



(12) 发明专利申请



(10) 申请公布号 CN 114502575 A

(43) 申请公布日 2022. 05. 13

(21) 申请号 202080054069.6

(22) 申请日 2020.06.09

(30) 优先权数据

62/859,539 2019.06.10 US

62/866,374 2019.06.25 US

62/915,523 2019.10.15 US

62/960,487 2020.01.13 US

62/987,858 2020.03.10 US

63/010,970 2020.04.16 US

(85) PCT国际申请进入国家阶段日

2022.01.26

(86) PCT国际申请的申请数据

PCT/US2020/036846 2020.06.09

(87) PCT国际申请的公布数据

W02020/251954 EN 2020.12.17

(71) 申请人 同源药物公司

地址 美国马萨诸塞州

(72) 发明人 T.B.圣马丁 A.B.西摩尔 H.鲁宾

(74) 专利代理机构 北京市柳沈律师事务所

11105

专利代理师 张文辉

(51) Int.Cl.

C07K 14/47 (2006.01)

A61K 48/00 (2006.01)

A61K 45/06 (2006.01)

A61P 25/28 (2006.01)

C12N 15/86 (2006.01)

C12N 7/00 (2006.01)

权利要求书9页 说明书37页

序列表115页 附图30页

(54) 发明名称

用于ARSA基因转移的腺相关病毒组合物和其使用方法

(57) 摘要

本文提供一种腺相关病毒 (AAV) 组合物,其可在细胞中表达芳基硫酸酯酶A (ARSA) 多肽,由此恢复ARSA基因功能。还提供使用所述AAV组合物的方法,和用于制备所述AAV组合物的包装系统。

1. 一种用于在细胞中表达芳基硫酸酯酶A (ARSA) 多肽的方法, 所述方法包含用重组腺相关病毒 (rAAV) 转导所述细胞, 所述重组腺相关病毒包含:

(a) AAV衣壳, 其包含AAV衣壳蛋白; 和

(b) 转移基因组, 其包含可操作地连接于以沉默方式改变的ARSA编码序列的转录调节元件。

2. 根据权利要求1所述的方法, 其中所述细胞是神经元和/或神经胶质细胞, 任选地其中所述细胞是中枢神经系统和/或周围神经系统的神经元和/或神经胶质细胞。

3. 根据权利要求1所述的方法, 其中所述细胞为选自由以下组成的群组的中枢神经系统区域的细胞: 脊髓、运动皮质、感觉皮质、海马、壳核、小脑, 任选地小脑核, 和其任何组合。

4. 根据权利要求1所述的方法, 其中所述细胞为选自由以下组成的群组的细胞: 运动神经元、星形胶质细胞、少突胶质细胞、中枢神经系统中的大脑皮质的细胞、周围神经系统的感觉神经元、施万细胞 (Schwann cell) 和其任何组合。

5. 根据权利要求1至4中任一项所述的方法, 其中所述细胞在哺乳动物受试者中并且以有效转导所述受试者中的所述细胞的量向所述受试者施用所述AAV。

6. 一种用于治疗患有异染性脑白质营养不良 (MLD) 的受试者的方法, 所述方法包含向所述受试者施用有效量的rAAV, 所述rAAV包含:

(a) AAV衣壳, 其包含衣壳蛋白; 和

(b) 转移基因组, 其包含可操作地连接于以沉默方式改变的ARSA编码序列的转录调节元件。

7. 根据前述权利要求中任一项所述的方法, 其中所述以沉默方式改变的ARSA编码序列编码SEQ ID NO:23中所阐述的氨基酸序列。

8. 根据权利要求7所述的方法, 其中所述以沉默方式改变的ARSA编码序列包含SEQ ID NO:14、62或72中所阐述的核苷酸序列。

9. 根据前述权利要求中任一项所述的方法, 其中所述转录调节元件包含选自由以下组成的群组的一种或多种元件: 巨细胞病毒 (CMV) 增强子元件、鸡 β -肌动蛋白 (CBA) 启动子、小型鸡 β -肌动蛋白 (SmCBA) 启动子、钙调蛋白1 (CALM1) 启动子、蛋白脂质蛋白质1 (PLP1) 启动子、胶质细胞原纤维酸性蛋白 (GFAP) 启动子、突触蛋白2 (SYN2) 启动子、金属硫蛋白3 (MT3) 启动子和其任何组合。

10. 根据前述权利要求中任一项所述的方法, 其中所述转录调节元件包含与选自由以下组成的群组的序列至少90%一致的核苷酸序列: SEQ ID NO:25、32、36、54、55和58。

11. 根据前述权利要求中任一项所述的方法, 其中所述转录调节元件包含选自由以下组成的群组的核苷酸序列: SEQ ID NO:25、32、36、54、55和58。

12. 根据前述权利要求中任一项所述的方法, 其中所述转录调节元件从5'到3'包含SEQ ID NO:58、25和32中所阐述的核苷酸序列。

13. 根据前述权利要求中任一项所述的方法, 其中所述转录调节元件包含SEQ ID NO:36中所阐述的核苷酸序列。

14. 根据前述权利要求中任一项所述的方法, 其中所述转移基因组还包含位于所述以沉默方式改变的ARSA编码序列3'的聚腺苷酸化序列。

15. 根据权利要求14所述的方法, 其中所述聚腺苷酸化序列为外源性聚腺苷酸化序列。

16. 根据权利要求15所述的方法, 其中所述外源性聚腺苷酸化序列为SV40聚腺苷酸化序列。

17. 根据权利要求16所述的方法, 其中所述SV40聚腺苷酸化序列包含SEQ ID NO: 42中所阐述的核苷酸序列。

18. 根据前述权利要求中任一项所述的方法, 其中所述转移基因组还包含填充序列。

19. 根据前述权利要求中任一项所述的方法, 其中所述转移基因组还包含位于所述以沉默方式改变的ARSA编码序列3'的填充序列。

20. 根据权利要求18或19所述的方法, 其中所述填充序列位于所述聚腺苷酸化序列3'。

21. 根据前述权利要求中任一项所述的方法, 其中所述转移基因组包含选自以下组成的群组的序列: SEQ ID NO: 41、44、46、65、67、75和79。

22. 根据前述权利要求中任一项所述的方法, 其中所述转移基因组还包含所述基因组5'的5'反向末端重复(5' ITR)核苷酸序列, 和所述基因组3'的3'反向末端重复(3' ITR)核苷酸序列。

23. 根据权利要求22所述的方法, 其中所述5' ITR核苷酸序列与SEQ ID NO: 18具有至少95%序列一致性, 并且所述3' ITR核苷酸序列与SEQ ID NO: 19具有至少95%序列一致性。

24. 根据权利要求22所述的方法, 其中所述5' ITR核苷酸序列与SEQ ID NO: 26具有至少95%序列一致性, 并且所述3' ITR核苷酸序列与SEQ ID NO: 27具有至少95%序列一致性。

25. 根据权利要求22所述的方法, 其中所述5' ITR核苷酸序列与SEQ ID NO: 18具有至少95%序列一致性, 并且所述3' ITR核苷酸序列与SEQ ID NO: 57具有至少95%序列一致性。

26. 根据前述权利要求中任一项所述的方法, 其中所述转移基因组包含选自以下组成的群组的核苷酸序列: SEQ ID NO: 47、48、49、68、69、76和80。

27. 根据权利要求5至25中任一项所述的方法, 其中异染性脑白质营养不良与芳基硫酸酯酶A (ARSA) 基因突变相关。

28. 根据权利要求6至27中任一项所述的方法, 其中所述受试者是人类受试者。

29. 根据前述权利要求中任一项所述的方法, 其中所述衣壳蛋白包含与SEQ ID NO: 2、3、4、6、7、10、11、12、13、15、16或17的氨基酸203-736的氨基酸序列具有至少95%序列一致性的氨基酸序列。

30. 根据权利要求29所述的方法, 其中: 所述衣壳蛋白中对应于SEQ ID NO: 16的氨基酸206的氨基酸为C; 所述衣壳蛋白中对应于SEQ ID NO: 16的氨基酸296的氨基酸为H; 所述衣壳蛋白中对应于SEQ ID NO: 16的氨基酸312的氨基酸为Q; 所述衣壳蛋白中对应于SEQ ID NO: 16的氨基酸346的氨基酸为A; 所述衣壳蛋白中对应于SEQ ID NO: 16的氨基酸464的氨基酸为N; 所述衣壳蛋白中对应于SEQ ID NO: 16的氨基酸468的氨基酸为S; 所述衣壳蛋白中对应于SEQ ID NO: 16的氨基酸501的氨基酸为I; 所述衣壳蛋白中对应于SEQ ID NO: 16的氨基酸505的氨基酸为R; 所述衣壳蛋白中对应于SEQ ID NO: 16的氨基酸590的氨基酸为R; 所述衣壳蛋白中对应于SEQ ID NO: 16的氨基酸626的氨基酸为G或Y; 所述衣壳蛋白中对应于SEQ ID NO: 16的氨基酸681的氨基酸为M; 所述衣壳蛋白中对应于SEQ ID NO: 16的氨基酸687的氨基酸为R; 所述衣壳蛋白中对应于SEQ ID NO: 16的氨基酸690的氨基酸为K; 所述衣壳蛋白中对应于SEQ ID NO: 16的氨基酸706的氨基酸为C; 或所述衣壳蛋白中对应于SEQ ID NO: 16的氨基酸718的氨基酸为G。

31. 根据权利要求30所述的方法, 其中:

(a) 所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸626的氨基酸为G, 并且所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸718的氨基酸为G;

(b) 所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸296的氨基酸为H, 所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸464的氨基酸为N, 所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸505的氨基酸为R, 并且所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸681的氨基酸为M;

(c) 所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸505的氨基酸为R, 并且所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸687的氨基酸为R;

(d) 所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸346的氨基酸为A, 并且所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸505的氨基酸为R; 或

(e) 所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸501的氨基酸为I, 所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸505的氨基酸为R, 并且所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸706的氨基酸为C。

32. 根据权利要求30所述的方法, 其中所述衣壳蛋白包含SEQ ID NO:2、3、4、6、7、10、11、12、13、15、16或17的氨基酸203-736的氨基酸序列。

33. 根据前述权利要求中任一项所述的方法, 其中所述衣壳蛋白包含与SEQ ID NO:2、3、4、5、6、7、9、10、11、12、13、15、16或17的氨基酸138-736的氨基酸序列具有至少95%序列一致性的氨基酸序列。

34. 根据权利要求33所述的方法, 其中: 所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸151的氨基酸为R; 所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸160的氨基酸为D; 所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸206的氨基酸为C; 所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸296的氨基酸为H; 所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸312的氨基酸为Q; 所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸346的氨基酸为A; 所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸464的氨基酸为N; 所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸468的氨基酸为S; 所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸501的氨基酸为I; 所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸505的氨基酸为R; 所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸590的氨基酸为R; 所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸626的氨基酸为G或Y; 所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸681的氨基酸为M; 所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸687的氨基酸为R; 所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸690的氨基酸为K; 所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸706的氨基酸为C; 或所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸718的氨基酸为G。

35. 根据权利要求34所述的方法, 其中:

(a) 所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸626的氨基酸为G, 并且所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸718的氨基酸为G;

(b) 所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸296的氨基酸为H, 所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸464的氨基酸为N, 所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸505的氨基酸为R, 并且所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸681的氨基酸为M;

(c) 所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸505的氨基酸为R,并且所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸687的氨基酸为R;

(d) 所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸346的氨基酸为A,并且所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸505的氨基酸为R;或

(e) 所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸501的氨基酸为I,所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸505的氨基酸为R,并且所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸706的氨基酸为C。

36. 根据权利要求34所述的方法,其中所述衣壳蛋白包含SEQ ID NO:2、3、4、5、6、7、9、10、11、12、13、15、16或17的氨基酸138-736的氨基酸序列。

37. 根据前述权利要求中任一项所述的方法,其中所述衣壳蛋白包含与SEQ ID NO:2、3、4、5、6、7、8、9、10、11、12、13、15、16或17的氨基酸1-736的氨基酸序列具有至少95%序列一致性的氨基酸序列。

38. 根据权利要求37所述的方法,其中:所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸2的氨基酸为T;所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸65的氨基酸为I;所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸68的氨基酸为V;所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸77的氨基酸为R;所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸119的氨基酸为L;所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸151的氨基酸为R;所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸160的氨基酸为D;所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸206的氨基酸为C;所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸296的氨基酸为H;所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸312的氨基酸为Q;所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸346的氨基酸为A;所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸464的氨基酸为N;所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸468的氨基酸为S;所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸501的氨基酸为I;所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸505的氨基酸为R;所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸590的氨基酸为R;所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸626的氨基酸为G或Y;所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸681的氨基酸为M;所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸687的氨基酸为R;所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸690的氨基酸为K;所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸706的氨基酸为C;或所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸718的氨基酸为G。

39. 根据权利要求38所述的方法,其中:

(a) 所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸2的氨基酸为T,并且所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸312的氨基酸为Q;

(b) 所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸65的氨基酸为I,并且所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸626的氨基酸为Y;

(c) 所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸77的氨基酸为R,并且所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸690的氨基酸为K;

(d) 所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸119的氨基酸为L,并且所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸468的氨基酸为S;

(e) 所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸626的氨基酸为G,并且所述衣壳蛋

白中对应于SEQ ID NO:16的氨基酸718的氨基酸为G;

(f) 所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸296的氨基酸为H,所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸464的氨基酸为N,所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸505的氨基酸为R,并且所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸681的氨基酸为M;

(g) 所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸505的氨基酸为R,并且所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸687的氨基酸为R;

(h) 所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸346的氨基酸为A,并且所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸505的氨基酸为R;或

(i) 所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸501的氨基酸为I,所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸505的氨基酸为R,并且所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸706的氨基酸为C。

40. 根据权利要求38所述的方法,其中所述衣壳蛋白包含SEQ ID NO:2、3、4、5、6、7、8、9、10、11、12、13、15、16或17的氨基酸1-736的氨基酸序列。

41. 一种rAAV,其包含:

(a) AAV衣壳,其包含AAV衣壳蛋白;和

(b) 转移基因组,其包含可操作地连接于以沉默方式改变的ARSA编码序列的转录调节元件。

42. 根据权利要求41所述的rAAV,其中所述以沉默方式改变的ARSA编码序列编码SEQ ID NO:23中所阐述的氨基酸序列。

43. 根据权利要求42所述的rAAV,其中所述以沉默方式改变的ARSA编码序列包含SEQ ID NO:14中所阐述的核苷酸序列。

44. 根据权利要求42所述的rAAV,其中所述以沉默方式改变的ARSA编码序列包含SEQ ID NO:62或72中所阐述的核苷酸序列。

45. 根据权利要求41至44中任一项所述的rAAV,其中所述转录调节元件包含选自由以下组成的群组的一种或多种元件:巨细胞病毒(CMV)增强子元件、鸡 β -肌动蛋白(CBA)启动子、小型鸡 β -肌动蛋白(SmCBA)启动子、钙调蛋白1(CALM1)启动子、蛋白脂质蛋白质1(PLP1)启动子、胶质细胞原纤维酸性蛋白(GFAP)启动子、突触蛋白2(SYN2)启动子、金属硫蛋白3(MT3)启动子及其任何组合。

46. 根据权利要求45所述的rAAV,其中所述转录调节元件包含与选自由以下组成的群组的序列至少90%一致的核苷酸序列:SEQ ID NO:25、32、36、54、55和58。

47. 根据权利要求45所述的rAAV,其中所述转录调节元件包含选自由以下组成的群组的核苷酸序列:SEQ ID NO:25、32、36、54、55和58。

48. 根据权利要求45所述的rAAV,其中所述转录调节元件从5'到3'包含SEQ ID NO:58、25和32中所阐述的核苷酸序列。

49. 根据权利要求45所述的rAAV,其中所述转录调节元件包含SEQ ID NO:36中所阐述的核苷酸序列。

50. 根据权利要求41至49中任一项所述的rAAV,其中所述转移基因组还包含位于所述以沉默方式改变的ARSA编码序列3'的聚腺苷酸化序列。

51. 根据权利要求50所述的rAAV,其中所述聚腺苷酸化序列是外源性聚腺苷酸化序列。

52. 根据权利要求51所述的rAAV,其中所述外源性聚腺苷酸化序列是SV40聚腺苷酸化序列。

53. 根据权利要求52所述的rAAV,其中所述SV40聚腺苷酸化序列包含SEQ ID NO:42中所阐述的核苷酸序列。

54. 根据权利要求41至53中任一项所述的rAAV,其中所述转移基因组包含选自由以下组成的群组的序列:SEQ ID NO:41、44、46、65、67、75和79。

55. 根据权利要求41至54中任一项所述的rAAV,其中所述转移基因组还包含所述基因组5'的5'反向末端重复(5' ITR)核苷酸序列,和所述基因组3'的3'反向末端重复(3' ITR)核苷酸序列。

56. 根据权利要求55所述的rAAV,其中所述5' ITR核苷酸序列与SEQ ID NO:18具有至少95%序列一致性,并且所述3' ITR核苷酸序列与SEQ ID NO:19具有至少95%序列一致性。

57. 根据权利要求41至55中任一项所述的rAAV,其中所述转移基因组包含选自由以下组成的群组的核苷酸序列:SEQ ID NO:47、48、49、68、69、76和80。

58. 根据权利要求41至55中任一项所述的rAAV,其中所述转移基因组的核苷酸序列由选自由以下组成的群组的核苷酸序列组成:SEQ ID NO:47、48、49、68、69、76和80。

59. 根据权利要求41至58中任一项所述的rAAV,其中所述转移基因组的核苷酸序列由SEQ ID NO:48中所阐述的核苷酸序列组成。

60. 根据权利要求41至59中任一项所述的rAAV,其中所述衣壳蛋白包含与SEQ ID NO:2、3、4、6、7、10、11、12、13、15、16或17的氨基酸203-736的氨基酸序列具有至少95%序列一致性的氨基酸序列。

61. 根据权利要求60所述的rAAV,其中:所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸206的氨基酸为C;所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸296的氨基酸为H;所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸312的氨基酸为Q;所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸346的氨基酸为A;所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸464的氨基酸为N;所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸468的氨基酸为S;所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸501的氨基酸为I;所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸505的氨基酸为R;所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸590的氨基酸为R;所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸626的氨基酸为G或Y;所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸681的氨基酸为M;所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸687的氨基酸为R;所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸690的氨基酸为K;所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸706的氨基酸为C;或所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸718的氨基酸为G。

62. 根据权利要求61所述的rAAV,其中:

(a) 所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸626的氨基酸为G,并且所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸718的氨基酸为G;

(b) 所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸296的氨基酸为H,所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸464的氨基酸为N,所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸505的氨基酸为R,并且所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸681的氨基酸为

M;

(c) 所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸505的氨基酸为R,并且所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸687的氨基酸为R;

(d) 所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸346的氨基酸为A,并且所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸505的氨基酸为R;或

(e) 所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸501的氨基酸为I,所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸505的氨基酸为R,并且所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸706的氨基酸为C。

63. 根据权利要求61所述的rAAV,其中所述衣壳蛋白包含SEQ ID NO:2、3、4、6、7、10、11、12、13、15、16或17的氨基酸203-736的氨基酸序列。

64. 根据权利要求41至63中任一项所述的rAAV,其中所述衣壳蛋白包含与SEQ ID NO:2、3、4、5、6、7、9、10、11、12、13、15、16或17的氨基酸138-736的氨基酸序列具有至少95%序列一致性的氨基酸序列。

65. 根据权利要求64所述的rAAV,其中:所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸151的氨基酸为R;所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸160的氨基酸为D;所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸206的氨基酸为C;所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸296的氨基酸为H;所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸312的氨基酸为Q;所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸346的氨基酸为A;所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸464的氨基酸为N;所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸468的氨基酸为S;所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸501的氨基酸为I;所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸505的氨基酸为R;所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸590的氨基酸为R;所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸626的氨基酸为G或Y;所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸681的氨基酸为M;所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸687的氨基酸为R;所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸690的氨基酸为K;所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸706的氨基酸为C;或所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸718的氨基酸为G。

66. 根据权利要求65所述的rAAV,其中:

(a) 所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸626的氨基酸为G,并且所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸718的氨基酸为G;

(b) 所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸296的氨基酸为H,所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸464的氨基酸为N,所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸505的氨基酸为R,并且所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸681的氨基酸为M;

(c) 所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸505的氨基酸为R,并且所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸687的氨基酸为R;

(d) 所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸346的氨基酸为A,并且所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸505的氨基酸为R;或

(e) 所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸501的氨基酸为I,所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸505的氨基酸为R,并且所述衣壳蛋白中对应于SEQ ID NO:16

的氨基酸706的氨基酸为C。

67. 根据权利要求65所述的rAAV, 其中所述衣壳蛋白包含SEQ ID NO: 2、3、4、5、6、7、9、10、11、12、13、15、16或17的氨基酸138-736的氨基酸序列。

68. 根据权利要求41至67中任一项所述的rAAV, 其中所述衣壳蛋白包含与SEQ ID NO: 2、3、4、5、6、7、8、9、10、11、12、13、15、16或17的氨基酸1-736的氨基酸序列具有至少95%序列一致性的氨基酸序列。

69. 根据权利要求68所述的rAAV, 其中: 所述衣壳蛋白中对应于SEQ ID NO: 16的氨基酸2的氨基酸为T; 所述衣壳蛋白中对应于SEQ ID NO: 16的氨基酸65的氨基酸为I; 所述衣壳蛋白中对应于SEQ ID NO: 16的氨基酸68的氨基酸为V; 所述衣壳蛋白中对应于SEQ ID NO: 16的氨基酸77的氨基酸为R; 所述衣壳蛋白中对应于SEQ ID NO: 16的氨基酸119的氨基酸为L; 所述衣壳蛋白中对应于SEQ ID NO: 16的氨基酸151的氨基酸为R; 所述衣壳蛋白中对应于SEQ ID NO: 16的氨基酸160的氨基酸为D; 所述衣壳蛋白中对应于SEQ ID NO: 16的氨基酸206的氨基酸为C; 所述衣壳蛋白中对应于SEQ ID NO: 16的氨基酸296的氨基酸为H; 所述衣壳蛋白中对应于SEQ ID NO: 16的氨基酸312的氨基酸为Q; 所述衣壳蛋白中对应于SEQ ID NO: 16的氨基酸346的氨基酸为A; 所述衣壳蛋白中对应于SEQ ID NO: 16的氨基酸464的氨基酸为N; 所述衣壳蛋白中对应于SEQ ID NO: 16的氨基酸468的氨基酸为S; 所述衣壳蛋白中对应于SEQ ID NO: 16的氨基酸501的氨基酸为I; 所述衣壳蛋白中对应于SEQ ID NO: 16的氨基酸505的氨基酸为R; 所述衣壳蛋白中对应于SEQ ID NO: 16的氨基酸590的氨基酸为R; 所述衣壳蛋白中对应于SEQ ID NO: 16的氨基酸626的氨基酸为G或Y; 所述衣壳蛋白中对应于SEQ ID NO: 16的氨基酸681的氨基酸为M; 所述衣壳蛋白中对应于SEQ ID NO: 16的氨基酸687的氨基酸为R; 所述衣壳蛋白中对应于SEQ ID NO: 16的氨基酸690的氨基酸为K; 所述衣壳蛋白中对应于SEQ ID NO: 16的氨基酸706的氨基酸为C; 或所述衣壳蛋白中对应于SEQ ID NO: 16的氨基酸718的氨基酸为G。

70. 根据权利要求69所述的rAAV, 其中:

(a) 所述衣壳蛋白中对应于SEQ ID NO: 16的氨基酸2的氨基酸为T, 并且所述衣壳蛋白中对应于SEQ ID NO: 16的氨基酸312的氨基酸为Q;

(b) 所述衣壳蛋白中对应于SEQ ID NO: 16的氨基酸65的氨基酸为I, 并且所述衣壳蛋白中对应于SEQ ID NO: 16的氨基酸626的氨基酸为Y;

(c) 所述衣壳蛋白中对应于SEQ ID NO: 16的氨基酸77的氨基酸为R, 并且所述衣壳蛋白中对应于SEQ ID NO: 16的氨基酸690的氨基酸为K;

(d) 所述衣壳蛋白中对应于SEQ ID NO: 16的氨基酸119的氨基酸为L, 并且所述衣壳蛋白中对应于SEQ ID NO: 16的氨基酸468的氨基酸为S;

(e) 所述衣壳蛋白中对应于SEQ ID NO: 16的氨基酸626的氨基酸为G, 并且所述衣壳蛋白中对应于SEQ ID NO: 16的氨基酸718的氨基酸为G;

(f) 所述衣壳蛋白中对应于SEQ ID NO: 16的氨基酸296的氨基酸为H, 所述衣壳蛋白中对应于SEQ ID NO: 16的氨基酸464的氨基酸为N, 所述衣壳蛋白中对应于SEQ ID NO: 16的氨基酸505的氨基酸为R, 并且所述衣壳蛋白中对应于SEQ ID NO: 16的氨基酸681的氨基酸为M;

(g) 所述衣壳蛋白中对应于SEQ ID NO: 16的氨基酸505的氨基酸为R, 并且所述衣壳蛋

白中对应于SEQ ID NO:16的氨基酸687的氨基酸为R;

(h) 所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸346的氨基酸为A,并且所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸505的氨基酸为R;或

(i) 所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸501的氨基酸为I,所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸505的氨基酸为R,并且所述衣壳蛋白中对应于SEQ ID NO:16的氨基酸706的氨基酸为C。

71. 根据权利要求69所述的rAAV,其中所述衣壳蛋白包含SEQ ID NO:2、3、4、5、6、7、8、9、10、11、12、13、15、16或17的氨基酸1-736的氨基酸序列。

72. 一种药物组合物,其包含根据权利要求41至71中任一项所述的rAAV。

73. 一种多核苷酸,其包含SEQ ID NO:14、62和72中所阐述的核酸序列。

74. 一种用于制备rAAV的包装系统,其中所述包装系统包含

(a) 第一核苷酸序列,其编码一种或多种AAV Rep蛋白;

(b) 第二核苷酸序列,其编码根据权利要求41至71中任一项所述的AAV的衣壳蛋白;和

(c) 第三核苷酸序列,其包含根据权利要求41至71中任一项所述的AAV的rAAV基因组序列。

75. 根据权利要求74所述的包装系统,其中所述包装系统包含有包含所述第一核苷酸序列和所述第二核苷酸序列的第一载体,以及包含所述第三核苷酸序列的第二载体。

76. 根据权利要求74或75所述的包装系统,其还包含第四核苷酸序列,所述第四核苷酸序列包含一种或多种辅助病毒基因。

77. 根据权利要求76所述的包装系统,其中所述第四核苷酸序列包含于第三载体内。

78. 根据权利要求74至77中任一项所述的包装系统,其中所述第四核苷酸序列包含一种或多种来自选自以下组成的群组的病毒的基因:腺病毒、疱疹病毒、痘苗病毒和巨细胞病毒(CMV)。

79. 根据权利要求74至78中任一项所述的包装系统,其中所述第一载体、第二载体和/或所述第三载体为质粒。

80. 一种用于rAAV的重组制备的方法,所述方法包含在借以产生所述rAAV的条件下将根据权利要求74至79中任一项所述的包装系统引入细胞中。

81. 根据权利要求41至71中任一项所述的rAAV、根据权利要求72所述的药物组合物或根据权利要求73所述的多核苷酸,其用作药剂。

82. 根据权利要求41至71中任一项所述的rAAV、根据权利要求72所述的药物组合物或根据权利要求73所述的多核苷酸,其用于治疗MLD。

83. 根据权利要求41至71中任一项所述的rAAV、根据权利要求72所述的药物组合物或根据权利要求73所述的多核苷酸,其用于治疗患有MLD的受试者的方法,所述方法包含向所述受试者施用有效量的所述rAAV、所述药物组合物或所述多核苷酸。

用于ARSA基因转移的腺相关病毒组合物和其使用方法

[0001] 相关申请的交叉引用

[0002] 本申请案要求2019年6月10日提交的美国临时申请案第62/859,539号、2019年6月25日提交的美国临时申请案第62/866,374号、2019年10月15日提交的美国临时申请案第62/915,523号、2020年1月13日提交的美国临时申请案第62/960,487号、2020年3月10日提交的美国临时申请案第62/987,858号及2020年4月16日提交的美国临时申请案第63/010,970号的权益,其各自以全文引用的方式并入本文中。

[0003] 序列表

[0004] 本申请案含有序列表,其已以ASCII格式、以电子方式提交并以全文引用的方式并入本文中(所述ASCII副本创建于2020年4月16日,命名为“705151_HMW-030-6_ST25.txt”并且大小为295,995个字节)。

技术领域

背景技术

[0005] 异染性脑白质营养不良 (Metachromatic leukodystrophy, MLD) 是一种具有较高的未满足的医学需求的致死性溶酶体贮积症。此神经退化性疾病以三种形式(婴儿晚期、青少年及成年人)存在并且由溶酶体酶芳基硫酸酯酶-A (ARSA) 不足引起。ARSA位于称为溶酶体的细胞结构中,其在溶酶体中有助于分解硫苷脂。此酶的不足可引起脑部、脊髓和周边器官中硫苷脂的大量积聚,这导致髓鞘的严重损伤,髓鞘是神经纤维的主要保护层髓鞘产生细胞中的硫苷脂积聚引起整个神经系统(包括脑部、脊髓以及连接脑部及脊髓与肌肉及检测感觉(如触觉、疼痛、热及声音)的感觉细胞的神经)中的白质的进行性损坏。因此,MLD的特征在于中枢神经系统并且接着周围神经系统的进行性轴突髓鞘脱失。这引起获得性功能和/或技能的丧失、张力减退、共济失调、癫痫、失明、听力丧失和过早地死亡。

[0006] 在患有异染性脑白质营养不良的人中,白质损伤引起心智功能及运动技能,如行走能力的进行性退化。受影响的个体还出现四肢感觉丧失、失禁、癫痫、麻痹、无法说话、失明和听力丧失。最终,此类个体丧失对其周围环境的感知并变得无反应。尽管神经问题是异染性脑白质营养不良的主要特征,但已报告硫苷脂积聚对其它器官和组织的影响,最常涉及胆囊。

[0007] 可藉由几种治疗来管理MLD。例如,用于减少MLD的病征和症状以及缓解相关疼痛的药物。已表明造血干细胞移植可通过引入健康细胞以帮助替代患病细胞来推迟MLD的进展。其它治疗包括用于促进肌肉及关节灵活性以及保持活动范围的物理、职业及言语疗法。然而,尚未能治愈MLD。

[0008] 大部分患有MLD的个体具有芳基硫酸酯酶A (ARSA) 基因的突变,并且已鉴别超过110种不同的引起MLD的ARSA突变。每100人中便发现1人具有携带者突变,并且在美国每40,000名活新生儿中便有1名或在全世界每160,000名活新生儿中便有1名受影响。

[0009] 基因疗法提供治愈MLD的独特机会。逆转录病毒载体(包括慢病毒载体)能够将核

酸整合到宿主细胞基因组中,由于其对基因组的非靶向插入而产生安全性问题。例如,存在载体破坏肿瘤抑制基因或激活癌基因从而引起恶性肿瘤的风险。实际上,在藉由用 γ 逆转录病毒载体转导CD34⁺骨髓前体来进行的用于治疗X连锁重度联合免疫缺损(SCID)的临床试验中,十名患者中的四名产生白血病(Hacein-Bey-Abina等人,《临床研究杂志(J Clin Invest.)》(2008) 118(9):3132-42)。另一方面,非整合载体通常在体内经历表达量不足或表达持续时间不足。

[0010] 因此,所属领域中需要改进的基因疗法组合物和方法,其可有效及安全地恢复MLD患者中的ARSA基因功能。

发明内容

[0011] 本文中提供可恢复细胞中的ARSA基因功能的腺相关病毒(AAV)组合物,以及使用其治疗与ARSA基因功能降低相关的疾病(例如MLD)的方法。还提供用于制备腺相关病毒组合物的包装系统。

[0012] 因此,在一个方面中,本公开提供一种用于在细胞中表达芳基硫酸酯酶A(ARSA)多肽的方法,所述方法包含用重组腺相关病毒(rAAV)转导细胞,所述重组腺相关病毒包含:(a) AAV衣壳,其包含AAV衣壳蛋白(例如进化枝F衣壳蛋白);和(b)转移基因组,其包含可操作地连接于以沉默方式改变的ARSA编码序列的转录调节元件。

[0013] 在某些实施例中,细胞为神经元和/或神经胶质细胞。在某些实施例中,细胞为中枢神经系统和/或周围神经系统的神经元和/或神经胶质细胞。在某些实施例中,细胞为选自以下组成的群组的中枢神经系统区域的细胞:脊髓、运动皮质、感觉皮质、海马、壳核、小脑(任选地,小脑核)和其任何组合。在某些实施例中,细胞为选自以下组成的群组的细胞:运动神经元、星形胶质细胞、少突胶质细胞、中枢神经系统中的大脑皮质的细胞、周围神经系统的感觉神经元、施万细胞(Schwann cell)和其任何组合。在某些实施例中,细胞是在哺乳动物受试者中,并且以有效转导受试者中的细胞的量向受试者施用AAV。

[0014] 在另一方面中,本公开提供一种用于治疗患有异染性脑白质营养不良(MLD)的受试者的方法,所述方法包含向受试者施用有效量的rAAV,其包含:(a) AAV衣壳,其包含AAV衣壳蛋白(例如进化枝F衣壳蛋白);和(b)转移基因组,其包含可操作地连接于以沉默方式改变的ARSA编码序列的转录调节元件。

[0015] 在某些实施例中,以沉默方式改变的ARSA编码序列编码SEQ ID NO:23中所阐述的氨基酸序列。在某些实施例中,以沉默方式改变的ARSA编码序列包含SEQ ID NO:14、62或72中所阐述的核苷酸序列。

[0016] 在某些实施例中,转录调节元件包含选自以下组成的群组的一种或多种元件:巨细胞病毒(CMV)增强子元件、鸡 β -肌动蛋白(CBA)启动子、小型鸡 β -肌动蛋白(SmCBA)启动子、钙调蛋白1(CALM1)启动子、蛋白脂质蛋白质1(PLP1)启动子、胶质细胞原纤维酸性蛋白(GFAP)启动子、突触蛋白2(SYN2)启动子、金属硫蛋白3(MT3)启动子及其任何组合。在某些实施例中,转录调节元件包含与选自以下组成的群组的序列至少90%一致的核苷酸序列:SEQ ID NO:25、32、36、54、55和58。在某些实施例中,转录调节元件包含选自以下组成的群组的核苷酸序列:SEQ ID NO:25、32、36、54、55和58。在某些实施例中,转录调节元件自5'到3'包含SEQ ID NO:58、25和32中所阐述的核苷酸序列。在某些实施例中,转录调节元件

包含SEQ ID NO:36中所阐述的核苷酸序列。

[0017] 在某些实施例中,转移基因组还包含位于以沉默方式改变的ARSA编码序列3'的聚腺苷酸化序列。在某些实施例中,聚腺苷酸化序列是外源性聚腺苷酸化序列。在某些实施例中,外源性聚腺苷酸化序列是SV40聚腺苷酸化序列。在某些实施例中,SV40聚腺苷酸化序列包含SEQ ID NO:42中所阐述的核苷酸序列。

[0018] 在某些实施例中,转移基因组还包含填充序列。在某些实施例中,转移基因组还包含位于以沉默方式改变的ARSA编码序列3'的填充序列。在某些实施例中,填充序列位于聚腺苷酸化序列3'。

[0019] 在某些实施例中,转移基因组包含选自由以下组成的群组的序列:SEQ ID NO:41、44、46、65、67和75。

[0020] 在某些实施例中,转移基因组还包含基因组5'的5'反向末端重复(5' ITR)核苷酸序列和基因组3'的3'反向末端重复(3' ITR)核苷酸序列。在某些实施例中,5' ITR核苷酸序列与SEQ ID NO:18具有至少95%序列一致性,并且3' ITR核苷酸序列与SEQ ID NO:19具有至少95%序列一致性。在某些实施例中,5' ITR核苷酸序列与SEQ ID NO:26具有至少95%序列一致性,并且3' ITR核苷酸序列与SEQ ID NO:27具有至少95%序列一致性。在某些实施例中,5' ITR核苷酸序列与SEQ ID NO:18具有至少95%序列一致性,并且3' ITR核苷酸序列与SEQ ID NO:57具有至少95%序列一致性。

[0021] 在某些实施例中,转移基因组包含选自由以下组成的群组的核苷酸序列:SEQ ID NO:47、48、49、68、69和76。

[0022] 在某些实施例中,异染性脑白质营养不良与芳基硫酸酯酶A (ARSA) 基因突变相关。在某些实施例中,受试者是人类受试者。

[0023] 在某些实施例中,衣壳蛋白包含与SEQ ID NO:2、3、4、6、7、10、11、12、13、15、16或17的氨基酸203-736的氨基酸序列具有至少95%序列一致性的氨基酸序列。在某些实施例中,衣壳蛋白中对应于SEQ ID NO:16的氨基酸206的氨基酸为C;衣壳蛋白中对应于SEQ ID NO:16的氨基酸296的氨基酸为H;衣壳蛋白中对应于SEQ ID NO:16的氨基酸312的氨基酸为Q;衣壳蛋白中对应于SEQ ID NO:16的氨基酸346的氨基酸为A;衣壳蛋白中对应于SEQ ID NO:16的氨基酸464的氨基酸为N;衣壳蛋白中对应于SEQ ID NO:16的氨基酸468的氨基酸为S;衣壳蛋白中对应于SEQ ID NO:16的氨基酸501的氨基酸为I;衣壳蛋白中对应于SEQ ID NO:16的氨基酸505的氨基酸为R;衣壳蛋白中对应于SEQ ID NO:16的氨基酸590的氨基酸为R;衣壳蛋白中对应于SEQ ID NO:16的氨基酸626的氨基酸为G或Y;衣壳蛋白中对应于SEQ ID NO:16的氨基酸681的氨基酸为M;衣壳蛋白中对应于SEQ ID NO:16的氨基酸687的氨基酸为R;衣壳蛋白中对应于SEQ ID NO:16的氨基酸690的氨基酸为K;衣壳蛋白中对应于SEQ ID NO:16的氨基酸706的氨基酸为C;或衣壳蛋白中对应于SEQ ID NO:16的氨基酸718的氨基酸为G。

[0024] 在某些实施例中,(a) 衣壳蛋白中对应于SEQ ID NO:16的氨基酸626的氨基酸为G,并且衣壳蛋白中对应于SEQ ID NO:16的氨基酸718的氨基酸为G;(b) 衣壳蛋白中对应于SEQ ID NO:16的氨基酸296的氨基酸为H,衣壳蛋白中对应于SEQ ID NO:16的氨基酸464的氨基酸为N,衣壳蛋白中对应于SEQ ID NO:16的氨基酸505的氨基酸为R,并且衣壳蛋白中对应于SEQ ID NO:16的氨基酸681的氨基酸为M;(c) 衣壳蛋白中对应于SEQ ID NO:16的氨基酸505

的氨基酸为R,并且衣壳蛋白中对应于SEQ ID NO:16的氨基酸687的氨基酸为R;(d)衣壳蛋白中对应于SEQ ID NO:16的氨基酸346的氨基酸为A,并且衣壳蛋白中对应于SEQ ID NO:16的氨基酸505的氨基酸为R;或(e)衣壳蛋白中对应于SEQ ID NO:16的氨基酸501的氨基酸为I,衣壳蛋白中对应于SEQ ID NO:16的氨基酸505的氨基酸为R,并且衣壳蛋白中对应于SEQ ID NO:16的氨基酸706的氨基酸为C。

[0025] 在某些实施例中,衣壳蛋白包含SEQ ID NO:2、3、4、6、7、10、11、12、13、15、16或17的氨基酸203-736的氨基酸序列。

[0026] 在某些实施例中,衣壳蛋白包含与SEQ ID NO:2、3、4、5、6、7、9、10、11、12、13、15、16或17的氨基酸138-736的氨基酸序列具有至少95%序列一致性的氨基酸序列。在某些实施例中,衣壳蛋白中对应于SEQ ID NO:16的氨基酸151的氨基酸为R;衣壳蛋白中对应于SEQ ID NO:16的氨基酸160的氨基酸为D;衣壳蛋白中对应于SEQ ID NO:16的氨基酸206的氨基酸为C;衣壳蛋白中对应于SEQ ID NO:16的氨基酸296的氨基酸为H;衣壳蛋白中对应于SEQ ID NO:16的氨基酸312的氨基酸为Q;衣壳蛋白中对应于SEQ ID NO:16的氨基酸346的氨基酸为A;衣壳蛋白中对应于SEQ ID NO:16的氨基酸464的氨基酸为N;衣壳蛋白中对应于SEQ ID NO:16的氨基酸468的氨基酸为S;衣壳蛋白中对应于SEQ ID NO:16的氨基酸501的氨基酸为I;衣壳蛋白中对应于SEQ ID NO:16的氨基酸505的氨基酸为R;衣壳蛋白中对应于SEQ ID NO:16的氨基酸590的氨基酸为R;衣壳蛋白中对应于SEQ ID NO:16的氨基酸626的氨基酸为G或Y;衣壳蛋白中对应于SEQ ID NO:16的氨基酸681的氨基酸为M;衣壳蛋白中对应于SEQ ID NO:16的氨基酸687的氨基酸为R;衣壳蛋白中对应于SEQ ID NO:16的氨基酸690的氨基酸为K;衣壳蛋白中对应于SEQ ID NO:16的氨基酸706的氨基酸为C;或衣壳蛋白中对应于SEQ ID NO:16的氨基酸718的氨基酸为G。

[0027] 在某些实施例中,(a)衣壳蛋白中对应于SEQ ID NO:16的氨基酸626的氨基酸为G,并且衣壳蛋白中对应于SEQ ID NO:16的氨基酸718的氨基酸为G;(b)衣壳蛋白中对应于SEQ ID NO:16的氨基酸296的氨基酸为H,衣壳蛋白中对应于SEQ ID NO:16的氨基酸464的氨基酸为N,衣壳蛋白中对应于SEQ ID NO:16的氨基酸505的氨基酸为R,并且衣壳蛋白中对应于SEQ ID NO:16的氨基酸681的氨基酸为M;(c)衣壳蛋白中对应于SEQ ID NO:16的氨基酸505的氨基酸为R,并且衣壳蛋白中对应于SEQ ID NO:16的氨基酸687的氨基酸为R;(d)衣壳蛋白中对应于SEQ ID NO:16的氨基酸346的氨基酸为A,并且衣壳蛋白中对应于SEQ ID NO:16的氨基酸505的氨基酸为R;或(e)衣壳蛋白中对应于SEQ ID NO:16的氨基酸501的氨基酸为I,衣壳蛋白中对应于SEQ ID NO:16的氨基酸505的氨基酸为R,并且衣壳蛋白中对应于SEQ ID NO:16的氨基酸706的氨基酸为C。

[0028] 在某些实施例中,衣壳蛋白包含SEQ ID NO:2、3、4、5、6、7、9、10、11、12、13、15、16或17的氨基酸138-736的氨基酸序列。

[0029] 在某些实施例中,衣壳蛋白包含与SEQ ID NO:2、3、4、5、6、7、8、9、10、11、12、13、15、16或17的氨基酸1-736的氨基酸序列具有至少95%序列一致性的氨基酸序列。在某些实施例中,衣壳蛋白中对应于SEQ ID NO:16的氨基酸2的氨基酸为T;衣壳蛋白中对应于SEQ ID NO:16的氨基酸65的氨基酸为I;衣壳蛋白中对应于SEQ ID NO:16的氨基酸68的氨基酸为V;衣壳蛋白中对应于SEQ ID NO:16的氨基酸77的氨基酸为R;衣壳蛋白中对应于SEQ ID NO:16的氨基酸119的氨基酸为L;衣壳蛋白中对应于SEQ ID NO:16的氨基酸151的氨基酸为

R;衣壳蛋白中对应于SEQ ID NO:16的氨基酸160的氨基酸为D;衣壳蛋白中对应于SEQ ID NO:16的氨基酸206的氨基酸为C;衣壳蛋白中对应于SEQ ID NO:16的氨基酸296的氨基酸为H;衣壳蛋白中对应于SEQ ID NO:16的氨基酸312的氨基酸为Q;衣壳蛋白中对应于SEQ ID NO:16的氨基酸346的氨基酸为A;衣壳蛋白中对应于SEQ ID NO:16的氨基酸464的氨基酸为N;衣壳蛋白中对应于SEQ ID NO:16的氨基酸468的氨基酸为S;衣壳蛋白中对应于SEQ ID NO:16的氨基酸501的氨基酸为I;衣壳蛋白中对应于SEQ ID NO:16的氨基酸505的氨基酸为R;衣壳蛋白中对应于SEQ ID NO:16的氨基酸590的氨基酸为R;衣壳蛋白中对应于SEQ ID NO:16的氨基酸626的氨基酸为G或Y;衣壳蛋白中对应于SEQ ID NO:16的氨基酸681的氨基酸为M;衣壳蛋白中对应于SEQ ID NO:16的氨基酸687的氨基酸为R;衣壳蛋白中对应于SEQ ID NO:16的氨基酸690的氨基酸为K;衣壳蛋白中对应于SEQ ID NO:16的氨基酸706的氨基酸为C;或衣壳蛋白中对应于SEQ ID NO:16的氨基酸718的氨基酸为G。

[0030] 在某些实施例中,(a)衣壳蛋白中对应于SEQ ID NO:16的氨基酸2的氨基酸为T,并且衣壳蛋白中对应于SEQ ID NO:16的氨基酸312的氨基酸为Q;(b)衣壳蛋白中对应于SEQ ID NO:16的氨基酸65的氨基酸为I,并且衣壳蛋白中对应于SEQ ID NO:16的氨基酸626的氨基酸为Y;(c)衣壳蛋白中对应于SEQ ID NO:16的氨基酸77的氨基酸为R,并且衣壳蛋白中对应于SEQ ID NO:16的氨基酸690的氨基酸为K;(d)衣壳蛋白中对应于SEQ ID NO:16的氨基酸119的氨基酸为L,并且衣壳蛋白中对应于SEQ ID NO:16的氨基酸468的氨基酸为S;(e)衣壳蛋白中对应于SEQ ID NO:16的氨基酸626的氨基酸为G,并且衣壳蛋白中对应于SEQ ID NO:16的氨基酸718的氨基酸为G;(f)衣壳蛋白中对应于SEQ ID NO:16的氨基酸296的氨基酸为H,衣壳蛋白中对应于SEQ ID NO:16的氨基酸464的氨基酸为N,衣壳蛋白中对应于SEQ ID NO:16的氨基酸505的氨基酸为R,并且衣壳蛋白中对应于SEQ ID NO:16的氨基酸681的氨基酸为M;(g)衣壳蛋白中对应于SEQ ID NO:16的氨基酸505的氨基酸为R,并且衣壳蛋白中对应于SEQ ID NO:16的氨基酸687的氨基酸为R;(h)衣壳蛋白中对应于SEQ ID NO:16的氨基酸346的氨基酸为A,并且衣壳蛋白中对应于SEQ ID NO:16的氨基酸505的氨基酸为R;或(i)衣壳蛋白中对应于SEQ ID NO:16的氨基酸501的氨基酸为I,衣壳蛋白中对应于SEQ ID NO:16的氨基酸505的氨基酸为R,并且衣壳蛋白中对应于SEQ ID NO:16的氨基酸706的氨基酸为C。

[0031] 在某些实施例中,衣壳蛋白包含SEQ ID NO:2、3、4、5、6、7、8、9、10、11、12、13、15、16或17的氨基酸1-736的氨基酸序列。

[0032] 在另一方面中,本公开提供rAAV,其包含:(a)AAV衣壳,其包含AAV衣壳蛋白(例如进化枝F衣壳蛋白);和(b)转移基因组,其包含可操作地连接于以沉默方式改变的ARSA编码序列的转录调节元件。

[0033] 在某些实施例中,以沉默方式改变的ARSA编码序列编码SEQ ID NO:23中所阐述的氨基酸序列。在某些实施例中,以沉默方式改变的ARSA编码序列包含SEQ ID NO:14中所阐述的核苷酸序列。在某些实施例中,以沉默方式改变的ARSA编码序列包含SEQ ID NO:62或72中所阐述的核苷酸序列。

[0034] 在某些实施例中,转录调节元件包含选自由以下组成的群组的一种或多种元件:巨细胞病毒(CMV)增强子元件、鸡 β -肌动蛋白(CBA)启动子、小型鸡 β -肌动蛋白(SmCBA)启动子、钙调蛋白1(CALM1)启动子、蛋白脂质蛋白质1(PLP1)启动子、胶质细胞原纤维酸性蛋白

(GFAP) 启动子、突触蛋白2 (SYN2) 启动子、金属硫蛋白3 (MT3) 启动子和其任何组合。在某些实施例中,转录调节元件包含与选自由以下组成的群组的序列至少90%一致的核苷酸序列:SEQ ID NO:25、32、36、54、55和58。在某些实施例中,转录调节元件包含选自由以下组成的群组的核苷酸序列:SEQ ID NO:25、32、36、54、55和58。在某些实施例中,转录调节元件自5'到3'包含SEQ ID NO:58、25和32中所阐述的核苷酸序列。在某些实施例中,转录调节元件包含SEQ ID NO:36中所阐述的核苷酸序列。

[0035] 在某些实施例中,转移基因组还包含位于以沉默方式改变的ARSA编码序列3'的聚腺苷酸化序列。在某些实施例中,聚腺苷酸化序列是外源性聚腺苷酸化序列。在某些实施例中,外源性聚腺苷酸化序列是SV40聚腺苷酸化序列。在某些实施例中,SV40聚腺苷酸化序列包含SEQ ID NO:42中所阐述的核苷酸序列。

[0036] 在某些实施例中,转移基因组包含选自由以下组成的群组的序列:SEQ ID NO:41、44、46、65、67和75。

[0037] 在某些实施例中,转移基因组还包含基因组5'的5'反向末端重复(5' ITR)核苷酸序列和基因组3'的3'反向末端重复(3' ITR)核苷酸序列。在某些实施例中,5' ITR核苷酸序列与SEQ ID NO:18具有至少95%序列一致性,并且3' ITR核苷酸序列与SEQ ID NO:19具有至少95%序列一致性。

[0038] 在某些实施例中,转移基因组包含选自由以下组成的群组的核苷酸序列:SEQ ID NO:47、48、49、68、69和76。在某些实施例中,转移基因组的核苷酸序列由选自由以下组成的群组的核苷酸序列组成:SEQ ID NO:47、48、49、68、69及76。在某些实施例中,转移基因组的核苷酸序列由SEQ ID NO:48中所阐述的核苷酸序列组成。

[0039] 在某些实施例中,衣壳蛋白包含与SEQ ID NO:2、3、4、6、7、10、11、12、13、15、16或17的氨基酸203-736的氨基酸序列具有至少95%序列一致性的氨基酸序列。在某些实施例中,衣壳蛋白中对应于SEQ ID NO:16的氨基酸206的氨基酸为C;衣壳蛋白中对应于SEQ ID NO:16的氨基酸296的氨基酸为H;衣壳蛋白中对应于SEQ ID NO:16的氨基酸312的氨基酸为Q;衣壳蛋白中对应于SEQ ID NO:16的氨基酸346的氨基酸为A;衣壳蛋白中对应于SEQ ID NO:16的氨基酸464的氨基酸为N;衣壳蛋白中对应于SEQ ID NO:16的氨基酸468的氨基酸为S;衣壳蛋白中对应于SEQ ID NO:16的氨基酸501的氨基酸为I;衣壳蛋白中对应于SEQ ID NO:16的氨基酸505的氨基酸为R;衣壳蛋白中对应于SEQ ID NO:16的氨基酸590的氨基酸为R;衣壳蛋白中对应于SEQ ID NO:16的氨基酸626的氨基酸为G或Y;衣壳蛋白中对应于SEQ ID NO:16的氨基酸681的氨基酸为M;衣壳蛋白中对应于SEQ ID NO:16的氨基酸687的氨基酸为R;衣壳蛋白中对应于SEQ ID NO:16的氨基酸690的氨基酸为K;衣壳蛋白中对应于SEQ ID NO:16的氨基酸706的氨基酸为C;或衣壳蛋白中对应于SEQ ID NO:16的氨基酸718的氨基酸为G。

[0040] 在某些实施例中,(a)衣壳蛋白中对应于SEQ ID NO:16的氨基酸626的氨基酸为G,并且衣壳蛋白中对应于SEQ ID NO:16的氨基酸718的氨基酸为G;(b)衣壳蛋白中对应于SEQ ID NO:16的氨基酸296的氨基酸为H,衣壳蛋白中对应于SEQ ID NO:16的氨基酸464的氨基酸为N,衣壳蛋白中对应于SEQ ID NO:16的氨基酸505的氨基酸为R,并且衣壳蛋白中对应于SEQ ID NO:16的氨基酸681的氨基酸为M;(c)衣壳蛋白中对应于SEQ ID NO:16的氨基酸505的氨基酸为R,并且衣壳蛋白中对应于SEQ ID NO:16的氨基酸687的氨基酸为R;(d)衣壳蛋

白中对应于SEQ ID NO:16的氨基酸346的氨基酸为A,并且衣壳蛋白中对应于SEQ ID NO:16的氨基酸505的氨基酸为R;或(e)衣壳蛋白中对应于SEQ ID NO:16的氨基酸501的氨基酸为I,衣壳蛋白中对应于SEQ ID NO:16的氨基酸505的氨基酸为R,并且衣壳蛋白中对应于SEQ ID NO:16的氨基酸706的氨基酸为C。

[0041] 在某些实施例中,衣壳蛋白包含SEQ ID NO:2、3、4、6、7、10、11、12、13、15、16或17的氨基酸203-736的氨基酸序列。

[0042] 在某些实施例中,衣壳蛋白包含与SEQ ID NO:2、3、4、5、6、7、9、10、11、12、13、15、16或17的氨基酸138-736的氨基酸序列具有至少95%序列一致性的氨基酸序列。在某些实施例中,衣壳蛋白中对应于SEQ ID NO:16的氨基酸151的氨基酸为R;衣壳蛋白中对应于SEQ ID NO:16的氨基酸160的氨基酸为D;衣壳蛋白中对应于SEQ ID NO:16的氨基酸206的氨基酸为C;衣壳蛋白中对应于SEQ ID NO:16的氨基酸296的氨基酸为H;衣壳蛋白中对应于SEQ ID NO:16的氨基酸312的氨基酸为Q;衣壳蛋白中对应于SEQ ID NO:16的氨基酸346的氨基酸为A;衣壳蛋白中对应于SEQ ID NO:16的氨基酸464的氨基酸为N;衣壳蛋白中对应于SEQ ID NO:16的氨基酸468的氨基酸为S;衣壳蛋白中对应于SEQ ID NO:16的氨基酸501的氨基酸为I;衣壳蛋白中对应于SEQ ID NO:16的氨基酸505的氨基酸为R;衣壳蛋白中对应于SEQ ID NO:16的氨基酸590的氨基酸为R;衣壳蛋白中对应于SEQ ID NO:16的氨基酸626的氨基酸为G或Y;衣壳蛋白中对应于SEQ ID NO:16的氨基酸681的氨基酸为M;衣壳蛋白中对应于SEQ ID NO:16的氨基酸687的氨基酸为R;衣壳蛋白中对应于SEQ ID NO:16的氨基酸690的氨基酸为K;衣壳蛋白中对应于SEQ ID NO:16的氨基酸706的氨基酸为C;或衣壳蛋白中对应于SEQ ID NO:16的氨基酸718的氨基酸为G。

[0043] 在某些实施例中,(a)衣壳蛋白中对应于SEQ ID NO:16的氨基酸626的氨基酸为G,并且衣壳蛋白中对应于SEQ ID NO:16的氨基酸718的氨基酸为G;(b)衣壳蛋白中对应于SEQ ID NO:16的氨基酸296的氨基酸为H,衣壳蛋白中对应于SEQ ID NO:16的氨基酸464的氨基酸为N,衣壳蛋白中对应于SEQ ID NO:16的氨基酸505的氨基酸为R,并且衣壳蛋白中对应于SEQ ID NO:16的氨基酸681的氨基酸为M;(c)衣壳蛋白中对应于SEQ ID NO:16的氨基酸505的氨基酸为R,并且衣壳蛋白中对应于SEQ ID NO:16的氨基酸687的氨基酸为R;(d)衣壳蛋白中对应于SEQ ID NO:16的氨基酸346的氨基酸为A,并且衣壳蛋白中对应于SEQ ID NO:16的氨基酸505的氨基酸为R;或(e)衣壳蛋白中对应于SEQ ID NO:16的氨基酸501的氨基酸为I,衣壳蛋白中对应于SEQ ID NO:16的氨基酸505的氨基酸为R,并且衣壳蛋白中对应于SEQ ID NO:16的氨基酸706的氨基酸为C。

[0044] 在某些实施例中,衣壳蛋白包含SEQ ID NO:2、3、4、5、6、7、9、10、11、12、13、15、16或17的氨基酸138-736的氨基酸序列。

[0045] 在某些实施例中,衣壳蛋白包含与SEQ ID NO:2、3、4、5、6、7、8、9、10、11、12、13、15、16或17的氨基酸1-736的氨基酸序列具有至少95%序列一致性的氨基酸序列。在某些实施例中,衣壳蛋白中对应于SEQ ID NO:16的氨基酸2的氨基酸为T;衣壳蛋白中对应于SEQ ID NO:16的氨基酸65的氨基酸为I;衣壳蛋白中对应于SEQ ID NO:16的氨基酸68的氨基酸为V;衣壳蛋白中对应于SEQ ID NO:16的氨基酸77的氨基酸为R;衣壳蛋白中对应于SEQ ID NO:16的氨基酸119的氨基酸为L;衣壳蛋白中对应于SEQ ID NO:16的氨基酸151的氨基酸为R;衣壳蛋白中对应于SEQ ID NO:16的氨基酸160的氨基酸为D;衣壳蛋白中对应于SEQ ID

NO:16的氨基酸206的氨基酸为C;衣壳蛋白中对应于SEQ ID NO:16的氨基酸296的氨基酸为H;衣壳蛋白中对应于SEQ ID NO:16的氨基酸312的氨基酸为Q;衣壳蛋白中对应于SEQ ID NO:16的氨基酸346的氨基酸为A;衣壳蛋白中对应于SEQ ID NO:16的氨基酸464的氨基酸为N;衣壳蛋白中对应于SEQ ID NO:16的氨基酸468的氨基酸为S;衣壳蛋白中对应于SEQ ID NO:16的氨基酸501的氨基酸为I;衣壳蛋白中对应于SEQ ID NO:16的氨基酸505的氨基酸为R;衣壳蛋白中对应于SEQ ID NO:16的氨基酸590的氨基酸为R;衣壳蛋白中对应于SEQ ID NO:16的氨基酸626的氨基酸为G或Y;衣壳蛋白中对应于SEQ ID NO:16的氨基酸681的氨基酸为M;衣壳蛋白中对应于SEQ ID NO:16的氨基酸687的氨基酸为R;衣壳蛋白中对应于SEQ ID NO:16的氨基酸690的氨基酸为K;衣壳蛋白中对应于SEQ ID NO:16的氨基酸706的氨基酸为C;或衣壳蛋白中对应于SEQ ID NO:16的氨基酸718的氨基酸为G。

[0046] 在某些实施例中,(a)衣壳蛋白中对应于SEQ ID NO:16的氨基酸2的氨基酸为T,并且衣壳蛋白中对应于SEQ ID NO:16的氨基酸312的氨基酸为Q;(b)衣壳蛋白中对应于SEQ ID NO:16的氨基酸65的氨基酸为I,并且衣壳蛋白中对应于SEQ ID NO:16的氨基酸626的氨基酸为Y;(c)衣壳蛋白中对应于SEQ ID NO:16的氨基酸77的氨基酸为R,并且衣壳蛋白中对应于SEQ ID NO:16的氨基酸690的氨基酸为K;(d)衣壳蛋白中对应于SEQ ID NO:16的氨基酸119的氨基酸为L,并且衣壳蛋白中对应于SEQ ID NO:16的氨基酸468的氨基酸为S;(e)衣壳蛋白中对应于SEQ ID NO:16的氨基酸626的氨基酸为G,并且衣壳蛋白中对应于SEQ ID NO:16的氨基酸718的氨基酸为G;(f)衣壳蛋白中对应于SEQ ID NO:16的氨基酸296的氨基酸为H,衣壳蛋白中对应于SEQ ID NO:16的氨基酸464的氨基酸为N,衣壳蛋白中对应于SEQ ID NO:16的氨基酸505的氨基酸为R,并且衣壳蛋白中对应于SEQ ID NO:16的氨基酸681的氨基酸为M;(g)衣壳蛋白中对应于SEQ ID NO:16的氨基酸505的氨基酸为R,并且衣壳蛋白中对应于SEQ ID NO:16的氨基酸687的氨基酸为R;(h)衣壳蛋白中对应于SEQ ID NO:16的氨基酸346的氨基酸为A,并且衣壳蛋白中对应于SEQ ID NO:16的氨基酸505的氨基酸为R;或(i)衣壳蛋白中对应于SEQ ID NO:16的氨基酸501的氨基酸为I,衣壳蛋白中对应于SEQ ID NO:16的氨基酸505的氨基酸为R,并且衣壳蛋白中对应于SEQ ID NO:16的氨基酸706的氨基酸为C。

[0047] 在某些实施例中,衣壳蛋白包含SEQ ID NO:2、3、4、5、6、7、8、9、10、11、12、13、15、16或17的氨基酸1-736的氨基酸序列。

[0048] 在另一方面中,本公开提供一种药物组合物,其包含本文中所描述的rAAV。

[0049] 在另一方面中,本公开提供一种多核苷酸,其包含SEQ ID NO:14、62和72中所阐述的核酸序列。

[0050] 在另一方面中,本公开提供一种用于制备rAAV的包装系统,其中所述包装系统包含(a)第一核苷酸序列,其编码一种或多种AAV Rep蛋白;(b)第二核苷酸序列,其编码如权利要求41至71中任一项所述的AAV的衣壳蛋白;和(c)第三核苷酸序列,其包含如权利要求41至71中任一项所述的AAV的rAAV基因组序列。

[0051] 在某些实施例中,包装系统包含有包含第一核苷酸序列和第二核苷酸序列的第一载体,以及包含第三核苷酸序列的第二载体。

[0052] 在某些实施例中,包装系统还包含有包含一种或多种辅助病毒基因的第四核苷酸序列。在某些实施例中,第四核苷酸序列包含于第三载体内。在某些实施例中,第四核苷酸

序列包含一种或多种来自选自以下组成的群组的病毒的基因：腺病毒、疱疹病毒、痘苗病毒和巨细胞病毒 (CMV)。

[0053] 在某些实施例中，第一载体、第二载体和/或第三载体为质粒。

[0054] 在另一方面中，本公开提供一种用于rAAV的重组制备的方法，所述方法包含在可产生rAAV的条件下将本文中所描述的包装系统引入细胞中。

[0055] 在另一方面中，本公开提供本文中所描述的rAAV，其用于如本文中所描述的用于在细胞中表达芳基硫酸酯酶A (ARSA) 多肽的方法中。

[0056] 在另一方面中，本公开提供本文中所描述的rAAV，其用于如本文中所描述的用于治疗患有异染性脑白质营养不良 (MLD) 的受试者的方法中。

附图说明

[0057] 图1A、1B、1C和1D分别为T-001、pHMI-5000、pHMI-5003和pHMI-hARSA1-TC-002载体的载体图。

[0058] 图2A、2B和2C。图2A为展示在用媒剂对照物或包装于AAVHSC15衣壳中的pHMI-5000处理的ARSA (-/-) 小鼠中，使用抗LAMP-1抗体藉由免疫组织化学研究的来源于LAMP-1免疫反应性的总像素强度的定量的图 (dWM: 背部白质; vWM: 腹部白质; 和vGM: 腹部灰质)。图2B为展示随时间推移，在对照组小鼠 (WT/Het) 和ARSA (-/-) 小鼠的脑中测量到的C18:0硫苷脂含量的图。图2C为展示在用 4×10^{13} vg/kg的剂量 (剂量-4) 的包装于AAVHSC15衣壳中的pHMI-hARSA1-TC-002或媒剂对照物处理的ARSA (-/-) 小鼠中的硫苷脂含量的变化 (以与年龄匹配野生型对照物相比的倍数形式) 的图。图2D为图的集合，其展示在用 4×10^{13} vg/kg或 6×10^{13} vg/kg的剂量的包装于AAVHSC15衣壳中的pHMI-5000或媒剂对照物处理的ARSA (-/-) 小鼠的前脑、中脑和后脑中的C18:0及C18:1硫苷脂同种型含量的变化 (以与年龄匹配野生型对照物相比的倍数形式)。图2E为图的集合，其展示在用 4×10^{13} vg/kg的剂量的包装于AAVHSC15衣壳中的pHMI-5000或媒剂对照物处理的ARSA (-/-) 小鼠的前脑、中脑和后脑中的C18:0及C18:1硫苷脂同种型含量的变化 (以与年龄匹配野生型对照物相比的倍数形式)。图2F为图的集合，其展示在用 4×10^{13} vg/kg的剂量的包装于AAVHSC15衣壳中的pHMI-5000或媒剂对照物处理的ARSA (-/-) 小鼠的前脑、中脑和后脑中的C24:0及C24:1硫苷脂同种型含量的变化 (以与年龄匹配野生型对照物相比的倍数形式)。图2G为图的集合，其展示在用 4×10^{13} vg/kg的剂量的包装于AAVHSC15衣壳中的pHMI-5000或媒剂对照物处理的ARSA (-/-) 小鼠的前脑、中脑和后脑中的总硫苷脂同种型含量的变化 (以与年龄匹配野生型对照物相比的倍数形式)。

[0059] 图3A和3B。图3A为展示在对照组小鼠 (WT/Het) 和ARSA (-/-) 小鼠中，在四周时测量的髓鞘和淋巴细胞蛋白质 (MAL) mRNA转录物的含量的图。图3B为展示与年龄匹配野生型小鼠和用媒剂处理的ARSA (-/-) 小鼠相比，在用 4×10^{13} vg/kg的剂量 (剂量-4) 的包装于AAVHSC15衣壳中的pHMI-5000处理的ARSA (-/-) 小鼠中检测到的MAL转录物的含量的图。图3C为展示在施用 4×10^{13} vg/kg的包装于AAVHSC15衣壳中的pHMI-5000或媒剂对照物之后第12或52周，在野生型小鼠或ARSA (-/-) 小鼠中检测到的MAL转录物拷贝数的图。

[0060] 图4为展示ARSA (-/-) 小鼠的脑中的以每个转导的细胞计的载体基因组的数目与以每ng cDNA计的hARSA的拷贝数之间的相关性的曲线。

[0061] 图5为展示在静脉内施用包装于AAV9或AAVHSC15衣壳中的转移载体pHMI-5000 (在

每种情况下以 2×10^{13} vg/kg的剂量施用)之后,ARSA (-/-)小鼠的脑中的以每个转导的细胞计的载体基因组的数目的图。

[0062] 图6为展示在静脉内施用包装于AAV9或AAVHSC15衣壳中并以所指示的剂量施用的转移载体pHMI-5000之后,在ARSA (-/-)小鼠的脑中测量到的正常人类ARSA酶活性水平的百分比的图。

[0063] 图7为展示在静脉内施用包装于AAV9或AAVHSC15中的转移载体pHMI-5000 (在每种情况下以 4×10^{13} vg/kg的剂量)之后,ARSA (-/-)小鼠的脑中以每个细胞计的载体基因组的数目的图。

[0064] 图8为展示在静脉内 (IV) 或鞘内 (IT) 施用包装于AAVHSC15中的转移载体pHMI-5000之后,后脑及中脑中的正常人类ARSA酶活性的百分比的图。

[0065] 图9A、9B、9C和9D。图9A为展示在以所指示的剂量向ARSA (-/-)小鼠静脉内施用包装于AAVHSC15衣壳中的转移载体pHMI-5000之后,在脑中实现的正常hARSA活性的百分比的图。图9B为展示在以所指示的剂量静脉内施用包装于AAVHSC15衣壳中的转移载体pHMI-5000之后,ARSA (-/-)小鼠的大脑中的以每个细胞计的载体基因组的数目的图。图9C为展示在给药后12周的过程中,给药 4×10^{13} vg/kg的包装于AAVHSC15衣壳中的pHMI-5000的新生ARSA (-/-)小鼠中的hARSA酶活性水平的图。图9D为展示给药 4×10^{13} vg/kg的包装于AAVHSC15衣壳中的pHMI-5000的成年ARSA (-/-)小鼠的大脑中的ARSA酶活性水平 (经由hARSA转录物分析) 的图。图9E为展示在施用单次静脉内 4×10^{13} vg/kg的剂量的包装于AAVHSC15衣壳中的pHMI-5000的ARSA (-/-)小鼠的脑中,以每 μ g基因组DNA计的载体基因组的数目的图。图9F为展示在施用单次静脉内 4×10^{13} vg/kg的剂量的包装于AAVHSC15衣壳中的pHMI-5000的ARSA (-/-)小鼠的脑中,以每ng RNA计的ARSA转录物的拷贝数的图。

[0066] 图10A和10B分别为TC-013.pHMIA2和TC-015.pKITR载体的载体图。

[0067] 图11为展示在施用转移载体pHMI-5000 (CBA启动子)、TC-013.pHMIA2 (CALM1启动子) 和TC-015.pKITR (smCBA启动子) (在每种情况下包装于AAVHSC15衣壳中并以 4×10^{13} vg/kg的剂量静脉内施用) 的小鼠ARSA (-/-)小鼠的脑中,以每个细胞计的转导的病毒基因组的数目的图。

[0068] 图12为展示在施用转移载体pHMI-5000 (CBA启动子) 和TC-015.pKITR (smCBA启动子) (在每种情况下包装于AAVHSC15衣壳中并以 4×10^{13} vg/kg的剂量静脉内施用) 的小鼠ARSA (-/-)小鼠的脑中检测到的正常人类ARSA酶活性的百分比的图。

[0069] 图13为免疫印迹的照片,其展示使用抗hARSA抗体的小鼠的脑中hARSA的表达。向ARSA (-/-)小鼠施用转移载体pHMI-5000 (CBA启动子) 和TC-015.pKITR (smCBA启动子),在每种情况下包装于AAVHSC15衣壳中并分别以 4×10^{13} vg/kg和 8×10^{13} vg/kg的剂量静脉内施用 (对于每种载体,n=5只小鼠)。

[0070] 图14为转移载体pHMI-5004的载体图。

[0071] 图15为转移载体pHMI-5005的载体图。

[0072] 图16为展示用所指示的剂量的包装于AAVHSC15衣壳中的pHMI-5005处理或用媒剂对照物处理的非人类灵长类动物中的丙氨酸氨基转移酶 (ALT) 含量的图。

[0073] 图17为展示给药包装于AAVHSC15衣壳中的pHMI-5005的非人类灵长类动物的中枢神经系统 (CNS) 和脑脊髓液 (CSF) 中的ARSA活性的图。

具体实施方式

[0074] 本文中提供可恢复细胞中的ARSA基因功能的腺相关病毒(AAV)组合物,以及使用其治疗与ARSA基因功能降低相关的疾病(例如MLD)的方法。还提供用于制备腺相关病毒组合物的包装系统。

[0075] I. 定义

[0076] 如本文所用,术语“复制缺陷型腺相关病毒”是指包含缺少Rep和Cap基因的基因组的AAV。

[0077] 如本文所用,术语“ARSA基因”是指芳基硫酸酯酶A基因。由国家生物技术信息中心(National Center for Biotechnology Information;NCBI)基因ID 410鉴别人类ARSA基因。ARSA mRNA的示例性核苷酸序列提供为SEQ ID NO:14。ARSA多肽的示例性氨基酸序列提供为SEQ ID NO:23。

[0078] 如本文所用,术语“转移基因组”是指包含可操作地连接于外源性转录调节元件的编码序列的重组AAV基因组,所述外源性转录调节元件在转移基因组被引入细胞中时介导编码序列的表达。在某些实施例中,转移基因组未整合在细胞的染色体DNA中。所属领域的技术人员将了解,转移基因组中的包含可操作地连接于ARSA编码序列的转录调节元件的部分相对于ARSA编码序列的转录方向可为有义或反义定向。

[0079] 如本文所用,术语“进化枝F衣壳蛋白”是指分别与在本文SEQ ID NO:1的氨基酸1-736、138-736和203-736中所阐述的VP1、VP2或VP3氨基酸序列具有至少90%一致性的AAV VP1、VP2或VP3衣壳蛋白。

[0080] 如本文所用,两个核苷酸序列之间或两个氨基酸序列之间的“一致性百分比”是藉由将所比对的序列对之间的匹配数目乘以100并除以所比对的区域的长度(包括内部空位)来计算。一致性评分仅对完美匹配进行计数,并且不考虑氨基酸彼此之间的相似性程度。长度中仅包括内部空位,而不包括序列末端处的空位。

[0081] 如本文所用,术语“与ARSA基因突变相关的疾病或病症”是指由ARSA基因的突变引起、由ARSA基因的突变加剧或在遗传学上与ARSA基因的突变相关的任何疾病或病症。在某些实施例中,与ARSA基因突变相关的疾病或病症为异染性脑白质营养不良(MLD)。

[0082] 如本文所用,术语“编码序列”是指互补DNA(cDNA)的编码多肽的部分,其开始于起始密码子并且终止于终止密码子。基因可由于替代性剪接、替代性翻译起始和群体内的变化而具有一个或多个编码序列。编码序列可以是野生型或密码子改变型。示例性野生型ARSA编码序列阐述于SEQ ID NO:24中。

[0083] 如本文所用,术语“以沉默方式改变”是指在不改变由编码序列或经填充子插入的编码序列编码的多肽的氨基酸序列的情况下,基因的编码序列或经填充子插入的编码序列的变化(例如通过核苷酸取代)。此类沉默变化的有利之处在于,其可提高编码序列的翻译效率,和/或在编码序列被转导到细胞中时防止与内源性基因的对应序列重组。

[0084] 在本公开中,相对于起始密码子的第一核苷酸指定ARSA基因中的核苷酸位置。起始密码子的第一核苷酸是位置1;位于起始密码子的第一核苷酸5'的核苷酸具有负数;位于起始密码子的第一核苷酸3'的核苷酸具有正数。人类ARSA基因的示例性核苷酸1为NCBI参考序列的核苷酸374:NG_009260.2(区域:5028-10426),并且人类ARSA基因的示例性核苷酸3为NCBI参考序列的核苷酸376:NG_009260.2(区域:5028-10426)。与起始密码子5'相邻的

核苷酸是核苷酸-1。

[0085] 在本公开中,相对于涵盖起始密码子的第一核苷酸的外显子指定ARSA基因中的外显子和内含子,所述第一核苷酸为NCBI参考序列之核苷酸374:NG_009260.2(区域:5028-10426)。涵盖起始密码子的第一核苷酸的外显子是外显子1。外显子3'到外显子1为5'到3':外显子2、外显子3等。内含子3'到外显子1为5'到3':内含子1、内含子2等。因此,ARSA基因自5'到3'包含:外显子1、内含子1、外显子2、内含子2、外显子3等。人类ARSA基因的示例性外显子1为NCBI参考序列的核苷酸374-597:NG_009260.2(区域:5028-10426)。人类ARSA基因的示例性内含子1为NCBI参考序列的核苷酸598-746:NG_009260.2(区域:5028-10426)。

[0086] 如本文所用,术语“转录调节元件”或“TRE”是指顺式作用核苷酸序列,例如DNA序列,其藉由RNA聚合酶调控(例如控制、增加或减少)可操作地连接的核苷酸序列的转录以形成RNA分子。TRE依靠一种或多种反式作用分子,如转录因子来调控转录。因此,一个TRE在其与不同的反式作用分子接触时(例如当其位于不同类型的细胞中时)可以不同方式调控转录。TRE可包含一种或多种启动子元件和/或增强子元件。所属领域的技术人员将了解,基因中的启动子和增强子元件可在位置上靠近,并且术语“启动子”可指包含启动子元件和增强子元件的序列。因此,术语“启动子”不排除序列中的增强子元件。启动子和增强子元件无需来源于相同基因或物种,并且每个启动子或增强子元件的序列可与基因组中的对应内源性序列一致或大体上一致。

[0087] 如本文所用,术语“可操作地连接”用于描述TRE与待转录的编码序列之间的连接。通常,使基因表达处于包含一种或多种启动子和/或增强子元件的TRE的控制下。如果编码序列的转录由TRE控制或影响,那么编码序列“可操作地连接”于TRE。TRE的启动子和增强子元件可自编码序列呈任何定向和/或距离,只要获得所需转录活性即可。在某些实施例中,TRE位于编码序列的上游。

[0088] 如本文所用,术语“核糖体跳跃元件”是指编码短肽序列的核苷酸序列,其能够引起由一个mRNA分子的翻译产生两个肽链。在某些实施例中,核糖体跳跃元件编码包含 $X_1X_2EX_3$ NPGP的共有基元之肽,其中 X_1 为D或G, X_2 为V或I,并且 X_3 为任何氨基酸(SEQ ID NO:34)。在某些实施例中,核糖体跳跃元件编码明脉扁刺蛾(*Thosea asigna*)病毒2A肽(T2A)、猪捷申病毒(porcine teschovirus)-1 2A肽(P2A)、口蹄疫病毒2A肽(F2A)、马鼻肺炎病毒2A肽(E2A)、细胞质多角体病毒2A肽(BmCPV 2A)或家蚕(*B.mori*) 2A肽的软化病病毒(BmIFV 2A)。T2A肽和P2A肽的示例性氨基酸序列分别阐述于SEQ ID NO:37和38中。T2A元件和P2A元件的示例性核苷酸序列分别阐述于SEQ ID NO:66和63中。在某些实施例中,核糖体跳跃元件编码在N端还包含Gly-Ser-Gly的序列的肽,任选地其中Gly-Ser-Gly的序列由GGCAGCGGA的核苷酸序列编码。尽管不希望受理论约束,但假设核糖体跳跃元件通过以下发挥功能:终止第一肽链的翻译和再起始第二肽链的翻译;或由核糖体跳跃元件编码的肽序列中的肽键通过编码的肽的内在蛋白酶活性进行裂解,或环境(例如胞质溶胶)中的另一种蛋白酶。

[0089] 如本文所用,术语“核糖体跳跃肽”是指由核糖体跳跃元件编码的肽。

[0090] 如本文所用,术语“聚腺苷酸化序列”是指当转录成RNA时构成聚腺苷酸化信号序列的DNA序列。聚腺苷酸化序列可以是原生的(例如来自ARSA基因)或外源的。外源性聚腺苷酸化序列可以是哺乳动物或病毒性聚腺苷酸化序列(例如SV40聚腺苷酸化序列)。

[0091] 如本文所用,“外源性聚腺苷酸化序列”是指与ARSA基因(例如人类ARSA基因)的内

源性聚腺苷酸化序列不一致或大体上一致的聚腺苷酸化序列。在某些实施例中,外源性聚腺苷酸化序列为相同物种(例如人类)中的非ARSA基因的聚腺苷酸化序列。在某些实施例中,外源性聚腺苷酸化序列为不同物种(例如病毒)的聚腺苷酸化序列。

