(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 116096866 A
(43)申请公布日 2023.05.09

(21)申请号 202180038773.7

(22)申请日 2021.03.30

(30)优先权数据
63/003,168 2020.03.31 US
63/154,341 2021.02.26 US

(85)PCT国际申请进入国家阶段日
2022.11.28

(86)PCT国际申请的申请数据
PCT/US2021/024993 2021.03.30

(87)PCT国际申请的公布数据
WO2021/202604 EN 2021.10.07

(71)申请人 萨那生物科技公司
地址 美国华盛顿
申请人 旗舰创业创新五公司

(72)发明人 K·M·特鲁多 C·班多罗
L·P·马肯锡 J·V·薛华
G·A·冯-马尔扎恩
J·R·鲁本斯 M·T·米

(74)专利代理机构 北京坤瑞律师事务所 11494
专利代理师 封新琴

(51)Int.Cl.
C12N 7/00(2006.01)

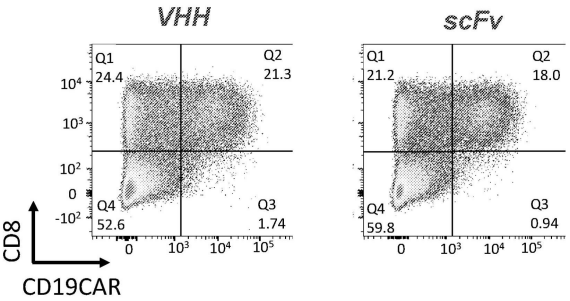
权利要求书14页 说明书257页
序列表（电子公布） 附图6页

(54)发明名称

靶向脂质颗粒及其组合物和用途

(57)摘要

本文提供了脂质颗粒，其包含围绕内腔或空腔的脂质双层、亨尼帕病毒F蛋白分子或其生物活性部分和靶向包膜蛋白，所述靶向包膜蛋白包含亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分和结合结构域诸如单结构域抗体(sdAb)可变域。本文还提供了包含与结合结构域诸如sdAb可变域融合或连接的G蛋白的靶向包膜蛋白以及编码此类蛋白质的多核苷酸。还提供了含有此类靶向脂质颗粒的生产者细胞和组合物以及制备和使用所述靶向脂质颗粒的方法。



CN 116096866 A

1.一种靶向脂质颗粒,其包含:

(a)包围内腔的脂质双层,

(b)亨尼帕病毒F蛋白分子或其生物活性部分;以及

(c)靶向包膜蛋白,所述靶向包膜蛋白包含(i)亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分和(ii)单结构域抗体(sdAb)可变域,其中所述sdAb可变域与所述G蛋白或其生物活性部分的C末端连接,和/或其中所述sdAb经由肽接头与所述G蛋白或其生物活性部分连接,其中所述sdAb结合靶细胞的细胞表面分子,

其中所述F蛋白分子或其生物活性部分和所述靶向包膜蛋白包埋在所述脂质双层中。

2.如权利要求1所述的靶向脂质颗粒,其中所述细胞表面分子是蛋白质、聚糖、脂质或低分子量分子。

3.如权利要求1或权利要求2所述的靶向脂质颗粒,其中所述靶细胞选自由以下组成的组:肿瘤浸润淋巴细胞、T细胞、赘生性或肿瘤细胞、病毒感染细胞、干细胞、中枢神经系统(CNS)细胞、造血干细胞(HSC)、肝细胞或完全分化细胞。

4.如权利要求1-3中任一项所述的靶向脂质颗粒,其中所述靶细胞选自由以下组成的组:CD3+T细胞、CD4+T细胞、CD8+T细胞、肝细胞、造血干细胞、CD34+造血干细胞、CD105+造血干细胞、CD117+造血干细胞、CD105+内皮细胞、B细胞、CD20+B细胞、CD19+B细胞、癌细胞、CD133+癌细胞、EpCAM+癌细胞、CD19+癌细胞、Her2/Neu+癌细胞、GluA2+神经元、GluA4+神经元、NKG2D+自然杀伤细胞、SLC1A3+星形胶质细胞、SLC7A10+脂肪细胞或CD30+肺上皮细胞。

5.如权利要求1-4中任一项所述的靶向脂质颗粒,其中所述单结构域抗体结合存在于肝细胞上的抗原或其部分。

6.如权利要求1-5中任一项所述的靶向脂质颗粒,其中所述细胞表面分子或抗原选自由以下组成的组:ASGR1、ASGR2和TM4SF。

7.如权利要求1-4中任一项所述的靶向脂质颗粒,其中所述单结构域抗体结合存在于T细胞上的抗原或其部分。

8.如权利要求1-4和7中任一项所述的靶向脂质颗粒,其中所述细胞表面分子或抗原是CD8或CD4。

9.如权利要求1-4中任一项所述的靶向脂质颗粒,其中所述细胞表面分子或抗原是LDL-R。

10.一种靶向脂质颗粒,其包含:

(a)包围内腔的脂质双层,

(b)亨尼帕病毒F蛋白分子或其生物活性部分;以及

(c)靶向包膜蛋白,所述靶向包膜蛋白包含(i)亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分和(ii)结合结构域,其中所述结合结构域与所述G蛋白或其生物活性部分的C末端连接,并且其中所述结合结构域结合选自由以下组成的组的细胞表面分子:ASGR1、ASGR2、TM4SF5、CD8、CD4和LDL-R,

其中所述F蛋白分子或其生物活性部分和所述靶向包膜蛋白包埋在所述脂质双层中。

11.如权利要求10所述的靶向脂质颗粒,其中所述结合结构域通过接头与所述G蛋白连接。

12.如权利要求11所述的靶向脂质颗粒,其中所述接头是肽接头。

13.如权利要求1-12中任一项所述的靶向脂质颗粒,其中所述脂质颗粒是慢病毒载体。

14.一种慢病毒载体,其包含:

(a)亨尼帕病毒F蛋白分子或其生物活性部分;和

(b)靶向包膜蛋白,所述靶向包膜蛋白包含(i)亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分和(ii)结合结构域,其中所述结合结构域与所述G蛋白或其生物活性部分的C末端连接,并且其中所述结合结构域结合CD4;以及

(c)包含编码嵌合抗原受体(CAR)的核酸的运载物,其中所述CAR包含(i)结合CD19的细胞外抗原结合结构域,任选地其中所述细胞外抗原结合结构域是scFv,(ii)跨膜结构域和(iii)包含CD3ζ信号传导结构域的细胞内信号传导区。

15.如权利要求14所述的慢病毒载体,其中所述CAR的细胞内信号传导区还包含4-1BB共刺激信号传导结构域。

16.如权利要求14或权利要求15所述的慢病毒载体,其中所述慢病毒载体能够将编码所述CAR的核酸递送至T细胞,任选地其中所述T细胞在受试者体内。

17.一种慢病毒载体,其包含:

(a)亨尼帕病毒F蛋白分子或其生物活性部分;和

(b)靶向包膜蛋白,所述靶向包膜蛋白包含(i)亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分和(ii)结合结构域,其中所述结合结构域与所述G蛋白或其生物活性部分的C末端连接,并且其中所述结合结构域结合选自由以下组成的组的细胞表面分子:ASGR1、ASGR2和TM4SF5。

18.如权利要求17所述的慢病毒载体,其中所述慢病毒载体能够靶向肝细胞。

19.如权利要求17和权利要求18所述的慢病毒载体,其还包含用于递送至肝细胞的外源药剂和/或其中所述慢病毒载体能够将所述外源药剂递送至肝细胞,任选地其中所述肝细胞在受试者体内。

20.如权利要求14-19中任一项所述的慢病毒载体,其中所述结合结构域通过接头与所述G蛋白连接,任选地其中所述接头是肽接头。

21.如权利要求10-13中任一项所述的靶向脂质颗粒或如权利要求14-20中任一项所述的慢病毒载体,其中所述结合结构域是单结构域抗体或是单链可变片段(scFv)。

22.如权利要求1-9、12和13中任一项所述的靶向脂质颗粒或如权利要求20或权利要求21所述的慢病毒载体,其中所述肽接头包含多达65个氨基酸的长度,任选地2至65个氨基酸的长度。

23.如权利要求1-9、12、13和22中任一项所述的靶向脂质颗粒或如权利要求20-22中任一项所述的慢病毒载体,其中所述肽接头是包含GS、GGS、GGGGS(SEQ ID NO:43)、GGGGGS(SEQ ID NO:41)或它们的组合的柔性接头,或其中所述肽接头包含(GGS)n,其中n为1至10;(GGGGS)n(SEQ ID NO:42),其中n为1至10;或(GGGGGS)n(SEQ ID NO:27),其中n为1至6。

24.如权利要求1-13、22和23中任一项所述的靶向脂质颗粒或如权利要求14-23中任一项所述的慢病毒载体,其中所述G蛋白或其生物活性部分是野生型尼帕病毒G(NiV-G)蛋白或亨德拉病毒G蛋白,或者是其功能活性变体或生物活性部分。

25.如权利要求1-13和22-24中任一项所述的靶向脂质颗粒或如权利要求14-24中任一项所述的慢病毒载体,其中所述G蛋白或其生物活性部分是野生型NiV-G蛋白或其功能活性

变体或生物活性部分。

26.如权利要求24或权利要求25所述的靶向脂质颗粒或慢病毒载体，其中所述NiV-G蛋白或其功能活性变体或生物活性部分包含SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44所示的氨基酸序列、或与SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44具有至少为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、至少为或约84％、至少为或约85％、至少为或约86％、至少为或约87％、至少为或约88％、至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。

27.如权利要求24-26中任一项所述的靶向脂质颗粒或慢病毒颗粒，其中所述NiV-G蛋白是截短的生物活性部分，并且缺乏在所述野生型NiV-G蛋白(SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44)的N末端处或附近的多达40个连续氨基酸残基。

28.如权利要求24-27中任一项所述的靶向脂质颗粒或慢病毒颗粒，其中所述NiV-G蛋白是在野生型NiV-G的N末端截短的生物活性部分，并且具有SEQ ID NO:10-15、35-40或45-50中任一者所示的序列、或与SEQ ID NO:10-15、35-40或45-50具有至少为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、至少为或约84％、至少为或约85％、至少为或约86％、至少为或约87％、至少为或约88％、至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。

29.如权利要求24-28中任一项所述的靶向脂质颗粒或慢病毒载体，其中所述NiV-G蛋白是在所述野生型NiV-G的N末端处或附近具有截短的生物活性部分，所述截短选自由以下组成的组：在所述野生型NiV-G蛋白的N末端处或附近的5个氨基酸截短、在所述野生型NiV-G蛋白的N末端处或附近的10个氨基酸截短、在所述N末端处或附近的15个氨基酸截短、在所述野生型NiV-G蛋白的N末端处或附近的20个氨基酸截短、在所述野生型NiV-G蛋白的N末端处或附近的25个氨基酸截短、在所述野生型NiV-G蛋白的N末端处或附近的30个氨基酸截短、或在所述野生型NiV-G蛋白的N末端处或附近的34个氨基酸截短，任选地其中所述野生型NiV-G蛋白如SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44所示。

30.如权利要求24-29中任一项所述的靶向脂质颗粒或慢病毒载体，其中所述NiV-G蛋白是在所述野生型NiV-G蛋白(SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44)的N末端处或附近具有34个氨基酸截短的生物活性部分。

31.如权利要求1-13和22-30中任一项所述的靶向脂质颗粒或如权利要求14-30中任一项所述的慢病毒载体，其中所述G蛋白或其生物活性部分是表现出与肝配蛋白B2或肝配蛋白B3的结合降低的突变体NiV-G蛋白。

32.如权利要求31所述的靶向脂质颗粒或慢病毒载体，其中所述突变体NiV-G蛋白包含：

一个或多个氨基酸取代，所述一个或多个氨基酸取代对应于选自由参考SEQ ID NO:28所示的编号的E501A、W504A、Q530A和E533A组成的组的氨基酸取代。

33.如权利要求31或权利要求32所述的靶向脂质颗粒或慢病毒载体，其中所述突变体NiV-G蛋白包含：

参考SEQ ID NO:28所示的编号的氨基酸取代E501A、W504A、Q530A和E533A。

34.如权利要求31-33中任一项所述的靶向脂质颗粒或慢病毒载体，其中所述突变体NiV-G蛋白或生物活性部分具有SEQ ID NO:16所示的氨基酸序列、或与SEQ ID NO:16具有为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。

35.如权利要求1-13和22-34中任一项所述的靶向脂质颗粒或如权利要求14-34中任一项所述的慢病毒载体，其中所述F蛋白或其生物活性部分是野生型尼帕病毒F(NiV-F)蛋白或亨德拉病毒F蛋白或是其功能活性变体或生物活性部分。

36.如权利要求1-13和22-35中任一项所述的靶向脂质颗粒或如权利要求14-35中任一项所述的慢病毒载体，其中所述F蛋白或其生物活性部分是野生型NiV-F蛋白或其功能活性变体或生物活性部分。

37.如权利要求1-13和22-36中任一项所述的靶向脂质颗粒或如权利要求14-36中任一项所述的慢病毒载体，其中所述NiV-F蛋白或其功能活性变体或生物活性部分包含SEQ ID NO:2所示的氨基酸序列、或与SEQ ID NO:2具有为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。

38.如权利要求1-13和22-37中任一项所述的靶向脂质颗粒或如权利要求14-37中任一项所述的慢病毒载体，其中所述NiV-F蛋白是其生物活性部分，所述生物活性部分包含：

i)在所述野生型NiV-F蛋白(SEQ ID NO:2)的C末端处或附近的20个氨基酸截短；和/或

ii)在N-连接糖基化位点上的点突变。

39.如权利要求38所述的靶向脂质颗粒或慢病毒载体，其中所述NiV-F蛋白或所述生物活性部分具有SEQ ID NO:5或SEQ ID NO:7所示的序列、或与SEQ ID NO:5或SEQ ID NO:7具有为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。

40.如权利要求1-13和22-37中任一项所述的靶向脂质颗粒或如权利要求14-37中任一项所述的慢病毒载体，其中所述NiV-F蛋白是其生物活性部分，所述生物活性部分在所述野生型NiV-F蛋白(SEQ ID NO:2)的C末端处或附近具有22个氨基酸截短。

41.如权利要求40所述的靶向脂质颗粒或慢病毒载体，其中所述NiV-F蛋白或所述生物活性部分具有SEQ ID NO:8或SEQ ID NO:23所示的序列、或由编码与SEQ ID NO:8或SEQ ID NO:23具有为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约

94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的序列的核苷酸序列编码的氨基酸序列。

42.如权利要求1-13和22-41中任一项所述的靶向脂质颗粒或如权利要求14-41中任一项所述的慢病毒载体，其中所述F蛋白包含SEQ ID NO:23所示的序列，并且所述G蛋白包含SEQ ID NO:16所示的序列。

43.如权利要求1-13和22-42中任一项所述的靶向脂质颗粒，其中所述脂质双层来源于用于产生逆转录病毒或逆转录病毒样颗粒的宿主细胞的膜。

44.如权利要求43所述的靶向脂质颗粒，其中所述宿主细胞选自由以下组成的组：CHO细胞、BHK细胞、MDCK细胞、C3H 10T1/2细胞、FLY细胞、Psi-2细胞、BOSC 23细胞、PA317细胞、WEHI细胞、COS细胞、BSC 1细胞、BSC 40细胞、BMT 10细胞、VERO细胞、W138细胞、MRC5细胞、A549细胞、HT1080细胞、293细胞、293T细胞、B-50细胞、3T3细胞、NIH3T3细胞、HepG2细胞、Saos-2细胞、Huh7细胞、HeLa细胞、W163细胞、211细胞和211A细胞。

45.如权利要求1-13和22-44中任一项所述的靶向脂质颗粒，其中所述脂质双层是或包含病毒包膜。

46.如权利要求1-13和22-45中任一项所述的靶向脂质颗粒，其中所述靶向脂质颗粒包含除所述F蛋白分子和所述G蛋白之外的一种或多种病毒组分。

47.如权利要求46所述的靶向脂质颗粒，其中所述一种或多种病毒组分来自逆转录病毒。

48.如权利要求43-47中任一项所述的靶向脂质颗粒，其中所述逆转录病毒是慢病毒或慢病毒样颗粒。

49.如权利要求1-13和22-48中任一项所述的靶向脂质颗粒或如权利要求14-42中任一项所述的慢病毒载体，其中所述脂质颗粒或慢病毒载体还包含外源药剂。

50.如权利要求49所述的靶向脂质颗粒或慢病毒载体，其中所述外源药剂存在于所述内腔中。

51.如权利要求49或权利要求50所述的靶向脂质颗粒或慢病毒载体，其中所述外源药剂是蛋白质或核酸，任选地其中所述核酸是DNA或RNA。

52.如权利要求49-51中任一项所述的靶向脂质颗粒或慢病毒载体，其中所述外源药剂是编码用于递送至所述靶细胞的运载物的核酸。

53.如权利要求49-52中任一项所述的靶向脂质颗粒或慢病毒载体，其中所述外源药剂编码治疗剂或诊断剂。

54.如权利要求49-53中任一项所述的靶向脂质颗粒或慢病毒载体，其中所述外源药剂编码膜蛋白，任选地其中所述膜蛋白是用于靶向由疾病或病症表达或与疾病或病症相关的细胞的抗原受体。

55.如权利要求53所述的靶向脂质颗粒或慢病毒载体，其中所述膜蛋白是嵌合抗原受体(CAR)。

56.如权利要求52-55中任一项所述的靶向脂质颗粒或慢病毒载体，其中所述靶细胞是T细胞。

57.如权利要求49-53中任一项所述的靶向脂质颗粒或慢病毒载体，其中所述外源药剂是包含有效负载基因的核酸，所述有效负载基因用于纠正遗传缺陷，任选地所述靶细胞中

的遗传缺陷,任选地其中所述遗传缺陷与肝脏细胞或肝细胞相关。

58.一种多核苷酸,其包含编码以下的核酸序列:(i)亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分和(ii)单结构域抗体(sdAb)可变域,其中所述sdAb可变域与所述G蛋白或其生物活性部分的C末端连接。

59.如权利要求58所述的多核苷酸,其中所述单结构域抗体结合存在于靶细胞上的细胞表面分子。

60.如权利要求59所述的多核苷酸,其中所述细胞表面分子是蛋白质、聚糖、脂质或低分子量分子。

61.如权利要求59或权利要求60所述的多核苷酸,其中所述靶细胞选自由以下组成的组:肿瘤浸润淋巴细胞、T细胞、赘生性或肿瘤细胞、病毒感染细胞、干细胞、中枢神经系统(CNS)细胞、造血干细胞(HSC)、肝细胞或完全分化细胞。

62.如权利要求59-61中任一项所述的多核苷酸,其中所述靶细胞选自由以下组成的组:CD3+T细胞、CD4+T细胞、CD8+T细胞、肝细胞、造血干细胞、CD34+造血干细胞、CD105+造血干细胞、CD117+造血干细胞、CD105+内皮细胞、B细胞、CD20+B细胞、CD19+B细胞、癌细胞、CD133+癌细胞、EpCAM+癌细胞、CD19+癌细胞、Her2/Neu+癌细胞、GluA2+神经元、GluA4+神经元、NKG2D+自然杀伤细胞、SLC1A3+星形胶质细胞、SLC7A10+脂肪细胞或CD30+肺上皮细胞。

63.如权利要求59-62中任一项所述的多核苷酸,其中所述细胞表面分子或抗原选自由以下组成的组:ASGR1、ASGR2、TM4SF5、CD8、CD4和低密度脂蛋白受体(LDL-R)。

64.一种包含核酸序列的多核苷酸,所述核酸序列编码(i)亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分和(ii)结合选自由ASGR1、ASGR2、TM4SF5、CD4、CD8和低密度脂蛋白受体(LDL-R)组成的组的细胞表面分子的结合结构域。

65.如权利要求64所述的多核苷酸,其中所述结合结构域是单结构域抗体(sdAb)或是单链可变片段(scFv)。

66.如权利要求58-65中任一项所述的多核苷酸,其中所述核酸序列是第一核酸序列,并且所述多核苷酸还包含编码亨尼帕病毒F蛋白分子或其生物活性部分的第二核酸序列。

67.如权利要求66所述的多核苷酸,其中所述多核苷酸包含IRES或在所述第一核酸序列与所述第二核酸序列之间的编码连接肽的序列,任选地,其中所述连接肽是自切割肽或引起核糖体跳跃的肽,任选地T2A肽。

68.如权利要求58-67中任一项所述的多核苷酸,其还包含至少一个启动子,所述至少一个启动子可操作地连接以控制所述核酸的表达,任选地所述第一核酸序列和所述第二核酸序列的表达。

69.如权利要求58-68中任一项所述的多核苷酸,其中所述sdAb可变域或所述结合结构域通过编码的肽接头与所述G蛋白连接。

70.如权利要求69所述的多核苷酸,其中所述编码的肽接头包含多达65个氨基酸的长度,任选地2至65个氨基酸的长度。

71.如权利要求69或权利要求70所述的多核苷酸,其中所述编码的肽接头包含GS、GGS、GGGGS(SEQ ID NO:43)、GGGGGS(SEQ ID NO:41)以及它们的组合,或其中所述编码的肽接头包含(GGS)n,其中n为1至10;(GGGGS)n(SEQ ID NO:42),其中n为1至10;或(GGGGGS)n(SEQ ID NO:27),其中n为1至4。

72.如权利要求58-71中任一项所述的多核苷酸，其中所述编码的G蛋白是野生型尼帕病毒G(NiV-G)蛋白或亨德拉病毒G蛋白，或是其功能活性变体或生物活性部分，任选地其中所述变体是表现出对天然结合配偶体的结合降低的其变体。

73.如权利要求58-72中任一项所述的多核苷酸，其中所述编码的G蛋白是野生型NiV-G蛋白或其功能活性变体或生物活性部分。

74.如权利要求72或权利要求73所述的多核苷酸，其中所述NiV-G蛋白或其功能活性变体或生物活性部分包含SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44所示的氨基酸序列，或包含与SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44具有至少为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。

75.如权利要求72-74中任一项所述的多核苷酸，其中所述NiV-G蛋白是截短的生物活性部分，并且缺乏在所述野生型NiV-G蛋白(SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44)的N末端处或附近的多达40个连续氨基酸残基。

76.如权利要求72-75中任一项所述的多核苷酸，其中所述NiV-G蛋白是在野生型NiV-G的N末端截短的生物活性部分，并且包含SEQ ID NO:10-15、35-40或45-50中任一者所示的序列、或与SEQ ID NO:10-15、35-40或45-50具有至少为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。

77.如权利要求72-76中任一项所述的多核苷酸，其中所述NiV-G蛋白是在所述野生型NiV-G的N末端处或附近具有截短的生物活性部分，所述截短选自由以下组成的组：在所述野生型NiV-G蛋白的N末端处或附近的5个氨基酸截短、在所述野生型NiV-G蛋白的N末端处或附近的10个氨基酸截短、在所述N末端处或附近的15个氨基酸截短、在所述野生型NiV-G蛋白的N末端处或附近的20个氨基酸截短、在所述野生型NiV-G蛋白的N末端处或附近的25个氨基酸截短、在所述野生型NiV-G蛋白的N末端处或附近的30个氨基酸截短、或在所述野生型NiV-G蛋白的N末端处或附近的34个氨基酸截短，任选地其中所述野生型NiV-G蛋白如SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44所示。

78.如权利要求72-77中任一项所述的多核苷酸，其中所述NiV-G蛋白是在所述野生型NiV-G蛋白(SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44)的N末端处或附近具有34个氨基酸截短的生物活性部分。

79.如权利要求58-78中任一项所述的多核苷酸，其中所述G蛋白是表现出与肝配蛋白B2或肝配蛋白B3的结合降低的突变体NiV-G蛋白。

80.如权利要求79所述的多核苷酸，其中所述突变体NiV-G蛋白包含：一个或多个氨基酸取代，所述一个或多个氨基酸取代对应于选自由参考SEQ ID NO:28所示的编号的E501A、W504A、Q530A和E533A组成的组的氨基酸取代。

81.如权利要求79或权利要求80所述的多核苷酸，其中所述突变体NiV-G蛋白包含参考

SEQ ID NO:28所示的编号的氨基酸取代E501A、W504A、Q530A和E533A。

82.如权利要求79-81中任一项所述的多核苷酸，其中所述突变体NiV-G蛋白包含SEQ ID NO:16所示的氨基酸序列、或与SEQ ID NO:16具有至少为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。

83.如权利要求66-82中任一项所述的多核苷酸，其中所述F蛋白或其生物活性部分是野生型尼帕病毒F(NiV-F)蛋白或亨德拉病毒F蛋白或是其功能活性变体或生物活性部分。

84.如权利要求66-83中任一项所述的多核苷酸，其中所述F蛋白或其生物活性部分是野生型NiV-F蛋白或其功能活性变体或生物活性部分。

85.如权利要求83或权利要求84所述的多核苷酸，其中所述NiV-F蛋白或其功能活性变体或生物活性部分包含SEQ ID NO:2所示的氨基酸序列、或与SEQ ID NO:2具有为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。

86.如权利要求85或权利要求85所述的多核苷酸，其中所述NiV-F蛋白是其生物活性部分，所述生物活性部分包含：

i)在所述野生型NiV-F蛋白(SEQ ID NO:2)的C末端处或附近的20个氨基酸截短；和/或

ii)在N-连接糖基化位点上的点突变。

87.如权利要求86所述的多核苷酸，其中所述NiV-F蛋白或所述生物活性部分具有SEQ ID NO:5或SEQ ID NO:7所示的序列、或与SEQ ID NO:5或SEQ ID NO:7具有为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。

88.如权利要求85或权利要求86所述的多核苷酸，其中所述NiV-F蛋白是在所述野生型NiV-F蛋白(SEQ ID NO:2)的C末端处或附近具有22个氨基酸截短的其生物活性部分。

89.如权利要求88所述的多核苷酸，其中所述NiV-F蛋白或所述生物活性部分具有SEQ ID NO:8或SEQ ID NO:23所示的序列、或由编码与SEQ ID NO:8或SEQ ID NO:23具有为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的序列的核苷酸序列编码的氨基酸序列。

90.如权利要求66-89中任一项所述的多核苷酸，其中所述F蛋白包含SEQ ID NO:23所示的序列，并且所述G蛋白包含SEQ ID NO:16所示的序列。

91.一种载体,其包含如权利要求58-90中任一项所述的多核苷酸。

92.如权利要求91所述的载体,其中所述载体是哺乳动物载体、病毒载体或人工染色体,任选地其中所述人工染色体是细菌人工染色体(BAC)。

93.一种质粒,其包含如权利要求58-90中任一项所述的多核苷酸。

94.如权利要求93所述的质粒,其还包含编码用于慢病毒产生的蛋白质的一种或多种核酸。

95.一种细胞,其包含如权利要求58-90中任一项所述的多核苷酸或如权利要求91或权利要求92所述的载体或如权利要求93或94所述的质粒。

96.一种制备靶向脂质颗粒的方法,所述靶向脂质颗粒包含亨尼帕病毒F蛋白分子或其生物活性部分和靶向包膜蛋白,所述靶向包膜蛋白包含亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分和单结构域抗体(sdAb)可变域,所述方法包括:

a)提供包含编码亨尼帕病毒F蛋白分子或其生物活性部分的核酸和编码靶向包膜蛋白的核酸的细胞,所述靶向包膜蛋白包含亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分和单结构域抗体(sdAb)可变域;

b)在允许产生靶向脂质颗粒的条件下培养所述细胞,以及

c)从所述细胞中分离、富集或纯化所述靶向脂质颗粒,从而制得所述靶向脂质颗粒。

97.一种制备假型化慢病毒载体的方法,所述方法包括:

a)提供生产者细胞,所述生产者细胞包含慢病毒病毒核酸、编码亨尼帕病毒F蛋白分子或其生物活性部分的核酸和编码靶向包膜蛋白的核酸,所述靶向包膜蛋白包含亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分和单结构域抗体;

b)在允许产生慢病毒载体的条件下培养所述细胞,以及

c)从所述细胞中分离、富集或纯化所述慢病毒载体,从而制得所述假型化慢病毒载体。

98.一种制备靶向脂质颗粒的方法,所述靶向脂质颗粒包含亨尼帕病毒F蛋白分子或其生物活性部分和靶向包膜蛋白,所述靶向包膜蛋白包含亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分和结合结构域,所述方法包括:

a)提供包含编码亨尼帕病毒F蛋白分子或其生物活性部分的核酸和编码靶向包膜蛋白的核酸的细胞,所述靶向包膜蛋白包含亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分和结合结构域,其中所述结合结构域:

(i)结合细胞表面分子,所述细胞表面分子选自由以下组成的组:ASGR1、ASGR2和TM4SF5,任选地人ASGR1、人ASGR2和人ASGR2;

(ii)结合细胞表面分子,所述细胞表面分子选自由以下组成的组:CD4或CD8,任选地人CD4或人CD8;或

(iii)结合细胞表面分子,所述细胞表面分子是低密度脂蛋白受体(LDL-R),任选地人LDL-R;

b)在允许产生靶向脂质颗粒的条件下培养所述细胞,以及

c)从所述细胞中分离、富集或纯化所述靶向脂质颗粒,从而制得所述靶向脂质颗粒。

99.一种制备假型化慢病毒载体的方法,所述方法包括:

a)提供生产者细胞,所述生产者细胞包含慢病毒病毒核酸、编码亨尼帕病毒F蛋白分子或其生物活性部分的核酸和编码靶向包膜蛋白的核酸,所述靶向包膜蛋白包含亨尼帕病毒

包膜附着糖蛋白G(G蛋白)或其生物活性部分和结合结构域,其中所述结合结构域:

(i)结合细胞表面分子,所述细胞表面分子选自由以下组成的组:ASGR1、ASGR2和TM4SF5,任选地人ASGR1、人ASGR2和人ASGR2;

(ii)结合细胞表面分子,所述细胞表面分子选自由以下组成的组:CD4或CD8,任选地人CD4或人CD8;或

(iii)结合细胞表面分子,所述细胞表面分子是低密度脂蛋白受体(LDL-R),任选地人LDL-R;

b)在允许产生慢病毒载体的条件下培养所述生产者细胞,以及

c)从所述细胞中分离、富集或纯化所述慢病毒载体,从而制得所述假型化慢病毒载体。

100.如权利要求98或权利要求99所述的方法,其中所述结合结构域是单结构域抗体或是单链可变片段(scFv)。

101.一种制备靶向脂质颗粒的方法,所述靶向脂质颗粒包含亨尼帕病毒F蛋白分子或其生物活性部分和靶向包膜蛋白,所述方法包括:

a)提供包含如权利要求58-90中任一项所述的多核苷酸或如权利要求91或权利要求92所述的载体或如权利要求93或权利要求94所述的质粒的细胞;

b)在允许产生靶向脂质颗粒的条件下培养所述细胞,以及

c)从所述细胞中分离、富集或纯化所述靶向脂质颗粒,从而制得所述靶向脂质颗粒。

102.一种制备假型化慢病毒载体的方法,所述方法包括:

a)提供包含慢病毒病毒核酸和如权利要求58-90中任一项所述的多核苷酸或如权利要求91或权利要求92所述的载体或如权利要求93或权利要求94所述的质粒的生产者细胞;

b)在允许产生慢病毒载体的条件下培养所述细胞,以及

c)从所述细胞中分离、富集或纯化所述慢病毒载体,从而制得所述假型化慢病毒载体。

103.如权利要求101或权利要求102所述的方法,其中在步骤(b)之前,所述方法还包括向所述细胞提供编码亨尼帕病毒F蛋白分子或其生物活性部分的多核苷酸。

104.如权利要求96-103所述的方法,其中所述细胞是哺乳动物细胞。

105.如权利要求96-104中任一项所述的方法,其中所述细胞是包含病毒核酸,任选地逆转录病毒核酸或慢病毒核酸的生产者细胞,并且所述靶向脂质颗粒是病毒颗粒或病毒样颗粒,任选地逆转录病毒颗粒或逆转录病毒样颗粒,任选地慢病毒颗粒或慢病毒样颗粒。

106.一种生产者细胞,其包含如权利要求58-90中任一项所述的多核苷酸或如权利要求91或权利要求92所述的载体或如权利要求93或权利要求94所述的质粒。

107.如权利要求106所述的生产者细胞,其还包含编码亨尼帕病毒F蛋白或其生物活性部分的核酸。

108.如权利要求106或权利要求107所述的生产者细胞,其中所述细胞还包含病毒核酸,任选地其中所述病毒核酸是慢病毒核酸。

109.一种生产者细胞,其包含(i)病毒核酸和(ii)编码亨尼帕病毒F蛋白分子或其生物活性部分的核酸和(iii)编码靶向包膜蛋白的核酸,所述靶向包膜蛋白包含亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分和单结构域抗体(sdAb)可变域,任选地其中所述病毒核酸是慢病毒核酸。

110.如权利要求109所述的生产者细胞,其中所述单结构域抗体结合存在于靶细胞上

的细胞表面分子。

111.如权利要求110所述的生产者细胞,其中所述细胞表面分子是蛋白质、聚糖、脂质或低分子量分子。

112.如权利要求110或权利要求111所述的生产者细胞,其中所述靶细胞选自由以下组成的组:肿瘤浸润淋巴细胞、T细胞、赘生性或肿瘤细胞、病毒感染细胞、干细胞、中枢神经系统(CNS)细胞、造血干细胞(HSC)、肝细胞或完全分化细胞。

113.如权利要求110-112中任一项所述的生产者细胞,其中所述靶细胞选自由以下组成的组:CD3+T细胞、CD4+T细胞、CD8+T细胞、肝细胞、造血干细胞、CD34+造血干细胞、CD105+造血干细胞、CD117+造血干细胞、CD105+内皮细胞、B细胞、CD20+B细胞、CD19+B细胞、癌细胞、CD133+癌细胞、EpCAM+癌细胞、CD19+癌细胞、Her2/Neu+癌细胞、GluA2+神经元、GluA4+神经元、NKG2D+自然杀伤细胞、SLC1A3+星形胶质细胞、SLC7A10+脂肪细胞或CD30+肺上皮细胞。

114.一种生产者细胞,其包含(i)病毒核酸和(ii)编码亨尼帕病毒F蛋白分子或其生物活性部分的核酸和(iii)编码靶向包膜蛋白的核酸,所述靶向包膜蛋白包含亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分和结合结构域,其中所述结合结构域:

(i)结合细胞表面分子,所述细胞表面分子选自由以下组成的组:ASGR1、ASGR2和TM4SF5,任选地人ASGR1、人ASGR2和人ASGR2;

(ii)结合细胞表面分子,所述细胞表面分子选自由以下组成的组:CD4或CD8,任选地人CD4或人CD8;或

(iii)结合细胞表面分子,所述细胞表面分子是低密度脂蛋白受体(LDL-R),任选地人LDL-R;

任选地其中所述病毒核酸是慢病毒核酸。

115.如权利要求114所述的生产者细胞,其中所述结合结构域是单结构域抗体或是单链可变片段。

116.如权利要求109-115中任一项所述的生产者细胞,其中所述F蛋白或其生物活性部分是野生型尼帕病毒F(NiV-F)蛋白或亨德拉病毒F蛋白或是其功能活性变体或生物活性部分。

117.如权利要求109-116中任一项所述的生产者细胞,其中所述F蛋白或其生物活性部分是野生型NiV-F蛋白或其功能活性变体或生物活性部分。

118.如权利要求109-117中任一项所述的生产者细胞,其中所述亨尼帕病毒F蛋白分子或其生物活性部分包含:

(i)SEQ ID NO:2、SEQ ID NO:5、SEQ ID NO:7、SEQ ID NO:8或SEQ ID NO:23所示的序列;

(ii)与SEQ ID NO:2、SEQ ID NO:5、SEQ ID NO:7、SEQ ID NO:8或SEQ ID NO:23具有为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、至少为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的氨基酸序列。

119.如权利要求109-118中任一项所述的生产者细胞,其中所述编码的G蛋白是野生型尼帕病毒G(NiV-G)蛋白或亨德拉病毒G蛋白,或是其功能活性变体或生物活性部分,任选地其中所述变体是表现出对天然结合配偶体的结合降低的其变体。

120.如权利要求109-119中任一项所述的生产者细胞,其中所述编码的G蛋白是野生型NiV-G蛋白或其功能活性变体或生物活性部分。

121.如权利要求109-120中任一项所述的生产者细胞,其中所述亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分包含:

(i)SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44所示的序列;

(ii)与SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44具有为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、至少为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的氨基酸序列。

122.如权利要求109-121中任一项所述的生产者细胞,其中所述亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分包含:

(i)SEQ ID NO:10-16、35-40或45-51所示的序列;

(ii)与SEQ ID NO:10-16、35-40或45-51具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约87%、至少为或约88%、或至少为或约89%、为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、至少为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的氨基酸序列。

123.如权利要求109-122中任一项所述的生产者细胞,其中:

与已将相同的包膜蛋白掺入其膜(例如,质膜)中但与替代靶向部分融合的参考生产者细胞相比,所述生产者细胞具有更多的靶向包膜蛋白的膜(例如质膜)表达,任选地其中所述替代靶向部分是单链可变片段(scFv);和/或

所述生产者细胞具有每平方微米至少20个蛋白质(例如,至少50个、100个、200个、500个、1000个、2000个、5000个或10,000个蛋白质)的所述靶向包膜蛋白在所述生产者细胞的膜(例如,质膜)上的表达;和/或所述靶向包膜蛋白构成所述生产者细胞的总膜(例如,质膜)蛋白的至少0.1%(例如,至少0.2%、0.5%、1%、2%、3%、4%、5%、6%、7%、8%、9%或10%)(例如,按总蛋白重量计)。

124.一种靶向脂质颗粒或假型化慢病毒载体,其通过如权利要求96-105中任一项所述的方法或由如权利要求106-122中任一项所述的生产者细胞产生。

125.如权利要求1-13、21-57和124中任一项所述的靶向脂质颗粒,其中与已将相同包膜蛋白掺入相似脂质双层中但与替代靶向部分融合的参考脂质颗粒相比,所述靶向脂质颗粒具有更多的所述靶向包膜蛋白表达,任选地其中所述替代靶向部分是单链可变片段(scFv)。

126.如权利要求1-13、21-57、124和125中任一项所述的靶向脂质颗粒,如权利要求13-42、49-57和124中任一项所述的慢病毒载体,其中:

转导后在靶细胞中的滴度为或大于1x 10^6个转导单位(TU)/mL、为或大于2x 10^6TU/mL、

为或大于3x 10^6TU/mL、为或大于4x 10^6TU/mL、为或大于5x 10^6TU/mL、为或大于6x 10^6TU/mL、为或大于7x 10^6TU/mL、为或大于8x 10^6TU/mL、为或大于9x 10^6TU/mL、或为或大于1x 10^7TU/mL;和/或

所述靶向包膜蛋白以至少约(0.001、0.002、0.005、0.01、0.02、0.05、0.1、0.2或0.5)个靶向包膜蛋白/nm^2的密度存在于所述靶向脂质颗粒的表面上。

127.一种组合物,其包含如权利要求1-57和124-126中任一项所述的多个靶向脂质颗粒或多个慢病毒载体。

128.如权利要求127所述的组合物,其还包含药学上可接受的载体。

129.如权利要求127或权利要求128所述的组合物,其中:

在所述组合物中的脂质颗粒或慢病毒载体群体中,大于为或约50%、大于为或约55%、大于为或约60%、大于为或约65%、大于为或约70%、或大于为或约75%对靶向包膜蛋白呈表面阳性;和/或

所述靶向包膜蛋白以至少约(0.001、0.002、0.005、0.01、0.02、0.05、0.1、0.2或0.5)个靶向包膜蛋白/nm^2的平均密度存在于所述靶向脂质颗粒的表面上。

130.一种转导细胞的方法,所述方法包括用如权利要求13-42、49-57和124中任一项所述的慢病毒载体或包含如权利要求13-42、49-57和124中任一项所述的一种慢病毒载体或多种慢病毒载体的组合物转导细胞。

131.如权利要求130所述的方法,其中:

所述慢病毒载体或靶向脂质颗粒的所述靶向包膜蛋白靶向CD4,并且所述细胞是CD4+细胞;或

所述慢病毒载体的所述靶向包膜蛋白靶向CD8,并且所述细胞是CD8+细胞;或

所述慢病毒载体的所述靶向包膜蛋白靶向ASGR1、ASGR2或TM4SF5,并且所述细胞是肝细胞。

132.一种向受试者(例如,人类受试者)递送外源药剂的方法,所述方法包括向所述受试者施用如权利要求49-57、125和126中任一项所述的靶向脂质颗粒或慢病毒载体,其中所述靶向脂质颗粒或慢病毒载体包含所述外源药剂。

133.一种向受试者(例如,人类受试者)递送外源药剂的方法,所述方法包括向所述受试者施用如权利要求127-129中任一项所述的组合物,其中所述多种的靶向脂质颗粒或慢病毒载体包含所述外源药剂。

134.一种将嵌合抗原受体(CAR)递送至细胞的方法,所述方法包括使细胞与如权利要求14-16和19-42中任一项所述的慢病毒载体或如权利要求55或权利要求56所述的靶向脂质颗粒接触,其中所述慢病毒载体或靶向脂质颗粒包含编码所述CAR的核酸。

135.一种将嵌合抗原受体(CAR)递送至细胞的方法,所述方法包括使细胞与如权利要求127-129中任一项所述的组合物接触,其中所述多种的慢病毒载体或靶向脂质颗粒包含编码所述CAR的核酸。

136.一种将外源药剂递送至肝细胞的方法,所述方法包括使细胞与如权利要求17-42中任一项所述的慢病毒载体或如权利要求57所述的靶向脂质颗粒或慢病毒载体接触。

137.一种将外源药剂递送至肝细胞的方法,所述方法包括使细胞与如权利要求127-129中任一项所述的组合物接触,其中所述多种的慢病毒载体或靶向脂质颗粒包含用于递

送至所述肝细胞的外源药剂。

138.如权利要求134-137中任一项所述的方法，其中所述接触用慢病毒载体或靶向脂质颗粒转导所述细胞，任选地其中所述接触在受试者体内进行。

139.一种治疗受试者(例如，人类受试者)的疾病或病况的方法，所述方法包括向所述受试者施用如权利要求1-13、21-57和124-126中任一项所述的靶向脂质颗粒或如权利要求13-42、49-57、124和126中任一项所述的慢病毒载体或如权利要求127-129中任一项所述的组合物。

140.一种使哺乳动物细胞与靶向脂质颗粒融合的方法，所述方法包括向所述受试者施用如权利要求1-13、21-57和124-126中任一项所述的靶向脂质颗粒、如权利要求13-42、49-57、124和126中任一项所述的慢病毒载体或如权利要求127-129中任一项所述的组合物。

141.如权利要求140所述的方法，其中使所述哺乳动物细胞与所述靶向脂质颗粒的融合将外源药剂递送至受试者(例如，人类受试者)。

靶向脂质颗粒及其组合物和用途

[0001] 相关申请的交叉引用

[0002] 本申请要求于2020年3月31日提交的标题为“靶向脂质颗粒及其组合物和用途(Targeted Lipid Particles and Compositions and Uses Thereof)”的美国临时申请63/003,168和于2021年2月26日提交的标题为“靶向脂质颗粒及其组合物和用途(Targeted Lipid Particles and Compositions and Uses Thereof)”的美国临时申请63/154,341的优先权,出于所有目的将其各自的内容通过引用整体并入。

[0003] 通过引用并入序列表

[0004] 本申请连同序列表一起以电子格式提交。序列表以名为186152003640SeqList.TXT的文档形式提供,该文档创建于2021年3月29日,大小为2,071,683字节。电子格式的序列表中的信息以引用方式整体并入。

技术领域

[0005] 本公开涉及脂质颗粒,所述脂质颗粒包含围绕内腔或空腔的脂质双层(lipid bilayer)、亨尼帕病毒(henipavirus)F蛋白分子或其生物活性部分和靶向包膜蛋白,所述靶向包膜蛋白包含亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分和结合结构域(诸如单结构域抗体(sdAb)可变域)。本公开还提供了包含与结合结构域(诸如sdAb可变域)融合或连接的G蛋白的靶向包膜蛋白以及编码此类蛋白质的多核苷酸。还公开了含有此类靶向脂质颗粒的生产者细胞和组合物以及制备和使用所述靶向脂质颗粒的方法。

背景技术

[0006] 脂质颗粒(包括病毒样颗粒和病毒载体)通常用于将外源药剂递送至细胞。然而,将脂质颗粒递送至某些靶细胞可能具有挑战性。对于慢病毒载体,可以通过用异源包膜蛋白假型化来改变宿主范围。某些再靶向包膜蛋白可能不够稳定或在脂质颗粒表面上表达。对用于靶向期望的细胞的改进的脂质颗粒(包括病毒样颗粒和病毒载体)存在需求。所提供的公开解决了这种需求。

发明内容

[0007] 本文提供了靶向脂质颗粒,所述靶向脂质颗粒包含(a)围绕内腔的脂质双层,(b)亨尼帕病毒F蛋白分子或其生物活性部分,和(c)靶向包膜蛋白,所述靶向包膜蛋白包含(i)亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分和(ii)单结构域抗体(sdAb)可变域,其中所述sdAb可变域与G蛋白或其生物活性部分的C末端连接,其中所述F蛋白分子或其生物活性部分和所述靶向包膜蛋白包埋于所述脂质双层中。在一些实施方案中,所述单结构域抗体通过接头与G蛋白连接。在一些实施方案中,所述接头是肽接头。

[0008] 本文提供了靶向脂质颗粒,所述靶向脂质颗粒包含(a)围绕内腔的脂质双层,(b)亨尼帕病毒F蛋白分子或其生物活性部分;和(c)靶向包膜蛋白,所述靶向包膜蛋白包含通过肽接头与单结构域抗体(sdAb)可变域连接的亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其

生物活性部分，其中所述单结构域抗体与靶细胞的细胞表面分子结合，其中所述F蛋白分子或其生物活性部分和所述靶向包膜蛋白包埋在脂质双层中。在一些实施方案中，所述F蛋白分子或其生物活性部分的N末端暴露在脂质双层的外侧。在一些实施方案中，所述G蛋白的C末端暴露在脂质双层的外侧。

[0009] 在一些实施方案中，所述单结构域抗体结合存在于靶细胞上的细胞表面分子。在一些实施方案中，所述细胞表面分子是蛋白质、聚糖、脂质或低分子量分子。在一些任何实施方案中，所述单结构域抗体结合存在于靶细胞上的抗原或其部分。在一些实施方案中，所述抗原是细胞表面分子或细胞表面分子的含有由所述单结构域抗体识别的表位的部分。在一些任何实施方案中，所述靶细胞选自由以下组成的组：肿瘤浸润淋巴细胞、T细胞、赘生性或肿瘤细胞、病毒感染细胞、干细胞、中枢神经系统(CNS)细胞、造血干细胞(HSC)、肝细胞或完全分化细胞。在一些实施方案中，所述靶细胞选自由以下组成的组：CD3+T细胞、CD4+T细胞、CD8+T细胞、肝细胞、造血干细胞、CD34+造血干细胞、CD105+造血干细胞、CD117+造血干细胞、CD105+内皮细胞、B细胞、CD20+B细胞、CD19+B细胞、癌细胞、CD133+癌细胞、EpCAM+癌细胞、CD19+癌细胞、Her2/Neu+癌细胞、GluA2+神经元、GluA4+神经元、NKG2D+自然杀伤细胞、SLC1A3+星形胶质细胞、SLC7A10+脂肪细胞或CD30+肺上皮细胞。在一些任何实施方案中，所述靶细胞是肝细胞。在一些任何实施方案中，所述细胞表面分子或抗原选自由以下组成的组：ASGR1、ASGR2和TM4SF5。

[0010] 在一些任何实施方案中，所述靶细胞是T细胞。在一些任何实施方案中，所述细胞表面分子或抗原是CD8或CD4。

[0011] 在一些任何实施方案中，所述细胞表面分子或抗原是LDL-R。

[0012] 本文提供了靶向脂质颗粒，所述靶向脂质颗粒包含(a)围绕内腔的脂质双层，(b)亨尼帕病毒F蛋白分子或其生物活性部分，和(c)靶向包膜蛋白，所述靶向包膜蛋白包含(i)亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分和(ii)结合结构域，其中所述结合结构域与所述G蛋白或其生物活性部分的C末端连接，并且其中所述结合结构域结合选自由以下组成的组的细胞表面分子：ASGR1、ASGR2和TM4SF5，任选地人ASGR1、人ASGR2和人ASGR2，

[0013] 其中所述F蛋白分子或其生物活性部分和所述靶向包膜蛋白包埋在脂质双层中。

[0014] 本文提供了靶向脂质颗粒，所述靶向脂质颗粒包含(a)围绕内腔的脂质双层，(b)亨尼帕病毒F蛋白分子或其生物活性部分，和(c)靶向包膜蛋白，所述靶向包膜蛋白包含(i)亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分和(ii)结合结构域，其中所述结合结构域与所述G蛋白或其生物活性部分的C末端连接，并且其中所述结合结构域结合选自由以下组成的组的细胞表面分子：CD8和CD4，任选地人CD8或人CD4，其中所述F蛋白分子或其生物活性部分和所述靶向包膜蛋白包埋在脂质双层中。

[0015] 本文提供了靶向脂质颗粒，所述靶向脂质颗粒包含(a)围绕内腔的脂质双层，(b)亨尼帕病毒F蛋白分子或其生物活性部分，和(c)靶向包膜蛋白，所述靶向包膜蛋白包含(i)亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分和(ii)结合结构域，其中所述结合结构域与所述G蛋白或其生物活性部分的C末端连接，并且其中所述结合结构域结合细胞表面分子，所述细胞表面分子是低密度脂蛋白受体(LDL-R)，任选地人LDL-R，其中所述F蛋白分子或其生物活性部分和所述靶向包膜蛋白包埋在脂质双层中。

[0016] 在一些任何实施方案中,所述脂质颗粒是慢病毒载体。在一些任何实施方案中,所述结合结构域通过接头与G蛋白连接。在一些任何实施方案中,所述接头是肽接头。

[0017] 本文提供了慢病毒载体,所述慢病毒载体包含靶向细胞表面分子的结合结构域,所述细胞表面分子选自由以下组成的组:ASGR1、ASGR2和TM4SF5,任选地人ASGR1、人ASGR2和人TM4SF5,其中所述慢病毒载体用再靶向病毒融合蛋白假型化,所述再靶向病毒融合蛋白包含:(a)亨尼帕病毒F蛋白分子或其生物活性部分;和(b)包含与亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分连接的结合结构域的靶向包膜蛋白。

[0018] 本文提供了慢病毒载体,所述慢病毒载体包含靶向细胞表面分子的结合结构域,所述细胞表面分子选自由以下组成的组:CD8和CD4,任选地人CD8和人CD4,其中所述慢病毒载体用再靶向病毒融合蛋白假型化,所述再靶向病毒融合蛋白包含:(a)亨尼帕病毒F蛋白分子或其生物活性部分;和(b)包含与亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分连接的结合结构域的靶向包膜蛋白。

[0019] 本文提供了慢病毒载体,所述慢病毒载体包含靶向低密度脂蛋白受体(LDL-R)的结合结构域,任选地其中所述LDL-R是人LDL-R,其中所述慢病毒载体用再靶向病毒融合蛋白假型化,所述再靶向病毒融合蛋白包含:(a)亨尼帕病毒F蛋白分子或其生物活性部分;和(b)包含与亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分连接的结合结构域的靶向包膜蛋白。

[0020] 在一些任何实施方案中,所述结合结构域与G蛋白或其生物活性部分的C末端连接。

[0021] 本文提供了慢病毒载体,所述慢病毒载体包含(a)亨尼帕病毒F蛋白分子或其生物活性部分;和(b)靶向包膜蛋白,所述靶向包膜蛋白包含(i)亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分和(ii)结合结构域,其中所述结合结构域与所述G蛋白或其生物活性部分的C末端连接,并且其中所述结合结构域结合CD4;和(c)包含编码嵌合抗原受体(CAR)的核酸的运载物,其中所述CAR包含(i)结合细胞外抗原(例如,CD19或BCMA)的细胞外抗原结合结构域和(ii)细胞内信号传导区、CD3ζ信号传导结构域和任选地4-1BB或CD28共刺激信号传导结构域。在一些实施方案中,所述CAR的细胞外抗原结合结构域是scFv。

[0022] 在一些任何实施方案中,所述慢病毒载体能够将编码所述CAR的核酸递送至T细胞。在一些实施方案中,所述T细胞在受试者的体内。

[0023] 本文提供了慢病毒载体,所述慢病毒载体包含:(a)亨尼帕病毒F蛋白分子或其生物活性部分;和(b)靶向包膜蛋白,所述靶向包膜蛋白包含(i)亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分和(ii)结合结构域,其中所述结合结构域与所述G蛋白或其生物活性部分的C末端连接,并且其中所述结合结构域结合ASGR1;其中所述慢病毒载体能够靶向肝细胞。在一些任何实施方案中,所述慢病毒载体还包含用于递送至肝细胞的外源药剂。

[0024] 在一些任何实施方案中,所述慢病毒载体能够将所述外源药剂递送至肝细胞,任选地其中所述肝细胞在受试者的体内。

[0025] 在一些任何实施方案中,所述结合结构域通过接头与所述G蛋白连接。在一些任何实施方案中,所述接头是肽接头。在一些任何实施方案中,所述结合结构域是单结构域抗体。在一些任何实施方案中,所述结合结构域是单链可变片段(scFv)。

[0026] 在一些任何实施方案中,所述肽接头包含多达65个氨基酸的长度。在一些任何实

施方案中，所述肽接头包含多达50个氨基酸的长度。在一些任何实施方案中，所述肽接头包含从或从约2至65个氨基酸、2至60个氨基酸、2至56个氨基酸、2至52个氨基酸、2至48个氨基酸、2至44个氨基酸、2至40个氨基酸、2至36个氨基酸、2至32个氨基酸、2至28个氨基酸、2至24个氨基酸、2至20个氨基酸、2至18个氨基酸、2至14个氨基酸、2至12个氨基酸、2至10个氨基酸、2至8个氨基酸、2至6个氨基酸、6至65个氨基酸、6至60个氨基酸、6至56个氨基酸、6至52个氨基酸、6至48个氨基酸、6至44个氨基酸、6至40个氨基酸、6至36个氨基酸、6至32个氨基酸、6至28个氨基酸、6至24个氨基酸、6至20个氨基酸、6至18个氨基酸、6至14个氨基酸、6至12个氨基酸、6至10个氨基酸、6至8个氨基酸、8至65个氨基酸、8至60个氨基酸、8至56个氨基酸、8至52个氨基酸、8至48个氨基酸、8至44个氨基酸、8至40个氨基酸、8至36个氨基酸、8至32个氨基酸、8至28个氨基酸、8至24个氨基酸、8至20个氨基酸、8至18个氨基酸、8至14个氨基酸、8至12个氨基酸、8至10个氨基酸、10至65个氨基酸、10至60个氨基酸、10至56个氨基酸、10至52个氨基酸、10至48个氨基酸、10至44个氨基酸、10至40个氨基酸、10至36个氨基酸、10至32个氨基酸、10至28个氨基酸、10至24个氨基酸、10至20个氨基酸、10至18个氨基酸、10至14个氨基酸、10至12个氨基酸、12至65个氨基酸、12至60个氨基酸、12至56个氨基酸、12至52个氨基酸、12至48个氨基酸、12至44个氨基酸、12至40个氨基酸、12至36个氨基酸、12至32个氨基酸、12至28个氨基酸、12至24个氨基酸、12至20个氨基酸、12至18个氨基酸、12至14个氨基酸、14至65个氨基酸、14至60个氨基酸、14至56个氨基酸、14至52个氨基酸、14至48个氨基酸、14至44个氨基酸、14至40个氨基酸、14至36个氨基酸、14至32个氨基酸、14至28个氨基酸、14至24个氨基酸、14至20个氨基酸、14至18个氨基酸、18至65个氨基酸、18至60个氨基酸、18至56个氨基酸、18至52个氨基酸、18至48个氨基酸、18至44个氨基酸、18至40个氨基酸、18至36个氨基酸、18至32个氨基酸、18至28个氨基酸、18至24个氨基酸、18至20个氨基酸、20至65个氨基酸、20至60个氨基酸、20至56个氨基酸、20至52个氨基酸、20至48个氨基酸、20至44个氨基酸、20至40个氨基酸、20至36个氨基酸、20至32个氨基酸、20至28个氨基酸、20至26个氨基酸、20至24个氨基酸、24至65个氨基酸、24至60个氨基酸、24至56个氨基酸、24至52个氨基酸、24至48个氨基酸、24至44个氨基酸、24至40个氨基酸、24至36个氨基酸、24至32个氨基酸、24至30个氨基酸、24至28个氨基酸、28至65个氨基酸、28至60个氨基酸、28至56个氨基酸、28至52个氨基酸、28至48个氨基酸、28至44个氨基酸、28至40个氨基酸、28至36个氨基酸、28至34个氨基酸、28至32个氨基酸、32至65个氨基酸、32至60个氨基酸、32至56个氨基酸、32至52个氨基酸、32至48个氨基酸、32至44个氨基酸、32至40个氨基酸、32至38个氨基酸、32至36个氨基酸、36至65个氨基酸、36至60个氨基酸、36至56个氨基酸、36至52个氨基酸、36至48个氨基酸、36至44个氨基酸、36至40个氨基酸、40至65个氨基酸、40至60个氨基酸、40至56个氨基酸、40至52个氨基酸、40至48个氨基酸、40至44个氨基酸、44至65个氨基酸、44至60个氨基酸、44至56个氨基酸、44至52个氨基酸、44至48个氨基酸、48至65个氨基酸、48至60个氨基酸、48至56个氨基酸、48至52个氨基酸、50至65个氨基酸、50至60个氨基酸、50至56个氨基酸、50至52个氨基酸、54至65个氨基酸、54至60个氨基酸、54至56个氨基酸、58至65个氨基酸、58至60个氨基酸、或60至65个氨基酸。在一些任何实施方案中，所述肽接头包含长度为3个、4个、5个、6个、7个、8个、9个、10个、11个、12个、13个、14个、15个、16个、17个、18个、19,20个、21个、22个、23个、24个、25个、26个、27个、28个、29个、30个、31个、32个、33个、34个、35个、36个、37个、38个、39个、40个、41

个、42个、43个、44个、45个、46个、47个、48个、49个、50个、51个、52个、53个、54个、55个、56个、57个、58个、59个、60个、61个、62个、63个、64个或65个氨基酸的多肽。在一些任何实施方案中，其中所述肽接头是柔性接头，所述柔性接头包含GS、GGS、GGGGS(SEQ ID NO:43)、GGGGGS(SEQ ID NO:41)或它们的组合。在一些任何实施方案中，所述肽接头包含(GGS)n，其中n为1至10。在一些任何实施方案中，所述肽接头包含(GGGGS)n(SEQ ID NO:42)，其中n为1至10。在一些任何实施方案中，所述肽接头包含(GGGGGS)n(SEQ ID NO:27)，其中n为1至6。

[0027] 在一些任何实施方案中，所述G蛋白或其生物活性部分是野生型尼帕病毒G(NiV-G)蛋白或亨德拉病毒(Hendra virus)G蛋白。在一些任何实施方案中，所述G蛋白或其生物活性部分是野生型NiV-G蛋白或其功能活性变体或生物活性部分。在一些任何实施方案中，突变体NiV-G蛋白或其功能活性变体或生物活性部分包含与SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44具有至少为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、至少为或约84％、至少为或约85％、至少为或约86％、至少为或约87％、至少为或约88％、至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。

[0028] 在一些任何实施方案中，NiV-G蛋白是截短的生物活性部分，并且缺乏在野生型NiV-G蛋白(SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44)的N末端处或附近的多达40个连续氨基酸残基。

[0029] 在一些任何实施方案中，NiV-G蛋白是在野生型NiV-G的N末端截短的生物活性部分，并且具有SEQ ID NO:10-15、35-40或45-50中任一者所示的序列、或与SEQ ID NO:10-15、35-40或45-50具有至少为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、至少为或约84％、至少为或约85％、至少为或约86％、至少为或约87％、至少为或约88％、至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。

[0030] 在一些任何实施方案中，NiV-G蛋白是在野生型NiV-G蛋白(SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44)的N末端处或附近具有5个氨基酸截短的生物活性部分。在一些任何实施方案中，NiV-G蛋白或生物活性部分具有SEQ ID NO:10所示的氨基酸序列、或与SEQ ID NO:10具有至少为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。在一些任何实施方案中，NiV-G蛋白或生物活性部分具有SEQ ID NO:35所示的氨基酸序列、或与SEQ ID NO:35具有至少为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。在一些任何实施方案中，NiV-G蛋白或生物活性部分具有SEQ ID NO:45所示的氨基酸序列、或与SEQ ID NO:45具

有至少为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。

[0031] 在一些任何实施方案中，NiV-G蛋白是在野生型NiV-G蛋白(SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44)的N末端处或附近具有10个氨基酸截短的生物活性部分。在一些任何实施方案中，NiV-G蛋白或生物活性部分具有SEQ ID NO:36所示的氨基酸序列、或与SEQ ID NO:36具有至少为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。在一些任何实施方案中，NiV-G蛋白或生物活性部分具有SEQ ID NO:11所示的氨基酸序列、或与SEQ ID NO:11具有至少为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。在一些任何实施方案中，NiV-G蛋白或生物活性部分具有SEQ ID NO:46所示的氨基酸序列、或与SEQ ID NO:46具有至少为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。

[0032] 在一些任何实施方案中，NiV-G蛋白或生物活性部分具有在野生型NiV-G蛋白(SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44)的N末端处或附近的15个氨基酸截短。在一些任何实施方案中，NiV-G蛋白或生物活性部分具有SEQ ID NO:12所示的氨基酸序列、或与SEQ ID NO:12具有至少为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。在一些任何实施方案中，NiV-G蛋白或生物活性部分具有SEQ ID NO:37所示的氨基酸序列、或与SEQ ID NO:37具有至少为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。在一些任何实施方案中，NiV-G蛋白或生物活性部分具有SEQ ID NO:47所示的氨基酸序列、或与SEQ ID NO:47具有至少为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少

为或约85%、至少为或约86%、或至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的氨基酸序列。

[0033] 在一些任何实施方案中,NiV-G蛋白是在野生型NiV-G蛋白(SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44)的N末端处或附近具有20个氨基酸截短的生物活性部分。在一些任何实施方案中,NiV-G蛋白或生物活性部分具有SEQ ID NO:13所示的氨基酸序列、或与SEQ ID NO:13具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的氨基酸序列。在一些任何实施方案中,NiV-G蛋白或生物活性部分具有SEQ ID NO:38所示的氨基酸序列、或与SEQ ID NO:38具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的氨基酸序列。在一些任何实施方案中,NiV-G蛋白或生物活性部分具有SEQ ID NO:48所示的氨基酸序列、或与SEQ ID NO:48具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的氨基酸序列。

[0034] 在一些任何实施方案中,NiV-G蛋白是在野生型NiV-G蛋白(SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44)的N末端处或附近具有25个氨基酸截短的生物活性部分。在一些任何实施方案中,NiV-G蛋白具有SEQ ID NO:14所示的氨基酸序列、或与SEQ ID NO:14具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的氨基酸序列。在一些任何实施方案中,NiV-G蛋白或生物活性部分具有SEQ ID NO:39所示的氨基酸序列、或与SEQ ID NO:39具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的氨基酸序列。在一些任何实施方案中,NiV-G蛋白或生物活性部分具有SEQ ID NO:49所示的氨基酸序列、或与SEQ ID NO:49具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约

90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。

[0035] 在一些任何实施方案中，NiV-G蛋白是在野生型NiV-G蛋白(SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44)的N末端处或附近具有30个氨基酸截短的生物活性部分。在一些任何实施方案中，NiV-G蛋白或生物活性部分具有SEQ ID NO:15所示的氨基酸序列、或与SEQ ID NO:15具有至少为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。在一些任何实施方案中，NiV-G蛋白或生物活性部分具有SEQ ID NO:40所示的氨基酸序列、或与SEQ ID NO:40具有至少为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。

[0036] 在一些任何实施方案中，NiV-G蛋白是在野生型NiV-G蛋白(SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44)的N末端处或附近具有34个氨基酸截短的生物活性部分。在一些任何实施方案中，NiV-G蛋白或生物活性部分具有SEQ ID NO:22所示的序列、或与SEQ ID NO:22具有至少为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、至少为或约84％、至少为或约85％、至少为或约86％、至少为或约87％、至少为或约88％、至少为或约89％、为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。在一些任何实施方案中，NiV-G蛋白或生物活性部分具有SEQ ID NO:53所示的序列、或与SEQ ID NO:53具有至少为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、至少为或约84％、至少为或约85％、至少为或约86％、至少为或约87％、至少为或约88％、至少为或约89％、为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％的序列同一性的氨基酸序列。

[0037] 在一些任何实施方案中，所述G蛋白、其生物活性部分是功能活性变体，所述功能活性变体是表现出与肝配蛋白(Ephrin)B2或肝配蛋白B3的结合降低的突变体NiV-G蛋白。

[0038] 在一些任何实施方案中，突变体NiV-G蛋白包含一个或多个氨基酸取代，所述一个或多个氨基酸取代对应于选自由参考SEQ ID NO:28所示的编号的E501A、W504A、Q530A和E533A组成的组的氨基酸取代。在一些任何实施方案中，突变体NiV-G蛋白包含参考SEQ ID NO:28所示的编号的氨基酸取代E501A、W504A、Q530A和E533A。

[0039] 在一些任何实施方案中，突变体NiV-G蛋白或生物活性部分具有SEQ ID NO:16所示的氨基酸序列、或与SEQ ID NO:16具有至少为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、至少为或约84％、至少为或约85％、至少为或约86％、至少为或约87％、至少为或约88％、至少为或约89％、至少为或约90％、至少为或约91％、至少为或约

92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。在一些任何实施方案中，突变体NiV-G蛋白或生物活性部分具有SEQ ID NO:51所示的氨基酸序列、或与SEQ ID NO:51具有至少为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、至少为或约84％、至少为或约85％、至少为或约86％、至少为或约87％、至少为或约88％、至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。

[0040] 在一些任何实施方案中，F蛋白或其生物活性部分是野生型尼帕病毒F(NiV-F)蛋白或亨德拉病毒F蛋白或是其功能活性变体或生物活性部分。在一些任何实施方案中，所述F蛋白或其生物活性部分是野生型NiV-F蛋白或其功能活性变体或生物活性部分。在一些任何实施方案中，NiV-F蛋白或其功能活性变体或生物活性部分包含SEQ ID NO:2所示的氨基酸序列、或与SEQ ID NO:2具有至少为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、至少为或约84％、至少为或约85％、至少为或约86％、至少为或约87％、至少为或约88％、至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。

[0041] 在一些任何实施方案中，NiV-F蛋白是在野生型NiV-F蛋白(SEQ ID NO:2)的C末端处或附近具有20个氨基酸截短的其生物活性部分。

[0042] 在一些任何实施方案中，NiV-F蛋白或生物活性部分具有SEQ ID NO:5所示的序列、或与SEQ ID NO:5具有至少为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、至少为或约84％、至少为或约85％、至少为或约86％、至少为或约87％、至少为或约88％、至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。

[0043] 在一些任何实施方案中，NiV-F蛋白是其生物活性部分，所述生物活性部分包含：i)在野生型NiV-F蛋白(SEQ ID NO:2)的C末端处或附近的20个氨基酸截短；和ii)N-连接糖基化位点上的点突变。

[0044] 在一些任何实施方案中，NiV-F蛋白或生物活性部分具有SEQ ID NO:7所示的序列、或与SEQ ID NO:7具有至少为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、至少为或约84％、至少为或约85％、至少为或约86％、至少为或约87％、至少为或约88％、至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。

[0045] 在一些任何实施方案中，NiV-F蛋白是在野生型NiV-F蛋白(SEQ ID NO:2)的C末端处或附近具有22个氨基酸截短的其生物活性部分。

[0046] 在一些任何实施方案中，NiV-F蛋白或生物活性部分具有SEQ ID NO:8所示的序列、或由编码与SEQ ID NO:8具有至少为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、至少为或约84％、至少为或约85％、至少为或约86％、至少为或约87％、至少为

或约88%、至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的序列的核苷酸序列编码的氨基酸序列。

[0047] 在一些任何实施方案中,NiV-F蛋白或生物活性部分具有SEQ ID NO:23所示的序列、或与SEQ ID NO:23具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、至少为或约84%、至少为或约85%、至少为或约86%、至少为或约87%、至少为或约88%、至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的氨基酸序列。在一些任何实施方案中,F蛋白或其生物活性部分包含F1亚基或其促融合(fusogenic)部分。

[0048] 在一些任何实施方案中,F蛋白包含SEQ ID NO:23所示的序列,并且G蛋白包含SEQ ID NO:16所示的序列。

[0049] 在一些任何实施方案中,F蛋白由或基本上由SEQ ID NO:23所示的序列组成,和/或G蛋白由或基本上由SEQ ID NO:16所示的序列组成。

[0050] 在一些任何实施方案中,F1亚基是F0前体的蛋白水解切割部分。在一些任何实施方案中,F1亚基包含SEQ ID NO:4所示的氨基酸序列、或与SEQ ID NO:4具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、至少为或约84%、至少为或约85%、至少为或约86%、至少为或约87%、至少为或约88%、至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的氨基酸序列。

[0051] 在一些任何实施方案中,所述脂质双层来源于用于产生逆转录病毒或逆转录病毒样颗粒的宿主细胞的膜。在一些任何实施方案中,宿主细胞选自由以下组成的组:CHO细胞、BHK细胞、MDCK细胞、C3H 10T1/2细胞、FLY细胞、Psi-2细胞、BOSC 23细胞、PA317细胞、WEHI细胞、COS细胞、BSC 1细胞、BSC 40细胞、BMT 10细胞、VERO细胞、W138细胞、MRC5细胞、A549细胞、HT1080细胞、293细胞、293T细胞、B-50细胞、3T3细胞、NIH3T3细胞、HepG2细胞、Saos-2细胞、Huh7细胞、HeLa细胞、W163细胞、211细胞和211A细胞。在一些任何实施方案中,所述宿主细胞包含293T细胞。在一些任何实施方案中,所述脂质双层是或包含病毒包膜。在一些任何实施方案中,所述逆转录病毒样颗粒是复制缺陷的。

[0052] 在一些任何实施方案中,所述靶向脂质颗粒包含除F蛋白分子和G蛋白之外的一种或多种病毒组分。在一些任何实施方案中,所述一种或多种病毒组分来自逆转录病毒。在一些任何实施方案中,所述逆转录病毒是慢病毒。在一些任何实施方案中,所述一种或多种病毒组分包含选自Gag、Pol、Rev和Tat中的一种或多种的病毒包装蛋白。在一些任何实施方案中,所述一种或多种病毒组分包含以下核酸序列中的一种或多种(例如,全部):5'LTR(例如,包含U5并且缺乏功能性U3结构域)、Psi包装元件(Psi)、中央多嘌呤区(cPPT)/中央终止序列(CTS)(例如,DNA瓣)、聚A尾序列、转录后调控元件(例如,WPRE)、Rev反应元件(RRE)和3'LTR(例如,包含U5并且缺乏功能性U3)。

[0053] 在一些任何实施方案中,所述靶向脂质颗粒是慢病毒载体。

[0054] 在一些任何实施方案中,所述靶向脂质颗粒或所述慢病毒载体是复制缺陷的。

[0055] 在一些任何实施方案中,所述靶向脂质颗粒或所述慢病毒载体还包含外源药剂。在一些任何实施方案中,所述靶向脂质颗粒还包含外源药剂。在一些实施方案中,所述慢病毒载体还包含外源药剂。

[0056] 在一些任何实施方案中,所述外源药剂存在于内腔中。在一些任何实施方案中,所述外源药剂是蛋白质或核酸。在一些实施方案中,所述核酸是DNA或RNA。

[0057] 在一些任何实施方案中,所述外源药剂是编码用于递送至靶细胞的运载物的核酸。在一些任何实施方案中,所述外源药剂编码治疗剂或诊断剂。

[0058] 在一些任何实施方案中,所述外源药剂编码膜蛋白。在一些实施方案中,所述膜蛋白是靶向由疾病或病症表达或与疾病或病症相关的细胞的抗原受体。在一些实施方案中,所述膜蛋白是嵌合抗原受体(CAR)。在一些实施方案中,所述CAR包含(i)结合细胞外抗原(例如,CD19或BCMA)的细胞外抗原结合结构域,任选地其中所述细胞外抗原结合结构域是scFv,(ii)跨膜结构域和(iii)细胞内信号传导区,所述细胞内信号传导区包含CD3ζ信号传导结构域和任选地共刺激信号传导结构域(例如,4-1BB或CD28共刺激信号传导结构域)。在一些实施方案中,所述靶细胞是T细胞。在一些实施方案中,在所述靶细胞上的细胞表面分子是CD4或CD8。在一些实施方案中,所述结合结构域是结合CD4(例如人CD4)的scFv。在一些实施方案中,所述结合结构域是结合CD4(例如人CD4)的单结构域抗体。在一些实施方案中,所述结合结构域是结合CD8(例如人CD8)的scFv。在一些实施方案中,所述结合结构域是结合CD8(例如人CD8)的单结构域抗体。

[0059] 在一些任何实施方案中,所述外源药剂是包含用于校正遗传缺陷(任选地靶细胞中的遗传缺陷)的有效负载基因的核酸。在一些实施方案中,所述遗传缺陷与肝脏细胞或肝细胞相关。在一些实施方案中,所述靶细胞是肝细胞。在一些实施方案中,所述细胞表面分子是选自由ASGR1、ASGR2和TM4SF5组成的组的分子。在一些实施方案中,所述结合结构域是结合ASGR1(例如人ASGR1)的scFv。在一些实施方案中,所述结合结构域是结合ASGR1(例如人ASGR1)的单结构域抗体。在一些实施方案中,所述结合结构域是结合ASGR2(例如人ASGR2)的scFv。在一些实施方案中,所述结合结构域是结合ASGR2(例如人ASGR2)的单结构域抗体。在一些实施方案中,所述结合结构域是结合TM4SF5(例如人TM4SF5)的scFv。在一些实施方案中,所述结合结构域是结合TM4SF5(例如人TM4SF5)的单结构域抗体。

[0060] 在一些任何实施方案中,所述单结构域抗体结合存在于靶细胞上的细胞表面分子。在一些任何实施方案中,所述细胞表面分子是蛋白质、聚糖、脂质或低分子量分子。在一些任何实施方案中,所述靶细胞选自由以下组成的组:肿瘤浸润淋巴细胞、T细胞、赘生性或肿瘤细胞、病毒感染细胞、干细胞、中枢神经系统(CNS)细胞、造血干细胞(HSC)、肝细胞或完全分化细胞。在一些任何实施方案中,所述靶细胞选自由以下组成的组:CD3+T细胞、CD4+T细胞、CD8+T细胞、肝细胞、造血干细胞、CD34+造血干细胞、CD105+造血干细胞、CD117+造血干细胞、CD105+内皮细胞、B细胞、CD20+B细胞、CD19+B细胞、癌细胞、CD133+癌细胞、EpCAM+癌细胞、CD19+癌细胞、Her2/Neu+癌细胞、GluA2+神经元、GluA4+神经元、NKG2D+自然杀伤细胞、SLC1A3+星形胶质细胞、SLC7A10+脂肪细胞或CD30+肺上皮细胞。

[0061] 在一些任何实施方案中,所述单结构域抗体结合存在于靶细胞上的抗原或其部分。在一些任何实施方案中,所述细胞表面分子或抗原选自由以下组成的组:ASGR1、ASGR2和TM4SF5。在一些实施方案中,所述抗原或其部分是人ASGR1。在一些实施方案中,所述抗原

或其部分是人ASGR2。在一些实施方案中,所述抗原或其部分是人TM4SF5。

[0062] 本文提供了多核苷酸,所述多核苷酸包含编码(i)亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分和(ii)结合选自由ASGR1、ASGR2和TM4SF5组成的组的细胞表面分子的结合结构域的核酸序列。在一些实施方案中,所述细胞表面分子是人ASGR1。在一些实施方案中,所述细胞表面分子是人ASGR2。在一些实施方案中,所述细胞表面分子是人TM4SF5。在一些任何实施方案中,所述细胞表面分子或抗原是CD8或CD4。

[0063] 本文提供了核酸序列,所述核酸序列编码(i)亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分和(ii)结合选自由CD4和CD8组成的组的细胞表面分子的结合结构域的。在一些实施方案中,所述细胞表面分子是人CD4。在一些实施方案中,所述细胞表面分子是人CD8。在一些实施方案中,所述细胞表面分子或抗原是低密度脂蛋白受体(LDL-R)。在一些实施方案中,所述细胞表面分子或抗原是人LDL-R。

[0064] 本文提供了多核苷酸,所述多核苷酸包含编码(i)亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分和(ii)结合低密度脂蛋白受体(LDL-R)的结合结构域的核酸序列。在一些实施方案中,所述结合结构域结合人LDL-R。在一些任何实施方案中,所述结合结构域是单结构域抗体(sdAb)。在一些任何实施方案中,所述结合结构域是单链可变片段(scFv)。

[0065] 本文提供了多核苷酸,所述多核苷酸包含编码(i)亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分和(ii)单结构域抗体(sdAb)可变域的核酸序列,其中所述sdAb可变域与G蛋白或其生物活性部分的C末端连接。在一些任何实施方案中,所述多核苷酸还包含(iii)编码亨尼帕病毒F蛋白分子或其生物活性部分的核酸序列。

[0066] 在一些实施方案中,所述核酸序列是第一核酸序列,并且所述多核苷酸还包含编码亨尼帕病毒F蛋白分子或其生物活性部分的第二核酸序列。在一些实施方案中,所述多核苷酸包含IRES或在第一核酸序列与第二核酸序列之间的编码连接肽的序列。在一些实施方案中,所述连接肽是自切割肽或引起核糖体跳跃的肽,任选地T2A肽。

[0067] 在一些任何实施方案中,所述多核苷酸包含至少一个可操作地连接以控制所述核酸的表达的启动子。在一些任何实施方案中,所述启动子可操作地连接以控制第一核酸序列和第二核酸序列的表达。在一些任何实施方案中,所述启动子是组成型启动子。在一些任何实施方案中,所述启动子是诱导型启动子。

[0068] 在一些任何实施方案中,sdAb可变域通过编码的肽接头与G蛋白连接。在一些实施方案中,结合结构域通过编码的肽接头与G蛋白连接。在一些任何实施方案中,所述编码的肽接头包含多达25个氨基酸的长度。在一些任何实施方案中,所述编码的肽接头包含多达65个氨基酸的长度。在一些任何实施方案中,所述编码的肽接头包含从或从约2至65个氨基酸、2至60个氨基酸、2至56个氨基酸、2至52个氨基酸、2至48个氨基酸、2至44个氨基酸、2至40个氨基酸、2至36个氨基酸、2至32个氨基酸、2至28个氨基酸、2至24个氨基酸、2至20个氨基酸、2至18个氨基酸、2至14个氨基酸、2至12个氨基酸、2至10个氨基酸、2至8个氨基酸、2至6个氨基酸、6至65个氨基酸、6至60个氨基酸、6至56个氨基酸、6至52个氨基酸、6至48个氨基酸、6至44个氨基酸、6至40个氨基酸、6至36个氨基酸、6至32个氨基酸、6至28个氨基酸、6至24个氨基酸、6至20个氨基酸、6至18个氨基酸、6至14个氨基酸、6至12个氨基酸、6至10个氨基酸、6至8个氨基酸、8至65个氨基酸、8至60个氨基酸、8至56个氨基酸、8至52个氨基酸、8至

48个氨基酸、8至44个氨基酸、8至40个氨基酸、8至36个氨基酸、8至32个氨基酸、8至28个氨基酸、8至24个氨基酸、8至20个氨基酸、8至18个氨基酸、8至14个氨基酸、8至12个氨基酸、8至10个氨基酸、10至65个氨基酸、10至60个氨基酸、10至56个氨基酸、10至52个氨基酸、10至48个氨基酸、10至44个氨基酸、10至40个氨基酸、10至36个氨基酸、10至32个氨基酸、10至28个氨基酸、10至24个氨基酸、10至20个氨基酸、10至18个氨基酸、10至14个氨基酸、10至12个氨基酸、12至65个氨基酸、12至60个氨基酸、12至56个氨基酸、12至52个氨基酸、12至48个氨基酸、12至44个氨基酸、12至40个氨基酸、12至36个氨基酸、12至32个氨基酸、12至28个氨基酸、12至24个氨基酸、12至20个氨基酸、12至18个氨基酸、12至14个氨基酸、14至65个氨基酸、14至60个氨基酸、14至56个氨基酸、14至52个氨基酸、14至48个氨基酸、14至44个氨基酸、14至40个氨基酸、14至36个氨基酸、14至32个氨基酸、14至28个氨基酸、14至24个氨基酸、14至20个氨基酸、14至18个氨基酸、18至65个氨基酸、18至60个氨基酸、18至56个氨基酸、18至52个氨基酸、18至48个氨基酸、18至44个氨基酸、18至40个氨基酸、18至36个氨基酸、18至32个氨基酸、18至28个氨基酸、18至24个氨基酸、18至20个氨基酸、20至65个氨基酸、20至60个氨基酸、20至56个氨基酸、20至52个氨基酸、20至48个氨基酸、20至44个氨基酸、20至40个氨基酸、20至36个氨基酸、20至32个氨基酸、20至28个氨基酸、20至26个氨基酸、20至24个氨基酸、24至65个氨基酸、24至60个氨基酸、24至56个氨基酸、24至52个氨基酸、24至48个氨基酸、24至44个氨基酸、24至40个氨基酸、24至36个氨基酸、24至32个氨基酸、24至30个氨基酸、24至28个氨基酸、28至65个氨基酸、28至60个氨基酸、28至56个氨基酸、28至52个氨基酸、28至48个氨基酸、28至44个氨基酸、28至40个氨基酸、28至36个氨基酸、28至34个氨基酸、28至32个氨基酸、32至65个氨基酸、32至60个氨基酸、32至56个氨基酸、32至52个氨基酸、32至48个氨基酸、32至44个氨基酸、32至40个氨基酸、32至38个氨基酸、32至36个氨基酸、36至65个氨基酸、36至60个氨基酸、36至56个氨基酸、36至52个氨基酸、36至48个氨基酸、36至44个氨基酸、36至40个氨基酸、40至65个氨基酸、40至60个氨基酸、40至56个氨基酸、40至52个氨基酸、40至48个氨基酸、40至44个氨基酸、44至65个氨基酸、44至60个氨基酸、44至56个氨基酸、44至52个氨基酸、44至48个氨基酸、48至65个氨基酸、48至60个氨基酸、48至56个氨基酸、48至52个氨基酸、50至65个氨基酸、50至60个氨基酸、50至56个氨基酸、50至52个氨基酸、54至65个氨基酸、54至60个氨基酸、54至56个氨基酸、58至65个氨基酸、58至60个氨基酸、或60至65个氨基酸。

[0069] 在一些任何实施方案中，所述编码的肽接头包含长度为3个、4个、5个、6个、7个、8个、9个、10个、11个、12个、13个、14个、15个、16个、17个、18个、19,20个、21个、22个、23个、24个、25个、26个、27个、28个、29个、30个、31个、32个、33个、34个、35个、36个、37个、38个、39个、40个、41个、42个、43个、44个、45个、46个、47个、48个、49个、50个、51个、52个、53个、54个、55个、56个、57个、58个、59个、60个、61个、62个、63个、64个或65个氨基酸的多肽。在一些任何实施方案中，所述编码的肽接头包含GS、GGS、GGGGS(SEQ ID NO:43)、GGGGGS(SEQ ID NO:41)以及它们的组合。在一些任何实施方案中，所述编码的肽接头包含(GGS)n，其中n为1至10。在一些任何实施方案中，所述编码的肽接头包含(GGGGS)n(SEQ ID NO:42)，其中n为1至10。在一些任何实施方案中，所述编码的肽接头包含(GGGGGS)n(SEQ ID NO:27)，其中n为1至4。在一些任何实施方案中，编码G蛋白的序列是野生型尼帕病毒G(NiV-G)蛋白或亨德拉病毒G蛋白，或是其功能活性变体或生物活性部分。在一些实施方案中，所述变体是表现出

对天然结合配偶体的结合降低的其变体。在一些任何实施方案中，编码G蛋白的核酸序列是野生型尼帕病毒G(NiV-G)蛋白或亨德拉病毒G蛋白，或是表现出与天然结合配偶体的结合降低的其变体。在一些实施方案中，编码的G蛋白是野生型NiV-G蛋白或其功能活性变体或生物活性部分。在一些任何实施方案中，编码G蛋白的核酸序列是野生型NiV-G蛋白。在一些任何实施方案中，编码G蛋白的核酸序列是表现出与肝配蛋白B2或肝配蛋白B3的结合降低的突变体NiV-G蛋白。

[0070] 在一些任何实施方案中，NiV-G蛋白或其功能活性变体或生物活性部分包含SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44所示的氨基酸序列，或包含与SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44具有至少为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。在一些任何实施方案中，NiV-G蛋白是截短的生物活性部分，并且缺乏在野生型NiV-G蛋白(SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44)的N末端处或附近的多达40个连续氨基酸残基。在一些任何实施方案中，NiV-G蛋白是在野生型NiV-G的N末端截短的生物活性部分，并且包含SEQ ID NO:10-15、35-40或45-50中任一者所示的序列、或与SEQ ID NO:10-15、35-40或45-50具有至少为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。

[0071] 在一些任何实施方案中，NiV-G蛋白是在野生型NiV-G蛋白(SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44)的N末端处或附近包含5个氨基酸截短的生物活性部分。在一些任何实施方案中，NiV-G蛋白或生物活性部分包含SEQ ID NO:10所示的氨基酸序列、或与SEQ ID NO:10具有至少为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。在一些任何实施方案中，NiV-G蛋白或生物活性部分包含SEQ ID NO:35所示的氨基酸序列、或与SEQ ID NO:35具有至少为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。在一些任何实施方案中，NiV-G蛋白或生物活性部分包含SEQ ID NO:45所示的氨基酸序列、或与SEQ ID NO:45具有至少为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一

性的氨基酸序列。

[0072] 在一些任何实施方案中,NiV-G蛋白是在野生型NiV-G蛋白(SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44)的N末端处或附近包含10个氨基酸截短的生物活性部分。在一些任何实施方案中,突变体NiV-G蛋白或生物活性部分包含SEQ ID NO:11所示的氨基酸序列、或与SEQ ID NO:11具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的氨基酸序列。在一些任何实施方案中,NiV-G蛋白或生物活性部分包含SEQ ID NO:36所示的氨基酸序列、或与SEQ ID NO:36具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的氨基酸序列。在一些任何实施方案中,NiV-G蛋白或生物活性部分包含SEQ ID NO:46所示的氨基酸序列、或与SEQ ID NO:46具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的氨基酸序列。

[0073] 在一些任何实施方案中,NiV-G蛋白是在野生型NiV-G蛋白(SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44)的N末端处或附近包含15个氨基酸截短的生物活性部分。在一些任何实施方案中,NiV-G蛋白或生物活性部分包含SEQ ID NO:12所示的氨基酸序列、或与SEQ ID NO:12具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的氨基酸序列。在一些任何实施方案中,NiV-G蛋白或生物活性部分包含SEQ ID NO:37所示的氨基酸序列、或与SEQ ID NO:37具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的氨基酸序列。在一些任何实施方案中,NiV-G蛋白或生物活性部分包含SEQ ID NO:47所示的氨基酸序列、或与SEQ ID NO:47具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的氨基酸序列。

[0074] 在一些任何实施方案中,NiV-G蛋白是在野生型NiV-G蛋白(SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44)的N末端处或附近包含20个氨基酸截短的生物活性部分。在一些任何实施方案中,NiV-G蛋白或生物活性部分包含SEQ ID NO:13所示的氨基酸序列、或与SEQ ID NO:13具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的氨基酸序列。在一些任何实施方案中,NiV-G蛋白或生物活性部分包含SEQ ID NO:38所示的氨基酸序列、或与SEQ ID NO:38具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的氨基酸序列。在一些任何实施方案中,NiV-G蛋白或生物活性部分包含SEQ ID NO:48所示的氨基酸序列、或与SEQ ID NO:48具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的氨基酸序列。

[0075] 在一些任何实施方案中,NiV-G蛋白是在野生型NiV-G蛋白(SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44)的N末端处或附近包含25个氨基酸截短的生物活性部分。在一些任何实施方案中,NiV-G蛋白或生物活性部分包含SEQ ID NO:14所示的氨基酸序列、或与SEQ ID NO:14具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的氨基酸序列。在一些任何实施方案中,NiV-G蛋白或生物活性部分包含SEQ ID NO:39所示的氨基酸序列、或与SEQ ID NO:39具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的氨基酸序列。在一些任何实施方案中,NiV-G蛋白或生物活性部分包含SEQ ID NO:49所示的氨基酸序列、或与SEQ ID NO:49具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的氨基酸序列。

[0076] 在一些任何实施方案中,NiV-G蛋白是在野生型NiV-G蛋白(SEQ ID NO:9、SEQ ID

NO:28或SEQ ID NO:44)的N末端处或附近包含30个氨基酸截短的生物活性部分。在一些任何实施方案中,NiV-G蛋白或生物活性部分包含SEQ ID NO:15所示的氨基酸序列、或与SEQ ID NO:15具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的氨基酸序列。在一些任何实施方案中,NiV-G蛋白或生物活性部分包含SEQ ID NO:40所示的氨基酸序列、或与SEQ ID NO:40具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的氨基酸序列。在一些任何实施方案中,NiV-G蛋白或生物活性部分包含SEQ ID NO:50所示的氨基酸序列、或与SEQ ID NO:50具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的氨基酸序列。

[0077] 在一些任何实施方案中,NiV-G蛋白是在野生型NiV-G蛋白(SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44)的N末端处或附近具有34个氨基酸截短的生物活性部分。在一些任何实施方案中,NiV-G蛋白或生物活性部分具有SEQ ID NO:22所示的氨基酸序列、或与SEQ ID NO:22具有为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的氨基酸序列。在一些任何实施方案中,NiV-G蛋白或生物活性部分具有SEQ ID NO:53所示的氨基酸序列、或与SEQ ID NO:53具有为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的氨基酸序列。

