以确

定转染后72小时观察到的蛋白水平和转染后24小时观察到的蛋白水平的比率,从而确定蛋白表达的稳定性。

[0123] 翻译效率;确定翻译效率的测定

[0124] 如上文详细所述,翻译效率是包含可读框(ORF)的核酸分子(例如,mRNA)的特性,并且通常由所述核酸分子所包含的特定UTR元件(特定的5'-UTR元件或特定的3'-UTR元件)提供。翻译效率通常通过从ORF翻译的蛋白的量测定。在这一方面,本发明提供通用方法以及具体的测定法。例如,通过实验定量ORF编码的蛋白,可以比较多种(例如,两种以上)核酸分子的翻译效率。

[0125] 通常,为了评估翻译效率,使用基于细胞的方法。如本文限定那样在体内或体外实验检测本发明的人工核酸分子的翻译效率(即,体外是指(“活的”)细胞和/或组织,包括活受试者的组织;细胞包括在特定细胞系、原代细胞、组织或受试者中的细胞中,优选哺乳动物细胞,例如,人细胞和小鼠细胞,特别优选人细胞系HeLa、HEPG2和U-937和小鼠细胞系NIH3T3、JAWSII和L929,此外特别优选原代细胞,特别优选的实施方案是人皮肤成纤维细胞(HDF)),在将本发明的人工核酸分子注射/转染到靶细胞/组织中后确定所编码的蛋白的表达,并与由参比核酸诱导的蛋白表达进行比较。用于确定蛋白表达的定量方法在本领域中是已知的(例如,蛋白质印迹,FACS,ELISA,质谱)。在本情形中特别有用的是确定报告蛋白的表达,所述报告蛋白如荧光素酶、绿色荧光蛋白(GFP)或分泌的碱性磷酸酶(SEAP)。因此,将本发明所述的人工核酸或参比核酸例如通过转染或注射引入靶组织或细胞,优选在哺乳动物表达系统中,如在哺乳动物细胞中,例如HDF、L929、HepG2和/或HeLa细胞中。表达起始后或引入核酸分子后的数小时或数天(例如6,12,24,48或72小时),收集靶细胞样品并通过FACS测量和/或裂解。之后,裂解产物可以用于使用多种方法,例如蛋白质印迹、FACS、ELISA、质谱或通过荧光或发光测量检测表达的蛋白(并且由此确定蛋白表达的效率)。

[0126] 因此,如果将在特定时间点(例如表达起始后或引入核酸分子后6,12,24,48或72小时)本发明所述的人工核酸分子的蛋白表达与从参比核酸分子的蛋白表达相比,两种核酸分子分开引入到靶组织/细胞中,在特定时间点后收集组织/细胞的样品,根据对特定检测方法(例如蛋白质印迹,ELISA,荧光或发光测量等,如本领域中已知的)调整的特定方案制备蛋白裂解产物,并且通过选择的检测方法检测蛋白。作为细胞裂解产物中表达的蛋白量的测量的备选方案——或者,除了在裂解收集的细胞之前测量细胞裂解产物中的蛋白量或平行使用等分部分——蛋白量也可以通过使用FACS分析确定。

[0127] 除非上下文另外指明,其他通用方面对应于上文关于蛋白产生所述的方面。

[0128] 优选地,所述至少一个3'-非翻译区元件(3'-UTR元件)和/或所述至少一个5'-非翻译区元件(5'-UTR元件)为本发明的人工核酸分子提供高翻译效率。“提供”意指,在缺少所述至少一个3'-非翻译区元件(3'-UTR元件)和/或所述至少一个5'-非翻译区元件(5'-UTR元件)时,翻译效率不高,如本文定义。

[0129] 术语“高翻译效率”可以用于比较本发明的人工核酸分子的翻译效率与参比核酸分子的翻译效率。术语“参比核酸分子”用在本文中意指——除了不同的3'-UTR和/或5'-UTR——所述参比核酸分子与本发明包含所述3'-UTR元件和/或所述5'-UTR元件的人工核酸分子是相当的,优选是相同的。由此,人工核酸分子的翻译效率可以与参比核酸分子的翻译效率进行比较。本文提供了用于比较翻译效率的各种方法。

[0130] 除非上下文另外指明,上文关于确定蛋白产生所述的所有方面也适用于比较翻译效率的方法。由于用于比较翻译效率的方法涉及比较两种核酸分子,进行两个平行实验,这些实验间唯一的区别是所述核酸分子的性质。换言之,在各实验间所有其他条件,诸如例如,生长培养基、温度、pH,都是相同的。基于常识和本文提供的指导,技术人员常规选择适当的条件,要求是为两个平行实验选择相同的条件。通常,平行实验中的一个涉及参比核酸分子(其可以是野生型的或人工的),而另一个涉及人工核酸分子(也称为受试核酸分子)。该方法允许比较受试核酸分子是否以高翻译效率为特征;即,其是否是本发明所述的人工核酸分子。用于比较翻译效率的方法的具体方面如下:

[0131] 在该方法中,所述参比核酸分子包含至少一个可读框(ORF),其与人工核酸分子的至少一个可读框(ORF)相同;并且所述参比核酸分子不包含人工核酸分子的至少一个3'-非翻译区元件(3'-UTR元件)和/或人工核酸分子的至少一个5'-非翻译区元件(5'-UTR元件)。通过包括下述步骤的方法比较所述人工核酸分子的翻译效率与所述参比核酸分子的翻译效率:

[0132] (i) 用所述人工核酸分子转染哺乳动物细胞,并且在转染后特定时间点(例如,24或48小时)测量所述人工核酸分子的ORF编码的蛋白的表达量,

[0133] (ii) 用所述参比核酸分子转染哺乳动物细胞,并且在转染后相同时间点测量所述参比核酸分子的ORF编码的蛋白的表达量,

[0134] (iii) 计算从所述人工核酸分子表达的蛋白量与从所述参比核酸分子表达的蛋白量的比率,

[0135] 其中在(iii)中计算的比率 $\geq 1$ ,优选 $>1$ 。

[0136] 由此,通过比较从本发明所述的受试核酸分子产生的相对蛋白量与从所述参比核酸分子产生的相对蛋白量,可以确定本发明所述的人工核酸分子的翻译效率较所述参比核酸分子的翻译效率增加的倍数。

[0137] “转染后”用于表示已经用所述受试核酸分子和所述参比核酸分子转染了细胞的时间点。

[0138] 优选地,与从分别包含参比3'-和/或5'-UTR的参比核酸分子的蛋白产生相比,在本发明所述的人工核酸分子中的至少一个3'-和/或5'-UTR元件使所述人工核酸分子的翻译效率增加至少1.2倍,优选至少1.5倍,更优选至少2倍,甚至更优选至少2.5倍(即,在(iii)中计算的比率至少为1.2,优选至少为1.5,更优选至少为2,甚至更优选至少为2.5)。换言之,对于在相同的时间点,与参比核酸分子的翻译效率相比,所述翻译效率增加的倍数至少为1.2,优选至少为1.5,更优选至少2,甚至更优选至少为2.5。

[0139] 所述受试核酸分子与所述参比核酸分子不同的5'-UTR或3'-UTR赋予了所述高翻译效率。

[0140] 在本发明所述的比较翻译效率的方法的优选实施方案中,(i)所述参比核酸分子包含所述人工核酸分子中不存在的3'-UTR元件,或(ii)所述参比核酸分子包含所述人工核酸分子中不存在的5'-UTR元件。在更优选的实施方案智能光,(i)所述参比核酸分子包含来源于白蛋白基因的3'-UTR(优选来源于对应SEQ ID NO:207所述的序列的3'-UTR)的3'-UTR元件,或(ii)所述参比核酸分子包含来源于TOP基因的5'-UTR(优选缺少5'端寡嘧啶束的人大核糖体蛋白32的5'-UTR、优选对应SEQ ID NO:208所述的序列的5'-UTR)的5'-UTR元件。

[0141] 在备选的优选实施方案中, (i) 所述参比核酸分子不包含任意的3' -非翻译区元件(3' -UTR元件), 或(ii) 所述参比核酸分子不包含任意的5' -非翻译区元件(5' -UTR元件)。

[0142] 在优选实施方案中, 用于转染所述人工核酸分子和所述参比核酸分子的步骤中的细胞是哺乳动物细胞, 优选选自HDF、L929、HEP2G和HeLa细胞的组。可以进一步优选地, 在多于一种类型的哺乳动物细胞中平行进行所述方法, 诸如在任意两种、优选任意三种、更优选全部四种上述细胞系的组合中平行进行所述方法。当在多于一种类型的哺乳动物细胞中平行进行所述方法时, 则当在至少一个所述的哺乳动物细胞类型中在步骤(iii) 计算的比率 $\geq 1$ , 优选 $>1$ 时, 可以认为受试核酸分子具有高翻译效率的特征。然而, 优选地, 当在至少两种类型、如至少三种类型、如四种类型的所用的哺乳动物细胞中在步骤(iii) 中计算的比率 $\geq 1$ , 优选 $>1$ 时, 认为受试核酸分子具有高翻译效率的特征。

[0143] 在转染后特定时间点(例如, 12h, 24h, 36h, 48h或72h) 测量所述人工核酸分子和所述参比核酸分子的ORF编码的蛋白的表达量。优选地, 在转染后24小时测量ORF编码的蛋白的表达量。

[0144] 本发明还提供可以用于本发明的方法的特别优选的参比构建体。尤其优选所述参比核酸分子是SEQ ID NO:205的mRNA(图1A)。在该情形中, 进一步优选所述参比核酸分子与所述人工核酸分子的不同之处仅在于至少一个3' -UTR元件被至少一个另外的3' -UTR元件(优选一个3' -UTR元件) 替换, 或至少一个5' -UTR元件被至少一个另外的5' -UTR元件(优选一个5' -UTR元件) 替换。在每种情形中, 这允许直接检测所述另外的3' -UTR元件或另外的5' -UTR元件是否分别提供高翻译效率。

[0145] 特别优选地, 确定翻译效率的方法进一步具有下述特征:

[0146] (a) 在转染步骤, 所述细胞为选自HDF、L929、HEP2G和HeLa细胞的组的哺乳动物细胞; 和

[0147] (b) 测量的时间点是转染后24小时; 和

[0148] (c) 所述参比核酸分子是SEQ ID NO:205的mRNA(图1A); 和

[0149] 优选地, (d) 所述参比核酸分子与所述人工核酸分子不同仅在于(i) 至少一个3' -UTR元件(优选一个3' -UTR元件) 被至少一个另外的3' -UTR元件(优选一个3' -UTR元件) 替换, 或(ii) 至少一个5' -UTR元件(优选一个5' -UTR元件) 被至少一个另外的5' -UTR元件(优选一个5' -UTR元件) 替换; 换言之, 除了至少一个5' -UTR元件与图1A所示的5' -UTR不同, 或至少一个3' -UTR元件与图1A所示的3' -UTR不同之外, 所述受试核酸分子与所述参比核酸分子相同。图1B显示了5' -UTR元件替换的一个实例; 图1C显示了3' -UTR元件替换的一个实例。

[0150] 当(a) 至(d) 全实现时, 所述方法也称为“标准的确定的翻译效率的测定”, 或简称“确定翻译效率的测定”。特别优选地, 本发明的人工核酸分子的特征在于通过该标准的确定的翻译效率的测定确定的高翻译效率。该标准测定允许直接鉴定特别有用的人工核酸分子。

[0151] 在本发明实施例3中使用标准测定。由此, 本发明人发现一组新型人工核酸共有高翻译效率的有利特征, 即, 其翻译效率高于图1A的参比核酸分子的翻译效率, 参见实施例3.5。这是非常令人惊讶和有利的发现, 特别是根据背景, 所述参比核酸分子包含来源于白蛋白基因的3' -UTR(对应于SEQ ID NO:207所述的DNA序列的3' -UTR) 的3' -UTR元件和来源于TOP基因的5' -UTR(特别是缺少5' 端寡嘧啶束的人大核糖体蛋白32的5' -UTR(对应于SEQ ID NO:208所述的序列的5' -UTR)) 的5' -UTR元件。已知此类3' -UTR元件和5' -UTR元件在细

胞中蛋白翻译的情形中是有利的(对于来源于白蛋白基因的3'-UTR,参见W02013/143700;对于5'-TOP-UTR,参见W02013/143700)。本发明提供了进一步的改善。

[0152] 在一些实施方案中,人工核酸分子的ORF编码可以被定量的蛋白,优选是报告蛋白。这些实施方案非常适合比较本发明所述的翻译效率的方法。在其他实施方案中,人工核酸分子的ORF不编码可以定量的蛋白,优选不编码报告蛋白。例如,当使用人工核酸分子编码可以定量的蛋白的ORF通过本发明的方法确定了具体的5'-UTR或3'-UTR提供高翻译效率时,所述5'-UTR或3'-UTR可以与任意ORF组合。例如,尽管在标准的确定翻译效率的测定中,ORF编码报告蛋白,但是本发明的人工核酸不限于在标准测定中所用的ORF。

[0153] 这些方法还可以叠加使用;即,已经通过本文的方法确定为具有高翻译效率特征的人工核酸分子本身可以用作本文中下一轮方法的“参照构建体”。这允许检测其他人工核酸分子释放具有甚至更高的翻译效率特征(或者在其他人工核酸分子中包含的UTR是否提供甚至更高的翻译效率)。这允许鉴定甚至进一步改进的人工核酸分子和UTR。

#### [0154] 稳定mRNA

[0155] 在一些实施方案中,本发明所述的人工核酸分子中所述至少一个3'-UTR元件和/或所述至少一个5'-UTR元件源自稳定mRNA。由此,“源自”稳定mRNA意为,所述至少一个3'-UTR元件和/或所述至少一个5'-UTR元件与稳定mRNA的3'-UTR元件和/或5'-UTR元件具有至少50%,优选至少60%,优选至少70%,更优选至少75%,更优选至少80%,更优选至少85%,甚至更优选至少90%,甚至更优选至少95%,并且特别优选至少98%的序列同一性。优选地,稳定mRNA是天然存在的mRNA,并且,由此,稳定mRNA的3'-UTR元件和/或5'-UTR元件是指天然存在的mRNA的3'-UTR和/或5'-UTR,或其片段或变体。此外,源自稳定mRNA的3'-UTR元件和/或5'-UTR元件还优选指与天然存在的3'-UTR元件和/或5'-UTR元件相比是修饰的,例如以增加RNA稳定性,甚至进一步和/或延长和/或增加蛋白产生的3'-UTR元件和/或5'-UTR元件。其并不是说,所述修饰是优选的,例如与天然存在的(未修饰的)3'-UTR元件和/或5'-UTR元件相比,其不破坏RNA稳定性。尤其是,如本文中使用的术语mRNA是指mRNA分子,然而,其还可以指如本文所定义的mRNA种类。

[0156] 优选地,mRNA的稳定性,即mRNA降解和/或半衰期在标准条件下,例如用于使用的某细胞系的标准条件(标准培养基,孵育等)评估。

[0157] 如本文中使用的术语“稳定mRNA”通常是指具有缓慢mRNA降解的mRNA。由此,“稳定mRNA”通常具有长的半衰期。mRNA的半衰期是降解50%的体内或体外存在的mRNA分子所需的时间。因此,mRNA的稳定性通常在体内或体外评估。由此,体外尤其是指(“活”)细胞和/或组织,包括活受试者的组织。细胞尤其包括细胞系、原代细胞、组织或受试者中的细胞。在具体实施方案中,允许细胞培养的细胞类型可以适于本发明。特别优选的是哺乳动物细胞,例如人细胞和小鼠细胞。在特别优选的实施方案中,使用人细胞系HeLa、HEPG2和U-937以及小鼠细胞系NIH3T3、JAWSII和L929。此外,原代细胞是特别优选的,在特别优选的实施方案中,可以使用人皮肤成纤维细胞(HDF)。备选地,还可以使用受试者的组织。

[0158] 优选地,“稳定mRNA”的半衰期是至少5小时,至少6小时,至少7小时,至少8小时,至少9小时,至少10小时,至少11小时,至少12小时,至少13小时,至少14小时和/或至少15小时。研究的mRNA的半衰期可以通过本领域技术人员已知的不同方法确定。通常,研究的mRNA的半衰期通过确定降解常数确定,由此推测通常理想的体内(或如上文定义体外)状态,其



中研究的mRNA的转录可以完全“关闭”(或至少调至不可检测水平)。在该理想状态下,通常推测,mRNA降解按照一阶(first-order)动力学。因此,mRNA的降解通常可以通过以下等式描述:

$$[0159] \quad A(t) = A_0 * e^{-\lambda t}$$

[0160]  $A_0$ 是研究的mRNA在时间0,即降解开始前的量(或浓度), $A(t)$ 是研究的mRNA在降解期间时间 $t$ 的量(或浓度),并且 $\lambda$ 是降解常数。由此,如果研究的mRNA在0的量(或浓度)( $A_0$ )和研究的mRNA在降解过程期间某时间 $t$ 的量(或浓度)( $A(t)$ 和 $t$ )已知,可以计算降解常数 $\lambda$ 。基于降解常数 $\lambda$ ,可以通过以下等式计算半衰期 $t_{1/2}$ :

$$[0161] \quad t_{1/2} = \ln 2 / \lambda。$$

[0162] 因为按照定义在 $t_{1/2} A(t) / A_0 = 1/2$ 。由此,为了评估研究的mRNA的半衰期,通常在体内(或如上文定义的体外)RNA降解过程中确定mRNA的量或浓度。

[0163] 为了确定体内(或如上文所定义的体外)RNA降解过程中mRNA的量或浓度,可以使用各种方法,这是技术人员已知的。所述方法的非限制性实例包括例如利用转录抑制剂如放线菌素D的一般性抑制转录,使用诱导型启动子特异启动瞬时转录,例如c-fos血清-诱导型启动子系统和Tet-off调节启动子系统,以及例如通过4-硫尿核苷(4sU)、5-乙炔基尿苷(EU)或5'-溴-尿苷(BrU)的动力学标记技术,例如脉冲标记。关于如何在RNA降解过程中确定mRNA的量或浓度的进一步详述和优选的实施方案在下文,在鉴别本发明所述的3'-UTR元件和/或至少一个5'-UTR元件的方法的情况中概述。如何在RNA降解过程中确定mRNA量或浓度的各个描述和优选的实施方案也应用于此。

[0164] 优选地,与平均mRNA相比,在本发明的意义中的“稳定mRNA”具有较慢的mRNA降解,优选在体内(或如上文所定义的体外)评估。例如,“平均mRNA降解”可以通过研究多种mRNA种类,优选100种,至少300种,至少500种,至少1000种,至少2000种,至少3000种,至少4000种,至少5000种,至少6000种,至少7000种,至少8000种,至少9000种,至少10000种,至少11000种,至少12000种,至少13000种,至少14000种,至少15000种,至少16000种,至少17000种,至少18000种,至少19000种,至少20000种,至少21000种,至少22000种,至少23000种,至少24000种,至少25000种,至少26000种,至少27000种,至少28000种,至少29000种,至少30000种mRNA种类的mRNA降解来评估。特别优选的是,评估整个转录组,或转录组的尽可能多的mRNA种类。例如,这可以通过使用提供整个转录本覆盖的微阵列实现。

[0165] 如本文中使用的“mRNA种类”对应于基因组转录单元,即通常对应于基因。由此,例如,由于mRNA加工,在一个“mRNA种类”中可以存在不同转录本。例如,mRNA种类可以由微阵列上的点表示。因此,微阵列提供有利的工具以确定例如在mRNA降解期间某时间点的多种mRNA种类的量。然而,还可以使用技术人员已知的其他技术,例如RNA-seq,定量PCR等。

[0166] 在本发明中,特别优选的是,稳定mRNA特征为其中所述mRNA在第二时间点的量与所述mRNA在第一时间点的量的比是至少0.5(50%),至少0.6(60%),至少0.7(70%),至少0.75(75%),至少0.8(80%),至少0.85(85%),至少0.9(90%),或至少0.95(95%)的mRNA降解。由此,在降解过程中所述第二时间点晚于所述第一时间点。

[0167] 优选地,选择这样的第一时间点,从而仅考虑经历降解过程的mRNA,即避开出现的mRNA——例如在进行的转录中。例如,如果使用动力学标记技术,例如脉冲标记,优选选择这样的第一时间点,从而完成将标记引入mRNA,即不发生持续地将标记引入mRNA。由此,如

果使用动力学标记,所述第一时间点可以是实验标记过程结束之后,例如将细胞与标记孵育结束后至少10min,至少20min,至少30min,至少40min,至少50min,至少60min,至少70min,至少80min,或至少90min。

[0168] 例如,所述第一时间点可以优选是在诱导型启动子的情况下终止转录(例如通过转录抑制剂),终止启动子诱导后或在终止脉冲或标记供应后,例如标记结束后0至6h。更优选地,所述第一时间点可以是在诱导型启动子的情况下终止转录(例如通过转录抑制剂),终止启动子诱导后或在终止脉冲或标记供应后,例如标记结束后30min至5h,甚至更优选1h至4h,并且特别优选约3h。

[0169] 优选地,在mRNA降解过程中尽可能晚地选择第二时间点。然而,如果考虑多种mRNA种类,优选选择这样的第二时间点,从而相当多的量的多种mRNA种类,优选至少10%的mRNA种类,仍然以可检测的量,即以高于0的量存在。优选地,所述第二时间点是转录结束或实验标记过程结束后至少5h,至少6h,至少7h,至少8h,至少9h,至少10h,至少11h,至少12h,至少13h,至少14h,或至少15h。

[0170] 由此,所述第一时间点和所述第二时间点之间的时间间隔优选在上述界限内尽可能大。因此,所述第一时间点和所述第二时间点之间的时间间隔优选是至少4h,至少5h,至少6h,至少7h,至少8h,至少9h,至少10h,至少11h,或至少12h。

[0171] 此外,可能的是,本发明所述的人工核酸分子的所述至少一个3'-UTR元件和/或所述至少一个5'-UTR元件通过如本文中所述的用于鉴别本发明所述的3'-UTR元件和/或5'-UTR元件的方法鉴别。特别优选的是,本发明所述的人工核酸分子中的所述至少一个3'-UTR元件和/或所述至少一个5'-UTR元件通过如本文中所述的用于鉴别为人工核酸分子提供高翻译效率的3'-UTR元件和/或5'-UTR元件的方法鉴别。

[0172] 本发明的人工核酸分子的优选的实施方案

[0173] 优选地,本发明所述的人工核酸分子中的所述至少一个3'-UTR元件和/或至少一个5'-UTR元件包含下述核酸序列或由下述核酸序列组成:所述核酸序列源自真核蛋白编码基因的3'-UTR和/或5'-UTR,优选源自脊椎动物蛋白编码基因的3'-UTR和/或5'-UTR,更优选源自哺乳动物蛋白编码基因(例如源自小鼠和人蛋白编码基因)的3'-UTR和/或5'-UTR,甚至更优选源自灵长类或啮齿类蛋白编码基因的3'-UTR和/或5'-UTR,尤其是人或鼠蛋白编码基因的3'-UTR和/或5'-UTR。

[0174] 通常,要理解的是,本发明所述的人工核酸分子中的至少一个3'-UTR元件包含或由以下核酸序列组成:优选源自天然(天然状态)存在的3'-UTR的核酸序列,而本发明所述的人工核酸分子中的至少一个5'-UTR元件包含或由以下核酸序列组成:优选源自天然(天然状态)存在的5'-UTR的核酸序列。

[0175] 优选地,所述至少一个可读框与至少一个3'-UTR元件和/或与至少一个5'-UTR元件异源。在该情况下,术语“异源的”是指,人工核酸分子包含的两个序列元件,如可读框和3'-UTR元件和/或可读框和5'-UTR元件不天然地(天然状态)以此组合存在。它们通常是重组的。优选地,3'-UTR元件和/或5'-UTR元件源自与可读框不同的基因。例如,ORF可以源自与3'-UTR元件和/或与至少一个5'-UTR元件不同的基因,例如编码不同的蛋白或相同但属于不同种类的蛋白等。即,可读框源自与3'-UTR元件源自的和/或至少一个5'-UTR元件源自的基因不同的基因。在优选的实施方案中,ORF不编码人或植物(例如,拟南芥

(*Arabidopsis*)) 核糖体蛋白, 优选不编码人核糖体蛋白S6 (RPS6)、人核糖体蛋白L36a-样 (RPL36AL) 或拟南芥核糖体蛋白S16 (RPS16)。在进一步优选的实施方案中, 可读框 (ORF) 不编码核糖体蛋白S6 (RPS6)、核糖体蛋白L36a-样 (RPL36AL) 或核糖体蛋白S16 (RPS16)。

[0176] 在具体实施方案中, 优选的是, 可读框不编码报告蛋白, 例如, 选自由以下各项组成的组的报告蛋白: 球蛋白 (特别是 $\beta$ -球蛋白), 荧光素酶蛋白, GFP蛋白或其变体, 例如, 呈现与球蛋白、荧光素酶蛋白或GFP蛋白至少70%序列同一性的变体。由此, 特别优选的是, 可读框不编码GFP蛋白。还特别优选的是, 可读框 (ORF) 不编码报告基因或不源自报告基因, 其中报告基因优选不选自由以下各项组成的组: 球蛋白 (特别是 $\beta$ -球蛋白), 荧光素酶蛋白,  $\beta$ -葡萄糖苷酸酶 (GUS) 和GFP蛋白或其变体, 优选不选自EGFP, 或上述基因的任一种的变体, 所述变体通常呈现与这些报告基因中的任一种, 优选球蛋白、荧光素酶蛋白或GFP蛋白的至少70%序列同一性。

[0177] 甚至更优选地, 3' -UTR元件和/或5' -UTR元件与包含在本文中定义的人工核酸中的任意其他元件异源。例如, 如果本发明所述的人工核酸包含来自给定基因的3' -UTR元件, 其优选不包含任意其他核酸序列, 尤其不包含来自相同基因的功能性核酸序列 (例如编码或调节序列元件), 包括其在所述基因的ORF的5' 和3' 末端的调节序列。因此, 例如, 如果本发明所述的人工核酸包含来自给定基因的5' -UTR元件, 其优选不包含任意其他核酸序列, 尤其包含来自相同基因的功能核酸序列 (例如编码或调节序列元件), 包括其在所述基因的ORF的5' 和3' 末端的调节序列。

[0178] 此外, 优选的是, 本发明所述的人工核酸包含至少一个可读框, 至少一个3' -UTR (元件) 和至少一个5' -UTR (元件), 其中至少一个3' -UTR (元件) 是本发明所述的3' -UTR元件和/或至少一个5' -UTR (元件) 是本发明所述的5' -UTR元件。在所述优选的本发明所述的包含至少一个可读框、至少一个3' -UTR (元件) 和至少一个5' -UTR (元件) 的人工核酸中, 特别优选的是, 至少一个可读框、至少一个3' -UTR (元件) 和至少一个5' -UTR (元件) 中的每个是异源的, 即至少一个3' -UTR (元件) 和至少一个5' -UTR (元件) 以及可读框和3' -UTR (元件) 或5' -UTR (元件) 分别都不天然 (天然状态) 以此组合存在。这意为, 人工核酸分子包含ORF, 3' -UTR (元件) 和5' -UTR (元件), 它们彼此都是异源的, 例如由于它们每个都源自不同基因 (以及它们的5' 和3' UTR's), 它们是重组的。在另一优选的实施方案中, 3' -UTR (元件) 不源自病毒基因的3' -UTR (元件) 或不是病毒来源的。

[0179] 优选地, 至少一个3' -UTR元件和/或至少一个5' -UTR元件与ORF功能性连接。这意为, 优选地, 3' -UTR元件和/或至少一个5' -UTR元件与ORF相关, 从而其可以发挥功能, 如对于编码的肽或蛋白的表达的增强或稳定功能或对于人工核酸分子的稳定功能。优选地, ORF和3' -UTR元件以5'  $\rightarrow$  3' 方向关联和/或5' -UTR元件和ORF以5'  $\rightarrow$  3' 方向关联。由此, 优选地, 人工核酸分子通常包含结构5' -[5' -UTR元件] - (任选的) -接头-ORF - (任选的) -接头-[3' -UTR元件] -3', 其中所述人工核酸分子可以仅包含5' -UTR元件而没有3' -UTR元件, 仅3' -UTR元件而没有5' -UTR元件, 或包含3' -UTR元件和5' -UTR元件二者。此外, 接头可以存在或不存在。例如, 接头可以是一个或多个核苷酸, 如1-50个或1-20个核苷酸的链, 例如, 包含一个或多个限制性酶识别位点 (限制性位点) 或由一个或多个限制性酶识别位点 (限制性位点) 组成。

[0180] 优选地, 所述至少一个3' -UTR元件和/或所述至少一个5' -UTR元件包含或由以下

核酸序列组成:源自选自由以下各项组成的组的基因的转录本的3'-UTR和/或5'-UTR的核酸序列:ZNF460,TGM2,IL7R,BGN,TK1,RAB3B,CBX6,FZD2,COL8A1,NDUFS7,PHGDH,PLK2,TSP0,PTGS1,FBX032,NID2,ATP5D,EXOSC4,NOL9,UBB4B,VPS18,ORMDL2,FSCN1,TMEM33,TUBA4A,EMP3,TMEM201,CRIP2,BRAT1,SERPINH1,CD9,DPYSL2,CDK9,TFRC,PSMB3 5'-UTR,FASN,PSMB6,PRSS56,KPNA6,SFT2D2,PARD6B,LPP,SPARC,SCAND1,VASN,SLC26A1,LCLAT1,FBXL18,SLC35F6,RAB3D,MAP1B,VMA21,CYBA,SEZ6L2,PCOLCE,VTN,ALDH16A1,RAVER1,KPNA6,SERINC5,JUP,CPN2,CRIP2,EPT1,PNP0,SSSCA1,POLR2L,LIN7C,UQCR10,PYCRL,AMN,MAP1S,NDUFS7,PHGDH,TSP0,ATP5D,EXOSC4,TUBB4B,TUBA4A,EMP3,CRIP2,BRAT1,CD9,CDK9,PSMB3,PSMB6,PRSS56,SCAND1,AMN,CYBA,PCOLCE,MAP1S,VTN,ALDH16A1(全部优选是人的)和Dpysl2,Ccnd1,Acox2,Cbx6,Ubc,Ldlr,Nudt22,Pcyox11,Ankrd1,Tmem37,Tspyl4,Slc7a3,Cst6,Aacs,Nosip,Itga7,Ccnd2,Ebp,Sf3b5,Fasn,Hmgcs1,Osr1,Lmnb1,Vma21,Kif20a,Cdca8,Slc7a1,Ubqln2,Prps2,Shmt2,Aurkb,Fignl1,Cad,Anln,Slfn9,Ncaph,Pole,Uhrf1,Gjal,Fam64a,Kif2c,Tspan10,Scand1,Gpr84,Fads3,Cers6,Cxcr4,Gprc5c,Fen1,Cspg4,Mrpl34,Comtd1,Armc6,Emr4,Atp5d,1110001J03Rik,Csf2ra,Aarsd1,Kif22,Cth,Tpgs1,Ccl17,Alkbh7,Ms4a8a,Acox2,Ubc,Slpi,Pcyox11,Igf2bp1,Tmem37,Slc7a3,Cst6,Ebp,Sf3b5,Plk1,Cdca8,Kif22,Cad,Cth,Pole,Kif2c,Scand1,Gpr84,Tpgs1,Ccl17,Alkbh7,Ms4a8a,Mrpl34,Comtd1,Armc6,Atp5d,1110001J03Rik,Nudt22,Aarsd1(全部优选是小鼠的)。

[0181] 优选地,本发明所述的人工核酸分子的至少一个3'-UTR元件和/或至少一个5'-UTR元件包含基因转录本的3'-UTR和/或5'-UTR的“功能片段”、“功能变体”或“变体的功能片段”或由基因转录本的3'-UTR和/或5'-UTR的“功能片段”、“功能变体”或“变体的功能片段”组成。

[0182] 优选地,所述至少一个5'-UTR元件包含选自由下述组成的组的源自基因转录本的5'-UTR的核酸序列:ZNF460-5'-UTR,TGM2-5'-UTR,IL7R-5'-UTR,BGN-5'-UTR,TK1-5'-UTR,RAB3B-5'-UTR,CBX6-5'-UTR,FZD2-5'-UTR,COL8A1-5'-UTR,NDUFS7-5'-UTR,PHGDH-5'-UTR,PLK2-5'-UTR,TSP0-5'-UTR,PTGS1-5'-UTR,FBX032-5'-UTR,NID2-5'-UTR,ATP5D-5'-UTR,EXOSC4-5'-UTR,NOL9-5'-UTR,UBB4B-5'-UTR,VPS18-5'-UTR,ORMDL2-5'-UTR,FSCN1-5'-UTR,TMEM33-5'-UTR,TUBA4A-5'-UTR,EMP3-5'-UTR,TMEM201-5'-UTR,CRIP2-5'-UTR,BRAT1-5'-UTR,SERPINH1-5'-UTR,CD9-5'-UTR,DPYSL2-5'-UTR,CDK9-5'-UTR,TFRC-5'-UTR,PSMB3 5'-UTR,FASN-5'-UTR,PSMB6-5'-UTR,PRSS56-5'-UTR,KPNA6-5'-UTR,SFT2D2-5'-UTR,PARD6B-5'-UTR,LPP-5'-UTR,SPARC-5'-UTR,SCAND1-5'-UTR,VASN-5'-UTR,SLC26A1-5'-UTR,LCLAT1-5'-UTR,FBXL18-5'-UTR,SLC35F6-5'-UTR,RAB3D-5'-UTR,MAP1B-5'-UTR,VMA21-5'-UTR,CYBA-5'-UTR,SEZ6L2-5'-UTR,PCOLCE-5'-UTR,VTN-5'-UTR,ALDH16A1-5'-UTR,RAVER1-5'-UTR,KPNA6-5'-UTR,SERINC5-5'-UTR,JUP-5'-UTR,CPN2-5'-UTR,CRIP2-5'-UTR,EPT1-5'-UTR,PNP0-5'-UTR,SSSCA1-5'-UTR,POLR2L-5'-UTR,LIN7C-5'-UTR,UQCR10-5'-UTR,PYCRL-5'-UTR,AMN-5'-UTR,MAP1S-5'-UTR,(全部优选是人的)和Dpysl2-5'-UTR,Ccnd1-5'-UTR,Acox2-5'-UTR,Cbx6-5'-UTR,Ubc-5'-UTR,Ldlr-5'-UTR,Nudt22-5'-UTR,Pcyox11-5'-UTR,Ankrd1-5'-UTR,Tmem37-5'-UTR,Tspyl4-5'-UTR,Slc7a3-5'-UTR,Cst6-5'-UTR,Aacs-5'-UTR,Nosip-5'-UTR,Itga7-5'-UTR,Ccnd2-5'-UTR,

Ebp-5'-UTR, Sf3b5-5'-UTR, Fasn-5'-UTR, Hmgcs1-5'-UTR, Osr1-5'-UTR, Lmnb1-5'-UTR, Vma21-5'-UTR, Kif20a-5'-UTR, Cdca8-5'-UTR, Slc7a1-5'-UTR, Ubqln2-5'-UTR, Prps2-5'-UTR, Shmt2-5'-UTR, Aurkb-5'-UTR, Fignl1-5'-UTR, Cad-5'-UTR, Anln-5'-UTR, Slfn9-5'-UTR, Ncaph-5'-UTR, Pole-5'-UTR, Uhrf1-5'-UTR, Gja1-5'-UTR, Fam64a-5'-UTR, Kif2c-5'-UTR, Tspan10-5'-UTR, Scand1-5'-UTR, Gpr84-5'-UTR, Fads3-5'-UTR, Cers6-5'-UTR, Cxcr4-5'-UTR, Gprc5c-5'-UTR, Fen1-5'-UTR, Cspg4-5'-UTR, Mrpl34-5'-UTR, Comtd1-5'-UTR, Armc6-5'-UTR, Emr4-5'-UTR, Atp5d-5'-UTR, 1110001J03Rik-5'-UTR, Csf2ra-5'-UTR, Aarsd1-5'-UTR, Kif22-5'-UTR, Cth-5'-UTR, Tpgs1-5'-UTR, Cc117-5'-UTR, Alkbh7-5'-UTR, Ms4a8a-5'-UTR (全部优选是小鼠的)。

[0183] 优选地, 所述至少一个3'-UTR元件包含选自由下述组成的组的源自基因转录本的3'-UTR的核酸序列或由选自由下述组成的组的源自基因转录本的3'-UTR的核酸序列组成: NDUFS7-3'-UTR, PHGDH-3'-UTR, TSPO-3'-UTR, ATP5D-3'-UTR, EXOSC4-3'-UTR, TUBB4B-3'-UTR, TUBA4A-3'-UTR, EMP3-3'-UTR, CRIP2-3'-UTR, BRAT1-3'-UTR, CD9-3'-UTR, CDK9-3'-UTR, PSMB3-3'-UTR, PSMB6-3'-UTR, PRSS56-3'-UTR, SCAND1-3'-UTR, AMN-3'-UTR, CYBA-3'-UTR, PCOLCE-3'-UTR, MAP1S-3'-UTR, VTN-3'-UTR, ALDH16A1-3'-UTR (全部优选是人的) 和 Acox2-3'-UTR, Ubc-3'-UTR, Slpi-3'-UTR, Pcyox11-3'-UTR, Igf2bp1-3'-UTR, Tmem37-3'-UTR, Slc7a3-3'-UTR, Cst6-3'-UTR, Ebp-3'-UTR, Sf3b5-3'-UTR, Plk1-3'-UTR, Cdca8-3'-UTR, Kif22-3'-UTR, Cad-3'-UTR, Cth-3'-UTR, Pole-3'-UTR, Kif2c-3'-UTR, Scand1-3'-UTR, Gpr84-3'-UTR, Tpgs1-3'-UTR, Cc117-3'-UTR, Alkbh7-3'-UTR, Ms4a8a-3'-UTR, Mrpl34-3'-UTR, Comtd1-3'-UTR, Armc6-3'-UTR, Atp5d-3'-UTR, 1110001J03Rik-3'-UTR, Nudt22-3'-UTR, Aarsd1-3'-UTR (全部优选是小鼠的)。

[0184] 还优选地是, 本发明的人工核酸包含 (i) 至少一个优选的5'-UTR和 (i) 至少一个优选的3'-UTR二者。

[0185] 在特别优选的实施方案中, 所述至少一个5'-UTR元件包含选自由下述组成的组的源自基因转录本的5'-UTR的核酸序列: ZNF460-5'-UTR, TGM2-5'-UTR, IL7R-5'-UTR, COL8A1-5'-UTR, NDUFS7-5'-UTR, PLK2-5'-UTR, FBX032-5'-UTR, ATP5D-5'-UTR, TUBB4B-5'-UTR, ORMDL2-5'-UTR, FSCN1-5'-UTR, CD9-5'-UTR, PYSL2-5'-UTR, PSMB3-5'-UTR, PSMB6-5'-UTR, KPNA6-5'-UTR, SFT2D2-5'-UTR, LCLAT1-5'-UTR, FBXL18-5'-UTR, SLC35F6-5'-UTR, VMA21-5'-UTR, SEZ6L2-5'-UTR, PCOLCE-5'-UTR, VTN-5'-UTR, ALDH16A1-5'-UTR, KPNA6-5'-UTR, JUP-5'-UTR, CPN2-5'-UTR, PNPO-5'-UTR, SSSCA1-5'-UTR, POLR2L-5'-UTR, LIN7C-5'-UTR, UQCR10-5'-UTR, PYCRL-5'-UTR, AMN-5'-UTR, MAP1S-5'-UTR (全部是人的), Dpysl2-5'-UTR, Acox2-5'-UTR, Ubc-5'-UTR, Nudt22-5'-UTR, Pcyox11-5'-UTR, Ankrd1-5'-UTR, Tspyl4-5'-UTR, Slc7a3-5'-UTR, Aacs-5'-UTR, Nosip-5'-UTR, Itga7-5'-UTR, Ccnd2-5'-UTR, Ebp-5'-UTR, Sf3b5-5'-UTR, Fasn-5'-UTR, Hmgcs1-5'-UTR, Osr1-5'-UTR, Lmnb1-5'-UTR, Vma21-5'-UTR, Kif20a-5'-UTR, Cdca8-5'-UTR, Slc7a1-5'-UTR, Ubqln2-5'-UTR, Prps2-5'-UTR, Shmt2-5'-UTR, Fignl1-5'-UTR, Cad-5'-UTR, Anln-5'-UTR, Slfn9-5'-UTR, Ncaph-5'-UTR, Pole-5'-UTR, Uhrf1-5'-UTR, Gja1-5'-UTR, Fam64a-5'-UTR, Tspan10-5'-UTR, Scand1-5'-UTR, Gpr84-5'-UTR, Cers6-5'-UTR, Cxcr4-5'-UTR, Gprc5c-5'-UTR, Fen1-5'-UTR, Cspg4-5'-UTR, Mrpl34-5'-UTR, Comtd1-5'-UTR, Armc6-5'-UTR, Emr4-5'-UTR,

Atp5d-5'-UTR, Csf2ra-5'-UTR, Aarsd1-5'-UTR, Cth-5'-UTR, Tpgs1-5'-UTR, Cc117-5'-UTR, Alkbh7-5'-UTR, Ms4a8a-5'-UTR (全部是小鼠的)。表明此类UTR元件有助于高翻译效率。

[0186] 在特别优选的实施方案中,所述至少一个3'-UTR元件包含选自由下述组成的组的源自基因转录本的3'-UTR的核酸序列:NDUFS7-3'-UTR, PHGDH-3'-UTR, TSP0-3'-UTR, ATP5D-3'-UTR, EXOSC4-3'-UTR, TUBB4B-3'-UTR, TUBA4A-3'-UTR, EMP3-3'-UTR, CRIP2-3'-UTR, BRAT1-3'-UTR, CD9-3'-UTR, CDK9-3'-UTR, PSMB3-3'-UTR, PSMB6-3'-UTR, PRSS56-3'-UTR, SCAND1-3'-UTR, AMN-3'-UTR, CYBA-3'-UTR, PCOLCE-3'-UTR, MAP1S-3'-UTR, VTN-3'-UTR, ALDH16A1-3'-UTR (全部优选是人的) 和Acox2-3'-UTR, Ubc-3'-UTR, Slpi-3'-UTR, Pcyox11-3'-UTR, Igf2bp1-3'-UTR, Tmem37-3'-UTR, Slc7a3-3'-UTR, Cst6-3'-UTR, Ebp-3'-UTR, Sf3b5-3'-UTR, Plk1-3'-UTR, Cdca8-3'-UTR, Kif22-3'-UTR, Cad-3'-UTR, Cth-3'-UTR, Pole-3'-UTR, Kif2c-3'-UTR, Scand1-3'-UTR, Gpr84-3'-UTR, Tpgs1-3'-UTR, Cc117-3'-UTR, Alkbh7-3'-UTR, Ms4a8a-3'-UTR, Mrpl34-3'-UTR, Comtd1-3'-UTR, Armc6-3'-UTR, Atp5d-3'-UTR, 1110001J03Rik-3'-UTR, Nudt22-3'-UTR, Aarsd1-3'-UTR, (全部优选是小鼠的)。表明此类UTR元件有助于高翻译效率。

[0187] 特别优选的实施方案是可组合的,以使本发明的人工核酸包含(i)至少一个特别优选的5'-UTR和(i)至少一个特别优选的3'-UTR二者。

[0188] 短语“源自基因的转录本的3'-UTR和/或5'-UTR的核酸序列”优选是指基于基因的转录本或其片段或部分(优选天然存在的基因或其片段或部分)的3'-UTR序列和/或5'-UTR序列的核酸序列。在该情形中,术语天然存在的与术语野生型同义使用。该短语包括对应于整个3'-UTR序列和/或整个5'-UTR序列的序列,即基因的转录本的全长3'-UTR和/或5'-UTR序列,和对应于基因的转录本的3'-UTR序列和/或5'-UTR序列的片段的序列。优选地,基因的转录本的3'-UTR和/或5'-UTR的片段由与基因的转录本的全长3'-UTR和/或5'-UTR中的一段连续核苷酸对应的一段连续核苷酸组成,其代表基因的转录本的全长3'-UTR和/或5'-UTR的至少5%,10%,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%,甚至更优选至少80%,并且最优选至少90%。在本发明的意义中,所述片段优选是本文所述的功能性片段。优选地,片段保留对与3'-UTR和/或5'-UTR或其片段连接的ORF的翻译的调节功能。

[0189] 在基因的转录本的3'-UTR和/或5'-UTR的范畴中术语“基因的转录本的3'-UTR的变体和/或5'-UTR的变体”和“其变体”是指天然存在的基因的转录本的3'-UTR和/或5'-UTR的变体,优选是指脊椎动物基因的转录本的3'-UTR和/或5'-UTR的变体,更优选是指哺乳动物基因的转录本的3'-UTR和/或5'-UTR的变体,甚至更优选是指灵长类基因,尤其是如上文所述的人基因的转录本的3'-UTR和/或5'-UTR的变体。所述变体可以是基因的转录本的修饰的3'-UTR和/或5'-UTR。例如,与变体源自的天然存在的3'-UTR和/或5'-UTR相比,3'-UTR的变体和/或5'-UTR的变体可以呈现一个或多个核苷酸缺失、插入、添加和/或置换。优选地,基因的转录本的3'-UTR的变体和/或5'-UTR的变体与变体源自的天然存在的3'-UTR和/或5'-UTR至少40%,优选至少50%,更优选至少60%,更优选至少70%,甚至更优选至少80%,甚至更优选至少90%,最优选至少95%相同。优选地,变体是本文中所述的功能性变体。

[0190] 短语“源自基因的转录本的3'-UTR的变体和/或5'-UTR的变体的核酸序列”优选是指基于如上文所述的基因的转录本的3'-UTR序列和/或5'-UTR的变体或其片段或部分的核酸序列。该短语包括对应于基因的转录本的3'-UTR和/或5'-UTR的变体的整个序列的序列，即基因的转录本的全长变体3'-UTR序列和/或全长变体5'-UTR序列，和对应于基因的转录本的变体3'-UTR序列的片段和/或变体5'-UTR序列的片段的序列。优选地，基因的转录本的3'-UTR和/或5'-UTR的变体的片段由与基因的转录本的3'-UTR和/或5'-UTR的全长变体中的一段连续的核苷酸对应的一段连续的核苷酸组成，其代表基因的转录本的3'-UTR和/或5'-UTR的全长变体的至少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%，甚至更优选至少80%，并且最优选至少90%。在本发明的意义中，所述变体的片段优选是本文中所述的变体的功能性片段。

[0191] 在本发明的情况下，术语“功能性变体”、“功能性片段”和“变体的功能性片段”（还称为“功能性变体片段”）意为基因的转录本的3'-UTR和/或5'-UTR的片段、3'-UTR和/或5'-UTR的变体或3'-UTR和/或5'-UTR的变体的片段满足所述变体、片段、或变体的片段源自的基因的转录本天然存在的3'-UTR和/或5'-UTR的至少一种功能，优选多于一种功能。所述功能可以，例如，稳定mRNA和/或增强、稳定和/或延长从mRNA的蛋白产生和/或增加从mRNA的蛋白表达或总蛋白产生（优选在哺乳动物细胞中，如人细胞中）。优选地，3'-UTR和/或5'-UTR的功能涉及ORF编码的蛋白的翻译。更优选地，功能包括增强与3'-UTR和/或5'-UTR或其片段或变体连接的ORF的翻译效率。特别优选的是，在本发明的情况下，变体、片段和变体片段满足与包含参比3'-UTR和/或参比5'-UTR或缺少3'-UTR和/或5'-UTR的mRNA相比，优选在哺乳动物细胞，如人细胞中稳定mRNA的功能，和/或与包含参比3'-UTR和/或参比5'-UTR或缺少3'-UTR和/或5'-UTR的mRNA相比，优选在哺乳动物细胞中，如在人细胞中，增强、稳定和/或延长从mRNA的蛋白产生的功能，和/或与包含参比3'-UTR和/或参比5'-UTR或缺少3'-UTR和/或5'-UTR的mRNA相比，优选在哺乳动物细胞中，如在人细胞中，增加从mRNA的蛋白产生的功能。参比3'-UTR和/或参比5'-UTR可以是，例如，天然与ORF组合存在的3'-UTR和/或5'-UTR。此外，与变体、片段、或变体片段源自的野生型3'-UTR和/或野生型5'-UTR相比，基因的转录本的3'-UTR和/或5'-UTR的功能性变体、功能性片段或功能性变体片段优选对包含3'-UTR和/或5'-UTR的所述变体、片段或变体片段的mRNA的翻译效率不具有显著减少的效果。在本发明的情况下基因的转录本的3'-UTR和/或5'-UTR的“功能性片段”、“功能性变体”或“变体的功能性片段”的特别优选的功能是通过表达携带如上文所述的功能性片段、功能性变体或变体的功能性片段的mRNA增强、稳定和/或延长蛋白产生。

[0192] 优选地，就变体、片段或变体片段源自的基因的转录本天然存在的3'-UTR和/或5'-UTR所展现的翻译效率而言，功能性变体、功能性片段或功能性变体片段所展现的一种或多种功能的效率，如提供翻译效率，增加至少5%，更优选至少10%，更优选至少20%，更优选至少30%，更优选至少40%，更优选至少50%，更优选至少60%，甚至更优选至少70%，甚至更优选至少80%，最优选至少90%。

[0193] 在本发明的情况下，基因的转录本的3'-UTR和/或5'-UTR的片段或基因的转录本的3'-UTR和/或5'-UTR的变体优选呈现至少约3个核苷酸，优选至少约5个核苷酸，更优选至少约10, 15, 20, 25或30个核苷酸，甚至更优选至少约50个核苷酸，最优选至少约70个核苷酸的长度。优选地，基因的转录本的3'-UTR和/或5'-UTR的所述片段或基因的转录本的3'-UTR

和/或5' -UTR的变体是如上文所述的功能性片段。在优选的实施方案中,基因的转录本或其片段或变体的3' -UTR和/或5' -UTR呈现3至约500个核苷酸,优选5至约150个核苷酸,更优选10至100个核苷酸,甚至更优选15至90个,最优选20至70个核苷酸的长度。通常,5' -UTR元件和/或3' -UTR元件特征为少于500,400,300,200,150个或少于100个核苷酸。

[0194] 5' -UTR元件

[0195] 优选地,至少一个5' -UTR元件包含这样的核酸序列或由这样的核酸序列组成,所述核酸序列与分别选自SEQ ID NO:1-151或相应RNA序列组成的组的核酸序列具有至少约1,2,3,4,5,10,15,20,30或40%,优选至少约50%,优选至少约60%,优选至少约70%,更优选至少约80%,更优选至少约90%,甚至更优选至少约95%,甚至更优选至少约99%的同一性,或其中至少一个5' -UTR元件包含这样的核酸序列的片段或由这样的核酸序列的片段组成,所述核酸序列的片段与分别选自SEQ ID NO:1-151或相应RNA序列组成的组的核酸序列具有至少约40%,优选至少约50%,优选至少约60%,优选至少约70%,更优选至少约80%,更优选至少约90%,甚至更优选至少约95%,甚至更优选至少约99%的同一性。

[0196] 在以下详细显示的序列中,SEQ ID NO:1-136可以认为是野生型5' -UTR序列,SEQ ID NO:136-151可以认为是人工5' -UTR序列。

[0197] SEQ ID NO:1-智人(Homo sapiens)ZNF460 5' -UTR

[0198] NM\_006635.3

GAAACAGTGTGGGGCCTAGAGCGCTGGGTGGGCGCGTTCTGCGGCCTGAGCAGGGACGGGTAGTGA  
AGCGGTTACGCCCTTCTTCGCGTCTTGGCGGGAGCCTGACGCCCCGCTTCTCCCTAACGAGGTG  
[0199] TCCCACCGGCGCCCCGAGGCCTAGGCCTCCGCAGCCGCCCTCCGTCTCCTCAGCCCCGACGCTG  
CGCCTTGGGCCTTGTGCGCATTTTTTTCGGGGGAAACTGAGGCTCGGAGTGCGAAAGTCAGCCGA  
GGTCGCCCCGCCAGGACAGAGAAGGGCTGTGGTCGGCTGATCCGCGGCATTCCCGGG

[0200] SEQ ID NO:2-智人TGM2 5' -UTR

[0201] NM\_004613.2

ATAAGTTAGCGCCGCTCTCCGCCTCGGCAGTGCCAGCCGCCAGTGGTCGCACTTGGAGGGTCTCGC  
[0202] CGCCAGTGGAAGGAGCCACCGCCCCCGCCGACC

[0203] SEQ ID NO:3-智人IL7R 5' -UTR

[0204] NM\_002185.3

ATCTAAGCTTCTCTGTCTTCCTCCCTCCCTCCCTTCTCTTACTCTCATTCATTTTCATACACACTG  
[0205] GCTCACACATCTACTCTCTCTCTATCTCTCTCAGA

[0206] SEQ ID NO:4-智人BGN 5' -UTR

[0207] NM\_001711.4

CCTTTCTCCCTCCCCGCCCTCTCCCCGCTGTCCCCTCCCCGTCGGCCCGCCTGCCAGCCTTTAG  
[0208] CCTCCCGCCGCGCCTCTGTCTCCCTCTCTCCACAACTGCCAGGAGTGAGTAGCTGCTTTCGG  
TCCGCCGGACACACCGGACAGATAGACGTGCGGACGGCCACCACCCAGCCCGCCAACTAGTCAG  
CCTGCGCCTGGCGCCTCCCTCTCCAGGTCCATCCGCC

[0209] SEQ ID NO:5-智人TK1 5' -UTR

[0210] NM\_003258.4



- [0211] GCGCACGTCCCGGATTCTCTCCACGAGGGGGCGGGCTGCGGCCAAATCTCCCGCCAGGTCAGCGGC  
CGGGCGCTGATTGGCCCCATGGCGGCGGGGCCGGCTCGTGATTGGCCAGCACGCCGTGGTTTAAAG  
CGGTGCGGCGGGGAACCAGGGGCTTACTGCGGGACGGCCTTGAGAGTACTCGGGTTCTGTGAACCTT  
CCCGGAGGCGCA
- [0212] SEQ ID NO:6-智人RAB3B 5' -UTR
- [0213] NM\_002867.3  
AGACTCCGCCCTTGGGCGGGGCCTGGATGCGGCCGGAGCGGAGCAGTGCTGGAGCGGGAGCCTCAG  
CCCTCAGGCGCCACTGTGAGGACCTGACCGGACCAGACCATCCCGCAGCGCCCCGCCCGGCCCCC  
TCCGCGCCCTCCCGACGCCAGGTCCTGCCGTCCCGCCGACCGTCCGGGAGCGAACCCTGCTCCCG  
CACTCGGAGTCCGCG
- [0214] CACTCGGAGTCCGCG
- [0215] SEQ ID NO:7-智人CBX6 5' -UTR
- [0216] NM\_014292.3
- [0217] GAGCGGTGCCGCACCGGCCGCGGGCGCAGGGAGTATTATGGGCTGTGGGTGCCGCTGAGCAAG
- [0218] SEQ ID NO:8-智人FZD2 5' -UTR
- [0219] NM\_001466.3  
GAGGAGAGAGGGCAGCAGCGCGCGGTGTCTCCGGCTGCTCAGTCCGACCGCGCAAGCAAGCGGGC  
AGGCGCACCGCCCCCTCCCCCGCCGGCCTCCCCAACTCTGCGGCCGCGAGTAAAGTTTGCAAAGA  
GGCGCGGGAGGCGGCAGCCGACGAGGAGGCGCGGGGAAGAAGCGCAGTCTCCGGGTGTTGGGGGC  
GGGGGCGGGGGGGGCGCCAAGGAGCCGGGTGGGGGGCGGCGGCCAGC
- [0220] GGGGGCGGGGGGGGCGCCAAGGAGCCGGGTGGGGGGCGGCGGCCAGC
- [0221] SEQ ID NO:9-智人COL8A1 5' -UTR
- [0222] NM\_001850.4  
ATCACAGCCCTTCCCCGATCCTCTCCGTGGGAGCCAGCGAGCCTCTCTCCCTGATCTTACGTGCTC  
AAGGGAGCTCACACGTTACCAACTCACCTTGAAGTCATCTCAAGAACAAAAGACAACCTGAAAGA  
AGCTGTTGTGAAGGCAGAGCAGCATCTGCTGAAGAGACAGAAACCAGCCCCAGAGGTGTCACAGGA  
AGGCACCAGCAAGGACATTGGTCTTTGATTGATTGAGCAGTCCTGTCAAGTATAAATGTG
- [0223] AGGCACCAGCAAGGACATTGGTCTTTGATTGATTGAGCAGTCCTGTCAAGTATAAATGTG
- [0224] SEQ ID NO:10-智人NDUFS7 5' -UTR
- [0225] NM\_024407.4
- [0226] AAGGAGAACGGACCTCAGAGGTTGTCTGAAGGCCGAGGCCAAG
- [0227] SEQ ID NO:11-智人PHGDH 5' -UTR
- [0228] NM\_006623.3  
GCAGGGATTGGAACCTCAGAGCCGCGAGGAGGAGGCGGAGTCGCGGAGAGTTTGAGTATTTCCG  
TCCAATCAAAGGAGACTGTAAGAGGAGGAGGAGGAGGAGATGACTGGGGAGCGGGAGCTGGAGAA  
TACTGCCCAGTTACTCTAGCGCGCCAGGCCGAACCGCAGCTTCTTGGCTTAGGTACTTCTACTCAC  
AGCGGCCGATTCCGAGGCCAACTCCAGCA
- [0229] AGCGGCCGATTCCGAGGCCAACTCCAGCA
- [0230] SEQ ID NO:12-智人PLK2 5' -UTR
- [0231] NM\_006622.3  
TCCGCCCCCTTCCCGCCTCCCGTATATAAGACTTCGCCGAGCGCTCTCACTCGCACAAAGTGGACC  
GGGGTGTGGGTGCTAGTCGGCACCAAGGGTGGCAGGACCACGGCCGGCTCGGACGTGTG  
ACCGCGCCTAGGGGGTGGCAGCGGGCAGTGCGGGGCGGCAAGGCGACC
- [0232] ACCGCGCCTAGGGGGTGGCAGCGGGCAGTGCGGGGCGGCAAGGCGACC

- [0233] SEQ ID NO:13-智人TSP0 5' -UTR  
[0234] NM\_000714.5  
GGCGGCTGGGAGGGGCGGGGCGGATGCGGGGACAGCGGCCTGGCTAACTCCTGCCAGGCAGTGCCC  
[0235] TTCCCGGAGCGTGCCCTCGCCGCTGAGCTCCCCTGAACAGCAGCTGCAGCAGCC  
[0236] SEQ ID NO:14-智人PTGS1 5' -UTR  
[0237] NM\_000962.3  
ATGGGCTGGAGCTCCGGGCAGTGTGCGAGGCGCACGCACAGGAGCCTGCACTCTGCGTCCCGCACC  
[0238] CCAGCAGCCGCGCC  
[0239] SEQ ID NO:15-智人FBX032 5' -UTR  
[0240] NM\_058229.3  
GCTGCGGCCGCTGCGGGGATAAATACTGCGGCAGCTACTGCCGCGCAGCACTCCCGGAGCCTGCA  
[0241] ACGCTTGAGATCCTCTCCGCGCCCGCCACCCCGCAGGGTGCCCCGCGCCGTTCCCGCCGCCCGCC  
GCCCCCGTCGCGGGGCCCTGCACCCCGAGCATCCGCCCCGGGTGGCACGTCCCCGAGCCCACCAGG  
CCGGCCCCGTCTCCCCATCCGTCTAGTCCGCTCGCGGTGCC  
[0242] SEQ ID NO:16-智人NID2 5' -UTR  
[0243] NM\_007361.3  
GCCTTAGAAAAGTTAACGAGAACCAGATGTGGTGGCCACTGCCGAACTTTCTCAGAGCCGGTGATT  
[0244] GGTCCCCAGCCGAGGGCCTCAGCCAATTAGCTTGCTGGGTGGGCTGGAGTCCCGCCCCGCCAGG  
CGCCCCGCGGAGATCCAGGTTGAGGCTGGCGCGGCGCGGAGAGTGGGCTGGAGGCCGGGGCGGGAC  
GCGTTGTGCAGCGGGTAAGCGCACGGCCGAGCGAGC  
[0245] SEQ ID NO:17-智人ATP5D 5' -UTR  
[0246] NM\_001687.4  
CAGACGTCCCTGCGCGTCGTCTCTCCTCGCCCTCCAGGCCGCCCGCGCCGCGCCGGAGTCCGCTGTC  
[0247] CGCCAGCTACCCGCTTCCTGCCGCCCGCCGCTGCC  
[0248] SEQ ID NO:18-智人EXOSC4 5' -UTR  
[0249] NM\_019037.2  
ATGGCGGACCTCCGGAACCGTAGATTCCGGGCGGTTCGGAGCCGCCGGGAGCTGTAGTTCTCCCGC  
[0250] GGCTCAGAGAAGTAGGCAGAGAGCGGACCTGGCGGCCCGGCAGC  
[0251] SEQ ID NO:19-智人NOL9 5' -UTR  
[0252] NM\_024654.4  
GGCTCGGAGGCCCTCTTTTCGGCAGCTGGGAGGGCAGCGGGGCGGACCGACGGGATCTCGCGAGA  
[0253] TTCTCGCGCGGTAGGCCCTGAGGACCCAGC  
[0254] SEQ ID NO:20-智人TUBB4B 5' -UTR  
[0255] NM\_006088.5  
ATATAAGCGTTGGCGGAGCGTCGGTTGTAGCACTCTGCGCGCCCGCTCTTCTGCTGCTGTTGTCT  
[0256] ACTTCCTCCTGCTTCCCCGCCGCCGCCGCCGCCATC  
[0257] SEQ ID NO:21-智人VPS18 5' -UTR  
[0258] NM\_020857.2

- [0259] GCCCGCGTCACGGGGGCGGGAGTCAGCTGAGCTGCCGGGGCGAGGTTGGGATCACCTGGCACCGGC  
TGAAGGGAGCCTGTGATTTTTTTGTAGCGGGGCGGGGAGTAAGGTGCAAGACTGCGCCAGATTCA  
AGGACGAGGGCTGCCGATTATCTCGCTGCATAAGGCAAGAGCAAGAGGATCCTCAGGATTTTAAA  
GAGGAGGCGACGGCTGCAGGTTCCCAGGATCTGTCAGAGGCTGGGGAGTTACAGCTTCCATTCTGG  
GGCGACGGGGACCCCGGGGGGTAGCCCTTTTGTAAATCCCCAGGCCCGGACAAAGAGCCCAGAGG  
CCGGGCACC
- [0260] SEQ ID NO:22-智人ORMDL2 5' -UTR
- [0261] NM\_014182
- [0262] AAGGGGCGAAGGAGAGGCGTTACTTCCTGGAGACTTCAGGTGTGGTAGCCGGCGCCGCGCCATA  
GCCGGACGGGGATCTGAGCTGGCAGG
- [0263] SEQ ID NO:23-智人FSCN1 5' -UTR
- [0264] NM\_003088.3
- [0265] AGCTGGGCTTTGTGGAGCGCTGCGGAGGGTGCGTGCGGGCCGCGGCAGCCGAACAAAGGAGCAGGG  
GCGCCGCGCAGGGACCCGCCACCCACCTCCCGGGGCGCGCAGCGGCCTCTCGTCTACTGCCACC
- [0266] SEQ ID NO:24-智人TMEM33 5' -UTR
- [0267] NM\_018126.2
- [0268] ACGGAAATGAAAGGAGCACTTCCGGGTTTCGGCAATAACCTGGAGCCGGCGGCGTAGGTTGGCTCTT  
TAGGGCTTCACCCCGAAGCTCCACCTTCGCTCCCGTCTTTCTGGAAACACCGCTTTGATCTCGGCG  
GTGCGGGACAGGTACCTCCCGGCTGCTGCGGGTGCCCTGGATCCAGTCGGCTGCACCAGGCGAGCG  
AGACCCTTCCCTGGTGGAGGCTCAGAGTTCCGGCAGGGTGATCCGGCCTGTGTGTGGCGCGAGGC  
AGGGAAGCCGGTACCCGGGTCTTGCCCCAGCGCTGACGTTTTCTCTCCCCTTTCTTCTCTCTTCG  
CGGTTGCGGCGTCGCAGACGCTAGTGTGAGCCCC
- [0269] SEQ ID NO:25-智人TUBA4A5' -UTR
- [0270] NM\_006000.2
- [0271] ATAAGGGCGGTGCGGCACTGCAGCTAGCGCAGTTCTCACTGAGACCTGTCACCCGACTCAACGTG  
AGACGCACCGCCCGGACTCACC
- [0272] SEQ ID NO:26-智人EMP3 5' -UTR
- [0273] NM\_001425.2
- [0274] CGGGAGCAAGAGAGAAGGAGGCCAGACAGTGAGGGCAGGAGGGAGAGAAGAGACGCAGAAGGAGA  
GCGAGCGAGAGAGAAAGGGTTCTGGATTGGAGGGGAGAGCAAGGGAGGGAGGAAGGCGGTGAGAGA  
GGCGGGGGCCTCGGGAGGGTGAAAGGAGGGAGGAGAAGGGCGGGGCACGGAGGCCCGAGCGAGGGA  
CAAGACTCCGACTCCAGCTCTGACTTTTTTCGCGGCTCTCGGCTTCCACTGCAGCC
- [0275] SEQ ID NO:27-智人TMEM201 5' -UTR
- [0276] NM\_001010866.3
- [0277] CTACCCGCCTACCCGCCTACCCCCCTGCCGGCCTGCCGTCCTTCCACGCGGAGAGCC
- [0278] SEQ ID NO:28-智人CRIP2 5' -UTR
- [0279] NM\_001312.3
- [0280] GGACAGCCGGGCAGGCGGGGCTGGGCGCGGGCGGCGGGCGGCCCGGAGGAGAACGGGCGGAGGGCGC  
GGGCCGACCGGGCGCACCGACC

- [0281] SEQ ID NO:29-智人BRAT1 5' -UTR  
[0282] NM\_152743.3  
ACCGGATGCTCGGCATGAACCACTAGGCGCCTGGCGGGGGTGATCTGTCGGAGCGACCGGCTTGGC  
GCCTGCCTGTCCCCAGCCCCTCTCAGCTTGAACCTCCTTCCTTCAAGTCTGGGCCCTCGAGGCTTCC  
[0283] AGAGCGGCCTCCAGGGGTGCAGTCTCAGTTCCCCACGCCAGCCGTCTCCGTCTCCGCCTCCTCCG  
GGCCTGGCAGGTGGCACTGTCCGGAGGCGGAGCCTTGGGCGAGGGGTGGTTGCGGCGGAGGACGCA  
ACCGAGCGGGCCTGCGGCCTCACC  
[0284] SEQ ID NO:30-智人SERPINH1 5' -UTR  
[0285] NM\_001235.3  
AGTAGGACCCAGGGGCCGGGAGGCGCCGGCAGAGGGAGGGGCCGGGGCCGGGGAGGTTTTGAGGG  
AGGTCTTTGGCTTTTTTTGGCGGAGCTGGGGCGCCCTCCGGAAGCGTTTCCAACCTTCCAGAAGTT  
[0286] TCTCGGGACGGGCAGGAGGGGGTGGGGACTGCCATATATAGATCCCGGGAGCAGGGGAGCGGGCTA  
AGAGTAGAATCGTGTGCGGGCTCGAGAGCGAGAGTCACGTCCCGGCGCTAGCCCAGCCCGACCCAG  
GCCCCACGTGGTGCACGCAAACCACTTCCTGGCC  
[0287] SEQ ID NO:31-智人CD9 5' -UTR  
[0288] NM\_001769.3  
CTTTTCCCGGCACATGCGCACCGCAGCGGGTTCGCGCGCCCTAAGGAGTGGCACTTTTTAAAAGTGC  
[0289] AGCCGGAGACCAGCCTACAGCCGCCTGCATCTGTATCCAGCGCCAGGTCCCGCCAGTCCCAGCTGC  
GCGCGCCCCCAGTCCCGCACCCGTTTCGGCCCAGGCTAAGTTAGCCCTCACC  
[0290] SEQ ID NO:32-智人DPYSL2 5' -UTR  
[0291] NM\_001386.5  
ATATCCCAGGATCGCGGCCAATCGCTGCTCGTCTCTCTCGAAGCGGATGGCTTTTGCCTGAGAGGA  
AAGGGAGTGGCTGGCGGCGCATGCGCCACGGTGGCCGACTTGAACCGAGGCTTTTATTGCTGTAGT  
TTATTTCCACCCCTTCCCTCCTGTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTTTT  
[0292] TTTTCCGCCCTAGCTGGGGCTGTGTTGGAGGAGAGGAAGAAAGAGAGACAGAGGATTGCATTCATC  
CGTTACGTTCTTGAAATTTCCCTAATAGCAAGACCAGCGAAGCGGTTGCACCCTTTTCAATCTTGCA  
AAGGAAAAAAACAAAACAAAACAAAAAAACCCAAGTCCCCCTTCCCGGCAGTTTTTGCCTTAAAGC  
TGCCCTCTTGAAATTAATTTTTTCCCAGGAGAGAG  
[0293] SEQ ID NO:33-智人CDK9 5' -UTR  
[0294] NM\_001261.3  
AAGTGGCCGTGGAGGCGGAAGTGGCGCGGCCGCGGAGGGGCCTGGAGTGCGGCGGCGGGGACCC  
[0295] GGAGCAGGAGCGGCGGCAGCAGCGACTGGGGGCGGCGGCGGCGCGTGGAGGCGGCC  
[0296] SEQ ID NO:34-智人SSSCA1 5' -UTR  
[0297] NM\_006396.1  
[0298] CCGGCGGTGACAACGGCAAC  
[0299] SEQ ID NO:35-智人POLR2L 5' -UTR  
[0300] NM\_021128.4  
[0301] AGTCTGGGACGCGCCCGGCC  
[0302] SEQ ID NO:36-智人LIN7C 5' -UTR

- [0303] NM\_018362
- [0304] CTGTGGGTCTGTAGGTTAAGGGAGAAG
- [0305] SEQ ID NO:37-智人UQCR10 5' -UTR
- [0306] NM\_001003684.1
- [0307] GCGGTGGCGCGAGTTGGACTGTGAAGAAAC
- [0308] SEQ ID NO:38-智人TFRC 5' -UTR
- [0309] NM\_001128148.1  
ACGCACAGCCCCCTGGGGGCCGGGGCGGGGCCAGGCTATAAACCGCCGGTTAGGGGCCGCCATC
- [0310] CCCTCAGAGCGTCGGGATATCGGGTGGCGGCTCGGGACGGAGGACGCGCTAGTGTTCTTCTGTGTG  
GCAGTTCAGA
- [0311] SEQ ID NO:39-智人PSMB3 5' -UTR
- [0312] NM\_002795.2  
GAGCGGTTGCGCAGTGAAGGCTAGACCCGGTTTACTGGAATTGCTCTGGCGATCGAGGGGTCCTAG
- [0313] TACACCGCAATC
- [0314] SEQ ID NO:40-智人FASN 5' -UTR
- [0315] NM\_004104.4  
GAGAGACGGCAGCGGCCCCGGCCTCCCTCTCCGCCGCGCTTCAGCCTCCCGCTCCGCCGCGCTCCA
- [0316] GCCTCGCTCTCCGCCGCCCCGACCGCCGCCCGCGCCCTCACCAGAGCAGCC
- [0317] SEQ ID NO:41-智人PSMB6 5' -UTR
- [0318] NM\_002798.2
- [0319] GTGACAGAGCGCTTTACGACAGTTGCTTTGAGGCAGTACCGGAGGAGAAAAG
- [0320] SEQ ID NO:42-智人PYCRL 5' -UTR
- [0321] NM\_023078.3
- [0322] CTAGATCTGTGGGCGGGGCGCGGCCTGTGG
- [0323] SEQ ID NO:43-智人PRSS56 5' -UTR
- [0324] NM\_001195129.1  
GATGATTTGAGGGACAAGAATTCAGTGCCCGGGGGCCGAAAGGCAGCAGAAGGCGGGCACCAAAGG
- [0325] ATAGGCACCCGGAAGGTGGACTCCGAGGAGGAGAGAGGACAGGGGTCTCTACCCCAGCTCCTGGT  
CACC
- [0326] SEQ ID NO:44-智人KPNA6 5' -UTR
- [0327] NM\_012316.4  
CTACAGATCCGCCATATTGTCTACTGAAAGCTGCCGCTGAAGCTGCCGCCGTTGCCTCCGCCGCCA
- [0328] AGAGTGAGCGAGCGGACCCGCG
- [0329] SEQ ID NO:45-智人SFT2D2 5' -UTR
- [0330] NM\_199344.2
- [0331] CCGTCAACTTAGCGAGCGCAACAGGCTGCCGCTGAGGAGCTGGAGCTGGTGGGGACTGGGCCGCA
- [0332] SEQ ID NO:46-智人PARD6B 5' -UTR
- [0333] NM\_032521.2

- [0334] GAGGGAGGGAGCTGCTTCCCCGCTGCCGCGCCACCAGTCCGACCCTCGGTCCCGCCGTGTGAGCA  
GCTGGTGGAGTGGAGCTCAGCGCGGACGCCGAGCTGCGGCCGCCCTCTGCAGGTGCCTGTGAG  
GAGGCGCCCGGGCCGAACCGCTTCCGAGATCCCCAGTCGCGCACTCGCTCCCCGCGCTCCTGAG  
GGGCCGCCCGGCCGAGGAGGCCGTCGCGGGGCTCGGCGTTCAGC
- [0335] SEQ ID NO:47-智人LPP 5' -UTR  
GAGGAAGGAGGGGAAAGGCACCAACCAGCAGCCGCTCCAGCCCTGCCGAAGTTTCACTTTTGTGTC  
TGTGGGTTCGTTCCCGGTCCCGCGCGAGCTTCCCGGGATAAGTAGCTTTAGCGATCGGGAGACA  
GCCGGGCGCTGCAAGTGGGAACCTTGAGGCTCAGAGACAGAGCAGAAGACAGAACCTGGTCTTCTG  
ATTCCCTGTGTTCTGCTTTTTTTCATTGTTCCACTGGACGCTCATCAGAGGGAAGATCTTTTTCTCTC  
AATTGCATTGCAGTTGGCTGAACCTTGTGTCAACCATCCATTCCAGGAGCCAGCTATCTGAGATTCAACA
- [0336] SEQ ID NO:48-智人SPARC 5' -UTR
- [0338] NM\_003118.3  
GGGAGAAGGAGGAGGCCGGGGGAAGGAGGAGACAGGAGGAGGAGGGACCACGGGGTGGAGGGGAGA  
TAGACCCAGCCCAGAGCTCTGAGTGGTTTCCTGTTGCCTGTCTCTAAACCCCTCCACATTCCCGCG  
GTCTTCAGACTGCCCGGAGAGCGCGCTCTGCCTGCCGCTGCCTGCCTGCCACTGAGGGTCCCA  
GCACC
- [0339] SEQ ID NO:49-智人SCAND1 5' -UTR  
CTTCCGGGAGCCCGCAAGCGGCTTCCGGGTGCTCGCGCGCCGACCTGGACGCAGAGAAGCCAGAGA  
CTTTCGTTCCGGCTGCCGAGGCGCTTCGCTGGTGCAGGTAAGCTCCGCACACTCTCGGCCGGTC  
CCGAGTCCGACTCCCTCAAGGGTGACGCGAGCTCTGCCCTTTAACCGGAAACGTCTCCCTGCTCAC  
CCCACCCCGCGCAGACGCAGTGCTGAGCACACAGCTACCGGACAAAGAGTGACGCCCGGAGCTGG  
AGTT
- [0340] SEQ ID NO:50-智人VASN 5' -UTR
- [0341] NM\_138440.2  
GACTCCGGAGCCCGAGCCCGGGCGGGTGGACGCGGACTCGAACGCAGTTGCTTCGGGACCCAGGA  
CCCCCTCGGGCCCGACCCGCCAGGAAAGACTGAGGCCGCGGCCCTGCCCCGCCCGGCTCCCTGCGCC  
GCCGCCGCTCCCGGGACAGAAG
- [0342] SEQ ID NO:51-智人SLC26A1 5' -UTR
- [0343] NM\_022042.3  
ACAGACCACTGCCTGCAGGTTGGCGCCACCACCCCCACTCTCCCCGCTGCTCGCGGGAGCCAGAGG  
GCCCTGCGGTCTCGGTGGTCTTGCCAGCCCCTCGTCATCCAGGGCCCTCCGCGCCTGTGAGGAC  
TCCCTCAGGTCGGCCACGGGACCTGACGCAACAGG
- [0344] SEQ ID NO:52-智人LCLAT1 5' -UTR
- [0345] NM\_182551.3  
GGCCGACGCCTCCGCGTTACGGGATGAATTAACGGCGGGTTCGCGACGGAGGTTGTGACCCCTAC  
GGAGCCCCAGCTTGCCACGCACCCCACTCGGCGTCGCGCGGCGTGCCCTGCTTGTACAGGTGGG  
AGGCTGGAACATCAGGCTGAAAAACAGAGTGGGTACTCTCTTCTGGGAAGCTGGCAACAAATGGA  
TGATGTGATAT
- [0346] SEQ ID NO:52-智人LCLAT1 5' -UTR
- [0347] NM\_182551.3  
GGCCGACGCCTCCGCGTTACGGGATGAATTAACGGCGGGTTCGCGACGGAGGTTGTGACCCCTAC  
GGAGCCCCAGCTTGCCACGCACCCCACTCGGCGTCGCGCGGCGTGCCCTGCTTGTACAGGTGGG  
AGGCTGGAACATCAGGCTGAAAAACAGAGTGGGTACTCTCTTCTGGGAAGCTGGCAACAAATGGA  
TGATGTGATAT