[0092] 如本文所用,在向受试者施用AAV的情形下,术语“有效量”是指达到所期望的预防或治疗效果的AAV的量。

[0093] II.腺相关病毒组合物

[0094] 在一个方面中,本文中提供新颖重组AAV(例如复制缺陷型AAV)组合物,其适用于在具有降低的或以其它方式缺乏ARSA基因功能的细胞中表达ARSA多肽。在某些实施例中,本文中所公开的rAAV包含:AAV衣壳,其包含衣壳蛋白(例如进化枝F衣壳蛋白);和转移基因组,其包含可操作地连接于ARSA编码序列(例如以沉默方式改变的ARSA编码序列)的转录调节元件,从而实现经AAV转导的细胞中的ARSA的染色体表达。

[0095] 来自所属领域中已知的任何衣壳的衣壳蛋白可用于本文中所公开的rAAV组合物中,包括(但不限于)来自AAV1、AAV2、AAV3、AAV4、AAV5、AAV6、AAV7、AAV8或AAV9血清型的衣壳蛋白。例如,在某些实施例中,衣壳蛋白包含与SEQ ID NO:2、3、4、5、6、7、8、9、10、11、12、13、15、16或17的氨基酸203-736的氨基酸序列具有至少80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%序列一致性的氨基酸序列。在某些实施例中,衣壳蛋白包含与SEQ ID NO:2、3、4、5、6、7、8、9、10、11、12、13、15、16或17的氨基酸203-736的氨基酸序列具有至少80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%序列一致性的氨基酸序列,其中:衣壳蛋白中对应于SEQ ID NO:16的氨基酸206的氨基酸为C;衣壳蛋白中对应于SEQ ID NO:16的氨基酸296的氨基酸为H;衣壳蛋白中对应于SEQ ID NO:16的氨基酸312的氨基酸为Q;衣壳蛋白中对应于SEQ ID NO:16的氨基酸346的氨基酸为A;衣壳蛋白中对应于SEQ ID NO:16的氨基酸464的氨基酸为N;衣壳蛋白中对应于SEQ ID NO:16的氨基酸468的氨基酸为S;衣壳蛋白中对应于SEQ ID NO:16的氨基酸501的氨基酸为I;衣壳蛋白中对应于SEQ ID NO:16的氨基酸505的氨基酸为R;衣壳蛋白中对应于SEQ ID NO:16的氨基酸590的氨基酸为R;衣壳蛋白中对应于SEQ ID NO:16的氨基酸626的氨基酸为G或Y;衣壳蛋白中对应于SEQ ID NO:16的氨基酸681的氨基酸为M;衣壳蛋白中对应于SEQ ID NO:16的氨基酸687的氨基酸为R;衣壳蛋白中对应于SEQ ID NO:16的氨基酸690的氨基酸为K;衣壳蛋白中对应于SEQ ID NO:16的氨基酸706的氨基酸为C;或衣壳蛋白中对应于SEQ ID NO:16的氨基酸718的氨基酸为G。在某些实施例中,衣壳蛋白中对应于SEQ ID NO:16的氨基酸626的氨基酸为G,并且衣壳蛋白中对应于SEQ ID NO:16的氨基酸718的氨基酸为G。在某些实施例中,衣壳蛋白中对应于SEQ ID NO:16的氨基酸296的氨基酸为H,衣壳蛋白中对应于SEQ ID NO:16的氨基酸464的氨基酸为N,衣壳蛋白中对应于SEQ ID NO:16的氨基酸505的氨基酸为R,并且衣壳蛋白中对应于SEQ ID NO:16的氨基酸681的氨基酸为M。在某些实施例中,衣壳蛋白中对应于SEQ ID NO:16的氨基酸505的氨基酸为R,并且衣壳蛋白中对应于SEQ ID NO:16的氨基酸687的氨基酸为R。在某些实施例中,衣壳蛋白中对应于SEQ ID NO:16的氨基酸346的氨基酸为A,并且衣壳蛋白中对应于SEQ ID NO:16的氨基酸505的氨基酸为R。在某些实施例中,衣壳蛋白中对应于SEQ ID NO:16的氨基酸501的氨基酸为I,衣壳蛋白中对应于SEQ ID NO:16的氨基酸505的氨基酸为R,并且衣壳蛋白中对应于

SEQ ID NO:16的氨基酸706的氨基酸为C。在某些实施例中,衣壳蛋白包含SEQ ID NO:2、3、4、6、7、10、11、12、13、15、16或17的氨基酸203-736的氨基酸序列。

[0096] 例如,在某些实施例中,衣壳蛋白包含与SEQ ID NO:2、3、4、5、6、7、8、9、10、11、12、13、15、16或17的氨基酸138-736的氨基酸序列具有至少80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%序列一致性的氨基酸序列。在某些实施例中,衣壳蛋白包含与SEQ ID NO:2、3、4、5、6、7、8、9、10、11、12、13、15、16或17的氨基酸138-736的氨基酸序列具有至少80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%序列一致性的氨基酸序列,其中:衣壳蛋白中对应于SEQ ID NO:16的氨基酸151的氨基酸为R;衣壳蛋白中对应于SEQ ID NO:16的氨基酸160的氨基酸为D;衣壳蛋白中对应于SEQ ID NO:16的氨基酸206的氨基酸为C;衣壳蛋白中对应于SEQ ID NO:16的氨基酸296的氨基酸为H;衣壳蛋白中对应于SEQ ID NO:16的氨基酸312的氨基酸为Q;衣壳蛋白中对应于SEQ ID NO:16的氨基酸346的氨基酸为A;衣壳蛋白中对应于SEQ ID NO:16的氨基酸464的氨基酸为N;衣壳蛋白中对应于SEQ ID NO:16的氨基酸468的氨基酸为S;衣壳蛋白中对应于SEQ ID NO:16的氨基酸501的氨基酸为I;衣壳蛋白中对应于SEQ ID NO:16的氨基酸505的氨基酸为R;衣壳蛋白中对应于SEQ ID NO:16的氨基酸590的氨基酸为R;衣壳蛋白中对应于SEQ ID NO:16的氨基酸626的氨基酸为G或Y;衣壳蛋白中对应于SEQ ID NO:16的氨基酸681的氨基酸为M;衣壳蛋白中对应于SEQ ID NO:16的氨基酸687的氨基酸为R;衣壳蛋白中对应于SEQ ID NO:16的氨基酸690的氨基酸为K;衣壳蛋白中对应于SEQ ID NO:16的氨基酸706的氨基酸为C;或衣壳蛋白中对应于SEQ ID NO:16的氨基酸718的氨基酸为G。在某些实施例中,衣壳蛋白中对应于SEQ ID NO:16的氨基酸626的氨基酸为G,并且衣壳蛋白中对应于SEQ ID NO:16的氨基酸718的氨基酸为G。在某些实施例中,衣壳蛋白中对应于SEQ ID NO:16的氨基酸296的氨基酸为H,衣壳蛋白中对应于SEQ ID NO:16的氨基酸464的氨基酸为N,衣壳蛋白中对应于SEQ ID NO:16的氨基酸505的氨基酸为R,并且衣壳蛋白中对应于SEQ ID NO:16的氨基酸681的氨基酸为M。在某些实施例中,衣壳蛋白中对应于SEQ ID NO:16的氨基酸505的氨基酸为R,并且衣壳蛋白中对应于SEQ ID NO:16的氨基酸687的氨基酸为R。在某些实施例中,衣壳蛋白中对应于SEQ ID NO:16的氨基酸346的氨基酸为A,并且衣壳蛋白中对应于SEQ ID NO:16的氨基酸505的氨基酸为R。在某些实施例中,衣壳蛋白中对应于SEQ ID NO:16的氨基酸501的氨基酸为I,衣壳蛋白中对应于SEQ ID NO:16的氨基酸505的氨基酸为R,并且衣壳蛋白中对应于SEQ ID NO:16的氨基酸706的氨基酸为C。在某些实施例中,衣壳蛋白包含SEQ ID NO:2、3、4、5、6、7、9、10、11、12、13、15、16或17的氨基酸138-736的氨基酸序列。

[0097] 例如,在某些实施例中,衣壳蛋白包含与SEQ ID NO:2、3、4、5、6、7、8、9、10、11、12、13、15、16或17的氨基酸1-736的氨基酸序列具有至少80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%序列一致性的氨基酸序列。在某些实施例中,衣壳蛋白包含与SEQ ID NO:2、3、4、5、6、7、8、9、10、11、12、13、15、16或17的氨基酸1-736的氨基酸序列具有至少80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%序列一致性的氨基酸序列,其中:衣壳蛋白中对应于SEQ ID NO:16的氨基酸2的氨基酸为T;衣

壳蛋白中对应于SEQ ID NO:16的氨基酸65的氨基酸为I;衣壳蛋白中对应于SEQ ID NO:16的氨基酸68的氨基酸为V;衣壳蛋白中对应于SEQ ID NO:16的氨基酸77的氨基酸为R;衣壳蛋白中对应于SEQ ID NO:16的氨基酸119的氨基酸为L;衣壳蛋白中对应于SEQ ID NO:16的氨基酸151的氨基酸为R;衣壳蛋白中对应于SEQ ID NO:16的氨基酸160的氨基酸为D;衣壳蛋白中对应于SEQ ID NO:16的氨基酸206的氨基酸为C;衣壳蛋白中对应于SEQ ID NO:16的氨基酸296的氨基酸为H;衣壳蛋白中对应于SEQ ID NO:16的氨基酸312的氨基酸为Q;衣壳蛋白中对应于SEQ ID NO:16的氨基酸346的氨基酸为A;衣壳蛋白中对应于SEQ ID NO:16的氨基酸464的氨基酸为N;衣壳蛋白中对应于SEQ ID NO:16的氨基酸468的氨基酸为S;衣壳蛋白中对应于SEQ ID NO:16的氨基酸501的氨基酸为I;衣壳蛋白中对应于SEQ ID NO:16的氨基酸505的氨基酸为R;衣壳蛋白中对应于SEQ ID NO:16的氨基酸590的氨基酸为R;衣壳蛋白中对应于SEQ ID NO:16的氨基酸626的氨基酸为G或Y;衣壳蛋白中对应于SEQ ID NO:16的氨基酸681的氨基酸为M;衣壳蛋白中对应于SEQ ID NO:16的氨基酸687的氨基酸为R;衣壳蛋白中对应于SEQ ID NO:16的氨基酸690的氨基酸为K;衣壳蛋白中对应于SEQ ID NO:16的氨基酸706的氨基酸为C;或衣壳蛋白中对应于SEQ ID NO:16的氨基酸718的氨基酸为G。在某些实施例中,衣壳蛋白中对应于SEQ ID NO:16的氨基酸2的氨基酸为T,并且衣壳蛋白中对应于SEQ ID NO:16的氨基酸312的氨基酸为Q。在某些实施例中,衣壳蛋白中对应于SEQ ID NO:16的氨基酸65的氨基酸为I,并且衣壳蛋白中对应于SEQ ID NO:16的氨基酸626的氨基酸为Y。在某些实施例中,衣壳蛋白中对应于SEQ ID NO:16的氨基酸77的氨基酸为R,并且衣壳蛋白中对应于SEQ ID NO:16的氨基酸690的氨基酸为K。在某些实施例中,衣壳蛋白中对应于SEQ ID NO:16的氨基酸119的氨基酸为L,并且衣壳蛋白中对应于SEQ ID NO:16的氨基酸468的氨基酸为S。在某些实施例中,衣壳蛋白中对应于SEQ ID NO:16的氨基酸626的氨基酸为G,并且衣壳蛋白中对应于SEQ ID NO:16的氨基酸718的氨基酸为G。在某些实施例中,衣壳蛋白中对应于SEQ ID NO:16的氨基酸296的氨基酸为H,衣壳蛋白中对应于SEQ ID NO:16的氨基酸464的氨基酸为N,衣壳蛋白中对应于SEQ ID NO:16的氨基酸505的氨基酸为R,并且衣壳蛋白中对应于SEQ ID NO:16的氨基酸681的氨基酸为M。在某些实施例中,衣壳蛋白中对应于SEQ ID NO:16的氨基酸505的氨基酸为R,并且衣壳蛋白中对应于SEQ ID NO:16的氨基酸687的氨基酸为R。在某些实施例中,衣壳蛋白中对应于SEQ ID NO:16的氨基酸346的氨基酸为A,并且衣壳蛋白中对应于SEQ ID NO:16的氨基酸505的氨基酸为R。在某些实施例中,衣壳蛋白中对应于SEQ ID NO:16的氨基酸501的氨基酸为I,衣壳蛋白中对应于SEQ ID NO:16的氨基酸505的氨基酸为R,并且衣壳蛋白中对应于SEQ ID NO:16的氨基酸706的氨基酸为C。在某些实施例中,衣壳蛋白包含SEQ ID NO:2、3、4、5、6、7、8、9、10、11、12、13、15、16或17的氨基酸1-736的氨基酸序列。

[0098] 在某些实施例中,AAV衣壳包含以下中的两个或更多个:(a)包含SEQ ID NO:2、3、4、6、7、10、11、12、13、15、16或17的氨基酸203-736的氨基酸序列的衣壳蛋白;(b)包含SEQ ID NO:2、3、4、5、6、7、9、10、11、12、13、15、16或17的氨基酸138-736的氨基酸序列的衣壳蛋白;和(c)包含SEQ ID NO:2、3、4、5、6、7、8、9、10、11、12、13、15、16或17的氨基酸1-736的氨基酸序列的衣壳蛋白。在某些实施例中,AAV衣壳包含:(a)具有由SEQ ID NO:2、3、4、6、7、10、11、12、13、15、16或17的氨基酸203-736组成的氨基酸序列的衣壳蛋白;(b)具有由SEQ ID NO:2、3、4、5、6、7、9、10、11、12、13、15、16或17的氨基酸138-736组成的氨基酸序列的衣

壳蛋白；和(c)具有由SEQ ID NO:2、3、4、5、6、7、8、9、10、11、12、13、15、16或17的氨基酸1-736组成的氨基酸序列的衣壳蛋白。

[0099] 在某些实施例中, AAV衣壳包含以下中的一种或多种: (a) 包含与SEQ ID NO:8的氨基酸序列203-736具有至少80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%序列一致性的氨基酸序列的衣壳蛋白; (b) 包含与SEQ ID NO:8的氨基酸序列138-736具有至少80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%序列一致性的氨基酸序列的衣壳蛋白; 和(c) 包含与SEQ ID NO:8的氨基酸序列1-736具有至少80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%序列一致性的氨基酸序列的衣壳蛋白。在某些实施例中, AAV衣壳包含以下中的一种或多种: (a) 包含SEQ ID NO:8的氨基酸203-736的氨基酸序列的衣壳蛋白; (b) 包含SEQ ID NO:8的氨基酸138-736的氨基酸序列的衣壳蛋白; 和(c) 包含SEQ ID NO:8的氨基酸1-736的氨基酸序列的衣壳蛋白。在某些实施例中, AAV衣壳包含以下中的两种或更多种: (a) 包含SEQ ID NO:8的氨基酸203-736的氨基酸序列的衣壳蛋白; (b) 包含SEQ ID NO:8的氨基酸138-736的氨基酸序列的衣壳蛋白; 和(c) 包含SEQ ID NO:8的氨基酸1-736的氨基酸序列的衣壳蛋白。在某些实施例中, AAV衣壳包含: (a) 具有由SEQ ID NO:8的氨基酸203-736组成的氨基酸序列的衣壳蛋白; (b) 具有由SEQ ID NO:8的氨基酸138-736组成的氨基酸序列的衣壳蛋白; 和(c) 具有由SEQ ID NO:8的氨基酸1-736组成的氨基酸序列的衣壳蛋白。

[0100] 在某些实施例中, AAV衣壳包含以下中的一种或多种: (a) 包含与SEQ ID NO:11的氨基酸序列203-736具有至少80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%序列一致性的氨基酸序列的衣壳蛋白; (b) 包含与SEQ ID NO:11的氨基酸序列138-736具有至少80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%序列一致性的氨基酸序列的衣壳蛋白; 和(c) 包含与SEQ ID NO:11的氨基酸序列1-736具有至少80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%序列一致性的氨基酸序列的衣壳蛋白。在某些实施例中, AAV衣壳包含以下中的一种或多种: (a) 包含SEQ ID NO:11的氨基酸203-736的氨基酸序列的衣壳蛋白; (b) 包含SEQ ID NO:11的氨基酸138-736的氨基酸序列的衣壳蛋白; 和(c) 包含SEQ ID NO:11的氨基酸1-736的氨基酸序列的衣壳蛋白。在某些实施例中, AAV衣壳包含以下中的两种或更多种: (a) 包含SEQ ID NO:11的氨基酸203-736的氨基酸序列的衣壳蛋白; (b) 包含SEQ ID NO:11的氨基酸138-736的氨基酸序列的衣壳蛋白; 和(c) 包含SEQ ID NO:11的氨基酸1-736的氨基酸序列的衣壳蛋白。在某些实施例中, AAV衣壳包含: (a) 具有由SEQ ID NO:11的氨基酸203-736组成的氨基酸序列的衣壳蛋白; (b) 具有由SEQ ID NO:11的氨基酸138-736组成的氨基酸序列的衣壳蛋白; 和(c) 具有由SEQ ID NO:11的氨基酸1-736组成的氨基酸序列的衣壳蛋白。

[0101] 在某些实施例中, AAV衣壳包含以下中的一种或多种: (a) 包含与SEQ ID NO:13的氨基酸序列203-736具有至少80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%序列一致性的氨基酸序列的衣

壳蛋白；(b) 包含与SEQ ID NO:13的氨基酸序列138-736具有至少80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%序列一致性的氨基酸序列的衣壳蛋白；和(c) 包含与SEQ ID NO:13的氨基酸序列1-736具有至少80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%序列一致性的氨基酸序列的衣壳蛋白。在某些实施例中, AAV衣壳包含以下中的一种或多种: (a) 包含SEQ ID NO:13的氨基酸203-736的氨基酸序列的衣壳蛋白；(b) 包含SEQ ID NO:13的氨基酸138-736的氨基酸序列的衣壳蛋白；和(c) 包含SEQ ID NO:13的氨基酸1-736的氨基酸序列的衣壳蛋白。在某些实施例中, AAV衣壳包含以下中的两种或更多种: (a) 包含SEQ ID NO:13的氨基酸203-736的氨基酸序列的衣壳蛋白；(b) 包含SEQ ID NO:13的氨基酸138-736的氨基酸序列的衣壳蛋白；和(c) 包含SEQ ID NO:13的氨基酸1-736的氨基酸序列的衣壳蛋白。在某些实施例中, AAV衣壳包含: (a) 具有由SEQ ID NO:13的氨基酸203-736组成的氨基酸序列的衣壳蛋白；(b) 具有由SEQ ID NO:13的氨基酸138-736组成的氨基酸序列的衣壳蛋白；和(c) 具有由SEQ ID NO:13的氨基酸1-736组成的氨基酸序列的衣壳蛋白。

[0102] 在某些实施例中, AAV衣壳包含以下中的一种或多种: (a) 包含与SEQ ID NO:16的氨基酸序列203-736具有至少80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%序列一致性的氨基酸序列的衣壳蛋白；(b) 包含与SEQ ID NO:16的氨基酸序列138-736具有至少80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%序列一致性的氨基酸序列的衣壳蛋白；和(c) 包含与SEQ ID NO:16的氨基酸序列1-736具有至少80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%序列一致性的氨基酸序列的衣壳蛋白。在某些实施例中, AAV衣壳包含以下中的一种或多种: (a) 包含SEQ ID NO:16的氨基酸203-736的氨基酸序列的衣壳蛋白；(b) 包含SEQ ID NO:16的氨基酸138-736的氨基酸序列的衣壳蛋白；和(c) 包含SEQ ID NO:16的氨基酸1-736的氨基酸序列的衣壳蛋白。在某些实施例中, AAV衣壳包含以下中的两种或更多种: (a) 包含SEQ ID NO:16的氨基酸203-736的氨基酸序列的衣壳蛋白；(b) 包含SEQ ID NO:16的氨基酸138-736的氨基酸序列的衣壳蛋白；和(c) 包含SEQ ID NO:16的氨基酸1-736的氨基酸序列的衣壳蛋白。在某些实施例中, AAV衣壳包含: (a) 具有由SEQ ID NO:16的氨基酸203-736组成的氨基酸序列的衣壳蛋白；(b) 具有由SEQ ID NO:16的氨基酸138-736组成的氨基酸序列的衣壳蛋白；和(c) 具有由SEQ ID NO:16的氨基酸1-736组成的氨基酸序列的衣壳蛋白。

[0103] 适用于本文中所公开的AAV组合物中的转移基因组通常包含可操作地连接于ARSA编码序列的转录调节元件(TRE)。在某些实施例中, 转移基因组包含TRE和ARSA编码序列5'的5'反向末端重复(5' ITR)核苷酸序列, 以及TRE和ARSA编码序列3'的3'反向末端重复(3' ITR)核苷酸序列。

[0104] 在某些实施例中, ARSA编码序列包含ARSA基因的全部或大体上全部编码序列。在某些实施例中, 转移基因组包含编码SEQ ID NO:23的核苷酸序列并可任选地还包含位于ARSA编码序列3'的外源性聚腺苷酸化序列。在某些实施例中, 编码SEQ ID NO:23的核苷酸序列为野生型的(例如具有SEQ ID NO:24中所阐述的序列)。在某些实施例中, 编码SEQ ID

NO:23的核苷酸序列是以沉默方式改变(例如具有SEQ ID NO:14、62或72中所阐述的序列)。

[0105] 在某些实施例中,ARSA编码序列编码包含ARSA蛋白质的全部或大体上全部氨基酸序列的多肽。在某些实施例中,ARSA编码序列编码野生型ARSA蛋白质(例如人类ARSA蛋白质)的氨基酸序列。在某些实施例中,ARSA编码序列编码突变型ARSA蛋白质(例如人类ARSA蛋白质)的氨基酸序列,其中突变型ARSA多肽为野生型ARSA多肽的功能等效物,即可充当野生型ARSA多肽。在某些实施例中,功能上等效的ARSA多肽还包含至少一种未在野生型ARSA多肽中发现的特征,例如抵抗蛋白质降解的能力。

[0106] 在某些实施例中,适用于本文中所公开的AAV组合物中的转移基因组通常包含转录调节元件(TRE),其可操作地连接于编码ARSA和/或SUMF1的编码序列。硫酸酯酶修饰因子1(SUMF1)基因编码一种酶,所述酶通过使底物硫酸酯酶中的半胱氨酸残基氧化成活性位点3-氧代丙氨酸残基(其亦称为C- α -甲酰甘氨酸)来催化硫酸酯的水解。与SUMF1相关的疾病包括多种硫酸酯酶缺乏症和异染性脑白质营养不良。

[0107] 在某些实施例中,SUMF1编码序列包含SUMF1基因的全部或大体上全部编码序列。在某些实施例中,转移基因组包含编码SEQ ID NO:29的核苷酸序列并可任选地还包含位于SUMF1编码序列3'的外源性聚腺苷酸化序列。在某些实施例中,编码SEQ ID NO:29的核苷酸序列为野生型(例如具有SEQ ID NO:64中所阐述的序列)。在某些实施例中,编码SEQ ID NO:29的核苷酸序列是以沉默方式改变。

[0108] 在某些实施例中,SUMF1编码序列编码包含SUMF1蛋白质的全部或大体上全部氨基酸序列的多肽。在某些实施例中,SUMF1编码序列编码野生型SUMF1蛋白质(例如人类SUMF1蛋白质(hSUMF1))的氨基酸序列。在某些实施例中,SUMF1编码序列编码突变型SUMF1蛋白质(例如人类SUMF1蛋白质)的氨基酸序列,其中突变型SUMF1多肽为野生型SUMF1多肽的功能等效物,即可充当野生型SUMF1多肽。在某些实施例中,功能上等效的SUMF1多肽还包含至少一种未在野生型SUMF1多肽中发现的特征,例如抵抗蛋白质降解的能力。

[0109] 在某些实施例中,转移基因组被设计成表达hARSA和hSUMF1并包含核苷酸序列,所述核苷酸序列包含编码hARSA的第一编码序列和编码hSUMF1的第二编码序列。在某些实施例中,编码hARSA的第一编码序列和编码hSUMF1的第二编码序列是由核糖体跳跃元件分离。可使用所属领域中已知的任何核糖体跳跃元件,例如本文中其它地方描述的核糖体跳跃元件。在某些实施例中,包含编码hARSA的第一编码序列和编码hSUMF1的第二编码序列的核苷酸序列包含SEQ ID NO:30中所阐述的核苷酸序列。

[0110] 在某些实施例中,适用于本文中所公开的AAV组合物中的转移基因组通常包含转录调节元件(TRE),其可操作地连接于编码ARSA和/或SapB的编码序列。鞘脂激活蛋白原(PSAP)基因编码高保守性前原蛋白,所述前原蛋白被蛋白水解处理以产生四种主要裂解产物,包括鞘脂激活蛋白A、B、C和D。前体蛋白质的每个结构域的长度为大致80个氨基酸残基,其中半胱氨酸残基和糖基化位点的放置几乎相同。鞘脂激活蛋白A-D主要位于溶酶体隔室,其在溶酶体隔室中促进具有短寡糖基团的鞘糖脂的分解代谢。前体蛋白质以分泌蛋白质和整合膜蛋白质两者的形式存在并且具有神经营养活性。此基因中的突变与戈谢病(Gaucher disease)和异染性脑白质营养不良相关。已表明鞘脂激活蛋白B(SapB)可刺激半乳糖-脑苷脂硫酸酯由ARSA进行的水解、GM1神经节苷脂由 β -半乳糖苷酶进行的水解和球形三酰神经酰胺由 α -半乳糖苷酶A进行的水解。已表明SapB与鞘脂水解酶的底物形成溶解复合物。

[0111] 在某些实施例中, SapB编码序列包含SapB基因的全部或大体上全部编码序列。在某些实施例中, 转移基因组包含编码SEQ ID NO:33的核苷酸序列并可任选地还包含位于SapB编码序列3'的外源性聚腺苷酸化序列。在某些实施例中, 编码SEQ ID NO:33的核苷酸序列为野生型(例如具有SEQ ID NO:73中所阐述的序列)。在某些实施例中, 编码SEQ ID NO:33的核苷酸序列是以沉默方式改变。

[0112] 在某些实施例中, SapB编码序列编码包含SapB蛋白质的全部或大体上全部氨基酸序列的多肽。在某些实施例中, SapB编码序列编码野生型SapB蛋白质(例如人类SapB蛋白质(hSapB))的氨基酸序列。在某些实施例中, SapB编码序列编码突变型SapB蛋白质(例如人类SapB蛋白质)的氨基酸序列, 其中突变型SapB多肽为野生型SapB多肽的功能等效物, 即, 可充当野生型SapB多肽。在某些实施例中, 功能上等效的SapB多肽还包含至少一种未在野生型SapB多肽中发现的特征, 例如抵抗蛋白质降解的能力。

[0113] 在某些实施例中, 转移基因组被设计成表达hARSA和hSapB并包含核苷酸序列, 所述核苷酸序列包含编码hARSA的第一编码序列和编码hSapB的第二编码序列。在某些实施例中, 编码hARSA的第一编码序列和编码hSapB的第二编码序列是由核糖体跳跃元件分离。可使用所属领域中已知的任何核糖体跳跃元件, 例如本文中其它地方描述的核糖体跳跃元件。在某些实施例中, 包含编码hARSA的第一编码序列和编码hSapB的第二编码序列的核苷酸序列包含SEQ ID NO:74中所阐述的核苷酸序列。

[0114] 转移基因组可用于表达任何哺乳动物细胞(例如人类细胞)中的ARSA、SUMF1和/或SapB。因此, TRE可在任何哺乳动物细胞(例如人类细胞)中具有活性。在某些实施例中, TRE在广泛的人类细胞中具有活性。此类TRE可包含组成型启动子和/或增强子元件, 包括巨细胞病毒(CMV)启动子/强化子(例如包含与SEQ ID NO:58至少90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%一致的核苷酸序列)、SV40启动子、鸡 β 肌动蛋白(CBA)启动子(例如包含与SEQ ID NO:59或25至少90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%一致的核苷酸序列)、smCBA启动子(例如包含与SEQ ID NO:55至少90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%一致的核苷酸序列)、人类延长因子1 α (EF1 α)启动子(例如包含与SEQ ID NO:40至少90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%一致的核苷酸序列)、包含转录因子结合位点的小鼠细小病毒(MVM)内含子(例如包含与SEQ ID NO:35至少90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%一致的核苷酸序列)、人类磷酸甘油酸激酶(PGK1)启动子、人类泛蛋白C(Ubc)启动子、人类 β 肌动蛋白启动子、人类神经元特异性烯醇酶(ENO2)启动子、人类 β -葡萄糖苷酸酶(GUSB)启动子、兔 β -球蛋白元件(例如包含与SEQ ID NO:60至少90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%一致的核苷酸序列)、人类钙调蛋白1(CALM1)启动子(例如包含与SEQ ID NO:54至少90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%一致的核苷酸序列)和/或人类甲基-CpG结合蛋白质2(MeCP2)启动子。这些TRE中的任一个可以任何顺序组合以驱动有效转录。例如, 转移基因组可包含CMV增强子、CBA启动子和来自兔 β -球蛋白基因的外显子3的剪接受体, 统称为CAG启动子(例如包含与SEQ ID NO:28至少90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%一致的核苷酸序列)。例如, 转移基因组可包含CMV增强子和CBA启动子的杂交物, 接着为剪接供体和剪接受体, 统称为CASI启动子区域(例如包含与SEQ ID NO:63至少90%、91%、

92%、93%、94%、95%、96%、97%、98%、99%或100%一致的核苷酸序列)。

[0115] 或者, TRE可为组织特异性TRE, 即, 其在特定组织和/或器官中具有活性。组织特异性TRE包含一个或多个组织特异性启动子和/或增强子元件, 和任选地一个或多个组成型启动子和/或增强子元件。所属领域的技术人员将理解, 可通过所属领域中熟知的方法从组织中特异性表达的基因分离组织特异性启动子和/或增强子元件。

[0116] 在某些实施例中, TRE为脑特异性(例如神经元特异性、神经胶质细胞特异性、星形胶质细胞特异性、少突胶质细胞特异性、微神经胶质细胞特异性和/或中枢神经系统特异性)。示例性脑特异性TRE可包含一个或多个来自(但不限于)人类胶质细胞原纤维酸性蛋白(GFAP)启动子、人类突触蛋白1(SYN1)启动子、人类突触蛋白2(SYN2)启动子、人类金属硫蛋白3(MT3)启动子和/或人类蛋白脂质蛋白质1(PLP1)启动子的元件。更多的脑特异性启动子元件公开于W0 2016/100575A1中, 其以全文引用的方式并入本文中。

[0117] 在某些实施例中, 转移基因组包含两个或更多个TRE, 任选地包含以上所公开的TRE中的至少一个。所属领域的技术人员将理解, 这些TRE中的任一个可以任何顺序组合, 并且组成型TRE和组织特异性TRE的组合可驱动有效和组织特异性转录。

[0118] 在某些实施例中, 转移载体还包含非编码填充序列(例如包含与SEQ ID NO:39至少90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%一致的核苷酸序列)。非编码填充序列可用于将载体的尺寸保持在适当的限制内以用于有效DNA包装, 并且因此可用于提高DNA包装的功效。所属领域的技术人员将认识到, 填充序列的性质可对载体的功能具有影响, 并且因此将选择最合适的填充序列来使用。

[0119] 在某些实施例中, 转移载体还包含位于ARSA编码序列5'或插入ARSA编码序列中的内含子。此类内含子可增加转基因表达, 例如藉由减少转录沉默和增强从细胞核到细胞质的mRNA输出。在某些实施例中, 转移基因组从5'到3'包含: 非编码外显子、内含子和ARSA编码序列。在某些实施例中, 将内含子序列插入ARSA编码序列中, 任选地其中在连接两个原生外显子的核苷酸间键处插入内含子。在某些实施例中, 在连接原生外显子1与外显子2的核苷酸间键处插入内含子。

[0120] 内含子可包含ARSA基因的原生内含子序列、来自不同物种的内含子序列或来自相同物种的不同基因和/或合成内含子序列。所属领域的技术人员将了解, 合成内含子序列可被设计成藉由引入所属领域中已知的任何共有剪接基元来介导RNA剪接(例如Sibley等人, (2016)《自然综述遗传学(Nature Reviews Genetics)》, 17, 407-21, 其以全文引用的方式并入本文中)。示例性内含子序列提供于Lu等人(2013)《分子疗法(Molecular Therapy)》21(5):954-63, 和Lu等人(2017)《人类基因疗法(Hum. Gene Ther.)》, 28(1):125-34中, 其以全文引用的方式并入本文中。在某些实施例中, 转移基因组包含SV40内含子(例如包含与SEQ ID NO:31至少90%、91%、92%、93%、94%、95%、96%、97%、98%或99%一致的核苷酸序列)或小鼠细小病毒(MVM)内含子(例如包含与SEQ ID NO:35至少90%、91%、92%、93%、94%、95%、96%、97%、98%或99%一致的核苷酸序列)。在某些实施例中, 转移基因组包含SV40内含子(例如包含SEQ ID NO:31中所阐述的核苷酸序列)或小鼠细小病毒(MVM)内含子(例如包含SEQ ID NO:35中所阐述的核苷酸序列)。在某些实施例中, 转移基因组包含嵌合内含子序列, 其包含鸡和兔序列的组合, 部分包含未转录的鸡ACTB(cACTB)启动子、全部cACTB外显子1、部分cACTB内含子1、部分兔HBB2(rHBB2)内含子2和部分rHBB2外显子3(例如

SEQ ID NO:32)。在某些实施例中,转移基因组包含嵌合内含子序列(例如包含与SEQ ID NO:32至少90%、91%、92%、93%、94%、95%、96%、97%、98%或99%一致的核苷酸序列)。在某些实施例中,转移基因组包含嵌合内含子序列(例如包含SEQ ID NO:32中所阐述的核苷酸序列)。

[0121] 在某些实施例中,转移基因组包含TRE,其包含CMV增强子、CBA启动子和嵌合内含子序列(例如包含与SEQ ID NO:36至少90%、91%、92%、93%、94%、95%、96%、97%、98%或99%一致的核苷酸序列)。在某些实施例中,转移基因组包含有包含SEQ ID NO:36的TRE。

[0122] 在某些实施例中,本文中所公开的转移基因组还包含转录终止子(例如聚腺苷酸化序列)。在某些实施例中,转录终止子位于ARSA编码序列的3'。转录终止子可为任何有效终止转录的序列并且所属领域的技术人员将理解,可从任何表达于需要ARSA编码序列的转录的细胞中的基因分离此类序列。在某些实施例中,转录终止子包含聚腺苷酸化序列。在某些实施例中,聚腺苷酸化序列与人类ARSA基因的内源性聚腺苷酸化序列一致或大体上一致。在某些实施例中,聚腺苷酸化序列是外源性聚腺苷酸化序列。在某些实施例中,聚腺苷酸化序列是SV40聚腺苷酸化序列(例如包含SEQ ID NO:31、42、43或45中所阐述的核苷酸序列,或与其互补的核苷酸序列)。在某些实施例中,聚腺苷酸化序列包含SEQ ID NO:42中所阐述的序列。

[0123] 在某些实施例中,转移基因组从5'到3'包含:TRE、ARSA编码序列和聚腺苷酸化序列。在某些实施例中,TRE与SEQ ID NO:25、32、36、54、55和/或58中的任一个具有至少90%、91%、92%、93%、94%、95%、96%、97%、98%或99%序列一致性;ARSA编码序列与SEQ ID NO:14、24、62或72具有至少90%、91%、92%、93%、94%、95%、96%、97%、98%或99%序列一致性;和/或聚腺苷酸化序列与SEQ ID NO:42、43和45中的任一个具有至少90%、91%、92%、93%、94%、95%、96%、97%、98%或99%序列一致性。

[0124] 在某些实施例中,TRE包含SEQ ID NO:36中所阐述的序列;ARSA编码序列包含SEQ ID NO:14中所阐述的序列;和/或聚腺苷酸化序列包含SEQ ID NO:42中所阐述的序列。在某些实施例中,TRE从5'到3'包含SEQ ID NO:58中所阐述的序列、SEQ ID NO:25中所阐述的序列和SEQ ID NO:32中所阐述的序列。

[0125] 在某些实施例中,TRE包含SEQ ID NO:54中所阐述的序列;ARSA编码序列包含SEQ ID NO:62中所阐述的序列;和/或聚腺苷酸化序列包含SEQ ID NO:42中所阐述的序列。在某些实施例中,TRE包含SEQ ID NO:55中所阐述的序列;ARSA编码序列包含SEQ ID NO:62中所阐述的序列;和/或聚腺苷酸化序列包含SEQ ID NO:42中所阐述的序列。

[0126] 在某些实施例中,TRE包含SEQ ID NO:36中所阐述的序列;ARSA编码序列包含SEQ ID NO:72中所阐述的序列;和/或聚腺苷酸化序列包含SEQ ID NO:42中所阐述的序列。在某些实施例中,TRE从5'到3'包含SEQ ID NO:58中所阐述的序列、SEQ ID NO:25中所阐述的序列和SEQ ID NO:32中所阐述的序列。

[0127] 在某些实施例中,转移基因组还包含hSUMF1编码序列。在某些实施例中,转移基因组从5'到3'包含:TRE、ARSA编码序列、2A元件和hSUMF1编码序列。在某些实施例中,TRE与SEQ ID NO:25、32、36、54、55和/或58具有至少90%、91%、92%、93%、94%、95%、96%、97%、98%或99%序列一致性;ARSA编码序列与SEQ ID NO:62具有至少90%、91%、92%、93%、94%、95%、96%、97%、98%或99%序列一致性;2A组件与SEQ ID NO:63具有至少

90%、91%、92%、93%、94%、95%、96%、97%、98%或99%序列一致性；并且hSUMF1序列与SEQ ID NO:64具有至少90%、91%、92%、93%、94%、95%、96%、97%、98%或99%序列一致性。在某些实施例中，还包含hSUMF1编码序列的转移基因组从5'到3'包含：包含SEQ ID NO:54或55中所阐述的序列的TRE、包含SEQ ID NO:62中所阐述的序列的hARSA编码序列、包含SEQ ID NO:63中所阐述的序列的2A元件和包含SEQ ID NO:64中所阐述的序列的hSUMF1编码序列。在某些实施例中，hARSA-2A-hSUMF1编码序列包含SEQ ID NO:30中所阐述的序列。

[0128] 在某些实施例中，转移基因组还包含hSapB编码序列。在某些实施例中，转移基因组从5'到3'包含：TRE、ARSA编码序列、2A元件和hSapB编码序列。在某些实施例中，TRE与SEQ ID NO:25、32、36、54、55和/或58具有至少90%、91%、92%、93%、94%、95%、96%、97%、98%或99%序列一致性；ARSA编码序列与SEQ ID NO:72具有至少90%、91%、92%、93%、94%、95%、96%、97%、98%或99%序列一致性；2A元件与SEQ ID NO:63具有至少90%、91%、92%、93%、94%、95%、96%、97%、98%或99%序列一致性；并且hSapB序列与SEQ ID NO:73具有至少90%、91%、92%、93%、94%、95%、96%、97%、98%或99%序列一致性。在某些实施例中，还包含hSapB编码序列的转移基因组从5'到3'包含：包含SEQ ID NO:36中所阐述的序列的TRE、包含SEQ ID NO:72中所阐述的序列的hARSA编码序列、包含SEQ ID NO:63中所阐述的序列的2A元件和包含SEQ ID NO:74中所阐述的序列的hSapB编码序列。在某些实施例中，hARSA-2A-hSapB编码序列包含SEQ ID NO:74中所阐述的序列。

[0129] 在某些实施例中，转移基因组包含与SEQ ID NO:41、44、46、65、67或75至少80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%一致的序列。在某些实施例中，转移基因组包含SEQ ID NO:41、44、46、65、67或75中所阐述的核苷酸序列。在某些实施例中，转移基因组的核苷酸序列由SEQ ID NO:41、44、46、65、67或75中所阐述的核苷酸序列组成。在某些实施例中，转移基因组包含SEQ ID NO:44中所阐述的核苷酸序列。在某些实施例中，转移基因组的核苷酸序列由SEQ ID NO:44中所阐述的核苷酸序列组成。

[0130] 在某些实施例中，本文中所公开的转移基因组还包含TRE5'的5'反向末端重复(5' ITR)核苷酸序列，和ARSA编码序列3'的3'反向末端重复(3' ITR)核苷酸序列。来自任何AAV血清型或其变体的ITR序列均可用于本文中所公开的转移基因组中。5'和3' ITR可来自相同血清型的AAV或不同血清型的AAV。用于本文中所公开的转移基因组中的示例性ITR阐述于本文中的SEQ ID NO:18-21、26和27中。

[0131] 在某些实施例中，5' ITR或3' ITR来自AAV2。在某些实施例中，5' ITR和3' ITR均来自AAV2。在某些实施例中，5' ITR核苷酸序列与SEQ ID NO:18具有至少90% (例如至少95%、至少96%、至少97%、至少98%、至少99%或100%) 序列一致性，或3' ITR核苷酸序列与SEQ ID NO:19具有至少90% (例如至少95%、至少96%、至少97%、至少98%、至少99%或100%) 序列一致性。在某些实施例中，5' ITR核苷酸序列与SEQ ID NO:18具有至少90% (例如至少95%、至少96%、至少97%、至少98%、至少99%或100%) 序列一致性，并且3' ITR核苷酸序列与SEQ ID NO:19具有至少90% (例如至少95%、至少96%、至少97%、至少98%、至少99%或100%) 序列一致性。在某些实施例中，转移基因组包含SEQ ID NO:41、44、46、65、67或75中的任一个中所阐述的核苷酸序列，5' ITR核苷酸序列具有SEQ ID NO:18的序列，并且3' ITR核苷酸序列具有SEQ ID NO:19的序列。

[0132] 在某些实施例中,5' ITR或3' ITR来自AAV5。在某些实施例中,5' ITR和3' ITR均来自AAV5。在某些实施例中,5' ITR核苷酸序列与SEQ ID NO:20具有至少90% (例如至少95%、至少96%、至少97%、至少98%、至少99%或100%) 序列一致性,或3' ITR核苷酸序列与SEQ ID NO:21具有至少90% (例如至少95%、至少96%、至少97%、至少98%、至少99%或100%) 序列一致性。在某些实施例中,5' ITR核苷酸序列与SEQ ID NO:20具有至少90% (例如至少95%、至少96%、至少97%、至少98%、至少99%或100%) 序列一致性,并且3' ITR核苷酸序列与SEQ ID NO:21具有至少90% (例如至少95%、至少96%、至少97%、至少98%、至少99%或100%) 序列一致性。在某些实施例中,转移基因组包含SEQ ID NO:46-50中的任一个中所阐述的核苷酸序列、具有SEQ ID NO:20的序列的5' ITR核苷酸序列和具有SEQ ID NO:21的序列的3' ITR核苷酸序列。

[0133] 在某些实施例中,5' ITR核苷酸序列和3' ITR核苷酸序列彼此大体上互补 (例如除在5'或3' ITR中的1、2、3、4或5个核苷酸位置处的失配以外彼此互补)。

[0134] 在某些实施例中,5' ITR或3' ITR被修饰以降低或消除被Rep蛋白解析(resolution) (“不可解析的ITR”)。在某些实施例中,不可解析的ITR包含在末端解析位点的核苷酸序列中的插入、缺失或取代。此类修饰允许在转移基因组在感染的细胞中复制后形成AAV的自身互补型双链DNA基因组。示例性不可解析的ITR序列是所属领域中已知的 (参见例如美国专利案第7,790,154号和第9,783,824号中提供的序列,其以全文引用的方式并入本文中)。在某些实施例中,5' ITR包含与SEQ ID NO:26至少95%、96%、97%、98%或99%一致的核苷酸序列。在某些实施例中,5' ITR由与SEQ ID NO:26至少95%、96%、97%、98%或99%一致的核苷酸序列组成。在某些实施例中,5' ITR由SEQ ID NO:26中所阐述的核苷酸序列组成。在某些实施例中,3' ITR包含与SEQ ID NO:27至少95%、96%、97%、98%或99%一致的核苷酸序列。在某些实施例中,5' ITR由与SEQ ID NO:27至少95%、96%、97%、98%或99%一致的核苷酸序列组成。在某些实施例中,3' ITR由SEQ ID NO:27中所阐述的核苷酸序列组成。在某些实施例中,5' ITR由SEQ ID NO:26中所阐述的核苷酸序列组成,并且3' ITR由SEQ ID NO:27中所阐述的核苷酸序列组成。在某些实施例中,5' ITR由SEQ ID NO:26中所阐述的核苷酸序列组成,并且3' ITR由SEQ ID NO:19中所阐述的核苷酸序列组成。

[0135] 在某些实施例中,3' ITR由来源于野生型AAV2基因组序列的其它核苷酸序列侧接。在某些实施例中,3' ITR由来源于与野生型AAV2 ITR相邻的野生型AAV2序列的另一个37bp序列侧接。参见例如Savy等人,《人类基因疗法方法 (Human Gene Therapy Methods)》(2017) 28 (5):277-289 (其以全文引用的方式并入本文中)。在某些实施例中,另外的37bp序列位于3' ITR内部。在某些实施例中,37bp序列由SEQ ID NO:56中所阐述的序列组成。在某些实施例中,3' ITR包含与SEQ ID NO:57至少95%、96%、97%、98%或99%一致的核苷酸序列。在某些实施例中,3' ITR包含SEQ ID NO:57中所阐述的核苷酸序列。在某些实施例中,3' ITR的核苷酸序列由与SEQ ID NO:57至少95%、96%、97%、98%或99%一致的核苷酸序列组成。在某些实施例中,3' ITR的核苷酸序列由SEQ ID NO:57中所阐述的核苷酸序列组成。

[0136] 在某些实施例中,转移基因组从5'到3'包含:5' ITR;内部组件,其从5'到3'包含:TRE、任选地非编码外显子和内含子、ARSA编码序列和聚腺苷酸化序列,如本文中所公开;不可解析的ITR;与内部元件互补的核苷酸序列;和3' ITR。此类转移基因组可在感染之后和复制之前形成AAV的自互补、双链DNA基因组。

[0137] 在某些实施例中,转移基因组从5'到3'包含:5' ITR、TRE、ARSA编码序列、聚腺苷酸化序列和3' ITR。在某些实施例中,5' ITR与SEQ ID:18、20或26具有至少90%、91%、92%、93%、94%、95%、96%、97%、98%或99%序列一致性;TRE与SEQ ID NO:25、32、36、54、55和/或58中的任一个具有至少90%、91%、92%、93%、94%、95%、96%、97%、98%或99%序列一致性;ARSA编码序列与SEQ ID NO:14、24、62或72具有至少90%、91%、92%、93%、94%、95%、96%、97%、98%或99%序列一致性;聚腺苷酸化序列与SEQ ID NO:42、43和45中的任一个具有至少90%、91%、92%、93%、94%、95%、96%、97%、98%或99%序列一致性;和/或3' ITR与SEQ ID:19、21、27或57具有至少90%、91%、92%、93%、94%、95%、96%、97%、98%或99%序列一致性。在某些实施例中,5' ITR包含选自由SEQ ID NO:18、20和26组成的群组的核苷酸序列或由其组成;TRE包含选自由以下组成的群组的核苷酸序列:SEQ ID NO:25、32、36、54、55和/或58;ARSA编码序列包含SEQ ID NO:14、24、62或72中所阐述的序列;聚腺苷酸化序列包含选自由SEQ ID NO:42、43和45组成的群组的核苷酸序列;和/或3' ITR包含选自由SEQ ID NO:19、21、27或57组成的群组的核苷酸序列或由其组成。

[0138] 在某些实施例中,5' ITR包含SEQ ID NO:18中所阐述的序列或由其组成;TRE包含SEQ ID NO:36中所阐述的序列;ARSA编码序列包含SEQ ID NO:14、24、62或72中所阐述的序列;聚腺苷酸化序列包含SEQ ID NO:42中所阐述的序列;和/或3' ITR包含SEQ ID NO:19中所阐述的序列或由其组成。

[0139] 在某些实施例中,转移基因组包含与SEQ ID NO:47、48、49、68、69或76至少80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%一致的序列。在某些实施例中,转移基因组包含SEQ ID NO:47、48、49、68、69或76中所阐述的核苷酸序列。在某些实施例中,转移基因组的核苷酸序列由SEQ ID NO:47、48、49、68、69或76中所阐述的核苷酸序列组成。在某些实施例中,转移基因组包含SEQ ID NO:48中所阐述的核苷酸序列。在某些实施例中,转移基因组的核苷酸序列由SEQ ID NO:48中所阐述的核苷酸序列组成。

[0140] 在某些实施例中,rAAV包含:(a) AAV衣壳蛋白,其包含SEQ ID NO:16的氨基酸203-736的氨基酸序列,和转移基因组,其从5'到3'包含以下基因元件:5' ITR元件(例如SEQ ID NO:18的5' ITR)、增强子元件(例如SEQ ID NO:58的增强子元件)、启动子序列(例如SEQ ID NO:25的启动子序列)、嵌合内含子序列(例如SEQ ID NO:32的嵌合内含子序列)、以沉默方式改变的人类ARSA编码序列(例如SEQ ID NO:14的hARSA编码序列)、SV40聚腺苷酸化序列(例如SEQ ID NO:42的SV40聚腺苷酸化序列)和3' ITR元件(例如SEQ ID NO:19的3' ITR);(b) AAV衣壳蛋白,其包含SEQ ID NO:16的氨基酸138-736的氨基酸序列,和转移基因组,其从5'到3'包含以下基因元件:5' ITR元件(例如SEQ ID NO:18的5' ITR)、增强子元件(例如SEQ ID NO:58的增强子元件)、启动子序列(例如SEQ ID NO:25的启动子序列)、嵌合内含子序列(例如SEQ ID NO:32的嵌合内含子序列)、以沉默方式改变的人类ARSA编码序列(例如SEQ ID NO:14的hARSA编码序列)、SV40聚腺苷酸化序列(例如SEQ ID NO:42的SV40聚腺苷酸化序列)和3' ITR元件(例如SEQ ID NO:19的3' ITR);和/或(c) AAV衣壳蛋白,其包含SEQ ID NO:16的氨基酸序列,和转移基因组,其从5'到3'包含以下基因元件:5' ITR元件(例如SEQ ID NO:18的5' ITR)、增强子元件(例如SEQ ID NO:58的增强子元件)、启动子序列(例如SEQ ID NO:25的启动子序列)、嵌合内含子序列(例如SEQ ID NO:32的嵌合内含子序列)、以沉默方式改变的人类ARSA编码序列(例如SEQ ID NO:14的hARSA编码序列)、SV40聚腺苷酸化序列(例如SEQ ID NO:42的SV40聚腺苷酸化序列)和3' ITR元件(例如SEQ ID NO:19的3' ITR)。

化序列(例如SEQ ID NO:42的SV40聚腺苷酸化序列)和3' ITR元件(例如SEQ ID NO:19的3' ITR)。

[0141] 在某些实施例中,rAAV包含:(a) AAV衣壳蛋白,其包含SEQ ID NO:16的氨基酸203-736的氨基酸序列,和转移基因组,其包含SEQ ID NO:41、44、46、47、48、49、65、67、68、69、75或76中的任一个中所阐述的核苷酸序列;(b) AAV衣壳蛋白,其包含SEQ ID NO:16的氨基酸138-736的氨基酸序列,和转移基因组,其包含SEQ ID NO:41、44、46、47、48、49、65、67、68、69、75或76中的任一个中所阐述的核苷酸序列;和/或(c) AAV衣壳蛋白,其包含SEQ ID NO:16的氨基酸序列,和转移基因组,其包含SEQ ID NO:41、44、46、47、48、49、65、67、68、69、75或76中的任一个中所阐述的核苷酸序列。

[0142] 在另一方面中,本文中提供多核苷酸,其包含与SEQ ID NO:41、44、46、47、48、49、65、67、68、69、75或76中所阐述的核酸序列至少80% (例如至少85%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%)一致的核酸序列。在某些实施例中,多核苷酸包含SEQ ID NO:41、44、46、47、48、49、65、67、68、69、75或76中所阐述的核酸序列。在某些实施例中,多核苷酸由SEQ ID NO:41、44、46、47、48、49、65、67、68、69、75或76中所阐述的核酸序列组成。在某些实施例中,多核苷酸包含SEQ ID NO:44或48中所阐述的核酸序列。在某些实施例中,多核苷酸由SEQ ID NO:44或48中所阐述的核酸序列组成。

[0143] 本文中亦提供多核苷酸,其包含与SEQ ID NO:14、62或72中所阐述的核酸序列至少80% (例如至少85%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%)一致的核酸序列。在某些实施例中,多核苷酸包含SEQ ID NO:14、62或72中所阐述的核酸序列。在某些实施例中,多核苷酸由SEQ ID NO:14、62或72中所阐述的核酸序列组成。在某些实施例中,多核苷酸包含SEQ ID NO:14中所阐述的核酸序列。在某些实施例中,多核苷酸由SEQ ID NO:14中所阐述的核酸序列组成。

[0144] 在另一方面中,本公开提供一种药物组合物,其包含如本文中所公开的AAV和药学上可接受的赋形剂、佐剂、稀释剂、媒剂或载剂或其组合。“药学上可接受的载剂”包括任何在与组合物的活性成分组合时使所述成分保留生物活性并不引起破坏性生理学反应(如不希望的免疫反应)的材料。药学上可接受的载剂包括水、磷酸盐缓冲盐水、乳液(如油/水乳液)和湿润剂。包含此类载剂的组合物通过熟知的常规方法调配,如以下中所阐述的那些方法:《雷明顿氏医药科学(Remington's Pharmaceutical Sciences)》,当前版,马克出版公司(Mack Publishing Co.),宾夕法尼亚州伊斯顿(Easton Pa.)18042,USA;A.Gennaro(2000)《雷明顿:药剂学之科学及实践(Remington:The Science and Practice of Pharmacy)》,第20版,Lippincott,Williams,&Wilkins;《医药剂型及药物递送系统(Pharmaceutical Dosage Forms and Drug Delivery Systems)》(1999)H.C.Ansel等人,第7版,Lippincott,Williams,&Wilkins;及《医药赋形剂手册(Handbook of Pharmaceutical Excipients)》(2000)A.H.Kibbe等人,第3版,Amer.Pharmaceutical Assoc中所阐述的方法。

[0145] 在另一方面中,本公开提供多核苷酸,其包含编码人类ARSA蛋白质或其片段的编码序列,其中所述编码序列已以沉默方式改变以与野生型人类ARSA基因具有小于100% (例如小于95%、90%、85%、80%、75%、70%、65%、60%、55%或50%)一致性。在某些实施例中,多核苷酸包含SEQ ID NO:14、62或72中所阐述的序列。在某些实施例中,多核苷酸由SEQ

ID NO:14、62或72中所阐述的序列组成。多核苷酸可包含DNA、RNA、修饰的DNA、修饰的RNA或其组合。在某些实施例中，多核苷酸为表达载体。

[0146] III. 使用方法

[0147] 在另一方面中，本公开提供用于在细胞中表达ARSA多肽的方法。所述方法通常包含用如本文中所公开的rAAV转导细胞。此类方法在恢复ARSA表达方面是高效的。因此，在某些实施例中，本文中所公开的方法涉及用如本文中所公开的rAAV转导细胞。

[0148] 本文中所公开的方法可应用于任何在ARSA基因中具有突变的细胞。所属领域中的技术人员将了解，对需要活性内源性ARSA的细胞尤其感兴趣。因此，在某些实施例中，将所述方法应用于任何损失内源性ARSA活性的细胞。在某些实施例中，将所述方法应用于神经元和/或神经胶质细胞。在某些实施例中，尤其感兴趣的是需要活性内源性ARSA的神经元和/或神经胶质细胞。在某些实施例中，将所述方法应用于中枢神经系统的细胞和/或周围神经系统的细胞。在某些实施例中，尤其感兴趣的是需要活性内源性ARSA的中枢神经系统和/或周围神经系统的细胞。在某些实施例中，尤其感兴趣的是前脑、中脑、后脑、脊髓和其任何组合中的细胞。在某些实施例中，尤其感兴趣的是选自由以下组成的群组的中枢神经系统区域的细胞：脊髓、运动皮质、感觉皮质、丘脑、海马、壳核、小脑（例如小脑核）和其任何组合。在某些实施例中，尤其感兴趣的是脑中的脑桥和髓质、脊髓的上升束和其任何组合中的细胞。在某些实施例中，尤其感兴趣的是选自由以下组成的群组的中枢神经系统区域中的需要活性内源性ARSA的细胞：脊髓、运动皮质、感觉皮质、丘脑、海马、壳核、小脑（例如小脑核）和其任何组合。在某些实施例中，尤其感兴趣的是中枢神经系统（CNS）中的运动神经元和星形胶质细胞分布、CNS中的少突胶质细胞（上升纤维）、CNS中的大脑皮质的细胞群体以及周围神经系统（PNS）中之感觉神经元。在某些实施例中，尤其感兴趣的是少突胶质细胞，如脊髓的背束中的细胞。在某些实施例中，尤其感兴趣的是中枢神经系统中的神经胶质分布，包括（但不限于）星形胶质细胞、少突胶质细胞、施万细胞和其任何组合。在某些实施例中，尤其感兴趣的是运动神经元、星形胶质细胞、少突胶质细胞、中枢神经系统中的大脑皮质的细胞、周围神经系统中的感觉神经元、周围神经系统中的神经胶质细胞（例如施万细胞）和其任何组合。

[0149] 本文中所公开的方法可体外进行以用于研究目的或可离体或体内进行以用于治疗目的。

[0150] 在某些实施例中，待转导的细胞在哺乳动物受试者中并以可有效转导受试者中的细胞的量向受试者施用AAV。因此，在某些实施例中，本公开提供一种用于治疗患有与ARSA基因突变相关的疾病或病症的受试者的方法，所述方法通常包含向受试者施用有效量的如本文中所公开的rAAV。受试者可为具有ARSA突变的人类受试者、非人类灵长类动物受试者（例如食蟹猕猴）或啮齿动物受试者（例如小鼠）。可使用本文中所公开的方法治疗任何与ARSA基因突变相关的疾病或病症。合适的疾病或病症包括（但不限于）异染性脑白质营养不良。

[0151] 在某些实施例中，前述方法使用包含以下的rAAV：(a) AAV衣壳蛋白，其包含SEQ ID NO:16的氨基酸203-736的氨基酸序列，和转移基因组，其从5'到3'包含以下基因元件：5' ITR元件（例如SEQ ID NO:18的5' ITR）、增强子元件（例如SEQ ID NO:58的增强子元件）、启动子序列（例如SEQ ID NO:25的启动子序列）、嵌合内含子序列（例如SEQ ID NO:32的嵌合

内含子序列)、以沉默方式改变的人类ARSA编码序列(例如SEQ ID NO:14的hARSA编码序列)、SV40聚腺苷酸化序列(例如SEQ ID NO:42的SV40聚腺苷酸化序列)和3' ITR元件(例如SEQ ID NO:19的3' ITR); (b) AAV衣壳蛋白,其包含SEQ ID NO:16的氨基酸138-736的氨基酸序列,和转移基因组,其从5'到3'包含以下基因元件:5' ITR元件(例如SEQ ID NO:18的5' ITR)、增强子元件(例如SEQ ID NO:58的增强子元件)、启动子序列(例如SEQ ID NO:25的启动子序列)、嵌合内含子序列(例如SEQ ID NO:32的嵌合内含子序列)、以沉默方式改变的人类ARSA编码序列(例如SEQ ID NO:14的hARSA编码序列)、SV40聚腺苷酸化序列(例如SEQ ID NO:42的SV40聚腺苷酸化序列)和3' ITR元件(例如SEQ ID NO:19的3' ITR); 和/或(c) AAV衣壳蛋白,其包含SEQ ID NO:16的氨基酸序列,和转移基因组,其从5'到3'包含以下基因元件:5' ITR元件(例如SEQ ID NO:18的5' ITR)、增强子元件(例如SEQ ID NO:58的增强子元件)、启动子序列(例如SEQ ID NO:25的启动子序列)、嵌合内含子序列(例如SEQ ID NO:32的嵌合内含子序列)、以沉默方式改变的人类ARSA编码序列(例如SEQ ID NO:14的hARSA编码序列)、SV40聚腺苷酸化序列(例如SEQ ID NO:42的SV40聚腺苷酸化序列)和3' ITR元件(例如SEQ ID NO:19的3' ITR)。

[0152] 在某些实施例中,前述方法使用包含以下的rAAV: (a) AAV衣壳蛋白,其包含SEQ ID NO:16的氨基酸203-736的氨基酸序列,和转移基因组,其包含SEQ ID NO:41、44、46、47、48、49、65、67、68、69、75或76中的任一个中所阐述的核苷酸序列; (b) AAV衣壳蛋白,其包含SEQ ID NO:16的氨基酸138-736的氨基酸序列,和转移基因组,其包含SEQ ID NO:41、44、46、47、48、49、65、67、68、69、75或76中的任一个中所阐述的核苷酸序列; 和/或(c) AAV衣壳蛋白,其包含SEQ ID NO:16的氨基酸序列,和转移基因组,其包含SEQ ID NO:41、44、46、47、48、49、65、67、68、69、75或76中的任一个中所阐述的核苷酸序列。

[0153] 本文中所公开的方法的有利之处尤其在于其能够体内和体外在细胞中以高效率表达ARSA蛋白质。在某些实施例中,ARSA蛋白质的表达量为在ARSA基因中不具有突变的相同类型的细胞中的内源性ARSA蛋白质的表达量的至少10%、20%、30%、40%、50%、60%、70%、80%、90%或100%。在某些实施例中,ARSA蛋白质的表达量比在ARSA基因中不具有突变的相同类型的细胞中的内源性ARSA蛋白质的表达量高至少1.1、1.2、1.3、1.4、1.5、1.6、1.7、1.8、1.9、2、3、4、5、6、7、8、9或10倍。可使用任何用于测定ARSA蛋白质的表达量的方法,包括(但不限于)ELISA、免疫印迹法(Western blotting)、免疫染色和质谱。

[0154] 在某些实施例中,用本文中所公开的AAV组合物转导细胞可如本文所提供来进行或通过所属领域中的技术人员已知的任何转导方法来进行。在某些实施例中,可使细胞在以下感染倍率(MOI)下与AAV接触:50,000;100,000;150,000;200,000;250,000;300,000;350,000;400,000;450,000;或500,000,或任何提供细胞的最佳转导的MOI。

[0155] 可通过任何适当的途径向受试者施用本文中所公开的AAV组合物,包括(但不限于)静脉内、鞘内、腹膜内、皮下、肌内、鼻内、局部或皮内途径。在某些实施例中,将组合物调配成用于通过静脉内注射或皮下注射施用。

[0156] IV. AAV包装系统

[0157] 在另一方面中,本公开提供用于本文中所公开的重组腺相关病毒(rAAV)的重组制备的包装系统。此类包装系统通常包含:第一核苷酸,其编码一种或多种AAV Rep蛋白;第二核苷酸,其编码如本文中所公开的AAV中的任一个的衣壳蛋白;和第三核苷酸序列,其包含

如本文中所公开的rAAV基因组中的任一个,其中包装系统在细胞中具有操作性以用于将转移基因组包裹在衣壳中以形成AAV。

[0158] 在某些实施例中,包装系统包含第一载体,其包含编码一种或多种AAV Rep蛋白的第一核苷酸序列和编码AAV衣壳蛋白的第二核苷酸序列,以及第二载体,其包含有包含rAAV基因组的第三核苷酸序列。如在如本文中所描述的包装系统的情形中所使用,“载体”是指一种核酸分子,其为用于将核酸引入细胞的媒介(例如质粒、病毒、粘质体、人工染色体等)。

[0159] 任何AAV Rep蛋白均可用于本文中所公开的包装系统中。在包装系统的某些实施例中,Rep核苷酸序列编码AAV2 Rep蛋白。合适的AAV2 Rep蛋白包括(但不限于)Rep 78/68或Rep 68/52。在包装系统的某些实施例中,编码AAV2 Rep蛋白的核苷酸序列包含编码与SEQ ID NO:22的AAV2 Rep氨基酸序列具有最小序列一致性百分比的蛋白质的核苷酸序列,其中跨越AAV2 Rep蛋白的氨基酸序列的长度,最小序列一致性百分比为至少70%(例如至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%)。在包装系统的某些实施例中,AAV2 Rep蛋白具有SEQ ID NO:22中所阐述的氨基酸序列。

[0160] 在包装系统的某些实施例中,包装系统还包含第四核苷酸序列,其包含一种或多种辅助病毒基因。在包装系统的某些实施例中,包装系统还包含第三载体,例如辅助病毒载体,其包含有包含一种或多种辅助病毒基因的第四核苷酸序列。第三载体可为独立的第三载体、与第一载体整合在一起或与第二载体整合在一起。

[0161] 在包装系统的某些实施例中,辅助病毒选自由以下组成的群组:腺病毒、疱疹病毒(包含单纯性疱疹病毒(HSV))、痘病毒(如痘苗病毒)、巨细胞病毒(CMV)和杆状病毒。在包装系统的某些实施例中,其中辅助病毒是腺病毒,腺病毒基因组包含一种或多种选自由以下组成的群组的腺病毒RNA基因:E1、E2、E4和VA。在包装系统的某些实施例中,其中辅助病毒是HSV,HSV基因组包含选自由以下组成的群组的HSV基因中的一个或多个:UL5/8/52、ICP0、ICP4、ICP22和UL30/UL42。

[0162] 在包装系统的某些实施例中,第一、第二和/或第三载体包含于一种或多种质粒内。在某些实施例中,第一载体和第三载体包含于第一质粒内。在某些实施例中,第二载体和第三载体包含于第二质粒内。

[0163] 在包装系统的某些实施例中,第一、第二和/或第三载体包含于一种或多种重组辅助病毒内。在某些实施例中,第一载体和第三载体包含于重组辅助病毒内。在某些实施例中,第二载体和第三载体包含于重组辅助病毒内。

[0164] 在另一方面中,本公开提供一种用于如本文中所描述的AAV的重组制备的方法,其中所述方法包含在可操作以将rAAV基因组包裹于衣壳中以形成如本文中所描述的rAAV的条件下,用如本文中所描述的包装系统转染或转导细胞。用于rAAV的重组制备的示例性方法包括短暂转染(例如用一种或多种含有如本文中所描述的第一和第二以及任选地第三载体的转染质粒)、病毒感染(例如用一种或多种重组辅助病毒,如腺病毒、痘病毒(如痘苗病毒)、疱疹病毒(包括HSV、巨细胞病毒或杆状病毒,其含有如本文中所描述的第一和第二以及任选地第三载体)和稳定的生产细胞系转染或感染(例如用稳定的生产细胞,如哺乳动物或昆虫细胞,其含有如本文中所描述的编码一种或多种AAV Rep蛋白的Rep核苷酸序列和/或编码一种或多种衣壳蛋白的Cap核苷酸序列,并且其中以质粒或重组辅助病毒形式传递如本文中所描述的转移基因组)。

[0165] 因此,本公开提供用于制备重组AAV(rAAV)的包装系统,其中所述包装系统包含第一核苷酸序列,其编码一种或多种AAV Rep蛋白;第二核苷酸序列,其编码本文中所描述的AAV中的任一个的衣壳蛋白;第三核苷酸序列,其包含本文中所描述的AAV中的任一个的rAAV基因组序列;和任选地第四核苷酸序列,其包含一种或多种辅助病毒基因。

[0166] V. 实例

[0167] 本文中所公开的重组AAV载体活体外和活体内介导高效基因转移。以下实例说明使用如本文中所公开的基于AAV的载体进行的ARSA基因(其在某些人类疾病(如异染性脑白质营养不良)中突变)的表达的有效恢复。这些实例以说明性方式而非限制性方式提供。

[0168] 实例1:人类ARSA转移载体

[0169] 此实例提供用于被载体转导的细胞(例如人类细胞或小鼠细胞)中的人类ARSA(hARSA)的表达的人类ARSA转移载体T-001、pHMI-5000、pHMI-5003和pHMI-hARSA1-TC-002。

[0170] a) T-001

[0171] 如图1A中所示,ARSA转移载体TC-001从5'到3'包含以下基因元件:5' ITR元件、转录调节元件,其包含CMV增强子元件、鸡β-肌动蛋白启动子和嵌合内含子序列;野生型人类ARSA编码序列;SV40聚腺苷酸化序列;以及3' ITR元件。这些元件的序列阐述于表1中。此载体能够在被载体转导的细胞(例如人类细胞或小鼠细胞)中表达人类ARSA蛋白质。

[0172] b) pHMI-5000

[0173] 如图1B中所示,ARSA转移载体pHMI-5000从5'到3'包含以下基因元件:5' ITR元件;转录调节元件,其包含CMV增强子元件、鸡β-肌动蛋白启动子和嵌合内含子序列;以沉默方式改变的人类ARSA编码序列;SV40聚腺苷酸化序列;以及3' ITR元件。这些元件的序列阐述于表1中。此载体能够在被载体转导的细胞(例如人类细胞或小鼠细胞)中表达人类ARSA蛋白质。

[0174] c) pHMI-5003

[0175] 如图1C中所示,ARSA转移载体pHMI-5003从5'到3'包含以下基因元件:5' ITR元件;转录调节元件,其包含CMV增强子元件、鸡β-肌动蛋白启动子和嵌合内含子序列;以沉默方式改变的人类ARSA编码序列;SV40聚腺苷酸化序列;非编码填充序列,以及3' ITR元件。这些元件的序列阐述于表1中。此载体能够在被载体转导的细胞(例如人类细胞或小鼠细胞)中表达人类ARSA蛋白质。

[0176] d) pHMI-hARSA1-TC-002

[0177] 如图1D中所示,ARSA转移载体pHMI-hARSA1-TC-002从5'到3'包含与pHMI-5000相同的基因元件。这些元件的序列阐述于表1中。pHMI-hARSA1-TC-002与pHMI-5000之间的差异位于载体主链序列中。此载体能够在被载体转导的细胞(例如人类细胞或小鼠细胞)中表达人类ARSA蛋白质。

[0178] 表1:人类ARSA转移载体T-001、pHMI-5000、pHMI-5003和pHMI-hARSA1-TC-002中的基因元件

基因元件	T-001	pHMI-5000	pHMI-5003	pHMI-hARSA1-TC-002
	SEQ ID NO:			
5'ITR 元件	18	18	18	18
增强子元件	58	58	58	58
启动子序列	25	25	25	25
内含子序列	32	32	32	32
转录调节元件	36	36	36	36
人类 ARSA 编码序列	24	14	14	14
[0179] SV40 聚腺苷酸化序列	42	42	42	42
填充序列	N/A	N/A	39	N/A
3'ITR 元件	19	19	19	19
转移基因组 (从启动子到聚腺苷酸化序列)	41	44	46	44
转移基因组 (从5'ITR 到 3'ITR)	47	48	49	48
完全载体序列	50	51	52	53

[0180] 本文中所公开的载体可包装于AAV衣壳中,如(但不限于)AAVHSC5、AAVHSC7、AAVHSC15或AAVHSC17衣壳。可将包装的病毒颗粒施用于野生型动物或ARSA缺失型动物。

[0181] 实例2:ARSA (-/-) 小鼠模型中的ARSA基因转移

[0182] 为了研究小鼠中ARSA基因转移的作用,产生ARSA (-/-) 小鼠模型。ARSA (-/-) 小鼠模型为通过将新霉素卡匣插入小鼠ARSA基因的外显子4而产生的ARSA基因剔除小鼠(参见Hess等人,《美国国家科学院院刊(Proc.Natl.Acad.Sci.U.S.A.)》1996,93(25):14821-14826,其以全文引用的方式并入本文中)。与人类相比,ARSA (-/-) 小鼠发展类似的但程度较轻的异染性脑白质营养不良(MLD)。ARSA (-/-) 小鼠未展示广泛的髓鞘脱失的证据。

[0183] 可使用各种生物标记研究MLD。例如,可测量脑中硫苷脂的含量。已报告少突胶质细胞(C24:0)和神经元(C18:0)硫苷脂的增加,其中积聚随动物年龄而增加。可测量髓鞘和淋巴细胞蛋白质(MAL)mRNA转录物的含量。MAL由少突胶质细胞和施万细胞表达,使神经胶质-轴突接合区稳定并与MLD的病变有关。已报告ARSA (-/-) 小鼠中的MAL转录物的含量降低。溶酶体相关膜蛋白质(LAMP-1)为另一种可用于研究MLD的生物标记。已使用抗LAMP-1抗体,通过对ARSA (-/-) 和野生型小鼠中的脊髓组织进行免疫组织化学来研究LAMP-1免疫反应性,表明ARSA (-/-) 小鼠中的LAMP-1免疫反应性提高。图2A展示来源于通过对来自ARSA (-/-) 小鼠的脊髓组织进行免疫组织化学(IHC)而研究的LAMP-1免疫反应性的总像素强度的定量。在用媒介剂对照物或包装于AAVHSC15衣壳中的pHMI-5000处理的ARSA (-/-) 小鼠中使用抗LAMP-1抗体进行IHC。如图2A中所示,在给药(4e13 vg/kg)的包装于AAVHSC15衣壳中的pHMI-5000)后第12周时,检测到与给药媒介剂对照物的ARSA (-/-) 动物相比显著降低的LAMP-1含量。

[0184] 将脑组织称重并在Precellys珠粒均质器中,在250uL水中均质化,并且移出均质物的10uL等分试样以用于Pierce BCA蛋白质分析定量。将760uL乙腈添加到各均质物中并将混合物第二次均质化。均质物在14,000×g下离心15分钟并移出离心机澄清的上清液并

在75%乙腈中稀释5倍以用于RapidFire-MS分析法。使用C19:0硫苷脂 (Matreya目录号1888) 作为内标并用Sciex API4000三重四极杆质谱仪以MRM模式与C18:0、C18:1、C24:0和C24:1硫苷脂一起监测。各样品与8种不同浓度的C19:0硫苷脂IS一起注射8次以产生各样品的独特标准曲线,使用其计算各分析物的浓度。图2B展示随时间推移,对照组小鼠 (WT/Het) 和ARSA (-/-) 小鼠的脑中的C18:0硫苷脂含量。对照组为野生型动物 (ARSA (+/+)) 与杂合动物 (ARSA (+/-)) 的混合。如图2B中所示,ARSA (-/-) 小鼠的脑中的C18:0硫苷脂含量随时间推移而累积,而对照组小鼠的脑中的C18:0硫苷脂含量随时间推移而保持基本不变。由两只对照组小鼠和两只ARSA (-/-) 小鼠的分析产生图2B中的数据。为了研究ARSA缺失型小鼠中ARSA基因递送对硫苷脂积聚的作用,用4e13 vg/kg的包装于AAVHSC15衣壳中的pHMI-hARSA1-TC-002处理ARSA (-/-) 小鼠 (图2C)。如图2C中所示,与用媒介剂对照物处理的ARSA (-/-) 小鼠相比,在给药后七个月时观察到被处理的ARSA (-/-) 小鼠中的脑部硫苷脂含量的显著降低。

[0185] 在用4e13 vg/kg和6e13 vg/kg的包装于AAVHSC15衣壳中的pHMI-5000或媒介剂对照物处理后七个月时测定ARSA (-/-) 小鼠的前脑、中脑和后脑中的C18:0及C18:1硫苷脂同种型含量 (图2D)。硫苷脂同种型含量呈现为与相同年龄的野生型对照动物相比的倍数。如图2D中所示,与用媒介剂对照物处理的ARSA (-/-) 小鼠相比,在给药后七个月时观察到被处理的ARSA (-/-) 小鼠的所有三个脑部区域中的脑部硫苷脂含量的显著降低。所使用的方法和材料与上述相同。使用非配对T测试 (unpaired T-test) 分析数据。

[0186] 在用4e13 vg/kg的包装于AAVHSC15衣壳中的pHMI-5000或媒介剂对照物处理后第52周测定ARSA (-/-) 小鼠的前脑、中脑和后脑中的C18:0和C18:1硫苷脂同种型含量 (图2E)、C24:0及C24:1硫苷脂同种型含量 (图2F) 以及总硫苷脂同种型含量 (图2G)。所使用的方法和材料与上述相同。使用非配对T测试 (unpaired T-test) 分析数据。

[0187] 图3A展示对照组小鼠 (WT/Het) 和ARSA (-/-) 小鼠中,在四周时MAL转录物的含量。对照组为野生型动物 (ARSA (+/+)) 与杂合动物 (ARSA (+/-)) 的混合。用Trizol提取,接着进行Qiagen RNEasy管柱纯化来制备小鼠总RNA。RNA用作使用赛默飞世尔高容量cDNA试剂盒 (ThermoFisher High Capacity cDNA Kit) 进行的cDNA合成的模板以产生转录物。使用液滴式数字PCR和对小鼠髓鞘以及淋巴细胞蛋白质 (MAL) 具有特异性的引子/探针集合评估MAL转录物,其中针对小鼠HPRT1标准化拷贝数。如所展示,在四周时,与杂合小鼠相比,ARSA (-/-) 小鼠中的MAL转录物的含量降低。由五只对照组小鼠和六只ARSA (-/-) 小鼠的分析产生图3中的数据。为了研究ARSA缺失型小鼠中ARSA基因传递对MAL转录物含量的影响,用4e13vg/kg的包装于AAVHSC15衣壳中的pHMI-5000处理ARSA (-/-) 小鼠 (图3B)。如图3B中所示,与野生型小鼠和用媒介剂处理的ARSA (-/-) 小鼠相比,在给药后三个月时观察到被处理的ARSA (-/-) 小鼠中的MAL转录物含量的显著增加。

[0188] 测定用4e13 vg/kg的包装于AAVHSC15衣壳中的pHMI-5000处理的ARSA (-/-) 小鼠中的MAL转录物拷贝数的水平 (图3C)。图3C展示在野生型小鼠中或在被施用媒介剂对照物或4e13 vg/kg的包装于AAVHSC15衣壳中的pHMI-5000的ARSA (-/-) 小鼠中,在给药后第12或52周时检测到的MAL转录物的拷贝数。所使用的方法和材料与上述相同。使用非配对T测试 (unpaired T-test) 分析数据。在图3C中,各动物组之间的统计显著性如下:12周媒介剂对比被处理的动物, $p=0.0012$;12周处理的对比野生型动物, $p<0.0001$;52周媒介剂对比被处理的

动物, $p=0.0004$; 以及52周处理的对比野生型动物, 不显著。

[0189] 为了研究是否可获得hARSA活性的治疗水平, 向ARSA (-/-) 小鼠施用包装于AAV9衣壳中的转移载体T-001 (参见PCT公开案第W02002/052052号, 其以全文引用的方式并入本文中)。从未处理的对照ARSA (-/-) 小鼠和被施用包装于AAV9衣壳中的转移载体T-001的ARSA (-/-) 小鼠获得的脑薄片的抗ARSA免疫反应性表明, 在 $2e13$ 载体基因组/公斤体重 (vg/kg) 的剂量下获得达到治疗水平 (10%) 的hARSA酶活性。处被处理的ARSA (-/-) 小鼠获得的脑薄片的抗ARSA免疫反应性亦展示脑中的ARSA酶活性的剂量依赖性增加。