[0078] 在一些任何实施方案中,G蛋白是表现出与肝配蛋白B2或肝配蛋白B3的结合降低的突变体NiV-G蛋白。在一些任何实施方案中,所述突变体NiV-G蛋白包含:一个或多个氨基酸取代,所述一个或多个氨基酸取代对应于选自由参考SEQ ID NO:28所示的编号的E501A、W504A、Q530A和E533A组成的组的氨基酸取代。在一些任何实施方案中,所述突变体NiV-G蛋白包含参考SEQ ID NO:28所示的编号的氨基酸取代E501A、W504A、Q530A和E533A。

[0079] 在一些任何实施方案中,所述突变体NiV-G蛋白包含:i)在N末端处或附近的截短;和ii)选自由E501A、W504A、Q530A和E533A组成的组的点突变。在一些任何实施方案中,所述突变体NiV-G蛋白包含SEQ ID NO:16所示的氨基酸序列、或与SEQ ID NO:16具有至少为或

约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。在一些任何实施方案中，所述突变体NiV-G蛋白包含SEQ ID NO:51所示的氨基酸序列、或与SEQ ID NO:51具有至少为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。

[0080] 在一些任何实施方案中，F蛋白或其生物活性部分是野生型尼帕病毒F(NiV-F)蛋白或亨德拉病毒F蛋白或是其功能活性变体或生物活性部分。在一些任何实施方案中，F蛋白或其生物活性部分是野生型NiV-F蛋白或其功能活性变体或生物活性部分。在一些任何实施方案中，NiV-F蛋白或其功能活性变体或生物活性部分包含SEQ ID NO:2所示的氨基酸序列、或与SEQ ID NO:2具有为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。

[0081] 在一些任何实施方案中，NiV-F蛋白是在野生型NiV-F蛋白(SEQ ID NO:2)的C末端处或附近具有20个氨基酸截短的其生物活性部分。在一些任何实施方案中，NiV-F蛋白或生物活性部分具有SEQ ID NO:5所示的序列或与SEQ ID NO:5具有为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。在一些任何实施方案中，NiV-F蛋白是包含以下的其生物活性部分：i)在野生型NiV-F蛋白(SEQ ID NO:2)的C末端处或附近的20个氨基酸截短；和ii)在N-连接的糖基化位点上的点突变。

[0082] 在一些任何实施方案中，NiV-F蛋白或生物活性部分具有SEQ ID NO:7所示的序列、或与SEQ ID NO:7具有为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。

[0083] 在一些任何实施方案中，NiV-F蛋白是在野生型NiV-F蛋白(SEQ ID NO:2)的C末端处或附近具有22个氨基酸截短的其生物活性部分。在一些任何实施方案中，NiV-F蛋白或生物活性部分具有SEQ ID NO:8所示的序列、或由编码与SEQ ID NO:8具有为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或

约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的序列的核苷酸序列编码的氨基酸序列。

[0084] 在一些任何实施方案中，NiV-F蛋白具有SEQ ID NO:23所示的序列、或与SEQ ID NO:23具有为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。在一些任何实施方案中，F蛋白包含SEQ ID NO:23所示的序列，并且G蛋白包含SEQ ID NO:16所示的序列。在一些任何实施方案中，F蛋白由或基本上由SEQ ID NO:23所示的序列组成，并且G蛋白由或基本上由SEQ ID NO:16所示的序列组成。

[0085] 本文提供了载体，所述载体包含本文所述的任何实施方案的多核苷酸。在一些任何实施方案中，所述载体是哺乳动物载体、病毒载体或人工染色体，任选地其中所述人工染色体是细菌人工染色体(BAC)。

[0086] 本文提供了质粒，所述质粒包含本文所述的任何实施方案的多核苷酸。在一些任何实施方案中，所述质粒还包含编码用于慢病毒产生的蛋白质的一种或多种核酸。

[0087] 本文提供了细胞，所述细胞包含本文所述的任何实施方案的多核苷酸或本文所述的任何实施方案的载体或本文所述的任何实施方案的质粒。

[0088] 本文提供了制备靶向脂质颗粒的方法，所述靶向脂质颗粒包含亨尼帕病毒F蛋白分子或其生物活性部分和靶向包膜蛋白，所述靶向包膜蛋白包含亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分和单结构域抗体(sdAb)可变域，所述方法包括：a)提供包含编码亨尼帕病毒F蛋白分子或其生物活性部分的核酸和编码靶向包膜蛋白的核酸的细胞，所述靶向包膜蛋白包含亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分和单结构域抗体(sdAb)可变域；b)在允许产生靶向脂质颗粒的条件下培养所述细胞，和c)从所述细胞分离、富集或纯化所述靶向脂质颗粒，从而制得所述靶向脂质颗粒。

[0089] 本文提供了制备假型化慢病毒载体的方法，所述方法包括：a)提供包含慢病毒病毒核酸、编码亨尼帕病毒F蛋白分子或其生物活性部分的核酸和编码靶向包膜蛋白的核酸的生产者细胞，所述靶向包膜蛋白包含亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分和单结构域抗体；b)在允许慢病毒载体产生的条件下培养细胞，和c)从所述细胞分离、富集或纯化慢病毒载体，从而制得所述假型化慢病毒载体。

[0090] 在一些任何实施方案中，所述单结构域抗体结合存在于靶细胞上的细胞表面分子。在一些任何实施方案中，所述细胞表面分子是蛋白质、聚糖、脂质或低分子量分子。在一些任何实施方案中，所述靶细胞选自由以下组成的组：肿瘤浸润淋巴细胞、T细胞、赘生性或肿瘤细胞、病毒感染细胞、干细胞、中枢神经系统(CNS)细胞、造血干细胞(HSC)、肝细胞或完全分化细胞。在一些任何实施方案中，所述靶细胞选自由以下组成的组：CD3+T细胞、CD4+T细胞、CD8+T细胞、肝细胞、造血干细胞、CD34+造血干细胞、CD105+造血干细胞、CD117+造血干细胞、CD105+内皮细胞、B细胞、CD20+B细胞、CD19+B细胞、癌细胞、CD133+癌细胞、EpCAM+癌细胞、CD19+癌细胞、Her2/Neu+癌细胞、GluA2+神经元、GluA4+神经元、NKG2D+自然杀伤细胞、SLC1A3+星形胶质细胞、SLC7A10+脂肪细胞或CD30+肺上皮细胞。在一些任何实施方案

中,所述单结构域抗体结合存在于靶细胞上的抗原或其部分。

[0091] 本文提供了制备靶向脂质颗粒的方法,所述靶向脂质颗粒包含亨尼帕病毒F蛋白分子或其生物活性部分和靶向包膜蛋白,所述靶向包膜蛋白包含亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分和结合结构域,所述方法包括:a)提供包含编码亨尼帕病毒F蛋白分子或其生物活性部分的核酸和编码靶向包膜蛋白的核酸的细胞,所述靶向包膜蛋白包含亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分和结合结构域,其中所述结合结构域(i)结合选自由以下组成的组的细胞表面分子:ASGR1、ASGR2和TM4SF5,任选地人ASGR1、人ASGR2和人ASGR2;(ii)结合选自由以下组成的组的细胞表面分子:CD4或CD8,任选地人CD4或人CD8;或(iii)结合作为低密度脂蛋白受体(LDL-R)、任选地人LDL-R的细胞表面分子;b)在允许产生靶向脂质颗粒的条件下培养所述细胞,和c)从所述细胞分离、富集或纯化所述靶向脂质颗粒,从而制得所述靶向脂质颗粒。

[0092] 本文提供了制备假型化慢病毒载体的方法,所述方法包括a)提供包含慢病毒病毒核酸、编码亨尼帕病毒F蛋白分子或其生物活性部分的核酸和编码靶向包膜蛋白的核酸的生产者细胞,所述靶向包膜蛋白包含亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分和结合结构域,其中所述结合结构域(i)结合选自由以下组成的组的细胞表面分子:ASGR1、ASGR2和TM4SF5,任选地人ASGR1、人ASGR2和人ASGR2;(ii)结合选自由以下组成的组的细胞表面分子:CD4或CD8,任选地人CD4或人CD8;或(iii)结合作为低密度脂蛋白受体(LDL-R)、任选地人LDL-R的细胞表面分子;b)在允许慢病毒载体产生的条件下培养生产者细胞,和c)从所述细胞分离、富集或纯化慢病毒载体,从而制得假型化慢病毒载体。

[0093] 在一些任何实施方案中,所述结合结构域是单结构域抗体。在一些任何实施方案中,所述结合结构域是单链可变片段(scFv)。在一些任何实施方案中,所述细胞表面分子选自由以下组成的组:ASGR1、ASGR2和TM4SF5。在一些任何实施方案中,所述细胞表面分子是CD8或CD4,在一些任何实施方案中,所述细胞表面分子是LDL-R。

[0094] 本文提供了制备包含亨尼帕病毒F蛋白分子或其生物活性部分和靶向包膜蛋白的靶向脂质颗粒的方法,所述方法包括:a)提供包含本文提供的任何实施方案的多核苷酸、本文所述的任何实施方案的载体或本文所述的任何实施方案的质粒的细胞;b)在允许产生靶向脂质颗粒的条件下培养所述细胞,和c)从所述细胞分离、富集或纯化所述靶向脂质颗粒,从而制得所述靶向脂质颗粒。

[0095] 本文提供了制备假型化慢病毒载体的方法,所述方法包括:a)提供生产者细胞,所述生产者细胞包含慢病毒病毒核酸和本文列出的任何实施方案的多核苷酸或本文列出的任何实施方案的载体;b)在允许慢病毒载体产生的条件下培养所述细胞;和c)从所述细胞分离、富集或纯化慢病毒载体,从而制得假型化慢病毒载体。在一些任何实施方案中,在步骤(b)之前,所述方法还包括向所述细胞提供编码亨尼帕病毒F蛋白分子或其生物活性部分的多核苷酸。

[0096] 在一些任何实施方案中,所述细胞是哺乳动物细胞。

[0097] 在一些任何实施方案中,所述细胞是包含病毒核酸的生产者细胞。在一些任何实施方案中,所述病毒核酸是逆转录病毒核酸或慢病毒核酸,并且所述靶向脂质颗粒是病毒颗粒或病毒样颗粒。在一些任何实施方案中,所述病毒颗粒或病毒样颗粒是逆转录病毒颗粒或逆转录病毒样颗粒。在一些实施方案中,所述病毒颗粒或病毒样颗粒是慢病毒颗粒或

慢病毒样颗粒。

[0098] 在一些任何实施方案中,所述病毒核酸缺乏参与病毒复制的一或多个基因。在一些任何实施方案中,所述病毒核酸包含编码选自Gag、Pol、Rev和Tat中的一种或多种的病毒包装蛋白的核酸。在一些任何实施方案中,所述病毒核酸包含:以下核酸序列中的一种或多种(例如,全部):5'LTR(例如,包含U5并且缺乏功能性U3结构域)、Psi包装元件(Psi)、中央多嘌呤区(cPPT)/中央终止序列(CTS)(例如,DNA瓣)、聚A尾序列、转录后调控元件(例如,WPRE)、Rev反应元件(RRE)和3'LTR(例如,包含U5并且缺乏功能性U3)。

[0099] 本文提供了生产者细胞,所述生产者细胞包含本文列出的任何实施方案的多核苷酸或本文列出的任何实施方案的载体或本文所述的任何实施方案的质粒。

[0100] 在一些任何实施方案中,所述生产者细胞还包含编码亨尼帕病毒F蛋白或其生物活性部分的核酸。

[0101] 在一些任何实施方案中,所述细胞还包含病毒核酸。在一些任何实施方案中,所述病毒核酸是慢病毒核酸。本文提供了生产者细胞,所述生产者细胞包含(i)病毒核酸和(ii)编码亨尼帕病毒F蛋白分子或其生物活性部分的核酸和(iii)编码靶向包膜蛋白的核酸,所述靶向包膜蛋白包含亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分和单结构域抗体(sdAb)可变域,任选地其中所述病毒核酸是慢病毒核酸。在一些任何实施方案中,所述单结构域抗体结合存在于靶细胞上的细胞表面分子。在一些任何实施方案中,所述细胞表面分子是蛋白质、聚糖、脂质或低分子量分子。

[0102] 在一些任何实施方案中,所述靶细胞选自由以下组成的组:肿瘤浸润淋巴细胞、T细胞、赘生性或肿瘤细胞、病毒感染细胞、干细胞、中枢神经系统(CNS)细胞、造血干细胞(HSC)、肝细胞或完全分化细胞。在一些任何实施方案中,所述靶细胞选自由以下组成的组:CD3+T细胞、CD4+T细胞、CD8+T细胞、肝细胞、造血干细胞、CD34+造血干细胞、CD105+造血干细胞、CD117+造血干细胞、CD105+内皮细胞、B细胞、CD20+B细胞、CD19+B细胞、癌细胞、CD133+癌细胞、EpCAM+癌细胞、CD19+癌细胞、Her2/Neu+癌细胞、GluA2+神经元、GluA4+神经元、NKG2D+自然杀伤细胞、SLC1A3+星形胶质细胞、SLC7A10+脂肪细胞或CD30+肺上皮细胞。在一些任何实施方案中,所述单结构域抗体结合存在于靶细胞上的抗原或其部分。

[0103] 本文提供了生产者细胞,所述生产者细胞包含(i)病毒核酸和(ii)编码亨尼帕病毒F蛋白分子或其生物活性部分的核酸和(iii)编码靶向包膜蛋白的核酸,所述靶向包膜蛋白包含亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分和结合结构域,其中所述结合结构域(i)结合选自由以下组成的组的细胞表面分子:ASGR1、ASGR2和TM4SF5,任选地人ASGR1、人ASGR2和人ASGR2;(ii)结合选自由以下组成的组的细胞表面分子:CD4或CD8,任选地人CD4或人CD8;或(iii)结合作为低密度脂蛋白受体(LDL-R)、任选地人LDL-R的细胞表面分子。在一些任何实施方案中,所述病毒核酸是慢病毒核酸。

[0104] 在一些任何实施方案中,所述细胞表面分子或抗原选自ASGR1、ASGR2和TM4SF5。在一些任何实施方案中,所述细胞表面分子或抗原是CD8或CD4。在一些任何实施方案中,所述细胞表面分子或抗原是LDL-R。

[0105] 在一些任何实施方案中,所述病毒核酸缺乏一个或多个参与病毒复制的基因。在一些任何实施方案中,所述病毒核酸包含编码选自Gag、Pol、Rev和Tat中的一种或多种的病毒包装蛋白的核酸。

[0106] 在一些任何实施方案中,所述病毒核酸包含以下核酸序列中的一种或多种(例如,全部):5'LTR(例如,包含U5并且缺乏功能性U3结构域)、Psi包装元件(Psi)、中央多嘌呤区(cPPT)/中央终止序列(CTS)(例如,DNA瓣)、聚A尾序列、转录后调控元件(例如,WPRE)、Rev反应元件(RRE)和3'LTR(例如,包含U5并且缺乏功能性U3)。

[0107] 在一些任何实施方案中,所述亨尼帕病毒F蛋白分子或其生物活性部分包含:(i)SEQ ID NO:2所示的序列;(ii)与SEQ ID NO:2具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、至少为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的氨基酸序列。在一些任何实施方案中,所述亨尼帕病毒F蛋白分子或其生物活性部分包含(i)SEQ ID NO:5所示的序列;(ii)与SEQ ID NO:5具有至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、至少为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的氨基酸序列。

[0108] 在一些任何实施方案中,所述亨尼帕病毒F蛋白分子或其生物活性部分包含(i)SEQ ID NO:7所示的序列;(ii)与SEQ ID NO:7具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、至少为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的氨基酸序列。在一些任何实施方案中,所述亨尼帕病毒F蛋白分子或其生物活性部分包含(i)由编码SEQ ID NO:8所示的序列的核苷酸序列编码的序列;(ii)由编码与SEQ ID NO:8具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、至少为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的序列的核苷酸序列编码的氨基酸序列。

[0109] 在一些任何实施方案中,所述亨尼帕病毒F蛋白分子或其生物活性部分包含(i)SEQ ID NO:23所示的序列;(ii)与SEQ ID NO:23具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、至少为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的氨基酸序列。

[0110] 在一些任何实施方案中,所述亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分包含:(i)SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44所示的序列;(ii)与SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、至少为或约96%、至少为或约97%、至少为

或约98%、或至少为或约99%序列同一性的氨基酸序列。

[0111] 在一些任何实施方案中,所述亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分包含:(i)SEQ ID NO:10所示的序列;(ii)与SEQ ID NO:10具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、至少为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的氨基酸序列。

[0112] 在一些任何实施方案中,所述亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分包含:(i)SEQ ID NO:35所示的序列;(ii)与SEQ ID NO:35具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、至少为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的氨基酸序列。

[0113] 在一些任何实施方案中,所述亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分包含:(i)SEQ ID NO:45所示的序列;(ii)与SEQ ID NO:45具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、至少为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的氨基酸序列。

[0114] 在一些任何实施方案中,所述亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分包含:(i)SEQ ID NO:11所示的序列;(ii)与SEQ ID NO:11具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、至少为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的氨基酸序列。

[0115] 在一些任何实施方案中,所述亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分包含:(i)SEQ ID NO:36所示的序列;(ii)与SEQ ID NO:36具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、至少为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的氨基酸序列。

[0116] 在一些任何实施方案中,所述亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分包含:(i)SEQ ID NO:46所示的序列;(ii)与SEQ ID NO:46具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、至少为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的氨基酸序列。

[0117] 在一些任何实施方案中,所述亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分包含:(i)SEQ ID NO:12所示的序列;(ii)与SEQ ID NO:12具有至少为或约80%、至少

为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、至少为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。

[0118] 在一些任何实施方案中，所述亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分包含：(i)SEQ ID NO:37所示的序列；(ii)与SEQ ID NO:37具有至少为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、至少为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。

[0119] 在一些任何实施方案中，所述亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分包含：(i)SEQ ID NO:47所示的序列；(ii)与SEQ ID NO:47具有至少为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、至少为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。

[0120] 在一些任何实施方案中，所述亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分包含(i)SEQ ID NO:13所示的序列；(ii)与SEQ ID NO:13具有至少为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、至少为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。

[0121] 在一些任何实施方案中，所述亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分包含：(i)SEQ ID NO:38所示的序列；(ii)与SEQ ID NO:38具有至少为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、至少为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。

[0122] 在一些任何实施方案中，所述亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分包含：(i)SEQ ID NO:48所示的序列；(ii)与SEQ ID NO:48具有至少为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、至少为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。

[0123] 在一些任何实施方案中，所述亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分包含(i)SEQ ID NO:14所示的序列；(ii)与SEQ ID NO:14具有至少为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、至少为或约

96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的氨基酸序列。

[0124] 在一些任何实施方案中,所述亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分包含:(i)SEQ ID NO:39所示的序列;(ii)与SEQ ID NO:39具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、至少为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的氨基酸序列。

[0125] 在一些任何实施方案中,所述亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分包含:(i)SEQ ID NO:49所示的序列;(ii)与SEQ ID NO:49具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、至少为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的氨基酸序列。

[0126] 在一些任何实施方案中,所述亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分包含(i)SEQ ID NO:15所示的序列;(ii)与SEQ ID NO:15具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、至少为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的氨基酸序列。

[0127] 在一些任何实施方案中,所述亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分包含:(i)SEQ ID NO:40所示的序列;(ii)与SEQ ID NO:40具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、至少为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的氨基酸序列。

[0128] 在一些任何实施方案中,所述亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分包含:(i)SEQ ID NO:50所示的序列;(ii)与SEQ ID NO:50具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、至少为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的氨基酸序列。

[0129] 在一些任何实施方案中,所述亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分包含(i)SEQ ID NO:16所示的序列;(ii)与SEQ ID NO:16具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、至少为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的氨基酸序列。

[0130] 在一些任何实施方案中,所述亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分包含(i)SEQ ID NO:51所示的序列;(ii)与SEQ ID NO:51具有至少为或约80%、至少为

或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、至少为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。

[0131] 在所提供的实施方案的一些方面中，与已将相同包膜蛋白掺入相似脂质双层中但与替代靶向部分融合的参考脂质颗粒相比，靶向脂质颗粒具有更多的靶向包膜蛋白表达，任选地其中所述替代靶向部分是单链可变片段(scFv)。在一些任何实施方案中，所述表达增加了为或大于5％、10％、20％、30％、40％、50％、60％、70％、80％、90％、100％、125％、150％、200％、300％、400％、500％或更多。在一些实施方案中，所述表达增加了为或大于1.5倍、2倍、3倍、4倍、5倍、6倍、7倍、8倍、9倍、10倍、15倍、20倍、30倍或更多，优选地为或约或大于10倍或更多。在一些任何实施方案中，转导后在靶细胞中的滴度为或大于$1x10^6$个转导单位(TU)/mL、为或大于$2x10^6$TU/mL、为或大于$3x10^6$TU/mL、为或大于$4x10^6$TU/mL、为或大于$5x10^6$TU/mL、为或大于$6x10^6$TU/mL、为或大于$7x10^6$TU/mL、为或大于$8x10^6$TU/mL、为或大于$9x10^6$TU/mL、或为或大于$1x10^7$TU/mL。本文还提供了一种组合物，其中在脂质颗粒群体中，大于为或约50％、大于为或约55％、大于为或约60％、大于为或约65％、大于为或约70％、或大于为或约75％对靶向包膜蛋白呈表面阳性。在一些任何实施方案中，靶向包膜蛋白以至少约(0.001、0.002、0.005、0.01、0.02、0.05、0.1、0.2或0.5)个靶向包膜蛋白/nm^2的密度存在于靶向脂质颗粒的表面上。

[0132] 本文提供了从本文提供的任何实施方案的生产者细胞产生的病毒载体颗粒或病毒样颗粒。

[0133] 本文提供了包含本文提供的任何实施方案的多个靶向脂质颗粒的组合物。在一些实施方案中，所述组合物还包含药学上可接受的载体。在一些任何实施方案中，靶向脂质颗粒包含小于1μm的平均直径。在一些任何实施方案中，所述组合物还包含以至少约(0.001、0.002、0.005、0.01、0.02、0.05、0.1、0.2或0.5)个靶向包膜蛋白/nm^2的平均密度存在于靶向脂质颗粒表面上的靶向包膜蛋白。

[0134] 本文提供了与已将相同的包膜蛋白掺入其膜(例如质膜)中但与替代靶向部分融合的参考生产者细胞相比含有更多的靶向包膜蛋白的膜(例如质膜)表达的生产者细胞，任选地其中所述替代靶向部分是单链可变片段(scFv)。在一些实施方案中，所述表达增加了为或大于5％、10％、20％、30％、40％、50％、60％、70％、80％、90％、100％、125％、150％、200％、300％、400％、500％或更多。在一些实施方案中，所述表达增加了为或大于1.5倍、2倍、3倍、4倍、5倍、6倍、7倍、8倍、9倍、10倍、15倍、20倍、30倍或更多，优选地为或约或大于10倍或更多。在一些实施方案中，所述生产者细胞具有每平方微米至少20个蛋白质(例如，至少50个、100个、200个、500个、1000个、2000个、5000个或10,000个蛋白质)的所述靶向包膜蛋白在所述生产者细胞的膜(例如，质膜)上的表达。在一些任何实施方案中，靶向包膜蛋白包含生产者细胞的总膜(例如，质膜)蛋白的至少0.1％(例如，至少0.2％、0.5％、1％、2％、3％、4％、5％、6％、7％、8％、9％或10％)(例如，按总蛋白重量计)。

[0135] 本文提供了转导细胞的方法，所述方法包括用本文所述的任何病毒载体或用本文所述的任何组合物转导细胞。在一些任何实施方案中，慢病毒载体或靶向脂质颗粒的靶向包膜蛋白靶向CD4，并且细胞是CD4+细胞。在一些任何实施方案中，慢病毒载体的靶向包膜

蛋白靶向CD8，并且细胞是CD8+细胞。在一些任何实施方案中，慢病毒载体的靶向包膜蛋白靶向ASGR1、ASGR2或TM4SF5，并且细胞是肝细胞。

[0136] 本文提供了向受试者(例如，人类受试者)递送外源药剂的方法，所述方法包括向所述受试者施用本文提供的任何实施方案的靶向脂质颗粒或本文提供的任何实施方案的组合物，其中所述靶向脂质颗粒或慢病毒载体包含所述外源药剂。

[0137] 本文提供了向受试者(例如，人类受试者)递送外源药剂的方法，所述方法包括向所述受试者施用本文所述的任何组合物，其中所述多种的靶向脂质颗粒或慢病毒载体包含所述外源药剂。

[0138] 本文提供了将嵌合抗原受体(CAR)递送至细胞的方法，所述方法包括使细胞与本文所述的慢病毒载体或本文所述的任何实施方案的靶向脂质颗粒接触，其中所述慢病毒载体或靶向脂质颗粒包含编码CAR的核酸。

[0139] 本文提供了将嵌合抗原受体(CAR)递送至细胞的方法，所述方法包括使细胞与本文所述的任何组合物接触，其中所述多种的慢病毒载体或靶向脂质颗粒包含编码CAR的核酸。

[0140] 本文提供了将外源药剂递送至肝细胞的方法，所述方法包括使细胞与本文所述的任何慢病毒载体或本文所述的任何实施方案的靶向脂质颗粒或慢病毒载体接触。

[0141] 本文提供了将外源药剂递送至肝细胞的方法，所述方法包括使细胞与本文所述的任何组合物接触，其中所述多种的慢病毒载体或靶向脂质颗粒包含用于递送至肝细胞的外源药剂。在一些任何实施方案中，所述接触用慢病毒载体或靶向脂质颗粒转导细胞。

[0142] 本文提供了治疗受试者(例如，人类受试者)的疾病或病况的方法，所述方法包括向所述受试者施用本文提供的任何实施方案的靶向脂质颗粒或本文提供的任何实施方案的组合物。

[0143] 本文提供了将哺乳动物细胞与靶向脂质颗粒融合的方法，所述方法包括向所述受试者施用本文提供的任何实施方案的靶向脂质颗粒或本文提供的任何实施方案的组合物。在一些任何实施方案中，哺乳动物细胞与靶向脂质颗粒的融合将所述外源药剂递送至受试者(例如，人类受试者)。在一些任何实施方案中，哺乳动物细胞与靶向脂质颗粒的融合治疗受试者(例如，人类受试者)的疾病或病况。在一些任何实施方案中，慢病毒载体或靶向脂质颗粒的靶向包膜蛋白靶向CD4并且细胞是CD4+细胞。在一些任何实施方案中，慢病毒载体的靶向包膜蛋白靶向CD8，并且细胞是CD8+细胞。在一些任何实施方案中，慢病毒载体的靶向包膜蛋白靶向ASGR1、ASGR2或TM4SF5，并且细胞是肝细胞。

[0144] 在一些任何实施方案中，与已将相同包膜蛋白掺入相似脂质双层中但与替代靶向部分融合的参考脂质颗粒相比，靶向脂质颗粒具有更多的靶向包膜蛋白表达。在一些实施方案中，替代靶向部分是单链可变片段(scFv)。在一些任何实施方案中，所述表达增加了为或大于5％、10％、20％、30％、40％、50％、60％、70％、80％、90％、100％、125％、150％、200％、300％、400％、500％或更多。在一些任何实施方案中，所述表达增加了为或大于1.5倍、2倍、3倍、4倍、5倍、6倍、7倍、8倍、9倍、10倍、15倍、20倍、30倍或更多，优选地为或约或大于10倍或更多。

[0145] 在一些任何实施方案中，转导后在靶细胞中的滴度为或大于$1x10^6$个转导单位(TU)/mL、为或大于$2x10^6$TU/mL、为或大于$3x10^6$TU/mL、为或大于$4x10^6$TU/mL、为或大于

$5x10^6$TU/mL、为或大于$6x10^6$TU/mL、为或大于$7x10^6$TU/mL、为或大于$8x10^6$TU/mL、为或大于$9x10^6$TU/mL、或为或大于$1x10^7$TU/mL。

[0146] 在一些任何实施方案中，在组合物中的脂质颗粒或慢病毒载体群体中，大于为或约50％、大于为或约55％、大于为或约60％、大于为或约65％、大于为或约70％、或大于为或约75％对靶向包膜蛋白呈表面阳性。在一些任何实施方案中，靶向包膜蛋白以至少约(0.001、0.002、0.005、0.01、0.02、0.05、0.1、0.2或0.5)个靶向包膜蛋白/nm^2的密度存在于靶向脂质颗粒的表面上。

[0147] 本文提供了组合物，其包含本文所述的任何实施方案的多个靶向脂质颗粒或本文所述的任何实施方案的多个慢病毒载体，其中所述靶向包膜蛋白以至少约(0.001、0.002、0.005、0.01、0.02、0.05、0.1、0.2或0.5)个靶向包膜蛋白/nm^2的平均密度存在于所述靶向脂质颗粒的表面上。

[0148] 在一些任何实施方案中，与已将相同的包膜蛋白掺入其膜(例如质膜)中但与替代靶向部分融合的参考生产者细胞相比，生产者细胞具有更多的靶向包膜蛋白的膜(例如质膜)表达，任选地其中所述替代靶向部分是单链可变片段(scFv)。在一些任何实施方案中，所述表达增加了为或大于5％、10％、20％、30％、40％、50％、60％、70％、80％、90％、100％、125％、150％、200％、300％、400％、500％或更多。在一些任何实施方案中，所述表达增加了为或大于1.5倍、2倍、3倍、4倍、5倍、6倍、7倍、8倍、9倍、10倍、15倍、20倍、30倍或更多倍，优选地为或约或大于10倍或更多倍。在一些任何实施方案中，所述生产者细胞具有每平方微米至少20个蛋白质(例如，至少50个、100个、200个、500个、1000个、2000个、5000个或10,000个蛋白质)的所述靶向包膜蛋白在所述生产者细胞的膜(例如质膜)上的表达。在一些任何实施方案中，靶向包膜蛋白构成生产者细胞的总膜(例如，质膜)蛋白的至少0.1％(例如，至少0.2％、0.5％、1％、2％、3％、4％、5％、6％、7％、8％、9％或10％)(例如，按总蛋白重量计)。

具体实施方式

[0149] 本文提供了靶向脂质颗粒，所述靶向脂质颗粒包含围绕内腔或空腔的脂质双层和靶向包膜蛋白，所述靶向包膜蛋白含有(1)亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分和(2)结合结构域，诸如单结构域抗体(sdAb)可变域，其中所述靶向包膜蛋白包埋在所述脂质颗粒的脂质双层中。在特定的实施方案中，结合结构域(诸如单结构域抗体)是具有结合(诸如特异性地结合)期望的靶分子的能力的抗体。示例性结合结构域描述于章节II.A.2中。在一些实施方案中，靶向脂质颗粒还含有包埋在脂质双层中的亨尼帕病毒融合(F)蛋白分子或其生物活性部分。在特定的实施方案中，脂质颗粒可以是病毒样颗粒、病毒或病毒载体(诸如慢病毒载体)。

[0150] 在一些实施方案中，G蛋白和F蛋白中的一者或两者来自亨德拉(HeV)或尼帕(NiV)病毒，或者是其生物活性部分或是其变体或突变体。在特定的实施方案中，G蛋白和F蛋白两者都来自亨德拉(HeV)或尼帕(NiV)病毒。在一些实施方案中，融合和附着糖蛋白介导尼帕病毒的细胞进入。

[0151] F蛋白(诸如NiV-F)是I类融合蛋白，其具有与许多家族(例如，HIV-1gp41或流感病毒血凝素[HA])的融合蛋白共同的结构和功能特征，诸如具有疏水性融合肽的胞外结构域

和两个七残基重复区域(White JM等人2008.Crit Rev Biochem Mol Biol 43:189-219)。F蛋白作为无活性前体F_0合成,并且通过蛋白水解切割成两个二硫键连接的亚基F_1和F_2而被激活(Moll M.等人2004.J.Virol.78(18):9705-9712)。

[0152] G蛋白是亨尼帕病毒(例如,尼帕病毒或亨德拉病毒)的附着蛋白,其是含有N末端胞质尾、跨膜结构域、细胞外茎和球状头的II型跨膜糖蛋白(Liu,Q.等人2015.Journal of Virology,89(3):1838-1850)。附着蛋白NiV-G识别受体肝配蛋白B2和肝配蛋白B3。受体与NiV-G的结合触发一系列的构象变化,所述一系列的构象变化最终导致NiV-F的触发,所述触发暴露NiV-F的融合肽,从而允许另一系列的构象变化,所述另一系列的构象变化导致病毒-细胞膜融合(Stone J.A.等人2016.J Virol.90(23):10762-10773)。肝配蛋白B2先前被鉴定为主要的NiV受体(Negrete等人,2005),并且肝配蛋白B3被鉴定为替代受体(Negrete等人,2006)。事实上,NiV-G对肝配蛋白B2和B3具有高亲和力,亲和结合常数(Kd)在皮摩尔范围内(Negrete等人,2006)(分别对于细胞表面表达的肝配蛋白B2和B3,Kd=0.06nM和0.58nM)。

[0153] 靶向脂质颗粒的转导效率可以通过在NiV-F和NiV-G中的一者或两者中工程化高促融合突变来改善。先前已经描述了若干种此类突变(参见例如,Lee等人,2011,Trends in Microbiology)。这可以用于例如保持NiV-G对肝配蛋白B2和/或B3的特异性和皮摩尔亲和力。另外,已经鉴定了完全消除肝配蛋白B2和B3结合但不影响该NiV-G与NiV-F的结合的NiV-G中的突变。改善脂质颗粒的靶向的方法可以通过将结合分子与G蛋白(例如Niv-G,包括具有消除肝配蛋白B2和肝配蛋白B3结合的突变的Niv-G)融合来实现。这可以允许改变的G蛋白向性,从而允许通过添加针对不同细胞表面分子的结合分子来靶向不是肝配蛋白B2+的期望的细胞类型。

[0154] 虽然已经产生了掺入与G蛋白融合的此类结合分子的再靶向脂质颗粒,但是本文发现一些结合分子当与G蛋白(例如NiV-G)融合时在脂质颗粒的表面上比其他分子表达得更好。例如,发现单结构域抗体(sdAb)(诸如VHH)可以比单链可变片段(scFv)表达好10倍。不希望受理论束缚,表达的增加可能是由于再靶向G蛋白在脂质颗粒表面上的稳定性增加。与类似含有针对相同靶分子的替代结合结构域(例如scFv)的脂质颗粒相比,这种更多的表达可以改善脂质颗粒靶向靶分子(例如细胞表面分子)的能力。

[0155] 因此,本文提供了靶向脂质颗粒,所述靶向脂质颗粒含有与针对或能够结合靶细胞上的细胞表面分子的sdAb可变域连接的亨尼帕病毒的G蛋白(例如亨德拉或尼帕病毒,例如NiV-G)。sdAb可变域可包括仅VL或VH sdAb、纳米抗体、骆驼科动物VHH结构域、鲨鱼IgNAR或其片段的那些。在一些实施方案中,所述sdAb是VHH。

[0156] 在所提供的实施方案的各个方面,靶向脂质颗粒可以被工程化以表达亨尼帕病毒F蛋白分子或其生物活性部分;和靶向包膜蛋白,所述靶向包膜蛋白包含(i)亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分和(ii)单结构域抗体(sdAb)可变域,其中所述F蛋白分子或其生物活性部分和所述靶向包膜蛋白包埋在脂质双层中。在一些实施方案中,sdAb可变域与G蛋白或其生物活性部分的C末端连接。在一些实施方案中,sdAb可变域通过接头与G蛋白连接。

[0157] 还提供了另外含有一种或多种外源药剂的靶向脂质颗粒,诸如用于将诊断剂或治疗剂递送至细胞,包括在体内施用于受试者之后。本文还提供了靶向脂质颗粒的方法和用

途，诸如在诊断和治疗方法中。还提供了多核苷酸，用于工程化、制备和产生靶向脂质非细胞颗粒的方法，含有所述颗粒的组合物，以及含有和用于使用、产生和施用所述颗粒的试剂盒和装置。

[0158] 本申请中提及的所有出版物(包括专利文献、科学文章和数据库)出于所有目的通过引用整体并入，其程度如同每篇单独的出版物单独地通过引用并入一样。如果本文阐述的定义与通过引用并入本文的专利、申请、公开的申请和其他出版物中阐述的定义相反或在其他方面不一致，则本文阐述的定义优先于通过引用并入本文的定义。

[0159] 本文所使用的章节标题仅用于组织目的，而不应被解释为限制所描述的主题。

附图说明

[0160] 图1A-图1C描绘了用含有scFv或VHH结合形态的构建体转染的细胞的特征。图1A描绘了用含有scFV或VHH结合形态的构建体转染的细胞的表面表达，其通过流式细胞术进行分析并且描绘为中值荧光强度(MFI)，通过His+细胞％所定量。图1B描绘了用含有scFV或VHH结合形态的构建体转染的细胞与可溶性hCD4-Fc蛋白的结合，其通过流式细胞术进行分析并且描绘为中值荧光强度(MFI)，通过Fc+细胞％所定量。图1C描绘了与scFv结合形态相比用含有VHH结合形态的构建体转染的细胞在293细胞上的靶向结合序列的表面表达，其通过流式细胞术进行分析并且描绘为中值荧光强度(MFI)，如通过His+细胞％所定量。空载体和没有结合剂结构域的表达载体被用作阴性对照。

[0161] 图2描绘了含有scFV或VHH结合形态的四种示例性构建体在经负选择以富集T细胞、被解冻并用抗CD3/抗CD28激活的来自外周血的PanT细胞上的转导效力。通过流式细胞术分析细胞，并且通过为GFP+的CD4阳性细胞％确定滴度。

[0162] 图3A-图3B描绘了使用腹膜内注射到NOD-scid-IL2rγ^{无效}小鼠中的激活的PBMC在体内模型中的CD8再靶向假型化慢病毒的转导效率，如通过流式细胞术所分析。描绘了CD8再靶向假型化慢病毒对CD8+(图3A)或CD8-(图3B)T细胞的转导效率，并且通过为GFP+的CD8阳性或阴性细胞％确定滴度。

[0163] 图4A-图4B描绘了含有嵌合抗原受体(CAR)的CD8再靶向假型化慢病毒在体外实现白血病细胞杀伤的能力。图4A示出了在转导后4天检测CD8+细胞上的CD19+CAR表达的能力。图4B示出了在温育后18小时评估的Nalm6细胞的消除，通过流式细胞术所分析。

[0164] I.定义

[0165] 除非另外定义，否则本文所使用的所有技术术语、符号和其他技术和科学术语或术语学旨在具有与所要求保护的主题所属领域的普通技术人员通常理解的相同的含义。在一些情况下，为了清楚和/或为了便于参考，本文定义了具有通常理解的含义的术语，并且本文包括此类定义不应解释为表示与本领域通常理解的含义的显著差异。

[0166] 除非另有定义，本文所使用的所有技术和科学术语、首字母缩略词和缩写具有与本发明所属领域的普通技术人员通常理解的相同的含义。除非另有说明，化学和生物化学名称的缩写和符号是按照IUPAC-IUB命名法。除非另有说明，所有数值范围包括定义所述范围的值以及介于中间的所有整数值。

[0167] 如本文所用，冠词“一个”和“一种”是指该冠词的语法对象中的一个/一种或多于一个/多于一种(即，至少一个/至少一种)。例如，“一个要素”意指一个要素或多于一个要

素。

[0168] 如本文所用,术语“约”将为本领域普通技术人员所理解,并且将在其使用的上下文中在一定程度上变化。如本文所用,“约”当指可测量的值诸如量、短暂的持续时间等时,意指涵盖与指定值相差±20%或±10%、更优选±5%、甚至更优选±1%、以及还更优选±0.1%的变化,因为此类变化适于执行所公开的方法。

[0169] 如本文所用,“脂质颗粒”是指包含围绕内腔或空腔的两亲性脂质双层的任何生物或合成颗粒。典型地,脂质颗粒不含有核。脂质颗粒的实例包括固体颗粒诸如纳米颗粒、病毒衍生的颗粒或细胞衍生的颗粒。此类脂质颗粒包括但不限于病毒颗粒(例如慢病毒颗粒)、病毒样颗粒、病毒载体(例如慢病毒载体)外来体、去核细胞、各种囊泡(诸如微泡、膜囊泡、细胞外膜囊泡、质膜囊泡、巨质膜囊泡)、凋亡小体、有丝分裂颗粒(mitoparticle)、核红细胞(pyrenocyte)或溶酶体。在一些实施方案中,脂质颗粒可以是融合体(fusosome)。在一些实施方案中,脂质颗粒不是血小板。

[0170] 如本文所用,“生物活性部分”(诸如关于蛋白质诸如G蛋白或F蛋白)是指表现出或保留全长蛋白质的活性或特性的蛋白质部分。例如,当各自包埋在脂质双层中时,F蛋白的生物活性部分保留与G蛋白结合的促融合活性。例如,当各自包埋在脂质双层中时,G蛋白的生物活性部分保留与F蛋白结合的促融合活性。保留的活性包括全长或野生型F蛋白或G蛋白的活性的10%-150%或更多。F蛋白和G蛋白的生物活性部分的实例包括胞质结构域的截短,例如多达1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、11个、12个、13个、14个、15个、20,25个、30个、35个或更多个连续氨基酸的截短,参见例如Khetawat和Broder 2010 Virology Journal 7:312;Witting等人2013 Gene Therapy 20:997-1005;公布的国际专利申请号WO/2013/148327。

[0171] 如本文所用,“融合体”是指含有围绕内腔或空腔的两亲性脂质双层和与两亲性脂质双层相互作用的融合剂的颗粒。在实施方案中,融合体包含核酸。在一些实施方案中,融合体是膜包围的制剂。在一些实施方案中,融合体来源于源细胞。

[0172] 如本文所用,“融合体组合物”是指包含一种或多种融合体的组合物。

[0173] 如本文所用,“融合剂”是指在两个膜包围的内腔之间产生相互作用的药剂或分子。在实施方案中,融合剂促进膜的融合。在其他实施方案中,融合剂在两个内腔(例如,逆转录病毒载体的内腔和靶细胞的细胞质)之间产生连接,例如小孔。在一些实施方案中,融合剂包含两种或更多种蛋白质的复合物,例如,其中两种蛋白质单独都不具有促融合活性。在一些实施方案中,融合剂包含靶向结构域。

[0174] 如本文所用,“再靶向融合剂”是指包含具有不是融合剂的天然存在形式的一部分的序列的靶向部分的融合剂。在实施方案中,融合剂包含相对于融合剂的天然存在形式中的靶向部分不同的靶向部分。在实施方案中,融合剂的天然存在形式缺乏靶向结构域,并且再靶向融合剂包含融合剂的天然存在形式不存在的靶向部分。在实施方案中,融合剂被修饰以包含靶向部分。在实施方案中,相对于融合剂的天然存在形式,融合剂在靶向部分的外侧包含一个或多个序列改变,例如在跨膜结构域、促融合活性结构域或胞质结构域中。

[0175] 如本文所用,“靶向包膜蛋白”是指含有与靶向期望的细胞类型上的分子单结构域抗体(sdAb)可变域连接的亨尼帕病毒G蛋白的多肽,所述单结构域抗体可变域诸如仅VL或VH sdAb、纳米抗体、骆驼科动物VHH结构域、鲨鱼IgNAR或其片段。在一些此类实施方案中,

所述连接可以直接进行或通过接头(诸如肽接头)间接进行。

[0176] 如本文所用,“靶向脂质颗粒”是指含有包埋在脂质双层中的靶向包膜蛋白的脂质颗粒。

[0177] 如本文所用,“逆转录病毒核酸”是指含有单独或与辅助细胞、辅助病毒或辅助质粒组合包装到逆转录病毒或逆转录病毒载体中的至少最小序列要求的核酸。在一些实施方案中,逆转录病毒核酸还包含或编码外源药剂、阳性靶细胞特异性调控元件、非靶细胞特异性调控元件或阴性TCSRE。在一些实施方案中,逆转录病毒核酸包含5'LTR(例如,以促进整合)、U3(例如,以激活病毒基因组RNA转录)、R(例如,Tat结合区)、U5、3'LTR(例如,以促进整合)、包装位点(例如,psi(Ψ))、RRE(例如,以结合Rev并促进核输出)中的一个或多个(例如,全部)。逆转录病毒核酸可包含RNA(例如,当为病毒体的一部分时)或DNA(例如,当被引入源细胞中时或在受体细胞中逆转录之后)。在一些实施方案中,使用包含gag、pol和env中的一种或多种(例如,全部)的辅助细胞、辅助病毒或辅助质粒来包装逆转录病毒核酸。

[0178] 如本文所用,“靶细胞”是指期望靶向脂质颗粒递送外源药剂的类型的细胞。在实施方案中,靶细胞是特定组织类型或种类的细胞,例如免疫效应细胞,例如T细胞。在一些实施方案中,靶细胞是患病细胞,例如癌细胞。在一些实施方案中,与非靶细胞相比,融合剂(例如再靶向融合剂)导致将外源药剂优先递送至靶细胞。

[0179] 如本文所用,“非靶细胞”是指不期望靶向脂质颗粒递送外源药剂的类型的细胞。在一些实施方案中,非靶细胞是特定组织类型或种类的细胞。在一些实施方案中,非靶细胞是非患病细胞,例如非癌细胞。在一些实施方案中,与靶细胞相比,融合剂(例如再靶向融合剂)导致外源药剂向非靶细胞的递送较少。

[0180] 如本文所用,“单结构域抗体”或“sdAb”是指具有单个单体结构域抗原结合/识别结构域的抗体。此类抗体包括纳米抗体、骆驼科动物抗体(例如VHH)或鲨鱼抗体(例如IgNAR)。在一些实施方案中,sdAb的可变域包含三个CDR和四个框架区,命名为FR1、CDR1、FR2、CDR2、FR3、CDR3和FR4。在一些实施方案中,sdAb可变域可以在N末端或C末端处被截短,使得其仅包含部分FR1和/或FR4,或者缺少这些框架区中的一个或两个,只要sdAb可变域基本上保持抗原结合和特异性即可。

[0181] 术语“CDR”表示本领域技术人员通过至少一种鉴定方式定义的互补决定区。给定CDR或FR的精确氨基酸序列边界可以使用许多熟知的方案中的任一种容易地确定,所述许多熟知的方案包括由以下文献所描述的那些:Kabat等人(1991),“Sequences of Proteins of Immunological Interest,”第5版Public Health Service,National Institutes of Health,Bethesda,MD(“Kabat”编号方案);Al-Lazikani等人,(1997)JMB 273,927-948(“Chothia”编号方案);MacCallum等人,J.Mol.Biol.262:732-745(1996),“Antibody-antigen interactions:Contact analysis and binding site topography,”J.Mol.Biol.262,732-745.”(“Contact”编号方案);Lefranc MP等人,“IMGT unique numbering for immunoglobulin and T cell receptor variable domains and Ig superfamily V-like domains,”Dev Comp Immunol,2003年1月;27(1):55-77(“IMGT”编号方案);Honegger A和Plückthun A,“Yet another numbering scheme for immunoglobulin variable domains:an automatic modeling and analysis tool,”J Mol Biol,2001年6月8日;309(3):657-70,(“Aho”编号方案);以及Martin等人,“Modeling

antibody hypervariable loops:a combined algorithm,"PNAS,1989,86(23):9268-9272,("AbM"编号方案)。

[0182] 给定CDR或FR的边界可以取决于用于鉴定的方案而变化。例如,Kabat方案是基于结构比对,而Chothia方案是基于结构信息。Kabat和Chothia方案两者的编号是基于最常见的抗体区序列长度,其中插入通过插入字母表示,例如"30a",并且缺失出现在一些抗体中。这两种方案将某些插入和缺失("插入缺失(indel)")置于不同的位置,导致不同的编号。Contact方案是基于复杂晶体结构的分析,并且在许多方面类似于Chothia编号方案。AbM方案是基于由Oxford Molecular的AbM抗体建模软件使用的介于Kabat和Chothia定义之间的折衷。

[0183] 在一些实施方案中,可以根据Chothia编号方案、Kabat编号方案、Kabat和Chothia的组合、AbM定义和/或contact定义中的任一种来定义CDR。sdAb可变域包含三个CDR,命名为CDR1、CDR2和CDR3。下面表1列出了分别由Kabat、Chothia、AbM和Contact方案鉴定的CDR-H1、CDR-H2、CDR-H3的示例性位置边界。对于CDR-H1,使用Kabat和Chothia编号方案列出了残基编号。FR位于CDR之间,例如FR-H1定位于CDR-H1之前,FR-H2定位于CDR-H1与CDR-H2之间,FR-H3定位于CDR-H2与CDR-H3之间等等。应当注意,因为所示的Kabat编号方案在H35A和H35B处放置插入,所以当使用所示的Kabat编号惯例编号时,Chothia CDR-H1环的末端在H32与H34之间变化,这取决于环的长度。

[0184]

表 1. 根据各种编号方案的 CDR 边界。

CDR	Kabat	Chothia	AbM	Contact
CDR-H1 (Kabat 编号 [1])	H31--H35B	H26--H32..34	H26--H35B	H30--H35B
CDR-H1 (Chothia 编号 [2])	H31--H35	H26--H32	H26--H35	H30--H35
CDR-H2	H50--H65	H52--H56	H50--H58	H47--H58
CDR-H3	H95--H102	H95--H102	H95--H102	H93--H101

[0185] 1-Kabat等人(1991),"Sequences of Proteins of Immunological Interest,"第5版Public Health Service,National Institutes of Health,Bethesda,MD

[0186] 2-Al-Lazikani等人,(1997)JMB 273,927-948

[0187] 因此,除非另有说明,否则给定抗体或其区域(诸如其可变区)的"CDR"或"互补决定区"或单独指定的CDR(例如CDR-H1、CDR-H2、CDR-H3)应理解为涵盖如任何前述方案所定义的(或特定的)互补决定区。例如,当陈述特定CDR(例如CDR-H3)含有给定sdAb氨基酸序列中的相应CDR的氨基酸序列时,应当理解这样的CDR具有sdAb内的相应CDR(例如CDR-H3)的序列,如任何前述方案所定义。应当理解,任何抗体(诸如sdAb)包括CDR,并且这可以根据任何其他前述编号方案或本领域技术人员已知的其他编号方案来鉴定。

[0188] 如本文所用,术语"特异性地结合"靶分子诸如抗原意指结合分子诸如单结构域抗体与特定靶分子的反应或缔合比与替代分子的反应或缔合更频繁、更迅速、持续时间更长和/或亲和力更高。如果以比结合其他分子更大的亲和力、亲合力、更容易地和/或以更长的持续时间结合,则结合分子(诸如sdAb可变域)"特异性地结合"靶分子。应当理解,特异性地结合第一靶标的结合分子诸如sdAb可以特异性地结合或不特异性地结合第二靶标。因此,

“特异性结合”不一定需要(尽管其可包括)排他性结合。

[0189] 如本文所用,关于肽、多肽或抗体序列的“氨基酸序列同一性百分比(%)”和“同源性”被定义为在比对序列并引入缺口(如果需要)以实现最大序列同一性百分比并且不将任何保守取代视为序列同一性的一部分之后,候选序列中与特定肽或多肽序列中的氨基酸残基相同的氨基酸残基的百分比。用于确定氨基酸序列同一性百分比的目的的比对可以以本领域技术内的各种方式(例如使用公众可获得的计算机软件诸如BLAST、BLAST-2、ALIGN或MEGALIGNTM(DNASTAR)软件)来实现。本领域技术人员可以确定用于测量比对的适当参数,包括在被比较的序列的全长上实现最大比对所需的任何算法。

[0190] 氨基酸取代可包括但不限于多肽中的一个氨基酸用另一种氨基酸替换。示例性取代显示于表2中。可以将氨基酸取代引入感兴趣的抗体中,并且筛选产物的期望活性,例如保留/改善的结合。

[0191] 表2

[0192]

原始残基	示例性取代
Ala(A)	Val;Leu;Ile
Arg(R)	Lys;Gln;Asn
Asn(N)	Gln;His;Asp;Lys;Arg
Asp(D)	Glu;Asn
Cys(C)	Ser;Ala
Gln(Q)	Asn;Glu
Glu(E)	Asp;Gln
Gly(G)	Ala
His(H)	Asn;Gln;Lys;Arg
Ile(I)	Leu;Val;Met;Ala;Phe;正亮氨酸
Leu(L)	正亮氨酸;Ile;Val;Met;Ala;Phe
Lys(K)	Arg;Gln;Asn
Met(M)	Leu;Phe;Ile
Phe(F)	Trp;Leu;Val;Ile;Ala;Tyr
Pro(P)	Ala
Ser(S)	Thr
Thr(T)	Val;Ser
Trp(W)	Tyr;Phe
Tyr(Y)	Trp;Phe;Thr;Ser
Val(V)	Ile;Leu;Met;Phe;Ala;正亮氨酸

[0193] 可以根据共同的侧链特性对氨基酸进行分组:

[0194] (1)疏水性:正亮氨酸、Met、Ala、Val、Leu、Ile;

[0195] (2)中性亲水性:Cys、Ser、Thr、Asn、Gln;

[0196] (3)酸性:Asp、Glu;

[0197] (4)碱性:His、Lys、Arg;

[0198] (5)影响链取向的残基:Gly、Pro;

[0199] (6)芳香族:Trp、Tyr、Phe。

[0200] 非保守取代将需要将这些类别之一的成员交换成另一类别。

[0201] 关于蛋白质的位置的术语“对应于”,诸如叙述核苷酸或氨基酸位置“对应于”所公开的序列(诸如序列表中所示的)中的核苷酸或氨基酸位置,是指基于结构序列比对或使用标准比对算法(诸如GAP算法)与所公开的序列比对后鉴定的核苷酸或氨基酸位置。例如,相似序列(例如片段或物种变体)的对应残基可以通过结构比对方法与参考序列比对来确定。通过比对序列,本领域技术人员可以鉴定对应的残基,例如,使用保守的和相同的氨基酸残基作为指导。

[0202] 如本文所用,术语“分离的”是指已与通常在自然界中发现或产生的至少一些组分分离的分子。例如,当多肽与产生它的细胞的至少一些组分分离时,所述多肽被称为“分离的”。当多肽在表达后由细胞分泌的情况下,从产生所述多肽的细胞中物理地分离含有所述多肽的上清液被认为是“分离”所述多肽。类似地,当多核苷酸不是其通常在自然界中发现的较大多核苷酸(例如在DNA多核苷酸的情况下为基因组DNA或线粒体DNA)的一部分时或与产生所述多核苷酸的细胞的至少一些组分分离(例如在RNA多核苷酸的情况下)时,将所述多核苷酸称为“分离的”。因此,在宿主细胞内的载体中包含的DNA多核苷酸可以被称为“分离的”。

[0203] 如本文所用,术语“有效量”意指足以显著且正面地改变待治疗的症状和/或病症(例如,提供正面临床反应)的药物组合物的量。用于药物组合物的活性成分的有效量将随所治疗的具体病症、病症的严重程度、治疗的持续时间、并行疗法的性质、所使用的具体活性成分、所使用的具体药学上可接受的赋形剂和/或载体和主治医师知识和专业知识内的类似因素而变化。

[0204] 如本文所用的关于靶向脂质颗粒的“外源药剂”是指在对应的野生型病毒中既不包含也不编码的药剂或由对应的野生型源细胞产生的融合剂。在一些实施方案中,外源药剂不是天然存在的,诸如具有相对于天然存在的蛋白质改变(例如,通过插入、缺失或取代)的序列的蛋白质或核酸。在一些实施方案中,外源药剂不天然存在于源细胞中。在一些实施方案中,外源药剂天然存在于源细胞中,但对于病毒而言是外源的。在一些实施方案中,外源药剂不天然存在于受体细胞中。在一些实施方案中,外源药剂天然存在于受体细胞中,但不以期望的水平或在期望时间存在。在一些实施方案中,外源药剂包含RNA或蛋白质。

[0205] 如本文所用,“启动子”是指当与基因编码序列可操作地连接时驱动基因转录的顺式调节DNA序列。启动子可包含转录因子结合位点。在一些实施方案中,启动子与基因远端的一个或多个增强子协同工作。

[0206] 如本文所用,组合物是指两种或更多种产物、物质或化合物(包括细胞)的任何混合物。其可以是溶液、悬浮液、液体、粉末、糊剂、水性、非水性或它们的任何组合。

[0207] 如本文所用,术语“药学上可接受的”是指不消除化合物的生物活性或特性并且相对无毒的材料(诸如载体或稀释剂),即,所述材料可以施用于个体而不引起不期望的生物效应或以有害方式与包含其的组合物的任何组分相互作用。

[0208] 如本文所用,术语“药物组合物”是指本发明的至少一种化合物与其他化学组分(诸如载体、稳定剂、稀释剂、分散剂、悬浮剂、增稠剂和/或赋形剂)的混合物。药物组合物有助于将化合物施用于生物体。本领域中存在多种施用化合物的技术,包括但不限于静脉内、

口服、气雾剂、肠胃外、眼部、肺部和局部施用。

[0209] 如本文所用的“疾病”或“病况”是指需要和/或期望治疗的病症。

[0210] 如本文所用，术语“治疗(treat、treating或treatment)”是指改善疾病或病况，例如减缓或阻止或减少疾病或病况的发展或减少其至少一种临床症状。出于本公开的目的，改善疾病或病况可包括获得有益或期望的临床结果，所述有益或期望的临床结果包括但不限于以下中的任何一项或多项：减轻一种或多种症状、减小疾病程度、预防或延迟疾病的扩散(例如，转移，例如，转移至肺部或淋巴结)、预防或延迟疾病的复发、延迟或减慢疾病进展、改善疾病状态、抑制疾病或疾病进展、抑制或减慢疾病或其进展、阻止其发展、和缓解(无论是部分还是全部)。

[0211] 术语“个体”和“受试者”在本文中可互换使用，是指动物；例如哺乳动物。术语患者包括人和兽医受试者。在一些实施方案中，提供了治疗哺乳动物的方法，所述哺乳动物包括但不限于人、啮齿动物、猿、猫科动物、犬类、马科动物、牛、猪、绵羊、山羊、哺乳动物实验室动物、哺乳动物农场动物、哺乳动物运动动物和哺乳动物宠物。受试者可以是雄性或雌性，并且可以是任何合适的年龄，包括婴儿、少年、青少年、成年和老年受试者。在一些实例中，“个体”或“受试者”是指需要治疗疾病或病况的个体或受试者。在一些实施方案中，接受治疗的受试者可以是患者，表明所述受试者已经被鉴定为患有与治疗有关的病况或有足够风险患上所述病况的事实。在特定实施方案中，受试者是人，诸如人患者。

[0212] II.靶向脂质颗粒(例如慢病毒载体)

[0213] 本文提供了靶向脂质颗粒，所述靶向脂质颗粒包含亨尼帕病毒F蛋白分子或其生物活性部分和靶向包膜蛋白，所述靶向包膜蛋白包含(i)亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分和(ii)结合结构域，其中所述结合结构域与所述G蛋白或所述生物活性部分的C末端连接，其中(i)和(ii)中的每一者暴露在所述靶向脂质颗粒的外表面上。在一些实施方案中，所述结合结构域是单结构域抗体。在一些实施方案中，所述结合结构域是单链可变片段。在特定的实施方案中，所提供的脂质颗粒表现出促融合活性，其由促进与靶细胞结合的靶向包膜蛋白介导并且含有G蛋白或其生物活性部分、以及参与促进脂质颗粒的两个内腔与靶细胞膜的合并或融合的F糖蛋白。

[0214] 本文提供了靶向脂质颗粒，所述靶向脂质颗粒包含亨尼帕病毒F蛋白分子或其生物活性部分和靶向包膜蛋白，所述靶向包膜蛋白包含(i)亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分和(ii)单结构域抗体(sdAb)可变域，其中所述单结构域抗体与所述G蛋白或所述生物活性部分的C末端连接，其中(i)和(ii)中的每一者暴露在所述靶向脂质颗粒的外表面上。在特定的实施方案中，所提供的脂质颗粒表现出促融合活性，其由促进与靶细胞结合的靶向包膜蛋白介导并且含有G蛋白或其生物活性部分、以及参与促进脂质颗粒的两个内腔与靶细胞膜的合并或融合的F糖蛋白。

[0215] 在一些任何实施方案中，所述靶向脂质颗粒是病毒颗粒或病毒样颗粒。在一些方面，此类靶向脂质颗粒含有病毒核酸，诸如逆转录病毒核酸，例如慢病毒核酸。在特定的实施方案中，任何所提供的靶向脂质颗粒(诸如病毒颗粒或病毒样颗粒)都是复制缺陷的。在一些实施方案中，靶向脂质颗粒是慢病毒载体，其中所述慢病毒载体用亨尼帕病毒F蛋白和靶向包膜蛋白假型化。

[0216] 例如，本文提供了假型化慢病毒载体，所述假型化慢病毒载体包含亨尼帕病毒F蛋

白分子或其生物活性部分和靶向包膜蛋白，所述靶向包膜蛋白包含(i)亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分和(ii)结合结构域，其中所述结合结构域与所述G蛋白或所述生物活性部分的C末端连接，其中(i)和(ii)中的每一者暴露在所述靶向脂质颗粒的外表面上。在一些实施方案中，所述结合结构域是单结构域抗体。在一些实施方案中，所述结合结构域是单链可变片段。

[0217] 在一些实施方案中，与掺入相似的包膜蛋白但融合与除sdAb可变域之外的替代靶向部分(诸如单链可变片段(scFv))融合的参考脂质颗粒(例如参考慢病毒载体)相比，本文提供的靶向脂质颗粒(例如靶向慢病毒载体)具有增加的或更多的靶向包膜蛋白的表达。在一些实施方案中，在用转移质粒、包膜质粒和gag-pol质粒共转染包装细胞后，通过将脂质颗粒(例如慢病毒颗粒)假型化来产生此类靶向脂质颗粒。

[0218] 在一些实施方案中，与参考脂质颗粒(例如参考慢病毒载体)(例如含有相似包膜蛋白但与scFv融合的参考脂质颗粒)相比，表达增加了为或大于5％、10％、20％、30％、40％、50％、60％、70％、80％、90％、100％、125％、150％、200％、300％、400％、500％或更多。在一些实例中，与参考脂质颗粒(例如参考慢病毒载体)(例如含有相似包膜蛋白但与scFv融合的参考脂质颗粒)相比，表达增加了1.5倍、2倍、3倍、4倍、5倍、6倍、7倍、8倍、9倍、10倍、15倍、20倍、30倍或更多倍。在一些实施方案中，可以使用流式细胞术(例如FAC)在体外测定表达。在一些实施方案中，表达可以描绘为靶向包膜蛋白在靶向脂质颗粒(例如靶向慢病毒载体)表面上的数量或密度。在一些实施方案中，表达可以描绘为靶向包膜蛋白在靶向脂质颗粒(例如靶向慢病毒载体)表面上的表面表达的平均荧光强度(MFI)。在一些实施方案中，表达可以描绘为在群体中脂质颗粒(例如慢病毒载体)对靶向包膜蛋白呈表面阳性的百分比。

[0219] 在一些实施方案中，在靶向脂质颗粒(例如靶向慢病毒载体)群体中，大于为或约50％的脂质颗粒对靶向包膜蛋白呈表面阳性。例如，在所提供的靶向脂质颗粒(例如靶向慢病毒载体)群体中，群体中大于为或约55％、大于为或约60％、大于为或约65％、大于为或约70％、大于为或约75％的细胞对靶向包膜蛋白呈表面阳性。

[0220] 在一些实施方案中，与引入相同靶细胞的参考脂质颗粒(例如参考慢病毒载体)的滴度相比，在引入靶细胞(诸如通过转导(例如转导细胞))后的靶向脂质颗粒的滴度增加，所述参考脂质颗粒掺入相似包膜蛋白但与除sdAb可变域之外的替代靶向部分(例如单链可变片段(scFv))融合。典型地，替代靶向部分识别或结合与靶向脂质颗粒的靶向包膜蛋白的sdAb可变域相同的靶分子。在一些实施方案中，与参考脂质颗粒(例如参考慢病毒载体)(例如含有相似包膜蛋白但与scFv融合的参考脂质颗粒)的滴度相比，滴度增加了为或大于5％、10％、20％、30％、40％、50％、60％、70％、80％、90％、100％、125％、150％、200％、300％、400％、500％或更多。在一些实例中，与参考脂质颗粒(例如参考慢病毒载体)(例如含有相似包膜蛋白但与scFv融合的参考脂质颗粒)的滴度相比，滴度增加了1.5倍、2倍、3倍、4倍、5倍、6倍、7倍、8倍、9倍、10倍、15倍、20倍、30倍或更多倍。在一些实施方案中，靶细胞(例如转导细胞)中的靶向脂质颗粒的滴度大于为或约$1x10^6$个转导单位(TU)/mL。例如，靶细胞(例如转导细胞)中的靶向脂质颗粒的滴度大于为或约$2x10^6$TU/mL、大于为或约$3x10^6$TU/mL、大于为或约$4x10^6$TU/mL、大于为或约$5x10^6$TU/mL、大于为或约$6x10^6$TU/mL、大于为或约$7x10^6$TU/mL、大于为或约$8x10^6$TU/mL、大于为或约$9x10^6$TU/mL、或大于为或约

$1x10^7$TU/mL。

[0221] A.靶向包膜蛋白(例如亨尼帕病毒加结合结构域)

[0222] 在一些实施方案中,靶向脂质颗粒(例如慢病毒载体)包括暴露在靶向脂质颗粒(例如慢病毒载体)表面上的靶向包膜蛋白。

[0223] 在一些实施方案中,靶向包膜蛋白含有亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分和结合靶细胞上的细胞表面分子的结合结构域。在一些实施方案中,结合结构域是单结构域抗体(sdAb)。在一些实施方案中,结合结构域是单链可变片段(scFv)。结合结构域可以直接或间接地与G蛋白连接。在特定的实施方案中,结合结构域与G蛋白或其生物活性部分的C末端(C末端氨基酸)连接。连接可以通过肽接头(诸如柔性肽接头)进行。

[0224] 1.G蛋白

[0225] 在一些实施方案中,靶向包膜蛋白含有亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分和单结构域抗体(sdAb)可变域或其生物活性部分。在一些实施方案中,sdAb结合靶细胞上的细胞表面分子。sdAb可变域可以直接或间接地与G蛋白连接。在特定的实施方案中,sdAb可变域与G蛋白或其生物活性部分的C末端(C末端氨基酸)连接。连接可以通过肽接头(诸如柔性肽接头)进行。

[0226] 在一些实施方案中,结合结构域(例如sdAb)结合细胞的细胞表面抗原。在一些实施方案中,细胞表面抗原是一种类型的细胞的特征。在一些实施方案中,细胞表面抗原是多于一种类型的细胞的特征。

[0227] 在一些实施方案中,结合结构域(例如sdAb)可变域结合细胞表面分子或抗原。在一些实施方案中,细胞表面分子是ASGR1、ASGR2、TM4SF5、CD8、CD4或低密度脂蛋白受体(LDL-R)。在一些实施方案中,细胞表面分子是ASGR1。在一些实施方案中,细胞表面分子是ASGR2。在一些实施方案中,细胞表面分子是TM4SF5。在一些实施方案中,细胞表面分子是CD8。在一些实施方案中,细胞表面分子是CD4。在一些实施方案中,细胞表面分子是LDL-R。

[0228] 在一些实施方案中,G蛋白是亨尼帕病毒G蛋白或其生物活性部分。在一些实施方案中,亨尼帕病毒G蛋白是亨德拉(HeV)病毒G蛋白、尼帕(NiV)病毒G蛋白(NiV-G)、雪松(Cedar)(CedPV)病毒G蛋白、墨江(Mojiang)病毒G蛋白、蝙蝠副粘病毒G蛋白或其生物活性部分。表3提供了G蛋白的非限制性实例。

[0229] 附着G蛋白是II型跨膜糖蛋白,其含有N末端胞质尾(例如对应于SEQ ID NO:9的氨基酸1-49)、跨膜结构域(例如对应于SEQ ID NO:9的氨基酸50-70)以及含有细胞外茎(例如对应于SEQ ID NO:9的氨基酸71-187)和球状头(对应于SEQ ID NO:9的氨基酸188-602)的细胞外结构域。N末端胞质结构域在脂质双层的内部内腔中,并且C末端部分是暴露在脂质双层外侧的细胞外结构域。已显示C末端区中的茎的区域(例如,对应于NiV-G的氨基酸159-167)参与与F蛋白的相互作用和F蛋白融合的触发(Liu等人2015J of Virology 89:1838)。在野生型G蛋白中,球状头介导受体与亨尼帕病毒进入受体肝配蛋白B2和肝配蛋白B3的结合,但不是膜融合所必需的(Brandel-Tretheway等人Journal of Virology.2019.93(13)e00577-19)。在本文的特定实施方案中,G蛋白的向性通过G蛋白或其生物活性片段(例如胞质截短)与sdAb可变域的连接而改变。G蛋白与结合配偶体的结合可以触发由相容的F蛋白或其生物活性部分介导的融合。本文所公开的G蛋白序列主要公开为包括翻译起始所必需的N末端甲硫氨酸的表达序列。由于这样的N末端甲硫氨酸通常被共翻译或翻译后切割,所

以本文所公开的所有G蛋白序列的成熟蛋白序列也被认为缺乏N末端甲硫氨酸。

[0230] G糖蛋白在亨尼帕病毒物种之间高度保守。例如,NiV和HeV病毒的G蛋白共享79％的氨基酸同一性。研究已显示G蛋白与不同物种的F蛋白之间的高度相容性,如通过异型融合激活所证明(Brandel-Tretheway等人Journal of Virology.2019)。如下面进一步描述,再靶向脂质颗粒可以含有来自不同物种的异源G和F蛋白。

[0231] 表3.亨尼帕病毒蛋白G序列簇。第1列,Genbank ID包括作为簇的质心序列的病毒整个基因组序列的Genbank ID。第2列,CDS的核苷酸提供与整个基因组中基因的CDS对应的核苷酸。第3列,完整基因名称,提供基因的完整名称,包括Genbank ID、病毒物种、毒株和蛋白质名称。第4列,序列,提供基因的氨基酸序列。第5列,序列/簇#,提供了与该质心序列聚类的序列的数量。第6列提供所描述的序列的SEQ ID号。

[0232]

Genbank ID	CDS 的核苷酸	全序列 ID	序列	序列/簇#	SEQ ID NO	SEQ ID NO (没有 N 末端甲硫氨酸)
AF017149	8913-10727	gb:AF017149|生物体: 亨德拉病毒|毒株名称: 未知-AF017149|蛋白质名称: 糖蛋白|基因符号: G	MMADSKLVSLNNNLSGKIKDQGKVIKNYYGTMDIKKINDGLLDSKILGAFNTVIALLGSIIIIVMNIMIIQNYTRTTDNQALIKESLQSVQQQIKALTDKIGTEIGPKVSLIDTSSTITIPANIGLLGSKISQSTSSINENVNDKCKFTLPPLKIHECNISCPNPLPFREYRPISQGVSDLVGLPNQICLQKTTSTILKPRLISYTLPINTREGVCITDPLLAVDNGFFAYSHLEKIGSCTRGIAKQRIIGVGEVLDRGDKVPSMFMTNVWTPPNPSTIHHCSSTYHEDFYYTLCAVSHVGDPILNSTSWTESLSLIRLAVRPKSDSGDYNQKYIAITKVERGKYDKVMPYGPSGIKQGDTLYFPAVGFLPRTEFQYNDSNCPIIHCKYSKAENCRLSMGVNSKSHYILRSGLLKYNLSLGGDIILQFIEIADNRLTIGSPSKIYNSLGQPVFYQASYSWDTMIKLGDVDTVDPLRVQWRNNSVISRPGQSQCPRFNVCPEVCWEGTYNDAFLIDRLNWVSAGVYLNSNQTAENPVFAVFKDNEILYQVPLAEDDTNAQKTITDCFLLENVIWCISLVEIYDTGDSVIRPKLFAVKIPAQCSES	14	18	52
AF212302	8943-10751	gb:AF212302|生物体: 尼帕病毒|毒株名称: 未知-AF212302|蛋白质名称: 附着糖蛋白|基因符号: G	MPAENKKVRFENTTSDKGKIPSKVIKSYYGTMDIKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQNYTRSTDNQAVIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSKISQSTASINENVNEKCKFTLPPLKIHECNISCPNPLPFREYRPQTEGVSNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMDEGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVWTPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILNSTYWSGSLMMTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQGDTLYFPAVGFLVRTEFKYNDSNCPITKCQYSKPENCRLSMGIRPNSHYILRSGLLKYNLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVFYQASFSWDTMIKFGDVLTVNPLVVNWRNNTVISRPGQSQCPRFNTCPEICWEGVYNDAFLIDRINWISAGVFLDSNQTAENPVFTVFKDNEIL	14	28	44

[0233]

			YRAQLASEDTNAQKTITNCFLLK NKIWCISLVEIYDTGDNVIRPKLF AVKIPEQCT			
JQ00 1776	8170-1 0275	gb:JQ0017 76:8170-10 275|生物体: 雪松病毒|毒株名称: CG1a|蛋白质名称: 附着糖蛋白|基因符号: G	MLSQLQKNYLDNSNQQGDKMN NPDKKLSVNFNPLELDKGQKDLN KSYYVKNKNYNVSNLLNESLHDI KFCIYCIFSLLIIITIINIITISIVITRLK VHEENNGMESPNLQSIQDSLSSLT NMINTEITPRIGILVTATSVTLSSSI NYVGTKTNQLVNELKDYITKSCG FKVPELKLHECNISCADPKISKSA MYSTNAYAELAGPPKIFCKSVSK DPDFRLKQIDYVIPVQQDRSICMN NPLLDISDGFFTYIHYEGINSCKKS DSFKVLLSHGEIVDRGDYRPSLYL LSSHYHPYSMQVINCVPVTCNQS SFVFCHISNNTKTLDNSDYSSDEY YITYFNGIDRPKTKKIPINNMTAD NRYIHFTFSGGGGVCLGEEFIIPVT TVINTDVFTHDYCESFNCSVQTG KSLKEICSESLRSPTNSSRYNLNGI MIISQNNMTDFKIQLNGITYNKLS FGSPGRLSKTLGQVLYYQSSMSW DTYLKAGFVEKWKPFTPNWMNN TVISRPNQGNCPRYHKCPEICYGG TYNDIAPLDLGKDMYVSVILDSD QLAENPEITVFNSTTILYKERVSK DELNTRSTTTSCFLFLDEPWCISV LETNRFNGKSIRPEIYSYKIPKYC	3	29	54
NC_ 0252 56	9117-1 1015	gb:NC_02 5256:9117-11015|生物体: 蝙蝠副粘病毒Eid_hel/GH-M74a/GHA/2009|毒株名称: BatPV/Eid_hel/GH-M74a/GHA/2009|蛋白质名称: 糖蛋白|基因符号: G	MPQKTVEFINMNSPLERGVSTLS DKKTLNQSKITKQGYFGLGSHSE RNWKKQKNQNDHYMTVSTMILE ILVVLGIMFNLIVLTMVYYQNDNI NQRMAELTSNITVLNLNLNQLTN KIQREIIPRITLIDTATTITIPSAITYI LATLTTRISELLPSINQKCEFKTPT LVLNDCRINCTPPLNPSDGVKMSS LATNLVAHGPSPCRNFSSVPTIYY YRIPGLYNRTALDERCILNPRLTIS STKFAYVHSEYDKNCTRGFKYYE LMTFGEILEGPEKEPRMFSRSFYS PTNAVNYHSCTPIVTVNEGYFLCL ECTSSDPLYKANLSNSTFHLVILR HNKDEKIVSMPSFNLSTDQEYVQI IPAEGGGTAESGNLYFPCIGRLLH KRVTHPLCKKSNCSRTDDESCLK SYYNQGSPQHQVVNCLIRIRNAQ RDNPTWDVITVDLTNTYPGSRSRI FGSFSKPMLYQSSVSWHTLLQVA EITDLDKYQLDWLDTPYISRPGGS ECPFGNYCPTVCWEGTYNDVYSL TPNNDLFVTVYLKSEQVAENPYF AIFSRDQILKEFPLDAWISSARTTT ISCFMFNNEIWCIAALEITRLNDDI	2	30	55

[0234]

			IRPIYYSFWLPTDCRTPYPHTGKM TRVPLRSTYNY			
NC_025352	8716-11257	gb:NC_025352:8716-11257|生物体: 墨江病毒|毒株名称: Tongguan1|蛋白质名称: 附着糖蛋白|基因符号: G	MATNRDNTITSAEVSQEDKVKKY YGVETAEKVADSISGNKVFILMN TLLILTGAIITITLNITNLTAAKSQQ NMLKIIQDDVNAKLEMFVNLDQL VKGEIKPKVSLINTAVSVSIPGQIS NLQTKFLQKYVYLEESITKQCTC NPLSGIFPTSGPTYPPTDKPDDDTT DDDKVDTTIKPIEYPKPDGCNRTG DHFTMEPGANFYTVPNLGPASSN SDECYTNPSFSIGSSIYMFSQEIRK TDCTAGEILSIQIVLGRIVDKGQQ GPQASPLLVWAVPNPKIINSCAVA AGDEMGWVLCSVTLTAASGEPIP HMFDGFWLYKLEPDTEVVSYRIT GYAYLLDKQYDSVFIGKGGGIQK GNDLYFQMYGLSRNRQSFKALCE HGSCLGTGGGGYQVLCDRAVMS FGSEESLITNAYLKVNDLASGKPV IIGQTFPPSDSYKGSNGRMYTIGD KYGLYLAPSSWNRYLRFGITPDIS VRSTTWLKSQDPIMKILSTCTNTD RDMCPEICNTRGYQDIFPLSEDSE YYTYIGITPNNGGTKNFVAVRDS DGHIASIDILQNYYSITSATISCFM YKDEIWCIAITEGKKQKDNPQRIY AHSYKIRQMCYNMKSATVTVGN AKNITIRRY	2	31	56

[0235] 在一些实施方案中,G蛋白具有SEQ ID NO:9、18、28、29、30、31、44、52或54-56中任一者所示的序列,或者是具有与SEQ ID NO:9、18、28、29、30、31、44、52或54-56中任一者至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、至少为或约84%、至少为或约85%、至少为或约86%、至少为或约87%、至少为或约88%、至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、至少为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%相同的序列的其功能活性变体或生物活性部分。在特定的实施方案中,所述G蛋白或功能活性变体或生物活性部分是保留与亨尼帕病毒F蛋白(诸如章节I.B中所述的F蛋白(例如NiV-F或HeV-F))结合的促融合活性的蛋白质。促融合活性包括G蛋白与亨尼帕病毒F蛋白结合以促进或帮助两个膜内腔(诸如在其脂质双层中包埋了亨尼帕病毒F和G蛋白的靶向脂质颗粒的内腔)和靶细胞(例如含有被靶向包膜蛋白识别或结合的表面受体或分子的细胞)的细胞质融合的活性。在一些实施方案中,所述F蛋白和G蛋白来自相同的亨尼帕病毒属物种(例如NiV-G和NiV-F)。在一些实施方案中,所述F蛋白和G蛋白来自不同的亨尼帕病毒属物种(例如NiV-G和HeV-F)。

[0236] 在特定的实施方案中,所述G蛋白具有SEQ ID NO:9、SEQ ID NO:28、SEQ ID NO:18、SEQ ID NO:30、SEQ ID NO:31、SEQ ID NO:44、SEQ ID NO:52或SEQ ID NO:54-56所示的氨基酸序列,或者是保留促融合活性的其功能活性变体或其生物活性部分。在一些实施方案中,所述功能活性变体包含与SEQ ID NO:9、SEQ ID NO:28、SEQ ID NO:18、SEQ ID NO:30、SEQ ID NO:31、SEQ ID NO:44、SEQ ID NO:52或SEQ ID NO:54-56具有至少为或约80%、

至少为或约85％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性并且保留与亨尼帕病毒F蛋白(例如,NiV-F或HeV-F)结合的促融合活性的氨基酸序列。在一些实施方案中,所述生物活性部分具有如下氨基酸序列,所述氨基酸序列与SEQ ID NO:9、SEQ ID NO:28、SEQ ID NO:18、SEQ ID NO:30、SEQ ID NO:31、SEQ ID NO:44、SEQ ID NO:52或SEQ ID NO:54-56具有至少或至少约80％、至少或至少约85％、至少或至少约90％、至少或至少约91％、至少或至少约92％、至少或至少约93％、至少或至少约94％、至少或至少约95％、为或约96％、至少或至少约97％、至少或至少约98％、或至少或至少约99％序列同一性并且保留与亨尼帕病毒F蛋白(例如,NiV-F或HeV-F)结合的促融合活性。

[0237] 提及保留促融合活性包括如下的(与亨尼帕病毒F蛋白结合的)活性,所述活性为诸如SEQ ID NO:9、SEQ ID NO:28、SEQ ID NO:18、SEQ ID NO:30、SEQ ID NO:31、SEQ ID NO:44、SEQ ID NO:52或SEQ ID NO:54-56所示的对应野生型G蛋白的结合水平或程度的为或约10％或与为或约150％或更高之间,诸如对应野生型G蛋白的促融合活性水平或程度的至少或至少约10％、诸如对应野生型G蛋白的促融合活性水平或程度的至少或至少约15％、诸如对应野生型G蛋白的促融合活性水平或程度的至少或至少约20％、诸如对应野生型G蛋白的促融合活性水平或程度的至少或至少约25％、诸如对应野生型G蛋白的促融合活性水平或程度的至少或至少约30％、诸如对应野生型G蛋白的促融合活性水平或程度的至少或至少约35％、诸如对应野生型G蛋白的促融合活性水平或程度的至少或至少约40％、诸如对应野生型G蛋白的促融合活性水平或程度的至少或至少约45％、诸如对应野生型G蛋白的促融合活性水平或程度的至少或至少约50％、诸如对应野生型G蛋白的促融合活性水平或程度的至少或至少约55％、诸如对应野生型G蛋白的促融合活性水平或程度的至少或至少约60％、诸如对应野生型G蛋白的促融合活性水平或程度的至少或至少约65％、诸如对应野生型G蛋白的促融合活性水平或程度的至少或至少约70％、诸如对应野生型G蛋白的促融合活性水平或程度的至少或至少约75％、诸如对应野生型G蛋白的促融合活性水平或程度的至少或至少约80％、诸如对应野生型G蛋白的促融合活性水平或程度的至少或至少约85％、诸如对应野生型G蛋白的促融合活性水平或程度的至少或至少约90％、诸如对应野生型G蛋白的促融合活性水平或程度的至少或至少约95％、诸如对应野生型G蛋白的促融合活性水平或程度的至少或至少约100％、或诸如对应野生型G蛋白的促融合活性水平或程度的至少或至少约120％。

[0238] 在一些实施方案中,G蛋白是含有一个或多个氨基酸突变(诸如一个或多个氨基酸插入、缺失、取代或截短)的功能活性变体或生物活性部分的突变体G蛋白。在一些实施方案中,本文所述的突变涉及与参考G蛋白序列相比氨基酸的氨基酸插入、缺失、取代或截短。在一些实施方案中,参考G蛋白序列是G蛋白的野生型序列或其生物活性部分。在一些实施方案中,其功能活性变体或生物活性部分是野生型亨德拉(HeV)病毒G蛋白、野生型尼帕(NiV)病毒G蛋白(NiV-G)、野生型雪松(CedPV)病毒G蛋白、野生型墨江病毒G蛋白、野生型蝙蝠副粘病毒G蛋白或其生物活性部分。在一些实施方案中,野生型G蛋白具有SEQ ID NO:9、18、28、29、30、31、SEQ ID NO:44、SEQ ID NO:52或SEQ ID NO:54-56中任一者所示的序列。

[0239] 在一些实施方案中,所述G蛋白是突变体G蛋白,其是野生型亨德拉(HeV)病毒G蛋

白、野生型尼帕(NiV)病毒G蛋白(NiV-G)、野生型雪松(CedPV)病毒G蛋白、野生型墨江病毒G蛋白、野生型蝙蝠副粘病毒G蛋白的N末端和/或C末端截短片段的生物活性部分。在特定的实施方案中,所述截短是全部或部分胞质结构域的N末端截短。在一些实施方案中,所述突变体G蛋白是被截短的生物活性部分,并且缺乏在野生型G蛋白的N末端处或附近的多达49个连续氨基酸残基,诸如SEQ ID NO:9、18、28、29、30、31、SEQ ID NO:44、SEQ ID NO:52或SEQ ID NO:54-56中任一者所示的野生型G蛋白。在一些实施方案中,突变体F蛋白被截短并且缺乏在野生型G蛋白的N末端处的多达49个连续氨基酸,诸如多达49个、48个、47个、46个、45个、44个、43个、42个、41个、40个、30个、38个、37个、36个、35个、34个、33个、32个、31个、30个、29个、28个、27个、26个、25个、24个、23个、22个、21个、20个、19个、18个、17个、16个、15个、14个、13个、12个、11个、10个、9个、8个、7个、6个、5个、4个、3个、2个或1个连续氨基酸。

[0240] 在一些实施方案中,所述G蛋白是野生型尼帕病毒G(NiV-G)蛋白或亨德拉病毒G蛋白,或是其功能活性变体或生物活性部分。在一些实施方案中,所述G蛋白是具有SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44所示序列的NiV-G蛋白,或是与SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、至少为或约84%、至少为或约85%、至少为或约86%、至少为或约87%、至少为或约88%、至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、为或约96%、至少为或约97%、至少为或约98%、至少为或约99%序列同一性的氨基酸序列的其功能变体或生物活性部分。

[0241] 在一些实施方案中,所述G蛋白是突变体NiV-G蛋白,其是野生型NiV-G的生物活性部分。在一些实施方案中,所述生物活性部分是N末端截短的片段。在一些实施方案中,突变体NiV-G蛋白被截短并且缺乏在野生型NiV-G蛋白(SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44)的N末端处或附近的多达5个连续氨基酸残基、在野生型NiV-G蛋白(SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44)的N末端处或附近的多达6个连续氨基酸残基、在野生型NiV-G蛋白(SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44)的N末端处或附近的多达7个连续氨基酸残基、在野生型NiV-G蛋白(SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44)的N末端处或附近的多达8个连续氨基酸残基、在野生型NiV-G蛋白(SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44)的N末端处或附近的多达9个连续氨基酸残基、在野生型NiV-G蛋白(SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44)的N末端处或附近的多达10个连续氨基酸残基、在野生型NiV-G蛋白(SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44)的N末端处或附近的多达11个连续氨基酸残基、在野生型NiV-G蛋白(SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44)的N末端处或附近的多达12个连续氨基酸残基、在野生型NiV-G蛋白(SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44)的N末端处或附近的多达13个连续氨基酸残基、在野生型NiV-G蛋白(SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44)的N末端处或附近的多达14个连续氨基酸残基、在野生型NiV-G蛋白(SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44)的N末端处或附近的多达15个连续氨基酸残基、在野生型NiV-G蛋白(SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44)的N末端处或附近的多达16个连续氨基酸残基、在野生型NiV-G蛋白(SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44)的N末端处或附近的多达17个连续氨基酸残基、在野生型NiV-G蛋白(SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44)的N末端处或附近的多达18个连续氨基酸残基、在野生型NiV-G蛋白(SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44)的N末端处或附近的

多达19个连续氨基酸残基、在野生型NiV-G蛋白(SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44)的N末端处或附近的多达20个连续氨基酸残基、在野生型NiV-G蛋白(SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44)的N末端处或附近的多达21个连续氨基酸残基、在野生型NiV-G蛋白(SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44)的N末端处或附近的多达22个连续氨基酸残基、在野生型NiV-G蛋白(SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44)的N末端处或附近的多达23个连续氨基酸残基、在野生型NiV-G蛋白(SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44)的N末端处或附近的多达24个连续氨基酸残基、在野生型NiV-G蛋白(SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44)的N末端处或附近的多达25个连续氨基酸残基、在野生型NiV-G蛋白(SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44)的N末端处或附近的多达26个连续氨基酸残基、在野生型NiV-G蛋白(SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44)的N末端处或附近的多达27个连续氨基酸残基、在野生型NiV-G蛋白(SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44)的N末端处或附近的多达28个连续氨基酸残基、在野生型NiV-G蛋白(SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44)的N末端处或附近的多达29个连续氨基酸残基、在野生型NiV-G蛋白(SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44)的N末端处或附近的多达30个连续氨基酸残基、在野生型NiV-G蛋白(SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44)的N末端处或附近的多达31个连续氨基酸残基、在野生型NiV-G蛋白(SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44)的N末端处或附近的多达32个连续氨基酸残基、在野生型NiV-G蛋白(SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44)的N末端处或附近的多达33个连续氨基酸残基、在野生型NiV-G蛋白(SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44)的N末端处或附近的多达34个连续氨基酸残基、在野生型NiV-G蛋白(SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44)的N末端处或附近的多达35个连续氨基酸残基、在野生型NiV-G蛋白(SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44)的N末端处或附近的多达36个连续氨基酸残基、在野生型NiV-G蛋白(SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44)的N末端处或附近的多达37个连续氨基酸残基、在野生型NiV-G蛋白(SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44)的N末端处或附近的多达38个连续氨基酸残基、在野生型NiV-G蛋白(SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44)的N末端处或附近的多达39个连续氨基酸残基、在野生型NiV-G蛋白(SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44)的N末端处或附近的多达40个连续氨基酸残基、在野生型NiV-G蛋白(SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44)的N末端处或附近的多达41个连续氨基酸残基、在野生型NiV-G蛋白(SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44)的N末端处或附近的多达42个连续氨基酸残基、在野生型NiV-G蛋白(SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44)的N末端处或附近的多达43个连续氨基酸残基、在野生型NiV-G蛋白(SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44)的N末端处或附近的多达44个连续氨基酸残基、或在野生型NiV-G蛋白(SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44)的N末端处或附近的多达45个连续氨基酸残基。

[0242] 在一些实施方案中,所述NiV-G蛋白是不含有胞质结构域的生物活性部分。在一些实施方案中,没有胞质结构域的NiV-G蛋白由SEQ ID NO:32编码。

[0243] 在一些实施方案中,突变体NiV-G蛋白包含SEQ ID NO:10-15、35-40、45-50、22、53或SEQ ID NO:32中任一者所示的序列,或是具有与SEQ ID NO:10-15、35-40、45-50、22、53或SEQ ID NO:32具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约

83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列的其功能变体。

[0244] 在一些实施方案中，所述突变体NiV-G蛋白具有在野生型NiV-G蛋白(SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44)的N末端处或附近的5个氨基酸截短，诸如SEQ ID NO:10所示或与SEQ ID NO:10具有至少为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的其功能变体，或诸如SEQ ID NO:35所示或与SEQ ID NO:35具有至少为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的其功能变体，或诸如SEQ ID NO:45所示或与SEQ ID NO:45具有至少为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的其功能变体。在一些实施方案中，所述突变体NiV-G蛋白具有在野生型NiV-G蛋白(SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44)的N末端处或附近的10个氨基酸截短，诸如SEQ ID NO:11所示或与SEQ ID NO:11具有至少为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的其功能变体，或诸如SEQ ID NO:36所示或与SEQ ID NO:36具有至少为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的其功能变体，或诸如SEQ ID NO:46所示或与SEQ ID NO:46具有至少为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的其功能变体。

