



(12) 发明专利申请

(10) 申请公布号 CN 119630786 A

(43) 申请公布日 2025. 03. 14

(21) 申请号 202380051164.4

(22) 申请日 2023.05.02

(30) 优先权数据

63/337564 2022.05.02 US

(85) PCT国际申请进入国家阶段日

2024.12.27

(86) PCT国际申请的申请数据

PCT/US2023/066470 2023.05.02

(87) PCT国际申请的公布数据

W02023/215727 EN 2023.11.09

(71) 申请人 加利福尼亚大学董事会

地址 美国加利福尼亚州

(72) 发明人 K·柯林斯 张筱竹 B·万特瑞克

H·E·厄普顿 S·帕姆

J·麦金太尔

(74) 专利代理机构 北京泛华伟业知识产权代理

有限公司 11280

专利代理师 郭广迅

(51) Int.Cl.

C12N 9/12 (2006.01)

C12N 15/90 (2006.01)

A61K 48/00 (2006.01)

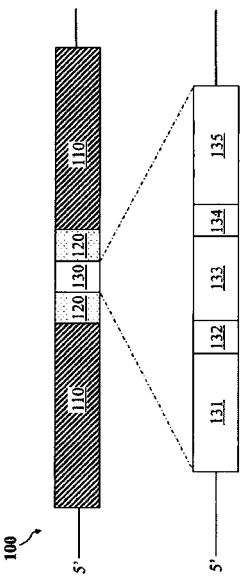
权利要求书5页 说明书281页 附图19页

(54) 发明名称

用于位点特异性基因组修饰的多组件系统

(57) 摘要

本发明包括用于通过逆转录酶相关的方法进行模块化基因编辑的系统、组合物和方法。特别提供使用修饰的核苷酸和肽的系统和方法。将转基因引入真核生物基因组为改善、纠正和/或改变基因表达提供巨大的机会,同时用于治疗或减轻疾病症状。成功的转基因插入将实现从功能丧失突变进行挽救,抑制功能获得性突变,外源性控制RNA和/或蛋白表达,引入同种型表达特异性,工程化的基因和蛋白表达,以及其它有用的结果。



1. 一种用于基因组编辑的系统,其包括:

(i) 至少一种逆转录酶构建体(RTC),所述RTC包含至少一个逆转录酶模块(RTC:RT模块)、至少一个逆转录酶构建体5'模块(RTC:5'模块)和/或至少一个逆转录酶构建体3'模块(RTC:3'模块),所述至少一个逆转录酶模块(RTC:RT模块)包含编码逆转录酶(RT)的mRNA,以及

(ii) 至少一种基因插入构建体(GIC),所述GIC包括至少一个适合由所述至少一种RTC编码的多肽进行逆转录的RNA模板,其中所述至少一种基因插入构建体包含至少一个任选的GIC:5'模块、至少一个GIC:有效载荷模块和至少一个GIC:3'模块。

2. 权利要求1所述的系统,其中:

i) 所述RTC 5'模块包含5'非翻译区域(5'-UTR)、Kozak序列或内部核糖体进入位点、非天然翻译起始密码子和/或5'帽;

ii) 所述RT模块包含编码来自选自白喉带鹀(*Zonotrichia albicollis*, ZoAl)、栗耳草雀(*Taeniopygia guttata*, TaGu)、白喉鵪(*Tinamus guttatus*, TiGu)、青鳉(*Oryzias latipes*, OrLa)和赤拟谷盗(*Tribolium castaneum*, (谱系B), TriCasB)的生物体的RT的mRNA;

iii) 所述RTC 3'模块包含逆转录酶翻译终止密码子、3'非翻译区域(3'UTR)和聚-A尾;

iv) 所述GIC:5'模块包含源自天然逆转录因子5'区的序列、rRNA序列、核糖酶序列、折叠基序序列和/或RNA聚合酶终止子序列;

(v) 所述GIC:有效载荷模块包含至少一个转基因ORF或非编码RNA(ncRNA)序列、转基因启动子序列、转基因5'非翻译序列、转基因3'非翻译序列、转基因多聚腺苷酸化信号序列和/或转基因ncRNA加工序列;以及

(iv) 所述GIC:3'模块包含逆转录酶识别序列、rRNA序列和/或A-区序列。

3. 权利要求1或2中任一项所述的系统,其中所述至少一种逆转录酶来自非长末端重复(非LTR)逆转录因子或其修饰的变体。

4. 权利要求1-3中任一项所述的系统,其中所述至少一种逆转录酶包含至少一个DNA结合域、至少一个RNA结合域、至少一个cDNA合成域、至少一个核酸内切酶结构域及其任意组合。

5. 权利要求1至4中任一项所述的系统,其中所述逆转录酶来自鸟类物种。

6. 权利要求5所述的系统,其中所述逆转录酶来自白喉带鹀(ZoAl)、栗耳草雀(TaGu)或白喉鵪(TiGu)。

7. 权利要求6所述的系统,其中所述逆转录酶包含与SEQ ID NO:18、SEQ ID NO:20、SEQ ID NO:27、SEQ ID NO:29或SEQ ID NO:25具有至少90%同一性的氨基酸序列。

8. 权利要求1-7中任一项所述的系统,其中所述至少一个逆转录模块包含或编码至少一个图2-5中图示说明的结构或其任意组合。

9. 权利要求1-8中任一项所述的系统,其中所述至少一种逆转录酶构建体包含、编码至少一个选自SEQ ID NO 1-57及其任意组合的序列或由至少一个选自SEQ ID NO 1-57及其任意组合的序列编码。

10. 权利要求2至9中任一项所述的系统,其中所述任选的至少一个GIC:5'模块rRNA序

列包含或编码1至30nt的受试者rRNA。

11.权利要求10所述的系统,其中所述rRNA序列包含选自SEQ ID NO:250-276的序列,或相对于选自SEQ ID NO:250-276的序列具有一个、两个或三个核苷酸变化的序列。

12.权利要求11所述的系统,其中所述GIC:5' 模块不包含rRNA序列。

13.权利要求2至12中任一项所述的系统,其中所述GIC:5' 模块核糖酶序列包含至少一种自剪切核糖酶,任选地其中所述自剪切核糖酶包含丁型肝炎病毒 (HDV) 核糖酶折叠。

14.根据权利要求13所述的系统,其中所述HDV核糖酶包含与选自SEQ ID NO:102-127和129-154的序列具有至少90%同一性的序列。

15.权利要求2至12中任一项所述的系统,其中所述GIC:5' 模块核糖酶序列包含来自至少一个非长末端重复逆转录因子的5' 区的核糖酶。

16.权利要求15所述的系统,其中所述核糖酶包含与选自SEQ ID NO:64-65、67、75-76、86、89-101和128的序列具有至少90%同一性的序列。

17.权利要求2至16中任一项所述的系统,其中所述GIC:5' 模块折叠基序序列包含至少一个自主折叠RNA序列基序,任选地其中所述自主折叠RNA序列基序包含至少一个发夹基序、至少一个茎环基序、至少一个配对的茎4基序或其任意组合。

18.权利要求17所述的系统,其中所述折叠基序序列包含SEQ ID NO 278或279,或与SEQ ID NO 278或279具有至少90%同一性的序列。

19.权利要求2至18中任一项所述的系统,其中所述GIC:5' 模块包含与选自SEQ ID NO 60-154的序列具有至少90%同一性的序列。

20.权利要求2至19中任一项所述的系统,其中所述GIC:3' 模块逆转录酶识别序列包含至少一个与至少一个逆转录酶相互作用的序列。

21.权利要求20所述的系统,其中所述GIC:3' 模块逆转录酶识别序列来自天然逆转录因子的3' 区。

22.权利要求20或21中任一项所述的系统,其中所述至少一个GIC:3' 模块逆转录酶识别序列包含与选自SEQ ID NO 200-224的序列具有至少90%同一性的序列。

23.权利要求2至22中任一项所述的系统,其中所述GIC:3' 模块rRNA序列包含1至30nt的rRNA。

24.权利要求23所述的系统,其中所述rRNA序列选自SEQ ID NO 280-289,或其包含一个或两个核苷酸取代的序列。

25.权利要求2至24中任一项所述的系统,其中所述GIC:3' 模块A-区序列包含1至50个腺嘌呤碱基。

26.权利要求2至25中任一项所述的系统,其中所述GIC:3' 模块包含与选自SEQ ID NO 300-329的序列具有至少90%同一性的序列或其任意组合。

27.权利要求2至25中任一项所述的系统,其中所述GIC:3' 模块包含来自ZoA1、TaGu、GeFo或TiGu的3' UTR序列。

28.权利要求27所述的系统,其中所述3' UTR序列包含与选自SEQ ID NO 202-205或SEQ ID NO 222-224的序列具有至少90%同一性的序列。

29.权利要求2至28中任一项所述的系统,其中所述至少一个转基因序列包含或编码至少一个用于插入受试者基因组的感兴趣序列。

30. 权利要求29所述的系统,其中所述转基因序列包含或编码至少一种mRNA、微小RNA、siRNA、rRNA、tRNA、长非编码RNA、小胞质RNA、小核RNA、小核仁RNA、小卡哈尔体(Cajal)RNA、环状RNA、调节RNA、肽、多肽、蛋白、抑制蛋白和/或控制至少一种转基因的表达的序列。

31. 权利要求30所述的系统,其中所述转基因编码选自hTERT、hPAH、hFactor VIII、具有可变大B结构域的突变的hFactor VIII或Factor IX的蛋白。

32. 权利要求2至31中任一项所述的系统,其中所述转基因启动子序列包含至少一个促进转基因在受试者基因组中的表达的序列。

33. 权利要求2至32中任一项所述的系统,其中所述转基因5'非翻译序列包含至少一个转基因mRNA 5'非翻译区。

34. 权利要求2至33中任一项所述的系统,其中所述转基因3'非翻译序列包含至少一个转基因mRNA 3'非翻译区。

35. 权利要求2至34中任一项所述的系统,其中所述转基因多聚腺苷酸化信号序列包含至少一种转基因多聚腺苷酸化信号。

36. 权利要求2至35中任一项所述的系统,其中所述转基因非编码RNA(ncRNA)加工序列包含用于至少一种转基因表达的ncRNA的至少一种终止信号、至少一种3'加工信号及其任意组合。

37. 权利要求2至36中任一项所述的系统,其中所述至少一种GIC:有效载荷模块包含或编码至少一个与选自SEQ ID NO 411-422或SEQ ID NO 499-536的序列具有至少90%同一性的序列或其任意组合。

38. 权利要求1-37中任一项所述的系统,其中所述至少一个GIC:5'模块和至少一个GIC:3'模块中的至少一个包含或编码至少一个序列,所述至少一个序列源自与其它至少一个GIC:5'模块和至少一个GIC:3'模块中的至少一个不同的非长末端重复逆转录因子的物种。

39. 权利要求1-38中任一项所述的系统,其中所述至少一种基因插入构建体包含或编码至少一个图6-9图示说明的结构及其任意组合。

40. 权利要求1至39中任一项所述的系统,其包含两种不同的基因插入构建体,所述两种不同的基因插入构建体包含GIC:有效载荷模块,所述GIC:有效载荷模块包含不同的转基因ORF。

41. 权利要求40所述的系统,其中所述两种不同的GIC存在于相同的RNA模板上。

42. 权利要求40所述的系统,其中所述两种不同的GIC存在于不同的RNA模板上。

43. 权利要求1至42中任一项所述的系统,其中所述系统包括:

(i) 至少一种逆转录酶构建体,其中所述至少一种逆转录酶构建体包含至少一个与选自SEQ ID NO 1-57的序列具有至少90%同一性的序列或由至少一种与选自SEQ ID NO 1-57的序列具有至少90%同一性的序列编码;

(ii) 至少一种基因插入构建体,其中所述至少一种基因插入构建体包含:

GIC:5'模块,其包含与选自SEQ ID NO:60-154的序列具有至少90%同一性的序列;

rRNA序列,其包含选自SEQ ID NO:250-276的序列,或相对于选自SEQ ID NO:250-276的序列具有一个、两个或三个核苷酸变化的序列;或不包含rRNA序列;

GIC:有效载荷模块,其包含至少一个转基因序列;以及

- GIC:3' 模块,其包含与选自SEQ ID NO 300-329的序列具有至少90%同一性的序列;
- GIC:3' 模块逆转录酶识别序列,其包含与选自SEQ ID NO 200-224的序列具有至少90%同一性的序列;
- 选自SEQ ID NO 280-289的GIC:3' 模块rRNA序列,或其包含一个或两个核苷酸取代的序列;
- GIC:3' 模块A-区序列,其包含1至100个腺嘌呤碱基。
- 44.权利要求43所述的系统,其中所述GIC:有效载荷模块包含至少一个与选自SEQ ID NO 411-422或499-536的序列具有至少90%同一性的序列。
- 45.权利要求1-44中任一项所述的系统,其中所述至少一种逆转录酶构建体和至少一种基因插入构建体中的至少一个包含或编码至少一个序列,所述至少一个序列源自与其它至少一种逆转录酶构建体和至少一种基因插入构建体中的至少一个不同物种的逆转录因子。
- 46.权利要求1-45中任一项所述的系统,其中所述RTC和/或GIC RNA包含至少一个修饰的尿嘧啶,或所述RTC和/或GIC RNA包含100%修饰的尿嘧啶。
- 47.权利要求46所述的系统,其中所述修饰的尿嘧啶选自5-甲基尿苷、5-甲氧基尿苷、假尿苷、N1-甲基假尿苷和/或2-硫代尿苷。
- 48.一种用于将至少一个转基因插入受试者基因组的方法,其包括向受试者给予有效量的至少一种权利要求1-47中任一项所述的基因插入系统 (GIS)。
- 49.权利要求48所述的方法,其中所述转基因被插入受试者基因组中的一个或多个靶位点,任选地其中所述一个或多个靶位点包括至少一个安全港位点。
- 50.权利要求49所述的方法,其中所述任选的至少一个安全港位点包含至少一个核糖体DNA (rDNA) 序列,任选地其中所述至少一个核糖体DNA序列包括至少一个28S rDNA序列。
- 51.权利要求48-50中任一项所述的方法,其包括给予用至少一种递送剂配制的至少一种基因插入系统。
- 52.权利要求51所述的方法,其中所述至少一种递送剂是至少一种纳米颗粒,任选地其中所述至少一种纳米颗粒包括至少一种脂质纳米颗粒。
- 53.权利要求48-52中任一项所述的方法,其中所述转基因以大于90%的靶位点特异性被插入。
- 54.权利要求53所述的方法,其中所述RTC RNA编码来自白喉带鹀 (ZoA1)、栗耳草雀 (TaGu) 或白喉鵂 (TiGu) 的RT,或包含与SEQ ID NO:18、SEQ ID NO:20、SEQ ID NO:27、SEQ ID NO:29或SEQ ID NO:25具有至少90%同一性的氨基酸序列。
- 55.权利要求48-54中任一项所述的方法,其中所述转基因在靶位点表达3个月或更长时间。
- 56.一种药物组合物,其包含至少一种权利要求1-47所述的基因插入系统,任选的至少一种赋形剂、至少一种递送剂、至少一种佐剂和其任意组合中的至少一种。
- 57.一种治疗有其需要的受试者的治疗指征的方法,其包括给予有效量的至少一种权利要求1-47所述的基因插入系统或权利要求56所述的药物组合物,任选地包括至少一种权利要求48-52所述的方法。
- 58.权利要求57所述的方法,其中所述治疗指征是由端粒酶活性丧失引起的。

59. 权利要求57或58中任一项所述的方法,其中所述至少一种基因插入系统包含至少一个TERT转基因。

60. 一种用于制造基因插入系统的试剂盒,其包含权利要求1-47所述的基因插入系统,任选的权利要求56所述的药物组合物,和任选地进一步包含缓冲液、DNA质粒或用于制造所述基因插入系统或药物组合物的方案。

61. 一种方法,其包括从头设计5' 模块,所述5' 模块招募用于第二链产生切口的宿主机器,以及因此的第二链合成。

62. 权利要求61所述的方法,其通过从头设计5' 模块以 (a) 包含预定长度和位置的rRNA, (b) 具有增强的RZ折叠,和/或 (c) 招募宿主细胞机器来提供插入增益的效率。

63. 一种用于将至少一个转基因插入细胞的基因组的方法,其包括使所述细胞与权利要求1-47中任一项所述的至少一种基因插入系统 (GIS) 接触。

64. 权利要求63所述的方法,其中所述转基因被插入受试者基因组中的一个或多个靶位点,任选地其中所述一个或多个靶位点包括至少一个安全港位点。

65. 权利要求64所述的方法,其中所述任选的至少一个安全港位点包含至少一个核糖体DNA (rDNA) 序列,任选地其中所述至少一个核糖体DNA序列包括至少一个28S rDNA序列。

66. 权利要求63-65中任一项所述的方法,其包括给予用至少一种递送剂配制的至少一种基因插入系统。

67. 权利要求66所述的方法,其中所述至少一种递送剂是至少一种纳米颗粒,任选地其中所述至少一种纳米颗粒包括至少一种脂质纳米颗粒。

68. 权利要求63-67中任一项所述的方法,其中所述转基因以大于90%的靶位点特异性被插入。

69. 权利要求68所述的方法,其中所述RTC RNA编码来自白喉带鸚 (ZoA1)、栗耳草雀 (TaGu) 或白喉鵒 (TiGu) 的RT,或包含与SEQ ID NO:18、SEQ ID NO:20、SEQ ID NO:27、SEQ ID NO:29或SEQ ID NO:25具有至少90%同一性的氨基酸序列。

70. 权利要求63-69中任一项所述的方法,其中所述转基因在靶位点表达3个月或更长时间。

71. 权利要求63-70中任一项所述的方法,其中所述RTC与GIC的摩尔比为约10:1至1:20。

72. 权利要求63-71中任一项所述的方法,其中所述方法是体外方法、离体方法或体内方法。

73. 权利要求63-72中任一项所述的方法,其中所述细胞选自原代细胞、转化的细胞、上皮细胞、成纤维细胞、人细胞、猴细胞和小鼠细胞。

74. 权利要求63-73中任一项所述的方法,其中所述细胞为异体细胞或自体细胞。

75. 权利要求74所述的方法,其中所述自体细胞是HLA匹配的细胞。

用于位点特异性基因组修饰的多组件系统

[0001] 相关申请的交叉引用

[0002] 本申请要求2022年5月2日提交的第63/337,564号美国临时申请的优先权,其公开内容在此通过引用整体并入,用于所有目的。

[0003] 对序列表的引用

[0004] 本申请与电子格式的序列表一起提交。序列表文件命名为[[XXXX.xml]],创建于[[XX]],大小为[[XX]]字节。序列表的电子格式信息通过引用整体并入本文。

[0005] 政府支持声明

[0006] 本发明是在美国国立卫生研究院(the National Institutes of Health)授予的拨款号为GM139306和HL156819的政府支持下完成的。政府对本发明享有一定权利。

[0007] 公开内容的背景

[0008] 将转基因引入真核生物基因组为改善、纠正和/或改变基因表达提供巨大的机会,同时用于治疗或减轻疾病症状。成功的转基因插入将实现从功能丧失突变进行挽救,抑制功能获得性突变,外源性控制RNA和/或蛋白表达,引入同种型表达特异性,工程化的基因和蛋白表达,以及其它有用的结果。

[0009] 目前将遗传物质引入细胞用于插入基因组的方法仍有重大障碍要克服。例如,将DNA递送到靶细胞的方法需要DNA穿过细胞的细胞质,这经常诱发破坏性或有害的免疫反应。此外,用于通过同源重组(HR)将引入的DNA位点特异性整合到基因组中的方法引入潜在地致突变的双链DNA断裂,并破坏整合位点处的受试者基因组和表观基因组。在高等真核生物中,尤其是在有丝分裂后的细胞中,该DNA整合经常不是位点特异性的,因为在大部分细胞周期中,HR受到抑制,有利于非同源末端连接。

[0010] 一种有效的、位点特异性的将转基因插入活细胞基因组的方法具有针对DNA长度的灵活性,在在对细胞质中的DNA没有潜在影响(potential)的情况下完成,这将对人、动物、微生物和植物生物学的巨大贡献,具有强大的研究和临床应用。

[0011] 一种方法将是引入作为RNA的转基因序列,该RNA可作用于通过逆转录酶(RT)合成互补DNA(cDNA)的模板。然而,目前还没有鉴定出这样的分子信号,即其能够实现被引入哺乳动物细胞的RNA被复制为用于将转基因插入基因组中的模板。

[0012] 被称为非长末端重复(LTR)逆转录因子(RE)或也称为非LTR逆转录转座子的一类基因提供令人兴奋的有潜力的解决方案。这些基因能够在它们的宿主基因组内自我扩增。它们通过表达非LTR逆转录转座子RT蛋白(RT)发挥作用,该非LTR逆转录转座子RT蛋白使用其自身的逆转录因子转录RNA作为模板和基因组DNA中的切口(由RT蛋白的核酸内切酶(EN)结构域催化)作为cDNA合成启动的引物(RT引物扩展)结合并合成cDNA。该过程被称为靶点引发的逆转录(TPRT),在基因组中增加另一个双链DNA逆转录因子的拷贝。

[0013] W02022/155055描述一种用于在人类基因组中插入位点特异性安全位点(safe-harbor)转基因的双组件系统。所述两个组件是非LTR逆转录因子逆转录酶(RT)和与该RT匹配的模板RNA,该RT经工程化以实现全长转基因插入,而不是天然逆转录因子所倾向的5'插入截断。第一条插入的DNA链的合成机制是靶点引发的逆转录(TPRT),由模板RNA 3'模块引

导,并被该3'模块的非天然3'尾部部分增强。5'模块的作用是提供模板RNA生物稳定性,提高模板RNA结合RT蛋白的生物利用率,以及指导第二链合成。

[0014] 通过创建部分源自逆转录因子序列的生物聚合物构建体,本公开内容提供用于将转基因插入和表达至真核生物,特别是人细胞基因组中的组合物和方法。

[0015] 公开内容的概述

[0016] 本发明提供蛋白和核苷酸以及修饰的蛋白和多核苷酸的组合物、方法和/或用途,它们使用源自非长末端重复(非LTR)逆转录转座子的组件实现靶点引发的逆转录(TPRT)转基因插入到受试者基因组中。

[0017] 本发明提供一种用于基因组编辑的系统,该系统包括(i)至少一种逆转录酶构建体(RTC),所述RTC包含编码多肽的多核苷酸,所述多肽具有逆转录多核苷酸模板的酶活性;以及(ii)至少一种基因插入构建体(GIC),所述GIC包含至少一种适合由至少一种RTC编码的多肽进行逆转录的多核苷酸模板。

[0018] 在一些实施方案中,用于基因组编辑的系统包含:

[0019] (i)至少一种逆转录酶构建体(RTC),所述RTC包含至少一个逆转录酶模块(RTC:RT模块)、至少一个逆转录酶构建体5'模块(RTC:5'模块)和/或至少一个逆转录酶构建体3'模块(RTC:3'模块),所述逆转录酶模块(RTC:RT模块)包含编码逆转录酶(RT)的mRNA,以及

[0020] (ii)至少一种基因插入构建体(GIC),所述GIC包含至少一个适合由至少一种RTC编码的多肽进行逆转录的RNA模板,其中所述至少一种基因插入构建体包含至少一个GIC:5'模块、至少一个GIC:有效载荷模块和/或至少一个GIC:3'模块。

[0021] 在一些实施方案中,RT模块包含编码来自生物体的RT的mRNA,所述生物体选自鸟类、节肢动物、鱼类、被囊动物或其它动物,包括哺乳动物和人。

[0022] 在一些实施方案中,用于基因组编辑的系统包含:

[0023] i)RTC 5'模块,其包含5'非翻译区(5'-UTR)、Kozak序列、非天然翻译起始密码子和/或5'帽;

[0024] ii)RT模块,其包含编码来自选自白喉带鹀(*Zonotrichia albicollis*, ZoAl)、栗耳草雀(*Taeniopygia guttata*, TaGu)、白喉鵪(*Tinamus guttatus*, TiGu)、青鳉(*Oryzias latipes*, OrLa)和赤拟谷盗(*Tribolium castaneum*, (谱系B), TriCasB)的生物体的RT的mRNA;

[0025] iii)RTC 3'模块,其包含逆转录酶翻译终止密码子、3'非翻译区(3'UTR)和聚-A尾;

[0026] iv)GIC:5'模块,其包含源自天然逆转录因子5'区的序列、rRNA序列、核糖酶序列、折叠基序序列和/或RNA聚合酶I终止子序列;

[0027] (v)GIC:有效载荷模块,其包含至少一个转基因ORF或非编码RNA(ncRNA)序列、转基因启动子序列或内部核糖体进入位点(IRES)、转基因5'非翻译序列、转基因3'非翻译序列、转基因多聚腺苷酸化信号序列和/或转基因ncRNA加工序列;以及

[0028] (iv)GIC:3'模块,其包含逆转录酶识别序列、rRNA序列和/或A-区(A-Tract)序列。

[0029] 在一些实施方案中,至少一种逆转录酶构建体包含至少一种生物聚合物,所述生物聚合物包括至少一种核酸、至少一种氨基酸及其任意组合。在一些实施方案中,上述(i)的RTC多核苷酸包含编码逆转录酶的mRNA。在一些实施方案中,上述(ii)的GIC多核苷酸模

板包含RNA。在一些实施方案中,上述(i)的多核苷酸包含编码逆转录酶的mRNA,并且上述(ii)的GIC多核苷酸模板包含分开的(不同的)RNA。在一些实施方案中,GIC包含与编码(i)的RT的mRNA不同的RNA模板。

[0030] 在一些实施方案中,至少一种逆转录酶构建体包含至少一个逆转录酶开放阅读框(ORF)模块(RTC:RT-模块)、任选的至少一个逆转录酶构建体5'非翻译区(UTR)模块(RTC:5'模块)、任选的至少一个逆转录酶构建体3'UTR模块(RTC:3'模块)及其任意组合。

[0031] 在一些实施方案中,至少一个逆转录酶模块包含或编码至少一种逆转录酶。

[0032] 在一些实施方案中,至少一个逆转录酶模块包含或编码至少一种来自非长末端重复(非LTR)逆转录因子的逆转录酶。

[0033] 在一些实施方案中,至少一种逆转录酶包含或编码非天然翻译起始密码子。

[0034] 在一些实施方案中,至少一种逆转录酶包含至少一个DNA结合域、至少一个RNA结合域、至少一个cDNA合成域、至少一个核酸内切酶结构域及其任意组合。

[0035] 在一些实施方案中,至少一个逆转录酶结构域、至少一个受试者DNA结合域、至少一个模板RNA结合域和至少一个核酸内切酶结构域中的至少一个及其任意组合源自一种逆转录酶,该逆转录酶不同于其它至少一个逆转录酶结构域、至少一个受试者DNA结合域、至少一个模板RNA结合域和至少一个核酸内切酶结构域中的至少一个。

[0036] 在一些实施方案中,至少一个逆转录酶构建体5'模块包含或编码至少一个RNA聚合酶启动子、至少一个5'非翻译区(5'-UTR)、至少一个Kozak序列、至少一个5'帽及其任意组合。

[0037] 在一些实施方案中,至少一个逆转录酶构建体3'模块包含或编码至少一个逆转录酶翻译终止密码子、至少一个3'非翻译区(3'UTR)、至少一个聚-A区和/或尾及其任意组合。

[0038] 在一些实施方案中,至少一个逆转录模块包含或编码至少一个图2-5图示说明的结构或其任意组合。

[0039] 在一些实施方案中,至少一种逆转录酶构建体包含、编码SEQ ID NO 1-57中的至少一个或由SEQ ID NO 1-57中的至少一个编码。在一些实施方案中,至少一种逆转录酶构建体包含编码RT蛋白的mRNA,所述RT蛋白来自选自TriCasB、NaViB、OrLa、ZoAl、TiGu、TaGu、GeFo、DroSi、BoMo、DrMerc、DrMe、GaAc、PuPu、AdVa、HyMaA、CiIn、LiPo、TriCan、LeCo及其任意组合的一种。

[0040] 在一些实施方案中,至少一种基因插入构建体包含或编码至少一种核酸生物聚合物。在一些实施方案中,基因插入构建体包含模板RNA。

[0041] 在一些实施方案中,至少一种基因插入构建体包含或编码至少一个任选的GIC:5'模块、至少一个GIC:有效载荷模块、至少一个任选的GIC:3'模块及其任意组合。

[0042] 在一些实施方案中,至少一个GIC:5'模块包含或编码至少一个源自天然逆转录因子5'区域的序列、任选的至少一个GIC:5'模块rRNA序列、任选的至少一个GIC:5'模块核糖酶(RZ)序列、任选的至少一个GIC:5'模块折叠基序序列或其任意组合。

[0043] 在一些实施方案中,任选的至少一个GIC:5'模块rRNA序列包含或编码1至30nt的受试者rRNA。

[0044] 在一些实施方案中,任选的至少一个GIC:5'模块核糖酶序列包含或编码至少一种自剪切核糖酶,任选地其中所述自剪切核糖酶包括肝炎δ病毒(HDV)核糖酶。

[0045] 在一些实施方案中,任选的至少一个GIC:5' 模块核糖酶序列包含或编码源自至少一个非长末端重复逆转录因子的5' 区的核糖酶。在一些实施方案中,任选的至少一个GIC:5' 模块折叠基序序列包含或编码至少一个自主折叠RNA序列基序,任选地其中所述自主折叠RNA序列基序包含在RZ内的至少一个发夹基序、至少一个茎环基序、至少一个配对的茎基序或其任意组合。

[0046] 在一些实施方案中,GIC:5' 模块包含或编码SEQ ID NO 60-154中的至少一个,或与SEQ ID NO 60-154中的至少一个具有至少90%同一性(例如,至少90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%同一性)的序列。在一些实施方案中,GIC:5' 模块包含来自选自以下种类的序列:OrLa、TriCasB、TriCasA、ZoAl、TiGu、DroSi、LeCo、CiIn、FoRa、TriCan、HDV-28、HDV-24、HDV-21、HDV-13、HDV-36或其任意组合。

[0047] 在一些实施方案中,至少一个GIC:3' 模块包含或编码至少一个GIC:3' 模块逆转录酶识别序列、任选的至少一个GIC:3' 模块rRNA序列、任选的至少一个GIC:3' 模块A-区序列或其任意组合。

[0048] 在一些实施方案中,至少一个GIC:3' 模块逆转录酶识别序列包含或编码至少一个与至少一种逆转录酶相互作用的序列。在一些实施方案中,至少一个GIC:3' 模块逆转录酶识别序列包含选自SEQ ID NO 200-224的序列。

[0049] 在一些实施方案中,至少一个GIC:3' 模块逆转录酶识别序列源自天然逆转录因子的3' 区。

[0050] 在一些实施方案中,任选的至少一个GIC:3' 模块rRNA序列包含或编码1至30nt的rRNA。

[0051] 在一些实施方案中,任选的至少一个GIC:3' 模块A-区序列包含或编码1至50个腺嘌呤碱基的序列。

[0052] 在一些实施方案中,至少一个GIC:3' 模块包含或编码SEQ ID NO 200-224中的至少一个或SEQ ID NO 300-329中的至少一个。在一些实施方案中,GIC:3' 模块包括来自选自以下种类的序列:OrLa、TriCasB、TaGu、GeFo、ZoAl、NaViB、DroSi、PuPu、LiPo、BoMo、GaAc、LeCo、CiIn、DrMe、DrNa、DrMer、TriCan、AdVa、HyMaA或其任意组合。

[0053] 在一些实施方案中,至少一个GIC:有效载荷模块包含或编码至少一个转基因ORF序列、任选的至少一个转基因启动子序列、任选的至少一个转基因5' 非翻译序列、任选的至少一个转基因3' 非翻译序列、任选的至少一个转基因多聚腺苷酸化信号序列、任选的至少一个转基因非编码RNA(ncRNA)、任选的至少一个ncRNA加工序列和/或其它供选择的3' 末端加工或稳定信号,或其任意组合。

[0054] 在一些实施方案中,至少一个转基因序列包含或编码至少一个用于插入受试者基因组的感兴趣序列。

[0055] 在一些实施方案中,至少一个转基因启动子序列包含或编码至少一个促进转基因在受试者基因组中的表达的序列。

[0056] 在一些实施方案中,至少一种GIC:有效载荷模块包含或编码至少一个转基因5' 非翻译序列,所述转基因5' 非翻译序列包含或编码至少一个转基因mRNA 5' 非翻译区。

[0057] 在一些实施方案中,至少一个转基因3' 非翻译序列包含或编码至少一个转基因mRNA 3' 非翻译区。

[0058] 在一些实施方案中,至少一个转基因多聚腺苷酸化信号序列包含或编码至少一种转基因多聚腺苷酸化信号。

[0059] 在一些实施方案中,至少一个转基因非编码RNA(ncRNA)加工序列和/或其它供选择的3'末端加工或稳定信号包含或编码用于至少一个转基因表达的ncRNA的至少一种终止信号、至少一种3'加工信号,及其任意组合。

[0060] 在一些实施方案中,至少一个GIC:有效载荷模块包含或编码与SEQ ID NO 411-422或SEQ ID NO 499-536中的至少一个具有至少90%同一性(例如,至少90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%同一性)的序列或其任意组合。

[0061] 在一些实施方案中,至少一个GIC:5'模块和至少一个GIC:3'模块中的至少一个包含或编码至少一个序列,所述序列源自与其它至少一个GIC:5'模块和至少一个GIC:3'模块中的至少一个不同的非长末端重复逆转录因子种类。

[0062] 在一些实施方案中,至少一个基因插入构建体包含或编码图中,例如图6-9中图示说明的至少一个结构及其任意组合。

[0063] 在一些实施方案中,所述系统包括:(i)至少一种逆转录酶构建体,其中所述至少一个逆转录酶构建体包含、编码至少一个与选自SEQ ID NO 1-57的序列具有至少90%同一性的序列,或由至少一个与选自SEQ ID NO 1-57的序列具有至少90%同一性的序列编码,和(ii)至少一种基因插入构建体,其中至少一种基因插入构建体包含至少一个与选自SEQ ID NO 60-154、250-276、278-279、280-289、300-329、400-404、405-407、411-422或499-536的序列具有至少90%的同一性(例如至少90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%同一性)的序列。在一些实施方案中,转染以产生RT蛋白的mRNA序列是从质粒分离出来的并编码蛋白氨基酸序列。

[0064] 在一些实施方案中,所述系统包括:

[0065] (i)至少一种逆转录酶构建体,其中所述至少一种逆转录酶构建体包含至少一个与选自SEQ ID NO 1-57的序列具有至少90%同一性(例如至少90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%同一性)的序列或由至少一个与选自SEQ ID NO 1-57的序列具有至少90%同一性(例如至少90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%同一性)的序列编码;以及

[0066] (ii)至少一种基因插入构建体,其中所述至少一种基因插入构建体包含:

[0067] GIC:5'模块,其包含与选自SEQ ID NO:60-154的序列具有至少90%同一性(例如,至少90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%同一性)的序列;

[0068] 包含选自SEQ ID NO:250-276的序列的rRNA序列,或相对于选自SEQ ID NO:250-276的序列有一个、两个或三个核苷酸变化的序列;或不包含rRNA序列;

[0069] 包含至少一个转基因序列的GIC:有效载荷模块;以及

[0070] GIC:3'模块,其包含与选自SEQ ID NO 300-329的序列具有至少90%同一性(例如,至少90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%同一性)的序列;

[0071] GIC:3'模块逆转录酶识别序列,其包含与选自SEQ ID NO 200-224的序列具有至少90%同一性(例如,至少90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%同一性)的序列;

[0072] 选自SEQ ID NO 280-289的GIC:3'模块rRNA序列,或其包含一个、两个或三个核苷

酸取代的序列;以及

[0073] 包含1至100个腺嘌呤碱基的GIC:3' 模块A-区序列。

[0074] 在一些实施方案中,RTC 5' 模块5' UTR包含与SEQ ID NO:58具有至少90%同一性(例如,至少90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%同一性)的序列。

[0075] 在一些实施方案中,RTC 3' 模块3' UTR包含与SEQ ID NO:59具有至少90%同一性(例如,至少90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%同一性)的序列。

[0076] 在一些实施方案中,所述系统包括基因插入构建体合成构建体(GIC:合成构建体),其包含或编码本文所述的至少一种基因插入构建体。

[0077] 在一些实施方案中,至少一种逆转录酶构建体和至少一种基因插入构建体中的至少一个包含或编码至少一种序列,所述至少一种序列源自与其它至少一种逆转录酶构建体和至少一种基因插入构建体中的至少一个不同种类的逆转录因子。

[0078] 在一些实施方案中,用于基因组编辑的系统包括至少一种(i)至少一种本文所述的逆转录酶构建体和(ii)至少一种本文所述的基因插入构建体的组合。

[0079] 还提供一种用于将至少一种转基因插入受试者基因组的方法,该方法包括向受试者给予有效量的至少一种本公开内容的基因插入系统(GIS)。

[0080] 在一些实施方案中,将转基因插入受试者基因组中的一个或多个靶位点,任选地其中所述一个或多个靶位点包括至少一个安全港位点。

[0081] 在一些实施方案中,任选的至少一个安全港位点包含至少一个核糖体DNA(rDNA)序列,任选地其中所述至少一个核糖体DNA序列包括至少一个28S rDNA序列。

[0082] 在一些实施方案中,至少一种方法包括给予至少一种用至少一种递送剂配制的基因插入系统。

[0083] 在一些实施方案中,至少一种递送剂是至少一种纳米颗粒,任选地,其中至少一种纳米颗粒包括至少一种脂质纳米颗粒。

[0084] 还提供一种药物组合物,该药物组合物包含至少一种权利要求所述的基因插入系统,任选的至少一种赋形剂、至少一种递送剂、至少一种佐剂中的至少一种及其任意组合。

[0085] 还提供一种治疗有其需要的受试者的治疗指征的方法,其包括向受试者给予有效量的至少一种本公开内容的基因插入系统或至少一种本公开内容的药物组合物。

[0086] 在一些实施方案中,治疗指征是由端粒酶活性丧失引起的。

[0087] 在一些实施方案中,至少一种基因插入系统包含至少一个TERT转基因。

[0088] 还提供一种用于制造本公开内容的基因插入系统的试剂盒。在一些实施方案中,试剂盒包含本公开内容的药物组合物。在一些实施方案中,试剂盒任选地进一步包含缓冲液、DNA质粒或用于制备所述基因插入系统或药物组合物的方案。

[0089] 还提供一种方法,其包括从头设计5' 模块,该5' 模块招募宿主机器用于产生第二链切口,并因此的第二链合成。在实施方案中,该方法通过从头设计5' 模块提供插入增益的效率,以(a)包括预定的rRNA长度和位置(本文描述的), (b)具有增强的RZ折叠,和/或(c)招募宿主细胞机器。

[0090] 另一方面,本公开内容提供一种用于将至少一个转基因插入细胞的基因组的方法

法,该方法包括使细胞与本公开内容的至少一种基因插入系统 (GIS) 接触。

[0091] 在一些实施方案中,转基因被插入到受试者基因组中的一个或多个靶位点,任选地其中所述一个或多个靶位点包括至少一个安全港位点。在一些实施方案中,任选的至少一个安全港位点包含至少一个核糖体DNA (rDNA) 序列,任选地其中至少一个核糖体DNA序列包括至少一个28SrDNA序列。

[0092] 在一些实施方案中,该方法包括给予至少一种用至少一种递送剂配制的基因插入系统。在一些实施方案中,至少一种递送剂是至少一种纳米颗粒,任选地其中所述至少一种纳米颗粒包括至少一种脂质纳米颗粒。

[0093] 在一些实施方案中,转基因以大于90%在靶的靶位点特异性(例如,大于90%、91%、92%、93%、94%、95%、96%、97%、98%或99%靶位点特异性)插入。

[0094] 在一些实施方案中,RTC包含编码来自白喉带鹀 (ZoA1)、栗耳草雀 (TaGu) 或白喉鵀 (TiGu) 的RT的RNA,或包含与SEQ ID NO:18、SEQ ID NO:20、SEQ ID NO:27、SEQ ID NO:29或SEQ ID NO:25具有至少90%同一性的氨基酸序列。

[0095] 在一些实施方案中,转基因在靶位点表达3个月或更长时间。

[0096] 在一些实施方案中,细胞与GIS接触,其中RTC与GIC的摩尔比为约10:1至1:20。

[0097] 在一些实施方案中,所述方法是体外方法、离体方法或体内方法。

[0098] 在一些实施方案中,细胞选自原代细胞、转化细胞、上皮细胞、成纤维细胞、人细胞、猴细胞和小鼠细胞。

[0099] 在一些实施方案中,细胞是异体细胞或自体细胞。在一些实施方案中,自体细胞是HLA匹配的细胞。

[0100] 本发明包括本文所记载的具体实施方案的所有组合,如同已经着力地记载了每种组合。

[0101] 附图简述

[0102] 图1是图示说明包含靶插入位点和天然逆转录因子的示例受试者基因组的图。展开视图 (底部) 图示说明R2天然逆转录因子的示例性组件结构。

[0103] 图2是图示说明示例逆转录酶构建体 (RTC) 的结构的图。

[0104] 图3是图示说明本发明的RT蛋白的示例性结构域的图。

[0105] 图4是描绘RT蛋白结构域的示例性来源生物体的图示,RT蛋白结构域包含DNA结合域 (DB)、RNA结合域 (RB)、逆转录酶 (RT) 结构域和核酸内切酶 (EN) 结构域。还图示说明描绘一小组RT蛋白结构域的示例组合的图。结构域特性由在其中发现野生型RT的生物体定义,A1是白喉带鹀 (*Zonotrichia albicollis*),A2是栗耳草雀 (*Taeniopygia guttata*),A3是白喉鵀 (*Tinamus guttatus*),A4是中嘴地雀 (*Geospiza fortis*),B1是九刺棘鱼 (*Pungitius pungitius*),B2是青鳉 (*Oryzias latipes*),B3是三刺鱼 (*Gasterosteus aculeatus*),C1是蝇蛹金小蜂 (*Nasonia vitripennis*),C2是黑腹果蝇 (*Drosophila melanogaster*),C3是赤拟谷盗 (*Tribolium castaneum*) (谱系B)、C4是家蚕 (*Bombyx mori*)、C5是拟果蝇 (*Drosophila simulans*)、C6是黄斑果蝇 (*Drosophila mercatorum*)、D1是库氏蝌蚪虾 (*Lepidurus couseii*)、D2是欧洲蝌蚪虾 (*Triops cancriformis*)、E1是大乳头水螅 (*Hydra magnipapillata*)、E2是美洲鲎 (*Limulus polyphemus*)、E3是污轮虫 (*Adineta vaga*) 和E4是玻璃海鞘 (*Ciona intestinalis*)。

[0106] 图5是一组图示说明本发明的一系列示例性RTC的图,其包含序列,所述序列包含或编码包含RT翻译起始密码子(M)的RT蛋白(RT)。RTC可包含5'非翻译序列(5'-UTR)、翻译终止密码子(SC)和/或3'非翻译序列(3'-UTR)。

[0107] 图6是图示说明示例基因插入构建体(中间)的结构图。展开视图显示示例5'模块(左下)、3'模块(右下)和有效载荷模块(上)的结构。

[0108] 图7是描绘GIC5'模块(5'M)组件、3'模块(3'M)组件和RTC RT模块(RT)组件的示例性来源生物体的图示。此外,还图示说明描绘一小组可能的示例GIC的图,所述GIC具有5'和3'模块的潜在组合,5'和3'模块在带有配对的逆转录酶构建体(配对的RT)的有效载荷模块的侧翼。模块特性由在其中发现野生型逆转录因子和/或逆转录酶的生物体定义,其中A1是白喉带鹀(*Zonotrichia albicollis*),A2是栗耳草雀(*Taeniopygia guttata*),A3是白喉鵂(*Tinamus guttatus*),A4是中嘴地雀(*Geospiza fortis*),B1是九刺棘鱼(*Pungitius pungitius*),B2是青鳉(*Oryzias latipes*),B3是三刺鱼(*Gasterosteus aculeatus*),C1是蝇蛹金小蜂(*Nasonia vitripennis*),C2是黑腹果蝇(*Drosophila melanogaster*),C3是赤拟谷盗(*Tribolium castaneum*) (谱系B)、C4是家蚕(*Bombyx mori*)、C5是拟果蝇(*Drosophila simulans*)、C6是黄斑果蝇(*Drosophila mercatorum*)、D1是库氏蝌蚪虾(*Lepidurus couseii*)、D2是欧洲蝌蚪虾(*Triops cancriformis*)、E1是大乳头水螅(*Hydra magnipapillata*)、E2是美洲螯(*Limulus polyphemus*)、E3是污轮虫(*Adineta vaga*)和E4是玻璃海鞘(*Ciona intestinalis*)。

[0109] 图8是图示说明由本发明的基因插入系统(GIS)插入转基因后示例受试者基因组的结构的图。

[0110] 图9是图示说明示例GIC合成构建体的结构的图。

[0111] 图10是通过变性PAGE凝胶解析的放射性DNA合成产物的图像。黑色实线框表示具有预期产物长度的凝胶区域。泳道号对应于实施例10的表3中详述的被测试的各种RT蛋白。泳道1的反应含有从不表达RT蛋白的细胞纯化的阴性对照。

[0112] 图11A是描绘用于测试RT蛋白结合模板RNA的特异性的示例实验设计的草图,所述模板RNA来自同源和非同源R2元件3'UTR。

[0113] 图11B显示测定家蚕、拟果蝇和青鳉RT对于同源和非同源模板3'UTR的选择性的斑点印迹结果。

[0114] 图12A和图12B显示TPRT反应产物的变性PAGE凝胶的结果。箭头指示正确的TPRT产物的预期大小。泳道B包含家蚕RT的反应产物,泳道D包含拟果蝇RT的反应产物,泳道O包含青鳉的反应产物,泳道N包含无酶的反应产物。图12A显示反应结果,所述反应含有所示RT蛋白与含有拟果蝇模板3'UTR的模板(单独标记的泳道)或与含有拟果蝇模板3'UTR和4nt rRNA的模板(用R4标记的泳道)的反应产物。图12B显示反应结果,所述反应含有所示RT蛋白与含有青鳉模板3'UTR(单独标记的泳道)或与含有青鳉模板3'UTR与4nt rRNA(用R4标记的泳道)的模板的反应产物。

[0115] 图13显示采用所示模板来自家蚕RT的TPRT反应产物的变性PAGE凝胶的结果。箭头指示正确的TPRT产物的预期大小,圆圈标记内部启动产生的产物的长度。

[0116] 图14A和图14B显示采用所示模板来自青鳉RT的TPRT反应产物的变性PAGE凝胶的结果。

[0117] 图15显示采用所示模板来自赤拟谷盗RT的TPRT反应产物的变性PAGE凝胶的结果。箭头指示预期的TPRT产物长度。

[0118] 图16显示来自源于白喉带鸮的RT蛋白的TPRT反应产物的变性PAGE凝胶的结果。实施例17中的表8给出用于每个所示泳道的GIC特性。TPRT产物的预期长度用实线框表示(顶部),沉淀回收对照的预期长度用具有虚线轮廓的框表示(中间),放射性标记的靶位点寡核苷酸的预期长度用点-点-虚线图案描画的框表示(底部)。

[0119] 图17显示来自源自栗耳草雀的RT蛋白的TPRT反应产物的变性PAGE凝胶的结果。泳道1含有长度参考梯,泳道2仅含有RT蛋白(无模板RNA),实施例19中的表11给出用于每个其它所示泳道的GIC特性。TPRT产物的预期长度用实线框表示(顶部),沉淀回收对照的预期长度用具有虚线轮廓的框表示(中间),放射性标记的靶位点寡核苷酸的预期长度用以点-点-虚线图案描画的框表示(底部)。

[0120] 图18A和图18B显示赤拟谷盗RT蛋白以指定的模板插入模板转基因后基因组DNA的PCR扩增产物。在图18A中,预期产物长度用框表示。所有正确插入PCR产物应大小相同。在图18B中,预期产物长度用箭头指示。没有5' 模块(3)相对于有5' 模块(5_3)的模板的正确插入PCR产物长度不同。

[0121] 图19显示基因组DNA的PCR扩增结果。上图对应预期的3' 连接的扩增,下图对应预期的5' 连接的扩增。标记“L”的泳道含有参考长度梯,标记1和9的泳道含有未转染源自TriCasB的RT表达质粒或GIC的PCR产物,标记2-8的泳道含有转染实施例21表13中所述的GIC而没有转染RT表达质粒后的PCR产物,而标记10-16的泳道含有转染实施例21表13中所述的GIC和RT表达质粒两者后的PCR产物。一些预期的PCR产物长度用星号标记。包括的所有星号参见SuppFIGS。

[0122] 图20显示基因组DNA的PCR扩增结果。标记A-J的泳道含有具有如所预期的大小的PCR产物,用于如实施例24表16中所示在共转染RTC mRNA和GIS RNA后检测预期的5' 连接。

[0123] 图21显示转基因GFP阴性克隆细胞群(顶部2个图)和转基因GFP阳性克隆细胞群(底部2个图)的示例性FACS分析结果。

[0124] 公开内容的详述

[0125] I. 引言

[0126] 除非禁用或另有说明,否则在这些描述和整个说明书中,术语“一(a)”和“一(an)”表示一或多,术语“或”表示和/或。应理解,本文所述的实施例和实施方案仅用于说明目的,将向本领域技术人员提示各种根据其的修改或改变,所述修改或改变应包括在本申请的精神和范围以及所附权利要求的范围内。本文引用的所有公开文献、专利和专利申请,包括其中的引文,在此通过引用整体并入用于所有目的。

[0127] 本发明提供用于基因组编辑和/或基因修饰的系统和方法,其包括将转基因插入受试者基因组中。所述系统在本文中被称作基因插入系统(GIS),可包括至少2种组件(即,2组件GIS), (a)至少一种逆转录酶(RT)构建体(RTC),其包含或编码至少一种逆转录酶,以及(b)至少一种分别表达的基因插入构建体(GIC),其包含或编码待用作逆转录模板的RNA构建体。本文所用的术语“构建体”可以指任何人工设计或合成的生物聚合物。所述生物聚合物可以例如由核酸(例如,DNA或RNA)、氨基酸或其任意组合组成。在一些实施方案中,(a)和(b)两者都是RNA构建体。在一些实施方案中,(a)是氨基酸构建体(即,蛋白),而(b)是RNA构

建体。

[0128] 还提供能够进行靶点引发的逆转录 (TPRT) 的工程化的RTC。本文所用的术语“靶点引发的逆转录”是指其中逆转录酶使用靶位点处可用的DNA 3' 端作为引物引发cDNA合成的任何过程。

[0129] 此外,所提供的系统和方法可以实现将转基因插入受试者DNA中的序列特异性位置(本文称为靶位点),诸如安全港位点。本文所用的术语“安全港”和“安全港位点”是指受试者基因组中的任何位点,其中受试者DNA序列的破坏,例如通过插入异源序列对受试者DNA序列的破坏不负面影响受试者细胞的功能。本文所用的示例性安全港位点位于编码核糖体RNA(rRNA)的受试者基因组部分内,包括由RNA聚合酶I转录的rRNA前体,其由本文称为核糖体DNA(rDNA)基因座的序列编码,包含编码5.8S、18S或28S rRNA的序列。

[0130] 本公开内容表明,仅递送RNA可对DNA转基因插入细胞,例如人细胞的基因组的安全港位置进行编程。在一些实施方案中,将编码待插入的转基因的RNA模板和编码逆转录酶的信使RNA两者递送至细胞,所述编码逆转录酶的信使RNA是将RNA模板转化为基因组DNA所必需的。预期通过利用无毒、高效、细胞类型靶向的RNA递送机制的持续创新,仅递送RNA将更容易转化为人的基因疗法。

[0131] 在一些方面,逆转录酶(RT)的基于质粒的表达与转染的RNA模板相结合。在一些实施方案中,包含R2逆转录因子序列的天然或自然部分的转基因模板5' 模块与RT以异源组合来使用,这提供全长位点特异性序列插入而非截短的逆转录因子序列插入的优点。在一些实施方案中,模板RNA包含3' 模块,其中逆转录因子3' UTR序列来自与RT相同的种类。在一些实施方案中,3' UTR还包含3' 聚-A区,其提高靶位点特异性插入效率。

[0132] 与先前的系统和方法相比,本公开内容提供以下改进和优点。本发明人证明:

[0133] (i) 来自鸟类的RT蛋白具有显著的转基因插入活性,使得超过20%的转染细胞具有功能性表达的转基因。鸟类RT对复制包含鸟类3' UTR随后是3' 聚-A区的模板RNA具有超选择性;

[0134] (ii) 鸟类R2逆转录因子3' UTR和RT蛋白的异源组合可以比天然组合更有效;

[0135] (iii) 非天然的、从头创建且优化的5' 模块可以更有效,实现位点特异性插入效率提高一个或多个数量级。

[0136] (iv) 来自赤拟谷盗的天然5' 模块(TriCasA) (TCA、TCA5、TCARZ等) 来自与鸟类RT蛋白完全不同的进化枝的R2逆转录因子,可以更有效;

[0137] (v) 使用共转染的2-RNA系统,而不是质粒表达RT,然后转染模板RNA进行转基因插入递送;

[0138] (vi) 2-RNA转染可以每个细胞插入多种转基因,实现在单次RNA给药中的基因递送的多重化。这实现将多种治疗性转基因插入同一细胞的基因组中,包括编码治疗性蛋白或治疗性蛋白的不同亚基的转基因,或治疗性蛋白和RNA的组合;

[0139] (vii) 2-RNA递送实现在广泛的细胞类型中的转基因表达,包括原代细胞系和非分裂或缓慢分裂的细胞,包括小鼠和猴以及人细胞;

[0140] (viii) 基因组测序证实插入的位点特异性;和

[0141] (ix) 插入的转基因表达盒具有数月的表达稳定性。

[0142] 产生逆转录因子的组件

[0143] 本发明的RTC和/或GIC可包含组件(可互换地称为模块),所述组件可以源自至少一种非长末端重复逆转录因子(非LTR)的部分和/或是自然界中未知的。不希望受理论的束缚,图1图示说明(顶部)包含天然逆转录因子100的受试者基因组,在这种情况下逆转录因子是非长末端重复逆转录因子(非LTR)。从图中可以看出,受试者DNA 110可以包含至少一个靶向插入位点120,并且在靶向插入位点处可以存在天然逆转录因子130。可以在展开视图(底部)中进一步审视示例天然逆转录因子的架构。这里,逆转录因子5'区131在翻译起始位点132之前。逆转录因子5'区通常不翻译成氨基酸生物聚合物,并且可以包含被逆转录因子RT识别的核酸序列和/或在后来插入期间影响天然逆转录因子的第二链合成。翻译起始位点132是将被翻译成氨基酸的第一核苷酸。逆转录因子逆转录酶开放阅读框133编码逆转录酶,所述逆转录酶可以识别、结合并使用逆转录因子RNA转录物作为用于逆转录的模板。逆转录因子逆转录酶开放阅读框延伸至翻译终止位点134但不包含翻译终止位点134。逆转录因子3'区135通常不翻译成氨基酸生物聚合物,并且可以包含由天然逆转录因子RT识别的核酸序列。区域131和135可以存在或可以不存在,并且如果存在,则可以包含重复周围靶位点序列和/或不由逆转录因子RNA模板编码的序列。

[0144] 可从其获得GIS组件的合适逆转录因子包括但不限于非LTR逆转录因子,例如RLE型或APE型或Penelope型非LTR逆转录因子。RLE型非LTR逆转录转座子可以来自许多进化枝中的任一个,包括但不限于R2、R4、CRE、Genie、HERO、NeSL。APE型非LTR逆转录转座子可以来自许多进化枝中的任一个,包括但不限于I、R1、L1、Tx1、CR1、Rex1、Jockey、L2、Tad、RTE、RTEX、Ingi、Vingi、TRAS、SART或其任意组合。在一些实施方案中,GIS组件可以源自插入rDNA的逆转录因子,即所谓的R元件,诸如R1或R2进化枝的逆转录因子。在一些实施方案中,R2进化枝逆转录因子可以具有典型的R2逆转录因子插入位点特异性,或者可以源自较大R2进化枝中的R8和/或R9逆转录因子,其相对于典型的R2逆转录因子具有改变的靶序列,或者可以源自看起来已失去靶位点特异性的R2NS逆转录因子。

[0145] GIS组件可以源自在任何物种中发现的逆转录因子的部分或结构域,包括与受试者具有远缘进化关系的那些。例如,可以由其获得GIS组件的合适逆转录因子可包括在鸟类(白喉带鹀(*Zonotrichia albicollis*)、栗耳草雀(*Taeniopygia guttata*)、白喉鵪(*Tinamus guttatus*)和中嘴地雀(*Geospiza fortis*))、鱼类(例如,九刺棘鱼(*Pungitius pungitius*)、青鳉(*Oryzias latipes*)、斑马鱼(*Danio rerio*)、黑点青鳉(*Oryzias melastigma*)、海七鳃鳗(*Petromyzon marinus*)、褐鳟(*Salmo trutta*)、大西洋鲑(*Salmo salar*)或三刺鱼(*Gasterosteus aculeatus*))、昆虫(黄斑果蝇(*Drosophila mercatorum*)、黑腹果蝇(*Drosophila melanogaster*)、蝇蛹金小蜂(*Nasonia vitripennis*)、赤拟谷盗(*Tribolium castaneum*)、拟果蝇(*Drosophila simulans*)、东方蜜蜂(*Apis cerana*)和家蚕(*Bombyx mori*))、甲壳动物(例如库氏蝌蚪虾(*Lepidurus couesii*)和欧洲蝌蚪虾(*Triops cancriformis*))、其它无脊椎动物(例如美洲鲎(*Limulus polyphemus*)、大乳头水螅(*Hydra magnipapillata*)或污轮虫(*Adineta vaga*))、脊索动物(例如,玻璃海鞘(*Ciona intestinalis*))、包括哺乳动物,及其任意组合中发现的那些。

[0146] 在一些实施方案中,GIS组件可源自本文公开的任何序列的部分或结构域。

[0147] II. 基因插入系统(GIS)组合物

[0148] 本发明用于将遗传物质(例如,转基因)插入受试者基因组的系统在整个本公开内

容中被称为基因插入系统 (GIS)。GIS可由多种生物聚合物构建体组成,所述生物聚合物构建体被共同给予以通过靶点引发逆转录 (TPRT) 插入至少一种转基因。这些生物聚合物构建体可以是氨基酸生物聚合物、核酸生物聚合物、含有氨基酸和核酸两者的混合生物聚合物,或其任意组合。在一些实施方案中,GIS由至少2种生物聚合物构建体,至少一种逆转录酶构建体 (RTC) 和至少一种基因插入构建体 (GIC) 组成。在这样的实施例中,RTC包含用于诸如通过包含或编码逆转录酶进行逆转录的工具,并且GIC包含或编码至少一种RNA序列,所述RNA序列可被RTC用作cDNA合成的模板。

[0149] 本发明的生物聚合物构建体本身由多个模块组成,使得可以根据需要组合所述模块以改变系统,用于期望的功能。本文所用的术语“模块”是指构建体的一部分,其由其功能(例如,蛋白的功能域)或其序列(例如,氨基酸或核酸序列)定义。

[0150] 逆转录酶构建体 (RTC)

[0151] 本发明的GIS包含至少一种RTC,所述RTC包含或编码活性RT蛋白,诸如源自非LTR逆转录因子的RT。本文所用的术语“RTC”是指生物聚合物构建体,其包含或编码至少一种逆转录酶 (RT)。在一些实施方案中,用于本发明的GIS的至少一种RTC可包含氨基酸生物聚合物,包括但不限于多肽、蛋白、前蛋白或其任意组合。在一些实施方案中,用于本发明的GIS的至少一种RTC可包含核酸生物聚合物,包括但不限于RNA、DNA或其任意组合。在一些实施方案中,至少一种RTC可包含至少一种mRNA构建体。

[0152] RTC架构

[0153] 本发明的RTC可包含至少一种RTC:逆转录酶模块 (RTC:RT-模块)、至少一种任选的逆转录酶构建体5' 模块 (RTC:5' 模块)、至少一种任选的逆转录酶构建体3' 模块 (RTC:3' 模块) 及其任意组合。在RTC的一些实施例中,RTC:5' 模块和RTC:3' 模块可以是任选的,并且一种或两种可以不存在。在一些实施方案中,至少一种RTC可包含线性RNA生物聚合物,或者作为线性RNA生物聚合物递送给受试者。在一些实施方案中,至少一种RTC可包含mRNA生物聚合物,或者作为mRNA生物聚合物递送给受试者。

[0154] 现在转向图2,提供示例性线性RNA生物聚合物(例如,mRNA) RTC 200的架构。如图示说明的,对于mRNA生物聚合物RTC,RTC:5' 模块210是RTC的任选组件,当存在时,其可包含改变RTC的免疫原性和/或控制RTC:RT模块220的表达的序列。例如,RTC:5' 模块可包含或编码至少一个5' 帽(例如TriLink Clean Cap AG,m7(3'OMeG)(5') ppp(5')(2'OMeA)pG)、至少一个5' 非翻译区(5'-UTR)、至少一个Kozak序列、至少一个启动子及其任意组合。起始密码子,已知的启动翻译的核酸的3-核苷酸序列,标记RTC:RT模块的5' 端。RTC:RT-模块(详述如下)包含起始密码子,并从起始密码子延伸到终止密码子,并不包含终止密码子。任选的RTC:3' 模块230,当存在时,包含终止密码子并从终止密码子延伸到RTC 3' 端。RTC:3' 模块,当存在时,可包含改变RTC的免疫原性和/或控制RTC:RT-模块的表达的序列。例如,RTC:3' 模块可包含或编码翻译终止密码子、3' UTR、(多个)多聚腺苷酸序列、多聚腺苷酸化信号或其任意组合。

[0155] 在一些实施方案中,至少一种RTC可包含质粒,或者作为质粒递送给受试者。在一些实施方案中,至少一种RTC可包含mRNA或前mRNA,或者作为mRNA或前mRNA递送给受试者。在一些实施方案中,至少一种RTC可包含蛋白,或者作为蛋白递送给受试者。在一些实施方案中,至少一种RTC可包含前蛋白,或者作为前蛋白递送至受试者。

[0156] RTC:RT模块

[0157] RTC的RT模块包含或编码至少一种具有逆转录活性的化合物或组合物,其具体但非限制性的实例是一类被称为逆转录酶(RT)的酶蛋白。在一些实施方案中,RT模块可包含或编码源自逆转录因子基因中发现的至少一种RT(即逆转录因子RT)的生物聚合物。在一些实施方案中,RTC:RT模块包含或编码至少一种源自非长末端重复逆转录因子的逆转录酶。

[0158] 逆转录酶

[0159] 本文所用的术语“逆转录酶(RT)”以其最广泛的含义使用,以指代任何具有逆转录活性的生物聚合物。在一些实施方案中,用于本发明的RT可以是或源自白喉带鹀(*Zonotrichia albicollis*)、栗耳草雀(*Taeniopygia guttata*)、白喉鵪(*Tinamus guttatus*)、中嘴地雀(*Geospiza fortis*)、九刺棘鱼(*Pungitius pungitius*)、青鳉(*Oryzias latipes*)、斑马鱼(*Danio rerio*)、黑点青鳉(*Oryzias melastigma*)、海七鳃鳗(*Petromyzon marinus*)、褐鳟(*Salmo trutta*)、大西洋鲑(*Salmo salar*)或三刺鱼(*Gasterosteus aculeatus*)、黄斑果蝇(*Drosophila mercatorum*)、黑腹果蝇(*Drosophila melanogaster*)、蝇蛹金小蜂(*Nasonia vitripennis*)、赤拟谷盗(*Tribolium castaneum*)、拟果蝇(*Drosophila simulans*)、东方蜜蜂(*Apis cerana*)、家蚕(*Bombyx mori*)、库氏蝌蚪虾(*Lepidurus couesii*)、欧洲蝌蚪虾(*Triops cancriformis*)、美洲鲎(*Limulus polyphemus*)、大乳头水螅(*Hydra magnipapillata*)、污轮虫(*Adineta vaga*)、玻璃海鞘(*Ciona intestinalis*)、其它鸟类、其它节肢动物、其它鱼类、其它被囊动物、其它动物(包括哺乳动物和人)或类似的基因组的非LTR RT。

[0160] 在一些实施方案中,用于本公开内容的GIS的至少一个RTC:RT模块可包含、编码SEQ ID NO 1-57中的至少一个或由SEQ ID NO 1-57中的至少一个编码。在一些实施方案中,至少一个RTC:RT模块可包含、编码与SEQ ID NO 1-57中的至少一个具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列或由与SEQ ID NO 1-57中的至少一个具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列编码。在一些实施方案中,RTC:RT-模块包含与SEQ ID NO 1-57具有至少90%、91%、92%、93%、94%、95%、96%、97%、98%或99%同一性的非天然或非自然序列。

[0161] 在一些实施方案中,至少一种RTC:RT-模块可包含、编码与SEQ ID NO 17-21(ZoA1 RT序列)中的至少一个具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列或由与SEQ ID NO 17-21(ZoA1 RT序列)中的至少一个具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列编码。

[0162] 在一些实施方案中,至少一种RTC:RT模块可包含、编码与SEQ ID NO 26-29(TaGu RT序列)中的至少一个具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列或由与SEQ ID NO 26-29(TaGu RT序列)中的至少一个具有至少100%、99%、98%、

97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列编码。

[0163] 在一些实施方案中,至少一种RTC:RT模块可包含、编码与SEQ ID NO 1-5(TriCasB RT序列)中的至少一个具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列或由与SEQ ID NO 1-5(TriCasB RT序列)中的至少一个具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列编码。

[0164] 在一些实施方案中,RTC:RT模块可包含或编码通过合适的TPRT测定显示对TPRT有活性的蛋白。合适的TPRT测定的非限制性实例包括(i)用编码RT蛋白的表达质粒转染细胞群,所述表达质粒具有合适的用于亲和纯化的标签(例如,FLAG标签),(ii)裂解细胞群并通过本领域已知的适当方法收集和纯化表达的蛋白产物,(iii)通过本领域已知的任何方法(例如,T7 RNA聚合酶)制备重组模板RNA,(iv)将纯化的RT蛋白、重组模板和包含靶位点寡核苷酸双链DNA的核苷酸溶液与末端放射性标记的下游链在促进通过RT的逆转录的培养基中结合,以及(v)通过本领域已知的任何合适的方法(例如,变性PAGE)收集和分析产物。

[0165] 适合用于本发明的RT可以由多个功能域组成。在一些实施方案中,诸如图3中所图示说明的,至少一种逆转录酶300包含至少一个DNA结合域310、至少一个RNA结合域320、至少一个cDNA合成域330、至少一个核酸内切酶结构域340及其任意组合。注意,对于此图示说明,仅呈现一种可能的结构域配置。在一些实施方案中,任何所描述的结构域可以以不同的频率存在于RT中和/或所述结构域可以以任何顺序存在。在一些实施方案中,DNA和RNA结合域可以来自不同于RT的多肽类型或真核生物基因组中未知的序列(例如,从头设计的DNA或RNA结合域)。

[0166] 起始密码子

[0167] 可通过本领域已知的各种方法将至少一个非天然翻译起始密码子添加到编码RT的核酸序列中。可将非天然翻译起始密码子添加到源自非LTR逆转录因子的序列的任何位置,这产生功能性RT。例如,可在距野生型非LTR逆转录因子中的已知参考点(例如,距天然逆转录因子RT ORF中的氨基酸序列基序)约1、10、50、100、150、200、250、300、350、400、450、500、550、600、650、700、750、800、850、900、950、1000或更多个碱基处添加至少一个非天然起始密码子。翻译起始密码子的定位可以根据多肽长度、序列组成、活性、生物稳定性、缺少聚集或定位的优化来选择,和/或除了对于在感兴趣的靶细胞中设计最佳的或受调节的蛋白表达的领域中的从业者来说显而易见的其它考虑之外,为赋予编码蛋白的mRNA改进的生物稳定性而选择。

[0168] 翻译起始密码子可以是已知的通过核糖体启动翻译的任何3个核苷酸,其依赖于或独立于mRNA中的另一个序列或结构。在一些实施方案中,非天然翻译起始密码子是AUG。

[0169] RTC:5' 模块

[0170] 本发明的RTC可包含至少一个RTC:5' 模块。一般而言,RTC:5' 模块包含未翻译的生物聚合物组件,作为非限制性实例,其可以改变GIC的免疫原性、帮助将GIC定位到被靶向的细胞内区域、控制或改变GIC的RTC:RT模块的表达、标记GIC以进行识别、帮助纯化GIC、控制GIC的降解、实现外源或内源调节GIC活性和/或功能,及其任意组合。

[0171] 在一些实施方案中,至少一个RTC:5' 模块可包含或编码至少一个5' UTR。在一些实施方案中,至少一个RTC:5' 模块可包含或编码至少一个5' 帽。在一些实施方案中,至少一个RTC:5' 模块可包含或编码至少一个微小RNA结合序列。在一些实施方案中,至少一个RTC:5' 模块可包含或编码至少一个RNA聚合酶启动子。

[0172] 在一些实施方案中,至少一个用于本公开内容的GIS的RTC:5' 模块包含SEQ ID NO 58的5' UTR。

[0173] 在实施方案中,我们对转染的mRNA使用了一个5' 和一个3' UTR,所述mRNA取自向WHO报告的BioNTech疫苗序列。我们还使用了它们的模板编码的聚A区域(而不是使用转录后的聚A聚合酶),该区域由A30-10 nt接头-A70组成,随后是Type IIS限制位点,以剪切用于mRNA转录的模板,而无任何另外的3' nt。所有mRNA均用TriLink AG clean capm7 (3'OMeG) (5') ppp (5') (2'OMeA) pG) 加帽。选择UTR进行组织特异性RT表达,例如用以施加细胞类型特异性的翻译控制。

[0174] 在一些实施方案中,RTC:5' 模块可包含、编码与SEQ ID NO 58具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列或由与SEQ ID NO 58具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列编码。

[0175] RTC:3' 模块

[0176] 本发明的RTC可包含至少一个RTC:3' 模块。一般而言,RTC:3' 模块包含未翻译的生物聚合物组件,作为非限制性实例,其可以改变GIC的免疫原性、帮助将GIC定位到被靶向的细胞内区域、控制或改变GIC的RTC:RT模块的表达、标记GIC用于识别、帮助纯化GIC、控制GIC的降解、实现外源或内源调节GIC活性和/或功能,及其任意组合。

[0177] 在一些实施方案中,至少一个RTC:3' 模块可包含至少一个3' UTR。在一些实施方案中,至少一个RTC:3' 模块可包含或编码至少一个聚-A区或聚-A尾。在一些实施方案中,至少一个RTC:3' 模块可包含或编码至少一个微小RNA结合序列。

[0178] 在一些实施方案中,用于本公开内容的GIS的至少一个RTC:3' 模块包含SEQ ID NO 59的3' UTR和聚-A尾。

[0179] 在一些实施方案中,RTC:3' 模块包含与SEQ ID NO 59具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的3' UTR。

[0180] RTC的模块化

[0181] 本发明的RTC可通过组合至少一个RTC:RT模块、任选的至少一个RTC:5' 模块和/或任选的至少一种RTC:3' 模块的任意组合而被设计用于期望的功能或活性。在一些实施方案中,RTC包含至少一个RTC:5' 模块。在一些实施方案中,RTC包含至少一个RTC:3' 模块。在一些实施方案中,RTC包含至少一个RTC:RT模块。在一些实施方案中,RTC包含至少一个RTC:5' 模块、至少一个RTC:RT模块和至少一个RTC:3' 模块。在一些实施方案中,RTC包含至少一个RTC:5' 模块和至少一个RTC:RT模块。在一些实施方案中,RTC包含至少一个RTC:RT模块和至少一个RTC:3' 模块。

[0182] 在一些实施方案中,本发明的RTC可不包含至少一个RTC:5' 模块和至少一个RTC:

3' 模块。在一些实施方案中,本发明的RTC可不包含至少一个RTC:5' 模块或至少一个RTC:3' 模块。在一些实施方案中,本发明的RTC可不包含至少一个RTC:5' 模块。在一些实施方案中,本发明的RTC可不包含至少一个RTC:3' 模块。

[0183] 在一些实施方案中,至少一个RTC可包含以下的任意组合:(a)选自、编码SEQ ID NO 58中的任一个或由SEQ ID NO 58中的任一个编码的至少一个RTC:5' 模块,(b)选自、编码SEQ ID NO 1-57中的任一个或由SEQ ID NO 1-57中的任一个编码的至少一个RTC:RT-模块,和/或(c)选自、编码SEQ ID NO 59中的任一个或由SEQ ID NO 59中的任一个编码的至少一个RTC:3' 模块。

[0184] 示例性RTC

[0185] 用于本发明的RTC可包含、编码SEQ ID NO 1-57中的至少一个或由SEQ ID NO 1-57中的至少一个编码。在一些实施方案中,RTC可包含、编码与SEQ ID NO 1-57中的至少一个具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列或由与SEQ ID NO 1-57中的至少一个具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列编码。

[0186] 在一些实施方案中,至少一种RTC可包含、编码与SEQ ID NO 17-21中的至少一个具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列或由与SEQ ID NO 17-21中的至少一个具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列编码。

[0187] 在一些实施方案中,至少一种RTC可包含、编码与SEQ ID NO 26-29中的至少一个具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列或由与SEQ ID NO 26-29中的至少一个具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列编码。

[0188] 在一些实施方案中,至少一种RTC可包含、编码与SEQ ID NO 24-25中的至少一个具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列或由与SEQ ID NO 24-25中的至少一个具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列编码。

[0189] 在一些实施方案中,至少一种RTC可包含、编码与SEQ ID NO 1-5中的至少一个具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列或由与SEQ ID NO 1-5中的至少一个具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列编码。

[0190] 在一些实施方案中,至少一种RTC可包含、编码与SEQ ID NO 35-37中的至少一个具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列或由与SEQ ID NO 35-37中的至少一个具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列编码。

[0191] 在一些实施方案中,至少一种RTC可包含、编码与SEQ ID NO 32-34中的至少一个具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列或由与SEQ ID NO 32-34中的至少一个具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列编码。

[0192] 在一些实施方案中,至少一种RTC包含图5中图示说明的结构。

[0193] RTC调节元件

[0194] 本发明的RTC还可包含任意数量的调节元件,其可位于RTC模块中的任何RTC模块内。本文所用的术语“调节元件”是指实现对其所属的生物聚合物的表达或活性的控制的任何序列、区域或结构域。

[0195] 例如,基于RNA的RTC可包含任意数量的微小RNA(miRNA)或小干扰RNA(siRNA)结合位点。不希望受理论的束缚,基于存在的RNAi转录组,这些RNA干扰(RNAi)结合位点的存在可阻止RT蛋白在特定细胞类型中的表达。这样,本发明的GIS可以从受试者细胞类型去靶向。本文所用的术语“miRNA或siRNA结合位点”是指分别与至少一个miRNA或siRNA互补的RNA序列。

[0196] 在一些实施方案中,RTC可包含至少一个miRNA和/或siRNA结合位点,所述结合位点与GIS插入的转基因中包含的或由其编码的至少一种miRNA和/或siRNA互补。一般而言,这可使本发明的GIS能够自调节单次给予GIS所进行的转基因插入的次数和/或防止在初次给药后重复插入转基因。这样,GIS可具有增加的对给定受试者重新给药或共同给药的能力。

[0197] 基因插入构建体(GIC)

[0198] 本发明的GIS包含至少一种GIC,其一般包含或编码至少一个旨在插入受试者基因组中的感兴趣序列(即“有效载荷序列”)。本文所用的术语“GIC”是指任何生物聚合物构建体,其包含或编码至少一个RNA序列,使得RNA序列被至少一种RTC:RT模块包含或编码的至少一个RT识别,并可用作逆转录的模板。在一些实施方案中,用于本发明的GIS的至少一种GIC可包含核酸生物聚合物,包括但不限于RNA、DNA或其任意组合。

[0199] GIC架构

[0200] 本发明的基因插入构建体(GIC)可包含或编码至少一个GIC:5'模块、至少一个GIC:有效载荷模块、至少一个GIC:3'模块及其任意组合。在一些实施方案中,至少一种GIC可包含质粒或以质粒形式递送给受试者。在一些实施方案中,至少一种GIC可包含线性RNA或以线性RNA形式递送给受试者。

[0201] 在一些实施方案中,至少一个GIC:5'模块是任选的。在一些实施方案中,至少一个

GIC:3' 模块可以是任选的。在一些实施方案中,本发明的GIC可包含或编码至少一个GIC:有效载荷模块,并且不包含或编码至少一个GIC:5' 模块和/或至少一个GIC:3' 模块。

[0202] 如图6中所示,其描绘示例性线性RNA GIC 400,任选的GIC:5' 模块410从5' GIC序列末端延伸到GIC:5' 模块末端420。GIC:有效载荷模块430朝向GIC:5' 模块(当存在时)的3',并延伸到GIC:有效载荷模块末端440。最后,GIC:3' 模块450延伸到3' GIC末端。这些特征中的每一个都将在下文详细讨论。

[0203] GIC:5' 模块

[0204] 用于本公开内容的GIC的GIC:5' 模块可包含或编码至少一个源自天然逆转录因子5' 区的序列。不希望受理论的束缚,5' 模块可包含或编码RNA序列,所述RNA序列与RT的至少一个RNA结合域相互作用,在转基因插入期间影响第二链合成,降低GIC的免疫原性,提供对GIC稳定性和/或纯化有用的特征,及其任意组合。

[0205] GIC:5' 模块架构

[0206] 在实施方案中,5' 模块包含或含有5' rRNA序列和核糖酶(RZ)序列。在一些实施方案中,5' rRNA序列和RZ序列不一定完全分开。在一些实施方案中,5' 模块包含“折叠序列”,其可以与RZ序列分开。在一些实施方案中,GIC:5' 模块可任选地包含或编码至少一个GIC:5' 模块rRNA序列(或其它靶位点序列)、任选的至少一个GIC:5' 模块核糖酶(RZ)序列、任选的至少一个GIC:5' 模块折叠序列及其任意组合。

[0207] 回到图6,GIC:5' 模块410的展开视图(左下)图示说明一个示例性GIC:5' 模块的架构。GIC:5' rRNA序列411,当存在于5' 模块的5' 端时,可包含或编码与位于靶插入位点的5' 或以其它方式位于靶插入位点附近的受试者DNA序列互补的RNA序列。GIC:5' 模块核糖酶(RZ)序列412,当存在时,可包含至少一个具有自剪切RZ折叠的RNA序列,其可以或可以不自剪切以从转录的5' 前导序列释放功能性GIC。GIC:5' 模块RZ序列将折叠,并且在有活性时将剪切,使得GIC:5' rRNA序列作为RZ的一部分被包含在GIC的5' 端处或其附近。任选的GIC:5' 模块折叠基序序列413可包含至少一个具有预测的或证实的自主折叠的RNA序列,这可用于在物理和/或动力学上将GIC:5' 模块RZ的折叠与有效载荷序列的折叠分开。另外,在区域414内或位置420处,位于GIC:5' 模块410和有效载荷模块430之间,可添加GIC序列以终止或以其它方式调节从靶位点侧翼的(多个)内源性细胞启动子序列启动的转录。在一些实施方案中,(多个)靶位点侧翼的内源性细胞启动子序列可用于有效载荷表达,这是可在位置420和/或440添加(多个)GIC序列以调节有效载荷表达(例如,以启动翻译或终止含有有效载荷序列的宿主启动子RNA转录物的转录)的情况中的一个实例。另外,区域414可含有RNA聚合酶(RNAP)终止序列以防止RNA聚合酶从靶插入位点处的基因读穿(readthrough)。在一些实施方案中,RNAP是RNAP I(Pol I),并且当GIC有效载荷模块被整合到核糖体DNA基因靶位点中时,终止序列阻止Pol I读穿转录。在一些实施方案中,RNAP终止子序列包含序列5'-AGGTCGACCAGATGTCCGAGGTCGACCAGTTGTCCG-3'(SEQ ID NO:537)。

[0208] GIC:5' 模块rRNA序列

[0209] 至少一个GIC:5' 模块rRNA序列是GIC:5' 模块的任选组件。当存在时,其可包含或编码人核糖体RNA(rRNA)序列或与位于靶插入位点的5' 处的至少一个受试者DNA序列同源和/或互补的其它序列。不希望受理论的束缚,该rRNA序列可通过招募至少一种内源性DNA修复机制指导插入的cDNA转基因的第二链合成。在一些实施方案中,GIC:5' 模块rRNA序列

位于GIC:5' 模块RZ序列的5' 处。在一些实施方案中,GIC:5' 模块不包含包含rRNA基因组序列的序列。

[0210] 在一些实施方案中,至少一个GIC:5' 模块rRNA序列可包含或编码约1至36nt的rRNA。在一些实施方案中,至少一个GIC:5' 模块rRNA序列可包含或编码约1至30nt的rRNA。在一些实施方案中,至少一个GIC:5' 模块rRNA序列可包含或编码约1至28nt的rRNA。在一些实施方案中,至少一个GIC:5' 模块rRNA序列可包含或编码约1至26nt的rRNA。在一些实施方案中,至少一个GIC:5' 模块rRNA序列可包含或编码约1至13nt的rRNA。在一些实施方案中,至少一个GIC:5' 模块rRNA序列可包含或编码约1至11nt的rRNA。

[0211] 在一些实施方案中,至少一个GIC:5' 模块rRNA序列可包含或编码约1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35或36nt的rRNA。

[0212] 在一些实施方案中,至少一个GIC:5' 模块rRNA序列可包含或编码约30nt的rRNA。在一些实施方案中,至少一个GIC:5' 模块rRNA序列可包含或编码约36nt的rRNA。在一些实施方案中,至少一个GIC:5' 模块rRNA序列可包含或编码约28nt的rRNA。在一些实施方案中,至少一个GIC:5' 模块rRNA序列可包含或编码约26nt的rRNA。在一些实施方案中,至少一个GIC:5' 模块rRNA序列可包含或编码约13nt的rRNA。在一些实施方案中,至少一个GIC:5' 模块rRNA序列可包含或编码约11nt的rRNA。在一些实施方案中,GIC:5' 模块rRNA序列包含5' G核苷酸。

[0213] 在一些实施方案中,至少一个GIC:5' 模块rRNA序列可包含、编码SEQ ID NO 250-276中的至少一个或由SEQ ID NO 250-276中的至少一个编码。在一些实施方案中,至少一个GIC:5' 模块rRNA序列可包含、编码与SEQ ID NO 250-276中的至少一个具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%同源性的序列或由与SEQ ID NO 250-276中的至少一个具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%同源性的序列编码。在一些实施方案中,至少一个GIC:5' 模块rRNA序列包含相对于选自SEQ ID NO:250-276的序列具有一个、两个或三个核苷酸变化或取代的序列。

[0214] 在一些实施方案中,至少一个GIC:5' 模块rRNA序列可包含、编码与SEQ ID NO 252中的至少一个具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%同源性的序列或由与SEQ ID NO 252中的至少一个具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%同源性的序列编码。在一些实施方案中,至少一个GIC:5' 模块rRNA序列包含相对于SEQ ID NO 252具有一个、两个或三个核苷酸变化的序列。

[0215] 在一些实施方案中,至少一个GIC:5' 模块rRNA序列可包含、编码与SEQ ID NO 254中的至少一个具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%同源性的序列或由与SEQ ID NO 254中的至少一个具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%同源性的序列编码。在一些实施方案中,至少一个GIC:5' 模块rRNA序列包含相对于SEQ ID NO 254具有一个、两个或三个核苷酸变化的序列。

[0216] 在一些实施方案中,至少一个GIC:5' 模块rRNA序列可包含、编码与SEQ ID NO 255中的至少一个具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%同源性的序列或由与SEQ ID NO 255中的至少一个具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%同源性的序列编码。在一些实施方案中,至少一个GIC:5'

模块rRNA序列包含相对于SEQ ID NO 255具有一个、两个或三个核苷酸变化的序列。

[0217] GIC:5' 模块RZ序列

[0218] GIC:5' 模块RZ序列是GIC:5' 模块的任选组件,当存在时,其包含或编码至少一个自剪切核糖酶或具有自剪切核糖酶的折叠的序列(统一描述为RZ)。不希望受理论的束缚,该基序可将GIC的5' OH末端,诸如自剪切产生的5' 末端,埋入稳定的三级结构中,这可降低对外源RNA的先天免疫反应,减少依赖于5' 单磷酸酯以启动剪切的5' -3' 核酸外切酶使GIC衰变,并降低受试者细胞将GIC识别为mRNA或其它不期望的RNA类型而不是模板RNA的机会。

[0219] 在一些实施方案中,至少一个GIC:5' 模块RZ序列包含或编码源自至少一个非LTR逆转录因子的5' 区的核糖酶。在一些实施方案中,至少一个GIC:5' 模块RZ序列包含或编码源自来自三刺鱼(*G.aculeatus*)、美洲鲨(*L.polyphemus*)、九刺棘鱼(*P.pungitis*)、蝇蛹金小蜂(*N.vitripennis*)、中嘴地雀(*Gfortis*)、青鳉(*O.latipes*)、白喉带鹀(*Z.albicollis*)、栗耳草雀(*T.guttata*)、赤拟谷盗(*T.castaneum*) (例如来自R2家族A或B)、白喉鵂(*T.guttatus*)、其它鸟类、其它节肢动物、其它鱼类、其它被囊动物、其它动物等的基因组的非LTR逆转录因子的5' 区的核糖酶。

[0220] 在一些实施方案中,GIC:5' 模块RZ序列包含或编码具有形成丁型肝炎病毒(HDV)RZ二级和三级结构潜力的RZ,其可从自然界中发现的序列修饰和/或从头设计而不使用已知的基因组序列。在一些实施方案中,桥接配对的茎P1和P2的HDV折叠RZ序列可描述为连接点(J)1/2,其部分或全部由期望长度的靶位点序列例如5' rRNA或通过形成茎环而另外的保护的期望靶位点序列组成。在一些实施方案中,HDV折叠RZ配对的茎4(P4)设计可实现非变性GIC纯化,例如通过结合到PP7或MS2噬菌体衣壳蛋白的天然或修饰的序列。在一些实施方案中,RZ的序列被设计和优化以最小化或消除供选择的非生产性折叠。在一些实施方案中,RZ的序列被设计和优化以最小化尿苷核苷酸的数量。在一些实施方案中,RZ的序列被设计和优化以能够用在模板RNA合成过程中用掺入的核苷酸类似物完全或部分地替换规范核糖核苷酸。

[0221] 在一些实施方案中,至少一个GIC:5' 模块RZ序列可包含、编码SEQ ID NO 60-154中的至少一个或由SEQ ID NO 60-154中的至少一个编码。在一些实施方案中,RZ序列自发折叠为活性RZ。在一些实施方案中,RZ序列在5' 端包含内部rRNA序列。在一些实施方案中,RZ序列向5' 或3' 延伸。在一些实施方案中,RZ序列包含无催化活性的RZ序列。在一些实施方案中,至少一个GIC:5' 模块RZ序列可包含、编码与SEQ ID NO 60-154中的至少一个具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列或由与SEQ ID NO 60-154中的至少一个具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列编码。在一些实施方案中,GIC:5' 模块RZ序列包含与SEQ ID NO 60-154具有至少90%、91%、92%、93%、94%、95%、96%、97%、98%或99%同一性的非天然或非自然序列。

[0222] 在一些实施方案中,至少一个GIC:5' 模块RZ序列可包含、编码与SEQ ID NO 60中的至少一个具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列或由与SEQ ID NO 60中的至少一个具有至少100%、99%、98%、97%、96%、95%、90%、85%、

80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列编码。

[0223] 在一些实施方案中,至少一个GIC:5' 模块RZ序列可包含、编码与SEQ ID NO 64中的至少一个具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列或由与SEQ ID NO 64中的至少一个具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列编码。

[0224] 在一些实施方案中,至少一个GIC:5' 模块RZ序列可包含、编码与SEQ ID NO 67中的至少一个具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列或由与SEQ ID NO 67中的至少一个具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列编码。

[0225] 在一些实施方案中,至少一个GIC:5' 模块RZ序列可包含、编码与SEQ ID NO 101中的至少一个具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列或由与SEQ ID NO 101中的至少一个具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列编码。

[0226] 在一些实施方案中,至少一个GIC:5' 模块RZ序列可包含、编码与SEQ ID NO 121中的至少一个具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列或由与SEQ ID NO 121中的至少一个具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列编码。

[0227] 在一些实施方案中,至少一个GIC:5' 模块RZ序列可包含、编码与SEQ ID NO 122中的至少一个具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列或由与SEQ ID NO 122中的至少一个具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列编码。

[0228] 在一些实施方案中,至少一个GIC:5' 模块RZ序列可包含、编码与SEQ ID NO 137中至少一个具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列或由与SEQ ID NO 137中至少一个具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列编码。

[0229] GIC:5' 模块折叠序列

[0230] GIC:5' 模块折叠序列是5' 模块的任选组件,当存在时,其包含至少一个具有特定设计结构的RNA序列基序。在一些实施方案中,自主折叠RNA序列基序包含至少一个发夹基序,其例如可存在于RZ之后,以通过与随后转录的有效载荷区域碱基配对来使RZ序列免于错误折叠。在一些实施方案中,设计用于改善生产性模板RNA折叠的5' 模块区域可以与有效载荷模块或3' 模块中的另一模板RNA区域碱基配对或以其它方式直接或间接相互作用。在一些实施方案中,指导模板RNA折叠的至少一个RNA序列基序可包含至少一个将蛋白桥结合到另一茎环基序的茎环基序。在一些实施方案中,5' 模块折叠序列可以有利于模板RNA与RT编码mRNA的配对,例如以促进单独递送载体中RT编码mRNA和模板RNA的1:1化学计量的共包装。在一些实施方案中,5' 模块折叠序列可以有利于模板RNA与内源性靶细胞RNA的配对,例如出于模板RNA稳定、定位和/或其它有用结果的目的。

[0231] 在一些实施方案中,至少一个GIC:5' 模块折叠序列可包含、编码SEQ ID NO 278-279中的至少一个或由SEQ ID NO 278-279中的至少一个编码。在一些实施方案中,至少一个GIC:5' 模块折叠序列可包含、编码与SEQ ID NO 278-279中的至少一个具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列或由与SEQ ID NO 278-279中的至少一个具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列编码。在一些实施方案中,GIC:5' 模块折叠序列包含与SEQ ID NO 278-279具有至少90%、91%、92%、93%、94%、95%、96%、97%、98%或99%同一性的非天然或非自然序列。

[0232] 在一些实施方案中,至少一个GIC:5' 模块折叠序列可包含、编码与SEQ ID NO 278中的至少一个具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列或由与SEQ ID NO 278中的至少一个具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列编码。

[0233] 在一些实施方案中,至少一个GIC:5' 模块折叠序列可包含、编码与SEQ ID NO 279中的至少一个具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列或由与SEQ ID NO 279中的至少一个具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列编码。

[0234] GIC:5' 模块的模块化

[0235] 所公开的5' 模块组件可以以组合方式彼此互换地使用,以设计具有特定GIS所需或期望的功能性的5' 模块。

[0236] 在一些实施方案中,至少一个GIC:5' 模块包含至少一个GIC:5' 模块rRNA序列。在一些实施方案中,至少一个GIC:5' 模块包含至少一个GIC:5' 模块RZ序列。在一些实施方案中,至少一个GIC:5' 模块包含至少一个GIC:5' 模块折叠序列。在一些实施方案中,至少一个GIC:5' 模块包含至少一个GIC:5' 模块rRNA序列和至少一个GIC:5' 模块RZ序列。在一些实施方案中,至少一个GIC:5' 模块包含至少一个GIC:5' 模块rRNA序列和至少一个GIC:5' 模块RZ

序列和至少一个GIC:5' 模块折叠序列。

[0237] 在一些实施方案中,至少一个GIC:5' 模块可包含以下的任意组合:(a)选自、编码SEQ ID NO 250-276中的任一个或由SEQ ID NO 250-276中的任一个编码的至少一个GIC:5' 模块rRNA序列,(c)选自、编码SEQ ID NO 60-154中的任一个或由SEQ ID NO 60-154中的任一个编码的至少一个GIC:5' 模块RZ序列,和/或(d)选自、编码SEQ ID NO 278-279中的任一个或由SEQ ID NO 278-279中的任一个编码的至少一个GIC:5' 模块折叠序列。

[0238] 示例性GIC:5' 模块

[0239] 在一些实施方案中,至少一个GIC:5' 模块可包含、编码SEQ ID NO 60-154中的至少一个或由SEQ ID NO 60-154中的至少一个编码。在一些实施方案中,至少一个GIC:5' 模块可包含、编码与SEQ ID NO 60-154中的至少一个具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列或由与SEQ ID NO 60-154中的至少一个具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列编码。在一些实施方案中,GIC:5' 模块包含与SEQ ID NO 60-154具有至少90%、91%、92%、93%、94%、95%、96%、97%、98%或99%同一性的非天然或非自然序列。

[0240] 在一些实施方案中,至少一个GIC:5' 模块可包含、编码与SEQ ID NO 60、61、77和79-83中的至少一个具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列或由与SEQ ID NO 60、61、77和79-83中的至少一个具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列编码。

[0241] 在一些实施方案中,至少一个GIC:5' 模块可包含、编码与SEQ ID NO 62和63中的至少一个具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列或由与SEQ ID NO 62和63中的至少一个具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列编码。

[0242] 在一些实施方案中,至少一个GIC:5' 模块可包含、编码与SEQ ID NO 121中的至少一个具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列或由与SEQ ID NO 121中的至少一个具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列编码。

[0243] 在一些实施方案中,至少一个GIC:5' 模块可包含、编码与SEQ ID NO 117-119中的至少一个具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列或由与SEQ ID NO 117-119中的至少一个具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同

源性的序列编码。

[0244] GIC:3' 模块

[0245] 用于本公开内容的GIC的3' 模块可包含或编码至少一个源自天然逆转录因子3' UTR的序列。一般而言,3' 模块包含促进RT识别和结合GIC、定位有效载荷模块进行逆转录以及稳定GIC RNA的组件。

[0246] GIC:3' 模块架构

[0247] 在一些实施方案中,GIC:3' 模块可包含或编码至少一个GIC:3' 模块RT识别序列、任选的至少一个GIC:3' 模块rRNA序列、任选的至少一个GIC:3' 模块A-区序列及其任意组合。

[0248] 再次转向图6。展开视图(右下)图示说明示例GIC:3' 模块450的架构。GIC:3' 模块的5' 端处是GIC:3' 模块RT识别序列451,其可含有或编码由至少一个RT识别或结合的序列。当存在时,GIC:3' 模块rRNA序列452可以是GIC:3' 模块RT识别序列的3' 端,并且可包含或编码与靶位点区域同源的序列,例如可与TPRT引物3' 端碱基配对的28S rRNA核苷酸。最后,当存在时,GIC:3' 模块A-区序列453可包含腺苷富集或串联腺苷序列,其长度可能受限,例如为10至60nt,并且可以在GIC:3' 模块的3' 端。

[0249] GIC:3' 模块RT识别序列

[0250] GIC:3' 模块RT识别序列可包含或编码至少一个与至少一种逆转录酶相互作用或被至少一种逆转录酶识别的序列。不希望受理论的束缚,GIC:3' 模块RT识别序列中的至少一个RNA序列可以至少暂时与RT的至少一个模板RNA结合域诸如逆转录因子RT结合。GIC:3' 模块RT识别序列的长度和序列同一性还可起到将RT定位在GIC上的作用,使得RT逆转录的第一个核苷酸是待插入的转基因的预期3' 端。将理解的是,GIC:3' 模块RT识别序列在本文中可称为GIC:3' 模块3' UTR。

[0251] 在一些实施方案中,至少一个GIC:3' 模块RT识别序列源自天然逆转录因子的3' 区或包含天然逆转录因子的3' 区。在一些实施方案中,至少一个GIC:3' 模块RT识别序列源自来自三刺鱼(*G. aculeatus*)、黑腹果蝇(*D. melanogaster*)、美洲鲎(*L. polyphemus*)、九刺棘鱼(*P. pungitis*)、蝇蛹金小蜂(*N. vitripennis*)、中嘴地雀(*G. fortis*)、青鳉(*O. latipes*)、白喉带鹀(*Z. albicollis*)、栗耳草雀(*T. guttata*)、赤拟谷盗(*T. castaneum*)、白喉鹀(*T. guttatus*)、拟果蝇(*D. simulans*)、家蚕(*B. mori*)、污轮虫(*Adineta vaga*)、其它鸟类、其它节肢动物、其它鱼类、其它被囊动物、其它动物等的基因组的非LTR逆转录因子的3' 区。在一些实施方案中,至少一个GIC:3' 模块RT识别序列通过增加折叠的稳定性或同质性而从天然逆转录因子的3' 区修饰。在一些实施方案中,至少一个GIC:3' 模块RT识别序列针对RT相互作用的期望亲和力和/或特异性、或针对赋予作为逆转录模板的期望功能的另一机制进行设计和/或选择。在一些实施方案中,至少一个GIC:3' 模块RT识别序列被设计和/或选择为不与内源性靶细胞组件相互作用或影响内源性靶细胞组件和/或对宿主细胞有有害影响。

[0252] 在一些实施方案中,至少一个GIC:3' 模块RT识别序列(或GIC:3' 模块3' UTR序列)可包含、编码SEQ ID NO 200-224中的至少一个或由SEQ ID NO 200-224中的至少一个编码。在一些实施方案中,至少一个GIC:3' 模块RT识别序列可包含、编码与SEQ ID NO 200-221中的至少一个具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、

70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列或由与SEQ ID NO 200-221中的至少一个具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列编码。在一些实施方案中,GIC:3' 模块RT识别序列是与SEQ ID NO 200-224具有至少90%、91%、92%、93%、94%、95%、96%、97%、98%或99%同一性的非天然或非自然序列。

[0253] 在一些实施方案中,至少一个GIC:3' 模块RT识别序列可包含、编码与SEQ ID NO 202具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列或由与SEQ ID NO 202具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列编码。

[0254] 在一些实施方案中,至少一个GIC:3' 模块RT识别序列可包含、编码与SEQ ID NO 204、222、223或224中的至少一个具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列或由与SEQ ID NO 204、222、223或224中的至少一个具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列编码。

[0255] 在一些实施方案中,至少一个GIC:3' 模块RT识别序列可包含、编码与SEQ ID NO 203具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列或由与SEQ ID NO 203具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列编码。

[0256] 在一些实施方案中,GIC:3' 模块包含来自与RTC构建体编码的RT不同的物种的RT识别序列。例如,在一些实施方案中,RT识别序列可来自一种鸟类物种,而RT可来自另一种鸟类物种。在一些实施方案中,RT识别序列来自选自白喉带鹀(*Zonotrichia albicollis*)、栗耳草雀(*Taeniopygia guttata*)、白喉鵪(*Tinamus guttatns*)或中嘴地雀(*Geospiza fortis*)之一的鸟类,并且RT选自不同的鸟类物种(例如,白喉带鹀(*Zonotrichia albicollis*),栗耳草雀(*Taeniopygia guttata*)、白喉鵪(*Tinamus guttatus*)或中嘴地雀(*Geospiza fortis*))。在一些实施方案中,RTC构建体编码的RT选自白喉带鹀(*Zonotrichia albicollis*)、栗耳草雀(*Taeniopygia guttata*)、白喉鵪(*Tinamus guttatus*)或中嘴地雀(*Ggospiza fortis*)之一,并且RT识别序列选自不同的鸟类物种(例如,白喉带鹀(*Zonotrichia albicollis*),栗耳草雀(*Taeniopygia guttata*)、白喉鵪(*Tinamus guttatus*)或中嘴地雀(*Geospiza fortis*))。在一些实施方案中,RTC构建体编码的RT选自与SEQ ID NO 18或20具有至少90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%同一性的氨基酸序列,并且RT识别序列选自与SEQ ID NO 203、204、205或222-224具有至少90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%同一性的氨基酸序列。在一些实施方案中,RTC构建体编码的RT选自与SEQ ID NO:27或29具有至少90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%同一性的氨基酸序列,并且RT

识别序列选自与SEQ ID NO 202、204、205或222-224具有至少90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%同一性的氨基酸序列。在一些实施方案中,RTC构建体编码的RT选自与SEQ ID NO 25具有至少90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%同一性的氨基酸序列,并且RT识别序列选自与SEQ ID NO 202、203、204或222-224具有至少90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%同一性的氨基酸序列。在一些实施方案中,RTC构建体编码的RT选自与SEQ ID NO 31具有至少90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%同一性的氨基酸序列,并且RT识别序列选自与SEQ ID NO 202、203或205具有至少90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%同一性的氨基酸序列。

[0257] GIC:3' 模块rRNA序列

[0258] GIC:3' 模块rRNA序列,或在非rDNA靶位点处将紧接靶位点切口的下游的TPRT引物碱基配对的序列是3' 模块的任选组件,当存在时,其可包含人核糖体RNA(rRNA)序列。不希望受理论的束缚,GIC:3' 模块rRNA序列的长度和序列特性影响本文公开的GIS将转基因插入受试者基因组的准确度和效率。例如,选择一些GIC:3' 模块rRNA序列长度可以导致逆转录的内部启动,有效地缩短插入的转基因,或者可使得插入到脱靶位点,这两者都会降低在预期靶位点转基因插入的效率和特异性。RTC和GIC被工程化为要求GIC:3' 模块rRNA序列与紧接靶位点切口下游的引物序列有特定长度的碱基配对。这在靶位点使用中加入另外的保真度和另外的精确转基因插入连接的效率。GIC:3' rRNA的最佳长度小于20nt,具体为4nt,靶位点切口处所有4bp的形成具有强烈刺激。因此,如果RTC随机产生切口,使用4nt GIC:3' rRNA,则只有1/256个切口将具有最佳转基因插入。

[0259] 在一些实施方案中,至少一个GIC:3' 模块rRNA序列可包含或编码约1至30nt的rRNA。在一些实施方案中,至少一个GIC:3' 模块rRNA序列可包含或编码约1至20nt的rRNA。在一些实施方案中,至少一个GIC:3' 模块rRNA序列可包含或编码约1至10nt的rRNA。在一些实施方案中,至少一个GIC:3' 模块rRNA序列可包含或编码约1至5nt的rRNA。

[0260] 在一些实施方案中,至少一个GIC:3' 模块rRNA序列可包含或编码约1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29或30nt的rRNA的一部分。

[0261] 在一些实施方案中,至少一个GIC:3' 模块rRNA序列可包含或编码约20nt的rRNA。在一些实施方案中,至少一个GIC:3' 模块rRNA序列可包含或编码约4nt的rRNA。在一些实施方案中,至少一个GIC:3' 模块rRNA序列可包含或编码约10nt的rRNA。

[0262] 在一些实施方案中,至少一个GIC:3' 模块rRNA序列可包含SEQ ID NO 280-285中的至少一个。在一些实施方案中,至少一个GIC:3' 模块rRNA序列选自SEQ ID NO 280-289,或是其包含一个、两个或三个核苷酸取代的序列。

[0263] GIC:3' 模块A-区序列

[0264] GIC:3' 模块A-区序列是3' 模块的任选组件,当存在时,其包含具有串联腺苷(A)的末端序列区。不希望受理论的束缚,GIC:3' 模块A-区序列可稳定或保护GIC免受进一步的3' 加工,但不利于识别、核糖核蛋白组装、运输和细胞对作为mRNA的GIC的与翻译相关的衰变。此外,至少一个GIC:3' 模块A-区序列可保护GIC免于被一般单链RNA结合蛋白结合,并有助于将GIC:3' rRNA序列定位以与靶位点引物碱基配对。为了清楚起见,A-区序列不等同于天

然mRNA聚-A尾序列,其通常为约大于100-200nt的串联A。

[0265] 在一些实施方案中,任选的至少一个GIC:3' 模块A-区序列包含或编码约1至50个腺苷的序列。例如,任选的GIC:3' 模块A-区序列可包含或编码约1至50个腺苷、约5至50个腺苷、约10至50个腺苷、约15至50个腺苷、约20至50个腺苷、约25至50个腺苷、约30至50个腺苷、约35至50个腺苷、约40至50个腺苷、约45至50个腺苷、约1至45个腺苷、约5至45个腺苷、约10至45个腺苷、约15至45个腺苷、约20至45个腺苷、约25至45个腺苷、约30至45个腺苷、约35至45个腺苷、约40至45个腺苷、约1至40个腺苷、约5至40个腺苷、约10至40个腺苷、约15至40个腺苷、约20至40个腺苷、约25至40个腺苷、约30至40个腺苷、约35至40个腺苷、约1至35个腺苷、约5至35个腺苷、约10至35个腺苷、约15至35个腺苷、约20至35个腺苷、约25至35个腺苷、约30至35个腺苷、约1至30个腺苷、约5至30个腺苷、约10至30个腺苷、约15至30个腺苷、约20至30个腺苷、约25至30个腺苷、约1至25个腺苷、约5至25个腺苷、约10至25个腺苷、约15至25个腺苷、约20至25个腺苷、约1至20个腺苷、约5至20个腺苷、约10至20个腺苷、约15至20个腺苷、约1至15个腺苷、约5至15个腺苷、约10至15个腺苷、约1至10个腺苷、约5至10个腺苷或约1至5个腺苷的序列。在一些实施方案中,GIC:3' 模块A-区序列包含约1至100、1至90、1至80、1至70或1至60个腺苷。

[0266] 在一些实施方案中,任选的至少一个GIC:3' 模块A-区序列包含或编码约20至25个腺苷的序列。

[0267] 在一些实施方案中,任选的至少一个GIC:3' 模块A-区序列包含或编码约1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49或50个腺苷的序列。在一些实施方案中,GIC:3' 模块A-区序列包含22个腺苷。

[0268] GIC:3' 模块的模块化

[0269] 所公开的3' 模块组件可以以组合方式彼此互换地使用,以设计具有特定GIS所需或期望的功能的3' 模块。

[0270] 在一些实施方案中,至少一个GIC:3' 模块包含至少GIC:3' 模块RT识别序列。在一些实施方案中,至少一个GIC:3' 模块包含至少一个GIC:3' 模块rRNA序列。在一些实施方案中,至少一个GIC:3' 模块包含至少一个GIC:3' 模块A-区序列。在一些实施方案中,至少一个GIC:3' 模块包含至少GIC:3' 模块RT识别序列和至少一个GIC:3' 模块rRNA序列。在一些实施方案中,至少一个GIC:3' 模块包含至少一个GIC:3' 模块RT识别序列和至少一个GIC:3' 模块A-区序列。在一些实施方案中,至少一个GIC:3' 模块包含至少一个GIC:3' 模块RT识别序列、至少一个GIC:3' 模块rRNA序列和至少一个GIC:3' 模块A-区序列。

[0271] 在一些实施方案中,至少一个GIC:3' 模块可包含以下的任意组合:(a)选自、编码SEQ ID NO 200-221中的任一个或由SEQ ID NO 200-221中的任一个编码的至少一个GIC:3' 模块RT识别序列,(b)选自、编码SEQ ID NO 280-289中的任一个或由SEQ ID NO 280-289中的任一个编码的至少一个GIC:3' 模块rRNA序列,和/或(c)至少一个GIC:3' 模块A-区序列。

[0272] 示例性GIC:3' 模块

[0273] 在一些实施方案中,至少一个GIC:3' 模块可包含、编码SEQ ID NO 300-329中的至少一个或由SEQ ID NO 300-329中的至少一个编码。在一些实施方案中,至少一个3' 模块可

包含、编码与选自SEQ ID NO 300-329的至少一个序列具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列或由与选自SEQ ID NO 300-329的至少一个序列具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列编码。在一些实施方案中,至少一个GIC:3' 模块包含与选自SEQ ID NO 300-329的序列具有至少90%同一性(例如,至少90%、91%、92%、93%、94%、95%、96%、97%、98%或99%同一性)的序列或其任意组合。在一些实施方案中,GIC:3' 模块包含与SEQ ID NO 300-329具有至少90%、91%、92%、93%、94%、95%、96%、97%、98%或99%同一性的非天然或非自然序列。

[0274] 在一些实施方案中,至少一个GIC:3' 模块可包含、编码与SEQ ID NO 313-320中的至少一个具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列或由与SEQ ID NO 313-320中的至少一个具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列编码。

[0275] 在一些实施方案中,至少一个GIC:3' 模块可包含与选自以下的序列具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%的同源性的序列:“GACGGTAGC TAGGTTTCGA AGGCAGCCAC AAGCCAAAGA TAGGTAGGGT GCTCATAGTG AGTAGGGACA GTGCCTTTTG ATTCACAACG CGTCAATACC ATCTGACACG GATACCCTTA CCGGACTTGT CATGATCTCC CAGACTTGTC CAAGGTGGAC GGGCCACCTT TACTTAACCC GGAAAAGGAA CATATATTAA TTATATGTGT TCGGAAAA”(SEQ ID NO: 222),“CCGGACTTGT CATGATCTCC CAGACTTGTC CAAGGTGGAC GGGCCACCTT TACTTAACCC GGAAAAGGAA CATATATTAA TTATATGTGT TCGGAAAA”(SEQ ID N:223)和“CAAGGTGGAC GGGCCACCTT TACTTAACCC GGAAAAGGAA CATATATTAA TTATATGTGT TCGGAAAA”(SEQ ID NO: 224)。在一些实施方案中,这些序列进一步包括3' 序列TAGCaaaaaaaaaaaaaaaaaaaaa (SEQ ID NO:538)。

[0276] 在一些实施方案中,至少一个GIC:3' 模块可包含、编码与SEQ ID NO 315具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列或由与SEQ ID NO 315具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列编码。

[0277] 在一些实施方案中,至少一个GIC:3' 模块可包含、编码与SEQ ID NO 307具有至少95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列或由与SEQ ID NO 307具有至少95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列编码。

[0278] 在一些实施方案中,至少一个GIC:3' 模块可包含、编码与SEQ ID NO 316具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列或由与SEQ ID NO 316具有至

少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列编码。

[0279] GIC:有效载荷模块

[0280] 用于本发明的GIC的GIC:有效载荷模块包含或编码至少一个有效载荷序列,所述有效载荷序列将用作用于逆转录的模板的一部分,并通过本文公开的GIS插入到受试者基因组中。本文所用的术语“有效载荷序列”或简称为“有效载荷”是指旨在通过本发明的至少一种GIS插入靶基因组的任何生物聚合物序列。本发明的有效载荷序列可包含至少一种转基因。

[0281] 本文所用的术语“转基因”以其最广泛的意义用于指通过本发明的GIS插入受试者基因组的任何基因序列。例如,转基因可包含通常不在受试者基因组中发现的序列或通常在受试者基因组中存在但不在靶插入位点存在的序列。转基因可包括但不限于包含或编码期望的表达产物的序列(例如,至少一种mRNA、微小RNA、siRNA、rRNA、tRNA、长非编码RNA、小胞质RNA、小核RNA、小核仁RNA、小卡哈尔体RNA、环状RNA、肽、多肽和/或蛋白)和/或控制至少一种转基因表达的序列。在一些实施方案中,转基因编码选自端粒酶逆转录酶(TERT,例如,人TERT)、苯丙氨酸羟化酶(PAH,例如,人PAH)、因子VIII(例如,人因子VIII)、具有可变大小B结构域的突变的因子VIII(例如,hFactor VIIIN6和hFactor VIIIN6突变体)或因子IX(例如,人因子IX)的蛋白。在一些实施方案中,转基因编码调节RNA。在一些实施方案中,转基因编码另一种蛋白的抑制剂。在一些实施方案中,抑制剂是单链抗体。在一些实施方案中,转基因编码可用于治疗选自表X中的基因的疾病的蛋白。

[0282] 表X.代表性转基因。

[0283]

疾病	位置	基因名称
全色盲(ACHM)	CNGB3	环核苷酸门控的离子通道的 β 3亚基
全色盲(ACHM)	CNGA3	环核苷酸门控的离子通道的 β 3亚基
肾上腺脑白质营养不良	ABCD1	ALDP 蛋白
眼皮肤白化病2型	OCA2	眼皮肤白化病2型(OCA2)
β 地中海贫血	HBB	血红蛋白亚基 β
布鲁加综合征(Brugada Syndrome)	SCN5A	钠电压门控通道阿尔法亚基5
卡纳万病(Canavan disease)	ASPA	天门冬酰基酶
腓骨肌萎缩症	PMP22	外周髓鞘蛋白22
无脉络膜症(CHM)	REP1	Rab 护送蛋白1
慢性肉芽肿病(CGD)	CYBA	p22-phox(吞噬细胞氧化酶): α 亚

[0284]

		基
CILD1, 伴有或不伴有内脏逆位 (Kartagener 综合征)	DNAI1	动力蛋白轴丝中间链 1
经典的埃勒斯-当洛综合征 (Ehlers Danlos) (cEDS)	COL5A1/2	V 型胶原蛋白
颅骨锁骨发育不良 (CCD)	RUNX2	RUNX 家族转录因子 2
先天性耳聋 (出生时出现)	GJB2	间隙连接蛋白 β 2
克 - 纳 综 合 征 (Crigler-Najjar syndrome), I 型	UGT1A1	胆红素尿苷二磷酸葡萄糖醛酸转移酶
囊性纤维化	CFTR	CF 跨膜传导调节因子
家族性腺瘤样息肉病	APC	WNT 信号通路的 APC 调节因子
范科尼贫血 (Fanconi anemia)	FANCE	FA 互补组 E
脆性 X 综合征	FMR1	脆弱 X 信使核糖核蛋白 1
戈谢病 (Gaucher disease) 1 型	GBA	葡萄糖苷酰鞘氨醇酶 β 1
血色素沉着病 (铁超载)	HFE	稳态铁调节因子
血友病 A	F8	凝血因子 VIII
亨廷顿病	HTT	亨廷顿蛋白 (<i>HTT</i>)
B 型高胆固醇血症	APOB	载脂蛋白 B
低磷性佝偻病	PHEX	磷酸调节内肽酶同源物, X-连锁型
膝关节综合征	COL2A1	II 型胶原蛋白的 α -1 链
利伯氏先天性黑朦 (Leber congenital amaurosis, LCA)	CEP290	中心体蛋白 290 kDa
利伯氏先天性黑朦 (LCA)	CRB1	克鲁姆蛋白家族 (Crumbs) 成员 1, 光感受器形态发生相关的
利伯氏先天性黑朦 (LCA)	GUCY2D	鸟苷酸环化酶 2D, 膜 (视网膜特异性)
莱伯遗传性视神经病变 (Leber Hereditary Optic Neuropathy, LHON)	ND4	NADH 脱氢酶 4
莱伯遗传性视神经病变 (LHON)	ND1	NADH 脱氢酶 1
雷 - 尼 综 合 征 (Lesch-Nyhan)	HPRT1	次黄嘌呤-鸟嘌呤磷酸核糖转移酶

[0285]

syndrome, LNS)		
马凡综合征 (Marfan syndrome)	FBN1	原纤维蛋白 1
中链酰基辅酶 A 脱氢酶缺乏症	ACADM	中链酰基辅酶 A 脱氢酶
黏多糖贮积症 (MPS)	IDUA	α -L-艾杜糖醛酸酶
贝克尔型 (Becker type) 肌肉萎缩症	DMD	抗肌萎缩蛋白
杜氏型 (Duchenne type) 肌肉萎缩症	DMD	抗肌萎缩蛋白
强直性肌营养不良 1 型	DMPK	强直性肌营养不良—蛋白激酶
强直性肌营养不良 2 型	CNBP	CCHC 型锌指核酸结合蛋白
神经纤维瘤 2 型	NF2	Moesin-Ezrin-Radixin 样 (MERLIN) 肿瘤抑制因子
神经纤维瘤 1 型	NF1	神经纤维瘤病 I 型蛋白 (NF1)
尼曼 - 匹克病 (Niemann-Pick disease) A 型和 B 型	SMPD1	神经鞘磷脂酶
帕金森病	GBA	葡糖苷神经酰胺酶 β 1
苯丙酮尿症 (PKU)	PAH	苯丙氨酸羟化酶 (PAH)
多囊肾病 1 型和 2 型	PKD2	多囊肾病 2
呼吸窘迫综合征, 表面活性蛋白 B (SP-B) 缺乏症	SFTPC	表面活性剂, 肺相关蛋白 C
视网膜色素变性视野	EYS	Eyes Shut Homolog 基因
雷特综合征 (Rett's syndrome)	MECP2	甲基-CpG 结合蛋白 2
视紫红质介导的常染色体显性遗传型视网膜色素变性 (RHO-adRP)	PRPH2	外周蛋白 2
视紫红质介导的常染色体显性遗传型视网膜色素变性 (RHO-adRP)	PRPF31	前体 mRNA 加工因子 31
视紫红质介导的常染色体显性遗传型视网膜色素变性 (RHO-adRP)	RHO	视紫红质
镰状细胞贫血	HBB	血红蛋白 β 亚基
非梗阻性生精功能障碍	USP9Y	泛素特异性肽酶 9Y
脊髓性肌肉萎缩症	SMN1	运动神经元 1 的存活, 端粒的
斯特格病 (Stargardt disease)	ABCA4	ATP 结合盒亚族 A 成员 4

[0286]	泰-萨克斯病 (Tay-Sachs disease)	HEXA	己糖胺酶 A
	乌舍尔综合征 (Usher Syndrome)	MYO7A	肌球蛋白 VIIA
	卵黄状黄斑营养不良 (Best 病)	BEST1	斑萎蛋白-1
	希佩尔-林道综合征 (von Hippel-Lindau, VHL)	VHL	希佩尔-林道综合征泛素化复合物
	X 连锁视网膜色素变性 (XLRP)	RPGR	视网膜色素变性 GTP 酶调节因子
	X 连锁视网膜色素变性 (XLRP)	RP2	视网膜色素变性 2
	X 连锁视网膜劈裂症 (XLRS)	RS1	视网膜劈裂蛋白
	α 1-抗胰蛋白酶缺乏症 (COPD, 肺气肿、肝病)	SERPINA1	α 1-抗胰蛋白酶

[0287] GIC:有效载荷模块架构

[0288] GIC:有效载荷模块可包含至少一个(例如,一个、两个或三个或更多个)转基因序列,以及任选的至少一个转基因启动子序列、任选的至少一个转基因5' 非翻译序列、任选的至少一个转基因3' 非翻译序列、任选的至少一种转基因多聚腺苷酸化信号或聚-A尾序列、任选的至少一个转基因非编码RNA(ncRNA)加工序列,及其任意组合。

[0289] 再次转到图6,示例性有效载荷模块430的架构在顶部展开视图中进行图示说明。当存在时,任选的转基因启动子序列431可包含或编码至少一种启动子,该启动子可控制受试者细胞对插入的转基因的表达。任选的转基因5' UTR序列432可包含或编码当插入的转基因表达时编码转基因mRNA的5' UTR的序列。有效载荷模块的转基因序列433可包含至少一个用于通过公开的GIS进行逆转录和插入的转基因序列,例如,该序列可包含或编码感兴趣的基因的ORF。任选的转基因3' UTR序列434可包含或编码至少一个表达的转基因的mRNA的3' UTR。类似地,任选的转基因多聚腺苷酸化信号序列435可包含或编码表达的转基因的mRNA的多聚腺苷酸化信号。最后,任选的转基因非编码RNA(ncRNA)加工序列436可包含或编码转基因表达的nrRNA的终止和/或3' 加工信号。

[0290] 转基因启动子序列和RNAP II 5' UTR序列

[0291] 当存在时,转基因启动子序列可包含或编码至少一个启动子序列,所述启动子序列包含促进转基因在受试者基因组中的表达的手段。本领域已知许多这样的促进基因和/或转基因的表达的手段,包括将已知的启动子序列插入到感兴趣的基因的5'。本领域技术人员将理解,启动子序列的特性可基于转基因的特性和其它用途特定的因素来选择,因此,在本公开内容的实践中可使用任何合适的启动子。

[0292] 用于本公开内容的示例性启动子可以是组成型或诱导型。在一些实施方案中,转基因启动子序列可包含或编码至少一种RNA聚合酶I-III(RNAP I、RNAP II或III)的启动子。在一些实施方案中,代替启动子或除了启动子之外,至少一个转基因的相同区域可包含或编码至少一种核糖酶或其它基序,以使转基因RNA转录物能够从宿主细胞rDNA RNAP I转录中释放出来。

[0293] 在一些实施方案中,至少一个转基因启动子序列包含或编码至少一种人U1 snRNA启动子。在一些实施方案中,至少一个转基因启动子序列包含或编码至少一种人U3 snRNA

启动子。在一些实施方案中,至少一个转基因启动子序列包含或编码至少一种人U6 snRNA启动子。在一些实施方案中,至少一个转基因启动子序列包含或编码至少一种人tRNA启动子。

[0294] 当存在时,转基因5' UTR序列包含或编码插入的转基因的至少一个mRNA 5' UTR。一般而言,该序列包含或编码这样的序列,即当插入的转基因由细胞表达时,该序列不被细胞核糖体翻译成氨基酸生物聚合物。这些序列包括例如与转基因天然相关的5' UTR、转基因的非天然5' UTR(包括源自逆转录因子5' 序列的序列)、可以未发现与任何已知野生型基因相关的“合成的”5' UTR,及其任意组合。

[0295] 本领域技术人员将理解,转基因5' UTR序列的选择将取决于转基因的特性和其它用途特定因素,因此任何已知或发现的5' UTR序列都可以适合用于有效载荷模块的转基因5' 序列。

[0296] 在一些实施方案中,至少一个转基因启动子序列可包含、编码SEQ ID NO 400-404或408-409中的至少一个或由SEQ ID NO 400-404或408-409中的至少一个编码。在一些实施方案中,至少一个转基因启动子序列可包含、编码与SEQ ID NO 400-404或408-409中的至少一个具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列或由与SEQ ID NO 400-404或408-409中的至少一个具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列编码。

[0297] 在一些实施方案中,至少一个转基因启动子序列可包含、编码与SEQ ID NO 400具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列或由与SEQ ID NO 400具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列编码。

[0298] 在一些实施方案中,至少一个转基因启动子序列可包含、编码与SEQ ID NO 401具有至少95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列或由与SEQ ID NO 401具有至少95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列编码。

[0299] 在一些实施方案中,至少一个转基因启动子序列可包含、编码与SEQ ID NO 403具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列或由与SEQ ID NO 403具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列编码。

[0300] 在一些实施方案中,至少一个转基因启动子序列包含与SEQ ID NO 404具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列。

[0301] 在一些实施方案中,至少一个转基因启动子序列包含与SEQ ID NO 408具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、

45%、40%、35%、30%、25%、20%、10%或5%同源性的序列。

[0302] 在一些实施方案中,至少一个转基因启动子序列包含与SEQ ID NO 409具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列。

[0303] 在一些实施方案中,GIC:有效载荷模块包含位于转基因启动子序列的5'处的RNA聚合酶(RNAP)终止子序列。在一些实施方案中,RNAP是RNAP I(Pol I),并且当GIC有效载荷模块被整合到核糖体DNA基因靶位点中时,终止序列阻止Pol I读穿转录。在一些实施方案中,RNAP终止子序列包含序列5'-AGGTCGACCAGATGTCCGAGGTCGACCAGTTGTCCG-3'(SEQ ID NO:537)。

[0304] 转基因序列

[0305] 有效载荷模块的转基因序列包含或编码至少一个用于插入受试者基因组中的感兴趣序列。本文所用的术语“感兴趣序列”是指包含或编码至少一种期望表达产物的生物聚合物序列。在一些实施方案中,转基因编码选自hTERT、hPAH、hFactor VIII、具有可变大小B结构域的突变的hFactor VIII(例如,hFactor VIII N6和hFactor VIII N6突变体)或因子IX(例如,人因子IX)的蛋白。在一些实施方案中,转基因编码调节RNA。在一些实施方案中,转基因编码另一种蛋白的抑制剂。在一些实施方案中,抑制剂是单链抗体。在一些实施方案中,转基因编码可用于治疗选自表X中的基因的疾病的蛋白。

[0306] 任何感兴趣序列都可以适合用于本公开内容的实践,而限于获得序列的来源(即其来源物种或者序列是自然的还是人工的)或序列的长度。

[0307] 在一些实施方案中,至少一个转基因序列可包含、编码SEQ ID NO 411-422中的至少一个或由SEQ ID NO 411-422中的至少一个编码。在一些实施方案中,至少一个转基因序列可包含、编码与SEQ ID NO 411-422中的至少一个具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列或由与SEQ ID NO 411-422中的至少一个具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列编码。

[0308] 在一些实施方案中,至少一个转基因序列可包含、编码与SEQ ID NO 419或420中的至少一个具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列或由与SEQ ID NO 419或420中的至少一个具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列编码。

[0309] 在一些实施方案中,至少一个转基因序列可包含、编码与SEQ ID NO 421-422中的至少一个具有至少95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列或由与SEQ ID NO 421-422中的至少一个具有至少95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列编码。

[0310] 在一些实施方案中,至少一个转基因序列可包含、编码与SEQ ID NO 518-536中的至少一个具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、

60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列或由与SEQ ID NO 518-536中的至少一个具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列编码。

[0311] 转基因3' UTR序列和多聚腺苷酸化信号

[0312] 当存在时,转基因3' UTR序列包含或编码插入的转基因的至少一个mRNA 3' UTR。一般而言,该序列包含或编码当插入的转基因由细胞表达时不被细胞核糖体翻译成氨基酸生物聚合物的序列。这些序列可包括例如与转基因天然相关的3' UTR、转基因非天然的3' UTR(包括源自逆转录因子的3' 序列的序列)、不与任何已知野生型基因相关的“合成的”3' UTR及其任意组合。

[0313] 本领域技术人员将理解,转基因3' UTR序列的选择将取决于转基因的特性和其它用途特定因素,因此任何已知或被发现的3' UTR序列都可以适合用于有效载荷模块的转基因3' 序列。

[0314] 当存在时,转基因多聚腺苷酸化信号序列包含或编码至少一种转基因mRNA多聚腺苷酸化信号。任何已知或被发现的合适的多聚腺苷酸化信号可用于本公开内容的模板模块。为了清楚起见,插入的转基因中存在或编码的至少一种转基因多聚腺苷酸化信号使RNAP II能够在转基因的mRNA或ncRNA表达产物上附加聚-A尾。

[0315] 在一些实施方案中,至少一个转基因3' UTR序列可包含选自SEQ ID NO 405-407中的至少一个的序列。在一些实施方案中,至少一个转基因3' UTR序列可包含与至少一个SEQ ID NO405-407具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列。

[0316] 在一些实施方案中,至少一个转基因3' UTR序列可包含、编码与SEQ ID NO 405具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列或由与SEQ ID NO405具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列编码。

[0317] 在一些实施方案中,至少一个转基因3' UTR序列可包含、编码与SEQ ID NO406具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列或由与SEQ ID NO 406具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列编码。

[0318] 在一些实施方案中,至少一个转基因3' UTR序列可包含、编码与SEQ ID NO 407具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列或由与SEQ ID NO407具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列编码。

[0319] 转基因非编码RNA(ncRNA)加工序列

[0320] 当存在时,转基因ncRNA加工序列包含或编码控制转基因表达的ncRNA的表达或加工的序列,诸如转移RNA(tRNA)、rRNA、微小RNA、siRNA、snRNA等。在一些实施方案中,至少一

个非编码RNA(ncRNA)加工序列包含或编码至少一个转基因表达的ncRNA的至少一种终止信号、至少一种3'加工信号及其任意组合。

[0321] 在一些实施方案中,至少一个转基因ncRNA加工序列包含或编码至少一种MALAT1 3'加工和/或保护信号。在一些实施方案中,至少一个转基因ncRNA加工序列包含或编码至少一个RNA三链形成末端保护结构。在一些实施方案中,至少一个转基因ncRNA加工序列包含或编码至少一个核酸内切酶招募结构、位点或基序。在一些实施方案中,至少一个转基因ncRNA加工序列包含或编码至少一个多聚胸苷区。在一些实施方案中,至少一个转基因RNA 3'终止和/或加工序列包含RNAP I的SalI终止盒。

[0322] 有效载荷模块的模块化

[0323] 所公开的GIC:有效载荷模块组件可以以组合方式彼此互换地使用,以设计具有特定GIS所需或期望的功能的3'模块。

[0324] 在一些实施方案中,至少一个GIC:有效载荷模块可包含或编码至少一个转基因序列。在一些实施方案中,至少一个GIC:有效载荷模块可任选地包含或编码至少一个转基因启动子序列。在一些实施方案中,至少一个GIC:有效载荷模块可任选地包含或编码至少一个转基因5'UTR序列。在一些实施方案中,至少一个GIC:有效载荷模块可任选地包含或编码至少一个转基因3'UTR序列。在一些实施方案中,至少一个GIC:有效载荷模块可任选地包含或编码至少一个转基因多聚腺苷酸化信号序列。在一些实施方案中,至少一个GIC:有效载荷模块可任选地包含或编码至少一个转基因ncRNA加工序列。

[0325] 在一些实施方案中,至少一个GIC:有效载荷模块可包含或编码至少一个转基因序列、至少一个转基因启动子序列、至少一个转基因5'UTR序列、至少一个转基因3'UTR序列、至少一个转基因多聚腺苷酸化信号序列和/或至少一个ncRNA加工序列。

[0326] 在一些实施方案中,至少一个GIC:有效载荷模块可包含以下的任意组合:(a)至少一个转基因启动子序列和5'UTR序列,其选自SEQ ID NO 400-404中的任一个,或与SEQ ID NO 400-404中的任一个具有至少90%同一性(例如至少90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%同一性)的序列,(b)至少一个转基因序列,其选自、编码SEQ ID NO 411-422或SEQ ID NO 499-536中的任一个或由SEQ ID NO 411-422或SEQ ID NO 499-536中的任一个编码,或与SEQ ID NO 411-422和SEQ ID NO 499-536中的任一个具有至少90%同一性(例如至少90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%同一性)的序列,以及(c)至少一个选自SEQ ID NO 405-407的转基因3'UTR序列和多聚腺苷酸化信号,或与SEQ ID NO 405-407具有至少90%同一性(例如,至少90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%同一性)的序列。

[0327] 示例性GIC:有效载荷模块

[0328] 在一些实施方案中,至少一个GIC:有效载荷模块可包含、编码选自SEQ ID NO499-536的至少一个序列或由选自SEQ ID NO 499-536的至少一个序列编码。在一些实施方案中,至少一个GIC:有效载荷模块可包含、编码与选自SEQ ID NO499-536的至少一个序列具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列或由与选自SEQ ID NO 499-536的至少一个序列具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的

序列编码。

[0329] 在一些实施方案中,至少一个GIC:有效载荷模块可包含、编码与SEQ ID NO 419、420、518或519具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列或由与SEQ ID NO419、420、518或519具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列编码。

[0330] 在一些实施方案中,至少一个GIC:有效载荷模块可包含、编码与SEQ ID NO 421、422、520或521具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列或由与SEQ ID NO421、422、520或521具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列编码。

[0331] 在一些实施方案中,至少一种GIC:有效载荷模块可包含、编码与SEQ ID NO522、523、524或525具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列或由与SEQ ID NO522、523、524或525具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列编码。

[0332] GIC的模块化

[0333] 所公开的GIC组件(即,GIC:5' 模块、GIC:3' 模块和GIC:有效载荷模块)可以以组合方式彼此互换地使用,以设计具有特定GIS所需或期望的功能的GIC。

[0334] 在一些实施方案中,至少一种GIC包括至少一个GIC:5' 模块。在一些实施方案中,至少一种GIC包括至少一个GIC:有效载荷模块。在一些实施方案中,至少一种GIC包括至少一个GIC:3' 模块。在一些实施方案中,至少一种GIC包括至少一个GIC:5' 模块和至少一个GIC:有效载荷模块。在一些实施方案中,至少一种GIC包括至少一个GIC:5' 模块和至少一个GIC:3' 模块。在一些实施方案中,至少一种GIC包括至少一个GIC:5' 模块、至少一个GIC:有效载荷模块和至少一个GIC:3' 模块。

[0335] 在一些实施方案中,至少一种GIC包括至少一个GIC:5' 模块,该模块包含源自与GIC:3' 模块RT识别序列相同物种的逆转录因子的GIC:5' 模块RE序列。在一些实施方案中,至少一种GIC包含至少一个GIC:5' 模块,该模块包含源自与GIC:3' 模块RT识别序列不同物种的逆转录因子的GIC:5' 模块RE序列。在一些实施方案中,至少一种GIC包括至少一个GIC:5' 模块,该模块包含真核生物学非天然的GIC:5' 模块序列,并且通常可用于至少一种含有任何GIC:3' 模块RT识别序列的GIC。

[0336] 在一些实施方案中,GIC包含图7中图示说明的GIC:5' 模块序列来源和GIC:3' 模块序列来源的组合。在图7中,A1是白喉带鹀(*Zonotrichia albicollis*),A2是栗耳草雀(*Taeniopygia guttata*),A3是白喉鵂(*Tinamus guttatus*),A4中嘴地雀(*Geospiza fortis*),B1是九刺棘鱼(*Pungitius pungitius*),B2是青鳉(*Oryzias latipes*),B3是三刺鱼(*Gasterosteus aculeatus*),C1是蝇蛹金小蜂(*Nasonia vitripennis*),C2是黑腹果蝇

(*Drosophila melanogaster*), C3是赤拟谷盗(*Tribolium castaneum*), C4是家蚕(*Bombyx mori*), C5是拟果蝇(*Drosophila simulans*), C6是黄斑果蝇(*Drosophila mercatorum*), D1是库氏蝌蚪虾(*Lepidurus couseii*), D2是欧洲蝌蚪虾(*Triops cancriformis*), E1是大乳头水螅(*Hydra magnipapillata*), E2是美洲鲎(*Limulus polyphemus*), E3是污轮虫(*Adineta vaga*), 并且E4是玻璃海鞘(*Ciona intestinalis*)。

[0337] 在一些实施方案中,至少一种GIC可包含、编码以下的任意组合或由以下的任意组合编码:(a)至少一个GIC:5'模块,其选自、编码以下序列或由以下序列编码:选自SEQ ID NO 250-276的任意序列,或相对于SEQ ID NO:250-276、SEQ ID NO:60-154具有一个、两个或三个核苷酸变化或取代的序列,或与SEQ ID NO60-154、SEQ ID NO 278-279具有至少90%同一性(例如至少90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%同一性)的序列,或与SEQ ID NO 278-279具有至少90%同一性(例如至少90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%同一性)的序列,(b)至少一个GIC:有效载荷模块,其选自、编码以下序列或由以下序列编码:选自SEQ ID NO411-422或499-525之一的任何序列,或与SEQ ID NO411-422或499-522具有至少90%同一性(例如,至少90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%同一性)的序列,和/或(c)至少一个GIC:3'模块,其选自、编码以下序列或由以下序列编码:选自SEQ ID NO 300-329之一的任何序列,或与SEQ ID NO 300-329具有至少90%同一性(例如,至少90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%同一性)的序列。