[0190] 实例3: ARSA (-/-) 小鼠模型中的ARSA基因转移

[0191] 此实例提供与人类ARSA转移载体pHMI-5000的用途相关的实验数据。如本文中所描述, 转移载体pHMI-5000包含以沉默方式改变的人类ARSA编码序列, 表明其呈现显著改进的ARSA蛋白质的表达。

[0192] 图4为展示脑部中以每个被转导的细胞计的载体基因组的数目与以每ng cDNA计的hARSA的拷贝数之间的相关性的曲线。使用来自Qiagen的QIAamp Fast DNA Tissue试剂盒制备小鼠基因组DNA。藉由液滴式数字PCR和对经密码子优化的人类ARSA载体基因组的编码区具有特异性的引子/探针集合测定VG计数, 其中针对内源性小鼠基因组序列进行标准化。如本文中所描述制备小鼠总RNA并使用液滴式数字PCR以及用于测定VG计数的相同引子/探针集合评估ARSA转录物, 其中针对小鼠GUSB标准化拷贝数。如所示出, 对于使用包装于AAVHSC15衣壳中的转移载体pHMI-5000转导的细胞, 所检测到的以每个被转导的细胞计的载体基因组的数目与以每ng cDNA计的hARSA的拷贝数强相关 ($R^2=0.9332$)。

[0193] 发现在AAVHSC15与AAV9衣壳介导的传递之间的比较中, AAVHSC15在脑部中显著优于AAV9。图5展示在 $2e13$ vg/kg的包装于AAV9或AAVHSC15衣壳中的转移载体pHMI-5000的剂量下, 脑中的以每个被转导的细胞计的载体基因组的数目。如所示出, 与AAV9衣壳相比, 当转移载体pHMI-5000包装于AAVHSC15衣壳中时, 观察到高十倍的以每个细胞计的载体基因组计数。图6展示对于以所指示的剂量施用的包装于AAV9或AAVHSC15衣壳中的转移载体pHMI-5000, 所测量的正常人类ARSA酶活性水平的百分比。图7展示在以 $4e13$ vg/kg施用包装于AAV9或AAVHSC15中的转移载体pHMI-5000的小鼠中, 以每个被转导的脑细胞计的载体基因组的数目。

[0194] 与包装于AAV9衣壳中的pHMI-5000相比, 包装于AAVHSC15衣壳中的pHMI-5000表明更强并且更广泛的脑部和脊髓表达概况。抗ARSA免疫反应性实验表明, 在 $3e13$ vg/kg的剂量下, 在各种情况下, 与静脉内施用包装于AAV9衣壳中的pHMI-5000的小鼠相比, 在静脉内施用包装于AAVHSC15衣壳中的pHMI-5000的小鼠的脑薄片检测到显著更高的量。

[0195] 为了评估投药途径对脑中hARSA的生物分布的影响, 分别在 $4e13$ vg/kg和 $4e12$ vg/kg的剂量下经由静脉内 (IV) 和鞘内 (IT) 途径施用包装于AAVHSC15衣壳中的转移载体pHMI-5000。在向ARSA (-/-) 小鼠静脉内施用包装于AAVHSC15中的pHMI-5000之后, 在重要的中枢神经系统区域中存在抗ARSA免疫反应性。广泛检测到抗小鼠ARSA (mARSA) 或人类ARSA (hARSA), 包括 (但不限于) 运动和感觉皮质、海马 (CA3区域)、壳核和小脑。在包装于AAVHSC15中的转移载体pHMI-5000的IV或IT投药之后, 后脑和中脑中的正常人类ARSA酶活性的百分比的定量展示于图8中。

[0196] 在持续4周以 $4e13$ vg/kg施用包装于AAVHSC15衣壳中的pHMI-5000的ARSA (-/-) 小

鼠中,在脑部的重要生理学区域以及中枢神经系统(CNS)的整个后尾轴中检测到hARSA的生物相关分布。使用抗hARSA抗体检测hARSA,并且在脊髓、运动皮质、丘脑、海马和小脑核中检测到。亦在以下中检测到hARSA:CNS中的运动神经元和星形胶质细胞分布;CNS中的少突胶质细胞(在上升纤维中大量检测到);CNS中的大脑皮质的细胞群;以及周围神经系统(PNS)中的感觉神经元和施万细胞。最早可在处理后第2周检测到类似的生物学分布。

[0197] 在以 2×10^{13} vg/kg施用包装于AAVHSC15衣壳中的pHMI-5000的小鼠中,观察到与在以 4×10^{13} vg/kg或更高的剂量进行投药的小鼠中所发现相同的组织学分布。在此等实验中,以溶酶体特有的点状模式在细胞质中检测到hARSA。

[0198] 如图9A和9B中所示,在给药后第4周,被处理的ARSA(-/-)小鼠的大脑中的人类ARSA酶活性的生理学水平得到恢复。使用来自ARSA(-/-)小鼠的脑溶解物评估hARSA酶活性。剂量范围发现研究表明,hARSA酶活性与包装于AAVHSC15衣壳中的转移载体pHMI-5000的静脉内投药的剂量相关。在被处理的动物中,但未在媒介对照性动物中检测到酶活性。对于所测试的剂量,酶活性水平(约40-145%)显著高于治疗目标(约10-15%),如先前在临床上测定(参见Patil和Maegawa,《药物设计和发展与疗法(Drug Des.Devel.Ther.)》2013,7:729-745)。图9A展示通过以所指示的剂量向ARSA(-/-)小鼠施用包装于AAVHSC15衣壳中的转移载体pHMI-5000而实现的正常hARSA活性的百分比。如所示出,获得hARSA活性的剂量依赖性反应。图9B展示以所指示的剂量施用包装于AAVHSC15衣壳中的转移载体pHMI-5000的ARSA(-/-)小鼠的脑中以每个细胞计的载体基因组的数目。对于 1×10^{13} vg/kg、 4×10^{13} vg/kg和 6×10^{13} vg/kg剂量, $n=5$ 只小鼠。对于 2×10^{13} vg/kg剂量, $n=4$ 只小鼠。所有小鼠为5周龄并且均是雄性。在图9C中,使用比色芳基硫酸酯酶A特异性分析法评估ARSA酶活性,所述分析法测量由可溶性底物对硝基儿茶酚-硫酸酯(pNCS)引起的硫酸酯的裂解。使用芳基硫酸酯酶A特异性免疫沉淀步骤消除由竞争酶引起的硫酸酯的非特异性裂解。通过各两位正常人类男性和女性的额叶皮质中的ARSA酶活性的分析来测定脑中的正常人类ARSA酶活性。人类额叶皮质样品购自BioiVT并且在每个ARSA酶活性分析盘上与测试样品一起一式三份地操作。数据表示为以每小时每mg蛋白质计的脱硫酸化pNCS的平均量(以ng为单位)的百分比。图9C展示最早在处理第1周并且直到处理后第12周,单次静脉内 4×10^{13} vg/kg剂量的包装于AAVHSC15衣壳中的pHMI-5000引起在新生ARSA(-/-)小鼠的大脑中检测到hARSA酶活性,其水平超过既定人类治疗目标(10-15%)(如虚线所指示)。在给药后第1、2、3、4和12周收集材料。对于每个时间点, $n=6$ 只小鼠,3只雄性和3只雌性,8周龄。

[0199] 在图9D中,用Trizol提取,接着进行Qiagen RNEasy管柱纯化来制备小鼠总RNA。RNA用作使用赛默飞世尔高容量cDNA试剂盒进行的cDNA合成的模板以产生转录物。使用液滴式数字PCR个对经密码子优化的人类ARSA转录物具有特异性的引子/探针集合评估ARSA转录物,其中针对小鼠GUSB标准化拷贝数。图9D展示最早在处理第1周,单次静脉内 4×10^{13} vg/kg剂量的包装于AAVHSC15衣壳中的pHMI-5000引起在成年ARSA(-/-)小鼠的大脑中检测到正常水平的hARSA酶活性(经由hARSA转录物分析法)。在给药后第2到3周观察到hARSA酶活性的峰值水平,接着稳定的平线区持续到处理后第52周(以超过既定人类治疗目标(10-15%)的水平)。在给药后第1、2、3、4、8、12、26和52周收集材料。图9E显示在施用单次静脉内 4×10^{13} vg/kg的剂量的包装于AAVHSC15衣壳中的pHMI-5000的ARSA(-/-)小鼠的大脑中,以每ug基因组DNA计的载体基因组的数目。在给药后第1、2、3、8、12、26和52周收集材料。图9F展

示在施用单次静脉内 4×10^{13} vg/kg的剂量的包装于AAVHSC15衣壳中的pHMI-5000的ARSA (-/-)小鼠的大脑中,以每ng RNA计的ARSA转录物的拷贝数。在给药后第4、8、12、26和52周收集材料。

[0200] 实例4:人类ARSA转移载体

[0201] 此实例提供用于在被载体转导的细胞(例如人类细胞或小鼠细胞)中的hARSA的表达的人类ARSA转移载体TC-013.pHMIA2和TC-015.pKITR。除表达hARSA以外,这些载体被设计成还表达人类SUMF1。hARSA和hSUMF1的编码序列由2A元件分离。在某些实施例中,核糖体跳跃元件(例如2A元件)编码在N端还包含Gly-Ser-Gly的序列的肽,任选地其中Gly-Ser-Gly的序列由GGCAGCGGA的核苷酸序列编码。尽管不希望受理论约束,但假设核糖体跳跃元件通过以下发挥功能:终止第一肽链的翻译和再起始第二肽链的翻译;或由核糖体跳跃元件编码的肽序列中的肽键通过编码的肽的内在蛋白酶活性进行裂解,或环境(例如胞质溶胶)中的另一种蛋白酶。

[0202] a) TC-013.pHMIA2

[0203] 如图10A中所示,ARSA转移载体TC-013.pHMIA2从5'到3'包含以下基因元件:5' ITR元件、包含CALM1启动子的转录调节元件;以沉默方式改变的人类ARSA编码序列;2A元件;以沉默方式改变的人类SUMF1编码序列;以及3' ITR元件。这些元件的序列阐述于表2中。此载体能够在被载体转导的细胞(例如人类细胞或小鼠细胞)中表达人类ARSA蛋白质和人类SUMF1蛋白质。

[0204] b) TC-015.pKITR

[0205] 如图10B中所示,ARSA转移载体TC-015.pKITR从5'到3'包含以下基因元件:5' ITR元件、包含smCBA启动子的转录调节元件;以沉默方式改变的人类ARSA编码序列;2A元件;以沉默方式改变的人类SUMF1编码序列;以及3' ITR元件。这些元件的序列阐述于表2中。此载体能够在被载体转导的细胞(例如人类细胞或小鼠细胞)中表达人类ARSA蛋白质和人类SUMF1蛋白质。

[0206] 表2:人类ARSA转移载体TC-013.pHMIA2和TC-015.pKITR中的基因元件

基因元件	TC-013.pHMIA2	TC-015.pKITR
	SEQ ID NO:	
5'ITR 元件	18	18
启动子序列	54	55
转录调节元件	54	55
人类 ARSA 编码序列	62	62
[0207] 2A 元件	63	63
人类 SUMF1 编码序列	64	64
hARSA-2A-hSUMF1 序列	30	30
3'ITR 元件	19	19
转移基因组(从启动子到 SUMF1 编码序列)	65	67
转移基因组(从 5'ITR 到 3'ITR)	68	69
完全载体序列	70	71

[0208] 本文中所公开的载体可包装于AAV衣壳中,如(但不限于)AAVHSC5、AAVHSC7、AAVHSC15或AAVHSC17衣壳。可将包装的病毒颗粒施用于野生型动物或ARSA缺失型动物。

[0209] 为了评估脑中启动子对hARSA表达的影响,将转移载体pHMI-5000、TC-013.pHMIA2

和TC-015.pKitr包装于AAVHSC15衣壳中并且静脉内施用ARSA (-/-) 小鼠。在以 4×10^{13} vg/kg的剂量施用pHMI-5000载体 (鸡 β -肌动蛋白 (CBA) 启动子) 和以 8×10^{13} vg/kg的剂量施用TC-015.pKitr (smCBA启动子) 的情况下在脑中检测到hARSA表达和酶活性, 其中以每个细胞计数的病毒基因组类似。与所测试的其它启动子相比, CBA启动子在最低剂量下引起hARSA的最高表达。图11展示pHMI-5000 (CBA启动子)、TC-013.pHMIA2 (CALM1启动子) 和TC-015.pKitr (smCBA启动子) 的以每个细胞计的被转导的病毒基因组的数目, 在每种情况下包装于AAVHSC15衣壳中并以 4×10^{13} vg/kg的剂量施用 (对于每种载体, $n=5$ 只小鼠)。图12展示所检测的pHMI-5000 (CBA启动子) 和TC-015.pKitr (smCBA启动子) 的正常人类ARSA酶活性的百分比, 在每种情况下包装于AAVHSC15衣壳中并以 4×10^{13} vg/kg的剂量施用 (对于每种载体, $n=5$ 只小鼠)。图13展示对于包装于AAVHSC15衣壳中并以 4×10^{13} vg/kg的剂量施用的pHMI-5000 (CBA启动子) 和包装于AAVHSC15衣壳中并以 8×10^{13} vg/kg的剂量施用的TC-015.pKitr (smCBA启动子), 在免疫印迹法中, 可在使用抗hARSA抗体的小鼠的脑中检测到hARSA的表达 (对于每种载体, $n=5$ 只小鼠)。

[0210] 实例5: 人类ARSA转移载体

[0211] 此实例提供用于在被载体转导的细胞 (例如人类细胞或小鼠细胞) 中的hARSA的表达的人类ARSA转移载体pHMI-5004。除表达hARSA以外, 此载体被设计成还表达人类鞘脂激活蛋白B (SapB)。hARSA和SapB的编码序列由2A元件分离。

[0212] 如图14中所示, ARSA转移载体pHMI-5004从5'到3'包含以下基因元件: 5' ITR元件; 转录调节元件, 其包含CMV增强子元件、鸡 β -肌动蛋白启动子和嵌合内含子序列; 以沉默方式改变的人类ARSA编码序列; 2A元件; 野生型人类SapB编码序列; 以及3' ITR元件。这些元件的序列阐述于表3中。此载体能够在被载体转导的细胞 (例如人类细胞或小鼠细胞) 中表达人类ARSA和/或SapB蛋白质。

[0213] 表3: 人类ARSA转移载体pHMI-5004中的基因元件

基因元件	SEQ ID NO:
5' ITR元件	18
增强子元件	58
启动子序列	25
内含子序列	32
转录调节元件	36
人类ARSA编码序列	72
2A元件	63
人类SapB编码序列	73
hARSA-2A-hSapB序列	74
SV40聚腺苷酸化序列	42
3' ITR元件	19
转移基因组 (从启动子到聚腺苷酸化序列)	75
转移基因组 (从5' ITR到3' ITR)	76
完全载体序列	77

[0215] 实例6: 非人类灵长类动物中的ARSA基因转移

[0216] 为了研究非人类灵长类动物中单次剂量的AAVHSC介导的ARSA基因传递的作用,根据表4和5中所阐述的实验设计向六只雄性未处理的幼年食蟹猕猴给药。

[0217] 表4:用于非人类灵长类动物研究的实验设计

组	途径	动物/组	给药 天数	剂量 (vg/kg)	体积 (mL/kg)	浓度 (vg/mL)	尸检
1	IV	2只雄性	1	0	5.0	-	第 28/29 天
2	IV	2只雄性	1	4e13	5.0	1.2e13	
3	CM	2只雄性	1	静脉内剂量的大致 10%, 基于动物体重以 固定剂量形式提供 (约 6e12 vg/kg)	0.5 mL	储备溶液 为 1.98 vg/mL	

[0219] 表5:用于非人类灵长类动物研究的实验设计

动物	重量 (kg)	治疗	途径	剂量 (vg/kg)	Vg/动物
18C42	1.38	媒介剂	IV	0	0
18C17	1.55	媒介剂	IV	0	0
18C21	1.28	AAVHSC15-pHMI-5005	IV	4e13	5.12e13
18C27	1.28	AAVHSC15-pHMI-5005	IV	4e13	5.12e13
18C13	1.9	AAVHSC15-pHMI-5005	CM	4e12	7.6e12
18C7	1.74	AAVHSC15-pHMI-5005	CM	4e12	6.96e12

[0221] 如图15中所示,ARSA转移载体pHMI-5005从5'到3'包含以下基因元件:5' ITR元件;转录调节元件,其包含CMV增强子元件、鸡β-肌动蛋白启动子和嵌合内含子序列;以沉默方式改变的人类ARSA编码序列;V5标签;以及3' ITR元件。这些元件的序列阐述于表6中。此载体能够在被载体转导的细胞(例如人类细胞或小鼠细胞)中表达人类ARSA蛋白质。

[0222] 表6:人类ARSA转移载体pHMI-5005中的基因元件

基因元件	SEQ ID NO:
5' ITR元件	18
增强子元件	58
启动子序列	25
内含子序列	32
转录调节元件	36
人类ARSA编码序列	14
V5标签	78
SV40聚腺苷酸化序列	42
3' ITR元件	19
转移基因组 (从启动子到聚腺苷酸化序列)	79
转移基因组 (从5' ITR到3' ITR)	80
完全载体序列	81

[0224] pHMI-5005是被V5标记的ARSA转移载体。根据表4和5中所阐述的实验设计向非人类灵长类动物(NHP)施用包装于AAVHSC15衣壳中的pHMI-5005。在第0天,经由经头部/隐静

脉进行1-2分钟缓慢推注静脉内注射(IV)或直接注射到小脑延髓池(CM)中来进行投药。每天两次针对死亡和濒死迹象来进行存活率检查。在每天上午以及在给药当天在完成给药之后(15min)和给药后4小时进行临床观察。在即将给药之前以及在给药后第1、2和4周获得用于血液学和临床化学的血液。在第28和29天的尸检时,在脑脊髓液(CSF)和血液收集之后,向动物灌注1.0L低温生理食盐水以去除血球。在尸检时收集脑、肝脏、脊髓(子宫颈和腰)、子宫颈和腰背根神经节(DRG)、三叉神经节、肾脏、坐骨神经、外周淋巴结、脾、心脏、肺和睾丸。

[0225] 对于生物分析,在即将给药之前以及在第1、2和4周收集血清以用于V5 Elisa (0.5mL全血,处理成血清/分成两个等分试样)。在给药前收集0.5mL CSF(来自第3组CM给药组动物)和在尸检时收集1-2mL(对于所有动物)。在尸检之前,从全血收集15mL外周血液单核细胞(PBMC)。

[0226] 图16展示在施用包装于AAVHSC15衣壳中的pHMI-5005的NHP中,丙氨酸氨基转移酶(ALT)的含量升高。升高的ALT在给药后第14天恢复到基线水平。

[0227] 在给药后第28和29天处死接受单次静脉内剂量的 4×10^{13} vg/kg的包装于AAVHSC15中的pHMI-5005的NHP(第2组动物)。在经处死的第2组动物的中枢神经系统(CNS)和脑脊髓液(CSF)中检测人类ARSA酶活性水平(图17)。如图17中所示,检测到高于治疗阈值(野生型人类脑含量的15%)的hARSA活性水平,如由虚线所指示。动物18C27(第2组)的CNS和周围神经系统(PNS)中的免疫荧光染色证实存在hARSA(经由V5标签检测)并且存在于包括背根神经节、脊椎运动神经元和小脑的特定区域中。

[0228] ***

[0229] 本发明不应限于本文所描述的具体实施例的范围。实际上,根据前文的描述和附图,除了所描述的修改之外,本发明的各种修改对于所属领域技术的人员来说也将变得显而易见。此类修改旨在落入所附权利要求书的范围内。

[0230] 本文所引用的全部参考文献(例如,出版物或专利或专利申请)均以全文引用的方式并且出于所有目的并入本文中,其程度如同每个单独的参考文献(例如,出版物或专利或专利申请)出于所有目的被特别地或单独地指出以全文引用的方式并入一样。其它实施例处于所附权利要求书内。

序列表

<110> HOMOLOGY MEDICINES, INC.

<120> 用于ARSA基因转移的腺相关病毒组合物和其使用方法

<130> 706508: HMW-030PC

<150> US 62/859,539

<151> 2019-06-10

<150> US 62/866,374

<151> 2019-06-25

<150> US 62/915,523

<151> 2019-10-15

<150> US 62/960,487

<151> 2020-01-13

<150> US 62/987,858

<151> 2020-03-10

<150> US 63/010,970

<151> 2020-04-16

<160> 81

<170> PatentIn version 3.5

<210> 1

<211> 736

<212> PRT

<213> 腺相关AAV9

<400> 1

```

Met Ala Ala Asp Gly Tyr Leu Pro Asp Trp Leu Glu Asp Asn Leu Ser
1           5           10           15
Glu Gly Ile Arg Glu Trp Trp Ala Leu Lys Pro Gly Ala Pro Gln Pro
          20           25           30
Lys Ala Asn Gln Gln His Gln Asp Asn Ala Arg Gly Leu Val Leu Pro
          35           40           45
Gly Tyr Lys Tyr Leu Gly Pro Gly Asn Gly Leu Asp Lys Gly Glu Pro
          50           55           60
Val Asn Ala Ala Asp Ala Ala Ala Leu Glu His Asp Lys Ala Tyr Asp
65           70           75           80
Gln Gln Leu Lys Ala Gly Asp Asn Pro Tyr Leu Lys Tyr Asn His Ala
          85           90           95
Asp Ala Glu Phe Gln Glu Arg Leu Lys Glu Asp Thr Ser Phe Gly Gly
          100          105          110
Asn Leu Gly Arg Ala Val Phe Gln Ala Lys Lys Arg Leu Leu Glu Pro

```

115	120	125
Leu Gly Leu Val Glu Glu Ala	Ala Lys Thr Ala Pro Gly Lys Lys Arg	
130	135	140
Pro Val Glu Gln Ser Pro Gln Glu Pro Asp Ser Ser Ala Gly Ile Gly		
145	150	155
Lys Ser Gly Ala Gln Pro Ala Lys Lys Arg Leu Asn Phe Gly Gln Thr		
165	170	175
Gly Asp Thr Glu Ser Val Pro Asp Pro Gln Pro Ile Gly Glu Pro Pro		
180	185	190
Ala Ala Pro Ser Gly Val Gly Ser Leu Thr Met Ala Ser Gly Gly Gly		
195	200	205
Ala Pro Val Ala Asp Asn Asn Glu Gly Ala Asp Gly Val Gly Ser Ser		
210	215	220
Ser Gly Asn Trp His Cys Asp Ser Gln Trp Leu Gly Asp Arg Val Ile		
225	230	235
Thr Thr Ser Thr Arg Thr Trp Ala Leu Pro Thr Tyr Asn Asn His Leu		
245	250	255
Tyr Lys Gln Ile Ser Asn Ser Thr Ser Gly Gly Ser Ser Asn Asp Asn		
260	265	270
Ala Tyr Phe Gly Tyr Ser Thr Pro Trp Gly Tyr Phe Asp Phe Asn Arg		
275	280	285
Phe His Cys His Phe Ser Pro Arg Asp Trp Gln Arg Leu Ile Asn Asn		
290	295	300
Asn Trp Gly Phe Arg Pro Lys Arg Leu Asn Phe Lys Leu Phe Asn Ile		
305	310	315
Gln Val Lys Glu Val Thr Asp Asn Asn Gly Val Lys Thr Ile Ala Asn		
325	330	335
Asn Leu Thr Ser Thr Val Gln Val Phe Thr Asp Ser Asp Tyr Gln Leu		
340	345	350
Pro Tyr Val Leu Gly Ser Ala His Glu Gly Cys Leu Pro Pro Phe Pro		
355	360	365
Ala Asp Val Phe Met Ile Pro Gln Tyr Gly Tyr Leu Thr Leu Asn Asp		
370	375	380
Gly Ser Gln Ala Val Gly Arg Ser Ser Phe Tyr Cys Leu Glu Tyr Phe		
385	390	395
Pro Ser Gln Met Leu Arg Thr Gly Asn Asn Phe Gln Phe Ser Tyr Glu		
405	410	415
Phe Glu Asn Val Pro Phe His Ser Ser Tyr Ala His Ser Gln Ser Leu		
420	425	430

Asp Arg Leu Met Asn Pro Leu Ile Asp Gln Tyr Leu Tyr Tyr Leu Ser		
435	440	445
Lys Thr Ile Asn Gly Ser Gly Gln Asn Gln Gln Thr Leu Lys Phe Ser		
450	455	460
Val Ala Gly Pro Ser Asn Met Ala Val Gln Gly Arg Asn Tyr Ile Pro		
465	470	475
Gly Pro Ser Tyr Arg Gln Gln Arg Val Ser Thr Thr Val Thr Gln Asn		
	485	490
Asn Asn Ser Glu Phe Ala Trp Pro Gly Ala Ser Ser Trp Ala Leu Asn		
	500	505
Gly Arg Asn Ser Leu Met Asn Pro Gly Pro Ala Met Ala Ser His Lys		
	515	520
Glu Gly Glu Asp Arg Phe Phe Pro Leu Ser Gly Ser Leu Ile Phe Gly		
	530	535
Lys Gln Gly Thr Gly Arg Asp Asn Val Asp Ala Asp Lys Val Met Ile		
545	550	555
Thr Asn Glu Glu Glu Ile Lys Thr Thr Asn Pro Val Ala Thr Glu Ser		
	565	570
Tyr Gly Gln Val Ala Thr Asn His Gln Ser Ala Gln Ala Gln Ala Gln		
	580	585
Thr Gly Trp Val Gln Asn Gln Gly Ile Leu Pro Gly Met Val Trp Gln		
	595	600
Asp Arg Asp Val Tyr Leu Gln Gly Pro Ile Trp Ala Lys Ile Pro His		
610	615	620
Thr Asp Gly Asn Phe His Pro Ser Pro Leu Met Gly Gly Phe Gly Met		
625	630	635
Lys His Pro Pro Pro Gln Ile Leu Ile Lys Asn Thr Pro Val Pro Ala		
	645	650
Asp Pro Pro Thr Ala Phe Asn Lys Asp Lys Leu Asn Ser Phe Ile Thr		
	660	665
Gln Tyr Ser Thr Gly Gln Val Ser Val Glu Ile Glu Trp Glu Leu Gln		
	675	680
Lys Glu Asn Ser Lys Arg Trp Asn Pro Glu Ile Gln Tyr Thr Ser Asn		
	690	695
Tyr Tyr Lys Ser Asn Asn Val Glu Phe Ala Val Asn Thr Glu Gly Val		
705	710	715
Tyr Ser Glu Pro Arg Pro Ile Gly Thr Arg Tyr Leu Thr Arg Asn Leu		
	725	730

<210> 2

<211> 736

<212> PRT

<213> 人工序列

<220>

<223> AAV分离株

<400> 2

```

Met Thr Ala Asp Gly Tyr Leu Pro Asp Trp Leu Glu Asp Asn Leu Ser
1           5           10           15
Glu Gly Ile Arg Glu Trp Trp Ala Leu Lys Pro Gly Ala Pro Gln Pro
          20           25           30
Lys Ala Asn Gln Gln His Gln Asp Asn Ala Arg Gly Leu Val Leu Pro
          35           40           45
Gly Tyr Lys Tyr Leu Gly Pro Gly Asn Gly Leu Asp Lys Gly Glu Pro
          50           55           60
Val Asn Ala Ala Asp Ala Ala Ala Leu Glu His Asp Lys Ala Tyr Asp
65           70           75           80
Gln Gln Leu Lys Ala Gly Asp Asn Pro Tyr Leu Lys Tyr Asn His Ala
          85           90           95
Asp Ala Glu Phe Gln Glu Arg Leu Lys Glu Asp Thr Ser Phe Gly Gly
          100          105          110
Asn Leu Gly Arg Ala Val Phe Gln Ala Lys Lys Arg Leu Leu Glu Pro
          115          120          125
Leu Gly Leu Val Glu Glu Ala Ala Lys Thr Ala Pro Gly Lys Lys Arg
          130          135          140
Pro Val Glu Gln Ser Pro Gln Glu Pro Asp Ser Ser Ala Gly Ile Gly
145          150          155          160
Lys Ser Gly Ala Gln Pro Ala Lys Lys Arg Leu Asn Phe Gly Gln Thr
          165          170          175
Gly Asp Thr Glu Ser Val Pro Asp Pro Gln Pro Ile Gly Glu Pro Pro
          180          185          190
Ala Ala Pro Ser Gly Val Gly Ser Leu Thr Met Ala Ser Gly Gly Gly
          195          200          205
Ala Pro Val Ala Asp Asn Asn Glu Gly Ala Asp Gly Val Gly Ser Ser
          210          215          220
Ser Gly Asn Trp His Cys Asp Ser Gln Trp Leu Gly Asp Arg Val Ile
225          230          235          240
Thr Thr Ser Thr Arg Thr Trp Ala Leu Pro Thr Tyr Asn Asn His Leu
          245          250          255
Tyr Lys Gln Ile Ser Asn Ser Thr Ser Gly Gly Ser Ser Asn Asp Asn

```

260	265	270
Ala Tyr Phe Gly Tyr Ser Thr Pro Trp Gly Tyr Phe Asp Phe Asn Arg		
275	280	285
Phe His Cys His Phe Ser Pro Arg Asp Trp Gln Arg Leu Ile Asn Asn		
290	295	300
Asn Trp Gly Phe Arg Pro Lys Gln Leu Asn Phe Lys Leu Phe Asn Ile		
305	310	315
Gln Val Lys Glu Val Thr Asp Asn Asn Gly Val Lys Thr Ile Ala Asn		
325	330	335
Asn Leu Thr Ser Thr Val Gln Val Phe Thr Asp Ser Asp Tyr Gln Leu		
340	345	350
Pro Tyr Val Leu Gly Ser Ala His Glu Gly Cys Leu Pro Pro Phe Pro		
355	360	365
Ala Asp Val Phe Met Ile Pro Gln Tyr Gly Tyr Leu Thr Leu Asn Asp		
370	375	380
Gly Ser Gln Ala Val Gly Arg Ser Ser Phe Tyr Cys Leu Glu Tyr Phe		
385	390	395
Pro Ser Gln Met Leu Arg Thr Gly Asn Asn Phe Gln Phe Ser Tyr Glu		
405	410	415
Phe Glu Asn Val Pro Phe His Ser Ser Tyr Ala His Ser Gln Ser Leu		
420	425	430
Asp Arg Leu Met Asn Pro Leu Ile Asp Gln Tyr Leu Tyr Tyr Leu Ser		
435	440	445
Lys Thr Ile Asn Gly Ser Gly Gln Asn Gln Gln Thr Leu Lys Phe Ser		
450	455	460
Val Ala Gly Pro Ser Asn Met Ala Val Gln Gly Arg Asn Tyr Ile Pro		
465	470	475
Gly Pro Ser Tyr Arg Gln Gln Arg Val Ser Thr Thr Val Thr Gln Asn		
485	490	495
Asn Asn Ser Glu Phe Ala Trp Pro Gly Ala Ser Ser Trp Ala Leu Asn		
500	505	510
Gly Arg Asn Ser Leu Met Asn Pro Gly Pro Ala Met Ala Ser His Lys		
515	520	525
Glu Gly Glu Asp Arg Phe Phe Pro Leu Ser Gly Ser Leu Ile Phe Gly		
530	535	540
Lys Gln Gly Thr Gly Arg Asp Asn Val Asp Ala Asp Lys Val Met Ile		
545	550	555
Thr Asn Glu Glu Glu Ile Lys Thr Thr Asn Pro Val Ala Thr Glu Ser		
565	570	575

Tyr Gly Gln Val Ala Thr Asn His Gln Ser Ala Gln Ala Gln		
580	585	590
Thr Gly Trp Val Gln Asn Gln Gly Ile Leu Pro Gly Met Val Trp Gln		
595	600	605
Asp Arg Asp Val Tyr Leu Gln Gly Pro Ile Trp Ala Lys Ile Pro His		
610	615	620
Thr Asp Gly Asn Phe His Pro Ser Pro Leu Met Gly Gly Phe Gly Met		
625	630	635
Lys His Pro Pro Pro Gln Ile Leu Ile Lys Asn Thr Pro Val Pro Ala		
645	650	655
Asp Pro Pro Thr Ala Phe Asn Lys Asp Lys Leu Asn Ser Phe Ile Thr		
660	665	670
Gln Tyr Ser Thr Gly Gln Val Ser Val Glu Ile Glu Trp Glu Leu Gln		
675	680	685
Lys Glu Asn Ser Lys Arg Trp Asn Pro Glu Ile Gln Tyr Thr Ser Asn		
690	695	700
Tyr Tyr Lys Ser Asn Asn Val Glu Phe Ala Val Asn Thr Glu Gly Val		
705	710	715
Tyr Ser Glu Pro Arg Pro Ile Gly Thr Arg Tyr Leu Thr Arg Asn Leu		
725	730	735

<210> 3

<211> 736

<212> PRT

<213> 人工序列

<220>

<223> AAV分离株

<400> 3

Met Ala Ala Asp Gly Tyr Leu Pro Asp Trp Leu Glu Asp Asn Leu Ser		
1	5	10
Glu Gly Ile Arg Glu Trp Trp Ala Leu Lys Pro Gly Ala Pro Gln Pro		
20	25	30
Lys Ala Asn Gln Gln His Gln Asp Asn Ala Arg Gly Leu Val Leu Pro		
35	40	45
Gly Tyr Lys Tyr Leu Gly Pro Gly Asn Gly Leu Asp Lys Gly Glu Pro		
50	55	60
Val Asn Ala Ala Asp Ala Ala Ala Leu Glu His Asp Lys Ala Tyr Asp		
65	70	75
Gln Gln Leu Lys Ala Gly Asp Asn Pro Tyr Leu Lys Tyr Asn His Ala		
85	90	95

Asp	Ala	Glu	Phe	Gln	Glu	Arg	Leu	Lys	Glu	Asp	Thr	Ser	Phe	Gly	Gly			
																100	105	110
Asn	Leu	Gly	Arg	Ala	Val	Phe	Gln	Ala	Lys	Lys	Arg	Leu	Leu	Glu	Pro			
																115	120	125
Leu	Gly	Leu	Val	Glu	Glu	Ala	Ala	Lys	Thr	Ala	Pro	Gly	Lys	Lys	Arg			
																130	135	140
Pro	Val	Glu	Gln	Ser	Pro	Gln	Glu	Pro	Asp	Ser	Ser	Ala	Gly	Ile	Gly			
																145	150	155
Lys	Ser	Gly	Ala	Gln	Pro	Ala	Lys	Lys	Arg	Leu	Asn	Phe	Gly	Gln	Thr			
																165	170	175
Gly	Asp	Thr	Glu	Ser	Val	Pro	Asp	Pro	Gln	Pro	Ile	Gly	Glu	Pro	Pro			
																180	185	190
Ala	Ala	Pro	Ser	Gly	Val	Gly	Ser	Leu	Thr	Met	Ala	Ser	Gly	Gly	Gly			
																195	200	205
Ala	Pro	Val	Ala	Asp	Asn	Asn	Glu	Gly	Ala	Asp	Gly	Val	Gly	Ser	Ser			
																210	215	220
Ser	Gly	Asn	Trp	His	Cys	Asp	Ser	Gln	Trp	Leu	Gly	Asp	Arg	Val	Ile			
																225	230	235
Thr	Thr	Ser	Thr	Arg	Thr	Trp	Ala	Leu	Pro	Thr	Tyr	Asn	Asn	His	Leu			
																245	250	255
Tyr	Lys	Gln	Ile	Ser	Asn	Ser	Thr	Ser	Gly	Gly	Ser	Ser	Asn	Asp	Asn			
																260	265	270
Ala	Tyr	Phe	Gly	Tyr	Ser	Thr	Pro	Trp	Gly	Tyr	Phe	Asp	Phe	Asn	Arg			
																275	280	285
Phe	His	Cys	His	Phe	Ser	Pro	Arg	Asp	Trp	Gln	Arg	Leu	Ile	Asn	Asn			
																290	295	300
Asn	Trp	Gly	Phe	Arg	Pro	Lys	Arg	Leu	Asn	Phe	Lys	Leu	Phe	Asn	Ile			
																305	310	315
Gln	Val	Lys	Glu	Val	Thr	Asp	Asn	Asn	Gly	Val	Lys	Thr	Ile	Ala	Asn			
																325	330	335
Asn	Leu	Thr	Ser	Thr	Val	Gln	Val	Phe	Thr	Asp	Ser	Asp	Tyr	Gln	Leu			
																340	345	350
Pro	Tyr	Val	Leu	Gly	Ser	Ala	His	Glu	Gly	Cys	Leu	Pro	Pro	Phe	Pro			
																355	360	365
Ala	Asp	Val	Phe	Met	Ile	Pro	Gln	Tyr	Gly	Tyr	Leu	Thr	Leu	Asn	Asp			
																370	375	380
Gly	Ser	Gln	Ala	Val	Gly	Arg	Ser	Ser	Phe	Tyr	Cys	Leu	Glu	Tyr	Phe			
																385	390	395
Pro	Ser	Gln	Met	Leu	Arg	Thr	Gly	Asn	Asn	Phe	Gln	Phe	Ser	Tyr	Glu			

				405					410					415			
Phe	Glu	Asn	Val	Pro	Phe	His	Ser	Ser	Tyr	Ala	His	Ser	Gln	Ser	Leu		
				420					425					430			
Asp	Arg	Leu	Met	Asn	Pro	Leu	Ile	Asp	Gln	Tyr	Leu	Tyr	Tyr	Leu	Ser		
				435					440					445			
Lys	Thr	Ile	Asn	Gly	Ser	Gly	Gln	Asn	Gln	Gln	Thr	Leu	Lys	Phe	Ser		
				450					455					460			
Val	Ala	Gly	Pro	Ser	Asn	Met	Ala	Val	Gln	Gly	Arg	Asn	Tyr	Ile	Pro		
465									470					475			480
Gly	Pro	Ser	Tyr	Arg	Gln	Gln	Arg	Val	Ser	Thr	Thr	Val	Thr	Gln	Asn		
				485						490				495			
Asn	Asn	Ser	Glu	Phe	Ala	Trp	Pro	Gly	Ala	Ser	Ser	Trp	Ala	Leu	Asn		
				500					505					510			
Gly	Arg	Asn	Ser	Leu	Met	Asn	Pro	Gly	Pro	Ala	Met	Ala	Ser	His	Lys		
				515					520					525			
Glu	Gly	Glu	Asp	Arg	Phe	Phe	Pro	Leu	Ser	Gly	Ser	Leu	Ile	Phe	Gly		
				530						535				540			
Lys	Gln	Gly	Thr	Gly	Arg	Asp	Asn	Val	Asp	Ala	Asp	Lys	Val	Met	Ile		
545									550					555			560
Thr	Asn	Glu	Glu	Glu	Ile	Lys	Thr	Thr	Asn	Pro	Val	Ala	Thr	Glu	Ser		
				565						570				575			
Tyr	Gly	Gln	Val	Ala	Thr	Asn	His	Gln	Ser	Ala	Gln	Ala	Gln	Ala	Gln		
				580					585					590			
Thr	Gly	Trp	Val	Gln	Asn	Gln	Gly	Ile	Leu	Pro	Gly	Met	Val	Trp	Gln		
				595					600					605			
Asp	Arg	Asp	Val	Tyr	Leu	Gln	Gly	Pro	Ile	Trp	Ala	Lys	Ile	Pro	His		
				610					615					620			
Thr	Gly	Gly	Asn	Phe	His	Pro	Ser	Pro	Leu	Met	Gly	Gly	Phe	Gly	Met		
625									630					635			640
Lys	His	Pro	Pro	Pro	Gln	Ile	Leu	Ile	Lys	Asn	Thr	Pro	Val	Pro	Ala		
				645						650				655			
Asp	Pro	Pro	Thr	Ala	Phe	Asn	Lys	Asp	Lys	Leu	Asn	Ser	Phe	Ile	Thr		
				660					665					670			
Gln	Tyr	Ser	Thr	Gly	Gln	Val	Ser	Val	Glu	Ile	Glu	Trp	Glu	Leu	Gln		
				675					680					685			
Lys	Glu	Asn	Ser	Lys	Arg	Trp	Asn	Pro	Glu	Ile	Gln	Tyr	Thr	Ser	Asn		
				690					695					700			
Tyr	Tyr	Lys	Ser	Asn	Asn	Val	Glu	Phe	Ala	Val	Asn	Thr	Gly	Gly	Val		
705									710					715			720

Tyr Ser Glu Pro Arg Pro Ile Gly Thr Arg Tyr Leu Thr Arg Asn Leu
 725 730 735
 <210> 4
 <211> 736
 <212> PRT
 <213> 人工序列
 <220>
 <223> AAV分离株
 <400> 4
 Met Ala Ala Asp Gly Tyr Leu Pro Asp Trp Leu Glu Asp Asn Leu Ser
 1 5 10 15
 Glu Gly Ile Arg Glu Trp Trp Ala Leu Lys Pro Gly Ala Pro Gln Pro
 20 25 30
 Lys Ala Asn Gln Gln His Gln Asp Asn Ala Arg Gly Leu Val Leu Pro
 35 40 45
 Gly Tyr Lys Tyr Leu Gly Pro Gly Asn Gly Leu Asp Lys Gly Glu Pro
 50 55 60
 Ile Asn Ala Ala Asp Ala Ala Ala Leu Glu His Asp Lys Ala Tyr Asp
 65 70 75 80
 Gln Gln Leu Lys Ala Gly Asp Asn Pro Tyr Leu Lys Tyr Asn His Ala
 85 90 95
 Asp Ala Glu Phe Gln Glu Arg Leu Lys Glu Asp Thr Ser Phe Gly Gly
 100 105 110
 Asn Leu Gly Arg Ala Val Phe Gln Ala Lys Lys Arg Leu Leu Glu Pro
 115 120 125
 Leu Gly Leu Val Glu Glu Ala Ala Lys Thr Ala Pro Gly Lys Lys Arg
 130 135 140
 Pro Val Glu Gln Ser Pro Gln Glu Pro Asp Ser Ser Ala Gly Ile Gly
 145 150 155 160
 Lys Ser Gly Ala Gln Pro Ala Lys Lys Arg Leu Asn Phe Gly Gln Thr
 165 170 175
 Gly Asp Thr Glu Ser Val Pro Asp Pro Gln Pro Ile Gly Glu Pro Pro
 180 185 190
 Ala Ala Pro Ser Gly Val Gly Ser Leu Thr Met Ala Ser Gly Gly Gly
 195 200 205
 Ala Pro Val Ala Asp Asn Asn Glu Gly Ala Asp Gly Val Gly Ser Ser
 210 215 220
 Ser Gly Asn Trp His Cys Asp Ser Gln Trp Leu Gly Asp Arg Val Ile
 225 230 235 240

Thr	Thr	Ser	Thr	Arg	Thr	Trp	Ala	Leu	Pro	Thr	Tyr	Asn	Asn	His	Leu
				245					250					255	
Tyr	Lys	Gln	Ile	Ser	Asn	Ser	Thr	Ser	Gly	Gly	Ser	Ser	Asn	Asp	Asn
			260						265					270	
Ala	Tyr	Phe	Gly	Tyr	Ser	Thr	Pro	Trp	Gly	Tyr	Phe	Asp	Phe	Asn	Arg
			275						280					285	
Phe	His	Cys	His	Phe	Ser	Pro	Arg	Asp	Trp	Gln	Arg	Leu	Ile	Asn	Asn
			290						295					300	
Asn	Trp	Gly	Phe	Arg	Pro	Lys	Arg	Leu	Asn	Phe	Lys	Leu	Phe	Asn	Ile
305						310					315				320
Gln	Val	Lys	Glu	Val	Thr	Asp	Asn	Asn	Gly	Val	Lys	Thr	Ile	Ala	Asn
						325					330				335
Asn	Leu	Thr	Ser	Thr	Val	Gln	Val	Phe	Thr	Asp	Ser	Asp	Tyr	Gln	Leu
						340								350	
Pro	Tyr	Val	Leu	Gly	Ser	Ala	His	Glu	Gly	Cys	Leu	Pro	Pro	Phe	Pro
						355								365	
Ala	Asp	Val	Phe	Met	Ile	Pro	Gln	Tyr	Gly	Tyr	Leu	Thr	Leu	Asn	Asp
						370								380	
Gly	Ser	Gln	Ala	Val	Gly	Arg	Ser	Ser	Phe	Tyr	Cys	Leu	Glu	Tyr	Phe
385						390									400
Pro	Ser	Gln	Met	Leu	Arg	Thr	Gly	Asn	Asn	Phe	Gln	Phe	Ser	Tyr	Glu
						405									415
Phe	Glu	Asn	Val	Pro	Phe	His	Ser	Ser	Tyr	Ala	His	Ser	Gln	Ser	Leu
						420								430	
Asp	Arg	Leu	Met	Asn	Pro	Leu	Ile	Asp	Gln	Tyr	Leu	Tyr	Tyr	Leu	Ser
						435								445	
Lys	Thr	Ile	Asn	Gly	Ser	Gly	Gln	Asn	Gln	Gln	Thr	Leu	Lys	Phe	Ser
						450								460	
Val	Ala	Gly	Pro	Ser	Asn	Met	Ala	Val	Gln	Gly	Arg	Asn	Tyr	Ile	Pro
465						470									480
Gly	Pro	Ser	Tyr	Arg	Gln	Gln	Arg	Val	Ser	Thr	Thr	Val	Thr	Gln	Asn
						485								495	
Asn	Asn	Ser	Glu	Phe	Ala	Trp	Pro	Gly	Ala	Ser	Ser	Trp	Ala	Leu	Asn
						500								510	
Gly	Arg	Asn	Ser	Leu	Met	Asn	Pro	Gly	Pro	Ala	Met	Ala	Ser	His	Lys
						515								525	
Glu	Gly	Glu	Asp	Arg	Phe	Phe	Pro	Leu	Ser	Gly	Ser	Leu	Ile	Phe	Gly
						530								540	
Lys	Gln	Gly	Thr	Gly	Arg	Asp	Asn	Val	Asp	Ala	Asp	Lys	Val	Met	Ile

545	550	555	560
Thr Asn Glu Glu Glu Ile Lys Thr Thr Asn Pro Val Ala Thr Glu Ser			
	565	570	575
Tyr Gly Gln Val Ala Thr Asn His Gln Ser Ala Gln Ala Gln Ala Gln			
	580	585	590
Thr Gly Trp Val Gln Asn Gln Gly Ile Leu Pro Gly Met Val Trp Gln			
	595	600	605
Asp Arg Asp Val Tyr Leu Gln Gly Pro Ile Trp Ala Lys Ile Pro His			
	610	615	620
Thr Tyr Gly Asn Phe His Pro Ser Pro Leu Met Gly Gly Phe Gly Met			
625	630	635	640
Lys His Pro Pro Pro Gln Ile Leu Ile Lys Asn Thr Pro Val Pro Ala			
	645	650	655
Asp Pro Pro Thr Ala Phe Asn Lys Asp Lys Leu Asn Ser Phe Ile Thr			
	660	665	670
Gln Tyr Ser Thr Gly Gln Val Ser Val Glu Ile Glu Trp Glu Leu Gln			
	675	680	685
Lys Glu Asn Ser Lys Arg Trp Asn Pro Glu Ile Gln Tyr Thr Ser Asn			
	690	695	700
Tyr Tyr Lys Ser Asn Asn Val Glu Phe Ala Val Asn Thr Glu Gly Val			
705	710	715	720
Tyr Ser Glu Pro Arg Pro Ile Gly Thr Arg Tyr Leu Thr Arg Asn Leu			
	725	730	735
<210> 5			
<211> 736			
<212> PRT			
<213> 人工序列			
<220>			
<223> AAV分离株			
<400> 5			
Met Ala Ala Asp Gly Tyr Leu Pro Asp Trp Leu Glu Asp Asn Leu Ser			
1	5	10	15
Glu Gly Ile Arg Glu Trp Trp Ala Leu Lys Pro Gly Ala Pro Gln Pro			
	20	25	30
Lys Ala Asn Gln Gln His Gln Asp Asn Ala Arg Gly Leu Val Leu Pro			
	35	40	45
Gly Tyr Lys Tyr Leu Gly Pro Gly Asn Gly Leu Asp Lys Gly Glu Pro			
	50	55	60
Val Asn Ala Ala Asp Ala Ala Ala Leu Glu His Asp Lys Ala Tyr Asp			

65	70	75	80
Gln Gln Leu Lys Ala Gly Asp Asn Pro Tyr Leu Lys Tyr Asn His Ala			
	85	90	95
Asp Ala Glu Phe Gln Glu Arg Leu Lys Glu Asp Thr Ser Phe Gly Gly			
	100	105	110
Asn Leu Gly Arg Ala Val Phe Gln Ala Lys Lys Arg Leu Leu Glu Pro			
	115	120	125
Leu Gly Leu Val Glu Glu Ala Ala Lys Thr Ala Pro Gly Lys Lys Arg			
	130	135	140
Pro Val Glu Gln Ser Pro Gln Glu Pro Asp Ser Ser Ala Gly Ile Asp			
145	150	155	160
Lys Ser Gly Ala Gln Pro Ala Lys Lys Arg Leu Asn Phe Gly Gln Thr			
	165	170	175
Gly Asp Thr Glu Ser Val Pro Asp Pro Gln Pro Ile Gly Glu Pro Pro			
	180	185	190
Ala Ala Pro Ser Gly Val Gly Ser Leu Thr Met Ala Ser Gly Gly Gly			
	195	200	205
Ala Pro Val Ala Asp Asn Asn Glu Gly Ala Asp Gly Val Gly Ser Ser			
210	215	220	
Ser Gly Asn Trp His Cys Asp Ser Gln Trp Leu Gly Asp Arg Val Ile			
225	230	235	240
Thr Thr Ser Thr Arg Thr Trp Ala Leu Pro Thr Tyr Asn Asn His Leu			
	245	250	255
Tyr Lys Gln Ile Ser Asn Ser Thr Ser Gly Gly Ser Ser Asn Asp Asn			
	260	265	270
Ala Tyr Phe Gly Tyr Ser Thr Pro Trp Gly Tyr Phe Asp Phe Asn Arg			
	275	280	285
Phe His Cys His Phe Ser Pro Arg Asp Trp Gln Arg Leu Ile Asn Asn			
290	295	300	
Asn Trp Gly Phe Arg Pro Lys Arg Leu Asn Phe Lys Leu Phe Asn Ile			
305	310	315	320
Gln Val Lys Glu Val Thr Asp Asn Asn Gly Val Lys Thr Ile Ala Asn			
	325	330	335
Asn Leu Thr Ser Thr Val Gln Val Phe Thr Asp Ser Asp Tyr Gln Leu			
	340	345	350
Pro Tyr Val Leu Gly Ser Ala His Glu Gly Cys Leu Pro Pro Phe Pro			
	355	360	365
Ala Asp Val Phe Met Ile Pro Gln Tyr Gly Tyr Leu Thr Leu Asn Asp			
370	375	380	

Gly Ser Gln Ala Val Gly Arg Ser Ser Phe Tyr Cys Leu Glu Tyr Phe															
385							390					395			400
Pro Ser Gln Met Leu Arg Thr Gly Asn Asn Phe Gln Phe Ser Tyr Glu															
							405					410			415
Phe Glu Asn Val Pro Phe His Ser Ser Tyr Ala His Ser Gln Ser Leu															
							420					425			430
Asp Arg Leu Met Asn Pro Leu Ile Asp Gln Tyr Leu Tyr Tyr Leu Ser															
							435					440			445
Lys Thr Ile Asn Gly Ser Gly Gln Asn Gln Gln Thr Leu Lys Phe Ser															
							450					455			460
Val Ala Gly Pro Ser Asn Met Ala Val Gln Gly Arg Asn Tyr Ile Pro															
465							470					475			480
Gly Pro Ser Tyr Arg Gln Gln Arg Val Ser Thr Thr Val Thr Gln Asn															
							485					490			495
Asn Asn Ser Glu Phe Ala Trp Pro Gly Ala Ser Ser Trp Ala Leu Asn															
							500					505			510
Gly Arg Asn Ser Leu Met Asn Pro Gly Pro Ala Met Ala Ser His Lys															
							515					520			525
Glu Gly Glu Asp Arg Phe Phe Pro Leu Ser Gly Ser Leu Ile Phe Gly															
							530					535			540
Lys Gln Gly Thr Gly Arg Asp Asn Val Asp Ala Asp Lys Val Met Ile															
545							550					555			560
Thr Asn Glu Glu Glu Ile Lys Thr Thr Asn Pro Val Ala Thr Glu Ser															
							565					570			575
Tyr Gly Gln Val Ala Thr Asn His Gln Ser Ala Gln Ala Gln Ala Gln															
							580					585			590
Thr Gly Trp Val Gln Asn Gln Gly Ile Leu Pro Gly Met Val Trp Gln															
							595					600			605
Asp Arg Asp Val Tyr Leu Gln Gly Pro Ile Trp Ala Lys Ile Pro His															
							610					615			620
Thr Asp Gly Asn Phe His Pro Ser Pro Leu Met Gly Gly Phe Gly Met															
625							630					635			640
Lys His Pro Pro Pro Gln Ile Leu Ile Lys Asn Thr Pro Val Pro Ala															
							645					650			655
Asp Pro Pro Thr Ala Phe Asn Lys Asp Lys Leu Asn Ser Phe Ile Thr															
							660					665			670
Gln Tyr Ser Thr Gly Gln Val Ser Val Glu Ile Glu Trp Glu Leu Gln															
							675					680			685
Lys Glu Asn Ser Lys Arg Trp Asn Pro Glu Ile Gln Tyr Thr Ser Asn															

690	695	700
Tyr Tyr Lys Ser Asn Asn Val Glu Phe Ala Val Asn Thr Glu Gly Val		
705	710	715
Tyr Ser Glu Pro Arg Pro Ile Gly Thr Arg Tyr Leu Thr Arg Asn Leu		
725	730	735
<210> 6		
<211> 736		
<212> PRT		
<213> 人工序列		
<220>		
<223> AAV分离株		
<400> 6		
Met Ala Ala Asp Gly Tyr Leu Pro Asp Trp Leu Glu Asp Asn Leu Ser		
1	5	10
Glu Gly Ile Arg Glu Trp Trp Ala Leu Lys Pro Gly Ala Pro Gln Pro		
20	25	30
Lys Ala Asn Gln Gln His Gln Asp Asn Ala Arg Gly Leu Val Leu Pro		
35	40	45
Gly Tyr Lys Tyr Leu Gly Pro Gly Asn Gly Leu Asp Lys Gly Glu Pro		
50	55	60
Val Asn Ala Ala Asp Ala Ala Ala Leu Glu His Asp Lys Ala Tyr Asp		
65	70	75
Gln Gln Leu Lys Ala Gly Asp Asn Pro Tyr Leu Lys Tyr Asn His Ala		
85	90	95
Asp Ala Glu Phe Gln Glu Arg Leu Lys Glu Asp Thr Ser Phe Gly Gly		
100	105	110
Asn Leu Gly Arg Ala Val Leu Gln Ala Lys Lys Arg Leu Leu Glu Pro		
115	120	125
Leu Gly Leu Val Glu Glu Ala Ala Lys Thr Ala Pro Gly Lys Lys Arg		
130	135	140
Pro Val Glu Gln Ser Pro Gln Glu Pro Asp Ser Ser Ala Gly Ile Gly		
145	150	155
Lys Ser Gly Ala Gln Pro Ala Lys Lys Arg Leu Asn Phe Gly Gln Thr		
165	170	175
Gly Asp Thr Glu Ser Val Pro Asp Pro Gln Pro Ile Gly Glu Pro Pro		
180	185	190
Ala Ala Pro Ser Gly Val Gly Ser Leu Thr Met Ala Ser Gly Gly Gly		
195	200	205
Ala Pro Val Ala Asp Asn Asn Glu Gly Ala Asp Gly Val Gly Ser Ser		

210	215	220
Ser Gly Asn Trp His Cys Asp Ser Gln Trp Leu Gly Asp Arg Val Ile		
225	230	235
Thr Thr Ser Thr Arg Thr Trp Ala Leu Pro Thr Tyr Asn Asn His Leu		240
	245	250
Tyr Lys Gln Ile Ser Asn Ser Thr Ser Gly Gly Ser Ser Asn Asp Asn		255
	260	265
Ala Tyr Phe Gly Tyr Ser Thr Pro Trp Gly Tyr Phe Asp Phe Asn Arg		270
	275	280
Phe His Cys His Phe Ser Pro Arg Asp Trp Gln Arg Leu Ile Asn Asn		285
290	295	300
Asn Trp Gly Phe Arg Pro Lys Arg Leu Asn Phe Lys Leu Phe Asn Ile		
305	310	315
Gln Val Lys Glu Val Thr Asp Asn Asn Gly Val Lys Thr Ile Ala Asn		320
	325	330
Asn Leu Thr Ser Thr Val Gln Val Phe Thr Asp Ser Asp Tyr Gln Leu		335
	340	345
Pro Tyr Val Leu Gly Ser Ala His Glu Gly Cys Leu Pro Pro Phe Pro		350
355	360	365
Ala Asp Val Phe Met Ile Pro Gln Tyr Gly Tyr Leu Thr Leu Asn Asp		
370	375	380
Gly Ser Gln Ala Val Gly Arg Ser Ser Phe Tyr Cys Leu Glu Tyr Phe		
385	390	395
Pro Ser Gln Met Leu Arg Thr Gly Asn Asn Phe Gln Phe Ser Tyr Glu		400
	405	410
Phe Glu Asn Val Pro Phe His Ser Ser Tyr Ala His Ser Gln Ser Leu		415
	420	425
Asp Arg Leu Met Asn Pro Leu Ile Asp Gln Tyr Leu Tyr Tyr Leu Ser		430
	435	440
Lys Thr Ile Asn Gly Ser Gly Gln Asn Gln Gln Thr Leu Lys Phe Ser		445
450	455	460
Val Ala Gly Ser Ser Asn Met Ala Val Gln Gly Arg Asn Tyr Ile Pro		
465	470	475
Gly Pro Ser Tyr Arg Gln Gln Arg Val Ser Thr Thr Val Thr Gln Asn		480
	485	490
Asn Asn Ser Glu Phe Ala Trp Pro Gly Ala Ser Ser Trp Ala Leu Asn		495
	500	505
Gly Arg Asn Ser Leu Met Asn Pro Gly Pro Ala Met Ala Ser His Lys		510
515	520	525

Glu Gly Glu Asp Arg Phe Phe Pro Leu Ser Gly Ser Leu Ile Phe Gly		
530	535	540
Lys Gln Gly Thr Gly Arg Asp Asn Val Asp Ala Asp Lys Val Met Ile		
545	550	555 560
Thr Asn Glu Glu Glu Ile Lys Thr Thr Asn Pro Val Ala Thr Glu Ser		
	565	570 575
Tyr Gly Gln Val Ala Thr Asn His Gln Ser Ala Gln Ala Gln Ala Gln		
	580	585 590
Thr Gly Trp Val Gln Asn Gln Gly Ile Leu Pro Gly Met Val Trp Gln		
	595	600 605
Asp Arg Asp Val Tyr Leu Gln Gly Pro Ile Trp Ala Lys Ile Pro His		
	610	615 620
Thr Asp Gly Asn Phe His Pro Ser Pro Leu Met Gly Gly Phe Gly Met		
625	630	635 640
Lys His Pro Pro Pro Gln Ile Leu Ile Lys Asn Thr Pro Val Pro Ala		
	645	650 655
Asp Pro Pro Thr Ala Phe Asn Lys Asp Lys Leu Asn Ser Phe Ile Thr		
	660	665 670
Gln Tyr Ser Thr Gly Gln Val Ser Val Glu Ile Glu Trp Glu Leu Gln		
	675	680 685
Lys Glu Asn Ser Lys Arg Trp Asn Pro Glu Ile Gln Tyr Thr Ser Asn		
	690	695 700
Tyr Tyr Lys Ser Asn Asn Val Glu Phe Ala Val Asn Thr Glu Gly Val		
705	710	715 720
Tyr Ser Glu Pro Arg Pro Ile Gly Thr Arg Tyr Leu Thr Arg Asn Leu		
	725	730 735

<210> 7

<211> 736

<212> PRT

<213> 人工序列

<220>

<223> AAV分离株

<400> 7

Met Ala Ala Asp Gly Tyr Leu Pro Asp Trp Leu Glu Asp Asn Leu Ser		
1	5	10 15
Glu Gly Ile Arg Glu Trp Trp Ala Leu Lys Pro Gly Ala Pro Gln Pro		
	20	25 30
Lys Ala Asn Gln Gln His Gln Asp Asn Ala Arg Gly Leu Val Leu Pro		
	35	40 45

Gly Tyr Lys Tyr Leu Gly Pro Gly Asn Gly Leu Asp Lys Gly Glu Pro		
50	55	60
Val Asn Ala Ala Asp Ala Ala Ala Leu Glu His Asp Lys Ala Tyr Asp		
65	70	75 80
Gln Gln Leu Lys Ala Gly Asp Asn Pro Tyr Leu Lys Tyr Asn His Ala		
85	90	95
Asp Ala Glu Phe Gln Glu Arg Leu Lys Glu Asp Thr Ser Phe Gly Gly		
100	105	110
Asn Leu Gly Arg Ala Val Phe Gln Ala Lys Lys Arg Leu Leu Glu Pro		
115	120	125
Leu Gly Leu Val Glu Glu Ala Ala Lys Thr Ala Pro Gly Lys Lys Arg		
130	135	140
Pro Val Glu Gln Ser Pro Gln Glu Pro Asp Ser Ser Ala Gly Ile Gly		
145	150	155 160
Lys Ser Gly Ala Gln Pro Ala Lys Lys Arg Leu Asn Phe Gly Gln Thr		
165	170	175
Gly Asp Thr Glu Ser Val Pro Asp Pro Gln Pro Ile Gly Glu Pro Pro		
180	185	190
Ala Ala Pro Ser Gly Val Gly Ser Leu Thr Met Ala Ser Gly Gly Gly		
195	200	205
Ala Pro Val Ala Asp Asn Asn Glu Gly Ala Asp Gly Val Gly Ser Ser		
210	215	220
Ser Gly Asn Trp His Cys Asp Ser Gln Trp Leu Gly Asp Arg Val Ile		
225	230	235 240
Thr Thr Ser Thr Arg Thr Trp Ala Leu Pro Thr Tyr Asn Asn His Leu		
245	250	255
Tyr Lys Gln Ile Ser Asn Ser Thr Ser Gly Gly Ser Ser Asn Asp Asn		
260	265	270
Ala Tyr Phe Gly Tyr Ser Thr Pro Trp Gly Tyr Phe Asp Phe Asn Arg		
275	280	285
Phe His Cys His Phe Ser Pro Arg Asp Trp Gln Arg Leu Ile Asn Asn		
290	295	300
Asn Trp Gly Phe Arg Pro Lys Arg Leu Asn Phe Lys Leu Phe Asn Ile		
305	310	315 320
Gln Val Lys Glu Val Thr Asp Asn Asn Gly Val Lys Thr Ile Ala Asn		
325	330	335
Asn Leu Thr Ser Thr Val Gln Val Phe Thr Asp Ser Asp Tyr Gln Leu		
340	345	350
Pro Tyr Val Leu Gly Ser Ala His Glu Gly Cys Leu Pro Pro Phe Pro		

355	360	365
Ala Asp Val Phe Met Ile Pro Gln Tyr Gly Tyr Leu Thr Leu Asn Asp		
370	375	380
Gly Ser Gln Ala Val Gly Arg Ser Ser Phe Tyr Cys Leu Glu Tyr Phe		
385	390	395
Pro Ser Gln Met Leu Arg Thr Gly Asn Asn Phe Gln Phe Ser Tyr Glu		
405	410	415
Phe Glu Asn Val Pro Phe His Ser Ser Tyr Ala His Ser Gln Ser Leu		
420	425	430
Asp Arg Leu Met Asn Pro Leu Ile Asp Gln Tyr Leu Tyr Tyr Leu Ser		
435	440	445
Lys Thr Ile Asn Gly Ser Gly Gln Asn Gln Gln Thr Leu Lys Phe Ser		
450	455	460
Val Ala Gly Pro Ser Asn Met Ala Val Gln Gly Arg Asn Tyr Ile Pro		
465	470	475
Gly Pro Ser Tyr Arg Gln Gln Arg Val Ser Thr Thr Val Thr Gln Asn		
485	490	495
Asn Asn Ser Glu Phe Ala Trp Pro Gly Ala Ser Ser Trp Ala Leu Asn		
500	505	510
Gly Arg Asn Ser Leu Met Asn Pro Gly Pro Ala Met Ala Ser His Lys		
515	520	525
Glu Gly Glu Asp Arg Phe Phe Pro Leu Ser Gly Ser Leu Ile Phe Gly		
530	535	540
Lys Gln Gly Thr Gly Arg Asp Asn Val Asp Ala Asp Lys Val Met Ile		
545	550	555
Thr Asn Glu Glu Glu Ile Lys Thr Thr Asn Pro Val Ala Thr Glu Ser		
565	570	575
Tyr Gly Gln Val Ala Thr Asn His Gln Ser Ala Gln Ala Arg Ala Gln		
580	585	590
Thr Gly Trp Val Gln Asn Gln Gly Ile Leu Pro Gly Met Val Trp Gln		
595	600	605
Asp Arg Asp Val Tyr Leu Gln Gly Pro Ile Trp Ala Lys Ile Pro His		
610	615	620
Thr Asp Gly Asn Phe His Pro Ser Pro Leu Met Gly Gly Phe Gly Met		
625	630	635
Lys His Pro Pro Pro Gln Ile Leu Ile Lys Asn Thr Pro Val Pro Ala		
645	650	655
Asp Pro Pro Thr Ala Phe Asn Lys Asp Lys Leu Asn Ser Phe Ile Thr		
660	665	670

Gln Tyr Ser Thr Gly Gln Val Ser Val Glu Ile Glu Trp Glu Leu Gln
 675 680 685
 Lys Glu Asn Ser Lys Arg Trp Asn Pro Glu Ile Gln Tyr Thr Ser Asn
 690 695 700
 Tyr Tyr Lys Ser Asn Asn Val Glu Phe Ala Val Asn Thr Glu Gly Val
 705 710 715 720
 Tyr Ser Glu Pro Arg Pro Ile Gly Thr Arg Tyr Leu Thr Arg Asn Leu
 725 730 735
 <210> 8
 <211> 736
 <212> PRT
 <213> 人工序列
 <220>
 <223> AAV分离株
 <400> 8
 Met Ala Ala Asp Gly Tyr Leu Pro Asp Trp Leu Glu Asp Asn Leu Ser
 1 5 10 15
 Glu Gly Ile Arg Glu Trp Trp Ala Leu Lys Pro Gly Ala Pro Gln Pro
 20 25 30
 Lys Ala Asn Gln Gln His Gln Asp Asn Ala Arg Gly Leu Val Leu Pro
 35 40 45
 Gly Tyr Lys Tyr Leu Gly Pro Gly Asn Gly Leu Asp Lys Gly Glu Pro
 50 55 60
 Val Asn Ala Val Asp Ala Ala Ala Leu Glu His Asp Lys Ala Tyr Asp
 65 70 75 80
 Gln Gln Leu Lys Ala Gly Asp Asn Pro Tyr Leu Lys Tyr Asn His Ala
 85 90 95
 Asp Ala Glu Phe Gln Glu Arg Leu Lys Glu Asp Thr Ser Phe Gly Gly
 100 105 110
 Asn Leu Gly Arg Ala Val Phe Gln Ala Lys Lys Arg Leu Leu Glu Pro
 115 120 125
 Leu Gly Leu Val Glu Glu Ala Ala Lys Thr Ala Pro Gly Lys Lys Arg
 130 135 140
 Pro Val Glu Gln Ser Pro Gln Glu Pro Asp Ser Ser Ala Gly Ile Gly
 145 150 155 160
 Lys Ser Gly Ala Gln Pro Ala Lys Lys Arg Leu Asn Phe Gly Gln Thr
 165 170 175
 Gly Asp Thr Glu Ser Val Pro Asp Pro Gln Pro Ile Gly Glu Pro Pro
 180 185 190

Ala Ala Pro Ser Gly Val Gly Ser Leu Thr Met Ala Ser Gly Gly Gly		
195	200	205
Ala Pro Val Ala Asp Asn Asn Glu Gly Ala Asp Gly Val Gly Ser Ser		
210	215	220
Ser Gly Asn Trp His Cys Asp Ser Gln Trp Leu Gly Asp Arg Val Ile		
225	230	235
Thr Thr Ser Thr Arg Thr Trp Ala Leu Pro Thr Tyr Asn Asn His Leu		
	245	250
		255
Tyr Lys Gln Ile Ser Asn Ser Thr Ser Gly Gly Ser Ser Asn Asp Asn		
	260	265
		270
Ala Tyr Phe Gly Tyr Ser Thr Pro Trp Gly Tyr Phe Asp Phe Asn Arg		
	275	280
		285
Phe His Cys His Phe Ser Pro Arg Asp Trp Gln Arg Leu Ile Asn Asn		
	290	295
		300
Asn Trp Gly Phe Arg Pro Lys Arg Leu Asn Phe Lys Leu Phe Asn Ile		
305	310	315
		320
Gln Val Lys Glu Val Thr Asp Asn Asn Gly Val Lys Thr Ile Ala Asn		
	325	330
		335
Asn Leu Thr Ser Thr Val Gln Val Phe Thr Asp Ser Asp Tyr Gln Leu		
	340	345
		350
Pro Tyr Val Leu Gly Ser Ala His Glu Gly Cys Leu Pro Pro Phe Pro		
	355	360
		365
Ala Asp Val Phe Met Ile Pro Gln Tyr Gly Tyr Leu Thr Leu Asn Asp		
	370	375
		380
Gly Ser Gln Ala Val Gly Arg Ser Ser Phe Tyr Cys Leu Glu Tyr Phe		
385	390	395
		400
Pro Ser Gln Met Leu Arg Thr Gly Asn Asn Phe Gln Phe Ser Tyr Glu		
	405	410
		415
Phe Glu Asn Val Pro Phe His Ser Ser Tyr Ala His Ser Gln Ser Leu		
	420	425
		430
Asp Arg Leu Met Asn Pro Leu Ile Asp Gln Tyr Leu Tyr Tyr Leu Ser		
	435	440
		445
Lys Thr Ile Asn Gly Ser Gly Gln Asn Gln Gln Thr Leu Lys Phe Ser		
	450	455
		460
Val Ala Gly Pro Ser Asn Met Ala Val Gln Gly Arg Asn Tyr Ile Pro		
465	470	475
		480
Gly Pro Ser Tyr Arg Gln Gln Arg Val Ser Thr Thr Val Thr Gln Asn		
	485	490
		495
Asn Asn Ser Glu Phe Ala Trp Pro Gly Ala Ser Ser Trp Ala Leu Asn		

500	505	510
Gly Arg Asn Ser Leu Met Asn Pro Gly Pro Ala Met Ala Ser His Lys		
515	520	525
Glu Gly Glu Asp Arg Phe Phe Pro Leu Ser Gly Ser Leu Ile Phe Gly		
530	535	540
Lys Gln Gly Thr Gly Arg Asp Asn Val Asp Ala Asp Lys Val Met Ile		
545	550	555
Thr Asn Glu Glu Glu Ile Lys Thr Thr Asn Pro Val Ala Thr Glu Ser		
565	570	575
Tyr Gly Gln Val Ala Thr Asn His Gln Ser Ala Gln Ala Gln Ala Gln		
580	585	590
Thr Gly Trp Val Gln Asn Gln Gly Ile Leu Pro Gly Met Val Trp Gln		
595	600	605
Asp Arg Asp Val Tyr Leu Gln Gly Pro Ile Trp Ala Lys Ile Pro His		
610	615	620
Thr Asp Gly Asn Phe His Pro Ser Pro Leu Met Gly Gly Phe Gly Met		
625	630	635
Lys His Pro Pro Pro Gln Ile Leu Ile Lys Asn Thr Pro Val Pro Ala		
645	650	655
Asp Pro Pro Thr Ala Phe Asn Lys Asp Lys Leu Asn Ser Phe Ile Thr		
660	665	670
Gln Tyr Ser Thr Gly Gln Val Ser Val Glu Ile Glu Trp Glu Leu Gln		
675	680	685
Lys Glu Asn Ser Lys Arg Trp Asn Pro Glu Ile Gln Tyr Thr Ser Asn		
690	695	700
Tyr Tyr Lys Ser Asn Asn Val Glu Phe Ala Val Asn Thr Glu Gly Val		
705	710	715
Tyr Ser Glu Pro Arg Pro Ile Gly Thr Arg Tyr Leu Thr Arg Asn Leu		
725	730	735

<210> 9

<211> 736

<212> PRT

<213> 人工序列

<220>

<223> AAV分离株

<400> 9

Met Ala Ala Asp Gly Tyr Leu Pro Asp Trp Leu Glu Asp Asn Leu Ser		
1	5	10
Glu Gly Ile Arg Glu Trp Trp Ala Leu Lys Pro Gly Ala Pro Gln Pro		

20	25	30
Lys Ala Asn Gln Gln His Gln Asp Asn Ala Arg Gly Leu Val Leu Pro		
35	40	45
Gly Tyr Lys Tyr Leu Gly Pro Gly Asn Gly Leu Asp Lys Gly Glu Pro		
50	55	60
Val Asn Ala Ala Asp Ala Ala Ala Leu Glu His Asp Lys Ala Tyr Asp		
65	70	75
Gln Gln Leu Lys Ala Gly Asp Asn Pro Tyr Leu Lys Tyr Asn His Ala		
85	90	95
Asp Ala Glu Phe Gln Glu Arg Leu Lys Glu Asp Thr Ser Phe Gly Gly		
100	105	110
Asn Leu Gly Arg Ala Val Phe Gln Ala Lys Lys Arg Leu Leu Glu Pro		
115	120	125
Leu Gly Leu Val Glu Glu Ala Ala Lys Thr Ala Pro Gly Lys Lys Arg		
130	135	140
Pro Val Glu Gln Ser Pro Arg Glu Pro Asp Ser Ser Ala Gly Ile Gly		
145	150	155
Lys Ser Gly Ala Gln Pro Ala Lys Lys Arg Leu Asn Phe Gly Gln Thr		
165	170	175
Gly Asp Thr Glu Ser Val Pro Asp Pro Gln Pro Ile Gly Glu Pro Pro		
180	185	190
Ala Ala Pro Ser Gly Val Gly Ser Leu Thr Met Ala Ser Gly Gly Gly		
195	200	205
Ala Pro Val Ala Asp Asn Asn Glu Gly Ala Asp Gly Val Gly Ser Ser		
210	215	220
Ser Gly Asn Trp His Cys Asp Ser Gln Trp Leu Gly Asp Arg Val Ile		
225	230	235
Thr Thr Ser Thr Arg Thr Trp Ala Leu Pro Thr Tyr Asn Asn His Leu		
245	250	255
Tyr Lys Gln Ile Ser Asn Ser Thr Ser Gly Gly Ser Ser Asn Asp Asn		
260	265	270
Ala Tyr Phe Gly Tyr Ser Thr Pro Trp Gly Tyr Phe Asp Phe Asn Arg		
275	280	285
Phe His Cys His Phe Ser Pro Arg Asp Trp Gln Arg Leu Ile Asn Asn		
290	295	300
Asn Trp Gly Phe Arg Pro Lys Arg Leu Asn Phe Lys Leu Phe Asn Ile		
305	310	315
Gln Val Lys Glu Val Thr Asp Asn Asn Gly Val Lys Thr Ile Ala Asn		
325	330	335

Asn	Leu	Thr	Ser	Thr	Val	Gln	Val	Phe	Thr	Asp	Ser	Asp	Tyr	Gln	Leu
Pro	Tyr	Val	Leu	Gly	Ser	Ala	His	Glu	Gly	Cys	Leu	Pro	Pro	Phe	Pro
Ala	Asp	Val	Phe	Met	Ile	Pro	Gln	Tyr	Gly	Tyr	Leu	Thr	Leu	Asn	Asp
Gly	Ser	Gln	Ala	Val	Gly	Arg	Ser	Ser	Phe	Tyr	Cys	Leu	Glu	Tyr	Phe
Pro	Ser	Gln	Met	Leu	Arg	Thr	Gly	Asn	Asn	Phe	Gln	Phe	Ser	Tyr	Glu
Phe	Glu	Asn	Val	Pro	Phe	His	Ser	Ser	Tyr	Ala	His	Ser	Gln	Ser	Leu
Asp	Arg	Leu	Met	Asn	Pro	Leu	Ile	Asp	Gln	Tyr	Leu	Tyr	Tyr	Leu	Ser
Lys	Thr	Ile	Asn	Gly	Ser	Gly	Gln	Asn	Gln	Gln	Thr	Leu	Lys	Phe	Ser
Val	Ala	Gly	Pro	Ser	Asn	Met	Ala	Val	Gln	Gly	Arg	Asn	Tyr	Ile	Pro
Gly	Pro	Ser	Tyr	Arg	Gln	Gln	Arg	Val	Ser	Thr	Thr	Val	Thr	Gln	Asn
Asn	Asn	Ser	Glu	Phe	Ala	Trp	Pro	Gly	Ala	Ser	Ser	Trp	Ala	Leu	Asn
Gly	Arg	Asn	Ser	Leu	Met	Asn	Pro	Gly	Pro	Ala	Met	Ala	Ser	His	Lys
Glu	Gly	Glu	Asp	Arg	Phe	Phe	Pro	Leu	Ser	Gly	Ser	Leu	Ile	Phe	Gly
Lys	Gln	Gly	Thr	Gly	Arg	Asp	Asn	Val	Asp	Ala	Asp	Lys	Val	Met	Ile
Thr	Asn	Glu	Glu	Glu	Ile	Lys	Thr	Thr	Asn	Pro	Val	Ala	Thr	Glu	Ser
Tyr	Gly	Gln	Val	Ala	Thr	Asn	His	Gln	Ser	Ala	Gln	Ala	Gln	Ala	Gln
Thr	Gly	Trp	Val	Gln	Asn	Gln	Gly	Ile	Leu	Pro	Gly	Met	Val	Trp	Gln
Asp	Arg	Asp	Val	Tyr	Leu	Gln	Gly	Pro	Ile	Trp	Ala	Lys	Ile	Pro	His
Thr	Asp	Gly	Asn	Phe	His	Pro	Ser	Pro	Leu	Met	Gly	Gly	Phe	Gly	Met
Lys	His	Pro	Pro	Pro	Gln	Ile	Leu	Ile	Lys	Asn	Thr	Pro	Val	Pro	Ala