[0245] 在一些实施方案中，所述突变体NiV-G蛋白具有在野生型NiV-G蛋白(SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44)的N末端处或附近的15个氨基酸截短，诸如SEQ ID NO:12所示或具有与SEQ ID NO:12具有至少为或约80％、至少为或约81％、至少为或约82％、至少

为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的氨基酸序列的其功能变体，或诸如SEQ ID NO:37所示或与SEQ ID NO:37具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的其功能变体，或诸如SEQ ID NO:47所示或与SEQ ID NO:47具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的其功能变体。在一些实施方案中，所述突变体NiV-G蛋白具有在野生型NiV-G蛋白(SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44)的N末端处或附近的20个氨基酸截短，诸如SEQ ID NO:13所示或与SEQ ID NO:13具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的其功能变体，或诸如SEQ ID NO:38所示或与SEQ ID NO:38具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的其功能变体，或诸如SEQ ID NO:48所示或与SEQ ID NO:48具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的其功能变体。在一些实施方案中，所述突变体NiV-G蛋白具有在野生型NiV-G蛋白(SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44)的N末端处或附近的25个氨基酸截短，诸如SEQ ID NO:14所示或与SEQ ID NO:14具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的其功能变体，或诸如SEQ ID NO:39所示或与SEQ ID NO:39具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约

93%、至少为或约94%、至少为或约95%、为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的其功能变体，或诸如SEQ ID NO:49所示或与SEQ ID NO:49具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的其功能变体。在一些实施方案中，所述突变体NiV-G蛋白具有在野生型NiV-G蛋白(SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44)的N末端处或附近的30个氨基酸截短，诸如SEQ ID NO:15所示或与SEQ ID NO:15具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的其功能变体，或诸如SEQ ID NO:40所示或与SEQ ID NO:40具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的其功能变体，或诸如SEQ ID NO:50所示或与SEQ ID NO:50具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的其功能变体。在一些实施方案中，所述突变体NiV-G蛋白具有在野生型NiV-G蛋白(SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44)的N末端处或附近的34个氨基酸截短，诸如SEQ ID NO:22所示或与SEQ ID NO:22具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的其功能变体，或诸如SEQ ID NO:53所示或与SEQ ID NO:53具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的其功能变体。在一些实施方案中，所述突变体NiV-G蛋白缺乏野生型NiV-G蛋白(SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44)的N末端胞质结构域，诸如SEQ ID NO:32所示或与SEQ ID NO:32具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、为或约96%、至少为或约97%、

至少为或约98％、或至少为或约99％序列同一性的其功能变体。

[0246] 在一些实施方案中，所述突变体G蛋白是具有SEQ ID NO:18或52所示序列的突变体HeV-G蛋白，或是具有与SEQ ID NO:18或52具有至少为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、至少为或约84％、为或约85％、至少为或约86％、至少为或约87％、为或约88％、至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列的其功能变体或生物活性部分。

[0247] 在一些实施方案中，所述G蛋白是突变体HeV-G蛋白，其是野生型HeV-G的生物活性部分。在一些实施方案中，所述生物活性部分是N末端截短的片段。在一些实施方案中，所述突变体HeV-G蛋白被截短并且缺乏在野生型HeV-G蛋白(SEQ ID NO:18或52)的N末端处或附近的多达5个连续氨基酸残基、在野生型HeV-G蛋白(SEQ ID NO:18或52)的N末端处或附近的多达6个连续氨基酸残基、在野生型HeV-G蛋白(SEQ ID NO:18或52)的N末端处或附近的多达7个连续氨基酸残基、在野生型HeV-G蛋白(SEQ ID NO:18或52)的N末端处或附近的多达8个连续氨基酸残基、在野生型HeV-G蛋白(SEQ ID NO:18或52)的N末端处或附近的多达9个连续氨基酸残基、在野生型HeV-G蛋白(SEQ ID NO:18或52)的N末端处或附近的多达10个连续氨基酸残基、在野生型HeV-G蛋白(SEQ ID NO:18或52)的N末端处或附近的多达11个连续氨基酸残基、在野生型HeV-G蛋白(SEQ ID NO:18或52)的N末端处或附近的多达12个连续氨基酸残基、在野生型HeV-G蛋白(SEQ ID NO:18或52)的N末端处或附近的多达13个连续氨基酸残基、在野生型HeV-G蛋白(SEQ ID NO:18或52)的N末端处或附近的多达14个连续氨基酸残基、在野生型HeV-G蛋白(SEQ ID NO:18或52)的N末端处或附近的多达15个连续氨基酸残基、在野生型HeV-G蛋白(SEQ ID NO:18或52)的N末端处或附近的多达16个连续氨基酸残基、在野生型HeV-G蛋白(SEQ ID NO:18或52)的N末端处或附近的多达17个连续氨基酸残基、在野生型HeV-G蛋白(SEQ ID NO:18或52)的N末端处或附近的多达18个连续氨基酸残基、在野生型HeV-G蛋白(SEQ ID NO:18或52)的N末端处或附近的多达19个连续氨基酸残基、在野生型HeV-G蛋白(SEQ ID NO:18或52)的N末端处或附近的多达20个连续氨基酸残基、在野生型HeV-G蛋白(SEQ ID NO:18或52)的N末端处或附近的多达21个连续氨基酸残基、在野生型HeV-G蛋白(SEQ ID NO:18或52)的N末端处或附近的多达22个连续氨基酸残基、在野生型HeV-G蛋白(SEQ ID NO:18或52)的N末端处或附近的多达23个连续氨基酸残基、在野生型HeV-G蛋白(SEQ ID NO:18或52)的N末端处或附近的多达24个连续氨基酸残基、在野生型HeV-G蛋白(SEQ ID NO:18或52)的N末端处或附近的多达25个连续氨基酸残基、在野生型HeV-G蛋白(SEQ ID NO:18或52)的N末端处或附近的多达26个连续氨基酸残基、在野生型HeV-G蛋白(SEQ ID NO:18或52)的N末端处或附近的多达27个连续氨基酸残基、在野生型HeV-G蛋白(SEQ ID NO:18或52)的N末端处或附近的多达28个连续氨基酸残基、在野生型HeV-G蛋白(SEQ ID NO:18或52)的N末端处或附近的多达29个连续氨基酸残基、在野生型HeV-G蛋白(SEQ ID NO:18或52)的N末端处或附近的多达30个连续氨基酸残基、在野生型HeV-G蛋白(SEQ ID NO:18或52)的N末端处或附近的多达31个连续氨基酸残基、在野生型HeV-G蛋白(SEQ ID NO:18或52)的N末端处或附近的多达32个连续氨基酸残基、在野生型HeV-G蛋白(SEQ ID NO:18或52)的N末端处或附近的多达33个连续氨基酸残基、在野生型HeV-G蛋白(SEQ ID NO:18或52)的N末端处或附近的多达34个连续氨基酸残

基、在野生型HeV-G蛋白(SEQ ID NO:18或52)的N末端处或附近的多达35个连续氨基酸残基、在野生型HeV-G蛋白(SEQ ID NO:18或52)的N末端处或附近的多达36个连续氨基酸残基、在野生型HeV-G蛋白(SEQ ID NO:18或52)的N末端处或附近的多达37个连续氨基酸残基、在野生型HeV-G蛋白(SEQ ID NO:18或52)的N末端处或附近的多达38个连续氨基酸残基、在野生型HeV-G蛋白(SEQ ID NO:18或52)的N末端处或附近的多达39个连续氨基酸残基、在野生型HeV-G蛋白(SEQ ID NO:18或52)的N末端处或附近的多达40个连续氨基酸残基、在野生型HeV-G蛋白(SEQ ID NO:18或52)的N末端处或附近的多达41个连续氨基酸残基、在野生型HeV-G蛋白(SEQ ID NO:18或52)的N末端处或附近的多达42个连续氨基酸残基、在野生型HeV-G蛋白(SEQ ID NO:18或52)的N末端处或附近的多达43个连续氨基酸残基、在野生型HeV-G蛋白(SEQ ID NO:18或52)的N末端处或附近的多达44个连续氨基酸残基、或在野生型HeV-G蛋白(SEQ ID NO:18或52)的N末端处或附近的多达45个连续氨基酸残基。在一些实施方案中,所述HeV-G蛋白是不含有胞质结构域的生物活性部分。在一些实施方案中,所述突变体HeV-G蛋白缺乏在野生型HeV-G蛋白的N末端胞质结构域(SEQ ID NO:18或52),诸如SEQ ID NO:33所示或与SEQ ID NO:33具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的其功能变体。

[0248] 在一些实施方案中,所述G蛋白或其功能活性变体或生物活性部分结合肝配蛋白B2或肝配蛋白B3。在一些方面,所述G蛋白具有SEQ ID NO:9、SEQ ID NO:18或SEQ ID NO:28、SEQ ID NO:29、SEQ ID NO:44、SEQ ID NO:30或SEQ ID NO:31中任一者所示的氨基酸序列,或是能够结合肝配蛋白B2或肝配蛋白B3的其功能活性变体或其生物活性部分。在一些实施方案中,所述功能活性变体或生物活性部分具有与SEQ ID NO:9、SEQ ID NO:18或SEQ ID NO:28、SEQ ID NO:29、SEQ ID NO:44、SEQ ID NO:30或SEQ ID NO:31具有至少约80%、至少约85%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的氨基酸序列或其功能活性变体或生物活性部分,并且保留与肝配蛋白B2或B3的结合。提及保留与肝配蛋白B2或B3的结合包括如下结合,所述结合为对应野生型G蛋白的结合水平或程度的至少或至少约5%,所述对应野生型G蛋白诸如SEQ ID NO:9、SEQ ID NO:18或SEQ ID NO:28、SEQ ID NO:29、SEQ ID NO:44、SEQ ID NO:30或SEQ ID NO:31所示或其功能活性变体或生物活性部分;对应野生型G蛋白的结合水平或程度的10%,所述对应野生型G蛋白诸如SEQ ID NO:9、SEQ ID NO:18或SEQ ID NO:28、SEQ ID NO:29、SEQ ID NO:44、SEQ ID NO:30或SEQ ID NO:31所示或其功能活性变体或生物活性部分;对应野生型G蛋白的结合水平或程度的15%,所述对应野生型G蛋白诸如SEQ ID NO:9、SEQ ID NO:18或SEQ ID NO:28、SEQ ID NO:29、SEQ ID NO:44、SEQ ID NO:30或SEQ ID NO:31所示或其功能活性变体或生物活性部分;对应野生型G蛋白的结合水平或程度的20%,所述对应野生型G蛋白诸如SEQ ID NO:9、SEQ ID NO:18或SEQ ID NO:28、SEQ ID NO:29、SEQ ID NO:44、SEQ ID NO:30或SEQ ID NO:31所示或其功能活性变体或生物活性部分;对应野生型G蛋白的结合水平或程度的25%,所述对应野生型G蛋白诸如SEQ ID NO:9、SEQ ID NO:18或SEQ ID NO:28、

SEQ ID NO:29、SEQ ID NO:44、SEQ ID NO:30或SEQ ID NO:31所示或其功能活性变体或生物活性部分;对应野生型G蛋白的结合水平或程度的30%,所述对应野生型G蛋白诸如SEQ ID NO:9、SEQ ID NO:18或SEQ ID NO:28、SEQ ID NO:29、SEQ ID NO:44、SEQ ID NO:30或SEQ ID NO:31所示或其功能活性变体或生物活性部分;对应野生型G蛋白的结合水平或程度的35%,所述对应野生型G蛋白诸如SEQ ID NO:9、SEQ ID NO:18或SEQ ID NO:28、SEQ ID NO:29、SEQ ID NO:44、SEQ ID NO:30或SEQ ID NO:31所示或其功能活性变体或生物活性部分;对应野生型G蛋白的结合水平或程度的40%,所述对应野生型G蛋白诸如SEQ ID NO:9、SEQ ID NO:18或SEQ ID NO:28、SEQ ID NO:29、SEQ ID NO:44、SEQ ID NO:30或SEQ ID NO:31所示或其功能活性变体或生物活性部分;对应野生型G蛋白的结合水平或程度的45%,所述对应野生型G蛋白诸如SEQ ID NO:9、SEQ ID NO:18或SEQ ID NO:28、SEQ ID NO:29、SEQ ID NO:44、SEQ ID NO:30或SEQ ID NO:31所示或其功能活性变体或生物活性部分;对应野生型G蛋白的结合水平或程度的50%,所述对应野生型G蛋白诸如SEQ ID NO:9、SEQ ID NO:18或SEQ ID NO:28、SEQ ID NO:29、SEQ ID NO:44、SEQ ID NO:30或SEQ ID NO:31所示或其功能活性变体或生物活性部分;对应野生型G蛋白的结合水平或程度的55%,所述对应野生型G蛋白诸如SEQ ID NO:9、SEQ ID NO:18或SEQ ID NO:28、SEQ ID NO:29、SEQ ID NO:44、SEQ ID NO:30或SEQ ID NO:31所示或其功能活性变体或生物活性部分;对应野生型G蛋白的结合水平或程度的60%,所述对应野生型G蛋白诸如SEQ ID NO:9、SEQ ID NO:18或SEQ ID NO:28、SEQ ID NO:29、SEQ ID NO:44、SEQ ID NO:30或SEQ ID NO:31所示或其功能活性变体或生物活性部分;对应野生型G蛋白的结合水平或程度的65%,所述对应野生型G蛋白诸如SEQ ID NO:9、SEQ ID NO:18或SEQ ID NO:28、SEQ ID NO:29、SEQ ID NO:44、SEQ ID NO:30或SEQ ID NO:31所示或其功能活性变体或生物活性部分;对应野生型G蛋白的结合水平或程度的70%,所述对应野生型G蛋白诸如SEQ ID NO:9、SEQ ID NO:18或SEQ ID NO:28、SEQ ID NO:29、SEQ ID NO:44、SEQ ID NO:30或SEQ ID NO:31所示或其功能活性变体或生物活性部分;诸如对应野生型G蛋白的结合水平或程度的至少或至少约75%,所述对应野生型G蛋白诸如SEQ ID NO:9、SEQ ID NO:18或SEQ ID NO:28、SEQ ID NO:29、SEQ ID NO:44、SEQ ID NO:30或SEQ ID NO:31所示或其功能活性变体或生物活性部分;诸如对应野生型G蛋白的结合水平或程度的至少或至少约80%,所述对应野生型G蛋白诸如SEQ ID NO:9、SEQ ID NO:18或SEQ ID NO:28、SEQ ID NO:29、SEQ ID NO:44、SEQ ID NO:30或SEQ ID NO:31所示或其功能活性变体或生物活性部分;诸如对应野生型G蛋白的结合水平或程度的至少或至少约85%,所述对应野生型G蛋白诸如SEQ ID NO:9、SEQ ID NO:18或SEQ ID NO:28、SEQ ID NO:29、SEQ ID NO:44、SEQ ID NO:30或SEQ ID NO:31所示或其功能活性变体或生物活性部分;诸如对应野生型G蛋白的结合水平或程度的至少或至少约90%,所述对应野生型G蛋白诸如SEQ I DNO:9、SEQ ID NO:18或SEQ ID NO:28、SEQ ID NO:29、SEQ ID NO:44、SEQ ID NO:30或SEQ ID NO:31所示或其功能活性变体或生物活性部分;诸如对应野生型蛋白的结合水平或程度的至少或至少约95%,所述对应野生型蛋白诸如SEQ ID NO:9、SEQ ID NO:18或SEQ ID NO:28、SEQ ID NO:29、SEQ ID NO:44、SEQ ID NO:30或SEQ ID NO:31所示或其功能活性变体或生物活性部分。在一些实施方案中,所述G蛋白是NiV-G或其功能活性变体或生物活性部分,并且结合肝配蛋白B2或肝配蛋白B3。在一些方面,所述NiV-G具有SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44所示的氨基酸序列,或是能够结合肝配蛋白B2或肝配

蛋白B3的其功能活性变体或其生物活性部分。在一些实施方案中，所述功能活性变体或生物活性部分具有与SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44具有至少约80％、至少约85％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列，并且保留与肝配蛋白B2或B3的结合。示例性生物活性部分包括缺乏胞质结构域的全部或部分(例如1个或多个，诸如1个至49个连续N末端氨基酸残基)的N末端截短变体，例如SEQ ID NO:10-15、35-40、45-50和32中任一者所示。提及保留与肝配蛋白B2或B3的结合包括如下结合，所述结合为对应野生型NiV-G(诸如SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44所示)的结合水平或程度的至少或至少约5％、对应野生型NiV-G(诸如SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44所示)的结合水平或程度的10％、对应野生型NiV-G(诸如SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44所示)的结合水平或程度的15％、对应野生型NiV-G(诸如SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44所示)的结合水平或程度的20％、对应野生型NiV-G(诸如SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44所示)的结合水平或程度的25％、对应野生型NiV-G(诸如SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44所示)的结合水平或程度的30％、对应野生型NiV-G(诸如SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44所示)的结合水平或程度的35％、对应野生型NiV-G(诸如SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44所示)的结合水平或程度的40％、对应野生型NiV-G(诸如SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44所示)的结合水平或程度的45％、对应野生型NiV-G(诸如SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44所示)的结合水平或程度的50％、对应野生型NiV-G(诸如SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44所示)的结合水平或程度的55％、对应野生型NiV-G(诸如SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44所示)的结合水平或程度的60％、对应野生型NiV-G(诸如SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44所示)的结合水平或程度的65％、对应野生型NiV-G(诸如SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44所示)的结合水平或程度的70％、诸如对应野生型NiV-G(诸如SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44所示)的结合水平或程度的至少或至少约75％、诸如对应野生型NIV-G(诸如SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44所示)的结合水平或程度的至少或至少约80％、诸如对应野生型NiV-G(诸如SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44所示)的结合水平或程度的至少或至少约85％、诸如对应野生型NiV-G(诸如SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44所示)的结合水平或程度的至少或至少约90％、或诸如对应野生型NiV-G(诸如SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44所示)的结合水平或程度的至少或至少约95％。

[0249] 在一些实施方案中，所述G蛋白是HeV-G或其功能活性变体或生物活性部分，并且结合肝配蛋白B2或肝配蛋白B3。在一些方面，所述HeV-G具有SEQ ID NO:18或52所示的氨基酸序列，或是能够结合肝配蛋白B2或肝配蛋白B3的其功能活性变体或其生物活性部分。在一些实施方案中，所述功能活性变体或生物活性部分具有与SEQ ID NO:18或52具有至少约80％、至少约85％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列，并且保留与肝配蛋白B2或B3的结合。示例性生物活性部分包括缺乏胞质结构域的全部或部分(例如1个或多个，诸如1个至49个连续N末端氨基

酸残基)的N末端截短变体,例如SEQ ID NO:33中任一者所示。关于保留与肝配蛋白B2或B3的结合包括如下结合,所述结合为对应野生型HeV-G(诸如SEQ ID NO:18或52所示)的结合水平或程度的至少或至少约5%、对应野生型HeV-G(诸如SEQ ID NO:18或52所示)的结合水平或程度的10%、对应野生型HeV-G(诸如SEQ ID NO:18或52所示)的结合水平或程度的15%、对应野生型HeV-G(诸如SEQ ID NO:18或52所示)的结合水平或程度的20%、对应野生型HeV-G(诸如SEQ ID NO:18或52所示)的结合水平或程度的25%、对应野生型HeV-G(诸如SEQ ID NO:18或52所示)的结合水平或程度的30%、对应野生型HeV-G(诸如SEQ ID NO:18或52所示)的结合水平或程度的35%、对应野生型HeV-G(诸如SEQ ID NO:18或52所示)的结合水平或程度的40%、对应野生型HeV-G(诸如SEQ ID NO:18或52所示)的结合水平或程度的45%、对应野生型HeV-G(诸如SEQ ID NO:18或52所示)的结合水平或程度的50%、对应野生型HeV-G(诸如SEQ ID NO:18或52所示)的结合水平或程度的55%、对应野生型HeV-G(诸如SEQ ID NO:18或52所示)的结合水平或程度的60%、对应野生型HeV-G(诸如SEQ ID NO:18或52所示)的结合水平或程度的65%、对应野生型HeV-G(诸如SEQ ID NO:18或52所示)的结合水平或程度的70%、诸如对应野生型HeV-G(诸如SEQ ID NO:18或52所示)的结合水平或程度的至少或至少约75%、诸如对应野生型NIV-G(诸如SEQ ID NO:18或52所示)的结合水平或程度的至少或至少约80%、诸如对应野生型HeV-G(诸如SEQ ID NO:18或52所示)的结合水平或程度的至少或至少约85%、诸如对应野生型HeV-G(诸如SEQ ID NO:18或52所示)的结合水平或程度的至少或至少约90%、或诸如对应野生型HeV-G(诸如SEQ ID NO:18或52所示)的结合水平或程度的至少或至少约95%。

[0250] 在一些实施方案中,所述G蛋白或其生物是突变体G蛋白,其表现出对野生型G蛋白的天然结合配偶体的降低的结合。在一些实施方案中,所述突变体G蛋白或其生物活性部分是野生型Niv-G的突变体,并且表现出与天然结合配偶体肝配蛋白B2或肝配蛋白B3中的一者或两者的降低的结合。在一些实施方案中,所述突变体G蛋白或生物活性部分(诸如突变体NiV-G蛋白)表现出与天然结合配偶体的降低的结合。在一些实施方案中,与肝配蛋白B2或肝配蛋白B3的降低的结合降低了大于为或约5%、为或约10%、为或约15%、为或约20%、为或约25%、为或约30%、为或约40%、为或约50%、为或约60%、为或约70%、为或约80%、为或约90%、或为或约100%。

[0251] 在一些实施方案中,本文所述的突变可以改善转导效率。在一些实施方案中,本文所述的突变允许特异性靶向不是肝配蛋白B2或肝配蛋白B3的其他期望的细胞类型。在一些实施方案中,本文所述的突变导致至少部分不能结合至少一种天然受体,这降低了与肝配蛋白B2或肝配蛋白B3中的至少一种的结合。在一些实施方案中,本文所述的突变干扰天然受体识别。

[0252] 在一些实施方案中,所述G蛋白含有在参与与肝配蛋白B2和肝配蛋白B3中的一者或两者相互作用的残基中的一个或多个氨基酸取代。在一些实施方案中,氨基酸取代对应于参考SEQ ID NO:28所示的编号的突变E501A、W504A、Q530A和E533A。

[0253] 在一些实施方案中,所述G蛋白是含有一个或多个氨基酸取代的G蛋白,所述一个或多个氨基酸取代选自由以下组成的组:E501A、W504A、Q530A和E533A(参考SEQ ID NO:28所示的编号)。在一些实施方案中,所述G蛋白是含有一个或多个氨基酸取代的突变体G蛋白,所述一个或多个氨基酸取代选自由以下组成的组:E501A、W504A、Q530A和E533A(参考

SEQ ID NO:28)，并且是含有N末端截短的其生物活性部分。在一些实施方案中，所述突变体NiV-G蛋白或其生物活性部分被截短，并且缺乏在野生型NiV-G蛋白(SEQ ID NO:28)的N末端处或附近的多达5个连续氨基酸残基、在野生型NiV-G蛋白(SEQ ID NO:28)的N末端处或附近的6个连续氨基酸残基、在野生型NiV-G蛋白(SEQ ID NO:28)的N末端处或附近的7个连续氨基酸残基、在野生型NiV-G蛋白(SEQ ID NO:28)的N末端处或附近的8个连续氨基酸残基、在野生型NiV-G蛋白(SEQ ID NO:28)的N末端处或附近的9个连续氨基酸残基、在野生型NiV-G蛋白(SEQ ID NO:28)的N末端处或附近的多达10个连续氨基酸残基、在野生型NiV-G蛋白(SEQ ID NO:28)的N末端处或附近的11个连续氨基酸残基、在野生型NiV-G蛋白(SEQ ID NO:28)的N末端处或附近的12个连续氨基酸残基、在野生型NiV-G蛋白(SEQ ID NO:28)的N末端处或附近的13个连续氨基酸残基、在野生型NiV-G蛋白(SEQ ID NO:28)的N末端处或附近的14个连续氨基酸残基、在野生型NiV-G蛋白(SEQ ID NO:28)的N末端处或附近的多达15个连续氨基酸残基、在野生型NiV-G蛋白(SEQ ID NO:28)的N末端处或附近的16个连续氨基酸残基、在野生型NiV-G蛋白(SEQ ID NO:28)的N末端处或附近的17个连续氨基酸残基、在野生型NiV-G蛋白(SEQ ID NO:28)的N末端处或附近的18个连续氨基酸残基、在野生型NiV-G蛋白(SEQ ID NO:28)的N末端处或附近的19个连续氨基酸残基、在野生型NiV-G蛋白(SEQ ID NO:28)的N末端处或附近的多达20个连续氨基酸残基、在野生型NiV-G蛋白(SEQ ID NO:28)的N末端处或附近的21个连续氨基酸残基、在野生型NiV-G蛋白(SEQ ID NO:28)的N末端处或附近的22个连续氨基酸残基、在野生型NiV-G蛋白(SEQ ID NO:28)的N末端处或附近的23个连续氨基酸残基、在野生型NiV-G蛋白(SEQ ID NO:28)的N末端处或附近的24个连续氨基酸残基、在野生型NiV-G蛋白(SEQ ID NO:28)的N末端处或附近的多达25个连续氨基酸残基、在野生型NiV-G蛋白(SEQ ID NO:28)的N末端处或附近的26个连续氨基酸残基、在野生型NiV-G蛋白(SEQ ID NO:28)的N末端处或附近的27个连续氨基酸残基、在野生型NiV-G蛋白(SEQ ID NO:28)的N末端处或附近的28个连续氨基酸残基、在野生型NiV-G蛋白(SEQ ID NO:28)的N末端处或附近的29个连续氨基酸残基、在野生型NiV-G蛋白(EQ ID NO:28)的N末端处或附近的多达30个连续氨基酸残基、在野生型NiV-G蛋白(SEQ ID NO:28)的N末端处或附近的多达31个连续氨基酸残基、在野生型NiV-G蛋白(SEQ ID NO:28)的N末端处或附近的32个连续氨基酸残基、在野生型NiV-G蛋白(SEQ ID NO:28)的N末端处或附近的33个连续氨基酸残基、在野生型NiV-G蛋白(SEQ ID NO:28)的N末端处或附近的34个连续氨基酸残基、在野生型NiV-G蛋白(SEQ ID NO:28)的N末端处或附近的35个连续氨基酸残基、在野生型NiV-G蛋白(EQ ID NO:28)的N末端处或附近的多达36个连续氨基酸残基、在野生型NiV-G蛋白(EQ ID NO:28)的N末端处或附近的多达37个连续氨基酸残基、在野生型NiV-G蛋白(EQ ID NO:28)的N末端处或附近的多达38个连续氨基酸残基、在野生型NiV-G蛋白(EQ ID NO:28)的N末端处或附近的多达39个连续氨基酸残基、或在野生型NiV-G蛋白(EQ ID NO:28)的N末端处或附近的多达40个连续氨基酸残基。

[0254] 在一些实施方案中，突变体NiV-G蛋白具有SEQ ID NO:16或51所示的氨基酸序列或与SEQ ID NO:16或51具有至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。在特定的实施方案中，所述G蛋白具有SEQ ID NO:16或51所示的氨基酸序列。

[0255] 在一些实施方案中，靶向包膜蛋白含有G蛋白或功能活性变体或生物活性部分和sdAb可变域，其中所述靶向包膜蛋白表现出对与野生型G蛋白的天然结合配偶体不同的另一种分子的增加的结合。在一些实施方案中，所述分子可以是在期望的靶细胞表面上表达的蛋白质。在一些实施方案中，与其他分子的增加的结合增加了大于为或约25％、为或约30％、为或约40％、为或约50％、为或约60％、为或约70％、为或约80％、为或约90％、或为或约100％。在特定的实施方案中，与其中赋予新的或不同的结合活性的野生型G蛋白的结合相比，所述结合赋予再靶向结合。

[0256] 2.结合结构域

[0257] 在一些实施方案中，结合结构域可以是结合靶细胞上的细胞表面分子的任何药剂。在一些实施方案中，结合结构域可以是抗体或抗体部分或片段。

[0258] 可以调节结合结构域以具有不同的结合强度。例如，具有不同结合强度的scFv和抗体可以用于改变嵌合附着蛋白对展示高或低的靶抗原量的细胞的融合活性，例如，具有不同亲和力的DARPin可以用于改变对展示高或低的靶抗原量的细胞的融合活性。还可以调节结合结构域以靶向靶配体上的不同区域，这将影响与展示所述靶标的细胞的融合速率。

[0259] 结合结构域可包含人源化抗体分子、完整IgA、IgG、IgE或IgM抗体；双特异性或多特异性抗体(例如，Zybodies®等)；抗体片段，诸如Fab片段、Fab'片段、F(ab')2片段、Fd'片段、Fd片段和分离的CDR或其集合；单链Fv；多肽-Fc融合物；单结构域抗体(例如，鲨鱼单结构域抗体诸如IgNAR或其片段)；骆驼科动物抗体；掩蔽抗体(例如，Probodies®)；Small Modular ImmunoPharmaceuticals(“SMIPsTM”)；单链或串联双抗体(TandAb®)；VHH；Anticalins®；Nanobodies®；微抗体(minibodies)；BiTE®s；锚蛋白重复蛋白或DARPINs®；Avimers®；DART；TCR样抗体；Adnectins®；Affilins®；Trans-bodies®；Affibodies®；TrimerX®；MicroProteins；Fynomers®；Centyrins®；和KALBITOR®s。靶向部分还可包括抗体或其抗原结合片段(例如，Fab、Fab'、F(ab')2、Fv片段、scFv抗体片段、二硫键连接的Fv(sdFv)、由VH和CH1结构域组成的Fd片段、线性抗体、单结构域抗体诸如sdAb(VL或VH)、纳米抗体或骆驼科动物VHH结构域)、抗原结合纤连蛋白III型(Fn3)支架诸如纤连蛋白多肽微抗体、配体、细胞因子、趋化因子或T细胞受体(TCR)。

[0260] 在一些实施方案中，所述结合结构域是单链分子。在一些实施方案中，所述结合结构域是单结构域抗体。在一些实施方案中，所述结合结构域是单链可变片段。在特定的实施方案中，所述结合结构域含有人的或人源化的抗体可变序列。

[0261] 在一些实施方案中，所述结合结构域是单结构域抗体。在一些实施方案中，所述单结构域抗体可以是人的或人源化的。在一些实施方案中，所述单结构域抗体或其部分是天然存在的。在一些实施方案中，所述单结构域抗体或其部分是合成的。

[0262] 在一些实施方案中，所述单结构域抗体是其互补决定区是单结构域多肽的一部分的抗体。在一些实施方案中，所述单结构域抗体是仅有重链的抗体可变域。在一些实施方案中，所述单结构域抗体不包含轻链。

[0263] 在一些实施方案中，缺少轻链的重链抗体被称为VHH。在一些实施方案中，所述单结构域抗体具有12-15kDa的分子量。在一些实施方案中，所述单结构域抗体包括骆驼科动

物抗体或鲨鱼抗体。在一些实施方案中，所述单结构域抗体分子来源于骆驼科物种中产生的抗体，所述骆驼科物种例如骆驼、美洲驼羊、单峰骆驼、羊驼、骆马和原驼。在一些实施方案中，所述单结构域抗体被称为免疫球蛋白新抗原受体(IgNAR)并且来源于软骨鱼。在一些实施方案中，通过将人或小鼠IgG的二聚体可变域分裂成单体并骆驼源化关键残基来产生所述单结构域抗体。

[0264] 在一些实施方案中，所述单结构域抗体可以从噬菌体展示文库产生。在一些实施方案中，所述噬菌体展示文库产生自用各种抗原免疫的骆驼科动物的VHH库，如以下文献所述：Arbabi等人，FEBS Letters，414，521-526(1997)；Lauwereys等人，EMBO J.，17，3512-3520(1998)；Decanniere等人，Structure，7，361-370(1999)。在一些实施方案中，产生包含未免疫的骆驼科动物的抗体片段的噬菌体展示文库。在一些实施方案中，通过将多样性引入一个或多个支架而合成产生单结构域抗体人单结构域抗体文库。

[0265] 在一些实施方案中，所述单结构域抗体的C末端与所述G蛋白或其生物活性部分的C末端连接。在一些实施方案中，所述单结构域抗体的N末端暴露在脂质双层的外表面上。在一些实施方案中，所述单结构域抗体的N末端结合靶细胞的细胞表面分子。在一些实施方案中，所述单结构域抗体特异性地结合存在于靶细胞上的细胞表面分子。在一些实施方案中，所述细胞表面分子是蛋白质、聚糖、脂质或低分子量分子。

[0266] 在一些实施方案中，靶细胞的细胞表面分子是抗原或其部分。在一些实施方案中，所述单结构域抗体或其部分是具有能够选择性地结合特定抗原的单体单结构域抗原结合/识别结构域的抗体。在一些实施方案中，所述单结构域抗体结合存在于靶细胞上的抗原。

[0267] 示例性细胞包括多形核细胞(也称为PMN、PML、PMNL或粒细胞)、干细胞、胚胎干细胞、神经干细胞、间充质干细胞(MSC)、造血干细胞(HSC)、人肌源性干细胞、肌肉来源的干细胞(MuStem)、胚胎干细胞(ES或ESC)、角膜缘上皮干细胞、心肌源性干细胞、心肌细胞、祖细胞、免疫效应细胞、淋巴细胞、巨噬细胞、树突细胞、自然杀伤细胞、T细胞、细胞毒性T淋巴细胞、同种细胞、常驻心脏细胞、诱导多能干细胞(iPS)、脂肪来源的或表型修饰的干细胞或祖细胞、CD133+细胞、醛脱氢酶阳性细胞(ALDH+)、脐带血(UCB)细胞、外周血干细胞(PBSC)、神经元、神经祖细胞、胰腺β细胞、胶质细胞或肝细胞。

[0268] 在一些实施方案中，所述靶细胞是靶组织的细胞。所述靶组织可包括肝、肺、心脏、脾、胰腺、胃肠道、肾、睾丸、卵巢、脑、生殖器官、中枢神经系统、外周神经系统、骨骼肌、内皮、内耳或眼睛。

[0269] 在一些实施方案中，所述靶细胞是肌肉细胞(例如骨骼肌细胞)、肾细胞、肝脏细胞(例如肝细胞)或心肌细胞(cadiac cell)(例如心肌细胞(cardiomyocyte))。在一些实施方案中，所述靶细胞是心脏细胞(例如心肌细胞(例如休眠心肌细胞))、成肝细胞(例如胆管成肝细胞)、上皮细胞、T细胞(例如幼稚T细胞)、巨噬细胞(例如肿瘤浸润巨噬细胞)或成纤维细胞(例如心脏成纤维细胞)。

[0270] 在一些实施方案中，所述靶细胞是肿瘤浸润淋巴细胞、T细胞、赘生性或肿瘤细胞、病毒感染细胞、干细胞、中枢神经系统(CNS)细胞、造血干细胞(HSC)、肝细胞或完全分化细胞。在一些实施方案中，所述靶细胞选自CD3+T细胞、CD4+T细胞、CD8+T细胞、肝细胞、造血干细胞、CD34+造血干细胞、CD105+造血干细胞、CD117+造血干细胞、CD105+内皮细胞、B细胞、CD20+B细胞、CD19+B细胞、癌细胞、CD133+癌细胞、EpCAM+癌细胞、CD19+癌细胞、Her2/Neu+

癌细胞、GluA2+神经元、GluA4+神经元、NKG2D+自然杀伤细胞、SLC1A3+星形胶质细胞、SLC7A10+脂肪细胞或CD30+肺上皮细胞。

[0271] 在一些实施方案中，所述靶细胞是抗原呈递细胞、MHC II类+细胞、专职性抗原呈递细胞、非典型抗原呈递细胞、巨噬细胞、树突细胞、骨髓树突细胞、浆细胞样树突细胞、CD11c+细胞、CD11b+细胞、脾细胞、B细胞、肝细胞、内皮细胞或非癌细胞。

[0272] 在一些实施方案中，所述细胞表面分子是CD8、CD4、脱唾液酸糖蛋白受体2(ASGR2)、跨膜4L6家族成员5(TM4SF5)、低密度脂蛋白受体(LDLR)或脱唾液酸糖蛋白1(ASGR1)中的任一种。

[0273] 在一些实施方案中，所述G蛋白或其功能活性变体或生物活性部分与sdAb可变域直接连接。在一些实施方案中，所述靶向包膜蛋白是具有以下结构的融合蛋白：(N'-单结构域抗体-C')-(C'-G蛋白-N')。

[0274] 在一些实施方案中，所述G蛋白或其功能活性变体或生物活性部分通过接头间接连接至sdAb可变域。在一些实施方案中，所述接头是肽接头。在一些实施方案中，所述接头是化学接头。

[0275] 在一些实施方案中，所述接头是肽接头，并且所述靶向包膜蛋白是融合蛋白，其含有通过肽接头与sdAb可变域连接的G蛋白或其功能活性变体或生物活性部分。在一些实施方案中，所述靶向包膜蛋白是具有以下结构的融合蛋白：(N'-单结构域抗体-C')-接头-(C'-G蛋白-N')。

[0276] 在一些实施方案中，所述肽接头为多达65个氨基酸的长度。在一些实施方案中，所述肽接头包含从或从约2至65个氨基酸、2至60个氨基酸、2至56个氨基酸、2至52个氨基酸、2至48个氨基酸、2至44个氨基酸、2至40个氨基酸、2至36个氨基酸、2至32个氨基酸、2至28个氨基酸、2至24个氨基酸、2至20个氨基酸、2至18个氨基酸、2至14个氨基酸、2至12个氨基酸、2至10个氨基酸、2至8个氨基酸、2至6个氨基酸、6至65个氨基酸、6至60个氨基酸、6至56个氨基酸、6至52个氨基酸、6至48个氨基酸、6至44个氨基酸、6至40个氨基酸、6至36个氨基酸、6至32个氨基酸、6至28个氨基酸、6至24个氨基酸、6至20个氨基酸、6至18个氨基酸、6至14个氨基酸、6至12个氨基酸、6至10个氨基酸、6至8个氨基酸、8至65个氨基酸、8至60个氨基酸、8至56个氨基酸、8至52个氨基酸、8至48个氨基酸、8至44个氨基酸、8至40个氨基酸、8至36个氨基酸、8至32个氨基酸、8至28个氨基酸、8至24个氨基酸、8至20个氨基酸、8至18个氨基酸、8至14个氨基酸、8至12个氨基酸、8至10个氨基酸、10至65个氨基酸、10至60个氨基酸、10至56个氨基酸、10至52个氨基酸、10至48个氨基酸、10至44个氨基酸、10至40个氨基酸、10至36个氨基酸、10至32个氨基酸、10至28个氨基酸、10至24个氨基酸、10至20个氨基酸、10至18个氨基酸、10至14个氨基酸、10至12个氨基酸、12至65个氨基酸、12至60个氨基酸、12至56个氨基酸、12至52个氨基酸、12至48个氨基酸、12至44个氨基酸、12至40个氨基酸、12至36个氨基酸、12至32个氨基酸、12至28个氨基酸、12至24个氨基酸、12至20个氨基酸、12至18个氨基酸、12至14个氨基酸、14至65个氨基酸、14至60个氨基酸、14至56个氨基酸、14至52个氨基酸、14至48个氨基酸、14至44个氨基酸、14至40个氨基酸、14至36个氨基酸、14至32个氨基酸、14至28个氨基酸、14至24个氨基酸、14至20个氨基酸、14至18个氨基酸、18至65个氨基酸、18至60个氨基酸、18至56个氨基酸、18至52个氨基酸、18至48个氨基酸、18至44个氨基酸、18至40个氨基酸、18至36个氨基酸、18至32个氨基酸、18至28个氨基酸、18至24个氨基

酸、18至20个氨基酸、20至65个氨基酸、20至60个氨基酸、20至56个氨基酸、20至52个氨基酸、20至48个氨基酸、20至44个氨基酸、20至40个氨基酸、20至36个氨基酸、20至32个氨基酸、20至28个氨基酸、20至26个氨基酸、20至24个氨基酸、24至65个氨基酸、24至60个氨基酸、24至56个氨基酸、24至52个氨基酸、24至48个氨基酸、24至44个氨基酸、24至40个氨基酸、24至36个氨基酸、24至32个氨基酸、24至30个氨基酸、24至28个氨基酸、28至65个氨基酸、28至60个氨基酸、28至56个氨基酸、28至52个氨基酸、28至48个氨基酸、28至44个氨基酸、28至40个氨基酸、28至36个氨基酸、28至34个氨基酸、28至32个氨基酸、32至65个氨基酸、32至60个氨基酸、32至56个氨基酸、32至52个氨基酸、32至48个氨基酸、32至44个氨基酸、32至40个氨基酸、32至38个氨基酸、32至36个氨基酸、36至65个氨基酸、36至60个氨基酸、36至56个氨基酸、36至52个氨基酸、36至48个氨基酸、36至44个氨基酸、36至40个氨基酸、40至65个氨基酸、40至60个氨基酸、40至56个氨基酸、40至52个氨基酸、40至48个氨基酸、40至44个氨基酸、44至65个氨基酸、44至60个氨基酸、44至56个氨基酸、44至52个氨基酸、44至48个氨基酸、48至65个氨基酸、48至60个氨基酸、48至56个氨基酸、48至52个氨基酸、50至65个氨基酸、50至60个氨基酸、50至56个氨基酸、50至52个氨基酸、54至65个氨基酸、54至60个氨基酸、54至56个氨基酸、58至65个氨基酸、58至60个氨基酸、或60至65个氨基酸。在一些实施方案中，所述肽接头是长度为3个、4个、5个、6个、7个、8个、9个、10个、11个、12个、13个、14个、15个、16个、17个、18个、19个、20个、21个、22个、23个、24个、25个、26个、27个、28个、29个、30个、31个、32个、33个、34个、35个、36个、37个、38个、39个、40个、41个、42个、43个、44个、45个、46个、47个、48个、49个、50个、51个、52个、53个、54个、55个、56个、57个、58个、59个、60个、61个、62个、63个、64个或65个氨基酸的多肽。

[0277] 在特定的实施方案中，所述接头是柔性肽接头。在一些此类实施方案中，所述接头为1-20个氨基酸，诸如主要由甘氨酸构成的1-20个氨基酸。在一些实施方案中，所述接头为1-20个氨基酸，诸如主要由甘氨酸和丝氨酸构成的1-20个氨基酸。在一些实施方案中，所述接头是含有氨基酸甘氨酸和丝氨酸的柔性肽接头，称为GS-接头。在一些实施方案中，所述肽接头包括序列GS、GGS、GGGGS(SEQ IDNO:43)、GGGGGS(SEQ ID NO:41)或它们的组合。在一些实施方案中，所述多肽接头具有序列(GGS)n，其中n为1至10。在一些实施方案中，所述多肽接头具有序列(GGGGS)n(SEQ ID NO:42)，其中n为1至10。在一些实施方案中，所述多肽接头具有序列(GGGGGS)n(SEQ ID NO:27)，其中n为1至6。

[0278] 3.多核苷酸

[0279] 本文提供了包含编码靶向包膜蛋白的核酸序列的多核苷酸。在一些实施方案中，所述多核苷酸包含编码G蛋白或其生物活性部分的核酸序列。在一些实施方案中，所述多核苷酸还包含编码单结构域抗体(sdAb)可变域或其生物活性部分的核酸序列。所述多核苷酸可包含编码上文所述的任何靶向包膜蛋白的核苷酸序列。所述多核苷酸可以是合成的核酸。还提供了含有任何所提供的多核苷酸的表达载体。

[0280] 在一些任何实施方案中，天然或合成核酸的表达通常通过将编码感兴趣的基因的核酸可操作地连接到启动子并将构建体整合到表达载体中来实现。在一些实施方案中，载体可以适合于在真核生物中复制和整合。在一些实施方案中，克隆载体含有可用于表达期望的核酸序列的转录和翻译终止子、起始序列和启动子。在一些任何实施方案中，质粒包含适合于在细胞中表达的启动子。

[0281] 在一些实施方案中,所述多核苷酸含有至少一个启动子,其可操作地连接以控制含有所述G蛋白和单结构域抗体(sdAb)可变域的靶向包膜蛋白的表达。对于靶向包膜蛋白的表达,每个启动子中的至少一个模块起定位RNA合成的起始位点的作用。其最著名的实例是TATA盒,但在一些启动子中缺乏TATA盒,诸如哺乳动物末端脱氧核苷酸转移酶基因的启动子和SV40基因的启动子,覆盖起始位点本身的离散元件有助于固定起始位置。

[0282] 在一些实施方案中,另外的启动子元件例如增强子调节转录起始的频率。在一些实施方案中,另外的启动子元件位于起始位点上游30-110bp的区域,但是最近已显示许多启动子也含有起始位点下游的功能元件。在一些实施方案中,启动子元件之间的间隔通常是柔性的,使得当元件相对于彼此反转或移动时保持启动子功能。在一些实施方案中,胸苷激酶(tk)启动子,启动子元件之间的间隔可以在活性开始下降之前增加到相隔50bp。在一些实施方案中,取决于启动子,单个元件可以协同地或独立地起作用以激活转录。

[0283] 启动子可以是与基因或多核苷酸序列天然相关的启动子,这可以通过分离位于编码区段和/或外显子上游的5′非编码序列来获得。这种启动子可以被称为“内源的”。类似地,增强子可以是与多核苷酸序列天然相关的增强子,其位于该序列的下游或上游。可替代地,通过将编码多核苷酸区段置于重组或异源启动子的控制下将获得某些优点,所述重组或异源启动子是指在其天然环境中通常不与多核苷酸序列相关的启动子。重组或异源增强子也指在其天然环境中通常不与多核苷酸序列相关的增强子。此类启动子或增强子可包括其他基因的启动子或增强子,和从任何其他原核、病毒或真核细胞分离的启动子或增强子,以及非“天然存在的”启动子或增强子,即含有不同转录调节区的不同元件和/或改变表达的突变。除了合成产生启动子和增强子的核酸序列之外,还可以使用重组克隆和/或核酸扩增技术(包括PCR)结合本文公开的组合物来产生序列(美国专利号4,683,202和5,928,906)。

[0284] 在一些实施方案中,合适的启动子是立即早期巨细胞病毒(CMV)启动子序列。在一些实施方案中,启动子序列是能够驱动与其可操作地连接的任何多核苷酸序列高水平表达的强组成型启动子序列。在一些实施方案中,合适的启动子是延伸生长因子-1a(EF-1 a)。在一些实施方案中,也可以使用其他组成型启动子序列,包括但不限于猿猴病毒40(SV40)早期启动子、小鼠乳腺肿瘤病毒(MMTV)、人免疫缺陷病毒(HIV)长末端重复序列(LTR)启动子、MoMuLV启动子、禽白血病病毒启动子、爱泼斯坦-巴尔病毒(Epstein-Barr virus)立即早期启动子、劳斯肉瘤病毒(Rous sarcoma virus)启动子、以及人基因启动子(诸如但不限于肌动蛋白启动子、肌球蛋白启动子、血红蛋白启动子和肌酸激酶启动子)。

[0285] 在一些实施方案中,启动子是诱导型启动子。在一些实施方案中,诱导型启动子提供了分子开关,当需要表达时,所述分子开关能够开启其可操作地连接的多核苷酸序列的表达,或者当不需要表达时,关闭所述表达。在一些实施方案中,诱导型启动子包含金属硫蛋白启动子、糖皮质激素启动子、孕酮启动子和四环素启动子。

[0286] 在一些实施方案中,外源控制的诱导型启动子可以用于调节G蛋白和单结构域抗体(sdAb)可变域的表达。例如,辐射诱导型启动子、热诱导型启动子和/或药物诱导型启动子可以用于在例如靶向区域中选择性地驱动转基因表达。在此类实施方案中,转基因表达的位置、持续时间和水平可以通过施用外源诱导源来调节。

[0287] 在一些实施方案中,使用药物诱导型启动子调节含有G蛋白和单结构域抗体

(sdAb)可变域的靶向包膜蛋白的表达。例如,在一些情况下,启动子、增强子或反式激活因子包含Lac操纵子序列、四环素操纵子序列、半乳糖操纵子序列、多西环素操纵子序列、雷帕霉素操纵子序列、他莫昔芬操纵子序列或激素反应性操纵子序列或其类似物。在一些情形中,诱导型启动子包含四环素反应元件(TRE)。在一些实施方案中,诱导型启动子包含雌激素反应元件(ERE),其可以在他莫昔芬的存在下激活基因表达。在一些情形中,药物诱导型元件诸如TRE可以与所选启动子组合以在药物诸如多西环素的存在下增强转录。在一些实施方案中,药物诱导型启动子是小分子诱导型启动子。

[0288] 可以修饰任何所提供的多核苷酸以去除CpG基序和/或优化特定物种(诸如人、犬、猫科动物、马科动物、绵羊、牛等物种)中用于翻译的密码子。在一些实施方案中,多核苷酸针对人类密码子使用进行优化(即,人类密码子优化的)。在一些实施方案中,修饰多核苷酸以去除CpG基序。在其他实施方案中,修饰所提供的多核苷酸以去除CpG基序并进行密码子优化,诸如人类密码子优化。密码子优化和CpG基序检测和修饰的方法是熟知的。通常,多核苷酸优化增强转基因表达,增加转基因稳定性并保留所编码多肽的氨基酸序列。

[0289] 为了评价靶向包膜蛋白的表达,待引入细胞的表达载体还可以含有选择标记基因或报告基因或二者,以促进表达颗粒例如病毒颗粒的鉴定和选择。在其他实施方案中,选择标记可以携带在单独的DNA片段上并用于共转染程序中。选择标记和报告基因两者都可以与适当的调节序列侧接,以便能够在宿主细胞中表达。有用的选择标记是本领域已知的,并且包括例如抗生素抗性基因,诸如neo等。

[0290] 将报告基因用于鉴定潜在转染的细胞和用于评价调节序列的功能性。编码容易测定的蛋白质的报告基因是本领域熟知的。通常,报告基因是不存在于受体生物体或组织中或不由受体生物体或组织表达并且编码其表达由一些容易检测的特性例如酶活性证明的蛋白质的基因。在将DNA引入受体细胞之后的合适时间测定报告基因的表达。

[0291] 合适的报告基因可包括编码萤光素酶、β半乳糖苷酶、氯霉素乙酰转移酶、分泌的碱性磷酸酶或绿色荧光蛋白基因的基因(参见例如Ui-Tei等人,2000,FEBS Lett.479:79-82)。合适的表达系统是熟知的,并且可以使用熟知的技术来制备或商购获得。可以使用独特的内部限制性位点或通过部分消化非独特的限制性位点来产生内部缺失构建体。然后可以将构建体转染到显示高水平的期望的多核苷酸和/或多肽表达的细胞中。通常,具有显示最高水平的报告基因表达的最小5′侧翼区的构建体被鉴定为启动子。此类启动子区可以与报告基因连接并用于评价药剂调节启动子驱动的转录的能力。

[0292] B.融合剂(例如亨尼帕病毒F蛋白)

[0293] 在一些实施方案中,所述靶向脂质颗粒包含一种或多种融合剂。在一些实施方案中,所述靶向脂质颗粒含有外源性或过表达的融合剂。在一些实施方案中,所述融合剂被置于脂质双层中。在一些实施方案中,所述融合剂促进靶向脂质颗粒与膜的融合。在一些实施方案中,所述膜是浆细胞膜。

[0294] 在一些实施方案中,所述融合剂包含基于蛋白质的融合剂、基于脂质的融合剂和基于化学品的融合剂。在一些实施方案中,所述靶向脂质颗粒包含含有蛋白质融合剂的第一融合剂和含有脂质融合剂或化学品融合剂的第二融合剂。在一些实施方案中,所述融合剂结合靶细胞表面上的融合剂结合配偶体。

[0295] 在一些实施方案中,所述融合剂包含具有疏水性融合肽结构域的蛋白质。在一些

实施方案中，所述融合剂包含亨尼帕病毒F蛋白分子或其生物活性部分。在一些实施方案中，亨尼帕病毒F蛋白是亨德拉(Hev)病毒F蛋白、尼帕(NiV)病毒F蛋白、雪松(CedPV)病毒F蛋白、墨江病毒F蛋白或蝙蝠副粘病毒F蛋白或它们的生物活性部分。

[0296] 表4提供了F蛋白的非限制性实例。在一些实施方案中，F蛋白分子或其生物活性部分的N末端疏水性融合肽结构域暴露在脂质双层的外侧。

[0297] 亨尼帕病毒的F蛋白被编码为含有信号肽(例如对应于SEQ IDNO:1的氨基酸残基1-26)的F_0前体。在切割信号肽后，成熟F_0(例如SEQ ID NO:2)被转运至细胞表面，然后被组织蛋白酶L内吞和切割(例如在SEQ ID NO:1的氨基酸109-110之间)为成熟融合剂亚基F1(例如对应于SEQ ID NO:1的氨基酸110-546;如SEQ ID NO:4所示)和F2(例如对应于SEQ ID NO:1的氨基酸残基27-109;如SEQ IDNO:3所示)。F1和F2亚基通过二硫键结合并循环回细胞表面。F1亚基含有位于F1亚基的N末端的融合肽结构域(例如，对应于SEQ IDNO:1的氨基酸110-129)，在这里其能够插入细胞膜以驱动融合。在特定的情况下，融合活性被F蛋白与G蛋白的缔合所阻断，直到G与靶分子接合，导致其从F解离并暴露融合肽以介导膜融合。

[0298] 在不同的亨尼帕病毒物种中，F蛋白的序列和活性是高度保守的。例如，NiV和HeV病毒的F蛋白共享89％的氨基酸序列同一性。此外，在一些情况下，亨尼帕病毒F蛋白表现出与来自其他物种的G蛋白的相容性，以触发融合(Brandel-Tretheway等人Journal of Virology.2019.93(13):e00577-19)。在一些方面或所提供的再靶向脂质颗粒中，F蛋白与G蛋白是异源的，即F和G蛋白或生物活性部分来自不同的亨尼帕病毒物种。例如，F蛋白来自亨德拉病毒，并且G蛋白来自尼帕病毒。在其他方面，F蛋白可以是含有来自不同亨尼帕病毒物种的F蛋白区域的嵌合F蛋白。在一些实施方案中，将F蛋白的氨基酸残基的区域从亨尼帕病毒的一个物种转换为另一个物种可以导致与包含氨基酸插入的物种的G蛋白融合。(Brandel-Tretheway等人2019)。在一些情况下，嵌合F蛋白含有来自一种亨尼帕病毒物种的细胞外结构域和来自不同亨尼帕病毒物种的跨膜和/或细胞质结构域。例如，F蛋白含有亨德拉病毒的细胞外结构域和尼帕病毒的跨膜/细胞质结构域。本文所公开的F蛋白序列主要公开为包括N末端信号序列的表达序列。由于这样的N末端信号序列通常被共翻译或翻译后切割，所以本文所公开的所有F蛋白序列的成熟蛋白序列也被认为缺乏N末端信号序列。

[0299] 表4.亨尼帕病毒F序列簇。第1列，Genbank ID包括作为簇的质心序列的病毒整个基因组序列的Genbank ID。第2列，CDS的核苷酸提供与整个基因组中基因的CDS对应的核苷酸。第3列，完整基因名称，提供基因的完整名称，包括Genbank ID、病毒物种、毒株和蛋白质名称。尼帕病毒F蛋白与亨德拉病毒的F蛋白为>80％相同，并且发现在相同的序列簇中。第4列，序列，提供基因的氨基酸序列。第5列，序列/簇#，提供了与该质心序列聚类的序列的数量。第6列提供所描述的序列的SEQ ID号。

[0300]

Genbank ID	CDS的核苷酸	全基因名称	序列	序列/簇#	SEQ ID NO	SEQ ID NO (没有信号序列)
AF017149	6618-8258	gb:AF017149|生物体：亨德拉病毒|毒株名	MATQEVRLKCLLCGIIVLVLSLEGL GILHYEKLSKIGLVKGITRKYKIKSN PLTKDIVIKMIPNVSNVSKCTGTVME NYKSRLTGILSPIKGAIELYNNNTHD LVGDVKLAGVVMAGIAIGIATAAQI TAGVALYEAMKNADNINKLKSSIES	29	17	59

[0301]

			称：未知-AF017149|蛋白质名称：融合|基因符号:F	TNEAVVKLQETAEKTVYVLTALQDYINTNLVPTIDQISCKQTELALDLALSKYLSDLLFVFGPNLQDPVSNSMTIQAISQAFGGNYETLLRTLGYATEDFDDLLESDSIAGQIVYVDLSSYYIIVRVYFPILTEIQQAYVQELLPVSFNNDNSEWISIVPNFVLIRNTLISNIEVKYCLITKKSVICNQDYATPMTASVRECLTGSTDKCPRELVVSSHVPRFALSGGVLFANCISVTCQCQTTGRAISQSGEQTLLMIDNTTCTTVVLGNIIISLGKYLGSINYNSESIAVGPPVYTDKVDISSQISSMNQSLQQSKDYIKEAQKILDTVNPSLISMLSMIILYVLSIAALCIGLITFISFVIVEKKRGNYSRLDDRQVRPVSNGDLYYIGT			
Q9IH63		附加于簇中：sp|Q9IH63|FUS_NIPAV 融合糖蛋白F0 OS=尼帕病毒	MVVILDKRCYCNLLILILMISECSVGILHYEKLSKIGLVKGVTRKYKIKSNPLTKDIVIKMIPNVSNMSQCTGSVMENYKTRLNGILTPIKGALEIYKNNTHDLVGDVRLAGVIMAGVAIGIATAAQITAGVALYEAMKNADNINKLKSSIESTNEAVVKLQETAEKTVYVLTALQDYINTNLVPTIDKISCKQTELSLDLALSKYLSDLLFVFGPNLQDPVSNSMTIQAISQAFGGNYETLLRTLGYATEDFDDLLESDSITGQIIYVDLSSYYIIVRVYFPILTEIQQAYIQELLPVSFNNDNSEWISIVPNFILVRNTLISNIEIGFCLITKRSVICNQDYATPMTNNMRECLTGSTEKCPRELVVSSHVPRFALSNGVLFANCISVTCQCQTTGRAISQSGEQTLLMIDNTTCPTAVLGNVIISLGKYLGSVNYNSEGIAIGPPVFTDKVDISSQISSMNQSLQQSKDYIKEAQRLLDTVNPSLISMLSMIILYVLSIASLCIGLITFISFIIVEKKRNTYSRLEDRRVRPTSSGDLYYIGT	1	2
JQ001776	6129-8166	gb:JQ001776:6129-8166|生物体：雪松病毒|毒株名称:CG1a|蛋白质名称：融合糖蛋白|基因符号：F	MSNKRTTVLIIISYTLFYLNNAAIVGFDFDKLNKIGVVQGRVLNYKIKGDPMTKDLVLKFIPNIVNITECVREPLSRYNETVRRLLLPIHNMLGLYLNNTNAKMTGLMIAGVIMGGIAIGIATAAQITAGFALYEAKKNTENIQKLTDSIMKTQDSIDKLTDSVGTSILILNKLQTYINNQLVPNLELLSCRQNKIEFDLMLTKYLVDLMTVIGPNINNPVNKDMTIQSLSLLFDGNYDIMMSELGYTPQDFLDLIESKSITGQIIYVDMENLYVVIRTYLPTLIEVPDAQIYEFNKITMSSNGGEYLSTIPNFILIRGNYMSNIDVATCYMTKASVICNQDYSLPMSQNLRSCYQGETEYCPVEAVIASHSPRFALTNGVIFANCINTICRCQDNGKTITQNINQFVSMI	3	24	57

[0302]

			DNSTCNDVMVDKFTIKVGKYMGRK DINNINIQIGPQIIIDKVDLSNEINKM NQSLKDSIFYLREAKRILDSVNISLIS PSVQLFLIIISVLSFIILLIIIVYLYCKSK HSYKYNKFIDDPDYYNDYKRERING KASKSNNIYYVGD			
NC_025352	5950-8712	gb:NC_025352:5950-8712|生物体:墨江病毒|毒株名称:Tongguan1|蛋白质名称:融合蛋白|基因符号:F	MALNKNMFSSLFLGYLLVYATTVQ SSIHYDSLSKVGVIKGLTYNYKIKGS PSTKLMVVKLIPNIDSVKNCTQKQY DEYKNLVRKALEPVKMAIDTMLNN VKSGNNKYRFAGAIMAGVALGVAT AATVTAGIALHRSNENAQAIANMKS AIQNTNEAVKQLQLANKQTLAVIDT IRGEINNNIIPVINQLSCDTIGLSVGIR LTQYYSEIITAFGPALQNPVNTRITIQ AISSVFNGNFDELLKIMGYTSGDLYE ILHSELIRGNIIDVDVDAGYIALEIEFP NLTLVPNAVVQELMPISYNIDGDEW VTLVPRFVLTRTTLLSNIDTSRCTITD SSVICDNDYALPMSHELIGCLQGDTS KCAREKVVSSYVPKFALSDGLVYA NCLNTICRCMDTDTPISQSLGATVSL LDNKRCSVYQVGDVLISVGSYLGD GEYNADNVELGPPIVIDKIDIGNQLA GINQTLQEAEDYIEKSEEFLKGVNPS IITLGSMVVLYIFMILIAIVSVIALVLS IKLTVKGNVVRQQFTYTQHVPSME NINYVSH	2	25	60
NC_025256	6865-8853	gb:NC_025256:6865-8853|生物体:蝙蝠副粘病毒Eid_hel/GH-M74a/GHA/2009|毒株名称:BatPV/Eid_hel/GH-M74a/GHA/2009|蛋白质名称:融合蛋白|基因符号:F	MKKKTDNPTISKRGHNHSRGIKSRA LLRETDNYSNGLIVENLVRNCHHPS KNNLNYTKTQKRDSTIPYRVEERKG HYPKIKHLIDKSYKHIKRGKRRNGH NGNIITIILLLILILKTQMSEGAIHYET LSKIGLIKGITREYKVKGTPSSKDIVI KLIPNVTGLNKCTNISMENYKEQLD KILIPINNIIELYANSTKSAPGNARFA GVIIAGVALGVAAAAQITAGIALHE ARQNAERINLLKDSISATNNAVAEL QEATGGIVNVITGMQDYINTNLVPQI DKLQCSQIKTALDISLSQYYSEILTVF GPNLQNPVTTSMSIQAISQSFGGNID LLLNLLGYTANDLLDLLESKSITGQI TYINLEHYFMVIRVYYPIMTTISNAY VQELIKISFNVDGSEWVSLVPSYILIR NSYLSNIDISECLITKNSVICRHDFAM PMSYTLKECLTGDTEKCPREAVVTS YVPRFAISGGVIYANCLSTTCQCYQT GKVIAQDGSQTLMMIDNQTCSIVRIE EILISTGKYLGSQEYNTMHVSVGNP VFTDKLDITSQISNINQSIEQSKFYLD KSKAILDKINLNLIGSVPISILFIIAILS LILSIITFVIVMIIVRRYNKYTPLINSD PSSRRSTIQDVYIIPNPGEHSIRSAAR SIDRDRD	2	26	58

[0303] 在一些实施方案中,F蛋白由编码SEQ ID NO:1、2、17、24、25、26或57-60中任一者所示的序列的核苷酸序列编码,或者是与SEQ ID NO:1、2、17、24、25、26或57-60中任一者至少为或约80%、至少为或约85%、至少为或约90%、至少为或约91%、至少为或约92%、至少

为或约93％、至少为或约94％、至少为或约95％、至少为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％相同的其功能活性变体或生物活性部分。在特定的实施方案中，所述F蛋白或其功能活性变体或生物活性部分保留与亨尼帕病毒G蛋白(诸如章节I.A中所述的G蛋白(例如NiV-G或HeV-G))结合的促融合活性。促融合活性包括F蛋白与亨尼帕病毒G蛋白结合以促进或帮助两个膜内腔(诸如在其脂质双层中包埋了亨尼帕病毒F和G蛋白的靶向脂质颗粒的内腔)和靶细胞(例如含有被靶向包膜蛋白识别或结合的表面受体或分子的细胞)的细胞质融合的活性。在一些实施方案中，所述F蛋白和G蛋白来自相同的亨尼帕病毒属物种(例如NiV-G和NiV-F)。在一些实施方案中，所述F蛋白和G蛋白来自不同的亨尼帕病毒属物种(例如NiV-G和HeV-F)。在特定的实施方案中，所述F蛋白或功能活性变体或生物活性部分保留由组织蛋白酶L切割的切割位点(例如，对应于SEQ ID NO:1的氨基酸109-110之间的切割位点)。

[0304] 在特定的实施方案中，所述F蛋白具有SEQ ID NO:1、SEQ ID NO:2、SEQ ID NO:17、SEQ ID NO:24、SEQ ID NO:25、SEQ ID NO:26、SEQ ID NO:57、SEQ ID NO:57、SEQ ID NO:58、SEQ ID NO:59或SEQ ID NO:60所示的氨基酸序列，或是保留促融合活性的其功能活性变体或其生物活性部分。在一些实施方案中，所述功能活性变体包含如下氨基酸序列，所述氨基酸序列与SEQ ID NO:1、SEQ ID NO:2、SEQ ID NO:17、SEQ ID NO:24、SEQ ID NO:25、SEQ ID NO:26、SEQ ID NO:57、SEQ ID NO:57、SEQ ID NO:58、SEQ ID NO:59或SEQ ID NO:60具有至少为或约80％、至少为或约85％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性并且保留与亨尼帕病毒G蛋白(例如，NiV-G或HeV-G)结合的促融合活性。在一些实施方案中，所述生物活性部分具有如下氨基酸序列，所述氨基酸序列与SEQ ID NO:1、SEQ ID NO:2、SEQ ID NO:17、SEQ ID NO:24、SEQ ID NO:25、SEQ ID NO:26、SEQ ID NO:57、SEQ ID NO:57、SEQ ID NO:58、SEQ ID NO:59或SEQ ID NO:60具有至少为或约80％、至少为或约85％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性并且保留与亨尼帕病毒G蛋白(例如，NiV-G或HeV-G)结合的促融合活性。

[0305] 提及保留促融合活性包括如下的(与亨尼帕病毒F蛋白结合的)活性，所述活性为诸如SEQ ID NO:1、SEQ ID NO:2、SEQ ID NO:17、SEQ ID NO:24、SEQ ID NO:25、SEQ ID NO:26、SEQ ID NO:57、SEQ ID NO:57、SEQ ID NO:58、SEQ ID NO:59或SEQ ID NO:60所示的对应野生型F蛋白的结合水平或程度的为或约10％或与为或约150％或更高之间，诸如对应野生型F蛋白的促融合活性水平或程度的至少或至少约10％、诸如对应野生型F蛋白的促融合活性水平或程度的至少或至少约15％、诸如对应野生型F蛋白的促融合活性水平或程度的至少或至少约20％、诸如对应野生型F蛋白的促融合活性水平或程度的至少或至少约25％、诸如对应野生型F蛋白的促融合活性水平或程度的至少或至少约30％、诸如对应野生型F蛋白的促融合活性水平或程度的至少或至少约35％、诸如对应野生型F蛋白的促融合活性水平或程度的至少或至少约40％、诸如对应野生型F蛋白的促融合活性水平或程度的至少或至少约45％、诸如对应野生型F蛋白的促融合活性水平或程度的至少或至少约50％、诸如对应野生型F蛋白的促融合活性水平或程度的至少或至少约55％、诸如对应野生型F蛋白的促

融合活性水平或程度的至少或至少约60％、诸如对应野生型F蛋白的促融合活性水平或程度的至少或至少约65％、诸如对应野生型F蛋白的促融合活性水平或程度的至少或至少约70％、诸如对应野生型F蛋白的促融合活性水平或程度的至少或至少约75％、诸如对应野生型F蛋白的促融合活性水平或程度的至少或至少约80％、诸如对应野生型F蛋白的促融合活性水平或程度的至少或至少约85％、诸如对应野生型F蛋白的促融合活性水平或程度的至少或至少约90％、诸如对应野生型F蛋白的促融合活性水平或程度的至少或至少约95％、诸如对应野生型F蛋白的促融合活性水平或程度的至少或至少约100％、或诸如对应野生型F蛋白的促融合活性水平或程度的至少或至少约120％。

[0306] 在一些实施方案中，所述F蛋白是含有一个或多个氨基酸突变(诸如一个或多个氨基酸插入、缺失、取代或截短)的功能活性片段或生物活性部分的突变体F蛋白。在一些实施方案中，本文所述的突变涉及与参考F蛋白序列相比氨基酸的氨基酸插入、缺失、取代或截短。在一些实施方案中，参考F蛋白序列是F蛋白的野生型序列或其生物活性部分。在一些实施方案中，突变体F蛋白或其生物活性部分是野生型亨德拉(Hev)病毒F蛋白、尼帕(NiV)病毒F蛋白、雪松(CedPV)病毒F蛋白、墨江病毒F蛋白或蝙蝠副粘病毒F蛋白的突变体。在一些实施方案中，野生型F蛋白由编码SEQ ID NO:1、2、17、24、25、26或57-60中任一者的核苷酸序列编码。

[0307] 在一些实施方案中，突变体F蛋白是野生型F蛋白的生物活性部分，其是N末端和/或C末端截短片段。在一些实施方案中，突变体F蛋白或其野生型F蛋白的生物活性部分包含一个或多个氨基酸取代。在一些实施方案中，本文所述的突变可以改善转导效率。在一些实施方案中，本文所述的突变可以增加促融合能力。示例性突变包括所描述的任何突变，参见例如Khetawat和Broder 2010 Virology Journal 7:312;Witting等人2013Gene Therapy 20:997-1005;公开的国际专利申请号WO/2013/148327。

[0308] 在一些实施方案中，突变体F蛋白是截短的生物活性部分，并且缺乏在野生型F蛋白的C末端处或附近的多达20个连续氨基酸残基，诸如由编码SEQ ID NO:1、17、24、25或26中任一者所示的F蛋白的核苷酸序列编码的野生型F蛋白。在一些实施方案中，突变体F蛋白被截短并且缺乏在野生型F蛋白的C末端处的多达19个连续氨基酸，诸如多达18个、17个、16个、15个、14个、13个、12个、11个、10个、9个、8个、7个、6个、5个、4个、3个、2个或1个连续氨基酸。

[0309] 在一些实施方案中，所述F蛋白或其功能活性变体或生物活性部分包含F1亚基或其融合剂部分。在一些实施方案中，所述F1亚基是F0前体的蛋白水解切割部分。在一些实施方案中，所述F0前体是非活性的。在一些实施方案中，所述F0前体的切割形成二硫键连接的F1+F2异二聚体。在一些实施方案中，所述切割暴露融合肽并产生成熟的F蛋白。在一些实施方案中，所述切割发生在单个碱性残基处或周围。在一些实施方案中，所述切割发生在NiV-F蛋白的精氨酸109处。在一些实施方案中，所述切割发生在亨德拉病毒F蛋白的赖氨酸109处。

[0310] 在一些实施方案中，所述F蛋白是野生型尼帕病毒F(NiV-F)蛋白或是其功能活性变体或生物活性片段。在一些实施方案中，所述F_0前体由编码SEQ ID NO:1所示的序列的核苷酸序列编码。所述编码核酸可以编码具有序列MVVILDKRCY CNLLILILMI SECSVG(SEQ ID NO:34)的信号肽序列。在一些实施方案中，所述F蛋白具有SEQ ID NO:2所示的序列。在一些

实例中,所述F蛋白被切割为包含SEQ ID NO:4所示的序列的F1亚基和包含SEQ ID NO:3所示的序列的F2亚基。

[0311] 在一些实施方案中,所述F蛋白是由编码SEQ ID NO:1所示的序列的核苷酸序列编码的NiV-F蛋白,或者是具有如下氨基酸序列的其功能活性变体或生物活性部分,所述氨基酸序列与SEQ ID NO:1具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、至少为或约84%、至少为或约85%、为或约86%、至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性。在一些实施方案中,所述NiV-F蛋白具有SEQ ID NO:2所示的序列,或者是具有如下氨基酸序列的其功能活性变体或生物活性部分,所述氨基酸序列与SEQ ID NO:2具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、至少为或约84%、至少为或约85%、为或约86%、至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性。在特定的实施方案中,所述F蛋白或其功能活性变体或生物活性部分保留被组织蛋白酶L切割的切割位点(例如,对应于SEQ ID NO:1的氨基酸109-110之间的切割位点)。

[0312] 在一些实施方案中,所述F蛋白或其功能活性变体或生物活性部分包括具有SEQ ID NO:4所示的序列的F1亚基或与SEQ ID NO:4具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、至少为或约84%、至少为或约85%、为或约86%、至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的氨基酸序列。

[0313] 在一些实施方案中,所述F蛋白或其功能活性变体或生物活性部分包括具有SEQ ID NO:3所示的序列的F2亚基或与SEQ ID NO:3具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、至少为或约84%、至少为或约85%、为或约86%、至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的氨基酸序列。

[0314] 在一些实施方案中,所述F蛋白是突变体NiV-F蛋白,所述突变体NiV-F蛋白是被截短的并且缺乏在野生型NiV-F蛋白(例如SEQ ID NO:2所示)的C末端处或附近的多达20个连续氨基酸残基的其生物活性部分。在一些实施方案中,所述突变体NiV-F蛋白包含SEQ ID NO:5所示的氨基酸序列。在一些实施方案中,所述突变体NiV-F蛋白具有与SEQ ID NO:5具有至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的序列。在一些实施方案中,所述突变体F蛋白含有具有SEQ ID NO:6所示序列的F1蛋白。在一些实施方案中,所述突变体F蛋白具有与SEQ ID NO:6具有至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的序列。

[0315] 在一些实施方案中,所述F蛋白是突变体NiV-F蛋白,所述突变体NiV-F蛋白是包含在野生型NiV-F蛋白(SEQ ID NO:2)的C末端处或附近的20个氨基酸截短以及在N-连接糖基化位点的点突变的其生物活性部分。在一些实施方案中,所述突变体NiV-F蛋白包含SEQ ID NO:7所示的氨基酸序列。在一些实施方案中,所述突变体NiV-F蛋白具有与SEQ ID NO:7具有至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的序列。

[0316] 在一些实施方案中,所述F蛋白是突变体NiV-F蛋白,所述突变体NiV-F蛋白是包含在野生型NiV-F蛋白(SEQ ID NO:2)的C末端处或附近的22个氨基酸截短的其生物活性部分。在一些实施方案中,所述NiV-F蛋白由编码SEQ ID NO:8所示的序列的核苷酸序列编码。在一些实施方案中,所述NiV-F蛋白由编码与SEQ ID NO:8具有至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的序列的核苷酸序列编码。在特定的实施方案中,所述变体F蛋白是具有SEQ ID NO:23所示的氨基酸序列的突变体Niv-F蛋白。在一些实施方案中,所述NiV-F蛋白由与SEQ ID NO:23具有至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的序列编码。

[0317] C.脂质双层

[0318] 在一些实施方案中,所述靶向脂质颗粒包括围绕内腔或空腔的天然来源的两亲性脂质双层。在一些实施方案中,靶向脂质颗粒包含脂质双层作为最外的表面。在一些实施方案中,所述脂质双层围绕内腔。在一些实施方案中,所述内腔是水性的。在一些实施方案中,所述内腔与脂质双层内部的亲水性头部基团接触。在一些实施方案中,所述内腔是胞质。在一些实施方案中,所述胞质含有存在于源细胞中的细胞组分。在一些实施方案中,所述胞质不含有存在于源细胞中的组分。在一些实施方案中,所述内腔是空腔。在一些实施方案中,所述空腔含有水性环境。在一些实施方案中,所述空腔不含有水性环境。

[0319] 在一些方面,所述脂质双层在产生含脂质颗粒的过程中来源于源细胞。在章节I.E.中提供了用于产生含脂质颗粒的示例性方法。在一些实施方案中,所述脂质双层包括产生脂质双层的细胞的膜组分,例如磷脂,膜蛋白等。在一些实施方案中,所述脂质双层包括胞质,所述胞质包括在产生微囊泡的细胞中发现的组分,例如溶质、蛋白质、核酸等,但不是细胞的所有组分,例如它们缺乏细胞核。在一些实施方案中,所述脂质双层被认为是外来体样的。脂质双层的大小可以变化,并且在一些情形中具有从30至300nm(诸如从30至150nm以及包括从40至100nm)范围内的直径。

[0320] 在一些实施方案中,所述脂质双层是病毒包膜。在一些实施方案中,所述病毒包膜获自源细胞。在一些实施方案中,所述病毒包膜由来自源细胞质膜的病毒衣壳获得。在一些实施方案中,所述脂质双层获自除宿主细胞的质膜之外的膜。在一些实施方案中,病毒包膜脂质双层包埋有病毒蛋白,包括病毒糖蛋白。

[0321] 在其他方面,所述脂质双层包括合成脂质复合物。在一些实施方案中,所述合成脂质复合物是脂质体。在一些实施方案中,所述脂质双层是以磷脂双层膜和内部水性介质为

特征的囊状结构。在一些实施方案中，所述脂质双层具有由水性介质分开的多个脂质层。在一些实施方案中，当磷脂悬浮于过量的水溶液中时，脂质双层自发形成。在一些实例中，脂质组分在形成封闭结构之前经历自身重排，并且在脂质双层之间截留水和溶解的溶质。

[0322] 在一些实施方案中，靶向包膜蛋白和融合剂(诸如上文所述的任何一种，包括相对于源细胞而言是外源的或过表达的任何一种)被置于脂质双层中。

[0323] 在一些实施方案中，所述靶向脂质颗粒包含几种不同类型的脂质。在一些实施方案中，所述脂质是两亲性脂质。在一些实施方案中，所述两亲性脂质是磷脂。在一些实施方案中，所述磷脂包括磷脂酰胆碱、磷脂酰乙醇胺、磷脂酰肌醇和磷脂酰丝氨酸。在一些实施方案中，所述脂质包括磷脂诸如磷酸胆碱和磷酸肌醇。在一些实施方案中，所述脂质包括DMPC、DOPC和DSPC。

[0324] 在一些实施方案中，所述双层可以由一种或多种相同或不同类型的脂质构成。在一些实施方案中，所述源细胞包括选自以下的细胞：CHO细胞、BHK细胞、MDCK细胞、C3H 10T1/2细胞、FLY细胞、Psi-2细胞、BOSC 23细胞、PA317细胞、WEHI细胞、COS细胞、BSC 1细胞、BSC 40细胞、BMT 10细胞、VERO细胞、W138细胞、MRC5细胞、A549细胞、HT1080细胞、293细胞、293T细胞、B-50细胞、3T3细胞、NIH3T3细胞、HepG2细胞、Saos-2细胞、Huh7细胞、HeLa细胞、W163细胞、211细胞和211A细胞。

[0325] D.外源药剂

[0326] 在实施方案中，所述靶向脂质颗粒(诸如慢病毒载体)还包含相对于源细胞是外源的药剂(下文也称为“运载物”或“有效负载”)。在一些实施方案中，所述外源药剂是蛋白质或核酸(例如DNA、染色体(例如人类人工染色体)、RNA(例如mRNA或miRNA))。在一些实施方案中，所述外源药剂是编码蛋白质的核酸。所述蛋白质可以是靶向递送至靶细胞所需的任何蛋白质。在一些实施方案中，所述蛋白质是治疗剂或诊断剂。在一些实施方案中，所述蛋白质是由疾病或病症表达或与疾病或病症相关的用于靶向细胞的抗原受体，例如嵌合抗原受体(CAR)或T细胞受体(TCR)。提及编码蛋白质的核酸的编码序列在本文中也称为有效负载基因。在一些实施方案中，所述外源药剂或编码所述外源药剂的核酸存在于非细胞颗粒的内腔中。

[0327] 在一些实施方案中，所述外源药剂或运载物包含或编码胞质蛋白。在一些实施方案中，所述外源药剂或运载物包含或编码膜蛋白。在一些实施方案中，所述外源药剂或运载物包含或编码治疗剂。在一些实施方案中，所述治疗剂选自一种或多种蛋白质，例如酶、跨膜蛋白、受体、抗体；核酸，例如DNA、染色体(例如人类人工染色体)、RNA、mRNA、siRNA、miRNA或小分子。

[0328] 在实施方案中，所述外源药剂以至少或不超过10、20、50、100、200、500、1,000、2,000、5,000、10,000、20,000、50,000、100,000、200,000、500,000、1,000,000、5,000,000、10,000,000、50,000,000、100,000,000、500,000,000或1,000,000,000个拷贝存在。在实施方案中，例如由于用siRNA或基因编辑酶处理源细胞(例如哺乳动物源细胞)，所述靶向脂质颗粒具有改变的(例如增加的或降低的)水平的一种或多种内源分子，例如蛋白质或核酸(例如，在一些实施方案中，相对于源细胞是内源的，并且在一些实施方案中，相对于靶细胞是内源的)。在实施方案中，所述内源分子以至少或不超过10、20、50、100、200、500、1,000、2,000、5,000、10,000、20,000、50,000、100,000、200,000、500,000、1,000,000、5,000,000、

10,000,000、50,000,000、100,000,000、500,000,000或1,000,000,000个拷贝存在。在实施方案中，所述内源分子(例如RNA或蛋白质)以比其在源细胞中的浓度大至少1、2、3、4、5、10、20、50、100、500、10^3、$5.0x10^3$、10^4、$5.0x10^4$、10^5、$5.0x10^5$、10^6、$5.0x10^6$、$1.0x10^7$、$5.0x10^7$或$1.0x10^8$的浓度存在。在实施方案中，所述内源分子(例如RNA或蛋白质)以比其在源细胞中的浓度小至少1、2、3、4、5、10、20、50、100、500、10^3、$5.0x10^3$、10^4、$5.0x10^4$、10^5、$5.0x10^5$、10^6、$5.0x10^6$、$1.0x10^7$、$5.0x10^7$或$1.0x10^8$的浓度存在。

[0329] 在一些实施方案中，所述靶向脂质颗粒向靶细胞递送由融合体所包含的运载物(例如治疗剂，例如外源性治疗剂)的至少10％、20％、30％、40％、50％、60％、70％、80％、90％、95％、96％、97％、98％、或99％。在一些实施方案中，与靶细胞融合的靶向脂质颗粒向靶细胞递送由与靶细胞融合的脂质颗粒所包含的运载物(例如治疗剂，例如外源性治疗剂)的平均至少10％、20％、30％、40％、50％、60％、70％、80％、90％、95％、96％、97％、98％、或99％。在一些实施方案中，靶向脂质颗粒组合物向靶组织递送由靶向脂质颗粒组合物所包含的运载物(例如治疗剂，例如外源性治疗剂)的至少10％、20％、30％、40％、50％、60％、70％、80％、90％、95％、96％、97％、98％、或99％。

[0330] 在一些实施方案中，所述外源药剂或运载物在靶向脂质颗粒来源的细胞中不天然表达。在一些实施方案中，所述外源药剂或运载物在靶向脂质颗粒来源的细胞中天然表达。在一些实施方案中，通过在脂质颗粒来源的细胞中的表达(例如从通过转染、转导或电穿孔引入的DNA或mRNA的表达)将所述外源药剂或运载物装载到所述靶向脂质颗粒中。在一些实施方案中，所述外源药剂或运载物由整合到基因组中的DNA表达或以附加体维持。在一些实施方案中，所述外源药剂或运载物的表达是组成型的。在一些实施方案中，所述外源药剂或运载物的表达是诱导型的。在一些实施方案中，紧接在产生所述靶向脂质颗粒之前诱导所述外源药剂或运载物的表达。在一些实施方案中，在融合剂表达的同时诱导所述外源药剂或运载物的表达。

[0331] 在一些实施方案中，通过电穿孔到脂质颗粒本身中或电穿孔到融合体来源的细胞中将所述外源药剂或运载物装载到脂质颗粒中。在一些实施方案中，通过转染(例如编码所述运载物的DNA或mRNA)到脂质颗粒本身中或到脂质颗粒来源的细胞中将所述外源药剂或运载物装载到脂质颗粒中。

[0332] 在一些实施方案中，所述外源药剂或运载物可包括一种或多种核酸序列、一种或多种多肽、核酸序列和/或多肽的组合、一种或多种细胞器以及它们的任何组合。在一些实施方案中，所述外源药剂或运载物可包括一种或多种细胞组分。在一些实施方案中，所述外源药剂或运载物包括一种或多种胞质和/或核组分。

[0333] 在一些实施方案中，所述外源药剂或运载物包括核酸，例如DNA、nDNA(核DNA)、mtDNA(线粒体DNA)、编码蛋白质的DNA、基因、操纵子、染色体、基因组、转座子、反转录转座子、病毒基因组、内含子、外显子、经修饰的DNA、mRNA(信使RNA)、tRNA(转移RNA)、经修饰的RNA、微小RNA、siRNA(小干扰RNA)、tmRNA(转移信使RNA)、rRNA(核糖体RNA)、mtRNA(线粒体RNA)、snRNA(小核RNA)、小核仁RNA(snoRNA)、SmY RNA(mRNA反式剪接RNA)、gRNA(向导RNA)、TERC(端粒酶RNA组分)、aRNA(反义RNA)、顺式-NAT(顺式-天然反义转录物)、CRISPR RNA(crRNA)、IncRNA(长非编码RNA)、piRNA(piwi相互作用RNA)、shRNA(短发夹RNA)、tasiRNA(反式作用siRNA)、eRNA(增强子RNA)、卫星RNA、pcRNA(编码蛋白质的RNA)、dsRNA(双链

RNA)、RNAi(干扰RNA)、circRNA(环状RNA)、重编程RNA、适体以及它们的任何组合。在一些实施方案中,所述核酸是野生型核酸。在一些实施方案中,所述蛋白质是突变体核酸。在一些实施方案中,所述核酸是多个核酸序列的融合物或嵌合体。

[0334] 在一些实施方案中,所述外源药剂或运载物可包括核酸。例如,所述外源药剂或运载物可包含增强内源蛋白表达的RNA,或抑制内源蛋白的蛋白质表达的siRNA或miRNA。例如,所述内源蛋白可以调节在靶细胞中的结构或功能。在一些实施方案中,所述运载物可包含编码调节在靶细胞中的结构或功能的工程化蛋白质的核酸。在一些实施方案中,所述外源药剂或运载物是靶向调节在靶细胞中的结构或功能的转录激活因子的核酸。

[0335] 在一些实施方案中,所述外源药剂或运载物是或编码多肽,例如酶、结构多肽、信号传导多肽、调节多肽、转运多肽、感觉多肽、运动多肽、防御多肽、贮存多肽、转录因子、抗体、细胞因子、激素、分解代谢多肽、合成代谢多肽、蛋白水解多肽、代谢多肽、激酶、转移酶、水解酶、裂解酶、异构酶、连接酶、酶调节多肽、蛋白结合多肽、脂质结合多肽、膜融合多肽、细胞分化多肽、表观遗传多肽、细胞死亡多肽、核转运多肽、核酸结合多肽、重编程多肽、DNA编辑多肽、DNA修复多肽、DNA重组多肽、转座酶多肽、DNA整合多肽、靶向核酸内切酶(例如锌指核酸酶、转录激活物样核酸酶(TALEN)、cas9及其同源物)、重组酶以及它们的任何组合。在一些实施方案中,所述蛋白质靶向细胞中的蛋白质用于降解。在一些实施方案中,所述蛋白质靶向细胞中的蛋白质以通过将蛋白质定位到蛋白酶体而降解。在一些实施方案中,所述蛋白质是野生型蛋白质。在一些实施方案中,所述蛋白质是突变体蛋白质。在一些实施方案中,所述蛋白质是融合蛋白或嵌合蛋白。

[0336] 在一些实施方案中,所述外源药剂或运载物是小分子,例如离子(例如Ca^{2+}、C1-、Fe^{2+})、碳水化合物、脂质、活性氧物质、活性氮物质、类异戊二烯、信号传导分子、血红素、多肽辅因子、电子接受化合物、电子供给化合物、代谢物、配体以及它们的任何组合。在一些实施方案中,所述小分子是与细胞中的靶标相互作用的药物。在一些实施方案中,所述小分子靶向细胞中的蛋白质用于降解。在一些实施方案中,所述小分子靶向细胞中的蛋白质以通过将蛋白质定位到蛋白酶体而降解。在一些实施方案中,所述小分子是蛋白水解靶向嵌合体分子(PROTAC)。

[0337] 在一些实施方案中,所述外源药剂或运载物包括蛋白质、核酸或代谢物的混合物,例如多种多肽、多种核酸、多种小分子;核酸、多肽和小分子的组合;核糖核蛋白复合物(例如Cas9-gRNA复合物);多种转录因子、多种表观遗传因子、重编程因子(例如Oct4、Sox2、cMyc和Klf4);多种调节RNA;以及它们的任何组合。

[0338] 在一些实施方案中,所述外源药剂或运载物包括一种或多种细胞器,例如线粒体(chondrisomes)、线粒体(mitochondria)、溶酶体、细胞核、细胞膜、细胞质、内质网、核糖体、液泡、内体、剪接体、聚合酶、衣壳、顶体、自体吞噬体、中心粒、糖体、乙醛酸循环体、氢化酶体、黑素小体、纺锤剩体、肌原纤维、刺丝囊、过氧化物酶体、蛋白酶体、囊泡、应激颗粒、细胞器网络、以及它们的任何组合。

[0339] 在一些实施方案中,所述外源药剂是或编码胞质蛋白,例如在受体细胞中产生并定位于受体细胞的细胞质的蛋白质。在一些实施方案中,所述外源药剂是或编码分泌蛋白,例如由受体细胞产生和分泌的蛋白质。在一些实施方案中,所述外源药剂是或编码核蛋白,例如在受体细胞中产生并被输入受体细胞的细胞核的蛋白质。在一些实施方案中,所述外

源药剂是或编码细胞器蛋白(例如线粒体蛋白),例如在受体细胞中产生并被输入受体细胞的细胞器(例如线粒体)中的蛋白质。在一些实施方案中,所述蛋白质是野生型蛋白质或突变体蛋白质。在一些实施方案中,所述蛋白质是融合蛋白或嵌合蛋白。