[0338] 示例性GIC

[0339] 在一些实施方案中,至少一种GIC可包含、编码SEQ ID NO 411-422或499-525中的至少一个或由SEQ ID NO 411-422或499-525中的至少一个编码。在一些实施方案中,至少一种GIC可包含、编码与SEQ ID NO411-422或499-536中的至少一个具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列或由与SEQ ID NO 411-422或499-536中的至少一个具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列编码。

[0340] 在一些实施方案中,至少一种GIC可包含、编码与SEQ ID NO 419、420、518或519具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列或由与SEQ ID NO 419、420、518或519具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列编码。

[0341] 在一些实施方案中,至少一种GIC可包含、编码与SEQ ID NO 421、422、520或521具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列或由与SEQ ID NO421、422、520或521具有至少100%、99%、98%、97%、96%、95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列编码。

[0342] 在一些实施方案中,至少一种GIC可包含、编码与SEQ ID NO522、523、524或525具有至少95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列或由与SEQ ID NO 522、523、524或525具有至少95%、

90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、10%或5%同源性的序列编码。

[0343] GIS设计和模块化

[0344] 所公开的GIS组件(即RTC和GIC)可以以组合方式彼此互换地使用,以设计具有所需或期望的功能性的GIS。

[0345] 在一些实施方案中,至少一种GIS可包含至少一种RTC。在一些实施方案中,至少一中GIS可包含至少一种GIC。在一些实施方案中,至少一种GIS可包含至少RTC和至少一种GIC。

[0346] GIS生物聚合物的组成

[0347] 包含GIS组件的生物聚合物的组合物可以从本文公开的那些中以组合方式选择,以设计具有所需或期望的功能性的GIS。

[0348] 在一些实施方案中,可以将至少一种RTC作为RNA生物聚合物引入至少一个受试者。在一些实施方案中,可以将至少一种RTC作为mRNA生物聚合物引入至少一个受试者。

[0349] 在一些实施方案中,可以将至少一种GIC作为RNA生物聚合物引入至少一个受试者。在一些实施方案中,可以将至少一种GIC作为线性RNA生物聚合物引入至少一个受试者。

[0350] 在一些实施方案中,可以将至少一种RTC作为RNA生物聚合物引入至少一个受试者,并且可以将至少一种GIC作为RNA生物聚合物引入至少一个受试者。

[0351] 在一些实施方案中,可以将至少一种RTC作为mRNA生物聚合物引入至少一个受试者,并且可以将至少一种GIC作为RNA生物聚合物引入至少一个受试者。

[0352] 在一些实施方案中,可以将至少一种RTC和/或至少一种GIC作为DNA生物聚合物引入至少一个受试者。在一些实施方案中,可以将至少一种RTC和/或至少一种GIC作为质粒引入至少一个受试者。

[0353] 在一些实施方案中,可以将至少一种RTC作为氨基酸生物聚合物引入至少一个受试者。在一些实施方案中,可以将至少一种RTC作为蛋白引入至少一个受试者。

[0354] 在一些实施方案中,可以将至少一种RTC作为氨基酸生物聚合物引入至少一个受试者,并且可以将至少一种GIC作为RNA生物聚合物引入至少一个受试者。在一些实施方案中,可以将至少一种RTC作为质粒引入至少一个受试者,并且可以将至少一种GIC作为RNA生物聚合物引入至少一个受试者。

[0355] 在一些实施方案中,可以将至少一种RTC作为质粒引入至少一个受试者,并且可以将至少一种GIC作为质粒引入至少一个受试者。在一些实施方案中,可以将至少一种RTC作为RNA(例如,mRNA)引入至少一个受试者,并且可以将至少一种GIC作为质粒引入至少一个受试者。

[0356] 配对的RT

[0357] 可以通过设计或选择GIS的GIC、RTC或两者中的至少一个的组成针对期望的功能来优化本发明的GIS,以控制GIC和RTC之间的相互作用。例如,改变GIC和/或RTC的组成可以实现全长有效载荷插入的效率、速率和/或保真度的变化,如通过使用PCR、测序检测插入和/或通过有效载荷转基因表达来监测的;有效载荷插入的靶位点选择的序列特异性和/或染色体位置,如通过测序、杂交或插入的DNA的基因组位置的其它可视化监测的;RTC仅利用给予的GIC作为逆转录模板的选择性;等等。术语“配对的RT”在本文中用于指与特定GIC序

列组合给予的特定RTC:RT模块序列。

[0358] 不希望受理论的束缚,改变RTC和GIC的相互作用可以通过选择RTC:RT模块和GIC:5' 模块和/或GIC:3' 模块来实现。例如,可以通过选择源自相同或不同物种的逆转录因子的组件来改变RTC对GIC的特异性。如本文所用,如果两个GIS组件源自相同物种的逆转录因子,则称它们为同源的。相反,如果两个GIS组件源自不同物种的逆转录因子,则称它们为异源的。

[0359] 在一些实施方案中,至少一个RTC:RT模块包含或编码至少一个序列,所述至少一个序列源自与至少一个源自逆转录因子的GIC:5' 模块和/或GIC:3' 模块序列不同物种的逆转录因子(本文称为“异源配对的RT”)。

[0360] 在一些实施方案中,RTC和GIC两者中源自逆转录因子的所有序列均源自相同物种的逆转录因子(本文称为“同源配对的RT”)。

[0361] 在一些实施方案中,与同源配对的RT相比,异源配对的RT可具有增加的特异性。

[0362] 本文所用的术语“特异性”是指配对的RT将有效和/或优先利用预期模板RNA进行转基因插入的可能性。

[0363] 在一些实施方案中,至少一种GIS可包含GIC和配对的RT的至少一种组合,如图7图说明的。

[0364] 示例性GIS

[0365] 在一些实施方案中,至少一种GIS可包含、编码以下的任意组合或由以下的任意组合编码:(a) 选自、编码以下序列或由以下序列编码的至少一种RTC:选自SEQ ID NO1-59之一的任意序列或与SEQ ID NO1-59之一具有至少90%同一性(例如,至少90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%同一性)的序列,和(b) 选自、编码以下序列或由以下序列编码的至少一种GIC:包含SEQ ID NO250-276之一的任意序列,或相对于SEQ ID NO:250-276;SEQ ID NO 60-154具有一个、两个或三个核苷酸变化或取代的序列,或与SEQ ID NO 60-154、SEQ ID NO278-279具有至少90%同一性(例如,至少90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%同一性)的序列,或与SEQ ID NO278-279;SEQ ID NO 411-422或499-536具有至少90%同一性(例如,至少90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%同一性)的序列,或与SEQ ID NO 411-422或499-536;和/或SEQ ID NO300-329具有至少90%同一性(例如,至少90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%同一性)的序列,或与SEQ ID NO 300-329具有至少90%同一性(例如,至少90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%同一性)的序列。

[0366] III. 制剂和递送机制

[0367] 核酸

[0368] 在一些实施方案中,RTC构建体或GIC构建体可含有一种或多种修饰的核苷酸,诸如但不限于核碱基修饰、糖修饰的核苷酸和/或骨架修饰。在一些实施方案中,RTC构建体或GIC构建体可含有组合修饰,例如组合的核碱基和骨架修饰。

[0369] 在一些实施方案中,修饰的核苷酸可为核碱基修饰的核苷酸。修饰的碱基是指通过替换或添加一个或多个基团或原子而已经被修饰的核苷酸碱基,诸如但不限于腺嘌呤、胞嘧啶、胸腺嘧啶、鸟嘌呤、尿嘧啶、黄嘌呤、肌苷和腺苷。在一些实施方案中,修饰的核苷酸

可为骨架修饰的核苷酸。

[0370] RTC构建体和/或GIC构建体可相对于参考序列,特别是感兴趣序列包含一个或多个替换、插入和/或添加、缺失和共价修饰,这些都包括在本发明的范围内。

[0371] 在一些实施方案中,RTC构建体和/或GIC构建体包含一种或多种转录后修饰(例如,加帽、剪切、多聚腺苷酸化、剪接、聚-A序列、甲基化、酰化、磷酸化、赖氨酸和精氨酸残基的甲基化、乙酰化、硫醇基团和酪氨酸残基的亚硝基化等)。

[0372] RTC构建体和/或GIC构建体可包含任何有用的修饰,诸如对糖、核碱基或核苷间键(例如,对连接磷酸酯/对磷酸二酯键/对磷酸二酯骨架)的修饰。

[0373] 在一些实施方案中,修饰可包括化学或细胞诱导的修饰。例如,Lewis和Pan在Nat Reviews Mol Cell Biol,2017,18:202-210中的“RNA修饰和结构协同引导RNA-蛋白相互作用(RNA modifications and structures cooperate to guide RNA-protein interactions)”中描述了一些细胞内RNA修饰的非限制性实例。

[0374] 在一些实施方案中,对RNA的化学修饰可以增强免疫逃避。RNA可通过本领域中公认的方法合成和/或修饰。

[0375] 在一些实施方案中,至少一种RNA构建体可包含至少一种修饰的尿嘧啶。尿嘧啶修饰的实例包括5-甲基-尿苷、5-甲氧基-尿苷、假尿苷、N1-甲基-假尿苷和/或2-硫尿苷。在一些实施方案中,至少一种RNA构建体可包含至少一种修饰的腺苷。腺苷修饰的实例包括2,6--二氨基嘌呤脱氧核苷酸。

[0376] 在一些实施方案中,糖修饰(例如,在2'位置或4'位置)或一个或多个RNA的糖替换以及骨架修饰可包括磷酸二酯键的修饰或替换。

[0377] 递送机制

[0378] 本发明的基因插入系统(GIS)可通过本领域已知的任何递送机制引入受试者。本文所用的“递送机制”是指用于将GIS、GIS的组件或GIS的产物引入受试者的方法或组合物。递送机制的非限制性实例包括递送载体、直接转染(诸如使用转染剂)、植入先前用GIS转染的细胞,及其任意组合。

[0379] 递送载体

[0380] 在一些实施方案中,本发明的GIS可配制在递送载体中。一般而言,递送载体可通过保护GIS组件免于在细胞外环境中降解、促进受试者细胞吸收、增强内体逃逸及其任意组合来促进受试者细胞的体内或体外转染。递送载体可包括但不限于纳米颗粒,包括基于脂质的纳米颗粒(例如,脂质纳米颗粒(LNP)、脂质体和胶束)和非脂质纳米颗粒(例如,病毒样颗粒(VLP)和聚合物递送颗粒)。

[0381] 纳米颗粒

[0382] 在一些实施方案中,递送载体可包含至少一种纳米颗粒。一般而言,本文所用的术语“纳米颗粒”可指尺寸范围为10-1000nm的任何颗粒,例如颗粒可以为10、15、20、25、30、35、40、45、50、55、60、65、70、75、80、85、90、95、100、105、110、115、120、125、130、135、140、145、150、155、160、165、170、175、180、185、190、195、200、205、210、215、220、225、230、235、240、245、250、255、260、265、270、275、280、285、290、295、300、305、310、315、320、325、330、335、340、345、350、355、360、365、370、375、380、385、390、395、400、405、410、415、420、425、430、435、440、445、450、455、460、465、470、475、480、485、490、495、500、505、510、515、520、

525、530、535、540、545、550、555、560、565、570、575、580、585、590、595、600、605、610、615、620、625、630、635、640、645、650、655、660、665、670、675、680、685、690、695、700、705、710、715、720、725、730、735、740、745、750、755、760、765、770、775、780、785、790、795、800、805、810、815、820、825、830、835、840、845、850、855、860、865、870、875、880、885、890、895、900、905、910、915、920、925、930、935、940、945、950、955、960、965、970、975、980、985、990、995或1000nm。

[0383] 脂质基颗粒

[0384] 在一些实施方案中,递送载体可包含至少一种脂质基纳米颗粒,包括但不限于脂质纳米颗粒(LNP)、脂质体、胶束及其任意组合。

[0385] 脂质纳米颗粒

[0386] 在一些实施方案中,递送载体可以是脂质纳米颗粒(LNP)。一般而言,LNP具有外部脂质层,包括暴露于非LNP环境的亲水性外表面、非水性或水性内部空间(即,分别为胶束样和囊泡样LNP)和至少一个疏水性膜间空间。LNP膜可以是非层状的或层状的,并且可以由1、2、3、4、5或超过5层组成。LNP可以是固体或半固体。在一些实施方案中,至少一种负荷或有效载荷(诸如GIS)可存在于LNP的内部空间、膜间空间中、外部表面上或其任意组合。

[0387] 本文中有用的LNP是本领域已知的,并且通常包括可离子化(阳离子)脂质、磷脂、胆固醇和聚合物缀合的脂质。不希望受理论的束缚,胆固醇促进膜融合,并有助于LNP稳定性,磷脂可有助于内体逃逸,并为LNP双层提供结构,聚合物的缀合脂质减少LNP聚集,并“保护”LNP免受免疫细胞的非特异性内吞作用,并且可离子化(阳离子)脂质增强内体逃逸,并复合带负电荷的负荷(诸如GIS的多核苷酸)。

[0388] 在一些实施方案中,本发明的GIS可结合到LNP中。在一些实施方案中,脂质纳米颗粒可由至少一种阳离子脂质(例如,可离子化的阳离子脂质)、至少一种非阳离子脂质(例如,磷脂)、至少一种固醇(例如,胆固醇)、至少一种聚合物缀合的脂质(例如,PEG-脂质)或其任意组合组成。在一些实施方案中,脂质纳米颗粒可由至少一种阳离子脂质(例如,可离子化的阳离子脂质)、至少一种非阳离子脂质(例如,磷脂)、至少一种固醇(例如,胆固醇)和至少一种聚合物缀合的脂质(例如,PEG-脂质)组成。在一些实施方案中,LNP可由至少一种阳离子脂质(例如,可离子化的阳离子脂质)、至少一种非阳离子脂质和至少一种固醇(例如,胆固醇)组成。在一些实施方案中,LNP可由至少一种阳离子脂质(例如,可离子化的阳离子脂质)、至少一种非阳离子脂质(例如,磷脂)和至少一种聚合物缀合的脂质(例如,PEG-脂质)组成。在一些实施方案中,LNP可由至少一种非阳离子脂质(例如,磷脂)、至少一种固醇(例如,胆固醇)和至少一种聚合物缀合的脂质(例如,PEG-脂质)组成。在一些实施方案中,LNP可由至少一种阳离子脂质(例如,可离子化的阳离子脂质)和至少一种非阳离子脂质(例如,磷脂)组成。在一些实施方案中,LNP可由至少一种阳离子脂质(例如,可离子化的阳离子脂质)和至少一种固醇组成。在一些实施方案中,LNP可由至少一种阳离子脂质(例如,可离子化的阳离子脂质)和至少一种聚合物缀合的脂质(例如,PEG-脂质)组成。在一些实施方案中,LNP可由至少一种非阳离子脂质(例如,磷脂)和至少一种固醇(例如,胆固醇)组成。在一些实施方案中,LNP可由至少一种非阳离子脂质(例如,磷脂)和至少一种聚合物缀合的脂质(例如,PEG-脂质)组成。在一些实施方案中,LNP可由至少一种固醇(例如胆固醇)和至少一种聚合物缀合的脂质(例如PEG-脂质)组成。在一些实施方案中,LNP可由至少一种阳离子脂

质(例如可离子化的阳离子脂质)组成。在一些实施方案中,LNP可由至少一种非阳离子脂质(例如磷脂)组成。在一些实施方案中,LNP可由固醇(例如胆固醇)组成。在一些实施方案中,LNP可由聚合物缀合的脂质(例如PEG-脂质)组成。

[0389] 本文所述的LNP可使用本领域已知的技术形成。作为非限制性实例,将含有脂质的有机溶液与含有GIS的酸性水溶液在微流体通道中混合,结果形成载有GIS的递送载体。

[0390] 胶束

[0391] 在一些实施方案中,递送载体包含至少一种胶束。在一些实施方案中,胶束可由与脂质纳米颗粒相同的任何或所有组分组成,主要区别在于它们的制造方法。本文所用的“胶束”是指不具有水性颗粒内空间的小颗粒。不希望受理论的束缚,胶束的颗粒内空间不包含任何另外的脂质-头基,而是由包含胶束膜和可能相关的GIS的脂质的疏水尾部占据。

[0392] 脂质体

[0393] 在一些实施方案中,递送载体包括至少一种脂质体。在一些实施方案中,脂质体可由与脂质纳米颗粒相同的任何或所有组分和相同的组分量组成,主要区别在于它们的制造方法。本文使用的“脂质体”是指由至少一层包围水性纳米颗粒内空间的脂质双层膜组成的小囊泡。此外,脂质体与细胞外囊泡的不同之处在于它们通常不是来自祖细胞/宿主细胞。脂质体的直径可能为数百纳米,包括一系列由狭窄水性空间隔开的同心双层(即(大)多层囊泡(MLV)),直径可能小于50nm(小单室囊泡(SUV)),直径可能为50至500nm(大单层囊泡(LUV))。

[0394] 外泌体

[0395] 在一些实施方案中,递送载体包括至少一种外泌体。一般而言,“外泌体”是指具有内吞起源的小型的、膜结合的细胞外囊泡。外泌体膜通常由双层脂质和层状物组成,具有水性纳米颗粒间空间。除了设计的组件之外,外泌体将倾向于还包括它们所源自的宿主/祖细胞膜的组件。不希望受理论的束缚,外泌体通常在多囊泡体与细胞质膜融合后从宿主/祖细胞释放到细胞外环境中。

[0396] 病毒样颗粒

[0397] 在一些实施方案中,递送载体包括至少一种病毒样颗粒(VLP)。一般而言,病毒样颗粒是一种非感染性囊泡,其主要由源自可装载GIS的病毒的蛋白衣壳、外壳、壳或鞘(在本文中全部应理解为可互换使用的等同物)组成。在一些实施方案中,VLP可使用细胞机制以表达病毒衣壳蛋白序列来合成,然后其自组装并结合GIS。在一些实施方案中,VLP可通过提供衣壳和GIS组件而不表达相关的细胞机制并使它们自组装来形成。

[0398] 可得到VLP的病毒科和种的非限制性实例包括细小病毒科、逆转录病毒科、黄病毒科、副粘病毒科、腺相关病毒、HIV、丙型肝炎病毒、HPV、噬菌体或其任意组合。

[0399] 聚合物递送颗粒

[0400] 在一些实施方案中,递送载体可包含至少一种聚合物递送颗粒。本文所用的“聚合物递送颗粒”是指由通过各种连接基团与GIS部分缀合的可溶性聚合物组成的非聚集性递送颗粒。在一些实施方案中,聚合物递送剂可包含本文所述的任何聚合物。

[0401] 在一些实施方案中,递送载体可包含核酸纳米颗粒(NANP)。一般而言,“核酸纳米颗粒”是由非编码核酸序列形成的小颗粒,其相互作用以形成能够携带负荷(例如,GIS组件)的三维结构。

[0402] 封装

[0403] 在一些实施方案中,递送载体可完全封装本文所公开的GIS。在一些实施方案中,递送载体可部分封装本文所公开的GIS。在一些实施方案中,在最终制剂中,基本上0%的存在的GIS暴露于递送载体外部的环境(即,GIS被完全封装)。在一些实施方案中,GIS与递送载体相连,但至少部分暴露于递送载体外部的环境。

[0404] 在一些实施方案中,递送载体的特征可在于封装效率,即未暴露于递送载体外部的环境的GIS的百分比。为了清楚起见,约100%的封装效率是指这样的递送载体制剂,其中基本上所有GIS都被递送载体完全封装,而约0%的封装率是指这样的递送载体,其中基本上没有GIS被封装在递送载体中,诸如采用其中GIS与递送载体的外表面结合的递送载体。在一些实施方案中,递送载体可具有小于约100%、小于约95%、小于约85%、小于约80%、小于约75%、小于约70%、小于约65%、小于约60%、小于约55%、小于约50%、小于约45%、小于约40%、小于约35%、小于约30%、小于约25%、小于约20%、小于约15%、小于约10%或小于5%的封装效率。在一些实施方案中,递送载体可具有约90至100%、80至100%、70至100%、60至100%、50至100%、40至100%、30至100%、20至100%、10至100%、80至90%、70至90%、60至90%、50至90%、40至90%、30至90%、20至90%、10至90%、70至80%、60至80%、50至80%、40至80%、30至80%、20至80%、10至80%、60至70%、50至70%、40至70%、30至70%、20至70%、10至70%、40至50%、30至50%、20至50%、10至50%、30至40%、20至40%、10至40%、20至30%、10至30%和10至20%的封装效率。

[0405] 递送载体纳米颗粒的物理特性

[0406] 在一些实施方案中,递送载体可通过它们的形状表征。在一些实施方案中,递送载体可以是但不限于基本上球形、基本上棒形(即圆柱形)或基本上盘形。

[0407] 在一些实施方案中,递送载体可通过它们的尺寸表征。在一些实施方案中,递送载体的尺寸可定义为其直径。本文中使用的与递送载体尺寸相关的“直径”是指递送载体其最大圆形横截面的直径。在一些实施方案中,递送载体的直径可为30nm至约150nm。例如,递送载体的直径范围可为约40至150nm、50至150nm、60至150nm、约70至150nm或80至150nm、90至150nm、100至150nm、110至150nm、120至150nm、130至150nm、140至150nm、30至30至140nm、40至140nm、50至140nm、60至140nm、70至140nm、80至140nm、90至140nm、100至140nm、110至140nm、120至140nm、130至140nm、140至140nm、30至140nm、40至130nm、50至130nm、60至130nm、70至130nm、80至130nm、90至130nm、100至130nm、110至130nm、120至130nm、30至120nm、40至120nm、50至120nm、60至120nm、70至120nm、80至120nm、90至120nm、100至120nm、110至120nm、30至110nm、40至110nm、50至110nm、60至110nm、70至110nm、80至110nm、90至110nm、100至110nm、30至100nm、40至100nm、50至100nm、60至100nm、70至100nm、80至100nm、90至100nm、30至90nm、40至90nm、50至90nm、60至90nm、70至90nm、80至90nm、30至80nm、40至80nm、50至80nm、60至80nm、70至80nm、30至70nm、40至70nm、50至70nm、60至70nm、30至60nm、40至60nm、50至60nm、30至50nm、40至50nm和30至40nm。

[0408] 在一些实施方案中,可通过测量群中颗粒的物理特性(例如,大小、形状或质量)的均匀性来表征递送载体群,例如由相同制剂得到的所有递送载体。在一些实施方案中,均匀性可表示为群的多分散性指数(PI)。在一些实施方案中,均匀性可表示为群的差异(**D**)。本文所用的术语“多分散性指数”和“差异”应理解为等同的并且可以互换使用。

[0409] 在一些实施方案中,由给定制剂得到的递送载体群将具有约0.1至1的PI。在一些实施方案中,由给定制剂得到的递送载体群将具有约0.1至1、0.1至0.8、0.1至0.6、0.1至0.4、0.1至0.2、0.2至1、0.2至0.8、0.2至0.6、0.2至0.4、0.4至1、0.4至0.8、0.4至0.6、0.6至1、0.6至0.8和0.8至1的PI。在一些实施方案中,由给定制剂得到的递送载体群将具有小于约1、小于约0.5、小于约0.4、小于约0.3,小于约0.2、小于约0.1的PI。

[0410] 递送靶向

[0411] 在一些实施方案中,用GIS配制的递送载体可促进GIS定位到本文所述的任何靶向区域、组织、细胞或生理系统(即,递送载体“靶向”指定位置)。在一些实施方案中,靶向可通过给定的递送载体结构组件的制剂来实现。在一些实施方案中,递送载体可包含靶向剂。

[0412] 靶向剂

[0413] 在一些实施方案中,递送载体可包含至少一种靶向剂。本文所用的术语靶向剂在一些实施方案中可指特异性结合特定类型或类别的细胞和/或其它特定类型的化合物的部分、化合物、抗体等(例如,靶向特异性细胞或细胞类型的部分)。在一些实施方案中,靶向剂可对某些靶细胞的表面、靶细胞表面抗原、靶细胞受体或其组合具有亲和力(即,对其具有特异性)。

[0414] 在一些实施方案中,靶向剂可指在暴露于特定类型或类别的物质和/或细胞时具有特定作用(例如,裂解)的试剂,并且该作用可驱动递送载体靶向特定类型或类别的细胞。

[0415] 在一些实施方案中,术语靶向剂可指可为递送载体的一部分且在递送载体对靶点的特异性中发挥作用的试剂,尽管试剂本身可以对特定类型或类别的细胞本身具有特异性或可以对其不具有特异性。

[0416] 在一些实施方案中,递送载体中存在至少一种靶向剂可增加递送载体递送的GIS的细胞摄取效率(例如,总量或速率)。在一些实施方案中,递送载体中存在至少一种靶向剂可增加递送载体递送的GIS的细胞摄取特异性(例如,总量或速率)。本文所用的“特异性”是指靶细胞比非靶细胞更高的细胞摄取效率。

[0417] 在一些实施方案中,合适的靶向剂可包括但不限于一种或多种小分子靶向剂(例如,碳水化合物部分)、抗体、抗体样分子、肽、维生素(例如,叶酸)、糖(例如,乳糖和半乳糖)、人工亲和分子(例如,肽模拟物或适体)、抗体片段、单链可变片段(scFv)、细胞表面受体(例如,T细胞受体(TCR)、B细胞受体(BCR)或嵌合抗原受体(CAR))及其任意组合。

[0418] 在一些实施方案中,靶向剂可靶向的细胞表面抗原可包括靶细胞的任何细胞表面分子。合适的细胞表面分子的实例包括但不限于细胞表面上的蛋白、糖、脂质或其它抗原。在一些实施方案中,细胞表面抗原经历内化。

[0419] 在一些具体实施方案中,递送载体可包含多于一种的靶向剂。

[0420] 在一些实施方案中,至少一种靶向剂可结合到纳米颗粒的脂质膜中。在一些实施方案中,至少一种靶向剂可存在于纳米颗粒的外表面上。在一些实施方案中,至少一种靶向剂可与纳米颗粒的脂质组分缀合。在一些实施方案中,至少一种靶向剂可与纳米颗粒的聚合物组分缀合。在一些实施方案中,包含靶向剂残基的单体(例如,靶向剂的可聚合衍生物,诸如肽的(烷基)丙烯酸衍生物)可共聚以形成形成递送载体的聚合物缀合的脂质。在一些实施方案中,至少一种靶向剂可通过至少一种靶向剂、纳米颗粒膜和纳米颗粒内部或外部的水环境之间的疏水和亲水相互作用被锚定到纳米颗粒上。在一些实施方案中,至少一种

靶向剂与纳米颗粒膜的肽/蛋白组分缀合。在一些实施方案中,至少一种靶向剂与合适的接头部分缀合,该接头部分与纳米颗粒膜的组分缀合。在一些实施方案中,任何力和键的组合都可导致靶向剂与纳米颗粒相连。

[0421] 在一些实施方案中,一种或多种靶向剂可通过连接部分与至少一种递送载体的聚合物偶联。在一些实施方案中,连接部分可以是可剪切的连接部分(例如,包含可剪切的键)。在一些实施方案中,连接部分可包含可被特定酶(例如磷酸酶或蛋白酶)剪切的键。在一些实施方案中,连接部分可包含在细胞内pH、氧化还原电位或其它细胞内参数改变时可剪切的键。在一些实施方案中,连接部分可包含可在暴露于基质金属蛋白酶(MMP)时被剪切的键。

[0422] 直接转染

[0423] 在一些实施方案中,本文公开的GIS可直接转染到靶细胞中,而不使用递送载体。在一些实施方案中,本文公开的GIS可使用本领域已知的任何技术转染到靶细胞中。这样的技术可包括但不限于化学转染方法(例如磷酸钙暴露)、物理转染方法(例如电穿孔、微注射和生物弹道颗粒递送)。在一些实施方案中,可利用脂质介导的转染剂诸如但不限于 lipofectamine、lipofectamine 2000及其任意组合进行直接转染。

[0424] 转染的细胞的植入

[0425] 在一些实施方案中,本发明的GIS可在体外引入细胞群(例如,通过本文所述的直接转染)用于随后植入受试者。在一些实施方案中,用于植入的细胞群可以是干细胞。在一些实施方案中,用于植入的细胞群可以来自受试者。在一些实施方案中,可通过本领域已知的任何方法进行植入。

[0426] IV. 药物组合物和给药途径

[0427] 本发明提供用于向受试者给予GIS的药物组合物。在一些实施方案中,本发明提供用作治疗指征的治疗中的药物的药物组合物。在一些实施方案中,药物组合物包含至少一种活性成分(例如,本发明的GIS)和至少一种药学上可接受的赋形剂、佐剂、载体、稀释剂或其任意组合。在一些实施方案中,药物组合物被配制用于至少一种给药途径。在一些实施方案中,药物组合物被配制用于递送指定剂量的至少一种活性成分(例如,GIS),任选地按照指定的时间表递送。

[0428] 本文所用的术语“药物组合物”是指包含至少一种活性成分和任选的一种或多种药学上可接受的赋形剂的组合物。本文所用的短语“活性成分”通常指GIS、GIS携带的用于插入受试者基因组的基因有效载荷或本文所述的GIS携带的基因有效载荷的表达产物中的任一种。

[0429] 药物制剂和组合物

[0430] GIS可使用一种或多种赋形剂配制以:(1)增加GIS或包含GIS的递送机制的稳定性;(2)增加细胞转染或转导;(3)允许持续或延迟地将GIS引入受试者的细胞;(4)改变生物分布(例如,使GIS靶向特定组织或细胞类型);(5)增加编码基因的表达;(6)改变编码蛋白的释放曲线;和/或(7)实现GIS和/或GIS有效载荷的可调节表达。

[0431] 不受限制地,制剂可包含盐水、脂质体、脂质纳米颗粒、聚合物、肽、蛋白、用GIS转染的细胞(例如,用于转移或移植到受试者体内)及其任意组合。

[0432] 在一些实施方案中,本文所述的药物组合物的制剂可通过药理学领域中已知或今

后开发的任何方法来制备。一般而言,这样的制备方法包括将活性成分与赋形剂和/或一种或多种其它辅助成分结合的步骤。

[0433] 本文所述的GIS和药物组合物的制剂可通过药理学领域中已知或今后开发的任何方法来制备。一般而言,这样的制备方法包括将活性成分与赋形剂和/或一种或多种其它辅助成分结合的步骤,然后,如果必要和/或期望,将产品分割、成型和/或包装成期望的单剂量或多剂量单位。

[0434] 本文所述的药物组合物可批量、作为单个单位剂量和/或作为多个单个单位剂量进行制备、包装和/或出售。本文所用的“单位剂量”是指包含预定量活性成分的药物组合物的离散的量。活性成分的量通常等于将给予受试者的活性成分的剂量和/或这样的剂量的便利的份数,例如这样的剂量的一半或三分之一。

[0435] 在一些实施方案中,赋形剂被批准用于人和用于兽医用途。在一些实施方案中,赋形剂可由美国食品药品监督管理局(United States Food and Drug Administration)批准。在一些实施方案中,赋形剂可符合美国药典(USP)、欧洲药典(EP)、英国药典和/或国际药典的标准。在一些实施方案中,药学上可接受的赋形剂的纯度可为至少100%、至少99%、至少98%、至少97%、至少96%或95%。在一些实施方案中,赋形剂可为药物级。

[0436] 在一些实施方案中,药学上可接受的赋形剂、活性成分和/或任何另外的成分的相对量在本发明的药物组合物中可不同。在一些实施方案中,相对量可根据所治疗的受试者的尺寸、状况和/或特性而不同。在一些实施方案中,相对量可根据组合物的给药途径而不同。例如,组合物可包含0.1%至100%(例如,0.1%至99%、0.5至50%、1-30%、5-80%或至少80%(重量/重量))的活性成分。

[0437] 赋形剂、稀释剂和非活性成分

[0438] 在一些实施方案中,药物组合物可包含本领域已知或发现的任何赋形剂。合适的赋形剂的实例包括但不限于任何和所有防腐剂、等渗剂、增稠剂或乳化剂、溶剂、分散介质、稀释剂或其它液体载体、分散或悬浮助剂、表面活性剂及其组合。在一些实施方案中,可基于赋形剂对期望特定剂型的适用性来选择赋形剂。

[0439] 在一些实施方案中,本文描述的制剂可包含至少一种非活性成分。本文所用的术语“非活性成分”是指制剂中包含的一种或多种对药物组合物的活性成分的活性没有贡献的试剂。在一些实施方案中,药物组合物中的非活性成分中的一些或所有可以获得美国食品药品监督管理局(FDA)的批准,或药物组合物中的非活性成分均可以不获得美国食品药品监督管理局(FDA)的批准。

[0440] 在一些实施方案中,本文公开的药物制剂可包含阳离子或阴离子。在一些实施方案中,药物制剂包含金属阳离子,诸如但不限于 Ca^{2+} 、 Zn^{2+} 、 Mn^{2+} 、 Cu^{2+} 、 Mg^{+} 及其任意组合。在一些实施方案中,药物制剂可包含与金属阳离子络合的聚合物。

[0441] 在一些实施方案中,药物组合物可包含一种或多种药学上可接受的盐。本文所用的“药学上可接受的盐”是指所公开化合物的衍生物,其中通过将现有的酸或碱部分转化为其盐形式(例如,通过使游离碱基与合适的有机酸反应)来修饰母体化合物。本发明的药学上可接受的盐包括例如由无毒的无机或有机酸形成的任何母体化合物的常规无毒盐。药学上可接受的盐包括但不限于酸性残基诸如羧酸的碱盐或有机盐;以及碱性残基诸如胺的矿物盐或有机酸盐。

[0442] 在一些实施方案中,药物组合物可包含至少一种溶剂。在一些实施方案中,当水是溶剂时,溶剂化物通常称为“水合物”。

[0443] 给药途径

[0444] GIS,包括包含本文所述GIS的药物组合物可通过使GIS成功整合到受试者细胞中的任何递送途径给予。可接受的给药途径包括但不限于耳部(耳内或通过耳)、胆道灌注、颊部(朝向面颊)、心脏灌注、骶管阻滞、结膜、皮肤、牙科(对一颗或多颗牙)、牙冠内、诊断、滴耳剂、电渗透、宫颈内、鼻窦内、气管内、灌肠、肠内(进入肠道)、表皮(应用到皮肤上)、硬膜外(进入硬脑膜)、羊膜外给药、体外、滴眼剂(到结膜上)、胃肠、血液透析、浸润、吹入(鼻吸)、间质、腹内、羊膜内、动脉内(进入动脉)、关节内、胆道内、支气管内、滑囊内、心内(进入心脏)、软骨内(在软骨内)、尾内(在马尾神经内)、海绵体内注射(进入病理腔内)、腔内(进入阴茎根部)、脑内(进入大脑)、脑室内(进入脑室)、脑池内(在小脑延髓池内)、角膜内(在角膜内)、冠状动脉内(在冠状动脉内)、海绵体内(在阴茎海绵体的可扩张空间内)、皮内(进入皮肤本身内)、椎间盘内(在椎间盘内)、导管内(在腺体导管内)、十二指肠内(在十二指肠内)、硬膜内(在硬膜内或硬膜下)、表皮内(至表皮)、食管内(至食管)、胃内(在胃内)、龈内(在齿龈内)、回肠内(在小肠远端内)、病灶内(在局部病灶内或直接引入局部病灶)、管腔内(在管腔内)、淋巴内(在淋巴内)、髓内(在骨的骨髓腔内)、脑膜内(在脑膜内)、肌肉内(进入肌肉内)、心肌内(在心肌内)、眼内(在眼内)、骨内输注(进入骨髓内)、卵巢内(在卵巢内)、脑实质内(进入脑组织内)、心包内(在心包内)、腹膜内(输注或注射到腹膜内)、胸膜内(在胸膜内)、前列腺内(在前列腺内)、肺内(在肺或其支气管内)、窦内(在鼻窦或眶周窦内)、脊柱内(在脊柱内)、滑膜内(在关节滑膜腔内)、腱内(在腱内)、睾丸内(在睾丸内)、鞘内(进入椎管)、鞘内(在脑脊液轴任何层面的脑脊液内)、胸腔内(在胸腔内)、管腔内(在器官的小管内)、肿瘤内(在肿瘤内)、鼓室内(在中耳(aurus media)内)、宫内、阴道内给药、血管内(一个或多个血管内)、静脉内(进入静脉)、静脉推注、静脉滴注、脑室内(在脑室内)、膀胱内输注、玻璃体内(通过眼)、离子电渗疗法(借助于电流,其中将可溶性盐的离子迁移到身体组织内)、冲洗(浸泡或冲洗开放性伤口或体腔)、喉部(直接在喉部)、鼻腔给药(通过鼻)、鼻胃管(通过鼻并进入胃部)、神经阻滞、封闭敷裹技术(局部途径给药,然后用封闭区域的敷料覆盖)、眼部(至外眼)、口服(通过口)、口咽(直接至口和咽部)、肠胃外、经皮、关节周、硬膜外、神经周、牙周、光分离置换疗法(photopheresis)、直肠、呼吸道(在呼吸道内通过口腔或鼻腔吸入以达到局部或全身效果)、眼球后(脑桥后或眼球后)、软组织、蛛网膜下腔、结膜下、皮下(在皮肤下)、唇下、舌下、粘膜下、局部、经皮、经皮(通过完整皮肤扩散进行全身分布)、经粘膜(通过粘膜扩散)、经胎盘(通过或穿过胎盘)、经气管(通过气管壁)、经鼓室(穿过或通过鼓室)、经阴道、输尿管(至输尿管)、尿道(至尿道)、阴道和脊柱。

[0445] 在一些实施方案中,药物组合物可以以实现它们穿过血管屏障、血脑屏障或其它上皮屏障的方式给予。GIS可以以任何合适的形式给予,合适的形式包括但不限于液体溶液、悬浮液、固体形式、适合溶解在液体溶液中的固体形式、能够悬浮在液体溶液中的固体形式及其任意组合。

[0446] 在一些实施方案中,GIS可通过多位点给药途径递送至受试者。可以在2、3、4、5个或多于5个位点处给予受试者。

[0447] 在一些实施方案中,GIS可通过单一途径给药递送至受试者。

[0448] 在一些实施方案中,可使用推注输注给予受试者GIS。

[0449] 在一些实施方案中,可使用持续递送(即输注)在数分钟、数小时或数天时间段内向受试者给予GIS。输注速率可根据任何递送参数而改变,所述递送参数包括但不限于受试者的性质、期望的分布、使用的制剂等。

[0450] 在一些实施方案中,GIS可通过肌肉内递送途径递送,肌肉内递送途径包括但不限于皮下注射或静脉内注射。

[0451] 在一些实施方案中,GIS可通过口服给药递送,口服给药包括但不限于消化道给药或口腔给药。

[0452] 在一些实施方案中,GIS可通过眼内递送途径递送,眼内递送途径包括但不限于玻璃体内注射或施加滴眼剂。

[0453] 在一些实施方案中,GIS可通过鼻内递送途径递送,鼻内递送途径包括但不限于滴鼻剂或鼻喷雾剂。

[0454] 在一些实施方案中,GIS可通过外周注射给予受试者,外周注射包括但不限于肌肉内、腹膜内、静脉内、结膜或关节注射。

[0455] 在一些实施方案中,GIS可通过注射入脑脊液途径递送,注射入脑脊液途径包括但不限于鞘内和脑室内给药。

[0456] 在一些实施方案中,GIS可通过全身递送途径递送,全身递送途径包括但不限于血管内给药。

[0457] 在一些实施方案中,GIS可通过脑实质内给药给予受试者。

[0458] 在一些实施方案中,GIS可通过局部给药给予受试者。

[0459] 在一些实施方案中,GIS可通过颅内递送给予受试者。

[0460] 在一些实施方案中,GIS可通过肌内给药给予受试者。

[0461] 在一些实施方案中,GIS可通过静脉内给药给予受试者。

[0462] 在一些实施方案中,GIS可通过皮下给药给予受试者。

[0463] 在一些实施方案中,GIS可通过多于一种的给药途径递送。

[0464] 注射和肠胃外给药

[0465] 在一些实施方案中,本文所述的药物组合物可肠胃外给药。用于肠胃外和口服给药的液体剂型包括但不限于药学上可接受的溶液、乳剂、微乳剂、酞剂、悬浮液和/或糖浆。除活性成分外,液体剂型还可包含本领域常用的惰性稀释剂,例如增溶剂、水或其它溶剂和乳化剂(例如聚乙二醇、丙二醇、1,3-丁二醇、四氢糠醇、异丙醇、乙醇、碳酸乙酯、乙酸乙酯、苯甲醇、苯甲酸苄酯、二甲基甲酰胺、油、甘油和脱水山梨醇的脂肪酸酯)及其任意组合。示例性油可包括棉籽油、花生油、玉米油、胚芽油、橄榄油、蓖麻油和芝麻油及其混合物。在一些实施方案中,药物组合物包含增溶剂,诸如醇、油、二醇、**CREMOPHOR®**、改性油、聚山梨醇酯、聚合物、环糊精和/或其组合。在一些实施方案中,包含表面活性剂,诸如羟丙基纤维素。

[0466] 在一些实施方案中,注射制剂可包含无菌可注射水性或油性悬浮液。可按照已知技术使用合适的润湿剂、分散剂和/或悬浮剂配制用于注射的无菌溶液。无菌注射制剂可以是无毒的、肠胃外可接受的稀释剂和/或溶剂中的无菌可注射悬浮液、溶液和/或乳剂。在一些实施方案中,无菌注射制剂可以是1,3-丁二醇溶液。在一些实施方案中,可接受的载体和

溶剂包括但不限于林格氏溶液、U.S.P.、水、等渗氯化钠溶液和无菌不挥发油。在一些实施方案中,不挥发油可包括任何温和的不挥发油(例如合成的单甘油酯或双甘油酯)。在一些实施方案中,脂肪酸诸如油酸可用于制备可注射剂。

[0467] 在一些实施方案中,注射制剂可通过保留细菌的过滤器过滤和/或通过加入灭菌剂进行灭菌。在一些实施方案中,灭菌剂可为无菌固体组合物的形式,其可在使用前溶解或分散在无菌可注射介质诸如无菌水中。

[0468] 经常期望减缓来自皮下或肌肉注射剂的活性成分的吸收,以延长活性成分的效果。在一些实施方案中,通过将药物组合物溶解或悬浮在油载体中实现肠胃外给予的药物组合物的延迟吸收。在一些实施方案中,可通过使用水溶性差的无定形或结晶材料的液体悬浮液完成减缓活性成分的吸收。活性成分的吸收速率取决于溶解速率,而溶解速率又可取决于晶体大小和晶形。

[0469] 口服给药

[0470] 在一些实施方案中,本文所述的药物组合物和/或制剂可以口服给药。用于口服给药的固体剂型包括片剂、胶囊、粉剂、丸剂和颗粒剂。一般而言,对于固体剂型,活性成分与至少一种惰性的、药学上可接受的赋形剂混合,惰性的、药学上可接受的赋形剂包括但不限于磷酸二钙或柠檬酸钠、粘合剂(例如羧甲基纤维素、藻酸盐、明胶、聚乙烯吡咯烷酮、蔗糖和阿拉伯胶)、填充剂或增量剂(例如淀粉、乳糖、蔗糖、葡萄糖、甘露醇和硅酸)、崩解剂(例如琼脂、碳酸钙、马铃薯或木薯淀粉、藻酸、某些硅酸盐和碳酸钠)、吸收加速剂(例如季铵化合物)、保湿剂(例如甘油)、溶解减缓剂(例如石蜡)、吸收剂(例如高岭土和膨润土)、润湿剂(例如鲸蜡醇和单硬脂酸甘油酯)、润滑剂(例如滑石粉、硬脂酸钙、硬脂酸镁、固体聚乙二醇、十二烷基硫酸钠)及其任意组合。在片剂、胶囊和丸剂的情况下,剂型可包含缓冲剂。

[0471] 用于口服给药的液体剂型可包括上述用于肠胃外给药的那些。除了惰性稀释剂之外,口服组合物还可包含佐剂,诸如乳化剂、润湿剂、悬浮剂、调味剂、甜味剂和/或芳香剂。

[0472] 局部或经皮给药

[0473] 在一些实施方案中,本文所述的药物组合物和/或制剂可以配制用于局部给药。皮肤可以是递送的理想靶位点,因为它易于进入。在一些实施方案中,将本文所述的药物组合物递送至皮肤或通过皮肤的途径包括但不限于局部应用(例如,用于美容应用和/或局部/区域治疗)、皮内注射(例如,用于美容应用和/或局部/区域治疗)和全身递送(例如,用于治疗影响皮肤和皮外区域的皮肤病)。

[0474] 在一些实施方案中,本文所述的药物组合物和/或制剂可使用各种敷料绷带(例如,粘合剂绷带)或(例如,伤口敷料)递送,以有效和/或方便地实施本文所述的方法。在一些实施方案中,敷料或绷带可包含足够量的本文所述的药物组合物,以允许用户进行多种治疗。

[0475] 用于局部和/或经皮给药的剂型可包括洗剂、乳膏、软膏、凝胶、喷雾剂、糊剂、粉末、溶液、吸入剂和/或贴剂。通常,局部和/或经皮给药可通过在无菌条件下将活性成分与药学上可接受的赋形剂、缓冲剂和/或任何所需的防腐剂混合来配制。

[0476] 在一些实施方案中,可使用经皮贴剂。经皮贴剂可具有另外的优势,即提供本文所述药物组合物向身体的受控递送。一般而言,经皮贴剂可通过在适当的介质中溶解和/或分配本文所述药物组合物制备。在一些实施方案中,递送速率可通过将药物组合物分散在聚

合物基体和/或凝胶中、提供速率控制膜或其任意组合来控制。

[0477] 在一些实施方案中,适合局部给药的制剂可包括液体和/或半液体制剂(例如,搽剂和洗剂)、水包油和/或油包水乳剂(例如,软膏、乳膏和/或糊剂)、溶液和/或悬浮液,及其任意组合。

[0478] 眼部或耳部给药

[0479] 在一些实施方案中,本文所述的药物组合物可以是适合眼部给药、耳部给药或两者的制剂。一般而言,这样的制剂可以是滴眼剂和/或滴耳剂的形式,包括但不限于活性成分在水性和/或油性液体赋形剂中的溶液和/或悬浮液。在一些实施方案中,这样的滴剂可包含盐、缓冲剂、本文所述的一种或多种其它任何另外的成分及其组合。在一些实施方案中,可眼部给药的制剂包括脂质体制剂和/或微晶形式的活性成分。在一些实施方案中,药物组合物可通过视网膜下给药。

[0480] 肺部给药

[0481] 在一些实施方案中,本文所述的药物组合物可在适合肺部给药的制剂中。在一些实施方案中,肺部给药是通过颊腔进行的。在一些实施方案中,药物组合物可包含含有活性成分的干颗粒。在一些实施方案中,用于肺部给药的干颗粒的直径范围可为约0.5-7nm或约1-6nm。

[0482] 在一些实施方案中,可使用自推进溶剂/粉末分配容器给予药物组合物。一般而言,活性成分可溶解和/或悬浮在密封容器中的低沸点推进剂中。在一些实施方案中,药物组合物可为干粉形式,用于使用包括干粉储存器的装置给予,推进剂流可被引导至该干粉储存器以分散这样的粉末。在一些利用干粉的实施方案中,粉末可包含颗粒,其中按重量计至少98%的颗粒具有大于0.5nm的直径,并且按数量计至少95%的颗粒具有小于7nm的直径。在一些实施方案中,按重量计至少95%的颗粒具有大于1nm的直径,并且按数量计至少90%的颗粒具有小于6nm的直径。在一些实施方案中,包含粉末的干药物组合物可包含固体细粉稀释剂(例如,糖),并且可以以单位剂量形式提供以方便使用。

[0483] 在一些实施方案中,低沸点推进剂包括在大气压下沸点低于65°F的液体推进剂。在一些实施方案中,推进剂可占药物组合物的50%至99.9%(重量/重量),并且活性成分可占药物组合物的0.1%至20%(重量/重量)。在一些实施方案中,推进剂可包含另外的成分,其包括但不限于液体非离子表面活性剂、固体阴离子表面活性剂、固体稀释剂(包括例如具有与包含活性成分的颗粒相同数量级的粒度的固体稀释剂)及其任意组合。

[0484] 在一些实施方案中,配制用于肺部递送的药物组合物可以是溶液、悬浮液及其组合的液滴形式。当作为溶液、悬浮液或其组合制备、包装和/或出售时,可使用任何雾化和/或喷雾装置给予这样的制剂。在一些实施方案中,溶液和/或悬浮液可以是无菌的。示例性溶液和/或悬浮液包括水性和/或稀释的酒精组合物。在一些实施方案中,配制用于肺部递送的药物组合物可包含调味剂(例如,糖精钠)、挥发油、表面活性剂、缓冲剂、防腐剂(例如,甲基羟基苯甲酸酯)及其任意组合。在一些实施方案中,通过这种给药途径提供的液滴的平均直径可在约0.1nm至约200nm的范围内。

[0485] 鼻内、鼻腔或颊部给药

[0486] 在一些实施方案中,本文所述的药物组合物可通过鼻内、鼻腔或两者给予。在一些实施方案中,用于鼻内递送的药物组合物可包括本文所述的用于肺部递送的那些。在一些

实施方案中,用于鼻内给药的药物组合物包括粗粉,其平均粒径为约0.2 μm 至500 μm ,所述粗粉包含活性成分。在一些实施方案中,药物组合物可通过从保持靠近鼻的粉末容器中快速吸入通过鼻通道来给药,即以吸鼻烟的方式。示例性药物制剂可包含约0.1% (重量/重量) 至100% (重量/重量) 的活性成分,并且可包含本文所述的一种或多种另外的成分。

[0487] 在一些实施方案中,药物组合物可为适合颊部给药的制剂,其包括但不限于片剂、锭剂及其任意组合。一般而言,这种片剂或锭剂可使用常规方法制成,并且可包含0.1% - 20% (重量/重量) 活性成分(作为非限制性实例给出)、任何口服可溶解或口服可降解组合物的组合,以及任选的一种或多种本文所述的另外的成分。在一些实施方案中,适合颊部给药的药物组合物可包括粉末、气溶胶化的溶液和/或悬浮液、或雾化溶液和/或悬浮液的任意组合,其包含具有约0.1nm-200nm的分散平均粒度和/或液滴尺寸的活性成分。在一些实施方案中,用于颊部给药的药物组合物还可包含一种或多种本文所述的任何另外的成分。

[0488] 储库给药

[0489] 在一些实施方案中,本文所述的药物组合物配制在储库中用于延长释放。在一些实施方案中,本文所述的药物组合物在空间上保留在靶组织内或靶组织附近。

[0490] 可注射的储库形式通常通过在可生物降解聚合物(例如聚丙交酯-聚乙交酯)中形成药物组合物的微胶囊基质来制备。一般而言,可通过改变药物组合物与聚合物的比例以及所用的特定聚合物的性质来控制药物组合物的释放速率。合适的可生物降解聚合物包括但不限于聚(原酸酯)和聚(酸酐)。储库注射制剂是通过将药物组合物包裹在与身体组织相容的脂质体或微乳剂中来制备的。

[0491] 直肠和阴道给药

[0492] 在一些实施方案中,本文所述的药物组合物可通过直肠给药、通过阴道给药或其任意组合给药。一般而言,用于直肠或阴道给药的组合物是栓剂,其可通过将活性成分与合适的无刺激性赋形剂(例如,聚乙二醇、可可脂或栓剂蜡)混合来制备,所述赋形剂在环境温度下为固体但在体温下为液体。栓剂在直肠或阴道腔中熔化释放活性成分。

[0493] 剂量

[0494] GIS和/或包含GIS的药物组合物可以在受试者中产生期望效果(例如,期望治疗效果、研究结果等)的任何量(即剂量)给予。在一些实施方案中,可基于受试者参数(例如,受试者尺寸、状态或特点)、效果参数(例如,所要求的反应程度、治疗有效阈值、效果持续时间或存在的副作用)或其任意组合来确定期望的剂量。在一些实施方案中,可在初次给药之前确定适当的剂量,任选地基于测试至少一种受试者参数的至少一次测定确定。在一些实施方案中,可在初次给药之后确定适当的剂量,任选地基于测试至少一种效果参数的至少一次测定确定。在一些实施方案中,剂量可在整个给药过程中保持不变。在一些实施方案中,剂量可在给药过程中改变一次、两次或许多次。

[0495] 在一些实施方案中,剂量可描述为活性成分的质量与受试者的质量之比(例如,以mg/kg为单位)。例如,剂量可以为0.1至100、1至100、2至100、3至100、4至100、5至100、6至100、7至100、8至100、9至100、10至100、15至100、20至100、25至100、30至100、35至100、40至100、45至100、50至100、55至100、60至100、65至100、70至100、75至100、80至100、85至100、90至100、95至100、0.1至95、1至95、2至95、3至95、4至95、5至95、6至95、7至95、8至95、9至95、10至95、15至95、20至95、25至95、30至95、35至95、40至95、45至95、50至95、55至95、60至

95、65至95、70至95、75至95、80至95、85至95、90至95、0.1至90、1至90、2至90、3至90、4至90、5至90、6至90、7至90、8至90、9至90、10至90、15至90、20至90、25至90、30至90、35至90、40至90、45至90、50至90、55至90、60至90、65至90、70至90、75至90、80至90、85至90、0.1至85、1至85、2至85、3至85、4至85、5至85、6至85、7至85、8至85、9至85、10至85、15至85、20至85、25至85、30至85、35至85、40至85、45至85、50至85、55至85、60至85、65至85、70至85、75至85、80至85、0.1至80、1至80、2至80、3至80、4至80、5至80、6至80、7至80、8至80、9至80、10至80、15至80、20至80、25至80、30至80、35至80、40至80、45至80、50至80、55至80、60至80、65至80、70至80、75至80、0.1至75、1至75、2至75、3至75、4至75、5至75、6至75、7至75、8至75、9至75、10至75、15至75、20至75、25至75、30至75、35至75、40至75、45至75、50至75、55至75、60至75、65至75、70至75、0.1至70、1至70、2至70、3至70、4至70、5至70、6至70、7至70、8至70、9至70、10至70、15至70、20至70、25至70、30至70、35至70、40至70、45至70、50至70、55至70、60至70、65至70、0.1至65、1至65、2至65、3至65、4至65、5至65、6至65、7至65、8至65、9至65、10至65、15至65、20至65、25至65、30至65、35至65、40至65、45至65、50至65、55至65、60至65、0.1至60、1至60、2至60、3至60、4至60、5至60、6至60、7至60、8至60、9至60、10至60、15至60、20至60、25至60、30至60、35至60、40至60、45至60、50至60、55至60、0.1至55、1至55、2至55、3至55、4至55、5至55、6至55、7至55、8至55、9至55、10至55、15至55、20至55、25至55、30至55、35至55、40至55、45至55、50至55、0.1至50、1至50、2至50、3至50、4至50、5至50、6至50、7至50、8至50、9至50、10至50、15至50、20至50、25至50、30至50、35至50、40至50、45至50、0.1至45、1至45、2至45、3至45、4至45、5至45、6至45、7至45、8至45、9至45、10至45、15至45、20至45、25至45、30至45、35至45、40至45、0.1至40、1至40、2至40、3至40、4至40、5至40、6至40、7至40、8至40、9至40、10至40、15至40、20至40、25至40、30至40、35至40、0.1至35、1至35、2至35、3至35、4至35、5至35、6至35、7至35、8至35、9至35、10至35、15至35、20至35、25至35、30至35、0.1至30、1至30、2至30、3至30、4至30、5至30、6至30、7至30、8至30、9至30、10至30、15至30、20至30、25至30、0.1至25、1至25、2至25、3至25、4至25、5至25、6至25、7至25、8至25、9至25、10至25、15至25、20至25、0.1至20、1至20、2至20、3至20、4至20、5至20、6至20、7至20、8至20、9至20、10至20、15至20、0.1至15、1至15、2至15、3至15、4至15、5至15、6至15、7至15、8至15、9至15、10至15、0.1至10、1至10、2至10、3至10、4至10、5至10、6至10、7至10、8至10、9至10、0.1至9、1至9、2至9、3至9、4至9、5至9、6至9、7至9、8至9、0.1至8、1至8、2至8、3至8、4至8、5至8、6至8、7至8、0.1至7、1至7、2至7、3至7、4至7、5至7、6至7、0.1至6、1至6、2至6、3至6、4至6、5至6、0.1至5、1至5、2至5、3至5、4至5、0.1至4、1至4、2至4、3至4、0.1至3、1至3、2至3、0.1至2、1至2或0.1至1mg/kg。

[0496] 给药方案

[0497] GIS和/或包含GIS的药物组合物可以以在受试者中产生期望的效果(例如期望的治疗效果、研究结果等)的任何频率(即给药方案)给予。在一些实施方案中,给药方案可通过用于确定本文所述剂量的任何方法确定。在一些实施方案中,GIS可给予仅一次。

[0498] 在一些实施方案中,GIS可给予多于一次。例如,GIS可给予2、3、4、5、6、7、8、9、10或更多次。在一些实施方案中,GIS可在治疗受试者的治疗指征的过程中间歇和/或连续给予。在一些实施方案中,GIS可在受试者的一生中重复给予。

[0499] V.使用方法

[0500] GIS制剂递送的靶标区域、组织或细胞

[0501] 本文提供用于将本文所述的药物组合物和/或制剂递送至受试者的至少一个靶位置的方法,通过将至少一种靶点(包括一个或多个靶细胞),诸如生理系统、解剖位置、器官、组织、细胞类型、细胞群等与本文所述的至少一种药物组合物和/或制剂接触。

[0502] 本文所述的药物组合物和/或制剂包含足够的活性成分(例如,本发明的GIS),使得在位于靶点处的至少一种细胞中产生感兴趣的效果(例如,将至少一种转基因插入受试者基因组中)。

[0503] 在一些实施方案中,本文所述的药物组合物和/或制剂通常包含一种或多种细胞渗透剂,尽管也考虑了“裸”制剂(诸如没有细胞渗透剂或其它试剂),具有或不具有药学上可接受的载体。

[0504] 生理系统

[0505] 在一些实施方案中,本文所述的药物组合物和/或制剂靶向生理系统。

[0506] 在一些实施方案中,生理系统可包括听觉系统、心血管系统、中枢神经系统、化学受体系统、循环系统、消化系统、内分泌系统、排泄系统、外分泌系统、生殖系统(genital)、外皮系统、淋巴系统、肌肉系统、肌肉骨骼系统、神经系统、周围神经系统、肾脏系统、生殖系统(reproductive)、呼吸系统、泌尿系统和视觉系统。

[0507] 在一些实施方案中,本文所述的药物组合物和/或制剂靶向胺前体摄取和脱羧(APUD)系统(一系列具有内分泌功能并分泌多种小胺或多肽激素的细胞),诸如但不限于垂体组织、甲状旁腺组织、甲状腺组织、支气管组织、肾上腺髓质组织、胰腺组织、胃和肠、颈动脉体和化学受体系统组织。

[0508] 器官

[0509] 在一些实施方案中,本文所述的药物组合物和/或制剂靶向器官。器官包括肛管、动脉、升结肠、膀胱、骨髓、脑、支气管、细支气管、尿道球腺、毛细血管、盲肠、小脑、大脑半球、大脑、宫颈、脉络丛、阴蒂、脑神经、降结肠、间脑、十二指肠、耳、肠神经系统、附睾、食道、外生殖器官、输卵管、胆囊、神经节、味觉器官、肠道相关的淋巴组织、心脏、回肠、内生殖器官、间质、空肠、关节、肾脏、大肠、喉、韧带、肝脏、肺、淋巴结、淋巴管、乳腺、延髓、肠系膜、中脑、口腔、呼吸肌、鼻腔、神经、嗅觉器官、卵巢、胰腺、腮腺、阴茎、咽、胎盘、脑桥、前列腺、直肠、唾液腺、阴囊、精囊、乙状结肠、骨骼、皮肤、小肠、脊神经、脾脏、胃、皮下组织、舌下腺、颌下腺、牙齿、肌腱、睾丸、脑干、脊髓、脑室系统、胸腺、舌、扁桃体、气管、横结肠、输尿管、尿道、子宫、阴道、输精管、静脉和外阴。

[0510] 在一些实施方案中,本文所述的药物组合物和/或制剂靶向一只眼睛或两只眼睛。

[0511] 在一些实施方案中,本文所述的药物组合物和/或制剂靶向肝脏。

[0512] 在一些实施方案中,本文所述的药物组合物和/或制剂靶向脑。

[0513] 细胞

[0514] 在一些实施方案中,本文所述的药物组合物和/或制剂靶向特定细胞和/或细胞类型。

[0515] 细胞包括脂肪细胞、肾上腺素能神经细胞、 α 细胞、无长突细胞、成釉细胞、前晶状体上皮细胞、前/中间垂体细胞、顶泌汗腺细胞、星形胶质细胞、柯蒂氏器听觉内毛细胞、柯蒂氏器听觉外毛细胞、b细胞、巴托林氏腺细胞、角膜的基底细胞(干细胞)、舌、口腔、鼻腔、

远端肛管、远端尿道和远端阴道、嗅觉上皮的基底细胞、篮状细胞、嗜碱性粒细胞和前体、 β 细胞、贝茨细胞、骨髓网状组织成纤维细胞、柯蒂氏器的边界细胞、边界细胞、鲍曼氏腺细胞、棕色脂肪细胞、布伦纳氏腺细胞、尿道球腺细胞、灌木状细胞、c细胞、卡哈尔-雷茨乌斯细胞、心肌细胞、心肌细胞、车轮细胞、产生糖皮质激素的束状带细胞、产生盐皮质激素的球状带细胞、产生雄激素的网状带细胞、肾上腺皮质细胞、成牙骨质细胞、泡心细胞、耳内耳腺细胞、吊灯细胞、颈动脉体细胞的化学感受器球细胞、主细胞、胆碱能神经元、嗜铬细胞、棒状细胞(club cells)、冷敏感的初级感觉神经元、结缔组织巨噬细胞(所有类型)、角膜成纤维细胞(角膜角质细胞)、分泌孕酮的破裂的卵泡的黄体细胞、皮质毛干细胞、促皮质素细胞、含晶状体蛋白的晶状体纤维细胞、角质毛干细胞、细胞毒性T细胞、d细胞、 δ 细胞、树突状细胞、双束细胞、导管细胞,外分泌汗腺透明细胞、外分泌汗腺暗细胞、输出管细胞、弹性软骨细胞、内皮细胞、肠神经胶质细胞、肠嗜铬细胞、肠嗜铬样细胞、肠内分泌细胞、嗜酸性粒细胞及前体、室管膜细胞、表皮基底细胞、表皮朗格汉斯细胞、附睾基底细胞、附睾主细胞、上皮网状细胞、 ϵ 细胞、红细胞、纤维软骨细胞、叉状神经元、小凹细胞、g细胞、胆囊上皮细胞、生殖细胞、利特氏腺细胞(gland of litter cell)、眼睑中的莫尔腺细胞、神经胶质细胞、高尔基细胞、性腺基质细胞、促性腺激素细胞、颗粒细胞(granule cell)、颗粒细胞(granulosa cell)、颗粒黄体细胞、网格细胞和头部方向细胞。

[0516] 在一些实施方案中,细胞可以是癌细胞。在一些实施方案中,细胞可以是非癌细胞。

[0517] 在一些实施方案中,真核细胞可以是干细胞。本领域已知多种干细胞类型,其中任何一种或所有可用于本公开内容的实践。示例干细胞包括但不限于胚胎干细胞、造血干细胞、神经干细胞、表皮神经嵴干细胞、可诱导多能干细胞、乳腺干细胞、肠干细胞、间充质干细胞、嗅觉成体干细胞、睾丸细胞和祖细胞(例如,神经、成血管细胞、成骨细胞、成软骨细胞、胰腺、表皮等)。在一些实施方案中,干细胞可以是源自取自受试者的细胞的干细胞系。

[0518] 在一些实施方案中,真核细胞是人、非人灵长动物和/或其它哺乳动物,包括小鼠和/或大鼠的循环系统中发现的细胞。示例性循环系统细胞包括但不限于血小板、浆细胞、红细胞、B细胞、T细胞、自然杀伤细胞、巨噬细胞、中性粒细胞、其前体细胞等。在一些实施方案中,至少一种真核细胞可源自这些循环真核细胞中的任何一种。

[0519] 在一些实施方案中,至少一种真核细胞是自然杀伤细胞,或自然杀伤细胞的前体或祖细胞。

[0520] 在一些实施方案中,至少一种真核细胞是B细胞,或B细胞前体或祖细胞。

[0521] 在一些实施方案中,真核细胞可以是植物细胞。在一些实施方案中,植物细胞是单子叶植物或双子叶植物的细胞,所述植物包括但不限于西葫芦、木本植物诸如针叶树和落叶树、小麦、芜菁、番茄、烟草、向日葵、甘蔗、甜菜、草莓、菠菜、大豆、高粱、黑麦、水稻、覆盆子、油菜籽、萝卜、南瓜、马铃薯(包括红薯)、李子、菠萝、花生、豌豆、木瓜、燕麦、甜瓜、芒果、玉米、莴苣、扁豆、草本植物、大麻、草、花、桉树、黄瓜、棉花、咖啡、柑橘、菊苣、樱桃、芹菜、花椰菜、胡萝卜、油菜、卷心菜、西兰花、芸苔属植物、黑莓、豆类、大麦、香蕉、鳄梨、芦笋、拟南芥和其它结果植物、观赏植物、杏仁、苜蓿、多年生草、饲料作物、其它蔬菜、其它核果(例如桃、油桃、杏、梨、李子等)、其它仁果(例如苹果、梨等)、其它水果、其它球茎类蔬菜(例如大蒜、洋葱、韭菜等)、其它农作物、多年生植物部分(例如球茎;块茎;根;冠;茎;匍匐茎;分蘖;

嫩枝;插穗,包括未生根的插穗、生根的插穗、愈伤组织插穗或愈伤组织产生的小植株;顶端分生组织等)及其任意组合或杂交体。本文所用的术语“植物”是指植物的所有物理部分,包括种子、幼苗、树苗、根、块茎、茎、秆、叶和果实。

[0522] 肿瘤

[0523] 在一些实施方案中,本文所述的药物组合物和/或制剂靶向肿瘤。肿瘤可以是良性肿瘤、恶性肿瘤前肿瘤或恶性肿瘤。

[0524] 转基因的插入

[0525] 本发明提供用于将转基因引入受试者,例如人受试者的方法。在一些实施方案中,所述方法包括将有效量的至少一种本文所述的GIS引入受试者。在一些实施方案中,所述方法包括将有效量的至少一种包含转基因的GIS引入受试者。

[0526] 在一些实施方案中,所述方法可包括将转基因插入一个或多个靶插入位点。现在转到图8,其中图示说明具有插入的转基因的受试者基因组区域500。在该示例中,受试者基因组DNA包括靶插入位点120和周围的基因组DNA 110。为清楚起见,应理解靶插入位点是受试者DNA的一部分。5' 连接510标记受试者DNA和插入的转基因520之间的过渡点,在转基因的5' 端;该连接510可以具有在受试者基因组中和在模板RNA 5' 端处均存在的任何上游靶位点序列的部分或全部的重复。相反,3' 连接530标记转基因的3' 端和受试者DNA之间的过渡点;该连接530可以具有在受试者基因组和在模板RNA 3' 模块中都存在的任何下游靶位点序列的部分或全部的重复。连接510和/或530也可以含有另外的(多种)核苷酸,所述(多种)另外的核苷酸诸如可源自酶从模板-产物双链体解离之前,通过RT向尚未延伸的引物或向cDNA 3' 端的非模板核苷酸添加。

[0527] 靶插入位点

[0528] 在一些实施方案中,一个或多个靶插入位点包括安全港位点。本文所用的术语“安全港位点”是指受试者基因组中插入转基因不导致细胞功能的意外中断的位置。一般而言,如果(a)在该位点插入遗传物质不改变受试者基因的表达,或(b)在该位点插入遗传物质改变基因的表达,但该改变不改变正常的受试者细胞功能(例如,由于受试者基因组中被破坏基因的大量重复),则受试者基因组中的位点可被确定为安全港位点。作为情况(b)的非限制性实例,编码核糖体RNA(rRNA)的基因在基因组中重复的次数如此之多,使得一些rRNA基因的破坏不扰乱正常的细胞功能。

[0529] 在一些实施方案中,至少一个安全港位点和/或靶插入位点包含至少一个核糖体DNA(rDNA)序列。本文所用的术语“核糖体DNA”是指编码rRNA的任何基因。在一些实施方案中,至少一个安全港位点和/或靶插入位点包含至少一个28S rDNA序列。

[0530] 转基因

[0531] 本发明的方法和组合物可用于插入任何有效载荷序列(即,转基因),而不受有效载荷序列的长度或来源的限制。

[0532] 在一些实施方案中,转基因包括治疗活性基因。本文所用的术语“治疗活性基因”是指具有用于治疗、改善或预防至少一种治疗指征的表达产物的任何基因。

[0533] 在一些实施方案中,至少一种转基因可包括至少一种端粒酶逆转录酶(TERT)基因。在一些实施方案中,至少一种转基因可包括至少一种因子VIII短型基因。在一些实施方案中,至少一种转基因可包括至少一种苯丙氨酸羟化酶(PAH)基因。

[0534] 在一些实施方案中,至少一种转基因是报告基因。本文所用的术语“报告基因”是指具有可通过任何测定检测的表达产物的任何基因。

[0535] 在一些实施方案中,至少一个报告基因可包含或编码,但不限于,至少一种绿色荧光蛋白 (GFP)、至少一种红色荧光蛋白 (RFP)、荧光素酶 (LUC)、 β -半乳糖苷酶 (LacZ)、氯霉素乙酰转移酶 (cat) 等。

[0536] 非野生型转基因

[0537] 本领域技术人员将理解,尽管给出的转基因的许多主要实例参考天然或野生型序列,但本文公开的GIS绝不限于插入野生型或天然存在的基因或基因序列的部分。本发明的GIS可用于插入例如这样的基因,所述基因源自野生型基因、包含仅野生型基因的部分、是来自不同野生型基因的部分的组装体和/或是未知其序列存在于自然界的基因。此外,本发明的GIS可用于插入转基因,所述转基因的表达产物通常不存在于受试者细胞中和/或通常不是基因表达的结果。

[0538] 转基因调节元件

[0539] 在一些实施方案中,本发明的GIS可用于插入至少一种包含或编码至少一种调节元件的转基因。例如,转基因可被设计和/或工程化以在转基因表达产物中包含任意数量的miRNA和/或siRNA结合区。通常,miRNA和/或siRNA的包含可实现从在它们的转录组中包括互补miRNA或siRNA的细胞类型进行转基因表达的去靶向。

[0540] 在一些实施方案中,转基因可包含或编码第一表达产物和第二表达产物两者(或更多),所述第一表达产物包含或编码至少一种miRNA和/或siRNA,所述第二表达产物包含或编码与所述第一表达产物互补的至少一种miRNA和/或siRNA结合位点。不希望受理论的束缚,这可以阻止第二种表达产物的长期表达。

[0541] 抗体

[0542] 本文所用的术语“抗体”以最广泛的意义指代并具体涵盖各种实施方案,包括但不限于单克隆抗体、多克隆抗体、多特异性抗体(例如,由至少两种完整抗体形成的双特异性抗体)和抗体片段(例如,双抗体),只要它们表现出期望的生物活性(例如,“功能性”)。抗体主要是基于氨基酸的分子,其是包含至少一个源自已知或亲本抗体序列的氨基酸区域的单体或多聚体多肽。抗体可包含氨基酸基序,其招募一个或多个内源性或非天然修饰(包括但不限于添加糖部分、荧光部分、化学标签等)。出于本文的目的,“抗体”可包括重链和轻链可变域以及Fc区。

[0543] 本发明的GIS可用于插入包含或编码至少一种或多种功能性抗体的转基因。

[0544] 治疗指征的治疗

[0545] 本发明提供用于治疗或预防有其需要的受试者的至少一种治疗指征的方法。在一些实施方案中,所述方法包括向受试者引入有效量的至少一种本文所述的GIS。在一些实施方案中,所述方法包括向受试者引入有效量的至少一种包含至少一种治疗活性转基因的GIS。

[0546] 在一些实施方案中,至少一种治疗指征包括至少一种功能丧失遗传病症。在一些实施方案中,至少一种用于治疗至少一种治疗指征的方法包括给予至少一种转基因,所述转基因将受试者从功能丧失遗传病症拯救出来。本文所用的术语“拯救”是指向受试者提供至少一种组合物,所述组合物使受试者能够行使其原本缺失的天然功能。

[0547] 在一些实施方案中,至少一种方法包括通过向受试者给予有效量的包含至少一种TERT转基因的GIS来拯救受试者的端粒酶活性不足。

[0548] 在一些实施方案中,本发明的方法和组合物可用于治疗或预防受试者的端粒酶功能不足引起的病症。在一些实施方案中,至少一种方法包括向显示端粒酶活性不足的受试者给予治疗有效量的至少一种包含至少一种TERT基因的GIS。在一些实施方案中,至少一种方法包括给予治疗有效量的至少一种GIS,其包含怀疑因端粒酶活性不足而患上疾病的受试者的至少一种TERT基因。

[0549] 异源基因的调节

[0550] 本发明的GIS,包括本文所述的制剂和药物组合物,可用于调节异源基因的表达的方法。为了清楚起见,术语“异源基因”在本文中用于调节基因表达时,是指受试者基因组中除被GIS插入的基因之外的任何基因。

[0551] 一般而言,用于调节异源基因表达的方法可包括使用本发明的GIS插入其表达产物作用于另一基因的表达途径的序列。例如,插入基因的表达产物可影响异源基因转录成mRNA、异源基因mRNA翻译成多肽、异源基因mRNA在细胞质中的降解或失活的速率,或类似的任意组合。

[0552] 在一些实施方案中,本发明的至少一种GIS可用于插入转基因,所述转基因包含或编码至少一种微小RNA(miRNA)。在一些实施方案中,适合用于实施本公开内容的miRNA可包括本领域已知或尚未发现的任何miRNA。在一些实施方案中,至少一种GIS可用于插入转基因,所述转基因包含或编码至少一种人工miRNA,其中所述人工miRNA被设计成与受试者中存在的至少一种基因表达产物结合。本文所用的术语“人工miRNA”用于指其序列已被改变或设计成与期望靶序列结合的miRNA。人工miRNA可通过本领域已知的各种方法设计。

[0553] 在一些实施方案中,本发明的至少一种GIS可用于插入转基因,所述转基因包含或编码至少一种小干扰RNA(siRNA)。本文所用的术语“小干扰RNA”是指具有与靶基因的至少一部分基本相同的核苷酸序列的双链核糖核酸(dsRNA)。通常,siRNA的长度通常为21-25nt,但可以更短或更长,并通过促进靶基因的mRNA的降解来干扰(抑制)靶基因表达。任何已知或尚未发现的siRNA都可适合用于本发明。

[0554] 在一些实施方案中,本发明的至少一种GIS可用于插入转基因,所述转基因包含或编码至少一种人工siRNA。本文所用的术语“人工siRNA”是指其序列已经被设计为补充至少一种感兴趣的基因的siRNA。

[0555] 在一些实施方案中,本发明的至少一种GIS可用于插入转基因,所述转基因包含或编码至少一种转录因子(TF)。本文所用的术语“转录因子”是指结合DNA并改变或影响至少一种基因的转录的任何多肽。任何已知或尚未发现的TF都可适合用于本发明。

[0556] 本发明的GIS可用于插入转基因,该转基因包含或编码miRNA、siRNA和/或TF的任意组合。例如,至少一种GIS可用于插入转基因,所述转基因包含或编码以下任一种:至少一种miRNA和至少一种siRNA;至少一种miRNA和至少一种TF;至少一种siRNA和至少一种TF;或至少一种miRNA、至少一种siRNA和至少一种TF。

[0557] 预防性应用

[0558] 在一些实施方案中,本文所述的药物组合物和/或制剂可用于预防疾病或稳定治疗指征的进展。

[0559] 在一些实施方案中,本文所述的药物组合物和/或制剂可用作预防剂,以防止将来的治疗指征。

[0560] 在一些实施方案中,本文所述的药物组合物和/或制剂可用于阻止治疗指征的进一步进展。

[0561] 疫苗

[0562] 在一些实施方案中,本文所述的药物组合物和/或制剂可用作疫苗,和/或以类似于疫苗的方式使用。本文所用的“疫苗”是提高对特定治疗指征或感染原的免疫力的生物制剂。

[0563] 在一些实施方案中,本文所述的药物组合物和/或制剂可用作疫苗,和/或以类似于疫苗的方式用于治疗领域,诸如但不限于皮肤病学、CNS、心血管、肿瘤学、内分泌学、免疫学、呼吸和抗感染。

[0564] 抗原

[0565] 本发明的GIS可用于插入转基因,所述转基因包含或编码至少一种抗原,所述抗原可任选地由至少一种受试者细胞激发或呈现在其表面上。本文所用的术语“抗原”是指引起生物体中免疫反应的组合物。例如,引起受试者生物体产生针对所述组合物的抗体的组合物,特别是,所述抗体进而引起受试者生物体中的适应性免疫反应。抗原可以是任何免疫原性物质,包括例如多肽、蛋白、多糖、核酸、脂质等。在一些实施方案中,抗原可源自感染原,包括但不限于细菌、病毒、原生动物、真菌、朊病毒等。

[0566] 在一些实施方案中,抗原可包括感染原的部分或亚单位,例如,外壳、外壳组分、外壳蛋白、外壳多肽、表面组分、表面蛋白、表面多肽、荚膜组分、细胞壁组分、鞭毛、菌毛、毒素或类毒素。

[0567] 在一些实施方案中,本发明的至少一种GIS可用于插入包含或编码至少一种抗原的转基因,以针对至少一种治疗指征对受试者进行疫苗接种。

[0568] 研究

[0569] 在一些实施方案中,本文所述的药物组合物和/或制剂可用于诊断目的或作为本文公开的任何治疗指征的研究工具。

[0570] 在一些实施方案中,本文所述的药物组合物和/或制剂可用于任何研究实验,例如体内或体外实验。

[0571] 在一些实施方案中,本文所述的药物组合物和/或制剂可用于检测用于研究的生物标志物。

[0572] 在一些实施方案中,本文所述的药物组合物和/或制剂可用于培养的细胞。培养的细胞可来自本领域技术人员已知的任何来源,并且可作为非限制性实例,来自稳定细胞系、动物模型或人患者或对照受试者。

[0573] 在一些实施方案中,本文所述的药物组合物和/或制剂可用于在动物模型(即小鼠、大鼠、兔、猫、狗、非人灵长动物、豚鼠、果蝇、雪貂、秀丽隐杆线虫(*C.elegans*)、斑马鱼或本领域已知的用于研究目的的任何其它动物)中的体内实验。

[0574] 在一些实施方案中,本文所述的药物组合物和/或制剂可用于干细胞和/或细胞分化。

[0575] 在一些实施方案中,本文所述的药物组合物和/或制剂可用于人研究实验或人临

床试验。

[0576] 本发明提供用于对受试者进行科学和/或医学研究的方法。在一些实施方案中,所述方法包括向受试者引入有效量的至少一种本文所述的GIS。在一些实施方案中,所述方法包括向受试者引入有效量的至少一种包含至少一种报告转基因的GIS。

[0577] 单独和联合疗法

[0578] 在一些实施方案中,本文所述的药物组合物和/或制剂可用于治疗疾病的单独疗法或联合疗法。

[0579] 在一些实施方案中,本文所述的药物组合物和/或制剂可用作单独疗法。在一些实施方案中,本文所述的药物组合物和/或制剂可用于联合疗法。联合疗法可以是与一种或多种神经保护剂联合,所述神经保护剂为诸如小分子化合物、生长因子和激素,所述神经保护剂的对神经元退化的神经保护作用已经过测试。

[0580] 在一些实施方案中,本文所述的药物组合物和/或制剂可以与一种或多种其它治疗剂联合使用。“联合”并不意在暗示必须同时给予药剂和/或一起配制药剂用于递送,尽管这些递送方法属于本发明的范围。本文所述的药物组合物和/或制剂以及其它治疗剂可以与一种或多种其它期望的治疗或医疗程序同时、在其之前或之后给予。一般而言,每种药剂将按照为该药剂确定的剂量和/或时间表给予。

[0581] 可与本文所述药物组合物和/或制剂联合使用的治疗剂可以是小分子化合物,其是抗氧化剂、抗炎剂、抗凋亡剂、钙调节剂、抗谷氨酸能药物、结构蛋白抑制剂、参与肌肉功能的化合物和参与金属离子调节的化合物。

[0582] 体内GIC合成

[0583] 本发明提供用于合成GIS生物聚合物例如GIC生物聚合物的方法。在一些实施方案中,所述方法包括向受试者细胞群给予至少一种GIC合成构建体,维持细胞群足够长的时间以使至少一种GIS合成构建体由受试者细胞表达,并通过本领域已知的这样的方法收集和纯化GIS合成构建体表达产物。

[0584] 在一些实施方案中,至少一种GIC合成构建体包含或编码本发明的GIC。在一些实施方案中,至少一种GIC合成构建体包含或编码GIC和用于体内合成至少一种重组RNA的手段。这样的手段可包括提供或编码RNA聚合酶启动子、用于选择和纯化重组RNA的序列、互补GIC序列和重组后RNA生产加工信号。在一些实施方案中,至少一种GIC合成构建体以DNA质粒的形式给予,其实现通过内源性细胞机制产生编码的RNA。

[0585] 图9中图示说明示例性GIC合成构建体600。在构建体的5'端,RNAP模块610可包括任何合适的RNA聚合酶启动子(例如T7 RNAP启动子)。当存在时,任选的5'前导模块620位于RNAP模块的3'处,并且可包括改善模板5'模块折叠和自剪切,和/或允许快速去除具有免疫原性和/或使转录不稳定的5'端的GIC转录物(例如,由于RZ自剪切失败而导致的)的组件。在用作GIC之前,任何表达的5'前导模块RNA都在RZ自剪切位点630处被剪切。5'模块补充(compliment)640、模板模块补充(compliment)650和3'模块补充(compliment)660分别编码GIC5'模块、模板模块和3'模块。最后,线性化限制酶位点670可在3'端,其是通过限制酶剪切的点,提供GIC RNA的线性化并确保所有多余的载体组件保留在载体上。

[0586] VII. 列举的实施方案

[0587] 实施方案1. 一种用于基因组编辑的系统,其包括(i)至少一种逆转录酶构建体

(RTC),所述RTC包含编码具有用于逆转录多核苷酸模板的酶活性的多肽的多核苷酸,以及(ii)至少一种基因插入构建体(GIC),所述GIC包括至少一种适合由所述至少一种RTC编码的多肽进行逆转录的多核苷酸模板。

[0588] 实施方案2.实施方案1的系统,其中所述至少一种逆转录酶构建体包含至少一种生物聚合物,所述生物聚合物包括至少一种核酸、至少一种氨基酸及其任意组合。

[0589] 实施方案3.实施方案1或2中任一项的系统,其中所述至少一种逆转录酶构建体包含至少一个逆转录酶模块(RTC:RT-模块)、任选的至少一个逆转录酶构建体5'模块(RTC:5'模块)、任选的至少一个逆转录酶构建体3'模块(RTC:3'模块)及其任意组合。

[0590] 实施方案4.实施方案3的系统,其中所述至少一个逆转录酶模块包含或编码至少一种逆转录酶。

[0591] 实施方案5.实施方案3或4中任一项的系统,其中所述至少一个逆转录酶模块包含或编码至少一种源自非长末端重复(非LTR)逆转录因子的逆转录酶。

[0592] 实施方案6.实施方案4或5中任一项的系统,其中所述至少一种逆转录酶包含或编码非天然翻译起始密码子。

[0593] 实施方案7.实施方案4-6中任一项的系统,其中所述至少一种逆转录酶包括至少一个DNA结合域、至少一个RNA结合域、至少一个cDNA合成域、至少一个核酸内切酶结构域及其任意组合。

[0594] 实施方案8.实施方案7的系统,其中所述至少一个逆转录酶结构域、至少一个受试者DNA结合域、至少一个模板RNA结合域和至少一个核酸内切酶域结构及其任意组合中的至少一个源自与其它至少一个逆转录酶结构域、至少一个受试者DNA结合域、至少一个模板RNA结合域和至少一个核酸内切酶结构域中的至少一个不同的逆转录酶种类。

[0595] 实施方案9.实施方案3的系统,其中所述任选的至少一个逆转录酶构建体5'模块包含或编码至少一个RNA聚合酶启动子、至少一个5'非翻译区域(5'-UTR)、至少一个Kozak序列、至少一个5'帽及其任意组合。

[0596] 实施方案10.实施方案3的系统,其中所述任选的至少一个逆转录酶构建体3'模块包含或编码至少一个逆转录酶翻译终止密码子、至少一个3'非翻译区(3'UTR)、至少一个聚-A尾及其任意组合。

[0597] 实施方案11.实施方案1-10中任一项的系统,其中所述至少一个逆转录模块包含或编码图2-5中图示说明的至少一种结构或其任意组合。

[0598] 实施方案12.实施方案1-11中任一项的系统,其中所述至少一种逆转录酶构建体包含、编码SEQ ID NO 1-57中的至少一个及其任意组合,或由SEQ ID NO 1-57中的至少一个及其任意组合编码。

[0599] 实施方案13.实施方案1的系统,其中所述至少一种基因插入构建体包含或编码至少一种核酸生物聚合物。

[0600] 实施方案14.实施方案1或13中任一项的系统,其中所述至少一种基因插入构建体包含或编码至少一个任选的GIC:5'模块、至少一个GIC:有效载荷模块、至少一个任选的GIC:3'模块及其任意组合。

[0601] 实施方案15.实施方案14的系统,其中所述至少一个GIC:5'模块包含或编码至少一个源自天然逆转录因子5'区的序列、任选的至少一个GIC:5'模块rRNA序列、任选的至少

一个GIC:5' 模块核糖酶序列、任选的至少一个GIC:5' 模块折叠基序序列或其任意组合。

[0602] 实施方案16.实施方案15的系统,其中所述任选的至少一个GIC:5' 模块rRNA序列包含或编码1至30nt的受试者rRNA。

[0603] 实施方案17.实施方案15的系统,其中所述任选的至少一个GIC:5' 模块核糖酶序列包含或编码至少一种自剪切核糖酶,任选地其中所述自剪切核糖酶包括丁型肝炎病毒核糖酶。

[0604] 实施方案18.实施方案17的系统,其中所述任选的至少一个GIC:5' 模块核糖酶序列包含或编码源自至少一个非长末端重复逆转录因子的5' 区的核糖酶。

[0605] 实施方案19.实施方案15的系统,其中所述任选的至少一个GIC:5' 模块折叠基序序列包含或编码至少一个自主折叠RNA序列基序,任选地其中所述自主折叠RNA序列基序包含至少一个发夹基序、至少一个茎环基序、至少一个配对的茎4基序或其任意组合。

[0606] 实施方案20.实施方案14-19中任一项的系统,其中所述GIC:5' 模块包含或编码SEQ ID NO 60-154、250-276或278-279中的至少一个或其任意组合。

[0607] 实施方案21.实施方案14的系统,其中所述至少一个GIC:3' 模块包含或编码至少一个GIC:3' 模块逆转录酶识别序列、任选的至少一个GIC:3' 模块rRNA序列、任选的至少一个GIC:3' 模块A-区序列、或其任意组合。

[0608] 实施方案22.实施方案21的系统,其中所述至少一个GIC:3' 模块逆转录酶识别序列包含或编码至少一个与至少一种逆转录酶相互作用的序列。

[0609] 实施方案23.实施方案21或22中任一项的系统,其中至少一个GIC:3' 模块逆转录酶识别序列源自天然逆转录因子的3' 区。

[0610] 实施方案24.实施方案21的系统,其中所述任选的至少一个GIC:3' 模块rRNA序列包含或编码1至30nt的rRNA。

[0611] 实施方案25.实施方案21的系统,其中所述任选的至少一个GIC:3' 模块A-区序列包含或编码1至50个腺嘌呤碱基的序列。

[0612] 实施方案26.实施方案14或实施方案21-25中任一项的系统,其中所述至少一个GIC:3' 模块包含或编码SEQ ID NO 300-329中的至少一个或其任意组合。

[0613] 实施方案27.实施方案14的系统,其中所述至少一个GIC:有效载荷模块包含或编码至少一个转基因序列、任选的至少一个转基因启动子序列、任选的至少一个转基因5' 非翻译序列、任选的至少一个转基因3' 非翻译序列、任选的至少一个转基因多聚腺苷酸化信号序列、任选的至少一个转基因非编码RNA(ncRNA)加工序列,或其任意组合。

[0614] 实施方案28.实施方案27的系统,其中所述至少一个转基因序列包含或编码至少一个用于插入受试者基因组中的感兴趣的序列。

[0615] 实施方案29.实施方案27的系统,其中至少一个转基因启动子序列包含或编码至少一个促进转基因在受试者基因组中的表达的序列。

[0616] 实施方案30.实施方案27的系统,其包含至少一个转基因5' 非翻译序列,该序列包含或编码至少一个转基因mRNA 5' 非翻译区。

[0617] 实施方案31.实施方案27的系统,其中至少一个转基因3' 非翻译序列包含或编码至少一个转基因mRNA 3' 非翻译区。

[0618] 实施方案32.实施方案27的系统,其中至少一个转基因多聚腺苷酸化信号序列包

含或编码至少一种转基因多聚腺苷酸化信号。

[0619] 实施方案33. 实施方案27的系统, 其中至少一个转基因非编码RNA (ncRNA) 加工序列包含或编码至少一个转基因表达的ncRNA的至少一种终止信号、至少一种3' 加工信号及其任意组合。

[0620] 实施方案34. 实施方案14或实施方案27-33中任一项的系统, 其中至少一个GIC: 有效载荷模块包含或编码SEQ ID NO 499-525中的至少一个或其任意组合。

[0621] 实施方案35. 实施方案13-34中任一项的系统, 其中所述至少一个GIC: 5' 模块和至少一个GIC: 3' 模块中的至少一个包含或编码至少一个源自与其它至少一个GIC: 5' 模块和至少一个GIC: 3' 模块中的至少一个不同种类的非长末端重复逆转录因子的序列。

[0622] 实施方案36. 实施方案1或实施方案13-35中任一项的系统, 其中所述至少一种基因插入构建体包含或编码图6-9中图示说明的至少一种结构及其任意组合。

[0623] 实施方案37. 实施方案1或实施方案13-36中任一项的系统, 其中该系统包括: (i) 至少一种逆转录酶构建体, 其中所述至少一种逆转录酶构建体由SEQ ID NO 1-57中的至少一个组成或编码, 以及 (ii) 至少一种基因插入构建体, 其中至少一种基因插入构建体由SEQ ID NO 60-154、250-276、278-279、280-289、300-329、400-404、405-407、411-422或499-536中的至少一个序列组成或编码。

[0624] 实施方案38. 实施方案1或实施方案13-37中任一项的系统, 其包括基因插入构建体合成构建体 (GIC: 合成构建体), 其包含或编码实施方案13-37中描述的至少一种基因插入构建体。

[0625] 实施方案39. 实施方案1-38中任一项的系统, 其中至少一种逆转录酶构建体和至少一种基因插入构建体中的至少一个包含或编码至少一个源自与其它至少一种逆转录酶构建体和至少一种基因插入构建体中的至少一个不同种类的逆转录因子的序列。

[0626] 实施方案40. 实施方案1-39中任一项的系统, 其中所述用于基因组编辑的系统包含以下的至少一种组合: (i) 实施方案2-12中描述的至少一种逆转录酶构建体, 和 (ii) 实施方案13-37中描述的至少一种基因插入构建体。

[0627] 实施方案41. 一种用于将至少一个转基因插入受试者基因组的方法, 其包括给予有效量的实施方案1-40中的至少一种基因插入系统 (GIS)。

[0628] 实施方案42. 实施方案41的方法, 其中所述转基因被插入受试者基因组中的一个或多个靶位点, 任选地其中所述一个或多个靶位点包括至少一个安全港位点。

[0629] 实施方案43. 实施方案42的方法, 其中所述任选的至少一个安全港位点包含至少一个核糖体DNA (rDNA) 序列, 任选地其中所述至少一个核糖体DNA序列包括至少一个28S rDNA序列。

[0630] 实施方案44. 实施方案40-43中任一项的方法, 其包括给予用至少一种递送剂配制的至少一种基因插入系统。

[0631] 实施方案45. 实施方案44的方法, 其中所述至少一种递送剂是至少一种纳米颗粒, 任选地其中所述至少一种纳米颗粒包括至少一种脂质纳米颗粒。

[0632] 实施方案46. 一种药物组合物, 其包含至少一种实施方案1-40的基因插入系统, 以及任选的至少一种赋形剂、至少一种递送剂、至少一种佐剂及其任意组合中的至少一种。

[0633] 实施方案47. 一种治疗有其需要的受试者的治疗指征的方法, 其包括给予有效量

的至少一种实施方案1-40的基因插入系统或至少一种实施方案46的药物组合物,任选地包括至少一种实施方案41-45的方法。

[0634] 实施方案48.实施方案47的方法,其中所述治疗指征由端粒酶活性丧失引起。

[0635] 实施方案49.实施方案46或47中任一项的方法,其中所述至少一种基因插入系统包含至少一种TERT转基因。

[0636] 实施方案50.一种用于制备基因插入系统的试剂盒,其包括实施方案1-40的基因插入系统的方法,任选的实施方案46的药物组合物,并且任选地进一步包含缓冲液、DNA质粒或用于制备所述基因插入系统或药物组合物的方案。

[0637] VIII.定义

[0638] 28S rDNA:本文所用的术语“28S rDNA”是指编码真核细胞质核糖体的大亚基(LSU)的大结构性核糖体RNA(rRNA)的受试者基因组的部分。

[0639] 3' 连接:本文所用的术语“3' 连接”是指插入的序列的3' 端与受试者基因组的5' 端连接的位置。

[0640] 3' 区:本文所用的术语“3' 区”是指位于开放阅读框3' 处的逆转录因子基因的部分。

[0641] 5' 连接:本文所用的术语“5' 连接”是指受试者基因组的3' 端与插入的序列的3' 端连接的位置。

[0642] 5' 区:本文所用的术语“5' 区”是指位于开放阅读框5' 处的逆转录因子基因的部分。

[0643] 活性:本文所用的术语“活性”是指事情发生或完成的条件。本公开内容的蛋白和核酸可以具有活性,并且该活性可以涉及一个或多个生物事件。

[0644] 适应的:本文所用的术语“适应的”是指改变蛋白或氨基酸序列用以改变、添加或去除性质和/或活性。

[0645] 测定:当在本文中用作动词时,术语“测定”以其最广泛的意义使用,是指通过本领域已知的任何合适方法进行测试的行为。当在本文中用作名词时,术语“测定”是指用于确定测定对象的性质、状态和/或活性的测试。

[0646] 生物性质:本文所用的术语“生物性质”和“性质”是指生物体、生理系统、器官、组织、细胞或分子的任何特征或活性,所述特征或活性可被测量或观察。

[0647] 负荷:在递送载体的上下文中,术语“负荷”和“有效载荷”通常是指旨在用于递送至受试者细胞、组织、器官或生理系统、递送到受试者细胞、组织、器官或生理系统上或递送到受试者细胞、组织、器官或生理系统附近的任何化合物或结构(例如,本发明的GIS)。

[0648] 细胞:本文所用的术语“细胞”被赋予其最广泛的可能的含义,是指任何活的膜约束的结构。

[0649] 细胞过程:本文所用的术语“细胞过程”及其语法上等同词语是指在细胞水平上进行的任何过程,其可以限于或可以不限于单个细胞。

[0650] 特性:本文所用的术语“特性”是指通常属于人、地点或事物的特征或品质,用于识别它。术语“特性”和“性质”具有相同的含义,可以互换使用。

[0651] 赋予:本文所用的术语“赋予”及其语法上等同词语是指向受试者添加特征的过程。

[0652] 构建体:本文所用的名词“构建体”是指人工设计的生物聚合物。示例生物聚合物包括DNA、RNA和多肽。一般而言,本文所述的构建体被设计用于GIS。

[0653] 降解:本文所用的“降解”是指组合物的功能随时间而丧失。

[0654] 递送:本文所用的术语“递送”是指在活细胞或生物体中递送化合物、物质、实体、部分、负荷或有效载荷的行为或方式。除非另有说明,术语“递送”和“生物递送”可互换使用。

[0655] 递送系统:本文所用的术语“递送系统”是指任何组合物、方法或其组合,当其与本发明的GIS一起配制时,将GIS的组件递送到靶细胞的细胞质中。递送系统的非限制性实例包括由递送载体组成的系统和用于直接转染的系统。

[0656] 源自:本文所用的术语“源自”是指从特定来源分离或获得的核酸或蛋白序列,诸如非长末端重复(非LTR)逆转录转座子。该术语包括从特定来源分离或获得的天然序列。该术语还包括具有相同或相似功能性质的来自原始来源的序列的人造变体,例如,变体可包含已从原始来源修饰,以具有与原始来源分子相比改善的功能性质的核酸或氨基酸序列。

[0657] 设计的:本文所用的术语“设计的”是指组合物已从它们的天然或当前状态改变以具有新的和期望的性质和/或活性。

[0658] DNA和RNA:本文所用的术语“RNA”或“RNA分子”或“核糖核酸分子”是指核糖核苷酸的聚合物;术语“DNA”或“DNA分子”或“脱氧核糖核酸分子”是指脱氧核糖核苷酸的聚合物。DNA和RNA可以自然合成,例如分别通过DNA复制和DNA的转录而自然合成;或化学合成。DNA和RNA可以是单链的(即分别为ssRNA或ssDNA),或者多链的(例如,双链的,即分别为dsRNA和dsDNA)。本文所用的术语“mRNA”或“信使RNA”是指编码一个或多个多肽链的氨基酸序列的单链RNA。如果使用脱氧核糖核苷酸叙述RNA序列,则任何胸苷(“T”)都可以用尿苷(“U”)或尿苷类似物替换,以将DNA序列转换为RNA序列。

[0659] DNA修复:本文所用的术语“DNA修复”是指在细胞中进行的纠正对细胞的基因组的损伤的任何内源性过程。

[0660] 有效的:在涉及转基因插入时,本文所用的术语“有效的”及其语法上等同词语是指RT蛋白、GIC:5' 模块和GIC:3' 模块的给定组合在期望靶位点实现全长有效载荷模块的插入的有效性。

[0661] 元件:本文所用的术语“元件”是指分子或系统的任何离散组件,或方法的单个步骤。

[0662] 表达产物:本文所用的术语“表达产物”是指从感兴趣的序列转录的RNA(例如,mRNA)或从感兴趣的序列转录的mRNA翻译的多肽。

[0663] 封装:本文所用的术语“封装”是指封闭、包围或包裹。

[0664] 编码:本文所用的术语“编码”广义上是指任何利用聚合物大分子中的信息来指导产生不同于第一分子的第二分子的过程。第二分子可具有与第一分子的化学性质不同的化学结构。

[0665] 核酸内切酶:本文所用的术语“核酸内切酶”是指任何蛋白或蛋白的一部分,其通过分开除两个末端核苷酸之外的核苷酸来剪切多核苷酸链。

[0666] 外泌体:本文所用的“外泌体”是哺乳动物细胞分泌的囊泡或参与RNA降解的复合物。

[0667] 离体:术语“离体”是指从供体受试者中取出细胞,使用本文所述方法修饰细胞,并将细胞加回至受体受试者。该术语包括从同一个体受试者获得的自体细胞(即,同一受试者既是未修饰细胞的供体又是离体修饰细胞的受体),以及从与受体受试者不同个体的供体受试者获得的同种异体细胞。同种异体供体和受体可以是HLA匹配的。

[0668] 促进:本文所用的术语“促进”以其最广泛的含义使用,是指通过添加指定的元件使得动作或过程更有可能发生。

[0669] 保真度:本文所用的术语“保真度”是指将感兴趣的基因插入到受试者基因组中的准确度。术语“高保真度”对应于感兴趣的基因被插入时在核苷酸一致性(identity)、序列长度和靶位点位置方面有相对较少数量的错误。例如,如果模板RNA含有大约5000个核苷酸并且可以被RT蛋白复制以产生cDNA而不产生碱基对错配,则基因插入具有高保真度。根据转基因插入的目的,可以发生有限数量的错配,并且保真度仍然足够高的以产生功能性转基因。

[0670] 侧翼:本文所用的术语“侧翼”是指将一个元件定位在另一个元件的5' (5' 侧翼)或3' (3' 侧翼)。被称为在侧翼的元件可以直接彼此连接,或者可以有穿插在它们之间的其它元件。

[0671] 制剂:本文所用的“制剂”包含至少一种本文所述的GIS组件,以及至少一种递送剂、药学上可接受的赋形剂或两者。

[0672] 功能性/有活性的:关于生物分子,本文所用的术语“功能性”是指其表现出作为其特征的性质和/或活性的形式的生物分子。

[0673] 基因:本文所用的术语“基因”以其最广泛的意义用于指形成或可能形成染色体一部分的不同核苷酸序列,其顺序决定多肽或核酸分子中单体的顺序。

[0674] 基因插入构建体:本文所用的术语“基因插入构建体”或GIC是指包含RT蛋白的RNA模板的RNA构建体。

[0675] 基因插入系统:本文所用的术语“基因插入系统”或“GIS”是可用于通过逆转录,包括TPRT将基因序列(转基因)插入受试者基因组的特定位置的组件(模块)的系统。

[0676] GIC:3' 模块:本文所用的术语“3' 模块”是指GIC的一部分,其包含至少一个源自或功能性地取代逆转录因子基因的3' 区的元件。

[0677] GIC:5' 模块:本文所用的术语“GIC:5' 模块”是指促进全长转基因插入的GIC的一部分,其可以源自或可以不源自逆转录因子基因的5' 区。

[0678] 生成:本文所用的动词“生成”及其变位以其最广泛的意义上用于指导致指定产物存在的任何过程。

[0679] 基因组:本文所用的术语“基因组”以其最广泛的意义用于指细胞中存在的所有遗传物质。

[0680] HDV RZ折叠:本文所用的术语“HDV RZ折叠”是指可以采用丁型肝炎病毒(HDV)核糖酶的折叠并保留核糖酶功能的任何RNA序列。

[0681] 异源的:本文所用的术语“异源的”是指任何基因或蛋白序列或结构,其被放入通常不产生该基因或蛋白序列或结构的细胞中。该术语还包括本公开内容的RTC或GIC的单个元件、模块或部分,其包含来自不同物种的核酸(DNA或RNA)序列或氨基酸序列。例如,RTC或GIC的5' 模块可包含来自一种(或第一)鸟类物种的序列,并且相同RTC或GIC的3' 模块可包

含来自不同(或第二)鸟类物种的序列。

[0682] 同源重组:本文所用的术语“同源重组”是指依赖于转基因和受试者基因组之间的序列同源性的任何转基因插入的过程。

[0683] 体外:本文所用的术语“体外”用于指在活细胞或生物体外进行的反应或过程。

[0684] 体内:本文所用的术语“体内”用于指在活细胞或生物体内部或表面上进行的反应或过程。

[0685] 无活性的:关于生物分子,本文所用的术语“无活性的”是指不表现出作为其特征的性质和/或活性的形式的生物分子。

[0686] 无活性成分:本文所用的术语“无活性成分”是指一种或多种对制剂中包含的药物组合物的活性成分的活性没有贡献的试剂。在一些实施方案中,可用于本发明的制剂的所有或一些无活性成分可以由美国食品药品监督管理局(FDA)批准,或可用于本发明的制剂的无活性成分可以由美国食品药品监督管理局(FDA)批准。

[0687] 诱导:本文所用的术语“诱导”及其语法上等同词汇是指导致所述结果,而对过程的步骤没有任何具体限制的过程。

[0688] 引入:本文所用的术语“引入”是指将遗传物质,经常是DNA添加到细胞中。

[0689] 插入:本文所用的术语“插入”是指将核苷酸添加到DNA序列中。

[0690] 连接:本文所用的术语“连接”是指受试者基因组中受试者的插入位点DNA与插入的转基因的cDNA连接的位置。

[0691] 至少一:本文所用的术语“至少一”是指一、二、三、四、五或更多个修饰的对象,例如,本公开内容的构建体、模块或序列。

[0692] 脂质纳米颗粒:本文所用的“脂质纳米颗粒”或“LNP”是指包含一种或多种脂质(例如阳离子脂质、非阳离子脂质、PEG修饰的脂质)的递送载体。

[0693] 脂质体:本文所用的“脂质体”通常指由布置成一个或多个球形双层或双层的脂质(例如两亲性脂质)组成的囊泡。

[0694] 功能丧失:本文所用的术语“功能丧失”指受试者基因中的任何变化,其导致缺乏野生型基因的功能的改变的基因产物。

[0695] 修饰的:本文所用的“修饰的”指分子的改变的状态或结构。分子可以以多种方式进行修饰,该多种方式包括化学、结构和功能方式。

[0696] 模块化系统:本文所用的“模块化系统”指可以被分成多组的强烈相互作用的部分的系统,所述部分是彼此相对自主的。

[0697] 基序:本文所用的术语“基序”是指具有可识别结构的生物聚合物的任何序列,该结构可以由或可以不由独特的化学或生物功能定义。

[0698] 天然的:本文所用的术语“天然的”是指野生型或自然存在的化合物、生物分子(例如,蛋白或核酸)或组合物。

[0699] 非LTR逆转录因子逆转录酶:本文所用的术语“非LTR逆转录因子逆转录酶(RT)”是指源自非LTR逆转录因子的具有逆转录活性的蛋白。

[0700] 非LTR逆转录因子:本文所用的术语“非LTR逆转录因子”是指一类不含有长末端重复的逆转录因子基因(又名逆转录转座子)。

[0701] 外部:关于插入位点,本文所用的术语“外部”是指距离插入位点的5' 或3' 超过约

60bp的基因组的任何部分。

[0702] 配对的RT:本文所用的术语“配对的RT”是指逆转录酶(RT)与至少一个包含插入有效载荷模块的模块的组合。模块可以与其配对的RT同源,这意味着RT和模块中的所有元件源自相同的逆转录因子基因。模块可以与其配对的RT异源,意味着模块的至少一个元件不是源自与RT相同的逆转录因子基因。

[0703] 有效载荷:除了在递送载体的上下文中使用时,术语“有效载荷”可以指包含在旨在插入受试者基因组的基因插入系统(GIS)中的任何核酸序列(例如,感兴趣的基因)。

[0704] 同源性百分比:术语“同源性百分比”或“%同源性”是指两个核酸或氨基酸序列之间一致或相同的序列的量。术语“同源性百分比”可以与本文定义的术语“同一性百分比”或“序列同一性百分比”互换使用。

[0705] 本文所用的“同一性百分比”或“序列同一性百分比”或“同源性百分比”是通过在比较窗口上比较两个最佳比对的序列来确定的,其中与参考序列(其不包含添加或缺失)相比,比较窗口中的序列部分可以包含添加或缺失(即间隙),用于实现两个序列的最佳比对。百分比计算如下:确定两个序列中出现相同核酸碱基或氨基酸残基的位置数以得到匹配的位置数,将匹配的位置的数量除以比较窗口中的总位置数,并将结果乘以100以得到序列同一性百分比。

[0706] 在两个或多个核酸或多肽序列的上下文中,术语“相同的”、“同一性”或“同源性”是指两个或更多个相同的序列或子序列。当在比较窗口或指定区域内进行比较和最大对应性比对时,如使用以下序列比较算法之一或通过手动比对和目视检查测量的,如果序列具有指定百分比的相同核苷酸或氨基酸残基(例如,在指定区域内至少20%、至少25%、至少30%、至少35%、至少40%、至少45%、至少50%、至少55%、至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%同一性、至少96%同一性、至少97%同一性、至少98%同一性、至少99%同一性、至少99.5%同一性或至少99.9%同一性),则序列彼此“实质上相同”。因此,除非另有说明,本文提供的所有核酸和氨基酸序列都包含与参考序列实质上相同的序列。

[0707] 对于序列比较,通常一个序列充当参考序列,将测试序列与其进行比较。当使用序列比较算法时,将测试序列和参考序列输入计算机,如果必需,则指定子序列坐标,并指定序列算法程序参数。通常使用默认程序参数,或者可以指定替代参数。然后,序列比较算法基于程序参数计算测试序列相对于参考序列的序列同一性或相似性百分比。

[0708] 适合用于确定序列同一性和序列相似性百分比的算法是BLAST和BLAST 2.0算法,它们分别在Altschul等人(Nuc.Acids Res.25:3389-402,1977)和Altschul等人(J.Mol.Biol.215:403-10,1990)中进行了描述。执行BLAST分析的软件可通过美国国家生物技术信息中心(the National Center for Biotechnology Information)(<http://www.ncbi.nlm.nih.gov/>)公开获得。该算法涉及首先通过识别查询序列中长度为W的短词来识别高得分序列对(HSP),所述短词在与数据库序列中相同长度的词对齐时匹配或满足一些正值阈值得分T。T称为邻域词得分阈值(Altschul等人,上文)。这些初始邻域词命中充当启动搜索以查找包含它们的更长HSP的种子。词命中沿着每个序列在两个方向上扩展,累积比对得分尽可能地增加。对于核苷酸序列,使用参数M(一对匹配残基的奖励分数;始终>0)和N(错配残基的惩罚分数;始终<0)来计算累积得分。对于氨基酸序列,使用评分矩阵来

计算累积得分。在以下情况时,每个方向的词命中扩展停止:累积比对得分从其最大实现值下降量X;由于累积一个或多个负得分残基比对,累积得分变为零或以下;或者到达任一序列的末端。BLAST算法参数W、T和X决定比对的灵敏度和速度。BLASTN程序(用于核苷酸序列)默认使用字长(W) 11、期望值(E) 10、M=5、N=-4和两条链的比较。对于氨基酸序列,BLASTP程序默认使用字长3、期望值(E) 10,以及BLOSUM62评分矩阵(参见Henikoff和Henikoff, Proc.Natl.Acad.Sci.USA 89:10915,1989),比对值(B) 50,期望值(E) 10,M=5,N=-4。

[0709] BLAST算法还对两个序列之间的相似性进行统计分析(参见,例如,Karlin和Altschul,Proc.Natl.Acad.Sci.USA 90:5873-87,1993)。BLAST算法提供的一个相似性度量是最小总概率(P(N)),其提供两个核苷酸或氨基酸序列之间将偶然发生匹配的概率的指示。例如,如果测试核酸与参考核酸的比较中最小总概率小于约0.2,通常小于约0.01,更通常小于约0.001,则认为核酸与参考序列相似。

[0710] 肽:本文所用的“肽”是指长度小于或等于50个氨基酸的氨基酸链(chain)或链(strand),例如长度为约5、10、15、20、25、30、35、40、45或50个氨基酸的氨基酸链或链。

[0711] 药物组合物:本文所用的术语“药物组合物”是指包含至少一种活性成分和任选的一种或多种药学上可接受的赋形剂的组合物。

[0712] 聚腺苷酸:本文所用的术语“聚腺苷酸”是指任意长度的腺苷核苷酸序列。

[0713] 聚腺苷酸尾:本文所用的术语“聚腺苷酸尾”或“聚-A尾”用于指长度为约80个或更多个核苷酸的腺苷核苷酸序列。

[0714] 聚腺苷酸区:本文所用的术语“聚腺苷酸区”、“聚A-区”和“A-区”(均缩写为PA)是等效的并且可互换使用,以指长度为约1-50个核苷酸的腺苷核苷酸序列。

[0715] 启动子:本文所用的术语“启动子”是指启动转录的蛋白所结合的任何DNA序列。

[0716] 前蛋白:本文所用的术语“蛋白前体”、“前蛋白”和“前肽”指可通过翻译后修饰转变为活性形式的无活性蛋白。

[0717] 保护:本文所用的术语“保护”及其语法上等同词语是指防止生物聚合物的全部或部分降解的任何组合物或过程。

[0718] 蛋白:本文所用的“蛋白”用于指长度超过50个氨基酸的氨基酸生物聚合物。本文所述蛋白的非限制性实例是酶、逆转录酶和核酸内切酶。

[0719] 区域:本文所用的术语“区域”是指核苷酸或氨基酸序列的一部分。区域可以是未知长度的或不确定长度的,在这种情况下,它由其所涉及的功能或其相对于序列中其它元件的位置指定。

[0720] 逆转录因子/逆转录转座子:本文所用的术语“逆转录因子”和“逆转录转座子”可互换地指一类能够通过RNA中间体复制到它们自身基因组内的新位置的真核基因。

[0721] 逆转录酶:本文所用的术语“逆转录酶”是指能够从RNA模板序列合成cDNA的任何蛋白。

[0722] 逆转录酶构建体:本文所用的术语“逆转录酶构建体”(RTC),如前所述,是指包含或编码至少一种RT的生物聚合物构建体。

[0723] RTC:RT模块:本文所用的术语“RTC:RT模块”或“逆转录酶模块”是指包含或编码至少一种RT的生物聚合物构建体。

[0724] 核糖体DNA:本文所用的术语“核糖体DNA(rDNA)”是指编码由RNAP I合成的前体核

糖体RNA的受试者基因组的部分。

[0725] 核糖体RNA:本文所用的术语“核糖体RNA(rRNA)”是指核糖体的非编码RNA组件。

[0726] 片段:本文所用的术语“片段”是指序列的一部分。例如,核苷酸序列的片段可包含小于其全长的基因的任何部分。

[0727] 选择性的:本文所用的术语“选择性的”和“选择性”是指倾向于结合至非常有限种类的其它分子、其它分子的结构、蛋白或遗传序列的分子,包括但不限于酶、酶蛋白和基因。

[0728] 自剪切核糖酶:本文所用的术语“自剪切核糖酶”是指一类催化序列特异性分子内(或分子间)剪切的RNA。

[0729] 选择性:本文所用的“选择性”是指RT有效利用异源配对的GIC 5'或3'模块的可能性。

[0730] 序列:本文所用的术语“序列”是指从N末端到C末端给出的氨基酸顺序,或生物聚合物的5'到3'给出的核苷酸顺序。

[0731] 位点特异性的:本文所用的短语“位点特异性的”是指基因座,例如约60bp序列的基因座。

[0732] 稳定性:本文所用的术语“稳定性”是指组合物随时间保持其性质的能力。

[0733] 成功的TPRT:本文所用的短语“成功的TPRT”是指使用通过靶位点形成切口制成的引物合成cDNA和/或插入转基因。

[0734] 适合:本文所用的术语“适合”是指有效的、能工作的或适于特定目的或用途的任何事物。

[0735] 合成的:本文所用的术语“合成的”是指通过人工生产、制备和/或制造的任何事物。本发明的多核苷酸或多肽或其它分子的合成可以是化学合成或酶合成。

[0736] 靶细胞:本文所用的短语“靶细胞”是指任何一个或多个感兴趣的细胞。细胞可以存在于生物体的体外、体内、原位或在组织或器官中。生物体可以是动物,优选哺乳动物,更优选人,最优选患者。

[0737] 靶点引发的逆转录:本文所用的术语“靶点引发的逆转录”是指逆转录酶使用靶位点处基因组内嵌的带切口的DNA 3'端作为引物来启动cDNA合成的任何过程。

[0738] 模板:本文所用的术语“模板”和“RNA模板”是指通过RT转录成cDNA的RNA序列。

[0739] 模板末端:本文所用的术语“模板末端”是指RNA模板的5'端或3'端。

[0740] 有治疗活性的:本文所用的术语“有治疗活性的”是指治疗或缓解受试者的治疗指征的基因或基因产物。

[0741] 转录:本文所用的术语“转录”是指RNA聚合酶使用DNA分子作为模板形成或合成RNA分子。

[0742] 转染:本文所用的术语“转染”是指将外源核酸引入细胞的方法。转染方法包括但不限于化学方法、物理处理和阳离子脂质或混合物。

[0743] 转基因:本文所用的术语“转基因”是指插入受试者基因组的任何基因。

[0744] 翻译:本文所用的术语“翻译”是指核糖体基于RNA模板形成多肽分子。

[0745] 治疗和预防:本文所用的术语“治疗”或“预防”以及由此衍生的词语不一定要求是100%或完全的治疗或预防。而是存在不同程度的治疗或预防,本领域技术人员认识到所述治疗或预防具有潜在的益处或治疗效果。此外,“预防”可包括延迟疾病、症状或病症的发

作。

[0746] 未修饰的:本文所用的术语“未修饰的”是指在以任何方式被改变之前的任何物质、化合物或分子。未修饰的可以但并非总是指生物分子的野生型或天然形式。分子可以经历一系列修饰,其中每个修饰的分子可以用作用于后续修饰的“未修饰的”起始分子。

[0747] 载体:本文所用的术语“载体”是转运、转导或以其它方式充当异源分子的载体的任何分子或部分。

[0748] IX. 等效物和范围

[0749] 本领域技术人员将认识到或能够仅使用常规实验来确定根据本文所述公开内容的具体实施方案的许多等效方案。本发明的范围并非旨在限于上述描述,而是如所附权利要求中所述。

[0750] 在权利要求中,除非有相反的指示或另外从上下文中明显可见,否则冠词诸如“(a)”、“一(an)”和“所述(the)”可指一或多于一。除非有相反的指示或另外从上下文中明显可见,否则如果一个、多于一个或所有组成员存在于、用于给定产品或方法或以其它方式与给定产品或方法相关,则在一个或多个组成员之间包括“或”的权利要求或描述被视为已满足要求。本公开内容包括其中组中的恰好一个成员存在于、用于给定产品或方法或以其它方式与给定产品或方法相关的实施方案。本公开内容包括其中多于一个或整个组成员存在于、用于给定产品或方法或以其它方式与给定产品或方法相关的实施方案。

[0751] 还应注意的是,术语“包含”旨在是开放的,并且允许但不要求包括另外的要素或步骤。当本文使用术语“包含”时,术语“由……组成”因此也被涵盖和公开。

[0752] 给出范围时,包括端点。此外,应理解,除非另有指示或另外从上下文和本领域技术人员的理解明显可见,否则以范围表示的值可采取本公开内容的不同实施方案中所述范围内的任何具体值或子范围,直至范围下限单位的十分之一,除非上下文另有明确规定。

[0753] 另外,应理解,本发明的落入现有技术的任何具体实施方案都可以明确排除在任何一项或多项权利要求之外。由于这样的实施方案被视为本领域技术人员已知的,因此即使本文未明确阐述所述排除,也可以将它们排除。本公开内容的组合物的任何具体实施方案(例如,任何抗生素、治疗或活性成分;任何生产方法;任何使用方法;等)都可以出于任何原因从任何一项或多项权利要求中排除,无论是否与现有技术的存在有关。

[0754] 应当理解,所使用的词语是描述性词语,而不是限制性词语,并且可以在所附权利要求的范围内进行改变,而不背离本公开内容在其更广泛方面的真正范围和精神。

[0755] 尽管已经就几个描述的实施方案以一定长度和以一定详细程度描述本发明,但这并非意味着本发明应局限于任何这样的细节或实施方案或任何具体实施方案,而是其应参考所附权利要求进行解释,以根据现有技术提供对这样的权利要求的最广泛的可能解释,从而有效地涵盖本公开内容的预期范围。

[0756] 通过以下非限制性实施例进一步说明本发明。

[0757] X. 实施例

[0758] 实施例1. 基因插入构建体(GIC)体外转录(IVT) - 无有效载荷的GIC

[0759] 小于约1000nt的GIC RNA生物聚合物,诸如用于采用体外纯化RT的TPRT测定的RNA,通常如下通过体外RNA转录(IVT)反应制备。

[0760] 使用Q5 DNA聚合酶(NEB)通过PCR生成用于RNA转录的GIC DNA模板,并通过柱净化

(column clean-up) (Bio Basic) 纯化。

[0761] 使用T7 RNA聚合酶(RNAP)通过产生等效纯化RNA的两种方案之一进行IVT反应。第一种方法使用纯化的反应组分,于25uL反应溶液中转录1ug的DNA模板,该反应溶液含有40mM Tris pH7.9、2.5mM亚精胺、26mM MgCl₂、0.01% Triton X-100、约30mM DTT、8mM GTP、4mM所有其它rNTP、0.5uL RiboLock(Thermo Scientific)、0.5uL无机焦磷酸酶(NEB)、0.5uL T7 RNAP(在细菌中过表达后纯化,并以50mg/mL储存在20mM KPO₄ pH7.5、100mM NaCl、50%甘油、10mM DTT、0.1mM EDTA、0.2%NaN₃中)。反应在37℃下孵育3-4小时,然后加入1uL DNase RQ1(Promega)、1.5uL 20mM CaCl₂和2uL H₂O。第二种方法按照制造商的说明使用NEB HiScribe T7 Kit,每20uL反应溶液使用1ug消化的质粒。反应在37℃下孵育2小时,然后加入1uL DNase RQ1(Promega)、1.5uL 20mM CaCl₂和2uL H₂O。

[0762] 然后按照本领域已知的常规程序通过脱盐(罗氏迷你快速旋转柱)、有机萃取和沉淀纯化产物RNA。

[0763] 实施例2基因插入构建体(GIC)体外转录(IVT)-具有转基因有效载荷的GIC

[0764] 含有转基因表达盒有效载荷的GIC RNA生物聚合物如下通过体外RNA转录(IVT)反应制备。

[0765] 将GIC DNA转录模板序列克隆到pUC57-mini骨架(SEQ ID NO 350)中,T7 RNAP启动子在预期的GIC RNA模板的上游,BbsI位点在预期的GIC RNA模板的下游。纯化的质粒DNA通过在37℃下用BbsI-HF(NEB)消化4小时来线性化。然后,消化的质粒通过Qiagen PCR纯化柱纯化并在无核酸酶的水中洗脱。

[0766] IVT反应采用NEB HiScribe T7 Kit进行,每20uL反应溶液用1ug消化的质粒。具体而言,每个IVT反应有2uL各种rNTP、2uL的10X缓冲液、2uL的T7聚合酶混合物、1ug消化的质粒和ddH₂O,在37℃下孵育2小时。

[0767] IVT后,通过无核糖核酸酶的脱氧核糖核酸酶I在37℃下处理30分钟去除DNA模板。接下来,通过加入等体积的25:24:1的苯酚:氯仿:异戊醇,pH6.7(PCI),剧烈涡旋,离心,用10%体积的3M乙酸钠(pH5)和3体积的100%乙醇使水层沉淀,纯化合成的RNA。在70%乙醇中洗涤三次后,将RNA团粒风干并溶解在pH6.5的1mM柠檬酸钠中。

[0768] 实施例3.用于TPRT测定的逆转录酶(RT)蛋白制备

[0769] RT蛋白通过人细胞中的瞬时表达产生,并如下纯化。

[0770] 将编码所示RT(GenScript)的密码子优化的ORF克隆到pcDNA3.1N-DYK质粒(GenScript)的Kpn I和XbaI位点之间,以与载体编码的N端FLAG标签(SEQ ID NO.351)融合。KpnI位点在FLAG标签和RT氨基酸序列之间添加甘氨酸-苏氨酸接头。XbaI位点位于3' UTR起始处附近的(多个)翻译终止密码子之后。使用Lipofectamine 3000(Invitrogen)反向转染12ug质粒DNA。首先,将DNA与500uL的OPTI-MEM和24uL的P3000轻柔混合。然后将500uL的OPTI-MEM和24uL的Lipofectamine混合在一起并添加到DNA混合物中。将Lipofectamine/DNA复合物在室温下孵育10分钟,并添加到如下制备的细胞中。简言之,对于每次转染,将1个10cm培养皿中的80%汇合的HEK 293T细胞(以下称293T)分到Lipofectamine/DNA复合物上,并80%汇合度时重新铺板(replated)。18-24小时后,将细胞胰蛋白酶化以将它们从平板中移出,在5mL培养基中重悬,并在15mL锥形管中以约2000g低速离心3分钟。用含有1mM PMSF的PBS清洗团粒,将其转移到1.5mL管中,并在4℃下以2000g

重新成团1分钟。

[0771] 将细胞团悬浮在4x团粒体积的1X低渗裂解缓冲液[HLB;20mM HEPES (pH8)、2mM $MgCl_2$ 、200uM EGTA、10%甘油、1mM DTT、0.2%丝氨酸蛋白酶抑制剂混合物(SPIC,Sigma)、1mM PMSF]中并置于冰上5分钟以使细胞膨胀。然后通过3次循环的在液氮中快速冷冻样品并在室温水浴中解冻来裂解细胞。然后将样品移至400mM NaCl,轻柔涡旋,并再置于冰上5分钟。然后将样品在4℃下以17000g离心5分钟。收集上清液并通过添加等体积的含有0.2% NP-40的1X HLB将NaCl的浓度降低至200mM并将NP-40的浓度升高至0.1%。将样品轻柔涡旋并在4℃下以17000g离心10分钟。

[0772] 将澄清的上清液收集到新管中,并添加20uL封闭和平衡的FLAG抗体树脂(Sigma)。将样品在4℃下旋转2小时以将蛋白免疫沉淀。然后用IP缓冲液(1X HLB、200mM NaCl、0.1% NP-40)总共洗涤FLAG树脂4次(2次快速,2次在4℃下旋转5分钟)。最后一次洗涤后,用30G针去除所有缓冲液,并将树脂重悬在40uL IP缓冲液中。通过添加50ng/uL三FLAG肽(Sigma)并在室温下孵育1小时来部分洗脱蛋白。洗脱的蛋白在液氮中快速冷冻并储存在-80℃下以供后续使用。

[0773] 实施例4.RTC mRNA产生

[0774] RNA(mRNA)RTC生物聚合物如下制备。

[0775] 通过PCR扩增编码RT (GenScript)的密码子优化的ORF,以在ORF之前附加BamHI位点,并在终止ORF的终止密码子之后附加XhoI位点。BamHI位点位于N端FLAG标签和RT ORF之间的框架内,并且它在该连接处添加甘氨酸-丝氨酸接头。

[0776] 将RT ORF克隆到pUC57-mini (SEQ ID NO 350)中的5' UTR (SEQ ID NO 58)和3' UTR和模板编码的聚腺苷酸尾 (SEQ ID NO 59)之间,T7 RNAP启动子序列位于上游,BbsI位点位于下游。然后用BbsI将mRNA转录模板质粒线性化,并按照实施例2所述将其重新纯化。AG Clean cap mRNA合成和使用二氧化硅膜的纯化由商业供应商 (TriLink) 进行,或使用TriLink试剂和方案进行,通常使用5-甲氧基尿苷核糖核苷三磷酸 (5moU) 100%替换尿苷核糖核苷三磷酸 (U)。将通过5moU与N1-甲基假尿苷的100%尿苷替换进行比较,表明具有任一种修饰的核苷酸的mRNA的功能相当。

[0777] 实施例5.体外RT活性筛选

[0778] 使用退火为RNA模板的DNA引物,如下在体外测试候选蛋白的逆转录酶活性,这是领域标准RT测定。

[0779] 按照实施例3制备RT蛋白。引物DNA寡核苷酸 (SEQ ID NO 352购自IDT) 和模板RNA (SEQ ID NO 353) 通过实施例1的第一方案生成。

[0780] 对于每个筛选反应,通过将样品加热至65℃3分钟并将样品放在冰上至少5分钟,使2uL的8uM DNA寡核苷酸和2uL的4uM模板RNA退火。

[0781] 形成非放射性主混合物,其含有以下成分:2uL的10X RT缓冲液 (50mM $MgCl_2$ 、250mM Tris (pH7.5) 和750mM KCl)、2uL的100mM DTT、2uL的20%PEG-6K和5uL的无核酸酶 H_2O 。

[0782] 还形成放射性主混合物,其含有以下成分:1uL的10mM dA、dC和dTTP;1uL的2mM dGTP;4uL的上述退火的DNA-RNA和1uL的 $^{32}P\alpha$ -dGTP (Perkin Elmer)。

[0783] 对于每个反应,将11uL的非放射性主混合物、2uL的候选RT蛋白和7uL的放射性主

混合物混合,使每个反应体积达到20 μ L。使反应在37℃下进行30分钟,然后在70℃下热灭活5分钟。将80 μ L的终止溶液(50mM Tris (pH7.5)、20mM EDTA和0.2% SDS)加入反应中,该终止溶液含有100nt寡核苷酸(SEQ ID NO 290),该寡核苷酸先前使用 γ ³²P ATP和T4多核苷酸激酶(NEB)进行5'端放射性标记,然后通过PCI提取纯化和浓缩DNA,随后进行乙醇沉淀(干冰乙醇浴)。DNA在台式离心机中以14000g成团20分钟,用75%乙醇洗涤一次,风干,并重悬于5 μ L H₂O+5 μ L 2x甲酰胺上样缓冲液中。

[0784] 使样品在9%尿素-PAGE变性凝胶上行进,干燥,在磷光成像仪屏幕上曝光,并于第二天在Typhoon Trio成像系统上成像。

[0785] 实施例6. 体外TPRT活性测定

[0786] 按照实施例3制备RT蛋白。TPRT的模板RNA通过如实施例1所述的IVT反应制备。RT蛋白和模板RNA与靶位点寡核苷酸双链结合,双链DNA长度为64或84bp(分别为SEQ ID NO.291和SEQ ID NO.292),其中底链使用 γ ³²P ATP和T4多核苷酸激酶(NEB)在镁反应缓冲液中在37℃下进行5'-末端放射性标记30分钟。通过变性PAGE来解析产物,并使用Typhoon Trio成像系统对凝胶进行成像。

[0787] 实施例7. 细胞培养和基于RNA的RTC和GIC的共转染

[0788] 将所示的哺乳动物细胞系在转染前立即铺到6孔板上,密度为每孔125-250万个细胞。

[0789] 在125 μ L的Opti-MEM中稀释5 μ L的Messenger Max并孵育10分钟。

[0790] 将RTC mRNA和GIC RNA(分别按照实施例4和2制备)以指定的摩尔比混合,然后在125 μ L的Opti-MEM中稀释。然后将Opti-MEM溶液中的Messenger Max和Opti-MEM溶液中的GIS RNA充分混合并在室温下孵育5分钟。

[0791] 将所得混合物逐滴加入6孔板中的一个孔的细胞中,将板放回细胞孵育箱,等待足够时间后再分析细胞。

[0792] 实施例8. FACS分析

[0793] 流式细胞术分析

[0794] 转染后一天(除非另有说明),通过胰蛋白酶化将细胞收获到含有5%FBS的DMEM培养基中,然后在Attune NxT流式细胞仪(Thermo)或同等设备上进行分析。通过前向和侧向散射对活单细胞进行门控。Attune上的mCherry通道为YL2,在561nm激发,发射滤光片为620/15nm。Attune上的eGFP通道为BL1,在488nm激发,发射滤光片为530/30nm。使用FlowJo 10.8.1分析流式细胞术结果。将仅采用GIC RNA、无RT mRNA的转染用作背景对照;量化时从信号中减去背景。

[0795] 细胞分选

[0796] 转染后一天(除非另有说明),通过胰蛋白酶化将细胞收获到含有5%FBS的DMEM培养基中,并在超纯模式或同等模式下在具有130 μ m芯片的Sony SH800分选器上进行分选。通过离心收集分选的细胞并用PBS清洗。

[0797] 实施例9. 基于RNA的RTC和GIC组合物

[0798] 用于转染的RTC mRNA按照实施例4产生,并在表1中描述。

[0799] 表1: 2-RNA组件GIS RTC

[0800]	RTC标识	RTC:RT-模块来源生物体	SEQ ID NO.