				645				650				655			
Asp	Pro	Pro	Thr	Ala	Phe	Asn	Lys	Asp	Lys	Leu	Asn	Ser	Phe	Ile	Thr
				660				665				670			
Gln	Tyr	Ser	Thr	Gly	Gln	Val	Ser	Val	Glu	Ile	Glu	Trp	Glu	Leu	Gln
				675				680				685			
Lys	Glu	Asn	Ser	Lys	Arg	Trp	Asn	Pro	Glu	Ile	Gln	Tyr	Thr	Ser	Asn
				690				695				700			
Tyr	Tyr	Lys	Ser	Asn	Asn	Val	Glu	Phe	Ala	Val	Asn	Thr	Glu	Gly	Val
705				710				715				720			
Tyr	Ser	Glu	Pro	Arg	Pro	Ile	Gly	Thr	Arg	Tyr	Leu	Thr	Arg	Asn	Leu
				725				730				735			
<210> 10															
<211> 736															
<212> PRT															
<213> 人工序列															
<220>															
<223> AAV分离株															
<400> 10															
Met	Ala	Ala	Asp	Gly	Tyr	Leu	Pro	Asp	Trp	Leu	Glu	Asp	Asn	Leu	Ser
1				5				10						15	
Glu	Gly	Ile	Arg	Glu	Trp	Trp	Ala	Leu	Lys	Pro	Gly	Ala	Pro	Gln	Pro
				20				25				30			
Lys	Ala	Asn	Gln	Gln	His	Gln	Asp	Asn	Ala	Arg	Gly	Leu	Val	Leu	Pro
				35				40				45			
Gly	Tyr	Lys	Tyr	Leu	Gly	Pro	Gly	Asn	Gly	Leu	Asp	Lys	Gly	Glu	Pro
				50				55				60			
Val	Asn	Ala	Ala	Asp	Ala	Ala	Ala	Leu	Glu	His	Asp	Lys	Ala	Tyr	Asp
65				70				75				80			
Gln	Gln	Leu	Lys	Ala	Gly	Asp	Asn	Pro	Tyr	Leu	Lys	Tyr	Asn	His	Ala
				85				90				95			
Asp	Ala	Glu	Phe	Gln	Glu	Arg	Leu	Lys	Glu	Asp	Thr	Ser	Phe	Gly	Gly
				100				105				110			
Asn	Leu	Gly	Arg	Ala	Val	Phe	Gln	Ala	Lys	Lys	Arg	Leu	Leu	Glu	Pro
				115				120				125			
Leu	Gly	Leu	Val	Glu	Glu	Ala	Ala	Lys	Thr	Ala	Pro	Gly	Lys	Lys	Arg
				130				135				140			
Pro	Val	Glu	Gln	Ser	Pro	Gln	Glu	Pro	Asp	Ser	Ser	Ala	Gly	Ile	Gly
145				150				155				160			
Lys	Ser	Gly	Ala	Gln	Pro	Ala	Lys	Lys	Arg	Leu	Asn	Phe	Gly	Gln	Thr

				165					170					175			
Gly	Asp	Thr	Glu	Ser	Val	Pro	Asp	Pro	Gln	Pro	Ile	Gly	Glu	Pro	Pro		
				180					185					190			
Ala	Ala	Pro	Ser	Gly	Val	Gly	Ser	Leu	Thr	Met	Ala	Ser	Cys	Gly	Gly		
				195					200					205			
Ala	Pro	Val	Ala	Asp	Asn	Asn	Glu	Gly	Ala	Asp	Gly	Val	Gly	Ser	Ser		
				210					215					220			
Ser	Gly	Asn	Trp	His	Cys	Asp	Ser	Gln	Trp	Leu	Gly	Asp	Arg	Val	Ile		
225					230				235					240			
Thr	Thr	Ser	Thr	Arg	Thr	Trp	Ala	Leu	Pro	Thr	Tyr	Asn	Asn	His	Leu		
				245					250					255			
Tyr	Lys	Gln	Ile	Ser	Asn	Ser	Thr	Ser	Gly	Gly	Ser	Ser	Asn	Asp	Asn		
				260					265					270			
Ala	Tyr	Phe	Gly	Tyr	Ser	Thr	Pro	Trp	Gly	Tyr	Phe	Asp	Phe	Asn	Arg		
				275					280					285			
Phe	His	Cys	His	Phe	Ser	Pro	Arg	Asp	Trp	Gln	Arg	Leu	Ile	Asn	Asn		
				290					295					300			
Asn	Trp	Gly	Phe	Arg	Pro	Lys	Arg	Leu	Asn	Phe	Lys	Leu	Phe	Asn	Ile		
305					310				315					320			
Gln	Val	Lys	Glu	Val	Thr	Asp	Asn	Asn	Gly	Val	Lys	Thr	Ile	Ala	Asn		
				325					330					335			
Asn	Leu	Thr	Ser	Thr	Val	Gln	Val	Phe	Thr	Asp	Ser	Asp	Tyr	Gln	Leu		
				340					345					350			
Pro	Tyr	Val	Leu	Gly	Ser	Ala	His	Glu	Gly	Cys	Leu	Pro	Pro	Phe	Pro		
				355					360					365			
Ala	Asp	Val	Phe	Met	Ile	Pro	Gln	Tyr	Gly	Tyr	Leu	Thr	Leu	Asn	Asp		
				370					375					380			
Gly	Ser	Gln	Ala	Val	Gly	Arg	Ser	Ser	Phe	Tyr	Cys	Leu	Glu	Tyr	Phe		
385					390				395					400			
Pro	Ser	Gln	Met	Leu	Arg	Thr	Gly	Asn	Asn	Phe	Gln	Phe	Ser	Tyr	Glu		
				405					410					415			
Phe	Glu	Asn	Val	Pro	Phe	His	Ser	Ser	Tyr	Ala	His	Ser	Gln	Ser	Leu		
				420					425					430			
Asp	Arg	Leu	Met	Asn	Pro	Leu	Ile	Asp	Gln	Tyr	Leu	Tyr	Tyr	Leu	Ser		
				435					440					445			
Lys	Thr	Ile	Asn	Gly	Ser	Gly	Gln	Asn	Gln	Gln	Thr	Leu	Lys	Phe	Ser		
				450					455					460			
Val	Ala	Gly	Pro	Ser	Asn	Met	Ala	Val	Gln	Gly	Arg	Asn	Tyr	Ile	Pro		
465					470				475					480			

Gly	Pro	Ser	Tyr	Arg	Gln	Gln	Arg	Val	Ser	Thr	Thr	Val	Thr	Gln	Asn	
				485					490					495		
Asn	Asn	Ser	Glu	Phe	Ala	Trp	Pro	Gly	Ala	Ser	Ser	Trp	Ala	Leu	Asn	
				500					505					510		
Gly	Arg	Asn	Ser	Leu	Met	Asn	Pro	Gly	Pro	Ala	Met	Ala	Ser	His	Lys	
				515					520					525		
Glu	Gly	Glu	Asp	Arg	Phe	Phe	Pro	Leu	Ser	Gly	Ser	Leu	Ile	Phe	Gly	
				530					535					540		
Lys	Gln	Gly	Thr	Gly	Arg	Asp	Asn	Val	Asp	Ala	Asp	Lys	Val	Met	Ile	
545					550					555					560	
Thr	Asn	Glu	Glu	Glu	Ile	Lys	Thr	Thr	Asn	Pro	Val	Ala	Thr	Glu	Ser	
				565					570					575		
Tyr	Gly	Gln	Val	Ala	Thr	Asn	His	Gln	Ser	Ala	Gln	Ala	Gln	Ala	Gln	
				580					585					590		
Thr	Gly	Trp	Val	Gln	Asn	Gln	Gly	Ile	Leu	Pro	Gly	Met	Val	Trp	Gln	
				595					600					605		
Asp	Arg	Asp	Val	Tyr	Leu	Gln	Gly	Pro	Ile	Trp	Ala	Lys	Ile	Pro	His	
				610					615					620		
Thr	Asp	Gly	Asn	Phe	His	Pro	Ser	Pro	Leu	Met	Gly	Gly	Phe	Gly	Met	
625					630					635					640	
Lys	His	Pro	Pro	Pro	Gln	Ile	Leu	Ile	Lys	Asn	Thr	Pro	Val	Pro	Ala	
				645					650					655		
Asp	Pro	Pro	Thr	Ala	Phe	Asn	Lys	Asp	Lys	Leu	Asn	Ser	Phe	Ile	Thr	
				660					665					670		
Gln	Tyr	Ser	Thr	Gly	Gln	Val	Ser	Val	Glu	Ile	Glu	Trp	Glu	Leu	Gln	
				675					680					685		
Lys	Glu	Asn	Ser	Lys	Arg	Trp	Asn	Pro	Glu	Ile	Gln	Tyr	Thr	Ser	Asn	
				690					695					700		
Tyr	Tyr	Lys	Ser	Asn	Asn	Val	Glu	Phe	Ala	Val	Asn	Thr	Glu	Gly	Val	
705					710					715					720	
Tyr	Ser	Glu	Pro	Arg	Pro	Ile	Gly	Thr	Arg	Tyr	Leu	Thr	Arg	Asn	Leu	
				725					730					735		

<210> 11

<211> 736

<212> PRT

<213> 人工序列

<220>

<223> AAV分离株

<400> 11

Met	Ala	Ala	Asp	Gly	Tyr	Leu	Pro	Asp	Trp	Leu	Glu	Asp	Asn	Leu	Ser
1				5					10					15	
Glu	Gly	Ile	Arg	Glu	Trp	Trp	Ala	Leu	Lys	Pro	Gly	Ala	Pro	Gln	Pro
			20					25					30		
Lys	Ala	Asn	Gln	Gln	His	Gln	Asp	Asn	Ala	Arg	Gly	Leu	Val	Leu	Pro
		35					40					45			
Gly	Tyr	Lys	Tyr	Leu	Gly	Pro	Gly	Asn	Gly	Leu	Asp	Lys	Gly	Glu	Pro
	50					55					60				
Val	Asn	Ala	Ala	Asp	Ala	Ala	Ala	Leu	Glu	His	Asp	Arg	Ala	Tyr	Asp
65				70						75				80	
Gln	Gln	Leu	Lys	Ala	Gly	Asp	Asn	Pro	Tyr	Leu	Lys	Tyr	Asn	His	Ala
				85					90					95	
Asp	Ala	Glu	Phe	Gln	Glu	Arg	Leu	Lys	Glu	Asp	Thr	Ser	Phe	Gly	Gly
			100					105					110		
Asn	Leu	Gly	Arg	Ala	Val	Phe	Gln	Ala	Lys	Lys	Arg	Leu	Leu	Glu	Pro
		115					120					125			
Leu	Gly	Leu	Val	Glu	Glu	Ala	Ala	Lys	Thr	Ala	Pro	Gly	Lys	Lys	Arg
	130					135					140				
Pro	Val	Glu	Gln	Ser	Pro	Gln	Glu	Pro	Asp	Ser	Ser	Ala	Gly	Ile	Gly
145				150					155					160	
Lys	Ser	Gly	Ala	Gln	Pro	Ala	Lys	Lys	Arg	Leu	Asn	Phe	Gly	Gln	Thr
			165						170				175		
Gly	Asp	Thr	Glu	Ser	Val	Pro	Asp	Pro	Gln	Pro	Ile	Gly	Glu	Pro	Pro
		180					185					190			
Ala	Ala	Pro	Ser	Gly	Val	Gly	Ser	Leu	Thr	Met	Ala	Ser	Gly	Gly	Gly
		195					200				205				
Ala	Pro	Val	Ala	Asp	Asn	Asn	Glu	Gly	Ala	Asp	Gly	Val	Gly	Ser	Ser
	210					215					220				
Ser	Gly	Asn	Trp	His	Cys	Asp	Ser	Gln	Trp	Leu	Gly	Asp	Arg	Val	Ile
225				230					235					240	
Thr	Thr	Ser	Thr	Arg	Thr	Trp	Ala	Leu	Pro	Thr	Tyr	Asn	Asn	His	Leu
			245						250				255		
Tyr	Lys	Gln	Ile	Ser	Asn	Ser	Thr	Ser	Gly	Gly	Ser	Ser	Asn	Asp	Asn
		260					265						270		
Ala	Tyr	Phe	Gly	Tyr	Ser	Thr	Pro	Trp	Gly	Tyr	Phe	Asp	Phe	Asn	Arg
	275						280				285				
Phe	His	Cys	His	Phe	Ser	Pro	Arg	Asp	Trp	Gln	Arg	Leu	Ile	Asn	Asn
	290					295					300				
Asn	Trp	Gly	Phe	Arg	Pro	Lys	Arg	Leu	Asn	Phe	Lys	Leu	Phe	Asn	Ile

305	310	315	320
Gln Val Lys Glu Val Thr Asp Asn Asn Gly Val Lys Thr Ile Ala Asn			
	325	330	335
Asn Leu Thr Ser Thr Val Gln Val Phe Thr Asp Ser Asp Tyr Gln Leu			
	340	345	350
Pro Tyr Val Leu Gly Ser Ala His Glu Gly Cys Leu Pro Pro Phe Pro			
	355	360	365
Ala Asp Val Phe Met Ile Pro Gln Tyr Gly Tyr Leu Thr Leu Asn Asp			
	370	375	380
Gly Ser Gln Ala Val Gly Arg Ser Ser Phe Tyr Cys Leu Glu Tyr Phe			
385	390	395	400
Pro Ser Gln Met Leu Arg Thr Gly Asn Asn Phe Gln Phe Ser Tyr Glu			
	405	410	415
Phe Glu Asn Val Pro Phe His Ser Ser Tyr Ala His Ser Gln Ser Leu			
	420	425	430
Asp Arg Leu Met Asn Pro Leu Ile Asp Gln Tyr Leu Tyr Tyr Leu Ser			
	435	440	445
Lys Thr Ile Asn Gly Ser Gly Gln Asn Gln Gln Thr Leu Lys Phe Ser			
	450	455	460
Val Ala Gly Pro Ser Asn Met Ala Val Gln Gly Arg Asn Tyr Ile Pro			
465	470	475	480
Gly Pro Ser Tyr Arg Gln Gln Arg Val Ser Thr Thr Val Thr Gln Asn			
	485	490	495
Asn Asn Ser Glu Phe Ala Trp Pro Gly Ala Ser Ser Trp Ala Leu Asn			
	500	505	510
Gly Arg Asn Ser Leu Met Asn Pro Gly Pro Ala Met Ala Ser His Lys			
	515	520	525
Glu Gly Glu Asp Arg Phe Phe Pro Leu Ser Gly Ser Leu Ile Phe Gly			
	530	535	540
Lys Gln Gly Thr Gly Arg Asp Asn Val Asp Ala Asp Lys Val Met Ile			
545	550	555	560
Thr Asn Glu Glu Glu Ile Lys Thr Thr Asn Pro Val Ala Thr Glu Ser			
	565	570	575
Tyr Gly Gln Val Ala Thr Asn His Gln Ser Ala Gln Ala Gln Ala Gln			
	580	585	590
Thr Gly Trp Val Gln Asn Gln Gly Ile Leu Pro Gly Met Val Trp Gln			
	595	600	605
Asp Arg Asp Val Tyr Leu Gln Gly Pro Ile Trp Ala Lys Ile Pro His			
610	615	620	

Thr	Asp	Gly	Asn	Phe	His	Pro	Ser	Pro	Leu	Met	Gly	Gly	Phe	Gly	Met
625						630				635					640
Lys	His	Pro	Pro	Pro	Gln	Ile	Leu	Ile	Lys	Asn	Thr	Pro	Val	Pro	Ala
					645					650					655
Asp	Pro	Pro	Thr	Ala	Phe	Asn	Lys	Asp	Lys	Leu	Asn	Ser	Phe	Ile	Thr
			660							665					670
Gln	Tyr	Ser	Thr	Gly	Gln	Val	Ser	Val	Glu	Ile	Glu	Trp	Glu	Leu	Gln
			675							680					685
Lys	Lys	Asn	Ser	Lys	Arg	Trp	Asn	Pro	Glu	Ile	Gln	Tyr	Thr	Ser	Asn
			690												700
Tyr	Tyr	Lys	Ser	Asn	Asn	Val	Glu	Phe	Ala	Val	Asn	Thr	Glu	Gly	Val
705						710									720
Tyr	Ser	Glu	Pro	Arg	Pro	Ile	Gly	Thr	Arg	Tyr	Leu	Thr	Arg	Asn	Leu
						725									735

<210> 12

<211> 736

<212> PRT

<213> 人工序列

<220>

<223> AAV分离株

<400> 12

Met	Ala	Ala	Asp	Gly	Tyr	Leu	Pro	Asp	Trp	Leu	Glu	Asp	Asn	Leu	Ser
1				5					10					15	
Glu	Gly	Ile	Arg	Glu	Trp	Trp	Ala	Leu	Lys	Pro	Gly	Ala	Pro	Gln	Pro
			20						25					30	
Lys	Ala	Asn	Gln	Gln	His	Gln	Asp	Asn	Ala	Arg	Gly	Leu	Val	Leu	Pro
			35						40					45	
Gly	Tyr	Lys	Tyr	Leu	Gly	Pro	Gly	Asn	Gly	Leu	Asp	Lys	Gly	Glu	Pro
			50						55					60	
Val	Asn	Ala	Ala	Asp	Ala	Ala	Ala	Leu	Glu	His	Asp	Lys	Ala	Tyr	Asp
65						70									80
Gln	Gln	Leu	Lys	Ala	Gly	Asp	Asn	Pro	Tyr	Leu	Lys	Tyr	Asn	His	Ala
						85									95
Asp	Ala	Glu	Phe	Gln	Glu	Arg	Leu	Lys	Glu	Asp	Thr	Ser	Phe	Gly	Gly
						100									110
Asn	Leu	Gly	Arg	Ala	Val	Phe	Gln	Ala	Lys	Lys	Arg	Leu	Leu	Glu	Pro
						115									125
Leu	Gly	Leu	Val	Glu	Glu	Ala	Ala	Lys	Thr	Ala	Pro	Gly	Lys	Lys	Arg
						130									140

Pro Val Glu Gln Ser Pro Gln Glu Pro Asp Ser Ser Ala Gly Ile Gly			
145	150	155	160
Lys Ser Gly Ala Gln Pro Ala Lys Lys Arg Leu Asn Phe Gly Gln Thr			
	165	170	175
Gly Asp Thr Glu Ser Val Pro Asp Pro Gln Pro Ile Gly Glu Pro Pro			
	180	185	190
Ala Ala Pro Ser Gly Val Gly Ser Leu Thr Met Ala Ser Gly Gly Gly			
	195	200	205
Ala Pro Val Ala Asp Asn Asn Glu Gly Ala Asp Gly Val Gly Ser Ser			
	210	215	220
Ser Gly Asn Trp His Cys Asp Ser Gln Trp Leu Gly Asp Arg Val Ile			
225	230	235	240
Thr Thr Ser Thr Arg Thr Trp Ala Leu Pro Thr Tyr Asn Asn His Leu			
	245	250	255
Tyr Lys Gln Ile Ser Asn Ser Thr Ser Gly Gly Ser Ser Asn Asp Asn			
	260	265	270
Ala Tyr Phe Gly Tyr Ser Thr Pro Trp Gly Tyr Phe Asp Phe Asn Arg			
	275	280	285
Phe His Cys His Phe Ser Pro His Asp Trp Gln Arg Leu Ile Asn Asn			
	290	295	300
Asn Trp Gly Phe Arg Pro Lys Arg Leu Asn Phe Lys Leu Phe Asn Ile			
305	310	315	320
Gln Val Lys Glu Val Thr Asp Asn Asn Gly Val Lys Thr Ile Ala Asn			
	325	330	335
Asn Leu Thr Ser Thr Val Gln Val Phe Thr Asp Ser Asp Tyr Gln Leu			
	340	345	350
Pro Tyr Val Leu Gly Ser Ala His Glu Gly Cys Leu Pro Pro Phe Pro			
	355	360	365
Ala Asp Val Phe Met Ile Pro Gln Tyr Gly Tyr Leu Thr Leu Asn Asp			
	370	375	380
Gly Ser Gln Ala Val Gly Arg Ser Ser Phe Tyr Cys Leu Glu Tyr Phe			
385	390	395	400
Pro Ser Gln Met Leu Arg Thr Gly Asn Asn Phe Gln Phe Ser Tyr Glu			
	405	410	415
Phe Glu Asn Val Pro Phe His Ser Ser Tyr Ala His Ser Gln Ser Leu			
	420	425	430
Asp Arg Leu Met Asn Pro Leu Ile Asp Gln Tyr Leu Tyr Tyr Leu Ser			
	435	440	445
Lys Thr Ile Asn Gly Ser Gly Gln Asn Gln Gln Thr Leu Lys Phe Asn			

450	455	460
Val Ala Gly Pro Ser Asn Met Ala Val Gln Gly Arg Asn Tyr Ile Pro		
465	470	475
Gly Pro Ser Tyr Arg Gln Gln Arg Val Ser Thr Thr Val Thr Gln Asn		
485	490	495
Asn Asn Ser Glu Phe Ala Trp Pro Arg Ala Ser Ser Trp Ala Leu Asn		
500	505	510
Gly Arg Asn Ser Leu Met Asn Pro Gly Pro Ala Met Ala Ser His Lys		
515	520	525
Glu Gly Glu Asp Arg Phe Phe Pro Leu Ser Gly Ser Leu Ile Phe Gly		
530	535	540
Lys Gln Gly Thr Gly Arg Asp Asn Val Asp Ala Asp Lys Val Met Ile		
545	550	555
Thr Asn Glu Glu Glu Ile Lys Thr Thr Asn Pro Val Ala Thr Glu Ser		
565	570	575
Tyr Gly Gln Val Ala Thr Asn His Gln Ser Ala Gln Ala Gln Ala Gln		
580	585	590
Thr Gly Trp Val Gln Asn Gln Gly Ile Leu Pro Gly Met Val Trp Gln		
595	600	605
Asp Arg Asp Val Tyr Leu Gln Gly Pro Ile Trp Ala Lys Ile Pro His		
610	615	620
Thr Asp Gly Asn Phe His Pro Ser Pro Leu Met Gly Gly Phe Gly Met		
625	630	635
Lys His Pro Pro Pro Gln Ile Leu Ile Lys Asn Thr Pro Val Pro Ala		
645	650	655
Asp Pro Pro Thr Ala Phe Asn Lys Asp Lys Leu Asn Ser Phe Ile Thr		
660	665	670
Gln Tyr Ser Thr Gly Gln Val Ser Met Glu Ile Glu Trp Glu Leu Gln		
675	680	685
Lys Glu Asn Ser Lys Arg Trp Asn Pro Glu Ile Gln Tyr Thr Ser Asn		
690	695	700
Tyr Tyr Lys Ser Asn Asn Val Glu Phe Ala Val Asn Thr Glu Gly Val		
705	710	715
Tyr Ser Glu Pro Arg Pro Ile Gly Thr Arg Tyr Leu Thr Arg Asn Leu		
725	730	735

<210> 13

<211> 736

<212> PRT

<213> 人工序列

<220>

<223> AAV分离株

<400> 13

```

Met Ala Ala Asp Gly Tyr Leu Pro Asp Trp Leu Glu Asp Asn Leu Ser
1           5           10           15
Glu Gly Ile Arg Glu Trp Trp Ala Leu Lys Pro Gly Ala Pro Gln Pro
           20           25           30
Lys Ala Asn Gln Gln His Gln Asp Asn Ala Arg Gly Leu Val Leu Pro
           35           40           45
Gly Tyr Lys Tyr Leu Gly Pro Gly Asn Gly Leu Asp Lys Gly Glu Pro
           50           55           60
Val Asn Ala Ala Asp Ala Ala Ala Leu Glu His Asp Lys Ala Tyr Asp
65           70           75           80
Gln Gln Leu Lys Ala Gly Asp Asn Pro Tyr Leu Lys Tyr Asn His Ala
           85           90           95
Asp Ala Glu Phe Gln Glu Arg Leu Lys Glu Asp Thr Ser Phe Gly Gly
           100          105          110
Asn Leu Gly Arg Ala Val Phe Gln Ala Lys Lys Arg Leu Leu Glu Pro
           115          120          125
Leu Gly Leu Val Glu Glu Ala Ala Lys Thr Ala Pro Gly Lys Lys Arg
           130          135          140
Pro Val Glu Gln Ser Pro Gln Glu Pro Asp Ser Ser Ala Gly Ile Gly
145          150          155          160
Lys Ser Gly Ala Gln Pro Ala Lys Lys Arg Leu Asn Phe Gly Gln Thr
           165          170          175
Gly Asp Thr Glu Ser Val Pro Asp Pro Gln Pro Ile Gly Glu Pro Pro
           180          185          190
Ala Ala Pro Ser Gly Val Gly Ser Leu Thr Met Ala Ser Gly Gly Gly
           195          200          205
Ala Pro Val Ala Asp Asn Asn Glu Gly Ala Asp Gly Val Gly Ser Ser
           210          215          220
Ser Gly Asn Trp His Cys Asp Ser Gln Trp Leu Gly Asp Arg Val Ile
225          230          235          240
Thr Thr Ser Thr Arg Thr Trp Ala Leu Pro Thr Tyr Asn Asn His Leu
           245          250          255
Tyr Lys Gln Ile Ser Asn Ser Thr Ser Gly Gly Ser Ser Asn Asp Asn
           260          265          270
Ala Tyr Phe Gly Tyr Ser Thr Pro Trp Gly Tyr Phe Asp Phe Asn Arg
           275          280          285

```

Phe	His	Cys	His	Phe	Ser	Pro	Arg	Asp	Trp	Gln	Arg	Leu	Ile	Asn	Asn
290						295					300				
Asn	Trp	Gly	Phe	Arg	Pro	Lys	Arg	Leu	Asn	Phe	Lys	Leu	Phe	Asn	Ile
305					310					315					320
Gln	Val	Lys	Glu	Val	Thr	Asp	Asn	Asn	Gly	Val	Lys	Thr	Ile	Ala	Asn
				325					330					335	
Asn	Leu	Thr	Ser	Thr	Val	Gln	Val	Phe	Thr	Asp	Ser	Asp	Tyr	Gln	Leu
			340					345					350		
Pro	Tyr	Val	Leu	Gly	Ser	Ala	His	Glu	Gly	Cys	Leu	Pro	Pro	Phe	Pro
	355					360						365			
Ala	Asp	Val	Phe	Met	Ile	Pro	Gln	Tyr	Gly	Tyr	Leu	Thr	Leu	Asn	Asp
370						375					380				
Gly	Ser	Gln	Ala	Val	Gly	Arg	Ser	Ser	Phe	Tyr	Cys	Leu	Glu	Tyr	Phe
385					390					395					400
Pro	Ser	Gln	Met	Leu	Arg	Thr	Gly	Asn	Asn	Phe	Gln	Phe	Ser	Tyr	Glu
			405						410					415	
Phe	Glu	Asn	Val	Pro	Phe	His	Ser	Ser	Tyr	Ala	His	Ser	Gln	Ser	Leu
			420						425				430		
Asp	Arg	Leu	Met	Asn	Pro	Leu	Ile	Asp	Gln	Tyr	Leu	Tyr	Tyr	Leu	Ser
	435					440						445			
Lys	Thr	Ile	Asn	Gly	Ser	Gly	Gln	Asn	Gln	Gln	Thr	Leu	Lys	Phe	Ser
450						455						460			
Val	Ala	Gly	Pro	Ser	Asn	Met	Ala	Val	Gln	Gly	Arg	Asn	Tyr	Ile	Pro
465					470					475					480
Gly	Pro	Ser	Tyr	Arg	Gln	Gln	Arg	Val	Ser	Thr	Thr	Val	Thr	Gln	Asn
				485					490					495	
Asn	Asn	Ser	Glu	Phe	Ala	Trp	Pro	Arg	Ala	Ser	Ser	Trp	Ala	Leu	Asn
			500					505					510		
Gly	Arg	Asn	Ser	Leu	Met	Asn	Pro	Gly	Pro	Ala	Met	Ala	Ser	His	Lys
	515					520						525			
Glu	Gly	Glu	Asp	Arg	Phe	Phe	Pro	Leu	Ser	Gly	Ser	Leu	Ile	Phe	Gly
530						535						540			
Lys	Gln	Gly	Thr	Gly	Arg	Asp	Asn	Val	Asp	Ala	Asp	Lys	Val	Met	Ile
545					550					555					560
Thr	Asn	Glu	Glu	Glu	Ile	Lys	Thr	Thr	Asn	Pro	Val	Ala	Thr	Glu	Ser
				565					570					575	
Tyr	Gly	Gln	Val	Ala	Thr	Asn	His	Gln	Ser	Ala	Gln	Ala	Gln	Ala	Gln
			580					585					590		
Thr	Gly	Trp	Val	Gln	Asn	Gln	Gly	Ile	Leu	Pro	Gly	Met	Val	Trp	Gln

595	600	605
Asp Arg Asp Val Tyr Leu Gln Gly Pro Ile Trp	Ala Lys Ile Pro His	
610	615	620
Thr Asp Gly Asn Phe His Pro Ser Pro Leu Met	Gly Gly Phe Gly Met	
625	630	635
Lys His Pro Pro Pro Gln Ile Leu Ile Lys Asn Thr	Pro Val Pro Ala	
645	650	655
Asp Pro Pro Thr Ala Phe Asn Lys Asp Lys Leu Asn Ser	Phe Ile Thr	
660	665	670
Gln Tyr Ser Thr Gly Gln Val Ser Val Glu Ile Glu Trp	Glu Leu Gln	
675	680	685
Lys Glu Asn Ser Lys Arg Trp Asn Pro Glu Ile Gln Tyr Thr	Ser Asn	
690	695	700
Tyr Tyr Lys Ser Asn Asn Val Glu Phe Ala Val Asn Thr Glu Gly	Val	
705	710	715
Tyr Ser Glu Pro Arg Pro Ile Gly Thr Arg Tyr Leu Thr Arg Asn Leu		
725	730	735

<210> 14

<211> 1527

<212> DNA

<213> 人工序列

<220>

<223> 合成多核苷酸

<400> 14

```

atgtctatgg gggctcctcg ctccctgctg ctggcactgg ccgccgggct ggctgtcgca 60
agaccaccta atatcgtcct gattttttgca gacgatctgg gatacggcga cctgggatgc 120
tatggccacc caagctccac cacacccaac ctggaccagc tggcagcagg aggcctgcgg 180
ttcaccgact tctacgtgcc agtgagcctg tgcacccctt ccagagccgc cctgctgaca 240
ggcaggctgc cagtgcgcac gggcatgtat cctggcgctg tggtgccatc tagcaggggc 300
ggcctgccac tggaggaggt gaccgtggca gaggtgctgg cagccagagg ctacctgaca 360
ggaatggccg gcaagtggca cctggggagt ggaccagagg gaggccttctt gccccctcac 420
cagggttcc accggtttct gggcatcctt tattctcag accagggcc atgccagaac 480
ctgacctgtt ttccaccagc aacaccatgc gacggaggat gtgatcaggg cctggtgcca 540
atcccactgc tggcaaactt gagcgtggag gcacagcctc catggctgcc tggcctggag 600
gcaagataca tggccttcgc ccacgacctg atggcagatg cacagcggca ggatagacct 660
ttctttctgt actatgcctc ccaccacacc cactatccac agttcagcgg ccagtccttt 720
gccgagaggt ccggaagggg accattcggc gactctctga tggagctgga tgccgccgtg 780
ggcaccctga tgacagcaat cggcgacctg ggctgctgg aggagacact ggtcatcttc 840
accgccgata acggccctga gacaatgcgg atgtctagag gcggatgcag cggcctgctg 900

```

agatgtggca agggaaccac atacgagga ggcgtgcgcg agcctgccct ggcatttttg 960
 ccaggacaca tcgcacctgg agtgaccac gagctggcct cctctctgga cctgctgcca 1020
 aacttgcccg ccctggcagg agcacctctg ccaaattgta ccctggacgg ctctgatctg 1080
 agcccactgc tgctgggaac cggcaagtcc ctaggcagt ctctgttctt ttaccctcc 1140
 tatcctgatg aggtgcgggg cgtgtttgcc gtgagaaccg gcaagtacaa ggcccacttc 1200
 ttacacagg gctctgcccc cagcgacacc acagcagatc cagcatgcca cgccagctcc 1260
 tctctgaccg cacacgagcc acctctgctg tacgacctgt ccaaggatcc cggcgagaac 1320
 tataatctgc tgggaggagt ggcaggagca acccctgagg tgctgcaggc cctgaagcag 1380
 ctgcagctgc tgaaggcaca gctggacgca gcagtacat tcggcccaag ccaggtggcc 1440
 agaggcgagg atcccgccct gcagatctgt tgccaccccg gctgcacccc aagacctgcc 1500
 tgttgccatt gccccgaccc acacgcc 1527

<210> 15

<211> 736

<212> PRT

<213> 人工序列

<220>

<223> AAV分离株

<400> 15

Met	Ala	Ala	Asp	Gly	Tyr	Leu	Pro	Asp	Trp	Leu	Glu	Asp	Asn	Leu	Ser
1				5					10					15	
Glu	Gly	Ile	Arg	Glu	Trp	Trp	Ala	Leu	Lys	Pro	Gly	Ala	Pro	Gln	Pro
			20					25					30		
Lys	Ala	Asn	Gln	Gln	His	Gln	Asp	Asn	Ala	Arg	Gly	Leu	Val	Leu	Pro
		35				40					45				
Gly	Tyr	Lys	Tyr	Leu	Gly	Pro	Gly	Asn	Gly	Leu	Asp	Lys	Gly	Glu	Pro
	50				55					60					
Val	Asn	Ala	Ala	Asp	Ala	Ala	Ala	Leu	Glu	His	Asp	Lys	Ala	Tyr	Asp
65				70					75					80	
Gln	Gln	Leu	Lys	Ala	Gly	Asp	Asn	Pro	Tyr	Leu	Lys	Tyr	Asn	His	Ala
			85					90					95		
Asp	Ala	Glu	Phe	Gln	Glu	Arg	Leu	Lys	Glu	Asp	Thr	Ser	Phe	Gly	Gly
		100						105					110		
Asn	Leu	Gly	Arg	Ala	Val	Phe	Gln	Ala	Lys	Lys	Arg	Leu	Leu	Glu	Pro
		115					120					125			
Leu	Gly	Leu	Val	Glu	Glu	Ala	Ala	Lys	Thr	Ala	Pro	Gly	Lys	Lys	Arg
	130					135					140				
Pro	Val	Glu	Gln	Ser	Pro	Gln	Glu	Pro	Asp	Ser	Ser	Ala	Gly	Ile	Gly
145				150					155					160	
Lys	Ser	Gly	Ala	Gln	Pro	Ala	Lys	Lys	Arg	Leu	Asn	Phe	Gly	Gln	Thr

				165					170					175			
Gly	Asp	Thr	Glu	Ser	Val	Pro	Asp	Pro	Gln	Pro	Ile	Gly	Glu	Pro	Pro		
				180					185					190			
Ala	Ala	Pro	Ser	Gly	Val	Gly	Ser	Leu	Thr	Met	Ala	Ser	Gly	Gly	Gly		
				195				200						205			
Ala	Pro	Val	Ala	Asp	Asn	Asn	Glu	Gly	Ala	Asp	Gly	Val	Gly	Ser	Ser		
				210				215						220			
Ser	Gly	Asn	Trp	His	Cys	Asp	Ser	Gln	Trp	Leu	Gly	Asp	Arg	Val	Ile		
225					230				235					240			
Thr	Thr	Ser	Thr	Arg	Thr	Trp	Ala	Leu	Pro	Thr	Tyr	Asn	Asn	His	Leu		
				245					250					255			
Tyr	Lys	Gln	Ile	Ser	Asn	Ser	Thr	Ser	Gly	Gly	Ser	Ser	Asn	Asp	Asn		
				260					265					270			
Ala	Tyr	Phe	Gly	Tyr	Ser	Thr	Pro	Trp	Gly	Tyr	Phe	Asp	Phe	Asn	Arg		
				275					280					285			
Phe	His	Cys	His	Phe	Ser	Pro	Arg	Asp	Trp	Gln	Arg	Leu	Ile	Asn	Asn		
				290					295					300			
Asn	Trp	Gly	Phe	Arg	Pro	Lys	Arg	Leu	Asn	Phe	Lys	Leu	Phe	Asn	Ile		
305					310					315				320			
Gln	Val	Lys	Glu	Val	Thr	Asp	Asn	Asn	Gly	Val	Lys	Thr	Ile	Ala	Asn		
				325					330					335			
Asn	Leu	Thr	Ser	Thr	Val	Gln	Val	Phe	Thr	Asp	Ser	Asp	Tyr	Gln	Leu		
				340					345					350			
Pro	Tyr	Val	Leu	Gly	Ser	Ala	His	Glu	Gly	Cys	Leu	Pro	Pro	Phe	Pro		
				355					360					365			
Ala	Asp	Val	Phe	Met	Ile	Pro	Gln	Tyr	Gly	Tyr	Leu	Thr	Leu	Asn	Asp		
				370					375					380			
Gly	Ser	Gln	Ala	Val	Gly	Arg	Ser	Ser	Phe	Tyr	Cys	Leu	Glu	Tyr	Phe		
385					390					395				400			
Pro	Ser	Gln	Met	Leu	Arg	Thr	Gly	Asn	Asn	Phe	Gln	Phe	Ser	Tyr	Glu		
				405					410					415			
Phe	Glu	Asn	Val	Pro	Phe	His	Ser	Ser	Tyr	Ala	His	Ser	Gln	Ser	Leu		
				420					425					430			
Asp	Arg	Leu	Met	Asn	Pro	Leu	Ile	Asp	Gln	Tyr	Leu	Tyr	Tyr	Leu	Ser		
				435					440					445			
Lys	Thr	Ile	Asn	Gly	Ser	Gly	Gln	Asn	Gln	Gln	Thr	Leu	Lys	Phe	Ser		
				450					455					460			
Val	Ala	Gly	Pro	Ser	Asn	Met	Ala	Val	Gln	Gly	Arg	Asn	Tyr	Ile	Pro		
465					470					475				480			

Gly	Pro	Ser	Tyr	Arg	Gln	Gln	Arg	Val	Ser	Thr	Thr	Val	Thr	Gln	Asn	
				485					490					495		
Asn	Asn	Ser	Glu	Phe	Ala	Trp	Pro	Arg	Ala	Ser	Ser	Trp	Ala	Leu	Asn	
				500					505					510		
Gly	Arg	Asn	Ser	Leu	Met	Asn	Pro	Gly	Pro	Ala	Met	Ala	Ser	His	Lys	
				515					520					525		
Glu	Gly	Glu	Asp	Arg	Phe	Phe	Pro	Leu	Ser	Gly	Ser	Leu	Ile	Phe	Gly	
				530					535					540		
Lys	Gln	Gly	Thr	Gly	Arg	Asp	Asn	Val	Asp	Ala	Asp	Lys	Val	Met	Ile	
545					550					555					560	
Thr	Asn	Glu	Glu	Glu	Ile	Lys	Thr	Thr	Asn	Pro	Val	Ala	Thr	Glu	Ser	
				565					570					575		
Tyr	Gly	Gln	Val	Ala	Thr	Asn	His	Gln	Ser	Ala	Gln	Ala	Gln	Ala	Gln	
				580					585					590		
Thr	Gly	Trp	Val	Gln	Asn	Gln	Gly	Ile	Leu	Pro	Gly	Met	Val	Trp	Gln	
				595					600					605		
Asp	Arg	Asp	Val	Tyr	Leu	Gln	Gly	Pro	Ile	Trp	Ala	Lys	Ile	Pro	His	
				610					615					620		
Thr	Asp	Gly	Asn	Phe	His	Pro	Ser	Pro	Leu	Met	Gly	Gly	Phe	Gly	Met	
625					630					635					640	
Lys	His	Pro	Pro	Pro	Gln	Ile	Leu	Ile	Lys	Asn	Thr	Pro	Val	Pro	Ala	
				645					650					655		
Asp	Pro	Pro	Thr	Ala	Phe	Asn	Lys	Asp	Lys	Leu	Asn	Ser	Phe	Ile	Thr	
				660					665					670		
Gln	Tyr	Ser	Thr	Gly	Gln	Val	Ser	Val	Glu	Ile	Glu	Trp	Glu	Arg	Gln	
				675					680					685		
Lys	Glu	Asn	Ser	Lys	Arg	Trp	Asn	Pro	Glu	Ile	Gln	Tyr	Thr	Ser	Asn	
				690					695					700		
Tyr	Tyr	Lys	Ser	Asn	Asn	Val	Glu	Phe	Ala	Val	Asn	Thr	Glu	Gly	Val	
705					710					715					720	
Tyr	Ser	Glu	Pro	Arg	Pro	Ile	Gly	Thr	Arg	Tyr	Leu	Thr	Arg	Asn	Leu	
				725					730					735		

<210> 16

<211> 736

<212> PRT

<213> 人工序列

<220>

<223> AAV分离株

<400> 16

Met	Ala	Ala	Asp	Gly	Tyr	Leu	Pro	Asp	Trp	Leu	Glu	Asp	Asn	Leu	Ser
1				5					10					15	
Glu	Gly	Ile	Arg	Glu	Trp	Trp	Ala	Leu	Lys	Pro	Gly	Ala	Pro	Gln	Pro
			20					25					30		
Lys	Ala	Asn	Gln	Gln	His	Gln	Asp	Asn	Ala	Arg	Gly	Leu	Val	Leu	Pro
		35					40					45			
Gly	Tyr	Lys	Tyr	Leu	Gly	Pro	Gly	Asn	Gly	Leu	Asp	Lys	Gly	Glu	Pro
	50					55					60				
Val	Asn	Ala	Ala	Asp	Ala	Ala	Ala	Leu	Glu	His	Asp	Lys	Ala	Tyr	Asp
65				70						75				80	
Gln	Gln	Leu	Lys	Ala	Gly	Asp	Asn	Pro	Tyr	Leu	Lys	Tyr	Asn	His	Ala
				85					90					95	
Asp	Ala	Glu	Phe	Gln	Glu	Arg	Leu	Lys	Glu	Asp	Thr	Ser	Phe	Gly	Gly
			100					105					110		
Asn	Leu	Gly	Arg	Ala	Val	Phe	Gln	Ala	Lys	Lys	Arg	Leu	Leu	Glu	Pro
		115					120					125			
Leu	Gly	Leu	Val	Glu	Glu	Ala	Ala	Lys	Thr	Ala	Pro	Gly	Lys	Lys	Arg
	130					135					140				
Pro	Val	Glu	Gln	Ser	Pro	Gln	Glu	Pro	Asp	Ser	Ser	Ala	Gly	Ile	Gly
145				150					155					160	
Lys	Ser	Gly	Ala	Gln	Pro	Ala	Lys	Lys	Arg	Leu	Asn	Phe	Gly	Gln	Thr
			165						170				175		
Gly	Asp	Thr	Glu	Ser	Val	Pro	Asp	Pro	Gln	Pro	Ile	Gly	Glu	Pro	Pro
		180					185					190			
Ala	Ala	Pro	Ser	Gly	Val	Gly	Ser	Leu	Thr	Met	Ala	Ser	Gly	Gly	Gly
		195					200					205			
Ala	Pro	Val	Ala	Asp	Asn	Asn	Glu	Gly	Ala	Asp	Gly	Val	Gly	Ser	Ser
	210					215					220				
Ser	Gly	Asn	Trp	His	Cys	Asp	Ser	Gln	Trp	Leu	Gly	Asp	Arg	Val	Ile
225				230					235					240	
Thr	Thr	Ser	Thr	Arg	Thr	Trp	Ala	Leu	Pro	Thr	Tyr	Asn	Asn	His	Leu
			245						250				255		
Tyr	Lys	Gln	Ile	Ser	Asn	Ser	Thr	Ser	Gly	Gly	Ser	Ser	Asn	Asp	Asn
		260					265						270		
Ala	Tyr	Phe	Gly	Tyr	Ser	Thr	Pro	Trp	Gly	Tyr	Phe	Asp	Phe	Asn	Arg
	275						280					285			
Phe	His	Cys	His	Phe	Ser	Pro	Arg	Asp	Trp	Gln	Arg	Leu	Ile	Asn	Asn
	290					295					300				
Asn	Trp	Gly	Phe	Arg	Pro	Lys	Arg	Leu	Asn	Phe	Lys	Leu	Phe	Asn	Ile

305	310	315	320
Gln Val Lys Glu Val Thr Asp Asn Asn Gly Val Lys Thr Ile Ala Asn			
	325	330	335
Asn Leu Thr Ser Thr Val Gln Val Phe Ala Asp Ser Asp Tyr Gln Leu			
	340	345	350
Pro Tyr Val Leu Gly Ser Ala His Glu Gly Cys Leu Pro Pro Phe Pro			
	355	360	365
Ala Asp Val Phe Met Ile Pro Gln Tyr Gly Tyr Leu Thr Leu Asn Asp			
	370	375	380
Gly Ser Gln Ala Val Gly Arg Ser Ser Phe Tyr Cys Leu Glu Tyr Phe			
385	390	395	400
Pro Ser Gln Met Leu Arg Thr Gly Asn Asn Phe Gln Phe Ser Tyr Glu			
	405	410	415
Phe Glu Asn Val Pro Phe His Ser Ser Tyr Ala His Ser Gln Ser Leu			
	420	425	430
Asp Arg Leu Met Asn Pro Leu Ile Asp Gln Tyr Leu Tyr Tyr Leu Ser			
	435	440	445
Lys Thr Ile Asn Gly Ser Gly Gln Asn Gln Gln Thr Leu Lys Phe Ser			
	450	455	460
Val Ala Gly Pro Ser Asn Met Ala Val Gln Gly Arg Asn Tyr Ile Pro			
465	470	475	480
Gly Pro Ser Tyr Arg Gln Gln Arg Val Ser Thr Thr Val Thr Gln Asn			
	485	490	495
Asn Asn Ser Glu Phe Ala Trp Pro Arg Ala Ser Ser Trp Ala Leu Asn			
	500	505	510
Gly Arg Asn Ser Leu Met Asn Pro Gly Pro Ala Met Ala Ser His Lys			
	515	520	525
Glu Gly Glu Asp Arg Phe Phe Pro Leu Ser Gly Ser Leu Ile Phe Gly			
	530	535	540
Lys Gln Gly Thr Gly Arg Asp Asn Val Asp Ala Asp Lys Val Met Ile			
545	550	555	560
Thr Asn Glu Glu Glu Ile Lys Thr Thr Asn Pro Val Ala Thr Glu Ser			
	565	570	575
Tyr Gly Gln Val Ala Thr Asn His Gln Ser Ala Gln Ala Gln Ala Gln			
	580	585	590
Thr Gly Trp Val Gln Asn Gln Gly Ile Leu Pro Gly Met Val Trp Gln			
	595	600	605
Asp Arg Asp Val Tyr Leu Gln Gly Pro Ile Trp Ala Lys Ile Pro His			
610	615	620	

Thr	Asp	Gly	Asn	Phe	His	Pro	Ser	Pro	Leu	Met	Gly	Gly	Phe	Gly	Met
625						630				635					640
Lys	His	Pro	Pro	Pro	Gln	Ile	Leu	Ile	Lys	Asn	Thr	Pro	Val	Pro	Ala
					645					650					655
Asp	Pro	Pro	Thr	Ala	Phe	Asn	Lys	Asp	Lys	Leu	Asn	Ser	Phe	Ile	Thr
			660						665				670		
Gln	Tyr	Ser	Thr	Gly	Gln	Val	Ser	Val	Glu	Ile	Glu	Trp	Glu	Leu	Gln
			675					680					685		
Lys	Glu	Asn	Ser	Lys	Arg	Trp	Asn	Pro	Glu	Ile	Gln	Tyr	Thr	Ser	Asn
			690				695						700		
Tyr	Tyr	Lys	Ser	Asn	Asn	Val	Glu	Phe	Ala	Val	Asn	Thr	Glu	Gly	Val
705						710					715				720
Tyr	Ser	Glu	Pro	Arg	Pro	Ile	Gly	Thr	Arg	Tyr	Leu	Thr	Arg	Asn	Leu
				725					730						735

<210> 17

<211> 736

<212> PRT

<213> 人工序列

<220>

<223> AAV分离株

<400> 17

Met	Ala	Ala	Asp	Gly	Tyr	Leu	Pro	Asp	Trp	Leu	Glu	Asp	Asn	Leu	Ser
1				5					10					15	
Glu	Gly	Ile	Arg	Glu	Trp	Trp	Ala	Leu	Lys	Pro	Gly	Ala	Pro	Gln	Pro
			20					25					30		
Lys	Ala	Asn	Gln	Gln	His	Gln	Asp	Asn	Ala	Arg	Gly	Leu	Val	Leu	Pro
			35				40					45			
Gly	Tyr	Lys	Tyr	Leu	Gly	Pro	Gly	Asn	Gly	Leu	Asp	Lys	Gly	Glu	Pro
			50				55				60				
Val	Asn	Ala	Ala	Asp	Ala	Ala	Ala	Leu	Glu	His	Asp	Lys	Ala	Tyr	Asp
65					70					75					80
Gln	Gln	Leu	Lys	Ala	Gly	Asp	Asn	Pro	Tyr	Leu	Lys	Tyr	Asn	His	Ala
				85					90					95	
Asp	Ala	Glu	Phe	Gln	Glu	Arg	Leu	Lys	Glu	Asp	Thr	Ser	Phe	Gly	Gly
			100						105				110		
Asn	Leu	Gly	Arg	Ala	Val	Phe	Gln	Ala	Lys	Lys	Arg	Leu	Leu	Glu	Pro
			115					120					125		
Leu	Gly	Leu	Val	Glu	Glu	Ala	Ala	Lys	Thr	Ala	Pro	Gly	Lys	Lys	Arg
			130				135					140			

Pro Val Glu Gln Ser Pro Gln Glu Pro Asp Ser Ser Ala Gly Ile Gly			
145	150	155	160
Lys Ser Gly Ala Gln Pro Ala Lys Lys Arg Leu Asn Phe Gly Gln Thr			
	165	170	175
Gly Asp Thr Glu Ser Val Pro Asp Pro Gln Pro Ile Gly Glu Pro Pro			
	180	185	190
Ala Ala Pro Ser Gly Val Gly Ser Leu Thr Met Ala Ser Gly Gly Gly			
	195	200	205
Ala Pro Val Ala Asp Asn Asn Glu Gly Ala Asp Gly Val Gly Ser Ser			
	210	215	220
Ser Gly Asn Trp His Cys Asp Ser Gln Trp Leu Gly Asp Arg Val Ile			
225	230	235	240
Thr Thr Ser Thr Arg Thr Trp Ala Leu Pro Thr Tyr Asn Asn His Leu			
	245	250	255
Tyr Lys Gln Ile Ser Asn Ser Thr Ser Gly Gly Ser Ser Asn Asp Asn			
	260	265	270
Ala Tyr Phe Gly Tyr Ser Thr Pro Trp Gly Tyr Phe Asp Phe Asn Arg			
	275	280	285
Phe His Cys His Phe Ser Pro Arg Asp Trp Gln Arg Leu Ile Asn Asn			
	290	295	300
Asn Trp Gly Phe Arg Pro Lys Arg Leu Asn Phe Lys Leu Phe Asn Ile			
305	310	315	320
Gln Val Lys Glu Val Thr Asp Asn Asn Gly Val Lys Thr Ile Ala Asn			
	325	330	335
Asn Leu Thr Ser Thr Val Gln Val Phe Thr Asp Ser Asp Tyr Gln Leu			
	340	345	350
Pro Tyr Val Leu Gly Ser Ala His Glu Gly Cys Leu Pro Pro Phe Pro			
	355	360	365
Ala Asp Val Phe Met Ile Pro Gln Tyr Gly Tyr Leu Thr Leu Asn Asp			
	370	375	380
Gly Ser Gln Ala Val Gly Arg Ser Ser Phe Tyr Cys Leu Glu Tyr Phe			
385	390	395	400
Pro Ser Gln Met Leu Arg Thr Gly Asn Asn Phe Gln Phe Ser Tyr Glu			
	405	410	415
Phe Glu Asn Val Pro Phe His Ser Ser Tyr Ala His Ser Gln Ser Leu			
	420	425	430
Asp Arg Leu Met Asn Pro Leu Ile Asp Gln Tyr Leu Tyr Tyr Leu Ser			
	435	440	445
Lys Thr Ile Asn Gly Ser Gly Gln Asn Gln Gln Thr Leu Lys Phe Ser			

450	455	460
Val Ala Gly Pro Ser Asn Met Ala Val Gln Gly Arg Asn Tyr Ile Pro		
465	470	475
Gly Pro Ser Tyr Arg Gln Gln Arg Val Ser Thr Thr Val Thr Gln Asn		
485	490	495
Asn Asn Ser Glu Ile Ala Trp Pro Arg Ala Ser Ser Trp Ala Leu Asn		
500	505	510
Gly Arg Asn Ser Leu Met Asn Pro Gly Pro Ala Met Ala Ser His Lys		
515	520	525
Glu Gly Glu Asp Arg Phe Phe Pro Leu Ser Gly Ser Leu Ile Phe Gly		
530	535	540
Lys Gln Gly Thr Gly Arg Asp Asn Val Asp Ala Asp Lys Val Met Ile		
545	550	555
Thr Asn Glu Glu Glu Ile Lys Thr Thr Asn Pro Val Ala Thr Glu Ser		
565	570	575
Tyr Gly Gln Val Ala Thr Asn His Gln Ser Ala Gln Ala Gln Ala Gln		
580	585	590
Thr Gly Trp Val Gln Asn Gln Gly Ile Leu Pro Gly Met Val Trp Gln		
595	600	605
Asp Arg Asp Val Tyr Leu Gln Gly Pro Ile Trp Ala Lys Ile Pro His		
610	615	620
Thr Asp Gly Asn Phe His Pro Ser Pro Leu Met Gly Gly Phe Gly Met		
625	630	635
Lys His Pro Pro Pro Gln Ile Leu Ile Lys Asn Thr Pro Val Pro Ala		
645	650	655
Asp Pro Pro Thr Ala Phe Asn Lys Asp Lys Leu Asn Ser Phe Ile Thr		
660	665	670
Gln Tyr Ser Thr Gly Gln Val Ser Val Glu Ile Glu Trp Glu Leu Gln		
675	680	685
Lys Glu Asn Ser Lys Arg Trp Asn Pro Glu Ile Gln Tyr Thr Ser Asn		
690	695	700
Tyr Cys Lys Ser Asn Asn Val Glu Phe Ala Val Asn Thr Glu Gly Val		
705	710	715
Tyr Ser Glu Pro Arg Pro Ile Gly Thr Arg Tyr Leu Thr Arg Asn Leu		
725	730	735

<210> 18

<211> 145

<212> DNA

<213> 人工序列

<220>

<223> AAV2 5' ITR

<400> 18

ttggccactc cctctctgcg cgctcgctcg ctactgagg ccgggcgacc aaaggtcgcc 60
cgacgcccgg gctttgcccg ggcggcctca gtgagcgagc gagcgcgag agagggagtg 120
gccaactcca tcactagggg ttcct 145

<210> 19

<211> 145

<212> DNA

<213> 人工序列

<220>

<223> AAV2 3' ITR

<400> 19

aggaacccct agtgatggag ttggccactc cctctctgcg cgctcgctcg ctactgagg 60
ccgggcgacc aaaggtcgcc cgacgcccgg gctttgcccg ggcggcctca gtgagcgagc 120
gagcgcgag agagggagtg gccaa 145

<210> 20

<211> 167

<212> DNA

<213> 人工序列

<220>

<223> AAV5 5' ITR

<400> 20

ctctcccccc tgtcgcttc gctcgctcg tggctcgttt gggggggtgg cagctcaaag 60
agctgccaga cgacggccct ctggccgtcg ccccccaaa cgagccagcg agcgagcgaa 120
cgcgacaggg gggagagtgc cacactctca agcaaggggg ttttgta 167

<210> 21

<211> 167

<212> DNA

<213> 人工序列

<220>

<223> AAV5 3' ITR

<400> 21

tacaaaacct ctttgcttga gagtgtggca ctctcccccc tgtcgcttc gctcgctcg 60
tggctcgttt gggggggtgg cagctcaaag agctgccaga cgacggccct ctggccgtcg 120
ccccccaaa cgagccagcg agcgagcgaa cgcgacaggg gggagag 167

<210> 22

<211> 621

<212> PRT

<213> 人工序列

<220>

<223> AAV2 Rep

<400> 22

```

Met Pro Gly Phe Tyr Glu Ile Val Ile Lys Val Pro Ser Asp Leu Asp
1           5           10           15
Glu His Leu Pro Gly Ile Ser Asp Ser Phe Val Asn Trp Val Ala Glu
          20           25           30
Lys Glu Trp Glu Leu Pro Pro Asp Ser Asp Met Asp Leu Asn Leu Ile
          35           40           45
Glu Gln Ala Pro Leu Thr Val Ala Glu Lys Leu Gln Arg Asp Phe Leu
          50           55           60
Thr Glu Trp Arg Arg Val Ser Lys Ala Pro Glu Ala Leu Phe Phe Val
65           70           75           80
Gln Phe Glu Lys Gly Glu Ser Tyr Phe His Met His Val Leu Val Glu
          85           90           95
Thr Thr Gly Val Lys Ser Met Val Leu Gly Arg Phe Leu Ser Gln Ile
          100          105          110
Arg Glu Lys Leu Ile Gln Arg Ile Tyr Arg Gly Ile Glu Pro Thr Leu
          115          120          125
Pro Asn Trp Phe Ala Val Thr Lys Thr Arg Asn Gly Ala Gly Gly Gly
          130          135          140
Asn Lys Val Val Asp Glu Cys Tyr Ile Pro Asn Tyr Leu Leu Pro Lys
145          150          155          160
Thr Gln Pro Glu Leu Gln Trp Ala Trp Thr Asn Met Glu Gln Tyr Leu
          165          170          175
Ser Ala Cys Leu Asn Leu Thr Glu Arg Lys Arg Leu Val Ala Gln His
          180          185          190
Leu Thr His Val Ser Gln Thr Gln Glu Gln Asn Lys Glu Asn Gln Asn
          195          200          205
Pro Asn Ser Asp Ala Pro Val Ile Arg Ser Lys Thr Ser Ala Arg Tyr
          210          215          220
Met Glu Leu Val Gly Trp Leu Val Asp Lys Gly Ile Thr Ser Glu Lys
225          230          235          240
Gln Trp Ile Gln Glu Asp Gln Ala Ser Tyr Ile Ser Phe Asn Ala Ala
          245          250          255
Ser Asn Ser Arg Ser Gln Ile Lys Ala Ala Leu Asp Asn Ala Gly Lys
          260          265          270
Ile Met Ser Leu Thr Lys Thr Ala Pro Asp Tyr Leu Val Gly Gln Gln

```

275	280	285
Pro Val Glu Asp Ile Ser Ser Asn Arg Ile Tyr Lys Ile Leu Glu Leu		
290	295	300
Asn Gly Tyr Asp Pro Gln Tyr Ala Ala Ser Val Phe Leu Gly Trp Ala		
305	310	315
Thr Lys Lys Phe Gly Lys Arg Asn Thr Ile Trp Leu Phe Gly Pro Ala		
325	330	335
Thr Thr Gly Lys Thr Asn Ile Ala Glu Ala Ile Ala His Thr Val Pro		
340	345	350
Phe Tyr Gly Cys Val Asn Trp Thr Asn Glu Asn Phe Pro Phe Asn Asp		
355	360	365
Cys Val Asp Lys Met Val Ile Trp Trp Glu Glu Gly Lys Met Thr Ala		
370	375	380
Lys Val Val Glu Ser Ala Lys Ala Ile Leu Gly Gly Ser Lys Val Arg		
385	390	395
Val Asp Gln Lys Cys Lys Ser Ser Ala Gln Ile Asp Pro Thr Pro Val		
405	410	415
Ile Val Thr Ser Asn Thr Asn Met Cys Ala Val Ile Asp Gly Asn Ser		
420	425	430
Thr Thr Phe Glu His Gln Gln Pro Leu Gln Asp Arg Met Phe Lys Phe		
435	440	445
Glu Leu Thr Arg Arg Leu Asp His Asp Phe Gly Lys Val Thr Lys Gln		
450	455	460
Glu Val Lys Asp Phe Phe Arg Trp Ala Lys Asp His Val Val Glu Val		
465	470	475
Glu His Glu Phe Tyr Val Lys Lys Gly Gly Ala Lys Lys Arg Pro Ala		
485	490	495
Pro Ser Asp Ala Asp Ile Ser Glu Pro Lys Arg Val Arg Glu Ser Val		
500	505	510
Ala Gln Pro Ser Thr Ser Asp Ala Glu Ala Ser Ile Asn Tyr Ala Asp		
515	520	525
Arg Tyr Gln Asn Lys Cys Ser Arg His Val Gly Met Asn Leu Met Leu		
530	535	540
Phe Pro Cys Arg Gln Cys Glu Arg Met Asn Gln Asn Ser Asn Ile Cys		
545	550	555
Phe Thr His Gly Gln Lys Asp Cys Leu Glu Cys Phe Pro Val Ser Glu		
565	570	575
Ser Gln Pro Val Ser Val Val Lys Lys Ala Tyr Gln Lys Leu Cys Tyr		
580	585	590

Ile His His Ile Met Gly Lys Val Pro Asp Ala Cys Thr Ala Cys Asp
 595 600 605
 Leu Val Asn Val Asp Leu Asp Asp Cys Ile Phe Glu Gln
 610 615 620
 <210> 23
 <211> 509
 <212> PRT
 <213> 智人
 <400> 23
 Met Ser Met Gly Ala Pro Arg Ser Leu Leu Leu Ala Leu Ala Ala Gly
 1 5 10 15
 Leu Ala Val Ala Arg Pro Pro Asn Ile Val Leu Ile Phe Ala Asp Asp
 20 25 30
 Leu Gly Tyr Gly Asp Leu Gly Cys Tyr Gly His Pro Ser Ser Thr Thr
 35 40 45
 Pro Asn Leu Asp Gln Leu Ala Ala Gly Gly Leu Arg Phe Thr Asp Phe
 50 55 60
 Tyr Val Pro Val Ser Leu Cys Thr Pro Ser Arg Ala Ala Leu Leu Thr
 65 70 75 80
 Gly Arg Leu Pro Val Arg Met Gly Met Tyr Pro Gly Val Leu Val Pro
 85 90 95
 Ser Ser Arg Gly Gly Leu Pro Leu Glu Glu Val Thr Val Ala Glu Val
 100 105 110
 Leu Ala Ala Arg Gly Tyr Leu Thr Gly Met Ala Gly Lys Trp His Leu
 115 120 125
 Gly Val Gly Pro Glu Gly Ala Phe Leu Pro Pro His Gln Gly Phe His
 130 135 140
 Arg Phe Leu Gly Ile Pro Tyr Ser His Asp Gln Gly Pro Cys Gln Asn
 145 150 155 160
 Leu Thr Cys Phe Pro Pro Ala Thr Pro Cys Asp Gly Gly Cys Asp Gln
 165 170 175
 Gly Leu Val Pro Ile Pro Leu Leu Ala Asn Leu Ser Val Glu Ala Gln
 180 185 190
 Pro Pro Trp Leu Pro Gly Leu Glu Ala Arg Tyr Met Ala Phe Ala His
 195 200 205
 Asp Leu Met Ala Asp Ala Gln Arg Gln Asp Arg Pro Phe Phe Leu Tyr
 210 215 220
 Tyr Ala Ser His His Thr His Tyr Pro Gln Phe Ser Gly Gln Ser Phe
 225 230 235 240

Ala	Glu	Arg	Ser	Gly	Arg	Gly	Pro	Phe	Gly	Asp	Ser	Leu	Met	Glu	Leu			
																245	250	255
Asp	Ala	Ala	Val	Gly	Thr	Leu	Met	Thr	Ala	Ile	Gly	Asp	Leu	Gly	Leu			
																260	265	270
Leu	Glu	Glu	Thr	Leu	Val	Ile	Phe	Thr	Ala	Asp	Asn	Gly	Pro	Glu	Thr			
																275	280	285
Met	Arg	Met	Ser	Arg	Gly	Gly	Cys	Ser	Gly	Leu	Leu	Arg	Cys	Gly	Lys			
																290	295	300
Gly	Thr	Thr	Tyr	Glu	Gly	Gly	Val	Arg	Glu	Pro	Ala	Leu	Ala	Phe	Trp			
																305	310	315
Pro	Gly	His	Ile	Ala	Pro	Gly	Val	Thr	His	Glu	Leu	Ala	Ser	Ser	Leu			
																325	330	335
Asp	Leu	Leu	Pro	Thr	Leu	Ala	Ala	Leu	Ala	Gly	Ala	Pro	Leu	Pro	Asn			
																340	345	350
Val	Thr	Leu	Asp	Gly	Phe	Asp	Leu	Ser	Pro	Leu	Leu	Leu	Gly	Thr	Gly			
																355	360	365
Lys	Ser	Pro	Arg	Gln	Ser	Leu	Phe	Phe	Tyr	Pro	Ser	Tyr	Pro	Asp	Glu			
																370	375	380
Val	Arg	Gly	Val	Phe	Ala	Val	Arg	Thr	Gly	Lys	Tyr	Lys	Ala	His	Phe			
																385	390	395
Phe	Thr	Gln	Gly	Ser	Ala	His	Ser	Asp	Thr	Thr	Ala	Asp	Pro	Ala	Cys			
																405	410	415
His	Ala	Ser	Ser	Ser	Leu	Thr	Ala	His	Glu	Pro	Pro	Leu	Leu	Tyr	Asp			
																420	425	430
Leu	Ser	Lys	Asp	Pro	Gly	Glu	Asn	Tyr	Asn	Leu	Leu	Gly	Gly	Val	Ala			
																435	440	445
Gly	Ala	Thr	Pro	Glu	Val	Leu	Gln	Ala	Leu	Lys	Gln	Leu	Gln	Leu	Leu			
																450	455	460
Lys	Ala	Gln	Leu	Asp	Ala	Ala	Val	Thr	Phe	Gly	Pro	Ser	Gln	Val	Ala			
																465	470	475
Arg	Gly	Glu	Asp	Pro	Ala	Leu	Gln	Ile	Cys	Cys	His	Pro	Gly	Cys	Thr			
																485	490	495
Pro	Arg	Pro	Ala	Cys	Cys	His	Cys	Pro	Asp	Pro	His	Ala						
																500	505	

<210> 24

<211> 1527

<212> DNA

<213> 智人

<400> 24

atgtccatgg gggcaccgcg gtccctcctc ctggcccttg ctgctggcct ggccgttgcc 60
 cgtccgcccc acatcgtgct gatctttgcc gacgacctcg gctatgggga cctgggctgc 120
 tatgggcacc ccagctctac cactcccaac ctggaccagc tggcggcggg agggctgcgg 180
 ttcacagact tctacgtgcc tgtgtctctg tgcacaccct ctagggccgc cctcctgacc 240
 ggccggctcc cggttcggat gggcatgtac cctggcgtcc tggtgcccag ctcccggggg 300
 ggccctgcccc tggaggaggt gaccgtggcc gaagtcctgg ctgcccagg ctacctaca 360
 ggaatggccg gcaagtggca ccttgggggtg gggcctgagg gggccttcct gccccccat 420
 cagggcttcc atcgatttct aggcaccccg tactcccacg accagggccc ctgccagaac 480
 ctgacctgct tcccgccgac cactccttgc gacggtggct gtgaccagg cctgggtccc 540
 atccccactgt tggccaacct gtccgtggag gcgcagcccc cctggctgcc cggactagag 600
 gcccgtaca tggttttcgc ccatgacctc atggccgacg cccagcgcca ggatcgcccc 660
 ttcttctgt actatgcctc tcaccacacc cactaccctc agttcagtgg gcagagcttt 720
 gcagagcggt caggccgcgg gccatttggg gactccctga tggagctgga tgcagctgtg 780
 gggaccctga tgacagccat aggggacctg gggctgcttg aagagacgt ggtcatcttc 840
 actgcagaca atggacctga gaccatgcgt atgtcccag gcggctgctc cgtctcttg 900
 cgggtgtgaa agggaaacgac ctacgagggc ggtgtccgag agcctgcctt ggccttcttg 960
 ccaggtcata tcgtccccgg cgtgaccac gagctggcca gctccctgga cctgctgcct 1020
 accctggcag ccctggctgg ggccccactg cccaatgtca ccttgatgg ctttgacctc 1080
 agccccctgc tgctgggcac aggcaagagc cctcggcagt ctctcttctt ctaccctgc 1140
 taccagacg aggtccgtgg ggtttttgct gtgcggactg gaaagtacaa ggctcacttc 1200
 ttcaccaggg gctctgcccc cagtataacc actgcagacc ctgcctgcca cgcctccagc 1260
 tctctgactg ctcatgagcc cccgctgctc tatgacctgt ccaaggacct tggtgagaac 1320
 tacaacctgc tgggggggtg ggccggggcc acccagagg tgctgcaagc cctgaaacag 1380
 cttcagctgc tcaaggcccc gttagacgca gctgtgacct tcggccccag ccagggtggc 1440
 cggggcgagg accccgccct gcagatctgc tgtcatcctg gctgcacccc ccgcccagct 1500
 tgctgccatt gccagatcc ccatgcc 1527

<210> 25

<211> 278

<212> DNA

<213> 人工序列

<220>

<223> 合成多核苷酸

<400> 25

tcgaggtgag cccacgttc tgettcactc tccccatctc cccccctcc ccacccccaa 60
 ttttgtattt atttatTTTT taattatTTT gtgcagcgat gggggcgggg gggggggggg 120
 ggcgcgcgcc aggcggggcg gggcgggggc aggggcgggg cggggcgagg cggagaggtg 180
 cggcggcagc caatcagagc ggcgcgctcc gaaagtttcc ttttatggcg aggcggcggc 240
 ggcgggcgcc ctataaaaag cgaagcgcg gcggggcg 278

<210> 26

<211> 106

<212> DNA

<213> 人工序列

<220>

<223> 5' ITR

<400> 26

ctgcgcgctc gctcgtcac tgaggccgcc cgggcaaagc ccgggcgtcg ggcgaccttt 60
ggtcgccccg cctcagttag cgagcgagcg cgcagagagg gagggtg 106

<210> 27

<211> 143

<212> DNA

<213> 人工序列

<220>

<223> 3' ITR

<400> 27

aggaacccct agtgatggag ttggccactc cctctctgcg cgctcgtcg ctcactgagg 60
ccgggcgacc aaaggtcgcc cgacgcccgg gctttgcccg ggcggcctca gtgagcgagc 120
gagcgcgcag agagggagtg gcc 143

<210> 28

<211> 1873

<212> DNA

<213> 人工序列

<220>

<223> 合成多核苷酸

<400> 28

gatcttcaat attggccatt agccatatta ttcatgtgtt atatagcata aatcaatatt 60
ggctattggc cattgcatac gttgtatcta tatcataata tgtacattta tattggctca 120
tgtccaatat gaccgccatg ttggcattga ttattgacta gttattaata gtaatcaatt 180
acgggggtcat tagttcatag cccatatatg gaggttccg cgccattga cgtcaataat gacgtatgtt 240
ggcccgccctg gctgaccgcc caacgacccc cgccattga cgtcaataat gacgtatgtt 300
cccatagtaa cgccaatagg gactttccat tgacgtcaat gggtggagta tttacggtaa 360
actgcccact tggcagtaca tcaagtgtat catatgcaa gtccgcccc tattgacgtc 420
aatgacggta aatggccccg ctggcattat gccagtaga tgaccttac ggactttcct 480
acttggcagt acatctacgt attagtcac gctattacca tggtcgaggt gagccccacg 540
ttctgcttca ctctcccat ctccccccc tcccccccc caattttgta tttatttatt 600
ttttaattat tttgtgcagc gatgggggcg gggggggggg gggggcgcg gccagcggg 660
gcggggcggg gcgaggggcg gggcggggag aggcggagag gtgcggcggc agccaatcag 720
agcggcgcg tccgaaagt tccttttat gcgagcggg ggcggcggg gccctataaa 780
aagcgaagcg cgcgggcggc gggagtcgct gcgacgtgc cttcgcccc tgccccgctc 840

cgccgcccgc tcgcgcccgc cgccccggct ctgactgacc gcgttactcc cacaggtgag 900
 cgggcgggac ggcccttctc ctccgggctg taattagcgc ttggtttaat gacggcttgt 960
 ttcttttctg tggctgcgtg aaagccttga ggggctccgg gagggccctt tgtgcggggg 1020
 ggagcgggctc ggggggtgcg tgcgtgtgtg tgtgcgtggg gagcgccgcg tgcggcccgc 1080
 gctgcccggc ggctgtgagc gctgcgggcg cggcgcgggg ctttgtgcmc tccgcagtgt 1140
 gcgcgagggg agcgcggccg ggggcgggtc cccgcgggtc ggggggggct gcgaggggaa 1200
 caaaggctgc gtgcgggggtg tgtgcgtggg ggggtgagca gggggtgtgg gcgcggcggt 1260
 cgggctgtaa cccccccctg cccccctc cccgagttgc tgagcacggc ccggcttcgg 1320
 gtgcgggggt ccgtacgggg cgtggcgcgg ggctcgccgt gccgggcggg ggggtggcggc 1380
 aggtgggggt gccgggcggg gcggggccgc ctccggccgg ggagggtcgc ggggaggggc 1440
 gcgcgggccc ccggagcgcc ggcggtgtc gaggcgcggc gagccgcagc cattgccttt 1500
 tatggtaatc gtgcgagagg gcgcagggac ttctttgtc ccaaactctg gcggagccga 1560
 aatctgggag gcgccgccg accccctcta gcgggcgcgg ggccaagcgg tgcggcgccg 1620
 gcaggaagga aatgggcggg gagggccttc gtgcgtgcc gcgccccgt ccccttctcc 1680
 ctctccagcc tcggggctgt ccgcgggggg acggctgcct tcggggggga cggggcaggg 1740
 cggggttcgg cttctggcgt gtgaccggcg gctctagagc ctctgctaac catgttcatt 1800
 ccttcttctt tttcctacag ctctgggca acgtgctggt tattgtgctg tctcatcatt 1860
 ttggcaaaga att 1873

<210> 29

<211> 374

<212> PRT

<213> 智人

<400> 29

Met	Ala	Ala	Pro	Ala	Leu	Gly	Leu	Val	Cys	Gly	Arg	Cys	Pro	Glu	Leu
1			5					10						15	
Gly	Leu	Val	Leu	Leu	Leu	Leu	Leu	Ser	Leu	Leu	Cys	Gly	Ala	Ala	
			20					25						30	
Gly	Ser	Gln	Glu	Ala	Gly	Thr	Gly	Ala	Gly	Ala	Gly	Ser	Leu	Ala	Gly
			35					40						45	
Ser	Cys	Gly	Cys	Gly	Thr	Pro	Gln	Arg	Pro	Gly	Ala	His	Gly	Ser	Ser
			50					55						60	
Ala	Ala	Ala	His	Arg	Tyr	Ser	Arg	Glu	Ala	Asn	Ala	Pro	Gly	Pro	Val
65								70						75	
Pro	Gly	Glu	Arg	Gln	Leu	Ala	His	Ser	Lys	Met	Val	Pro	Ile	Pro	Ala
								85						90	
Gly	Val	Phe	Thr	Met	Gly	Thr	Asp	Asp	Pro	Gln	Ile	Lys	Gln	Asp	Gly
								100						105	
Glu	Ala	Pro	Ala	Arg	Arg	Val	Thr	Ile	Asp	Ala	Phe	Tyr	Met	Asp	Ala
								115						120	
															125

Tyr Glu Val Ser Asn Thr Glu Phe Glu Lys Phe Val Asn Ser Thr Gly		
130	135	140
Tyr Leu Thr Glu Ala Glu Lys Phe Gly Asp Ser Phe Val Phe Glu Gly		
145	150	155 160
Met Leu Ser Glu Gln Val Lys Thr Asn Ile Gln Gln Ala Val Ala Ala		
165	170	175
Ala Pro Trp Trp Leu Pro Val Lys Gly Ala Asn Trp Arg His Pro Glu		
180	185	190
Gly Pro Asp Ser Thr Ile Leu His Arg Pro Asp His Pro Val Leu His		
195	200	205
Val Ser Trp Asn Asp Ala Val Ala Tyr Cys Thr Trp Ala Gly Lys Arg		
210	215	220
Leu Pro Thr Glu Ala Glu Trp Glu Tyr Ser Cys Arg Gly Gly Leu His		
225	230	235 240
Asn Arg Leu Phe Pro Trp Gly Asn Lys Leu Gln Pro Lys Gly Gln His		
245	250	255
Tyr Ala Asn Ile Trp Gln Gly Glu Phe Pro Val Thr Asn Thr Gly Glu		
260	265	270
Asp Gly Phe Gln Gly Thr Ala Pro Val Asp Ala Phe Pro Pro Asn Gly		
275	280	285
Tyr Gly Leu Tyr Asn Ile Val Gly Asn Ala Trp Glu Trp Thr Ser Asp		
290	295	300
Trp Trp Thr Val His His Ser Val Glu Glu Thr Leu Asn Pro Lys Gly		
305	310	315 320
Pro Pro Ser Gly Lys Asp Arg Val Lys Lys Gly Gly Ser Tyr Met Cys		
325	330	335
His Arg Ser Tyr Cys Tyr Arg Tyr Arg Cys Ala Ala Arg Ser Gln Asn		
340	345	350
Thr Pro Asp Ser Ser Ala Ser Asn Leu Gly Phe Arg Cys Ala Ala Asp		
355	360	365
Arg Leu Pro Thr Met Asp		
370		
<210> 30		
<211> 2718		
<212> DNA		
<213> 人工序列		
<220>		
<223> 合成多核苷酸		
<400> 30		

```

atgagcatgg gcgccccag aagcctgtta cttgcttttag ctgctggcct tgcagtggca 60
aggcccccta acatcgtgct gatctttgca gatgacttgg gatatgggga tcttggttgt 120
tatggccacc catcaagcac aactcccaat ctggatcagt tggctgcagg aggtctgagg 180
tttacagact tttatgttcc agtctccctg tgcactcctt ctcgggctgc cctgcttact 240
gggaggctcc ctgtgagaat gggatatgtac cctggagtgt tgggtccatc cagcagggga 300
gggctgcccc tggaagaggt gacagtggca gaggtgctgg cagcacgagg ctatctgact 360
ggcatggcag gcaagtggca cctgggtgta gggccagagg gtgctttcct gcctcccat 420
cagggctttc ataggtttct gggaatccca tactctcatg accaaggacc ctgccagaac 480
ctcacctgtt tccccctgc aacacatgt gatggggct gtgatcaagg tctggttct 540
ataccactgc ttgctaattt ttcagtggaa gctcaaccac cctggctgcc tggcttgagg 600
gctagataca tggccttcgc acatgatctg atggcagatg cccagagaca agataggcct 660
ttcttctct actatgcac tcaccacacc cactatctc agttctcagg ccaatcattt 720
gctgagcgta gtggcagggg cccatttggg gacagtttga tggaactgga tgccgcagtt 780
ggtaccctca tgacagcaat aggggactta ggtttgctgg aggaaacatt ggtaattttc 840
acagctgata atggccctga gacaatgaga atgtctaggg gaggtgctc tgggtcttctg 900
aggtgtggta aagggactac atatgaggga ggagtggagg aaccagctct tgccttttgg 960
ccaggtcaca tagcccctgg agttacacat gaactagctt cttccctgga cttgcttct 1020
acactggcag ccctggcagg tgcccctct cctaattgaa ctttagatgg atttgacctc 1080
tctccactac ttttagggac agggaaaagt ccaaggcagt cttattctt ctatccttcc 1140
taccagatg aggtgagggg tgtttttgcc gtgaggactg ggaaatacaa agctcatttt 1200
tttaccagg gatcagctca ttcagacacc acagctgatc ctgcctgtca tgccagcagt 1260
agcttgacag cacatgagcc tcccttactg tatgacctga gcaaggacc aggggagaac 1320
tataacctgc ttgggggggt tgctggggcc acccagaag tgcttcaggc actaaagcag 1380
ctgcaactgc ttaaagcaca gttggatgct gcagtacct ttggcccttc ccaggtggcc 1440
agaggcgagg atcccgccct gcagatctgc tgccaccag gctgcacacc cagacctgcc 1500
tgctgtcact gccccgacc acacgccggc agcggagcta ctaacttcag cctgctgaag 1560
caggctggag acgtggagga gaaccctgga cctatggctg cccagccct ggggctgggtg 1620
tgtggcagat gccctgagct gggcctgggt ctgcttctcc tgctgctgag cctcctgtgt 1680
ggtgctgctg gctctcagga agcagggaca ggagcaggag caggttctct ggctggctca 1740
tgcggttgtg ggacccccca gaggccagg gctcatgggt cctctgcagc tgcccacagg 1800
tactcaaggg aagcaaatgc ccctggcccc gtacctgggg aaaggcaact tgctcactcc 1860
aagatggttc ctatccctgc aggagttttt actatgggaa ctgatgacc tcagatcaag 1920
caggatgggtg aagcaccagc taggagagtc acaattgatg cttctatat ggatgcctat 1980
gaagtgtcaa acacagaatt tgagaaattt gtaaacagca ctggatacct tacagaggct 2040
gagaaatttg gtgacagttt tgtttttgaa ggcatgctaa gtgagcaggt gaagaccaat 2100
atccaacagg cagtggctgc agccccctgg tggctgcctg ttaaaggagc caattggaga 2160
caccagagg gaccagactc aactatctc cacaggcctg accacctgt gctgcatgtg 2220
tcttggaatg atgcagtggc atactgcacc tgggctggga aaaggttacc aacagaggca 2280
gaatgggagt attcctgccg ggggtggactg cacaacagac tgttcccctg gggcaataag 2340