[0340] 在一些实施方案中,所述外源药剂能够被递送至肝细胞或肝脏细胞。在一些实施方案中,可以递送所述外源药剂或运载物以治疗肝细胞或肝脏细胞中的疾病或病况。

[0341] 在一些实施方案中,所述外源药剂由选自以下的基因编码:OTC、CPS1、NAGS、BCKDHA、BCKDHB、DBT、DLD、MUT、MMAA、MMAB、MMACHC、MMADHC、MCEE、PCCA、PCCB、UGT1A1、ASS1、PAH、PAL、ATP8B1、ABCB11、ABCB4、TJP2、IVD、GCDH、ETFA、ETFB、ETFDH、ASL、D2HGDH、HMGCL、MCCC1、MCCC2、ABCD4、HCFC1、LNBRD1、ARG1、SLC25A15、SLC25A13、ALAD、CPOX、HMBS、PPOX、BTD、HLCS、PC、SLC7A7、CPT2、ACADM、ACADS、ACADVL、AGL、G6PC、GBE1、PHKA1、PHKA2、PHKB、PHKG2、SLC37A4、PMM2、CBS、FAH、TAT、GALT、GALK1、GALE、G6PD、SLC3A1、SLC7A9、MTHFR、MTR、MTRR、ATP7B、HPRT1、HJV、HAMP、JAG1、TTR、AGXT、LIPA、SERPING1、HSD17B4、UROD、HFE、LPL,GRHPR、HOGA1、LDLR、ACAD8、ACADSB、ACAT1、ACSF3、ASPA、AUH、DNAJC19、ETHE1、FBP1、FTCD、GSS、HIBCH、IDH2、L2HGDH、MLYCD、OPA3、OPLAH、OXCT1、POLG、PPM1K、SERAC1、SLC25A1、SUCLA2、SUCLG1、TAZ、AGK、CLPB、TMEM70、ALDH18A1、OAT、CA5A、GLUD1、GLUL、UMPS、SLC22A5、CPT1A、HADHA、HADH、SLC52A1、SLC52A2、SLC52A3、HADHB、GYS2、PYGL、SLC2A2、ALG1、ALG2、ALG3、ALG6、ALG8、ALG9、ALG11、ALG12、ALG13、ATP6V0A2、B3GLCT、CHST14、COG1、COG2、COG4、COG5、COG6、COG7、COG8、DOLK、DHDDS、DPAGT1、DPM1、DPM2、DPM3、G6PC3、GFPT1、GMPPA、GMPPB、MAGT1、MAN1B1、MGAT2、MOGS、MPDU1、MPI、NGLY1、PGM1、PGM3、RFT1、SEC23B、SLC35A1、SLC35A2、SLC35C1、SSR4、SRD5A3、TMEM165、TRIP11、TUSC3、ALG14、B4GALT1、DDOST、NUS1、RPN2、SEC23A、SLC35A3、ST3GAL3、STT3A、STT3B、AGA、ARSA、ARSB、ASAH1、ATP13A2、CLN3、CLN5、CLN6、CLN8、CTNS、CTSA、CTSD、CTSF、CTSK、DNAJC5、FUCA1、GAA、GALC、GALNS、GLA、GLB1、GM2A、GNPTAB、GNPTG、GNS、GRN、GUSB、HEXA、HEXB、HGSNAT、HYAL1、IDS、IDUA、KCTD7、LAMP2、MAN2B1、MANBA、MCOLN1、MFSD8、NAGA、NAGLU、NEU1、NPC1、NPC2、SGSH、PPT1、PSAP、SLC17A5、SMPD1、SUMF1、TPP1、AHCY、GNMT、MAT1A、GCH1、PCBD1、PTS、QDPR、SPR、DNAJC12、ALDH4A1、PRODH、HPD、GBA、HGD、AMN、CD320、CUBN、GIF、TCN1、TCN2、PREPL、PHGDH、PSAT1、PSPH、AMT,GCSH、GLDC、LIAS、NFU1、SLC6A9、SLC2A1、ATP7A、AP1S1、CP、SLC33A1、PEX7、PHYH、AGPS、GNPAT、ABCD1、ACOX1、PEX1、PEX2、PEX3、PEX5、PEX6、PEX10、PEX12、PEX13、PEX14、PEX16、PEX19、PEX26、AMACR、ADA、ADSL、AMPD1、GPHN、MOCOS、MOCS1、PNP、XDH、SUOX、OGDH、SLC25A19、DHTKD1、SLC13A5、FH、DLAT、MPC1、PDHA1、PDHB、PDHX、PDP1、ABCC2、SLCO1B1、SLCO1B3、HFE2、ADAMTS13、PYGM、COL1A2、TNFRSF11B、TSC1、TSC2、DHCR7、PGK1、VLDLR、KYNU、F5、C3、COL4A1、CFH、SLC12A2、GK、SFTPC、CRTAP、P3H1、COL7A1、PKLR、TALDO1、TF、EPCAM、VHL、GC、SERPINA1、ABCC6、F8、F9、ApoB、PCSK9、LDLRAP1,ABCG5、ABCG8、LCAT、SPINK5或GNE。

[0342] 在一些实施方案中,所述外源药剂由选自以下的基因编码:OTC、CPS1、NAGS、BCKDHA、BCKDHB、DBT、DLD、MUT、MMAA、MMAB、MMACHC、MMADHC、MCEE、PCCA、PCCB、UGT1A1、ASS1、PAL、PAH、ATP8B1、ABCB11、ABCB4、TJP2、IVD、GCDH、ETFA、ETFB、ETFDH、ASL、D2HGDH、HMGCL、MCCC1、MCCC2、ABCD4、HCFC1、LMBRD1、ARG1、SLC25A15、SLC25A13、ALAD、CPOX、HMBS、PPOX、BTD、HLCS、PC、SLC7A7、CPT2、ACADM、ACADS、ACADVL、AGL、G6PC、GBE1、PHKA1、PHKA2、PHKB、PHKG2、SLC37A4、PMM2、CBS、FAH、TAT、GALT、GALK1、GALE、G6PD、SLC3A1、SLC7A9、MTHFR、MTR、

MTRR、ATP7B、HPRT1、HJV、HAMP、JAG1、TTR、AGXT、LIPA、SERPING1、HSD17B4、UROD、HFE、LPL、GRHPR、HOGA1或LDLR。在一些实施方案中，所述外源药剂是酶苯丙氨酸解氨酶(PAL)。

[0343] 在一些实施方案中，可以递送所述外源药剂或运载物以治疗表5中列出的疾病或适应症。在一些实施方案中，所述适应症对肝脏细胞或肝细胞具有特异性。

[0344] 在一些实施方案中，所述外源药剂包含下面表5的蛋白质。在一些实施方案中，所述外源药剂包含表5的任何蛋白质的野生型人序列、其功能片段(例如，其酶活性片段)或其功能变体。在一些实施方案中，所述外源药剂包含与表5的氨基酸序列具有至少70％、75％、80％、85％、90％、95％、96％、97％、98％、或99％同一性的氨基酸序列，例如表5第4列的Uniprot蛋白质登录号序列或表5第5列的氨基酸序列。在一些实施方案中，编码外源药剂的有效负载基因编码与表5的氨基酸序列具有至少70％、75％、80％、85％、90％、95％、96％、97％、98％、或99％同一性的氨基酸序列。在一些实施方案中，编码外源药剂的有效负载基因具有与表5的核酸序列(例如表5第3列的Ensemble基因登录号)具有至少70％、75％、80％、85％、90％、95％、96％、97％、98％、或99％同一性的核酸序列。

[0345] 表5.第一列列出了可以根据本文的方法和用途递送以治疗第六列中的适应症的外源药剂。表5的每个Uniprot登录号通过引入整体并入本文。

[0346]

基因	Entrez登录号	Ensembl基因登录号(所示的ENSG0000+编号)	Uniprot蛋白质登录号	氨基酸序列(第一个Uniprot登录号)SEQ ID NO	疾病/障碍	类别
OTC	5009	0036473	P00480	61	鸟氨酸氨甲酰	尿素循

[0347]

					基转移酶(OTC)缺乏症	环障碍
CPS1	1373	0021826	P31327、Q6PEK7、B7ZAW0、A0A024R454	62	氨甲酰基-磷酸合成酶 I (CPSI)缺乏症	尿素循环障碍
NAGS	162417	0161653	Q8N159、Q2NKP2	63	N-乙酰谷氨酸合酶(NAGS)缺乏症	尿素循环障碍
BCKDHA	593	0248098	A0A024R0K3、P12694、Q59EI3	64	枫糖尿症(MSUD);经典枫糖尿症(CMSUD)	有机酸血症
BCKDHB	594	0083123	A0A140VKB3、P21953、B4E2N3、B7ZB80	65	枫糖尿症(MSUD);经典枫糖尿症(CMSUD)	有机酸血症
DBT	1629	0137992	P11182	66	枫糖尿症(MSUD);经典枫糖尿症(CMSUD)	有机酸血症
DLD	1738	0091140	A0A024R713、P09622、E9PEX6	67	枫糖尿症(MSUD)二氢硫辛酰胺脱氢酶缺乏症	尿素循环障碍
MUT	4594	0146085	A0A024RD82、B2R6K1、P22033	68	由于甲基丙二酰 CoA 变位酶缺乏引起的甲基丙二酸血症	有机酸血症
MMAA	166785	0151611	Q8IVH4	69	钴胺素 A 缺乏症(甲基丙二酸血症)	有机酸血症
MMAB	326625	0139428	Q96EY8	70	钴胺素 B 缺乏症(甲基丙二酸血症)	有机酸血症
MMACHC	25974	0132763	A0A0C4DGU2、Q9Y4U1	71	钴胺素 C 缺乏症(甲基丙二酸血症);甲基丙二酸血症伴有高胱氨酸尿症	有机酸血症
MMADHC	27249	0168288	Q9H3L0	72	钴胺素 D 缺乏症(甲基丙二酸血症);甲基丙二酸血症伴有高胱氨酸尿症;高胱氨酸尿症;	有机酸血症

[0348]

					钴胺素 C 缺乏症	
MCEE	84693	0124370	Q96PE7	73	甲基丙二酸血症; 钴胺素 D 缺乏症	有机酸血症
PCCA	5095	0175198	P05165	74	丙酸血症	有机酸血症
PCCB	5096	0114054	P05166	75	丙酸血症	有机酸血症
UGT1A1	54658	0241635	P22309、Q5DT03	76	克里格勒-纳贾尔综合征 1 型 克里格勒-纳贾尔综合征 2 型、吉尔伯特综合征	
ASS1	445	0130707	P00966、Q5T6L4	77	瓜氨酸血症 I 型	尿素循环障碍
PAH	5053	0171759	A0A024RBG4、P00439	78	苯丙氨酸羟化酶缺乏症	氨基酸代谢病
PAL				79	苯丙氨酸羟化酶缺乏症	氨基酸代谢病
ATP8B1	5205	0081923	O43520	80	进行性家族性肝内胆汁淤积 1 型	
ABCB11	8647	0073734,0276582	O95342	81	进行性家族性肝内胆汁淤积 2 型; 进行性家族性肝内胆汁淤积 3 型	
ABCB4	5244	0005471	P21439	82	进行性家族性肝内胆汁淤积 3 型; 进行性家族性肝内胆汁淤积 2 型	
TJP2	9414	0119139	B7Z2R3、Q9UDY2、B7Z954	83	进行性家族性肝内胆汁淤积 4 型	
IVD	3712	0128928	P26440、A0A0A0MT83	84	异戊酸血症(IVD)	有机酸血症
GCDH	2639	0105607	A0A024R7F9、Q92947	85	戊二酸血症 I 型	有机酸血症
ETFA	2108	0140374	A0A0S2Z3L0、P13804	86	多酰基 CoA 脱氢酶缺乏症(也称为戊二酸尿	有机酸血症

[0349]

					症 II 型)	
ETFB	2109	0105379	P38117	87	多酰基 CoA 脱氢酶缺乏症(也称为戊二酸尿症 II 型)	有机酸血症
ETFDH	2110	0171503	B4DEQ0、Q16134	88	多酰基 CoA 脱氢酶缺乏症(也称为戊二酸尿症 II 型)	有机酸血症
ASL	435	0126522	A0A024RDL8、P04424、A0A0S2Z316	89	精氨基琥珀酸裂合酶(ASL)缺乏症	尿素循环障碍
D2HGDH	728294	0180902	B3KSR6、B4E3K7、B5MCV2、Q8N465	90	D-2-羟基戊二酸尿症 I 型	有机酸血症
HMGCL	3155	0117305	P35914	91	3-羟基-3-甲基戊二酰 CoA 裂解酶(3HMG)缺乏症	有机酸血症 尿素循环障碍
MCCC1	56922	0078070	Q68D27、Q96RQ3、A0A0S2Z693、E9PHF7	92	3-甲基巴豆酰 CoA 羧化酶(3MCC)缺乏症	有机酸血症
MCCC2	64087	0131844,0281742,0275300	A0A140VK29、Q9HCC0	93	3-甲基巴豆酰 CoA 羧化酶(3MCC)缺乏症	有机酸血症
ABCD4	5826	0119688	A0A024R6B9、O14678、A0A024R6C8	94	甲基丙二酸血症伴有高胱氨酸尿症	有机酸血症
HCFC1	3054	0172534	P51610、A6NEM2	95	甲基丙二酸血症伴有高胱氨酸尿症	有机酸血症
LMBRD1	55788	0168216	Q9NUN5	96	甲基丙二酸血症伴有高胱氨酸尿症	有机酸血症
ARG1	383	0118520	P05089	97	精氨酸酶(ARG1)缺乏症	尿素循环障碍(Urea cycle disorder)
SLC25A15	10166	0102743	Q9Y619	98	高氨血症-高鸟氨酸血症-同型瓜氨酸尿症(HHH)综合征	尿素循环障碍
SLC25A13	10165	0004864	Q9UJS0	99	维生素 P 缺乏症瓜氨酸血症 II 型	尿素循环障碍

[0350]

ALAD	210	0148218	P13716	100	急性肝卟啉症	卟啉症
CPOX	1371	0080819	P36551	101	急性肝卟啉症	卟啉症
HMBS	3145	0256269,0281702	P08397	102	急性肝卟啉症;急性间歇性卟啉症	卟啉症
PPOX	5498	0143224	P50336、B4DY76	103	急性肝卟啉症	卟啉症
BTD	686	0169814	P43251	104	生物素酶缺乏症	有机酸血症
HLCS	3141	0159267	P50747	105	全羧化酶合成酶缺乏症	有机酸血症
PC	5091	0173599	P11498 A0A024R5C5	106	丙酮酸羧化酶缺乏症	尿素循环障碍
SLC7A7	9056	0155465	Q9UM01 A0A0S2Z502	107	赖氨酸尿性蛋白耐受不良	尿素循环障碍
CPT2	1376	0157184	P23786 A0A140VK13 A0A1B0GTB8	108	肉碱棕榈基转移酶 II 型(CPT II)缺乏症	脂肪酸氧化
ACADM	34	0117054	P11310 A0A0S2Z366、B7Z911、Q5HYG7、Q5T4U5、B4DJE7	109	中链酰基 CoA 脱氢酶(MCAD)缺乏症	脂肪酸氧化
ACADS	35	0122971	P16219 E5KSD5、B4DUH1、E9PE82	110	短链酰基 CoA (SCAD)脱氢酶缺乏症	脂肪酸氧化
ACADVL	37	0072778	P49748 B3KPA6	111	极长链酰基 CoA 脱氢酶(VLCAD)缺乏症	脂肪酸氧化
AGL	178	0162688	P35573 A0A0S2A4E4	112	GSD III (Cori / Forbe 疾病或脱支酶(Debrancher))	肝糖原贮积症
G6PC	2538	0131482	P35575	113	GSDIa (冯·吉尔克病)	肝糖原贮积症
GBE1	2632	0114480	Q04446 Q59ET0	114	GSD IV (安德森病，分支酶)	肝糖原贮积症
PHKA1	5255	0067177	P46020	115	GSD IXa	
PHKA2	0044446 5256	5256 0044446	P46019	116	GSD IXa	肝糖原贮积症
PHKB	5257	0102893	Q93100	117	GSD IXb	肝糖原贮积症
PHKG2	5261	0156873	P15735	118	GSD IXc	肝糖原贮积症

	SLC37A4	2542	02815000137700	O43826 A0A024R3H9、A8K0S7、A0A024R3L1、B4DUH2	119	GSDIb. c, d	肝糖原贮积症
	PMM2	5373	0140650	O15305、A0A0S2Z4J6、Q59F02	120	PMM2-CDG	糖基化障碍
	CBS	102724560,875	0160200	P35520、P0DN79、Q9NTF0、B7Z2D6	121	胱硫醚β合酶缺乏症(经典高胱氨酸尿症);高胱氨酸尿症	氨基酸代谢病
	FAH	2184	0103876	P16930	122	酪氨酸血症I型	氨基酸代谢病
	TAT	6898	0198650	P17735、A0A140VKB7	123	酪氨酸血症II型 酪氨酸血症III型	氨基酸代谢病
	GALT	2592	0213930	P07902、A0A0S2Z3Y7、B2RAT6	124	半乳糖血症 由于半乳糖-1-磷酸尿苷酰转移酶(GALT)缺乏症	碳水化合物障碍(Carbohydrate disorder)
[0351]	GALK1	2584	0108479	P51570	125	半乳糖血症	碳水化合物障碍
	GALE	2582	0117308	Q14376	126	半乳糖血症	碳水化合物障碍
	G6PD	2539	0160211	P11413	127	葡萄糖-6-磷酸脱氢酶(G6PD)缺乏症	碳水化合物障碍
	SLC3A1	6519	0138079	Q07837、A0A0S2Z4E1、B8ZZK1	128	胱氨酸尿症	氨基酸代谢病
	SLC7A9	11136	0021488	P82251	129	胱氨酸尿症	氨基酸代谢病
	MTHFR	4524	0177000	P42898、Q59GJ6、Q81U67	130	高胱氨酸尿症	氨基酸代谢病
	MTR	4548	0116984	Q99707	131	高胱氨酸尿症	氨基酸代谢病
	MTRR	4552	0124275	Q9UBK8	132	高胱氨酸尿症	氨基酸代谢病
	ATP7B	540	0123191	P35670、A0A024RDX3、B7ZLR4、B7ZLR3、	133	威尔逊病(Wilson Disease) 铜代谢障碍	金属转运障碍

[0352]

			E7ET55			
HPRT1	3251	0165704	P00492、A0A140VJL3	134	莱施-奈恩综合征 嘌呤代谢障碍	嘌呤代谢障碍
HJV	148738	0168509	Q6ZVN8	135	血色素沉着症，2A 型	
HAMP	57817	0105697	P81172	136	血色素沉着症 2B 型：原发性血色素沉着症	
JAG1	182	0101384	P78504、Q99740	137	Alagille 综合征 1	
TTR	7276	0118271	P02766、E9KL36	138	家族性 TTR 淀粉样变性；家族性淀粉样多神经病	
AGXT	189	0172482	P21549	139	原发性高草酸尿症 I 型	
LIPA	3988	0107798	P38571 A0A0A0MT32	140	溶酶体酸性脂肪酶缺乏症	溶酶体贮积症
SERPING1	710	0149131	P05155、A0A0S2Z4J1、B2R659、E7EWE5、B3KSP2、G5E9S2	141	遗传性血管性水肿	
HSD17B4	3295	0133835	P51659	142	D-双功能蛋白缺乏症 X-连锁肾上腺脑白质营养不良	过氧化物酶体障碍
UROD	7389	0126088	P06132	143	迟发性皮肤卟啉症	
HFE	3077	0010704	Q30201	144	迟发性皮肤卟啉症	
LPL	4023	0175445	P06858、A0A1B1RVA9	145	脂蛋白脂酶缺乏症(高脂蛋白血症 Ia 型；Buerger-Gruetz 综合征或家族性高乳糜微粒血症)	
GRHPR	9380	0137106	Q9UBQ7	146	原发性高草酸尿症 II 型	
HOGA1	112817	0241935	Q86XE5	147	原发性高草酸尿症 III 型	
LDLR	3949	0130164	P01130、A0A024R7D5	148	纯合性家族性高胆固醇血症	

[0353]

ACAD8	27034	0151498	Q9UKU7	149	异丁酰基 CoA 脱氢酶(IBD)缺乏症	有机酸血症
ACADSB	36	0196177	P45954、A0A0S2Z3P9	150	短支链酰基 CoA 脱氢酶(SBCAD)缺乏症	有机酸血症
ACAT1	38	0075239	A0A140VJX1、P24752	151	β-酮硫解酶缺乏症	有机酸血症
ACSF3	197322	0176715	Q4G176、F5H5A1	152	丙二酸和甲基丙二酸联合酸尿症	有机酸血症
ASPA	443	0108381	P45381、Q6FH48	153	卡纳万病	有机酸血症
AUH	549	0148090	Q13825、B4DYI6	154	3-甲基戊烯二酸血症 I 型	有机酸血症
DNAJC19	131118	0205981	Q96DA6、A0A0S2Z5X1	155	扩张型心肌病伴有共济失调综合征(导致 3-甲基戊烯二酸尿症)	有机酸血症
ETHE1	23474	0105755	A0A0S2Z580、O95571、A0A0S2Z5N8、A0A0S2Z5B3、B2RCZ7	156	乙基丙二酸脑病	有机酸血症
FBP1	2203	0165140	P09467、Q2TU34	157	果糖 1,6-二磷酸酶缺乏症	有机酸血症
FTCD	10841	0160282,0281775	O95954	158	谷氨酸亚胺甲基转移酶缺乏症(FIGLU	有机酸血症
GSS	2937	0100983	P48637、V9HWJ1	159	谷胱甘肽合成酶缺乏症	有机酸血症
HIBCH	26275	0198130	A0A140VJL0、Q6NVY1	160	3-羟基异丁酰基 CoA 水解酶缺乏症	有机酸血症
IDH2	3418	0182054	P48735、B4DSZ6	161	D-2-羟基戊二酸尿症 II 型	有机酸血症
L2HGDH	79944	0087299	Q9H9P8	162	L-2-羟基戊二酸尿症	有机酸血症
MLYCD	23417	0103150	O95822	163	丙二酸血症	有机酸血症
OPA3	80207	0125741	Q9H6K4、B4DK77	164	Costeff 综合征/3-甲基戊烯二酸尿症 III 型	有机酸血症
OPLAH	26873	0178814	O14841	165	5-氧脯氨酸酶缺乏症	有机酸血症

	OXCT1	5019	0083720	A0A024R040、P55809	166	SCOT 缺乏症	有机酸血症
	POLG	5428	0140521	E5KNU5、P54098	167	3-甲基戊烯二酸尿症	有机酸血症
	PPM1K	152926	0163644	Q8N3J5	168	枫糖尿症(MSUD)，变异型	有机酸血症
	SERAC1	84947	0122335	Q96JX3	169	Megdel 综合征	有机酸血症
	SLC25A1	6576	0100075	D9HTE9、B4DP62、P53007	170	D,L-2-羟基戊二酸尿症	有机酸血症
	SUCLA2	8803	0136143	E5KS60、Q9P2R7、Q9Y4T0	171	琥珀酸 CoA 连接酶缺乏症，甲基丙二酸尿症	有机酸血症
	SUCLG1	8802	0163541	P53597	172	琥珀酸 CoA 连接酶缺乏症，甲基丙二酸尿症	有机酸血症
	TAZ	6901	0102125	A0A0S2Z4K0、Q16635、A6XNE1、A0A0S2Z4E6、A0A0S2Z4K9、A0A0S2Z4F4	173	Barth 综合征	有机酸血症
[0354]	AGK	55750	0006530,0262327	A4D1U5、Q53H12	174	3-甲基戊烯二酸尿症	有机酸血症
	CLPB	81570	0162129	Q9H078、A0A140VK11	175	3-甲基戊烯二酸尿症	有机酸血症
	TMEM70	54968	0175606	Q9BUB7	176	3-甲基戊烯二酸尿症	有机酸血症
	ALDH18A1	5832	0059573	P54886	177	ALDH18A1 相关皮肤松弛症	尿素循环障碍
	OAT	4942	0065154	A0A140VJQ4、P04181	178	环状萎缩(OAT)	尿素循环障碍
	CA5A	763	0174990	P35218	179	碳酸酐酶缺乏症	尿素循环障碍
	GLUD1	2746	0148672	P00367、E9KL48	180	谷氨酸脱氢酶缺乏症	尿素循环障碍
	GLUL	2752	0135821	A8YXX4、P15104	181	谷氨酰胺合成酶缺乏症	尿素循环障碍
	UMPS	7372	0114491	A8K5J1、P11172	182	乳清酸尿症	尿素循环障碍
	SLC22A5	6584	0197375	O76082	183	肉碱-酰基肉碱移位酶(CACT)缺乏症	脂肪酸氧化
	CPT1A	1374	0110090	P50416、A0A024R5F4、B2RAQ8、	184	肉碱软脂酰转移酶 I 型(CPT I)缺乏症	脂肪酸氧化

			Q8WZ48			
HADHA	3030	0084754	E9KL44、P40939	185	长链 3-羟酰基 CoA 脱氢酶 (LCHAD)缺乏症	脂肪酸氧化
HADH	3033	0138796	Q16836、B3KTT6	186	中/短链酰基 CoA 脱氢酶 (M/SCHAD)缺乏症	脂肪酸氧化
SLC52A1	55065	0132517	Q9NWF4	187	核黄素转运蛋白缺乏症	脂肪酸氧化
SLC52A2	79581	0185803	Q9HAB3	188	核黄素转运蛋白缺乏症	脂肪酸氧化
SLC52A3	113278	0101276	K0A6P4、Q9NQ40	189	核黄素转运蛋白缺乏症	脂肪酸氧化
HADHB	3032	0138029	P55084、F5GZQ3	190	三功能蛋白缺乏症	脂肪酸氧化
GYS2	2998	0111713	P54840	191	GSD 0 (糖原合酶，肝同工型)	肝糖原贮积症
PYGL	5836	0100504	P06737	192	GSD VI (Hers 病)	肝糖原贮积症
SLC2A2	6514	0163581	P11168、Q6PAU8	193	Fanconi-Bickel 综合征	肝糖原贮积症
ALG1	56052	0033011	Q9BT22	194	ALG1-CDG	糖基化障碍
ALG2	85365	0119523	A0A024R184、Q9H553	195	ALG2 相关肌无力综合征	糖基化障碍
ALG3	10195	0214160	Q92685、C9J7S5	196	ALG3-CDG	糖基化障碍
ALG6	29929	0088035	Q9Y672	197	ALG6-CDG	糖基化障碍
ALG8	79053	0159063	Q9BVK2、A0A024R5K5	198	ALG8-CDG	糖基化障碍
ALG9	79796	0086848	Q9H6U8	199	ALG9-CDG	糖基化障碍
ALG11	440138	0253710	Q2TAA5	200	ALG11-CDG	糖基化障碍
ALG12	79087	0182858	A0A024R4V6、Q9BV10	201	ALG12-CDG	糖基化障碍
ALG13	79868	0101901	Q9NP73、A0A087WX43、A0A087WT15	202	ALG13-CDG	糖基化障碍
ATP6V0A2	23545	0185344	Q9Y487	203	ATP6V0A2 相关皮肤松弛症	糖基化障碍
B3GLCT	145173	0187676	Q6Y288	204	B3GLCT-CDG	糖基化障碍
CHST	113189	0169105	Q8NCH0	205	CHST14-CDG	糖基化

[0355]

[0356]

14						障碍
COG1	9382	0166685	Q8WTW3	206	COG1-CDG	糖基化障碍
COG2	22796	0135775	Q14746、B1ALW7	207	COG2-CDG	糖基化障碍(Glycosylation disorder)
COG4	25839	0103051	A0A0A0MS45、Q8N8L9、Q9H9E3、J3KNI1	208	COG4-CDG	糖基化障碍
COG5	10466	0164597,0284369	Q9UP83	209	COG5-CDG	糖基化障碍
COG6	57511	0133103	A0A140VJG7、Q9Y2V7、A0A024RDW5	210	COG6-CDG	糖基化障碍
COG7	91949	0168434	A0A0S2Z652、P83436	211	COG7-CDG	糖基化障碍
COG8	84342	0272617	A0A024R6Z6、Q96MW5	212	COG8-CDG	糖基化障碍
DOLK	22845	0175283	A0A0S2Z597、Q9UPQ8	213	DOLK-CDG	糖基化障碍
DHDDS	79947	0117682	Q86SQ9	214	DHDDS-CDG	糖基化障碍
DPAGT1	1798	0172269	A0A024R3H8、Q9H3H5	215	DPAGT1-CDG	糖基化障碍
DPM1	8813	0000419	O60762、Q5QPK2、A0A0S2Z4Y5	216	DPM1-CDG	糖基化障碍
DPM2	8818	0136908	O94777	217	DPM2-CDG	糖基化障碍
DPM3	54344	0179085	A0A140VJI4、Q9P2X0、Q86TM7	218	DPM3-CDG	糖基化障碍
G6PC3	92579	0141349	Q9BUM1	219	先天性中性粒细胞减少症	糖基化障碍
GFPT1	2673	0198380	Q06210	220	先天性肌无力综合征	糖基化障碍
GMPPA	29926	0144591	A0A024R482、Q96IJ6	221	GMPPA-CDG	糖基化障碍
GMPPB	29925	0173540	Q9Y5P6	222	先天性肌营养不良症、先天性肌无力综合征和抗肌萎缩相关糖蛋白病(dystroglycanopathy)	糖基化障碍

[0357]

MAGT1	84061	0102158	A0A087WU53、Q9H0U3	223	MAGT1-CDG;X连锁免疫缺陷伴有镁缺乏、爱泼斯坦-巴尔病毒感染和瘤形成(XMEN)综合征	糖基化障碍
MAN1B1	11253	0177239	Q9UKM7	224	MAN1B1-CDG	糖基化障碍
MGAT2	4247	0168282	Q10469	225	MGAT2-CDG	糖基化障碍
MOGS	7841	0115275	Q13724、Q58F09	226	MOGS-CDG	糖基化障碍
MPDU1	9526	0129255	J3QW43、O75352、A0A0S2Z4W8、B4DLH7	227	MPDU1-CDG	糖基化障碍
MPI	4351	0178802	H3BPP3、Q8NHZ6、B4DW50、F5GX71、P34949、H3BPB8	228	MPI-CDG	糖基化障碍
NGLY1	55768	0151092	Q96IV0	229	NGLY1-CDG	糖基化障碍
PGM1	5236	0079739	B7Z6C2、P36871、B4DDQ8	230	PGM1-CDG	糖基化障碍
PGM3	5238	0013375	O95394、A0A087WT27	231	PGM3-CDG	糖基化障碍
RFT1	91869	0163933	Q96AA3	232	RFT1-CDG	糖基化障碍
SEC23B	10483	0101310	Q15437、B4DJW8	233	SEC23B-CDG	糖基化障碍
SLC35A1	10559	0164414	P78382	234	SLC35A1-CDG	糖基化障碍
SLC35A2	7355	0102100	P78381、A6NFI1、A6NKM8、B4DE15	235	SLC35A2-CDG	糖基化障碍
SLC35C1	55343	0181830	Q96A29、B3KQH0	236	SLC35C1-CDG	糖基化障碍
SSR4	6748	0180879	P51571	237	SSR4-CDG	糖基化障碍
SRD5A3	79644	0128039	Q9H8P0	238	SRD5A3-CDG	糖基化障碍
TMEM165	55858	0134851	Q9HC07	239	TMEM165-CDG	糖基化障碍

[0358]

TRIP11	9321	0100815	Q15643	240	TRIP11-CDG	糖基化障碍
TUSC3	7991	0104723	Q13454	241	TUSC3-CDG	糖基化障碍
ALG14	199857	0172339	Q96F25	242	ALG14-CDG	糖基化障碍
B4GALT1	2683	0086062	P15291、W6MEN3	243	B4GALT1-CDG	糖基化障碍
DDOST	1650	0244038	A0A024RAD5、P39656	244	DDOST-CDG	糖基化障碍
NUS1	116150	0153989	Q96E22	245	NUS1-CDG	糖基化障碍
RPN2	6185	0118705	P04844	246	RPN2-CDG	糖基化障碍
SEC23A	10484	0100934	Q15436	247	SEC23A-CDG	糖基化障碍
SLC35A3	23443	0117620	Q9Y2D2、A0A1W2PRT7、A0A1W2PSD1、A0A1W2PQL8	248	SLC35A3-CDG	糖基化障碍
ST3GAL3	6487	0126091	Q11203	249	ST3GAL3-CDG	糖基化障碍
STT3A	3703	0134910	P46977	250	STT3A-CDG	糖基化障碍
STT3B	201595	0163527	Q8TCJ2	251	STT3B-CDG	糖基化障碍
AGA	175	0038002	P20933	252	天冬氨酰葡糖胺尿症	溶酶体贮积症
ARSA	410	0100299	A0A0C4DFZ2、B4DVI5、P15289	253	异染性脑白质营养不良症	溶酶体贮积症
ARSB	411	0113273	A0A024RAJ9、P15848、A8K4A0	254	粘多糖贮积症 VI 型	溶酶体贮积症
ASAH1	427	0104763	A8K0B6、Q13510、Q53H01	255	法伯病	溶酶体贮积症
ATP13A2	23400	0159363	Q8N4D4、Q9NQ11、Q8NBS1	256	神经元蜡样质脂褐质沉积症 12 (CLN12)、Kufor-Rakeb 综合征(KRS)	溶酶体贮积症
CLN3	1201	0188603,0261832	A0A024QZB8、Q13286、B4DMY6、Q2TA70、B4DFF3	257	神经元蜡样质脂褐质沉积症 3 (CLN3)	溶酶体贮积症
CLN5	1203	0102805	A0A024R644、O75503	258	神经元蜡样质脂褐质沉积症 5 (CLN5)	溶酶体贮积症

[0359]

CLN6	54982	0128973	A0A024R601、Q9NWW5	259	神经元蜡样质脂褐质沉积症 6 (CLN6)	溶酶体贮积症
CLN8	2055	0182372,0278220	A0A024QZ57、Q9UBY8	260	神经元蜡样质脂褐质沉积症 8 (CLN8)	溶酶体贮积症
CTNS	1497	0040531	A0A0S2Z3I9、O60931、A0A0S2Z3K3	261	胱氨酸病	溶酶体贮积症
CTSA	5476	0064601	P10619、X6R8A1、B4E324、X6R5C5	262	半乳糖唾液酸苷贮积症	溶酶体贮积症
CTSD	1509	0117984	P07339、V9HWI3	263	神经元蜡样质脂褐质沉积症 10 (CLN10)	溶酶体贮积症
CTSF	8722	0174080	Q9UBX1	264	神经元蜡样质脂褐质沉积症 13 (CLN13)	溶酶体贮积症
CTSK	1513	0143387	P43235	265	致密性成骨不全症	溶酶体贮积症
DNAJC5	80331	0101152	Q6AHX3、Q9H3Z4	266	神经元蜡样质脂褐质沉积症 4 (CLN4)	溶酶体贮积症
FUCA1	2517	0179163	P04066、B5MDC5	267	岩藻糖苷沉积症	溶酶体贮积症
GAA	2548	0171298	P10253	268	庞贝病	溶酶体贮积症
GALC	2581	0054983	A0A0A0MQV0、P54803	269	克拉伯病	溶酶体贮积症
GALNS	2588	0141012	P34059、Q96I49、Q6YL38	270	粘多糖贮积症 IVa 型	溶酶体贮积症
GLA	2717	0102393	P06280、Q53Y83	271	法布里病	溶酶体贮积症
GLB1	2720	0170266	P16278、B7Z6Q5	272	GM1 神经节苷脂贮积病、粘多糖贮积症 IVb	溶酶体贮积症
GM2A	2760	0196743	P17900	273	GM2 神经节苷脂贮积病，AB 变体	溶酶体贮积症
GNPTAB	79158	0111670	Q3T906	274	粘脂贮积症 II 型 α/β、粘脂贮积症 III 型 α/β	溶酶体贮积症
GNPTG	84572	0090581	Q9UJJ9	275	粘脂贮积症 III γ	溶酶体贮积症
GNS	2799	0135677	A0A024RBC5、P15586、Q7Z3X3	276	粘多糖贮积症 IIID 型	溶酶体贮积症
GRN	2896	0030582	P28799	277	神经元蜡样质	溶酶体

[0360]

					脂褐质沉积症 11 (CLN11)、额颞叶痴呆	贮积症
GUSB	2990	0169919	P08236	278	粘多糖贮积症 VII 型	溶酶体贮积症
HEXA	3073	0213614	A0A0S2Z3W3、P06865、B4DVA7、H3BP20	279	泰-萨二氏病	溶酶体贮积症
HEXB	3074	0049860	A0A024RAJ6、P07686、Q5URX0	280	山德霍夫病 (Sandhoff disease)	溶酶体贮积症
HGSNAT	138050	0165102	Q68CP4、Q8IVU6	281	粘多糖贮积症 IIIC 型	溶酶体贮积症
HYAL1	3373	0114378	A0A024R2X3、Q12794、B3KUI5、A0A0S2Z3Q0	282	粘多糖贮积症 IX 型	溶酶体贮积症
IDS	3423	0010404	P22304、B4DGD7	283	粘多糖贮积症 II 型	溶酶体贮积症
IDUA	3425	0127415	P35475	284	粘多糖贮积症 I 型	溶酶体贮积症
KCTD7	154881	0243335	Q96MP8、A0A024RDN7	285	神经元蜡样质脂褐质沉积症 14 (CLN14)	溶酶体贮积症
LAMP2	3920	0005893	P13473	286	Danon 病	溶酶体贮积症
MAN2B1	4125	0104774	O00754、A8K6A7	287	α-甘露糖苷贮积症	溶酶体贮积症
MANBA	4126	0109323	O00462	288	β-甘露糖苷贮积症	溶酶体贮积症
MCOLN1	57192	0090674	Q9GZU1	289	粘脂贮积症 IV 型	溶酶体贮积症
MFSD8	256471	0164073	Q8NHS3	290	神经元蜡样质脂褐质沉积症 7 (CLN7)	溶酶体贮积症
NAGA	4668	0198951	A0A024R1Q5、P17050	291	辛德勒病 (Schindler disease)	溶酶体贮积症
NAGLU	4669	0108784	A0A140VJE4、P54802	292	粘多糖贮积症 IIIB	溶酶体贮积症
NEU1	4758	0204386,0227315,0227129,0223957,0234846,0184494,0228691,0234343	Q5JQI0、Q99519	293	粘脂贮积症 I 型、唾液贮积症 I 型	溶酶体贮积症

[0361]

NPC1	4864	0141458	O15118	294	尼曼-皮克病 C 型	溶酶体贮积症
NPC2	10577	0119655	A0A024R6C0、P61916、G3V3E8	295	尼曼-皮克病 C 型	溶酶体贮积症
SGSH	6448	0181523	P51688	296	粘多糖贮积症 IIIA	溶酶体贮积症
PPT1	5538	0131238	P50897	297	神经元蜡样质脂褐质沉积症 1 (CLN1)	溶酶体贮积症
PSAP	5660	0197746	P07602、A0A024QZQ2	298	前列腺血清酸性磷酸酶缺乏症、SapA 缺乏症(Krabbe 变体)、SapB 缺乏症(MLD 变体)、SapC 缺乏症(Gaucher 变体)	溶酶体贮积症
SLC17A5	26503	0119899	Q9NRA2	299	婴儿唾液酸贮积病、萨拉病	溶酶体贮积症
SMPD1	6609	0166311	P17405、Q59EN6、E9LUE8、Q8IUN0、E9LUE9	300	尼曼-皮克病 A 和 B 型	溶酶体贮积症
SUMF1	285362	0144455	Q8NBK3	301	多硫酸酯酶缺乏症	溶酶体贮积症
TPP1	1200	0166340	O14773	302	神经元蜡样质脂褐质沉积症 2 (CLN2)	溶酶体贮积症
AHCY	191	0101444	P23526、Q1RMG2	303	高蛋氨酸血症	氨基酸代谢病
GNMT	27232	0124713	A0A0S2Z5F2、Q14749、V9HW60	304	高蛋氨酸血症	氨基酸代谢病
MAT1A	4143	0151224	Q00266	305	高蛋氨酸血症	氨基酸代谢病
GCH1	2643	0131979	A0A024R642、P30793、Q8IZH9	306	BH4 辅因子缺乏症	氨基酸代谢病
PCBD1	5092	0166228	P61457	307	BH4 辅因子缺乏症	氨基酸代谢病
PTS	5805	0150787	Q03393	308	BH4 辅因子缺乏症	氨基酸代谢病
QDPR	5860	0151552	A0A140VKA9、P09417	309	BH4 辅因子缺乏症	氨基酸代谢病
SPR	6697	0116096	P35270	310	BH4 辅因子缺乏症	氨基酸代谢病
DNAJ	56521	0108176	Q6IAH1、	311	苯丙氨酸、酪氨	氨基酸

	C12			Q9UKB3		酸和色氨酸羟化酶热休克共陪伴蛋白缺乏症	代谢病
	ALDH4A1	8659	0159423	P30038、A0A024RAD8	312	高脯氨酸血症	氨基酸代谢病
	PRODH	5625	0100033	O43272	313	高脯氨酸血症	氨基酸代谢病
	HPD	3242	0158104	P32754	314	酪氨酸血症 II 型	氨基酸代谢病
	GBA	2629	0177628,0262446	A0A068F658、P04062、B7Z6S9	315	戈谢病	
	HGD	3081	0113924	Q93099、B3KW64	316	黑酸尿症	
	AMN	81693	0166126	Q9BXJ7、B3KP64	317	甲基丙二酸血症合并高胱氨酸尿症	有机酸血症
	CD320	51293	0167775	Q9NPF0	318	甲基丙二酸血症合并高胱氨酸尿症	有机酸血症
	CUBN	8029	0107611	O60494	319	甲基丙二酸血症合并高胱氨酸尿症	有机酸血症
[0362]	GIF	2694	0134812	P27352	320	甲基丙二酸血症合并高胱氨酸尿症	有机酸血症
	TCN1	6947	0134827	P20061	321	甲基丙二酸血症合并高胱氨酸尿症	有机酸血症
	TCN2	6948	0185339	P20062	322	甲基丙二酸血症合并高胱氨酸尿症	有机酸血症
	PREPL	9581	0138078	Q4J6C6	323	胱氨酸尿症	氨基酸代谢病
	PHGDH	26227	0092621	O43175	324	丝氨酸生物合成障碍	氨基酸代谢病
	PSAT1	29968	0135069	A0A024R280、Q9Y617、A0A024R222	325	丝氨酸生物合成障碍	氨基酸代谢病
	PSPH	5723	0146733	A0A024RDL3、P78330	326	丝氨酸生物合成障碍	氨基酸代谢病
	AMT	275	0145020	A0A024R2U7、P48728	327	甘氨酸脑病	氨基酸代谢病
	GCSH	2653	0140905	P23434	328	甘氨酸脑病	氨基酸代谢病
	GLDC	2731	0178445	P23378	329	甘氨酸脑病	氨基酸代谢病

[0363]

LIAS	11019	0121897	O43766、Q6P5Q6、B4E0L7、A0A024R9W0、A0A1W2PQE9、A0A1X7SBR7	330	甘氨酸脑病	氨基酸代谢病
NFU1	27247	0169599	Q9UMS0	331	甘氨酸脑病	氨基酸代谢病
SLC6A9	6536	0196517	P48067、B7Z3W8、B7Z589	332	甘氨酸脑病	氨基酸代谢病
SLC2A1	6513	0117394	P11166、Q59GX2	333	葡萄糖转运蛋白 1 型缺乏症	碳水化合物障碍
ATP7A	538	0165240	B4DRW0、Q04656、Q762B6	334	ATP7A 相关障碍 铜代谢障碍	金属转运障碍
AP1S1	1174	0106367	A0A024QYT6、P61966	335	铜代谢障碍	金属转运障碍
CP	1356	0047457	A5PL27、P00450	336	铜代谢障碍	金属转运障碍
SLC33A1	9197	0169359	O00400	337	铜代谢障碍	金属转运障碍(Metal transport disorder)
PEX7	5191	0112357	O00628、Q6FGN1	338	成人雷弗苏姆病(Refsum Disease) 肢根型点状软骨发育不良系列	过氧化物酶体障碍(Peroxisomal disorder)
PHYH	5264	0107537	O14832	339	成人雷弗苏姆病	过氧化物酶体障碍
AGPS	8540	0018510	O00116、B7Z3Q4	340	肢根型点状软骨发育不良系列	过氧化物酶体障碍
GNPAT	8443	0116906	O15228	341	肢根型点状软骨发育不良系列	过氧化物酶体障碍
ABCD1	215	0101986	P33897	342	X-连锁肾上腺脑白质营养不良	过氧化物酶体障碍
ACOX1	51	0161533	Q15067	343	X-连锁肾上腺脑白质营养不良	过氧化物酶体障碍
PEX1	5189	0127980	O43933、	344	X-连锁肾上腺	过氧化

[0364]

			A0A0C4DG33、B4DER6		脑白质营养不良	物酶体障碍
PEX2	5828	0164751	P28328	345	X-连锁肾上腺脑白质营养不良	过氧化物酶体障碍
PEX3	8504	0034693	P56589	346	X-连锁肾上腺脑白质营养不良	过氧化物酶体障碍
PEX5	5830	0139197	A0A0S2Z480、P50542、B4DR50、A0A0S2Z4F3、A0A0S2Z4H1、B4E0T2	347	X-连锁肾上腺脑白质营养不良	过氧化物酶体障碍
PEX6	5190	0124587	A0A024RD09、Q13608	348	X-连锁肾上腺脑白质营养不良	过氧化物酶体障碍
PEX10	5192	0157911	A0A024R068、O60683、A0A024R0A4	349	X-连锁肾上腺脑白质营养不良	过氧化物酶体障碍
PEX12	5193	0108733	O00623	350	X-连锁肾上腺脑白质营养不良	过氧化物酶体障碍
PEX13	5194	0162928	Q92968	351	X-连锁肾上腺脑白质营养不良	过氧化物酶体障碍
PEX14	5195	0142655	O75381	352	X-连锁肾上腺脑白质营养不良	过氧化物酶体障碍
PEX16	9409	0121680	Q9Y5Y5	353	X-连锁肾上腺脑白质营养不良	过氧化物酶体障碍
PEX19	5824	0162735	P40855、A0A0S2Z497	354	X-连锁肾上腺脑白质营养不良	过氧化物酶体障碍
PEX26	55670	0215193	A0A024R100、Q7Z412、A0A0S2Z5M7、Q7Z2D7	355	X-连锁肾上腺脑白质营养不良	过氧化物酶体障碍
AMACR	23600	0242110	Q9UHK6	356	Zellweger 谱系障碍	过氧化物酶体障碍
ADA	100	0196839	A0A0S2Z381、P00813、F5GWI4	357	嘌呤代谢障碍	嘌呤代谢障碍
ADSL	158	0239900	P30566、X5D8S6、X5D7W4、	358	嘌呤代谢障碍	嘌呤代谢障碍

[0365]

			A0A1B0GWJ0			
AMPD1	270	0116748	P23109	359	嘌呤代谢障碍	嘌呤代谢障碍
GPHN	10243	0171723	Q9NQX3	360	嘌呤代谢障碍	嘌呤代谢障碍
MOCOS	55034	0075643	Q96EN8	361	嘌呤代谢障碍	嘌呤代谢障碍
MOCS1	4337	0124615	A0A024RD17、Q9NZB8	362	嘌呤代谢障碍	嘌呤代谢障碍
PNP	4860	0198805	P00491、V9HWH6	363	嘌呤代谢障碍	嘌呤代谢障碍
XDH	7498	0158125	P47989	364	嘌呤代谢障碍	嘌呤代谢障碍
SUOX	6821	0139531	A0A024RB79、P51687	365	嘌呤代谢障碍	嘌呤代谢障碍
OGDH	4967	0105953	A0A140VJQ5、Q02218、B4E3E9、E9PCR7、E9PDF2	366	2-酮戊二酸脱氢酶缺乏症	丙酮酸代谢和三羧酸循环缺陷
SLC25A19	60386	0125454	Q5JPC1、Q9HC21	367	2-酮戊二酸脱氢酶缺乏症	丙酮酸代谢和三羧酸循环缺陷
DHTKD1	55526	0181192	Q96HY7	368	2-酮戊二酸脱氢酶缺乏症	丙酮酸代谢和三羧酸循环缺陷
SLC13A5	284111	0141485	Q68D44、Q86YT5	369	柠檬酸转运蛋白缺乏症	丙酮酸代谢和三羧酸循环缺陷
FH	2271	0091483	A0A0S2Z4C3、P07954	370	延胡索酸酶缺乏症	丙酮酸代谢和三羧酸循环缺陷
DLAT	1737	0150768	P10515、Q86YI5	371	丙酮酸脱氢酶缺乏症	丙酮酸代谢和三羧酸循环缺陷
MPC1	51660	0060762	Q5TI65、Q9Y5U8	372	丙酮酸脱氢酶缺乏症	丙酮酸代谢和

						三羧酸循环缺陷
PDHA1	5160	0131828	A0A024RBX9、P08559	373	丙酮酸脱氢酶缺乏症	丙酮酸代谢和三羧酸循环缺陷
PDHB	5162	0168291	P11177	374	丙酮酸脱氢酶缺乏症	丙酮酸代谢和三羧酸循环缺陷
PDHX	8050	0110435	O00330	375	丙酮酸脱氢酶缺乏症	丙酮酸代谢和三羧酸循环缺陷
PDP1	54704	0164951	Q9P0J1、Q6P1N1、A0A024R9C0	376	丙酮酸脱氢酶缺乏症	丙酮酸代谢和三羧酸循环缺陷
ABCC2	1244	0023839	Q92887	377	杜宾-约翰逊综合征	
SLCO1B1	10599	0134538	A0A024RAU7、Q05CV5、Q9Y6L6	378	罗托综合征(Rotor Syndrome)	
SLCO1B3	28234	0111700	B3KP78、Q9NPD5	379	罗托综合征	
HFE2	148738	0168509	Q6ZVN8、A8K466、A0A024R4F5	380	血色素沉着症，2A 型	
ADAMTS13	11093	0160323,0281244	Q76LX8	381	由于 ADAMTS-13 缺乏引起的先天性血栓性血小板减少性紫癜	
PYGM	5837	0068976	P11217	382	麦卡德尔氏病(McArdle's Disease)	
COL1A2	1278	0164692	A0A0S2Z3H5、P08123	383	埃勒斯-当洛斯综合征，心脏瓣膜 型	
TNFRSF11B	4982	0164761	O00300	384	幼年型佩吉特病	
TSC1	7248	0165699	Q86WV8、Q92574、	385	结节性硬化症	

[0366]

[0367]

			X5D9D2、Q32NF0			
TSC2	7249	0103197	P49815、X5D7Q2、B3KWH7、Q5HYF7、H3BMQ0、X5D2U8	386	结节性硬化症	
DHCR7	1717	0172893	A0A024R5F7、Q9UBM7	387	Smith-Lemli-Opitz 综合征	
PGK1	5230	0102144	P00558、V9HWF4	388	D-甘油酸血症	
VLDLR	7436	0147852	P98155、Q5VVF5	389	平衡失调综合征	
KYNU	8942	0115919	Q16719	390	由于羟基犬尿酸尿症引起的脑病	
F5	2153	0198734	P12259	391	因子 V 缺乏症	
C3	718	0125730	B4DR57、P01024、V9HWA9	392	非典型溶血性尿毒症综合征伴有 C3 异常	
COL4A1	1282	0187498	A5PKV2、F5H5K0、P02462	393	常染色体显性家族性血尿症-视网膜小动脉迂曲-挛缩	
CFH	3075	0000971	A0A024R962、P08603、A0A0D9SG88	394	非典型溶血性尿毒症综合征	
SLC12A2	6558	0064651	P55011、Q53ZR1、B7ZM24	395	巴特综合征 I 型(新生儿)	
GK	2710	0198814	B4DH54、P32189	396	甘油激酶缺乏症	
SFTPC	6440	0168484	A0A0A0MTC9、P11686、A0A0S2Z4Q0、E5RI64	397	慢性呼吸窘迫伴有表面活性物质代谢缺陷	
CRTAP	10491	0170275	O75718	398	成骨不全 VII	
P3H1	64175	0117385	Q32P28	399	成骨不全 VIII	
COL7A1	1294	0114270	Q02388、Q59F16	400	常染色体隐性营养不良大疱性表皮松解症	
PKLR	5313	0143627	P30613	401	丙酮酸激酶缺乏症	
TALDO1	6888	0177156	A0A140VK56、P37837	402	转醛醇酶缺乏症	
TF	7018	0091513	A0PJA6、	403	无转铁蛋白血	

[0368]

			P02787、Q06AH7		症(家族性低转铁蛋白血症)	
EPCAM	4072	0119888	P16422	404	肠上皮发育不良	
VHL	7428	0134086	A0A024R2F2、P40337、A0A0S2Z4K1	405	家族性红细胞增多症 2 型；希-林二氏病	
GC	2638	0145321	P02774	406	维生素 D 缺乏症	
SERPINA1	5265	0197249,0277377	E9KL23、P01009	407	α-1 抗胰蛋白酶缺乏症	
ABCC6	368	0091262,0275331	O95255	408	弹性假黄瘤	
F8	2157	0185010	P00451	409	血友病 A	
F9	2158	0101981	P00740	410	血友病 B	
ApoB	338	0084674	P04114	411	家族性高胆固醇血症	
PCSK9	255738	0169174	Q8NBP7	412	家族性高胆固醇血症	
LDLRAP1	26119	0157978	B3KR97、Q5SW96	413	家族性高胆固醇血症	
ABCG5	64240	0138075	Q9H222	414	谷甾醇血症	
ABCG8	64241	0143921	Q9H221	415	谷甾醇血症	
LCAT	3931	0213398	A0A140VK24、P04180	416	卵磷脂胆固醇酰基转移酶缺乏症	
SPINK5	11005	0133710	Q9NQ38	417	内瑟顿综合征	
GNE	10020	0159921	Q9Y223	418	包涵体肌病 2	

[0369] 在一些实施方案中，所述靶向脂质颗粒或慢病毒载体含有能够靶向T细胞的外源药剂。在一些实施方案中，所述能够靶向T细胞的外源药剂是嵌合抗原受体(CAR)、T细胞受体、整联蛋白、离子通道、成孔蛋白、Toll样受体、白介素受体、细胞粘附蛋白或转运蛋白。

[0370] 在一些实施方案中，所述CAR是或包含含有抗原结合结构域、跨膜结构域和信号传导结构域(例如，一个、两个或三个信号传导结构域)的第一代CAR。在一些实施方案中，所述CAR包含含有抗原结合结构域、跨膜结构域和至少三个信号传导结构域的第三代CAR。在一些实施方案中，第四代CAR包含抗原结合结构域、跨膜结构域、三个或四个信号传导结构域和在CAR成功信号传导时诱导细胞因子基因表达的结构域。在一些实施方案中，所述抗原结合结构域是或包含scFv或Fab。

[0371] 在一些实施方案中，CAR抗原结合结构域是或包含抗体或其抗原结合部分。在一些实施方案中，CAR抗原结合结构域是或包含scFv或Fab。在一些实施方案中，CAR抗原结合结构域包含以下的scFv或Fab片段：T细胞α链抗体；T细胞β链抗体；T细胞γ链抗体；T-细胞δ链抗体；CCR7抗体；CD3抗体；CD4抗体；CD5抗体；CD7抗体；CD8抗体；CD11b抗体；CD11c抗体；CD16抗体；CD19抗体；CD20抗体；CD21抗体；CD22抗体；CD25抗体；CD28抗体；CD34抗体；CD35

抗体;CD40抗体;CD45RA抗体;CD45RO抗体;CD52抗体;CD56抗体;CD62L抗体;CD68抗体;CD80抗体;CD95抗体;CD117抗体;CD127抗体;CD133抗体;CD137(4-1BB)抗体;CD163抗体;F4/80抗体;IL-4Ra抗体;Sca-1抗体;CTLA-4抗体;GITR抗体GARP抗体;LAP抗体;粒酶B抗体;LFA-1抗体;MR1抗体;uPAR抗体;或运铁蛋白受体抗体。

[0372] 在一些实施方案中,CAR结合结构域结合细胞的细胞表面抗原。在一些实施方案中,细胞表面抗原是一种类型的细胞的特征。在一些实施方案中,细胞表面抗原是多于一种类型的细胞的特征。

[0373] 在一些实施方案中,CAR的抗原结合结构域靶向表征T细胞的抗原。在一些实施方案中,所述表征T细胞的抗原选自表征T细胞的细胞表面受体、膜转运蛋白(例如,主动或被动转运蛋白,例如离子通道蛋白、成孔蛋白等)、跨膜受体、膜酶和/或细胞粘附蛋白。在一些实施方案中,表征T细胞的抗原可以是G蛋白偶联受体、受体酪氨酸激酶、酪氨酸激酶相关受体、受体样酪氨酸磷酸酶、受体丝氨酸/苏氨酸激酶、受体鸟苷酸环化酶、组氨酸激酶相关受体、AKT1、AKT2、AKT3、ATF2、BCL10、CALM1、CD3D(CD3δ)、CD3E(CD3ε)、CD3G(CD3γ)、CD4、CD8、CD28、CD45、CD80(B7-1)、CD86(B7-2)、CD247(CD3ζ)、CTLA4(CD152)、ELK1、ERK1(MAPK3)、ERK2、FOS、FYN、GRAP2(GADS)、GRB2、HLA-DRA、HLA-DRB1、HLA-DRB3、HLA-DRB4、HLA-DRB5、HRAS、IKBKA(CHUK)、IKBKB、IKBKE、IKBKG(NEMO)、IL2、ITPR1、ITK、JUN、KRAS2、LAT、LCK、MAP2K1(MEK1)、MAP2K2(MEK2)、MAP2K3(MKK3)、MAP2K4(MKK4)、MAP2K6(MKK6)、MAP2K7(MKK7)、MAP3K1(MEKK1)、MAP3K3、MAP3K4、MAP3K5、MAP3K8、MAP3K14(NIK)、MAPK8(JNK1)、MAPK9(JNK2)、MAPK10(JNK3)、MAPK11(p38β)、MAPK12(p38γ)、MAPK13(p38δ)、MAPK14(p38α)、NCK、NFAT1、NFAT2、NFKB1、NFKB2、NFKBIA、NRAS、PAK1、PAK2、PAK3、PAK4、PIK3C2B、PIK3C3(VPS34)、PIK3CA、PIK3CB、PIK3CD、PIK3R1、PKCA、PKCB、PKCM、PKCQ、PLCY1、PRF1(穿孔素)、PTEN、RAC1、RAF1、RELA、SDF1、SHP2、SLP76、SOS、SRC、TBK1、TCRA、TEC、TRAF6、VAV1、VAV2或ZAP70。

[0374] 在一些实施方案中,CAR的抗原结合结构域靶向表征疾病的抗原。在一些实施方案中,所述疾病或病况与CD4+T细胞相关。在一些实施方案中,所述疾病或病况与CD8+T细胞相关。

[0375] 在一些实施方案中,CAR跨膜结构域至少包含以下的跨膜区:T细胞受体的α、β或ζ链、CD28、CD3ε、CD45、CD4、CD5、CD8、CD9、CD16、CD22、CD33、CD37、CD64、CD80、CD86、CD134、CD137、CD154或其功能变体。在一些实施方案中,所述跨膜结构域至少包含以下的跨膜区:CD8α、CD8β、4-1BB/CD137、CD28、CD34、CD4、FcεRIγ、CD16、OX40/CD134、CD3ζ、CD3ε、CD3γ、CD3δ、TCRα、TCRβ、TCRζ、CD32、CD64、CD64、CD45、CD5、CD9、CD22、CD37、CD80、CD86、CD40、CD40L/CD154、VEGFR2、FAS和FGFR2B或其功能变体。

[0376] 在一些实施方案中,所述CAR包含选自以下中的一种或多种的至少一个信号传导结构域:B7-1/CD80、B7-2/CD86、B7-H1/PD-L1、B7-H2、B7-H3、B7-H4、B7-H6、B7-H7、BTLA/CD272、CD28、CTLA-4、Gi24/VISTA/B7-H5、ICOS/CD278、PD-1、PD-L2/B7-DC、PDCD6)、4-1BB/TNFSF9/CD137、4-1BB配体/TNFSF9、BAFF/BLyS/TNFSF13B、BAFF R/TNFRSF13C、CD27/TNFRSF7、CD27配体/TNFSF7、CD30/TNFRSF8、CD30配体/TNFSF8、CD40/TNFRSF5、CD40/TNFSF5、CD40配体/TNFSF5、DR3/TNFRSF25、GITR/TNFRS F18、GITR配体/TNFSF18、HVEM/TNFRSF14、LIGHT/TNFSF14、淋巴毒素-α/TNF-β、OX40/TNFRSF4、OX40配体/TNFSF4、RELT/

TNFRSF19L、TACI/TNFRSF13B、TL1A/TNFSF15、TNF-α、TNF RII/TNFRSF1B)、2B4/CD244/SLAMF4、BLAME/SLAMF8、CD2、CD2 F-10/SLAMF9、CD48/SLAMF2、CD58/LFA-3、CD84/SLAMF5、CD229/SLAMF3、CRACC/SLAMF7、NTB-A/SLAMF6、SLAM/CD150)、CD2、CD7、CD53、CD82/Kai-1、CD90/Thy1、CD96、CD160、CD 200、CD300a/LMIR1、HLA I类、HLA-DR、Ikaros、整联蛋白α4/CD49d、整联蛋白α4β1、整联蛋白α4β7/LPAM-1、LAG-3、TCL1A、TCL1B、CRTAM、DAP12、Dectin-1/CLEC7A、DPPIV/CD26、Eph B6、TIM-1/KIM-1/HAVCR、TIM-4、TSLP、TSLP R、淋巴细胞功能相关抗原1(LFA-1)、NKG2C、CD3ζ结构域、基于免疫受体酪氨酸的激活基序(ITAM)、CD27、CD28、4-1BB、CD134/OX40、CD30、CD40、PD-1、ICOS、淋巴细胞功能相关抗原1(LFA-1)、CD2、CD7、LIGHT、NKG2C、B7-H3、特异性地结合CD83的配体、或其功能片段。

[0377] 在一些实施方案中,所述CAR包含CD3ζ结构域或基于免疫受体酪氨酸的激活基序(ITAM)或其功能变体。在一些实施方案中,所述CAR包含(i)CD3ζ结构域或基于免疫受体酪氨酸的激活基序(ITAM)或其功能变体;和(ii)CD28结构域或4-1BB结构域或其功能变体。在一些实施方案中,所述CAR包含(i)CD3ζ结构域或基于免疫受体酪氨酸的激活基序(ITAM)或其功能变体;(ii)CD28结构域或其功能变体;和(iii)4-1BB结构域或CD134结构域或其功能变体。在一些实施方案中,所述CAR包含(i)CD3ζ结构域或基于免疫受体酪氨酸的激活基序(ITAM)或其功能变体;(ii)CD28结构域或4-1BB结构域或其功能变体,和/或(iii)4-1BB结构域或CD134结构域或其功能变体。在一些实施方案中,所述CAR包含(i)CD3ζ结构域或基于免疫受体酪氨酸的激活基序(ITAM)或其功能变体;(ii)CD28结构域或其功能变体;(iii)4-1BB结构域或CD134结构域或其功能变体;和(iv)细胞因子或共刺激配体转基因。

[0378] 在某些实施方案中,所述细胞内信号传导结构域包含与CD3(例如CD3-ζ)细胞内结构域连接的CD28跨膜和信号传导结构域。在一些实施方案中,所述细胞内信号传导结构域包含与CD3ζ细胞内结构域连接的嵌合CD28和CD137(4-1BB、TNFRSF9)共刺激结构域。

[0379] 在一些实施方案中,所述CAR涵盖在胞质部分中的一个或多个(例如两个或更多个)共刺激结构域和激活结构域(例如初级激活结构域)。示例性CAR包括CD3-ζ、CD28和4-1BB的细胞内组分。

[0380] 在一些实施方案中,所述细胞内信号传导结构域包括4-1BB信号传导结构域和CD3-ζ信号传导结构域的细胞内组分。在一些实施方案中,所述细胞内信号传导结构域包括CD28信号传导结构域和CD3ζ信号传导结构域的细胞内组分。

[0381] 在一些实施方案中,所述CAR包含与抗原(例如肿瘤抗原)结合的细胞外抗原结合结构域(例如抗体或抗体片段,诸如scFv)、间隔子(例如含有铰链结构域,诸如本文所述的任何一种)、跨膜结构域(例如本文所述的任何一种)和细胞内信号传导结构域(例如任何细胞内信号传导结构域,诸如本文所述的初级信号传导结构域或共刺激信号传导结构域)。在一些实施方案中,所述细胞内信号传导结构域是或包括初级胞质信号传导结构域。在一些实施方案中,所述细胞内信号传导结构域另外包括共刺激分子的细胞内信号传导结构域(例如,共刺激结构域)。CAR的示例性组分的实例描述于表6中。在所提供的方面,CAR中的每个组分的序列可包括表6中列出的任何组合。

表 6: CAR 组分和示例性序列

组分	序列	SEQ ID NO
细胞外结合结构域		
抗 CD19 scFv (FMC63)	DIQMTQTTSSLSASLGDRVTISCRASQDISK YLNWYQQKPDGTVKLLIYHTSRLHSGVPSR FSGSGSGTDYSLTISNLEQEDIATYFCQQGN TLPYTFGGGTKLEITGSTSGSGKPGSGEGST KGEVKLQESGPGLVAPSQSLSVTCTVSGVS LPDYGVSWIRQPPRKGLEWLGVIWGSETTY YNSALKSRLTIIKDNSKSQVFLKMNSLQTD DTAIYYCAKHYYYGGSYAMDYWGQGTSV TVSS	419
抗 CD19 scFv (FMC63)	DIQMTQTTSSLSASLGDRVTISCRASQDISK YLNWYQQKPDGTVKLLIYHTSRLHSGVPSR FSGSGSGTDYSLTISNLEQEDIATYFCQQGN TLPYTFGGGTKLEITGGGGSGGGGSGGGGS EVKLQESGPGLVAPSQSLSVTCTVSGVSLP DYGVSWIRQPPRKGLEWLGVIWGSETTYY NSALKSRLTIIKDNSKSQVFLKMNSLQTDD TAIYYCAKHYYYGGSYAMDYWGQGTSVT VSS	420
间隔子(例如铰链)		
IgG4 铰链	ESKYGPPCPPCP	421
CD8 铰链	TTTPAPRPPTPAPTIASQPLSLRPE	422
CD28	IEVMYPPPYLDNEKSNGTIIHVKGKHLCPSP LFPGPSKP	423
跨膜		
CD8	ACRPAAGGAVHTRGLDFACDIYIWAPLAG TCGVLLLSLVITLYC	424
CD28	FWVLVVVGGVLACYSLLVTVAFIIFWV	425
CD28	FWVLVVVGGVLACYSLLVTVAFIIFWV	426
共刺激结构域		
CD28	RSKRSRLLHSDYMNMTPRRPGPTRKHYQP YAPPRDFAAYRS	427
4-1BB	KRGRKKLLYIFKQPFMRPVQTTQEEDGCSC RFPEEEEGGCEL	428
初级信号传导结构域		
CD3ζ	RVKFSRSADAPAYQQGQNQLYNELNLGRR EEYDVLDKRRGRDPEMGGKPRRKNPQEGL YNELQKDKMAEAYSEIGMKGERRRGKGH DGLYQGLSTATKDTYDALHMQALPPR	429
CD3ζ	RVKFSRSADAPAYKQGQNQLYNELNLGRR EEYDVLDKRRGRDPEMGGKPRRKNPQEGL YNELQKDKMAEAYSEIGMKGERRRGKGH DGLYQGLSTATKDTYDALHMQALPPR	430

[0382]

[0383]

[0384] 在一些实施方案中,所述CAR还包含一个或多个间隔子,例如,其中所述间隔子是抗原结合结构域与跨膜结构域之间的第一间隔子。在一些实施方案中,所述第一间隔子包括免疫球蛋白恒定区或其变体或修饰版本的至少一部分。在一些实施方案中,所述间隔子是跨膜结构域与信号传导结构域之间的第二间隔子。在一些实施方案中,所述第二间隔子是寡肽,例如,其中所述寡肽包含甘氨酸-丝氨酸双联体。

[0385] 除了本文所述的CAR之外,各种嵌合抗原受体和编码其的核苷酸序列是已知的,并且将适合于如本文所述在体内和体外融合体递送和重编程靶细胞。参见例如WO2013040557;WO2012079000;WO2016030414;Smith T等人,Nature Nanotechnology.2017.(DOI:10.1038/NNANO.2017.57),其公开内容通过引用整体并入本文。

[0386] 在一些实施方案中,将包含CAR或编码CAR的核酸(例如,DNA、gDNA、cDNA、RNA、pre-MRNA、mRNA、miRNA、siRNA等)的靶向脂质颗粒递送至靶细胞。在一些实施方案中,所述靶细胞是效应细胞,例如表达一种或多种Fc受体并介导一种或多种效应子功能的免疫系统细胞。在一些实施方案中,所述靶细胞可包括但不限于单核细胞、巨噬细胞、嗜中性粒细胞、树突细胞、嗜酸性粒细胞、肥大细胞、血小板、大颗粒淋巴细胞、朗格汉斯细胞、自然杀伤(NK)细胞、T淋巴细胞(例如T细胞),γδT细胞、B淋巴细胞(例如B细胞)中的一种或多种,并且可以来自任何生物体,包括但不限于人、小鼠、大鼠、兔和猴。

[0387] E.产生靶向脂质颗粒的方法

[0388] 本文提供了靶向脂质颗粒,所述靶向脂质颗粒包含脂质双层、被脂质双层包围的内腔、靶向包膜蛋白和融合剂,其中所述靶向包膜蛋白和融合剂包埋在脂质双层内。在一些实施方案中,所述靶向脂质颗粒可以是病毒颗粒、病毒样颗粒、纳米颗粒、囊泡、外来体、树状聚合物、慢病毒、病毒载体、去核细胞、微泡、膜囊泡、细胞外膜囊泡、质膜囊泡、巨质膜囊泡、凋亡小体、有丝分裂颗粒、核红细胞、溶酶体、另一种膜包围的囊泡、或慢病毒载体、基于病毒的颗粒、病毒样颗粒(VLP)或细胞来源的颗粒。

[0389] 1.病毒样颗粒

[0390] 本文提供了源自病毒的靶向脂质颗粒,诸如病毒颗粒或病毒样颗粒,包括源自逆转录病毒或慢病毒的那些。在一些实施方案中,两亲性脂质的靶向脂质颗粒双层是或包含病毒包膜。在一些实施方案中,两亲性脂质的靶向脂质颗粒双层是或包含源自生产者细胞的脂质。在一些实施方案中,病毒包膜可包含融合剂,例如对病毒内源的融合剂或假型化融合剂。在一些实施方案中,靶向脂质颗粒的内腔或空腔包含病毒核酸,例如逆转录病毒核酸,例如慢病毒核酸。在一些实施方案中,病毒核酸可以是病毒基因组。在一些实施方案中,靶向脂质颗粒还包含一种或多种病毒非结构蛋白,例如在其空腔或内腔中。在一些实施方案中,靶向脂质颗粒是或包含病毒样颗粒(VLP)。在一些实施方案中,VLP不包含包膜。在一些实施方案中,VLP包含包膜。

[0391] 在一些实施方案中,病毒颗粒或病毒样颗粒(诸如逆转录病毒或逆转录病毒样颗粒)包含gag多蛋白、聚合酶(例如pol)、整合酶(例如功能或非功能变体)、蛋白酶和融合剂中的一种或多种。在一些实施方案中,靶向脂质颗粒还包含rev。在一些实施方案中,一种或多种上述蛋白质在逆转录病毒基因组中编码,并且在一些实施方案中,一种或多种上述蛋白质以反式例如由辅助细胞、辅助病毒或辅助质粒提供。在一些实施方案中,靶向脂质颗粒

核酸(例如,逆转录病毒核酸)包含一种或多种以下核酸序列:5'LTR(例如,包含U5并且缺乏功能性U3结构域)、Psi包装元件(Psi)、可操作地连接至有效负载基因的中央多嘌呤区(cPPT)启动子、有效负载基因(任选地在开放阅读框之前包含内含子)、聚A尾序列、WPRE和3'LTR(例如,包含U5并且缺乏功能性U3)。在一些实施方案中,靶向脂质颗粒核酸还包含一个或多个绝缘子元件。在一些实施方案中,识别位点位于聚A尾序列与WPRE之间。

[0392] 在一些实施方案中,靶向脂质颗粒包含由自组装成衣壳的病毒蛋白形成的超分子复合物。在一些实施方案中,靶向脂质颗粒是来源于病毒衣壳的病毒颗粒或病毒样颗粒。在一些实施方案中,靶向脂质颗粒是来源于病毒核衣壳的病毒颗粒或病毒样颗粒。在一些实施方案中,靶向脂质颗粒包含来源于核衣壳的保留包装核酸的性质。在一些实施方案中,病毒颗粒或病毒样颗粒仅包含病毒结构糖蛋白。在一些实施方案中,靶向脂质颗粒不含有病毒基因组。

[0393] 在一些实施方案中,靶向脂质颗粒在表达期间包装来自宿主细胞的核酸。在一些实施方案中,所述核酸不编码参与病毒复制的任何基因。在特定的实施方案中,靶向脂质颗粒是病毒样颗粒,例如逆转录病毒样颗粒,诸如慢病毒样颗粒,其是复制缺陷的。

[0394] 在一些情况下,靶向脂质颗粒是在形态上与野生型感染性病毒不可区分的病毒颗粒。在一些实施方案中,所述病毒颗粒将整个病毒蛋白质组作为抗原呈递。在一些实施方案中,所述病毒颗粒仅将蛋白质组的一部分作为抗原呈递。

[0395] 在一些实施方案中,利用来自副粘病毒科(Paramyxoviridae)家族内的病毒的蛋白质(例如包膜蛋白)产生病毒颗粒或病毒样颗粒。在一些实施方案中,所述副粘病毒科家族包含亨尼帕病毒属中的成员。在一些实施方案中,亨尼帕病毒是或包含亨德拉(HeV)或尼帕(NiV)病毒。在特定的实施方案中,病毒颗粒或病毒样颗粒掺入如章节I.A.和1.B.所述的靶向包膜蛋白和融合剂。

[0396] 在一些实施方案中,病毒颗粒或病毒样颗粒可以在多种细胞培养系统中产生,所述多种细胞培养系统包括细菌、哺乳动物细胞系、昆虫细胞系、酵母和植物细胞。

[0397] 在一些实施方案中,病毒颗粒或病毒样颗粒的组装通过核心蛋白与病毒基因组内的独特衣壳化序列(例如具有茎环结构的UTR)的结合来启动。在一些实施方案中,所述核心与所述衣壳化序列的相互作用促进寡聚化。

[0398] 在一些实施方案中,靶向脂质颗粒是包含缺少或缺乏病毒RNA的序列(这可能是从所述序列中除去或清除病毒RNA的结果)的病毒样颗粒。在一些实施方案中,这可以通过使用gag上的内源包装信号结合位点来实现。在一些实施方案中,所述内源包装信号结合位点在pol上。在一些实施方案中,待递送的RNA将含有同源包装信号。在一些实施方案中,位于待递送的RNA上的异源结合结构域(其与gag异源)和位于gag或pol上的同源结合位点可以用于确保待递送的RNA的包装。在一些实施方案中,所述异源序列可以是非病毒的或者其可以是病毒的,在这种情况下其可以源自不同的病毒。在一些实施方案中,所述载体颗粒可以用于递送治疗性RNA,在这种情况下不需要功能性整合酶和/或逆转录酶。在一些实施方案中,所述载体颗粒也可以用于递送感兴趣的治疗基因,在这种情况下通常包括pol。

[0399] a.转移载体

[0400] 在一些实施方案中,逆转录病毒核酸包含以下中的一种或多种(例如,全部):5'启动子(例如,用于控制整个包装的RNA的表达)、5'LTR(例如,其包括R(聚腺苷酸化尾信号)

和/或包括引物激活信号的U5)、引物结合位点、psi包装信号、用于核输出的RRE元件、直接位于转基因上游以控制转基因表达的启动子、转基因(或其他外源药剂元件)、多嘌呤区和3'LTR(例如,其包括突变的U3、R和U5)。在一些实施方案中,逆转录病毒核酸还包含cPPT、WPRE和/或绝缘子元件中的一种或多种。

[0401] 逆转录病毒通常通过将其基因组RNA逆转录成线性双链DNA拷贝并随后将其基因组DNA共价整合到宿主基因组中来复制。适合用于特定实施方案的示例性逆转录病毒包括但不限于:莫洛尼鼠白血病病毒(M-MuLV)、莫洛尼鼠肉瘤病毒(MoMSV)、哈维鼠肉瘤病毒(HaMuSV)、鼠乳腺肿瘤病毒(MuMTV)、长臂猿白血病病毒(GaLV)、猫白血病病毒(FLV)、泡沫病毒、Friend鼠白血病病毒、鼠干细胞病毒(MSCV)和劳斯肉瘤病毒(RSV)和慢病毒。

[0402] 在一些实施方案中,所述逆转录病毒是γ逆转录病毒。在一些实施方案中,所述逆转录病毒是ε逆转录病毒。在一些实施方案中,所述逆转录病毒是α逆转录病毒。在一些实施方案中,所述逆转录病毒是β逆转录病毒。在一些实施方案中,所述逆转录病毒是δ逆转录病毒。在一些实施方案中,所述逆转录病毒是慢病毒。在一些实施方案中,所述逆转录病毒是泡沫逆转录病毒。在一些实施方案中,所述逆转录病毒是内源性逆转录病毒。

[0403] 例示性的慢病毒包括但不限于:HIV(人免疫缺陷病毒;包括HIV1型和HIV 2型);维士那-梅地病毒(VMV)病毒;山羊关节炎-脑炎病毒(CAEV);马传染性贫血病毒(EIAV);猫免疫缺陷病毒(FIV);牛免疫缺陷病毒(BIV);和猿猴免疫缺陷病毒(SIV)。在一些实施方案中,使用基于HIV的载体骨架(即HIV顺式作用序列元件)。

[0404] 在一些实施方案中,本文的载体是能够转移或转运另一核酸分子的核酸分子。转移的核酸通常与载体核酸分子连接,例如插入载体核酸分子中。载体可包括在细胞中指导自主复制的序列,或可包括足以允许整合到宿主细胞DNA中的序列。有用的载体包括例如质粒(例如DNA质粒或RNA质粒)、转座子、粘粒、细菌人工染色体和病毒载体。有用的病毒载体包括例如复制缺陷逆转录病毒和慢病毒。

[0405] 在一些实施方案中,病毒载体包含包括病毒来源的核酸元件的核酸分子(例如转移质粒),所述病毒来源的核酸元件通常促进核酸分子的转移或整合到细胞基因组或介导核酸转移的病毒颗粒中。病毒颗粒通常将包括各种病毒组分,并且有时除了核酸外还包括宿主细胞组分。在一些实施方案中,病毒载体包含例如能够将核酸转移到细胞中或转移到转移核酸(例如作为裸DNA)中的病毒颗粒或病毒性颗粒。在一些实施方案中,病毒载体和转移质粒包含主要来源于病毒的结构和/或功能遗传元件。逆转录病毒载体可包含病毒载体或质粒,其含有主要来源于逆转录病毒的结构和功能遗传元件或其部分。慢病毒载体可包含含有结构和功能遗传元件或其部分的病毒载体或质粒,包括主要来源于慢病毒的LTR。

[0406] 在实施方案中,慢病毒载体(例如慢病毒表达载体)可包含慢病毒转移质粒(例如作为裸DNA)或感染性慢病毒颗粒。关于诸如克隆位点、启动子、调控元件、异源核酸等元件,应理解这些元件的序列可以在慢病毒颗粒中以RNA形式存在,并且可以在DNA质粒中以DNA形式存在。

[0407] 在一些实施方案中,在本文所述的载体中,与对应的野生型病毒相比,可能不存在有助于复制或为复制所必需的一个或多个蛋白质编码区的至少一部分。在一些实施方案中,病毒载体是复制缺陷的。在一些实施方案中,载体能够转导靶非分裂宿主细胞和/或将其基因组整合到宿主基因组中。

[0408] 在一些实施方案中,野生型逆转录病毒基因组的结构通常包含5'长末端重复序列(LTR)和3'LTR,在它们之间或在它们内定位有使得基因组能够被包装的包装信号、引物结合位点、使得能够整合到宿主细胞基因组中的整合位点、以及编码促进病毒颗粒组装的包装组分的gag、pol和env基因。更复杂的逆转录病毒具有附加的特征,诸如HIV中的rev和RRE序列,其使得整合的原病毒的RNA转录物能够从感染靶细胞的细胞核有效输出到细胞质。在原病毒中,病毒基因在两个末端侧接称为长末端重复序列(LTR)的区域。在一些实施方案中,LTR参与原病毒整合和转录。在一些实施方案中,LTR充当增强子-启动子序列并且可以控制病毒基因的表达。在一些实施方案中,逆转录病毒RNA的衣壳化通过位于病毒基因组5'末端的psi序列发生。

[0409] 在一些实施方案中,LTR是可以分成称为U3、R和U5的三个元件的相似序列。U3来源于对RNA的3'末端独特的序列。R来源于在RNA两个末端重复的序列,并且U5来源于对RNA的5'末端独特的序列。这三种元件的尺寸在不同的逆转录病毒中可以有相当大的变化。

[0410] 在一些实施方案中,对于病毒基因组,转录起始位点通常在一个LTR中的U3与R之间的边界处,而聚(A)添加(终止)位点在另一个LTR中的R与U5之间的边界处。U3含有原病毒的大部分转录控制元件,其包括对细胞和在一些情况下病毒转录激活因子蛋白作出应答的启动子和多个增强子序列。在一些实施方案中,逆转录病毒包含编码参与调节基因表达的蛋白质的以下基因中的任何一种或多种:tat、rev、tax和rex。

[0411] 在一些实施方案中,结构基因包括gag、pol和env,gag编码病毒的内部结构蛋白。在一些实施方案中,Gag蛋白被蛋白水解加工为成熟蛋白MA(基质)、CA(衣壳)和NC(核衣壳)。在一些实施方案中,pol基因编码逆转录酶(RT),其含有介导基因组复制的DNA聚合酶、相关的RNA酶H和整合酶(IN)。在一些实施方案中,env基因编码病毒体的表面(SU)糖蛋白和跨膜(TM)蛋白,它们形成与细胞受体蛋白特异性地相互作用的复合物。在一些实施方案中,所述相互作用通过病毒膜与细胞膜的融合促进感染。

[0412] 在一些实施方案中,复制缺陷逆转录病毒载体基因组gag、pol和env可以不存在或没有功能。在一些实施方案中,RNA两个末端的R区通常是重复序列。在一些实施方案中,U5和U3分别代表RNA基因组的5'和3'末端的独特序列。

[0413] 在一些实施方案中,逆转录病毒还可以含有编码除gag、pol和env之外的蛋白质的附加基因。附加基因的实例包括(在HIV中)vif、vpr、vpx、vpu、tat、rev和nef中的一种或多种。EIAV具有(尤其)附加基因S2。在一些实施方案中,由附加基因编码的蛋白质具有多种功能,其中一些可以是由细胞蛋白质提供的功能的复制。例如,在EIAV中,tat充当病毒LTR的转录激活因子(Derse和Newbold 1993Virology 194:530-6;Maury等人1994Virology 200:632-42)。其与称为TAR的稳定的茎环RNA二级结构结合。Rev通过rev反应元件(RRE)调节和协调病毒基因的表达(Martarano等人1994J.Virol.68:3102-11)。

[0414] 在一些实施方案中,除蛋白酶、逆转录酶和整合酶之外,非灵长类慢病毒还含有编码脱氧尿苷三磷酸酶(dUTP酶)的第四pol基因产物。在一些实施方案中,这在这些慢病毒感染某些不分裂或缓慢分裂的细胞类型的能力中起作用。

[0415] 在实施方案中,重组慢病毒载体(RLV)是具有足够的逆转录病毒遗传信息以允许在包装组分的存在下将RNA基因组包装成能够感染靶细胞的病毒颗粒的载体。在一些实施方案中,靶细胞的感染可包括逆转录和整合到靶细胞基因组中。在一些实施方案中,RLV通

常携带将由载体递送至靶细胞的非病毒编码序列。在一些实施方案中,RLV不能独立复制以在靶细胞内产生感染性逆转录病毒颗粒。在一些实施方案中,RLV缺乏功能性gag-pol和/或env基因和/或参与复制的其他基因。在一些实施方案中,所述载体可以被配置为分裂内含子载体,例如,如PCT专利申请WO 99/15683中所述,所述专利申请通过引用整体并入本文。

[0416] 在一些实施方案中,慢病毒载体包含最小的病毒基因组,例如,病毒载体已被操纵以除去非必需元件并保留必需元件,从而提供感染、转导和递送感兴趣的核苷酸序列至靶宿主细胞所需的功能,例如,如WO 98/17815中所述,所述专利通过引用整体并入本文。

[0417] 在一些实施方案中,最小慢病毒基因组可包含例如(5')R-U5-一个或多个第一核苷酸序列-U3-R(3')。在一些实施方案中,用于在源细胞内产生慢病毒基因组的质粒载体还可包括可操作地连接到慢病毒基因组以指导基因组在源细胞中转录的转录调节控制序列。在一些实施方案中,所述调节序列可包含与转录的逆转录病毒序列相关的天然序列,例如5'U3区,或者它们可包含异源启动子,诸如另一种病毒启动子,例如CMV启动子。在一些实施方案中,慢病毒基因组包含附加序列以促进有效的病毒产生。在一些实施方案中,在HIV的情况下,可包括rev和RRE序列。在一些实施方案中,可替代地或组合地,可以使用密码子优化,例如,编码外源药剂的基因可以是密码子优化的,例如,如WO 01/79518中所述,所述专利通过引用整体并入本文。在一些实施方案中,也可以使用执行与rev/RRE系统相似或相同功能的替代序列。在一些实施方案中,在Mason Pfizer猴病毒中发现了rev/RRE系统的功能类似物。在一些实施方案中,这被称为CTE并且包含在基因组中的据信与感染细胞中的因子相互作用的RRE型序列。细胞因子可以被认为是rev类似物。在一些实施方案中,CTE可以用作rev/RRE系统的替代物。在一些实施方案中,HTLV-I的Rex蛋白质可以在功能上替代HIV-I的Rev蛋白质。Rev和Rex具有与IRE-BP相似的效应。

[0418] 在一些实施方案中,逆转录病毒核酸(例如慢病毒核酸,例如灵长类或非灵长类慢病毒核酸)(1)包含缺失的gag基因,其中gag中的缺失除去了gag编码序列的约核苷酸350或354下游的一个或多个核苷酸;(2)在所述逆转录病毒核酸中不存在一个或多个辅助基因;(3)缺乏tat基因但包含位于5'LTR的末端与gag的ATG之间的前导序列;和(4)(1)、(2)和(3)的组合。在一个实施方案中,慢病毒载体包含所有特征(1)和(2)和(3)。在WO 99/32646中更详细地描述了这种策略,所述专利通过引用整体并入本文。

[0419] 在一些实施方案中,灵长类慢病毒最小系统不需要HIV/SIV附加基因vif、vpr、vpx、vpu、tat、rev和nef用于载体产生或用于转导分裂和非分裂细胞。在一些实施方案中,EIAV最小载体系统不需要S2用于载体产生或用于转导分裂和非分裂细胞。

[0420] 在一些实施方案中,附加基因的缺失可以允许产生没有与慢病毒(例如HIV)感染中的疾病相关的基因的载体。在一些实施方案中,tat与疾病相关。在一些实施方案中,附加基因的缺失允许载体包装更多的异源DNA。在一些实施方案中,可以省略功能未知的基因,诸如S2,从而降低引起不希望的效应的风险。最小慢病毒载体的实例公开于WO 99/32646和WO 98/17815中。

[0421] 在一些实施方案中,逆转录病毒核酸至少缺乏tat和S2(如果其是EIAV载体系统),也可能缺乏vif、vpr、vpx、vpu和nef。在一些实施方案中,逆转录病毒核酸还缺乏rev、RRE或两者。

[0422] 在一些实施方案中,逆转录病毒核酸包含vpx。Vpx多肽结合并诱导SAMHD1限制因

子的降解,所述SAMHD1限制因子降解细胞质中的游离dNTP。在一些实施方案中,游离dNTP在细胞质中的浓度随着Vpx降解SAMHD1而增加并且逆转录活性增加,因此促进逆转录病毒基因组的逆转录和整合到靶细胞基因组中。

[0423] 在一些实施方案中,不同的细胞在它们对特定密码子的使用方面不同。在一些实施方案中,此密码子偏倚对应于细胞类型中特定tRNA的相对丰度的偏倚。在一些实施方案中,通过改变序列中的密码子以使它们适合与对应tRNA的相对丰度相匹配,其可能增加表达。在一些实施方案中,有可能通过故意选择已知对应tRNA在特定细胞类型中罕见的密码子来降低表达。在一些实施方案中,附加程度的翻译控制是可用的。密码子优化的另外描述见于例如WO 99/41397中,其通过引用整体并入本文。

[0424] 在一些实施方案中,病毒(包括HIV和其他慢病毒)使用大量罕见密码子,并且通过将这些改变为对应于常用的哺乳动物密码子,可以实现包装组分在哺乳动物生产者细胞中的增加的表达。

[0425] 在一些实施方案中,密码子优化具有许多其他优点。在一些实施方案中,由于它们序列的改变,编码包装组分的核苷酸序列可以从它们中减少或清除RNA不稳定性序列(INS)。同时,保留包装组分的氨基酸序列编码序列,使得由所述序列编码的病毒组分保留相同或至少足够相似,使得包装组分的功能不受损害。在一些实施方案中,密码子优化还克服了输出的Rev/RRE要求,使得优化的序列为Rev非依赖性的。在一些实施方案中,密码子优化还减少了载体系统内不同构建体之间的同源重组(例如在gag-pol和env开放阅读框中重叠的区域之间)。在一些实施方案中,密码子优化导致病毒滴度增加和/或安全性提高。