F-ZoAl RT mRNA	白喉带鹀	19
F-TaGu RT mRNA	栗耳草雀	28
F-TriCasB RT mRNA	赤拟谷盗	3
OrLa-3F RT mRNA	青鳉	10
ZoAl RT mRNA(未标记的)	白喉带鹀	21
ZoAl_catdead RT mRNA	白喉带鹀	23
TriCasB RT mRNA(未标记的)	赤拟谷盗	5

[0801] 用于转染的GIC RNA按照实施例1和2产生,并在表2中描述。

[0802] 表2:2-RNA组件GIS GIC

GIC 标识	5'模块来源生物体	转基因启动子区域和 5' UTR	转基因	转基因 3' UTR 和聚 A 信号区	3'模块来源生物体	GIC SEQ ID NO.
TriCas_ZoAl	赤拟谷盗	CBh	NLSeGFP	SV40LPA	白喉带鹀	
TriCas_GeFo	赤拟谷盗	CBh	NLSeGFP	SV40LPA	中嘴地雀	
TriCas_TaGu	赤拟谷盗	CBh	NLSeGFP	SV40LPA	栗耳草雀	
TriCasFlipZoAl	赤拟谷盗	CBh_Flip	GFP	SV40LPA	白喉带鹀	
TriCasBsiZoAl	赤拟谷盗	CBh_Bsi	GFP	SV40LPA	白喉带鹀	
TCA5_ZoAl	赤拟谷盗	CBh	GFP	SV40LPA	白喉带鹀	
TCA5_GeFo	赤拟谷盗	mPGK	GFP	SV40LPA	中嘴地雀	
TCA5_TaGu	赤拟谷盗	CBh_Bsi	GFP	SV40LPA	栗耳草雀	
TCA5_TiGu	赤拟谷盗	CBh_Bsi	GFP	SV40LPA	白喉鹀	
TCARZ_GeFo	赤拟谷盗	CBh_Bsi	GFP	SV40LPA	中嘴地雀	
TCARZ_Cherry_GeFo	赤拟谷盗	CBh_Bsi	mCherry	SV40LPA	中嘴地雀	
HDVgu5_GeFo	无	CBh_Bsi	GFP	SV40LPA	中嘴地雀	
HDVgu5b_GeFo	无	CBh_Bsi	GFP	SV40LPA	中嘴地雀	
HDVgu5c_GeFo	无	CBh_Bsi	GFP	SV40LPA	中嘴地雀	
HDVgu5d_GeFo	无	CBh_Bsi	GFP	SV40LPA	中嘴地雀	
HDVac11_GeFo	无	CBh_Bsi	GFP	SV40LPA	中嘴地雀	
HDVac11b_GeFo	无	CBh_Bsi	GFP	SV40LPA	中嘴地雀	
HDVac12_GeFo	无	CBh_Bsi	GFP	SV40LPA	中嘴地雀	
HDVac12b_GeFo 等	无	CBh_Bsi	GFP	SV40LPA	中嘴地雀	

[0805] 实施例10. 候选蛋白逆转录活性筛选

[0806] 按照实施例3制备针对逆转录筛选的候选R2家族逆转录因子蛋白(见表3),并按照实施例5测试逆转录活性。一些TPRT或RT蛋白仅在一小组测定中被检测为具有活性(显示为

低/无)。

[0807] 表3:候选蛋白逆转录酶活性

SEQ ID NO.	源自的物种	物种参考代码	图 10 泳道	RT 活性
			#	
47	黄斑果蝇	DrMerc	15	无
57	库氏蝌蚪虾	LeCoB	11	无
55	欧洲蝌蚪虾	TriCan	12	无
43	玻璃海鞘	Ciln	3	无
51	三刺鱼	GaAc	19	无
49	黑腹果蝇	DrMe	14	低/无
45	美洲鲎	LiPo	13	低/无
53	九刺棘鱼	PuPu	16	低
7	蝇蛹金小蜂 (谱系 B)	NaviB	9	高
9	青鳉	OrLa	8	低
18	白喉带鹀	ZoAl	10	低/中等
27	栗耳草雀	TaGu	18	高
2	赤拟谷盗 (谱系 B)	TriCasB	5	高
25	白喉鸫	TiGu	17	低
33	拟果蝇	DroSi	4	高
36	家蚕	BoMo	2	高
39	污轮虫	AdVa	7	中等
41	大乳头水螅	HyMa	6	无
31	中嘴地雀	GeFo	NS	低/无

[0810] RT活性在物种间差异很大。如从图10中的PAGE图像结果看出,对于候选RT蛋白TriCasB、DroSi、TaGu、NaViB、BoMo、OrLa、AdVa(当相对蛋白表达归一化时)、ZoAl、LiPo(可变地可检测产物)、PuPu和TiGu以及GeFo(可变地可检测产物),在深色实线框中观察到预期长度的初始逆转录产物。对于CiIn、LeCoB、TriCan、DroMer、DroMe、HyMa和GaAc,未检测到可重现的RT产物。有时检测到DrMe和GeFo的活性非常低。预期产品长度处的条带不透明度,结合使用针对RT蛋白FLAG表位标签的抗体通过免疫印迹检测的纯化的蛋白的量,实现对逆转录活性水平进行比较估计以及将候选蛋白分选为具有高、中、低或无(所使用的测定法检测不到)逆转录活性的候选蛋白,如表3所见。一般而言,候选蛋白TriCasB、DroSi、TaGu、NaViB和BoMo表现出最高水平的逆转录酶能力,因此是用于包含在本发明的RTC中的强力候选者。

[0811] 实施例11. 针对3' 模块特异性的体内RT测定

[0812] 用不同的质粒组合转染9个HEK293T细胞群,所述质粒组合由表达由家蚕(B.mori)(SEQ ID NO.35)、拟果蝇(D.simulans)(SEQ ID NO.32)和青鳉(O.latipes)(SEQ ID NO.8)修饰的RT蛋白ORF的pcDNA3.1骨架质粒之一以及表达来自家蚕(B.mori)(SEQ ID 209)、拟

果蝇 (*D.simulans*) (SEQ ID NO.210) 或青鳉 (*O.latipes*) (SEQ ID NO.200) R2元件的3' UTR RNA的另外的质粒组成 (参见图11A)。每种RT蛋白与每种3' UTR RNA共表达。

[0813] 在给予足够的时间使RT蛋白质被转录和翻译并与转录的3' UTR RNA结合后,细胞被裂解,并且任何RT蛋白+RNA模板复合物都通过FLAG免疫纯化 (Sigma FLAG抗体树脂) 进行纯化。纯化每种输入细胞裂解物中存在的RNA和与每个免疫纯化的样品相关的RNA。将每个输入RNA样品和每个RT结合的RNA样品的等量等分试样以斑点网格的形式固定在Hybond N+膜 (Cytiva) 上。一起探测含有每种类型的3' UTR RNA的斑点的膜上3' UTR RNA的存在,如通过与使用T4多核苷酸激酶 (NEB) 进行³²P5' 端放射性标记的互补寡核苷酸探针杂交所检测的。换句话说,探测来自表达家蚕 (*B.mori*) R2 3' UTR的细胞的样本的家蚕 (*B.mori*) 3' UTR序列 (家蚕 (*B.mori*) 3' UTR探针为CATCATGGATTAGGATCGGAAGACCCCCG (SEQ ID NO.539); GTACGCCGGCGAAATTGGATCAGTAGATG (SEQ ID NO.540) 和GAGAAACAGACGGGCCTGATCTACACCC (SEQ ID NO.541)。探测表达拟果蝇 (*D.simulans*) R2 3' UTR RNA的样本的拟果蝇 (*D.simulans*) 3' UTR序列 (拟果蝇 (*D.simulans*) 3' UTR探针为CTATCTGAACCGAAGTTCCGCAACCCTACGTAC (SEQ ID NO.542), CACTGCGTGTGGTCAGTTTTCTAGCATGCACG (SEQ ID NO.543), 和GATGTTATGCCAAGACAGCAAGCAAATGTTTTGAACCAAACG) (SEQ ID NO.544)。探测表达青鳉 (*O.latipes*) R2 3' UTR RNA的样本的青鳉 (*O.latipes*) 3' UTR序列 (青鳉 (*O.latipes*) 3' UTR探针为TTGAGGCGAGTCACCACTCGCTTTCCGG (SEQ ID NO.545) 和GTGTCCGTCACGGGGACGACATCCGAGTG) (SEQ ID NO.546)。

[0814] 如图11B所见,修饰的家蚕 (*B.mori*) RT蛋白不仅结合其同源3' UTR,还结合拟果蝇 (*D.simulans*) 和青鳉 (*O.latipes*) R2元件的3' UTR序列,而修饰的拟果蝇 (*D.simulans*) 和青鳉 (*O.latipes*) 蛋白具有更高的选择性。家蚕 (*B.mori*) RT具有此处所述的发现,显示在人细胞中相对无差别的RNA相互作用。

[0815] 实施例12.家蚕 (*B.mori*)、拟果蝇 (*D.simulans*) 和青鳉 (*O.latipes*) 的体外TPRT特异性

[0816] 按照实施例3制备来自家蚕 (*B.mori*) (SEQ ID NO.36)、拟果蝇 (*D.simulans*) (SEQ ID NO.33) 和青鳉 (*O.latipes*) (SEQ ID NO.9) 的RT蛋白。按照实施例1制备GIC,其包含源自青鳉 (*O.latipes*) 3' UTR (SEQ ID NO.200) 的带有或不带有3' 附加的4nt rRNA序列 (SEQ ID 280) “R4”的GIC;RT识别序列和源自拟果蝇 (*D.simulans*) 3' UTR (SEQ ID NO.210) 的带有或不带有3' 附加的4nt rRNA序列 (SEQ ID 280) “R4”的GIC;RT识别序列。

[0817] 体外TPRT测定按照实施例6进行,以测试每种RT利用每种GIC的能力。

[0818] 源自拟果蝇 (*D.simulans*) 的RT蛋白不使用包含源自青鳉 (*O.latipes*) 3' UTR的GIC;RT识别序列的GIC,源自青鳉 (*O.latipes*) RT的RT蛋白不使用包含源自拟果蝇 (*D.simulans*) 3' UTR的GIC;RT识别序列的GIC进行TPRT。然而,源自家蚕 (*B.mori*) 的RT蛋白可以将两者用于TPRT (图12)。

[0819] 与其它修饰的R2 RT蛋白相反,家蚕 (*B.mori*) RT蛋白在TPRT期间具有无差别的模板复制 (即,它对其同源GIC没有选择性)。例如,源自青鳉 (*O.latipes*) 或拟果蝇 (*D.simulans*) 的RT对它们的同源GIC;RT识别序列有选择性,因此在设计更具选择性的GIS时可以是优选的。

[0820] 实施例13.RT特异性的系统发育筛选

[0821] 分别按照实施例3和1制备源自各种物种逆转录因子的RT蛋白和GIC,所述GIC包括源自各种物种天然逆转录因子3' UTR的GIC:RT识别序列,如表4所概述的。为了进行这种体外TRPT比较,所有GIC:RT识别序列均具有3'附加的“R4”4nt rRNA序列(SEQ ID 280),并且如果必需,还具有5'附加的(多个)鸟苷,用于T7 RNAP转录起始。

[0822] 按照实施例6进行体外TPRT测定,以测试每种RT识别给定GIC:RT识别序列的能力。变性PAGE凝胶上预期产物长度处的条带的不透明度实现对靶点引发的逆转录活性水平进行比较估计,并将候选蛋白分选为具有高、中、低或无(通过测定检测不到的)靶点引发逆转录活性的候选蛋白。

[0823] TPRT测定的结果总结在如下表4中。每个数据行都用所使用的RT蛋白标记,包括RT序列所源自的来源生物体。每个数据列都用所使用的GIC标记,包括GIC:RT识别序列所源自的来源生物体。带有减号(-)的单元格表示对于给定的RT和GIC的组合未观察到预期长度的产物。带有加号和减号(+/-)的单元格表示在至少一些测定中观察到的预期长度的产物的量几乎不可检测。带有单个加号(+)的单元格表示观察到少量预期长度的产物,两个加号(++)表示观察到中等量预期长度的产物,三个加号(+++)表示观察到大量预期长度的产物。

[0824] 与源自家蚕(*Bombyx mori*)的RT蛋白相比,源自栗耳草雀(*Taeniopygia guttata*)、青鳉(*Oryzias latipes*)、白喉带鹀(*Z.albicollis*)、白喉鵪(*Tinamus guttatus*)、赤拟谷盗(*Tribolium castaneum*) (R2谱系B)和拟果蝇(*Drosophila simulans*)的RT蛋白对包含它们的同源GIC:RT识别序列的GIC更具选择性。因此,与源自家蚕(*B.mori*)的RT蛋白相比,源自栗耳草雀(*T.guttata*)、青鳉(*O.latipes*)、白喉带鹀(*Z.albicollis*)、白喉鵪(*T.guttatus*)、赤拟谷盗(*T.castaneum*)和/或拟果蝇(*D.simulans*)的RT蛋白可更优选包含在本发明的GIS中,用以最小化或防止非预期模板序列插入受试者基因组。

[0825] 此外,源自白喉带鹀(*Z.albicollis*)、栗耳草雀(*T.guttata*)和/或白喉鵪(*T.guttatus*)的RT蛋白对源自鸟类物种的GIC:RT识别序列具有高度特异性。因此,源自白喉带鹀(*Z.albicollis*)、栗耳草雀(*T.guttata*)和/或白喉鵪(*T.guttatus*)的RT蛋白可优选包含在本发明的GIS中,因为它们可防止非预期模板序列插入受试者基因组,同时实现灵活地设计3'模块。

[0826] 表4:RT特异性

[0827]

		GIC: 3'模块 RT 识别序列（源自所示的来源）																												
SEQ ID NO.		217	212	213	214	211	215	208	207	206	204			202	203			210	209	218	219									
	来源 代码	DrMerc-GIC	LeCoB-GIC	TriCan-GIC	Ciln-GIC	GaAc-GiC	DrMe-GiC	LiPo-GIC	PuPu-GIC	NaviB-GIC	GeFo-GIC	OrLa-GIC		ZoAl-GIC	TaGu-GIC		TriCasB-GIC	201	TiGu-GIC	205		DroSi-GiC	210	BoMo-GiC	209		AdVa-GiC	218	HyMa-GiC	219
7	NaviB-RT	-	-	-	-	-	-	+	+	++	+	+/-		++	+		+++		++		+/-		+/-		+/-		+/-		-	
9	OrLa-RT	-	+	-	-	-	-	+/-	+	+/-	+	++		+	+		+/-		++		-		+		+/-		+/-		-	
18	ZoAl-RT	-	-	-	-	-	-	-	-	-	++	-		+	++		-		++		-		-		-		-		-	
27	TaGu-RT	-	-	-	-	-	-	-	-	-	+++	-		+++	+++		-		+++		-		-		-		-		-	
2	TriCasB-RT	+/-	-	-	-	-	+/-	+/-	++	++	++	++		++	++		++		++		+		+		+/-		-		-	
25	TiGu-	ND	ND	ND	ND	ND	ND	ND	ND	ND	+	-		+/-	+/-		-		+		ND		ND		ND		ND		ND	
	RT																													
33	DroSi-RT	+/-	ND	ND	+/-	ND	+/-	ND	ND	ND	ND	+/-		ND	ND		+		ND		++		+		+		+		ND	
36	BoMo-RT	+	+	+	-	-	++	++	++	++	++	++		++	++		++		++		+++		+++		+++		++		+/-	

[0828]

[0829] 实施例14. 3' 模块工程化对源自家蚕(B.mori)的RT TPRT的影响

[0830] 按照实施例3制备源自家蚕(B.mori)的RT蛋白(SEQ ID NO 36)。按照实施例1制备含有表5中所述的具有5' 和/或3' 侧翼序列的BoMo 3' UTR序列(SEQ ID 209)的GIC。

[0831] 表5:源自家蚕(B.mori)的GIC

[0832] *表示为T7 RNAP转录起始而添加的5' 鸟苷

[0833]	模板标记	GIC: 5' rRNA 长度	源自 RE 的 序列来源	3'受试者 rRNA 长度	A 区长 度
	GG*-BM3UTR-R3	0 nt	家蚕	3 nt (SEQ ID 286)	0 nt
	R26_BM3UTR	26 nt (SEQ ID 254)	家蚕	0 nt	0 nt
	GG*_BM3UTR_R4	0 nt	家蚕	4 nt (SEQ ID 280)	0 nt
	GGG*-R4_BM3UTR_R4	4 nt (SEQ ID 275)	家蚕	4 nt (SEQ ID 280)	0 nt
	R26_BM3UTR_R4	26 nt (SEQ ID 254)	家蚕	4 nt (SEQ ID 280)	0 nt
	R26_BM3UTR_R4_PA	26 nt (SEQ ID 254)	家蚕	4 nt (SEQ ID 280)	22 nt
	R26_BM3UTR_R20	26 nt (SEQ ID 254)	家蚕	20 nt (SEQ ID 285)	0 nt

[0834] 体外TPRT测定按照实施例6所述进行,其源自家蚕(B.mori)的RT蛋白分别与每个模板和64或84bp靶位点DNA双链(分别为SEQ ID 291和292)结合。箭头标记预期TPRT产物长度的区域,用于预期3'连接形成。

[0835] 如图13所示,自家蚕(B.mori)3' UTR RNA的3'端的序列延伸不显著影响家蚕(B.mori)RT的靶点引发的逆转录(TPRT)效率。具体而言,TPRT的模板上3'侧翼rRNA不是必需的。3'添加4nt rRNA增加了TPRT产物长度的均一性,但没有增加实际TPRT产物长度,正如如果将整个模板RNA复制到cDNA中将预期的那样。相反,另外的4nt模板长度可以与有切口的靶位点引物碱基配对,用以启动cDNA合成。

[0836] 3' rRNA长度增加到20nt通过实现内部启动(圆圈标记的位置)降低3'连接保真度,相比之下,使用具有仅4nt的3' rRNA的模板RNA进行TPRT合成的精度更高(箭头标记高保真度3'连接形成区域)。因此,相对于4nt 3'侧翼rRNA序列,20nt 3'侧翼rRNA序列是不利的。值得注意的是,3'侧翼rRNA可通过至少22nt的腺苷(PA)区延伸,而不损失正确产物合成的效率或精度。

[0837] 实施例15. 3'模块工程化对源自青鳉(O.latipes)的RT的TPRT效率的影响

[0838] 按照实施例3制备源自青鳉(O.latipes)的RT蛋白(SEQ ID NO 9)。按照实施例16制备含有表6中所述的具有5'和/或3'侧翼序列的OrLa3' UTR序列(SEQ ID 200)的GIC。

[0839] 表6:源自青鳉(O.latipes)的GIC

[0840] *表示为T7 RNAP转录起始而添加的5'鸟苷

模板标记	GIC: 5' rRNA 长度	源自 RE 的区域	GIC: 3'模块 rRNA 序列 长度	GIC: 3'模块 A-区序列长度
R26_OL	26 nt (SEQ ID 254)	青鳉	0 nt	0 nt
R4_OL_R4	4 nt (SEQ ID 275)	青鳉	4 nt (SEQ ID 280)	0 nt
R26_OL_R4	26 nt (SEQ ID 254)	青鳉	4 nt (SEQ ID 280)	0 nt
R26_OL_R20	26 nt (SEQ ID 254)	青鳉	20 nt (SEQ ID 285)	0 nt
R26_OL_R4_PA	26 nt (SEQ ID 254)	青鳉	4 nt (SEQ ID 280)	22 nt
GG*-R0-OL3-R0	0 nt	青鳉	0 nt	0 nt
GG*-R0-OL3-R4	0 nt	青鳉	4 nt (SEQ ID 280)	0 nt
GG*-R0-OL3-R8	0 nt	青鳉	8 nt (SEQ ID 287)	0 nt
GG*-R0-OL3-R12	0 nt	青鳉	12 nt (SEQ ID 288)	0 nt
GG*-R0-OL3-R16	0 nt	青鳉	16 nt (SEQ ID 289)	0 nt
GG*-R0-OL3-R20	0 nt	青鳉	20 nt (SEQ ID 285)	0 nt

[0843] 体外TPRT测定按照实施例6所述进行,其中源自青鳉 (*O. latipes*) 的RT蛋白分别与每个模板结合。产物形成表明源自青鳉 (*O. latipes*) 的RT具有TPRT生化活性。

[0844] 如图14 (A) 所示,缺乏rRNA的3' 延伸的青鳉 (*O. latipes*) 3' UTR未被青鳉 (*O. latipes*) RT有效用于TPRT,这与图13中的结果不同,图13表明在无3' 侧翼rRNA的情况下,家蚕 (*B. mori*) RT使用家蚕 (*B. mori*) 3' UTR RNA进行有效TPRT。与家蚕 (*B. mori*) 组件一样,3' 侧翼rRNA可通过至少22nt的聚腺苷酸 (PA) 区延伸,而不抑制青鳉 (*O. latipes*) RT TPRT,并且产物长度的均一性增加。

[0845] 进行了第二组TPRT测定,以系统地检查不同3' 受试者rRNA长度的影响。

[0846] 如图14 (B) 所示,这些结果证实上述观察到的结果。缺乏3' rRNA延伸导致青鳉 (*O. latipes*) RT活性差且内部启动不当,而4nt的rRNA的存在足以刺激TPRT并改善3' 连接精度。因此,在本发明的GIC中,在GIC3' 模块rRNA序列中包含仅4nt的3' 受试者rRNA可以是优选的。GIC 3' rRNA序列长度的增加未相应地增加TPRT产物的长度,表明GIC 3' rRNA序列未被复制;相反,它可与有切口的靶位点引物DNA碱基配对用以启动cDNA合成。

[0847] 实施例16. 3' 模块工程化对源自赤拟谷盗 (*T. castaneum*) 的RT的TPRT效率的影响

[0848] 按照实施例3制备来自赤拟谷盗 (*T. castaneum*) 的RT蛋白 (SEQ ID NO.2)。按照实施例1制备含有表7中所述的具有5' 和/或3' 侧翼序列的TriCasB 3' UTR序列 (SEQ ID 201) 的GIC。

[0849] 表7:源自赤拟谷盗(*T.castaneum*)的GIC

[0850] 模板名称	GIC: 5' rRNA 序列	源自 RE 的	GIC: 3'模块 rRNA	GIC: 3'模块
	长度	区域	序列长度	A-区序列长度
R25-TC_UTR-R4	25 nt (SEQ ID 276)	赤拟谷盗	4 nt (SEQ ID 280)	0 nt
[0851] R25-TC_UTR-R4_PA	25 nt (SEQ ID 276)	赤拟谷盗	4 nt (SEQ ID 280)	22 nt
	R25-TC_UTR-R10	25 nt (SEQ ID 276)	赤拟谷盗	10 nt (SEQ ID 284)
				0 nt

[0852] 体外TPRT测定按照实施例6所述进行,其中源自赤拟谷盗(*T.castaneum*)的RT蛋白分别与每个模板结合。箭头指示预期TPRT产物的位置。靶位点DNA被检测为图像底部的暗条带。产物形成表明源自赤拟谷盗(*T.castaneum*)的RT对TPRT具有生化活性。

[0853] 如图15可见的,通过添加超过4nt的GIC:3' 模块rRNA序列,产物合成没有明显改善,并且3' 侧翼rRNA可通过至少22nt的聚腺苷酸(PA)延伸,而不抑制正确的产物合成。

[0854] 实施例17. 3' 模块工程化对源自白喉带鹀(*Z.albicollis*)和栗耳草雀(*T.guttata*)的RT的TPRT效率的影响

[0855] 按照实施例3制备源自白喉带鹀(*Z.albicollis*)的RT蛋白(SEQ ID NO 18)。按照实施例1制备含有表8中所述的具有5' 和/或3' 侧翼序列的白喉带鹀(*Z.albicollis*, ZoA1) 3' UTR(SEQ ID 202)或白喉鹀(*T.guttatus*, TiGu) 3' UTR(SEQ ID 205)或栗耳草雀(*T.guttata*, TaGu) 3' UTR(SEQ ID 203)的3' 模块RT识别序列的GIC。

[0856] 表8:鸟类R2 GIC

[0857] 模板名称	图 16 泳道 #	GIC: 5' rRNA 序列 长度	GIC: 3'模 块 RT 识别 序列来源	GIC: 3'模块 rRNA 序列长度	GIC: 3' 模块 A- 区长度
R26(-28)-ZA3-R0	1	26 nt (SEQ ID 254)	白喉带鹀	0 nt	0 nt
R26(-28)-ZA3-R4	2	26 nt (SEQ ID 254)	白喉带鹀	4 nt (SEQ ID 280)	0 nt
R26(-28)-ZA3-R20	3	26 nt (SEQ ID 254)	白喉带鹀	20 nt (SEQ ID 285)	0 nt
R26(-28)-ZA3-R4PA	4	26 nt (SEQ ID 254)	白喉带鹀	4 nt (SEQ ID 280)	22 nt
R26(-28)-TiG3-R0	5	26 nt (SEQ ID 254)	白喉鹀	0 nt	0 nt
产物丢失	6				
R26(-28)-TiG3-R20	7	26 nt (SEQ ID 254)	白喉鹀	20 nt (SEQ ID 285)	0 nt
R26(-28)-TiG3-R4PA	8	26 nt (SEQ ID 254)	白喉鹀	4 nt (SEQ ID 280)	22 nt
R28(-28)-TaG3-R0	9	28 nt (SEQ ID 252)	栗耳草雀	0 nt	0 nt
[0858] R28(-28)-TaG3-R4	10	28 nt (SEQ ID 252)	栗耳草雀	4 nt (SEQ ID 280)	0 nt
	R28(-28)-TaG3-R20	11	28 nt (SEQ ID 252)	栗耳草雀	20 nt (SEQ ID 285)
	R28(-28)-TaG3-R4PA	12	28 nt (SEQ ID 252)	栗耳草雀	4 nt (SEQ ID 280)
					22 nt

[0859] 体外TPRT测定按照实施例6所述进行,其中源自白喉带鹀(*Z.albicollis*)的RT蛋白

白分别与每个模板结合。实线框包围TPRT产物,虚线框包围沉淀恢复对照,虚线和点混合轮廓的框包围64bp靶位点DNA。这些结果表明源自白喉带鹀(*Z.albicollis*)的RT对靶点引发的逆转录具有生化活性。

[0860] 如图16可见的,源自白喉带鹀(*Z.albicollis*)的RT蛋白未有效利用具有缺少GIC:3' 模块rRNA序列的3' 模块设计的GIC,因此显示在GIC 3' rRNA序列可以碱基配对的靶位点处cDNA合成效率增加。GIC 3' rRNA序列长度的增加不增加TPRT产物的长度,表明GIC 3' rRNA序列未被复制;它必须与带切口的靶位点引物碱基配对用以启动cDNA合成。使用包含4nt的3' rRNA序列和A区22nt尾或20nt的rRNA序列的GIC产生的TPRT产物合成量最高。最后,源自白喉带鹀(*Z.albicollis*)的RT蛋白能够利用含有源自所测试的几种鸟类的GIC:3' 模块RT识别序列的GIC。使用源自栗耳草雀(*T.guttata*) (SEQ ID 27)的RT蛋白进行了平行实验,结果源自栗耳草雀(*T.guttata*)的鸟类RT蛋白可利用含有源自几种鸟类的GIC:3' 模块RT识别序列的GIC,并且在其利用含有GIC:3' rRNA序列的GIC中具有选择性。

[0861] 这些结果进一步支持GIS可以包含源自白喉带鹀(*Z.albicollis*)或栗耳草雀(*T.guttata*)的RT蛋白与源自各种鸟类的GIC:3' 模块RT识别序列的组合,其与具有或不具有GIC:3' 模块A-区序列的GIC:3' 模块rRNA序列一起以改变TPRT反应效率。在GIC:3' 模块rRNA序列不能与带切口的靶位点引物碱基配对的情况下,没有观察到cDNA合成。如果可以与GIC:3' 模块rRNA碱基配对的切口下游的靶位点序列被改变为不同的序列(突变靶位点; SEQ ID 296),仍观察到剪切,但TPRT因GIC:3' 模块rRNA与引物链3' 端碱基配对失败而被阻断。因此,只有采用在正确的靶位点序列处的切口,产生与GIC:3' 模块rRNA序列匹配的引物3' 端,TPRT对于cDNA合成才有效。使用突变靶位点,如果将GIC:3' 模块rRNA序列改变为将与切口产生的引物3' 端碱基配对的序列,则通过TRPT的cDNA合成得以挽救。这表明GIC:3' 模块RNA序列的作用机制是与引物链的3' 末端区域碱基配对。

[0862] 实施例18.GIC:3' 模块尾工程化对转基因插入人类基因组的体内影响

[0863] 部分A:源自栗耳草雀(*T.guttata*)的RTC:RT模块

[0864] 按照实施例4制备源自栗耳草雀(*T.guttata*)的RTC mRNA(SEQ ID NO 28)。按照实施例2制备包含GFP转基因表达盒有效载荷并具有相同GIC:5' 模块和GIC:3' 模块RT识别序列(TCA5_CBhBsi_GFP_GeFo3)的GIC RNA并在表9中列举。

[0865] 使用Lipofectamine Messenger Max将hTERT RPE-1细胞与RTC和所示的GIC(1:1摩尔比)共转染,然后在24小时后收获。通过FACS分析确定每次处理中GFP阳性细胞的百分比,结果报告于表9。

[0866] 表9:体内3' 模块尾工程化影响

[0867]

GIC: 3'模块 rRNA 长度 (nt)	GIC: 3'模块 A-区 长度 (nt)	GIC SEQ ID NO	GFP 阳性细胞百分比
0	0	500	0.12
0	22	501	0.17
4	0	502	4.05
4	22	503	15.67
20	0	504	6.84
20	22	505	4.23

[0868] 这些结果表明,与所测试的其它组合相比,使用包含4nt的GIC:3' 模块rRNA序列和22nt的A-区序列的GIC:3' 模块得到显著更高的转基因插入率。值得注意的是,包含至少4nt的GIC:3' 模块rRNA序列的其它组合确实实现在哺乳动物细胞系中成功插入并表达转基因。然而,采用20nt的GIC:3' 模块rRNA序列,22nt长度的A-区序列是抑制性的。

[0869] B部分:源自栗耳草雀(*T.guttata*)和白喉带鹀(*Z.albicollis*)的RTC:RT模块的比较

[0870] 按照实施例4制备源自栗耳草雀(*T.guttata*) (SEQ ID NO 28) 或白喉带鹀(*Z.albicollis*) (SEQ ID NO 19) 的RTC mRNA。按照实施例2制备包含GFP转基因表达盒有效载荷和相同GIC:5' 模块和GIC:3' 模块RT识别序列(TCA5_CBhBsi_GFP_GeFo3)的GIC RNA,如表10所列举的。

[0871] 使用Lipofectamine Messenger Max将hTERT RPE-1细胞与RTC和所示的GIC(摩尔比1:3)共转染,然后在24小时后收获。通过FACS分析确定每种处理的GFP阳性细胞的百分比和GFP阳性细胞中GFP表达的中位强度,如表10所示。

[0872] 表10:体内另外的3' 模块尾工程化影响

[0873]

RTC: RT- 模块来源 生物体	GIC: 3' 模 块 rRNA 长 度 (nt)	GIC: 3' 模 块 A-区长 度 (nt)	GIC SEQ ID NO	GFP 阳性细胞 百分比 (%)	GFP 强度 (单独的 GIC 背景以上的相 对荧光单位)
栗耳草雀	0	0	500	0.093	1705
栗耳草雀	0	22	501	0.17	2098
栗耳草雀	4	0	502	2.84	4570
栗耳草雀	4	22	503	14.708	9011
栗耳草雀	20	0	504	5.342	5003
栗耳草雀	20	22	505	2.235	3835
白喉带鹀	0	0	500	0	0
白喉带鹀	0	22	501	0.25	2183
白喉带鹀	4	0	502	3.83	4260
白喉带鹀	4	22	503	13.608	7364
白喉带鹀	20	0	504	4.972	4315
白喉带鹀	20	22	505	2.075	3145

[0874] 这些结果证实了A部分中所见的关于源自栗耳草雀 (*T.guttata*) 的RTC mRNA的那些结果。此外,它们表明源自白喉带鹀 (*Z.albicollis*) 的RTC mRNA在GIC:3' 模块rRNA序列和A区长度方面表现出与源自栗耳草雀 (*T.guttata*) 的RTC mRNA相同的效率模式。使用最佳的R4A22模板,源自白喉带鹀 (*Z.albicollis*) 的RTC:RT模块在转基因插入方面的效率仅略低于源自栗耳草雀 (*T.guttata*) 的RTC:RT模块。

[0875] 源自栗耳草雀 (*T.guttata*) 和白喉带鹀 (*Z.albicollis*) 两者的RTC:RT模块都是本发明的GIS的可用组件。两者都显示出利用具有可变长度的GIC:3' 模块rRNA和/或GIC:3' 模块A-区的GIC的能力,其中潜在的最佳的GIC组成包含约4nt长度的GIC:3' 模块rRNA序列和约22nt长度的GIC:3' 模块A-区序列。

[0876] 实施例19. 3' 模块工程化对源自栗耳草雀 (*T.guttata*) 的RT的TPRT效率的影响

[0877] 按照实施例3制备源自栗耳草雀 (*T.guttata*) 的RT蛋白 (SEQ ID NO 27)。按照实施例1制备含有不同GIC:3' 模块RT识别序列的GIC,添加或不添加用于T7 RNAP转录起始的(多个)5' 鸟苷,以及添加GIC:3' 模块rRNA序列R4 (SEQ ID 280),如表11所述。

[0878] 表11:栗耳草雀 (*T.guttata*) RT对GIC:3' 模块RT识别序列的特异性

[0879] *表示添加用于T7 RNAP转录起始的(多个)5' 鸟苷

模板标记	图 17 泳道#	GIC: 3'模块 RT 识别序列 来源和 SEQ ID (#)	GIC: 3'模块 rRNA 序列长度	SEQ ID NO.
无模板对照	2	NA	NA	NA
GGG*-HM3-R4	3	大乳头水螅 (219)	4 nt	
GGG*-AV3-R4	4	污轮虫 (218)	4 nt	
G*-LP3-R4	5	美洲鲎 (208)	4 nt	
G*-ZA3-R4	6	白喉带鹀 (202)	4 nt	
G*-TiG3-R4	7	白喉鹀 (205)	4 nt	
G*-TaG3-R4	8	栗耳草雀 (203)	4 nt	
G*-GF3-R4	9	中嘴地雀 (204)	4 nt	
GA3-R4	10	三刺鱼 (211)	4 nt	
OL3-R4	11	青鳉 (200)	4 nt	
G*-PP3-R4	12	九刺棘鱼 (207)	4 nt	
GGG*-TCasB3-R4	13	赤拟谷盗 (201)	4 nt	
G*-NVB3-R4	14	蝇蛹金小蜂 (206)	4 nt	
GGG*-CI3-R4	15	玻璃海鞘 (214)	4 nt	
BM3-R4	16	家蚕 (209)	4 nt	
G*-LCB3-R4	17	库氏蝌蚪虾 (212)	4 nt	
G*-TCan3-R4	18	欧洲蝌蚪虾 (213)	4 nt	
G*-DS3-5iA-R4	19	拟果蝇 (210)	4 nt	
GG*-DMer3-R4	20	黄斑果蝇 (217)	4 nt	
G*-DMel3-5iA-R4	21	黑腹果蝇 (215)	4 nt	
GG*-DN3-R4	22	辉颜果蝇 (<i>D. nasuta</i>) (216)	4 nt	

[0881] 按实施例6所述进行体外TPRT测定,其中源自栗耳草雀(*T. guttata*)的RT蛋白分别与每个模板结合。模板序列由逆转录因子3' UTR序列组成,如果必需支持T7 RNAP转录则添加(多个)5' 鸟苷,并且GIC:3' 模块rRNA序列长度为4nt,以及没有GIC:3' 模块A-区rRNA序列。实线框包围预期的TPRT产物,虚线框包围沉淀恢复对照,虚线和点线混合轮廓的框包围剩余的完整64bp靶位点DNA。

[0882] 如图17所示,源自栗耳草雀(*T. guttata*)的RT蛋白能够识别具有源自各种鸟类的GIC:3' 模块RT识别序列的GIC,在存在包含来自非鸟类的GIC:3' 模块RT识别序列的GIC的情况下,观察到的TPRT活性非常小或没有。此外,采用源自栗耳草雀(*T. guttata*)的RT蛋白和源自中嘴地雀(*G. fortis*)的GIC与最短的所测试的鸟类GIC:3' 模块RT识别序列的组合,观察到高TPRT活性。

[0883] 因此,可以优选设计至少一种本发明的GIS以包含至少一个包含或编码至少一种

源自栗耳草雀 (*T.guttata*) 的RT蛋白的RTC:RT模块和至少一种包含或编码至少一个源自中嘴地雀 (*G.fortis*) 的GIC:3' 模块RT识别序列的GIC,其特别是将给予非鸟受试者。该组合可使GIS既高效地将其有效载荷序列插入受试者基因组,又对其GIC具有高度特异性。

[0884] 实施例20. 5' 和3' 模块工程化对赤拟谷盗 (*T.castaneum*) RT的转基因插入人类基因组的效率的体内影响

[0885] 按照实施例3用质粒转染293T细胞,以表达从赤拟谷盗 (*T.castaneum*) R2的三个谱系之一修饰的蛋白,其中合成序列ORF呈现用于翻译的单个AUG起始密码子 (SEQ ID NO.1)。一些细胞没有用质粒平行转染作为阴性对照。48小时后,使用lipofectamine3000用按照实施例1制备的纯化GIC RNA以表12中所述的组合转染这些细胞。第二次转染后第1天从转染细胞中纯化基因组DNA。

[0886] 表12:源自赤拟谷盗 (*T.castaneum*) 的GIC

模板标记	GIC: 5'模块 rRNA 序列长度*	GIC: 5'模块 RE 序列来源**	GIC: 3'模块 RT 识别序列来源	GIC: 3'模块 rRNA 序列长度	GIC: 3'模块 A-区序列长度	SEQ ID NO.
R25-TCB3-R4	25nt	NA	赤拟谷盗	4 nt	0 nt	330
R25-TCB3-R10	25nt	NA	赤拟谷盗	10 nt	0 nt	331
R25-TCB3-R4-PA	25nt	NA	赤拟谷盗	4 nt	22 nt	332
R25*-TCB5_TCB3-R4	25nt*	赤拟谷盗	赤拟谷盗	4 nt	0 nt	333
R25*-TCB5_TCB3-R10	25nt*	赤拟谷盗	赤拟谷盗	10 nt	0 nt	334
R25*-TCB5_TCB3-R4-PA	25nt*	赤拟谷盗	赤拟谷盗	4 nt	22 nt	335
R25*-TCB5_TCB3-PA	25nt*	赤拟谷盗	赤拟谷盗	0 nt	22 nt	336
R25*-TCB5_TCB3-R10PA	25nt*	赤拟谷盗	赤拟谷盗	10 nt	22 nt	337
• 25 nt rRNA 中的 3' 13 nt 包含在 GIC: 5'模块中, 并且在自剪切后仍将保留。5' 12 nt 将被移除。 • **TriCasB 5'模块序列自天然序列修饰, 以在靶位点第一切口上游包含 13 nt 的 rRNA, 其与人类基因组匹配, 而不是与较短的天然 rRNA 长度和进化上改变的 rRNA 序列匹配。						

[0888] 在图18A证明的一项实验中,GIC具有赤拟谷盗 (*T.castaneum*) R2谱系B 5' 模块和赤拟谷盗 (*T.castaneum*) R2谱系B 3' 模块 ("5_3UTR"),并且在GIC:3' 模块rRNA长度 (0、R4或R10) 和GIC:3' 模块22nt A-区 (PA) 存在或不存在方面不同。使用来自不同细胞群的一致量的基因组DNA (正向引物:CTCCTGACCACTAGCTCACTGACTAATTTTAAAC (SEQ ID NO:547)) 和反向引物:CCACTTATTCTACACCTCTCATGTCTCTTCACCG (SEQ ID NO:548)) 进行PCR以检测转基因插入3' 连接。PCR产物DNA在非变性琼脂糖凝胶上解析并用溴化乙锭检测。

[0889] 在用GIC:3' 模块22nt A-区 (PA) 转染的细胞中,尤其是在采用长度为4nt的GIC:3' 模块rRNA转染的细胞中,具有针对预期的3' 连接所预计的尺寸的连接PCR产物最为丰富。没有GIC:3' 模块rRNA的GIC:3' 模块A-区不足以检测到转基因插入,这有利于排除加腺苷尾的人宿主细胞mRNA作为转基因合成的潜在模板。

[0890] 在图18B证明的不同实验中,GIC具有赤拟谷盗 (*T.castaneum*) R2谱系B 3' 模块,其具有或没有赤拟谷盗 (*T.castaneum*) R2谱系B 5' 模块(分别为“5_3”或“3”)。GIC在GIC:3' 模块rRNA长度(R4或R10)和/或GIC:3' 模块A-区 (PA) 存在或不存在方面也有所不同。使用3' 插入连接引物(正向引物:CTCCTGACCAACTAGCTCACTGACTAATTTTAAAC (SEQ ID NO:547) 和反向引物:CCACTTATTCTACACCTCTCATGTCTCTTCACCG (SEQ ID NO:548))或5' 插入连接引物(正向引物:CTAGCAGCCGACTTAGAACTGGTGCGG (SEQ ID NO:549) 和反向引物:CTTCGTCTTCGGAATCCATGTCCATAGC (SEQ ID NO:550))使用来自不同细胞群的一致量的基因组DNA进行PCR以检测转基因插入3' 和5' 连接。PCR产物DNA在1X TAE中的非变性琼脂糖凝胶上解析,用溴化乙锭检测,并在BioRad分子成像仪ChemIDoc XRS+上成像。

[0891] 在左图中,在用长度为4nt的GIC:3' 模块rRNA和GIC:3' 模块A-区 (PA) 转染的细胞中,具有针对完美的3' 连接(用箭头指示的)所预期的尺寸的PCR产物最为丰富。此外,赤拟谷盗 (*T.castaneum*) R2谱系B 5' 模块的存在已经增加了3' 连接产物,表明更多的被插入的转基因。对于R4_PA GIC,如果检测到任何大小不正确的PCR产物,也是极少的,表明3' 连接形成的高保真度。然而,用其它GIC转染的细胞有另外的3' 连接PCR产物。

[0892] 在右图中,对于具有或不具有5' 模块的GIC,针对全长转基因的5' 连接所预期尺寸的PCR产物大小不同,在每种情况下都用箭头指示。全长转基因插入的5' 连接的PCR产物在用长度为4nt的GIC:3' 模块rRNA和GIC:3' 模块A-区 (PA) 转染的细胞中最丰富。此外,尽管5' 连接PCR产物长度较长(这将偏向于效率较低的PCR),但赤拟谷盗 (*T.castaneum*) R2谱系B 5' 模块的存在增加了5' 连接产物的量和均一性,表明插入的转基因更多,插入保真度更高。

[0893] 仅当RT蛋白表达和RNA模板转染两者都发生时,5' 和3' 连接形成两者才可检测到。在无模板RNA情况下表达RT蛋白的细胞或用模板RNA转染而无RT蛋白表达的细胞不显示或显示少量非特异性PCR产物。

[0894] 这些结果表明,较短长度的GIC:3' 模块rRNA序列,诸如4nt长的序列可以为本发明的GIS提供优异的TPRT活性,包括更高的反应产量和更具特异性的转基因连接形成(5' 和3' 连接两者)。

[0895] 实施例21. 5' 模块RZ工程化对赤拟谷盗 (*T.castaneum*) RT将转基因插入人类基因组的效率的体内影响

[0896] 按照实施例3转染293T细胞以表达源自赤拟谷盗 (*T.castaneum*) 的RT蛋白 (SEQ ID1)。随后,使用Lipofectamine3000用按照实施例1制备的GIC RNA以表13中所述的组合转染这些细胞。所有GIC构建体包含源自赤拟谷盗 (*T.castaneum*) 的GIC:3' 模块RT识别序列、长度为4nt的GIC:3' 模块rRNA序列和长度为22nt的GIC:3' 模块A-区序列 (SEQ ID 338)。GIC构建体在GIC:5' 模块方面不同。

[0897] 表13:具有供选择的RZ的源自赤拟谷盗 (*T.castaneum*) 的GIC

模板标记	GIC: 5'模块 rRNA 序列长度**	图 19 泳 道#s	GIC: 5'模块 RZ 序列来源/修饰	GIC: 5'模块 RE 序列来源	SEQ ID NO.
TriCasB_5 (SEQ ID 62)	13 (SEQ ID 266)	2 & 10	赤拟谷盗/无额外*	赤拟谷盗	
TriCasB_5rzdead (SEQ ID 63)	13 (SEQ ID 266)	3 & 11	赤拟谷盗/灭活的	赤拟谷盗	
TriCasB_5RZ (SEQ ID 64)	13 (SEQ ID 266)	4 & 12	赤拟谷盗/无额外*	赤拟谷盗	
TriCasB_5RZmin (SEQ ID 65)	13 (SEQ ID 266)	5 & 13	赤拟谷盗/缩短的 5RZ	赤拟谷盗	
TriCasB_5RZmin+down (SEQ ID 145)	13 (SEQ ID 266)	6 & 14	赤拟谷盗/代替 TriCasB 5 的天然 RZ 区的缩短的 5RZ	赤拟谷盗	
OrLa_5L (SEQ ID 60)	26 (SEQ ID 254)	7 & 15	青鳉/无	青鳉	
DroSi_5 (SEQ ID 70)	0	8 & 16	拟果蝇/无	拟果蝇	
[0899] *TriCasB 5'模块序列自天然序列修饰, 以在靶位点第一切口上游包含 13 nt 的 rRNA, 其与人类基因组匹配, 而不是与较短的天然长度的 rRNA 和进化上改变的 rRNA 序列匹配。 **自剪切后的 5' rRNA 长度					

[0900] 按照实施例20使用2个不同的来自转染细胞池的基因组DNA的PCR扩增来检测3' 插入连接(上图)和5' 插入连接(下图)。PCR引物:3' 连接:正向引物:CTCCTGACCACTAGCTCACT GACTAATTTTAAAC(SEQ ID NO:547),反向引物:CCACTATTCTACACCTCTCATGTCTCTTCACCG(SEQ ID NO:548);5' 连接:正向引物:CCAGGGGAATCCGACTGTTAATTAAAACAAAGC(SEQ ID NO:551),反向引物:GCGACTCGCATCACTGACTTTAATTGGTTG(SEQ ID NO:552)。

[0901] 如图19中所观察到的,具有源自赤拟谷盗(*T. castaneum*)谱系B或青鳉(*O. latipes*)R2逆转录因子的5' 模块组件的GIC支持最多的转基因插入和连接保真度,这由预期长度的占主导地位的单个PCR产物证明,该预期长度是针对具有精确的3' 和5' 连接形成的全长转基因插入。赤拟谷盗(*T. castaneum*)谱系B 5' 模块RZ活性位点的单个nt变化杀死RZ活性(TriCasB_5rzdead),严重降低了转基因插入效率并损害插入保真度。此外,相对于仅包含源自赤拟谷盗(*T. castaneum*)的完整5' 模块序列的RZ区域(TriCasB_5RZ)的GIC,包含全长赤拟谷盗(*T. castaneum*)GIC:5' 模块RE序列(TriCasB_5)的GIC产生了优异的转基因插入。然而,仅具有长度最小化版本的赤拟谷盗(*T. castaneum*)RZ(TriCasB_5RZmin)的

GIC表现与GIC“TriCasB_5”相当,优于“TriCasB_5RZ”,并且优于“TriCasB_5RZmin+down”,所述“TriCasB_5RZmin+down”在从“TriCasB_5”中移出的RZ下游具有来自赤拟谷盗(*T.castaneum*) 5' UTR的加回序列,以形成“TriCasB_5RZ”。

[0902] 最后,尽管包含青鳉(*O.latipes*) 5' 模块组件(OrLa_5L)的GIC在与源自赤拟谷盗(*T.castaneum*)的RT蛋白组合时表现与“TriCasB_5”一样好,其GIC:3' 模块组件源自赤拟谷盗(*T.castaneum*),但对于拟果蝇(*D.simulans*) 5' 模块组件(DroSi_5)情况并非如此。拟果蝇(*D.simulans*) 5' 模块RZ自剪切活性去除初始GIC转录物中5' UTR的5' 处的所有序列,包括任何5' rRNA。没有在自剪切RZ内受保护的5' rRNA,初始第一链cDNA合成仍可发生,但相对于使用“TriCasB_5”或“OrLa_5L”的GIC,5' 连接形成和稳定转基因插入所必需的第二链合成效率和精度降低。这从5' PCR连接产物长度的模糊分布是明显可见的(图19,下图泳道16)。

[0903] 实施例22:GIC:5' 模块rRNA长度

[0904] 按照实施例4制备F-ZoA1 RT(SEQ ID NO 19)的RTC mRNA。按照实施例2制备包含GFP转基因表达盒(SEQ ID 506,CBhBsi_GPF_GeFo_R4A22)的GIC RNA,其仅在5' 模块的序列上不同。开发了经优化采用自剪切HDV RZ折叠的从头设计的GIC:5' 模块序列,其强制自剪切的GIC 5' 端处在靶位点切口上游的rRNA序列的特定位置,例如在位置-28(HDV-28)或在位置-13(HDV-13)或在允许+1鸟苷要求的另一位置,并且通过经验验证产生T7 RNAP转录物自剪切。

[0905] 此外,经优化采用自剪切HDV RZ折叠的从头设计的GIC:5' 模块序列通过给定每个自剪切位置的GIC:5' 模块中存在的rRNA序列的量进行定制。例如,在相对于TPRT切口的位置-28处诱导自剪切的GIC:5' 模块可包含28nt的5' rRNA,或者通过从其3' 边界修剪rRNA序列,可包含另一长度的rRNA,诸如25、26或27nt。

[0906] 使用Lipofectamine Messenger Max用以1:3摩尔比混合的RTC mRNA和所示的GIC RNA共转染hTERT RPE-1细胞。24小时后通过流式细胞术分析转染的细胞池以检测GFP+细胞百分比。通过FACS分析确定GFP阳性细胞的百分比,如表14所报告的。

[0907] 表14:GIC:5' 模块rRNA序列长度的影响

GIC: 5'模块 RZ 序列 ID	GIC: 5'模块 rRNA 起始序列位置	GIC: 5'模块 rRNA 序列长度	RZ 自剪切效率	GFP 阳性细胞百分比	每个自剪切的 GIC 的归一化 GFP+细胞百分比
HDV-28(26)gu1 (SEQ ID 107)	-28	26	76	12.6	17
HDV-28(26)ac2 (SEQ ID 109)	-28	26	58	10.3	18

[0909]	HDV-28(28)ac2b (SEQ ID 113)	-28	28	57	9.5	17
	HDV-28(27)ac2c (SEQ ID 114)	-28	27	59	9.2	16
	HDV-28(25)ac2d (SEQ ID 115)	-28	25	56	10.9	19
	HDV-13(13)ac11 (SEQ ID 116)	-13	13	~100	2.7	2.7
	HDV-13(11)ac11b (SEQ ID 118)	-13	11	~100	4.9	4.9

[0910] 结果揭示了成功插入转基因的几个主题。首先,设计的RZ相对于天然RZ对于转基因插入的目的而言是高度有效的。其次,对于5' 侧翼rRNA序列中任何给定的RZ剪切位点(例如-28或-13),GIC:5' rRNA序列的长度有影响,其可通过包含小于最大rRNA序列而不是最大rRNA序列(例如,在RZ骨架序列“ac2”系列中,将具有26或25nt的rRNA的ac2(ac2和ac2d分别为归一化的%18或19GFP+)与具有28或27nt的rRNA序列的ac2(ac2b和ac2c分别为归一化的%17或16GFP+)进行比较)改善转基因插入。第三,RZ剪切的上游位点影响转基因插入效率(例如,在HDV-28(28)或HDV-13(13)中,当与延伸至底链切口的rRNA序列相匹配时,或者在HDV-28(26)或HDV-13(11)中,当通过在5' 模块rRNA与底链切口位点之间留出间隙来改善效率时,HDV-13RZ的5' 模块在转基因插入效率方面次于HDV-28RZ的5' 模块)。

[0911] 实施例23.GIC:5' 模块工程

[0912] 按照实施例4制备F-ZoA1 RT的RTC mRNA(SEQ ID NO 19)。按照实施例2制备包含GFP转基因表达盒(SEQ ID 506,CBhBsi_GPF_GeFo_R4A22)的GIC RNA,仅在5' 模块的序列上不同,如表20中所列举的。

[0913] 使用Lipofectamine Messenger Max用以1:3摩尔比混合的RTC mRNA和所示的GIC RNA共转染hTERT RPE-1细胞。24小时后通过FACS分析转染的细胞池以检测GFP+细胞百分比。通过FACS分析确定每个处理中的GFP+细胞的百分比,如表15所示。

[0914] 表15:工程化的GIC:5' 模块组件

[0915]

5'模块	RZ 前的 T7 RNAP 转录 物 5'前导*	GIC: 5'模 块 rRNA 序列长度	GIC: 5' RZ 自剪切 效率 (%)	GFP 阳 性细胞 百分比
HDV-28(26)gu1 (SEQ ID 107)	PP7hp	26 nt	60	3.2
HDV-28(28)gu5b (SEQ ID 121)	PP7hp	28 nt	80	4.4
HDV-28(28)NL (SEQ ID 121)	无	28 nt	0	1.8
HDV-28(28)_rzdead (SEQ ID 126)	PP7hp	28 nt	0	0.44
-28(28) (No RZ) (SEQ ID 252)	无	28 nt	0	0.022
TCARZ-28(28) (SEQ ID 67)	PP7hp	28 nt	87	3.9
TCA5-28(28) (SEQ ID 62)	PP7hp	28 nt	89	3.2
TCA5_rzdead (SEQ ID 63)	PP7hp	28 nt	0	0.29
[01] *PP7hp 表示存在用于与噬菌体 PP7 衣壳蛋白结合的共识序列的发夹茎环				

[0916] 结果支持了几个结论。首先,模板RNA中上游rRNA的存在不支持有效的转基因插入,除非其包含在有效折叠的RZ中(将5'模块“-28(28)(无RZ)”与任何有RZ活性的5'模块诸如TCA5或TCARZ或从头设计的HDV-28变体进行比较)。其次,如果T7 RNAP转录物具有5'前导序列以促进RZ自剪切,则至少一些自剪切5'模块RZ折叠序列支持更高的转基因插入效率(将HDV-28(28)NL(无前导)的转基因插入效率与HDV-28(28)gu5b中存在PP7噬菌体发夹前导序列的情况下产生的相同RZ剪切的模板RNA序列进行比较)。第三,具有RZ和前导序列的5'模块的最佳转基因插入效率需要具有催化活性的RZ(比较rzdead和有RZ活性的5'模块版本)。

[0917] 实施例24. 2-RNA组件GIS:5' 连接保真度

[0918] 按照实施例4制备RTC mRNA RTC。按照实施例2制备GIC RNA,如表16所述。

[0919] 表16:用于2-RNA组件5' 连接测定的GIC

[0920]

RTC mRNA	GIC 标识符	图 20 中的 泳道符号	5'模块来源 生物体	3'模块来 源生物体	GIC SEQ ID NO.
无		A			
青鳉 (SEQ ID 10)	TCA5_OrLa3	B	赤拟谷盗	青鳉	339
白喉带鹀 (SEQ ID 19)	TCA5_ZoAl3	C	赤拟谷盗	白喉带鹀	340
赤拟谷盗 (SEQ ID 3)	TCA5_TCB3	D	赤拟谷盗	赤拟谷盗	341
赤拟谷盗 未标记 (SEQ ID 5)	TCA5_TCB3	E	赤拟谷盗	赤拟谷盗	341
无		F			
青鳉 (SEQ ID 10)	OrLa5L_OrLa3	G	青鳉	青鳉	343
白喉带鹀 (SEQ ID 19)	OrLa5L_ZoAl3	H	青鳉	白喉带鹀	344
赤拟谷盗 (SEQ ID 3)	OrLa5L_TCB3	I	青鳉	赤拟谷盗	345
赤拟谷盗 未标记 (SEQ ID 5)	OrLa5L_TCB3	J	青鳉	赤拟谷盗	345

[0921] 所有RNA均在pH6.5的1mM柠檬酸钠的最终缓冲液中制备。6孔板的每个孔中,总RNA量固定为2.5ug。如果包含荧光蛋白的内标mRNA(spike-in mRNA)作为转染效率对照(来自Trilink的mCherry mRNA,100%5moU代替U),则将50ng的该mRNA添加到混合的RTC mRNA和GIC RNA中。

[0922] 293T细胞用RTC mRNA和GIC RNA转染,大致根据实施例7所述,但使用Lipofectamine3000而不是MessengerMax,并使用1:1摩尔比的RTC:GIC。每个RTC mRNA都用GIC RNA构建体转染,该构建体包含(i)源自赤拟谷盗(*T.castaneum*)谱系A或青鳉(*O.latipes*)的5'模块,以及(ii)源自与RT蛋白相同物种的3'模块,并且如果相关的话,源自相同物种的逆转录因子谱系(例如,赤拟谷盗(*T.castaneum*)R2谱系B组件TriCasB RT与TriCasB3' UTR“TCB”配对,不同于赤拟谷盗(*T.castaneum*)R2谱系A5'模块“TCA5”)。

[0923] 24小时后,为了提取基因组DNA,使用200ul变性RIPA缓冲液(150mM NaCl、50mM Tris pH7.5、1mM EDTA、1%TX-100、0.5%脱氧胆酸钠、0.1%SDS和1mM DTT)裂解细胞团。加入10ul核糖核酸酶A,将样品在37℃下孵育30分钟。然后加入5ul蛋白酶K,将样品在50℃下孵育过夜。加入等体积的PCI溶液(苯酚:氯仿:异戊醇25:24:1)。在涡旋(vertexing)以及5分钟的快速离心后,提取水层。加入1ul糖原(20ug/ul)、10%体积的5M氯化钠和3体积的100%乙醇。在混合并在-20℃下孵育30分钟后,将样品在4℃下离心30分钟。将基因组DNA团粒在70%乙醇中洗涤三次。风干后,将团粒溶解在TE缓冲液中。500ng基因组DNA用于插入连接的PCR测定。在PCR后,将6ul上样染料与25ul的PCR反应混合,将混合物的一半上样到在1X TAE缓冲液中含有溴化乙锭的1.2%琼脂糖凝胶的孔中。电泳后,凝胶在BioRad分子成像仪ChemiDoc XRS+上成像,如图20所示。

[0924] 上述分析表明,两种RNA组件GIS系统可以在人类基因组的预期靶位点插入全长转基因。

[0925] 与利用源自青鳉(*O.latipes*)或赤拟谷盗(*T.castaneum*)谱系B的表达的RT蛋白和GIC:3'模块RT识别序列的系统相比,利用源自白喉带鹀(*Z.albicollis*)的表达的RT蛋白和相应的GIC:3'模块RT识别序列产生了更多预期尺寸的PCR产物,并且与利用源自青鳉(*O.latipes*)的表达的RT蛋白和GIC:3'模块RT识别序列的系统相比,使用源自赤拟谷盗(*T.castaneum*)谱系B的表达的RT蛋白和相应的GIC:3'模块RT识别序列的系统产生了更多预期大小的PCR产物。

[0926] 使用具有两个GIC:5'模块组件中的每一个的GIC将每个RTC进行比较,表明来自青鳉(*O.latipes*)R2 5'区域的“OrLa5L”和来自赤拟谷盗(*T.castaneum*)R2谱系A 5'区域的“TCA5”两者均能实现全长转基因插入。这些GIC:5'模块与任何所测试的GIC:3'模块RT识别序列配对,结果没有改变。本实施例表明GIS仅递送RNA。

[0927] 实施例25. 2-RNA GIS在多种细胞系中的递送

[0928] 按照实施例4制备F-ZoA1 RT(SEQ ID NO 19)的RTC mRNA。按照实施例2制备包含GFP转基因表达盒(TCA5_CBh_NLSGPF_ZoA13_R4A22或TCA5_CBh_NLSGPF_GeFo3_R4A22,分别为SEQ ID 507和508)的GIC RNA,如表17所述。

[0929] 将SK-HEP1、293T、HCT 116、hTERT RPE-1、HeLa、Huh7、IMR-90和HaCaT人细胞系以及Cos7和Vero猴细胞系和C2C12小鼠细胞系进行培养,并按照实施例7用与GIC RNA以1:3的mRNA:模板RNA摩尔比混合的RTC mRNA共转染。24小时后,通过流式细胞术分析转染的细胞

池以检测GFP+细胞的百分比,结果在表17A和B中给出。

[0930] 表17:通过2-RNA递送GIS的转基因插入的细胞类型组

[0931]		SK-HEP1	293T	hTERT RPE-1	HeLa	HCT 116	Huh7
	GIC: 3' ZoA1	1.15 %	0.19 %	2.12 %	0.26 %	0.36 %	1.02 %

[0932] 表17B:通过2-RNA递送GIS的转基因插入的另外的细胞类型组

[0933]		IMR-90	HaCaT	hTERT RPE-1	C2C12	Cos7	Vero
	GIC: 3' GeFo	0.52 %	3.26 %	2.59 %	2.77 %	1.08 %	0.52 %

[0934] 所有群均显示至少一定百分比的表达GFP的细胞,表明RTC和GIC的两种组合在将GFP表达转基因插入受试者基因组方面至少具有最低限度的效果。此外,与源自人癌症的细胞系诸如HeLa或293T相比,在hTERT RPE-1原代人细胞系中观察到相对高的GFP+细胞百分比。

[0935] 进行了另外的实验,所述试验证明2-RNA GIS在多种细胞系中的递送。编码F-ZoA1 RT(用N1甲基假尿苷制成)的RTC mRNA分别与两种不同的GIS RNA模板共转染:i) 5' TCA5_RNA PIterm1_sylac0_CBh启动子_eGFP_SV40LPA_sylac0_GeFo3_R4A22,由常规尿苷核苷酸组成,或ii) 5' TCARZ_CMV*启动子_eGFP_minpA_GeFo3_R4A22,包含用于表达转基因RNA的修饰CMV启动子和包含假U核苷酸。在2-RNA递送后的第1天(或第1天和第3天),通过流式细胞术确定转基因的表达。编码mCherry (TriLink)的mRNA被共转染,作为一种比较相对于%细胞GFP+的总体转染效率的方法。结果显示在下表17C和17D中。

[0936] 表17C:使用包含常规尿苷核苷酸的RNA模板5' TCA5_RNAPIterm1_sylac0_CBhpromoter_eGFP_SV40LPA_sylac0_GeFo3_R4A22,通过2-RNA递送GIS插入转基因的另外的细胞类型组。

[0937]	细胞系	第 1 天		第 3 天
		GFP%	mCherry%	GFP%
	RPEhTERT	20.71	92.657	18.89
	ARPE19	19.64	91.32	17.1
	293T	0.21	90.866	2.34

[0938]

HaCat	3.74	84.801	2.47
Hela	0.92	62.78	0.77
Huh7	0	98.12	11.68
IMR90	6.07	75.12	8.51
MRC5	5.42	82.99	5.7
Cos7	3.41	95.18	4.66
Vero	1.91	91.938	2.38
C2C12	9.26	96.98	5.69
G8	1.9	84.338	1.04
C26	1.03	77.744	1.26

[0939] 表17D:使用包含用于表达转基因RNA的修饰的CMV启动子并且包含假U核苷酸的RNA模板5' TCARZ_CMV*启动子_eGFP_minpA_GeFo3_R4A22,通过2-RNA递送GIS插入转基因的另外的细胞类型组。

[0940]

细胞系	GFP% 平均	标准偏差 GFP	GFP 中位强度	mCherry%平均	标准偏差 mCherry
RPE	61.31	0.8386	44537	90.61333	0.1528
ARPE-19	53.57	0.3512	61436	90.57333	0.2517
293T	9.567	0.2003	3511	74.125	0.1732
Hela	11.52	0.1	5827	51.47667	0.6506
IMR90	38.01	0.0577	27271	66.22	0.755
MRC5	40.52	0.1155	28822	71.65333	0.5859
Vero	10.06	0.3	4071	83.20333	0.6429
C2C12	30.53	1.701	5560	78.00667	1.5044

[0941] SD=标准偏差

[0942] 上述数据表明2-RNA递送在来自人、猴和小鼠的多种细胞类型中起作用。数据还表明,修饰的CMV启动子和假U核苷酸的组合增加表达转基因的细胞的百分比。

[0943] 实施例26.RTC和GIC组合

[0944] 培养hTERT RPE-1细胞系,并根据实施例9所述以1:3的RTC与GIC之比用ZoA1 RT mRNA、ZoA1 RT-dead mRNA或TaGu RT mRNA RTC (分别为SEQ ID 19、24和28)之一和TCA5_ZoA13、TCA5_GeFo3或TCA5_TaGu3 GICs RNA (分别为SEQ ID 509、503、510)之一转染。

[0945] 5天后,收获并如前所述对细胞群计数,并确定GFP阳性细胞的百分比和GFP阳性细胞群的中位强度,并在表18中报告。

[0946] 表18:RTC和GIC组合

[0947]

RTC	GIC	GFP阳性细胞百分比(%)

F-ZoA1 RT mRNA	TCA5_TaGu3	2.38
F-TaGu RT mRNA	TCA5_TaGu3	3.56
F-ZoA1 RT mRNA	TCA5_ZoA13	11.75
F-TaGu RT mRNA	TCA5_ZoA13	13.28
F-ZoA1 RT mRNA	TCA5_GeFo3	11.71
F-TaGu RT mRNA	TCA5_GeFo3	13.87

[0948] 给予的RTC(ZoA1 RT mRNA或TaGu RT mRNA)与GIC TCA5_ZoA13或TCA5_GeFo3的任何组合都导致表达GFP的细胞百分比显著更高。这表明,具有源自白喉带鹀(*Z.albicollis*)或中嘴地雀(*G.fortis*)的3' 模块RT识别序列的GIC优选与源自白喉带鹀(*Z.albicollis*)或栗耳草雀(*T.guttata*)的RTC:RT模块配对,以实现更高的转基因插入百分比。此外,所有组合都导致稳定的插入(如通过PCR检测5' 和3' 连接插入位点确定的)和转基因表达。ZoA1 RT-dead mRNA与任何GIC构建体的组合都没有导致GFP荧光高于背景。

[0949] 实施例27.细胞系的RTC与GIC比率和3' 模块

[0950] A部分

[0951] 如上所述,培养hTERT RPE-1、SK-HEP1和HeLa人细胞系,并用ZoA1 RT mRNA RTC和TCA5_ZoA13或TCA5_GeFo3 GICs RNA转染。

[0952] 5天后,收获群并如前所述对其计数。表19显示表达eGFP的细胞的百分比(%)。

[0953] 表19:RTC与GIC比率

细胞系	GIC	RTC 与 GIC 的比率					
		无 RTC	1: 1	1: 3	1: 5	1: 8	1: 10
[0954] hTERT RPE-1	TCA5_ZoA13	0.01%	2.47%	2.8%	2.63%	N.A.	2.3%
hTERT RPE-1	TCA5_GeFo3	0.04%	1.96%	2.48%	2.57%	2.34%	2.28%
SK-HEP1	TCA5_ZoA13	0.04%	0.38%	0.58%	0.62%	0.64%	0.7%

[0955] B部分

[0956] 如上述培养、转染、收获和分析SK-HEP 1和HeLa细胞系,并在表20中进行了描述。RTC与GIC的比率如表20所示变化。

[0957] 表20:RTC与GIC的比率

细胞系	GIC	RTC 与 GIC 的比率					
		No RTC	3: 1	2: 1	1: 1	1: 2	1: 3
[0958] SK-HEP 1	TCA5_ZoA13	0.09	1.07	1.58	2.44	3.23	3.60
HeLa	TCA5_ZoA13	0.04	0.15	0.20	0.26	0.27	0.32

[0959] 表20B显示使用培养的hTERT RPE-1人细胞和用F-TaGu mRNA RTC和F-ZoA1 mRNA RTC(两者均用5moU制成)以及如上所述的TCA5_ZoA13或TCA5_GeFo3 GICs RNA转染的类似实验的结果。

[0960] 表20B:RTC与GIC比率

[0961]

RTC mRNA/GIC RNA	摩尔比:	GFP%	GFP 强度
TaGu/ZoAl3	10: 1	6.27	3846
	3: 1	11.26	5562
	1: 1	14.36	6412
	1: 3	14.36	6347
	1: 6	14.76	6746
	1: 12	13.36	6156
	1: 20	11.26	5773
TaGu/GeFo3	10: 1	4.5	3405
	3: 1	9.1	4330
	1: 1	12.21	4841

[0962]

	1: 3	13.91	5146
	1: 6	13.31	5413
	1: 12	13.51	5600
	1: 20	11.61	5323
ZoAl/ZoAl3	10: 1	1.89	3014
	3: 1	5.35	4540
	1: 1	9.96	5676
	1: 3	11.46	6347
	1: 6	11.06	6521
	1: 12	9.96	5911
	1: 20	7.82	5487
ZoAl/GeFo3	10: 1	0.77	3014
	3: 1	4.25	3393
	1: 1	8.12	4330
	1: 3	10.31	5233
	1: 6	10.71	5146
	1: 12	9.91	5146
	1: 20	7.84	4744

[0963] 产生最有效转基因插入的RTC与GIC的比例有所变化,但最佳的是是GIC RNA比RTC RNA多的摩尔比。

[0964] 这些结果表明,2组件GIS将转基因插入特定受试者的理想比例可能需要通过实验确定,而不是根据组件或受试者身份预测。对于旨在给予hTERT RPE-1细胞的GIS,其包含RTC和GIC,所述RTC包括源自白喉带鹀(*Z.albicollis*)的RT模块,所述GIC包括源自白喉带

鹱(*Z.albicollis*)的GIC:3' 模块RT识别序列,1:3(RTC:GIC)的比例可以是优选的。对于旨在给予hTERT RPE-1细胞的GIS,其包含RTC和GIC,所述RTC包含源自白喉带鹱(*Z.albicollis*)的RTC:RT模块,所述GIC包含源自中嘴地雀(*G.fortis*)的GIC:3' 模块RT识别序列,1:5(RTC:GIC)的比例可以是优选的。对于旨在给予SK-HEP1或HeLa细胞的GIS,其包含RTC和GIC,所述RTC包含源自白喉带鹱(*Z.albicollis*)的RTC:RT模块,所述GIC包含源自白喉带鹱(*Z.albicollis*)的GIC:RT识别序列,1:3(RTC:GIC)的比例可以是优选的。

[0965] 实施例29:转基因表达的耐久性

[0966] 按照实施例4制备编码F-ZoA1 RT(SEQ ID NO 19)的RTC mRNA。按照实施例2制备包括GFP转基因表达盒TCA5_CBh_NLSGFP_ZoA13(SEQ ID NO 507)的GIC RNA。

[0967] 将RTC和GIC构建体共转染到实施例7中描述的293T细胞培养物中,并在转染后第3天进行分选以富集GFP+细胞,1天后使用Fusion Aria分选板支架将其分选以将单个GFP阳性细胞分离到96孔板的单个孔中。在增殖约3周后,筛选单个孔中的活的GFP阳性细胞系,然后将其转移到主24孔板中并每周分裂两次。通过具有GFP荧光强度的单一峰分布(图21),37个细胞系被认为是克隆的。每个细胞系具有不同的绝对GFP强度,与GFP阴性克隆细胞系明显不同(图21)。在连续培养期间,大约每周使用Attune NxT声学聚焦细胞仪筛选细胞的等分试样。在作为克隆细胞系传代2个月后,自初次转染已近3个月,37个中的仅一个显示GFP强度有所下降,而且下降为仅约50%。

[0968] 这些结果表明,通过本发明的GIS插入哺乳动物细胞基因组的转基因可以稳定表达3个月或更长时间。

[0969] 实施例30:多个转基因的插入和表达

[0970] 按照实施例4制备F-ZoA1 RT(SEQ ID NO 19)的RTC mRNA。按照实施例2制备具有GFP转基因表达盒(TCA5_CBhBsi_GFP_GeFo3,SEQ ID NO 503)和mCherry转基因表达盒(TCA5_CBhBsi_mCherry_GeFo3,SEQ ID NO 512)的GIC RNA。

[0971] 用RTC mRNA和2个GIC构建体之一或两者的等量混合物共转染hTERT RPE-1细胞,其中RTC mRNA与总GIC模板RNA的摩尔比为1:3。对于对照,一些细胞未转染(阴性对照),仅用RTC转染(RTC对照),或仅用GFP或mCherry GIC转染(仅GFP和mCherry模板的对照)。细胞还用RTC和三种GIC:GFP、mCherry之一或两者的等量混合物转染。24小时后,通过流式细胞术测定细胞的GFP和mCherry表达。表达预期转基因产物的细胞的百分比记录在表21中。

[0972] 表21:2种转基因的插入和表达

[0973]	转染的组分	仅 GFP 阳性的细胞百分比	仅 mCherry 阳性的细胞百分比	细胞 GFP 和 mCherry 阳性百分比
	无	0.0041	0.0055	0.0014
	仅 RTC	0.026	0.024	0
	仅 GFP GIC	0.043	0.0020	0.0061
	仅 mCherry GIC	0.024	0.010	0.017
	RTC + GFP GIC	21.7	0.29	0.044
	RTC + mCherry GIC	0.3	15.3	0.038
	RTC + GFP 和 mCherry GIC	5.43	3.54	8.94

[0974] 这些结果表明,本发明的GIS可以将包含在单个GIC中的多于一个的转基因插入到受试者基因组中,使得两个转基因都可以由受试者细胞表达。作为推论,可以使用单个GIC将多个转基因插入基因组中,导致受试者细胞的有效载荷表达水平更高。如果不希望有多个转基因拷贝,则转基因有效载荷可包含负反馈机制,使用本领域技术人员已知的策略,在第一个转基因插入之后停止另外的转基因插入。

[0975] 进行了另外的实验,其中将两种不同的GIC模板RNA混合在一起并转染到细胞中,或者使用编码两种不同转基因的单个GIC模板RNA(称为串联模板)。用编码F-ZoAl RT (SEQ ID 19, 包含5moU) 或RT催化失效的 (dead) ZoAl (“ZoAl RTD” SEQ ID 23, 包含N1甲基假尿苷) 的RTC mRNA和单个转基因模板TCARZ_SV40*_GFP_GeFo3 (SEQ ID NO:529) 和TCARZ_CMV*_mCherry_GeFo3 (SEQ ID NO:531) 或串联模板TCARZ_SV40*_GFP_minPA_CMV*_mCherry_SV40LPA_GeFo3 (SEQ ID NO:533) 共转染细胞。结果显示在下表21B中。eGFP+mCherry mRNA是阳性对照。

[0976] 表21B.mCherry、eGFP或mCherry和eGFP两者均为阳性的细胞百分比。

[0977]	转染的组分	仅 mCherry 阳性的细胞百分比	仅 GFP 阳性的细胞百分比	细胞 GFP 和 mCherry 阳性百分比
	ZoAl RTD + CMV-mCherry	0.01 ± 0.004	0.01 ± 0.002	0.0004 ± 0.0004
[0978]	ZoAl RTD + SV40-eGFP	0.06 ± 0.006	0.005 ± 0.003	0.005 ± 0.002
	ZoAl RTD + 串联 SV40-eGFP_CMV-mCherry	0.005 ± 0.002	0.03 ± 0.003	0.008 ± 0.004
	ZoAl + CMV-mCherry	69.4 ± 0.88	0.02 ± 0.0009	0.06 ± 0.006
	ZoAl + SV40-eGFP	0.002 ± 0.0009	68.0 ± 0.38	0.04 ± 0.005
	ZoAl + SV40-eGFP + CMV-mCherry	7.77 ± 0.54	3.3 ± 0.23	52.2 ± 2.8
	ZoAl + 串联 SV40-eGFP_CMV-mCherry	12.5 ± 0.20	0.6 ± 0.02	45.6 ± 0.73
	eGFP + mCherry mRNA	0.4 ± 0.05	0.3 ± 0.2	97.4 ± 0.2

[0979] 平均±SEM, n=3。

[0980] 数据表明,两种不同的转基因RNA可以被成功地插入同一细胞,并且两种不同的转基因RNA可以成功地以同一GIC模板RNA递送。

[0981] 实施例31:通过GIS募集内源性修复机制

[0982] A部分-通过RNA干扰敲低MUS81

[0983] 按照实施例4制备F-ZoAl (SEQ ID NO 19) 的RTC mRNA。按照实施例2制备包括GFP转基因表达盒TCA5_CBhBsi_GFP_GeFo3 (SEQ ID NO 503) 的GIC RNA。表22中描述的经验证的抗MUS81siRNA和抗MSH2siRNA购自ThermoFisher Scientific。沉默基因阴性对照1号 (Silencer Select Negative Control No.1) siRNA购自Invitrogen。

[0984] 表22:siRNA双链设计

[0985]	siRNA ID	靶基因	正义序列	反义序列
	s37038	MUS81	CGCGCUUCGUAUUUCAGAAtt (SEQ ID NO: 553)	UUCUGAAAUACGAAGCGCGtg (SEQ ID NO: 554)
	s37039	MUS81	UGACCUCUCCAAACCCUCUtt (SEQ ID NO: 555)	AGAGGGUUUGGAGAGGUCAtg (SEQ ID NO: 556)
	s37040	MUS81	GGGAGCACCUGAAUCCUAAtt (SEQ ID NO: 557)	UUAGGAUUCAGGUGCUCCCgg (SEQ ID NO: 558)
	s8966	MSH2	GGAUAUUACUUUCGUGUAAtt	UUACACGAAAGUAAUAUCCaa
[0986]			(SEQ ID NO: 559)	(SEQ ID NO: 560)
	s8967	MSH2	CGUCGAUUCCCAGAUCUUAAtt (SEQ ID NO: 561)	UAAGAUCUGGGAAUCGACGaa (SEQ ID NO: 562)
	s8968	MSH2	GAAUCGCAAGGAUAUGAUAtt (SEQ ID NO: 563)	UAUCAUAUCCUUGCGAUUCtc (SEQ ID NO: 564)

[0987] 每条siRNA双链均有退火的正义和反义链,小写字母表示突出部分。每次siRNA处理混合三条siRNA双链。

[0988] 通过将两个管组合在一起制备用于转染的siRNA混合物:一个管含有与37.5μl的Lipofectamine 3000混合的625μl的OptiMEM(Gibco),一个管含有与375pmol siRNA混合的625μl的OptiMEM。将针对任何靶点的三种不同siRNA汇集在一起,并使用375pmol的沉默基因选择阴性对照1号siRNA(Invitrogen)作为阴性对照。

[0989] 孵育10分钟后,将1.25ml的siRNA-脂质复合物混合物加入板中,然后加入约450万个hTERT RPE-1细胞(相当于贴上时约75%的汇合),使孔中的培养基总体积达到10ml(最终浓度为37.5nM siRNA)。24小时后,将细胞以1:3分开,在siRNA引入后2天达到约60%的汇合,这时然后用2-RNA组合转染它们。进行qRT-PCR以测量转染后72小时的靶mRNA敲低效率。

[0990] 首先用抗MUS81、抗MSH2 siRNA或乱序的siRNA转染hTERT RPE-1细胞以用作对照。一(1)或两(2)天后,细胞或者不用GIS转染(阴性对照),或者仅用GIC转染,或者如上所述用RTC和GIC共转染。

[0991] 最后的转染后24小时,收获细胞,并根据实施例8所述,通过FACS分析确定表达GFP的细胞百分比,并在表23中报告。

[0992] 表23:内源性修复敲低对GIS功能的影响

[0993]	转染的 siRNA	转染的 GIS	转染之间的天数	GFP 阳性细胞百分比
	乱序的	无	1	0.0016
	乱序的	无	2	0.073
	乱序的	仅 GIC	1	0.029
	乱序的	仅 GIC	2	0.028
	乱序的	RTC + GIC	1	4.57

[0994]

乱序的	RTC + GIC	2	1.48
siMSH2	无	1	0.024
siMSH2	无	2	未测试
siMSH2	仅 GIC	1	0.01
siMSH2	仅 GIC	2	0.017
siMSH2	RTC + GIC	1	4.36
siMSH2	RTC + GIC	2	1.72
siMUS81	无	1	0.043
siMUS81	无	2	0.034
siMUS81	仅 GIC	1	0.021
siMUS81	仅 GIC	2	0.06
siMUS81	RTC + GIC	1	3.06
siMUS81	RTC + GIC	2	0.22

[0995] B部分-MUS81缺陷型细胞系中的GIS活性

[0996] 按照实施例4制备F-ZoA1 RT (SEQ ID NO 19)的RTC mRNA。按照实施例2制备包括GFP转基因表达盒TCA5_CBhBsi_GFP_GeFo3 (SEQ ID NO 503)的GIC RNA。

[0997] 按照先前描述,共转染野生型或MUS81阴性突变体HTC 116细胞系。转染后24小时收获细胞,并通过FACS分析确定表达GFP的细胞的百分比,如表24所报告的。

[0998] 表24:MUS81阴性细胞系中的GIS活性

[0999]

细胞系	转染的GIS	GFP阳性细胞百分比
HCT116 MUS81+	无	0.00677
HCT116 MUS81+	无	0.028
HCT116 MUS81+	仅GIC	0.26
HCT116 MUS81-	仅GIC	0.01
HCT116 MUS81-	RTC+GIC	0.00721
HCT116 MUS81-	RTC+GIC	0.036

[1000] 这些结果表明,MUS81活性是通过本发明的GIS的转基因插入最大效率所需要的。因此,本文所述的GIS可以招募内源性基因组修复机制(例如,MUS81)以实现成功的转基因插入。

[1001] 鉴于MSH2活性的丧失,另一种已知在基因组修复中起作用的酶未显著阻碍GIS的转基因插入率,本发明的GIS可以选择性地招募用于转基因插入的MUS81。应该注意的是,已知MUS81在任何天然逆转录因子或转基因插入机制中不发挥作用。

[1002] C部分:RNA干扰敲低与报告基因共转染

[1003] 按照实施例4制备RTC mRNA (SEQ ID NO F-ZoA1 RT19)。按照实施例2制备包含GFP转基因表达盒TCA5_CBhBsi_GFP_GeFo3 (SEQ ID NO 503)的GIC RNA。购买编码mCherry的

mRNA (TriLink#L-7203)、抗MUS81siRNA (ThermoFisher Silencer Select ID编号s37038、s37039、s37040的等量混合物) 和沉默基因选择阴性对照1号siRNA (Invitrogen)。

[1004] 首先用抗MUS81或阴性对照siRNA转染hTERT RPE-1细胞。两(2)天后,细胞或者不用GIS转染(阴性对照),或者仅用GIC转染,或者用RTC、GIC和mCherry mRNA共转染。所有转染均使用Lipofectamine MessengerMax进行。mCherry mRNA被设计成通过经典的帽依赖性mRNA翻译(即,无需GIS活性)来翻译mCherry,并在GFP插入效率降低时用作转染效率的对照。

[1005] 最后的转染后24小时,收获细胞并通过FACS分析确定表达GFP和mCherry的细胞的百分比,如表25中所报告的(括号中包括相对于背景的GFP阳性细胞百分比,在适用的情况下)。

[1006] 表:25:MUS81的siRNA敲低

[1007]	转染的 siRNA	转染的构建体	GFP 阳性细 胞百分比	GFP 中 位强度	mCherry 阳 性 细胞的百分比	mCherry 中位强度
	乱序的	无	0.00279	1968	0.00558	187
	乱序的	仅 GIC	0.025	2299	0.034	321
	乱序的	RTC + GIC + mCherry(mRNA)	3.8	8126	90.666	2307
	siMus81	无	0.069	2163	0.05	291
	siMus81	仅 GIC	0.17	2331	0.22	337
	siMus81	RTC + GIC + mCherry(mRNA)	0.39	2705	88.78	2460

[1008] 这些结果证实,在部分A、B和C中可重复观察到的MUS81缺失或耗竭的细胞中GFP转基因表达急剧下降不是由于MUS81敲低对用RNA转染细胞的能力或含有GIS的hTERT RPE-1细胞翻译转染mRNA的能力的任何影响。

[1009] 实施例32. 具有不同启动子的模板模块

[1010] 培养hTERT RPE-1细胞,并用F-ZoA1 RT mRNA RTC19)转染,GIC含有具有不同表达环境(context)的GFP ORF+/-N末端核定位序列(NLS) (SEQ ID 513-517)。测试的转录启动子包括CBh、EFS和mPGK (SEQ ID 400-402或408-409)。有效载荷盒转录的方向或者与RNAPII同向,或者是与RNAPII转录汇聚的反向“翻转”方向;“翻转”方向还将RNAPII转录终止信号盒的定位从RNAPII启动子上游移除。

[1011] 在转染后第1天和5天通过FACS监测GFP合成(表26A)。几个比较特别有趣。首先,同向取向的CBh_GFP或CBh_NLSGFP和汇聚取向的[CBh_NLSGFP]翻转在转染后第1天具有相似的%GFP细胞,但4天后,汇聚取向的[CBh_NLSGFP]翻转GFP%细胞减少,而同向取向的转基因的GFP信号保持高。这表明,RNAPII表达盒之前的同向转录和/或RNAPII转录终止信号有利于持续的转基因表达,而当期望瞬时表达时,翻转环境是有利的。其次,使用mPGK和EFS启动子可检测的GFP转基因表达表明,不同的启动子可用于生产性转基因表达。

[1012] 表26A:最佳表达的转基因启动子和环境

[1013]	启动子和 ORF	GIC SEQ ID	第 1 天的 GFP 阳性细胞 百分比	第 5 天的 GFP 阳性细胞 百分比
	CBh_GFP	513	4.041	4.46
	CBh_NLSGFP	514	3.192	2.2
	[CBh_NLSGFP]翻转	515	3.934	0.6
	mPGK_GFP	516	0.614	0.16
	EFS_GFP	517	0.963	0.57

[1014] 使用含有其它转基因转录启动子的GIC进行了另外的实验。测试了具有CpG突变和 neo3 5' UTR(CMV*,SEQ ID NO 408)的修饰的巨细胞病毒启动子,并测试了具有改善的TATA盒(SV40*,SEQ ID NO 409)的修饰的猿猴病毒40启动子。这些用于GIC以插入GFP表达转基因。hTERT RPE-1细胞与ZoA1 RTC mRNA和GIC构建体之一共转染,RTC mRNA与总GIC模板RNA的摩尔比为1:3。24小时后,通过流式细胞术测定细胞的GFP表达。表达预期转基因产物的细胞的百分比显示在表26B中。

[1015] 表26B:用于最佳表达的转基因启动子

[1016]	启动子_报告蛋白	GIC SEQ ID	常规 U(U)	第 1 天的 GFP+ 细胞百分比	第 1 天的 mCherry+ 细胞百分比
	CBh_GFP	513	U	20.7	n.a.
	CMV*_GFP	528	U	50.7	n.a.
	SV40*_GFP	529	U	44.8	n.a.
	CBh_mCherry	512	U	n.a.	19.9
	CMV*_mCherry	531	U	n.a.	33.5
	SV40*_mCherry	532	U	n.a.	16.6

[1017] 实施例33. 从基因组DNA插入的转基因测序以确定插入位点特异性

[1018] 按照实施例4制备F-ZoA1 RT(SEQ ID NO 19)或F-TaGu RT(SEQ ID NO 28)的RTC mRNA。按照实施例2制备含有5' 模块TCA5(TCA5_CBhBsi_GFP_GeFo3,SEQ ID NO 503)或5' 模块TCARZ(TCARZ_CBhBsi_GFP_GeFo3,SEQ ID NO 526)的GFP转基因表达盒的GIC RNA。

[1019] 用RTC mRNA和GIC RNA共转染了hTERT RPE-1细胞,RTC mRNA与GIC模板RNA的摩尔比为1:3。24小时后,根据实施例8所述对细胞进行分选以富集GFP+群。根据实施例24所述收获富集的GFP+细胞进行基因组DNA纯化。加州大学伯克利分校功能基因组学实验室(University of California,Berkeley Functional Genomics Laboratory)和 Vincent J.Coates基因组学测序实验室(Vincent J.Coates Genomics Sequencing Laboratory)分别提交了1ug的DNA进行标准文库制备和Illumina全基因组鸟枪法(WGS)测序。使用Kapa Hyper Prep试剂和独特的双索引Y型接头用1个PCR循环进行人WGS制备。在NovaSeq 6000 S4上以30倍覆盖率进行测序,使用150bp配对末端读段。

[1020] 在修剪接头后,将读段映射到包含转基因序列的自定义重叠群(custom contig)

上。任何具有唯一地被映射到自定义重叠群 (SEQ ID NO 360) 的转基因序列区域并且还具有读段的未映射部分 (“剪切” 部分) 的区域的读段都被评价为转基因和基因组的候选连接序列。首先将候选转基因3' 连接读段映射到转基因序列, 所述转基因序列的侧翼是精确预期下游靶位点 (SEQ ID NO 361), 以对 “靶位点处” 插入 (绝大多数) 计数。然后将任何与精确靶位点不匹配的候选3' 连接的剪切区域映射到整个人rDNA共识骨架上, 以对不精确连接但仍以rDNA为靶点的插入 (“rDNA但不是精确的靶位点”) 计数。

[1021] 任何未映射到rDNA的剪切区域都被映射到人类基因组组装体GRCh39。来自ZoA1 RTC转染的候选脱靶插入连接读段 (“不确定的”) 没有插入的转基因3' 端标志, 表明它们是在广泛测序文库扩增过程中人为的序列重排。没有明显的脱靶插入位点。来自TaGu RTC转染的七个候选脱靶插入连接读段将预期的转基因3' 端连接到除rDNA之外的人类基因组序列, 得到的最大脱靶插入频率小于1%。

[1022] 表27. 基于基因组测序的插入位点特异性

	RTC mRNA (seq ID)	GIC SEQ ID	在靶位点	rDNA 但不是精确的靶位点	不确定的 (文库生产人工产物或脱靶)
[1023]	ZoA1 (19)	503	531	1	1
	ZoA1 (19)	526	1033	3	1
	TaGu (28)	526	964	5	7

[1024] 实施例34. 具有尿苷类似物的RTC mRNA和GIC RNA

[1025] 使用尿苷或修饰的尿苷核苷酸根据实施例4制备F-ZoA1 RT (SEQ ID NO 19) 的RTC mRNA。使用尿苷或修饰的尿苷核苷酸根据实施例2制备具有GFP转基因表达盒的GIC模板RNA。每个实验的RNA包含100%的所列的尿苷类似物, 或者如果列出了两种尿苷, 则是每种50%的混合物。下表显示用2个不同的RNA转染的结果, 一个是ZoA1 RT的mRNA, 另一个是具有GFP转基因表达盒的GIC模板RNA。转染后第1天收获细胞, 并通过流式细胞术测定GFP阳性细胞的百分比。

[1026] 表28显示包含所示尿苷类似物的F-ZoA1 mRNA和具有未修饰尿苷 (尿苷三磷酸核糖核苷酸 “regU”) 的GIC模板RNA TCA5_CBhBsi_GFP_GeFo3_R4A22 (SEQ ID 503) 的数据。

[1027] 表28. 具有尿苷类似物的ZoA1 RT的RTC mRNA

[1028]

ZoA1 mRNA 尿 苷核苷酸	GFP%平均值	标准偏差	GFP 中位强度平均值	标准偏差
regU	7.17633333	0.79286401	5645.33333	133.881789
regU:N1mpsU 50: 50 混合物	13.2296667	0.56862407	9568.66667	520.66528
N1mpsU	13.2296667	0.66583281	9354	733.130957
N1mpsU:psU 50: 50 混合物	12.2963333	0.45092498	9160.33333	580.542275
psU	12.463	0.43588989	9086.33333	933.837423
5mU	11.163	0.7	7338.33333	354.425357
5moU	12.9296667	0.37859389	8715.66667	177.902595

[1029] 缩写:尿苷三磷酸核糖核苷酸(regU)、5-甲氧基尿苷三磷酸核糖核苷酸(5moU)、5-甲基尿苷三磷酸核糖核苷酸(5mU)、假尿苷三磷酸核糖核苷酸(psU)、N1-甲基假尿苷三磷酸核糖核苷酸(N1mpsU)。

[1030] 表29显示包含5moU的F-ZoA1 mRNA和包含所示尿苷类似物的GIC模板RNA TCA5_CBhBsi_GFP_GeFo3_R4A22(SEQ ID 503)的数据。

[1031] 表29.具有尿苷类似物的GIC模板RNA

[1032]

GIC 模板 RNA 尿 苷核苷酸	GFP%平均值	标准偏差	GFP 中位强度平均值	标准偏差
regU	12.9296667	0.37859389	8715.66667	177.902595
5moU	1.04333333	0.0321455	1171.66667	41.0528115
5mU	17.81	0.26457513	5845.66667	113.160653
psU	41.44	0.45825757	6311.33333	86.3153134
N1mpsU	30.1433333	0.75055535	3959.66667	307.034743

[1033] 表30显示用N1甲基假尿苷和六种包含psU的不同GIC模板RNA(表达GFP或mCherry的转基因,每种都具有CBh、CMV*或SV40*启动子)制成的ZoA1 mRNA(SEQ ID21)的数据,其中SEQ ID如表中所示。这些结果是与表26B中的结果并行确定的。比较这两个表表明,使用psU模板的转基因递送效率比使用常规U模板更好。

[1034] 表30:编码不同启动子的GIC模板RNA受益于假尿苷。

[1035]

启动子_报告蛋白	GIC SEQ ID	假 尿 苷 (psU)	第 1 天的 GFP+ 细胞百分比	第 1 天的 mCherry+ 细胞百分比
CBh_GFP	513	psU	38	n.a.
CMV*_GFP	528	psU	81.9	n.a.
SV40*_GFP	529	psU	65	n.a.
CBh_mCherry	512	psU	n.a.	42.2
CMV*_mCherry	531	psU	n.a.	70.1
SV40*_mCherry	532	psU	n.a.	51.9

[1036] 结果表明,当编码RT蛋白的RTC mRNA包含修饰的尿苷核苷酸时,观察到转基因表达的增加。同样,当GIC模板RNA包含修饰的尿苷核苷酸时,当尿苷为psU或N1mpsU时观察到转基因表达的增加。

[1037] 实施例35.具有截短的GeFo 3' UTR的包含尿苷类似物的GIC3' 模块RNA增加表达转基因的细胞的频率。

[1038] 本实施例表明,具有截短的GeFo 3' UTR和包含尿苷类似物的模板RNA的GIC 3' 模块增加转基因表达的频率。用5moU合成了F-ZoA1 mRNA (SEQ ID19),用常规U或假U合成了GIC模板RNA (TCARZ_CBh_GFP_GeFo3_R4A22,SEQ ID 526)。GIC模板RNA包含全长3' UTR (GeFo3,SEQ ID NO 204)或三个不同的截短的3' UTR (GeFo217,SEQ ID NO 222;GeFo98,SEQ ID NO 223;和GeFo68,SEQ ID NO 224)。结果显示在下表31中。

[1039] 表31.使用包含截短GeFo 3' UTR和假U的GIC模板RNA的转基因表达。

[1040]

GIC 3' UTR	GFP%	GFP 中位强度
GeFo3 regU	8.54	2855
GeFo217 regU	9.7	3129
GeFo98 regU	10.4	3846

[1041]

GeFo68 regU	9.33	3487
GeFo3 psU	25.86	2642
GeFo217 psU	28.16	2875
GeFo98 psU	29.26	3558
GeFo68 psU	17.96	2001

[1042] 数据表明,与全长3' UTR相比,包含截短的3' UTR的GIC RNA模板增加了表达功能性转基因蛋白的细胞的频率。数据还表明,与用regU合成的模板相比,包含假U的GIC RNA模板增加了表达功能性转基因蛋白的细胞的频率。

序列表

如果 RNA 序列包含 “T”，则实际的核糖核苷酸将是 “U”。

1. R2A TriCasB* N-DYK 质粒 ORF 序列:[* = 自天然 aa 修饰]

ATGGATTACAAGGATGACGACGATAAGGGTACCCTGCCCTTTTCAGAGCAGAAGCTGCGGCATCTGTCT
GAATGCCGGTAAAGGTAACTTCCGCGCGCTGAGCCTGGACGACGAAGAACGGCACCTGCGTGAACG
GCATCCACTGTCCCTGATCCTCTATAAATGCAGCGATTGCAAGGGCCAGTACAGATCCAAGAGGGCCG
CCCTGTGCCACGCCCCCAAGTGCACCGGACCGACCCCTGATCCTCAAGGCAATGCCCTGCGTTGTCAT
CTGTGTGGTCTTGTGTTGTAAGCCAGAGTGGTGTACCCAGCATTTACGTCATAGACACCCTCTGGT
CAGAAACACCCAGCGGGCAGCTGAAGAAAGCGGTAGAGCCGAACGTGCTGCACTGCCTCGGCCTCT
GCGTCGTAACACCCGTTTCGGTTCAGCGAAGAGGATGAAGCAAAAATGCTGGAGCTGGAAGTGC GG
TTCCAGAACGAGCGTTGTGTTGCAAAATGCATGCTGCCGTTTTTTCCGAATAGAACTTGCAAGCAGAT
CCGTGATAAGCGTAATACCGATGCATATAACGGAGAAGAGAACTGTACTTCGAGGGCGTCCGGGTGC
AGGACCCTGCAGGCGCCGAGGACAGCGTTCTCCCTGTTGTGAAACCGACGAACCCGCCGAGGAAA
ATATCCGCTGGAGTACCCGAGCTGCCTGGCGATGAAGAGGGTGCTCCTGCCTGCAGCCAGACTATT
[1043] CTGAACACAGAAGGTCCGGATGGACTGGGCAGCCCACCGGTGCCC GTTGAAGAAGAAATGGCAAGT
TCGGGTAGCACATCTAATAACGTGGATAACCGTTGGAGAGAAAAGCATTATCACAGCTGCACTCGGCGT
TGAAATTCCGAAAGCAATCAGCCAAGAGCCCGCCGCGTTATTTCAGGAGCTGCAGGATGCTCTGCGC
GAGGCCGTGATCGGCGTGTTCGCGCAGGATCGCCTGGACGAGATGTACGAGCGGGTACTCAAAGTGG
TCAACCCGGATGATACACAGGAACGCCCCGAAACGTCAAAGAAAGAAAGGCAAGTCTCGTAATGCCTT
CCGCCGTTATGTGTACAGCCAGACCCAGGACCTGTTCAAGAAAAATCCTGGACAGCTGGCACGTTAT
GTTAGGGAAGATGTGAGATGGCTGGAACAGGGCCGGGTGCAGTTGCAGAGAGATGATATTGAAAGAA
TGTACAACAAGCTGTGGGGCACCAAGCCGGATGTGCTGCCTCCCCACTGGGATTATCCACTGCCACTC
GACACCGCTGATGTTCTGACCCCGATTGAGCTGAAAGAAGTCCGGAAGAAGATATCTCAGACGAAAC
TCAAATCTGCAGCAGGGCCCCGACGGTCTGCAGAAAAGACATCTGGTGCCTGCGGTTGTGCAAGAAAT
TCTGCGCCTGCTGTATAACCTGCTGATGTGTTGTGCAATGCAGCCTACACAGTGGCGGATGAACCGTA
CCCAACTGTTACTGAAGCAGGGTAAAGATCCTCTGGATGTCGCTAGTTATCGTCCTATAACCATCTCCA
GCATCCTTTGTCGCTGTACTGGGGTATAATCGACCAGAAGCTGCGTGAGCATGTTTCGTTTCCACCCAC
GTCAGAAAGGCTTCGTGAGCGAGGCAGGTTGTTTAAATAATGTGCAAAATCTGAATGAACTGCTGCGT
CACAGCAAGGGCCAGCACAAAAATCTGGTTGCCGTGTGCCTGGATGTTTCTAAAGCATTTGATACCGT
TCCTCATAGCATCCTCGGCCCCGCCCTGCGCATGAAGGGCCTGCCGGAACAGGTCGTTTCGTCTGGTTG
AAGATAGCTACAAAGATCTGCATACTGTCGTTAAACAGGGGACCGCAGAAGTGACGCTGAGCCTGCA

[1044]

GCGTGGAGTGAAGCAGGGCGACCCCTGAGCCCCTTCCTGTTCAACGCCGTGCTGGAGCCGCTGCTG
CTGCAGCTGGAAAGCCATCCTGGTTATAAAGTGGGCGGTGAACTGGCCTCTGTTAGCTGTATGGCCTT
TGCAGATGATATCTTTCTGATTGCAGCTAATGTTCCGCAGGCCTGTACCCTGCTGAGGGTCACGGAAG
ATTATCTGGAAAGACTGGGCATGCGTATCAGCGCCCCATAATGTACCAGCTTTGAAATCCGTCGGACC
AAAGATAGCTGGTATGTTGCAGATCCGGGGCTTACACTGACCAAAGGAGAACGTATCCCTGTGCTGCTG
TGTGGATGCCGTTTTTAGCTACCTGGGTGTTGAAATTAGCCCTTGGGCAGGTATCACCAGCGAGGGCA
TCGAACGGGATTGGCGGGGTACACTGCATCGTGTGCAACGCCTGCCGCTGAAGCCCCACCAGAAACT
GGAAGTATCAGCAGATACCTGGTTCCTCATTTTCTGTATAAACTGGTGGTGACCATCCCTAGCATAAC
CCTGATTAGACAGCTGGATCAGGAAGTGCGGGTGTGGTGAAGCAGATCTGTCATCTGCCTCAGAGC
ACCGCCGACGGCATGATCTATTGTCGGAGAGTGGACGGCGGTCTGGGTATTCCGAAGCTGGAATTTGT
TACCGTGACCAGCATACTGAAAGCAGGCCTGAAATTTAGAGATAGCCAGGACAAAATCATGCAGGCA
CTCTGGCTGGCATCAGGTATGAGCAGCCGTCTGAACAGCCTGGCCAAGGCGACCAGAGTACAACCTT
GGCCCCGAACAATATTAAGATCTGGACAGACATAAAGTTGCTCGTAAGAAAGAAGAACTGGCCCG
ATGGGCCAGTTTGACCAGCCAGGGTAAAAGCGTGAAAAGCTTCGCCGGCAGCCGTACCGCCAATGCA
TGGCTGATTAACAAGAAGTTACTGAAGCCCTCTACCTTTATCAGCGCCTTAAGACTGAGAGGCAATGT
CGCTGGAGACCGTGTGGCCCTGAATAGAGCAATCCCGCAGGCCAACCTGATGTGCAGACGTTGCGGT
AGCCAGAGGGAACTCTGGGCCACATCCTGGGTATCTGTACCAGCACCAAAGCCCTACGTATTTACG
CCATGATGAGATCAAGAATCTGATCGTGGACGAAGCAGCAAAGAAGGACGACGAAGTGGCTGTTACA
CTGGAGCCAACCATTCGTCACCCTGTTCTGTTGTAACCTGAAACCGGACCTGGTGGTTCAAAACAGAG
AAGGCGTGACGTTGTTGACGTGACAGTGAGACACGAGGATGGCAACCTGCTTGACAGGGACGTC
AGGATAAACTGGACAAGTACGAAGTGCTGCTGCCGATTCTGCAAGAAAGACTGGGTGCTCCTACCGG
TGAGGTTCTGCCGATTGTTGTTGGCACCCGTGGCGCCATGCCTAAAGAGACAGTGGAAGCCTTGAAG
AACTGCGCATTACCGACCGGCAGACCCTGCTCAGATCAGCCTGATTGCCCTGAGAATGTCTGTGAA
AATTTATCATACCTTCATGGACTATGCAACGCCAGACCGCGTCCGGGCGGGGTGCAAACTACCCCC
ACAGATGATAA

2. R2A TriCasB* N-DYK 氨基酸序列: [* = 自天然 aa 修饰]

MDYKDDDDKGTLPFQSRSCGICLNAGKGNFRALSLDDEERHLRERHPLSLILYKCSACKGQYRSKRAAL
CHAPKCTGTPDPQGNALRCHLCGLVCKSQSGVTQHLRHRHPLVRNTQRAAEESGRAERAALPRPLRRN
TRSVFSEEDAKMLELEVRFNQNERCVAKCMLPFFPNRTCKQIRDKRNTDAYKRRRELYFEGVRVQDPAGA
EDSVLPVVETDEPAEENIPLEYPELPGDEEGAPACSQTILNTEGPDGLGSPVPVEEEMASSGSTSNVDTG
WRESIITAALGVEIPKAISQEPAAVIQELQDALREAVIGVFPQDRLEMYERVLKVNPDDTQERPKRQRK

KGKSRNAFRRYVYSQTQDLFKKNPGQLARYVREDVRWLEQGRVQLQRDDIERMYNKLWGTPDVLPP
HWDYPLPLDTADVLTPIELKEVRKRISQTKLKSAAGPDGLQKRHLVRRVVQEILRLLYNLLMCCAMQPTQ
WRMNRTQLLLKQGKDPLDVASYRPITISSILCRLYWGIDQKLREHVRFHPRQKGFVSEAGCFNNVQILNE
LLRHSKGQHKNLVAVCLDVSKAFDTPHSILGPALRMKGLPEQVVRLVEDSYKDLHTVVKQGTAEVTL
LQRGVKQGDPLSPFLFNAVLEPLLLQLESHPGYKVGELASVSCMAFADDIFLIAANVPQACTLLRVTE
YLERLGMRIAPKCTSFIRPTKDSWYVADPGLTLTKGERIPVAAVDAVFSYLGVEISPWAGITSEGIERDW
RGTLHRVQRLPLKPHQKLELISRYLVPFLYKLVVTIPSTILRQLDQELRVVVKQICHLPQSTADGMIYCR
VDGGLGIPKLEIVTVTSILKAGLKFRDSQDKIMQALWLASGMSSRLNSLAKATRVQPWPPNNIKDLDRHK
VARKKEELARWASLTSQGKSVKSFAGSRTANAWLINKKLLKPSTFISALRLRGNVAGDRVALNRAIPQANL
MCRRCGSQRETLGHILGICTSTKALRISRHDEIKNLIVDEAAKKDDEVAVTLEPTIRHPVRGNLKPDLVVQ
NREGVYVVDVTVRHEDGNLLAQGRQDKLDKYEVLPLILQERLGAPTGEVLPVVVGTGAMPKETVEAL
KKLRITDRQTLLTISLIALRMSVKIYHTFMDYANARPRPGGGANYPHR**

3. R2A TriCasB* N-DYK 转染的 mRNA 序列:[* = 自天然 aa 修饰]

[1045]

AGGAAUAAACUAGUAUUCUUCUGGUCCCCACAGACUCAGAGAGAACCCGCCACCAUGGAUUACA
AGGAUGACGACGAUAAGGGAUCCUGCCUUUCAGAGCAGAAGCUGCGGCAUCUGUCUGAAUGC
CGGUAAAGGUAACUCCGCGCGCUGAGCCUGGACGACGAAGAACGGCACCUGCGUGAACGGCAUC
CACUGUCCUGAUCCUCUAUAAAUGCAGCGAUUGCAAGGGCCAGUACAGAUCCAAGAGGGCCGCC
CUGUGCCACGCCCCAAGUGCACCCGACCGACCCUGAUCCUCAAGGCAAUGCCUGCGUUGUCA
UCUGUGUGGUCUUGUUUGUAAAAGCCAGAGUGGUGUUACCCAGCAUUUACGUCAUAGACACCCU
CUGGUCAGAAACACCCAGCGGGCAGCUGAAGAAAGCGGUAGAGCCGAACGUGCUGCACUGCCUCG
GCCUCUGCGUCGUAACACCCGUUCGGUUUUCAGCGAAGAGGAUGAAGCAAAAUGCUGGAGCUG
GAAGUGCGGUUCCAGAACGAGCGUUGUGUUGCAAAAUGCAUGCUGCCGUUUUUCCGAAUAGAA
CUUGCAAGCAGAUCCGUGAUAAAGCGUAAUACCGAUGCAUAAACGGAGAAGAGAACUGUACUU
CGAGGGCGUCCGGGUGCAGGACCCUGCAGGCGCCGAGGACAGCGUUCUCCUGUUGGAAACCG
ACGAACCCGCCGAGGAAAAUUAUCCGUGGAGUACCCCGAGCUGCCUGGCGAUGAAGAGGGUGC
UCCUGCCUGCAGCCAGACUAUUCUGAACACAGAAGGUCCGGAUGGACUGGGCAGCCCACCGGUGC
CCGUUGAAGAAGAAAUGGCAAGUUCGGGUAGCACAUCAAUAACGUGGAUACCGGUUGGAGAGA
AAGCAUUAUCACAGCUGCACUCGGCGUUGAAAUCCGAAAGCAAUCAGCCAAGAGCCCGCCGCCG
UUAUUCAGGAGCUGCAGGAUGCUCUGCGCGAGGCCGUGAUCGGCGUGUUUCCGCGAGGAUCGCCU
GGACGAGAUUGACGAGCGGGUACUAAAGUGGUCAACCCGGAUGAUACACAGGAACGCCCGAAA
CGUCAAGAAAGAAAGGCAAGUCUGUAAUGCCUCCGCCGUUAUGUGUACAGCCAGACCCAGG

[1046]

ACCUGUUCAAGAAAAAUCCUGGACAGCUGGCACGUUAUGUUAGGGAAGAUGUGAGAUGGCUGGA
ACAGGGCCGGGUGCAGUUGCAGAGAGAUGAUUUGAAAGAAUGUACAACAAGCUGUGGGGCACC
AAGCCGGAUGUGCUGCCUCCCCACUGGGAUUAUCCACUGCCACUCGACACCGCUGAUGUUCUGAC
CCCGAUUGAGCUGAAAGAAGUCCGGAAGAAUUAUCUCAGACGAAACUAAAUCUGCAGCAGGG
CCCGACGGUCUGCAGAAAAGACAUCUGGUGCGUCGCGUUGUGCAAGAAAUUCUGCGCCUGCUGU
AUAACCUGCUGAUGUGUUGUGCAAUGCAGCCUACACAGUGGCGGAUGAACCGUACCCAACUGUU
ACUGAAGCAGGGUAAAAGAUCCUCUGGAUGUCGCUAGUUAUCGUCCUAUAACCAUCUCCAGCAUC
CUUUGUCGUCUGUACUGGGGUAAUAAUCGACCAGAAGCUGCGUGAGCAUGUUCGUUCCACCCAC
GUCAGAAAGGCUUCGUGAGCGAGGCAGGUUGUUUAAUAAUGUGCAAAUUCUGAAUGAACUGCU
GCGUCACAGCAAGGGCCAGCACAAAAUCUGGUUGCCGUGUGCCUGGAUGUUUCUAAAGCAUUU
GAUACCGUUCCUCAUAGCAUCCUCGGCCCCGCCUGCGCAUGAAGGGCCUGCCGGAACAGGUCGU
UCGUCUGGUUGAAGAUAGCUACAAAGAUUCUGCAUACUGUCGUUAAACAGGGGACCGCAGAAGUG
ACGCUGAGCCUGCAGCGUGGAGUGAAGCAGGGCGACCCCCUGAGCCCCUCCUGUUAACGCCGU
GCUGGAGCCGCUGCUGCUGCAGCUGGAAAGCCAUCCUGGUUAUAAAGUGGGCGGUGAACUGGCC
UCUGUUAGCUGUAUGGCCUUUGCAGAUGAUUCUUUCUGAUUUCAGCUAAUGUUCGCAGGCCU
GUACCCUGCUGAGGGUCACGGAAGAUUAUCUGGAAAGACUGGGCAUGCGUAUCAGCGCCCCUAA
AUGUACCAGCUUUGAAAUCCGUCCGACCAAAGAUAGCUGGUUUGUUGCAGAUCCGGGGCUUACA
CUGACCAAAGGAGAACGUAUCCCUUGUCGUGCUGUGGAUGCCGUUUUAGCUACCUGGGUGUUG
AAAUUAGCCCUUGGGCAGGUUUCACCAGCGAGGGCAUCGAACGGGAUUGGCGGGGUACACUGCA
UCGUGUGCAACGCCUGCCGCUGAAGCCCCACCAGAAACUGGAACUGAUCAGCAGAUACCUGGUUC
CUCAUUUUCUGUAUAAACUGGUGGUGACCAUCCCUAGCAUAACCCUGAUUAGACAGCUGGAUCA
GGAACUGCGGGUUGUGGUGAAGCAGAUUCUGUCAUCUGCCUCAGAGCACCGCCGACGGCAUGAUC
UAUUGUCGGAGAGUGGACGGCGGUCUGGGUAUUCGGAAGCUGGAAAUUGUUAACCGUGACCAGCA
UACUGAAAGCAGGCCUGAAAUUUAGAGAUAGCCAGGACAAAUCAUGCAGGCACUCUGGCUGGC
AUCAGGUUAGAGCAGCCGUCUGAACAGCCUGGCCAAGGCGACCAGAGUACAACCUUGGCCCCCGA
ACAAUAAUAAAGAUUCUGGACAGACAUAAGUUGCUCGUAAAGAAAGAAGAACUGGCCCCAUGGGC
CAGUUUGACCAGCCAGGGUAAAAGCGUGAAAAGCUUCGCCGGCAGCCGUACCGCCAAUGCAUGGC
UGAUUAACAAGAAGUUACUGAAGCCCUUACCUUUAUCAGCGCCUUAAGACUGAGAGGCAAUGU
CGCUGGAGACCGUGUGGCCCUGAUAGAGCAAUCCCGCAGGCCAACCUGAUGUGCAGACGUUGCG
GUAGCCAGAGGGAAACUCUGGGCCACAUCUGGGUAUCUGUACCAGCACCAAAGCCCUACGUUU
UCACGCCAUGAUGAGAUCAAGAAUCUGAUCGUGGACGAAGCAGCAAAGAAGGACGACGAAGUGG
CUGUUACACUGGAGCCAACCAUUCGUCACCCUGUUCGUGGUAAACCUGAAACCGGACCUGGUGGU

CAAAACAGAGAAGGCGUGUACGUUGUACGUGACAGUGAGACACGAGGAUGGCAACCUGCUUG
CACAGGGACGUCAGGAUAAACUGGACAAGUACGAAGUGCUGCUGCCGAUUCUGCAAGAAAGACU
GGGUGCUCCUACCGGUGAGGUUCUGCCGAUUGUUGUUGGCACCCGUGGCGCCAUGCCUAAAGAG
ACAGUGGAAGCCUUGAAGAAACUGCGCAUUACCGACCGGCAGACCCUGCUCACGAUCAGCCUGAU
UGCCCUGAGAAUGUCUGUGAAAAUUUAUCAUACCUUCAUGGACUAUGCAAACGCCAGACCGCGU
CCGGGCGGCGGUGCAAACUACCCCCACAGAUGAUGACCUCGAGCUGGUACUGCAUGCACGCAAUG
CUAGCUGCCCCUUUCCCGUCCUGGGUACCCCGAGUCUCCCCGACCUCGGGUCCAGGUAUGCUC
CCACCUCACCUGCCCCACUCACCACCUCUGCUAGUUCAGACACCUCCCAAGCACGCAGCAAUGC
AGCUCAAAACGCUUAGCCUAGCCACACCCCCACGGGAAACAGCAGUGAUUAACCUUUAGCAAUAA
ACGAAAGUUUAACUAAGCUAUACUAACCCAGGGUUGGUCAAUUUCGUGCCAGCCACACCCUGG
AGCUAGCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGCAUUGACUAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

4. R2A TriCasB*氨基酸序列:[* = 自天然 aa 修饰]

[1047] MLPFQSRSCGICLNAGKGNFRALSLDDEERHLRERHPLSLILYKSDCKGQYRSKRAALCHAPKCTGPTP
DPQGNALRCHLCGLVCKSQSGVTQHRLRHRHPLVRNTQRAAEESGRAERAALPRPLRRNTRSVFSEDEA
KMLELEVRVFQNERCVAKCMLPFFPNRTCKQIRDKRNTDAYKRRRELYFEGVRVQDPAGAEDSVLPVET
DEPAEENIPLEYPELPGDEEGAPACSQTILNTEGPDGLGSPVPVEEEMASSGSTSNVDTGWRESIITAAL
GVEIPKAISQEPAAVIQELQDALREAVIGVFPQDRLDEMYERVVKVNPDDTQERPKRQRKKGKSRNAFR
RYVYSQTQDLFKKNPGQLARYVREDVRWLEQGRVQLQRDDIERMYNKLWGTKPDVLPHPWDYPLPLDT
ADVLTPIELKEVRKRISQTKLKSAGPDGLQKRHLVRRVVQEILRLLYNLLMCCAMQPTQWRMNRQTLL
LKQGKDPLDVASYRPITISSILCRLYWGIIQKLREHVRFHPRQKGFVSEAGCFNNVQILNELLRHSKGQH
KNLVAVCLDVSKAFDTPHSILGPALRMKGLPEQVVRLVEDSYKDLHTVVKQGTAEVTLSQLRGVKQGD
PLSPFLFNAVLEPLLLQLESHPGYKVGELASVSCMAFADDIFLIAANVPQACTLLRVTEDYLERLGMRIS
APKCTSFEIRPTKDSWYVADPGLTLTKGERIPVAAVDAVFSYLGVESPWAGITSEGIERDWRGTLHRVQRL
PLKPHQKLELISRYLVPHFLYKLVVTIPSITLIRQLDQELRVVVKQICHLPQSTADGMIYCRRVDGGLGIPKL
EIVTVTSILKAGLKFRDSQDKIMQALWLASGMSSRLNSLAKATRVQPWPNNIKDLDRHKVARKKEELAR
WASLTSQGKSVKSFAGSRTANAWLINKLLKPSTFISALRLRGNVAGDRVALNRAIPQANLMCRRCGSQR
ETLGHILGICTSTKALRISRHDEIKNLIVDEAAKKDDEVAVTLEPTIRHPVRGNLKPDLVVQNREGVYVVD
VTVRHEDGNLLAQGRQDKLDKYEVLPLILQERLGAPTGEVLPVVGTRGAMPKETVEALKKLRTDRQT
LLTISLIALRMSVKIYHTFMDYANARPRPGGGANYPHR**

5. R2A TriCasB* 转染的 mRNA 序列： [* = 自天然 aa 修饰]

AGGAAUAAACUAGUAUUCUUCUGGUCCCCACAGACUCAGAGAGAACCCGCCACCAUGCUGCCUUU
UCAGAGCAGAAGCUGCGGCAUCUGUCUGAAUGCCGGUAAAGGUAACUCCGCGCGCUGAGCCUG
GACGACGAAGAACGGCACCUGCGUGAACGGCAUCCACUGUCCCUGAUCCUCUAUAAAUGCAGCGA
UUGCAAGGGCCAGUACAGAUCCAAGAGGGCCGCCUGUGCCACGCCCCAAGUGCACCGGACCGA
CCCCUGAUCCUCAAGGCAAUGCCCUGCGUUGUCAUCUGUGUGGUCUUGUUUGUAAAAGCCAGAG
UGGUGUUACCCAGCAUUUACGUCAUAGACACCCUCUGGUCAGAAACACCCAGCGGGCAGCUGAAG
AAAGCGGUAGAGCCGAACGUGCUGCACUGCCUCGCCCUCUGCGUCGUAACACCCGUUCGGUUUUC
AGCGAAGAGGAUGAAGCAAAAUGCUGGAGCUGGAAGUGCGGUUCCAGAACGAGCGUUGUGUUG
CAAAAUGCAUGCUGCCGUUUUUUCCGAUAGAACUUGCAAGCAGAUCCGUGAUAAAGCGUAAUAC
CGAUGCAUAAUAAACGGAGAAGAGAACUGUACUUCGAGGGCGUCCGGGUGCAGGACCCUGCAGGC
GCCGAGGACAGCGUUCUCCUGUUGUGGAAACCGACGAACCCGCCGAGGAAAAUAUUCGCUGGA
GUACCCCGAGCUGCCUGGCGAUGAAGAGGGUGCUCCUGCCUGCAGCCAGACUUAUCUGAACACAG
AAGGUCCGGAUGGACUGGGCAGCCCACCGGUGCCCGUUGAAGAAGAAAUGGCAAGUUCGGGUAG
CACAUCUAAUAAACGUGGAUACCGGUUGGAGAGAAAGCAUUAUCACAGCUGCACUCGGCGUUGAA
AUUCCGAAAGCAAUCAGCCAAGAGCCCGCCCGGUUAUUCAGGAGCUGCAGGAUGCUCUGCGCGA
[1048] GGCCGUGAUCGGCGUGUUUCCGCAGGAUCGCCUGGACGAGAUGUACGAGCGGGUACUCAAGUG
GUCAACCCGGAUGAUACACAGGAACGCCGAAACGUCAAAGAAAGAAAGGCAAGUCUCGUAAUG
CCUUCGCCCGUUAUGUGUACAGCCAGACCCAGGACCUGUUAAGAAAAUCCUGGACAGCUGGCA
CGUUAUGUUAGGGAAGAUGUGAGAUGGCUGGAACAGGGCCGGGUGCAGUUGCAGAGAGAUGAUA
UUGAAAGAAUGUACAACAAGCUGUGGGGCACCAAGCCGGAUGUGCUGCCUCCCCACUGGGAUUA
UCCACUGCCACUCGACACCGCUGAUGUUCUGACCCCGAUUGAGCUGAAAGAAGUCCGGAAAAAGAA
UAUCUCAGACGAAACUAAAUCUGCAGCAGGGCCCCACGGUCUGCAGAAAAGACAUCUGGUGCG
UCGCGUUGUGCAAGAAAUUCUGCGCCUGCUGUAUAACCUGCUGAUGUGUUGUGCAAUGCAGCCU
ACACAGUGGCGGAUGAACCGUACCCAACUGUUACUGAAGCAGGGUAAAGAUCUCUGGAUGUCG
CUAGUUAUCGUCCUAUAACCAUCUCCAGCAUCCUUUGUCGUCUGUACUGGGGUAAUAAUCGACCA
GAAGCUGCGUGAGCAUGUUCGUUUCACCCACGUCAGAAAGGCUUCGUGAGCGAGGCAGGUUGU
UUUAAUAAUGUGCAAAUUCUGAAUGAACUGCUGCGUCACAGCAAGGGCCAGCACAAAAAUCUGG
UUGCCGUGUGCCUGGAUGUUUCUAAAGCAUUUGAUACCGUUCUUAUAGCAUCCUCGGCCCCGCC
CUGCGCAUGAAGGGCCUGCCGGAACAGGUCGUUCGUCUGGUUGAAGAUAGCUACAAAGAUCUGC
AUACUGUCGUUAAACAGGGGACCGCAGAAGUGACGCUGAGCCUGCAGCGUGGAGUGAAGCAGGG
CGACCCCCUGAGCCCCUCCUGUUAACGCCGUGCUGGAGCCGUGCUGCUGCAGCUGGAAAGCC

[1049]

AUCCUGGUUAUAAAAGUGGGCGGUGAACUGGCCUCUGUUAGCUGUAUGGCCUUUGCAGAUUAU
CUUUCUGAUUGCAGCUAAUGUCCGCAGGCCUGUACCCUGCUGAGGGUCACGGAAGAUUAUCUG
GAAAGACUGGGCAUGCGUAUCAGCGCCCCUAAAUGUACCAGCUUUGAAAUCCGUCCGACCAAAG
AUAGCUGGUUAUGUUGCAGAUCCGGGGCUUACACUGACCAAAGGAGAACGUAUCCCCUGUCGCUGC
UGUGGAUGCCGUUUUAGCUACCUGGGUGUUGAAAUAGCCCUUGGGCAGGUUAUCACCAGCGAG
GGCAUCGAACGGGAUUGGCGGGGUACACUGCAUCGUGUGCAACGCCUGCCGCUGAAGCCCCACCA
GAAACUGGAACUGAUCAGCAGAUACCUGGUUCCUCAUUUCUGUAUAAACUGGUGGUGACCAUC
CCUAGCAUAACCCUGAUUAGACAGCUGGAUCAGGAACUGCGGGUUGUGGUGAAGCAGAUUCUGUC
AUCUGCCUCAGAGCACCGCCGACGGCAUGAUCUAUUGUCGGAGAGUGGACGGCGGUCUGGGUAU
UCCGAAGCUGGAAAUUGUUACCGUGACCAGCAUACUGAAAGCAGGCCUGAAAUUUAGAGAUAGC
CAGGACAAAAUCAUGCAGGCACUCUGGCUGGCAUCAGGU AUGAGCAGCCGUCUGAACAGCCUGG
CCAAGGCGACCAGAGUACAACCUUGGCCCCCGAACAUAUUAAGAUCUGGACAGACAUAAGU
UGCUCGUAAGAAAGAAGAACUGGCCCGAUGGGCCAGUUUGACCAGCCAGGGUAAAAGCGUGAAA
AGCUUCGCCGGCAGCCGUACCGCCAAUGCAUGGCUGAUUAACAAGAAGUUACUGAAGCCCUUAC
CUUUAUCAGCGCCUUAAGACUGAGAGGCAAUGUCGCUGGAGACCGUGUGGCCCUGAAUAGAGCA
AUCCCGCAGGCCAACCCUGAUGUGCAGACGUUGCGGUAGCCAGAGGGAAACUCUGGGCCACAUCU
GGUAUCUGUACCAGCACCAAAGCCCUACGUAUUUCACGCCAUGAUGAGAUAAGAAUCUGAUC
GUGGACGAAGCAGCAAAGAAGGACGACGAAGUGGCUGUUACACUGGAGCCAACCAUUCGUCACC
CUGUUCGUGGUAACCUGAAACCGGACCUGGUGGUUCAAACAGAGAAGGCGUGUACGUUGUUGA
CGUGACAGUGAGACACGAGGAUGGCAACCUGCUUGCACAGGGACGUCAGGAUAAACUGGACAAG
UACGAAGUGCUGCUGCCGAUUCUGCAAGAAAGACUGGGUGCUCCUACCGGUGAGGUUCUGCCGA
UUGUUGUUGGCACCCGUGGCGCCAUGCCUAAAAGAGACAGUGGAAGCCUUGAAGAAACUGCGCAU
UACCGACCGGCAGACCCUGCUCACGAUCAGCCUGAUUGCCUGAGAAUGUCUGUGAAAAUUUAUC
AUACCUUCAUGGACUAUGCAAACGCCAGACCGCGUCCGGGCGGCGGUGCAAACUACCCCCACAGA
UGAUGACUCGAGCUGGUACUGCAUGCACGCAAUGCUAGCUGCCCCUUUCCCGUCCUGGGUACCCC
GAGUCUCCCCCGACCUCGGGUCCCAGGU AUGCUCCACCUCACCUGCCCCACUACCACCUCUGC
UAGUCCAGACACCUCCCAAGCACGCAGCAAUGCAGCUAAAACGCUUAGCCUAGCCACACCCCC
ACGGGAAACAGCAGUGAUUAACCUUUAGCAAUAAACGAAAGUUUAACUAAGCUAUACUAACCCC
AGGGUUGGUCAAUUUCGUGCCAGCCACACCCUGGAGCUAGCAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAGCAUAGACUAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAA

6. R2A NaviB N-DYK 质粒 ORF 序列:

ATGGATTACAAGGATGACGACGATAAGGGTACCATGGAGATCGTGGCCATCGCCAAGAATCCACAGG
CATGCCCCAAAGTGCTGTCAGGGAGGAACCCAGCTGCTGTGCATGGGCAGCTGGGAGCTGTCCAGAC
ACATCAACAAGGAGCACCCCAGCGTGGACGTGACATGGGTGTGCGGCGCCTGTGACGGGAGATGTAC
CACACTGCGCAGCTGGTCCTGCCACGTGCCTCACTGTAAGGGCCGGCAGGAGCCTAAGGATCTGCCA
TTCAAGTGCGAGCACTGTTCTCTGAGCTTTGACAGCCAGATCGGCCTGTCCCAGCACGAGAGGCACG
TGCACCCAGAGGTGCGGAATGATAAGAGAGCCGCCGAGGCCAACAAGCCAAAGGGCAAGAGCGGC
AGGAGGCCCTTCCATCTGGTCTGATGAGGACCTGCTGCTGATCAGGGAGCTGGAGTCCGAGTACCACG
GAGCCCGGAATATCAACGAGAAGATCGCAGAGCACTTCCCTGACAGGACCGGAAGGCAGGTGTCCG
ATGCACGGAGAAGGAAGGACTATGCCGCCCTGAGGGGCGCGGAGGACCACAGGGACCTGCAGAGG
GAGTGGAGGCAATCGAGGAGGTGGATGAGGGCGAGATCCCAGAGGGCGAGGAGCTGGTGGCCACA
GACGGAGCAGCACTGGAGTCTGGACCACCTGAGAATGGAGGCAGCGCACCTGCAGAGCAGGTGAAC
GCACCAGCACTGGAGAGCTCCTCTCAGCAGGATAGGGAGTGCTCTCCAGCAGTGGGCAGCGATGAG
CAGATCGAGGACAGCTCCGACGATGACGAGTTTTCTGACGCACTGGGAGAGATCAGCCTGCCAGAGC
CTCTGTCCGTGGAGAGGACCACAATCAGCCCACCACCTAGGGATGACTGGAAGGGACCTATGCGGTG
GGAGATCTGTAATGCCTCCGAGGAGGCCGGCTCTTACGCCAACTGGGTGACCGGACTGCAGGAGCTG
GTGCGCAACAATGCCCTGTCTGAGATCGGCCTGGATAGCCTGTATGACCAGCTGATCCAGATCATGCG
GCACCCAGCGATGACAATGAGCAGGACAGGCTGCAGCTGAACGCAAGAGGACCACCAAGGAGGG
GACACCGCAAGAATAGAAGGAGGAGGAGACTGACAGCAGCAGATAGGAAGAGATTGCGCTTTGCCC
GGTGCCAGGACCTGTGGAACAATAACCTAAGAAGCTGGCCGAGCTGGTCATCGCCAACGATCTGTC
CATCTGCAGAGGAGGCAGGCACCAGGAAGAACCGAGACACAGACCCTGTACAATGAGCTGTGGGG
AAGGGTGGGACCAAAACATCGAGGCACCAAGGAGAACCGAGGACCCAATCCCCGTGTCCCGCATCTT
CACACCAATCACCCCCAGGAGATCATGGGCCGGATCAGGCGCATCAAGAACGATTCTGCAGCAGGA
CCTGACGGAGTGACAAAGGATGACCTGAGGGGCGCGGCGTGTCTATCGCACTGAGCAAGCTGTTTA
ATAGCATCCTGCTGGCCGGCTACTATCCAAAGGCCTGGCGGGAGAACAGAACCACACTGCTGCCTAA
GCCAGAGAAGGACCCCGCCGACGTGAAGAATTGGAGACCCATCACCATCTCTAGCATGGTGTCCAGG
GTGTATTCTGGCCTGCTGGATCAGCGGGTGAGAGCCGTGATCAAGCAGTGCGACAGGCAGAAGGGCT
TCACAGAGGAGAATGGCTGTTTTTCCAACATCCAGCTGCTGGATGACGCCGTGTCTAACGCCAAGAA
GGCCGGCGGCGTGATCACCATCCTGGATGTGAGCAAGGCCCTTCGACACAGTGCCACACGCCGTGATC
CAGGGCTGCCTGGAGAAGAAGGGCATCCCCGAGACCGTGGCCGCCTACATCTCCTCTATGTATCGCGA
TTGTAGCACAGCAATCAGGACCAGGTCCGGCGACGTGAAGATCGGCATGAAGAGGGGAGTGAAGCA
GGGCGACCCTCTGTCCCCACTGATCTTTAATCTGGTGCTGGAGCCACTGCTGGAGAGGCTGCAGGAG

[1050]

[1051]

ACCAGCGGCGTGGAGATCGAGGGCATGAATCTGTCTGCGCCGCCTTCGCCGATGACATCGTGTGTTT
TGCAAACACAGCACCTGAGGCAGGAAGGCAGCTGCGCATGGTGGCCGATTACCTGGGCAGACTGGA
CATGTCCCTGTCTGTGAGCAAGTGCATCGCCGTGGAGTACGTGCCTCACAGGAAGACATGGTATACCA
AGAACCCAGGCCTGGAAGTGAATGGCAACGCCGTGCCCTCCATCTCTCCTAGCGAGACCTTCAAGTA
CCTGGGCGCCAAGGTGAGCCCATGGAAGGGCCTGCTGGAGGGCTTCGAGTCCGATGCCTTTCGCGAA
GTGATCTCTCGGGTGCAGAGACTGCCCCTGAAGCCTATGCAGAAGGTGGACCTGCTGCAGATGTATAT
CTTCCCCAGATACACCTACGGCCTGATCACCAGCCCTCCAGCCAAGGCCGTGCTGAAGACAATCGATA
GGATCATCCGGACCAGAATCAAGGAGATCCTGCACCTGCCAGAGTCCGTGAGCTCCTCTTTTCTGTAC
ACCCCAAGGAAGCAGGGAGGACTGGGACTGCTGGAGGTGGAGAAGATGGTGTGATCGCAGCACTG
AGAAATGGACTGAGGGCAAGGCAGTCTCACGACCCAGTGACCAGGGCCGCCATGAATTCTAACGCC
GCCGATGACAGACTGAAGAGCTACGCCGATGCACTGAGGCTGCACTGGCCACTGACCACAAAGGAG
CTGGACACATACAAGTATCAGCTGAGACTGTCCTATGCCCAGAAGTGGGCAGAGCAGAAGTGGCAGG
GACAGGGAGTGGAGGAGTTCGCACAGGACCCCGTGGGCAATTCTTGGCTGCAGAGATACGATCTGCT
GCCTGCCAGCAGGTATATCGACGCCATCAAGCTGCGGACAAACACCTACCCACCAGAGCCCTGATG
AAGATCATCGATGGCAGGGTGGACAGCTCCTGCCGCAAGTGTGAGGGCTCTAGCGAGACCCTGGGAC
ACATCCTGGGAAGGTGTAGATATACAAAGGATAAGAGGATCAGCCGCCACAATGAGATCAAGGACCT
GCTGAAGGCCCCGCTGGCCAAGAACCACCAGGTCATGGATGAGCCTCAGATCACCGTGCGGGGCCA
GAGATTCAAGCCAGACCTGGTGGTGAAGACAAATGAGGGCCGGGTGCACGTGATCGATGTGACAGT
GAGGTACGAGCACCGCACCTATCTGGACGAGGGCCGCACAGAGAAGATCGGCAAGTACCGGCAGAT
CCTGAGCACCTGCGGAGAGATCTGCACTCCAACGCCGAGGAAGTGATCCCCATCGTGATCGGCTCC
AGAGGAGCAATCCCTAGGGAGACCAGGAAGGCACTGTCTAAGCTGGGCATCGGCAAGTCCGACTGG
CTGACAATCTCTCTGATCGCCCTGAGGTCTCTCTGGAGATCGTGAACGCCTTTATGGATGACTAG

7. R2A NaviB N-DYK 氨基酸序列:

MDYKDDDDKGTMEIVAIKPNQACPKCLQGGTQLLCMGSWELSRHINKEHPSVDVTWVCGACQRRCT
TLRSWSCHVPCHKGRQEPKDLFPKCEHCSLSFDSQIGLSQHERHVVHPEVRNDKRAAEANKPKGKSGRRP
SIWSEDELLIRELESEYHGARNINEKIAEHFPDRTRGRQVSDARRRKDYAALRGRGGPQGPAEGVEAIEEV
DEGEIPEGEELVATDGAALESGPPENGGSAPAEQVNAPALESSSQDRECSPAVGSDEQIEDSSDDDEFSDA
LGEISLPEPLSVERTTISPPRDDWKGPMRWEICNASEEAGSYANWVTGLQELVRNNALSEIGLDSLYDQLI
QIMRHPSDDNEQDRLQLNARGPPRRGHRKNRRRRRLTAADRKRFAFARCQDLWNNNPKKLAELVIANDL
SILQRRQAPGRTEQTLYNELWGRVGPNIAPRRTEPIPVSRIFTPTPQEIMGRIRRIKND SAAGPDGVTK
DDL RGRGVSIASLKFNSILLAGYYPKAWRENRTLLPKPEKDPADVKNWRPITISSMVSRYVSGLLDQVR

RAVIKQCDRQKGFTEENGCFNSIQLLDDAVSNAKKAGGVITILDVSKAFDTPHAVIQGCLEKKGIPETVA
AYISSMYRDCSTAIRTRSGDVKIGMKRGVKQGDPLSPLIFNLVLEPLLERLQETSGVEIEGMNLSCAAFAFAD
DIVCFANTAPEAGRQLRMVADYLGRLDMSLSVSKCIAVEYVPHRKTWYTKNPGLEVNGNAVPSISPSETF
KYLGAKVSPWKGLLEGFESDAFREVISRVQRLPLKPMQKVDLLQMYIFPRYTYGLITSPPAKAVLKTIDRII
RTRIKEILHLPESVSSFLYTPRKQGGGLGLEVEKMLVIAALRNGLRARQSHDPVTRAAMNSNAADDRLK
SYADALRLHWPLTTKELDTYKYLRLSYAQKWAEQKWQGGQVEEFAQDPVGNLQRYDLLPASRYID
AIKLRTNTYPTRALMKIIDGRVDSSCRKCGSSETLGHILGRCRYTKDKRISRHEIKDLLKARLAKNHQV
MDEPQITVRGQRFKPDLVVKTNEGRVHVIDVTVRYEHRTYLDEGRTEKIGKYRQILSTLRRDLHSNAEEVI
PIVIGSRGAIPRETRKALSKLGIGKSDWLTISLIALRSSLEIVNAFMDD*