```

ctgcaaccta aaggacagca ttatgccaat atttggcagg gagagtcccc agtcacaaac 2400
 actggtgagg atggcttcca gggaactgcc cctgtggatg ctttcccacc caatggctat 2460
 gggttgtaca atatagttgg gaatgcctgg gagtggactt ctgactggtg gacggtccat 2520
 cacagtgtgg aagagacact gaacccaaag gggccccctt caggcaagga cagagtcaag 2580
 aaaggtggct cttatatgtg tcacagaagc tattgctaca gatataggtg tgctgcaaga 2640
 agtcagaaca cccctgacag ctcagctagc aatctgggat ttagatgtgc agcagataga 2700
 ctccccacca tggactga 2718

<210> 31

<211> 93

<212> DNA

<213> 人工序列

<220>

<223> 合成多核苷酸

<400> 31

ctctaaggta aatataaaat ttttaagtgt ataatgtgtt aaactactga ttctaattgt 60
 ttctctcttt tagattccaa cctttggaac tga 93

<210> 32

<211> 1017

<212> DNA

<213> 人工序列

<220>

<223> 合成多核苷酸

<400> 32

ggagtcgctg cgcgctgcct tcgccccgtg ccccgctccg ccgcccgcctc gcgcccgcctc 60
 ccccggtctt gactgaccgc gttactccca caggtgagcg ggcgggacgg cccttctcct 120
 ccgggctgta attagcgctt ggtttaatga cggcttggtt cttttctgtg gctgctgaa 180
 agccttgagg ggctccggga gggccctttg tgcgggggga gcggctcggg ggggtgcgtgc 240
 gtgtgtgtgt gcgtggggag cgccgcgtgc ggctccgcgc tgcccggcgg ctgtgagcgc 300
 tgcgggcgcg gcgcggggct ttgtgcgtc gcagtggtgc gcgaggggag cgcgccggg 360
 ggcggtgccc cgcggtgcgg ggggggctgc gaggggaaca aaggctgcgt gcgggggtgtg 420
 tgcgtggggg ggtgagcagg ggggtgtggc gcgtcggtcg ggctgcaacc cccctgcac 480
 cccctcccc gagttgctga gcacggcccg gcttcgggtg cggggctccg tacggggcgt 540
 ggcgcggggc tcgccgtgcc gggcgggggg tggcggcagg tgggggtgcc gggcggggcg 600
 gggccgcctc gggccgggga gggtcgggg gaggggcgcg gcggccccc gagcgccggc 660
 ggctgtcgag gcgcggcgag ccgcagccat tgccttttat ggtaatcgtg cgagagggcg 720
 cagggaactt ctttgtccca aatctgtgcg gagccgaaat ctgggaggcg ccgccgcacc 780
 ccctctagcg ggcgcggggc gaagcgggtc ggcccgga ggaaggaaat gggcggggag 840
 ggcttcgtg cgtcgccgcg ccgccgtccc cttctccctc tccagcctcg gggctgtccg 900
 cggggggacg gctgccttcg ggggggacgg ggcagggcgg ggttcggctt ctggcgtgtg 960

accggcggct ctagagcctc tgctaaccat gttcatgcct tcttcttttt cctacag 1017

<210> 33

<211> 79

<212> PRT

<213> 智人

<400> 33

Gly Asp Val Cys Gln Asp Cys Ile Gln Met Val Thr Asp Ile Gln Thr

1 5 10 15

Ala Val Arg Thr Asn Ser Thr Phe Val Gln Ala Leu Val Glu His Val

20 25 30

Lys Glu Glu Cys Asp Arg Leu Gly Pro Gly Met Ala Asp Ile Cys Lys

35 40 45

Asn Tyr Ile Ser Gln Tyr Ser Glu Ile Ala Ile Gln Met Met Met His

50 55 60

Met Gln Pro Lys Glu Ile Cys Ala Leu Val Gly Phe Cys Asp Glu

65 70 75

<210> 34

<211> 8

<212> PRT

<213> 人工序列

<220>

<223> 合成多核苷酸

<220>

<221> MISC_FEATURE

<222> (1) .. (1)

<223> Xaa为D或G

<220>

<221> MISC_FEATURE

<222> (2) .. (2)

<223> Xaa为V或I

<220>

<221> MISC_FEATURE

<222> (4) .. (4)

<223> Xaa为任何氨基酸

<400> 34

Xaa Xaa Glu Xaa Asn Pro Gly Pro

1 5

<210> 35

<211> 92

<212> DNA

<213> 人工序列

<220>

<223> 合成多核苷酸

<400> 35

aagaggtaag ggtttaaggg atggttggtt ggtgggggtat taatgtttaa ttacctggag 60
cacctgcctg aaatcacttt ttttcaggtt gg 92

<210> 36

<211> 1676

<212> DNA

<213> 人工序列

<220>

<223> 合成多核苷酸

<400> 36

ggcattgatt attgactagt tattaatagt aatcaattac ggggtcatta gttcatagcc 60
catatatgga gttccgcgtt acataactta cggtaaatgg cccgcctggc tgaccgcca 120
acgacccccg cccattgacg tcaataatga cgtatgttcc catagtaacg ccaataggga 180
ctttccattg acgtcaatgg gtggagtatt tacggtaaac tgcccacttg gcagtacatc 240
aagtgtatca tatgccaaagt ccgcccccta ttgacgtcaa tgacggtaaa tggcccgctt 300
ggcattatgc ccagtacatg accttacggg actttcctac ttggcagtac atctacgtat 360
tagtcatcgc tattaccatg gtcgaggtga gccccacgtt ctgcttcaact ctccccatct 420
cccccccctc cccaccccca attttgtatt tatttatttt ttaattattt tgtgcagcga 480
tgggggcggg gggggggggg gggcgcgcg caggcggggc ggggcggggc gaggggcggg 540
gcggggcgag gcggagaggt gcggcggcag ccaatcagag cggcgcgctc cgaaagtttc 600
cttttatggc gaggcggcgg cggcggcggc cctataaaaa gcgaagcgcg cggcgggcgg 660
gagtcgctgc gcgctgcctt cgccccgtgc cccgctccgc cgccgcctcg cgccgcccgc 720
cccggtcttg actgaccgcg ttactccac aggtgagcgg gcgggacggc ctttctcctc 780
cgggctgtaa ttagcgcttg gtttaatgac ggcttgtttc ttttctgtgg ctgcgtgaaa 840
gccttgaggg gctccgggag ggccctttgt gcggggggag cggctcgggg ggtgcgtgcg 900
tgtgtgtgtg cgtggggagc gccgcgtgcg gctccgcgct gccggcggc tgtgagcgct 960
gcgggcgcgg gcgggggctt tgtgcgctcc gcagtgtgcg cgaggggagc gcggccgggg 1020
gcggtgcccc gcggtgcggg gggggctgcg aggggaacaa aggctgcgtg cggggtgtgt 1080
gcgtgggggg gtgagcaggg ggtgtgggcg cgtcggtcgg gctgcaacc cccctgcacc 1140
cccctccccg agttgctgag cacggcccgg cttcgggtgc ggggctccgt acggggcgctg 1200
gcgcgggggt cgccgtgccg ggccgggggt ggccggcaggt gggggtgccg ggccggggcg 1260
ggccgcctcg ggccggggag ggctcggggg aggggcgcgg cgcccccg agcgcggcg 1320
gctgtcgagg cgcggcgagc cgcagccatt gccttttatg gtaatcgtgc gagaggcgcg 1380
agggacttcc tttgtcccaa atctgtgcgg agccgaaatc tgggagcgcg cgccgcaccc 1440
cctctagcgg gcgcggggcg aagcggtgcg gcgccggcag gaaggaaatg ggcggggagg 1500

gccttcgtgc gtcgccgcgc cgccgtcccc ttctccctct ccagcctcgg ggctgtccgc 1560
ggggggacgg ctgccttcgg gggggacggg gcagggcggg gttcggcttc tggcgtgtga 1620
ccggcggctc tagagcctct gctaaccatg ttcatgcctt cttctttttc ctacag 1676

<210> 37

<211> 16

<212> PRT

<213> 人工序列

<220>

<223> T2A肽

<400> 37

Glu Gly Arg Gly Ser Leu Leu Thr Cys Gly Asp Val Glu Glu Asn Pro

1 5 10 15

<210> 38

<211> 16

<212> PRT

<213> 人工序列

<220>

<223> P2A肽

<400> 38

Ala Thr Asn Phe Ser Leu Leu Lys Gln Ala Gly Asp Val Glu Glu Asn

1 5 10 15

<210> 39

<211> 540

<212> DNA

<213> 人工序列

<220>

<223> 合成多核苷酸

<400> 39

cctgcaggct caccagtgtt tgtgactggg aactetccct gccaaatatt ggcataatgc 60
tgtcctttag gttgcagctt attgcccag gggaacagtc tgttgtgcag tccaccccg 120
caggaatact ccattctgc ctctgttggg aaccttttcc cagcccagggt gcagtatgcc 180
actgcatcat tccaggacac atgcagcaca ggggtggtcag gcctgtggag gatagttgag 240
tctgggccct ctgggtgtct ccaattggct cctttaacag gcagccacca gggggctgca 300
gccactgcct gttggatatt ggtcttcacc tgctcactta gcatgccttc aaaaacaaaa 360
ctgtcaccaa atttctcagc ctctgtaagg tatccagtgc tgtttacaaa tttctcaaat 420
tctgtgtttg acacttcata ggcatccata tagaaggcat caattgtgac tctcctagct 480
gggtgcttcac catcctgctt gatctgaggg tcatcagttc ccatagtaaa aactcctgca 540

<210> 40

<211> 1168

<212> DNA

<213> 人工序列

<220>

<223> 合成多核苷酸

<400> 40

```

cgtgaggctc cgggtgcccgt cagtgggcag agcgcacatc gcccacagtc cccgagaagt 60
tggggggagg ggtcggcaat tgaaccggtg cctagagaag gtggcgcggg gtaaactggg 120
aaagtgatgt cgtgtactgg ctccgccttt ttcccagggtg tgggggagaa ccgtatataa 180
gtgcagtagt cgccgtgaac gttctttttc gcaacggggtt tgccgccaga acacaggtaa 240
gtgccgtgtg tggttcccgc gggcctggcc tctttacggg ttatggccct tgcgtgcctt 300
gaattacttc cacctggctc cagtacgtga ttcttgatcc cgagctggag ccaggggagg 360
gccttgcgct ttaggagccc cttcgctcgc tgcttgagtt gaggcctggc ctgggcgctg 420
gggccgccgc gtgcgaatct ggtggcacct tcgcgcctgt ctgcgtgctt tcgataagtc 480
tctagccatt taaaattttt gatgacctgc tgcgacgctt tttttctggc aagatagtct 540
tgtaaattgc ggccaggatc tgcacactgg tatttcgggtt tttggggccg cgggcggcga 600
cggggcccgt gcgtcccagc gcacatgttc ggcgaggcgg ggcctgcgag cgcggccacc 660
gagaatcgga cgggggtagt ctcaagctgg ccggcctgct ctggtgcctg gcctcgcgcc 720
gccgtgtatc gccccgccct gggcggcaag gctggcccgg tcggcaccag ttgcgtgagc 780
ggaaagatgg ccgcttcccg gccctgctcc agggggctca aaatggagga cgcggcgctc 840
gggagagcgg gcgggtgagt caccacaca aaggaaaggg gcctttccgt cctcagccgt 900
cgcttcatgt gactccacgg agtaccgggc gccgtccagg cacctcgatt agttctggag 960
cttttggagt acgtcgtctt taggttgggg ggaggggttt tatgcgatgg agtttcccca 1020
cactgagtgg gtggagactg aagttaggcc agcttggcac ttgatgtaat tctccttgga 1080
at ttgcccctt tttgagtttg gatcttggtt cattctcaag cctcagacag tggttcaaag 1140
tttttttctt ccatttcagg tgtcgtga 1168

```

<210> 41

<211> 3416

<212> DNA

<213> 人工序列

<220>

<223> 合成多核苷酸

<400> 41

```

ggcattgatt attgactagt tattaatagt aatcaattac ggggtcatta gttcatagcc 60
catatatgga gttccgcgtt acataactta cggtaaattg cccgcctggc tgaccgcccc 120
acgacccccg cccattgacg tcaataatga cgtatgttcc catagtaacg ccaataggga 180
ctttccattg acgtcaatgg gtggagtatt tacggtaaac tgcccacttg gcagtacatc 240
aagtgtatca tatgccaagt ccgcccccta ttgacgtcaa tgacggtaaa tggccccgct 300
ggcatttatgc ccagtacatg accttacggg actttcctac ttggcagtac atctacgtat 360
tagtcatcgc tattaccatg gtcgaggtga gcccacgtt ctgcttcact ctccccatct 420

```

```

ccccccctc cccaccccca attttgtatt tatttatttt ttaattattt tgtgcagcga 480
tgggggcggg gggggggggg gggcgcgcg caggcggggc ggggcggggc gaggggcggg 540
gcggggcgag gcggagaggt gcggcggcag ccaatcagag cggcgcgctc cgaaagtttc 600
cttttatggc gaggcggcgg cggcggcggc cctataaaaa gcgaagcgcg cggcgggcgg 660
gagtcgctgc gcgctgcctt cgccccgtgc cccgctccgc cgccgcctcg cgccgcccgc 720
cccggtcttg actgaccgcg ttactcccac aggtgagcgg gcgggacggc ctttctcctc 780
cgggctgtaa ttagcgcttg gtttaatgac ggcttgtttc ttttctgtgg ctgctgaaa 840
gccttgaggg gctccgggag ggccctttgt gcggggggag cggctcgggg ggtgcgtgcg 900
tgtgtgtgtg cgtggggagc gccgcgtgcg gctccgcgt gcccggcggc tgtgagcgt 960
gcgggcgcgg cgcggggctt tgtgcgctcc gcagtgtgcg cgaggggagc gcggccgggg 1020
gcggtgcccc gcggtgcggg gggggctgcg aggggaacaa aggctgcgtg cggggtgtgt 1080
gcgtgggggg gtgagcaggg ggtgtgggcg cgtcggtcgg gctgcaacc cccctgcacc 1140
cccctccccg agttgctgag cacggcccgg cttcgggtgc ggggctccgt acggggcgtg 1200
gcgcgggggt cgccgtgccg ggccgggggt ggccggcaggt gggggtgccg ggccgggcgg 1260
ggccgcctcg ggccggggag ggctcggggg aggggcgcgg cggcccccgg agcgcggcg 1320
gctgtcgagg cgcggcgagc cgagccatt gccttttatg gtaatcgtgc gagaggcg 1380
agggacttcc tttgtcccaa atctgtgcgg agccgaaatc tgggagggcg cgccgcacc 1440
cctctagcgg gcgcggggcg aagcgggtgc gcgccggcag gaaggaaatg ggccgggagg 1500
gccttcgtgc gtcgccgcgc cgccgtcccc ttctccctct ccagcctcg ggctgtccgc 1560
ggggggacgg ctgccttcgg gggggacggg gcagggcggg gttcggcttc tggcgtgtga 1620
ccggcggtc tagagcctct gctaaccatg ttcatgcctt cttctttttc ctacagctcc 1680
tgggcaacgt gctggttatt gtgctgtctc atcattttgg caaagaattc cgccaccatg 1740
tccatggggg caccgcggtc cctcctcctg gccctggctg ctggcctggc cgttgcccgt 1800
ccgccaaca tcgtgctgat ctttgccgac gacctcggct atggggacct gggctgctat 1860
gggcacccca gctctaccac tcccaacctg gaccagctgg cggcgggagg gctgcggttc 1920
acagacttct acgtgcctgt gtctctgtgc acaccctcta gggccgccct cctgaccggc 1980
cggctcccgg ttcggatggg catgtaccct ggctcctgg tggccagctc ccgggggggc 2040
ctgcccctgg aggaggtgac cgtggccgaa gtcttggtg cccgaggcta cctcacagga 2100
atggccggca agtggcacct tggggtgggg cctgaggggg ccttcctgcc ccccatcag 2160
ggcttccatc gattttctag catcccgtac tcccacgacc agggcccctg ccagaacctg 2220
acctgcttcc cgccggccac tccttgcgac ggtggctgtg accagggcct ggtccccatc 2280
ccactgttgg ccaacctgtc cgtggaggcg cagccccct ggctgcccgg actagaggcc 2340
cgctacatgg ctttcgcca tgacctcatg gccgacgcc agcgcaggga tcgccccttc 2400
ttcctgtact atgcctctca ccacaccac taccctcagt tcagtgggca gagctttgca 2460
gagcgttcag gccgcgggcc atttggggac tcctgatgg agctggatgc agctgtgggg 2520
accctgatga cagccatagg ggacctgggg ctgcttgaag agacgtggt catcttact 2580
gcagacaatg gacctgagac catgcgtatg tcccaggcg gctgctccgg tctcttgccg 2640
tgtggaaagg gaacgaccta cgaggcggt gtccgagagc ctgccttggc cttctggcca 2700
ggtcatatcg ctcccgcggt gaccacgag ctggccagct ccctggacct gctgcctacc 2760

```

ctggcagccc tggctggggc cccactgccc aatgtcacct tggatggctt tgacctcage 2820
cccctgctgc tgggcacagg caagagccct cggcagtctc tcttcttcta cccgtcctac 2880
ccagacgagg tccgtggggg ttttgctgtg cggactggaa agtacaaggc tcacttcttc 2940
accagaggct ctgcccacag tgataccact gcagaccctg cctgccacgc ctccagctct 3000
ctgactgctc atgagcccc gctgctctat gacctgtcca aggaccctgg tgagaactac 3060
aacctgctgg ggggtgtggc cggggccacc ccagaggtgc tgcaagccct gaaacagctt 3120
cagctgctca agggccagtt agacgcagct gtgaccttcg gccccagcca ggtggcccgg 3180
ggcgaggacc ccgccctgca gatctgctgt catcctggct gcaccccccg cccagcttgc 3240
tgccattgcc cagatcccca tgcctgagat tctagagtcg agccgcggac tagtaacttg 3300
tttattgcag cttataatgg ttacaaataa agcaatagca tcacaaattt cacaaataaa 3360
gcattttttt cactgcattc tagttgtggg ttgtccaaac tcatcaatgt atctta 3416

<210> 42

<211> 122

<212> DNA

<213> 人工序列

<220>

<223> 合成多核苷酸

<400> 42

aacttgttta ttgcagctta taatggttac aaataaagca atagcatcac aaatttcaca 60
aataaagcat ttttttcact gcattctagt tgtggtttgt ccaaactcat caatgtatct 120
ta 122

<210> 43

<211> 133

<212> DNA

<213> 人工序列

<220>

<223> 合成多核苷酸

<400> 43

tgctttatatt gtgaaatttg tgatgctatt gctttatttg taaccattat aagctgcaat 60
aaacaagtta acaacaacaa ttgcattcat tttatgtttc aggttcaggg ggaggtgtgg 120
gaggtttttt aaa 133

<210> 44

<211> 3416

<212> DNA

<213> 人工序列

<220>

<223> 合成多核苷酸

<400> 44

ggcattgatt attgactagt tattaatagt aatcaattac ggggtcatta gttcatagcc 60

```

catatatgga gttccgcgtt acataactta cggtaaattgg cccgcctggc tgaccgcccc 120
acgacccccg cccattgacg tcaataatga cgtatgttcc catagtaacg ccaataggga 180
ctttccattg acgtcaatgg gtggagtatt tacggtaaac tgcccacttg gcagtacatc 240
aagtgtatca tatgccaaagt ccgcccccta ttgacgtcaa tgacggtaaa tggcccgctt 300
ggcattatgc ccagtacatg accttacggg actttcctac ttggcagtac atctacgtat 360
tagtcatcgc tattaccatg gtcgaggtga gccccacgtt ctgcttcaact ctccccatct 420
cccccccctc cccacccccca attttgtatt tatttatttt ttaattattt tgtgcagcga 480
tgggggcggg gggggggggg gggcgcgcg caggcggggc ggggcggggc gaggggcggg 540
gcggggcgag gcggagaggt gcggcggcag ccaatcagag cggcgcgctc cgaaagtttc 600
cttttatggc gaggcggcgg cggcggcggc cctataaaaa gcgaagcgcg cggcgggcgg 660
gagtcgctgc gcgctgcctt cgccccgtgc cccgctccgc cgccgcctcg cgccgcccgc 720
cccggctctg actgaccgcg ttactcccac aggtgagcgg gcgggacggc ctttctctc 780
cgggctgtaa ttagcgcttg gtttaatgac ggcttgtttc ttttctgttg ctgcgtgaaa 840
gccttgaggg gctccgggag ggccctttgt gcggggggag cggctcgggg ggtgcgtgcg 900
tgtgtgtgtg cgtggggagc gccgcgtgcg gctccgcgct gcccggcggc tgtgagcgct 960
gcgggcgcgg cgcggggctt tgtgcgtcc gcagtgtgcg cgaggggagc gcggccgggg 1020
gcggtgcccc gcggtgcggg gggggctgcg aggggaacaa aggctgcgtg cggggtgtgt 1080
gcgtgggggg gtgagcaggg ggtgtgggcg cgtcggtcgg gctgcaacc cccctgcacc 1140
cccctccccg agttgctgag cacggcccgg cttcgggtgc ggggctccgt acggggcgctg 1200
gcgcggggct cgccgtgccg ggccgggggt ggccggcaggt gggggtgccg ggccgggcgg 1260
ggccgcctcg ggccggggag ggctcggggg aggggcgcgg cggcccccg agcgccggcg 1320
gctgtcgagg cgcggcgagc cgagccatt gccttttatg gtaatcgtgc gagaggcgcg 1380
agggacttcc tttgtcccaa atctgtgcgg agccgaaatc tgggaggcg cgccgcaccc 1440
cctctagcgg gcgcggggcg aagcgggtgc gcgccggcag gaaggaaatg ggccgggagg 1500
gccttcgtgc gtcgccgcgc cgccgtcccc ttctccctct ccagcctcg ggctgtccgc 1560
ggggggacgg ctgccttcgg gggggacggg gcagggcggg gttcggcttc tggcgtgtga 1620
ccggcggtc tagagcctct gtaaccatg ttcatgcctt cttctttttc ctacagctcc 1680
tgggcaacgt gctggttatt gtgtgtctc atcatttttg caaagaattc cgccaccatg 1740
tctatggggg ctctcgctc cctgtctgtg gcactggccg ccgggctggc tgtcgcaaga 1800
ccaccttaata tcgtcctgat ttttgagac gatctgggat acggcgacct gggatgctat 1860
ggccacccaa gctccaccac acccaacctg gaccagctgg cagcaggagg cctgcggttc 1920
accgacttct acgtgccagt gagcctgtgc acccctcca gagccgcct gctgacaggc 1980
aggctgccag tgcgcatggg catgtatcct ggcgtgctgg tgccatctag caggggcggc 2040
ctgccactgg aggaggtgac cgtggcagag gtgctggcag ccagaggcta cctgacagga 2100
atggccggca agtggcacct gggagtggga ccagaggag ccttctgcc cctcaccag 2160
ggcttccacc ggtttctggg catcccttat tctcagacc agggcccatg ccagaacctg 2220
acctgttttc caccagcaac accatgcgac ggaggatgtg atcaggcct ggtgccaatc 2280
ccactgctgg caaatctgag cgtggaggca cagcctccat ggctgcctgg cctggaggca 2340
agatacatgg ccttcgcccc cgacctgat gcagatgcac agcggcagga tagaccttc 2400

```

ttctctgtact atgcctccca ccacacccac tatccacagt tcagcggcca gtcctttgcc 2460
 gagaggtccg gaaggggacc attcggcgac tctctgatgg agctggatgc cgccgtgggc 2520
 accctgatga cagcaatcgg cgacctgggc ctgctggagg agacactggg catcttcacc 2580
 gccgataacg gccctgagac aatgcggatg tctagaggcg gatgcagcgg cctgctgaga 2640
 tgtggcaagg gaaccacata cgagggaggc gtgcgcgagc ctgccctggc attttggcca 2700
 ggacacatcg cacctggagt gacccacgag ctggcctcct ctctggacct gctgccaaca 2760
 ctggccgccc tggcaggagc acctctgcca aatgtgacct tggacggctt cgatctgagc 2820
 ccaactgctgc tgggaaccgg caagtcccct aggcagtctc tgttctttta cccctcctat 2880
 cctgatgagg tgcggggcgt gtttgccgtg agaaccggca agtacaaggc ccacttcttt 2940
 acacagggtc ctgcccacag cgacaccaca gcagatccag catgccacgc cagctcctct 3000
 ctgaccgcac acgagccacc tctgctgtac gacctgtcca aggatcccgg cgagaactat 3060
 aatctgctgg gaggagtggc aggagcaacc cctgaggtgc tgcaggccct gaagcagctg 3120
 cagctgctga aggcacagct ggacgcagca gtgacattcg gccaagcca ggtggccaga 3180
 ggcgaggatc ccgccctgca gatctgttgc caccgccgtt gcacccaag acctgcctgt 3240
 tgccattgcc ccgaccaca cgcctaagat tctagagtcg agccgcggac tagtaacttg 3300
 ttatttgcag cttataatgg ttacaaataa agcaatagca tcacaaattt cacaataaaa 3360
 gcattttttt cactgcattc tagttgtggt ttgtccaaac tcatcaatgt atctta 3416

<210> 45

<211> 198

<212> DNA

<213> 人工序列

<220>

<223> 合成多核苷酸

<400> 45

gatccagaca tgataagata cattgatgag tttggacaaa ccacaactag aatgcagtga 60
 aaaaaatgct ttatttgtga aatttgtgat gctattgctt tatttgtaac cattataagc 120
 tgcaataaac aagttaacaa caacaattgc attcatttta tgtttcaggt tcagggggag 180
 gtgtgggagg ttttttaa 198

<210> 46

<211> 3416

<212> DNA

<213> 人工序列

<220>

<223> 合成多核苷酸

<400> 46

ggcatgtatt attgactagt tattaatagt aatcaattac ggggtcatta gttcatagcc 60
 catatatgga gttccgcgtt acataactta cggtaaatgg cccgcctggc tgaccgcca 120
 acgacccccg cccattgacg tcaataatga cgtatgttcc catagtaacg ccaataggga 180
 ctttccattg acgtcaatgg gtggagtatt tacggtaaac tgcccacttg gcagtacatc 240

```

aagtgtatca tatgccaagt ccgcccccta ttgacgtcaa tgacggtaaa tggccccgct 300
ggcattatgc ccagtacatg accttacggg actttcctac ttggcagtac atctacgtat 360
tagtcatcgc tattaccatg gtcgaggtga gccccacgtt ctgcttcaact ctccccatct 420
cccccccctc cccacccccca attttgtatt tattttatit ttaattatit tgtgcagcga 480
tgggggcggg gggggggggg gggcgcgcg caggcggggc ggggcggggc gaggggcggg 540
gcggggcgag gcggagaggt gcggcggcag ccaatcagag cggcgcgctc cgaaagtttc 600
cttttatggc gaggcggcgg cggcggcggc cctataaaaa gcgaagcgcg cggcgggcgg 660
gagtcgctgc gcgctgcctt cgccccgtgc cccgctccgc cgccgcctcg cgccgcccgc 720
cccggtcttg actgaccgcg ttactccac aggtgagcgg gcgggacggc ctttctctc 780
cgggctgtaa ttagcgcttg gtttaatgac ggcttgtttc ttttctgtgg ctgcgtgaaa 840
gccttgaggg gctccgggag ggccctttgt gcggggggag cggctcgggg ggtgcgtgcg 900
tgtgtgtgtg cgtggggagc gccgctgcg gctccgcgt gcccggcggc tgtgagcgt 960
gcgggcgcgg cgcggggctt tgtgcgtcc gcagtgtgc cgaggggagc gcggccgggg 1020
gcggtgcccc gcggtgcggg gggggctgc aggggaacaa aggctgcgtg cggggtgtgt 1080
gcgtgggggg gtgagcaggg ggtgtgggcg cgtcggtcgg gctgcaacc cccctgcacc 1140
cccctccccg agttgctgag cacggccccg cttcgggtgc ggggctccgt acggggcgtg 1200
gcgcggggct cgccgtgccg ggcggggggt ggcggcaggt gggggtgccg ggcggggcgg 1260
ggccgcctcg ggccggggag ggctcggggg aggggcgcgg cgcccccg agcgccggcg 1320
gctgtcgagg cgcggcgagc cgcagccatt gccttttatg gtaatcgtgc gagaggcg 1380
agggacttcc tttgtcccaa atctgtgcgg agccgaaatc tgggaggcgc cgccgcacc 1440
cctctagcgg gcgcggggcg aagcgggtgc gcgccggcag gaaggaaatg ggcggggagg 1500
gccttcgtgc gtcgccgcgc cgccgtcccc ttctccctct ccagcctcgg ggctgtccgc 1560
ggggggacgg ctgccttcgg gggggacggg gcagggcggg gttcggcttc tggcgtgtga 1620
ccggcggtc tagagcctct gctaaccatg ttcatgcctt cttctttttc ctacagctcc 1680
tgggcaacgt gctggttatt gtgctgtctc atcatttttg caaagaattc cgccaccatg 1740
tctatggggg ctctcgtc cctgctgtg gactggccg ccgggctggc tgtcgcaaga 1800
ccaccttaata tcgtcctgat ttttgacag gatctgggat acggcgacct gggatgctat 1860
ggccacccaa gctccaccac acccaacctg gaccagctgg cagcaggagg cctgcggttc 1920
accgacttct acgtgccagt gagcctgtgc acccctcca gagccgcct gctgacaggc 1980
aggctgccag tgcgcatggg catgtatcct ggcgtgctgg tgccatctag caggggcggc 2040
ctgccactgg aggaggtgac cgtggcagag gtgctggcag ccagaggcta cctgacagga 2100
atggccggca agtggcacct gggagtggga ccagaggag ccttctgcc ccctcaccag 2160
ggcttccacc ggtttctggg catcccttat tctcagacc agggcccatg ccagaacctg 2220
acctgttttc caccagcaac accatgcgac ggaggatgtg atcagggcct ggtgccaatc 2280
ccactgctgg caaatctgag cgtggaggca cagcctccat ggctgcctgg cctggaggca 2340
agatacatgg ccttcgcccc cgacctgat gcagatgcac agcggcagga tagaccttcc 2400
tttctgtact atgcctcccc ccacaccac tatccacagt tcagcgcca gtcctttgcc 2460
gagaggctcg gaaggggacc attcggcgac tctctgatgg agctggatgc cgccgtgggc 2520
accctgatga cagcaatcgg cgacctgggc ctgctggagg agacactgg catcttcacc 2580

```

gccgataacg gccctgagac aatgcggatg tctagaggcg gatgcagcgg cctgctgaga 2640
 tgtggcaagg gaaccacata cgagggaggc gtgcgcgagc ctgccctggc attttggcca 2700
 ggacacatcg cacctggagt gaccacagag ctggcctcct ctctggacct gctgccaaca 2760
 ctggccgccc tggcaggagc acctctgcca aatgtgacct tggacggctt cgatctgagc 2820
 ccactgctgc tgggaaccgg caagtcccct aggagctctc tgttctttta cccctcctat 2880
 cctgatgagg tgcggggcgt gtttgccgtg agaaccggca agtacaaggc ccacttcttt 2940
 acacagggtc ctgcccacag cgacaccaca gcagatccag catgccacgc cagctcctct 3000
 ctgaccgcac acgagccacc tctgctgtac gacctgtcca aggatcccgg cgagaactat 3060
 aatctgctgg gaggagtggc aggagcaacc cctgaggtgc tgcaggccct gaagcagctg 3120
 cagctgctga aggcacagct ggacgcagca gtgacattcg gccaagcca ggtggccaga 3180
 ggcgaggatc ccgcccctga gatctgttgc caccgccgtt gcacccaag acctgcctgt 3240
 tgccattgcc ccgaccaca cgctaagat tctagagtcg agccgcggac tagtaacttg 3300
 ttattgcag ctataatgg ttacaaataa agcaatagca tcacaaattt cacaaataa 3360
 gcattttttt cactgcattc tagttgtggt ttgtccaaac tcataatgt atctta 3416

<210> 47

<211> 3949

<212> DNA

<213> 人工序列

<220>

<223> 合成多核苷酸

<400> 47

ttggccactc cctctctgcg cgctcgctcg ctactgagg ccgggcgacc aaaggtcgcc 60
 cgacgcccgg gctttgcccg ggcgccctca gtgagcgagc gagcgcgagc agagggagtg 120
 gccaactcca tcactagggg ttcttgagg ggtggagtcg tgacgtgaat tacgtcatag 180
 ggtaggggag gtcctgcaga tcttcaatat tggccattag ccatattatt cattggttat 240
 atagcataaa tcaatatagg ctattggcca ttgcatacgt tgtatctata tcataatatg 300
 tacatttata ttggctcatg tccaatatga ccgccatgtt ggcatgatt attgactagt 360
 tattaatagt aatcaattac ggggtcatta gttcatagcc catatatgga gttccgcgtt 420
 acataactta cggtaaattg ccgcctggc tgaccgccc acgacccccg ccattgacg 480
 tcaataatga cgtatgttcc catagtaacg ccaataggga ctttccattg acgtcaatgg 540
 gtggagtatt tacggtaaac tgcccacttg gcagtacatc aagtgtatca tatgccaagt 600
 ccgcccccta ttgacgtcaa tgacggtaaa tggccgcctt ggcatatgc ccagtacatg 660
 accttacggg actttcctac ttggcagtac atctacgtat tagtcatcgc tattaccatg 720
 gtcgaggtga gccccacgtt ctgttcaact ctccccatct cccccccctc cccaccccca 780
 attttgtatt tatttatttt ttaattattt tgtgcagcga tgggggcggg gggggggggg 840
 gggcgcgcg caggcggggc ggggcggggc gaggggcggg gcggggcgag gcggagaggt 900
 gcggcggcag ccaatcagag cggcgcgctc cgaaagtctt cttttatggc gaggcggcgg 960
 cggcgggcgg cctataaaaa gcgaagcgcg cggcgggcgg gagtcgctgc gcgctgcctt 1020
 cgccccgtgc cccgctccgc cgccgcctcg cgccgcccgc cccggtcttg actgaccgcg 1080

```

ttactccac aggtgagcgg gcgggacggc cttctcctc cgggctgtaa ttagcgcttg 1140
gtttaatgac ggcttgttc tttctgtgg ctgcgtgaaa gccttgaggg gctccgggag 1200
ggccctttgt gcggggggag cggctcgggg ggtgcgtgcg tgtgtgtgtg cgtggggagc 1260
gccgcgtgcg gctccgcgct gccggcggcg tgtgagcgtc gcgggcgcgg cgcggggctt 1320
tgtgcgtcc gcagtgtgcg cgaggggagc gcggccgggg gcggtgcccc gcggtgcggg 1380
gggggctgcg aggggaacaa aggctgcgtg cggggtgtgt gcgtgggggg gtgagcaggg 1440
ggtgtgggcg cgtcggtcgg gctgcaacct cccctgcacc cccctccccg agttgctgag 1500
cacggccccg cttcgggtgc ggggctccgt acggggcgtg gcgcggggct cgccgtgccg 1560
ggcggggggg ggcggcaggt gggggtgccg ggcggggcgg ggccgcctcg ggccggggag 1620
ggctcggggg aggggcgcgg cggcccccg agcgccggcg gctgtcagag cgcggcgagc 1680
cgcagccatt gccttttatg gtaatcgtgc gagagggcgc agggacttcc tttgtcccaa 1740
atctgtgcgg agccgaaatc tgggaggcgc cgccgcacct cctctagcgg gcgcggggcg 1800
aagcgggtgc gcgccgcag gaaggaaatg ggcggggagg gccttcgtgc gtcgccgcgc 1860
cgccgtcccc ttctccctct ccagcctcgg ggctgtccgc ggggggacgg ctgccttcgg 1920
gggggacggg gcagggcggg gttcggttc tggcgtgtga ccggcggtc tagagcctct 1980
gctaaccatg ttcatgcctt cttctttttc ctacagctcc tgggcaacgt gctggttatt 2040
gtgctgtctc atcattttgg caaagaattc cgccaccatg tccatggggg caccgcggtc 2100
cctcctcctg gccctggctg ctggcctggc cgttgcccg cgcaccaaca tcgtgctgat 2160
ctttgccgac gacctcggct atggggacct gggctgctat gggcaccca gctctaccac 2220
tcccaacctg gaccagctgg cggcgggagg gctgcggttc acagacttct acgtgcctgt 2280
gtctctgtgc acacctcta gggccgcct cctgaccggc cggtccccg ttcgatggg 2340
catgtaccct ggcgctctgg tgcccagctc ccgggggggc ctgcccctgg aggaggtgac 2400
cgtggccgaa gtctggctg cccgaggcta cctcacagga atggccggca agtggcacct 2460
tggggtgggg cctgaggggg cttcctgcc ccccatcag ggcttccatc gatttctagg 2520
catcccgtag tcccacgacc agggccccct ccagaacctg acctgcttcc cgccggccac 2580
tccttgcgac ggtggctgtg accagggcct ggtccccatc cactgttg ccaacctgtc 2640
cgtggaggcg cagccccct ggctgcccg actagaggcc cgctacatgg ctttcgcca 2700
tgacctcatg gccgacgccc agcgccagga tcgccccctt ttctgtact atgcctctca 2760
ccacaccac tacctcagt tcagtgggca gagctttgca gagcgttcag gcccggggc 2820
atttggggac tccctgatgg agctggatgc agctgtgggg accctgatga cagccatagg 2880
ggacctgggg ctgcttgaag agacgtggt catcttact gcagacaatg gacctgagac 2940
catgcgtatg tcccgaggcg gctgtccgg tctcttgagg tgtggaaagg gaacgacct 3000
cgagggcggt gtccgagagc ctgccttggc cttctggcca ggtcatatcg ctcccgcgct 3060
gaccacagag ctggccagct cctggacct gctgcctacc ctggcagccc tggctggggc 3120
cccactgccc aatgtcacct tggatggctt tgacctcagc cccctgctgc tgggcacagg 3180
caagagccct cggcagtctc tctttctta cccgtctac ccagacgagg tccgtgggg 3240
ttttgctgtg cggactggaa agtacaaggc tcaattctc acccagggt ctgcccacag 3300
tgataccact gcagacctg cctgccagc ctccagctct ctgactgtc atgagcccc 3360
gctgtcttat gacctgtcca aggaccctg tgagaactac aacctgctg ggggtgtggc 3420

```

cggggccacc ccagaggtgc tgcaagccct gaaacagctt cagctgctca aggcccagtt 3480
 agacgcagct gtgaccttcg gccccagcca ggtggccccg ggcgaggacc ccgccctgca 3540
 gatctgctgt catcctggct gcaccccccg cccagcttgc tgccattgcc cagatcccca 3600
 tgcctgagat tctagagtcg agccgcggac tagtaacttg tttattgcag cttataatgg 3660
 ttacaaataa agcaatagca tcacaaatth cacaaataaa gcattttttt cactgcattc 3720
 tagttgtggt ttgtccaaac tcatcaatgt atcttaggtc tagatacgta gataagtagc 3780
 atggcggggt aatcattaac tacaaggaac ccctagtgat ggagttggcc actccctctc 3840
 tgcgcgctcg ctcgctcact gaggccgggc gaccaaaggt cgcccagacg ccgggctttg 3900
 cccgggcggc ctcagtgagc gagcgagcgc gcagagaggg agtggccaa 3949

<210> 48

<211> 3949

<212> DNA

<213> 人工序列

<220>

<223> 合成多核苷酸

<400> 48

ttggccactc cctctctgcg cgctcgctcg ctactgagg ccgggcgacc aaaggtcgcc 60
 cgacgccccg gctttgcccc ggcggcctca gtgagcgagc gagcgcgag agaggagtg 120
 gccaaactca tcactagggg ttcctggagg ggtggagtcg tgacgtgaat tacgtcatag 180
 ggtaggggag gtccctgcaga tcttcaatat tggccattag ccatattatt cattggttat 240
 atagcataaa tcaatatagg ctattggcca ttgcatacgt tgtatctata tcataatatg 300
 tacatttata ttggctcatg tccaatatga ccgcatgtt ggcatgatt attgactagt 360
 tattaatagt aatcaattac ggggtcatta gttcatagcc catatatgga gttccgcgtt 420
 acataactta cggtaaatgg cccgcctggc tgaccgcca acgacccccg cccattgacg 480
 tcaataatga cgtatgttcc catagtaacg ccaatagga ctttcattg acgtcaatgg 540
 gtggagtatt tacggtaaac tgcccacttg gcagtacatc aagtgtatca tatgccaagt 600
 ccgcccccta ttgacgtcaa tgacggtaaa tggcccgctt ggcatatgc ccagtacatg 660
 accttacggg actttcctac ttggcagtac atctacgtat tagtcatcgc tattaccatg 720
 gtcgaggtag gccccacgtt ctgcttcaact ctccccatct cccccccctc cccaccccca 780
 attttgtatt tattttatth ttaattatth tgtgcagcga tgggggcggg gggggggggg 840
 gggcgcgcg caggcggggc ggggcggggc gaggggcggg gcggggcgag gcggagaggt 900
 gcggcggcag ccaatcagag cggcgcgctc cgaaagtth cttttatggc gaggcggcgg 960
 cggcggcggc cctataaaaa gcgaagcgcg cggcgggcgg gagtcgctgc gcgctgcctt 1020
 cgccccgtgc cccgctccgc cgccgctcgc cgccgcccgc cccggtctg actgaccgcg 1080
 ttactccac aggtgagcgg gcgggacggc ctttctctc cgggctgtaa ttagcgcttg 1140
 gtttaatgac ggcttgtht ttttctgttg ctgcgtgaaa gccttgaggg gctccgggag 1200
 ggccctttgt gcggggggag cggctcgggg ggtgcgtgcg tgtgtgtgtg cgtggggagc 1260
 gccgcgtgcg gctccgcgct gcccgcggcg tgtgagcgtc gcgggcgcgg cgcggggctt 1320
 tgtgcgctcc gcagtgtgcg cgaggggagc gcggccgggg gcggtgcccc gcggtgcggg 1380

```

gggggctgcg aggggaacaa aggctgcgtg cggggtgtgt gcgtgggggg gtgagcaggg 1440
ggtgtgggcg cgtcggtcgg gctgcaaccc cccctgcacc cccctccccg agttgctgag 1500
cacggccccg cttcgggtgc ggggctccgt acggggcgtg gcgcggggct cgccgtgccg 1560
ggcggggggg ggcggcaggt gggggtgccg ggcggggcgg ggccgcctcg ggccggggag 1620
ggctcggggg aggggcgcgg cggcccccg agcgccggcg gctgtcgagg cgcggcgagc 1680
cgcagccatt gccttttatg gtaatcgtgc gagaggcgcg agggacttcc tttgtccaa 1740
atctgtgcgg agccgaaatc tgggaggcgc cgccgcaccc cctctagcgg gcgcggggcg 1800
aagcgggtgc gcgccggcag gaaggaaatg ggcggggagg gccttcgtgc gtcgccgcgc 1860
cgccgtcccc ttctccctct ccagcctcgg ggctgtccgc ggggggacgg ctgccttcgg 1920
gggggacggg gcagggcggg gttcggcttc tggcgtgtga cggcgggctc tagagcctct 1980
gctaaccatg ttcatgcctt cttctttttc ctacagctcc tgggcaacgt gctggttatt 2040
gtgctgtctc atcattttgg caaagaattc cgccaccatg tctatggggg ctctctgctc 2100
cctgtgtctg gcactggccg ccgggctggc tgtcgcaaga ccacctaata tcgtcctgat 2160
ttttgcagac gatctgggat acggcgacct gggatgctat ggccaccaa gctccaccac 2220
accaacctg gaccagctgg cagcaggagg cctgcggttc accgacttct acgtgccagt 2280
gagcctgtgc accccctcca gagccgcctt gctgacaggc aggctgccag tgcgcatggg 2340
catgtatcct ggctgtctgg tgccatctag caggggcggc ctgccactgg aggagtgac 2400
cgtggcagag gtgctggcag ccagaggcta cctgacagga atggccggca agtggcacct 2460
gggagtggga ccagagggag ccttcctgcc ccctcaccag ggcttcacc ggtttctggg 2520
catcccttat tctcacgacc agggcccatg ccagaacctg acctgttttc caccagcaac 2580
accatgcgac ggaggatgtg atcagggcct ggtgccaatc ccaactgctgg caaatctgag 2640
cgtggaggca cagcctccat ggctgcctgg cctggaggca agatacatgg ccttcgcca 2700
cgacctgatg gcagatgcac agcggcagga tagaccttc tttctgtact atgcctcca 2760
ccacaccac tatccacagt tcagcgcca gtcctttgcc gagaggtccg gaaggggacc 2820
attcggcgac tctctgatgg agctggatgc cgccgtgggc accctgatga cagcaatcgg 2880
cgacctgggc ctgctggagg agacactggt catcttcacc gccgataacg gccctgagac 2940
aatgcggatg tctagaggcg gatgcagcgg cctgctgaga tgtggcaagg gaaccacata 3000
cgagggaggc gtgcgcgagc ctgccctggc attttgcca ggacacatcg cacctggagt 3060
gacccacgag ctggcctcct ctctggacct gctgccaaca ctggccgccc tggcaggagc 3120
acctctgcca aatgtgacct tggacggctt cgatctgagc ccaactgctgc tgggaaccgg 3180
caagtcccct aggcagtctc tgttctttta cccctctat cctgatgagg tgcggggcgt 3240
gtttgccgtg agaaccggca agtacaaggc ccacttcttt acacagggt ctgcccacag 3300
cgacaccaca gcagatccag catgccacgc cagctcctct ctgaccgcac acgagccacc 3360
tctgtgttac gacctgtcca aggatcccgg cgagaactat aatctgctgg gaggagtggc 3420
aggagcaacc cctgaggtgc tgcaggccct gaagcagctg cagctgctga aggcacagct 3480
ggacgcagca gtgacattcg gcccgaagca ggtggccaga ggcgaggatc ccgccctgca 3540
gatctgttgc caccgccgt gcacccaag acctgcctgt tgccattgcc ccgaccaca 3600
cgcctaagat tctagagtcg agccgcggac tagtaacttg tttattgcag cttataatgg 3660
ttacaaataa agcaatagca tcacaaattt cacaaataaa gcattttttt cactgcattc 3720

```

tagttgtggt ttgtccaaac tcatcaatgt atcttaggtc tagatacgta gataagtagc 3780
 atggcggggtt aatcattaac tacaaggaaac ccctagtgtat ggagttggcc actccctctc 3840
 tgcgcgctcg ctcgctcact gaggccgggc gaccaaaggt cgcgcgacgc ccgggctttg 3900
 cccgggcggc ctcagtgagc gagcgagcgc gcagagaggg agtggccaa 3949

<210> 49

<211> 4500

<212> DNA

<213> 人工序列

<220>

<223> 合成多核苷酸

<400> 49

ttggccactc cctctctgcg cgtcgtctcg ctactgagg ccgggcgacc aaaggtcgcc 60
 cgacgcccgg gctttgcccc ggcgccctca gtgagcgagc gagcgcgcag agagggagtg 120
 gccaaactcca tcactagggg ttcttgagg ggtggagtgc tgacgtgaat tacgtcatag 180
 ggtaggggag gtctgcata tgcggccgcg atcttcaata ttggccatta gccatattat 240
 tcattgggta tatagcataa atcaatattg gctattggcc attgcatacg ttgtatctat 300
 atcataatat gtacatttat attggctcat gtccaatatg accgccatgt tggcattgat 360
 tattgactag ttattaatag taatcaatta cggggtcatt agttcatagc ccatatatgg 420
 agttccgctg tacataactt acggtaaatg gccgcctgg ctgaccgccc aacgaccccc 480
 gccattgac gtcaataatg acgtatgttc ccatagtaac gccaataggg actttccatt 540
 gacgtcaatg ggtggagtat ttacggtaaa ctgccactt ggagtagcat caagtgtatc 600
 atatgccaaag tccgccccct attgacgtca atgacggtaa atggcccgcc tggcattatg 660
 ccagtagcat gacctacgg gactttccta cttggcagta catctacgta ttagtcatcg 720
 ctattaccat ggtagagggt agccccacgt tctgcttcac tctccccatc tccccccct 780
 cccaccccc aattttgtat ttatttattt tttaattatt ttgtgcagcg atgggggcgg 840
 gggggggggg ggggcgcgcg ccaggcgggg cggggcgggg cgaggggcgg ggcggggcga 900
 ggcggagagg tgcggcggca gccaatcaga gcggcgcgct ccgaaagttt ctttttatgg 960
 cgaggcggcg gcggcggcgg ccctataaaa agcgaagcgc gcggcggcg ggagtcgctg 1020
 cgcgctgcct tcgccccgtg ccccgctccg ccgcccctc gcgcccggcg ccccggtct 1080
 gactgaccgc gttactccca caggtgagcg ggcgggacgg cccttctcct ccgggctgta 1140
 attagcgctt ggtttaatga cggttggtt cttttctgtg gctgcgtgaa agccttgagg 1200
 ggctccggga gggccctttg tgcgggggga gcggctcggg ggggtgcgtgc gtgtgtgtgt 1260
 gcgtggggag cgccgcgtgc ggctccgcgc tgcccggcgg ctgtgagcgc tgcgggcgcg 1320
 gcgcggggct ttgtgcgtc cgagtggtgc gcgaggggag cgcgccggg ggcggtgccc 1380
 cgcggtgcgg ggggggctgc gaggggaaca aaggctgcgt gcggggtgtg tgcgtggggg 1440
 ggtgagcagg ggggtgtggg gcgtcggtcg ggctgcaacc cccctgcac cccctcccc 1500
 gagttgctga gcacggcccc gcttcgggtg cggggctccg tacggggcgt ggcgcggggc 1560
 tcgccgtgcc gggcgggggg tggcggcagg tgggggtgcc gggcggggcg gggccgcctc 1620
 gggccgggga gggctcgggg gaggggcgcg gcggccccc gagcgccggc ggctgtcgag 1680

gcgcggcgag ccgcagccat tgccttttat ggtaatcgtg cgagagggcg cagggacttc 1740
 ctttgtccca aatctgtgcg gagccgaaat ctgggaggcg ccgccgcacc ccctctagcg 1800
 ggcgcggggc gaagcgggtgc ggcgccggca ggaaggaaat gggcggggag ggccttcgtg 1860
 cgtcgcccgcg ccgccgtccc ctctccctc tccagcctcg gggctgtccg cggggggacg 1920
 gctgccttcg ggggggacgg ggcagggcgg ggttcggctt ctggcgtgtg accggcggct 1980
 ctagagcctc tgctaaccat gttcatgcct tcttcttttt cctacagctc ctgggcaacg 2040
 tgctgggttat tgtgctgtct catcattttg gcaaagaatt ccgccaccat gtctatgggg 2100
 gtcctcgtc ccctgctgct ggcactggcc gccgggctgg ctgtcgcaag accacctaata 2160
 atcgtcctga tttttgcaga cgatctggga tacggcgacc tgggatgcta tggccaccca 2220
 agctccacca caccacacct ggaccagctg gcagcaggag gcctgcggtt caccgacttc 2280
 tacgtgccag tgagcctgtg caccctctcc agagccgcc tgctgacagg caggctgcc 2340
 gtgcgcatgg gcatgtatcc tggcgtgctg gtgccatcta gcaggggcgg cctgccactg 2400
 gaggaggtga ccgtggcaga ggtgctggca gccagaggct acctgacagg aatggccggc 2460
 aagtggcacc tgggagtggg accagaggga gccttctgc cccctacca gggtctccac 2520
 cggtttctgg gcatccctta ttctcacgac cagggcccat gccagaacct gacctgtttt 2580
 ccaccagcaa caccatgcga cggaggatgt gatcagggcc tggtgccaat cccactgctg 2640
 gcaaatctga gcgtggaggc acagcctcca tggctgcctg gcctggaggc aagatacatg 2700
 gccttcgccc acgacctgat ggcagatgca cagcggcagg atagaccttt ctttctgtac 2760
 tatgcctccc accacaccca ctatccacag ttcagcggcc agtcctttgc cgagaggtcc 2820
 ggaaggggac cattcggcga ctctctgatg gagctggatg ccgccgtggg caccctgatg 2880
 acagcaatcg gcgacctggg cctgctggag gagacactgg tcattcttcac cgccgataac 2940
 ggccctgaga caatgcggat gtctagaggc ggatgcagcg gcctgctgag atgtggcaag 3000
 ggaaccacat acgagggagg cgtgcgcgag cctgccttgg cattttggcc aggacacatc 3060
 gcacctggag tgaccacga gctggcctcc tctctggacc tgctgccaac actggccgcc 3120
 ctggcaggag cacctctgcc aaatgtgacc ctggacggct tcgatctgag cccactgctg 3180
 ctgggaaccg gcaagtcccc taggcagtct ctgttctttt accctccta tcctgatgag 3240
 gtgcggggcg tgtttgccgt gagaaccggc aagtacaagg cccacttctt tacacagggc 3300
 tctgcccaca gcgacaccac agcagatcca gcatgccacg ccagctctc tctgaccga 3360
 cacgagccac ctctgctgta cgacctgtcc aaggatcccc gcgagaacta taatctgctg 3420
 ggaggagtgg caggagcaac ccctgagggtg ctgcaggccc tgaagcagct gcagctgctg 3480
 aaggcacagc tggacgcagc agtgacattc ggcccaagcc aggtggccag aggcgaggat 3540
 cccgccctgc agatctgttg ccacccggc tgcacccaa gacctgcctg ttgccattgc 3600
 cccgaccac acgcctaaga ttctagagtc gagccgcgga ctagtaactt gtttattgca 3660
 gcttataatg gttacaaata aagcaatagc atcacaatt tcacaaataa agcatttttt 3720
 tcaactgcatt ctagttgtgg tttgtccaaa ctcatcaatg tatcttacct gcaggtcac 3780
 cagtgtttgt gactgggaac tctccctgcc aaatattggc ataatgctgt ctttaggtt 3840
 gcagcttatt gcccagggg aacagtctgt tgtgcagtc accccggcag gaatactccc 3900
 attctgcctc tgttggtaac ctttcccag cccaggtgca gtatgccact gcatcattcc 3960
 aggacacatg cagcacaggg tggtcaggcc tgtggaggat agttgagtct ggtccctctg 4020

ggtgtctcca attggctcct ttaacaggca gccaccaggg ggctgcagcc actgcctggt 4080
 ggatattggg cttcacctgc tcacttagca tgccttcaaa aacaaaactg tcaccaaatt 4140
 tctcagcctc tgtaagggtat ccagtgtgtt ttacaaattt ctcaaattct gtgtttgaca 4200
 cttcataggc atccatatag aaggcatcaa ttgtgactct cctagctggg gcttcacat 4260
 cctgcttgat ctgagggtca tcagttccca tagtaaaaac tcctgcaggt ctagatacgt 4320
 agataagtag catggcgggt taatcattaa ctacaaggaa cccctagtga tggagttggc 4380
 cactccctct ctgcgcgtc gctcgtcac tgaggccggg cgaccaaagg tcgcccgcg 4440
 cccgggcttt gcccgggcgg cctcagttag cgagcgagcg cgcagagagg gaggggc 4500

<210> 50

<211> 6612

<212> DNA

<213> 人工序列

<220>

<223> 合成多核苷酸

<400> 50

cgccagggtt ttcccagtca cgacgttgta aaacgacggc cagtgccaaag cttgcatgcc 60
 tgcatttggc cactccctct ctgcgcgtc gctcgtcac tgaggccggg cgaccaaagg 120
 tcgcccgcg cccgggcttt gcccgggcgg cctcagttag cgagcgagcg cgcagagagg 180
 gaggggc 240
 cataggggta gggagggtcct gcagatcttc aatattggcc attagccata ttattcattg 300
 gttatatagc ataaatcaat attggctatt ggccattgca tacgttgtat ctatatcata 360
 atatgtacat ttatatggc tcatgtccaa tatgaccgcc atgttggcat tgattattga 420
 ctagttatta atagtaatca attacggggg cattagttca tagcccatat atggagttcc 480
 gcgttacata acttacggta aatggcccg cttggctgacc gccaacgac cccgcccac 540
 tgacgtcaat aatgacgtat gttcccatag taacgccaat agggactttc cattgacgtc 600
 aatgggtgga gtatttacgg taaactgccc acttggcagt acatcaagtg tatcatatgc 660
 caagtccgcc ccctattgac gtcaatgacg gtaaatggcc cgcctggcat tatgccaggt 720
 acatgacctt acgggacttt cctacttggc agtacatcta cgtattagtc atcgctatta 780
 ccatggctga ggtgagcccc acgttctgct tcaactctcc catctcccc cctccccac 840
 cccaatttt gtatttattt attttttaat tttttgtgc agcgatggg gcggggggg 900
 ggggggggcg cgcgccaggc ggggcggggc ggggcgaggg gcggggcggg gcgaggcgga 960
 gagggtcggc ggcagccaat cagagcggcg cgctccgaaa gtttcctttt atggcgaggc 1020
 ggcggcggcg gcggccctat aaaaagcgaa gcgcgcggcg ggcgggagtc gctgcgcgt 1080
 gccttcgccc cgtgccccgc tccgcccgc ctcgcgcgc cccgccccgc ctctgactga 1140
 ccgcgttact cccacaggtg agcgggcggg acggcccttc tcctccgggc tgtaattagc 1200
 gcttgggtta atgacggctt gtttcttttc tgtggctgcg tgaaagcctt gaggggctcc 1260
 gggaggggccc tttgtgcggg gggagcggct cgggggtgc gtgcgtgtgt gtgtgcgtgg 1320
 ggagcgccgc gtgcggctcc gcgctgccc gcggctgtga gcgctgcgg gcggcgcgg 1380
 ggctttgtgc gctccgcagt gtgcgcgagg ggagcgcggc cgggggcggg gcccgcgg 1440

```

gcgggggggg ctgcgagggg aacaaaggct gcgtgcgggg tgtgtgcgtg ggggggtgag 1500
caggggggtgt gggcgcgctcg gtcgggctgc aacccccctt gcacccccct ccccgagttg 1560
ctgagcacgg cccggcttcg ggtgcggggc tccgtacggg gcgtggcgcg gggctcgccg 1620
tgccggggcgg ggggtggcgg caggtggggg tgccggggcgg ggcggggccc cctcggggccg 1680
gggagggctc gggggagggg cgcggcgggc cccggagcgc cggcggctgt cgaggcgcg 1740
cgagccgcag ccattgcctt ttatggtaat cgtgcgagag ggcgaggga cttcctttgt 1800
cccaaattctg tgcggagccg aaatctggga ggcgccggcg caccctctct agcgggcgcg 1860
gggcgaagcg gtgcggcgcc ggcaggaagg aaatgggcgg ggagggcctt cgtgcgtcgc 1920
cgcgccggcg tccccttctc cctctccagc ctcggggctg tccgcggggg gacggctgcc 1980
ttcggggggg acggggcagg gcgggggttcg gcttctggcg tgtgaccggc ggctctagag 2040
cctctgctaa ccatgttcat gccttcttct ttttctaca gctcctgggc aacgtgctgg 2100
ttattgtgct gtctcatcat tttggcaaag aattccgcca ccatgtccat gggggcaccg 2160
cggtccttcc tcctggccct ggetgctggc ctggccgttgc cccgtccgcc caacatcgtg 2220
ctgatctttg ccgacgacct cggctatggg gacctgggct gctatgggca cccagctct 2280
accactccca acctggacca gctggcgggc ggagggtgc ggttcacaga cttctacgtg 2340
cctgtgtctc tgtcacacc ctctagggc gccctcctga ccggccggct cccggttcgg 2400
atgggcatgt accctggcgt cctggtgccc agtccccggg ggggcctgcc cctggaggag 2460
gtgaccgtgg ccgaagtcct ggctgcccga ggctacctca caggaatggc cggcaagtgg 2520
caccttgggg tggggcctga gggggccttc ctgcccccc atcagggtt ccatcgattt 2580
ctaggcatcc cgtactccca cgaccagggc ccctgccaga acctgacctg cttcccgccg 2640
gccactcctt gcgacggtgg ctgtgaccag ggctgtgtcc ccatccact gttggccaac 2700
ctgtccgtgg aggcgcagcc cccctggctg cccggactag agggccgcta catggcttct 2760
gcccattgacc tcatggccga cgcccagcgc caggatcgcc ctttcttct gtactatgcc 2820
tctcaccaca ccactaccc tcagttcagt gggcagagct ttgcagagcg ttcaggccgc 2880
gggccatttg gggactccct gatggagctg gatgcagctg tggggaccct gatgacagcc 2940
ataggggacc tggggctgct tgaagagacg ctggtcattt tctactgaga caatggacct 3000
gagaccatgc gtatgtcccg aggcggctgc tccggtctct tgcggtgtgg aaagggaacg 3060
acctacgagg gcggtgtccg agagcctgcc ttggccttct ggccaggtea tatcgctccc 3120
ggcgtgacct acgagctggc cagctccctg gacctgctgc ctaccctggc agccctggct 3180
ggggccccac tgcccaatgt caccttggat ggctttgacc tcagccccct gctgctgggc 3240
acaggcaaga gccctcggca gtctctcttc ttctaccgt cctaccaga cgaggtccgt 3300
ggggtttttg ctgtgcggac tggaaagtac aaggtcact tcttcacca gggctctgcc 3360
cacagtgata cactgcaga cctgctgc cacgctcca gctctctgac tgctcatgag 3420
cccccgctgc tctatgacct gtccaaggac cctggtgaga actacaacct gctggggggt 3480
gtggccgggg ccaccccaga ggtgctgcaa gccctgaaac agcttcagct gctcaaggcc 3540
cagttagacg cagctgtgac cttcggcccc agccaggtag cccggggcga ggaccccgcc 3600
ctgcagatct gctgtcatcc tggtgcacc cccgcccag cttgctgcca ttgcccagat 3660
cccatgcct gagattctag agtcgagccg cggactagta acttgtttat tgcagcttat 3720
aatggttaca aataaagcaa tagcatcaca aatttcacaa ataaagcatt tttttcactg 3780

```

cattctagtt gtggtttgtc caaactcatc aatgtatctt aggtctagat acgtagataa 3840
gtagcatggc gggftaatca ttaactacaa ggaaccccta gtgatggagt tggccactcc 3900
ctctctgcgc gctcgcctgc tctactgaggc cgggcgacca aaggtcgccc gacgcccggg 3960
ctttgcccgg gcggcctcag tgagcgagcg agcgcgcaga gagggagtgg ccaaagatcc 4020
ccgggtaccg agctcgaatt cgtaatcatg tcatagctgt ttcctgtgtg aaattgttat 4080
ccgctcaciaa ttccacaciaa catacgagcc ggaagcataa agtgtaaagc ctgggggtgcc 4140
taatgagtga gctaactcac attaatgtcg ttgcgctcac tgcccgttt ccagtcggga 4200
aacctgtcgt gccagctgca ttaatgaatc ggccaacgcg cggggagagg cggtttgcgt 4260
attggcgaac ttttgcctgag ttgaaggatc agatcacgca tcttcccagc aacgcagacc 4320
gttccgtggc aaagcaaaaag ttcaaaaatca gtaaccgtca gtgccgataa gttcaaagtt 4380
aaacctggtg ttgataccaa cattgaaacg ctgatcgaaa acgcgctgaa aaacgctgct 4440
gaatgtgcga gcttcttccg cttcctcgtc cactgactcg ctgcgctcgg tcgttcggct 4500
gcggcgagcg gtatcagctc actcaaaggc ggtaatacgg ttatccacag aatcagggga 4560
taacgcagga aagaacatgt gagcaaaagg ccagcaaaag gccaggaacc gtaaaaaggc 4620
cgcgttgctg gcgtttttcc ataggtccg cccccctgac gagcatcaca aaaatcgacg 4680
ctcaagtcag aggtggcgaa acccgacagg actataaaga taccaggcgt tccccctgg 4740
aagctccctc gtgcgctctc ctgttccgac cctgccgctt accgatacc tgtccgcctt 4800
tctcccttcg ggaagcgtgg cgcttttctc atgctcacgc tntaggtatc tcagttcgg 4860
gtaggtcggt cgctccaagc tgggctgtgt gcacgaacct cccgttcagc ccgaccgctg 4920
cgcttatcc ggtaactatc gtcttgagtc caaccggta agacacgact tatcgccact 4980
ggcagcagcc actggtaaca ggattagcag agcgaggtat gtaggcggtg ctacagagtt 5040
cttgaagtgg tggcctaact acggctacac tagaaggaca gtatttggtg tctgcgctct 5100
gctgaagcca gttaccttcg gaaaaagagt tggtagctct tgatccggca aacaaaccac 5160
cgctggtagc ggtggttttt ttgtttgcaa gcagcagatt acgcgcagaa aaaaaggatc 5220
tcaagaagat cctttgatct tttctacggg gtctgacgct cagtggaacg atccgtcgag 5280
aggtctgcct cgtgaagaag gtgttgctga ctcataccag gcctgaatcg ccccatcatc 5340
cagccagaaa gtgagggagc cacggttgat gagagctttg ttgtaggtgg accagttggt 5400
gattttgaac ttttgccttg ccacggaacg gtctgcgttg tcgggaagat gcgtgatctg 5460
atccttcaac tcagcaaaaag ttcgatttat tcaacaaagc cacgttgtgt ctcaaaatct 5520
ctgatgttac attgcacaag ataaaaatat atcatcatga acaataaaac tgtctgctta 5580
cataaacagt aatacaaggg gtgttatgag ccatattcaa cgggaaacgt cttgctcgaa 5640
gccgcgatta aattccaaca tggatgctga tttatatggg tataaatggg ctgcgcgataa 5700
tgtcgggcaa tcaggtgcga caatctatcg attgtatggg aagcccgatg cgccagagtt 5760
gtttctgaaa catggcaaaag gtagcgttgc caatgatgtt acagatgaga tggtcagact 5820
aaactggctg acggaattta tgcctcttcc gaccatcaag cattttatcc gtactcctga 5880
tgatgcattg ttactacca ctgcgatccc cgggaaaaca gcattccagg tattagaaga 5940
atatcctgat tcaggtgaaa atattgttga tgcgctggca gtgttctgc gccggttgca 6000
ttcgattcct gtttgaatt gtccttttaa cagcgatcgc gtatttcgtc tcgctcaggc 6060
gcaatcacga atgaataacg gtttggttga tgcgagtgt tttgatgacg agcgtaatgg 6120

ctggcctgtt gaacaagtct ggaaagaaat gcataagctt ttgccattct caccggattc 6180
 agtcgtcact catggtgatt tctcacttga taaccttatt ttgacgagg ggaaattaat 6240
 aggttgtatt gatgttggac gagtcggaat cgcagaccga taccaggatc ttgccatcct 6300
 atggaactgc ctcggtgagt tttctccttc attacagaaa cggctttttc aaaaatatgg 6360
 tattgataat cctgatatga ataaattgca gtttcatttg atgctcgatg agtttttcta 6420
 atcagaattg gttaattggt tgtaacactg gcagagcatt acgctgactt gacgggacgg 6480
 cggctttgtt gaataaatcg cattcgccat tcaggctgcg caactgttgg gaaggcgat 6540
 cgggtcgggc ctcttcgcta ttacgccagc tggcgaaagg gggatgtgct gcaaggcgat 6600
 taagttgggt aa 6612

<210> 51

<211> 5792

<212> DNA

<213> 人工序列

<220>

<223> 合成多核苷酸

<400> 51

gtcagggtggc acttttcggg gaaatgtggc atgcctgcat ttggccactc cctctctgcg 60
 cgctcgtcgc ctactgagg ccgggcgacc aaaggtcgcc cgacgcccgg gctttgccc 120
 ggcggcctca gtgagcgagc gagcgcgagc agagggagtg gccaaactcca tctactagggg 180
 ttcttgaggg ggtggagtcg tgacgtgaat tacgtcatag ggtagggag gtcctgcaga 240
 tcttcaatat tggccattag ccatattatt cattggttat atagcataaa tcaatattgg 300
 ctattggcca ttgcatacgt tgtatctata tcataatatg tacatttata ttggctcatg 360
 tccaatatga ccgccatgtt ggcatgtgatt attgactagt tattaatagt aatcaattac 420
 ggggtcatta gttcatagcc catatatgga gttccgcgtt acataactta cggtaaatgg 480
 cccgcctggc tgaccgccc acgacccccg ccattgacg tcaataatga cgtatgttcc 540
 catagtaacg ccaataggga ctttcattg acgtcaatgg gtggagtatt tacggtaaac 600
 tgcccacttg gcagtacatc aagtgtatca tatgccaagt ccgcccccta ttgacgtcaa 660
 tgacggtaaa tggcccgctt ggcatatgc ccagtacatg accttacggg actttcctac 720
 ttggcagtac atctacgtat tagtcatcgc tattaccatg gtcgaggtga gccccacgtt 780
 ctgcttcaact ctccccatct cccccctc cccacccca attttgtatt tatttttttt 840
 ttaattattt tgtgcagcga tgggggcggg gggggggggg gggcgcgcg caggcggggc 900
 ggggcggggc gaggggcggg gcggggcgag gcggagaggt gcggcggcag ccaatcagag 960
 cggcgcgctc cgaaagtctt cttttatggc gaggcggcgg cggcggcggc cctataaaaa 1020
 gcgaagcgcg cggcgggcgg gagtcgctgc gcgctgcctt cgccccgtgc cccgctccgc 1080
 cgccgcctcg cgccgcccgc cccggctctg actgaccgcg ttactccac aggtgagcgg 1140
 gcgggacggc ccttctcctc cgggctgtaa ttagcgcttg gtttaatgac ggcttgtttc 1200
 ttttctgtgg ctgcgtgaaa gccttgaggg gctccgggag ggccctttgt gcggggggag 1260
 cggctcgggg ggtgcgtgcg tgtgtgtgtg cgtggggagc gccgcgtgcg gctccgcgct 1320
 gcccggcggc tgtgagcgct gcgggcgcgg cgcggggctt tgtgcgctcc gcagtgtgcg 1380

cgaggggagc gcggccgggg gcggtgcccc gcggtgcggg gggggctgcg aggggaacaa 1440
 aggctgcgtg cgggggtgtgt gcgtgggggg gtgagcaggg ggtgtgggcg cgtcggctcg 1500
 gctgcaaccc cccctgcacc cccctccccg agttgctgag cacggcccgg cttcgggtgc 1560
 ggggctccgt acggggcggt gcgcggggct cgccgtgccg ggcggggggt ggcggcaggt 1620
 gggggtgccg ggcggggcgg ggccgcctcg ggccggggag ggctcggggg aggggcgcgg 1680
 cggcccccg agcgccggcg gctgtcagg cgcgggcagc cgcagccatt gccttttatg 1740
 gtaatcgtgc gagaggcgcg agggacttcc tttgtccaa atctgtgcgg agccgaaatc 1800
 tgggaggcgc cgccgcaccc cctctagcgg gcgcggggcg aagcggtgcg gcgccggcag 1860
 gaaggaaatg ggcggggagg gccttcgtgc gtgcgcgcg cgccgtcccc ttctccctct 1920
 ccagcctcgg ggctgtccgc ggggggacgg ctgccttcgg gggggacggg gcagggcggg 1980
 gttcggcttc tggcgtgtga ccggcggtc tagagcctc gctaaccatg ttcatgcctt 2040
 cttctttttc ctacagctcc tgggcaacgt gctggttatt gtgctgtctc atcattttgg 2100
 caaagaattc cgccaccatg tctatggggg ctctcgtc cctgctgctg gcactggccg 2160
 ccgggctggc tgtcgcaaga ccacctaata tcgtcctgat ttttgagac gatctgggat 2220
 acggcgacct gggatgctat ggccacccaa gtcaccac acccaacctg gaccagctgg 2280
 cagcaggagg cctgcggttc accgacttct acgtgccagt gagcctgtgc accccctcca 2340
 gagccgccct gctgacaggc aggtgccag tgcgcatgg catgtatcct ggcgtgctgg 2400
 tgccatctag caggggcggc ctgccactgg aggagtgac cgtggcagag gtgctggcag 2460
 ccagaggcta cctgacagga atggccggca agtggcacct gggagtggga ccagagggag 2520
 ccttcctgcc cctcaccag ggcttcacc ggtttctgg catcccttat tctcagacc 2580
 agggcccatg ccagaacctg acctgtttc caccagcaac accatgcgac ggaggatgtg 2640
 atcagggcct ggtgccaatc cactgctgg caaatctgag cgtggaggca cagcctccat 2700
 ggctgcctgg cctggaggca agatacatgg ccttcgcca cgacctgat gcagatgcac 2760
 agcggcagga tagaccttc tttctgtact atgcctcca ccacaccac tatccacagt 2820
 tcagcgcca gtcctttgcc gagaggtccg gaaggggacc attcggcgac tctctgatgg 2880
 agctggatgc cgccgtgggc accctgatga cagcaatcgg cgacctgggc ctgctggagg 2940
 agacactggt catcttcacc gccgataacg gccctgagac aatgcggatg tctagaggcg 3000
 gatgcagcgg cctgctgaga tgtggcaagg gaaccacata cgagggaggc gtgcgcgagc 3060
 ctgccctggc attttgcca ggacacatc cacctggagt gaccacagag ctggcctcct 3120
 ctctggacct gctgccaaca ctggccgcc tggcaggagc acctctgcca aatgtgacc 3180
 tggacggctt cgatctgagc cactgctgc tgggaaccgg caagtcccct aggcagtctc 3240
 tgtttcttta cccctcctat cctgatgagg tgcggggcgt gtttgccgtg agaaccggca 3300
 agtacaaggc cacttcttt acacagggt ctgcccacag cgacaccaca gcagatccag 3360
 catgccacgc cagctcctct ctgaccgcac acgagccacc tctgctgtac gacctgtcca 3420
 aggatcccgg cgagaactat aatctgctgg gaggagtggc aggagcaacc cctgaggtgc 3480
 tgcaggccct gaagcagct cagctgctga aggcacagct ggacgcagca gtgacattcg 3540
 gccaagcca ggtggccaga ggcgaggatc ccgccctgca gatctgttgc cccccggct 3600
 gcacccaag acctgcctgt tgccattgcc ccgaccaca cgcctaagat tctagagtcg 3660
 agccgcggac tagtaactg tttattgcag cttataatgg ttacaaataa agcaatagca 3720

```

tcacaaatTT cacaataaaa gcattttttt cactgcattc tagttgtggt ttgtccaaac 3780
tcatcaatgt atcttaggtc tagatacgta gataagtagc atggcggggt aatcattaac 3840
tacaaggaac ccctagtgat ggagttggcc actccctctc tgcgcgctcg ctgcgtcact 3900
gaggccgggc gaccaaaggt cgcgcgacgc ccgggctttg cccgggcggc ctcagtgagc 3960
gagcgagcgc gcagagaggg agtggccaaa gatccccggg taccgaggac gaattctcta 4020
gatatcgctc aatactgacc atttaaatca tacctgacct ccatagcaga aagtcaaaag 4080
cctccgaccg gaggcttttg acttgatcgg cacgtaagag gttccaactt tcaccataat 4140
gaaataagat cactaccggg cgtatttttt gagttatcga gattttcagg agctaaggaa 4200
gctaaaatga gccatattca acgggaaacg tcttgctcga ggccgcgatt aaattccaac 4260
atggatgctg atttatatgg gtataaatgg gctcgcgata atgtcgggca atcaggtgcg 4320
acaatctatc gattgtatgg gaagcccgat gcgccagagt tgtttctgaa acatggcaaa 4380
ggtagcggtg ccaatgatgt tacagatgag atggtcaggc taaactggct gacggaatTT 4440
atgcctcttc cgaccatcaa gcattttatc cgtactctg atgatgcag gttactcacc 4500
actgcgatcc cagggaanaac agcattccag gtattagaag aatatactga ttcaggtgaa 4560
aatattgttg atgcgctggc agtgttcctg cgccggttgc attcgattcc tgtttgtaat 4620
tgtcctttta acggcgatcg cgtatttcgt ctgcgtcagg cgcaatcacg aatgaataac 4680
ggtttggttg gtgcgagtga ttttgatgac gagcgtaatg gctggcctgt tgaacaagtc 4740
tggaagaaaa tgcataagct tttgccattc tcaccgatt cagtcgtcac tcatggtgat 4800
ttctcacttg ataaccttat tttgacgag gggaaattaa taggttgat tcatggtgat 4860
cgagtcggaa tcgcagaccg ataccaggat cttgccatcc tatggaactg cctcggtgag 4920
ttttctcctt cattacagaa acggcttttt caaaaatatg gtattgataa tcctgatatg 4980
aataaattgc agtttcactt gatgctcgat gagtttttct gagggcccaa atgtaatcac 5040
ctggctcacc ttcgggtggg ctttctgcg ttgctggcgt ttttccatag gctccgcccc 5100
cctgacgagc atcacaaaaa tcgatgctca agtcagaggt ggcgaaaccg gacaggacta 5160
taaagatacc aggcgtttcc ccctggaagc tccctcgtgc gctctcctgt tccgaccctg 5220
ccgcttaccg gatacctgtc cgcttttctc ctttcgggaa gcgtggcgct ttctcatagc 5280
tcacgctgta ggtatctcag ttcggtgtag gtcgttcgct ccaagctggg ctgtgtgcac 5340
gaaccccccg ttcagcccga ccgctgcgcc ttatccggtg actatcgtct tgagtccaac 5400
ccggttaagac acgacttata gccactggca gcagccactg gtaacaggat tagcagagcg 5460
aggatatgtag gcggtgctac agagtctctg aagtgggtggc ctaactacgg ctacactaga 5520
agaacagtat ttggtatctg cgctctgctg aagccagtta cctcggaana agagttggta 5580
gctcttgatc cggcaanaaa accaccgtg gtagcggtgg tttttttgtt tgcaagcagc 5640
agattacgcg cagaaaaaaa ggatctcaag aagatccttt gattttctac cgaagaaagg 5700
cccaccgctg aaggtgagcc agtgagttga ttgcagtcca gttacgctgg agtctgaggc 5760
tcgtcctgaa tgatatcaag cttgaattcg tt 5792