[0426] 在一些实施方案中,只有与INS有关的密码子被密码子优化。在其他实施方案中,除了包含gag-pol移码位点的序列外,所述序列全部是密码子优化的。

[0427] gag-pol基因包含编码gag-pol蛋白的两个重叠阅读框。两种蛋白质的表达取决于翻译期间的移码。由于翻译期间核糖体“滑动”而发生这种移码。这种滑动被认为至少部分是由核糖体-停滞RNA二级结构引起的。这种二级结构存在于gag-pol基因中的移码位点的下游。对于HIV,重叠区从gag起点(其中核苷酸1是gag ATG的A)下游的核苷酸1222延伸到gag的末尾(nt 1503)。因此,跨越移码位点的281bp片段和两个阅读框的重叠区优选不是密码子优化的。在一些实施方案中,保留该片段将能够更有效地表达gag-pol蛋白。对于EIAV,重叠的起点位于nt 1262(其中核苷酸1是gag ATG的A)。重叠的末尾位于nt 1461。为了确保移码位点和gag-pol重叠被保留,可以保留从nt 1156到1465的野生型序列。

[0428] 在一些实施方案中,例如为了适应方便的限制性位点,可以从最佳密码子使用中进行衍生,并且可以将保守氨基酸改变引入gag-pol蛋白中。

[0429] 在一些实施方案中,密码子优化是基于哺乳动物系统中具有较差密码子使用的密码子。可以改变第三个碱基,有时也可以改变第二和第三个碱基。

[0430] 在一些实施方案中,由于遗传密码的简并性,可以理解,技术人员可以获得许多gag-pol序列。此外,描述了许多逆转录病毒变体,它们可以用作产生密码子优化的gag-pol序列的起点。慢病毒基因组可以是相当可变的。例如,存在许多仍有功能的HIV-I的准种。这也是EIAV的情况。这些变体可以用于增强转导过程的特定部分。HIV-I变体的实例可以在os Alamos国家实验室维护的HIV数据库中找到。EIAV克隆的细节可以在国家卫生研究院(National Institutes of Health)维护的NCBI数据库中找到。

[0431] 在一些实施方案中,密码子优化的gag-pol序列的策略可以用于任何逆转录病毒,例如EIAV、FIV、BIV、CAEV、VMR、SIV、HIV-I和HIV-2。此外,该方法可以用于增加来自HTLV-I、HTLV-2、HFV、HSRV和人内源性逆转录病毒(HERV)、MLV和其他逆转录病毒的基因的表达。

[0432] 在实施方案中,所述逆转录病毒载体包含包装信号,所述包装信号在仍保留env序列的载体中包含gag的255至360个核苷酸,或在剪接供体突变gag和env缺失的特定组合中包含gag的约40个核苷酸。在一些实施方案中,逆转录病毒载体包括包含一个或多个缺失的gag序列,例如所述gag序列包含可从N末端衍生的约360个核苷酸。

[0433] 在一些实施方案中,逆转录病毒载体、辅助细胞、辅助病毒或辅助质粒可包含逆转录病毒结构蛋白和辅助蛋白,例如gag、pol、env、tat、rev、vif、vpr、vpu、vpx或nef蛋白或其他逆转录病毒蛋白。在一些实施方案中,逆转录病毒蛋白来源于相同的逆转录病毒。在一些实施方案中,逆转录病毒蛋白来源于多于一种逆转录病毒,例如2种、3种、4种或更多种逆转录病毒。

[0434] 在一些实施方案中,gag和pol编码序列通常在天然慢病毒中被组织为Gag-Pol前体。gag序列编码55-kD Gag前体蛋白,也称为p55。p55在成熟过程期间被病毒编码的蛋白酶(pol基因的产物)切割成四种较小的蛋白质,命名为MA(基质[p17])、CA(衣壳[p24])、NC(核衣壳[p9])和p6。pol前体蛋白被病毒编码的蛋白酶从Gag上切割开,并进一步消化以分离蛋白酶(p10)、RT(p50)、RNA酶H(p15)和整合酶(p31)活性。

[0435] 在一些实施方案中,慢病毒载体是整合缺陷的。在一些实施方案中,pol是整合酶缺陷的,诸如由于整合酶基因中的突变而编码。例如,pol编码序列可以含有整合酶中的失活突变,诸如通过突变参与催化活性的一个或多个氨基酸,即突变天冬氨酸64、天冬氨酸116和/或谷氨酸152中的一个或多个。在一些实施方案中,整合酶突变是D64V突变。在一些实施方案中,整合酶中的突变允许将病毒RNA包装成慢病毒。在一些实施方案中,整合酶中的突变允许将病毒蛋白包装成慢病毒。在一些实施方案中,整合酶中的突变降低了插入诱变的可能性。在一些实施方案中,整合酶中的突变降低了产生具有复制能力的重组体(RCR)的可能性(Wanisch等人2009.Mol Ther.1798):1316-1332)。在一些实施方案中,天然Gag-Pol序列可以用于辅助载体(例如辅助质粒或辅助病毒),或者可以进行修饰。这些修饰包括嵌合Gag-Pol,其中Gag和Pol序列获自不同的病毒(例如不同的物种、亚种、病毒株、分化体等),和/或其中所述序列已被修饰以改善转录和/或翻译和/或减少重组。

[0436] 在一些实施方案中,逆转录病毒核酸包括编码gag蛋白的150-250个(例如168个)核苷酸部分的多核苷酸,所述多核苷酸(i)包括相对于野生型INS1减少RNA核输出限制的突变INS1抑制序列,(ii)含有导致移码和提前终止的两个核苷酸插入,和/或(iii)不包括gag的INS2、INS3和INS4抑制序列。

[0437] 在一些实施方案中,本文所述的载体是包含逆转录病毒(例如慢病毒)序列和非慢病毒病毒序列的杂合载体。在一些实施方案中,杂合载体包含用于逆转录、复制、整合和/或包装的逆转录病毒(例如慢病毒)序列。

[0438] 在一些实施方案中,大部分或全部病毒载体骨架序列源自慢病毒,例如HIV-1。然而,应当理解,可以使用或组合许多不同来源的逆转录病毒和/或慢病毒序列,并且可以容纳某些慢病毒序列中的许多取代和改变而不损害转移载体执行本文所述的功能的能力。Naldini等人,(1996a、1996b和1998);Zufferey等人,(1997);Dull等人,1998,美国专利号

6,013,516和5,994,136中描述了多种慢病毒载体,其中许多可以适于产生逆转录病毒核酸。

[0439] 在一些实施方案中,在原病毒的每个末端,通常发现长末端重复序列(LTR)。LTR通常包含位于逆转录病毒核酸末端的结构域,在其天然序列背景下其是正向重复序列并且含有U3、R和U5区。LTR通常促进逆转录病毒基因的表达(例如,基因转录物的促进、起始和聚腺苷酸化)和病毒复制。LTR可包含许多调节信号,包括转录控制元件、聚腺苷酸化信号和用于病毒基因组复制和整合的序列。病毒LTR通常分为三个区域,称为U3、R和U5。U3区通常含有增强子和启动子元件。U5区通常是在引物结合位点与R区之间的序列,并且可以含有聚腺苷酸化序列。R(重复)区可以侧接U3和U5区。LTR通常由U3、R和U5区构成并且可以出现在病毒基因组的5'和3'末端。在一些实施方案中,与5'LTR相邻的是用于逆转录基因组(tRNA引物结合位点)和用于将病毒RNA有效包装成颗粒(Psi位点)的序列。

[0440] 在一些实施方案中,包装信号可包含位于逆转录病毒基因组内介导病毒RNA插入到病毒衣壳或颗粒中的序列,参见例如Clever等人,1995.J.of Virology,第69卷第4期;第2101-2109页。几种逆转录病毒载体使用最小包装信号(psi[Ψ]序列)用于病毒基因组的衣壳化。

[0441] 在各种实施方案中,逆转录病毒核酸包含经修饰的5'LTR和/或3'LTR。LTR中的任一个或两个可包含一个或多个修饰,包括但不限于一个或多个缺失、插入或取代。通常对3'LTR进行修饰以通过使病毒成为复制缺陷的(例如不能完全、有效地复制从而不产生感染性病毒体的病毒(例如复制缺陷慢病毒子代))来改善慢病毒或逆转录病毒系统的安全性。

[0442] 在一些实施方案中,载体是自我失活(SIN)载体,例如复制缺陷载体,例如逆转录病毒或慢病毒载体,其中右(3')LTR增强子-启动子区(称为U3区)已被修饰(例如通过缺失或取代)以防止病毒转录超过第一轮病毒复制。这是因为在病毒复制期间右(3')LTR U3区可以用作左(5')LTR U3区的模板,并且因此不存在U3增强子-启动子会抑制病毒复制。在实施方案中,3'LTR被修饰使得U5区被除去、改变或例如被外源聚(A)序列替换。3'LTR、5'LTR或3'和5'LTR两者可以是经修饰的LTR。

[0443] 在一些实施方案中,5'LTR的U3区被异源启动子替换以在病毒颗粒产生期间驱动病毒基因组的转录。可以使用的异源启动子的实例包括例如病毒猿猴病毒40(SV40)(例如早期或晚期)、巨细胞病毒(CMV)(例如立即早期)、莫洛尼鼠白血病病毒(MoMLV)、劳斯肉瘤病毒(RSV)和单纯疱疹病毒(HSV)(胸苷激酶)启动子。在一些实施方案中,启动子能够以Tat非依赖性方式驱动高水平的转录。在某些实施方案中,异源启动子在控制病毒基因组转录的方式方面具有附加的优点。例如,异源启动子可以是诱导型的,使得只有当存在诱导因子时才发生全部或部分病毒基因组的转录。诱导因子包括但不限于一种或多种化学化合物或培养宿主细胞的生理条件诸如温度或pH。

[0444] 在一些实施方案中,病毒载体包含TAR(反式激活反应)元件,例如位于慢病毒(例如HIV)LTR的R区中。该元件与慢病毒反式激活因子(tat)遗传元件相互作用以增强病毒复制。然而,例如在其中5'LTR的U3区被异源启动子替换的实施方案中,该元件不是必需的。

[0445] 在一些实施方案中,R区(例如逆转录病毒LTR内起始于封端基团的起点(即转录的起点)并紧接在多聚A片段的起点之前终止的区域)可以侧接U3和U5区。R区在将新生DNA从基因组的一端转移到另一端的逆转录期间起作用。

[0446] 在一些实施方案中,逆转录病毒核酸还可包含FLAP元件,例如其序列包括逆转录病毒(例如HIV-1或HIV-2)的中央多嘌呤区和中央终止序列(cPPT和CTS)的核酸。合适的FLAP元件在美国专利号6,682,907中和在Zennou等人,2000,Cell,101:173中有所描述,所述文献通过引用整体并入本文。在HIV-1逆转录期间,正链DNA在中央多嘌呤区(cPPT)的中央起始和在中央终止序列(CTS)的中央终止可以导致三链DNA结构的形成:HIV-1中央DNA瓣。在一些实施方案中,逆转录病毒或慢病毒载体骨架包含在编码外源药剂的基因上游或下游的一个或多个FLAP元件。例如,在一些实施方案中,转移质粒包括FLAP元件,例如源自或分离自HIV-1的FLAP元件。

[0447] 在实施方案中,逆转录病毒或慢病毒核酸包含一个或多个输出元件,例如调节将RNA转录物从细胞核转运到细胞质的顺式作用转录后调控元件。RNA输出元件的实例包括但不限于人免疫缺陷病毒(HIV)rev反应元件(RRE)(参见例如Cullen等人,1991.J.Virol.65:1053;以及Cullen等人,1991.Cell 58:423)和乙型肝炎病毒转录后调控元件(HPRE),所述文献通过引用整体并入本文。通常,将RNA输出元件置于基因的3'UTR内,并且可以作为一个或多个拷贝插入。

[0448] 在一些实施方案中,通过将转录后调控元件、聚腺苷酸化位点和转录终止信号中的一个或多个(例如,全部)掺入载体中来增加异源序列的表达。多种转录后调控元件可以增加异源核酸表达于蛋白质,例如土拨鼠肝炎病毒转录后调控元件(WPRE;Zufferey等人,1999,J.Virol.,73:2886);存在于乙型肝炎病毒(HPRE)中的转录后调控元件(Huang等人,Mol.Cell.Biol.,5:3864);等等(Liu等人,1995,Genes Dev.,9:1766),所述文献中的每一篇通过引用整体并入本文。在一些实施方案中,本文所述的逆转录病毒核酸包含转录后调控元件诸如WPRE或HPRE。

[0449] 在一些实施方案中,本文所述的逆转录病毒核酸缺乏或不包含转录后调控元件诸如WPRE或HPRE。

[0450] 在一些实施方案中,可包括指导异源核酸转录物终止和聚腺苷酸化的元件,例如以增加外源药剂的表达。可以在聚腺苷酸化信号的下游发现转录终止信号。在一些实施方案中,载体包含编码外源药剂的多核苷酸3'的聚腺苷酸化序列。聚A位点可包含指导RNA聚合酶II终止和聚腺苷酸化新生RNA转录物的DNA序列。聚腺苷酸化序列可以通过向编码序列的3'末端添加聚A尾来促进mRNA稳定性,并且因此有助于增加翻译效率。可以用于逆转录病毒核酸中的聚A信号的例示性实例包括AATAAA、ATTAAA、AGTAAA、牛生长激素聚A序列(BGHpA)、兔β珠蛋白聚A序列(rβgpA)或另一种合适的异源或内源聚A序列。

[0451] 在一些实施方案中,逆转录病毒或慢病毒载体还包含一个或多个绝缘子元件,例如本文所述的绝缘子元件。

[0452] 在各种实施方案中,载体包含与编码外源药剂的多核苷酸可操作地连接的启动子。载体可以具有一个或多个LTR,其中任一个LTR包含一个或多个修饰,诸如一个或多个核苷酸取代、添加或缺失。载体还可包含一个或多个增加转导效率的辅助元件(例如cPPT/FLAP)、病毒包装(例如Psi(Ψ)包装信号,RRE)和/或增加外源基因表达的其他元件(例如聚(A)序列),并且可以任选地包含WPRE或HPRE。

[0453] 在一些实施方案中,慢病毒核酸例如从5'至3'包含以下中的一种或多种(例如全部):启动子(例如CMV)、R序列(例如包含TAR)、U5序列(例如用于整合)、PBS序列(例如用于

逆转录)、DIS序列(例如用于基因组二聚化)、psi包装信号、部分gag序列、RRE序列(例如用于核输出)、cPPT序列(例如用于核输入)、驱动外源药剂表达的启动子、编码外源药剂的基因、WPRE序列(例如,用于有效的转基因表达)、PPT序列(例如,用于逆转录)、R序列(例如,用于聚腺苷酸化和终止)和U5信号(例如,用于整合)。

[0454] b.包装载体和生产者细胞

[0455] 大规模病毒颗粒产生通常可用于实现所需的病毒滴度。可以通过将转移载体转染到包装细胞系中来产生病毒颗粒,所述包装细胞系包含病毒结构和/或辅助基因,例如gag、pol、env、tat、rev、vif、vpr、vpu、vpx或nef基因或其他逆转录病毒基因。

[0456] 在一些实施方案中,包装载体是缺乏包装信号并包含编码一个、两个、三个、四个或更多个病毒结构和/或辅助基因的多核苷酸的表达载体或病毒载体。通常,包装载体包含在生产者细胞中,并且通过转染、转导或感染引入细胞中。可以通过转染、转导或感染将逆转录病毒(例如慢病毒)转移载体引入生产者细胞系中,以产生源细胞或细胞系。通过标准方法将包装载体引入人细胞或细胞系,所述标准方法包括例如磷酸钙转染、脂质转染或电穿孔。在一些实施方案中,将包装载体与显性选择标记诸如新霉素、潮霉素、嘌呤霉素、杀稻瘟菌素、博来霉素(zeocin)、胸苷激酶、DHFR、Gln合成酶或ADA一起引入细胞中,随后在适当药物的存在下进行选择并分离克隆。选择标记基因可以例如通过IRES或自切割病毒肽与由包装载体编码的基因物理连接。

[0457] 在一些实施方案中,生产者细胞系包括不含有包装信号但稳定或瞬时表达可以包装病毒颗粒的病毒结构蛋白和复制酶(例如gag、pol和env)的细胞系。可以使用任何合适的细胞系,例如哺乳动物细胞,例如人细胞。可以使用的合适细胞系包括例如CHO细胞、BHK细胞、MDCK细胞、C3H 10T1/2细胞、FLY细胞、Psi-2细胞、BOSC 23细胞、PA317细胞、WEHI细胞、COS细胞、BSC 1细胞、BSC 40细胞、BMT 10细胞、VERO细胞、W138细胞、MRC5细胞、A549细胞、HT1080细胞、293细胞、293T细胞、B-50细胞、3T3细胞、NIH3T3细胞、HepG2细胞、Saos-2细胞、Huh7细胞、HeLa细胞、W163细胞、211细胞和211A细胞。在实施方案中,包装细胞是293细胞、293T细胞或A549细胞。

[0458] 在一些实施方案中,源细胞系包括能够产生重组逆转录病毒颗粒的细胞系,所述重组逆转录病毒颗粒包含生产者细胞系和含有包装信号的转移载体构建体。以下文献例示了制备病毒储备液的方法:例如Y.Soneoka等人(1995)Nucl.Acids Res.23:628-633,以及N.R.Landau等人(1992)J.Virol.66:5110-5113,其通过引用并入本文。可以从生产者细胞收集感染性病毒颗粒,例如通过细胞裂解或收集细胞培养物的上清液。任选地,可以富集或纯化所收集的病毒颗粒。

[0459] 在一些实施方案中,源细胞包含编码可以包装病毒颗粒的病毒结构蛋白和复制酶(例如gag、pol和env)的一种或多种质粒。在一些实施方案中,编码gag、pol和env前体中的至少两种的序列在同一质粒上。在一些实施方案中,编码gag、pol和env前体的序列在不同的质粒上。在一些实施方案中,编码gag、pol和env前体的序列具有相同的表达信号,例如启动子。在一些实施方案中,编码gag、pol和env前体的序列具有不同的表达信号,例如不同的启动子。在一些实施方案中,gag、pol和env前体的表达是诱导型的。在一些实施方案中,在相同时间或在不同时间转染编码病毒结构蛋白和复制酶的质粒。在一些实施方案中,在与包装载体相同的时间或不同的时间转染编码病毒结构蛋白和复制酶的质粒。

[0460] 在一些实施方案中，源细胞系包含一个或多个稳定整合的病毒结构基因。在一些实施方案中，稳定整合的病毒结构基因的表达是诱导型的。

[0461] 在一些实施方案中，病毒结构基因的表达在转录水平上受到调节。在一些实施方案中，病毒结构基因的表达在翻译水平上受到调节。在一些实施方案中，病毒结构基因的表达在翻译后水平受到调节。

[0462] 在一些实施方案中，病毒结构基因的表达受四环素(Tet)依赖性系统调节，其中Tet调节的转录阻遏物(Tet-R)结合在启动子中包含的DNA序列并且通过空间位阻抑制转录(Yao等人，1998；Jones等人，2005)。在添加多西环素(dox)后，Tet-R被释放，从而允许转录。多种其他合适的转录调节启动子、转录因子和小分子诱导物适合用于调节病毒结构基因的转录。

[0463] 在一些实施方案中，将在Tet调节的启动子的控制下并与抗生素抗性盒偶联的第三代慢病毒组分、人免疫缺陷病毒1型(HIV)、Rev、Gag/Pol和包膜分别整合到源细胞基因组中。在一些实施方案中，源细胞仅在基因组中整合了一个拷贝的Rev、Gag/Pol和包膜蛋白中的每一种。

[0464] 在一些实施方案中，编码外源药剂的核酸(例如，编码外源药剂的逆转录病毒核酸)也被整合到源细胞基因组中。

[0465] 在一些实施方案中，本文所述的逆转录病毒核酸不能进行逆转录。在实施方案中，这样的核酸能够瞬时表达外源药剂。逆转录病毒或VLP可能包含无效的逆转录酶蛋白，或者可能不包含逆转录酶蛋白。在实施方案中，逆转录病毒核酸包含无效的引物结合位点(PBS)和/或att位点。在实施方案中，一种或多种病毒辅助基因(包括rev、tat、vif、nef、vpr、vpu、vpx和S2或其功能等效物)在逆转录病毒核酸中是无效的或缺失的。在实施方案中，选自S2、rev和tat的一个或多个辅助基因在逆转录病毒核酸中是无效的或缺失的。

[0466] 2.细胞衍生的颗粒

[0467] 本文提供了包含天然来源的膜的靶向脂质颗粒。在一些实施方案中，天然来源的膜包括由细胞或组织制备的膜囊泡。在一些实施方案中，靶向脂质颗粒包含可从细胞获得的囊泡。在一些实施方案中，靶向脂质颗粒包含微泡、外来体、膜包围体、凋亡小体(来自凋亡细胞)、颗粒(其可以来源于例如血小板)、核外粒(可来源于例如血清中的嗜中性粒细胞和单核细胞)、前列腺小体(prostatosome)(可获自前列腺癌细胞)或心小体(可来源于心脏细胞)。

[0468] 在一些实施方案中，源细胞是内皮细胞、成纤维细胞、血细胞(例如，巨噬细胞、嗜中性粒细胞、粒细胞、白细胞)、干细胞(例如，间充质干细胞、脐带干细胞、骨髓干细胞、造血干细胞、诱导多能干细胞(例如，来源于受试者细胞的诱导多能干细胞)、胚胎干细胞(例如，来自胚胎卵黄囊、胎盘、脐带、胎儿皮肤、青少年皮肤、血液、骨髓、脂肪组织、红细胞生成组织、造血组织的干细胞))、成肌细胞、实质细胞(例如，肝细胞)、肺泡细胞、神经元(例如，视网膜神经元细胞)、前体细胞(例如，视网膜前体细胞、成髓细胞、骨髓前体细胞、胸腺细胞、性母细胞、成巨核细胞、幼巨核细胞、成黑素细胞、成淋巴细胞、骨髓前体细胞、正成红细胞或成血管细胞)、祖细胞(例如，心脏祖细胞、卫星细胞、放射状胶质细胞、骨髓基质细胞、胰腺祖细胞、内皮祖细胞、母细胞)或无限增殖化细胞(例如，HeEa、HEK293、HFF-1、MRC-5、WI-38、IMR 90、IMR 91、PER.C6、HT-1080或BJ细胞)。在一些实施方案中，源细胞不是293细胞、

HEK细胞、人内皮细胞或人上皮细胞、单核细胞、巨噬细胞、树突细胞或干细胞。

[0469] 在一些实施方案中，靶向脂质颗粒具有<1、1-1.1、1.05-1.15、1.1-1.2、1.15-1.25、1.2-1.3、1.25-1.35或>1.35g/ml的密度。在一些实施方案中，靶向脂质颗粒组合物包含按蛋白质质量计少于0.01％、0.05％、0.1％、0.5％、1％、1.5％、2％、2.5％、3％、4％、5％或10％的源细胞或少于0.01％、0.05％、0.1％、0.5％、1％、1.5％、2％、2.5％、3％、4％、5％或10％的具有功能细胞核的细胞。

[0470] 在实施方案中，靶向脂质颗粒的大小或靶向脂质颗粒群体的平均大小为源细胞大小的小于约0.01％、0.05％、0.1％、0.5％、1％、2％、3％、4％、5％、10％、20％、30％、40％、50％、60％、70％、80％、90％。

[0471] 在一些实施方案中，靶向脂质颗粒包含细胞外囊泡，例如包含膜的细胞衍生囊泡，所述膜包围内部空间并且具有比其来源的细胞更小的直径。在实施方案中，细胞外囊泡具有20nm至1000nm的直径。在实施方案中，靶向脂质颗粒包含凋亡小体、细胞片段、通过直接或间接操作来源于细胞的囊泡、囊泡化细胞器和由活细胞产生的囊泡(例如通过直接质膜出芽或晚期内体与质膜融合)。在实施方案中，细胞外囊泡源自活的或死的生物体、移植的组织或器官或培养的细胞。

[0472] 在实施方案中，靶向脂质颗粒包含纳米囊泡，例如细胞衍生的小(例如直径在20-250nm之间或直径在30-150nm之间)的囊泡，其包含包围内部空间的膜并且其通过直接或间接操作从所述细胞产生。在一些情形中，纳米囊泡的产生可以导致源细胞的破坏。纳米泡囊可包含脂质或脂肪酸和多肽。

[0473] 在实施方案中，靶向脂质颗粒包含外来体。在实施方案中，外来体是细胞衍生的小(例如，直径在20-300nm之间或直径在40-200nm之间)的囊泡，其包含包围内部空间的膜并且其通过直接质膜出芽或通过晚期内体与质膜的融合从所述细胞产生。在实施方案中，外来体的产生不导致源细胞的破坏。在实施方案中，外来体包含脂质或脂肪酸和多肽。示例性外来体和其他膜包围体在WO/2017/161010、WO/2016/077639、US20160168572、US20150290343和US20070298118中也有所描述，所述专利中的每一篇通过引用整体并入本文。

[0474] 在一些实施方案中，靶向脂质颗粒源自具有遗传修饰的源细胞，所述遗传修饰导致免疫调节剂的表达增加。在一些实施方案中，免疫抑制剂在细胞的外表面上。在一些实施方案中，将免疫抑制剂掺入靶向脂质颗粒的外表面中。在一些实施方案中，靶向脂质颗粒包含通过共价或非共价键附着于固体颗粒表面的免疫调节剂。

[0475] c.a.细胞衍生颗粒的产生

[0476] 在一些实施方案中，靶向脂质颗粒通过诱导外来体、微泡、膜囊泡、细胞外膜囊泡、质膜囊泡、巨质膜囊泡、凋亡小体、有丝分裂小体、核红细胞、溶酶体或其他膜包围囊泡的出芽而产生。

[0477] 在一些实施方案中，通过诱导细胞去核来产生靶向脂质颗粒。去核可以使用如下测定来进行：所述测定诸如遗传、化学(例如，使用放线菌素D，参见Bayona-Bafaluy等人，“A chemical enucleation method for the transfer of mitochondrial DNA to ρ° cells” Nucleic Acids Res.2003年8月15日；31(16)：e98)、机械方法(例如，挤压或抽吸，参见Lee等人，“A comparative study on the efficiency of two enucleation methods in pig

somatic cell nuclear transfer:effects of the squeezing and the aspiration methods."Anim Biotechnol.2008;19(2):71-9)或它们的组合。

[0478] 在一些实施方案中,通过诱导细胞破碎来产生靶向脂质颗粒。在一些实施方案中,细胞破碎可以使用以下方法来进行,所述方法包括但不限于:化学方法、机械方法(例如离心(例如超速离心或密度离心)、冷冻-解冻或超声处理)或它们的组合。

[0479] 在一些实施方案中,靶向脂质颗粒是微泡。在一些实施方案中,微泡具有约100nm至约2000nm的直径。在一些实施方案中,靶向脂质颗粒包含细胞空壳。在一些实施方案中,囊泡是质膜囊泡,例如巨质膜囊泡。

[0480] 在一些实施方案中,用于制备靶向脂质颗粒的源细胞在制备靶向脂质颗粒之后将不可用于测试。

[0481] 在一些实施方案中,通过与参考细胞比较来描述靶向脂质颗粒的特征。在实施方案中,参考细胞是源细胞。在实施方案中,参考细胞是HeLa、HEK293、HFF-1、MRC-5、WI-38、IMR 90、IMR 91、PER.C6、HT-1080或BJ细胞。在一些实施方案中,通过与参考细胞群(例如源细胞群或HeLa、HEK293、HFF-1、MRC-5、WI-38、IMR 90、IMR 91、PER.C6、HT-1080或BJ细胞群体)比较来描述靶向脂质颗粒群体的特征。

[0482] III.药物组合物

[0483] 在一些方面,本公开还提供了药物组合物,所述药物组合物包含本文所述的靶向脂质颗粒组合物和药学上可接受的载体。所述药物组合物可包含任何所描述的靶向脂质颗粒。

[0484] 在一些实施方案中,靶向脂质颗粒满足药物或药品生产质量管理规范(GMP)标准。在一些实施方案中,根据药品生产质量管理规范(GMP)制备靶向脂质颗粒。在一些实施方案中,靶向脂质颗粒具有低于预定参考值的病原体水平,例如基本上不含病原体。在一些实施方案中,靶向脂质颗粒具有低于预定参考值的污染物水平,例如基本上不含污染物。在一些实施方案中,靶向脂质颗粒具有低免疫原性。

[0485] 在一些实施方案中,本文提供了本发明的药物组合物或其盐用于实践本发明的方法的用途。这样的药物组合物可以由适合施用于受试者的形式的至少一种本发明的化合物或缀合物或其盐组成,或者所述药物组合物可包含至少一种本发明的化合物或缀合物或其盐以及一种或多种药学上可接受的载体、一种或多种附加成分或这些的一些组合。在一些实施方案中,本发明的化合物或缀合物可以以生理学上可接受的盐(诸如与生理学上可接受的阳离子或阴离子组合,如本领域中熟知的)的形式存在于药物组合物中。

[0486] 在一些实施方案中,可以施用可用于实施本发明方法的药物组合物以递送1ng/kg/天至100mg/kg/天的剂量。在另一个实施方案中,可以施用可用于实施本发明的药物组合物以递送1ng/kg/天至500mg/kg/天的剂量。

[0487] 在一些实施方案中,本发明的药物组合物中的活性成分、药学上可接受的载体和任何附加成分的相对量将取决于所治疗的受试者的身份、大小和病症并且进一步取决于组合物有待施用的途径而变化。在一些实施方案中,所述组合物可包含0.1%与100%(w/w)之间的活性成分。

[0488] 在一些实施方案中,可以适当地开发可用于本发明方法的药物组合物用于口服、直肠、阴道、肠胃外、局部、肺部、鼻内、颊、眼部或其他施用途径。在一些实施方案中,可用于

本发明方法中的组合物可以直接施用于哺乳动物的皮肤、阴道或任何其他组织。在一些实施方案中，配制品包括脂质体制剂、含有活性成分的重新密封的红细胞和基于免疫学的配制品。在一些实施方案中，施用途径对于本领域技术人员来说将是显而易见的，并且将取决于许多因素，包括所治疗的疾病的类型和严重程度，所治疗的兽医或人类受试者的类型和年龄等。

[0489] 在一些实施方案中，本文所述的药物组合物的配制品可以通过药理学领域中已知或此后开发的任何方法来制备。在一些实施方案中，制备方法包括以下步骤：使活性成分与载体或一种或多种其他辅助成分缔合，然后如果有必要或需要，使产品成形或包装成所需的单剂量或多剂量单位。

[0490] 在一些实施方案中，“单位剂量”是包含预定量的活性成分的离散量的药物组合物。在一些实施方案中，活性成分的量通常等于将向受试者施用的活性成分的剂量或这种剂量的方便部分(例如这种剂量的二分之一或三分之一)。在一些实施方案中，单位剂型可以用于单个日剂量或多个日剂量之一(例如，每天约1至4次或更多次)。在一些实施方案中，当使用多个日剂量时，单位剂型对于每个剂量可以是相同的或不同的。

[0491] 在一些实施方案中，虽然本文提供的药物组合物的描述主要涉及适合于向人类进行伦理施用的药物组合物，但是本领域技术人员将理解，此类组合物通常适合施用于所有种类的动物。在一些实施方案中，对适合于向人类施用的药物组合物进行修饰以使所述组合物适合施用于各种动物得到很好地理解，并且普通熟练的兽医药理学家可以仅利用普通的实验(如果有的话)来设计和进行这种修饰。在一些实施方案中，预期施用本发明的药物组合物的受试者包括人和其他灵长类动物、哺乳动物(包括商业上相关的哺乳动物，诸如牛、猪、马、绵羊、猫和狗)。

[0492] 在一些任何实施方案中，使用一种或多种药学上可接受的赋形剂或载体配制本发明的组合物。在一个实施方案中，本发明的药物组合物包含治疗有效量的本发明的化合物或缀合物和药学上可接受的载体。在一些实施方案中，有用的药学上可接受的载体包括但不限于甘油、水、盐水、乙醇和其他药学上可接受的盐溶液诸如磷酸盐和有机酸盐。这些和其他药学上可接受的载体的实例在Remington's Pharmaceutical Sciences(1991,Mack Publication Co.,New Jersey)中有所描述。

[0493] 在一些实施方案中，所述载体可以是含有例如水、乙醇、多元醇(例如，甘油、丙二醇和液态聚乙二醇等)、它们合适的混合物和植物油的溶剂或分散介质。在一些实施方案中，可以例如通过使用包衣诸如卵磷脂、在分散液的情况下通过维持所需的粒度以及通过使用表面活性剂来维持适当的流动性。在一些实施方案中，可以通过各种抗细菌剂和抗真菌剂例如对羟基苯甲酸酯、氯丁醇、苯酚、抗坏血酸、硫柳汞等来实现对微生物作用的预防。在一些实施方案中，优选的是在所述组合物中包括等渗剂，例如糖、氯化钠或多元醇诸如甘露醇和山梨糖醇。在一些实施方案中，可以通过在所述组合物中包括延迟吸收的药剂(例如单硬脂酸铝或明胶)来实现可注射组合物的延长吸收。在一个实施方案中，药学上可接受的载体不是单独的DMSO。

[0494] 在一些实施方案中，配制品可以与常规赋形剂混合使用，所述常规赋形剂即适合于口服、阴道、肠胃外、鼻、静脉内、皮下、肠内或本领域已知的任何其他合适的施用模式的药学上可接受的有机或无机载体物质。在一些实施方案中，药物制剂可以被灭菌，并且如果

需要可以与助剂混合，所述助剂例如润滑剂、防腐剂、稳定剂、润湿剂、乳化剂、用于影响渗透压的盐、缓冲剂、着色剂、调味剂和/或芳香物质等。在一些实施方案中，药物制剂还可以在需要时与其他活性剂例如其他镇痛剂组合。

[0495] 在一些实施方案中，“附加成分”包括但不限于以下中的一种或多种：赋形剂；表面活性剂；分散剂；惰性稀释剂；造粒剂和崩解剂；粘合剂；润滑剂；甜味剂；调味剂；着色剂；防腐剂；生理可降解组合物，诸如明胶；水性媒介物和溶剂；油性媒介物和溶剂；悬浮剂；分散剂或润湿剂；乳化剂，缓和剂；缓冲剂；盐；增稠剂；填充剂；乳化剂；抗氧化剂；抗生素；抗真菌剂；稳定剂；以及药学上可接受的聚合材料或疏水性材料。在一些实施方案中，可包括在本发明的药物组合物中的“附加成分”是本领域已知的并且描述于例如Genaro编辑(1985,Remington's Pharmaceutical Sciences,Mack Publishing Co.,Easton,Pa.)中，所述文献通过引用并入本文。

[0496] 在一些实施方案中，本发明的组合物可包含按组合物的总重量计约0.005％至2.0％的防腐剂。在一些实施方案中，防腐剂用于在暴露于环境中的污染物的情况下防止腐败。在一些实施方案中，可用于本发明的防腐剂的实例包括但不限于选自由以下组成的组的那些：苄醇、山梨酸、对羟基苯甲酸酯、咪脲以及它们的组合。在一些实施方案中，特别优选的防腐剂是约0.5％至2.0％苯甲醇和0.05％至0.5％山梨酸的组合。

[0497] 在一些实施方案中，所述组合物优选地包括抗氧化剂和抑制化合物降解的螯合剂。在一些实施方案中，用于一些化合物的抗氧化剂是BHT、BHA、α-生育酚和抗坏血酸，其范围优选地为按组合物的总重量计约0.01重量％至0.3重量％以及BHT范围更优选地为0.03重量％至0.1重量％。在一些实施方案中，螯合剂以按组合物的总重量计0.01重量％至0.5重量％的量存在。特别优选的螯合剂包括依地酸盐(例如依地酸二钠)和柠檬酸，其重量范围为按组合物的总重量计约0.01重量％至0.20重量％、更优选地范围为0.02重量％至0.10重量％。在一些实施方案中，螯合剂可用于螯合组合物中可能对配制品的保质期有害的金属离子。在一些实施方案中，如本领域技术人员已知的，因此可以用其他合适和等效的抗氧化剂和螯合剂代替。

[0498] 在一些实施方案中，可以使用常规方法制备液体悬浮液以实现活性成分悬浮于水性或油性媒介物中。在一些实施方案中，水性媒介物包括例如水和等渗盐水。在一些实施方案中，油性媒介物包括例如杏仁油、油性酯、乙醇、植物油(诸如花生油、橄榄油、芝麻油或椰子油)、分馏的植物油和矿物油诸如液体石蜡。在一些实施方案中，液体悬浮液还可包含一种或多种附加成分，包括但不限于助悬剂、分散剂或润湿剂、乳化剂、缓和剂、防腐剂、缓冲剂、盐、调味剂、着色剂和甜味剂。在一些实施方案中，油性悬浮液还可包含增稠剂。在一些实施方案中，悬浮剂包括但不限于山梨糖醇糖浆、氢化食用脂肪、海藻酸钠、聚乙烯吡咯烷酮、黄蓍胶、阿拉伯树胶和纤维素衍生物诸如羧甲基纤维素钠、甲基纤维素、羟丙基甲基纤维素。在一些实施方案中，分散剂或润湿剂包括但不限于天然存在的磷脂诸如卵磷脂、环氧烷烃与脂肪酸、与长链脂肪醇、与来源于脂肪酸和己糖醇的偏酯、或与来源于脂肪酸和己糖醇酐的偏酯的缩合产物(例如分别为聚氧乙烯硬脂酸酯、十七乙烯氧鲸蜡醇(heptadecaethyleneoxycetanol)、聚氧乙烯山梨糖醇单油酸酯和聚氧乙烯脱水山梨糖醇单油酸酯)。已知的乳化剂包括但不限于卵磷脂和阿拉伯胶。已知的防腐剂包括但不限于对羟基苯甲酸甲酯、对羟基苯甲酸乙酯或对羟基苯甲酸正丙酯、抗坏血酸和山梨酸。已知的甜

味剂包括例如甘油、丙二醇、山梨糖醇、蔗糖和糖精。用于油性悬浮液的已知增稠剂包括例如蜂蜡、硬石蜡和鲸蜡醇。

[0499] 在一些实施方案中，活性成分在水性或油性溶剂中的液体溶液可以以与液体悬浮液基本相同的方式制备，主要区别在于活性成分溶解而不是悬浮在溶剂中。如本文所用，“油性”液体是包含含碳液体分子并且表现出比水小的极性特征的液体。在一些实施方案中，本发明的药物组合物的液体溶液可包含关于液体悬浮液所述的每种组分，应当理解悬浮剂不一定有助于活性成分在溶剂中的溶解。在一些实施方案中，水性溶剂包括例如水和等渗盐水。在一些实施方案中，油性溶剂包括例如杏仁油、油性酯、乙醇、植物油(诸如花生油、橄榄油、芝麻油或椰子油)、分馏的植物油和矿物油诸如液体石蜡。

[0500] 在一些实施方案中，本发明的药物制剂的粉末和颗粒配制品可以使用已知方法制备。在一些实施方案中，配制品可以直接施用于受试者，用于例如形成片剂，填充胶囊或通过向其中添加水性或油性媒介物来制备水性或油性悬浮液或溶液。在一些任何实施方案中，配制品还可包含分散剂或润湿剂、悬浮剂和防腐剂中的一种或多种。这些配制品中还可包含附加的赋形剂，诸如填充剂和甜味剂、调味剂或着色剂。

[0501] 在一些实施方案中，本发明的药物组合物还可以以水包油乳液或油包水乳液的形式制备、包装或销售。在一些实施方案中，油相可以是植物油诸如橄榄油或花生油、矿物油诸如液体石蜡、或这些的组合。在一些实施方案中，组合物还包含一种或多种乳化剂，诸如天然存在的树胶(诸如阿拉伯树胶或黄蓍胶)、天然存在的磷脂(诸如大豆磷脂或卵磷脂)、来源于脂肪酸和己糖醇酐的组合的酯或偏酯(诸如脱水山梨糖醇单油酸酯)以及此类偏酯与环氧乙烷的缩合产物(诸如聚氧乙烯脱水山梨糖醇单油酸酯)。在一些实施方案中，乳液还可以含有附加成分，包括例如甜味剂或调味剂。

[0502] IV.治疗方法

[0503] 在一些实施方案中，可以将本文提供的靶向脂质颗粒或如本文所述的其药物组合物施用于受试者，例如哺乳动物，例如人类。在此类实施方案中，受试者可能有风险患上特定疾病或病症，可能具有特定疾病或病症的症状，或可能被诊断或鉴定为患有特定疾病或病症。在一个实施方案中，受试者患有癌症。在一个实施方案中，受试者患有感染性疾病。在一些实施方案中，靶向脂质颗粒含有编码用于治疗受试者的疾病或病症的外源药剂的核酸序列。例如，外源药剂是靶向赘生性细胞或对赘生性细胞的蛋白质特异的外源药剂，并且将靶向脂质颗粒施用于受试者以治疗受试者的肿瘤或癌症。在另一个实例中，外源药剂是炎性介质或免疫分子，诸如细胞因子，并且将靶向脂质颗粒施用于受试者以治疗希望调节(例如增加)免疫反应的任何病症，诸如癌症或感染性疾病。在一些实施方案中，靶向脂质颗粒以实现治疗疾病、病症或病况的有效量或剂量施用。本文提供了任何所提供的靶向脂质颗粒在此类方法和治疗中的用途、以及在制备用以执行此类治疗方法的药物中的用途。在一些实施方案中，通过向患有、已经患有或怀疑患有所述疾病或病症或病况的受试者施用靶向脂质颗粒或包含靶向脂质颗粒的组合物来执行所述方法。在一些实施方案中，所述方法由此治疗受试者的疾病或病症或病况。本文还提供了任何组合物(诸如本文提供的药物组合物)用于治疗与由外源药剂靶向或由外源药剂提供的特定基因或蛋白质相关的疾病、病症或病况的用途。

[0504] 在一些实施方案中，所提供的方法或用途涉及药物组合物的施用，所述施用包括

口服、吸入、经皮或肠胃外(包括静脉内、肿瘤内、腹膜内、肌内、腔内和皮下)施用。在一些实施方案中,靶向脂质颗粒可以单独施用或配制为药物组合物。在一些实施方案中,可以将本文所述的靶向脂质颗粒或组合物施用于受试者,例如哺乳动物,例如人类。在一些任何实施方案中,受试者可能有风险患上特定疾病或病症(例如,本文所述的疾病或病症),可能具有特定疾病或病症(例如,本文所述的疾病或病症)的症状,或可能被诊断或鉴定为患有特定疾病或病症(例如,本文所述的疾病或病症)。在一些实施方案中,所述疾病是疾病或病况(disorder)。

[0505] 在一些实施方案中,可以将靶向脂质颗粒以单位剂量组合物的形式施用,所述单位剂量组合物诸如单位剂量口服、肠胃外、经皮或吸入组合物。在一些实施方案中,所述组合物通过混合制备并且适于口服、吸入、经皮或肠胃外施用,并且因此可以呈片剂、胶囊、口服液体制剂、粉末、颗粒、锭剂、可重构粉末、可注射和可输注溶液或悬浮液或栓剂或气雾剂的形式。

[0506] 在一些实施方案中,施用方案可能影响有效量的组成。在一些实施方案中,可以在疾病诊断之前或之后将治疗配制品施用于受试者。在一些实施方案中,可以每天或依序施用几个分剂量以及交错剂量,或者可以连续输注剂量,或者可以是推注注射。在一些实施方案中,如治疗或预防情况的紧急程度所指示,可以成比例地增加或减少治疗配制品的剂量。

[0507] 在一些实施方案中,可以使用已知程序以有效预防或治疗疾病的剂量和时间段进行本发明的组合物向受试者(优选地哺乳动物,更优选地人类)的施用。在一些实施方案中,实现治疗效果所必需的治疗化合物的有效量可以根据如下因素而变化,所述因素诸如所用的特定化合物的活性;施用的时间;化合物的排泄速率;治疗的持续时间;与化合物组合使用的其他药物、化合物或材料;所治疗的受试者的疾病或病况的状态、年龄、性别、体重、状况、一般健康和既往医疗史;以及医学领域熟知的类似因素。在一些实施方案中,可以调整给药方案以提供最佳治疗反应。在一些实施方案中,可以每天施用几个分剂量,或者如治疗情况的紧急程度所指示,可以成比例减少所述剂量。在一些实施方案中,本发明的治疗化合物的有效剂量范围为约1至5,000mg/kg体重/天。本领域普通技术人员将能够研究相关因素并确定治疗化合物的有效量而无需过多的实验。

[0508] 在一些实施方案中,化合物可以每天若干次的频率施用于受试者,或者其可以更少频率地施用,诸如每天一次、每周一次、每两周一次、每月一次、或甚至更少频率(诸如每数月一次或甚至每年一次或更少频率)。在一些实施方案中,在非限制性实例中,可以每天、每隔一天、每2天、每3天、每4天或每5天施用每天给药的化合物的量。在一些实施方案中,对于每隔一天施用,可以在星期一开始5mg/天的剂量,在星期三施用第一个随后5mg/天的剂量,在星期五施用第二个随后5mg/天的剂量,依此类推。剂量的频率对于本领域技术人员来说将是显而易见的,并且将取决于许多因素,诸如但不限于所治疗的疾病的类型和严重程度、动物的类型和年龄等。

[0509] 在一些实施方案中,本发明的药物组合物中的活性成分的剂量水平可以改变,以便获得对于特定受试者、组合物和施用模式可有效实现期望的治疗反应而对受试者无毒的活性成分的量。

[0510] 具有本领域普通技术的医生(例如医师或兽医)可以容易地确定和开出所需药物组合物的有效量。在一些实施方案中,医师或兽医可以以低于实现期望的治疗效果所需的

水平开始药物组合物中采用的本发明化合物的剂量，并逐渐增加所述剂量直至实现期望的效果。

[0511] 在一些实施方案中，尤其有利的是将化合物配制成易于施用和剂量均匀的剂量单位形式。在一些实施方案中，如本文所用的剂量单位形式是指适合作为用于待治疗受试者的单位剂量的物理离散单位；每个单位含有经计算可产生期望的治疗效果的预定量的治疗化合物以及期望的药物媒介物。在一些实施方案中，本发明的剂量单位形式由以下项规定并且直接取决于以下项：(a)治疗化合物的独特特征和待实现的特定治疗效果，和(b)本领域中混配/配制这种治疗化合物用于治疗受试者的疾病所固有的限制。

[0512] 在一些实施方案中，术语“容器”包括用于容纳药物组合物的任何贮器。在一些实施方案中，容器是含有药物组合物的包装。在其他实施方案中，容器不是含有药物组合物的包装，即容器是贮器，诸如含有包装的药物组合物或未包装的药物组合物和药物组合物的使用说明书的盒子或小瓶。应当理解，药物组合物的使用说明书可包含在含有药物组合物的包装上，并且因此所述说明书形成增加的与包装产品的功能关系。在一些实施方案中，说明书可包含关于化合物执行其预期功能(例如治疗或预防受试者的疾病或向受试者递送成像剂或诊断剂)的能力的信息。

[0513] 在一些实施方案中，本文公开的任何组合物的施用途径包括口服、鼻、直肠、肠胃外、舌下、经皮、经粘膜(例如舌下、舌、(经)颊、(经)尿道、阴道(例如经阴道和阴道周围)、鼻(内)和(经)直肠)、膀胱内、肺内、十二指肠内、胃内、鞘内、皮下、肌内、真皮内、动脉内、静脉内、支气管内、吸入和局部施用。

[0514] 在一些任何实施方案中，合适的组合物和剂型包括例如片剂、胶囊、囊片、丸剂、凝胶帽、锭剂、分散体、悬浮液、溶液、糖浆、颗粒、珠粒、经皮贴片、凝胶、粉末、小丸、乳浆剂、锭剂、乳膏、糊剂、硬膏剂、洗剂、原片、栓剂、用于鼻或口服施用的液体喷雾剂、用于吸入的干粉或气溶胶化配制品、用于膀胱内施用的组合物和配制品等。

[0515] 在一些实施方案中，包含外源药剂或运载物的靶向脂质颗粒组合物可以用于将这种外源药剂或运载物递送至细胞组织或受试者。在一些实施方案中，通过施用本文所述的靶向脂质颗粒组合物递送运载物可以改变细胞蛋白质表达水平。在某些实施方案中，所施用的组合物指导一种或多种运载物(例如多肽或mRNA)的上调(通过在细胞中表达、在细胞中递送或在细胞内诱导)，所述一种或多种运载物提供在递送多肽的细胞中基本上不存在或降低的功能活性。在一些实施方案中，缺失的功能活性本质上可以是酶活性、结构活性或调节活性。在一些实施方案中，所施用的组合物指导一种或多种多肽的上调，其(例如协同地)增加在多肽被上调的细胞中存在但基本上缺乏的功能活性。在一些任何实施方案中，所施用的组合物指导一种或多种运载物(例如多肽、siRNA或miRNA)的下调(通过在细胞中表达、在细胞中递送或在细胞内诱导)，所述一种或多种运载物抑制在递送多肽、siRNA或miRNA的细胞中存在或上调的功能活性。在一些任何实施方案中，上调的功能活性本质上可以是酶活性、结构活性或调节活性。在一些实施方案中，所施用的组合物指导一种或多种多肽的下调，其(例如协同地)降低在多肽被下调的细胞中存在或被上调的功能活性。在一些实施方案中，所施用的组合物指导某些功能活性的上调和其他功能活性的下调。

[0516] 在一些任何实施方案中，靶向脂质颗粒组合物(例如，包含线粒体或DNA的组合物)介导对靶细胞的作用，并且所述作用持续至少1、2、3、4、5、6或7天、2、3或4周、或1、2、3、6或

12个月。在一些实施方案中(例如,其中靶向脂质颗粒组合物包含外源蛋白),所述作用持续少于1、2、3、4、5、6或7天、2、3或4周、或1、2、3、6或12个月。

[0517] 在一些任何实施方案中,本文所述的靶向脂质颗粒组合物离体递送至细胞或组织,例如人类细胞或组织。在实施方案中,组合物改善离体细胞或组织的功能,例如改善细胞活力、呼吸或其他功能(例如本文所述的另一种功能)。

[0518] 在一些实施方案中,将组合物递送至处于损伤状态(例如,来自创伤、疾病、缺氧、缺血或其他损伤)的离体组织。

[0519] 在一些实施方案中,将组合物递送至离体移植物(例如用于移植的组织外植体或组织,例如人类静脉、肌肉骨骼移植物诸如骨或肌腱、角膜、皮肤、心脏瓣膜、神经;或分离或培养的器官,例如待移植到人类中的器官,例如人类心脏、肝、肺、肾、胰腺、肠、胸腺、眼睛)。在一些实施方案中,在移植之前、期间和/或之后将组合物递送至组织或器官。

[0520] 在一些实施方案中,将组合物与细胞(例如细胞制剂)一起递送、施用或与细胞(例如细胞制剂)接触。在一些实施方案中,细胞制剂可以是细胞治疗制剂(旨在施用于人类受试者的细胞制剂)。在实施方案中,细胞制剂包含表达嵌合抗原受体(CAR)(例如表达重组CAR)的细胞。表达CAR的细胞可以是例如T细胞、自然杀伤(NK)细胞、细胞毒性T淋巴细胞(CTL)、调节性T细胞。在实施方案中,细胞制剂是神经干细胞制剂。在实施方案中,细胞制剂是间充质干细胞(MSC)制剂。在实施方案中,细胞制剂是造血干细胞(HSC)制剂。在实施方案中,细胞制剂是胰岛细胞制剂。

[0521] 在一些实施方案中,可以将本文所述的靶向脂质颗粒组合物施用于受试者,例如哺乳动物,例如人类。在此类实施方案中,受试者可能有风险患上特定疾病或病症(例如,本文所述的疾病或病症),可能具有特定疾病或病症(例如,本文所述的疾病或病症)的症状,或可能被诊断或鉴定为患有特定疾病或病症(例如,本文所述的疾病或病症)。

[0522] 在一些实施方案中,靶向脂质颗粒的来源来自施用过靶向脂质颗粒组合物的同一受试者。在其他实施方案中,他们是不同的。在一些实施方案中,靶向脂质颗粒和受体组织的来源可以是自体的(来自同一受试者)或异源的(来自不同受试者)。在一些实施方案中,用于本文所述的靶向脂质颗粒组合物的供体组织可以是与受体组织不同的组织类型。在一些实施方案中,供体组织可以是肌肉组织,并且受体组织可以是结缔组织(例如脂肪组织)。在其他实施方案中,供体组织和受体组织可以是相同或不同的类型,但来自不同的器官系统。

[0523] 在一些实施方案中,可以将本文所述的靶向脂质颗粒组合物施用于患有癌症、自身免疫性疾病、感染性疾病、代谢疾病、神经退行性疾病或遗传疾病(例如酶缺陷)的受试者。在一些实施方案中,受试者需要再生。

[0524] 在一些实施方案中,靶向脂质颗粒与抑制膜融合的蛋白质的抑制剂共同施用。例如,Suppressyn是抑制细胞-细胞融合的人蛋白(Sugimoto等人,"A novel human endogenous retroviral protein inhibits cell-cell fusion"Scientific Reports 3:1462(DOI:10.1038/srep01462))。在一些实施方案中,靶向脂质颗粒与sypressyn的抑制剂(例如siRNA或抑制性抗体)共同施用。

[0525] V.示例性实施方案

[0526] 所提供的实施方案包括:

[0527] 1.一种靶向脂质颗粒,其包含:

[0528] (a)包围内腔的脂质双层,

[0529] (b)亨尼帕病毒F蛋白分子或其生物活性部分;以及

[0530] (c)靶向包膜蛋白,所述靶向包膜蛋白包含(i)亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分和(ii)单结构域抗体(sdAb)可变域,其中所述sdAb可变域与所述G蛋白或其生物活性部分的C末端连接,

[0531] 其中所述F蛋白分子或其生物活性部分和所述靶向包膜蛋白包埋在脂质双层中。

[0532] 2.如实施方案1所述的靶向脂质颗粒,其中所述单结构域抗体经由接头与所述G蛋白连接。

[0533] 3.如实施方案2所述的靶向脂质颗粒,其中所述接头是肽接头。

[0534] 4.一种靶向脂质颗粒,其包含:

[0535] (a)包围内腔的脂质双层,

[0536] (b)亨尼帕病毒F蛋白分子或其生物活性部分;以及

[0537] (c)靶向包膜蛋白,所述靶向包膜蛋白包含经由肽接头与单结构域抗体(sdAb)可变域连接的亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分,其中所述单结构域抗体结合靶细胞的细胞表面分子,

[0538] 其中所述F蛋白分子或其生物活性部分和所述靶向包膜蛋白包埋在脂质双层中。

[0539] 5.如实施方案1-4中任一项所述的靶向脂质颗粒,其中所述F蛋白分子或其生物活性部分的N末端暴露在脂质双层的外侧。

[0540] 6.如实施方案1-5中任一项所述的靶向脂质颗粒,其中所述G蛋白的C末端暴露在所述脂质双层的外侧。

[0541] 7.如实施方案1-6中任一项所述的靶向脂质颗粒,其中所述单结构域抗体结合存在于靶细胞上的细胞表面分子。

[0542] 8.如实施方案7所述的靶向脂质颗粒,其中所述细胞表面分子是蛋白质、聚糖、脂质或低分子量分子。

[0543] 9.如实施方案7所述的靶向脂质颗粒,其中所述靶细胞选自由以下组成的组:肿瘤浸润淋巴细胞、T细胞、赘生性或肿瘤细胞、病毒感染细胞、干细胞、中枢神经系统(CNS)细胞、造血干细胞(HSC)、肝细胞或完全分化细胞。

[0544] 10.如实施方案9所述的靶向脂质颗粒,其中所述靶细胞选自由以下组成的组:CD3+T细胞、CD4+T细胞、CD8+T细胞、肝细胞、造血干细胞、CD34+造血干细胞、CD105+造血干细胞、CD117+造血干细胞、CD105+内皮细胞、B细胞、CD20+B细胞、CD19+B细胞、癌细胞、CD133+癌细胞、EpCAM+癌细胞、CD19+癌细胞、Her2/Neu+癌细胞、GluA2+神经元、GluA4+神经元、NKG2D+自然杀伤细胞、SLC1A3+星形胶质细胞、SLC7A10+脂肪细胞或CD30+肺上皮细胞。

[0545] 11.如前述实施方案中任一项所述的靶向脂质颗粒,其中所述单结构域抗体结合存在于靶细胞上的抗原或其部分。

[0546] 12.如实施方案3-11中任一项所述的靶向脂质颗粒,其中所述肽接头包含多达65个氨基酸的长度。

[0547] 13.如实施方案3-11中任一项所述的靶向脂质颗粒,其中所述肽接头包含从或从约2至65个氨基酸、2至60个氨基酸、2至56个氨基酸、2至52个氨基酸、2至48个氨基酸、2至44

个氨基酸、2至40个氨基酸、2至36个氨基酸、2至32个氨基酸、2至28个氨基酸、2至24个氨基酸、2至20个氨基酸、2至18个氨基酸、2至14个氨基酸、2至12个氨基酸、2至10个氨基酸、2至8个氨基酸、2至6个氨基酸、6至65个氨基酸、6至60个氨基酸、6至56个氨基酸、6至52个氨基酸、6至48个氨基酸、6至44个氨基酸、6至40个氨基酸、6至36个氨基酸、6至32个氨基酸、6至28个氨基酸、6至24个氨基酸、6至20个氨基酸、6至18个氨基酸、6至14个氨基酸、6至12个氨基酸、6至10个氨基酸、6至8个氨基酸、8至65个氨基酸、8至60个氨基酸、8至56个氨基酸、8至52个氨基酸、8至48个氨基酸、8至44个氨基酸、8至40个氨基酸、8至36个氨基酸、8至32个氨基酸、8至28个氨基酸、8至24个氨基酸、8至20个氨基酸、8至18个氨基酸、8至14个氨基酸、8至12个氨基酸、8至10个氨基酸、10至65个氨基酸、10至60个氨基酸、10至56个氨基酸、10至52个氨基酸、10至48个氨基酸、10至44个氨基酸、10至40个氨基酸、10至36个氨基酸、10至32个氨基酸、10至28个氨基酸、10至24个氨基酸、10至20个氨基酸、10至18个氨基酸、10至14个氨基酸、10至12个氨基酸、12至65个氨基酸、12至60个氨基酸、12至56个氨基酸、12至52个氨基酸、12至48个氨基酸、12至44个氨基酸、12至40个氨基酸、12至36个氨基酸、12至32个氨基酸、12至28个氨基酸、12至24个氨基酸、12至20个氨基酸、12至18个氨基酸、12至14个氨基酸、14至65个氨基酸、14至60个氨基酸、14至56个氨基酸、14至52个氨基酸、14至48个氨基酸、14至44个氨基酸、14至40个氨基酸、14至36个氨基酸、14至32个氨基酸、14至28个氨基酸、14至24个氨基酸、14至20个氨基酸、14至18个氨基酸、18至65个氨基酸、18至60个氨基酸、18至56个氨基酸、18至52个氨基酸、18至48个氨基酸、18至44个氨基酸、18至40个氨基酸、18至36个氨基酸、18至32个氨基酸、18至28个氨基酸、18至24个氨基酸、18至20个氨基酸、20至65个氨基酸、20至60个氨基酸、20至56个氨基酸、20至52个氨基酸、20至48个氨基酸、20至44个氨基酸、20至40个氨基酸、20至36个氨基酸、20至32个氨基酸、20至28个氨基酸、20至26个氨基酸、20至24个氨基酸、24至65个氨基酸、24至60个氨基酸、24至56个氨基酸、24至52个氨基酸、24至48个氨基酸、24至44个氨基酸、24至40个氨基酸、24至36个氨基酸、24至32个氨基酸、24至30个氨基酸、24至28个氨基酸、28至65个氨基酸、28至60个氨基酸、28至56个氨基酸、28至52个氨基酸、28至48个氨基酸、28至44个氨基酸、28至40个氨基酸、28至36个氨基酸、28至34个氨基酸、28至32个氨基酸、32至65个氨基酸、32至60个氨基酸、32至56个氨基酸、32至52个氨基酸、32至48个氨基酸、32至44个氨基酸、32至40个氨基酸、32至38个氨基酸、32至36个氨基酸、36至65个氨基酸、36至60个氨基酸、36至56个氨基酸、36至52个氨基酸、36至48个氨基酸、36至44个氨基酸、36至40个氨基酸、40至65个氨基酸、40至60个氨基酸、40至56个氨基酸、40至52个氨基酸、40至48个氨基酸、40至44个氨基酸、44至65个氨基酸、44至60个氨基酸、44至56个氨基酸、44至52个氨基酸、44至48个氨基酸、48至65个氨基酸、48至60个氨基酸、48至56个氨基酸、48至52个氨基酸、50至65个氨基酸、50至60个氨基酸、50至56个氨基酸、50至52个氨基酸、54至65个氨基酸、54至60个氨基酸、54至56个氨基酸、58至65个氨基酸、58至60个氨基酸、或60至65个氨基酸。

[0548] 14.如实施方案3-11中任一项所述的靶向脂质颗粒，其中所述肽接头包含长度为3个、4个、5个、6个、7个、8个、9个、10个、11个、12个、13个、14个、15个、16个、17个、18个、19,20个、21个、22个、23个、24个、25个、26个、27个、28个、29个、30个、31个、32个、33个、34个、35个、36个、37个、38个、39个、40个、41个、42个、43个、44个、45个、46个、47个、48个、49个、50个、51个、52个、53个、54个、55个、56个、57个、58个、59个、60个、61个、62个、63个、64个或65

个氨基酸的多肽。

[0549] 15.如实施方案3-14中任一项所述的靶向脂质颗粒,其中所述肽接头是柔性接头,所述柔性接头包含GS、GGS、GGGGS(SEQ ID NO:43)、GGGGGS(SEQ ID NO:41)或它们的组合。

[0550] 16.如实施方案3-15中任一项所述的靶向脂质颗粒,其中所述肽接头包含(GGS)n,其中n为1至10。

[0551] 17.如实施方案3-15中任一项所述的靶向脂质颗粒,其中所述肽接头包含(GGGGS)n(SEQ ID NO:42),其中n为1至10。

[0552] 18.如实施方案3-15中任一项所述的靶向脂质颗粒,其中所述肽接头包含(GGGGGS)n(SEQ ID NO:27),其中n为1至6。

[0553] 19.如实施方案1-18中任一项所述的靶向脂质颗粒,其中所述G蛋白或其生物活性部分是野生型尼帕病毒G(NiV-G)蛋白或亨德拉病毒G蛋白。

[0554] 20.如实施方案1-19中任一项所述的靶向脂质颗粒,其中所述G蛋白或其生物活性部分是野生型NiV-G蛋白或其功能活性变体或生物活性部分。

[0555] 21.如实施方案20所述的靶向脂质颗粒,其中所述突变体NiV-G蛋白或其功能活性变体或生物活性部分包含与SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、至少为或约84%、至少为或约85%、至少为或约86%、至少为或约87%、至少为或约88%、至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的氨基酸序列。

[0556] 22.如实施方案21所述的靶向脂质颗粒,其中所述NiV-G蛋白是截短的生物活性部分,并且缺乏在野生型NiV-G蛋白(SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44)的N末端处或附近的多达40个连续氨基酸残基。

[0557] 23.如实施方案1-18中任一项所述的靶向脂质颗粒,其中所述NiV-G蛋白是在野生型NiV-G的N末端截短的生物活性部分,并且具有SEQ ID NO:10-15、35-40或45-50中任一者所示的序列、或与SEQ ID NO:10-15、35-40或45-50具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、至少为或约84%、至少为或约85%、至少为或约86%、至少为或约87%、至少为或约88%、至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的氨基酸序列。

[0558] 24.如实施方案21-23中任一项所述的靶向脂质颗粒,其中所述NiV-G蛋白具有在所述野生型NiV-G蛋白(SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44)的N末端处或附近的5个氨基酸截短。

[0559] 25.如实施方案24所述的靶向脂质颗粒,其中所述NiV-G蛋白具有SEQ ID NO:10所示的氨基酸序列、或与SEQ ID NO:10具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的氨基酸序列。

[0560] 26.如实施方案24所述的靶向脂质颗粒,其中所述NiV-G蛋白具有SEQ ID NO:35所示的氨基酸序列、或与SEQ ID NO:35具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的氨基酸序列。

[0561] 27.如实施方案24所述的靶向脂质颗粒,其中所述NiV-G蛋白具有SEQ ID NO:45所示的氨基酸序列、或与SEQ ID NO:45具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的氨基酸序列。

[0562] 28.如实施方案21-23中任一项所述的靶向脂质颗粒,其中所述NiV-G蛋白具有在所述野生型NiV-G蛋白(SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44)的N末端处或附近的10个氨基酸截短。

[0563] 29.如实施方案28所述的靶向脂质颗粒,其中所述NiV-G蛋白具有SEQ ID NO:11所示的氨基酸序列、或与SEQ ID NO:11具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的氨基酸序列。

[0564] 30.如实施方案28所述的靶向脂质颗粒,其中所述NiV-G蛋白具有SEQ ID NO:36所示的氨基酸序列、或与SEQ ID NO:36具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的氨基酸序列。

[0565] 31.如实施方案28所述的靶向脂质颗粒,其中所述NiV-G蛋白具有SEQ ID NO:46所示的氨基酸序列、或与SEQ ID NO:46具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的氨基酸序列。

[0566] 32.如实施方案21-23中任一项所述的靶向脂质颗粒,其中所述NiV-G蛋白具有在所述野生型NiV-G蛋白(SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44)的N末端处或附近的15个氨基酸截短。

[0567] 33.如实施方案32所述的靶向脂质颗粒,其中所述NiV-G蛋白具有SEQ ID NO:12所示的氨基酸序列、或与SEQ ID NO:12具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约

87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。

[0568] 34.如实施方案32所述的靶向脂质颗粒，其中所述NiV-G蛋白具有SEQ ID NO:37所示的氨基酸序列、或与SEQ ID NO:37具有至少为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。

[0569] 35.如实施方案32所述的靶向脂质颗粒，其中所述NiV-G蛋白具有SEQ ID NO:47所示的氨基酸序列、或与SEQ ID NO:47具有至少为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。

[0570] 36.如实施方案21-23中任一项所述的靶向脂质颗粒，其中所述NiV-G蛋白具有在所述野生型NiV-G蛋白(SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44)的N末端处或附近的20个氨基酸截短。

[0571] 37.如实施方案36所述的靶向脂质颗粒，其中所述NiV-G蛋白具有SEQ ID NO:13所示的氨基酸序列、或与SEQ ID NO:13具有至少为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。

[0572] 38.如实施方案36所述的靶向脂质颗粒，其中所述NiV-G蛋白具有SEQ ID NO:38所示的氨基酸序列、或与SEQ ID NO:38具有至少为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。

[0573] 39.如实施方案36所述的靶向脂质颗粒，其中所述NiV-G蛋白具有SEQ ID NO:48所示的氨基酸序列、或与SEQ ID NO:48具有至少为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。

[0574] 40.如实施方案21-23中任一项所述的靶向脂质颗粒，其中所述NiV-G蛋白具有在所述野生型NiV-G蛋白(SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44)的N末端处或附近的25个氨基酸截短。

[0575] 41.如实施方案40所述的靶向脂质颗粒,其中所述NiV-G蛋白具有SEQ ID NO:14所示的氨基酸序列、或与SEQ ID NO:14具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的氨基酸序列。

[0576] 42.如实施方案40所述的靶向脂质颗粒,其中所述NiV-G蛋白具有SEQ ID NO:39所示的氨基酸序列、或与SEQ ID NO:39具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的氨基酸序列。

[0577] 43.如实施方案40所述的靶向脂质颗粒,其中所述NiV-G蛋白具有SEQ ID NO:49所示的氨基酸序列、或与SEQ ID NO:49具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的氨基酸序列。

[0578] 44.如实施方案21-23中任一项所述的靶向脂质颗粒,其中所述NiV-G蛋白具有在所述野生型NiV-G蛋白(SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44)的N末端处或附近的30个氨基酸截短。

[0579] 45.如实施方案44所述的靶向脂质颗粒,其中所述NiV-G蛋白具有SEQ ID NO:15所示的氨基酸序列、或与SEQ ID NO:15具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的氨基酸序列。

[0580] 46.如实施方案44所述的靶向脂质颗粒,其中所述NiV-G蛋白具有SEQ ID NO:40所示的氨基酸序列、或与SEQ ID NO:40具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的氨基酸序列。

[0581] 47.如实施方案44所述的靶向脂质颗粒,其中所述NiV-G蛋白具有SEQ ID NO:50所示的氨基酸序列、或与SEQ ID NO:50具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的氨基酸序列。

[0582] 48.如实施方案21-23中任一项所述的靶向脂质颗粒，其中所述NiV-G蛋白具有在所述野生型NiV-G蛋白(SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44)的N末端处或附近的34个氨基酸截短。

[0583] 49.如实施方案48所述的靶向脂质颗粒，其中所述NiV-G蛋白具有SEQ ID NO:22所示的氨基酸序列、或与SEQ ID NO:22具有为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。

[0584] 50.如实施方案48所述的靶向脂质颗粒，其中所述NiV-G蛋白具有SEQ ID NO:53所示的氨基酸序列、或与SEQ ID NO:53具有为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。

[0585] 51.如实施方案1-48中任一项所述的靶向脂质颗粒，其中所述G蛋白或其生物活性部分是表现出与肝配蛋白B2或肝配蛋白B3的结合降低的突变体NiV-G蛋白。

[0586] 52.如实施方案51所述的靶向脂质颗粒，其中所述突变体NiV-G蛋白包含：

[0587] 一个或多个氨基酸取代，所述一个或多个氨基酸取代对应于选自由参考SEQ ID NO:28所示的编号的E501A、W504A、Q530A和E533A组成的组的氨基酸取代。

[0588] 53.如实施方案51或实施方案52所述的靶向脂质颗粒，其中所述突变体NiV-G蛋白具有SEQ ID NO:16所示的氨基酸序列、或与SEQ ID NO:16具有为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。

[0589] 54.如实施方案51或实施方案52所述的靶向脂质颗粒，其中所述突变体NiV-G蛋白具有SEQ ID NO:51所示的氨基酸序列、或与SEQ ID NO:51具有为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。

[0590] 55.如实施方案1-54中任一项所述的靶向脂质颗粒，其中所述F蛋白或其生物活性部分是野生型尼帕病毒F(NiV-F)蛋白或亨德拉病毒F蛋白或是其功能活性变体或生物活性部分。

[0591] 56.如实施方案1-55中任一项所述的靶向脂质颗粒，其中所述F蛋白或其生物活性部分是野生型NiV-F蛋白或其功能活性变体或生物活性部分。

[0592] 57.如实施方案1-56中任一项所述的靶向脂质颗粒，其中所述NiV-F蛋白或其功能活性变体或生物活性部分包含SEQ ID NO:2所示的氨基酸序列、或与SEQ ID NO:2具有为或

约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。

[0593] 58.如实施方案1-57中任一项所述的靶向脂质颗粒，其中所述NiV-F蛋白是在所述野生型NiV-F蛋白(SEQ ID NO:2)的C末端处或附近具有20个氨基酸截短的其生物活性部分。

[0594] 59.如实施方案58所述的靶向脂质颗粒，其中所述NiV-F蛋白具有与SEQ ID NO:5具有为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。

[0595] 60.如实施方案1-57中任一项所述的靶向脂质颗粒，其中所述NiV-F蛋白是其生物活性部分，所述生物活性部分包含：

[0596] i)在所述野生型NiV-F蛋白(SEQ ID NO:2)的C末端处或附近的20个氨基酸截短；和

[0597] ii)在N-连接糖基化位点上的点突变。

[0598] 61.如实施方案60所述的靶向脂质颗粒，其中所述NiV-F蛋白具有与SEQ ID NO:7具有为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。

[0599] 62.如实施方案1-57中任一项所述的靶向脂质颗粒，其中所述NiV-F蛋白是在所述野生型NiV-F蛋白(SEQ ID NO:2)的C末端处或附近具有22个氨基酸截短的其生物活性部分。

[0600] 63.如实施方案62所述的靶向脂质颗粒，其中所述NiV-F蛋白具有由编码与SEQ ID NO:8具有为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的序列的核苷酸序列编码的氨基酸序列。

[0601] 64.如实施方案63所述的靶向脂质颗粒，其中所述NiV-F蛋白具有与SEQ ID NO:23具有为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性

的氨基酸序列。

[0602] 65.如实施方案1-57中任一项所述的靶向脂质颗粒,其中所述F蛋白或其生物活性部分包含F1亚基或其促融合部分。

[0603] 66.如实施方案65所述的靶向脂质颗粒,其中所述F1亚基是所述F0前体的蛋白水解切割部分。

[0604] 67.如实施方案66所述的靶向脂质颗粒,其中所述F1亚基包含SEQ ID NO:4所示的序列、或与SEQ ID NO:4具有为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的氨基酸序列。

[0605] 68.如实施方案1-67中任一项所述的靶向脂质颗粒,其中所述脂质双层来源于用于产生逆转录病毒或逆转录病毒样颗粒的宿主细胞的膜。

[0606] 69.如实施方案1-60中任一项所述的靶向脂质颗粒,其中所述脂质双层是或包含病毒包膜。

[0607] 70.如实施方案68所述的靶向脂质颗粒,其中所述逆转录病毒样颗粒是复制缺陷的。

[0608] 71.如实施方案1-70中任一项所述的靶向脂质颗粒,其中所述靶向脂质颗粒包含除F蛋白分子和G蛋白之外的一种或多种病毒组分。

[0609] 72.如实施方案71所述的靶向脂质颗粒,其中所述一种或多种病毒组分来自逆转录病毒。

[0610] 73.如实施方案72所述的靶向脂质颗粒,其中所述逆转录病毒是慢病毒。

[0611] 74.如实施方案71-73中任一项所述的靶向脂质颗粒,其中所述一种或多种病毒组分包含选自Gag、Pol、Rev和Tat中的一种或多种的病毒包装蛋白。

[0612] 75.如实施方案71-74中任一项所述的靶向脂质颗粒,其中所述一种或多种病毒组分包含以下核酸序列中的一种或多种(例如,全部):5'LTR(例如,包含U5并且缺乏功能性U3结构域)、Psi包装元件(Psi)、中央多嘌呤区(cPPT)/中央终止序列(CTS)(例如,DNA瓣)、聚A尾序列、转录后调控元件(例如,WPRE)、Rev反应元件(RRE)和3'LTR(例如,包含U5并且缺乏功能性U3)。

[0613] 76.如实施方案1-75中任一项所述的靶向脂质颗粒,其中所述脂质颗粒还包含外源药剂。

[0614] 77.如实施方案76所述的靶向脂质颗粒,其中所述外源药剂存在于所述内腔中。

[0615] 78.如实施方案77所述的靶向脂质颗粒,其中所述外源药剂是蛋白质或核酸,任选地其中所述核酸是DNA或RNA。

[0616] 79.如实施方案76-78中任一项所述的靶向脂质颗粒,其中所述外源药剂编码治疗剂或诊断剂。

[0617] 80.如实施方案68-79中任一项所述的靶向脂质颗粒,其中所述宿主细胞选自由以下组成的组:CHO细胞、BHK细胞、MDCK细胞、C3H 10T1/2细胞、FLY细胞、Psi-2细胞、BOSC 23细胞、PA317细胞、WEHI细胞、COS细胞、BSC 1细胞、BSC 40细胞、BMT 10细胞、VERO细胞、W138

细胞、MRC5细胞、A549细胞、HT1080细胞、293细胞、293T细胞、B-50细胞、3T3细胞、NIH3T3细胞、HepG2细胞、Saos-2细胞、Huh7细胞、HeLa细胞、W163细胞、211细胞和211A细胞。

[0618] 81.如实施方案68-80中任一项所述的靶向脂质颗粒，其中所述宿主细胞包含293T细胞。

[0619] 82.一种多核苷酸，其包含编码以下的核酸序列：(i)亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分和(ii)单结构域抗体(sdAb)可变域，其中所述sdAb可变域与所述G蛋白或其生物活性部分的C末端连接。

[0620] 83.如实施方案82所述的多核苷酸，其还包含(iii)编码亨尼帕病毒F蛋白分子或其生物活性部分的核酸序列。

[0621] 84.如实施方案82或实施方案83所述的多核苷酸，其还包含可操作地连接以控制所述核酸表达的至少一个启动子。

[0622] 85.如实施方案83-84中任一项所述的多核苷酸，其中所述启动子是组成型启动子。

[0623] 86.如实施方案83-85中任一项所述的多核苷酸，其中所述启动子是诱导型启动子。

[0624] 87.如实施方案82-86中任一项所述的多核苷酸，其中所述sdAb可变域经由编码的肽接头与所述G蛋白连接。

[0625] 88.如实施方案86-87中任一项所述的多核苷酸，其中所述编码的肽接头包含多达65个氨基酸的长度。

[0626] 89.如实施方案86-87中任一项所述的多核苷酸，其中所述编码的肽接头包含从或从约2至65个氨基酸、2至60个氨基酸、2至56个氨基酸、2至52个氨基酸、2至48个氨基酸、2至44个氨基酸、2至40个氨基酸、2至36个氨基酸、2至32个氨基酸、2至28个氨基酸、2至24个氨基酸、2至20个氨基酸、2至18个氨基酸、2至14个氨基酸、2至12个氨基酸、2至10个氨基酸、2至8个氨基酸、2至6个氨基酸、6至65个氨基酸、6至60个氨基酸、6至56个氨基酸、6至52个氨基酸、6至48个氨基酸、6至44个氨基酸、6至40个氨基酸、6至36个氨基酸、6至32个氨基酸、6至28个氨基酸、6至24个氨基酸、6至20个氨基酸、6至18个氨基酸、6至14个氨基酸、6至12个氨基酸、6至10个氨基酸、6至8个氨基酸、8至65个氨基酸、8至60个氨基酸、8至56个氨基酸、8至52个氨基酸、8至48个氨基酸、8至44个氨基酸、8至40个氨基酸、8至36个氨基酸、8至32个氨基酸、8至28个氨基酸、8至24个氨基酸、8至20个氨基酸、8至18个氨基酸、8至14个氨基酸、8至12个氨基酸、8至10个氨基酸、10至65个氨基酸、10至60个氨基酸、10至56个氨基酸、10至52个氨基酸、10至48个氨基酸、10至44个氨基酸、10至40个氨基酸、10至36个氨基酸、10至32个氨基酸、10至28个氨基酸、10至24个氨基酸、10至20个氨基酸、10至18个氨基酸、10至14个氨基酸、10至12个氨基酸、12至65个氨基酸、12至60个氨基酸、12至56个氨基酸、12至52个氨基酸、12至48个氨基酸、12至44个氨基酸、12至40个氨基酸、12至36个氨基酸、12至32个氨基酸、12至28个氨基酸、12至24个氨基酸、12至20个氨基酸、12至18个氨基酸、12至14个氨基酸、14至65个氨基酸、14至60个氨基酸、14至56个氨基酸、14至52个氨基酸、14至48个氨基酸、14至44个氨基酸、14至40个氨基酸、14至36个氨基酸、14至32个氨基酸、14至28个氨基酸、14至24个氨基酸、14至20个氨基酸、14至18个氨基酸、18至65个氨基酸、18至60个氨基酸、18至56个氨基酸、18至52个氨基酸、18至48个氨基酸、18至44个氨基酸、18至40个氨基

酸、18至36个氨基酸、18至32个氨基酸、18至28个氨基酸、18至24个氨基酸、18至20个氨基酸、20至65个氨基酸、20至60个氨基酸、20至56个氨基酸、20至52个氨基酸、20至48个氨基酸、20至44个氨基酸、20至40个氨基酸、20至36个氨基酸、20至32个氨基酸、20至28个氨基酸、20至26个氨基酸、20至24个氨基酸、24至65个氨基酸、24至60个氨基酸、24至56个氨基酸、24至52个氨基酸、24至48个氨基酸、24至44个氨基酸、24至40个氨基酸、24至36个氨基酸、24至32个氨基酸、24至30个氨基酸、24至28个氨基酸、28至65个氨基酸、28至60个氨基酸、28至56个氨基酸、28至52个氨基酸、28至48个氨基酸、28至44个氨基酸、28至40个氨基酸、28至36个氨基酸、28至34个氨基酸、28至32个氨基酸、32至65个氨基酸、32至60个氨基酸、32至56个氨基酸、32至52个氨基酸、32至48个氨基酸、32至44个氨基酸、32至40个氨基酸、32至38个氨基酸、32至36个氨基酸、36至65个氨基酸、36至60个氨基酸、36至56个氨基酸、36至52个氨基酸、36至48个氨基酸、36至44个氨基酸、36至40个氨基酸、40至65个氨基酸、40至60个氨基酸、40至56个氨基酸、40至52个氨基酸、40至48个氨基酸、40至44个氨基酸、44至65个氨基酸、44至60个氨基酸、44至56个氨基酸、44至52个氨基酸、44至48个氨基酸、48至65个氨基酸、48至60个氨基酸、48至56个氨基酸、48至52个氨基酸、50至65个氨基酸、50至60个氨基酸、50至56个氨基酸、50至52个氨基酸、54至65个氨基酸、54至60个氨基酸、54至56个氨基酸、58至65个氨基酸、58至60个氨基酸、或60至65个氨基酸。

[0627] 90.如实施方案86-87中任一项所述的多核苷酸,其中所述编码的肽接头包含长度为3个、4个、5个、6个、7个、8个、9个、10个、11个、12个、13个、14个、15个、16个、17个、18个、19,20个、21个、22个、23个、24个、25个、26个、27个、28个、29个、30个、31个、32个、33个、34个、35个、36个、37个、38个、39个、40个、41个、42个、43个、44个、45个、46个、47个、48个、49个、50个、51个、52个、53个、54个、55个、56个、57个、58个、59个、60个、61个、62个、63个、64个或65个氨基酸的多肽。

[0628] 91.如实施方案86-87中任一项所述的多核苷酸,其中所述编码的肽接头包含GS、GGS、GGGGS(SEQ ID NO:43)、GGGGGS(SEQ ID NO:41)以及它们的组合。

[0629] 92.如实施方案86-87中任一项所述的多核苷酸,其中所述编码的肽接头包含(GGS)n,其中n为1至10。

[0630] 93.如实施方案86-87中任一项所述的多核苷酸,其中所述编码的肽接头包含(GGGGS)n(SEQ ID NO:42),其中n为1至10。

[0631] 94.如实施方案86-87中任一项所述的多核苷酸,其中所述编码的肽接头包含(GGGGGS)n(SEQ ID NO:27),其中n为1至4。

[0632] 95.如实施方案86-87中任一项所述的多核苷酸,其中编码所述G蛋白的核酸序列是野生型尼帕病毒G(NiV-G)蛋白或亨德拉病毒G蛋白,或是表现出与天然结合配偶体的结合降低的其变体。

[0633] 96.如实施方案82-95中任一项所述的多核苷酸,其中编码所述G蛋白的核酸序列是野生型NiV-G蛋白。

[0634] 97.如实施方案82-95中任一项所述的多核苷酸,其中编码所述G蛋白的核酸序列是表现出与肝配蛋白B2或肝配蛋白B3的结合降低的突变体NiV-G蛋白。

[0635] 98.如实施方案97所述的多核苷酸,其中编码所述突变体NiV-G蛋白的核酸序列包含与SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44具有至少为或约80%、至少为或约81%、

至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。

[0636] 99.如实施方案82-95和97中任一项所述的多核苷酸，其中编码所述突变体NiV-G蛋白的核酸序列包含SEQ ID NO:10-15、35-40或45-50中任一者所示的序列、或与SEQ ID NO:10-15、35-40或45-50具有至少为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。

[0637] 100.如实施方案97-99中任一项所述的多核苷酸，其中编码所述突变体NiV-G蛋白的核酸序列包含在所述野生型NiV-G蛋白(SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44)的N末端处或附近的5个氨基酸截短。

[0638] 101.如实施方案100所述的多核苷酸，其中编码所述突变体NiV-G蛋白的核酸序列包含SEQ ID NO:10所示的氨基酸序列、或与SEQ ID NO:10具有至少为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。

[0639] 102.如实施方案100所述的多核苷酸，其中编码所述突变体NiV-G蛋白的核酸序列包含SEQ ID NO:35所示的氨基酸序列、或与SEQ ID NO:35具有至少为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。

[0640] 103.如实施方案100所述的多核苷酸，其中编码所述突变体NiV-G蛋白的核酸序列包含SEQ ID NO:45所示的氨基酸序列、或与SEQ ID NO:45具有至少为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。

[0641] 104.如实施方案97-99中任一项所述的多核苷酸，其中编码所述突变体NiV-G蛋白的核酸序列包含在所述野生型NiV-G蛋白(SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44)的N末端处或附近的10个氨基酸截短。

[0642] 105.如实施方案104所述的多核苷酸，其中编码所述突变体NiV-G蛋白的核酸序列包含SEQ ID NO:11所示的氨基酸序列、或与SEQ ID NO:11具有至少为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或

约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。

[0643] 106.如实施方案104所述的多核苷酸，其中编码所述突变体NiV-G蛋白的核酸序列包含SEQ ID NO:36所示的氨基酸序列、或与SEQ ID NO:36具有至少为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。

[0644] 107.如实施方案104所述的多核苷酸，其中编码所述突变体NiV-G蛋白的核酸序列包含SEQ ID NO:46所示的氨基酸序列、或与SEQ ID NO:46具有至少为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。

[0645] 108.如实施方案97-99中任一项所述的多核苷酸，其中编码所述突变体NiV-G蛋白的核酸序列包含在所述野生型NiV-G蛋白(SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44)的N末端处或附近的15个氨基酸截短。

[0646] 109.如实施方案108所述的多核苷酸，其中编码所述突变体NiV-G蛋白的核酸序列包含SEQ ID NO:12所示的氨基酸序列、或与SEQ ID NO:12具有至少为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。

[0647] 110.如实施方案108所述的多核苷酸，其中编码所述突变体NiV-G蛋白的核酸序列包含SEQ ID NO:37所示的氨基酸序列、或与SEQ ID NO:37具有至少为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。

[0648] 111.如实施方案108所述的多核苷酸，其中编码所述突变体NiV-G蛋白的核酸序列包含SEQ ID NO:47所示的氨基酸序列、或与SEQ ID NO:47具有至少为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。

[0649] 112.如实施方案97-99中任一项所述的多核苷酸，其中编码所述突变体NiV-G蛋白的核酸序列包含在所述野生型NiV-G蛋白(SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44)的N末端处或附近的20个氨基酸截短。

[0650] 113.如实施方案112所述的多核苷酸，其中编码所述突变体NiV-G蛋白的核酸序列

包含SEQ ID NO:13所示的氨基酸序列、或与SEQ ID NO:13具有至少为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。

[0651] 114.如实施方案112所述的多核苷酸，其中编码所述突变体NiV-G蛋白的核酸序列包含SEQ ID NO:38所示的氨基酸序列、或与SEQ ID NO:38具有至少为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。

[0652] 115.如实施方案112所述的多核苷酸，其中编码所述突变体NiV-G蛋白的核酸序列包含SEQ ID NO:48所示的氨基酸序列、或与SEQ ID NO:48具有至少为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。

[0653] 116.如实施方案97-99中任一项所述的多核苷酸，其中编码所述突变体NiV-G蛋白的核酸序列包含在所述野生型NiV-G蛋白(SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44)的N末端处或附近的25个氨基酸截短。

[0654] 117.如实施方案116所述的多核苷酸，其中编码所述突变体NiV-G蛋白的核酸序列包含SEQ ID NO:14所示的氨基酸序列、或与SEQ ID NO:14具有至少为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。

[0655] 118.如实施方案116所述的多核苷酸，其中编码所述突变体NiV-G蛋白的核酸序列包含SEQ ID NO:39所示的氨基酸序列、或与SEQ ID NO:39具有至少为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。

[0656] 119.如实施方案116所述的多核苷酸，其中编码所述突变体NiV-G蛋白的核酸序列包含SEQ ID NO:49所示的氨基酸序列、或与SEQ ID NO:49具有至少为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。

[0657] 120.如实施方案97-99中任一项所述的多核苷酸，其中编码所述突变体NiV-G蛋白

的核酸序列包含在所述野生型NiV-G蛋白(SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44)的N末端处或附近的30个氨基酸截短。

[0658] 121.如实施方案120所述的多核苷酸,其中编码所述突变体NiV-G蛋白的核酸序列包含SEQ ID NO:15所示的氨基酸序列、或与SEQ ID NO:15具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的氨基酸序列。

[0659] 122.如实施方案120所述的多核苷酸,其中编码所述突变体NiV-G蛋白的核酸序列包含SEQ ID NO:40所示的氨基酸序列、或与SEQ ID NO:40具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的氨基酸序列。

[0660] 123.如实施方案120所述的多核苷酸,其中编码所述突变体NiV-G蛋白的核酸序列包含SEQ ID NO:50所示的氨基酸序列、或与SEQ ID NO:50具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的氨基酸序列。

[0661] 124.如实施方案97-99中任一项所述的多核苷酸,其中编码所述突变体NiV-G蛋白的核酸序列包含:

[0662] i)在所述N末端处或附近的截短;以及

[0663] ii)选自由以下组成的组的点突变:E501A、W504A、Q530A和E533A。

[0664] 125.如实施方案124所述的多核苷酸,其中编码所述突变体NiV-G蛋白的核酸序列包含SEQ ID NO:16所示的氨基酸序列、或与SEQ ID NO:16具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的氨基酸序列。

[0665] 126.如实施方案124所述的多核苷酸,其中编码所述突变体NiV-G蛋白的核酸序列包含SEQ ID NO:51所示的氨基酸序列、或与SEQ ID NO:51具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的氨基酸序列。

[0666] 127.一种载体,其包含如实施方案82-126中任一项所述的多核苷酸。

[0667] 128.如实施方案127所述的载体,其中所述载体是哺乳动物载体、病毒载体或人工染色体,任选地其中所述人工染色体是细菌人工染色体(BAC)。

[0668] 129.一种细胞,其包含如实施方案82-126中任一项所述的多核苷酸或如实施方案127或实施方案128所述的载体。

[0669] 130.一种制备靶向脂质颗粒的方法,所述靶向脂质颗粒包含亨尼帕病毒F蛋白分子或其生物活性部分和靶向包膜蛋白,所述靶向包膜蛋白包含亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分和单结构域抗体(sdAb)可变域,所述方法包括:

[0670] a)提供包含编码亨尼帕病毒F蛋白分子或其生物活性部分的核酸和编码靶向包膜蛋白的核酸的细胞,所述靶向包膜蛋白包含亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分和单结构域抗体(sdAb)可变域;

[0671] b)在允许产生靶向脂质颗粒的条件下培养所述细胞,以及

[0672] c)从所述细胞中分离、富集或纯化所述靶向脂质颗粒,从而制得所述靶向脂质颗粒。

[0673] 131.一种制备靶向脂质颗粒的方法,所述靶向脂质颗粒包含亨尼帕病毒F蛋白分子或其生物活性部分和靶向包膜蛋白,所述靶向包膜蛋白包含亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分和单结构域抗体(sdAb)可变域,所述方法包括:

[0674] a)提供包含如实施方案82-126中任一项所述的多核苷酸或如实施方案127或实施方案128所述的载体的细胞;

[0675] b)提供包含编码亨尼帕病毒F蛋白分子或其生物活性部分的多核苷酸的细胞;

[0676] c)在允许产生靶向脂质颗粒的条件下培养所述细胞,以及

[0677] d)从所述细胞中分离、富集或纯化所述靶向脂质颗粒,从而制得所述靶向脂质颗粒。

[0678] 132.如实施方案130或实施方案131所述的方法,其中所述细胞是哺乳动物细胞。

[0679] 133.如实施方案130-131中任一项所述的方法,其中所述细胞是生产者细胞,并且所述靶向脂质颗粒是病毒颗粒或病毒样颗粒,任选地逆转录病毒颗粒或逆转录病毒样颗粒,任选地慢病毒颗粒或慢病毒样颗粒。

[0680] 134.一种生产者细胞,其包含(i)病毒核酸和(ii)编码亨尼帕病毒F蛋白分子或其生物活性部分的核酸和(iii)编码靶向包膜蛋白的核酸,所述靶向包膜蛋白包含亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分和单结构域抗体(sdAb)可变域,任选地其中所述病毒核酸是慢病毒核酸。

[0681] 135.如实施方案134所述的生产者细胞,其中所述病毒核酸缺乏一个或多个参与病毒复制的基因。

[0682] 136.如实施方案134或实施方案135所述的生产者细胞,其中所述病毒核酸包含编码选自Gag、Pol、Rev和Tat中的一种或多种的病毒包装蛋白的核酸。

[0683] 137.如实施方案134-136中任一项所述的生产者细胞,其中所述病毒核酸包含:

[0684] 以下核酸序列中的一种或多种(例如,全部):5'LTR(例如,包含U5并且缺乏功能性U3结构域)、Psi包装元件(Psi)、中央多嘌呤区(cPPT)/中央终止序列(CTS)(例如,DNA瓣)、聚A尾序列、转录后调控元件(例如,WPRE)、Rev反应元件(RRE)和3'LTR(例如,包含U5并且缺乏功能性U3)。

[0685] 138.如实施方案134-137中任一项所述的生产者细胞,其中所述亨尼帕病毒F蛋白分子或其生物活性部分包含:

[0686] (i)SEQ ID NO:2所示的序列；

[0687] (ii)与SEQ ID NO:2具有为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、至少为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。

[0688] 139.如实施方案134-137中任一项所述的生产者细胞，其中所述亨尼帕病毒F蛋白分子或其生物活性部分包含：

[0689] (i)SEQ ID NO:5所示的序列；

[0690] (ii)与SEQ ID NO:5具有为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、至少为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。

[0691] 140.如实施方案134-137中任一项所述的生产者细胞，其中所述亨尼帕病毒F蛋白分子或其生物活性部分包含：

[0692] (i)SEQ ID NO:7所示的序列；

[0693] (ii)与SEQ ID NO:7具有为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、至少为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。

[0694] 141.如实施方案134-137中任一项所述的生产者细胞，其中所述亨尼帕病毒F蛋白分子或其生物活性部分包含：

[0695] (i)由编码SEQ ID NO:8所示序列的核苷酸序列编码的序列；

[0696] (ii)由编码与SEQ ID NO:8具有为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、至少为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的序列的核苷酸序列编码的氨基酸序列。

[0697] 142.如实施方案134-137中任一项所述的生产者细胞，其中所述亨尼帕病毒F蛋白分子或其生物活性部分包含：

[0698] (i)SEQ ID NO:23所示的序列；

[0699] (ii)与SEQ ID NO:23具有为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、至少为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。

[0700] 143.如实施方案134-142中任一项所述的生产者细胞，其中所述亨尼帕病毒包膜

附着糖蛋白G(G蛋白)或其生物活性部分包含：

[0701] (i)SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44所示的序列；

[0702] (ii)与SEQ ID NO:9、SEQ ID NO:28或SEQ ID NO:44具有为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、至少为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。

[0703] 144.如实施方案134-142中任一项所述的生产者细胞，其中所述亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分包含：

[0704] (i)SEQ ID NO:10所示的序列；

[0705] (ii)与SEQ ID NO:10具有至少为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、至少为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。

[0706] 145.如实施方案134-142中任一项所述的生产者细胞，其中所述亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分包含：

[0707] (i)SEQ ID NO:35所示的序列；

[0708] (ii)与SEQ ID NO:35具有至少为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、至少为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。

[0709] 146.如实施方案134-142中任一项所述的生产者细胞，其中所述亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分包含：

[0710] (i)SEQ ID NO:45所示的序列；

[0711] (ii)与SEQ ID NO:45具有至少为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、至少为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。

[0712] 147.如实施方案134-142中任一项所述的生产者细胞，其中所述亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分包含：

[0713] (i)SEQ ID NO:11所示的序列；

[0714] (ii)与SEQ ID NO:11具有为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、至少为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。

[0715] 148.如实施方案134-142中任一项所述的生产者细胞，其中所述亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分包含：

[0716] (i)SEQ ID NO:36所示的序列；

[0717] (ii)与SEQ ID NO:36具有为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、至少为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。

[0718] 149.如实施方案134-142中任一项所述的生产者细胞，其中所述亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分包含：

[0719] (i)SEQ ID NO:46所示的序列；

[0720] (ii)与SEQ ID NO:46具有为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、至少为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。

[0721] 150.如实施方案134-142中任一项所述的生产者细胞，其中所述亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分包含：

[0722] (i)SEQ ID NO:12所示的序列；

[0723] (ii)与SEQ ID NO:12具有为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、至少为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。

[0724] 151.如实施方案134-142中任一项所述的生产者细胞，其中所述亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分包含：

[0725] (i)SEQ ID NO:37所示的序列；

[0726] (ii)与SEQ ID NO:37具有为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、至少为或约96％、至少为或约97％、至少为或约98％、或至少为或约99％序列同一性的氨基酸序列。

[0727] 152.如实施方案134-142中任一项所述的生产者细胞，其中所述亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分包含：

[0728] (i)SEQ ID NO:47所示的序列；

[0729] (ii)与SEQ ID NO:47具有为或约80％、至少为或约81％、至少为或约82％、至少为或约83％、为或约84％、至少为或约85％、至少为或约86％、或至少为或约87％、至少为或约88％、或至少为或约89％、至少为或约90％、至少为或约91％、至少为或约92％、至少为或约93％、至少为或约94％、至少为或约95％、至少为或约96％、至少为或约97％、至少为或约

98%、或至少为或约99%序列同一性的氨基酸序列。

[0730] 153.如实施方案134-142中任一项所述的生产者细胞，其中所述亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分包含：

[0731] (i)SEQ ID NO:13所示的序列；

[0732] (ii)与SEQ ID NO:13具有为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、至少为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的氨基酸序列。

[0733] 154.如实施方案134-142中任一项所述的生产者细胞，其中所述亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分包含：

[0734] (i)SEQ ID NO:38所示的序列；

[0735] (ii)与SEQ ID NO:38具有为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、至少为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的氨基酸序列。

[0736] 155.如实施方案134-142中任一项所述的生产者细胞，其中所述亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分包含：

[0737] (i)SEQ ID NO:48所示的序列；

[0738] (ii)与SEQ ID NO:48具有为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、至少为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的氨基酸序列。

[0739] 156.如实施方案134-142中任一项所述的生产者细胞，其中所述亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分包含：

[0740] (i)SEQ ID NO:14所示的序列；

[0741] (ii)与SEQ ID NO:14具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、至少为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的氨基酸序列。

[0742] 157.如实施方案134-142中任一项所述的生产者细胞，其中所述亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分包含：

[0743] (i)SEQ ID NO:39所示的序列；

[0744] (ii)与SEQ ID NO:39具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为

或约93%、至少为或约94%、至少为或约95%、至少为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的氨基酸序列。

[0745] 158.如实施方案134-142中任一项所述的生产者细胞，其中所述亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分包含：

[0746] (i)SEQ ID NO:49所示的序列；

[0747] (ii)与SEQ ID NO:49具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、至少为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的氨基酸序列。

[0748] 159.如实施方案134-142中任一项所述的生产者细胞，其中所述亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分包含：

[0749] (i)SEQ ID NO:15所示的序列；

[0750] (ii)与SEQ ID NO:15具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、至少为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的氨基酸序列。

[0751] 160.如实施方案134-142中任一项所述的生产者细胞，其中所述亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分包含：

[0752] (i)SEQ ID NO:40所示的序列；

[0753] (ii)与SEQ ID NO:40具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、至少为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的氨基酸序列。

[0754] 161.如实施方案134-142中任一项所述的生产者细胞，其中所述亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分包含：

[0755] (i)SEQ ID NO:50所示的序列；

[0756] (ii)与SEQ ID NO:50具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约87%、至少为或约88%、或至少为或约89%、至少为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、至少为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的氨基酸序列。

[0757] 162.如实施方案134-142中任一项所述的生产者细胞，其中所述亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分包含：

[0758] (i)SEQ ID NO:16所示的序列；

[0759] (ii)与SEQ ID NO:16具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约87%、至少为

或约88%、或至少为或约89%、为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、至少为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的氨基酸序列。

[0760] 163.如实施方案134-142中任一项所述的生产者细胞,其中所述亨尼帕病毒包膜附着糖蛋白G(G蛋白)或其生物活性部分包含:

[0761] (i)SEQ ID NO:51所示的序列;

[0762] (ii)与SEQ ID NO:51具有至少为或约80%、至少为或约81%、至少为或约82%、至少为或约83%、为或约84%、至少为或约85%、至少为或约86%、或至少为或约87%、至少为或约88%、或至少为或约89%、为或约90%、至少为或约91%、至少为或约92%、至少为或约93%、至少为或约94%、至少为或约95%、至少为或约96%、至少为或约97%、至少为或约98%、或至少为或约99%序列同一性的氨基酸序列。

[0763] 164.一种病毒载体颗粒或病毒样颗粒,其由如实施方案134-163中任一项所述的生产者细胞产生。

[0764] 165.一种组合物,其包含如实施方案1-81和173-176中任一项所述的多个靶向脂质颗粒。

[0765] 166.如实施方案165所述的组合物,其还包含药学上可接受的载体。

[0766] 167.如实施方案165或实施方案166所述的药物组合物,其中所述靶向脂质颗粒包含小于1μm的平均直径。

[0767] 168.一种将外源药剂递送至受试者(例如,人类受试者)的方法,所述方法包括向所述受试者施用如实施方案1-81和173-176中任一项所述的靶向脂质颗粒或如实施方案165-167和177中任一项所述的组合物。

[0768] 169.一种治疗受试者(例如,人类受试者)的疾病或病况的方法,所述方法包括向所述受试者施用如实施方案1-81和173-176中任一项所述的靶向脂质颗粒或如实施方案165-167和177中任一项所述的组合物。

[0769] 170.一种使哺乳动物细胞与靶向脂质颗粒融合的方法,所述方法包括向所述受试者施用如实施方案1-81和173-176中任一项所述的靶向脂质颗粒或如实施方案165-167和177中任一项所述的组合物。

[0770] 171.如实施方案170所述的方法,其中所述哺乳动物细胞与所述靶向脂质颗粒的融合使外源药剂递送至受试者(例如,人类受试者)。

[0771] 172.如实施方案170或实施方案171所述的方法,其中所述哺乳动物细胞与所述靶向脂质颗粒的融合治疗受试者(例如,人类受试者)的疾病或病况。

[0772] 173.如实施方案1-81中任一项所述的靶向脂质颗粒,其中与已将相同包膜蛋白掺入相似脂质双层中但与替代靶向部分融合的参考脂质颗粒相比,所述靶向脂质颗粒具有更多的所述靶向包膜蛋白表达,任选地其中所述替代靶向部分是单链可变片段(scFv)。

[0773] 174.如实施方案173所述的靶向脂质颗粒,其中所述表达增加了为或大于5%、10%、20%、30%、40%、50%、60%、70%、80%、90%、100%、125%、150%、200%、300%、400%、500%或更多。

[0774] 175.如实施方案173所述的靶向脂质颗粒,其中所述表达增加了为或大于1.5倍、2倍、3倍、4倍、5倍、6倍、7倍、8倍、9倍、10倍、15倍、20倍、30倍或更多,优选地为或约或大于10

倍或更多。

[0775] 176.如实施方案1-81和173-175中任一项所述的靶向脂质颗粒或如实施方案164所述的病毒载体颗粒或病毒样颗粒,其中转导后在靶细胞中的滴度为或大于$1x10^6$个转导单位(TU)/mL、为或大于$2x10^6$TU/mL、为或大于$3x10^6$TU/mL、为或大于$4x10^6$TU/mL、为或大于$5x10^6$TU/mL、为或大于$6x10^6$TU/mL、为或大于$7x10^6$TU/mL、为或大于$8x10^6$TU/mL、为或大于$9x10^6$TU/mL、或为或大于$1x10^7$TU/mL。

[0776] 177.如实施方案165-167中任一项所述的组合物,其中在所述组合物中的脂质颗粒群体中,大于为或约50%、大于为或约55%、大于为或约60%、大于为或约65%、大于为或约70%、或大于为或约75%对所述靶向包膜蛋白呈表面阳性。

[0777] 178.如实施方案1-81和173-176中任一项所述的靶向脂质颗粒,其中所述靶向包膜蛋白以至少约(0.001、0.002、0.005、0.01、0.02、0.05、0.1、0.2或0.5)个靶向包膜蛋白/nm^2的密度存在于所述靶向脂质颗粒的表面上。

[0778] 179.一种组合物,其包含如实施方案1-81、173-176和178中任一项所述的多个靶向脂质颗粒,其中所述靶向包膜蛋白以至少约(0.001、0.002、0.005、0.01、0.02、0.05、0.1、0.2或0.5)个靶向包膜蛋白/nm^2的平均密度存在于所述靶向脂质颗粒的表面上。

[0779] 180.如实施方案134-163中任一项所述的生产者细胞,其中与已将相同的包膜蛋白掺入其膜(例如,质膜)中但与替代靶向部分融合的参考生产者细胞相比,所述生产者细胞具有更多的所述靶向包膜蛋白的膜(例如质膜)表达,任选地其中所述替代靶向部分是单链可变片段(scFv)。

[0780] 181.如实施方案180所述的生产者细胞,其中所述表达增加了为或大于5%、10%、20%、30%、40%、50%、60%、70%、80%、90%、100%、125%、150%、200%、300%、400%、500%或更多。

[0781] 182.如实施方案180所述的生产者细胞,其中所述表达增加了为或大于1.5倍、2倍、3倍、4倍、5倍、6倍、7倍、8倍、9倍、10倍、15倍、20倍、30倍或更多,优选地为或约或大于10倍或更多。

[0782] 183.如实施方案134-163和180-182中任一项所述的生产者细胞,其中所述生产者细胞具有每平方微米至少20个蛋白质(例如,至少50个、100个、200个、500个、1000个、2000个、5000个或10,000个蛋白质)的所述靶向包膜蛋白在所述生产者细胞的膜(例如,质膜)上的表达。

[0783] 184.如实施方案134-163和180-183中任一项所述的生产者细胞,其中所述靶向包膜蛋白构成所述生产者细胞的总膜(例如,质膜)蛋白的至少0.1%(例如,至少0.2%、0.5%、1%、2%、3%、4%、5%、6%、7%、8%、9%或10%)的(例如,按总蛋白重量计)。

[0784] 实施例

[0785] 包括以下实施例仅用于说明目的,并不旨在限制本发明的范围。

[0786] 实施例1:含有靶向结合剂的生产者细胞的产生和表征。

[0787] 本实施例描述了NiVG靶向结合序列的产生和评估,其中NiVG与scFv或VHH结合形态连接。

[0788] A.针对CD4的结合形态。

[0789] 产生含有抗人细胞受体CD4的scFv或VHH结合形态的示例性再靶向NivG融合剂构

建体。对于每种结合形态，评估含有独特CDR3的四种不同序列。将每个示例性结合剂序列进行密码子优化，并作为与编码NiVG(Gc Δ34;Bender等人2016 PLoS Pathol 12(6):e1005641)的序列的融合物克隆到表达载体中。所得载体编码含有NiVG的NivG靶向结构域(SEQ ID NO:16)、柔性接头和结合结构域，随后是用于检测的6xHis-标签(NivG-接头-scFv-6xHis)。

[0790] 亚克隆后，使用转染试剂将5μg每种示例性构建体转染到HEK293细胞中。将pcDNA3.1质粒(空载体)和没有结合剂结构域的表达载体(NiVG-接头-无结合剂)用作阴性对照。

[0791] 在转染后48小时，收获细胞并将100,000个细胞与50nM或300nM具有人Fc标签的可溶性人CD4蛋白(hCD4-Fc)在4℃下温育1小时。温育后，洗涤细胞并且用与Alexa-647缀合的抗His抗体共染色以检测NivG-结合剂的表面表达，并且用与Alexa-488缀合的抗人Fc抗体共染色以检测与可溶性hCD4-Fc蛋白的结合。

[0792] 通过流式细胞术分析细胞，并且基于阴性对照空载体(pcDNA3.1)设置His(表面表达)和Fc(CD4-蛋白结合)的门控。用含有VHH结合形态的构建体转染的细胞的中值荧光强度(MFI)的评估证明，与用含有scFv结合形态的构建体转染的细胞相比，由His+细胞％定量的表面表达更高(图1A)，由Fc+细胞％定量的与可溶性hCD4-Fc蛋白的结合更高(图1B)。

[0793] B.针对多种细胞受体的结合形态

[0794] 产生示例性构建体，其含有如上所述的scFv和VHH结合形态，但含有针对其他细胞受体hCD8、CD4、ASGR2、TM4SF5、LDLR或ASGR1的独特序列。对含有不同细胞受体的每种结合形态评估各自含有独特CDR3的多个序列。在如上所述亚克隆到NivG-接头--6xHis表达载体中之后，将5μg的每种示例性构建体转染到约HEK293细胞中。将pcDNA3.1质粒(空载体)和没有结合结构域的表达载体(NiVG-接头-无结合剂)用作阴性对照。

[0795] 转染后48小时，收获细胞并且将100,000个细胞洗涤并用与Alexa-647缀合的抗-His抗体染色以检测NivG-结合剂的表面表达。通过流式细胞术分析细胞，并且基于阴性对照空载体(pcDNA3.1)设置His(表面表达)的门控。将中值荧光强度(MFI)归一化至设定为100的NivG-无结合剂对照。与scFv结合形态相比，用含有VHH结合形态的构建体转染的细胞证明了靶向结合序列在293细胞上的表面表达更高，如由His+细胞％定量的(图1C)。

[0796] 实施例2:用靶向结合剂假型化的慢病毒的产生和表征。

[0797] 本实施例描述了用NivG再靶向融合剂假型化的慢病毒的产生和原代人T细胞转导的评估。

[0798] A.NivG假型化慢病毒的产生。

[0799] 将293细胞以5.4x 10^6接种到10cm平皿中并静置24小时。在铺板后24小时，使用聚乙烯亚胺(PEI)用以下质粒转染细胞:含有与scFv或VHH结合形态连接的hCD4靶向结合序列的NivG假型化载体(NivG-接头-hCD4-结合形态)、含有编码NivF序列NivFdel22(SEQ ID NO:8;或没有信号序列的SEQ ID NO:23;Bender等人2016PLoS)的核苷酸序列的载体、含有空骨架、HIV-1pol、HIV-1gag、HIV-1Rev、HIV-1Tat、AmpR启动子和SV40启动子的包装质粒、以及编码在SFFV启动子控制下的增强的绿色荧光蛋白(eGFP)的慢病毒报告质粒pLenti-SFFV-eGFP。使用上文所述的质粒连同4μg VSV-G产生阳性对照细胞。

[0800] B.原代人T细胞的NivG假型化慢病毒转导效率。

[0801] 将经阴性选择以富集T细胞的来自外周血的PanT细胞(StemCellTech,加拿大温哥华)解冻并用抗CD3/抗CD28激活2天。如上所述产生的浓缩慢病毒从0.05稀释度开始连续稀释6倍,其中在稀释系列中总共有4个点。将慢病毒添加到100,000个PanT细胞中并通过在25C下以1000g离心感染90分钟进行转导。转导的PanT细胞在转导后第2天和第5天分裂,并且在转导后第7天,收获细胞并用缀合Alexa-647的抗人CD4抗体染色。通过流式细胞术分析细胞,并且通过为GFP+的CD4阳性细胞%确定滴度。用含有VHH结合形态的构建体转染的细胞表现出比在原代人T细胞上含有scFv结合形态的构建体增加10倍的滴度(图2)。

[0802] 实施例3.用CD8靶向结合剂假型化的慢病毒的体内递送。

[0803] 本实施例描述了用CD8 NivG再靶向融合剂假型化的慢病毒的产生和原代人T细胞转导的体内评估。

[0804] 基本上如实施例2所述产生CD8再靶向NivG融合剂。再靶向NivG假型化融合剂包含含有NiVG(SEQ ID NO:16)的NivG靶向结构域、柔性接头和示例性CD8结合结构域(VHH或scFv结合形态)。

[0805] 将来自人外周血单核细胞(PBMC)的T细胞用抗CD3/抗CD28激活3天。温育3天后,将$1x10^7$个细胞腹膜内注射到NOD-scid-IL2rγ^{无效}小鼠中。注射后一天,小鼠通过腹膜内注射接受$1x10^7$个转导单位(TU)的如上所述产生的CD8 NivG假型化慢病毒、或无慢病毒载体(LVV)对照。在CD8 NivG假型化慢病毒注射后第7天,收获腹膜细胞并通过流式细胞术分析,并且通过呈GFP+的CD8阳性或阴性细胞%确定滴度。CD8再靶向假型化慢病毒表现出CD8+T细胞的显著体内转导(图3A)和CD8-T细胞的最小转导(图3B)。这些结果表明CD8靶向假型化慢病毒介导的递送允许将转基因特异性递送至预期的细胞类型(例如CD8+T细胞)。

[0806] 实施例4.含有具有CD8靶向结合剂的假型化慢病毒的嵌合抗原受体(CAR)的体外评估。

[0807] 本实施例描述了用CD8再靶向融合剂假型化并表达针对CD19的嵌合抗原受体(CD19CAR)的慢病毒的体外肿瘤杀伤活性。基本上如实施例3所述产生慢病毒,除了将编码eGFP或CD19CAR的质粒转染到293生产者细胞中。CD19CAR含有针对CD19的抗scFv和含有4-1BB和CD3-ζ的细胞内组分的细胞内信号传导结构域。

[0808] 用抗CD3/抗CD28试剂激活人外周血单核细胞(PBMC),并以不同浓度范围(10-10,000个转导单位/孔)用表达CD19+CAR或GFP的CD8再靶向NivG慢病毒进行转导。在第3天将RFP+Nalm6白血病细胞添加培养物中,并且在18小时时通过流式细胞术评估Nalm6细胞的清除。

[0809] 如图4A所示,在转导后4天用两种CD8再靶向融合剂特异性地检测CD8+细胞中的CD19+CAR表达。表达CD19CAR的转导的CD8+T细胞也介导CD19+Nalm6白血病细胞杀伤的强效和慢病毒剂量依赖性增加,而相反,转导以表达GFP的细胞不表现出靶细胞杀伤(图4B)。

[0810] 这些结果证明具有编码CD19CAR的转基因的CD8再靶向假型化慢病毒将CD19CAR递送至人CD8+T细胞以介导PBMC的复杂混合物中CD8+T细胞的特异性转导,并且通过体外白血病细胞杀伤显示出剂量依赖性抗肿瘤反应。

[0811] 本发明的范围并不旨在限于所公开的特定实施方案,这些实施方案是例如为了说明本发明的各个方面而提供的。根据本文的描述和教导,对所描述的组合物和方法的各种修改将变得显而易见。可以在不脱离本公开的真实范围和精神的情况下实践此类变化,并

且此类变化旨在落入本公开的范围内。

[0812] 序列

[0813]

#	序列	注释
1	MVVILDKRCY CNLLILILMI SECSVGILHY EKLSKIGLVK GVTRKYKIKS NPLTKDIVIK MIPNVSNMSQ CTGSVMENYK TRLNGILTPI KGALEIYKNN THDLVGDVRL AGVIMAGVAI GIATAAQITA GVALYEAMKN ADNINKLKSS IESTNEAVVK LQETAEKTVY VLTALQDYIN TNLVPTIDKI SCKQTELSLD LALSKYLSDL LFVFGPNLQD PVSNSMTIQA ISQAFGGNYE TLLRTLGYAT EDFDDLLESD SITGQIIYVD LSSYYIIVRV YFPILTEIQQ AYIQELLPVS FNNDNSEWIS IVPNFILVRN TLISNIEIGF CLITKRSVIC NQDYATPMTN NMRECLTGST EKCPRELVVS SHVPRFALSN GVLFANCISV TCQCQTTGRA ISQSGEQTLL MIDNTTCPTA VLGNVIISLG KYLGSVNYNS EGIAIGPPVF TDKVDISSQI SSMNQSLQQS KDYIKEAQRL LDTVNPSLIS MLSMIILYVL SIASLCIGLI TFISFIIVEK KRNTYSRLED RRVRPTSSGD LYYIGT	具有信号序列的尼帕病毒 NiV-F (aa 1-546) Uniprot Q9IH63

[0814]

2	ILHY EKLSKIGLVK GVTRKYKIKS NPLTKDIVIK MIPNVSNMSQ CTGSVMENYK TRLNGILTPI KGALEIYKNN THDLVGDVRL AGVIMAGVAI GIATAAQITA GVALYEAMKN ADNINKLKSS IESTNEAVVK LQETAEKTVY VLTALQDYIN TNLVPTIDKI SCKQTELSLD LALSKYLSDL LFVFGPNLQD PVSNSMTIQA ISQAFGGNYE TLLRTLGYAT EDFDDLLESD SITGQIIYVD LSSYYIIVRV YFPILTEIQQ AYIQELLPVS FNNDNSEWIS IVPNFILVRN TLISNIEIGF CLITKRSVIC NQDYATPMTN NMRECLTGST EKCPRELVVS SHVPRFALSN GVLFANCISV TCQCQTTGRA ISQSGEQTLL MIDNTTCPTA VLGNVIISLG KYLGSVNYNS EGIAIGPPVF TDKVDISSQI SSMNQSLQQS KDYIKEAQRL LDTVNPSLIS MLSMIILYVL SIASLCIGLI TFISFIIVEK KRNTYSRLED RRVRPTSSGD LYYIGT	尼帕病毒 NiV-F F0 (aa 27-546)
3	ILHYEKLSKIGLVKGVTRKYKIKSNPLTKDIVIKMIPNVSNMSQCTGSVME NYKTRLNGILTPIKGALEIYKNNTHDLVGDVR	尼帕病毒 NiV-F F2 (aa 27-109)
4	LAGVIMAGVAIGIATAAQITAGVALYEAMKNADNINKLKSSIESTNEAVVK LQETAEKTVYVLTALQDYINTNLVPTIDKISCKQTELSLDLALSKYLSDLL FVFGPNLQDPVSNSMTIQAISQAFGGNYETLLRTLGYATEDFDDLLESDSI TGQIIYVDLSSYYIIVRVYFPILTEIQQAYIQELLPVSFNNDNSEWISIVP NFILVRNTLISNIEIGFCLITKRSVICNQDYATPMTNNMRECLTGSTEKCP RELVVSSHVPRFALSNGVLFANCISVTCQCQTTGRAISQSGEQTLLMIDNT TCPTAVLGNVIISLGKYLGSVNYNSEGIAIGPPVFTDKVDISSQISSMNQS LQQSKDYIKEAQRLLDTVNPSLISMLSMIILYVLSIASLCIGLITFISFII VEKKRNTYSRLEDRRVRPTSSGDLYYIGT	尼帕病毒 NiV F F1 (aa 110-546)
5	ILHY EKLSKIGLVK GVTRKYKIKS NPLTKDIVIK MIPNVSNMSQ CTGSVMENYK TRLNGILTPI KGALEIYKNN THDLVGDVRL AGVIMAGVAI GIATAAQITA GVALYEAMKN ADNINKLKSS IESTNEAVVK LQETAEKTVY VLTALQDYIN TNLVPTIDKI SCKQTELSLD LALSKYLSDL LFVFGPNLQD PVSNSMTIQA ISQAFGGNYE TLLRTLGYAT EDFDDLLESD SITGQIIYVD LSSYYIIVRV YFPILTEIQQ AYIQELLPVS FNNDNSEWIS IVPNFILVRN TLISNIEIGF CLITKRSVIC NQDYATPMTN NMRECLTGST EKCPRELVVS SHVPRFALSN GVLFANCISV TCQCQTTGRA ISQSGEQTLL MIDNTTCPTA VLGNVIISLG KYLGSVNYNS EGIAIGPPVF TDKVDISSQI SSMNQSLQQS KDYIKEAQRL LDTVNPSLIS MLSMIILYVL SIASLCIGLI TFISFIIVEK KRNTGT	尼帕病毒 NiV-F F0 T234 截短(aa 525-544)
6	LAGVIMAGVAIGIATAAQITAGVALYEAMKNADNINKLKSSIESTNEAVVK LQETAEKTVYVLTALQDYINTNLVPTIDKISCKQTELSLDLALSKYLSDLL FVFGPNLQDPVSNSMTIQAISQAFGGNYETLLRTLGYATEDFDDLLESDSI TGQIIYVDLSSYYIIVRVYFPILTEIQQAYIQELLPVSFNNDNSEWISIVP NFILVRNTLISNIEIGFCLITKRSVICNQDYATPMTNNMRECLTGSTEKCP RELVVSSHVPRFALSNGVLFANCISVTCQCQTTGRAISQSGEQTLLMIDNT TCPTAVLGNVIISLGKYLGSVNYNSEGIAIGPPVFTDKVDISSQISSMNQS LQQSKDYIKEAQRLLDTVNPSLISMLSMIILYVLSIASLCIGLITFISFII VEKKRNTGT	尼帕病毒 NiV F F1 (aa 110-546)截短 (aa 525-544)
7	ILHY EKLSKIGLVK GVTRKYKIKS NPLTKDIVIK MIPNVSNMSQ CTGSVMENYK TRLNGILTPI KGALEIYKNQ THDLVGDVRL AGVIMAGVAI GIATAAQITA GVALYEAMKN ADNINKLKSS IESTNEAVVK LQETAEKTVY VLTALQDYIN TNLVPTIDKI SCKQTELSLD LALSKYLSDL LFVFGPNLQD PVSNSMTIQA ISQAFGGNYE TLLRTLGYAT EDFDDLLESD SITGQIIYVD LSSYYIIVRV YFPILTEIQQ AYIQELLPVS FNNDNSEWIS IVPNFILVRN TLISNIEIGF CLITKRSVIC NQDYATPMTN NMRECLTGST EKCPRELVVS SHVPRFALSN GVLFANCISV TCQCQTTGRA ISQSGEQTLL MIDNTTCPTA VLGNVIISLG KYLGSVNYNS EGIAIGPPVF TDKVDISSQI SSMNQSLQQS KDYIKEAQRL LDTVNPSLIS MLSMIILYVL SIASLCIGLI TFISFIIVEK KRNTGT	N-连接糖基化位点上的尼帕病毒 NiV-F F0 T234 截短(aa 525-544)和突变

	8	MVVILDKRCY CNLLILILMI SECSVGILHY EKLSKIGLVK GVTRKYKIKS NPLTKDIVIK MIPNVSNMSQ CTGSVMENYK TRLNGILTPI KGALEIYKNN THDLVGDVRL AGVIMAGVAI GIATAAQITA GVALYEAMKN ADNINKLKSS IESTNEAVVK LQETAEKTVY VLTALQDYIN TNLVPTIDKI SCKQTELSLD LALSKYLSDL LFVFGPNLQD PVSNSMTIQA ISQAFGGNYE TLLRTLGYAT EDFDDLLESD SITGQIIYVD LSSYYIIVRV YFPILTEIQQ AYIQELLPVS FNNDNSEWIS IVPNFILVRN TLISNIEIGF CLITKRSVIC NQDYATPMTN NMRECLTGST EKCPRELVVS SHVPRFALSN GVLFANCISV TCQCQTTGRA ISQSGEQTLL MIDNTTCPTA VLGNVIISLG KYLGSVNYNS EGIAIGPPVF TDKVDISSQI SSMNQSLQQS KDYIKEAQRL LDTVNPSLIS MLSMIILYVL SIASLCIGLI TFISFIIVEK KRNT	在胞质尾截短的 NiV 融合糖蛋白(FcΔ22) (具有信号序列)
	9	MGPAENKKVR FENTTSDKGK IPSKVIKSYY GTMDIKKINE GLLDSKILSA FNTVIALLGS IVIIVMNIMI IQNYTRSTDN QAVIKDALQG IQQQIKGLAD KIGTEIGPKV SLIDTSSTIT IPANIGLLGS KISQSTASIN ENVNEKCKFT LPPLKIHECN ISCPNPLPFR EYRPQTEGVS NLVGLPNNIC LQKTSNQILK PKLISYTLPV VGQSGTCITD PLLAMDEGYF AYSHLERIGS CSRGVSKQRI IGVGEVLDRG DEVPSLFMTN VWTPPNPNTV YHCSAVYNNE FYYVLCAVST VGDPILNSTY WSGSLMMTRL AVKPKSNGGG YNQHQLALRS IEKGRYDKVM PYGPSGIKQG DTLYFPAVGF LVRTEFKYND SNCPITKCQY SKPENCRLSM GIRPNSHYIL RSGLLKYNLS DGENPKVVFI EISDQRLSIG SPSKIYDSLG QPVFYQASFS WDTMIKFGDV LTVNPLVVNW RNNTVISRPG QSQCPRFNTC PEICWEGVYN DAFLIDRINW ISAGVFLDSN QTAENPVFTV FKDNEILYRA QLASEDTNAQ KTITNCFLLK NKIWCISLVE IYDTGDNVIR PKLFAVKIPE QC	NiVG 蛋白附着糖蛋白(602 aa)
[0815]	10	MGKVR FENTTSDKGK IPSKVIKSYY GTMDIKKINE GLLDSKILSA FNTVIALLGS IVIIVMNIMI IQNYTRSTDN QAVIKDALQG IQQQIKGLAD KIGTEIGPKV SLIDTSSTIT IPANIGLLGS KISQSTASIN ENVNEKCKFT LPPLKIHECN ISCPNPLPFR EYRPQTEGVS NLVGLPNNIC LQKTSNQILK PKLISYTLPV VGQSGTCITD PLLAMDEGYF AYSHLERIGS CSRGVSKQRI IGVGEVLDRG DEVPSLFMTN VWTPPNPNTV YHCSAVYNNE FYYVLCAVST VGDPILNSTY WSGSLMMTRL AVKPKSNGGG YNQHQLALRS IEKGRYDKVM PYGPSGIKQG DTLYFPAVGF LVRTEFKYND SNCPITKCQY SKPENCRLSM GIRPNSHYIL RSGLLKYNLS DGENPKVVFI EISDQRLSIG SPSKIYDSLG QPVFYQASFS WDTMIKFGDV LTVNPLVVNW RNNTVISRPG QSQCPRFNTC PEICWEGVYN DAFLIDRINW ISAGVFLDSN QTAENPVFTV FKDNEILYRA QLASEDTNAQ KTITNCFLLK NKIWCISLVE IYDTGDNVIR PKLFAVKIPE QC	NiVG 蛋白附着糖蛋白截短的 Δ5
	11	MGNTTSDKGK IPSKVIKSYY GTMDIKKINE GLLDSKILSA FNTVIALLGS IVIIVMNIMI IQNYTRSTDN QAVIKDALQG IQQQIKGLAD KIGTEIGPKV SLIDTSSTIT IPANIGLLGS KISQSTASIN ENVNEKCKFT LPPLKIHECN ISCPNPLPFR EYRPQTEGVS NLVGLPNNIC LQKTSNQILK PKLISYTLPV VGQSGTCITD PLLAMDEGYF AYSHLERIGS CSRGVSKQRI IGVGEVLDRG DEVPSLFMTN VWTPPNPNTV YHCSAVYNNE FYYVLCAVST VGDPILNSTY WSGSLMMTRL AVKPKSNGGG YNQHQLALRS IEKGRYDKVM PYGPSGIKQG DTLYFPAVGF LVRTEFKYND SNCPITKCQY SKPENCRLSM GIRPNSHYIL RSGLLKYNLS DGENPKVVFI EISDQRLSIG SPSKIYDSLG QPVFYQASFS WDTMIKFGDV LTVNPLVVNW RNNTVISRPG QSQCPRFNTC PEICWEGVYN DAFLIDRINW ISAGVFLDSN QTAENPVFTV FKDNEILYRA QLASEDTNAQ KTITNCFLLK NKIWCISLVE IYDTGDNVIR PKLFAVKIPE QC	NiVG 蛋白附着糖蛋白截短的 Δ10

[0816]

12	MGKGK IPSKVIKSYY GTMDIKKINE GLLDSKILSA FNTVIALLGS IVIIVMNIMI IQNYTRSTDN QAVIKDALQG IQQQIKGLAD KIGTEIGPKV SLIDTSSTIT IPANIGLLGS KISQSTASIN ENVNEKCKFT LPPLKIHECN ISCPNPLPFR EYRPQTEGVS NLVGLPNNIC LQKTSNQILK PKLISYTLPV VGQSGTCITD PLLAMDEGYF AYSHLERIGS CSRGVSKQRI IGVGEVLDRG DEVPSLFMTN VWTPPNPNTV YHCSAVYNNE FYYVLCAVST VGDPILNSTY WSGSLMMTRL AVKPKSNGGG YNQHQLALRS IEKGRYDKVM PYGPSGIKQG DTLYFPAVGF LVRTEFKYND SNCPITKCQY SKPENCRLSM GIRPNSHYIL RSGLLKYNLS DGENPKVVFI EISDQRLSIG SPSKIYDSLG QPVFYQASFS WDTMIKFGDV LTVNPLVVNW RNNTVISRPG QSQCPRFNTC PEICWEGVYN DAFLIDRINW ISAGVFLDSN QTAENPVFTV FKDNEILYRA QLASEDTNAQ KTITNCFLLK NKIWCISLVE IYDTGDNVIR PKLFAVKIPE QC	NiVG 蛋白附着糖蛋白截短的 Δ15
13	MGSKVIKSYY GTMDIKKINE GLLDSKILSA FNTVIALLGS IVIIVMNIMI IQNYTRSTDN QAVIKDALQG IQQQIKGLAD KIGTEIGPKV SLIDTSSTIT IPANIGLLGS KISQSTASIN ENVNEKCKFT LPPLKIHECN ISCPNPLPFR EYRPQTEGVS NLVGLPNNIC LQKTSNQILK PKLISYTLPV VGQSGTCITD PLLAMDEGYF AYSHLERIGS CSRGVSKQRI IGVGEVLDRG DEVPSLFMTN VWTPPNPNTV YHCSAVYNNE FYYVLCAVST VGDPILNSTY WSGSLMMTRL AVKPKSNGGG YNQHQLALRS IEKGRYDKVM PYGPSGIKQG DTLYFPAVGF LVRTEFKYND SNCPITKCQY SKPENCRLSM GIRPNSHYIL RSGLLKYNLS DGENPKVVFI EISDQRLSIG SPSKIYDSLG QPVFYQASFS WDTMIKFGDV LTVNPLVVNW RNNTVISRPG QSQCPRFNTC PEICWEGVYN DAFLIDRINW ISAGVFLDSN QTAENPVFTV FKDNEILYRA QLASEDTNAQ KTITNCFLLK NKIWCISLVE IYDTGDNVIR PKLFAVKIPE QC	NiVG 蛋白附着糖蛋白截短的 Δ20
14	MGSYY GTMDIKKINE GLLDSKILSA FNTVIALLGS IVIIVMNIMI IQNYTRSTDN QAVIKDALQG IQQQIKGLAD KIGTEIGPKV SLIDTSSTIT IPANIGLLGS KISQSTASIN ENVNEKCKFT LPPLKIHECN ISCPNPLPFR EYRPQTEGVS NLVGLPNNIC LQKTSNQILK PKLISYTLPV VGQSGTCITD PLLAMDEGYF AYSHLERIGS CSRGVSKQRI IGVGEVLDRG DEVPSLFMTN VWTPPNPNTV YHCSAVYNNE FYYVLCAVST VGDPILNSTY WSGSLMMTRL AVKPKSNGGG YNQHQLALRS IEKGRYDKVM PYGPSGIKQG DTLYFPAVGF LVRTEFKYND SNCPITKCQY SKPENCRLSM GIRPNSHYIL RSGLLKYNLS DGENPKVVFI EISDQRLSIG SPSKIYDSLG QPVFYQASFS WDTMIKFGDV LTVNPLVVNW RNNTVISRPG QSQCPRFNTC PEICWEGVYN DAFLIDRINW ISAGVFLDSN QTAENPVFTV FKDNEILYRA QLASEDTNAQ KTITNCFLLK NKIWCISLVE IYDTGDNVIR PKLFAVKIPE QC	NiVG 蛋白附着糖蛋白截短的 Δ25
15	MGTMDIKKINE GLLDSKILSA FNTVIALLGS IVIIVMNIMI IQNYTRSTDN QAVIKDALQG IQQQIKGLAD KIGTEIGPKV SLIDTSSTIT IPANIGLLGS KISQSTASIN ENVNEKCKFT LPPLKIHECN ISCPNPLPFR EYRPQTEGVS NLVGLPNNIC LQKTSNQILK PKLISYTLPV VGQSGTCITD PLLAMDEGYF AYSHLERIGS CSRGVSKQRI IGVGEVLDRG DEVPSLFMTN VWTPPNPNTV YHCSAVYNNE FYYVLCAVST VGDPILNSTY WSGSLMMTRL AVKPKSNGGG YNQHQLALRS IEKGRYDKVM PYGPSGIKQG DTLYFPAVGF LVRTEFKYND SNCPITKCQY SKPENCRLSM GIRPNSHYIL RSGLLKYNLS DGENPKVVFI EISDQRLSIG SPSKIYDSLG QPVFYQASFS WDTMIKFGDV LTVNPLVVNW RNNTVISRPG QSQCPRFNTC PEICWEGVYN DAFLIDRINW ISAGVFLDSN QTAENPVFTV FKDNEILYRA QLASEDTNAQ KTITNCFLLK NKIWCISLVE IYDTGDNVIR PKLFAVKIPE QC	NiVG 蛋白附着糖蛋白截短的 Δ30

[0817]

16	MKKINEGLLDSKILSA FNTVIALLGS IVIIVMNIMI IQNYTRSTDN QAVIKDALQG IQQQIKGLAD KIGTEIGPKV SLIDTSSTIT IPANIGLLGS KISQSTASIN ENVNEKCKFT LPPLKIHECN ISCPNPLPFR EYRPQTEGVS NLVGLPNNIC LQKTSNQILK PKLISYTLPV VGQSGTCITD PLLAMDEGYF AYSHLERIGS CSRGVSKQRI IGVGEVLDRG DEVPSLFMTN VWTPPNPNTV YHCSAVYNNE FYYVLCAVST VGDPILNSTY WSGSLMMTRL AVKPKSNGGG YNQHQLALRS IEKGRYDKVM PYGPSGIKQG DTLYFPAVGF LVRTEFKYND SNCPITKCQY SKPENCRLSM GIRPNSHYIL RSGLLKYNLS DGENPKVVFI EISDQRLSIG SPSKIYDSLG QPVFYQASFS WDTMIKFGDV LTVNPLVVNW RNNTVISRPG QSQCPRFNTC PAICAEGVYN DAFLIDRINW ISAGVFLDSN ATAANPVFTV FKDNEILYRA QLASEDTNAQ KTITNCFLLK NKIWCISLVE IYDTGDNVIR PKLFAVKIPE QCT	NiVG 蛋白附着糖蛋白截短的和突变的 (E501A、W504A、Q530A、E533A) NiV G 蛋白(Gc Δ 34)
17	MATQEVRLKC LLCGIIVLVL SLEGLGILHY EKLSKIGLVK GITRKYKIKS NPLTKDIVIK MIPNVSNVSK CTGTVMENYK SRLTGILSPI KGAIELYNNN THDLVGDVKL AGVVMAGIAI GIATAAQITA GVALYEAMKN ADNINKLKSS IESTNEAVVK LQETAEKTVY VLTALQDYIN TNLVPTIDQI SCKQTELALD LALSKYLSDL LFVFGPNLQD PVSNSMTIQA ISQAFGGNYE TLLRTLGYAT EDFDDLLESD SIAGQIVYVD LSSYYIIVRV YFPILTEIQQ AYVQELLPVS FNNDNSEWIS IVPNFVLIRN TLISNIEVKY CLITKKSVIC NQDYATPMTA SVRECLTGST DKCPRELVVS SHVPRFALSG GVLFANCISV TCQCQTTGRA ISQSGEQTLL MIDNTTCTTV VLGNIIISLG KYLGSINYNS ESIAVGPPVY TDKVDISSQI SSMNQSLQQS KDYIKEAQKI LDTVNPSLIS MLSMIILYVL SIAALCIGLI TFISFVIVEK KRGNYSRLDD RQVRPVSNGD LYYIGT	亨德拉病毒 F 蛋白 Uniprot O89342 (具有信号序列)
18	MMADSKLVSL NNNLSGKIKD QGKVIKNYYG TMDIKKINDG LLDSKILGAF NTVIALLGSI IIIVMNIMII QNYTRTTDNQ ALIKESLQSV QQQIKALTDK IGTEIGPKVS LIDTSSTITI PANIGLLGSK ISQSTSSINE NVNDKCKFTL PPLKIHECNI SCPNPLPFRE YRPISQGVSD LVGLPNQICL QKTTSTILKP RLISYTLPIN TREGVCITDP LLAVDNGFFA YSHLEKIGSC TRGIAKQRII GVGEVLDRGD KVPSMFMTNV WTPPNPSTIH HCSSTYHEDF YYTLCAVSHV GDPILNSTSW TESLSLIRLA VRPKSDSGDY NQKYIAITKV ERGKYDKVMP YGPSGIKQGD TLYFPAVGFL PRTEFQYNDS NCPIIHCKYS KAENCRLSMG VNSKSHYILR SGLLKYNLSL GGDIILQFIE IADNRLTIGS PSKIYNSLGQ PVFYQASYSW DTMIKLGDVD TVDPLRVQWR NNSVISRPGQ SQCPRFNVCP EVCWEGTYND AFLIDRLNWV SAGVYLNSNQ TAENPVFAVF KDNEILYQVP LAEDDTNAQK TITDCFLLEN VIWCISLVEI YDTGDSVIRP KLFAVKIPAQ CSES	亨德拉病毒 G 蛋白 Uniprot O89343

	19	MVVILDKRCY CNLLILILMI SECSVGILHY EKLSKIGLVK GVTRKYKIKS NPLTKDIVIK MIPNVSNMSQ CTGSVMENYK TRLNGILTPI KGALEIYKNN THDLVGDVRL AGVIMAGVAI GIATAAQITA GVALYEAMKN ADNINKLKSS IESTNEAVVK LQETAEKTVY VLTALQDYIN TNLVPTIDKI SCKQTELSLD LALSKYLSDL LFVFGPNLQD PVSNSMTIQA ISQAFGGNYE TLLRTLGYAT EDFDDLLESD SITGQIIYVD LSSYYIIVRV YFPILTEIQQ AYIQELLPVS FNNDNSEWIS IVPNFILVRN TLISNIEIGF CLITKRSVIC NQDYATPMTN NMRECLTGST EKCPRELVVS SHVPRFALSN GVLFANCISV TCQCQTTGRA ISQSGEQTLL MIDNTTCPTA VLGNVIISLG KYLGSVNYNS EGIAIGPPVF TDKVDISSQI SSMNQSLQQS KDYIKEAQRL LDTVNPSLIS MLSMIILYVL SIASLCIGLI TFISFIIVEK KRNTGT	尼帕病毒 NiV-F F0 T234 截短(aa 525-544)(具有信号序列)
	20	MVVILDKRCY CNLLILILMI SECSVGILHY EKLSKIGLVK GVTRKYKIKS NPLTKDIVIK MIPNVSNMSQ CTGSVMENYK TRLNGILTPI KGALEIYKNQ THDLVGDVRL AGVIMAGVAI GIATAAQITA GVALYEAMKN ADNINKLKSS IESTNEAVVK LQETAEKTVY VLTALQDYIN TNLVPTIDKI SCKQTELSLD LALSKYLSDL LFVFGPNLQD PVSNSMTIQA ISQAFGGNYE TLLRTLGYAT EDFDDLLESD SITGQIIYVD LSSYYIIVRV YFPILTEIQQ AYIQELLPVS FNNDNSEWIS IVPNFILVRN TLISNIEIGF CLITKRSVIC NQDYATPMTN NMRECLTGST EKCPRELVVS SHVPRFALSN GVLFANCISV TCQCQTTGRA ISQSGEQTLL MIDNTTCPTA VLGNVIISLG KYLGSVNYNS EGIAIGPPVF TDKVDISSQI SSMNQSLQQS KDYIKEAQRL LDTVNPSLIS MLSMIILYVL SIASLCIGLI TFISFIIVEK KRNTGT	N-连接糖基化位点上的尼帕病毒 NiV-F F0 T234 截短(aa 525-544)和突变(具有信号序列)
[0818]	21	MVVILDKRCY CNLLILILMI SECSVGILHY EKLSKIGLVK GVTRKYKIKS NPLTKDIVIK MIPNVSNMSQ CTGSVMENYK TRLNGILTPI KGALEIYKNN THDLVGDVRL AGVIMAGVAI GIATAAQITA GVALYEAMKN ADNINKLKSS IESTNEAVVK LQETAEKTVY VLTALQDYIN TNLVPTIDKI SCKQTELSLD LALSKYLSDL LFVFGPNLQD PVSNSMTIQA ISQAFGGNYE TLLRTLGYAT EDFDDLLESD SITGQIIYVD LSSYYIIVRV YFPILTEIQQ AYIQELLPVS FNNDNSEWIS IVPNFILVRN TLISNIEIGF CLITKRSVIC NQDYATPMTN NMRECLTGST EKCPRELVVS SHVPRFALSN GVLFANCISV TCQCQTTGRA ISQSGEQTLL MIDNTTCPTA VLGNVIISLG KYLGSVNYNS EGIAIGPPVF TDKVDISSQI SSMNQSLQQS KDYIKEAQRL LDTVNPSLIS MLSMIILYVL SIASLCIGLI TFISFIIVEK KRNT	在胞质尾截短的 NiV 融合糖蛋白(FcΔ22) (具有信号序列)
	22	MKKINEGLLDSKILSA FNTVIALLGS IVIIVMNIMI IQNYTRSTDN QAVIKDALQG IQQQIKGLAD KIGTEIGPKV SLIDTSSTIT IPANIGLLGS KISQSTASIN ENVNEKCKFT LPPLKIHECN ISCPNPLPFR EYRPQTEGVS NLVGLPNNIC LQKTSNQILK PKLISYTLPV VGQSGTCITD PLLAMDEGYF AYSHLERIGS CSRGVSKQRI IGVGEVLDRG DEVPSLFMTN VWTPPNPNTV YHCSAVYNNE FYYVLCAVST VGDPILNSTY WSGSLMMTRL AVKPKSNGGG YNQHQLALRS IEKGRYDKVM PYGPSGIKQG DTLYFPAVGF LVRTEFKYND SNCPITKCQY SKPENCRLSM GIRPNSHYIL RSGLLKYNLS DGENPKVVFI EISDQRLSIG SPSKIYDSLG QPVFYQASFS WDTMIKFGDV LTVNPLVVNW RNNTVISRPG QSQCPRFNTC PEICWEGVYN DAFLIDRINW ISAGVFLDSN QTAENPVFTV FKDNEILYRA QLASEDTNAQ KTITNCFLLK NKIWCISLVE IYDTGDNVIR PKLFAVKIPE QCT	NiVG 蛋白附着糖蛋白截短的(Gc Δ 34)

[0819]

23	ILHY EKLSKIGLVK GVTRKYKIKS NPLTKDIVIK MIPNVSNMSQ CTGSVMENYK TRLNGILTPI KGALEIYKNN THDLVGDVRL AGVIMAGVAI GIATAAQITA GVALYEAMKN ADNINKLKSS IESTNEAVVK LQETAEKTVY VLTALQDYIN TNLVPTIDKI SCKQTELSLD LALSKYLSDL LFVFGPNLQD PVSNSMTIQA ISQAFGGNYE TLLRTLGYAT EDFDDLLESD SITGQIIYVD LSSYYIIVRV YFPILTEIQQ AYIQELLPVS FNNDNSEWIS IVPNFILVRN TLISNIEIGF CLITKRSVIC NQDYATPMTN NMRECLTGST EKCPRELVVS SHVPRFALSN GVLFANCISV TCQCQTTGRA ISQSGEQTLL MIDNTTCPTA VLGNVIISLG KYLGSVNYNS EGIAIGPPVF TDKVDISSQI SSMNQSLQQS KDYIKEAQRL LDTVNPSLIS MLSMIILYVL SIASLCIGLI TFISFIIVEK KRNT	在胞质尾截短的成熟 NiV 融合糖蛋白 (FcΔ22)
24	MSNKRTTVLIIISYTLFYLNNAAIVGFDFDKLNKIGVVQGRVLNYKIKGDP MTKDLVLKFIPNIVNITECVREPLSRYNETVRRLLLPIHNMLGLYLNNTNA KMTGLMIAGVIMGGIAIGIATAAQITAGFALYEAKKNTENIQKLTDSIMKT QDSIDKLTDSVGTSILILNKLQTYINNQLVPNLELLSCRQNKIEFDLMLTK YLVDLMTVIGPNINNPVNKDMTIQSLSLLFDGNYDIMMSELGYTPQDFLDL IESKSITGQIIYVDMENLYVVIRTYLPTLIEVPDAQIYEFNKITMSSNGGE YLSTIPNFILIRGNYMSNIDVATCYMTKASVICNQDYSLPMSQNLRSCYQG ETEYCPVEAVIASHSPRFALTNGVIFANCINTICRCQDNGKTITQNINQFV SMIDNSTCNDVMVDKFTIKVGKYMGRKDINNINIQIGPQIIIDKVDLSNEI NKMNQSLKDSIFYLREAKRILDSVNISLISPSVQLFLIIISVLSFIILLII IVYLYCKSKHSYKYNKFIDDPDYYNDYKRERINGKASKSNNIYYVGD	gb:JQ001776:6129-8166|生物体：雪松病毒|毒株名称：CG1a|蛋白质名称：融合糖蛋白|基因符号：F (具有信号序列)
25	MALNKNMFSSLFLGYLLVYATTVQSSIHYDSLSKVGVIKGLTYNYKIKGSP STKLMVVKLIPNIDSVKNCTQKQYDEYKNLVRKALEPVKMAIDTMLNNVKS GNNKYRFAGAIMAGVALGVATAATVTAGIALHRSNENAQAIANMKSAIQNT NEAVKQLQLANKQTLAVIDTIRGEINNNIIPVINQLSCDTIGLSVGIRLTQ YYSEIITAFGPALQNPVNTRITIQAISSVFNGNFDELLKIMGYTSGDLYEI LHSELIRGNIIDVDVDAGYIALEIEFPNLTLVPNAVVQELMPISYNIDGDE WVTLVPRFVLTRTTLLSNIDTSRCTITDSSVICDNDYALPMSHELIGCLQG DTSKCAREKVVSSYVPKFALSDGLVYANCLNTICRCMDTDTPISQSLGATV SLLDNKRCSVYQVGDVLISVGSYLGDGEYNADNVELGPPIVIDKIDIGNQL AGINQTLQEAEDYIEKSEEFLKGVNPSIITLGSMVVLYIFMILIAIVSVIA LVLSIKLTVKGNVVRQQFTYTQHVPSMENINYVSH	gb:NC_025352:5950-8712|生物体：墨江病毒|毒株名称：Tongguan1|蛋白质名称：融合蛋白|基因符号：F (具有信号序列)
26	MKKKTDNPTISKRGHNHSRGIKSRALLRETDNYSNGLIVENLVRNCHHPSK NNLNYTKTQKRDSTIPYRVEERKGHYPKIKHLIDKSYKHIKRGKRRNGHNG NIITIILLLILILKTQMSEGAIHYETLSKIGLIKGITREYKVKGTPSSKDI VIKLIPNVTGLNKCTNISMENYKEQLDKILIPINNIIELYANSTKSAPGNA RFAGVIIAGVALGVAAAAQITAGIALHEARQNAERINLLKDSISATNNAVA ELQEATGGIVNVITGMQDYINTNLVPQIDKLQCSQIKTALDISLSQYYSEI LTVFGPNLQNPVTTSMSIQAISQSFGGNIDLLLNLLGYTANDLLDLLESKS ITGQITYINLEHYFMVIRVYYPIMTTISNAYVQELIKISFNVDGSEWVSLV PSYILIRNSYLSNIDISECLITKNSVICRHDFAMPMSYTLKECLTGDTEKC PREAVVTSYVPRFAISGGVIYANCLSTTCQCYQTGKVIAQDGSQTLMMIDN QTCSIVRIEEILISTGKYLGSQEYNTMHVSVGNPVFTDKLDITSQISNINQ SIEQSKFYLDKSKAILDKINLNLIGSVPISILFIIAILSLILSIITFVIVM IIVRRYNKYTPLINSDPSSRRSTIQDVYIIPNPGEHSIRSAARSIDRDRD	gb:NC_025256:6865-8853|生物体：蝙蝠副粘病毒 Eid_hel/GH-M74a/GHA/2009|毒株名称：BatPV/Eid_hel/GH-M74a/GHA/2009|蛋白质名称：融合蛋白|基因符号：F (具有信号序列)
27	(GGGGGS)n，其中 n 为 1 至 6	肽接头

[0820]

28	MPAENKKVRFENTTSDKGKIPSKVIKSYYGTMDIKKINEGLLDSKILSAFN TVIALLGSIVIIVMNIMIIQNYTRSTDNQAVIKDALQGIQQQIKGLADKIG TEIGPKVSLIDTSSTITIPANIGLLGSKISQSTASINENVNEKCKFTLPPL KIHECNISCPNPLPFREYRPQTEGVSNLVGLPNNICLQKTSNQILKPKLIS YTLPVVGQSGTCITDPLLAMDEGYFAYSHLERIGSCSRGVSKQRIIGVGEV LDRGDEVPSLFMTNVWTPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILN STYWSGSLMMTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIK QGDTLYFPAVGFLVRTEFKYNDSNCPITKCQYSKPENCRLSMGIRPNSHYI LRSGLLKYNLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVFYQASFS WDTMIKFGDVLTVNPLVVNWRNNTVISRPGQSQCPRFNTCPEICWEGVYND AFLIDRINWISAGVFLDSNQTAENPVFTVFKDNEILYRAQLASEDTNAQKT ITNCFLLKNKIWCISLVEIYDTGDNVIRPKLFAVKIPEQCT	gb:AF212302|生物体：尼帕病毒|毒株名称：未知-AF212302|蛋白质名称：附着糖蛋白|基因符号：G (Uniprot Q9IH62)
29	MLSQLQKNYLDNSNQQGDKMNNPDKKLSVNFNPLELDKGQKDLNKSYYVKN KNYNVSNLLNESLHDIKFCIYCIFSLLIIITIINIITISIVITRLKVHEEN NGMESPNLQSIQDSLSSLTNMINTEITPRIGILVTATSVTLSSSINYVGTK TNQLVNELKDYITKSCGFKVPELKLHECNISCADPKISKSAMYSTNAYAEL AGPPKIFCKSVSKDPDFRLKQIDYVIPVQQDRSICMNNPLLDISDGFFTYI HYEGINSCKKSDSFKVLLSHGEIVDRGDYRPSLYLLSSHYHPYSMQVINCV PVTCNQSSFVFCHISNNTKTLDNSDYSSDEYYITYFNGIDRPKTKKIPINN MTADNRYIHFTFSGGGGVCLGEEFIIPVTTVINTDVFTHDYCESFNCSVQT GKSLKEICSESLRSPTNSSRYNLNGIMIISQNNMTDFKIQLNGITYNKLSF GSPGRLSKTLGQVLYYQSSMSWDTYLKAGFVEKWKPFTPNWMNNTVISRPN QGNCPRYHKCPEICYGGTYNDIAPLDLGKDMYVSVILDSDQLAENPEITVF NSTTILYKERVSKDELNTRSTTTSCFLFLDEPWCISVLETNRFNGKSIRPE IYSYKIPKYC	gb:JQ001776:8170-10275|生物体：雪松病毒|毒株名称：CG1a|蛋白质名称：附着糖蛋白|基因符号：G
30	MPQKTVEFINMNSPLERGVSTLSDKKTLNQSKITKQGYFGLGSHSERNWKK QKNQNDHYMTVSTMILEILVVLGIMFNLIVLTMVYYQNDNINQRMAELTSN ITVLNLNLNQLTNKIQREIIPRITLIDTATTITIPSAITYILATLTTRISE LLPSINQKCEFKTPTLVLNDCRINCTPPLNPSDGVKMSSLATNLVAHGPSP CRNFSSVPTIYYYRIPGLYNRTALDERCILNPRLTISSTKFAYVHSEYDKN CTRGFKYYELMTFGEILEGPEKEPRMFSRSFYSPTNAVNYHSCTPIVTVNE GYFLCLECTSSDPLYKANLSNSTFHLVILRHNKDEKIVSMPSFNLSTDQEY VQIIPAEGGGTAESGNLYFPCIGRLLHKRVTHPLCKKSNCSRTDDESCLKS YYNQGSPQHQVVNCLIRIRNAQRDNPTWDVITVDLTNTYPGSRSRIFGSFS KPMLYQSSVSWHTLLQVAEITDLDKYQLDWLDTPYISRPGGSECPFGNYCP TVCWEGTYNDVYSLTPNNDLFVTVYLKSEQVAENPYFAIFSRDQILKEFPL DAWISSARTTTISCFMFNNEIWCIAALEITRLNDDIIRPIYYSFWLPTDCR TPYPHTGKMTRVPLRSTYNY	gb:NC_025256:9117-11015|生物体：蝙蝠副粘病毒Eid_hel/GH-M74a/GHA/2009|毒株名称：BatPV/Eid_hel/GH-M74a/GHA/2009|蛋白质名称：糖蛋白|基因符号：G
31	MATNRDNTITSAEVSQEDKVKKYYGVETAEKVADSISGNKVFILMNTLLIL TGAIITITLNITNLTAAKSQQNMLKIIQDDVNAKLEMFVNLDQLVKGEIKP KVSLINTAVSVSIPGQISNLQTKFLQKYVYLEESITKQCTCNPLSGIFPTS GPTYPPTDKPDDDTTDDDKVDTTIKPIEYPKPDGCNRTGDHFTMEPGANFY TVPNLGPASSNSDECYTNPSFSIGSSIYMFSQEIRKTDCTAGEILSIQIVL GRIVDKGQQGPQASPLLVWAVPNPKIINSCAVAAGDEMGWVLCSVTLTAAS GEPIPHMFDGFWLYKLEPDTEVVSYRITGYAYLLDKQYDSVFIGKGGGIQK GNDLYFQMYGLSRNRQSFKALCEHGSCLGTGGGGYQVLCDRAVMSFGSEES LITNAYLKVNDLASGKPVIIGQTFPPSDSYKGSNGRMYTIGDKYGLYLAPS SWNRYLRFGITPDISVRSTTWLKSQDPIMKILSTCTNTDRDMCPEICNTRG YQDIFPLSEDSEYYTYIGITPNNGGTKNFVAVRDSDGHIASIDILQNYYSI TSATISCFMYKDEIWCIAITEGKKQKDNPQRIYAHSYKIRQMCYNMKSATV TVGNAKNITIRRY	gb:NC_025352:8716-11257|生物体：墨江病毒|毒株名称：Tongguan1|蛋白质名称：附着糖蛋白|基因符号：G

[0821]

32	FNTVIALLGS IVIIVMNIMI IQNYTRSTDN QAVIKDALQG IQQQIKGLAD KIGTEIGPKV SLIDTSSTIT IPANIGLLGS KISQSTASIN ENVNEKCKFT LPPLKIHECN ISCPNPLPFR EYRPQTEGVS NLVGLPNNIC LQKTSNQILK PKLISYTLPV VGQSGTCITD PLLAMDEGYF AYSHLERIGS CSRGVSKQRI IGVGEVLDRG DEVPSLFMTN VWTPPNPNTV YHCSAVYNNE FYYVLCAVST VGDPILNSTY WSGSLMMTRL AVKPKSNGGG YNQHQLALRS IEKGRYDKVM PYGPSGIKQG DTLYFPAVGF LVRTEFKYND SNCPITKCQY SKPENCRLSM GIRPNSHYIL RSGLLKYNLS DGENPKVVFI EISDQRLSIG SPSKIYDSLG QPVFYQASFS WDTMIKFGDV LTVNPLVVNW RNNTVISRPG QSQCPRFNTC PEICWEGVYN DAFLIDRINW ISAGVFLDSN QTAENPVFTV FKDNEILYRA QLASEDTNAQ KTITNCFLLK NKIWCISLVE IYDTGDNVIR PKLFAVKIPE QC	NivG 蛋白附着糖蛋白无胞质尾 Uniprot Q9IH62
33	FNTVIALLGSI IIIVMNIMII QNYTRTTDNQ ALIKESLQSV QQQIKALTDK IGTEIGPKVS LIDTSSTITI PANIGLLGSK ISQSTSSINE NVNDKCKFTL PPLKIHECNI SCPNPLPFRE YRPISQGVSD LVGLPNQICL QKTTSTILKP RLISYTLPIN TREGVCITDP LLAVDNGFFA YSHLEKIGSC TRGIAKQRII GVGEVLDRGD KVPSMFMTNV WTPPNPSTIH HCSSTYHEDF YYTLCAVSHV GDPILNSTSW TESLSLIRLA VRPKSDSGDY NQKYIAITKV ERGKYDKVMP YGPSGIKQGD TLYFPAVGFL PRTEFQYNDS NCPIIHCKYS KAENCRLSMG VNSKSHYILR SGLLKYNLSL GGDIILQFIE IADNRLTIGS PSKIYNSLGQ PVFYQASYSW DTMIKLGDVD TVDPLRVQWR NNSVISRPGQ SQCPRFNVCP EVCWEGTYND AFLIDRLNWV SAGVYLNSNQ TAENPVFAVF KDNEILYQVP LAEDDTNAQK TITDCFLLEN VIWCISLVEI YDTGDSVIRP KLFAVKIPAQ CSES	亨德拉病毒 G 蛋白 Uniprot O89343 无胞质尾
34	MVVILDKRCY CNLLILILMI SECSVG	信号序列
35	MKVR FENTTSDKGK IPSKVIKSYY GTMDIKKINE GLLDSKILSA FNTVIALLGS IVIIVMNIMI IQNYTRSTDN QAVIKDALQG IQQQIKGLAD KIGTEIGPKV SLIDTSSTIT IPANIGLLGS KISQSTASIN ENVNEKCKFT LPPLKIHECN ISCPNPLPFR EYRPQTEGVS NLVGLPNNIC LQKTSNQILK PKLISYTLPV VGQSGTCITD PLLAMDEGYF AYSHLERIGS CSRGVSKQRI IGVGEVLDRG DEVPSLFMTN VWTPPNPNTV YHCSAVYNNE FYYVLCAVST VGDPILNSTY WSGSLMMTRL AVKPKSNGGG YNQHQLALRS IEKGRYDKVM PYGPSGIKQG DTLYFPAVGF LVRTEFKYND SNCPITKCQY SKPENCRLSM GIRPNSHYIL RSGLLKYNLS DGENPKVVFI EISDQRLSIG SPSKIYDSLG QPVFYQASFS WDTMIKFGDV LTVNPLVVNW RNNTVISRPG QSQCPRFNTC PEICWEGVYN DAFLIDRINW ISAGVFLDSN QTAENPVFTV FKDNEILYRA QLASEDTNAQ KTITNCFLLK NKIWCISLVE IYDTGDNVIR PKLFAVKIPE QCT	NiVG 蛋白附着糖蛋白截短的 Δ5

[0822]

36	MNTTSDKGK IPSKVIKSYY GTMDIKKINE GLLDSKILSA FNTVIALLGS IVIIVMNIMI IQNYTRSTDN QAVIKDALQG IQQQIKGLAD KIGTEIGPKV SLIDTSSTIT IPANIGLLGS KISQSTASIN ENVNEKCKFT LPPLKIHECN ISCPNPLPFR EYRPQTEGVS NLVGLPNNIC LQKTSNQILK PKLISYTLPV VGQSGTCITD PLLAMDEGYF AYSHLERIGS CSRGVSKQRI IGVGEVLDRG DEVPSLFMTN VWTPPNPNTV YHCSAVYNNE FYYVLCAVST VGDPILNSTY WSGSLMMTRL AVKPKSNGGG YNQHQLALRS IEKGRYDKVM PYGPSGIKQG DTLYFPAVGF LVRTEFKYND SNCPITKCQY SKPENCRLSM GIRPNSHYIL RSGLLKYNLS DGENPKVVFI EISDQRLSIG SPSKIYDSLG QPVFYQASFS WDTMIKFGDV LTVNPLVVNW RNNTVISRPG QSQCPRFNTC PEICWEGVYN DAFLIDRINW ISAGVFLDSN QTAENPVFTV FKDNEILYRA QLASEDTNAQ KTITNCFLLK NKIWCISLVE IYDTGDNVIR PKLFAVKIPE QCT	NiVG 蛋白附着糖蛋白截短的 Δ10
37	MKGK IPSKVIKSYY GTMDIKKINE GLLDSKILSA FNTVIALLGS IVIIVMNIMI IQNYTRSTDN QAVIKDALQG IQQQIKGLAD KIGTEIGPKV SLIDTSSTIT IPANIGLLGS KISQSTASIN ENVNEKCKFT LPPLKIHECN ISCPNPLPFR EYRPQTEGVS NLVGLPNNIC LQKTSNQILK PKLISYTLPV VGQSGTCITD PLLAMDEGYF AYSHLERIGS CSRGVSKQRI IGVGEVLDRG DEVPSLFMTN VWTPPNPNTV YHCSAVYNNE FYYVLCAVST VGDPILNSTY WSGSLMMTRL AVKPKSNGGG YNQHQLALRS IEKGRYDKVM PYGPSGIKQG DTLYFPAVGF LVRTEFKYND SNCPITKCQY SKPENCRLSM GIRPNSHYIL RSGLLKYNLS DGENPKVVFI EISDQRLSIG SPSKIYDSLG QPVFYQASFS WDTMIKFGDV LTVNPLVVNW RNNTVISRPG QSQCPRFNTC PEICWEGVYN DAFLIDRINW ISAGVFLDSN QTAENPVFTV FKDNEILYRA QLASEDTNAQ KTITNCFLLK NKIWCISLVE IYDTGDNVIR PKLFAVKIPE QCT	NiVG 蛋白附着糖蛋白截短的 Δ15
38	MSKVIKSYY GTMDIKKINE GLLDSKILSA FNTVIALLGS IVIIVMNIMI IQNYTRSTDN QAVIKDALQG IQQQIKGLAD KIGTEIGPKV SLIDTSSTIT IPANIGLLGS KISQSTASIN ENVNEKCKFT LPPLKIHECN ISCPNPLPFR EYRPQTEGVS NLVGLPNNIC LQKTSNQILK PKLISYTLPV VGQSGTCITD PLLAMDEGYF AYSHLERIGS CSRGVSKQRI IGVGEVLDRG DEVPSLFMTN VWTPPNPNTV YHCSAVYNNE FYYVLCAVST VGDPILNSTY WSGSLMMTRL AVKPKSNGGG YNQHQLALRS IEKGRYDKVM PYGPSGIKQG DTLYFPAVGF LVRTEFKYND SNCPITKCQY SKPENCRLSM GIRPNSHYIL RSGLLKYNLS DGENPKVVFI EISDQRLSIG SPSKIYDSLG QPVFYQASFS WDTMIKFGDV LTVNPLVVNW RNNTVISRPG QSQCPRFNTC PEICWEGVYN DAFLIDRINW ISAGVFLDSN QTAENPVFTV FKDNEILYRA QLASEDTNAQ KTITNCFLLK NKIWCISLVE IYDTGDNVIR PKLFAVKIPE QCT	NiVG 蛋白附着糖蛋白截短的 Δ20
39	MSYY GTMDIKKINE GLLDSKILSA FNTVIALLGS IVIIVMNIMI IQNYTRSTDN QAVIKDALQG IQQQIKGLAD KIGTEIGPKV SLIDTSSTIT IPANIGLLGS KISQSTASIN ENVNEKCKFT LPPLKIHECN ISCPNPLPFR EYRPQTEGVS NLVGLPNNIC LQKTSNQILK PKLISYTLPV VGQSGTCITD PLLAMDEGYF AYSHLERIGS CSRGVSKQRI IGVGEVLDRG DEVPSLFMTN VWTPPNPNTV YHCSAVYNNE FYYVLCAVST VGDPILNSTY WSGSLMMTRL AVKPKSNGGG YNQHQLALRS IEKGRYDKVM PYGPSGIKQG DTLYFPAVGF LVRTEFKYND SNCPITKCQY SKPENCRLSM GIRPNSHYIL RSGLLKYNLS DGENPKVVFI EISDQRLSIG SPSKIYDSLG QPVFYQASFS WDTMIKFGDV LTVNPLVVNW RNNTVISRPG QSQCPRFNTC PEICWEGVYN DAFLIDRINW ISAGVFLDSN QTAENPVFTV FKDNEILYRA QLASEDTNAQ KTITNCFLLK NKIWCISLVE IYDTGDNVIR PKLFAVKIPE QCT	NiVG 蛋白附着糖蛋白截短的 Δ25

[0823]

40	MTMDIKKINE GLLDSKILSA FNTVIALLGS IVIIVMNIMI IQNYTRSTDN QAVIKDALQG IQQQIKGLAD KIGTEIGPKV SLIDTSSTIT IPANIGLLGS KISQSTASIN ENVNEKCKFT LPPLKIHECN ISCPNPLPFR EYRPQTEGVS NLVGLPNNIC LQKTSNQILK PKLISYTLPV VGQSGTCITD PLLAMDEGYF AYSHLERIGS CSRGVSKQRI IGVGEVLDRG DEVPSLFMTN VWTPPNPNTV YHCSAVYNNE FYYVLCAVST VGDPILNSTY WSGSLMMTRL AVKPKSNGGG YNQHQLALRS IEKGRYDKVM PYGPSGIKQG DTLYFPAVGF LVRTEFKYND SNCPITKCQY SKPENCRLSM GIRPNSHYIL RSGLLKYNLS DGENPKVVFI EISDQRLSIG SPSKIYDSLG QPVFYQASFS WDTMIKFGDV LTVNPLVVNW RNNTVISRPG QSQCPRFNTC PEICWEGVYN DAFLIDRINW ISAGVFLDSN QTAENPVFTV FKDNEILYRA QLASEDTNAQ KTITNCFLLK NKIWCISLVE IYDTGDNVIR PKLFAVKIPE QCT	NiVG 蛋白附着糖蛋白截短的 Δ30
41	GGGGGS	肽接头
42	(GGGGS)n，其中 n 为 1 至 10	肽接头
43	GGGGS	肽接头
44	PAENKKVR FENTTSDKGK IPSKVIKSYY GTMDIKKINE GLLDSKILSA FNTVIALLGS IVIIVMNIMI IQNYTRSTDN QAVIKDALQG IQQQIKGLAD KIGTEIGPKV SLIDTSSTIT IPANIGLLGS KISQSTASIN ENVNEKCKFT LPPLKIHECN ISCPNPLPFR EYRPQTEGVS NLVGLPNNIC LQKTSNQILK PKLISYTLPV VGQSGTCITD PLLAMDEGYF AYSHLERIGS CSRGVSKQRI IGVGEVLDRG DEVPSLFMTN VWTPPNPNTV YHCSAVYNNE FYYVLCAVST VGDPILNSTY WSGSLMMTRL AVKPKSNGGG YNQHQLALRS IEKGRYDKVM PYGPSGIKQG DTLYFPAVGF LVRTEFKYND SNCPITKCQY SKPENCRLSM GIRPNSHYIL RSGLLKYNLS DGENPKVVFI EISDQRLSIG SPSKIYDSLG QPVFYQASFS WDTMIKFGDV LTVNPLVVNW RNNTVISRPG QSQCPRFNTC PEICWEGVYN DAFLIDRINW ISAGVFLDSN QTAENPVFTV FKDNEILYRA QLASEDTNAQ KTITNCFLLK NKIWCISLVE IYDTGDNVIR PKLFAVKIPE QC	NiVG 蛋白附着糖蛋白(602 aa) 没有 N 末端甲硫氨酸
45	KVR FENTTSDKGK IPSKVIKSYY GTMDIKKINE GLLDSKILSA FNTVIALLGS IVIIVMNIMI IQNYTRSTDN QAVIKDALQG IQQQIKGLAD KIGTEIGPKV SLIDTSSTIT IPANIGLLGS KISQSTASIN ENVNEKCKFT LPPLKIHECN ISCPNPLPFR EYRPQTEGVS NLVGLPNNIC LQKTSNQILK PKLISYTLPV VGQSGTCITD PLLAMDEGYF AYSHLERIGS CSRGVSKQRI IGVGEVLDRG DEVPSLFMTN VWTPPNPNTV YHCSAVYNNE FYYVLCAVST VGDPILNSTY WSGSLMMTRL AVKPKSNGGG YNQHQLALRS IEKGRYDKVM PYGPSGIKQG DTLYFPAVGF LVRTEFKYND SNCPITKCQY SKPENCRLSM GIRPNSHYIL RSGLLKYNLS DGENPKVVFI EISDQRLSIG SPSKIYDSLG QPVFYQASFS WDTMIKFGDV LTVNPLVVNW RNNTVISRPG QSQCPRFNTC PEICWEGVYN DAFLIDRINW ISAGVFLDSN QTAENPVFTV FKDNEILYRA QLASEDTNAQ KTITNCFLLK NKIWCISLVE IYDTGDNVIR PKLFAVKIPE QC	NiVG 蛋白附着糖蛋白截短的 Δ5 没有 N 末端甲硫氨酸

[0824]

46	NTTSDKGK IPSKVIKSYY GTMDIKKINE GLLDSKILSA FNTVIALLGS IVIIVMNIMI IQNYTRSTDN QAVIKDALQG IQQQIKGLAD KIGTEIGPKV SLIDTSSTIT IPANIGLLGS KISQSTASIN ENVNEKCKFT LPPLKIHECN ISCPNPLPFR EYRPQTEGVS NLVGLPNNIC LQKTSNQILK PKLISYTLPV VGQSGTCITD PLLAMDEGYF AYSHLERIGS CSRGVSKQRI IGVGEVLDRG DEVPSLFMTN VWTPPNPNTV YHCSAVYNNE FYYVLCAVST VGDPILNSTY WSGSLMMTRL AVKPKSNGGG YNQHQLALRS IEKGRYDKVM PYGPSGIKQG DTLYFPAVGF LVRTEFKYND SNCPITKCQY SKPENCRLSM GIRPNSHYIL RSGLLKYNLS DGENPKVVFI EISDQRLSIG SPSKIYDSLG QPVFYQASFS WDTMIKFGDV LTVNPLVVNW RNNTVISRPG QSQCPRFNTC PEICWEGVYN DAFLIDRINW ISAGVFLDSN QTAENPVFTV FKDNEILYRA QLASEDTNAQ KTITNCFLLK NKIWCISLVE IYDTGDNVIR PKLFAVKIPE QC	NiVG 蛋白附着糖蛋白截短的 Δ10 没有 N 末端甲硫氨酸
47	KGK IPSKVIKSYY GTMDIKKINE GLLDSKILSA FNTVIALLGS IVIIVMNIMI IQNYTRSTDN QAVIKDALQG IQQQIKGLAD KIGTEIGPKV SLIDTSSTIT IPANIGLLGS KISQSTASIN ENVNEKCKFT LPPLKIHECN ISCPNPLPFR EYRPQTEGVS NLVGLPNNIC LQKTSNQILK PKLISYTLPV VGQSGTCITD PLLAMDEGYF AYSHLERIGS CSRGVSKQRI IGVGEVLDRG DEVPSLFMTN VWTPPNPNTV YHCSAVYNNE FYYVLCAVST VGDPILNSTY WSGSLMMTRL AVKPKSNGGG YNQHQLALRS IEKGRYDKVM PYGPSGIKQG DTLYFPAVGF LVRTEFKYND SNCPITKCQY SKPENCRLSM GIRPNSHYIL RSGLLKYNLS DGENPKVVFI EISDQRLSIG SPSKIYDSLG QPVFYQASFS WDTMIKFGDV LTVNPLVVNW RNNTVISRPG QSQCPRFNTC PEICWEGVYN DAFLIDRINW ISAGVFLDSN QTAENPVFTV FKDNEILYRA QLASEDTNAQ KTITNCFLLK NKIWCISLVE IYDTGDNVIR PKLFAVKIPE QC	NiVG 蛋白附着糖蛋白截短的 Δ15 没有 N 末端甲硫氨酸
48	SKVIKSYY GTMDIKKINE GLLDSKILSA FNTVIALLGS IVIIVMNIMI IQNYTRSTDN QAVIKDALQG IQQQIKGLAD KIGTEIGPKV SLIDTSSTIT IPANIGLLGS KISQSTASIN ENVNEKCKFT LPPLKIHECN ISCPNPLPFR EYRPQTEGVS NLVGLPNNIC LQKTSNQILK PKLISYTLPV VGQSGTCITD PLLAMDEGYF AYSHLERIGS CSRGVSKQRI IGVGEVLDRG DEVPSLFMTN VWTPPNPNTV YHCSAVYNNE FYYVLCAVST VGDPILNSTY WSGSLMMTRL AVKPKSNGGG YNQHQLALRS IEKGRYDKVM PYGPSGIKQG DTLYFPAVGF LVRTEFKYND SNCPITKCQY SKPENCRLSM GIRPNSHYIL RSGLLKYNLS DGENPKVVFI EISDQRLSIG SPSKIYDSLG QPVFYQASFS WDTMIKFGDV LTVNPLVVNW RNNTVISRPG QSQCPRFNTC PEICWEGVYN DAFLIDRINW ISAGVFLDSN QTAENPVFTV FKDNEILYRA QLASEDTNAQ KTITNCFLLK NKIWCISLVE IYDTGDNVIR PKLFAVKIPE QC	NiVG 蛋白附着糖蛋白截短的 Δ20 没有 N 末端甲硫氨酸
49	SYY GTMDIKKINE GLLDSKILSA FNTVIALLGS IVIIVMNIMI IQNYTRSTDN QAVIKDALQG IQQQIKGLAD KIGTEIGPKV SLIDTSSTIT IPANIGLLGS KISQSTASIN ENVNEKCKFT LPPLKIHECN ISCPNPLPFR EYRPQTEGVS NLVGLPNNIC LQKTSNQILK PKLISYTLPV VGQSGTCITD PLLAMDEGYF AYSHLERIGS CSRGVSKQRI IGVGEVLDRG DEVPSLFMTN VWTPPNPNTV YHCSAVYNNE FYYVLCAVST VGDPILNSTY WSGSLMMTRL AVKPKSNGGG YNQHQLALRS IEKGRYDKVM PYGPSGIKQG DTLYFPAVGF LVRTEFKYND SNCPITKCQY SKPENCRLSM GIRPNSHYIL RSGLLKYNLS DGENPKVVFI EISDQRLSIG SPSKIYDSLG QPVFYQASFS WDTMIKFGDV LTVNPLVVNW RNNTVISRPG QSQCPRFNTC PEICWEGVYN DAFLIDRINW ISAGVFLDSN QTAENPVFTV FKDNEILYRA QLASEDTNAQ KTITNCFLLK NKIWCISLVE IYDTGDNVIR PKLFAVKIPE QC	NiVG 蛋白附着糖蛋白截短的 Δ25 没有 N 末端甲硫氨酸

	50	TMDIKKINE GLLDSKILSA FNTVIALLGS IVIIVMNIMI IQNYTRSTDN QAVIKDALQG IQQQIKGLAD KIGTEIGPKV SLIDTSSTIT IPANIGLLGS KISQSTASIN ENVNEKCKFT LPPLKIHECN ISCPNPLPFR EYRPQTEGVS NLVGLPNNIC LQKTSNQILK PKLISYTLPV VGQSGTCITD PLLAMDEGYF AYSHLERIGS CSRGVSKQRI IGVGEVLDRG DEVPSLFMTN VWTPPNPNTV YHCSAVYNNE FYYVLCAVST VGDPILNSTY WSGSLMMTRL AVKPKSNGGG YNQHQLALRS IEKGRYDKVM PYGPSGIKQG DTLYFPAVGF LVRTEFKYND SNCPITKCQY SKPENCRLSM GIRPNSHYIL RSGLLKYNLS DGENPKVVFI EISDQRLSIG SPSKIYDSLG QPVFYQASFS WDTMIKFGDV LTVNPLVVNW RNNTVISRPG QSQCPRFNTC PEICWEGVYN DAFLIDRINW ISAGVFLDSN QTAENPVFTV FKDNEILYRA QLASEDTNAQ KTITNCFLLK NKIWCISLVE IYDTGDNVIR PKLFAVKIPE QC	NiVG 蛋白附着糖蛋白截短的 Δ30 没有 N 末端甲硫氨酸
	51	KKINEGLLDSKILSA FNTVIALLGS IVIIVMNIMI IQNYTRSTDN QAVIKDALQG IQQQIKGLAD KIGTEIGPKV SLIDTSSTIT IPANIGLLGS KISQSTASIN ENVNEKCKFT LPPLKIHECN ISCPNPLPFR EYRPQTEGVS NLVGLPNNIC LQKTSNQILK PKLISYTLPV VGQSGTCITD PLLAMDEGYF AYSHLERIGS CSRGVSKQRI IGVGEVLDRG DEVPSLFMTN VWTPPNPNTV YHCSAVYNNE FYYVLCAVST VGDPILNSTY WSGSLMMTRL AVKPKSNGGG YNQHQLALRS IEKGRYDKVM PYGPSGIKQG DTLYFPAVGF LVRTEFKYND SNCPITKCQY SKPENCRLSM GIRPNSHYIL RSGLLKYNLS DGENPKVVFI EISDQRLSIG SPSKIYDSLG QPVFYQASFS WDTMIKFGDV LTVNPLVVNW RNNTVISRPG QSQCPRFNTC PAICAEGVYN DAFLIDRINW ISAGVFLDSN ATAANPVFTV FKDNEILYRA QLASEDTNAQ KTITNCFLLK NKIWCISLVE IYDTGDNVIR PKLFAVKIPE QCT	NiVG 蛋白附着糖蛋白截短的和突变的 (E501A、W504A、Q530A、E533A) NiV G 蛋白(Gc Δ 34)没有 N 末端甲硫氨酸
[0825]	52	MADSKLVSL NNNLSGKIKD QGKVIKNYYG TMDIKKINDG LLDSKILGAF NTVIALLGSI IIIVMNIMII QNYTRTTDNQ ALIKESLQSV QQQIKALTDK IGTEIGPKVS LIDTSSTITI PANIGLLGSK ISQSTSSINE NVNDKCKFTL PPLKIHECNI SCPNPLPFRE YRPISQGVSD LVGLPNQICL QKTTSTILKP RLISYTLPIN TREGVCITDP LLAVDNGFFA YSHLEKIGSC TRGIAKQRII GVGEVLDRGD KVPSMFMTNV WTPPNPSTIH HCSSTYHEDF YYTLCAVSHV GDPILNSTSW TESLSLIRLA VRPKSDSGDY NQKYIAITKV ERGKYDKVMP YGPSGIKQGD TLYFPAVGFL PRTEFQYNDS NCPIIHCKYS KAENCRLSMG VNSKSHYILR SGLLKYNLSL GGDIILQFIE IADNRLTIGS PSKIYNSLGQ PVFYQASYSW DTMIKLGDVD TVDPLRVQWR NNSVISRPGQ SQCPRFNVCP EVCWEGTYND AFLIDRLNWV SAGVYLNSNQ TAENPVFAVF KDNEILYQVP LAEDDTNAQK TITDCFLLEN VIWCISLVEI YDTGDSVIRP KLFAVKIPAQ CSES	亨德拉病毒 G 蛋白 Uniprot O89343 没有 N 末端甲硫氨酸
	53	KKINEGLLDSKILSA FNTVIALLGS IVIIVMNIMI IQNYTRSTDN QAVIKDALQG IQQQIKGLAD KIGTEIGPKV SLIDTSSTIT IPANIGLLGS KISQSTASIN ENVNEKCKFT LPPLKIHECN ISCPNPLPFR EYRPQTEGVS NLVGLPNNIC LQKTSNQILK PKLISYTLPV VGQSGTCITD PLLAMDEGYF AYSHLERIGS CSRGVSKQRI IGVGEVLDRG DEVPSLFMTN VWTPPNPNTV YHCSAVYNNE FYYVLCAVST VGDPILNSTY WSGSLMMTRL AVKPKSNGGG YNQHQLALRS IEKGRYDKVM PYGPSGIKQG DTLYFPAVGF LVRTEFKYND SNCPITKCQY SKPENCRLSM GIRPNSHYIL RSGLLKYNLS DGENPKVVFI EISDQRLSIG SPSKIYDSLG QPVFYQASFS WDTMIKFGDV LTVNPLVVNW	NiVG 蛋白附着糖蛋白截短的(Gc Δ 34)没有 N 末端甲硫氨酸

[0826]