8. R2A OrLaSyn C-3F 质粒 ORF 序列:

[1052]

ATGGGCACCGATACCGTTTATGTGGGTCAAGATTATCCAAGCGGCCTGTCCAAGCGCGTTCGGGCTCG
TCTGGTTGCTGGTCCAATGCTGCGCGAGCGCAGCTGCCATGCCACGTGTTCCGTGCTGGTCACATGT
GGAATTGGCGCACCGAGCTGCCGAGCGGTCTGTTGGGACCAGCCGGCCCTGGAGAAGAGCCGCGTGC
TGACCCGTAGCGTGGCTACCGCCACCGATCCGGAGATCACCAGCTACCCGGGCAAGAGCGTGAGCAC
CAGCACCCAGGTGCAGGAAGAGGACTGGTGCAGCCGTGAAAGCGGCTGGATCAGCCGGGTCTGGC
TCCAGAGGAACCGAGCGTGGTGAGCGAGATTACCGCTAGCATGGTGGCTACCATGCGTGTGGCTACC
GAGGAAGTGGTGCTGGAACCGCAGCCGGAGCAGGTGGTGACCATTCTGCCAGAACATGGTCGTAAC
GTGCCGCCGGGTCTGGCTGAACAGGACACCGCGAGCCCGATTGAGGTGAGCGTGCTGCTGCCGGATC
TGGCGGAGAACTGCCCAGTGTGCGGTGTGCCGAGCGGCGGTCTGCGTCTGCTGGGCAAGCACTTCGC
GGTGCGTCATGCTGGTGTGCCAGTGACCTACGAGTGCCGCAAATGCGCGTGCGGTAGCCCGAACAGC
CACAGCATCAGCTGCCATGTGCCGAAGTGCCGCGGTCTGTGCTCGTATGCCGAGCGGCGACCCGGGTA
TTGCTTGCGATCTGTGCGAAGCGCGCTTTGCTACCGAGGTGGGTGTGGCCAGCACAAAGCGTCACGT
GCACCCGGTGGAATGGAACAAAGTGCGCCTGGAGCGTCGCGGTGCCCGTGCGGTGGCATCAAGGC
GACCAAAGTGTGGAGCGTGGCTGAAGTGGAACCCCTGATCCGTCTGATTCGTGAGCACGGCGACAGC
GGTGCGACCTATCAGCTGATCGCTGATGAACTGGGTCTGGCAAGACCGCCGAGCAGGTGCGTAGCA
AGAAACGCCTGCTGCGTATTGACACCGCCAGCAACAGCCCGGACGATGCGGAAGTGAGGAAGAGC
GCCTGGAGAGCCTGGCCGTGCGTAGCAGCAGCCGTAGCCCGCCGAGCCTGGTGGCCACCCGCGTGC
GTGAAGCGGTGGCTCGTGGTGAAAGCGAGGGTGGCGAAGAGATCCGCGCCATTGCCGCGCTGATCC
GTGACGTGGATCAGAACCCGTGCCTGATTGAAACCAGCGCCAGCGACATCATTAGCAAGCTGGGCCG
TCGCGTGACGGCCCGAAACGCCCCGCTCCGGTGGTGGTGAACAGACCCAGGAGAAGGGTTGGGT
GCGTCGCTGGCCCGTCGAAACGCGAATACCGTGAGGCTCAGTACCTGTATAGCCGCGACCAAGCT

[1053]

CGTCTGGCTGCGCAGATCCTGGACGGCGCCGCTAGCCAGGAGTGCGCCCTGCCAGTGGACCAGGTGT
ATGGTGCTTTCCGTGAGAAAGTGGGAAACCGTGGGCCAGTTCCACGGTCTGGGCGAATTCGCACCGG
TGCGCGTGCTGATAACTGGGAGTTCTACAGCCCGATCCTGGCCGCGGAAGTGAAGGAGAACCTGATG
CGCATGGCTAATGGTACCGCCCCAGGTCCAGACCGTATTAGCAAGAAAGCCCTGCTGGATTGGGACCC
GCGTGGTGAACAGCTGGCCCGTCTGTATACCACCTGGCTGATCGGTGGCGTGATTCCGCGCGTGTTCA
AGGAATGCCGTACCAAGCTGCTGCCGAAGAGCAGCGACCCGGTGGAGCTGCAGGATATCGGTGGCTG
GCGTCCGGTGACCATTGGTAGCATGGTGACCCGCCTGTTTAGCCGTATCCTGACCATGCGTCTGACCC
GCGCGTGCCCGATTAAACCCGCGCCAGCGTGCTTCTGGCCAGCAGCAGCGGTTGCGCCGAGAACCT
GCTGATCTTTGACGAGATTGTGCGTCGCAGCCGTCGCGATGGTGGCCCACTGGCCGTGGTGTTCGTGG
ACTTTGCCCCGTGCCTTCGACAGCATCAGCCACGAACACATTCTGTGCGTGCTGGAAGAGGGTGGCCT
GGACCGTCACGTGATCGGTCTGATTCTGTAACAGCTACGTGGATTGCGTGACCCGTGTGGGCTGCGTGG
AGGGTATGACCCCGCCGATCCAGATGAAAGTGGGCGTGAAACAGGGTGACCCGATGAGCCCGCTGCT
GTTTAACCTGGCGATGGACCCGTGATCCACAAGCTGGAAACCGCTGGTACCGGCCTGAAATGGGGC
GACCTGAGCATTGCCACCCTGGCTTTCGCCGACGATCTGGTGCTGGTGAGCGATAGCGAAGAGGGTAT
GGGCGGTAGCCTGGGCATTCTGGAGAAGTTCTGCCAGCTGACCGGTCTGCGTGTCAGCCGCGTAAG
TGCCACGGCTTCTTTATGGATAAAGGTGTGGTGAACGGTTGCGGCACCTGGGAAATCTGCGGCAGCC
CGATCCACATGATTCCGCCGGGTGAGAGCGTGCGCTACCTGGGTGTGCAAGTGGGTCCGGGTCTGTGG
TGTGATGGAGCCAGACCTGATTCCGACCGTGACACCTGGATCGAACGCATTAGCGAAGCCCCGCTG
AAACCGAGCCAGCGCATGCGTGTGCTGAACAGCTTCGCTCTGCCGCGTATCATCTACCAGGCCGACCT
GGGCAAGGTGACCGTGACCAAACTGGCCCAGATCGATGGTATTGTGCGTAAGGCGGTGAAGAAATGG
CTGCACCTGAGCCCCAGCACCTGCAACGGCCTGCTGTACAGCCGCAACCGTGATGGTGGCCTGGGTC
TGCTGAAGCTGGAGCGTCTGATCCCGAGCGTGCGCACCAAAACGTATCTACCGCATGAGCCGTAGCCC
GGACATCTGGACCCGTCGCATGACCAGCCACAGCGTGAGCAAGAGCGATTGGGAAATGCTGTGGGTG
CAGGCGGGTGGCGAGCGTGCGACGCGCCCCGGTTATGGGTGCGGTGGAAGCGGCTCCAACCGACGTG
GAGCGTAGCCCGGACTACCCAGATTGGCGTCGCGAAGAGAACCTGGCCTGGAGCGCGCTGCGTGTGC
AGGGCGTGGGTGCCGACCAGTTCCGCGGCGACCGTACCAGCAGCAGCTGGATCGCTGAACCAGCCA
GCGTGGGCTTCGCCCAGCGTCACTGGCTGGCGGCTCTGGCGCTGCGTGCTGGCGTGTATCCAACCCG
CGAATTCCTGGCCCGTGGCAAGGAGAAGAGCGGTGCTGCCTGCCGTCGCTGCCAGCTCGTCTGGAG
AGCTGCAGCCACATTCTGGGCCAGTGCCCGTTTGTGCAGGCTAACCGCATTGCCCGTCACAACAAAG
TGTGCGTGCTGCTGGCTACCGAAGCCGAGCGTTTCGGCTGGACCGTGATCCGCGAATTCGTCTGGAG
GACGCTGCCGGCGGCCTGAAGATCCCGGATCTGGTGTGCAAGAAAGCCGACACCGTGCTGATTGTGG
ACGTGACCGTGCGCTACGAAATGGACGGTGAAACCCTGAAGCGTGCGGCTAGCGAGAAGGTGAAGC

ACTATCTGCCGGTGGGTCAGCAGATCACCGATAAAGTGGGTGGCCGCTGCTTCAAAGTGATGGGCTTT
CCGGTGGGTGCTCGTGGCAAGTGGCCGGCCAGCAACAATACCGTGCTGGCCGAACTGGGCGTGCCG
GCTGGTCGCATGCGTACCTTCGCGCGTCTGGTGAGCCGTCGCACCCTGCTGTATAGCCTGGACATTCT
GCGCGACTTCATGCGTGAACCAGCAGGTCTGGCACCCGTGTGGCACTGATTCCAGCAGCAACCGGC
GCAGCAAATATGGGGAGCGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAA
GGATGACGATGACAAGAAGTGATAA

9. R2A OrLaSyn C-3F 氨基酸序列:

MGTDTVYVGQDYPSGLSKRVPARLVAGPMLRERSCHAHVFRAGHMWNWRTSLPSGRWDQPALEKSRVL
TRSVATATDPEITSYPGKSVSTSTQVQEEDWCSRESGWISPLAPEEPSVVSEITASMVATMRVATEEVVLEP
QPEQVVITLPEHGRNVPPGLAEQDTASPIEVSVLLPDLAENCPLCGVPSGGLRLLGKHFAVRHAGVPVITYE
CRKCAWRSPNSHSISCHVPKCRGRARMPSGDPGIACDLCEARFATEVGVAQHCRHVHPVEWVNVRLERR
GARGGGIKATKLWSVAEVELIRLIREHGDGATYQLIADELGRGKTAEQVRSKKRLLRIDTASNPDDAE
VEEERLESLAVRSSSRSPSLVATRVREAVARGESEGGEEIRAIALIRDVDQNPCLIETSASDIISKLGRDVD
GPKRPRPVVREQTQEKGWVRRRLARRKREYEAQYLYSRDQARLAAQILDGAASQECALPVDQVYGAFR
[1054] EKWETVGQFHGLGEFRTGARADNWEFYSPILAAEVKENLMRMANGTAPGPDRISKKALLDWDPRGEQL
ARLYTTWLIGGVIPRVFKECRTKLLPKSSDPVELQDIGGWRPVTIGSMVTRLFSRILTMRLTRACPINRQR
GFLASSSGCAENLLIFDEIVRRSRDGGPLAVVFDFARAFDSISHEHILCVLEEGGLDRHVIGLIRNSYVD
CVTRVGCVEGMTPIQMKVGKQGDPMSPLLFNLAMDPLIHKLETAGTGLKWGDLSIATLAFADDLVLV
SDSEEGMGRSLGILEKFCQLTGLRVQPRKCHGFFMDKGVVNGCGTWEICGSPHIMPPGESVRYLGVQVG
PGRGVMEPDLIPTVHTWIERISEAPLKPSQRMRLNSFALPRIYQADLGKVTVTKLAQIDGIVRKAVKKW
LHLSPSTCNGLLYSRNRDGGGLGLLKLERLIPSVRTKRIYRMSRSPDIWTRRMTSHSVSKSDWEMLWVQAG
GERGSAPVMGAVEAAPT DVERS PDY PDWRREENLAWSALRVQGVGADQFRGDR TSSSWIAEPASVGFA
QRHWAALALRAGVYPTREFLAGKEKSGAACRRCPARLESCSHILGQCPFVQANRIARHNKVCVLLAT
EAERFGWTVIREFRLEDAAGGLKIPDLVCKKADTVLIVDVTVRYEMDGETLKRAASEKVKHYPVGQOI
TDKVGGRCFKVMGFVVGARGKWPA SNNTVLAELGVPAGRMRTFARLVSRRTL LYSLDILRDFMREPAGR
GTRVALIPAATGAANMGSDYKDHDGDYKDHDIDYKDDDDKK**

10. R2A OrLaSyn C-3F 转染的 mRNA 序列:

AGGAAUAAACUAGUAUUCUUCUGGUCCCCACAGACUCAGAGAGAACCCGCCACCAUGGGCACCGA
UACCGUUUAUGUGGGUCAAGAUUAUCCAAGCGGCCUGUCCAAGCGCGUUCGGGUCGUCUGGUU
GCUGGUCCAAUGCUGCGGAGCGCAGCUGCCAUGCCCACGUGUUCGGUGCUGGUCACAUGUGGAA

[1055]

UUGGCGCACCAGCCUGCCGAGCGGUCGUUGGGACCAGCCGGCCCUGGAGAAGAGCCGCGUGCUGA
CCCGUAGCGUGGCUACCGCCACCGAUCCGGAGAUCACCAGCUACCCGGGCAAGAGCGUGAGCACC
AGCACCAGGUGCAGGAAGAGGACUGGUGCAGCCGUGAAAGCGGCUGGAUCAGCCGGGUCUGG
CUCCAGAGGAACCGAGCGUGGUGAGCGAGAUUACCGCUAGCAUGGUGGCUACCAUGCGUGUGGC
UACCGAGGAAGUGGUGCUGGAACCGCAGCCGGAGCAGGUGGUGACCAUUCUGCCAGAACAUGGU
CGUAACGUGCCCGCCGGGUCUGGCUGAACAGGACACCGCGAGCCCGAUUGAGGUGAGCGUGCUGCU
GCCGGAUCUGGCGGAGAACUGCCACUGUGCGGUGUGCCGAGCGGGGUCUGCGUCUGCUGGGCA
AGCACUUCGCGGUGCGUCAUGCUGGUGUGCCAGUGACCUACGAGUGCCGCAA AUGCGCGUGGCG
UAGCCCGAACAGCCACAGCAUCAGCUGCCAUGUGCCGAAGUGCCGCGGUCGUGCUCGUAUGCCGA
GCGGCGACCCGGUAUUGCUUGCGAUCUGUGCGAAGCGCGCUUUGCUACCGAGGUGGGUGUGGC
CCAGCACAAGCGUCACGUGCACCCGGUGGAAUGGAACAAAGUGCGCCUGGAGCGUCGCGGUGCCC
GUGGCGGUGGCAUCAAGGCGACCAAACUGUGGAGCGUGGCUGAAGUGGAAACCCUGAUCCGUCU
GAUUCGUGAGCACGGCGACAGCGUGCGACCUAUCAGCUGAUCGUGAUGAACUGGGUCGUGGC
AAGACCGCCGAGCAGGUGCGUAGCAAGAAACGCCUGCUGCGUAUUGACACCGCCAGCAACAGCCC
GGACGAUGCGGAAGUGGAGGAAGAGCGCCUGGAGAGCCUGGCCGUGCGUAGCAGCAGCCGUAGC
CCGCCGAGCCUGGUGGCCACCCGCGUGCGUGAAGCGGUGGCUCGUGGUGAAAGCGAGGGUGGCG
AAGAGAUCCGCGCCAUUGCCGCGUGAUCGUGACGUGGAUCAGAACCCGUGCCUGAUUGAAACC
AGCGCCAGCGACAUAUAGCAAGCUGGGCCGUCGCGUGGACGGCCGAAACGCCCGCGUCCGGU
GGUGCGUGAACAGACCCAGGAGAAGGGUUGGGUGCGUCGCCUGGCCCGUCGAAACGCGAAUAC
CGUGAGGCUCAGUACCGUAUAGCCGCGACCAGGCUCGUCUGGCUGCGCAGAUCCUGGACGGCGC
CGCUAGCCAGGAGUGCGCCCUGCCAGUGGACCAGGUGUAUGGUGCUUCCGUGAGAAGUGGGAA
ACCGUGGGCCAGUUCCACGGUCUGGGCGAAUUCGCAACCGGUGCGCGUGCUGAUAAACUGGGAGU
UCUACAGCCCGAUCCUGGCCGCGGAAGUGAAGGAGAACCUGAUGCGCAUGGCUAAUGGUACCGCC
CCAGGUCCAGACCGUAUAGCAAGAAAGCCCUGCUGGAUUGGGACCCGCGUGGUGAACAGCUGG
CCCGUCUGUAUACCACCGUGCUGAUCGGUGGCGUGAUUCCGCGCGUGUUAAGGAAUGCCGUACC
AAGCUGCUGCCGAAGAGCAGCGACCCGGUGGAGCUGCAGGAUAUCGGUGGCUGGCGUCCGGUGA
CCAUUGGUAGCAUGGUGACCCGCCUGUUUAGCCGUAUCCUGACCAUGCGUCUGACCCGCGCGUGC
CCGAUUAACCCGCGCCAGCGUGGCUUCCUGGCCAGCAGCAGCGGUUGCGCCGAGAACCUGCUGAU
CUUUGACGAGAUUGUGCGUCGAGCCGUCGCGAUGGUGGCCACUGGCCGUGGUGUUCGUGGAC
UUUGCCCGUGCCUUCGACAGCAUCAGCCACGAACACAUCUGUGCGUGCUGGAAGAGGGUGGCCU
GGACCGUCACGUGAUCGGUCUGAUUCGUAACAGCUACGUGGAUUGCGUGACCCGUGUGGGCUGC
GUGGAGGGUAUGACCCCGCCGAUCCAGAUGAAAGUGGGCGUGAAACAGGGUGACCCGAUGAGCC

[1056]

CGCUGCUGUUUAACCUUGGCGAUGGACCCGCGUAUCCACAAGCUGGAAACCGCUGGUACCGGCCUG
AAAUGGGGCGACCUGAGCAUUGCCACCCUGGCUUUCGCCGACGAUCUGGUGCUGGUGAGCGAUA
GCGAAGAGGGUAUGGGCCGUAGCCUGGGCAUUCUGGAGAAGUUCUGCCAGCUGACCGGUCUGCG
UGUGCAGCCGCGUAAGUGCCACGGCUUCUUUAUGGAUAAAGGUGUGGUGAACGGUUGCGGCACC
UGGAAAUCUGCGGCAGCCCGAUCCACAUGAUUCCGCCGGGUGAGAGCGUGCGCUACCUGGGUG
UGCAAGUGGGUCCGGGUCGUGGUGUGAUGGAGCCAGACCUGAUUCCGACCGUGCACACCUGGAU
CGAACGCAUUAAGCGAAGCCCCGCGUAAACCGAGCCAGCGCAUGCGUGUGCUGAACAGCUUCGCUC
UGCCGCGUAUCAUCUACCAGGCCGACCUGGGCAAGGUGACCGUGACCAAACUGGCCCAGAUUGAU
GGUAUUGUGCGUAAGGCGGUGAAGAAAUGGCUGCACCUGAGCCCGAGCACCUGCAACGGCCUGC
UGUACAGCCGCAACCGUGAUGGUGGCCUGGGUCUGCUGAAGCUGGAGCGUCUGAUCCCGAGCGU
GCGCACCAAACGUAUCUACCGCAUGAGCCGUAGCCCGGACAUCUGGACCCGUCGCAUGACCAGCC
ACAGCGUGAGCAAGAGCGAUUGGGAAAUGCUGUGGGUGCAGGCGGGUGGCGAGCGUGGCAGCGC
CCCGUUAUGGGUGCGGUGGAAGCGGCUCCAACCGACGUGGAGCGUAGCCCGGACUACCCAGAUU
GGCGUCGCGAAGAGAACCUGGCCUGGAGCGCGCUGCGUGUGCAGGGCGUGGGUGCCGACCAGUU
CCGCGGCGACCGUACCAGCAGCAGCUGGAUCGCUGAACCAGCCAGCGUGGGCUUCGCCAGCGUC
ACUGGCUGGGCGCUCUGGCGCUGCGUGCUGGCGUGUAUCCAACCCGCGAAUUCUGGGCCCGUGGC
AAGGAGAAGAGCGGUGCUGCCUGCCGUCGUGCCAGCUCGUCUGGAGAGCUGCAGCCACAUUCU
GGGCCAGUGCCCGUUUGUGCAGGCUAACCGCAUUGCCCGUCACAACAAAGUGUGCGUGCUGCUGG
CUACCGAAGCCGAGCGUUUCGGCUGGACCGUGAUCCGCGAAUUCGUCUGGAGGACGCUGCCGGC
GGCCUGAAGAUCCCGGAUCUGGUGUGCAAGAAAGCCGACACCGUGCUGAUUGUGGACGUGACCG
UGCGCUACGAAAUGGACGGUGAAACCCUGAAGCGUGCGGCUAGCGAGAAGGUGAAGCACUAUCU
GCCGGUGGGUCAGCAGAUACCGAUAAAGUGGGUGGGCCGCGUCUCAAAGUGAUGGGCUUUCG
GUGGGUGCUCGUGGCAAGUGGCCGGCCAGCAACAAUACCGUGCUGGCCGAACUGGGCGUGCCGGC
UGGUCGCAUGCGUACCUUCGCGCGUCUGGUGAGCCGUCGCACCCUGCUGUAUAGCCUGGACAUC
UGCGCGACUUAUGCGUGAACCAGCAGGUCGUGGCACCCGUGUGGCACUGAUUCCAGCAGCAACC
GGCGCAGCAAAUAUGGGGAGCGACUACAAAGACCAUGACGGUGAUUAUAAAGAUCAUGACAUCG
AUUACAAGGAUGACGAUGACAAGAAGUGAUGACUCGAGCUGGUACUGCAUGCACGCAAUGCUAG
CUGCCCCUUCGCGUCCUGGGUACCCCGAGUCUCCCCGACCUCGGGUCCCAGGUAUGCUCCCACC
UCCACCUGCCCCACUACCAACCUCUGCUAGUCCAGACACCUCCCAAGCACGCAGCAAUGCAGCU
CAAAACGCUUAGCCUAGCCACACCCCCACGGGAAACAGCAGUGAUUAACCUUUAAGCAAUAAACGA
AAGUUUAACUAAGCUAUACUAAACCCAGGGUUGGUCAAUUCGUGCCAGCCACACCCUGGAGCU
AGCAAAAAAAAAAAAAAAAAAAAAAAAAAAGCAUAUGACUAAAAAAAAAAAAAAAAAAAAA

AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

11. R2A OrLa 天然 C-3F 质粒 ORF 序列:

ATGGGAACAGATACAGTGTATGTCGGCCAGGACTACCCTTCTGGCTTATCAAAACGGGTACCAGCACG
GTTAGTGGCGGGACCGATGCTGCGAGAGCGAAGCTGTCACGCCCATGTGTTTAGGGCTGGACACATG
TGGAACCTGGCGAACCAGCCTTCCGAGCGGGCGCTGGGACCAGCCCCTTTGGAGAAGTCTCGGGTC
CTAACCCGGTTCGGTGGCGACGGCCACCGACCCCGAAATTACCTCTTACCCAGGAAAGTCCGTATCGA
CAAGTACGCAGGTTTCAGGAGGAGGACTGGTGTAGCCGGGAGAGCGGGTGGATCTCGCCAGGACTTG
CTCCTGAAGAACCCTCGGTGGTGTCCGAAATTACAGCCTCCATGGTAGCGACAATGAGGGTAGCAAC
CGAGGAGGTTCGTGCTGGAACCAAGCCTGAACAGGTTCGTACAATACTGCCGGAGCATGGTCGAAA
CGTTCTTCCGGGGCTGGCAGAACAGGACACCGCCAGCCCCATAGAAGTCTCGGTGCTCTCTCCAGAC
CTCGCTGAGAACTGCCATTGTGTGGCGTGCCGAGCGGGGGCCTACGCTTGCTCGGGAAGCATTG
CTGTCCGACATGCGGGGGTGCCTGTAACGTATGAGTGCCGTAAGTGTGCGTGGCGGAGCCCCAACAG
CCACTCAATCTCGTGTACGTCCCCAAATGCCGGGGGCGTGCGCGGATGCCAGTGGCGATCCAGGG
ATCGCCTGCGATCTCTGTGAAGCCCGTTTGCCACGGAGGTTGGGGTCGCCAACACAAGCGGCACG
TTCATCCGGTGGAGTGGAACAAGGTGAGGCTGGAAGGAGAGGTGCGCGCGGAGGGGGAATTAAGG
CGACGAAGCTCTGGAGTGTAGCGGAGGTAGAGACGCTAATCCGGCTCATCCGTGAGCACGGAGATTC
AGGTGCCACTTACCAGCTCATTGCCGATGAGCTGGGAAGGGGCAAGACGGCCGAACAGGTGAGGAG
TAAAAAGAGGCTCCTGCGCATAGATACGGCAAGCAATAGCCCAGATGATGCAGAGGTTGAGGAGGAG
AGGTTGGAATCTCTGGCGGTTCCGTCCTCGTCACGGTCACCCCGAGCCTGGTGGCGACCAGGGTCA
GGGAGGCAGTTGCCAGGGGTGAATCAGAAGGTGGCGAGGAGATCAGGGCTATTGCTGCTCTCATTAG
GGACGTAGATCAGAATCCTTGCTGATTGAAACCTCGGCGTCGGACATCATCTCGAAGCTGGGAAGG
AGGGTGGATGGGCCCCAAGAGACCCAGGCCCGTTGTCAGAGAACAGACCCAAGAGAAGGGATGGGTA
AGGCGGCTTGCCCGGCGGAAAAGGGAGTACAGAGAAGCGCAGTACCTGTACTCAAGGGATCAAGCA
AGGCTGGCGGCCAGATCCTCGATGGTGCCGCCAGCCAGGAATGCGCCCTCCCGGTGGACCAGGTCT
ACGGAGCGTTCCGTGAGAAATGGGAAACCGTAGGGCAGTTCCACGGACTTGGTGAGTTCCGGACGG
GTGCACGCGCAGACAACTGGGAGTTCTACTCTCCAATTCTGGCGGTGAGGTGAAAGAAAACCTAAT
GAGAATGGCTAACGGCACGGCCCCGGGACCAGACAGGATAAGCAAAAAGGCTCTGCTTGACTGGGA
CCCCCGGGGTGAGCAACTGGCACGGCTGTACACGACGTGGCTGATCGGTGGGGTCATACCAAGGGTC
TTCAAGGAGTGCAGGACTAAGCTGCTACCGAAATCCAGCGACCCGGTCGAGTTGCAGGACATCGGTG
GATGGAGGCCGGTGACGATTGGGTGATGGTGACTAGGCTGTTTCAGTCGGATTCTAACGATGAGGCTA
ACCCGAGCCTGTCCGATCAATCCGAGGCAGCGCGGTTTCTTGGCCTCCTCGAGTGGATGCGCGGAAA

[1057]

[1058]

ACCTGTTGATCTTTGACGAGATCGTCAGGCGCTCGAGGCGGGACGGGGGGCCGCTGGCAGTGGTGTT
TGTGGACTTTGCGAGGGCCTTTGACTCCATCTCACATGAACATATCCTGTGTGTTCTCGAAGAAGGCG
GGCTTGACAGGCACGTTATCGGGTTGATCCGAAACTCGTACGTGGATTGCGTGACCAGGGTGGGTTGT
GTCGAGGGCATGACACCACCAATACAAATGAAGGTTGGAGTGAAGCAGGGAGACCCCATGTCCCCCT
TGCTCTTCAACCTGGCTATGGATCCCCTCATCCATAAACTCGAGACGGCCGGAACCTGGACTGAAATGG
GGCGATCTTTCAATCGCCACGCTGGCCTTTGCCGACGATCTGGTGCTGGTGAGTGAAGTCTGAGGAAG
GCATGGGGAGGAGTCTCGGGATTTTGAGAAAGTTTGGCAACTGACTGGGCTGAGGGTTCAGCCCAG
GAAGTGTCACGGTTTCTTTATGGACAAGGGCGTGGTGAACGGCTGTGGAACCTGGGAAATCTGTGGG
TCACCGATCCACATGATTCCCCGGGGGAATCAGTTCGTTATTTGGGAGTCCAGGTAGGCCCGGGGCG
CGGCGTGATGGAACCGGATCTTATCCCTACGGTCCACACGTGGATCGAAAGGATCTCGGAGGCTCCTC
TAAAGCCCTCACAACGCATGAGGGTTTTGAACTCATTCGCTCTCCCCCGGATAATTTACCAGGCCGAT
CTAGGGAAGGTTACGGTAACCAAATTGGCCCAGATAGATGGGATTGTCCGGAAGGCTGTGAAGAAGT
GGCTCCATTTGTACCATCCACGTGCAATGGACTGCTGTATTACGGAACCGCGACGGTGGTTTGGGC
CTCCTAAAGCTGGAAAGACTAATCCCATCCGTGCGCACGAAGCGTATCTATCGGATGTCCAGGTCTCC
GGATATCTGGACACGGCGAATGACCAGCCATTCTGTGTCAAATCTGACTGGGAGATGTTGTGGGTCC
AAGCGGGAGGTGAGAGGGGCAGTGCACCTGTAATGGGTGCCGTGGAGGCTGCCCCGACCGATGTGG
AGAGATCGCCAGACTACCCAGACTGGCGGCGTGAGGAAAACCTGGCATGGTCGGCCCTGCGGGTGC
AGGGTGTGGGTGCAGACCAGTTTCGAGGCGACAGGACCAGCAGCTCTTGGATCGCCGAGCCCGCTTC
GGTTGGGTTCGCGCAGCGCCACTGGTTGGCTGCCCTGGCGCTGAGGGCTGGGGTGTATCCGACTCGG
GAGTTTCTGGCTCGGGTAAGGAAAAGTCAGGAGCAGCTTGCAGACGCTGCCCCGCCAGGTTGGAA
TCATGTTACACATACTTGGGCAATGTCCGTTTCGTTACAGGCGAACAGAATTGCGAGGCACAACAAGGT
GTGTGTGCTCTTGGCCACGGAGGCGGAGAGGTTTCGGCTGGACGTAATAAGGGAGTTCCGTCTTGAG
GACGCCGCTGGCGGTCTCAAGATACCCGACCTGGTTTGCAAGAAGGCCGACACAGTTCTCATTGTCG
ACGTGACCGTCCGGTACGAGATGGATGGAGAGACGCTAAAAAGGGCCGCATCGGAGAAGGTGAAAC
ACTATCTCCAGTAGGGCAACAGATAACGGACAAGGTTCGGAGGGCGTTGCTTTAAAGTCATGGGGTT
CCCTGTAGGTGCTAGGGGAAAGTGGCCGGCGAGCAACAACACAGTTTTGGCTGAGTTAGGCGTCCCT
GCAGGTCGGATGAGGACCTTTGCCAGGCTGGTGAGCCGGAGGACTCTTCTTTATCTTTGGATATATTG
AGGGACTTCATGCGTGAGCCGGCCGGCAGGGGAACTCGGGTTGCTCTCATCCCTGCGGCAACGGGTG
CCGCGAATATGGGGAGCGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAG
GATGACGATGACAAGAAGTGA

12. R2A OrLa 天然 C-3F 氨基酸序列:

[1059]

MGTDTVYVGQDYPSGLSKRVPARLVAGPMLRERSCHAHVFRAGHMWNWRTSLPSGRWDQPALEKSRVL
TRSVATATDPEITSYPGKSVSTSTQVQEEDWCSRESGWISPLAPEEPSVVSEITASMVATMRVATEEVVLEP
QPEQVVITLPEHGRNVPPGLAEQDTASPIEVSVLLPDLAENCPLCGVPSGGLRLLGKHFAVRHAGVPVITYE
CRKCAWRSPNSHSISCHVPKCRGRARMPSGDPGIACDLCEARFATEVGVAQHKRHVHPVEWNKVRLERR
GARGGGIKATKLWSVAEVETLIRLIREHGDSGATYQLIADELGRGKTAEQVRSKRLLRIDTASNPDDAE
VEEERLESLAVRSSSRSPSLVATRVREAVARGESEGGEEIRAIAALIRDVDQNPCLIETSASDIISKLGRDVD
GPKRPRPVVREQTQEKGWVRRLARRKREYREAQYLYSRDQARLAAQILDGAASQECALPVDQVYGAFR
EKWETVGQFHGLGEFRTGARADNWEFYSPILAAEVKENLMRMANGTAPGPDRISKKALLDWDPRGEQL
ARLYTTWLIGGVIPRVFKECRTKLLPKSSDPVELQDIGGWRPVTIGSMVTRLFSRILTMRLTRACPINRQR
GFLASSSGCAENLLIFDEIVRRSRRDGGPLAVVVFDFARAFDSISHEHILCVLEEGGLDRHVIGLIRNSYVD
CVTRVGCVEGMTPIQMKVGKQGDPMSPLLFNLAMDPLIHKLETAGTGLKWGDLSIATLAFADDLVLV
SDSEEGMGRSLGILEKFCQLTGLRVQPRKCHGFFMDKGVVNGCGTWEICGSPHIMPPGESVRYLGVQVG
PGRGVMEDLIPTVHTWIERISEAPLKPSQRMVNLNSFALPRIYQADLGKVTVTKLAQIDGIVRKAVKKW
LHLSPTCNGLLYSRNRDGGGLGLLKLERLIPSVRTKRIYRMSRSPDIWTRRMTSHSVSKSDWEMLWVQAG
GERGSAPVMGAVEAAPTDDVERSPDYPDWRREENLAWSALRVQGVGADQFRGDRTSSSWIAEPASVGFA
QRHWLAALALRAGVYPTREFLAGKEKSGAACRRCPARLESCSHILGQCPFVQANRIARHNKVCVLLAT
EAERFGWTVIREFRLEDAAGGLKIPDLVCKKADTVLIVDVTVRYEMDGETLKRAASEKVKHYPVGQQI
TDKVGGRCFKVMGFPVGARGKWPAASNNTVLAELGVPAGRMRTFARLVSRRTLLYSLDILRDFMREPAGR
GTRVALIPAATGAANMGSDYKDHDGDYKDHDIDYKDDDDKK*

13. R2A OrLa Syn N-DYK 质粒 ORF 序列:

ATGGATTACAAGGATGACGACGATAAGGGTACCGATACCGTTTATGTGGGTCAAGATTATCCAAGCGG
CCTGTCCAAGCGCGTTCCGGCTCGTCTGGTTGCTGGTCCAATGCTGCGCGAGCGCAGCTGCCATGCCC
ACGTGTTCCGTGCTGGTCACATGTGGAATTGGCGCACCAGCCTGCCGAGCGGTCTGGGACCAGCC
GGCCCTGGAGAAGAGCCGCGTGCTGACCCGTAGCGTGGCTACCGCCACCGATCCGGAGATCACCAGC
TACCCGGGCAAGAGCGTGAGCACCAGCACCAGGTGCAGGAAGAGGACTGGTGCAGCCGTGAAAG
CGGCTGGATCAGCCCGGTCTGGCTCCAGAGGAACCGAGCGTGGTGAGCGAGATTACCGCTAGCATG
GTGGCTACCATGCGTGTGGCTACCGAGGAAGTGGTGCTGGAACCGCAGCCGGAGCAGGTGGTGACCA
TTCTGCCAGAACATGGTCGTAACGTGCCGCCGGGTCTGGCTGAACAGGACACCGCGAGCCCGATTGA
GGTGAGCGTGCTGCTGCCGGATCTGGCGGAGAACTGCCCACTGTGCGGTGTGCCGAGCGGCGGTCTG
CGTCTGCTGGGCAAGCACTTCGCGGTGCGTCATGCTGGTGTGCCAGTGACCTACGAGTGCCGCAAAT
GCGCGTGGCGTAGCCCGAACAGCCACAGCATCAGCTGCCATGTGCCGAAGTGCCGCGGTCTGTGCTCG

[1060]

TATGCCGAGCGGCGACCCGGGTATTGCTTGCGATCTGTGCGAAGCGCGCTTTGCTACCGAGGTGGGTG
TGGCCCAGCACAAAGCGTCACGTGCACCCGGTGGAATGGAACAAAGTGCGCCTGGAGCGTCGCGGTG
CCCGTGGCGGTGGCATCAAGGCGACCAAACTGTGGAGCGTGGCTGAAGTGAAACCCTGATCCGTCT
GATTCGTGAGCACGGCGACAGCGGTGCGACCTATCAGCTGATCGCTGATGAACTGGGTCTGGCAAG
ACCGCCGAGCAGGTGCGTAGCAAGAAACGCCTGCTGCGTATTGACACCGCCAGCAACAGCCCGGAC
GATGCGGAAGTGGAGGAAGAGCGCCTGGAGAGCCTGGCCGTGCGTAGCAGCAGCCGTAGCCCGCCG
AGCCTGGTGGCCACCCGCGTGCGTGAAGCGGTGGCTCGTGGTGAAAGCGAGGGTGGCGAAGAGATC
CGCGCCATTGCCGCGCTGATCCGTGACGTGGATCAGAACCCGTGCCTGATTGAAACCAGCGCCAGCG
ACATCATTAGCAAGCTGGGCCGTCGCGTGGACGGCCCGAAACGCCCGCGTCCGGTGGTGCCTGAACA
GACCCAGGAGAAGGGTTGGGTGCGTCGCCTGGCCCGTCGCAAACGCGAATACCGTGAGGCTCAGTA
CCTGTATAGCCGCGACCAAGGCTCGTCTGGCTGCGCAGATCCTGGACGGCGCCGCTAGCCAGGAGTGC
GCCCTGCCAGTGGACCAGGTGTATGGTGCTTTCCGTGAGAAGTGGGAAACCGTGGGCCAGTTCCACG
GTCTGGGCGAATTTGCGACCGGTGCGCGTGCTGATAACTGGGAGTTCTACAGCCCGATCCTGGCCGCG
GAAGTGAAGGAGAACCTGATGCGCATGGCTAATGGTACCGCCCCAGGTCCAGACCGTATTAGCAAGA
AAGCCCTGCTGGATTGGGACCCGCGTGGTGAACAGCTGGCCCGTCTGTATACCACCTGGCTGATCGGT
GGCGTGATTCCGCGCGTGTTCAAGGAATGCCGTACCAAGCTGCTGCCGAAGAGCAGCGACCCGGTGG
AGCTGCAGGATATCGGTGGCTGGCGTCCGGTGACCATTGGTAGCATGGTGACCCGCCTGTTTAGCCGT
ATCCTGACCATGCGTCTGACCCGCGCGTGCCCGATTAACCCGCGCCAGCGTGGCTTCCTGGCCAGCAG
CAGCGGTTGCGCCGAGAACCTGCTGATCTTTGACGAGATTGTGCGTCCGAGCCGTGCGGATGGTGCC
CCACTGGCCGTGGTGTTCTGTGGACTTTGCCCCGTGCCCTTCGACAGCATCAGCCACGAACACATTCTGTG
CGTGCTGGAAGAGGGTGGCCTGGACCGTCACGTGATCGGTCTGATTCTGTAACAGCTACGTGGATTGC
GTGACCCGTGTGGGCTGCGTGGAGGGTATGACCCCGCCGATCCAGATGAAAGTGGGCGTGAAACAG
GGTGACCCGATGAGCCCGCTGCTGTTTAACCTGGCGATGGACCCGCTGATCCACAAGCTGGAAACCG
CTGGTACCGGCCTGAAATGGGGCGACCTGAGCATTGCCACCCCTGGCTTTTCGCCGACGATCTGGTGCTG
GTGAGCGATAGCGAAGAGGGTATGGGCCGTAGCCTGGGCATTCTGGAGAAGTTCTGCCAGCTGACCG
GTCTGCGTGTGCAGCCGCGTAAGTGCCACGGCTTCTTTATGGATAAAGGTGTGGTGAACGGTTGCGGC
ACCTGGGAAATCTGCGGCAGCCCGATCCACATGATTCCGCCGGGTGAGAGCGTGCGCTACCTGGGTG
TGCAAGTGGGTCCGGGTCTGGTGTGATGGAGCCAGACCTGATTCCGACCGTGACACCTGGATCGA
ACGCATTAGCGAAGCCCCGCTGAAACCGAGCCAGCGCATGCGTGTGCTGAACAGCTTCGCTCTGCCG
CGTATCATCTACCAGCCGACCTGGGCAAGGTGACCGTGACCAAACTGGCCCAGATCGATGGTATTGT
GCGTAAGGCGGTGAAGAAATGGCTGCACCTGAGCCCGAGCACCTGCAACGGCCTGCTGTACAGCCG
CAACCGTGATGGTGGCCTGGGTCTGCTGAAGCTGGAGCGTCTGATCCCGAGCGTGCGCACCAAAACGT

ATCTACCGCATGAGCCGTAGCCCGGACATCTGGACCCGTCGCATGACCAGCCACAGCGTGAGCAAGA
GCGATTGGGAAATGCTGTGGGTGCAGGCGGGTGGCGAGCGTGGCAGCGCCCCGGTTATGGGTGCGGT
GGAAGCGGCTCCAACCGACGTGGAGCGTAGCCCGGACTACCCAGATTGGCGTCGCGAAGAGAACCT
GGCCTGGAGCGCGCTGCGTGTGCAGGGCGTGGGTGCCGACCAGTTCCGCGGCGACCGTACCAGCAG
CAGCTGGATCGCTGAACCAGCCAGCGTGGGCTTCGCCCAGCGTCACTGGCTGGCGGCTCTGGCGCTG
CGTGCTGGCGTGTATCCAACCCGCGAATTCCTGGCCCGTGGCAAGGAGAAGAGCGGTGCTGCCTGCC
GTCGCTGCCCAGCTCGTCTGGAGAGCTGCAGCCACATTCTGGGCCAGTGCCCGTTTGTGCAGGCTAA
CCGCATTGCCCCTCACAACAAAGTGTGCGTGTGCTGGCTACCGAAGCCGAGCGTTTCGGCTGGACC
GTGATCCGCGAATTTCTGCTGGAGGACGCTGCCGGCGGCGCTGAAGATCCCGGATCTGGTGTGCAAGA
AAGCCGACACCGTGCTGATTGTGGACGTGACCGTGCGTACGAAATGGACGGTGAAACCTGAAGC
GTGCGGCTAGCGAGAAGGTGAAGCACTATCTGCCGGTGGGTCAGCAGATCACCGATAAAGTGGGTGG
CCGCTGCTTCAAAGTGATGGGCTTTCCGGTGGGTGCTCGTGGCAAGTGCCCGGCCAGCAACAATACC
GTGCTGGCCGAACCTGGGCGTGCCGGCTGGTTCGATGCGTACCTTCGCGCGTCTGGTGAGCCGTCGCA
CCCTGCTGTATAGCCTGGACATTCTGCGCGACTTCATGCGTGAACCAGCAGGTCGTGGCACCCGTGTG
GCACTGATTCCAGCAGCAACCGGCGCAGCAAATTGATAA

[1061]

14. R2A OrLa Syn N-DYK 氨基酸序列:

MDYKDDDDKGTDTVYVGQDYPSGLSKRVPARLVAGPMLRERSCHAHVFRAGHMWNWRTSLPSGRWD
QPALEKSRVLTRS VATATDPEITSYPGKSVSTSTQVQEEDWCSRESGWISPLAPEEPSVVSEITASMVATMR
VATEEVVLEPQPEQVVTLPEHGRNVPPGLAEQDTASPIEVSVLLPDLAENCPLCGVPSGGLRLLGKHFAV
RHAGVPVTYECKERCAWRSPNSHSISCHVPKCRGRARMPSGDPGIACDLCEARFATEVGVAQHKRHVHPV
EWNKVRLERRGARGGGIKATKLWSVAEVETLIRLIREHGD SGATYQLIADELGRGKTAEQVRSKKRLLRI
DTASNSPDDAEVEEERLESLAVRSSSRSPSLVATRVREAVARGESEGGEEIRAI AALIRDVDQNPCLIETSAS
DIISKLGRRVDGPKRPRPVVREQTQEKGWVRRRLARRKREYREAYLYSRDQARLAAQILDGAASQECAL
PVDQVYGAFREK WETVGQFHGLGEFRTGARADNWEFYSPILAAEVKENLMRMANGTAPGPDRIKKAL
LDWDPRGEQLARLYTTWLIGGVIPRVFKECRTKLLPKSSDPVELQDIGGWRPVTIGSMVTRLFSRILTMRL
TRACPINPRQRGFLASSSGCAENLLIFDEIVRRSRRDGGPLAVVFVDFARAFDSISHEHILCVLEEGGLDRH
VIGLIRNSYVDCVTRVGCVEGMTPIQMKVGVKQGDPMSPLLFNLA MDPLIHKLETAGTGLKWGDL SIAT
LAFADDLVLVSDSEEGMGRSLGILEKFCQLTGLRVQPRKCHGFFMDKG VVNGCGTWEICGSPHIMIPPGES
VRYLG VQVGPGRGVMEPDLIPTVHTWIERISEAPLKPSQRMRLNSFALPRIIYQADLGKVTVTKLAQIDG
IVRKAVKKWLHLSPTCNGLLYSRNRDGGGLLLKLERLIPSVRTKRIYRMSRSPDIWTRRMTSHSVSKSD
WEMLWVQAGGERGSAPVMGAVEAAPT DVERS PDYDWRREENLAWSALRVQGVGADQFRGDR TSSS

WIAEPASVGFAQRHWLAALALRAGVYPTREFLARGKEKSGAACRRCPARLESCSHILGQCPFVQANRIAR
HNKVCVLLATEAERFGWTVIREFRLEDAAGGLKIPDLVCKKADTVLIVDVTVRYEMDGETLKRAASEKV
KHYLPVGQQITDKVGGRCFKVMGFPVGARGKWPA SNNTVLAELGVPAGRMRTFARLVSRRTLTLYSLDIL
RDFMREPAGRGTRVALIPAATGAAN**

15. R2A OrLa_s Syn N-DYK 质粒 ORF 序列:

[1062] ATGGATTACAAGGATGACGACGATAAGGGTACCGATCTGGCGGAGAACTGCCCACTGTGCGGTGTGC
CGAGCGGCGGTCTGCGTCTGCTGGGCAAGCACTTCGCGGTGCGTCATGCTGGTGTGCCAGTGACCTA
CGAGTGCCGCAAATGCGCGTGGCGTAGCCCGAACAGCCACAGCATCAGCTGCCATGTGCCGAAGTGC
CGCGGTCTGCTCGTATGCCGAGCGGCGACCCGGGTATTGCTTGCGATCTGTGCGAAGCGCGCTTTGC
TACCGAGGTGGGTGTGGCCCAGCACAAAGCGTCACGTGCACCCGGTGGAATGGAACAAAGTGCGCCT
GGAGCGTCGCGGTGCCCCGTGGCGGTGGCATCAAGGCGACCAAACTGTGGAGCGTGGCTGAAGTGGA
AACCTGATCCGTCTGATTCTGTGAGCACGGCGACAGCGGTGCGACCTATCAGCTGATCGCTGATGAAC
TGGGTCGTGGCAAGACCGCCGAGCAGGTGCGTAGCAAGAAACGCCTGCTGCGTATTGACACCGCCA
GCAACAGCCCGGACGATGCGGAAAGTGGAGGAAGAGCGCCTGGAGAGCCTGGCCGTGCGTAGCAGCA
GCCGTAGCCCGCCGAGCCTGGTGGCCACCCGCGTGCGTGAAGCGGTGGCTCGTGGTGAAAGCGAGG
GTGGCGAAGAGATCCGCGCCATTGCCGCGCTGATCCGTGACGTGGATCAGAACCCGTGCCTGATTGA
AACAGCGCCAGCGACATCATTAGCAAGCTGGGCCGTGCGGTGGACGGCCCGAAACGCCCGCGTCC
GGTGGTGCGTGAACAGACCCAGGAGAAGGGTTGGGTGCGTCGCCTGGCCCGTCGCAAACGCGAATA
CCGTGAGGCTCAGTACCTGTATAGCCGCGACCAGGCTCGTCTGGCTGCGCAGATCCTGGACGGCGCC
GCTAGCCAGGAGTGCGCCCTGCCAGTGGACCAGGTGTATGGTGCTTTCCGTGAGAAGTGGGAAACCG
TGGGCCAGTTCCACGGTCTGGGCGAATTTGCGACCGGTGCGCGTGCTGATAACTGGGAGTTCTACAG
CCCGATCCTGGCCGCGGAAGTGAAGGAGAACCTGATGCGCATGGCTAATGGTACCGCCCCAGGTCCA
GACCGTATTAGCAAGAAAGCCCTGCTGGATTGGGACCCGCGTGGTGAACAGCTGGCCCGTCTGTATAC
CACCTGGCTGATCGGTGGCGTGATTCCGCGCGTGTTCAGGAATGCCGTACCAAGCTGCTGCCGAAG
AGCAGCGACCCGGTGGAGCTGCAGGATATCGGTGGCTGGCGTCCGGTGACCATTTGGTAGCATGGTGA
CCCGCCTGTTTAGCCGTATCCTGACCATGCGTCTGACCCGCGCGTGCCGATTAACCCGCGCCAGCGT
GGCTTCCTGGCCAGCAGCAGCGTTGCGCCGAGAACCTGCTGATCTTTGACGAGATTGTGCGTCGCA
GCCGTGCGGATGGTGGCCCACTGGCCGTGGTGTTCTGTGGAATTTGCCCGTGCCCTTCGACAGCATCAGC
CACGAACACATTCTGTGCGTGCTGGAAGAGGGTGGCCTGGACCGTCACGTGATCGGTCTGATTCGTA
ACAGCTACGTGGATTGCGTGACCCGTGTGGGCTGCGTGGAGGGTATGACCCCGCCGATCCAGATGAA
AGTGGGCGTGAAACAGGGTGACCCGATGAGCCCGCTGCTGTTTAACCTGGCGATGGACCCGCTGATC

[1063]

CACAAGCTGGAAACCGCTGGTACCGGCCTGAAATGGGGCGACCTGAGCATTGCCACCCTGGCTTTTCG
CCGACGATCTGGTGCTGGTGAGCGATAGCGAAGAGGGTATGGGCCGTAGCCTGGGCATTCTGGAGAA
GTTCTGCCAGCTGACCGGTCTGCGTGTGCAGCCGCGTAAGTGCCACGGCTTCTTTATGGATAAAGGTG
TGGTGAACGGTTGCGGCACCTGGGAAATCTGCGGCAGCCCGATCCACATGATTCGCGCCGGGTGAGAG
CGTGCGCTACCTGGGTGTGCAAGTGGGTCCGGGTCTGGTGTGATGGAGCCAGACCTGATTCCGACC
GTGCACACCTGGATCGAACGCATTAGCGAAGCCCCGCTGAAACCGAGCCAGCGCATGCGTGTGCTGA
ACAGCTTCGCTCTGCCGCGTATCATCTACCAGGCCGACCTGGGCAAGGTGACCGTGACCAAACCTGGC
CCAGATCGATGGTATTGTGCGTAAGGCGGTGAAGAAATGGCTGCACCTGAGCCCGAGCACCTGCAAC
GGCCTGCTGTACAGCCGCAACCGTGATGGTGGCCTGGGTCTGCTGAAGCTGGAGCGTCTGATCCCGA
GCGTGCGCACCAAACGTATCTACCGCATGAGCCGTAGCCCGGACATCTGGACCCGTCGCATGACCAG
CCACAGCGTGAGCAAGAGCGATTGGGAAATGCTGTGGGTGCAGGCGGGTGGCGAGCGTGGCAGCGC
CCCGGTTATGGGTGCGGTGGAAGCGGCTCCAACCGACGTGGAGCGTAGCCCGGACTACCCAGATTGG
CGTCGCGAAGAGAACCTGGCCTGGAGCGCGCTGCGTGTGCAGGGCGTGGGTGCCGACCAGTTCCGC
GGCGACCGTACCAGCAGCAGCTGGATCGCTGAACCAGCCAGCGTGGGCTTCGCCAGCGTCACTGGC
TGGCGGCTCTGGCGCTGCGTGTGGCGTGTATCCAACCCGCGAATTCCTGGCCCGTGGCAAGGAGAA
GAGCGGTGCTGCCTGCCGTCGCTGCCAGCTCGTCTGGAGAGCTGCAGCCACATTCTGGGCCAGTGC
CCGTTTGTGCAGGCTAACCGCATTGCCCGTCACAACAAAGTGTGCGTGCTGCTGGCTACCGAAGCCG
AGCGTTTCGGCTGGACCGTGATCCGCGAATTCGTCGTGGAGGACGCTGCCGGCGGCCTGAAGATCCC
GGATCTGGTGTGCAAGAAAGCCGACACCGTGCTGATTGTGGACGTGACCGTGCGCTACGAAATGGAC
GGTGAAACCCTGAAGCGTGCGGCTAGCGAGAAGGTGAAGCACTATCTGCCGGTGGGTGAGCAGATC
ACCGATAAAGTGGGTGGCCGCTGCTTCAAAGTGATGGGCTTTCCGGTGGGTGCTCGTGGCAAGTGGC
CGGCCAGCAACAATACCGTGCTGGCCGAACTGGGCGTGCCGGCTGGTCGCATGCGTACCTTCGCGCG
TCTGGTGAGCCGTCGCACCCTGCTGTATAGCCTGGACATTCTGCGCGACTTCATGCGTGAACCAGCAG
GTCGTGGCACCCGTGTGGCACTGATTCCAGCAGCAACCGGCGCAGCAAATTGATAA

16. R2A OrLa_s Syn N-DYK 氨基酸序列:

MDYKDDDDKGTDLAENCPLCGVPSGGLRLLGKHFAVRHAGVPVTYEKRKCAWRSPNSHSISCHVPKCR
GRARMPSGDPGIACDLCEARFATEVGVAQHKRHVHPVEWNVRLERRGARGGGIKATKLWSVAEVETLI
RLIREHGD SGATYQLIADELGRGKTAEQVRSKKRLLRIDTASNPDDAEVEEERLESLAVRSSSRSPSLVA
TRVREAVARGESEGGEEIRAIAALIRDVDQNPCLietsASDIISKLGRRVDGPKRPRPVVREQTQEKGWVRR
LARRKREYREAQYLYSRDQARLAAQILDGAASQECALPVDQVYGAFREKWETVGQFHGLGEFRTGARA
DNWEFYSPILAAEVKENLMRMANGTAPGPDRISKKALLDWDPRGEQLARLYTTWLIGGVIPRVFKECRT

KLLPKSSDPVELQDIGGWRPVTIGSMVTRLFSRILTMRLTRACPINPRQRGFLASSSGCAENLLIFDEIVRRS
RRDGGPLAVVFVDFARAFDSISHEHILCVLEEGGLDRHVIGLIRNSYVDCVTRVGCVEGMTPIQMKVGV
KQGDPMSPLLFNLAMDPLIHKLETAGTGLKWGDLSIATLAFADDLVLVSDSEEGMGRSLGILEKFCQLTGL
RVQPRKCHGFFMDKGVVNGCGTWEICGSPHIMPPGESVRYLGVQVGPGRGVMEPDLIPTVHTWIERISE
APLKPSQRMRLVNSFALPRIYQADLGKVTVTKLAQIDGIVRKAVKKWLHLSPSTCNGLLYSRNRDGGGLG
LLKLERLIPSVRTKRIYRMSRSPDIWTRMTSHSVSKSDWEMLWVQAGGERGSAPVMGAVEAAPT DVER
SPDYPDWRREENLAWSALRVQGVGADQFRGDR TSSSWIAEPASVGFAQRHWAALALRAGVYPTREFL
ARGKEKSGAACRRCPARLESCSHILGQCPFVQANRIARHNKVCVLLATEAERFGWTVIREFRLEDAAGGL
KIPDLVCKKADTVLIVDVT VRYEMDGETLKRAASEKVKHYPVGQQTDKVGGRCFKVMGFPVGARGK
WPASNNTVLAELGVPAGRMRTFARLVSRRTLLYSLDILRDFMREPAGRGTRVALIPAATGAAN**

17. R2A ZoAl N-DYK 质粒 ORF 序列:

[1064]

ATGGATTACAAGGATGACGACGATAAGGGTACCATGAACATCGTGAAGGTGACCGTGCCCGACAAGA
ATCCCCCTTGCTCTTGCTGTGGCGTGCGCCTGAACAGCGTGCTGGCACTGATCGAGCACCTGAAGGG
CTCCACCGGAAGGAGAAGGGTGTGCTTCCGCTGTGCCAAGTGCGGCCGGGAGAACTTTAATCACCAC
AGCACCGTGCTGCTACTACGCCAAGTGCAAGGGCCACAGATCGAGAGACCAACCGTGCGGCGAGTGG
ATCTGTGAGGTGTGCGGCAGGGATTTCACCACAAAGATCGGCCTGGGCCAGCACAAAGCGCCACATGC
ACGCCATGGTGCGCAACCAGGAGCGGATCGACGCCAGCCAGCCAAAGGAGACCTCCAATCGCGGCG
CCCACAAGCGGTGCTGGACAAAGGAGGAGGAGGAGCTGCTGATGAAGCTGGAGGTGCAGTTTGAGA
ACCACAAGAACATCAATAAGCTGATCGCCGAGCAGCTGACCACAAAGACCGCCAAGCAGATCTCCGA
TAAGCGCCGGATGCTGCTGAAGAAGGGCAGAGGCACCACAGGCAACCTGGAGACAGAGCCTGGCAT
GAGCCACCAGTCCCAGGCCAAGGTGAAGGACAATGGA CTGGGAGGCGATCACCTGCCAGGAGGACC
TGTGGTGGACAAGGGCACCATCGGCAAGCCAGGCCAGCACCTGGACACCGATAATTCTCACCAGATC
ACAGCAGGCAAGAAGAAGGGAGGAGGACTGCAGGCCAGATACAGAAGGCGCATCATGAAGAGGCT
GGCCGCCGGCACCATCAACATCTTCCCAAAGGTGTTAAGGAGCTGATCAACGACCAGGAGGCCAGG
CCCCTGATCAATCAGACCACAGAGGACTGTTTCGGCCTGCTGGATTCCGCCTGCCAGATCAGAACCGC
CCTGAGGGAGAAGGGCAAGTCTCAGGAGGAGCGGCCAGAAAGCAGTACCAGAAGTGGATGAAGA
AGCGCGCCATCAAGCGGGGCGACTATCTGCGGTTCCAGAGACTGTTTCACCTGGATAGAGGCAAGCT
GGCCAGGATCATCCTGGACAACACCGAGTCCCTGTCTTGTGATATCAGCCCTCCGAGATCTACAGCG
TGTTCAAGGCCCGCTGGGAGACACCAGGCCACTTCAATGGCCTGGGCGATTTTGAGATCAAGGGCAA
GGCCAACAATAAGGCCTTCCGGGACTTTATCACCGCCAAGGAGATCGAGAAGAACGTGAGAGAGATG
TCTAAGGGCAGCGCACCGAGCCTGATGGAATCGCCCTGGGCGACATCAAGAAGATGGACCCCGGCT

[1065]

ATTCTCGGACCGCCGAGCTGTTAATCTGTGGCTGACAGCCGGCGACATCCCAGATATGGTGAGGGGA
TGCAGGACCGTGCTGATCCCAAAGAGCACCACACCCGAGCGCCTGAAGGACATCAACAATTGGCGG
CCCATCACAATCGGCTCCATCCTGCTGAGACTGTTCTCTAGGATCATCACCGCCAGAATGACAAAGGC
ATGTCCACTGAACCCAAGGCAGAGGGGCTTTATCTCCGACCAGGCTGCTCTGAGAACCTGAAGCTG
CTGCAGAGCATCATCAGGACCGCCAAGAATGAGCACAAGCCCCTGGGCGTGATCTTCGTGGACATCG
CCAAGGCCTTTGATACAGTGTCCCACCAGCACATCATCCACGTGCTGCAGCAGCGGAGAGTGGAACCC
CCACATCGTGGGCCTGGTGAACAATATGTACAAGGACATCTCCACCTATGTGACCACAAAGAAGAAC
ACCCACACAGACAAGATCCAGATCAGAGTGGGCGTGAAGCAGGGCGATCCTCTGTCTCCACTGCTGT
TCAATCTGGCCATGGACCCACTGCTGTGTAAGCTGGAGGAGAGCGGCAAGGGCTTCCACAGGGGCCA
GAGCTCCATCACCGCCATGGCCTTTGCCGACGATCTGGTGCTGCTGTCTGATAGCTGGGAGAACATGA
AGGAGAATATCAAGATCCTGGAGACATTCTGTAATCTGACCGGCCTGAAGACACAGGGCCAGAAGTG
CCACGGCTTTTACATCAAGCCTACCAAGGACTCCTATACAATCAACAATTGCCCAGCCTGGACCATCA
ACGGCACACCTCTGAACATGATCAATCCAGGCGAGTCTGAGAAGTACCTGGGACTGCAGATCGACCC
ATGGACCGGCGTGGCCAAGTATGATCTGAGCACAAGCTGAAGATCTGGCTGGAGTCCATCGACAGG
GCCCCCTGAAGCCTCTGCAGAAGCTGGATATCCTGAAGACCTACACAATCCCCGCTGACCTATCT
GGCCGATCACAGCGAGATGAAGGCAGGAGCACTGGAGGCACTGGACCAGCAGATCAGGACCGCCGT
GAAGGATTGGCTGCACCTGCCTTCCTGCACATGTGACGCCATCCTGTACGTGTCTACCAGGGATGGAG
GACTGGGAGTGACAAAGCTGGCAGGACTGATCCCATCTGTGCAGGCAAGGAGGCTGCACAGGATCG
CACAGAGCCCTGACGAGACCATGAAGGATTTCTTGAGAAGGCCAGATGGAGAAGATGTACGAGA
AGCTGTGGGTGCAGGCCGGCGGCAAGAAGAAGGGCATGCCATCCATCTGGGAGGCCCTGCCCATGAC
AGTGCCTCCAACCAACACAGGCAATCTGTCTGAGTGGGAGGCCCAAATCCCAAGAGCAAGTATCCT
AAGCCATGCGACTGGCGGAGAAAGGAGCTGAAGAAGTGGACCAAGCTGGAGAGCCAGGGCAGAGG
CGTGAAGAACTTTAGGAATGACACAATCTCCAACGATTGGATTCACTACTATAGGCGCATCCCACACC
GCAAGCTGCTGACCGCCATCCAGCTGCGCGCCAACGTGTACCCAACAAGGGAGTTCTGGCACGGG
GCAGAGGCGATAATTATGTGAAGTTTTGTGCGCACTGCGAGGCCGACCTGGAGACCTGTGGCCACAT
CATCGGCTTCTGCCCCGTGACAAAGGACGCCCGGATCAAGAGACACAACAGGATCTGTGATAGACTG
TGCGAGGAGGCAGCAAAGAGGGAGTGGGTGGTGTTTAAGGAGCCCCACCTGAGGGACGCCACCACA
GAGCTGTTCAAGCCTGATGTGATCTTTGTGAAGGAGGACAGGGCACTGGTGGTGGATGTGACCGTGC
GGTACGAGTCCGCCAAGACCACACTGGAGGCTGCCGCCATGGAGAAGGTGGATAAGTATAAGCACCT
GGAGGCCGAGGTGAAGGAGCTGACAAATGCCAAGGACGTGGTGTTCATGGGATTTCCTACTGGGAGC
AAGGGGCAAGTTCTACAAGGGCAACTTTAATCTGCTGGAGACCCTGGGCCTGCCTAAGACAAGACAG
CTGTCTGTGGCCAAGACCCTGAGCACCTACGCCCTGATGTCTAGCGTGGACATCGTGACATGTTTCG

CTCCAGGTCTCGCAAGCCTAACGTGTAG

18. R2A ZoAl N-DYK 氨基酸序列:

MDYKDDDDKGTMNIVKVTVPDKNPPCCGVRLNSVLALIEHLKGSHGRRVCFRCAKCGRENFNHHSTVCHYAKCKGPQIERPPVGEWICEVCGRDFTTKIGLGQHKRHMHAMVRNQERIDASQPKETSNRGAHKRCWTKEEEEELLMKLEVQFENHKNINKLIAEQLTTKTAKQISDKRRMLLKKGRGTTGNLETEPGMSHQSQAKVKDNGLGGDHLPGGPVVDKGTIGKPGQHLDTDNSHQITAGKKKGGGLQARYRRRIMKRLAAGTINIFPKVFKELINDQEARPLINQTTEDCFGLLDSACQIRTALREKGKSQEERPRKQYQKWMKKRAIKRGDYLRFRQLRFLHLDRGKLARIILDNTESLSCDISPSEIYSVFKARWETPGHFNGLGDFEIKGKANNKAFRDFITAKEIEKNVREMSKGSAPGPDGIALGDIKKMDPGYSRTAELFNLWLTAGDIPDMVRGCRTVLIPKSTTPERLKDINNRWPITIGSILLRLFSRIITARMTKACPLNPRQRFISAPGCSENKLLQSIIRTAKNEHKPLGVIFVDIAKAFTVSHQHIIHVLQRRRVDPHIVGLVNNMYKDISTYVTTKKNTHTDKIQIRVGVKQGDPLSPLLFNLAMDPLLCKLEESGKGFHRGQSSITAMAFADDLVLLSDSWENMKENIKILETFCNLTGLKTQGQKCHGFYIKPTKDSYTINNCPAWTINGTPLNMINPGESEKYLGLQIDPWTGVAKYDLSTKLKIWLESIDRAPLKPLQKLDILKTYTIPRLTYLADHSEMKAGALEALDQIRTAVKDWLHLPSCDAILYVSTRDGGLGVTKLAGLIPSVQARRLHRIAQSPDETMKDFLEKAQMEKMYEKLWVQAGGKKKGMPSSIWEALPMTVPPTNTGNLSEWEAPNPKSKYPKPCDWRRKELKKWTKLESQGRGVKNFRNDTISNDWIQYYRRIPHRKLLTAIQLRANVYPTREFLAGRGRGDNYVKFCRHCEADLETCGHIIGFCPVTKDARIKRNICDRLCEEAAKREWVVFKEPHLRDATTLEFKPDVIFVKEDRALVVDVTVRYESAKTITLEAAAMEKVDKYKHLEAEVKELTNAKDVFVMGFPLGARGKFYKGNFNLLETGLPKTRQLSVAKTLSTYALMSSVDIVHMFASRSRKPNV*

[1066]

19. R2A ZoAl N-DYK 转染的 mRNA 序列:

AGGAAUAAACUAGUAUUCUUCUGGUCCCCACAGACUCAGAGAGAACCCGCCACCAUGGAUUACAAGGAUGACGACGAUAAGGGAUCCAACAUCGUGAAGGUGACCGUGCCCGACAAGAAUCCCCCUUGUCCUUGCUGUGGCGUGCGCCUGAACAGCGUGCUGGCACUGAUCGAGCACCUGAAGGGCUCCCACGGAAGGAGAAGGGUGUGCUUCCGUGUGCCAAGUGCGGCCGGGAGAACUUUAAUACACACAGCACCGUGUGUCACUACGCCAAGUGCAAGGGCCACAGAUUCGAGAGACCACCCGUGGGCGAGUGGAUCUGUGAGGUGUGCGGCAGGGAUUUACACACAAAGAUCCGCCUGGGCCAGCACAAAGCGCCACAUGCACGCCAUGGUGCGCAACCAGGAGCGGAUCGACGCCAGCCAGCCCAAGGAGACCUCCAAUCGCGGGCCACACAAGCGGUGCUGGACAAAGGAGGAGGAGGAGCUGCUGAUGAAGCUGGAGGUGCAGUUUGAGAACACACAAGAACAUCAAUAAGCUGAUCGCCGAGCAGCUGACCACAAAGACCGCCAAGCAGAUUCGGAUAAGCGCCGGAUCGUGCUGAAGAAGGGCAGAGGCACCACAGGCAACCUGGAGACAGAGCCU

[1067]

GGCAUGAGCCACCAGUCCAGGCCAAGGUGAAGGACAAUGGACUGGGAGGCGAUCACCUGCCAG
GAGGACCUGUGGUGGACAAGGGCACCAUCGGCAAGCCAGGCCAGCACCUGGACACCGAUAAUUCU
CACCAGAUCACAGCAGGCAAGAAGAAGGGAGGAGGACUGCAGGCCAGAUACAGAAGGCGCAUCA
UGAAGAGGCUGGCCGCCGGCACCAUCAACAUUUCCCAAAGGUGUUUAAGGAGCUGAUCAACGA
CCAGGAGGCCAGGCCCCUGAUCAAUCAGACCACAGAGGACUGUUUCGGCCUGCUGGAUUCGCCU
GCCAGAUCAGAACCGCCCUGAGGGAGAAGGGCAAGUCUCAGGAGGAGCGGCCCAGAAAGCAGUA
CCAGAAGUGGAUGAAGAAGCGCGCCAUAAGCGGGGCGACUAUCUGCGGUUCCAGAGACUGUUU
CACCUGGAUAGAGGCAAGCUGGCCAGGAUCAUCCUGGACAACACCGAGUCCUGUCUUGUGAUA
UCAGCCCUUCCGAGAUCUACAGCGUGUUAAGGCCCGCUGGGAGACACCAGGCCACUCAAUGGC
CUGGGCGAUUUUGAGAUCAAGGGCAAGGCCAACAAUAAGGCCUUCGGGACUUUAUACCGCCA
AGGAGAUUCGAGAAGAACGUGAGAGAGAUGUCUAAGGGCAGCGCACCAGGACCUGAUGGAAUCGC
CCUGGGCGACAUCAAGAAGAUAGGACCCCGGCUAUUCUCGGACCGCCGAGCUGUUUAAUCUGUGGC
UGACAGCCGGCGACAUCCAGAUUAGGUGAGGGGAUGCAGGACCGUGCUGAUCCCAAAGAGCAC
CACACCCGAGCGCCUGAAGGACAUCAACAAUUGGCGGCCCAUCACAAUCGGCUCCAUCCUGCUGA
GACUGUUCUCUAGGAUCAUCACCGCCAGAAUGACAAAGGCAUGUCCACUGAACCCAAGGCAGAG
GGGCUUUUUCUCGACACCAGGCUGCUCUGAGAACCUGAAGCUGCUGCAGAGCAUCAUCAGGACCG
CCAAGAAUGAGCACAAAGCCCCUGGGCGUGAUCUUCGUGGACAUCGCCAAGGCCUUUGAUACAGU
GUCCACCAGCACAUCAUCCACGUGCUGCAGCAGCGGAGAGUGGACCCCCACAUCGUGGGCCUGG
UGAACAAUAUGUACAAGGACAUCUCCACCUAUGUGACCACAAAGAAGAACACCCACACAGACAA
GAUCCAGAUCAGAGUGGGCGUGAAGCAGGGCGAUCCUCUGUCUCCACUGCUGUCAAUCUGGCC
AUGGACCCACUGCUGUGUAAGCUGGAGGAGAGCGGCAAGGGCUUCCACAGGGGCCAGAGCUCCA
UCACCGCCAUGGCCUUUGCCGACGAUCUGGUGCUGCUGUCUGAUAGCUGGGAGAACAUGAAGGA
GAAUAUCAAGAUCUGGAGACAUCUGUAAUCUGACCGGCCUAAAGACACAGGGCCAGAAGUGC
CACGGCUUUUACAUCAAGCCUACCAAGGACUCCUAUACAAUCAACAAUUGCCCAGCCUGGACCAU
CAACGGCACACCUCUGAACAUGAUCAAUCCAGGCGAGUCUGAGAAGUACCUGGGACUGCAGAU
GACCCAUGGACCGGCGUGGCCAAGUAUGAUCUGAGCACAAAGCUGAAGAUCUGGCUGGAGUCCA
UCGACAGGGCCCCCUGAAGCCUCUGCAGAAGCUGGAUAUCCUGAAAACCUACACAAUCCCCCGC
CUGACCUAUCUGGCCGAUCACAGCGAGAUGAAGGCAGGAGCACUGGAGGCACUGGACCAGCAGA
UCAGGACCGCCGUGAAGGAUUGGCUGCACCUGCCUCCUGCACAUGUGACGCCAUCCUGUACGUG
UCUACCAGGGAUGGAGGACUGGGAGUGACAAAGCUGGCAGGACUGAUCCCAUCUGUGCAGGCAA
GGAGGCUGCACAGGAUCGCACAGAGCCCUGACGAGACCAUGAAGGAUUUCCUGGAGAAGGCCCA
GAUGGAGAAGAUGUACGAGAAGCUGUGGGUGCAGGCCGGCGGCAAGAAGAAGGGCAUGCCAUC

[1068]

AUCUGGGAGGCCUGCCCAUGACAGUGCCUCCAACCAACACAGGCAAUCUGUCUGAGUGGGAGGC
CCCCAAUCCCAAGAGCAAGUAUCCUAAGCCAUGCGACUGGCGGAGAAAGGAGCUGAAGAAGUGG
ACCAAGCUGGAGAGCCAGGGCAGAGGCGUGAAGAACUUUAGGAAUGACACAAUCUCCAACGAUU
GGAUUCAGUACUAUAGGCGCAUCCACACCGCAAGCUGCUGACCGCCAUCCAGCUGCGCGCCAAC
GUGUACCCAACAAGGGAGUUCCUGGCACGGGGCAGAGGCGAUAAUUAUGUGAAGUUUUGUCGGC
ACUGCGAGGCCGACCUGGAGACCUGUGGCCACAUCAUCGGCUUCUGCCCCGUGACAAAGGACGCC
CGGAUCAAGAGACACAACAGGAUCUGUGAUAGACUGUGCGAGGAGGCAGCAAAGAGGGGAGUGGG
UGGUGUUUAAGGAGCCCCACCUGAGGGACGCCACCACAGAGCUGUUAAGCCUGAUGUGAUCUU
UGUGAAGGAGGACAGGGCACUGGUGGUGGAUGUGACCGUGCGGUACGAGUCCGCCAAGACCACA
CUGGAGGCUGCCGCAUGGAGAAGGUGGAUAAGUAUAAGCACCUGGAGGCCGAGGUGAAGGAGC
UGACAAAUGCCAAGGACGUGGUGUUAUGGGAUUUCACUGGGAGCAAGGGGCAAGUUCUACAA
GGGCAACUUUAAUCUGCUGGAGACCCUGGGCCUGCCUAAAGACAAGACAGCUGUCUGUGGCCAAG
ACCCUGAGCACCUCAGCCCUGAUGUCUAGCGUGGACAUCGUGCACAUGUUCGCCUCCAGGUCUCG
CAAGCCUAACGUGUGAUGACCUCGAGCUGGUACUGCAUGCACGCAAUGCUAGCUGCCCCUUUCCC
GUCCUGGGUACCCCCGAGUCUCCCCCGACCUCGGGUCCAGGUAUGCUCACCUCUCCACCUGCCCCA
CUCACCACCUCUGCUAGUCCAGACACCUCCCAAGCACGCGAGCAAUGCAGCUAAAACGCUUAGC
CUAGCCACACCCCCACGGGAAACAGCAGUGAUUAACCUUAGCAAUAAACGAAAGUUUAACUAA
GCUAUACUAAACCCAGGGUUGGUCAAUUUCGUGCCAGCCACACCCUGGAGCUAGCAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAGCAUAUGACUAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

20. R2A ZoAl 氨基酸序列:

MNIVKVTVPDKNPPCPCGVRNLSVLALIEHLKGSHGRRRVCFRAKCGRENFNHHSTVCHYAKCKGPQ
IERPPVGEWICEVCGRDFTTKIGLGQHKRHMHAMVRNQRIDASQPKETSNRGAHKRCWTKEEEELMK
LEVQFENHKNINKLIAEQLTTKTAKQISDKRRMLLKKGRGTTGNLETEPGMSHSQAKVKDNLGGDHL
PGGPVVDKGTIGKPGQHLDTDNSHQITAGKKKGGGLQARYRRRIMKRLAAGTNIFPKVFKELINDQEAR
PLINQTTEDCFGLLDACQIRTALREKGSQEERPRKQYQKWMKKRAIKRGDYLRFRQLFHLDRGKLARI
ILDNTESLSCDISPSEIYSVFKARWETPGHFNLGDFEIKGKANNKAFRDFITAKEIEKNVREMSKGSAPGP
DGIALGDIKKMDPGYSRTAELFNLWLTAGDIPDMVRGCRTVLIPKSTTPERLKDINNWRPITIGSILLRFLSR
IITARMTKACPLNPRQGRFISAPGCSENLLQSIIRTAKNEHKPLGVIFVDIAKAFDTVSHQHIIHVLQQR
VDPHIVGLVNNMYKDISTYVTTKKNTHTDKIQIRVGKQGDPLSPLLFNLAMDPLLCKLEESGKGFHRGQ
SSITAMAFADDLVLLSDSWENMKENIKILETFCNLTGLKTQGQKCHGFYIKPTKDSYTINNCPAWTINGTP

LNMINPGESEKYLGLQIDPWTGVAKYDLSTKLKIWLESIDRAPLKPLQKLDILKTYTIPRLTYLADHSEMK
AGALEALDQQIRTAVKDWLHLPSCDAILYVSTRDGGLGVTKLAGLIPSVQARRLHRIAQSPDETMKDF
LEKAQMEKMYEKLWVQAGGKKKGMPSEALPMTVPPTNTGNLSEWEAPNPKSKYPKPCDWRRKELK
KWTKLESQGRGVKNFRNDTISNDWIQYYRRIPHRKLLTAIQLRANVYPTREFLARGRDNVVKFGRHCE
ADLETCGHIIGFCPVTKDARIKRHNRCIDRLCEEAAKREWVVFKEPHLRDATTFLKPDVIFVKEDRALVV
DVTVRYESAKTTLAAAAMEKVDKYKHLEAEVKELTNAKDVVFMGFPLGARGKFYKGNFNLLETGLPK
TRQLSVAKTLSTYALMSSVDIVHMFASRSRKPV**

21. R2 ZoAL 转染的 mRNA 序列:

[1069]

AGGAAUAAACUAGUAUUCUUCUGGUCCCCACAGACUCAGAGAGAACCCGCCACCAUGAACAUUCGU
GAAGGUGACCGUGCCCGACAAGAAUCCCCUUGUCCUUGCUGUGGCGUGCGCCUGAACAGCGUGC
UGGCACUGAUCGAGCACCUGAAGGGCUCCACGGAAGGAGAAGGGUGUGCUUCCGCUGUGCCAA
GUGCGGCCGGGAGAACUUAAUCACCACAGCACCUGUGUCACUACGCCAAGUGCAAGGGCCCCAC
AGAUCGAGAGACCACCCGUGGGCGAGUGGAUCUGUGAGGUGUGCGGCAGGGAUUUCACCACAAA
GAUCGGCCUGGGCCAGCACAAGCGCCACAUGCACGCCAUGGUGCGCAACCAGGAGCGGAUCGACG
CCAGCCAGCCCAAGGAGACCUCAAUCGCGGCGCCACAAGCGGUGCUGGACAAAGGAGGAGGAG
GAGCUGCUGAUGAAGCUGGAGGUGCAGUUUGAGAACCACAAGAACAUAUAAGCUGAUCGCCG
AGCAGCUGACCACAAAGACCGCCAAGCAGAUUCUGGAUAAGCGCCGGAUCGUGCUGAAGAAGGGC
AGAGGCACCACAGGCAACCUGGAGACAGAGCCUGGCAUGAGCCACCAGUCCCAGGCCAAGGUGAA
GGACAAUGGACUGGGAGGCGAUCACCUGCCAGGAGGACCUGUGGUGGACAAGGGCACCAUCGGC
AAGCCAGGCCAGCACCUGGACACCGAUAAUUCUACACAGCAGGCAAGAAGAAGGGAG
GAGGACUGCAGGCCAGAUACAGAAGGCGCAUCAUGAAGAGGCUGGCCGCCGGCACCAUCAACAUC
UCCCCAAAGGUGUUUAAGGAGCUGAUAACGACCAGGAGGCCAGGCCCCUGAUAUAUCAGACCAC
AGAGGACUGUUUCGGCCUGCUGGAUUCGCCUGCCAGAUCAAGCCGCCUGAGGGAGAAGGGC
AAGUCUCAGGAGGAGCGGCCAGAAAGCAGUACCAGAAGUGGAUGAAGAAGCGCGCCAUCAAGC
GGGGCGACUAUCUGCGGUUCCAGAGACUGUUUACCCUGGAUAGAGGCAAGCUGGCCAGGAUCAU
CCUGGACAACACCGAGUCCCUGUCUUGUGAUUACAGCCCUUCCGAGAUUACAGCGUGUUAAGG
CCCGCUGGGAGACACCAGGCCACUCAAUGGCCUGGGCGAUUUUGAGAUCAAGGGCAAGGCCAAC
AAUAAGGCCUUCGGGACUUUAUACCCGCCAAGGAGAUCAAGAAGACGUGAGAGAGAUUGUCUA
AGGGCAGCGCACACAGGACCUGAUGGAAUCGCCCUGGGCGACAUAAGAAGAUGGACCCCGGCUAU
UCUCGGACCGCCGAGCUGUUAAUCUGUGGCUGACAGCCGGCGACAUCCAGAUUAGGUGAGGG
GAUGCAGGACCGUGCUGAUCCCAAAGAGCACACCCGAGCGCCUGAAGGACAUAACAUAUUGG

[1070]

CGGCCCAUCACAAUCGGCUCCAUCCUGGCUGAGACUGUUCUCUAGGAUCAUACCCGCCAGAAUGAC
AAAGGCAUGUCCACUGAACCCAAGGCAGAGGGGCUUUAUCUCCGCACCAGGCUGCUCUGAGAACC
UGAAGCUGCUGCAGAGCAUCAUCAGGACCGCCAAGAAUGAGCACAAGCCCCUGGGCGUGAUCUUC
GUGGACAUCGCCAAGGCCUUUGAUACAGUGUCCACCAGCACAUAUCCACGUGCUGCAGCAGCG
GAGAGUGGACCCCCACAUCGUGGGCCUGGUGAACAAUAUGUACAAGGACAUCUCCACCUAUGUG
ACCACAAAGAAGAACACCCACACAGACAAGAUAUCCAGAUCAAGAGUGGGCGUGAAGCAGGGCGAUC
CUCUGUCUCCACUGCUGUCAAUCUGGGCAUGGACCCACUGCUGUGUAAGCUGGAGGAGAGCGGC
AAGGGCUUCCACAGGGGCCAGAGCUCCAUCACCGCCAUGGCCUUUGCCGACGAUCUGGUGCUGCU
GUCUGAUAGCUGGGAGAACAUAGAAGGAGAAUAUCAAGAUCCUGGAGACAUCUGUAAUCUGACC
GGCCUAAAGACACAGGGGCCAGAAGUGCCACGGCUUUUACAUAAGCCUACCAAGGACUCCUAUAC
AAUCAACAAUUGCCCAGCCUGGACCAUCAACGGCACACCUCUGAACAUGAUCAAUCCAGGGCGAGU
CUGAGAAGUACCUGGGACUGCAGAUCCGACCAUGGACCGGCGUGGCCAAGUAUGAUCUGAGCAC
AAAGCUGAAGAUUCUGGCUGGAGUCCAUCGACAGGGCCCCCUGAAGCCUCUGCAGAAGCUGGAU
AUCCUGAAAACCUACACAAUCCCCCGCCUGACCUAUCUGGCCGAUCACAGCGAGAUGAAGGCAGG
AGCACUGGAGGCACUGGACCAGCAGAUCAAGACCGCCGUGAAGGAUUGGCUGCACCUGCCUUCU
GCACAUGUGACGCCAUCCUGUACGUGUCUACCAGGGAUGGAGGACUGGGAGUGACAAAGCUGGC
AGGACUGAUCCCAUCUGUGCAGGCAAGGAGGCUGCACAGGAUCGCACAGAGCCCUGACGAGACCA
UGAAGGAUUUCCUGGAGAAGGCCCAGAUUGGAGAAGAUACGAGAAGCUGUGGGUGCAGGCCGG
CGGCAAGAAGAAGGGCAUGCCAUCUUGGGAGGCCUGCCCAUGACAGUGCCUCCAACCAACA
CAGGCAAUCUGUCUGAGUGGGAGGCCCCAAAUCCCAAGAGCAAGUAUCCUAAGCCAUGCGACUG
GCGGAGAAAGGAGCUGAAGAAGUGGACCAAGCUGGAGAGCCAGGGCAGAGGCGUGAAGAACUUU
AGGAAUGACACAAUCUCCAACGAUUGGAUUCAGUACUAUAGGCGCAUCCACACCGCAAGCUGCU
GACCGCCAUCCAGCUGCGCGCCAACGUGUACCCAACAAGGGAGUUCUGGCACGGGGCAGAGGCG
AUAAUUAUGUGAAGUUUUGUCGGCACUGCGAGGCGGACCUGGAGACCUGUGGCCACAUAUCGG
CUUCUGCCCCGUGACAAAGGACGCCCCGAUCAAGAGACACAACAGGAUCUGUGAUAGACUGUGC
GAGGAGGCAGCAAAGAGGGAGUGGGUGGUGUUUAAGGAGCCCCACCUGAGGGACGCCACCACAG
AGCUGUUAAGCCUGAUGUGAUCUUUGUGAAGGAGGACAGGGCACUGGUGGUGGAUGUGACCGU
GCGGUACGAGUCCGCCAAGACCACACUGGAGGCUGCCGCCAUGGAGAAGGUGGAUAAGUAUAAG
CACCUGGAGGCCGAGGUGAAGGAGCUGACAAAUGCCAAGGACGUGGUGUUAUGGGAUUCCAC
UGGGAGCAAGGGGCAAGUUCUACAAGGGCAACUUUAAUCUGCUGGAGACCCUGGGCCUGCCUAA
GACAAGACAGCUGUCUGUGGCCAAGACCCUGAGCACCUACGCCUGAUGUCUAGCGUGGACAUCG
UGCACAUGUUCGCCUCCAGGUCUCGCAAGCCUAACGUGUGAUGACCUCGAGCUGGUACUGCAUGC

ACGCAAUGCUAGCUGCCCCUUUCCCGUCCUGGGUACCCCGAGUCUCCCCGACCUCGGGUCCCAG
GUAUGCUCCCACCUCACCUGCCCCACUCACCACCUCUGCUAGUUCACAGACACCUCCCAAGCACGC
AGCAAUGCAGCUAAAAACGCUUAGCCUAGCCACACCCCCACGGGAAACAGCAGUGAUUAACCUUU
AGCAAUAAACGAAAGUUUAACUAAGCUAUACUAACCCCAGGGUUGGUCAAUUUCGUGCCAGCCA
CACCCUGGAGCUAGCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGCAUUGACUAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

22. R2 ZoAl_catdead 氨基酸序列:

[1071] MNIVKVTVPDKNPPCPCGVRNLSVLALIEHLKGSHGRRRVCFRCAKCGRENFNHHSTVCHYAKCKGPQ
IERPPVGEWICEVCGRDFTTKIGLGQHKRHHAMVRNQRIDASQPKETSNRGAHKRCWTKEEEELMK
LEVQFENHKNINKLIAEQLTTKTAKQISDKRRMLLKKGRGTTGNLETEPGMSHQSAKVKNLGGDHL
PGGPVVDKGTIGKPGQHLDTDNSHQITAGKKKGGGLQARYRRRIMKRLAAGTINIFPKVFKELINDQEAR
PLINQTTEDCFGLLDSACQIRTALREKGKSQEERPRKQYQKWMKKRAIKRGDYLRQRLFHLDRGKLARI
ILDNTESLSCDISPSEIYSVFKARWETPGHFNGLDGFEIKGKANNKAFRDFITAKEIEKNVREMSKGSAPGP
DGIALGDIKMDPGYSRTAELFNLWLTAGDIPDMVRGCRTVLIPKSTTPERLKDINNWRPITIGSILLRLFSR
IITARMTKACPLNPRQRFISAPGCSENKLLQSIIRTAKNEHKPLGVIFVDIAKAFDTVSHQHIIHVLQRRR
VDPHIVGLVNNMYKDISTYVTTKKNTHTDKIQIRVGVKQGDPLSPLLFNLAMDPLLCKLEESGKGFHRGQ
SSITAMAFAAALVLLSDSWENMKENIKILETFCNLTGLKTQGQKCHGFYIKPTKDSYTINNCPAWTINGTP
LNMINPGESEKYLGLQIDPWTGVAKYDLSTKLKIWLESIDRAPLKPLQKLDILKTYTIPRLTYLADHSEMK
AGALEALDQQIRTAVKDWLHLPSCDAILYVSTRDGGGLGVTKLAGLIPSVQARRLHRIAQSPDETMKDF
LEKAQMEKMYEKLWVQAGGKKKGMPSEIWEALPMTVPPTNTGNLSEWEAPNPKSKYPKPCDWRRKELK
KWTKLESQGRGVKNFRNDTISNDWIQYYRRIPHRKLLTAIQLRANVYPTREFLARGRDNYVKFCRHCE
ADLETCGHIIGFCPVTKDARIKRHNRICDRLCEEAAKREWVVFKEPHLRDATTLEFKPDVIFVKEDRALVV
DVTVRYESAKTTLAAAMEKVDKYKHLEAEVKELTNAKDVVFMGFPLGARGKFYKGNFNLLETGLPK
TRQLSVAKTLSTYALMSSVDIVHMFASRSRKPVN**

23. R2 ZoAl_catdead 转染的 mRNA 序列:

AGGAAUAAACUAGUAUUCUUCUGGUCCCCACAGACUCAGAGAGAACCCGCCACCAUGAACAUUCGU
GAAGGUGACCGUGCCCCACAAGAAUCCCCCUUGUCCUUGCUGUGGCGUGCGCCUGAACAGCGUGC
UGGCACUGAUCGAGCACCUGAAGGGCUCCCACGGAAGGAGAAGGGUGUGCUUCCGCUGUGCCAA
GUGCGGCCGGGAGAACUUUAAUCACCACAGCACCGUGUGUCACUACGCCAAGUGCAAGGGCCCAC
AGAUCGAGAGACCACCCGUGGGCGAGUGGAUCUGUGAGGUGUGCGGCAGGGAUUUACACACAAA

[1072]

GAUCGGCCUGGGCCAGCACAAAGCGCCACAUGCACGCCAUGGUGCGCAACCAGGAGCGGAUCGACG
CCAGCCAGCCCAAGGAGACCUCCAUAUCGCGGCGCCACAAGCGGUGCUGGACAAAGGAGGAGGAG
GAGCUGCUGAUGAAGCUGGAGGUGCAGUUUGAGAACCACAAGAACAUAUAAGCUGAUCGCCG
AGCAGCUGACCACAAAGACCGCCAAGCAGAUUCUGGAUAAGCGCCGGAUCGUGCUGAAGAAGGGC
AGAGGCACCACAGGCAACCUGGAGACAGAGCCUGGCAUGAGCCACCAGUCCCAGGCCAAGGUGAA
GGACAAUGGACUGGGAGGCGAUCACCUGCCAGGAGGACCUGUGGUGGACAAGGGCACCAUCGGC
AAGCCAGGCCAGCACCUGGACACCGAUAAUUCUACCCAGAUACAGCAGGCAAGAAGAAGGGAG
GAGGACUGCAGGCCAGAUACAGAAGGCGCAUCAUGAAGAGGCUGGCCGCCGGCACCAUCAACAUC
UUCCCAAAGGUGUUUAAGGAGCUGAUAACGACCAGGAGGCCAGGCCCCUGAUAUAUCAGACCAC
AGAGGACUGUUUCGGCCUGCUGGAUUCGCCUGCCAGAUACAACCGCCCUGAGGGAGAAGGGC
AAGUCUCAGGAGGAGCGGCCAGAAAGCAGUACCAGAAGUGGAUGAAGAAGCGCGCCAUCAAGC
GGGGCGACUAUCUGCGGUUCCAGAGACUGUUUACCUGGAUAGAGGCAAGCUGGCCAGGAUCAU
CCUGGACAACACCGAGUCCCUGUCUUGUGAUUAUCAGCCCUUCCGAGAUUAACAGCGUGUUAAGG
CCCGCUGGGAGACACCAGGCCACUUAUGGCCUGGGCGAUUUUGAGAUCAAGGGCAAGGCCAAC
AAUAAGGCCUUCGGGACUUUAUCACCGCCAAGGAGAUCAAGAAGACGUGAGAGAGAUUGUCUA
AGGGCAGCGCACCCAGGACCUGAUGGAUUCGCCCUGGGCGACAUAAGAAGAUGGACCCCCGGCUAU
UCUCGGACCGCCGAGCUGUUUAUCUGUGGCUGACAGCCGGCGACAUCCAGAUUAGGUGAGGG
GAUCGAGGACCGUGCUGAUCCCAAAGAGCACACCCGAGCGCCUGAAGGACAUAACAUAUUGG
CGGCCAUACAAUCGGCUCCAUCUGCUGAGACUGUUCUCUAGGAUCAUACCGCCAGAAUGAC
AAAGGCAUGUCCACUGAACCCAAGGCAGAGGGGCUUAUCUCCGCACCAGGCUGCUCUGAGAACC
UGAAGCUGCUGCAGAGCAUCAUCAGGACCGCCAAGAAUGAGCACAAGCCCCUGGGCGUGAUCUUC
GUGGACAUCGCCAAGGCCUUUGAUACAGUGUCCCACCAGCACAUCAUCCACGUGCUGCAGCAGCG
GAGAGUGGACCCCCACAUCGUGGGCCUGGUGAACAAUAUGUACAAGGACAUCUCCACCUAUGUG
ACCACAAAGAAGAACACCCACACAGACAAGAUAUCCAGAUCAAGUGGGCGUGAAGCAGGGCGAUC
CUCUGUCUCCACUGCUGUCAAUCUGGCCAUGGACCCACUGCUGUGUAAGCUGGAGGAGAGCGGC
AAGGGCUUCCACAGGGGCCAGAGCUCCAUCACCGCCAUGGCCUUUGCCGCCGACUGGUGCUGCU
GUCUGAUAGCUGGGAGAACAUAGAAGGAGAAUAUCAAGAUAUCCUGGAGACAUUCUGUAAUCUGACC
GGCCUAAAGACACAGGGCCAGAAGUGCCACGGCUUUUACAUAAGCCUACCAAGGACUCCUAUAC
AAUCAACAAUUGCCCAGCCUGGACCAUCAACGGCACACCUCUGAACAUGAUAUCCAGGCGAGU
CUGAGAAGUACCUGGGACUGCAGAUCAAGCCAUUGGACCGGCGUGGCCAAGUAUGAUCUGAGCAC
AAAGCUGAAGAUCUGGCUGGAGUCCAUCGACAGGGCCCCCUGAAGCCUCUGCAGAAGCUGGAU
AUCCUGAAAACCUACACAAUCCCCCGCCUGACCUAUCUGGCCGAUCACAGCGAGAUGAAGGCAGG

[1073]

AGCACUGGAGGCACUGGACCAGCAGAUACAGGACCGCCGUGAAGGAUUGGCUGCACCUGCCUUCCU
GCACAUGUGACGCCAUCCUGUACGUGUCUACCAGGGAUGGAGGACUGGGAGUGACAAAGCUGGC
AGGACUGAUCCCAUCUGUGCAGGCAAGGAGGCUGCACAGGAUCGCACAGAGCCCUGACGAGACCA
UGAAGGAUUUCCUGGAGAAGGCCAGAUUGGAGAAGAUGUACGAGAAGCUGUGGGUGCAGGCCGG
CGGCAAGAAGAAGGGCAUGCCAUCUUGGGAGGCCUGCCCAUGACAGUGCCUCCAACCAACA
CAGGCAAUCUGUCUGAGUGGGAGGCCCAAAUCCCAAGAGCAAGUAUCCUAAGCCAUGCGACUG
GCGGAGAAAGGAGCUGAAGAAGUGGACCAAGCUGGAGAGCCAGGGCAGAGGCGUGAAGAACUUU
AGGAAUGACACAAUCUCCAACGAUUGGAUUCAGUACUUAUAGGCGCAUCCACACCGCAAGCUGCU
GACCGCCAUCCAGCUGCGCGCCAACGUGUACCCAACAAGGGAGUUCCUGGCACGGGGCAGAGGCG
AUAUUUAUGUGAAGUUUUGUCGGCACUGCGAGGCCGACCUGGAGACCUGUGGCCACAUCUCCG
CUUCUGCCCCGUGACAAAGGACGCCCCGAUCAAGAGACACAACAGGAUCUGUGAUAGACUGUGC
GAGGAGGCAGCAAAGAGGGAGUGGGUGGUGUUUAAGGAGCCCCACCUGAGGGACGCCACCACAG
AGCUGUUAAGCCUGAUGUGAUCUUUGUGAAGGAGGACAGGGCACUGGUGGUGGAUGUGACCGU
GCGGUACGAGUCCGCCAAGACCACACUGGAGGCUGCCGCCAUGGAGAAGGUGGAUAAGUAUAAG
CACCUGGAGGCCGAGGUGAAGGAGCUGACAAUGCCAAGGACGUGGUGUUAUGGGAUUCCAC
UGGGAGCAAGGGGCAAGUUCUACAAGGGCAACUUUAUCUGCUGGAGACCCUGGGCCUGCCUAA
GACAAGACAGCUGUCUGUGGCCAAGACCCUGAGCACCACGCCCUGAUGUCUAGCGUGGACAUCG
UGCACAUGUUCGCCUCCAGGUCUCGCAAGCCUAACGUGUGAUGACCUCGAGCUGGUACUGCAUGC
ACGCAAUGCUAGCUGCCCCUUUCCCGUCCUGGGUACCCCGAGUCUCCCCCGACCUCGGGUCCCAG
GUAUGCUCUCCACCUCACCUGCCCCACUCACCACCUCUGCUAGUUCAGACACCUCUCCAAAGCACGC
AGCAAUGCAGCUCAAAACGCUUAGCCUAGCCACACCCCCACGGGAAACAGCAGUGAUUAACCUUU
AGCAAUAAACGAAAGUUUAACUAAGCUAUACUAACCCCAGGGUUGGUCAAUUUCGUGCCAGCCA
CACCCUGGAGCUAGCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGCAUAUGACUAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

24. R2A TiGu N-DYK 质粒 ORF 序列:

ATGGATTACAAGGATGACGACGATAAGGGTACCATGGTGGGCGGAAGCCCTCCTGGCGAGGGCGTTA
GAAGCCCTGGAATCGCCTCTCCTAGCGGTACAGCCGTGCAGCACGACTTCGGCTCTCCAATCCTGGTG
CCTGGCGCTGAAGCTGCCGAAGTTTCTACACCTGTGGTGAAAAGTGCCCCAGGACCACCCAGCCTGTC
CATGCTGCGGCACACGGGTGGTGAAGGTGACAGCCCTGTCTGAGCACCTCCGGCGCGCCACGGGA
GAAAGCGGGTGCTGTTCCAGTGCTCTCGGTGCGGCAGAATGAACGAGAAGCACCATTTCTATCGCCTG
CCACTTCCCCAAGTGCCGGGGACCTCCTGTGGAGGAAGGCCCCCTGGGCGCTCCTGAATGGTGCTGT

[1074]

GAAGAGTGC GGACAGAAAGTTCAACACAAAGTCCGGCCTGAGCCAGCATAAGCGCAGCGTGACCCCC
CTGACCAGAAATGTGGAAAGAATTGAAGCCGCCCCGCCCCAAGGGTAAGGGCAAAAGAGGAGCCCCAT
AAGGGCTGTTGGACAGAGGCCGAGGTGGCACAGCTCATCGAACTGGAGGGCAGATTCAAGAACCAG
AGATTCATCAACAAGCTGATTGCGGAGCACCTGCCTTCTAAATCCGCCAAGCAGATCAGCGACAAGA
GACGGCAGCTGGCCGCTGCTACCAAGACCTCTAGCCCCGAGAAGAGGGTGACCAGCAGCACAAAGCG
GAGAGAGCAGTCCTGAGGTGGAGAAAGTGGAGGGAATAAAACGGGAGTACCGTAGACGGGTGGGC
GAGTGGCTGTGCGCCGGTAGTCTGAACGACCAGACCAGTTTTTCAGAAAATCCTGGAAGATGTGGAGA
GCGGCAGCGAGATCGTGACCGGGCCACTGGAGGAACTGGCCAGCTTCGCCAGGGGCAAGCTGGCCG
CCGCTAGAGTGC GGCAACCACCGGAAGCATCCCCCGAGGCCGTGCCTGCGAGAGAGGAGCAGAGAT
GGATGAAGCGCAGAGTCGGACGACGGGGACTGTACCTGAGATTCCAGAGACTGTTCCGCCCTGGACC
GGAGAAAGCTAGCCGGCATCATCTGGACGATGTGGAATCTATCAAGTGCCCTCTGCCTATGGAAGAA
GTGGCCGACGTGTTCCGGAGGAGGTGGGAAGAAGTGGCCCCTTTACCGGGTCAGGCAGCTTCAGG
TCTCTGGGCAAGGCCGACAACGGCGCCTTCAAGCCTATGATCAGCGCCAAGGAAGTTATGAAGAACG
TGCGGGAGATGAGCCGACGGAGCGCTCCTGGACCTGATGGCCTGTCCCTGAGGGATCTCATGAAGAT
CGATCCTCAGGGCTCCAGAATGGCCGAACTGTTCAATCTGTGGCTGCTGGCTGGAAGAGTGCCAGAC
CAGGTGAAGGCCGGCAGAACCGTCCTGATTCTTAAGAGCGCCGATCCTGGGAAGATCGGCAATATCG
ACAACTGGCGGCCTATCACCATCGGAAGCGTGCTGCTTAGAATGTTACAGCAGAATCCTGAGTGCCAG
ACTGCGGAGAGCCTGTCTATCAACAGACGGCAGCGGGGCTTTATCGCCGCCCCCGGCTGCAGCGAG
AACCTGAAGTTACTGCAAGCCCTCATTAAAGAGCGCCAAGCGGGACCACAGAACACTGGGCGTGGTGT
TCGTGGACCTGGCGAAGGCCTTCGACAGCGTGAACCACCAGCACATCTTCCAAGTGCTGGTACAGAA
GGGGGTGGATGGCCACATCATCGACATCCTGAGAGATCTTTACACCAACGCCGGCACCTACCTGGAAT
CCGGCTCTCAGAGAAGCGGCTTTATCAAGATCCTGAGAGGCGTCAAGCAGGGCGACCCACTGAGCCC
AATCCTGTTTAACTGGCCCTGGACCCCTGCTATGTAGACTGGAGGACCGGGGCTGGGATATAAGT
ACGGCGATCAGCAAATCACCAGCCTGGCATTGCGCGATGATCTGGCCCTACTGAGCGACAGCTGGGA
GGGCATGCAGCAAAGCATCCGGGTCGTGCAAGAGTTTTGCCAGAGGACAGGCCTGCGAGTGCAAGC
TCCAAAGTGCCACGGCTTCCTGATCAGACCTACCAAAGAATCTTACACCATTAACTGCGACCCCTT
GGACAATCGCCGACATGCAGCTGGATATGATCGACCCAGGCTCCAGCGAAAAATACCTGGGCCTGGG
CATCGACCCCTTGATCGGCCTGAGCCGTCCTGAGCTGTCCGAAGTGCTGACCAGATGGGTGAAGAAC
ATCGGAGGCGCCCCCTGAAGCCCCTGCAAAAAGTGGACATCCTGAGAAGCTACGCCCTGCCTAGAC
TGCTGTTTATCGCCGACCACGCCGGCTGTCCGCCACCTGCCTGCACTCTCTGGATCTTCCATCAGA
AGCGCTGTGAAGGGCTGGCTGCACCTGCCTCCCAGCACCTGTGATGCCATCATCTACGTGAGCTACAA
GGACGGCGGACTGGGCCTGCCAGACTGGCAAGCCTGATCCCTAACGTGCAGGCAAGACGGCTGGT

GAGAATCGCCCAAAGCGAGGATGACGTGATCCGCAGCGTGGTGCTGCAGGAGGGAATCCAGGAGGA
GATCAGAAAAGTGTGGATCTCTGCCGGCGGCAGACCTGAGAAGGTCCCTAGCGTGACCGGCGAATTC
CCCGTGATGGAAGCCCAGGCCGCCGACGAGGCACTGTCCGAGTGGGAGAGACGGGCCCCCTCGCACC
ATCTACCCCATCCCTTGTAATGGAGAAAGCGCGAGATGGAAAAGTGGACGAATCTCAAAATCTCAGG
GCCACGGCATTAGAAATTTGAGAATGACAGAATCAGCAACGACTGGCTGCTGCACTACGGGAGAAT
CCCTCACAGAAAGCTGATCACAGCTATCCAGCTGAGAGCCAACGTGTATCCTACAAGAGAGTTCCTG
GCTAGAGGCCTTGCGGAGGGCGCTCCTAGAGGCTGCAGACACTGTCCCGCCGAATGGGAATCCTGCA
GCCACATCATCGGCTACTGCCCTGCCGTGCAGGAGGCCAGAATCAAGAGACACAACGACATCTGCGG
CGTGCTTGCTGAAGAGGCCAGAAAGCTGGGATGGGTGATCTTCATCGAGCCTCACTTGAGAGATAATA
CCAACGAGCTGTTCAAACCCGACCTGGTGTGGTGAAGGATCTTGCGCCAAGGTGGTGGACGTGAC
CATCAGATACGAGAGCGGGCTAACCACCCTGTCAGATGCTGCCGCTGAGAAAGCACGAAAATACCAG
CACCTGGCGGGCGAGGTGAGAGCCCTGACTAGCGCCACAACCGTGGACTTTCTGGGCTTCCCCATCG
GTGCCCGGGGCAAATGGTACGTTGGAAACAACGGCCTCCTGAGCGACCTGGGCTTCAGCACCAAGTAG
AGTTGTTTCGGATCGCCCGGGCCCTGAGCAAGAAGGCCCTGCTGAGCTCTGTGGACATAATCCACATCT
TTGCTAGCAGAGCCAGACAGGCCAGACCAGCGAATAG

[1075]

25. R2A TiGu N-DYK 氨基酸序列:

MDYKDDDDDKGTMVGGSPPEGVVRSPGIASPSGTAVQHDFGSPILVPGAEAAEVSTPVVKVPQDHPACPC
GTRVVKVLTALSEHLRRAHGRKRVLFQCSRCGRMNEKHHSIACHFPKCRGPPVEEGPLGAPEWCCEECGQ
KFNTKSGLSQHKRSVHPLTRNVERIEAARPKGKGKRGAAHKGCWTEAEVAQLIELEGRFKNQRFINKLIAE
HLPSKSAKQISDKRRQLAAATKTSSPEKRVTSSTSGESSPEVEKVEGIKREYRRRVGEWLCAGSLNDQTSF
QKILEDVESGSEIVTGPLEELASFARGKLAAARVRHHRKHPAEAVPAREEQRWMKRRVGRRGLYLRFQRL
FALDRRKLAGIILDDVESIKCPLPMEEVADVFRRRWEEVAPFTGSGSFRSLGKADNGAFKPMISAKEVMK
NVREMSRRSAPGPDGLSLRDLMKIDPQGSMAELFNLWLLAGRVPDQVKAGRTVLIPKSADPGKIGNIDN
WRPITIGSVLLRMFSRILSARLRRACPINRRQRGFIAAPGCSENKLLQALIKSAKRDHRTLGVVFVDLAK
AFDSVNHQHIFQVLVQKGVDPGHIDILRDLYTNAGTYLESQSQRSGFIKILRGVKQGDPLSPILFNLALDPL
LCRLEDRLGLGYKYGDQQTSLAFADDLALLSDSWEGMQQSIRVVEEFCQRTGLRVQAPKCHGFLIRPTKE
SYTINDCDPWTIADMQLDMIDPGSSEKYLGLGIDPWIGLSRPELSEVLTRWVKNIGGAPLKPLQKVDILRS
YALPRLLFIADHAGLSATCLHSLDLSIRSAVKGWLHLPSTCDAIIVSYKDGGGLGLPRLASLIPNVQARRL
VRIAQSEDDVIRSVVLQEGIQEEIRKVVISAGGRPEKVPSVTGEFPVMEAQAADSEWERRAPRTTYP
CKWRKREMNWTNLKSQGHGIRNFENDRISNDWLLHYGRIPHRKLITAIQLRANVYPTREFLARGLGEG
APRGCRHCPAEWESCSHIIGYCPAVQEARIKRHNDCGVLAEARKLGWVIFIEPHLRDNTNELFKPDLVL

VKGSCAKVVDVTTRYESGLTTLSDAAAEKARKYQHLAGEVRALTSATTVDLGFPIGARGKWYVGNNGL
LSDLGFSTSRVVRIARALSKKALLSSVDIIHIFASRARQAQTSE*

26. R2A TaGu N-DYK 质粒 ORF 序列:

[1076] ATGGATTACAAGGATGACGACGATAAGGGTACCGAGAAGGTGATGGTCACCGTTCCGGACAAGAACC
CTCCATGTCCATGCTGCGGCACCCGGGTGAACAGCGTGCTGAACCTTATCGAGCACCTGAAGGTCAG
CCACGGAAAGCGGGCGTGTGCTTCCGGTGC GCGAAGTGTGGGAAGGAAAATAGCAACTACCACAG
CGTGGTGTGCCATTTCCCAAAATGTAGAGGACCTGAGACAGAGAAAGCTCCAGCCGGCGAGTGGATT
TGTGAAGTGTGCAACAGAGACTTCACCACCAAAATCGGCCTGGGCCAGCATAAGAGATTAGCTCACC
CCGCTGTTCCGAATCAGGAGAGAATCGTCGCCAGCCAGCCTAAGGAACTAGCAACAGAGGCGCTC
ACAAAAGATGCTGGACCAAAGAGGAGGAGGAACTGCTGATCAGACTGGAAGCCCAGTTTGAGGGCA
ACAAGAACATCAACAAGCTGATCGCAGAACACATCACCACAAAGACCGCCAAGCAGATCTCCGATAA
GAGGAGACTGCTGAGCCGCAAGCCTGCCGAGGAGCCTAGAGAGGAACCTGGCACATGCCACCACAC
CAGACGGGCAGCTGCCAGCCTGAGAACAGAGCCCGAGATGTCCACCACGCTCAGGCCGAAGATCG
GGATAATGGCCCTGGCCGGAGACCCCTGCCTGGTAGAGCCGCCGCCGGCGGCAGAACAAATGGACGA
GATCAGGAGACACCCTGACAAGGGCAATGGACAGCAGCGGCCTACCAAACAGAAGTCTGAGGAGCA
GCTGCAAGCCTACTACAAAAAGACACTGGAGGAGCGGCTGAGCGCCGGAGCCCTGAATACCTTCCCC
CGCGCCTTCAAGCAAGTGATGGAAGGCAGAGATATCAAGCTGGTGATCAACCAGACAGCCCAGGACT
GCTTCGGCTGTCTGGAGAGCATCTCTCAGATCAGAACAGCGACCAGAGATAAGAAAGATACCGTGAC
CAGAGAGAAGCACCCGAAGAAGCCATTCCAAAAGTGGATGAAGGACCGGGCCATAAAGAAGGGCA
ACTACCTGAGGTTCCAGAGATTGTTCTACCTGGACCGAGGCAAACTGGCCAAGATCATCTGGACGAT
ATCGAATGCCTGAGCTGCGACATCCCCCTGTCTGAGATCTACAGCGTGTTTAAGACAAGATGGGAAAC
CACAGGAAGCTTCAAATCTCTGGGCGATTTTAAAACCTACGGCAAGGCCGACAACACTGCCTTCAGA
GAACTGATTACAGCTAAGGAAATCGAGAAAAACGTGCAGGAGATGAGCAAGGGCAGCGCCCCTGGC
CCTGACGGTATCACCTGGGAGATGTGGTGAAAATGGACCCCGAGTTCAGCCGCACAATGGAGATCT
TCAACCTGTGGCTGACCACTGGCAAAATCCCTGACATGGTGCGGGGATGCCGCACCGTGCTGATACCT
AAGAGCAGCAAGCCAGATAGACTGAAGGATATCAACAACTGGCGGCCTATCACAATCGGCAGCATCC
TGCTCCGGCTGTTTAGCAGAAATCGTAACCGCCCGTCTGAGTAAGGCCTGTCTCTGAACCCCAGACAG
CGGGGCTTCATCCGGGGCCGCCGGCTGCAGCGAGAACCTCAAACCTGCTGCAGACCATCATCTGGTCGG
CCAAACGAGAGCACAGACCTCTCGGCGTGGTGTTCTGTGGACATCGCCAAAGCTTTTGATACCGTCTC
CCACCAGCACATCATCCACGCCCTGCAGCAGCGGGAAGTGGACCCTCACATCGTGGGACTGGTGTCC
AACATGTACGAGAACATCAGCACTTATATCACCACCAAGAGAAACACCCACACAGACAAGATCCAGA

[1077]

TCAGAGTGGGCGTTAAGCAGGGCGATCCTATGTCTCCTCTGCTGTTCAACCTGGCTATGGACCCTCTG
CTCTGTAAACTGGAGGAGTCTGGCAAGGGATACACAGAGGCCAAAGCTCCATCACCGCTATGGCCT
TTGCCGACGACCTGGTGCTGCTGTCTGATTCTTGGGAGAACATGAATACCAACATAAGCATCCTGGAA
ACATTCTGCAACCTTACAGGCCTGAAAACCCAGGGACAAAAGTGCCACGGCTTTTACATCAAGCCTA
CCAAAGACAGCTACACCATCAACGACTGCGCCGCCTGGACCATCAACGGCACCCCTTTAAACATGAT
CGACCCTGGAGAAAGCGAGAAGTATCTGGGCCTGCAGTTCGACCCCTGGATCGGCATCGCTAGAAGC
GGCCTGAGCACAAAGCTGGACTTTTGCTGCAGAGAATTGACCAGGCCCTCTGAAGCCTCTGCAGA
AAACCGACATCCTGAAAACCTACACAATCCCTCGGCTGATCTACATTGCCGATCACAGCGAGGTAAAG
ACCGCCCTGCTGGAAACCCTGGACCAAAAGATCAGAACCGCCGTGAAGGAATGGCTCCACCTGCCCC
CCTGCACCTGCGACGCTATCCTGTACAGCAGCACCAGGGACGGCGGCCTGGGCATCACCAGCTGGC
GGGCCTGATCCCCCTCCGTCCAGGCTAGACGGCTGCATAGAATCGCCAGAGCAGCGATGACACAATG
AAGTGCTTTATGGAAAAGGAAAAGATGGAACAGCTGCACAAGAAGCTGTGGATTTCAGGCCGGTGGC
GACAGAGAGAACATTCTAGCATCTGGGAGGCGCCGCTAGTAGCGAGCCTCCCAACAACGTGTCTA
CCAATTCTGAGTGGGAAGCCCCTACACAGAAGGACAAGTTCCCTAAGCCTTGTAATTGGAGAAAGAA
CGAGTTCAAGAAAGTGGACAAAAGCTGGCCTCTCAGGGCCGGGAATTGTGAATTCGAGCGGGACAA
GATCAGCAATCACTGGATTTCAGTACTACAGAAGAATCCACACAGAAAGCTGCTGACGGCCCTTCAG
CTGAGAGCCAACGTGTACCCACGCGGGAGTTCCTGGCCAGAGGTAGACAGGACCAGTACATCAAG
GCCTGCAGACATTGTGATGCTGATATCGAGTCTTGCGCCACATCATCGGCAACTGCCCTGTGACACA
GGACGCGAGAATCAAAAGACACAACCTACATCTGCGAGCTGCTGCTGGAAGAGGCCAAGAAGAAGGA
CTGGGTGGTGTTCAGGAACCCACATCAGAGACAGCAATAAGGAACCTCTATAAACCTGACCTGATC
TTCGTGAAGGACGCCCGGGCCCTGGTCGTGGACGTGACCGTGAGATACGAGGCCGCCAAGTCTAGCC
TGGAGGAGGCTGCCGCCGAGAAAGTGCGGAAGTACAAGCACCTTGAAACAGAAGTGCGGCACCTGA
CCAACGCCAAGGACGTACATTCGTGGGCTTCCCCCTGGGCGCCAGAGGCAAATGGCACCAGGATAA
CTTCAAGCTGCTGACAGAGCTGGGCCTGAGCAAGTCCCGCCAAGTGAAGATGGCCGAGACCTTCAG
CACCGTGGCCCTGTTCTCCTCTGTGGACATCGTGCACATGTTTCGCTAGCAGAGCCAGAAAGAGCATG
GTGATGTAG

27. R2A TaGu N-DYK 氨基酸序列:

MDYKDDDDKGTEKVMVTVDPKNPPCPCGTRVNSVLNIEHLKVSHGKRGVCFRCAKCGKENSNYHS
VVCHFPGKCRGPETEKAPAGEWICEVCNRDFTTKIGLGQHKRLAHPAVRNQERIVASQPKETSNRGAHKRC
WTKEEEEELLIRLEAQFEGNKNINKLIAEHITTKAKQISDKRRLLSRKPAEEPREEPGTCHHTRRAAASLRT
EPEMSHHAQAEDRDNGPGRRLPLGRAAAAGGRMTDEIRHPDKGNGQQRPTKQKSEEQLQAYYKKTLEE

RLSAGALNTFPRAFKQVMEGRDIKLVINQTAQDCFGCLESISQIRTATRDKKDTVREKHPKKPFQKWMK
DRAIKKGNLYLRFQRLFYLDRGKLAKIILDDIECLSCDIPLSEIYSVFKTRWETTGSFKSLGDFKTYGKADNT
AFRELITAKEIEKNVQEMSKGSAPGPDGITLGDVVKMDPEFSRTMEIFNLWLTTGKIPDMVRGCRTVLIPK
SSKPDRLKDINNWRPITIGSILLRLFSRIVTARLSKACPLNPRQGRGFIRAAGCSENKLLQTIWSAKREHRP
LGVVFDIAKAFDTVSHQHIIHALQQREVDPHIVGLVSNMYENISTYITTKRNTHTDKIQIRVGVKQGDPM
SPLLNLAMDPLLCKLEESGKGYHRGQSSITAMAFADDLVLLSDSWENMNTNISILETFCNLTLGLKTQGG
KCHGFYIKPTKDSYNTINDCAAWTINGTPLNMIDPGESEKYLGLQFDPWIGIARSLSTKLDFWLQRIDQAP
LKPLQKTDILKTYTIPRLIYIADHSEVKTALLETLDQKIRTAVKEWLHLPCTCDAILYSSSTRDGGGLGITKLA
GLIPSVQARRLHRIAQSSDDTMKCFMEKEKMEQLHKKLWIQAGGDRENIPSIWEAPPSSEPPNNVSTNSE
WEAPTQKDKFPKPCNWRKNEFKKWTKLASQGRGIVNFERDKISNHWIQQYYRIPHRKLLTALQLRANVY
PTREFLAGRQDQYIKACRHCADIESCAHIIGNCPVTQDARIKRHNYICELLLEEAKKKDWVVFKEPHIR
DSNKELYKPDILFVKDARALVVDVTVRYEAAKSSLEEAEEKVRKYKHLETEVRHLTNAKDVTTFVGFPL
GARGKWHQDNFKLLTELGLSKSRQVKMAETFTVALFSSVDIVHMFASRARKSMVM*

28. R2A TaGu N-DYK 转染的 mRNA 序列:

[1078]

AGGAAUAAACUAGUAUUCUUCUGGUCCCCACAGACUCAGAGAGAACCCGCCACCAUGGAUUACA
AGGAUGACGACGAUAAGGGAUCCGAGAAGGUGAUGGUCACCGUCCGGACAAGAACCCUCCAUG
UCCAUGCUGCGGCACCCGGGUGAACAGCGUGCUGAACCUUAUCGAGCACCUGAAGGUCAGCCACG
GAAAGCGGGGCGUGUGCUUCCGGUGCGCGAAGUGUGGGAAGGAAAAUAGCAACUACCACAGCGU
GGUGUGCCAUUCCCCAAAUGUAGAGGACCUGAGACAGAGAAAGCUCCAGCCGGCGAGUGGAUU
UGUGAAGUGUGCAACAGAGACUUCACCACCAAAUUCGGCCUGGGCCAGCAUAAGAGAUUAGCUC
ACCCCGCUGUUCGGAAUCAGGAGAGAAUCGUCGCCAGCCAGCCUAAGGAAACUAGCAACAGAGGC
GCUCACAAAAGAUGCUGGACCAAAGAGGAGGAGGAACUGCUGAUCAGACUGGAAGCCCAGUUUG
AGGGCAACAAGAACAUCAACAAGCUGAUCGCAGAACACAUCACCACAAAGACCGCCAAGCAGAU
UCCGAUAAGAGGAGACUGCUGAGCCGCAAGCCUGCCGAGGAGCCUAGAGAGGAACCUGGCACAU
GCCACCACACCAGACGGGCAGCUGCCAGCCUGAGAACAGAGCCCCGAGAUGUCCACCACGCUCAG
GCCGAAGAUCCGGGAUAAUGGCCCUGGCCGGAGACCCUGCCUGGUAGAGCCGCCGCCGGCGGCAG
AACAAUGGACGAGAUCAAGGAGACACCCUGACAAGGGCAAUGGACAGCAGCGGCCUACCAAACAG
AAGUCUGAGGAGCAGCUGCAAGCCUACUACAAAAGACACUGGAGGAGCGGCUGAGCGCCGGAG
CCCUGAAUACCUUCCCCCGGCCUUAAGCAAGUGAUGGAAGGCAGAGAUUAAGCUGGUGAUC
AACCAGACAGCCCAGGACUGCUUCGGCUGUCUGGAGAGCAUCUCUCAGAUCAAGACAGCGACCAG
AGAUAAGAAAGAUACCGUGACCAGAGAGAAGCACCCGAAGAAGCCAUCCAAAAGUGGAUGAAG

[1079]

GACCGGGCCAUAAGAAGGGCAACUACCUGAGGUUCCAGAGAUUGUUCUACCUGGACCGAGGCA
AACUGGCCAAGAUAUCCUGGACGAUAUCGAAUGCCUGAGCUGCGACAUCCCCUGUCUGAGAUC
UACAGCGUGUUUAAAGACAAGAUGGGAAACCACAGGAAGCUUCAAUUCUCUGGGCGAUUUUAAAA
CCUACGGCAAGGCCGACAACACUGCCUUCAGAGAACUGAUUACAGCUAAGGAAAUCGAGAAAAA
CGUGCAGGAGAUGAGCAAGGGCAGCGCCCCUGGCCUGACGGUAUACCCUGGGAGAUGUGGUG
AAAAUGGACCCCGAGUUCAGCCGCACAAUGGAGAUCUUAACCUGUGGCUGACCACUGGCAAAA
UCCCUGACAUGGUGCGGGGAUGCCGCACCGUGCUGAUACCUAAGAGCAGCAAGCCAGAUAGACU
GAAGGAUAUCAACAACUGGCGGCCUAUCACAAUCGGCAGCAUCCUGCUCCGGCUGUUUAGCAGA
AUCGUAACCGCCCGUCUGAGUAAGGCCUGUCCUCUGAACCCAGACAGCGGGGCUUCAUCCGGGC
CGCCGGCUGCAGCGAGAACCUCAAACUGCUGCAGACCAUCAUCUGGUCGGCCAAACGAGAGCACA
GACCUCUCGGCGUGGUGUUCGUGGACAUCGCCAAAGCUUUUGAUACCGUCUCCACCAGCACAUC
AUCCACGCCCGCAGCAGCGGGAAGUGGACCCUCACAUCGUGGGACUGGUGUCCAACAUGUACGA
GAACAUCAGCACUUUAUACACCACCAAGAGAAACACCCACACAGACAAGAUAUCCAGAUAGAGUGG
GCGUUAAGCAGGGCGAUCCUAUGUCUCCUCUGCUGUUAACCUGGCUAUGGACCCUCUGCUCUGU
AAACUGGAGGAGUCUGGCAAGGGAUACCACAGAGGCCAAAGCUCCAUCACCGCUAUGGCCUUUG
CCGACGACCUGGUGCUGCUGUCUGAUUCUUGGAGAAACUAUGAAUACCAACAUAAAGCAUCCUGGA
AACAUUCUGCAACCUUACAGGCCUGAAAACCCAGGGACAAAAGUGCCACGGCUUUUACAUAAGC
CUACCAAAGACAGCUACACCAUCAACGACUGCGCCGCCUGGACCAUAACGGCACCCCUUUAAAC
AUGAUCGACCCUGGAGAAAGCGAGAAGUAUCUGGGCCUGCAGUUCGACCCUGGAUCGGCAUCG
CUAGAAGCGGCCUGAGCACAAGCUGGACUUUUGGCUGCAGAGAAUUGACCAGGCCCCUCUGAA
GCCUCUGCAGAAAACCGACAUCCUGAAAACCUACACAAUCCUCGGCUGAUCUACAUUGCCGAUC
ACAGCGAGGUAAAAGACCGCCUGCUGGAAACCCUGGACCAAAGAUCAGAACCGCCGUGAAGGA
AUGGCUCCACCUGCCCCCUGCACCUGCGACGCUAUCCUGUACAGCAGCACCAGGGACGGCGGCC
UGGGCAUACCAAGCUGGCGGGCCUGAUCCCCUCCGUCCAGGCUAGACGGCUGCAUAGAAUCGCC
CAGAGCAGCGAUGACACAAUGAAGUGCUUUUUGGAAAAGGAAAAGAUGGAACAGCUGCACAAGA
AGCUGUGGAUUCAGGCCGGUGGCGACAGAGAGAACAUCCUAGCAUCUGGGAGGCGCCGCCUAG
UAGCGAGCCUCCCAACAACGUGUCUACCAAUUCUGAGUGGGAAGCCCCUACACAGAAGGACAAGU
UCCCUAAGCCUUGUAAUUGGAGAAAGAACGAGUUAAGAAGUGGACAAAGCUGGCCUCUCAGGG
CCGGGGAUUGUGAAUUUCGAGCGGGACAAGAUCAGCAAUCACUGGAUUCAGUACUACAGAAGA
AUCCACACAGAAAGCUGCUGACGGCCCUUACAGCUGAGAGCCAACGUGUACCCACGCGGGAGUU
CCUGGCCAGAGGUAGACAGGACCAGUACAUAAGGCCUGCAGACAUUGUGAUGCUGAUUUCGAG
UCUUGCGCCCACAUCAUCGGCAACUGCCCUGUGACACAGGACGCGAGAAUCAAAGACACAACUA

CAUCUGCGAGCUGCUGCUGGAAGAGGCCAAGAAGAAGGACUGGGUGGUGUUAAGGAACCCAC
AUCAGAGACAGCAAUAAGGAACUCUAUAAACCUGACCUGAUCUUCGUGAAGGACGCCCCGGGCC
UGGUCGUGGACGUGACCGUGAGAUACGAGGCCGCCAAGUCUAGCCUGGAGGAGGCUGCCGCCGA
GAAAGUGCGGAAGUACAAGCACCUUGAAACAGAAGUGCGGCACCUGACCAACGCCAAGGACGUC
ACAUUCGUGGGCUUCCCCUGGGCGCCAGAGGCCAAAUGGCACCAGGAUAACUUAAGCUGCUGAC
AGAGCUGGGCCUGAGCAAGUCCCGCCAAGUGAAGAUGGCCGAGACCUUCAGCACCGUGGCCUGU
UCUCCUCUGUGGACAUCGUGCACAUGUUCGCUAGCAGAGCCAGAAAGAGCAUGGUGAUGUAGUG
ACCUCGAGCUGGUACUGCAUGCACGCAAUGCUAGCUGCCCCUUUCCCGUCCUGGGUACCCCGAGU
CUCCCCCGACCUCGGGUCCCAGGU AUGCUCCACCUCACCUGCCCCACUCACCACCUCUGCUAGU
UCCAGACACCUCCCAAGCACGCAGCAAUGCAGCUCAAAACGCUUAGCCUAGCCACACCCCCACGG
GAAACAGCAGUGAUUAACCUUUAGCAAUAAACGAAAGUUUAACUAAGCUAUACUAACCCCAGGG
UUGGUCAAUUUCGUGCCAGCCACACCCUGGAGCUAGCAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAGCAUAUGACUAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAA

29. R2A TaGu 氨基酸序列:

[1080]

MVTVPDKNPPCPCGTRVNSVLNLIHLKVSHGKRGVCFRAKCGKENSNYHSVCHFPKCRGPETEK
PAGEWICEVCNRDFTTKIGLGQHKRLAHPAVRNQERIVASQPKETSNRGAHKRCWTKEEEELLIRLEAQF
GNKNINKLIAEHITTKTAKQISDKRRLSRKPAEPPREPGTCHHTRRAAASLRTEPEMSHHAQAEDRDNG
PGRRLPLGRAAAGGRMTDEIRRHDPDKGNGQQRPTKQKSEEQLQAYYKKTLEERLSAGALNTFPRAFKQV
MEGRDIKLVINQTAQDCFGCLESISQIRTATRDKKDVTREKHPKKPFQKWMKDRAIKKGNLYRFQRLFY
LDRGKLAKIILDDIECLSCDIPLSEIYSVFKTRWETTSFKSLGDFKTYGKADNTAFRELITAKEIEKNVQE
MSKGSAPGPDGITLGDVVKMDPEFSRTMEIFNLWLTGKIPDMVRGCRTVLIPKSSKPDRLKDINNWRPIT
IGSILLRLFSRIVTARLSKACPLNPRQGRGFIARAAGCSENKLLQTIIWSAKREHRPLGVVFVDIAKAFDTVSH
QHIIHALQQREVDPHIVGLVSNMYENISTYITTKRNTHTDKIQRVGVKQGDPMSPLLFNAMDPLLCKLE
ESGKGYHRGQSSITAMAFADDLVLLSDSWENMNTNISILETFCNLTGLKTQGQKCHGFYIKPTKDSYTND
CAAWTINGTPLNMIDPGESEKYLGLQFDPWIGIARSLSTKLDFWLQRIDQAPLKPLQKTDILKTYTIPRLI
YIADHSEVKTALLETLDQKIRTAVKEWLHLPCTCDAILYSSTRDGGLGITKLAGLIPSVQARRLHRIAQSS
DDTMKCFMEKEKMEQLHKKLWIQAGGDRENIPSIWEAPPSSEPPNNVSTNSEWEAPTQDKDFPKPCNWR
KNEFKKWTKLASQGRGIVNFERDKISNHWIQYYRRIPHRKLLTALQLRANVYPTREFLARGRQDQYIKAC
RHCDADIESCAHIIGNCPVTQDARIKRHNYICELLLEEAKKDWVVFKEPHIRDSNKELYKPDIFVKDAR
ALVVDVTVRYEAAKSSLEEAEEKVRKYKHLETEVRHLTNAKDVTTFVGFP LGARGKWHQDNFKLLTEL

GLSKSRQVKMAETFFSTVALFSSVDIVHMFASRARKSMVM*

30. R2A GeFo N-DYK 质粒 ORF 序列:

[1081] ATGGATTACAAGGATGACGACGATAAGGGTACCATGGGCAACATCGTCATGGTGACCTTACCAAACAA
GAACCTCCATGTCCCTGTTGCCGGGTGCGGCTGCACAGCGTGCTGGCTCTGATCGAGCATCTGAAG
GGCAGCCACGGAAAAAACGGGCATGCTTCCGGTGCGTGAAATGCGGACGTGAGAACTTCAACTAC
CACAGCACCGTGTGCCACATCGCCAAGTGCAAGGGCCCCAAAGTAGAAAAAGCCCCTGTGGGCGAG
TGGATCTGTGAAGTTTGCGGCCGAGATTTACCACCAAGATCGGCCTGGGACAACACAAGAGACTGG
CTCATCTCTTGTGCGGAACCAGGAGCGGATCGACGCTTCTCAGCCTAAGGAAACCAGCAACAGAGG
CGCCATAAGAGATGTTGGACAAAAGAGGAAGAAGAGATGCTGATTAAGCTGGAGGTGCAGTTCGA
AGGCCACAGAAACATCAACAAGCTCATCGCTGAGCACCTGACCACAAAGACCAGCAAGCAGATCTC
TGACAAAAGGAGACTGCTGCCCAGGAAGCAGCTGACAGACCTGAGCAAGGGGGTGGCCGGCCAGA
AGGTGCTGGACCCTGGCCTGAGCCACCAGCCTCAGCTGGGCGTGGTGGACAATGGCCTGGGCGGAG
GACACCTACCTGGCGGGCCGGCCGCGCAAGGCAGAACGATCGAACCTCTTGGCCACCACCTTGATAA
GGACAACGGCCACCGGGAGATCGCCGACCAGCACAAGGCCGGCAGACTGCAGGCCCACTACAGAAA
GAAGATCCGGAAGCGGCTGAGCGAGGGCATGATCTCCAACCTCCCCGAGGTGTTTCGAGCAACTGCTG
GACTGCCAGGAGGCCAGCCTCTGATTAACCAGGCCGCACAGGACTGCTTCGGCTGCCTGGATTCTG
CCAGCCAGATTAGAAAAGCTCTGCGGAAGCAGAACCCAGAAGGATCAGGGCGACCAGCCTAAGC
GGCCCGCCAGAAGTGGATGAAGAAGAGAGCCGTCAAGCGGGGCCACTTCCTGAGATTCCAGAAAC
TGTTCCACCTGGATCGGGGCAAGCTGGCTAAGATCATCTTGGACGACGTGGAATGCCTGTCTTGTGAT
ATCCCTCCAAGTGAGATCTACAGCGTGTTCAGGCCAGATGGGAGACACCCGGACAGTTCGCTGGCC
TGGGCGACTTCGAGATCAACAGAAAGGCAAACAATAAAGCTTTCAGAGATCTGATTACCGCGAAAGA
GATCCTGAAGAACGTGAGAGAAATGACAAAGGGCTCTGCCCCTGGTCCTGATGGCATCGCTCTGGGC
GACATCAGAAAGATGGACCCGAGTACACCAGAACTGCCGAGCTGTTAATCTGTGGCTGACATCTG
GCGAGATCCCCGACATGGTGAGAGGCTGTAGAACAGTGCTGATCCCCAAGAGCAGCAAACTGAGC
GGCTCAAGGACATCAATAATTGGAGACCCATCACCATCGGGAGCATCCTGCTGCGGCTGTTTCAGCAGA
ATCATCACCGCTAGACTGACAAAAGCCTGCCCTCTGAATCCAAGACAGAGAAGCTTTATCAGCGCCG
CCGGATGTAGCGAGAACCTGAAGCTCCTGCAGACCATCATCCGACCGCCAAGAACGAGCACC GGCC
ACTGGGCGTTGTGTTTCGTGGACATCGCCAAGGCCTTCGACACCGTGAGCCACCAGCACATCATCCAC
GTGCTGCAGAGAAGAAGAGTGGACCCACACATCATCGGCCTCGTGAAGAACATGTACAAAGATATCA
GCACCGTCATCACCACCAAGAAGAACACCTATACAGACAAGATTCAGATCCAGGTGGGCGTGAAGCA
AGGCGATCCCCTGAGCCCTCTCCTTTTAACTGGCTATGGACCCCTGCTGTGCAAGCTGGAGGAAC