```

<210> 52

<211> 6342

<212> DNA

<213> 人工序列

<220>

<223> 合成多核苷酸

<400> 52

```
tttccatagg ctccgcccc ctgacgagca tcacaaaaat cgatgctcaa gtcagaggtg 60
gcgaaacccg acaggactat aaagatacca ggcgtttccc cctggaagct ccctcgtgcg 120
ctctcctggt ccgaccctgc cgcttaccgg atacctgtcc gcctttctcc cttcggaag 180
cgtggcgctt tctcatagct cacgctgtag gtatctcagt tcggtgtagg tcgttcgctc 240
caagctgggc tgtgtgcacg aacccccctg tcagcccgac cgctgcgcct tatccggtaa 300
ctatcgtctt gagtccaacc cggtaaagaca cgacttatcg ccactggcag cagccactgg 360
taacaggatt agcagagcga ggtatgtagg cggtgctaca gagttcttga agtgggtggc 420
taactacggc tacactagaa gaacagtatt tggtatctgc gctctgctga agccagttac 480
ctcggaaaaa gaggttgtag ctcttgatcc ggcaaaaaa ccaccgctgg tagcgggtgt 540
ttttttgttt gcaagcagca gattacgcgc agaaaaaag gatctcaaga agatcctttg 600
atcttctacc gaagaaaggc ccaccctga aggtgagcca gtgagttgat tgcagtcag 660
ttacgctgga gtctgaggct cgtcctgaat gatatcaagc ttgaattcgt gtcaggtggc 720
acttttcggg gaaatgtggc atgcctgcat ttggccactc cctctctgcg cgctcgtcgc 780
ctcactgagg cggggcgacc aaaggtcgcc cgacgcccgg gctttgcccg ggcgccctca 840
gtgagcgagc gagcgcgagc agaggagtg gccaaactca tcactagggg ttcctggagg 900
ggtggagtcg tgacgtgaat tacgtcatag ggtagggag gtccctgcata tgcggccgcg 960
atcttcaata ttggccatta gccatattat tcattggtta tatagcataa atcaatattg 1020
gctattggcc attgcatacg ttgtatctat atcataatat gtacatttat attggctcat 1080
gtccaatatg accgccatgt tggcattgat tattgactag ttattaatag taatcaatta 1140
cggggtcatt agttcatagc ccataatagg agttccgcgt tacataactt acggtaaatg 1200
gcccgcctgg ctgaccgccc aacgaccccc gccattgac gtcaataatg acgtatgttc 1260
ccatagtaac gccaataggg actttccatt gacgtcaatg ggtggagtat ttacggtaaa 1320
ctgcccactt ggcagtacat caagtgtatc atatgccaag tccgccccct attgacgtca 1380
atgacggtaa atggcccgcc tggcattatg ccagtagcat gaccttacgg gactttccta 1440
cttggcagta catctacgta ttagtcatcg ctattaccat ggtcgaggtg agccccacgt 1500
tctgcttcac tctccccatc tccccccct cccaccccc aattttgtat ttattttatt 1560
tttaattatt ttgtgcagcg atggggggcg gggggggggg ggggcgcgcg ccaggcgggg 1620
cggggcgggg cgagggggcg ggcggggcga ggcggagagg tgcggcgga gccaatcaga 1680
gcggcgcgct ccgaaagttt ctttttatgg cgaggcgcg gcggcgcgcc ccctataaaa 1740
agcgaagcgc gcggcgggcg ggagtcgctg cgcgctgcct tcgccccgtg ccccgctccg 1800
ccgccgcctc gcgccgccc ccccggtct gactgaccgc gttactcca caggtgagcg 1860
ggcgggacgg cccttctcct ccgggctgta attagcgtt ggtttaatga cggtttgttt 1920
cttttctgtg gctgcgtgaa agccttgagg ggctccggga gggccctttg tgcgggggga 1980
gcggctcggg ggggtcgtgc gtgtgtgtgt gcgtggggag cgccgcgtgc ggctccgcgc 2040
tgcccgcgcg ctgtgagcgc tgcgggcgcg gcgcggggct ttgtgcgtc cgcagtgtgc 2100
gcgaggggag cgcggccggg ggcggtgcc cgcggtgcg ggggggctgc gaggggaaca 2160
```

aaggctgcgt gcgggggtgtg tgcgtggggg ggtgagcagg ggggtgtgggc gcgtcggtcg 2220
ggctgcaacc cccctgcac cccctcccc gagttgctga gcacggcccc gcttcgggtg 2280
cggggctccg tacggggcgt ggcgcggggc tcgccgtgcc gggcgggggg tggcggcagg 2340
tgggggtgcc gggcggggag gggccgcctc gggccgggga gggctcgggg gaggggcgcg 2400
gcggcccccg gagcgccggc ggctgtcgag gcgcggcgag ccgcagccat tgccttttat 2460
ggtaatcgtg cgagagggcg cagggacttc ctttgtccca aatctgtgcg gagccgaaat 2520
ctgggaggcg ccgccgcacc ccctctagcg ggcgcggggc gaagcgggtc ggcgccggca 2580
ggaaggaaat gggcggggag ggccttcgtg cgtcgccgcg ccgccgtccc cttctccctc 2640
tccagcctcg gggctgtccg cggggggacg gctgccttcg ggggggacgg ggcaggcgcg 2700
ggttcggctt ctggcgtgtg accggcggtc ctagagcctc tgctaaccat gttcatgcct 2760
tcttcttttt cctacagctc ctgggcaacg tgctggttat tgtgtgtgtc catcattttg 2820
gcaaagaatt ccgccaccat gtctatgggg gctcctcgt ccttgcgtgc ggcactggcc 2880
gccgggctgg ctgtcgcaag accaccta atcgtcctga tttttgcaga cgatctggga 2940
tacggcgacc tgggatgcta tggccacca agctccacca cacccaacct ggaccagctg 3000
gcagcaggag gcctgcggtt caccgacttc tacgtgccag tgagcctgtg caccctcc 3060
agagccgccc tgctgacagg caggctgcca gtgcgcattg gcatgtatcc tggcgtgctg 3120
gtgccatcta gcaggggagc cctgccactg gaggaggtga ccgtggcaga ggtgctggca 3180
gccagaggct acctgacagg aatggccggc aagtggcacc tgggagtggg accagaggga 3240
gccttctgc cccctacca gggcttcac cggtttcttg gcatccctta ttctcacgac 3300
cagggcccat gccagaacct gacctgtttt ccaccagcaa caccatgcga cggaggatgt 3360
gatcagggcc tgggtgccaat cccactgctg gcaaacttga gcgtggaggc acagcctcca 3420
tggctgcctg gcctggaggc aagatacatg gccttcgcc acgacctgat ggcagatgca 3480
cagcggcagg atagacctt ctttctgtac tatgcctccc accacacca ctatccacag 3540
ttcagcggcc agtcctttgc cgagaggtcc ggaaggggac cattcggcga ctctctgatg 3600
gagctggatg ccgccgtggg caccctgatg acagcaatcg gcgacctggg cctgctggag 3660
gagacactgg tcatcttcac cgccgataac ggccctgaga caatgcggat gtctagaggc 3720
ggatgcagcg gcctgctgag atgtggcaag ggaaccacat acgaggagg cgtgcgcgag 3780
cctgccctgg cattttgccc aggacacatc gcacctggag tgaccacga gctggcctcc 3840
tctctggacc tgctgccaac actggccgcc ctggcaggag cacctctgcc aaatgtgacc 3900
ctggacggct tcgatctgag cccactgctg ctgggaaccg gcaagtcccc taggcagtct 3960
ctgttctttt acccctccta tctgatgag gtgcggggcg tgtttgccgt gagaaccggc 4020
aagtacaagg cccacttctt tacacagggc tctgccaca gcgacaccac agcagatcca 4080
gcatgccacg ccagctcctc tctgaccgca cagagccac ctctgctgta cgacctgtcc 4140
aaggatcccc gcgagaacta taatctgtg ggaggagtgg caggagcaac ccctgaggtg 4200
ctgcaggccc tgaagcagct gcagctgctg aaggcacagc tggacgcagc agtgacattc 4260
ggcccaagcc aggtggccag aggcgaggat cccgcctgc agatctgttg ccaccccgcc 4320
tgcaccccaa gacctgcctg ttgccattgc cccgaccac acgcctaaga ttctagagtc 4380
gagccgcgga ctagtaactt gtttattgca gcttataatg gttacaaata aagcaatagc 4440
atcacaaatt tcacaaataa agcatttttt tcaactgcatt ctagttgtgg tttgtccaaa 4500

ctcatcaatg tatcttacct gcaggctcac cagtgtttgt gactgggaac tctccctgcc 4560
aaatatggc ataatgctgt cctttagggt gcagcttatt gcccagggg aacagtctgt 4620
tgtgcagtc accccggcag gaatactccc attctgcctc tgttggtaac cttttccag 4680
cccagggtgca gtatgccact gcatcattcc aggacacatg cagcacaggg tggtcaggcc 4740
tgtggaggat agttgagtct ggtccctctg ggtgtctcca attggctcct ttaacaggca 4800
gccaccaggg ggctgcagcc actgcctgtt ggatattggt cttcacctgc tcacttagca 4860
tgccttcaaa aacaaaaactg tcaccaaatt tctcagcctc tgtaaggat ccagtgtgt 4920
ttacaaattt ctcaaattct gtgtttgaca cttcataggc atccatatag aaggcatcaa 4980
ttgtgactct cctagctggt gcttcacat cctgcttgat ctgagggtca tcagttccca 5040
tagtaaaaac tcctgcaggt ctagatacgt agataagtag catggcgggt taatcattaa 5100
ctacaaggaa cccctagtga tggagttggc cactccctct ctgcgcctc gctcgctcac 5160
tgaggccggg cgaccaaagg tcgcccgcg cccgggcttt gcccgggagg cctcagtgag 5220
cgagcgagcg cgagagagg gagtggccaa agatccccg gtaccgagga cgaattctct 5280
agatatcgct caatactgac catttaaate atacctgacc tccatagcag aaagtcaaaa 5340
gcctccgacc ggaggctttt gacttgatcg gcacgtaaga ggttccaact ttcaccataa 5400
tgaaataaga tcactaccgg gcgtattttt tgagttatcg agattttcag gagctaagga 5460
agctaaaatg agccatattc aacgggaaac gtcttgctcg aggccgcgat taaattccaa 5520
catggatgct gatattatg ggtataaatg ggctcgcgat aatgtcgggc aatcaggtgc 5580
gacaatctat cgattgtatg ggaagcccga tgcgccagag ttgtttctga aacatggcaa 5640
aggtagcggt gccaatgatg ttacagatga gatggtcagg ctaaactggc tgacggaatt 5700
tatgcctctt ccgaccatca agcattttat ccgtactcct gatgatgat ggttactcac 5760
cactgcgatc ccagggaaaa cagcattcca ggtattagaa gaatatcctg attcaggtga 5820
aaatatgtt gatgcgctgg cagtgttcct gcgccggtt cattcgattc ctgtttgtaa 5880
ttgtcctttt aacggcgatc gcgtatttcg tctcgctcag gcgcaatcac gaatgaataa 5940
cggtttgggt ggtgcgagt attttgatga cgagcgtaat ggctggcctg ttgaacaagt 6000
ctggaaagaa atgcataagc ttttgccatt ctcaccggat tcagtcgtca ctcaggtga 6060
tttctcactt gataacctt ttttgacga ggggaaatta ataggttgta ttgatgttg 6120
acgagtcgga atcgagacc gataccagga tcttgccatc ctatggaact gcctcggtga 6180
gttttctcct tcattacaga aacggctttt tcaaaaatat ggtattgata atcctgatat 6240
gaataaattg cagtttact tgatgctcga tgagtttttc tgagggccca aatgtaatca 6300
cctggctcac cttcgggtgg gcctttctgc gttgctggcg tt 6342

<210> 53

<211> 6612

<212> DNA

<213> 人工序列

<220>

<223> 合成多核苷酸

<400> 53

cgccagggtt ttcccagtca cgacgttgta aaacgacggc cagtgccaaag cttgcatgcc 60

```

tgcatttggc cactccctct ctgcgcgtc gctcgctcac tgaggccggg cgaccaaagg 120
tcgcccgcacg cccgggcttt gcccgggcgg cctcagtgag cgagcgagcg cgcagagagg 180
gagtggccaa ctccatcact aggggttcct ggaggggtgg agtcgtgacg tgaattacgt 240
catagggtta gggaggctct gcagatcttc aatattggcc attagccata ttattcattg 300
gttatatagc ataaatcaat attggctatt ggccattgca tacgttgtat ctatatcata 360
atatgtacat ttatatattg tcatgtccaa tatgaccgcc atgttggcat tgattattga 420
ctagttatta atagtaatca attacggggg cattagtcca tagcccatat atggagttcc 480
gcgttacata acttacggta aatggcccgc ctggctgacc gcccaacgac cccgcgccat 540
tgacgtcaat aatgacgtat gttcccatag taacgccaat agggactttc cattgacgtc 600
aatgggtgga gtatttacgg taaactgccc acttggcagt acatcaagtg tatcatatgc 660
caagtccgcc ccctattgac gtcaatgacg gtaaattggcc cgcctggcat tatgccaggt 720
acatgacctt acgggacttt cctacttggc agtacatcta cgtattagtc atcgctatta 780
ccatggctga ggtgagcccc acgttctgct tcactctccc catctcccc ccctccccac 840
ccccaatttt gtatttattt attttttaat tattttgtgc agcgatgggg gcgggggggg 900
ggggggggcg cgcgccaggc ggggcggggc ggggcgaggg gcggggcggg gcgaggcgga 960
gaggtgcggc ggcagccaat cagagcggcg cgctccgaaa gtttcctttt atggcgaggc 1020
ggcgggcgcg gcggccctat aaaaagcgaa gcgcgcggcg ggcgggagtc gctgcgcgct 1080
gccttcgccc cgtgccccgc tccgccgccg cctcgcgccg cccgccccgg ctctgactga 1140
ccgcgttact cccacaggtg agcgggcggg acggcccttc tcctccgggc tgtaattagc 1200
gcttggttta atgacggctt gtttcttttc tgtggctgcg tgaaagcctt gaggggctcc 1260
gggagggccc tttgtgcggg gggagcggct cggggggtgc gtgcgtgtgt gtgtgcgtgg 1320
ggagcgccgc gtgcggctcc gcgctgcccg gcggctgtga gcgctgcggg cgcggcgcgg 1380
ggctttgtgc gctccgcagt gtgcgcgagg ggagcgcggc cgggggcggg gccccgcggt 1440
gcgggggggg ctgcgagggg aacaaaggct gcgtgcgggg tgtgtgcgtg ggggggtgag 1500
cagggggtgt gggcgcgtcg gtcgggctgc aacccccct gcacccccct ccccgagttg 1560
ctgagcacgg cccggcttcg ggtgcggggc tccgtacggg gcgtggcgcg gggctcgccg 1620
tgccgggcgg ggggtggcgg caggtggggg tgccgggcgg ggcggggccg cctcgggccg 1680
gggagggctc gggggagggg cgcggcgggc cccggagcgc cggcggctgt cgaggcgcgg 1740
cgagccgcag ccattgcctt ttatggtaat cgtgcgagag ggcgcaggga cttcctttgt 1800
cccaaactctg tgcggagccg aaatctggga ggcgccgccg caccctctct agcgggcgcg 1860
gggcgaagcg gtgcggcgcc ggcaggaagg aaatgggcgg ggagggcctt cgtgcgtcgc 1920
cgcgccgccg tccccttctc cctctccagc ctcggggctg tccgcggggg gacggctgcc 1980
ttcggggggg acggggcagg gcgggggttc gcttctggcg tgtgaccggc ggctctagag 2040
cctctgctaa ccatgttcat gccttcttct ttttctaca gctcctgggc aacgtgctgg 2100
ttattgtgct gtctcatcat tttggcaaag aattccgcca ccatgtctat gggggctcct 2160
cgctccctgc tgctggcact ggccgccggg ctggctgtcg caagaccacc taatatcgte 2220
ctgatttttg cagacgatct gggatacggc gacctgggat gctatggcca cccaagctcc 2280
accacacca acctggacca gctggcagca ggaggcctgc ggttcaccga cttctacgtg 2340
ccagtgagcc tgtgcacccc ctccagagcc gccctgctga caggcaggct gccagtgcgc 2400

```

```

atgggcatgt atcctggcgt gctggtgcca tctagcaggg gcggcctgcc actggaggag 2460
gtgaccgtgg cagaggtgct ggcagccaga ggctacctga caggaatggc cggcaagtgg 2520
cacctgggag tgggaccaga gggagccttc ctgccccctc accagggttt ccaccggttt 2580
ctgggcatcc cttatttctca cgaccagggc ccatgccaga acctgacctg ttttccacca 2640
gcaacaccat gcgacggagg atgtgatcag ggcctggtgc caatcccact gctggcaaat 2700
ctgagcgtgg aggcacagcc tccatggctg cctggcctgg aggcaagata catggccttc 2760
gcccacgacc tgatggcaga tgcacagcgg caggatagac ctttctttct gtactatgcc 2820
tcccaccaca cccactatcc acagtccagc ggccagtcct ttgccgagag gtccggaagg 2880
ggaccattcg gcgactctct gatggagctg gatgccggcg tgggcaccct gatgacagca 2940
atcggcgacc tgggcctgct ggaggagaca ctggtcattc tcaccgccga taacggccct 3000
gagacaatgc ggatgtctag aggcggatgc agcggcctgc tgagatgtgg caagggaacc 3060
acatacgagg gaggcgtgcg cgagcctgcc ctggcatttt ggccaggaca catcgcacct 3120
ggagtgaccc acgagctggc ctctctcttg gacctgctgc caacactggc cgccctggca 3180
ggagcacctc tgccaaatgt gaccctggac ggcttcgac tgagcccact gctgctggga 3240
accggcaagt cccctaggca gtctctgttc ttttaccctt cctatcctga tgaggtgcgg 3300
ggcgtgtttg cctgagaac cggcaagtac aaggcccact tctttacaca gggctctgcc 3360
cacagcgaca ccacagcaga tccagcatgc cagccagct cctctctgac cgcacacgag 3420
ccacctctgc tgtacgacct gtccaaggat cccggcgaga actataatct gctgggagga 3480
gtggcaggag caaccctga ggtgctgcag gccctgaagc agctgcagct gctgaaggca 3540
cagctggacg cagcagtgc attcggcca agccaggtgg ccagaggcga ggatcccgcc 3600
ctgcagatct gttgccaccc cggctgcacc ccaagacctg cctgttgcca ttgccccgac 3660
ccacacgcct aagattctag agtcgagccg cggactagta acttgtttat tgcagcttat 3720
aatggttaca aataaagcaa tagcatcaca aatttcacaa ataaagcatt tttttactg 3780
cattctagtt gtggtttgtc caaactcatc aatgtatctt aggtctagat acgtagataa 3840
gtagcatggc ggggttaatca ttaactacaa ggaaccctta gtgatggagt tggccactcc 3900
ctctctgcgc gctcgtcgc tcactgaggc cgggcgacca aaggtcggcc gacggccggg 3960
ctttgcccgg gcggcctcag tgagcgagcg agcgcgcaga gagggagtgg ccaaagatcc 4020
ccgggtaccg agctcgaatt cgtaatcatg tcatagctgt ttctgtgtg aaattgttat 4080
ccgctcaca ttccacacaa catacgagcc ggaagcataa agtgtaaagc ctgggggtgcc 4140
taatgagtga gctaactcac attaatgctg ttgcgtcac tgcccgcttt ccagtcggga 4200
aacctgtcgt gccagctgca ttaatgaatc ggccaacgcg cggggagagg cggtttgct 4260
attggcgaac ttttgtgag ttgaaggatc agatcacgca tcttcccgac aacgcagacc 4320
gttccgtggc aaagcaaaaag ttcaaaatca gtaaccgtca gtgccgataa gttcaaagtt 4380
aaacctggtg ttgatacaca cattgaaacg ctgatcgaaa acgcgtgaa aaacgtgct 4440
gaatgtgcga gcttcttccg cttctctgct cactgactcg ctgcgtcgg tcgttcggct 4500
gcggcgagcg gtatcagctc actcaaaggc ggtaatacgg ttatccacag aatcagggga 4560
taacgcagga aagaacatgt gagcaaaagg ccagcaaaag gccaggaacc gtaaaaaggc 4620
cgcgttgctg gcgtttttcc ataggtccg cccccctgac gagcatcaca aaaatcgacg 4680
ctcaagtcag aggtggcgaa acccgacagg actataaaga taccaggcgt tccccctgg 4740

```

aagctccctc gtgcgctctc ctgttccgac cctgccgctt accggatacc tgtccgcctt 4800
tctcccttcg ggaagcgtgg cgctttctca atgctcacgc ttaggtatc tcagttcggt 4860
gtaggtcggt cgctccaagc tgggctgtgt gcacgaacct cccgttcagc ccgaccgctg 4920
cgccttatcc ggtaactatc gtcttgagtc caaccggta agacacgact tatcgccact 4980
ggcagcagcc actggtaaca ggattagcag agcgaggtat gtaggcggtg ctacagagtt 5040
cttgaagtgg tggcctaact acggctacac tagaaggaca gtatttggtg tctgcgctct 5100
gctgaagcca gttaccttcg gaaaaagagt tggtagctct tgatccggca aacaaaccac 5160
cgctggtagc ggtgggtttt ttgtttgcaa gcagcagatt acgcgcagaa aaaaaggatc 5220
tcaagaagat cctttgatct tttctacggg gtctgacgt cagtggaaac atccgtcgag 5280
aggtctgcct cgtgaagaag gtgttgctga ctataccag gcctgaatcg ccccatcatc 5340
cagccagaaa gtgaggagc cacggttgat gagagctttg ttgtaggtgg accagttggt 5400
gattttgaac ttttgctttg ccacggaacg gtctgcgttg tcgggaagat gcgtgatctg 5460
atccttcaac tcagcaaaag ttcgatttat tcaacaaagc cacgttgtgt ctcaaatct 5520
ctgatgttac attgcacaag ataaaaatat atcatcatga acaataaac tgtctgctta 5580
cataaacagt aatacaagg gtgttatgag ccatattcaa cgggaaacgt cttgctcgaa 5640
gccgcgatta aattccaaca tggatgctga tttatatggg tataaatggg ctgcgcataa 5700
tgtcgggcaa tcaggtgcga caatctatcg attgtatggg aagcccgatg cgccagagtt 5760
gtttctgaaa catggcaaag gtagcgttg caatgatgtt acagatgaga tggtcagact 5820
aaactggctg acggaattta tgcctcttcc gaccatcaag cattttatcc gtactcctga 5880
tgatgcatgg ttactacca ctgcgatccc cgggaaaaca gcattccagg tattagaaga 5940
atatcctgat tcaggtgaaa atattgttga tgcgctggca gtgttcctgc gccggttgca 6000
ttcgattcct gtttgtaatt gtccttttaa cagcgatcgc gtatttcgtc tcgctcaggc 6060
gcaatcacga atgaataacg gtttggttga tgcgagtgat tttgatgacg agcgtaatgg 6120
ctggcctggt gaacaagtct ggaaagaaat gcataagctt ttgccattct caccgattc 6180
agtcgtcact catggtgatt tctcacttga taaccttatt ttgacgagg ggaaattaat 6240
aggttgtatt gatgttgac gagtcggaat cgcagaccga taccaggatc ttgccatcct 6300
atggaactgc ctcggtgagt tttctcctt attacagaaa cggctttttc aaaaatatgg 6360
tattgataat cctgatatga ataaattgca gtttcatttg atgctcgatg agtttttcta 6420
atcagaattg gttaattggt tgtaaacact gcagagcatt acgctgactt gacgggacgg 6480
cggctttggt gaataaatcg cattcgccat tcaggctgcg caactgttgg gaaggcgat 6540
cgggtcgggc ctcttcgcta ttacgccagc tggcgaaagg gggatgtgct gcaaggcgat 6600
taagttgggt aa 6612

<210> 54

<211> 918

<212> DNA

<213> 人工序列

<220>

<223> 合成多核苷酸

<400> 54

ggcatcctaa aaaatattca gtggaaacgt aaaaacatta aagactgatt aaacatcgca 60
 gcatgacaca gatttagcaa ctgagcataa ataatttgac tcggatactg ctccaaaatc 120
 cgaagaggac caatttcttc caggaggaca actacctcgt cctctgcaga cccctctcct 180
 cggcagctga aggagtgtgg ccaatctgcc tccacctccc cgcggaaccc ctactctcag 240
 gacctcctgc agcaccccaa actggaagtg gccgctgcag acccaaggac gaggggcacg 300
 cgggagccgg cagccctagt ggagcggtt ggagtggtga ggtgggaggg tcaccaggt 360
 ggggtgaggc tggggtaggt agcggagtga acggcttccg aagctctggg ccgccccag 420
 gttggactaa gcaggcgctc tgtcttcgcc cccgccagg gtgggcgtct cctgaggact 480
 ccccgccaca cctgaccgga gaccgcgcgc ccagcctaga acgcttcccc gaccagcgt 540
 agggccgccc cgactggcgg gcgagggctc gcgggaggcc tggcgaaccc gggggcgga 600
 ccaggcgggc aaggcccggc tgccgcagcg ccgctctgcg cgaggcggt ccgccgcggc 660
 ggagggatac ggcgacccat atatatatcg cggggcgag actcgcgctc cggcagtgt 720
 gctgggagtg tcgtggacgc cgtgccgta ctcgtagtca ggcggcggcg caggcgcgcg 780
 cggcggcata gcgcacagcg cgccttagca gcagcagcag cagcagcggc atcggaggt 840
 ccccgccgt cgagcccc gcgctggtgc agccacctc gctccctctg ctcttctc 900
 cttcgtcgc accaagag 918

<210> 55

<211> 953

<212> DNA

<213> 人工序列

<220>

<223> 合成多核苷酸

<400> 55

aattcggtag cctagttatt aatagtaatc aattacgggg tcattagttc atagcccata 60
 tatggagttc cgcggtacat aacttacggg aaatggcccc cctggctgac cgcccacga 120
 ccccgcccca ttgacgtcaa taatgacgta tgttcccata gtaacgcaa tagggacttt 180
 ccattgacgt caatgggtgg actatttacg gtaaacgcc cacttgagcag tacatcaagt 240
 gtatcatatg ccaagtacgc cccctattga cgtcaatgac ggtaaatggc ccgcctggca 300
 ttatgcccag tacatgacct tatgggactt tctacttgag cagtacatct acgtattagt 360
 catcgctatt accatggtag aggtgagccc caggttctgc ttcactctcc ccatctcccc 420
 cccctcccca cccccaattt tgtattttatt ttttttttaa ttattttgtg cagcgatggg 480
 ggcggggggg gggggggggc gcgcgccagg cggggcgagg cggggcgagg ggcgggcgcg 540
 ggcgaggcgg agagggtgcgg cggcagccaa tcagagcggc gcgctccgaa agtttctttt 600
 tatggcgagg cggcggcggc ggcgcccta taaaaagcga agcgcgcggc gggcgggagt 660
 cgctgcgacg ctgccttcgc cccgtgcccc gctccgccgc cgctcgcgc cgcccgcccc 720
 ggctctgact gaccgcgtta ctcccacagg tgagcgggag ggacggccct tctctccgg 780
 gctgtaatta gcgcttggtt taatgacggc ttgtttcttt tctgtggctg cgtgaaagcc 840
 ttgaggggct ccgggagcta gagcctctgc taaccatgtt catgccttct tcttttctct 900
 acagctcctg ggcaacgtgc tggttattgt gctgtctcat cattttggca aag 953

<210> 56

<211> 37

<212> DNA

<213> 人工序列

<220>

<223> 合成多核苷酸

<400> 56

gtagataagt agcatggcgg gttaatcatt aactaca 37

<210> 57

<211> 180

<212> DNA

<213> 人工序列

<220>

<223> 3' ITR

<400> 57

gtagataagt agcatggcgg gttaatcatt aactacaagg aaccctagt gatggagttg 60
gccactccct ctctgcgcgc tcgctcgtc actgaggccg ggcgaccaa ggtcgcccga 120
cgcccgggct ttgcccgggc ggcctcagt agcgagcgag cgcgagaga gggagtggcc 180

<210> 58

<211> 380

<212> DNA

<213> 人工序列

<220>

<223> 合成多核苷酸

<400> 58

ggcattgatt attgactagt tattaatagt aatcaattac ggggtcatta gttcatagcc 60
catatatgga gttccgcgtt acataactta cggtaaattg cccgcctggc tgaccgcca 120
acgacccccg cccattgacg tcaataatga cgtatgttcc catagtaacg ccaataggga 180
ctttccattg acgtcaatgg gtggagtatt tacggtaaac tgcccacttg gcagtacatc 240
aagtgtatca tatgccaagt ccgccccta ttgacgtcaa tgacggtaaa tggccgcct 300
ggcattatgc ccagtacatg accttacggg actttcctac ttggcagtac atctacgtat 360
tagtcatcgc tattaccatg 380

<210> 59

<211> 1246

<212> DNA

<213> 人工序列

<220>

<223> 合成多核苷酸

<400> 59

tcgaggtgag cccacacgttc tgcttcactc tccccatctc cccccctcc ccacccccaa 60
 ttttgtatattt atttatattt taattatttt gtgcagcgcg gggggcgggg gggggggggg 120
 ggcgcgcgcc aggcggggcg gggcggggcg agggcgggg cggggcgagg cggagaggtg 180
 cggcggcagc caatcagagc ggcgcgctcc gaaagtttcc ttttatggcg aggcggcggc 240
 ggcgcgggcc ctataaaaag cgaagcgcgc ggcgggcggg agtcgctgcg cgctgccttc 300
 gccccgtgcc ccgctccgcc gccgcctcgc gccgcccccc cgggctctga ctgaccgcgt 360
 tactcccaca ggtgagcggg cgggacggcc cttctcctcc gggctgtaat tagcgcttgg 420
 tttaatgacg gcttgtttct tttctgtggc tgcgtgaaag ccttgagggg ctccgggagg 480
 gccctttgtg cgggggggagc ggctcggggg gtgcgtgcgt gtgtgtgtgc gtggggagcg 540
 ccgctgtcgg ctccgcgtg cccggcggt gtgagcgtg cgggcgcggc gcggggcttt 600
 gtgcgtccg cagtgtgcgc gaggggagcg cggccggggg cggtgccccg cggtgcgggg 660
 ggggctgcga ggggaacaaa ggctgcgtgc ggggtgtgtg cgtggggggg tgagcagggg 720
 gtgtggggcg gtcggtcggg ctgcaacccc ccctgcaccc ccctccccga gttgctgagc 780
 acggccccggc ttcgggtgcg gggctccgta cggggcgctg gcgggggctc gccgtgccgg 840
 gcgggggggtg gcggcaggtg ggggtgccgg gcggggcggg gccgcctcgg gccggggagg 900
 gctcggggga ggggcgcggc ggcccccgga gcgccggcg ctgtcgaggc gcggcgagcc 960
 gcagccattg ccttttatgg taatcgtgcg agagggcga gggacttcct ttgtcccaaa 1020
 tctgtgcgga gccgaaatct gggaggcgcc gccgcacccc ctctagcggg cgcggggcga 1080
 agcggtgcgg cgccggcagg aaggaaatgg gcggggaggg ccttcgtgcg tcgccgcgcc 1140
 gccgtcccct tctccctctc cagcctcggg gctgtccgcg gggggacggc tgccttcggg 1200
 ggggacgggg cagggcgggg ttcggcttct ggcgtgtgac cggcgg 1246

<210> 60

<211> 95

<212> DNA

<213> 人工序列

<220>

<223> 合成多核苷酸

<400> 60

cctctgctaa ccatgttcat gcctttcttct ttttctaca gctcctgggc aacgtgctgg 60
 ttattgtgct gtctcatcat tttggcaaag aattc 95

<210> 61

<211> 1061

<212> DNA

<213> 人工序列

<220>

<223> 合成多核苷酸

<400> 61

tagggaggtc ctgcacgtta cataacttac ggtaaatggc ccgcctggct gaccgccccaa 60
 cgacccccgc ccattgacgt caataatgac gtatgttccc atagtaacgc caatagggac 120

ttccattga cgtcaatggg tggagtattht acggtaaaact gcccacttgg cagtacatca 180
 agtgtatcat atgccaagta cgtcccttat tgacgtcaat gacggtaaat ggcccgctg 240
 gcattatgcc cagtacatga ccttatggga ctttctact tggcagtaca tctacgtatt 300
 agtcatcgct attaccatgg tcgaggtgag cccacgttc tgcttctc tccccatctc 360
 cccccctcc ccacccccaa ttttgtattht atttattht taattattht gtgcagcgat 420
 gggggcgggg gggggggggg gcgcgcgcca ggcgggggcg ggcgggggcg gggcggggc 480
 gggcgaggc ggagaggtgc ggcggcagcc aatcagagcg gcgcgctccg aaagtthtct 540
 tttatggcga ggcggcggcg gcggcgcccc tataaaaagc gaagcgcgcg gcggcgggga 600
 gtcgtgcgc gctgccttcg ccccggtgcc cgctccgccc ccgcctcgcg ccggcgccc 660
 cggtcttgac tgaccgcgtt actaaaacag gtaagtccg cctccgcgc gggtthtggc 720
 gcctcccgcg ggcgcccc tctcacggc gagcgctgc acgtcagacg aaggcgcgag 780
 cgagcgctct gatccttcg cccggacgt caggacagcg gcccgctgct cataagactc 840
 ggcttagaa cccagtatc agcagaagga cattthtagga cgggacttgg gtgactctag 900
 ggcaactggtt tttttccag agagcggaac aggcgaggaa aagtagtccc tttcggcgga 960
 tttcgggag ggatctccgt gggcggtga acgcatga tgcctctact aaccatgttc 1020
 atgtthtctt tttthtcta caggtcctgg gtgacgaaca g 1061

<210> 62

<211> 1527

<212> DNA

<213> 人工序列

<220>

<223> 合成多核苷酸

<400> 62

atgagcatgg gcgccccag aagcctgta cttgctthtag ctgctggcct tgcagtggca 60
 agggcccccta acatcgctgatccttgca gatgacttgg gatatgggga tcttggttgt 120
 tatggccacc catcaagcac aactcccaat ctggatcagt tggctgcagg aggtctgagg 180
 tttacagact tttatgttcc agtctccctg tgcactcctt ctgggctgc cctgcttact 240
 gggaggctcc ctgtgagaat gggtatgtac cctggagtgt tggctccatc cagcagggga 300
 gggctgcccc tggaagaggt gacagtggca gaggtgctgg cagcacgagg ctatctgact 360
 ggcatggcag gcaagtggca cctgggtgta gggccagagg gtgctthtct gcctcccat 420
 cagggttht ataggttht gggaatccca tactctcatg accaaggacc ctgccagaac 480
 ctacactgtt tccccctgc aacacatgt gatggggct gtgatcaagg tctggttct 540
 ataccactgc ttgctaactt ttcagtggaa gctcaaccac cctggctgcc tggcttgag 600
 gctagataca tggccttcgc acatgatctg atggcagatg ccagagaca agataggcct 660
 ttttctct actatgcatc tcaccacacc cactatctc agttctcagg ccaatcattt 720
 gctgagcgta gtggcagggg ccctthtggg gacagttga tggaactgga tgccgcagtt 780
 ggtaccctca tgacagcaat aggggactta ggtthtctgg aggaacatt ggtaatttht 840
 acagctgata atggccctga gacaatgaga atgtctaggg gaggtgctc tggctthtctg 900
 aggtgtggta aagggactac atatgaggga ggagtgagg aaccagctct tgcctthtgg 960

ccaggtcaca tagccccctgg agttacacat gaactagctt cttccctgga cttgcttctt 1020
 aactggcag cctggcagg tgccccctctc cctaattgtaa ctttagatgg atttgacctc 1080
 tctccactac ttttagggac agggaaaagt ccaaggcagt ccttattctt ctatccttcc 1140
 taccagatg aggtgagggg tgtttttgcc gtgaggactg ggaaatacaa agctcatttt 1200
 tttaccagagg gatcagctca ttcagacacc acagctgac ctgcctgtca tgccagcagt 1260
 agcttgacag cacatgagcc tcccttactg tatgacctga gcaaggaccc aggggagAAC 1320
 tataacctgc ttgggggggt tgctggggcc acccagaag tgcttcaggc actaaagcag 1380
 ctgcaactgc ttaaagcaca gttggatgct gcagtacct ttggcccttc ccaggtggcc 1440
 agaggcgagg atccccccct gcagatctgc tgccaccag gctgcacacc cagacctgcc 1500
 tgctgtcact gccccgaccc acacgcc 1527

<210> 63

<211> 57

<212> DNA

<213> 人工序列

<220>

<223> 合成多核苷酸

<400> 63

gctactaact tcagcctgct gaagcaggct ggagacgtgg aggagaaccc tggacct 57

<210> 64

<211> 1122

<212> DNA

<213> 智人

<400> 64

atggctgccc cagccctggg gctggtgtgt ggcagatgcc ctgagctggg cctggtgctg 60
 cttctcctgc tgctgagcct cctgtgtggt gctgctggct ctcaggaagc agggacagga 120
 gcaggagcag gttctctggc tggctcatgc ggttgtggga cccccagag gccaggggct 180
 catgggtcct ctgcagctgc ccacaggtac tcaagggaag caaatgcccc tggccccgta 240
 cctggggaaa ggcaacttgc tcaactccaag atggttccta tccctgcagg agtttttact 300
 atgggaactg atgaccctca gatcaagcag gatggtgaag caccagctag gagagtcaca 360
 attgatgcct tctatatgga tgcctatgaa gtgtcaaaca cagaatttga gaaatttgta 420
 aacagcactg gataccttac agaggctgag aaatttggtg acagttttgt ttttgaaggc 480
 atgctaagtg agcaggtgaa gaccaatatc caacaggcag tggctgcagc cccctggtgg 540
 ctgcctgtta aaggagccaa ttggagacac ccagagggac cagactcaac tatectccac 600
 aggctgacc accctgtgct gcatgtgtcc tggaatgatg cagtggcata ctgcacctgg 660
 gctgggaaaa ggttaccaac agaggcagaa tgggagtatt cctgccgggg tggactgcac 720
 aacagactgt tcccctgggg caataagctg caacctaaag gacagcatta tgccaatatt 780
 tggcaggagg agttcccagt cacaaacact ggtgaggatg gcttccaggg aactgcccct 840
 gtggatgctt tcccaccaa tggctatggg ttgtacaata tagttgggaa tgcctgggag 900
 tggacttctg actggtggac ggtccatcac agtgtggaag agacactgaa cccaaagggg 960

ccccctcag gcaaggacag agtcaagaaa ggtggctctt atatgtgtca cagaagctat 1020
 tgctacagat ataggtgtgc tgcaagaagt cagaacaccc ctgacagctc agctagcaat 1080
 ctgggattta gatgtgcagc agatagactc cccacatgg ac 1122

<210> 65

<211> 3739

<212> DNA

<213> 人工序列

<220>

<223> 合成多核苷酸

<400> 65

ggcatcctaa aaaatattca gtggaaacgt aaaaacatta aagactgatt aaacatcgca 60
 gcatgacaca gatttagcaa ctgagcataa ataatttgac tcggatactg ctccaaaatc 120
 cgaagaggac caattttcttc caggaggaca actacctcgt cctctgcaga cccctctcct 180
 cggcagctga aggagtgtgg ccaatctgcc tccacctccc cgcggacccc ctactctcag 240
 gacctcctgc agcaccccaa actggaagtg gccgctgcag acccaaggac gaggggcacg 300
 cgggagccgg cagccctagt ggagcggttg gagatgttga ggtgggaggg tcaccaggt 360
 ggggtgaggc tggggtaggt agcggagtga acggcttccg aagctctggg ccgccccag 420
 gttggactaa gcaggcgctc tgtcttcgcc cccgcccagg gtgggcgtct cctgaggact 480
 ccccgccaca cctgaccgga gaccgcgcgc ccagcctaga acgcttcccc gaccagcgt 540
 agggccgccg cgactggcgg gcgagggtcg gcgggaggcc tggcgaacct gggggcgga 600
 ccaggcgggc aaggcccggc tgccgcagcg ccgctctgcg cgaggcggct ccgcccgggc 660
 ggagggatac ggcgcaccat atatatatcg cggggcgagc actcgcgctc cggcagtgg 720
 gctgggagtgc tcgtggacgc cgtgccgtta ctctagtcga ggcggcggcg caggcggcg 780
 cggcggcata gcgcacagcg cgccttagca gcagcagcag cagcagcggc atcgaggta 840
 cccccgccgt cgcagcccc gcgctgggtc agccaccctc gctccctctg ctcttcctcc 900
 ctctcgtcgc accaagaggt aagggtttta gggatggttg gttggtggg tattaatgtt 960
 taattacctg gagcacctgc ctgaaatcac ttttttcag gttgggccac ccgccccac 1020
 catgagcatg ggcgccccca gaagcctgtt acttgcttta gctgctggcc ttgcagtggc 1080
 aaggccccct aacatcgtgc tgatctttgc agatgacttg ggatatgggg atcttggttg 1140
 ttatggccac ccatcaagca caactcccaa tctggatcag ttggctgcag gaggtctgag 1200
 gtttacagac ttttatgttc cagtctccct gtgcactcct tctcgggctg ccctgcttac 1260
 tgggaggctc cctgtgagaa tgggtatgta ccctggagtgt ttggtcccat ccagcagggg 1320
 agggtgccc ctggaagagg tgacagtggc agaggtgctg gcagcacgag gctatctgac 1380
 tggcatggca ggcaagtggc acctgggtgt agggccagag ggtgctttcc tgccctccca 1440
 tcagggttt cataggtttc tgggaatccc atactctcat gaccaaggac cctgccagaa 1500
 cctcacctgt tccccccctg caacaccatg tgatgggggc tgtgatcaag gtctggttcc 1560
 tataccactg cttgctaate tttcagtga agctcaacca ccctggctgc ctggcttgga 1620
 ggctagatac atggccttcg cacatgatct gatggcagat gccagagac aagataggcc 1680
 tttcttcctc tactatgcat ctccaccac ccactatcct cagttctcag gccaatcatt 1740

tgctgagcgt agtggcaggg gccatttgg ggacagtttg atggaactgg atgccgcagt 1800
 tggtaccctc atgacagcaa taggggactt aggtttgctg gaggaacat tggtaatTTT 1860
 cacagctgat aatggccctg agacaatgag aatgtctagg ggaggctgct ctggtcttct 1920
 gaggtgtggg aaagggacta catatgaggg aggagtgagg gaaccagctc ttgccttttg 1980
 gccaggtcac atagcccctg gagttacaca tgaactagct tcttccctgg acttgcttcc 2040
 tacactggca gccctggcag gtgcccctct ccctaagtga acttttagatg gatttgacct 2100
 ctctccacta ctttttaggga cagggaaaag tccaaggcag tccttattct tctatccttc 2160
 ctaccagat gaggtgaggg gtgtttttgc cgtgaggact gggaaataca aagctcattt 2220
 ttttaccag ggatcagctc attcagacac cacagctgat cctgcctgtc atgccagcag 2280
 tagcttgaca gcacatgagc ctcccttact gtatgacctg agcaaggacc caggggagaa 2340
 ctataacctg cttggggggg ttgctggggc caccacagaa gtgcttcagg cactaaagca 2400
 gctgcaactg cttaaagcac agttggatgc tgcagtgacc tttggccctt ccaggtggc 2460
 cagaggcgag gatccccgcc tgcagatctg ctgccacca ggctgcacac ccagacctgc 2520
 ctgctgtcac tgccccgacc cacacgccgg cagcggagct actaacttca gcctgctgaa 2580
 gcaggctgga gacgtggagg agaaccctgg acctatggct gccccagccc tggggctggg 2640
 gtgtggcaga tgccctgagc tgggcctggg gctgcttctc ctgctgctga gcctcctgtg 2700
 tggtgctgct ggctctcagg aagcagggac aggagcagga gcaggttctc tggctggctc 2760
 atgcggttgt gggaccccc agaggccagg ggctcatggg tcctctgcag ctgccacag 2820
 gtactcaagg gaagcaaatg cccctggccc cgtacctggg gaaaggcaac ttgctcactc 2880
 caagatgggt cctatccctg caggagtttt tactatggga actgatgacc ctcagatcaa 2940
 gcaggatggg gaagcaccag ctaggagagt cacaattgat gccttctata tggatgccta 3000
 tgaagtgtca aacacagaat ttgagaaatt tgtaaacagc actggatacc ttacagaggc 3060
 tgagaaattt ggtgacagtt ttgtttttga aggcatgcta agtgagcagg tgaagaccaa 3120
 tatccaacag gcagtggctg cagccccctg gtggctgcct gttaaaggag ccaattggag 3180
 acaccagag ggaccagact caactatcct ccacaggcct gaccaccctg tgctgcatgt 3240
 gtcctggaat gatgcagtgg catactgcac ctgggctggg aaaaggttac caacagaggc 3300
 agaatgggag tattcctgcc ggggtggact gcacaacaga ctgttcccct ggggcaataa 3360
 gctgcaacct aaaggacagc attatgccaa tatttggcag ggagagttcc cagtcacaaa 3420
 cactggtgag gatggcttcc agggaactgc ccctgtggat gctttccac ccaatggcta 3480
 tgggttgtac aatatagttg ggaatgctg ggagtggact tctgactggg ggacgggtcca 3540
 tcacagtgtg gaagagacac tgaacccaaa ggggcccccc tcaggcaagg acagagtcaa 3600
 gaaagggtgc tcttatatgt gtcacagaag ctattgctac agatataggt gtgctgcaag 3660
 aagtcagaac acccctgaca gctcagctag caatctggga tttagatgtg cagcagatag 3720
 actccccacc atggactga 3739

<210> 66

<211> 54

<212> DNA

<213> 人工序列

<220>

<223> 合成多核苷酸

<400> 66

gagggcagag gaagtccttct aacatgcggt gacgtggagg agaatcccgg ccct 54

<210> 67

<211> 3686

<212> DNA

<213> 人工序列

<220>

<223> 合成多核苷酸

<400> 67

```

aattcggtac cctagttatt aatagtaatc aattacgggg tcattagttc atagcccata 60
tatggagttc cgcgttacat aacttacggt aaatggcccc cctggctgac cgcccaacga 120
ccccgcccc ttgacgtcaa taatgacgta tgttcccata gtaacgcaa tagggacttt 180
ccattgacgt caatgggtgg actattttacg gtaaaactgcc cacttggcag tacatcaagt 240
gtatcatatg ccaagtacgc cccctattga cgtcaatgac ggtaaattggc ccgcctggca 300
ttatgcccag tacatgacct tatgggactt tcctacttgg cagtacatct acgtattagt 360
catcgtctatt accatggtcg aggtgagccc cacgttctgc ttcactctcc ccatctcccc 420
cccctcccca cccccaattt tgtattttatt tatttttttaa ttattttgtg cagcgatggg 480
ggcggggggg gggggggggc gcgcgccagg cggggcgagg cggggcgagg ggcggggcgg 540
ggcgaggcgg agaggtgcgg cggcagccaa tcagagcggc gcgctccgaa agtttctttt 600
tatggcgagg cggcgggcgg ggcggcccta taaaaagcga agcgcgcggc gggcgggagt 660
cgctgcgacg ctgccttcgc cccgtgcccc gctccgccgc cgcctcgcgc cgcccccccc 720
ggctctgact gaccgcgtta ctcccacagg tgagcgggag ggacggccct tctcctccgg 780
gctgtaatta gcgcttggtt taatgacggc ttgtttcttt tctgtggctg cgtgaaagcc 840
ttgaggggct ccgggagcta gagcctctgc taacctgtt catgccttct tcttttctct 900
acagctcctg ggcaacgtgc tggttattgt gctgtctcat cattttggca aaggctagcg 960
ccgccaccat gagcatgggc gccccagaa gcctgttact tgcttttagct gctggccttg 1020
cagtggcaag gccccctaac atcgtgctga tctttgcaga tgacttggga tatggggatc 1080
ttggttggtt tggccacca tcaagcaca ctcccaatct ggatcagttg gctgcaggag 1140
gtctgaggtt tacagacttt tatgttccag tctccctgtg cactccttct cgggctgccc 1200
tgcttacttg gaggtccct gtgagaatgg gtatgtacce tggagtgttg gtcccatcca 1260
gcaggggagg gctgccccct gaagaggtga cagtggcaga ggtgctggca gcacgaggct 1320
atctgactgg catggcaggc aagtggcacc tgggtgtagg gccagagggt gctttcctgc 1380
ctccccatca gggctttcat aggtttcttg gaatcccata ctctcatgac caaggacct 1440
gccagaacct cacctgtttc cccctgcaa caccatgtga tgggggctgt gatcaaggtc 1500
tggttcctat accactgctt gctaattctt cagtgggaag tcaaccaccc tggtgcctg 1560
gcttggaggc tagatacatg gccttcgcac atgatctgat ggcagatgcc cagagacaag 1620
ataggccttt ctctctctac tatgcatctc accacacca ctatcctcag ttctcaggcc 1680
aatcatttgc tgagcgtagt ggcaggggcc catttgggga cagtttgatg gaactggatg 1740

```

ccgcagttgg taccctcatg acagcaatag gggacttagg tttgctggag gaaacattgg 1800
taattttcac agctgataat ggccctgaga caatgagaat gtctagggga ggctgctctg 1860
gtcttctgag gtgtggtaaa gggactacat atgagggagg agtgaggga ccagctcttg 1920
ccttttggcc aggtcacata gcccctggag ttacacatga actagcttct tccctggact 1980
tgcttcttac actggcagcc ctggcaggtg cccctctccc taatgtaact ttagatggat 2040
ttgacctctc tccactactt ttagggacag ggaaaagtcc aaggcagtcc ttattcttct 2100
atccttccta cccagatgag gtgaggggtg tttttgccgt gaggactggg aaatacaaag 2160
ctcatttttt taccagggga tcagctcatt cagacaccac agctgatcct gcctgtcatg 2220
ccagcagtag cttgacagca catgagcctc cttactgta tgacctgagc aaggaccag 2280
gggagaacta taacctgctt ggggggggtg ctggggccac cccagaagtg cttcaggcac 2340
taaagcagct gcaactgctt aaagcacagt tggatgctgc agtgaccttt ggcccttccc 2400
aggtggccag aggcgaggat cccgccctgc agatctgctg ccaccaggc tgcacacca 2460
gacctgcctg ctgtcactgc cccgaccac acgccggcag cggagctact aacttcagcc 2520
tgctgaagca ggctggagac gtggaggaga accctggacc tatggctgcc ccagccctgg 2580
ggctgggtgtg tggcagatgc cctgagctgg gcctgggtgt gcttctctg ctgctgagcc 2640
tctgtgtgg tgctgctggc tctcaggaag caggacagg agcaggagca ggttctctgg 2700
ctggctcatg cggttgtggg accccccaga ggccaggggc tcatgggtcc tctgcagctg 2760
cccacaggta ctcaaggga gcaaatgcc ctggccccgt acctggggaa aggcaacttg 2820
ctcactccaa gatggttct atccctgcag gagtttttac tatgggaact gatgacctc 2880
agatcaagca ggatggtgaa gcaccagcta ggagagtcac aattgatgcc ttctatatgg 2940
atgcctatga agtgtcaaac acagaatttg agaaatttgt aaacagcact ggatacctta 3000
cagaggctga gaaatttggg gacagttttg tttttgaagg catgctaagt gagcaggtga 3060
agaccaatat ccaacaggca gtggctgcag cccctgggtg gctgcctgtt aaaggagcca 3120
attggagaca cccagaggga ccagactcaa ctatcctcca caggcctgac caccctgtgc 3180
tgcatgtgtc ctggaatgat gcagtggcat actgcacctg ggctgggaaa aggttaccaa 3240
cagaggcaga atgggagtat tcctgccggg gtggactgca caacagactg ttcccctggg 3300
gcaataagct gcaacctaaa ggacagcatt atgccaatat ttggcaggga gagttccag 3360
tcacaaacac tggtaggat ggcttccagg gaactgcccc tgtggatgct ttcccacca 3420
atggctatgg gttgtacaat atagttggga atgcctggga gtggacttct gactgggtga 3480
cggtccatca cagtgtggaa gagacactga acccaaagg gccccctca ggcaaggaca 3540
gagtcaagaa aggtggctct tatatgtgtc acagaagcta ttgctacaga tataggtgtg 3600
ctgcaagaag tcagaacacc cctgacagct cagctagcaa tctgggattt agatgtgcag 3660
cagatagact ccccacatg gactga 3686

<210> 68

<211> 4346

<212> DNA

<213> 人工序列

<220>

<223> 合成多核苷酸

<400> 68

```

ttggccactc cctctctgcg cgctcgctcg ctactgagg ccgggcgacc aaaggtcgcc 60
cgacgccccg gctttgcccc ggcggcctca gtgagcgagc gagcgcgag agagggagtg 120
gccaactcca tcactagggg ttcctggagg ggtggagtcg tgacgtgaat tacgtcatag 180
ggttagggag gtcctgcata tgcggccgcg gcatacctaaa aaatattcag tggaaacgta 240
aaaacattaa agactgatta aacatcgag catgacacag atttagcaac tgagcataaa 300
taatttgact cggatactgc tccaaaatcc gaagaggacc aatttcttcc aggaggacaa 360
ctacctcgtc ctctgcagac ccctctctc ggagctgaa ggagtgtggc caatctgcct 420
ccacctcccc gcggaccccc tactctcagg acctctgca gcaccccaaa ctggaagtgg 480
ccgctgcaga cccaaggacg aggggcacgc gggagccggc agccctagtg gagcggttgg 540
agatgttgag gtgggagggt caccaggtg gggtaggct ggggtaggta gcggagtga 600
cggcttccga agctctgggc cgccccagg ttggactaag caggcgctct gtcttcgccc 660
ccgccagggt tgggcgtctc ctgaggactc cccgccacac ctgaccgag accgcgcgc 720
cagcctagaa cgcttccccg acccagcgta gggccgcgc gactggcggg cgagggtcgg 780
cgggaggcct ggcaacccg ggggcgggac caggcgggca aggcccggt gccgcagcgc 840
cgctctgcmc gaggcggctc cgccgcggcg gagggatacg gcgcaccata tatatatcmc 900
ggggcgcaga ctgcgctcc ggagtggtg ctgggagtg ctgggacgcc gtgccgttac 960
tcgtagtcag gcggcggcmc aggcggcggc ggcgcatag cgcacagcmc gccttagcag 1020
cagcagcagc agcagcgga tcggaggtac cccgcgcgtc gcagccccg cgctggtgca 1080
gccaccctcg ctccctctgc tcttctccc ttcgctcga ccaagaggta agggtttaag 1140
ggatggttgg ttggtgggtt attaatgttt aattacctgg agcacctgcc tgaatact 1200
ttttttcagg ttgggccacc cgccgccacc atgagcatgg gcgccccag aagcctgtta 1260
cttgctttag ctgctggcct tgcagtgga agggcccta acatcgctg gatctttgca 1320
gatgacttgg gatattggga tcttggttgt tatggccacc catcaagcac aactcccaat 1380
ctggatcagt tggctgcagg aggtctgagg ttacagact ttatgttcc agtctccctg 1440
tgcactcctt ctcgggctgc cctgcttact gggaggctcc ctgtgagaat gggatatgtac 1500
cctggagtgt tgggtccatc cagcagggga gggctgcccc tggaagagg gacagtggca 1560
gaggtgctgg cagcacgagg ctatctgact ggcatggcag gcaagtggca cctgggtgta 1620
gggccagagg gtgctttcct gcctcccat cagggtttc ataggtttct gggaatccca 1680
tactctcatg accaaggacc ctgccagaac ctacctgtt tccccctgc aacaccatgt 1740
gatgggggct gtgatcaagg tctggttct atacctgc ttgctaactc ttcagtggaa 1800
gctcaaccac cctggctgcc tggttggag gctagataca tggccttcgc acatgatctg 1860
atggcagatg cccagagaca agataggcct ttcttctct actatgcac tcaccacacc 1920
cactatctc agttctcagg ccaatcattt gctgagcga gtggcagggg ccattttggg 1980
gacagtttga tggaaactga tgccgcagtt ggtacctca tgacagcaat aggggactta 2040
ggtttgctgg aggaacatt ggtaattttc acagctgata atggccctga gacaatgaga 2100
atgtctaggg gaggtgctc tggcttctg aggtgtgga aaggactac atatgaggga 2160
ggagtgaggg aaccagctct tgcttttgg ccaggtcaca tagccctgg agttacacat 2220
gaactagctt ctccctgga ctgcttctc aactggcag ccctggcagg tgccctctc 2280

```

```

cctaattgtaa ctttagatgg atttgacctc tctccactac ttttagggac agggaaaagt 2340
ccaaggcagt ccttattctt ctatccttcc taccagatg aggtgagggg tgtttttgcc 2400
gtgaggactg ggaaatacaa agctcatitt tttaccagg gatcagctca ttcagacacc 2460
acagctgata ctgcctgtca tgccagcagt agcttgacag cacatgagcc tcccttactg 2520
tatgacctga gcaaggaccc aggggagaac tataacctgc ttgggggggt tgctggggcc 2580
accccagaag tgcttcaggc actaaagcag ctgcaactgc ttaaagcaca gttggatgct 2640
gcagtgaact ttggcccttc ccaggtggcc agaggcgagg atcccgccct gcagatctgc 2700
tgccacccag gctgcacacc cagacctgcc tgctgtcact gccccgaccc acacgccggc 2760
agcggagcta ctaacttcag cctgctgaag caggtggag acgtggagga gaacctgga 2820
cctatggctg cccagccct ggggctggtg tgtggcagat gccctgagct gggcctggtg 2880
ctgcttctcc tgctgtgag cctcctgtgt ggtgctgctg gctctcagga agcagggaca 2940
ggagcaggag caggttctct ggtggtctca tgcggttgtg ggacccccca gaggccaggg 3000
gctcatgggt cctctgcagc tgcccacagg tactcaaggg aagcaaagtc ccctggcccc 3060
gtacctgggg aaaggcaact tgctcactcc aagatggttc ctatccctgc aggagttttt 3120
actatgggaa ctgatgaccc tcagatcaag caggatggtg aagcaccagc taggagagtc 3180
acaattgatg ctttctatat ggatgcctat gaagtgtcaa acacagaatt tgagaaattt 3240
gtaaacagca ctggatacct tacagaggct gagaaatttg gtgacagttt tgtttttgaa 3300
ggcatgctaa gtgagcaggc gaagaccaat atccaacagg cagtggctgc agccccctgg 3360
tggtgcctg ttaaaggagc caattggaga caccagagg gaccagactc aactatcctc 3420
cacaggcctg accaccctgt gctgcatgtg tcctggaatg atgcagtggc atactgcacc 3480
tggtgctgga aaagggtacc aacagaggca gaatgggagt attcctgccg ggggtgactg 3540
cacaacagac tgttcccctg gggcaataag ctgcaaccta aaggacagca ttatgccaat 3600
atttggcagg gagagtccc agtcacaaac actggtgagg atggcttcca gggaactgcc 3660
cctgtggatg ctttcccacc caatggctat gggttgtaca atatagttgg gaatgcctgg 3720
gagtggactt ctgactggtg gacgttccat cacagtgtgg aagagacact gaacccaaag 3780
gggccccct caggcaagga cagagtcaag aaaggtggct cttatatgtg tcacagaagc 3840
tattgctaca gatataggtg tgctgcaaga agtcagaaca cccctgacag ctgagctagc 3900
aatctgggat ttagatgtgc agcagataga ctccccacca tggactgaga tccagacatg 3960
ataagataca ttgatgagtt tggacaaacc acaactagaa tgcagtgaag aaaatgcttt 4020
atttgtgaaa tttgtgatgc tattgttta tttgtaacca ttataagctg caataaaca 4080
gttaacaaca acaattgcat tcattttatg tttcaggttc agggggagggt gtgggagggt 4140
ttttaaacct gcaggcttag atacgtagat aagtagcatg gcgggttaat cattaactac 4200
aaggaacccc tagtgatgga gttggccact ccctctctgc gcgctcgctc gctcactgag 4260
gccgggcgac caaaggtcgc ccgacgcccg ggctttgccc gggcggcctc agtgagcgag 4320
cgagcgcgca gagagggagt ggccaa 4346

```

<210> 69

<211> 4492

<212> DNA

<213> 人工序列

<220>

<223> 合成多核苷酸

<400> 69

```
ttggccactc cctctctgcg cgctcgctcg ctactgagg ccgggcgacc aaaggtcgcc 60
cgacgccccg gctttgcccg ggcggcctca gtgagcgagc gagcgcgag agagggagtg 120
gccaactcca tcactagggg ttcctggagg ggtggagtcg tgacgtgaat tacgtcatag 180
ggttagggag gtcctgcata tgcggccgca cctaggtcat tctggcctcc ccctccctca 240
aggccagtca ttctggcctg tccttccccg aaggccagtc attctggcct cccctcccc 300
caaggccagt cattctggcc ttccctccc ttaaggccag agtactatcg attcacacaa 360
aaaaccaaca cactattgca atgaaaataa atttcttta ttaagcttaa ttcggtaccc 420
tagttattaa tagtaatcaa ttacggggtc attagttcat agcccatata tggagttccg 480
cgttacataa cttacggtaa atggcccgcc tggtgaccg cccaacgacc cccgcccatt 540
gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca 600
atgggtggac tatttacggt aaactgccc cttggcagta catcaagtgt atcatatgcc 660
aagtagccc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgccagta 720
catgacctta tgggactttc ctacttgga gtacatctac gtattagtca tcgctattac 780
catggtcgag gtgagcccca cgttctgctt cactctcccc atctcccccc cctccccacc 840
cccaattttg tatttattta ttttttaatt attttgtgca gcgatggggg cggggggggg 900
ggggggggcg gcgccaggcg gggcggggcg gggcgagggg cggggcgggg cgaggcggag 960
aggtgcggcg gcagccaatc agagcggcg gctccgaaag tttcctttta tggcgaggcg 1020
gcggcggcg cgccctata aaaagcgaag cgcgcgggcg gcgggagtcg ctgcgacgt 1080
gccttcgccc cgtgccccgc tccgccgcc cctcgcgcc cccgccccgg ctctgactga 1140
ccgcgttact cccacagtg agcgggcggg acggcccttc tcctccgggc tgtaattagc 1200
gcttggttta atgacggctt gtttctttc tgtggctgcg tgaaagcctt gaggggctcc 1260
gggagctaga gcctctgcta accatgttca tgcttcttc ttttctctac agctcctggg 1320
caacgtgctg gttattgtgc tgtctcatca ttttgcaaa ggctagcgcc gccaccatga 1380
gcatgggcgc cccagaagc ctgttacttg ctttagctgc tggccttgca gtggcaaggc 1440
cccctaacat cgtgctgac tttgcagatg acttgggata tggggatctt ggttggtatg 1500
gccacccatc aagcacaaact cccaatctgg atcagttggc tgcaggaggt ctgaggttta 1560
cagactttta tgttccagtc tcctgtgca ctcttctcg ggctgccctg cttactggga 1620
ggctccctgt gagaatgggt atgtaccctg gagtgttggg cccatccagc aggggagggc 1680
tgcccctgga agaggtgaca gtggcagagg tgctggcagc acgaggctat ctgactggca 1740
tggcaggcaa gtggcacctg ggtgtagggc cagagggtgc tttctgcct ccccatcagg 1800
gctttcatag gtttctggga atcccatact ctcatgacca aggacctgc cagaacctca 1860
cctgtttccc ccctgcaaca ccatgtgatg ggggctgtga tcaaggtctg gttctatac 1920
cactgcttgc taatctttca gtggaagctc aaccacctg gctgcctggc ttggaggcta 1980
gatacatggc cttcgacat gatctgatg cagatgccc gagacaagat aggcctttct 2040
tcctctacta tgcatctcac cacaccact atcctcagtt ctcaggccaa tcatttgctg 2100
agcgtagtgg caggggccc tttggggaca gtttgatgga actggatgcc gcagttggta 2160
```

```

ccctcatgac agcaataggg gacttaggtt tgctggagga aacattggta attttcacag 2220
ctgataatgg ccctgagaca atgagaatgt ctaggggagg ctgctctggt cttctgaggt 2280
gtggtaaagg gactacatat gagggaggag tgaggggaacc agctcttgcc ttttgccag 2340
gtcacatagc ccctggagtt acacatgaac tagcttcttc cctggacttg cttcctacac 2400
tggcagccct ggcaggtgcc cctctcccta atgtaacttt agatggattt gacctctctc 2460
cactactttt agggacaggg aaaagtccaa ggcagtcctt attcttctat ctttcctacc 2520
cagatgaggt gaggggtgtt tttgccgtga ggactgggaa atacaaagct cattttttta 2580
cccagggatc agctcattca gacaccacag ctgatcctgc ctgtcatgcc agcagtagct 2640
tgacagcaca tgagcctccc ttactgtatg acctgagcaa ggaccaggg gagaactata 2700
acctgcttgg gggggttgtc ggggccaccc cagaagtgtc tcaggcacta aagcagctgc 2760
aactgcttaa agcacagttg gatgtgcag tgacctttgg ccttcccag gtggccagag 2820
gcgaggatcc cgccctgcag atctgtgcc acccaggtg cacaccaga cctgcctgct 2880
gtcactgccc cgaccacac gccggcagcg gagctactaa cttcagcctg ctgaagcagg 2940
ctggagacgt ggaggagaac cctggaccta tggctgcccc agccctgggg ctggtgtgtg 3000
gcagatgccc tgagctgggc ctggtgtgtc ttctctgct gctgagcctc ctgtgtggtg 3060
ctgctggctc tcaggaagca gggacaggag caggagcagg ttctctggct ggctcatgcg 3120
gttgtgggac ccccagagg ccaggggtc atgggtcctc tgcagctgcc cacaggtact 3180
caagggaagc aaatgcccct ggccccgtac ctggggaaag gcaacttgct cactccaaga 3240
tggttcctat ccctgcagga gtttttacta tgggaactga tgaccctcag atcaagcagg 3300
atggtgaagc accagctagg agagtcacaa ttgatgcctt ctatatgat gcctatgaag 3360
tgtcaaacac agaatttgag aaatttgtaa acagcactgg ataccttaca gaggtgaga 3420
aatttggtga cagttttgtt tttgaaggca tgctaagtga gcaggtgaag accaatatcc 3480
aacaggcagt ggctgcagcc ccctggtggc tgcctgttaa aggagccaat tggagacacc 3540
cagagggacc agactcaact atcctccaca ggcctgacca ccctgtgctg catgtgtcct 3600
ggaatgatgc agtggcatac tgcacctggg ctgggaaaag gttaccaaca gaggcagaat 3660
gggagtattc ctgccggggt ggactgcaca acagactgtt cccctggggc aataagctgc 3720
aacctaaagg acagcattat gccaatattt ggcagggaga gttcccagtc acaaactg 3780
gtgaggatgg cttccaggga actgcccctg tggatgcttt cccaccaat ggctatgggt 3840
tgtacaatat agttgggaat gcctgggagt ggacttctga ctggtggacg gtccatcaca 3900
gtgtggaaga gacactgaac ccaaaggggc cccctcagg caaggacaga gtcaagaaag 3960
gtggctctta tatgtgtcac agaagctatt gctacagata taggtgtgct gcaagaagtc 4020
agaacacccc tgacagctca gctagcaatc tgggatttag atgtgcagca gatagactcc 4080
ccaccatgga ctgagatcca gacatgataa gatacttga tgagtttgga caaaccacaa 4140
ctagaatgca gtgaaaaaaaa tgctttattt gtgaaatttg tgatgctatt gctttatttg 4200
taaccattat aagctgcaat aaacaagtta acaacaacaa ttgcattcat tttatgtttc 4260
aggttcaggg ggaggtgtgg gaggtttttt aaacctgcag gtctagatac gtagataagt 4320
agcatggcgg gttaatcatt aactacaagg aaccctagt gatggagttg gccactccct 4380
ctctgcgcgc tcgctcgctc actgaggccg ggcgaccaa ggtcgcccga cgccgggct 4440
ttgcccgggc ggcctcagtg agcgagcgag cgcgagaga gggagtggcc aa 4492

```

<210> 70

<211> 7537

<212> DNA

<213> 人工序列

<220>

<223> 合成多核苷酸

<400> 70

```

aaagcgggca gtagcgcaa cgcaattaat gtaggttagc tctactcatta ggcaccccag 60
gctttacact ttatgcttcc ggctcgtatg ttgtgtggaa ttgtgagcgg ataacaattt 120
cacacaggaa acagctatga ccatgattac gccaaagctta gatccccggg taccgagctc 180
gaattcactg gccgtcgttt tacaacgtcg tgactgggaa aaccttggcg ttaccaact 240
taatcgccct gcagcacatc cccctttcgc cagctggcgt aatagcgaag aggcccgcac 300
cgatcgccct tcccaacagt tgcgcagcct gaatggcgaa tggcgccctga tgcggtattt 360
tctccttacg catctgtgcg gtatttcaca ccgcatatgg tgcactctca gtacaatctg 420
ctctgatgcc gcatagttaa gccagccccg acacccgcca acacccgtg acgcgccctg 480
acgggcttgt ctgctcccgg catccgctta cagacaagct gtgaccgtct ccgggagctg 540
catgtgtcag aggttttcac cgtcatcacc gaaacgcgcg agacgaaagg gcctcgtgat 600
acgcctattt ttataggtta atgtcatgat aataatggtt tcttagacgt caggtggcac 660
ttttcgggga aatgtggcat gcctgcattt ggccactccc tctctgcgcg ctgcctcgt 720
cactgaggcc gggcgaccaa aggtcgcccc acgcccgggc tttgcccggg cggcctcagt 780
gagcgagcga gcgcgcagag agggagtggc caactccatc actaggggtt cctggagggg 840
tgtagtcgtg acgtgaatta cgtcataggg ttagggaggt cctgcatatg cggccgcggc 900
atcctaaaaa atattcagtg gaaacgtaaa aacattaaag actgattaaa catcgagca 960
tgacacagat ttagcaactg agcataaata atttgactcg gatactgctc caaaatccga 1020
agaggaccaa tttcttccag gaggacaact acctcgctct ctgcagacct ctctcctcgg 1080
cagctgaagg agtgtggcca atctgcctcc acctccccgc ggacccccta ctctcaggac 1140
ctcctgcagc accccaaact ggaagtggcc gctgcagacc caaggacgag gggcacgcgg 1200
gagccggcag ccctagtgga gcggttgag atgttgaggt gggagggtca cccaggtggg 1260
gtgaggctgg gtaggtagc ggagtgaacg gcttccgaag ctctgggccg ccccaggtt 1320
ggactaagca ggcgctctgt cttcgcccc gccagggtg ggcgtctcct gaggactccc 1380
cgccacacct gacccgagac cgcgcgccc gccatgaacg cttccccgac ccagcgtagg 1440
gcccgccgca ctggcgggag agggtcggcg ggaggcctgg cgaacccggg ggcgggacca 1500
ggcgggcaag gcccggctgc cgcagcgccg ctctgcgcga ggcggctccg ccgcggcgga 1560
gggatacggc gcacatatata tatatcgcg ggcgagact cgcgctccg cagtgggtgct 1620
gggagtgtcg tggacgccgt gccgttactc gtagtcaggc ggcggcgag gcggcgcg 1680
cgccatagcg cacagcgcg cttagcagca gcagcagcag cagcggcatc ggaggtaccc 1740
ccgccgtcgc agcccccgcg ctggtgcagc caccctcgt ccctctgctc ttctctcctt 1800
cgctcgcacc aagaggtaag ggtttaagg atggttggtt ggtggggtat taatgtttaa 1860
ttacctggag cacctgcctg aaatcacttt ttttcaggtt gggccacccg ccgccacat 1920

```

gagcatgggc gccccagaa gcctgttact tgcttttagct gctggccttg cagtggcaag 1980
 gccccctaac atcgtgctga tctttgcaga tgacttggga tatggggatc ttggttgta 2040
 tggccaccca tcaagcacia ctcccaatct ggatcagttg gctgcaggag gtctgaggtt 2100
 tacagacttt tatgttccag tctccctgtg cactccttct cgggctgccc tgcttactgg 2160
 gaggctccct gtgagaatgg gtatgtaccc tggagtgttg gtcccatcca gcaggggagg 2220
 gctgccccctg gaagaggtga cagtggcaga ggtgctggca gcacagggt atctgactgg 2280
 catggcaggc aagtggcacc tgggtgtagg gccagagggt gctttcctgc ctcccatca 2340
 gggctttcat aggtttctgg gaatcccata ctctcatgac caaggaccct gccagaacct 2400
 cacctgtttc cccctgcaa caccatgtga tgggggctgt gatcaaggct tggttcctat 2460
 accactgctt gctaattctt cagtggagc tcaaccaccc tggctgcctg gcttggaggc 2520
 tagatacatg gccttcgcac atgatctgat ggcagatgcc cagagacaag ataggccttt 2580
 cttcctctac tatgcatctc accacaccca ctatctcag ttctcaggcc aatcatttgc 2640
 tgagcgtagt ggcaggggcc catttgggga cagtttgatg gaactggatg ccgcagttgg 2700
 taccctcatg acagcaatag gggacttagg tttgctggag gaaacattgg taattttcac 2760
 agctgataat ggccctgaga caatgagaat gtctagggga ggctgctctg gtcttctgag 2820
 gtgtggtaaa gggactacat atgagggagg agtgagggaa ccagctcttg ctttttgccc 2880
 aggtcacata gcccctggag ttacacatga actagcttct tccctggact tgcttctac 2940
 actggcagcc ctggcagggt cccctctccc taatgtaact ttagatgat ttgacctctc 3000
 tccactactt ttagggacag ggaaaagtcc aaggcagtc ttattcttct atccttctca 3060
 cccagatgag gtgaggggtg tttttgccgt gaggactggg aaatacaaag ctcatttttt 3120
 taccagggga tcagctcatt cagacaccac agctgatcct gcctgtcatg ccagcagtag 3180
 cttgacagca catgagcctc cttactgta tgacctgagc aaggaccag gggagaacta 3240
 taacctgctt ggggggggtg ctggggccac cccagaagt ctacaggcac taaagcagct 3300
 gcaactgctt aaagcacagt tggatgctgc agtgacctt ggcccttccc aggtggccag 3360
 aggcgaggat cccgccctgc agatctgctg ccaccaggc tgcacacca gacctgcctg 3420
 ctgtcactgc cccgaccac acgccggcag cggagctact aacttcagcc tgctgaagca 3480
 ggctggagac gtggaggaga accctggacc tatggctgcc ccagccctgg ggctgggtgtg 3540
 tggcagatgc cctgagctgg gcctggtgct gcttctctg ctgctgagcc tctgtgtgg 3600
 tgctgctggc tctcaggaag caggacagg agcaggagca ggttctctgg ctggctcatg 3660
 cggttggtgg accccccaga ggccaggggc tcatgggtcc tctgcagctg cccacaggta 3720
 ctcaagggaa gcaaatgccc ctggccccgt acctggggaa aggcaactg ctcactccaa 3780
 gatggttcct atccctgcag gagtttttac tatgggaact gatgacctc agatcaagca 3840
 ggatggtgaa gcaccagcta ggagagtcac aattgatgcc ttctatatgg atgcctatga 3900
 agtgtcaaac acagaatttg agaaatttgt aaacagcact ggatacctta cagaggctga 3960
 gaaatttggt gacagttttg tttttgaagg catgctaagt gagcaggtga agaccaatat 4020
 ccaacaggca gtggctgcag cccctggtg gctgcctgtt aaaggagcca attggagaca 4080
 cccagaggga ccagactcaa ctatcctcca caggcctgac caccctgtgc tgcatgtgtc 4140
 ctggaatgat gcagtggcat actgcacctg ggctgggaaa aggttaccaa cagaggcaga 4200
 atgggagtat tcctgccggg gtggactgca caacagactg tccccctggg gcaataagct 4260

gcaacctaaa ggacagcatt atgccaatat ttggcaggga gagttcccag tcacaaacac 4320
 tgggtgaggat ggcttccagg gaactgcccc tgtggatgct ttcccacca atggctatgg 4380
 gttgtacaat atagttagga atgcctggga gtggacttct gactggtgga cgggtccatca 4440
 cagtgtggaa gagacactga acccaaaggg gccccctca ggcaaggaca gagtcaagaa 4500
 aggtggctct tataatgtgtc acagaagcta ttgctacaga tataggtgtg ctgcaagaag 4560
 tcagaacacc cctgacagct cagctagcaa tctgggattt agatgtgcag cagatagact 4620
 ccccaccatg gactgagatc cagacatgat aagatacatt gatgagtttg gacaaaccac 4680
 aactagaatg cagtgaaaaa aatgctttat ttgtgaaatt tgtgatgcta ttgctttatt 4740
 tgtaaccatt ataagctgca ataaacaagt taacaacaac aattgcattc attttatgtt 4800
 tcaggttcag ggggaggtgt gggaggtttt ttaaactgc aggtctagat acgtagataa 4860
 gtagcatggc ggggttaatca ttaactacaa ggaacccta gtgatggagt tggccactcc 4920
 ctctctgctc gctcgtcctc tcaactgaggc cgggcgacca aaggctcgcc gacgcccggg 4980
 ctttgcccgg gcggcctcag tgagcgagcg agcgcgaga gagggagtgg ccaaagatcc 5040
 ccgggtaccg agctcgaatt cactggccgt cgttttacaa cgtcgtgact gggaaaaccc 5100
 tggcgttacc caacttaatc gccttgacgc acatccccct ttcgccagct ggcgtaatat 5160
 cgaagaggcc cgcaccgatc gcccttccca acagttgcgc agcctgaatg gcgaatggcg 5220
 cctgatgcgg tattttctcc ttacgcatct gtgcggtatt tcacaccgca tatggtgcac 5280
 tctcagtaca atctgctctg atgccgcata gttaagccag ccccgacacc cgccaacacc 5340
 cgctgacgcg ccctgacggg cttgtctgct cccggcatcc gcttacagac aagctgtgac 5400
 cgtctccggg agctgcatgt gtcagaggtt ttcaccgtca tcaccgaaac gcgcgagacg 5460
 aaagggcctc gtgatacgcc tatttttata ggtaaatgtc atgataataa tggttttcta 5520
 gacgtcaggt ggcacttttc ggggaaatgt gcgcggaacc cctatttggt tatttttcta 5580
 aatacattca aatatgtatc cgctcatgag acaataaccc tgataaatgc ttcaataata 5640
 ttgaaaaagg aagagtatga gtattcaaca tttccgtgtc gcccttattc ctttttttgc 5700
 ggcattttgc cttcctgttt ttgctcacc agaaacgctg gtgaaagtaa aagatgctga 5760
 agatcagttg ggtgcacgag tgggttacat cgaactggat ctcaacagcg gtaagatcct 5820
 tgagagtttt cgccccgaag aacgttttcc aatgatgagc acttttaaag ttctgctatg 5880
 tggcgcggta ttatcccgtt ttgacgccgg gcaagagcaa ctcggtcgcc gcatacacta 5940
 ttctcagaat gacttggtg agtactcacc agtcacagaa aagcatctta cggatggcat 6000
 gacagtaaga gaattatgca gtgctgccat aaccatgagt gataaactg cggccaactt 6060
 acttctgaca acgatcggag gaccgaagga gctaaccgt tttttgcaca acatggggga 6120
 tcatgtaact cgccttgatc gttgggaacc ggagctgaat gaagccatac caaacgacga 6180
 gcgtgacacc acgatgcctg tagcaatggc aacaacgttg cgcaactat taactggcga 6240
 actacttact ctagcttccc ggcaacaatt aatagactgg atggaggcgg ataaagttgc 6300
 aggaccactt ctgcgtcgg cccttccggc tggctggttt attgctgata aatctggagc 6360
 cggtgagcgt ggggtctcgc gtatcattgc agcactgggg ccagatggta agccctcccg 6420
 tatcgtagtt atctacacga cggggagtca ggcaactat gatgaacgaa atagacagat 6480
 cgctgagata ggtgcctcac tgattaagca ttgtaactg tcagaccaag tttactcata 6540
 tatacttttag attgatttaa aacttcattt ttaatttaaa aggatctagg tgaagatcct 6600

ttttgataat ctcatgacca aaatccctta acgtgagttt tcgttccact gagcgtcaga 6660
 ccccgtagaa aagatcaaag gatcttcttg agatcctttt tttctgcgcg taatctgctg 6720
 cttgcaaaca aaaaaaccac cgctaccagc ggtggtttgt ttgccgcatc aagagctacc 6780
 aactcttttt ccgaaggtaa ctggcttcag cagagcgagc ataccaaata ctgttcttct 6840
 agtgtagccg tagttaggcc accacttcaa gaactctgta gcaccgccta catacctcgc 6900
 tctgctaata ctgttaccag tggctgctgc cagtggcgat aagtcgtgtc ttaccgggtt 6960
 ggactcaaga cgatagttac cggataaggc gcagcggtcg ggctgaacgg ggggttcgtg 7020
 cacacagccc agcttggagc gaacgaccta caccgaactg agatacctac agcgtgagct 7080
 atgagaaaagc gccacgcttc ccgaaggag aaaggcggac aggtatccgg taagcggcag 7140
 ggtcggaaca ggagagcgca cgaggagct tccaggggga aacgcctggt atctttatag 7200
 tcctgtcggg tttcgccacc tctgacttga gcgtcgattt ttgtgatgct cgtcaggggg 7260
 gcggagccta tggaaaaacg ccagcaacgc ggctttttta cggttcctgg ccttttgctg 7320
 gccttttgct cacatgttct ttctgcgtt atccctgat tctgtggata accgtattac 7380
 cgcctttgag tgagctgata ccgctcgccg cagccgaacg accgagcgca gcgagtcagt 7440
 gagcgaggaa gcggaagagc gcccaatacg caaacgcct ctccccgcgc gttggccgat 7500
 tcattaatgc agctggcacg acaggtttcc cgactgg 7537