	RNNTVISRPG QSQCPRFNTC PEICWEGVYN DAFLIDRINW ISAGVFLDSN QTAENPVFTV FKDNEILYRA QLASEDTNAQ KTITNCFLLK NKIWCISLVE IYDTGDNVIR PKLFAVKIPE QCT	
54	LSQLQKNYLDNSNQQGDKMNNPDKKLSVNFNPLELDKGQKDLNKSYYVKNK NYNVSNLLNESLHDIKFCIYCIFSLLIIITIINIITISIVITRLKVHEENN GMESPNLQSIQDSLSSLTNMINTEITPRIGILVTATSVTLSSSINYVGTKT NQLVNELKDYITKSCGFKVPELKLHECNISCADPKISKSAMYSTNAYAELA GPPKIFCKSVSKDPDFRLKQIDYVIPVQQDRSICMNNPLLDISDGFFTYIH YEGINSCKKSDSFKVLLSHGEIVDRGDYRPSLYLLSSHYHPYSMQVINCVP VTCNQSSFVFCHISNNTKTLDNSDYSSDEYYITYFNGIDRPKTKKIPINNM TADNRYIHFTFSGGGGVCLGEEFIIPVTTVINTDVFTHDYCESFNCSVQTG KSLKEICSESLRSPTNSSRYNLNGIMIISQNNMTDFKIQLNGITYNKLSFG SPGRLSKTLGQVLYYQSSMSWDTYLKAGFVEKWKPFTPNWMNNTVISRPNQ GNCPRYHKCPEICYGGTYNDIAPLDLGKDMYVSVILDSDQLAENPEITVFN STTILYKERVSKDELNTRSTTTSCFLFLDEPWCISVLETNRFNGKSIRPEI YSYKIPKYC	gb:JQ001776:8170-10275|生物体：雪松病毒|毒株名称：CG1a|蛋白质名称：附着糖蛋白|基因符号：G 没有 N 末端甲硫氨酸
55	PQKTVEFINMNSPLERGVSTLSDKKTLNQSKITKQGYFGLGSHSERNWKKQ KNQNDHYMTVSTMILEILVVLGIMFNLIVLTMVYYQNDNINQRMAELTSNI TVLNLNLNQLTNKIQREIIPRITLIDTATTITIPSAITYILATLTTRISEL LPSINQKCEFKTPTLVLNDCRINCTPPLNPSDGVKMSSLATNLVAHGPSPC RNFSSVPTIYYYRIPGLYNRTALDERCILNPRLTISSTKFAYVHSEYDKNC TRGFKYYELMTFGEILEGPEKEPRMFSRSFYSPTNAVNYHSCTPIVTVNEG YFLCLECTSSDPLYKANLSNSTFHLVILRHNKDEKIVSMPSFNLSTDQEYV QIIPAEGGGTAESGNLYFPCIGRLLHKRVTHPLCKKSNCSRTDDESCLKSY YNQGSPQHQVVNCLIRIRNAQRDNPTWDVITVDLTNTYPGSRSRIFGSFSK PMLYQSSVSWHTLLQVAEITDLDKYQLDWLDTPYISRPGGSECPFGNYCPT VCWEGTYNDVYSLTPNNDLFVTVYLKSEQVAENPYFAIFSRDQILKEFPLD AWISSARTTTISCFMFNNEIWCIAALEITRLNDDIIRPIYYSFWLPTDCRT PYPHTGKMTRVPLRSTYNY	gb:NC_025256:9117-11015|生物体：蝙蝠副粘病毒 Eid_hel/GH-M74a/GHA/2009|毒株名称：BatPV/Eid_hel/GH-M74a/GHA/2009|蛋白质名称：糖蛋白|基因符号：G 没有 N 末端甲硫氨酸
56	ATNRDNTITSAEVSQEDKVKKYYGVETAEKVADSISGNKVFILMNTLLILT GAIITITLNITNLTAAKSQQNMLKIIQDDVNAKLEMFVNLDQLVKGEIKPK VSLINTAVSVSIPGQISNLQTKFLQKYVYLEESITKQCTCNPLSGIFPTSG PTYPPTDKPDDDTTDDDKVDTTIKPIEYPKPDGCNRTGDHFTMEPGANFYT VPNLGPASSNSDECYTNPSFSIGSSIYMFSQEIRKTDCTAGEILSIQIVLG RIVDKGQQGPQASPLLVWAVPNPKIINSCAVAAGDEMGWVLCSVTLTAASG EPIPHMFDGFWLYKLEPDTEVVSYRITGYAYLLDKQYDSVFIGKGGGIQKG NDLYFQMYGLSRNRQSFKALCEHGSCLGTGGGGYQVLCDRAVMSFGSEESL ITNAYLKVNDLASGKPVIIGQTFPPSDSYKGSNGRMYTIGDKYGLYLAPSS WNRYLRFGITPDISVRSTTWLKSQDPIMKILSTCTNTDRDMCPEICNTRGY QDIFPLSEDSEYYTYIGITPNNGGTKNFVAVRDSDGHIASIDILQNYYSIT SATISCFMYKDEIWCIAITEGKKQKDNPQRIYAHSYKIRQMCYNMKSATVT VGNAKNITIRRY	gb:NC_025352:8716-11257|生物体：墨江病毒|毒株名称：Tongguan1|蛋白质名称：附着糖蛋白|基因符号：G 没有 N 末端甲硫氨酸
57	DFDKLNKIGVVQGRVLNYKIKGDPMTKDLVLKFIPNIVNITECVREPLSRY NETVRRLLLPIHNMLGLYLNNTNAKMTGLMIAGVIMGGIAIGIATAAQITA GFALYEAKKNTENIQKLTDSIMKTQDSIDKLTDSVGTSILILNKLQTYINN QLVPNLELLSCRQNKIEFDLMLTKYLVDLMTVIGPNINNPVNKDMTIQSLS LLFDGNYDIMMSELGYTPQDFLDLIESKSITGQIIYVDMENLYVVIRTYLP TLIEVPDAQIYEFNKITMSSNGGEYLSTIPNFILIRGNYMSNIDVATCYMT KASVICNQDYSLPMSQNLRSCYQGETEYCPVEAVIASHSPRFALTNGVIFA NCINTICRCQDNGKTITQNINQFVSMIDNSTCNDVMVDKFTIKVGKYMGRK DINNINIQIGPQIIIDKVDLSNEINKMNQSLKDSIFYLREAKRILDSVNIS LISPSVQLFLIIISVLSFIILLIIIVYLYCKSKHSYKYNKFIDDPDYYNDY KRERINGKASKSNNIYYVGD	gb:JQ001776:6129-8166|生物体：雪松病毒|毒株名称：CG1a|蛋白质名称：融合糖蛋白|基因符号：F (没有信号序列)

[0827]

58	SRALLRETDNYSNGLIVENLVRNCHHPSKNNLNYTKTQKRDSTIPYRVEER KGHYPKIKHLIDKSYKHIKRGKRRNGHNGNIITIILLLLILILKTQMSEGAI HYETLSKIGLIKGITREYKVKGTPSSKDIVIKLIPNVTGLNKCTNISMENY KEQLDKILIPINNIIELYANSTKSAPGNARFAGVIIAGVALGVAAAAQITA GIALHEARQNAERINLLKDSISATNNAVAELQEATGGIVNVITGMQDYINT NLVPQIDKLQCSQIKTALDISLSQYYSEILTVFGPNLQNPVTTSMSIQAIS QSFGGNIDLLLNLLGYTANDLLDLLESKSITGQITYINLEHYFMVIRVYYP IMTTISNAYVQELIKISFNVDGSEWVSLVPSYILIRNSYLSNIDISECLIT KNSVICRHDFAMPMSYTLKECLTGDTEKCPREAVVTSYVPRFAISGGVIYA NCLSTTCQCYQTGKVIAQDGSQTLMMIDNQTCSIVRIEEILISTGKYLGSQ EYNTMHVSVGNPVFTDKLDITSQISNINQSIEQSKFYLDKSKAILDKINLN LIGSVPISILFIIAILSLILSIITFVIVMIIVRRYNKYTPLINSDPSSRRS TIQDVYIIPNPGEHSIRSAARSIDRDRD	gb:NC_025256:6865-8853|生物体：蝙蝠副粘病毒Eid_hel/GH-M74a/GHA/2009|毒株名称：BatPV/Eid_hel/GH-M74a/GHA/2009|蛋白质名称：融合蛋白|基因符号：F (没有信号序列)
59	ILHY EKLSKIGLVK GITRKYKIKS NPLTKDIVIK MIPNVSNVSK CTGTVMENYK SRLTGILSPI KGAIELYNNN THDLVGDVKL AGVVMAGIAI GIATAAQITA GVALYEAMKN ADNINKLKSS IESTNEAVVK LQETAEKTVY VLTALQDYIN TNLVPTIDQI SCKQTELALD LALSKYLSDL LFVFGPNLQD PVSNSMTIQA ISQAFGGNYE TLLRTLGYAT EDFDDLLESD SIAGQIVYVD LSSYYIIVRV YFPILTEIQQ AYVQELLPVS FNNDNSEWIS IVPNFVLIRN TLISNIEVKY CLITKKSVIC NQDYATPMTA SVRECLTGST DKCPRELVVS SHVPRFALSG GVLFANCISV TCQCQTTGRA ISQSGEQTLL MIDNTTCTTV VLGNIIISLG KYLGSINYNS ESIAVGPPVY TDKVDISSQI SSMNQSLQQS KDYIKEAQKI LDTVNPSLIS MLSMIILYVL SIAALCIGLI TFISFVIVEK KRGNYSRLDD RQVRPVSNGD LYYIGT	亨德拉病毒F蛋白Uniprot O89342 (没有信号序列)
60	IHYDSLSKVGVIKGLTYNYKIKGSPSTKLMVVKLIPNIDSVKNCTQKQYDE YKNLVRKALEPVKMAIDTMLNNVKSGNNKYRFAGAIMAGVALGVATAATVT AGIALHRSNENAQAIANMKSAIQNTNEAVKQLQLANKQTLAVIDTIRGEIN NNIIPVINQLSCDTIGLSVGIRLTQYYSEIITAFGPALQNPVNTRITIQAI SSVFNGNFDELLKIMGYTSGDLYEILHSELIRGNIIDVDVDAGYIALEIEF PNLTLVPNAVVQELMPISYNIDGDEWVTLVPRFVLTRTTLLSNIDTSRCTI TDSSVICDNDYALPMSHELIGCLQGDTSKCAREKVVSSYVPKFALSDGLVY ANCLNTICRCMDTDTPISQSLGATVSLLDNKRCSVYQVGDVLISVGSYLGD GEYNADNVELGPPIVIDKIDIGNQLAGINQTLQEAEDYIEKSEEFLKGVNP SIITLGSMVVLYIFMILIAIVSVIALVLSIKLTVKGNVVRQQFTYTQHVPS MENINYVSH	gb:NC_025352:5950-8712|生物体：墨江病毒|毒株名称：Tongguan1|蛋白质名称：融合蛋白|基因符号：F (没有信号序列)
61	MLFNLRILLNNAAFRNGHNFMVRNFRCGQPLQNKVQLKGRD LLTLKNFTGEEIKYMLWLSADLKFRIKQKGEYLPLLQGKSLG MIFEKRSTRTRLSTETGFALLGGHPCFLTTQDIHLGVNESLTDT ARVLSSMADAVLARVYKQSDLDTLAKEASIPIINGLSDLYHPI QILADYLTLQEHYSSLKGLTLSWIGDGNNILHSIMMSAAKFG MHLQAATPKGYEPDASVTKLAEQYAKENGTKLLLTNDPLEA AHGGNVLITDTWISMGQEEEKKKRLQAFQGYQVTMKTAKV AASDWTFLHCLPRKPEEVDDEVFYSPRSLVFPEAENRKWTIM AVMVSLLTDYSPQLQKPKF	OTC
62	MTRILTAFKVVRTLKTGFGFTNVTAHQKWKFSRPGIRLLSVK AQTAHIVLEDGTKMKGYSFGHPSSVAGEVVFNTGLGGYPEAI	CPS1

[0828]

	TDPAYKGQILTMANPIIGNGGAPDTTALDELGLSKYLESNGIK VSGLLVLDYSKDYNHWLATKSLGQWLQEEKVPAIYGVDTR MLTKIIRDKGTMLGKIEFEGQPVDFVDPNKQNLIAEVSTKDV KVYGKGNPTKVVAVDCGIKNNVIRLLVKRGAEVHLVPWNH DFTKMEYDGILIAGGPGNPALAEPLIQNVRKILESDRKEPLFGI STGNLITGLAAGAKTYKMSMANRGQNQPVLNITNKQAFITAQ NHGYALDNTLPAGWKPLFVNVNDQTNEGIMHESKPFFAVQF HPEVTPGPIDTEYLFDSFFSLIKKGKATTITSVLPKPALVASRV EVSKVLILGSGGLSIGQAGEFDYSGSQAVKAMKEENVKTVLM NPNIASVQTNEVGLKQADTVYFLPITPQFVTEVIKAEQPDGLIL GMGGQTALNCGVELFKRGVLKEYGVKVLGTSVESIMATEDR QLFSDKLNEINEKIAPSFAVESIEDALKAADTIGYPVMIRSAYA LGGLGSGICPNRETLMDLSTKAFAMTNQILVEKSVTGWKEIE YEVVRDADDNCVTVCNMENVDAMGVHTGDSVVVAPAQTLS NAEFQMLRRTSINVVRHLGIVGECNIQFALHPTSMEYCIIEVN ARLSRSSALASKATGYPLAFIAAKIALGIPLPEIKNVVSGKTSA CFEPSLDYMVTKIPRWDLDRFHGTSSRIGSSMKSVGEVMAIG RTFEESFQKALRMCHPSIEGFTPRLPMNKEWPSNLDLRKELSE PSSTRIYAIAKAIDDNMSLDEIEKLTYIDKWFLYKMRDILNME KTLKGLNSESMTEETLKRAKEIGFSDKQISKCLGLTEAQTREL RLKKNIHPWVKQIDTLAAEYPSVTNYLYVTYNGQEHDVNFD DHGMMVLGCGPYHIGSSVEFDWCAVSSIRTLRQLGKKTVVV NCNPETVSTDFDECDKLYFEELSLERILDIYHQEACGGCIISVG GQIPNNLAVPLYKNGVKIMGTSPLQIDRAEDRSIFSAVLDELK VAQAPWKAVNTLNEALEFAKSVDYPCLLRPSYVLSGSAMNV VFSEDEMKKFLEEATRVSQEHPVVLTKFVEGAREVEMDAVG KDGRVISHAISEHVEDAGVHSGDATLMLPTQTISQGAIEKVKD ATRKIAKAFAISGPFNVQFLVKGNDVLVIECNLRASRSFPFVS KTLGVDFIDVATKVMIGENVDEKHLPTLDHPIIPADYVAIKAP MFSWPRLRDADPILRCEMASTGEVACFGEGIHTAFLKAMLST GFKIPQKGILIGIQQSFRPRFLGVAEQLHNEGFKLFATEATSDW LNANNVPATPVAWPSQEGQNPSLSSIRKLIRDGSIDLVINLPNN NTKFVHDNYVIRRTAVDSGIPLLTNFQVTKLFAEAVQKSRKV DSKSLFHYRQYSAGKAA	
63	MATALMAVVLRAAAVAPRLRGRGGTGGARRLSCGARRRAA RGTSPGRRLSTAWSQPQPPPEEYAGADDVSQSPVAEEPSWVP SPRPPVPHESPEPPSGRSLVQRDIQAFLNQCGASPGEARHWLT QFQTCHHSADKPFAVIEVDEEVLKCQQGVSSLAFALAFLQRM DMKPLVVLGLPAPTAPSGCLSFWEAKAQLAKSCKVLVDALR HNAAAAVPFFGGGSVLRAAEPAPHASYGGIVSVETDLLQWC LESGSIPILCPIGETAARRSVLLDSLEVTASLAKALRPTKIIFLN NTGGLRDSSHKVLSNVNLPADLDLVCNAEWVSTKERQQMRL IVDVLSRLPHHSSAVITAASTLLTELFSNKGSGTLFKNAERML RVRSLDKLDQGRLVDLVNASFGKKLRDDYLASLRPRLHSIYV SEGYNAAAILTMEPVLGGTPYLDKFVVSSSRQGQGSGQMLW ECLRRDLQTLFWRSRVTNPINPWYFKHSDGSFSNKQWIFFWF GLADIRDSYELVNHAKGLPDSFHKPASDPGS	NAGS
64	MAVAIAAARVWRLNRGLSQAALLLLRQPGARGLARSHPPRQ QQQFSSLDDKPQFPGASAEFIDKLEFIQPNVISGIPIYRVMDRQ GQIINPSEDPHLPKEKVLKLYKSMTLLNTMDRILYESQRQGRI SFYMTNYGEEGTHVGSAAALDNTDLVFGQYREAGVLMYRD YPLELFMAQCYGNISDLGKGRQMPVHYGCKERHFVTISSPLA TQIPQAVGAAYAAKRANANRVVICYFGEGAASEGDAHAGFN FAATLECPIIFFCRNNGYAISTPTSEQYRGDGIAARGPGYGIMSI	BCKDHA

		RVDGNDVFAVYNATKEARRRAVAENQPFLIEAMTYRIGHHS TSDDSSAYRSVDEVNYWDKQDHPISRLRHYLLSQGWWDEEQ EKAWRKQSRRKVMEAFEQAERKPKPNPNLLFSDVYQEMPAQ LRKQQESLARHLQTYGEHYPLDHFDK	
	65	MAVVAAAAGWLLRLRAAGAEGHWRRLPGAGLARGFLHPA ATVEDAAQRRQVAHFTFQPDPEPREYGQTQKMNLFQSVTSA LDNSLAKDPTAVIFGEDVAFGGVFRCTVGLRDKYGKDRVFN TPLCEQGIVGFGIGIAVTGATAIAEIQFADYIFPAFDQIVNEAA KYRYRSGDLFNCGSLTIRSPWGCVGHGALYHSQSPEAFFAHC PGIKVVIPRSPFQAKGLLLSCIEDKNPCIFFEPKILYRAAAEEVP IEPYNIPLSQAEVIQEGSDVTLVAWGTQVHVIREVASMAKEKL GVSCEVIDLRTIIPWDVDTICKSVIKTGRLLISHEAPLTGGFASE ISSTVQEECFLNLEAPISRVCGYDTPFPHIFEPFYIPDKWKCYD ALRKMINY	BCKDHB
	66	MAAVRMLRTWSRNAGKLICVRYFQTCGNVHVLKPNYVCFF GYPSFKYSHPHHFLKTTAALRGQVVQFKLSDIGEGIREVTVKE WYVKEGDTVSQFDSICEVQSDKASVTITSRYDGVIKKLYYNL DDIAYVGKPLVDIETEALKDSEEDVVETPAVSHDEHTHQEIKG RKTLATPAVRRLAMENNIKLSEVVGSGKDGRILKEDILNYLE KQTGAILPPSPKVEIMPPPPKPKDMTVPILVSKPPVFTGKDKTE PIKGFQKAMVKTMSAALKIPHFGYCDEIDLTELVKLREELKPI AFARGIKLSFMPFFLKAASLGLLQFPILNASVDENCQNITYKA SHNIGIAMDTEQGLIVPNVKNVQICSIFDIATELNRLQKLGSVG QLSTTDLTGGTFTLSNIGSIGGTFAKPVIMPPEVAIGALGSIKAI PRFNQKGEVYKAQIMNVSWSADHRVIDGATMSRFSNLWKSY LENPAFMLLDLK	DBT
[0829]	67	MQSWSRVYCSLAKRGHFNRISHGLQGLSAVPLRTYADQPIDA DVTVIGSGPGGYVAAIKAAQLGFKTVCIEKNETLGGTCLNVG CIPSKALLNNSHYYHMAHGKDFASRGIEMSEVRLNLDKMME QKSTAVKALTGGIAHLFKQNKVVHVNGYGKITGKNQVTATK ADGGTQVIDTKNILIATGSEVTPFPGITIDEDTIVSSTGALSLKK VPEKMVVIGAGVIGVELGSVWQRLGADVTAVEFLGHVGGVG IDMEISKNFQRILQKQGFKFKLNTKVTGATKKSDGKIDVSIEA ASGGKAEVITCDVLLVCIGRRPFTKNLGLEELGIELDPRGRIPV NTRFQTKIPNIYAIGDVVAGPMLAHKAEDEGIICVEGMAGGA VHIDYNCVPSVIYTHPEVAWVGKSEEQLKEEGIEYKVGKFPF AANSRAKTNADTDGMVKILGQKSTDRVLGAHILGPGAGEMV NEAALALEYGASCEDIARVCHAHPTLSEAFREANLAASFGKSI NF	DLD
	68	MLRAKNQLFLLSPHYLRQVKESSGSRLIQQRLLHQQQPLHPE WAALAKKQLKGKNPEDLIWHTPEGISIKPLYSKRDTMDLPEE LPGVKPFTRGPYPTMYTFRPWTIRQYAGFSTVEESNKFYKDNI KAGQQGLSVAFDLATHRGYDSDNPRVRGDVGMAGVAIDTV EDTKILFDGIPLEKMSVSMTMNGAVIPVLANFIVTGEEQGVPK EKLTGTIQNDILKEFMVRNTYIFPPEPSMKIIADIFEYTAKHMP KFNSISISGYHMQEAGADAILELAYTLADGLEYSRTGLQAGLT IDEFAPRLSFFWGIGMNFYMEIAKMRAGRRLWAHLIEKMFQP KNSKSLLLRAHCQTSGWSLTEQDPYNNIVRTAIEAMAAVFGG TQSLHTNSFDEALGLPTVKSARIARNTQIIIQEESGIPKVADPW GGSYMMECLTNDVYDAALKLINEIEEMGGMAKAVAEGIPKL RIEECAARRQARIDSGSEVIVGVNKYQLEKEDAVEVLAIDNTS VRNRQIEKLKKIKSSRDQALAERCLAALTECAASGDGNILAL AVDASRARCTVGEITDALKKVFGEHKANDRMVSGAYRQEFG ESKEITSAIKRVHKFMEREGRRPRLLVAKMGQDGHDRGAKVI	MUT

		ATGFADLGFDVDIGPLFQTPREVAQQAVDADVHAVGISTLAA GHKTLVPELIKELNSLGRPDILVMCGGVIPPQDYEFLFEVGVS NVFGPGTRIPKAAVQVLDDIEKCLEKKQQSV	
[0830]	69	MPMLLPHPHQHFLKGLLRAPFRCYHFIFHSSTHLGSGIPCAQP FNSLGLHCTKWMLLSDGLKRKLCVQTTLKDHTEGLSDKEQR FVDKLYTGLIQGQRACLAEAITLVESTHSRKKELAQVLLQKV LLYHREQEQSNKGKPLAFRVGLSGPPGAGKSTFIEYFGKMLT ERGHKLSVLAVDPSSCTSGGSLLGDKTRMTELSRDMNAYIRP SPTRGTLGGVTRTTNEAILLCEGAGYDIILIETVGVGQSEFAVA DMVDMFVLLLPPAGGDELQGIKRGIIEMADLVAVTKSDGDLI VPARRIQAEYVSALKLLRKRSQVWKPKVIRISARSGEGISEM WDKMKDFQDLMLASGELTAKRRKQQKVWMWNLIQESVLE HFRTHPTVREQIPLLEQKVLIGALSPGLAADFLLKAFKSRD	MMAA
	70	MAVCGLGSRLGLGSRLGLRGCFGAARLLYPRFQSRGPQGVE DGDRPQPSSKTPRIPKIYTKTGDKGFSSTFTGERRPKDDQVFE AVGTTDELSSAIGFALELVTEKGHTFAEELQKIQCTLQDVGSA LATPCSSAREAHLKYTTFKAGPILELEQWIDKYTSQLPPLTAFI LPSGGKISSALHFCRAVCRRAERRVVPLVQMGETDANVAKFL NRLSDYLFTLARYAAMKEGNQEKIYMKNDPSAESEGL	MMAB
	71	MFDRALKPFLQSCHLRMLTDPVDQCVAYHLGRVRESLPELQI EIIADYEVHPNRRPKILAQTAAHVAGAAYYYQRQDVEADPW GNQRISGVCIHPRFGGWFAIRGVVLLPGIEVPDLPPRKPHDCV PTRADRIALLEGFNFHWRDWTYRDAVTPQERYSEEQKAYFST PPAQRLALLGLAQPSEKPSSPSPDLPFTTPAPKKPGNPSRARS WLSPRVSPPASPGP	MMACHC
	72	MANVLCNRARLVSYLPGFCSLVKRVVNPKAFSTAGSSGSDES HVAAAPPDICSRTVWPDETMGPFGPQDQRFQLPGNIGFDCHL NGTASQKKSLVHKTLPDVLAEPLSSERHEFVMAQYVNEFQG NDAPVEQEINSAETYFESARVECAIQTCPELLRKDFESLFPEVA NGKLMILTVTQKTKNDMTVWSEEVEIEREVLLEKFINGAKEI CYALRAEGYWADFIDPSSGLAFFGPYTNNTLFETDERYRHLG FSVDDLGCCKVIRHSLWGTHVVVGSIFTNATPDSHIMKKLSG N	MMADHC
	73	MARVLKAAAANAVGLFSRLQAPIPTVRASSTSQPLDQVTGSV WNLGRLNHVAIAVPDLEKAAAFYKNILGAQVSEAVPLPEHG VSVVFVNLGNTKMELLHPLGRDSPIAGFLQKNKAGGMHHICI EVDNINAAVMDLKKKKIRSLSEEVKIGAHGKPVIFLHPKDCG GVLVELEQA	MCEE
	74	MAGFWVGTAPLVAAGRRGRWPPQQLMLSAALRTLKHVLYY SRQCLMVSRNLGSVGYDPNEKTFDKILVANRGEIACRVIRTC KKMGIKTVAIHSDVDASSVHVKMADEAVCVGPAPTSKSYLN MDAIMEAIKKTRAQAVHPGYGFLSENKEFARCLAAEDVVFIG PDTHAIQAMGDKIESKLLAKKAEVNTIPGFDGVVKDAEEAVR IAREIGYPVMIKASAGGGGKGMRIAWDDEETRDGFRLSSQEA ASSFGDDRLLIEKFIDNPRHIEIQVLGDKHGNALWLNERECSIQ RRNQKVVEEAPSIFLDAETRRAMGEQAVALARAVKYSSAGT VEFLVDSKKNFYFLEMNTRLQVEHPVTECITGLDLVQEMIRV AKGYPLRHKQADIRINGWAVECRVYAEDPYKSFGLPSIGRLS QYQEPLHLPGVRVDSGIQPGSDISIYYDPMISKLITYGSDRTEA LKRMADALDNYVIRGVTHNIALLREVIINSRFVKGDISTKFLS DVYPDGFKGHMLTKSEKNQLLAIASSLFVAFQLRAQHFQENS RMPVIKPDIANWELSVKLHDKVHTVVASNNGSVFSVEVDGS KLNVTSTWNLASPLLSVSVDGTQRTVQCLSREAGGNMSIQFL	PCCA

[0831]

	GTVYKVNILTRLAAELNKFMLEKVTEDTSSVLRSPMPGVVVA VSVKPGDAVAEGQEICVIEAMKMQNSMTAGKTGTVKSVHCQ AGDTVGEGDLLVELE	
75	MAAALRVAAVGARLSVLASGLRAAVRSLCSQATSVNERIEN KRRTALLGGGQRRIDAQHKRGKLTARERISLLLDPGSFVESD MFVEHRCADFGMAADKNKFPGDSVVTGRGRINGRLVYVFSQ DFTVFGGSLSGAHAQKICKIMDQAITVGAPVIGLNDSGGARIQ EGVESLAGYADIFLRNVTASGVIPQISLIMGPCAGGAVYSPAL TDFTFMVKDTSYLFITGPDVVKSVTNEDVTQEELGGAKTHTT MSGVAHRAFENDVDALCNLRDFFNYLPLSSQDPAPVRECHDP SDRLVPELDTIVPLESTKAYNMVDIIHSVVDEREFFEIMPNYA KNIIVGFARMNGRTVGIVGNQPKVASGCLDINSSVKGARFVR FCDAFNIPLITFVDVPGFLPGTAQEYGGIIRHGAKLLYAFAEAT VPKVTVITRKAYGGAYDVMSSKHLCGDTNYAWPTAEIAVM GAKGAVEIIFKGHENVEAAQAEYIEKFANPFPAAVRGFVDDII QPSSTRARICCDLDVLASKKVQRPWRKHANIPL	PCCB
76	MAVESQGGRPLVLGLLLCVLGPVVSHAGKILLIPVDGSHWLS MLGAIQQLQQRGHEIVVLAPDASLYIRDGAFYTLKTYPVPFQ REDVKESFVSLGHNVFENDSFLQRVIKTYKKIKKDSAMLLSG CSHLLHNKELMASLAESSFDVMLTDPFLPCSPIVAQYLSLPTV FFLHALPCSLEFEATQCPNPFSYVPRPLSSHSDHMTFLQRVKN MLIAFSQNFLCDVVYSPYATLASEFLQREVTVQDLLSSASVW LFRSDFVKDYPRPIMPNMVFVGGINCLHQNPLSQEFEAYINAS GEHGIVVFSLGSMVSEIPEKKAMAIADALGKIPQTVLWRYTG TRPSNLANNTILVKWLPQNDLLGHPMTRAFITHAGSHGVYESI CNGVPMVMMPLFGDQMDNAKRMETKGAGVTLNVLEMTSE DLENALKAVINDKSYKENIMRLSSLHKDRPVEPLDLAVFWVE FVMRHKGAPHLRPAAHDLTWYQYHSLDVIGFLLAVVLTVAF ITFKCCAYGYRKCLGKKGRVKKAHKSKTH	UGT1A1
77	MSSKGSVVLAYSGGLDTSCILVWLKEQGYDVIAYLANIGQKE DFEEARKKALKLGAKKVFIEDVSREFVEEFIWPAIQSSALYED RYLLGTSLARPCIARKQVEIAQREGAKYVSHGATGKGNDQV RFELSCYSLAPQIKVIAPWRMPEFYNRFKGRNDLMEYAKQHG IPIPVTPKNPWSMDENLMHISYEAGILENPKNQAPPGLYTKTQ DPAKAPNTPDILEIEFKKGVPVKVTNVKDGTTHQTSLELFMY LNEVAGKHGVGRIDIVENRFIGMKSRGIYETPAGTILYHAHLD IEAFTMDREVRKIKQGLGLKFAELVYTGFWHSPECEFVRHCIA KSQERVEGKVQVSVLKGQVYILGRESPLSLYNEELVSMNVQG DYEPTDATGFININSLRLKEYHRLQSKVTAK	ASS1
78	MSTAVLENPGLGRKLSDFGQETSYIEDNCNQNGAISLIFSLKE EVGALAKVLRLFEENDVNLTHIESRPSRLKKDEYEFFTHLDKR SLPALTNIIKILRHDIGATVHELSRDKKKDTVPWFPRTIQELDR FANQILSYGAELDADHPGFKDPVYRARRKQFADIAYNYRHG QPIPRVEYMEEEKKTWGTVFKTLKSLYKTHACYEYNHIFPLL EKYCGFHEDNIPQLEDVSQFLQTCTGFRLRPVAGLLSSRDFLG GLAFRVFHCTQYIRHGSKPMYTPEPDICHELLGHVPLFSDRSF AQFSQEIGLASLGAPDEYIEKLATIYWFTVEFGLCKQGDSIKA YGAGLLSSFGELQYCLSEKPKLLPLELEKTAIQNYTVTEFQPL YYVAESFNDAKEKVRNFAATIPRPFSVRYDPYTQRIEVLDNT QQLKILADSINSEIGILCSALQKIK	PAH
79	MAKTLSQAQSKTSSQQFSFTGNSSANVIIGNQKLTINDVARVA RNGTLVSLTNNTDILQGIQASCDYINNAVESGEPIYGVTSGFG GMANVAISREQASELQTNLVWFLKTGAGNKLPLADVRAAML	PAL

[0832]

	LRANSHMRGASGIRLELIKRMEIFLNAGVTPYVYEFGSIGASG DLVPLSYITGSLIGLDPSFKVDFNGKEMDAPTALRQLNLSPLT LLPKEGLAMMNGTSVMTGIAANCVYDTQILTAIAMGVHALD IQALNGTNQSFHPFIHNSKPHPGQLWAADQMISLLANS QLVRDELDGKHDYRDHELIQDRYSLRCLPQYLGPIVDGISQIA KQIEIEINSVTDNPLIDVDNQASYHGGNFLGQYVGMGMDHLR YYIGLLAKHLDVQIALLASPEFSNGLPPSLLGNRERKVNMGL KGLQICGNSIMPLLTFYGNSIADRFPTHAEQFNQNINSQGYTS ATLARRSVDIFQNYVAIALMFGVQAVDLRTYKKTGHYDARA SLSPATERLYSAVRHVVGQKPTSDRPYIWNDNEQGLDEHIARI SADIAAGGVIVQAVQDILPSLH	
80	MSTERDSETTFDEDSQPNDEVVPYSDDETEDELDDQGSAVEP EQNRVNREAEENREPFRKECTWQVKANDRKYHEQPHFMNT KFLCIKESKYANNAIKTYKYNAFTFIPMNLFEQFKRAANLYFL ALLILQAVPQISTLAWYTTLVPLLVVLGVTAIKDLVDDVARH KMDKEINNRTCEVIKDGRFKVAKWKEIQVGDVIRLKKNDFVP ADILLLSSSEPNSLCYVETAELDGETNLKFKMSLEITDQYLQR EDTLATFDGFIECEEPNNRLDKFTGTLFWRNTSFPLDADKILL RGCVIRNTDFCHGLVIFAGADTKIMKNSGKTRFKRTKIDYLM NYMVYTIFVVLILLSAGLAIGHAYWEAQVGNSSWYLYDGED DTPSYRGFLIFWGYIIVLNTMVPISLYVSVEVIRLGQSHFINWD LQMYYAEKDTPAKARTTTLNEQLGQIHYIFSDKTGTLTQNIM TFKKCCINGQIYGDHRDASQHNHNKIEQVDFSWNTYADGKL AFYDHYLIEQIQSGKEPEVRQFFFLLAVCHTVMVDRTDGQLN YQAASPDEGALVNAARNFGFAFLARTQNTITISELGTERTYNV LAILDFNSDRKRMSIIVRTPEGNIKLYCKGADTVIYERLHRMN PTKQETQDALDIFANETLRTLCLCYKEIEEKEFTEWNKKFMA ASVASTNRDEALDKVYEEIEKDLILLGATAIEDKLQDGVPETI SKLAKADIKIWVLTGDKKETAENIGFACELLTEDTTICYGEDI NSLLHARMENQRNRGGVYAKFAPPVQESFFPPGGNRALIITGS WLNEILLEKKTKRNKILKLKFPRTEEERRMRTQSKRRLEAKK EQRQKNFVDLACECSAVICCRVTPKQKAMVVDLVKRYKKAI TLAIGDGANDVNMIKTAHIGVGISGQEGMQAVMSSDYSFAQF RYLQRLLLVHGRWSYIRMCKFLRYFFYKNFAFTLVHFWYSFF NGYSAQTAYEDWFITLYNVLYTSLPVLLMGLLDQDVSDKLSL RFPGLYIVGQRDLLFNYKRFFVSLLHGVLTSMILFFIPLGAYLQ TVGQDGEAPSDYQSFAVTIASALVITVNFQIGLDTSYWTFVNA FSIFGSIALYFGIMFDFHSAGIHVLFPSAFQFTGTASNALRQPYI WLTIILAVAVCLLPVVAIRFLSMTIWPSESDKIQKHRKRLKAE EQWQRRQQVFRRGVSTRRSAYAFSHQRGYADLISSGRSIRKK RSPLDAIVADGTAEYRRTGDS	ATP8B1
81	MSDSVILRSIKKFGEENDGFESDKSYNNDKKSRLQDEKKGDG VRVGFFQLFRFSSSTDIWLMFVGSLCAFLHGIAQPGVLLIFGT MTDVFIDYDVELQELQIPGKACVNNTIVWTNSSLNQNMTNGT RCGLLNIESEMIKFASYYAGIAVAVLITGYIQICFWVIAAARQI QKMRKFYFRRIMRMEIGWFDCNSVGELNTRFSDDINKINDAI ADQMALFIQRMTSTICGFLLGFFRGWKLTLVIISVSPLIGIGAA TIGLSVSKFTDYELKAYAKAGVVADEVISSMRTVAAFGGEKR EVERYEKNLVFAQRWGIRKGIVMGFFTGFVWCLIFLCYALAF WYGSTLVLDEGEYTPGTLVQIFLSVIVGALNLGNASPCLEAFA TGRAAATSIFETIDRKPIIDCMSEDGYKLDRIKGEIEFHNVTFH YPSRPEVKILNDLNMVIKPGEMTALVGPSGAGKSTALQLIQRF YDPCEGMVTVDGHDIRSLNIQWLRDQIGIVEQEPVLFSTTIAE NIRYGREDATMEDIVQAAKEANAYNFIMDLPQQFDTLVGEG	ABCB11

		GGQMSGGQKQRVAIARALIRNPKILLLDMATSALDNESEAM VQEVLSKIQHGHTIISVAHRLSTVRAADTIIGFEHGTAVERGTH EELLERKGVYFTLVTLQSQGNQALNEEDIKDATEDDMLARTF SRGSYQDSLRASIRQRSKSQLSYLVHEPPLAVVDHKSTYEEDR KDKDIPVQEEVEPAPVRRILKFSAPEWPYMLVGSVGAAVNGT VTPLYAFLFSQILGTFSIPDKEEQRSQINGVCLLFVAMGCVSLF TQFLQGYAFAKSGELLTKRLRKFGFRAMLGQDIAWFDDLRN SPGALTTRLATDASQVQGAAGSQIGMIVNSFTNVTVAMIIAFS FSWKLSLVILCFFPFLALSGATQTRMLTGFASRDKQALEMVG QITNEALSNIRTVAGIGKERRFIEALETELEKPFKTAIQKANIYG FCFAFAQCIMFIANSASYRYGGYLISNEGLHFSYVFRVISAVVL SATALGRAFSYTPSYAKAKISAARFFQLLDRQPPISVYNTAGE KWDNFQGKIDFVDCKFTYPSRPDSQVLNGLSVSISPGQTLAFV GSSGCGKSTSIQLLERFYDPDQGKVMIDGHDSKKVNVQFLRS NIGIVSQEPVLFACSIMDNIKYGDNTKEIPMERVIAAAKQAQL HDFVMSLPEKYETNVGSQGSQLSRGEKQRIAIARAIVRDPKIL LLDEATSALDTESEKTVQVALDKAREGRTCIVIAHRLSTIQNA DIIAVMAQGVVIEKGTHEELMAQKGAYYKLVTTGSPIS	
[0833]	82	MDLEAAKNGTAWRPTSAEGDFELGISSKQKRKKTKTVKMIG VLTLFRYSDWQDKLFMSLGTIMAIAHGSGLPLMMIVFGEMTD KFVDTAGNFSFPVNFSLSLLNPGKILEEEMTRYAYYYSGLGA GVLVAAYIQVSFWTLAAGRQIRKIRQKFFHAILRQEIGWFDIN DTTELNTRLTDDISKISEGIGDKVGMFFQAVATFFAGFIVGFIR GWKLTLVIMAISPILGLSAAVWAKILSAFSDKELAAYAKAGA VAEEALGAIRTVIAFGGQNKELERYQKHLENAKEIGIKKAISA NISMGIAFLLIYASYALAFWYGSTLVISKEYTIGNAMTVFFSILI GAFSVGQAAPCIDAFANARGAAYVIFDIIDNNPKIDSFSERGH KPDSIKGNLEFNDVHFSYPSRANVKILKGLNLKVQSGQTVAL VGSSGCGKSTTVQLIQRLYDPDEGTINIDGQDIRNFNVNYLREI IGVVSQEPVLFSTTIAENICYGRGNVTMDEIKKAVKEANAYEF IMKLPQKFDTLVGERGAQLSGGQKQRIAIARALVRNPKILLLD EATSALDTESEAEVQAALDKAREGRTTIVIAHRLSTVRNADVI AGFEDGVIVEQGSHSELMKKEGVYFKLVNMQTSGSQIQSEEF ELNDEKAATRMAPNGWKSRLFRHSTQKNLKNSQMCQKSLD VETDGLEANVPPVSFLKVLKLNKTEWPYFVVGTVCAIANGGL QPAFSVIFSEIIAIFGPGDDAVKQQKCNIFSLIFLFLGIISFFTFFL QGFTFGKAGEILTRRLRSMAFKAMLRQDMSWFDDHKNSTGA LSTRLATDAAQVQGATGTRLALIAQNIANLGTGIIISFIYGWQL TLLLLAVVPIIAVSGIVEMKLLAGNAKRDKKELEAAGKIATEA IENIRTVVSLTQERKFESMYVEKLYGPYRNSVQKAHIYGITFSI SQAFMYFSYAGCFRFGAYLIVNGHMRFRDVILVFSAIVFGAV ALGHASSFAPDYAKAKLSAAHLFMLFERQPLIDSYSEEGLKP DKFEGNITFNEVVFNYPTRANVPVLQGLSLEVKKGQTLALVG SSGCGKSTVVQLLERFYDPLAGTVFVDFGFQLLDGQEAKKLN VQWLRAQLGIVSQEPILFDCSIAENIAYGDNSRVVSQDEIVSA AKAANIHPFIETLPHKYETRVGDKGTQLSGGQKQRIAIARALI RQPQILLLDEATSALDTESEKVVQEALDKAREGRTCIVIAHRL STIQNADLIVVFQNGRVKEHGTHQQLLAQKGIYFSMVSVQAG TQNL	ABCB4
	83	MPVRGDRGFPPRRELSGWLRAPGMEELIWEQYTVTLQKDSK RGFGIAVSGGRDNPHFENGETSIVISDVLPGGPADGLLQENDR VVMVNGTPMEDVLHSFAVQQLRKSGKVAAIVVKRPRKVQV AALQASPPLDQDDRAFEVMDEFDGRSFRSGYSERSRLNSHGG RSRSWEDSPERGRPHERARSRERDLSRDRSRGRSLERGLDQD	TJP2

[0834]

	HARTRDRSRGRSLERGLDHDFGPSRDRDRDRSRGRSIDQDYE RAYHRAYDPDYERAYSPEYRRGARHDARSRGPRSRSREHPHS RSPSPEPRGRPGPIGVLLMKSRANEEYGLRLGSQIFVKEMTRT GLATKDGNLHEGDIILKINGTVTENMSLTDARKLIEKSRGKLQ LVVLRDSQQTLINIPSLNDSDSEIEDISEIESNRSFSPEERRHQYS DYDYHSSSEKLKERPSSREDTPSRLSRMGATPTPFKSTGDIAG TVVPETNKEPRYQEDPPAPQPKAAPRTFLRPSPEDEAIYGPNT KMVRFKKGDSVGLRLAGGNDVGIFVAGIQEGTSAEQEGLQE GDQILKVNTQDFRGLVREDAVLYLLEIPKGEMVTILAQSRAD VYRDILACGRGDSFFIRSHFECEKETPQSLAFTRGEVFRVVDT LYDGKLGNWLAVRIGNELEKGLIPNKSRAEQMASVQNAQRD NAGDRADFWRMRGQRSGVKKNLRKSREDLTAVVSVSTKFP AYERVLLREAGFKRPVVLFGPIADIAMEKLANELPDWFQTAK TEPKDAGSEKSTGVVRLNTVRQIIEQDKHALLDVTPKAVDLL NYTQWFPIVIFFNPDSRQGVKTMRQRLNPTSNKSSRKLFDQA NKLKKTCAHLFTATINLNSANDSWFGSLKDTIQHQQGEAVW VSEGKMEGMDDDPEDRMSYLTAMGADYLSCDSRLISDFEDT DGEGGAYTDNELDEPAEEPLVSSITRSSEPVQHEESIRKPSPEP RAQMRRAASSDQLRDNSPPPAFKPEPPKAKTQNKEESYDFSK SYEYKSNPSAVAGNETPGASTKGYPPPVAAKPTFGRSILKPST PIPPQEGEEVGESSEEQDNAPKSVLGKVKIFEKMDHKARLQR MQELQEAQNARIEIAQKHPDIYAVPIKTHKPDPGTPQHTSSRP PEPQKAPSRPYQDTRGSYGSDAEEEEYRQQLSEHSKRGYYGQ SARYRDTEL	
84	MATATRLLGWRVASWRLRPPLAGFVSQRAHSLLPVDDAING LSEEQRQLRQTMAKFLQEHLAPKAQEIDRSNEFKNLREFWKQ LGNLGVLGITAPVQYGGSGLGYLEHVLVMEEISRASGAVGLS YGAHSNLCINQLVRNGNEAQKEKYLPKLISGEYIGALAMSEP NAGSDVVSMKLKAEKKGNHYILNGNKFWITNGPDADVLIVY AKTDLAAVPASRGITAFIVEKGMPGFSTSKKLDKLGMRGSNT CELIFEDCKIPAANILGHENKGVYVLMSGLDLERLVLAGGPLG LMQAVLDHTIPYLHVREAFGQKIGHFQLMQGKMADMYTRL MACRQYVYNVAKACDEGHCTAKDCAGVILYSAECATQVAL DGIQCFGGNGYINDFPMGRFLRDAKLYEIGAGTSEVRRLVIGR AFNADFH	IVD
85	MALRGVSVRLLSRGPGLHVLRTWVSSAAQTEKGGRTQSQLA KSSRPEFDWQDPLVLEEQLTTDEILIRDTFRTYCQERLMPRILL ANRNEVFHREIISEMGELGVLGPTIKGYGCAGVSSVAYGLLA RELERVDSGYRSAMSVQSSLVMHPIYAYGSEEQRQKYLPQLA KGELLGCFGLTEPNSGSDPSSMETRAHYNSSNKSYTLNGTKT WITNSPMADLFVVWARCEDGCIRGFLLEKGMRGLSAPRIQGK FSLRASATGMIIMDGVEVPEENVLPGASSLGGPFGCLNNARY GIAWGVLGASEFCLHTARQYALDRMQFGVPLARNQLIQKKL ADMLTEITLGLHACLQLGRLKDQDKAAPEMVSLLKRNNCGK ALDIARQARDMLGGNGISDEYHVIRHAMNLEAVNTYEGTHDI HALILGRAITGIQAFTASK	GCDH
86	MFRAAAPGQLRRAASLLRFQSTLVIAEHANDSLAPITLNTITA ATRLGGEVSCLVAGTKCDKVAQDLCKVAGIAKVLVAQHDV YKGLLPEELTPLILATQKQFNYTHICAGASAFGKNLLPRVAAK LEVAPISDIIAIKSPDTFVRTIYAGNALCTVKCDEKVKVFSVRG TSFDAAATSGGSASSEKASSTSPVEISEWLDQKLTKSDRPELT GAKVVVSGGRGLKSGENFKLLYDLADQLHAAVGASRAAVD AGFVPNDMQVGQTGKIVAPELYIAVGISGAIQHLAGMKDSKT IVAINKDPEAPIFQVADYGIVADLFKVVPEMTEILKKK	ETFA

[0835]

87	MAELRVLVAVKRVIDYAVKIRVKPDRTGVVTDGVKHSMNPF CEIAVEEAVRLKEKKLVKEVIAVSCGPAQCQETIRTALAMGA DRGIHVEVPPAEAERLGPLQVARVLAKLAEKEKVDLVLLGK QAIDDDCNQTGQMTAGFLDWPQGTFASQVTLEGDKLKVERE IDGGLETLRLKLPAVVTADLRLNEPRYATLPNIMKAKKKKIE VIKPGDLGVDLTSKLSVISVEDPPQRTAGVKVETTEDLVAKLK EIGRI	ETFB
88	MLVPLAKLSCLAYQCFHALKIKKNYLPLCATRWSSTSTVPRIT THYTIYPRDKDKRWEGVNMERFAEEADVVIVGAGPAGLSAA VRLKQLAVAHEKDIRVCLVEKAAQIGAHTLSGACLDPGAFKE LFPDWKEKGAPLNTPVTEDRFGILTEKYRIPVPILPGLPMNNH GNYIVRLGHLVSWMGEQAEALGVEVYPGYAAAEVLFHDDG SVKGIATNDVGIQKDGAPKATFERGLELHAKVTIFAEGCHGH LAKQLYKKFDLRANCEPQTYGIGLKELWVIDEKNWKPGRVD HTVGWPLDRHTYGGSFLYHLNEGEPLVALGLVVGLDYQNPY LSPFREFQRWKHHPSIRPTLEGGKRIAYGARALNEGGFQSIPK LTFPGGLLIGCSPGFMNVPKIKGTHTAMKSGILAAESIFNQLTS ENLQSKTIGLHVTEYEDNLKNSWVWKELYSVRNIRPSCHGVL GVYGGMIYTGIFYWILRGMEPWTLKHKGSDFERLKPAKDCTP IEYPKPDGQISFDLLSSVALSGTNHEHDQPAHLTLRDDSIPVNR NLSIYDGPEQRFCPAGVYEFVPVEQGDGFRLQINAQNCVHCK TCDIKDPSQNINWVVPEGGGGPAYNGM	ETFDH
89	MASESGKLWGGRFVGAVDPIMEKFNASIAYDRHLWEVDVQ GSKAYSRGLEKAGLLTKAEMDQILHGLDKVAEEWAQGTFKL NSNDEDIHTANERRLKELIGATAGKLHTGRSRNDQVVTDLRL WMRQTCSTLSGLLWELIRTMVDRAEAERDVLFPGYTHLQRA QPIRWSHWILSHAVALTRDSERLLEVRKRINVLPLGSGAIAGN PLGVDRELLRAELNFGAITLNSMDATSERDFVAEFLFWASLC MTHLSRMAEDLILYCTKEFSFVQLSDAYSTGSSLMPQKKNPD SLELIRSKAGRVFGRCAGLLMTLKGLPSTYNKDLQEDKEAVF EVSDTMSAVLQVATGVISTLQIHQENMGQALSPDMLATDLA YYLVRKGMPFRQAHEASGKAVFMAETKGVALNQLSLQELQT ISPLFSGDVICVWDYGHSVEQYGALGGTARSSVDWQIRQVRA LLQAQQA	ASL
90	MVGGSVPVFDEIILSTARMNRVLSFHSVSGILVCQAGCVLEEL SRYVEERDFIMPLDLGAKGSCHIGGNVATNAGGLRFLRYGSL HGTVLGLEVVLADGTVLDCLTSLRKDNTGYDLKQLFIGSEGT LGIITTVSILCPPKPRAVNVAFLGCPGFAEVLQTFSTCKGMLGE ILSAFEFMDAVCMQLVGRHLHLASPVQESPFYVLIETSGSNAG HDAEKLGHFLEHALGSGLVTDGTMATDQRKVKMLWALRER ITEALSRDGYVYKYDLSLPVERLYDIVTDLRARLGPHAKHVV GYGHLGDGNLHLNVTAEAFSPSLLAALEPHVYEWTAGQQGS VSAEHGVGFRKRDVLGYSKPPGALQLMQQLKALLDPKGILN PYKTLPSQA	D2HGDH
91	MAAMRKALPRRLVGLASLRAVSTSSMGTLPKRVKIVEVGPR DGLQNEKNIVSTPVKIKLIDMLSEAGLSVIETTSFVSPKWVPQ MGDHTEVLKGIQKFPGINYPVLTPNLKGFEAAVAAGAKEVVI FGAASELFTKKNINCSIEESFQRFDAILKAAQSANISVRGYVSC ALGCPYEGKISPAKVAEVTKKFYSMGCYEISLGDTIGVGTPGI MKDMLSAVMQEVPLAALAVHCHDTYGQALANTLMALQMG VSVVDSSVAGLGGCPYAQGASGNLATEDLVYMLEGLGIHTG VNLQKLLEAGNFICQALNRKTSSKVAQATCKL	HMGCL
92	MAAASAVSVLLVAAERNRWHRLPSLLLPPRTWVWRQRTMK	MCCC1

[0836]

	YTTATGRNITKVLIANRGEIACRVMRTAKKLGVQTVAVYSEA DRNSMHVDMADEAYSIGPAPSQQSYLSMEKIIQVAKTSAAQA IHPGCGFLSENMEFAELCKQEGIIFIGPPPSAIRDMGIKSTSKSI MAAAGVPVVEGYHGEDQSDQCLKEHARRIGYPVMIKAVRG GGGKGMRIVRSEQEFQEQLESARREAKKSFNDDAMLIEKFVD TPRHVEVQVFGDHHGNAVYLFERDCSVQRRHQKIIEEAPAPG IKSEVRKKLGEAAVRAAKAVNYVGAGTVEFIMDSKHNFCFM EMNTRLQVEHPVTEMITGTDLVEWQLRIAAGEKIPLSQEEITL QGHAFEARIYAEDPSNNFMPVAGPLVHLSTPRADPSTRIETGV RQGDEVSVHYDPMIAKLVVWAADRQAALTKLRYSLRQYNIV GLHTNIDFLLNLSGHPEFEAGNVHTDFIPQHHKQLLLSRKAAA KESLCQAALGLILKEKAMTDTFTLQAHDQFSPFSSSSGRRLNI SYTRNMTLKDGKNNVAIAVTYNHDGSYSMQIEDKTFQVLGN LYSEGDCTYLKCSVNGVASKAKLIILENTIYLFSKEGSIEIDIPV PKYLSSVSSQETQGGPLAPMTGTIEKVFVKAGDKVKAGDSLM VMIAMKMEHTIKSPKDGTVKKVFYREGAQANRHTPLVEFEE EESDKRESE	
93	MWAVLRLALRPCARASPAGPRAYHGDSVASLGTQPDLGSAL YQENYKQMKALVNQLHERVEHIKLGGGEKARALHISRGKLL PRERIDNLIDPGSPFLELSQFAGYQLYDNEEVPGGGIITGIGRVS GVECMIIANDATVKGGAYYPVTVKKQLRAQEIAMQNRLPCIY LVDSGGAYLPRQADVFPDRDHFGRTFYNQAIMSSKNIAQIAV VMGSCTAGGAYVPAMADENIIVRKQGTIFLAGPPLVKAATGE EVSAEDLGGADLHCRKSGVSDHWALDDHHALHLTRKVVRN LNYQKKLDVTIEPSEEPLFPADELYGIVGANLKRSFDVREVIA RIVDGSRFTEFKAFYGDTLVTGFARIFGYPVGIVGNNGVLFSE SAKKGTHFVQLCCQRNIPLLFLQNITGFMVGREYEAEGIAKD GAKMVAAVACAQVPKITLIIGGSYGAGNYGMCGRAYSPRFL YIWPNARISVMGGEQAANVLATITKDQRAREGKQFSSADEAA LKEPIIKKFEEEGNPYYSSARVWDDGIIDPADTRLVLGLSFSAA LNAPIEKTDFGIFRM	MCCC2
94	MAVAGPAPGAGARPRLDLQFLQRFLQILKVLFPSWSSQNALM FLTLLCLTLLEQFVIYQVGLIPSQYYGVLGNKDLEGFKTLTFL AVMLIVLNSTLKSFDQFTCNLLYVSWRKDLTEHLHRLYFRGR AYYTLNVLRDDIDNPDQRISQDVERFCRQLSSMASKLIISPFTL VYYTYQCFQSTGWLGPVSIFGYFILGTVVNKTLMGPIVMKLV HQEKLEGDFRFKHMQIRVNAEPAAFYRAGHVEHMRTDRRLQ RLLQTQRELMSKELWLYIGINTFDYLGSILSYVVIAIPIFSGVY GDLSPAELSTLVSKNAFVCIYLISCFTQLIDLSTTLSDVAGYTH RIGQLRETLLDMSLKSQDCEILGESEWGLDTPPGWPAAEPAD TAFLLERVSISAPSSDKPLIKDLSLKISEGQSLLITGNTGTGKTS LLRVLGGLWTSTRGSVQMLTDFGPHGVLFLPQKPFFTDGTLR EQVIYPLKEVYPDSGSADDERILRFLELAGLSNLVARTEGLDQ QVDWNWYDVLSPGEMQRLSFARLFYLQPKYAVLDEATSALT EEVESELYRIGQQLGMTFISVGHRQSLEKFHSLVLKLCGGGR WELMRIKVE	ABCD4
95	MASAVSPANLPAVLLQPRWKRVVGWSGPVPRPRHGHRAVAI KELIVVFGGGNEGIVDELHVYNTATNQWFIPAVRGDIPPGCA AYGFVCDGTRLLVFGGMVEYGKYSNDLYELQASRWEWKRL KAKTPKNGPPPCPRLGHSFSLVGNKCYLFGGLANDSEDPKNN IPRYLNDLYILELRPGSGVVAWDIPITYGVLPPPRESHTAVVYT EKDNKKSKLVIYGGMSGCRLGDLWTLDIDTLTWNKPSLSGV APLPRSLHSATTIGNKMYVFGGWVPLVMDDVKVATHEKEW KCTNTLACLNLDTMAWETILMDTLEDNIPRARAGHCAVAINT	HCFC1

[0837]

	RLYIWSGRDGYRKAWNNQVCCKDLWYLETEKPPPPARVQL VRANTNSLEVSWGAVATADSYLLQLQKYDIPATAATATSPTP NPVPSVPANPPKSPAPAAAAPAVQPLTQVGITLLPQAAPAPPT TTTIQVLPTVPGSSISVPTAARTQGVPAVLKVTGPQATTGTPL VTMRPASQAGKAPVTVTSLPAGVRMVVPTQSAQGTVIGSSPQ MSGMAALAAAAAAATQKIPPSSAPTVLSVPAGTTIVKTMAVTP GTTTLPATVKVASSPVMVSNPATRMLKTAAAQVGTSVSSAT NTSTRPIITVHKSGTVTVAQQAQVVTTVVGGVTKTITLVKSPI SVPGGSALISNLGKVMSVVQTKPVQTSAVTGQASTGPVTQIIQ TKGPLPAGTILKLVTSADGKPTTIITTTQASGAGTKPTILGISSV SPSTTKPGTTTIIKTIPMSAIITQAGATGVTSSPGIKSPITIITTKV MTSGTGAPAKIITAVPKIATGHGQQGVTQVVLKGAPGQPGTI LRTVPMGGVRLVTPVTVSAVKPAVTTLVVKGTTGVTTLGTV TGTVSTSLAGAGGHSTSASLATPITTLGTIATLSSQVINPTAITV SAAQTTLTAAGGLTTPTITMQPVSQPTQVTLITAPSGVEAQPV HDLPVSILASPTTEQPTATVTIADSGQGDVQPGTVTLVCSNPP CETHETGTTNTATTTVVANLGGHPQPTQVQFVCDRQEAAAS LVTSTVGQQNGSVVRVCSNPPCETHETGTTNTATTATSNMAG QHGCSNPPCETHETGTTNTATTAMSSVGANHQRDARRACAA GTPAVIRISVATGALEAAQGSKSQCQTRQTSATSTTMTVMAT GAPCSAGPLLGPSMAREPGGRSPAFVQLAPLSSKVRLSSPSIK DLPAGRHSHAVSTAAMTRSSVGAGEPRMAPVCESLQGGSPST TVTVTALEALLCPSATVTQVCSNPPCETHETGTTNTATTSNAG SAQRVCSNPPCETHETGTTHTATTATSNGGTGQPEGGQQPPA GRPCETHQTTSTGTTMSVSVGALLPDATSSHRTVESGLEVAA APSVTPQAGTALLAPFPTQRVCSNPPCETHETGTTHTATTVTS NMSSNQDPPPAASDQGEVESTQGDSVNITSSSAITTTVSSTLTR AVTTVTQSTPVPGPSVPPPEELQVSPGPRQQLPPRQLLQSAST ALMGESAEVLSASQTPELPAAVDLSSTGEPSSGQESAGSAVV ATVVVQPPPPTQSEVDQLSLPQELMAEAQAGTTTLMVTGLTP EELAVTAAAEAAAQAAATEEAQALAIQAVLQAAQQAVMGT GEPMDTSEAAATVTQAELGHLSAEGQEGQATTIPIVLTQQEL AALVQQQQLQEAQAQQQHHHLPTEALAPADSLNDPAIESNC LNELAGTVPSTVALLPSTATESLAPSNTFVAPQPVVVASPAKL QAAATLTEVANGIESLGVKPDLPPPPSKAPMKKENQWFDVG VIKGTNVMVTHYFLPPDDAVPSDDDLGTVPDYNQLKKQELQ PGTAYKFRVAGINACGRGPFSEISAFKTCLPGFPGAPCAIKISK SPDGAHLTWEPPSVTSGKIIEYSVYLAIQSSQAGGELKSSTPAQ LAFMRVYCGPSPSCLVQSSSLSNAHIDYTTKPAIIFRIAARNEK GYGPATQVRWLQETSKDSSGTKPANKRPMSSPEMKSAPKKS KADGQ	
96	MATSGAASAELVIGWCIFGLLLLAILAFCWIYVRKYQSRRESE VVSTITAIFSLAIALITSALLPVDIFLVSYMKNQNGTFKDWANA NVSRQIEDTVLYGYYTLYSVILFCVFFWIPFVYFYYEEKDDDD TSKCTQIKTALKYTLGFVVICALLLLVGAFVPLNVPNNKNSTE WEKVKSLFEELGSSHGLAALSFSISSLTLIGMLAAITYTAYGM SALPLNLIKGTRSAAYERLENTEDIEEVEQHIQTIKSKSKDGRP LPARDKRALKQFEERLRTLKKRERHLEFIENSWWTKFCGALR PLKIVWGIFFILVALLFVISLFLSNLDKALHSAGIDSGFIIFGANL SNPLNMLLPLLQTVFPLDYILITIIIMYFIFTSMAGIRNIGIWFFW IRLYKIRRGRTRPQALLFLCMILLLIVLHTSYMIYSLAPQYVM YGSQNYLIETNITSDNHKGNSTLSVPKRCDADAPEDQCTVTR TYLFLHKFWFFSAAYYFGNWAFLGVFLIGLIVSCCKGKKSVIE GVDEDSDISDDEPSVYSA	LMBRD1

[0838]

97	MSAKSRTIGIIGAPFSKGQPRGGVEEGPTVLRKAGLLEKLKEQ ECDVKDYGDLPFADIPNDSPFQIVKNPRSVGKASEQLAGKVA EVKKNGRISLVLGGDHSLAIGSISGHARVHPDLGVIWVDAHT DINTPLTTTSGNLHGQPVSFLLKELKGKIPDVPGFSWVTPCISA KDIVYIGLRDVDPGEHYILKTLGIKYFSMTEVDRLGIGKVMEE TLSYLLGRKKRPIHLSFDVDGLDPSFTPATGTPVVGGLTYREG LYITEEIYKTGLLSGLDIMEVNPSLGKTPEEVTRTVNTAVAITL ACFGLAREGNHKPIDYLNPPK	ARG1
98	MKSNPAIQAAIDLTAGAAGGTACVLTGQPFDTMKVKMQTFP DLYRGLTDCCLKTYSQVGFRGFYKGTSPALIANIAENSVLFM CYGFCQQVVRKVAGLDKQAKLSDLQNAAAGSFASAFAALVL CPTELVKCRLQTMYEMETSGKIAKSQNTVWSVIKSILRKDGP LGFYHGLSSTLLREVPGYFFFFGGYELSRSFFASGRSKDELGP VPLMLSGGVGGICLWLAVYPVDCIKSRIQVLSMSGKQAGFIR TFINVVKNEGITALYSGLKPTMIRAFPANGALFLAYEYSRKLM MNQLEAY	SLC25A15
99	MAAAKVALTKRADPAELRTIFLKYASIEKNGEFFMSPNDFVT RYLNIFGESQPNPKTVELLSGVVDQTKDGLISFQEFVAFESVL CAPDALFMVAFQLFDKAGKGEVTFEDVKQVFGQTTIHQHIPF NWDSEFVQLHFGKERKRHLTYAEFTQFLLEIQLEHAKQAFVQ RDNARTGRVTAIDFRDIMVTIRPHVLTPFVEECLVAAAGGTTS HQVSFSYFNGFNSLLNNMELIRKIYSTLAGTRKDVEVTKEEFV LAAQKFGQVTPMEVDILFQLADLYEPRGRMTLADIERIAPLEE GTLPFNLAEAQRQKASGDSARPVLLQVAESAYRFGLGSVAG AVGATAVYPIDLVKTRMQNQRSTGSFVGELMYKNSFDCFKK VLRYEGFFGLYRGLLPQLLGVAPEKAIKLTVNDFVRDKFMHK DGSVPLAAEILAGGCAGGSQVIFTNPLEIVKIRLQVAGEITTGP RVSALSVVRDLGFFGIYKGAKACFLRDIPFSAIYFPCYAHVKA SFANEDGQVSPGSLLLAGAIAGMPAASLVTPADVIKTRLQVA ARAGQTTYSGVIDCFRKILREEGPKALWKGAGARVFRSSPQF GVTLLTYELLQRWFYIDFGGVKPMGSEPVPKSRINLPAPNPDH VGGYKLAVATFAGIENKFGLYLPLFKPSVSTSKAIGGGP	SLC25A13
100	MQPQSVLHSGYFHPLLRAWQTATTTLNASNLIYPIFVTDVPD DIQPITSLPGVARYGVKRLEEMLRPLVEEGLRCVLIFGVPSRVP KDERGSAADSEESPAIEAIHLLRKTFPNLLVACDVCLCPYTSH GHCGLLSENGAFRAEESRQRLAEVALAYAKAGCQVVAPSDM MDGRVEAIKEALMAHGLGNRVSVMSYSAKFASCFYGPFRDA AKSSPAFGDRRCYQLPPGARGLALRAVDRDVREGADMLMV KPGMPYLDIVREVKDKHPDLPLAVYHVSGEFAMLWHGAQA GAFDLKAAVLEAMTAFRRAGADIIITYYTPQLLQWLKEE	ALAD
101	MALQLGRLSSGPCWLVARGGCGGPRAWSQCGGGGLRAWSQ RSAAGRVCRPPGPAGTEQSRGLGHGSTSRGGPWVGTGLAAA LAGLVGLATAAFGHVQRAEMLPKTSGTRATSLGRPEEEEDEL AHRCSSFMAPPVTDLGELRRRPGDMKTKMELLILETQAQVCQ ALAQVDGGANFSVDRWERKEGGGGISCVLQDGCVFEKAGVS ISVVHGNLSEEAAKQMRSRGKVLKTKDGKLPFCAMGVSSVIH PKNPHAPTIHFNYRYFEVEEADGNKQWWFGGGCDLTPTYLN QEDAVHFHRTLKEACDQHGPDLYPKFKKWCDDYFFIAHRGE RRGIGGIFFDDLDSPSKEEVFRFVQSCARAVVPSYIPLVKKHC DDSFTPQEKLWQQLRRGRYVEFNLLYDRGTKFGLFTPGSRIES ILMSLPLTARWEYMHSPSENSKEAEILEVLRHPRDWVR	CPOX
102	MSGNGNAAATAEENSPKMRVIRVGTRKSQLARIQTDSVVAT LKASYPGLQFEIIAMSTTGDKILDTALSKIGEKSLFTKELEHAL	HMBS

[0839]

	EKNEVDLVVHSLKDLPTVLPPGFTIGAICKRENPHDAVVFHPK FVGKTLETLPEKSVVGTSSLRRAAQLQRKFPHLEFRSIRGNLN TRLRKLDEQQEFSAIILATAGLQRMGWHNRVGQILHPEECMY AVGQGALGVEVRAKDQDILDLVGVLHDPETLLRCIAERAFLR HLEGGCSVPVAVHTAMKDGQLYLTGGVWSLDGSDSIQETMQ ATIHVPAQHEDGPEDDPQLVGITARNIPRGPQLAAQNLGISLA NLLLSKGAKNILDVARQLNDAH	
103	MGRTVVVLGGGISGLAASYHLSRAPCPPKVVLVESSERLGGW IRSVRGPNGAIFELGPRGIRPAGALGARTLLLVSELGLDSEVLP VRGDHPAAQNRFLYVGGALHALPTGLRGLLRPSPPFSKPLFW AGLRELTKPRGKEPDETVHSFAQRRLGPEVASLAMDSLCRGV FAGNSRELSIRSCFPSLFQAEQTHRSILLGLLLGAGRTPQPDSA LIRQALAERWSQWSLRGGLEMLPQALETHLTSRGVSVLRGQP VCGLSLQAEGRWKVSLRDSSLEADHVISAIPASVLSELLPAEA APLARALSAITAVSVAVVNLQYQGAHLPVQGFGHLVPSSEDP GVLGIVYDSVAFPEQDGSPPGLRVTVMLGGSWLQTLEASGCV LSQELFQQRAQEAAATQLGLKEMPSHCLVHLHKNCIPQYTLG HWQKLESARQFLTAHRLPLTLAGASYEGVAVNDCIESGRQA AVSVLGTEPNS	PPOX
104	MAHAHIQGGRRAKSRFVVCIMSGARSKLALFLCGCYVVALG AHTGEESVADHHEAEYYVAAVYEHPSILSLNPLALISRQEALE LMNQNLDIYEQQVMTAAQKDVQIIVFPEDGIHGFNFTRTSIYP FLDFMPSPQVVRWNPCLEPHRFNDTEVLQRLSCMAIRGDMFL VANLGTKEPCHSSDPRCPKDGRYQFNTNVVFSNNGTLVDRY RKHNLYFEAAFDVPLKVDLITFDTPFAGRFGIFTCFDILFFDPAI RVLRDYKVKHVVYPTAWMNQLPLLAAIEIQKAFAVAFGINV LAANVHHPVLGMTGSGIHTPLESFWYHDMENPKSHLIIAQVA KNPVGLIGAENATGETDPSHSKFLKILSGDPYCEKDAQEVHC DEATKWNVNAPPTFHSEMMYDNFTLVPVWGKEGYLHVCSN GLCCYLLYERPTLSKELYALGVFDGLHTVHGTYYIQVCALVR CGGLGFDTCGQEITEATGIFEFHLWGNFSTSYIFPLFLTSGMTL EVPDQLGWENDHYFLRKSRLSSGLVTAALYGRLYERD	BTD
105	MEDRLHMDNGLVPQKIVSVHLQDSTLKEVKDQVSNKQAQIL EPKPEPSLEIKPEQDGMEHVGRDDPKALGEEPKQRRGSASGSE PAGDSDRGGGPVEHYHLHLSSCHECLELENSTIESVKFASAEN IPDLPYDYSSSLESVADETSPEREGRRVNLTGKAPNILLYVGS DSQEALGRFHEVRSVLADCVDIDSYILYHLLEDSALRDPWTD NCLLLVIATRESIPEDLYQKFMAYLSQGGKVLGLSSSFTFGGF QVTSKGALHKTVQNLVFSKADQSEVKLSVLSSGCRYQEGPV RLSPGRLQGHLENEDKDRMIVHVPFGTRGGEAVLCQVHLELP PSSNIVQTPEDFNLLKSSNFRRYEVLREILTTLGLSCDMKQVP ALTPLYLLSAAEEIRDPLMQWLGKHVDSEGEIKSGQLSLRFVS SYVSEVEITPSCIPVVTNMEAFSSEHFNLEIYRQNLQTKQLGK VILFAEVTPTTMRLLDGLMFQTPQEMGLIVIAARQTEGKGRG GNVWLSPVGCALSTLLISIPLRSQLGQRIPFVQHLMSVAVVEA VRSIPEYQDINLRVKWPNDIYYSDLMKIGGVLVNSTLMGETF YILIGCGFNVTNSNPTICINDLITEYNKQHKAELKPLRADYLIA RVVTVLEKLIKEFQDKGPNSVLPLYYRYWVHSGQQVHLGSA EGPKVSIVGLDDSGFLQVHQEGGEVVTVHPDGNSFDMLRNLI LPKRR	HLCS
106	MLKFRTVHGGLRLLGIRRTSTAPAASPNVRRLEYKPIKKVMV ANRGEIAIRVFRACTELGIRTVAIYSEQDTGQMHRQKADEAY LIGRGLAPVQAYLHIPDIIKVAKENNVDAVHPGYGFLSERADF AQACQDAGVRFIGPSPEVVRKMGDKVEARAIAIAAGVPVVPG	PC

[0840]

	TDAPITSLHEAHEFSNTYGFPIIFKAAYGGGGRGMRVVHSYEE LEENYTRAYSEALAAFGNGALFVEKFIEKPRHIEVQILGDQYG NILHLYERDCSIQRRHQKVVEIAPAAHLDPQLRTRLTSDSVKL AKQVGYENAGTVEFLVDRHGKHYFIEVNSRLQVEHTVTEEIT DVDLVHAQIHVAEGRSLPDLGLRQENIRINGCAIQCRVTTEDP ARSFQPDTGRIEVFRSGEGMGIRLDNASAFQGAVISPHYDSLL VKVIAHGKDHPTAATKMSRALAEFRVRGVKTNIAFLQNVLN NQQFLAGTVDTQFIDENPELFQLRPAQNRAQKLLHYLGHVM VNGPTTPIPVKASPSPTDPVVPAVPIGPPPAGFRDILLREGPEGF ARAVRNHPGLLLMDTTFRDAHQSLLATRVRTHDLKKIAPYV AHNFSKLFSMENWGGATFDVAMRFLYECPWRRLQELRELIP NIPFQMLLRGANAVGYTNYPDNVVFKFCEVAKENGMDVFRV FDSLNYLPNMLLGMEAAGSAGGVVEAAISYTGDVADPSRTK YSLQYYMGLAEELVRAGTHILCIKDMAGLLKPTACTMLVSSL RDRFPDLPLHIHTHDTSGAGVAAMLACAQAGADVVDVAADS MSGMTSQPSMGALVACTRGTPLDTEVPMERVFDYSEYWEGA RGLYAAFDCTATMKSGNSDVYENEIPGGQYTNLHFQAHSMG LGSKFKEVKKAYVEANQMLGDLIKVTPSSKIVGDLAQFMVQ NGLSRAEAEAQAEELSFPRSVVEFLQGYIGVPHGGFPEPFRSK VLKDLPRVEGRPGASLPPLDLQALEKELVDRHGEEVTPEDVL SAAMYPDVFAHFKDFTATFGPLDSLNTRLFLQGPKIAEEFEVE LERGKTLHIKALAVSDLNRAGQRQVFFELNGQLRSILVKDTQ AMKEMHFHPKALKDVKGQIGAPMPGKVIDIKVVAGAKVAK GQPLCVLSAMKMETVVTSPMEGTVRKVHVTKDMTLEGDDLI LEIE	
107	MVDSTEYEVASQPEVETSPLGDGASPGPEQVKLKKEISLLNG VCLIVGNMIGSGIFVSPKGVLIYSASFGLSLVIWAVGGLFSVFG ALCYAELGTTIKKSGASYAYILEAFGGFLAFIRLWTSLLIIEPTS QAIIAITFANYMVQPLFPSCFAPYAASRLLAAACICLLTFINCA YVKWGTLVQDIFTYAKVLALIAVIVAGIVRLGQGASTHFENS FEGSSFAVGDIALALYSALFSYSGWDTLNYVTEEIKNPERNLP LSIGISMPIVTIIYILTNVAYYTVLDMRDILASDAVAVTFADQIF GIFNWIIPLSVALSCFGGLNASIVAASRLFFVGSREGHLPDAIC MIHVERFTPVPSLLFNGIMALIYLCVEDIFQLINYYSFSYWFFV GLSIVGQLYLRWKEPDRPRPLKLSVFFPIVFCLCTIFLVAVPLY SDTINSLIGIAIALSGLPFYFLIIRVPEHKRPLYLRRIVGSATRYL QVLCMSVAAEMDLEDGGEMPKQRDPKSN	SLC7A7
108	MVPRLLLRAWPRGPAVGPGAPSRPLSAGSGPGQYLQRSIVPT MHYQDSLPRLPIPKLEDTIRRYLSAQKPLLNDGQFRKTEQFCK SFENGIGKELHEQLVALDKQNKHTSYISGPWFDMYLSARDSV VLNFNPFMAFNPDPKSEYNDQLTRATNMTVSAIRFLKTLRAG LLEPEVFHLNPAKSDTITFKRLIRFVPSSLSWYGAYLVNAYPL DMSQYFRLFNSTRLPKPSRDELFTDDKARHLLVLRKGNFYIFD VLDQDGNIVSPSEIQAHLKYILSDSSPAPEFPLAYLTSENRDIW AELRQKLMSSGNEESLRKVDSAVFCLCLDDFPIKDLVHLSHN MLHGDGTNRWFDKSFNLIIAKDGSTAVHFEHSWGDGVAVLR FFNEVFKDSTQTPAVTPQSQPATTDSTVTVQKLNFELTDALKT GITAAKEKFDATMKTLTIDCVQFQRGGKEFLKKQKLSPDAVA QLAFQMAFLRQYGQTVATYESCSTAAFKHGRTETIRPASVYT KRCSEAFVREPSRHSAGELQQMMVECSKYHGQLTKEAAMG QGFDRHLFALRHLAAAKGIILPELYLDPAYGQINHNVLSTSTL SSPAVNLGGFAPVVSDGFGVGYAVHDNWIGCNVSSYPGRNA REFLQCVEKALEDMFDALEGKSIKS	CPT2
109	MAAGFGRCCRVLRSISRFHWRSQHTKANRQREPGLGFSFEFT	ACADM

[0841]

	EQQKEFQATARKFAREEIIPVAAEYDKTGEYPVPLIRRAWELG LMNTHIPENCGGLGLGTFDACLISEELAYGCTGVQTAIEGNSL GQMPIIIAGNDQQKKKYLGRMTEEPLMCAYCVTEPGAGSDV AGIKTKAEKKGDEYIINGQKMWITNGGKANWYFLLARSDPD PKAPANKAFTGFIVEADTPGIQIGRKELNMGQRCSDTRGIVFE DVKVPKENVLIGDGAGFKVAMGAFDKTRPVVAAGAVGLAQ RALDEATKYALERKTFGKLLVEHQAISFMLAEMAMKVELAR MSYQRAAWEVDSGRRNTYYASIAKAFAGDIANQLATDAVQI LGGNGFNTEYPVEKLMRDAKIYQIYEGTSQIQRLIVAREHIDK YKN	
110	MAAALLARASGPARRALCPRAWRQLHTIYQSVELPETHQML LQTCRDFAEKELFPIAAQVDKEHLFPAAQVKKMGGLGLLAM DVPEELGGAGLDYLAYAIAMEEISRGCASTGVIMSVNNSLYL GPILKFGSKEQKQAWVTPFTSGDKIGCFALSEPGNGSDAGAA STTARAEGDSWVLNGTKAWITNAWEASAAVVFASTDRALQ NKGISAFLVPMPTPGLTLGKKEDKLGIRGSSTANLIFEDCRIPK DSILGEPGMGFKIAMQTLDMGRIGIASQALGIAQTALDCAVN YAENRMAFGAPLTKLQVIQFKLADMALALESARLLTWRAAM LKDNKKPFIKEAAMAKLAASEAATAISHQAIQILGGMGYVTE MPAERHYRDARITEIYEGTSEIQRLVIAGHLLRSYRS	ACADS
111	MQAARMAASLGRQLLRLGGGSSRLTALLGQPRPGPARRPYA GGAAQLALDKSDSHPSDALTRKKPAKAESKSFAVGMFKGQL TTDQVFPYPSVLNEEQTQFLKELVEPVSRFFEEVNDPAKNDAL EMVEETTWQGLKELGAFGLQVPSELGGVGLCNTQYARLVEI VGMHDLGVGITLGAHQSIGFKGILLFGTKAQKEKYLPKLASG ETVAAFCLTEPSSGSDAASIRTSAVPSPCGKYYTLNGSKLWIS NGGLADIFTVFAKTPVTDPATGAVKEKITAFVVERGFGGITHG PPEKKMGIKASNTAEVFFDGVRVPSENVLGEVGSGFKVAMHI LNNGRFGMAAALAGTMRGIIAKAVDHATNRTQFGEKIHNFG LIQEKLARMVMLQYVTESMAYMVSANMDQGATDFQIEAAIS KIFGSEAAWKVTDECIQIMGGMGFMKEPGVERVLRDLRIFRIF EGTNDILRLFVALQGCMDKGKELSGLGSALKNPFGNAGLLLG EAGKQLRRRAGLGSGLSLSGLVHPELSRSGELAVRALEQFAT VVEAKLIKHKKGIVNEQFLLQRLADGAIDLYAMVVVLSRASR SLSEGHPTAQHEKMLCDTWCIEAAARIREGMAALQSDPWQQ ELYRNFKSISKALVERGGVVTSNPLGF	ACADVL
112	MGHSKQIRILLLNEMEKLEKTLFRLEQGYELQFRLGPTLQGK AVTVYTNYPFPGETFNREKFRSLDWENPTEREDDSDKYCKLN LQQSGSFQYYFLQGNEKSGGGYIVVDPILRVGADNHVLPLDC VTLQTFLAKCLGPFDEWESRLRVAKESGYNMIHFTPLQTLGL SRSCYSLANQLELNPDFSRPNRKYTWNDVGQLVEKLKKEWN VICITDVVYNHTAANSKWIQEHPECAYNLVNSPHLKPAWVLD RALWRFSCDVAEGKYKEKGIPALIENDHHMNSIRKIIWEDIFP KLKLWEFFQVDVNKAVEQFRRLLTQENRRVTKSDPNQHLTII QDPEYRRFGCTVDMNIALTTFIPHDKGPAAIEECCNWFHKRM EELNSEKHRLINYHQEQAVNCLLGNVFYERLAGHGPKLGPVT RKHPLVTRYFTFPFEEIDFSMEESMIHLPNKACFLMAHNGWV MGDDPLRNFAEPGSEVYLRRELICWGDSVKLRYGNKPEDCP YLWAHMKKYTEITATYFQGVRLDNCHSTPLHVAEYMLDAA RNLQPNLYVVAELFTGSEDLDNVFVTRLGISSLIREAMSAYNS HEEGRLVYRYGGEPVGSFVQPCLRPLMPAIAHALFMDITHDN ECPIVHRSAYDALPSTTIVSMACCASGSTRGYDELVPHQISVV SEERFYTKWNPEALPSNTGEVNFQSGIIAARCAISKLHQELGA KGFIQVYVDQVDEDIVAVTRHSPSIHQSVVAVSRTAFRNPKTS	AGL

[0842]

	FYSKEVPQMCIPGKIEEVVLEARTIERNTKPYRKDENSINGTPD ITVEIREHIQLNESKIVKQAGVATKGPNEYIQEIEFENLSPGSVII FRVSLDPHAQVAVGILRNHLTQFSPHFKSGSLAVDNADPILKI PFASLASRLTLAELNQILYRCESEEKEDGGGCYDIPNWSALKY AGLQGLMSVLAEIRPKNDLGHPFCNNLRSGDWMIDYVSNRLI SRSGTIAEVGKWLQAMFFYLKQIPRYLIPCYFDAILIGAYTTLL DTAWKQMSSFVQNGSTFVKHLSLGSVQLCGVGKFPSLPILSP ALMDVPYRLNEITKEKEQCCVSLAAGLPHFSSGIFRCWGRDT FIALRGILLITGRYVEARNIILAFAGTLRHGLIPNLLGEGIYARY NCRDAVWWWLQCIQDYCKMVPNGLDILKCPVSRMYPTDDS APLPAGTLDQPLFEVIQEAMQKHMQGIQFRERNAGPQIDRNM KDEGFNITAGVDEETGFVYGGNRFNCGTWMDKMGESDRAR NRGIPATPRDGSAVEIVGLSKSAVRWLLELSKKNIFPYHEVTV KRHGKAIKVSYDEWNRKIQDNFEKLFHVSEDPSDLNEKHPNL VHKRGIYKDSYGASSPWCDYQLRPNFTIAMVVAPELFTTEKA WKALEIAEKKLLGPLGMKTLDPDDMVYCGIYDNALDNDNY NLAKGFNYHQGPEWLWPIGYFLRAKLYFSRLMGPETTAKTIV LVKNVLSRHYVHLERSPWKGLPELTNENAQYCPFSCETQAW SIATILETLYDL	
113	MEEGMNVLHDFGIQSTHYLQVNYQDSQDWFILVSVIADLRN AFYVLFPIWFHLQEAVGIKLLWVAVIGDWLNLVFKWILFGQR PYWWVLDTDYYSNTSVPLIKQFPVTCETGPGSPSGHAMGTA GVYYVMVTSTLSIFQGKIKPTYRFRCLNVILWLGFWAVQLNV CLSRIYLAAHFPHQVVAGVLSGIAVAETFSHIHSIYNASLKKY FLITFFLFSFAIGFYLLLKGLGVDLLWTLEKAQRWCEQPEWV HIDTTPFASLLKNLGTLFGLGLALNSSMYRESCKGKLSKWLPF RLSSIVASLVLLHVFDSLKPPSQVELVFYVLSFCKSAVVPLAS VSVIPYCLAQVLGQPHKKSL	G6PC
114	MAAPMTPAARPEDYEAALNAALADVPELARLLEIDPYLKPY AVDFQRRYKQFSQILKNIGENEGGIDKFSRGYESFGVHRCAD GGLYCKEWAPGAEGVFLTGDFNGWNPFSYPYKKLDYGKWE LYIPPKQNKSVLVPHGSKLKVVITSKSGEILYRISPWAKYVVR EGDNVNYDWIHWDPEHSYEFKHSRPKKPRSLRIYESHVGISS HEGKVASYKHFTCNVLPRIKGLGYNCIQLMAIMEHAYYASFG YQITSFFAASSRYGTPEELQELVDTAHSMGIIVLLDVVHSHAS KNSADGLNMFDGTDSCYFHSGPRGTHDLWDSRLFAYSSWEI LRFLLSNIRWWLEEYRFDGFRFDGVTSMLYHHHGVGQGFSG DYSEYFGLQVDEDALTYLMLANHLVHTLCPDSITIAEDVSGM PALCSPISQGGGGFDYRLAMAIPDKWIQLLKEFKDEDWNMG DIVYTLTNRRYLEKCIAYAESHDQALVGDKSLAFWLMDAEM YTNMSVLTPFTPVIDRGIQLHKMIRLITHGLGGEGYLNFMGNE FGHPEWLDFPRKGNNESYHYARRQFHLTDDDLLRYKFLNNF DRDMNRLEERYGWLAAPQAYVSEKHEGNKIIAFERAGLLFIF NFHPSKSYTDYRVGTALPGKFKIVLDSDAAEYGGHQRLDHST DFFSEAFEHNGRPYSLLVYIPSRVALILQNVDLPN	GBE1
115	MRSRSNSGVRLDGYARLVQQTILCHQNPVTGLLPASYDQKD AWVRDNVYSILAVWGLGLAYRKNADRDEDKAKAYELEQSV VKLMRGLLHCMIRQVDKVESFKYSQSTKDSLHAKYNTKTCA TVVGDDQWGHLQLDATSVYLLFLAQMTASGLHIIHSLDEVNF IQNLVFYIEAAYKTADFGIWERGDKTNQGISELNASSVGMAK AALEALDELDLFGVKGGPQSVIHVLADEVQHCQSILNSLLPR ASTSKEVDASLLSVVSFPAFAVEDSQLVELTKQEIITKLQGRY GCCRFLRDGYKTPKEDPNRLYYEPAELKLFENIECEWPLFWT YFILDGVFSGNAEQVQEYKEALEAVLIKGKNGVPLLPELYSVP	PHKA1

[0843]

	PDRVDEEYQNPHTVDRVPMGKLPHMWGQSLYILGSLMAEGF LAPGEIDPLNRRFSTVPKPDVVVQVSILAETEEIKTILKDKGIY VETIAEVYPIRVQPARILSHIYSSLGCNNRMKLSGRPYRHMGV LGTSKLYDIRKTIFTFTPQFIDQQQFYLALDNKMIVEMLRTDL SYLCSRWRMTGQPTITFPISHSMLDEDGTSLNSSILAALRKMQ DGYFGGARVQTGKLSEFLTTSCCTHLSFMDPGPEGKLYSEDY DDNYDYLESGNWMNDYDSTSHARCGDEVARYLDHLLAHTA PHPKLAPTSQKGGLDRFQAAVQTTCDLMSLVTKAKELHVQN VHMYLPTKLFQASRPSFNLLDSPHPRQENQVPSVRVEIHLPRD QSGEVDFKALVLQLKETSSLQEQADILYMLYTMKGPDWNTE LYNERSATVRELLTELYGKVGEIRHWGLIRYISGILRKKVEAL DEACTDLLSHQKHLTVGLPPEPREKTISAPLPYEALTQLIDEAS EGDMSISILTQEIMVYLAMYMRTQPGLFAEMFRLRIGLIIQVM ATELAHSLRCSAEEATEGLMNLSPSAMKNLLHHILSGKEFGV ERSVRPTDSNVSPAISIHEIGAVGATKTERTGIMQLKSEIKQVE FRRLSISAESQSPGTSMTPSSGSFPSAYDQQSSKDSRQGQWQR RRRLDGALNRVPVGFYQKVWKVLQKCHGLSVEGFVLPSSTT REMTPGEIKFSVHVESVLNRVPQPEYRQLLVEAILVLTMLADI EIHSIGSIIAVEKIVHIANDLFLQEQKTLGADDTMLAKDPASGI CTLLYDSAPSGRFGTMTYLSKAAATYVQEFLPHSICAMQ	
116	MRSRSNSGVRLDGYARLVQQTILCYQNPVTGLLSASHEQKD AWVRDNIYSILAVWGLGMAYRKNADRDEDKAKAYELEQNV VKLMRGLLQCMMRQVAKVEKFKHTQSTKDSLHAKYNTATC GTVVGDDQWGHLQVDATSLFLLFLAQMTASGLRIIFTLDEVA FIQNLVFYIEAAYKVADYGMWERGDKTNQGIPELNASSVGM AKAALEAIDELDLFGAHGGRKSVIHVLPDEVEHCQSILFSMLP RASTSKEIDAGLLSIISFPAFAVEDVNLVNVTKNEIISKLQGRY GCCRFLRDGYKTPREDPNRLHYDPAELKLFENIECEWPVFWT YFIIDGVFSGDAVQVQEYREALEGILIRGKNGIRLVPELYAVPP NKVDEEYKNPHTVDRVPMGKVPHLWGQSLYILSSLLAEGFL AAGEIDPLNRRFSTSVKPDVVVQVTVLAENNHIKDLLRKHGV NVQSIADIHPIQVQPGRILSHIYAKLGRNKNMNLSGRPYRHIG VLGTSKLYVIRNQIFTFTPQFTDQHHFYLALDNEMIVEMLRIE LAYLCTCWRMTGRPTLTFPISRTMLTNDGSDIHSAVLSTIRKL EDGYFGGARVKLGNLSEFLTTSFYTYLTFLDPDCDEKLFDNA SEGTFSPDSDSDLVGYLEDTCNQESQDELDHYINHLLQSTSLR SYLPPLCKNTEDRHVFSAIHSTRDILSVMAKAKGLEVPFVPMT LPTKVLSAHRKSLNLVDSPQPLLEKVPESDFQWPRDDHGDVD CEKLVEQLKDCSNLQDQADILYILYVIKGPSWDTNLSGQHGV TVQNLLGELYGKAGLNQEWGLIRYISGLLRKKVEVLAEACTD LLSHQKQLTVGLPPEPREKIISAPLPPEELTKLIYEASGQDISIA VLTQEIVVYLAMYVRAQPSLFVEMLRLRIGLIIQVMATELARS LNCSGEEASESLMNLSPFDMKNLLHHILSGKEFGVERSVRPIH SSTSSPTISIHEVGHTGVTKTERSGINRLRSEMKQMTRRFSADE QFFSVGQAASSSAHSSKSARSSTPSSPTGTSSSDSGGHHIGWG ERQGQWLRRRRLDGAINRVPVGFYQRVWKILQKCHGLSIDG YVLPSSTTREMTPHEIKFAVHVESVLNRVPQPEYRQLLVEAIM VLTLLSDTEMTSIGGIIHVDQIVQMASQLFLQDQVSIGAMDTL EKDQATGICHFFYDSAPSGAYGTMTYLTRAVASYLQELLPNS GCQMQ	PHKA2
117	MAGAAGLTAEVSWKVLERRARTKRSGSVYEPLKSINLPRPDN ETLWDKLDHYYRIVKSTLLLYQSPTTGLFPTKTCGGDQKAKI QDSLYCAAGAWALALAYRRIDDDKGRTHELEHSAIKCMRGI LYCYMRQADKVQQFKQDPRPTTCLHSVFNVHTGDELLSYEE	PHKB