[1082]

ACGGCAAGGGATTTCACAGAGGCCAGAGCAAGATCACAGCTATGGCCTTTGCCGATGACCTCGTGCT
GCTGTCCGACAGCTGGGAGGACATGAACGCCAATATCAAGATACTGGAAACCTTCTGCGACCTGACA
GGCCTGAAAACTCAGGGACAGAAGTGTACGGTTTCTACATCAAACCTACCAAGGACAGCTACACAG
TGAACAACTGCGCCGCCTGGACGATCAACGGTACACCTCTGAACATGATTAACCTGGAGAGAGCGA
GAAGTACCTGGGCCTGCAGTTCGATCCCTGGGTAGGCATCGCCAAAACAAGCCTGCCTGAGAAGCTG
GACTTCTGGCTGGAACGCATTGATCGGGCCCCCTCTGAAGCCTTTCCAAAAGCTGGACATCCTGAAAA
CCTACACCATCCCTAGACTCACCTATGTGGCCGACCACTCTGAGATGAAGGCCGGCGCCCTGGAAGCC
CTGGATAGAACCATCCGGAGCGCCGTGAAGGACTGGCTGCACCTGCCTAGCAGCACCTGTGACGCGA
TCCTGTACACCAGTATGAAGGACGGTGGACTGGGCGTGACCAAACCTGGTGGGCCTGATCCCATCCGT
GCAGGCCAGACGACTGCACAGAATCGCCAAAGCCCCGAGGAAACCATGAAGGACTTCCTGGAGAA
GGCTCAGATGGAAAAAATGTACGAGAAGCTGTGGGTGCAGGCCGGAGGCAAGAGAAAGAGAATGCC
TTCTATCTGGGAGGCTCTGCCAGAAGTGGTCCCCTCCATCGATACCGCCACAACCTAGCGAGTGGGAGG
CCCCTAACCCCAAGTCCAAGTACCCTAGACCTTGTAAATGGCGCCGAAAGAGTTCAAGAAGTGGAC
CAAGCTGATTGCTCAAGGATGGGGCATCCGGTGCTTTAAAGGGGATAAGATCTCCAATAACTGGATCA
GGCATTATAGATACATTCTCACAGAAAGCTGCTGACAGCCATTCAACTGAGAGCCAGCGTCTACCCC
ACCCGGGAGTTCCTGGCCCGGGGCAGAGAAGATAACTGCGTGAAAAGCTGCCGGCACTGCGAGGCC
GCTGAGGAAAGCTGCGCCACATCATCGGAATGTGCCCCGTGGTGCGGGACGCCAGAATCAAGAGAC
ACAACAGAATCTGCGAGAGGCTCATGGAAGAAGCCGGCAAGCGCGATTGGACCGTGTTTCAGGAGC
CTCATATACGTGACGTGACAAAGGAACTGTACAAGCCCGACCTGATCTTCGTGAAAGAAGGCCTGGC
CCTGGTGGTGGACGTGACCATCCGGTTCGAAAGCACCAAGACCACCCTGGAAGAGGCAGCCGCTGA
GAAGGTGAACAAGTACAAGCACCTGGAACCGGAAGTCAGAAATCTGACAAACGCCAAGGACGTGAT
CTTCATGGGCTTTCTCTGGGGGCCAGAGGCCAATGGTACAACAAGAACTTCGAGCTGCTGGACACC
CTGGGCCTGCCTAGAAGCAGGCAGGACATCATCGCCAAAACCTGTCCACCGATGCTCTGATCTCTTC
CGTGATATTATCCACATGTTCGCCAGCAGAGGCAGACGGCAGCACGCCTGATGA

31. R2A GeFo N-DYK 氨基酸序列:

MDYKDDDDKGTMGNI VMVTL PNKNPPCPCRRVRLHSVLALIEHLK GSHGKKRACFR CVKCGRENFN YH
STVCHIAKCKGPKVEKAPVGEWICEVCGRDFTTKIGLGQHKRLAHPLVRN QERIDASQPKETS NRGAHKR
CWTKEEEEMLIKLE VQFEGHRNINKLIAEHLTTKTSKQISDKRRLLPRKQLTDLSKG VAGQKVLDPGLSHQ
PQLGVVDNGLGGGHLPGGPAAEGRTIEPLGHHLDKDNHREIADQHKAGRLQAHYRKKIRKRLSEGMIS
NFPEVFEQLLDCQEAQPLINQAAQDCFGCLDSASQIRKALRKQNTQKDQGDQPKRPAQKWMKKRAVKR
GHFLRFQKLFHLDRGKLAKIILDDVECLSDIPPSEIYSVFKARWETPGQFAGLGDFEINRKANNKAFRDLI

TAKEILKNVREMTKGSAPGPDGIALGDIRKMDPEYTRTAELFNLWLTSGEIPDMVRGCRTVLIPKSSKPER
LKDINNWRPITIGSILLRLFSRIITARLTKACPLNPRQRSFISAAGCSENKLLQTIIRTAKNEHRPLGVVFVDI
AKAFDTVSHQHIIHVLQRRRVDPHIIGLVKNMYKDISTVITTKKNITYTDKIQIQVGVKQGDPLSPLLFNLA
MDPLLCKLEEKGKGFHRGQSKITAMAFADDLVLLSDSWEDMNANIKILETFCDLTGLKTQGQKCHGFYI
KPTKDSYTVNNCAAWTINGTPLNMINPGESEKYLGLQFDPWVGIAKTSLEKLDWFLERIDRAPLKPFPK
LDILKTYTIPRLTYVADHSEMKAGALEALDRTIRSAVKDWLHLPSSCDAILYTSMKDGGGLGVTKLVGLIP
SVQARRLHRIAQSPEETMKDFLEKAQMEKMYEKLWVQAGGKRKRMPSEIWEALPEVPSIDTATTSEWEA
PNPKSKYPRPCNWRRKEFKKWTCLIAQGWGIRCFKGDKISNNWIRHYRYIPHRKLLTAIQLRASVYPTRE
FLARGREDNCVKSCRHCEAAEESCAHIIGMCPVVRDARIKRHNRICERLMEEAGKRDWTVFQEPHIRDV
TKELYKPDILFVKEGLALVVDVTIRFESTKTTLEEAABEKVNKYKHLETEVRNLTNAKDVIFMGFPLGAR
GQWYNKNFELDLTGLPRSRQDIIAKTLSTDALISSVDIIHMFASRGRRQHA**

32. R2D DroSi SynM C_3F 质粒 ORF 序列:

[1083]

ATGACCACCCGTCCGAGCGTGGATATCTTCCCGGAAGACCAGTACGAGCCGAATGCGGCCGCTACCCT
GAGCCGTGTGCCATGCACCGTGTGCGGTCTAGCTTTAACAGCAAACGCGGTCTGGGCGTGACATG
CGCAGCCGTCACCCGACGAGCTGGACGAGGAACGTCGCCGTGTGGATATCAAGGCGCGTTGGAGC
GAGGAAGAGAAATGGATGATGGCTCGCAAGGAAGTGGAGCTGACCGCCAACGGCCACAAGCACATG
AACAAACAGCTGGCTGTGTACTTCGCCAACCGTAGCGTGGAGGCCATTAAGAACTGCGCCAGCGTG
GTGACTATAAGGAAAAAATTGAGCAGATCCGTGGTCAGAGCGCTCTGGTGCCAGAAGTGGCTAACCT
GACCATTCGTGCTGCTCCGAGCCGTAGCGAGCAGAACACCAGGTGACCACCAGCGAAACCACCCC
GATCACCCCGTTCGAACAGAGCAACCGTGAGATTCTGCGTACCCTGCGTGTTACAGCCAGTGGAG
TGCCACAGCAAATGGCGTGCTCAGGAACTGCAGACCATCATTGACCGCGCCGAAGTGGAGGGCAAG
GAGACCACCCTGCAGTGCCTGAGCCTGTACCTGCTGGGTATTTTCCGGCGCAGGGCGTGCGTCATAC
CCTGACCCGTCACACGTCGTCGCGTAACCGTCGTGAAAGCCGTCGTCAGCAGTATGCTGTGGTG
CAGCGTAACTGGGATAAGCACAAAGGTCGCTGCATCAAAAAGCCTGCTGAACGGCACCGACGAGAGC
GTGATGCCGAGCCAGGAAGTGATGGTGCCGTATTGGCGTGAGGTGATGACCCAGCCAAGCCCAAGCA
GCTGCAGCGGTGAAGTGATTCAGATGGATCACAGCCTGGAGCGTGTGTGGAGCGCCATCACCGAACA
CGACCTGCGTGCTAGCCGCATTAGCCTGAGCAGCAGCCCAGGTCCAGATGGTATCACCCCAAAGAGC
GCCCCGTGAGGTGCCGAGCGGTATTATGCTGCGCATCATGAACCTGATTCTGTGGTGCGGCAACCTGCC
GCACAGCATTCGTCTGGCCCGCACCGTGTTCAATCCGAAGACCGTGACCGCGAAACGTCCGCAGGAC
TTTCGCCCAATCAGCGTGCCGAGCGTGCTGGTGCGTCAGCTGAACGCTATCCTGGCCACCCGCCTGA
ACAGCAGCATTAAGTGGGACCCGCGTCAGCGTGGTTTCCTGCCAACCGATGGCTGCGCCGACAACGC

[1084]

TACCATCGTGGACCTGGTGCTGCGTCACAGCCACAAACACTTCCGCAGCTGCTACATTGCCAACCTGG
ACGTGAGCAAGGCCTTCGACAGCCTGAGCCACGCTAGCATCTACGATACCCTGCGTGCCTATGGTGCG
CCGAAGGGCTTTGTGGACTACGTGCAGAACACCTATGAGGGCGGTGGCACCAGCCTGAACGGCGAC
GGCTGGAGCAGCGAAGAGTTCGTGCCGGCCCGTGGTGTGAAACAGGGCGATCCGCTGAGCCCCGATC
CTGTTCAACCTGGTTATGGACCGTCTGCTGCGCAACCTGCCGAGCGAGATCGGTGCTAAGGTGGGTA
ACGCCATTACCAACGCCGCGGCTTTCGCCGACGATCTGGTGCTGTTTGCGGAAACCCGTATGGGTCTG
CAGGTGCTGCTGGACAAGACCCTGGACTTCCTGAGCCTGGTGGGCCTGAAGCTGAACGCCGACAAAT
GCTTTACCGTGGGTATCAAGGGCCAGCCGAAGCAGAAATGCACCGTGCTGGAGGCGCAGAGCTTCTA
CGTGGGTAGCCGTGAGATCCCGAGCCTGAAACGCACCGATGAATGGAAGTATCTGGGTATTAACCTTA
CCGCTACCGGTGCGGTGCGTTGCAACCCAGCGGAGGACATTGGCCCCGAACTGCAGCGTCTGACCAA
GGCGCCGCTGAAACCGCAGCAGCGTATGTTGCTCTGCGCACCGTGCTGATCCCGCAGCTGTACCAC
AAGCTGGCGCTGGGTAGCGTGGCTATCGGCGTGCTGCGTAAGACCGATAAACTGATTCGCTACTATGT
GCGTCGTTGGCTGAACCTGCCACTGGATGTGCCAATTGCGTTCATTATGCTCCGCCGAAAAGCGGTG
GTCTGGGTATTCCAAGCCTGCGTTGGGTGGCCCCAATGCTGCGTCTGCGTCGCTGAGCAACATCAAA
TGGCCGCACCTGACCCAGAACGAGGTGGCTAGCAGCTTCTGGAAGCGGAGAAACAGCGTGCCCGT
GATCGTCTGCTGGCTGAACAGAACGAGCTGCTGAGCCGTCCGGCGATCGAGAAGTACTGGGCTAACA
AACTGTATCTGAGCGTGGATGGTAGCGGTCTGCGTGAAGCCGGTCACTGGGGTCCACAGCATGGTTG
GGTGAACCAGCCAACCCGTCTGCTGACCGGTAAAGAGTACATTGATGGCATCCGCTGCGCATTAAATG
CTCTGCCAACCAAGAGCCGTACCACCCGTGGTCGTCATGAACTGGAACGCCAGTGCCGTGCTGGTTG
CGATGCCCCGAAACCACCAACCACATCATGCAGAAATGCTATCGTAGCCATGGTCGTCGTGTGGCTC
GTCACAACTGCGTGGTGAACCGTATCAAGCGCGGTCTGGAAGAGCGTGGCTGCGTGGTGATTGTGGA
ACCGAGCCTGCAGTGCGAGAGCGGTCTGAACAAACCGGATCTGGTGGCTCTGCGTCAGGACCACATT
GATGTGATCGACATTCAGATCGTGACCGACGGCCACAGCATGGACGATGCCACCAGCGTAAGATCA
ACCGTTACGATCGCCCGGACATTCGCACCGAACTGCGTCGTCGTTTTGAGGCGGCCGGTGATATCGAA
TTTCACAGCGCGACCCTGAACTGGCGTGGTATCTGGAGCGGCCAGAGCGTGAAGCGCCTGATTGCTA
AAGGCCTGCTGAGCAAGTATGACAGCCACATCATTAGCGTGCAGGTGATGCGTGGTTCCCTGGGCTGT
TTCAAGCAATTCATGTATCTGAGCGGTTTCTCCCGTGA CTGGACCATGGGAGCGACTACAAAGACCA
TGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGACGATGACAAGAAGTAGTAA

33. R2D DroSi SynM C_3F 氨基酸序列:

MTTRPSVDIFPEDQYEPNAAATLSRVPC TVCGRSFNSKRGLGVHMRSRHPDELDEERRRVDIKARWSEEE
KWMMARKEVELTANGHKHMNKQLAVYFANRSVEAIKKLRQRGDYKEKIEQIRGQSALVPEVANLTIRRR

PSRSEQNHQVTTSETTPITPFEQSNREILRTLRGYSPVECHSKWRAQELQTIIDRAELEGKETTLQCLSLYLL
GIFPAQGVHRHTLTRPPRRPRNRRESRRQQYAVVQRNWDKHKGRCIKSLLNGTDESVMPSEQEVMVPYWRE
VMTQSPSSSCSGEVIQMDHSLERVWSAITEHDLRASRISLSSSPGPDGITPKSAREVPSGIMLRIMNLILWC
GNLPHSIRLARTVFIPKTVTAKRPQDFRPISVPSVLVRQLNAILATRLNSSINWDPRQRGFLPTDGCADNATI
VDLVLRHSKHFRSCYIANLDVSKAFDSLHASYDTRLRAYGAPKGFVDYVQNTYEGGGTSLNGDGWSS
EEFVPARGVKQGDPLSPILFNLVMDRLLRNLPSEIGAKVGNAITNAAAFADDLVLFAETRMGLQVLLDKT
LDLFLSLVGLKLNADKCFTVGIGQPKQKCTVLEAQSFYVGSREIPSLKRTDEWKYLGINFATGRVRCNPA
EDIGPKLQRLTKAPLKPQQRMFALRTVLIPQLYHKLALGSVAIGVLRKTDKLIRYYVRRWLNPLDVPPIAFI
HAPPKSGGLGIPSLRWVAPMLRLRRLSNIKWPHLTQNEVASSFLEAEKQRARDRLLAEQNELLSRPAIEKY
WANKLYLSVDGSGLREAGHWGPQHGWVNQPTRLLTGKEYIDGIRLRINALPTKSRTTRGRHELERQCRA
GCDAPETTNHIMQKCYRSHGRRVARHNCVVNRIKRGLEERGCVVIVEPSLQCESGLNKPDVALRQDHID
VIDIQIVTDGHSMDDAHQRKINRYDRPDIRTELRRRFEAAGDIEFHSATLNWRGIWSGQSVKRLLIAKGLLS
KYDSHIISVQVMRGLGCFKQFMYLSGFSRDWTMGSDYKDHDGDYKDHDIDYKDDDDKK**

34. R2D DroSi SynM C_3F 转染的 mRNA 序列:

[1085]

AGGAAUAAACUAGUAUUCUUCUGGUCCCCACAGACUCAGAGAGAACCCGCCACCAUGACCACCCG
UCCGAGCGUGGAUAUCUUCGGAGGACCAGUACGAGCCGAAUGCGGCCGCUACCCUGAGCCGUG
UGCCAUGCACCGUGUGCGGUCGUAGCUUUAACAGCAAACGCGGUCUGGGCGUGCACAUGCAGCAGC
CGUCACCCGGACGAGCUGGACGAGGAACGUCGCCGUGUGGAUAUCAAGGCGCGUUGGAGCGAGG
AAGAGAAAUGGAUGAUGGCUCGCAAGGAAGUGGAGCUGACCGCCAACGGCCACAAGCACAUCAA
CAAACAGCUGGCUGUGUACUUCGCCAACCGUAGCGUGGAGGCCAUUAAGAAACUGCGCCAGCGU
GGUGACUAUAAGGAAAAAAUUGAGCAGAUCCGUGGUCAGAGCGCUCUGGUGCCAGAAGUGGCUA
ACCUGACCAUUCGUCGUCGUCCGAGCCGUAGCGAGCAGAACACCAGGUGACCACCAGCGAAACC
ACCCCGAUCACCCCGUUCGAACAGAGCAACCGUGAGAUUCUGCGUACCCUGCGUGGUACAGCCC
AGUGGAGUGCCACAGCAAAUGGCGUGCUCAGGAACUGCAGACCAUCAUUGACCGCGCCGAACUG
GAGGGCAAGGAGACCACCCUGCAGUGCCUGAGCCUGUACCUGCUGGGUAUUUUUCCGGCGCAGG
GCGUGCGUCAUACCCUGACCCGUCCACCACGUCGUCCGCGUAACCGUCGUGAAAGCCGUCGUCAG
CAGUAUGCUGUGGUGCAGCGUAACUGGGUAAGCACAAAGGUCGUGCAUAAAAGCCUGCUGA
ACGGCACCGACGAGAGCGUGAUGCCGAGCCAGGAAGUGAUGGUGCCGUUUUGGCGUGAGGUGAU
GACCCAGCCAAGCCCAAGCAGCUGCAGCGGUGAAGUGAUUCAGAUGGAUCACAGCCUGGAGCGU
GUGUGGAGCGCCAUCACCGAACACGACCUGCGUGCUAGCCGCAUUAGCCUGAGCAGCAGCCCAGG
UCCAGAUGGUAUACCCCCAAAGAGCGCCCUGAGGUGCCGAGCGGUAUUUAGCUGCGCAUCAUG

[1086]

AACCUGAUUCUGUGGUGCGGCAACCUGCCGCACAGCAUUCGUCUGGCCCGCACCGUGUUCAUUCC
GAAAACCGUGACCGCGAAACGUCCGCAGGACUUUCGCCCAAUCAGCGUGCCGAGCGUGCUGGUGC
GUCAGCUGAACGCUAUCCUGGCCACCCGCCUGAACAGCAGCAUUAACUGGGACCCGCGUCAGCGU
GGUUUCCUGCCAACCGAUUGGUGCGCGGACAACGCUACCAUCGUGGACCUGGUGCUGCGUCACAG
CCACAAACACUCCGCAGCUGCUACAUUGCCAACCUGGACGUGAGCAAGGCCUUCGACAGCCUGA
GCCACGCUAGCAUCUACGAUACCCUGCGUGCCUAUGGUGCGCCGAAGGGCUUUGUGGACUACGUG
CAGAACACCUAUGAGGGCGGUGGCACCAGCCUGAACGGCGACGGCUGGAGCAGCGAAGAGUUCG
UGCCGGCCCCGUGGUGUGAAACAGGGCGAUCCGCUGAGCCCGAUCCUGUUAACCUGGUUAUGGAC
CGUCUGCUGCGCAACCUGCCGAGCGAGAUCCGUGCUAAGGUGGGUAACGCCAUUACCAACGCCGC
GGCUUUCGCCGACGAUCUGGUGCUGUUUGCGGAAACCCGUAUGGGUCUGCAGGUGCUGCUGGAC
AAGACCCUGGACUUCUGAGCCUGGUGGGCCUGAAGCUGAACGCCGACAAUUGCUUUACCGUGG
GUAUCAAGGGCCAGCCGAAGCAGAAAUGCACCGUGCUGGAGGCGCAGAGCUUCUACGUGGGUAG
CCGUGAGAUCCCGAGCCUGAAACGCACCGAUGAAUGGAAGUAUCUGGGUAUUAAACUUUACCGCU
ACCGGUCGCGUGCGUUGCAACCCAGCGGAGGACAUUGGCCCGAAACUGCAGCGUCUGACCAAGGC
GCCGCUGAAACCGCAGCAGCGUAUGUUCGCUUCGCGACCGUGCUGAUCCCGCAGCUGUACCACA
AGCUGGGCGUGGGUAGCGUGGCUAUCGGCGUGCUGCGUAAGACCGAUAAACUGAUUCGCUACUA
UGUGCGUCGUUGGCUGAACCCUGCCACUGGAUGUGCCAAUUGCGUUCAUUCAUGCUCGCCGAAA
AGCGGUGGUCUGGGUAUUCCAAGCCUGCGUUGGGUGGCCCCAAUGCUGCGUCUGCGUCGUCUGA
GCAACAUCAAAUGGCCCGACCCUGACCCAGAACGAGGUGGCUAGCAGCUUUCUGGAAGCGGAGAA
ACAGCGUGCCCGUGAUCGUCUGCUGGCUGAACAGAACGAGCUGCUGAGCCGUCCGGCGAUCGAGA
AGUACUGGGCUAACAACUGUAUCUGAGCGUGGAUGGUAGCGGUCUGCGUGAAGCCGGUCACUG
GGGUCCACAGCAUGGUUGGGUGAACAGCCAACCCGUCUGCUGACCCGUAAAGAGUACAUUGAU
GGCAUCCGUCUGCGCAUUAAGCUCUGCCAACCAAGAGCCGUACCAACCCGUGGUCGUCAUGAACU
GGAACGCCAGUGCCGUGCUGGUUGCGAUGCCCCGAAACCACCAACCAUCAUGCAGAAAUGCU
AUCGUAGCCAUGGUCGUCGUGUGGCUCGUCACAACUGCGUGGUGAACCGUAUCAAGCGCGGUCU
GGAAGAGCGUGGCUGCGUGGUGAUUGUGGAACCGAGCCUGCAGUGCGAGAGCGGUCUGAACAAA
CCGGAUCUGGUGGCUCUGCGUCAGGACCACAUUGAUGUGAUCGACAUCAGAUUGGACCGACG
GCCACAGCAUGGACGAUGCCCACCAGCGUAAGAUAACCGUUACGAUCGCCCCGGACAUCGCACC
GAACUGCGUCGUCGUUUUGAGGCGGCGGUGAUUUCGAAUUUCACAGCGCGACCCUGAACUGGC
GUGGUAUCUGGAGCGGCCAGAGCGUGAAGCGCCUGAUUGCUAAAGGCCUGCUGAGCAAGUAUGA
CAGCCACAUAUUAGCGUGCAGGUGAUGCGUGGUUCCUGGGCUGUUUCAAGCAAUUCAGUAU
CUGAGCGGUUUCUCCCGUGACUGGACCAUGGGGAGCGACUACAAAGACCAUGACGGUGAUUAUA

AAGAUC AUGACAUCGAUUACAAGGAUGACGAUGACAAGAAGUGAUGACUCGAGCUGGUACUGCA
UGCACGCAAUGCUAGCUGCCCCUUUCCCGUCCUGGGUACCCCGAGUCUCCCCGACCUCGGGUCC
CAGGUAUGCUCCCACCUCACCUGCCCCACUCACCACCUCUGCUAGUCCAGACACCUCCCAAGCA
CGCAGCAAUGCAGCUCAAAACGCUUAGCCUAGCCACACCCCCACGGGAAACAGCAGUGAUUAACC
UUUAGCAAUAAACGAAAGUUUAACUAAGCUAUACUAACCCAGGGUUGGUCAAUUUCGUGCCAG
CCACACCCUGGAGCUAGCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGCAUAUGACUAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
A

35. R2D BoMo D69 C_3F 质粒 ORF 序列:

[1087]

ATGAAGAAAAGCAACAAGGAGAACCGTCCGGAAGCGAGCGGTCTGCCGCTGGAGAGCGAACGTAC
CGGCGATAACCCGACCGTGCGTGGTAGCGCGGGTGCGGACCCGGTTGGTCAGGATGCGCCGGGTGG
ACCTGCCAATTCTGCGAGCGTACCTTTAGCACCAACCGTGGTCTGGGCGTGACAAGCGTCGTGCGC
ACCCGGTTGAAACCAACACCGACGCGGCGCCGATGATGGTGAAACGTCGTTGGCACGGCGAGGAAA
TCGATCTGCTGGCGCGTACCGAGGCGCGTCTGCTGGCGGAACGTGGCCAGTGACGCGGTGGCGACCT
GTTCCGGCGCGTCCCGGGTTTTGGTTCGTACCCTGGAGGCGATTAAAGGTCAACGTCGTCTGAACCG
TATCGTGCGCTGGTTTCAAGCGCATCTGGCGCGTTTTGGTAGCCAACCGGGTCCGAGCAGCGGTGGCT
GCAGCGCGGAGCCGATTTTCGTTCGTGCGAGCGGTGCGGAGGAAGCGGGCGAGGAACGTTGCGCGG
AAGATGCGGCGGCGTATGATCCGAGCGCGGTGGGTCAAATGAGCCCGGATGCGGCGCGTGTGCTGAG
CGAACTGCTGGAGGGTGCGGGTCGTTCGTGCGTGCCGTGCGATGCGTCCGAAGACCGCGGGTCGT
CGTAACGACCTGCACGACGATCGTACCGCGAGCGCGCACAAAGACCAGCCGTGAGAAACGTCGTGCG
GAGTACGCGCGTGTGCAAGAACTGTATAAGAAATGCCGTAGCCGTGCGGCGGCGGAAGTGATCGATG
GTGCGTGCGGTGGCGTTGGTCACAGCCTGGAGGAAATGGAAACCTACTGGCGTCCGATTCTGGAACG
TGTGAGCGACGCGCCGGGTCCGACCCCGAGGCGCTGCACGCGCTGGGTCTGCGGAATGGCACGG
TGGCAACCGTGATTATACCCAGCTGTGGAAGCCGATCAGCGTTGAGGAAATTAAGCGAGCCGTTTC
GACTGGCGTACCAGCCCGGGTCCGGATGGTATCCGTAGCGGCCAGTGGCGTGCGGTGCCGGTTCACC
TGAAGGCGGAAATGTTCAACGCGTGGATGGCGCGTGGCGAGATCCCGGAAATTCTGCGTCAATGCCG
TACCGTGTTTGTTCGAAAGTTGAGCGTCCGGGTGGCCCGGTGAATACCGTCCGATCAGCATTGCGA
GCATCCCCTGCGTCACTTCCACAGCATTCTGGCGCGTCTGCTGGCGTGCTGCCCGCCGGACGCG
CGTCAGCGTGGCTTTATCTGCGCGGATGGTACCCTGGAGAACAGCGCGGTGCTGGACGCGGTTCTGG
GTGATAGCCGTAAAGAACTGCGTGAATGCCACGTGGCGGTTCTGGACTTCGCGAAGGCGTTTGATAC
CGTGAGCCACGAGGCGCTGGTTGAACTGCTGCGTCTGCGTGGCATGCCGAGCAGTTCTGCGGTTAC

[1088]

ATTGCGCACCTGTATGACACCGCGAGCACCACCCTGGCGGTGAACAACGAAATGAGCAGCCCGGTGA
AAGTTGGCCGTGGTGTTCGTCAAGGCGACCCGCTGAGCCCGATCCTGTTTAACGTGGTTATGGATCTG
ATTCTGGCGAGCCTGCCGGAGCGTGTGGGTTACCGTCTGGAGATGGAAGTGGTTAGCGCGCTGGCGT
ATGCGGACGATCTGGTGCTGCTGGCGGGCAGCAAGGTTGGTATGCAGGAAAGCATCAGCGCGGTGGA
CTGCGTTGGCCGTCAAATGGGTCTGCGTCTGAACTGCCGTAAAAGCGCGGTGCTGAGCATGATCCCG
GATGGTCACCGTAAGAAACACCACTACCTGACCGAGCGTACCTTCAACATTGGTGGCAAGCCGCTGC
GTCAGGTGAGCTGCGTTGAACGTTGGCGTTATCTGGGCGTGGACTTTGAGGCGAGCGGTTGCGTTAC
CCTGGAACACAGCATCAGCAGCGCGCTGAACAACATTAGCCGTGCGCCGCTGAAACCGCAGCAACG
TCTGGAGATCCTGCGTGCGCACCTGATTCCGCGTTTCCAGCACGGCTTTGTTCTGGGTAACATCAGCG
ACGATCGTCTGCGTATGCTGGATGTGCAGATTTCGTAAGGCGGTTGGTCAATGGCTGCGTCTGCCGGCG
GACGTGCCGAAAGCGTACTATCACGCGGCGGTTCAAGATGGTGGCCTGGCGATCCCGAGCGTGCGTG
CGACCATCCCGGACCTGATTGTTTCGTCGTTTTGGTGGCCTGGATAGCAGCCCGTGGAGCGTGGCGCGT
GCGGCGGCGAAGAGCGACAAAATTCGTAAGAACTGCGTTGGGCGTGGAAGCAGCTGCGTCTGTTTC
AGCCGTGTGGATAGCACCACCAACGTCCGAGCGTTCGTCTGTTTTGGCGTGAGCACCTGCACGCGA
GCGTTGACGGTCGTGAGCTGCGTGAAAGCACCCGTACCCCGACCAGCACCAATGGATTCTGTAACG
TTGCGCGCAGATTACCGGTCGTGATTTCGTGCAATTTGTTACACCCACATCAACGCGCTGCCGAGCC
GTATTCTGTGGCAGCCGTGGCCGTCGTGGTGGCGGTGAGAGCAGCCTGACCTGCCGTGCGGGTTGCAA
AGTGCGTGAAACCACCGCGCACATCCTGCAGCAATGCCACCGTACCCACGGCGGTCTATCCTGCGT
CACAACAAGATTGTGAGCTTCGTTGCGAAGGCGATGGAGGAAAACAAATGGACCGTGGAGCTGGAA
CCGCGTCTGCGTACCAGCGTTGGCCTGCGTAAACCGGACATCATTGCGAGCCGTGATGGCGTGGGTGT
TATCGTGGACGTTCAAGTGGTTAGCGGTCAACGTAGCCTGGATGAGCTGCACCGTGAAAAGCGTAAC
AAATACGGCAACCACGGTGAGCTGGTTGAGCTGGTTGCGGGCCGTCTGGGTCTGCCGAAAGCGGAGT
GCGTGCGTGCGACCAGCTGCACCATTAGCTGGCGTGGCGTTTGGAGCCTGACCAGCTATAAAGAGCT
GCGTAGCATCATTGGTCTGCGTGAACCGACCCTGCAGATCGTGCCGATTCTGGCGCTGCGTGGCAGCC
ACATGAACTGGACCCGTTTAAACCAAATGACCAGCGTGATGGGTGGCGGTGTTGGTGGTGGAGGTAG
CGGGGGCAGTGGAGGGATGGGGAGCGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATC
GATTACAAGGATGACGATGACAAGAAGTAATAA

36. R2D BoMo D69 C_3F 氨基酸序列:

MKKSNNKENRPEASGLPLESERTGDNPTVRGSAGADPVGQDAPGWTCQFCERTFSTNRGLGVHKRRAHP
VETNTDAAPMMVKRRWHGEEIDLLARTEARLLAERGQCSGGDLFGALPGFGRTLEAIKGQRRREPYRAL
VQAHLARFGSQPGPSSGGCSAEPDFRRASGAEEAGEERCAEDAAAYDPSAVGQMSPDAARVLSELLEGA

GRRRACRAMRPKTAGRRNDLHDDRTASAHKTSRQKRRAEYARVQELYKKCRSRAAAEVIDGACGGVGH
SLEEMETYWRPILERVSDAPGPTPEALHALGRAEWHGGNRDYTQLWKPISVEEIKASRFDWRTSPGPDGI
RSGQWRAVPVHLKAEMFNAWMARGEIPEILRQCRTVFVPKVERPGGPGEYRPISIASIPLRHFHSILARRLL
ACCPDARQRGFICADGTLENSAVLDAVLGDSRKKLRECHVAVLDFAKAFDTSHEALVELLRLRGMPEQ
FCGYIAHLYDTASTTLAVNNEMSSPVKVGRGVRQGDPLSPILFNVVMDLILASLPERVGYRLEMELVSALA
YADDLVLLAGSKVGMQESISAVDCVGRQMGLRLNCRKSAVLSMIPDGHRKKHHYLTERTFNIGGKPLRQ
VSCVERWRYLGVDFEASGCVTLEHSSALNNISRAPLKPQQRLEILRAHLIPRFQHGFLGNISDDRLRM
LDVQIRKAVGQWLRLPADVPKAYYHAAVQDGGLAIPSVRATIPDLIVRRFGGLDSSPWSVARAAAKSDKI
RKKLRWAWKQLRRFSRVDSTTQRPSVRLFWREHLHASVDGRELRESTRTPSTKWIRERCAQITGRDFVQ
FVHTHINALPSRIRGSRGRRGGGESSLTCRAGCKVRETTAHILQQCHRTHGGRILRHNKIVSFVAKAMEEN
KWTVELEPRLRTSVGLRKPDIASRDGVGVVDVQVVSQQRSLDELHREKRKNKYGNHGELVELVAGRLGL
PKAECVRATSCTISWRGVWSLTSYKELRSIIGLREPTLQIVPILALRGSHMNWTRFNQMTSVMGGGVGGG
GSGGSGGMGSDYKDHDGDYKDHDIDYKDDDDKK

37. R2D BoMo D69 N-DYK 转染的 mRNA 序列:

[1089]

AGGAAUAAACUAGUAUUCUUCUGGUCCCCACAGACUCAGAGAGAACCCGCCACCAUGGAUUACA
AGGAUGACGACGAUAAGGGAUCCAAGAAAAGCAACAAGGAGAACCGUCCGGAAGCGAGCGGUCU
GCCGUGGAGAGCGAACGUACCGGCGAUAAACCCGACCGUGCGUGGUAGCGCGGGUGCGGACCCGG
UUGGUCAGGAUGCGCCGGGUUGGACCUGCCAAUUCUGCGAGCGUACCUUUAGCACCAACCGUGG
UCUGGGCGUGCACAAGCGUCGUGCGCACCCGGUUGAAACCAACACCGACGCGGCGCCGAUGAUGG
UGAAACGUCGUUGGCACGGCGAGGAAAUCGAUCUGCGUGGCGCGUACCGAGGCGCGUCUGCUGGC
GGAACGUGGCCAGUGCAGCGGUGGCGACCUGUUCGGCGCGCUGCCGGGUUUUGGUCGUACCCUG
GAGGCGAUUAAAGGUCAACGUCGUCGUGAACCGUAUCGUGCGCUGGUUCAGGCGCAUCUGGCGC
GUUUUGGUAGCCAACCGGUCCGAGCAGCGGUGGCUGCAGCGCGGAGCCGAUUUUCGUCGUGC
GAGCGGUGCGGAGGAAGCGGGCGAGGAACGUUGCGCGGAAGAUGCGGCGGCGUAUGAUCCGAGC
GCGGUGGGUCAAAUGAGCCCGGAUGCGGCGCGUGUGCUGAGCGAACUGCUGGAGGGUGCGGGUC
GUCGUCGUGCGUGCCGUGCGAUGCGUCCCAAAACCGCGGGUCGUCGUAACGACCUGCACGACGAU
CGUACCGCGAGCGCGCACAAGACCAGCCGUCAGAAACGUCGUGCGGAGUACGCGCGUGUGCAAGA
ACUGUAUAAGAAAUGCCGUAGCCGUGCGGCGGCGGAAGUGAUCGAUGGUGCGUGCGGUGGGCGUU
GGUCACAGCCUGGAGGAAAUGGAAACCUACUGGCGUCCGAUUCUGGAACGUGUGAGCGACGCGC
CGGGUCCGACCCCGGAGGCGCUGCACGCGCUGGGUCGUGCGGAUUGGCACGGUGGCAACCGUGAU
UAUACCCAGCUGUGGAAGCCGAUCAGCGUUGAGGAAAUAAGCGAGCCGUUUCGACUGGCGUA

[1090]

CCAGCCCGGGUCCGGAUGGUAUCCGUAGCGGCCAGUGGCGUGCGGUGCCGGUUCACCUGAAGGCG
GAAAUGUUAACGCGUGGAUGGCGCGUGGCGAGAUCCCGGAAAUUCUGCGUCAUAGCCGUACCG
UGUUUGUUCGAAAGUUGAGCGUCCGGGUGGCCCCGGGUGAAUACCGUCCGAUCAGCAUUGCGAG
CAUCCCGCUGCGUCACUCCACAGCAUUCUGGCGCGUCGUCUGCUGGCGUGCUGCCCGCCGGACG
CGCGUCAGCGUGGCUUUAUCUGCGCGGAUGGUACCCUGGAGAACAGCGCGGUGCUGGACGCGGU
UCUGGGUGAUAGCCGUAAGAAACUGCGUGAAUGCCACGUGGCGGUUCUGGACUUCGCGAAGGCG
UUUGAUACCGUGAGCCACGAGGCGCUGGUUGAACUGCUGCGUCUGCGUGGCAUGCCGGAGCAGU
UCUGCGGUUACAUUGCGCACCUGUAUGACACCGCGAGCACCAACCCUGGCGGUGAACAACGAAAUG
AGCAGCCCGGUGAAAGUUGGCCGUGGUGUUCGUCAAGGCGACCCGCGUGAGCCCGAUCCUGUUUA
ACGUGGUUAUGGAUCUGAUUCUGGCGAGCCUGCCGGAGCGUGUGGGUUAACGUCUGGAGAUGGA
ACUGGUUAGCGCGCUGGCGUAUGCGGACGAUCUGGUGCUGCUGGCGGGCAGCAAGGUUGGUAUG
CAGGAAAGCAUCAGCGCGGUGGACUGCGUUGGCCGUCAAAUGGGUCUGCGUCUGAACUGCCGUA
AAAGCGCGGUGCUGAGCAUGAUCCCGGAUGGUCACCGUAAGAAACACCACUACCUGACCGAGCGU
ACCUUCAACAUUGGUGGCAAGCCGCGUCGUCAGGUGAGCUGCGUUGAACGUUGGCGUUAUCUGG
GCGUGGACUUUGAGGCGAGCGGUUGCGUUAACCCUGGAACACAGCAUCAGCAGCGCGCUGAACAA
CAUUAAGCCGUGCGCCCGUGAAACCGCAGCAACGUCUGGAGAUCUGCGUGCGCACCUGAUUCCGC
GUUUCAGCACGGCUUUGUUCUGGGUAAACAUACGCGACGAUCGUCUGCGUAUGCUGGAUGUGCA
GAUUCGUAAGGCGGUUGGUCAAUGGCUGCGUCUGCCGGCGGACGUGCCGAAAGCGUACUAUCAC
GCGGCGGUUCAAGAUGGUGGCCUGGCGAUCCCGAGCGUGCGUGCGACCAUCCCGGACCUGAUUGU
UCGUCGUUUUGGUGGCCUGGAUAGCAGCCCGUGGAGCGUGGCGCGUGCGGGCGGAAGAGCGAC
AAAAUUCGUAAGAAACUGCGUUGGGCGUGGAAGCAGCUGCGUCGUUUCAGCCGUGUGGAUAGCA
CCACCAACGUCCGAGCGUUCGUCUGUUUUGGCGUGAGCACCUGCACGCGAGCGUUGACGGUCGU
GAGCUGCGUGAAAGCACCCGUACCCCGACCAGCACCAAAUGGAUUCGUGAACGUUGCGCGCAGAU
UACCGGUCGUGAUUUCGUGCAAUUUGUUCACACCCACAUCAACGCGCUGCCGAGCCGUAUUCGUG
GCAGCCGUGGCCGUCGUGGUGGCGGUGAGAGCAGCCUGACCUGCCGUGCGGGUUGCAAAGUGCG
UGAAACCACCGCGCACAUCCUGCAGCAAUGCCACCGUACCCACGGCGGUCGUAUCCUGCGUCACA
ACAAGAUUGUGAGCUUCGUUGCGAAGGCGAUGGAGGAAAAACAAUGGACCGUGGAGCUGGAACC
GCGUCUGCGUACCAGCGUUGGCCUGCGUAAACCGGACAUCAUUGCGAGCCGUGAUGGCGUGGGU
GUUAUCGUGGACGUUCAGGUGGUUAGCGGUCAACGUAGCCUGGAUGAGCUGCACCGUGAAAAGC
GUAACAAAUACGGCAACCACGGUGAGCUGGUUGAGCUGGUUGCGGGCCGUCUGGGUCUGCCGAA
AGCGGAGUGCGUGCGUGCGACCAGCUGCACCAUAGCUGGCGUGGCGUUUGGAGCCUGACCAGC
UAUAAAGAGCUGCGUAGCAUCAUUGGUCUGCGUGAACCGACCCUGCAGAUUCGUGCCGAUUCUGG

CGCUGCGUGGCAGCCACAUGAACUGGACCCGUUUUAACCAAUGACCAGCGUGAUGGGUGGCGG
UGUUGGUUAGUGACCUCGAGCUGGUACUGCAUGCACGCAAUGCUAGCUGCCCCUUUCCCGUCCUG
GGUACCCCGAGUCUCCCCGACCUCGGGUCCCAGGUAUGCUCCCACCUCACCUGCCCCACUCACC
ACCUCUGCUAGUUCAGACACCUCCCAAGCACGCAGCAAUGCAGCUCAAAACGCUUAGCCUAGCC
ACACCCCCACGGGAAACAGCAGUGAUUAACCUUUAGCAAUAAACGAAAGUUUAACUAAGCUAUA
CUAACCCAGGGUUGGUCAAUUUCGUGCCAGCCACACCCUGGAGCUAGCAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAGCAUAUGACUAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

38. R9 AdVa 质粒 ORF 序列:

[1091]

ATGGATTACAAGGATGACGACGATAAGGGTACCGCATTCAACATTCCCTTACCCTATCGGTCTGTGTTGAT
TGTGAAGTGTGCCATGAGGATATAGGTCAGATGAACCTGCCGATTCTGTGAACACGCAGTGAGCGTGC
ACAACATCAATAAATTCAACTACCTGTGTGACGTGTGCAGCAAAAGCTACGACACCATCAACTCCGTT
AAGGCCATTATGTTGCATGTCGTCGTCAGAAGAATGCTTCTAGTACGACCGCGGTTCCACCAACGT
GATTAATAACAACCAGCTGGCAATTAACACAAACCAAGTGATTAGCCGTAATCCGCTACAGTGCGTGG
AATGCCTGATGAAGCAGGTGGATTTCTACGCCAAGGACACCAAAGCCCTGGTTACCCACATGCGGAC
CAAACATGCCGCCGCCTACGAAGAGTCCAAGAAAGTTGCCACCAGAAGAGTTGCATGGAGCCCGGA
TGAGGACCAGATTCTGGCAGAACTGGAAGTGAAGCTGAAAAAGATCCAGAAAGGACAGCTGCTGAG
CCGTCTGGTGGTTCGAATATAACAAATGCGCCGATAAAAGCAAAGCACCGAGCAGGTGCAAGATGCC
ATTCGCACGCGGAGACAACAGCACGACTACAACTGCTGCTTAGAAGCCTGCAGAGCCAACAGCCTC
CTGTGGGTAGCGAGGATAGTGATTCCGACATCAGCTCCAGCAATAATAATCCGCTTACAACCACCCATA
ATGTTACCCCGACCCCTGACAGCAGCAATGTTGTGCTGCTGATCCAAAAGATTCTGGGAAAAGCGTGGA
CAGCATCGTGAAAATTACAAATCTGAACTGAACACCAATATGCTGAATGCAGCAAGCGCATTTATCA
ACCAGAACATAACATGGACCCGTTGGAGTTAAGCATGAGAGGCATTGAGGAGGACGTTAAAGCCAT
TAGAGATAAAGAACTACAGAAACCGACACGTAATGTCCCTAGTAGCACCACCAGCAGAAAGCCGACC
AGAAATGCCAAGCGGCTGGAAAAAGCAAGAAGTATGGTTATTATCAGCATCTTTATTACAACAACAA
AAAGAACTGGTGGCTGAGATTCTGGATGGCGAGACATCTGGCGCCAAGCCTCCGCCTATGAATCTG
GTTGAGGATTATTATCGTAATATTTGGAGCAGATCTACCATTGATGATTCTCCCGTCAACAATATCAAAA
CTGTTAATTCCGATTCTATCTTTGCCCCAATTTCTCGTGATGAAATCAAAGCTCTGAGCAATACCAA
GAAAGATTCCGCCGAGGACCTGATGCCGTTACCATTAAGAAGCAAAAGCTATAATCGATAATCTGT
ACGTGGCCTATAATATCTGGCTGGGTGTTCAAGGCATTCCGGAACAAGCTGAAGCTGAATAAAACCATC
CTGATTCCGAAAGGCAACAGCGACCTGAGCCTGCTGAAAAATTGGCGTCCGATTACCATCTCTTCCAT

[1092]

TATTCTGCGTGTTTATAATCGTCTGCTGGCCTATAGAATGAATAAAATCTTCAAAACAAATGACAAGCA
AGTTGGTTTCAAGCCTGTTAATGGCTGCGGTATTAACATTAGCTGGCTGCATAGCCTGCTGAAGCATGC
ACGTCTGAATAAAAATAGCATCTACGCTTGTCTGGTTGACGTGTCCAAGGCCTTTGATAGCGTGAGCC
ACCAGAGCATCGTGAGAGCTCTGACCATGAATGGAGCACCGTCTCTGCTGGTTAAACTCATTATGGAT
CAGTACACTAATGTTAATACCGTAATCACATGTAGCGGTAGCATCAGCAACAAGATTAATATTAGCAGC
GGTGTCAAGCAAGGTGATCCTCTGAGTTCTCTGCTGTTCAACTTGGTTATTGATGAACTGTTGACGT
GATCAAGGATCAGTATGGATATACTATCGACAACATTGGGACCACCAATGCCAGATGCTTTGCAGATGA
TCTGACCCTGATCTCTAGCAGCCGTATGGGCATGAACAAACTGCTGGAAGTGAACAACAAAGTTCTTCA
AGGAACGTGGTCTGAATGTTAACCCAGCAAGTGCATGTCCATCGGTATGAGCAAGGGTTACAAGGG
TAAAAAAGCAAAATTGAAAGCGAGCCACTGTTTAGCATTACCGACGCACAGATCCCCATGCTGGGC
TATATTGACAAAACAACACGTTATCTGGGTGTAAACTTCACTTCCATTGGCGCAATCGACGCCAAGAG
GATTAAAAAGGACCTGCAGGATACTCTGGACAAGCTGGAACACCTGAAACTCAAAGCACAGTGTA
AATGGACCTGCTACGTACCTACATGATCCCTCGGTTTATGTTTCAGCTCATAACACCGAGCTGTATCCT
AAGCTGCTGATCAAAATGGATATCCTCATCAGAAAGCTGGCTAAACGTATACTTCATCTGCCTATCAGC
ACCAGCAGCGAATTTTTTATCTGCCGTTTAAAGAAGCGGACTGCAGCTGACAAGCCTGAAAGAAG
CAGTGGGTCTGGCCAAAATTAAACTGCATAAAAAAATTATGTCCTCAAACGACCCGATGCTGTGCTAT
CTGATCGAGAGCCAGCGTAGTCGTATTGTTGAACACTTCATGAAGGATCTGAAGCTGGGCGACAGCCT
GACCCTGAACGAGATGAATAATATTAAGAATGTTTTATGAAAGAGAAGCGAATTAGCTTCGCCCAGA
AAATCCACGGCGTGGGCTTTGAAGTTTTCAGCTCCAGCCCGTTAACCAACCAGTGGATCAACGGTGA
GATCAAAACCATGACCACAAAACCTATATAAACAGCATCAAGTTGCGTACCAACACCCTGGAAACC
CGCGTAACGACCTCACGTGGTCTGAACATCATCAAAACCTGTCGGAGATGTCATGTTGCCGATGAAAG
TCTGATGCACGTGCTGCAGTGTGTAGCTCAACAAAGGGCCTCCGGTACAGCCGTCATCATAAAATTT
GCGCAAAAGTGGCCAACAAACTGGTGATGAACGGCTATGGTGTGTTTCGTGAAAAGAGCTATCCTGA
TCCTAACAATAGCGGTAGCTACCTGCGTCCCGATATCATCGCCGTGAAGAACGGTCATGTGATTGTGCT
GGATGTGACCGTAGTTTACGAGGTGACCGGTGCAACCTTTATCAACGCCTACCAGACCAAAATCAACA
AGTACAACGCAATTATGGTGCAGATCGAACAGATGTTTAATTGTGTGAATGGCGAGCTACATGGCTTA
GTCATTGGCTCCCGAGGCAGCATAACACAGCCAGCTGCACATCTGGCACCAGATGGGTTTTAGCAG
TATTGAACTGAAATATGTGGCAATCGGCTGTATGGAAGATTCTCTGCGCATTATGAGCACCTTCAGCAA
GGCCATCACCTGATAA

39. R9 AdVa 氨基酸序列:

MDYKDDDDKGTA FNIPYPIGRVDCEVCHEDIGQMNLPIREHAVSVHNINKFNLYLCQLCSKSYDTINSVKA

HYVACRRQKNASSTTAVPTNVINNNQLAINTNQVISRNPLQCVECLMKQVDFYAKDTKALVTHMRKHA
AAYEESKKVATRRVAWSPDEDQILAELEVKLKKIQKGQLSRLVVEYNKCADKSKAPSRSKDAIRTRRQQ
HDYKLLLRSLQSQPPVGSESDSDISSNNNPLTTTHNVTPDSSNVLLIQKIRESVDSIVKITNLKLNT
NMLNAASAFINQNNNMDPLELSMRGIEEDVKAIRDKEQLKPTRNVPSSTTSRKPTRNAKRLEKSKKYGY
YQHLYYNNKKLVAEILDGETSGAKPPPMNLVEDYYRNIWSRSTIDDSPVNNIKTVNSDSIFAPISRDEIKL
ALSNTKKDSAAGPDAVTIKEAKAIIDNLYVAYNIWLGVQGIPEQLKLNKTILIPKGNSDLSLLKNWRPITISS
IILRVYNRLLAYRMNKIFKTNDKQVGFKPVNGCGINISWLHSLK HARLNKNSIYACLVDVSKAFDSVSH
QSIVRALTMNGAPSLLVKLIMDQYTNVNTVITCSGSISNKINISSGVKQGDPLSLLFNLVIDELFDVIKDQY
GYTIDNIGTTNARCFADDLTLISSSRMGMNKLLELTTFKFKERGLNVNPSKCMSIGMSKGYKGKSKIESE
PLFSITDAQIPMLGYIDKTTRYLGVNFTSIGAIDAKRIKKDLQDTLDKLEHLKKAQCKMDLLRTYMIPRF
MFQLIHTELYPKLLIKMDILIRKLAKRILHLPISTSSEFFYLPFKEGGLQLTSLKEAVGLAKIKLHKIMSSND
PMLCYLIESQSRSRIVEHFMKDLKLGDSLTLNEMNNIKECFMKEKRISFAQKIHG VGFEVFSSSPLTNQWIN
GEIKTMTTKTYINSIKLRTNTLETRVTTSRGLNIIKTCRRCHVADESLMHVLQCCSSTKGLRYSRHHKICAK
VANKLVMNGYGVFREKSYDPNNSGSYLRPDIIAVKNGHVIVLDVTVVYEVTGATFINAYQTKINKYNAI
MVQIEQMFNCVNGELHGLVIGSRGSIHHSQLHIWHQMGFSSIELKYVAIGCMEDSLRIMSTFSKAIT**

[1093]

40. R8 HyMaA 质粒 ORF 序列:

ATGGATTACAAGGATGACGACGATAAGGGTACCGTGATTATACACCCGGGCACCGATGGTATTTGGTGT
TGTAACCGAATGTGTGGAAATCCATAATAGCGGCAAGGATCTGAAGCGACACCTGGCCAAACGCCATCC
CTCGGTCAACCATCTCCGGCTACAAGTGCAACCTGTGTCCGTTTGTAGCGAGCGGCAACTGAGTGTTG
GTACACATTTGCGTTACTGTCGTGGTGTTAAGGAAGTTGTGAAAAGAGAATTCGCATGTGCATCCTGT
TCCTTTAGCAGCGACACCTTCAGCGGCCTTCAGGTGCACATGCAGCGTAAGCACATCGCAGAGTGGA
ATGATCAGCTGAAGGAAAAGACAGAGTTCGCCTGGACCGACCGGAACTCCGTGAACTGGCTGAAA
AGGAGCTGACCACACCGTCTTCCGGTACAACAAGATCTTCTACGCCGCTCTGGGCACCTCCAGGAC
CTACGATGCAGTGCGGAAAATTCGCTATAATGATCGATACAAGAGCGCCATCGCCGAAATGAGAAGCC
AGATTGCAGATGCAGCAGCAGCCGCCAAGAACGTGATGTGGAGCGTGGTCTCGTGAGCGCCACACA
GCGACCGTGGAAGAGAGATGCTGCCTGTTGTTGAAACCAAAAGCGACATCCAGGTTAATAACGATAT
CAAAAAGACATTGAGCTGACCCCTAATAGCCGCCAGAAACAGACCAATCTGGCCCTGGCCAGACCT
GCCGTGATCGAGGTGGAAGAGGATCTGGGCAGACAGGACGTGAAGCAGTACCTGGCGAGCCTGAGA
CAGGATGATTACACCAGCCCTGCTGAGCGTAGCATTTTTGCATACTGCCGCGAAGAAACAAACTGGA
GCGCCACGAAGCGGCAAGTGCTGAAAATAAGTCGGACCACCAGAGGTCTGCGCCAGCCGAAAAAAG
TGCGTCCATTTGAGTTTCCGGAGGGCTTCAAACCCAACAGAAATATGAGGAAGTGGCGCAAATATCGT

[1094]

TTTCTGCAGGAATGCTACCGTGAAAAACGTGCAGAAACAGTGAGCAAAATCCTGGACGGCACCTTCA
TTGACGAACCCGAGGAAGAAATTCGTCCTGAACTGGAGGAGGTTTCAGAGAATGTACATCGACCGTCT
GGAAAAGCGTACACAGCTGGATACCACAAAATCGTTCAGACAGATGAAGTGTCTGCCTGCAGAGC
TACGGCAGAATCACGATCGGTGAAGTGAGAGATGCTCTGGGCGCAAGCAAAAAGGACAGCGCATCC
GGTCCGGATGGGCTGCTGCTGCAAGACGTGCGGAGGCTGGGACCTCTGCTGCTGTGCAATATCTTCA
ACATGTGGTATCTGCACGGCATCCCGGTGGAAGAGAACAGATGTAGGACAATCCTGCTGTATAAAAGC
GGTGATCGACATCTGGCCAGCAACTACCGTCCGGTTACCATCGGCAACATGCTGAACCGCCTGTATGC
AAAAATATGGGACAAAAGAATTAGAAAAAATGTGCGTCTGCATGTTAGACAGAAAGCCTTCATCCCG
GTTGATGGATGTTTTGAGAATGTAAAGACCAATTCAGTGCGTTCTACAGAGCTATCGTAAGCGCAAATTA
GAACACAACGTGGTTTTTATTGATCTGGCCAAAGCATTTGATACCGTGCTGCATGACAGCATCCGGAA
AGCCCTGTGGCGTAAGGGGGTGCTAGCGGCGTGGTGAAGGTTGTTGATAGCCTGTACGCCGGAGCC
GTAACATCTATCAGCGTTGGCAAGACCAAAAACGAGAAGCATTTCATCAACAGCGGTGTCAAACAGG
GCTGCCCTTTGTACCGCTGCTGTTAATCTGATTCTGGATGAGTTAGCAGAACGGATTGAAGCCACA
GGCTGCGGCCTTGATCTTGACGGCCACGTTCTGTCTAGCATGGCCTTTGCCGACGACTATGTTCTGCTG
GCTAAGGATAGCGTGGAATGAACGAAGTATTGCGGTTTGTCTACCTTCTTCAAAGAGAAGGGCTT
GTCTGTGAATCCTGGCAAATGTCAGAGCCTGCGCGTGCTGCCAGTAAAAGAGAAAAAACGCAGCATG
AAGGTACTGGTGCGTCCTCATCGTTGGTGGCGTATTAAGGACCAGGATGTGGATATTCCCAGCATGAC
CTACGATTCTCTCGGCAAGTATCTGGGGGTAGCATCGACCCACCGGTAAAATCGCCCTGCCTATTGA
GGAATGGAAGAACTGGATGACCAAACTGAAAGAGTGCAAGCTGAAACCGGAACAGAAAGTGAAGA
TTCTGAAAGAAGTGGTGTGTAGCAGAGTCAATTATGTTCTGCGTATGAGCGAATGTGGTATTAGCGAA
CTGCGTAGCTGGACCAGATTTGTTAGAAATTGGGCAAAAACATCATCCATCTGCCTACGTGGTGCAG
CAGCGATTGGATTCAATCCATTAAAGGTCTGGGTATTCTGATGTGTCTAAGGGCATTGTTATCCAACG
GATGAGAGCTAGTGAGAAGATGAGCACCAGCGAAGATGGCATCGTTCGTGTTGTGGGAGCACGCCTG
GTTCAGAAGAATCGTGTTCTGTGGGAAAAAGCCGGTTTTGAAGGCATTGAACTGAAAGCTGCACGTA
GACACTGCGAAGTTGAGAGACTGAACAACATTGGCAATATCACCACGGAGTAGCCCTGAAGACCAT
CGCAGCCGTGTCCTCTGTGAACCGATACTGGATGATTGAGGACAATCTGAAAAGCGGTAATAAAATCC
TGGTGTGGAAGGCAATGGCAGGTGCTATCCCGACCAAAATCAATCTGTACGGGGCGTTGCCGACCA
GACCCTTAAGAAATGTCGACGTTGTGGTCTAACAGCGGAAACCGATGGTCATATTCTGGCTGGCTGCC
ACACCAGCTCTGATGCATATAGCAAACGACATAATATGCTGTGTGATAAACTGGCCAAAGAGCTGAAA
CTGAATGGCGGTCCGAATCGGCGTGTTTGGCGGGAACGGACCTGCTTTACCAGCACAGGTCTGCTGTTA
TCGTCCGGATATTATTGTAAGATGATAGCAAAATCACAGTCATCGACATGACCTGCCCCGTATGAAAA
ATCCGAAGGTCATCTGATCCAGTGTGAAAGCGCGAAAGTGACCAAATGAACCTCTGAAACTGGAT

AAGTATTGGACAAGAGAGCTGGAAGGCGCCAATGGTATTGTGGCAGAAAAAGTTGAGCTGATGGGCC
TGGCCATCGGTGCAATCGGCACCATCATGCGTTCTACCCTGCGCAAAGTGTGTGAAGTGAAGAGCGGT
AGAATTGTGAGACGTCTGCAGATGATTGCATGTAACAACAGCGCCCAGATCATCAAGGGTCATTTGTC
GCGCGCCACACGCCGTAATCTGAGATGATAA

41. R8 HyMaA 氨基酸序列:

MDYKDDDDKGTVIIHPGTDGIWCCTECVEIHNSGKDLKRHLAKRHPSVTISGYKCNLCPFVSEERQLSVGT
HLRYCRGVKEVVKREFACASCSFSSDTFSGLQVHMQRKHIAEWNDQLKEKTEFAWTDRELRELAEKELT
TPSFRYNKIFYAALGTSRTYDAVRKIRYNDRYKSAIAEMRSQIADAAAAAQERDVERGLVSAHSDRGKEM
LPVVETKSDIQVNNDIKKDIELTPNSRQKQTNLALARPVIEVEEDLGRQDVKQYLASLRQDDYTSPAERS
IFAYCREETNWSATKRQVLKISRTRGLRQPKKVRPFEPFEGFKPNRNMWKWKYRFLQECYREKRAETV
SKILDGTFIDEPEEEIRPELEEVQRMVIDRLEKRTQLDTTKIVQTDEVFCLQSYGRITIGEVRDALGASKKD
SASGPDGLLLQDVRRLGPLLLCNIFNMWYLGHPVEENRCRTILLYKSGDRHLASNYRPVTIGNMLNRLY
AKIWDKRIRKNVRLHVRQKAFIPVDGCFENVKTIQCVLQSYRKRLKLEHNVFIDLAKAFDFTVLHDSIRKA
LWRKGVPSGVVKVVDLSYAGAVTSISVGKTKTRSICINSGVKQGCPSPLLFNILDELAERIEATGCGLDL
DGHVLSSMAFADDYVLLAKDSVEMNELIRVCSTFFKEKGLSVNPGKCQSLRVLVPVKEKKRSMKVLVRPH
[1095] RWWRIKDQDVIDPSMTYDSLGLVGSIDPTGKIALPIEWWKNWMTKLKECKLKPEQKVILKEVVCSR
VNYVLRMSECGISELRSWTRFVRNWKNIHLPTWCSSDWIHSIKGLGIPDVSKGIVIQMRASEKMSTSE
DGIVRVVGARLVQKNRVLWEKAGFEGIELKAARRHCEVERLNNIGNITNGVALKTIAAVSSVNRYWMIED
NLKSGNKILVWKAMAGAIPTKINLSRGVADQTLKKCRRCGLTAETDGHILAGCHTSSDAYSKRHNMLCD
KLAKELKLNGGPNNRVWRERTCTSTGRRYRPDIIVKDDSKITVIDMTCPEYKSEGHILQCESAKVTKYEP
LKLDKYWTRELEGANGIVAEEKVELMGLAIGAIGTIMRSTLRKLCCLKSGRIVRRLQMIACNNSAQIIKGHL
SRATRRNLR**

42. R2D Cifn 质粒 ORF 序列

ATGGATTACAAGGATGACGACGATAAGGGTACCAATACCCTGGGAGAAAACGTGGACGCCTCAGCCA
CCACCTCCTCTGCAAACCCTCTGTCTCAGGAACCGGGCTGGTGCGAGAGCTGTAGCAAAGTGTTCAG
GTCCCAGCGTGGTCTGCGGGTACACCAGCGCAGCAAGCACCCCGAACTGTACCACAGTCAGAATCAG
CCGCTGCCACGTTCAAAAGCACGTTGGAGCGACGAAGAAATGGTTATATTGCCCGTGAAGAGATTG
CTAATCGAAAAATCAGATTCATTAATCAGCATCTGCACAAAGTGTTCGCGACCGTACCCTGGAGAGC
ATCAAGGGCCTGAGGGGTAAAAACGTTAGATATGCTCGTATCATGGCCGACCTGGAAGCCGAGATGAC
CAGCCAGCCAGAAGCAGCAACCTCACTGTGTACCGAGACAAGCGAGAACCTGGCAAGCAGCAATGT

[1096]

TCTGCCTCAGACCCGTGGCTGGGCAGAGAACCTGGTTGAAAATATCGATACTGCCCATCTGGCCAACC
TGGGTCCGCTGAGCCAGTTTGAGCCCGGGAAACCGTCGTCCAGTACCAAAGAAGCCATTAACACCGA
ATATAATGATTGGATCTCTAAGTGGCTGCCTAGCGGTGCCGCCACAGAGAAAGACGCGCCAACCCCC
CCAGCACGAAGCTGAATGCACGTGCAACCAGAAGACTGCAATACAGTAGAATCCAGAATCTGTACAA
ACTGAACAGATCTGCCTGTGCACAAGAGGTTCTGTCCGGTGCCTGGAAAAGTTCAGAGCGGAGAACTG
AATTTGAAAGAGGTTTCAGCCGTTTTGGGAAAAGATGTTCCGGAAAAGAAAGCGCAAAAGACAGACGG
AAGCCTAAGCCCACAGGTGAAGTGCTGTGGGGTCTGATGGAACCGTGACCATCGCAGAAGTTGGCA
GCACACTGAAGTCCACGACCCCTAGCGCACCAGGTCCGGATAAACTAACACTGGATGGTGTTAAGCG
TATCCCGATCGCGAACTGGTTAGCCACTACAATCTGTGGCTGTATGCAGGTATCAGCCTGAAGGCCT
GCGGGAAGGCATCACCACACTGATTCCGAAAATTAAAGGCACAAGAGATCCCGCGAAACTGCGTCCG
ATCACCGTGAGTAGTTTCATCTGCCGGATCTTTCATCGTTGTCTGGCTCAGCGTATGGAAACCAGCCTG
CCTCTGGGTGAACGTCAAAAAGCATTGAGAAAGGTGGACGGCATTGTGCATAATATCTGGTCCCTGCG
GAGCCTGATCCACAACAGCAAAGACAATCTGAAAGAGCTGAACATTACCTTTCTGGATGTTTCGAAA
GCCTTCGACAGCATCTCTCATAAGAGCCTGGGAATTGCCGCGGCAAGGCTGGGCCTGCCTCCTCCGCT
AATTACTTACATCAGCAACCTGTATCCGAATTGCAGCACCAAACTTAAGGTGAATGGCAAAATTTCCA
AACCATAGAGGTCCGTCTGTGGCGTGCAGACAGGGTGATCCTCTGAGTCCTCTGCTGTTTAATGCCGTG
ATGGATTGGGCCCTGAGCGAGCTGGACCCTAGAGTAGGTGTTTCAGATCGGCGAACAGCGGATTAACC
ACCTGGCCTTTGACAGATGATATTATCTGGTTAGCAGTACCAAGATTGGCATGGTGAGCTCTATCAATA
CCCTGAGCCGTCATCTGGCCAAAAGTGGACTGGAGATCAGTGCCGGTAAAGAAGGAAAAATCGGCCA
GCATGGCTATTGTTGTGATGGTAAAAAGAAGATGTGGACCGTTGATCCGCTGCCAGATTTAAAGTG
AATAGCCAAAAAATCCGGCACTGAGCATACCCAGCAGTACAAGTACCTGGGGATTAATATCGATGC
ACAGGGCGCCCGGAACGACGCCGCCAGAAATTCTGACCGAGGGCCTAGCCGAACTGAGCCGTGCTCC
GTTAAAACCGCAGCAGAGACTGTATCTGCTCCGCGTGCATCTGTTACCCAAGCTGCAGCATGGTCTGG
TGCTCTCTAGCTGCGCCAAGCGTGCTCTGACCTACTTAGACAAGAGCGTTCGTTTCAGCCATCAGGAGA
TGGCTGACCCTGCCGAAAGATACACCGACAGCTTTTTATCATGCAAAAGCATGCGACGGAGGTCTGG
GTATAACCCGGCTGGAACATACCATTCCTATCCTAAAGCGTAACAGAATGATGAAGCTGACCTTAAGC
GAAGATCCGGTTATCATGGAACCTCGTGAAGCTCACATACTTCACCAATTTATTACACAAGTACAGCAA
CGTGAAACTGCTGAACAGCTGGCCTGTGACCGATAAGGATAGCCTGGCACGGGCTGAGGCAAGCATG
CTGCACACCTCTGTGGATGGCCGGGGCCTGAGCAACTGCAGCGATGTGCCTAGACAGAGCGATTGGG
TTACCAATGGTGCTAGCCTGCTGAGCGGCAGAGATTTTATTGGTGCAATCAAAGTTCGTGGTAATCTGC
TGCCTACCAAAGTGAGCGCAGCTCGTGGACGTCAACGTGAAATTACCTGCGATTGTTGTCGTCGTCCT
GAGAGTCTGGGCCACATCCTACAGACATGTCCGCGACCTGGGGCCCTCGCATTAGCCGGCACGATA

GCCTGCTGAAGAGAGTTTCGCAACCAGGCCTGCCTTAAAACTGGACCCCTATCATCGAACCGAGCAT
TCCTACAAACATTGGTTTACGGAGACCGGATCTGGTGCTGGCGAAAGGCAATATCGCTTTCCTGGTGG
ACGCCACCGTGGTGGCCGATAATGCAAACATGCAGCTGCAGCATGAAGCCAAAGTGGAAAAATATAA
CAACAGCGACATCAAGGAATGGATCAAAGTTTCATTGCCCGGCGTCGACGAGGTTCTGTGACCAGC
CTGACAGCTAATTGGCGGGGCTGTCTGTATGGCGGTAGCGCTAGCTTCCTAACCGAGGACCTGGGTCT
GCCGAAGGCGGAGCTCAGCCTGCTCAGCGCCAAGATTAACGAGAAAGGCTACTACCTGTGGTGTGCA
CATTATAGAGGCACCGCAAGACTGTGGAATCGTCCTCTGAGATCCTGATAA

43. R2D Ciln 氨基酸序列

[1097] MDYKDDDDKGTNTLGENVDASATTSSANPLSQEPGWCESSKLFKSQRGLRVHQRSKHPELYHSQNQPL
PRSKARWSDEEMVIFAREEIANRKIRFINQHLHKVFPHTLESIKGLRGKNVRYARIMADLEAEMTSQPEA
ATSLCTETSENLASSNVLPQTRGWAENLVENIDTAHLANLGPLSQFEPGKPSSTKEAINTYNDWISKWL
PSGAAHRERRANPPSTKLNARATRLQYSRIQNLKLNRSACAQEVLSGAWKVQSGELNLKEVQPFWEK
MFRKESAKDRRKPKPTGEVLWGLMEPLTIAEVGSTLKSTTPSAPGPKLTLTGDKVRIPIAELVSHYNLWLY
AGYQPEGLREGITTLPKIKGTRDPAKLRPITVSSFICRIFHRCLAQRMETSLPLGERQKAFRKVDGICHNIW
SLRSLIHNKSKDNLKELNITFLDVRKAFDSISHKSLGIAAARLGLPPPLITYISNLYPNCSTKLKVNGKISKPIE
VRRGVRQGDPLSPLLFNAVMDWALSELDPVGVQIGEQRINHLAFADDIILVSSTKIGMVSSINTLSRHLLA
KSGLEISAGKEGKSASMAIVDGGKKMWTVDPVPRFKVNSQKIPALSITQQYKYLGINIDAQGARNDA
RILTEGLAELSRAPLKPQQRLLRVHLLPKLQHGLVLSSCAKRALTYLDKSVRSAIRRWLTLPKDTPTAFY
HAKACDGGGLITRLEHTIPILKRNRMKLTSEDVIMELVKLTYFTNLLHKYSNVKLLNSWPVTDKDSL
ARAEASMLHTSVDGRGLSNCSVDPRQSDWVTNGASLLSGRDFIGAIVRGNLLPTKVSAARGRQREITC
DCCRRPESLGHILQTCPTWGPRI SRHDSLLKVRNQACLKNWTPHIEPSIPTNIGLRRLPDVLAKGNIAFLV
DATVVADNANMQLQHEAKVEKYNNSDIKEWIKVHCPGVDEVVTSLTANWRGCLYGGASFLTEDLGL
PKAELSLLSAKINEKGYLWCAHYRG TARLWNRPLRS**

44. R2A LiPo 质粒 ORF 序列

ATGGGGAGCGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGACGA
TGACAAGAAGGGTGGAGGTAGCGGGGGCAGTGGAGGGCCCAAGAAGAAGAGGAAGGTGATGTGCG
TGACCGAGGGTGAATGCCTGCCGCCGACCGACAACCTGCAACCCGAGCGTGGAGAACCAGCTGCCGT
GCGTGACCGAGGGTCGCTTCGAACGTGTGGCAGCCTGGTGACCGTGCCTCTGCCATTTGCAAGGT
GGCCTGCGACCTGTGCAGCAAGGAGTTCCTGACCTACAGCAAATTTGCTGTGCACCAGGCCAACTTC
CACAAACAGCGAAACCCAGGCTTGCTGCACCTATTGCGGCAAGAGCGATGGCAACCACCACAGCATTG

[1098]

CTTGCCATGTGCCAAAATGCCCGTGGCGTCGCACCGTGACCTTCGCCGCCAACCTGAGCAACTTTCTG
TGCGACCTGTGCAACGATAGCTTTAAGACCAAAAGCGGTCTGAGCCAGCACAAAGCGCCACAAACAC
CCGTGCAGCCGTAATGCCGAACGCATCCTGAGCCTGGGCGTGCGTACCCCAAGCGCCCGTCCACGTC
AGGTGGTGTGGAGCGAGGAAGAGACCCGCACCCTGCGTGAGGTGGAAGTGGTGTACAGCGGTCAGA
AGAACATTAACGTGCTGTGCGCCGGTCACCTGCCAGGCAAGACCAGCAAACAAGTGAGCGACAAAC
GTCGCGATCTGCACCGCATTCGTAGCAGCAACGTGCACGGTACCCCGACCAACCCAGAGCCGTGGCGA
CCCGGTGGAGCAGGTGGAAGAGTACGAAGAGCTGGATTGGGAGGGTATGCACCCGTTCCCGGACCC
GGATAGCAAGTTTTCGAGCTATCTGGACCAGCTGCGCGATCAGAAAGGCCTGACCGAGCCGGTGTGG
CAGGAGATCGAAATTGTGGCTCAGGAGTGGGTGGAAAACCTGGCCACGTGCAGAGCAGCTGGAAC
CACGAGCGTACCACCAAGCAGGTGCCAGAAAATAATACCCCGGCTCGTCGCCCGTTCAAGCGTCGCC
TGCACCGTGTGGAGCGTTACAAACGCTTTCAGCGTATGTATGACCTGCAGCGCAAGCGTCTGGCCGA
AGAGATCCTGGATGGCCGTGAAGCGGTGACCTGCAACCTGAAGAAAGAAGAGATTAAAGACCACTA
CGATCAGGTGTATGGTGTGAGCAACGACCGTGTGAGCCTGGATGATTGCCCGCTCCGCCGGGTGCC
AACAACACCGACCTGCTGAAGCCGTTACCCCGACCGAGGTGATGGATAGCCTGCAGGGTATGAAGA
ACGGTGTCTCCGGGCCCCGATAAAATCACCTGCCGTTTCTGCAGAAGCGTCTGAAAAACGGTATTCA
CGTGAGCCTGGCCAACGTGTTCAACCTGTGGCAGTTTAGCGGCCGTATCCCGGAGTGCATGAAAAGC
AACCGCAGCGTGCTGATTCCGAAGGGCAAAAGCAACCTGCGCGACGTGCGTAACTGGCGCCCCGATC
ACCATTAGCAGCATCGTGCTGCGTCTGTACACCCGCATTCTGGCTCGTCGCCTGGAACGTGCCGTGCA
GATCAATCCACGTCAGCGTGGTTTTGTGCCACAGGCGGGCTGCCCGGATAACATTTTTCTGCTGCAGA
GCGCGATGCGTCGCGCTAAGCGTAAAGGTACCCTGGCTCTGGGCCTGCTGGACCTGAGCAAGGCCTT
CGATACCGTGGGCCACAAACACCTGCTGACCAGCCTGGAGCGTTTCGCTGTGCACCCGCACTTTGTG
CGCATCGTGGAAGACATGTACAGCGGTTGCAGCACCAGCTTTCGTGTGGGCAGCCAGAGCACCCGTC
CGATTGTGCTGATGCGCGGTGTGAAGCAGGGCGACCCGATGAGCCCGATCCTGTTTAACATCGCCCTG
GACCCGCTGCTGCGTCAGCTGGAAGAGGAAAGCCGTGGTTTTCATGTTTCGTGAGGGTCAGGCCCCGG
TGAGCAGCCTGGCTTATGCCGACGATATGGCTCTGCTGGCCAAGGACCACGCTAGCCTGCAGAGCATG
CTGGGTACCGTGGATAAATTCTGCAGCGGCAACGGTCTGGGCCTGAACATCGCGAAGAGCGCCGGTC
TGCTGATTGCGGGCGCTAACAAAACCTTTACCGTGAACGACTGCCCAGCTGGCTGGTGAACGGCGA
GACCTGCCGATGATCGGCCCCGAACAGACCTACCGTTATCTGGGTGCTAGCATCTGCCCGTGGACCG
GTATTAACAGCGGCCCGGTGAAGCCGACCCTGGAGAAATGGATCGCCAACATTACCGAAAGCCCGCT
GAAGCCACACCAGCGTGTGGACATTCTGTGCAAATACGCGCTGCCGCGTCTGTTCTATCAGCTGGAGC
TGGGTACCCTGAACTTTAAGGAACTGAAAGAACTGGATAGCATGGTGAAGCAGGCGGTGAAGCGTTG
GTGCCATCTGCCAGCTTGCACCGCCGATGGTCTGCTGTACAGCCGTCACCGTGATGGCGGTCTGGCCG

TGGTGAAGCTGGAAAGCCTGGTGCCGTGAAGATCAAAACCAACCTGCGTCTGGTGCACAGCA
CCGACCCGGTGATTAGCAGCCTGGCCGAGAGCGATGGTCTGGTGGGTGCTATCGAAGGTATTGCTCAG
AAAGCGGGTCTGCCAATCCCAACCCCAGACCAGCGTAGCGGTACCTACCACAGCAACTGGCGCGATA
TGGAGCGTCGCAGCTGGGAACGTCTGGCCCTGCACGGTCAGGGTGTGGAGCTGTTCAAAGGTAGCC
GTAGCGCGAACCCTGGCTGCCACGTCCAGTGGGCATGAAGCCGCACCACTGGGTGAAATGCCTGGC
GATGCGTGCTAACGTGTATCCGACCAAGCGTGGTCTGAGCCGCGGCAACCTGAGCAAGAACAAAGAC
AGCGCCAAATGCCGTGGTTGCACCAGCATGCGTGAGACCTGTGCCACCTGAGCGGCCAGTGCCCGA
AGCTGAAAAGCATGCGTATCCGTGCGCCACAACAAAATCTGCGAGCACCTGATTGCGGAAGCTAGCTT
CAAGGGTTGGAAAGTGCTGCAGGAGCCGACCCTGGTGACCGACAACGGCGAACGTGCGCGTCCGGA
TCTGATCTTTCACCGTGACGATAAGGCGGTGGTGGTGGACGTGACCGTGCGCTACGAGATTAGCAAG
GATACCTGCGTGAAAGCTACGCGAGCAAAGTGCGCCGTTATGGCTGCCTGACCGAGCAGATCAAAG
ACCTGACCGGTGCGACCAGCGTGGTGTTCACGGCTTTCGATGGGTGCGCGTGGCGCCTGGTTCCC
AGAAAGCAGCGACGTGATGGCTGATCTGAACATCCGTAGCAAGTACTTCGAGGAATTTCTGTGCCGC
CGTACCATTCTGTATACCCTGGATCTGCTGTGGAAAAGCAACAACGAGCAGTATCTGGAACGTCTGGC
GCCGTAATAA

[1099]

45. R2A LiPo 氨基酸序列

MGSDYKDHDGDYKDHDIDYKDDDDKKGGGSGSGGPKKKRKVMCVTEGECLPPTDNCNPSVENQLPC
VTEGRFERVGS�VTVRLPFRKVACDLCSKEFLTYSKFAVHQANFHNSETQACCTYCGKSDGNHHSIACHV
PKCPWRRITVFAANLSNFLCDLCNDSFKTKSGLSQHKRHKHPCSRNAERILSLGVRTPSARPRQVWVSEE
ETRTLREVEVVYSGQKNINVLCAHLPKGTSKQVSDKRRDLHRIRSSNVHGTPTTQSRGDPVEQVEEYEE
LDWEGMHFPDPDSKFCSYLDQLRDQKGLTEPVWQEIEIVAQEWVENLAHVQSSWNHERTTKQVPENN
TPARRPFKRRLHRVERYKRFQRMVDLQKRRLAEEILDGREAVTCNLKKEIKDHYDQVYGVSNDRVSLD
DCPRPPGANNTDLLKPFTPTVMDSLQGMKNGAPGPKITLPFLQKRLKNGIHVSLANVFNLWQFSGRIP
ECMKSNSRVLPKGKSNLRDVRNWRPITISSIVLRLYTRILARRLERAVQINPRQRGFVPQAGCRDNIFLLQ
SAMRRAKRKGTALGLLDLSKAFDTVGHKHLTSLERFAVHPHFVRIVEDMYSGCSTSFRVGSQSTRPIVL
MRGVKQGDPMSPILFNIALDPLLRLQLEESRGFMFREGQAPVSSLAYADDMALLAKDHASLQSMGLTVD
KFCSGNGLGLNIAKSAGLLIRGANKTFTVNDGPSWLVNGETLPMIGPEQTYRYLGASICPWTGINS GPVKP
TLEKWIANITESPLKPHQRVDILCKYALPRLFYQLELGTNLFKELKELDSMVKQAVKRWCHLPACTADGL
LYSRHRDGG LAVVKLES LVPCLKIKTNLRLVHSTDPVISSLAESDGLVGAIEGIAQKAGLPIPTDQSRGTY
HSNWRDMERRSWERLALHGQGVELFKGSRSANHWLPRPVGMPKPHHWVKCLAMRANVYPTKRGLSRG
NLSKNKDSAKCRGCTSMRETLCHLSGQCPKLKSMRIRRH NKICEHLIAEASFKGWKVLQEPTLVTDNGE

RRRPDLIFHRDDKAVVVDVTVRYEISKDTLREAYASKVRRYGCLTEQIKDLTGATSVVFHGFPMGARGAW
FPESSDVMADLNIRSKYFEEFLCRRTILYTLDLLWKSNNQYLERLAP**

46. R2D DrMerc 质粒 ORF 序列

[1100] ATGGATTACAAGGATGACGACGATAAGGGTACCATGCACCAGGTGCGGAACGCCGATGCCAATCTGC
AGACCATGTTCCCCTGCAGAGAGTGTGAGCGCAGCTTTCGGACAAAGATCGGACTGGGAGTGCACAT
GAGGCACAGGCACAAGGACGAGCTGGATACCGCCCGGAGAAGGGTGGACGTGAAGGCCAGATGGA
ATGAGGAGGAGCTGTCCATGATGGCCAGGAAGGAGCTGGAGCTGACCGCAAACGGAGAGCGCTTCA
TCAATAAGAAGCTGGCCGAGATCTTTACAAACAGAAGCGTGGATGCCATCAAGAAGTGCAGACAGAG
GGACAATTACAAGGCCAAGATCGAGCAGCTGCAGGGACAGGCAGCACTGATCTCCGAGGCAAACGA
GCCACCTACCACACAGAGGAGGCCATCTCTGAGCGAGCTGGAGGTGACCCCAAGCTCCTCTCACTCC
GTGCCAATCGCACCACCACCTATCCACTCTGACGATATCCTGCTGCAGGAGCTGCAGGGCATGAGCCC
TGTGGCCGTGAGAAGGTCCTGGAGAGTGGAGGTGCTGCAGAGCATCATCGATAGGGCACACATCTCC
GGCAAGGAGGCAACCCTGCAGTGTCTGAGCAACTACCTGCTGGAGATCTTCCCTAACAGAAATGACA
GGCCAAGCTCCGCAACAGTGCCAGCACGCCGGCCTAGGAATAGAAGGATCTCTCGCCGGCAGCAGTA
CGCCCGCTGCATCAAGAGCCTGCTGGACGGAACCGATGAGTCCGCACTGCCTAACCAAGTCTATCATGG
AGCCATATTGGAGGCAGGTCATGACCCAGCCATCCCCCTCTCTGTGTTCCAACACAGTGCCTCGGAAG
GGCAATATGCAGGAGGGCGTGTGGTCTCCAATCACCAGCAGAGATCTGCAGGTGCACAAGGTGCCCC
TGACATCTAGCCCTGGCCCAGACGGCATCACCTCCCAGACAGCCAGATCTATCCCAATCGGCATCATG
CTGAGGATCGTGAATCTGATCCTGTGGTGCGGCGATCTGCCAGTGCCTTTCGCGATGGCAAGGACCAT
CTTTATCCCTAAGACAGTGAGAGCCAACAGGCCACAGGACTTCCGCCCAATCAGCGTGCCCTCCATCG
TGGTGCGCCAGCTGAATGCCATCCTGGCCTCTCGGCTGACCGCAGCCGTGAGCTGGGACCCAGGCA
GAGGGGCTTTCTGCCTACCGACGGATGCGCCGATAACGCCACAATCGTGACCTGGTGCTGCGCGATC
ACCACAAGCGGTACGCCAGCTGTTATATCGCCACCCTGGATGTGAGCAAGGCCTTTGACTCCGTGGCC
CACGATGCCGTGTTTAACACCGTGACAGCCTACGGCGCCCCCAAGAGCTTCGTGGACTACGTGAGAA
GGTGGTATTCCGGCGGGCGGCACATATTTCAATGGCGGCGATTGGAGATCCGAGGAGTTTGTGCCAGCA
AGGGGAGTGAAGCAGGGCGACCCACTGTCCCCGTGCTGTTTAACCTGATCATCGATCGCCTGCTGC
GGTCTCTGCCTAAGGACATCGGCGTGACGTGGGCAACGCCAAAGTGAATGCCTGTGCCTTCGCCGA
CGATCTGATGCTGTTTGCCAGCACCCCTAAGGGCCTGCAGGAGCTGCTGAACACCACAGTGAAGTTC
CTGTCTCTGTGGGCCTGACCCTGAATGCCGACAAGTGCTTTACAATCTCTATCAAGGGCCAGCCAAA
GCAGAAGGTGACCGTGGTGGAGCAGAGGACATTCTGTATCGGCAGAGCCAGGGTGCAGCTGAAGCG
CTCCGAGGAGTGGAAGTACCTGGGCATCCACTTTACAGCAGATGGAAGGGCAAGGTATAACCCAAGC

[1101]

GAGGACATCGGCCCTAAGCTGGAGAGGCTGATGCAGTCCCCCTCTGAAGCCACAGCAGAAGCTGTTCTG
CACTGAGGACCGTGCTGGTGCCTCAGCTGTACCACAAGCTGACACTGGGCAGCGTGGCACTGGGAG
TGCTGAGGAAGTGCATAAGCTGGTGAGATCTTTCGCCAGGAAGCTGCTGGGACTGCCACTGGACGT
GAGCGTGGCCTTTTATCACGCACCACACTCTTGTTGGAGGACTGGGCATCCCAAGCGTGCCTGGATC
GCACCCATGCTGAGAACCAAGAGGCTGGCCGGCATCAATTGGCCACACCTGGAGCAGTCCGAGGTG
GCCTCTGCCTTTCTGAGCGAGGAGCTGAGGAGGGCAAGAGATAGGGCAAAGGCAGGAGTGAACGAG
CTGCTGAGCCAGCCCAAGATCGACACCTACTGGGCCGATCGGCTGTATACATCCGTGGACGGCAATGG
CCTGAGAGAGGCAAGAAGATACGCACCACAGCACGGATGGGTGTCCCAGCCTACCCGGCTGATGTCT
GGCAAGGCCTATCGCACAGGCATCCAGCTGCGGATCAACGCCCTGCCTACCCGCTCTCGGACCACAA
GAGGCAGGCACGAGATGAACAGGCAGTGCAGGGCAGGATGTGATGCACCAAGCCACAATCACGTGC
TGCAGAGATGCCACAGGACACACGGCTCTCGCGTGAGCCGGCACAATGGCGTGGTGAGCTACCTGA
AGAAGGGCCTGGAGACCAGAGGCTACACAGTGTATTCCGAGCAGTCTCTGCACGGCCAGAACCGCG
TGTATAAGCCCCGATATCGTGGCCTTCCGGCACGACTCCACCATCGTGGTGGATGCCAGGTGGTGACA
GACGGACTGGACCTGGATAGGGCACACCAGTCTAAGGTGGAGATCTACAATCGCCAGGACCTGCTGA
CCCACTGAGGTCCGTGTATAGGGCAAGGGAGAACATCGAGGTGGTGTCTGCCACCCTGAATTGGAG
AGGCATCTGGAGCTTTCAGTCCATCACCAGACTGAGGACACTGGGCATCCTGACCGCCGGCGACTCC
AACGTGATCAGCTCCAGAGTGGTGTCTGGCAGGGTGTACAGCTTCAAGACATTCATGTTTCACGCCGG
CTTTCACCGGGGCATGGCCTAG

47. R2D DrMerc 氨基酸序列

MDYKDDDDKGTMHQVRNADANLQTMFPCRECERSFRKIGLGVHMRHRHKDELDTARRRVDVKARW
NEEELSMMARKELELTANGERFINKKLAEIFTNRSVDAIKKCRQRDNYKAKIEQLQGQAALISEANEPPTT
QRRPSLSELEVTPSSSHSVPIAPPPHSDILLQELQGMSPVAVRRSWRVEVLQSIIDRAHISGKEATLQCLSN
YLLEIFPNRNDRPSSATVPARRPRNRRISRRQQYARCIKSLLDGTDESALPNQSIMEPYWRQVMTQPSPLC
SNTVPRKGNMQEGVWSPITSRDLQVHKVPLTSSPGPDGITSQTARSIPIGIMLRIVNLILWCGDLPVPFRMA
RTIFIPKTVRANRPQDFRPISVPSIVVRQLNAILASRLTAAVSWDPRQRGFLPTDGCADNATVLDLVRDHH
KRYASCYIATLDVSKAFDSVAHDAVFNTVTAYGAPKSFVDYVRRWYSGGGTYFNNGDWRSEEFVPARGV
KQGDPLSPVLFNLIIDRLRLSLPKDIGVHVGNKVNACAFADDLMLFASTPKGLQELLNTTVKFLSSVGLT
LNADKCFITISIKGQPKQKVTVVEQRTFCIGRARVQLKRSEEWKYLGIHFTADGRARYNPSEDIGPKLERLM
QSPLKPQQLFALRTVLVPQLYHKLTLGSVALGVLRKCDKLVRSFARKLLGLPLDVSVAFYHAPHSCGGL
GIPSVRWIAPMLRTKRLAGINWPHLEQSEVASAFLSEELRRARDRAKAGVNELLSQPKIDTYWADRLYTS
VDGNGLREARRYAPQHGWVSQPTRLMSGKAYRTGIQLRINALPTRSRTTRGRHEMNRRQCRAAGCDAPSHN

HVLQRCHRTHGSRVSRHNGVVSYLKKGLETRGYTVYSEQSLHGQNRVYKPDIVAFRHDSTIVVDAQVVT
DGLDLDRAHQSKVEIYNRQDLLTTLRSVYRARENIEVVSATLNWRGIWSFQSITRLRTLGLITAGDSNVISS
RVVSGRVYSFKTFMFHAGFHRGMA*

48. R2D DrMe 质粒 ORF 序列

[1102] ATGGGGAGCGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGACGA
TGACAAGAAGGGTGGAGGTAGCGGGGGCAGTGGAGGGCCCAAGAAGAAGAGGAAGGTGatgACCAC
CCGCCCCGAGCGTTGACATCTTCCCAGAGGACCAATACGAACCAAATGCCGCAGCGACCCTGAGCCGT
GTGCCGTGTACCGTGTGCGGTCGTAGCTTCAACAGCAAACGTGGTCTGGGTGTGCACATGCGCAGCC
GTCACCCAGACGAGCTGGATGAGGAACGTCGCCGTGTGGACATCAAGGCCCGTTGGAGCGATGAGG
AAAAATGGATGATGGCTCGCAAGGAGGTGGAACCTGACCGCCAACGGCTGCAAGCACATTAACAAAC
AGCTGGCGGTGTACTTTGCTAACCGTAGCGTGGAAGCCATCAAGAACTGCGCCAGCGTGGTGACTA
TAAGGAGAAAATCGAACAGATTCTGTGGTCAGAGCGCCCTGGCTCCAGAGGTGGCTAACCTGACCATC
CGCCGTGCCCCGAGCCGAGCGAACAGGATCACCAGGTGACCACCAGCGAAACCACCCCGATTACC
CCGTTTGAGCAGAGCAACCGTGAAATCCTGCGTACCCTGCGCGGTTACAGCCCGGTGGAATGCCACA
GCAAATGGCGTGCTCAGGAGCTGCAGACCATCATTGACCGCGCCCACCTGGAGGGCAAGGAAACCA
CCCTGCAGTGCCTGAGCCTGTACCTGCTGGGTATCTTTCCGGCGCAGGGCGTGCGTCACACCCTGACC
CGCCCCGCCACGTCGCCCCGTAACCGTCGCGAGAGCCGTCGCCAGCAGTATGCTGTGGTGCAGCGTA
ACTGGGACAAGCACAAAGGTCGCTGCATCAAGAGCCTGCTGAACGGCACCGATGAAAGCGTGATGC
CGAGCCAGGAGATCATGGTGCCGTATTGGCGTGAAGTGATGACCCAGCCAAGCCCAAGCAGCTGCAG
CGGTGAAGTGATCCAGATGGACCACAGCCTGGAACGCGTGTGGAGCGCGATCACCGAGCAGGATCTG
CGTGCTAGCCGCGTGAGCCTGAGCAGCAGCCCAGGTCCAGACGGTATCACCCCAAAGAGCGCCCGTG
AGGTGCCAAGCGGTATCATGCTGCGCATTATGAACCTGATCCTGTGGTGCGGCAACCTGCCGCACAGC
ATTCGTCTGGCCCGCACCGTGTTTCATCCCGAAGACCGTGACCGCGAAACGTCCGCAGGATTTTCGCCC
AATCAGCGTGCCAAGCGTGCTGGTGCGTCAGCTGAACGCGATTCTGGCTACCCGCCTGAACAGCAGC
ATCAACTGGGACCCGCGTCAGCGTGGTTTCTGCCAACCGACGGTTGCGCTGATAACGCTACCATTGT
GGATCTGGTGCTGCGTCACAGCCACAAACACTTCCGACGTGCTACATCGCCAACCTGGACGTGAGC
AAGGCGTTTGATAGCCTGAGCCACGCTAGCATCTACGACACCCTGCGTGCTTATGGTGCCCCGAAGGG
CTTTGTGGATTACGTGCAGAACACCTATGAGGGTGGTGGCACCAGCCTGAACGGTGACGGCTGGAGC
AGCGAGGAATTCGTGCCAGCTCGTGGTGTGAAACAGGGCGACCCGCTGAGCCCGATCCTGTTCAACC
TGGTTATGGATCGTCTGCTGCGCACCCCTGCCGAGCGAAATTGGTGCGAAAGTGGGCAACGCTATCACC
AACGCCGCGGCTTTGCGCCAGCATCTGGTGCTGTTTGCGGAAACCCGTATGGGTCTGCAGGTGCTGCT

[1103]

GGACAAAACCTGGATTTCCTGAGCATTGTGGGCCTGAAGCTGAACGCCGACAAATGCTTTACCGTG
GGTATCAAGGGCCAGCCGAAGCAGAAATGCACCGTGCTGGAAGCGCAGAGCTTCTACGTGGGTAGC
AGCGAAATTCCGAGCCTGAAACGTACCGACGAGTGGAAGTATCTGGGTATCAACTTTACCGCTACCGG
TCGCGTGC GTTGAACCCAGCGGAGGATATCGGCCCCGAAACTGCAGCGTCTGACCAAGGCGCCGCTG
AAACCGCAGCAGCGTCTGTTGCTCTGCGCACCGTGCTGATTCCGCAGCTGTACCACAAGCTGGCCC
TGGGTAGCGTGGCGATTGGCGTGCTGCGTAAGACCGACAACTGATCCGCTACTATGTGCGTCGCTGG
CTGAACCTGCCGCTGGATGTGCCAATTGCCCTTCGTGCACGCTCCACCAAAGAGCGGTGGCCTGGGCA
TCCCGAGCCTGCGTTGGGTGGCTCCAATGCTGCGTCTGCGTCGCTGAGCAACATCAAATGGCCGCA
CCTGACCCAGAACGAGGTGGCTAGCAGCTTTCTGGAGGCTGAAAAGCAGCGTGCTCGTGACCGCCT
GCTGGCTGAGCAGAACGAACCTGCTGAGCCGTCCGGCGATTGAAAAGTACTGGGCTAACAACTGTAT
CTGAGCGTGGATGGTAGCGGCCTGCGTGAGGGTGGTCACTACGGTCCACAGCACGGTTGGGTGAGCC
AGCCGACCCGCTGCTGACCGGTAAAGAATATATGGACGGCATTCTGCTGCGCATCAACGCTCTGCCA
ACCAAGAGCCGTACCACCGTGGTCGCCACGAGCTGGAACGTCAGTGCCGTGCTGGTTGCGATGCTC
CAGAAACCACCAACCACATCATGCAGAAATGCTACCGTAGCCACGGTCGTCGCGTGGCTCGTCACAA
CTGCGTGGTGAACCGTATTAAGCGCGGTCTGGAGGAACGTGGCTGCGTTGTGATCGTGGAGCCGAGC
CTGCACTGCGAAAGCGGTCTGAACAAACCGGACCTGGTGGCTCTGCGCCAGAACCACATCGACGTG
ATTGATACCCAGATCGTGACCGATGGCCACAGCATGGACGATGCCACCAGCGTAAGATTAACCGTTA
CGACCGTCCAGATATCCGTACCGAGCTGCGTCGCCGTTTCGAAGCTGCTGGCGACATTGAGTTTCACA
GCGCGACCCTGAACTGGCGTGGTATTTGGAGCGGCCAGAGCGTGAAGCGCCTGATCGCTAAAGGTCT
GCTGAGCAAGTATGATAGCCACATCATTAGCGTGCAGGTTATGCGTGGCTCCCTGGGTTGCTTCAAAC
AGTTCATGTATCTGAGCGGCTTTTCCCGTGATTGGACCTAATAA

49. R2D DrMe 氨基酸序列

MGSDYKDHDGDYKDHDIDYKDDDDKKGGGSGGSGGPKKKRKVM TTRPSVDIFPEDQYEPNAAATLSR
VPCTVCGRSFNSKRGLGVHMRSRHPDELDEERRRVDIKARWSDEEKWMMARKEVELTANGCKHINKQL
AVYFANRSVEAIKKLRQRGDYKEKIEQIRGQSALAPEVANLTIRRRPSRSEQDHQVTTSETTPITPFEQSNR
EILRTLRYSPVECHSKWRAQELQTIIDRAHLEGKETTLQCLSLYLLGIFPAQGV RHTLTRPPRRPRNRRES
RRQQYAVVQRNWDKHKGRCIKSLLNGTDESVMPSQEIMVPYWREVM TQSPSSCSGEVIQMDHSLERV
WSAITEQDLRASRVSLSSSPGPDGITPKSAREVPSGIMLRIMNLILWCGNLPHSIRLARTVFIPKTVTAKRPQ
DFRPISVPSVLVRQLNAILATRLNSSINWDPRQRGFLPTDGCADNATIVDLVLRHSHKHFRSCYIANLDVSK
AFDSLSHASIYDTL RAYGAPKGFVDYVQNTYEGGGTSLNGDGWSSEEFVPARGVKQGDPLSPILFNLVM
DRLLRTL PSEIGAKVGNAITNAAAFADDLV LFAETRMGLQVLLDKTLDFLSIVGLKLNADKCF TVGIKGQP

KQKCTVLEAQSFYVGSSEIPSLKRTDEWKYLGINFATGRVRCNPAEDIGPKLQRLTKAPLKPQQLFALR
TVLIPQLYHKLALGSVAIGVLRKTDKLRYYVRRWLNPLDVPIAFVHAPPKSGGLGIPSLRWVAPMLRLR
RLSNIKWPHLTQNEVASSFLEAEKQRARDRLAEQNELLSRPAIEKYWANKLYLSVDGSGLREGGHYGPQ
HWVVSQPTRLLTGKEYMDGIRLRINALPTKSRTTRGRHELERQCRAAGCDAPETTNHIMQKCYRSHGRRV
ARHNCVVNRIKRGLEERGCVVIVEPSLQCESGLNKPDLVALRQNHIDVIDTQIVTDGHSMDDAHQRKINR
YDRPDIRTELRRRFEAAAGDIEFHSATLNWRGIWSGQSVKRLIAKGLLSKYDSHISVQVMRGS LGCFKQFM
YLSGFSRDWT**

50. R2A GaAc 质粒 ORF 序列

[1104]

ATGGATTACAAGGATGACGACGATAAGGGTACCCATATGGCCGGTCCTAGCACCAGATCTAGAAGGAA
GGGCTGCCTGCCCCCTGTGGACTTCGAGAACCCTAAGAAGAGAACAAGACTCATGGCCAAGATGAC
AAATGGCAACCCTACCAGCCACGTGCCGTGCCCCGCACCCTGCAGCAATGGCCACGAGGGCGGCGG
AAGAGTGGCTGTGATCGAGGGCCGGCTTCCAGAACTGAGCGGCAGCAGAATTAGCGGGATTACGCT
GCCCTACCTGTGGAAACCAGCTTCGTGGGACAAAACACCGGCAGAGGGGCGGACGGAGATGCCAAC
GCCAACAGCAGCCCTCCCAGCCCCAACCTGGGCGGCAGCGTGGGCATGGTGCTGCAGTGCGCGATG
GCACACCTCCTCTGGGCGGCGCTGGCGAGGACCACAGCCGGGAATGCGCCGGCGGCAACACCCCTC
TGTGGATGCTGGAGGATAGCTTCCGGTGCGACTACTGCCCTAGAGAGTTCCGAACAAGAGCCGGAAG
AAGCCTGCACATGCGGCGGGCTCACCTGGCTGAGTACGACGGCGCTGGATTTTGTCTGGGGCGAAAGA
CTGTCCGAGTTGCGCGCTACCCGGCTGTGGTCTACCGAGGAAACCAAGAAGCTGGCTGTCTTTGTG
AACGCGGTGTGCCCAGCCCTTCTGAGTGCAGGGCCATCGCCGCCAGCCTCGGCGCCGGTAAGACACA
CCACCAGGTGAGGAGCAAGTGCCGACTGGTGTGTTGAGGCTATCAGACGGAGAGAGCTGCTGGAAGT
GGCCGCCGCCACCGAGAGACTGGAAAAGAGCGCACGGCGGAAGCAGCCTGCCGTGCCCCCAGCACC
CGTGACGGCGTGAGAGGCGTGCTTAGAGGCCTGCTGGGCAAGAGAGTGCCGCGGGAGGGTGGCAC
CACAGGAAGCACCTCCGCCCATTGTGAGAAGGGACGACTGCCGGCAGGGCGCTGTGGCCTCCGC
CAGCTTGAACCTGATCCGGCGGCTGGGCCGCAAGGCTACAGGAAGAAGCGGACGTAGGAGAGTCCT
GGGCAGACCTCCCAGAATGGATGTGCGCCGGAGCGTGCGTATGAGAAGAATGCGGCGGTTCTGTAC
AGACTGGCTCGGCTGGGCTGGGCCAAGCTGGCCATGTTCTGTTGGATGGCCAGATGGGCGCTTCTT
GTCCTGTTCCACTGGTGGAAGTGAGCGCCGTGTTTACAGAGCGGTGGAGCATCGTCAGAGCCTTCCT
GGGCCTGGGACAGTTCGGCGGCTTCGGCACCGCCGACAACGCGGGCTTTGGCAAGTTGATAGACCCT
GCGGAGGTGAGAGCTCACCTGCAGAGCATCAAAAACAGAAGCAGCCCTGGCCCTGACGGCATCACC
AAGGTGGCCCTGTCTAAGTGGGACCCTGAGGGAATCAAGCTGGCCACATGTACAGCACCTGGCTGG
TGTCTGCTGGCATCCCAAAGGTGTTCAAGAAATGCAGAACAACTGATCCCTAAGACCGGCGACGT

[1105]

GTCCCTGCATGGCGACGTGGGGCAGTGGCGGCCTATCACAATCGCCTCCCTGGTGCTGAGACTGTACT
CCCGCATCCTGACCGAGAGAATGACCGTGGCCTGCCCTAGCCATCCTAGACAAAGAGGATTCATCGCC
AGCCCCGGCTGTTCTGAAAACCTGATGCTGCTTGAAGGCTGTATGAGCCTGAGCAAGGCCGGAACG
GCTCCCTGGCTGTTGTGTTGTGGACTTCGCCAAGGCTTTCGACACCGTGAGCCACGAGCACCTGCT
GAGCGTTTTAGTTCAGAAGGGCCTGGACCAGCACATGGTGGAAGTATCAAGGACAGCTATGAGAAC
AGCGTGACCAAGGTGCATTGCCAGGAGGGATGCTCTACCGACATCGCCATGAAAGTGGGCGTGAAAC
AGGGCGATAGTATGTCTCCACTGCTGTTCAACCTGGCCCTAGACCCTCTGATCCAGCAGCTGGAAAGA
GAAGGCAGGGGCTTCCCCGTGAACGGAAGTCTATCACCGCCATGGCCTTTGCCGATGACCTGGCCA
TCGTGAGCGACAGCTGGGAGGGCATGAGAGCCAATCTGGACATCCTGGTGGAAGTCTGTGAGCTGAC
CGAATGCGGACCCAGCCAGCAAGTGCCACGGTTTCCTGATTGAAAAGAGCGGCTCCCGTAGCTAC
AAGGTGAACAGATGCGAGCCTTGGCTGCTGAACGATACCGCCCTGCACATGGTGGGCCCCAAGGAGT
CCATCAAGTACCTGGGGGTGCAAGTGAATCCCTGGACAGGCATCTTCGCCGAAGATACCGTGGCTAA
GCTGAGACAGTGGGTGGTGGCCATCAGCAAGACACCTCTGAGACCTCTGGATAAGGTTAGCCTGCTG
TGTCAGTTCCCGTGCCTAGGGTGATCTTCGTGGCCGACCACTGTATGCTGAGCGCCAAAGCCCTGAC
AGAGATGGACAGAAGCATCCGGCAGGCCGTAAAGAGATGGCTGCACCTAGCTCGGTGTACCACCAAC
GGCCTGCTCTACAGCCGGAAGTCATCCGGCGGCCTGGGCATCCCTAAACTGAGCATGATCGTCCCTGC
GATGCAGGCCAGACGGCTGCTGGGGCTGAGCCGCTCTAAAGACGAGACTGTAAGATGGATGTTTCTG
GAAACCACAGATCACGTGGCCTTCGAGAGAGCCTGGCTTCGGGGCCGGCGGAAGCCCCGACGAGGTG
CCAGAGCTGGGCCCTGATCTGGTGGAAGGATCTCCAGCCGAAGGCAATGCCGATCCCGTGTCTACAG
TGAGACCCAGAAAGAGAATCGTGCCATGTGACTGGAGACAAGTGGAGTTCGACAGATGGGCCGGCC
AACTGGTTTCAGGGAAAGGGAATCAGAACATTCGAGGCCGATAAGATCTCTAATTGCTGGCTGTATGAC
TACCCTCCTAACAAGCTCAAGCCTGGCGATTCACCGCCGCCGTGCAGCTGAGAGCCAACGTGTACC
CCACAAGAGAGCTCGCCGGCAGAGGCAGAACCGACACAATCGACGTGTGCTGCCGGCACTGCCGCG
AGGCTCCTGAAACCTGTTGGCACATCCTGGCCCTGTGCCCCAAGGTGAAGCGGTGCAGAATCCAGAG
ACACCACAAGGTCTGCCAGGTGCTGGTGCCGAGGCCGAGAGACACGGCTGGGAAGTGGAAAGAG
AGAAACGGTGGATGCTGCCAGCGGCGAATGCGTGGCCCCCGATCTGATCTGCTGGCTGGACGAACT
GGCACTGATCGTGGACGTAACCGTTAGATACGAGTTCGACGAGGAATCCCTGGAAAGAGCCAGAATC
GAGAAAGAATGCAAGTACCGGCCCCCTGATCCCCGTGATCAGAGCCTCTCGTGTGCAGACCAAGAAAAG
TCACAGTTTACGGCTTTCCTCTGGGCGCCCGTGGTAAGTGGCCTGCCAAAAACGAGCTGCTGCTGGC
CGACCTGGGCCTGTCAAAGGCCAGGACCAGATCTTCGCCAAACTTCTGTCCCGGAGAGTGCTTCTG
CACAGCCTGGATGTCATGAGAACCTTCATGCGCTAG

51. R2A GaAc 氨基酸序列

MDYKDDDDKGTHMAGPSTRSRKGCCLPPVDFENPKKRTLMAMKMTNGNPTSHVPCAPCSNGHEGGGR
VAVIEGRLPELSGSRISGIQPALPVETSFVGQSTGRGADGDANANSSPPSPNLGGSVGMVPAVRDGTPLGR
PGEDHSRECAAGNTPLWMLEDSFRCDYCPREFGTRAGRSLHMRRHLAEYDGAGFCWGERLSEFAATR
LWSTEETKKLAVFCERGVPSPSECRAIAASLGAGKTHHQVRSKCRLVFEAIRRELLEVAAATERLEKSAR
RKQPAVPPAPVHGVRGVLRLGLLGKRVPREGGTTGSTSARIVRRDDCRQGAVASASLNLIRRLGRKATGRS
GRRRVLGRPPRMDVRRSVRMRMRRLFLYRLARLGWAKLAMFVLDGQMGASCPVPLVEVSAVFRERWSI
VRAFLGLGQFGFGTADNAGFGKLIDPAEVRHLQSIKNRSSPGPDGITKVALSKWDPEGIKLAHMYSTW
LVSAGIPKVFKKCRTTLIPKTGDVSLHGDVGQWRPITIASLVLRLYSRILTERMTVACPSHPRQRGFIASPGC
SENLMLEGCMSLSKAGNGSLAVVFVDFAKAFDTSHEHLLSVLVQKGLDQHMVELIKDSYENSVTKVH
CQEGCSTDIAMKVGKQGDMSPLLFNLALDPLIQQLEREGRGFPVNGKSITAMAFADDLAIVSDSWEG
MRANLDILVDFCELTGMRTQPSKCHGFLIEKSGSRSYKVNRCPEWLLNDTALHMGPKESIKYLGVQVNP
WTGIFAEDTVAKLRQWVVAISKTPRLPLDKVSLLCQFAVPRVIFVADHCMLSAKALTEMDRSIRQAVKRW
LHLARCTTNGLLYSRKSSGGLGIPKLSMIVPAMQARRLLGLSRSKDETVRWMFLETTDHFVAFERAWLRA
GGSPDEVPELGPDLVEGSPAEGNADPVSTVRPRKRIVPCDWRQVEFDRWAGQLVQKGKIRTFEADKISNC
WLYDYPPNKLKPGDFTAQVQLRANVYPTRELAGRGRDTIDVCCRHCGEAPETCWHILALCPKVKRCRI
QRHHKVCQVLVAEAEERHGWEVEREKRWMLPSGECVAPDLICWLDELALIVDVTVRYEFDEESLERARIE
KECKYRPLIPVIRASRVQTKKVTVYGFPLGARGKWPKNELLADLGLSKARTRSAKLLSRRVLLHSLD
VMRTFMR*

[1106]

52. R2A PuPu 质粒 ORF 序列

ATGGATTACAAGGATGACGACGATAAGGGTACCATGAACATCCCCACCCCTGTGGGCATCGAGGTGCT
GCTGCCTGACCCTAATCCTAACTGCCCTCTGTGCGGCGTCCAGGTGGGCAAGGTGGGAATGCTGGGC
AAACACTTCGCCAGCCGCCACGCCAGAGTGCCCGTGTCTTATAAGTGTGCGGCGTGTGGCAGGACAA
ACAGCAACTCCCATTCTATCAGCTGTACGTGCCTAAGTGCAAGGGACTCGGCCTGGAAGAGACAGG
AAGAAAGGAAGGCGAATTTGGCTGCGAGCTGTGTTCAGGACGGTTCAGAACCGGCGTGGGGCTGAC
CCAACACCGGCGGCTGGCTCACCCTGTGGAATGGAACATGGAAGAGGTGGCTCTACCCAGCCTAGA
GGAGGCCTGAAGGGCACAAAGCTGTGGTCCGTGGACGAGCTGAAAAGCCTGATCGAGCTATCCGTG
CGGTTTCGCTGGCGAACCCAATCATAAGATCAACAAGCTGATCGCTGATGAACCTGGTAGCGGCAAGA
CCGCCGAGCAGGTCAGAAGCAAGCGCCGGCTGCTGCAAAAGGAGATGGCCAGCAACGGCCCCCA
TCGCTGCCGCTGAAGAGAGCGGCACCTCTCGGGCCCAGGCCGATCCGATCGTCCCCAGAGCCCTCC
TAACGTGGCCAAGACACGGCTGATGGAGGCCATCAACACCGAGGGCGGCACCGAGGGCGAGAACGA

[1107]

GCGGAGAGCCATCGCCGACCTGATCAGAGACGTGGACCAGAACCCCGCCTGATCGAGTCTAGCGCC
CTGGAGCTGATCCAGAGACTGGGAAAGGGCGCCACCAGAAATAGAGCCGAAAACGGCAGAGAGATC
ACCAGAGCTAAGAAAGGATGGGAAAGACGGCTGAGCCAGAAGCGCAGAGAATTCAGAGTGCACCA
GAGACTCTACGAGAAGGACCCAACCAAGCTGGCAGCGCTGGTGCTGGACGGCGATGCCTGTGCTGA
GTGCACCATCCCTGTGACCCAAATCGACGAAGAATACAGACGGCGCTGGGAGGCCGGCGGCGAGTTC
CACGGGCTGGGCAACTTCAGATCTGCCGGAGGCGCCGATAATAGCGAGTTTACAACCCCATCCTGG
CCACCGAGGTGTCCGAGAACCTGATCAAGATGAAAAACGGCTCTGCCCCTGGCCCTGATGGCATCTC
CAAGAAGGCCCTGCTTCAATGGGACCCTGCCGGCGAACAACCTGGCTAGACTCTACACTGCCTGGCTG
GTGCACGGCGTGATCCACGGGCCTTCAAGGAATGTAGAACCAAGCTGCTGCCCAAGAGCGGTGAC
GCGGATGAGCTGAAGGACGTCAACGGCTGGAGACCTATCACCATCGGCTCCCTAGTGCTGAGACTTT
TCAGCAGAATCCTGACAATGCGGCTGACCAGAGCCTGCACAATGAACCCTAGACAGAGAGGTTTTCT
GGCCTCTAGCAATGGTTGCGCCGACAATCTGATGGTGCTGAATGGCGTGATCAAGCGGAGCCGGGAA
GAACGGGCTCCTCTGGCCGTCGTGTTCTGTGGACTTCGCCAAGGCCTTTGATAGCATCTCACACGAGCA
CTTTCTGCGGGTGCTGGAGCAGCGGCAGCTGGATAGACATGTGATCGAACTGATTAAGAACAGCTAC
GTGGACTGCGTGACAAGAGTGGGATCTGGCGGCGAACGCACCAGCCCCATCAACATGAAAGTGGGA
GTCAAACAGGGCGACCCAATGAGCCCCCTGCTGTTCAACCTGGCTATGGACCCTTTGATCCAAGCCCT
GGAGGACGGCCAGAGCGGCCTGTCTTGGGATGGCAGATCCATCAGCACCCCTGGCCTTCGCCGACGAC
TTAGTGCTGCTGAGCAGCAGCGTGCGGGCATGAGAAGAAACCTGGAGATCCTGGAAGAATTCTGCC
AGCTGACCGGCCTGAAGGTGCAGCCTAAGAAGTGCCACGGCTTCTACTACGACAAGGGCGTGATCAA
CGACTGTGAGCCTTGGACCATCGGCGGCACCAGCATCCACCTGGTGGGCCCTGGCGAGAGCGTGAAG
TACCTGGGAGTTTCAGGTGGGCCCCAGAAGAGGCATCACAATGCCTGATCTCTCGCCACAGCTGCAGG
AGTGGATCGTGCGCATCAATAAGGCACCACTGAAGCCTACGCAGAGGGTGAAAGTGCTGAATAGCTT
CGCCCTGCCTCGACTAATCTACCAGGCTGACCACTGCGACGTGCAGGGCACCGCCCTGGCCACACTG
GACGGAATGGTGAGAAAGGCCGTGAAAACATGGCTGCACCTGGCCCCTAGCACTTGTAACGGCCTGC
TGACAGCAGAAATAGAGACAGCGGCCTGGGCATCGTGAGACTGGAAAGACTGATCCCTAGCCTGCA
GGTCAAGAGAATCCACAGACTGTCCCGGAGCCAGGACCTGATGACCCGGTTAATTACCGTGAAAACC
GTCACCCAGTCTGAGTGGCAGCGCAAGTGGCAAATGGTCGGCGGCGATCCTGACAACAGTCCAGTGT
TGGAAGTCGAGGAGGGATTCCCCGTGGACGGCAGAAGCCTGGTTGTGACCGATTGGAGGCGGGGAG
AGAACGAGGCCTGGCTGGCCCTCCAGGTGCAGGGCATGGGCGCTGGTGAGTTTAAGGGCGATAAAAT
CAGCAACAGCTGGCTGTTTCGACCCCACTAAGGTGGGCTTCAGACAGCGGCACTACATCGCCGCCCTG
GCCCTTCGGACCAACACCTACCCACAGGCGAGTTCCTGGCACGCGGCAGAGATAAGGCCGGCGCCG
CATGCAGACGGTGCCAGGCCCGGCTGGAGACATGCAGCCACATACTCGGGCAGTGCCCTTCGTGAA

AGGCAGCAGAATCAGACGGCACACAGAATCTGCGAACTGCTGGCCAACGAAGCCGAGAGAGTGGG
CTGGAAGGCTACCAGAGAGTTTACAGAGTGATTACACCTGAGGGCGGAGTACGGATTCCAGACCTGGTG
TGCACCAAGGGCGATAAGGCCCTGGTGTGACGTGACAATTAGATTTCGAGATGGCCCCTGACACCC
TGTTGCGCGCTGCCGCCGAAAAGGTGAACCACTACCAGCCTGTCGCCTCTCAGATTGCCAGCATGGT
GGGCGCTAAAGAAGTGAAAGTCATGGGATTCCCTATCGGCGCCAGAGGAAAGTGGCCTTCTTGCAAC
AACAAAGTGCTGACAGAGCTGGGCGTTACCGTGGGCAGACGGGCCACCTTTGCCAGACTGGTGAGT
AGACGCTCCTTGCTGTATAGCTTGGATATCCTGCGGGCTTTCGTGTGCGAGCCCGCCCTGTAG

53. R2A PuPu 氨基酸序列

[1108] MDYKDDDDKGTMNIPTPVGIEVLLPDPNPNCLCGVQVGKVGMLGKHFA SRHARVAVSYKCRRCGRTN
SNSHSISCHVPKCKGLGLEETGRKEGEFGCELCSGRFRTGVGLTQHRRLAHPVEWNMEKVGSTQPRGGL
KGTKLWSVDELKSLIELSVRFAGEPNHKINKLIADELGSGKTAEQVRSKRRLQKEMASNGPPIAAAEESG
TSRAQADPIVPQSPPNVAKTRLMEAINTEGGTEGENERRAIADLIRDVDQNPGLIESSALELIQRLGKGATR
NRAENGREITRAKKGWERRLSQKRREFRVHQRLYEKDP TKLAALVLDGDACAECTIPVTQIDEEYRRRW
EAGGEFHLGNFRSAGGADNSEFHNPI LATEVSENLIKMKNGSAPGPDGISKALLQWDPAGEQLARLYT
AWLVHGVIPRAFKECRTKLLPKSGDADELKDVNGWRPITIGSLVLR LFSRILTMRLTRACTMNPRQGF LA
SSNGCADNLMVLNGVIKRSREERAPLAVVFVDFAKAFDSISHEHFLRVLEQRQLDRHVIELIKNSYVDCVT
RVGSGGERTSPINMKVGKQGDPMSPLLFN LAMDPLIQALEDGQSGLSWDGRSISTLAFADDLVLLSSSV
RGMRRNLEILEEFCQLTGLKVQPKKCHGFYDYDKGVINDCEPWTIGGTSIHLVGPESVKYLGVQVGPRRG
ITMPDLSPQLQEWIVRINKAPLKPTQRVKVLNSFALPRLIYQADHCDVQGTALATLDGMVRKAVK TWLHL
APSTCNGLLYSRNRD SGLGIVRLERLIPSLQVKRIHRLSRSQDLMT RLITVKT VQTQSEWQRKWQMVGGDP
DNSPVLEVEEGFPVDGRSLVVDWRRGENEAWLALQVQGMGAGEFKGDKISNSWLFDP TKVGFRQRHY
IAALALRTNTYPTGEFLARGRDKAGAACRRCQARLETCSHILGQCPFVKGSRRRHHRICELLANEAE RVG
WKATREFRVITPEGGVRIPDLVCTKGDKALVLDVTIRFEMAPDTLLRAAAEKVNHYQPVASQIASMVGAK
EVKVMGFPIGARGKWPSCNNKVLTELGVTVGRRATFARLVSRRLYS LDILRAFVCEPAL*

54. TriCan 质粒 ORF 序列

ATGGATTACAAGGATGACGACGATAAGGGTACCCATATGAGTCAGAAGAGAAGACCCGAAAAAGCCG
TCCCCGATGAAGGAGCCACCGCCACGATGTCGCCAGCCAGACAAGAGTAAATGTAGCGTGTGCGG
CGAGACATTCAAGGGCCCAGCCAGCGTGACCATGCACATGGTGAAGAAGCACCCCGTGGAGTTTAAT
GAGCTGAAGATGGCCAAGAAGCCCGTGCCTAAGAAGGTGAGATGGTCCGAGGAGGAGATCTTCCAG
CTGGCCAGGACAGAGGCAGAGCTGACCCTGCAGGGCGTGCGCTTCATCAACGTGGAGCTGCAGAAG

[1109]

ATCTTTCCTGCCAGAGAGATCGAGGGCATCAAGGGCCAGAGGAAGCTGGCCAAGTACAAGGAGCTG
GTGAAGGACCAGCTGGATGAGATCGGCAGGGCCCCAAATCCCCCTGAGCAGGAGATCGGCGAGGAC
GTGCCAAGCCCCTTCAAGGCATGGCTGGAGCTGCTGCTGGCCCTGCCAAAGACCCCCAATGATTTTCT
GGAGCACAAGCTGGACAACATCATCGTGCAGGGCCCTGAAGGAGGACGTGAACTCTGATCAGGTGTTT
AACGATCTGAATAGCTACCTGAAGCTGATCCTGGAGCCATCTGGAAGGGCAAAGAGCGTGCCCGGAG
AGATCATCCACGGCGACCCTTCTGGCAGCGCCAAGACCTCCGTGACAAAGGGCCCTAAGCCAGCCAC
AGTGAGCTCCTCTCGGAAGAAGCGGAGAGATGCCGAGTTTGCCAGAATCCAGCGGCTGTATAGGAAG
AACCGCACATCTTGTCATCAATACCATCCTGGACGGCAACACCAGAGAGCACGAGGGCCCCCAAGAATA
TGGAGGGCTTCTGGCGCGAGATCTTTGAGCGGGAGAGCCCTGACGATCCAGACGATCCCGATATCTTC
CTGGAGGAGGAGGCCTCCGACATCTGGAAGTACATCTCTTTTTATGAGATGTGCAACCTGTATCCACC
ACCTTCCACAGCACCAGGACCTGACGGCTTCAGCTCCAAGGATCTGAGGCGCATGACCCCCAGGGTG
CTGAACAAGATCCTGAATCTGCTGCTGCACCTGCGCGATCTGCCTCAGATCCTGAAGAGCCACCGGA
CAGTGCTGATCCCAAAGACAGATCTGCCTACCAAGCCAGGCGACTTCCGCCCTATCACCATCAGCAAC
ATCCTGGTGCGCCACCTGAACAAGATCCTGGCCAATCGGGTGTCCACCTGATCCCATCAATGAGCG
GCAGAAGGCCTTTCTGCCTATCGATGGCTGCGCCGAGAACATCTTCACACTGGACTTTATCCTGCACC
ACGCCAGAACCAAGATCAAGTCCCTGTCTATGGCCATCCTGGACATCAGCAAGGCCTTCGATAGCGTG
TCCCACCACTCCATCTTTCGGGCCCTGAGAGAGGCCAGGTGCCCAATCGGCTTCATCAAGTTTATCGA
GAATTGCTACGGCGGCTGTTTACCAAGCTGTTTTGTGGCGGCGTGAAGTATCCCTCTGAGGTGAGCA
TGAATAGGGGCGTGAAGCAGGGCGATCCCCTGTCTCCCGTGCTGTTCAACCTGGTCATCGACGGCCT
GATCCGCCAGATCCCAAGCGCCCTGGGCTTTAACGTGTCTGACCAGGTGAAGGTGAGCTGCATCGCCT
ACGCCGACGATCTGATCCTGATCGCCACCACACGGGCGGCGCTGAAGACACTGCTGGATCTGACCAA
TAGCTATCTGGCCAAGAGAGGCCTGTCCCTGAACCCGACAAGTGTTCCGCCCTGTCTATCGTGGCCA
GCGGCAAGCAGAAGCTGGTGATACATCGCCTCTAGCGAGCACTTCGATCTGGCCGGCCAGAAGATGAG
AAACCTGAATGTGGGCGACTCCTGGAGGTATCTGGGCATCCAGTTTTCTCACCTGGGAAGGGCAGAG
AAGGTGACACCAGATCTGACCTGTCTGATCAATAGACTGCAGAAGGCCCCCTCTGAAGCTGCAGCAGA
AGCTGTACGCCCTGAGAATCTATCTGATCCCCAGGCTGATCCACGGCCTGACACTGTCCAAGACCAAC
CTGGGCGAGCTGAAGACCCTGGACAAGCTGATCCGCAAGTACATCCGGGCCTGGCTGCACCTGCCAG
ACGATACACCCATGGGCTACTTCTATACCCCTCTGAAGGCAGGAGGCCTGGGCCTGCCAAGCCTGCGG
CTGGTCATCCTGAACAATCGCCTGGAGAGAATCCTGAGAATGAAGGCCTCCCAGGATATCATCGTGAG
GACAATCGCCGAGTCTGAGACACTGGGCGTGGAGATCCGCAAGCTGCACGACCTGCTGTCCATCGAT
GGCACAATCCTGGACACCTCCGTGAAGATCCACTCTTTTTGGGCCGAGAGACTGTACTCCTCTTATGA
CGGCAAGTGCCTGTGCAACTCTGCCAATTTCCACCCGGCAATAAGTGGATCGGCGAGGATAGCCTG

AACCAGCGGTCCCACATCTTTGCCGACTGCCTGAAGCTGAGAATCAACGCCCTGCCTACACGGTCCA
GAACCGCAAGGGGCCGCCCTCTGAAGGATAAGCCATGCAGAGCCGGCTGTAGGAATTCTGACGGCGT
GAAAGTGATCGAGACACTGAACCACATCACCCAGGTGTGCGAGAGGACACACGGAGCAAGGGTGAA
GAGGCACGACAGGCTGGTGGATTTTCGAGTGAAGGGCCTGCAGCGGCCCCACAGGGTGGTGCTGAA
GGAGCCCCACTACAAGACCGTGAATGGCGTGAGGAAGCCCGACATCGTGATCAAGATCCCTGACCAC
ACATATATCTGCGATTTCCAGGTGGTGAGCGACACCTCCTGTCTGGAGCTGGAGTTTCGCAAGAAGGC
CCTGAAGTACGCCGAGGATAAGGGCCTGTGCGACCAGCTGACCAGGGATCACCCAGGAGAGCTGAG
CTTCACCGCCATCACCTTCAACACCCGCGGCCTGATCGCCAAGAGCTCCGTGACAGCCCTGAGGAAG
CTGGGAATGCCTCCACGGTCCATCATGACCCTGCAGAAGATCTGTATGGAGGGGTCACTGGAAATCTG
GAGAATCTTCAATCAGACTACTGCTATGGCACGCAACTAG

55. TriCan 氨基酸序列

[1110]

MDYKDDDDKGTHMSQKRRPEKAVPDEGATAHDVAQPDKSKCSVCGETFKGPASVTMHMVKKHPVEFN
ELKMAKKPVPKKVRWSEEEIFQLARTEAELTLQGVRFINVELQKIFPAREIEGIKQQRKLAKYKELVKDQL
DEIGRAPNPPEQEIGEDVPSPFKAWLELLLALPKTPNDFLEHKLDNIIVQALKEDVNSDQVFNDLNSYLKLI
LEPSGRAKSVPGEIHHGDPGSAKTSVTAKPKPATVSSSRKKRRDAEFARIQRLYRKNRTSCINTILDGNTRE
HEAPKNMEGFWREIFERESPDDPDDPDIFLEEEASDIWKYISFYEMCNLYPPPSTAPGPDGFSSKDLRRMTP
RVLNKLNLHLRLDLPQILKSHRTVLIPKTDLPKPGDFRPITISNILVRHLNKLANRVSHLIPINERQKAFL
PIDGCAENIFTLDFILHHARTKIKSLMAILDISKAFDSVSHHSIFRALREARCPIGFIKFIENCYGGCFTKLFC
GGVKYPSEVSMNRGVKQGDPLSPVLFNLVIDGLIRQIPALGFNVSDQVKVSCIAYADDLILIAATTRAGLKT
LLDLTNSYLAKRGLSLNPDKCSALSIVASGKQKLVIYASSEHFDLAGQKMRNLNVGDSWRYLGIQFSLHG
RAEKVTPDLTCLINRLQKAPLKLQKLYALRIYLPRLIHGLTSLKTNLGELKTLDKLIRKYIRAWLHLPDD
TPMGYFYTPLKAGGLGLPSLRLVILNNRLERILRMKASQDIIVRTIAESETLGVEIRKLHDLLSIDGTILDT
VKIHSFWAERLYSSYDGKCLCNSANFPNGK WIGEDSLNQRSHIFADCLKLRINALPTRSRTARGRPLKDK
PCRAGCRNSDGKVIETLNHITQVCERTHGARVKRHDRLVDFAVKGLQRPHRVVLKEPHYKTVNGVRKP
DIVIKIPDHTYICDFQVVS DTS CLELEFRKKALKYAEDKGLCDQLTRDHPGELSFTAITFNTRGLIAKSSVTA
LRKLGMPPRSIMTLQKICMEGSLEIWRFNQTTAMARN*

56. LeCoB 质粒 ORF 序列

ATGGATTACAAGGATGACGACGATAAGGGTACCATGAGCGGCAAGAGCTCCAAGCCTCGCACCGTGT
CTAGCGGCTCCTCTAGCCAGGAGACACCCCTTCCGGCTCTAATGCCTGCGACATCTGTGGCAAGTGC
TTCATGAAGCCCGTGGGCCTGTCTCTGCACATGAGCAAGGTGCACCCTACCCAGTACCACGCCCCGCT

[1111]

GGAGAAGAACCAGCCAAAGGCCAAGAAGTTCCGGTGGACAGACGAGGATCTGTATTTTCTGGCCAA
GAAGGAGGCAGAGCTGCTGCTGCTGGGAGGCATCAAGTTTATGAATAAGGAGCTGGCCGAGTTCTTT
CCCGAGAAGTCTGTGGATCAGATCAAGGGCCAGCGGAGAAGCGAGACCTACAAGCAGCAGGTGGTG
AGCATCCACTCCGAGCTGCTGAAGCTGCAGGCAGTGGCAGACTCCCCACCACCTTCTCGCATCCCCG
CAAAGGAGGTGTCCGCATGGCTGGATTTCTGCTGGCCCTGCCTAAGACCAAGAATAAGTTTCTGAG
GACAAGCTGGATCAGCTGATCAGAACCGCACAGGAGGGAACACCAGTGCTGAACGACCTGGATCTG
TATCTGAGGGAGGTGCTGGTGCAGCCAACCAGACAGGGAGAGAGGCAGGCCAAAGCCACTGCCACCC
CCTAAGTCCTCTCGGGAGAAGAGGGACCGCGAGTACGCCAGAGTGCAGAATTTCTATAGGAAGAACA
AGACCGCCTGTGTGAACGCCATCCTGGATGGCAATAAGAAGTGCAGAGAACAAGATCCCTGACATCGA
TGAGTTCTGGAAGGCCATCTTTGAGAGCCAGTCCCCACCAGACGCAGAGCCCGTGAGCTACGTGGTG
GATGAGGAGCCTAAGAATATCTGGTCTTGGATCAGCTTCTTTGAGATGAACAGAAATTATCCCGACAC
CTCTACAAGCCCAGGACCTGATGGAGTGACCGCAAGGATGCTGAGATCCATCCCTGCCAGGGTGCTG
AACAAGCTGCTGAATCTGCTGCTGTTTCATCGAGGACCTGCCCGCCGTGTTAAGTGTACCCGACCGT
GCTGATCCCTAAGGTGGATAACCCAGCACTGCCAGGAGAGTTCCGGCCAATCACAATCAGCTCCATCA
TCGTGAGACAGCTGAATAAGATCATCGCAGCAAGGGTGAGCGAGGGAGTGCCTATCAACCCACGCCA
GAAGGCCTTTCGGCAGATCGACGGCTGCGCCGAGAACGTGTTCCCTGCTGGACTTTATCCTGAGGGAT
GCCAAGACCAAGATCAAGAGCCTGTCTCTGGCCACAGTGGACATCAAGAAGGCCTTCGATAGCGTGT
CCCACCACAGCATCTTTAGGGCAATCAGGGGAGCAAGGTGTCCAGAGAACCTGGTGAATTACATCCA
GAATTCCTTATAGCGGCTGCACCACACAGATCTCCGTGGGCGGCTCCATCTCTACCACAAAGATCCTGA
TGAACAGAGGCGTGAAGCAGGGCGATCCCCTGAGCCCTGTGCTGTTCAACCTGGTCATCAACGAGAT
CATCAGGAAGCTGCCTGCCTCCATCGGCTACCCAATCAATTCCGAGCTGTCTATCAACTGCATCGCCTA
TGCCGACGATCTGATCCTGGTGGCCAATACCAGAGAGGGCCTGAAGCTGCTGCTGAACCTGCTGAAT
GAGGAGCTGCCCAAGAGGGGCCTGGAGCTGAACGCCTCTAAGTGTTTTGGCCTGAGCCTGACAGCC
CTGGGCAAGCTGAAGAAGACCCACCTGTGCACAAGCGACCAGCTGGATCTGCACGGCACCCCTGATC
AAGAATCTGACAGCCGAGGAGAGCTGGGTGTACCTGGGCGTGCCATTCTCCACATCGGCCGCAGCA
AGTCCTTTTCTCCCGACCTGGAGGCCCTGCTGAACAAGCTGCAGAAGTCCCCACTGAAGCTGCAGCA
GAAGCTGTTTCGCCCTGCGCGTGTATCTGATCCCAAGGCTGCTGCACGGACTGGTGCTGTCTAGGGTGG
CAATCGGCGAGCTGAAGATCATGGACAAGCTGATCCTGAAGCACCTGAGGGTGTGGCTGAGGCTGCC
AAAGGATACCCCCCTGGGCTTCTTTTACTCCCCAGTGAAGCTGGGCGGCCTGGGCATCAAGAACCTG
CGCACAAATGTGCTGAAGTGTAGAAAGCAGAGGATCGAGCGCATGCTGGTGTCCCCAGACGATGTGG
TGCGGCTGGTGGCAGAGTCTGAGATCTTTCTGAAGGAGACCGACAAGCTGAAGGATCTGCTGACAAT
CAATGGCATGTGCCTGGACACCAGAAACGTGCCTAGGACAGGCAAGAACAATAAGTTCTGGTCCGAG