<210> 71

<211> 6335

<212> DNA

<213> 人工序列

<220>

<223> 合成多核苷酸

<400> 71

gtcagggtggc acttttcggg gaaatgtggc atgcctgcat ttggccactc cctctctgcg 60
 cgctcgctcg ctactgagg ccgggcgacc aaaggtcgcc cgacgcccgg gctttgccc 120
 ggcgccctca gtgagcgagc gagcgcgag agaggagtg gccaaactca tctactaggg 180
 ttcttgagg ggtggagtcg tgacgtgaat tacgtcatag ggttagggag gtctgcata 240
 tgcggccgca cctaggtcat tctggcctcc ccctccctca aggccagtca ttctggcctg 300
 tccttccccg aaggccagtc attctggcct cccctcccc caaggccagt cattctggcc 360
 ttccccctcc ttaaggccag agtactatcg attcacacaa aaaaccaaca cactattgca 420
 atgaaaataa atttccttta ttaagcttaa ttcggtaccc tagttattaa tagtaatcaa 480
 ttacggggtc attagttcat agcccatata tggagttccg cgttacataa cttacggtaa 540
 atgccccgcc tggctgaccg cccaacgacc cccgccatt gacgtcaata atgacgtatg 600
 ttcccatagt aacgccaata gggactttcc attgacgtca atgggtggac tatttacggt 660
 aaactgccc cttggcagta catcaagtgt atcatatgcc aagtacgcc cctattgacg 720
 tcaatgacgg taaatggccc gcctggcatt atgccagta catgacctta tgggactttc 780
 ctacttgga gtacatctac gtattagtca tcgctattac catggtcgag gtgagcccca 840
 cgttctgctt cactctcccc atctcccccc cctccccacc cccaattttg tattttattta 900
 ttttttaatt attttgtgca gcgatggggg cggggggggg gggggggcgc gcgccaggcg 960

```

gggcggggcg gggcgagggg cggggcgggg cgaggcggag aggtgcggcg gcagccaatc 1020
agagcggcgc gctccgaaag tttcctttta tggcgaggcg gcggcgggcg cggccctata 1080
aaaagcgaag cgcgcgggcg gcgggagtcg ctgcgacgct gccttcgccc cgtgccccgc 1140
tccgccgccg cctcgcgccg cccgccccgg ctctgactga ccgcgttact cccacaggtg 1200
agcgggcggg acggcccttc tcctccgggc tgtaattagc gcttggttta atgacggctt 1260
gtttcttttc tgtggctgcg tgaaagcctt gaggggctcc gggagctaga gcctctgcta 1320
accatgttca tgccttcttc tttttcctac agctcctggg caacgtgctg gttattgtgc 1380
tgtctcatca ttttgcaaaa ggctagcgcc gccacatga gcatgggcg cccagaagc 1440
ctgttacttg ctttagctgc tggccttgca gtggcaaggc cccctaacat cgtgctgac 1500
tttgcagatg acttgggata tggggatctt ggttggtatg gccacccatc aagcacaact 1560
cccaatctgg atcagttggc tgcaggaggt ctgaggttta cagactttta tgttcagtc 1620
tccctgtgca ctcttctcg ggtgccttg cttactggga ggctccctgt gagaatgggt 1680
atgtaccctg gagtgttggc cccatccagc aggggaggcg tgcccctgga agaggtgaca 1740
gtggcagagg tgctggcagc acgaggtat ctgactggca tggcaggcaa gtggcacctg 1800
ggtgtagggc cagagggtgc tttcctgcct ccccatcagg gctttcatag gtttctggga 1860
atcccatact ctcatgacca aggaccctgc cagaacctca cctgtttccc ccctgcaaca 1920
ccatgtgatg ggggctgtga tcaaggtctg gttcctatac cactgcttgc taatctttca 1980
gtggaagctc aaccaccctg gctgcctggc ttggaggcta gatacatggc cttcgcacat 2040
gatctgatgg cagatgccc gagacaagat aggcctttct tcctctacta tgcattctac 2100
cacaccact atcctcagtt ctgaggcaa tcatttgctg agcgtagtgg caggggccc 2160
tttggggaca gtttgatgga actggatgcc gcagttggta ccctcatgac agcaatagg 2220
gacttaggtt tgctggagga aacattggta attttcacag ctgataatgg ccctgagaca 2280
atgagaatgt ctaggggagg ctgctctggc cttctgaggt gtggtaaagg gactacatat 2340
gagggaggag tgagggaacc agctcttgcc ttttgccag gtcacatagc ccctggagtt 2400
acacatgaac tagcttcttc cctggacttg cttcctacac tggcagccct ggcaggtgcc 2460
cctctcccta atgtaacttt agatggattt gacctctctc cactactttt agggacagg 2520
aaaagtccaa ggcagtcctt attcttctat ctttctacc cagatgaggt gaggggtgtt 2580
tttgccgtga ggactgggaa atacaaagct cattttttta cccagggatc agctcattca 2640
gacaccacag ctgacctgc ctgtcatgcc agcagtagct tgacagcaca tgagcctccc 2700
ttactgtatg acctgagcaa ggaccaggg gagaactata acctgcttgg gggggttgct 2760
ggggccaccc cagaagtgtc tcaggcacta aagcagctgc aactgcttaa agcacagttg 2820
gatgctgcag tgaccttgg ccttcccag gtggccagag gcgaggatcc cgccctgcag 2880
atctgctgcc acccaggtg cacaccaga cctgcctgct gtcactgccc cgaccacac 2940
gccggcagcg gagtactaa cttcagcctg ctgaagcagg ctggagacgt ggaggagaac 3000
cctggacctg tggctgcccc agccctgggg ctggtgtgtg gcagatgccc tgagctgggc 3060
ctggtgctgc ttctctgct gctgagctc ctgtgtggtg ctgctggctc tcaggaagca 3120
gggacaggag caggagcagg ttctctggct ggctcatgcg gttgtgggac ccccagagg 3180
ccaggggctc atgggtcctc tgcagctgcc cacaggtact caagggaagc aaatgccct 3240
ggccccgtac ctggggaaa gcaacttgct cactccaaga tggttcctat ccctgcagga 3300

```

gtttttacta tgggaactga tgaccctcag atcaagcagg atggtgaagc accagctagg 3360
 agagtcacaa ttgatgcctt ctatatggat gcctatgaag tgtcaaacac agaatttgag 3420
 aaatttgtaa acagcactgg ataccttaca gaggctgaga aatttggtga cagttttgtt 3480
 tttgaaggca tgctaagtga gcaggtgaag accaatatcc aacaggcagt ggctgcagcc 3540
 ccctgggtggc tgcctgttaa aggagccaat tggagacacc cagagggacc agactcaact 3600
 atcctccaca ggcctgacca ccctgtgctg catgtgtcct ggaatgatgc agtggcatac 3660
 tgcacctggg ctgggaaaaag gttaccaaca gaggcagaat gggagtattc ctgccggggt 3720
 ggactgcaca acagactgtt cccctggggc aataagctgc aacctaaagg acagcattat 3780
 gccaatattt ggcagggaga gttcccagtc acaaacactg gtgaggatgg cttccaggga 3840
 actgcccctg tggatgcttt cccaccaat ggctatgggt tgtacaatat agttgggaat 3900
 gcctgggagt ggacttctga ctggtggacg gtccatcaca gtgtggaaga gacactgaac 3960
 ccaaaggggc cccctcagg caaggacaga gtcaagaaag gtggctctta tatgtgtcac 4020
 agaagctatt gctacagata taggtgtgct gcaagaagtc agaacacccc tgacagctca 4080
 gctagcaatc tgggatttag atgtgcagca gatagactcc ccacatgga ctgagatcca 4140
 gacatgataa gatacattga tgagtttga caaaccacaa ctagaatgca gtgaaaaaaa 4200
 tgctttattt gtgaaatttg tgatgctatt gctttatttg taaccattat aagctgcaat 4260
 aaacaagtta acaacaacaa ttgcattcat tttatgtttc aggttcaggg ggaggtgtgg 4320
 gaggtttttt aaacctgcag gtctagatac gtagataagt agcatggcgg gttaatcatt 4380
 aactacaagg aacccctagt gatggagttg gccactccct ctctgcgcgc tcgctcgctc 4440
 actgaggccg ggcgaccaa ggtcgccga cgccgggct ttgcccgggc ggcctcagt 4500
 agcgagcgag cgcgcagaga gggagtggcc aaagatcccc gggtagcgag gacgaattct 4560
 ctagatatcg ctcaatactg accatttaaa tcatactga cctccatagc agaaagtcaa 4620
 aagcctccga cgggaggctt ttgacttgat cggcacgtaa gaggttccaa ctttcacat 4680
 aatgaaataa gatcactacc gggcgtattt tttgagttat cgagattttc aggagctaag 4740
 gaagctaaaa tgagccatat tcaacgggaa acgtcttgct cgaggccgcg attaaattcc 4800
 aacatggatg ctgatttata tgggtataaa tgggctcgcg ataatgtcgg gcaatcaggt 4860
 gcgacaatct atcgattgta tgggaagccc gatgcgccag agttgtttct gaaacatggc 4920
 aaaggtagcg ttgccaatga tgttacagat gagatggtca ggctaaactg gctgacggaa 4980
 tttatgcctc ttccgaccat caagcatttt atccgtactc ctgatgatgc atggttactc 5040
 accactgcga tcccagggaa aacagcattc caggtattag aagaatatcc tgattcaggt 5100
 gaaaaatattg ttgatgcgct ggagtggtc ctgcgccggg tgcattcgat tcctgtttgt 5160
 aattgtcctt ttaacggcga tcgcgtattt cgtctcgctc aggcgcaatc acgaatgaat 5220
 aacggtttgg ttggtgcgag tgattttgat gacgagcgta atggctggcc tgttgaacaa 5280
 gtctggaaaag aaatgcataa gcttttgcca ttctcaccgg attcagtcgt cactcatggt 5340
 gattttctcac ttgataacct tatttttgac gaggggaaat taataggttg tattgatgtt 5400
 ggacgagtcg gaatcgcaga ccgataccag gatcttgcca tcctatggaa ctgcctcggt 5460
 gagttttctc cttcattaca gaaacggctt tttcaaaaat atggtattga taatcctgat 5520
 atgaataaat tgcagtttca cttgatgctc gatgagtttt tctgagggcc caaatgtaat 5580
 cacctggctc accttcgggt gggcctttct gcgttgctgg cgtttttcca taggctccgc 5640

ccccctgacg agcatcacia aaatcgatgc tcaagtcaga ggtggcgaaa cccgacagga 5700
ctataaagat accaggcggt tccccctgga agctccctcg tgcgctctcc tgttccgacc 5760
ctgccgctta ccggatacct gtccgccttt ctcccttcgg gaagcgtggc gctttctcat 5820
agctcacgct gtaggtatct cagttcggtg taggtcgttc gctccaagct gggctgtgtg 5880
cacgaacccc ccgttcagcc cgaccgctgc gccttatccg gtaactatcg tcttgagtcc 5940
aaccgggtaa gacacgactt atcgccactg gcagcagcca ctggtaacag gattagcaga 6000
gcgaggatg taggcgggtg tacagagttc ttgaagtggg ggcctaacta cggctacact 6060
agaagaacag tatttggtat ctgcgctctg ctgaagccag ttacctcgga aaaagagttg 6120
gtagctcttg atccggcaaa caaaccaccg ctggtagcgg tggttttttt gtttgcaagc 6180
agcagattac gcgcagaaaa aaaggatctc aagaagatcc tttgattttc taccgaagaa 6240
aggcccaccc gtgaaggatga gccagtgaat tgattgcagt ccagttacgc tggagtctga 6300
ggctcgtcct gaatgatata aagcttgaat tcgtt 6335

<210> 72

<211> 1527

<212> DNA

<213> 人工序列

<220>

<223> 合成多核苷酸

<400> 72

atgtctatgg gggctcctcg ctccctgctg ctggcactgg ccgccgggct ggctgtcgca 60
agaccaccta atatcgtcct gatttttgca gacgatctgg gatacggcga cctgggatgc 120
tatggccacc caagctccac cacaccaac ctggaccagc tggcagcagg aggctgcgg 180
ttcaccgact tctacgtgcc agtgagcctg tgcacccct ccagagccgc cctgctgaca 240
ggcaggctgc cagtgcgcat gggcatgtat cctggcgtgc tggtgccatc tagcaggggc 300
ggcctgccac tggaggaggt gaccgtggca gaggtgctgg cagccagagg ctacctgaca 360
ggaatggccg gcaagtggca cctgggagtg ggaccagagg gaggcttcct gcccctcac 420
cagggcttcc accggtttct gggcatccct tattctcacg accagggcc atgccagaac 480
ctgacctgtt ttccaccagc aacaccatgc gacggaggat gtgatcaggg cctggtgcca 540
atcccactgc tggcaaatct gagcgtggag gcacagcctc catggctgcc tggcctggag 600
gcaagataca tggccttcgc ccacgacctg atggcagatg cacagcggca ggatagacct 660
ttctttctgt actatgcctc ccaccacacc cactatccac agttcagcgg ccagtccttt 720
gccgagaggt ccggaagggg accattcggc gactctctga tggagctgga tgccgccgtg 780
ggcaccctga tgacagcaat cggcgacctg ggctgctgg aggagacact ggtcatcttc 840
accgccgata acggccctga gacaatgcgg atgtctagag gcggatgcag cggcctgctg 900
agatgtggca aggaaccac atacgaggga ggcgtgcgcg agcctgccct ggcatttttg 960
ccaggacaca tcgcacctgg agtgaccac gagctggcct cctctctgga cctgctgcca 1020
aactggccg ccctggcagg agcacctctg ccaaatgtga ccctggacgg cttegatctg 1080
agcccactgc tgctgggaac cggcaagtcc ctaggcagt ctctgttctt ttaccctcc 1140
tatcctgatg aggtgcgggg cgtgtttgcc gtgagaaccg gcaagtacaa gcccacttc 1200

tttacacagg gctctgcca cagcgacacc acagcagatc cagcatgcca cgccagctcc 1260
tctctgaccg cacacgagcc acctctgctg tacgacctgt ccaaggatcc cggcgagaac 1320
tataatctgc tgggaggagt ggcaggagca acccctgagg tgctgcaggc cctgaagcag 1380
ctgcagctgc tgaaggcaca gctggacgca gcagtacat tcggcccaag ccaggtggcc 1440
agaggcgagg atcccgccct gcagatctgc tgccaccag gctgcacacc cagacctgcc 1500
tgctgtcact gccccgaccc acacgcc 1527

<210> 73

<211> 237

<212> DNA

<213> 智人

<400> 73

ggggacgttt gccaggactg cattcagatg gtgactgaca tccagactgc tgtacggacc 60
aactccacct ttgtccaggc cttggtggaa catgtcaagg aggagtgtga ccgcctgggc 120
cctggcatgg ccgacatatg caagaactat atcagccagt attctgaaat tgctatccag 180
atgatgatgc acatgcaacc caaggagatc tgtgcgctgg ttgggttctg tgatgag 237

<210> 74

<211> 1833

<212> DNA

<213> 人工序列

<220>

<223> 合成多核苷酸

<400> 74

atgtctatgg gggctcctcg ctccctgctg ctggcactgg ccgccgggct ggctgtcgca 60
agaccaccta atatcgtcct gatttttgca gacgatctgg gatacggcga cctgggatgc 120
tatggccacc caagctccac cacaccaac ctggaccagc tggcagcagg aggcctgcgg 180
ttcaccgact tctacgtgcc agtgagcctg tgcacccct ccagagccgc cctgctgaca 240
ggcaggctgc cagtgcgcat gggcatgtat cctggcgctg tggtgccatc tagcaggggc 300
ggcctgccac tggaggaggt gaccgtggca gaggtgctgg cagccagagg ctacctgaca 360
ggaatggccg gcaagtggca cctgggagtg ggaccagagg gagccttcct gccccctcac 420
cagggccttc accggtttct gggcatccct tattctcacg accagggcc atgccagaac 480
ctgacctgtt ttccaccagc aacaccatgc gacggaggat gtgatcaggg cctggtgcc 540
atcccactgc tggcaaattc gagcgtggag gcacagcctc catggctgcc tggcctggag 600
gcaagataca tggccttcgc ccacgacctg atggcagatg cacagcggca ggatagacct 660
ttctttctgt actatgcctc ccaccacacc cactatccac agttcagcgg ccagtccttt 720
gccgagaggt ccggaagggg accattcggc gactctctga tggagctgga tgccgccgtg 780
ggcaccctga tgacagcaat cggcgacctg ggcctgctgg aggagacact ggatcatcttc 840
accgccgata acggccctga gacaatgcgg atgtctagag gcggatgcag cggcctgctg 900
agatgtggca agggaaccac atacgaggga ggcgtgcgag agcctgccct ggcatttttg 960
ccaggacaca tcgcacctgg agtgaccac gagctggcct cctctctgga cctgctgcc 1020

acactggccg ccctggcagg agcacctctg ccaaattgtga ccctggacgg cttcgatctg 1080
 agcccactgc tgctgggaac cggcaagtcc ctaggcagt ctctgttctt ttaccctcc 1140
 tatcctgatg aggtgcgggg cgtgtttgcc gtgagaaccg gcaagtacaa ggcccacttc 1200
 ttacacagg gctctgcccc cagcgacacc acagcagatc cagcatgcca cgccagctcc 1260
 tctctgaccg cacacgagcc acctctgctg tacgacctgt ccaaggatcc cggcgagaac 1320
 tataatctgc tgggaggagt ggcaggagca acccctgagg tgctgcaggc cctgaagcag 1380
 ctgcagctgc tgaaggcaca gctggacgca gcagtacat tcggcccaag ccaggtggcc 1440
 agaggcgagg atcccgccct gcagatctgc tgccaccag gctgcacacc cagacctgcc 1500
 tgctgtcact gccccgaccc acacgccggc agcggagcta ctaacttcag cctgctgaag 1560
 caggctggag acgtggagga gaacctgga cctggggacg tttgccagga ctgcattcag 1620
 atggtgactg acatccagac tgctgtacgg accaactcca cttttgtcca ggccttggtg 1680
 gaacatgtca aggaggagtg tgaccgctg ggccctggca tggccgacat atgcaagaac 1740
 tatatcagcc agtattctga aattgtatc cagatgatga tgcacatgca acccaaggag 1800
 atctgtgcgc tggttgggtt ctgtgatgag tga 1833

<210> 75

<211> 3698

<212> DNA

<213> 人工序列

<220>

<223> 合成多核苷酸

<400> 75

ggcatgtatt attgactagt tattaatagt aatcaattac ggggtcatta gttcatagcc 60
 catatatgga gttccgcgtt acataactta cggtaaattgg cccgcctggc tgaccgcccc 120
 acgacccccg cccattgacg tcaataatga cgtatgttcc catagtaacg ccaataggga 180
 ctttccattg acgtcaatgg gtggagtatt tacggtaaac tgcccacttg gcagtacatc 240
 aagtgtatca tatgccaaagt ccgcccccta ttgacgtcaa tgacggtaaa tggcccgctt 300
 ggcatattgc ccagtacatg accttacggg actttcctac ttggcagtac atctacgtat 360
 tagtcatcgc tattaccatg gtcgaggtga gcccacgtt ctgcttact ctccccatct 420
 cccccccctc cccacccccca attttgtatt tattttatctt ttaattatct tgtgcagcga 480
 tgggggcggg gggggggggg gggcgcgcg caggcggggc ggggcggggc gaggggcggg 540
 gcggggcgag gcggagaggt gcggcggcag ccaatcagag cggcgcgctc cgaaagtctc 600
 cttttatggc gaggcggcgg cggcggcggc cctataaaaa gcgaagcgcg cggcgggcgg 660
 gagtcgctgc gcgctgcctt cgccccgtgc cccgctccgc cgccgcctcg cgccgcccgc 720
 cccggctctg actgaccgcg ttactccac aggtgagcgg gcgggacggc ctttctctc 780
 cgggctgtaa ttagcgcttg gtttaatgac ggcttgcttc ttttctgttg ctgcgtgaaa 840
 gccttgaggg gctccgggag ggccctttgt gcggggggag cggctcgggg ggtgcgtgcg 900
 tgtgtgtgtg cgtggggagc gccgcgtgcg gctccgcgct gccggcggc tgtgagcgt 960
 gcgggcgcgg cgcggggctt tgtgcgtcc gcagtgtgcg cgaggggagc gcggccgggg 1020
 gcggtgcccc gcggtgcggg gggggctgcg aggggaacaa aggctgcgtg cggggtgtgt 1080

gcgtgggggg gtgagcaggg ggtgtgggcg cgtcggtcgg gctgcaaccc cccctgcacc 1140
 cccctccccg agttgtctgag cacggccccg cttegggtgc ggggctccgt acggggcgtg 1200
 gcgcgggggct cgccgtgccg ggcgggggggt ggcggcaggt gggggtgccg ggcggggcgg 1260
 ggccgcctcg ggccggggag ggctcggggg aggggcgcgg cggcccccg agcgccggcg 1320
 gctgtcgagg cgcggcgagc cgcagccatt gccttttatg gtaatcgtgc gagaggcgcg 1380
 agggacttcc tttgtcccaa atctgtgcgg agccgaaatc tgggaggcg cgccgcaccc 1440
 cctctagcgg gcgcggggcg aagcgggtcg gcgccggcag gaaggaaatg ggcggggagg 1500
 gccttcgtgc gtcgccgcgc cgccgtcccc ttctccctct ccagcctcgg ggctgtccgc 1560
 ggggggacgg ctgccttcgg gggggacggg gcagggcggg gttcggcttc tggcgtgtga 1620
 ccggcggctc tagagcctct gtaaacatg ttcatgcctt cttctttttc ctacagctcc 1680
 tgggcaacgt gctggttatt gtgtgtctc atcatttttg caaagaattc cgccaccatg 1740
 tctatggggg ctctcgtc cctgtgtctg gactggccg ccgggctggc tgtcgcaaga 1800
 ccacctataa tcgtcctgat ttttgagac gatctgggat acggcgacct gggatgctat 1860
 ggccacccaa gctccaccac acccaacctg gaccagctgg cagcaggagg cctgcggttc 1920
 accgacttct acgtgccagt gagcctgtgc accccctcca gagccgcct gctgacaggc 1980
 aggctgccag tgcgcatggg catgtatcct ggctgtctgg tgccatctag caggggcggc 2040
 ctgccactgg aggaggtgac cgtggcagag gtgctggcag ccagaggcta cctgacagga 2100
 atggccggca agtggcacct gggagtggga ccagaggag ccttcctgcc ccctcaccag 2160
 ggcttccacc ggtttctggg catcccttat tctcacgacc agggcccatg ccagaacctg 2220
 acctgttttc caccagcaac accatgcgac ggaggatgtg atcagggcct ggtgccaatc 2280
 ccaactgctg caaatctgag cgtggaggca cagcctccat ggctgcctgg cctggaggca 2340
 agatacatgg ccttcgcca cgacctgatg gcagatgcac agcggcagga tagaccttc 2400
 tttctgtact atgcctcca ccacaccac tatccacagt tcagcggcca gtcctttgcc 2460
 gagagggtcc gaaggggacc attcggcgac tctctgatgg agctggatgc cgccgtgggc 2520
 accctgatga cagcaatcgg cgacctgggc ctgctggagg agacactggt catcttcacc 2580
 gccgataacg gccctgagac aatgcggatg tctagaggcg gatgcagcg cctgctgaga 2640
 tgtggcaagg gaaccacata cgaggaggc gtgcgcgagc ctgccctggc attttgcca 2700
 ggacacatcg cacctggagt gaccacgag ctggcctcct ctctggacct gctgccaaca 2760
 ctggccgccc tggcaggagc acctctgcca aatgtgacce tggacggctt cgatctgagc 2820
 ccaactgctg tgggaaccgg caagtccct aggcagtctc tgttctttta cccctctat 2880
 cctgatgagg tgcggggcgt gtttgccgtg agaaccggca agtacaaggc ccacttcttt 2940
 acacagggt ctgcccacag cgacaccaca gcagatccag catgccacgc cagctcctct 3000
 ctgaccgcac acgagccacc tctgtgttac gacctgtcca aggatcccgg cgagaactat 3060
 aatctgctgg gaggagtggc aggagcaacc cctgaggtgc tgcaggccct gaagcagctg 3120
 cagctgctga aggcacagct ggacgcagca gtgacattcg gccaagcca ggtggccaga 3180
 ggcgaggatc ccgccctgca gatctgtgc caccaggt gcacaccag acctgcctgc 3240
 tgtactgcc ccgaccaca cgccggcagc ggagctacta acttcagct gctgaagcag 3300
 gctggagacg tggaggagaa ccctggacct ggggacgttt gccaggactg cattcagatg 3360
 gtgactgaca tccagactgc tgtacggacc aactccacct ttgtccaggc cttggtggaa 3420

catgtcaagg aggagtgtga ccgcctgggc cctggcatgg ccgacatatg caagaactat 3480
 atcagccagt attctgaaat tgctatccag atgatgatgc acatgcaacc caaggagatc 3540
 tgtgcgctgg ttgggttctg tgatgagtga actagtaact tgtttattgc agcttataat 3600
 gggtacaaat aaagcaatag catcacaaat ttcacaaata aagcattttt ttcactgcat 3660
 tctagtgttg gtttgtccaa actcatcaat gtatctta 3698

<210> 76

<211> 4231

<212> DNA

<213> 人工序列

<220>

<223> 合成多核苷酸

<400> 76

ttggccactc cctctctgcg cgtcgtctcg ctactgagg ccgggcgacc aaaggtcgcc 60
 cgacgcccgg gctttgcccg ggcggcctca gtgagcgagc gagcgcgag agagggagtg 120
 gccaaactcca tcactagggg ttcttgagg ggtggagtcg tgacgtgaat tacgtcatag 180
 ggtaggggag gtccctgcaga tcttcaatat tggccattag ccatattatt cattggttat 240
 atagcataaa tcaatatattg ctattggcca ttgcatacgt tgtatctata tcataatatg 300
 tacatttata ttggctcatg tccaatatga ccgccatgtt ggcatgtatt attgactagt 360
 tattaatagt aatcaattac ggggtcatta gttcatagcc catatatgga gttccgcgtt 420
 acataactta cggtaaattg ccgcctggc tgaccgcca acgacccccg cccattgacg 480
 tcaataatga cgtatgttcc catagtaacg ccaatagga ctttcattg acgtcaatgg 540
 gtggagtatt tacggtaaac tgcccacttg gcagtacatc aagtgtatca tatgccaagt 600
 ccgcccccta ttgacgtcaa tgacggtaaa tggccgcct ggcatatgc ccagtacatg 660
 accttacggg actttcctac ttggcagtac atctacgtat tagtcatcgc tattaccatg 720
 gtcgaggtag gccccacgtt ctgcttact ctccccatct cccccccctc cccaccccca 780
 attttgtatt tatttatatt ttaattattt tgtgcagcga tgggggcggg gggggggggg 840
 gggcgcgcg caggcggggc ggggcggggc gaggggcggg gcggggcgag gcggagaggt 900
 gcggcggcag ccaatcagag cggcgcgctc cgaaagtctt cttttatggc gaggcggcgg 960
 cggcggcggc cctataaaaa gcgaagcgcg cggcgggcgg gagtcgctgc gcgctgcctt 1020
 cgccccgtgc cccgctccgc cgccgctcgc cgccgcccgc cccggctctg actgaccgcg 1080
 ttactccac aggtgagcgg gcgggacggc ctttctctc cgggctgtaa ttagcgcttg 1140
 gttaaatgac ggcttgtttc ttttctgttg ctgcgtgaaa gccttgagg gctccgggag 1200
 ggccctttgt gcggggggag cggtcgggg ggtgcgtgcg tgtgtgtgtg cgtggggagc 1260
 gccgcgtgcg gctccgcgct gccggcggc tgtgagcgt gcgggcgcgg cgcggggctt 1320
 tgtgcgtcc gcagtgtgcg cgaggggagc gcggccgggg gcggtgcccc gcggtgcggg 1380
 gggggctgcg aggggaacaa aggtgcgtg cggggtgtgt gcgtgggggg gtgagcaggg 1440
 ggtgtgggcg cgtcggtcgg gctgcaacc cccctgcacc cccctcccc agttgctgag 1500
 cacggcccgg cttcggtgc ggggctccgt acggggcgtg gcgcggggct cgccgtgccg 1560
 ggcggggggt ggcggcaggt gggggtgccg ggcggggcgg ggccgcctcg ggccggggag 1620

```

ggctcggggg aggggcgcg cgcccccg agcgccggcg gctgtcgagg cgcggcgagc 1680
cgcagccatt gccttttatg gtaatcgtgc gagagggcgc agggacttcc tttgtcccaa 1740
atctgtgcgg agccgaaatc tgggaggcgc cgccgcaccc cctctagcgg gcgcggggcg 1800
aagcgggtgc gcgccggcag gaaggaaatg ggcggggagg gccttcgtgc gtcgccgcgc 1860
cgccgtcccc ttctccctct ccagcctcgg ggctgtccgc ggggggacgg ctgccttcgg 1920
gggggacggg gcagggcggg gttcggcttc tggcgtgtga ccggcggtc tagagcctct 1980
gctaaccatg ttcatgcctt cttctttttc ctacagctcc tgggcaacgt gctggttatt 2040
gtgctgtctc atcatttttg caaagaattc cgccaccatg tctatggggg ctccctcgctc 2100
cctgtgtctg gcactggccg ccgggctggc tgtcgcaaga ccacctaata tcgtcctgat 2160
ttttgcagac gatctgggat acggcgacct gggatgctat ggccaccaa gctccaccac 2220
accaacctg gaccagctgg cagcaggagg cctgcggttc accgacttct acgtgccagt 2280
gagcctgtgc accccctcca gagccgcct gctgacaggc aggtgccag tgcgcatggg 2340
catgtatcct ggctgtctgg tgccatctag caggggcggc ctgccactgg aggagtgac 2400
cgtggcagag gtgctggcag ccagaggcta cctgacagga atggccggca agtggcacct 2460
gggagtggga ccagagggag ccttcctgcc ccctcaccag ggcttcacc ggtttctggg 2520
catcccttat tctcacgacc agggcccatg ccagaacctg acctgttttc caccagcaac 2580
accatgcgac ggaggatgtg atcagggcct ggtgccaatc ccactgctgg caaatctgag 2640
cgtggaggca cagcctccat ggctgcctgg cctggaggca agatacatgg ccttcgcca 2700
cgacctgatg gcagatgcac agcggcagga tagaccttc tttctgtact atgcctcca 2760
ccacaccac tatccacagt tcagcggcca gtcctttgcc gagaggtccg gaaggggacc 2820
attcggcgac tctctgatgg agctggatgc cgccgtgggc accctgatga cagcaatcgg 2880
cgacctgggc ctgctggagg agacactggt catcttcacc gccgataacg gccctgagac 2940
aatgcggatg tctagaggcg gatgcagcgg cctgctgaga tgtggcaagg gaaccacata 3000
cgagggaggc gtgcgcgagc ctgccctggc attttggcca ggacacatcg cacctggagt 3060
gaccacagag ctggcctcct ctctggacct gctgccaaca ctggccgcc tggcaggagc 3120
acctctgcca aatgtgacct tggacggctt cgatctgagc ccactgctgc tgggaaccgg 3180
caagtccctt aggcagtctc tgttctttta cccctcctat cctgatgagg tgcggggcgt 3240
gtttgccgtg agaaccggca agtacaaggc ccacttcttt acacagggt ctgcccacag 3300
cgacaccaca gcagatccag catgccacgc cagctcctct ctgaccgcac acgagccacc 3360
tctgtgttac gacctgtcca aggatcccg cgagaactat aatctgctgg gaggagtggc 3420
aggagcaacc cctgaggtgc tgcaggccct gaagcagctg cagctgctga aggcacagct 3480
ggacgcagca gtgacattcg gccaagcca ggtggccaga ggcgaggatc ccgccctgca 3540
gatctgctgc caccaggt gcacaccag acctgcctgc tgtcactgcc ccgaccaca 3600
cgccggcagc ggagctacta acttcagcct gctgaagcag gctggagacg tggaggagaa 3660
ccctggacct ggggacgttt gccaggactg cattcagatg gtgactgaca tccagactgc 3720
tgtacggacc aactccacct ttgtccaggc cttggtggaa catgtcaagg aggagtgtga 3780
ccgcctgggc cctggcatgg ccgacatat caagaactat atcagccagt attctgaaat 3840
tgctatccag atgatgatgc acatgcaacc caaggagatc tgtgcgctgg ttgggttctg 3900
tgatgagtga actagtaact tgtttattgc agcttataat ggttacaat aaagcaatag 3960

```

catcacaat ttcacaaata aagcattttt ttcactgcat tctagttgtg gtttgtccaa 4020
actcatcaat gtatcttagg tctagatacg tagataagta gcatggcggg ttaatcatta 4080
actacaagga acccctagtg atggagttgg ccactccctc tctgcgcgct cgctcgctca 4140
ctgaggccgg gcgaccaaag gtcgcccgc gcccgggctt tgcccgggcg gcctcagtga 4200
gcgagcgagc gcgcagagag ggagtggcca a 4231

<210> 77

<211> 6073

<212> DNA

<213> 人工序列

<220>

<223> 合成多核苷酸

<400> 77

tttccatagg ctccgcccc ctgacgagca tcacaaaaat cgatgctcaa gtcagaggtg 60
gcgaaacccg acaggactat aaagatacca ggcgtttccc cctggaagct ccctcgtgcg 120
ctctcctgtt ccgaccctgc cgcttaccgg atacctgtcc gcctttctcc cttecggaag 180
cgtggcgctt tctcatagct cacgctgtag gtatctcagt tcggtgtagg tcgttcgctc 240
caagctgggc tgtgtgcacg aacccccctg tcagcccgcac cgctgcgcct tatccggtaa 300
ctatcgtctt gagtccaacc cggtaagaca cgacttatcg ccaactggcag cagccactgg 360
taacaggatt agcagagcga ggtatgtagg cgggtgctaca gagttcttga agtggtggcc 420
taactacggc tacactagaa gaacagtatt tggtatctgc gctctgctga agccagttac 480
ctcggaaaaa gagttggtag ctcttgatcc ggcaaaaaa ccaccgctgg tagcgggtgt 540
ttttttgttt gcaagcagca gattacgcgc agaaaaaag gatctcaaga agatcctttg 600
atcttctacc gaagaaaggc ccaccgtga aggtgagcca gtgagttgat tgcagtccag 660
ttacgctgga gtctgaggct cgtcctgaat gatatacagc ttgaattcgt gtcaggtggc 720
acttttcggg gaaatgtggc atgcctgcat ttggccactc cctctctgcg cgctcgctcg 780
ctcactgagg ccgggcgacc aaaggtcgcc cgacgcccgg gctttgcccg ggcggcctca 840
gtgagcgagc gagcgcgagc agagggagtg gccaaactcca tcaactagggg ttcttgagg 900
ggtggagtcg tgacgtgaat tacgtcatag ggtagggag gtccctgcaga tcttcaatat 960
tgccatttag ccatattatt cattgggtat atagcataaa tcaatattgg ctattggcca 1020
ttgcatacgt tgtatctata tcataatatg tacatttata ttggctcatg tccaatatga 1080
ccgccatgtt ggcattgatt attgactagt tattaatagt aatcaattac ggggtcatta 1140
gttcatagcc catatatgga gttccgcgtt acataactta cggtaaattg cccgcctggc 1200
tgaccgcca acgaccccc cccattgacg tcaataatga cgtatgttcc catagtaacg 1260
ccaataggga ctttccattg acgtcaatgg gtggagtatt tacggtaaac tgcccacttg 1320
gcagtacatc aagtgtatca tatgccaaat ccgccccta ttgacgtcaa tgacggtaaa 1380
tgggccgcct ggcattatgc ccagtacatg accttacggg acttttctac ttggcagtac 1440
atctacgtat tagtcatcgc tattaccatg gtcgaggtga gcccacgtt ctgcttact 1500
ctccccatct cccccccctc cccaccccca attttgtatt tttttatatt ttaattatatt 1560
tgtgcagcga tgggggcggg gggggggggg gggcgcgcg caggcggggc ggggcggggc 1620

gaggggcggg gcggggcgag gcggagaggt gcggcggcag ccaatcagag cggcgcgctc 1680
 cgaaagtttc cttttatggc gaggcggcgg cggcggcggc cctataaaaa gcgaagcgcg 1740
 cggcggggcg gagtcgctgc gcgctgcctt cgccccgtgc cccgctccgc cgccgcctcg 1800
 cgccgccccgc cccggctctg actgaccgcg ttactccac aggtgagcgg gcgggacggc 1860
 ccttctcctc cgggctgtaa ttagcgcttg gtttaatgac ggcttgtttc ttttctgtgg 1920
 ctgcgtgaaa gccttgaggg gctccgggag ggccctttgt gcggggggag cggctcgggg 1980
 ggtgcgtgcg tgtgtgtgtg cgtggggagc gccgcgtgcg gctccgcgct gcccgcggc 2040
 tgtgagcgct gcgggcgcgg cgcggggctt tgtgcgctcc gcagtgtgcg cgaggggagc 2100
 gcggccgggg gcggtgcccc gcggtgcggg gggggctgcg aggggaacaa aggctgcgtg 2160
 cggggtgtgt gcgtgggggg gtgagcaggg ggtgtgggag cgtcggtcgg gctgcaacct 2220
 cccctgcacc cccctccccg agttgctgag cacggcccg cttcgggtgc ggggctccgt 2280
 acggggcggt gcgcggggct cgccgtgccg ggcggggggt ggcggcaggt gggggtgccg 2340
 ggcgggggcg ggccgcctcg ggccggggag ggctcggggg aggggcgcgg cgcccccg 2400
 agcgccggcg gctgtcgagg cgcggcgagc cgcagccatt gccttttatg gtaatcgtgc 2460
 gagagggcgc agggacttcc tttgtcccaa atctgtgcgg agccgaaatc tgggagcgcg 2520
 cgccgcaccc cctctagcgg gcgcggggcg aagcgggtgc gcgccggcag gaaggaaatg 2580
 ggcggggagg gccttcgtgc gtcgccgcgc cgccgtcccc ttctccctct ccagcctcgg 2640
 ggctgtccgc ggggggacgg ctgccttcgg gggggacggg gcagggcggg gttcggcttc 2700
 tggcgtgtga ccggcggtc tagagcctct gtaaacatg ttcatgcctt cttctttttc 2760
 ctacagctcc tgggcaacgt gctggttatt gtgctgtctc atcatttttg caaagaattc 2820
 cgccaccatg tctatggggg ctctcgtc cctgctgctg gacttgccg ccgggctggc 2880
 tgtcgcaaga ccacctaata tcgtcctgat tttgcagac gatctgggat acggcgacct 2940
 gggatgctat ggccacccaa gctccaccac acccaacctg gaccagctgg cagcaggagg 3000
 cctgcggttc accgacttct acgtgccagt gagcctgtgc accccctcca gagccgcct 3060
 gctgacaggc aggctgccag tgcgcatggg catgtatcct ggctgctgg tgccatctag 3120
 caggggcggc ctgccactgg aggaggtgac cgtggcagag gtgctggcag ccagaggcta 3180
 cctgacagga atggccggca agtggcacct gggagtggga ccagaggagg ctttctgcc 3240
 cctcaccag ggcttcacc ggtttctggg catcccttat tctcacgacc agggcccatg 3300
 ccagaacctg acctgttttc caccagcaac accatgcgac ggaggatgtg atcagggcct 3360
 ggtgccaatc cactgctgg caaatctgag cgtggaggca cagcctccat ggctgcctgg 3420
 cctggaggca agatacatgg ctttcgcca cgacctgat gcagatgcac agcggcagga 3480
 tagaccttc tttctgtact atgctccca ccacaccac tatccacagt tcagcgcca 3540
 gtcctttgcc gagaggtccg gaaggggacc attcggcgac tctctgatgg agctggatgc 3600
 cgccgtgggc accctgatga cagcaatcgg cgacctgggc ctgctggagg agacactggt 3660
 catcttcacc gccgataacg gccctgagac aatgcgatg tctagaggcg gatgcagcgg 3720
 cctgctgaga tgtggcaagg gaaccacata cgagggaggc gtgcgcgagc ctgccctggc 3780
 attttgcca ggacacatcg cacctggagt gaccacag ctggcctcct ctctggacct 3840
 gctgccaaca ctggccgcc tggcaggagc acctctgcca aatgtgacct tggacggctt 3900
 cgatctgagc cactgctgc tgggaaccgg caagtccct aggcagtctc tgttctttta 3960

ccccctctat cctgatgagg tgcggggcgt gtttgccgtg agaaccggca agtacaaggc 4020
ccacttcttt acacagggct ctgcccacag cgacaccaca gcagatccag catgccacgc 4080
cagctcctct ctgaccgcac acgagccacc tctgctgtac gacctgtcca aggatcccgg 4140
cgagaactat aatctgctgg gaggagtggc aggagcaacc cctgaggtgc tgcaggccct 4200
gaagcagctg cagctgctga aggcacagct ggacgcagca gtgacattcg gccaagcca 4260
ggtggccaga ggcgaggatc ccgccctgca gatctgctgc caccaggct gcacaccag 4320
acctgcctgc tgtcactgcc ccgaccaca cgccggcagc ggagctacta acttcagcct 4380
gctgaagcag gctggagacg tggaggagaa ccctggacct ggggacgttt gccaggactg 4440
cattcagatg gtgactgaca tccagactgc tgtacggacc aactccacct ttgtccaggc 4500
cttggtggaa catgtcaagg aggagtgtga ccgcctgggc cctggcatgg ccgacatatg 4560
caagaactat atcagccagt attctgaaat tgctatccag atgatgatgc acatgcaacc 4620
caaggagatc tgtgcgctgg ttgggttctg tgatgagtga actagtaact tgttttattgc 4680
agcttataat ggttacaaat aaagcaatag catcacaat ttcacaaata aagcattttt 4740
ttcactgcat tctagtgtg gtttgtccaa actcatcaat gtatcttagg tctagatacg 4800
tagataagta gcatggcggg ttaatcatta actacaagga acccctagtg atggagttgg 4860
ccactccctc tctgcgcgct cgctcgtca ctgaggccgg gcgaccaaag gtcgcccgc 4920
gcccgggctt tgcccgggcg gcctcagtga gcgagcgagc gcgcagagag ggagtggcca 4980
aagatccccg ggtaccgagg acgaattctc tagatatcgc tcaatactga ccattttaat 5040
catacctgac ctccatagca gaaagtcaaa agcctccgac cggaggcttt tgacttgatc 5100
ggcacgtaag aggttccaac tttcaccata atgaaataag atcactaccg ggcgtatttt 5160
ttgagttatc gagattttca ggagctaagg aagctaaaat gagccatatt caacgggaaa 5220
cgtcttgctc gaggccgcga ttaaattcca acatggatgc tgatttatat gggataaat 5280
gggctcgcga taatgtcggg caatcaggtg cgacaatcta tcgattgtat gggaagcccc 5340
atgcgccaga gttgtttctg aaacatggca aaggtagcgt tgccaatgat gttacagatg 5400
agatggtcag gctaaactgg ctgacggaat ttatgcctct tccgaccatc aagcatttta 5460
tccgtactcc tgatgatgca tggttactca cactgcgat ccagggaaa acagcattcc 5520
aggattaga agaatatcct gattcaggtg aaaatattgt tgatgcgctg gcagtgttcc 5580
tgcccggtt gcattcgatt cctgtttgta attgtccttt taacggcgat cgcgtatttc 5640
gtctcgtca ggcgcaatca cgaatgaata acggtttggt tgggtgcgagt gattttgatg 5700
acgagcgtaa tggctggcct gttgaacaag tctggaaaga aatgcataag cttttgccat 5760
tctcaccgga ttcagtcgtc actcatggtg atttctact tgataacctt atttttgacg 5820
aggggaaatt aataggttgt attgatgtg gacgagtcgg aatcgcagac cgataccagg 5880
atcttgccat cctatggaac tgctcgggtg agttttctcc ttcattacag aaacggcttt 5940
ttcaaaaata tggattgat aatcctgata tgaataaatt gcagtttcac ttgatgctcg 6000
atgagttttt ctgagggcc aaatgtaatc acctggctca cttcgggtg ggcctttctg 6060
cgttgctggc gtt 6073

<210> 78

<211> 42

<212> DNA

<213> 人工序列

<220>

<223> 合成核酸序列

<400> 78

ggaaaaccaa taccaaaccc tctattagga ttggactcaa ca 42

<210> 79

<211> 3458

<212> DNA

<213> 人工序列

<220>

<223> 合成核酸序列

<400> 79

```

ggcattgatt attgactagt tattaatagt aatcaattac ggggtcatta gttcatagcc 60
catatatgga gttccgcgtt acataactta cggtaaatgg cccgcctggc tgaccgcca 120
acgacccccg cccattgacg tcaataatga cgtatgttcc catagtaacg ccaataggga 180
ctttccattg acgtcaatgg gtggagtatt tacggtaaac tgcccacttg gcagtacatc 240
aagtgtatca tatgccaaagt ccgcccccta ttgacgtcaa tgacggtaaa tggcccgctt 300
ggcattatgc ccagtacatg accttacggg actttcctac ttggcagtac atctacgtat 360
tagtcatcgc tattaccatg gtcgaggtga gcccacgtt ctgcttcaact ctccccatct 420
cccccccctc cccaccccca attttgtatt tatttatttt ttaattattt tgtgcagcga 480
tgggggcggg gggggggggg gggcgcgcg caggcggggc ggggcggggc gaggggcggg 540
gcggggcgag gcggagaggt gcggcggcag ccaatcagag cggcgcgctc cgaaagtttc 600
cttttatggc gaggcggcgg cgggcgcggc cctataaaaa gcgaagcgcg cggcgggcgg 660
gagtcgctgc gcgctgcctt cgccccgtgc cccgctccgc cgccgcctcg cgccgcccgc 720
cccggtcttg actgaccgcg ttactccac aggtgagcgg gcgggacggc ctttctcctc 780
cgggctgtaa ttagcgcttg gtttaatgac ggcttgtttc ttttctgtgg ctgcgtgaaa 840
gccttgaggg gctccgggag ggccctttgt gcggggggag cggctcgggg ggtgcgtgcg 900
tgtgtgtgtg cgtggggagc gccgcgtgcg gctccgcgct gcccggcggc tgtgagcgct 960
gcgggcgcgg cgcggggctt tgtgcgctcc gcagtgtgcg cgaggggagc gcggccgggg 1020
gcggtgcccc gcggtgcggg gggggctgcg aggggaacaa aggctgcgtg cggggtgtgt 1080
gcgtgggggg gtgagcaggg ggtgtgggcg cgtcggtcgg gctgcaacct cccctgcacc 1140
cccctccccg agttgctgag cacggcccgg cttcgggtgc ggggctccgt acggggcggtg 1200
gcgcggggct cgccgtgccg ggccgggggt ggccggcaggt gggggtgccg ggccgggcgg 1260
ggccgcctcg ggccggggag ggctcggggg aggggcgcgg cgcccccg agcgccggcg 1320
gctgtcgagg cgcggcgagc cgcagccatt gccttttatg gtaatcgtgc gagaggcgcg 1380
agggaacttc tttgtcccaa atctgtgcgg agccgaaatc tgggaggcgc cgccgcaccc 1440
cctctagcgg gcgcggggcg aagcggtgcg gcgccggcag gaaggaaatg ggccgggagg 1500
gccttcgtgc gtcgccgcgc cgccgtcccc ttctccctct ccagcctcgg ggctgtccgc 1560
ggggggacgg ctgccttcgg gggggacggg gcagggcggg gttcggcttc tggcgtgtga 1620

```

ccggcggctc tagagcctct gctaaccatg ttcatgcctt cttctttttc ctacagctcc 1680
tgggcaacgt gctgggttatt gtgctgtctc atcatttttg caaagaattc cgccaccatg 1740
tctatggggg ctcctcgctc cctgctgctg gcactggccg ccgggctggc tgtcgcaaga 1800
ccaccttaata tcgtcctgat ttttgacagc gatctgggat acggcgacct gggatgctat 1860
ggccacccaa gctccaccac acccaacctg gaccagctgg cagcaggagg cctgcggttc 1920
accgacttct acgtgccagt gagcctgtgc accccctcca gagccgccct gctgacaggc 1980
aggctgccag tgcgcatggg catgtatcct ggctgtctgg tgccatctag caggggaggc 2040
ctgccactgg aggaggtgac cgtggcagag gtgctggcag ccagaggcta cctgacagga 2100
atggccggca agtggcacct gggagtggga ccagaggagg ccttcctgcc ccctcaccag 2160
ggcttccacc ggtttctggg catcccttat tctcacgac agggcccatg ccagaacctg 2220
acctgttttc caccagcaac accatgcgac ggaggatgtg atcagggcct ggtgccaatc 2280
ccactgctgg caaatctgag cgtggaggca cagcctccat ggctgcctgg cctggaggca 2340
agatacatgg ccttcgcccc cgacctgatg gcagatgcac agcggcagga tagacctttc 2400
tttctgtact atgcctcccc ccacaccac tatccacagt tcagcggcca gtcccttgcc 2460
gagaggtccg gaaggggacc attcggcgac tctctgatgg agctggatgc cgccgtgggc 2520
accctgatga cagcaatcgg cgacctgggc ctgctggagg agacactggt catcttcacc 2580
gccgataacg gccctgagac aatgcggatg tctagaggcg gatgcagcgg cctgctgaga 2640
tgtggcaagg gaaccacata cgaggaggc gtgcgcgagc ctgccctggc attttgcca 2700
ggacacatcg cacctggagt gaccacgag ctggcctcct ctctggacct gctgccaaca 2760
ctggccgccc tggcaggagc acctctgcca aatgtgacct tggacggctt cgatctgagc 2820
ccactgctgc tgggaaccgg caagtcccct aggcagtctc tgtttcttta cccctcctat 2880
cctgatgagg tgcggggcgt gtttgccgtg agaaccggca agtacaaggc ccacttcttt 2940
acacagggct ctgcccacag cgacaccaca gcagatccag catgccacgc cagctcctct 3000
ctgaccgcac acgagccacc tctgctgtac gacctgtcca aggatcccgg cgagaactat 3060
aatctgctgg gaggagtggc aggagcaacc cctgagggtg tgcaggccct gaagcagctg 3120
cagctgctga aggcacagct ggacgcagca gtgacattcg gccaagcca ggtggccaga 3180
ggcgaggatc ccgcccgtga gatctgttgc caccgccgct gcacccaag acctgcctgt 3240
tgccattgcc ccgaccaca cgccggaaaa ccaataccaa accctctatt aggattggac 3300
tcaacataag attctagagt cgagccggcg actagtaact tgtttattgc agcttataat 3360
ggttacaaat aaagcaatag catcacaat ttcacaaata aagcattttt ttactgcat 3420
tctagtgtg gtttgtccaa actcatcaat gtatctta 3458

<210> 80

<211> 3991

<212> DNA

<213> 人工序列

<220>

<223> 合成核酸序列

<400> 80

ttggccactc cctctctgcg cgctcgctcg ctactgagg ccgggcgacc aaaggtcgcc 60

cgacgcccgg gctttgcccg ggcggcctca gtgagcgagc gagcgcgag agagggagtg 120
gccaactcca tcactagggg ttcctggagg ggtggagtcg tgacgtgaat tacgtcatag 180
ggttagggag gtcctgcaga tcttcaatat tggccattag ccatattatt cattggttat 240
atagcataaa tcaatattgg ctattggcca ttgcatacgt tgtatctata tcataaatatg 300
tacatttata ttggctcatg tccaatatga ccgccatgtt ggcatlgatt attgactagt 360
tattaatagt aatcaattac ggggtcatta gttcatagcc catatatgga gttccgcgtt 420
acataactta cggtaaatgg cccgcctggc tgaccgccc aacgacccccg cccattgacg 480
tcaataatga cgtatgttcc catagtaacg ccaataggga ctttccattg acgtcaatgg 540
gtggagtgatt tacggtaaac tgcccacttg gcagtacatc aagtgtatca tatgccaagt 600
ccgcccccta ttgacgtcaa tgacggtaaa tggcccgctt ggcatatgc ccagtacatg 660
accttacggg actttcctac ttggcagtac atctacgtat tagtcatgcg tattaccatg 720
gtcgagtgta gccccacgtt ctgttcaact ctccccatct cccccccctc cccaccccc 780
attttgtatt tatttatatt ttaattattt tgtgcagcga tgggggaggg gggggggggg 840
gggcgcgcgc caggcggggc ggggcggggc gaggggaggg gcggggcgag gcggagaggt 900
gcggcggcag ccaatcagag cggcgcgctc cgaaagtctt cttttatggc gaggcggcgg 960
cggcggcggc cctataaaaa gcgaagcgcg cggcggcgag gagtcgctgc gcgctgcctt 1020
cgccccgtgc cccgctccgc cgccgctcgc cgccgcccgc cccggtctgc actgaccgcg 1080
ttactccac aggtgagcgg gcgggacggc cttctctc cgggctgtaa ttagcgcttg 1140
gtttaatgac ggcttgctt ttttctgttg ctgcgtgaaa gccttgaggg gctccgggag 1200
ggccctttgt gcggggggag cggctcgggg ggtgcgtgcg tgtgtgtgtg cgtggggagc 1260
gccgcgtgcg gctccgcgct gccggcggcg tgtgagcgtc gcgggcgcg gcgggggctt 1320
tgtgcgtcc gcagtgtgcg cgaggggagc gcggccgggg gcggtgcccc gcggtgcggg 1380
gggggctgcg aggggaacaa aggctgcgtg cggggtgtgt gcgtgggggg gtgagcaggg 1440
ggtgtgggag cgtcggctcg gctgcaacc cccctgcacc cccctcccc agttgctgag 1500
cacggccccg cttcgggtgc ggggctccgt acggggcgtg gcgcggggct cgccgtgccg 1560
ggcggggggg ggccgcaggt gggggtgccg ggccggggcg ggccgcctcg ggccggggag 1620
ggctcggggg aggggcgcgg cgcccccg agcgccggcg gctgtcagg cgcggcgagc 1680
cgagccatt gccttttatg gtaatcgtgc gagaggcgc agggacttcc tttgtccaa 1740
atctgtgcgg agccgaaatc tgggaggcgc cgccgcacc cctctagcgg gcgcggggcg 1800
aagcgggtgc gcgccggc gaaggaaatg ggccggggag gccttcgtgc gtcgccgcg 1860
cgccgtcccc ttctccctct ccagcctcgg ggctgtccgc ggggggacgg ctgccttcgg 1920
gggggacggg gcagggcggg gttcggttc tggcgtgtga ccggcggtc tagagcctct 1980
gctaaccatg ttcatgcctt cttctttttc ctacagctc tgggcaacgt gctggttatt 2040
gtgctgtctc atcattttgg caaagaattc cgccaccatg tctatggggg ctctctgctc 2100
cctgctgctg gcaactggccg ccgggctggc tgtcgcaaga ccacctaata tcgtctgat 2160
ttttgcagac gatctgggat acggcgacct gggatgctat ggccaccaa gctccaccac 2220
acccaacctg gaccagctgg cagcaggagg cctgcggttc accgacttct acgtgccagt 2280
gagcctgtgc accccctcca gagccgcctt gctgacaggc aggctgccag tgcgcatggg 2340
catgtatcct ggctgtctgg tgccatctag caggggcggc ctgccactgg aggaggtgac 2400

cgtggcagag gtgctggcag ccagaggcta cctgacagga atggccggca agtggcacct 2460
 gggagtggga ccagaggag ccttcctgcc ccctcaccag ggcttccacc ggtttctggg 2520
 catcccttat tctcacgacc agggcccatg ccagaacctg acctgttttc caccagcaac 2580
 accatgcgac ggaggatgtg atcagggcct ggtgccaatc cactgtctgg caaatctgag 2640
 cgtggaggca cagcctccat ggctgcctgg cctggaggca agatacatgg ccttcgcccc 2700
 cgacctgatg gcagatgcac agcggcagga tagacctttc tttctgtact atgcctcccc 2760
 ccacaccac tatccacagt tcagcggcca gtcttttgcc gagaggtccg gaaggggacc 2820
 attcggcgac tctctgatgg agctggatgc cgccgtgggc accctgatga cagcaatcgg 2880
 cgacctgggc ctgctggagg agacactggt catcttcacc gccgataacg gccctgagac 2940
 aatgcggatg tctagaggcg gatgcagcgg cctgctgaga tgtggcaagg gaaccacata 3000
 cgagggaggc gtgcgcgagc ctgccctggc attttggcca ggacacatcg cacctggagt 3060
 gaccacagag ctggcctcct ctctggacct gctgccaaca ctggccgccc tggcaggagc 3120
 acctctgcca aatgtgacct tggacggctt cgatctgagc cactgtctgc tgggaaccgg 3180
 caagtcccct aggcagtctc tgttctttta cccctcctat cctgatgagg tgcggggcgt 3240
 gtttgccgtg agaaccggca agtacaaggc ccacttcttt acacagggtc ctgcccacag 3300
 cgacaccaca gcagatccag catgccacgc cagctcctct ctgaccgcac acgagccacc 3360
 tctgtgttac gacctgtcca aggatcccg cgagaactat aatctgctgg gaggagtggc 3420
 aggagcaacc cctgagggtc tgcaggccct gaagcagctg cagctgctga aggcacagct 3480
 ggacgcagca gtgacattcg gcccagcca ggtggccaga ggcgaggatc ccgccctgca 3540
 gatctgttgc caccgccgct gcacccaag acctgcctgt tgccattgcc ccgaccaca 3600
 cgccggaaaa ccaataccaa accctctatt aggattggac tcaacataag attctagagt 3660
 cgagccgcgg actagtaact tgtttattgc agcttataat ggttacaaat aaagcaatag 3720
 catcacaat ttcacaaata aagcattttt ttcactgcat tctagttgtg gtttgtccaa 3780
 actcatcaat gtatcttagg tctagatacg tagataagta gcatggcggg ttaatcatta 3840
 actacaagga acccctagtg atggagtgg ccactccctc tctgcgcgct cgctcgctca 3900
 ctgaggccgg gcgaccaaag gtcgcccgc gcccgggctt tgcccgggcg gcctcagtga 3960
 gcgagcgagc gcgcagagag ggagtggcca a 3991

<210> 81

<211> 6654

<212> DNA

<213> 人工序列

<220>

<223> 合成核酸序列

<400> 81

cgccagggtt ttcccagtca cgacgttgta aaacgacggc cagtgccaaag cttgcatgcc 60
 tgcatttggc cactccctct ctgcgcgtc gctcgtcac tgaggccggg cgaccaaagg 120
 tcgcccagc cccgggcttt gcccgggcgg cctcagttag cgagcgagcg gcgagagagg 180
 gagtggccaa ctccatcact aggggttcct ggagggtgg agtcgtgacg tgaattacgt 240
 catagggtta gggaggctct gcagatcttc aatattggcc attagccata ttattcattg 300

```

gttatatagc ataaatcaat attggctatt ggccattgca tacgttgtat ctatatcata 360
atatgtacat ttatatggc tcatgtccaa tatgaccgcc atgttggcat tgattattga 420
ctagttatta atagtaatca attacggggc cattagtcca tagcccatat atggagttcc 480
gcgttacata acttacggta aatggccgc ctggctgacc gccaacgac cccgcccatt 540
tgacgtcaat aatgacgtat gttcccatag taacgccaat agggactttc cattgacgtc 600
aatgggtgga gtattttacgg taaactgccc acttggcagt acatcaagtg tatcatatgc 660
caagtccgcc ccctattgac gtcaatgacg gtaaatggcc cgcctggcat tatgcccagt 720
acatgacctt acgggacttt cctacttggc agtacatcta cgtattagtc atcgctatta 780
ccatggctga ggtgagcccc acgttctgct tcaactctcc catctcccc cctccccac 840
ccccaatttt gtattttattt attttttaat tattttgtgc agcgatgggg gcgggggggg 900
ggggggggcg cgcgccaggc ggggcggggc ggggcgaggg gcggggcggg gcgaggcgga 960
gaggtgcggc ggcagccaat cagagcggcg cgctccgaaa gtttctttt atggcgaggc 1020
ggcggcgggc gcggccctat aaaaagcgaa gcgcgcggcg ggcgggagtc gctgcgcgt 1080
gccttcgccc cgtgccccgc tccgccgcgc cctcgcgcgc cccgcccccg ctctgactga 1140
ccgcgttact cccacaggtg agcgggcggg acggcccttc tcctccgggc tgtaattagc 1200
gcttggttta atgacggctt gtttctttt tgtggctgcg tgaaagcctt gaggggctcc 1260
gggagggccc tttgtgcggg gggagcggct cggggggtgc gtgctgtgt gtgtgcgtgg 1320
ggagcgccgc gtgcggctcc gcgctgccc gcgctgtga gcgctgcgg cgcggcgcgg 1380
ggctttgtgc gctccgcagt gtgcgcgagg ggagcgcggc cgggggcggg gccccgcgg 1440
gcgggggggg ctgcgagggg aacaaaggct gcgtgcgggg tgtgtgcgtg ggggggtgag 1500
cagggggtgt gggcgcgctg gtcgggctgc aacccccct gcacccccct ccccgagttg 1560
ctgagcacgg cccggcttcg ggtgcggggc tccgtacggg gcgtggcgcg gggctcgccg 1620
tgccgggcgg ggggtggcgg caggtggggg tgccgggcgg ggcggggccg cctcgggcgg 1680
gggagggctc gggggagggg cgcggcgggc cccggagcgc cggcggtgt cgaggcgcg 1740
cgagccgcag ccattgcctt ttatggtaat cgtgcgagag ggcgcaggga cttcctttgt 1800
cccaaacttg tgcggagccg aaatctggga ggcgccgcc caccctctt agcgggcgcg 1860
ggcggaagcg gtgcggcgcc ggcaggaagg aaatgggcgg ggagggcctt cgtgcgtcgc 1920
cgcgcgcgcg tccccttct cctctccagc ctcggggctg tccgcggggg gacggctgcc 1980
ttcggggggg acggggcagg gcggggttcg gcttctggcg tgtgaccggc ggctctagag 2040
cctctgctaa ccatgttcat gccttcttct ttttctaca gctcctgggc aacgtgctgg 2100
ttattgtgct gtctcatcat tttggcaaag aattccgcca ccatgtctat gggggctcct 2160
cgctccctgc tgctggcact ggccgccggg ctggctgtcg caagaccacc taatatcgtc 2220
ctgatttttg cagacgatct gggatacggc gacctgggat gctatggcca cccaagctcc 2280
accacacca acctggacca gctggcagca ggaggcctgc ggttcaccga cttctacgtg 2340
ccagtgagcc tgtgcacccc ctccagagcc gccctgctga caggcaggct gccagtgcgc 2400
atgggcatgt atcctggcgt gctgggtcca tctagcagg gcggcctgcc actggaggag 2460
gtgaccgtgg cagaggtgct ggcagccaga ggctacctga caggaatggc cggcaagtgg 2520
cacctgggag tgggaccaga gggagccttc ctgccccctc accagggtt ccaccggttt 2580
ctgggcatcc cttattctca cgaccagggc ccatgccaga acctgacctg ttttcacca 2640

```

```

gcaacacccat gcgacggagg atgtgatcag ggcctggtgc caatcccact gctggcaaatt 2700
ctgagcgtgg aggcacagcc tccatggctg cctggcctgg aggcaagata catggccttc 2760
gcccacgacc tgatggcaga tgcacagcgg caggatagac ctttctttct gtactatgcc 2820
tcccaccaca cccactatcc acagttcagc ggccagtcct ttgccgagag gtccggaagg 2880
ggaccattcg gcgactctct gatggagctg gatgccgccg tgggcaccct gatgacagca 2940
atcggcgacc tgggcctgct ggaggagaca ctggtcattc tcaccgccga taacggccct 3000
gagacaatgc ggatgtctag aggcggatgc agcggcctgc tgagatgtgg caagggaacc 3060
acatacgagg gaggcgtgcg cgagcctgcc ctggcatttt ggccaggaca catcgcacct 3120
ggagtgaccc acgagctggc ctctctctg gacctgctgc caacactggc cgccctggca 3180
ggagcacctc tgccaaatgt gacctggac ggcttcgac tgagcccact gctgctggga 3240
accggcaagt cccctaggca gtctctgttc ttttaccct cctatcctga tgaggtgcgg 3300
ggcgtgtttg ccgtgagaac cggcaagtac aaggcccact tctttacaca gggctctgcc 3360
cacagcgaca ccacagcaga tccagcatgc cagccagct cctctctgac cgcacacgag 3420
ccacctctgc tgtacgacct gtccaaggat cccggcgaga actataatct gctgggagga 3480
gtggcaggag caaccctga ggtgctgcag gccctgaagc agctgcagct gctgaaggca 3540
cagctggacg cagcagtac attcggcca agccaggtgg ccagaggcga ggatcccgcc 3600
ctgcagatct gttgccaccc cggtgcacc ccaagacct cctgttgcca ttgccccgac 3660
ccacacgccg gaaaaccaat accaaaccct ctattaggat tggactcaac ataagattct 3720
agagtcgagc cgcggactag taacttgttt attgcagctt ataatggtta caaataaagc 3780
aatagcatca caaatttcac aaataaagca tttttttcac tgcattctag ttgtggtttg 3840
tccaaactca tcaatgtatc ttaggtctag atacgtagat aagtagcatg gcgggttaat 3900
cattaactac aaggaacccc tagtgatgga gttggccact ccctctctgc gcgctcgctc 3960
gctcactgag gccggggcgac caaaggtcgc ccgacgccc ggctttgccc gggcggcctc 4020
agtgagcgag cgagcgcgca gagaggaggt ggccaaagat ccccggttac cgagctcgaa 4080
ttcgtaatca tgtcatagct gtttcctgtg tgaaattgtt atccgctcac aattccacac 4140
aacatacgag ccggaagcat aaagtgtaaa gcctgggggtg cctaattgagt gagctaactc 4200
acattaattg cgttgcgctc actgcccgtt ttccagtcgg gaaacctgtc gtgccagctg 4260
cattaatgaa tcggccaacg cgcggggaga ggcggtttgc gtattggcga acttttgctg 4320
agttgaagga tcagatcacg catcttccc acaacgcaga ccgttccgtg gcaaagcaaa 4380
agttcaaaat cagtaaccgt cagtgccgat aagttcaag ttaaacctgg tgttgatacc 4440
aacattgaaa cgctgatcga aaacgcgctg aaaaacgctg ctgaatgtgc gagcttcttc 4500
cgcttcctcg ctactgact cgctgcgctc ggtcgttcgg ctgcggcgag cggtatcagc 4560
tactcaaaag gcggtataac ggttatccac agaatcaggg gataacgcag gaaagaacat 4620
gtgagcaaaa ggccagcaaa aggccaggaa ccgtaaaaag gccgcgttgc tggcgttttt 4680
ccataggctc cgccccctg acgagcatca caaaaatcga cgctcaagtc agaggtggcg 4740
aaacccgaca ggactataaa gataccaggc gtttccccct ggaagctccc tcgtgcgctc 4800
tcctgttccg accctgccgc ttaccggata cctgtccgcc tttctccctt cggaagcgt 4860
ggcgctttct caatgctcac gctgtaggta tctcagttcg gtgtaggtcg ttcgctccaa 4920
gctgggctgt gtgcacgaac cccccgttca gcccgaccgc tgcgccttat ccggtacta 4980

```

tcgtcttgag tccaacccgg taagacacga cttatcgcca ctggcagcag ccaactggtaa 5040
caggattagc agagcgaggt atgtaggcgg tgctacagag ttcttgaagt ggtggcctaa 5100
ctacggctac actagaagga cagtatttgg tatctgcgct ctgctgaagc cagttacctt 5160
cggaaaaaga gttggtagct cttgatccgg caaacaacc accgctggta gcggtggttt 5220
ttttgtttgc aagcagcaga ttacgcgcag aaaaaagga tctcaagaag atcctttgat 5280
cttttctacg gggctctgacg ctcagtggaa cgatccgtcg agaggtctgc ctcgtgaaga 5340
agggtgttgct gactcatacc aggcctgaat cgccccatca tccagccaga aagtgaggga 5400
gccacggttg atgagagctt tgtttaggtt ggaccagttg gtgattttga acttttgctt 5460
tgccacggaa cggctctgcgt tgtcgggaag atgcgtgac tgatccttca actcagcaaa 5520
agttcgattt attcaacaaa gccacgttgt gtctcaaaat ctctgatgtt acattgcaca 5580
agataaaaaat atatcatcat gaacaataaa actgtctgct tacataaaca gtaatacaag 5640
gggtgttatg agccatattc aacgggaaac gtcttgctcg aagccgcgat taaattccaa 5700
catggatgct gatttatatg ggtataaatg ggctcgcgat aatgtcgggc aatcaggtgc 5760
gacaatctat cgattgtatg ggaagcccga tgcgccagag ttgtttctga aacatggcaa 5820
aggtagcgtt gccaatgatg ttacagatga gatggtcaga ctaaactggc tgacggaatt 5880
tatgcctctt ccgaccatca agcattttat ccgtactcct gatgatgcat ggttactcac 5940
cactgcgac cccgggaaaa cagcattcca ggtattagaa gaatatcctg attcaggtga 6000
aaatatgtt gatgcgctgg cagtgttcct gcgccggtt cattcgattc ctgtttgtaa 6060
ttgtcctttt aacagcgatc gcgtatttcg tctcgctcag gcgcaatcac gaatgaataa 6120
cggtttggtt gatgcgagtg attttgatga cgagcgtaat ggctggcctg ttgaacaagt 6180
ctggaaaaga atgcataagc ttttgccatt ctcaccgat tcagtcgtca ctcaggtga 6240
tttctcactt gataacctta tttttgacga ggggaaatta ataggttgta ttgatgttg 6300
acgagtcgga atcgagacc gataccagga tcttgccatc ctatggaact gcctcggtga 6360
gttttctcct tcattacaga aacggctttt tcaaaaatat ggtattgata atcctgatat 6420
gaataaattg cagtttcatt tgatgctcga tgagttttt taatcagaat tggtaattg 6480
gttgtaacac tggcagagca ttacgctgac ttgacgggac ggcggctttg ttgaataaat 6540
cgcatcgcg attcaggctg cgcaactgtt gggaaggcg atcgggtcgg gcctcttcgc 6600
tattacgcca gctggcgaaa gggggatgtg ctgcaaggcg attaagttgg gtaa 6654