- [0351] SEQ ID NO:53-智人FBXL18 5' -UTR  
[0352] NM\_024963.4  
[0353] GCGGTGGACGCGCCGGCTTCGAGCATCCCTAGCCGGGCAGGTGGGAGGCACGGGGTTGCGGATCCC  
GCGGCCGCGGTTTCGACCCGCCGGCGAC
- [0354] SEQ ID NO:54-智人SLC35F6 5' -UTR  
[0355] NM\_017877.3  
[0356] GGAAGCGCTCGCGCAGGAGACCCCGGGTGACGGGGCCGGCGCCGCTAACTGGAGCGAACCCAGC  
GTCCGCCGAC
- [0357] SEQ ID NO:55-智人RAB3D 5' -UTR  
[0358] NM\_004283.3  
[0359] CCTTCCTCCGCCTTCTGGGCGGAGCCCGCGCGGGATCCGGGTGGCTGCAGGCTGCTGGCTTCTGCG  
GCTGCGGGGTCGGGGTCGCGGCCAGGGCCAAGCCGCAGCGAGTTCACAGGCGGAACCCCTGCAGGC  
GGCGCCCCCTACGCGAGGTACCCCTGGGAAGGAGCGCAGCCACCCGGCCCCCTCCGCATCCGAGC  
AGGACGCCCCGTCTCCTCTCCCTGAGGATTTTCAGGTCTCCCTGTCCCAGGAGGCTTGTGCCAAG
- [0360] SEQ ID NO:56-智人MAP1B 5' -UTR  
[0361] NM\_005909.3  
[0362] GCCTAGTCTCCATATAAAAGCGGCGCCGCCTCCCCGCCCTCTCTCACTCCCCGCTCCTCTCCGCCG  
CGCACTCTCCGCGGCGCTGGGAGAGGGCGGAGGGGGAGGCGGCGCGCGGCCAGAGGAGGGGGGA  
CGCAGGGGGCGGAGCGGAGACAGTACCTTCGGAGATAATCCTTTCTCCTGCCGCGAGTGAGAGGAG  
CGGCCGAGCGAGACACTTCGCCGAGGCACAGCAGCCGGCAGG
- [0363] SEQ ID NO:57-智人VMA21 5' -UTR  
[0364] NM\_001017980.3  
[0365] GCACTTCCGGCGCGAACCCTACTTCCGGTGCGAACCCTCGGCCGTTCCCTCGCGGAGCTTACT  
GAGCGCGGCCGCCGAGCCAGCTCCGCCGCCGAGCGCCTGTGCCGGCACGGCTACACC
- [0366] SEQ ID NO:58-智人AMN 5' -UTR  
[0367] NM\_030943.3  
[0368] GTCTCCTGGTGGGGTGCAAGGAGCCGAGGCGAG
- [0369] SEQ ID NO:59-智人CYBA5' -UTR  
[0370] NM\_000101.3  
[0371] GGCGGGGTTCCGGCCGGGAGCGCAGGGGCGGCAGTGCGCGCCTAGCAGTGTCAGCCGGGTTCGTG  
TCGCC
- [0372] SEQ ID NO:60-智人SEZ6L2 5' -UTR  
[0373] NM\_012410.3  
[0374] ACCACAGAGCCGCGAGAAGAGGACAGAGGAGACTGAGCAAAGGGGGGTGGGCTCCAGGCGACCCCT  
AGCCCAATTCTGCCCCCTCCATCCCAAGGGGCAGAGAAATTGTCTTTCTTTGCTGACTCCTACGAGG  
AAAAAAAAAAAAAAAAAAAAACAATTAAAGGGAAAGATAAACGGAGACGGAGGAAAGGTGGCAGC  
CAGATTACTTAGAGAGGCACAGAGGAGAGAGATCGGGGTGAGTCGCC
- [0375] SEQ ID NO:61-智人PCOLCE 5' -UTR  
[0376] NM\_002593.3

- GACCTAGAGAGGTCCCAGGACACGCCACTGTCCCGCCTTCCCCATTGCCCCGCCCCACTGGCCAGTC
- [0377] CCCACGCCCACACACCCAAGGCTGCCCCATCTGGCGCTGATTATCCTGCTGCTGCCGCCACCGCTG  
CTGCTGCTCTGCAAATTCAGCTGCTGCCTCTGTCTTGAGGACCCCAGCGCCTTTCCCCCGGGGCC
- [0378] SEQ ID NO:62-智人MAP1S 5' -UTR
- [0379] NM\_018174.4
- [0380] GGCCCGAAG
- [0381] SEQ ID NO:63-智人VTN 5' -UTR
- [0382] NM\_000638.3
- GAGCAACAGAGCAGCAGAAAAGGCAGTTCCTCTTCTCCAGTGCCCTCCTTCCCTGTCTCTGCCTC
- [0383] TCCCTCCCTTCCCTCAGGCATCAGAGCGGAGACTTCAGGGAGACCAGAGCCCAGCTTGCCAGGCACT  
GAGCTAGAAGCCCTGCC
- [0384] SEQ ID NO:64-智人ALDH16A1 5' -UTR
- [0385] NM\_001145396.1
- ATTCCCATTAGCCCCGCCCCCTTTGGGCTGGAACCGGAGGTGTCGCTCTTCGGACCTCAAGGTTCCC
- [0386] CTTAACACAGAGCGCCCCGCAGTCTTCGCGGAAAGCGTTCGGGGTAGGCG
- [0387] SEQ ID NO:65-智人RAVER1 5' -UTR
- [0388] NM\_133452.2
- TTAGCCAACGGGGCAAGGGGGCGGGGAAGGAGGTGGGGTTTCTCCCAGCCAATCGACGGGCGCGCC
- [0389] CTCGTTGCCGCTCT
- [0390] SEQ ID NO:66-智人KPNA6 5' -UTR (与SEQ ID NO:44一致)
- [0391] NM\_012316.4
- CTACAGATCCGCCATATTGTCTACTGAAAGCTGCCGCTGAAGCTGCCGCCGTTGCCTCCGCCGCCA
- [0392] AGAGTGAGCGAGCGGACCCGCG
- [0393] SEQ ID NO:67-智人SERINC5 5' -UTR
- [0394] NM\_178276.5
- GTTTTTACCCAGGCCTCGGCGCCTAGGCGCTTCGCCGAGGCTGATCTTCGTTCAAGTGTGAGCTGC
- [0395] GGCTGAGCCCAGCGCTCGAGGCGCGAGGCAGCCAGGAGGGCCCGTGCGGCGCGGGGAGCCAGCGAG  
CGCGCCTTCGGCATTGGCCGCCGCG
- [0396] SEQ ID NO:68-智人JUP 5' -UTR
- [0397] NM\_002230.2
- TGAGTTGCTGCCTTGCCAGAGTCCGGAGCAGCCGCCGCCGACCACGCCGAGCTCAGTTCGCTGT
- [0398] CCGCGCCGGCTCCCACCCCGCCCCGACCCGACCCGGCCCGGTCAGGCCCCATACTCAGTAGCCAC  
G
- [0399] SEQ ID NO:69-智人CPN2 5' -UTR
- [0400] NM\_001080513.2
- GGATTGAGCTGACCACAGGCCACACCAGACTCCTCTCTGCTCCTGAGGAAGACAGGGCAGCCCGGC
- [0401] GCCACCCGCTCGGCCCTCACGAAG
- [0402] SEQ ID NO:70-智人CRIP2 5' -UTR



- [0403] NM\_001312.3  
GGACAGCCGGGCAGGCGGGGCTGGGCGCGGGCGGGCGGCCCCGGAGGAGAACGGGCGGAGGGCGC
- [0404] GGGCCGACCGGGCGCACCGACC
- [0405] SEQ ID NO:71-智人EPT1 5'-UTR
- [0406] NM\_033505.2  
ACGCGTGCGCGGTGGGCGGAGCGCGGCTCTCCTACCTTCTCGGGCAGCCCAGTCTTTGCCATCCTT
- [0407] GCCCAGCCGGTGTGGTGCTTGTGTGTCACAGCCTTGTAGCCGGGAGTCGCTGCCGAGTGGGCGCTC  
AGTTTTCGGGTCGTC
- [0408] SEQ ID NO:72-智人PNP0 5'-UTR
- [0409] NM\_018129.3  
CATTGGCTCCGAGGACTTAGGACCTGTTAGCTTGGTTGGGCGACTGGCAAATCCTTCCTTCCCCGG
- [0410] GG TAGAAGTCCAGGGTGAGAAATTGGTTCCGAAC TCAAGGAACCCAGTGCCGGGCCACAGCCGGG  
TCACGTGGCCGGCGGCCCCCC
- [0411] SEQ ID NO:73-小家鼠 (*Mus musculus*) Dpysl2 5'-UTR
- [0412] NM\_009955.3  
CTCTCTCCCTCTTTCTTTTTTCCGCCCTAGCTGGATCTGTGTTGGAGGAGAGGAAAAGAGACAGAG
- [0413] GATTGCATCTGTTTTGAAATCTCCTAATAGCAAGACCAGTTAAGGGATTGTACCTTTTTCTACAA  
ATATAAATATATATATATTTTTAAACCAAGTCTTTTTTTTTCCGGCTATCTTTGCTTTAAAGCTGTC  
CTCTTGAGATTACTTCCTCCCGCCCCCGGAGAG
- [0414] SEQ ID NO:74-小家鼠 Cend1 5'-UTR
- [0415] NM\_007631.2  
TTTTCTCTGCCCGGCTTTGATCTCTGCTTAACAACAGTAACGTCACACGGACTACAGGGGAGTTTT  
GTTGAAGTTGCAAAGTCTTGCAGCCTCCAGAGGGCTGTGCGCGCAGTAGCAGAGAGCTACAGACTC
- [0416] CGCGCGCTCCGGAGACCGGCAGTACAGCGCGAGGCAGCGCGCTCAGCAGCCGCCACCGGAGCCCA  
ACCGAGACCACAGCCCTCCCCAGACGGCCGCGCC
- [0417] SEQ ID NO:75-小家鼠 Acox2 5'-UTR
- [0418] NM\_001161667.1  
CGCTGTCCCAGGACAGAGGTGAGTCAAGAGTTTAGTCCCAGGCCAGCCAGAGAACTGACAGAACAT
- [0419] ACCAAAGGTTCTTGGTACCTCCCGGTGCTCAGAGCAGACCCTAAAGGAAGCCAAGTCTTCCTGAG  
ACAGGCAGATCCAGG
- [0420] SEQ ID NO:76-小家鼠 Cbx6 (*Npcd*) 5'-UTR
- [0421] NM\_001013360.2  
GAGCGCGCGCCCCGCCGAGCGCCCCGCCGGGGGCTGAGCGCTGGGGCGCGTGCGCGAGCGGT
- [0422] GCAGCACCGGCCGCGGGAGCAGGGAGTATTATGGGCTGTGGGTGCCGCTGAGCAAG
- [0423] SEQ ID NO:77-小家鼠 Ubc 5'-UTR
- [0424] AGTTCGCTGGAGACTGCGAGTTCGCTCTGCTGTGTGAGGACTGCCGCCACCACCGCTGACG
- [0425] SEQ ID NO:78-小家鼠 Ldlr 5'-UTR
- [0426] NM\_001252659.1

- [0427] GCAGACTCCTCCCCCGCCTGGAAACCTCGCCCCTAGTACTGGGAATGACTCTGGGCGTGCGGCGTA  
GTTTGCAGCCGGGACACCGTGAGGCTTGCGAGCCCAGATTGCGAGCCGAGACACCGTGGGGCCCCGC  
GATCCAGTGTTTGCAGCGGGAACATTTCGGGGTCTGTGATCCGAGTGAGGACGCAACGCAGAAGCT  
AAGG
- [0428] SEQ ID NO:79-小家鼠Nudt22 5' UTR
- [0429] GGCCGGGCCAGCCGCTAAATTCCGAGATCAAGCTTCCAGATCCTGGGAGTAGATCAAGCGTGCCA  
GGTTGGAGAGCTGCCCTGTCAGACT
- [0430] SEQ ID NO:80-小家鼠Pcyox11 5' -UTR
- [0431] NM\_172832.4
- [0432] GAGCGCCAACCGCTAGAGGCCGATCCGGAGCGTGCTGCCCCGGTCACCACCCGCC
- [0433] SEQ ID NO:81-小家鼠Ankrd1 5' -UTR
- [0434] NM\_013468.3
- [0435] CCACAGGGCCAGTTCAGGGGTTTCATCCACAAGAGAGAAAAACATAGACTCACGGCTGCCAAC
- [0436] SEQ ID NO:82-小家鼠Tmem37 5' -UTR
- [0437] NM\_019432.2
- [0438] GAGTGCGACAGCTAGGCCAGGCGACAGCGCTGCCTACCAGAGCGCAGC
- [0439] SEQ ID NO:83-小家鼠Tspyl4 5' -UTR
- [0440] NM\_030203.2
- [0441] AGAAGGAAAAGGGGTGGAGCTAAGCACTCACTGCGGTTTTGCGCTGCGTCTGCAGAGGACAAGGAA  
AGCTCTGCAGGGCTGTCAGCTGCCAAA
- [0442] SEQ ID NO:84-小家鼠Slc7a3 5' -UTR
- [0443] GGGCGCTTGGCTTGCAAGGACCCTGAGCTGCGGCATTGAAGCACACCCAACCAACTCGACTGAAG  
TCAGCCTCACTGAACCGGATCTGAGAATCTTCTCTCTCTGGGCTTGCCAGGGCTCTCCGAACCTAG  
CTAGCATCCTCTTCAATTCCAACCTAGA
- [0444] SEQ ID NO:85-小家鼠Cst6 5' -UTR
- [0445] CGCGAAGGCTGTAAAAGCCGCGCAGGATGGAAGTCCAGACACTGAATCCGCGGCT
- [0446] SEQ ID NO:86-小家鼠Aacs 5' -UTR
- [0447] NM\_030210.1
- [0448] GAGTCTCGCGCTGTGGTTCGTCGGCGCACCGCTGATCCGCTCCACGCCTTGCGCTCTCCGCTCTCA  
GCCAAAGCCCGGCAGCCCCGGCCACGCAGCTCCGCAACC
- [0449] SEQ ID NO:87-小家鼠Nosip 5' -UTR
- [0450] NM\_001163684.1
- [0451] CTCCTGTCGGGCGGAAGTAGGAGGAGTAGAGTTTAAAAACAGTACTCTTTTCCGGTTCGGGACGT  
AGTTGAAGCAACGACAAGCCGGATAACCGCTCTTGAGACAGG
- [0452] SEQ ID NO:88-小家鼠Itga7 5' -UTR
- [0453] NM\_008398.2
- [0454] GTACTTAGCTGGTCCTGGGGCAGCAGCGGGAAGGGACTGAAGAGGGGTGCTGAGGTGAAAGATTGG  
AAGACCCGAAAGATCAGCTAGATTTTCGGAAACCAGAAACACCCTCGGGGAGACCTGGGGCTCCCG  
GCGCGCGACGATTTCTCGCACTAGCTGGGAGAGCGTTGATCCC

- [0455] SEQ ID NO:89-小家鼠Cend2 5' -UTR  
[0456] NM\_009829.3  
TGCCTGAGCGAGAGAGAGAGCGAGCTGAGGAGAGCCGGGCAGTTCGGAGGGAAGGACCGGTGCGA  
GTCAGGCGGCCCTTGAGGCTCCGCTCGCCACCTTCCACTCTTCTCTCTCTCTCCCTCTCTCTC  
[0457] TTTGCCATTTCTTTCTCTCCCAAATCTCCATTAGCCAAAGGAAGGAGGTAAGGGAAGCACTCC  
CCGACCCCCCGCACCTCCAAAAAATAATAATAATAAAAAAATTTACAGTCGGGACCGAGTGGTG  
GCCGGCTGGCT  
[0458] SEQ ID NO:90-小家鼠Ebp 5' -UTR  
CGGAACTGGGGCTATTAGGGAGCCTGCAGGTCTTCTGGAAAGTCGCAAGCCTGTGTTAGGAAGTC  
[0459] GCCCTGCGATCGCGCCGCCTGGGGTTTTCTGTCCCTTTGTCTCGTTTATGTACGAAGCTGCCAG  
CGGGCCATAGAAAC  
[0460] SEQ ID NO:91-小家鼠Sf3b5 5' -UTR  
[0461] NM\_009829.3  
GAAAAGACTGTCTACTCCTGCGTCTCGGTGGCGTCTTTCTCCCGCGCCTGCACGAACTGAGGTTTT  
[0462] GCGTGGCGGCGGCGGCACCGGCAGCGGCAGCGTCTCTCACTTGAACGCCGCGAGCGGCAGCTTCTC  
GTCTGTGTCCTGACCTCGGAGCCTACGAGCAGAGCGGCGCG  
[0463] SEQ ID NO:92-小家鼠Fasn 5' -UTR  
GCAGCGGCGTCCCGTCCAGTTCGCCTGCCGCGCTCCTCGCTTGTCGTCTGCCTCCAGAGCCCAGAC  
[0464] AGAGAAGAGCC  
[0465] SEQ ID NO:93-小家鼠Hmgcs1 5' -UTR  
CCCCGGCGCTGCCGAGTCGGGTGGGTGCGGTGGCTATAAAGCTGCGGAGGGCGGGAGGCACAGTCTG  
[0466] CGGTCTCCTTGCTTTGCTCGTTCTTCTTCCAGGGTCTGATCCCCCTTTGGTGGCTGAAGGAGGAACC  
GGTGACCGACCTGGAGACCACAGTTCTCTGTCCTTCACACAGCTCTTTCACC  
[0467] SEQ ID NO:94-小家鼠Osr1 5' -UTR  
[0468] NM\_011859.3  
TAGGCGGCTCGGTGCTAAGGGATGAGTGAGACGTAGCTCGGCCTCTCCCCAGCAAGTTCTCGCCGG  
CACTGGTTCAGATTCAGAAAGGGAGCCAGGAGCGAGGCTCAGCAGACCCCGGCGGGTAGAACTCCG  
GGGCTGTGAAGCGCTCCCGCGTGTCCGACTCTTGAGTACACGCGAGCTACAGAAGCGTCAGAAGTC  
[0469] TAGTTCGCCAGGGGTGCCAGTCGGCTCCAGAGTCTCTGGTCACTCAAGTCCAGCGCAGCGACCCTC  
ACAGACGCGCTCTGCCCTGGGACCGCGGCGGAACAAGATATTTGAAATCATTGCGGTTCCAGCGA  
CAGAA  
[0470] SEQ ID NO:95-小家鼠Lmnb1 5' -UTR  
[0471] NM\_011121.3  
GCCCCGCTCTCGCCGCCCTCCTCCTCGGCCCGCCGCGCTCTGAGCAGCCCGAGAGGAAACAAAG  
TGCTGCGGGCGGGAGACTCGCGTCCGCCGCGCACAGCCGTCTGCGTCTCCCGGTGCCCTGGCCTC  
[0472] TTCCCGCGCGCGTCTGCGAGTGTGCGTGTACACTACAAAGGGCGTCTGGCGGGCGATCGCGGCCC  
TCCCGCTTCGCTCTTTGTGCGGTAGCCCCGCCGCCACCGCCAGCCAGGTCCGCTCGATCCTCACC  
GGCCTGTGGTTTTGTACCTTCGGTCCCGCCGCCGCC  
[0473] SEQ ID NO:96-小家鼠Aarsd1 5' -UTR

- [0474] NM\_144829.1  
[0475] CCGGCTTTTCTGCGCGCGCGGGTCTCTCCGGCCTAGGAGACCGCG
- [0476] SEQ ID NO:97-小家鼠Vma21 5' -UTR  
[0477] NM\_001081356.2  
[0478] GCTTCGGCAGGGGCACTTCCGGCGCGACCGGCCTAGACTGTCACCCTGTTAGCCGCCGCTGAGCCT  
CCGCTAACTACCCGGCGCCTGCTGCTGCCCTCCCGGCCACGCC
- [0479] SEQ ID NO:98-小家鼠Kif20a 5' -UTR  
[0480] AACAAAGCTGTTCTCGTGGGCGGAGTTGTGCTCTCCAGCCGCAGGAGTCTTGCTGCGCGGGTGGG  
TAGAAGCTGGGATCAGAGAAGCGGCGAGGCAGACAGTCTTCGGACATAGACTGACTCTGCTTGTC
- [0481] SEQ ID NO:99-小家鼠Cdca8 5' -UTR  
[0482] NM\_026560.4  
TGTGCCCCGTTGAGTTTGAATTGGGTGGCGGTTAACCGAGGAGCCGCCCGTCCCTTAGTTGGAGCTG  
TGAGGGTTCCTCAGACTGTGTTTTGGGACCTGCAGGTAGGTTTCGGCAGAGTTCTGGAAACCTAGA  
[0483] CTCCAACGACTGAACTTTCTCAGCTCTCCGACCGCTCACACCCTCTCCCCGTCTCAGTCGCGGAGC  
CGGCTGCTTGGCCCCCTCGCTCGACGCAGCCAGGCGCC
- [0484] SEQ ID NO:100-小家鼠Slc7a1 5' -UTR  
[0485] NM\_007513.4  
GGTGCCCTGGAAGCTGACGCGCAGCACTGGCCGTGGCCGTGGGGCCCCGCGGAGGGCGGCGCGCGGC  
[0486] TGATGAAACCGGCTCGGATTCCGCCCCGCTGCGCCATCCCCCTCAGCTAGCAGGTGTGAGAGGCTTT  
CTACCCGCGGTCTCCACACAGCTCAACATCTTGCCGCCTCCTCCGAGCCTGAAGCTACCGTGGACT  
CTGCTGTGGCGTCTTGGCCCCCAGGTGCGGATCCTCCCCAGTGAGAAGTCCCACGAGTCTTACAGC  
[0487] AGATTCGCTCAGCACA
- [0488] SEQ ID NO:101-小家鼠Ubqln2 5' -UTR  
[0489] NM\_018798.2  
CGGAGACGGCCTGCAGGACCTGCTCTCTCAGCCCTCAGCCGAGGCCTACGCCGAGCCGAGTGCGCA  
[0490] GCCGACGACCGGGAGGAGCCGCAGCCTTCAACTCTGAGGTACTGTGATCCGCGCTGCCCGCCGGGC  
CGCCCCAGTCCGCTGCTGCGGCACCTCCTTCCCTCGCGCCCTCTTCGCTCGCCAGCGCCTTCCCTG  
TGAGCCTGCGTCACCGCGGCCGCC
- [0491] SEQ ID NO:102-小家鼠Prps2 5' -UTR  
[0492] NM\_026662.4  
AGCCCAGGCCACCGCAGCAGCAGCAACAGCCGAGCAACGGTAGCAGTAGTCTGCATCGCAGT  
[0493] CCCTTTCTCCTTCTCCAGCGCGCTCCTCAGTCCCCGGTCACC
- [0494] SEQ ID NO:103-小家鼠Shmt2 5' -UTR  
[0495] NM\_026662.4  
CTCTTTCACCTCGAACTTCACGGGGGCAATTTTCTCGCGCACGCGTTCTAAGAGCAGCTGGTTTTTCG  
[0496] ACCCCGGTACTACACCGATACAGAGTTAGTGGCTAGTCCTCCTGTGCCTCCTGTAGCG
- [0497] SEQ ID NO:104-小家鼠Kif22 5' -UTR  
[0498] NM\_145588.1  
[0499] GGCCGAAGAAGGAAGGCT

- [0500] SEQ ID NO:105-小家鼠Aurkb 5' -UTR  
[0501] NM\_011496.1  
GGAGATTCGAAAGCGTCCGGGTCGCGGGGTAAACCGGTTCTCCGTGTGCGAGCGCCTAGTGGCGTA  
GGCTGCGGCTTTGCGGGGAACTGCGGGGGCTGCAGTGGTCCACGGGGCTGATCGGGTTCCGTGGG
- [0502] CGGATCCACGTGCCCGCTATCCGCCTGGAAGGAGAGGTGCAGGAGTACCCCCGACCTTGGCTGCGT  
GCTGACTCGCTTCCTTCTGCCCCGCCAGGCTTGCACTCCCCGGGGATCTGCCTCTGCATCTCTTGC  
CTTCGCTGTTGTTTCCCTCTCTGTCCAGCTCCCCTCCCGCTCTCGCCCTGGAGA
- [0503] SEQ ID NO:106-小家鼠Fignl1 5' -UTR  
GTCAC TTCGCGCAGGCGCACTTCGAAAGGCGCGCTTTTCGGTGGCTGCGGTCCCGGCAGGGAGCAC
- [0504] AGCTCATCCTGTGCTGATAACATCGAGAAGTGTTTCAGTGCCTGGTAAAGTACATAGACCTTGCTTC  
ACTTGGAAC TCGGCCTTGATTTCTGCCGTTGGTCATAATCAGCAGAGTTCTCTCTAAACCTTTGAC
- [0505] SEQ ID NO:107-小家鼠Cad 5' -UTR  
[0506] NM\_023525.2  
AGTCGTGCTTCAGCTCACTACGCTTAGGGCTCTGGCTTGCCGCTCCCGCCTGCTCTCCAGCGCCCC
- [0507] GCGCAGCGAGCCACGTGGACCAACTCCGGCGCGCGGTGTTTCGCTTGGTTCCAGTGGGGCTCGCCGC  
GCCTCCCGCGTCTGCGTGCTTGCCCTGTCTCAGCTCCGACCCG
- [0508] SEQ ID NO:108-小家鼠Anln 5' -UTR  
[0509] NM\_028390.3  
GTGTCGTCCGGGGCGCTGAAATTCAAATTTTGAACGGCCGTGGTAGCCTACCGACTCCGTGGGTGC
- [0510] GGAGGGCAGAGCCGACTGGGTGTGAGAGCGCCCGCCGCTCGACTGCAGTCTCCTCCAGGAGCTG  
CGCCGAGCCTGCACTCACTTCTTTCTCTTCTGAGTTTGAACCGTCGGACCCACCGTCTAGCCGT  
CCACTGGTGAGGCCTGGGGCG
- [0511] SEQ ID NO:109-小家鼠Slfn9 5' -UTR  
GGGAATTCCACGCCACCTCCGCGCGCTCTGCGCTCTGGGATCCGGAGCGACCAGGACCTGGTGAGA
- [0512] CCCTCAGCTCCCCTCCACCTTCCCCGAGGTCCAGCACACCAGAACTGGAACCTGAGCAGCCCAGAA  
GCCAGGGTGGCACCCTGTGTCTCTCCTGTCTGAAGACCCGGATATTTTCTCAGACTTGGCGACAC  
GTTCCTTTAAAAGATCAGC
- [0513] SEQ ID NO:110-小家鼠Ncaph 5' -UTR  
[0514] NM\_144818.3  
AAAGACCACGCCCCAGTGACGTACGCGGCGGTTACCGCGCTTGCGCTCGCGATTTAAAACTTAC
- [0515] TCCGGAGACGTGGAGAGCAAG
- [0516] SEQ ID NO:111-小家鼠Cth 5' -UTR  
[0517] NM\_145953.2  
[0518] AGCCAAAGCAACACCTCGCACTCCTGCCCCAGC
- [0519] SEQ ID NO:112-小家鼠Pole 5' -UTR  
[0520] NM\_011132.2  
GCCTCGCGAGAGCACGTGGAGAGCGCGCCAAATTCTCCCCGGAGCCTGAGGGAGCTTTGGAGCGTC
- [0521] GCA
- [0522] SEQ ID NO:113-小家鼠Uhrf1 5' -UTR

- [0523] NM\_001111079.1  
AATTGGGGTGGAAGTCTCCCGCAGCAGCCTGTGCACACTAATAAAAACGCCCTGAGTTTTCGCGGG  
AAAAAAAGTCCTAGCAGCTGGAAGGAACCCGCGCTCTAGTGCTCACTTGGGTCTTCAGCCACTCAC
- [0524] GCGGCTCCCTTCTGGGTACCCAGCCGAGAGCCCTAGCCTAGAACCAGGCGTTCCAAGGGAGAGG  
AGAGTGCGGATCGCCGCCGTGAGAGAGTACATCGGCATC
- [0525] SEQ ID NO:114-小家鼠Gja1 5' -UTR
- [0526] NM\_010288.3  
TTTTAAAAGCTCTGTGCTCCAAGTTAAAAACGCTTTTACGAGGTATCAGCACTTTTCTTTCATTG  
GGGGAAGGCGTGAGGGAAGTACCCAACAGCAGCAGACTTTGAACTTTAAACAGACAGGTCTGAG
- [0527] AGCCCGAACTCTCCTTTTCTTTGACTTCAGCCTCCAAGGAGTTCACCACTTTGGCGTGCCGGCT  
TCACTTTCATTAAGTGAAAGAGAGGTGCCCAGAC
- [0528] SEQ ID NO:115-小家鼠Fam64a 5' -UTR  
CGGTGGGCTAGGAGAGGGTGTGTGATCTTCGGGTCCGGGTATCGAGCAGGGAGGATCTAGCGGGCAG
- [0529] CGGAAACTGGACCAAGCAC
- [0530] SEQ ID NO:116-小家鼠Kif2c 5' -UTR
- [0531] NM\_134471.4  
TGACGTTGTAGGGAGGCTGGCGCGGGATTAACTGCAGCGGTTTAGGCGTTGTAAACACAGCG
- [0532] CAGTATTAGCAGAGTCGTGGTTTCCAAGCTTCTTTCATTTGTGTTGCCTGTTGTTGTTCTGAGTC  
C
- [0533] SEQ ID NO:117-小家鼠Tspan10 5' -UTR
- [0534] NM\_145363.2  
CACTGGAGGAGGAGCTTGCAGCTCTCCAGCTCTGGCTTATACAGTTCACAGAGAAGCCAAGGGACA
- [0535] CCGGGACACCGTTATTTTAGGACTGTAACCTGTTTCAGAGAGACCCTGGCCACTGCTCTCCGTGGTT  
TTCTCCAATTGTGTGG
- [0536] SEQ ID NO:118-小家鼠Scand1 5' -UTR  
TTCATTTCCGGCCACCGAGGAGCTTCGCGGGTCGAGACGCAGGCGAGGCGCCGGACTGCGAAACAA
- [0537] AGGGGGACGCCAACAGCCGTAGTC
- [0538] SEQ ID NO:119-小家鼠Gpr84 5' -UTR
- [0539] NM\_030720.1  
TTATATGTCCAGCTGGAAGCCTGGCTGTCCCTAGAAAAGCTGGAAGCCTGACTGCCCCCTCAAAGA
- [0540] CCTGCTCTTTAGGAGAGCTAGATATTGTTTACTGAAGACAAGTGTGAAAACCTGGGAACCTCAGTCT  
CCATC
- [0541] SEQ ID NO:120-小家鼠Tpqs1 5' -UTR
- [0542] BC138516.1
- [0543] AAATGGTCGTCGAGGGAAGGCGCCTCATCGCGCCGTGAATT
- [0544] SEQ ID NO:121-小家鼠Cc117 5' -UTR
- [0545] NM\_011332.3
- [0546] AAGACAGGCAGAAGGACCC
- [0547] SEQ ID NO:122-小家鼠Fads3 5' -UTR

- [0548] NM\_021890.3  
AAGGGAGGGAAAAGTTCCCTGCGCGGAGAGCCGGGCAGGCGCACGCTCTTACGGCGGCCGCAGCGG
- [0549] CAGGGCGGGGCCGGTGGGGCGGGCGGAGGAAGACCCCTGATTGCCACCTCGCCTCCCTCAGTGTCT  
CTCTTAGACCTTGGTCACGTACCGGGGTCCGGAGGACTTGTGTACAGCGGCA
- [0550] SEQ ID NO:123-小家鼠Cers6 5' -UTR
- [0551] NM\_172856.3  
AGCGCGCATCCCCAGTGCCCTGAGCTGCAGAGAGCTCGGAGGAGCGCGGGAGCAGCGACACCGGAG  
TGGACAAAGCAAG
- [0552] TGGACAAAGCAAG
- [0553] SEQ ID NO:124-小家鼠Alkbh7 5' -UTR
- [0554] NM\_027372.1  
AGCCTGCTGTGGTCAATCCCTGAAG
- [0555] AGCCTGCTGTGGTCAATCCCTGAAG
- [0556] SEQ ID NO:125-小家鼠Ms4a8a 5' -UTR
- [0557] GTAGAGAGGACTTTGGATCATTCTAA
- [0558] SEQ ID NO:126-小家鼠Cxcr4 5' -UTR
- [0559] NM\_009911.3  
AATTTTGTTCCTGGTGCAGCAGGTAGCAGTGAAACCTCTGAGGCGTTTGGTGTCCGGTAACCAC  
CACGGCTGTAGAGCGAGTGTGGCC
- [0560] CACGGCTGTAGAGCGAGTGTGGCC
- [0561] SEQ ID NO:127-小家鼠Gprc5c 5' -UTR
- [0562] CAGAAACTCCGATCGCCTTTCCCAACCCGAAAGTGCGCGTCGGCCGAGCCTGGAGGGACCCAGCTG  
AAGCCTGGCCTGGGAGCCAGG
- [0563] SEQ ID NO:128-小家鼠Fen1 5' -UTR
- [0564] NM\_007999.4  
AGATCACGTGACGAGAGCGCGGGCTTTGGAAGGCGGCGAAGCTGGGAACGATACTGAAAGAACGGG  
CTCGGGACTGTCCAGAGAACGCTGTGGACTCCAAACCACTGCTAGCTGCTTAAGGCTCGTGCATC  
GAGACGGGGTGAGGTCTCGCGGAAGCGTCTCTGAAAGCGGCAGAGCCGCGGGAACAGCACCGGGCA  
GCCCCGGGCTTGGGCCATTGCTCTGCTCCGAACATTCTCTTCGCCGGTAGGAAGAAGCCATTGCT  
CCTGTGCTACC
- [0565] GAGACGGGGTGAGGTCTCGCGGAAGCGTCTCTGAAAGCGGCAGAGCCGCGGGAACAGCACCGGGCA  
GCCCCGGGCTTGGGCCATTGCTCTGCTCCGAACATTCTCTTCGCCGGTAGGAAGAAGCCATTGCT  
CCTGTGCTACC
- [0566] SEQ ID NO:129-小家鼠Cspg4 5' -UTR
- [0567] NM\_139001.2  
GACTTGCGACTTGCGACTCGGTGCTGTCCAGATTAGCAGCGCTCTGTTCCCTTCGCCTTACAAGTCC  
AGGCACCCAGCCCCGCCGCG
- [0568] AGGCACCCAGCCCCGCCGCG
- [0569] SEQ ID NO:130-小家鼠Mrpl34 5' -UTR
- [0570] NM\_053162.2  
GAAGGCATTGTCAAGGAGCCCAGAGTGTAGGAAGTGGCATTGACCCGGACGACCGGACCATTGACC  
CGCCGCTGGAT
- [0571] CGCCGCTGGAT
- [0572] SEQ ID NO:131-小家鼠Comtd1 5' -UTR
- [0573] CGGCTCTGTGGGCAGGGCCCGGACAAGGTGGAGCCCTGTGCGCTTACTGGTCCCAGTCCCATCCTC  
GCTACTCCAGCACACGTGACCTCCGCCGCCGCTGCTAACCTGCACC
- [0574] SEQ ID NO:132-小家鼠Armc6 5' -UTR

- [0575] NM\_133972.2  
CGCTGCTCACTTCTCAGGGCCGAGTCTACCTAGGTGAAAACATCTCTTTTGTAGGGGAAGCGACAC  
CTAGGAGTAGCATTGAGAGTACTGCAAGAGTCTCAGGCCACGCGTATTCTGTGGCTCCCTTCGCT
- [0576] TTACCTGTGCTCGCGGTGCGCCCTCGTTGCCCCGAAAGGAGCCTCTCGGCTGAAGGAGGGACGCCG  
AGGCGAGGCGGCGCCTCTGCGCTTGCGCTTATATGGTGGCTCGAAGGAGCCAAAGGCAAAGTCTCG  
GGTGACAGCTGCGAGCCCCACCCTTTCCCCACGTGGCTGCGTAGACCCGGCAGTG
- [0577] SEQ ID NO:133-小家鼠Emr4 5' -UTR
- [0578] NM\_139138.3  
AACATTCCCTTGAGAAAGAGAGAACAAGATAAGCAGTGGTGCACCTCCTCCTTCATTGCTGCTGAGA
- [0579] ATGTTCCAGGCTGAGTGAGAAGTAAAAATTCATCATCTCTGAAGAACTCTTACCCAGCCCTGTTGA  
AGAAATTCCCAGA
- [0580] SEQ ID NO:134-小家鼠Atp5d 5' -UTR
- [0581] NM\_025313.2  
CCTTCGGAGAATCCTGTGCGCGTGCGTTCTTGTGGGAAGTGC GCCTCCAGAAGGCACCGCGCGTC
- [0582] GTCCCTTCTCCCTCCCTGAAGGCCGCTCGCTTGCCAGTGTGTGCGCCGCCCGCAAGCTAGAGTC  
CACTGACTTTTCCGCCACC
- [0583] SEQ ID NO:135-小家鼠1110001J03Rik 5' -UTR
- [0584] GCGGAGTGGGCGGGGAACACCTCGCCGAGGCAGTGAGGGACCAGGCTCTCCAAGGACAGAAAA
- [0585] SEQ ID NO:136-小家鼠Csf2ra 5' -UTR
- [0586] CCCAACCTGCAGATGAGGAAGAGGAAGCGGGAGACAGACGGACAGACGGACGACACAGACGCTGC  
GCCCAGGCCCTCACCATCCATCGCAGGAAGCCCCCTGTCTCAG

[0587] 这些野生型5' -UTR元件中的一些与NCBI GenBank鉴定的公众可获得的5' -UTR序列不同,参见表1(实施例1)。该表格反映了本发明人的测序结果。

[0588] 由此,本发明提供了选自由下述组成的组的5' -UTR元件:SEQ ID NO:45,SEQ ID NO:47,SEQ ID NO:49,SEQ ID NO:77,SEQ ID NO:79,SEQ ID NO:84,SEQ ID NO:85,SEQ ID NO:90,SEQ ID NO:92,SEQ ID NO:93,SEQ ID NO:98,SEQ ID NO:106,SEQ ID NO:109,SEQ ID NO:115,SEQ ID NO:118,SEQ ID NO:125,SEQ ID NO:127,SEQ ID NO:131,SEQ ID NO:135,并且可以用于本发明的任一方面。

[0589] 用于本发明的人工5' -UTR元件的实例

[0590] 在一些实施方案中,本发明所述的5' -UTR元件与野生型5' -UTR元件不同。此类5' -UTR元件称为“人工5' -UTR元件”。

[0591] 优选地,人工5' -UTR元件与野生型5' -UTR元件表现出一定程度的序列同一性,所述序列同一性(a)小于100%,同时(B)大于10%,大于20%,大于30%,大于40%,大于50%,大于60%,大于70%,大于80%,大于90%,大于95%,所述野生型5' -UTR元件选自由SEQ ID NOs:1-136组成的组。典型地,所述人工5' -UTR元件与作为其基础的野生型5' -UTR元件不同,其中至少一个核苷酸,如两个核苷酸,三个核苷酸,四个核苷酸,五个核苷酸,六个核苷酸,七个核苷酸,八个核苷酸,九个核苷酸,十个核苷酸,或多于十个核苷酸被交换(exchanged)。例如,在野生型5' -UTR元件包含认为是不利的核苷酸元件的情形中,所述核苷酸交换可以是推荐的。例如,在一些实施方案中,认为是不利的核苷酸元件选自:(i)内部



ATG三联体(即,除了本发明的核酸的可读框的起始密码子之外的ATG三联体)或(ii)限制性酶识别位点(切割位点),特别是由用于制备(克隆)本发明的人工核酸的工序的限制性酶识别(可切割)的限制性酶识别位点(切割位点)。因此,可以特异性引入特定的碱基(对于各个野生型碱基,以交换,优选置换而引入),以使所述人工5'-UTR元件不包含认为是不利的核苷酸元件。在具体的实施方案中,所述人工5'-UTR元件选自由SEQ ID NOs:137-151组成的组:

[0592] SEQ ID NO:137-人工序列(基于SEQ ID NO:3)

[0593] ATCTTAGCTTCTCTGTCTTCCTCCCTCCCTCCCTTCCTCTTACTCTCATTTCATTTTCATACACACTG  
GCTCACACATCTACTCTCTCTCTATCTCTCTCAGA

[0594] [该序列对应作为其基础的野生型序列,不同之处在于AAGCTT被TAGCTT替换,上述用下划线表示强调]

[0595] SEQ ID NO:138-人工序列(基于SEQ ID NO:9)

[0596] ATCACAGCCCTTCCCCGATCCTCTCCGTGGGAGCCAGCGAGCCTCTCTCCCTGATCTTACGTGCTC  
AAGGGAGCTCACACGTTACCAACTCACCTTGAAGTCATCTCAAGAACAAAAGACAACCTGAAAGA  
AGCTGTTGTGAAGGCAGAGCAGCATCTGCTGAAGAGACAGAAACCAGCCCCAGAGGTGTCACAGGA  
AGGCACCAGCAAGGACATTGGTCTTTGATTTGATTTCAGCAGTCCTGTCAAGTATAATAGTG

[0597] [该序列对应作为其基础的野生型序列,不同之处在于ATG优选地被TAG替换,上述用下划线表示强调]

[0598] SEQ ID NO:139-人工序列(基于SEQ ID NO:14)

[0599] TAGGGCTGGAGCTCCGGGCAGTGTGCGAGGCGCACGCACAGGAGCCTGCACTCTGCGTCCCGCACC  
CCAGCAGCCGCGCC

[0600] [该序列对应作为其基础的野生型序列,不同之处在于ATG优选地被TAG替换,上述用下划线表示强调]

[0601] SEQ ID NO:140-人工序列(基于SEQ ID NO:17)

[0602] CAGTCGTCCCTGCGCGTCGTCTCTCGCCCTCCAGGCCGCCGCGCCGCGCCGGAGTCCGCTGTC  
CGCCAGCTACCCGCTTCTGCGCCCGCCGCTGCC

[0603] [该序列对应作为其基础的野生型序列,不同之处在于GACGTC被GTCGTC替换,上述用下划线表示强调]

[0604] SEQ ID NO:141-人工序列(基于SEQ ID NO:18)

[0605] TAGGCGGACCTCCGGAACCGTAGATTCCGGGCGGTTCGGAGCCGCCGGGAGCTGTAGTTCTCCCGC  
GGCTCAGAGAAGTAGGCAGAGAGCGGACCTGGCGGCCGGGCAGC

[0606] [该序列对应作为其基础的野生型序列,不同之处在于ATG被TAG替换,上述用下划线表示强调]

[0607] SEQ ID NO:142-人工序列(基于SEQ ID NO:29)

[0608] ACCGGTAGTTAGCTCGGCTTAGAACCACTAGGCGCCTGGCGGGGGTGATCTGTTCGGAGCGACCGGCTT  
GGCGCCTGCCTGTCCCCAGCCCCTCTCAGCTTGAACCTCCTTCAAGTCTGGGCCCTCGAGGCT  
TCCAGAGCGGCCTCCAGGGGTGCAGTCTCAGTTCCCCACGCCAGCCGTCTCCGTCTCCGCCTCCT  
CCGGGCCTGGCAGGTGGCACTGTCCGGAGGCGGAGCCTTGGGCGAGGGGTGGTTGCGGCGGAGGAC  
GCAACCGAGCGGGCCTGCGGCCTCAC

[0609] [该序列对应作为其基础的野生型序列,不同之处在于两个ATG分别被TAG替换,上述两处下划线表示强调]

[0610] SEQ ID NO:143-人工序列(基于SEQ ID NO:31)

CTTTTCCCGGCACTAGCGCACCGCAGCGGGTCGCGCGCCCTAAGGAGTGGCACTTTTAAAAAGTGC

[0611] AGCCGGAGACCAGCCTACAGCCGCCTGCATCTGTATCCAGCGCCAGGTCCCGCCAGTCCCAGCTGC  
GCGCGCCCCCAGTCCCGCACCCGTTCTGGCCAGGCTAAGTTAGCCCTCAC

[0612] [该序列对应作为其基础的野生型序列,不同之处在于ATG被TAG替换,上述用下划线表示强调]

[0613] SEQ ID NO:144-人工序列(基于SEQ ID NO:42)

[0614] CTTGATCTGTGGCGGGGCGCGGCCTGTGG

[0615] [该序列对应作为其基础的野生型序列,不同之处在于AGATCT被TGATCT替换,上述用下划线表示强调]

[0616] SEQ ID NO:145-人工序列(基于SEQ ID NO:43)

GTAGATTTGAGGGACAACAATTCAGTGCCCGGGGGCCGAAAGGCAGCAGAAGGCGGGCACCAAAGG

[0617] ATAGGCACCCGGAAGGTGGACTCCGAGGAGGAGAGAGGACAGGGGTCTCTCACCCAGCTCCTGGT  
CACC

[0618] [该序列对应作为其基础的野生型序列,不同之处在于,(i) ATG被TAG替换,并且(ii) GAATTC被CAATTC替换,上述分别用下划线表示强调]

[0619] SEQ ID NO:146-人工序列(基于SEQ ID NO:52)

GGCCGGACGCCTCCGCGTTACGGGATGAATTAACGGCGGGTTCCGCACGGAGTTGTGACCCCTAC

[0620] GGAGCCCCAGCTTGCCACGCACCCCACTCGGCGTCGCGCGGCGTGCCCTGCTTGTACAGGTGGG  
AGGCTGGAACATATCAGGCTGAAAAACAGAGTGGGTACTCTCTTCTGGGAAGCTGGCAACAATAGGT  
AGTAGTGATAT

[0621] [该序列对应作为其基础的野生型序列,不同之处在于,三个ATG分别被TAG替换,上述分别用下划线表示强调]

[0622] SEQ ID NO:147-人工序列(基于SEQ ID NO:109)

GGCAATTCCACGCCACCTCCGCGCGCTCTGCGCTCTGGGATCCGAGCGACCAGGACCTGGTGAGA

[0623] CCCTCAGCTCCCCTCCACCTTCCCCGAGGTCCAGCACACCAGAACTGGAACCTGAGCAGCCCAGAA  
GCCAGGGTGGCACCACTGTGTCTCTCCTGTCTGAAGACCCGGATATTTTCTCAGACTTGGCGACAC  
GTTCCCTTTAAAGATCAGC

[0624] [该序列对应作为其基础的野生型序列,不同之处在于GAATTC被CAATTC替换,上述用下划线表示强调]

[0625] SEQ ID NO:148-人工序列(基于SEQ ID NO:110)

AAAGACCACGCCCCAGTCACGTCACGCGGCGGTTACCGCGCTTGGCGCTCGCGATTTAAACTTAC

[0626] TCCGGAGACGTGGAGAGCAAG

[0627] [该序列对应作为其基础的野生型序列,不同之处在于GACGTC被CACGTC替换,上述用下划线表示强调]

[0628] SEQ ID NO:149-人工序列(基于SEQ ID NO:119)

- TTATTAGTCCAGCTGGAAGCCTGGCTGTCCCTAGAAAAGCTGGAAGCCTGACTGCCCTCAAAAGA
- [0629] CCTGCTCTTTAGGAGAGCTAGATATTGTTTACTGAAGACAAGTGTGAAACTGGGAACCTCAGTCT  
CCATC
- [0630] [该序列对应作为其基础的野生型序列,不同之处在于ATG被TAG替换,上述用下划线表示强调]
- [0631] SEQ ID NO:150-人工序列(基于SEQ ID NO:120)
- [0632] AATAGGTCGTCGAGGGAAGGCGCCTCATCGCGCGTGAATT
- [0633] [该序列对应作为其基础的野生型序列,不同之处在于ATG被TAG替换,上述用下划线表示强调]
- [0634] SEQ ID NO:151-人工序列(基于SEQ ID NO:136)
- [0635] CCCAACCTGCAGTAGAGGAAGAGGAAGCGGGAGACAGACGGACAGACGGACGGACACAGACGCTGC  
GCCCAGGCCTCACCATCCATCGCAGGAAGCCCCCTGTCTCAG
- [0636] [该序列对应作为其基础的野生型序列,不同之处在于ATG被TAG替换,上述用下划线表示强调]
- [0637] 3' -UTR元件
- [0638] 优选地,所述至少一个3' -UTR元件包含下述核酸序列或由下述核酸序列组成:所述核酸序列与选自SEQ ID NO:152-204组成的组的核酸序列或对应的DNA或RNA序列分别具有至少约1,2,3,4,5,10,15,20,30或40%,优选至少约50%,优选至少约60%,优选至少约70%,更优选至少约80%,更优选至少约90%,甚至更优选至少约95%,甚至更优选至少约99%的同一性,或者其中所述至少一个3' -UTR元件包含下述核酸序列的片段或由下述核酸序列的片段组成:所述核酸序列的片段与选自SEQ ID NO:152-204组成的组的核酸序列或对应的DNA或RNA序列分别具有至少约40%,优选至少约50%,优选至少约60%,优选至少约70%,更优选至少约80%,更优选至少约90%,甚至更优选至少约95%,甚至更优选至少约99%的同一性。
- [0639] 在下述详细显示的序列中,SEQ ID NO:152-203可以认为是野生型3' -UTR序列,并且SEQ ID NO:204可以认为是人工5' -UTR序列。
- [0640] SEQ ID NO:152-智人NDUFS7 3' -UTR
- [0641] NM\_024407.4
- [0642] CGCCGCCGCCGCCGCCGCCGAGCCTGTCGCCGTCCTGTCCCCAGCCTGCTTGTGTCCCGTGAGGT  
TGTCATAAACCTGCCCTCGGGCTGCCGCCTCCC
- [0643] SEQ ID NO:153-智人PHGDH 3' -UTR
- [0644] NM\_006623.3
- CCTTGAGGCTCACTGGTCCCTGCCTCTGGGGCTTTTCTGAAGAAACCCACCCACTGTGATCAATAG
- [0645] GGAGAGAAAATCCACATTCTTGGGCTGAACGCGGGCCTCTGACACTGCTTACACTGCACTCTGACC  
CTGTAGTACAGCAATAACCGTCTAATAAAGAGCCTACCCCCAACTCCTTCTGC
- [0646] SEQ ID NO:154-智人TSPO 3' -UTR
- [0647] NM\_000714.5

- [0648] GTGCCCCGCCCCACCAGGGACTGCAGCTGCACCAGCAGGTGCCATCACGCTTGTGATGTGGTGGCCG  
TCACGCTTTCATGACCACTGGGCGCTGCTAGTCTGTGAGGGCCTTGGCCCAGGGGTGAGCAGAGCTT  
CAGAGGTGGCCCCACCTGAGCCCCACCCGGGAGCAGTGTCTGTGCTTCTGCATGCTTAGAGCA  
TGTTCTTGGAACATGGAATTTTATAAGCTGAATAAAGTTTTTGACTTCCTTT
- [0649] SEQ ID NO:155-智人ATP5D 3' -UTR
- [0650] NM\_001687.4  
GCGGTGCGTACCCGGTGTCCCGAGGCCCCGGCCAGGGGCTGGGCAGGGATGCCAGGTGGGCCCAGCC  
AGCTCCTGGGGTCCCGGCCACCTGGGGAAGCCGCGCCTGCCAAGGAGGCCACCAGAGGGCAGTGCA  
GGCTTCTGCCTGGGCCCCAGGCCCTGCCTGTGTTGAAAGCTCTGGGGACTGGGCCAGGGAAGCTCC  
TCCTCAGCTTTGAGCTGTGGCTGCCACCCATGGGGCTCTCCTTCCGCCTCTCAAGATCCCCCAGC  
CTGACGGGCCGCTTACCATCCCCCTGCCCCTGCAGAGCCAGCCGCCAAGGTTGACCTCAGCTTCGG  
AGCCACCTCTGGATGAACTGCCCCCAGCCCCCGCCCCATTAAAGACCCGGAAGCCTG
- [0651] SEQ ID NO:156-智人EXOSC4 3' -UTR
- [0652] NM\_019037.2
- [0653] CCACCCAGCCACCCATGTCCAGAATAAAACCCTCCTCTGCCCACAC
- [0654] SEQ ID NO:157-智人TUBB4B 3' -UTR
- [0655] NM\_006088.5  
AGCCTTCAGTCACTGGGGAAGCAGGGAAGCAGTGTGAACTCTTTATTCACTCCAGCCTGTCTTG  
TGGCCTGTCCCACTGTGTGCACTTGCTGTTTTCCCTGTCCACATCCATGCTGTACAGACACCACCA  
TTAAAGCATTTTCATAGTG
- [0656] SEQ ID NO:158-智人TUBA4A3' -UTR
- [0657] NM\_006000.1  
AGCAGCTGCCTGGAGCCTATTCATATGTTTATTGCAAAATCCTTTCGAAATAAACAGTTTCCTTG  
CACGGTT
- [0658] SEQ ID NO:159-智人EMP3 3' -UTR
- [0659] GCGCCCCGCTCGCTCGGCTGCCCCGCCCCCTTCCCGCCCCCCTCGCCGCGCGTCCTCCAAAAA  
TAAACCTTAACCGCGG
- [0660] SEQ ID NO:160-智人CRIP2 3' -UTR
- [0661] NM\_001312.3  
GCTACAGCGGCTCTCATGATGTGGGCTCACCTGCGCCCCAGACCCTGCAGGGGCCCCCCTGCTTGG  
CTCTGCTGGGAGAGTGCTCAGCCGCCAGTCCTGCCTGCAAGCCCAGGGCGAGTATTGGAGGAGGG  
GCAGCCACGGGCAGAGCACCATGCCATCCCCGAGTCTCTGGTGTGTCTGCCCCCTCTGGCATCCT  
CTGGGCGTCCCATGATCCCTTCTGTGTCTGCGTGTCCGAATCCCCGTGTGACCCTGTCCCAGCATT  
TTCCCGCCGACCCTGCGTGTCCCCGTGGCGCTGTCCGCTCTCCCTCTCCTGCTGCCCACCCACCTG  
CCAGTGTTATTTATGCTCCCTTCGTGGGTGATGGCCACGCCCTCACCATGTCCCTGGCAGAGGGCT  
TCCCTCCGGGATCCCCTGCCTGGTGCCACACTGCCTCGCAAGCGCTCGCCACCCTCACGTGGCTC  
ACCTGCTGTTGAGCCTTGCTGTCAATAAACGTTTTGAGGATTGCAGGATTGTC
- [0662] SEQ ID NO:161-智人BRAT1 3' -UTR
- [0663] NM\_152743.3

65

- [0686] GCGGTGGAGCTGCGGGCGGCCAGGGCCGGGCGCTCTGTGCGGACTGGGGCCATGATCGGGCCCCGGG  
GGCCTGAGCCTGGGACCCACCCCGTGTAAATGAAAATGAGTTTTGGCAGCGCCTGTGGTC
- [0687] SEQ ID NO:168-智人AMN 3' -UTR
- [0688] NM\_030943.3  
GCGGCCGCCTGACCGTCGACCTTGGGGCTCTCCACCCGCTCTGGCCCCAGTCGAACTGGGGGCTAG
- [0689] CCACCTCCTCGTCCAGCCCCAAACCTCCCCTTCTTTCCCCCTCCTCCGGGGGCCAAGGACAGGG  
TGGCCTTACTCAGTAAAGGTGTTTCCTGC
- [0690] SEQ ID NO:169-智人CYBA 3' -UTR
- [0691] NM\_000101.3
- [0692] CCTCGCCCCGGACCTGCCCTCCCGCCAGGTGCACCCACCTGCAATAAATGCAGCGAAGCCGGG
- [0693] SEQ ID NO:170-智人PCOLCE 3' -UTR
- [0694] NM\_002593.3  
GACGCAGGCCAGCCCCGGCCCCCTAGCCCTCAGGCCTTCTTTCTTATCCAATAAATGTTTCTTAAT  
GAGG
- [0695] SEQ ID NO:171-智人MAP1S 3' -UTR
- [0696] NM\_018174.4  
CCCCATCGCCGACACGCCCCCACTCAGCCCAGCCCGCCTGTCCCTAGATTAGCCACATCAGAAA  
TAAACTGTGACTACACTTGGCAA
- [0697] SEQ ID NO:172-智人VTN 3' -UTR
- [0698] NM\_000638.3  
GAGTCAGAGCCCACATGGCCGGGCCCTCTGTAGCTCCCTCCTCCCATCTCCTTCCCCCAGCCCAAT  
AAAGGTCCCTTAGCCCCGAGTTTAAA
- [0699] SEQ ID NO:173-智人ALDH16A1 3' -UTR
- [0700] NM\_001145396.1  
TGCCTGAGCGCCACCTACTGCATTTTGGACACCTCACACCAAGGGGAGATGCACCCACAGACACC  
TGGGACTTTCCCCTTCTGGTTCCCTGTGTCTCCCAATAAACTCTCTGACCAACCCTAGCTGTGCTTC  
TGCGAGAAGAAAGGGTGTAGCAACTTCTGGCAGATATGAGGCTTTTTTCTTTTTTTTTTTTTTTT  
TGAGACAACGTCTGGCTCTGTACCCAGGCTGGAGCGCAGTGGCACAATCTCGGCCCCACTGCAGCC
- [0701] TCGACCTCTGGGGCTCAAGGGATCCTCATGCCTCAGCCTCATATGTAGCTGGGGCCACAGACATGC  
ACCACCACACCTGGCTCGAGGCCATTTTAGTTCTGAGGTTGAGCAGCTCAGGAGCCGGCTCCAGCA  
CGGTGCTGTGTTTGTGAAACAGAGAAAGGGGACCCCCGAGGACCCAGACAGGGCCTTAGGACTCT  
CATATCTTCTTGTCTTCTCCATCTGGTGGCTCTTGCTCTCTGGTTTTCTCTACCTTTTCATGGCCC  
CAGAATCCATATGCAGTAAAGGAACTCTCTGGAATAAAATTAAAGTCCTCCT
- [0702] SEQ ID NO:174-小家鼠Acox2 3' -UTR
- [0703] NM\_001161667.1  
AAAGCCAAAGGATTAGGACCAAGCAGCACCATGGCCTTCCTATGGCACATATGCATATAAAGAAT  
TTAAAGCACGGGGTGGCGTGCGGCTTGTTTCAGATCAGCGAGTAAACTGGTACATGAAAGGATGTTT  
ATCATATTATTCTGCTACTGAGTACATCTGAAACTTTCCCTTGCTGTTTGTAGTACGTATTTGGC  
TAAATGCTAGAATTTTGCTTTAAATACAGCAAAGCTAATAAACTTGTTAGTAACTA

67

- [0727] SEQ ID NO:182-小家鼠Cst6 3'-UTR  
CCGGCCCCAGGGTGACAGAGTCAGGGCCCCATGGGAACAAGCATGGTGGAGGCGCCTCAGGGCCAC
- [0728] GGGCCACGTCTGTTGCAATAAATCTGCAGCTCTGCTTCTTGC
- [0729] SEQ ID NO:183-小家鼠Ebp 3'-UTR  
AGAGCCGGAGAGTGGGCTGAGTCCCTGCTGATGAGGCGCTCTCCTCACTTCCCAGAAGACTCAAAT  
CTTCTTCCTCCTCGCACAGGTTGGGGGGGTCAGAACTATCGGACTTGTCCTCACTCAAACATGATGA
- [0730] GTGAGCACACAAAGGGCCAGAGTCAGAAAAGGAAGGGAACAAGTTGAGCTACTGCTAGGAACCTGA  
GGGAAAAGATGGATGAGAAGGTGGCAAGTCCGTGACGGTGGCAACTTCCCAAACCAGGAAATAAAA  
TGTCTTTTACTATAA
- [0731] SEQ ID NO:184-小家鼠Sf3b5 3'-UTR
- [0732] NM\_009829.3  
GGTGAGCGCTGCCCAGTCTTCCCCGTGTGCCGGCTGCGAGCCTCCTTGCTCCTGCATCTCGACCAT  
TCCGTGTTGGCTGTGTGCGCTGACCTGCGTACCTGTGGAGGATTGGAACAAGTCATGGAGAGACT
- [0733] GTCCGGGTCCGCTCCTTGTGAAGTGTGCAGAAGGAGTGATCCCAGCATCGGCAAGCGAGGGAGAAG  
ACTGCACGAGAGTGATGCGCATTTTCGAGTCTGCTTTTCGATAGTTGATGTCTTCTTGCCCT
- [0734] SEQ ID NO:185-小家鼠Plk1 3'-UTR
- [0735] NM\_011121.3  
GTCTCTCTCCTTCGGACTGGTGCCCCCTTCACTCTGCTAGCTCTGAGCCTGCACTGTAGGCTCCTGG  
GGGCTGCTGTTCAAGTGCCCCCTGGCCCTGGGGGCTGGGCAGGAGCAGGGCCCCCTTTGAAGGGGTTG
- [0736] CTGTGTGTAAGTTATTTTGTACATGTCTGGGTGTGGGTTTACATCTGCTTCCCTGCCCTCGCTCAG  
CCCACTGTATGAATTGTATAAATGTTTCTATTAAATTGGGACTGCTCTTTCCTTAGCTTTATATAT  
ATTAAAATGTGTACATCTCTC
- [0737] SEQ ID NO:186-小家鼠Aarsd1 3'-UTR
- [0738] NM\_144829.1
- [0739] GGGGCCAGGACGTCGTCCCTGTGACCAACAGTAAAATATTGTGACTC
- [0740] SEQ ID NO:187-小家鼠Cdca8 3'-UTR
- [0741] NM\_026560.4  
GAGGACAACAGGACACACAGTGGCAGCAGGGACTGTGGTAGCAGAGTGACACATCTGTCCTTCTT  
CTGTGGGGTCCTTCACTGCCAACACCTGCAACGGTGCTTTGTCTCTCTGACAGCTATGGTGTCTTG  
CTGCACACTTCTAGTTAGTGGGAATTTTAGACGGGGAACACAGGGCTAGTCAGGGCCTTTGTGTGC  
TTGGTGTGGAGTGACTGAGAACCGTCTATGGTTCAAGGTCCCACTGGGGATAAACTGCTTAGAGCA
- [0742] CTGTCCTAGAGGGCAAGTGATAGCCTTCGCCTCCGGGCCAGGCAGGCTATGCAGTCAGCAGTAGGG  
TCTGTGCTCCATGCGGGTCCAGGCGCACGGCTCTCCTATTCTGTTGTCAATTTGTGCCCTCTATGGG  
CAGGTGTGTTTCAAGTTGGTTTTCTGTTGCTGAGGCTTTCATACACATCAGTTACCATCTCAGCTG  
ATTTGTCTACTGAAAGCTTGCTGTTTTCAATAAATCTTAGTTTGCCATGGTTTTTAAGTC
- [0743] SEQ ID NO:188-小家鼠Kif22 3'-UTR
- [0744] NM\_145588.1
- [0745] GACACTACCTTCTTTTAAAAATCCTTGTATAGTGACGTGTTGTGTAAATACAGTTTTTTATTCC
- [0746] SEQ ID NO:189-小家鼠Cad 3'-UTR



- [0747] NM\_023525.2  
GTCCCAGCTTCCTTTCTTGTCTCTTCAGGCCAGCTGCTGGGCAAGGAATCCCGATGCCCCCCCCA  
CGGGGGCAGCACACTTAGCAGACAATCCTGGAGCATACAGATAGCCCAAGCACATAGATGTACTAA
- [0748] ACTGGGACCATGCTTTAGGCTCCAGACGAGAGTCTTTTCTAACTTGGGGATGCAGTCTCAGATACT  
GGGGGCCCTCTGCCTCATCTCCATTCTTACACCTCAAATCTGTACAGTTACTTTTGTACTGACTG  
TAATAAACAGCCAAACAGTT
- [0749] SEQ ID NO:190-小家鼠Cth 3' -UTR
- [0750] NM\_145953.2  
AAGTTCGAGTCAAAGCTGTCATTCCAGTGCTGCCATCAGGAGCAGCATCCAAGGGGCCAAATCTTC  
GGAATAACTGGACAGATCATTACGGAGCATACGCAGAACTGCACTGAATATTTTAAGGCCCTAATG  
AGTTTACAGCTGTAACTTCCATGGATCTTCCCTTAAGAACTGTCTTCTGTTTATCTTCTAACTAA  
CAGGTTGTTCTGTTAGTATCATTTTCGGTAATTTTGCTATATTTGTGTCCAAGGAAGTAAGAGTTGT  
TCTGTTTTGGGATCATGTTGTCTTTTTCTTTCTTCAGTAGCTTATGATATATTTTAATCATGTTT  
ACAATGTTACAATTTAGTATTGATGTTTTATGAAGTTAAATTATTAAATGAATGGTCTTAAATCCA  
CTGTGGTGGTTTTTTTTTTGAAAAATTATATAATTACCATAAGCCAAAAATCAAATATTTGGAATAC  
CTACTGTGAAATTCAAGAGATTAAAGGTTGTACTTGATACTTGTTATTTTTCTTAAATAAATCTAG  
TTATCAAGTTATTTTTCTTAAATAAATCTAGTTATCAAATGT
- [0751]
- [0752] SEQ ID NO:191-小家鼠Pole 3' -UTR
- [0753] NM\_011132.2  
CAAGCCTAGGCTAAAGACACTTTGGTATCCCACACCTACTGCCTGCTCCAAAAGGCAGAACCACTG  
ACCACCTTGCTTTTCCAACTCATGAGCACAGCCCAGAAGGAACAGAAGACTTCTGCTAACGTCAT
- [0754] CATGCCATAAACAGACAGAAGCAGGGAATGGCTCTATCCCTAGCTGCCTGCTAAGTAAACACGGTT  
TTGAAGCGTC
- [0755] SEQ ID NO:192-小家鼠Kif2c 3' -UTR
- [0756] NM\_134471.4  
TGACTTCAAATAAAGATCTGTTTGATATCACACTAGCCTCTTCCCCTCCCCAGAAAACCTTTGGGCA  
CCTTCTGGCTTTGGTTAGGAAGTGGGAGGTTGGGTAAATTTCAACCTCAGGGCATCATGC  
CCAGGAAAGCTGGGGAGAGGCGATGTTCTTTTCTCAGTTTGGAGCGGGAAGAAGGAGTTTAAATG  
ACTGTGCTTCCTCTTCTCCTCCCCATGGAAGGGGGGGGGATCCCTGGCTCAGCATCCACTGCTGG  
TGGAAACCTCGCTGCGAGGTGCCGCTGTGAGGTGTGCCGCTGTGAGGTGCGTGCTCTGGAAGAGAG  
GCGGAGCTTGTCGCTTCCTGCTGCCTCTGCCTGCAGTACCACCCTCGGTCTAAGCACTCATGATT  
TTATACTTTTCAAAAAAGTTTTAAGACTTCTTTGGGTTTTTTGTTTTTGAAGCCTAAGAA  
AGTCACTAGTTCTTGCCTTTATGTATTTATATGCTGTACCTAACAATAAAGAAAGAAGAAAAAAA  
CAAA
- [0757]
- [0758] SEQ ID NO:193-小家鼠Scand1 3' -UTR
- [0759] NM\_020255.3  
GCGGCGGAGCCTGGGGCTCCCTGGACTGACACCCTACCCCTGAGTTAATAAAAGTTGAATTTTGAC
- [0760] AGCTCCCA
- [0761] SEQ ID NO:194-小家鼠Gpr84 3' -UTR

- [0762] NM\_030720.1  
AGCTATTTAAAGCTAGTAGTCCATTCACCAGGACAGGCCAAACATCCGGAACCAGAGTGGCCTGCA  
GAGGACAGGACAGGAGCCCTTCAGTCCTTGGGTATTTACAGACAACCTCAGTGGTATAGAGGTAC
- [0763]  
ACCACTTTCCCATTCAGGAATCAGTGCCTCAACCCTGTGTGACCCAAGGAGTGTGGTTAATTATTA  
ATAAAGACATTGCATCCCCCCTC
- [0764] SEQ ID NO:195-小家鼠Tpgs1 3' -UTR
- [0765] NM\_148934.2  
GACCACGATGCACCCACATCCCGCCACCTGCCCAGGACTTGGGACCTGCAAACCTGATGTCCCGGCA
- [0766] TTCTGCACAGGGTCATCACCCAGGAAGGGGTACCCGCTTGGACTTTGCAGGACCAGCACCTGCAGC  
TCAATCCCCTCTGTACTGGCCTGCCTACCCCACTCACAGAAATAAAATCTAAACGAC
- [0767] SEQ ID NO:196-小家鼠Cc117 3' -UTR
- [0768] NM\_011332.3  
CCTTCCCGCTGAGGCATTTGGAGACGCCAGGGCTGCTGTCCATGGTTTCAACATAAAACGGCCTGT
- [0769] GACCAGCAGAGCCCAAGAGCAGCCACAGAGCAGAAGTCCCTGTTCCCTTTTTTATGGACTCTTATG  
CACTACAGGCCGAACACAAAAAAGCAACGGAATAAAGCCTTCCTCCCTC
- [0770] SEQ ID NO:197-小家鼠Alkbh7 3' -UTR
- [0771] CCTCCCACATATCCTCCATATTTGTGAATAAAGTAGGGAAAATGCTGTCAT
- [0772] SEQ ID NO:198-小家鼠Ms4a8a 3' -UTR
- [0773] NM\_022430.2  
GTGAACCTGAAGATTCTAGAGACCAAGTGACATCCTCTCCTACCTAGACTCCTATAAACCAAGTTC  
TTCCTTTCTGACGAAGGGTAAATATCTTTCTTGTGGCCTAAATTATAGACTCTTGCTTCAACTCA
- [0774] CCCTGGAAAAATCTCTATTAACGAGATGGGAGATTGAAATGGATTCAAATAAAGATGCTCTAGC  
AGGA
- [0775] SEQ ID NO:199-小家鼠Mrp134 3' -UTR
- [0776] NM\_053162.2  
GGAGCGCAACGTCGCAGGGAATAGACTCAGACGCCGGTAGAGTTGCCCAGGGCGAGAAGTTGGGG  
CAAGCTGAAAGTGGGGGTCTGTCCCTTCGTAGGTTTTTAGACGCATCCCCGTCGACGGCTGCTTA
- [0777] CCAACCACCAGCACCTGCGATACCTCATCCATTTCTGTGAATATATTTGAAGGCATCTTTGCAAGA  
CGGGGTCATTAAACACAGACACCACCCCTACGTTTCCAGCCATTCAAAGTATTACCGACTTAAT  
AAATATTCACGTTTCTTGAACCACT
- [0778] SEQ ID NO:200-小家鼠Comtd1 3' -UTR  
GGTTAACACGAAGCTTAGGGCCTGGGTGTGAGAACCTAAGACCCCCAGCCAGTGACCTGACTTTTA
- [0779] AACCTGACAATAAAGGTACTGGACACGCG
- [0780] SEQ ID NO:201-小家鼠Armc6 3' -UTR
- [0781] NM\_133972.2

CCTTAGCTCAGAGAACTCTGACTCCAGAACGTGGGTGGCTATGGGGCCCTGCCTCTCTGTCTCCTCCT  
TCCCATGCTACCCAGAGCACAAAGTGTGCTAGCAGGGCAGGGGGTGGGGAGGGCAGAAGCAGTGATG  
ACCTGGTGGCCCACCACTGCCAGTGTCCCTTCTACCACAGCTGCTCTCTGCAGCTTTGGGACAAAG  
[0782] GGCACATGGAAGTGCAGTGTCTGTCTGGAGTTTGAGGCCTCCACAAGCTGGTGTCCCAGCAGGGC  
CAGCCTGCAATCCGGGCACCACTCCAAGCAGTTGATAACCTGCTCCTTGGGCAGGCCCCCTGGGT  
AGAGGTGCTTCAGCCAAGGCTGATTCCCCAAGATCCGCATGATGGGAGGAGGGAGGTGTCTCCTCC  
AAATTGTCCTCTGAACTCTGTATGTGTGCTGTGGTGCACACATACATGCTAGCTAAATAAACAAAC  
AAACAAATAA

[0783] SEQ ID NO:202-小家鼠Atp5d 3' -UTR

[0784] NM\_025313.2

GTGGTACCTACTGTCTGACACCCACGGGGAACTGAGCCAGGTCCAGGCCGATGAGAAGTTCCCAG  
TGGGCTGAAGTGGCCACCAGGGGTCAGCAGTGCTCCAGTTGCTGGGCTTAAAGCTTCTGGTGCCT  
[0785] GTCTGCCAGGTCATGGAGGATTCCCCAATCTGGCATCCCCACGATGCCTCTGGAGAGATGGCCTTG  
ATTGCCCCCTCAAAGCCACCTGAACCGTCGTCAACTTACCCAGCCTGTCTCCATTAAACACCAGGAA  
CCAACTGAG

[0786] SEQ ID NO:203-小家鼠1110001J03Rik 3' -UTR

[0787] NM\_025363.3

TGCAGAGAGTCCTCAGATGTTTCCTTCATTCAAGAGTTTAACCATTCTAACAATATGTAGTTATCA  
[0788] TTAAATCTTTTTTAAAGTGTG

[0789] 这些野生型3' -UTR元件中的一些与NCBI GenBank鉴定的公众可获得的3' -UTR序列不同,参见表2(实施例1)。该表格反映了本发明人的测序结果。

[0790] 由此,本发明提供了选自由下述组成的组的3' -UTR元件:SEQ ID NO:179,SEQ ID NO:180,SEQ ID NO:182,SEQ ID NO:183,SEQ ID NO:197,SEQ ID NO:200,并且可以用于本发明的任一方面。

[0791] 用于本发明的人工3' -UTR元件的实例

[0792] 在一些实施方案中,本发明所述的3' -UTR元件与野生型3' -UTR元件不同。此类3' -UTR元件称为“人工3' -UTR元件”。

[0793] 优选地,人工3' -UTR元件与野生型3' -UTR元件表现出一定程度的序列同一性,所述序列同一性(a)小于100%,同时(B)大于10%,大于20%,大于30%,大于40%,大于50%,大于60%,大于70%,大于80%,大于90%,大于95%,所述野生型3' -UTR元件选自由SEQ ID NOs:152-203组成的组。典型地,所述人工3' -UTR元件与作为其基础的野生型3' -UTR元件不同,其中至少一个核苷酸,如两个核苷酸,三个核苷酸,四个核苷酸,五个核苷酸,六个核苷酸,七个核苷酸,八个核苷酸,九个核苷酸,十个核苷酸,或多于十个核苷酸被交换(exchanged)。例如,在野生型3' -UTR元件包含认为是不利的核苷酸元件的情形中,所述核苷酸交换可以是推荐的。例如,在一些实施方案中,认为是不利的核苷酸元件是限制性酶识别位点(切割位点),特别是由用于制备(克隆)本发明的人工核酸的工序的限制性酶识别(可切割)的限制性酶识别位点(切割位点)。因此,可以特异性引入特定的碱基(对于各个野生型碱基,以交换,优选置换而引入),以使所述人工3' -UTR元件不包含认为是不利的核苷酸元件。在具体的实施方案中,所述人工3' -UTR元件是SEQ ID NO:204:

[0794] SEQ ID NO:204-人工序列(基于SEQ ID NO:192)

TGACTTCAAATAATGATCTGTTTGATATCACACTAGCCTCTTCCCCTCCCCAGAAAACCTTTGGGCA  
CCTTCTGGCTTTGGTTAGGAAGTGGAGTGGACAGGTTGGGTAAATTTCAACCTCAGGGCATCATGC  
CCAGGAAAGCTGGGGAGAGGCGATGTTCTTTTCTCAGTTTGGAGCGGGAAGAAGGAGTTTAAATG  
ACTGTGCTTCTCTTCTCCTCCCCATGGAAGGGGGGGGGATCCCTGGCTCAGCATCCACTGCTGG  
[0795] TGGAAACCTCGCTGCGAGGTGCCGCTGTGAGGTGTGCCGCTGTGAGGTGCGTGCTCTGGAAGAGAG  
GCGGAGCTTGTCCGCTTCTGCTGCCTCTGCCTGCAGTACCACCTCGGTCTAAGCACTCATGATT  
TTATACTTTCAGAAAAAGTTTTAAGACTTCTTTGGGTTTTTGTGTTTTGTTTGAAGCCTAAGAA  
AGTCACTAGTTCTTGCCTTTATGTATTTATATGCTGTACCTAACAATAAAGAAAGAAGAAAAAAA  
CAAA

[0796] [该序列对应作为其基础的野生型序列,不同之处在于,AGATCT被TGATCT替换,上述用下划线表示强调]

[0797] 新型的5'-UTR元件和新型的3'-UTR元件

[0798] 本发明还提供新型的5'-UTR元件和新型的3'-UTR元件,即本发明人发现分别在人细胞和小鼠细胞中表达但是不被公共数据库所知晓的5'-UTR元件和3'-UTR元件(见实施例1)。在本发明的一些实施方案中,选自由SEQ ID NO:45、SEQ ID NO:47、SEQ ID NO:49组成的组的任意5'-UTR可以是优选的。在本发明的一些实施方案中,选自由SEQ ID NO:77,SEQ ID NO:79,SEQ ID NO:84,SEQ ID NO:85,SEQ ID NO:90,SEQ ID NO:92,SEQ ID NO:93,SEQ ID NO:98,SEQ ID NO:106,SEQ ID NO:109,SEQ ID NO:115,SEQ ID NO:118,SEQ ID NO:125,SEQ ID NO:127,SEQ ID NO:131,SEQ ID NO:135组成的组的任意5'-UTR可以是优选的。在本发明的一些实施方案中,选自由SEQ ID NO:179,SEQ ID NO:180,SEQ ID NO:182,SEQ ID NO:183,SEQ ID NO:197,SEQ ID NO:200组成的组的任意3'-UTR可以是优选的。