AGACTGTACACCTCTTTTGACGGCAAGACACTGGCCTACAGCGAGTATTTACCCAGGGCGGCGGCT
GGATCAGAGAGGATAAGATCCTGCAGCCAGCCACGTGTTTGCCGAGTGTATCAAGCTGCGGATCAA
CGCCCTGCCTACCAAGAGCAGAGTGGCCACGGCAGGCCAACAAAGGACCGGTCCTGTAGAGCCGG
CTGCCTGGATGTGCAGAAGGTGCCTGCCATCGAGACCATCAATCACATCGCCAGGTGTGCCCAAGA
ACACACGGAGCAAGGATCAAGAGGCACGACAGGCTGGTGCAGTTCCTGTCCCTGAACCTGCGCAAG
AATCCTAAGCGGAACGTGCTGGTGGAGTACAATTTTAGAACCGTGGCCGGCATCAGGAAGCCAGACA
TCATCGTGATCGAGGATAACAAGGGCAGCAATCCTGGACGTGCAGGTGGTGGGCGATTCTAGCAACCT
GGAGATGGAGTACCTGGAGAAGAGCAGAAAAGTATTCCAATGATGCCACCCTGTCTATGAGGATCAAC
GCCCTGCAGAAGCTGTATCCCACCGTGACAAGCCTGACCTTCCACGCCGTGACCTTCAACAATAGGG
GCCTGATCGCCAAGTCTACAGTGGCAGCACTGAGGATGCTGGGAGTGCCTCCACGGTGTATCATGATC
CTGTGCGTGATCAGCCTGGAGAAGACCCTGGAAGTGTGGCGCATGTTCAACCAGAGCACAGCCTCCG
CCCGGAAGTAG

57. LeCoB 氨基酸序列

[1112]

MDYKDDDDKGTMSGKSSKPRTVSSGSSSQETPPSGSNACDICGKCFMKPVGLSLHMSKVHPTQYHARLE
KNQPKAKKFRWTDDELYFLAKKEAELLLGGIKFMNKELAEFFPEKSVDQIKGQRRSETYKQQVVSIIHSE
LLKLQAVADSPPPSRIPAKEVSAWLDLFLALPKTKNKFSEDKLDQLIRTAQEGTPVLNDLDLYLREVLVQPT
RQGERQAKPLPPPSSREKRDREYARVQNIFYRKNKTACVNAILDGNKKCENKIPDIDEFWKAIFESQSPPD
AEPVSYVVDEEPKNIWSWISFFEMNRNYPDTSTSPGPDGVTARMLRSIPARVLNKLNNLLFIEDLPAVFKC
HRTVLIPKVDNPALPGEFRPITISSIIVRQLNKIIAARVSEGV PINPRQKA FRQIDGCAENVFLDDFILRDAKT
KIKSLSLATVDIKKAFDSVSHHSIFRAIRGARCPENLVNYIQNSYSGCTTQISVGGSSISTTKILMNRGVKQG
DPLSPVLFNLVINEIIRKLPAISIGYPINSELINCIAYADDLILVANTREGLKLLNLLNEELPKRGLELNASKC
FGLSLTALGKLKKTHLCTSDQLDLHGTLIKNLTAESWVYLGVPFISHGRSKSFSPDLEALLNKLQKSPLK
LQQKLFALRVYLI PRLLHGLVLSRVAIGELKIMDKLILKHLRVWLRLPKDTP LGFFYSPVKLGGLGIKNLRT
NVLKCRKQRIERMLVSPDDVVRLVAESEIFLKETDKLKDLLTINGMCLDTRNVPRTGKNNKFWSERLYTS
FDGKTLAYSEYFTQGGGWIREDKILQPAHVFAECIKLRINALPTKSRVAHGRPTKDRSCRAGCLDVQKVPA
IETINHIAQVCPRTHGARIKRHDLVQFLSLNLRKNPKRNVLVEYNFRTVAGIRKPDIIIVIEDTRAAILDVQV
VGDSSNLEMEYLEKSRKYSNDATLSMRINALQKLYPTVTSLTFHAVTFNNRGLIAKSTVAALRMLGVPPR
CIMILCVISLEKTLEVWRMFNQSTASARK*

58. 转染的 RT mRNA 5' UTR

AGGAAUAAACUAGUAUUCUUCUGGUCCCCACAGACUCAGAGAGAACCCGCCACC

59. 转染的 RT mRNA 3' UTR 和聚 A 尾

UGAUGACCUCGAGCUGGUACUGCAUGCACGCAAUGCUAGCUGCCCCUUUCCCGUCCUGGGUACCC
CGAGUCUCCCCCGACCUCGGGUCCCAGGU AUGCUCCCACCUCACCUGCCCCACUCACCACCUCUG
CUAGUUC CAGACACCUC CCAAGCACGCAGCAAUGCAGCUAAAACGCUUAGCCUAGCCACACCCC
CACGGGAAACAGCAGUGAUUAACCUUUAGCAAUAAACGAAAGUUUAACUAAGCUAUACUAACCC
CAGGGUUGGUCAAUUUCGUGCCAGCCACACCCUGGAGCUAGCAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAGCAUAUGACUAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAA

60. OrLa5L 模块

GGCGGGAGTAACTATGACTCTCTTAACGCACAGGGGACACAGAGCCTGCCCAAGTACCGCTCCCGAG
GGAGCGGGAAACGGGGGGGTGACTATCCCCTGGGGTCCGGCGAGAGCGCTGGTCTACGGACCAGGG
GTGGCTGTGGGCAGGCTGCTCCTCAGGCCAGTTGATTAGTTACGCATGGGCTGTACCTCCACGTGGTC
CCGCTGGTAACGACTTGTCGGCTAAATCAGCCCGCCACCATCTGGGATATGGTTGACCGTCTAACCC
CAGTACTCAGGTCACAAACAAAATGGGAACAGATACAGTGTATGTCGGCCAGGACTACCCTTCTGGC
TTATCAAAACGGGTACCAGCACGGT

[1113]

61. OrLa5S 模块

GGCGGGAGTAACTATGACTCTCTTAACGCACAGGGGACACAGAGCCTGCCCAAGTACCGCTCCCGAG
GGAGCGGGAAACGGGGGGGTGACTATCCCCTGGGGTCCGGCGAGAGCGCTGGTCTACGGACCAGGG
GTGGCTGTGGGCAGGCTGCTCCTCAGGCCAGTTGATTAGTTACGCATGGGCTGTACCTCCACGTGGTC
CCGCTGGTAACGACTTGTCGGCTAAATCAGCCCGCCACCATCTGGGATATGGTTGACCGTCTAACCC
CAGTACTCAGGTCACAAACAAA

62. TriCasB5 [aka TCA5-28(28)]

GACTCTCTTAAGGGAAGACCCCGCCCATGAGGCTTGGAGAGTGTGATCCTGATCAGATCACACTTGA
AAAGTTATGCTGAGTACGTCCGCGTCGTGAGAGTCGGTAACTGTCCAGGATGGTCTGGGATAGGCTA
AACCTCAGCAGGGGAAAGTTGTAGGGGCCTGCCACCCCTACACTTTATTGGTATGGCATTTCGATACCC
CTAACGAAGCCTCGGACTTGGAGGAGCACGGTTCCTCCTCGTATTAGACCAGGAACCAACTG
TCCTGACAACCCCATTTGGACCTATGGGAGCGGACCATGCTATGGACATGGATTCCGAAGACGAAGCG
GGGGCACACGGACCCCCCGCCGATAGTGCTCACTTAACGTCAGGCGAACCCTTGAAATCATCTTG

63. TriCasB_5RZdead [aka TCA5_rzdead]

GACTCTCTTAAGGGAAGACCCCGCCCATGAGGCTTGGAGAGTGTGATCCTGATCAGATCACACTTGA
AAAGTTATGCTGAGTACGTCCGCGTCGTGAGAGTCGGTAACTGTCCCAGGATGGTCTGGGATAGGTTA
AACCTCAGCAGGGGAAAGTTGTAGGGCCTGCCACCCCTACACTTTATTGGTATGGCATTTCGATACCC
CTAACGAAGCCTCGGACTTGGAGGAGCACGGTTCCCTCCTCCTCGTATTAGACCAGGAACCAACTG
TCCTGACAACCCCATTTGGACCTATGGGAGCGGACCATGCTATGGACATGGATTCCGAAGACGAAGCG
GGGGCACACGGACCCCCCGCGGATAGTGCTCACTTAACGTCAGGCGAACCCCTTGAAATCATCTTG

64. TriCasB_5RZ

GACTCTCTTAAGGGAAGACCCCGCCCATGAGGCTTGGAGAGTGTGATCCTGATCAGATCACACTTGA
AAAGTTATGCTGAGTACGTCCGCGTCGTGAGAGTCGGTAACTGTCCCAGGATGGTCTGGGATAGGCTA
AACCTCAGCA

65. TriCasB_5RZmin

GACTCTCTTAAGGTGCTGAGTACGTCCGCGTCGTGAGAGTCGGTAACTGTCCCAGGATGGTCTGGGAT
AGGCTAAACTCAGCA

[1114]

66. TriCasA5(28)模块

GGCGGGAGTAACTATGACTCTCTTAAGGAGTCATAGAGCCAGAACCTCCTCGTGGTCCCGCTGGGCA
CAGGGATTAATTTTCTGTGGCAAATTTGACTGGCTTCAGAGAGCGTTTTCGAAGTGGACTGTGTGA
CTGCGTTCCCCCTTAGTTGCTATATCCGTTTCGATTAAACATCTCACCTCGACGTATAAGATCATT

67. TriCasA_RZ 模块[aka TCARZ-28(28)]

GGCGGG AGTAACTATGACTCTCTTAAGG AGTCATAG AGCCAGA ACC TCCTCGT GGT CCCGCT GG
GCACAGGGATTAATTTTCTGTGG CAAA TTTGACT GGCTTCAGA

68. ZoA15 模块

GGCGGGAGTAACTATGACTCTCTTAAGG cgacttgaga aggtctggtt acaactgggc atagct gcagagatcg cgcc
tcctcgtggc ccegct ggta agccctaac aggggtgacta agt cgatctctgc ccagtc caggagccgc tgggttcac cagcccagcg
attccttcca aattcggtga aacaaattcc tcggtaaaag ccgctggct tattgcctga aacctggccc ccggttca gacaggggca aagagtcgg
aagtggactg accaccacc ccgaaccga gagcgaatct ggtcatgacc caactgtccc aaatcctggt ccgtctctg gagcggggga

aggtgcacag ccactaccct tactcaggt atcgggtggc acccaaacct gtgaagagga ctttataaca tctagaccaa ccaaattacc
cggaattgaa tcagaattag gcccgtggt gaagtttct ttagagggtt acaggtcaga tcttaagggg gatgtgcaat ttgaggggat
tcattttcca gataattggg ggttactgga ggggttctt gaggtgtacg aacaactggc accacagcca aacgggggag acgagttaa tcatagtctc
ccaggggaca gggaggggga tgtactgag aaggatagca gcgaaaagga gaaggaggct gcaccagagg cattgccctc agtgcaaagg
gcccgcagtg aacagttgcc

69. TiGu5 模块

GGCGGGAGTAACTATGACTCTCTTAAGG ctggggaccg tggttacaac ccgggcttag ct gcagagacag tacctccc cgtgggt
ccc gcc ggacccc gtaacatcgg gtgactgaat ctgtctctgc ccggggagta gticctcctt gccctatga ccagcggctg
ccggctgctc aatagtattc taggcgtgaa atatagcgt agtcctagtg gttgtcttac tggccatag ccccttgctt caggggtcat
tcgcgaagtc tctcaggaga actgggggtg gtgttcttct gggtatagct aaaccacct aactgtgtcc gatccatggg
gtcctggatc gtgaatttcg ttccgtggc gactcagacg ggagaattcc ctgtggatac ggccaggagg gcacctgtgc cggtaacatc
ataccctgag tcggaatgcc acntaccgtt gcccctgaca tttgtaact cggatgtgac tatttgggga ggggttcgcc
ctgaaccggt ggactgctg ggwgatctc crgaggygta tgatgcactc ccaggggtgg ctgggcctcg ggaaycggtg
ggtgggagcc cgccgggrga aggggtcagg tcgccaggga ttgcgtccc ctcgtgact gcggtccaac atgatttgg gagtccatc ctcgtaccgg
[1115] gagccgaagc cgccgaggtt tctaccccg tagtgaaggt tcncaagac catccagcat

70. DroSi5 模块

GGGGATCTGGGGTAATTGCGAGCAGAGGGGGAGTATTTTCTGTAAATTCGTAAGTCATATCATATGGTG
TGCGGAAGGGGAATTTTACTCTGTAACTCACAAGTCTCTCTTTACTCAAGTCGACTCAAAACCTCCT
CGTGGTGGTCCCCGTAATGCTAAACTTGTATTAGCAGCTAATTTGAGCGGCAAAACTTTTCCGATGG
GCTGGTTACCCAGAGGAAATTTACTCATATTGGAACACGAACACAAATAACGAGCCTCGGATATCTTT
ACACAATCTG

71. TriCan5 模块

Ggtctggcatttgatcgtttccgctcctcgtggcgccagactgggtaagctgatttatcagtgagctaagagaaacgatcccgaggagtgccatctacctgcgctgc
acgtgatttcattgtgcacttggcgagttatcccttgggagctcgtcgttagcttttgag

72. LeCoB5 模块

ggggtagcaattgatcattccgctcctcgtggcgctaccctgggataacctcaagaaatttgacggtaagctaagaggaattgatcacaccgtgacgaatatca
tcagtcacagctgcacgaatccagatagatataaacaggcgagtaattctttcgaagctcgtcagtaattcttccag

73. FoRa5 模块

GGCGGGAGTAAGTATGACTCTCTTAAGGTAGTCCAAAAGCCTTGTCATCTTAGGATGACGCGCAGGTT
TTTGAGATGTATATTTTATAGTAGTAGTAAAATACGACTTGTTAAATTCTTAAGATAAAGAGTCATAGT
AGCAAGTGAAGACCTTTTCGTGGTTCCTGTCATAAGTACCTACGGGGAAACCAACTTCACCCCTGCG
AACATGGAGCTTAGAGGTTACCAGCCTCCTCGCTCGTCAATTAGCGAAGAGC

74. CiIn5 模块

GGTAACTATG ACTCTCTTAA cgacgggtgaa ccacctgtgc gcgggtgaag agctttagtgc tctgaacaa gaaatagctt gtgtgctgtc
cttctgggcg gtgcacatac ttcttaacct cccgaggcca tgcggggcggg ggctttagcc cccggcaggt ttaccatgc cggacgggtt cgagaggttag
aggccaaact aagagttcac cagcagactt cgcacgcggc tggccactgg ccgaagtta aacaacaggg ccgcatcttc
ccaaactcaa tatatggtgt taagtgaacc gtgccgatgg gtgaatggcc atgggtaagt tggctcttaa cggctcctgt ggagaagtgg cgaccctta
caatacttca gccttacct atgcccgggc aacttaggt ggacgtatat ctcccggga aaacatctta tctcatggat aaaaatatat acgaaaacac
tacttctca ggagggggac cgcttgcgg tgagaagacg catcgttccg atgtgatcat cctccaccg gggttcgcc cgtcgacgga
cactgcttcc

[1116]

75. OrLa 5RZop

ggcgggAGTAACTATGACTCTCTTAA cgcacagggg acacagagcc tgcccaagta ccgctcccga gggagcggga aacggggggg
tgactatccc ctgggggtccg gcgagagcgc tggcttacgg accaggggtg gctgtgggca ggctgctcct caggccagtt gattagtac
gcatgggtgt acc tcc acgt ggt cccgt gg ta acgaccttggtcgg cGaa acagccc

76. OrLa 5RZmin

ggcgggAGTAACTATGACTCTCTTAA gggctgt acc tcc acgt ggt cccgt ggtaacgaccttggtcgg cGaa acagccc

77. OrLa5L_MS2

GGCGGG AGTAACTATGACTCTCTTAA
CGCACAGGGGACACAGAGCCTGCCCAAGTACCGCTCCCGAGGGAGCGGGAACGGGGGGGTGACTA
TCCCCTGGGGTCCGGCGAGAGCGCTGGTCTACGGACCAGGGGTGGCTGTGGGCAGGCTGCTCCTCAG
GCCAGTTGATTAGTTACGCATGGGCTGTACCTCCACGTGGTCCCGCTGGTAACGACCAGGAACACCGG
TCGGCTAAATCAGCCCGCCACCATCTGGGATATGGTTGACCGTCTAACCCAGTACTCAGGTCACAA
ACAAAATGGGAACAGATACAGTGTATGTCGGCCAGGACTACCCTTCTGGCTTATCAAAACGGGTACCA
GCACGGT

78. Orla5S_PP7

GGCGGG AGTAACTATGACTCTCTTAA
CGCACAGGGGACACAGAGCCTGCCCAAGTACCGCTCCCGAGGGAGCGGGAAACGGGGGGGTGACTA
TCCCCTGGGGTCCGGCGAGAGCGCTGGTCTACGGACCAGGGGTGGCTGTGGGCAGGCTGCTCCTCAG
GCCAGTTGATTAGTTACGCATGGGCTGTACCTCCACGTGGTCCCGCTGGTA
ACGACACGACATAGGAGTCGGTCGG CTAAATCAGCCC
GCCCACCATCTGGGATATGGTTGACCGTCTAACCCAGTACTCAGGTCACAAACAAA

79. Orla5L_MS2(CAAA)

GGCGGG AGTAACTATGACTCTCTTAA
CGCACAGGGGACACAGAGCCTGCCCAAGTACCGCTCCCGAGGGAGCGGGAAACGGGGGGGTGACTA
TCCCCTGGGGTCCGGCGAGAGCGCTGGTCTACGGACCAGGGGTGGCTGTGGGCAGGCTGCTCCTCAG
GCCAGTTGATTAGTTACGCATGGGCTGTACCTCCACGTGGTCCCGCTGGTA
ACGACCAGGAACACCGGTCGG CAAAATCAGCCC
GCCCACCATCTGGGATATGGTTGACCGTCTAACCCAGTACTCAGGTCACAAACAAA
ATGGGAACAGATACAGTGTATGTCGGCCAGGACTACCCTTCTGGCTTATCAAAACGGGTACCAGCACG
GT

[1117]

80. OrLa5L_P2

GGCGGG AGTAACTATGACTCTCTTAA
CGCACAGGGGACACAGAGCCTGCCCAAGTACCGCTCCCGAGGGAGCGGGAAACGGGGGGGTGACTA
TCCCCTGGGGTCCGGCGAGAGCGCTGGTCTACGGACCAGGGGTGGCTGTGGGCAGGCTGCTCCTCAG
GCCAGTTGATTAGTTACGCATGGGCTGTACCTCCACGTGGT CCCGCT GGTA
ACGACCAGGAACACCGGTCGG CTAA
ACAGCCCGCCACCATCTGGGATATGGTTGACCGTCTAACCCAGTACTCAGGTCACAAACAAA
ATGGGAACAGATACAGTGTATGTCGGCCAGGACTACCCTTCTGGCTTATCAAAACGGGTACCAGCACG
GT

81. Orla5L(28)

GGCGGG AGTAACTATGACTCTCTTAAGG
CGCACAGGGGACACAGAGCCTGCCCAAGTACCGCTCCCGAGGGAGCGGGAAACGGGGGGGTGACTA
TCCCCTGGGGTCCGGCGAGAGCGCTGGTCTACGGACCAGGGGTGGCTGTGGGCAGGCTGCTCCTCAG

GCCAGTTGATTAGTTACGCATGGGCTGTACCTCCACGTGGTCCCGCTGGTA
ACGACCAGGAACACCGGTCGG CTAA
ATCAGCCCGCCACCATCTGGGATATGGTTGACCGTCTAACCCAGTACTCAGGTCACAAACAAA
ATGGGAACAGATACAGTGTATGTCGGCCAGGACTACCCTTCTGGCTTATCAAAACGGGTACCAGCACG
GT

82. Orla5L(27)

GGCGGG AGTAACTATGACTCTCTTAAG
CGCACAGGGGACACAGAGCCTGCCCAAGTACCGTCCCGAGGGAGCGGGAAACGGGGGGGTGACTA
TCCCTTGGGGTCCGGCGAGAGCGCTGGTCTACGGACCAGGGGTGGCTGTGGGCAGGCTGCTCCTCAG
GCCAGTTGATTAGTTACGCATGGGCTGTACCTCCACGTGGTCCCGCTGGTA
ACGACCAGGAACACCGGTCGG CTAA
ATCAGCCCGCCACCATCTGGGATATGGTTGACCGTCTAACCCAGTACTCAGGTCACAAACAAA
ATGGGAACAGATACAGTGTATGTCGGCCAGGACTACCCTTCTGGCTTATCAAAACGGGTACCAGCACG
GT

[1118]

83. Orla5L(25)

GGCGGG AGTAACTATGACTCTCTTA
CGCACAGGGGACACAGAGCCTGCCCAAGTACCGTCCCGAGGGAGCGGGAAACGGGGGGGTGACTA
TCCCTTGGGGTCCGGCGAGAGCGCTGGTCTACGGACCAGGGGTGGCTGTGGGCAGGCTGCTCCTCAG
GCCAGTTGATTAGTTACGCATGGGCTGTACCTCCACGTGGTCCCGCTGGTA
ACGACCAGGAACACCGGTCGG CTAA
ATCAGCCCGCCACCATCTGGGATATGGTTGACCGTCTAACCCAGTACTCAGGTCACAAACAAA
ATGGGAACAGATACAGTGTATGTCGGCCAGGACTACCCTTCTGGCTTATCAAAACGGGTACCAGCACG
GT

84. #65的重复

85. TriCasB 5RZmin+5下游

gactctc ttaAGG tgctgagt acgtccgcgtcgt gagagtc ggta actgtcccaggatggctgggatagg ctaa actcagca
ggggaaagtgtagggcctgccaccctacactttattggatggcattcgatacccctaacgaagcctcgacttggaggagcacgggtccctcctcctcgatta

gaccaggaaccaactgtcctgacaacccattggacctatgggagcggaccatgctatggacatggattccgaagacgaagcgggggcacacggaccccccgcc
gatatgtctcacttaacgtcaggcgaacccctgaaatcatcttg

86. TriCasB_RZop

gactctc ttaAGG gctgagt acc tcc gcgt ggt gagagtc ggta acagccauugcacuccggctgg cgaa actcagc

87. TCA_MS2

GGCGGG AGTAACTATGACTCTCTTAAGG

AGTCATAG AGCCAGA ACC TCCTCGT GGT CCCGCT GG GCACAGAGGAACACCCTGTGG CAAA
TTTGACT GGCTTCAGA GAGCGTTTTT CGAAGTGGAC TGTGTGACTG CGTTCACCCC TTAGTTGCTA
TATCCGCTTC GATTAACATC TCACCTCGAC GTATAAGATC ATT

88. TriCasA5_catdead

GGCGGG AGTAACTATGACTCTCTTAAGG AGTCATAG AGCCAGA ACC TCCTCGT
GGT CCCGCT GG GCACAGGGATTAAATTTTTCTGTGG TAAA TTTGACT GGCTTCAGA GAGCGTTTTT
CGAAGTGGAC TGTGTGACTG CGTTCACCCC TTAGTTGCTA TATCCGCTTC GATTAACATC
TCACCTCGAC GTATAAGATC ATT

[1119]

89. TCARZ_MS2

GGCGGG AGTAACTATGACTCTCTTAAGG AGTCATAG AGCCAGA ACC TCCTCGT GGT CCCGCT GG
ACACAGAGGAACACCCTGTGG CAAA TTTGACT GGC

90. TCARZ_P1

GGCGGG AGTAACTATGACTCTCTTAAGG AGTCATAG AGCCAGA ACC TCCTCGT GGT CCCGCC GG
ACACAGAGGAACACCCTGTGG CAAA TTTGACT GGC

91. TCARZ_P2

GGCGGG AGTAACTATGACTCTCTTAAGG AGTCATAG AGCAGA ACC TCCTCGT GGT CCCGCT GG
ACACAGAGGAACACCCTGTGG CAAA TTTGCT GGC

92. TCARZ(27)

GGCGGG AGTAACTATGACTCTCTTAAG AGTCATAG AGCCAGA ACC TCCTCGT GGT CCCGCT GG

ACACAGAGGAACACCCTGTGG CAAA TTTGACT GGC

93. TCARZ(26)

GGCGGG AGTAACTATGACTCTCTTAA AGTCATAG AGCCAGA ACC TCCTCGT GGT CCCGCT GG

ACACAGAGGAACACCCTGTGG CAAA TTTGACT GGC

94. TCARZ(25)

GGCGGG AGTAACTATGACTCTCTT AGTCATAG AGCCAGA ACC TCCTCGT GGT CCCGCT GG

ACACAGAGGAACACCCTGTGG CAAA TTTGACT GGC

95. TCARZ_wob1

GGCGGG AGTAACTATGACTCTCTTAAGG AGTCATAG AGCAGA ACC CCCCCAC GGT CCCGCC GG

ACACAGAGGAACACCCTGTGG CAAA TTTGCT GGC

96. TCARZ_wob2

GGCGGG AGTAACTATGACTCTCTTAAGG AGTCATAG AGCAGA ACC CCCCCAC GGT CCCGCC GG

[1120]

GCAACACAGAGGAACACCCTGTGG CAAA TTTGCT GGC

97. TCARZ_wob3

GGCGGG AGTAACTATGACTCTCTTAAGG AGTCATAG AGCAGA GCC CCCCCAC GGC CCCGCC GG

GCAACACAGAGGAACACCCTGTGG CAAA TTTGCT GGC

98. TCARZ_wob4

GGCGGG AGTAACTATGACTCTCTTAAGG AGTCATAG AGCAGA GCC CCCCCAC GGC CCCGCC GG

GCAACACAGAGGAACACCCTGTGG CAAA TTTGCT GGC

99. TCARZ(28)_NL

GGCGGG AGTAACTATGACTCTCTTAAGG AGTCATAG AGCCAGA ACC TCCTCGT GGT CCCGCT GG

GCACAGGGATTAATTTTCTGTGG CAAA TTTGACT GGC TTCTAGA

100. TCARZ(28)_L8 [自剪切移除前 8 个 nt]

GGGTAAAC GGCGGG AGTAACTATGACTCTCTTAAGG AGTCATAG AGCCAGA ACC TCCTCGT GGT

CCCGCT GG GCACAGGGATTAATTTTCTGTGG CAAA TTTGACT GGC TTCTAGA
AGGTCGACCAGATGTCCGAG

101. TCARZ(26)_L8 [自剪切移除前 8 个 nt]

GGGTAAAC GCGGG AGTAACTATGACTCTCTTAA AGTCATAG AGCCAGA ACC TCCTCGT GGT
CCCGCT GG GCACAGGGATTAATTTTCTGTGG CAAA TTTGACT GGC TTCTAGA

102. hdvRZ-28

ggcgggAGTAACTATGACTCTCTTAA ggtccca gcc tcc tcgt ggc tccgcc ggta acagccauugcacuccggctgg cgaa
tgggacc

103. hdvRZ-13_1

gactctcTTAAGG ggtccca gcc tcctcgt ggc gagagtc ggta acagccauugcacuccggctgg cgaa tgggacc

104. hdvRZ-13_2

gactctcTTAAGG ggtccca gcc CCCTCAC ggc gagagtc ggta ACACAGAGGAACACCCTGTGG cgaa tgggacc

[1121]

105. hdvRZ-21

GTAACCTATGACTCTCTTAAGG ggtccca gcc tcc tcgt ggc tagttac ggta acagccauugcacuccggctgg cgaa tgggacc

106. hdvRZ+0

ggccggc at ggtccca gcc tcc tcgt ggc gccggcc ggta acagccauugcacuccggctgg cgaa tgggacc

107. hdvRZ-28(26)_gu1

GCGGG AGTAACTATGACTCTCTTAA AAAAGAGAATCATAGAA CGTCAGCA GCCTCCTCGC GGC
CCCGCC GGTAAGATTCCGAAAGGAATCG CGAA TGCTGACG AAAGTCTTC

108. hdvRZ-28(26)_ac1

GCGGG AGTAACTATGACTCTCTTAA AAAAGAGAATCATAGAA CGTCAGCA GCCCCCTCAC GGC
CCCGCC GGTAAGATTCCGAAAGGAATCG CGAA TGCTGACG AAAGTCTTC

109. hdvRZ-28(26)_ac2

GGCGGG AGTAACTATGACTCTCTTAA AAAAGAGAATCATAGAA CGTCAGCA GCCCCCTCAC GGC
CCCGCC GGTA ACACAGAGGAACACCCTGTGG CGAA TGCTGACG AAAGTCTTC

110. hdvRZ-13(13)_gu1

GACTCTC TTAAGG CGTCAGCAGCCTCC TCGCGGCCCC GCCGGTAAGA TTCCGAAAGG AATCG
CGAA TGCTGACG A AAGTCTTC

111. hdvRZ-13(13)_ac1

GACTCTC TTAAGG CGTC AGCAGCCCCC TCACGGCCCC GCCGGTAAGA TTCCGAAAGG AATCG
CGAA TGCTGACG A AAGTCTTC

112. hdvRZ-13(13)_ac2

GACTCTC TTAAGG CGTC AGCAGCCCCC TCACGGCCCC GCCGGTAAGA CAGAGGAACA
CCCTGTCG CGAA TGCTGACG A AAGTCTTC

113. hdvRZ-28(28)_ac2b

[1122]

GGCGGG AGTAA CTATGACTCTCTTAAGG AAGAGAATCATAG AA CGTCAGCA GCCCCCTCAC GGC
CCCGCC GGTA ACACAGAGGAACACCCTGTGG CGAA TGCTGACG A

114. hdvRZ-28(27)_ac2c

GGCGGG AGTAA CTATGACTCTCTTAAG A AAGAGAATCATAG AA CGTCAGCA GCCCCCTCAC GGC
CCCGCC GGTA ACACAGAGGAACACCCTGTGG CGAA TGCTGACG A

115. hdvRZ-28(25)_ac2d

GGCGGG AGTAA CTATGACTCTCTTA CAA AAGAGAATCATAG AA CGTCAGCA GCCCCCTCAC GGC
CCCGCC GGTA ACACAGAGGAACACCCTGTGG CGAA TGCTGACG A

116. hdvRZ-13(13)_ac11

GACTCTC TTAAGG A CGCAGCA GCCCCCTCAC GGC GAGAGTC GGTA
ACACAGAGGAACACCCTGTGG CGAA TGCTGCG A

117. hdvRZ-13(13)_ac12

GACTCTC TTAAGG A CGCAGCA GCCCCCTCAC GGC GAGAGTT GGTA
ACACAGAGGAACACCCTGTGG CGAA TGCTGCG A

118. hdvRZ-13(11)_ac11b

GACTCTC TTAAAA CGCAGCA GCCCCCTCAC GGC GAGAGTC GGTA
ACACAGAGGAACACCCTGTGG CGAA TGCTGCG A

119. hdvRZ-13(11)_ac12b

GACTCTC TTAAAA CGCAGCA GCCCCCTCAC GGC GAGAGTT GGTA
ACACAGAGGAACACCCTGTGG CGAA TGCTGCG A

120. HDVRZ-13(13)_ac5

GACTCTC TTAAGG A CGACGCG GCC CCCACAC GGC GAGAGTC GGCA
ACACAGAGGAACACCCTGTGG CGAA CGCGTCG A

121. HDVRZ-28(28)_gu5b

[1123]

GGCGGG AGTAA CTATGACTCTCTTAAGG AAAAGAGAATCATAGAA CGTCAGCA GCCTCCTCGC
GGC CCCGCC GGTA ACACAGAGGAACACCCTGTGG CGAA TGCTGACG A

122. HDV-28(26)_gu5

GGCGGG AGTAA CTATGACTCTCTTAA AAAAGAGAATCATAGAA CGTCAGCA GCCTCCTCGC GGC
CCCGCC GGTA ACACAGAGGAACACCCTGTGG CGAA TGCTGACG A TCTAGA

123. HDV-28(22)_gu5

GGCGGG AGTAA CTATGACTCTC AACAAAGAGAATCATAGAA CGTCAGCA GCCTCCTCGC GGC
CCCGCC GGTA ACACAGAGGAACACCCTGTGG CGAA TGCTGACG A TCTAGA

124. HDV-28(18)_gu5

GGCGGG AGTAA CTATGAC AACA GTCATAGAA CGTCAGCA GCCTCCTCGC GGC CCCGCC GGTA
ACACAGAGGAACACCCTGTGG CGAA TGCTGACG A TCTAGA

125. HDV-28(14)_gu5

GGCGGG AG TAACTA AACA TAGTTA GAA CGTCAGCA GCCTCCTCGC GGC CCCGCC GGTA
ACACAGAGGAACACCCTGTGG CGAA TGCTGACG A TCTAGA

126. HDVRZ-28(28)_gu5_NP2 [aka HDV-28(28)_rzdead]

GGCGGG AGTAA CTATGACTCTCTTAAGG AAAAGAGAATCATAGAA CGTCAGCA GCCTCCTCGC
GGC CCCGCC GGTA ACACAGAGGAACACCCTGTGG CGAA

127. HDVRZ-28(28)_gu5_CD

GGCGGG AGTAA CTATGACTCTCTTAAGG AAAAGAGAATCATAGAA CGTCAGCA GCCTCCTCGC
GGC CCCGCC GGTA ACACAGAGGAACACCCTGTGG AGAA TGCTGACG A

128. SL-28_26

GG CGGGAGT A ACTAT GA CTCTCTT AACA AAGAGAG AATAGT AACTCCCG TCTAGA

129. HDV-28(28)_gu5b_CDNL

[1124]

GGCGGG AGTAA CTATGACTCTCTTAAGG AAAAGAGAATCATAGAA CGTCAGCA GCCTCCTCGC
GGC CCCGCC GGTA ACACAGAGGAACACCCTGTGG AGAA TGCTGACG A TCTAGA

130. HDV-28(28)_gu5b_L8 [自剪切移除前 8 个 nt]

GGGTAAAC GGCGGG AGTAA CTATGACTCTCTTAAGG AAAAGAGAATCATAGAA CGTCAGCA
GCCTCCTCGC GGC CCCGCC GGTA ACACAGAGGAACACCCTGTGG CGAA TGCTGACG A TCTAGA

131. HDV-28(27)_gu5b_L8 [自剪切移除前 8 个 nt]

GGGTAAAC GGCGGG AGTAA CTATGACTCTCTTAAG AAAAGAGAATCATAGAA CGTCAGCA
GCCTCCTCGC GGC CCCGCC GGTA ACACAGAGGAACACCCTGTGG CGAA TGCTGACG A TTCTAGA

132. HDV-28(25)_gu5_L8 [自剪切移除前 8 个 nt]

GGGTAAAC GGCGGG AGTAA CTATGACTCTCTTAGCA AAGAGAATCATAGAA CGTCAGCA
GCCTCCTCGC GGC CCCGCC GGTA ACACAGAGGAACACCCTGTGG CGAA TGCTGACG A TTCTAGA

133. HDV-28(24)gu5_L8 [自剪切移除前 8 个 nt]

GGGTAAAC GGCGGG AGTAA CTATGACTCTCTTGAAA AAGAGAATCATAGAA CGTCAGCA

GCCTCCTCGC GGC CCCGCC GGTA ACACAGAGGAACACCCTGTGG CGAA TGCTGACG A TTCTAGA

134. HDV-28(23)gu5_L8 [自剪切移除前 8 个 nt]

GGGTAAAC GGC GGG AGTAA CTATGACTCTCTGAAA AGAGAATCATAGAA CGTCAGCA
GCCTCCTCGC GGC CCCGCC GGTA ACACAGAGGAACACCCTGTGG CGAA TGCTGACG A TTCTAGA

135. HDV-28(30)gu5_L8 [自剪切移除前 8 个 nt]

GGGTAAAC GGC GGG AGTAA CTATGACTCTCTTAAGG TA AAGAGAATCATAGAA CGTCAGCA
GCCTCCTCGC GGC CCCGCC GGTA ACACAGAGGAACACCCTGTGG CGAA TGCTGACG A TTCTAGA

136. HDV-28(32)gu5_L8 [自剪切移除前 8 个 nt]

GGGTAAAC GGC GGG AGTAA CTATGACTCTCTTAAGG TAGC TAAGAGAATCATAGAA CGTCAGCA
GCCTCCTCGC GGC CCCGCC GGTA ACACAGAGGAACACCCTGTGG CGAA TGCTGACG A TTCTAGA

137. HDV-28(28)_gu5b_L8 [自剪切移除前 8 个 nt]

[1125] GGGTAAAC GGC GGG AGTAA CTATGACTCTCTTAAGG AAAAGAGAATCATAGAA CGTCAGCA
GCCTCCTCGC GGC CCCGCC GGTA ACACAGAGGAACACCCTGTGG CGAA TGCTGACG A TCTAGA

138. HDV-36(36)

GGGTAAA CGGCGGGAGTAACTATGAC TCTCTTAAGG AAAAGAGAATCATAG AA
CGTCAGCA GCCTCCTCGC GGC TTTACCC GGTA ACACAGAGGAACACCCTGTGG CGAA TGCTGACG
A TTCTAGA

139. HDV-24(24)

GGAGTAA CTATGACTCTCTTAAGG AAGAGAATCATCAA CGTCAGCA GCCTCCTCGC GGC TTA
GGTA ACACAGAGGAACACCCTGTGG CGAA TGCTGACG A TTCTAGA

140. HDV-21(21)

GTAAC TA GACTCTCTTAAGG AAGAGAGTAAA CGTCAGCA GCCTCCTCGC GGC TAGTTAC GGTA
ACACAGAGGAACACCCTGTGG CGAA TGCTGACG A TTCTAGA

141. HDV-13(13)

GACTCTC TTAAGG CGTCAGCA GCCTCCTCGC GGC TAGTTAC GGTA
ACACAGAGGAACACCCTGTGG CGAA TGCTGACG A TTCTAGA

142. HDVgRZ

GGCCGGC AT GGTCCCA GCCTCCTCGCT GGC GCCGGCT GG GCA ACATTCCGAGGGGA CCG
TCCCCTCGGTAATGG CGAA TGGGACC TCTAGA

143. HDV-28(3co)

GGGTAAAC GCGGG AGTAA CTATGACTCTCTTAACA CCGAAC TGTTA GAGAATCATAGAA
CGTCAGCA GCCTCCTCGC GGC CCCGCC GGTA ACACAGAGGAACACCCTGTGG CGAA TGCTGACG
A TTCTAGA

144. HDV-28(26)_gu6

GCGGG AGTAA CTATGACTCTCTTAA AAAAGAGAATCATAGAA CGTCAGCG GCCTCCACGC GGC
CCCGCC GGA ACGCAGAGGAACACCCTGCGG CGAA CGCTGACG A TCTAGA

[1126]

145. TriCasB_5RZmin+down

gactctc ttaAGG tgctgagt acgtccgcgtgt gagagtc ggta actgtcccaggatggtctgggatagg ctaa actcagca
ggggaaagtgtagggcctgccaccctacactttattggtatggcattcgatacccctaacgaagcctcggacttgaggagcacgggtccctcctcgtatta
gaccaggaaccaactgtctgacaacccattggacctatgggagcggaccatgctatggacatggattccgaagacgaagcggggcacacggacccccgcc
gatagtgtctacttaacgtcaggcgaacccctgaaatcatcttg

146. HDV-28gu6_L8 [自剪切移除前 8 个 nt]

GGGTAAAC GCGGG AGTAA CTATGACTCTCTT AA AAAAGAGAATCATAGAA CGTCAGCG
GCCTCCACGC GGC CCCGCC GGA ACGCAGAGGAACACCCTGCGG CGAA CGCTGACG C TCTAGA

147. HDV-28gu7_NL

GCGGG AGTAA CTATGACTCTCTT AA AAAAGAGAATCATAGAC CGTCAGCG GCCTCCACGC GGC
CCCGCC GGA ACGCAGAGGAACACCCTGCGG CGAA CGCTGACG TCTAGA

148. HDV-28ac6_L8 [自剪切移除前 8 个 nt]

GGGTAAAC GCGGG AGTAA CTATGACTCTCTT AA AAAAGAGAATCATAGAA CGTCAGCG

GCCCCACAC GGC CCCGCC GGA ACGCAGAGGAACACCCTGCGG CGAA CGCTGACG A TTCTAGA

149. HDV-28ac7_L8 [自剪切移除前 8 个 nt]

GGGTAAAC GGCGGG AGTAA CTATGACTCTCTT AA AAAAGAGAATCATAGAC CGTCAGCG
GCCCCACAC GGC CCCGCC GGA ACGCAGAGGAACACCCTGCGG CGAA CGCTGACG TCTAGA

150. HDV-28ac7_NL

GGCGGG AGTAA CTATGACTCTCTT AA AAAAGAGAATCATAGAC CGTCAGCG GCCCCACAC GGC
CCCGCC GGA ACGCAGAGGAACACCCTGCGG CGAA CGCTGACG TCTAGA

151. HDV-28gu8

GGCGGG AGTAA CTATGACTCTCTT AA CAAAGAGAATCATAGAA CGTCAGCG GCCTCCACGC GGC
CCCGCC GGA ACGCAGAGGAACACCCTGCGG CGAA CGCTGACG C TCTAGA

152. HDV-28ac8

[1127]

GGCGGG AGTAA CTATGACTCTCTT AA CAAAGAGAATCATAGAA CGTCAGCG GCCCCACAC GGC
CCCGCC GGA ACGCAGAGGAACACCCTGCGG CGAA CGCTGACG G TCTAGA

153. HDV-28ac8_L8 [自剪切移除前 8 个 nt]

GGGTAAAC GGCGGG AGTAA CTATGACTCTCTT AA CAAAGAGAATCATAGAA CGTCAGCG
GCCCCACAC GGC CCCGCC GGA ACGCAGAGGAACACCCTGCGG CGAA CGCTGACG G TCTAGA

154. HDV-28ac9_L8 [自剪切移除前 8 个 nt]

GGGTAAAC GGCGGG AGTAA CTATGACTCTCTT AA CAAAGAGAATCATAGAA CGTCTGCG
GCCCCACAC GGC CCCGCC GGA ACGCAGAGGAACACCCTGCGG CGAA CGCAGACG G TCTAGA

200. OrLa 3UTR

TGAGGGGGACAGCTGGGAGTCTCGGCATGATTACAAATCTTGCGCTGCACTCGGATGTCGTCCCCGTG
ACGGACACATTAATCCGAAAGCGAGTGGTGACTCGCCTCAAG

201. TriCasB 3UTR

taaaatctcctgaccaactagctcactgactaattttaactgtcctgtcttactgttttacacgtgctctgtggcggggccatttacaccccgctgcaacacaacctgtaa
atacttgtagtctgttttatgtcctaatttattttaaacagatcttggccatggtctcggccaaccaattaaagtcagtgatgcgagtcgcaatgcggagcaagagac
ctaggcgtgtatttattgtgcatgcggcggcgagccggtcatctgtatggggagcaatggccggcgatacctccacgtggttccctgtgggtgcccgtcg
aggacggtaaccacgaaactcgttaaagtccttctacgagaaggaactccgggttaagattttccaagcctgtacacgtgattcccttgaacaagcaaagtgtgg
ttccctcgagaggggccagggtcaggagttcgaatagtggtgctgaagagttcatgctgggctacagtgtaggacgaagagtggttagtgatcgaaaaacacgtg
aatagctacccccgcctggcaccactagacaacaacaagggtacgacgtcttctgtcgaaagttcggcgcacaccgtaaaagg

202. ZoAl 3'UTR

taggtagtcacattgcactttctgtaacttgcactgggtgtggatgtgggcctggggtgtgggttatgggtatatatgtgggataattctgtgggaatgtccattcactg
tatgcctatcttttaataaaaaagacggtagctaggttcggaagcagccacaagccaatagccagttaggtagctcatagtggttaggtgacaggaacctttgactcag
aacgcgtccattaacatctagaacggaccaaaacttcggacatgcacggattaacggattgtccaaggtggacgggccacctttacttaacccggaagggaacata
tatagtatatgtgttcgtaata

203. TaGu 3'UTR

[1128] taatt cagggtattt agatgcttag ttttgtacc tttctgttt tgtttaggat ttgatagtg ttagtattt tatattttg tacgattgca
taatgttctt ttttatacag ttctgttta ataaaataga cgatagctag agacgttagg gcagccacaa gccagttagg tagcggatag taggtaggaa
cagactttta ctatttcata acgcgtcaat taccacgtga ttggaccaa ttcacgggat ttgtccaagg tggacgggcc acctttactt
aaccgggaaa aggaacatat ataattatg tgtgttcgat aaa

204. GeFo 3'UTR

tagg gtagataatc ttgtatagt ggggggggat ctcatgtacc gggtttctt tatttgattt tcaataaaac agacggtagc taggttcgca
aggcagccac aagccaaaga taggtagggt gctcatagtg agtagggaca gtgccttttg attcacaacg cgtcaatacc atctgacacg
gataccctta ccggacttgt catgatctcc cagactgtc caaggtggac gggccacctt tacttaacct ggaaaaggaa catatattaa
ttatatgtgt tcgaaaa

205. TiGu 3'UTR

tagggg gcttggcatt tctcattgcc tgctcctgaa aggatatggg tctgcgtcg cgtggtaggc agaccattc gtccgagtag ggggcttggc
agntccatt gcctgtgcc gaaaggacgt gggtcatctg gtctgtctgc ctacacctt ctagacttgt aacatctagt ctgtcaaca
gatcaaaatt cttcacacag acgaccgagc ttgtcagtc ttctgtacc cgcagaattt tgctctgtct ctccttggc tgtgtcctgg acgtgggact
attccatctc gtcccaaag cgcgtccaa ttataccgga ttgacaaag cggacggccc gcttataag ccgaaaagg tgcttgttaa aattgcaagg
ttcattaaat ag

206. NaviB 3'UTR

TGACCTGAACAAAACGTGTTGTCTTGTCTTGTCTAAACTATTTATTCGAAATAAGGGGAGGCTAACT
GCCTGCAAGTTGAACGCGAAAGTTAGACCTTCCACCTAAAGCCCCAAAAGTGATCGGGGAATGAATC
CGCGGGTGACCCAGAGTTGGGTAAACCTTGAAACGTTGGAGAAGCGGAAGAGAGTCCC GCCACC
GAGCATCGAGTGCTGCGGCGCCCGAATGAAACCGATCGCGGATGGTGCAAGTCGTAGGACGGGGCA
CGACCTAAGCCTCTGTACGGCGGCGAAGCCAGGAATCACCATGCAAAGGTGTGAACTGGGGCGGAT
ACCTCCACGGGGTTTCCCTGGGCATCGCGCGAGCGATGGCCAAAAGTCCGCTTCTCAGCTACAAAAC
AAAAATGGTATGAGACTTCGTAAACACTAATTTTTCCGAGCCTAGCAGGCTCCCTTGACAACGCTTATG
AATCTGGAAAAGGACACAAAGTGGAAGGCGCTGATGGTGACAAAAGTCAGTTGAGACTTGATA
TCAGTTGTTTTGACTAAGAATTTTATTATCGTTGACTTTTAAATATTTTATTATTGACTGTTAATACTGA
CTTGGGACCAAGTCATCTCTGTTACCCGGTACCGGTTCTGTTCATCAAACCGGAAAGTCCGTCCACG
TAATGTGGTAGACGCAGGAG

207. PuPu 3'UTR

TAG G GTTCTCCAC CCTCCGGCTG ACGAACAGCT GGTAACGGG GGGGCGGTGG
GGTGCCTCTC CAGCCGACTG ATACAGGAGT AAGGGACGGT GGGGTGCGAT CCAGGAAGCG
CAGCACCGCG ATGCCGAAAC TGATGTGCAG TATAACACAG AAAGCCTAAA GGGCCAAAAG

[1129]

208. Lipo 3'UTR

TAAATTTGTCTCTTTCCCAATGATGTCTACTAGCACGCTGCCGAAGCTAGATAGATTGAGGAATCTG
CGTAATCTGTAATGATTACGCCTCATGGGCATCTATCGGTAGCGTCGACCCTGACGTTAAATTGGGT
AATAAGAAATAT

209. BoMo 3'UTR

TGAGCCTTGACAGTAGTCCAGCGGTAAGGGGTGTAGATCAGGCCCCGTCTGTTTCTCCCCGGAGCTCG
CTCCCTTGGCTTCCCTTATATATTTTAACATCAGAAACAGACATTAAACATCTACTGATCCAATTCGCC
GGCGTACGGCCACGATCGGGAGGGTGGGAATCTCGGGGTCTTCCGATCCTAATCCATGATGATTACG
ACCTGAGTCACTAAAGACGATGGCATGATGATCCGGCGATGAAAA

210. DroSi 3'UTR

TAG CTAAAACGTT TGGTTCAAAA CATTTGCTTG CTGTCTTGGC ATAACATCAA TAAAGGCATA
AACATCGCAA AATAATGGTT ATATATAAAT GGCTATGAGG ATGGTTTTAG TACGTAGGCG

TTGCGGAACT TCGGTTTCAGA TAGAGCAATG AATCGTGCAT GCTAGGAAAA CTGACCACAC
GCAGTGTGG CAGCCCTAGT ATCTTTCGAT AGATTTCAT ACCTCCGCGA TCAAAAAAAA
AAAAAAAAAAAA

211. GaAc 3'UTR

[1130] GGAGGGGAGTAGGTCTCTACTCTGACCCGAAGGGCCCCCGTTTCAGACCTGATTCTAGGCTACCTG
TGCCTAATTGGGGGGTCCCAAAGAGATGTTGTCTGTTGTAGAAAGGGTTTGCGCCACTGACTGCACG
GAAGGGTGGGCCTCGACAGGTAGGGGTACATGACTCCGTGCTGCTCAGCAGACCCGCGCCTCTGAG
ACCGGGTAGGGCTACTTGAACAAGCGACGCCCTGGTGTATGTCCGTATCCTAACCTGGTTTGGGAAAG
CCGATACCGCAATGCCCGCCACAGGTGTCGCGCACCCACGGGATGACGTATGGGCCCCGGGGGAC
CTCATGGATACTCCACTGGACTTGACACAATCCTGGTGTACTGGATGCAGCGACGTTGGTGACATAAGC
AATCGCTAAGTCGGGGTAGGGGAGGTGGGGACCTCGGCACGGCTGTAGGAACGGGTGTATGGGCTCC
GGCAGCCGTCGTCACTCCCATACAACACAGGGGCTGCATCCTGGTGGCCGGTGCTAGTTGGTTCTGG
AAGCCCGCCCCGGGTGGTTCGAGAAGCAGGGTGCGCCAGGGTAGGTTTGGTATATCTGGGTCCGG
TGCGATACCTATCGATGGGCAGCGAGGGCCGCCTCGTGACGCGCTGTGTGGAGCTGGAGCCGGCCTG
GGTATGAACAGTTCTTGC GGATGTGGCGTAGCTAGATAGTACCCGTGGTTGTGGGCGTGGTGTGACCC
AAATGTTGTCTGTGTGCACATAGGCCAAGGGTTACGTGGGTGGCAGTCAGAAGCACCCGCACCTGG
AAGTGATTGCCCCGGGATCCCGGCTCTCTGTGAAGAGCTACCTTGAGGAAAGGTGTTCCGCTGGAAC
TCAAGACCCTACAGTAGGGGATATCAACTGGCTTTGAGGTGCTGTGATTCCGGAACCAGGGCGAGGG
CGAGTACTTAGAGCATGTCCAAAAGCCCGGGGAACGTTCCGGGGGCCTGCTTGGGTGCTTGGACCCA
CATCCGTAAACGATGGATCTCGCGTCGGCGCTCGGGAGAACTTCCCGCATGAACGCTGATTGCATGT
GAGAACGCCCCACGGCGGCGGGGCAGGCGCTCCCCCTGGGTGTAAGGCTCGGGGGGGTCACGGCT
CCGCTCTAAAAG

212. LeCoB 3'UTR

Tgataatcgctccatcctgcaactaatatgaatgcaaactgttaagtacattagtatactacactgatactaccctggatattttgacctatactacactggatcta
cctgacatatatttatctaaccacctacatgatgactcccgaggaaactctcacttaccttattaccacttggtctttttctcgttccttattactttgtcctttggttag
ggttctctggtttttggaacggcttccttagccggaattttgtctgatgtatcttctgtctcctgaaatatacgaccaggttgcgtcatttaggctctgggaaa

213. TriCan 3'UTR

Tagatgactgcctacccttttgcgccgaagaactactgaagactattgaactaaccaactgtgtaagtaagaactaatgcctcttttccctgcatgtatccctgatc
agtgactattttctttctgtgcgccctttgttagttatttcttttaatactagaataatcttttctcctgtaattgctttttgggtgtaggaacggctacctaagctgg

aagtgggaagtgttttcaatgtactttgtgattatagaataatgacccgaggtgcatgtttggcatttcctcgagaaa

214. CiIn 3UTR

tgacagtaatatgaaacatcacatctgaccggcacagaatcaccatgccgtaatgcacccaactaaggattccaatgggtaaaaaaaaaaaaaaaaaaaaaa
aaaaaaaaaaaaa

215. DrMe 3UTR

tagctaaatcgtttgggtcaaacatttgctgtcttggcataacatcaataaaggcataaacatcgcaaaataat ggtataatt aaatggctat gaggatggtt
ttagctagtaggcgttgcggaacttcggtt catatagagc aatgaatcgt gcatgctagg aaaactgacc acacacagtg ttggcagacc tagtatcttt
cgaagatttc catacctccg cgatcaaaaa aaaaaaaaaaaaaaaaaa

216. DrNa 3UTR

TAGCACCGAGGTTTACTTTCTCATTGAAAGTCAACTACACTGGCCAACATCTATTGTGTACAAAGTAA
AGGGAAAATAGTCATATGACGTGCCATCCAAATATGGCATTGTTGTGTGGGTAAATTCATTTCCTAC
AAAAAAAAAAAAAAAAAAAAAAAAA

[1131]

217. DrMerc 3UTR

TAGCAGCTCTGAAAGCGTATCCTACAAACGCTGCTAACTTGACACAACATCTTAAAAAGTTATAAAGAA
AGCACACGGGTGTCTTAGCTTATCAGATATCTCTGGTATAGCGAGTCATATAGCGCAAC AAAAAAAAAA
AAAAAAAAAAAAAAAAA

218. AdVa 3UTR

tgaac tagtctcctt ctctattag tcagtctaataaattttct tacattctac atctagtcc attattaaat tggatgatc agtgctatct ctgctacact
caatgcttaa tcgtatgta ttgacagtct gacactgat tactcttacg acatatgcac tgttgcttc agagaaacca ctgtcatat agtgaagtc ctcagtttc
tttgatata ttcttcttc attctcgtt ctcttttct actgtgtct tttatcagt ttttggtga aaaatgaga ataaataaag t

219. HyMaA 3UTR

tga taaatgcaa aagttgcttg ggctaaatga tacgtacgct agaaaaagcg acttgctgca cggatgacgg ttcacagag cccgatatgt gcatgtcaag
gcggcagggga gaatcactag
tgtagctgtt cttccatta cgacttacgc ggtaacgtg gcacgataga ttacaccag
gaaataatac gtgaagggtt ccacatata ctggagtta gatctatgag gaaacattt
gtaataagtc agtctggtaa cctggcgccg ctgttgagtc aaattaacta tgtcaatact

cattaagtta tcgactttga tatggcatgg ggtgattccg cgttatatca aagtcaaaca
tgatgattgc aatgagaaac taccacgctt ggtcacgttt gtgaggagaa catctcattc
aagcctcccg gatgcggca cccgctgaca tcttctggct tatgaaaatt ttcattaatt
tttgaagtc atgggaggct tgaaagc

220. DroSi3v2

GCTTGC TGCCTT GGCATAACATCAATA AAGGCA TAAAC TTCCC AAAA G AGGATGC TTTTAG
TACGTAGGCG TTGCGGAACT TCGGTTTCTAGA TAGAGCAATGAATCGTGCATGC TAGG AACA CCTAGT
ATCTTTGA AGATTT GCATCCT CC GGGAA CAAAAA AAAAA AAAAA AAAAA

221. DroSi3v3

GCTTGC TGCCTT GGCATAACATCAATA AAGGCA TAAAC TTCCC AAAA G AGGATGC TTTTAG
TACGTAGGCG TTGCGGAACT TCGGTTTCTAGA TAGAGCAATGAATCGTGCATGC TAGG AACA CCTAGT
ATCTTTGA AGATTT GCATCCT CC GGGAA CAAAAG AAAAG AAAAG AAAAA

[1132]

222. GEFO217

GACGGTAGC TAGGTTTCGCA AGGCAGCCAC AAGCCAAAGA TAGGTAGGGT GCTCATAGTG
AGTAGGGACA GTGCCTTTGT ATTCACAACG CGTCAATACC ATCTGACACG GATACCCTTA
CCGGACTTGT CATGATCTCC CAGACTTGTC CAAGGTGGAC GGGCCACCTT TACTTAACCC
GGAAAAGGAA CATATATTAA TTATATGTGT TCGGAAAA

223. GEFO98

CCGGACTTGT CATGATCTCC CAGACTTGTC CAAGGTGGAC GGGCCACCTT TACTTAACCC
GGAAAAGGAA CATATATTAA TTATATGTGT TCGGAAAA

224. GEFO68

CAAGGTGGAC GGGCCACCTT TACTTAACCC GGAAAAGGAA CATATATTAA TTATATGTGT
TCGGAAAA

250. -28(32)

GGCGGGAGTAACTATGACTCTCTTAAGGTAGC

251. -28(30)

GGCGGGAGTAACTATGACTCTCTTAAGGTA

252. -28(28)

GGCGGGAGTAACTATGACTCTCTTAAGG

253. -28(27)

GGCGGGAGTAACTATGACTCTCTTAAG

254. -28(26)

GGCGGGAGTAACTATGACTCTCTTAA

255. -28(25)

GGCGGGAGTAACTATGACTCTCTTA

[1133] 256. -28(24)

GGCGGGAGTAACTATGACTCTCTT

257. -28(23)

GGCGGGAGTAACTATGACTCTCT

258. -28(22)

GGCGGGAGTAACTATGACTCTC

259. -28(18)

GGCGGGAGTAACTATGAC

260. -28(14)

GGCGGGAGTAACTA

261. -24(24)

GGAGTAACTATGACTCTCTTAAGG

262. -24(22)

GGAGTAACTATGACTCTCTTAA

263. -24(21)

GGAGTAACTATGACTCTCTTA

264. -21(21)

GTA ACTATGACTCTCTTAAGG

265. -21(19)

GTA ACTATGACTCTCTTAA

266. -13(13)

[1134] GTA ACTATGACTCTCTTAAGG

267. -13(11)

GTA ACTATGACTCTCTTAAGG

268. -36(36)

GGGTAAACGGCGGGAGTAACTATGACTCTCTTAAGG

269. -36(34)

GGGTAAACGGCGGGAGTAACTATGACTCTCTTAA

270. -36(33)

GGGTAAACGGCGGGAGTAACTATGACTCTCTTA

271. -28(26)_L8

GGGTAAAC GGCGGG AGTAA CTATGACTCTCTTAA

272. -28(28)_L8

GGGTAAAC GGCGGG AGTAA CTATGACTCTCTTAAGG

273. -28(25)_L8

GGGTAAAC GGCGGG AGTAA CTATGACTCTCTTA

274. -28(24)_L8

GGGTAAAC GGCGGG AGTAA CTATGACTCTCTT

275. BoMo/OrLa R2 上游 R4

UUAA

276. R25

GGGAGTAACTATGACTCTCTTAAG

[1135]

278. OrLa_下游

TGGGGTCCGGCGAGAGCGCTGGTCTACGGACCAGGGGTGGCTGTGGGCAGGCTGCTCCTCAGGCCA
GTTGATTAGTTACGCATGGGCTGTACCTCCACGTGGTCCCGCTGGTAACGACTTGTCGGCTAAATCAG
CCCGCCACCATCTGGGATATGGTTGACCGTCTAACCCCACTCAGGTCACAAACAAAATGGGAAC
AGATACAGTGTATGTCGGCCAGGACTACCCCTTCTGGCTTATCAAAACGGGTACCAGCACGGT

279. TriCasA 下游

GAGCGTTTTTCGAAGTGGACTGTGTGACTGCGTTCCCCCTTAGTTGCTATATCCGCTTCGATTAAACAT
CTCACCTCGACGTATAAGATCATT

280. R4

UAGC

281. R4+1

GUAGC

282. R3

UAG

283. R5

UAGC

284. R10

UAGCCAAAUG

285. R20

UAGCCAAAUGCCUCGUCAUC

286. R3

UAG

[1136]

287. R8

UAGCCAAA

288. R12

UAGCCAAAUGCC

289. R16

UAGCCAAAUGCCUCGU

290. 沉淀回收对照

GAAATTAATACGACTCACTATAGGGAGAACTCACTATTTGTTTTTCGCGCCCAGTTGCAAAAAGTGTCG
GCCACCATGGGAACAGATACAGTGTATGTCGG

291. R2 64 bp 靶位点退火的寡核苷酸

顶部:

CAAGCGCGGGTAAACGGCGGGAGTAACTATGACTCTCTTAAGGTAGCCAAATGCCTCGTCATCT/3Cy5

Sp/

底部:

AGATGACGAGGCATTTGGCTACCTTAAGAGAGTCATAGTTACTCCCGCCGTTTACCCGCGCTTG/3Cy5S

p/

292. R2 84 bp 靶位点退火的寡核苷酸

顶部:

CAAGCGCGGGTAAACGGCGGGAGTAACTATGACTCTCTTAAGGTAGCCAAATGCCTCGTCATCTAATT

AGTGACGCGCATGAAT/3Phos/

底部:

ATTCATGCGCGTCACTAATTAGATGACGAGGCATTTGGCTACCTTAAGAGAGTCATAGTTACTCCCGCC

GTTTACCCGCGCTTG/3Phos/

293. R9 靶位点退火的寡核苷酸

顶部:

[1137]

CAAATCGATCGTCAAACATGCGTTAAGGGGCGAAAGACTAATCGAACCATCTAGTAGCTGGTTCCGTC

CGAAGTT

底部:

AACTTCGGACGGAACCAGCTACTAGATGGTTCGATTAGTCTTTCGCCCCCTTAACGCATGTTTGACGATC

GATTTG

294. HoSa R8 靶位点: [阴性对照 TPRT]

顶部:

TTTGGGTTCCGGGGGGAGTATGGTTGCAAAGCTGAAACTTAAAGGAATTGACGGAAAGGAGTGG/3Ph

os/

底部:

CCACTCCTTTCCGTCAATTCTTTAAGTTTCAGCTTTGCAACCATACTCCCCCGGAACCCAAA/3Cy5S

p/

295. HyMa R8 靶位点:

顶部:

TTTGATTCCGGGGTAGTATGGTTGCAAAGCTGAAACTTAAACGGAAGGGCACCACCAGGAGT/3Ph
os/

底部:

ACTCCTGGTGGTGCCCTTCCGTTTAAAGTTTCAGCTTTGCAACCATACTACCCCCGGAATCCAAA/3Phos/

296. R2-tsrc 靶位点

顶部:

CAAGCGCGGGTAAACGGCGGGAGTAACTATGACTCTCTGGCTACCTTAAAATGCCTCGTCATCT

底部:

AGATGACGAGGCATTTTAAGGTAGCCAGAGAGTCATAGTTACTCCCGCCGTTTACCCGCGCTTG

300. OrLa3_R4A22

TGAGGGGGACAGCTGGGAGTCTCGGCATGATTACAAATCTTGCGCTGCACTCGGATGTCGTCCCCGTG

ACGGACACATTAATCCGGAAAGCGAGTGGTGACTCGCCTCAAG TAGC

AAAAAAAAAAAAAAAAAAAAA

[1138]

301. OrLa3_R4A40

TGAGGGGGACAGCTGGGAGTCTCGGCATGATTACAAATCTTGCGCTGCACTCGGATGTCGTCCCCGTG

ACGGACACATTAATCCGGAAAGCGAGTGGTGACTCGCCTCAAG TAGC

AAAAAAAAAAAAAAAAAAAAA AAAAAAAAAAAAAAAAAAAAAA

302. OrLa3_R4A60

TGAGGGGGACAGCTGGGAGTCTCGGCATGATTACAAATCTTGCGCTGCACTCGGATGTCGTCCCCGTG

ACGGACACATTAATCCGGAAAGCGAGTGGTGACTCGCCTCAAG TAGC

AAAAAAAAAAAAAAAAAAAAA AAAAAAAAAAAAAAAAAAAAAA AAAAAAAAAAAAAAAAAAAAAA

303. OrLa3_R20A22

TGAGGGGGACAGCTGGGAGTCTCGGCATGATTACAAATCTTGCGCTGCACTCGGATGTCGTCCCCGTG

ACGGACACATTAATCCGGAAAGCGAGTGGTGACTCGCCTCAAG UAGCCAAAUGCCUCGUCAUC

AAAAAAAAAAAAAAAAAAAAA

304. OrLa3_R20A40

TGAGGGGGACAGCTGGGAGTCTCGGCATGATTACAAATCTTGCGCTGCACTCGGATGTCGTCCCCGTG
ACGGACACATTAATCCGGAAGCGAGTGGTGACTCGCCTCAAG UAGCCAAAUGCCUCGUCAUC
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

305. OrLa3_R20A60

TGAGGGGGACAGCTGGGAGTCTCGGCATGATTACAAATCTTGCGCTGCACTCGGATGTCGTCCCCGTG
ACGGACACATTAATCCGGAAGCGAGTGGTGACTCGCCTCAAG UAGCCAAAUGCCUCGUCAUC
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

306. TriCasB3_R4A22

taaaatctcctgaccaactagctcactgactaattttaactgtcctgtcttactgttttacagctgctctgtggcgggcccatttacaccccgtcgcaacacacactgtaa
atactgtgtatgtctgtttatgtcctaatttattttaacagatctggccatggctcggccaaccaattaaagtcagtgatgcgagtcgcaatcgggagcaagagac
ctaggcgtgtattatgtctggcatgcgccgggagccgtcatctgtatgggagcaatggccggcgatacctccacgtggttccctgtgggtggcccgctg
aggacggtaccagcgaaactccgtaagtccttcttacgagaagggaactccggttaagattttccaagcctgtacacgtgattcccttgaacaagcaaagtgtgg
ttccctcgagaggcccaggcaggagttcgcaatagtggtgctgcaagagttcatgtgggtacagtgtaggacgaagagtggttagtgatcgaaaatcacgtg
aatagctacccccgcctggcaccactagacaacaacaagggtacgacagctcttctgtcgaaagttcggcgccacacccgtaaaagg
TAGCAAAAAAAAAAAAAAAAAAAAAAAAAA

[1139]

307. ZoAl3_R4A22

taggtagtcacattgcactttctgtaactgcactgggtgtgggatgtggcctggggtgtgggttatgggtatatgtgggatattctgtgggaatgtccattcactg
tatgcctatcttttaataaaaagacggtagctaggttcggaagcagccacaagccaatagccagttaggtagctcatagtggttagtgtagcaggaacctttgactcag
aacgcgtccattaacatctagaacggaccaaacttcggacatgcaccgattaacggatttgcgaaggtggacgggccaccttacttaacccggaaagggaacata
tatagttatatgtgttcgtaata TAGC AAAAAAAAAAAAAAAAAAAAAAAAAA

308. ZoAl3_R4A40

taggtagtcacattgcactttctgtaactgcactgggtgtgggatgtggcctggggtgtgggttatgggtatatgtgggatattctgtgggaatgtccattcactg
tatgcctatcttttaataaaaagacggtagctaggttcggaagcagccacaagccaatagccagttaggtagctcatagtggttagtgtagcaggaacctttgactcag
aacgcgtccattaacatctagaacggaccaaacttcggacatgcaccgattaacggatttgcgaaggtggacgggccaccttacttaacccggaaagggaacata
tatagttatatgtgttcgtaata TAGC AAAAAAAAAAAAAAAAAAAAAAAAAA AAAAAAAAAAAAAAAAAAAAAA

309. ZoAl3_R4A60

taggtagtcacattgcactttctgtaactgcactgggtgtgggatgtggcctggggtgtgggttatgggtatatgtgggatattctgtgggaatgtccattcactg

tatgcctatcttttaataaaaagacggtagctaggttcggaagcagccacaagccaatagccagttaggtagctcatagtggttaggtgacaggaacctttgactcag
aacgcgtccattaacatctagaacggaccaaaacttcggacatgcaccgattaaccggatttgcgaaggtggacgggccacctttacttaacccggaaagggaacata
tatagttatatgtgttcgtaata TAGC AAAAAAAAAAAAAAAAAAAAAA AAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAA

310. ZoAl3_R10A22

taggtagtcacattgcactttctgaacttcactgggtgtgggatgtggcctggggtgtgggttatgggtatatatgtgggatattctgtgggaatgtccattcactg
tatgcctatcttttaataaaaagacggtagctaggttcggaagcagccacaagccaatagccagttaggtagctcatagtggttaggtgacaggaacctttgactcag
aacgcgtccattaacatctagaacggaccaaaacttcggacatgcaccgattaaccggatttgcgaaggtggacgggccacctttacttaacccggaaagggaacata
tatagttatatgtgttcgtaata UAGCCAAAUG AAAAAAAAAAAAAAAAAAAAAA AAAAAAAAAAAAAAAAAAAAAA

311. ZoAl3_R10A40

taggtagtcacattgcactttctgaacttcactgggtgtgggatgtggcctggggtgtgggttatgggtatatatgtgggatattctgtgggaatgtccattcactg
tatgcctatcttttaataaaaagacggtagctaggttcggaagcagccacaagccaatagccagttaggtagctcatagtggttaggtgacaggaacctttgactcag
aacgcgtccattaacatctagaacggaccaaaacttcggacatgcaccgattaaccggatttgcgaaggtggacgggccacctttacttaacccggaaagggaacata
tatagttatatgtgttcgtaata UAGCCAAAUG AAAAAAAAAAAAAAAAAAAAAA AAAAAAAAAAAAAAAAAAAAAA

[1140]

312. ZoAl3_R10A60

taggtagtcacattgcactttctgaacttcactgggtgtgggatgtggcctggggtgtgggttatgggtatatatgtgggatattctgtgggaatgtccattcactg
tatgcctatcttttaataaaaagacggtagctaggttcggaagcagccacaagccaatagccagttaggtagctcatagtggttaggtgacaggaacctttgactcag
aacgcgtccattaacatctagaacggaccaaaacttcggacatgcaccgattaaccggatttgcgaaggtggacgggccacctttacttaacccggaaagggaacata
tatagttatatgtgttcgtaata UAGCCAAAUG AAAAAAAAAAAAAAAAAAAAAA AAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAA

313. GeFo3_R4A22

tagg gtagataatc tttgtatagt ggggggggat ctcatgtacc gggtttcttt tatttgattt tcaataaac agacggtagc taggttcgca
aggcagccac aagccaaaga taggtagggt gtcatagtg agtagggaca gtgccttttg atcacaacg cgtcaatacc atctgacacg
gataccctta ccggacttgt catgatctcc cagacttgtc caaggtggac ggccacctt tacttaaccc ggaaaaggaa catatatata
ttatatgtgt tcggaata TAGC AAAAAAAAAAAAAAAAAAAAAA

315. GeFo3_R4

tagg gtagataatc tttgtatagt ggggggggat ctcatgtacc gggtttcttt tatttgattt tcaataaac agacggtagc taggttcgca

aggcagccac aagccaaaga taggtagggt gctcatagtg agtagggaca gtgccttttg attcacaacg cgtcaatacc atctgacacg
gataccctta ccggacttgt catgatctcc cagacttgc caaggtggac gggccacctt tacttaaccc ggaaaaggaa catatattaa
ttatatgtgt tcgaaaa UAGCCAAAUG

316. GeFo3_R10A15

tagg gtagataatc ttgtatagt ggggggggat ctcatgtacc gggtttcttt tatitgattt tcaataaaac agacggtagc taggttcgca
aggcagccac aagccaaaga taggtagggt gctcatagtg agtagggaca gtgccttttg attcacaacg cgtcaatacc atctgacacg
gataccctta ccggacttgt catgatctcc cagacttgc caaggtggac gggccacctt tacttaaccc ggaaaaggaa catatattaa
ttatatgtgt tcgaaaa UAGCCAAAUG AAAAAAAAAAAAAAAAAAAAAA

317. GeFo3_R10A22

tagg gtagataatc ttgtatagt ggggggggat ctcatgtacc gggtttcttt tatitgattt tcaataaaac agacggtagc taggttcgca
aggcagccac aagccaaaga taggtagggt gctcatagtg agtagggaca gtgccttttg attcacaacg cgtcaatacc atctgacacg
gataccctta ccggacttgt catgatctcc cagacttgc caaggtggac gggccacctt tacttaaccc ggaaaaggaa catatattaa
ttatatgtgt tcgaaaa UAGCCAAAUG AAAAAAAAAAAAAAAAAAAAAA

[1141]

318. GeFo3_R20

tagg gtagataatc ttgtatagt ggggggggat ctcatgtacc gggtttcttt tatitgattt tcaataaaac agacggtagc taggttcgca
aggcagccac aagccaaaga taggtagggt gctcatagtg agtagggaca gtgccttttg attcacaacg cgtcaatacc atctgacacg
gataccctta ccggacttgt catgatctcc cagacttgc caaggtggac gggccacctt tacttaaccc ggaaaaggaa catatattaa
ttatatgtgt tcgaaaa UAGCCAAAUGCCUCGUCAUC

319. GeFo3_R20A15

tagg gtagataatc ttgtatagt ggggggggat ctcatgtacc gggtttcttt tatitgattt tcaataaaac agacggtagc taggttcgca
aggcagccac aagccaaaga taggtagggt gctcatagtg agtagggaca gtgccttttg attcacaacg cgtcaatacc atctgacacg
gataccctta ccggacttgt catgatctcc cagacttgc caaggtggac gggccacctt tacttaaccc ggaaaaggaa catatattaa
ttatatgtgt tcgaaaa UAGCCAAAUGCCUCGUCAUC AAAAAAAAAAAAAAAAAA

320. GeFo3_R20A22

tagg gtagataatc ttgtatagt ggggggggat ctcatgtacc gggtttcttt tatitgattt tcaataaaac agacggtagc taggttcgca
aggcagccac aagccaaaga taggtagggt gctcatagtg agtagggaca gtgccttttg attcacaacg cgtcaatacc atctgacacg
gataccctta ccggacttgt catgatctcc cagacttgc caaggtggac gggccacctt tacttaaccc ggaaaaggaa catatattaa

ttatatgtgt tcggaaaa UAGCCAAAUGCCUCGUCAUC AAAAAAAAAAAAAAAAAAAAAA

321. TaGu3_R4A22

taatt caggttattt agatgcttag ttttgtacc tttctgttt tgtttaggat ttgatagtg ttagtattt tatattttg tacgattgca
taatgttctt tttatacag ttctgttta ataaaataga cgatagctag agacgttagg gcagccacaa gccagttagg tagcgtag taggtaggaa
cagacttta ctattcata acgctcaat taccactga ttggaccaa ttcacgggat ttgtccaagg tggacgggcc accttactt
aaccgggaaa aggaacatat ataattatg tgtgttcgat aaa TAGCAAAAAAAAAAAAAAAAAAAAAA

322. TiGu3_R4A22

tagggg gcttgccatt tctcattgcc tgctcctgaa aggatatggg tctgcgtcg cgtgtaggc agaccattc gtccagtag ggggcttggc
agtntccatt gccgtgccc gaaaggacgt gggcatctg gtctgtctc ctacacctt ctagactgt aacatctagt ctgtcaaca
gatcaaaatt ctacacag acgaccgagc ttgtcagtc ttctgtacc cgcagaatt ttgtcttct ctctttggc tgtgtcctgg acgtgggact
attccatctc gtccaaatg ccgctccaa ttataccgga ttgacaaag cggacggccc gcttataag ccggaaaagg tgcctgttaa aattgaagg
ttcattaat ag TAGC AAAAAAAAAAAAAAAAAAAAAA

[1142]

323. BoMo3_R4A22

TGAGCCTTGACAGTAGTCCAGCGGTAAGGGTGTAGATCAGGCCGCTCTGTTTCTCCCCGGAGCTCG
CTCCCTTGCGTTCCCTTATATATTTTAACATCAGAAACAGACATTAAACATCTACTGATCCAATTTTCGCC
GGCGTACGGCCACGATCGGGAGGGTGGGAATCTCGGGGTCTTCCGATCCTAATCCATGATGATTACG
ACCTGAGTCACTAAAGACGATGGCATGATGATCCGGCGATGAAAA TAGC
AAAAAAAAAAAAAAAAAAAAA

324. DroSi3_R4A22

TAG CTAAAACGTT TGGTTCAAAA CATTTGCTTG CTGTCTTGGC ATAACATCAA TAAAGGCATA
AACATCGCAA AATAATGGTT ATATATAAAT GGCTATGAGG ATGGTTTTAG TACGTAGGCG
TTGCGGAAC TCGGTTTCTAGA TAGAGCAATG AATCGTGCAT GCTAGGAAAA CTGACCACAC
GCAGTGTGG CAGCCCTAGT ATCTTTCGAT AGATTTCCAT ACCTCCGCGA TCAAAAAA
AAAAAAAAA AAAAA TAGC AAAAAAAAAAAAAAAAAAAAAA

325. DroSi3_R20A22

TAG CTAAAACGTT TGGTTCAAAA CATTTGCTTG CTGTCTTGGC ATAACATCAA TAAAGGCATA
AACATCGCAA AATAATGGTT ATATATAAAT GGCTATGAGG ATGGTTTTAG TACGTAGGCG

TTGCGGAACT TCGGTTTCAGA TAGAGCAATG AATCGTGCAT GCTAGGAAAA CTGACCACAC
GCAGTGTGG CAGCCCTAGT ATCTTTCGAT AGATTTCAT ACCTCCGCGA TCAAAAAAAA
AAAAAAAAAAAAAAAA UAGCCAAAUGCCUCGUCAUC AAAAAAAAAAAAAAAAAAAAAA

326. DroSi3_R20

TAG CTAAAACGTT TGGTTCAAAA CATTTGCTTG CTGTCTTGGC ATAACATCAA TAAAGGCATA
AACATCGCAA AATAATGGTT ATATATAAAT GGCTATGAGG ATGGTTTTAG TACGTAGGCG
TTGCGGAACT TCGGTTTCAGA TAGAGCAATG AATCGTGCAT GCTAGGAAAA CTGACCACAC
GCAGTGTGG CAGCCCTAGT ATCTTTCGAT AGATTTCAT ACCTCCGCGA TCAAAAAAAA
AAAAAAAAAAAAAAAA UAGCCAAAUGCCUCGUCAUC

327. DroSi3_R4

TAG CTAAAACGTT TGGTTCAAAA CATTTGCTTG CTGTCTTGGC ATAACATCAA TAAAGGCATA
AACATCGCAA AATAATGGTT ATATATAAAT GGCTATGAGG ATGGTTTTAG TACGTAGGCG
[1143] TTGCGGAACT TCGGTTTCAGA TAGAGCAATG AATCGTGCAT GCTAGGAAAA CTGACCACAC
GCAGTGTGG CAGCCCTAGT ATCTTTCGAT AGATTTCAT ACCTCCGCGA TCAAAAAAAA
AAAAAAAAAAAAAAAA TAGC

328. DroSi3v2_R20

GCTTGC TGCCCTT GGCATAACATCAATA AAGGCA TAAAC TTCCC AAAA G AGGATGC TTTTAG
TACGTAGGCG TTGCGGAACT TCGGTTTCAGA TAGAGCAATGAATCGTGCATGC TAGG AACA CCTAGT
ATCTTTGA AGATTT GCATCCT CC GGGAA CAAAAAAAAAAAAAAAAAAAAA
UAGCCAAAUGCCUCGUCAUC

329. DroSi3v3_R20

GCTTGC TGCCCTT GGCATAACATCAATA AAGGCA TAAAC TTCCC AAAA G AGGATGC TTTTAG
TACGTAGGCG TTGCGGAACT TCGGTTTCAGA TAGAGCAATGAATCGTGCATGC TAGG AACA CCTAGT
ATCTTTGA AGATTT GCATCCT CC GGGAA CAAAAGAAAAGAAAAGAAAAA
UAGCCAAAUGCCUCGUCAUC

330. R25-TCB3-R4

GGG AGT AAC TAT GAC TCT CTT AAG
GTAAATCTCCTGACCAACTAGCTCACTGACTAATTTTAAACTGTCCTGTCTTACTTGTTTTACACGTG
CTCTGTGGCGGGGCCATTTACACCCCGTCGCAACACAACCTGTAAATACTTGTGTATGTCTGTTTATGT
CCTAATTTATTATTTTAAACAGATCTTGGCCATGGTCTCGGCCAACCAATTAAAGTCAGTGATGCGAGT
CGCAATGCGGAGCAAGAGACCTAGGCGTGTATTTATTGCTGGCATGCGGCGCCGGAGCCGGTCATCTG
CTATGGGGAGCAATGGCCGGGCGGATACCTCCACGTGGTTCCTGTGGGTGGCCGTCGAGGACGGT
AACCAGCGAAACTCCGTAAAGTCCTTCTTACGAGAAGGAACTCCGGTTAAAGATTTTCCAAGCCTGT
ACACGTGATTCCCTTGGAACAAGCAAAGTGTGGTTCCTCGAGAGGGCCCAGGTCAGGAGTTCGCAA
TAGTGGGCTGCAAGAGTTCATGCTGGGCTACAGTGTGAGGACGAAGAGTGGGTAGTGATCGAAAAT
CACGTGAATAGCTACCCCCGCCTGGCACCCTAGACAACAACAAGGGGTACGACAGCTCTTCTGTC
GAAAGTTCGGGCGCACACCCGTAAAAGGTAGc

331. R25-TCB3-R10

[1144]

GGGAGTAACTATGACTCTCTTAAGGTAAAATCTCCTGACCAACTAGCTCACTGACTAATTTTAAACTGT
CCTGTCTTACTTGTTTTACACGTGCTCTGTGGCGGGGCCATTTACACCCCGTCGCAACACAACCTGTA
AATACTTGTGTATGTCTGTTTATGTCCTAATTTATTATTTTAAACAGATCTTGGCCATGGTCTCGGCCAAC
CAATTAAAGTCAGTGATGCGAGTCGCAATGCGGAGCAAGAGACCTAGGCGTGTATTTATTGCTGGCAT
GCGGCGCCGGAGCCGGTCATCTGCTATGGGGAGCAATGGCCGGGCGGATACCTCCACGTGGTTCCT
GTGGGTGGCCCGTCGAGGACGGTAACCAGCGAAACTCCGTAAAGTCCTTCTTACGAGAAGGAACTCC
GGTTAAAGATTTTTCCAAGCCTGTACACGTGATTCCCTTGGAACAAGCAAAGTGTGGTTCCTCGAGA
GGGCCAGGTCAGGAGTTCGCAATAGTGGGCTGCAAGAGTTCATGCTGGGCTACAGTGTGAGGACGA
AGAGTGGGTAGTGATCGAAAATCACGTGAATAGCTACCCCCGCCTGGCACCCTAGACAACAACA
AGGGGTACGACAGCTCTTCTGTGCAAGTTCGGGCGCACACCCGTAAAAGGTAGGccaaatg

332. R25-TCB3-R4-PA

GGGAGTAACTATGACTCTCTTAAGGTAAAATCTCCTGACCAACTAGCTCACTGACTAATTTTAAACTGT
CCTGTCTTACTTGTTTTACACGTGCTCTGTGGCGGGGCCATTTACACCCCGTCGCAACACAACCTGTA
AATACTTGTGTATGTCTGTTTATGTCCTAATTTATTATTTTAAACAGATCTTGGCCATGGTCTCGGCCAAC
CAATTAAAGTCAGTGATGCGAGTCGCAATGCGGAGCAAGAGACCTAGGCGTGTATTTATTGCTGGCAT
GCGGCGCCGGAGCCGGTCATCTGCTATGGGGAGCAATGGCCGGGCGGATACCTCCACGTGGTTCCT
GTGGGTGGCCCGTCGAGGACGGTAACCAGCGAAACTCCGTAAAGTCCTTCTTACGAGAAGGAACTCC

GGTTAAAGATTTTTC CAAGCCTGTACACGTGATTCCTTGGAAACAAGCAAAGTGTGGTTCCCTCGAGA
GGGCCCAGGTCAGGAGTTCGCAATAGTGGGCTGCAAGAGTTCATGCTGGGCTACAGTGTGAGGACGA
AGAGTGGGTAGTGATCGCAAAATCACGTGAATAGCTACCCCCGCCTGGCACCCTAGACAACAACA
AGGGGTACGACAGCTCTTCTGTGCGAAAGTTCGGGCGCACACCCGTAAAAGGTAGc aaaaa aaaaa aaaaa
aaaaa aa

- 333. R25*-TCB5_TCB3-R4 [5' 12 nt 将通过 RZ 自剪切被移除]

GGGAGTAACTAT

[1145]

GACTCTCTTAAGGGAAGACCCGCCCATGAGGCTTGGAGAGTGTGATCCTGATCAGATCACACTTGA
AAAGTTATGCTGAGTACGTCCGCGTCGTGAGAGTCGGTAACTGTCCCAGGATGGTCTGGGATAGGCTA
AACCTCAGCAGGGGAAAGTTGTAGGGGCTGCCACCCCTACACTTTATTGGTATGGCATTTCGATACCC
CTAACGAAGCCTCGGACTTGGAGGAGCACGGTTCCTCTCCTCGTATTAGACCAGGAACCAACTG
TCCTGACAACCCCATTTGGACCTATGGGAGCGGACCATGCTATGGACATGGATTCCGAAGACGAAGCG
GGGGCACACGGACCCCCCGCCGATAGTGCTCACTTAACGTCAGGCGAACCCCTTGAAATCATCTTGTA
AAATCTCCTGACCAACTAGCTCACTGACTAATTTTAAACTGTCTGTCTTACTTGTTTTACAGTGCTC
TGTGGCGGGGCCATTTACACCCCGTCGCAACACAACCTGTAAATACTTGTGTATGTCTGTTTATGTCCT
AATTTATTATTTTAAACAGATCTTGGCCATGGTCTCGGCCAACCAATTAAGTCAGTGATGCGAGTCGC
AATGCGGAGCAAGAGACCTAGGCGTGTATTATTGCTGGCATGCGGCGCCGGAGCCGGTCACTGCTA
TGGGGAGCAATGGCCGGGCGGATACCTCCACGTGGTTCCTGTGGGTGGCCCGTCGAGGACGGTAAC
CAGCGAAACTCCGTAAAGTCTTCTTACGAGAAGGAACTCCGGTTAAAGATTTTTC CAAGCCTGTACA
CGTGATTCCCTTGGAAACAAGCAAAGTGTGGTTCCTCGAGAGGGCCAGGTCAGGAGTTCGCAATAG
TGGGCTGCAAGAGTTCATGCTGGGCTACAGTGTGAGGACGAAGAGTGGGTAGTGATCGCAAAATCAC
GTGAATAGCTACCCCCGCCTGGCACCCTAGACAACAACAAGGGGTACGACAGCTCTTCTGTGCGAA
AGTTCGGGCGCACACCCGTAAAAGGTAGc

- 334. R25*-TCB5_TCB3-R10 [5' 12 nt 将通过 RZ 自剪切被移除]

GGGAGTAACTA

TGACTCTCTTAAGGGAAGACCCGCCCATGAGGCTTGGAGAGTGTGATCCTGATCAGATCACACTTGA
AAAGTTATGCTGAGTACGTCCGCGTCGTGAGAGTCGGTAACTGTCCCAGGATGGTCTGGGATAGGCTA
AACCTCAGCAGGGGAAAGTTGTAGGGGCTGCCACCCCTACACTTTATTGGTATGGCATTTCGATACCC
CTAACGAAGCCTCGGACTTGGAGGAGCACGGTTCCTCTCCTCGTATTAGACCAGGAACCAACTG
TCCTGACAACCCCATTTGGACCTATGGGAGCGGACCATGCTATGGACATGGATTCCGAAGACGAAGCG

GGGGCACACGGACCCCCCGCCGATAGTGCTCACTTAACGTCAGGCGAACCCCTTGAAATCATCTTGTA
AAATCTCCTGACCAACTAGCTCACTGACTAATTTTAAACTGTCCTGTCTTACTTGTTTTACACGTGCTC
TGTGGCGGGGCCATTTACACCCCGTCGCAACACAACCTGTAAATACTTGTGTATGTCTGTTTATGTCCT
AATTTATTATTTTAAACAGATCTTGGCCATGGTCTCGGCCAACCAATTAAAGTCAGTGATGCGAGTCGC
AATGCGGAGCAAGAGACCTAGGCGTGATTTATTGCTGGCATGCGGCGCCGGAGCCGGTCATCTGCTA
TGGGGAGCAATGGCCGGGCGGATACCTCCACGTGGTTCCCTGTGGGTGGCCCGTCGAGGACGGTAAC
CAGCGAAACTCCGTAAAGTCCTTCTTACGAGAAGGAACTCCGGTTAAAGATTTTCCAAGCCTGTACA
CGTGATTCCCTTGGAACAAGCAAAGTGTGGTTCCTCGAGAGGGCCCAGGTCAGGAGTTCGCAATAG
TGGGCTGCAAGAGTTCATGCTGGGCTACAGTGTGAGGACGAAGAGTGGGTAGTGATCGCAAAATCAC
GTGAATAGCTACCCCCCGCCTGGCACCCTAGACAACAACAAGGGGTACGACAGCTCTTCTGTCGAA
AGTTCGGGCGCACACCCGTAAAAGGTAGc

335. R25*-TCB5_TCB3-R4-PA [5' 12 nt 将通过 RZ 自剪切被移除]

GGGAGTAACTAT

[1146]

GACTCTCTTAAGGGAAGACCCCGCCCATGAGGCTTGGAGAGTGTGATCCTGATCAGATCACACTTGA
AAAGTTATGCTGAGTACGTCCGCGTCGTGAGAGTCGGTAACTGTCCCAGGATGGTCTGGGATAGGCTA
AACCTCAGCAGGGGAAAGTTGTAGGGGCCTGCCACCCCTACACTTTATTGGTATGGCATTGATACCC
CTAACGAAGCCTCGGACTTGGAGGAGCACGGTTCCCTCCTCCTCGTATTAGACCAGGAACCAACTG
TCCTGACAACCCCATTTGGACCTATGGGAGCGGACCATGCTATGGACATGGATTCCGAAGACGAAGCG
GGGGCACACGGACCCCCCGCCGATAGTGCTCACTTAACGTCAGGCGAACCCCTTGAAATCATCTTGTA
AAATCTCCTGACCAACTAGCTCACTGACTAATTTTAAACTGTCCTGTCTTACTTGTTTTACACGTGCTC
TGTGGCGGGGCCATTTACACCCCGTCGCAACACAACCTGTAAATACTTGTGTATGTCTGTTTATGTCCT
AATTTATTATTTTAAACAGATCTTGGCCATGGTCTCGGCCAACCAATTAAAGTCAGTGATGCGAGTCGC
AATGCGGAGCAAGAGACCTAGGCGTGATTTATTGCTGGCATGCGGCGCCGGAGCCGGTCATCTGCTA
TGGGGAGCAATGGCCGGGCGGATACCTCCACGTGGTTCCCTGTGGGTGGCCCGTCGAGGACGGTAAC
CAGCGAAACTCCGTAAAGTCCTTCTTACGAGAAGGAACTCCGGTTAAAGATTTTCCAAGCCTGTACA
CGTGATTCCCTTGGAACAAGCAAAGTGTGGTTCCTCGAGAGGGCCCAGGTCAGGAGTTCGCAATAG
TGGGCTGCAAGAGTTCATGCTGGGCTACAGTGTGAGGACGAAGAGTGGGTAGTGATCGCAAAATCAC
GTGAATAGCTACCCCCCGCCTGGCACCCTAGACAACAACAAGGGGTACGACAGCTCTTCTGTCGAA
AGTTCGGGCGCACACCCGTAAAAGGTAGc aaaaa aaaaa aaaaa aaaaa aa

336. R25*-TCB5_TCB3-PA [5' 12 nt 将通过 RZ 自剪切被移除]

GGGAGTAACTAT

GACTCTCTTAAGGGAAGACCCCGCCCATGAGGCTTGGAGAGTGTGATCCTGATCAGATCACACTTGA
AAAGTTATGCTGAGTACGTCCGCGTCGTGAGAGTCGGTAACTGTCCCAGGATGGTCTGGGATAGGCTA
AACCTCAGCAGGGGAAAGTTGTAGGGGCCTGCCACCCCTACACTTTATTGGTATGGCATTGATAACCC
CTAACGAAGCCTCGGACTTGGAGGAGCACGGTTCCCTCCTCCTCGTATTAGACCAGGAACCAACTG
TCCTGACAACCCCATTTGGACCTATGGGAGCGGACCATGCTATGGACATGGATTCCGAAGACGAAGCG
GGGGCACACGGACCCCCCGCCGATAGTGCTCACTTAACGTCAGGCGAACCCCTTGAAATCATCTTGTA
AAATCTCCTGACCAACTAGCTCACTGACTAATTTTAAACTGTCCTGTCTTACTTGTTTTACACGTGCTC
TGTGGCGGGGCCATTTACACCCCGTCGCAACACAACCTGTAAATACTTGTGTATGTCTGTTTATGTCCT
AATTTATTATTTTAAACAGATCTTGGCCATGGTCTCGGCCAACCAATTAAAGTCAGTGATGCGAGTCGC
AATGCGGAGCAAGAGACCTAGGCGTGTATTTATTGCTGGCATGCGGCGCCGGAGCCGGTCATCTGCTA
TGGGGAGCAATGGCCGGGCGGATACCTCCACGTGGTTCCCTGTGGGTGGCCCGTCGAGGACGGTAAC
CAGCGAAACTCCGTAAAGTCCTTCTTACGAGAAGGAACTCCGGTTAAAGATTTTCCAAGCCTGTACA
CGTGATTCCCTTGGAACAAGCAAAGTGTGGTTCCCTCGAGAGGGCCCAGGTCAGGAGTTCGCAATAG
TGGGCTGCAAGAGTTCATGCTGGGCTACAGTGTGAGGACGAAGAGTGGGTAGTGATCGCAAAATCAC
GTGAATAGCTACCCCCCGCCTGGCACCCTAGACAACAACAAGGGGTACGACAGCTCTTCTGTGCGAA
AGTTCGGGCGCACACCCGTAAAAGG aaaaa aaaaa aaaaa aaaaa aa

[1147]

337. R25*-TCB5_TCB3-R10PA [5' 12 nt 将通过 RZ 自剪切被移除]

GGGAGTAACTAT

GACTCTCTTAAGGGAAGACCCCGCCCATGAGGCTTGGAGAGTGTGATCCTGATCAGATCACACTTGA
AAAGTTATGCTGAGTACGTCCGCGTCGTGAGAGTCGGTAACTGTCCCAGGATGGTCTGGGATAGGCTA
AACCTCAGCAGGGGAAAGTTGTAGGGGCCTGCCACCCCTACACTTTATTGGTATGGCATTGATAACCC
CTAACGAAGCCTCGGACTTGGAGGAGCACGGTTCCCTCCTCCTCGTATTAGACCAGGAACCAACTG
TCCTGACAACCCCATTTGGACCTATGGGAGCGGACCATGCTATGGACATGGATTCCGAAGACGAAGCG
GGGGCACACGGACCCCCCGCCGATAGTGCTCACTTAACGTCAGGCGAACCCCTTGAAATCATCTTGTA
AAATCTCCTGACCAACTAGCTCACTGACTAATTTTAAACTGTCCTGTCTTACTTGTTTTACACGTGCTC
TGTGGCGGGGCCATTTACACCCCGTCGCAACACAACCTGTAAATACTTGTGTATGTCTGTTTATGTCCT
AATTTATTATTTTAAACAGATCTTGGCCATGGTCTCGGCCAACCAATTAAAGTCAGTGATGCGAGTCGC
AATGCGGAGCAAGAGACCTAGGCGTGTATTTATTGCTGGCATGCGGCGCCGGAGCCGGTCATCTGCTA
TGGGGAGCAATGGCCGGGCGGATACCTCCACGTGGTTCCCTGTGGGTGGCCCGTCGAGGACGGTAAC
CAGCGAAACTCCGTAAAGTCCTTCTTACGAGAAGGAACTCCGGTTAAAGATTTTCCAAGCCTGTACA

CGTGATTCCCTTGGAACAAGCAAAGTGTGGTTCCCTCGAGAGGGCCCAGGTCAGGAGTTCGCAATAG
TGGGCTGCAAGAGTTCATGCTGGGCTACAGTGTGAGGACGAAGAGTGGGTAGTGATCGCAAAATCAC
GTGAATAGCTACCCCCCGCTGGCACCCTAGACAACAACAAGGGGTACGACAGCTCTTCTGTCGAA
AGTTCGGGCGCACACCCGTAAAAGG TAGccaaatg aaaaa aaaaa aaaaa aaaaa aa

338. TCB3-R4-PA

TAAATCTCCTGACCAACTAGCTCACTGACTAATTTTAAACTGTCCTGTCTTACTTGTTTTACACGTGC
TCTGTGGCGGGGCCATTACACCCCGTCGCAACACAACCTGTAAATACTTGTGTATGTCTGTTTATGTC
CTAATTTATTATTTTAAACAGATCTTGGCCATGGTCTCGGCCAACCAATTAAAGTCAGTGATGCGAGTC
GCAATGCGGAGCAAGAGACCTAGGCGTGATTTATTGCTGGCATGCGGCGCCGGAGCCGGTCATCTGC
TATGGGGAGCAATGGCCGGGCGGATACCTCCACGTGGTTCCCTGTGGGTGGCCCCGTCGAGGACGGTA
ACCAGCGAAACTCCGTAAAGTCCTTCTTACGAGAAGGAACTCCGGTTAAAGATTTTCCAAGCCTGTA
CACGTGATTCCCTTGGAACAAGCAAAGTGTGGTTCCCTCGAGAGGGCCCAGGTCAGGAGTTCGCAAT
AGTGGGCTGCAAGAGTTCATGCTGGGCTACAGTGTGAGGACGAAGAGTGGGTAGTGATCGCAAAATC
ACGTGAATAGCTACCCCCCGCTGGCACCCTAGACAACAACAAGGGGTACGACAGCTCTTCTGTGCG
AAAGTTCGGGCGCACACCCGTAAAAGGTAGc aaaaa aaaaa aaaaa aaaaa aa

[1148]

339. TCA5_OrLa3 [前 54 个 nt 通过自剪切被移除]

gggagacggucaaccgcuaggagcggugaccgGAAUUCGGCGGGAGUAACUAU
GACUCUCUUAAGGAGUCAUAGAGCCAGAACCUCUGUGGUCCCGCUGGGCACAGGGAUUAAUU
UUUCUGUGGCAAAUUUGACUGGCUUCAGAGAGCGUUUUUCGAAGUGGACUGUGUGACUGCGUUC
CCCCCUUAGUUGCUAUAUCCGCUUCGAUUAACAUCUCACCUCGACGUUAAGAUAUUUCUAGAG
CGGCCGUGAGGGGGACAGCUGGGAGUCUCGGCAUGAUUACAAAUCUUGCGCUGCACUCGGAUG
UCGUCCCCGUGACGGACACAUAUAUCCGGAAAGCGAGUGGUGACUCGCCUCAAGUAGCAAAAAA
AAAAAAAAAAAAAAAAA

340. TCA5_ZoAl3 [前 54 个 nt 通过自剪切被移除]

gggagacggucaaccgcuaggagcggugaccgGAAUUCGGCGGGAGUAACUAU
GACUCUCUUAAGGAGUCAUAGAGCCAGAACCUCUGUGGUCCCGCUGGGCACAGGGAUUAAUU
UUUCUGUGGCAAAUUUGACUGGCUUCAGAGAGCGUUUUUCGAAGUGGACUGUGUGACUGCGUUC
CCCCCUUAGUUGCUAUAUCCGCUUCGAUUAACAUCUCACCUCGACGUUAAGAUAUUUCUAGAG
CGGCCGCUagguagucacauugcacuuucuguaacuugcacugggugugggauugggccuggggugggguuauaggguuauauaugggg

219

220

AAAAAAAAAA

346-349 无序列

350. pUC57-极微质粒

CCAATGATATCGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCG
CGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCTGACGAGCATCACAAAATCGACGCTCAAGTCAG
AGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCT
CTCCTGTTCCGACCCCTGCCGCTTACCGGATACCTGTCCGCTTTCTCCCTTCGGGAAGCGTGGCGCTTT
CTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTGCTCCAAGCTGGGCTGTGTGCAC
GAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAG
ACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGT
GCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGC
TCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCT
GGTAGCGGTGGTTTTTTTGTGTTGCAAGCAGCAGATTACGCGCAGAAAAAAGGATCTCAAGAAGATC
CTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACCTACGTTAAGGGATTTTGGTCATG
[1151] AGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAATAAATGAAGTTTTAAATCAATCTAAAGT
ATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGT
CTATTTGTTTCATCCATAGTTGCCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCA
TCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAA
CCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTA
ATTGTTGCCGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGC GCAACGTTGTTGCCATTGCTA
CAGGCATCGTGGTGTACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGG
CGAGTTACATGATCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAG
AAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCC
ATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCG
ACCGAGTTGCTCTTGCCCGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGC
TCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTG
ATGTAACCCACTCGTGACCCCACTGATCTTCAGCATCTTTTACTTTACCAGCGTTTCTGGGTGAGCA
AAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATA
CTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATG
TATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTC

351. pcDNA3.1 N-DYK 质粒

GACGGATCGGGAGATCTCCCGATCCCCTATGGTGCACCTCTCAGTACAATCTGCTCTGATGCCGCATAGT
TAAGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGC GCGAGCAAAATTTAAGCT
ACAACAAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGC
TTCGCGATGTACGGGCCAGATATACGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTAC
GGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCCGCCTGG
CTGACCGCCCAACGACCCCCGCCAATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAG
GGACTTTCCATTGACGTCAATGGGTGGAGTATTACGGTAAACTGCCCACTTGGCAGTACATCAAGTG
TATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCCGCCTGGCATTATGCCCAG
TACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGA
TGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTC AAGTCTCCAC
CCCATTGACGTCAATGGGAGTTTGT TTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAA
CTCCGCCCCATTGACGCAAATGGGCGGTAGGCGGTGACGGTGGGAGGTCTATATAAGCAGAGCTCTCT
GGCTAACTAGAGAACCCACTGCTTACTGGCTTATCGAAATTAATACGACTCACTATAGGGAGACCCAA
GCTGGCTAGCGTTTAAACTTAAGCTTGGTACCGAGCTCGGATCCGCCACCGAATTCTGCAGATATCCA
[1152] GCACAGTGGCGGCCGCTCGAGTCTAGAGGGCCCCGATTACAAGGATGACGACGATAAGTGATAAACCC
GCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCCTCCCCCGTGCTTTCCT
TGACCCTGGAAGGTGCCACTCCCACTGTCTTTTCTAATAAAATGAGGAAATTGCATCGCATTGTCTG
AGTAGGTGTCAATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGAC
AATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACCAGCTGGGGCT
CTAGGGGGTATCCCCACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAG
CGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTTCGCTTCTTCCCTTCCCTTCTCGCCAC
GTTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACG
GCACCTCGACCCCAAAAACTTGATTAGGGTGATGGTTACGTAGTGGGCCATCGCCCTGATAGACGG
TTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACA
CTCAACCCTATCTCGGTCTATTCTTTGATTATAAGGGATTTTGCCGATTTCGGCCTATTGGTTAAAAA
ATGAGCTGATTTAACA AAAAATTTAACGCGAATTAATTCTGTGGAATGTGTGTCAGTTAGGGTGTGGAA
AGTCCCCAGGCTCCCCAGCAGGCAGAAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCAGGTG
TGGAAGTCCCCAGGCTCCCCAGCAGGCAGAAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACC
ATAGTCCCGCCCCTA ACTCCGCCCCATCCCGCCCCCTAACTCCGCCCAGTTCCGCCCCATCTCCGCCCCAT
GGCTGACTAATTTTTTTTATTATGCAGAGGCCGAGGCCGCCTCTGCCTCTGAGCTATTCCAGAAGTAG

[1153]

TGAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAAAGCTCCCGGGAGCTTGTATATCCATTTTCGG
ATCTGATCAAGAGACAGGATGAGGATCGTTTCGCATGATTGAACAAGATGGATTGCACGCAGGTTCTC
CGGCCGCTTGGGTGGAGAGGCTATTCCGCTATGACTGGGCACAACAGACAATCGGCTGCTCTGATGC
CGCCGTGTTCCGGCTGTCAGCGCAGGGGCGCCCGGTTCTTTTTGTCAAGACCGACCTGTCCGGTGCC
CTGAATGAACTGCAGGACGAGGCAGCGCGGCTATCGTGGCTGGCCACGACGGGCGTTCTTTGCGCAG
CTGTGCTCGACGTTGTCACTGAAGCGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCGGGGCAGGA
TCTCTGTCACTCTACCTTGCTCCTGCCGAGAAAAGTATCCATCATGGCTGATGCAATGCGGCGGCTGCA
TACGCTTGATCCGGCTACCTGCCCATTGACCACCAAGCGAAACATCGCATCGAGCGAGCACGTA CTC
GGATGGAAGCCGGTCTTGTGATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGA
ACTGTTCCGCCAGGCTCAAGGCGCGCATGCCGACGGCGAGGATCTCGTCGTGACCCATGGCGATGCC
TGCTTGCCGAATATCATGGTGAAAATGGCCGCTTTTCTGGATTTCATCGACTGTGGCCGGCTGGGTGT
GGCGGACCGCTATCAGGACATAGCGTTGGCTACCCGTGATATTGCTGAAGAGCTTGCGGCGAATGGG
CTGACCGCTTCCTCGTGCTTTACGGTATCGCCGCTCCCGATTTCGACGCGCATCGCCTTCTATCGCCTTC
TTGACGAGTTCTTCTGAGCGGGACTCTGGGGTTCGAAATGACCGACCAAGCGACGCCAACCTGCCA
TCACGAGATTTGATTCCACCGCCGCTTCTATGAAAGGTTGGGCTTCGGAATCGTTTTCCGGGACGC
CGGCTGGATGATCCTCCAGCGCGGGGATCTCATGCTGGAGTTCTTCGCCACCCCAACTTGTTTATTGC
AGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTACAAATAAAGCATTTTTTTTCACTGCA
TTCTAGTTGTGGTTTGTCCAACTCATCAATGTATCTTATCATGTCTGTATACCGTCGACCTCTAGCTAG
AGCTTGGCGTAATCATGGTCATAGCTGTTTCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAAC
ATACGAGCCGGAAGCATAAAGTGTAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGC
GTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCTGTGCCAGCTGCATTAATGAATCGGCCAAC
GCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCTCTGCTCACTGACTCGCTGCGCTCG
GTCGTTCCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGG
GGATAACGCAGGAAAGAACATGTGAGCAAAAAGGCCAGCAAAAAGGCCAGGAACCGTAAAAAGGCCG
CGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCTGACGAGCATCACAAAATCGACGCTCAAGTCAG
AGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCT
CTCCTGTTCCGACCTGCCGCTTACCGGATACCTGTCCGCTTTCTCCCTTCGGGAAGCGTGCGCTTT
CTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAAGTCGTTCCGCTCCAAGCTGGGCTGTGTGCAC
GAACCCCCGTTTCAGCCCCACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGTAAG
ACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGT
GCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGC
TCTGCTGAAGCCAGTTACCTTCGGA AAAAGAGTTGGTAGCTCTTGATCCGGCAAAACAAACCACCGCT

[1154]

GGTAGCGGTGGTTTTTTTGTGTTGCAAGCAGCAGATTACGCGCAGAAAAAAGGATCTCAAGAAGATC
CTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGAACGAAAACCTCACGTTAAGGGATTTTGGTCATG
AGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAATAATGAAGTTTTAAATCAATCTAAAGT
ATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGT
CTATTTTCGTTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCA
TCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAA
CCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTA
ATTGTTGCCGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTA
CAGGCATCGTGGTGTACGCTCGTCGTTTGGTATGGCTTCATTACGCTCCGGTTCCCAACGATCAAGG
CGAGTTACATGATCCCCATGTTGTGCAAAAAAGCGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAG
AAGTAAGTTGGCCGCAGTGTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCC
ATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCG
ACCGAGTTGCTCTTGCCCGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAGTGC
TCATCATTGGAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCG
ATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTACCAGCGTTTCTGGGTGAGCA
AAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATA
CTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATG
TATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTC

352. 用于 RT 引物延伸测定的 DNA 引物

CAGCACTAGATTTTTGGGGTTGAATG

353. 用于 RT 引物延伸测定的模版 RNA

AUACCCGCUUAAUUCAUUCAGAUUCUGUAAUAGAACUGUCAUUAACCCCAAAAAUCUAGUGCUG
AUUAUAACCUUCACCAAUUAGGUUCAAUAAGUGGUAAUGCGGGACAAAAGACUAUCGACAUUUG
AUACACUAUUUAUCAUUGGAUGUCUUAUUUUUUU

360. 用于序列读段映射的转基因重叠群[TriCasA5_CBhBsi_GFP_GeFo3]

AGTCATAGAGCCAGAACCTCCTCGTGGTCCCGCTGGGCACAGGGATTAATTTTTCTGTGGCAAATTTG
ACTGGCTTCAGAGAGCGTTTTTCGAAGTGGACTGTGTGACTGCGTTCCCCCTTAGTTGCTATATCCG
CTTCGATTAACATCTCACCTCGACGTATAAGATCATTTCTAGAAGGTCGACCAGATGTCCGAGGTCGAC
CAGTTGTCCGTGTGGAATTGTGAGCGCTCACAATCCACACGgttacataacttacggtaaatggcccgctggctgaccgcca

[1155]

acgacccccgccattgacgtcaatagtaacgccaatagggacttccattgacgtcaatgggtggagtatttacggtaaactcccacttggcagtacatcaagtgtat
catatgccaagtacgccccctattgacgtcaatgacggtaaatggccccctggcattgtgccagttacatgacctatgggactttctacttggcagtacatctacgtat
ttagtcatcgctattaccatgggtcagggtgagccccacttctgcttctacttccccatctcccccccccccccccccaatttgtattttatttttaattttttgtgcag
cgatggggcgggggggggggggggcgcgccaggcgggcgggcgggcgaggggcgggcgggcgaggcgaggaggtgcggcgggcagcc
aatcagagcgggcgcgctccgaaagtcttctttatggcgaggcgggcgggcgggcgccctataaaaagcgaagcgcgcgggcgggcgaggcgctgcgacgc
tgccttcgccccgtccccgtccgcccgcctcgcgccgccccggctctgactgaccggttactccacaggtgagcgggcgggcagggcccttctcctc
cgggctgtaattagctgagcaagaggtaagggttaagggtggttgggtgggtggttaattgtaattacctggagcacctgcctgaaatcactttttcagggtgg
cgtacggccaccatggtgagcaaggcgaggagctgtaccggggtggtgccatcctggtcagctggacggcgagcgtaaacggccacaagttcagcgtgtcc
ggcgaggcgaggcgatgccacctacggcaagctgacctgaagttcatctgcaccaccggcaagctgcccgtgccctggccaccctctgaccaccctgac
ctacggcgtgcagtgcttaccgctaccccgaccacatgaagcagcagcacttctcaagtcgccatgccgaaggctacgtccaggagcgcaccatcttctca
aggacgacggcaactacaagacccgcgccgaggtgaagttcaggcgacacccctggtgaaccgcatcagctgaaggcagcactcaaggaggacggcga
catcctggggcacaagctggagtacaactacaacagccacaacgtctatatcatggcgacaagcagaagaacggcatcaagtgaaactcaagatccgccacaac
atcaggagcggcagcgtgcagctcgccgaccactaccagcagaacacccccatcgcgacggccccgtgctgctgcccgacaaccactacctgagcaccagtc
cgccctgagcaagaccccaacgagaagcgcatcacatggtcctgctgagttcgtgacgcgcccgggatcactctcgcatggacgagctgtacaagtaaaa
gcttGatccagacatgataagatacattgatgattggacaaccacaactagaatgcagtgaaaaaatgctttattgtgaaattgtgatgtattgctttattgtaac
cattataagctgcaataacaagtTGTGGAATTGTGAGCGCTCACAAATTCACAGCGCCGCTGAGGTAGATAATCT
TTGTATAGTGGGGGGGATCTCATGTACCGGGTTTCTTTTATTGATTTTCAATAAAACAGACGGTAGC
TAGGTTTCGAAGGCAGCCACAAGCCAAAGATAGGTAGGTTGCTCATAGTGAGTAGGGACAGTGCCTT
TTGATTCACAACGCGTCAATACCATCTGACACGGATACCCCTTACCGGACTTGTCATGATCTCCCAGACT
TGTCCAAGGTGGACGGGCCACCTTTACTTAACCCGGAAGGAACATATATTAATTATATGTGTTTCGGA
AAA

361. 用于序列读段映射的 rDNA 侧翼转基因序列重叠群

GCTGGGCGCGCGCCGCGGCTGGACGAGGCGCGCGCCCCCCCCACGCCCCGGGACCCCCCTCGCGG
CCCTCCCCCGCCCCACCCGCGCGCGCGCTCGCTCCCTCCCCACCCGCGCCCTCTCTCTCTCTCT
CCCCCGCTCCCCGTCTCCCCCTCCCCGGGGGAGCGCCGCTGGGGGCGCGGGGGGAGAAGG
GTCGGGCGGCGAGGGGCCGCGCGCGCGCCGCGGGGCGCGCGGGGCGAGGTCCCCGCGAGGG
GGCCCCGGGGACCCGGGGGGCCGGCGGCGCGGACTCTGGACGCGAGCCGGGCCCTTCCCGTG
GATCGCCCCAGCTGCGGCGGGCGTCGCGGCCGCCCCGGGGAGCCGGCGGCGCGCGCGCGCC
CCCACCCCCACCCACGTCTCGGTCTGCGCGCGCGTCCGCTGGGGCGGGAGCGGTCTGGGCGGCGGC
GGTCGGCGGGCGGCGGGGCGGGCGGTTTCGTCCCCCGCCCTACCCCCCGGCCCGTCCGCCCCC
GTTCCCCCTCCTCCTCGGCGCGCGCGGCGGCGGCGGAGGGGCCGCGGGCGCGGTC

[1156]

CCCCCGCCGGGTCCGCCCCGGGGCCGCGGTTCCGCGCGCGCCTCGCCTCGGCCGGCGCCTAGCAG
CCGACTTAGAACTGGTGCGGACCAGGGGAATCCGACTGTTTAATTAATAAAACAAAGCATCGCGAAGGCC
CGCGGCGGGTGTTGACGCGATGTGATTTCTGCCCAGTGCTCTGAATGTCAAAGTGAAGAAATTCATG
AAGCGCGGGTAAACGGCGGGAGTAACTATGACTCTCTTAAGGAGTCATAGAGCCAGAACCTCCTCGT
GGTCCCGCTGGGCACAGGGATTAATTTTCTGTGGCAAATTTGACTGGCTTCAGAGAGCGTTTTTCGA
AGTGGACTGTGTGACTGCGTTCCCCCTTAGTTGCTATATCCGCTTCGATTAAACATCTCACCTCGACGT
ATAAGATCATTTCTAGAAGGTCGACCAGATGTCCGAGGTCGACCAGTTGTCCGTGTGGAATTGTGAGC
GCTCACAATTCCACACggtacataacttacggtaaatggcccgcctggctgaccgcccaacgacccccgccattgacgtcaatagtaacccaatag
ggactttccattgacgtcaatgggtggagtattacggtaaactgccacttggcagtagcatcaagtgtatcatatgccaagtacccccctattgacgtcaatgacggta
aatggcccgcctggcattgtgccagtagcatgacctatgggactttcctacttggcagtagcatctacgtattagtcacgtctattaccatggctgaggtgagccccagtt
ctgcttcaactctccccatctccccccccctccccacccaatttgtatttatttatttttaatttttgtgcagcgatggggcgggggggggggggggcgcgccca
ggcggggcgggcgggcgaggggcgggcgaggcgagagggtgcggcgagccaatcagagcgcgctccgaaagtttctttatggcga
ggcggcgggcgggcgggcgccataaaaagcgaagcgcgcgggcgggcgaggtagtgcgtgcagcgctgccttcgccccgtccccgtccgccgcgcctcgcg
cgccccccccggtctgactgaccgcgttactccacaggtgagcgggcgggagcgcccttctcctccgggtgtaattagctgagcaagaggttaagg
gatggttggttggtgggtattaatgttaattacctggagcacctgcctgaaatcactttttcagggtgctgacggccaccatgggtgagcaaggcgaggagctgtt
caccgggggtggtgccatcctgctgagctggagcgcgacgtaaacgccacaagttcagcggtgcggcgaggcgaggcgatgccacctacggcaagctg
acctgaagttcatctgcaccaccggcaagctgcccgtgccctggccaccctcgtgaccacctgacctacggcgtgagctgttcagcgctaccccgaccacat
gaagcagcagacttcttaagtcgccatgccgaaggctacgtccaggcgcacatcttctcaaggacgacggcaactacaagaccgcgcggaggtga
gttcgaggcgacacctggtgaaccgcatcagctgaaggcagcagctcaaggaggacggcaacatcctggggcacaagctggagtacaactacaacagcc
acaacgtctatatcatggccgacaagcagaagaacggcatcaaggtgaactcaagatccgccacaacatcgaggacggcagcgtgcagctcgcgaccactacc
agcagaacacccccatggcgacggccccgtgctgctgcccgaaccactacgtgacccagtcgccctgagcaagaccccaacgagaagcgcgatca
catggtcctgctggagttcgtgaccgccgggatcactctggcatggagagctgtacaagtaaaagcttGatccagacatgataagatacattgatgagttgg
acaaaccacaactagaatgcagtgaaaaaatgcttatttgtgaaatttgtgatgctattgcttatttgaaccattataagctgcaataaacaagtTGTGGAATT
GTGAGCGCTCACAATTCCACAGCGCCGCTGAGGTAGATAATCTTTGTATAGTGGGGGGGGATCTCAT
GTACCGGGTTTCTTTTATTTGATTTTCAATAAAACAGACGGTAGCTAGGTTCGCAAGGCAGCCACAAG
CCAAAGATAGGTAGGTGCTCATAGTGAGTAGGGACAGTGCCTTTTTGATTACAACGCGTCAATACCA
TCTGACACGGATACCTTACCGGACTTGTCTATGATCTCCAGACTTGTCCAAGGTGGACGGGCCACCT
TTACTTAACCCGGAAGGAACATATATTAATTATATGTGTTTCGAAAAATAGCCAAATGCCTCGTCATCT
AATTAGTGACGCGCATGAATGGATGAACGAGATTCCCACTGTCCCTACCTACTATCCAGCGAAACCAC
AGCCAAGGGAACGGGCTTGGCGGAATCAGCGGGGAAAGAAGACCCTGTTGAGCTTGA
GGCACGGTGAAGAGACATGAGAGGTGTAGAATAAGTGGGAGGCCCCCGCGCCCCCGGTGTCCC
CGCGAGGGGCCCGGGCGGGGTCCGCGGCCCTGCGGGCCCGCGGTGAAATACCACTACTCTGATCGT

TTTTTCACTGACCCGGTGAGGCGGGGGGGCGAGCCCGAGGGGCTCTCGCTTCTGGCGCCAAGCGCCC
GCCCGGCCGGGCGCGACCCGCTCCGGGGACAGTGCCAGGTGGGGAGTTTGACTGGGGCGGTACACC
TGTCAAACGGTAACGCAGGTGTCCTAAGGCGAGCTCAGGGAGGACAGAAACCTCCCGTGGAGCAGA
AGGGCAAAAGCTCGCTTGATCTTGATTTTCAGTACGAATACAGACCGTGAAAGCGGGGCCTCACGAT
CCTTCTGACCTTTTGGGTTTTAAGCAGGAGGTGTCAGAAAAGTTACCACAGGGATAACTGGCTTGTGG
CGGCCAAGCGTTCATAGCGACGTCGCTTTTTGATCCTTCGATGTCGGCTCTTCCTATCATTGTGAAGCA
GAATTCGCCAAGCGTTGGATTGTTACCCACTAATAGGGAACGTGAGCTGGGTTTAGACCGTCGTGAG
ACAGGTTAGTTTTACCCTACTGATGATGTGTTGTTGCCATGGTAATCCTGCTCAGTACGAGAGGAACCG
CAGGTTACAGACA

400. CBh 启动子

[1157]

Cgt tacataact acggtaaatg gccgcctgg ctgaccgcc aacgacccc gcccatgac gtcaatagta acgccaatag ggactttcca
ttgacgtcaa tgggtggagt atttacggta aactgccac ttggcagtac atcaagtgtat tcatatgccca agtacgccc ctattgacgt caatgacggt
aaatggccc cctggcattg tgcccagtac atgacctat gggactttcc tacttggcag tacatctacg tattagtcac cgctattacc atggtcgagg
tgagccccac gttctgttc actctccca tctcccccc ctccccacc ccaattttgt atttattat ttttaatta tttgtgcag cgatgggggc
gggggggggg gggggggcgc cgccaggcgc ggcggggcgc ggcgaggcgc gggcggggc gagcgcgaga ggtgcggcg cagccaatca
gagcggcgcg ctccgaaagt ttcctttat ggcgaggcgc ggcggcgcgc ggccctataa aaagcgaagc gcgcggcggc cgggagtcgc
tgcgacgtg ccttcgccc gtgcccgcct ccgcccgc ctcgcgcgc ccgccccgc tctgactgac cgcgttact ccacaggtga
gcgggcggga cgccctct cctcgggct gtaattagct gagcaagagg taagggtta agggatggti ggttggtggg gtattaatgt ttaattacot
ggagcacctg cctgaaatca cttttttca ggttgg

401. EFS 启动子

Gggcagagcgcacatgcccacagtcccagagaagttgggggaggggtcggcaattgaaccggtgcctagagaaggtggcggggtaaactgggaaagtg
atgtcgtgtactggctccgcttttcccgagggtgggggagaaccgtatataagtgcagtagtcgccgtgaacgttcttttcgcaacgggttgcgcgagaca
Gg

403. EFSBsiWI 启动子

Gggcagagcgcacatgcccacagtcccagagaagttgggggaggggtcggcaattgaaccggtgcctagagaaggtggcggggtaaactgggaaagtg
atgtcgtgtactggctccgcttttcccgagggtgggggagaaccgtatataagtgcagtagtcgccgtgaacgttcttttcgcaacgggttgcgcgagaca
gcgtacg gccacatg

404. mPGK 启动子

TTCTACCGGGTAGGGGAGGCGCTTTTCCCAAGGCAGTCTGGAGCATGCGCTTTAGCAGCCCCGCTGG
GCACTTGGCGCTACACAAGTGGCCTCTGGCCTCGCACACATTCCACATCCACCGGTAGGCGCCAACC
GGCTCCGTTCTTTGGTGGCCCCCTTCGCGCCACCTTCTACTCCTCCCCTAGTCAGGAAGTTCCCCCCCCG
CCCCGCAGCTCGCGTCGTGCAGGACGTGACAAATGGAAGTAGCACGTCTCACTAGTCTCGTGCAGAT
GGACAGCACCGCTGAGCAATGGAAGCGGGTAGGCCCTTTGGGGCAGCGGCCAATAGCAGCTTTGCTCC
TTCGCTTTCTGGGCTCAGAGGCTGGGAAGGGGTGGGTCCGGGGGCGGGCTCAGGGGCGGGCTCAGG
GGCGGGGCGGGCGCCCCGAAGGTCCTCCGGAGGCCCGGCATTCTGCACGCTTCAAAAGCGCACGTCT
GCCGCGCTGTTCTCCTCTTCCTCATCTCCGGGCCTTTCGACCT

405. SV40 LpA

aactgtttattgcagcttataatggttacaaataagcaatagcatcacaatttcacaaataagcatttCttcactgcatttagttgtggtttgtccaaactcatcaatgt
atctta

406. min pA

[1158] AAATAAAATACGAAATGAATTCGTAATCATGTGTGTTGGTTTCTTGTGT

407. BGH pA

CTAGTTGCCAGCCATCTGTTGTTTGCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCC
ACTGTCCTTTCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGG
GGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCG
GTGGGCTCTATGG

408. CMV*

CGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCATTGACGTCAA
TAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTAC
GGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATG
ACGGTAAATGGCCCGCCTGGCATTATGCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACA
TCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGG
GGTTTGACTCACGGGGATTTCGAAGTCTCCACCCCATGACGTCAATGGGAGTTTGTGTTTGGCACCAA
AATCAACGGGACTTTCCAAAATGTCGTAACAACCTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGT

ACGGTGGGAGGTCTATATAAGCAGAGCTcgtttagtgaaccgCTTGTCTCGCTCCGGGGAACGCTCGGAAACT
CCCGGCCCGCCGCCACCCGCGTCTGTTCTGTTACACAAGGGAAGAAAAGCCGCTGCCGCACTCCGAGT
GTCGTACGGCCACC.

409. SV40*

Ctgtggaatgtgtcagftagggtgtgaaagtcgccaggtcccgaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccaggtgtgaaagtc
cccggctcccgaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccatagtcggccctaaactccgccatcccgccctaaactccgccagtt
ccgccattctccgcccattGgctgactaattGCGtCtatAtatgcagaggccgaggccgctcggcctctgagctattccagaagtagtgaggaggcCGTA
CGGCCACC.