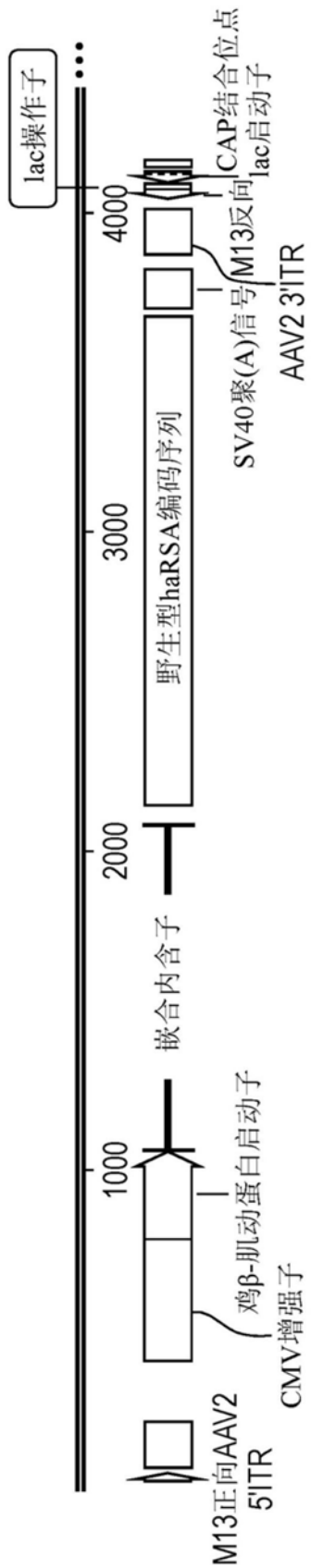


图1A

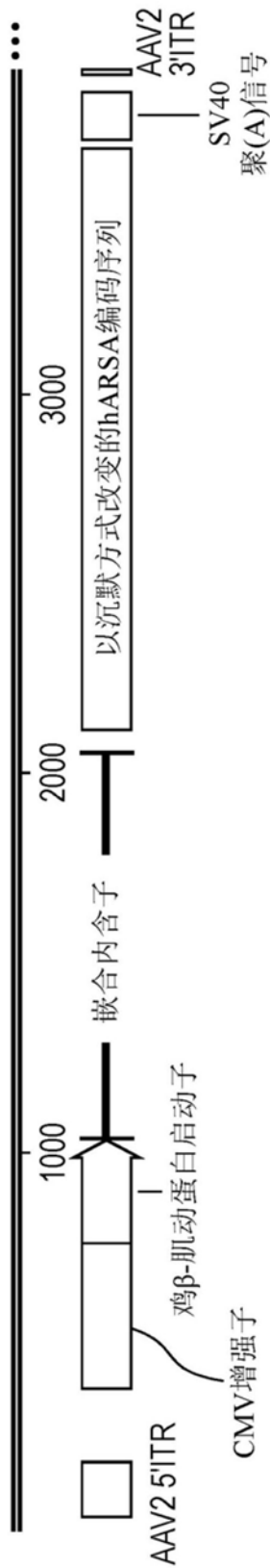


图1B

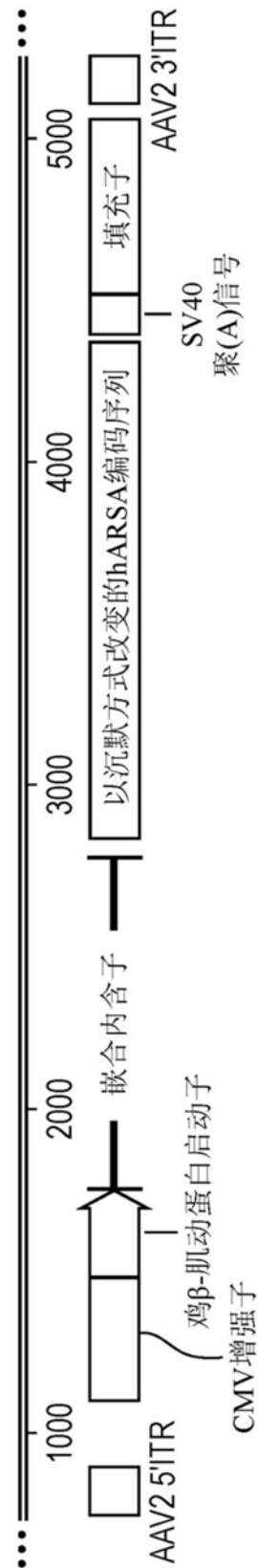


图1C

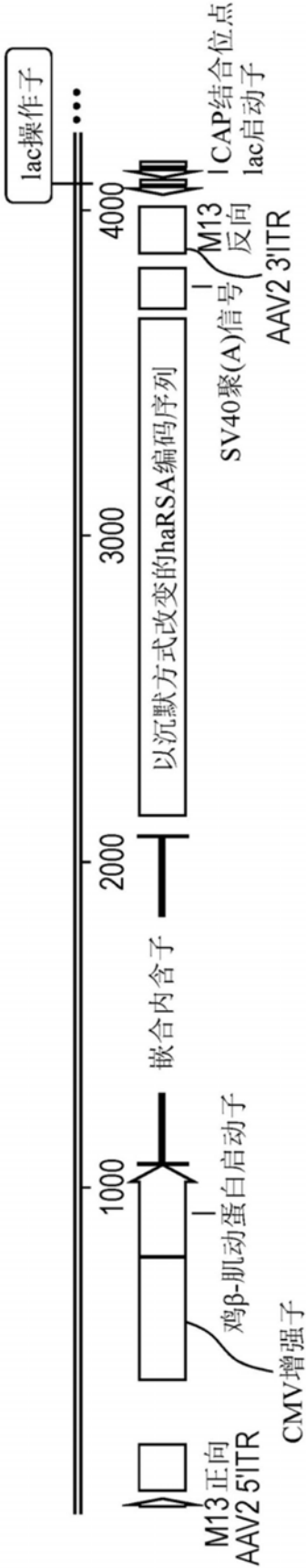


图1D

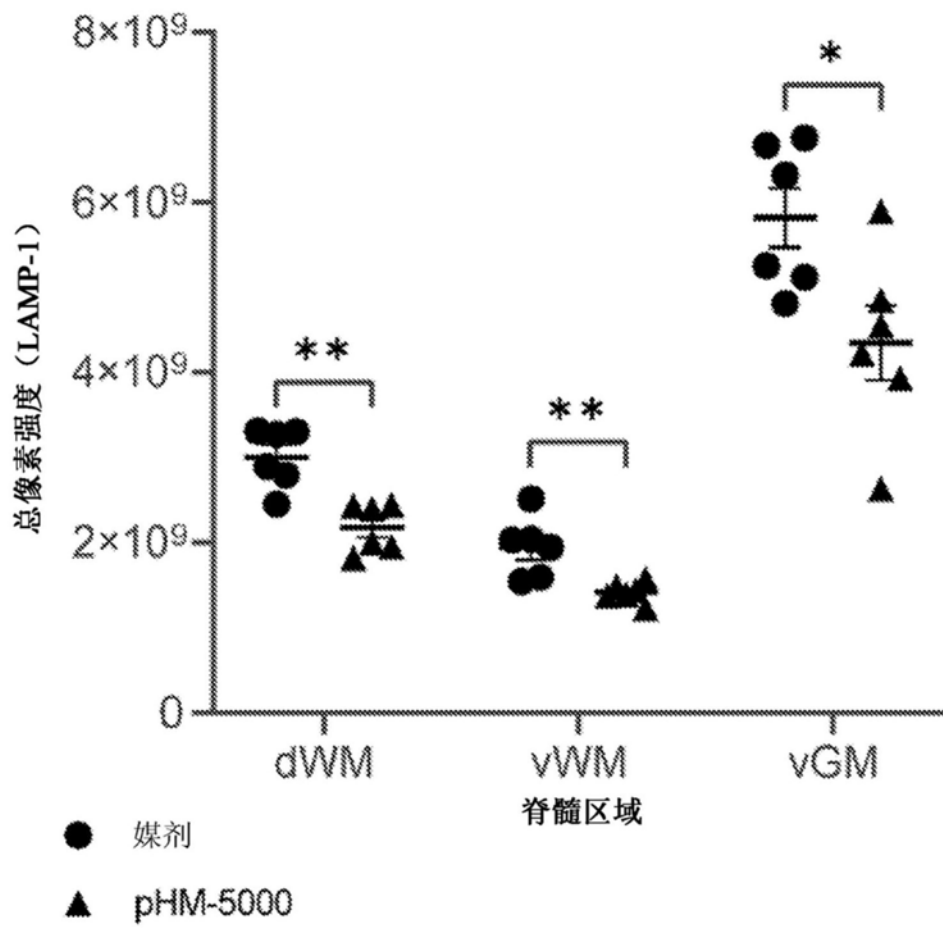


图2A

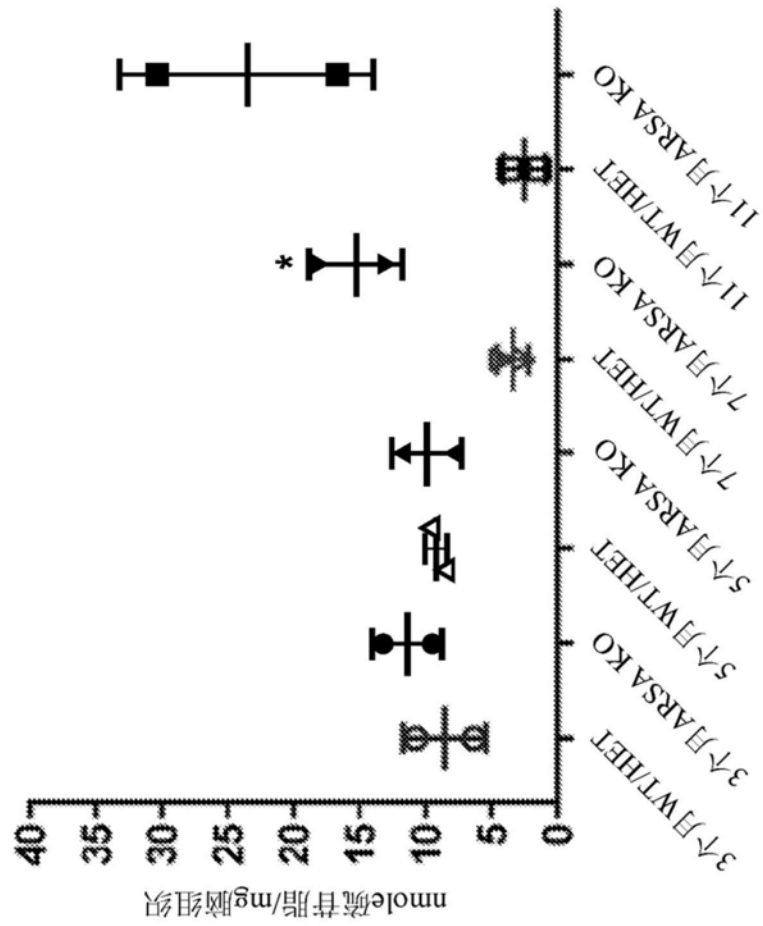


图2B

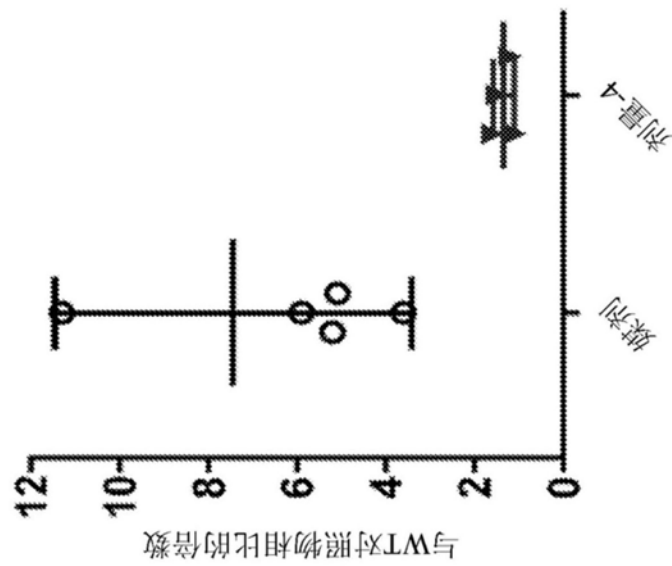


图2C

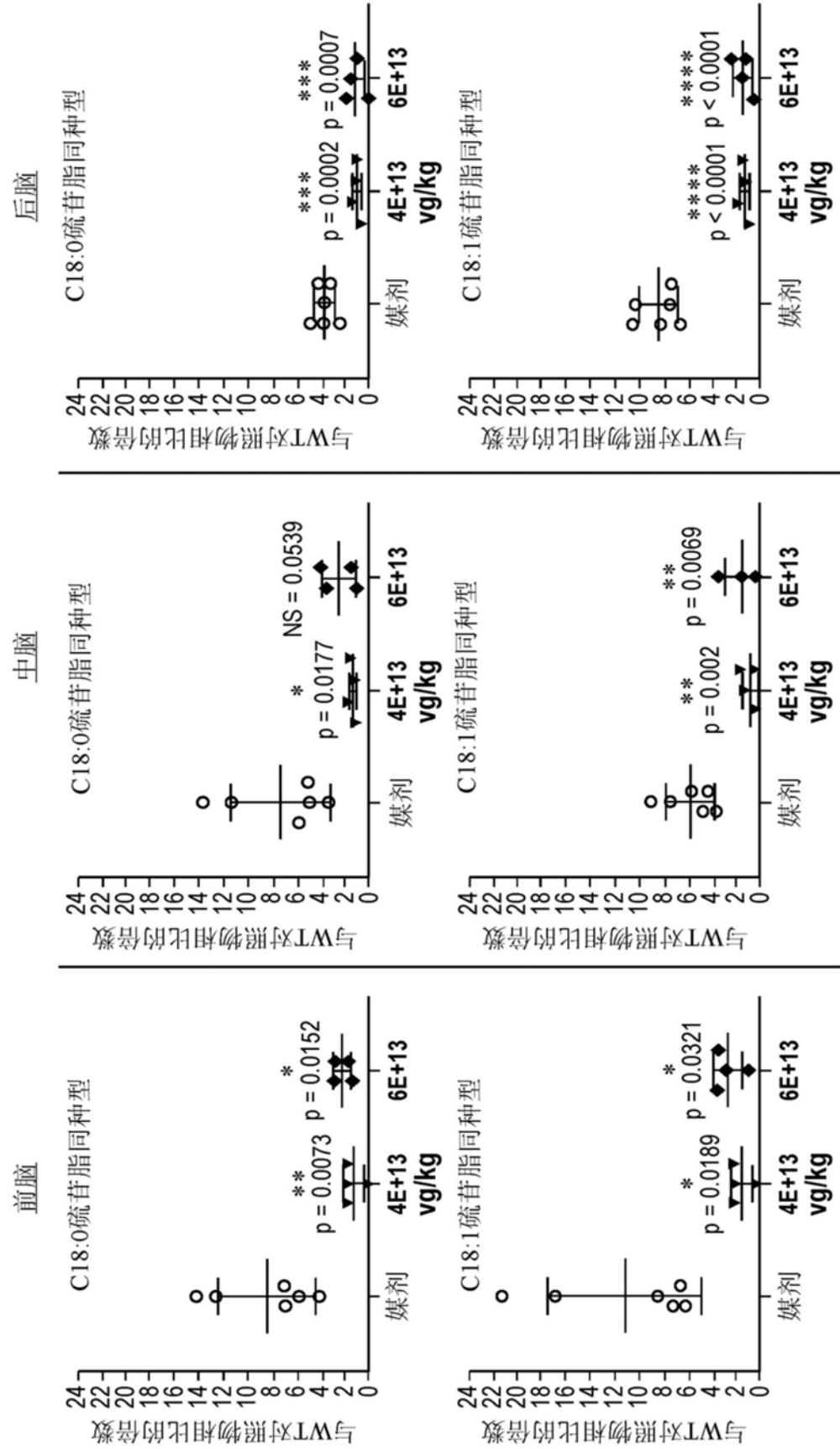


图2D

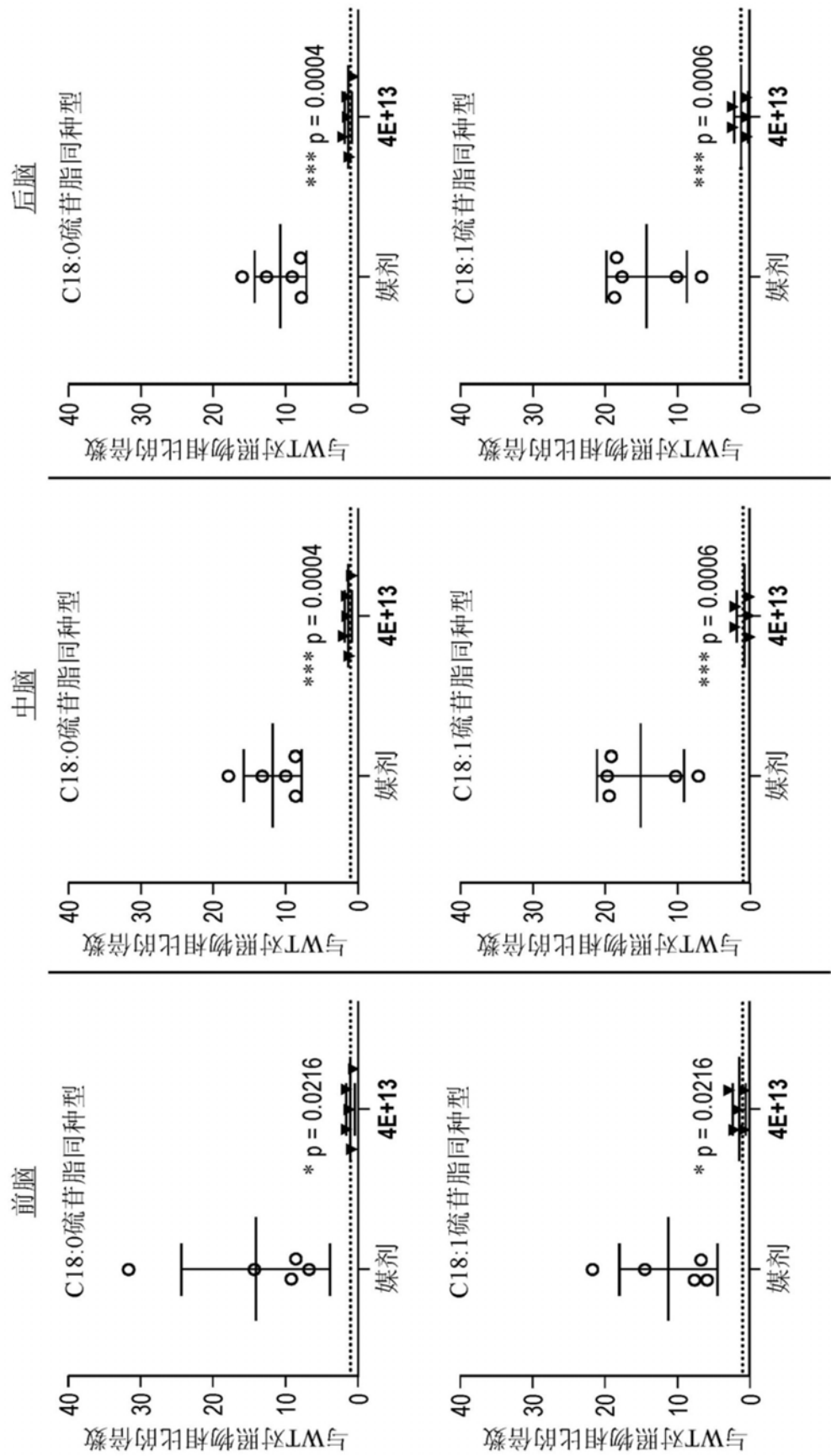


图2E

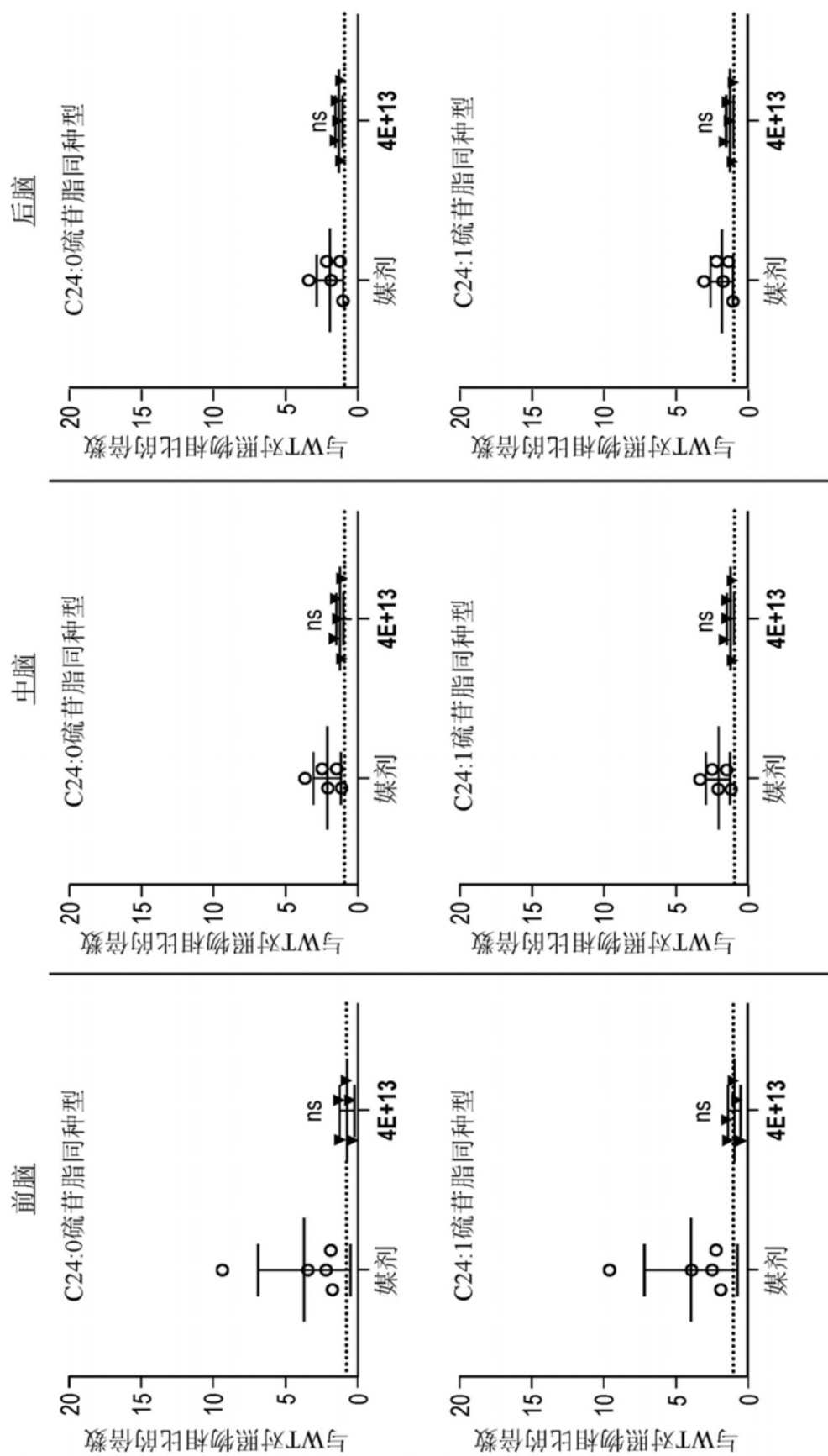


图2F

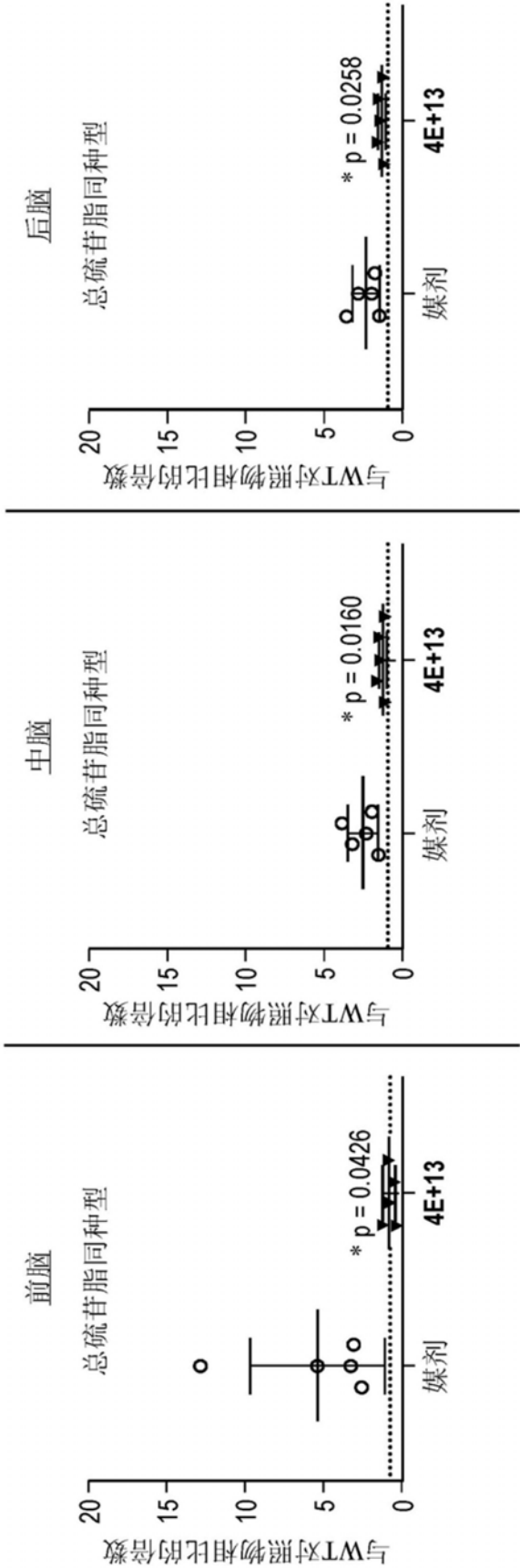


图2G

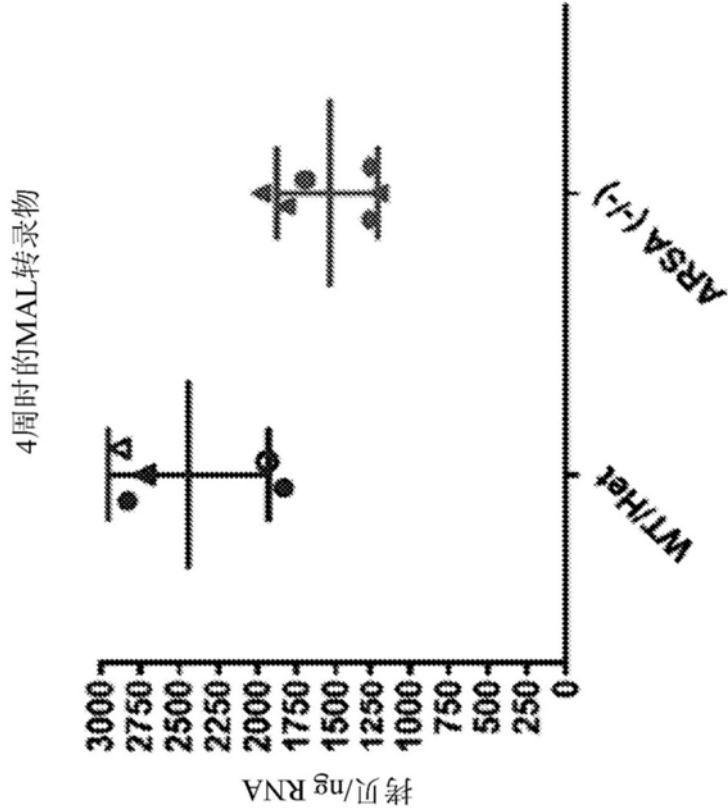


图3A

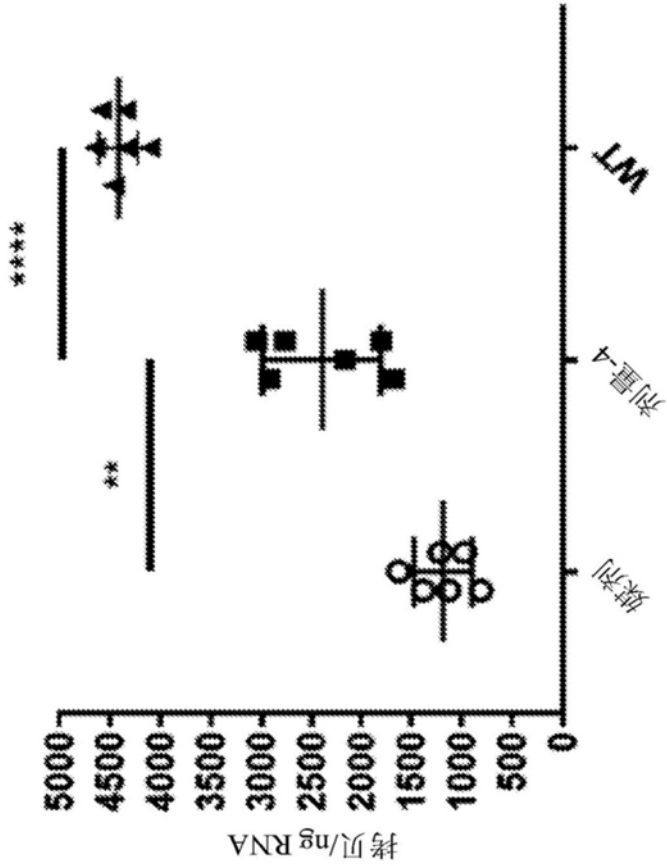


图3B

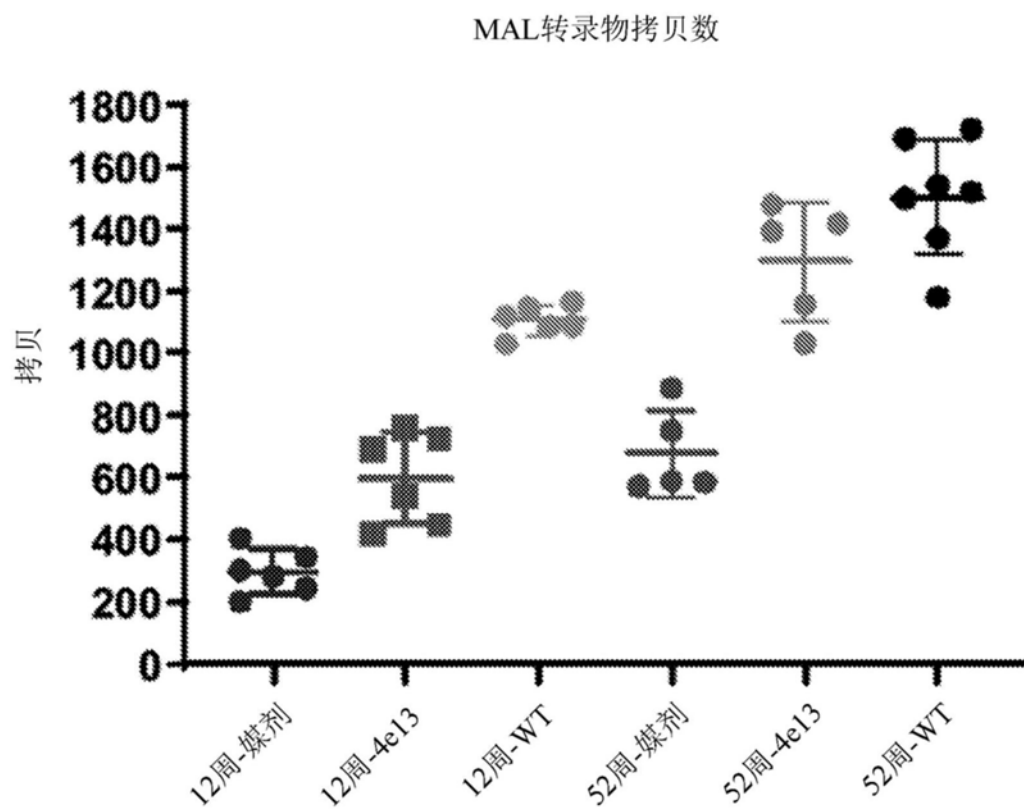


图3C

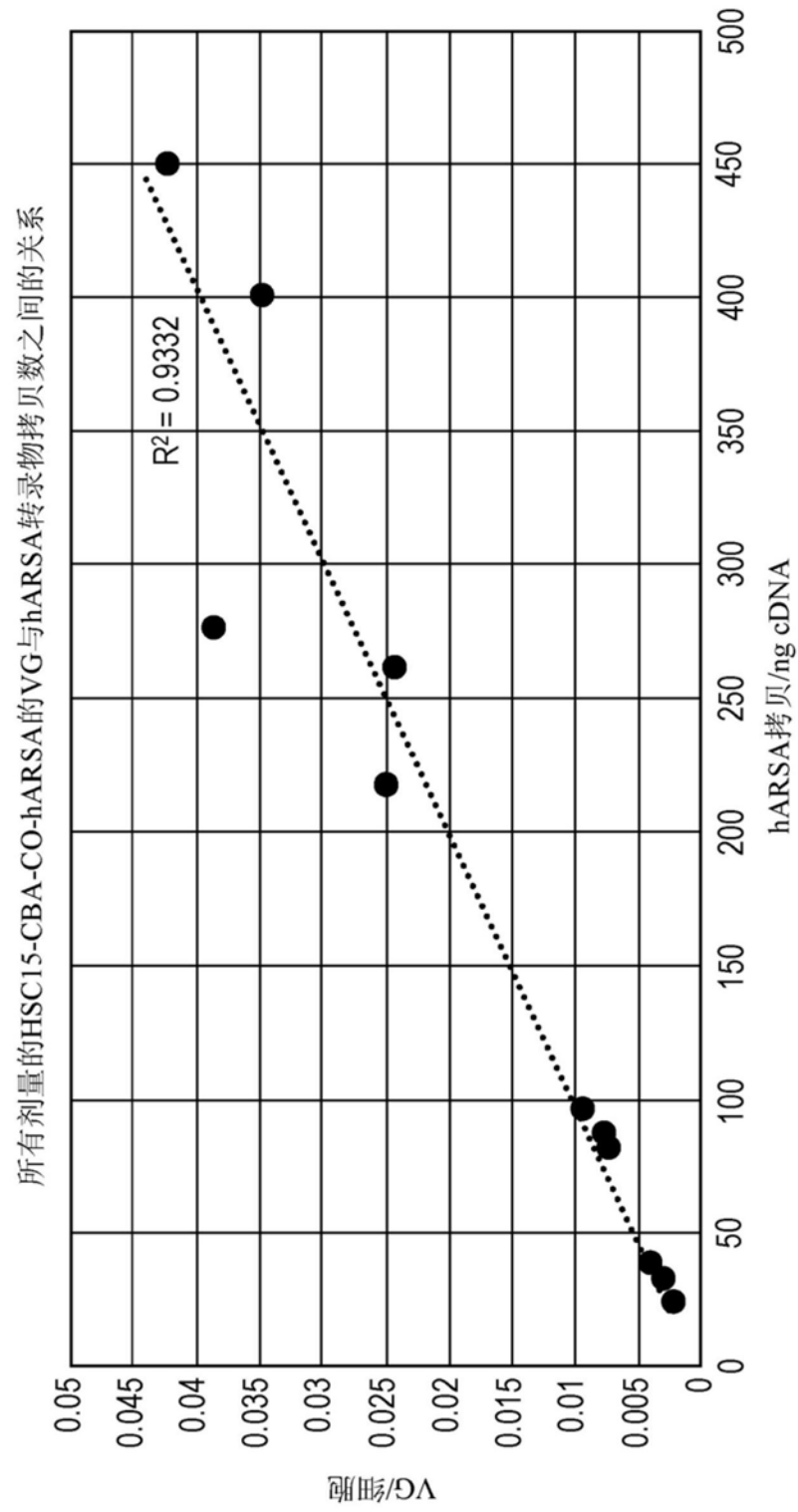


图4

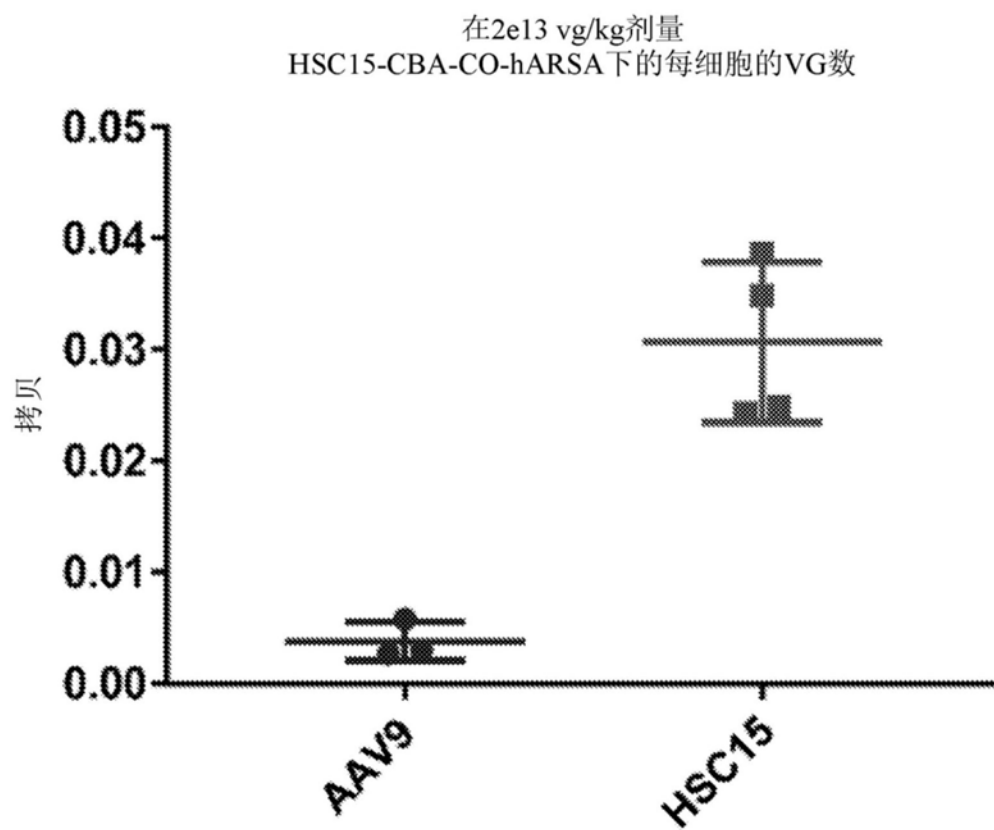


图5

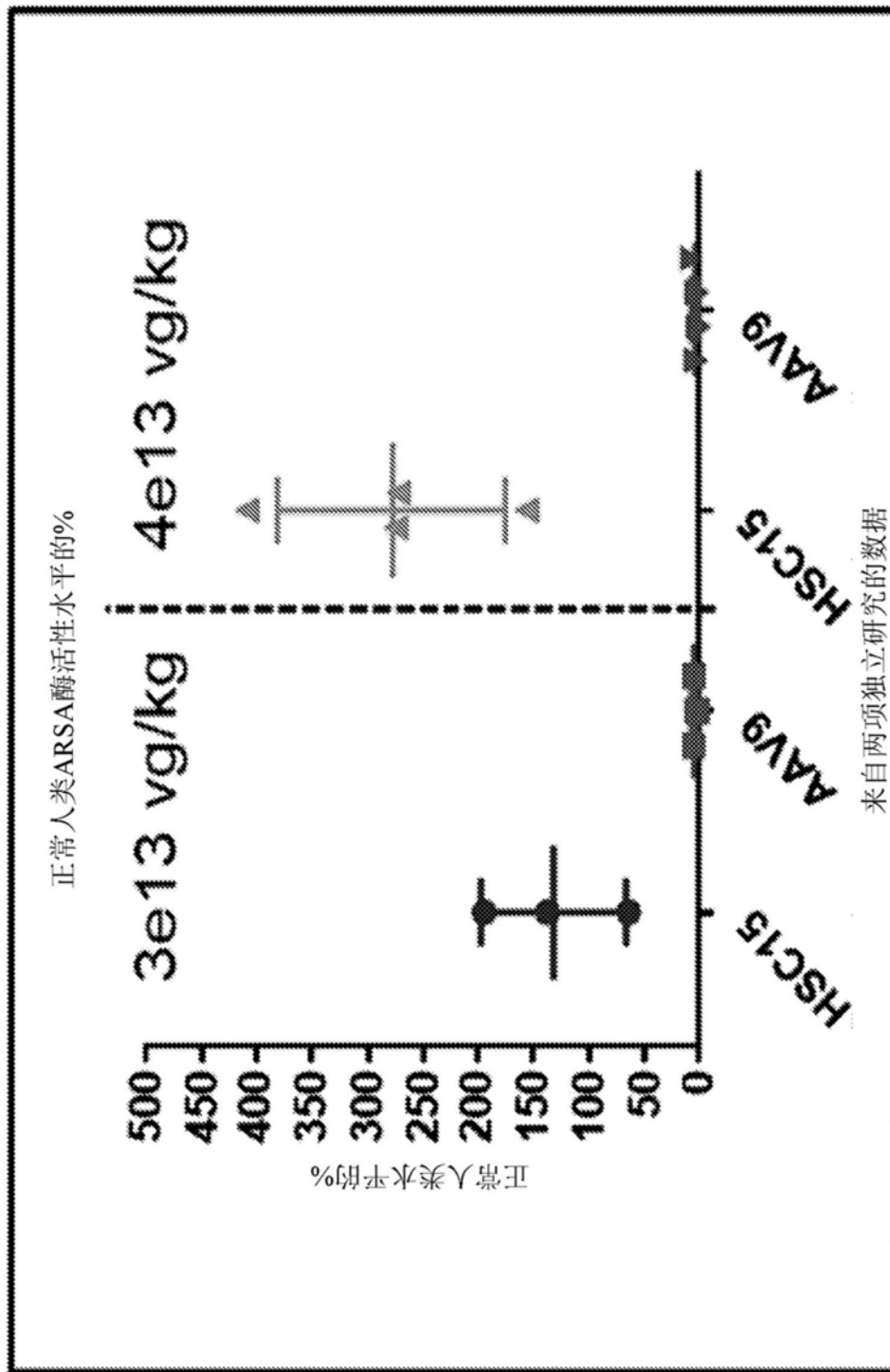


图6

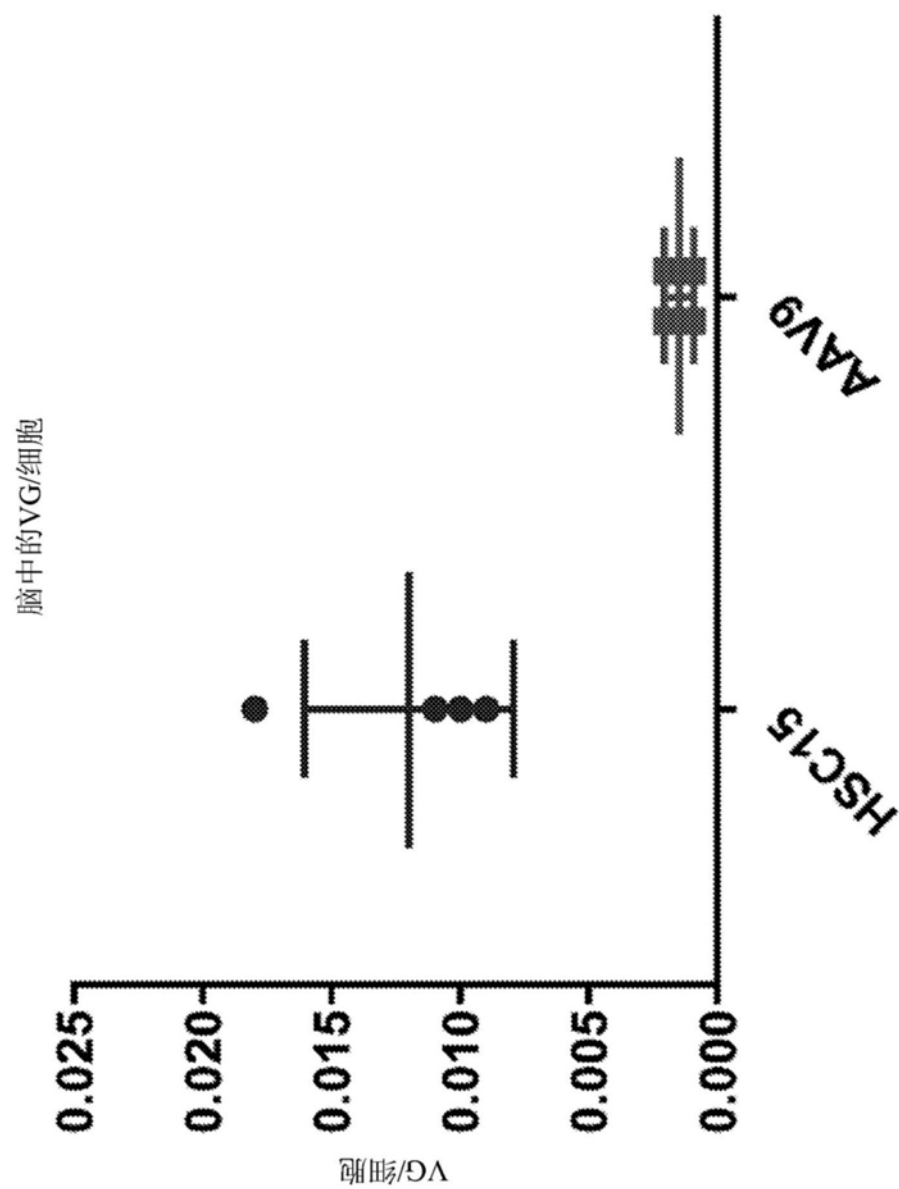


图7

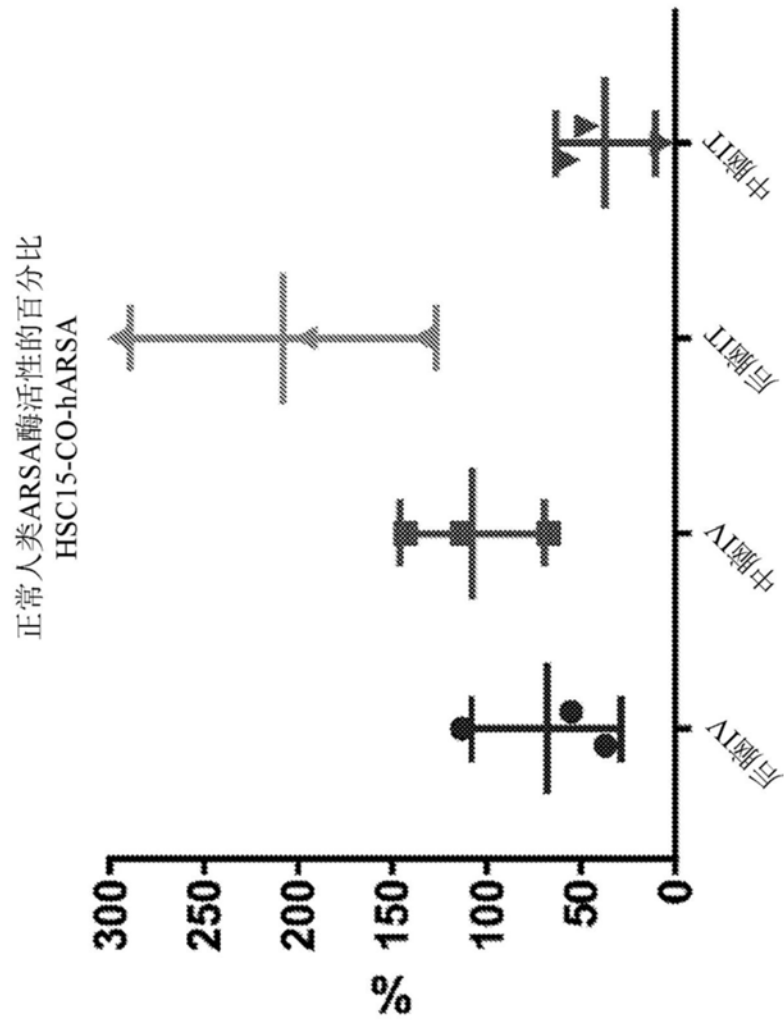


图8

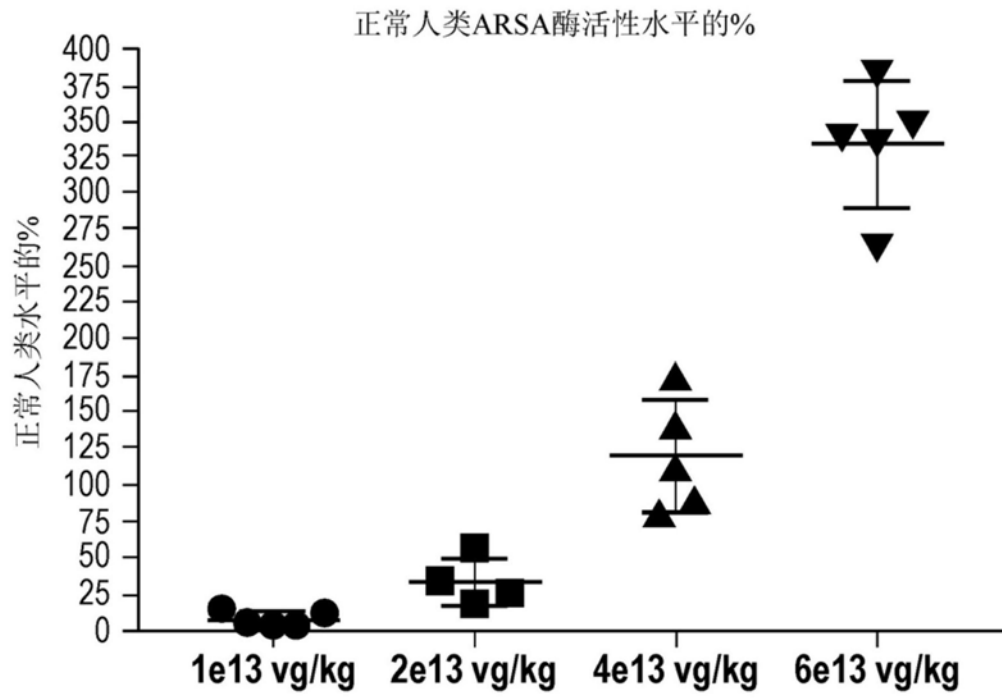


图9A

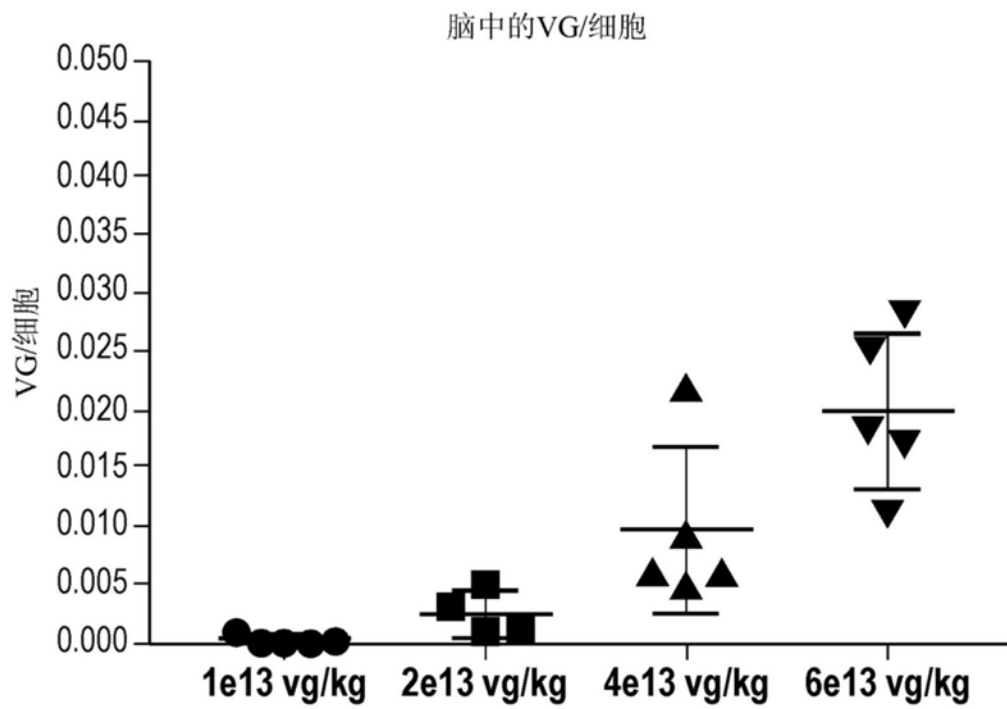


图9B

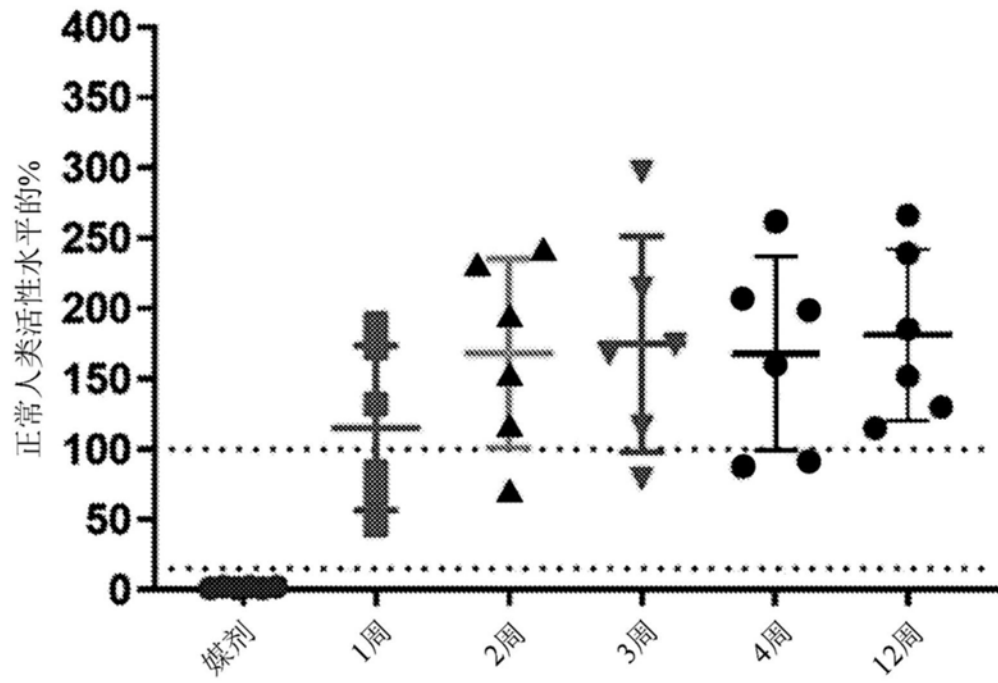


图9C

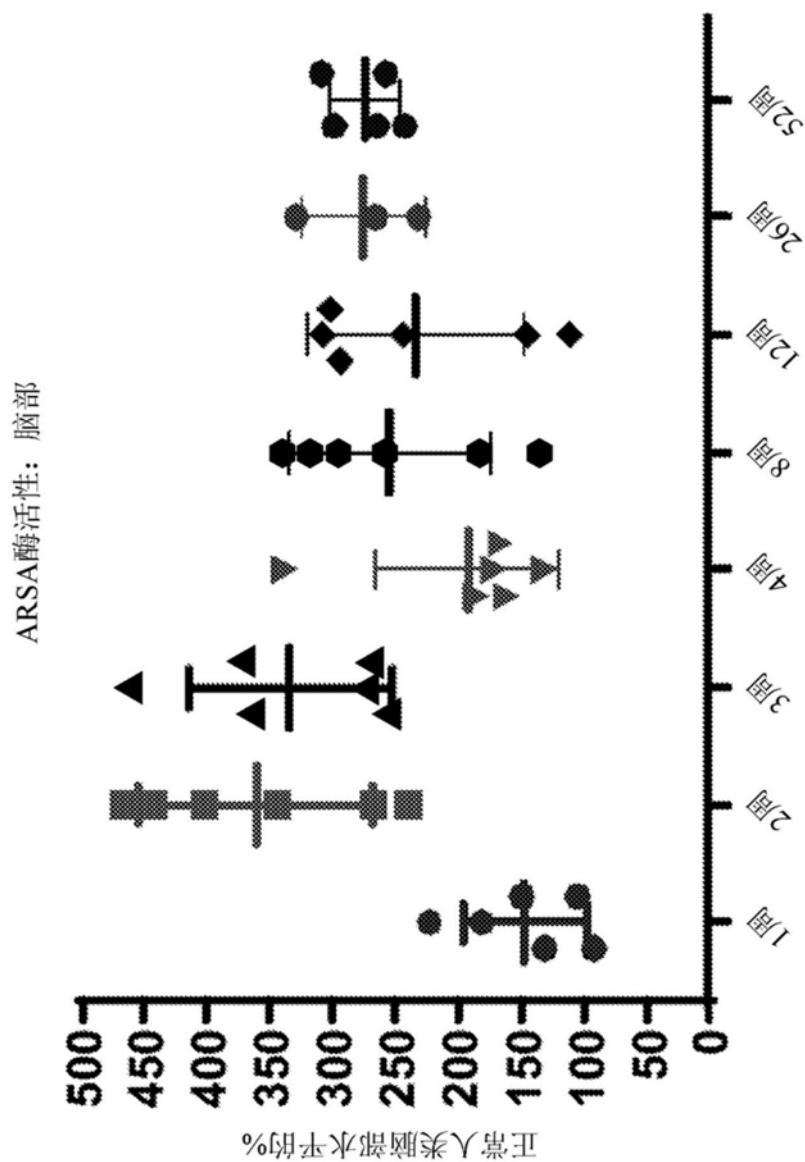


图9D

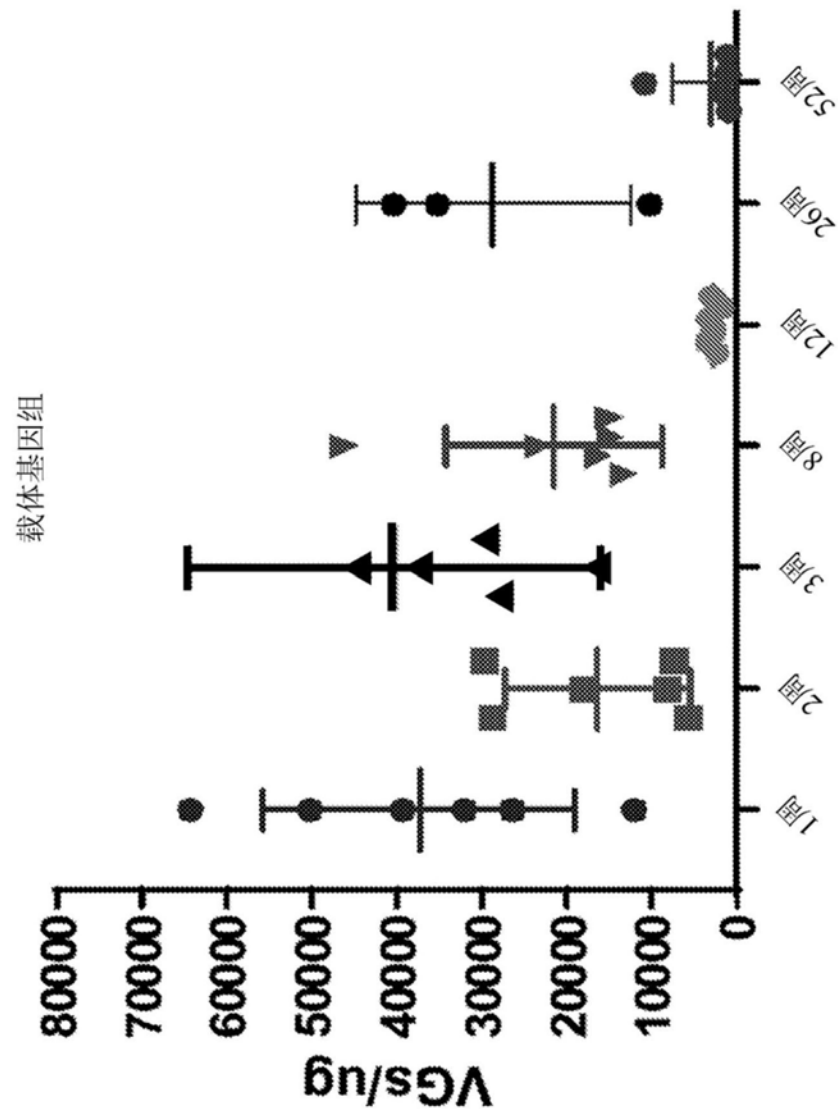


图9E

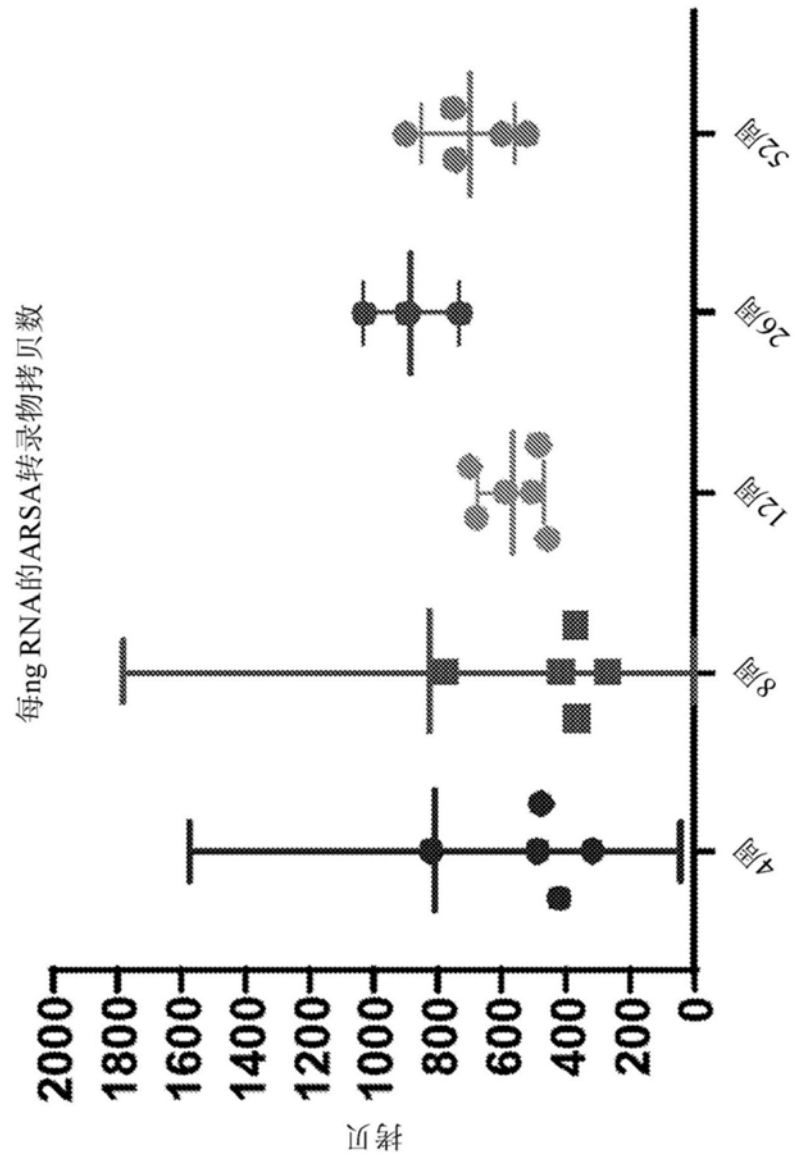


图9F

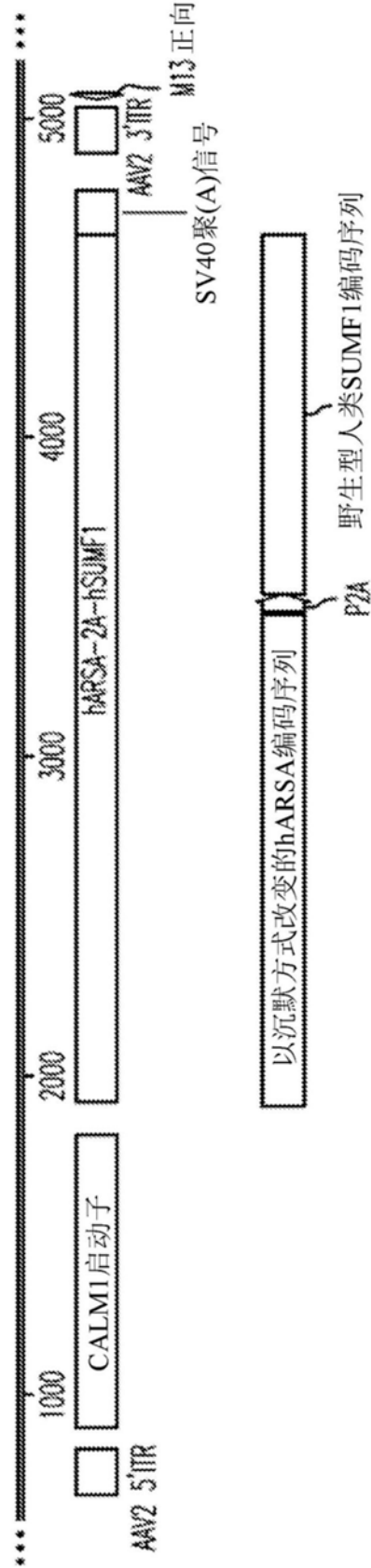


图10A

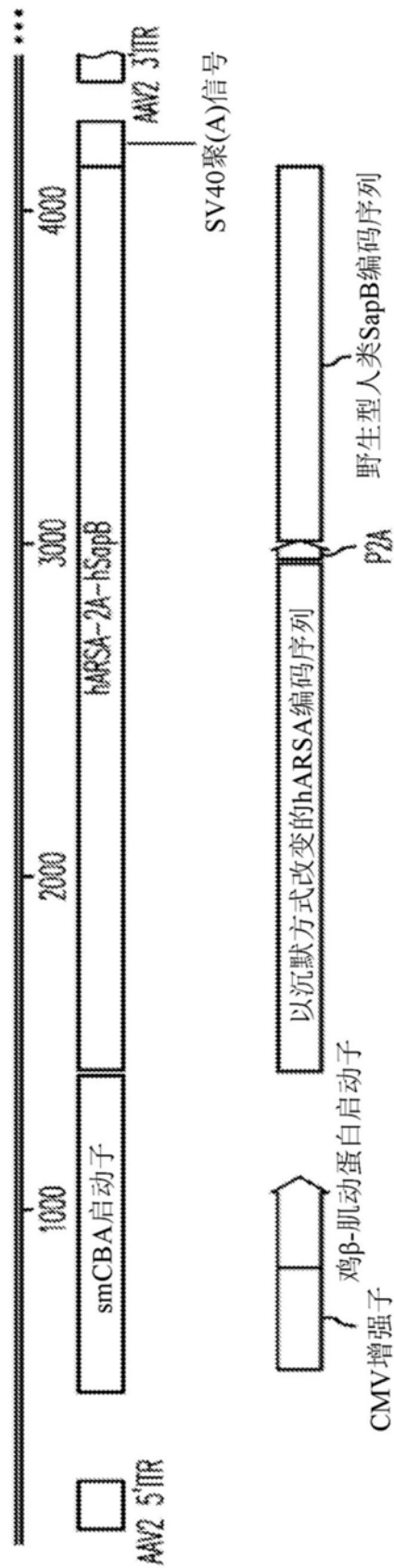


图10B

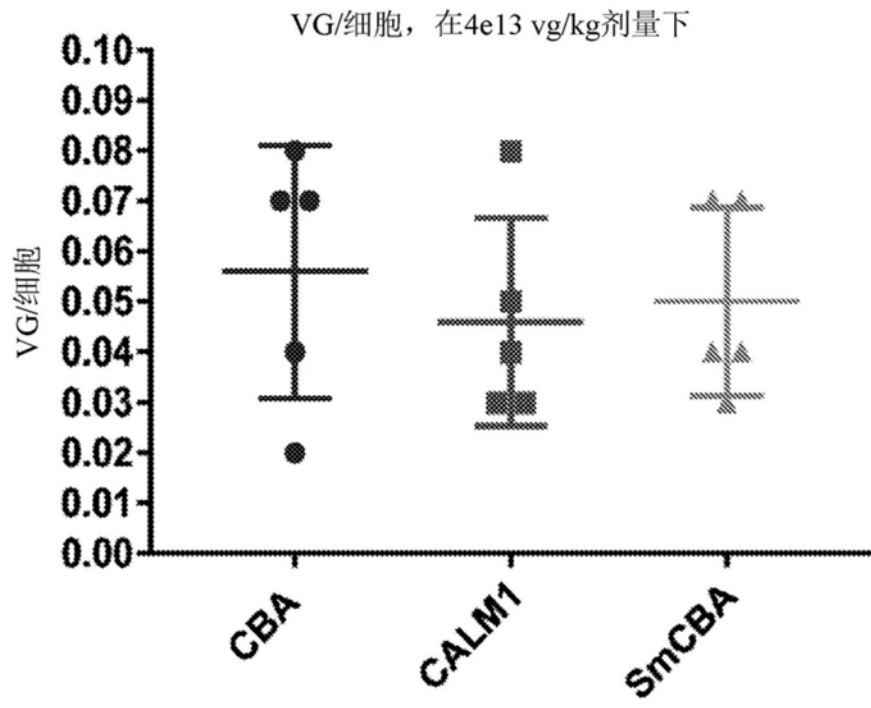


图11

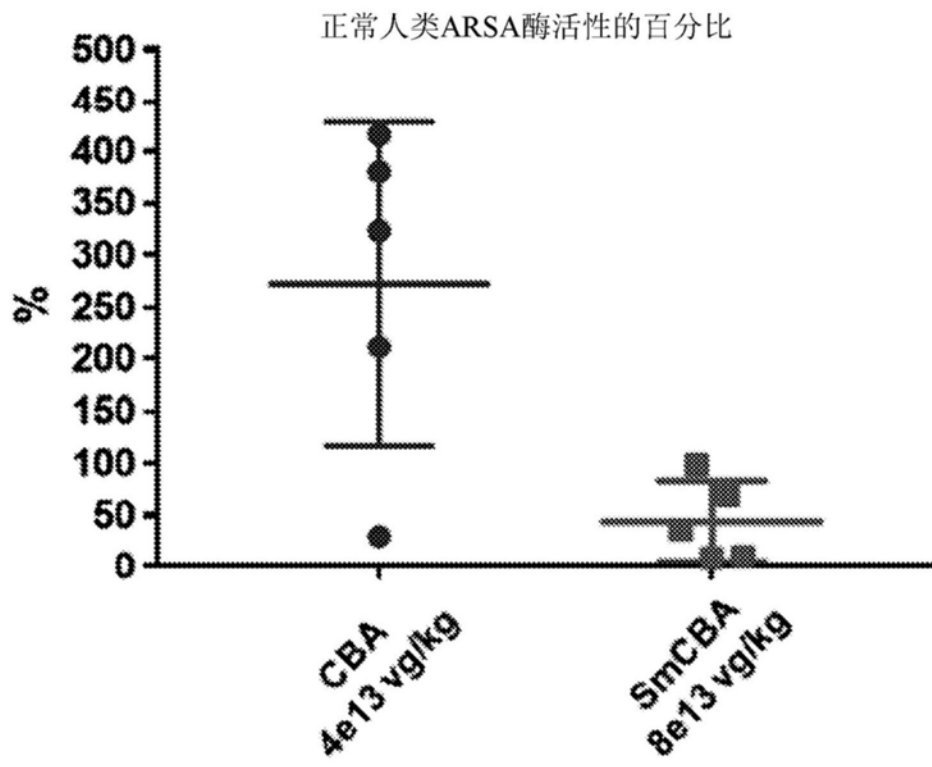


图12

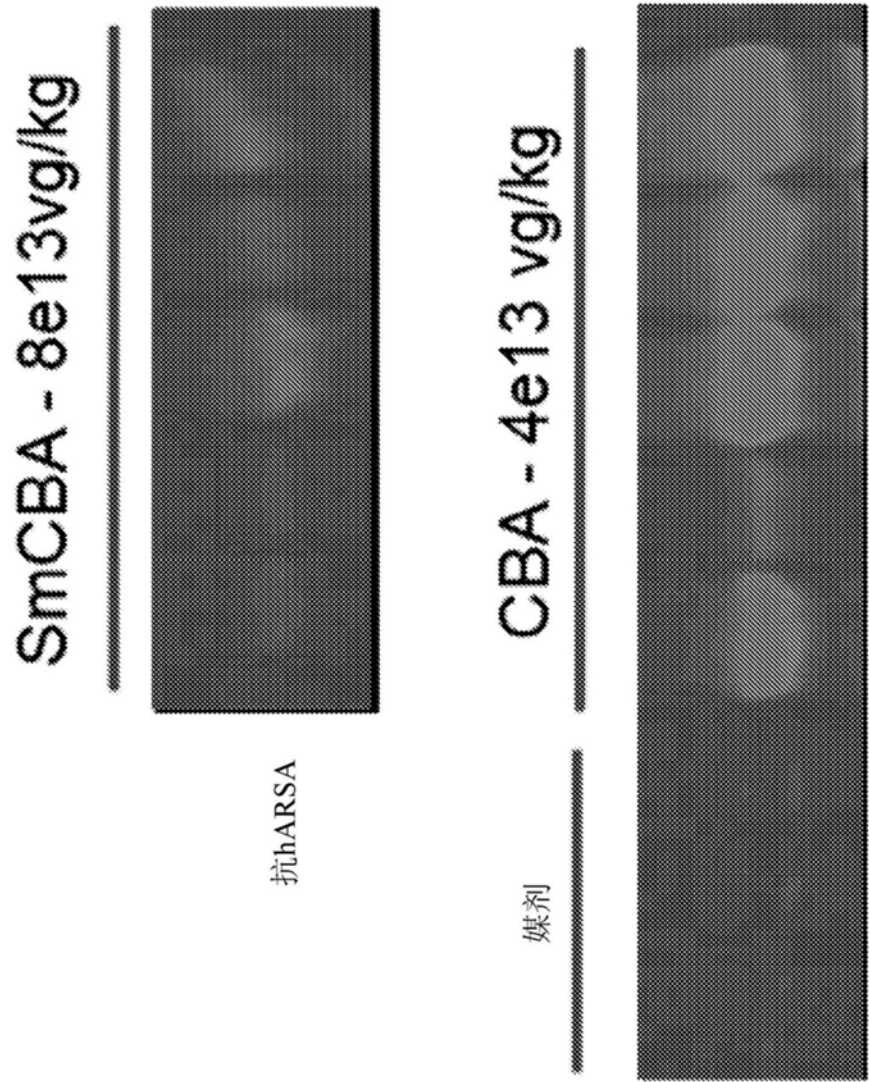


图13

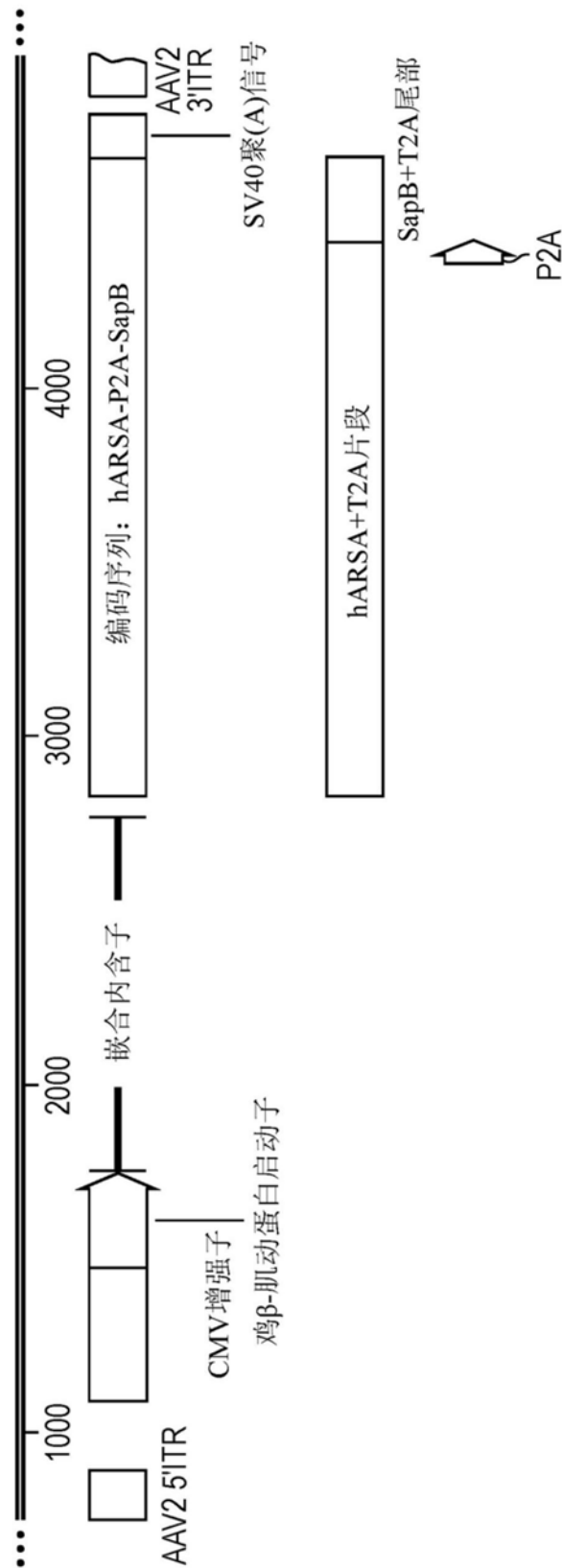


图14

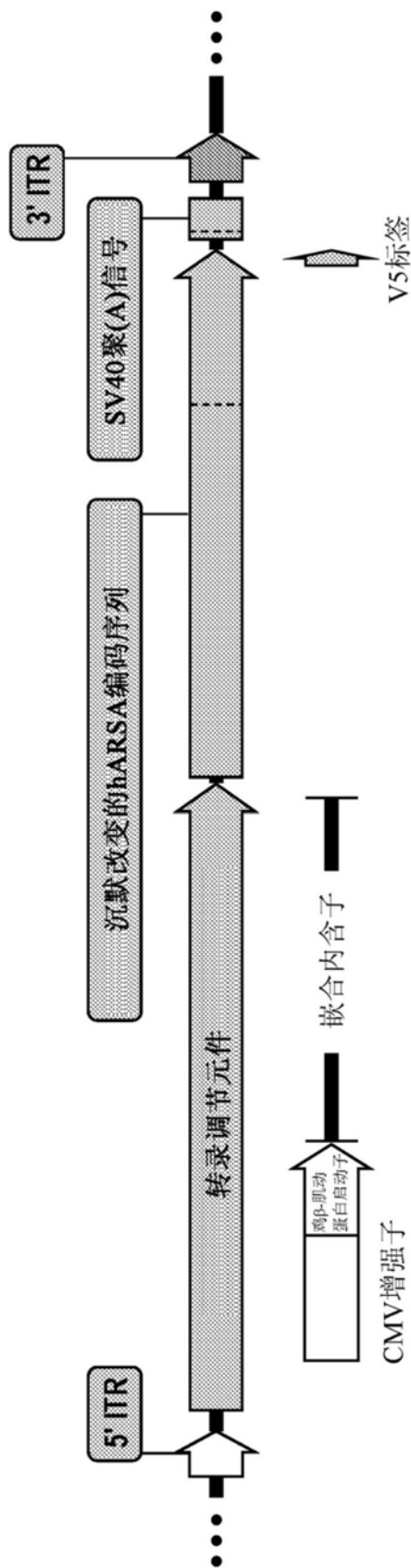


图15

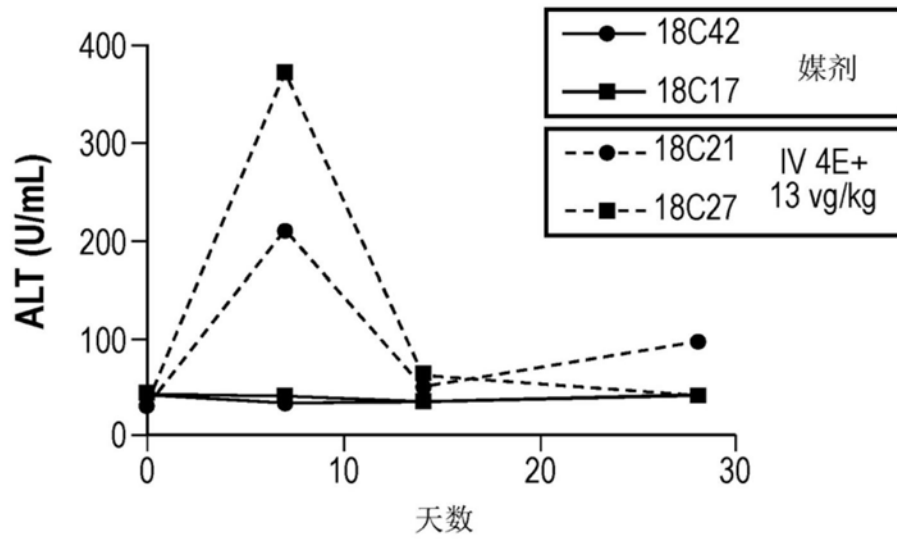


图16

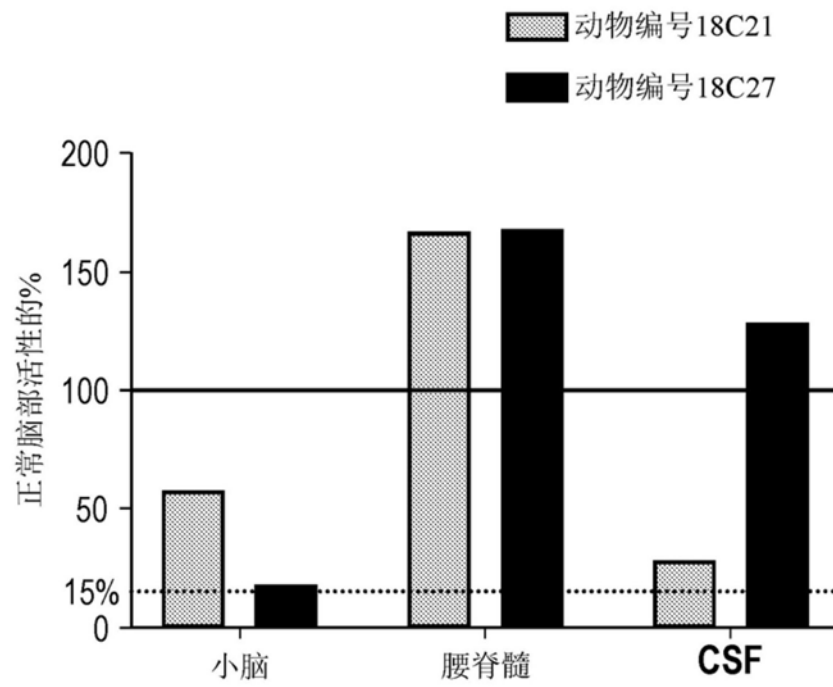


图17