[0799] 5'-UTR元件和优选的3'-UTR元件的说明

[0800] 优选地,本发明所述的人工核酸分子的至少一个3'-UTR元件包含这样的核酸序列或由这样的核酸序列组成:所述核酸序列与选自由以下各项组成的组的基因的转录本的3'-UTR序列具有至少约40%,优选至少约50%,优选至少约60%,优选至少约70%,更优选至少约80%,更优选至少约90%,甚至更优选至少约95%,甚至更优选至少约99%,最优选100%的同一性:NDUFS7-3'-UTR,PHGDH-3'-UTR,TSPO-3'-UTR,ATP5D-3'-UTR,EXOSC4-3'-UTR,TUBB4B-3'-UTR,TUBA4A-3'-UTR,EMP3-3'-UTR,CRIP2-3'-UTR,BRAT1-3'-UTR,PSMB3-3'-UTR,PSMB6-3'-UTR,SCAND1-3'-UTR,AMN-3'-UTR,CYBA-3'-UTR,PCOLCE-3'-UTR,MAP1S-3'-UTR,VTN-3'-UTR,ALDH16A1-3'-UTR(均为人的),Acox2-3'-UTR,Ubc-3'-UTR,Slpi-3'-UTR,Igf2bp1-3'-UTR,Tmem37-3'-UTR,Slc7a3-3'-UTR,Cst6-3'-UTR,Ebp-3'-UTR,Sf3b5-3'-UTR,Cdca8-3'-UTR,Kif22-3'-UTR,Cad-3'-UTR,Pole-3'-UTR,Kif2c-3'-UTR,Scand1-3'-UTR,Gpr84-3'-UTR,Tpgs1-3'-UTR,Cc117-3'-UTR,Alkbh7-3'-UTR,Ms4a8a-3'-UTR,Mrp134-3'-UTR,Comtd1-3'-UTR,Armc6-3'-UTR,Atp5d-3'-UTR,1110001J03Rik-3'-UTR,Nudt22-3'-UTR(均为小鼠的)。更优选地,本发明所述的人工核酸分子的至少一个3'-UTR元件包含这样的核酸序列或由这样的核酸序列组成:所述核酸序列与分别选自由下列各项或相应RNA序列组成的组的序列具有至少约40%,优选至少约50%,优选至少约60%,优选至少约70%,更优选至少约80%,更优选至少约90%,甚至更优选至少约95%,甚至更优选至少约99%,最优选100%的同一性:SEQ ID NO:152,SEQ ID NO:153,SEQ ID NO:154,SEQ ID NO:155,SEQ ID NO:156,SEQ ID NO:157,SEQ ID NO:158,SEQ ID NO:159,SEQ ID NO:160,

SEQ ID NO:161,SEQ ID NO:164,SEQ ID NO:165,SEQ ID NO:167,SEQ ID NO:168,SEQ ID NO:169,SEQ ID NO:170,SEQ ID NO:171,SEQ ID NO:172,SEQ ID NO:173;SEQ ID NO:174,SEQ ID NO:175,SEQ ID NO:176,SEQ ID NO:179,SEQ ID NO:180,SEQ ID NO:181,SEQ ID NO:182,SEQ ID NO:183,SEQ ID NO:184,SEQ ID NO:187,SEQ ID NO:188,SEQ ID NO:189,SEQ ID NO:191,SEQ ID NO:204(或为SEQ ID NO:192),SEQ ID NO:193,SEQ ID NO:194,SEQ ID NO:195,SEQ ID NO:196,SEQ ID NO:197,SEQ ID NO:198,SEQ ID NO:199,SEQ ID NO:200,SEQ ID NO:201,SEQ ID NO:202,SEQ ID NO:203,SEQ ID NO:177。

[0801] 优选地,本发明所述的人工核酸分子的至少一个5'-UTR元件包含这样的核酸序列或由这样的核酸序列组成:所述核酸序列与以下各项的转录本的5'-UTR序列具有至少约40%,优选至少约50%,优选至少约60%,优选至少约70%,更优选至少约80%,更优选至少约90%,甚至更优选至少约95%,甚至更优选至少约99%,最优选100%的同一性:ZNF460-5'-UTR,TGM2-5'-UTR,IL7R-5'-UTR,COL8A1-5'-UTR,NDUFS7-5'-UTR,PLK2-5'-UTR,FBX032-5'-UTR,ATP5D-5'-UTR,TUBB4B-5'-UTR,ORMDL2-5'-UTR,FSCN1-5'-UTR,CD9-5'-UTR,PYSL2-5'-UTR,PSMB3-5'-UTR,PSMB6-5'-UTR,KPNA6-5'-UTR,SFT2D2-5'-UTR,LCLAT1-5'-UTR,FBXL18-5'-UTR,SLC35F6-5'-UTR,VMA21-5'-UTR,SEZ6L2-5'-UTR,PCOLCE-5'-UTR,VTN-5'-UTR,ALDH16A1-5'-UTR,KPNA6-5'-UTR,JUP-5'-UTR,CPN2-5'-UTR,PNPO-5'-UTR,SSSCA1-5'-UTR,POLR2L-5'-UTR,LIN7C-5'-UTR,UQCR10-5'-UTR,PYCRL-5'-UTR,AMN-5'-UTR,MAP1S-5'-UTR(均为人的),Dpysl2-5'-UTR,Acox2-5'-UTR,Ubc-5'-UTR,Nudt22-5'-UTR,Pcyox11-5'-UTR,Ankrd1-5'-UTR,Tspyl4-5'-UTR,Slc7a3-5'-UTR,Aacs-5'-UTR,Nosip-5'-UTR,Itga7-5'-UTR,Ccnd2-5'-UTR,Ebp-5'-UTR,Sf3b5-5'-UTR,Fasn-5'-UTR,Hmgcs1-5'-UTR,Osrl-5'-UTR,Lmnb1-5'-UTR,Vma21-5'-UTR,Kif20a-5'-UTR,Cdca8-5'-UTR,Slc7a1-5'-UTR,Ubqln2-5'-UTR,Prps2-5'-UTR,Shmt2-5'-UTR,Fignl1-5'-UTR,Cad-5'-UTR,Anln-5'-UTR,Slfn9-5'-UTR,Ncaph-5'-UTR,Pole-5'-UTR,Uhrfl-5'-UTR,Gja1-5'-UTR,Fam64a-5'-UTR,Tspan10-5'-UTR,Scand1-5'-UTR,Gpr84-5'-UTR,Cers6-5'-UTR,Cxcr4-5'-UTR,Gprc5c-5'-UTR,Fen1-5'-UTR,Cspg4-5'-UTR,Mrpl34-5'-UTR,Comtd1-5'-UTR,Arm6c-5'-UTR,Emr4-5'-UTR,Atp5d-5'-UTR,Csf2ra-5'-UTR,Aarsd1-5'-UTR,Cth-5'-UTR,Tpgs1-5'-UTR,Cc117-5'-UTR,Alkbh7-5'-UTR,Ms4a8a-5'-UTR(均为小鼠的)。最优选地,本发明所述的人工核酸分子的至少一个5'-UTR元件包含这样的核酸序列或由这样的核酸序列组成:所述核酸序列与以下各项的序列分别具有至少约40%,优选至少约50%,优选至少约60%,优选至少约70%,更优选至少约80%,更优选至少约90%,甚至更优选至少约95%,甚至更优选至少约99%,最优选100%的同一性:SEQ ID NO:1,SEQ ID NO:2,SEQ ID NO:137(或为SEQ ID NO:3),SEQ ID NO:138(或为SEQ ID NO:9),SEQ ID NO:10,SEQ ID NO:12,SEQ ID NO:15,SEQ ID NO:140(或为SEQ ID NO:17),SEQ ID NO:20,SEQ ID NO:22,SEQ ID NO:23,SEQ ID NO:143(或为SEQ ID NO:31),SEQ ID NO:32,SEQ ID NO:39,SEQ ID NO:41,SEQ ID NO:44,SEQ ID NO:45,SEQ ID NO:146(或为SEQ ID NO:52),SEQ ID NO:53,SEQ ID NO:54,SEQ ID NO:57,SEQ ID NO:60,SEQ ID NO:61,SEQ ID NO:63,SEQ ID NO:64,SEQ ID NO:66,SEQ ID NO:68,SEQ ID NO:69,SEQ ID NO:72,SEQ ID NO:34,SEQ ID NO:35,SEQ ID NO:36,SEQ ID NO:37,SEQ ID NO:144(或为SEQ ID NO:42),SEQ ID NO:58,SEQ ID NO:62;SEQ ID NO:73,SEQ ID NO:75,SEQ ID NO:77,SEQ ID NO:79,SEQ ID NO:

80,SEQ ID NO:81,SEQ ID NO:83,SEQ ID NO:84,SEQ ID NO:86,SEQ ID NO:87,SEQ ID NO:88,SEQ ID NO:89,SEQ ID NO:90,SEQ ID NO:91,SEQ ID NO:92,SEQ ID NO:93,SEQ ID NO:94,SEQ ID NO:95,SEQ ID NO:97,SEQ ID NO:98,SEQ ID NO:99,SEQ ID NO:100,SEQ ID NO:101,SEQ ID NO:102,SEQ ID NO:103,SEQ ID NO:106,SEQ ID NO:107,SEQ ID NO:108,SEQ ID NO:147(或为SEQ ID NO:109),SEQ ID NO:148(或为SEQ ID NO:110),SEQ ID NO:112,SEQ ID NO:113,SEQ ID NO:114,SEQ ID NO:115,SEQ ID NO:117,SEQ ID NO:118,SEQ ID NO:149(或为SEQ ID NO:119),SEQ ID NO:123,SEQ ID NO:126,SEQ ID NO:127,SEQ ID NO:128,SEQ ID NO:129,SEQ ID NO:130,SEQ ID NO:131,SEQ ID NO:132,SEQ ID NO:133,SEQ ID NO:134,SEQ ID NO:151(或为SEQ ID NO:136),SEQ ID NO:96,SEQ ID NO:111,SEQ ID NO:150(基于SEQ ID NO:120),SEQ ID NO:121,SEQ ID NO:124,SEQ ID NO:125,或相对应的RNA序列。

[0802] 本发明所述的人工核酸分子的至少一个3'-UTR元件还可以包含这样的核酸序列的片段或由这样的核酸序列的片段组成:所述核酸序列的片段与基因的转录本的3'-UTR的核酸序列,如与SEQ ID NOs:152至204的序列的3'-UTR具有至少约40%,优选至少约50%,优选至少约60%,优选至少约70%,更优选至少约80%,更优选至少约90%,甚至更优选至少约95%,甚至更优选至少约99%,最优选100%的同一性,其中所述片段优选是如上文所述的功能性片段或功能性变体片段。所述片段优选呈现至少约3个核苷酸,优选至少约5个核苷酸,更优选至少约10,15,20,25或30个核苷酸,甚至更优选至少约50个核苷酸,最优选至少约70个核苷酸的长度。在优选的实施方案中,片段或其变体呈现3至约500个核苷酸,优选5至约150个核苷酸,更优选10至100个核苷酸,甚至更优选15至90个,最优选20至70个的长度。优选地,所述变体、片段或变体片段是3'-UTR的功能性变体、功能性片段或功能性变体片段,以由包含选自由下列各项组成的组的核酸序列的人工核酸分子呈现的蛋白产生延长效率的至少30%的效率,优选至少40%的效率,更优选至少50%,更优选至少60%,甚至更优选至少70%,甚至更优选至少80%,最优选至少90%的效率延长从本发明所述的人工核酸分子的蛋白产生:SEQ ID NO:152,SEQ ID NO:153,SEQ ID NO:154,SEQ ID NO:155,SEQ ID NO:156,SEQ ID NO:157,SEQ ID NO:158,SEQ ID NO:159,SEQ ID NO:160,SEQ ID NO:161,SEQ ID NO:164,SEQ ID NO:165,SEQ ID NO:167,SEQ ID NO:168,SEQ ID NO:169,SEQ ID NO:170,SEQ ID NO:171,SEQ ID NO:172,SEQ ID NO:173;SEQ ID NO:174,SEQ ID NO:175,SEQ ID NO:176,SEQ ID NO:179,SEQ ID NO:180,SEQ ID NO:181,SEQ ID NO:182,SEQ ID NO:183,SEQ ID NO:184,SEQ ID NO:187,SEQ ID NO:188,SEQ ID NO:189,SEQ ID NO:191,SEQ ID NO:204(或为SEQ ID NO:192),SEQ ID NO:193,SEQ ID NO:194,SEQ ID NO:195,SEQ ID NO:196,SEQ ID NO:197,SEQ ID NO:198,SEQ ID NO:199,SEQ ID NO:200,SEQ ID NO:201,SEQ ID NO:202,SEQ ID NO:203,SEQ ID NO:177。

[0803] 本发明所述的人工核酸分子的至少一个5'-UTR元件还可以包含这样的核酸序列的片段或由这样的核酸序列的片段组成,所述核酸序列的片段与基因的转录本的5'-UTR的核酸序列,如与SEQ ID NO:1至151的序列的5'-UTR具有至少约40%,优选至少约50%,优选至少约60%,优选至少约70%,更优选至少约80%,更优选至少约90%,甚至更优选至少约95%,甚至更优选至少约99%,最优选100%的同一性,其中所述片段优选是如上文所述的功能性片段或功能性变体片段。所述片段优选呈现至少约3个核苷酸,优选至少约5个核苷

酸,更优选至少约10,15,20,25或30个核苷酸,甚至更优选至少约50个核苷酸,最优选至少约70个核苷酸的长度。在优选的实施方案中,片段或其变体呈现3至约500个核苷酸,优选5至约150个核苷酸,更优选10至100个核苷酸,甚至更优选15至90个,最优选20至70个的长度。优选地,所述变体、片段或变体片段是5'-UTR的功能性变体、功能性片段或功能性变体片段,以由包含选自下列各项组成的组的核酸序列的人工核酸分子呈现的蛋白产生效率增加效率的至少30%的效率,优选以至少40%的效率,更优选至少50%,更优选至少60%,甚至更优选至少70%,甚至更优选至少80%,最优选至少90%的效率增加从本发明所述的人工核酸分子的蛋白产生:SEQ ID NO:1,SEQ ID NO:2,SEQ ID NO:137(或为SEQ ID NO:3),SEQ ID NO:138(或为SEQ ID NO:9),SEQ ID NO:10,SEQ ID NO:12,SEQ ID NO:15,SEQ ID NO:140(或为SEQ ID NO:17),SEQ ID NO:20,SEQ ID NO:22,SEQ ID NO:23,SEQ ID NO:143(或为SEQ ID NO:31),SEQ ID NO:32,SEQ ID NO:39,SEQ ID NO:41,SEQ ID NO:44,SEQ ID NO:45,SEQ ID NO:146(或为SEQ ID NO:52),SEQ ID NO:53,SEQ ID NO:54,SEQ ID NO:57,SEQ ID NO:60,SEQ ID NO:61,SEQ ID NO:63,SEQ ID NO:64,SEQ ID NO:66,SEQ ID NO:68,SEQ ID NO:69,SEQ ID NO:72,SEQ ID NO:34,SEQ ID NO:35,SEQ ID NO:36,SEQ ID NO:37,SEQ ID NO:144(或为SEQ ID NO:42),SEQ ID NO:58,SEQ ID NO:62;SEQ ID NO:73,SEQ ID NO:75,SEQ ID NO:77,SEQ ID NO:79,SEQ ID NO:80,SEQ ID NO:81,SEQ ID NO:83,SEQ ID NO:84,SEQ ID NO:86,SEQ ID NO:87,SEQ ID NO:88,SEQ ID NO:89,SEQ ID NO:90,SEQ ID NO:91,SEQ ID NO:92,SEQ ID NO:93,SEQ ID NO:94,SEQ ID NO:95,SEQ ID NO:97,SEQ ID NO:98,SEQ ID NO:99,SEQ ID NO:100,SEQ ID NO:101,SEQ ID NO:102,SEQ ID NO:103,SEQ ID NO:106,SEQ ID NO:107,SEQ ID NO:108,SEQ ID NO:147(或为SEQ ID NO:109),SEQ ID NO:148(或为SEQ ID NO:110),SEQ ID NO:112,SEQ ID NO:113,SEQ ID NO:114,SEQ ID NO:115,SEQ ID NO:117,SEQ ID NO:118,SEQ ID NO:149(或为SEQ ID NO:119),SEQ ID NO:123,SEQ ID NO:126,SEQ ID NO:127,SEQ ID NO:128,SEQ ID NO:129,SEQ ID NO:130,SEQ ID NO:131,SEQ ID NO:132,SEQ ID NO:133,SEQ ID NO:134,SEQ ID NO:151(或为SEQ ID NO:136),SEQ ID NO:96,SEQ ID NO:111,SEQ ID NO:150(基于SEQ ID NO:120),SEQ ID NO:121,SEQ ID NO:124,SEQ ID NO:125。

#### [0804] 进一步优选的实施方案

[0805] 优选地,本发明所述的人工核酸分子的所述至少一个3'-UTR元件和/或所述至少一个5'-UTR元件呈现至少约3个核苷酸,优选至少约5个核苷酸,更优选至少约10,15,20,25或30个核苷酸,甚至更优选至少约50个核苷酸,最优选至少约70个核苷酸的长度。所述至少一个3'-UTR元件和/或所述至少一个5'-UTR元件的长度上限可以是500个核苷酸或更少,例如400,300,200,150或100个核苷酸。对于其他实施方案,上限可以在50至100个核苷酸的范围内选择。例如,片段或其变体可以呈现3至约500个核苷酸,优选5至约150个核苷酸,更优选10至100个核苷酸,甚至更优选15至90个,最优选20至70个的长度。

[0806] 还优选的是,包含在所述人工核酸分子中的UTR元件包含(或由其组成)短于本文提供的选自SEQ ID NOs:1-204的任一序列但是仍然与选自SEQ ID NOs:1-204的任意序列的各个序列具有70%以上、80%以上、90%以上或95%以上的同一性程度的连续序列。在所述情形中,优选地,替换UTR元件包含在其整个长度上与选自SEQ ID NOs:1-204的任一序列相同的连续序列;然而,替换UTR元件较短,例如,具有的长度为其另外与之相同的且选自

SEQ ID NOs:1-204的序列的全长的70%以上、80%以上、90%以上或95%以上。

[0807] 此外,本发明所述的人工核酸分子可以包含多于一个3'-UTR元件和/或多于一个5'-UTR元件(如上文所述的)。例如,本发明所述的人工核酸分子可以包含一个、两个、三个、四个或更多个3'-UTR元件,和/或一个、两个、三个、四个或更多个5'-UTR元件,其中个体3'-UTR元件可以相同或它们可以不同,并且类似地,个体5'-UTR元件可以相同或它们可以不同。例如,本发明所述的人工核酸分子可以包含两个基本上相同的如上文所述的3'-UTR元件,例如两个3'-UTR元件包含这样的核酸序列的或由这样的核酸序列的组成:所述核酸序列源自基因的转录本的3'-UTR,如源自SEQ ID NO:152至204的序列,或源自如上文所述的基因的转录本的3'-UTR的片段或变体,其功能性变体,其功能性片段,或其功能性变体片段。因此,例如,本发明所述的人工核酸分子可以包含两个基本上相同的如上文所述的5'-UTR元件,例如两个包含这样的核酸序列或由这样的核酸序列组成的5'-UTR元件:所述核酸序列源自基因的转录本的5'-UTR,如源自SEQ ID NO:1至151的序列,或源自基因的转录本的5'-UTR的片段或变体,其功能性变体,其功能性片段,或其功能性变体片段,如上文所述。

[0808] 出人意料地,本发明人发现包含如上文所述的3'-UTR元件和/或如上文所述的5'-UTR元件的人工核酸分子可以表示或可以提供mRNA分子,所述mRNA分子提供高翻译效率。由此,本文中所述的3'-UTR元件和/或本文中所述的5'-UTR元件可以提高mRNA分子的翻译效率。

[0809] 尤其是,本发明所述的人工核酸分子可以包含(i)提供高翻译效率的至少一个3'-UTR元件和至少一个5'-UTR元件;(ii)提供高翻译效率的至少一个3'-UTR元件,但没有提供高翻译效率的5'-UTR元件;或(iii)提供高翻译效率的至少一个5'-UTR元件,但没有提供高翻译效率的3'-UTR元件。

[0810] 然而,尤其是在情况(ii)和(iii)中,但可能也在情况(i)中,本发明所述的人工核酸分子还可以包含一个或多个“其他3'-UTR元件和/或5'-UTR元件”,即不满足如上文所述的需要的3'-UTR元件和/或5'-UTR元件。例如,包含本发明所述的3'-UTR元件(即为所述人工核酸分子提供高翻译效率的3'-UTR元件)的本发明所述的人工核酸分子,可以另外包含任意另外的3'-UTR和/或任意另外的5'-UTR,尤其是另外的5'-UTR,例如5'-TOP UTR,或任意其他5'-UTR或5'-UTR元件。类似地,例如,包含本发明所述的5'-UTR元件(即为所述人工核酸分子提供高翻译效率的5'-UTR元件)的本发明所述的人工核酸分子,可以另外包含任意另外的3'-UTR和/或任意另外的5'-UTR,尤其是另外的3'-UTR,例如源自白蛋白基因的3'-UTR的3'-UTR,特别优选包含SEQ ID NO:206或207,尤其是SEQ ID NO:207的序列的3'-UTR,或任意其他3'-UTR或3'-UTR元件。

[0811] 如果除了提供高翻译效率的本发明的至少一个5'-UTR元件和/或本发明的至少一个3'-UTR元件以外,另外的3'-UTR(元件)和/或另外的5'-UTR(元件)存在于本发明所述的人工核酸分子中,则所述另外的5'-UTR(元件)和/或所述另外的3'-UTR(元件)可以与本发明的3'-UTR元件和/或本发明的5'-UTR元件相互作用,并且,由此,分别支持本发明的3'-UTR元件和/或本发明的5'-UTR元件的翻译效率效果。所述另外的3'-UTR和/或5'-UTR(元件)可以进一步支持稳定性和翻译效率。此外,如果二者(本发明的3'-UTR元件和本发明的5'-UTR元件)都存在于本发明所述的人工核酸分子中,本发明的5'-UTR元件和本发明的3'-UTR元件的翻译效率效果优选以协同的方式导致提供高翻译效率。



[0812] 优选地,所述另外的3'-UTR包含这样的核酸序列或由这样的核酸序列组成:所述核酸序列源自选自自由以下各项组成的组的基因的3'-UTR:白蛋白基因, $\alpha$ -珠蛋白基因, $\beta$ -珠蛋白基因,酪氨酸羟化酶基因,脂氧化酶基因,和胶原蛋白 $\alpha$ 基因,如胶原蛋白 $\alpha 1$ (I)基因,或源自选自自由以下各项组成的组的基因的3'-UTR的变体:白蛋白基因, $\alpha$ -珠蛋白基因, $\beta$ -珠蛋白基因,酪氨酸羟化酶基因,脂氧化酶基因,和胶原蛋白 $\alpha$ 基因,如胶原蛋白 $\alpha 1$ (I)基因,上述基因是根据专利申请W02013/143700的SEQ ID No.1369-1390,其公开内容通过引用结合于本文。在特别优选的实施方案中,所述另外的3'-UTR包含源自白蛋白基因,优选脊椎动物白蛋白基因,更优选哺乳动物白蛋白基因,最优选SEQ ID NO:206的人白蛋白基因的3'-UTR的核酸序列或由所述核酸序列组成:

[0813] SEQ ID NO:206:

[0814] CATCACATTT AAAAGCATCT CAGCCTACCA TGAGAATAAG AGAAAGAAAA TGAAGATCAA  
AAGCTTATTC ATCTGTTTTT CTTTTTCGTT GGTGTAAAGC CAACACCCTG TCTAAAAAAC  
ATAAATTTCT TTAATCATTT TGCCTCTTTT CTCTGTGCTT CAATTAATAA AAAATGGAAA  
GAATCT

[0815] (人白蛋白3'-UTR;对应于专利申请W0 2013/143700的SEQ ID No:1369)。

[0816] 在另一个特别优选的实施方案中,所述另外的3'-UTR包含下述核酸序列或由下述核酸序列组成:所述核酸序列源自 $\alpha$ -珠蛋白基因、优选脊椎动物的 $\alpha$ -或 $\beta$ -珠蛋白基因、更优选哺乳动物的 $\alpha$ -或 $\beta$ -珠蛋白基因、更优选专利申请W0 2013/143700中SEQ ID No.1370的人 $\alpha$ -或 $\beta$ -珠蛋白基因的3'-UTR(智人血红蛋白的3'-UTR, $\alpha 1$ (HBA1))或专利申请W0 2013/143700中SEQ ID No.1371的人 $\alpha$ -或 $\beta$ -珠蛋白基因的3'-UTR(智人血红蛋白的3'-UTR, $\alpha 2$ (HBA2))和/或专利申请W0 2013/143700中SEQ ID No.1372的人 $\alpha$ -或 $\beta$ -珠蛋白基因的3'-UTR(智人血红蛋白的3'-UTR, $\beta$ (HBB))。

[0817] 例如,所述另外的3'-UTR可以包含专利申请W0 2013/143700中SEQ ID No.1393的 $\alpha$ -珠蛋白基因的3'-UTR的中央 $\alpha$ -复合物-结合部分或由其组成。

[0818] 在该情况下特别优选的是,本发明的核酸分子包含源自根据专利申请W02013/143700的SEQ ID No.1369-1390的核酸或其片段、同源物或变体的另外的3'-UTR元件。

[0819] 最优选所述另外的3'-UTR包含源自SEQ ID NO:207的人白蛋白基因的片段的核酸序列:

[0820] SEQ ID NO:207:

[0821] CATCACATTTAAAAGCATCTCAGCCTACCATGAGAATAAGAGAAAGAAAATGAAGATCAATAGCTT  
ATTTCATCTCTTTTTCTTTTTTCGTTGGTGTAAAGCCAACACCCTGTCTAAAAAACATAAATTTCTTT  
AATCATTTTGCCTCTTTTCTCTGTGCTTCAATTAATAAAAAATGGAAAGAACCT

[0822] (白蛋白7 3'-UTR;对应于专利申请W0 2013/143700的SEQ ID No:1376)

[0823] 在该情况下特别优选的是,本发明的人工核酸分子的所述另外的3'-UTR包含SEQ ID NO:207的核酸序列或相应的RNA序列或由SEQ ID NO:207的核酸序列或相应的RNA序列组成。

[0824] 在一些实施方案中,所述另外的3'-UTR包含下述核酸序列或由下述核酸序列组成:所述核酸序列源自核糖体蛋白编码基因,优选如国际专利申请W0 2015/101414或W0 2015/101415所述,它们的公开内容通过引用结合在本文中。

[0825] 所述另外的3'-UTR还可以包含这样的核酸序列或由这样的核酸序列组成:所述核酸序列源自核糖体蛋白编码基因,其中另外的3'-UTR可以来源于的核糖体蛋白编码基因包

括,但不限于,核糖体蛋白L9 (RPL9),核糖体蛋白L3 (RPL3),核糖体蛋白L4 (RPL4),核糖体蛋白L5 (RPL5),核糖体蛋白L6 (RPL6),核糖体蛋白L7 (RPL7),核糖体蛋白L7a (RPL7A),核糖体蛋白L11 (RPL11),核糖体蛋白L12 (RPL12),核糖体蛋白L13 (RPL13),核糖体蛋白L23 (RPL23),核糖体蛋白L18 (RPL18),核糖体蛋白L18a (RPL18A),核糖体蛋白L19 (RPL19),核糖体蛋白L21 (RPL21),核糖体蛋白L22 (RPL22),核糖体蛋白L23a (RPL23A),核糖体蛋白L17 (RPL17),核糖体蛋白L24 (RPL24),核糖体蛋白L26 (RPL26),核糖体蛋白L27 (RPL27),核糖体蛋白L30 (RPL30),核糖体蛋白L27a (RPL27A),核糖体蛋白L28 (RPL28),核糖体蛋白L29 (RPL29),核糖体蛋白L31 (RPL31),核糖体蛋白L32 (RPL32),核糖体蛋白L35a (RPL35A),核糖体蛋白L37 (RPL37),核糖体蛋白L37a (RPL37A),核糖体蛋白L38 (RPL38),核糖体蛋白L39 (RPL39),大核糖体蛋白P0 (RPLP0),大核糖体蛋白P1 (RPLP1),大核糖体蛋白P2 (RPLP2),核糖体蛋白S3 (RPS3),核糖体蛋白S3A (RPS3A),核糖体蛋白S4,X-连锁的 (RPS4X),核糖体蛋白S4,Y-连锁的1 (RPS4Y1),核糖体蛋白S5 (RPS5),核糖体蛋白S6 (RPS6),核糖体蛋白S7 (RPS7),核糖体蛋白S8 (RPS8),核糖体蛋白S9 (RPS9),核糖体蛋白S10 (RPS10),核糖体蛋白S11 (RPS11),核糖体蛋白S12 (RPS12),核糖体蛋白S13 (RPS13),核糖体蛋白S15 (RPS15),核糖体蛋白S15a (RPS15A),核糖体蛋白S16 (RPS16),核糖体蛋白S19 (RPS19),核糖体蛋白S20 (RPS20),核糖体蛋白S21 (RPS21),核糖体蛋白S23 (RPS23),核糖体蛋白S25 (RPS25),核糖体蛋白S26 (RPS26),核糖体蛋白S27 (RPS27),核糖体蛋白S27a (RPS27a),核糖体蛋白S28 (RPS28),核糖体蛋白S29 (RPS29),核糖体蛋白L15 (RPL15),核糖体蛋白S2 (RPS2),核糖体蛋白L14 (RPL14),核糖体蛋白S14 (RPS14),核糖体蛋白L10 (RPL10),核糖体蛋白L10a (RPL10A),核糖体蛋白L35 (RPL35),核糖体蛋白L13a (RPL13A),核糖体蛋白L36 (RPL36),核糖体蛋白L36a (RPL36A),核糖体蛋白L41 (RPL41),核糖体蛋白S18 (RPS18),核糖体蛋白S24 (RPS24),核糖体蛋白L8 (RPL8),核糖体蛋白L34 (RPL34),核糖体蛋白S17 (RPS17),核糖体蛋白SA (RPSA),泛素A-52残基核糖体蛋白融合产物1 (UBA52),遍表达的Finkel-Biskis-Reilly鼠肉瘤病毒 (FBR-MuSV) (FAU),核糖体蛋白L22-样1 (RPL22L1),核糖体蛋白S17 (RPS17),核糖体蛋白L39-样 (RPL39L),核糖体蛋白L10-样 (RPL10L),核糖体蛋白L36a-样 (RPL36AL),核糖体蛋白L3-样 (RPL3L),核糖体蛋白S27-样 (RPS27L),核糖体蛋白L26-样1 (RPL26L1),核糖体蛋白L7-样1 (RPL7L1),核糖体蛋白L13a假基因 (RPL13AP),核糖体蛋白L37a假基因8 (RPL37AP8),核糖体蛋白S10假基因5 (RPS10P5),核糖体蛋白S26假基因11 (RPS26P11),核糖体蛋白L39假基因5 (RPL39P5),大核糖体蛋白P0假基因6 (RPLP0P6) 和核糖体蛋白L36假基因14 (RPL36P14)。

[0826] 优选地,所述另外的5' -UTR包含这样的核酸序列或由这样的核酸序列组成:所述核酸序列源自TOP基因的5' -UTR或源自TOP基因的5' -UTR的片段、同系物或变体。

[0827] 特别优选的是,5' -UTR元件不包含如上文所定义的TOP-基序或5' TOP。尤其是,优选的是TOP基因的5' -UTR是缺乏TOP基序的TOP基因的5' -UTR。

[0828] 源自TOP基因的5' -UTR的核酸序列源自真核TOP基因,优选植物或动物TOP基因,更优选脊索动物TOP基因,甚至更优选脊椎动物TOP基因,最优选哺乳动物TOP基因,如人TOP基因。

[0829] 例如,所述另外的5' -UTR优选选自包含这样的核酸序列或由这样的核酸序列组成的5' -UTR元件:所述核酸序列源自选自以下各项组成的组的核酸序列:专利申请W02013/

143700 (其公开内容通过引用结合在本文中) 的SEQ ID NOs.1-1363, SEQ ID NO.1395, SEQ ID NO.1421和SEQ ID NO.1422, 专利申请W02013/143700的SEQ ID NOs.1-1363、SEQ ID NO.1395、SEQ ID NO.1421和SEQ ID NO.1422的同系物、其变体, 或优选对应的RNA序列。术语“专利申请W02013/143700的SEQ ID NOs.1-1363, SEQ ID NO.1395, SEQ ID NO.1421和SEQ ID NO.1422的同系物”是指与根据专利申请W02013/143700的SEQ ID NOs.1-1363, SEQ ID NO.1395, SEQ ID NO.1421和SEQ ID NO.1422的序列同源的除智人之外的其他物种的序列。

[0830] 在优选的实施方案中, 所述另外的5' -UTR包含这样的核酸序列或由这样的核酸序列组成: 所述核酸序列源自选自专利申请W02013/143700的SEQ ID NOs.1-1363、SEQ ID NO.1395、SEQ ID NO.1421和SEQ ID NO.1422, 专利申请W02013/143700的SEQ ID NOs.1-1363、SEQ ID NO.1395、SEQ ID NO.1421和SEQ ID NO.1422的同系物, 其变体或相应的RNA序列的核酸序列的从核苷酸位置5 (即位于序列中位置5的核苷酸) 延伸到紧接起始密码子 (位于序列的3' 末端) 的5' 的核苷酸位置, 例如紧接ATG序列的5' 的核苷酸位置的核酸序列。特别优选的是, 所述另外的5' -UTR源自选自专利申请W02013/143700的SEQ ID NOs.1-1363, SEQ ID NO.1395, SEQ ID NO.1421和SEQ ID NO.1422, 专利申请W02013/143700的SEQ ID NOs.1-1363, SEQ ID NO.1395, SEQ ID NO.1421和SEQ ID NO.1422的同系物, 其变体, 或相应RNA序列的核酸序列的从紧接5' TOP的3' 的核苷酸位置延伸到紧接起始密码子的5' 的核苷酸位置 (位于序列的3' 末端), 例如紧接ATG序列的5' 的核苷酸位置的核酸序列。

[0831] 在特别是优选的实施方案中, 所述另外的5' -UTR包含这样的核酸序列或由这样的核酸序列组成: 所述核酸序列源自编码核糖体蛋白的TOP基因的5' -UTR或编码核糖体蛋白的TOP基因的5' -UTR的变体。例如, 所述5' -UTR元件包含这样的核酸序列或由这样的核酸序列组成: 所述核酸序列源自根据专利申请W02013/143700的SEQ ID NOs:170,232,244,259,1284,1285,1286,1287,1288,1289,1290,1291,1292,1293,1294,1295,1296,1297,1298,1299,1300,1301,1302,1303,1304,1305,1306,1307,1308,1309,1310,1311,1312,1313,1314,1315,1316,1317,1318,1319,1320,1321,1322,1323,1324,1325,1326,1327,1328,1329,1330,1331,1332,1333,1334,1335,1336,1337,1338,1339,1340,1341,1342,1343,1344,1346,1347,1348,1349,1350,1351,1352,1353,1354,1355,1356,1357,1358,1359,或1360任一项的核酸序列的5' -UTR, 相应RNA序列, 其同系物, 或其变体, 如本文所述, 优选缺少5' -TOP基序。如上文所述的, 从位置5延伸至紧接ATG的5' 的核苷酸 (其位于序列的3' 末端) 的序列对应于所述序列的5' -UTR。

[0832] 优选地, 所述另外的5' -UTR包含这样的核酸序列或由这样的核酸序列组成: 所述核酸序列源自编码核糖体大蛋白 (RPL) 的TOP基因的5' -UTR或编码核糖体大蛋白 (RPL) 的TOP基因的5' -UTR的同系物或变体。例如, 5' -UTR元件包含这样的核酸序列或由这样的核酸序列组成: 所述核酸序列源自根据专利申请W02013/143700的SEQ ID NOs:SEQ ID NOs:67,259,1284-1318,1344,1346,1348-1354,1357,1358,1421和1422中的任一个的核酸序列的5' -UTR, 相应RNA序列, 其同系物, 或其变体, 如本文所述, 优选缺少5' TOP基序。

[0833] 在特别优选的实施方案中, 5' -UTR元件包含这样的核酸序列或由这样的核酸序列组成: 所述核酸序列源自大核糖体蛋白32基因 (优选源自脊椎动物大核糖体蛋白32 (L32) 基因, 更优选源自哺乳动物大核糖体蛋白32 (L32) 基因, 最优选源自人大核糖体蛋白32 (L32)

基因)的5'-UTR,或源自大核糖体蛋白32基因(优选源自脊椎动物大核糖体蛋白32(L32)基因,更优选源自哺乳动物大核糖体蛋白32(L32)基因,最优选源自人大核糖体蛋白32(L32)基因)的5'-UTR的变体,,其中优选所述另外的5'-UTR不包含所述基因的5' TOP。

[0834] 因此,在特别是优选的实施方案中,所述另外的5'-UTR包含这样的核酸序列或由这样的核酸序列组成:所述核酸序列与SEQ ID NO:208(缺少5'末端寡嘧啶束的人大核糖体蛋白32的5'-UTR:GGCGCTGCCTACGGAGGTGGCAGCCATCTCCTTCTCGGCATC(SEQ ID NO:208);对应于专利申请W02013/143700的SEQ ID NO.1368)的核酸序列或优选相应RNA序列具有至少约40%,优选至少约50%,优选至少约60%,优选至少约70%,更优选至少约80%,更优选至少约90%,甚至更优选至少约95%,甚至更优选至少约99%的同一性,或其中所述另外的5'-UTR包含这样的核酸序列的片段或由这样的核酸序列的片段组成:所述核酸序列的片段与SEQ ID NO:208的核酸序列或更优选相应RNA序列具有至少约40%,优选至少约50%,优选至少约60%,优选至少约70%,更优选至少约80%,更优选至少约90%,甚至更优选至少约95%,甚至更优选至少约99%的同一性,其中,优选地,所述片段如上文所述,即,是代表全长5'-UTR的至少20%等的一段连续的核苷酸。优选地,所述片段呈现至少约20个核苷酸或更多,优选至少约30个核苷酸或更多,更优选至少约40个核苷酸或更多的长度。优选地,所述片段是本文所述的功能性片段。

[0835] 在一些实施方案中,所述人工核酸分子包含另外的5'-UTR,其包含这样的核酸序列或由这样的核酸序列组成:所述核酸序列源自选自以下的脊椎动物TOP基因(如哺乳动物,例如人TOP基因)的5'-UTR:RPSA,RPS2,RPS3,RPS3A,RPS4,RPS5,RPS6,RPS7,RPS8,RPS9,RPS10,RPS11,RPS12,RPS13,RPS14,RPS15,RPS15A,RPS16,RPS17,RPS18,RPS19,RPS20,RPS21,RPS23,RPS24,RPS25,RPS26,RPS27,RPS27A,RPS28,RPS29,RPS30,RPL3,RPL4,RPL5,RPL6,RPL7,RPL7A,RPL8,RPL9,RPL10,RPL10A,RPL11,RPL12,RPL13,RPL13A,RPL14,RPL15,RPL17,RPL18,RPL18A,RPL19,RPL21,RPL22,RPL23,RPL23A,RPL24,RPL26,RPL27,RPL27A,RPL28,RPL29,RPL30,RPL31,RPL32,RPL34,RPL35,RPL35A,RPL36,RPL36A,RPL37,RPL37A,RPL38,RPL39,RPL40,RPL41,RPLP0,RPLP1,RPLP2,RPLP3,RPLP0,RPLP1,RPLP2,EEF1A1,EEF1B2,EEF1D,EEF1G,EEF2,EIF3E,EIF3F,EIF3H,EIF2S3,EIF3C,EIF3K,EIF3EIP,EIF4A2,PABPC1,HNRNPA1,TPT1,TUBB1,UBA52,NPM1,ATP5G2,GNB2L1,NME2,UQCRB或其同系物或变体,其中优选所述另外的5'-UTR不包含TOP-基序或所述基因的5' TOP,并且其中任选地所述另外的5'-UTR在其5'-末端以位于5'末端寡嘧啶束(TOP)下游位置1,2,3,4,5,6,7,8,9或10处的核苷酸起始,并且其中进一步任选地,所述源自TOP基因的5'-UTR的另外的5'-UTR在其3'-末端以位于其所来源的基因的起始密码子(A(U/T)G)的上游位置1,2,3,4,5,6,7,8,9或10处的核苷酸终止。

[0836] 在某些实施方案中,另外的5'-UTR包含这样的核酸序列或由这样的核酸序列组成:所述核酸序列源自国际专利申请W0 2016/107877中所述的5'-UTR,其公开内容通过引用结合在本文中。

[0837] 本发明所述的人工核酸分子可以是RNA,如mRNA或病毒RNA或复制子,DNA,如DNA质粒或病毒DNA,或可以是修饰的RNA或DNA分子。其可以提供为具有有义链和反义链的双链分子,例如,为具有有义链和反义链的DNA分子。

[0838] 本发明所述的人工核酸分子还可以包含任选地5'-帽。任选的5'-帽优选位于ORF

的5' ,更优选本发明所述的人工核酸分子内的至少一个5' -UTR或任意另外的5' -UTR的5' 。

[0839] 优选地,本发明所述的人工核酸分子还包含聚腺苷酸(poly(A))序列和/或聚腺苷酸化信号。优选地,任选的聚腺苷酸序列位于至少一个3' -UTR元件或任意另外的3' -UTR的3' ,更优选任选的聚腺苷酸序列与3' -UTR元件的3' -末端连接。连接可以是直接或间接的,例如,通过一段2,4,6,8,10,20个等的核苷酸,如通过例如包含一个以上限制性位点或由一个以上限制性位点组成的1-50,优选1-20个核苷酸的接头。然而,即使本发明所述的人工核酸分子不包含3' -UTR,例如如果其仅包含至少一个5' -UTR元件,其优选仍然包含聚腺苷酸序列和/或聚腺苷酸化信号。

[0840] 在一个实施方案中,所述任选的聚腺苷酸化信号位于3' -UTR元件的3' 的下游。优选地,所述聚腺苷酸化信号包含共有序列NN(U/T)ANA,其中N=A或U,优选AA(U/T)AAA或A(U/T)(U/T)AAA。此种共有序列可以被多数动物和细菌细胞系统识别,例如被聚腺苷酸化因子,如与CstF、PAP、PAB2、CFI和/或CFII配合的切割/聚腺苷酸化特异性因子(CPSF)识别。优选地,所述聚腺苷酸化信号,优选所述共有序列NNUANA,位于3' -UTR元件元件或ORF(如果不存在3' -UTR元件)的3' -末端下游少于约50个碱基,更优选少于约30个碱基,最优选少于约25个碱基,例如21个碱基处。

[0841] 本发明所述的人工核酸分子(例如在3' -UTR元件(或ORF)的下游包含聚腺苷酸化信号的人工DNA分子)的转录将产生在其3' -UTR元件(或ORF)的下游包含聚腺苷酸化信号的成熟前-RNA。

[0842] 随后使用合适的转录系统将导致聚腺苷酸序列附着于成熟前-RNA。例如,本发明的人工核酸分子可以是包含如上文描述的3' -UTR元件和聚腺苷酸化信号的DNA分子,其可以在该DNA分子转录时导致RNA的聚腺苷酸化。因此,产生的RNA可以包含本发明的3' -UTR元件接着聚腺苷酸序列的组合。

[0843] 可能的转录系统是体外转录系统或细胞转录系统等。因此,本发明所述的人工核酸分子的转录,例如包含可读框、3' -UTR元件和/或5' -UTR元件和任选地聚腺苷酸化信号的人工核酸分子的转录,可以产生包含可读框、3' -UTR元件和任选地聚腺苷酸序列的mRNA分子。

[0844] 因此,本发明还提供人工核酸分子,其是包含可读框、如上文所述的3' -UTR元件和/或如上文所述的5' -UTR元件和任选地聚腺苷酸序列的mRNA分子。

[0845] 在另一实施方案中,本发明所述的人工核酸分子的3' -UTR不包含聚腺苷酸化信号或聚腺苷酸序列。还优选地,本发明所述的人工核酸分子不包含聚腺苷酸化信号或聚腺苷酸序列。更优选地,人工核酸分子的3' -UTR,或本发明的人工核酸分子本身,不包含聚腺苷酸化信号,尤其是其不包含聚腺苷酸化信号AAU/TAAA。

[0846] 在优选的实施方案中,本发明提供人工核酸分子,其是人工RNA分子,包含可读框和与选自下述序列组成的组的DNA序列对应的RNA序列:SEQ ID NO:1,SEQ ID NO:2,SEQ ID NO:137(或为SEQ ID NO:3),SEQ ID NO:138(或为SEQ ID NO:9),SEQ ID NO:10,SEQ ID NO:12,SEQ ID NO:15,SEQ ID NO:140(或为SEQ ID NO:17),SEQ ID NO:20,SEQ ID NO:22,SEQ ID NO:23,SEQ ID NO:143(或为SEQ ID NO:31),SEQ ID NO:32,SEQ ID NO:39,SEQ ID NO:41,SEQ ID NO:44,SEQ ID NO:45,SEQ ID NO:146(或为SEQ ID NO:52),SEQ ID NO:53,SEQ ID NO:54,SEQ ID NO:57,SEQ ID NO:60,SEQ ID NO:61,SEQ ID NO:63,SEQ ID NO:

64,SEQ ID NO:66,SEQ ID NO:68,SEQ ID NO:69,SEQ ID NO:72,SEQ ID NO:34,SEQ ID NO:35,SEQ ID NO:36,SEQ ID NO:37,SEQ ID NO:144(或为SEQ ID NO:42),SEQ ID NO:58,SEQ ID NO:62;SEQ ID NO:73,SEQ ID NO:75,SEQ ID NO:77,SEQ ID NO:79,SEQ ID NO:80,SEQ ID NO:81,SEQ ID NO:83,SEQ ID NO:84,SEQ ID NO:86,SEQ ID NO:87,SEQ ID NO:88,SEQ ID NO:89,SEQ ID NO:90,SEQ ID NO:91,SEQ ID NO:92,SEQ ID NO:93,SEQ ID NO:94,SEQ ID NO:95,SEQ ID NO:97,SEQ ID NO:98,SEQ ID NO:99,SEQ ID NO:100,SEQ ID NO:101,SEQ ID NO:102,SEQ ID NO:103,SEQ ID NO:106,SEQ ID NO:107,SEQ ID NO:108,SEQ ID NO:147(或为SEQ ID NO:109),SEQ ID NO:148(或为SEQ ID NO:110),SEQ ID NO:112,SEQ ID NO:113,SEQ ID NO:114,SEQ ID NO:115,SEQ ID NO:117,SEQ ID NO:118,SEQ ID NO:149(或为SEQ ID NO:119),SEQ ID NO:123,SEQ ID NO:126,SEQ ID NO:127,SEQ ID NO:128,SEQ ID NO:129,SEQ ID NO:130,SEQ ID NO:131,SEQ ID NO:132,SEQ ID NO:133,SEQ ID NO:134,SEQ ID NO:151(或为SEQ ID NO:136),SEQ ID NO:96,SEQ ID NO:111,SEQ ID NO:150(基于SEQ ID NO:120),SEQ ID NO:121,SEQ ID NO:124,SEQ ID NO:125或如上文所述的其片段。此外,还提供相应的人工DNA分子。

[0847] 在另一优选的实施方案中,本发明提供人工核酸分子,其是人工DNA分子,其包含可读框和选自由下述序列组成的组的序列:SEQ ID NO:1,SEQ ID NO:2,SEQ ID NO:137(或为SEQ ID NO:3),SEQ ID NO:138(或为SEQ ID NO:9),SEQ ID NO:10,SEQ ID NO:12,SEQ ID NO:15,SEQ ID NO:140(或为SEQ ID NO:17),SEQ ID NO:20,SEQ ID NO:22,SEQ ID NO:23,SEQ ID NO:143(或为SEQ ID NO:31),SEQ ID NO:32,SEQ ID NO:39,SEQ ID NO:41,SEQ ID NO:44,SEQ ID NO:45,SEQ ID NO:146(或为SEQ ID NO:52),SEQ ID NO:53,SEQ ID NO:54,SEQ ID NO:57,SEQ ID NO:60,SEQ ID NO:61,SEQ ID NO:63,SEQ ID NO:64,SEQ ID NO:66,SEQ ID NO:68,SEQ ID NO:69,SEQ ID NO:72,SEQ ID NO:34,SEQ ID NO:35,SEQ ID NO:36,SEQ ID NO:37,SEQ ID NO:144(或为SEQ ID NO:42),SEQ ID NO:58,SEQ ID NO:62;SEQ ID NO:73,SEQ ID NO:75,SEQ ID NO:77,SEQ ID NO:79,SEQ ID NO:80,SEQ ID NO:81,SEQ ID NO:83,SEQ ID NO:84,SEQ ID NO:86,SEQ ID NO:87,SEQ ID NO:88,SEQ ID NO:89,SEQ ID NO:90,SEQ ID NO:91,SEQ ID NO:92,SEQ ID NO:93,SEQ ID NO:94,SEQ ID NO:95,SEQ ID NO:97,SEQ ID NO:98,SEQ ID NO:99,SEQ ID NO:100,SEQ ID NO:101,SEQ ID NO:102,SEQ ID NO:103,SEQ ID NO:106,SEQ ID NO:107,SEQ ID NO:108,SEQ ID NO:147(或为SEQ ID NO:109),SEQ ID NO:148(或为SEQ ID NO:110),SEQ ID NO:112,SEQ ID NO:113,SEQ ID NO:114,SEQ ID NO:115,SEQ ID NO:117,SEQ ID NO:118,SEQ ID NO:149(或为SEQ ID NO:119),SEQ ID NO:123,SEQ ID NO:126,SEQ ID NO:127,SEQ ID NO:128,SEQ ID NO:129,SEQ ID NO:130,SEQ ID NO:131,SEQ ID NO:132,SEQ ID NO:133,SEQ ID NO:134,SEQ ID NO:151(或为SEQ ID NO:136),SEQ ID NO:96,SEQ ID NO:111,SEQ ID NO:150(基于SEQ ID NO:120),SEQ ID NO:121,SEQ ID NO:124,SEQ ID NO:125。

[0848] 因此,本发明提供一种人工核酸分子,所述人工核酸分子可以用作RNA分子(优选mRNA分子)的模板,其以高翻译效率为特征。换言之,所述人工核酸分子是可以用作模板用于生产mRNA的DNA。可获得的mRNA可以相应被翻译以生产由可读框编码的所需肽或蛋白。如果所述人工核酸分子是DNA,则其可以例如用作双链储存形式以连续和重复地体外或体内生产mRNA。由此,体外尤其是是指(“活”)细胞和/或组织,包括活受试者的组织。细胞尤

其包括细胞系,原代细胞,组织或受试者中的细胞。在具体实施方案中,允许细胞培养的细胞类型可以适于本发明。特别优选的是哺乳动物细胞,例如人细胞和小鼠细胞。在特别优选的实施方案中,使用人细胞系HeLa、HEPG2和U-937以及小鼠细胞系NIH3T3、JAWSII和L929。此外原代细胞是特别优选的,在特别优选的实施方案中,可以使用人皮肤成纤维细胞(HDF)。优选的细胞系的组包括

[0849] 备选地,还可以使用受试者的组织。

[0850] 在一个实施方案中,本发明所述的人工核酸分子还包含聚腺苷酸序列。例如,包含ORF(任选地接着3' UTR)的DNA分子,可以含有一段在得到的mRNA中可以转录为聚腺苷酸序列的胸腺嘧啶核苷酸。聚腺苷酸序列的长度可以不同。例如,所述聚腺苷酸序列可以具有约20个腺嘌呤核苷酸至高达约300个腺嘌呤核苷酸,优选约40至约200个腺嘌呤核苷酸,更优选从约50至约100个腺嘌呤核苷酸,如约60,70,80,90或100个腺嘌呤核苷酸的长度。最优选地,本发明的核酸包含约60至约70个核苷酸,最优选64个腺嘌呤核苷酸的聚腺苷酸序列。

[0851] 人工RNA-分子还可以通过在不需要从DNA前体转录的情况下通过化学合成的常规方法体外获得。

[0852] 在特别是优选的实施方案中,本发明所述的人工核酸分子是RNA分子,优选以5' -至-3' -方向包含可读框、如上文所述的3' -UTR元件和聚腺苷酸序列或以5' -至-3' -方向包含如上文所述的5' -UTR元件、可读框和聚腺苷酸序列的mRNA分子。

[0853] 在优选的实施方案中,可读框源自与本发明的人工核酸的3' -UTR元件和/或5' -UTR元件源自的基因不同的基因。在一些其他优选的实施方案中,可读框不编码选自以下各项组成的组的基因:ZNF460,TGM2,IL7R,BGN,TK1,RAB3B,CBX6,FZD2,COL8A1,NDUFS7,PHGDH,PLK2,TSP0,PTGS1,FBX032,NID2,ATP5D,EXOSC4,NOL9,UBB4B,VPS18,ORMDL2,FSCN1,TMEM33,TUBA4A,EMP3,TMEM201,CRIP2,BRAT1,SERPINH1,CD9,DPYSL2,CDK9,TFRC,PSMB3,FASN,PSMB6,PRSS56,KPNA6,SFT2D2,PARD6B,LPP,SPARC,SCAND1,VASN,SLC26A1,LCLAT1,FBXL18,SLC35F6,RAB3D,MAP1B,VMA21,CYBA,SEZ6L2,PCOLCE,VTN,ALDH16A1,RAVER1,KPNA6,SERINC5,JUP,CPN2,CRIP2,EPT1,PNP0,SSSCA1,POLR2L,LIN7C,UQCR10,PYCRL,AMN,MAP1S,NDUFS7,PHGDH,TSP0,ATP5D,EXOSC4,TUBB4B,TUBA4A,EMP3,CRIP2,BRAT1,CD9,CDK9,PSMB3,PSMB6,PRSS56,SCAND1,AMN,CYBA,PCOLCE,MAP1S,VTN,ALDH16A1(全部优选是人的)和Dpyl2,Ccnd1,Acox2,Cbx6,Ubc,Ldlr,Nudt22,Pcyox11,Ankrd1,Tmem37,Tspsyl4,Slc7a3,Cst6,Aacs,Nosip,Itga7,Ccnd2,Ebp,Sf3b5,Fasn,Hmgcs1,Osr1,Lmnbl,Vma21,Kif20a,Cdca8,Slc7a1,Ubqln2,Prps2,Shmt2,Aurkb,Fignl1,Cad,Anln,Slfn9,Ncaph,Pole,Uhrf1,Gjal,Fam64a,Kif2c,Tspan10,Scand1,Gpr84,Fads3,Cers6,Cxcr4,Gprc5c,Fen1,Cspg4,Mrpl34,Comtd1,Armc6,Emr4,Atp5d,1110001J03Rik,Csf2ra,Aarsd1,Kif22,Cth,Tpgs1,Ccl17,Alkbh7,Ms4a8a,Acox2,Ubc,Slpi,Pcyox11,Igf2bp1,Tmem37,Slc7a3,Cst6,Ebp,Sf3b5,Plkl1,Cdca8,Kif22,Cad,Cth,Pole,Kif2c,Scand1,Gpr84,Tpgs1,Ccl17,Alkbh7,Ms4a8a,Mrpl34,Comtd1,Armc6,Atp5d,1110001J03Rik,Nudt22,Aarsd1(全部优选是小鼠的),或其变体,条件是所述3' -UTR元件和/或5' -UTR元件是选自由SEQ ID NO:1-204的序列组成的组的序列。

[0854] 在优选的实施方案中,ORF不编码人或植物的核糖体蛋白,尤其是拟南芥的核糖体蛋白,尤其是不编码人核糖体蛋白S6(RPS6)、人核糖体蛋白L36a-样(RPL36AL)或拟南芥核

糖体蛋白S16 (RPS16)。在进一步的优选的实施方案中,可读框 (ORF) 不编码任何来源的核糖体蛋白S6 (RPS6)、核糖体蛋白L36a-样 (RPL36AL) 或核糖体蛋白S16 (RPS16)。

[0855] 在一个实施方案中,本发明提供人工DNA分子,其包含可读框,优选源自与3' -UTR元件和/或5' -UTR元件源自的基因的不同的基因的可读框;3' -UTR元件,所述3' -UTR元件包含这样的序列或由这样的序列组成:所述序列与选自由下述序列组成的组的DNA序列具有至少约60%,优选至少约70%,更优选至少约80%,更优选至少约90%,甚至更优选至少约95%;甚至更优选至少99%;甚至更优选100%的序列同一性:SEQ ID NO:152,SEQ ID NO:153,SEQ ID NO:154,SEQ ID NO:155,SEQ ID NO:156,SEQ ID NO:157,SEQ ID NO:158,SEQ ID NO:159,SEQ ID NO:160,SEQ ID NO:161,SEQ ID NO:164,SEQ ID NO:165,SEQ ID NO:167,SEQ ID NO:168,SEQ ID NO:169,SEQ ID NO:170,SEQ ID NO:171,SEQ ID NO:172,SEQ ID NO:173;SEQ ID NO:174,SEQ ID NO:175,SEQ ID NO:176,SEQ ID NO:179,SEQ ID NO:180,SEQ ID NO:181,SEQ ID NO:182,SEQ ID NO:183,SEQ ID NO:184,SEQ ID NO:187,SEQ ID NO:188,SEQ ID NO:189,SEQ ID NO:191,SEQ ID NO:204 (或为SEQ ID NO:192),SEQ ID NO:193,SEQ ID NO:194,SEQ ID NO:195,SEQ ID NO:196,SEQ ID NO:197,SEQ ID NO:198,SEQ ID NO:199,SEQ ID NO:200,SEQ ID NO:201,SEQ ID NO:202,SEQ ID NO:203,SEQ ID NO:177,和/或5' -UTR元件,所述5' -UTR元件包含这样的序列或由这样的序列组成:所述序列与选自由下述序列组成的组的DNA序列具有至少约60%,优选至少约70%,更优选至少约80%,更优选至少约90%,甚至更优选至少约95%;甚至更优选至少99%;甚至更优选100%的序列同一性:SEQ ID NO:1,SEQ ID NO:2,SEQ ID NO:137 (或为SEQ ID NO:3),SEQ ID NO:138 (或为SEQ ID NO:9),SEQ ID NO:10,SEQ ID NO:12,SEQ ID NO:15,SEQ ID NO:140 (或为SEQ ID NO:17),SEQ ID NO:20,SEQ ID NO:22,SEQ ID NO:23,SEQ ID NO:143 (或为SEQ ID NO:31),SEQ ID NO:32,SEQ ID NO:39,SEQ ID NO:41,SEQ ID NO:44,SEQ ID NO:45,SEQ ID NO:146 (或为SEQ ID NO:52),SEQ ID NO:53,SEQ ID NO:54,SEQ ID NO:57,SEQ ID NO:60,SEQ ID NO:61,SEQ ID NO:63,SEQ ID NO:64,SEQ ID NO:66,SEQ ID NO:68,SEQ ID NO:69,SEQ ID NO:72,SEQ ID NO:34,SEQ ID NO:35,SEQ ID NO:36,SEQ ID NO:37,SEQ ID NO:144 (或为SEQ ID NO:42),SEQ ID NO:58,SEQ ID NO:62;SEQ ID NO:73,SEQ ID NO:75,SEQ ID NO:77,SEQ ID NO:79,SEQ ID NO:80,SEQ ID NO:81,SEQ ID NO:83,SEQ ID NO:84,SEQ ID NO:86,SEQ ID NO:87,SEQ ID NO:88,SEQ ID NO:89,SEQ ID NO:90,SEQ ID NO:91,SEQ ID NO:92,SEQ ID NO:93,SEQ ID NO:94,SEQ ID NO:95,SEQ ID NO:97,SEQ ID NO:98,SEQ ID NO:99,SEQ ID NO:100,SEQ ID NO:101,SEQ ID NO:102,SEQ ID NO:103,SEQ ID NO:106,SEQ ID NO:107,SEQ ID NO:108,SEQ ID NO:147 (或为SEQ ID NO:109),SEQ ID NO:148 (或为SEQ ID NO:110),SEQ ID NO:112,SEQ ID NO:113,SEQ ID NO:114,SEQ ID NO:115,SEQ ID NO:117,SEQ ID NO:118,SEQ ID NO:149 (或为SEQ ID NO:119),SEQ ID NO:123,SEQ ID NO:126,SEQ ID NO:127,SEQ ID NO:128,SEQ ID NO:129,SEQ ID NO:130,SEQ ID NO:131,SEQ ID NO:132,SEQ ID NO:133,SEQ ID NO:134,SEQ ID NO:151 (或为SEQ ID NO:136),SEQ ID NO:96,SEQ ID NO:111,SEQ ID NO:150 (基于SEQ ID NO:120),SEQ ID NO:121,SEQ ID NO:124,SEQ ID NO:125;和聚腺苷酸化信号和/或聚腺苷酸序列。

[0856] 此外,本发明提供人工RNA分子,优选人工mRNA分子或人工病毒RNA分子,其包含可



读框,优选可读框源自与3'-UTR元件和/或5'-UTR元件源自的基因不同的基因;3'-UTR元件,其包含这样的序列或由这样的序列组成:所述序列与选自由下述序列组成的组的DNA序列对应的RNA序列具有至少约60%,优选至少约70%,更优选至少约80%,更优选至少约90%,甚至更优选至少约95%;甚至更优选至少99%;甚至更优选100%的序列同一性:SEQ ID NO:152,SEQ ID NO:153,SEQ ID NO:154,SEQ ID NO:155,SEQ ID NO:156,SEQ ID NO:157,SEQ ID NO:158,SEQ ID NO:159,SEQ ID NO:160,SEQ ID NO:161,SEQ ID NO:164,SEQ ID NO:165,SEQ ID NO:167,SEQ ID NO:168,SEQ ID NO:169,SEQ ID NO:170,SEQ ID NO:171,SEQ ID NO:172,SEQ ID NO:173;SEQ ID NO:174,SEQ ID NO:175,SEQ ID NO:176,SEQ ID NO:179,SEQ ID NO:180,SEQ ID NO:181,SEQ ID NO:182,SEQ ID NO:183,SEQ ID NO:184,SEQ ID NO:187,SEQ ID NO:188,SEQ ID NO:189,SEQ ID NO:191,SEQ ID NO:204(或为SEQ ID NO:192),SEQ ID NO:193,SEQ ID NO:194,SEQ ID NO:195,SEQ ID NO:196,SEQ ID NO:197,SEQ ID NO:198,SEQ ID NO:199,SEQ ID NO:200,SEQ ID NO:201,SEQ ID NO:202,SEQ ID NO:203,SEQ ID NO:177,和/或5'-UTR元件,其包含这样的序列或由这样的序列组成:所述序列与选自由下述序列组成的组的DNA序列对应的RNA序列具有至少约60%,优选至少约70%,更优选至少约80%,更优选至少约90%,甚至更优选至少约95%;甚至更优选至少99%;甚至更优选100%的序列同一性:SEQ ID NO:1,SEQ ID NO:2,SEQ ID NO:137(或为SEQ ID NO:3),SEQ ID NO:138(或为SEQ ID NO:9),SEQ ID NO:10,SEQ ID NO:12,SEQ ID NO:15,SEQ ID NO:140(或为SEQ ID NO:17),SEQ ID NO:20,SEQ ID NO:22,SEQ ID NO:23,SEQ ID NO:143(或为SEQ ID NO:31),SEQ ID NO:32,SEQ ID NO:39,SEQ ID NO:41,SEQ ID NO:44,SEQ ID NO:45,SEQ ID NO:146(或为SEQ ID NO:52),SEQ ID NO:53,SEQ ID NO:54,SEQ ID NO:57,SEQ ID NO:60,SEQ ID NO:61,SEQ ID NO:63,SEQ ID NO:64,SEQ ID NO:66,SEQ ID NO:68,SEQ ID NO:69,SEQ ID NO:72,SEQ ID NO:34,SEQ ID NO:35,SEQ ID NO:36,SEQ ID NO:37,SEQ ID NO:144(或为SEQ ID NO:42),SEQ ID NO:58,SEQ ID NO:62;SEQ ID NO:73,SEQ ID NO:75,SEQ ID NO:77,SEQ ID NO:79,SEQ ID NO:80,SEQ ID NO:81,SEQ ID NO:83,SEQ ID NO:84,SEQ ID NO:86,SEQ ID NO:87,SEQ ID NO:88,SEQ ID NO:89,SEQ ID NO:90,SEQ ID NO:91,SEQ ID NO:92,SEQ ID NO:93,SEQ ID NO:94,SEQ ID NO:95,SEQ ID NO:97,SEQ ID NO:98,SEQ ID NO:99,SEQ ID NO:100,SEQ ID NO:101,SEQ ID NO:102,SEQ ID NO:103,SEQ ID NO:106,SEQ ID NO:107,SEQ ID NO:108,SEQ ID NO:147(或为SEQ ID NO:109),SEQ ID NO:148(或为SEQ ID NO:110),SEQ ID NO:112,SEQ ID NO:113,SEQ ID NO:114,SEQ ID NO:115,SEQ ID NO:117,SEQ ID NO:118,SEQ ID NO:149(或为SEQ ID NO:119),SEQ ID NO:123,SEQ ID NO:126,SEQ ID NO:127,SEQ ID NO:128,SEQ ID NO:129,SEQ ID NO:130,SEQ ID NO:131,SEQ ID NO:132,SEQ ID NO:133,SEQ ID NO:134,SEQ ID NO:151(或为SEQ ID NO:136),SEQ ID NO:96,SEQ ID NO:111,SEQ ID NO:150(基于SEQ ID NO:120),SEQ ID NO:121,SEQ ID NO:124,SEQ ID NO:125;和聚腺苷酸化信号和/或聚腺苷酸序列。

[0857] 本发明提供人工核酸分子,优选人工mRNA,其可以以高翻译效率为特征。不受任何理论限制,高翻译效率可以由本发明所述的人工核酸分子(如人工mRNA分子)的降解的减少导致。因此,本发明的3'-UTR元件和/或本发明的5'-UTR元件可以阻止人工核酸分解和降解。

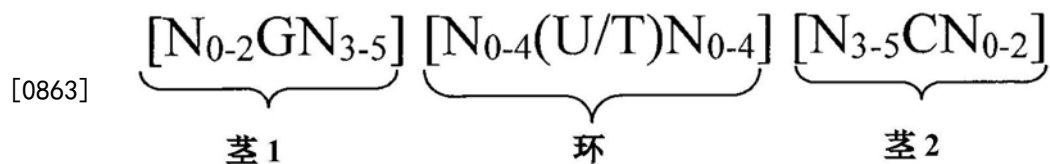
[0858] 优选地,人工核酸分子可以另外包含组蛋白茎环。由此,本发明所述的人工核酸分子可以,例如,以5'-至-3'-方向包含ORF、3'-UTR元件、任选的组蛋白茎环序列、任选的聚腺苷酸序列或聚腺苷酸化信号和任选的聚胞苷酸序列,或以5'-至-3'-方向包含5'-UTR元件、ORF、任选的组蛋白茎环序列、任选的聚腺苷酸序列或聚腺苷酸化信号和任选的聚胞苷酸序列,或以5'-至-3'-方向包含5'-UTR元件、ORF、3'-UTR元件、任选的组蛋白茎环序列、任选的聚腺苷酸序列或聚腺苷酸化信号和任选的聚胞苷酸序列。其还可以以5'-至-3'-方向包含ORF、3'-UTR元件、任选的聚腺苷酸序列、任选的聚胞苷酸序列和任选的组蛋白茎环序列,或以5'-至-3'-方向包含5'-UTR元件、ORF、任选的聚腺苷酸序列、任选的聚胞苷酸序列和任选的组蛋白茎环序列,或以5'-至-3'-方向包含5'-UTR元件、ORF、3'-UTR元件、任选的聚腺苷酸序列、任选的聚胞苷酸序列和任选的组蛋白茎环序列。

[0859] 在优选的实施方案中,本发明所述的人工核酸分子还包含至少一个组蛋白茎环序列。

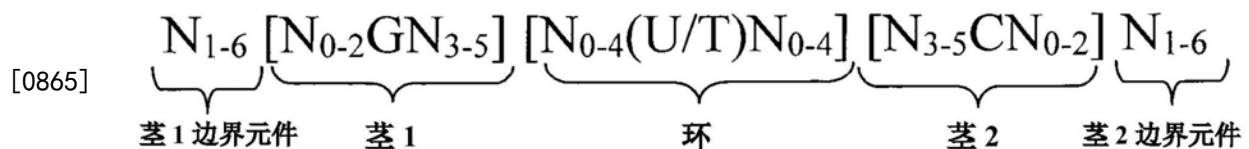
[0860] 所述组蛋白茎环序列优选选自如W0 2012/019780中公开的组蛋白茎环序列,其公开内容通过引用结合在本文中。

[0861] 适于在本发明中使用的组蛋白茎环序列优选选自下式(I)或(II)中的至少一个:

[0862] 式(I) (不具有茎边界元件的茎环序列):



[0864] 式(II) (具有茎边界元件的茎环序列):



[0866] 其中:

[0867] 茎1或茎2边界元件 $N_{1-6}$ 是1-6个、优选2-6个、更优选2-5个、甚至更优选3-5个、最优选4-5个或5个N的连续序列,其中每个N彼此独立地选自下述核苷酸:所述核苷酸选自A,U,T,G和C,或其核苷酸类似物;

[0868] 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5-7个核苷酸的连续序列;

[0869] 其中 $N_{0-2}$ 是0-2个、优选0-1个、更优选1个N的连续序列,其中每个N彼此独立地选自下述核苷酸:所述核苷酸选自A,U,T,G和C或其核苷酸类似物;

[0870] 其中 $N_{3-5}$ 是3-5个、优选4-5个、更优选4个N的连续序列,其中每个N彼此独立地选自下述核苷酸:所述核苷酸选自A,U,T,G和C或其核苷酸类似物,并且

[0871] 其中G是鸟苷或其类似物,并且可以任选地被胞苷或其类似物替代,条件是其在茎2中的互补核苷酸胞苷被鸟苷替代;

[0872] 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3-5个核苷酸、更优选4个核苷酸的连续序列;

[0873] 其中每个 $N_{0-4}$ 彼此独立地是0-4个、优选1-3个、更优选1-2个N的连续序列,其中每

个N彼此独立地选自下述核苷酸:所述核苷酸选自A,U,T,G和C或其核苷酸类似物;并且

[0874] 其中U/T表示尿苷或任选地胸苷;

[0875] 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5-7个核苷酸的连续序列;

[0876] 其中 $N_{3-5}$ 是3-5个、优选4-5个、更优选4个N的连续序列,其中每个N彼此独立地选自下述核苷酸:所述核苷酸选自A,U,T,G和C或其核苷酸类似物;

[0877] 其中 $N_{0-2}$ 是0-2个、优选0-1个、更优选1个N的连续序列,其中每个N彼此独立地选自下述核苷酸:所述核苷酸选自A,U,T,G或C,或其核苷酸类似物;并且

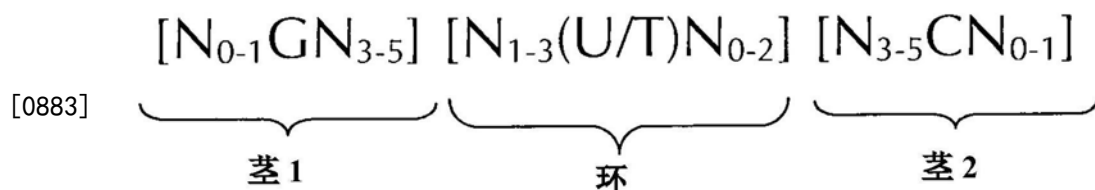
[0878] 其中C是胞苷或其类似物,并且可以任选地被鸟苷或其类似物替代,条件是其在茎1中的互补核苷鸟苷被胞苷替代;

[0879] 其中

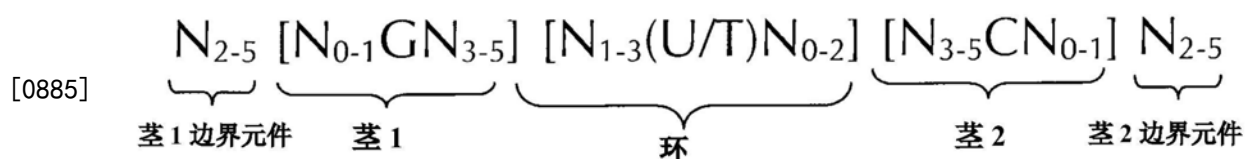
[0880] 茎1和茎2能够彼此碱基配对形成反向互补序列,其中碱基配对可以发生在茎1与茎2之间,例如,通过核苷酸A与U/T或G与C的沃森-克里克碱基配对或通过非沃森-克里克碱基配对,例如摆动碱基配对,反向沃森-克里克碱基配对,Hoogsteen碱基配对,反向Hoogsteen碱基配对,或者能够彼此碱基配对形成部分反向互补的序列,其中,基于一个茎中的一个或多个碱基在另一个茎的反向互补序列中不具有互补碱基,可以在茎1和茎2之间发生不完全的碱基配对。

[0881] 根据进一步优选的实施方案,可以根据以下具体式(Ia)或(IIa)中的至少一个选择组蛋白茎环序列:

[0882] 式(Ia)(不具有茎边界元件的茎环序列):



[0884] 式(IIa)(具有茎边界元件的茎环序列):

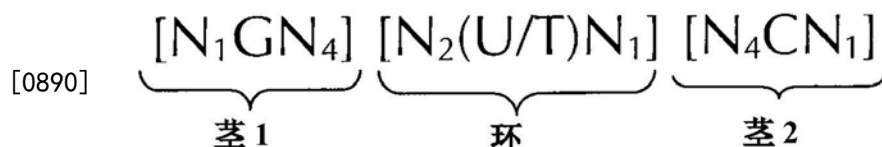


[0886] 其中:

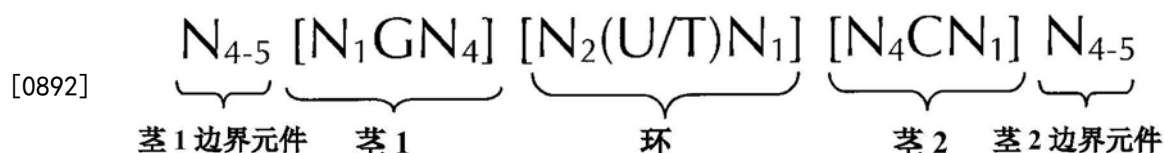
[0887] N,C,G,T和U如上文定义。

[0888] 按照第一方面的另一个更特别优选的实施方案,人工核酸分子序列可以包含至少一种按照下述具体的式(Ib)或(IIb)中的至少一个所述的组蛋白茎环序列:

[0889] 式(Ib)(不具有茎边界元件的茎环序列):



[0891] 式(IIb)(具有茎边界元件的茎环序列):



[0893] 其中：

[0894] N,C,G,T和U如上文定义。

[0895] 一个特定的优选的组蛋白茎环序列是SEQ ID NO:209: CAAAGGCTCTTTTCAGAGCCACCA,或更优选是SEQ ID NO:209的核酸序列的相对对应RNA序列。

[0896] 作为实例,单个元件可以以下述顺序存在于人工核酸分子中：

[0897] 5'-帽-5'-UTR(元件)-ORF-3'-UTR(元件)-组蛋白茎环-聚腺苷酸/聚胞苷酸序列；

[0898] 5'-帽-5'-UTR(元件)-ORF-3'-UTR(元件)-聚腺苷酸/聚胞苷酸序列-组蛋白茎环；

[0899] 5'-帽-5'-UTR(元件)-ORF-IRES-ORF-3'-UTR(元件)-组蛋白茎环-聚腺苷酸/聚胞苷酸序列；

[0900] 5'-帽-5'-UTR(元件)-ORF-IRES-ORF-3'-UTR(元件)-组蛋白茎环-聚腺苷酸/聚胞苷酸序列-聚腺苷酸/聚胞苷酸序列；

[0901] 5'-帽-5'-UTR(元件)-ORF-IRES-ORF-3'-UTR(元件)-聚腺苷酸/聚胞苷酸序列-组蛋白茎环；

[0902] 5'-帽-5'-UTR(元件)-ORF-IRES-ORF-3'-UTR(元件)-聚腺苷酸/聚胞苷酸序列-聚腺苷酸/聚胞苷酸序列-组蛋白茎环；

[0903] 5'-帽-5'-UTR(元件)-ORF-3'-UTR(元件)-聚腺苷酸/聚胞苷酸序列-聚腺苷酸/聚胞苷酸序列；

[0904] 5'-帽-5'-UTR(元件)-ORF-3'-UTR(元件)-聚腺苷酸/聚胞苷酸序列-聚腺苷酸/聚胞苷酸序列-组蛋白茎环；等等。

[0905] 在一些实施方案中,所述人工核酸分子还包含诸如5'-帽、聚胞苷酸序列和/或IRES-基序的元件。5'-帽可以在转录期间或转录后添加于RNA的5'末端。此外,尤其是如果所述核酸是mRNA的形式或编码mRNA,本发明的人工核酸分子可以被至少10个胞苷,优选至少20个胞苷,更优选至少30个胞苷的序列(所谓的“聚胞苷酸序列”)修饰。尤其是,特别是如果核酸是(m)RNA的形式或编码mRNA,本发明的人工核酸分子可以含有典型地约10至200个胞苷核苷酸,优选约10至100个胞苷核苷酸,更优选约10至70个胞苷核苷酸或甚至更优选约20至50或甚至20至30个胞苷核苷酸的聚胞苷酸序列。最优选地,本发明的核酸包含30个胞嘧啶的聚胞苷酸序列。由此,优选本发明所述的人工核酸分子,优选在5'-至-3'方向上,包含如上文所述的至少一个5'-UTR元件、ORF、如上文所述的至少一个3'-UTR元件、聚腺苷酸序列或聚腺苷酸化信号和聚胞苷酸序列,或在5'-至-3'方向,包含任选地另外的5'-UTR、ORF、如上文所述的至少一个3'-UTR元件、聚腺苷酸序列或聚腺苷酸化信号,和聚胞苷酸序列,或在5'-至-3'方向,包含如上文所述的至少一个5'-UTR元件、ORF、任选地另外的3'-UTR、聚腺苷酸序列或聚腺苷酸化信号,和聚胞苷酸序列。