408-410 无序列

411. NLS-GFP ORF 序列

[1159] ATG gcc cca aag aag aag cgg aag
gtcgtgagcaagggcgaggagctgttcaccggggtgtgccatcctgtgcagctggacggcgacgtaaacggccacaagttcagcgtgtccggcgaggcgga
ggcgcatgccacctacggcaagctgacctgaagttcatctgcaccaccggcaagctgccgtgccctggccaccctctgaccacctgacctacggcgtgca
gtgttcagccgctaccccgaccacatgaagcagcagcacttcttaagtcggccatgccggaaggctacgtccaggagcgcaccatcttcttaaggacgacggca
actacaagaccgcgccgaggtgaagttcagggcgacaccctgggtgaaccgcatcgagctgaaggcgatcgactcaaggaggacggcaacatcctggggcac
aagctggagtacaactacaacagccacaacgtctatatcatggccgacaagcagaagaacggcatcaaggtgaactcaagatccgccacaacatcgaggacggc
agcgtgagctcgccgaccactaccagcagaacaccccatcgcgacggcccgctgctgcccgacaaccactacctgagcaccagtcggccctgagcaa
agacccaacgagaagcgcatcacatgctcgtgagttcgtgaccgcccgggcatcactctcgcatggacgagctgtacaagtaa

412. NLS-GFP 氨基酸序列

MAPKKRKRVVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGDATYGKLTCLKFICTTGKLPVPWPTLV
TTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFK
EDGNILGHKLEYNYSNHNVIYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHY
LSTQSALSKDPNEKRDHMLLEFVTAAGITLGMDELYK-

413. GFP ORF 序列

atggtgagcaagggcgaggagctgttcaccggggtgtgccatcctgtgcagctggacggcgacgtaaacggccacaagttcagcgtgtccggcgaggcgga
ggcgcatgccacctacggcaagctgacctgaagttcatctgcaccaccggcaagctgccgtgccctggccaccctctgaccacctgacctacggcgtgca

gtgttcagccgctaccccgaccacatgaagcagcacgacttctcaagtcgccatgccgaaggctacgtccaggagcgcaccatcttctcaaggacgacggca
actacaagaccgcgccgaggtgaagttcgaggggcgacaccttggtgaaccgcacgcagctgaaggcgacgttcaaggaggacggcaacatcctggggcac
aagctggagtacaactacaacagccacaacgtctatatacatgcccgacaagcagaagaacggcatcaagggtgaacttcaagatccgccacaacatcgaggacggc
agcgtgacgctcgcggaccactaccagcagaacacccccatcgccgacggggcccgctgctgctcccgacaaccactacctgagcaccacgtccgcctgagcaa
agaccccaacgagaaagcgcgatcatatgtctcgtgagttctgtgacggccggggatcactctcggcgatggacgagctgtacaagtaa

414. GFP 氨基酸序列

MVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEDATYGKLTCLKICTTGKLPVPWPTLVTTILTYGVQ
CFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGH
KLEYNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQNTPIGDGPVLLPDNHYLSTQSALS
KDPNEKRDHMLVLEFYTAAGITLGMDELYK-

415. Tdtomato ORF 序列

atggtgtccaaaggcgaggaaagtgaatcaagggaattatgagattcaaggtcagaatggagagctatgaacggccacgagttcagagatcgaggagaaaggagaa
ggcagacccctacgagggcaccacaaccgctaaattaaaagtgaacaaggcgccgacctctgccgttcgctgggatctctgtcccccagttcatgtacggcagca
aggcctacgtgaagcaccctgctgacatccccgattacaaaaagctgtcttttctgagggctttaagtgaggagagagtgaacttcgaggacggaggcctggtga
ccgttaccacaagacagcagcctgcaagatggaacactgatctacaagtgaagatgcggggcactaatctccacagacggccccgtgatgcagaagaaaaccat
gggctgggaggcctctacagaacggctgtacccagagacggagtgctgaaggcgagatccaccaggcactgaagctgaagatggcgggacactacctggtgg
agttcaagaccattacatggccaagaagcctgtgcagctgcctggctactactacgtggacaccaagctggacatcacctcccacaatgaggattatacatctgtgga
acagtagcagagaaagcgaggcgagacaccacctgttctctggccacggcacaggctccaccggctctggcagcagcggcacagccagcagcgaggacaacaa
catggccgtgatcaaggagttcatgcggttcaaggtgcggatggaaggcagtatgaacggccatgagttcgaaatcgaggcggaaggcggaaggcagacctacga
gggcacacagacagctaagctgaaggtagcaaggcgagccctctgccattgcctgggacatcctgagccctcagtttatgtatggcagcaaggcctacgtgaag
caccgcccgacatccccgactacaagaagctgagcttccctgaggggtttcaatgggaaagggtgatgaacttcgaggacggcgggactggtgacagtgaccag
gacagcagcctgcaggatggcaccctgatctataagtggaagatgagaggcaccaactttccacctgacggggcccgtgatgcagaaaaaacatgggatgggag
gccagcaccgagcggctgtaccttagatggcgttctgaaggcgagatccaccaggccctgaagctgaaggacggcgggccactacctgttggaagttaagac
catctacatggccaagaaacctgtgcagctgcctgatactactactgtggacacaaagctggacatcacctctcacaaaggaggactataccatctgcgagcagtagca
acgcagcgaaggcgagacaccatctgttctgtacgggtatggacgagctgtacaagtaatga

416. Tdtomato 氨基酸序列

MVSKGEEVIK EFMRFKVRME GSMNGHEFEI EGEGERPYE GTQTAALKVT KGGPLPFAWD ILSPQFMY
GS KAYVKHPADI PDYKKLSFPE GFKWERVMNF EDGGLVTVTQ DSSLQDGTLI YKVKMRGTNF PPDGPV
MOKK TMGWEASTER LYPRDGVLKG EIHOALKLKD GGHYLVEFKT IYMAKKPVOL PGYYYYVDTKL DIT

SHNEDYT IVEQYERSEG RHHLFLGHGT GSTGSGSSGT ASSEDNNMAV IKEFMRFKVR MEGSMNGHEF EI
EGEGEGRP YEGTQTAKLK VTKGGPLPFA WDILSPQFMY GSKAYVKHPA DIPDYKKLSF PEGFKWERVM
NFEDGGLVTV TQDSSLQDGT LIYKVKMRGT NFPPDGPVMQ KKTMGWEAST ERLYPRDGLV KGEIHQAL
KL KDGGHYLVEF KTIYMAKKPV QLPGYYYVDT KLDITSHNED YTIVEQYERS EGRHHLFLYG MDELYK

417. mCherry ORF 序列

auggugagcaaggcgaggaggauaacauggccaucacuaagagaucaugcgcuucaaggugcacauaggagcuccgugaacggccacgaguucga
gaucgagggcgaggcgaggcgccccuacgaggcaccagaccgccaagcugaaggugaccaaggugggccccugcccuucgcuugggacaucuu
gucccccaguucauguacggcucaagggcuacgugaagcaccggcgacaucccgacuacuugaagcuguccuucccgaggcgcucaaguggga
gcgcgugaugaacuucgaggacggcgcgugggugaccgugaccaggaucuccuccugcaggacggcgaguucaucaagggugaagcugcgcgccac
caacuucccuccgacggccccgaaugcagaagaagaccauggcgugggaggccuccuccgagcgggauguaccccaggagcggcgccugaaggcgga
gaucgaagcagaggcugaagcugaaggcggcgccacuacgacgcugaggucagaaccaccuacaaggccaagaagcccgugcagcugcccgcgccuac
aacgucaucauagauggacauaccuccacacgaggacuacaccaucguggaacaguacgaacgcggcgaggcgccacuccaccggcggaugg
acgagcuguacaaguag

[1161]

418. mCherry 氨基酸序列

MVSKGEEDNMAIIKEFMRFKVHMEGVSNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDILSPQFM
YGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVTVTVTQDSSLQDGEFIYVKVLRGTNFPDGPVM
QKKTMGWEASSERMYPEDGALKGEIKQRLKLDGGHYDAEVKTTYKAKKPVQLPGAYNVNIKLDITSHN
EDYTIVEQYERAEGRHSTGGMDELYK-

419. 人 TERT ORF 序列

atggtgatgcctcgagcaccgggtgagagccgtgcggagcctgctgcggtctcactacagagaggtcctccccctggccacgttcgtgcggtctggtggccctc
agggctggcgactggtgcagcgcgggcaccggccgattccgcgcctggtggccagtgccctggtgtgtgcttgggatgccagacctcctcgtcggtcc
ttcttttagacaggtgtcttctgaaggagctggtggccgggtgctgcagcggtgtgcgagcgggagtaagaatgtgttagctttcgttttccctgctggacg
gcgcccagaggcgaccacctgaagccttactacctcgtgagaagctacctgcctaaccctgacagacgcccctgcggggcagcggcgctggggcctgcta
ctgagacgggtggagatgatgtgctggtgcacctgctggccagatgcgcccctttgtgtggtggcccttagctgtgctaccaggtgtgtgcccctccctgtac
cagctggggcgccgcacgcaggccagacccccctccgcacgcagcgcccttaggagacgtctgggctgcgaaagagcttggaaccacagtggtgcgagaagcc
ggcgctccccctggcctccctgcccctggagccagaagaaggagggtcggccagcagaagcctccgctgcccgaagcgccctagacggggcgccggccct
gaaccggagcggaccccggtcggccaggcgagttggggccaccccggaaggaccagaggacctagcgaccggggcttctgctgtgtctccccagccggcca
ggccgagggaagcaaccagcctggaggcgccctgagcggcaccagacactccacccttccgttggcagacagcaccacggggccccaagcaccagcaga
cctccgcgacctgggacacacctgtcctcctgtgtacgccgagaccaagcacttctgtactcctctggtgcgataaggaacagctccggccttagcttctgctgagct

[1162]

ctctgagaccagcttgaccggcgccgcagactcgtggaacaatctctgggcagcagaccttggatgccaggcaccctagacggctgcctagactgcccc
gagatactggcagatgctccactgttttgaactgctgggcaaccacgctcagtgctctacggcgttctgctgaaaaccactgccccctgagagccgtgtcacc
ccagccgccggcgtgtgcccagagaaaagcctcagggaagcgtggctgctcctgagggaaggatacagaccctagacgcctgggcaactgctgcgccaaca
tagcagcccttgccaggtgtacggcttcgtcagagctgtctgagaaggctgggtcccccgggactgtggggtccagacacaacgagcggagattcctgcggaac
aaaagaagtcatcagcctgggcaagcacgccaagctgtctctgaggagtgtacctggaagatgagcgtgagggtattgcgcgtggctgagaagatccctggag
tgggctgtgtccagccgagcacagactgcgagaggaaatcctggccaaatcctgcactggctgatgagcgtatatgtgtggaaactgctgcggagcttctct
acgtgaccgagacaacattcagaagaacagacttttcttaccggaagtctgtatggagcaagctgcagtcctatcggaatcagacagcacctgaagagagttcagc
tgcgggaactgagcggagcagaggtgaggcagcatagagaggttagacctgctctgttaccagcagactgagattatccaaaacctgatggcctgagacctat
cgtgaatatggactacgtggtggcgcccgaaaccttccggcgtgagaagcgggctgaaagactgacatccgggtgaaggccctgttcagcgtgtgaactacga
gagagctcgtggcctgggctactggcgccagcgtgctggcctggatgacatccacagagcctggagaacattcgtgctgagagtggggctcaagatcccc
accggagctgtacttcgttaaggtggacgtgacaggcgcatagcacaccatccctcaggacagactgaccgaggtgattgccagcattatcaagcccagaacacct
attcgtgagacgggtacgcccgtggtgcagaaagccgcacatggccacgtgctggaaagccttcaagtctcacgtgtccacctgaccgacctgcagccttcatgcg
gcaattcgtggcccacctccaggagacaagcctctgagagacgccgtggtgatcagcagtcagcagccttaataaggccttagcggcctgttcagcgtgttcc
tgagatttatgtccaccacgccgtgagaatcagaggcaaaagctacgtgcagtgccagggtatccccaaggatctatcctgagcacctgctgtgcagcctgtgt
acggcgacatggaaaacaagctgttcgcccgaatccggagagacggcctgctgctgcggctcgttgacgacttccctgctggtgacacctcacctgacacagccaa
aaccttctcagaacactgggtcggggcgctcccagtagtggtgctgctgaacctgagaagacagtggtgaacttccctgtggaagatgaggcactcggcgga
acagcgttcgtgcagatgcccgccagcctgttcccctgtgctgcggcctgctgctggacaccagaacctggaggtgcagagcgattacgcagctatgccagaa
ccagcatccgagccagcctaacctttaaccggggtcctaagccggcgagaatatgagaagaagctgttcggagtgctgaggcttaatgccacagcctgttctg
gacctgcagggtcaacagcctgcaaacagtgtgcaccaacaatctacaagatcctgtgttaaacgctacagatttcacgctgtgtgctgcagctgccttccaccagc
agggtgtggaagaaccttacgttttctgagagtgtctctgatacagccagttgtgtctacagatcctgaaggccaagaacgccggcatgagcctgggagccaagg
gcgctgctgggctctgcttctgagcgtgagcgtggtgctgcccaggccttctgctgaagctgaccagacatagagtacacctgctcctctcctggcagcc
tgagaaccgccagaccagctctctcggaagctgcctggcacaacctgaccgcctggaagctgctgccaatcctgcctgcctagcgactcaagaccatcctg
gactgataa

420. 人 TERT 氨基酸序列

MVMPRAPRCRAVRSLLRSHYREVLPLATFVRRLGPQGWRVLVQRGDPAAFRALVAQCLVCPWDARPPPA
APSFQVSVCLKELVARVLQRLCERGAKNVLAFGFALLDGARGGPPEAFTTSVRSYLPNTVTDALRGSGAW
GLLLRRVGDDVLVHLLARCALFVLVAPSCAYQVCGPLYQLGAATQARPPPHASGPRRRLGCERAWNHS
VREAGVPLGLPAPGARRRGGSASRSLPLKPRRGAPEPERTPVGQGSWAHPGRTRGPSDRGFCVVSPA
RPAEEATSLEGALSGTRHSHPSVGRQHHAGPPSTSRPPRPWDTPCPPVYAETKHFYSSGDKEQLRPSFLLS
SLRPSLTGARRLVETIFLGSRPWMPGTPRRLPRLPQRYWQMRPLFLELLGNHAQCPYGVLLKTHCPLRAA
VTPAAGVCAREKPQGSVAAPPEEDTDPRRLVQLLRQHSSPWQVYGFVRACLRRLVPPGLWGSRHNERF

LRNTKKFISLGKHAKLSLQELTWKMSVRDCAWLRRSPGVGCVPAAEHRLREEILAKFLHWLMSVYVVEL
LRSFFYVTETTFQKNRLFFYRKS VWSKLQSIGIRQHLKRVQLRELSEAEVRQHREARPALLTSRLRFIPKPD
GLRPVNM DYVVGARTFRREKRAERLTSRVKALF SVLNYERARRPGLLGASVLGLDDIHRWRTFVLRV
RAQDPPPELYFVKVDVTGAYDTIPQDRLTEVIASIIKPQNTYCVRRYAVVQKAAHGHVRKAFKSHVSTLTD
LQPYMRQFVAHLQETSPLRDAVVIEQSSSLNEASSGLFDVFLRFMCHHAVRIRGKSYVQCQGIPQGSILSTL
LCSLCYGD MENKLFAGIRRDGLLLRLVDDFLLVTPHLTHAKTFLRTLVRGVPEYGCVVNLRKTVVNFVPE
DEALGGTAFVQMPAHGLFPWCGLLLDTRTLEVQSDYSSYARTSIRASLTFNRGFKAGRNMRRKLFGLVRL
KCHSLFLDLQVNSLQTVCTNIYKILLQAYRFHACVLQLPFHQVWKNPTFFLRVISDTASLCYSILKAKN
AGMSLGAKGAAGPLPSEAVQWLCHQAFLKLTRHRVTVPLLSLRTAQTQLSRKLP GTTLTALEAAANP
ALPSDFKTILD--

421. 人 PAH ORF 序列

[1163]

atggtgatgagcaccgccgtgctggagaatcctggcctgggcagaaaagctgccgatttcggccaagagacaagctacattgaggataattgcaaccagaacggcg
ctatcagcctgatctttcttgaagagggaagtgggcgctctggccaaggctcctgaggctgttcgaggaaaacgacgtgaacctgacctatcgaaagcagacca
ggcggctgaagaaagatgagtatgagttctaccacacctggacaagagatctctccctgcctgacaaacatcatcaaaatcctaagacacgacattggcgccaccg
tgcacgagctgtcaagagataagaagaaggacaccgtgccttggttcctagaacctccaggagctggatagattcgccaatcagatcctgtcttacggcgccgaac
tggacgccgaccaccccgctcaaggacctgtgtaccgggcccagacgcaagcagttcgccgacatcgctataactaccggcatggccagcctatccctagagt
ggatacatggaagaagagaaaaaacctggggcaccgttttcaagacctgaaaagcctgtataagactcacgcctgtacgagtacaaccacatcttccctctgct
ggaaaagtactgtggctttcacgaggacaacatcccacagctggaggacgtgtcccagttcctgcagacctgtaccggcttcgggctggcccgctggcgccgactg
ctgagcagccgggacttctggcgccgactggcttttagagtgttccactgcacacagtacatcagacacggctccaagcctatgtacacacctgagcccacatctgc
cacgagctgtgggacacgtccctctgttagcgatagatctttcgccagtttagccaggagatcggcctggccagcttggcgccctgatgaatacatcgagaaa
ctggctacaatctactggttcacagtggagttcgccctctgcaagcaggggcagacgacatcaaggcctacggcgccggactgcttctagcttcggcgagctgcaatac
tgtctgtccgagaagcccaagctgtgcctctggaactggaaaagaccgccatccagaattacaccgtgaccgagttccagccctgtactacgtggtgctgagagttc
aacgacgctaaggaaaagggtcggaactcgccgccacaatccaagaccttcagcgtgcggtacgaccctacacccaagaatcgaggtgctggacaacacc
cagcagctgaagatcctggcagatagcatcaacagcgagatcggaatcctgtgcagcgccctgcagaagattaagtgataag

422. 人 PAH 氨基酸序列

MSTAVLENPGLGRKLSDFGQETSYIEDNCNQNGAISLIFSLKEEVGALAKVLRLFEENDV
NLTHIESRPSRLKKDEYEFFTHLDKRSLPALTNIKILRHDIGATVHEL SRDKKKDTPVW
FPRTIQELDRFANQILSYGAELDADHPGFKDPVYRARRKQFADIAYNRYRHGQPIPRVEYM
EEEKKTWGT VFKTLKSLYKTHACYEYNHIFLLEKYCGFHEDNIPQLEDVSQFLQTCTGF
RLRPVAGLLSSRDFLGGLAFRVFHTCTQYIRHGSKPMYTPEPDICHELLGHVPLFSDRSFA

QFSQEIGLASLGAPDEYIEKLATIWFTVEFGLCKQGDSIKAYGAGLLSSFGEQLQYCLSE
KPKLLPLELEKTAIQNYTVTEFQPLYVAESFNDAKEKVRNFAATIPRPFVRYDPYTQR
IEVLDNTQQLKILADSINSEIGILCSALQKIK

423-498 无序列

499. XbaI_RNAPI_term_rlacO_CbH_HindIII_Kozak_NLSGFP_HindIII_SV40LPA_rlacO_NotI

[1164] TCTAGAAAGTTCGACCAGATGTCCGAGGTCGACCAGTTGTCCGTGTGGAATTGTGAGCGCTCACAATTC
CACA Cgt tacataact acggtaaatg gccgcctgg ctgaccgcc aacgacccc gccattgac gtcaatagta acgccaatag ggactttcca
ttgacgtcaa tgggtggagt atttacggta aactgccac ttggcagtac atcaagtga tcatatgcca agtacgccc ctattgacgt caatgacggt
aaatggcccg cctggcattg tgcccagtac atgacctat gggactttcc tacttggcag tacatctacg tattagtcat cgctattacc atggtcgagg
tgagccccac gttctgcttc actctcccca tctccccc cccccccc ccaattttgt atttattat ttttaatta tttgtgcag cgatgggggc
gggggggggg gggggggcgc gccagggcgc ggcggggcgc ggcgaggggc gggcggggc gaggcggaga ggtgcggcgc cagccaatca
gagcggcgcgc ctccgaaagt ttcttttat ggcgagggcg cgcgcgccgc ggccctataa aaagcgaagc gcgcggcggc cgggagtcgc
tgcgacgtcg ccttcgccc gtgcccgc cgcgcgcgc ctcgcgcgc cgcggcggc tctgactgac cgcgttactc ccacaggtga
gcggggcgga cgcccttct cctccggct gtaattagct gagcaagagg taagggtta agggatggtt gttggtggg gtattaatgt ttaattacct
ggagcacctg cctgaaatca cttttttca ggttga agctt GCCACCATG gcc cca aag aag aag cgg aag
gtcgtgagcaaggcgaggagctgtcaccggggtggtgccatcctggtcagctggacggcgacgtaaacggccacaagttcagcgtgtccggcgaggcgga
ggcgatgccacctacggcaagctgacctgaagttcatctgcaccaccggcaagctgccgtgccctggcccaccctcgtgaccacctgacctacggcgtgca
gtgttcagccgctaccccgaccacatgaagcagcagacttctcaagtcgccatgccgaaggctacgtccaggagcgcaccatcttctcaaggacgacggca
actacaagaccgcgcggaggtgaagttcgaggcgacaccctggtgaacgcgcagctgaaggcgatcgaactcaaggaggacggcaacatcctggggcac
aagctggagtacaactacaacagccacaacgtctatatcatggccgacaagcagaagaacggcatcaaggtgaactcaagatccgccacaacatcgaggacggc
agcgtgcagctcgcgaccactaccagcagaacacccccatcgcgacggccccgtgctgctgcccgaaccactacctgagcaccagtcggccctgagcaa
agacccaacgagaagcgcatcacatgtcctgctggagttcgtgaccgcggccggatcactctcgcatggacgagctgtacaagtaaaagcttGatccaga
catgataagatacatgatgagtttgacaaccacaactagaatgcagtgaaaaaatgctttattgtgaatttggatgctattgctttattgtaaccattataagctgc
aataaacaagtTGTGGAATTGTGAGCGCTCACAATTCACAGCGGCCGC

500. TCA5_CbHbsi_GFP_GeFo3

gggagacggucaaccgcuaggagcggugaccgGAAUUCGGCGGGAGUAACUAUGACUCUCUUAAGGAGUCAUAG
AGCCAGAACCUCUCUGUGUCCCGCUGGGCACAGGGAUUAUUUUUCUGUGGCAAAUUUGACUG
GCUUCAGAGAGCGUUUUUCGAAGUGGACUGUGUGACUGCGUCCCCCUUAGUUGCUAUAUCCG
CUUCGAUUAACAUCUCACCUCGACGUUAAGAUAUUUCUAGAAGGUCGACCAGAUGUCCGAGG

[1165]

UCGACCAGUUGUCCGUGUGGAAUUGUGAGCGCUCACAAUCCACACGUUACAUAAACUUACGGUA
AAUGGCCCCGCCUGGCUGACCGCCCAACGACCCCCGCCAUUGACGUCAAUAGUAAACGCCAAUAGG
GACUUUCCAUUGACGUCAAUGGGUGGAGUAUUUACGGUAAACUGCCCACUUGGCAGUACAUCAA
GUGUAUCAUAUGCCAAGUACGCCCCCUAUUGACGUCAAUGACGGUAAAUGGCCCCGCCUGGCAUU
GUGCCCAGUACAUGACCUUAUGGGACUUUCCUACUUGGCAGUACAUCUACGUAAUAGUCAUCGC
UAUUACCAUGGUCGAGGUGAGCCCCACGUUCUGCUUCACUCUCCCCAUCUCCCCCCCCUCCCCACC
CCCAAUUUUGUAUUUAUUUAUUUUUAAUUUAUUUUGUGCAGCGAUGGGGGCGGGGGGGGGGGGG
GGGCGCGCGCCAGGCGGGGCGGGGCGGGGCGAGGGGCGGGGCGGGGCGAGGCGGAGAGGUGCGG
CGGCAGCCAAUCAGAGCGGCGCGCUCCGAAAGUUUCCUUUUAUGGCGAGGCGGCGGGCGGCGG
CCCUAUA AAAAGCGAAGCGCGCGGGCGGGGAGUCGUGCGACGCGUCCUUCGCCCCGUGCCCC
GCUCCGCCGCCGCCUCGCGCCGCCCGCCCCGGCUCUGACUGACCGCGUUACUCCACAGGUGAGCG
GGCGGGACGGCCCUUCUCCUCCGGGCGUGAAUUAGCUGAGCAAGAGGUAAGGGUUUAAGGGAUG
GUUGGUUGGUGGGGUAAUUAAGUUUAUUUACCUGGAGCACCUGCCUGAAAUCACUUUUUUUCAG
GUUGGCGUACGGCCACCAUGGUGAGCAAGGGCGAGGAGCUGUUCACCGGGGUGGUGCCCAUCCU
GGUCGAGCUGGACGGCGACGUAAACGGCCACAAGUUCAGCGUGUCCGGCGAGGGCGAGGGCGAU
GCCACCUACGGCAAGCUGACCCUGAAGUUCUUCUGCACCACCGGCAAGCUGCCCCGUGCCCUGGCC
CACCUCGUGACCACCCUGACCUACGGCGUGCAGUGCUUCAGCCGCUACCCCGACCACAUGAAGC
AGCACGACUUCUUAAGUCCGCCAUGCCCCGAAGGCUACGUCCAGGAGCGCACCAUCUUCUUAAG
GACGACGGCAACUACAAGACCCGCGCCGAGGUGAAGUUCGAGGGCGACACCCUGGUGAACC CGAU
CGAGCUGAAGGGCAUCGACUUAAGGAGGACGGCAACAUCUGGGGCACAAGCUGGAGUACAAC
UACAACAGCCACAACGUCUAUAUCAUGGCCGACAAGCAGAAGAACGGCAUCAAGGUGAACUUA
AGAUCGCCACAACAUCGAGGACGGCAGCGUGCAGCUCGCCGACCACUACCAGCAGAACACCCCC
AUCGGCGACGGCCCCGUGCUGCUGCCCCGACAACCACUACCUGAGCACCCAGUCCGCCCGAGCAA
AGACCCCAACGAGAAGCGCGAUCACAUGGUCCUGCUGGAGUUCGUGACCGCCGCCGGGAUCACUC
UCGGCAUGGACGAGCUGUACAAGUAAAAGCUUGAUCCAGACAUGAUAAAGAUACAUUGAUGAGUU
UGGACAAACCACAACUAGAAUGCAGUGAAAAAAUUGCUUUAUUUGUGAAAUUUGUGAUGCUAUU
GCUUUAUUUGUAACCAUUAUAAGCUGCAAUAAACAAGUUGUGGAAUUGUGAGCGCUCACAAUUC
CACAGCGGCCCGCUGAGGUAGAUAAUCUUUGUAUAGUGGGGGGGGAUCUCAUGUACCGGGUUUCU
UUUAUUUGAUUUUCAAUAAAACAGACGGUAGCUAGGUUCGCAAGGCAGCCACAAGCCAAAGUA
GGUAGGGUGCUCAUAGUGAGUAGGGACAGUGCCUUUUGAUUCACAACGCGUCAAUACCAUCUGA
CACGGAUACCCUUAACGGACUUGUCAUGAUCUCCAGACUUGUCCAAGGUGGACGGGCCACCUU
ACUUAACCCGAAAAGGAACAUUAUUAUUUAUUAUGUGUUCGGAAAA

501. TCA5_CBhBsi_GFP_GeFo3_A22

gggagacggucaaccgcuaggagcggugaccgGAAUUCGGCGGGAGUAACUAUGACUCUCUUAAGGAGUCAUAG
AGCCAGAACCUCUCGUGGUCCCGCUGGGCACAGGGAUUAUUUUUCUGUGGCAAAUUUGACUG
GCUUCAGAGAGCGUUUUUCGAAGUGGACUGUGUGACUGCGUUCUUUUUUUAGUUGCUAUAUCCG
CUUCGAUUAACAUCUCACCUCGACGUUAUAGAUAUUUCUAGAAGGUCGACCAGAUGUCCGAGG
UCGACCAGUUGUCCGUGUGGAAUUGUGAGCGCUCACAAUCCACACGUUACAUAACUUACGGUA
AAUGGCCCGCCUGGCUGACCGCCCAACGACCCCGCCCAUUGACGUCAAUAGUAACGCCAAUAGG
GACUUUCCAUUGACGUCAAUGGGUGGAGUAUUUACGGUAAACUGCCCACUUGGCAGUACAUCAA
GUGUAUCAUAUGCCAAGUACGCCCCCUAUUGACGUCAAUGACGGUAAAUGGCCCGCCUGGCAUU
GUGCCCAGUACAUGACCUUAUGGGACUUUCCUACUUGGCAGUACAUCUACGUAAUAGUCAUCGC
UAUUACCAUGGUCGAGGUGAGCCCCACGUUCUGCUUCACUCUCCCCAUCUCCCCCCCCUCCCCACC
CCCAAUUUUUGUAUUUAUUUAUUUUUAAUUUUUUUGUGCAGCGAUGGGGGCGGGGGGGGGGGGG
GGGCGCGCGCCAGGCGGGGCGGGGCGGGGCGAGGGGCGGGGCGGGGCGAGGCGGAGAGGUGCGG
CGGCAGCCAAUCAGAGCGGCGCGCUCCGAAAGUUUCCUUUUUUGGCGAGGCGGCGGGCGGCGG
CCCUAUA AAAAAGCGAAGCGCGCGGGCGGGGAGUCGCGGACGCGUGCCUUCGCCCCGUGCCCC
[1166] GCUCCGCCGCCGCCUCGCGCCGCCCGCCCCGGCUCUGACUGACCGCGUUACUCCCACAGGUGAGCG
GGCGGGACGGCCCUUCUCCUCCGGGCUGUAAUUAGCUGAGCAAGAGGUAAGGGUUUAAGGGAUG
GUUGGUUGGUGGGGUAAUUAUGUUUAAUUACCUGGAGCACCUGCCUGAAAUACAUUUUUUUCAG
GUUGGCGUACGGCCACCAUGGUGAGCAAGGGCGAGGAGCUGUUCACCGGGGUGGUGCCCAUCCU
GGUCGAGCUGGACGGCGACGUAAACGGCCACAAGUUCAGCGUGUCCGGCGAGGGCGAGGGCGAU
GCCACCUACGGCAAGCUGACCCUGAAGUUAUCUGCACCACCGGCAAGCUGCCCGUGCCCGUGGCC
CACCCUCGUGACCACCCUGACCUACGGCGUGCAGUGCUUCAGCCGCUACCCCGACCACAUGAAGC
AGCACGACUUCUUAAGUCCGCCAUGCCCGAAGGCUACGUCCAGGAGCGCACCAUCUUCUUAAG
GACGACGGCAACUACAAGACCCGCGCCGAGGUGAAGUUCGAGGGCGACACCCUGGUGAACCGCAU
CGAGCUGAAGGGCAUCGACUUAAGGAGGACGGCAACAUCUGGGGCACAAGCUGGAGUACAAC
UACAACAGCCACAACGUCUAUAUCAUGGCCGACAAGCAGAAGAACGGCAUCAAGGUGAACUUA
AGAUCCGCCACAACAUCGAGGACGGCAGCGUGCAGCUCGCCGACCACUACCAGCAGAACACCCCC
AUCGGCGACGGCCCCGUGCUGCUGCCCCGACAACCACUACCUGAGCACCCAGUCCGCCUGAGCAA
AGACCCCAACGAGAAGCGCGAUCACAUGGUCCUGCUGGAGUUCGUGACCGCCGCGGGGAUCACUC
UCGGCAUGGACGAGCUGUACAAGUAAAAGCUUGAUCCAGACAUGAUAAAGAUACAUGAUGAGUU
UGGACAAACCACAACUAGAAUGCAGUGAAAAAAAUGCUUUAUUUGUGAAAUUUGUGAUGCUAUU

GCUUUAUUUGUAACCAUUUAAGCUGCAAUAAACAAGUUGUGGAAUUGUGAGCGCUCACAAUUC
CACAGCGGCCGCUGAGGUAGAUAAUCUUUGUAUAGUGGGGGGGGAUCUCAUGUACCGGGUUUCU
UUUAUUUGAUUUUCAUAAAACAGACGGUAGCUAGGUUCGCAAGGCAGCCACAAGCCAAAGAU
GGUAGGGUGCUCAUAGUGAGUAGGGACAGUGCCUUUGAUUCACAACGCGUCAAUACCAUCUGA
CACGGAUACCCUUAACCGACUUGUCAUGAUCUCCAGACUUGUCCAAGGUGGACGGGCCACCUUU
ACUUAACCCGGAAGGAACAUUAUUAUUUAUUAUGUGUUCGGAAAAAAAAAAAAAAAAAAAA
AAAAA

502. TCA5_CBhBsi_GFP_GeFo3_R4

[1167] gggagacgguaaccgcuaggagcggugaccgGAAUUCGGCGGGAGUAACUAUGACUCUCUUAAGGAGUCAUAG
AGCCAGAACCUCUCGUGGUCCCGCUGGGCACAGGGAUUAUUUUUCUGUGGCAAAUUUGACUG
GCUUCAGAGAGCGUUUUUCGAAGUGGACUGUGUGACUGCGUUCUUUUUUAGUUGCUAUAUCCG
CUUCGAUUAACAUCUCACCUCGACGUUAAGAUAUUUCUAGAAGGUCGACCAGAUGUCCGAGG
UCGACCAGUUGUCCGUGUGGAAUUGUGAGCGCUCACAAUCCACACGUUACAUAAUACGUUA
AAUGGCCCGCCUGGCUGACCGCCCAACGACCCCGCCCAUUGACGUCAAUAGUAACGCCAAUAGG
GACUUUCCAUUGACGUCAAUGGGUGGAGUAUUUACGGUAAACUGCCCACUUGGCAGUACAUCAA
GUGUAUCAUAUGCCAAGUACGCCCCUAUUGACGUCAAUGACGGUAAAUGGCCCGCCUGGCAUU
GUGCCCAGUACAUGACCUUAUGGGACUUUCCUACUUGGCAGUACAUCUACGUUUAGUCAUCGC
UAUUACCAUGGUCGAGGUGAGCCCCACGUUCUGCUUCACUCUCCCCAUCUCCCCCCCCUCCCCACC
CCCAAUUUUUGUAUUUAUUUAUUUUUUAAUUAUUUUUGUGCAGCGAUGGGGGCGGGGGGGGGGGG
GGGCGCGCGCCAGGCGGGGCGGGGCGGGGCGAGGGGCGGGGCGAGGGCGGAGAGGUGCGG
CGGCAGCCAAUCAGAGCGGCGCGCUCCGAAAGUUUCCUUUUUUGGCGAGGCGGCGGCGGCGGCGG
CCCUAUA AAAAGCGAAGCGCGCGGGCGGGAGUCGUGCGACGCGUGCCUUCGCCCCGUGCCCC
GCUCCGCCGCCGCCUCGCGCCGCCCGCCCCGGCUCUGACUGACCGCGUUACUCCACAGGUGAGCG
GGCGGGACGGCCCUUCUCCUCCGGGCGUGAAUUAGCUGAGCAAGAGGUAAGGGUUUAAGGGAUG
GUUGGUUGGUGGGGUUUUAUGUUUAUUUACCUAGGAGCACCUGCCUGAAUACAUUUUUUUCAG
GUUGGCGUACGGCCACCAUGGUGAGCAAGGGCGAGGAGCUGUUCACCGGGGUGGUGCCCAUCCU
GGUCGAGCUGGACGGCGACGUAAACGGCCACAAGUUCAGCGUGUCCGGCGAGGGCGAGGGCGAU
GCCACCUACGGCAAGCUGACCCUGAAGUUAUCUGCACCACCGGCAAGCUGCCCCGUGCCCUGGCC
CACCCUCGUGACCACCCUGACCUACGGCGUGCAGUGCUUCAGCCGCUACCCCGACCAUGAAGC
AGCACGACUUCUUAAGUCCGCAUGCCCCGAAGGCUACGUCCAGGAGCGCACCAUCUUCUUAAG
GACGACGGCAACUACAAGACCCGCGCCGAGGUGAAGUUCGAGGGCGACCCUGGUGAACC GCAU

503. TCA5 CBhBsi GFP GeFo3 R4A22

238

[1169]

GUUGGCGUACGGCCACCAUGGUGAGCAAGGGCGAGGAGCUGUUCACCGGGGUGGUGCCCAUCCU
GGUCGAGCUGGACGGCGACGUAAACGGCCACAAGUUCAGCGUGUCCGGCGAGGGCGAGGGCGAU
GCCACCUACGGCAAGCUGACCCUGAAGUUCAUCUGCACCACCGGCAAGCUGCCCCGUGCCCUGGCC
CACCCUCGUGACCACCCUGACCUACGGCGUGCAGUGCUUCAGCCGCUACCCCGACCACAUGAAGC
AGCACGACUUCUUAAGUCCGCCAUGCCCCGAAGGCUACGUCCAGGAGCGCACCAUCUUCUUAAG
GACGACGGCAACUACAAGACCCGCGCCGAGGUGAAGUUCGAGGGCGACACCCUGGUGAACCGCAU
CGAGCUGAAGGGCAUCGACUUAAGGAGGACGGCAACAUCUGGGGCACAAGCUGGAGUACAAC
UACAACAGCCACAACGUCUAUAUCAUGGCCGACAAGCAGAAGAACGGCAUCAAGGUGAACUUA
AGAUCGCCCACAACAUCGAGGACGGCAGCGUGCAGCUCGCCGACCACUACCAGCAGAACACCCCC
AUCGGCGACGGCCCCGUGCUGCUGCCCCACAACCACUACCUGAGCACCCAGUCCGCCCCUGAGCAA
AGACCCCAACGAGAAGCGCGAUCACAUGGUCCUGCUGGAGUUCGUGACCGCCGCCGGGAUCACUC
UCGGCAUGGACGAGCUGUACAAGUAAAAGCUUGAUCCAGACAUGAUAAAGAUACAUUGAUGAGUU
UGGACAAACCACAACUAGAAUGCAGUGAAAAAAUUGCUUUUUUGUGAAAAUUUGUGAUGCUAUU
GCUUUUUUGUAACCAUUUAAGCUGCAAUAAACAAGUUGUGGAUUUGUGAGCGCUCACAAUUC
CACAGCGGCCCGUGAGGUAGAUAAUCUUUGUAUAGUGGGGGGGGAUCUCAUGUACCGGGUUUCU
UUUAUUUGAUUUUCAAUAAAACAGACGGUAGCUAGGUUCGCAAGGCAGCCACAAGCCAAAGUA
GGUAGGGUGCUCAUAGUGAGUAGGGACAGUGCCUUUUGAUUCACAACGCGUCAAUACCAUCUGA
CACGGAUACCCUUAACGGACUUGUCAUGAUCUCCAGACUUGUCCAAGGUGGACGGGCCACCUUU
ACUUAACCCGAAAAGGAACAUUAUUAUUUAUUAUGUGUUCGGAAAAUAGCAAAAAAAAAAAAAA
AAAAAAAAA

504. TCA5_CBhBsi_GFP_GeFo3_R20

gggagacggucaaccgcuaggagcggugaccgGAAUUCGGCGGGAGUAACU AUGACUCUCUUAAGGAGUCAUAG
AGCCAGAACCUCUCGUGGUCCCGCUGGGCACAGGGAUUAUUUUUCUGUGGCAAAUUUGACUG
GCUUCAGAGAGCGUUUUUCGAAGUGGACUGUGUGACUGCGUUCUUUUUUAGUUGCUAUAUCCG
CUUCGAUUAACAUCUCACCUCGACGUUAAGAUCAUUUCUAGAAGGUCGACCAGAUGUCCGAGG
UCGACCAGUUGUCCGUGUGGAAUUGUGAGCGCUCACAAUCCACACGUUACAUAACUUAACGGUA
AAUGGCCCCGCCUGGCUGACCGCCCAACGACCCCGCCCAUUGACGUCAAUAGUAACGCCAAUAGG
GACUUUCCAUUGACGUCAAUGGGUGGAGUAUUUACGGUAAACUGCCCACUUGGCAGUACAUCAA
GUGUAUCAUAUGCCAAGUACGCCCCCUAUUGACGUCAAUGACGGUAAAUGGCCCCGCCUGGCAUU
GUGCCCAGUACAUGACCUUAUGGGACUUCCUACUUGGCAGUACAUCUACGUAAUAGUCAUCGC
UAUUACCAUGGUCGAGGUGAGCCCCACGUUCUGCUUCACUCUCCCCAUCUCCCCCCCCUCCCCACC

[1170]

CCCAAUUUUGUAUUUAUUUAUUUUUUAAUUUAUUUUGUGCAGCGAUGGGGGCGGGGGGGGGGGGG
GGGCGCGCGCCAGGCGGGGCGGGGCGGGGCGAGGGGCGGGGCGGGGCGAGGCGGAGAGGUGCGG
CGGCAGCCAAUCAGAGCGGCGCGCUCCGAAAGUUUCCUUUUUUGGCGAGGCGGCGGGCGGGCGG
CCCUAUAAAAAGCGAAGCGCGCGGGGCGGGGAGUCGUGCGACGUGCCUUCGCCCCGUGCCCC
GCUCCGCCGCCGCCUCGCGCCGCCCGCCCCGGCUCUGACUGACCGCGUUACUCCACAGGUGAGCG
GGCGGGACGGCCCUUCUCCUCCGGGCGUGUAAUUAGCUGAGCAAGAGGUAAGGGUUUAAGGGAUG
GUUGGUUGGUGGGGUAAUAAUGUUUAAUUACCUGGAGCACCUGCCUGAAAUCACUUUUUUUUCAG
GUUGGCGUACGGCCACCAUGGUGAGCAAGGGCGAGGAGCUGUUCACCGGGGUGGUGCCCAUCCU
GGUCGAGCUGGACGGCGACGUAAACGGCCACAAGUUCAGCGUGUCCGGCGAGGGCGAGGGCGAU
GCCACCUACGGCAAGCUGACCCUGAAGUUCUUCGACCAACCGGCAAGCUGCCCGUGCCCGGGCC
CACCCUCGUGACCACCCUGACCUACGGCGUGCAGUGCUUCAGCCGCUACCCCGACCACAUGAAGC
AGCACGACUUCUUAAGUCCGCCAUGCCCGAAGGCUACGUCCAGGAGCGCACCAUCUUCUUAAG
GACGACGGCAACUACAAGACCCGCGCCGAGGUGAAGUUCGAGGGCGACACCCUGGUGAACCGCAU
CGAGCUGAAGGGCAUCGACUUAAGGAGGACGGCAACAUCUGGGGCGACAAGCUGGAGUACAAC
UACAACAGCCACAACGUCUAUAUCAUGGCCGACAAGCAGAAGAACGGCAUCAAGGUGAACUUA
AGAUCGCCACAACAUCGAGGACGGCAGCGUGCAGCUCGCCGACCACUACCAGCAGAACACCCCC
AUCGGCGACGGCCCCGUGCUGCUGCCCGACAACCACUACCUGAGCACCCAGUCCGCCCGAGCAA
AGACCCCAACGAGAAGCGCGAUCACAUGGUCCUGCUGGAGUUCGUGACCGCGCGGGGAUCACUC
UCGGCAUGGACGAGCUGUACAAGUAAAAGCUUGAUCCAGACAUGAUAAAGAUACAUUGAUGAGUU
UGGACAAACCACAACUAGAAUGCAGUGAAAAAAUAGCUUUUUUUGUGAAAUUUGUGAUGCUAUU
GCUUUUUUUGUAACCAUUUAAGCUGCAAUAAACAAGUUGUGGAAUUGUGAGCGCUCACAAUUC
CACAGCGGCCCGCUGAGGUAGAUAAUCUUUGUAUAGUGGGGGGGGAUCUCAUGUACCGGGUUUCU
UUUAUUUGAUUUUCAAUAAAACAGACGGUAGCUAGGUUCGCAAGGCAGCCACAAGCCAAAGUA
GGUAGGGUGCUCAUAGUGAGUAGGGACAGUGCCUUUUGAUUCACAACGCGUCAAUACCAUCUGA
CACGGAUACCCUUACCGGACUUGUCAUGAUCUCCAGACUUGUCCAAGGUGGACGGGCCACCUUU
ACUUAACCCGGAAGGAACAUAUAUUAAUUAUUGUGUUCGAAAA
UAGCCAAAUGCCUCGUCAUC

505. TCA5_CBhBsi_GFP_GeFo3_R20A22

gggagacgguaaccgcuaggagcggugaccgGAAUUCGGCGGGAGUAACUAUGACUCUCUUAAGGAGUCAUAG
AGCCAGAACCUCUCGUGGUCCCGCUGGGCACAGGGAUUAAUUUUUCUGUGGCAAAUUUGACUG
GCUUCAGAGAGCGUUUUUCGAAGUGGACUGUGUGACUGCGUCCCCCUUAGUUGCUAUAUCCG

[1171]

CUUCGAUUAAACAUCUCACCUCGACGUAAAGAUAUUUCUAGAAGGUCGACCAGAUGUCCGAGG
UCGACCAGUUGUCCGUGUGGAAUUGUGAGCGCUCACAAUCCACACGUUACAUAAACUUACGGUA
AAUGGCCCCGCCUGGCUGACCGCCCAACGACCCCCGCCAUUGACGUCAAUAGUAACGCCAAUAGG
GACUUUCCAUGACGUCAAUGGGUGGAGUAUUUACGGUAAACUGCCCACUUGGCAGUACAUA
GUGUAUCAUAUGCCAAGUACGCCCCUAUUGACGUCAAUGACGGUAAAUGGCCCCGCCUGGCAUU
GUGCCCAGUACAUGACCUUAUGGGACUUUCCUACUUGGCAGUACAUCUACGUAAUAGUCAUCGC
UAUUACCAUGGUCGAGGUGAGCCCCACGUUCUGCUUCACUCUCCCCAUCUCCCCCCCCUCCCCACC
CCCAAUUUUGUAUUUAUUUAUUUUUAAUUAAUUUUGUGCAGCGAUGGGGGCGGGGGGGGGGGGG
GGGCGCGCGCCAGGCGGGGCGGGGCGGGGCGAGGGGCGGGGCGGGGCGAGGCGGAGAGGUGCGG
CGGCAGCCAAUCAGAGCGGCGCGCUCCGAAAGUUUCCUUUUUUGGCGAGGCGGCGGGCGGCGG
CCCUAUAAAAAGCGAAGCGCGCGGGCGGGGAGUCGCGUGCGACGCGUGCCUUCGCCCCGUGCCCC
GCUCCGCCGCCGCCUCGCGCCGCCCGCCCCGGCUCUGACUGACCGCGUUACUCCACAGGUGAGCG
GGCGGGACGGCCUUCUCCUCCGGGCGUGAAUUAGCUGAGCAAGAGGUAAGGGUUUAAGGGAUG
GUUGGUUGGUGGGGUAAUAAUGUUAAUUACCUGGAGCACCUGCCUGAAAUCACUUUUUUUCAG
GUUGGCGUACGGCCACCAUGGUGAGCAAGGGCGAGGAGCUGUUCACCGGGGUGGUGCCCAUCCU
GGUCGAGCUGGACGGCGACGUAAACGGCCACAAGUUCAGCGUGUCCGGCGAGGGCGAGGGCGAU
GCCACCUACGGCAAGCUGACCCUGAAGUUAUCUGCACCACCGGCAAGCUGCCCCGUGCCCUGGCC
CACCCUCGUGACCACCCUGACCUACGGCGUGCAGUGCUUCAGCCGCUACCCCGACCACAUGAAGC
AGCACGACUUCUUAAGUCCGCCAUGCCCCGAAGGCUACGUCCAGGAGCGCACCAUCUUCUUAAG
GACGACGGCAACUACAAGACCCGCGCCGAGGUGAAGUUCGAGGGCGACACCCUGGUGAACCGCAU
CGAGCUGAAGGGCAUCGACUUAAGGAGGACGGCAACAUCUGGGGCACAAGCUGGAGUACAAC
UACAACAGCCACAACGUCUAUAUCAUGGCCGACAAGCAGAAGAACGGCAUCAAGGUGAACUUA
AGAUCGCCACAACAUCGAGGACGGCAGCGUGCAGCUCGCCGACCACUACCAGCAGAACACCCCC
AUCGGCGACGGCCCCGUGCUGCUGCCCCGACAACCACUACCUGAGCACCCAGUCCGCCUGAGCAA
AGACCCCAACGAGAAGCGCGAUCACAUGGUCCUGCUGGAGUUCGUGACCGCCGCCGGGAUCACUC
UCGGCAUGGACGAGCUGUACAAGUAAAAGCUUGAUCCAGACAUGAUAAAGAUACAUUGAUGAGUU
UGGACAAACCACAACUAGAAUGCAGUGAAAAAAUUGCUUUAAUUUGUGAAAUUUGUGAUGCUAUU
GCUUUAAUUUGUAACCAUUAUAAGCUGCAAUAAACAAGUUGUGGAAUUGUGAGCGCUCACAAUUC
CACAGCGGCCGUGAGGUAGAUAAUCUUUGUAUAGUGGGGGGGGAUCUCAUGUACCGGGUUUCU
UUUAAUUUGAUUUUCAAUAAAACAGACGGUAGCUAGGUUCGCAAGGCAGCCACAAGCCAAAGUA
GGUAGGGUGCUCAUAGUGAGUAGGGACAGUGCCUUUUGAUUCACAACGCGUCAAUACCAUCUGA
CACGGAUACCCUUAACGGACUUGUCAUGAUCUCCAGACUUGUCCAAGGUGGACGGGCCACCUU

ACUUAACCCGGAAGGAACAUUAUUAUUAUUAUGUGUUCGGAAAA
UAGCCAAAUGCCUCGUCAUC AAAAAAAAAAAAAAAAAAAAAA

506. CBhBsi_GPF_GeFo3_R4A22

[1172] AGGUCGACCAGAUGUCCGAGGUCGACCAGUUGUCCGUGUGGAAUUGUGAGCGCUCACAAUCCA
CACGUUACAUAAUACGGUAAAUGGCCCCGCGGCGUACCGCCCAACGACCCCCGCCCAUUGAC
GUCAAUAGUAACGCCAAUAGGGACUUUCCAUGACGUCAAUGGGUGGAGUAUUUACGGUAAACU
GCCCACUUGGCAGUACAUAAGUGUAUCAUAUGCCAAGUACGCCCCUAUUGACGUCAAUGACGG
UAAAUGGCCCCGCGGCAUUGUGCCCAGUACAUGACCUUAUGGGACUUUCCUACUUGGCAGUAC
AUCUACGUAAUAGUCAUCGCUAAUACCAUGGUCGAGGUGAGCCCCACGUUCUGCUACUCUCCC
CAUCUCCCCCCCCUCCCCACCCCCAAUUUUGUAUUUAUUUAUUUUUAAUUUAUUUUGUGCAGCGA
UGGGGGCGGGGGGGGGGGGGGGCGCGGCCAGGCGGGCGGGCGGGGCGAGGGCGGGGCGG
GGCGAGGCGGAGAGGUGCGGCGGCAGCCAAUCAGAGCGGCGCGCUCCGAAAGUUUCCUUUAUG
GCGAGGCGGCGGCGGCGGCGGCCUUA AAAAAGCGAAGCGCGCGGCGGGCGGGAGUCGUGCGAC
GCUGCCUUCGCCCCGUGCCCCGCUCCGCCGCCGCCUCGCGCCGCCCGCCCCGGCUCUGACUGACCG
CGUUACUCCCACAGGUGAGCGGGCGGGACGGCCCUUCUCCUCCGGGCGUGAAUUAGCUGAGCAAG
AGGUAAGGGUUUAAGGGAUGGUUGGUUGGUGGGGUAUUAAUGUUUAAUUACCUGGAGCACCUGC
CUGAAAUACAUUUUUUUCAGGUUGGCGUACGGCCACCAUGGUGAGCAAGGGCGAGGAGCUGUUC
ACCGGGGUGGUGCCCAUCCUGGUCGAGCUGGACGGCGACGUAAACGGCCACAAGUUCAGCGUGUC
CGGCGAGGGCGAGGGCGAUGCCACCUACGGCAAGCUGACCCUGAAGUUAUCUGCACCACCGGCA
AGCUGCCCCGUGCCCUGGCCCACCCUCGUGACCACCCUGACCUACGGCGUGCAGUGCUUCAGCCGC
UACCCCGACCACAUGAAGCAGCACGACUUCUUAAGUCCGCCAUGCCCGAAGGCUACGUCCAGGA
GCGCACCAUCUUCUUAAGGACGACGGCAACUACAAGACCCGCGCCGAGGUGAAGUUCGAGGGCG
ACACCCUGGUGAACCGCAUCGAGCUGAAGGGCAUCGACUUAAGGAGGACGGCAACAUCUGGG
GCACAAGCUGGAGUACAACUACAACAGCCACAACGUCUAUAUCAUGGCCGACAAGCAGAAGAAC
GGCAUCAAGGUGAACUUAAGAUCGCCACAACAUCGAGGACGGCAGCGUGCAGCUCGCCGACCA
CUACCAGCAGAACACCCCCAUCGGCGACGGCCCCGUGCUGCUGCCCCACAACCACUACCUGAGCA
CCCAGUCCGCCUGAGCAAAGACCCCAACGAGAAGCGCGAUCACAUGGUCCUGCUGGAGUUCGUG
ACCGCCGCCGGGAUCACUCUGGCAUGGACGAGCUGUACAAGUAAAAGCUUGAUCCAGACAUGA
UAAGAUACAUGAUGAGUUUGGACAAACCACAACUAGAAUGCAGUGAAAAAAUGCUUUUAUUUG
UGAAAUUUGUGAUGCUAUUGCUUUAUUUGUAACCAUUAUAAGCUGCAAUAAACAAGUUGUGGAA
UUUGAGCGCUCACAAUCCACAGCGGCCGCGUGAGGUAGAUAAUCUUUGUAUAGUGGGGGGGGA

UCUCAUGUACCGGUUUUCUUUUUUUUUGAUUUUCAAUAAAAACAGACGGUAGCUAGGUUCGCAAGG
CAGCCACAAGCCAAAGAUAGGUAGGGUGUCUAGUGAGUAGGGACAGUGCCUUUUGAUUCACA
ACGCGUCAAUACCAUCUGACACGGAUACCCUUACCGGACUUGUCAUGAUCUCCAGACUUGUCCA
AGGUGGACGGGCCACCUUUACUUAACCCGAAAAAGGAACAUUAUUAUUUAUUAUGUGUUCGGAA
AAUAGCAAAAAAAAAAAAAAAAAAAAAA

507. TCA5_CBh_NLSGPF_ZoAl3_R4A22

gggagacggucaaccgcuaggagcggugaccgGAAUUCGGCGGGAGUAAACUAUGACUCUCUUAAGGAGUCAUAG
AGCCAGAACCUCUCGUGGUCCCGCUGGGCACAGGGAUUAUUUUUCUGUGGCAAAUUUGACUG
GCUUCAGAGAGCGUUUUUCGAAGUGGACUGUGUGACUGCGUCCCCCUUAGUUGCUAUAUCCG
CUUCGAUUAACAUCUCACCUCGACGUUAAGAUCAUU

TCTAGAAGGTCGACCAGATGTCCGAGGTCGACCAGTTGTCCGTGTGGAATTGTGAGCGCTCACAATTC
CACA Cgt tacataact acggtaaatg gccgcctgg ctgaccgcc aacgaccccc gccattgac gtaaatagta acgcaatag ggactttcca
ttgacgtcaa tgggtggagt atttacggta aactgccac ttggcagtac atcaagtga tcatatgcca agtaccccc ctattgacgt caatgacggg
aaatggcccc cctggcattg tgcccagtac atgacctat gggactttcc tacttggcag tacatctacg tattagtcat cgctattacc atggtcgagg
tgagccccac gttctgctc acttccccca tctcccccc cteccaccc ccaattttgt atttattat ttttaatta ttttgtagc cgatgggggc
gggggggggg gggggggcgc cgccaggcgg ggcggggcgg ggcgaggggc ggggcggggc gaggcggaga ggtgcggcgg cagccaatca
gagcggcgcg ctccgaaagt ttcctttat ggcgaggcgg cgcgggcgcg gccctataa aaagcgaagc gcgcggcggg cgggagtcgc
tgcgacgctg ccttcgcccc gtgccccgt cgcggcggc ctcgcgccgc cgcggcggc tctgactgac cgcgttact ccaagggtga
gcggcgggga cgcccttct cctccgggct gtaattagct gagcaagagg taagggtta agggatggtt ggttggtggg gtattaatgt ttaattacct
ggagcacctg cctgaaatca ctttttca ggttga agctt GCCACCATG gcc cca aag aag aag cgg aag
gtcgtgagcaaggcgagcgagctgttaccggggtgtgccatcctgtgcgagctggacggcgacgtaaacggccacaagttcagcgtgtccggcgaggcgga
ggcgatgccacctacggcaagctgacctgaagttcatctgcaccaccggcaagctgccgtgccctggccaccctctgaccacctgacctacggcgtgca
gtgcttaccgctacccccgaccatgaagcagcagcacttcttaagtcgccatcccgaaggctacgtccaggagcgaccatcttcttaaggacgacggca
actacaagaccgcgccgaggtgaagttcaggggcgacacctggtgaaccgcatcgagctgaaggcgatcgacttcaaggaggacggcaacatcctggggcac
aagctggagtacaactacaagccacaacgtctatatcatggcgacaagcagaagaacggcatcaagggtgaacttcaagatccgcacaacatcgaggacggc
agcgtgcagctcgcgaccactaccagcagaacccccatcgcgacggccccgtgctgctgccgacaaccactacctgagcaccagtcgccccctgagcaa
agacccaacgagaagcgcatcacatggtcctgctggagtctgtgaccgccggggatcactctcgcatggacgagctgtacaagtaaaagcttGatccaga
catgataagatacattgatgatttgacaaccacaactagaatgcagtgaaaaaatgctttttgtgaaatttgatgctattgctttattgtaaccattataagctgc
aataaacaagtTGTGGAATTGTGAGCGCTCACAATTCACAGCGCCGCTaggtagtcacattgcacttctgtaactgactgggt
gtgggatgtggcctgggggtgtgggttatgggtatatgtggatattctggtgggaatgtccattcactgtatgcctatcttttaataaaaagacggtagctaggttc
gcgaagcagccacaagccaatagccagtagtagctcatagtggttaggtgacaggaaaccttgactcagaacgcgtccattaacatcagaacggaccaaaacttc

[1173]

ggacatgcaccgattaaccggatttgcgaagtgacgggcccacttacttaaccggaaagggaacatatatagttatgtgttcgtaata TAGC
AAAAAAAAAAAAAAAAAAAAA

508. TCA5_CBh_NLSGPF_GeFo3_R4A22

gggagacgggucaaccgcuaggagcggugaccgGAAUUCGGCGGGAGUAACUAUGACUCUCUUAAGGAGUCAUAG
AGCCAGAACCUCUCGUGGUCCCGCUGGGCACAGGGAUUAUUUUUCUGUGGCAAAUUUGACUG
GCUUCAGAGAGCGUUUUUCGAAGUGGACUGUGUGACUGCGUCCCCCUUAGUUGCUAUAUCCG
CUUCGAUUAACAUCUCACCUCGACGUUAUAAGAUAU
TCTAGAAGGTCGACCAGATGTCCGAGGTCGACCAGTTGTCCGTGTGGAATTGTGAGCGCTCACAATTC
CACA Cgt tacataact acggtaaatg gcccgcttg ctgaccgcc aacgacccc gccattgac gtcaatagta acgccaatag ggacttcca
ttgacgtcaa tgggtggagt attfacgga aactgccac ttggcagtac atcaagtga tcatatgcca agtacgccc ctattgacgt caatgacggt
aatggcccg cctggcattg tgcccagtac atgacctat gggacttcc tacttggcag tacatctacg tattagtcac cgctattacc atggtcgagg
tgagccccc gttctgctc actctccca tctccccc cccccccc ccaatttgt attatttat ttttaatta tttgtgcag cgatgggggc
gggggggggg gggggggcgc cgccaggcgg ggcggggcgg ggcgaggcgg gggcggggc gaggcggaga ggtgcggcgg cagccaatca
gagcggcgcg ctccgaaagt ttcctttat ggcgaggcgg cgcggcggc gccctataa aaagcgaagc gcgcgcggg cgagggtcgc
tgcgagcgtg ccttcgccc gtgcccgc cgcggcgc ctcgcgccg cgcggcggc tctgactgac cgcgttact ccaagggtga
gcggggggga cggccctct cctccgggt gtaattagct gagcaagagg taagggtta agggatggti ggttggtggg gtattaatgt ttaattacct
ggagcacctg cctgaaatca ctttttca ggttga agctt GCCACCATG gcc cca aag aag aag cgg aag
gtcgtgagcaaggcggaggagctgttaccgggtgtgcccactctgtcgtgagctggacggcgaacggccacaagttcagcgtgtccggcgaggcga
ggcgatgccacctagcgaagctgacctgaagtcatctgcaccaccggcaagctgcccgtgccctggccaccctctgaccacctgacctacggcgtga
gtgttcacggctaccgcaccacatgaagcagcagacttcttaagtcgccatgccgaaggctacgtccaggagcgcaccatcttcttaaggagcagggca
actacaagaccgcgccgaggtgaagttcagggcgacacctgtgtgaacgcacatcagctgaaggcgcactcaaggaggacggcaacatcctggggcac
aagctggagtacaactacaacagccacaacgtctatatcatggcgacaagcagaagaacggcatcaaggtgaactcaagatccgcacaacatcgaggacggc
agcgtgcagctcgcgaccactaccagcagaacacccccatggcgacggccccgtgctgctgcccacaaccactacctgagcaccagtcgcccctgagcaa
agacccaacgagaagcgcgatcacatggtcctgtggtgagttcgtgaccgcccgggatcactctggcatggagcagctgtacaagtaaagcttGatccaga
catgataagatacattgatgagttggacaaccacaactagaatgcagtgaaaaaatgcttatttgtgaaattgtgatgctattgcttatttgaaccattataagctgc
aataacaagtTGTGGAATTGTGAGCGCTCACAATTCACAGCGCCGCGUGAGGUAGAUAAUCUUUGUAU
AGUGGGGGGGGAUCUCAUGUACCGGGUUUCUUUAUUUGAUUUUCAUAAAACAGACGGUAGCU
AGGUUCGCAAGGCAGCCACAAGCCAAAGAUAGGUAGGGUGCUAUAGUGAGUAGGGACAGUGCC
UUUUGAUUCACAACGCGUCAAUACCAUCUGACACGGAUACCCUUAACCGGACUUGUCAUGAUCUCC
CAGACUUGUCCAAGGUGGACGGGCCACCUUACUUAACCCGGAAGGAACAUUAUUAUUAAUUAU
AUGUGUUCGAAAAUAGCAAAAAAAAAAAAAAAAAAAAAA

[1174]

509. TCA5_ZoAl3

gggagacggucaaccgcuaggagcggugaccgGAAUUCGGCGGGAGUAACUAUGACUCUCUUAAGGAGUCAUAG
AGCCAGAACCUCUCGUGGUCCCGCUGGGCACAGGGAUUAUUUUUCUGUGGCAAAUUUGACUG
GCUUCAGAGAGCGUUUUUCGAAGUGGACUGUGUGACUGCGUUCUUUUUUUAGUUGCUAUAUCCG
CUUCGAUUAACAUCUCACCUCGACGUUAUAGAUAUUUCUAGAAGGUCGACCAGAUGUCCGAGG
UCGACCAGUUGUCCGUGUGGAAUUGUGAGCGCUCACAAUCCACACGUUACAUAACUUACGGUA
AAUGGCCCCGCCUGGCUGACCGCCCAACGACCCCCGCCAUUGACGUCAAUAGUAACGCCAAUAGG
GACUUUCCAUUGACGUCAAUGGGUGGAGUAUUUACGGUAAACUGCCCACUUGGCAGUACAUA
GUGUAUCAUAUGCCAAGUACGCCCCUAUUGACGUCAAUGACGGUAAAUGGCCCCGCCUGGCAUU
GUGCCCAGUACAUGACCUUAUGGGACUUUCCUACUUGGCAGUACAUCUACGUAAUAGUCAUCGC
UAUUACCAUGGUCGAGGUGAGCCCCACGUUCUGCUUCACUCUCCCCAUCUCCCCCCCCUCCCCACC
CCCAAUUUUUGUAUUUAUUUAUUUUUAAUUUUUUUGUGCAGCGAUGGGGGCGGGGGGGGGGGGG
GGGCGCGCGCCAGGCGGGGCGGGGCGGGGCGAGGGGCGGGGCGGGGCGAGGCGGAGAGGUGCGG
CGGCAGCCAAUCAGAGCGGCGCGCUCCGAAAGUUUCCUUUUUUGGCGAGGCGGCGGCGGCGGCGG
CCCUAUAAAAAGCGAAGCGCGCGGGCGGGGAGUCGCGUGCGACGCGUCCUUCGCCCCGUGCCCC
[1175] GCUCCGCCGCCGCCUCGCGCCGCCCGCCCCGGCUCUGACUGACCGCGUUACUCCCACAGGUGAGCG
GGCGGGACGGCCCUUCUCCUCCGGGCGUGAAUUAGCUGAGCAAGAGGUAAGGGUUUAAGGGAUG
GUUGGUUGGUGGGGUAAUUAUGUUUAUUUACCUGGAGCACCUGCCUGAAAUACUUUUUUUUCAG
GUUGGCGUACGGCCACCAUGGUGAGCAAGGGCGAGGAGCUGUUCACCGGGGUGGUGCCCAUCCU
GGUCGAGCUGGACGGCGACGUAAACGGCCACAAGUUCAGCGUGUCCGGCGAGGGCGAGGGCGAU
GCCACCUACGGCAAGCUGACCCUGAAGUUAUCUGCACCACCGGCAAGCUGCCCCGUGCCCUGGCC
CACCCUCGUGACCACCCUGACCUACGGCGUGCAGUGCUUCAGCCGCUACCCCGACCACAUGAAGC
AGCACGACUUCUUAAGUCCGCCAUGCCCCGAAGGCUACGUCCAGGAGCGCACCAUCUUCUUAAG
GACGACGGCAACUACAAGACCCGCGCCGAGGUGAAGUUCGAGGGCGACACCCUGGUGAACCGCAU
CGAGCUGAAGGGCAUCGACUUAAGGAGGACGGCAACAUCUGGGGCACAAGCUGGAGUACAAC
UACAACAGCCACAACGUCUAUAUCAUGGCCGACAAGCAGAAGAACGGCAUCAAGGUGAACUUA
AGAUCCGCCACAACAUCGAGGACGGCAGCGUGCAGCUCGCCGACCACUACCAGCAGAACACCCCC
AUCGGCGACGGCCCCGUGCUGCUGCCCCACAACCACUACCUGAGCACCCAGUCCGCCUGAGCAA
AGACCCCAACGAGAAGCGCGAUCACAUGGUCCUGCUGGAGUUCGUGACCGCCGCCGGGAUCACUC
UCGGCAUGGACGAGCUGUACAAGUAAAAGCUUGAUCCAGACAUGAUAAAGAUACAUAUGAGUU
UGGACAAACCACAACUAGAAUGCAGUGAAAAAAUUGCUUUAUUUGUGAAAUUUGUGAUGCUAUU

GCUUUUAUUUGUAACCAUUUAAGCUGCAAUAAACAAGUUGUGGAAUUGUGAGCGCUCACAAUUC
CACAGCGGCCGCTaggtagtcacattgcactttctgtaactgcactgggtgtgggatgtggcctggggtgtgggttatgggtatatgtgggatattctg
gtgggaatgtccattcactgtatgcctatcttttaataaaaagacggtagctaggttcggaagcagccacaagccaatagccagtaggttagctcatagtggttaggt
gacaggaacctttgactcagaacgcgtccattaacatctagaacggaccaaaactcggacatgcaccgattaaccggattgtccaaggtggacgggccaccttact
aaccggaaaggaacatatatagttatgtgttcgtaata TAGC AAAAAAAAAAAAAAAAAAAAAA

510. TCA5_TaGu3

gggagacggucaaccgcuaggagcggugaccgGAAUUCGGCGGGAGUAACUAUGACUCUCUUAAGGAGUCAUAG
AGCCAGAACCUCUCUGUGUCCCGCUGGGCACAGGGAUUAUUUUUCUGUGGCAAAUUUGACUG
GCUUCAGAGAGCGUUUUUCGAAGUGGACUGUGUGACUGCGUUCUUUUUUUAGUUGCUAUAUCCG
CUUCGAUUAACAUCUCACCUCGACGUUAAGAUAUUUCUAGAAGGUCGACCAGAUUGUCCGAGG
UCGACCAGUUGUCCGUGUGGAAUUGUGAGCGCUCACAAUCCACACGUUACAUAAUUACGGUA
AAUGGCCCCGCCUGGCUGACCGCCCAACGACCCCCGCCAUUGACGUCAAUAGUAACGCCAAUAGG
GACUUUCCAUUGACGUCAAUGGGUGGAGUAUUUACGGUAAACUGCCCACUUGGCAGUACAUCAA
GUGUAUCAUAUGCCAAGUACGCCCCUAUUGACGUCAAUGACGGUAAAUGGCCCCGCCUGGCAUU
GUGCCCAGUACAUGACCUUAUGGGACUUUCCUACUUGGCAGUACAUCUACGUAAUAGUCAUCGC
[1176] UAUUACCAUGGUCGAGGUGAGCCCCACGUUCUGCUUCACUCUCCCCAUCUCCCCCCCCUCCCCACC
CCCAAUUUUUGUAUUUAUUUAUUUUUUAAUUAUUUUGUGCAGCGAUGGGGGCGGGGGGGGGGGGG
GGGCGCGCGCCAGGCGGGCGGGCGGGCGAGGGCGGGCGGGCGAGGCGGAGAGGUGCGG
CGGCAGCCAAUCAGAGCGGCGCGCUCCGAAAGUUUCCUUUUUUGGCGAGGCGGCGGCGGCGG
CCCUAUA AAAAGCGAAGCGCGCGGGCGGGAGUCGCGUGCGACGCGUCCUUGCCCCGUGCCCC
GCUCCGCCGCCGCCUCGCGCCGCCCGCCCCGGCUCUGACUGACCGCGUUACUCCACAGGUGAGCG
GGCGGGACGGCCUUCUCCUCCGGGCGUGAAUUAGCUGAGCAAGAGGUAAGGGUUUAAGGGAUG
GUUGGUUGGUGGGGUAAUUAUGUUUAUUUACCUGGAGCACCUGCCUGAAAUCACUUUUUUUUCAG
GUUGGCGUACGGCCACCAUGGUGAGCAAGGGCGAGGAGCUGUUCACCGGGGUGGUGCCCAUCCU
GGUCGAGCUGGACGGCGACGUAAACGGCCACAAGUUCAGCGUGUCCGGCGAGGGCGAGGGCGAU
GCCACCUACGGCAAGCUGACCCUGAAGUUAUCUGCACCACCGGCAAGCUGCCCCGUGCCCUGGCC
CACCCUCGUGACCACCCUGACCUACGGCGUGCAGUGCUUCAGCCGCUACCCCGACCACAUGAAGC
AGCACGACUUCUUAAGUCCGCCAUGCCCCGAAGGCUACGUCCAGGAGCGCACCAUCUUCUUAAG
GACGACGGCAACUACAAGACCCGCGCCGAGGUGAAGUUCGAGGGCGACACCCUGGUGAACCGCAU
CGAGCUGAAGGGCAUCGACUUAAGGAGGACGGCAACAUCUGGGGCACAAGCUGGAGUACAAC
UACAACAGCCACAACGUCUAUAUCAUGGCCGACAAGCAGAAGAACGGCAUCAAGGUGAACUUA

247

[1178]

gggagacggucaaccgcuaggagcggugaccgGAAUUCGGCGGGGAGUAAACU AUGACUCUCUU AAGGAGUCAUAG
AGCCGAAACCUCCUCGUGGUCCCGCUGGGCACAGGGAUUAAUUUUUCUGUGGCAAAUUGACUG
GCUUCAGAGAGCGUUUUUCGAAGUGGACUGUGUGACUGCGUUC CCCCCCUUAGUUGCUAUAUCCG
CUUCGAUUAACAUCUCACCUCGACGUAAAGAUCAUUUCUAGAAGGUCGACCAGAUGUCCGAGG
UCGACCAGUUGUCCGUGUGGAAUUGUGAGCGCUCACAAU UCCACACGUUACAUAACUUAACGGUA
AAUGGCCC GCCUGGCUGACCGCCCAACGACCCCCGCCCAUUGACGUCAAUAGUAACGCCAAUAGG
GACUUUCCAUUGACGUCAAUGGGUGGAGUAUUUACGGUAAACUGCCCACUUGGCAGUACAUCAA
GUGUAUCAUAUGCCAAGUACGCCCCCUAUUGACGUCAAUGACGGUAAAUGGCCCCGCCUGGCAUU
GUGCCCAGUACAUGACCUUAUGGGACUUUCCUACUUGGCAGUACAUCUACGUAAUAGUCAUCGC
UAUUACCAUGGUCGAGGUGAGCCCCACGUUCUGCUUCACUCUCCCCAUCUCCCCCCCCUCCCCACC
CCCAAUUUUGUAUUUAUUUAUUUUUAUUUAUUUUGUGCAGCGAUGGGGGCGGGGGGGGGGGGG
GGGCGCGCGCCAGGCGGGGCGGGGCGGGGCGAGGGGCGGGGCGGGGCGAGGCGGAGAGGUGCGG
CGGCAGCCAAUCAGAGCGGCGCGCUCCGAAAGUUUCCUUUAUGGCGAGGCGGCGGCGGCGGCGG
CCCUAUA AAAAGCGAAGCGCGCGGGCGGGGAGUCGCGACGCGUCCUUCGCCCCCGUGCCCC

gggagacgggucaaccgcguaggagcggugaccgGAAUUCGGCGGGAGUAACUAUGACUCUCUUAAGGAGUCAUAG
AGCCGAAACCUCCUCGUGGUCCCGCUGGGCACAGGGAUUAUUUUUCUGUGGCAAAUUUGACUG
GCUUCAGAGAGCGUUUUUCGAAGUGGACUGUGUGACUGCGUUCUUUUUUUUAGUUGCUAUAUCCG
CUUCGAUUAACAUCUCACCUCCGACGUAAAGAUCAUUUUCUAGAAGGUCGACCAGAUGUCCGAGG
UCGACCAGUUGUCCGUGUGGAAUUGUGAGCGCUCACAAUCCACACGuuacauaacuuacgguaaauggcccg
cuggcugaccgcccacgacccccgccauugacgucaauaguaaacgccauagggacuuuccauugacgucaauugggugggaguuuuacgguaaac
ugcccacuuugcaguuacaucaaguguaucuuuagccaaugacccccuuuugacgucaauugacgguaaauggcccgccggaauugugcccauuac

[illegible]

gggagagcggucaaccgcguaggagcggugaccgGAAUUCGGCGGAGUAACUAUGACUCUacCUUAAAGGAGUCAUAG
AGCCGAAACCUCCUCGUGGUCCCGCUGGGCACAGGGAUUAUUUUUUCUGUGGCAAAUUUGACUG
GCUUCAGAGAGCGUUUUUCGAAGUGGACUGUGUGACUGCGUUCUUUUUUUUAGUUGCUUAUUCG
CUUCGAUUAACAUCUCACCUUCGACGUAUAAGAUCAUUGCGGGCCGCUGUGGAAUUGUGAGCGCUC
ACAAUCCACAacuuuuuuuugcagcuuuuuuuuugguuacaaauaaagcauuagcaucacaaauuucacaaauaaagcauuuuuuuucacugc
auucuguuuggguuuguccaaacucaucaauguaucuuuaucaugucuggauCaagcuuuuacuuuguacagcucguccaugccgagagugaucccg
gcggcgguacacgaacuccagcaggaccaugugaucgcgcguucucguuuuggggucuuuugcucagggcgagacuggggugucagguagugguuugcggg
cagcagcacggggccgucgcgaugggggguguucugcugguaguggucggcgagcugcacgcugccgucuccgauguuugggcggaucuuugaagu
ucaccuuugaugccguuuucucguuuugcggccaugauauagacguuugggcguuuuguaguuguuacuccagcuuugugccccagggauguuugccgucc

[illegible]

[1182]

GGCGGGCGCCCGAAGGUCCUCCGGAGGCCCGGCAUUCUGCACGCUUCAAAGCGCACGUCUGCCG
CGCUGUUCUCCUCUCCUCAUCUCCGGGCCUUUCGACCUaagcuuGCCACCAUGgccccaaagaagaagcggaa
ggucgugagcaaggcgaggagcuguucaccggggugggucccauccugugcagcugggacggcgacguaaacggccacaaguucagcuguccgg
cgaggcgaggcggaugccaccuacggcaagcugaccugaaugucaucugcaccaccggcaagcugcccugcccugggccaccucgugaccacc
ugaccuacggcgugcagugcuucagccguaccccgaccacaugaagcagcagcguucuucaaguccccaugcccgaaggcuacguccaggagcg
accaucuuucaagagcagcgcaacuacaagaccggcgccgaggugaaguucgagggcgacaccucggugaaccgcaucgagcugaaggcgcaucga
cuucaaggaggacggcaacuuccggggcacaagcugggaguacaacuacaacgccacaacgucuauaucaugggcgaagcagaagaacggcauca
aggugaacuuaagauccgccacaacaucaggagcgagcgugcagcucggcaccacuaccagcagaacacccccaucggcgagggcccgugcug
cugcccgacaaccacuaccugagcaccaguccggcgagcgaagaccccaacgagaagcggaucacauagguccugcuggaguucgugaccggcg
cgggaucaucucggcaugagcagcuguacaaguaaaagcuuacuuguuuuuugcagcuuauaauaggguuacaaauaaagcaauagcauacaaau
uucacaaauaaagcauuuCuucacugcauucaguugugguuuguccaaacucauauaauaUGUGGAAUUGUGAGCGCU
CACAAUCCACAGCGGCCGCGUGAGGUAGAUAAUCUUUGUAUAGUGGGGGGGGAUCUCAUGUACC
GGGUUUCUUUUUAUUUGAUUUUCAAUAAAACAGACGGUAGCUAGGUUCGCAAGGCAGCCACAAGC
CAAAGAUAGGUAGGGUGCUCAUAGUGAGUAGGGACAGUGCCUUUUGAUUCACAACGCGUCAUA
CCAUCUGACACGGAUACCCUUAACCGGACUUGUCAUGAUCUCCAGACUUGUCCAAGGUGGACGGG
CCACCUUUACUUAACCCGGAAAAGGAACAUAUAUUAAUUAUAUGUGUUCGGAAAAUAGCAAAAA
AAAAAAAAAAAAAAAAA

517. EFS-GFP

gggagacggucaaccgcuaggagcggugaccgGAAUUCGGCGGGAGUAACUAUGACUCUCUUAAGGAGUCAUAG
AGCCAGAACCUCUCGUGGUCCCGCUGGGCACAGGGAUUAAUUUUUCUGUGGCAAAUUUGACUG
GCUUCAGAGAGCGUUUUUCGAAGUGGACUGUGUGACUGCGUUCUUUUUUAGUUGCUAUAUCCG
CUUCGAUUAACAUCUCACCUCGACGUUAAGAUCAUUGCGGCCGCGUGUGGAAUUGUGAGCGCUC
ACAAUCCACAAAAACCUCCACAUUCCCCUGAACCUGAAACAUAAAAUGAAUGCAAUUGUUG
UUGUUAACUUGUUUAUUGCAGCUUAUAAUGGUUACAAAUAAGCAAUAGCAUCACAAAUUUCAC
AAUAAAGCUAGAUGGUUACAAAUAAGCAAUAGCAUCACAAUUCACAAAUAACACCACGG
AAUUGUCAGUGCCCAACAGCCGAGCCCCUGUCCAGCAGCGGGCAAGGCAGGCGGCGAUGAGUCC
GCCGUGGCAAGAACUAACCAGGAUUUAUACAAGGAGGAGAAAUGAAAGCCAUACGGGAAGCAA
UAGCAUGAUACAAAGGCAUUAAGCAGCGUAUCCACAUAGCGUAAAAGGAGCAACAUAGUUAAG
AAUACCAGUCAAUUCUUACAAAUUUGUAAUCCAGAGGUUGAUUAUAGCUUACUUGUACAGC
UCGUCCAUGCCGAGAGUGAUCCCCGGCGGGUCACGAACUCCAGCAGGACCAUGUGAUCCGCGCUU
CUCGUUGGGGUCUUUGCUCAGGGCGGACUGGGUGCUCAGGUAGUGGUUGCGGGCAGCAGCACG

[illegible]

[1184]

GGGCGCGCGCCAGGCGGGCGGGCGGGGCGAGGGGCGGGGCGGGGCGAGGCGGAGAGGUGCGG
CGGCAGCCAAUCAGAGCGGCGCGCUCCGAAAGUUUCCUUUUAUGGCGAGGCGGCGGGCGGGCGG
CCCUAUAAAAAGCGAAGCGCGCGGGCGGGAGUCGCGACGCGUCCUUCGCCCCGUGCCCC
GCUCCGCCGCCGCCUCGCGCCGCCCGCCCCGGCUCUGACUGACCGGUUACUCCCACAGGUGAGCG
GGCGGGACGGCCCUUCUCCUCCGGGCGUGUAAUUAGCUGAGCAAGAGGUAAGGGUUUAAGGGAUG
GUUGGUUGGUGGGGUAUUAAGUUUAAUUACCUGGAGCACCUGCCUGAAAUCACUUUUUUCAG
GUUGGCGUACGgccaccauggugaugccucgagcaccgccggugcagagccugcggagccugcugcggucucacacagagagguccuccc
cugggccacguucgugcggcgucugggccucagggcugcgcagugcagcgcggcgaccccgcccauuccgcgcccugugggccagugccug
gugugugugccuugggauccagaccuccucgcgccuucuuuagacagguucuuugucgaaggagcugugggccgggugcugcagcg
gcuugcgcagcggggagcuaagaauuguuagcuuucgguuuucccugcugcagcggcgccagagggcaccaccuagagccuucacuaaccuggu
gagaagcuaccugccuacaccgugacagacgcccugcggggcagcggcgccugggcgccuacugagagcggugggagauaugugcugugca
ccugcugggcagauccgcccucuuugugcugggugggccuagcugugccuaccagguugugggccuccccuguaaccagcugggcgccgcacgag
ggcagacccccuccgcagcgccgcccugagagacgucugggcugcgaagagcuuggaaccacagugugcgagaagccggcgucccccugggcc
uccugccccuggagccagaagaagagggagcuccgacgagaagccucccgugcccgaagcgccuagacggggcgccgcccugaaaccgagcg
acccggugcgccagggcaguuugggccacccgggaaggaccagaggaccuagcgaccggggcuucugcugugucucccagcccggcagccgagg
aagcaaccagccuggagggcgcccugagcgccacagacacuccacccuuccguugcgagacagcaccacggcccaccaagcaccagcagaccuc
cgcgaccuugggacacaccuguccuccugugacgcccagaccaagcacuuccuguaucuccugcgauaaggaaacagcuccggccuagcuuccu
gcuagagcucucugagaccagcuugaccggcgcccagacucuguggaacaauuccugggcagcagaccuuggaugccagggcaccuccuagacgg
cugccuagacugccccagagauacuggcagauccgucacuguuuuggaacugcugggcaaccacgcucaguguccuacggcgcuucugcugaaaa
cccacugccccugagagccgcugaccccagccggcgugugcgccagagaaaagccuacgggaagcgugggcugcuccugaggaagaggauaca
gaccuagacgcccugugcaacugcugcgccaacauagcagccuugggcagguuacggcuucgucagagcuugugcagaagggcugggccccgg
gacuguggggcuccagacacaacgagcgagauuccugcggaacacaaagaauguacagccugggcaagcacgccaagcugucucgagggauu
gaccuggaagagagcgugagggauugcgugggcugagaagaucccuggagugggcugugugccagcccggagcacagacugcgagaggaaau
ccugggcaaaauuccgacugcugagcguauaugugggaaucgugcgagcucuuucuaugacggagacaacaauucagaagaacaga
cuuuucuuuaccggaagucuguaugagcagcugcaguccaucggaucagacagcaccugaagagauucagcugcggaacugagcgagcgag
aggugagggcagcauagagaggcuagaccugcucugcuuaccagcagacugagauucauccaaaaccugaugggcugagaccuauugauuagg
cuacgugggggcgcccgaaccuuccggcgugagaagcgggcugaagacugacauccgggugaaggcccuuacgugcugcugaacuacgagaga
gcucgucggccugggcuacugggcgccagcgugcugggccuggaugacauccacagagccuggagaacaauucgugcugagagugagggcucaagau
ccccaccggagcuguaucguuaagguggagcugacagcgcauacgacaccauccucagggacagacugaccgaggugauugccagcauuuacaa
ggcccgagaacaccuauugcgugagagcguaaccggugcgagaaagccgcacauggccagugcggaagccuuaagucucacguguccaccug
ccgaccugcagccuacauugcgcaauucgugcccaccuccagggagacaagcccucugagagacggugugaucgagcaguccagcagccuuaau
gaggccuacgcgccguucgacguguuuccugagauuauugccaccacggcgugagaauacagaggcaaaagcuacgugcagugccagggcaucc

[illegible]

[1186]

CGGCAGCCAAUCAGAGCGGCGCGCUCGAAAGUUUCCUUUUUAUGGCGAGGCGGCGGCGGCGGCGG
CCCUAUAAAAAGCGAAGCGCGCGGCGGCGGAGUCGCGUGCGACGCGUGCCUUCGCCCCGUGCCCC
GCUCCGCCGCCGCCUCGCGCCGCCCGCCCCGGCUCUGACUGACCGCGUACUCCCACAGGUGAGCG
GGCGGGACGGCCCUUCUCCUCCGGGCGUGUAAUUAGCUGAGCAAGAGGUAAGGGUUUAAGGGAUG
GUUGGUUGGUGGGGUAUUAAUGUUUAAUUACCUGGAGCACCUGCCUGAAAUCACUUUUUUUCAG
GUUGGCGUACGgccaccauggugaugccucgagcaccggugcagagccgugcgaggccugcugcgguucacacagagaggucccccc
cuggccacguucgugcgcgucugggccucagggcugggcgacugugcagcgcgaggccgcauuccgcccugggccagugccug
gugugugugccuuggggaugccagaccuccuccugcgccuucuuuuagacaggugucugucugaaggagcugugggcccgggugcugcagcg
gcuugcgagcggggagcuagaauuguuagcuuucgguuuugccugcuggacggcgccagaggcgagaccuccuagagccuacacuccuggu
gagaagcuaccugccuacaccgugacagagcccgcgggcgagcgccgucggcgcuacugagagggugggagauaugugcugugca
ccugcugggcagauccgcccucuuugugcugggcccuaugcuguccuaccagugugugcccucccuguaaccagcugggcgccgacgag
ggcagaccccccuccgacgcccagcgcccuaaggagagcugucggcgugcgaaagagcuuggaaccacagugugcgagaaggcgugccccugggcc
ucccugccccuggagccagaagaaggagggcuccgcccagcagaagccucccgugcccagggccuagacggggcgcccccgaaaccgagcg
acccggugggccaggcgaguuugggcccacccggaaaggaccagaggaccuagcgaccggggcuucugcguguguccccagcccggccagccagg
aaggcaaccagccuggaggcgcccugagcgccagacacuccaccuuccguuggcagacagcaccacggcccaccaagcaccagcagaccuc
cgcgaccuugggacacaccuuguccuugugacgcccagaccagcaccuuccguuacuccugcgauaaggaaagcuccggccuagcuuccu
gcuagagcucucugagaccagcuugaccggcgcccagacucuggaaacaauuccugggcagcagaccuuggaugccaggcaccuccuagcgg
cugccuagacugccccagagauacuggcagauccgucacuguuuuuggaucugggcaaccacgcucaguguccuacggcguuucugcugaaaa
cccacugccccugagagcccgugaccccagcccgccgugugcgcgagaaaaagccuaggggaagcgugggcugcuccugagggaaggauaca
gaccuagacgcccugugcaacugcugcgccaacauagcagccuuggcagggugacggcuucgucagagcuugucugagaaggcugggccccgg
gacuguggggcuccagacacaacgagcgagauuccugcggaacacaagaaguucauagccugggcaagcacgccaagcugucucugagggaau
gaccuggaagagagcgugagggaauugcgugcgugcugagaagaucuccugagugggcugugugccagccggcagcacagacugcgagaggaaau
ccuggccaaauuccugcacugcgugaugagcguaauuguguggaacugcugcgagcuucuuuacugacggagacaacuuucagaagaacaga
cuuuuucuuaccggaagucugauaggagcaagcugcaguccaaccggaucagacagcaccugaagagaguuacugcgggaacugagcgaggcag
aggugaggcagcauagagaggcuagaccugcucugcuuaccagcagacugagauuacuucaaaaccugaugccugagaccuauccugaaauagg
cuacgugggggcgcccgaaccuuccggcgugagaagcgggcugaaagacugacauccgggugaaggcccuuacggugcugaacuacgagaga
gcucgucggccugggcuacugggcgccagcgugcgggccuggaugacauccacagagccuggagaacauucgugcugagagugaggcucagau
ccccaccggagcugacuucguuaagguggagcugacaggcgcauacgacacauccucagagacagcugaccggugauugccagcauuucaa
gccccagaacaccuauugcgugagagcguaacggugggcagaaagccgacauugccacgugcggaagccuuaagucacguguccaccug
ccgaccugcagccuuaucgugcgcaauucgugcccaccuccagagacaagccuucugagagacggugggugaucgagcaguccagcagccuuaau
gaggccuacgaggccuugucagcuguuuccugagauuuauugccaccacggcgugagaauacagaggcaaaagcuacgugcagugccagggcaucc
cccaaggaucauuccugagcaccucugcugcagccugugcugacggcgacauggaaaacaagcuucgggaauccggagagagcgccugcugcu

cgggcucgugacgacuccugcugugacaccuaccugacacagccaaaaccuuccucagaacacugguucggggcgucccccgagucaggcgugc
gucuguaaccugagaaagacaguggugaacuucccugugggaagugaggcacucggcggaacagcguucgugcagaugcccgccaggccuguucc
ccugugcggccugcugcuggacaccagaaccuccggagggugcagagcgauuacagcagcuaugccagaaccagcauccgagccagccuaaccuuuaac
cggggcguucaaaagccggcagaaauaugagaagaagcguuucggagugcugagggcuuaaauaggccacagccuguuuccugaccugcaggucaacagcc
ugcaaacagugugcaccacaacaucaagaauccugcuguuacaagccuacagauuucacgcccugugugcugcagcugccuuuccaccagcagggugug
gaagaaccuacguuuuuccugagagugaucucugauacagccaguuugugcuacagcauccugaaggccaagaacgccggcgaugagccugggagcc
aaggggcgugcugggccucugccuucugaggccgugcagugggcugugccaccaggccuuucugcugaagcugaccagacauagagugaccuacgug
ccucuccugggcagccugagaaccgccagaccagcucucucgggaagcugccugggcacaaccucgaccgcccuggaagcugcugccaauccugcccu
gccuagcgacuucaagaccauccuggacugauAAGCUUGAUCCAGACAUGAUAAAGAUACAUUGAUGAGUUUGGAC
AAACCACAACUAGAAUGCAGUGAAAAAAUAGCUUUAUUUGUGAAAAUUUGUGAUGCUAUUUGCUUU
AUUUGUAACCAUUAUAAGCUGCAAUAAACAAGUUGUGGAAUUGUGAGCGCUCACAAUUCCACAG
CGGCCGCggggcagagcgacaucccccagucccgagaaguuuggggggaggggucggcaauugaaccggugccuagagaaggugggcgggg
guaaaacugggaagugaugugcugugacugggcuuccgcuuuuucccgaggguggggggagaaccguauuaaagugcaguagucggcugaacguucuu
uuucgcaacggguuuugccgccagaacacagcgcuacggccaccaugguguccaaaggcgaggaagugaucaagggaauuuagagauucaaggucaga
auggaagggaucuaugaacggccacgaguuucgagaucgaggggagaaggagaaggcgagaccuacgagggcacccaaaccgcuaaaauaaaagugacaaa
ggggcggaaccucugccguuucgcuugggauauccuguccccccaguucauguacggcgagcaaggccuacgugaagcaccucugcagacucccgauuac
aaaaagcugucuuuuccugagggcuuuuaagugggagagagugaugaaacuucgaggacggaggccugugaccguuacccaagacagcagccugcagaag
auggaacacugaucuacaaagugaagaugcgggggcacuaauuuccuccagacggccccgugaugcagaagaaaaccaugggcugggagggccucuaac
agaacggcguguaaccacagagacggagugcugaaggggcgagauuaccaggcgacugaagcugaagaagugggcgagacuaaccugguuggaguucaagacc
auuuacauaggccaagaaagccugugcagcugccuggcuaacuacuguggacaccaagcuggacauaccuccacaaugagggaauuauaccaucugggga
acaguacgagagaagcgaggggcagacaccaccuguuuccucggccacggcgacaggcuaccaccggcucugggcagcagcgggcacagccagcagcgaggaca
acaauaggccgugaucaagggaguucaugcgguucaaggugcggauggaaggcaguuagaacggccaugaguuucgaaucgagggcggaaggcggaag
gcagaccuuacgagggcacacagacagcuuaagcugaaggugaccaaggggagggccucugccauucgcccugggacauccugagcccucaguuuaugua
ugggcagcaaggccuacgugaagcaccgcccgacaucccgacuaaagaagcugagcuucccugaggguuucaaauggggaaggggugaugaaacuuc
gaggacggcgggacuuggugacagugacccaggacagcagccugcaggauggcaccucgaucauaaaggugaagaugagaggcaccaacuuccaccug
acggggcccugaugcagaaaaaaaccaugggaugggagggccagcaccgagcgggcuguaaccuagagauggcgguucugaaggggcgagauccaccaggcg
ccugaagcugaaggacggcgggccacuaccugggugaguucaagaccauacauaggccaagaaccugugcagcugccuggaucuaacuacugggac
acaaagcuggacauaccucucacaacgaggacuaauaccaucgucgagcaguuacgaacgcagcggaaggcgagacaccaucuguuuccuguauggga
cgagcuguaaaguaauggauuaauuaaaauaaaagauuuuauuguaauagauucuguguguuuguuuuuguguguggaauugugagcgcuca
caauuccacagcgggccgcUGAGGUAGAUAAUCUUUGUAUAGUGGGGGGGGAUCUCAUGUACCGGGUUUCU
UUUAUUUGAUUUUCAUAAAAACAGACGGUAGCUAGGUUCGCAAGGCAGCCACAAGCCAAAGUA
GGUAGGGUGCUCAUAGUGAGUAGGGACAGUGCCUUUUGAUUCACAACGCGUCAAUACCAUCUGA

CACGGAUACCCUUACCGGACUUGUCAUGAUCUCCAGACUUGUCCAAGGUGGACGGGCCACCUUU
ACUUAACCCGGAAAAGGAACAUUAUUAUUUAUAUGUGUUCGGAAAAUAGCAAAAAAAAAAAAA
AAAAAAAAA

520. TCA5_CBhBsi_PAH_GeFo3 RNA 序列

[1188] gggagacggucaaccgcguaggagcgugaccgGAAUUCGGCGGGAGUAACUAUGACUCUCUUAAGGAGUCAUAG
AGCCAGAACCUCUGUGGUCCCGCUGGGCACAGGGAUUAAUUUUUCUGUGGCAAAUUUGACUG
GCUUCAGAGAGCGUUUUUCGAAGUGGACUGUGUGACUGCGUUCUUUUUUAGUUGCUUAUACCG
CUUCGAUUAACAUCUCACCUCGACGUUAAGAUAUUUCUAGAAGGUCGACCAGAUGUCCGAGG
UCGACCAGUUGUCCGUGUGGAAUUGUGAGCGCUCACAAUCCACACGUUACAUACCUUACGGUA
AAUGGCCCCGCCUGGCUGACCGCCCAACGACCCCCGCCCAUUGACGUCAAUAGUAACGCCAAUAGG
GACUUUCCAUUGACGUCAAUGGGUGGAGUAUUUACGGUAAACUGGCCACUUGGCAGUACAUCAA
GUGUAUCAUAUGCCAAGUACGCCCCCUAUUGACGUCAAUGACGGUAAAUGGCCCCGCCUGGCAUU
GUGCCCAGUACAUGACCUUAUGGGACUUUCCUACUUGGCAGUACAUCUACGUAAUAGUCAUCGC
UAUUACCAUGGUCGAGGUGAGCCCCACGUUCUGCUUCACUCUCCCCAUCUCCCCCCCCUCCCCACC
CCCAAUUUUUGUAUUUAUUUAUUUUUUAAUUAUUUUUGUGCAGCGAUGGGGGCGGGGGGGGGGGGG
GGGCGCGCGCCAGGCGGGGCGGGGCGGGGCGAGGGGCGGGGCGGGGCGAGGCGGAGAGGUGCGG
CGGCAGCCAAUCAGAGCGGCGCGCUCCGAAAGUUUCCUUUUUAUGGCGAGGCGGCGGGCGGCGG
CCCUAUA AAAAAGCGAAGCGCGCGGGCGGGGAGUCGCGACGCGUGCCUUCGCCCCGUGCCCC
GCUCCGCCGCCGCCUCGCGCCGCCCGCCCCGGCUCUGACUGACCGCGUUACUCCACAGGUGAGCG
GGCGGGACGGCCCUUCUCCUCCGGGCGUAAUUAGCUGAGCAAGAGGUAAGGGUUUAAGGGAUG
GUUGGUUGGUGGGGUAAUAAUGUUUAAUUACCUGGAGCACCUGCCUGAAAUCAUUUUUUUCAG
GUUGGCGUACGgccaccAUGGUGAUGAGCACCGCCGUGCUGGAGAAUCCUGGCCUGGGCAGAAAGC
UGUCCGAUUUCGGCCAAGAGACAAGCUACAUAUGAGGAUAAUUGCAACCAGAACGGCGCUAUCAG
CCUGAUCUUUUUCUUUGAAAGAGGAAGUGGGCGCUCUGGCCAAGGUCCUGAGGCUGUUCGAGGAA
AACGACGUGAACCUGACCCAUUUCGAAAGCAGACCCAGCCGGCUGAAGAAAGAUGAGUAUGAGU
UCUUCACCCACCUGGACAAGAGAUUCUCUCCUGCCUGACAAACAUCAUAAAAUCCUAAGACAC
GACAUUGGCGCCACCGUGCACGAGCUGUCAAGAGAUAGAAGAAGGACACCGUGCCUUGGUUUC
CUAGAACCAUCCAGGAGCUGGAUAGAUUCGCCAAUCAGAUCCUGUCUACGGCGCCGAACUGGAC
GCCGACCACCCCGGCUUCAAGGACCCUGUGUACCGGGCCAGACGCAAGCAGUUCGCCGACAUCGC
CUAUAACUACCGGCAUGGCCAGCCUAUCCCUAGAGUGGAAUACAUGGAAGAAGAGAAAAAAACC
UGGGGCACCGUUUCAAGACCCUGAAAAGCCUGUAUAAGACUCACGCCUGCUACGAGUACAACCA

[1189]

CAUCUUCCCUCUGCUGGAAAAGUACUGUGGCUUUCACGAGGACAACAUCCCACAGCUGGAGGACG
UGUCCCAGUUCCUGCAGACCUGUACCGGCUUCCGGCUGCGGCCCCGUGGCCGGACUGCUGAGCAGC
CGGGACUUCCUGGGCGGACUGGCUUUUAGAGUGUCCACUGCACACAGUACAUCAGACACGGCUC
CAAGCCUAUGUACACACCUGAGCCCCGACAUCUGCCACGAGCUGCUGGGACACGUCCCCUCUGUUUA
GCGAUAGAUCUUUCGCCCAGUUUAGCCAGGAGAUCGGCCUGGCCAGCUUGGGCGCCCCUGAUGAA
UACAUCGAGAAACUGGCUACAAUCUACUGGUUCACAGUGGAGUUCGGCCUCUGCAAGCAGGGCG
ACAGCAUCAAGGCCUACGGCGCCGGACUGCUUUCUAGCUUCGGCGAGCUGCAAUACUGUCUGUCC
GAGAAGCCCAAGCUGCUGCCUCUGGAACUGGAAAAGACCGCCAUCCAGAAUUACACCGUGACCGA
GUUCCAGCCCCUGUACUACGUGGCUGAGAGUUUCAACGACGCUAAGGAAAAGGUGCGGAACUUC
GCCGCCACAAUCCCAAGACCUUUCAGCGUGCGGUACGACCCCUACACCCAAAGAAUCGAGGUGCU
GGACAACACCCAGCAGCUGAAGAUCCUGGCAGAUAGCAUCAACAGCGAGAUCGGAAUCCUGUGC
AGCGCCCUGCAGAAGAUUAAGUGAUAAAGCUUGAUCCAGACAUGAUAAAGAUACAUUGAUGAGUUU
GGACAAACCACAACUAGAAUGCAGUGAAAAAAAUGCUUUAAUUGUGAAAAUUGUGAUGCUAUUG
CUUUAAUUGUAACCAUUUAAGCUGCAAUAAACAAGUUGUGGAAUUGUGAGCGCUCACAAUCC
ACAGCGGCCCGUGAGGUAGAUAAUCUUUGUAUAGUGGGGGGGGAUCUCAUGUACCGGGUUUCUU
UUAAUUGAUUUUCAAUAAAACAGACGGUAGCUAGGUUCGCAAGGCAGCCACAAGCCAAAGAUAG
GUAGGGUGCUCAUAGUGAGUAGGGACAGUGCCUUUUGAUUCACAACGCGUCAAUACCAUCUGAC
ACGGAUACCCUUAACCGGACUUGUCAUGAUCUCCAGACUUGUCCAAGGUGGACGGGCCACCUUUA
CUUAACCCGGAAGGAACAUAUAUAAUUAUUGUGUUCGGAAAAUAGCAAAAAAAAAAAAAA
AAAAAAA

521. TCA5_CBhBsi_PAH_EFS_TdT_GeFo3 RNA 序列

gggagacggucaaccgcuaggagcggugaccgGAAUUCGGCGGGAGUAACUAUGACUCUCUUAAGGAGUCAUAG
AGCCAGAACCUCUCUGGUUCCCGCUGGGCACAGGGAUUAUUUUUCUGUGGCAAAUUGACUG
GCUUCAGAGAGCGUUUUUCGAAGUGGACUGUGUGACUGCGUUCUUUUUUAGUUGCUAUAUCCG
CUUCGAUUAACAUCUCACCUCGACGUUAAGAUAUUUCUAGAAGGUCGACCAGAUGUCCGAGG
UCGACCAGUUGUCCGUGUGGAAUUGUGAGCGCUCACAAUCCACACGUUACAUAACUUAACGGUA
AAUGGCCCCGCCUGGCUGACCGCCCAACGACCCCCGCCAUUGACGUCAAUAGUAAACGCCAAUAGG
GACUUUCCAUUGACGUCAAUGGGUGGAGUAUUUACGGUAAACUGCCCACUUGGCAGUACAUCAA
GUGUAUCAUAUGCCAAGUACGCCCCCUAUUGACGUCAAUGACGGUAAUUGGCCCGCCUUGCAUU
GUGCCCAGUACAUGACCUUAUGGGACUUCCUACUUGGCAGUACAUCUACGUAAUAGUCAUCGC
UAUUACCAUGGUCGAGGUGAGCCCCACGUUCUGCUUCACUCUCCCCAUCUCCCCCCCCUCCCCACC

[1190]

CCCAAUUUUUGUAUUUAUUUAUUUUUUAAUUUUUUUUGUGCAGCGAUGGGGGCGGGGGGGGGGGGG
GGGCGCGCGCCAGGCGGGGCGGGGCGGGGCGAGGGGCGGGGCGGGGCGAGGCGGAGAGGUGCGG
CGGCAGCCAAUCAGAGCGGCGCGCUCCGAAAGUUUCCUUUUUUGGCGAGGCGGCGGGCGGGCGG
CCCUAUAAAAAGCGAAGCGCGCGGGGCGGGGAGUCGUCGACGCGUGCCUUCGCCCCGUGCCCC
GCUCCGCCGCCGCCUCGCGCCGCCCGCCCCGGCUCUGACUGACCGCGUUACUCCACAGGUGAGCG
GGCGGGACGGCCCUUCUCCUCCGGGCGUGAAUUAGCUGAGCAAGAGGUAAGGGUUUAAGGGAUG
GUUGGUUGGUGGGGUAAUAAUGUUUAAUUACCUGGAGCACCUGCCUGAAAUCACUUUUUUUUCAG
GUUGGCGUACGgccaccAUGGUGAUGAGCACCGCCGUGCUGGAGAAUCCUGGCCUGGGCAGAAAGC
UGUCCGAUUUCGGCCAAGAGACAAGCUACAUUGAGGAUAAUUGCAACCAGAACGGCGCUAUCAG
CCUGAUCUUUUUCUUUGAAAGAGGAAGUGGGGCGCUCUGGCCAAGGUCCUGAGGCUGUUCGAGGAA
AACGACGUGAACCUGACCCAUAUCGAAAGCAGACCCAGCCGGCUGAAGAAAGAUGAGUAUGAGU
UCUUCACCCACCUGGACAAGAGAUUCUCCCCUGCCUGACAAACAUAUAAAAUCCUAAGACAC
GACAUUGGCGCCACCGUGCACGAGCUGUCAAGAGAUAAAGAAGAAGGACACCGUGCCUUGGUUUC
CUAGAACCAUCCAGGAGCUGGAUAGAUUCGCCAAUCAGAUCCUGUCUACGGCGCCGAACUGGAC
GCCGACCACCCCGGCUUCAAGGACCCUGUGUACCGGGCCAGACGCAAGCAGUUCGCCGACAUCGC
CUAUAACUACCGGCAUGGCCAGCCUAUCCCUAGAGUGGAAUACAUGGAAGAAGAGAAAAAAACC
UGGGGCACCGUUUCAAGACCCUGAAAAGCCUGUAUAAGACUCACGCCUGCUACGAGUACAACCA
CAUCUUCUCCUCUGCUGGAAAAGUACUGUGGCUUUCACGAGGACAACAUCCACAGCUGGAGGACG
UGUCCCAGUUCUGCAGACCUGUACCGGCUUCCGGCUGCGGCCCGUGGCCGGACUGCUGAGCAGC
CGGGACUUCUGGGCGGACUGGCUUUUAGAGUGUCCACUGCACACAGUACAUCAGACACGGCUC
CAAGCCUAUGUACACACCUGAGCCCACAUCUGCCACGAGCUGCUGGGACACGUCCCUCUGUUUA
GCGAUAGAUCUUUCGCCAGUUUAGCCAGGAGAUCCGGCCUGGCCAGCUUGGGCGCCCCUGAUGAA
UACAUCGAGAAACUGGCUACAAUCUACUGGUUCACAGUGGAGUUCGGCCUCUGCAAGCAGGGCG
ACAGCAUCAAGGCCUACGGCGCCGGACUGCUUUCUAGCUUCGGCGAGCUGCAAUACUGUCUGUCC
GAGAAGCCCAAGCUGCUGCCUCUGGAACUGGAAAAGACCGCCAUCCAGAAUUACACCGUGACCGA
GUUCCAGCCCCUGUACUACGUGGCUGAGAGUUUCAACGACGCUAAGGAAAAGGUGCGGAACUUC
GCCGCCACAAUCCCAAGACCUUUCAGCGUGCGGUACGACCCUACACCCAAAGAAUCGAGGUGCU
GGACAACACCCAGCAGCUGAAGAUCUGGCAGAUAGCAUCAACAGCGAGAUCGGAAUCCUGUGC
AGCGCCUGCAGAAGAUUAAGUGAUAAAGCUUGAUCCAGACAUGAUAAAGAUACAUUGAUGAGUUU
GGACAAACCACAACUAGAAUGCAGUGAAAAAAUAGCUUUUAUUUGUGAAAUUUGUGAUGCUAUUG
CUUUAAUUUGUAACCAUUAUAAGCUGCAAUAAACAAGUUGUGGAAUUGUGAGCGCUCACAAUCC
ACAGCGGCCGCgggcagagcgcaucgcccacagucccgagaaguugggggaggggucggcaauugaaccggugccuagagaaggugg

[1191]

cgcggggguaaacuggggaaagugaugucgugucacuggcuccuuuuuccgaggguggggggagaaccguauauaagugcaguagucgggugaac
guucuuuuucgcaacggguuugccgccagaacacagcgguacggccaccaugguuccaaaggcgagggaagugaucaaggaaauuauagagauucaag
gucagaauagggaaggaucuaugaacggccacgaguuucgagaucgaggggagaaggagaaggcagaccuacgagggcaccacaaaccguaaaaaaaagu
gacaaaggggcgaccucugccguucgucggggaauauccuguccccaguucaugacggcagcaaggccuacgugaagcaccucgugacaucccc
gaauacaaaaagcugucuuuuuccugagggcuuaagugggagagagugaugaacuucgaggacggaggccugugacccguuacccaagacagcagcc
ugcaagauggaacacugaucuacaagugaagugcggggcacuaauuuuccuccagacggcccgugaugcagaagaaaaccaugggcugggaggc
cucuacagaaggcguguaacccagagacggagugcugaaggggcagauuaccaggcacugaagcugaaggaugggcgacacuaccuguggaguuu
aagaccuuuacaugcccaagagccugugcagcugccugcuacuacuacgugacaccaagcuggacauaccuccacaauaggagauuauaccau
cguggaacaguacgagagaaggcgaggcgagacaccaccuuguccucggccacggcacaggcuccaccggcucuggcagcagcggcacagccagcagcg
aggacaacaacauaggccgugaucaaggaguucaugcgggucaaggugcggauuggaaggcaguaugaacggccaugaguuagaaucaggggcgagg
cgaaggcagaccuuacgagggcacacagacagcuagcugaaggugaccaaggaggagccucugccauucgucgggacauccugagccucaguuu
auguauggcagcaaggccuacgugaagcaccgccgacaucccgacuacaagaagcugagcuuccugaggguuucaauagggaaggguugauga
acuucgaggcagcgcgacugugacagugacccagggacagcagccugcaggauggcacccugaucuauaaggugaaggaugagaggcaccacuuucc
accugacggggccgugaugcagaaaaaaccaugggauaggaggccagcaccgagcgcguguaaccuagagauggcguucugaaggggcagauccac
caggcccguaagcugaaggacggcgccacuaccuguggaguucaagaccuacaucaugcccaagaaaccugugcagcugccuggaucuaucacg
uggacacaaagcuggacauaccucacacagagagacuuaccuacgucgagcaguuacgaacgagcgaaggcagacaccuacguuuccuguaagg
auaggacgagcuguacaagaaugauuaauaaaaaaaagauuuuauuagauuagauucuguguuuguuuuuguguguggaauugugagc
gcucacauuuaccacggcgccgcUGAGGUAGAUAAUCUUUGUAUAGUGGGGGGGGAUCUCAUGUACCGGGUU
UCUUUUUUUUGAUUUUUCAAUAAAACAGACGGUAGCUAGGUUCGCAAGGCAGCCACAAGCCAAAG
AUAGGUAGGGUGCUCAUAGUGAGUAGGGACAGUGCCUUUUGAUUCACAACGCGUCAAUACCAUC
UGACACGGAUACCCUUACCGGACUUGUCAUGAUCUCCCAGACUUGUCCAAGGUGGACGGGCCACC
UUUACUUAACCCGGAAGGAACAUUAUUAUUUAUUAUGUGUUCGGAAAAUAGCAAAAAAAAAA
AAAAAAAAAAAA

522. TCA5_CBhBsi_F8N6_GeFo3 RNA 序列

gggagacggucaaccgcuaggagcgguaccgGAAUUCGGCGGGAGUAACUAUGACUCUCUUAAGGAGUCAUAG
AGCCAGAACCUCUCGUGGUCCCGCUGGGCACAGGGAUUAUUUUUCUGUGGCAAAUUUGACUG
GCUUCAGAGAGCGUUUUUCGAAGUGGACUGUGUGACUGCGUCCCCCUUAGUUGCUAUAUCCG
CUUCGAUUAACAUCACACUCGACGUUAAGAUAUUUCUAGAAGGUCGACCAGAUGUCCGAGG
UCGACCAGUUGUCCGUGUGGAAUUGUGAGCGCUCACAAUCCACACGUUACAUAACUUAACGGUA
AAUGGCCCCGCCUGGCUGACCGCCCAACGACCCCCGCCAUUGACGUCAAUAGUAACGCCAAUAGG
GACUUUCCAUUGACGUCAAUGGGUGGAGUAUUUACGGUAAACUGCCCACUUGGCAGUACAUCAA

[1192]

GUGUAUCAUAUGCCAAGUACGCCCCUAUUGACGUCAAUGACGGUAAAUGGCCCGCCUGGCAU
GUGCCCAGUACAUGACCUUAUGGGACUUUCCUACUUGGCAGUACAUCUACGUAAUAGUCAUCGC
UAUUACCAUGGUCGAGGUGAGCCCCACGUUCUGCUUCACUCUCCCCAUCUCCCCCCCCUCCCCACC
CCCAAUUUUGUAUUUAUUUAUUUUUAAUUAUUUUGUGCAGCGAUGGGGGCGGGGGGGGGGGG
GGGCGCGCGCCAGGCGGGCGGGGCGGGGCGAGGGGCGGGGCGGGGCGAGGCGGAGAGGUGCGG
CGGCAGCCAAUCAGAGCGGCGCGCUCCGAAAGUUUCCUUUAUGGCGAGGCGGCGGCGGCGGCGG
CCCUAUA AAAAAGCGAAGCGCGCGGGCGGGGAGUCGCGUGCGACGCGGCCUUCGCCCCGUGCCCC
GCUCCGCCGCCGCCUCGCGCCGCCCGCCCCGGCUCUGACUGACCGCGUUACUCCCACAGGUGAGCG
GGCGGGACGGCCCUUCUCCUCCGGGCUGUAAUAGCUGAGCAAGAGGUAAGGGUUUAAGGGAUG
GUUGGUUGGUGGGGUAAUAAUGUUUAAUUACCUGGAGCACCUGCCUGAAAUCACUUUUUUCAG
GUUGGCGUACGgccaccaugcaaaucgagcuguccacuugcuuuccuguccugcugcgcuuuuguuucagugcuaccgcagauacua
ccugggagcgcuggagcguuucugggacuauugcagagcgauucuggcgagcugccagucgacgcagauuuccaccuagagugccuagaguuu
cccuuacaaccucugugguuacaagaaaaccuuguuuguggaguuaccgcaccaccuugcaauaucguuagccucgcccuccuuggaugggc
cugcugggccccaccauucaggcugaagguuacgacaccgugugauaccuccugaagaacaugccuccaccucugucucugcacgcccugcg
ugucguacuggaaggccagcgaggcgcguaauacgacgaccagaccagccaaagagagaaggaggaugauaaggguuuccaggugcgagccacac
cuacugugcgagcugcugaagagaacgcccuaugccucagaccucugugccugacauacgcuauuuauccacguugaccugugaaggac
cugaacagcgagcugaucggcgcccugcugugugcagagaaggcuuccugccaaaggaaaagacacagacacuacacaaguucacuuacuguuug
ccguguucgacgaggcgcaagagcuggcacagcgagaccaagaauagccugaugcaagacagagaugccgcuuccgcuagagcguggccuaaaugcac
accgugaacggcuacgucaacagauucugcccggccugauuggcugucaccgggaugccguguaucggcgacugaucggaugggcacacccccg
agguacacuaauuuccuggaaggccacaguuuccugugagaaccaccggcaggccucgucgagaucccccuaucacauuucugaccgcccag
accuugcuauggaccugggccaguuucugcuuuucugccacauacgucuccaccagcacgacggcauggaggguuacuguaagguggaauuucug
cugaggagccucagcugagaauagaagaauaagaggaaagccgaagauuacgacgacgaccugacggacagcgagauugauguugagga
ugauauagcccuuccuuuuccagauccggucagugcggaagaacaccuaagacaugggucacuaucgcccugaggaggaggaauugggac
uacgccccccugugcugccccugacgaccggagcuacaaucucaguaccugaacaacggccucaaagaauccgagaaaguuacaagaaggugag
auucauggcuuuacagauaacaauuaaaaccagagaagcuuccagcaugagagcggaauccugggaccucugcuguacggcgaggugggcgac
accucgugaucaucuuuaagaaccaggccagaccuuacaauaucuaccucacggcauuacagacgugagaccucuguacagccggcgcgcc
caaggcgugaagcaccugaaggacuucccauuccugccugagagauuucaaguuacaaguggaccugacaguggaagugcccuaccaagucu
gacccccggugccugacccgguaucucgaguuuuguaacauggagcgcgaccugccucuggacuaucggccauugcugaucugcuacaaag
aguccguggaccagagaggcaaccagaucaugagcgacaagcguaacgugauccuguuacgcuuuugaugagaauagaagcugguauucgaccga
gaacauccagcgguuccccccaccucggcgugcagcugggagaccugaguuccaggccagcaacaucagcacucuaauaacgcuacgugu
ucgacuccugcagcuagcguuugccugcacgaggugcuuacugguacaucugagcaucggcgcccagaccgacuuccuguccuguuucuc
aggcuacacaucaaaacauaagaugguuagaggacaccugaccuugccuuuccggagagaccguguuacagcaugagagaaccuggc

[1193]

cuuuggauccugggugucacaaagcgacuuucggaaaccggggaugaccgcccugcugaaagucagcucuuugcgacaagaacaccggcgacuacu
acggaggacagcuauaggagacauccgcuaccugcugagcaagaacaacgccaucgagccucggucguucagccagaacucuaaggcaccacccagcacc
cggcagaagcaaucaacgcaaccaccauuccugaaaaacgacaucaaaagaccgauccuugguucgucacagaaccccaaugcccaagauccaaaac
guguccucuuugaccugcugaugcugcugagacagagcccaacccucacggccugagccugucugaccugcagggagggccaaauacgagacuucuu
cugacgauccuagccccggggccaucgacagcaacaacagccugagcgagauacauuuccggccucagcuccaccacucugggagacaugguuuuc
acaccugaaagcggccugcagcugagacuagcagagagcuaggcaccaccgucgcccaccgaacugaaagcugcagcuucaagguaagcagcaccag
caacaaccugaucagcaccuuccuucugacaaccuugccggcgaccgacaacacagcagccugggcccccuccuuaugccugugcacuacgacag
ccagcugggauaccacacuguuuggaaagaaucuaagcccucugaccgaauucugcgccgcccucugagccugagcgaagagaacaacgaugcaaacugc
uggagagcggccugaugaacagccagagucacuuuggggcaagaacgugagcaguaccaggagauacccggacaaccugcagagcgaucaggga
ggagauccgaaugaugacacuaucagcggugagauaagaagaggacuuacacuaucagcagagcagaaaaccaaagcccucggagcuuccaga
agaaaaccagacacuaauuacugccgcuugggagagacugugggacuacggcgaugagcuccucuccacgugcugcggaaccgggcccagagcgg
aucuguuucccaguucaagaagucguguucaagaguuaccggcagcucuuuuaccagccucuguaugagggcgaauuaaacgaacaccugggc
cugcugggcccuuacacagagcagaagugaggauaacaucaugggaccuuacagaauacaggccucucggcccuacucuuuacuccucgcuca
ucagcuacgagggagcaccagagcgaaggcggccgaaccagaaagaacuuugugaagcccaacgaaccaaagacuuacuuuuggaagguagaacaccac
augggcccuuacaaagacgaguuugauugcaaggccuugggccuacuuacagcgaugugagccuggaaaaggauugccauagcggccuauugggcccu
ugcugggugccacacaaacacacuaauccugcccacgggagacaggugaccgugcagggauucgcccuguuuacaccuucgacgagacaaag
agcugguacuucacagaaaacauaggaaagaaauugcagagcccuugcaacauucagaugagggauccaccuucaaggaaaauacagauuccacgc
caucaacggauaacaucagacacacugccagcagcuguuuugggccaggaucagagaauccggugguaccugcucuaugggcucaaacgagaa
auccacuccauccacuuuagcggccacguguuacccgugagaaagaagaggaguuacaagaugggccuacaaucuguaaccccgccgugucgaga
cagugggaaaugcugccaagcaaggccgggacucggcgggggaugucuuuugggcagcaccugcagccgggaugagcacacuguuucgguau
acagcaacaagugccagacacccuugggcagcugcucacacagagaauuucagauacagccucugugcaguuacgacagugggcuuccaa
gucggccagacugcacuacuccggcuuccaucaacgccugagcacaagaagccuuucagcuggaucagggucgacucugggcuuccaagaucaucc
acggaaucagaccagggcgcccgagaaaucucgagccuacacucgaguucauuuacaguuacagccuggacggaaaagagugcgagacc
uacagaggaaauaguaccggcaccugaugguguuucggaugggacagcucaggaauaagcacaacauucaauccuccaaucauagccag
auacacagacuucacuuaccacuaucuaucagaagcagcugagaauggaguggaugggauugaccugaacagcugcagcaugccucuuaggc
auggaaucaaaagccauucugaugccagauaccgcuucuauguuacuuacacaaugucgccacaugguccuagcaaggccagacugcaccu
gcagggcaggucaauugccugggccucagggugaacaaccccaaggaaugggcugcagguccauuuccagaaaaccaaagaaagugaccgagugacc
accagggcgugaaaagccugcugaccagcauguaagcugaagagauuccugaucuacgucuccaggaugggccaccaguggaccuucuuucaga
acggcaagugaaaguguucaaggcaaccagagacuuacacccagucgugaacagccuggauccaccucuccuuaaccagauaccugagaauccau
ccccaguccggguccaccagauccgcccgggaugggagugcugggcugcagggcccaagaccugacuAAGCUUGAUCCAGACAUG
AUAAGAUACAUGAUGAGUUUGGACAAACCACAACUAGAAUGCAGUGAAAAAAUGCUUUAUUU
GUGAAAAUUUGUGAUGCUAUUGCUUUAUUUGUAACCAUUAUAAGCUGCAAUAAACAAGUUGUGGA

AUUGUGAGCGCUCACAAUCCACAGCGGCCGUCGAGGUAGAUAAUCUUUGUAUAGUGGGGGGGG
AUCUCAUGUACCGGGUUUCUUUUUUUUUUAUUUGAUUUUCAAUAAAACAGACGGUAGCUAGGUUCGCAAG
GCAGCCACAAGCCAAAGAUAGGUAGGGUGCUCAUAGUGAGUAGGGACAGUGCCUUUUGAUUCAC
AACCGGUCAAUACCAUCUGACACGGAUACCCUUACCGGACUUGUCAUGAUCUCCAGACUUGUCC
AAGGUGGACGGGCCACCUUUACUUAACCCGAAAAGGAACAUAUAUUAUUUAUAGUGUUCGGA
AAAUAGCAAAAAAAAAAAAAAAAAAAAAA

523. TCA5_CBhBsi_F8N6_EFS_TdT_GeFo3 RNA 序列

[1194] gggagacggucaaccgcuaggagcggugaccgGAAUUCGGCGGGAGUAAUAUGACUCUCUUAAGGAGUCAUAG
AGCCAGAACCUCUGUGGUCCCGCUGGGCACAGGGAUUAUUUUUCUGUGGCAAAUUUGACUG
GCUUCAGAGAGCGUUUUUCGAAGUGGACUGUGUGACUGCGUUCUUUUUUAGUUGCUAUAUCCG
CUUCGAUUAACAUCUCACCUCGACGUUAAGAUCAUUUUCUAGAAGGUCGACCAGAUGUCCGAGG
UCGACCAGUUGUCCGUGUGGAAUUGUGAGCGCUCACAAUCCACACGUUACAUAACUUAACGGUA
AAUGGCCCCGCCUGGCUGACCGCCCAACGACCCCCGCCAUUGACGUCAAUAGUAACGCCAAUAGG
GACUUUCCAUUGACGUCAAUGGGUGGAGUAUUUACGGUAAACUGCCCACUUGGCAGUACAUCAA
GUGUAUCAUAUGCCAAGUACGCCCCCUAUUGACGUCAAUGACGGUAAAUGGCCCCGCCUGGCAUU
GUGCCCAGUACAUGACCUUAUGGGACUUCCUACUUGGCAGUACAUCUACGUAAUAGUCAUCGC
UAUUACCAUGGUCGAGGUGAGCCCCACGUUCUGCUUCACUCUCCCCAUUCCCCCCCCUCCCCACC
CCCAAUUUUUGUAUUUAUUUAUUUUUUAAUUAUUUUUGUGCAGCGAUGGGGGCGGGGGGGGGGGG
GGGCGCGCGCCAGGCGGGGCGGGGCGGGGCGAGGGGCGGGGCGGGGCGAGGCGGAGAGGUGCGG
CGGCAGCCAAUCAGAGCGGCGCGCUCCGAAAGUUUCCUUUUUUGGCGAGGCGGCGGGCGGCGG
CCCUAUA AAAAGCGAAGCGCGCGGGCGGGGAGUCGCGUGCGACGCUGCCUUCGCCCCGUGCCCC
GCUCCGCCGCCGCCUCGCGCCGCCCGCCCCGGCUCUGACUGACCGCGUUACUCCACAGGUGAGCG
GGCGGGACGGCCCUUCUCCUCCGGGCGUGAAUUAGCUGAGCAAGAGGUAAGGGUUUAAGGGAUG
GUUGGUUGGUGGGGUAAUAAUGUUUAAUUACCUGGAGCACCUGCCUGAAAUCACUUUUUUUCAG
GUUGGCGUACGgccaccaugcaaaucgagcuguccacuugcuuuccugugccugcugcgcuuuuguuucagucuaaccgcagauacua
ccuggggagccgugggagcuuucuuugggacuauaugcagagcgauucggcgagcugccagucgacgccaauuccaccuagaguccuaagaguuu
cccuuuaacaccucugugguguacaagaaaaccuuguuugggaguuaccgaccaccuuguucaaauaucgcuaggccugggccuccuuggaugggc
cugcugggggcccauuacaggcugaaguguacgacaccgugguagacaccugaagaacauggccuccaccucugucucugcacgcccugcgcg
ugucguacuggaaggccagcgaggcgcgugaaucgacgaccagaccagccaagagagaaggaggauaagguuuccagguaggcagccacac
cuacguguggcagugcugaaagagaacggcccaaugccucagaccucugugccugacaacagcuauuuaucccaoguugaccugugaaggac
cugaacagcgagcugaucggcgccugcugugugcagagaaggcuuccugccaaggaaaagacacagacacuacacaagucaucuacuguuug

[1195]

ccgguuucgacgagggccaagagcuggcacagcgagaccaagaauagccugaugcaagacagagauccgcuagagcgugggccuaaaaugcac
accgugaacggcuacgucacagauucugcccggccugauuggcugucaccgggaaguccguguacuggcacugaucggaaugggcacaacccggg
agguccacucaauucuggaagccacacguuccugugagaaccaccggcagggccucgucgagaucccccuaucacauuucugaccggccag
accuugcucauggaccugggccaguuucugcuuuucugccacaucagcuccaccagcagcagggcaugggagggcuuacgugaagguaguuucuguc
cugagggagccucagcugagaaugaagaauaauaggaagccgaagauuacgacgacgaccugacggagacgagaguggaugauugugagguuugacga
ugauaauagcccuuccuuuuccagauccggucaguggcgaagaacaccuuaagacaugggugcacuacaucggcgugagggagggauugggac
uacgccccccugugcugggcccgacgacggagcuacaaucucaguaccugaaacagggcccucaaagaauccggcagaagaucagaagguag
auucaugggcuuauacagauaacaauuuaaaaccagagaagcuauccagcaugagagcggaauccugggaccucugcuguaacggcgagguaggcgac
accucgucgaucaucuuuaagaaccagggcagcagaccuacaaauaucacccucacggcauuacagacgugagaccucuguaacagccggcgugcc
caaggggcgugaagcaccugaaagacuuccaauccugccugagagauucuaaguaaaguggaccgugacaguggaagaugggccuacaaagucu
gacccccggugccugacccgguaucucgaguuuugugaacauggagcgcgaccugggccucuggacucaucggccauugcugaucugcuacaaag
aguccguggaccagagagggcaaccagaucaugagcgacaagcguaacgugauccugucagcguguuugaugagaauaagagcugguauucgaccga
gaacauccagcgguuccucccaaccugccggcgugcagcugggagggaccugaguuccagggcagcaacaucaugcacucuaauaacggcuacgugu
ucgacuccugcagcuuaagcguuugccugcagcagggugguuuacugguacauccugagcaucggcgccagaccgacuuccuguccguguuucucuc
aggcuacacauucaaacaauaagauuguguaugagggacaccugaccucguuuucccgagagaccguguuauagagcauggagaaccugggc
cuuuggauccugggcugucacaauagcgacuucgggaaccggggaaugaccggccugcugaagucagcucuuugcgacaagaacaccggcgacuacu
acgagggacagcuauagggacaucuccgcuaccugcugagcaagaacaacgccaucgagccucggucguuacagccagaacucuaaggcaccacgaccc
cggcagaagcaauuacgcaaccaccauccugaaaacgacaucgaaagaccgauccuugguuucgucacagaaccccauugcccaagauccaaaac
guguccucucugaccugcugaugcugcugagacagagcccaccccacggccugagccugucugaccugcagggagggccaaauacgagacuucuc
cugacgauccuagccccggggccaucgacagcaacaacagccugagcgagauagacuauuuccggccucagcuccaccacucuggagacaugguuuuc
acaccugaagggcccgucgagcugagacugaacgagaagcuagggcaccaccgucggccaggaacugaagaagcuggacuuaagguaaagcagcaccag
caacaaccugaucagcaccuuccuugacaaccuugccggcgacggacaacacagcagccugggccuccuuaauggccugcacuacgacag
ccagcuggaauaccacacuguuuggaagaaaucuaagccucugaccgaaucugcgccgcccucugagccugagcgaaagagaacaacgaugcaaacugc
uggagagcgccugaugaaacagccaggaucacuuggggcaagaaacgugagcaguuaccagggagauacccggacaaccugcagagcgauagga
ggagaucauuuugaugacacuacagcguggagauaagaagagggacuucgacauuacgacgagcagaaaaccaagcccucggagcuuccaga
agaaaaccagacacuauuucaucggcguggagagacuuggggacuacggcaugagcuccucuccacgugcugcggaaccggggccagagcggg
aucuguuuccucaguuaagaagugcuguuuccaagaguacaccgacggcucuuuuccagccucuguauagagggcgaauuaaacgaacaccugggc
cugcugggcccuuacaucagagcagaaguuaggaauaacauugguagccuucagaaaucagggccucugggccuacucuuuacuccucgcuca
ucagcuacgagagggaccagagcgaagcgccgaacccagaagaacuuucgugaagcccaagaaacgaacuuuuggaagguagaacaccac
augggcccuaccaaagacgaguuugauugcaagggccuacuuacggauguggagccuggaaaaggauguccauagcgggccuaguuugggcccu
ugcugguuguccacaaaacacucuaauccugcccacgggagacagguagccgugcaggaugucggccuguuucacacauucgacgagacaaag
agcugguacuucacagaaaacaggaagaaauugcagagcccuugcaacauucagauaggagauccaccuuaaggaauuacagauuccacgc

[1196]

caucaacggaucaucaugggacacacugccagggacugguuauaggccaggaucagagaaucggugguuaccugcucuaugggcucaaacgagaau
auccacuccauccacuuuagcggccacguguuaccgugagaaagaagaggaguuacaagauaggccuguaacaucguacccggcguguuaggaga
caguggaaaugcugccaagcaaggccgggaucugggcgugggaugucuauuaggcgagcaccugcacggcgaugagcacacuguuucugguau
acagcaacaagugccagacacccugggcaugguucugggucacaucaagaguuuucagauacacagccucugggucaguuacggaggggcuccaa
gcuaggccagacugcacuacuccggcuccaucaacggccugggagcacaaaagagccuuucagcuaggaucaaggucgauucugcugggcucaaugaucaucc
acggaaucagaccagggcgccggcagaauuucgagccucuaucucgaguucauuaucauguacagccugggaggaagaagugggagacc
uacagaggaaauaguaccggcaccugauagguguuucuggcaaugggagacgucaggaauuaagcacaaucuucaauccuccaaucagccag
auacaucagacuucacuccacacucuaucagaagcacgucgagaaugggaguggaugggagugaccugaacagcugcagcaugccucugggc
auggaaucaaaagccaucucugaugccagauaccgcuucuauguuacuuaccaacauuucgccaugggucccuaggccagacugcaccu
gcaggggcaguguaauagccugggcgccucaggguaacaaccccaaggaaugggcugcagggcgaauuccagaaaaccaugaaagugaccggagugacc
acccaggggcugaaaagccugcugaccagcaugacugaaaggaguuucugaucucuaagcuccaggaugggccaccagugggaccuccuucuuucaga
acggcagaagugaaaguguucaaggcaaccaggacagcuucacccagucgugaacagccuggauccaccucuccuuaccagauaccugagaauccau
cccaguccggguccaccagauccgucgggaugggagugucggcgucgagggcccaagaccuguaucAAGCUUGAUCCAGACAUG
AUAAGAUACAUGAUGAGUUUGGACAAACCACAACUAGAAUGCAGUGAAAAAAUGCUUUUUU
GUGAAAUUUGUGAUGCUAUUGCUUUUUUGUAACCAUUAUAAGCUGCAAUAAACAAGUUGUGGA
AUUGUGAGCGCUCACAAUCCACAGCGGCCGCGgggcagagcgacaucccccaguccccgagaaguuaggggggagggggu
cggaauuagaccggugccuagagaaggugcgcggggguaaacuggggaagugaugucguguaucuggcuccgcuuuuuccggaggguagggggaga
accguauuaagugcagugagucggcugaaagcuuuuuucgcaacggguuuugcccgacagacacagcguaacggccaccaguguccaaaggcgga
ggaaugaucaaggaaauuauagagauucaaggucagaauaggaaaggaucuaugaacggccacagauucgagauccaggggagaaggagaaggcagacc
uacgagggcacccaaaccgcuaaaauaaaagugacaaggggcgaccucugccguucggcuggggaauuccuguccccaguucauguacggcagcaa
ggccuacgugaagcaccucugcagaucccgauuacaaaaagcugucuuuuuccugaggcgcuuaagugggagagagugaugaauucgaggacgga
ggccugggugaccguuaccgaagacagcagccugcaaggaaggaacacugaucuaaaagugaagugcggggcacuaauuuccuccagacggcccggu
gaugcagaagaaaaccagggcuggggagggccuacagaacggcuguaacccagagacggagugcugaaggcgagauuaccaggcacugaagcug
aaagaugggcgacacuaccuggggaguucaagaccuuuacauaggccaagaagccugugcagcugccuggcuaucacugugacaccaagcugg
acauaccuccacaaugaggauuuuaccuacuggaacaguuacgagagaagcgaggggcagacaccaccuguuuccugggccagggcacagggcuccacc
ggcucuggcagcagcgcgacagccagcagcgaggaacaacaauaggccgugaucaaggaguucaugcgguucaaggugcggauggaaggcaguuaga
acggccaugaguucgaaaucgaggcggaaggcgagaccuuacgaggcgacacagacagcuaggcugaaggugaccaaggaggggccucugcc
auucggcgggacauccugagccucaguuuauuguagggcagaaggccuacugaaagcaccggccgacaucccgacuacaagaagcugagcuucc
cugaggguuucaauagggaaggguagaauucgagggagggcgacugggagacagugaccagagcagccugcaggaugggcaccuccagauca
uaagguagaagagaggcaccacuuuccaccugacggcccgugaugcagaaaaaaccaugggaugggaggcgaccagcgagcgcgcuacccu
agagaugggcuucgaaaggcgagauccaccagggccuagcugaaggagcgggccacuaccuggggaguucaagaccuacauaggccaaga
aaccugugcagcugccuggaucuaucacugggacacaaagcuggacauaccucacacagaggaucuuaccuacugcagcaguuacgaacgcagc

524. TCA5 CBhBsi F8N6 突变体 GeFo3 RNA 序列

[1197]

gggagacgggucaaccgcguaggagcgugaccgGAAUUCGGCGGGAGUAAACU AUGACUCUCUUAAGGAGUCAUAG
AGCCGAAACCUCCUCGUGGUCCCGCUGGGCACAGGGAUUAUUUUUCUGUGGCAAAUUGACUG
GCUUCAGAGAGCGUUUUUCGAAGUGGACUGUGUGACUGCGUUCUUUUUUAGUUGCUAUAUCCG
CUUCGAUUAACAUCUCACCUCCGACGUAAUAGAUAUUUCUAGAAGGUCGACCAGAUGUCCGAGG
UCGACCAGUUGUCCGUGUGGAAUUGUGAGCGCUCACAAUCCACACGUUACAUAACUACGGUA
AAUGGCCCCGCCUGGCUGACCGCCCAACGACCCCCGCCCAUUGACGUCAAUAGUAAACGCCAAUAGG
GACUUUCCAUAUGACGUCAAUGGGUGGAGUAUUUACGGUAAACUGCCCACUUGGCAGUACAUCAA
GUGUAUCAUAUGCCAAGUACGCCCCCUAUUGACGUCAAUGACGGUAAAUGGCCCCGCCUGGCAUU
GUGCCCAGUACAUGACCUUAUGGGACUUUCCUACUUGGCAGUACAUCUACGUAAUAGUCAUCGC
UAUUACCAUGGUCGAGGUGAGCCCCACGUUCUGCUUCACUCUCCCCAUCUCCCCCCCCUCCCCACC
CCCAAUUUUGUAUUUAUUUAUUUUUAAUUAUUUUUGUCAGCGAUGGGGGCGGGGGGGGGGGGG
GGGCGCGCGCCAGGCGGGGGCGGGGCGGGGCGAGGGGCGGGGCGAGGCGGAGAGGUGCGG
CGGCAGCCAAUCAGAGCGGCGCGCUCCGAAAGUUUCCUUUAUUGGCGAGGCGGCGGCGGCGGCGG
CCCUAUAAAAAGCGAAGCGCGGGCGGGCGGGAGUCGCUGCGACGCUGCCUUCGCCCCGUGCCCC
GCUCCGCCGCCGCCUCGCGCCGCCGCCCGGGCUCUGACUGACCGCGUUACUCCACAGGUGAGCG
GGCGGGACGGCCCUUCUCCUCCGGGCUGUAAUUAGCUGAGCAAGAGGUAAAGGGUUUAAGGGAUG
GUUGGUUGGUGGGGUAAUAAUGUUUAAUUACCUUGGAGCACCUGCCUGAAAUCACUUUUUUUCAG
GUUGGCGUACGgcccaccAUGCAAUUCGAGCUGUCCACUUGCUUCUUCUUGGCCUGCUGCGCUUUUG
UUUCAGUGCUACCCCGAGAUACUACCUGGGAGCCGUGGAGCUUUUCUUGGGACUAAUAGCAGAGC
GAUCUGGGCGAGCUGCCAGUCGACGCCAGAUUCCACCUAGAGUGCCUAAGAGUUUCCCCUUCAA
CACCUCUGUGGUGUACAAGAAAACCCUGUUUGUGGAGUUCACCGACCACCUGUUCAAUAUCGCU
AAGCCUCGGCCUCCUUGGAUGGGCCUGCUGGGCCCCACCAUUCAGGCUGAAGUGUACGACACCGU
GGUGAUCACCCUGAAGAACAUGGCCUCCCAACCCUGUCUCUCUGCACGCCGUCGGCGUGUCGUACU

[1198]

GGAAGGCCAGCGAGGGCGCUGAAUACGACGACCAGACCAGCCAAAGAGAGAAGGAGGAUGAUAA
GGUGUUCCCAGGUGGCAGCCACACCUACGUGUGGCAGGUGCUGAAAGAGAACGGCCCAAUGGCC
UCAGACCCUCUGUGCCUGACAUAACAGCUAUUUAUCCCACGUUGACCUGGUGAAGGACCUGAACAG
CGGACUGAUCGGCGCCUGCUGGUGUGCAGAGAAGGCUCCCUGGCCAAGGAAAAGACACAGACAC
UACACAAGUUCUACUUAACUGUUUGCCGUGUUCGACGAGGGCAAGAGCUGGCACAGCGAGACCAA
GAAUAGCCUGAUGCAAGACAGAGAUGCCGCCUCCGCUAGAGCGUGGCCUAAAAUGCACACCGUG
AACGGCUACGUCAACAGAUUCUCUGCCCGGCCUGAUUGGCUGUCACCGGAAGUCCGUGUACUGGCA
CGUGAUCGGAAUGGGCACAACCCCGGAGGUGCACUCAUUCUCCUGGAAGGCCACACGUUCCUGG
UGAGAAACCACCGGCAGGCCUCGCUCGAGAUUCUUUUUACACAUUUCUGACCGCCCAGACCUUG
CUCAUGGACCUGGGCCAGUUUCUGCUUUUCUGCCACAUCAGCUCCCACCAGCACGACGGCAUGGA
GGCUUACGUGAAGGUGGAUUCUUGUCCUGAGGAGCCUCAGCUGAGAAUGAAGAAUAAUGAGGAA
GCCGAAGAUUACGACGACGACCUGACGGACAGCGAGAUGGAUGUUGUGAGGUUUGACGAUGAUA
AUAGCCCUUCCUUUAUCCAGAUCCGGUCAGUGGCGAAGAAACACCCUAAGACAUGGGUGCACUAC
AUCGCCGCUGAGGAGGAGGAUUGGGACUACGCCCCCUGGUGCUGGCCCCUGACGACCGGAGCUA
CAAAUCUCAGUACCUGAACAACGGCCCUCAAAGAAUCGGCAGAAAGUACAAGAAGGUGAGAUUC
AUGGCUUAUACAGAUAGAAACAUUUAAAACCAGAGAAGCUAUCCAGCAUGAGAGCGGAUCCUGG
GACCUCUGCUGUACGGCGAGGUGGGCGACACCCUGCUGAUCAUCUUUAAGAACCAGGCCAGCAGA
CCUUACAUAUUCUACCCUCACGGCAUUAACAGACGUGAGACCUCUGUACAGCCGGCGGCUGCCCAA
GGGCGUGAAGCACCUGAAGGACUUCCCAAUCCUGCCUGGAGAGAUUCUUAAGUACAAGUGGACC
GUGACAGUGGAAGAUGGCCCACCAAGUCUGACCCCCGGUGCCUGACCCGGUACUACUCGAGUUU
UGUGAACAUGGAGCGCGACCUGGCCUCUGGACUCAUCGGCCCAUUGCUGAUCUGCUACAAAGAG
UCCGUGGACCAGAGAGGCAACCAGAUCAUGAGCGACAAGCGUAACGUGAUCCUGUUCAGCGUGU
UUGAUGAGAAUAGAAGCUGGUUUCUGACCGAGAACAUCCAGCGGUUCCUCCCCAACCCUGCCGGC
GUGCAGCUGGAGGACCCUGAGUUCAGGCCAGCAACAUCAUUGCACUCUAUUAACGGCUACGUGU
UCGACUCCCUGCAGCUAAGCGUUUGCCUGCACGAGGUGGCUUACUGGUACAUCUGAGCAUCGGC
GCCCAGACCGACUUCUGUCCGUGUUCUUCUCAGGCUACACAUUCAACAUAAGAUGGUGUAUG
AGGACACCCUGACCCUGUCCCCUUUCCGGAGAGACCGUGUUAUGAGCAUGGAGAACCCUGGC
CUUUGGAUCCUGGGCUGUCACAAUAGCGACUUUCGGAACCGGGGAAUGACCGCCCUGCUGAAAG
UCAGCUCUUGCGACAAGAACACCGGCGACUACUACGAGGACAGCUAUGAGGACAUCUCCGCCUAC
CUGCUGAGCAAGAACAACGCCAUCGAGCCUCGGUCGUUCAGCCAGAACCCCCUGUGCUGAAGAG
ACACCAGAGGGAGAUACCCGGACAACCCUGCAGAGCGAUACAGGAGGAGAUCAUUAUGAUGAC
ACUAUCAGCGUGGAGAUGAAGAAAGAGGACUUCGACAUCUACGACGAGGACGAAAACCAAAGCC