[0906] 例如,如果所述人工核酸分子编码两个以上肽或蛋白,内部核糖体进入位点(IRES)序列或IRES-基序可以分开一些可读框。如果人工核酸分子是双或多顺反子核酸分子,则IRES-序列可以尤其有用。

[0907] 此外,所述人工核酸分子可以包含另外的5'-元件,优选启动子,或含启动子的序列。所述启动子可以驱动或调节本发明所述的人工核酸分子的转录,例如本发明所述的人工DNA-分子的转录。

[0908] 优选地,本发明所述的人工核酸分子,优选所述可读框,是至少部分G/C修饰的。因此,可以通过修饰所述分子的G(鸟苷)/C(胞苷)含量在热力学上稳定本发明的人工核酸分子。与相应野生型序列的可读框的G/C含量相比,可以优选通过使用遗传密码的简并增加本发明所述的人工核酸分子的可读框的G/C含量。因此,与特定野生型序列的编码氨基酸序列相比,优选不通过G/C修饰来改变人工核酸分子的编码氨基酸序列。因此,与野生型编码序列相比,编码序列或整个人工核酸分子(例如mRNA)的密码子可以不同,从而在保持翻译的氨基酸序列的同时,它们包括增加的G/C核苷酸量。由于一些密码子编码一个且相同的氨基酸(所谓的遗传密码简并)的事实,易于改变密码子而不改变编码的肽/蛋白序列(所谓的备选密码子使用)。因此,可能具体引入某些密码子(替换各自的编码相同氨基酸的野生型密码子),对于受试者中RNA的稳定性和/或密码子使用是更有利的(所谓的密码子优化)。

[0909] 与其野生型编码区相比,取决于将由如本文限定的本发明的人工核酸分子的编码区编码的氨基酸,对于核酸序列,例如可读框的修饰,存在各种可能性。在由仅含有G或C核苷酸的密码子编码的氨基酸的情况下,不需要密码子修饰。因此,因为不存在A或U/T,关于Pro(CCC或CCG)、Arg(CGC或CGG)、Ala(GCC或GCG)和Gly(GGC或GGG)的密码子不需要修饰。

[0910] 相反,含有A和/或U/T核苷酸的密码子可以通过编码相同氨基酸但不含有A和/或U/T的其它密码子的置换被修饰。例如

[0911] Pro的密码子可以从CC(U/T)或CCA被修饰为CCC或CCG;

[0912] Arg的密码子可以从CG(U/T)或CGA或AGA或AGG被修饰为CGC或CGG;

[0913] Ala的密码子可以从GC(U/T)或GCA被修饰为GCC或GCG;

[0914] Gly的密码子可以从GG(U/T)或GGA被修饰为GGC或GGG。

[0915] 在其它情况下,尽管不能从密码子中消除A或(U/T)核苷酸,然而,通过使用含有更低含量A和/或(U/T)核苷酸的密码子降低A和(U/T)含量是可能的。这些的实例是:

[0916] Phe的密码子可以从(U/T)(U/T)(U/T)被修饰为(U/T)(U/T)C;

[0917] Leu的密码子可以从(U/T)(U/T)A,(U/T)(U/T)G,C(U/T)(U/T)或C(U/T)A被修饰为C(U/T)C或C(U/T)G;

[0918] Ser的密码子可以从(U/T)C(U/T)或(U/T)CA或AG(U/T)被修饰为(U/T)CC,(U/T)CG或AGC;

[0919] Tyr的密码子可以从(U/T)A(U/T)被修饰为(U/T)AC;

[0920] Cys的密码子可以从(U/T)G(U/T)被修饰为(U/T)GC;

[0921] His的密码子可以从CA(U/T)被修饰为CAC;

[0922] Gln的密码子可以从CAA被修饰为CAG;

[0923] Ile的密码子可以从A(U/T)(U/T)或A(U/T)A被修饰为A(U/T)C;

[0924] Thr的密码子可以从AC(U/T)或ACA被修饰为ACC或ACG;

[0925] Asn的密码子可以从AA(U/T)被修饰为AAC;

[0926] Lys的密码子可以从AAA被修饰为AAG;

[0927] Val的密码子可以从G(U/T)(U/T)或G(U/T)A被修饰为G(U/T)C或G(U/T)G;

[0928] Asp的密码子可以从GA (U/T) 被修饰为GAC;

[0929] Glu的密码子可以从GAA被修饰为GAG;

[0930] 终止密码子 (U/T) AA可以被修饰为 (U/T) AG或 (U/T) GA。

[0931] 另一方面,在Met (A (U/T) G) 和Trp ( (U/T) GG) 的密码子的情形中,不存在不改变编码的氨基酸序列的情况下序列修饰的可能性。

[0932] 可以单独或以所有可能的组合使用上文所列的置换,以与其特定野生型可读框(即原始序列)相比增加如本文限定的本发明的人工核酸分子的可读框的G/C含量。因此,例如,野生型序列中存在的Thr的所有密码子可以被修饰为ACC (或ACG)。

[0933] 优选地,与野生型编码区的G/C含量相比,如本文限定的本发明的人工核酸分子的可读框的G/C含量增加至少7%,更优选至少15%,尤其优选至少20%,而不改变编码的氨基酸序列,即,使用遗传密码的简并性。根据一个特定的实施方案,本发明的人工核酸分子的可读框或其片段、变体或衍生物的可置换密码子的至少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%被置换,从而增加所述可读框的G/C含量。

[0934] 在此种情况下,尤其优选与野生型可读框相比,将如本文限定的本发明的人工核酸分子的可读框的G/C含量增加至最大(即可置换密码子的100%),而不改变编码的氨基酸序列。

[0935] 此外,所述可读框优选至少部分被密码子优化。密码子优化是基于可以由细胞中转移RNA (tRNA) 出现的不同频率确定翻译效率的发现。因此,如果所谓的“稀有密码子”以增加的程度存在于如本文限定的本发明的人工核酸分子的编码区,相应修饰的核酸序列的翻译比存在编码相对“常见的”tRNA的密码子的情况下效率低。

[0936] 因此,与相应野生型编码区相比,本发明的人工核酸分子的可读框优选被修饰,从而编码在细胞中相对稀有的tRNA的野生型序列的至少一个密码子被交换为编码在细胞比较频繁并携带与相对稀有tRNA相同的氨基酸的tRNA的密码子。通过该修饰,如本文限定的本发明的人工核酸分子的可读框被修饰,从而频繁出现的tRNA可用的密码子可以替换与稀有tRNA对应的密码子。换言之,根据本发明,通过此种修饰,编码稀有tRNA的野生型可读框的所有密码子可以交换为编码在细胞中更频繁且携带与稀有tRNA相同的氨基酸的tRNA的密码子。哪些tRNA在细胞中相对频繁存在和相反哪些相对极少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。因此,优选地,优选就本发明所述的人工核酸分子将在其中被表达的系统而言,优选就本发明所述的人工核酸分子将在其中被翻译的系统而言,所述可读框是密码子优化的。优选地,可读框的密码子使用是根据哺乳动物密码子使用,更优选根据人密码子使用而密码子优化的。优选地,所述可读框是密码子优化和G/C-含量修饰的。

[0937] 为了进一步改善降解抗性,例如对由核酸外切酶或核酸内切酶导致的体内(或上文定义的体外)降解的抗性,和/或为了进一步改善本发明所述的人工核酸分子的蛋白表达的稳定性,所述人工核酸分子可以还包含修饰,如骨架修饰,糖修饰和/或碱基修饰,例如,脂修饰等。优选地,本发明所述的人工核酸分子的转录和/或翻译不被所述修饰显著损害。

[0938] 通常,本发明的人工核酸分子可以包含任何天然(=天然存在的)核苷酸,例如鸟苷、尿嘧啶、腺苷、和/或胞嘧啶或其类似物。在这方面,核苷酸类似物定义为天然存在的核

苷酸腺苷、胞嘧啶、胸苷、鸟苷和尿苷的天然或非天然存在的变体。因此,类似物是例如具有非天然存在的官能团(其优选为添加于天然存在的核苷酸或从天然存在的核苷酸缺失或其替代核苷酸的天然存在的官能团)化学衍生的核苷酸。因此,天然存在的核苷酸的各个成分可以被修饰,即碱基组分、糖(核糖)组分和/或形成RNA序列的骨架(见上)的磷酸盐组分。鸟苷、尿苷、腺苷、胸苷和胞嘧啶的类似物包括(但不意味任何限制)例如从化学上,例如通过乙酰化、甲基化、羟基化等改变的任何天然存在的或非天然存在的鸟苷,尿苷,腺苷,胸苷或胞嘧啶,包括1-甲基-腺苷,1-甲基-鸟苷,1-甲基-肌苷,2,2-二甲基-鸟苷,2,6-二氨基嘌呤,2'-氨基-2'-脱氧腺苷,2'-氨基-2'-脱氧胞苷,2'-氨基-2'-脱氧鸟苷,2'-氨基-2'-脱氧尿苷,2-氨基-6-氯嘌呤核糖苷,2-氨基嘌呤-核糖苷,2'-阿糖腺苷,2'-阿糖胞苷,2'-阿糖尿苷,2'-叠氮基-2'-脱氧腺苷,2'-叠氮基-2'-脱氧胞苷,2'-叠氮基-2'-脱氧鸟苷,2'-叠氮基-2'-脱氧尿苷,2-氯腺苷,2'-氟-2'-脱氧腺苷,2'-氟-2'-脱氧胞苷,2'-氟-2'-脱氧鸟苷,2'-氟-2'-脱氧尿苷,2'-氟胸苷,2-甲基-腺苷,2-甲基-鸟苷,2-甲基-硫代-N6-异戊烯基(isopenenyl)-腺苷,2'-0-甲基-2-氨基腺苷,2'-0-甲基-2'-脱氧腺苷,2'-0-甲基-2'-脱氧胞苷,2'-0-甲基-2'-脱氧鸟苷,2'-0-甲基-2'-脱氧尿苷,2'-0-甲基-5-甲基尿苷,2'-0-甲基肌苷,2'-0-甲基假尿苷,2-硫代胞苷,2-硫代-胞嘧啶,3-甲基-胞嘧啶,4-乙酰基-胞嘧啶,4-硫代尿苷,5-(羟基甲基)-尿嘧啶,5,6-二氢尿苷,5-氨基烯丙基胞苷,5-氨基烯丙基-脱氧-尿苷,5-溴尿苷,5-羧基甲基氨基甲基-2-硫代-尿嘧啶,5-羧基甲基氨基甲基-尿嘧啶,5-氯-阿糖-胞嘧啶,5-氟-尿苷,5-碘尿苷,5-甲氧基羰基甲基-尿苷,5-甲氧基-尿苷,5-甲基-2-硫代-尿苷,6-氮杂胞苷,6-氮杂尿苷,6-氯-7-去氮杂-鸟苷,6-氯嘌呤核糖苷,6-巯基-鸟苷,6-甲基-巯基嘌呤-核糖苷,7-去氮杂-2'-脱氧-鸟苷,7-去氮杂腺苷,7-甲基-鸟苷,8-氮杂腺苷,8-溴-腺苷,8-溴-鸟苷,8-巯基-鸟苷,8-氧代鸟苷,苯并咪唑-核糖苷, $\beta$ -D-甘露糖基-queosine,二氢-尿嘧啶,肌苷, $N^1$ -甲基腺苷, $N^6$ -([6-氨基己基]氨甲酰基甲基)-腺苷, $N^6$ -异丙烯基-腺苷, $N^6$ -甲基-腺苷, $N^7$ -甲基-黄苷,N-尿嘧啶-5-氧基乙酸甲酯,嘌呤霉素,queosine,尿嘧啶-5-氧基乙酸,尿嘧啶-5-氧基乙酸甲酯,wybutoxosine,黄苷和木糖-腺苷。本领域技术人员例如从美国专利4,373,071,US 4,401,796,US 4,415,732,US 4,458,066,US 4,500,707,US 4,668,777,US 4,973,679,US 5,047,524,US 5,132,418,US 5,153,319,US 5,262,530和5,700,642知晓所述类似物的制备。在如上文所述的类似物的情况下,特别的偏爱可以根据本发明针对那些增加编码的肽或蛋白的蛋白表达或增加本发明的人工核酸分子的免疫原性和/或不干扰已经引入的人工核酸分子的进一步修饰的类似物的某些实施方案给出。

[0939] 根据特定实施方案,本发明的人工核酸分子可以包含脂修饰。

[0940] 在优选的实施方案中,人工核酸分子优选从5'至3'方向包含以下元件:

[0941] 5'-UTR元件,其为所述人工核酸分子(优选SEQ ID NO:1-151任一项的核酸序列)提供高翻译效率;或另外的5'-UTR,优选5'-TOP UTR;

[0942] 至少一个可读框(ORF),其中,相对于野生型序列,所述ORF优选包含至少一个修饰;

[0943] 3'-UTR元件,其为所述人工核酸分子(优选SEQ ID NO:152-204任一项的核酸序列)提供高翻译效率;或另外的3'-UTR,优选白蛋白73'-UTR;

[0944] 聚腺苷酸序列,其优选包含64个腺苷酸;

[0945] 聚胞苷酸序列,其优选包含30个腺苷酸;

[0946] 组蛋白茎环序列。

[0947] 在尤其优选的实施方案中,本发明所述的人工核酸分子还可以包含一种以上以下描述的修饰:

[0948] 化学修饰:

[0949] 如本文中使用的就人工核酸分子而言的术语“修饰”可以指包含骨架修饰以及糖修饰或碱基修饰的化学修饰。

[0950] 在该情况下,如本文定义的人工核酸分子,优选RNA分子,可以含有核苷酸类似物/修饰,例如骨架修饰,糖修饰或碱基修饰。与本发明相关的骨架修饰是这样的修饰,其中包含在如本文定义的核酸分子中的核苷酸的骨架的磷酸酯被化学修饰。与本发明相关的糖修饰是如本文定义的核酸分子的核苷酸的糖的化学修饰。此外,与本发明相关的碱基修饰是核酸分子的核苷酸的碱基部分的化学修饰。在该情况下,核苷酸类似物或修饰优选选自可用于转录和/或翻译的核苷酸类似物。

[0951] 糖修饰:

[0952] 如本文所述的可以结合入人工核酸分子(优选RNA)中的修饰的核苷和核苷酸可以在糖部分中被修饰。“氧基”-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;其中2'羟基,例如,通过亚甲基桥连接于相同核糖的4'碳的“锁定”核酸(LNA);以及氨基(-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。

[0953] “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);或氨基可以通过接头连接于糖,其中所述接头包含原子C、N和O中的一种或多种。

[0954] 糖基团还可以含有一个以上与核糖中相应碳相比具有相反立体化学构型的碳。因此,修饰的核酸分子可以包括含有例如阿拉伯糖作为糖的核苷酸。

[0955] 骨架修饰:

[0956] 还可以在可以结合入如本文所述的人工核酸分子(优选RNA)的修饰的核苷和核苷酸中进一步修饰磷酸骨架。骨架的磷酸基团可以通过用不同取代基替代一个以上氧原子来修饰。此外,修饰的核苷和核苷酸可以包括用如本文所述的修饰的磷酸酯完全替代未修饰磷酸酯部分。修饰的磷酸基团的实例包括,但不限于,硫代磷酸根、硒代磷酸根(phosphoroselenates)、硼代磷酸根(borano phosphates)、硼代磷酸酯、氢磷酸酯、氨基磷酸酯(phosphoroamidates)、烷基或芳基磷酸酯和磷酸三酯。二硫代磷酸酯的两个未连接氧被硫替代。磷酸酯接头还可以通过用氮(桥接的氨基磷酸酯)、硫(桥接的硫代磷酸酯)和碳(桥接的亚甲基-磷酸酯)替代连接的氧而被修饰。

[0957] 碱基修饰:

[0958] 如本文所述的可以结合入人工核酸分子(优选RNA分子)的修饰的核苷和核苷酸还可以在核碱基部分被修饰。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,本文中所述的核苷和核苷酸可以在主沟表面被化学修饰。在一些实施方案中,主沟化学修饰可以包括氨基、巯基、烷基、或卤代基团。



[0959] 在本发明的尤其优选的实施方案中,核苷酸类似物/修饰选自碱基修饰,其优选选自2-氨基-6-氯嘌呤核糖苷-5'-三磷酸,2-氨基嘌呤-核糖苷-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-氯嘌呤核糖苷-5'-三磷酸,7-去氮杂腺苷-5'-三磷酸,7-去氮杂鸟苷-5'-三磷酸,8-氮杂腺苷-5'-三磷酸,8-叠氮基腺苷-5'-三磷酸,苯并咪唑-核糖苷-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。对核苷酸给出对于选自由5-甲基胞苷-5'-三磷酸,7-去氮杂鸟苷-5'-三磷酸,5-溴胞苷-5'-三磷酸,和假尿苷-5'-三磷酸组成的碱基修饰的核苷酸的组的碱基修饰的特别偏爱。

[0960] 在一些实施方案中,修饰的核苷包括吡啶-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-牛磺酸甲基尿苷,1-牛磺酸甲基-假尿苷,5-牛磺酸甲基-2-硫代-尿苷,1-牛磺酸甲基-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-去氮杂-假尿苷,2-硫代-1-甲基-1-去氮杂-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。

[0961] 在一些实施方案中,修饰的核苷包括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-去氮杂-假异胞苷,1-甲基-1-去氮杂-假异胞苷,折布拉林(zebularine),5-氮杂-折布拉林,5-甲基-折布拉林,5-氮杂-2-硫代-折布拉林,2-硫代-折布拉林,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。

[0962] 在其他实施方案中,修饰的核苷包括2-氨基嘌呤,2,6-二氨基嘌呤,7-去氮杂-腺嘌呤,7-去氮杂-8-氮杂-腺嘌呤,7-去氮杂-2-氨基嘌呤,7-去氮杂-8-氮杂-2-氨基嘌呤,7-去氮杂-2,6-二氨基嘌呤,7-去氮杂-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-甲基硫代-N6-(顺式-羟基异戊烯基)腺苷,N6-甘氨酸氨甲酰基腺苷,N6-苏氨酸氨甲酰基腺苷,2-甲基硫代-N6-苏氨酸氨甲酰基腺苷,N6,N6-二甲基腺苷,7-甲基腺嘌呤,2-甲基硫代-腺嘌呤,和2-甲氧基-腺嘌呤。

[0963] 在其他实施方案中,修饰的核苷包括肌苷,1-甲基-肌苷,wyosine,wybutosine,7-去氮杂-鸟苷,7-去氮杂-8-氮杂-鸟苷,6-硫代-鸟苷,6-硫代-7-去氮杂-鸟苷,6-硫代-7-去氮杂-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。

[0964] 在一些实施方案中,核苷酸可以在主沟表面被修饰,并且可以包括用甲基或卤代基团替代尿嘧啶的C-5上的氢。在具体的实施方案中,修饰的核苷是5'-O-(1-硫代磷酸)-腺苷,5'-O-(1-硫代磷酸)-胞苷,5'-O-(1-硫代磷酸)-鸟苷,5'-O-(1-硫代磷酸)-尿苷或5'-O-(1-硫代磷酸)-假尿苷。

[0965] 在进一步具体的实施方案中,人工核酸分子,优选RNA分子,可以包含选自以下各项的核苷修饰:6-氮杂-胞苷,2-硫代-胞苷, $\alpha$ -硫代-胞苷,假-异-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-去氮杂-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-叠氮基-腺苷,7-去氮杂-腺苷。

[0966] 脂修饰:

[0967] 根据进一步的实施方案,如本文定义的人工核酸分子,优选RNA,可以含有脂修饰。该脂-修饰的RNA典型地包含如本文定义的RNA。如本文定义的该脂-修饰的RNA分子典型地还包含至少一个与该RNA分子共价连接的接头,和至少一个与各自的接头共价连接的脂。备选地,脂-修饰的RNA分子包含至少一个如本文定义的RNA分子和至少一个与该RNA分子共价连接(无接头)的(双功能)脂。根据第三个备选方案,脂-修饰的RNA分子包含如本文定义的人工核酸分子,优选RNA分子,至少一个与该RNA分子共价连接的接头,和至少一个与各自的接头共价连接的脂,并且此外至少一个与该RNA分子共价连接(无接头)的(双功能)脂。在该情况下,尤其优选的是,脂修饰存在于线性RNA序列的末端。

[0968] 修饰的RNA的5'末端的修饰:

[0969] 根据本发明的另一优选的实施方案,如本文定义的人工核酸分子,优选RNA分子,可以通过加入所谓的"5'帽"结构来修饰。

[0970] 5'-帽是实体,通常是修饰的核苷酸实体,其通常在成熟mRNA的5'端"加帽"。5'-帽可以典型地由修饰的核苷酸形成,尤其是由鸟嘌呤核苷酸的衍生物形成。优选地,5'-帽与5'-末端经由5'-5'-三磷酸连接来连接。5'-帽可以是甲基化的,例如m7GpppN,其中N是携带5'-帽的核酸的末端5'核苷酸,典型地是RNA的5'-端。m7GpppN是由聚合酶II转录的mRNA中天然存在的5'-帽结构并且因此不被认为是包含在本发明所述的修饰的RNA中的修饰。这意为,本发明所述的人工核酸分子,优选RNA分子,可以包含m7GpppN作为5'-帽,但此外,所述人工核酸分子,优选RNA分子,包含至少一个如本文定义的进一步修饰。

[0971] 5'帽结构的进一步实例包括甘油基,反向脱氧非碱性残基(部分),4',5'亚甲基核苷酸,1-( $\beta$ -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, $\alpha$ -核苷酸,修饰的碱基核苷酸,苏-戊呋喃糖基核苷酸,无环3',4'-开环核苷酸,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'-反向核苷酸部分,3'-3'-反向非碱性部分,3'-2'-反向核苷酸部分,3'-2'-反向非碱性部分,1,4-丁二醇磷酸酯,3'-氨基磷酸酯,磷酸己酯,磷酸氨基己酯,3'-磷酸脂,3'硫代磷酸酯,二硫代磷酸酯,或桥接或非桥接的磷酸甲酯部分。这些修饰的5'-帽结构被认为是本发明所述的人工核酸分子(优选

RNA分子)中包含的至少一个修饰。

[0972] 尤其优选的修饰的5'-帽结构是CAP1 (m7G的相邻核苷酸的核糖的甲基化), CAP2 (m7G下游第二个核苷酸的核糖甲基化), CAP3 (m7G下游第三个核苷酸的核糖甲基化), CAP4 (m7G下游第四个核苷酸的核糖甲基化), ARCA (抗逆CAP类似物, 修饰的ARCA (例如硫代磷酸酯修饰的ARCA), 肌苷, N1-甲基-鸟苷, 2'-氟-鸟苷, 7-去氮杂-鸟苷, 8-氧代-鸟苷, 2-氨基-鸟苷, LNA-鸟苷, 和2-叠氮基-鸟苷。

[0973] 在优选的实施方案中, 至少一个可读框编码治疗性蛋白或肽。在另一实施方案中, 抗原由至少一个可读框编码, 如致病抗原、肿瘤抗原、致敏抗原或自身免疫抗原。其中, 编码抗原的人工核酸分子的施用用于针对涉及所述抗原的疾病的基因疫苗接种方法。

[0974] 在备选实施方案中, 抗体或抗原特异性T细胞受体或其片段由本发明所述的人工核酸分子的至少一个可读框编码。

[0975] 抗原:

[0976] 致病抗原:

[0977] 本发明所述的人工核酸分子可以编码这样的蛋白或肽, 所述蛋白或肽包含致病抗原或其片段、变体或衍生物。所述致病抗原来源于致病性生物体, 特别是细菌、病毒或原生动物(多细胞)致病性生物体, 其激发受试者、特别是哺乳动物受试者、更特别是人中的免疫学反应。更特别地, 致病抗原优选是表面抗原, 例如, 位于病毒或细菌或原生动物生物体的表面上的蛋白(或蛋白的片段, 例如, 表面抗原的外部部分)。

[0978] 致病抗原是优选来源于与传染病相关的病原体的肽或蛋白抗原, 其优选地选自来源于下述病原体的抗原: 鲍氏不动杆菌 (*Acinetobacter baumannii*), 无形体属 (*Anaplasma* genus), *Anaplasma phagocytophilum*, 猫钩虫 (*Ancylostoma braziliense*), 十二指肠钩虫 (*Ancylostoma duodenale*), 溶血隐秘杆菌 (*Arcanobacterium haemolyticum*), 人蛔虫 (*Ascaris lumbricoides*), 曲霉菌属 (*Aspergillus* genus), 星状病毒科 (*Astroviridae*), 巴贝虫属 (*Babesia* genus), 炭疽芽孢杆菌 (*Bacillus anthracis*), 蜡状芽孢杆菌 (*Bacillus cereus*), 汉氏巴尔通氏体 (*Bartonella henselae*), BK病毒, 人芽囊原虫 (*Blastocystis hominis*), 皮炎芽生菌 (*Blastomyces dermatitidis*), 百日咳博德特氏菌 (*Bordetella pertussis*), 伯氏疏螺旋体 (*Borrelia burgdorferi*), 疏螺旋体属 (*Borrelia* genus), 疏螺旋体属物种 (*Borrelia* spp), 布鲁杆菌属 (*Brucella* genus), 马来丝虫 (*Brugia malayi*), 布尼亚病毒科 (*Bunyaviridae* family), 洋葱伯克霍尔德氏菌 (*Burkholderia cepacia*) 和其他伯克霍尔德氏菌属物种 (*Burkholderia* species), 鼻疽伯克霍尔德氏菌 (*Burkholderia mallei*), 类鼻疽伯克霍尔德氏菌 (*Burkholderia pseudomallei*), 嵌杯病毒科 (*Caliciviridae* family), 弯曲杆菌属 (*Campylobacter* genus), 白色念珠菌 (*Candida albicans*), 念珠菌属物种 (*Candida* spp), 砂眼衣原体 (*Chlamydia trachomatis*), 肺炎衣原体 (*Chlamydophila pneumoniae*), 鹦鹉热衣原体 (*Chlamydophila psittaci*), CJD朊病毒, 华支睾吸虫 (*Clonorchis sinensis*), 肉毒梭菌 (*Clostridium botulinum*), 艰难梭菌 (*Clostridium difficile*), 产气荚膜梭菌 (*Clostridium perfringens*), 产气荚膜梭菌, 梭菌属物种 (*Clostridium* spp), 破伤风梭菌 (*Clostridium tetani*), 球孢菌属物种 (*Coccidioides* spp), 冠状病毒 (*coronaviruses*), 白喉棒杆菌 (*Corynebacterium diphtheriae*), 伯内特考克斯体 (*Coxiella burnetii*), 克

里米亚-刚果出血热病毒(Crimean-Congo hemorrhagic fever virus),新型隐球菌(*Cryptococcus neoformans*),隐孢子虫属(*Cryptosporidium* genus),巨细胞病毒(Cytomegalovirus,CMV),登革热病毒(Dengue viruses)(DEN-1,DEN-2,DEN-3和DEN-4),脆弱双核阿米巴(*Dientamoeba fragilis*),埃博拉病毒(Ebolavirus,EBOV),棘球属(*Echinococcus* genus),恰菲埃里希氏体(*Ehrlichia chaffeensis*),尤氏埃里希氏体(*Ehrlichia ewingii*),埃里希氏体属(*Ehrlichia* genus),溶组织内阿米巴(*Entamoeba histolytica*),肠球菌属(*Enterococcus* genus),肠道病毒属(*Enterovirus* genus),肠道病毒(Enteroviruses),主要为柯萨奇病毒A病毒(Coxsackie A virus)和肠道病毒71(EV71),表皮癣菌属物种(*Epidermophyton* spp),埃巴病毒(Epstein-Barr Virus,EBV),大肠杆菌(*Escherichia coli*)0157:H7,0111和0104:H4,肝片形吸虫(*Fasciola hepatica*)和巨大片形吸虫(*Fasciola gigantica*),FFI朊病毒,丝虫目超家族(*Filarioidea* superfamily),黄病毒(Flaviviruses),土拉热弗朗西丝菌(*Francisella tularensis*),梭杆菌属(*Fusobacterium* genus),白地霉(*Geotrichum candidum*),肠贾第虫(*Giardia intestinalis*),颚口线虫属物种(*Gnathostoma* spp),GSS朊病毒,瓜纳里托病毒(*Guanarito virus*),),杜克雷嗜血杆菌(*Haemophilus ducreyi*),流感嗜血菌(*Haemophilus influenzae*),幽门螺杆菌(*Helicobacter pylori*),亨尼病毒属(Henipavirus)(亨德拉病毒(Hendra virus)尼帕病毒(Nipah virus)),甲型肝炎病毒(Hepatitis A Virus),乙型肝炎病毒(Hepatitis B Virus,HBV),丙型肝炎病毒(Hepatitis C Virus,HCV),丁型肝炎病毒(Hepatitis D Virus),戊型肝炎病毒(Hepatitis E Virus),单纯疱疹病毒(Herpes simplex virus)1和2(HSV-1和HSV-2),荚膜组织胞浆菌(*Histoplasma capsulatum*),HIV(人免疫缺陷病毒(Human immunodeficiency virus)),威尼克何德霉(*Hortaea werneckii*),人博卡病毒(Human bocavirus,HBoV),人疱疹病毒6(HHV-6)和人疱疹病毒7(HHV-7),人偏肺病毒(Human metapneumovirus,hMPV),人乳头瘤病毒(Human papillomavirus,HPV),人副流感病毒(Human parainfluenza viruses,HPIV),日本脑炎病毒(Japanese encephalitis virus),JC病毒,胡宁病毒(Junin virus),金氏金氏菌(*Kingella kingae*),肉芽肿杆菌(*Klebsiella granulomatis*),库鲁朊病毒(Kuru prion),拉沙病毒(Lassa virus),嗜肺军团病杆菌(*Legionella pneumophila*),利什曼原虫属(*Leishmania* genus),钩端螺旋体属(*Leptospira* genus),单核细胞增生利斯特菌(*Listeria monocytogenes*),淋巴细胞性脉络丛脑膜炎病毒(Lymphocytic choriomeningitis virus,LCMV),Machupo病毒,马拉色霉菌属物种(*Malassezia* spp),马堡病毒(Marburg virus),麻疹病毒(Measles virus),横川后殖吸虫(*Metagonimus yokagawai*),微孢子虫门(Microsporidia phylum),传染性软疣病毒(Molluscum contagiosum virus,MCV),腮腺炎病毒(Mumps virus),麻风分枝杆菌(*Mycobacterium leprae*)和Mycobacterium lepromatosis,结核分枝杆菌(Mycobacterium tuberculosis),溃疡分枝杆菌(Mycobacterium ulcerans),肺炎支原体(Mycoplasma pneumoniae),福氏耐格里原虫(*Naegleria fowleri*),美洲板口线虫(*Necator americanus*),淋病奈瑟球菌(*Neisseria gonorrhoeae*),脑膜炎奈瑟球菌(*Neisseria meningitidis*),星状诺卡菌(*Nocardia asteroides*),诺卡菌属物种(*Nocardia* spp),旋盘尾丝虫(*Onchocerca volvulus*),恙虫热立克次体(*Orientia tsutsugamushi*),正粘病

毒科 (*Orthomyxoviridae* family) (流感 (*Influenza*)), 巴西副球孢子菌 (*Paracoccidioides brasiliensis*), 并殖吸虫属物种 (*Paragonimus* spp), 卫斯特曼并殖吸虫 (*Paragonimus westermani*), 细小病毒 (*Parvovirus*) B19, 巴斯德菌属 (*Pasteurella* genus), 疟原虫属 (*Plasmodium* genus), 杰氏肺囊虫 (*Pneumocystis jirovecii*), 脊髓灰质炎病毒 (*Poliovirus*), 狂犬病病毒 (*Rabies virus*), 呼吸道合胞病毒 (*Respiratory syncytial virus*, RSV), 鼻病毒属 (*Rhinovirus*), 鼻病毒 (*rhinoviruses*), 痘立克次体 (*Rickettsia akari*), 立克次体属 (*Rickettsia* genus), 普氏立克次体 (*Rickettsia prowazekii*), 立氏立克次体 (*Rickettsia rickettsii*), 斑疹伤寒立克次体 (*Rickettsia typhi*), 裂谷热病毒 (*Rift Valley fever virus*), 轮状病毒属 (*Rotavirus*), 风疹病毒 (*Rubella virus*), 萨比亚病毒 (*Sabia virus*), 沙门菌属 (*Salmonella* genus), 疥螨 (*Sarcoptes scabiei*), SARS冠状病毒 (*SARS coronavirus*), 血吸虫属 (*Schistosoma* genus), 志贺菌属 (*Shigella* genus), 辛诺柏病毒 (*Sin Nombre virus*), 汉坦病毒属 (*Hantavirus*), 申克孢子丝菌 (*Sporothrix schenckii*), 葡萄球菌属 (*Staphylococcus* genus), 葡萄球菌属, 无乳链球菌 (*Streptococcus agalactiae*), 肺炎链球菌 (*Streptococcus pneumoniae*), 酿脓链球菌 (*Streptococcus pyogenes*), 粪类圆线虫 (*Strongyloides stercoralis*), 绦虫属 (*Taenia* genus), 猪肉绦虫 (*Taenia solium*), 蜱传脑炎病毒 (*Tick-borne encephalitis virus*, TBEV), 犬弓蛔线虫 (*Toxocara canis*) 或猫弓蛔虫 (*Toxocara cati*), 鼠弓形虫 (*Toxoplasma gondii*), 苍白密螺旋体 (*Treponema pallidum*), 旋毛线虫 (*Trichinella spiralis*), 阴道毛滴虫 (*Trichomonas vaginalis*), 发癣菌属物种 (*Trichophyton* spp), 鞭形鞭虫 (*Trichuris trichiura*), 布鲁斯锥虫 (*Trypanosoma brucei*), 克鲁斯锥虫 (*Trypanosoma cruzi*), 尿素分解尿素支原体 (*Ureaplasma urealyticum*), 水痘-带状疱疹病毒 (*Varicella zoster virus*, VZV), 水痘-带状疱疹病毒 (VZV), 重型天花 (*Variola major*) 或类天花 (*Variola minor*), vCJD朊病毒, 委内瑞拉马脑炎病毒 (*Venezuelan equine encephalitis virus*), 霍乱弧菌 (*Vibrio cholerae*), 西尼罗河病毒 (*West Nile virus*), 西部马脑炎病毒 (*Western equine encephalitis virus*), 班氏吴策线虫 (*Wuchereria bancrofti*), 黄热病病毒 (*Yellow fever virus*), 小肠结肠炎耶尔森氏菌 (*Yersinia enterocolitica*), 鼠疫耶尔森氏菌 (*Yersinia pestis*) 和假结核耶尔森氏菌 (*Yersinia pseudotuberculosis*)。

[0979] 在这一情形中, 特别优选的是来自选自下述的病原体的抗原: 流感病毒 (*Influenza virus*), 呼吸道合胞病毒 (RSV), 单纯疱疹病毒 (HSV), 人乳头瘤病毒 (HPV), 人免疫缺陷病毒 (HIV), 疟原虫属 (*Plasmodium*), 金黄色葡萄球菌 (*Staphylococcus aureus*), 登革热病毒, 砂眼衣原体 (*Chlamydia trachomatis*), 巨细胞病毒 (CMV), 乙型肝炎病毒 (HBV), 结核分枝杆菌 (*Mycobacterium tuberculosis*), 狂犬病病毒和黄热病病毒。

[0980] 肿瘤抗原:

[0981] 在进一步的实施方案中, 本发明所述的人工核酸分子可以编码蛋白或肽, 其包含含有肿瘤抗原、所述肿瘤抗原的片段、变体或衍生物的肽或蛋白, 优选为, 其中所述肿瘤抗原是黑素细胞特异的抗原、睾丸癌抗原或肿瘤特异的抗原, 优选为CT-X抗原, 非-X CT-抗原, 针对CT-X抗原的结合伙伴或针对非-X CT-抗原或肿瘤特异的抗原的结合伙伴, 更优选为CT-X抗原, 针对非-X CT-抗原或肿瘤特异的抗原或所述肿瘤抗原的片段、变体或衍生物

的合作伙伴;并且其中各个核酸序列编码不同的肽或蛋白;并且其中至少一个核酸序列编码5T4、707-AP、9D7、AFP、AlbZIP HPG1、 $\alpha$ -5- $\beta$ -1-整联蛋白、 $\alpha$ -5- $\beta$ -6-整联蛋白、 $\alpha$ -辅肌动蛋白-4/m、 $\alpha$ -甲基酰基-辅酶A消旋酶、ART-4、ARTC1/m、B7H4、BAGE-1、BCL-2、bcr/abl、 $\beta$ -联蛋白/m、BING-4、BRCA1/m、BRCA2/m、CA 15-3/CA 27-29、CA 19-9、CA72-4、CA125、钙网蛋白、CAMEL、CASP-8/m、组织蛋白酶B、组织蛋白酶L、CD19、CD20、CD22、CD25、CDE30、CD33、CD4、CD52、CD55、CD56、CD80、CDC27/m、CDK4/m、CDKN2A/m、CEA、CLCA2、CML28、CML66、COA-1/m、coactosin-样蛋白、胶原XXIII、COX-2、CT-9/BRD6、Cten、细胞周期蛋白B1、细胞周期蛋白D1、cyp-B、CYPB1、DAM-10、DAM-6、DEK-CAN、EFTUD2/m、EGFR、ELF2/m、EMMPRIN、EpCam、EphA2、EphA3、ErbB3、ETV6-AML1、EZH2、FGF-5、FN、Frau-1、G250、GAGE-1、GAGE-2、GAGE-3、GAGE-4、GAGE-5、GAGE-6、GAGE7b、GAGE-8、GDEP、GnT-V、gp100、GPC3、GPNMB/m、HAGE、HAST-2、hepsin、Her2/neu、HERV-K-MEL、HLA-A\*0201-R17I、HLA-A11/m、HLA-A2/m、HNE、同源框NKX3.1、HOM-TES-14/SCP-1、HOM-TES-85、HPV-E6、HPV-E7、HSP70-2M、HST-2、hTERT、iCE、IGF-1R、IL-13Ra2、IL-2R、IL-5、不成熟层粘连蛋白受体、激肽释放酶-2、激肽释放酶-4、Ki67、KIAA0205、KIAA0205/m、KK-LC-1、K-Ras/m、LAGE-A1、LDLR-FUT、MAGE-A1、MAGE-A2、MAGE-A3、MAGE-A4、MAGE-A6、MAGE-A9、MAGE-A10、MAGE-A12、MAGE-B1、MAGE-B2、MAGE-B3、MAGE-B4、MAGE-B5、MAGE-B6、MAGE-B10、MAGE-B16、MAGE-B17、MAGE-C1、MAGE-C2、MAGE-C3、MAGE-D1、MAGE-D2、MAGE-D4、MAGE-E1、MAGE-E2、MAGE-F1、MAGE-H1、MAGEL2、乳腺球蛋白A、MART-1/melan-A、MART-2、MART-2/m、基质蛋白22、MC1R、M-CSF、ME1/m、间皮素(mesothelin)、MG50/PXDN、MMP11、MN/CA IX-抗原、MRP-3、MUC-1、MUC-2、MUM-1/m、MUM-2/m、MUM-3/m、I型肌球蛋白I/m、NA88-A、N-乙酰基葡萄糖胺基转移酶-V、Neo-PAP、Neo-PAP/m、NFYC/m、NGEP、NMP22、NPM/ALK、N-Ras/m、NSE、NY-ESO-1、NY-ESO-B、OA1、OFA-iLRP、OGT、OGT/m、OS-9、OS-9/m、骨钙素、骨桥蛋白、p15、p190小bcr-abl、p53、p53/m、PAGE-4、PAI-1、PAI-2、PAP、PART-1、PATE、PDEF、Pim-1-激酶、Pin-1、Pml/PAR $\alpha$ 、POTE、PRAME、PRDX5/m、prostein、蛋白酶-3、PSA、PSCA、PSGR、PSM、PSMA、PTPRK/m、RAGE-1、RBAF600/m、RHAMM/CD168、RU1、RU2、S-100、SAGE、SART-1、SART-2、SART-3、SCC、SIRT2/m、Sp17、SSX-1、SSX-2/HOM-MEL-40、SSX-4、STAMP-1、STEAP-1、存活素、存活素-2B、SYT-SSX-1、SYT-SSX-2、TA-90、TAG-72、TARP、TEL-AML1、TGF $\beta$ 、TGF $\beta$ RII、TGM-4、TPI/m、TRAG-3、TRG、TRP-1、TRP-2/6b、TRP/INT2、TRP-p8、酪氨酸酶、UPA、VEGFR1、VEGFR-2/FLK-1、WT1和淋巴血细胞的免疫球蛋白独特型或淋巴血细胞的T细胞受体独特型,或所述肿瘤抗原片段、变体或衍生物;优选为存活素或其同系物,MAGE-家族的抗原或其结合伙伴或所述肿瘤抗原的片段、变体或衍生物。在该情况下尤其优选的是肿瘤抗原NY-ESO-1、5T4、MAGE-C1、MAGE-C2、存活素、Muc-1、PSA、PSMA、PSCA、STEAP和PAP。

[0982] 在优选的实施方案中,人工核酸分子编码蛋白或肽,其包含治疗性蛋白或其片段、变体或衍生物。

[0983] 本文定义的治疗性蛋白是有益于任何遗传或获得性疾病的治疗或改善个体的状况的肽或蛋白。具体地,除其他功能外,治疗性蛋白在产生可以修饰并修复遗传错误、破坏癌细胞或病原体感染的细胞、治疗免疫系统病症、治疗代谢或内分泌病症的治疗剂方面起重要作用。例如,红细胞生成素(EPO)(一种蛋白激素)可以用于治疗患有红细胞缺乏症(erythrocyte deficiency)的患者,其是肾脏并发症的常见原因。此外,治疗性蛋白涵盖佐剂蛋白、治疗性抗体,以及例如用于绝经期妇女的治疗中的激素替代疗法。在最近的方法

中,利用患者的体细胞来使其重新编程为多能干细胞,所述多能干细胞替代有争议的干细胞疗法。此外,用于使体细胞重新编程或用于使干细胞分化的这些蛋白在本文中定义为治疗性蛋白。此外,治疗性蛋白可以用于其他目的,例如,伤口愈合、组织再生、血管发生等。此外,抗原特异的B细胞受体和其片段和变体在本文中定义为治疗性蛋白。

[0984] 因此,治疗性蛋白可以用于多种目的,包括多种疾病的治疗,所述疾病如例如,传染病,肿瘤(例如,癌症或肿瘤疾病),血液和血液形成器官的疾病,内分泌、营养和代谢病,神经系统的疾病,循环系统的疾病,呼吸系统的疾病,消化系统的疾病,皮肤和皮下组织的疾病,肌肉骨骼系统和结缔组织的疾病和生殖泌尿系统的疾病,不管其是遗传的还是获得性的。

[0985] 在这一情形中,可以特别用于代谢或内分泌病症的治疗的特别优选的治疗性蛋白选自下述(括号中是在治疗中使用所述治疗性蛋白的具体疾病):酸性鞘磷脂酶(尼曼-皮克病(Niemann-Pick disease)),Adipotide(肥胖(obesity)),Agalsidase- $\beta$ (人半乳糖苷酶A(human galactosidase A))(法布里病(Fabry disease);防止可能导致肾脏和心血管并发症的脂肪累积),Alglucosidase(庞贝病(Pompe disease)(糖原贮积症II型(glycogen storage disease type II))), $\alpha$ -半乳糖苷酶A( $\alpha$ -GAL A,Agalsidase $\alpha$ )(法布里病), $\alpha$ -葡萄糖苷酶(糖原贮积症(Glycogen storage disease,GSD),Morbus Pompe), $\alpha$ -L-艾杜糖苷酸酶(黏多糖贮积症(mucopolysaccharidoses,MPS),胡尔勒综合征(Hurler syndrome),沙伊综合征(Scheie syndrome)), $\alpha$ -N-乙酰葡萄糖苷酶(桑菲利波综合征(Sanfilippo syndrome)),双调蛋白(癌症,代谢紊乱(metabolic disorder)),血管生成素((Ang1,Ang2,Ang3,Ang4,ANGPTL2,ANGPTL3,ANGPTL4,ANGPTL5,ANGPTL6,ANGPTL7)(血管发生,稳定血管)), $\beta$ 动物纤维素(代谢紊乱(metabolic disorder)), $\beta$ -葡萄糖醛酸糖苷酶(斯赖综合征(Sly syndrome)),骨形态发生蛋白BMP(BMP1,BMP2,BMP3,BMP4,BMP5,BMP6,BMP7,BMP8a,BMP8b,BMP10,BMP15)(再生性作用,骨骼相关的病症,慢性肾病(chronic kidney disease,CKD)),CLN6蛋白(CLN6疾病-非典型性晚期婴儿期,晚期发作的变体,青少年早期,神经元蜡样脂褐质沉积症(Neural Ceroid Lipofuscinoses,NCL)),表皮生长因子(EGF)(伤口愈合,调节细胞生长、增殖和分化),Epigen(代谢紊乱),表皮调节素(代谢紊乱),成纤维细胞生长因子(FGF,FGF-1,FGF-2,FGF-3,FGF-4,FGF-5,FGF-6,FGF-7,FGF-8,FGF-9,FGF-10,FGF-11,FGF-12,FGF-13,FGF-14,FGF-16,FGF-17,FGF-17,FGF-18,FGF-19,FGF-20,FGF-21,FGF-22,FGF-23)(伤口愈合,血管发生,内分泌紊乱,组织再生),Galsulphase(黏多糖贮积症VI(Mucopolysaccharidosis VI)),葛瑞林(Ghrelin)(肠易激综合征(irritable bowel syndrome,IBS),肥胖,普拉德-威利综合征(Prader-Willi syndrome),II型糖尿病(type II diabetes mellitus)),葡萄糖脑苷脂酶(戈谢病(Gaucher's disease)),GM-CSF(再生作用,产生白血细胞,癌症),肝素-结合EGF-样生长因子(HB-EGF)(伤口愈合,心脏肥大(cardiac hypertrophy)和心脏发育与功能),肝细胞生长因子HGF(再生作用,伤口愈合),Hepcidin(铁代谢紊乱(iron metabolism disorders)), $\beta$ -地中海贫血(Beta-thalassemia)),人白蛋白(减少的白蛋白产生(低蛋白血症(hypoproteinaemia)),增加的白蛋白损失(肾病综合征(nephrotic syndrome)),低血容量症(hypovolaemia),高胆红素血症(hyperbilirubinaemia)),艾度硫酸酯酶(Idursulphase)(艾杜糖醛酸-2-硫酸酯酶(黏多糖贮积症II(Mucopolysaccharidosis II)(亨特综合征(Hunter syndrome))),整联

蛋白 $\alpha$ V $\beta$ 3、 $\alpha$ V $\beta$ 5和 $\alpha$ 5 $\beta$ 1(结合基质大分子和蛋白酶,血管发生),艾杜糖醛酸硫酸酯酶(亨特综合征(Hunter syndrome)),拉罗尼酶(Laronidase)(胡尔勒和胡尔勒-沙伊形式的黏多糖贮积症I(Hurler and Hurler-Scheie forms of mucopolysaccharidosis I)),N-乙酰半乳糖胺-4-硫酸酯酶(rhASB;galsulfase,芳基硫酸酯酶A(ARSA),芳基硫酸酯酶B(ARSB))(芳基硫酸酯酶B缺乏症,马洛托-拉梅综合征(Maroteaux-Lamy syndrome),黏多糖贮积症VI(mucopolysaccharidosis VI)),N-乙酰葡萄糖胺-6-硫酸酯酶(桑菲利波综合征(Sanfilippo syndrome)),神经生长因子(NGF,脑源性神经营养因子(BDNF),神经营养因子-3(NT-3),和神经营养因子4/5(NT-4/5)(再生作用,心血管病(cardiovascular diseases),冠状动脉粥样硬化(coronary atherosclerosis),肥胖,2型糖尿病,代谢综合征,急性冠状动脉综合征(acute coronary syndromes),痴呆(dementia),抑郁(depression),精神分裂症(schizophrenia),孤独症(autism),雷特综合征(Rett syndrome),神经性厌食症(anorexia nervosa),神经性贪食症(bulimia nervosa),伤口愈合,皮肤溃疡(skin ulcers),角膜溃疡(corneal ulcers),阿尔茨海默病(Alzheimer's disease)),神经调节蛋白(NRG1,NRG2,NRG3,NRG4)(代谢紊乱,精神分裂症),神经毡蛋白(NRP-1,NRP-2)(血管发生,轴突导向,细胞存活,迁移),肥胖抑制素(Obestatin)(肠易激综合征(irritable bowel syndrome,IBS),肥胖,普拉德-威利综合征(Prader-Willi syndrome),II型糖尿病),血小板衍生生长因子(PDGF(PDGF-A,PDGF-B,PDGF-C,PDGF-D)(再生作用,伤口愈合,血管发生中的紊乱(disorder in angiogenesis),动脉硬化(Arteriosclerosis),纤维化(Fibrosis),癌症),TGF $\beta$ 受体(内皮因子,TGF- $\beta$ 1受体,TGF- $\beta$ 2受体,TGF- $\beta$ 3受体)(肾纤维化(renal fibrosis),肾病(kidney disease),糖尿病(diabetes),终末期肾病(ultimately end-stage renal disease,ESRD),血管发生),血小板生成素(THPO)(巨核细胞生长和发育因子(Megakaryocyte growth and development factor,MGDF))(血小板病症(platelets disorders),献血用血小板(platelets for donation),骨髓抑制性化疗后血小板计数的恢复),转化生长因子(TGF(TGF- $\alpha$ ,TGF- $\beta$ (TGF $\beta$ 1,TGF $\beta$ 2和TGF $\beta$ 3))) (再生作用,伤口愈合,免疫性,癌症,心脏疾病,糖尿病,马方综合征(Marfan syndrome),Loeys-Dietz综合征(Loeys-Dietz syndrome)),VEGF(VEGF-A,VEGF-B,VEGF-C,VEGF-D,VEGF-E,VEGF-F和PIGF)(再生作用,血管发生,伤口愈合,癌症,渗透性),奈西立肽(Nesiritide)(急性代偿失调性充血性心力衰竭(Acute decompensated congestive heart failure)),胰蛋白酶(褥疮溃疡(Decubitus ulcer),静脉曲张性溃疡(varicose ulcer),焦痂清创(debridement of eschar),裂开性伤口(dehiscent wound),晒伤(sunburn),胎粪性肠梗阻(meconium ileus)),促肾上腺皮质激素(adrenocorticotrophic hormone,ACTH)(艾迪生病("Addison's disease),小细胞癌(Small cell carcinoma),肾上腺脑白质营养不良(Adrenoleukodystrophy),先天性肾上腺增生(Congenital adrenal hyperplasia),库欣综合征(Cushing's syndrome),纳尔逊综合征(Nelson's syndrome),婴儿痉挛(Infantile spasms)),心房钠尿肽(Atrial-natriuretic peptide,ANP)(内分泌紊乱(endocrine disorders)),胆囊收缩素(Cholecystokinin)(不同的),胃泌素(Gastrin)(低胃泌素血症(hypogastrinemia)),瘦蛋白(糖尿病,高甘油三酯血症(hypertriglyceridemia),肥胖),催产素(刺激人乳喂养,分娩的不进展(non-progression of parturition)),生长抑素(类癌综合征(carcinoid



syndrome)、急性静脉曲张破裂出血(acute variceal bleeding)和肢端肥大症(acromegaly),肝肾多囊性病(polycystic diseases of the liver and kidney),肢端肥大症和由神经内分泌瘤(neuroendocrine tumors)引起的症状的对症治疗),加压素(Vasopressin)(抗利尿激素)(尿崩症(diabetes insipidus)),降钙素(Calcitonin)(绝经后骨质疏松(Postmenopausal osteoporosis)),高钙血症(Hypercalcaemia),佩吉特病(Paget's disease),骨转移(Bone metastases),幻肢痛(Phantom limb pain),椎管狭窄(Spinal Stenosis)),Exenatide(对二甲双胍(metformin)和磺脲(sulphonylurea)治疗有耐药性的2型糖尿病),生长激素(Growth hormone,GH),生长素(somatotropin)(由于GH不足(GH deficiency)或慢性肾功能不全(chronic renal insufficiency)、普拉德-威利综合征(Prader-Willi syndrome)、特纳综合征(Turner syndrome)、艾滋消耗(AIDS wasting)或使用抗病毒疗法的恶病质(cachexia)导致的生长不足(Growth failure)),胰岛素(糖尿病,糖尿病酮症酸中毒(diabetic ketoacidosis),高钾血症(hyperkalaemia)),胰岛素样生长因子1IGF-1(具有GH基因缺失或严重的原发性IGF1不足(severe primary IGF1 deficiency)的儿童生长不足,神经变性疾病(neurodegenerative disease),心血管疾病(cardiovascular diseases),心力衰竭(heart failure)),林美卡舍明(Mecasermin rinfabate),IGF-1类似物(具有GH基因缺失或严重的原发性IGF1不足的儿童生长不足,神经变性疾病,心血管疾病,心力衰竭),美卡舍明(Mecasermin),IGF-1类似物(有GH基因缺失或严重的原发性IGF1不足的儿童生长不足,神经变性疾病,心血管疾病,心力衰竭),培维索孟(Pegvisomant)(肢端肥大症),普兰林肽(Pramlintide)(糖尿病,与胰岛素组合),特立帕肽(Teriparatide)(人甲状旁腺素残基1-34)(严重的骨质疏松症(Severe osteoporosis)),贝卡普勒明(Becaplermin)(糖尿病性溃疡(diabetic ulcers)辅助性清创术(Debridement)),Dibotermine- $\alpha$ (骨形态发生蛋白2)(脊柱融合手术(Spinal fusion surgery)),骨损伤修复(bone injury repair)),醋酸组氨瑞林(Histrelin acetate)(促性腺素释放激素(gonadotropin releasing hormone;GnRH))(早熟青春期(Precocious puberty)),奥曲肽(Octreotide)(肢端肥大症,VIP-分泌腺瘤(VIP-secreting adenoma)和转移性类癌肿瘤(metastatic carcinoid tumours)的症状缓和),和帕利夫明(Palifermin)(角质形成细胞生长因子(keratinocyte growth factor;KGF))(经历化疗的患者中的严重的口粘膜炎(Severe oral mucositis),伤口愈合)。

[0986] 这些和其他的蛋白应该理解为是治疗性的,原因在于其旨在通过以充分的量替代其功能蛋白的缺陷性内源产生而治疗受试者。因此,所述治疗性蛋白典型地是哺乳动物蛋白,特别是人蛋白。

[0987] 为了治疗血液病症、循环系统的疾病、呼吸系统的疾病、癌症或肿瘤疾病、传染病或免疫缺陷,可以使用下述治疗性蛋白:阿替普酶(Alteplase)(组织型纤溶酶原激活物(tissue plasminogen activator;tPA))(肺栓塞(Pulmonary embolism),心肌梗死(myocardial infarction),急性局部缺血性卒中(acute ischaemic stroke),中央静脉通路装置的阻塞(occlusion of central venous access devices)),阿尼普酶(Anistreplase)(溶栓(Thrombolysis)),抗凝血酶III(AT-III)(遗传性AT-III不足(Hereditary AT-III deficiency)),血栓栓塞(Thromboembolism)),比伐卢定(Bivalirudin)(冠状动脉血管成形术(coronary angioplasty)和肝素诱发的血小板减少

症 (heparin-induced thrombocytopaenia) 中减少的血液凝结危险), 达贝泊汀 $\alpha$  (Darbepoetin- $\alpha$ ) (治疗患有慢性肾功能不全 (chronic renal insufficiency) 和慢性肾衰竭 (肾功能不全) (+/-透析) 的患者的贫血症 (anaemia)), 屈曲克凝 $\alpha$  (Drotrecogin- $\alpha$ ) (活化的蛋白C) (具有高死亡危险的严重败血症 (Severe sepsis)), 促红细胞生成素, 依泊汀 $\alpha$  (Epoetin- $\alpha$ ), 促血红细胞生长素, 红细胞生成素 (erthropoietin) (慢性疾病的贫血症, myelodysplasia, 由于肾衰竭或化疗导致的贫血, 术前准备), 因子IX (血友病B (Haemophilia B)), 因子VIIa (患有血友病A或B的患者中的出血和因子VIII或因子IX的抑制剂), 因子VIII (血友病A (Haemophilia A)), 来匹卢定 (Lepirudin) (肝素诱发的血小板减少症 (Heparin-induced thrombocytopaenia)), 蛋白质C浓缩剂 (静脉血栓形成 (Venous thrombosis), 暴发性紫癜 (Purpura fulminans)), 瑞替普酶 (Retepase) (tPA的缺失突变蛋白) (急性心肌梗死 (acute myocardial infarction) 的管理, 心室功能的改善), 链激酶 (急性进展性透壁性心肌梗死 (Acute evolving transmural myocardial infarction), 肺栓塞 (pulmonary embolism), 深静脉血栓形成 (deep vein thrombosis), 动脉血栓形成 (arterial thrombosis) 或栓塞 (embolism), 动静脉插管闭塞 (occlusion of arteriovenous cannula)), 替奈普酶 (Tenecteplase) (急性心肌梗死), 尿激酶 (肺栓塞), 血管抑素 (癌症), 抗-CD22免疫毒素 (复发的CD33+急性髓性白血病 (Relapsed CD33+acute myeloid leukaemia)), 地尼白介素-毒素连接物 (Denileukin diftitox) (皮肤T细胞淋巴瘤 (Cutaneous T-cell lymphoma, CTCL)), Immunocyanin (膀胱和前列腺癌), MPS (金属泛调理素 (Metallopanstimulin)) (癌症), Aflibercept (非小细胞肺癌 (Non-small cell lung cancer, NSCLC), 转移性结直肠癌 (metastatic colorectal cancer, mCRC), 激素难治性转移性前列腺癌 (hormone-refractory metastatic prostate cancer), 湿性黄斑变性 (wet macular degeneration)), 内皮抑素 (Endostatin) (癌症, 炎性疾病, 如类风湿性关节炎 (rheumatoid arthritis) 以及克罗恩病 (Crohn's disease), 糖尿病视网膜病变 (diabetic retinopathy), 银屑病 (psoriasis) 和子宫内膜异位症 (endometriosis)), 胶原酶 (慢性皮肤溃疡 (chronic dermal ulcers) 和严重烧伤区域 (severely burned areas) 的清创术, 迪皮特朗挛缩 (Dupuytren's contracture), 佩罗尼病 (Peyronie's disease)), 人脱氧核糖核酸酶I, 链道酶 (dornase) (囊性纤维化 (Cystic fibrosis)); 减少所选的患有FVC的患者中的呼吸道感染, 大于预测的40%), 透明质酸酶 (用作佐剂以增加注射的药物的吸收和分散, 特别是眼科手术中的麻醉剂和特定的成像剂的吸收和分散), 木瓜蛋白酶 (坏死组织的清创或急性和慢性损伤中腐肉的软化, 诸如压力性溃疡 (pressure ulcers), 曲张性和糖尿病性溃疡 (varicose and diabetic ulcers), 烧伤 (burns), 术后伤口 (postoperative wounds), 藏毛囊肿伤口 (pilonidal cyst wounds), 痈 (carbuncles), 以及其他伤口), L-天冬酰胺酶 (急性淋巴性白血病 (Acute lymphocytic leukaemia), 其需要外源性天冬酰胺进行增殖), Peg-天冬酰胺酶 (急性淋巴性白血病, 其需要外源性天冬酰胺进行增殖), 拉布立酶 (Rasburicase) (患有白血病 (leukaemia)、淋巴瘤 (lymphoma) 和实体瘤的正在进行引起肿瘤溶解综合征 (tumour lysis syndrome) 的抗癌疗法的儿科患者), 人慢性促性腺素 (Human chorionic gonadotropin, HCG) (辅助的生殖), 人促卵泡激素 (follicle-stimulating hormone, FSH) (辅助的生殖), 促黄体素- $\alpha$  (Lutropin- $\alpha$ ) (具有促黄体激素不足的不育 (Infertility with luteinizing hormone deficiency)), 催乳素 (低泌乳素血

症(Hypoprolactinemia),血清催乳素不足(serum prolactin deficiency),妇女卵巢功能不全(ovarian dysfunction in women),焦虑(anxiety),动脉生成性勃起功能障碍(arteriogenic erectile dysfunction),早泄(premature ejaculation),精子减少症(oligozoospermia),精子活力不足(asthenospermia),精囊功能不良(hypofunction of seminal vesicles),男性的雄激素缺乏症(hypoandrogenism)), $\alpha$ -1-蛋白酶抑制剂(先天性抗胰蛋白酶缺乏(Congenital antitrypsin deficiency)),乳糖酶(由于乳糖消化不能(inability to digest lactose)引起的胀气(Gas)、胃气胀(bloating)、痛性痉挛(cramps)和腹泻(diarrhoea)),胰酶(脂肪酶,淀粉酶,蛋白酶(protease))(囊性纤维化,慢性胰腺炎(chronic pancreatitis),胰腺功能不全(pancreatic insufficiency),Billroth II胃分流术后(post-Billroth II gastric bypass surgery),胰管阻塞(pancreatic duct obstruction),脂肪泻(steatorrhoea),消化不良(poor digestion),胀气(gas),胃气胀(bloating)),腺苷脱氨酶(牛培加酶(pegademase bovine,PEG-ADA))(由于腺苷脱氨酶不足导致的严重组合性免疫缺陷疾病(Severe combined immunodeficiency disease)),阿巴他塞(Abatcept)(类风湿性关节炎(Rheumatoid arthritis)(尤其是当耐受TNF $\alpha$ 抑制时)),阿来法塞(Alefcept)(斑块状银屑病(Plaque Psoriasis)),阿那白滞素(Anakinra)(类风湿性关节炎),依那西普(Etanercept)(类风湿性关节炎,多关节病程幼年型类风湿关节炎(polyarticular-course juvenile rheumatoid arthritis),银屑病关节炎(psoriatic arthritis),强直性脊柱炎(ankylosing spondylitis),斑块状银屑病,强直性脊柱炎),白细胞介素-1(IL-1)受体拮抗剂,阿那白滞素(Anakinra)(与类风湿性关节炎相关的炎症和软骨降解(cartilage degradation)),胸腺素(Thymulin)(神经变性病,风湿病(rheumatism),神经性厌食症(anorexia nervosa)),TNF- $\alpha$ 拮抗剂(自身免疫病症,诸如类风湿性关节炎,强直性脊柱炎,克罗恩病,银屑病,化脓性汗腺炎(hidradenitis suppurativa),难治性哮喘(refractory asthma)),恩夫韦肽(Enfuvirtide)(HIV-1感染)和胸腺素 $\alpha$ 1(乙型和丙型肝炎(Hepatitis B和C))(括号中是在治疗中使用所述治疗性蛋白的具体疾病)。

[0988] (括号中是在治疗中使用所述治疗性蛋白的具体的疾病)

[0989] 在进一步的方面,本发明提供载体,所述载体包含:

[0990] a. 可读框(ORF)和/或克隆位点,例如用于插入可读框或包含可读框的序列的克隆位点;和

[0991] b. 至少一个3'-非翻译区元件(3'-UTR元件)和/或至少一个5'-非翻译区元件(5'-UTR元件),其中所述人工核酸分子以高翻译效率为特征。

[0992] 通常,本发明所述的载体可以包含如上文所述的本发明所述的人工核酸分子。尤其是,上述对于本发明所述的人工核酸分子的优选的实施方案也适用于本发明所述的载体包含的本发明所述的人工核酸分子。例如,在本发明的载体中,所述至少一个3'-UTR元件和/或所述至少一个5'-UTR元件和ORF如上文对于本发明所述的人工核酸分子所述,包括优选的实施方案。例如,在本发明所述的载体中,所述至少一个3'-UTR元件和/或所述至少一个5'-UTR元件源自的稳定mRNA可以优选特征为mRNA降解,其中所述mRNA在第二时间点的量与所述mRNA在第一时间点的量的比率是至少0.5(50%),至少0.6(60%),至少0.7(70%),至少0.75(75%),至少0.8(80%),至少0.85(85%),至少0.9(90%),或至少0.95(95%)。

[0993] 所述克隆位点可以是适于引入可读框或包含可读框的序列的任何序列,比如一种以上限制位点。因此,包含克隆位点的载体优选适于将可读框插入所述载体,优选适于将可读框插入5'-UTR元件的3'和/或3'-UTR元件的5'。优选所述克隆位点或所述ORF位于5'-UTR元件的3'和/或3'-UTR元件的5',优选紧邻5'-UTR元件的3'端和/或3'-UTR元件的5'端。例如,所述克隆位点或所述ORF可以紧接于所述5'-UTR元件的3'端和/或所述3'-UTR元件的5'端或它们可以通过如上文描述的用于本发明所述的人工核酸分子的一段核苷酸连接,比如通过一段2,4,6,8,10,20个等的核苷酸连接。

[0994] 优选地,本发明所述的载体适于例如,通过任选地将可读框或包含可读框的序列插入载体并转录所述载体生产本发明所述的人工核酸分子,优选适于生产本发明所述的人工mRNA。因此,优选地,所述载体包含转录所需的元件,如启动子,例如RNA聚合酶启动子。优选地,所述载体适于使用真核、原核、病毒或噬菌体转录系统,如真核细胞、原核细胞,或真核、原核、病毒或噬菌体体外转录系统转录。因此,例如,所述载体可以包含启动子序列,所述启动子序列被聚合酶,如RNA聚合酶,例如真核、原核、病毒、或噬菌体RNA聚合酶识别。在一个优选的实施方案中,所述载体包含噬菌体RNA聚合酶启动子,如SP6、T3或T7,优选T7启动子。优选地,所述载体适于使用基于噬菌体的体外转录系统,如基于T7 RNA聚合酶的体外转录系统体外转录。

[0995] 在另一优选的实施方案中,载体可以直接用于在细胞或组织中表达编码的肽或蛋白。为此,载体包含在那些细胞/组织中表达所必需的特定元件,例如特定启动子序列,如CMV启动子。

[0996] 所述载体还可以包含如上文对于本发明所述的人工核酸分子描述的聚腺苷酸序列和/或聚腺苷酸化信号。

[0997] 所述载体可以是RNA载体或DNA载体。优选地,所述载体是DNA载体。所述载体可以是技术人员已知的任何载体,如病毒载体或质粒载体。优选地,所述载体是质粒载体,优选DNA质粒载体。

[0998] 在一个优选的实施方案中,本发明所述的载体包含本发明所述的人工核酸分子。

[0999] 优选地,本发明所述的DNA载体包含与基因的转录本的3'-UTR的核酸序列,如与SEQ ID NOs:152-204的核酸序列具有至少约1,2,3,4,5,10,15,20,30或40%,优选至少约50%,优选至少约60%,优选至少约70%,更优选至少约80%,更优选至少约90%,甚至更优选至少约95%,甚至更优选至少约99%,最优选100%的同一性的核酸序列。

[1000] 优选地,本发明所述的DNA载体包含与基因的转录本的5'-UTR的核酸序列,如与SEQ ID NOs:1-151的核酸序列具有至少约1,2,3,4,5,10,15,20,30或40%,优选至少约50%,优选至少约60%,优选至少约70%,更优选至少约80%,更优选至少约90%,甚至更优选至少约95%,甚至更优选至少约99%,最优选100%同一性的核酸序列。

[1001] 优选地,本发明所述的DNA载体包含选自以下各项组成的组的序列:下述所示的5'-UTR元件的DNA序列:SEQ ID NO:1,SEQ ID NO:2,SEQ ID NO:137(或为SEQ ID NO:3),SEQ ID NO:138(或为SEQ ID NO:9),SEQ ID NO:10,SEQ ID NO:12,SEQ ID NO:15,SEQ ID NO:140(或为SEQ ID NO:17),SEQ ID NO:20,SEQ ID NO:22,SEQ ID NO:23,SEQ ID NO:143(或为SEQ ID NO:31),SEQ ID NO:32,SEQ ID NO:39,SEQ ID NO:41,SEQ ID NO:44,SEQ ID NO:45,SEQ ID NO:146(或为SEQ ID NO:52),SEQ ID NO:53,SEQ ID NO:54,SEQ ID NO:57,

SEQ ID NO:60,SEQ ID NO:61,SEQ ID NO:63,SEQ ID NO:64,SEQ ID NO:66,SEQ ID NO:68,SEQ ID NO:69,SEQ ID NO:72,SEQ ID NO:34,SEQ ID NO:35,SEQ ID NO:36,SEQ ID NO:37,SEQ ID NO:144(或为SEQ ID NO:42),SEQ ID NO:58,SEQ ID NO:62;SEQ ID NO:73,SEQ ID NO:75,SEQ ID NO:77,SEQ ID NO:79,SEQ ID NO:80,SEQ ID NO:81,SEQ ID NO:83,SEQ ID NO:84,SEQ ID NO:86,SEQ ID NO:87,SEQ ID NO:88,SEQ ID NO:89,SEQ ID NO:90,SEQ ID NO:91,SEQ ID NO:92,SEQ ID NO:93,SEQ ID NO:94,SEQ ID NO:95,SEQ ID NO:97,SEQ ID NO:98,SEQ ID NO:99,SEQ ID NO:100,SEQ ID NO:101,SEQ ID NO:102,SEQ ID NO:103,SEQ ID NO:106,SEQ ID NO:107,SEQ ID NO:108,SEQ ID NO:147(或为SEQ ID NO:109),SEQ ID NO:148(或为SEQ ID NO:110),SEQ ID NO:112,SEQ ID NO:113,SEQ ID NO:114,SEQ ID NO:115,SEQ ID NO:117,SEQ ID NO:118,SEQ ID NO:149(或为SEQ ID NO:119),SEQ ID NO:123,SEQ ID NO:126,SEQ ID NO:127,SEQ ID NO:128,SEQ ID NO:129,SEQ ID NO:130,SEQ ID NO:131,SEQ ID NO:132,SEQ ID NO:133,SEQ ID NO:134,SEQ ID NO:151(或为SEQ ID NO:136),SEQ ID NO:96,SEQ ID NO:111,SEQ ID NO:150(基于SEQ ID NO:120),SEQ ID NO:121,SEQ ID NO:124,SEQ ID NO:125,和下述所示的3'-UTR元件的DNA序列:SEQ ID NO:152,SEQ ID NO:153,SEQ ID NO:154,SEQ ID NO:155,SEQ ID NO:156,SEQ ID NO:157,SEQ ID NO:158,SEQ ID NO:159,SEQ ID NO:160,SEQ ID NO:161,SEQ ID NO:164,SEQ ID NO:165,SEQ ID NO:167,SEQ ID NO:168,SEQ ID NO:169,SEQ ID NO:170,SEQ ID NO:171,SEQ ID NO:172,SEQ ID NO:173;SEQ ID NO:174,SEQ ID NO:175,SEQ ID NO:176,SEQ ID NO:179,SEQ ID NO:180,SEQ ID NO:181,SEQ ID NO:182,SEQ ID NO:183,SEQ ID NO:184,SEQ ID NO:187,SEQ ID NO:188,SEQ ID NO:189,SEQ ID NO:191,SEQ ID NO:204(或为SEQ ID NO:192),SEQ ID NO:193,SEQ ID NO:194,SEQ ID NO:195,SEQ ID NO:196,SEQ ID NO:197,SEQ ID NO:198,SEQ ID NO:199,SEQ ID NO:200,SEQ ID NO:201,SEQ ID NO:202,SEQ ID NO:203,SEQ ID NO:177,或与所述序列的任一种具有至少约40%,优选至少约50%,优选至少约60%,优选至少约70%,更优选至少约80%,更优选至少约90%,甚至更优选至少约95%;甚至更优选至少约99%序列同一性的序列或其片段(如上文所述的),优选其功能性片段。

[1002] 优选地,本发明所述的RNA载体包含选自由以下各项组成的组的序列:根据对应于上文关于本发明所述的DNA载体所述的DNA序列的RNA序列的序列。

[1003] 优选地,所述载体是环状分子。优选地,所述载体是双链分子,如双链DNA分子。此种环状,优选双链DNA分子对于本发明的人工核酸分子可以作为储存形式方便使用。此外,其可以用于转染细胞,例如,培养的细胞。其还可以用于体外转录以获得本发明所述的人工RNA分子。

[1004] 优选地,所述载体,优选环状载体是线性化的,例如,通过限制性酶消化线性化的。在一个优选的实施方案中,所述载体包含切割位点,如限制性位点,优选独特的切割位点,所述切割位点紧位于ORF的3',或者——如果存在——紧位于3'-UTR元件的3',或者——如果存在——位于聚腺苷酸序列或聚腺苷酸化信号的3',或者——如果存在——位于聚胞苷酸序列的3',或者——如果存在——位于组蛋白茎环的3'。因此,优选通过线性化载体获得的产物在3'末端以ORF的3'末端,或者——如果存在——以3'-UTR元件的3'端,或者——如果存在——以聚腺苷酸序列或聚腺苷酸化信号的3'端,或者——如果存在——以聚胞苷酸

序列的3'端终止。在该实施方案中,其中本发明所述的载体包含本发明所述的人工核酸分子,限制位点,优选独特限制位点,优选紧位于人工核酸分子的3'端的3'。

[1005] 在另一方面,本发明涉及一种细胞,所述细胞包含本发明所述的人工核酸分子或本发明所述的载体。所述细胞可以是任何细胞,如细菌细胞,昆虫细胞,植物细胞,脊椎动物细胞,例如哺乳动物细胞。此种细胞可以,例如,用于本发明的载体的复制,例如,在细菌细胞中复制。此外,所述细胞可以用于转录本发明所述的人工核酸分子或载体和/或翻译本发明所述的人工核酸分子或载体的可读框。例如,所述细胞可以用于重组蛋白产生。

[1006] 例如,本发明所述的细胞可通过标准核酸转移方法,如标准转染、转导或转化方法获得。例如,可以将本发明所述的人工核酸分子或载体通过电穿孔、脂转染、例如基于阳离子脂和/或脂质体、磷酸钙沉淀、基于纳米粒子的转染、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转入所述细胞。

[1007] 优选地,所述细胞是哺乳动物细胞,如人受试者、家畜、实验室动物的细胞,如小鼠或大鼠细胞。优选所述细胞是人细胞。所述细胞可以是建立的细胞系的细胞,如CHO, BHK, 293T, COS-7, HeLa, HEPG2和HEK等,或所述细胞可以是原代细胞,比如人皮肤成纤维细胞(HDF)细胞等,优选分离自生物体的细胞。在一个优选的实施方案中,所述细胞是哺乳动物受试者,优选人受试者的分离的细胞。例如,所述细胞可以是优选哺乳动物受试者,优选人受试者的免疫细胞,如树突细胞,癌症或肿瘤细胞,或任何体细胞等。

[1008] 在另一方面,本发明提供一种药物组合物,所述药物组合物包含本发明所述的人工核酸分子,本发明所述的载体,或本发明所述的细胞。本发明所述的药物组合物可以用作,例如疫苗,例如用于基因接种。因此,所述ORF可以,例如,编码将要施用于患者用于疫苗接种的抗原。因此,在一个优选的实施方案中,本发明所述的药物组合物是疫苗。此外,本发明所述的药物组合物可以用于例如基因治疗。

[1009] 优选地,所述药物组合物还包含一种或多种药学可接受载体、稀释剂和/或赋形剂和/或一种或多种佐剂。在本发明范围内,药学可接受的载体典型地包括用于本发明的药物组合物的液体或非液体基础。在一个实施方案中,以液体形式提供所述药物组合物。在此种情况下,优选地,所述载体基于水,如无菌原水,等渗盐水或缓冲的(含水)溶液,例如磷酸盐,柠檬酸盐等缓冲的溶液。所述缓冲液可以就特定参考介质而言是高渗的、等渗的或低渗的,即就特定参考介质而言,所述缓冲液可以具有更高、相同的或更低的盐含量,其中优选可以使用之前提到的盐的此种浓度,其不由于渗透或其他浓度效应而导致哺乳动物细胞的破坏。参考介质例如在“体内”方法中存在的液体,如血液,淋巴,细胞内液,或其它液体,或例如可以在“体外”方法中用作参考介质的液体,如常见缓冲液或液体。此种常见缓冲液或液体是技术人员已知的。林格乳酸盐溶液作为液体基础是尤其优选的。

[1010] 对于本发明的药物组合物,也可以使用适于施用于患者的一种以上相容的固体或液体填充物或稀释剂或封装化合物。如本文使用的术语“相容的”优选意为本发明的药物组合物的这些组分能够以不发生基本上降低本发明的药物组合物在典型应用条件下的药学有效性的相互作用的方式与本文限定的本发明的人工核酸、载体或细胞混合。

[1011] 本发明所述的药物组合物还可以任选地包含一个以上另外的药学活性组分。在此种情况下,药学活性组分是对治愈、缓解或预防特定适应证或疾病呈现治疗效果的化合物。此种化合物包括,不暗含任何限制,肽或蛋白,核酸,(治疗上有活性的)低分子量有机或无

机化合物(分子量小于5000,优选小于1000),糖,抗原或抗体,现有技术中已知的治疗剂,抗原细胞,抗原细胞片段,细胞级分,细胞壁组分(例如多糖),修饰的、减毒的或灭活的(例如化学地或通过放射)病原体(病毒,细菌等)。

[1012] 此外,本发明的药物组合物可以包含用于人工核酸分子或载体(vector)的载体(carrier)。此种载体可以适于介导在生理可接受液体中的溶解,药学活性人工核酸分子或载体的转运和细胞摄取。因此,此种载体可以是适于储存和递送本发明所述的人工核酸分子或载体的组分。此种组分可以是,例如,阳离子或聚阳离子载体或可以用作转染或复合剂的化合物。

[1013] 在本发明所述的优选实施方案中,所述人工核酸分子或载体与一种或多种阳离子或聚阳离子化合物复合,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复合,例如,与鱼精蛋白、阳离子或聚阳离子多糖和/或阳离子或聚阳离子脂质复合。

[1014] 按照一个优选的实施方案,所述人工核酸分子或载体可以与脂质复合形成一种或多种脂质体、脂复合物或脂质纳米颗粒。因此,在一个实施方案中,本发明的药物组合物包含脂质体、脂复合物和/或脂质纳米颗粒,所述脂质体、脂复合物和/或脂质纳米颗粒包含所述人工核酸分子或载体,优选RNA,优选mRNA。

[1015] 由于它们的生物相容性和它们容易进行大规模生产,基于脂质的制剂已经日益被认为是一种最有希望的核酸、特别是RNA的递送系统。阳离子脂质作为用于递送RNA的合成物质已经进行了广泛研究。在混合在一起后,核酸被阳离子脂质凝聚,形成称为脂复合物的脂质/核酸复合物。这些脂质复合物能够保护遗传物质免受核酸酶的作用,并通过与带负电荷的细胞膜相互作用而将其递送到细胞中。脂复合物可以通过直接将带正电荷的脂质在生理pH下与带负电荷的核酸混合而制备。

[1016] 常规的脂质体由脂双层组成,所述脂双层可以由阳离子、阴离子或中性(磷)脂质和胆固醇组成,其包绕一个水性核心。脂双层和水性空间都可以分别掺入疏水或亲水的化合物。可以通过向脂质体表面添加亲水聚合物涂层(例如,聚乙二醇(PEG))赋予空间稳定性而改良脂质体特征和体内的表现。此外,脂质体可以通过在其表面上或在连接的PEG链的末端连接配体(例如,抗体、肽和碳水化合物)而用于特异性的靶向(Front Pharmacol.2015Dec 1;6:286)。

[1017] 脂质体通常是由包绕水性区室的磷脂双层构成的基于胶体液的和基于表面活性剂的递送系统。它们可以呈现为球形囊泡,并且尺寸可以在20nm至数微米范围内。基于阳离子脂质的脂质体能够通过静电相互作用与带负电荷的核酸复合,形成提供体内临床应用所需要的生物相容性、低毒性和大规模生产的可能性的复合物。脂质体可以与质膜融合用于摄入;一旦在细胞内部,脂质体通过内吞途径加工,并且然后将遗传物质从内体/载体释放到细胞质中。由于它们优越的生物相容性,脂质体长久以来被认为是药物递送载体,假定脂质体是生物膜的基本类似物,并且可以由天然的和合成的磷脂二者制备而成(Int J Nanomedicine.2014;9:1833-1843)。

[1018] 阳离子脂质体已经成为传统上最常用的非病毒递送系统,用于寡核苷酸,包括质粒DNA,反义寡聚体和siRNA/小发夹RNA-shRNA)。阳离子脂质,如DOTAP,(1,2-二油酰基-3-三甲基铵-丙烷)和DOTMA(N-[1-(2,3-二油酰基氧基)丙基]-N,N,N-三甲基-甲基硫酸铵),可以与带负电荷的核酸形成复合物或脂复合物,从而通过静电相互作用形成纳米颗粒,提

供高的体外转染效率。此外,开发了用于RNA递送的基于中性脂质的纳米脂质体,例如,基于中性的1,2-二油酰基-sn-甘油-3-磷脂酰胆碱(DOPC)的纳米脂质体。(Adv Drug Deliv Rev.2014Feb;66:110-116.)。

[1019] 因此,在一个实施方案中,所述人工核酸分子或载体,优选本发明所述的药物组合物的RNA,与阳离子脂质和/或中性脂质复合,并且由此形成脂质体、脂质纳米颗粒、脂复合物或基于中性脂质的纳米脂质体。

[1020] 在另外的实施方案中,在此种情况下,尤其优选的转染或复合剂是阳离子或聚阳离子化合物,包括鱼精蛋白,核仁蛋白(nucleoline),精胺或亚精胺,或其它阳离子肽或蛋白,如聚-L-赖氨酸(PLL),聚-精氨酸,碱性多肽,细胞穿透肽(CPP),包括HIV-结合肽,HIV-1Tat(HIV),源自Tat的肽,穿透肽(Penetratin),源自VP22的或类似物肽,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,Pep-1,L-寡聚物,降钙素肽,源自触角足的肽(尤其源自果蝇触角足(*Drosophila antennapedia*)),pAntp,pIsl,FGF,乳铁蛋白,Transportan,Buforin-2,Bac715-24,SynB,SynB(1),pVEC,源自hCT的肽,SAP,或组蛋白。

[1021] 此外,此种阳离子或聚阳离子化合物或载体可以是阳离子或聚阳离子肽或蛋白,其优选包含或另外修饰以包含至少一个SH部分。优选地,阳离子或聚阳离子载体选自具有以下总式(I)的阳离子肽:

[1022]  $\{(\text{Arg})_l; (\text{Lys})_m; (\text{His})_n; (\text{Orn})_o; (\text{Xaa})_x\}$ ; 式(I)

[1023] 其中 $l+m+n+o+x=3-100$ ,并且假如Arg(精氨酸)、Lys(赖氨酸)、His(组氨酸)和Orn(鸟氨酸)的总含量代表寡肽的所有氨基酸的至少10%,则 $l,m,n$ 或 $o$ 相互独立地是选自0,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21-30,31-40,41-50,51-60,61-70,71-80,81-90和91-100的任何数;并且Xaa是选自除了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;并且假设Xaa的总含量不超过寡肽的所有氨基酸的90%, $x$ 是选自0,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21-30,31-40,41-50,51-60,61-70,71-80,81-90的任何数。氨基酸Arg、Lys、His、Orn和Xaa的任一个可以置于肽的任何位置。在此种情况下,在7-30个氨基酸范围内的阳离子肽或蛋白是特别优选的。优选的阳离子肽和蛋白还记述在国际专利申请W0 2009/030481中,其公开内容通过引用结合在本文中。

[1024] 进一步地,当根据式 $\{(\text{Arg})_l; (\text{Lys})_m; (\text{His})_n; (\text{Orn})_o; (\text{Xaa})_x\}$ (式(I))限定的,如上文显示的,并包含或另外修饰以包含至少一个-SH部分的所述阳离子或聚阳离子肽或蛋白,可以,并不限于,选自分式(Ia):

[1025]  $\{(\text{Arg})_l; (\text{Lys})_m; (\text{His})_n; (\text{Orn})_o; (\text{Xaa}')_x (\text{Cys})_y\}$  分式(Ia)

[1026] 其中 $(\text{Arg})_l; (\text{Lys})_m; (\text{His})_n; (\text{Orn})_o$ ;和 $x$ 为如本文限定的,Xaa'是选自除了Arg、Lys、His、Orn或Cys之外的天然(=天然存在的)或非天然氨基酸的任何氨基酸,并且假如Arg(精氨酸)、Lys(赖氨酸)、His(组氨酸)和Orn(鸟氨酸)的总含量代表寡肽的所有氨基酸的至少10%, $y$ 是选自0,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21-30,31-40,41-50,51-60,61-70,71-80和81-90的任何数。在此情形中,W0 2012/013326的公开内容通过引用结合在本文中。进一步地,所述阳离子或聚阳离子肽可以选自分式(Ib):

[1027]  $\text{Cys}_1\{(\text{Arg})_l; (\text{Lys})_m; (\text{His})_n; (\text{Orn})_o; (\text{Xaa})_x\}\text{Cys}_2$  分式(Ib)



[1028] 其中经验式 $\{(\text{Arg})_1; (\text{Lys})_m; (\text{His})_n; (\text{Orn})_o; (\text{Xaa})_x\}$  (式(III))如本文限定,并且形成(半经验)式(III)所述的氨基酸序列的核心,并且其中 $\text{Cys}_1$ 和 $\text{Cys}_2$ 是邻近或位于 $(\text{Arg})_1; (\text{Lys})_m; (\text{His})_n; (\text{Orn})_o; (\text{Xaa})_x$ 末端的半胱氨酸。在此情形中,WO 2011/026641的公开内容通过引用结合在本文中。

[1029] 可以用作转染或复合剂的特别优选的阳离子或聚阳离子化合物可以包括阳离子多糖,例如,壳聚糖,聚凝胺(polybrene),阳离子聚合物,例如,聚乙烯亚胺(PEI),阳离子脂质,例如,DOTMA:[1-(2,3-二油酰氧基)丙基]-N,N,N-三甲基氯化铵,DMRIE,二- $\text{C}_{14}$ -脒,DOTIM,SAINT,DC-Chol,BGTC,CTAP,DOPC,DODAP,DOPE:二油酰磷脂酰乙醇胺,DOSPA,DODAB,DOIC,DMEPC,DOGS:双十八基酰胺基甘氨酰精胺(Dioctadecylamidoglycylspermin),DIMRI:二肉豆蔻酰-氧丙基二甲基羟乙基溴化铵,DOTAP:二油酰基氧基-3-(三甲基铵基)丙烷,DC-6-14:0,0-双十四酰基-N-( $\alpha$ -三甲基铵基乙酰基)二乙醇胺氯化物,CLIP1:外消旋-[2(2,3-双十八基氧基丙基)(2-羟乙基)]-二甲基氯化铵,CLIP6:外消旋-[2(2,3-双十六基氧基丙基-氧基甲氧基)乙基]三甲基铵,CLIP9:外消旋-[2(2,3-双十六基氧基丙基-氧基琥珀酰基氧基)乙基]-三甲基铵,oligofectamine,或阳离子或聚阳离子聚合物,例如,改性聚氨基酸,诸如 $\beta$ -氨基酸-聚合物或反向聚酰胺(reversed polyamides)等,改性的聚乙烯,诸如PVP(聚(N-乙基-4-溴化乙烯基吡啶鎓))等,改性的丙烯酸酯,诸如pDMAEMA(聚(二甲基氨基乙基丙烯酸甲酯))等,改性的氨基胺(amidoamines),诸如pAMAM(聚(氨基胺))等,改性的聚 $\beta$ 氨基酯(PBAE),诸如二胺末端修饰的1,4丁二醇双丙烯酸酯-共-5-氨基-1-戊醇聚合物等,树枝状聚合物(dendrimers),诸如聚丙胺树枝状聚合物或基于pAMAM的树枝状聚合物等,聚亚胺,诸如PEI:聚(乙烯亚胺),聚(丙烯亚胺)等,聚烯丙基胺,基于糖骨架的聚合物,诸如基于环糊精的聚合物,基于葡聚糖的聚合物,壳聚糖等,基于甲硅烷骨架的聚合物,诸如PMOXA-PDMS共聚物等,由一个或多个阳离子嵌段(例如选自上文提及的阳离子聚合物)的组合和一个或多个亲水或疏水嵌段(例如,聚乙二醇)的组合组成的嵌段聚合物;等等。

[1030] 根据另一实施方案,本发明所述的药物组合物可以包含佐剂以增强药物组合物的免疫刺激性质。在该情况下,佐剂可以理解为任何适于支持本发明所述的药物组合物中包含的组分如人工核酸分子或载体的施用和递送的化合物。此外,所述佐剂可以(但不限于其中)起始或增加先天免疫系统的免疫应答,即非特异性免疫应答。换句话说,当施用时,本发明所述的药物组合物典型地起始针对人工核酸分子编码的抗原的适应性免疫应答。此外,由于将如本文定义的佐剂加入至本发明所述的药物组合物,本发明所述的药物组合物可以产生(支持性的)先天免疫应答。

[1031] 所述佐剂可以选自技术人员已知的和适于目前情形,即支持在哺乳动物中诱导免疫反应的佐剂。优选为,佐剂可以是选自但不限于由以下各项组成的组:TDM,MDP,胞壁酰二肽,普流罗尼克(pluronics),明矾溶液,氢氧化铝,ADJUMER<sup>TM</sup>(聚磷腈);磷酸铝凝胶;来自藻类的葡聚糖类;algammulin;氢氧化铝凝胶(明矾);高蛋白吸附的氢氧化铝凝胶;低粘度氢氧化铝凝胶;AF或SPT(角鲨烷乳液(5%)),吐温80(0.2%),普流罗尼克(Pluronic)L121(1.25%),磷酸缓冲盐水,pH 7.4);AVRIDINE<sup>TM</sup>(丙二胺);BAY R1005<sup>TM</sup>(N-(2-脱氧-2-L-亮氨酰氨基-b-D-吡喃葡糖基)-N-十八基-十二烷酰-酰胺氢乙酸盐);CALCITRIOL<sup>TM</sup>(1- $\alpha$ ,25-二羟基-维生素D3);磷酸钙凝胶;CAP<sup>TM</sup>(磷酸钙纳米颗粒);霍乱全毒素,霍乱-毒素-A1-蛋白-A-D-片段融合蛋白,霍乱毒素的亚基B;CRL 1005(嵌段共聚物P1205);含有细胞因子的

脂质体;DDA(溴化二甲基双十八基铵);DHEA(脱氢表雄酮);DMPC(二肉豆蔻酰磷脂酰胆碱);DMPG(二肉豆蔻酰磷脂酰甘油);DOC/明矾复合物(脱氧胆酸钠盐);完全弗氏佐剂;不完全弗氏佐剂; $\gamma$  菊糖;Gerbu佐剂(是下述的混合物:i)N-乙酰葡萄糖胺基-(P1-4)-N-乙酰胞壁酰基-L-丙氨酸-D-谷氨酰胺(GMDP),ii)氯化二甲基双十八基铵(DDA),iii)锌-L-脯氨酸盐复合物(ZnPro-8);GM-CSF);GMDP(N-乙酰葡萄糖胺基-(b1-4)-N-乙酰胞壁酰基-L-丙氨酸-D-异谷氨酰胺);咪喹莫特(imiquimod)(1-(2-甲基丙基)-1H-咪唑并[4,5-c]喹啉-4-胺);ImmTher™(N-乙酰葡萄糖胺基-N-乙酰胞壁酰基-L-Ala-D-异Glu-L-Ala-甘油二棕榈酸酯);DRVs(由脱水-再水合囊泡制备的免疫脂质体);干扰素- $\gamma$ ;白介素-1 $\beta$ ;白介素-2;白介素-7;白介素-12;ISCOMSTM;ISCOPREP 7.0.3.TM;脂质体;LOXORIBINE™(7-烯丙基-8-氧代鸟苷);LT口服佐剂(大肠杆菌(E.coli)不稳定的肠毒素-原毒素);任意组成的微球体和微粒;MF59TM;(角鲨烯-水乳液);MONTANIDE ISA 51™(纯化的不完全弗氏佐剂);MONTANIDE ISA 720™(可代谢的油佐剂);MPL™(3-Q-脱酰基-4'-单磷酸基脂质A);MTP-PE和MTP-PE脂质体((N-乙酰基-L-丙氨酸-D-异谷氨酰基-L-丙氨酸-2-(1,2-二棕榈酰-sn-甘油-3-(羟基磷酸基氧基))-乙基酰胺,一钠盐);MURAMETIDE™(Nac-Mur-L-Ala-D-Gln-OCH3);MURAPALMITINETM和D-MURAPALMITINE™(Nac-Mur-L-Thr-D-异GIn-sn-甘油二棕榈酰基);NAGO(神经氨酸酶-半乳糖氧化酶);任意组成的纳米球体或纳米颗粒;NISVs(非离子表面活性剂囊泡);PLEURAN™(-葡聚糖);PLGA,PGA和PLA(乳酸和乙醇酸的均聚体和共聚体;微球体/纳米球体);普流罗尼克(PLURONIC)L121TM;PMMA(聚甲基丙烯酸甲酯);PODDS™(类蛋白微球体);聚乙烯氨基甲酸酯衍生物;聚-rA:聚-rU(聚腺苷酸-多聚尿苷酸复合物);聚山醇酯80(吐温80);蛋白螺旋物(protein cochleates)(Avanti Polar Lipids,Inc.,Alabaster,AL);STIMULON™(QS-21);Quil-A(Quil-A皂苷);S-28463(4-氨基-otec-二甲基-2-乙氧基甲基-1H-咪唑并[4,5c]喹啉-1-乙醇);SAF-1™("Syntex佐剂制剂");仙台蛋白脂质体和仙台含脂质基质;Span-85(三油酸脱水山梨酯);Specol(Marcol 52,Span 85和吐温85的乳液);角鲨烯或Robane®(2,6,10,15,19,23-六甲基二十四烷和2,6,10,15,19,23-六甲-2,6,10,14,18,22-二十四碳六烯);硬脂酰酪氨酸(十八基酪氨酸盐酸盐);Theramid®(N-乙酰葡萄糖胺基-N-乙酰胞壁酰基-L-Ala-D-异Glu-L-Ala-二棕榈酰氧基丙酰胺);苏氨酸-MDP(Termurtide™或[thr 1]-MDP;N-乙酰胞壁酰基-L-苏氨酸-D-异谷氨酰胺);Ty颗粒(Ty-VLP或病毒样颗粒);Walter-Reed脂质体(包含吸附在氢氧化铝上的脂质A的脂质体),和脂肽,包括Pam3Cys,特别是铝盐,如Adju-phos,Alhydrogel,Rehydragel;乳液,包括CFA,SAF,IFA,MF59,Provax,TiterMax,Montanide,Vaxfectin;共聚物,包括Optivax(CRL1005),L121,Poloxamer4010),等;脂质体,包括Stealth,螺旋物,包括BIORAL;植物来源的佐剂,包括QS21,Quil A,Iscomatrix,ISCOM;适于共刺激的佐剂,包括蕃茄素(Tomatine),生物聚合物,包括PLG,PMM,菊糖;微生物来源的佐剂,包括罗莫肽(Romurtide),DETOX,MPL,CWS,甘露糖,CpG核酸序列,CpG7909,人TLR 1-10的配体,鼠TLR 1-13的配体,ISS-1018,IC31,咪唑并喹啉,聚肌胞(Ampligen),Ribi529,IMOxine,IRIVs,VLPs,霍乱毒素,不耐热毒素,Pam3Cys,鞭毛蛋白(Flagellin),GPI锚定子,LNFP III/Lewis X,抗微生物肽,UC-1V150,RSV融合蛋白,cdiGMP;和适于作为拮抗剂的佐剂,包括CGRP神经肽。

[1032] 合适的佐剂还可以选自阳离子或聚阳离子化合物,其中所述佐剂优选为基于将药

物组合物的人工核酸分子或载体与阳离子或聚阳离子化合物复合来制备。将药物组合物的人工核酸分子或载体与如本文定义的阳离子或聚阳离子化合物关联或复合优选提供佐剂性质并且赋予药物组合物的人工核酸分子或载体稳定的效果。尤其是如此优选的,此类阳离子或聚阳离子化合物选自阳离子或聚阳离子肽或蛋白,包括鱼精蛋白,核仁蛋白,精胺或亚精胺,或其它阳离子肽或蛋白,如聚-L-赖氨酸(PLL),聚精氨酸,碱性多肽,细胞穿透肽(CPP),包括HIV-结合肽,Tat,HIV-1Tat(HIV),源自Tat的肽,穿透肽(Penetratin),源自VP22的或类似物肽,HSV VP22(单纯疱疹病毒),MAP,KALA或蛋白转导结构域(PTD,PpT620,富含脯氨酸的肽,富含精氨酸的肽,富含赖氨酸的肽,MPG-肽,Pep-1,L-寡聚物,降钙素肽,源自触角足的肽(尤其是来自果蝇触角足),pAntp,pIsl,FGF,乳铁蛋白,Transportan,Buforin-2,Bac715-24,SynB,SynB(1),pVEC,源自hCT的肽,SAP,鱼精蛋白,精胺,亚精胺,或组蛋白。其他优选的阳离子或聚阳离子化合物可以包括阳离子多糖,例如壳聚糖,聚凝胺,阳离子聚合物,例如聚乙烯亚胺(PEI),阳离子脂,例如DOTMA:1-(2,3-二油酰氧基)丙基)-N,N,N-三甲基氯化铵,DMRIE,二-C14-脘,DOTIM,SAINT,DC-Chol,BGTC,CTAP,DOPC,DODAP,DOPE:二油酰基磷脂酰乙醇胺,DOSPA,DODAB,DOIC,DMEPC,DOGS:双十八烷基酰胺基甘氨酸精胺(Dioctadecylamidoglycylspermin),DIMRI:二肉豆蔻酰-氧基丙基二甲基羟乙基溴化铵,DOTAP:二油酰氧基-3-(三甲基胺基)丙烷,DC-6-14:0,0-双十四酰基-N-(-三甲基胺基乙酰)氯化二乙醇胺,CLIP1:外消旋-[2(2,3-双十八烷氧基丙基)(2-羟乙基)]-二甲基氯化铵,CLIP6:外消旋-[2(2,3-双十六烷氧基丙基-氧甲氧基)乙基]三甲基铵,CLIP9:外消旋-[2(2,3-双十六烷氧基丙基-氧琥珀酰氧)乙基]-三甲基铵,oligofectamine,或阳离子或聚阳离子聚合物,例如改性聚氨基酸,如-氨基酸-聚合物或反向聚酰胺等,改性的聚乙烯,如PVP(聚(N-乙基-4-溴化乙烯基吡啶鎓))等,改性的丙烯酸盐,如pDMAEMA(聚(二甲基氨基乙基丙烯酸甲酯))等,改性的酰胺基胺如pAMAM(聚(酰胺基胺))等,改性的聚β氨基酯(PBAE),如二胺末端修饰的1,4丁二醇二丙烯酸酯-共-5-氨基-1-戊醇聚合物等,树状聚合物,如聚丙基胺树状聚合物或基于pAMAM的树状聚合物等,聚亚胺,如PEI:聚(乙烯亚胺),聚(丙烯亚胺)等,聚烯丙基胺,基于糖骨架的聚合物,如基于环糊精的聚合物,基于葡聚糖的聚合物,壳聚糖等,基于硅烷骨架的聚合物,如PMOXA-PDMS共聚物等,由一个以上阳离子嵌段(例如选自如上文提到的阳离子聚合物)的组合和一个以上亲水或疏水嵌段(如聚乙二醇)的组合组成的嵌段聚合物;等等。

[1033] 此外,可以通过将组合物的人工核酸分子或载体、优选RNA复合而用作佐剂的优选的阳离子或聚阳离子蛋白或肽可以选自以下具有以下总式(I)的蛋白或肽:  $(\text{Arg})_1; (\text{Lys})_m; (\text{His})_n; (\text{Orn})_o; (\text{Xaa})_x$ , 其中  $1+m+n+o+x=8-15$ , 并且  $1, m, n$  或  $o$  可以彼此独立是选自  $0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14$  或  $15$  的任意数, 条件是 Arg、Lys、His 和 Orn 的总含量代表寡肽的所有氨基酸的至少 50%; 并且 Xaa 可以是选自除了 Arg、Lys、His 或 Orn 之外的天然(=天然存在)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  $4$  的任意数, 条件是 Xaa 的总含量不超过寡肽的所有氨基酸的 50%。在该情况下尤其优选的寡精氨酸是例如  $\text{Arg}_7, \text{Arg}_8, \text{Arg}_9, \text{Arg}_7, \text{H}_3\text{R}_9, \text{R}_9\text{H}_3, \text{H}_3\text{R}_9\text{H}_3, \text{YSSR}_9\text{SSY}, (\text{RKH})_4, \text{Y}(\text{RKH})_2\text{R}$  等。

[1034] 人工核酸或载体与阳离子或聚阳离子化合物的比例可以基于整个核酸复合物的氮/磷比(N/P-比)确定。例如,  $1\mu\text{g}$  RNA 通常含有约  $3\text{nmol}$  磷酸酯残基, 条件是 RNA 显示碱基的统计学分布。此外,  $1\mu\text{g}$  肽通常含有约  $x\text{ nmol}$  氮残基, 这取决于碱性氨基酸的分子量和数量。

当示例性地对 (Arg)<sub>9</sub> (分子量1424g/mol, 9个氮原子) 进行计算时, 1μg (Arg)<sub>9</sub> 含有约 700pmol (Arg)<sub>9</sub>, 并且因此  $700 \times 9 = 6300 \text{ pmol 碱性氨基酸} = 6.3 \text{ nmol 氮原子}$ 。对于约 1:1RNA/(Arg)<sub>9</sub> 的质量比, 可以计算 N/P 比为约 2。当示例性地针对鱼精蛋白 (当使用来自鲑鱼的鱼精蛋白时, 分子量约 4250g/mol, 21 个氮原子) 与 2μg RNA 以约 2:1 的质量比计算时, 对 RNA 计算为 6nmol 磷酸酯; 1μg 鱼精蛋白含有约 235pmol 鱼精蛋白分子, 并且因此  $235 \times 21 = 4935 \text{ pmol 碱性氮原子} = 4.9 \text{ nmol 氮原子}$ 。对于约 2:1RNA/鱼精蛋白的质量比, 可以计算出约 0.81 的 N/P 比。对于约 8:1RNA/鱼精蛋白的质量比, 可以计算约 0.2 的 N/P 比。在本发明的情况下, 关于复合物中核酸:肽的比例, N/P-比优选在约 0.1-10 的范围内, 优选在约 0.3-4 的范围内, 并且最优选在约 0.5-2 或 0.7-2 的范围内, 并且最优选在约 0.7-1.5 的范围内。

[1035] 专利申请 W02010/037539 (其公开内容通过引用并入本文) 描述了免疫刺激组合物和用于制备免疫刺激组合物的方法。因此, 在本发明的优选的实施方案中, 以两个分开的步骤获得组合物, 从而获得本发明所述的人工核酸分子的有效免疫刺激作用和有效翻译。其中, 通过在第一步骤中将佐剂组分的人工核酸分子或载体、优选 RNA 与阳离子或聚阳离子化合物以特定比例复合从而形成稳定复合体来制备所谓的“佐剂组分”。在该情况下, 重要的是, 在复合核酸后, 没有游离阳离子或聚阳离子化合物或仅有可忽略的少量留存在佐剂组分中。因此, 佐剂组分中核酸和阳离子或聚阳离子化合物的比例通常在核酸被整体复合并且在组合物中没有游离阳离子或聚阳离子化合物或仅有可忽略的少量残留的范围内选择。优选地, 佐剂组分的比例, 即核酸与阳离子或聚阳离子化合物的比例, 选自约 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) 的范围内, 并且最优选是约 3:1 (w/w) 至约 2:1 (w/w) 的比例。

[1036] 根据优选的实施方案, 在第二步, 将本发明所述的人工核酸分子或载体, 优选 RNA 分子, 加入至佐剂组分的复合的核酸分子, 优选 RNA 中, 从而形成本发明的 (免疫刺激) 组合物。其中, 本发明的人工核酸分子或载体, 优选 RNA 以游离核酸加入, 即未与其他化合物复合的核酸。加入之前, 游离的人工核酸分子或载体未复合并且将优选在加入佐剂组分时不经历任何可检测或显著的复合反应。

[1037] 此外, 合适的佐剂可以选自具有式 (II) 的核酸: G<sub>1</sub>X<sub>m</sub>G<sub>n</sub>, 其中: G 是鸟苷 (鸟嘌呤)、尿苷 (尿嘧啶) 或是鸟苷 (鸟嘌呤) 或尿苷 (尿嘧啶) 的类似物; X 是鸟苷 (鸟嘌呤)、尿苷 (尿嘧啶)、腺苷 (腺嘌呤)、胸苷 (胸腺嘧啶)、胞苷 (胞嘧啶) 或上述核苷酸 (核苷) 的类似物; 1 是 1 至 40 的整数, 其中当 1=1 时, G 是鸟苷 (鸟嘌呤) 或其类似物, 当 1>1 时, 至少 50% 的核苷酸 (核苷) 是鸟苷 (鸟嘌呤) 或其类似物; m 是整数并至少是 3; 其中当 m=3 时, X 是尿苷 (尿嘧啶) 或其类似物, 当 m>3 时, 存在至少 3 个连续的尿苷 (尿嘧啶) 或尿苷 (尿嘧啶) 类似物; n 是 1 至 40 的整数, 其中当 n=1 时, G 是鸟苷或其类似物, 当 n>1 时, 至少 50% 的核苷酸 (核苷) 是鸟苷 (鸟嘌呤) 或其类似物。

[1038] 此外, 其他合适的佐剂可以选自具有式 (III) 的核酸: C<sub>1</sub>X<sub>m</sub>C<sub>n</sub>, 其中: C 是胞苷 (胞嘧啶)、尿苷 (尿嘧啶) 或是胞苷 (胞嘧啶) 或尿苷 (尿嘧啶) 的类似物; X 是鸟苷 (鸟嘌呤)、尿苷 (尿嘧啶)、腺苷 (腺嘌呤)、胸苷 (胸腺嘧啶)、胞苷 (胞嘧啶) 上述核苷酸 (核苷) 的类似物; 1 是 1 至 40 的整数, 其中当 1=1 时, C 是胞嘧啶或其类似物, 当 1>1 时, 至少 50% 的核苷酸 (核苷) 是胞苷 (胞嘧啶) 或其类似物; m 是整数并且至少是 3; 其中当 m=3 时, X 是尿苷 (尿嘧啶) 或其类似物, 当 m>3 时, 存在至少 3 个连续的尿苷 (尿嘧啶) 或尿苷 (尿嘧啶) 的类似物; n 是 1 至 40 的整

数,其中当 $n=1$ 时,C是胞苷(胞嘧啶)或其类似物,当 $n>1$ 时,至少50%的核苷酸(核苷)是胞苷(胞嘧啶)或其类似物。

[1039] 本发明所述的药物组合物优选包含“安全且有效量”的药物组合物组分,尤其是如本文限定的本发明的人工核酸分子、载体和/或细胞。如本文使用的,“安全且有效量”意为足以显著诱导如本文限定的疾病或病症的积极改进的量。然而同时,“安全且有效量”优选避免严重的副作用并且允许益处和风险间的合理关系。这些限制的确定典型地处于合理的医学判断范围中。

[1040] 在另一个方面,本发明提供(在基因接种中)或在基因治疗中用作药物,例如,用作疫苗的本发明所述的人工核酸分子、本发明所述的载体、本发明所述的细胞、或本发明所述的药物组合物。

[1041] 本发明所述的人工核酸分子,本发明所述的载体,本发明所述的细胞,或本发明所述的药物组合物尤其适于任何医学应用,其使用肽、多肽或蛋白的治疗作用或效果,或其中需要特定的肽或蛋白的补充。因此,本发明提供本发明所述的人工核酸分子、本发明所述的载体、本发明所述的细胞、或本发明所述的药物组合物用于治疗或预防接受通过肽、多肽或蛋白的治疗作用或效果的治疗或接受通过补充特定肽、多肽或蛋白的治疗的疾病或病症。例如,本发明所述的人工核酸分子,本发明所述的载体,本发明所述的细胞,或本发明所述的药物组合物可以用于例如,通过基因接种或基因治疗而治疗或预防遗传病,自身免疫疾病,癌症或肿瘤相关疾病,传染病,慢性病等。

[1042] 具体地,从在要被治疗的受试者中治疗性肽、多肽或蛋白的增加和延长的存在获益的此种治疗性治疗尤其适合作为本发明范围内的医学应用,因为本发明的3'-UTR元件提供本发明的人工核酸分子或载体的编码的肽或蛋白的稳定和延长的表达,和/或本发明的5'-UTR元件提供本发明的人工核酸分子或载体的编码的肽或蛋白的增加的表达。因此,本发明所述的人工核酸分子、本发明所述的载体、本发明所述的细胞或本发明所述的药物组合物的尤其合适的医学应用是疫苗接种。因此,本发明提供用于接种受试者,优选哺乳动物受试者,更优选人受试者的本发明所述的人工核酸分子、本发明所述的载体、本发明所述的细胞,或本发明所述的药物组合物。优选的疫苗接种治疗是针对传染病,如细菌、原虫或病毒感染的接种,和抗肿瘤-接种。此种接种治疗可以是预防性的或治疗性的。

[1043] 根据要治疗或预防的疾病,可以选择ORF。例如,所述可读框可以编码必须提供给患有完全缺少或至少部分缺失蛋白的功能的患者,如患有遗传病的患者的蛋白。另外,所述可读框可以选自编码有利影响受试者的疾病或病状的肽或蛋白的ORF。此外,所述可读框可以编码有效下调天然肽或蛋白的病理过量生产或消除病理表达蛋白或肽的细胞的肽或蛋白。此种功能的缺少、缺失或过量生产可以,例如,发生在肿瘤和瘤形成、自身免疫疾病、过敏、感染、慢性病等的情况下。此外,所述可读框可以编码抗原或免疫原,例如编码病原体的表位或编码肿瘤抗原。因此,在优选的实施方案中,本发明所述的人工核酸分子或载体包含编码包含以下或由以下组成的氨基酸序列的ORF:抗原或免疫原,例如病原体的表位或肿瘤相关抗原,如上文描述的3'-UTR元件和/或上文所述的5'-UTR元件,和任选的其他组分,例如聚腺苷酸序列等。

[1044] 在医学应用的情况下,具体地,在疫苗接种的情况下,优选本发明所述的人工核酸分子是RNA,优选mRNA,原因在于,DNA携带引起抗-DNA免疫应答的风险并且倾向于插入基因

组DNA。然而,在一些实施方案中,例如,如果病毒递送载体,如腺病毒递送载体用于递送本发明所述的人工核酸分子或载体,例如,在基因治疗性治疗的情况下,所述人工核酸分子或所述载体是DNA分子可以是合乎需要的。

[1045] 本发明所述的人工核酸分子、本发明所述的载体、本发明所述的细胞或本发明所述的药物组合物可以口服,肠胃外,通过吸入喷雾,局部,直肠,经鼻,经颊,阴道,通过植入的储器或通过喷射注射施用。如本文使用的术语肠胃外包括皮下,静脉内,肌肉内,关节内,滑膜内,胸骨内,鞘内,肝内,病灶内,颅内,经皮,皮内,肺内,腹膜内,心脏内,动脉内,和舌下注射或输注技术。在优选的实施方案中,本发明所述的人工核酸分子、本发明所述的载体、本发明所述的细胞或本发明所述的药物组合物经由无针头注射(例如喷射注射)施用。

[1046] 优选地,本发明所述的人工核酸分子,本发明所述的载体,本发明所述的细胞,或本发明所述的药物组合物例如通过肠胃外注射,更优选通过皮下,静脉内,肌肉内,关节内,滑膜内,胸骨内,鞘内,肝内,病灶内,颅内,经皮,皮内,肺内,腹膜内,心脏内,动脉内,舌下注射或通过输注技术被肠胃外施用。尤其优选的是皮内和肌肉内注射。本发明的药物组合物无菌可注射形式可以是水质或油质悬液。可以根据本领域已知的技术,使用合适的分散或湿润剂和悬浮剂制备这些悬液。优选地,所述溶液或悬液经由无针注射(例如喷射注射)施用。

[1047] 本发明所述的人工核酸分子,本发明所述的载体,本发明所述的细胞,或本发明所述的药物组合物可以以任何口服可接受剂型口服施用,所述口服可接受剂型包括但不限于胶囊、片剂、水质悬液或溶液。

[1048] 尤其当治疗靶标包括容易通过局部应用可到达的区域或器官,例如包括皮肤疾病或任何其它可到达的上皮组织的疾病时,本发明所述的人工核酸分子,本发明所述的载体,本发明所述的细胞,或本发明所述的药物组合物还可以局部施用。容易制备用于这些区域或器官中的每个的合适的局部制剂。为了局部应用,可以以混悬或溶解在一种或多种载体中的合适的软膏配制本发明所述的人工核酸分子、本发明所述的载体、本发明所述的细胞或本发明所述的药物组合物。

[1049] 在一个实施方案中,作为药物的使用包括转染哺乳动物细胞,优选体外或离体转染哺乳动物细胞,更优选体外转染要由所述药物治疗的受试者的分离的细胞的步骤。如果所述使用包括体外转染分离的细胞,则作为药物的使用还可以包括向患者再施用转染的细胞。本发明的人工核酸分子或载体作为药物的使用还可以包括选择成功转染的分离的细胞的步骤。因此,如果载体还包含选择标记,其可以是有利的。而且,作为药物的使用可以包括体外转染分离的细胞和从这些细胞纯化表达产物,即编码的肽或蛋白。该纯化的肽或蛋白可以随后施用于需要其的受试者。

[1050] 本发明还提供一种治疗或预防如上文描述的疾病或病症的方法,所述方法包括向需要其的受试者施用本发明所述的人工核酸分子,本发明所述的载体,本发明所述的细胞,或本发明所述的药物组合物。

[1051] 此外,本发明提供一种治疗或预防疾病或病症的方法,所述方法包括用本发明所述的人工核酸分子或用本发明所述的载体转染细胞。所述转染可以在体外、离体或体内进行。在一个优选的实施方案中,体外进行细胞转染并且将转染的细胞施用于需要其的受试者,优选人患者。优选地,要体外转染的细胞是受试者的分离的细胞,优选人患者的分离的

细胞。因此,本发明提供一种治疗方法,所述方法包括从受试者,优选从人患者分离细胞,用本发明所述的人工核酸或本发明所述的载体转染所述分离的细胞,并将转染的细胞施用于所述受试者,优选人患者的步骤。

[1052] 治疗或预防根据本发明的病症的方法优选是如上文描述的疫苗接种方法或基因治疗方法。

[1053] 如上文所述的,本发明的3'-UTR元件和/或本发明的5'-UTR元件能够延长和/或增加从mRNA的蛋白产生。由此,在进一步的方面,本发明涉及用于增加和/或延长从人工核酸分子,优选从mRNA分子或载体的蛋白产生的方法,所述方法包括将可读框与3'-UTR元件和/或5'-UTR元件关联的步骤,其中3'-UTR元件和/或5'-UTR元件为得到的人工核酸分子提供高翻译效率,从而获得上文所述的本发明的人工核酸分子,优选mRNA分子或上文所述的本发明的载体。

[1054] 优选地,在增加和/或延长从人工核酸分子,优选从mRNA分子或载体的蛋白产生的方法中,根据本发明,所述3'-UTR元件和/或5'-UTR元件包含这样的核酸序列或由这样的核酸序列组成,所述核酸序列源自选自以下各项组成的组的基因的转录本的3'-UTR和/或5'-UTR: ZNF460, TGM2, IL7R, BGN, TK1, RAB3B, CBX6, FZD2, COL8A1, NDUFS7, PHGDH, PLK2, TSP0, PTGS1, FBX032, NID2, ATP5D, EXOSC4, NOL9, UBB4B, VPS18, ORMDL2, FSCN1, TMEM33, TUBA4A, EMP3, TMEM201, CRIP2, BRAT1, SERPINH1, CD9, DPYSL2, CDK9, TFRC, PSMB3, FASN, PSMB6, PRSS56, KPNA6, SFT2D2, PARD6B, LPP, SPARC, SCAND1, VASN, SLC26A1, LCLAT1, FBXL18, SLC35F6, RAB3D, MAP1B, VMA21, CYBA, SEZ6L2, PCOLCE, VTN, ALDH16A1, RAVER1, KPNA6, SERINC5, JUP, CPN2, CRIP2, EPT1, PNPO, SSSCA1, POLR2L, LIN7C, UQCR10, PYCRL, AMN, MAP1S, NDUFS7, PHGDH, TSP0, ATP5D, EXOSC4, TUBB4B, TUBA4A, EMP3, CRIP2, BRAT1, CD9, CDK9, PSMB3, PSMB6, PRSS56, SCAND1, AMN, CYBA, PCOLCE, MAP1S, VTN, ALDH16A1 (全部优选是人的) 和 Dpysl2, Ccnd1, Acox2, Cbx6, Ubc, Ldlr, Nudt22, Pcyox11, Ankrd1, Tmem37, Tspyl4, Slc7a3, Cst6, Aacs, Nosip, Itga7, Ccnd2, Ebp, Sf3b5, Fasn, Hmgcs1, Osr1, Lmnb1, Vma21, Kif20a, Cdca8, Slc7a1, Ubqln2, Prps2, Shmt2, Aurkb, Figl1, Cad, Anln, Slfn9, Ncaph, Pole, Uhrf1, Gja1, Fam64a, Kif2c, Tspan10, Scand1, Gpr84, Fads3, Cers6, Cxcr4, Gprc5c, Fen1, Cspg4, Mrpl34, Comtd1, Armc6, Emr4, Atp5d, 1110001J03Rik, Csf2ra, Aarsd1, Kif22, Cth, Tpgs1, Ccl17, Alkbh7, Ms4a8a, Acox2, Ubc, Slpi, Pcyox11, Igf2bp1, Tmem37, Slc7a3, Cst6, Ebp, Sf3b5, Plk1, Cdca8, Kif22, Cad, Cth, Pole, Kif2c, Scand1, Gpr84, Tpgs1, Ccl17, Alkbh7, Ms4a8a, Mrpl34, Comtd1, Armc6, Atp5d, 1110001J03Rik, Nudt22, Aarsd1 (全部优选是小鼠的); 优选选自下述组成的组: ZNF460-5'-UTR, TGM2-5'-UTR, IL7R-5'-UTR, BGN-5'-UTR, TK1-5'-UTR, RAB3B-5'-UTR, CBX6-5'-UTR, FZD2-5'-UTR, COL8A1-5'-UTR, NDUFS7-5'-UTR, PHGDH-5'-UTR, PLK2-5'-UTR, TSP0-5'-UTR, PTGS1-5'-UTR, FBX032-5'-UTR, NID2-5'-UTR, ATP5D-5'-UTR, EXOSC4-5'-UTR, NOL9-5'-UTR, UBB4B-5'-UTR, VPS18-5'-UTR, ORMDL2-5'-UTR, FSCN1-5'-UTR, TMEM33-5'-UTR, TUBA4A-5'-UTR, EMP3-5'-UTR, TMEM201-5'-UTR, CRIP2-5'-UTR, BRAT1-5'-UTR, SERPINH1-5'-UTR, CD9-5'-UTR, DPYSL2-5'-UTR, CDK9-5'-UTR, TFRC-5'-UTR, PSMB3 5'-UTR, FASN-5'-UTR, PSMB6-5'-UTR, PRSS56-5'-UTR, KPNA6-5'-UTR, SFT2D2-5'-UTR, PARD6B-5'-UTR, LPP-5'-UTR, SPARC-5'-UTR, SCAND1-5'-UTR, VASN-5'-UTR, SLC26A1-5'-UTR, LCLAT1-5'-UTR, FBXL18-5'-UTR, SLC35F6-5'-UTR, RAB3D-5'-

UTR,MAP1B-5'-UTR,VMA21-5'-UTR,CYBA-5'-UTR,SEZ6L2-5'-UTR,PCOLCE-5'-UTR,VTN-5'-UTR,ALDH16A1-5'-UTR,RAVER1-5'-UTR,KPNA6-5'-UTR,SERINC5-5'-UTR,JUP-5'-UTR,CPN2-5'-UTR,CRIP2-5'-UTR,EPT1-5'-UTR,PNPO-5'-UTR,SSSCA1-5'-UTR,POLR2L-5'-UTR,LIN7C-5'-UTR,UQCR10-5'-UTR,PYCRL-5'-UTR,AMN-5'-UTR,MAP1S-5'-UTR,(全部优选是人的)和Dpysl2-5'-UTR,Ccnd1-5'-UTR,Acox2-5'-UTR,Cbx6-5'-UTR,Ubc-5'-UTR,Ldlr-5'-UTR,Nudt22-5'-UTR,Pcyox11-5'-UTR,Ankrd1-5'-UTR,Tmem37-5'-UTR,Tspyl4-5'-UTR,Slc7a3-5'-UTR,Cst6-5'-UTR,Aacs-5'-UTR,Nosip-5'-UTR,Itga7-5'-UTR,Ccnd2-5'-UTR,Ebp-5'-UTR,Sf3b5-5'-UTR,Fasn-5'-UTR,Hmgcs1-5'-UTR,Osr1-5'-UTR,Lmnb1-5'-UTR,Vma21-5'-UTR,Kif20a-5'-UTR,Cdca8-5'-UTR,Slc7a1-5'-UTR,Ubqln2-5'-UTR,Prps2-5'-UTR,Shmt2-5'-UTR,Aurkb-5'-UTR,Fignl1-5'-UTR,Cad-5'-UTR,Anln-5'-UTR,Slfn9-5'-UTR,Ncaph-5'-UTR,Pole-5'-UTR,Uhrf1-5'-UTR,Gjal-5'-UTR,Fam64a-5'-UTR,Kif2c-5'-UTR,Tspan10-5'-UTR,Scand1-5'-UTR,Gpr84-5'-UTR,Fads3-5'-UTR,Cers6-5'-UTR,Cxcr4-5'-UTR,Gprc5c-5'-UTR,Fen1-5'-UTR,Cspg4-5'-UTR,Mrpl34-5'-UTR,Comtd1-5'-UTR,Armc6-5'-UTR,Emr4-5'-UTR,Atp5d-5'-UTR,1110001J03Rik-5'-UTR,Csf2ra-5'-UTR,Aarsd1-5'-UTR,Kif22-5'-UTR,Cth-5'-UTR,Tpgs1-5'-UTR,Cc117-5'-UTR,Alkbh7-5'-UTR,Ms4a8a-5'-UTR(全部优选是小鼠的);和NDUFS7-3'-UTR,PHGDH-3'-UTR,TSP0-3'-UTR,ATP5D-3'-UTR,EXOSC4-3'-UTR,TUBB4B-3'-UTR,TUBA4A-3'-UTR,EMP3-3'-UTR,CRIP2-3'-UTR,BRAT1-3'-UTR,CD9-3'-UTR,CDK9-3'-UTR,PSMB3-3'-UTR,PSMB6-3'-UTR,PRSS56-3'-UTR,SCAND1-3'-UTR,AMN-3'-UTR,CYBA-3'-UTR,PCOLCE-3'-UTR,MAP1S-3'-UTR,VTN-3'-UTR,ALDH16A1-3'-UTR(全部优选是人的)和Acox2-3'-UTR,Ubc-3'-UTR,Slpi-3'-UTR,Pcyox11-3'-UTR,Igf2bp1-3'-UTR,Tmem37-3'-UTR,Slc7a3-3'-UTR,Cst6-3'-UTR,Ebp-3'-UTR,Sf3b5-3'-UTR,Plk1-3'-UTR,Cdca8-3'-UTR,Kif22-3'-UTR,Cad-3'-UTR,Cth-3'-UTR,Pole-3'-UTR,Kif2c-3'-UTR,Scand1-3'-UTR,Gpr84-3'-UTR,Tpgs1-3'-UTR,Cc117-3'-UTR,Alkbh7-3'-UTR,Ms4a8a-3'-UTR,Mrpl34-3'-UTR,Comtd1-3'-UTR,Armc6-3'-UTR,Atp5d-3'-UTR,1110001J03Rik-3'-UTR,Nudt22-3'-UTR,Aarsd1-3'-UTR,(全部优选是小鼠的)。

[1055] 在本发明的情况下,术语“将人工核酸分子或载体与3'-UTR元件和/或5'-UTR元件关联”优选意为将人工核酸分子或载体与3'-UTR元件和/或与5'-UTR元件功能性地关联或将人工核酸分子或载体与3'-UTR元件和/或与5'-UTR元件功能性地组合。这意为,人工核酸分子或载体以及3'-UTR元件和/或5'-UTR元件,优选如上文所述的3'-UTR元件和/或5'-UTR元件,关联或偶联,从而发挥3'-UTR元件和/或5'-UTR元件的功能,例如,RNA和/或蛋白产生延长和/或增加功能。通常,这意为,3'-UTR元件和/或5'-UTR元件分别在可读框的3'和/或5',优选紧挨可读框的3'和/或紧挨可读框的5',整合到人工核酸分子或载体中,优选mRNA分子中,3'-UTR元件优选在可读框和聚腺苷酸序列或聚腺苷酸化信号之间。优选地,3'-UTR元件和/或5'-UTR元件作为3'-UTR和/或作为5'-UTR分别整合到人工核酸分子或载体中,优选mRNA中,即,从而所述3'-UTR元件和/或5'-UTR元件分别是人工核酸分子或载体、优选mRNA的3'-UTR和/或5'-UTR,即,使得5'-UTR在ORF的5'-末端前立即终止,并且3'-UTR从可读框的3'-侧延伸到聚腺苷酸序列或聚腺苷酸化信号的5'-侧,任选地通过短接头(如包含一个或多个限制性位点或由一个或多个限制性位点组成的序列)连接。由此,优选地,术语



“将人工核酸分子或载体与3'-UTR元件和/或5'-UTR元件关联”意为将3'-UTR元件和/或5'-UTR元件与位于人工核酸分子或载体内、优选mRNA分子内的可读框功能性地关联。3'-UTR和/或5'-UTR和ORF分别如上文关于本发明的人工核酸分子所述,例如,优选地,所述ORF和所述3'-UTR是异源的和/或所述ORF和所述5'-UTR是异源的,例如源自不同的基因(如上文所述的)。3'-UTR元件和/或5'-UTR元件的关联可以通过各个元件的从头连接实现,或通过修饰预先存在的核酸(模板)而实现。在后一情形中,例如,并且出于举例说明的目的,可以通过下述得到本发明的核酸分子:(i)使用与图1A所示的序列相应的核酸作为模板,(ii)通过将(a)下划线的5'-UTR元件(在图1A中下划线显示)或(b)3'-UTR元件(在图1A中双下划线显示)中的至少一个替换为(a)不同的5'-UTR元件或(b)不同的3'-UTR元件而修饰所述模板,由此将人工核酸分子与所述5'-UTR元件或3'-UTR元件关联。

[1056] 在另一方面中,本发明提供3'-UTR元件和/或5'-UTR元件,优选如上文所述的3'-UTR元件和/或如上文所述的5'-UTR元件用于增加和/或延长从人工核酸分子、优选从mRNA分子或载体的蛋白产生的用途,其中所述3'-UTR元件和/或5'-UTR元件包含这样的核酸序列或由这样的核酸序列组成,所述核酸序列源自选自以下各项组成的组的基因的转录本的3'-UTR和/或5'-UTR:ZNF460,TGM2,IL7R,BGN,TK1,RAB3B,CBX6,FZD2,COL8A1,NDUFS7,PHGDH,PLK2,TSP0,PTGS1,FBX032,NID2,ATP5D,EXOSC4,NOL9,UBB4B,VPS18,ORMDL2,FSCN1,TMEM33,TUBA4A,EMP3,TMEM201,CRIP2,BRAT1,SERPINH1,CD9,DPYSL2,CDK9,TFRC,PSMB3 5'-UTR,FASN,PSMB6,PRSS56,KPNA6,SFT2D2,PARD6B,LPP,SPARC,SCAND1,VASN,SLC26A1,LCLAT1,FBXL18,SLC35F6,RAB3D,MAP1B,VMA21,CYBA,SEZ6L2,PCOLCE,VTN,ALDH16A1,RAVER1,KPNA6,SERINC5,JUP,CPN2,CRIP2,EPT1,PNP0,SSSCA1,POLR2L,LIN7C,UQCR10,PYCR1,AMN,MAP1S,NDUFS7,PHGDH,TSP0,ATP5D,EXOSC4,TUBB4B,TUBA4A,EMP3,CRIP2,BRAT1,CD9,CDK9,PSMB3,PSMB6,PRSS56,SCAND1,AMN,CYBA,PCOLCE,MAP1S,VTN,ALDH16A1(全部优选是人的)和Dpys12,Ccnd1,Acox2,Cbx6,Ubc,Ldlr,Nudt22,Pcyox11,Ankrd1,Tmem37,Tspyl4,Slc7a3,Cst6,Aacs,Nosip,Itga7,Ccnd2,Ebp,Sf3b5,Fasn,Hmgcs1,Osr1,Lmnbl,Vma21,Kif20a,Cdca8,Slc7a1,Ubqln2,Prps2,Shmt2,Aurkb,Fignl1,Cad,Anln,Slfn9,Ncaph,Pole,Uhrf1,Gja1,Fam64a,Kif2c,Tspan10,Scand1,Gpr84,Fads3,Cers6,Cxcr4,Gprc5c,Fen1,Cspg4,Mrpl34,Comtd1,Armc6,Emr4,Atp5d,1110001J03Rik,Csf2ra,Aarsd1,Kif22,Cth,Tpgs1,Ccl17,Alkbh7,Ms4a8a,Acox2,Ubc,Slpi,Pcyox11,Igf2bp1,Tmem37,Slc7a3,Cst6,Ebp,Sf3b5,Plk1,Cdca8,Kif22,Cad,Cth,Pole,Kif2c,Scand1,Gpr84,Tpgs1,Ccl17,Alkbh7,Ms4a8a,Mrpl34,Comtd1,Armc6,Atp5d,1110001J03Rik,Nudt22,Aarsd1(全部优选是小鼠的);优选选自下述组成的组:ZNF460-5'-UTR,TGM2-5'-UTR,IL7R-5'-UTR,BGN-5'-UTR,TK1-5'-UTR,RAB3B-5'-UTR,CBX6-5'-UTR,FZD2-5'-UTR,COL8A1-5'-UTR,NDUFS7-5'-UTR,PHGDH-5'-UTR,PLK2-5'-UTR,TSP0-5'-UTR,PTGS1-5'-UTR,FBX032-5'-UTR,NID2-5'-UTR,ATP5D-5'-UTR,EXOSC4-5'-UTR,NOL9-5'-UTR,UBB4B-5'-UTR,VPS18-5'-UTR,ORMDL2-5'-UTR,FSCN1-5'-UTR,TMEM33-5'-UTR,TUBA4A-5'-UTR,EMP3-5'-UTR,TMEM201-5'-UTR,CRIP2-5'-UTR,BRAT1-5'-UTR,SERPINH1-5'-UTR,CD9-5'-UTR,DPYSL2-5'-UTR,CDK9-5'-UTR,TFRC-5'-UTR,PSMB3 5'-UTR,FASN-5'-UTR,PSMB6-5'-UTR,PRSS56-5'-UTR,KPNA6-5'-UTR,SFT2D2-5'-UTR,PARD6B-5'-UTR,LPP-5'-UTR,SPARC-

5'-UTR, SCAND1-5'-UTR, VASN-5'-UTR, SLC26A1-5'-UTR, LCLAT1-5'-UTR, FBXL18-5'-UTR, SLC35F6-5'-UTR, RAB3D-5'-UTR, MAP1B-5'-UTR, VMA21-5'-UTR, CYBA-5'-UTR, SEZ6L2-5'-UTR, PCOLCE-5'-UTR, VTN-5'-UTR, ALDH16A1-5'-UTR, RAVR1-5'-UTR, KPNA6-5'-UTR, SERINC5-5'-UTR, JUP-5'-UTR, CPN2-5'-UTR, CRIP2-5'-UTR, EPT1-5'-UTR, PNPO-5'-UTR, SSSCA1-5'-UTR, POLR2L-5'-UTR, LIN7C-5'-UTR, UQCR10-5'-UTR, PYCRL-5'-UTR, AMN-5'-UTR, MAP1S-5'-UTR, (全部优选是人的) 和 Dpys12-5'-UTR, Ccnd1-5'-UTR, Acox2-5'-UTR, Cbx6-5'-UTR, Ubc-5'-UTR, Ldlr-5'-UTR, Nudt22-5'-UTR, Pcyox11-5'-UTR, Ankrd1-5'-UTR, Tmem37-5'-UTR, Tspyl4-5'-UTR, Slc7a3-5'-UTR, Cst6-5'-UTR, Aacs-5'-UTR, Nosip-5'-UTR, Itga7-5'-UTR, Ccnd2-5'-UTR, Ebp-5'-UTR, Sf3b5-5'-UTR, Fasn-5'-UTR, Hmgcs1-5'-UTR, Osr1-5'-UTR, Lmnb1-5'-UTR, Vma21-5'-UTR, Kif20a-5'-UTR, Cdca8-5'-UTR, Slc7a1-5'-UTR, Ubqln2-5'-UTR, Prps2-5'-UTR, Shmt2-5'-UTR, Aurkb-5'-UTR, Fignl1-5'-UTR, Cad-5'-UTR, Anln-5'-UTR, Slfn9-5'-UTR, Ncaph-5'-UTR, Pole-5'-UTR, Uhrf1-5'-UTR, Gjal-5'-UTR, Fam64a-5'-UTR, Kif2c-5'-UTR, Tspan10-5'-UTR, Scand1-5'-UTR, Gpr84-5'-UTR, Fads3-5'-UTR, Cers6-5'-UTR, Cxcr4-5'-UTR, Gprc5c-5'-UTR, Fen1-5'-UTR, Cspg4-5'-UTR, Mrpl34-5'-UTR, Comtd1-5'-UTR, Armc6-5'-UTR, Emr4-5'-UTR, Atp5d-5'-UTR, 1110001J03Rik-5'-UTR, Csf2ra-5'-UTR, Aarsd1-5'-UTR, Kif22-5'-UTR, Cth-5'-UTR, Tpgs1-5'-UTR, Ccl17-5'-UTR, Alkbh7-5'-UTR, Ms4a8a-5'-UTR (全部优选是小鼠的); 和 NDUFS7-3'-UTR, PHGDH-3'-UTR, TSP0-3'-UTR, ATP5D-3'-UTR, EXOSC4-3'-UTR, TUBB4B-3'-UTR, TUBA4A-3'-UTR, EMP3-3'-UTR, CRIP2-3'-UTR, BRAT1-3'-UTR, CD9-3'-UTR, CDK9-3'-UTR, PSMB3-3'-UTR, PSMB6-3'-UTR, PRSS56-3'-UTR, SCAND1-3'-UTR, AMN-3'-UTR, CYBA-3'-UTR, PCOLCE-3'-UTR, MAP1S-3'-UTR, VTN-3'-UTR, ALDH16A1-3'-UTR (全部优选是人的) 和 Acox2-3'-UTR, Ubc-3'-UTR, Slpi-3'-UTR, Pcyox11-3'-UTR, Igf2bp1-3'-UTR, Tmem37-3'-UTR, Slc7a3-3'-UTR, Cst6-3'-UTR, Ebp-3'-UTR, Sf3b5-3'-UTR, Plk1-3'-UTR, Cdca8-3'-UTR, Kif22-3'-UTR, Cad-3'-UTR, Cth-3'-UTR, Pole-3'-UTR, Kif2c-3'-UTR, Scand1-3'-UTR, Gpr84-3'-UTR, Tpgs1-3'-UTR, Ccl17-3'-UTR, Alkbh7-3'-UTR, Ms4a8a-3'-UTR, Mrpl34-3'-UTR, Comtd1-3'-UTR, Armc6-3'-UTR, Atp5d-3'-UTR, 1110001J03Rik-3'-UTR, Nudt22-3'-UTR, Aarsd1-3'-UTR, (全部优选是小鼠的)。

[1057] 本发明所述的用途优选包括将人工核酸分子、载体或RNA与如上文所述的3'-UTR元件和/或与如上文所述的5'-UTR元件关联。

[1058] 还可以彼此分开制造和买卖本发明的药物组合物的化合物和成分。因此, 本发明还涉及试剂盒或多部件的试剂盒, 其包含本发明所述的人工核酸分子、本发明所述的载体、本发明所述的细胞和/或本发明所述的药物组合物。优选, 此种试剂盒或多部件的试剂盒可以另外包含使用说明书、用于转染的细胞、佐剂、用于施用所述药物组合物的工具、药学可接受的载体和/或用于溶解或稀释所述人工核酸分子、所述载体、所述细胞或所述药物组合物的药学可接受的溶液。

[1059] 用于鉴定3'-非翻译区元件(3'-UTR元件)和/或5'-非翻译区元件(5'-UTR元件)的方法

[1060] 在另一方面, 本说明书提供用于鉴定3'-非翻译区元件(3'-UTR元件)和/或5'-非翻译区元件(5'-UTR元件)的方法, 所述方法包括下述步骤:

[1061] a) 分析mRNA的稳定性,包括以下子步骤:

[1062] i. 确定所述mRNA在所述mRNA的降解过程中在第一时间点的量,

[1063] ii. 确定所述mRNA在所述mRNA的降解过程中在第二时间点的量,和

[1064] iii. 计算步骤(i)中确定所述mRNA的量与步骤(ii)中确定的所述mRNA的量的比率;

[1065] b) 选择在子步骤(iii)中计算的比率为至少0.5(50%),至少0.6(60%),至少0.7(70%),至少0.75(75%),至少0.8(80%),至少0.85(85%),至少0.9(90%),或至少0.95(95%)的稳定mRNA;和

[1066] c) 确定所述稳定mRNA的3'-和/或5'-UTR元件的核苷酸序列。

[1067] 由此,mRNA的稳定性优选在标准条件,例如用于使用的某种细胞系或细胞类型的标准条件(标准培养基,孵育等)下评估。

[1068] 为了分析mRNA的稳定性,通过确定所述mRNA的降解过程中所述mRNA在第一和第二时间点的量或浓度评价该mRNA的降解过程(参见步骤a)i.和a)ii.)。

[1069] 为了确定如上文所限定的体内或体外(即体外,尤其是指(“活”)细胞和/或组织,包括活受试者的组织;细胞尤其包括细胞系,原代细胞,组织或受试者中的细胞,优选是哺乳动物细胞,例如人细胞和小鼠细胞,并且特别优选是使用人细胞系HeLa、HEPG2和U-937以及小鼠细胞系NIH3T3, JAWSII和L929;此外原代细胞是特别优选的,在特别优选的实施方案中,是人皮肤成纤维细胞(HDF))RNA降解过程中mRNA的量或浓度,可以使用技术人员已知的各种方法。所述方法的非限制性实例包括例如利用转录抑制剂如放线菌素D的一般性抑制转录,使用诱导型启动子特异性启动瞬时转录,例如c-fos血清-诱导型启动子系统和Tet-off调节性启动子系统,以及动力学标记技术,例如脉冲标记。

[1070] 例如,如果步骤a)中利用转录抑制剂-介导的转录停滞以确定如上文所定义的体内或体外RNA降解过程中mRNA的量或浓度,可以使用转录抑制剂如放线菌素D(ActD),5,6-二氯-1-D-核糖呋喃糖基-苯并咪唑(DRB)或-鹅膏蕈碱( $\alpha$ -Am)。由此,为了评估mRNA降解,通常将转录抑制剂加入到细胞中,并且,由此通常抑制转录,并且在不干扰正在进行的转录的情况下观察到RNA降解。

[1071] 备选地,特异性启动瞬时转录的诱导型启动子可以用于步骤a),由此原理是提供激活转录并导致mRNA合成突发的刺激,随后去除所述刺激以关闭转录并监测mRNA的降解。由此,诱导型启动子使得能够严格控制,从而在狭窄的时间窗口完成转录的诱导和沉默。在哺乳动物细胞中,已知c-fos启动子对于此目的是有价值的,因为其可以响应血清添加而快速和暂时诱导,由此提供可靠且简单的实现转录的瞬时爆发的方式。Tet-off启动子系统提供另一种选择,其进一步加宽转录脉冲方法的应用来研究哺乳动物细胞中的mRNA周转率。

[1072] 然而,在本发明中,动力学标记技术在步骤a)中优选用于确定如上文所定义的体内或体外RNA降解过程中mRNA的量。在动力学标记中,通常标记RNA,因此标记特别包括标记的核苷酸和标记的核苷,并且标记的尿苷和标记的尿嘧啶是特别优选的。优选的标记的实例包括4-硫尿苷(4sU),2-硫尿苷,6-硫鸟苷,5-乙炔基尿苷(EU),5-溴-尿苷(BrU),生物素-16-氨基烯丙基尿苷,5-氨基烯丙基尿苷,5-氨基烯丙基胞苷等,其中4-硫尿苷(4sU),5-乙炔基尿苷(EU)或5'-溴-尿苷(BrU)是更优选的。特别优选的是4-硫尿苷(4sU)。4-硫尿苷(4sU)优选以100-500 $\mu$ M的浓度使用。此外,还可以使用放射标记的核苷酸,例如利用尿苷-

<sup>3</sup>H。还可以使用上述标记的核苷酸的组合,由此4-硫尿苷和6-硫鸟苷的组合是特别优选的。

[1073] 在动力学标记中,通常例如通过在转录过程中掺入标记的尿苷或尿嘧啶标记出现的RNA。过一会儿后,终止提供标记,并且然后可以在通常不抑制转录的情况下通过评估特异标记的RNA观察RNA降解。

[1074] 为了确定RNA降解过程中的mRNA的量,在步骤a)中,脉冲标记是优选的,并且脉冲追踪方法是特别是优选的。如本文中使用的,术语“脉冲标记”是指标记,例如上述标记用于测量活细胞内化合物的合成和/或降解的速率的技术。通常,细胞在短期内暴露于少量的标记,此为术语“脉冲”。在脉冲追踪方法中,脉冲-标记后,在暴露于标记所需时期后,通常加入与“脉冲”对应的大得多的量的未标记的化合物(例如,如果标记的尿苷用作脉冲,为未标记的尿苷)。标记的和未标记的化合物之间的竞争效果使标记的化合物的进一步摄取降低至可忽略的水平,此为术语“追踪”。

[1075] 为了确定mRNA的量或浓度,通常必须分离mRNA。不同的RNA分离技术是技术人员已知的,例如通过胍盐硫氰酸盐-苯酚-氯仿提取或通过基于二氧化硅柱的提取。此外,可以使用可商购试剂盒,例如Qiagen的RNeasy试剂盒。

[1076] 此外,可以需要提取步骤,尤其是如果使用动力学标记(与转录抑制剂相反,其中总RNA代表“降解的”RNA,因为转录通常被抑制)。在提取步骤中,标记的RNA(即代表“降解的”RNA)从总的分离的RNA提取。由此,提取方法可以根据使用的标记选择。例如,可以使用利用针对所述标记的抗体的免疫纯化。

[1077] 此外,例如,对于硫标记的,例如4-硫尿苷(4sU)-标记的RNA的提取,HPDP-生物素(通过可裂解的(可逆的)二硫键缀合的吡啶基二巯基-激活的,巯基-反应性生物素化试剂)可以与分离的“总RNA”孵育。该试剂特异性与4-硫尿苷(4sU)-标记的RNA中还原的巯基(-SH)特异性反应,形成可逆二硫键。生物素化允许硫-标记的例如4-硫尿苷(4sU)-标记的RNA与链霉亲和素结合,并且因此可以通过用二巯苏糖醇或β-巯基乙醇(或任意其他还原剂)还原二硫键而从总RNA中提取。

[1078] 在生物素-标记的核苷酸,例如生物素-16-氨基烯丙基尿苷的情况下,链霉亲和素可以直接用于从总RNA提取标记的RNA。

[1079] 例如,对于新转录的5-乙炔基尿苷(EU)-标记的细胞RNA从总RNA中的提取,可以在铜催化的环加成反应(常称为点击化学)中使用EU的生物素化,接着通过链霉亲和素亲和纯化。该方法可作为Click-iT新生RNA捕获试剂盒(Click-iT Nascent RNA Capture Kit,目录号C10365,Invitrogen)商购。制造商的该试剂盒的使用说明建议脉冲标记时间对于0.5mM EU剂量是30至60min,或对于0.1或0.2mM EU剂量是1至24h。

[1080] 例如,BrU-标记的RNA分子可以通过用抗-溴脱氧尿苷抗体(例如Clone.2B1,目录号MI-11-3,MBL)免疫纯化和蛋白G Sepharose提取。

[1081] mRNA的量或浓度,即转录本水平,可以随后通过本领域技术人员已知的各种方法测量。所述方法的非限制性实例包括微阵列分析,Northern印记分析,定量PCR或通过下一代测序(高通量测序)。特别优选的是微阵列分析和下一代测序。此外,全基因组方法/全转录组方法是特别优选的,例如在微阵列分析中,全基因组微阵列分析,例如Affymetrix人基因1.0ST或2.0ST或Affymetrix鼠基因1.0ST或2.0ST或通过下一代测序的全转录组分析。

[1082] 在步骤a)的子步骤i.和ii.中,在mRNA降解过程中的第一和第二时间点确定mRNA

的量。通常,这意为,尤其是在mRNA的降解过程中第一和第二时间点分离mRNA,以确定各个量。因此,“所述第一时间点”和“所述第二时间点”尤其是RNA降解过程中的时间点,在该时间点分离RNA以确定RNA量。通常,在RNA降解过程中“所述第二时间点”晚于“所述第一时间点”。

[1083] 优选地,这样选择所述第一时间点,使得仅考虑经历降解过程的mRNA,即避免新出现的mRNA——例如在正在进行的转录中。例如,如果使用动力学标记技术,例如脉冲标记,优选这样选择所述第一时间点,使得完成标记在mRNA中的结合,即没有持续的向mRNA中掺入标记发生。由此,如果使用动力学标记,所述第一时间点可以是在实验标记步骤结束后,例如细胞与标记孵育结束后至少10min,至少20min,至少30min,至少40min,至少50min,至少60min,至少70min,至少80min,或至少90min。

[1084] 例如,所述第一时间点可以优选是终止转录(例如,通过转录抑制剂)后,在诱导型启动子的情况下终止启动子诱导后或终止脉冲或标记供应后,例如标记结束后0至6h。更优选地,所述第一时间点可以是终止转录(例如,通过转录抑制剂)后,在诱导型启动子的情况下终止启动子诱导或终止脉冲或标记供应后,例如标记结束后30min至5h,甚至更优选1h至4h,并且特别优选约3h。

[1085] 优选地,在mRNA降解过程中尽可能晚地选择所述第二时间点。然而,如果考虑多种mRNA种类,优选这样选择所述第二时间点,使得仍然有大量的多种mRNA种类,优选至少10%的mRNA种类,以可检测量,即以高于0的量存在。优选地,所述第二时间点是终止转录(例如通过转录抑制剂)后,在诱导型启动子的情况下终止启动子诱导后或终止脉冲或标记供应后,例如标记结束后至少5h,至少6h,至少7h,至少8h,至少9h,至少10h,至少11h,至少12h,至少13h,至少14h,或至少15h。

[1086] 例如,所述第二时间点可以优选是终止转录(例如通过转录抑制剂)后,在诱导型启动子的情况下终止启动子诱导后或终止脉冲或标记供应后,例如标记结束后3至48h。更优选地,所述第二时间点可以是终止转录(例如通过转录抑制剂)后,在诱导型启动子的情况下终止启动子诱导后或终止脉冲或标记供应后,例如标记结束后6min至36h,甚至更优选10h至24h,并且特别优选约15h。

[1087] 由此,所述第一时间点和所述第二时间点之间的时间间隔优选在上述极限内尽可能大。因此,所述第一时间点和所述第二时间点之间的时间间隔优选是至少4h,至少5h,至少6h,至少7h,至少8h,至少9h,至少10h,至少11h,或至少12h,其中约12h的时间间隔是特别是优选的。通常,第二较晚时间点比所述第一时间点晚至少10分钟。

[1088] 在步骤a)的子步骤iii.中,计算步骤(i)中确定的mRNA的量与步骤(ii)中确定的mRNA的量的比率。为此,将如上文所述在所述第二时间点确定的mRNA的量(转录本水平)除以如上文所述在所述第一时间点确定的mRNA(转录本水平)的量。该比率防止无视在所述第一时间点已经仅以非常低的量存在的稳定mRNA(相对于以高的量存在的mRNA)。

[1089] 在步骤b)中,选择这样的mRNA,其在步骤a)的子步骤(iii)计算的比率为至少0.5(50%),至少0.6(60%),至少0.7(70%),至少0.75(75%),至少0.8(80%),至少0.85(85%),至少0.9(90%),或至少0.95(95%)。在本发明中所述mRNA被认为是特别稳定的mRNA。

[1090] 在步骤c)中,确定所述mRNA,即步骤b)中选择的mRNA的3'-和/或5'-UTR元件的核

核苷酸序列。为此,可以利用技术人员已知的不同方法,例如测序或从公众可获得的数据库如例如NCBI(National Center for Biotechnology Information)选择。例如,可以在数据库中检索步骤b)中选择的mRNA的mRNA序列,并且可以随后从数据库中存在的mRNA序列提取3'-和/或5'-UTR。

[1091] 尤其是,在上述用于鉴别3'-非翻译区元件(3'-UTR元件)和/或5'-非翻译区元件(5'-UTR元件)的方法中,术语“mRNA”和/或“稳定mRNA”还分别可以指本文中定义的mRNA种类和/或稳定的mRNA种类。

[1092] 此外,本发明中优选的是,与平均mRNA降解相比,“稳定mRNA”可以具有较慢的mRNA降解,优选如上文所定义的体内或体外评估的。由此,“平均mRNA降解”可以通过研究多种mRNA种类的mRNA降解评估。

[1093] 因此,在另一方面,本发明提供鉴别3'-非翻译区元件(3'-UTR元件)和/或5'-非翻译区元件(5'-UTR元件)的方法,其包括以下步骤:

[1094] a) 分析多种mRNA种类的稳定性,包括以下子步骤:

[1095] i. 确定所述多个mRNA种类的每个mRNA种类在所述mRNA种类的降解过程中的第一时间点的量,

[1096] ii. 确定所述多个mRNA种类的每个mRNA种类在所述mRNA种类的降解过程中的第二时间点的量,和

[1097] iii. 对于所述多种mRNA种类的每个mRNA种类计算步骤(i)中确定的所述mRNA种类的量与步骤(ii)中确定的所述mRNA种类的量的比率;

[1098] b) 根据子步骤(iii)中对于每个mRNA种类计算的比率将多种mRNA种类中的mRNA种类排序;

[1099] c) 选择具有在子步骤(iii)中计算的最高比率或多个最高比率的一种或多种mRNA种类;和

[1100] d) 确定所述mRNA的3'-和/或5'-UTR元件的核苷酸序列。

[1101] “mRNA种类”,如本文中使用的,对应于基因组转录单位,即通常对应于基因。由此,在一个“mRNA种类”中可以出现不同转录本,例如,由于mRNA加工导致。例如,mRNA种类可以由微阵列上的点代表。因此,微阵列提供确定多种mRNA种类,例如在mRNA降解过程中某时间点的量的有利的工具。然而,还可以使用技术人员已知的其他技术,例如RNA-seq(也称为全转录组鸟枪(Shotgun)测序,其是使用下一代测序在给定时间点从基因组揭示RNA存在和数量的快照的能力的技术),定量PCR等。

[1102] 优选地,“多种mRNA种类”是指至少100,至少300,至少500,至少1000,至少2000,至少3000,至少4000,至少5000,至少6000,至少7000,至少8000,至少9000,至少10000,至少11000,至少12000,至少13000,至少14000,至少15000,至少16000,至少17000,至少18000,至少19000,至少20000,至少21000,至少22000,至少23000,至少24000,至少25000,至少26000,至少27000,至少28000,至少29000,或至少30000种mRNA种类。特别优选的是,评估全转录组,或转录组的尽可能多的mRNA种类。例如,这可以通过使用提供全转录本覆盖的微阵列实现。

[1103] 该方法的步骤a)以及其子步骤i.至iii.基本上对应于之前所述的本发明的方法的步骤a)以及其子步骤i.至iii.,但区别仅在于在第一和第二时间点确定多种mRNA种类中

每个mRNA种类的量以及对于每个mRNA种类计算比率。因此,上文概述的详细方法和优选的实施方案也适用于这里,并且单个mRNA种类的比率(以及每个单个mRNA种类分别的比率)可以如上文对于“一个mRNA”概述而确定。

[1104] 然而,与上述方法相反,mRNA的稳定性不通过比率的绝对值评估,而是根据在步骤a)的子步骤(iii)中对于每个mRNA种类计算的比率通过多种mRNA种类的mRNA种类的排名评估。在子步骤c)中,然后选择具有在步骤a)的子步骤(iii)中计算的最高比例或多个最高比例的一种或多种mRNA种类。

[1105] 在该情况下,特别优选的是在步骤c)中选择0.1%,0.2%,0.3%,0.4%,0.5%,0.6%,0.7%,0.8%,0.9%,1%,2%,3%,4%,5%,10%,15%,20%最稳定的mRNA种类。备选地或此外,在步骤c)中,所述mRNA种类可以这样选择,其显示对应于从分析的所有mRNA种类中计算的平均比率的至少100%的在步骤a)的子步骤iii.中计算的比率。更优选选择这样的所述mRNA种类,其显示从分析的所有mRNA种类计算的平均比率的至少150%,甚至更优选至少200%并且最优选至少300%的比率。