[1199]

CUCGGAGCUUCCAGAAGAAAACCAGACACUAUUUCAUCGCCGCGUGUGGAGAGACUGUGGGACUA
CGGCAUGAGCUCCUCUCCCCACGUGCUGCGGAACCGGGCCCAGAGCGGAUCUGUUCCUCAGUUCA
AGAAAGUCGUGUCCAAGAGUUCACCGACGGCUCUUUUACCCAGCCUCUGUAUAGAGGCGAAUU
AAACGAACACCUGGGCCUGCUGGGCCCUUACAUCAGAGCAGAAGUUGAGGAUAACAUCUAGGGUG
ACCUUCAGAAAUCAGGCCUCUCGGCCCUACUCUUUCUACUCCUCGCUCAUCAGCUACGAGGAGGA
CCAGAGGCAAGGCGCCGAACCCAGAAAGAACUUCGUGAAGCCCAACGAAACCAAGACUUACUUUU
GGAAGGUGCAACACCACAUGGCCCUUACCAAAGACGAGUUUGAUUGCAAGGCCUGGGCCUACUUC
AGCGAUGUGGACCUGGAAAAGGAUGUCCAUAGCGGCCUGAUUUGGCCCUUGCUGGUGUGCCACA
CAAACACACUCAAUCCUGCCCACGGGAGACAGGUGACCGUGCAGGAGUUCGCCUGUUCUUCACC
AUCUUCGACGAGACAAAGAGCUGGUACUUCACAGAAAACAUGGAAAGAAAUUGCAGAGCCCCUU
GCAACAUUCAGAUGGAGGAUCCCACCUUCAAGGAAAAUACAGAUUCCACGCCAUCAACGGUAU
CAUCAUGGACACACUGCCAGGACUGGUUAUGGCCCAGGAUCAGAGAAUCCGGUGGUACCUGCUC
UCUAUGGGCUCAAACGAGAAUAUCCACUCCAUCCACUUUAGCGGCCACGUGUUCACCGUGAGAAA
GAAAGAGGAGUACAAGAUGGCCUGUACAAUCUGUACCCCGGCGUGUUCGAGACAGUGGAAAUG
CUGCCAAGCAAGGCCGGGAUCUGGCGGGUGGAAUGUCUUAUUGGCGAGCACCUGCACGCCGGAA
UGAGCACACUGUUUCUGGUUAUACAGCAACAAGUGCCAGACACCCUGGGCAUGGCUUCUGGUCAC
AUCAGAGAUUUUCAGAUACAGCCUCUGGUCAGUACGGACAGUGGGCUCCCAAGCUGGCCAGAC
UGCACUACUCCGGCUCCAUAACGCCUGGAGCACAAAAGAGCCUUUCAGCUGGAUCAAGGUCGAU
CUGCUGGCUCCAUGAUCAUCCACGGAUCAAGACCCAGGGCGCCCGGCAGAAAUUCUCGAGCCU
CUACAUCUCGCAGUUAUUAUCAUGUACAGCCUGGACGGAAAGAAGUGGCAGACCUACAGAGGA
AAUAGUACCGGCACCCUGAUGGUGUUCUUCGGCAAUGUGGACAGCUCAGGAAUUAAGCACAACA
UCUUCAAUCCUCCAUAUAGCCAGAUACAUCAGACUUCAUCCUACCCACUACUCUAUCAGAAGC
ACGCUGAGAAUGGAGUGGAUGGGAUGUGACCUGAACAGCUGCAGCAUGCCUCUUGGCAUGGAAU
CUAAAGCCAUCUCUGAUGCCAGAUACCGCUUCUAGUUACUUCACCAACAUGUUCGCCACAUGG
UCCCCUAGCAAGGCCAGACUGCACCCUGCAGGGCAGGUCUAAUGCCUGGCGGCCUCAGGUGAACAA
CCCCAAGGAAUGGCUGCAGGUCGAUUUCCAGAAAACCAUGAAAGUGACCGGAGUGACCACCCAG
GGCGUGAAAAGCCUGCUGACCAGCAUGUACGUGAAGGAGUUCUGAUCUCUAGCUCUCCAGGAUG
GCCACCAGUGGACCCUGUUCUUUCAGAACGGCAAAGUGAAAGUGUCCAAGGCAACCAGGACAG
CUUCACCCCAGUCGUGAACAGCCUGGAUCCACCUCUCCUUACCAGAUACCUGAGAAUCCAUCCCC
AGUCCUGGGUCCACCAGAUCCGCCUGCGGAUGGAGGUGCUGGGCUGCGAGGCCCAAGACCUGUAC
UAAGCUUGAUCCAGACAUGAUAGAUAUUGAUGAGUUUGGACAAACCACAACUAGAAUGCAG
UGAAAAAAUAGCUUUUUUGUGAAAAUUUGUGAUGCUAUUGCUUUUUUGUAACCAUUAUAAGCU

GCAAUAAACAAGUUGUGGAAUUGUGAGCGCUCACAAUCCACAGCGGCCGCUGAGGUAGAUAAU
CUUUGUAUAGUGGGGGGGAUCUCAUGUACCGGGUUUCUUUAUUUGAUUUUCAAUAAAACAGA
CGGUAGCUAGGUUCGCAAGGCAGCCACAAGCCAAAGAUAGGUAGGGUGCUCAUAGUGAGUAGGG
ACAGUGCCUUUUGAUUCACAACGCGUCAAUACCAUCUGACACGGAUACCCUUAACCGGACUUGUCA
UGAUCUCCCAGACUUGUCCAAGGUGGACGGGCCACCUUUAACUUAACCCGAAAAGGAACAUAUA
UUAUUUAUUGUGUUCGGAAAAUAGCAAAAAAAAAAAAAAAAAAAAAA

525. TCA5_CBhBsi_F8N6 突变体_EFS_TdT_GeFo3 RNA 序列

[1200] gggagacgguaaccgcuaggagcggugaccgGAAUUCGGCGGGAGUAAUUAUGACUCUCUUAAGGAGUCAUAG
AGCCAGAACCUCUCGUGGUCCCGCUGGGCACAGGGAAUAAUUUUUCUGUGGCAAAUUUGACUG
GCUUCAGAGAGCGUUUUUCGAAGUGGACUGUGUGACUGCGUUCUUUUUUUAGUUGCUAUUAUCCG
CUUCGAUUAACAUCUCACCUCGACGUUAAGAUCAUUUUCUAGAAGGUCGACCAGAUGUCCGAGG
UCGACCAGUUGUCCGUGUGGAAUUGUGAGCGCUCACAAUCCACACGUUACAUAAACUUAACGGUA
AAUGGCCCCGCCUGGCUGACCGCCCAACGACCCCCGCCAUUGACGUCAAUAGUAAACGCCAAUAGG
GACUUUCCAUUGACGUCAAUGGGUGGAGUAUUUACGGUAAACUGCCCACUUGGCAGUACAUCAA
GUGUAUCAUAUGCCAAGUACGCCCCCUAUUGACGUCAAUGACGGUAAUUGGCCCCGCCUGGCAUU
GUGCCCAGUACAUGACCUUAUGGGACUUUCCUACUUGGCAGUACAUCUACGUAAUAGUCAUCGC
UAUUACCAUGGUCGAGGUGAGCCCCACGUUCUGCUUCACUCUCCCCAUUCCCCCCCCUCCCCACC
CCCAAUUUUUGUAUUUAUUUAUUUUUAAUUAAUUUUGUGCAGCGAUGGGGGCGGGGGGGGGGGGG
GGGCGCGCGCCAGGCGGGGCGGGGCGGGGCGAGGGGCGGGGCGGGGCGAGGCGGAGAGGUGCGG
CGGCAGCCAAUCAGAGCGGCGCGCUCCGAAAGUUUCCUUUUUUGGCGAGGCGGCGGGCGGCGG
CCCUAUAAAAAGCGAAGCGCGCGGGCGGGGAGUCGCGUGCGACGCUGCCUUCGCCCCGUGCCCC
GCUCCGCCGCCGCCUCGCGCCGCCCGCCCCGGCUCUGACUGACCGCGUUAUCCACAGGUGAGCG
GGCGGGACGGCCCUUCUCCUCCGGGCGUAAUUAGCUGAGCAAGAGGUAAGGGUUUAAGGGAUG
GUUGGUUGGUGGGGUAAUAAUGUUUAAUUACCUGGAGCACCUGCCUGAAAUCACUUUUUUUCAG
GUUGGCGUACGgcccaccAUGCAAUUCGAGCUGUCCACUUGCUUCUCCUGUGCCUGCUGCGCUUUUG
UUUCAGUGCUACCCGCAGAUACUACCUGGGAGCCGUGGAGCUUUCUUGGGACUAUAUGCAGAGC
GAUCUGGGCGAGCUGCCAGUCGACGCCAGAUUCCACCUAGAGUGCCUAAGAGUUUUUUUUCAA
CACCUCUGUGGUGUACAAGAAAACCCUGUUUUGUGGAGUUCACCGACCACCUGUCAAUAUCGCU
AAGCCUCGGCCUCCUUGGAUGGGCCUGCUGGGCCCCACCAUUCAGGCUGAAGUGUACGACACCGU
GGUGAUCACCCUGAAGAACAUGGCCUCCACCCUGUCUCUCUGCACGCCGUCGGCGUGUCGUACU
GGAAGGCCAGCGAGGGCGCUGAAUACGACGACCAGACCAGCCAAAGAGAGAAGGAGGAUGAUAA

[1201]

GGUGUUCCAGGUGGCAGCCACACCUACGUGUGGCAGGUGCUGAAAGAGAACGGCCCAAUGGCC
UCAGACCCUCUGUGCCUGACAUAACAGCUAUUUUAUCCACGUGACCUGGUGAAGGACCUGAACAG
CGGACUGAUCGGCGCCCUUGCUGGUGUGCAGAGAAGGCUCCUGGCCAAGGAAAAGACACAGACAC
UACACAAGUUCAUCUUACUGUUUGCCGUGUUCGACGAGGGCAAGAGCUGGCACAGCGAGACCAA
GAAUAGCCUGAUGCAAGACAGAGAUGCCGCCUCCGCUAGAGCGUGGCCUAAAAUGCACACCGUG
AACGGCUACGUCAACAGAUUCUGCCCCGGCCUGAUUGGCUGUCACCGGAAGUCCGUGUACUGGCA
CGUGAUCGGAAUGGGCACAACCCCGAGGUGCACUCAAUUCUCCUGGAAGGCCACACGUUCCUGG
UGAGAAACCACCGGCAGGCCUCGCUAGAGAUUCUUUUUAUCACAUUUCUGACCGCCCAGACCUUG
CUCAUGGACCUGGGCCAGUUUCUGCUUUUCUGCCACAUCAGCUCCACCAGCACGACGGCAUGGA
GGCUUACGUGAAGGUGGAUUCUUGUCCUGAGGAGCCUCAGCUGAGAAUGAAGAAUAAUGAGGAA
GCCGAAGAUUACGACGACGACCUGACGGACAGCGAGAUGGAUGUUGUGAGGUUUGACGAUGAUA
AUAGCCCUUCCUUUAUCCAGAUCCGGUCAGUGGCGAAGAAACACCCUAAGACAUGGGUGCACUAC
AUCGCCGUGAGGAGGAGGAUUGGGACUACGCCCCCUGGUGCUGGCCCCUGACGACCGGAGCUA
CAAAUCUCAGUACCUGAACAACGGCCCUCAAAGAAUCGGCAGAAAGUACAAGAAGGUGAGAUUC
AUGGCUUAUACAGAUAGAAACAUUUAAAACCAGAGAAGCUAUCCAGCAUGAGAGCGGAUCCUGG
GACCUCUGCUGUACGGCGAGGUGGGCGACACCCUGCUGAUCAUCUUUAAGAACCAGGCCAGCAGA
CCUUACAUAUCUACCCUCACGGCAUUAACAGACGUGAGACCUCUGUACAGCCGGCGGCUGCCCCAA
GGGCGUGAAGCACCUGAAGGACUUCCCAAUCCUGCCUGGAGAGAUUCUUAAGUACAAGUGGACC
GUGACAGUGGAAGAUGGCCCUACCAAGUCUGACCCCCGGUGCCUGACCCGGUACUACUCGAGUUU
UGUGAACAUGGAGCGCGACCUGGCCUCUGGACUCAUCGGCCCAUUGCUGAUCUGCUACAAAGAG
UCCGUGGACCAGAGAGGCAACCAGAUCAUGAGCGACAAGCGUAACGUGAUCCUGUUCAGCGUGU
UUGAUGAGAAUAGAAGCUGGUUAUCUGACCGAGAACAUCCAGCGGUUCCUCCCCAACCCUGCCGGC
GUGCAGCUGGAGGACCCUGAGUUCAGGCCAGCAACAUC AUGCACUCUAUUAACGGCUACGUGU
UCGACUCCUGCAGCUAAGCGUUUGCCUGCACGAGGUGGCUUACUGGUACAUCUGAGCAUCGGC
GCCCAGACCGACUUCUGUCCGUGUUCUUCUCAGGCUACACAUUCAAAACAUAAGAUGGUGUAUG
AGGACACCCUGACCCUGUUCUUUUUCCGGAGAGACCGUGUUAUGAGCAUGGAGAACCCUGGC
CUUUGGAUCCUGGGCUGUCACAAUAGCGACUUUCGGAACCGGGGAAUGACCGCCCUGCUGAAAG
UCAGCUCUUGCGACAAGAACACCGGCGACUACUACGAGGACAGCUAUGAGGACAUCUCCGCCUAC
CUGCUGAGCAAGAACAACGCCAUCGAGCCUCGGUCGUUCAGCCAGAACCCCCCUGUGCUGAAGAG
ACACCAGAGGGAGAUACCCGGACAACCCUGCAGAGCGAUCAGGAGGAGAUCAUUAUGAUGAC
ACUAUCAGCGUGGAGAUGAAGAAAGAGGACUUCGACAUCUACGACGAGGACGAAAACCAAAGCC
CUCGGAGCUUCCAGAAGAAAACCAGACACUAUUUCAUCGCCGUGUGGAGAGACUGUGGGACUA

[1202]

CGGCAUGAGCUCCUCUCCCCACGUGCGUGCGGAACCGGGCCCAGAGCGGAUCUGUUCCUCAGUUCA
AGAAAGUCGUGUUCCAAGAGUUCACCGACGGCUCUUUUACCCAGCCUCUGUAUAGAGGCGAAUU
AAACGAACACCUGGGCCUGCUGGGCCCUUACAUCAGAGCAGAAGUUGAGGAUAAACAUCAUGGUG
ACCUUCAGAAAUCAGGCCUCUCGGCCCUACUCUUUCUACUCCUCGCUCAUCAGCUACGAGGAGGA
CCAGAGGCAAGGCGCCGAACCCAGAAAGAACUUCGUGAAGCCCAACGAAACCAAGACUUACUUUU
GGAAGGUGCAACACCACAUGGCCCCUACCAAAGACGAGUUUGAUUGCAAGGCCUGGGCCUACUUC
AGCGAUGUGGACCUGGAAAAGGAUGUCCAUAGCGGCCUGAUUGGCCCCUUGCUGGUGUGCCACA
CAAACACACUCAAUCCUGCCCACGGGAGACAGGUGACCGUGCAGGAGUUCGCCCUGUUCUUCACC
AUCUUCGACGAGACAAAGAGCUGGUACUUCACAGAAAACAUGGAAAGAAAUUGCAGAGCCCCUU
GCAACAUUCAGAUGGAGGAUCCCACCUUCAAGGAAAAUACAGAUUCCACGCCAUCAACGGAUA
CAUCAUGGACACACUGCCAGGACUGGUUAUGGCCCAGGAUCAGAGAAUCCGGUGGUACCUGCUC
UCUAUGGGCUCAAACGAGAAUAUCCACUCCAUCCACUUUAGCGGCCACGUGUUCACCGUGAGAAA
GAAAGAGGAGUACAAGAUGGCCCUGUACAUCUGUACCCCCGGCGUGUUCGAGACAGUGGAAAUG
CUGCCAAGCAAGGCCGGGAUCUGGCGGGUGGAAUGUCUUAUUGGCGAGCACCUGCACGCCGGAA
UGAGCACACUGUUUCUGGUUAACAGCAACAAGUGCCAGACACCCUGGGCAUGGCUUCUGGUCAC
AUCAGAGAUUUUCAGAUACAGCCUCUGGUCAGUACGGACAGUGGGCUCCCAAGCUGGCCAGAC
UGCACUACUCCGGCUCCAUCAACGCCUGGAGCACAAAAGAGCCUUUCAGCUGGAUCAAGGUCGAU
CUGCUGGCUCCAUGAUCAUCCACGGAAUCAAGACCCAGGGCGCCCGGCAGAAAUUCUGAGCCU
CUACAUCUCGCAGUUCAUUAUCAUGUACAGCCUGGACGGAAAGAAGUGGCAGACCUACAGAGGA
AAUAGUACCGGCACCCUGAUGGUGUUCUUCGGCAAUGUGGACAGCUCAGGAAUUAAGCACAACA
UCUUCAUCCUCCAUAUAGCCAGAUACAUCAGACUUAUCCUACCCACUACUUAUCAGAAGC
ACGCUGAGAAUGGAGUGGAUGGGAUGUGACCUGAACAGCUGCAGCAUGCCUCUUGGCAUGGAAU
CUAAAGCCAUCUCUGAUGCCCAGAUACCGCUUCUAGUUACUUCACCAACAUGUUCGCCACAUGG
UCCCCUAGCAAGGCCAGACUGCACCUGCAGGGCAGGUCUAAUGCCUGGCGGCCUCAGGUGAACAA
CCCCAAGGAAUGGCUGCAGGUCGAUUUCCAGAAAACCAUGAAAGUGACCGGAGUGACCACCCAG
GGCGUGAAAAGCCUGCUGACCAGCAUGUACGUGAAGGAGUUCUGAUCUCUAGCUCCCAGGAUG
GCCACCAGUGGACCCUGUUCUUUCAGAACGGCAAAGUGAAAGUGUCCAAGGCAACCAGGACAG
CUUCACCCCAGUCGUGAACAGCCUGGAUCCACCUCUCCUUACCAGAUACCUGAGAAUCCAUCCCC
AGUCCUGGGUCCACCAGAUUCGCCUGCGGAUGGAGGUGCUGGGCUGCGAGGCCCAAGACCUGUAC
UAAGCUUGAUCCAGACAUGAUAAAGAUACAUUGAUGAGUUUGGACAAACCACAACUAGAAUGCAG
UGAAAAAAUAGCUUUAUUUGUGAAAAUUUGUGAUGCUAUUGCUUUAUUUGUAACCAUUAUAGCU
GCAAUAAACAAGUUGUGGAAUUGUGAGCGCUCACAAUCCACAGCGGCCGCgggcagagcgcacaucgcca

[1203]

cagucccccgaggaaguuaggggggagggggucggcgaauugaaccggugccuagagaaggugggcgcggggguaaacuggggaagugauugucguguaucuggc
uccgccuuuuuuucccgaggguggggggagaaaccguauauaagugcaguaugucgccgugaacguuuuuuuucgcgaacggguuuugccgccagaacacaggg
cguacgggccaccaugguguccaaaggcgagggaagugaucaaggaaauuuuagagauucaaggugcagaauagggaaggaucauugaacgggccacgaguuucg
agauccgaggggagaaaggaggaaggcgagaccuacgaggggcacccaaaccgcuaaaauuaaaagugacaaaggggcgaggaccucugccguuocgcccugggauauc
cugucccccccaguucauguacggcgagcaaggccuacgugaagcaccucgugacauccccgauuacaaaaagcugucuuuuuccugaggggcuuuuaaguu
ggggagagagugaugaaacuucgaggacggaggccuggugaccguuacccaagacagcagccugcagaaguggaacacugaucaacaagugaagauugcg
ggggcacuaauuuuccuccagacggcccccgugaugcagaagaaaaccuaggggcgaggggccucucacagaacggcguguaccccagagacgggagugugcug
aaggggcgagauuaccaggcgacugaaagcugaaggaugggcgacacuaccuggguggaguucaagaccuuuacauggccaaagaagccugugcagcugc
cuggcuaacuacuacguggacaccaagcuggacauaccuccacaaugagggaauuauaccaucguggaacaguuacgagagaagcgaggggcagacaccac
cuguuuccucggccacggcacagggcuccaccgggucucgugcagcagcggcgacagccagcagcgaggacaacaacaugggccgugaucaaggagauucaugcg
guucaaggugcgaggauagggaaggcaguuagaacggccaugaguuucgaaaucgaggggcgaaaggcgaaaggcagaccuuacgaggggcacacagacagcuuag
cugaaggugaccaaggggaggccucugccauucggccuggggacauccugaggccucaguuuauuguauggcagcaaggccuacgugaagcaccgcccg
acaucggcgacuacaagaagcugagcuuccugaggguuuucaaauggggaaaggguugaauaacuucgaggacggcgggacuggugacagugacccaggga
cagcagccugcaggauggcaccucgaucauuaaggugaaagauagagggcaccaacuuccaccugacggggcccgugaugcagaaaaaaaccaugggga
gggaggccagcaccggagcgcguguaaccuagagaugggcguuucugaaggggcgagauccaccaggcccguaagcugaaggacggcgggccacuaccuggu
ggaguucaagaccuacuacauggccaaagaaaccugugcagcugccuggaucuaacuacugggacacaaagcuggacauaccucucacaacgaggga
auaccaucgucgagcaguuacgaacgcagcgaaaggcgagacaccaucguuuccuguaaggguauaggagcagcuguacaaguaagauuaauuaaaauaaaa
gaucuuuauuguuacuuagauucuguguuugguuuuuugugugugggaaauugugagcgcucacaaauuccacagcgcccgUGAGGUAGAU
AAUCUUUGUAUAGUGGGGGGGGAUCUCAUGUACCGGGUUCUUUUUAUUUGAUUUUCAUAAAAAC
AGACGGUAGCUAGGUUCGCAAGGCAGCCACAAGCCAAAGAUAGGUAGGGUGUCUCAUAGUGAGUA
GGGACAGUGCCUUUUGAUUCACAACGCGUCAAUACCAUCUGACACGGAUACCCUUAACGGACUUG
UCAUGAUCUCCAGACUUGUCCAAGGUGGACGGGCCACCUUUAACUUAACCCGGAAAAGGAACAU
AUUUUUUUUUUAUUGUGUUCGGAAAAUAGCAAAAAAAAAAAAAAAAAAAAAA

526. TCARZ CBhBsi GFP GeFo3 R4A22

gggagacggucaaccgcuaggagcggugaccgGAAUUCGGCGGGAGUAACUAUGACUCUCUUAAGGAGUCAUAG
AGCCGAAACCUCCUCGUGGUCCCGCUGGGCACAGGGAUUAUUUUUUCUGUGGCAAAUUUGACUG
GCUCUAGAAGGUCGACCAGAUGUCCGAGGUCGACCAGUUGUCCGUGUGGAAUUGUGAGCGCUCU
CAAUCCACACGUUACAUAACUUAACGGUAAAUGGCCCCGCCUGGCUGACCGCCCAACGACCCCCGC
CCAUUGACGUCAAUAGUAACGCCAAUAGGGACUUUCCAUUGACGUCAAUGGGUGGAGUAUUUAC
GGUAAACUGCCCCAUUGGCAGUACAUCAAGUGUAUCAUAUGCCAAGUACGCCCCCUAUUGACGUC
AAUGACGGUAAAUGGCCCCGCCUGGCAUUGUGCCCAGUACAUGACCUUAUGGGACUUUCCUACUU

[1204]

GGCAGUACAUCUACGUUUAGUCAUCGCUAUUACCAUGGUCGAGGUGAGCCCCACGUUCUGCUU
CACUCUCCCCAUCUCCCCCCCCUCCCCACCCCCAAUUUUGUAUUUAUUUAUUUUUAAUUAUUUU
GUGCAGCGAUGGGGGCGGGGGGGGGGGGGGGCGCGGCCAGGCGGGGCGGGGCGGGGCGAGGG
GCGGGGCGGGGCGAGGCGGAGAGGUGCGGCGGCAGCCAAUCAGAGCGGCGCGCUCCGAAAGUUU
CCUUUUAUGGCGAGGCGGCGGCGGCGGCGGCCCUAUA AAAAGCGAAGCGCGGCGGGGCGGGAG
UCGCUGCGACGCUGCCUUCGCCCCGUGCCCCGCUCCGCCGCCGCCUCGCGCCGCCGCCCGGCUC
UGACUGACCGGUUACUCCACAGGUGAGCGGGCGGGACGGCCCUUCUCCUCCGGGCGUGAAUUA
GCUGAGCAAGAGGUAAAGGUUUAAGGGAUGGUUGGUUGGUGGGGUAUUA AUGUUUAAUUAACCU
GGAGCACCUGCCUGAAAUCACUUUUUUCAGGUUGGCGUACGGCCACCAUGGUGAGCAAGGGCG
AGGAGCUGUUCACCGGGGUGGUGCCCAUCCUGGUCGAGCUGGACGGCGACGUA AACGGCCACAA
GUUCAGCGUGUCCGGCGAGGGCGAGGGCGAUGCCACCUACGGCAAGCUGACCCUGAAGUUCAUCU
GCACCACCGGCAAGCUGCCCCGUGCCCUGGCCACCCUCGUGACCACCCUGACCUACGGCGUGCAG
UGCUCAGCCGCUACCCCGACCACAUGAAGCAGCACGACUUCUUAAGUCCGCCAUGCCCGAAGG
CUACGUCCAGGAGCGCACCAUCUUCUUAAGGACGACGGCAACUACAAGACCCGCGCCGAGGUGA
AGUUCGAGGGCGACACCCUGGUGAACCGCAUCGAGCUGAAGGGCAUCGACUUAAGGAGGACGG
CAACAUCUGGGGCAAGCUGGAGUACAACUACAACAGCCACAACGUCUAUAUCAUGGCCGACA
AGCAGAAGAACGGCAUCAAGGUGAACUUAAGAUCGCCACAACAUCGAGGACGGCAGCGUGCA
GCUCGCCGACCACUACCAGCAGAACCCCCAUCGGCGACGGCCCCGUGCUGCUGCCCCGACAACC
ACUACCUGAGCACCCAGUCCGCCUGAGCAAAGACCCCAACGAGAAGCGCGAUCACAUGGUCCUG
CUGGAGUUCGUGACCGCCCGGGGAUCACUCUGGCAUGGACGAGCUGUACAAGUAAAAGCUUG
AUCCAGACAUGAUAAAGAUACAUUGAUGAGUUUGGACAAACCACAACUAGAAUGCAGUGAAAAA
AUGCUUUUUUGUGAAAAUUUGUGAUGCUAUUGCUUUUUUGUAACCAUUAUAAGCUGCAAUAAA
CAAGUUGUGGAAUUGUGAGCGCUCACAAUCCACAGCGCCGCUGAGGUAGAUAAUCUUUGUAU
AGUGGGGGGGGAUCUCAUGUACCGGGUUUCUUUAUUUGAUUUUCAAUAAAACAGACGGUAGCU
AGGUUCGCAAGGCAGCCACAAGCCAAAGAUAGGUAGGGUGCUCAUAGUGAGUAGGGACAGUGCC
UUUUGAUUCACAACGCGUCAAUACCAUCUGACACGGAUACCCUUAACCGGACUUGUCAUGAUCUC
CAGACUUGUCCAAGGUGGACGGGCCACCUUUAACCCGGAAGGAACAUAUAUUAAUUAU
AUGUGUUCGGAUUAGCAAAAAAAAAAAAAAAAAAAAAA

527. TCARZ_CBh_GFP_GeFo3

gggagacggucaaccgguaggagcgggaccggaauucGGCGGGAGUAAUAUGACUCUCUUAAGGAGUCAUAGAG
CCAGAACCUCUCGUGGUCCCGCUGGGCACAGGGAUUAUUUUUCUGUGGCAAAUUUGACUGGC

[1205]

UCUAGAAGGUCGACCAGAUGUCCGAGGUCGACCAGUUGUCCGUGUGGAAUUGUGAGCGCUCACA
AUUCCACACGUUACAUAACUACGGUAAAUGGCCCCGCCUGGCUGACCGCCCAACGACCCCCGCC
AUUGACGUCAAUAGUAACGCCAAUAGGGACUUUCCAUUGACGUCAAUGGGUGGAGUAUUUACGG
UAAACUGCCCACUUGGCAGUACAUAAGUGUAUCAUAUGCCAAGUACGCCCCCUAUUGACGUCAA
UGACGGUAAAUGGCCCCGCCUGGCAUUGUGCCCAGUACAUGACCUUAUGGGACUUUCCUACUUGG
CAGUACAUCUACGUAAUAGUCAUCGCUAAUACCAUGGUCGAGGUGAGCCCCACGUUCUGCUUCAC
UCUCCCCAUCUCCCCCCCCUCCCCACCCCCAAUUUUGUAUUUAUUUAUUUUUAAUUUAUUUUGUG
CAGCGAUGGGGGCGGGGGGGGGGGGGGGGGCGCGCGCCAGGCGGGCGGGGCGGGGCGAGGGGCG
GGGCGGGGCGAGGCGGAGAGGUGCGGCGGCAGCCAAUCAGAGCGGCGCGCUCCGAAAGUUUCCU
UUUAUGGCGAGGCGGCGGCGGCGGCCCCUAUAAAAAGCGAAGCGCGGGCGGGGCGGGAGUCG
CUGCGACGCUGCCUUCGCCCCGUGCCCCGCUCCGCCGCCGCCUCGCGCCCGCCCGCCCGGCUCUGA
CUGACCGCGUUAUCCACAGGUGAGCGGGCGGGACGGCCUUCUCCUCCGGGCUGUAAUUAGCU
GAGCAAGAGGUAAGGGUUUAAGGGAUGGUUGGUUGGUGGGGUAAUAAUGUUUAAUUACCUGGA
GCACCUGCCUGAAAUCACUUUUUUUCAGGUUGGCGUACGGCCACCAUGGUGAGCAAGGGCGAGG
AGCUGUUCACCGGGGUGGUGCCCAUCCUGGUCGAGCUGGACGGCGACGUAAACGGCCACAAGUUC
AGCGUGUCCGGCGAGGGCGAGGGCGAUGCCACCUACGGCAAGCUGACCCUGAAGUUAUCUGCAC
CACCGCAAGCUGCCCGUGCCCUGGCCCCACCCUCGUGACCACCCUGACCUACGGCGUGCAGUGCU
UCAGCCGCUACCCCGACCACAUGAAGCAGCACGACUUCUUAAGUCCGCCAUGCCCCGAAGGCUAC
GUCCAGGAGCGCACCAUCUUCUUAAGGACGACGGCAACUACAAGACCCGCGCCGAGGUGAAGUU
CGAGGGCGACACCCUGGUGAACCGBAUCGAGCUGAAGGGCAUCGACUUAAGGAGGACGGCAAC
AUCCUGGGGCACAAGCUGGAGUACAACUACAACAGCCACAACGUCUAUAUCAUGGCCGACAAGCA
GAAGAACGGCAUCAAGGUGAACUUAAGAUCGCGCCACAACAUCGAGGACGGCAGCGUGCAGCUC
GCCGACCACUACCAGCAGAACACCCCCAUCGGCGACGGCCCCGUGCUGCUGCCCGACAACCACUA
CCUGAGCACCCAGUCCGCCUGAGCAAAGACCCCAACGAGAAGCGCGAUCACAUGGUCCUGCUGG
AGUUCGUGACCGCCGCCGGGAUCACUCUCGGCAUGGACGAGCUGUACAAGUAAAAGCUUGAUCC
AGACAUGAUAAAGAUACAUGAUGAGUUUGGACAAACCACAACUAGAAUGCAGUGAAAAAAUUGC
UUUAUUUGUGAAAUUUGUGAUGCUAUUGCUUUUUUGUAACCAUUAAGCUGCAAUAAACAAG
UUUGGGAUUUGUGAGCGCUCACAAUCCACAGCGGCCGUGAGGUAGAUAAUCUUUGUAUAGUG
GGGGGGGAUCUAUGUACCGGGUUUCUUUUUUUGAUUUUCAUAAAACAGACGGUAGCUAGGU
UCGCAAGGCAGCCACAAGCCAAAGAUAGGUAGGGUGCUAUAGUGAGUAGGGACAGUGCCUUUU
GAUUCACAACCGGUCAAUACCAUCUGACACGGAUACCCUUACCGGACUUGUCAUGAUCUCCAGA
CUUGUCCAAGGUGGACGGGCCACCUUACUUAACCCGAAAAGGAACAUAAUUAUUUAUUGU

GUUCGGAAAAUAGCAAAAAAAAAAAAAAAAAAAAAA

528. TCARZ_CMV*_GFP_GeFo3

ggguaaacGGCGGGAGUAACUAUGACUCUCUUAAGGAGUCAUAGAGCCAGAACCUCUCGUGGUCCC
GCUGGGCACAGGGAUUAUUUUUCUGUGGCAAUUUGACUGGCUCUAGACGUUACAUAAACUAC
GGUAAAUGGCCCCGCCUGGCUGACCGCCCAACGACCCCCGCCCAUUGACGUCAAUAUGACGUAUG
UUCCCAUAGUAACGCCAAUAGGGACUUUCCAUUGACGUCAAUGGGUGGAGUAUUUACGGUAAAC
UGCCCACUUGGCAGUACAUCAAGUGUAUCAUAUGCCAAGUACGCCCCCUAUUGACGUCAAUGACG
GUAAAUGGCCCCGCCUGGCAUUAUGCCCAGUACAUGACCUUAUGGGACUUUCCUACUUGGCAGUA
CAUCUACGUAAUAGUCAUCGCUAAUACCAUGGUGAUGCGGUUUUGGCAGUACAUAUGGGCGU
GGAUAGGGGUUUGACUCACGGGGAUUUCCAAGUCUCCACCCCAUUGACGUCAAUGGGAGUUUGU
UUUGGCACCAAAUCAACGGGACUUUCCAAAUGUCGUAACAACUCCGCCCAUUGACGCAAAUG
GGCGGUAGGCGUGUACGGUGGGAGGUCUAUAUAAGCAGAGCUCGUUUAGUGAACCGCUUGUCUC
GCUCCGGGGAACGCUCGGAAACUCCCGCCGCCGCCACCCGCGUCUGUUCUGUUACACAAGGGAA
GAAAAGCCGUGCCGCACUCCGAGUGUCGUACGGCCACCAUGGUGAGCAAGGGCGAGGAGCUGU
UCACCGGGGUGGUGCCCAUCCUGGUCGAGCUGGACGGCGACGUAAACGGCCACAAGUUCAGCGUG
[1206] UCCGGCGAGGGCGAGGGCGAUGCCACCUACGGCAAGCUGACCCUGAAGUUAUCUGCACCACCGG
CAAGCUGCCCCGUGCCCCUGGCCACCCUCGUGACCACCCUGACCUACGGCGUGCAGUGCUUCAGCC
GCUACCCCGACCACAUGAAGCAGCACGACUUCUUAAGUCCGCCAUGCCCGAAGGCUACGUCCAG
GAGCGACCAUCUUCUUAAGGACGACGGCAACUACAAGACCCGCGCCGAGGUGAAGUUCGAGGG
CGACACCCUGGUGAACCGCAUCGAGCUGAAGGGCAUCGACUUAAGGAGGACGGCAACAUCUGG
GGCACAAGCUGGAGUACAACUACAACAGCCACAACGUCUAUAUCAUGGCCGACAAGCAGAAGAA
CGGCAUCAAGGUGAACUUAAGAUCGCCACAACAUCGAGGACGGCAGCGUGCAGCUCGCCGACC
ACUACCAGCAGAACACCCCCAUCGGCGACGGCCCCGUGCUGCUGCCCGACAACCACUACCUGAGC
ACCCAGUCCGCCUGAGCAAAGACCCCAACGAGAAGCGCGAUCACAUGGUCCUGCUGGAGUUCGU
GACCGCCGCCGGGAUCACUCUCGGCAUGGACGAGCUGUACAAGUAAAAUAAAAGAUCUUUAUGU
UCAUUAGAUCUGUGUGUUGGUUUUUGUGUGGCUGAGGUAGAUAAUCUUUGUAUAGUGGGGGGG
GAUCUCAUGUACCGGGUUUCUUUUUUUUGAUUUUCAAUAAAACAGACGGUAGCUAGGUUCGCAA
GGCAGCCACAAGCCAAAGAUAGGUAGGGUGCUCUAUGUGAGUAGGGACAGUGCCUUUUGAUUCA
CAACGCGUCAAUACCAUCUGACACGGAUACCCUUAACCGGACUUGUCAUGAUCUCCAGACUUGUC
CAAGGUGGACGGGCCACCUUUAUUAAACCCGGAAAAGGAACAUAUAUUAAUUAUAUGUGUUCGG
AAAAUAGCAAAAAAAAAAAAAAAAAAAAAA

529. TCARZ_SV40*_GFP_GeFo3

ggguaaacGGCGGGAGUAACUAUGACUCUCUUAAGGAGUCAUAGAGCCAGAACCUCUCGUGGUCCC
GCUGGGCACAGGGAUUAUUUUUCUGUGGCAAUUUGACUGGCUCUAGAGCUGUGGAAUGUGUG
UCAGUUAGGGUGUGGAAAGUCCCCAGGCUCCCCAGCAGGCAGAAGUAUGCAAAGCAUGCAUCUC
AAUUAGUCAGCAACCAGGUGUGGAAAGUCCCCAGGCUCCCCAGCAGGCAGAAGUAUGCAAAGCA
UGCAUCUCAUUAGUCAGCAACCAUAGUCCCCGCCCUAACUCCGCCCAUCCCGCCCCUAAUCUCG
CCCAGUUCCGCCCAUUCUCCGCCCAUGGCUGACUAAUUGCGUCUAUAUAUGCAGAGGCCGAGGC
CGCCUCGGCCUCUGAGCUAUUCCAGAAGUAGUGAGGAGGCCGUACGGCCACCAUGGUGAGCAAG
GGCGAGGAGCUGUUCACCGGGGUGGUGCCCAUCCUGGUCGAGCUGGACGGCGACGUAAACGGCC
ACAAGUUCAGCGUGUCCGGCGAGGGCGAGGGCGAUGCCACCUACGGCAAGCUGACCCUGAAGUUC
AUCUGCACCACCGGCAAGCUGCCCGUGCCCGGGCCACCCUCGUGACCACCCUGACCUACGGCGU
GCAGUGCUUCAGCCGCUACCCCGACCACAUGAAGCAGCACGACUUCUUAAGUCCGCCAUGCCCG
AAGGCUACGUCCAGGAGCGCACCAUCUUCUUAAGGACGACGGCAACUACAAGACCCGCGCCGAG
GUGAAGUUCGAGGGCGACACCCUGGUGAACCGCAUCGAGCUGAAGGGCAUCGACUUAAGGAGG
ACGGCAACAUCCUGGGGCACAAGCUGGAGUACAACUACAACAGCCACAACGUCUAUAUCAUGGCC
[1207] GACAAGCAGAAGAACGGCAUCAAGGUGAACUUAAGAUCGCCACAACAUCGAGGACGGCAGCG
UGCAGCUCGCCGACCACUACCAGCAGAACACCCCAUCGGCGACGGCCCCGUGCUGCUGCCCGAC
AACCACUACCUGAGCACCCAGUCCGCCUGAGCAAAGACCCCAACGAGAAGCGCGAUCACAUGGU
CCUGCUGGAGUUCGUGACCGCCGCCGGGAUCACUCUCGGCAUGGACGAGCUGUACAAGUAAAAU
AAAAGAUCUUUAUGUUCAUUAGAUCUGUGUGUUGGUUUUUGUGUGGCUGAGGUAGAUAAUCUU
UGUAUAGUGGGGGGGGAUCUCAUGUACCGGUUUUCUUUAUUUGAUUUUCAAUAAAACAGACGG
UAGCUAGGUUCGCAAGGCAGCCACAAGCCAAAGAUAGGUAGGGUGCUCUAGUGAGUAGGGACA
GUGCCUUUUGAUUCACAACGCGUCAAUACCAUCUGACACGGAUACCCUUAACCGGACUUGUCAUGA
UCUCCCAGACUUGUCCAAGGUGGACGGGCCACCUUACUUAACCCGGAAGGAACAUAUAUUA
AUUAUAUGUGUUCGGAAAAUAGCAAAAAAAAAAAAAAAAAAAAAA

530.

TCA5_CBh_mCherry_GeFo3

gggagacggucaaccgcuaggagcggugaccggaauucGGCGGGAGUAACUAUGACUCUCUUAAGGAGUCAUAGAG
CCAGAACCUCUCGUGGUCCCGCUGGGCACAGGGAUUAUUUUUCUGUGGCAAUUUGACUGGC
UUCAGAGAGCGUUUUUCGAAGUGGACUGUGUGACUGCGUCCCCCUUAGUUGCUAUAUCCGCU
UCGAUUAACAUCUCACCUCGACGUUAAGAUCAUUUCUAGAAGGUCGACCAGAUGUCCGAGGUC

GACCAGUUGUCCGUGUGGAAUUGUGAGCGCUCACAAUCCACACGUUACAUAAACUUACGGUAAA
UGGCCCCGCCUGGCUGACCGCCCAACGACCCCCGCCCAUUGACGUCAAUAGUAACGCCAAUAGGGA
CUUUCCAUUGACGUCAAUGGGUGGAGUAUUUACGGUAAACUGCCCACUUGGCAGUACAUCAAGU
GUAUCAUAUGCCAAGUACGCCCCCUAUUGACGUCAAUGACGGUAAAUGGCCCGCCUGGCAUUGU
GCCCAGUACAUGACCUUAUGGGACUUUCCUACUUGGCAGUACAUCUACGUAAUAGUCAUCGCUA
UUACCAUGGUCGAGGUGAGCCCCACGUUCUGCUUCACUCUCCCCAUCUCCCCCCCCUCCCCACCC
CAAUUUUGUAUUUAUUUAUUUUUAUUUAUUUUGUGCAGCGAUGGGGGCGGGGGGGGGGGGGGG
GGCGCGCGCCAGGCGGGGCGGGGCGGGGCGAGGGGCGGGGCGGGGCGAGGCGGAGAGGUGCGGC
GGCAGCCAAUCAGAGCGGCGCGCUCCGAAAGUUUCCUUUUUUAUGGCGAGGCGGCGGCGGCGGCGG
CCUAUAAAAAGCGAAGCGCGCGGGCGGGGAGUCGCUGCGACGCGUCCUUCGCCCCGUGCCCCG
CUCCGCCGCCGCCUCGCGCCGCCCGCCCCGGCUCUGACUGACCGCGUUACUCCACAGGUGAGCGG
GCGGGACGGCCCUUCUCCUCCGGGCUGUAAUAGCUGAGCAAGAGGUAAGGGUUUAAGGGAUGG
UUGGUUGGUGGGGUAAUAAUGUUUAAUUACCUGGAGCACCUGCCUGAAAUCACUUUUUUUACAGG
UUGGCGUACGGCCACCAUGGUGAGCAAGGGCGAGGAGGAUAACAUGGCCAUCAUCAAGGAGUUC
AUGCGCUUCAAGGUGCACAUGGAGGGCUCCGUGAACGGCCACGAGUUCGAGAUCGAGGGCGAGG
GCGAGGGCGCGCCCUACGAGGGCACCCAGACCGCCAAGCUGAAGGUGACCAAGGGUGGCCCCCUG
CCCUUCGCCUGGGACAUCUGUCCCCUCAGUUCAUGUACGGCUCCAAGGCCUACGUGAAGCACCC
CGCCGACAUCCCCGACUACUUGAAGCUGUCCUUCCCCCGAGGGCUUCAAGUGGGAGCGCGUGAUGA
ACUUCGAGGACGGCGGCGUGGUGACCGUGACCCAGGACUCCUCCCUGCAGGACGGCGAGUUCAUC
UACAAGGUGAAGCUGCGCGGCACCAACUCCCCUCCGACGGCCCCGUAAUGCAGAAGAAAAACCAU
GGGCGGGGAGGCCUCCUCCGAGCGGAUGUACCCCGAGGACGGCGCCCGUAAGGGCGAGAUCAAGC
AGAGGCUGAAGCUGAAGGACGGCGGCCACUACGACGCUAGGCUAAGACCACCUACAAGGCCAA
GAAGCCCGUGCAGCUGCCCGGCGCCUACAACGUCAACAUCAAGUUGGACAUCACCUCCCAACG
AGGACUACACCAUCGUGGAACAGUACGAACGCGCCGAGGGCCGCCACUCCACCGGCGGCAUGGAC
GAGCUGUACAAGUAGAAGCUUGAUCCAGACAUGAUAAAGAUACAUUGAUGAGUUUGGACAAACCA
CAACUAGAAUGCAGUGAAAAAAAUGCUUUUUUUGUGAAAAUUUGUGAUGCUAUUGCUUUUAUUUGU
AACCAUUUAAGCUGCAAUAAACAAGUUUGUGGAAUUGUGAGCGCUCACAAUCCACAGCGGCCG
CUGAGGUAGAUAAUCUUUGUAUAGUGGGGGGGGAUCUCAUGUACCGGGUUUCUUUUUAUUUGAUU
UUCAAUAAAACAGACGGUAGCUAGGUUCGCAAGGCAGCCACAAGCCAAAGAUAGGUAGGGUGCU
CAUAGUGAGUAGGGACAGUGCCUUUGAUUCACAACGCGUCAAUACCAUCUGACACGGAUACCC
UUACCGGACUUGUCAUGAUCUCCAGACUUGUCCAAGGUGGACGGGCCACCUUACUUAACCCGG
AAAAGGAACAUAAUAAUUAUUUAUUGUGUUCGGAAAAUAGCAAAAAAAAAAAAAAAAAAAAAA

531. TCARZ_CMV*_mCherry_GeFo3

ggguaaacGGCGGGAGUAAACUAUGACUCUCUUAAGGAGUCAUAGAGCCAGAACCUCUCGUGGUCCC
GCUGGGCACAGGGAUUAUUUUUCUGUGGCAAUUUGACUGGCUCUAGACGUUACAUAACUUAC
GGUAAAUGGCCCCGCCUGGCUGACCGCCCAACGACCCCCGCCCAUUGACGUCAAUAAUGACGUAAUG
UCCCCAUAGUAACGCCAAUAGGGACUUUCCAUGACGUCAAUGGGUGGAGUAUUUACGGUAAAC
UGCCCACUUGGCAGUACAUCAAGUGUAUCAUAUGCCAAGUACGCCCCCUAUUGACGUCAAUGACG
GUAAAUGGCCCCGCCUGGCAUUAUGCCCAGUACAUGACCUUAUGGGACUUUCCUACUUGGCAGUA
CAUCUACGUUUAGUCAUCGCUAUUACCAUGGUGAUGCGGUUUUGGCAGUACAUCAUUGGGCGU
GGAUAGGGGUUUUGACUCACGGGGAUUUCCAAGUCUCCACCCCAUUGACGUCAAUGGGAGUUUGU
UUUGGCACCAAAUACAACGGGACUUUCCAAAAUGUCGUAACAACUCCGCCCAUUGACGCAAAUG
GGCGGUAGGCGUGUACGGUGGGAGGUCUAUAUAAGCAGAGCUCGUUUAGUGAACCGCUUGUCUC
GCUCCGGGGAACGCUCGGAAACUCCCGGCCGCCGCCACCCGCGUCUGUUCUGUUACACAAGGGAA
GAAAAGCCGCUGCCGCACUCCGAGUGUCGUACGGCCACCAUGGUGAGCAAAGGCGAAGAAGUA
ACAUGGCGAUUAUUAAAGAAUUUAUGCGCUUUAAGUGCAUAUGGAAGGCAGCGUGAACGGCCA
UGAAUUUGAAAUUGAAGGCGAAGGCGAAGGCCGCCCGUAUGAAGGCACCCAGACCGCGAAACUG
AAAGUGACCAAAGGCGGCCCGCUGCCGUUUUGCGUGGGAUAUUCUGAGCCCGCAGUUUAUGUAUG
GCAGCAAAGCGUAUGUGAAACAUCGGCGGAUAUUCGGAUUAUCUGAAACUGAGCUUUCGGGA
AGGCUUUAAAUGGGAACGCGUGAUGAACUUUGAAGAUGGCGGCGUGGUGACCGUGACCCAGGAU
AGCAGCCUGCAGGAUGGCGAAUUUAUUUAUAAAGUGAAACUGCGCGGCACCAACUUUCCGAGCG
AUGGCCCCGUGAUGCAGAAAAAAACCAUGGGCUGGGAAGCGAGCAGCGAACGCAUGUAUCCGGA
AGAUGGCGCGCUGAAAGGCGAAAUUAAACAGCGCCUGAAACUGAAAGAUGGCGGCCAUUAUGAU
GCGGAAGUGAAAAACCACCUAUAAAGCGAAAAAAACCGGUGCAGCUGCCGGGCGCGUAUAACGUGA
ACAUUAAACUGGAUAUUACCAGCCAUAACGAAGAUUAUACCAUUGUGGAACAGUAUGAACGCGC
GGAAGGCCGCCAUAGCACCGGCGGCAUGGAUGAACUGUAUAAAUAGUGAAAGCUUAAUAAAAGA
UCUUUAUGUUCAUUAGAUCUGUGUGUUGGUUUUUGUGUGGCGGCCGCUAGGGUAGAUAAUCUUU
GUUAUGUGGGGGGGGAUCUCAUGUACCGGGUUUCUUUAUUUGAUUUUCAAUAAAACAGACGGU
AGCUAGGUUCGCAAGGCAGCCACAAGCCAAAGAUAGGUAGGGUGCUCAUAGUGAGUAGGGACAG
UGCCUUUUGAUUCACAACGCGUCAAUACCAUCUGACACGGAUACCCUUAACCGGACUUGUCAUGAU
CUCCCAGACUUGUCCAAGGUGGACGGGCCACCUUUAACCUUAACCCGGAAGGAACAUAUAUAA
UUUAUUGUGUUCGGAAAAUAGCAAAAAAAAAAAAAAAAAAAAAA

[1209]

532. TCARZ_SV40*_mCherry_GeFo3

ggguaaacGGCGGGAGUAACUAUGACUCUCUUAAGGAGUCAUAGAGCCAGAACCUCUCGUGGUCCC
GCUGGGCACAGGGAUUAUUUUUCUGUGGCAAAUUUGACUGGCUCUAGAGCUGUGGAAUGUGUG
UCAGUUAGGGUGUGGAAAGUCCCCAGGCUCCCCAGCAGGCAGAAGUAUGCAAAGCAUGCAUCUC
AAUUAGUCAGCAACCAGGUGUGGAAAGUCCCCAGGCUCCCCAGCAGGCAGAAGUAUGCAAAGCA
UGCAUCUCAUUAGUCAGCAACCAUAGUCCCGCCCCUACUCCGCCCAUCCCGCCCCUACUCCG
CCCAGUUCCGCCCAUUCUCCGCCCAUGGCUGACUAAUUGCGUCUAUAUAUGCAGAGGCCGAGGC
CGCCUCGGCCUCUGAGCUAUUCCAGAAGUAGUGAGGAGGCCGUACGGCCACCAUGGUGAGCAAA
GGCGAAGAAGAUACAUGGCGAUUAUUAAAGAAUUUAUGCGCUUUAAGUGCAUAUGGAAGGCA
GCGUGAACGGCCAUGAAUUUGAAAUGAAGGCGAAGGCGAAGGCCGCCCGUAUGAAGGCACCCA
GACCGCGAAACUGAAAAGUGACCAAAGGCGGCCCGCUGCCGUUUGCGUGGGAUAUUCUGAGCCCCG
AGUUUAUGUAUGGCAGCAAAGCGUAUGUGAAACAUCGGCGGAUAUUCGGGAUUAUCUGAAACU
GAGCUUUCGGGAAGGCUUUAUUGGGAACGCGUGAUGAACUUUGAAGAUGGCGGCGUGGUGACC
GUGACCCAGGAUAGCAGCCUGCAGGAUGGCGAAUUUAUUUAUAAAGUGAAACUGCGCGGCACCA
ACUUUCCGAGCGAUGGCCCGGUGAUGCAGAAAAAAACCAUGGGCUGGGAAGCGAGCAGCGAACG
CAUGUAUCCGGAAGAUGGCGCGCUGAAAAGGCGAAAUUAAACAGCGCCUGAAACUGAAAAGAUGGC
GGCCAUAUGAUGCGGAAGUGAAAACCAUUAUAAAGCGAAAAAACCGGUGCAGCUGCCGGGCG
CGUAUAACGUGAACAUAUAAACUGGAUAUUAACAGCCAUAACGAAGAUUAUACCAUUGUGGAACA
GUAUGAACGCGCGGAAGGCCGCAUAGCACCGGCGGCAUGGAUGAACUGUAUAAAUAGUGAAAG
CUUAAUAAAAGAUCUUUAUGUUAUUAAGAUUCUGUGUGUUGGUUUUUGUGUGGCGGCCGCUAGGG
UAGAUAAUCUUUGUAUAGUGGGGGGGGAUCUCAUGUACCGGUUUUCUUUAUUAUUGAUUUUCAAU
AAAACAGACGGUAGCUAGGUUCGCAAGGCAGCCACAAGCCAAAGAUAGGUAGGGUGCUCAUAGU
GAGUAGGGACAGUGCCUUUUGAUUCACAACGCGUCAUACCAUCUGACACGGAUACCCUUAACG
GACUUGUCAUGAUCUCCAGACUUGUCCAAGGUGGACGGGCCACCUUUAACUUAACCCGGAAGG
GAACAUAUAUAAUUAUUAUGUGUUCGGAAAAUAGCAAAAAAAAAAAAAAAAAAAAAA

[1210]

533. TCARZ_SV40*_GFP_minPA_CMV*_mCherry_SV40LPA_GeFo3

ggguaaacGGCGGGAGUAACUAUGACUCUCUUAAGGAGUCAUAGAGCCAGAACCUCUCGUGGUCCC
GCUGGGCACAGGGAUUAUUUUUCUGUGGCAAAUUUGACUGGCUCUAGACUGUGGAAUGUGUGU
CAGUUAGGGUGUGGAAAGUCCCCAGGCUCCCCAGCAGGCAGAAGUAUGCAAAGCAUGCAUCUCA
AUUAGUCAGCAACCAGGUGUGGAAAGUCCCCAGGCUCCCCAGCAGGCAGAAGUAUGCAAAGCAU
GCAUCUCAUUAGUCAGCAACCAUAGUCCCGCCCCUACUCCGCCCAUCCCGCCCCUACUCCGCC

[1211]

CAGUUCGCCCCAUUCUCCGCCCCAUGGCUGACUAAUUGCGUCUAUAUAUGCAGAGGCCGAGGCCG
CCUCGGCCUCUGAGCUAUUCCAGAAGUAGUGAGGAGGCCGUACGGCCACCAUGGUGAGCAAGGG
CGAGGAGCUGUUCACCGGGGUGGUGCCCAUCCUGGUCGAGCUGGACGGCGACGUAAACGGCCACA
AGUUCAGCGUGUCCGGCGAGGGCGAGGGCGAUGCCACCUACGGCAAGCUGACCCUGAAGUUAUC
UGCACCACCGGCAAGCUGCCCCGUGCCCUGGCCCCACCCUCGUGACCACCCUGACCUACGGCGUGCA
GUGCUUCAGCCGCUACCCCGACCACAUGAAGCAGCACGACUUCUUAAGUCCGCCAUGCCCCGAAG
GCUACGUCCAGGAGCGCACCAUCUUCUUAAGGACGACGGCAACUACAAGACCCGCGCCGAGGUG
AAGUUCGAGGGCGACACCCUGGUGAACCGCAUCGAGCUGAAGGGCAUCGACUUAAGGAGGACG
GCAACAUCUGGGGCAACAAGCUGGAGUACAACUACAACAGCCACAACGUCUAUAUCAUGGCCGAC
AAGCAGAAGAACGGCAUCAAGGUGAACUUAAGAUCGCCACAACAUCGAGGACGGCAGCGUGC
AGCUCGCCGACCACUACCAGCAGAACACCCCCAUCGGCGACGGCCCCGUGCUGCUGCCCCGACAAC
CACUACCUGAGCACCCAGUCCGCCUGAGCAAAGACCCCAACGAGAAGCGCGAUCACAUGGUCCU
GCUGGAGUUCGUGACCGCCGCCGGGAUCACUCUCGGCAUGGACGAGCUGUACAAGUAAAAUAAA
AGAUCUUUAUGUUAUUAAGAUUCUGUGUGUUGGUUUUUGUGUGGCGUUACAUAACUUAACGGUAAA
UGGCCCCGCCUGGCUGACCGCCCAACGACCCCCGCCAUUGACGUCAAUAAUGACGUUUGUCCCCA
UAGUAAACGCCAAUAGGGACUUUCCAUAUGACGUCAAUGGGUGGAGUAAUUACGGUAAACUGCCCA
CUUGGCAGUACAUAAGUGUAUCAUAUGCCAAGUACGCCCCCUAUAUGACGUCAAUGACGGUAAA
UGGCCCCGCCUGGCAUUAUGCCCAGUACAUGACCUUAUGGGACUUUCCUACUUGGCAGUACAUCUA
CGUAUUAGUCAUCGCUAUUAACCAUGGUGAUGCGGUUUUGGCAGUACAUAUAUGGGCGUGGAUAG
GGGUUUGACUCACGGGGAUUCCAAGUCUCCACCCCAUUGACGUCAAUGGGAGUUUGUUUUGGC
ACCAAAAUCAACGGGACUUUCCAAAAUGUCGUAAACAACUCCGCCCAUUGACGCAAAUGGGCGGU
AGGCGUGUACGGUGGGAGGUCUAUAUAAGCAGAGCUCGUUUAGUGAACCBCUUGUCUCGCUCCG
GGGAACGCUCGGAAACUCCCGCCGCCGCCACCCGCGUCUGUUCUGUUAACAAAGGGAAGAAAAG
CCGUGCCGCACUCCGAGUGUCGUACGGCCACCAUGGUGAGCAAAGGCGAAGAAGAUACAUGGC
GAUUUUAAAGAAUUUAUGCGCUUUAAAGUGCAUAUGGAAGGCAGCGUGAACGGCCAUGAAUUU
GAAAUUGAAGGCGAAGGCGAAGGCCGCCCGUAUGAAGGCACCCAGACCGCGAAACUGAAAGUGA
CCAAAGGCGGCCCGCUGCCGUUUGCGUGGGGAUAUUCUGAGCCCGCAGUUUAUGUAUGGCAGCAA
AGCGUAUGUGAAACAUCGGCGGAUAUUCGGGAUUAUCUGAAACUGAGCUUUCGGAAGGCUUU
AAAUGGGAACGCGUGAUGAACUUUGAAGAUGGCGGCGUGGUGACCGUGACCCAGGAUAGCAGCC
UGCAGGAUGGCGAAUUUAUUUAUAAAGUGAAACUGCGCGGCACCAACUUCCGAGCGAUGGCC
GGUGAUGCAGAAAAAAACCAUGGGCUGGGAAGCGAGCAGCGAACGCAUGUAUCCGGAAGAUGGC
GCGCUGAAAGGCGAAAUAACAGCGCCUGAAACUGAAAGAUGGCGGCCAUUAUGAUGCGGAAG

UGAAAACCACCUAUAAGCGAAAAAACGGUGCAGCUGCCGGGCGCGUAUAACGUGAACAUUAA
ACUGGAUAUUACCAGCCAUAACGAAGAUUAUACCAUUGUGGAACAGUAUGAACGCGCGGAAGGC
CGCCAUAGCACCGGCGGCAUGGAUGAACUGUAUAAUAGUGAAAGCUAAUAAAAAGAUUUUAU
GUUCAUUAGAUCUGUGUGUUGGUUUUUGUGUGGCGGCCGCGGUAGAUAAUCUUUGUAUAGUGGG
GGGGGAUCUCAUGUACCGGGUUCUUUUUAUUUGAUUUUCAUAAAACAGACGGUAGCUAGGUUC
GCAAGGCAGCCACAAGCCAAAGAUAGGUAGGGUGCUCAUAGUGAGUAGGGACAGUGCCUUUGA
UUCACAACGCGUCAAUACCAUCUGACACGGAUACCCUUAACCGGACUUGUCAUGAUCUCCAGACU
UGUCCAAGGUGGACGGGCCACCUUACUUAACCCGGAAGGAACAUAUAUAAUUAUAGUGU
UCGGAAAAUAGCAAAAAAAAAAAAAAAAAAAAAA

534. TCARZ_CBhBsi_GFP_GeFo217_R4A22

[1212]

gggagacggucaaccgcuaggagcggugaccgGAAUUCGGCGGGAGUAACUAUGACUCUCUUAAGGAGUCAUAG
AGCCAGAACCUCUGUGGUCCCGCUGGGCACAGGGAUUAUUUUUCUGUGGCAAAUUGACUG
GCUCUAGAAGGUCGACCAGAUGUCCGAGGUCGACCAGUUGUCCGUGUGGAAUUGUGAGCGCUCA
CAAUUCCACACGUUACAUAACUUAACGGUAAAUGGCCCCGCCUGGCUGACCGCCCAACGACCCCCGC
CCAUUGACGUCAAUAGUAACGCCAAUAGGGACUUUCCAUUGACGUCAAUGGGUGGAGUAUUUAC
GGUAAACUGCCCACUUGGCAGUACAUAAGUGUAUCAUAUGCCAAGUACGCCCCUAUUGACGUC
AAUGACGGUAAAUGGCCCCGCCUGGCAUUGUGCCCAGUACAUGACCUUAUGGGACUUUCCUACUU
GGCAGUACAUCUACGUAAUAGUCAUCGCUAAUACCAUGGUCGAGGUGAGCCCCACGUUCUGCUU
CACUCUCCCCAUCUCCCCCCCCUCCCCACCCCCAAUUUUGUAUUUAUUUAUUUUUAAUUAUUU
GUGCAGCGAUGGGGGCGGGGGGGGGGGGGGGCGCGGCCAGGCGGGGCGGGGCGGGGCGAGGG
GCGGGGCGGGGCGAGGCGGAGAGGUGCGGCGGCAGCCAAUCAGAGCGGCGCGCUCCGAAAGUUU
CCUUUAUUGGCGAGGCGGCGGCGGCGGCCCUAUAAGCGAAGCGCGCGGCGGGCGGGAG
UCGUGCGACGUGCCUUCGCCCCGUGCCCCGUCGCGCGGCCUUCGCGCGCGCGCGCGCGGUC
UGACUGACCGGCUUACUCCACAGGUGAGCGGGCGGGACGGCCUUCUCCUCCGGGCUGUAAUUA
GCUGAGCAAGAGGUAAGGGUUUAAGGGAUGGUUGGUUGGUGGGUAUUAAGUUUAAUUAACCU
GGAGCACCUGCCUGAAAUCACUUUUUUCAGGUUGGCGUACGGCCACCAUGGUGAGCAAGGGCG
AGGAGCUGUUCACCGGGGUGGUGCCAUCCUGGUCGAGCUGGACGGCGACGUAAACGGCCACAA
GUUCAGCGUGUCCGGCGAGGGCGAGGGCGAUGCCACCUACGGCAAGCUGACCCUGAAGUUAUCU
GCACCACCGGCAAGCUGCCCCGUGCCCUGGCCCACCCUCGUGACCACCCUGACCUACGGCGUGCAG
UGCUUCAGCCGCUACCCCGACCACAUGAAGCAGCACGACUUCUUAAGUCCGCCAUGCCCGAAGG
CUACGUCCAGGAGCGCACCAUCUUCUUAAGGACGACGGCAACUACAAGACCCGCGCCGAGGUGA

AGUUCGAGGGCGACACCCUGGUGAACCGCAUCGAGCUGAAGGGCAUCGACUUAAGGAGGACGG
CAACAUCCUGGGGCACAAGCUGGAGUACAACUACAACAGCCACAACGUCUUAUAUCAUGGCCGACA
AGCAGAAGAACGGCAUCAAGGUGAACUUAAGAUCGCCACAACAUCGAGGACGGCAGCGUGCA
GCUCGCCGACCACUACCAGCAGAACACCCCCAUCGGCGACGGCCCCGUGCUGCUGCCCCGACAACC
ACUACCUGAGCACCCAGUCCGCCUGAGCAAAGACCCCAACGAGAAGCGCGAUCACAUGGUCCUG
CUGGAGUUCGUGACCGCCGCCGGGAUCACUCUCGGCAUGGACGAGCUGUACAAGUAAAAGCUUG
AUCCAGACAUGAUAAAGAUACAUUGAUGAGUUUGGACAAACCACAACUAGAAUGCAGUGAAAAAA
AUGCUUUUUUGUGAAAAUUUGUGAUGCUAUUGCUUUUUUGUAACCAUUUAAGCUGCAAUAAA
CAAGUUGUGGAAUUGUGAGCGCUCACAAUCCACAGCGCCGCUGAGACGGUAGCUAGGUUCGC
AAGGCAGCCACAAGCCAAAGAUAGGUAGGGUGCUCAUAGUGAGUAGGGACAGUGCCUUUGAUU
CACAACGCGUCAAUACCAUCUGACACGGAUACCCUUAACGGACUUGUCAUGAUCUCCAGACUUG
UCCAAGGUGGACGGGCCACCUUACUUAACCCGGAAGGAACAUAUUAUUAUUGUGUUC
GGAAAAUAGCAAAAAAAAAAAAAAAAAAAAAA

535. TCARZ_CBhBsi_GFP_GeFo98_R4A22

[1213]

gggagacgguaaccgcuaggagcggugaccgGAAUUCGGCGGGAGUAACUAGACUCUCUUAAGGAGUCAUAG
AGCCAGAACCUCUCGUGGUCCCGCUGGGCACAGGGAUUAUUUUUCUGUGGCAAAUUUGACUG
GCUCUAGAAGGUCGACCAGAUGUCCGAGGUCGACCAGUUGUCCGUGUGGAAUUGUGAGCGCUCA
CAAUCCACACGUUACAUAACUUAACGGUAAAUGGCCCCGCCUGGCUGACCGCCCAACGACCCCCGC
CCAUUGACGUCAAUAGUAACGCCAAUAGGGACUUAUCCAUGACGUCAAUGGGUGGAGUAUUUAC
GGUAAACUGCCCACUUGGCAGUACAUAAGUGUAUCAUAUGCCAAGUACGCCCCCUAUUGACGUC
AAUGACGGUAAAUGGCCCCGCCUGGCAUUGUGCCCAGUACAUGACCUUAUGGGACUUUCCUACUU
GGCAGUACAUCUACGUAUUAGUCAUCGCUAUUACCAUGGUCGAGGUGAGCCCCACGUUCUGCUU
CACUCUCCCCAUCUCCCCCCCCUCCCCACCCCCAAUUUUGUAUUUAUUUAUUUUUAAUUUUUU
GUGCAGCGAUGGGGGCGGGGGGGGGGGGGGGGGCGCGGCCAGGCGGGGCGGGGCGGGGCGAGGG
GCGGGGCGGGGCGAGGCGGAGAGGUGCGGCGGCAGCCAAUCAGAGCGGCGCGCUCCGAAAGUUU
CCUUUUUAUGGCGAGGCGGCGGCGGCGGCCCUAUA AAAAGCGAAGCGCGGCGGGCGGGAG
UCGUGCGACGCGUGCCUUCGCCCCGUGCCCCGCUCCGCCGCCGCCUCGCGCCGCCGCCGCCGCCGUC
UGACUGACCGCGUUAUCCACAGGUGAGCGGGCGGGACGGCCCUUCUCCUCCGGGCUGUAAUUA
GCUGAGCAAGAGGUAAAGGUUUAAAGGAUGGUUGGUUGGUGGGUAUUAAGUUUAUUUACCU
GGAGCACCUGCCUGAAAUCACUUUUUUUCAGGUUGGCGUACGGCCACCAUGGUGAGCAAGGGCG
AGGAGCUGUUCACGGGGUGGUGCCCAUCCUGGUCGAGCUGGACGGCGACGUAAACGGCCACAA

GUUCAGCGUGUCCGGCGAGGGCGAGGGCGAUGCCACCUACGGCAAGCUGACCCUGAAGUUCAUCU
GCACCACCGGCAAGCUGCCCCGUGCCCGUGGGCCACCCUCGUGACCACCCUGACCUACGGCGUGCAG
UGCUUCAGCCGCUACCCCGACCACAUGAAGCAGCACGACUUCUUAAGUCCGCCAUGCCCGAAGG
CUACGUCCAGGAGCGCACCAUCUUCUUAAGGACGACGGCAACUACAAGACCCGCGCCGAGGUGA
AGUUCGAGGGCGACACCCUGGUGAACCGCAUCGAGCUGAAGGGCAUCGACUUAAGGAGGACGG
CAACAUCCUGGGGCACAAGCUGGAGUACAACUACAACAGCCACAACGUCUAUAUCAUGGCCGACA
AGCAGAAGAACGGCAUCAAGGUGAACUUAAGAUCGCCACAACAUCGAGGACGGCAGCGUGCA
GCUCGCCGACCACUACCAGCAGAACACCCCCAUCGGCGACGGCCCCGUGCUGCUGCCCCGACAACC
ACUACCUGAGCACCCAGUCCGCCCGUGAGCAAAGACCCCAACGAGAAGCGCGAUCACAUGGUCCUG
CUGGAGUUCGUGACCGCCCGCGGGAUCACUCUCGGCAUGGACGAGCUGUACAAGUAAAAGCUUG
AUCCAGACAUGAUAAAGAUACAUAUGAUGAGUUUGGACAAACCACAACUAGAAUGCAGUGAAAAAA
AUGCUUUUUUUGUGAAAAUUUGUGAUGCUAUUGCUUUUUUUGUAACCAUUUAAGCUGCAAUAAA
CAAGUUGUGGAAUUGUGAGCGCUCACAAUUCACAGCGGCCGUGACCGGACUUGUCAUGAUCU
CCCAGACUUGUCCAAGGUGGACGGGCCACCUUUACUUAACCCGGAAGGAACAUUAUUAAUU
AUAUGUGUUCGGAAAAUAGCAAAAAAAAAAAAAAAAAAAAAA

[1214]

536. TCARZ_CBhBsi_GFP_GeFo68_R4A22

gggagacggucaaccgcuaggagcgguagacgGAAUUCGGCGGGAGUAAACUAUGACUCUCUUAAGGAGUCAUAG
AGCCAGAACCUCUCGUGGUCCCGCUGGGCACAGGGAUUAUUUUUCUGUGGCAAAUUUGACUG
GCUCUAGAAGGUCGACCAGAUGUCCGAGGUCGACCAGUUGUCCGUGUGGAAUUGUGAGCGCUCA
CAAUUCACACGUUACAUAACUUAACGGUAAAUGGCCCGCCUGGCUGACCGCCCAACGACCCCGC
CCAUUGACGUCAAUAGUAACGCCAAUAGGGACUUUCCAUUGACGUCAAUGGGUGGAGUAUUUAC
GGUAAACUGCCCACUUGGCAGUACAUAAGUGUAUCAUAUGCCAAGUACGCCCCUAUUGACGUC
AAUGACGGUAAAUGGCCCGCCUGGCAUUGUGCCCAGUACAUGACCUUAUGGGACUUUCCUACUU
GGCAGUACAUCUACGUUUAGUCAUCGCUAUUACCAUGGUCGAGGUGAGCCCCACGUUCUGCUU
CACUCUCCCCAUCUCCCCCCCCUCCCCACCCCAAUUUUGUAUUUAUUUAUUUUUAAUUUUUU
GUGCAGCGAUGGGGGCGGGGGGGGGGGGGGGGGCGCGGCCAGGCGGGGCGGGGCGGGGCGAGGG
GCGGGGCGGGGCGAGGCGGAGAGGUGCGGCGGCAGCCAAUCAGAGCGGCGCGCUCCGAAAGUUU
CCUUUUUUGGCGAGGCGGCGGCGGCGGCGGCCCUAUA AAAAGCGAAGCGCGCGGCGGGCGGGAG
UCGCUGCGACGCUGCCUUCGCCCCGUGCCCCGCUCCGCCGCCGCCUCGCGCCGCCGCCCGGCUC
UGACUGACCGGCUUACUCCACAGGUGAGCGGGCGGGACGGCCCUUCUCCUCCGGGCUGUAAUUA
GCUGAGCAAGAGGUAAGGGUUUAAGGGAUGGUUGGUUGGUGGGUAUUAAUGUUUAAUUACCU

GGAGCACCUGCCUGAAAUCACUUUUUUUCAGGUUGGCGUACGGCCACCAUGGUGAGCAAGGGCG
AGGAGCUGUUCACCGGGGUGGUGCCCAUCCUGGUCGAGCUGGACGGCGACGUAAACGGCCACAA
GUUCAGCGUGUCCGGCGAGGGCGAGGGCGAUGCCACCUACGGCAAGCUGACCCUGAAGUUCAUCU
GCACCACCGGCAAGCUGCCCCGUGCCCUGGGCCACCCUCGUGACCACCCUGACCUACGGCGUGCAG
UGCUUCAGCCGCUACCCCGACCACAUGAAGCAGCAGACUUCUUAAGUCCGCCAUGCCCGAAGG
CUACGUCCAGGAGCGCACCAUCUUCUUAAGGACGACGGCAACUACAAGACCCGCGCCGAGGUGA
AGUUCGAGGGGCGACACCCUGGUGAACCGCAUCGAGCUGAAGGGCAUCGACUUAAGGAGGACGG
CAACAUCCUGGGGCACAAGCUGGAGUACAACUACAACAGCCACAACGUCUAUAUCAUGGCCGACA
AGCAGAAGAACGGCAUCAAGGUGAACUUAAGAUCGCCACAACAUCGAGGACGGCAGCGUGCA
GCUCGCCGACCACUACCAGCAGAACACCCCCAUCGGCGACGGCCCCGUGCUGCUGCCCGACAACC
ACUACCUGAGCACCCAGUCCGCCUGAGCAAAGACCCCAACGAGAAGCGCGAUCACAUGGUCCUG
CUGGAGUUCGUGACCGCCGCCGGGAUCACUCUCGGCAUGGACGAGCUGUACAAGUAAAAGCUUG
AUCCAGACAUGAUAAAGAUACAUUGAUGAGUUUGGACAAACCACAACUAGAAUGCAGUGAAAAAA
AUGCUUUAUUUGUGAAAUUUGUGAUGCUAUUGCUUUAUUUGUAACCAUUAUAAGCUGCAAUAAA
CAAGUUGUGGAAUUGUGAGCGCUCACAAUCCACAGCGGCCGCUGACAAGGUGGACGGGCCACC
[1215] UUUACUUAACCCGGAAGGAACAUAUAUUAUUAUGUGUUCGAAAAUAGCAAAAAAAAAA
AAAAAAAAAAAA

537. AGGTCGACCAGATGTCCGAGGTCGACCAGTTGTCCG

538. TAGCaAAAAAAAAAAAAAAAAAAAA

539. CATCATGGATTAGGATCGGAAGACCCCCG

540. GTACGCCGGCGAAATTGGATCAGTAGATG

541. GAGAAACAGACGGGCCTGATCTACACC)

542. CTATCTGAACCGAAGTTCCGCAACGCCTACGTAC

543. CACTGCGTGTGGTCAGTTTTTCCTAGCATGCACG

544. GATGTTATGCCAAGACAGCAAGCAAATGTTTTGAACCAAACG

545. TTGAGGCGAGTCACCACTCGCTTTCCGG

546. GTGTCCGTCACGGGGACGACATCCGAGTG

547. CTCCTGACCAACTAGCTCACTGACTAATTTTAAAC

548. CCACTTATTCTACACCTCTCATGTCTCTTCACCG

549. CTAGCAGCCGACTTAGAACTGGTGCGG

550. CTTCGTCTTCGGAATCCATGTCCATAGC

551. CCAGGGGAATCCGACTGTTTAATTAAAACAAAGC

[1216]

552. GCGACTCGCATCACTGACTTTAATTGGTTG

553. CGCGCUUCGUUUUCAGAA_{tt}

554. UUCUGAAAUACGAAGCGCG_{tg}

555. UGACCUCUCCAAACCCUCU_{tt}

556. AGAGGGUUUGGAGAGGUC_{atg}

557. GGGAGCACCUGAAUCCUA_{tt}

558. UUAGGAUUCAGGUGCUC_{CCCgg}

559. GGAUAUUACUUUCGUGUA_{tt}

560. UUACACGAAAGUAAUAUCCaa

561. CGUCGAUUCCCAGAUCUUA_{tt}

[1217] 562. UAAGAUCUGGGAAUCGACGaa

563. GAAUCGCAAGGAUAUGAU_{tt}

564. UAUCAUAUCCUUGCGAUUC_{tc}

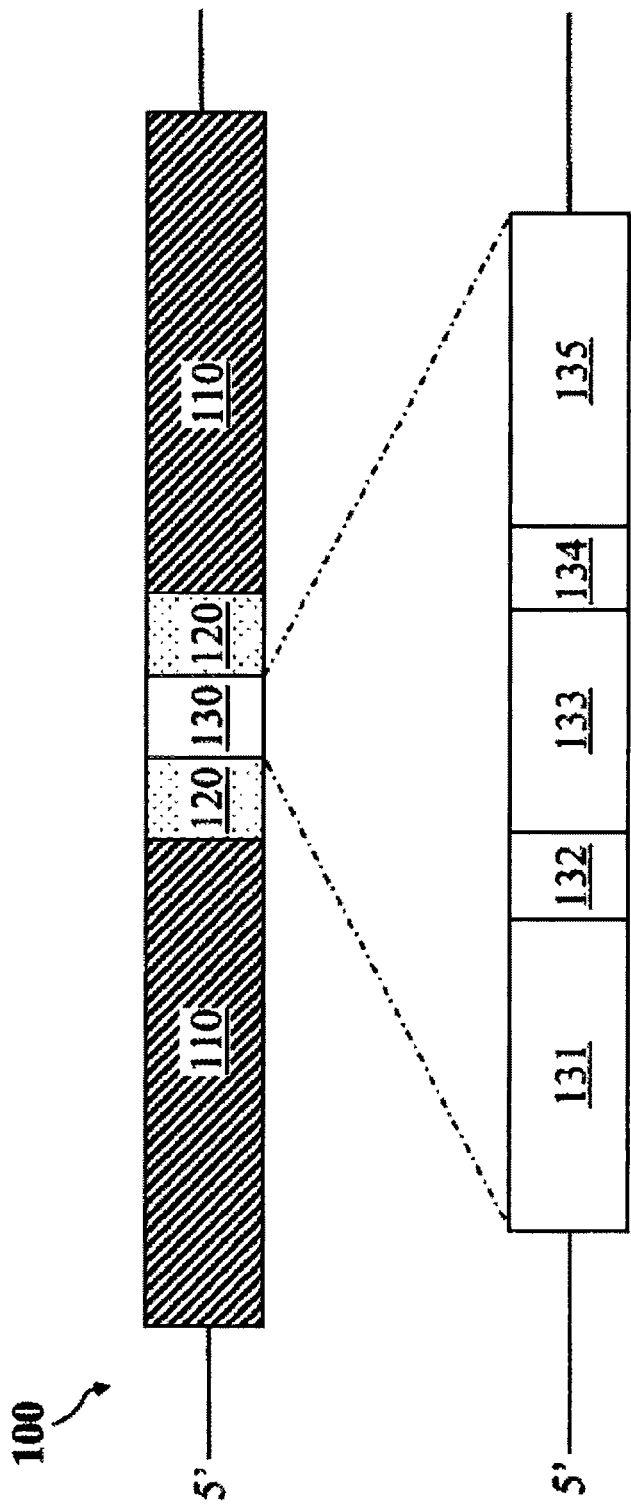


图1

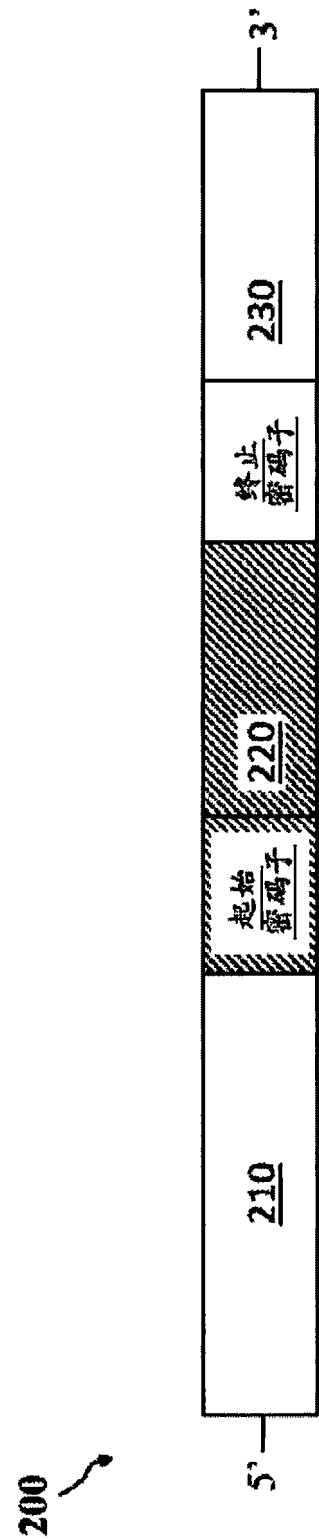


图2

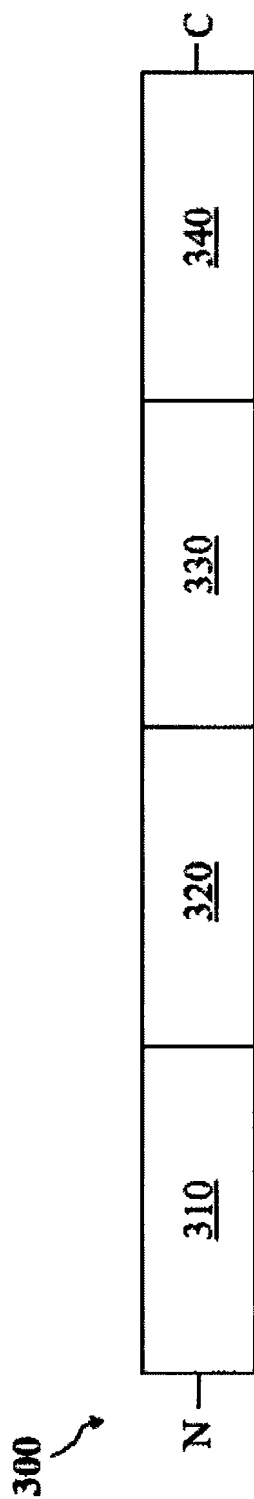
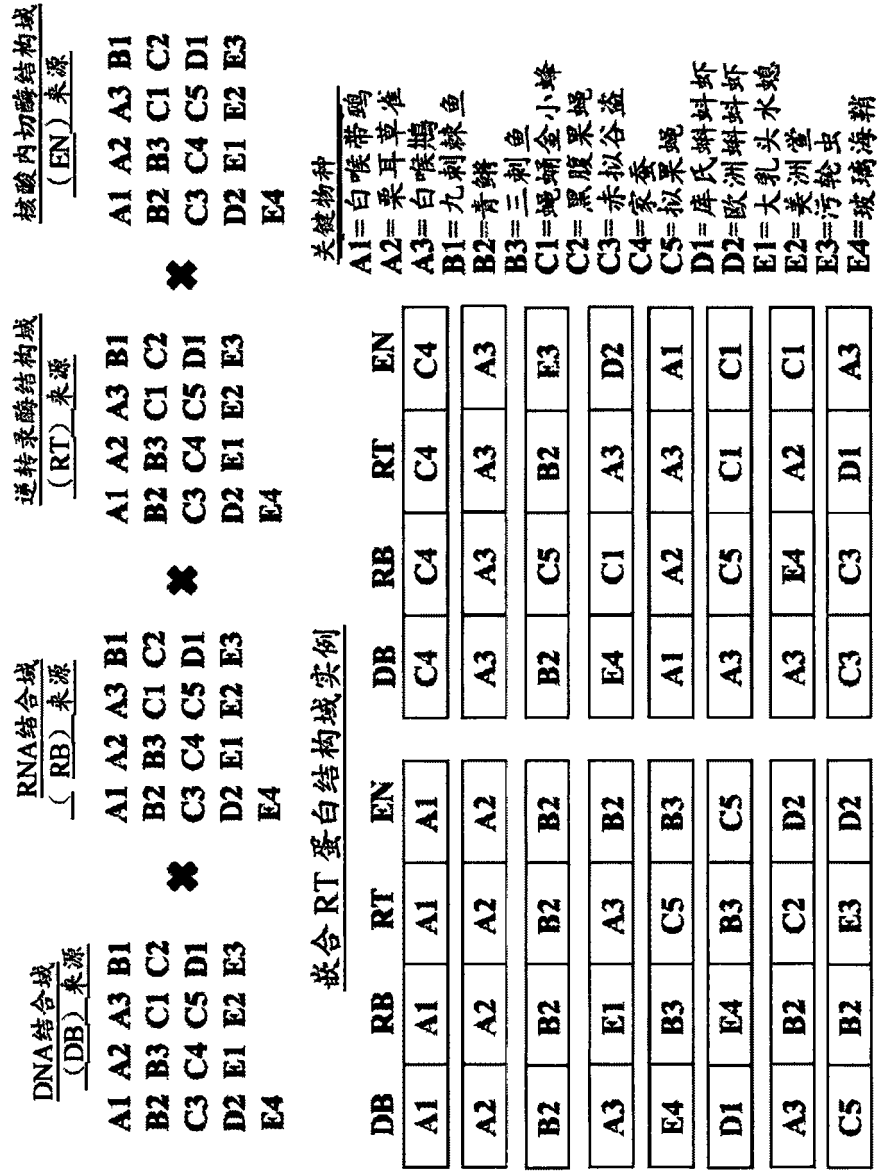


图3



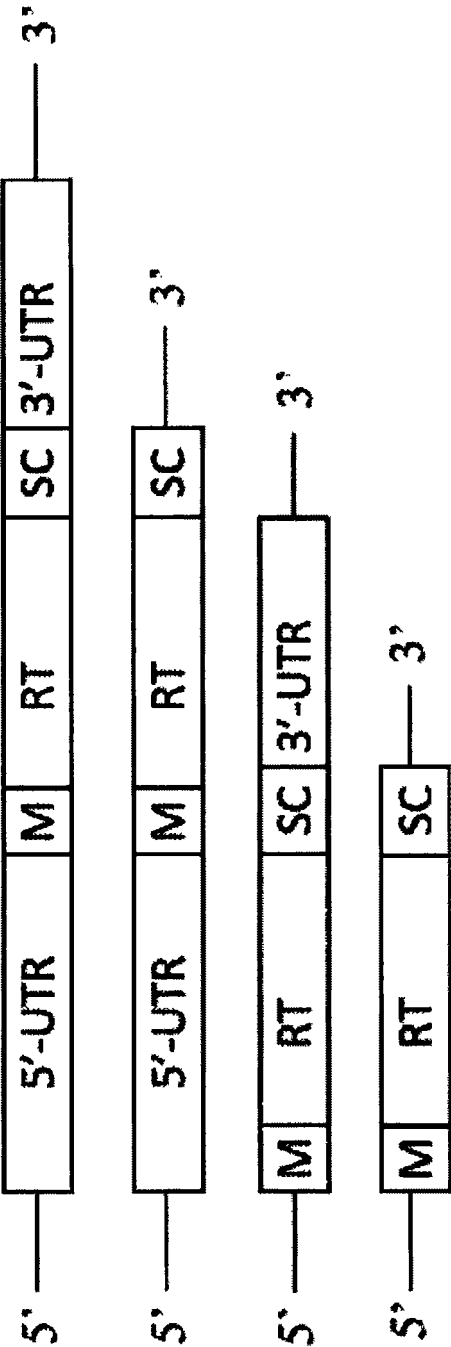


图5

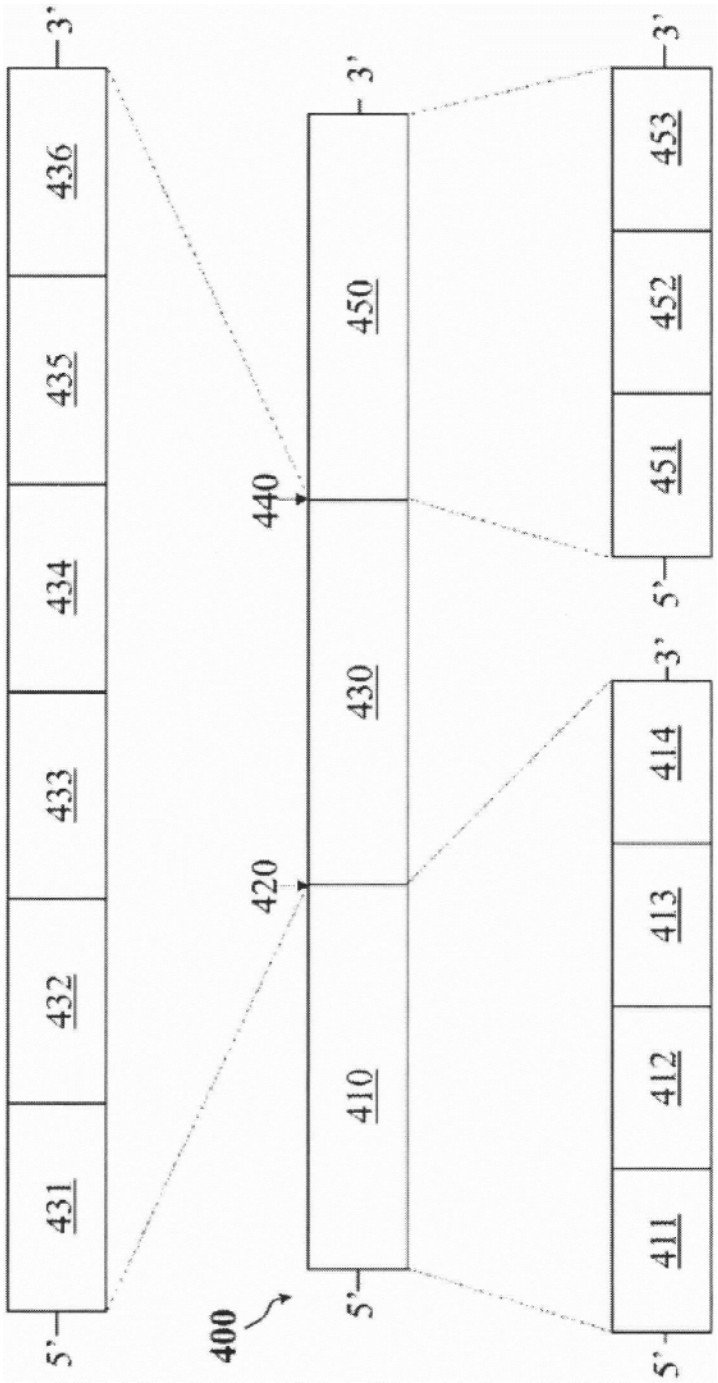


图6

5' 模块 (5'M) 来源				3' 模块 (3'M) 来源				RT 模块 (RT) 来源			
A1	A2	A3	B1	A1	A2	A3	B1	A1	A2	A3	B1
B2	B3	C1	C2	B2	B3	C1	C2	B2	B3	C1	C2
C3	C4	C5	D1	C3	C4	C5	D1	C3	C4	C5	D1
D2	E1	E2	E3	D2	E1	E2	E3	D2	E1	E2	E3
E4				E4				E4			

基因插入构建体				配对RT	关键物种
5'-A1 5' M	有效载荷	A1 3' M	3'	A1 RT	A1=白喉带鹀
5'-C4 5' M	有效载荷	C4 3' M	3'	C4 RT	A2=栗耳草雀
5'-B3 5' M	有效载荷	A2 3' M	3'	E4 RT	A3=白喉鵂
5'-A3 5' M	有效载荷	D2 3' M	3'	C3 RT	B1=九刺棘鱼
5'-D1 5' M	有效载荷	E4 3' M	3'	D2 RT	B2=青鳉
5'-E4 5' M	有效载荷	C4 3' M	3'	E1 RT	B3=三刺鱼
5'-A3 5' M	有效载荷	A1 3' M	3'	A1 RT	C1=蝇蛹金小蜂
5'-A1 5' M	有效载荷	A2 3' M	3'	A2 RT	C2=黑腹果蝇
5'-B2 5' M	有效载荷	A1 3' M	3'	A3 RT	C3=赤拟谷盗
5'-B3 5' M	有效载荷	E3 3' M	3'	B1 RT	C4=家蚕
5'-C5 5' M	有效载荷	B1 3' M	3'	D1 RT	C5=拟果蝇
5'-C3 5' M	有效载荷	E3 3' M	3'	E3 RT	D1=库氏蝌蚪虾
5'-E2 5' M	有效载荷	C2 3' M	3'	C1 RT	D2=欧洲蝌蚪虾
5'-C2 5' M	有效载荷	E4 3' M	3'	C4 RT	E1=大乳头水螅
5'-A2 5' M	有效载荷	A1 3' M	3'	A1 RT	E2=美洲鲎
5'-C3 5' M	有效载荷	A2 3' M	3'	A2 RT	E3=污轮虫
					E4=玻璃海鞘

图7

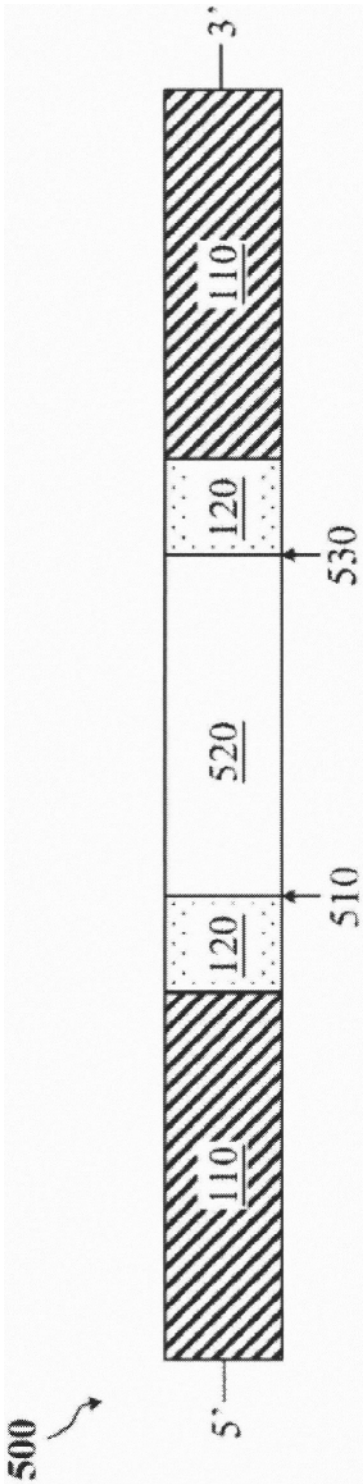


图8

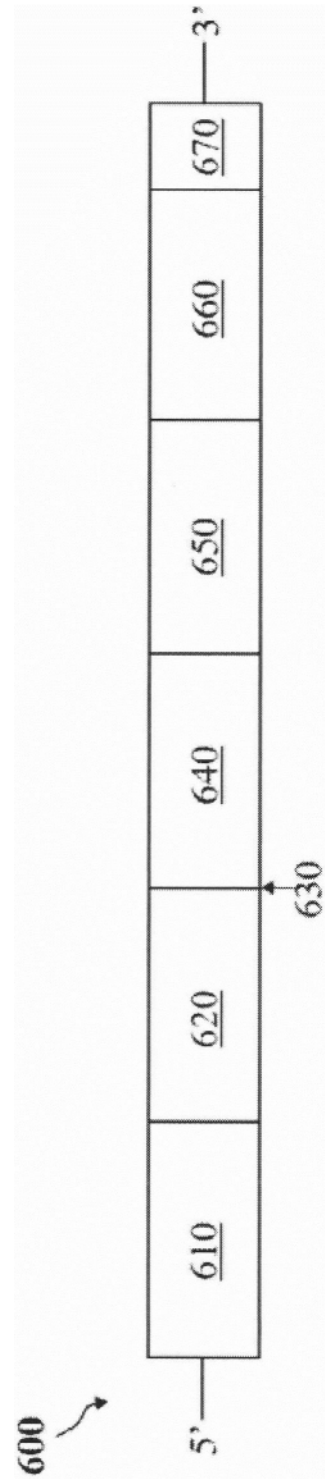


图9

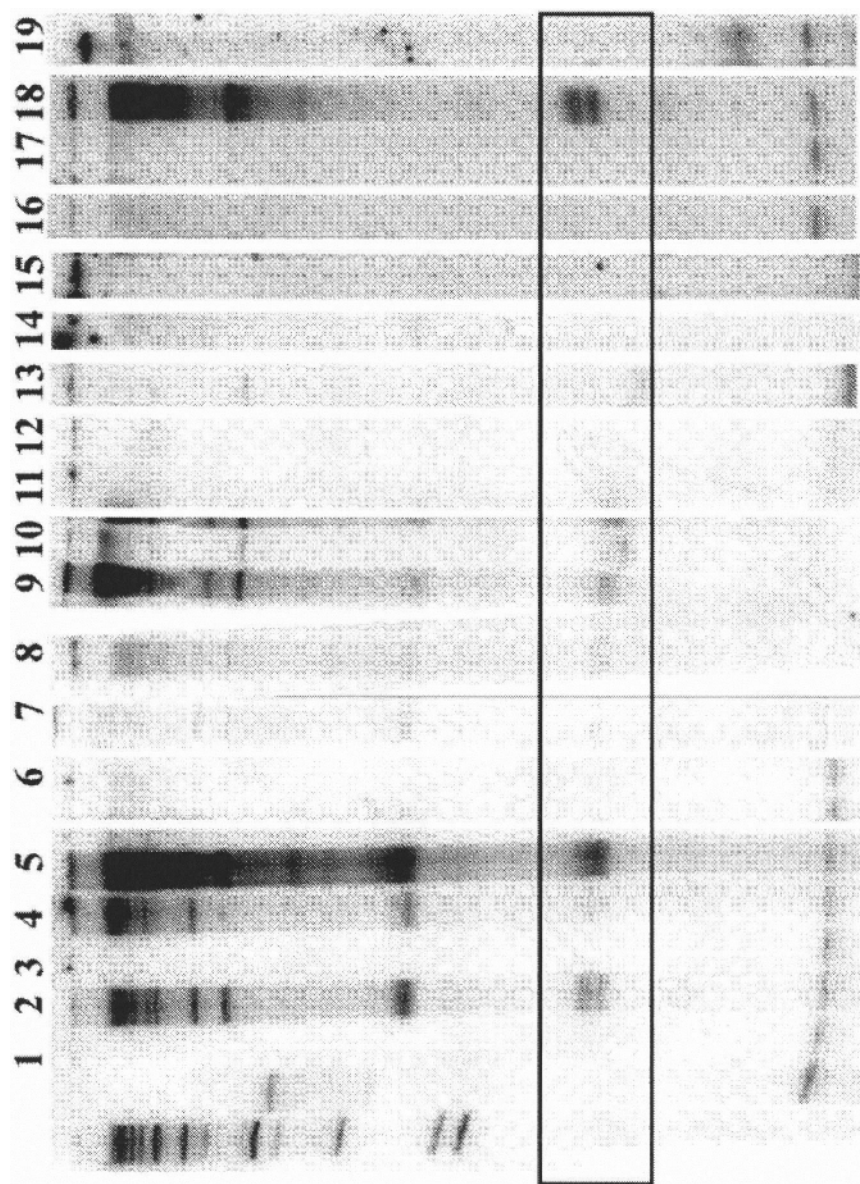


图10

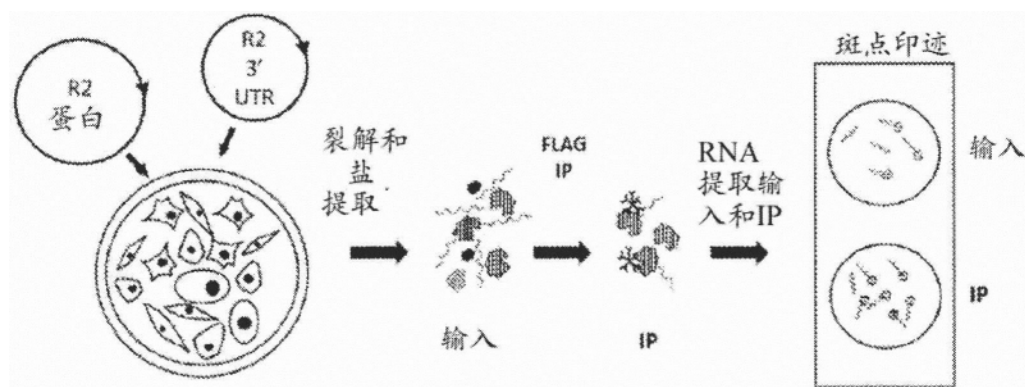


图11A

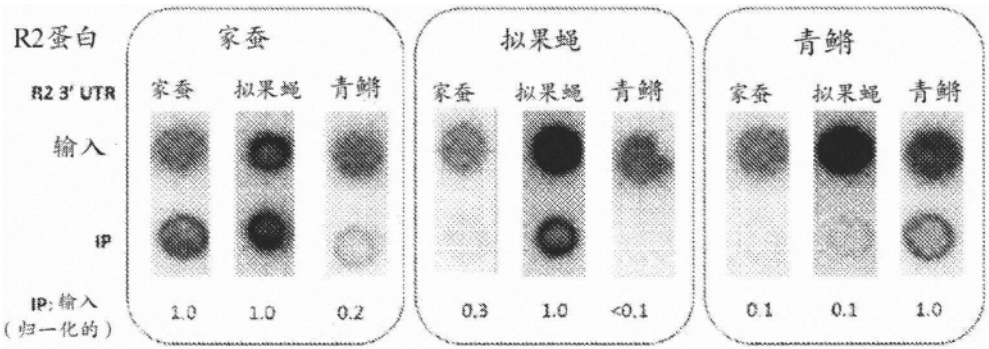


图11B

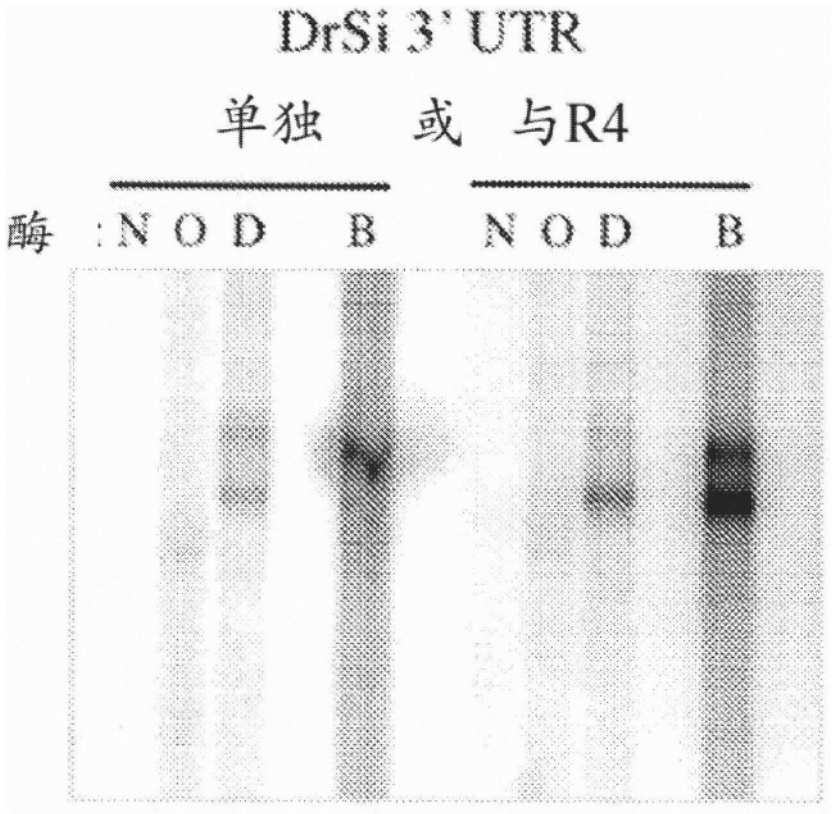


图12A

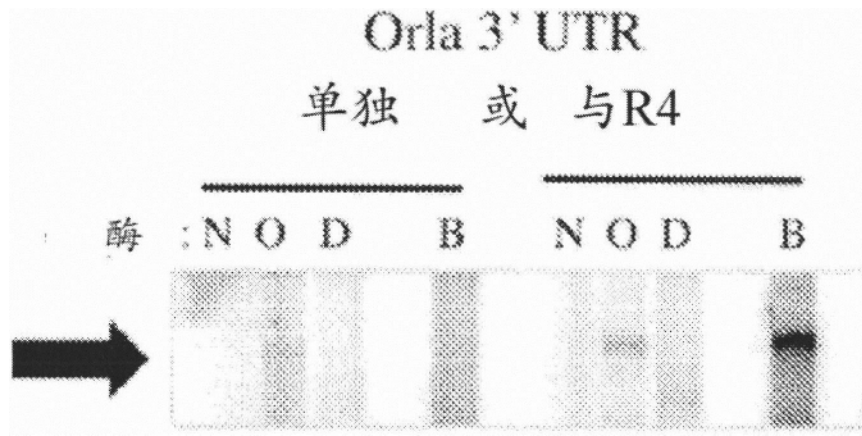


图12B

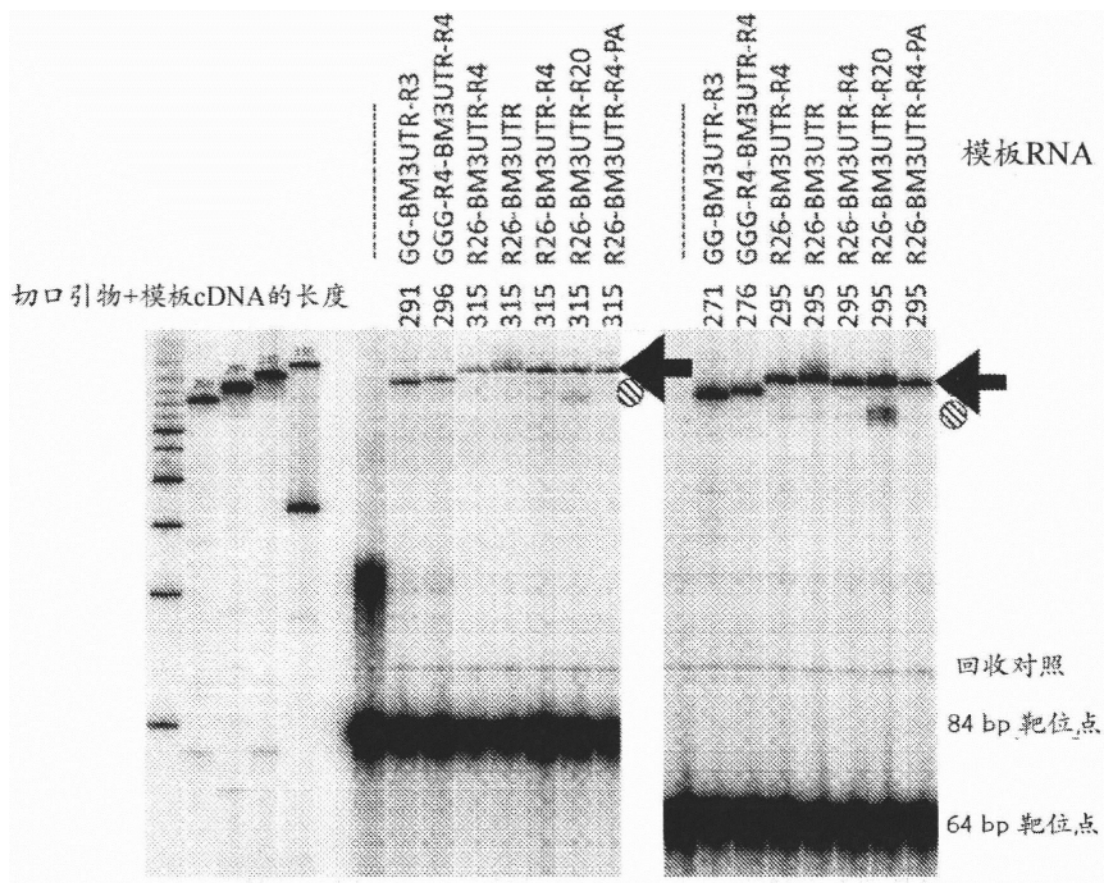


图13

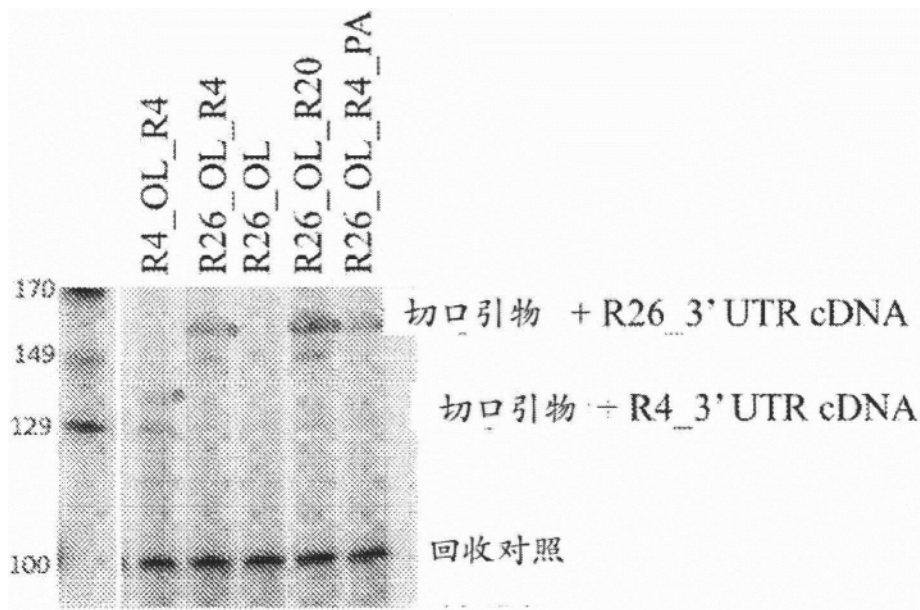


图14A

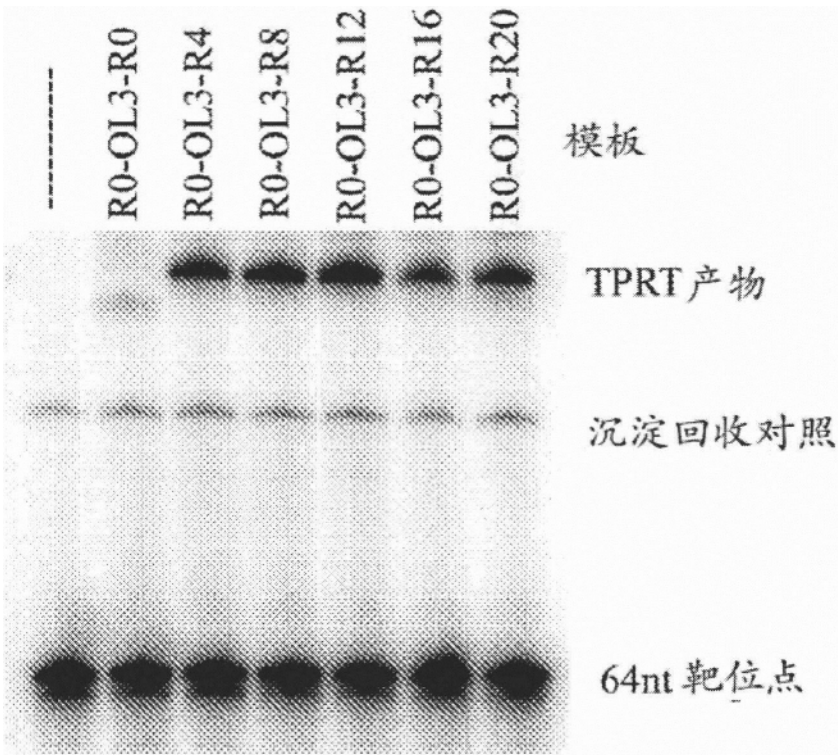


图14B

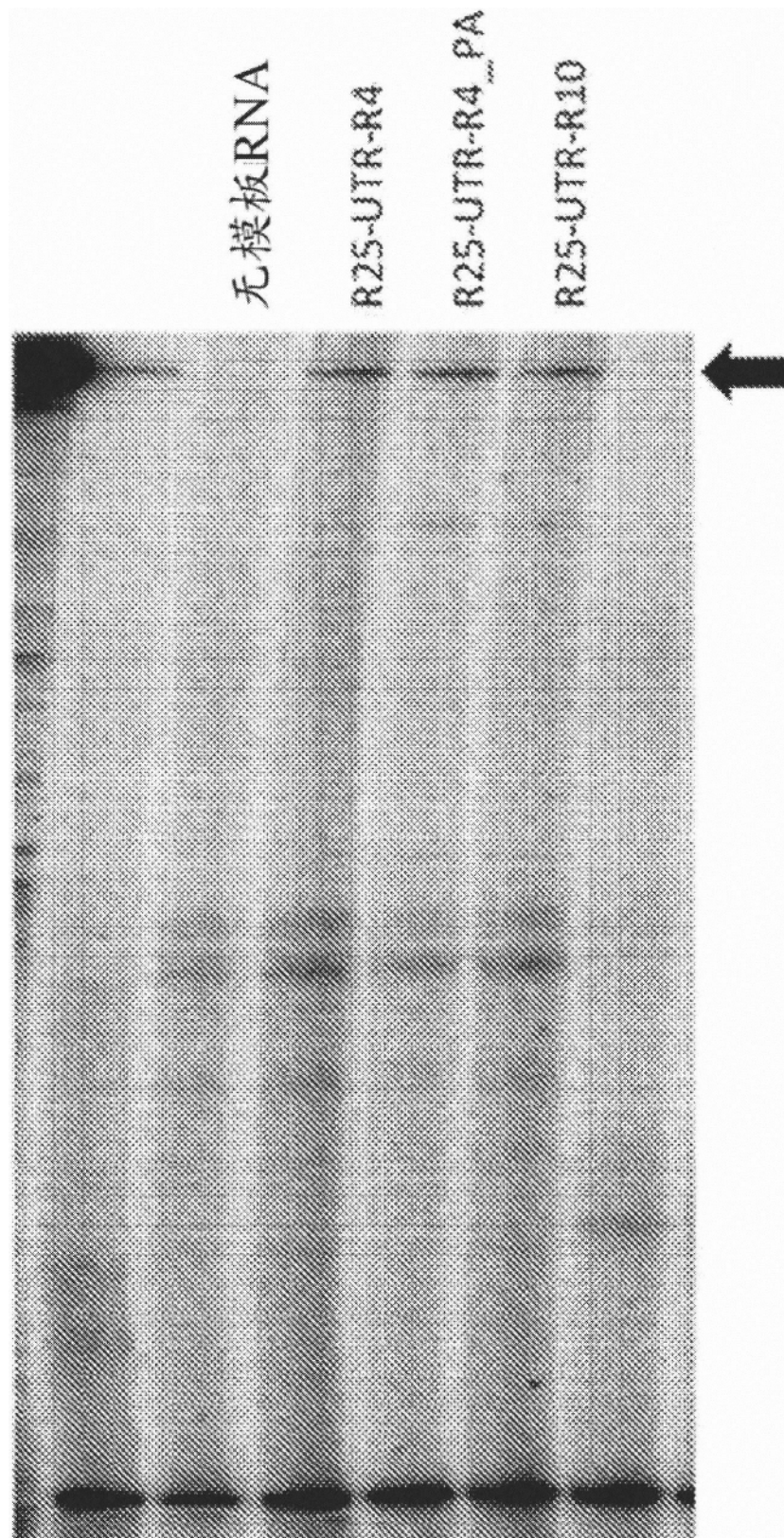


图15

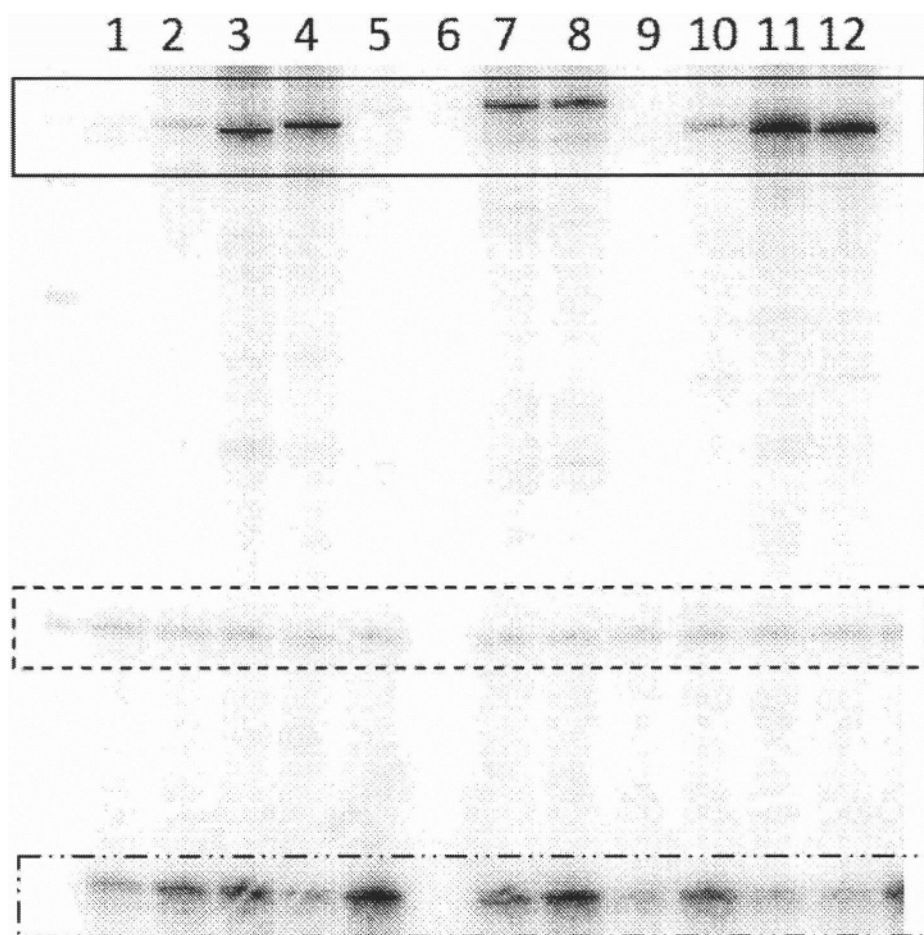


图16

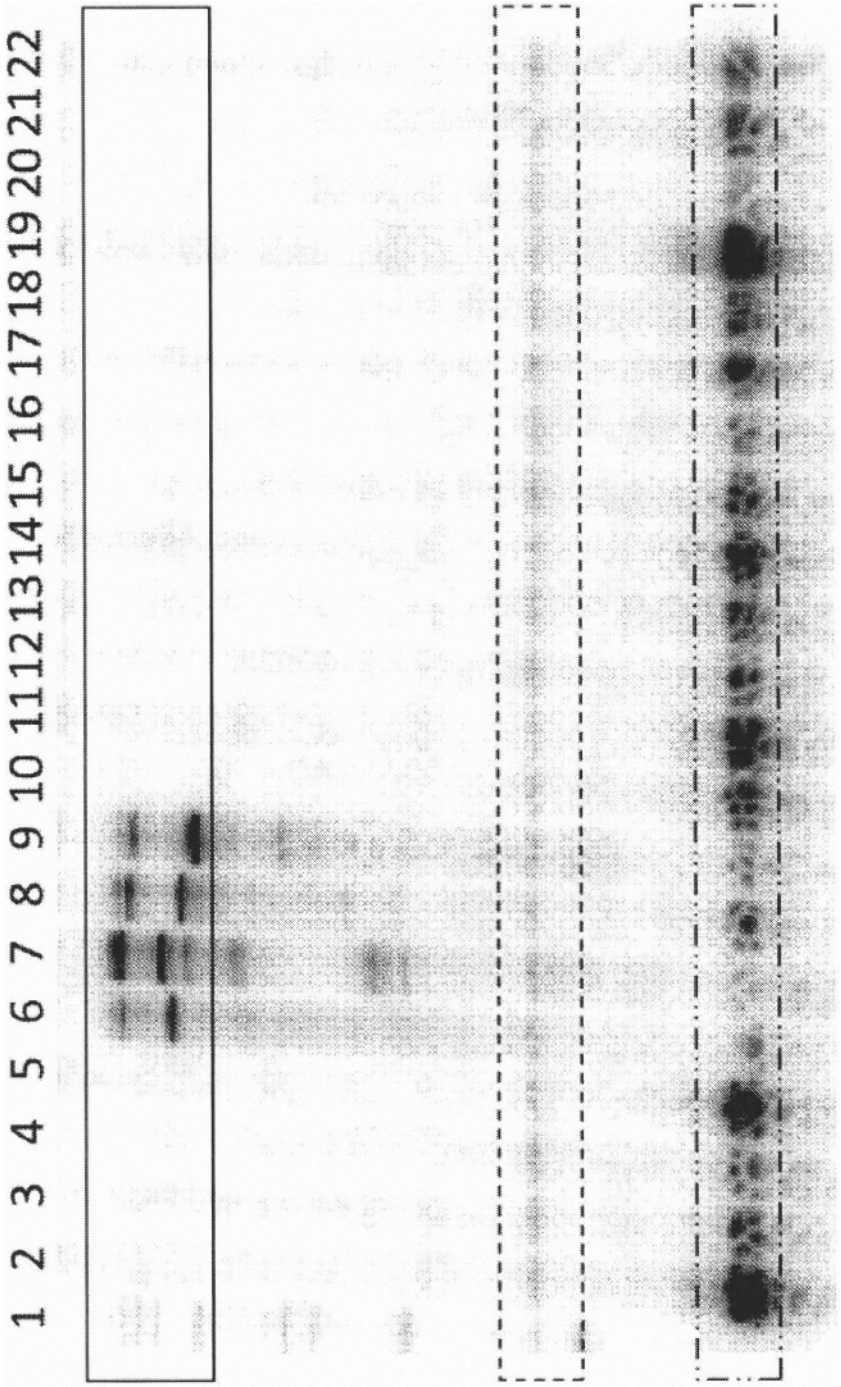


图17

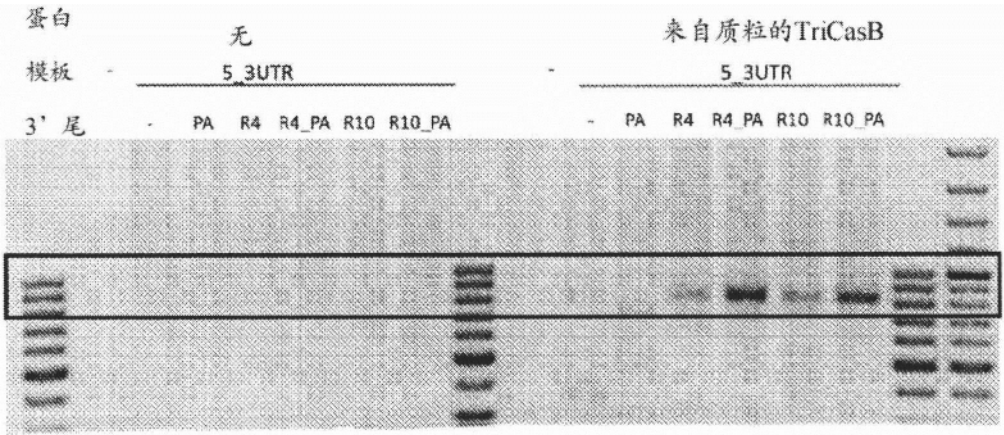


图18A

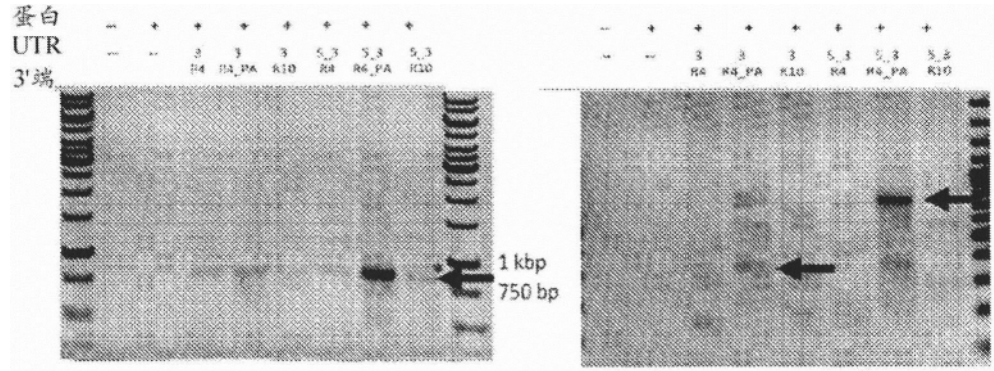


图18B

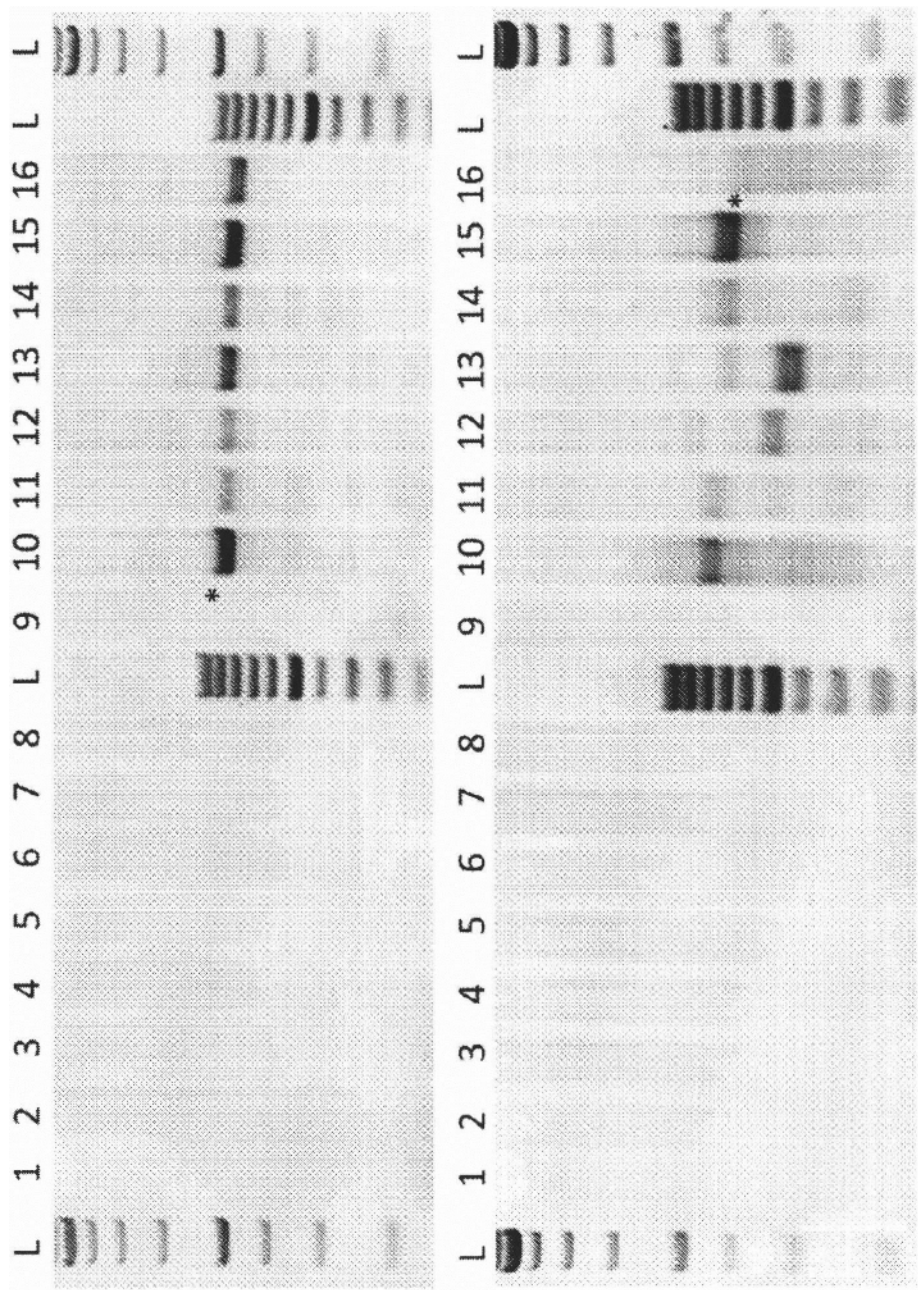


图19

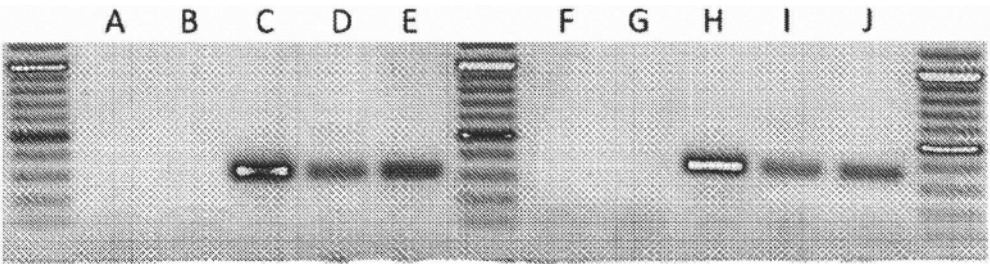


图20

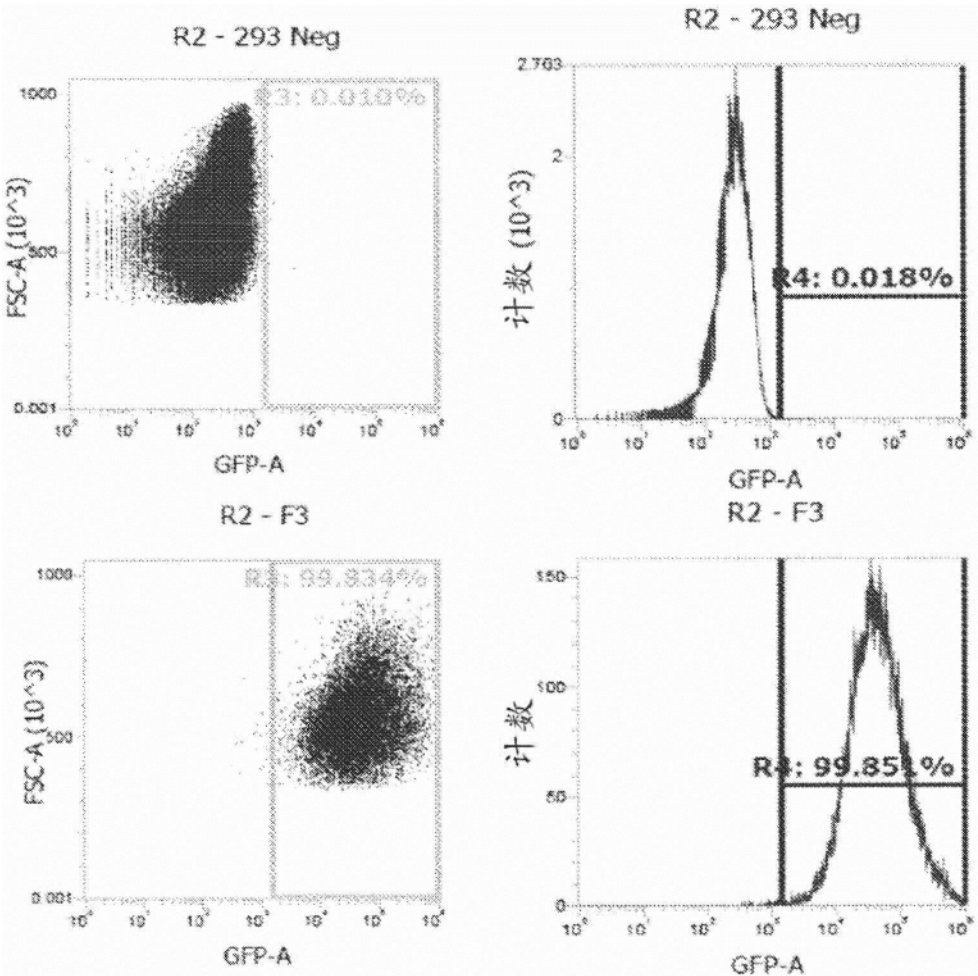


图21