[1106] 在步骤d)中,对于之前所述的本发明的方法的步骤c),如上文所述的确定步骤c)中选择的mRNA的3'-和/或5'-UTR元件的核苷酸序列。

[1107] 优选地,在两种上述根据本发明的鉴别3'-UTR元件和/或5'-UTR元件的方法中,所述第一时间点和所述第二时间点之间的时期是至少5h,优选至少6h,优选至少7h,更优选至少8h,更优选至少9h,甚至更优选至少10h,甚至更优选至少11h,并且特别优选至少12h。

[1108] 优选地,在两种上文所述的根据本发明的鉴别3'-UTR元件和/或5'-UTR元件的方法中,通过脉冲标记,优选使用脉冲追踪方法分析mRNA的稳定性。

[1109] 用于鉴定3'-非翻译区元件(3'-UTR元件)和/或5'-非翻译区元件(5'-UTR元件)的方法

[1110] 在另一方面,本发明还提供鉴别为人工核酸分子提供高翻译效率的3'-非翻译区元件(3'-UTR元件)和/或5'-非翻译区元件(5'-UTR元件)的方法,其包括以下步骤:

[1111] a) 根据上述方法中的任一种通过鉴别3'-UTR元件和/或5'-UTR元件的方法鉴别源自稳定mRNA的3'-UTR元件和/或5'-UTR元件;

[1112] b) 合成人工核酸分子,其包含至少一个可读框和对应于步骤a)中鉴别的3'-UTR元件和/或5'-UTR元件或步骤a)中鉴别的3'-UTR元件和/或5'-UTR元件包含的至少一个3'-UTR元件和/或至少一个5'-UTR元件;

[1113] c) 分析步骤b)中合成的人工核酸分子的至少一个可读框(ORF)编码的蛋白的表达;

[1114] d) 分析缺少3'-UTR元件和/或5'-UTR元件的参比人工核酸分子的至少一个可读框编码的蛋白的表达;

[1115] e) 将步骤c)中分析的人工核酸分子的蛋白表达与步骤d)中分析的参比人工核酸分子的蛋白表达进行比较;和

[1116] f) 如果与从步骤d)中分析的参比人工核酸分子的蛋白表达相比,从步骤c)中分析的人工核酸分子的蛋白表达延长和/或增加,则选择所述3'-UTR元件和/或5'-UTR元件。

[1117] 在该方法中,首先,3'-UTR元件和/或5'-UTR元件可通过如上文所述的根据本发明的方法鉴别。这使得能够通过技术人员已知的方法,例如通过PCR扩增合成3'-和/或5'-UTR

元件。用于所述PCR的引物可以优选包含用于克隆的限制性位点。备选地,3'-和/或5'-UTR元件可以例如通过化学合成或寡聚物退火合成。因此,在步骤b)中,合成人工核酸分子,其包含至少一个可读框和对应于步骤a)中鉴别的3'-UTR元件和/或5'-UTR元件或步骤a)中鉴别的3'-UTR元件和/或5'-UTR元件包含的至少一个3'-UTR元件和/或至少一个5'-UTR元件。尤其是,如果3'-和/或5'-UTR元件满足各自的需求,即如果它们延长和/或增加蛋白表达,则通常将所述至少一个3'-UTR元件和/或至少一个5'-UTR元件与可读框组合,这产生包含本发明所述的3'-和/或5'-UTR元件的人工核酸。为了测试它,步骤a)中鉴别的3'-和/或5'-UTR元件,或其PCR片段或合成的序列分别可以克隆入特定载体,优选表达载体,从而评估从各个ORF的蛋白表达。

[1118] 随后如本文中所述在步骤c)中评估从包含所述至少一个3'-UTR元件和/或所述至少一个5'-UTR元件的人工核酸分子的蛋白表达,并且与在步骤d)中评估的从如本文所述在步骤e)中的各个缺少3'-UTR元件和/或5'-UTR元件的参比人工核酸分子的蛋白表达相比较。

[1119] 之后,在步骤f)中,选择这样的3'-UTR元件和/或5'-UTR元件,与在步骤d)中分析的从参比人工核酸分子的蛋白表达相比,其延长和/或增加在步骤c)中分析的从人工核酸分子的蛋白表达。本发明的核酸分子与参比核酸分子的蛋白表达的比较如本文中所述进行,尤其是在本发明的人工核酸分子的情况下。

[1120] 此外,本发明提供鉴别为人工核酸分子提供高翻译效率的3'-非翻译区元件(3'-UTR元件)和/或5'-非翻译区元件(5'-UTR元件)的特别优选的方法,其包括以下步骤:

[1121] a) 为细胞供给标记的核苷酸/将细胞与标记的核苷酸孵育,所述标记的核苷酸用于掺入新转录的RNA分子中(脉冲追踪标记);

[1122] b) 在第一时间点和至少一个第二较晚时间点分离细胞的总RNA;

[1123] c) 从步骤b)中分离的总RNA中提取标记的RNA分子;

[1124] d) 测量标记的RNA中包含的不同mRNA种类的量/转录本水平;

[1125] e) 计算在至少一个第二较晚时间点存在的mRNA种类的量/转录本水平与在所述第一时间点存在的mRNA种类的量/转录本水平的比率;

[1126] f) 根据步骤e)中确定的比率将mRNA种类排名;

[1127] g) 选择最稳定的mRNA种类;

[1128] h) 确定步骤g)中选择的最稳定mRNA种类的3'-和/或5'-UTR的核苷酸序列;

[1129] i) 合成步骤h)中确定的3'-和/或5'-UTR中包含的3'-和/或5'-UTR元件;

[1130] j) 将步骤i)中合成的3'-和/或5'-UTR元件与可读框组合,得到如本文所述的本发明所述的核酸;和

[1131] k) 任选地比较本发明的核酸中存在的可读框的表达与没有如本文所述的3'-和/或5'-UTR元件的参比核酸中存在的可读框的表达。

[1132] 由此,在步骤a)至k)所概述的各个限制内,上文对于本发明的方法所述的详情和优选的实施方案也适用于这里。

[1133] 尤其是,以下标记的核苷酸优选的用于在本发明的方法的步骤a)中供给细胞:4-硫尿苷(4sU),2-硫尿苷,6-硫鸟苷,5-乙炔基尿苷(EU),5-溴-尿苷(BrU),生物素-16-氨基烯丙基尿苷,5-氨基烯丙基尿苷,5-氨基烯丙基胞苷等。特别优选的是4-硫尿苷(4sU)。4-硫



尿苷优选以100-500 $\mu$ M的浓度使用。备选地,可以使用放射标记的核苷酸,例如尿苷-<sup>3</sup>H。可以使用上述标记的核苷酸的组合。特别优选的是4-硫尿苷和6-硫鸟苷的组合。

[1134] 步骤a)中细胞与标记的核苷酸的孵育可以变化。特别优选的是10分钟至24小时的孵育(供给时间)。特别优选的是2至6小时,更优选2至3小时。

[1135] 可以用于本发明的方法的细胞尤其包括细胞系,原代细胞,组织或受试者中的细胞。在具体实施方案中,允许细胞培养的细胞类型可以适于本发明的方法。特别优选的是哺乳动物细胞,例如人细胞和小鼠细胞。在特别是优选的实施方案中,使用人细胞系HeLa、HEPG2和U-937以及小鼠细胞系NIH3T3、JAWSII和L929。此外,原代细胞是特别优选的;在特别优选的实施方案中,可以特别使用人皮肤成纤维细胞(HDF)。备选地,标记的核苷酸也可以适于受试者的组织,并且孵育时间后,根据步骤c)分离组织的RNA。

[1136] 为了确定细胞(类型)的最稳定mRNA,在如上文所述的第一时间点,例如标记后0至6h,优选标记后3h,并且在如上文所述的第二较晚时间点,例如标记后3至48h,优选标记后10至24h,最优选15h,提取总RNA。第二较晚时间点比第一时间点晚至少10分钟。

[1137] 在步骤f)中,根据步骤e)中计算的比率将mRNA种类排名。在该情况下,特别优选的是选择0.1%,0.2%,0.3%,0.4%,0.5%,0.6%,0.7%,0.8%,0.9%,1%,2%,3%,4%,5%,10%,15%,20%最稳定的mRNA种类。

[1138] 在该情况下,还优选的是选择与第一时间相比,显示mRNA种类在第二较晚时间的至少50%(0,5倍),至少60%(0,6倍),至少70%(0,7倍),至少90%(0,9倍)或至少95%(0,95倍)转录本水平/量的这些mRNA种类。如果在标记后3小时(第一时间点)和15小时(第二时间点)分离RNA,该实施方案是特别优选的。

[1139] 备选地或另外,选择这些mRNA种类,其显示对应于从分析的所有mRNA种类计算的平均比率的至少100%的步骤e)中计算的比率。更优选选择这些mRNA种类,其显示从分析的所有mRNA种类计算的平均比率的至少150%,并且更优选至少200%,并且最优选至少300%的比率。

[1140] 在本发明的方法的进一步的步骤中,确定步骤g)中选择的最稳定mRNA种类的3'-和/或5'-UTR的核苷酸序列,并且在步骤i)中例如通过PCR扩增合成3'-和/或5'-UTR元件。用于PCR的引物可以优选包含用于克隆的限制性位点。备选地,可以合成(例如通过化学合成或寡聚物退火)3'-和/或5'-UTR元件。

[1141] 在本发明的方法的步骤j)中,将得到的PCR片段或合成的序列与可读框组合,得到包含本发明所述的3'-和/或5'-UTR元件的人工核酸。优选地,PCR片段或序列可以克隆到载体中。

[1142] 在特别优选的实施方案中,本发明提供用于鉴别3'-非翻译区元件(3'-UTR元件)和/或5'-非翻译区元件(5'-UTR元件)的包含步骤a)至k)的方法,其中3'-UTR元件和/或5'-UTR元件延长从包含至少一个所述3'-UTR元件和/或至少一个所述5'-UTR元件的人工核酸分子的蛋白产生。

[1143] 在进一步的方面,本发明还提供产生人工核酸分子的方法,其中合成人工核酸分子,所述人工核酸分子包含至少一个可读框和通过如上文所述的根据本发明的鉴别3'-UTR元件和/或5'-UTR元件的方法鉴别的至少一个3'-UTR元件和/或至少一个5'-UTR元件。所述人工核酸分子的合成通常通过技术人员已知的方法进行,例如克隆方法,例如通常已知的

或本文中所述的克隆方法。

[1144] 优选地,如本文所述的本发明所述的载体用于所述的用于产生人工核酸分子的本发明的方法中。

[1145] 优选地,通过所述用于产生人工核酸分子的方法产生的人工核酸分子是如本文所述的本发明所述的核酸分子。

[1146] 此外,本发明还提供可通过用于产生如本文所述的本发明所述的人工核酸分子的方法获得的人工核酸分子。

[1147] 以下附图、序列和实施例意在进一步说明本发明。它们不意在将本发明的主题限制于其中。

[1148] 图1显示可以通过体外转录获得的编码mRNA(即,人工核酸分子)的序列。

[1149] 使用以下缩写:

[1150] • PpLuc (GC):编码萤火虫 (*Photinus pyralis*) 荧光素酶的富GC的mRNA序列。PpLuc (GC) ORF以斜体突出显示。

[1151] • A64:具有64个腺苷酸的聚腺苷酸-序列

[1152] • C30:具有30个胞苷酸的聚胞苷酸-序列

[1153] • hSL:从 (Cakmakci, Lerner, Wagner, Zheng, & William F Marzluff, 2008. *Mol. Cell. Biol.* 28 (3):1182-94) 获得的组蛋白茎环序列

[1154] • 32L4:缺少5' 末端寡嘧啶束的人大核糖体蛋白32的5' 非翻译区5' -UTR的人工变体。所述5' -UTR来源于缺少5' 末端寡嘧啶束的人大核糖体蛋白32mRNA。

[1155] • 白蛋白7:人白蛋白的3' 非翻译区(3' -UTR)的人工变体,其中引入三个单点突变以去除T7终止信号和HindIII和XbaI限制性位点

[1156] 图1A:参比构建体,SEQ ID NO:205,即,32L4-PpLuc (GC)-白蛋白7-A64-C30-hSL (R3111)的mRNA序列。

[1157] 对于进行实验检测的人来源的UTR元件,参见表3。对于进行实验检测的小鼠来源的UTR元件,参见表4。

[1158] 图1B:人工核酸(SEQ ID NO:210),所述人工核酸包含与SEQ ID NO:1(单下划线)对应的测试的5' -UTR。除了所述下划线的元件之外,该图中显示的该序列的所有元件与SEQ ID NO:205相同。因此,SEQ ID NO:210与SEQ ID NO:205的区别在于5' -UTR元件是不同的。

[1159] [5' -UTR,对应于SEQ ID NO:1]-PpLuc (GC)-白蛋白7-A64-C30-hSL

[1160] 图1C:人工核酸(SEQ ID NO:211),所述人工核酸包含与SEQ ID NO:152(双下划线)对应的测试的3' -UTR。除了双下划线的元件之外,该图中显示的该序列的所有元件与SEQ ID NO:205相同。因此,SEQ ID NO:211与SEQ ID NO:205的区别在于3' -UTR元件是不同的。

[1161] 32L4-PpLuc (GC)-[3' -UTR,对应于SEQ ID NO:152]-A64-C30-hSL

[1162] 图2至9显示一式三份分析的mRNA的平均荧光素酶表达和SEM;即,针对RrLuc标准化的相对PpLuc表达(给出三次独立实验的平均值和平均值的标准误差(SEM))。与从图1所示的mRNA(参比构建体)的荧光素酶表达相比较,检验了人来源的5' -UTR元件和3' -UTR元件(表3)和小鼠来源的5' -UTR元件和3' -UTR元件(表4)对从mRNA的荧光素酶表达的影响。关于包含表3和4的测试的5' -UTR元件和3' -UTR元件的构建体的详细内容参见实施例3,特别是

3.3和3.4节。为了这一目的,将该图中所示的细胞系通过脂转染用不同的mRNA转染(图2:用具有表3所示的人来源的5'-UTR元件和3'-UTR元件的mRNA转染的HDF细胞;图3:用具有表4所示的小鼠来源的5'-UTR元件和3'-UTR元件的mRNA转染的HDF细胞;图4:用具有表3所示的人来源的5'-UTR元件和3'-UTR元件的mRNA转染的L929细胞;图5:用具有表4所示的小鼠来源的5'-UTR元件和3'-UTR元件的mRNA转染的L929细胞;图6:用具有表3所示的人来源的5'-UTR元件和3'-UTR元件的mRNA转染的HEPG2细胞;图7:用具有表4所示的小鼠来源的5'-UTR元件和3'-UTR元件的mRNA转染的HEPG2细胞;图8:用具有表3所示的人来源的5'-UTR元件和3'-UTR元件的mRNA转染的HeLa细胞;图9:用具有表4所示的小鼠来源的5'-UTR元件和3'-UTR元件的mRNA转染的HeLa细胞)。关于详细内容,参见实施例3,特别是3.5节。图2至9分别显示了96个柱状图,其对应于实施例3的测定的96个孔。在图2至9中,96个柱状图以连续的顺序显示,即,A1,B1,C1,D1,[...],F12,G12,H12。

[1163] RLU=相对光单位(荧光素酶活性测量)

## 实施例

### [1164] 实施例1:候选UTR元件

[1165] 实验评价在人和小鼠细胞中的mRNA表达。这导致鉴别了在各种细胞中表达的mRNA。通过数据库搜索确定了mRNA种类的5'-和/或3'-UTR的核苷酸序列,按照细胞转录组的高通量测序数据验证并调整,并且通过PCR扩增或通过寡聚体退火合成。

[1166] 由于所鉴别的mRNA来源于它们的天然细胞环境,所以这些mRNA理解为是野生型mRNA:

[1167] SEQ ID NO:1至72表示野生型人5'-UTR元件。

[1168] SEQ ID NO:73至136表示野生型小鼠5'-UTR元件。

[1169] SEQ ID NO:152至173表示野生型人3'-UTR元件。

[1170] SEQ ID NO:174至203表示野生型小鼠3'-UTR元件。

[1171] 尽管这些UTR元件中的一些是已知的(例如,在其野生型环境中),但是该方法还导致鉴别了新型的5'-UTR元件和新型的3'-UTR元件,即本发明人发现分别在人细胞和小鼠细胞中表达的但是不为公共数据库(NCBI)所知的5'-UTR元件和3'-UTR元件:

[1172] 具体地,在由SEQ ID NO:1-72的人5'-UTR元件组成的组中,下述野生型5'-UTR元件不同于公共数据库(NCBI)已知的5'-UTR元件:SEQ ID NO:45,SEQ ID NO:47,SEQ ID NO:49。

[1173] 在由SEQ ID NO:73-136的小鼠5'-UTR元件组成的组中,下述野生型5'-UTR元件不同于公共数据库(NCBI)已知的5'-UTR元件:SEQ ID NO:77,SEQ ID NO:79,SEQ ID NO:84,SEQ ID NO:85,SEQ ID NO:90,SEQ ID NO:92,SEQ ID NO:93,SEQ ID NO:98,SEQ ID NO:106,SEQ ID NO:109,SEQ ID NO:115,SEQ ID NO:118,SEQ ID NO:125,SEQ ID NO:127,SEQ ID NO:131,SEQ ID NO:135。

[1174] 在由SEQ ID NO:174-203的小鼠3'-UTR元件组成的组中,下述野生型3'-UTR元件不同于公共数据库(NCBI)已知的3'-UTR元件:SEQ ID NO:179,SEQ ID NO:180,SEQ ID NO:182,SEQ ID NO:183,SEQ ID NO:197,SEQ ID NO:200。

[1175] 表1和2显示了新鉴别的5'-UTR元件和新鉴别的3'-UTR元件分别与已知的5'-UTR

元件和已知的3'-UTR元件的差异。具体而言,这些新鉴别的元件之前分别未被记载为3'-UTR元件或5'-UTR元件。

[1176] 表1:新型野生型5'-UTR元件

本发明的 5'-UTR 元件 (SEQ ID NO)	与 NCBI 的 GenBank 的条目(登记号或版本) 的区别
SEQ ID NO: 45	5'-端的前两个核苷酸(CC)不是 NM_199344.2 的一部分
SEQ ID NO: 47	包含 NM_001167671.1 和 NM_001167672.1 二者的核苷酸
SEQ ID NO: 49	在 NCBI 没有被鉴别为 5'-UTR。
SEQ ID NO: 77	SEQ ID NO: 77 的一部分对应于 BC006680.1
SEQ ID NO: 79	SEQ ID NO: 79 的一部分对应于 NM_026675.2
SEQ ID NO: 84	与 NM_001301840 相似
SEQ ID NO: 85	在 NCBI 没有被鉴别为 5'-UTR。
SEQ ID NO: 90	在 NCBI 没有被鉴别为 5'-UTR。
[1177] SEQ ID NO: 92	在 NCBI 没有被鉴别为 5'-UTR。
SEQ ID NO: 93	包含 NM_145942.4 加另外的上游碱基
SEQ ID NO: 98	比 NM_009004.4 略短
SEQ ID NO: 106	相对于 NM_001163359.1 有缺失
SEQ ID NO: 109	在 NCBI 没有被鉴别为 5'-UTR。
SEQ ID NO: 115	相对于 NM_144526.3, 包含在 5'端的另外的碱基
SEQ ID NO: 118	在 NCBI 没有被鉴别为 5'-UTR。
SEQ ID NO: 125	在 NCBI 没有被鉴别为 5'-UTR。
SEQ ID NO: 127	明显短于 NM_001110337.1
SEQ ID NO: 131	在 NCBI 没有被鉴别为 5'-UTR。
SEQ ID NO: 135	在 NCBI 没有被鉴别为 5'-UTR。
SEQ ID NO: 136	包含 NM_009970.2 的元件; 和 XM_011247136.1

[1178] 表2:新型的野生型3'-UTR元件

	本发明的 3'-UTR 元件 (SEQ ID NO)	与 NCBI 的 GenBank 的条目(登记号或版本) 的区别
[1179]	SEQ ID NO: 179	短于 NM_009951.4 的 3'-UTR
	SEQ ID NO: 180	短于 NM_019432.2 的 3'-UTR
	SEQ ID NO: 182	在 NCBI 没有找到。
	SEQ ID NO: 183	在 NCBI 没有找到。
	SEQ ID NO: 197	明显短于 NM_027372.1 的 3'-UTR
	SEQ ID NO: 200	很大程度上对应于 BC049670.1

[1180] 人工5'-UTR元件和人工3'-UTR元件

[1181] 在一些情形中制备人工5'-UTR元件和人工3'-UTR元件。这通过使用修饰的引物以基于PCR的方法修饰野生型5'-UTR元件和3'-UTR元件而进行,如下所述:

[1182] (i) 将在5'-UTR序列中的一些ATG三联体(如果存在的话)转化为三联体TAG。(i)之后,存在特定限制性酶的特定切割位点(如果存在的话)是不合要求的,万一所述特定切割位点被认为干扰后续的克隆实验。如果在后续克隆实验中使用的限制性酶(如在实施例2和3中所述)将识别并切割野生型3'-UTR元件或5'-UTR元件,则认为野生型序列中存在的切割位点干扰后续的克隆实验。由于此类5'-UTR元件和3'-UTR元件的内部切割是不需要的,通过用互补核苷酸替换在所述特定限制性酶的切割位点内的一个核苷酸而去除所述特定限制性酶的切割位点,由此去除所述特定限制性酶的切割位点。

[1183] SEQ ID NO:137(基于SEQ ID NO:3),SEQ ID NO:138,(基于SEQ ID NO:9),SEQ ID NO:139(基于SEQ ID NO:14),SEQ ID NO:140-(基于SEQ ID NO:17),SEQ ID NO:141(基于SEQ ID NO:18),SEQ ID NO:142(基于SEQ ID NO:29),SEQ ID NO:143(基于SEQ ID NO:31),SEQ ID NO:144(基于SEQ ID NO:42),SEQ ID NO:145(基于SEQ ID NO:43),SEQ ID NO:146(基于SEQ ID NO:52),SEQ ID NO:147(基于SEQ ID NO:109),SEQ ID NO:148(基于SEQ ID NO:110),SEQ ID NO:149(基于SEQ ID NO:119),SEQ ID NO:150(基于SEQ ID NO:120),和SEQ ID NO:151(基于SEQ ID NO:136)表示人工5'-UTR元件。SEQ ID NO:204(基于SEQ ID NO:192)表示人工3'-UTR元件。与各自野生型序列的准确差异如上文所示。

[1184] 总之,SEQ ID NO:1-136可以认为是野生型5'-UTR序列,并且SEQ ID NO:137-151可以认为是人工5'-UTR序列;SEQ ID NO:152-203可以认为是野生型3'-UTR序列,并且SEQ ID NO:204可以认为是人工3'-UTR序列。

[1185] 总之,该实施例提供一些5'-UTR元件和3'-UTR元件(野生型和人工的),它们分别可以相对于参比UTR元件(如下文实施例2和3中所述)检测其对翻译效率的相对影响。

[1186] 实施例2:参比UTR元件和包含其的参比构建体

[1187] 图1A显示用于本发明的参比构建体的RNA序列。还构建了包含对应于图1A的RNA序列的DNA序列的用于体外转录的载体。所述载体包含T7启动子、32L4(大核糖蛋白32)的5'非

翻译区(5'-UTR)的人工变体、编码萤火虫荧光素酶的富GC的序列(可读框,ORF)(PpLuc(GC))、人白蛋白(ALB)白蛋白7的3'非翻译区(3'-UTR)的人工变体。在白蛋白7的3'存在A64聚腺苷酸序列,之后是C30和组蛋白茎环序列。所述组蛋白茎环序列之后是用于在体外转录前将质粒线性化的限制性位点。

[1188] 将与实施例1所示的序列相对应的PCR片段分别克隆到包含与图1的RNA序列相对应的DNA序列的载体中。当实施例1的野生型序列在5'-UTR中包含ATG三联体时,和/或在5'-UTR或3'-UTR中不需要的限制性位点时,使用在实施例1中鉴别的各个人工序列(例如,使用SEQ ID NO:137替代SEQ ID NO:3)。

[1189] 详细地,为了产生用于测试候选5'-UTR元件的构建体,图1A中的单下划线5'-UTR元件替换为要测试的候选(野生型或人工的)5'-UTR元件(实施例1)(其实例显示在图1B中);并且为了产生用于测试候选3'-UTR元件的构建体,图1A中的双下划线3'-UTR元件替换为要测试的候选(野生型或人工的)3'-UTR元件(实施例1)(其实例显示在图1C中)。得到的载体或质粒(包含例如与RNA序列(例如,图1A至C中所示的RNA)相对应的DNA)也可以称为DNA模板。

[1190] 将所述DNA模板线性化并且使用T7-RNA聚合酶在体外转录(WO2015101416)。然后,将DNA模板用DNA酶处理消化。mRNA转录本包含5'-帽结构,其通过向转录反应添加过量的N<sup>7</sup>-甲基-鸟苷-5'-三磷酸-5'-鸟苷而得到。将这样得到的mRNA纯化并重悬在水中。这样得到的用于测试UTR元件的mRNA对应于图1A中所示的mRNA(参比构建体),不同之处在于参比构建体的5'-UTR元件或3'-UTR元件被要测试的各个UTR元件替换。

[1191] 实施例3:候选UTR元件相对于参比构建体的翻译效率

[1192] 实验测试人来源的或小鼠来源的5'-UTR元件或人来源的或小鼠来源的3'-UTR元件。在本实施例的情形内,“人来源的”可以意指人人野生型序列(例如,SEQ ID NO:1)或基于人野生型序列置换一个或多个碱基的人工序列(例如,SEQ ID NO:137),术语“小鼠来源的”相应地使用。在实施例1中给出了表3和4的人工序列如何以野生型序列为基础的详细内容。

[1193] 实验测试任何测试的(5'-或3'-)UTR(表3和4,见下)是否对翻译效率具有有益的效果。这通过包括下述步骤的方法进行:

[1194] (i) 用人工核酸分子转染哺乳动物细胞,并且在转染后一个或两个特定时间点(24和48h)测量由所述人工核酸分子的ORF编码的蛋白的表达量,

[1195] (ii) 用参比核酸分子(图1A)转染哺乳动物细胞,并且在转染后相同的一个或两个时间点(24和48h)测量由所述参比核酸分子的ORF编码的蛋白的表达量,

[1196] (iii) 计算从所述人工核酸分子表达的蛋白的量与从所述参比核酸分子报道的蛋白的量的比率,

[1197] 其中在(iii)中计算的比率 $\geq 1$ 。

[1198] 该比率与高翻译效率相关。

[1199] 在本实施例中进行实验测试的所有UTR序列还称为“实验测试的UTR元件”或“测试的UTR元件”或“实验测试的UTR”或“测试的UTR”。

[1200] (i) 的人工核酸分子包含测试的UTR元件。具体地,所述人工核酸分子(mRNA)在结构上与参比核酸分子(图1A)没有区别,区别在于参比核酸分子的5'-UTR元件或3'-UTR元件

被测试UTR元件替换(关于示例性的实例,参见图1B和1C)。

[1201] 3.1人来源的UTR元件

[1202] 下表3中给出了进行实验测试的人来源的UTR元件的详细信息。实验测试相对于上文和在“构建体详情”一节中详细描述参比构建体进行。

[1203] 表3:人来源的测试的UTR元件。

[1204] (H.s.=智人(Homo sapiens)=人)

[1205]

测试的 UTR				相应的野生型 序列	测定中的 位置
野生型 5'- UTR (SEQ ID NO)	人工 5'-UTR (SEQ ID NO)	野生型 3'- UTR (SEQ ID NO)	人工 3'- UTR (SEQ ID NO)		
SEQ ID NO: 1 - <i>H.s.</i> ZNF460 5'-UTR				ZNF460-5'- UTR	A1
SEQ ID NO: 2 - <i>H.s.</i> TGM2 5'-UTR				TGM2-5'- UTR	B1
	SEQ ID NO: 137 -( 基于 SEQ ID NO: 3 )			IL7R-5'-UTR	C1
SEQ ID NO: 4 - <i>H.s.</i> BGN 5'- UTR				BGN-5'- UTR	D1
SEQ ID NO: 5 - <i>H.s.</i> TK1 5'- UTR				TK1-5'-UTR	E1
SEQ ID NO: 6 - <i>H.s.</i> RAB3B 5'-UTR				RAB3B-5'- UTR	F1
SEQ ID NO: 7 - <i>H.s.</i> CBX6 5'-UTR				CBX6-5'- UTR	G1
SEQ ID NO: 8 - <i>H.s.</i> FZD2 5'-UTR				FZD2-5'- UTR	H1
	SEQ ID NO: 138 -( 基于 SEQ ID NO: 9 - <i>H.s.</i> COL8A1 5'- UTR)			COL8A1-5'- UTR	A2
SEQ ID NO:				NDUFS7-5'-	B2

[1206]

10 - <i>H.s.</i> NDUFS7 5'-UTR				UTR	
SEQ ID NO: 11 - <i>H.s.</i> PHGDH 5'-UTR				PHGDH-5'-UTR	C2
SEQ ID NO: 12 - <i>H.s.</i> PLK2 5'-UTR				PLK2-5'-UTR	D2
SEQ ID NO: 13 - <i>H.s.</i> TSPO 5'-UTR				TSPO-5'-UTR	E2
	SEQ ID NO: 139 -( 基于 SEQ ID NO: 14 - <i>H.s.</i> PTGS1 5'-UTR)			PTGS1-5'-UTR	F2
SEQ ID NO: 15 - <i>H.s.</i> FBXO32 5'-UTR				FBXO32-5'-UTR	G2
SEQ ID NO: 16 - <i>H.s.</i> NID2 5'-UTR				NID2-5'-UTR	H2
	SEQ ID NO: 140 ( SEQ ID NO: 17 - <i>H.s.</i> ATP5D 5'-UTR)			ATP5D-5'-UTR	A3
	SEQ ID NO: 141 序列 (基于 SEQ ID NO: 18 - <i>H.s.</i> EXOSC4 5'-UTR)			EXOSC4-5'-UTR	B3
SEQ ID NO: 19 - <i>H.s.</i> NOL9 5'-UTR				NOL9-5'-UTR	C3
SEQ ID NO: 20 - <i>H.s.</i> TUBB4B 5'-UTR				TUBB4B-5'-UTR	D3
SEQ ID NO: 21 - <i>H.s.</i> VPS18 5'-UTR				VPS18-5'-UTR	E3
SEQ ID NO: 22 - <i>H.s.</i> ORMDL2 5'-UTR				ORMDL2-5'-UTR	F3
SEQ ID NO: 23 - <i>H.s.</i>				FSCN1-5'-UTR	G3



[1207]

FSCN1 5'-UTR					
SEQ ID NO: 24 - <i>H.s.</i> TMEM33 5'- UTR				TMEM33-5'- UTR	H3
SEQ ID NO: 25 - <i>H.s.</i> TUBA4A 5'- UTR				TUBA4A-5'- UTR	A4
SEQ ID NO: 26 - <i>H.s.</i> EMP3 5'-UTR				EMP3-5'- UTR	B4
SEQ ID NO: 27 - <i>H.s.</i> TMEM201 5'- UTR				TMEM201- 5'-UTR	C4
SEQ ID NO: 28 - <i>H.s.</i> CRIP2 5'-UTR				CRIP2-5'- UTR	D4
	SEQ ID NO: 142 ( 基于 SEQ ID NO: 29 - <i>H.s.</i> BRAT1 5'-UTR)			BRAT1-5'- UTR	E4
SEQ ID NO: 30 - <i>H.s.</i> SERPINH1 5'- UTR				SERPINH1- 5'-UTR	F4
	SEQ ID NO: 143 ( 基于 SEQ ID NO: 31 - <i>H.s.</i> CD9 5'-UTR)			CD9-5'-UTR	G4
SEQ ID NO: 32 - <i>H.s.</i> DPYSL2 5'- UTR				DPYSL2-5'- UTR	H4
SEQ ID NO: 33 - <i>H.s.</i> CDK9 5'-UTR				CDK9-5'- UTR	A5
SEQ ID NO: 38 - <i>H.s.</i> TFRC 5'-UTR				TFRC-5'- UTR	B5
SEQ ID NO: 39 - <i>H.s.</i> PSMB3 5'- UTR				PSMB3-5'- UTR	C5
SEQ ID NO: 40 - <i>H.s.</i> FASN 5'-UTR				FASN-5'- UTR	D5
SEQ ID NO:				PSMB6-5'-	E5

[1208]

41 - <i>H.s.</i> PSMB6 5'-UTR				UTR	
	SEQ ID NO: 145 ( 基于 SEQ ID NO: 43 - <i>H.s.</i> PRSS56 5'-UTR)			PRSS56-5'- UTR	F5
SEQ ID NO: 44 - <i>H.s.</i> KPNA6 5'- UTR				KPNA6-5'- UTR	G5
SEQ ID NO: 45 - <i>H.s.</i> SFT2D2 5'- UTR				SFT2D2-5'- UTR	H5
SEQ ID NO: 46 - <i>H.s.</i> PARD6B 5'- UTR				PARD6B-5'- UTR	A6
SEQ ID NO: 47 - <i>H.s.</i> LPP 5'-UTR				LPP-5'-UTR	B6
SEQ ID NO: 48 - <i>H.s.</i> SPARC 5'-UTR				SPARC-5'- UTR	C6
SEQ ID NO: 49 - <i>H.s.</i> SCAND1 5'- UTR				SCAND1-5'- UTR	D6
SEQ ID NO: 50 - <i>H.s.</i> VASN 5'-UTR				VASN-5'- UTR	E6
SEQ ID NO: 51 - <i>H.s.</i> SLC26A1 5'- UTR				SLC26A1- 5'-UTR	F6
	SEQ ID NO: 146 -( 基于 SEQ ID NO: 52 - <i>H.s.</i> LCLAT1 5'-UTR)			LCLAT1-5'- UTR	G6
SEQ ID NO: 53 - <i>H.s.</i> FBXL18 5'- UTR				FBXL18-5'- UTR	H6
SEQ ID NO: 54 - <i>H.s.</i> SLC35F6 5'- UTR				SLC35F6-5'- UTR	A7
SEQ ID NO:				RAB3D-5'-	B7

[1209]

55 - <i>H.s.</i> RAB3D 5'- UTR				UTR	
SEQ ID NO: 56 - <i>H.s.</i> MAP1B 5'- UTR				MAP1B-5'- UTR	C7
SEQ ID NO: 57 - <i>H.s.</i> VMA21 5'- UTR				VMA21-5'- UTR	D7
SEQ ID NO: 59 - <i>H.s.</i> CYBA 5'-UTR				CYBA-5'- UTR	E7
SEQ ID NO: 60 - <i>H.s.</i> SEZ6L2 5'- UTR				SEZ6L2-5'- UTR	F7
SEQ ID NO: 61 - <i>H.s.</i> PCOLCE 5'- UTR				PCOLCE-5'- UTR	G7
SEQ ID NO: 63 - <i>H.s.</i> VTN 5'-UTR				VTN-5'-UTR	H7
SEQ ID NO: 64 - <i>H.s.</i> ALDH16A1 5'-UTR				ALDH16A1- 5'-UTR	A8
SEQ ID NO: 65 - <i>H.s.</i> RAVER1 5'- UTR				RAVER1-5'- UTR	B8
SEQ ID NO: 66 - <i>H.s.</i> KPNA6 5'- UTR				KPNA6-5'- UTR	C8
SEQ ID NO: 67 - <i>H.s.</i> SERINC5 5'- UTR				SERINC5-5'- UTR	D8
SEQ ID NO: 68 - <i>H.s.</i> JUP 5'-UTR				JUP-5'-UTR	E8
SEQ ID NO: 69 - <i>H.s.</i> CPN2 5'-UTR				CPN2-5'- UTR	F8
SEQ ID NO: 70 - <i>H.s.</i>				CRIP2-5'- UTR	G8

[1210]

CRIP2 5'-UTR					
SEQ ID NO: 71 - <i>H.s.</i> EPT1 5'-UTR				EPT1-5'- UTR	H8
SEQ ID NO: 72 - <i>H.s.</i> PNPO 5'-UTR				PNPO-5'- UTR	A9
		SEQ ID NO: 152 - 智人 NDUFS7 3'- UTR		NDUFS7-3'- UTR	B9
		SEQ ID NO: 153 - 智人 PHGDH 3'- UTR		PHGDH-3'- UTR	C9
		SEQ ID NO: 154 - 智人 TSPO 3'-UTR		TSPO-3'- UTR	D9
		SEQ ID NO: 155 - <i>H.s.</i> ATP5D 3'- UTR		ATP5D-3'- UTR	E9
		SEQ ID NO: 156 - <i>H.s.</i> EXOSC4 3'- UTR		EXOSC4-3'- UTR	F9
		SEQ ID NO: 157 - <i>H.s.</i> TUBB4B 3'- UTR		TUBB4B-3'- UTR	G9
		SEQ ID NO: 158 - <i>H.s.</i> TUBA4A 3'- UTR		TUBA4A-3'- UTR	H9
		SEQ ID NO: 159 - <i>H.s.</i> EMP3 3'-UTR		EMP3-3'- UTR	A10
		SEQ ID NO: 160 - <i>H.s.</i> CRIP2 3'-UTR		CRIP2-3'- UTR	B10
		SEQ ID NO: 161 - <i>H.s.</i> BRAT1 3'-UTR		BRAT1-3'- UTR	C10
		SEQ ID NO: 162 - <i>H.s.</i>		CD9-3'-UTR	D10

[1211]

		CD9 3'-UTR			
		SEQ ID NO: 163 - <i>H.s.</i> CDK9 3'-UTR		CDK9-3'- UTR	E10
		SEQ ID NO: 164 - <i>H.s.</i> PSMB3 3'- UTR		PSMB3-3'- UTR	F10
		SEQ ID NO: 165 - <i>H.s.</i> PSMB6 3'- UTR		PSMB6-3'- UTR	G10
		SEQ ID NO: 166 - <i>H.s.</i> PRSS56 3'- UTR		PRSS56-3'- UTR	H10
		SEQ ID NO: 167 - <i>H.s.</i> SCAND1 3'- UTR		SCAND1-3'- UTR	A11
		SEQ ID NO: 168 - <i>H.s.</i> AMN 3'-UTR		AMN-3'- UTR	B11
		SEQ ID NO: 169 - <i>H.s.</i> CYBA 3'-UTR		CYBA-3'- UTR	C11
		SEQ ID NO: 170 - <i>H.s.</i> PCOLCE 3'- UTR		PCOLCE-3'- UTR	D11
		SEQ ID NO: 171 - <i>H.s.</i> MAP1S 3'- UTR		MAP1S-3'- UTR	E11
		SEQ ID NO: 172 - <i>H.s.</i> VTN 3'-UTR		VTN-3'-UTR	F11
		SEQ ID NO: 173 - <i>H.s.</i> ALDH16A1 3'-UTR		ALDH16A1- 3'-UTR	G11
SEQ ID NO: 34 - <i>H.s.</i> SSSCA1 5'- UTR				SSSCA1-5'- UTR	H11
SEQ ID NO: 35 - <i>H.s.</i> POLR2L 5'- UTR				POLR2L-5'- UTR	A12

[1212]

SEQ ID NO: 36 - <i>H.s.</i> LIN7C 5'-UTR				LIN7C-5'- UTR	B12
SEQ ID NO: 37 - <i>H.s.</i> UQCR10 5'- UTR				UQCR10- 5'-UTR	C12
	SEQ ID NO: 144 ( 基于 SEQ ID NO: 42 - <i>H.s.</i> PYCRL 5'-UTR)			PYCRL-5'- UTR	D12
SEQ ID NO: 58 - <i>H.s.</i> AMN 5'-UTR				AMN-5'- UTR	E12
SEQ ID NO: 62 - <i>H.s.</i> MAP1S 5'- UTR				MAP1S-5'- UTR	F12
				参比构建体 (图 1A)	G12
				参比构建体 (图 1A)	H12

[1213] 3.2小鼠来源的UTR元件

[1214] 下表4中给出了进行实验测试的小鼠来源的UTR元件的详细信息。实验测试相对于上文和在“构建体详情”一节中详细描述参比构建体进行。

[1215] 表4: 小鼠来源的测试的UTR元件。

[1216] (M.m. = 小家鼠 (*Mus musculus*) = 小鼠)

[1217]

测试的UTR				相应的野生型 序列	测定中的 位置
野生型 5'- UTR (SEQ ID NO)	人工 5'-UTR (SEQ ID NO)	野生型 3'- UTR (SEQ ID NO)	人工 3'- UTR (SEQ ID NO)		
SEQ ID NO: 73 - <i>M.m.</i> Dpysl2 5'-UTR				Dpysl2-5'- UTR	A1
SEQ ID NO: 74 - <i>M.m.</i> Ccnd1 5'-UTR				Ccnd1-5'- UTR	B1
SEQ ID NO: 75 - <i>M.m.</i> Acox2 5'-UTR				Acox2-5'- UTR	C1

[1218]

SEQ ID NO: 76 – <i>M.m.</i> Cbx6 (Npcd) 5'-UTR				Cbx6-5'-UTR	D1
SEQ ID NO: 77 – <i>M.m.</i> Ubc 5'- UTR				Ubc-5'-UTR	E1
SEQ ID NO: 78 – <i>M.m.</i> Ldlr 5'- UTR				Ldlr-5'-UTR	F1
SEQ ID NO: 79 – <i>M.m.</i> Nudt22 5'UTR				Nudt22-5'- UTR	G1
SEQ ID NO: 80 – <i>M.m.</i> Pcyox11 5'- UTR				Pcyox11-5'- UTR	H1
SEQ ID NO: 81 – <i>M.m.</i> Ankrd1 5'-UTR				Ankrd1-5'- UTR	A2
SEQ ID NO: 82 – <i>M.m.</i> Tmem37 5'- UTR				Tmem37-5'- UTR	B2
SEQ ID NO: 83 – <i>M.m.</i> Tspyl4 5'-UTR				Tspyl4-5'- UTR	C2
SEQ ID NO: 84 – <i>M.m.</i> Slc7a3 5'-UTR				Slc7a3-5'- UTR	D2
SEQ ID NO: 85 – <i>M.m.</i> Cst6 5'- UTR				Cst6-5'-UTR	E2
SEQ ID NO: 86 – <i>M.m.</i> Aacs 5'-UTR				Aacs-5'-UTR	F2
SEQ ID NO: 87 – <i>M.m.</i> Nosip 5'-UTR				Nosip-5'-UTR	G2
SEQ ID NO: 88 – <i>M.m.</i> Itga7 5'-UTR				Itga7-5'-UTR	H2
SEQ ID NO: 89 – <i>M.m.</i> Ccnd2 5'-UTR				Ccnd2-5'- UTR	A3
SEQ ID NO: 90 – <i>M.m.</i> Ebp 5'- UTR				Ebp-5'-UTR	B3
SEQ ID NO: 91 – <i>M.m.</i> Sf3b5				Sf3b5-5'-UTR	C3

[1219]

5'-UTR					
SEQ ID NO: 92 – <i>M.m.</i> Fasn 5'-UTR				Fasn-5'-UTR	D3
SEQ ID NO: 93 – <i>M.m.</i> Hmgcs1 5'-UTR				Hmgcs1-5'-UTR	E3
SEQ ID NO: 94 – <i>M.m.</i> Osr1 5'-UTR				Osr1-5'-UTR	F3
SEQ ID NO: 95 – <i>M.m.</i> Lmnb1 5'-UTR				Lmnb1-5'-UTR	G3
SEQ ID NO: 97 – <i>M.m.</i> Vma21 5'-UTR				Vma21-5'-UTR	H3
SEQ ID NO: 98 – <i>M.m.</i> Kif20a 5'-UTR				Kif20a-5'-UTR	A4
SEQ ID NO: 99 – <i>M.m.</i> Cdca8 5'-UTR				Cdca8-5'-UTR	B4
SEQ ID NO: 100 – <i>M.m.</i> Slc7a1 5'-UTR				Slc7a1-5'-UTR	C4
SEQ ID NO: 101 – <i>M.m.</i> Ubqln2 5'-UTR				Ubqln2-5'-UTR	D4
SEQ ID NO: 102 – <i>M.m.</i> Prps2 5'-UTR				Prps2-5'-UTR	E4
SEQ ID NO: 103 – <i>M.m.</i> Shmt2 5'-UTR				Shmt2-5'-UTR	F4
SEQ ID NO: 105 – <i>M.m.</i> Aurkb 5'-UTR				Aurkb-5'-UTR	G4
SEQ ID NO: 106 – <i>M.m.</i> Fignl1 5'-UTR				Fignl1-5'-UTR	H4
SEQ ID NO: 107 – <i>M.m.</i> Cad 5'-UTR				Cad-5'-UTR	A5
SEQ ID NO: 108 – <i>M.m.</i> Anln 5'-UTR				Anln-5'-UTR	B5
	SEQ ID NO: 147 ( 基于 SEQ			Slfn9-5'-UTR	C5



[1220]

	ID NO: 109 - <i>M.m.</i> Slfn9 5'-UTR)				
	SEQ ID NO: 148 -( 基于 SEQ ID NO: 110 - <i>M.m.</i> Ncaph 5'- UTR)			Ncaph-5'- UTR	D5
SEQ ID NO: 112 - <i>M.m.</i> Pole 5'-UTR				Pole-5'-UTR	E5
SEQ ID NO: 113 - <i>M.m.</i> Uhrf1 5'-UTR				Uhrf1-5'-UTR	F5
SEQ ID NO: 114 - <i>M.m.</i> Gja1 5'-UTR				Gja1-5'-UTR	G5
SEQ ID NO: 115 - <i>M.m.</i> Fam64a 5'- UTR				Fam64a-5'- UTR	H5
SEQ ID NO: 116 - <i>M.m.</i> Kif2c 5'-UTR				Kif2c-5'-UTR	A6
SEQ ID NO: 117 - <i>M.m.</i> Tspan10 5'- UTR				Tspan10-5'- UTR	B6
SEQ ID NO: 118 - <i>M.m.</i> Scand1 5'-UTR				Scand1-5'- UTR	C6
	SEQ ID NO: 149 - 人工序列 (基于 SEQ ID NO: 119 - <i>M.m.</i> Gpr84 5'-UTR)			Gpr84-5'- UTR	D6
SEQ ID NO: 122 - <i>M.m.</i> Fads3 5'-UTR				Fads3-5'-UTR	E6
SEQ ID NO: 123 - <i>M.m.</i> Cers6 5'-UTR				Cers6-5'-UTR	F6
SEQ ID NO: 126 - <i>M.m.</i> Cxcr4 5'-UTR				Cxcr4-5'-UTR	G6
SEQ ID NO: 127 - <i>M.m.</i> Gprc5c 5'-UTR				Gprc5c-5'- UTR	H6

[1221]

SEQ ID NO: 128 – <i>M.m.</i> Fen1 5'-UTR				Fen1-5'-UTR	A7
SEQ ID NO: 129 – <i>M.m.</i> Cspg4 5'-UTR				Cspg4-5'-UTR	B7
SEQ ID NO: 130 – <i>M.m.</i> Mrpl34 5'-UTR				Mrpl34-5'- UTR	C7
SEQ ID NO: 131 – <i>M.m.</i> Comtd1 5'- UTR				Comtd1-5'- UTR	D7
SEQ ID NO: 132 – <i>M.m.</i> Armc6 5'-UTR				Armc6-5'- UTR	E7
SEQ ID NO: 133 – <i>M.m.</i> Emr4 5'-UTR				Emr4-5'-UTR	F7
SEQ ID NO: 134 – <i>M.m.</i> Atp5d 5'-UTR				Atp5d-5'-UTR	G7
SEQ ID NO: 135 – <i>M.m.</i> 1110001J03Rik 5'-UTR				1110001J03R ik-5'-UTR	H7
	SEQ ID NO: 151 – 人工序列 (基于 SEQ ID NO: 136 – <i>M.m.</i> Csf2ra 5'-UTR)			Csf2ra-5'- UTR	A8
		SEQ ID NO: 174 – <i>M.m.</i> Acox2 3'- UTR		Acox2-3'- UTR	B8
		SEQ ID NO: 175 – <i>M.m.</i> Ubc 3'-UTR		Ubc-3'-UTR	C8
		SEQ ID NO: 176 – <i>M.m.</i> Slpi 3'-UTR		Slpi-3'-UTR	D8
		SEQ ID NO: 178 – <i>M.m.</i> Pcyox1l 3'- UTR		Pcyox1l-3'- UTR	E8
		SEQ ID NO: 179 – <i>M.m.</i> Igf2bp1 3'-		Igf2bp1-3'- UTR	F8

[1222]

		UTR			
		SEQ ID NO: 180 – <i>M.m.</i> Tmem37 3'- UTR		Tmem37-3'- UTR	G8
		SEQ ID NO: 181 – <i>M.m.</i> Slc7a3 3'- UTR (240)		Slc7a3-3'- UTR	H8
		SEQ ID NO: 182 – <i>M.m.</i> Cst6 3'-UTR		Cst6-3'-UTR	A9
		SEQ ID NO: 183 – <i>M.m.</i> Ebp 3'-UTR		Ebp-3'-UTR	B9
		SEQ ID NO: 184 – <i>M.m.</i> Sf3b5 3'-UTR		Sf3b5-3'-UTR	C9
		SEQ ID NO: 185 – <i>M.m.</i> Plk1 3'-UTR		Plk1-3'-UTR	D9
		SEQ ID NO: 187 – <i>M.m.</i> Cdca8 3'- UTR		Cdca8-3'- UTR	E9
		SEQ ID NO: 188 – <i>M.m.</i> Kif22 3'-UTR		Kif22-3'-UTR	F9
		SEQ ID NO: 189 – <i>M.m.</i> Cad 3'-UTR		Cad-3'-UTR	G9
		SEQ ID NO: 190 – <i>M.m.</i> Cth 3'-UTR		Cth-3'-UTR	H9
		SEQ ID NO: 191 – <i>M.m.</i> Pole 3'-UTR		Pole-3'-UTR	A10
			SEQ ID NO: 204 (基 于 SEQ ID NO: 192 – <i>M.m.</i> Kif2c 3'-UTR)	Kif2c-3'-UTR	B10
		SEQ ID NO: 193 – <i>M.m.</i> Scand1 3'- UTR		Scand1-3'- UTR	C10
		SEQ ID NO:		Gpr84-3'-	D10

[1223]

		194 – <i>M.m.</i> Gpr84 3'- UTR		UTR	
		SEQ ID NO: 195 – <i>M.m.</i> Tpgs1 3'-UTR		Tpgs1-3'-UTR	E10
		SEQ ID NO: 196 – <i>M.m.</i> Ccl17 3'-UTR		Ccl17-3'-UTR	F10
		SEQ ID NO: 197 – <i>M.m.</i> Alkbh7 3'- UTR		Alkbh7-3'- UTR	G10
		SEQ ID NO: 198 – <i>M.m.</i> Ms4a8a 3'- UTR		Ms4a8a-3'- UTR	H10
		SEQ ID NO: 199 – <i>M.m.</i> Mrpl34 3'- UTR		Mrpl34-3'- UTR	A11
		SEQ ID NO: 200 – <i>M.m.</i> Comtd1 3'- UTR		Comtd1-3'- UTR	B11
		SEQ ID NO: 201 – <i>M.m.</i> Armc6 3'- UTR		Armc6-3'- UTR	C11
		SEQ ID NO: 202 – <i>M.m.</i> Atp5d 3'-UTR		Atp5d-3'-UTR	D11
		SEQ ID NO: 203 – <i>M.m.</i> 1110001J03R ik 3'-UTR		1110001J03R ik-3'-UTR	E11
SEQ ID NO: 96 – <i>M.m.</i> Aarsd1 5'-UTR				Aarsd1-5'- UTR	F11
SEQ ID NO: 104 – <i>M.m.</i> Kif22 5'-UTR				Kif22-5'-UTR	G11
SEQ ID NO: 111 – <i>M.m.</i> Cth 5'-UTR				Cth-5'-UTR	H11
	SEQ ID NO: 150 ( 基于 SEQ ID NO: 120 – <i>M.m.</i> Tpgs1 5'-			Tpgs1-5'-UTR	A12

	UTR)				
SEQ ID NO: 121 – <i>M.m.</i> Ccl17 5'-UTR				Ccl17-5'-UTR	B12
SEQ ID NO: 124 – <i>M.m.</i> Alkbh7 5'-UTR				Alkbh7-5'- UTR	C12
SEQ ID NO: 125 – <i>M.m.</i> Ms4a8a 5'- UTR				Ms4a8a-5'- UTR	D12
[1224]		SEQ ID NO: 177 – <i>M.m.</i> Nudt22 3'- UTR		Nudt22-3'- UTR	E12
		SEQ ID NO: 186 – <i>M.m.</i> Aarsd1 3'- UTR		Aarsd1-3'- UTR	F12
				参比构建体 (图 1A)	G12
				参比构建体 (图 1A)	H12

### [1225] 3.3构建体详情

[1226] 当测试5' -UTR时,这通过用“测试的”5' -UTR——即,野生型5' -UTR序列(与表3或4中第一列中引用的DNA序列相对应的mRNA序列)或基于其的人工5' -UTR序列(与表3或4中第二列中引用的DNA序列相对应的mRNA序列)替换参比构建体(即,32L4-PpLuc (GC) -白蛋白7-A64-C30-hSL;图1)的5' -UTR而进行。参比构建体的其他序列元件保持不变。

[1227] 依次,当测试3' -UTR时,这通过用“测试的”3' -UTR——即,野生型3' -UTR序列(表3或4中第三列中引用的DNA序列相对应的mRNA序列)或基于其的人工3' -UTR序列(与表3或4中第四列中引用的DNA序列相对应的mRNA序列)替换参比构建体(即,32L4-PpLuc (GC) -白蛋白7-A64-C30-hSL;图1)的3' -UTR而进行。参比构建体的其他序列元件保持不变。

### [1228] 3.4.用于确定翻译效率的荧光素酶测定

[1229] 使用不同的哺乳动物(人,小鼠)细胞系,即,HDF,L929,HepG2和Hela,在细胞转染测定中实验测试翻译效率(见图2至9)。

[1230] 将人皮肤成纤维细胞(HDF)、L929细胞、HEPG2细胞和HeLa细胞以 $1 \times 10^4$ 个细胞/孔的密度接种在96孔平板中。次日,将细胞在Opti-MEM中洗涤并且随后用25ng/孔的在Opti-MEM中的Lipofectamine2000-复合的PpLuc-编码mRNA转染。未转染的细胞用作对照。将编码海肾(*Renilla reniformis*)荧光素酶(RrLuc)的mRNA与PpLuc mRNA一起转染以控制转染效率(1ng RrLuc mRNA/孔)。开始转染后90分钟,将Opti-MEM换成培养基。转染后24和48小时,吸出培养基,并且将细胞在100μl的被动裂解缓冲液(Passive Lysis buffer)(Promega)中裂解。将裂解产物在-80℃储存,直到测量荧光素酶活性。

[1231] 在Hidex Chameleon平板读数器中以相对光单位(RLU)测量荧光素酶活性。在双重荧光素酶测定中从单个样品顺序测量Ppluc和Rrluc的活性。在2秒测量时间使用20μl裂解

产物和50μl Beetle汁 (pjk GmbH) 首先测量PpLuc活性。1500ms延迟后,利用50μl Renilla 汁 (pjk GmbH) 测量RrLuc活性。

[1232] 这在96孔平板中进行,并且将每个测试的UTR元件(在包含其的构建体的构架中,见3.3) 分配在96孔平板上的一个位置。使用两个96孔平板,一个用于测试人来源的UTR元件,另一个用于测试小鼠来源的UTR元件。在每个平板中,参比构建体(对应于如1A所示的mRNA序列) 在位置G12和H12。测试的UTR元件分别在表3和4中最后一列所示的位置。举例说明,包含小鼠Aarsd1的3' -UTR (SEQ ID NO:186) 的构建体在位置F12 (见表4)。

[1233] 在细胞转染后24小时 (1天,24小时),和在细胞转染后48小时 (2天,48小时) 测量荧光素酶活性。

[1234] 该测定的结果在下文3.5中描述。

[1235] 3.5结果——确定与高翻译效率相关的UTR元件

[1236] 在转染到不同的哺乳动物细胞系 (HDF,L929,HepG2和HeLa) 中后24小时 (1天,24小时) 和在48小时 (2天,48小时) 后测量的测定的表达结果 (见上文3.4) 显示在图2至9中。注意,存在两组96孔平板,一组涉及人来源的UTR元件,并且一组涉及小鼠来源的UTR元件。

[1237] 尽管图2至9是指96孔平板中的特定位置 (例如,A1),但是包3和4将特定的测试 (5' -或3' -) UTR元件分配到96孔平板的特定位置。例如,从表3可以看出,在测试的人UTR元件的96孔平板的位置A1,测试ZNF460-5' -UTR元件 (SEQ ID NO:1),如此等等。

[1238] 确定了一组特别优选的UTR元件 (还称为“最终选择”)。该组包括特征在于在至少一种所测试的细胞系中在1天 (24h) 后测量的翻译效率等于或优于、优选优于参比构建体 (图1A) 的5' -UTR元件和3' -UTR元件。该组落入四个亚组: (i) 人来源的5' -UTR元件 (见表5), (ii) 人来源的3' -UTR元件 (见表6), (iii) 小鼠来源的5' -UTR元件 (见表7), (iv) 小鼠来源的3' -UTR元件 (见表8)。

[1239] 表5: 优选的人来源的5' -UTR元件

[1240]

人来源的 5'-UTR 元件		相应的野生型 序列	测定中的 位置
野生型 5'-UTR (SEQ ID NO)	人工 5'-UTR (SEQ ID NO)		
SEQ ID NO: 1 - <i>H.s.</i> ZNF460 5'-UTR		ZNF460-5'-UTR	A1
SEQ ID NO: 2 - <i>H.s.</i> TGM2 5'-UTR		TGM2-5'-UTR	B1
	SEQ ID NO: 137 - (基 于 SEQ ID NO: 3 - <i>H.s.</i> IL7R-5'-UTR)	IL7R-5'-UTR	C1
	SEQ ID NO: 138 - (基 于 SEQ ID NO: 9 - <i>H.s.</i> COL8A1 5'-UTR)	COL8A1-5'-UTR	A2
SEQ ID NO: 10 - <i>H.s.</i> NDUFS7 5'-UTR		NDUFS7-5'-UTR	B2
SEQ ID NO: 12 - <i>H.s.</i> PLK2 5'-UTR		PLK2-5'-UTR	D2
SEQ ID NO: 15 - <i>H.s.</i> FBXO32 5'-UTR		FBXO32-5'-UTR	G2
	SEQ ID NO: 140 (基于 SEQ ID NO: 17 - <i>H.s.</i> ATP5D 5'-UTR)	ATP5D-5'-UTR	A3
SEQ ID NO: 20 - <i>H.s.</i> TUBB4B 5'-UTR		TUBB4B-5'-UTR	D3
SEQ ID NO: 22 - <i>H.s.</i>		ORMDL2-5'-UTR	F3

[1241]

<b>ORMDL2 5'-UTR</b>			
<b>SEQ ID NO: 23 - <i>H.s.</i> FSCN1 5'-UTR</b>		FSCN1-5'-UTR	G3
	<b>SEQ ID NO: 143 (基于 SEQ ID NO: 31 - <i>H.s.</i> CD9 5'-UTR)</b>	CD9-5'-UTR	G4
<b>SEQ ID NO: 32 - <i>H.s.</i> DPYSL2 5'-UTR</b>		DPYSL2-5'-UTR	H4
<b>SEQ ID NO: 39 - <i>H.s.</i> PSMB3 5'-UTR</b>		PSMB3-5'-UTR	C5
<b>SEQ ID NO: 41 - <i>H.s.</i> PSMB6 5'-UTR</b>		PSMB6-5'-UTR	E5
<b>SEQ ID NO: 44 - <i>H.s.</i> KPNA6 5'-UTR</b>		KPNA6-5'-UTR	G5
<b>SEQ ID NO: 45 - <i>H.s.</i> SFT2D2 5'-UTR</b>		SFT2D2-5'-UTR	H5
	<b>SEQ ID NO: 146 - (基 于 SEQ ID NO: 52 - <i>H.s.</i> LCLAT1 5'-UTR)</b>	LCLAT1-5'-UTR	G6
<b>SEQ ID NO: 53 - <i>H.s.</i> FBXL18 5'-UTR</b>		FBXL18-5'-UTR	H6
<b>SEQ ID NO: 54 - <i>H.s.</i> SLC35F6 5'-UTR</b>		SLC35F6-5'-UTR	A7
<b>SEQ ID NO: 57 - <i>H.s.</i> VMA21 5'-UTR</b>		VMA21-5'-UTR	D7
<b>SEQ ID NO: 60 - <i>H.s.</i> SEZ6L2 5'-UTR</b>		SEZ6L2-5'-UTR	F7
<b>SEQ ID NO: 61 - <i>H.s.</i> PCOLCE 5'-UTR</b>		PCOLCE-5'-UTR	G7
<b>SEQ ID NO: 63 - <i>H.s.</i> VTN 5'-UTR</b>		VTN-5'-UTR	H7
<b>SEQ ID NO: 64 - <i>H.s.</i> ALDH16A1 5'-UTR</b>		ALDH16A1-5'-UTR	A8
<b>SEQ ID NO: 66 - <i>H.s.</i> KPNA6 5'-UTR</b>		KPNA6-5'-UTR	C8
<b>SEQ ID NO: 68 - <i>H.s.</i> JUP 5'-UTR</b>		JUP-5'-UTR	E8
<b>SEQ ID NO: 69 - <i>H.s.</i> CPN2 5'-UTR</b>		CPN2-5'-UTR	F8
<b>SEQ ID NO: 72 - <i>H.s.</i> PNPO 5'-UTR</b>		PNPO-5'-UTR	A9
<b>SEQ ID NO: 34 - <i>H.s.</i> SSSCA1 5'-UTR</b>		SSSCA1-5'-UTR	H11
<b>SEQ ID NO: 35 - <i>H.s.</i> POLR2L 5'-UTR</b>		POLR2L-5'-UTR	A12
<b>SEQ ID NO: 36 - <i>H.s.</i> LIN7C 5'-UTR</b>		LIN7C-5'-UTR	B12
<b>SEQ ID NO: 37 - <i>H.s.</i> UQCR10 5'-UTR</b>		UQCR10-5'-UTR	C12



[1242]		<b>SEQ ID NO: 144 (基于 SEQ ID NO: 42 - <i>H.s.</i> PYCRL 5'-UTR)</b>	PYCRL-5'-UTR	D12
	<b>SEQ ID NO: 58 - <i>H.s.</i> AMN 5'-UTR</b>		AMN-5'-UTR	E12
	<b>SEQ ID NO: 62 - <i>H.s.</i> MAP1S 5'-UTR</b>		MAP1S-5'-UTR	F12

[1243] 由此,认为来自由下述组成的组的人基因的5'-UTR提供高翻译效率:ZNF460-5'-UTR,TGM2-5'-UTR,IL7R-5'-UTR,COL8A1-5'-UTR,NDUFS7-5'-UTR,PLK2-5'-UTR,FBX032-5'-UTR,ATP5D-5'-UTR,TUBB4B-5'-UTR,ORMDL2-5'-UTR,FSCN1-5'-UTR,CD9-5'-UTR,PYSL2-5'-UTR,PSMB3-5'-UTR,PSMB6-5'-UTR,KPNA6-5'-UTR,SFT2D2-5'-UTR,LCLAT1-5'-UTR,FBXL18-5'-UTR,SLC35F6-5'-UTR,VMA21-5'-UTR,SEZ6L2-5'-UTR,PCOLCE-5'-UTR,VTN-5'-UTR,ALDH16A1-5'-UTR,KPNA6-5'-UTR,JUP-5'-UTR,CPN2-5'-UTR,PNPO-5'-UTR,SSSCA1-5'-UTR,POLR2L-5'-UTR,LIN7C-5'-UTR,UQCR10-5'-UTR,PYCRL-5'-UTR,AMN-5'-UTR,MAP1S-5'-UTR。

[1244] 此外,认为下述反映的5'-UTR提供高翻译效率:SEQ ID NO:1,SEQ ID NO:2,SEQ ID NO:137(或其野生型等价物SEQ ID NO:3),SEQ ID NO:138(或其野生型等价物SEQ ID NO:9),SEQ ID NO:10,SEQ ID NO:12,SEQ ID NO:15,SEQ ID NO:140(或其野生型等价物SEQ ID NO:17),SEQ ID NO:20,SEQ ID NO:22,SEQ ID NO:23,SEQ ID NO:143(或其野生型等价物SEQ ID NO:31),SEQ ID NO:32,SEQ ID NO:39,SEQ ID NO:41,SEQ ID NO:44,SEQ ID NO:45,SEQ ID NO:146-(或其野生型等价物SEQ ID NO:52),SEQ ID NO:53,SEQ ID NO:54,SEQ ID NO:57,SEQ ID NO:60,SEQ ID NO:61,SEQ ID NO:63,SEQ ID NO:64,SEQ ID NO:66,SEQ ID NO:68,SEQ ID NO:69,SEQ ID NO:72,SEQ ID NO:34,SEQ ID NO:35,SEQ ID NO:36,SEQ ID NO:37,SEQ ID NO:144(或其野生型等价物SEQ ID NO:42),SEQ ID NO:58,SEQ ID NO:62。

[1245] 表6:优选的人来源的3'-UTR元件

[1246]

测试的野生型 3'-UTR 元件 (SEQ ID NO)	相应的野生型 序列	测定中的 位置
SEQ ID NO: 152 - <i>H.s.</i> NDUFS7 3'-UTR	NDUFS7-3'-UTR	B9
SEQ ID NO: 153 - <i>H.s.</i> PHGDH 3'-UTR	PHGDH-3'-UTR	C9
SEQ ID NO: 154 - <i>H.s.</i> TSPO 3'-UTR	TSPO-3'-UTR	D9
SEQ ID NO: 155 - <i>H.s.</i> ATP5D 3'-UTR	ATP5D-3'-UTR	E9
SEQ ID NO: 156 - <i>H.s.</i> EXOSC4 3'-UTR	EXOSC4-3'-UTR	F9
SEQ ID NO: 157 - <i>H.s.</i> TUBB4B 3'-UTR	TUBB4B-3'-UTR	G9
SEQ ID NO: 158 - <i>H.s.</i> TUBA4A 3'-UTR	TUBA4A-3'-UTR	H9
SEQ ID NO: 159 - <i>H.s.</i> EMP3 3'-UTR	EMP3-3'-UTR	A10
SEQ ID NO: 160 - <i>H.s.</i> CRIP2 3'-UTR	CRIP2-3'-UTR	B10
SEQ ID NO: 161 - <i>H.s.</i> BRAT1 3'-UTR	BRAT1-3'-UTR	C10
SEQ ID NO: 164 - <i>H.s.</i> PSMB3 3'-UTR	PSMB3-3'-UTR	F10
SEQ ID NO: 165 - <i>H.s.</i> PSMB6 3'-UTR	PSMB6-3'-UTR	G10
SEQ ID NO: 167 - <i>H.s.</i> SCAND1 3'-UTR	SCAND1-3'-UTR	A11
SEQ ID NO: 168 - <i>H.s.</i> AMN 3'-UTR	AMN-3'-UTR	B11
SEQ ID NO: 169 - <i>H.s.</i> CYBA 3'-UTR	CYBA-3'-UTR	C11
SEQ ID NO: 170 - <i>H.s.</i> PCOLCE 3'-UTR	PCOLCE-3'-UTR	D11
SEQ ID NO: 171 - <i>H.s.</i> MAP1S 3'-UTR	MAP1S-3'-UTR	E11
SEQ ID NO: 172 - <i>H.s.</i> VTN 3'-UTR	VTN-3'-UTR	F11
SEQ ID NO: 173 - <i>H.s.</i> ALDH16A1 3'-UTR	ALDH16A1-3'-UTR	G11

[1247] 由此,认为来自由下述组成的组的人基因的3'-UTR提供高翻译效率:NDUFS7-3'-UTR,PHGDH-3'-UTR,TSPO-3'-UTR,ATP5D-3'-UTR,EXOSC4-3'-UTR,TUBB4B-3'-UTR,TUBA4A-3'-UTR,EMP3-3'-UTR,CRIP2-3'-UTR,BRAT1-3'-UTR,PSMB3-3'-UTR,PSMB6-3'-UTR,SCAND1-3'-UTR,AMN-3'-UTR,CYBA-3'-UTR,PCOLCE-3'-UTR,MAP1S-3'-UTR,VTN-3'-UTR,ALDH16A1-3'-UTR。

[1248] 此外,认为由下述反映的3'-UTR提供高翻译效率:SEQ ID NO:152,SEQ ID NO:153,SEQ ID NO:154,SEQ ID NO:155,SEQ ID NO:156,SEQ ID NO:157,SEQ ID NO:158,SEQ ID NO:159,SEQ ID NO:160,SEQ ID NO:161,SEQ ID NO:164,SEQ ID NO:165,SEQ ID NO:167,SEQ ID NO:168,SEQ ID NO:169,SEQ ID NO:170,SEQ ID NO:171,SEQ ID NO:172,SEQ ID NO:173。

[1249] 表7:优选的小鼠来源的5'-UTR元件

[1250]

测试的 UTR		相应的野生型 序列	测定中的 位置
野生型 5'-UTR 元件 (SEQ ID NO)	人工 5'-UTR (SEQ ID NO)		
SEQ ID NO: 73 - <i>M.m.</i> Dpysl2 5'-UTR		Dpysl2-5'-UTR	A1
SEQ ID NO: 75 - <i>M.m.</i> Acox2 5'-UTR		Acox2-5'-UTR	C1
SEQ ID NO: 77 - <i>M.m.</i> Ubc 5'-UTR		Ubc-5'-UTR	E1
SEQ ID NO: 79 - <i>M.m.</i> Nudt22 5'-UTR		Nudt22-5'-UTR	G1
SEQ ID NO: 80 - <i>M.m.</i> Pcyox11 5'-UTR		Pcyox11-5'-UTR	H1
SEQ ID NO: 81 - <i>M.m.</i> Ankrd1 5'-UTR		Ankrd1-5'-UTR	A2
SEQ ID NO: 83 - <i>M.m.</i> Tspyl4 5'-UTR		Tspyl4-5'-UTR	C2
SEQ ID NO: 84 - <i>M.m.</i> Slc7a3 5'-UTR		Slc7a3-5'-UTR	D2
SEQ ID NO: 86 - <i>M.m.</i> Aacs 5'-UTR		Aacs-5'-UTR	F2
SEQ ID NO: 87 - <i>M.m.</i> Nosip 5'-UTR		Nosip-5'-UTR	G2
SEQ ID NO: 88 - <i>M.m.</i> Itga7 5'-UTR		Itga7-5'-UTR	H2
SEQ ID NO: 89 - <i>M.m.</i> Cend2 5'-UTR		Cend2-5'-UTR	A3
SEQ ID NO: 90 - <i>M.m.</i> Ebp 5'-UTR		Ebp-5'-UTR	B3
SEQ ID NO: 91 - <i>M.m.</i> Sf3b5 5'-UTR		Sf3b5-5'-UTR	C3
SEQ ID NO: 92 - <i>M.m.</i> Fasn 5'-UTR		Fasn-5'-UTR	D3
SEQ ID NO: 93 - <i>M.m.</i> Hmgcs1 5'-UTR		Hmgcs1-5'-UTR	E3
SEQ ID NO: 94 - <i>M.m.</i> Osr1 5'-UTR		Osr1-5'-UTR	F3
SEQ ID NO: 95 - <i>M.m.</i> Lmnbl1 5'-UTR		Lmnbl1-5'-UTR	G3
SEQ ID NO: 97 - <i>M.m.</i> Vma21 5'-UTR		Vma21-5'-UTR	H3
SEQ ID NO: 98 - <i>M.m.</i>		Kif20a-5'-UTR	A4

[1251]

<b>Kif20a 5'-UTR</b>			
<b>SEQ ID NO: 99 – <i>M.m.</i> Cdca8 5'-UTR</b>		Cdca8-5'-UTR	B4
<b>SEQ ID NO: 100 – <i>M.m.</i> Slc7a1 5'-UTR</b>		Slc7a1-5'-UTR	C4
<b>SEQ ID NO: 101 – <i>M.m.</i> Ubqln2 5'-UTR</b>		Ubqln2-5'-UTR	D4
<b>SEQ ID NO: 102 – <i>M.m.</i> Prps2 5'-UTR</b>		Prps2-5'-UTR	E4
<b>SEQ ID NO: 103 – <i>M.m.</i> Shmt2 5'-UTR</b>		Shmt2-5'-UTR	F4
<b>SEQ ID NO: 106 – <i>M.m.</i> Fignl1 5'-UTR</b>		Fignl1-5'-UTR	H4
<b>SEQ ID NO: 107 – <i>M.m.</i> Cad 5'-UTR</b>		Cad-5'-UTR	A5
<b>SEQ ID NO: 108 – <i>M.m.</i> Anln 5'-UTR</b>		Anln-5'-UTR	B5
	<b>SEQ ID NO: 147 (基于 SEQ ID NO: 109 – <i>M.m.</i> Slfn9 5'-UTR)</b>	Slfn9-5'-UTR	C5
	<b>SEQ ID NO: 148 – (基于 SEQ ID NO: 110 – <i>M.m.</i> Ncaph 5'-UTR)</b>	Ncaph-5'-UTR	D5
<b>SEQ ID NO: 112 – <i>M.m.</i> Pole 5'-UTR</b>		Pole-5'-UTR	E5
<b>SEQ ID NO: 113 – <i>M.m.</i> Uhrfl 5'-UTR</b>		Uhrfl-5'-UTR	F5
<b>SEQ ID NO: 114 – <i>M.m.</i> Gja1 5'-UTR</b>		Gja1-5'-UTR	G5
<b>SEQ ID NO: 115 – <i>M.m.</i> Fam64a 5'-UTR</b>		Fam64a-5'-UTR	H5
<b>SEQ ID NO: 117 – <i>M.m.</i> Tspan10 5'-UTR</b>		Tspan10-5'-UTR	B6
<b>SEQ ID NO: 118 – <i>M.m.</i> Scand1 5'-UTR</b>		Scand1-5'-UTR	C6
	<b>SEQ ID NO: 149 (基于 SEQ ID NO: 119 – <i>M.m.</i> Gpr84 5'-UTR)</b>	Gpr84-5'-UTR	D6
<b>SEQ ID NO: 123 – <i>M.m.</i> Cers6 5'-UTR</b>		Cers6-5'-UTR	F6
<b>SEQ ID NO: 126 – <i>M.m.</i> Cxcr4 5'-UTR</b>		Cxcr4-5'-UTR	G6
<b>SEQ ID NO: 127 – <i>M.m.</i> Gpre5c 5'-UTR</b>		Gpre5c-5'-UTR	H6
<b>SEQ ID NO: 128 – <i>M.m.</i> Fen1 5'-UTR</b>		Fen1-5'-UTR	A7
<b>SEQ ID NO: 129 – <i>M.m.</i> Cspg4 5'-UTR</b>		Cspg4-5'-UTR	B7

[1252]

<b>SEQ ID NO: 130 – <i>M.m.</i> Mrpl34 5'-UTR</b>		Mrpl34-5'-UTR	C7
<b>SEQ ID NO: 131 – <i>M.m.</i> Comtd1 5'-UTR</b>		Comtd1-5'-UTR	D7
<b>SEQ ID NO: 132 – <i>M.m.</i> Armc6 5'-UTR</b>		Armc6-5'-UTR	E7
<b>SEQ ID NO: 133 – <i>M.m.</i> Emr4 5'-UTR</b>		Emr4-5'-UTR	F7
<b>SEQ ID NO: 134 – <i>M.m.</i> Atp5d 5'-UTR</b>		Atp5d-5'-UTR	G7
	<b>SEQ ID NO: 151 (基 于 SEQ ID NO: 136 – <i>M.m.</i> Csf2ra 5'-UTR)</b>	Csf2ra-5'-UTR	A8
<b>SEQ ID NO: 96 – <i>M.m.</i> Aarsd1 5'-UTR</b>		Aarsd1-5'-UTR	F11
<b>SEQ ID NO: 111 – <i>M.m.</i> Cth 5'-UTR</b>		Cth-5'-UTR	H11
	<b>SEQ ID NO: 150 (基 于 SEQ ID NO: 120 – <i>M.m.</i> Tpgs1 5'-UTR)</b>	Tpgs1-5'-UTR	A12
<b>SEQ ID NO: 121 – <i>M.m.</i> Ccl17 5'-UTR</b>		Ccl17-5'-UTR	B12
<b>SEQ ID NO: 124 – <i>M.m.</i> Alkbh7 5'-UTR</b>		Alkbh7-5'-UTR	C12
<b>SEQ ID NO: 125 – <i>M.m.</i> Ms4a8a 5'-UTR</b>		Ms4a8a-5'-UTR	D12

[1253] 由此,认为来自由下述组成的组的小鼠基因的5'-UTR提供高翻译效率:Dpys12-5'-UTR, Acox2-5'-UTR, Ubc-5'-UTR, Nudt22-5'-UTR, Pcyox11-5'-UTR, Ankrd1-5'-UTR, Tspyl4-5'-UTR, Slc7a3-5'-UTR, Aacs-5'-UTR, Nosip-5'-UTR, Itga7-5'-UTR, Ccnd2-5'-UTR, Ebp-5'-UTR, Sf3b5-5'-UTR, Fasn-5'-UTR, Hmgcs1-5'-UTR, Osr1-5'-UTR, Lmnb1-5'-UTR, Vma21-5'-UTR, Kif20a-5'-UTR, Cdca8-5'-UTR, Slc7a1-5'-UTR, Ubqln2-5'-UTR, Prps2-5'-UTR, Shmt2-5'-UTR, Figl1-5'-UTR, Cad-5'-UTR, Anln-5'-UTR, Slfn9-5'-UTR, Ncaph-5'-UTR, Pole-5'-UTR, Uhrf1-5'-UTR, Gja1-5'-UTR, Fam64a-5'-UTR, Tspan10-5'-UTR, Scand1-5'-UTR, Gpr84-5'-UTR, Cers6-5'-UTR, Cxcr4-5'-UTR, Gprc5c-5'-UTR, Fen1-5'-UTR, Cspg4-5'-UTR, Mrpl34-5'-UTR, Comtd1-5'-UTR, Armc6-5'-UTR, Emr4-5'-UTR, Atp5d-5'-UTR, Csf2ra-5'-UTR, Aarsd1-5'-UTR, Cth-5'-UTR, Tpgs1-5'-UTR, Ccl17-5'-UTR, Alkbh7-5'-UTR, Ms4a8a-5'-UTR。

[1254] 此外,认为由下述反映的5'-UTR提供高翻译效率:SEQ ID NO:73, SEQ ID NO:75, SEQ ID NO:77, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:147(或其野生型等价物SEQ ID NO:109), SEQ ID NO:148(或其野生型等价物SEQ ID NO:110), SEQ ID NO:112, SEQ ID NO:113, SEQ ID NO:114, SEQ ID NO:115, SEQ ID NO:117, SEQ ID NO:118, SEQ ID NO:149

(或其野生型等价物SEQ ID NO:119),SEQ ID NO:123,SEQ ID NO:126,SEQ ID NO:127,SEQ ID NO:128,SEQ ID NO:129,SEQ ID NO:130,SEQ ID NO:131,SEQ ID NO:132,SEQ ID NO:133,SEQ ID NO:134,SEQ ID NO:151(或其野生型等价物SEQ ID NO:136),SEQ ID NO:96,SEQ ID NO:111,SEQ ID NO:150(基于SEQ ID NO:120),SEQ ID NO:121,SEQ ID NO:124,SEQ ID NO:125。

[1255] 表8:优选的小鼠来源的3'-UTR元件

[1256]

测试的 UTR		相应的野生型 序列	测定中的 位置
野生型 3'-UTR (SEQ ID NO)	人工 3'-UTR (SEQ ID NO)		
SEQ ID NO: 174 – <i>M.m.</i> Acox2 3'-UTR		Acox2-3'-UTR	B8
SEQ ID NO: 175 – <i>M.m.</i> Ubc 3'-UTR		Ubc-3'-UTR	C8
SEQ ID NO: 176 – <i>M.m.</i> Slpi 3'-UTR		Slpi-3'-UTR	D8
SEQ ID NO: 179 – <i>M.m.</i> Igf2bp1 3'-UTR		Igf2bp1-3'-UTR	F8
SEQ ID NO: 180 – <i>M.m.</i> Tmem37 3'-UTR		Tmem37-3'-UTR	G8
SEQ ID NO: 181 – <i>M.m.</i> Slc7a3 3'-UTR (240)		Slc7a3-3'-UTR	H8
SEQ ID NO: 182 – <i>M.m.</i> Cst6 3'-UTR		Cst6-3'-UTR	A9

[1257]

SEQ ID NO: 183 – <i>M.m.</i> Ebp 3'-UTR		Ebp-3'-UTR	B9
SEQ ID NO: 184 – <i>M.m.</i> Sf3b5 3'-UTR		Sf3b5-3'-UTR	C9
SEQ ID NO: 187 – <i>M.m.</i> Cdca8 3'-UTR		Cdca8-3'-UTR	E9
SEQ ID NO: 188 – <i>M.m.</i> Kif22 3'-UTR		Kif22-3'-UTR	F9
SEQ ID NO: 189 – <i>M.m.</i> Cad 3'-UTR		Cad-3'-UTR	G9
SEQ ID NO: 191 – <i>M.m.</i> Pole 3'-UTR		Pole-3'-UTR	A10
	SEQ ID NO: 204 (基 于 SEQ ID NO: 192 – <i>M.m.</i> Kif2c 3'-UTR)	Kif2c-3'-UTR	B10
SEQ ID NO: 193 – <i>M.m.</i> Scand1 3'-UTR		Scand1-3'-UTR	C10
SEQ ID NO: 194 – <i>M.m.</i> Gpr84 3'-UTR		Gpr84-3'-UTR	D10
SEQ ID NO: 195 – <i>M.m.</i> Tpgs1 3'-UTR		Tpgs1-3'-UTR	E10
SEQ ID NO: 196 – <i>M.m.</i> Ccl17 3'-UTR		Ccl17-3'-UTR	F10
SEQ ID NO: 197 – <i>M.m.</i> Alkbh7 3'-UTR		Alkbh7-3'-UTR	G10
SEQ ID NO: 198 – <i>M.m.</i> Ms4a8a 3'-UTR		Ms4a8a-3'-UTR	H10
SEQ ID NO: 199 – <i>M.m.</i> Mrpl34 3'-UTR		Mrpl34-3'-UTR	A11
SEQ ID NO: 200 – <i>M.m.</i> Comtd1 3'-UTR		Comtd1-3'-UTR	B11
SEQ ID NO: 201 – <i>M.m.</i> Armc6 3'-UTR		Armc6-3'-UTR	C11
SEQ ID NO: 202 – <i>M.m.</i> Atp5d 3'-UTR		Atp5d-3'-UTR	D11
SEQ ID NO: 203 – <i>M.m.</i> 1110001J03Rik 3'-UTR		1110001J03Rik-3'- UTR	E11
SEQ ID NO: 177 – <i>M.m.</i> Nudt22 3'-UTR		Nudt22-3'-UTR	E12

[1258] 由此,认为来自由下述组成的组的小鼠基因的3'-UTR提供高翻译效率: Acox2-3'-UTR, Ubc-3'-UTR, Slpi-3'-UTR, Igf2bp1-3'-UTR, Tmem37-3'-UTR, Slc7a3-3'-UTR, Cst6-3'-UTR, Ebp-3'-UTR, Sf3b5-3'-UTR, Cdca8-3'-UTR, Kif22-3'-UTR, Cad-3'-UTR, Pole-3'-UTR, Kif2c-3'-UTR, Scand1-3'-UTR, Gpr84-3'-UTR, Tpgs1-3'-UTR, Ccl17-3'-UTR, Alkbh7-3'-UTR, Ms4a8a-3'-UTR, Mrpl34-3'-UTR, Comtd1-3'-UTR, Armc6-3'-UTR, Atp5d-3'-UTR, 1110001J03Rik-3'-UTR, Nudt22-3'-UTR。

[1259] 此外,认为由下述反映的3'-UTR提供高翻译效率: SEQ ID NO: 174, SEQ ID NO: 175, SEQ ID NO: 176, SEQ ID NO: 179, SEQ ID NO: 180, SEQ ID NO: 181, SEQ ID NO: 182, SEQ ID NO: 183, SEQ ID NO: 184, SEQ ID NO: 187, SEQ ID NO: 188, SEQ ID NO: 189, SEQ ID NO:

191,SEQ ID NO:204(或其野生型等价物SEQ ID NO:192),SEQ ID NO:193,SEQ ID NO:194,SEQ ID NO:195,SEQ ID NO:196,SEQ ID NO:197,SEQ ID NO:198,SEQ ID NO:199,SEQ ID NO:200,SEQ ID NO:201,SEQ ID NO:202,SEQ ID NO:203,SEQ ID NO:177。



## 序列表

[0001]	<110>	库瑞瓦格股份公司	
	<120>	新型人工核酸分子	
	<130>	CU01P191W01	
	<160>	211	
	<170>	PatentIn version 3.5	
	<210>	1	
	<211>	322	
	<212>	DNA	
	<213>	智人 (Homo sapiens)	
	<400>	1	
		gaaacagtgt ggggcctaga gcgctgggtg ggcgcgttct gcggcctgag cagggacggg	60
		tagtgaagcg gttacgcccc ttcttcgcgt cttggcgggg gcctgacgcc ccgcttctcc	120
		cctaacgagg tgteccaccg gcgcccgcgc aggcctaggc ctccgcagcc gccctccgtc	180
		tcctcagccc cgacgctgcg ccttgggcct tgtgcgcatt ttttcgggg gaaaactgag	240
		gctcggagtg cgaaagtcag ccgaggtcgc cccgcccagg acagagaagg gctgtggtcg	300
		gctgatccgc ggcattcccg gg	322
	<210>	2	
	<211>	100	
	<212>	DNA	
	<213>	智人 (Homo sapiens)	
	<400>	2	
		ataagttagc gccgctctcc gcctcggcag tgccagccgc cagtggtcgc acttggaggg	60
		tctcgccgcc agtgggaagga gccaccgcc cgcgccgacc	100
	<210>	3	
	<211>	103	
	<212>	DNA	
	<213>	智人 (Homo sapiens)	
	<400>	3	
		atctaagctt ctctgtcttc ctccctccct cccttctctt tactctcatt catttcatac	60
		acactggctc acacatctac tctctctctc tatctctctc aga	103
	<210>	4	
	<211>	236	
	<212>	DNA	
	<213>	智人 (Homo sapiens)	
	<400>	4	
		cttttctctc ctccccgcgc tctccccgct gtccccctcc cgtcggcccg cctgcccagc	60
		cttttagctc ccgcccgcgc cctctgtctc cctctctcca caaactgccc aggagtgagt	120
		agctgcttgc ggctcgccgc acacaccgga cagatagacg tgcggacggc ccaccacccc	180
		agcccgcaca ctagtcagcc tgcgcctggc gcctccccctc tccaggtcca tccgcc	236
	<210>	5	
	<211>	210	
	<212>	DNA	
	<213>	智人 (Homo sapiens)	
	<400>	5	
		gcgcacgtcc cggattctct ccacgagggg gcgggctgcg gccaaatctc ccgccaggtc	60
		agcgggccgg cgctgattgg ccccatggcg gcggggccgg ctctgtgattg gccagcacgc	120
		cgtggtttta agcggtcggc gcgggaacca ggggcttact gcgggacggc cttggagagt	180
		actcgggttc gtgaacttcc cggaggcgca	210
	<210>	6	
	<211>	213	
	<212>	DNA	
	<213>	智人 (Homo sapiens)	
	<400>	6	
		agactccgcc cttgggcggg gcctggatgc ggccggagcg gagcagtgcg ggagcgggag	60
		cctcagccct caggcgccac tgtgaggacc tgaccggacc agaccatccc gcagcgcccc	120

	gccccggccc cctccgcgcc ctcccgcgc caggtcctgc cgtcccgcgc accgtccggg	180
	agcgaacccg tcgtcccgc ctcggagtcc gcg	213
	<210> 7	
	<211> 63	
	<212> DNA	
	<213> 智人 (Homo sapiens)	
	<400> 7	
	gagcgggtgcc gcaccggccg cgggcgcagg gagtattatg ggctgtgggt gccgctgagc	60
	aag	63
	<210> 8	
	<211> 245	
	<212> DNA	
	<213> 智人 (Homo sapiens)	
	<400> 8	
	gaggagagag ggcagcagcg cgcggtgtct ccggctgctc agtccgaccg cggcaagcaa	60
	gcgggcaggc gcaccgcccc ctcccccgcc cggcctcccc aactctgcgg ccgcgagtaa	120
	agtttgcaaa gaggcgcggg aggcggcagc cgcagcgagg aggcggcggg gaagaagcgc	180
	agtctccggg ttgggggcgg gggcgggggg ggcgccaagg agccgggttg gggcggcgg	240
	ccagc	245
	<210> 9	
	<211> 259	
	<212> DNA	
	<213> 智人 (Homo sapiens)	
	<400> 9	
	atcacagccc ttccccgac cttccgttg gagccagcga gcctctctcc ctgatcttac	60
	gtgctcaagg gagctcacac gttcaccaac tcacccttga agtcatctca agaacaaaag	120
	acaactgaaa gaagctgttg tgaaggcaga gcagcatctg ctgaagagac agaaaccagc	180
	cccagagtg tcacaggaag gcaccagcaa ggacatttgt ctttgatttg attcagcagt	240
	cctgtcaagt ataaatgtg	259
	<210> 10	
	<211> 43	
	<212> DNA	
	<213> 智人 (Homo sapiens)	
	<400> 10	
	aaggagaacg gacctcagag gttgtctgaa ggccgaggcc aag	43
	<210> 11	
	<211> 227	
	<212> DNA	
	<213> 智人 (Homo sapiens)	
	<400> 11	
	gcagggattt ggcaacctca gagccgcgag gaggaggcgg agtcgcggag agtttgagta	60
	tttccgtcca atcaaaagga gactgtaaga ggaggaggag gaggagatga ctggggagcg	120
	ggagctggag aatactgccc agttactcta gcgcgccagg ccgaaccgca gttcttggc	180
	ttaggtactt ctactcacag cggccgattc cgaggccaac tccagca	227
	<210> 12	
	<211> 180	
	<212> DNA	
	<213> 智人 (Homo sapiens)	
	<400> 12	
	tccgccccct tccgcctcc ccgtatataa gacttcgccg agcgtctctca ctgcacaag	60
	tggaccgggg tggtgggtgc tagtcggcac cagaggcaag ggtgcgagga ccacggccgg	120
	ctcggacgtg tgaccgcgcc tagggggtgg cagcgggcag tgcggggcgg caaggcgacc	180
	<210> 13	
	<211> 120	
	<212> DNA	
	<213> 智人 (Homo sapiens)	
	<400> 13	

[0002]

	ggcggctggg aggggcgggg cggatgcggg gacagcggcc tggctaactc ctgccaggca	60
	gtgcccttcc cggagcgtgc cctcgccgct gagctcccct gaacagcagc tgcagcagcc	120
	<210> 14	
	<211> 80	
	<212> DNA	
	<213> 智人 (Homo sapiens)	
	<400> 14	
	atgggctgga gctccgggca gtgtgcgagg cgcacgcaca ggagcctgca ctctgcgtcc	60
	cgcacccag cagccgcgcc	80
	<210> 15	
	<211> 239	
	<212> DNA	
	<213> 智人 (Homo sapiens)	
	<400> 15	
	gctgcggcgg gctgcgggga taaatactgc ggagctact gccgcgcagc actcccgagg	60
	cctgcaacgc ttgagatcct ctccgcgcc gccacccgc aggggtgccc gcgccgttcc	120
	cgcgcgcccg ccgccccgt cgcgggcccc tgcacccga gcatccgcc cgggtggcac	180
	gtccccgagc ccaccaggcc ggccccgtct ccccatccgt ctagtccgt cgcggtgcc	239
	<210> 16	
	<211> 234	
	<212> DNA	
	<213> 智人 (Homo sapiens)	
	<400> 16	
	gccttagaaa agttaacgag aaccagatgt ggtggccact gccgaacttt ctgagagccg	60
	gtgattggtc cccagccgag ggccctcagcc aattagcttg ctgggtgggc ctggagttcc	120
	gccccgccc ggcccccgc gagatccagg ttccagagctg gcgcggcgcg gagagtgggc	180
	tggaggccgg ggccggacgc gttgtgcagc gggtaaagc acggccgagc gage	234
	<210> 17	
[0003]	<211> 101	
	<212> DNA	
	<213> 智人 (Homo sapiens)	
	<400> 17	
	cagacgtccc tgcggtcgt cctcctgcc ctccaggccg cccgcgcgc gccggagtcc	60
	gctgtccgcc agctacccgc ttctgcccgc ccgcgctgc c	101
	<210> 18	
	<211> 110	
	<212> DNA	
	<213> 智人 (Homo sapiens)	
	<400> 18	
	atggcggacc tccggaacc gtagattccg ggccgtcgga gccgccggga gctgtagttc	60
	tcccgcggct cagagaagta ggcagagagc ggacctggcg gccgggcagc	110
	<210> 19	
	<211> 96	
	<212> DNA	
	<213> 智人 (Homo sapiens)	
	<400> 19	
	ggctcggagg cccctctttt cggcagctgg gagggcagcg gggcggaccg acgggatctc	60
	gcgagattct cgcgcggtag gccctgagga cccagc	96
	<210> 20	
	<211> 102	
	<212> DNA	
	<213> 智人 (Homo sapiens)	
	<400> 20	
	atataagcgt tggcggagcg tcggtttag cactctgcgc gcccgctctt ctgctgctgt	60
	ttgtctactt cctcctgctt ccccgccgc gcccgccca tc	102
	<210> 21	
	<211> 339	

	<212> DNA	
	<213> 智人 (Homo sapiens)	
	<400> 21	
	gccccgctca cgggggcggg agtcagctga gctgccgggg cgagggtggg atcacctggc	60
	accggctgaa gggagcctgt gatttttttg tagcgggggc ggggagtaag gtgcaagact	120
	gcgccagatt caaggacgag ggctgccga ttatctcgct gcataaggca agagcaagag	180
	gacctcagg attttaaaga ggagcgacg gctgcaggtt cccaggatct gtcagaggct	240
	ggggagttac agcttccatt ctggggcgac ggggacccc ggggggtagc cttttgtaa	300
	tccccaggcc ccggacaaag agcccagagg ccgggcacc	339
	<210> 22	
	<211> 92	
	<212> DNA	
	<213> 智人 (Homo sapiens)	
	<400> 22	
	aaggggcaga aggagaggcg ttacttcttg gagacttcag gtgtggtagc cggcgccgcg	60
	cccatagccg gacggggatc tgagctggca gg	92
	<210> 23	
	<211> 132	
	<212> DNA	
	<213> 智人 (Homo sapiens)	
	<400> 23	
	agctgggctt tgtggagcgc tgcggagggt gcgtgcgggc cgcggcagcc gaacaaagga	60
	gcaggggcgc cgccgcaggg acccgccacc cacctcccgg ggccgcgcag cggcctctcg	120
	tctactgccca cc	132
	<210> 24	
	<211> 365	
	<212> DNA	
	<213> 智人 (Homo sapiens)	
[0004]	<400> 24	
	acggaaatga aaggagcact tccgggttcg gcaataacct ggagccggcg gcgtaggttg	60
	gctctttagg gcttcacccc gaagctccac cttcgctccc gtctttctgg aaacaccgct	120
	ttgatctcgg cgggtcgagg caggtacctc ccggtctctg cgggtgccct ggatccagtc	180
	ggctgcacca ggcgagcgag acccttcctt ggtggaggct cagagttccg gcaggggtgca	240
	tccggcctgt gtgtggcgcg aggcagggaa gccggtaccc gggtcctggc cccagcgctg	300
	acgtttttct tcccccttct tctctcttcg cgggttcggc gtcgcagacg ctagtgtgag	360
	cccc	365
	<210> 25	
	<211> 88	
	<212> DNA	
	<213> 智人 (Homo sapiens)	
	<400> 25	
	ataaggcggg tgcggcactg cagctagcgc agttctcact gagacctgtc accccgactc	60
	aacgtgagac gcaccgccc gactcacc	88
	<210> 26	
	<211> 254	
	<212> DNA	
	<213> 智人 (Homo sapiens)	
	<400> 26	
	cgggagcaag agagaaggag gcccagacag tgagggcagg agggagagaa gagacgcaga	60
	aggagagcga gcgagagaga aagggtcttg gattggaggg gagagcaagg gagggaggaa	120
	ggcggtgaga gaggcggggg cctcgggagg gtgaaaggag ggaggagaag ggcggggcac	180
	ggaggcccga gcgagggaca agactccgac tccagctctg acttttttcg cggctctcgg	240
	cttccactgc agcc	254
	<210> 27	
	<211> 57	
	<212> DNA	
	<213> 智人 (Homo sapiens)	

161

	ggacccggag caggagcggc ggcagcagcg actgggggcg gcggcggcgc gttggaggcg	120
	gcc	123
	<210> 34	
	<211> 20	
	<212> DNA	
	<213> 智人 (Homo sapiens)	
	<400> 34	
	ccggcgggtga caacggcaac	20
	<210> 35	
	<211> 21	
	<212> DNA	
	<213> 智人 (Homo sapiens)	
	<400> 35	
	agtctgggac gcgccgccgc c	21
	<210> 36	
	<211> 27	
	<212> DNA	
	<213> 智人 (Homo sapiens)	
	<400> 36	
	ctgtgggtct gtaggttaag ggagaag	27
	<210> 37	
	<211> 30	
	<212> DNA	
	<213> 智人 (Homo sapiens)	
	<400> 37	
	gcggtggcgc gagttggact gtgaagaaac	30
	<210> 38	
	<211> 142	
[0006]	<212> DNA	
	<213> 智人 (Homo sapiens)	
	<400> 38	
	acgcacagcc cccctggggg ccgggggagg ggccaggcta taaaccgccc gttaggggcc	60
	gccatccct cagagcgctg ggatatcggg tggcggctcg ggacggagga cgcgctagt	120
	ttcttctgtg tggcagttca ga	142
	<210> 39	
	<211> 78	
	<212> DNA	
	<213> 智人 (Homo sapiens)	
	<400> 39	
	gagcggttgc gcagtgaagg ctagaccggt ttactggaa ttgctctgga gatcgagggg	60
	tcctagtaca ccgcaatc	78
	<210> 40	
	<211> 117	
	<212> DNA	
	<213> 智人 (Homo sapiens)	
	<400> 40	
	gagagacggc agcggecccc gctccctct cgcgcgcgt tcagcctccc gctccgccgc	60
	gctccagcct cgctctccgc cgcgcgcacc gccgcgcgc ccctcaccag agcagcc	117
	<210> 41	
	<211> 51	
	<212> DNA	
	<213> 智人 (Homo sapiens)	
	<400> 41	
	gtgacagagc gctttacgac agttgctttg aggcagtacc ggaggagaaa g	51
	<210> 42	
	<211> 30	
	<212> DNA	

	<213> 智人 (Homo sapiens)	
	<400> 42	
	ctagatctgt gggcggggcg cggcctgtgg	30
	<210> 43	
	<211> 136	
	<212> DNA	
	<213> 智人 (Homo sapiens)	
	<400> 43	
	gatgatttga gggacaagaa ttcagtgtccc gggggccgaa aggcagcaga aggcgggcac	60
	caaaggatag gcacccggaa ggtggactcc gaggaggaga gaggacaggg gtctctcacc	120
	ccagctcctg gtcacc	136
	<210> 44	
	<211> 88	
	<212> DNA	
	<213> 智人 (Homo sapiens)	
	<400> 44	
	ctacagatcc gccatattgt ctactgaaag ctgccgtga agctgccgcc gttgcctccg	60
	ccgccaagag tgagcgagcg gacccgcg	88
	<210> 45	
	<211> 65	
	<212> DNA	
	<213> 智人 (Homo sapiens)	
	<400> 45	
	ccgtcaactt agcgagcgca acaggctgcc gctgaggagc tggagctggt ggggactggg	60
	ccgca	65
	<210> 46	
	<211> 243	
	<212> DNA	
[0007]	<213> 智人 (Homo sapiens)	
	<400> 46	
	gagggaggga gctgcttccc cgcttccgc gccaccagtc cgaccctcgg tcccgccgtg	60
	tgagcagctg gtggagtga gctcagcgcg gacgccggag ctgcggccgc cccctctgca	120
	ggtgcctgtg aggaggcgcc cgggcccga cgccttccg agatccccag tcgcgcactc	180
	gctccccgcg ctccctgagg gccgcccggc cggaggaggc cgtcgcgggg ctcggcgttc	240
	agc	243
	<210> 47	
	<211> 335	
	<212> DNA	
	<213> 智人 (Homo sapiens)	
	<400> 47	
	gaggaaggag ggggaaaggc accaaccagc agccgctcca gccctgccga agtttcaactt	60
	tttgtctgtg ggttcgttcc cgggtcccgc gcgagcttcc ccgggataag tagctttagc	120
	gatcgggaga cagccgggcg ctgcaagtgg gaactttgag gctcagagac agagcagaag	180
	acagaacctg gtcttctgat tccctgtgtt ctgctttttt cattgttcca ctggacgctc	240
	atcagaggga agatcttttt cctcaattgc attgcagttg gctgaacctt gtgtcaacca	300
	tccattccag gagccagcta tctgagattc caaca	335
	<210> 48	
	<211> 203	
	<212> DNA	
	<213> 智人 (Homo sapiens)	
	<400> 48	
	gggagaagga ggaggccggg ggaaggagga gacaggagga ggagggacca cggggtggag	60
	gggagataga ccagcccag agctctgagt ggtttcctgt tgccctgtctc taaacccctc	120
	cacattcccg cggtccttca gactgcccgg agagcgcgct ctgcctgccg cctgcctgcc	180
	tgccactgag ggttcccagc acc	203
	<210> 49	
	<211> 268	

	<212> DNA	
	<213> 智人 (Homo sapiens)	
	<400> 49	
	cttccgggag cccgcaagcg gcttcggggt gctcgcgcgc cgacctggac gcagagaagc	60
	cagagacttt cgcttccggc tgccgcaggc gcttcgctgg tgcaggtaag ctccgcacac	120
	tctcggccgg tcccgagtcc gactccctca aggggtgacgc gagctctgcc ctttaaccgg	180
	aaacgtctcc ctgctcacc cacccccgcg cagacgcagt gctgagcaca cagctaccgg	240
	acaaagagtg acgcccggag ctggagtt	268
	<210> 50	
	<211> 155	
	<212> DNA	
	<213> 智人 (Homo sapiens)	
	<400> 50	
	gactccggag cccgagcccc gggcgggtgg acgcggactc gaacgcagtt gcttcgggac	60
	ccaggacccc ctcgggcccc acccgccagg aaagactgag gccgcggcct gccccggccg	120
	gctccctgcg ccgcccgcgc ctcccgggac agaag	155
	<210> 51	
	<211> 167	
	<212> DNA	
	<213> 智人 (Homo sapiens)	
	<400> 51	
	acagaccact gcctgcaggt tggcgccacc acccccactc tccccgctgc tcgcgggagc	60
	cagagggccc tgcggtctct ggtggtcttg ccagccctc gtcatcccag ggccctccgc	120
	gcctgtgagg actccctcag gtcggccacg ggacctgacg caacagg	167
	<210> 52	
	<211> 209	
	<212> DNA	
	<213> 智人 (Homo sapiens)	
[0008]	<400> 52	
	ggccggagcg ctccgcgtta cgggatgaat taacggcggg ttccgcacgg aggttgtgac	60
	ccctacggag cccagcttg ccacgcacc ccactcggcg tcgcgcggcg tgccctgctt	120
	gtcacagggt ggaggctgga actatcaggc tgaaaaacag agtgggtact ctcttctggg	180
	aagctggcaa caaatgatg atgtgatat	209
	<210> 53	
	<211> 94	
	<212> DNA	
	<213> 智人 (Homo sapiens)	
	<400> 53	
	gcggtggacg cgccggcttc gagcatccct agccgggcag gtgggaggca cggggttgcg	60
	gatcccgcg ccgcggttcg gacccgccg cgac	94
	<210> 54	
	<211> 76	
	<212> DNA	
	<213> 智人 (Homo sapiens)	
	<400> 54	
	ggaagcgctc gcgcaggaga ccccgggtga cggggcccgg cgccgctaac tggagcgaac	60
	cccagcgctc gccgac	76
	<210> 55	
	<211> 261	
	<212> DNA	
	<213> 智人 (Homo sapiens)	
	<400> 55	
	cttctctccg ctttctgggc ggagcccgcg cgggatccgg gtggctgcag gctgctggct	60
	tctgcggctg cggggtcggg gtcgcggcca gggccaagcc gcagcgagtt cacaggcgga	120
	accctgcag gcggcgcccc ctacgcgagg tcacccctgg gaaggagcgc agcccacccg	180
	gccctccgc atccgagcag gacgccgtc tctctccct gaggatttca ggtctccctg	240
	tcccaggagg cttgtgcaa g	261



	<210>	56	
	<211>	241	
	<212>	DNA	
	<213>	智人 (Homo sapiens)	
	<400>	56	
		gcctagtctc catataaaag cggcgccgcc tccccgccct ctctcactcc ccgtcctct	60
		ccgccgcgca ctctccgcgg cgctgggaga gggcggaggg ggaggcggcg cgcggcgcca	120
		gaggaggggg gacgcagggg gcggagcgga gacagtacct tcggagataa tcctttctcc	180
		tgccgcagtg gagaggagcg gccggagcga gacacttcgc cgaggcacag cagccggcag	240
		g	241
	<210>	57	
	<211>	124	
	<212>	DNA	
	<213>	智人 (Homo sapiens)	
	<400>	57	
		gcacttcgg cgcgaaaccgc tacttccggg gcgaaccgcc tcggccgttc cctcgcgag	60
		cttactgagc gcggccgccg agcccagctc cgccgccgag cgctgtgcc ggcacggcta	120
		cacc	124
	<210>	58	
	<211>	33	
	<212>	DNA	
	<213>	智人 (Homo sapiens)	
	<400>	58	
		gtctcctggg ggggtgcaag gagccgaggc gag	33
	<210>	59	
	<211>	71	
	<212>	DNA	
	<213>	智人 (Homo sapiens)	
[0009]	<400>	59	
		ggcgggggttc ggccgggagc gcaggggcgg cagtgcgcgc ctagcagtgt cccagccggg	60
		ttcgtgtcgc c	71
	<210>	60	
	<211>	245	
	<212>	DNA	
	<213>	智人 (Homo sapiens)	
	<400>	60	
		accacagagc cgcgagaaga ggacagagga gactgagcaa aggggggttg gctccaggcg	60
		accctagcc caattctgcc cctccatccc aaggggcaga gaaattgtct ttctttgctg	120
		actcctacga ggaaaaaaaa aaaaaaaaaa aaaacaatta aagggaaga taaacggaga	180
		cggaggaaag gtggcagcca gattacttag agaggcacag aggagagaga tcggggtgag	240
		tcgcc	245
	<210>	61	
	<211>	198	
	<212>	DNA	
	<213>	智人 (Homo sapiens)	
	<400>	61	
		gacctagaga ggtcccagga cacgccactg tcccgccttc cccattgccc gccccactgg	60
		ccagtcccca cgcccacaca cccaagctg ccccatctgg cgctgattat cctgtgtctg	120
		ccgccaccgc tgtgtgtgt ctgcaaaatt cagctgtctc ctctgtcttg aggaccccag	180
		cgcctttccc ccggggcc	198
	<210>	62	
	<211>	9	
	<212>	DNA	
	<213>	智人 (Homo sapiens)	
	<400>	62	
		ggcccgaag	9
	<210>	63	

	<211> 149	
	<212> DNA	
	<213> 智人 (Homo sapiens)	
	<400> 63	
	gagcaaacag agcagcagaa aaggcagttc ctcttctcca gtgccctcct tccctgtctc	60
	tgcctctccc tcccttctctc aggcattcaga gcggagactt caggagacc agagcccagc	120
	ttgccaggca ctgagctaga agccctgcc	149
	<210> 64	
	<211> 116	
	<212> DNA	
	<213> 智人 (Homo sapiens)	
	<400> 64	
	attcccatta gccccgcccc ttgggctgg aaccggaggt gtcgtcttct ggacctcaag	60
	gttcccctta acacagagcg ccccgagtc ttgcggaaa gcgttcgggg taggcg	116
	<210> 65	
	<211> 80	
	<212> DNA	
	<213> 智人 (Homo sapiens)	
	<400> 65	
	ttagccaacg gggcaagggg ggcgggaagg aggtggggtt tctcccagcc aatcgacggg	60
	cgcgcctcg ttgccgtct	80
	<210> 66	
	<211> 88	
	<212> DNA	
	<213> 智人 (Homo sapiens)	
	<400> 66	
	ctacagatcc gccatattgt ctactgaaag ctgccgtga agctgccgce gttgcctccg	60
	ccgccaagag tgagcgagcg gacccgcg	88
[0010]	<210> 67	
	<211> 157	
	<212> DNA	
	<213> 智人 (Homo sapiens)	
	<400> 67	
	gtttttaccc aggcctcggc gcctaggcgc ttgcgcgagg ctgatcttct ttcaagtgtg	60
	agctgcggct gagcccagcg ctgcaggcgc gaggcagcca ggagggcccg tgcggcgcgg	120
	ggagccagcg agcgcgcctt cggcattggc cgccgcg	157
	<210> 68	
	<211> 133	
	<212> DNA	
	<213> 智人 (Homo sapiens)	
	<400> 68	
	tgagttgctg ccttgccag agtccggagc agccgccgcc cgaccacgcc gagtcagtt	60
	cgctgtccgc gccggctccc accccggccc gaccccgacc cggcccggtc agggccata	120
	ctcagtagcc acg	133
	<210> 69	
	<211> 90	
	<212> DNA	
	<213> 智人 (Homo sapiens)	
	<400> 69	
	ggattgagct gaccacaggc cacaccagac tcctctctgc tcctgaggaa gacagggcag	60
	cccggcgcca cccgctcggc cctcacgaag	90
	<210> 70	
	<211> 88	
	<212> DNA	
	<213> 智人 (Homo sapiens)	
	<400> 70	
	ggacagccgg gcaggcgggg ctgggcgcgg gcggcggcgg cccggaggag aacgggcgga	60

	gggcgcgggc cgaccgggcg caccgacc	88
	<210> 71	
	<211> 147	
	<212> DNA	
	<213> 智人 (Homo sapiens)	
	<400> 71	
	acgcgtgcgc ggtgggcgga gcgcggctct cctaccttct cgggcagccc agtctttgcc	60
	atccttgccc agccggtgtg gtgcttgtgt gtcacagcct ttagccggg agtcgctgcc	120
	gagtgggcgc tcagttttcg ggtcgtc	147
	<210> 72	
	<211> 153	
	<212> DNA	
	<213> 智人 (Homo sapiens)	
	<400> 72	
	cattggctcc gaggacttag gacctgttag cttggttggg cgactggcaa atccttcctt	60
	ccccgggta gaagtcagg gtgagaaatt ggttcgaac tcaaaggaac ccagtgccg	120
	gccacagccg ggtcacgtgg ccggcggccc ccc	153
	<210> 73	
	<211> 232	
	<212> DNA	
	<213> 小家鼠 (Mus musculus)	
	<400> 73	
	ctctctccct tttctttttt tccgccctag ctggatctgt gttggaggag aggaaaagag	60
	acagaggatt gcatctgttt tgaaatctcc taatagcaag accagttaag ggattgtacc	120
	tttttctac aaatataaat atatatatat tttaaaccaa gtcttttttt ttccggctat	180
	ctttgtttaa aagctgtcct cttgagatta ctctctccc ccccccggag ag	232
	<210> 74	
	<211> 232	
[0011]	<212> DNA	
	<213> 小家鼠 (Mus musculus)	
	<400> 74	
	ttttctctgc ccggttttga tctctgttta acaacagtaa cgtcacacgg actacagggg	60
	agttttgttg aagttgcaaa gtctgcagc ctccagaggg ctgtcggcgc agtagcagag	120
	agctacagac tccgcgcgt ccggagaccg gcagtacagc gcgaggcagc gcgcgtcagc	180
	agccgccacc ggagcccaac cgagaccaca gccctcccca gacggccgcg cc	232
	<210> 75	
	<211> 147	
	<212> DNA	
	<213> 小家鼠 (Mus musculus)	
	<400> 75	
	cgtgtccca ggacagaggt gactcaagag tttagtccca ggccagccag agaactgaca	60
	gaacatacca aaggttcttg gtacctccc gtgtcagag cagaccctaa aggaagccaa	120
	gtcttctctg agacaggcag atccagg	147
	<210> 76	
	<211> 122	
	<212> DNA	
	<213> 小家鼠 (Mus musculus)	
	<400> 76	
	gagcgcgcgc cccgccgagc gcccgccgc cgggggcctg agcgtgggg cgcgtgcgcg	60
	agcgtgcag caccggccgc gggagcagg agtattatgg gctgtgggtg ccgtgagca	120
	ag	122
	<210> 77	
	<211> 61	
	<212> DNA	
	<213> 小家鼠 (Mus musculus)	
	<400> 77	
	agttccgtgg agactgcgag ttccgtctgc tgtgtgagga ctgccccac caccgtgac	60

	g	61
	<210> 78	
	<211> 202	
	<212> DNA	
	<213> 小家鼠 (Mus musculus)	
	<400> 78	
	gcagactcct cccccgcctg gaaacctcgc ccctagtact gggaatgact ctgggcgtgc	60
	ggcgtagttt gcagccggga caccgtgagg cttgcgagcc cagattcgca gccgagacac	120
	cgtggggccc gcgatccagt gtttcagcgc ggaacatttc ggggtctgtg atccgagtga	180
	ggacgcaacg cagaagctaa gg	202
	<210> 79	
	<211> 91	
	<212> DNA	
	<213> 小家鼠 (Mus musculus)	
	<400> 79	
	ggccggggcca gccgctaaat tccgagatca agcttcaga tcctgggagt agatcaagcg	60
	tgtccaggtt ggagagctgc cctgtcagac t	91
	<210> 80	
	<211> 54	
	<212> DNA	
	<213> 小家鼠 (Mus musculus)	
	<400> 80	
	gagcgccaac cgctagaggc cgatccggag cgtgctgccc ggccaccacc cgcc	54
	<210> 81	
	<211> 63	
	<212> DNA	
	<213> 小家鼠 (Mus musculus)	
	<400> 81	
[0012]	ccacagggcc agttccaggg gttcatccac aagagagaaa aacatagact cacggctgcc	60
	aac	63
	<210> 82	
	<211> 48	
	<212> DNA	
	<213> 小家鼠 (Mus musculus)	
	<400> 82	
	gagtgcgaca gctaggccag gcgacagcgc tgcctaccag agcgcagc	48
	<210> 83	
	<211> 93	
	<212> DNA	
	<213> 小家鼠 (Mus musculus)	
	<400> 83	
	agaaggaaaa ggggtggagc taagcactca ctgcggtttt gcgctgcgtc tgcagaggac	60
	aaggaaagct ctgcagggtc gtcagctgcc aaa	93
	<210> 84	
	<211> 159	
	<212> DNA	
	<213> 小家鼠 (Mus musculus)	
	<400> 84	
	gggcgcttgg cttgcaagga ccctgagctg cggcattgaa gcacacccaa cccaactcga	60
	ctgaagtcag cctcactgaa ccggtctga gaattcttc tctctgggct tgccagggtc	120
	ctccgaacct agctagcate ctcttcaatt ccaactaga	159
	<210> 85	
	<211> 55	
	<212> DNA	
	<213> 小家鼠 (Mus musculus)	
	<400> 85	
	cgcgagggt gtaaaagccg cgcaggatgg aagtcagac actgaatccg cggtc	55

	<210>	86	
	<211>	105	
	<212>	DNA	
	<213>	小家鼠 (Mus musculus)	
	<400>	86	
		gagtctcgcg ctgtggttcg tcggcgcacc gctgatccgc tccacgcctt gcgctctccg	60
		ctctcagcca aagcccggca gccccggcca cgcagctccg caacc	105
	<210>	87	
	<211>	108	
	<212>	DNA	
	<213>	小家鼠 (Mus musculus)	
	<400>	87	
		ctctctgcgg gcggaagtag gaggagtaga gtttaaaaac agtactcttt ttccggttcg	60
		ggacgtagtt gaagcaacga caagccggat aaccgctctt gagacagg	108
	<210>	88	
	<211>	176	
	<212>	DNA	
	<213>	小家鼠 (Mus musculus)	
	<400>	88	
		gtacttagct ggctctgggg cagcagcggg aagggactga agaggggtgc tgaggtgaaa	60
		gatttgaaga cccggaaaga tcagctagat ttcggaaacc agaaacaccc tcggggagac	120
		ctggggctcc cggcgcgcga cgatttcctc gcactagctg ggagagcggt gatccc	176
	<210>	89	
	<211>	275	
	<212>	DNA	
	<213>	小家鼠 (Mus musculus)	
	<400>	89	
[0013]		tgccctgagcg agagaggaga gcgagctgag gagagccggg cagttcggag ggaaggaccg	60
		gtgcgagtcg ggcggccctt gaggtccgc tcgccacct tccactcttc tctctctctc	120
		tccctctctc tctttgccat ttctttcttc tcccaaattc cccattcagc caaaggaagg	180
		aggtaaggga agcactcccc gacccccccg cacctccaaa aaataataat aataaaaaaa	240
		atttacagtc gggaccgagt ggtggccggc tggtct	275
	<210>	90	
	<211>	146	
	<212>	DNA	
	<213>	小家鼠 (Mus musculus)	
	<400>	90	
		cggaactggg gctattaggg agcctgcagg tcttctctgga aagtcgcaag cctgtgttag	60
		gaagtcgccc tcgatcgcg ccgctgggg tttttctgtc cctttgtcct cgtttatgta	120
		cgaagctgcc agcgggccat agaaac	146
	<210>	91	
	<211>	173	
	<212>	DNA	
	<213>	小家鼠 (Mus musculus)	
	<400>	91	
		gaaaagactg tctactcctg cgtctcggtg gcgtctttct cccgcgcctg cacgaactga	60
		ggttttgcgt ggcgggcgcg gcaccggcag cggcagcgtc tctcacttga acgccgcgag	120
		cggcagcttc tcgtctgtgt cctgacctcg gacctacga gcagagcggc gcg	173
	<210>	92	
	<211>	77	
	<212>	DNA	
	<213>	小家鼠 (Mus musculus)	
	<400>	92	
		gcagcggcgt cccgtccagt tcgctgccg cgctcctcgc ttgtcgtctg cctccagagc	60
		ccagacagag aagagcc	77
	<210>	93	

	<211>	184	
	<212>	DNA	
	<213>	小家鼠 (Mus musculus)	
	<400>	93	
		cccgccgctg cccgagtcgg gtgggtcggg ggctataaag ctgcggaggg cgggaggcac	60
		agtctgcggg ctcccttgctt tgctcgttct tcttccaggg tctgatcccc tttggtggct	120
		gaaggaggaa ccggtgaccg acctggagac cacagttctc tgtccttcac acagctcttt	180
		cacc	184
	<210>	94	
	<211>	335	
	<212>	DNA	
	<213>	小家鼠 (Mus musculus)	
	<400>	94	
		taggcggctc ggtgctaagg gatgagttag acgtagctcg gcctctcccc agcaagttct	60
		cgccggcact ggttcagatt cagaaaggga gccaggagcg aggctcagca gaccccgcg	120
		ggtagaactc cggggctgtg aagcgtccc gcgtgtccga ctcttgagtg caccgcagct	180
		acagaagcgt cagaagtcta gttcgccagg ggtgccagtc ggctccagag tctctggtea	240
		ctcaagtcca gcgcagcgac cctcacagac gcgctctgcc ctgggaccgc ggcggaacaa	300
		gatattttaa atcattgcgg ttcccagcga cagaa	335
	<210>	95	
	<211>	300	
	<212>	DNA	
	<213>	小家鼠 (Mus musculus)	
	<400>	95	
		gcccgcctct cgccgccctc ctccctcgcc cgcccgccgc tctgagcagc ccgagaggaa	60
		acaaagtgtc gcgggcggga gactcgcgtc cgccgcgcac agccgtctgc gtctcccggc	120
		tgccctggcc tcttcccgcg cgcgtctgcg agtgtgcgtg tacactcaca aaggcgctct	180
		ggcgggcgat cgccgccctc ccgcttcgtc ctttgtgcgg tagccccgcc gccaccgcca	240
		gcccaggttc gctcgatcct caccggcctg tggtttgtac ctccggtccc gccgcccgcc	300
[0014]	<210>	96	
	<211>	45	
	<212>	DNA	
	<213>	小家鼠 (Mus musculus)	
	<400>	96	
		ccggttttct tgcgcgcgcg ggtctctccg gcctaggaga ccgcg	45
	<210>	97	
	<211>	109	
	<212>	DNA	
	<213>	小家鼠 (Mus musculus)	
	<400>	97	
		gcttcggcag gggeacttcc ggcgcgaccg gcctagactg tcacctgtt agccgccgct	60
		gagcctccgc taactaccgc gcgctgctg ctgcctccc ggccacgcc	109
	<210>	98	
	<211>	131	
	<212>	DNA	
	<213>	小家鼠 (Mus musculus)	
	<400>	98	
		aacaaagctg ttccctcgtgg gcggagtgtg gctctccagc cgcaggagtc ttgtgcgcg	60
		ggtgggtaga agctgggatc agagaagcgg cgaggcagac agtcttcgga catagactga	120
		ctctgcttgt c	131
	<210>	99	
	<211>	235	
	<212>	DNA	
	<213>	小家鼠 (Mus musculus)	
	<400>	99	
		tgtgccggtt gagtttgaat tgggtggcgg ttaaccgagg agccgccgct cccttagttg	60
		gagctgtgag ggttcctcag actgtgtttt gggacctgca ggtaggtttc ggcagagttc	120

	tggaaccta gactccaacg actgaacttt ctcagctctc cgaccgctca caccctctcc	180
	ccgtctcagt cgcggagccg gctgcttgge ccctcgctcg acgcagccag gcgcc	235
	<210> 100	
	<211> 280	
	<212> DNA	
	<213> 小家鼠 (Mus musculus)	
	<400> 100	
	ggtgccctgg aagctgacgc gcagcactgg ccgtggccgt ggggcccgcg gagggcggcg	60
	cgcggctgat gaaaccggct cggattccgc ccgcgtgcgc catccctca gctagcaggt	120
	gtgagaggct ttctaccgcg ggtctccaca cagctcaaca tcttgccgc tcctccgagc	180
	ctgaagctac cgtggactct gctgtggcgt cttggccccc aggtgcggat cctccccagt	240
	gagaagtccc acgagtctta cagcagattc gctcagcaca	280
	<210> 101	
	<211> 222	
	<212> DNA	
	<213> 小家鼠 (Mus musculus)	
	<400> 101	
	cggagacggc ctgcaggacc tgctctctca gccctcagcc gaggcctacg ccgagccgag	60
	tgccagccg acgaccggga ggagccgcag ccttcaactc tgaggtaactg tgatccgcgc	120
	tgcccgcgg gccgcccag tccgtgctg cggaacctcc ttccctcgcg cctctctgcg	180
	tcgccagcgc cttccctgtg agcctgcgtc accgcggccg cc	222
	<210> 102	
	<211> 108	
	<212> DNA	
	<213> 小家鼠 (Mus musculus)	
	<400> 102	
	agcccaggcc accgcagcag cagcagcaac agccgcagca acggtagcag tagtctgcat	60
	cgcagtcctt ttctccttct ccagcgcgt cctcagtcgc cggtcacc	108
[0015]	<210> 103	
	<211> 124	
	<212> DNA	
	<213> 小家鼠 (Mus musculus)	
	<400> 103	
	ctcttttact cgaacttcac gggggcaatt ttctcgcgca cgcgttctaa gacgagctgg	60
	ttttgaccc cggtactaca ccgatacaga gttagtggct agtcctcctg tgctcctgt	120
	agcg	124
	<210> 104	
	<211> 19	
	<212> DNA	
	<213> 小家鼠 (Mus musculus)	
	<400> 104	
	ggccgaagaa gggaaggct	19
	<210> 105	
	<211> 318	
	<212> DNA	
	<213> 小家鼠 (Mus musculus)	
	<400> 105	
	ggagattcga aagcgtccgg gtcgcggggt aaaccggttc tccgtgtgcg agcgccctagt	60
	ggcgtaggct gcggtttgc ggggaactgc gggggctgca gtggtccacg gggctgatcg	120
	ggttccgttg ggcggatcca cgtgcccgt atccgcctgg aaggagaggt gcaggagtac	180
	ccccgacctt ggctgcgtgc tgactcgctt cttctgccc gccagagctt gcactccccg	240
	gggatctgcc tctgcatctc ttgccttcgc tgttgtttcc ctctctgtcc agctccccctc	300
	ccgctctcgc cctggaga	318
	<210> 106	
	<211> 198	
	<212> DNA	
	<213> 小家鼠 (Mus musculus)	

	<400> 106	
	gtcacttcgc gcaggcgac ttcgaaagc gcgcttttcg gtggtgcgg tcccggcagg	60
	gagcacagct catcctgtgc tgataacatc gagaagtgtt cagtgcctgg taaagtacat	120
	agaccttgc tcaattggaa ctccgcttg atttctgccg ttggtcataa tcagcagagt	180
	tctctctaaa cctttgac	198
	<210> 107	
	<211> 175	
	<212> DNA	
	<213> 小家鼠 (Mus musculus)	
	<400> 107	
	agtcgtgctt cagctcacta cgcttagggc tctggcttgc cgctcccgcc tgctctccag	60
	cgccccgcgc agcgagccac gtggaccaac tccggcgccg ggtgttcgt tggttccagt	120
	ggggctcgcc gcgcctcccg cgtctgcgtg cttgcctgt ctcagctccg acccg	175
	<210> 108	
	<211> 219	
	<212> DNA	
	<213> 小家鼠 (Mus musculus)	
	<400> 108	
	gtgtcgtccg gggcgctgaa attcaaattt tgaacggccg tggtagccta ccgactccgt	60
	gggtgcggag ggcagagccg actgggtgtg agagcgcccg ccgcctcgac tgcagtctc	120
	ctccaggagc tgcgccgagc ctgcactcac ttctttctc ttcttgagtt tgaaccgtcg	180
	gaccaccgt ctagccgtcc actggtgagg cctggggcg	219
	<210> 109	
	<211> 217	
	<212> DNA	
	<213> 小家鼠 (Mus musculus)	
	<400> 109	
[0016]	gggaattcca cgccacctcc gcgcgctctg cgctctggga tccggagcga ccaggacctg	60
	gtgagacct cagctcccct ccacctccc cgaggtccag cacaccagaa ctggaacctg	120
	agcagcccag aagccagggt ggcaccactg tgtctctcct gtctgaagac ccgatatatt	180
	tctcagactt ggcgacacgt tctttaaaa gatcagc	217
	<210> 110	
	<211> 87	
	<212> DNA	
	<213> 小家鼠 (Mus musculus)	
	<400> 110	
	aaagaccacg cccagtgac gtcacgcggc ggttaccgcg cttggcgctc gcgatttaaa	60
	acttactccg gagacgtgga gagcaag	87
	<210> 111	
	<211> 33	
	<212> DNA	
	<213> 小家鼠 (Mus musculus)	
	<400> 111	
	agccaaagca acacctcgca ctctgcccc agc	33
	<210> 112	
	<211> 69	
	<212> DNA	
	<213> 小家鼠 (Mus musculus)	
	<400> 112	
	gcctcgcgag agcacgtgga gagcgcgcca aattctcccc ggagcctgag ggagctttgg	60
	agcgtcgca	69
	<210> 113	
	<211> 237	
	<212> DNA	
	<213> 小家鼠 (Mus musculus)	
	<400> 113	
	aattgggggtg gaagtctccc gcagcagcct gtgcacacta ataaaaacgc cctgagtttt	60



	cgcgggaaaa	aaagtcctag	cagctggaag	gaacccgcgc	tctagtgtctc	acttgggtct	120
	tcagccactc	acgcggctcc	cttctgggtc	acccagccgc	agagccctag	cctagaacca	180
	ggcgttccaa	gggagaggag	agtgcggatc	gccgccgtga	gagagtacat	cggcatc	237
	<210>	114					
	<211>	232					
	<212>	DNA					
	<213>	小家鼠 (Mus musculus)					
	<400>	114					
	ttttaaaagc	tctgtgtctc	aagttaaaaa	acgcttttac	gaggtatcag	cacttttctt	60
	tcattggggg	aaaggcgtga	gggaagtacc	caacagcagc	agactttgaa	actttaaaca	120
	gacaggctctg	agagccccgaa	ctctcctttt	cctttgactt	cagcctccaa	ggagttccac	180
	cactttggcg	tgccggcttc	actttcatta	agtgaagag	aggtgccag	ac	232
	<210>	115					
	<211>	85					
	<212>	DNA					
	<213>	小家鼠 (Mus musculus)					
	<400>	115					
	cggtgggcta	ggagagggtg	ttgatcttcg	ggtccgggta	tcgagcaggg	aggatctagc	60
	gggcagcgga	aactggacca	agcac				85
	<210>	116					
	<211>	133					
	<212>	DNA					
	<213>	小家鼠 (Mus musculus)					
	<400>	116					
	tgacgttgta	gggaggtctg	cgcgcgggat	ttaaactgca	gcggtttagg	cgttggttaac	60
	acagcgcagt	attagcagag	tcgtggttcc	caagcttctt	tcatttgtgt	tgctgttgt	120
	tgttctctgag	tcc					133
	<210>	117					
[0017]	<211>	148					
	<212>	DNA					
	<213>	小家鼠 (Mus musculus)					
	<400>	117					
	cactggagga	ggagettgca	gctctccage	tctggcttat	acagttcaca	gagaagccaa	60
	gggacaccgg	gacaccgtta	ttttaggact	gtaacctgtt	cagagagacc	ctggccactg	120
	ctctccgtgg	ttttctccaa	ttgtgtgg				148
	<210>	118					
	<211>	90					
	<212>	DNA					
	<213>	小家鼠 (Mus musculus)					
	<400>	118					
	ttcatttccg	gccaccgagg	agcttcgcgg	gtcgagacgc	aggcgaggcg	ccggactgcg	60
	aaacaaaggg	ggacgccaac	agccgtagtc				90
	<210>	119					
	<211>	137					
	<212>	DNA					
	<213>	小家鼠 (Mus musculus)					
	<400>	119					
	ttatatgtcc	agctggaagc	ctggctgtcc	ctagaaaagc	tggaagcctg	actgccccctc	60
	aaaagacctg	ctcttttagga	gagctagata	ttgtttactg	aagacaagtg	tgaaaactgg	120
	gaacctcagt	ctccatc					137
	<210>	120					
	<211>	41					
	<212>	DNA					
	<213>	小家鼠 (Mus musculus)					
	<400>	120					
	aaatggtcgt	cgagggaagg	cgcctcatcg	cgccgtgaat	t		41
	<210>	121					

	<211> 19	
	<212> DNA	
	<213> 小家鼠 (Mus musculus)	
	<400> 121	
	aagacaggca gaaggaccc	19
	<210> 122	
	<211> 184	
	<212> DNA	
	<213> 小家鼠 (Mus musculus)	
	<400> 122	
	aagggaggga aaagtccct gcgcggagag cggggcaggc gcacgtcttt acggcggccg	60
	cagcggcagg gcggggccgg tggggcgggc ggaggaagac ccctgattgc cacctcgcct	120
	ccctcagtgt ctctcttaga ccttggtcac gtaccggggt ccggaggact tgtgtacagc	180
	ggca	184
	<210> 123	
	<211> 79	
	<212> DNA	
	<213> 小家鼠 (Mus musculus)	
	<400> 123	
	agcgcgcac cccagtgcct tgagctgcag agagctcgga ggagcgcggg agcagcgaca	60
	ccggagtgga caaagcaag	79
	<210> 124	
	<211> 25	
	<212> DNA	
	<213> 小家鼠 (Mus musculus)	
	<400> 124	
	agcctgctgt ggtcaatccc tgaag	25
[0018]	<210> 125	
	<211> 26	
	<212> DNA	
	<213> 小家鼠 (Mus musculus)	
	<400> 125	
	gtagagagga ctttggatca ttctaa	26
	<210> 126	
	<211> 90	
	<212> DNA	
	<213> 小家鼠 (Mus musculus)	
	<400> 126	
	aattttgttg cctgggtgcag caggtagcag tgaaacctct gaggcgtttg gtgctccggt	60
	aaccaccacg gctgtagagc gagtgttgcc	90
	<210> 127	
	<211> 87	
	<212> DNA	
	<213> 小家鼠 (Mus musculus)	
	<400> 127	
	cagaaactcc gatcgccctt cccaaccgga aagtgcgcgt cggccgagcc tggagggacc	60
	cagctgaagc ctggcctggg agccagg	87
	<210> 128	
	<211> 275	
	<212> DNA	
	<213> 小家鼠 (Mus musculus)	
	<400> 128	
	agatcacgtg acgagagcgc gggctttgga aggcggcgaa gctgggaacg atactgaaag	60
	aacgggctcg ggactgtcca gagaacgctg tggactccaa accactgcta gctgcttaag	120
	gctcgtgcac tcgagacggg gtgaggtctc gcggaagcgt ctctgaaagc ggcagagccg	180
	cgggaacagc accgggcagc ccgggcttgg gccattcget ctgctccgaa cattcctctt	240

	cgccggtagg aagaagccat tgctcctgtg ctacc	275
	<210> 129	
	<211> 86	
	<212> DNA	
	<213> 小家鼠 (Mus musculus)	
	<400> 129	
	gacttgcgac ttgcgactcg gtgctgtcca gattcagcac gctctgttcc ttcgccttac	60
	aagtccaggc acccagcccc gccgcg	86
	<210> 130	
	<211> 77	
	<212> DNA	
	<213> 小家鼠 (Mus musculus)	
	<400> 130	
	gaaggcattg tcaaggagcc cagagtgtag gaactggcat tgaccgggac gaccggacca	60
	ttgaccggcc gctggat	77
	<210> 131	
	<211> 112	
	<212> DNA	
	<213> 小家鼠 (Mus musculus)	
	<400> 131	
	cggctctgtg ggcagggccc ggacaagggtg gagccctgtc ggcttactgg tcccagtc	60
	atcctcgcta ctccagcaca cgtgacctcc gccgccgtg ctaacctgca cc	112
	<210> 132	
	<211> 319	
	<212> DNA	
	<213> 小家鼠 (Mus musculus)	
	<400> 132	
[0019]	cgctgctcac ttctcagggc cgagtctacc taggtgaaaa ctatctcttt tgaggggaag	60
	cgacacctag gtagtcatt gagagtactg caagagtctc aggccacgc gtattctgtg	120
	gctcccttcg ctttacctgt gctcgcggtc gccctcgttg ccccgaaag gagcctctcg	180
	gctgaaggag ggacgccgag gcgaggcggc gcctctgcgc ttgcgcttat atggtggtc	240
	gaaggagcca aaggcaaagt ctcgggtgac agctgcgagc cccacccttt cccacgtgg	300
	ctgcgtagac ccggcagtg	319
	<210> 133	
	<211> 145	
	<212> DNA	
	<213> 小家鼠 (Mus musculus)	
	<400> 133	
	aacattcctt gagaaagaga gaacaagata agcagtgggtg cacttccctcc ttcatgtctg	60
	ctgagaatgt tccaggctga gtgagaagta aaaattcatt atctctgaag aactcttacc	120
	cagccctgtt gaagaaattc ccaga	145
	<210> 134	
	<211> 151	
	<212> DNA	
	<213> 小家鼠 (Mus musculus)	
	<400> 134	
	ccttcggaga atcctgtgcg cgtgcgttct tgtgggaact gcgcctccca gaaggcaccg	60
	cgcgtcgtcc ttctccctcc ctgaaggccg cctcgttgc ccagtgtgtc ggccgccgc	120
	gaagctagag tccactgact tttccgccac c	151
	<210> 135	
	<211> 65	
	<212> DNA	
	<213> 小家鼠 (Mus musculus)	
	<400> 135	
	gcggagtggg cgggggaaca cctcggccga ggcagtgagg gaccaggtc tccaaggaca	60
	gaaaa	65

	<210>	136	
	<211>	108	
	<212>	DNA	
	<213>	小家鼠 (Mus musculus)	
	<400>	136	
		cccaacctgc agatgaggaa gaggaagcgg gagacagacg gacagacgga cggacacaga	60
		cgctgcgccc aggcctcacc atccatcgca ggaagccccc tgtctcag	108
	<210>	137	
	<211>	103	
	<212>	DNA	
	<213>	人工序列 (Artificial Sequence)	
	<220>		
	<223>	人工核酸构建体	
	<400>	137	
		atcttagctt ctctgtcttc ctccctccct cccttcctct tactctcatt catttcatac	60
		acactggctc acacatctac tctctctctc tatctctctc aga	103
	<210>	138	
	<211>	259	
	<212>	DNA	
	<213>	人工序列 (Artificial Sequence)	
	<220>		
	<223>	人工核酸序列	
	<400>	138	
		atcacagccc ttccccgate ctctccgtgg gagccagcga gcctctctcc ctgatcttac	60
		gtgctcaagg gagctcacac gttcaccaac tcacccttga agtcattctca agaacaaaag	120
		acaactgaaa gaagctgttg tgaaggcaga gcagcatctg ctgaagagac agaaaccagc	180
		cccagaggtg tcacaggaag gcaccagcaa ggacattggt ctttgatttg attcagcagt	240
		cctgtcaagt ataatagtg	259
[0020]	<210>	139	
	<211>	80	
	<212>	DNA	
	<213>	人工序列 (Artificial Sequence)	
	<220>		
	<223>	人工核酸序列	
	<400>	139	
		tagggctgga gctccgggca gtgtgcgagg cgcacgcaca ggagcctgca ctctgcgtcc	60
		cgcacccag cagccgcgcc	80
	<210>	140	
	<211>	101	
	<212>	DNA	
	<213>	人工序列 (Artificial Sequence)	
	<220>		
	<223>	人工核酸序列	
	<400>	140	
		cagtcgtccc tgcgcgtcgt cctcctcgcc ctccaggccg cccgcgcgcg gccggagtcc	60
		gctgtccgcc agctaccgc ttctgccgc cgcgcgtgc c	101
	<210>	141	
	<211>	110	
	<212>	DNA	
	<213>	人工序列 (Artificial Sequence)	
	<220>		
	<223>	人工核酸序列	
	<400>	141	
		taggcggacc tccggaacc gtagattccg ggcggtcgga gccgccggga gctgtagttc	60
		tcccgcggct cagagaagta ggcagagagc ggacctggcg gccgggcagc	110
	<210>	142	
	<211>	291	

	<212> DNA	
	<213> 人工序列 (Artificial Sequence)	
	<220>	
	<223> 人工核酸序列	
	<400> 142	
	accggtagta gctcggctag aaccactagg cgcctggcgg ggtgatctg tcggagcgac	60
	cggcttggcg cctgcctgtc cccagccct ctcagcttga actccttct tcaagtctgg	120
	gccctcgagg cttccagagc ggctccagg ggtgcagtct cagttcccca cgccagccgt	180
	ctccgtctc cgcctcctc gggcctggca ggtggcactg tccggaggcg gagccttggg	240
	cgaggggtgg ttgcggcgga ggacgcaacc gagcgggcct gcggcctcac c	291
	<210> 143	
	<211> 184	
	<212> DNA	
	<213> 人工序列 (Artificial Sequence)	
	<220>	
	<223> 人工核酸序列	
	<400> 143	
	cttttcccg cactagcgca ccgcagcggg tcgcgcgccc taaggagtgg cactttttaa	60
	aagtgcagcc ggagaccagc ctacagccgc ctgcattctgt atccagcgcc aggtcccgcc	120
	agtccagct gcgcgcgccc cccagtcccg cacccttcg gccagcgta agttagccct	180
	cacc	184
	<210> 144	
	<211> 30	
	<212> DNA	
	<213> 人工序列 (Artificial Sequence)	
	<220>	
	<223> 人工核酸序列	
	<400> 144	
[0021]	cttgatctgt gggcggggcg cggcctgtgg	30
	<210> 145	
	<211> 136	
	<212> DNA	
	<213> 人工序列 (Artificial Sequence)	
	<220>	
	<223> 人工核酸序列	
	<400> 145	
	gtagatttga gggacaacaa ttcagtgcc gggggccgaa aggcagcaga aggcgggcac	60
	caaaggatag gcacccgga ggtggactcc gaggaggaga gaggacaggg gtctctcacc	120
	ccagctcctg gtcacc	136
	<210> 146	
	<211> 209	
	<212> DNA	
	<213> 人工序列 (Artificial Sequence)	
	<220>	
	<223> 人工核酸序列	
	<400> 146	
	ggccggacgc ctcccggtta cgggatgaat taacggcggg ttccgcacgg aggttgtgac	60
	ccctacggag cccagcttg cccacgcacc ccactcggcg tcgcgcggcg tgccctgctt	120
	gtcacagtg ggaggctgga actatcagge tgaaaaacag agtgggtact ctcttctggg	180
	aagctggcaa caataggtag tagtgatat	209
	<210> 147	
	<211> 217	
	<212> DNA	
	<213> 人工序列 (Artificial Sequence)	
	<220>	
	<223> 人工核酸序列	
	<400> 147	

	ggcaattcca cgccacctcc gcgcgctctg cgctctggga tccggagcga ccaggacctg	60
	gtgagaccct cagctcccct ccaccttccc cgaggtccag cacaccagaa ctggaacctg	120
	agcagcccag aagccagggt ggcaccactg tgtctctcct gtctgaagac ccgcatattt	180
	tctcagactt ggcgacacgt tcttttaaaa gatcage	217
	<210> 148	
	<211> 87	
	<212> DNA	
	<213> 人工序列 (Artificial Sequence)	
	<220>	
	<223> 人工核酸序列	
	<400> 148	
	aaagaccacg ccccagtcac gtcacgcggc ggttaccgcg cttggcgctc gcgatttaaa	60
	acttactccg gagacgtgga gagcaag	87
	<210> 149	
	<211> 137	
	<212> DNA	
	<213> 人工序列 (Artificial Sequence)	
	<220>	
	<223> 人工核酸序列	
	<400> 149	
	ttattagtcc agctggaagc ctggctgtcc ctagaaaagc tggaagcctg actgccccctc	60
	aaaagacctg ctcttttagga gagctagata ttgtttactg aagacaagtg tgaaaactgg	120
	gaacctcagt ctccatc	137
	<210> 150	
	<211> 41	
	<212> DNA	
	<213> 人工序列 (Artificial Sequence)	
	<220>	
[0022]	<223> 人工核酸构建体	
	<400> 150	
	aataggtcgt cgagggaagg cgcctcatcg cgccgtgaat t	41
	<210> 151	
	<211> 108	
	<212> DNA	
	<213> 人工序列 (Artificial Sequence)	
	<220>	
	<223> 人工核酸序列	
	<400> 151	
	cccaacctgc agtagaggaa gaggaagcgg gagacagacg gacagacgga cggacacaga	60
	cgctgcgccc aggcctcacc atccatcgca ggaagccccc tgtctcag	108
	<210> 152	
	<211> 100	
	<212> DNA	
	<213> 智人 (Homo sapiens)	
	<400> 152	
	cgccgcgcgc gccgcgcgcg gagcctgtcg ccgtcctgtc cccagcctgc ttgtgtcccg	60
	tgaggttgtc aataaacctg cctcgggct gccgcctccc	100
	<210> 153	
	<211> 185	
	<212> DNA	
	<213> 智人 (Homo sapiens)	
	<400> 153	
	ccttgagct cactggctcc tgccctctggg gcttttctga agaaaccac ccactgtgat	60
	caataggag agaaaatcca cattcttggg ctgaacgcgg gcctctgaca ctgcttacac	120
	tgcactctga cctgtagta cagcaataac cgtctaataa agagcctacc cccaactcct	180
	tctgc	185
	<210> 154	

	<211>	250	
	<212>	DNA	
	<213>	智人 (Homo sapiens)	
	<400>	154	
	gtgcccggcc	caccagggac	tgcagctgca ccagcaggtg ccatcacgct tgtgatgtgg 60
	tggccgtcac	gctttcatga	ccactgggcc tgctagtctg tcagggcctt ggcccagggg 120
	tcagcagagc	ttcagaggtg	gccccacctg agccccacc cgggagcagt gtcctgtgct 180
	ttctgcatgc	ttagagcatg	ttcttggaac atggaatttt ataagctgaa taaagttttt 240
	gacttccttt		250
	<210>	155	
	<211>	387	
	<212>	DNA	
	<213>	智人 (Homo sapiens)	
	<400>	155	
	gcggtgctga	cccgggtgtcc	cgaggcccgg ccaggggctg ggcagggatg ccaggtgggc 60
	ccagccgcat	cctgggggtcc	cgccacctg gggaagccgc gcctgccaaag gaggccacca 120
	gagggcagtg	caggcttctg	cctgggcccc aggccttgcc tgtgttgaaa gctctgggga 180
	ctgggccagg	gaagctcctc	ctcagctttg agctgtggct gccacccatg gggctctcct 240
	tccgctctc	aagatcccc	cagcctgacg ggccgcttac catccccctc gccctgcaga 300
	gccagccgcc	aaggttgacc	tcagcttcgg agccacctct ggatgaactg cccccagccc 360
	ccgccccatt	aaagaccg	aagcctg 387
	<210>	156	
	<211>	46	
	<212>	DNA	
	<213>	智人 (Homo sapiens)	
	<400>	156	
	ccaccagcc	acccatgtcc	agaataaaac cctcctctgc ccacac 46
[0023]	<210>	157	
	<211>	151	
	<212>	DNA	
	<213>	智人 (Homo sapiens)	
	<400>	157	
	agccttcagt	cactggggaa	agcaggggaag cagtgtgaac tctttattca ctcccagcct 60
	gtcctgtggc	ctgtcccact	gtgtgcactt gctgttttcc ctgtccacat ccatgtctga 120
	cagacaccac	cattaaagca	tttcatagt g 151
	<210>	158	
	<211>	73	
	<212>	DNA	
	<213>	智人 (Homo sapiens)	
	<400>	158	
	agcagctgcc	tggagcctat	tcactatgtt tattgcaaaa tcctttcgaa ataaacagtt 60
	tccttgacg	gtt	73
	<210>	159	
	<211>	83	
	<212>	DNA	
	<213>	智人 (Homo sapiens)	
	<400>	159	
	gcgccccgcc	tcgctcggt	gccccgcc cttcccgcc cccctcgccg cgcgtcctcc 60
	aaaaaataaa	accttaaccg	cgg 83
	<210>	160	
	<211>	517	
	<212>	DNA	
	<213>	智人 (Homo sapiens)	
	<400>	160	
	gtacagcgg	ctctcatgat	gtgggctcac ctgcgcccc gacctgcag gggccccct 60
	gcttggtct	gtgggagag	tgctcagccg cccagtctg cctgcaagcc cagggcgagt 120
	attggaggag	gggcagccac	gggcagagca ccatgccat ccccgagtct ctggtgtgtc 180

	tgccccctct	ggcaccctct	gggcgtccca	tgatcccttc	tgtgtctgcg	tgtccgaatc	240
	cccgtgtgac	cctgtcccag	cattttcccg	ccgaccctgc	gtgtccccgt	ggcgtgttcc	300
	gctctccctc	tcctgtctgc	cacccacctg	ccagtgttat	ttatgtctcc	ttcgtgggtg	360
	atggccacgc	cctcaccatg	tccttggcag	agggttcccc	tccgggatcc	cctgcctggt	420
	gcccaactg	cctcgcaagc	gctcgccacc	ctcacgtggc	tcacctgctg	ttgagccttg	480
	tgctgtcaat	aaacggtttg	aggattgcag	gattgtc			517
	<210>	161					
	<211>	259					
	<212>	DNA					
	<213>	智人 (Homo sapiens)					
	<400>	161					
	gcagaaccag	agtctgccac	tggggctcag	gaccaaggga	ggcagcacca	tgtccttctg	60
	tgggacactg	ccagccccag	ggctccagcc	cagcccgggtg	gatcctctgg	ggaagccagg	120
	accaggagag	aagcaaggtc	aagaaatccc	acagtttgat	gtattaaaga	aatgacttat	180
	ttctactcaa	aataaatggc	attgaagtct	ttctttaacc	ctttttgagt	taatttaata	240
	ataatgatct	gagacaagg					259
	<210>	162					
	<211>	443					
	<212>	DNA					
	<213>	智人 (Homo sapiens)					
	<400>	162					
	agtcagctta	catccctgag	caggaaagtt	tacccatgaa	gattgggtggg	atTTTTtgtt	60
	tgTTTTgttt	gTTTTgtttg	ttgTTTgttg	TTTgtTTTT	tgccactaat	TTtagtatTC	120
	attctgcatt	gctagataaa	agctgaagtt	actTTtatgtt	tgtctTTTTaa	tgcTTcatTC	180
	aatattgaca	TTTgtagTTg	agcggggggg	TTggTTTgct	TTggTTTata	TTTTTcagt	240
	tgTTTTgttt	tgcTTgttat	attaagcaga	aatcctgcaa	tgaaggtac	tatatTTgct	300
	agactctaga	caagatatTg	tacataaaag	aatTTTTTTg	tctTTaaata	gatacaaatg	360
	tctatcaact	ttaatcaagt	tgtaacttat	attgaagaca	atttgataca	taataaaaaa	420
[0024]	ttatgacaat	gtcctggact	ggt				443
	<210>	163					
	<211>	527					
	<212>	DNA					
	<213>	智人 (Homo sapiens)					
	<400>	163					
	gggccggcgc	ttgccactag	ggctcttgtg	TTTTTTTTct	tctgctatgt	gacttgcatc	60
	gtggagacag	ggcatttgag	TTTatatctc	tcatgcatat	TTtatTTaat	ccccaccctg	120
	ggctctggga	gcagcccgt	gagtggactg	gagtggagca	ttggctgaga	gaccaggagg	180
	gcactggagc	tgtcttgtcc	ttgctggttt	tctggatggt	tcccagaggg	TTTccatggg	240
	gtaggaggat	gggctcgccc	accagtgact	TTTTcTaaga	gctcccggcg	tggtggaaga	300
	ggggacaggt	ccctcaccca	cccacaatcc	tattctcggg	ctgagaaccc	tgcgtgggga	360
	cagggtctgc	ctcaggaatg	ggctgttttt	ggcctaaccc	tcagaaacac	tggggctggc	420
	acaaactctt	ggtttcttca	acaggagaat	TTTactgtgt	TTctTTTggt	tccattgttt	480
	ggagacattc	ctgggcacag	TTTggtccgt	tagaattaaa	agttgaa		527
	<210>	164					
	<211>	57					
	<212>	DNA					
	<213>	智人 (Homo sapiens)					
	<400>	164					
	ccctgttccc	agageccact	TTTTTTtctt	TTTTtgaaat	aaaatagcct	gtctttc	57
	<210>	165					
	<211>	81					
	<212>	DNA					
	<213>	智人 (Homo sapiens)					
	<400>	165					
	atcctgggat	tctagtatgc	aataagagat	gccctgtact	gatgcaaaat	TTaataaagt	60
	TTgtcacaga	gaatctttgt	a				81
	<210>	166					



	<211> 209	
	<212> DNA	
	<213> 智人 (Homo sapiens)	
	<400> 166	
	gccatgtctg ggcccccagc ccctggggag gacctactgc tcccaggggc tgagaggggt	60
	tcgggagcat aatgacaaac tgtcgtgcc ccagtggctg ggtgtgtgtg ggtgggatgg	120
	ggtgggggtc ctgggcccc cgtgtcttcc caggtttaca atcagagaat cacagctgct	180
	ttaataaatg ttatttataa tacacggaa	209
	<210> 167	
	<211> 128	
	<212> DNA	
	<213> 智人 (Homo sapiens)	
	<400> 167	
	gcggtggagc tgcgggcggc cagggccggg cgctctgtgc ggactggggc catgatcggg	60
	cccggggggc tgagcctggg accccacccc gtgttaatga aaaatgagtt ttggcagcgc	120
	ctgtggtc	128
	<210> 168	
	<211> 161	
	<212> DNA	
	<213> 智人 (Homo sapiens)	
	<400> 168	
	gcggccgcct gaccgtcgac cttggggctc tccaccegt ctggccccag tcgaactggg	60
	ggctagccac ctctctgtcc agccccaaa cctcccctc ctttccccct ctcgggggg	120
	ccaaggacag ggtggcctta ctcagtaaag gtgtttcctg c	161
	<210> 169	
	<211> 63	
	<212> DNA	
	<213> 智人 (Homo sapiens)	
[0025]	<400> 169	
	cctcgccccg gacctgcct cccgccaggt gcaccacct gcaataaatg cagcgaagcc	60
	ggg	63
	<210> 170	
	<211> 70	
	<212> DNA	
	<213> 智人 (Homo sapiens)	
	<400> 170	
	gacgcaggcc agccccggcc cctagccctc aggccttctt tcttatccaa ataaatgttt	60
	cttaatgagg	70
	<210> 171	
	<211> 89	
	<212> DNA	
	<213> 智人 (Homo sapiens)	
	<400> 171	
	ccccatgcgc gacacgcccc ccactcagcc cagcccgcct gtccctagat tcagccacat	60
	cagaaataaa ctgtgactac acttgcaa	89
	<210> 172	
	<211> 92	
	<212> DNA	
	<213> 智人 (Homo sapiens)	
	<400> 172	
	gagtcagagc ccacatggcc gggccctctg tagctccctc ctcccatctc cttccccag	60
	cccaataaag gtcccttagc cccgagtta aa	92
	<210> 173	
	<211> 580	
	<212> DNA	
	<213> 智人 (Homo sapiens)	

	<400> 173	
	tgcctgagcg ccacctactg cattttggac acctcacacc aaggggagat gcaccccaca	60
	gacacctggg actttccctt tctggttcct gtgtctccca ataaactctc tgaccaaccc	120
	tagctgtgct tctgcgagaa gaaaggggtg agcaacttct ggcagatatg aggccttttt	180
	cttttttttt tttttttttg agacaacgtc tggctctgtc acccaggtcg ggcgcagtg	240
	gcacaatctc ggcccactgc agcctcgacc tctggggctc aagggatcct catgcctcag	300
	cctcatatgt agctggggcc acagacatgc accaccacac ctggctcgag gccattttag	360
	ttctgaggtt gacgagctca ggagccggtt ccagcacggt gctgtgtttg tgaacagag	420
	aaaggggacc cccgaggacc ccagacaggg ccttaggact ctcatatctt cttgtcttct	480
	ccatctgggt gctcttgctc tctggttttc tctacctttt catggcccca gaatccatat	540
	gcagtaaagg aactctctgg aataaaatta aagtcctcct	580
	<210> 174	
	<211> 255	
	<212> DNA	
	<213> 小家鼠 (Mus musculus)	
	<400> 174	
	aaagccaaag gattcaggac caagcagcac catggccttc ctatggcaca tatgcatata	60
	aagaatttaa agcacggggt ggctgtcggc ttgttcagat cagcgagtaa actggtacat	120
	gaaaggatgt tcatcatatt attctgtctac tgagtacatc tgaaactttc cttgtctgt	180
	ttgtagtacg tattttggcta aatgctagaa ttttgcctta aatacagcaa aagctaataa	240
	acttgtagt aacta	255
	<210> 175	
	<211> 80	
	<212> DNA	
	<213> 小家鼠 (Mus musculus)	
	<400> 175	
	tgggggaggt gtcttagttt tccctatctt ttaagctgtt aacaagtctt attgcacttt	60
	gaataaagtt cttgcattcc	80
[0026]	<210> 176	
	<211> 245	
	<212> DNA	
	<213> 小家鼠 (Mus musculus)	
	<400> 176	
	gcttgatccc tgacattggc gccggtctg gactcgtgct cgggtgtgctc tggaaactac	60
	ttccctgtc ccaggcgtcc ctgctccggg ttccatggt cccggtccc tgtatcccag	120
	gcttgatccc tgtggaccag ggttactgtt ttaccactaa catctccttt tggctcagca	180
	ttcaccgatc tttagggaaa tgctgttga gagcaaataa ataaacgcat tcatttctct	240
	atgca	245
	<210> 177	
	<211> 25	
	<212> DNA	
	<213> 小家鼠 (Mus musculus)	
	<400> 177	
	agacaataaa ggactttatt cttgt	25
	<210> 178	
	<211> 475	
	<212> DNA	
	<213> 小家鼠 (Mus musculus)	
	<400> 178	
	gggctcccag gagagcccag ggactctgcg ccacactgaa gataactgag atggagcgca	60
	ctccagccgc gcaggactga cgaggccaca cccatgggct ctgtggcttt ctctgtctc	120
	atggattgaa cagggtcccc catgtgggce cgggcgatct gctgtctgct tatctaagga	180
	tcctcttaga ggtcgccttt ttttttttt tttttaagga gcgattaaga aaaggggaagg	240
	aaattcaagt cagtgttttg gctgtttttg tttttgttt gtttgttttt gtttgttttt	300
	tgtggtgttg ggttttgagg tttttgttt ggttggtttg gttttgggat cgtgtgttta	360
	ttttattgtt tttgttttgg ggggttggtc gtttgttgtt ttggggggtt gttttgtttt	420
	gtttgtttgt tggttactgt tttcaagaa gaaatataaa ttcactctcc aagct	475

	<210>	179	
	<211>	325	
	<212>	DNA	
	<213>	小家鼠 (Mus musculus)	
	<400>	179	
		ccccgcccc tctgtccca ttggctccaa gatcagcagg aggaacacag aactggaggg	60
		gcgggtggag ggccggtgtg ctcttcccag caggcctgag aatgagtggg aatcagggca	120
		tttgggcctg gctggagatc aggtttgcac actgtcttga gaacaatggt ccagtgagga	180
		atcctgatct ctgccecca attgagccag ctggccacag cccaccctt ggaatatcac	240
		cattgcaatc atagcttggg ttgcttttaa acgtggattg tcttgaagtt ctccagcctc	300
		catggaagga tgggtcagat cccag	325
	<210>	180	
	<211>	162	
	<212>	DNA	
	<213>	小家鼠 (Mus musculus)	
	<400>	180	
		gacctctggc tgagattgaa taggacaacc aatgaccttg acactgcctc ttaggcactt	60
		agctctagca atgccctgga agtctcttca gctgagctcc agggcaaagg cagaagggtg	120
		cctctgtacg acggcacagt gagctggata ggtagtcat gc	162
	<210>	181	
	<211>	240	
	<212>	DNA	
	<213>	小家鼠 (Mus musculus)	
	<400>	181	
		cctaaccata cctaattggt gtcttgttct cctatacaat aatggagagt actcttgacc	60
		ccacgaagag ctgggcctct catgtggtgt tagtgacgga tatgaaataa tgctctgtat	120
		ttcttgtgaa atgatggcta tgtctttgct gtttctcatt tgtttttaac ttgtatactt	180
		tgaaatagtc tctgtagtgt tttcctggct ttgaaaaaaa tgttattaaa ggaaaattgt	240
[0027]	<210>	182	
	<211>	108	
	<212>	DNA	
	<213>	小家鼠 (Mus musculus)	
	<400>	182	
		ccggccccag ggtgacagag tcagggcccc atgggaacaa gcatggtgga ggcgcctcag	60
		ggccacgggc cagctctgtt gcaataaatc tgcagctctg cttcttgc	108
	<210>	183	
	<211>	279	
	<212>	DNA	
	<213>	小家鼠 (Mus musculus)	
	<400>	183	
		agagccggag agtgggctga gtccctgctg atgaggcgt ctctcactt cccagaagac	60
		tcaaattctt ttctcctcg cacaggttgg gggggtcaga actatcggac ttgtccact	120
		caaacatgat gagtgagcac acaaagggcc agagtcagaa aaggaaggga acaagttgag	180
		ctactgctag gaacctgagg gaaaagatgg atgagaaggt ggcaagtccg tgacggtggc	240
		aacttccaa accaggaaat aaaatgtctt ttactataa	279
	<210>	184	
	<211>	258	
	<212>	DNA	
	<213>	小家鼠 (Mus musculus)	
	<400>	184	
		ggtgagcgt gccagttctt cccgtgtgc cggtgctgag cctccttgct cctgcatctc	60
		gaccattccg tgttggctgt gtcgctgac ctgcgtacct gtggaggatt cggaacaagt	120
		catggagaga ctgtccgggt ccgtccttg tgaactgtgc agaaggagt atccccagcat	180
		cggcaagcga gggagaagac tgcacgagag tgatgcgcat ttcgagtctg cttttcgata	240
		gttgatgtct tcttgcct	258
	<210>	185	
	<211>	285	

<212>	DNA	
<213>	小家鼠 (Mus musculus)	
<400>	185	
gtctctctcc	ttcggactgg	tgccccttca ctctgctagc tctgagcctg cactgtaggc 60
tectgggggc	tgtgtttcag	tgcccctggc cctgggggct gggcaggagc agggccccct 120
ttgaaggggt	tgtgtgtgt	aagttatitt gtacatgtct ggggtgtgggt ttacatctgc 180
ttccctgccc	tcgtcagcc	cactgtatga attgtataaa tgtttctatt aaattgggac 240
tgctctttcc	ttagctttat	atatattaaa atgtgtacat ctctc 285
<210>	186	
<211>	47	
<212>	DNA	
<213>	小家鼠 (Mus musculus)	
<400>	186	
ggggccagga	cgctgctcct	gtgaccaaca gtaaaatatt gtgactc 47
<210>	187	
<211>	521	
<212>	DNA	
<213>	小家鼠 (Mus musculus)	
<400>	187	
gaggacaaca	ggacacacag	tggcagcagg gactgtggta gcagagtga cacatctgtc 60
cttcttctgt	ggggctcttc	actgccaaca cctgcaacgg tgctttgtct ctctgacagc 120
tatggtgtct	tgtgtcacac	ttctagttag tgggaatttt agacggggaa cacagggcta 180
gtcagggcct	ttgtgtgctt	ggtgtggagt gactgagaac cgtctatggt tcaaggctcc 240
actggggata	aactgcttag	agcactgtcc tagagggcaa gtgtagcctt cgcctccggg 300
cccaggcagg	ctatgcagtc	agcagtaggg tctgtgctcc atgcgggtcc aggcgcacgg 360
ctctcctatt	ctgttgtcat	ttgtgccctc tatgggcagg tgtgtttcaa gttggttttc 420
tgttgctgag	gctttcatac	acatcagtta ccatctcagc tgatttgtct actgaaagct 480
tgctgttttc	aataaatctt	agtttgccat ggttttaagt c 521
[0028]		
<210>	188	
<211>	63	
<212>	DNA	
<213>	小家鼠 (Mus musculus)	
<400>	188	
gacactacct	tcttttaaaa	atccttgtat agtgacgtgt tgtgtaaata cagtttttat 60
tcc		63
<210>	189	
<211>	284	
<212>	DNA	
<213>	小家鼠 (Mus musculus)	
<400>	189	
gtcccagctt	cctttcttgt	ctcttcagge ccagctgctg ggcaaggaat cccgatgccc 60
ccccacggg	ggcagcacac	ttagcagaca atcctggagc atacagatag cccaagcaca 120
tagatgtact	aaactgggac	catgctttag gctccagacg agagtctttt ctaacttggg 180
gatgcagtct	cagatactgg	gggccctct gcctcatctc cattcttaca cctcaaactc 240
gtacagttac	ttttgtactg	actgtaataa acagccaaac agtt 284
<210>	190	
<211>	570	
<212>	DNA	
<213>	小家鼠 (Mus musculus)	
<400>	190	
aagttcgagt	caaagctgtc	attccagtgc tgccatcagg agcagcatcc aaggggccaa 60
atcttcggaa	taactggaca	gatcattacg gagcatagc agaactgcac tgaatatttt 120
aaggccctaa	tgagtttaca	gctgtaacct tccatggatc ttcccttaag aactgtcttc 180
tgtttatctt	ctaactaaca	ggttgttctg ttagtatacat ttcggtaatt ttgctatatt 240
tgtgtccaag	gaagtaagag	ttgttctgtt ttgggcatcat gttgtctttt ttctttcttc 300
agtagcttat	gatataattt	aatcatgttt acaatgttac aatttagtat tgatgtttta 360

	tgaagttaaa ttattaaatg aatggtctta aatccactgt ggtggttttt ttttgaaaaa	420
	ttatataatt accataagcc aaaaatcaaa tatttggaat acctactgtg aaattcaaga	480
	gattaaaggt tgtacttgat acttggtatt tttcttaaata aaatctagtt atcaagttat	540
	ttttcttaaa taaatctagt tatcaaatgt	570
	<210> 191	
	<211> 208	
	<212> DNA	
	<213> 小家鼠 (Mus musculus)	
	<400> 191	
	caagcctagg ctaaagacac tttggtatcc cacacctact gcctgctcca aaaggcagaa	60
	ccactgacca ctttgccttt ccaaactcat gagcacagcc cagaaggaac agaagacttc	120
	tgctaagtc atcatgccat aaacagacag aagcaggga tggtctctatc ctagctgcc	180
	tgctaagtaa acacggtttt gaagcgtc	208
	<210> 192	
	<211> 532	
	<212> DNA	
	<213> 小家鼠 (Mus musculus)	
	<400> 192	
	tgacttcaaa taaagatctg tttgatata cactagcctc ttccctccc cagaaaactt	60
	tgggcacctt ctggcttttg ttaggaactg agttggacag gttgggtaaa tttcaacctc	120
	agggcacat gccaggaaa gctggggaga ggcgatgtt ttttctcag tttggagcgg	180
	gaagaaggag ttttaatgac tgtgtctct cttctctcc ccatggaagg ggggggggat	240
	ccctggctca gcateactg ctggtggaaa cctcgtcgc aggtgccgct gtgaggtgtg	300
	ccgctgtgag gtgcgtgctc tggaagagag gcggagctt tccgtctct gctgcctctg	360
	cctgcagtac caccctcgt ctaagcactc atgattttat actttcagaa aaaagtttta	420
	agacttcttt gggttttttg tttttgttt gaagcctaag aaagtcacta gttcttgcc	480
	ttatgtattt atatgctgta cctaacaata aagaaagaag aaaaaaaca aa	532
	<210> 193	
[0029]	<211> 74	
	<212> DNA	
	<213> 小家鼠 (Mus musculus)	
	<400> 193	
	gcggcggagc ctggggctcc ctggactgac accctacccc tgagttaata aaagttgaat	60
	tttgacagct ccca	74
	<210> 194	
	<211> 222	
	<212> DNA	
	<213> 小家鼠 (Mus musculus)	
	<400> 194	
	agctatttaa agctagtagt ccattcacca ggacaggcca aacatccgga accagagtgg	60
	cctgcagagg acaggacagg agcccttcag tccttgggta tttcacagac aacctcagt	120
	gtatagaggt acaccacttt ccattcagg aatcagtgcg tcaacctgt gtgaccaag	180
	gagtgtggtt aattattaat aaagacattg catcccccc tc	222
	<210> 195	
	<211> 190	
	<212> DNA	
	<213> 小家鼠 (Mus musculus)	
	<400> 195	
	gaccacgatg caccacatc ccgccacctg ccaggactt gggacctgca aactgatgtc	60
	ccggcattct gcacagggtc atcaccagg aaggggtacc cgcttgact ttgcaggacc	120
	agcacctgca gctcaatccc ctctgtactg gcctgcctac cccactcaca gaaataaaat	180
	ctaaaacgac	190
	<210> 196	
	<211> 182	
	<212> DNA	
	<213> 小家鼠 (Mus musculus)	

	<400> 196	
	ccttcccgct gaggcatttg gagacgccag ggctgctgtc catggtttca acataaaacg	60
	gcctgtgacc agcagagccc aagagcagcc acagagcaga agtccctgtt cccittttta	120
	tggactctta tgcactacag gcgaacacaa aaaaaagcaa cggaataaag ccttctctcc	180
	tc	182
	<210> 197	
	<211> 51	
	<212> DNA	
	<213> 小家鼠 (Mus musculus)	
	<400> 197	
	cctcccacat atcctccata tttgtgaata aagtagggaa aatgctgtca t	51
	<210> 198	
	<211> 202	
	<212> DNA	
	<213> 小家鼠 (Mus musculus)	
	<400> 198	
	gtgaacctga agattctaga gaccaagtga catcctctcc tacctagact cctataaacc	60
	aagtctcttc tttcctgacg aagggtaaat atctttcttg tggcctaaat tatagactct	120
	tgttcaact caccctggaa aaatctctat taaaacgaga tgggagattg aaatggattc	180
	aaataaagat gctctagcag ga	202
	<210> 199	
	<211> 289	
	<212> DNA	
	<213> 小家鼠 (Mus musculus)	
	<400> 199	
	ggagcgcaac gtcgcaggga atagactcag acgccggtag agttgcccga gggcgagaag	60
	ttggggcaag ctgaaagtgg gggtctgtcc ctctgtaggt ttttagacgc atccccgtc	120
	gacggctgct taccaaccac cagcacctgc gatacctcat ccatttctgt gaatatattt	180
[0030]	gaaggcatct ttgcaagacg gggtcattaa aacacagaca ccaccctac gttccagcc	240
	attccaaagt attaccgact taataaatat tcacgtttct tgaaccact	289
	<210> 200	
	<211> 95	
	<212> DNA	
	<213> 小家鼠 (Mus musculus)	
	<400> 200	
	ggttaacacg aagcttaggg cctgggtgtg agaacctaaag acccccagcc agtgacctga	60
	cttttaaacc tgacaataaa ggtactggac acgcg	95
	<210> 201	
	<211> 472	
	<212> DNA	
	<213> 小家鼠 (Mus musculus)	
	<400> 201	
	ccttagctca gagaactctg actccagaac gtgggtggct atggggccct gcctctctgt	60
	cctccttccc atgctaccega gagcacaagt gtgctagcag ggcagggggg ggggagggca	120
	gaagcagtga tgacctggtg gccaccact gccagtgtcc ctctaccac agctgctctc	180
	tgcagctttg ggacaaaggg cacatggaac tgcagtgttc tgtctggagt ttgaggcctc	240
	cacaagctgg tgtcccagca gggccagcct gcaatccggg caccactcca agcagttgat	300
	aaccctgtct cttgggcagg cccctgggt agaggtgctt cagccaagge tgattcccca	360
	agatccgcac gatgggagga gggaggtgtc tcttccaaat tgcctctga actctgtatg	420
	tgtgctgtgg tgcacacata catgctagct aaataaacia acaaacaat aa	472
	<210> 202	
	<211> 273	
	<212> DNA	
	<213> 小家鼠 (Mus musculus)	
	<400> 202	
	gtggtaccta ctgtctgaca cccacgggga aactgagcca ggtccaggcc gatgagaagt	60

	tccagtgagg	ctgaagtggc	caccaggggt	cagcagtgct	ccagttgctg	ggcttaaagc	120
	ttcctgggtc	ctgtctgcca	ggtcatggag	gattcccca	tctggcatcc	ccacgatgcc	180
	tctggagaga	tggccttgat	tgccctcaa	agccacctga	accgtcgtca	acttaccag	240
	cctgtctcca	ttaaaccacca	ggaaccaact	gag			273
	<210>	203					
	<211>	87					
	<212>	DNA					
	<213>	小家鼠 (Mus musculus)					
	<400>	203					
	tgcagagagt	cctcagatgt	tccttcattc	aagagtttaa	ccattttctaa	caatatgtag	60
	ttatcattaa	atctttttta	aagtgtg				87
	<210>	204					
	<211>	532					
	<212>	DNA					
	<213>	人工序列 (Artificial Sequence)					
	<220>						
	<223>	人工核酸序列					
	<400>	204					
	tgacttcaaa	taatgatctg	tttgatatca	cactagcctc	ttccctccc	cagaaaactt	60
	tgggcacctt	ctggcttttg	ttaggaactg	agttggacag	gttgggtaaa	tttcaacctc	120
	agggcacat	gccagggaaa	gctggggaga	ggcgatgttc	ttttcctcag	tttgagcgg	180
	gaagaaggag	ttttaatgac	tgtgcttct	cttctcctcc	ccatggaagg	ggggggggat	240
	ccctggctca	gcateccactg	ctggtggaaa	cctcgtcgcg	aggtgccgct	gtgaggtgtg	300
	ccgctgtgag	gtgcgtgctc	tggaagagag	gcgagccttg	tccgcttctc	gctgcctctg	360
	cctgcagtac	caccctcgg	ctaagcactc	atgattttat	actttcagaa	aaaagtttta	420
	agacttcttt	gggttttttg	ttttgtttt	gaagcctaag	aaagtcacta	gttcttgcc	480
	ttatgtattt	atatgctgta	cctaacaata	aagaaagaag	aaaaaaaaa	aa	532
	<210>	205					
	<211>	2035					
	<212>	RNA					
	<213>	人工序列 (Artificial Sequence)					
	<220>						
	<223>	人工核酸序列					
	<400>	205					
	ggggcgcugc	cuacggaggu	ggcagccauc	uccuucucgg	caucaagcuu	gaggauggag	60
	gacgccaaga	acaucaagaa	gggcccggcg	cccuucuaac	cgcuggagga	cgggaccgcc	120
	ggcgagcagc	uccacaaggc	caugaagcgg	uacgcccugg	ugccgggcac	gaucgcuuc	180
	accgacgccc	acaucgaggu	cgacaucacc	uacgcggagu	acuucgagau	gagcugcgc	240
	cuggccgagg	ccaugaagcg	guacggccug	aacaccaacc	accggaucgu	ggugugcucg	300
	gagaacagcc	ugcaguucuu	caugccggug	cugggcgccc	ucuucaucgg	cguggccguc	360
	gccccggcga	acgacaucua	caacgagcgg	gagcugcuga	acagcauggg	gaucagccag	420
	ccgaccgugg	uguucgugag	caagaagggc	cugcagaaga	uccugaacgu	gcagaagaag	480
	cugcccauca	uccagaagau	caucaucaug	gacagcaaga	ccgacuacca	gggcuuccag	540
	ucgauguaca	cguucgugac	cagccaccuc	ccgccgggcu	ucaacgagua	cgacuucguc	600
	ccggagagcu	ucgaccggga	caagaccauc	gcccugauca	ugaacagcag	cggcagcacc	660
	ggccugccga	aggggguggc	ccugccgcac	cggaccgccu	gcgugcguu	cucgcacgcc	720
	cgggacccca	ucuucggcaa	ccagaucac	ccggacaccg	ccauccugag	cguggugccg	780
	uuccaccacg	gcuucggcau	guucacgacc	cugggcuacc	ucaucugcgg	cuuccgggug	840
	guccugaugu	accgguucga	ggaggagcug	uuccugcgga	gccugcagga	cuacaagauc	900
	cagagcgcgc	ugcucgugcc	gaccucguuc	agcuucucg	ccaagagcac	ccugaucgac	960
	aaguacgacc	ugucgaaccu	gcacgagauc	gccagcgggg	gcgccccgc	gagcaaggag	1020
	gugggcgagg	ccguggccaa	gcgguuccac	cucccgggca	uccgccaggg	cuacggccug	1080
	accgagacca	cgagcgcgau	ccugaucacc	cccagggggg	acgacaagcc	gggcgccgug	1140
	ggcaaggugg	ucccgguucuu	cgaggccaag	gugguggacc	uggacaccgg	caagaccucg	1200
	ggcgugaacc	agcggggcga	gcugugcgug	cggggggcga	ugaucaugag	cggcuacgug	1260
	aacaaccg	aggccacca	cgcccucauc	gacaaggacg	gcuggcugca	cagcggcgac	1320

[0031]



	aucgccuacu	gggacgagga	cgagcacuuc	uucaucgucg	accggcugaa	gucgcugauc	1380
	aaguacaagg	gcuaccaggu	ggcgccggcc	gagcuggaga	gcauccugcu	ccagcacccc	1440
	aacaucucg	acgccggcgu	ggccgggcug	ccggacgacg	acgccggcga	gcugccggcc	1500
	gcgguggugg	ugcuggagca	cggcaagacc	augacggaga	aggagaucgu	cgacuacgug	1560
	gccagccagg	ugaccaccgc	caagaagcug	cggggcggcg	ugguguucgu	ggacgagguc	1620
	ccgaagggcc	ugaccgggaa	gcucgacgcc	cggagaaucc	gcgagauccu	gaucaaggcc	1680
	aagaagggcg	gcaagaucgc	cguguaagac	uagugcauca	cauuuaaaag	caucucagcc	1740
	uaccaugaga	auaagagaaa	gaaaauaag	aucaauagcu	uauucaucuc	uuuuucuuuu	1800
	ucguuggugu	aaagccaaca	cccugucuaa	aaaacauaaa	uuucuuuau	cauuuugccu	1860
	cuuuucucug	ugcuucaauu	auuaaaaaau	ggaaagaacc	uagaucuaaa	aaaaaaaaaa	1920
	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	augcaucccc	1980
	ccccccccc	ccccccccc	cccccccaa	ggcucuuuuc	agagccacca	gaauu	2035
	<210>	206					
	<211>	186					
	<212>	DNA					
	<213>	智人 (Homo sapiens)					
	<400>	206					
	catcacattt	aaaagcatct	cagcctacca	tgagaataag	agaaagaaaa	tgaagatcaa	60
	aagcttattc	atctgttttt	cttttcgtt	ggtgtaaagc	caacaccctg	tctaaaaaac	120
	ataaatttct	ttaatcattt	tgctctttt	ctctgtgctt	caattaataa	aaatggaaa	180
	gaatct						186
	<210>	207					
	<211>	186					
	<212>	DNA					
	<213>	智人 (Homo sapiens)					
	<400>	207					
	catcacattt	aaaagcatct	cagcctacca	tgagaataag	agaaagaaaa	tgaagatcaa	60
	tagcttattc	atctcttttt	cttttcgtt	ggtgtaaagc	caacaccctg	tctaaaaaac	120
	ataaatttct	ttaatcattt	tgctctttt	ctctgtgctt	caattaataa	aaatggaaa	180
	gaacct						186
	<210>	208					
	<211>	42					
	<212>	DNA					
	<213>	智人 (Homo sapiens)					
	<400>	208					
	ggcgctgcct	acggaggtgg	cagccatctc	cttctcgga	tc		42
	<210>	209					
	<211>	24					
	<212>	DNA					
	<213>	人工序列 (Artificial Sequence)					
	<220>						
	<223>	人工核酸序列					
	<400>	209					
	caaaggctct	tttcagagcc	acca				24
	<210>	210					
	<211>	2316					
	<212>	RNA					
	<213>	人工序列 (Artificial Sequence)					
	<220>						
	<223>	人工核酸序列					
	<400>	210					
	ggggaaacag	uguggggccu	agagcgcugg	gugggcgcgu	ucugcggccu	gagcaggggac	60
	ggguagugaa	gcgguuacgc	cccuucucg	cgucuuggcg	ggagccugac	gccccgcuuc	120
	uccccuacg	agguguccca	ccggcgcccc	ccgaggccua	ggccuccgca	gccgccucc	180
	gucuccucag	ccccgacgu	gcgccuuggg	ccuugugcgc	auuuuuuucg	ggggaaaacu	240
	gaggcucgga	gugcgaaagu	cagccgaggu	cgcgccgcc	aggacagaga	agggcugugg	300
	ucggcugauc	cgcggcauuc	ccgggaagcu	ugaggaugga	ggacgccaag	aacaucaaga	360



	agggcccggc	gcccucucac	ccgcuggagg	acgggaccgc	cggcgagcag	cuccacaagg	420
	ccaugaagcg	guacgccug	gugccgggca	cgaucgccuu	caccgacgcc	cacaucgagg	480
	ucgacaucac	cuacgcggag	uacuu cgaga	ugagcgugcg	ccuggccgag	gccaugaaagc	540
	gguacggccu	gaacaccaac	caccggaucg	uggugugcuc	ggagaacagc	cugcaguucu	600
	ucaugccggu	gcugggcgcc	cucuucaucg	gcguggccgu	cggcccgccg	aacgacauuc	660
	acaacgagcg	ggagcugcug	aacagcaugg	ggaucagcca	gccgaccgug	guguucguga	720
	gcaagaagg	ccugcagaag	auccugaacg	ugcagaagaa	gcugcccauc	auccagaaga	780
	ucaucaucau	ggacagcaag	accgacuacc	agggcuucca	gucgauguac	acguucguga	840
	ccagccaccu	cccgcgcggc	uucacagagu	acgacuucgu	cccggagagc	uucgaccggg	900
	acaagaccu	cggccugauc	augaacagca	gcggcagcac	cggccugccg	aagggggugg	960
	ccugcccgca	ccggaccgcc	ugcgugcgcu	ucucgcacgc	ccgggacccc	aucuucggca	1020
	accagaucau	cccggacacc	gccauccuga	gcguggugcc	guuccaccac	ggcuucggca	1080
	uguucacgac	ccugggcuac	cucaucugcg	gcuuccgggu	gguccugaug	uaccgguucg	1140
	aggaggagcu	guuccugcgg	agccugcagg	acuacaagau	ccagagcgcg	cugcucgugc	1200
	cgaccucguu	cagcuucuu	gccaagagca	cccugaucga	caaguacgac	cugucgaacc	1260
	ugcagagau	cggcagcgcc	ggcgccccgc	ugagaagga	ggugggcgag	gccguggcca	1320
	agcgguucca	ccucccgggc	auccgccagg	gcuacggccu	gaccgagacc	acgagcgcca	1380
	uccugaucac	ccccgagggg	gacgacaagc	cgggcgcgcu	gggcaaggug	guccguucu	1440
	ucgaggccaa	ggugggugac	cuggacaccg	gcaagaccuu	gggcgugaac	cagcggggcg	1500
	agcugugcgu	gcggggggccg	augaucauga	gcggcuacgu	gaacaaccgc	gaggccacca	1560
	acggccucau	cgacaaggac	ggcuggcugc	acagcggcga	caucgccuac	ugggacgagg	1620
	acgagcacuu	cuucaucguc	gaccggcgua	agucgcugau	caaguacaag	ggcuaccagg	1680
	uggcgccggc	cgagcuggag	agcauccugc	uccagcaccc	caacaucuu	gacgccggcg	1740
	uggccgggcu	gccggacgac	gacgccggcg	agcugccggc	cgcgguggug	gugcuggagc	1800
	acggcaagac	caugacggag	aaggagaucc	ucgacuacgu	ggccagccag	gugaccaccg	1860
	ccaagaagcu	gcggggcggc	gugguguuuc	uggacgaggu	cccgaagggc	cugaccggga	1920
[0033]	agcucgacgc	ccggaagauc	cgcgagauc	ugaucaggc	caagaagggc	ggcaagaucg	1980
	ccguguaaga	cuagugcauc	acauuuuaaa	gcaucucagc	cuaccaugag	aaauagagaa	2040
	agaaaaugaa	gaucaauagc	uuuuuucuu	uuuuuucuu	uucguuggug	uaaagccaac	2100
	accugucua	aaaaacauaa	auuuuuuaa	ucauuuugcc	ucuuuucucu	gugcuucaau	2160
	uaauaaaaaa	uggaaagaac	cuagaucuaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	2220
	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaugcauccc	ccccccccc	ccccccccc	2280
	cccccccaa	aggcucuuuu	cagagccacc	agaauu			2316
	<210>	211					
	<211>	1948					
	<212>	RNA					
	<213>	人工序列 (Artificial Sequence)					
	<220>						
	<223>	人工核酸序列					
	<400>	211					
	ggggcgucg	cuacggaggu	ggcagccauc	uccuucucgg	caucaagcuu	gaggauggag	60
	gacgccaaga	acaucaagaa	gggcccggcg	cccuucuaac	cgcuggagga	cgggaccgcc	120
	ggcgagcagc	uccacaaggc	caugaagcgg	uacgcccugg	ugccgggcac	gaucgcuuc	180
	accgacgcc	acaucgaggu	cgacauacc	uacgcggagu	acuucgagau	gagcgugcgc	240
	cuggccgagg	ccaugaagcg	guacggccug	aacaccaacc	accggauugu	ggugugcucg	300
	gagaacagcc	ugcaguucuu	caugccggug	cugggcgccc	ucuucaucgg	cguggccguc	360
	gccccggcga	acgacauca	caacgagcgg	gagcugcuga	acagcauggg	gaucagccag	420
	ccgaccgugg	uguucgugag	caagaagggc	cugcagaaga	uccugaacgu	gcagaagaag	480
	cugcccauca	uccagaagau	caucaucaug	gacagcaaga	ccgacuacca	ggguuccag	540
	ucgauguaca	cguucgugac	cagccaccuc	ccgccgggcu	ucaacgagua	cgacuucguc	600
	ccggagagcu	ucgaccggga	caagaccauc	gccugauca	ugaacagcag	cggcagcacc	660
	ggccugccga	aggggggugg	ccugccgcac	cggaccgccu	gcgugcgcuu	cucgcacgcc	720
	cgggacccca	ucuuucggca	ccagaucauc	ccggacaccg	ccauccugag	cguggugccg	780
	uuccaccacg	gcuucggcau	guucacgacc	cugggcuacc	ucaucugcgg	cuuccgggug	840
	guccugaugu	accgguucga	ggaggagcug	uuccugcgga	gccugcagga	cuacaagauc	900
	cagagcgcg	ugcucgugcc	gaccucguuc	agcuucuuuc	ccaagagcac	ccugaucgac	960

[0034]	aaguacgacc	ugucgaaccu	gcacgagauc	gccagcgggg	gcgccccgcu	gagcaaggag	1020
	gugggcgagg	ccguggccaa	gcggguccac	cucccgggca	uccgccaggg	cuacggccug	1080
	accgagacca	cgagcgcgau	ccugaucacc	cccagggggg	acgacaagcc	gggcgccgug	1140
	ggcaaggugg	ucccgucuuu	cgaggccaag	gugguggacc	uggacaccgg	caagaccucg	1200
	ggcgugaacc	agcggggcga	gcugugcgug	cggggggccga	ugaucaugag	cggcuacgug	1260
	aacaacccgg	aggccaccaa	cggccucauc	gacaaggacg	gcuggcgugca	cagcggcgac	1320
	aucgccuacu	gggacgagga	cgagcacuuc	uucaucgucg	accggcugaa	gucgcugauc	1380
	aaguacaagg	gcuaccaggu	ggcggccggc	gagcuggaga	gcauccugcu	ccagcacccc	1440
	aacaucuuuc	acgccggcgu	ggccgggcu	ccggacgacg	acgccggcga	gcugccggcc	1500
	gcgguggugg	ugcuggagca	cggcaagacc	augacggaga	aggagaucgu	cgacuacgug	1560
	gccagccagg	ugaccaccgc	caagaagcug	cggggcggcg	ugguguucgu	ggacgagguc	1620
	ccgaagggcc	ugaccgggaa	gcucgacgcc	cgggaagaucc	gcgagaucuu	gaucaaggcc	1680
	aagaaggggc	gcaagaucgc	cguguaagac	uagucgccgc	cgccgccgcc	gccggagccu	1740
	gucgccgucc	uguccccagc	cugcuugugu	cccugaggu	ugucaauaaa	ccugcccucg	1800
	ggcugccgcc	ucccagaucu	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	1860
	aaaaaaaaaa	aaaaaaaaaa	aaaaucauc	ccccccccc	ccccccccc	ccccccccc	1920
	aaaggcucuu	uucagagcca	ccagaauu				1948

图 1A: 32L4 - PpLuc(GC) - 白蛋白7-A64 - C30 - hSL

SEQ ID NO: 205

GGGGCGCUGCCUACGGAGGUGGCAGCCAUCUCCUUCUCGGCAUCAAGCUUGAGGAUGGAG  
GACGCCAAGAACAUCAAGAAGGGCCCGGGCGCCUUCUACCCGCGUGGAGGACGGGACCGCC  
GGCGAGCAGCUCCACAAGGCCAUGAAGCGGUACGCCUUGGUGCCGGGCACGAUCGCCUUC  
ACCGACGCCCACAUUCGAGGUCGACAUCACCUACGCGGAGUACUUCGAGAUGAGCGUGCGC  
CUGGCCGAGGCCAUGAAGCGGUACGGCCUGAACACCAACCACCGGAUCGUGGUGUGCUCG  
GAGAACAGCCUGCAGUUCUUCUUGCCGGUGCUGGGCGCCUUCUUCUUCGCGUGGGCCGUC  
GCCCCGGGGAACGACAUCUACAACGAGCGGGAGCUGCUGAACAGCAUGGGGAUCAGCCAG  
CCGACCGUGGUGUUCGUGAGCAAGAAGGGCCUGCAGAAGAUCUGAACGUGCAGAAGAAG  
CUGCCCAUCAUCCAGAAGAUCAUCAUUAUGGACAGCAAGACCGACUACCAGGGCUUCCAG  
UCGAUGUACACGUUCGUGACCAGCCACCUCGCCCGGGCUUCAACGAGUACGACUUCGUC  
CCGGAGAGCUUCGACCGGGACAAGACCAUCGCCCUGAUCAUGAACAGCAGCGGCAGCACC  
GGCCUGCCGAAGGGGGUGGCCUGCCGCACCGGACCGCCUGCGUGCGCUUCUCGACGCC  
CGGGACCCCAUCUUCGGCAACCAGAUAUCCCGGACACCGCCAUCCUGAGCGUGGUGCCG  
UUCACCACGGCUUCGGCAUGUUCACGACCCUGGGCUACCUCUUCUGCGGCUUCCGGGUG  
GUCCUGAUGUACCGGUUCGAGGAGGAGCUGUUCUGCGGAGCCUGCAGGACUACAAGAUC  
CAGAGCGCGCUGCUCGUGCCGACCCUGUUCAGCUUCUUCGCCAAGAGCACCCUGAUCGAC  
AAGUACGACCUUGUCGAACCUGCACGAGAUUCGCCAGCGGGGGCGCCCCGUGAGCAAGGAG  
GUGGGCGAGGCCGUGGCCAAGCGGUUCCACCUCCCGGGCAUCCGCCAGGGCUACGGCCUG  
ACCGAGACCACGAGCGCAUCCUGAUCACCCCCGAGGGGGACGACAAGCCGGGCGCCGUG  
GGCAAGGUGGUCCCGUUCUUCGAGGCCAAGGUGGUGGACCUGGACACCGGCAAGACCCUG  
GGCGUGAACCAGCGGGGCGAGCUGUGCGUGCGGGGGCCGAUGAUCAUGAGCGGCUACGUG  
AACAACCCGGAGGCCACCAACGCCCUCAUCGACAAGGACGGCUGGCUGCACAGCGGCGAC  
AUCGCCUACUGGGACGAGGACGAGCACUUCUUCUUCGUGCACCAGGCUAGAGUCGUGAUC  
AAGUACAAGGGCUACCAGGUGGCGCCGGCCGAGCUGGAGAGCAUCCUGCUCCAGCACCCC  
AACAUUCUUCGACGCCGGCGUGGCCGGGCGUGCCGGACGACGACGCCGGCGAGCUGCCGGCC  
GCGGUGGUGGUGCUGGAGCACGGCAAGACCAUGACGGAGAAGGAGAUCGUCGACUACGUG  
GCCAGCCAGGUGACCACCGCCAAGAAGCUGCGGGGCGGCGUGGUGUUCGUGGACGAGGUC  
CCGAAGGGCCUGACCGGGAAGCUCGACGCCCGGAAGAUCGCGGAGAUCCUGAUCAGGCC  
AAGAAGGGCGGCAAGAUCGCCGUGUAAGACUAGUGCAUACAUUUAAAAGCAUCUCAGCC  
UACCAUGAGAAUAAGAGAAAAGAAAUGAAGAUCAAUAGCUUAUUCUUCUUCUUCUUCU  
UCGUUGGUGUAAAAGCCAACACCCUGUCUAAAAACAUAUUUUCUUUAAUUAUUUUGCCU  
CUUUUCUCUGUGCUUCAAUUAAUAAAAAUGGAAAGAACCUAGAUCUAAAAAAAAAAAAA  
AAAUGCAUCCCC  
CCCCCCCCCCCCCCCCCCCCCCCCCAAGGCUCUUUCAGAGCCACCAGAAU

图1A

图 1B: 人工核酸 (SEQ ID NO: 210).

[5'-UTR 对应于 SEQ ID NO: 1] - PpLuc(GC) - 白蛋白 7-A64 - C30 - hSL

GGGGAAACAGUGUGGGGCCUAGAGCGCUGGGUGGGCGCGUUCUGCGGCCUGAGCAGGGAC  
GGGUAGUGAAGCGGUUACGCCCUUCUUCGCGUCUUGGCGGGAGCCUGACGCCCCGCUUC  
UCCCCUAACGAGGUGUCCACCGGCGCCCGCCGAGGCCUAGGCCUCCGCAGCCGCCCUCC  
GUCUCCUCAGCCCCGACGCGCGCCUUGGGCCUUGUGCGCAUUUUUUUCGGGGGAAAACU  
GAGGCUCGGAGUGCGAAAGUCAGCCGAGGUCGCCCCGCCAGGACAGAGAAGGGCUGUGG  
UCGGCUGAUCCGCGGCAUUCGCGGAAGCUUGAGGAUGGAGGACGCCAAGAACAUCAAGA  
AGGGCCCGGCGCCCUUCUACCCGCGUGGAGGACGGGACCGCCGGCGAGCAGCUCCACAAGG  
CCAUGAAGCGGUACGCCUGGUGCCGGGCACGAUCGCCUUCACCGACGCCCACAUUGAGG  
UCGACAUCACCUACGCGGAGUACUUCGAGAUGAGCGUGCGCCUGGCGGAGGCCAUGAAGC  
GGUACGGCCUGAACACCAACCACCGGAUCGUGGUGUGCUCGGAGAACAGCCUGCAGUUCU  
UCAUGCCGGUGCUGGGCGCCCUUCUUCGCGCGUGGCGUCGCCCGGCGAACGACAUCU  
ACAACGAGCGGGAGCUGCUGAACAGCAUGGGGAUCAGCCAGCCGACCGUGGUGUUCGUGA  
GCAAGAAGGGCCUGCAGAAGAUCUGAACGUGCAGAAGAAGCUGCCCAUCAUCCAGAAGA  
UCAUCAUCAUGGACAGCAAGACCGACUACAGGGCUUCCAGUCGAUGUACACGUUCGUGA  
CCAGCCACCUCGCCGGGCUUCAACGAGUACGACUUCGUCGCCGAGAGCUUCGACCGGG  
ACAAGACCAUCGCCUGAUCAUGAACAGCAGCGGCAGCACCGGCCUGCCGAAGGGGGUGG  
CCUGCCGCACCGGACCGCCUGCGUGCGCUUCUCGCACGCCCGGGACCCCAUCUUCGGCA  
ACCAGAUAUCCCGGACACCGCCAUCUGAGCGUGGUGCCGUUCCACCACGGCUUCGGCA  
UGUUCACGACCCUGGGCUACCUCUUCGCGGCUUCCGGGUGGUCCUGAUGUACCGGUUCG  
AGGAGGAGCUGUCCUGCGGAGCCUGCAGGACUACAAGAUCAGAGCGCGCUGCUCGUGC  
CGACCCUGUUCAGCUUCUUCGCCAAGAGCACCCUGAUCGACAAGUACGACCUGUCGAACC  
UGCACGAGAUCGCCAGCGGGGGCGCCCCGUGAGCAAGGAGGUGGGCGAGGCCGUGGCCA  
AGCGGUUCCACCUCGCCGGCAUCCGCCAGGGCUACGGCCUGACCGAGACCACGAGCGCGA  
UCCUGAUCACCCCGAGGGGGACGACAAGCCGGGCGCCGUGGGCAAGGUGGUCCCGUUCU  
UCGAGGCCAAGGUGGUGGACCUGGACACCGGCAAGACCCUGGGCGUGAACAGCGGGGCG  
AGCUGUGCGUGCGGGGGCCGAUGAUGAUGAGCGGCUACGUGAACAACCCGGAGGCCACCA  
ACGCCCUCUUCGACAAGGACGGCUGGCUGCACAGCGGCGACAUCGCCUACUGGGACGAGG  
ACGAGCACUUCUUCUUCGUCGACCGGCUGAAGUCGUGAUGAUGAACAAGGGCUACACAGG  
UGGCGCCGGCCGAGCUGGAGAGCAUCCUGCUCCAGCACCCCAACAUCUUCGACGCCGGCG  
UGGCGGGGCGUGCCGGACGACGACCGCGGCGAGCUGCCGGCCGCGGUGGUGGUGCUGGAGC  
ACGGCAAGACCAUGACGGAGAAGGAGAUCGUCGACUACGUGGCCAGCCAGGUGACCACCG  
CCAAGAAGCUGCGGGGGCGCGUGGUGUUCGUGGACGAGGUCGCGAAGGGCCUGACCGGGA  
AGCUCGACGCCCGGAAGAUCGCGGAGAUCUGAUGAAGGCCAAGAAGGGCGGCAAGAUCG  
CCGUGUAAGACUAGUGCAUCACAUUUAAAAGCAUCUCAGCCUACCAUGAGAAUAAGAGAA  
AGAAAAUGAAGAUCAAUAGCUUAUUAUCUCUUUUUUCUUUUUCGUUGGUGUAAAGCCAAC  
ACCCUGUCUAAAAAACAUAAAUUUCUUUAAUCAUUUUGCCUCUUUCUCUGUGCUUCAU  
UAAUAAAAAUGGAAAGAACCUAGAUCUAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA  
AAAUGCAUCCCCCCCCCCCCCCCCC  
CCCCCCCCAAAGGCUCUUUCAGAGCCACCAGAAU

U

图1B

图 10: 人工核酸 (SEQ ID NO: 211).

32L4 - PpLuc(GC) - [3'-UTR对应于 SEQ ID NO: 1] - A64 - C30 - hSL

GGGGCGCUGCCUACGGAGGUGGCAGCCAUCUCCUUCUCGGCAUCAAGCUUGAGG  
AUGGAGGACGCCAAGAACAUCAAGAAGGGCCCCGGCGCCCUUCUACCCGCUUGAG  
GACGGGACCGCCGGCGAGCAGCUCACAAAGGCCAUGAAGCGGUACGCCUUGGUG  
CCGGGCACGAUCGCCUUCACCGACGCCACAUUCGAGGUCGACAUACCUACGCG  
GAGUACUUCGAGAUGAGCGUGCGCCUGGCCGAGGCCAUGAAGCGGUACGGCCUG  
AACACCAACCACCGGAUCGUGGUGUGUCUGGAGAACAGCCUGCAGUUCUUC AUG  
CCGGUGCUGGGCGCCCUUCUUCUUCGCGUGGGCCUGCGCCCCGGCGAACGACAUC  
UACAACGAGCGGGAGCUGCUGAACAGCAUGGGGAUCAGCCAGCCGACCGUGGUG  
UUCGUGAGCAAGAAGGGGCCUGCAGAAGAUCUGAACGUGCAGAAGAAGCUGCCC  
AUCAUCCAGAAGAUAUCAUCAUGGACAGCAAGACCGACUACCAGGGCUUCCAG  
UCGAUGUACACGUUCGUGACCAGCCACCUCGCCCGGGCUUCAACGAGUACGAC  
UUCGUCCCGGAGAGCUUCGACCGGGACAAGACCAUCGCCCUGAUCAUGAACAGC  
AGCGGCAGCACCGGCCUGCCGAAGGGGGUGGGCCUGCCGCACCGGACCGCCUGC  
GUGCGCUUCUCGCACGCCCCGGGACCCCAUCUUCGGCAACCAGAUCAUCCGGAC  
ACCGCCAUCCUGAGCGUGGUGCCGUUCCACCACGGCUUCGGCAUGUUCACGACC  
CUGGGCUACCUAUCUGCGGCUUCCGGGUGGUCCUGAUGUACCGGUUCGAGGAG  
GAGCUGUUCUGCGGAGCCUGCAGGACUACAAGAUCAGAGCGCGCUGCUCGUG  
CCGACCCUGUUCAGCUUCUUCGCCAAGAGCACCCUGAUCGACAAGUACGACCUG  
UCGAACCUGCACGAGAUCCGACGCGGGGGCGCCCCGUGAGCAAGGAGGUGGGC  
GAGGCCGUGGCCAAGCGGUUCCACCUCGCCGGCAUCCGCCAGGGCUACGGCCUG  
ACCGAGACCACGAGCGGAUCCUGAUCACCCCCGAGGGGGACGACAAGCCGGGC  
GCCGUGGGCAAGGUGGUCCCGUUCUUCGAGGCCAAGGUGGUGGACCUGGACACC  
GGCAAGACCCUGGGCGUGAACAGCGGGGCGAGCUGUGCGUGCGGGGGCCGAUG  
AUCAUGAGCGGCUACGUGAACAACCCGGAGGCCACCAACGCCCUCUACGACAAG  
GACGGCUGGCUGCACAGCGGCGACAUCGCCUACUGGGACGAGGACGAGCACUUC  
UUCAUCGUGACCGGCUGAAGUCGUGAUAAGUACAAGGGCUACCAGGUGGCG  
CCGGCCGAGCUGGAGAGCAUCCUGCUCCAGCACCCCAACAUCUUCGACGCCGGC  
GUGGCCGGGCGUGCCGGACGACGACGCCGGCGAGCUGCCGGCCGCGGUGGUGGUG  
CUGGAGCACGGCAAGACCAUGACGGAGAAGGAGAUCGUCGACUACGUGGCCAGC  
CAGGUGACCACCGCCAAGAAGCUGCGGGGCGGCGUGGUGUUCGUGGACGAGGUC  
CCGAAGGGCCUGACCGGGAAGCUCGACGCCCGGAAGAUCGCGGAGAUCCUGAUC  
AAGGCCAAGAAGGGCGGCAAGAUCGCCGUGUAAGACUAGUCGCGCGCCGCCCGCG  
CCGCCGGAGCCUGUCGCCGUCUGUCCCAAGCCUGCUUGUGUCCCGUGAGGUUG  
UCAUAUAACCUGCCUCGGGCGGCCGCCUCCCAGAUCUAAAAAAAAAAAAAAAAA  
AAAUGCAUC  
CCCCCCCCCCCCCCCCCCCCCCCCCCCCCAAGGCUCUUUCAGAGCCACCAG  
AAUU

图10

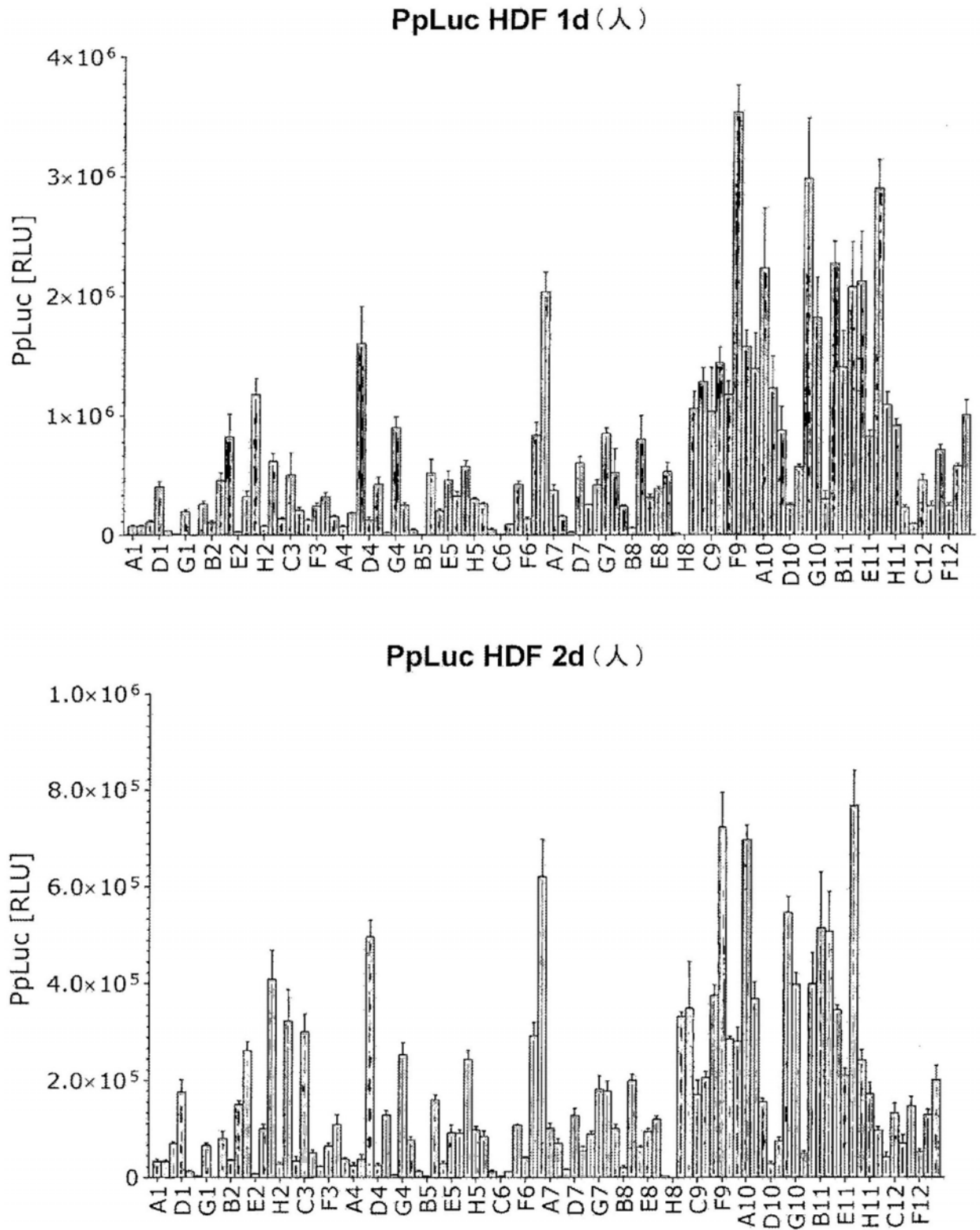


图2



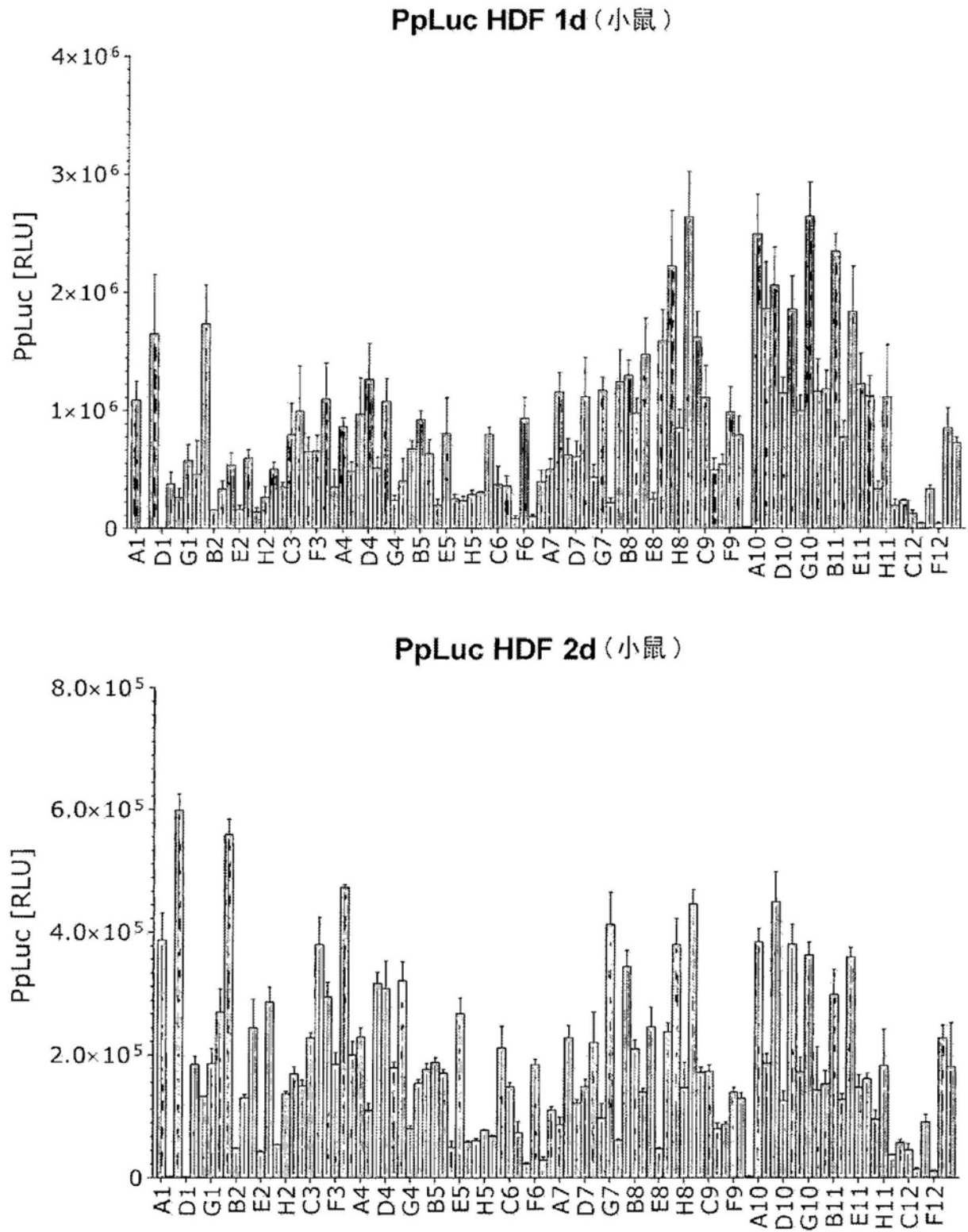


图3

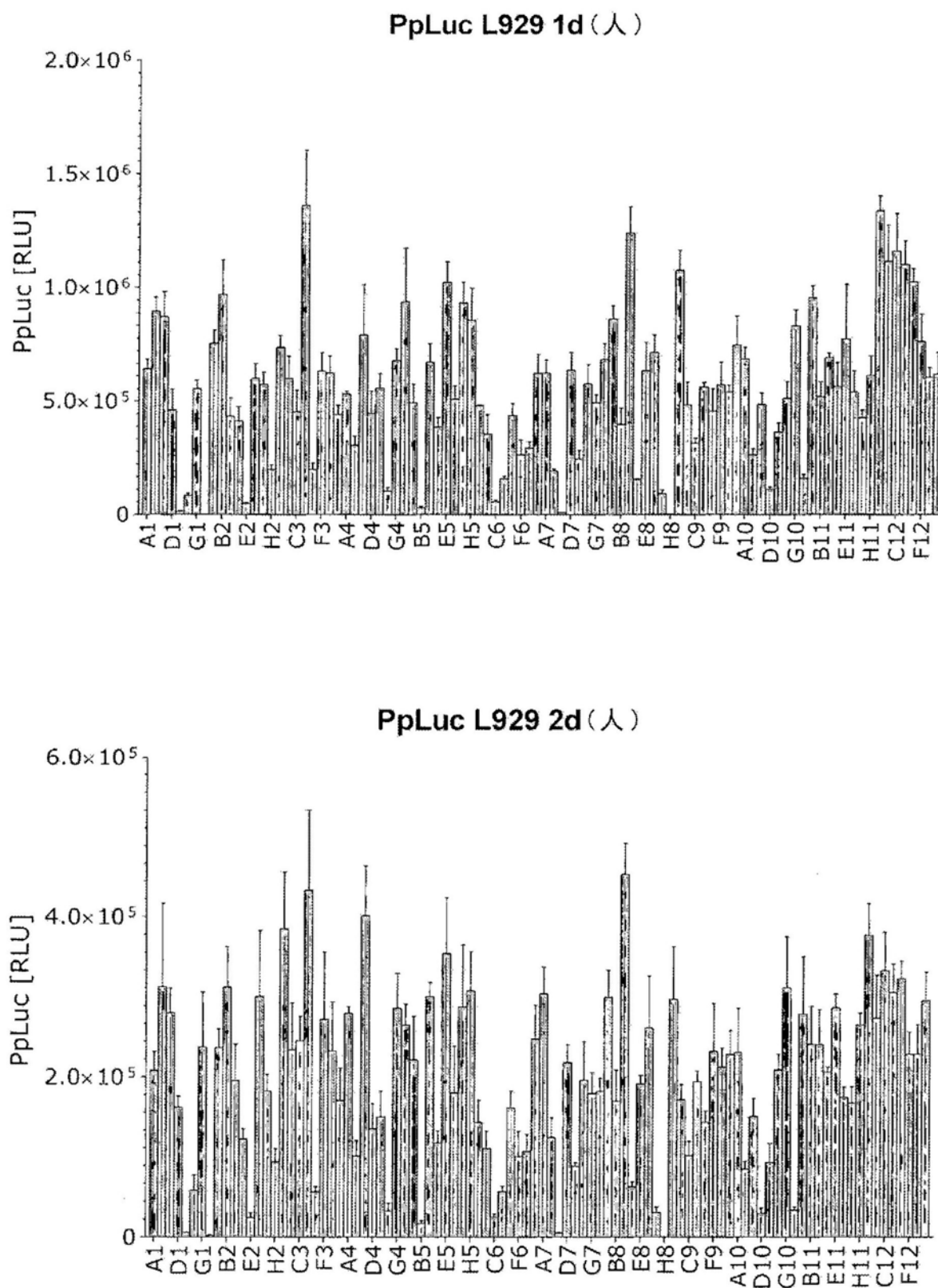


图4



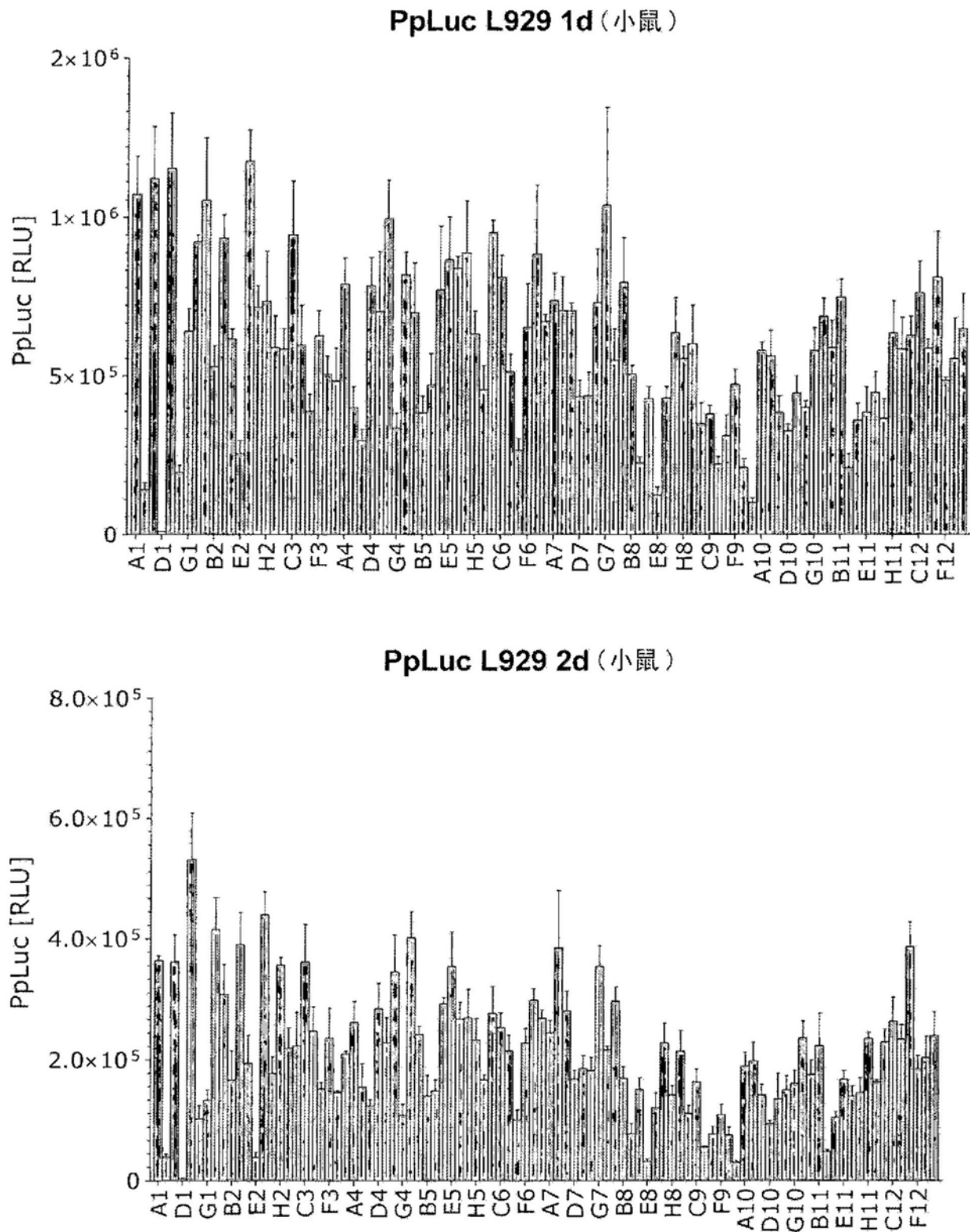


图5

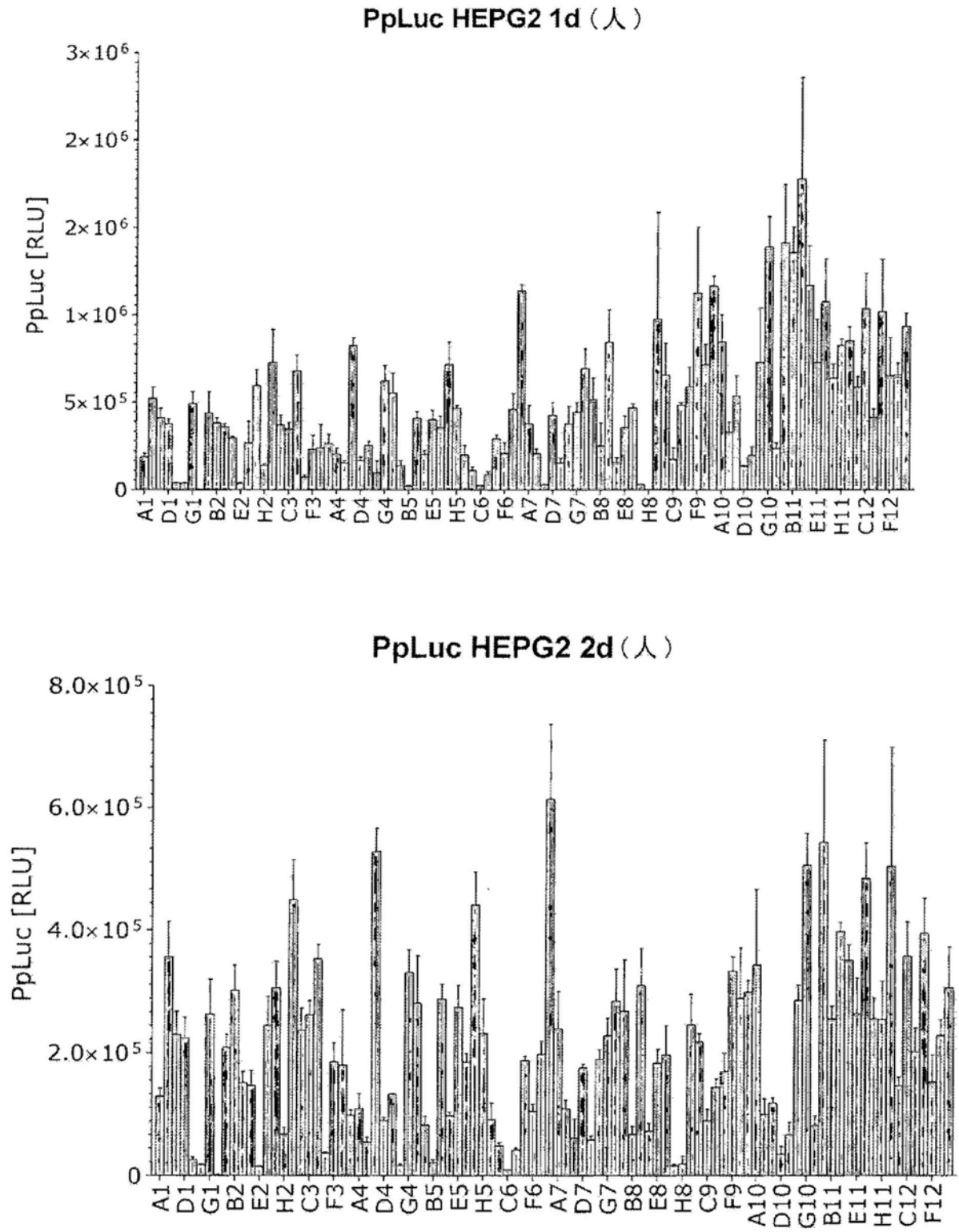


图6

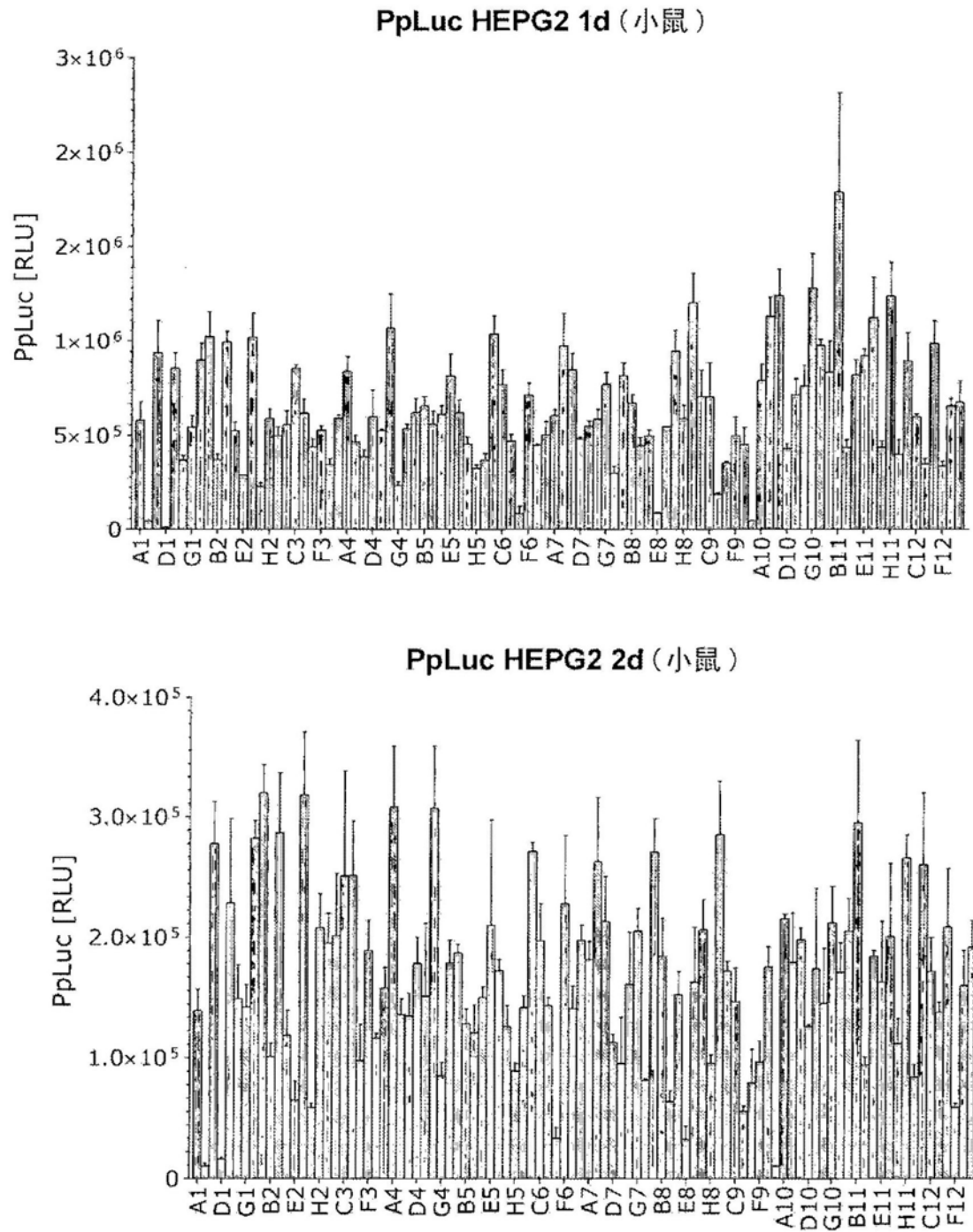


图7

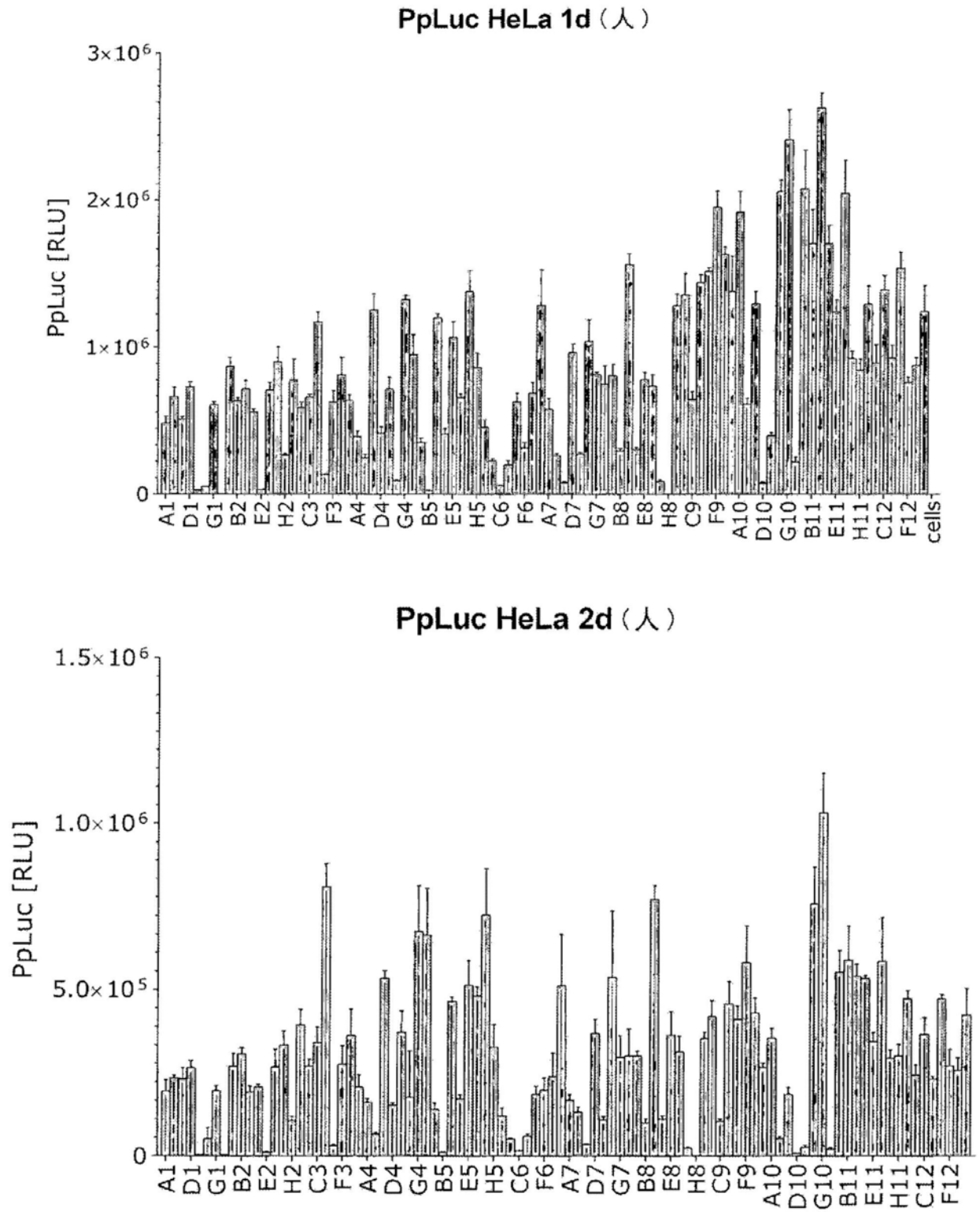


图8

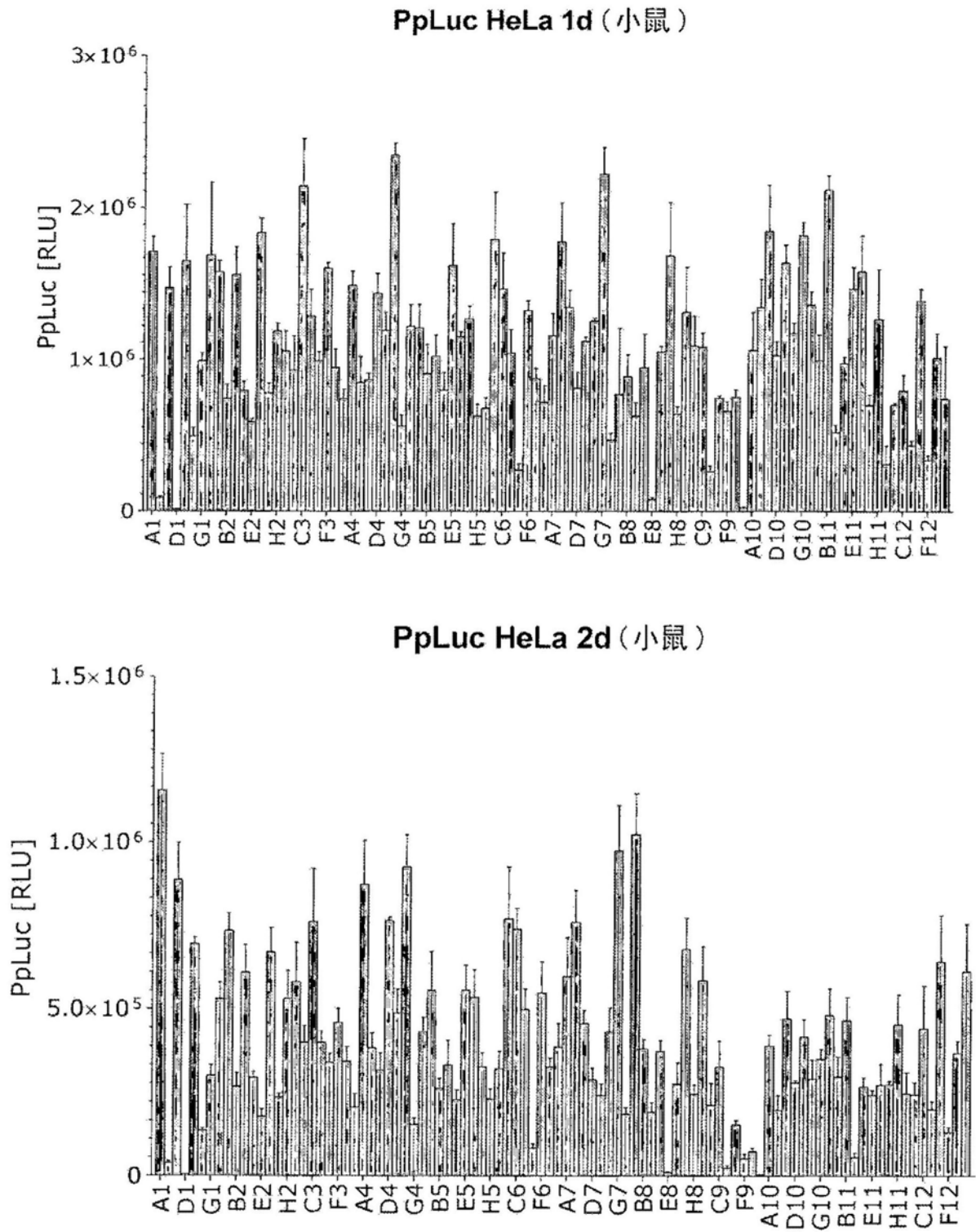


图9