[0844]

	YGHLQINAVSLYLLYLVEMISSGLQIIYNTDEVSFIQNLVFCVE RVYRVPDFGVWERGSKYNNGSTELHSSSVGLAKAALEAINGF NLFGNQGCSWSVIFVDLDAHNRNRQTLCSLLPRESRSHNTDA ALLPCISYPAFALDDEVLFSQTLDKVVRKLKGKYGFKRFLRD GYRTSLEDPNRCYYKPAEIKLFDGIECEFPIFFLYMMIDGVFRG NPKQVQEYQDLLTPVLHHTTEGYPVVPKYYYVPADFVEYEK NNPGSQKRFPSNCGRDGKLFLWGQALYIIAKLLADELISPKDI DPVQRYVPLKDQRNVSMRFSNQGPLENDLVVHVALIAESQR LQVFLNTYGIQTQTPQQVEPIQIWPQQELVKAYLQLGINEKLG LSGRPDRPIGCLGTSKIYRILGKTVVCYPIIFDLSDFYMSQDVF LLIDDIKNALQFIKQYWKMHGRPLFLVLIREDNIRGSRFNPILD MLAALKKGIIGGVKVHVDRLQTLISGAVVEQLDFLRISDTEEL PEFKSFEELEPPKHSKVKRQSSTPSAPELGQQPDVNISEWKDK PTHEILQKLNDCSCLASQAILLGILLKREGPNFITKEGTVSDHIE RVYRRAGSQKLWLAVRYGAAFTQKFSSSIAPHITTFLVHGKQ VTLGAFGHEEEVISNPLSPRVIQNIIYYKCNTHDEREAVIQQEL VIHIGWIISNNPELFSGMLKIRIGWIIHAMEYELQIRGGDKPAL DLYQLSPSEVKQLLLDILQPQQNGRCWLNRRQIDGSLNRTPT GFYDRVWQILERTPNGIIVAGKHLPQQPTLSDMTMYEMNFSL LVEDTLGNIDQPQYRQIVVELLMVVSIVLERNPELEFQDKVDL DRLVKEAFNEFQKDQSRLKEIEKQDDMTSFYNTPPLGKRGTC SYLTKAVMNLLLEGEVKPNNDDPCLIS	
118	MTLDVGPEDELPDWAAAKEFYQKYDPKDVIGRGVSSVVRRC VHRATGHEFAVKIMEVTAERLSPEQLEEVREATRRETHILRQ VAGHPHIITLIDSYESSSFMFLVFDLMRKGELFDYLTEKVALSE KETRSIMRSLLEAVSFLHANNIVHRDLKPENILLDDNMQIRLS DFGFSCHLEPGEKLRELCGTPGYLAPEILKCSMDETHPGYGKE VDLWACGVILFTLLAGSPPFWHRRQILMLRMIMEGQYQFSSP EWDDRSSTVKDLISRLLQVDPEARLTAEQALQHPFFERCEGS QPWNLTPRQRFRVAVWTVLAAGRVALSTHRVRPLTKNALLR DPYALRSVRHLIDNCAFRLYGHWVKKGEQQNRAALFQHRPP GPFPIMGPEEEGDSAAITEDEAVLVLG	PHKG2
119	MAAQGYGYYRTVIFSAMFGGYSLYYFNRKTFSFVMPSLVEEI PLDKDDLGFITSSQSAAYAISKFVSGVLSDQMSARWLFSSGLL LVGLVNIFFAWSSTVPVFAALWFLNGLAQGLGWPPCGKVLR KWFEPSQFGTWWAILSTSMNLAGGLGPILATILAQSYSWRST LALSGALCVVVSFLCLLLIHNEPADVGLRNLDPMPSEGKKGS LKEESTLQELLLSPYLWVLSTGYLVVFGVKTCCTDWGQFFLI QEKGQSALVGSSYMSALEVGGLVGSIAAGYLSDRAMAKAGL SNYGNPRHGLLLFMMAGMTVSMYLFRVTVTSDSPKLWILVL GAVFGFSSYGPIALFGVIANESAPPNLCGTSHAIVGLMANVGG FLAGLPFSTIAKHYSWSTAFWVAEVICAASTAAFFLLRNIRTK MGRVSKKAE	SLC37A4
120	MAAPGPALCLFDVDGTLTAPRQKITKEMDDFLQKLRQKIKIG VVGGSDFEKVQEQLGNDVVEKYDYVFPENGLVAYKDGKLL CRQNIQSHLGEALIQDLINYCLSYIAKIKLPKKRGTFIEFRNGM LNVSPIGRSCSQEERIEFYELDKKENIRQKFVADLRKEFAGKG LTFSIGGQISFDVFPDGWDKRYCLRHVENDGYKTIYFFGDKT MPGGNDHEIFTDPRTMGYSVTAPEDTRRICELLFS	PMM2
121	MPSETPQAEVGPTGCPHRSGPHSAKGSLEKGSPEDKEAKEPL WIRPDAPSRCTWQLGRPASESPHHHTAPAKSPKILPDILKKIG DTPMVRINKIGKKFGLKCELLAKCEFFNAGGSVKDRISLRMIE DAERDGTLKPGDTIIEPTSGNTGIGLALAAAVRGYRCIIVMPE KMSSEKVDVLRALGAEIVRTPTNARFDSPESHVGVAWRLKN	CBS

[0845]

	EIPNSHILDQYRNASNPLAHYDTTADEILQQCDGKLDMLVAS VGTGGTITGIARKLKEKCPGCRIIGVDPEGSILAEPEELNQTEQ TTYEVEGIGYDFIPTVLDRTVVDKWFKSNDEEAFTFARMLIA QEGLLCGGSAGSTVAVAVKAAQELQEGQRCVVILPDSVRNY MTKFLSDRWMLQKGFLKEEDLTEKKPWWWHLRVQELGLSA PLTVLPTITCGHTIEILREKGFDQAPVVDEAGVILGMVTLGNM LSSLLAGKVQPSDQVGKVIYKQFKQIRLTDTLGRLSHILEMDH FALVVHEQIQYHSTGKSSQRQMVFGVVTAIDLLNFVAAQERD QK	
122	MSFIPVAEDSDFPIHNLPYGVFSTRGDPRPRIGVAIGDQILDLSI IKHLFTGPVLSKHQDVFNQPTLNSFMGLGQAAWKEARVFLQ NLLSVSQARLRDDTELRKCAFISQASATMHLPATIGDYTDFYS SRQHATNVGIMFRDKENALMPNWLHLPVGYHGRASSVVVSG TPIRRPMGQMKPDDSKPPVYGACKLLDMELEMAFFVGPGNR LGEPIPISKAHEHIFGMVLMNDWSARDIQKWEYVPLGPFLGK SFGTTVSPWVVPMDALMPFAVPNPKQDPRPLPYLCHDEPYTF DINLSVNLKGEGMSQAATICKSNFKYMYWTMLQQLTHHSVN GCNLRPGDLLASGTISGPEPENFGSMLELSWKGTKPIDLGNGQ TRKFLLDGDEVIITGYCQGDGYRIGFGQCAGKVLPALLPS	FAH
123	MDPYMIQMSSKGNLPSILDVHVNVGGRSSVPGKMKGRKAR WSVRPSDMAKKTFNPIRAIVDNMKVKPNPNKTMISLSIGDPT VFGNLPTDPEVTQAMKDALDSGKYNGYAPSIGFLSSREEIASY YHCPEAPLEAKDVILTSGCSQAIDLCLAVLANPGQNILVPRPG FSLYKTLAESMGIEVKLYNLLPEKSWEIDLKQLEYLIDEKTAC LIVNNPSNPCGSVFSKRHLQKILAVAARQCVPILADEIYGDMV FSDCKYEPLATLSTDVPILSCGGLAKRWLVPGWRLGWILIHD RRDIFGNEIRDGLVKLSQRILGPCTIVQGALKSILCRTPGEFYH NTLSFLKSNADLCYGALAAIPGLRPVRPSGAMYLMVGIEMEH FPEFENDVEFTERLVAEQSVHCLPATCFEYPNFIRVVITVPEV MMLEACSRIQEFCEQHYHCAEGSQEECDK	TAT
124	MSRSGTDPQQRQQASEADAAAATFRANDHQHIRYNPLQDEW VLVSAHRMKRPWQGQVEPQLLKTVPRHDPLNPLCPGAIRAN GEVNPQYDSTFLFDNDFPALQPDAPSPGPSDHPLFQAKSARG VCKVMCFHPWSDVTLPLMSVPEIRAVVDAWASVTEELGAQY PWVQIFENKGAMMGCSNPHPHCQVWASSFLPDIAQREERSQ QAYKSQHGEPLLMEYSRQELLRKERLVLTSEHWLVLVPFWA TWPYQTLLLPRRHVRRLPELTPAERDDLASIMKKLLTKYDNL FETSFPYSMGWHGAPTGSEAGANWNHWQLHAHYYPPLLRS ATVRKFMVGYEMLAQAQRDLTPEQAAERLRALPEVHYHLG QKDRETATIA	GALT
125	MAALRQPQVAELLAEARRAFREEFGAEPELAVSAPGRVNLIG EHTDYNQGLVLPMALELMTVLVGSPRKDGLVSLLTTSEGAD EPQRLQFPLPTAQRSLEPGTPRWANYVKGVIQYYPAAPLPGFS AVVVSSVPLGGGLSSSASLEVATYTFLQQLCPDSGTIAARAQV CQQAEHSFAGMPCGIMDQFISLMGQKGHALLIDCRSLETSLV PLSDPKLAVLITNSNVRHSLASSEYPVRRRQCEEVARALGKES LREVQLEELEAARDLVSKEGFRRARHVVGEIRRTAQAAAALR RGDYRAFGRLMVESHRSLRDDYEVSCPELDQLVEAALAVPG VYGSRMTGGGFGGCTVTLLEASAAPHAMRHIQEHYGGTATF YLSQAADGAKVLCL	GALK1
126	MAEKVLVTGGAGYIGSHTVLELLEAGYLPVVIDNFHNAFRGG GSLPESLRRVQELTGRSVEFEEMDILDQGALQRLFKKYSFMA VIHFAGLKAVGESVQKPLDYYRVNLTGTIQLLEIMKAHGVKN	GALE

[0846]

	LVFSSSATVYGNPQYLPLDEAHPTGGCTNPYGKSKFFIEEMIR DLCQADKTWNAVLLRYFNPTGAHASGCIGEDPQGIPNNLMP YVSQVAIGRREALNVFGNDYDTEDGTGVRDYIHVVDLAKGH IAALRKLKEQCGCRIYNLGTGTGYSVLQMVQAMEKASGKKIP YKVVARREGDVAACYANPSLAQEELGWTAALGLDRMCEDL WRWQKQNPSGFGTQA	
127	MAEQVALSRTQVCGILREELFQGDAFHQSDTHIFIIMGASGDL AKKKIYPTIWWLFRDGLLPENTFIVGYARSRLTVADIRKQSEP FFKATPEEKLKLEDFFARNSYVAGQYDDAASYQRLNSHMNA LHLGSQANRLFYLALPPTVYEAVTKNIHESCMSQIGWNRIIVE KPFGRDLQSSDRLSNHISSLFREDQIYRIDHYLGKEMVQNLMV LRFANRIFGPIWNRDNIACVILTFKEPFGTEGRGGYFDEFGIIR DVMQNHLLQMLCLVAMEKPASTNSDDVRDEKVKVLKCISEV QANNVVLGQYVGNPDGEGEATKGYLDDPTVPRGSTTATFAA VVLYVENERWDGVPFILRCGKALNERKAEVRLQFHDVAGDI FHQQCKRNELVIRVQPNEAVYTKMMTKKPGMFFNPEESELD LTYGNRYKNVKLPDAYERLILDVFCGSQMHFVRSDELREAW RIFTPLLHQIELEKPKPIPYIYGSRGPTEADELMKRVGFQYEGT YKWVNPHKL	G6PD
128	MAEDKSKRDSIEMSMKGCQTNNGFVHNEDILEQTPDPGSSTD NLKHSTRGILGSQEPDFKGVQPYAGMPKEVLFQFSGQARYRI PREILFWLTVASVLVLIAATIAIIALSPKCLDWWQEGPMYQIYP RSFKDSNKDGNGDLKGIQDKLDYITALNIKTVWITSFYKSSLK DFRYGVEDFREVDPIFGTMEDFENLVAAIHDKGLKLIIDFIPNH TSDKHIWFQLSRTRTGKYTDYYIWHDCTHENGKTIPPNNWLS VYGNSSWHFDEVRNQCYFHQFMKEQPDLNFRNPDVQEEIKEI LRFWLTKGVDGFSLDAVKFLLEAKHLRDEIQVNKTQIPDTVT QYSELYHDFTTTQVGMHDIVRSFRQTMDQYSTEPGRYRFMG TEAYAESIDRTVMYYGLPFIQEADFPFNNYLSMLDTVSGNSV YEVITSWMENMPEGKWPNWMIGGPDSSRLTSRLGNQYVNV MNMLLFTLPGTPITYYGEEIGMGNIVAANLNESYDINTLRSKS PMQWDNSSNAGFSEASNTWLPTNSDYHTVNVDVQKTQPRSA LKLYQDLSLLHANELLLNRGWFCHLRNDSHYVVYTRELDGI DRIFIVVLNFGESTLLNLHNMISGLPAKMRIRLSTNSADKGSK VDTSGIFLDKGEGLIFEHNTKNLLHRQTAFRDRCFVSNRACYS SVLNILYTSC	SLC3A1
129	MGDTGLRKRREDEKSIQSQEPKTTSLQKELGLISGISIIVGTIIG SGIFVSPKSVLSNTEAVGPCLIIWAACGVLATLGALCFAELGT MITKSGGEYPYLMEAYGPIPAYLFSWASLIVIKPTSFAIICLSFS EYVCAPFYVGCKPPQIVVKCLAAAAILFISTVNSLSVRLGSYV QNIFTAAKLVIVAIIIISGLVLLAQGNTKNFDNSFEGAQLSVGAI SLAFYNGLWAYDGWNQLNYITEELRNPYRNLPLAIIIGIPLVT ACYILMNVSYFTVMTATELLQSQAVAVTFGDRVLYPASWIVP LFVAFSTIGAANGTCFTAGRLIYVAGREGHMLKVLSYISVRRL TPAPAIIFYGIIATIYIIPGDINSLVNYFSFAAWLFYGLTILGLIV MRFTRKELERPIKVPVVIPVLMTLISVFLVLAPIISKPTWEYLY CVLFILSGLLFYFLFVHYKFGWAQKISKPITMHLQMLMEVVPP EEDPE	SLC7A9
130	MVNEARGNSSLNPCLEGSASSGSESSKDSSRCSTPGLDPERHE RLREKMRRRLESGDKWFSLEFFPPRTAEGAVNLISRFDRMAA GGPLYIDVTWHPAGDPGSDKETSSMMIASTAVNYCGLETILH MTCCRQRLEEITGHLHKAKQLGLKNIMALRGDPIGDQWEEEE GGFNYAVDLVKHIRSEFGDYFDICVAGYPKGHPEAGSFEADL KHLKEKVSAGADFIITQLFFEADTFFRFVKACTDMGITCPIVPG	MTHFR

[0847]

	IFPIQGYHSLRQLVKLSKLEVPQEIKDVIEPIKDNDAAIRNYGIE LAVSLCQELLASGLVPGLHFYTLNREMATTEVLKRLGMWTE DPRRPLPWALSAHPKRREEDVRPIFWASRPKSYIYRTQEWDE FPNGRWGNSSSPAFGELKDYYLFYLKSKSPKEELLKMWGEEL TSEESVFEVFVLYLSGEPNRNGHKVTCLPWNDEPLAAETSLL KEELLRVNRQGILTINSQPNINGKPSSDPIVGWGPSGGYVFQK AYLEFFTSRETAEALLQVLKKYELRVNYHLVNVKGENITNAP ELQPNAVTWGIFPGREIIQPTVVDPVSFMFWKDEAFALWIER WGKLYEEESPSRTIIQYIHDNYFLVNLVDNDFPLDNCLWQVV EDTLELLNRPTQNARETEAP	
131	MSPALQDLSQPEGLKKTLRDEINAILQKRIMVLDGGMGTMIQ REKLNEEHFRGQEFKDHARPLKGNNDILSITQPDVIYQIIHKEY LLAGADIIETNTFSSTSIAQADYGLEHLAYRMNMCSAGVARK AAEEVTLQTGIKRFVAGALGPTNKTLSVSPSVERPDYRNITFD ELVEAYQEQAKGLLDGGVDILLIETIFDTANAKAALFALQNLF EEKYAPRPIFISGTIVDKSGRTLSGQTGEGFVISVSHGEPLCIGL NCALGAAEMRPFIEIIGKCTTAYVLCYPNAGLPNTFGDYDETP SMMAKHLKDFAMDGLVNIVGGCCGSTPDHIREIAEAVKNCK PRVPPATAFEGHMLLSGLEPFRIGPYTNFVNIGERCNVAGSRK FAKLIMAGNYEEALCVAKVQVEMGAQVLDVNMDDGMLDG PSAMTRFCNLIASEPDIAKVPLCIDSSNFAVIEAGLKCCQGKCI VNSISLKEGEDDFLEKARKIKKYGAAMVVMAFDEEGQATET DTKIRVCTRAYHLLVKKLGFNPNDIIFDPNILTIGTGMEEHNL YAINFIHATKVIKETLPGARISGGLSNLSFSFRGMEAIREAMHG VFLYHAIKSGMDMGIVNAGNLPVYDDIHKELLQLCEDLIWNK DPEATEKLLRYAQTQGTGGKKVIQTDEWRNGPVEERLEYAL VKGIEKHIIEDTEEARLNQKKYPRPLNIIEGPLMNGMKIVGDLF GAGKMFLPQVIKSARVMKKAVGHLIPFMEKEREETRVLNGT VEEEDPYQGTIVLATVKGDVHDIGKNIVGVVLGCNNFRVIDL GVMTPCDKILKAALDHKADIIGLSGLITPSLDEMIFVAKEMER LAIRIPLLIGGATTSKTHTAVKIAPRYSAPVIHVLDASKSVVVC SQLLDENLKDEYFEEIMEEYEDIRQDHYESLKERRYLPLSQAR KSGFQMDWLSEPHPVKPTFIGTQVFEDYDLQKLVDYIDWKPF FDVWQLRGKYPNRGFPKIFNDKTVGGEARKVYDDAHNMLN TLISQKKLRARGVVGFWPAQSIQDDIHLYAEAAVPQAAEPIAT FYGLRQQAEKDSASTEPYYCLSDFIAPLHSGIRDYLGLFAVAC FGVEELSKAYEDDGDDYSSIMVKALGDRLAEAFAEELHERVR RELWAYCGSEQLDVADLRRLRYKGIRPAPGYPSQPDHTEKLT MWRLADIEQSTGIRLTESLAMAPASAVSGLYFSNLKSKYFAV GKISKDQVEDYALRKNISVAEVEKWLGPILGYDTD	MTR
132	MGAASVRAGARLVEVALCSFTVTCLEVMRRFLLLYATQQGQ AKAIAEEICEQAVVHGFSADLHCISESDKYDLKTETAPLVVVV STTGTGDPPDTARKFVKEIQNQTLPVDFFAHLRYGLLGLGDSE YTYFCNGGKIIDKRLQELGARHFYDTGHADDCVGLELVVEP WIAGLWPALRKHFRSSRGQEEISGALPVASPASSRTDLVKSEL LHIESQVELLRFDDSGRKDSEVLKQNAVNSNQSNVVIEDFESS LTRSVPPLSQASLNIPGLPPEYLQVHLQESLGQEESQVSVTSAD PVFQVPISKAVQLTTNDAIKTTLLVELDISNTDFSYQPGDAFSV ICPNSDSEVQSLLQRLQLEDKREHCVLLKIKADTKKKGATLPQ HIPAGCSLQFIFTWCLEIRAIPKKAFLRALVDYTSDSAEKRRLQ ELCSKQGAADYSRFVRDACACLLDLLLAFPSCQPPLSLLLEHL PKLQPRPYSCASSSLFHPGKLHFVFNIVEFLSTATTEVLRKGVC TGWLALLVASVLQPNIHASHEDSGKALAPKISISPRTTNSFHLP DDPSIPIIMVGPGTGIAPFIGFLQHREKLQEQHPDGNFGAMWL	MTRR

[0848]

	FFGCRHKDRDYLFRKELRHFLKHGILTHLKVSFSRDAPVGEEE APAKYVQDNIQLHGQQVARILLQENGHIYVCGDAKNMAKDV HDALVQIISKEVGVEKLEAMKTLATLKEEKRYLQDIWS	
133	MPEQERQITAREGASRKILSKLSLPTRAWEPAMKKSFAFDNV GYEGGLDGLGPSSQVATSTVRILGMTCQSCVKSIEDRISNLKG IISMKVSLEQGSATVKYVPSVVCLQQVCHQIGDMGFEASIAEG KAASWPSRSLPAQEAVVKLRVEGMTCQSCVSSIEGKVRKLQ GVVRVKVSLSNQEAVITYQPYLIQPEDLRDHVNDMGFEAAIK SKVAPLSLGPIDIERLQSTNPKRPLSSANQNFNNSETLGHQGSH VVTLQLRIDGMHCKSCVLNIEENIGQLLGVQSIQVSLENKTAQ VKYDPSCTSPVALQRAIEALPPGNFKVSLPDGAEGSGTDHRSS SSHSPGSPPRNQVQGTCSTTLIAIAGMTCASCVHSIEGMISQLE GVQQISVSLAEGTATVLYNPSVISPEELRAAIEDMGFEASVVS ESCSTNPLGNHSAGNSMVQTTDGTPTSVQEVAPHTGRLPANH APDILAKSPQSTRAVAPQKCFLQIKGMTCASCVSNIERNLQKE AGVLSVLVALMAGKAEIKYDPEVIQPLEIAQFIQDLGFEAAV MEDYAGSDGNIELTITGMTCASCVHNIESKLTRTNGITYASVA LATSKALVKFDPEIIGPRDIIKIIEEIGFHASLAQRNPNAHHLDH KMEIKQWKKSFLCSLVFGIPVMALMIYMLIPSNEPHQSMVLD HNIIPGLSILNLIFFILCTFVQLLGGWYFYVQAYKSLRHRSANM DVLIVLATSIAYVYSLVILVVAVAEKAERSPVTFFDTPPMLFV FIALGRWLEHLAKSKTSEALAKLMSLQATEATVVTLGEDNLII REEQVPMELVQRGDIVKVVPGGKFPVDGKVLEGNTMADESL ITGEAMPVTKKPGSTVIAGSINAHGSVLIKATHVGNDTTLAQI VKLVEEAQMSKAPIQQLADRFSGYFVPFIIIMSTLTLVVWIVIG FIDFGVVQRYFPNPNKHISQTEVIIRFAFQTSITVLCIACPCSLG LATPTAVMVGTGVAAQNGILIKGGKPLEMAHKIKTVMFDKT GTITHGVPRVMRVLLLGDVATLPLRKVLAVVGTAEASSEHPL GVAVTKYCKEELGTETLGYCTDFQAVPGCGIGCKVSNVEGIL AHSERPLSAPASHLNEAGSLPAEKDAVPQTFSVLIGNREWLRR NGLTISSDVSDAMTDHEMKGQTAILVAIDGVLCGMIAIADAV KQEAALAVHTLQSMGVDVVLITGDNRKTARAIATQVGINKV FAEVLPSHKVAKVQELQNKGKKVAMVGDGVNDSPALAQAD MGVAIGTGTDVAIEAADVVLIRNDLLDVVASIHLSKRTVRRIR INLVLALIYNLVGIPIAAGVFMPIGIVLQPWMGSAAMAASSVS VVLSSLQLKCYKKPDLERYEAQAHGHMKPLTASQVSVHIGM DDRWRDSPRATPWDQVSYVSQVSLSSLTSDKPSRHSAAADD DGDKWSLLLNGRDEEQYI	HPRT1
134	MATRSPGVVISDDEPGYDLDLFCIPNHYAEDLERVFIPHGLIM DRTERLARDVMKEMGGHHIVALCVLKGGYKFFADLLDYIKA LNRNSDRSIPMTVDFIRLKSYCNDQSTGDIKVIGGDDLSTLTG KNVLIVEDIIDTGKTMQTLLSLVRQYNPKMVKVASLLVKRTP RSVGYKPDFVGFEIPDKFVVGYALDYNEYFRDLNHVCVISET GKAKYKA	
135	MGEPGQSPSPRSSHGSPPTLSTLTLLLLLCGHAHSQCKILRCN AEYVSSTLSLRGGGSSGALRGGGGGGRGGGVGSGGLCRALR SYALCTRRTARTCRGDLAFHSAVHGIEDLMIQHNCSRQGPTA PPPPRGPALPGAGSGLPAPDPCDYEGRFSRLHGRPPGFLHCAS FGDPHVRSFHHHFHTCRVQGAWPLLDNDFLFVQATSSPMAL GANATATRKLTIIFKNMQECIDQKVYQAEVDNLPVAFEDGSI NGGDRPGGSSLSIQTANPGNHVEIQAAYIGTTIIIRQTAGQLSF SIKVAEDVAMAFSAEQDLQLCVGGCPPSQRLSRSERNRRGAI TIDTARRLCKEGLPVEDAYFHSCVFDVLISGDPNFTVAAQAAL EDARAFLPDLEKLHLFPSDAGVPLSSATLLAPLLSGLFVLWLC	HJV

[0849]

	IQ	
136	MALSSQIWAACLLLLLLLASLTSGSVFPQQTGQLAELQPQDR AGARASWMPMFQRRRRRDTHFPICIFCCGCCHRSKCGMCCK T	HAMP
137	MRSPRTRGRSGRPLSLLLALLCALRAKVCGASGQFELEILSMQ NVNGELQNGNCCGGARNPGDRKCTRDECDTYFKVCLKEYQS RVTAGGPCSFGSGSTPVIGGNTFNLKASRGNDRNRIVLPFSFA WPRSYTLLVEAWDSSNDTVQPDSIIEKASHSGMINPSRQWQT LKQNTGVAHFEYQIRVTCDDYYYGFGCNKFCRPRDDFFGHY ACDQNGNKTCMEGWMGPECNRAICRQGCSPKHGSCKLPGD CRCQYGWQGLYCDKCIPHPGCVHGICNEPWQCLCETNWGG QLCDKDLNYCGTHQPCLNGGTCSNTGPDKYQCSCPEGYSGP NCEIAEHACLSDPCHNRGSCKETSLGFECECSPGWTGPTCSTN IDDCSPNNCSHGGTCQDLVNGFKCVCPPQWTGKTCQLDANE CEAKPCVNAKSCKNLIASYYCDCLPGWMGQNCDININDCLG QCQNDASCRDLVNGYRCICPPGYAGDHCERDIDECASNPCLN GGHCQNEINRFQCLCPTGFSGNLCQLDIDYCEPNPCQNGAQC YNRASDYFCKCPEDYEGKNCSHLKDHCRTTPCEVIDSCTVAM ASNDTPEGVRYISSNVCGPHGKCKSQSGGKFTCDCNKGFTGT YCHENINDCESNPCRNGGTCIDGVNSYKCICSDGWEGAYCET NINDCSQNPCHNGGTCRDLVNDFYCDCKNGWKGKTCHSRDS QCDEATCNNGGTCYDEGDAFKCMCPGGWEGTTCNIARNSSC LPNPCHNGGTCVVNGESFTCVCKEGWEGPICAQNTNDCSPHP CYNSGTCVDGDNWYRCECAPGFAGPDCRININECQSSPCAFG ATCVDEINGYRCVCPPGHSGAKCQEVSGRPCITMGSVIPDGA KWDDDCNTCQCLNGRIACSKVWCGPRPCLLHKGHSECPSGQ SCIPILDDQCFVHPCTGVGECRSSSLQPVKTKCTSDSYYQDNC ANITFTFNKEMMSPGLTTEHICSELRNLNILKNVSAEYSIYIAC EPSPSANNEIHVAISAEDIRDDGNPIKEITDKIIDLVSKRDGNSS LIAAVAEVRVQRRPLKNRTDFLVPLLSSVLTVAWICCLVTAF YWCLRKRRKPGSHTHSASEDNTTNNVREQLNQIKNPIEKHGA NTVPIKDYENKNSKMSKIRTHNSEVEEDDMDKHQQKARFAK QPAYTLVDREEKPPNGTPTKHPNWTNKQDNRDLESAQSLNR MEYIV	TTR
138	MASHRLLLLCLAGLVFVSEAGPTGTGESKCPLMVKVLDAVR GSPAINVAVHVFRKAADDTWEPFASGKTSESGELHGLTTEEE FVEGIYKVEIDTKSYWKALGISPFHEHAEVVFTANDSGPRRYT IAALLSPYSYSTTAVVTNPKE	TTR
139	MASHKLLVTPPKALLKPLSIPNQLLLGPGPSNLPPRIMAAGGL QMIGSMSKDMYQIMDEIKEGIQYVFQTRNPLTLVISGSGHCAL EAALVNVLEPGDSFLVGANGIWGQRAVDIGERIGARVHPMT KDPGGHYTLQEVEEGLAQHKPVLLFLTHGESSTGVLQPLDGF GELCHRYKCLLLVDSVASLGGTPLYMDRQGIDILYSGSQKAL NAPPGTSLISFSDKAKKKMYSRKTKPFSFYLDIKWLANFWGC DDQPRMYHHTIPVISLYSLRESLALIAEQGLENSWRQHREAA AYLHGRLQALGLQLFVKDPALRLPTVTTVAVPAGYDWRDIV SYVIDHFDIEIMGGLGPSTGKVLRIGLLGCNATRENVDRVTEA LRAALQHCPKKKL	AGXT
140	MKMRFLGLVVCLVLWTLHSEGSGGKLTAVDPETNMNVSEIIS YWGFPSEEYLVETEDGYILCLNRIPHGRKNHSDKGPKPVVFL QHGLLADSSNWVTNLANSSLGFILADAGFDVWMGNSRGNT WSRKHKTLSVSQDEFWAFSYDEMAKYDLPASINFILNKTGQE QVYYVGHSQGTTIGFIAFSQIPELAKRIKMFFALGPVASVAFC	LIPA

[0850]

	TSPMAKLGRLPDHLIKDLFGDKEFLPQSAFLKWLGTHVCTHV ILKELCGNLCFLLCGFNERNLNMSRVDVYTTHSPAGTSVQNM LHWSQAVKFQKFQAFDWGSSAKNYFHYNQSYPPTYNVKDM LVPTAVWSGGHDWLADVYDVNILLTQITNLVFHESIPEWEHL DFIWGLDAPWRLYNKIINLMRKYQ	
141	MASRLTLLTLLLLLLAGDRASSNPNATSSSSQDPESLQDRGEG KVATTVISKMLFVEPILEVSSLPTTNSTTNSATKITANTTDEPT TQPTTEPTTQPTIQPTQPTTQLPTDSPTQPTTGSFCPGPVTLCSD LESHSTEAVLGDALVDFSLKLYHAFSAMKKVETNMAFSPFSI ASLLTQVLLGAGENTKTNLESILSYPKDFTCVHQALKGFTTK GVTSVSQIFHSPDLAIRDTFVNASRTLYSSSPRVLSNNSDANLE LINTWVAKNTNNKISRLLDSLPSDTRLVLLNAIYLSAKWKTTF DPKKTRMEPFHFKNSVIKVPMMNSKKYPVAHFIDQTLKAKV GQLQLSHNLSLVILVPQNLKHRLEDMEQALSPSVFKAIMEKL EMSKFQPTLLTLPRIKVTTSQDMLSIMEKLEFFDFSYDLNLCG LTEDPDLQVSAMQHQTVLELTETGVEAAAAASAISVARTLLVF EVQQPFLFVLWDQQHKFPVFMGRVYDPRA	SERPING1
142	MGSPLRFDGRVVLVTGAGAGLGRAYALAFAERGALVVVND LGGDFKGVGKGSLAADKVVEEIRRRGGKAVANYDSVEEGEK VVKTALDAFGRIDVVVNNAGILRDRSFARISDEDWDIIHRVHL RGSFQVTRAAWEHMKKQKYGRIIMTSSASGIYGNFGQANYS AAKLGLLGLANSLAIEGRKSNIHCNTIAPNAGSRMTQTVMPE DLVEALKPEYVAPLVLWLCHESCEENGGLFEVGAGWIGKLR WERTLGAIVRQKNHPMTPEAVKANWKKICDFENASKPQSIQE STGSIIEVLSKIDSEGGVSANHTSRATSTATSGFAGAIGQKLPPF SYAYTELEAIMYALGVGASIKDPKDLKFIYEGSSDFSCLPTFG VIIGQKSMMGGGLAEIPGLSINFAKVLHGEQYLELYKPLPRAG KLKCEAVVADVLDKGSGVVIIMDVYSYSEKELICHNQFSLFL VGSGGFGGKRTSDKVKVAVAIPNRPPDAVLTDTTSLNQAALY RLSGDWNPLHIDPNFASLAGFDKPILHGLCTFGFSARRVLQQF ADNDVSRFKAIKARFAKPVYPGQTLQTEMWKEGNRIHFQTK VQETGDIVISNAYVDLAPTSGTSAKTPSEGGKLQSTFVFEEIGR RLKDIGPEVVKKVNAVFEWHITKGGNIGAKWTIDLKSGSGKV YQGPAKGAADTTIILSDEDFMEVVLGKLDPQKAFFSGRLKAR GNIMLSQKLQMILKDYAKL	HSD17B4
143	MEANGLGPQGFPELKNDTFLRAAWGEETDYTPVWCMRQAG RYLPEFRETRAAQDFFSTCRSPEACCELTLQPLRRFPLDAAIIFS DILVVPQALGMEVTMVPGKGPSFPEPLREEQDLERLRDPEVV ASELGYVFQAITLTRQRLAGRVPLIGFAGAPWTLMTYMVEGG GSSTMAQAKRWLYQRPQASHQLLRILTDALVPYLVGQVVAG AQALQLFESHAGHLGPQLFNKFALPYIRDVAKQVKARLREAG LAPVPMIIFAKDGHFALEELAQAGYEVVGLDWTVAPKKARE CVGKTVTLQGNLDPCALYASEEEIGQLVKQMLDDFGPHRYIA NLGHGLYPDMDPEHVGAFVDAVHKHSRLLRQN	UROD
144	MGPRARPALLLLMLLQTAVLQGRLLRSHSLHYLFMGASEQD LGLSLFEALGYVDDQLFVFYDHESRRVEPRTPWVSSRISSQM WLQLSQSLKGWDHMFTVDFWTIMENHNHSKESHTLQVILGC EMQEDNSTEGYWKYGYDGQDHLEFCPDTLDWRAAEPRAWP TKLEWERHKIRARQNRAYLERDCPAQLQQLLELGRGVLDQQ VPPLVKVTHHVTSSVTTLRCRALNYYPQNITMKWLKDKQPM DAKEFEPKDVLPNGDGTYQGWITLAVPPGEEQRYTCQVEHP GLDQPLIVIWEPSPSGTLVIGVISGIAVFVVILFIGILFIILRKRQG SRGAMGHYVLAERE	HFE

[0851]

145	MESKALLVLTLAVWLQSLTASRGGVAAADQRRDFIDIESKFA LRTPEDTAEDTCHLIPGVAESVATCHFNHSSKTFMVIHGWTV TGMYESWVPKLVAALYKREPDSNVIVVDWLSRAQEHYPVSA GYTKLVGQDVARFINWMEEEFNYPLDNVHLLGYSLGAHAAG IAGSLTNKKVNRITGLDPAGPNFEYAEAPSRLSPDDADFVDVL HTFTRGSPGRSIGIQKPVGHVDIYPNGGTFQPGCNIGEAIRVIA ERGLGDVDQLVKCSHERSIHLFIDSLLNEENPSKAYRCSSKEA FEKGLCLSCRKNRCNNLGYEINKVRAKRSSKMYLKTRSQMP YKVFHYQVKIHFSGTESETHTNQAFEISLYGTVAESENIPFTLP EVSTNKTYSFLIYTEVDIGELLMLKLKWKSDSYFSWSDWWSS PGFAIQKIRVKAGETQKKVIFCSREKVSHLQKGKAPAVFVKC HDKSLNKKSG	LPL
146	MRPVRLMKVFVTRRIPAEGRVALARAADCEVEQWDSDEPIP AKELERGVAGAHGLLCLLSDHVDKRILDAAGANLKVISTMS VGIDHLALDEIKKRGIRVGYTPDVLTDTTAELAVSLLLTTCRR LPEAIEEVKNGGWTSWKPLWLCGYGLTQSTVGIIGLGRIGQAI ARRLKPFGVQRFLYTGRQPRPEEAAEFQAEFVSTPELAAQSDF IVVACSLTPATEGLCNKDFFQKMKETAVFINISRGDVVNQDD LYQALASGKIAAAGLDVTSPEPLPTNHPLLTLKNCVILPHIGSA THRTRNTMSLLAANNLLAGLRGEPMPSELKL	GRHPR
147	MLGPQVWSSVRQGLSRSLSRNVGVWASGEGKKVDIAGIYPP VTTPFTATAEVDYGKLEENLHKLGTFPFRGFVVQGSNGEFPFL TSSERLEVVSRVRQAMPKNRLLLAGSGCESTQATVEMTVSM AQVGADAAMVVTPCYYRGRMSSAALIHHYTKVADLSPIPVV LYSVPANTGLDLPVDAVVTLSQHPNIVGMKDSGGDVTRIGLI VHKTRKQDFQVLAGSAGFLMASYALGAVGGVCALANVLGA QVCQLERLCCTGQWEDAQKLQHRLIEPNAAVTRRFGIPGLKK IMDWFGYYGGPCRAPLQELSPAEEEALRMDFTSNGWL	HOGA1
148	MGPWGWKLRWTVALLLAAAGTAVGDRCERNEFQCQDGKCI SYKWVCDGSAECQDGSDESQETCLSVTCKSGDFSCGGRVNR CIPQFWRCDGQVDCDNGSDEQGCPPKTCSQDEFRCHDGKCIS RQFVCDSDRDCLDGSDEASCPVLTCGPASFQCNSSTCIPQLW ACDNDPDCEDGSDEWPQRCRGLYVFQGDSSPCSAFEFHCLSG ECIHSSWRCDGGPDCKDKSDEENCAVATCRPDEFQCSDGNCI HGSRQCDREYDCKDMSDEVGCVNVTLCEGPNKFKCHSGECI TLDKVCNMARDCRDWSDEPIKECGTNECLDNNGGCSHVCND LKIGYECLCPDGFQLVAQRRCEDIDECQDPDTCSQLCVNLEG GYKCQCEEGFQLDPHTKACKAVGSIAYLFFTNRHEVRKMTL DRSEYTSLIPNLRNVVALDTEVASNRIYWSDLSQRMICSTQLD RAHGVSSYDTVISRDIQAPDGLAVDWIHSNIYWTDSVLGTVS VADTKGVKRKTLFRENGSKPRAIVVDPVHGFMYWTDWGTP AKIKKGGLNGVDIYSLVTENIQWPNGITLDLLSGRLYWVDSK LHSISSIDVNGGNRKTILEDEKRLAHPFSLAVFEDKVFWTDIIN EAIFSANRLTGSDVNLLAENLLSPEDMVLFHNLTQPRGVNWC ERTTLSNGGCQYLCLPAPQINPHSPKFTCACPDGMLLARDMR SCLTEAEAAVATQETSTVRLKVSSTAVRTQHTTTRPVPDTSRL PGATPGLTTVEIVTMSHQALGDVAGRGNEKKPSSVRALSIVLP IVLLVFLCLGVFLLWKNWRLKNINSINFDNPVYQKTTEDEVHI CHNQDGYSYPSRQMVSLEDDVA	LDLR

[0852]

149	MLWSGCRRFGARLGCLPGGLRVLVQTGHRS LTSCIDPSMGLNEEQKEFQKVAFDFAAREM APNMAEWDQKELFPVDVMRKAAQLGFGGVY IQTDVGGSGLSRLDTSVIFEALATGCTSTT AYISIHNMCAWMIDSFGNEE QRHKFCPPLCTMEKFASYCLTEPGSGSDAA SLLTSAKKQGDHYILNGSKAFISGAGESDI YVVMCRTGGPGPKGISCIVVEKGTPGLSFG KKEKKVGWNSQPTRAVIFEDCAVPVANRIG SEGQGFLIAVRGLNGGRINIASCSLGAAHA SVILTRDHLNVRKQFGEPLASNQYLQFTLADMATRLVAARLMVRN AAVALQEERKDAVALCSMAKLFATDECFAICNQALQMHGGYGYL KDYAVQQYVRDSRVHQILEGSNEVMRILISRSLLQE	ACAD8
150	MEGLAVRLLRGSRLLRRNFLTCLSSWKIPPHVSKSSQSEALLN ITNNGIHFAPLQTFTDEEMMIKSSVKKFAQEQIAPLVSTMDEN SKMEKSVIQGLFQQGLMGIEVDPEYGGTGASFLSTVLVIEELA KVDASVAVFCEIQNTLINTLIRKHGTEEQKATYLPQLTTEKVG SFCLSEAGAGSDSFALKTRADKEGDYYVLNGSKMWISSAEH AGLFLVMANVDPTIGYKGITSFLVDRDTPGLHIGKPENKLGLR ASSTCPLTFENVKVPEANILGQIGHGYKYAIGSLNEGRIGIAAQ MLGLAQGCFDYTIPYIKERIQFGKRLFDFQGLQHQVAHVATQ LEAARLLTYNAARLLEAGKPFIKEASMAKYYASEIAGQTTSK CIEWMGGVGYTKDYPVEKYFRDAKIGTIYEGASNIQLNTIAK HIDAEY	ACADSB
151	MAVLAALLRSGARSRSPLLRRLVQEIRYVERSYVSKPTLKEV VIVSATRTPIGSFLGSLSLLPATKLGSIAIQGAIEKAGIPKEEVK EAYMGNVLQGGEGQAPTRQAVLGAGLPISTPCTTINKVCASG MKAIMMASQSLMCGHQDVMVAGGMESMSNVPYVMNRGST PYGGVKLEDLIVKDGLTDVYNKIHMGSCAENTAKKLNIARNE QDAYAINSYTRSKAAWEAGKFGNEVIPVTVTVKGQPDVVVK EDEEYKRVDFSKVPKLKTVFQKENGTVTAANASTLNDGAAA LVLMTADAAKRLNVTPLARIVAFADAAVEPIDFPIAPVYAAS MVLKDVGLKKEDIAMWEVNEAFSLVVLANIKMLEIDPQKVN INGGAVSLGHPIGMSGARIVGHLTHALKQGEYGLASICNGGG GASAMLIQKL	ACAT1
152	MLPHVVLTFRRLGCALASCRLAPARHRGSGLLHTAPVARSDR SAPVFTRALAFGDRIALDQHGRHTYRELYSRSLRLSQEICRLC GCVGGDLREERVSFLCANDASYVVAQWASWMSGGVAVPLY RKHPAAQLEYVICDSQSSVVLASQEYLELLSPVVRKLGVPLLP LTPAIYTGAVEEPAEVPVPEQGWRNKGAMIIYTSGTTGRPKG VLSTHQNIRAVVTGLVHKWAWTKDDVILHVLPLHHVHGVV NALLCPLWVGATCVMMPEFSPQQVWEKFLSSETPRINVFMA VPTIYTKLMEYYDRHFTQPHAQDFLRAVCEEKIRLMVSGSAA LPLPVLEKWKNITGHTLLERYGMTEIGMALSGPLTTAVRLPG SVGTPLPGVQVRIVSENPQREACSYTIHAEGDERGTKVTPGFE EKEGELLVRGPSVFREYWNKPEETKSAFTLDGWFKTGDTVVF KDGQYWIRGRTSVDIIKTGGYKVSALEVEWHLLAHPSITDVA VIGVPDMTWGQRVTAVVTLREGHSLSHRELKEWARNVLAPY AVPSELVLVEEIPRNQMGKIDKKALIRHFHPS	ACSF3
153	MTSCHIAEEHIQKVAIFGGTHGNELTGVFLVKHWLENGAEIQ RTGLEVKPFITNPRAVKKCTRYIDCDLNRIFDLENLGKKMSED LPYEVRRAQEINHLFGPKDSEDSYDIIFDLHNTTSNMGCTLILE DSRNNFLIQMFHYIKTSLAPLPCYVYLIEHPSLKYATTRSIAKY PVGIEVGPQPQGVLRADILDQMRKMIKHALDFIHHFNEGKEFP PCAIEVYKIIEKVDYPRDENGEIA	ASPA

[0853]

	AIIHPNLQDQDWKPLHPGDPMFLTLDGKTIPLGGDCTVYPVF VNEAAYYEKKEAFAKTTKLTLNAKSIRCCLH	
154	MAAAVAAAPGALGSLHAGGARLVAACSAWLCPGLRLPGSL AGRRAGPAIWAQGWVPAAGGPAPKRGYSSEMKTEDELRVR HLEEENRGIVVLGINRAYGKNSLSKNLIKMLSKAVDALKSDK KVRTIIIRSEVPGIFCAGADLKERAKMSSSEVGPFVSKIRAVIN DIANLPVPTIAAIDGLALGGGLELALACDIRVAASSAKMGLVE TKLAIIPGGGGTQRLPRAIGMSLAKELIFSARVLDGKEAKAVG LISHVLEQNQEGDAAYRKALDLAREFLPQGPVAMRVAKLAIN QGMEVDLVTGLAIEEACYAQTIPTKDRLEGLLAFKEKRPPRY KGE	AUH
155	MASTVVAVGLTIAAAGFAGRYVLQAMKHMEPQVKQVFQSL PKSAFSGGYYRGGFEPKMTKREAALILGVSPTANKGKIRDAH RRIMLLNHPDKGGSPYIAAKINEAKDLLEGQAKK	DNAJC19
156	MAEAVLRVARRQLSQRGGSGAPILLRQMFEPVSCTFTYLLGD RESREAVLIDPVLETAPRDAQLIKELGLRLLYAVNTHCHADHI TGSGLLRSLLPGCQSVISRLSGAQADLHIEDGDSIRFGRFALET RASPGHTPGCVTFVLNDHSMAFTGDALLIRGCGRTDFQQGCA KTLYHSVHEKIFTLPGDCLIYPAHDYHGFTVSTVEEERTLNPR LTLSCEEFVKIMGNLNLPKPQQIDFAVPANMRCGVQTPTA	ETHE1
157	MADQAPFDTDVNTLTRFVMEEGRKARGTGELTQLLNSLCTA VKAISSAVRKAGIAHLYGIAGSTNVTGDQVKKLDVLSNDLV MNMLKSSFATCVLVSEEDKHAIIVEPEKRGKYVVCFDPLDGS SNIDCLVSVGTIFGIYRKKSTDEPSEKDALQPGRNLVAAGYAL YGSATMLVLAMDCGVNCFMLDPAIGEFILVDKDVKIKKKGKI YSLNEGYARDFDPAVTEYIQRKKFPPDNSAPYGARYVGSMV ADVHRTLVYGGIFLYPANKKSPNGKLRLLYECNPMAYVMEK AGGMATTGKEAVLDVIPTDIHQRAPVILGSPDDVLEFLKVYE KHSAQ	FBP1
158	MSQLVECVPNFSEGKNQEVIDAISGAITQTPGCVLLDVDAGPS TNRTVYTFVGPPECVVEGALNAARVASRLIDMSRHQGEHPR MGALDVCPFIPVRGVSVDECVLCAQAFGQRLAEELDVPVYL YGEAARMDSRRTLPAIRAGEYEALPKKLQQADWAPDFGPSSF VPSWGATATGARKFLIAFNINLLGTKEQAHRIALNLREQGRG KDQPGRLKKVQGIGWYLDEKNLAQVSTNLLDFEVTALHTVY EETCREAQELSLPVVGSQLVGLVPLKALLDAAAFYCEKENLFI LEEEQRI RLVVSRLGLDSLCPFSPKERIIEYLVPERGPERGLGSKSLRAFV GEVGARSAAPGGGSVAAAAAAMGAALGSMVGLMTYGRRQ FQSLDTTMRRLIPPFREASAKLTTLVDADAEAFTAYLEAMRLP KNTPEEKDRRTAALQEGLRRAVSVPLTLAETVASLWPALQEL ARCGNLACRSDLQVAAKALEMGVFGAYFNVLINLRDITDEAF KDQIHHRVSSLLQEAKTQAALVLDCLETRQE	FTCD
159	MATNWGSLLQDKQQLEELARQAVDRALAEGVLLRTSQEPTS SEVVSYAPFTLFPSLVPSALLEQAYAVQMDFNLLVDAVSQNA AFLEQTLSSTIKQDDFTARLFDIHKQVLKEGIAQTVFLGLNRS DYMFQRSADGSPALKQIEINTISASFGGLASRTPAVHRHVLSV LSKTKEAGKILSNNPSKGLALGIAKAWELYGSPNALVLLIAQE KERNIFDQRAIENELLARNIHVIRRTFEDISEKGSLDQDRRLFV DGQEIAVVYFRDGYMPRQYSLQNWEARLLLERSHAAKCPDI ATQLAGTKKVQQELSRPGMLEMLLPGQPEAVARLRATFAGL YSLDVGEEGDQAIAEALAAPSRFVLKPQREGGGNNLYGEEM VQALKQLKDSEERASYILMEKIEPEPFENCLLRPGSPARVVQC	GSS

[0854]

	ISELGIFGVYVRQEKTLVMNKHVGHLLRTKAIEHADGGVAAG VAVLDNPYPV	
160	MGQREMWRLMSRFNAFKRTNTILHHLRMSKHTDAAEEVLLE KKGCTGVITLNRPKFLNALTLNMIRQIYPQLKKWEQDPETFLII IKGAGGKAFCAGGDIRVISEAEKAKQKIAPVFFREEYMLNNA VGSCQKPYVALIHGITMGGGVGLSVHGQFRVATEKCLFAMP ETAIGLFPDVGGGYFLPRLQGKLGYFLALTGFRLKGRDVYRA GIATHFVDSEKLAMLEEDLLALKSPSKENIASVLENYHTESKI DRDKSFILEEHMDKINSCFSANTVEEIIENLQQDGSSFALEQLK VINKMSPTSLKITLRQLMEGSSKTLQEVLTMEYRLSQACMRG HDFHEGVRAVLIDKDQSPKWKPADLKEVTEEDLNNHFKSLG SSDLKF	HIBCH
161	MAGYLRVVRSLCRASGSRPAWAPAALTAPTSQEQPRRHYAD KRIKVAKPVVEMDGDEMTRIIWQFIKEKLILPHVDIQLKYFDL GLPNRDQTDDQVTIDSALATQKYSVAVKCATITPDEARVEEF KLKKMWKSPNGTIRNILGGTVFREPIICKNIPRLVPGWTKPITI GRHAHGDQYKATDFVADRAGTFKMVFTPKDGSGVKEWEVY NFPAGGVGMGMYNTDESISGFAHSCFQYAIQKKWPLYMSTK NTILKAYDGRFKDIFQEIFDKHYKTDFDKNKIWYEHRLIDDM VAQVLKSSGGFVWACKNYDGDVQSDILAQGFGSLGLMTSVL VCPDGKTIEAEAAHGTVTRHYREHQKGRPTSTNPIASIFAWTR GLEHRGKLDGNQDLIRFAQMLEKVCVETVESGAMTKDLAGC IHGLSNVKLNEHFLNTTDFLDTIKSNLDRALGRQ	IDH2
162	MVPALRYLVGACGRARGLFAGGSPGACGFASGRPRPLCGGS RSASTSSFDIVIVGGGIVGLASARALILRHPSLSIGVLEKEKDLA VHQTGHNSGVIHSGIYYKPESLKAKLCVQGAALLYEYCQQK GISYKQCGKLIVAVEQEEIPRLQALYEKGLQNGVPGLRLIQQE DIKKKEPYCRGLMAIDCPHTGIVDYRQVALSFAQDFQEAGGS VLTNFEVKGIEMAKESPSRSIDGMQYPIVIKNTKGEEIRCQYV VTCAGLYSDRISELSGCTPDPRIVPFRGDYLLLKPEKCYLVKG NIYPVPDSRFPFLGVHFTPRMDGSIWLGPNAVLAFKREGYRPF DFSATDVMDIIINSGLIKLASQNFSYGVTEMYKACFLGATVKY LQKFIPEITISDILRGPAGVRAQALDRDGNLVEDFVFDAGVGDI GNRILHVRNAPSPAATSSIAISGMIADEVQQRFEL	L2HGDH
163	MRGFGPGLTARRLLPLRLPPRPPGPRLASGQAAGALERAMDE LLRRAVPPTPAYELREKTPAPAEGQCADFVSFYGGLAETAQR AELLGRLARGFGVDHGQVAEQSAGVLHLRQQQREAAVLLQ AEDRLRYALVPRYRGLFHHISKLDGGVRFLVQLRADLLEAQA LKLVEGPDVREMNGVLKGMLSEWFSSGFLNLERVTWHSPCE VLQKISEAEAVHPVKNWMDMKRRVGPYRRCYFFSHCSTPGE PLVVLHVALTGDISSNIQAIVKEHPPSETEEKNKITAAIFYSISL TQQGLQG VELGTFLIKRVVKELQREFPHLGVFSSLSPIPGFTKWLLGLLNS QTKEHGRNELFTDSECKEISEITGGPINETLKLLLSSSEWVQSE KLVRALQTPLMRLCAWYLYGEKHRGYALNPVANFHLQNGA VLWRINWMADVSLRGITGSCGLMANYRYFLEETGPNSTSYL GSKIIKASEQVLSLVAQFQKNSKL	MLYCD
164	MVVGAFPMAKLLYLGIRQVSKPLANRIKEAARRSEFFKTYICL PPAQLYHWVEMRTKMRIMGFRGTVIKPLNEEAAAELGAELL GEATIFIVGGGCLVLEYWRHQAQQRHKEEEQRAAWNALRDE VGHLALALEALQAQVQAAPPQGALEELRTELQEVRAQLCNP GRSASHAVPASKK	OPA3
165	MGSPEGRFHFAIDRGGTFTDVFAQCPGGHVRVLKLLSEDPAN	OPLAH

[0855]

	YADAPTEGIRRILEQEAGMLLPRDQPLDSSHIASIRMGTTVAT NALLERKGERVALLVTRGFRDLLHIGTQARGDLFDLAVPMPE VLYEEVLEVDERVVLHRGEAGTGTPVKGRTGDLLEVQQPVD LGALRGKLEGLLSRGIRSLAVVLMHSYTWAQHEQQVGVLAR ELGFTHVSLSSEAMPMVRIVPRGHTACADAYLTPAIQRYVQG FCRGFQGQLKDVQVLFMRSDGGLAPMDTFSGSSAVLSGPAG GVVGYSATTYQQEGGQPVIGFDMGGTSTDVSRYAGEFEHVF EASTAGVTLQAPQLDINTVAAGGGSRLFFRSGLFVVGPESAG AHPGPACYRKGGPVTVTDANLVLGRLLPASFPCIFGPGENQPL SPEASRKALEAVATEVNSFLTNGPCPASPLSLEEVAMGFVRV ANEAMCRPIRALTQARGHDPSAHVLACFGGAGGQHACAIAR ALGMDTVHIHRHSGLLSALGLALADVVHEAQEPCSLLYAPET FVQLDQRLSRLEEQCVDALQAQGFPRSQISTESFLHLRYQGTD CALMVSAHQHPATA RSPRAGDFGAAFVERYMREFGFVIPERPVVVDDVRVRGTGRS GLRLEDAPKAQTGPPRVDKMTQCYFEGGYQETPVYLLAELG YGHKLHGPCLIIDSNSTILVEPGCQAEVTKTGDICISVGAEVPG TVGPQLDPIQLSIFSHRFMSIAEQMGRILQRTAISTNIKERLDFS CALFGPDGGLVSNAPHIPVHLGAMQETVQFQIQHLGADLHPG DVLLSNHPSAGGSHLPDLTVITPVFWPGQTRPVFYVASRGHH ADIGGITPGSMPPHSTMLQQEGAVFLSFKLVQGGVFQEEAVT EALRAPGKVPNCSGTRNLHDNLSDLRAQVAANQKGIQLVGE LIGQYGLDVVQAYMGHIQANAELAVRDMLRAFGTSRQARGL PLEVSSEDHMDDGSPIRLRVQISLSQGSAVFDFSGTGPEVFGN LNAPRAVTLSALIYCLRCLVGRDIPLNQGCLAPVRVVIPRGSIL DPSPEAAVVGGNVLTSQRVVDVILGAFGACAASQGCMNNVT LGNAHMGYYETVAGGAGAGPSWHGRSGVHSHMTNTRITDP EILESRYPVILRRFELRRGSGGRGRFRGGDGVTRELLFREEALL SVLTERRAFRPYGLHGGEPGARGLNLLIRKNGRTVNLGGKTS VTVYPGDVFCLHTPGGGGYGDPEDPAPPPGSPPQALAFPEHG SVYEYRRAQEAV	
166	MAALKLLSSGLRLCASARGSGATWYKGCVCSFSTSAHRHTK FYTDPVEAVKDIPDGATVLVGGFGLCGIPENLIDALLKTGVKG LTAVSNNAGVDNFGLGLLLRSKQIKRMVSSYVGENAEFERQ YLSGELEVELTPQGTLAERIRAGGAGVPAFYTPTGYGTLVQE GGSPIKYNKDGSVAIASKPREVREFNGQHFILEEAITGDFALV KAWKADRAGNVIFRKSARNFNLPMCKAAETTVVEVEEIVDIG AFAPEDIHIPQIYVHRLIKGEKYEKRIERLSIRKEGDGEAKSAK PGDDVRERIIKRAALEFEDGMYANLGIGIPLLASNFISPNITVH LQSENGVLGLGPYPRQHEADADLINAGKETVTILPGASFFSSD ESFAMIRGGHVDLTMLGAMQVSKYGDLANWMIPGKMVKG MGGAMDLVSSAKTKVVVTMEHSAKGNAHKIMEKCTLPLTG KQCVNRIITEKAVFDVDKKKGLTLIELWEGLTVDDVQKSTGC DFAVSPKLMPMQQIAN	OXCT1
167	MSRLLWRKVAGATVGPGPVPAPGRWVSSSVPASDPSDGQRR RQQQQQQQQQQQQPQQPQVLSSEGGQLRHNPLDIQMLSRG LHEQIFGQGGEMPGEAAVRRSVEHLQKHGLWGQPAVPLPDV ELRLPPLYGDNLDQHFRLLAQKQSLPYLEAANLLLQAQLPPK PPAWAWAEGWTRYGPEGEAVPVAIPEERALVFDVEVCLAEG TCPTLAVAISPSAWYSWCSQRLVEERYSWTSQLSPADLIPLEV PTGASSPTQRDWQEQLVVGHNVSFDRAHIREQYLIQGSRMRF LDTMSMHMAISGLSSFQRSLWIAAKQGKHKVQPPTKQGQKS QRKARRGPAISSWDWLDISSVNSLAEVHRLYVGGPPLEKEPR ELFVKGTMKDIRENFQDLMQYCAQDVWATHEVFQQQLPLFL	POLG

[0856]

	ERCPHPVTLAGMLEMGVSYLPVNQNWERYLAEAQGTYEELQ REMKKSLMDLANDACQLLSGERYKEDPWLWDLEWDLQEFK QKKAKKVKKEPATASKLPIEGAGAPGDPMDQEDLGPCSEEEE FQQDVMARACLQKLKGTTELLPKRPQHLPGHPGWYRKLCPR LDDPAWTPGPSLLSLQMRVTPKLMALTWDGFPLHYSERHGW GYLVPGRRDNLAKLPTGTTLESAGVVCPYRAIESLYRKHCLE QGKQQLMPQEAGLAEEFLLTDNSAIWQTVEELDYLEVEAEA KMENLRAAVPGQPLALTARGGPKDTQPSYHHGNGPYNDVDI PGCWFFKLPHKDGNSCNVGSPFAKDFLPKMEDGTLQAGPGG ASGPRALEINKMISFWRNAHKRISSQMVVWLPRSALPRAVIR HPDYDEEGLYGAILPQVVTAGTITRRAVEPTWLTASNARPDR VGSELKAMVQAPPGYTLVGADVDSQELWIAAVLGDAHFAG MHGCTAFGWMTLQGRKSRGTDLHSKTATTVGISREHAKIFN YGRIYGAGQPFAERLLMQFNHRLTQQEAAEKAQQMYAATK GLRWYRLSDEGEWLVRELNLPVDRTEGGWISLQDLRKVQRE TARKSQWKKWEVVAERAWKGGTESEMFNKLESIATSDIPRT PVLGCCISRALEPSAVQEEFMTSRVNWVVQSSAVDYLHLML VAMKWLFEEFAIDGRFCISIHDEVRYLVREEDRYRAALALQIT NLLTRCMFAYKLGLNDLPQSVAFFSAVDIDRCLRKEVTMDC KTPSNPTGMERRYGIPQGEALDIYQIIELTKGSLEKRSQPGP	
168	MSTAALITLVRSGGNQVRRRVLLSSRLLQDDRRVTPTCHSSTS EPRCSRFDPDGSGSPATWDNFGIWDNRIDEPILLPPSIKYGKPIP KISLENVGCASQIGKRKENEDRFDFAQLTDEVLYFAVYDGHG GPAAADFCHTHMEKCIMDLLPKEKNLETLLTLAFLEIDKAFSS HARLSADATLLTSGTTATVALLRDGIELVVASVGDSRAILCRK GKPMKLTIDHTPERKDEKERIKKCGGFVAWNSLGQPHVNGR LAMTRSIGDLDLKTSGVIAEPETKRIKLHHADDSFLVLTTDGI NFMVNSQEICDFVNQCHDPNEAAHAVTEQAIQYGTEDNSTA VVVPFGAWGKYKNSEINFSFSRSFASSGRWA	PPM1K
169	MSLAAYCVICCRRIGTSTSPPKSGTHWRDIRNIIKFTGSLILGGS LFLTYEVLALKKAVTLDTQVVEREKMKSYIYVHTVSLDKGE NHGIAWQARKELHKAVRKVLATSAKILRNPFADPFSTVDIED HECAVWLLLRKSKSDDKTTRLEAVREMSETHHWHDYQYRII AQACDPKTLIGLARSEESDLRFFLLPPPLPSLKEDSSTEEELRQ LLASLPQTELDECIQYFTSLALSESSQ SLAAQKGGLWCFGGNGLPYAESFGEVPSATVEMFCLEAIVKH SEISTHCDKIEANGGLQLLQRLYRLHKDCPKVQRNIMRVIGN MALNEHLHSSIVRSGWVSIMAEAMKSPHIMESSHAARILANL DRETVQEKYQDGVYVLHPQYRTSQPIKADVLFIHGLMGAAF KTWRQQDSEQAVIEKPMEDEDRYTTCWPKTWLAKDCPALRI ISVEYDTSLSDWRARCPMERKSIAFRSNELLRKLRAAGVGDR PVVWISHSMGGLLVKKMLLEASTKPEMSTVINNTRGIIFYSVP HHGSRLAEYSVNIRYLLFPSLEVKELSKDSPALKTLQDDFLEF AKDKNFQVLNFVETLPTYIGSMIKLHVVPVESADLGIGDLIPV DVNHLNICKPKKKDAFLYQRTLQFIREALAKDLEN	SERAC1
170	MPAPRAPRALAAAAPASGKAKLTHPGKAILAGGLAGGIEICIT FPTEYVKTQLQLDERSHPPRYRGIGDCVRQTVRSHGVLGLYR GLSSLLYGSIPKAAVRFGMFEFLSNHMRDAQGRLDSTRGLLC GLGAGVAEAVVVVCPMETIKVKFIHDQTSPNPKYRGFFHGVR EIVREQGLKGTYQGLTATVLKQGSNQAIRFFVMTSLRNWYR GDNPNKPMNPLITGVFGAIAGAASVFGNTPLDVIKTRMQGLE AHKYRNTWDCGLQILKKEGLKAFYKGTVPRLGRVCLDVAIV FVIYDEV VKLLNKVWKTD	SLC25A1

[0857]

171	MAASMFYGRLVAVATLRNHRPRTAQRAAAQVLGSSGLFNN HGLQVQQQQQRNLSLHEYMSMELLQEAGVSVPKGYVAKSP DEAYAIAKKLGSKDVVIKAQVLAGGRGKGTFESGLKGGVKI VFSPEEAKAVSSQMIGKKLFTKQTGEKGRICNQVLVCERKYP RREYYFAITMERSFQGPVLIGSSHGGVNIEDVAAESPEAIIKEPI DIEEGIKKEQALQLAQKMGFPPNIVESAAENMVKLYSLFLKY DATMIEINPMVEDSDGAVLCMDAKINFDSNSAYRQKKIFDLQ DWTQEDERDKDAAKANLNYIGLDGNIGCLVNGAGLAMATM DIIKLHGGTPANFLDVGGGATVHQVTEAFKLITSDKKVLAILV NIFGGIMRCDVIAQGIVMAVKDLEIKIPVVVRLQGTRVDDAK ALIADSGLKILACDDLDEAARMVVKLSEIVTLAKQAHVDVKF QLPI	SUCLA2
172	MTATLAAAADIATMVSGSSGLAAARLLSRSFLLPQNGIRHCS YTASRQHLYVDKNTKIICQGFTGKQGTFHSQQALEYGTKLVG GTTPGKGGQTHLGLPVFNTVKEAKEQTGATASVIYVPPPFAA AAINEAIEAEIPLVVCITEGIPQQDMVRVKHKLLRQEKTRLIGP NCPGVINPGECKIGIMPGHIHKKGRIGIVSRSGTLTYEAVHQTT QVGLGQSLCVGIGGDPFNGTDFIDCLEIFLNDSATEGIILIGEIG GNAEENAAEFLKQHNSGPNSKPVVSFIAGLTAPPGRRMGHAG AIIAGGKGGAKEKISALQSAGVVVSMSPAQLGTTIYKEFEKRK ML	SUCLG1
173	MPLHVKWPFPAVPPLTWTLASSVVMGLVGTYSCFWTKYMN HLTVHNREVLYELIEKRGPATPLITVSNHQSCMDDPHLWGIL KLRHIWNLKLMRWTPAAADICFTKELHSHFFSLGKCVPVCRG AEFFQAENEGKGVLDTGRHMPGAGKRREKGDGVYQKGMDF ILEKLNHGDWVHIFPEGKVNMSSEFLRFKWGIGRLIAECHLNP IILPLWHVGMNDVLPNSPPYFPRFGQKITVLIGKPFSALPVLER LRAENKSAVEMRKALTDFIQEEFQHLKTQAEQLHNHLQPGR	TAZ
174	MTVFFKTLRNHWKKTTAGLCLLTWGGHWLYGKHCDNLLRR AACQEAQVFGNQLIPPNAQVKKATVFLNPAACKGKARTLFE KNAAPILHLSGMDVTIVKTDYEGQAKKLLELMENTDVIIVAG GDGTLQEVVTGVLRRTDEATFSKIPIGFIPLGETSSLSHTLFAES GNKVQHITDATLAIVKGETVPLDVLQIKGEKEQPVFAMTGLR WGSFRDAGVKVSKYWYLGPLKIKAAHFFSTLKEWPQTHQAS ISYTGPTERPPNEPEETPVQRPSLYRRILRRLASYWAQPQDALS QEVSPEVWKDVQLSTIELSITTRNNQLDPTSKEDFLNICIEPDTI SKGDFITIGSRKVRNPKLHVEGTECLQASQCTLLIPEGAGGSFS IDSEEYEAMPVEVKLLPRKLQFFCDPRKREQMLTSPTQ	AGK
175	MLGSLVLRRKALAPRLLLRLLRSPTLRGHGGASGRNVTTGSL GEPQWLRVATGGRPGTSPALFSGRGAATGGRQGGRFDTKCL AAATWGRLPGPEETLPGQDSWNGVPSRAGLGMCALAAALV VHCYSKSPSNKDAALLEAARANNMQEVSRLLSEGADVNAKH RLGWTALMVAAINRNNSVVQVLLAAGADPNLGDDFSSVYKT AKEQGIHSLEDGGQDGASRHITNQWTSALEFRRWLGLPAGVL ITREDDFNNRLNNRASFKGCTALHYAVLADDYRTVKELLDG GANPLQRNEMGHTPLDYAREGEVMKLLRTSEAKYQEKQRK REAEERRRFPLEQRLKEHIIGQESAIATVGAA IRRKENGWYDEEHPLVFLFLGSSGIGKTELAKQTAKYMHKDA KKGFIRLDMSEFQERHEVAKFIGSPPGYVGHEEGGQLTKKLK QCPNAVVLFDEVDKAHPDVLTIMLQLFDEGRLTDGKGKTIDC KDAIFIMTSNVASDEIAQHALQLRQEALEMSRNRIAENLGDV QISDKITISKNFKENVIRPILKAHFRRDEFLGRINEIVYFLPFCHS ELIQLVNKELNFWAKRAKQRHNITLLWDREVADVLVDGYNV HYGARSIKHEVERRVVNQLAAAYEQDLLPGGCTLRITVEDSD	CLPB

[0858]

	KQLLKSPELPSPQAEKRLPKLRLEIIDKDSKTRRLDIRAPLHPE KVCNTI	
176	MLFLALGSPWAVELPLCGRRTALCAAAALRGPRASVSRASSS SGPSGPVAGWSTGPSGAARLLRRPGRAQIPVYWEGYVRFLNT PSDKSEDGRLIYTGNMARAVFGVKCFSYSTSLIGLTFLPYIFTQ NNAISESVPLPIQIIFYGIMGSFTVITPVLLHFITKGYVIRLYHEA TTDTYKAITYNAMLAETSTVFHQNDVKIPDAKHVFTTFYAKT KSLLVNPVLFPNREDYIHLMGYDKEEFILYMEETSEEKRHKD DK	TMEM70
177	MLSQVYRCGFQPFNQHLLPWVKCTTVFRSHCIQPSVIRHVRS WSNIPFITVPLSRTHGKSFAHRSELKHAKRIVVKLGSAVVTRG DECGLALGRLASIVEQVSVLQNQGREMMLVTSGAVAFGKQR LRHEILLSQSVRQALHSGQNQLKEMAIPVLEARACAAAGQSG LMALYEAMFTQYSICAAQILVTNLDFHDEQKRRNLNGTLHEL LRMNIVPIVNTNDAVVPPAEPNSDLQGVNVISVKDNDSLAAR LAVEMKTDLLIVLSDVEGLFDSPPGSDDAKLIDIFYPGDQQSV TFGTKSRVGMGGMEAKVKAALWALQGGTSVVIANGTHPKV SGHVITDIVEGKKVGTFFSEVKPAGPTVEQQGEMARSGGRML ATLEPEQRAEIIHHLADLLTDQRDEILLANKKDLEEAEGRLAA PLLKRLSLSTSKLNSLAIGLRQIAASSQDSVGRVLRRTRIAKNL ELEQVTVPIGVLLVIFESRPDCLPQVAALAIASGNGLLLKGGK EAAHSNRILHLLTQEALSIHGVKEAVQLVNTREEVEDLCRLD KMIDLIIPRGSSQLVRDIQKAAKGIPVMGHSEGICHMYVDSEA SVDKVTRLVRDSKCEYPAACNALETLLIHRDLLRTPLFDQIID MLRVEQVKIHAGPKFASYLTFSPSEVKSLRTEYGDLELCIEVV DNVQDAIDHIIHKYGSSHTDVIVTEDENTAEFFLQHVDSACVF WNASTRFSDGYRFGLGAEVGISTSRIHARGPVGLEGLLTTKW LLRGKDHVVSDFSEHGSLKYLHENLPIPQRNTN	ALDH18A1
178	MFSKLAHLQRFAVLSRGVHSSVASATSVATKKTVQGPPTSDD IFEREYKYGAHNYHPLPVALERGKGIYLWDVEGRKYFDFLSS YSAVNQGHCHPKIVNALKSQVDKLTLTSRAFYNNVLGEYEE YITKLFNYHKVLPMNTGVEAGETACKLARKWGYTVKGIQKY KAKIVFAAGNFWGRTLSAISSSTDPTSYDGFGPFMPGFDIIPYN DLPALERALQDPNVAAFMVEPIQGEAGVVVPDPGYLMGVRE LCTRHQVLFIADEIQTGLARTGRWLAVDYENVRPDIVLLGKA LSGGLYPVSAVLCDDDIMLTIKPGEHGSTYGGNPLGCRVAIA ALEVLEEENLAENADKLGIILRNELMKLPSDVVTAVRGKGLL NAIVIKETKDWDAWKVCLRLRDNGLLAKPTHGDIIRFAPPLVI KEDELRESIEIINKTILSF	OAT
179	MLGRNTWKTSAFSFLVEQMWAPLWSRSMRPGRWCSQRSCA WQTSNNTLHPLWTVPVSVPGGTRQSPINIQWRDSVYDPQLKP LRVSYEAASCLYIWNTGYLFQVEFDDATEASGISGGPLENHY RLKQFHFHWGAVNEGGSEHTVDGHAYPAELHLVHWNSVKY QNYKEAVVGENGLAVIGVFLKLGAHHQTLQRLVDILPEIKHK DARAAMRPFDPSTLLPTCWDYWTYAGSLTTPPLTESVTWIIQ KEPVEVAPSQLSAFRTLLFSALGEEEKMMVNNYRPLQPLMNR KVWASFQATNEGTRS	CA5A
180	MYRYLGEALLLSRAGPAALGSASADSAALLGWARGQPAAAP QPGLALAARRHYSEAVADREDDPNFFKMVEGFFDRGASIVED KLVEDLRTRESEEQKRNRVRGILRIIKPCNHVLSLSFPIRRDDG SWEVIEGYRAQHSQHRTPCKGGIRYSTDVSVDEVKALASLMT YKCAVVDVPFGGAKAGVKINPKNYTDNELEKITRRFTMELA KKGFIGPGIDVPAPDMSTGEREMSWIADTYASTIGHYDINAHA	GLUD1

[0859]

	CVTGKPISQGGIHGRISATGRGVFHGIENFINEASYMSILGMTP GFG DKTFVVQGFGNVGLHSMRYLHRFGAKCIAVGESDGSIWNPD GIDPKELEDFKLQHGSILGFPKAKPYEGSILEADCDILIPAASEK QLTKSNAPRVKAKIIAEGANGPTTPEADKIFLERNIMVIPDLYL NAGGVTVSYFEWLKNLNHVSYGRLTFKYERDSNYHLLMSVQ ESLERKFGKHGGTIPIVPTAEFQDRISGASEKDIVHSGLAYTME RSARQIMRTAMKYNLGLDLRTAAYVNAIEKVFKVYNEAGVT FT	
181	MTTSASSHLNKGIKQVYMSLPQGEKVQAMYIWIDGTGEGLR CKTRTLDSEPKCVEELPEWNFDGSSTLQSEGSNSDMYLVPAA MFRDPFRKDPNKLVLCEVFKYNRRPAETNLRHTCKRIMDMV SNQHPWFGMEQEYTLMGTDGHPFGWPSNGFPGPQGPYYCG VGADRAYGRDIVEAHYRACLYAGVKIAGTNAEVMPAQWEF QIGPCEGISMGDHLWVARFILHRVCEDFGVIATFDPKPIPGNW NGAGCHTNFSTKAMREENGLKYIEEAIEKLSKRHQYHIRAYD PKGGLDNARRLTGFHETSNINDFSAGVANRSASIRIPRTVGQE KKGYFEDRRPSANCDPFSVTEALIRTCLLNETGDEPFQYKN	GLUL
182	MAVARAALGPLVTGLYDVQAFKFGDFVLKSGLSSPIYIDLRGI VSRPRLLSQVADILFQTAQNAGISFDTVCGVPYTALPLATVICS TNQIPMLIRRKETKDYGTKRLVEGTINPGETCLIIEDVVTSGSS VLETVEVLQKEGLKVTDAIVLLDREQGGKDKLQAHGIRLHSV CTLSKMLEILEQQKKVDAETVGRVKRFIQENVFVAANHNGSP LSIKEAPKELSFGARAELPRIHPVA SKLLRLMQKKETNLCLSADVSLARELLQLADALGPSICMLKT HVDILNDFTLDVMKELITLAKCHEFLIFEDRKFADIGNTVKKQ YEGGIFKIASWADLVNAHVVPGSGVVKGLQEVGLPLHRGCL LIAEMSSTGSLATGDYTRAAVRMAEEHSEFVVGFISGSRVSM KPEFLHLTPGVQLEAGGDNLGQQYNSPQEVIGKRGSDIIIVGR GIISAADRLEAAEMYRKAAWEAYLSRLGV	UMPS
183	MRDYDEVTAFLGEWGPFQRLIFFLLSASIIPNGFTGLSSVFLIA TPEHRCRVPDAANLSSAWRNHTVPLRLRDGREVPHSCRRYRL ATIANFSALGLEPGRDVDLGQLEQESCLDGWEFSQDVYLSTIV TEWNLVCEDDWKAPLTISLFFVGVLLGSFISGQLSDRFGRKN VLFVTMGMQTGFSFLQIFSKNFEMFVVLFVLVGMGQISNYVA AFVLGTEILGKSVRIIFSTLGVCIFYAFGYMVLPLFAYFIRDWR MLLVALTMPGVLCVALWWFIPESPRWLISQGRFEEAEVIIRKA AKANGIVVPSTIFDPSELQDLSSKKQQSHNILDLLRTWNIRMV TIMSIMLWMTISVGYFGLSLDTPNLHGDIFVNCFLSAMVEVPA YVLAWLLLQYLPRRYSMATALFLGGSVLLFMQLVPPDLYYL ATVLVMVGKFGVTAAFSMVYVYTAELYPTVVRNMGVGVSS TASRLGSILSPYFVYLGAYDRFLPYILMGSLTILTAILTLFLPES FGTPLPDTIDQMLRVKGMKHRKTPSHTR MLKDGQERPTILKSTAF	SLC22A5
184	MAEAHQAVAFQFTVTPDGIDLRLSHEALRQIYLSGLHSWKKK FIRFKNGIITGVYPASPSSWLIVVVGVMTTMYAKIDPSLGIIAKI NRTLETANCMSSQTKNVVSGVLFGTGLWVALIVTMRYSLKV LLSYHGWMFTEHGKMSRATKIWMGMVKIFSGRKPMLYSFQT SLPRLPVPAVKDTVNRYLQSVRPLMKEEDFKRMTALAQDFA VGLGPRLQWYLKLKSWWATNYVSDWWEEYIYLRGRGPLM VNSNYYAMDLLYILPTHIQAARAGNAIHAILLYRRKLDREEIK PIRLLGSTIPLCSAQWERMFNTSRIPGEETDTIQHMRDSKHIVV YHRGRYFKVWLYHDGRLLKPREMEQQMQRILDNTSEPQPGE ARLAALTAGDRVPWARCRQAYFGRGKNKQSLDAVEKAAFF	CPT1A

[0860]

	VTLDETEEGYRSEDPDTSMDSYAKSLLHGRCYDRWFDKSFTF VVFKNGKMGLNAEHSWADAPIVAHLWEYVMSIDSLQLGYA EDGHCKGDINPNIPYPTRLQWDIPGECQEVIETSLNTANLLAN DVDFHSFPFVAFGKGIIKKCRTSPDAFVQLALQLAHYKDMGK FCLTYEASMTRLFREGRTETVRSCTTESCDFVRAMVDPAQTV EQRLKLFKLASEKHQHMYRLAMTGSGIDRHLFCLYVVSKYL AVESPFLKEVLSEPWRLSTSQTPQQQVELFDLENNPEYVSSGG GFGPVADDGYGVSYILVGENLINFHISSKFSCPETDSHRFGRH LKEAMTDIITLFGLSSNSKK	
185	MVACRAIGILSRFSAFRILRSRGYICRNFTGSSALLTRTHINYG VKGDVAVVRINSPNSKVNTLSKELHSEFSEVMNEIWASDQIRS AVLISSKPGCFIAGADINMLAACKTLQEVTQLSQEAQRIVEKL EKSTKPIVAAINGSCLGGGLEVAISCQYRIATKDRKTVLGTPE VLLGALPGAGGTQRLPKMVGVPAALDMMLTGRSIRADRAK KMGLVDQLVEPLGPGLKPPEERTIEYLEEVAITFAKGLADKKI SPKRDKGLVEKLTAYAMTIPFVRQQVYKKVEEKVRKQTKGL YPAPLKIIDVVKTGIEQGSDAGYLCESQKFGELVMTKESKAL MGLYHGQVLCKKNKFGAPQKDVKHLAILGAGLMGAGIAQV SVDKGLKTILKDATLTALDRGQQQVFKGLNDKVKKKALTSF ERDSIFSNLTGQLDYQGFEKADMVIEAVFEDLSLKHRVLKEV EAVIPDHCIFASNTSALPISEIAAVSKRPEKVIGMHYFSPVDKM QLLEIITTEKTSKDTSASAVAVGLKQGKVIIVVK DGPGFYTTRCLAPMMSEVIRILQEGVDPKKLDSLTTSFGFPVG AATLVDEVGVDVAKHVAEDLGKVFGERFGGGNPELLTQMV SKGFLGRKSGKGFYIYQEGVKRKDLNSDMDSILASLKLPPKSE VSSDEDIQFRLVTRFVNEAVMCLQEGILATPAEGDIGAVFGLG FPPCLGGPFRFVDLYGAQKIVDRLKKYEAAYGKQFTPCQLLA DHANSPNKKFYQ	HADHA
186	MAFVTRQFMRSVSSSSTASASAKKIIVKHVTVIGGGLMGAGI AQVAAATGHTVVLVDQTEDILAKSKKGIEESLRKVAKKKFAE NLKAGDEFVEKTLSTIATSTDAASVVHSTDLVVEAIVENLKV KNELFKRLDKFAAEHTIFASNTSSLQITSIANATTRQDRFAGLH FFNPVPVMKLVEVIKTPMTSQKTFESLVDFSKALGKHPVSCK DTPGFIVNRLLVPYLMEAIRLYERGDASKEDIDTAMKLGAGY PMGPFELLDYVGLDTTKFIVDGWHEMDAENPLHQPSPSLNKL VAENKFGKKTGEGFYKYK	HADH
187	MAAPTLGRLVLTHLLVALFGMGSWAAVNGIWVELPVVVKD LPEGWSLPSYLSVVVALGNLGLLVVTLWRQLAPGKGEQVPIQ VVQVLSVVGTALLAPLWHHVAPVAGQLHSVAFLTLALVLA MACCTSNVTFLPFLSHLPPPFLRSFFLGQGLSALLPCVLALVQ GVGRLECPPAPTNGTSGPPLDFPERFPASTFFWALTALLVTSA AAFRGLLLLLPSLPSVTTGGSGPELQLGSPGAEEEEKEEEEALP LQEPPSQAAGTIPGPDPEAHQLFSAHGAFLLGLMAFTSAVTNG VLPSVQSFSCLPYGRLAYHLAVVLGSAANPLACFLAMGVLCR SLAGLVGLSLLGMLFGAYLMALAILSPCPPLVGTTAGVVLVV LSWVLCLCVFSYVKVAASSLLHGGGRPALLAAGVAIQVGSLL GAGAMFPPTSIYHVFQSRKDCVDPCGP	SLC52A1
188	MAAPTPARPVLTHLLVALFGMGSWAAVNGIWVELPVVVKEL PEGWSLPSYVSVLVALGNLGLLVVTLWRRLAPGKDEQVPIRV VQVLGMVGTALLASLWHHVAPVAGQLHSVAFLALAFVLAL ACCASNVTFLPFLSHLPPRFLRSFFLGQGLSALLPCVLALVQG VGRLECPPAPINGTPGPPLDFLERFPASTFFWALTALLVASAA AFQGLLLLLPPPPSVPTGELGSGLQVGAPGAEEEVEESSPLQEP PSQAAGTTPGPDPKAYQLLSARSACLLGLLAATNALTNGVLP	SLC52A2

[0861]

	AVQSFSCLPYGRLAYHLAVVLGSAANPLACFLAMGVLCRSL AGLGGLSLLGVFCGGYLMALAVLSPCPPLVGTSAGVVLVVLS WVLCLGVFSYVKVAASSLLHGGGRPALLAAGVAIQVGSLLG AVAMFPPTSIYHVFHSRKDCADPCDS	
189	MAFLMHLLVCVFGMGSWVTINGLWVELPLLVMELPEGWYL PSYLTVVIQLANIGPLLVTLLHHFRPSCLSEVPIIFTLLGVGTVT CIIFAFLWNMTSWVLDGHHSIAFLVLTFFLALVDCTSSVTFLPF MSRLPTYYLTTFFVGEGLSGLLPALVALAQGSGLTTCVNVTEI SDSVPSPVPTRETDIAQGVPRALVSALPGMEAPLSHLESRYLP AHFSPLVFFLLLSIMMACCLVAFFV LQRQPRCWEASVEDLLNDQVTLHSIRPREENDLGPAGTVDSS QGQGYLEEKAAPCCPAHLAFIYTLVAFVNALTNGMLPSVQTY SCLSYGPVAYHLAATLSIVANPLASLVSMFLPNRSLLFLGVLS VLGTCFGGYNMAMAVMSPCPLLQGHWGGEVLIVASWVLFS GCLSYVKVMLGVVLRDLSRSALLWCGAAVQLGSLLGALLMF PLVNVLRLFSSADFCNLHCPA	SLC52A3
190	MTILTYPFKNLPTASKWALRFSIRPLSCSSQLRAAPAVQTKTK KTLAKPNIRNVVVVDGVRTPFLLSGTSYKDLMPHDLARAALT GLLHRTSVPKEVVDYIIFGTVIQEVKTSNVAREAALGAGFSDK TPAHTVTMACISANQAMTTGVGLIASGQCDVIVAGGVELMS DVPIRHSRKMRKLMLDLNKAKSMGQRLSLISKFRFNFLAPEL PAVSEFSTSETMGHSADRLAAAFAVSRLEQDEYALRSHSLAK KAQDEGLLSDVVPFKVPGKDTVTKDNGIRPSSLEQMAKLKPA FIKPY GTVTAANSSFLTDGASAMLIMAEEKALAMGYKPKAYLRDFM YVSQDPKDQLLLGPTYATPKVLEKAGLTMNDIDAFEFHEAFS GQILANFKAMDSDWFAENYMGRKTKVGLPPLEKFNNWGGS LSLGHPFGATGCRLVMAAANRLRKEGGQYGLVAACAAGGQ GHAMIVEAYPK	HADHB
191	MLRGRSLSVTSLGGLPQWEVEELPVEELLLFEVAWEVTNKV GGIYTVIQTKAKTTADEWGENYFLIGPYFEHNMKTQVEQCEP VNDAVRRAVDAMNKHGCQVHFGRWLIEGSPYVVLFDIGYSA WNLDRWKGDLWEACSVGIPYHDREANDMLIFGSLTAWFLKE VTDHADGKYVVAQFHEWQAGIGLILSRARKLPIATIFTTHATL LGRYLCAANIDFYNHLDKFNIDKEAGERQIYHRYCMERASVH CAHVFTTVSEITAIEAEHMLKRKPDVVTPNGLNVKKFSAVHE FQNLHAMYKARIQDFVRGHFYGHLDFDLEKTLFLFIAGRYEF SNKGADIFLESLSRLNFLLRMHKSDITVMVFFIMPAKTNNFNV ETLKGQAVRKQLWDVAHSVKEKFGKKLYDALLRGEIPDLND ILDRDDLTIMKRAIFSTQRQSLPPVTTHNMIDDSTDPILSTIRRI GLFNNRTDRVKVILHPEFLSSTSPLLPMDYEEFVRGCHLGVFP SYYEPWGYTPAECTVMGIPSVTTNLSGFGCFMQEHVADPTAY GIYIVDRRFRSPDDSCNQLTKFLYGFCKQSRRQRIIQRNRTERL SDLLDWRYLGRYYQHARHLTLSRAFPDKFHVELTSPPTTEGF KYPRPSSVPPSPSGSQASSPQSSDVEDEVEDERYDEEEEAERD RLNIKSPFSLSHVPHGKKKLHGEYKN	GYS2
192	MAKPLTDQEKRRQISIRGIVGVENVAELKKSFNRHLHFTLVK DRNVATTRDYYFALAHTVRDHLVGRWIRTQQHYYDKCPKR VYYLSLEFYMGRTLQNTMINLGLQNACDEAIYQLGLDIEELE EIEEDAGLGNGGLGRLAACFLDSMATLGLAAYGYGIRYEYGI FNQKIRDGWQVEEADDWLRYGNPWEKSRPEFMLPVHFYGK VEHTNTGTKWIDTQVVLALPYDTPVPGYMNNTVNTMRLWS ARAPNDFNLRDFNVGDYIQAVLDRNLAENISRVLYPNDNFFE GKELRLKQEYFVVAATLQDIIRRFKASKFGSTRGAGTVFDAFP	PYGL

		DQVAIQLNDTHPALAIPELMRIFVDIEKL PWSKAWELTQKTFAYTNHTVLPEALERWPVDLVEKLLPRHL EIIYEINQKHLDRIVALFPKDVDRLRRMSLIEEEGSKRINMAHL CIVGSHAVNGVAKIHSDIVKTKVFKDFSELEPDKFQNKTNGIT PRRWLLLCNPGLAELIAEKIGEDYVKDLSQLTKLHSFLGDDVF LRELAKVKQENKLKFSQFLETEYKVKINPSSMFDVQVKRIHE YKRQLLNCLHVITMYNRIKKDPKKLFVPRTVIIGGKAAPGYH MAKMIIKLITSVADVVNNDPMVGSKLKVIFLENYRVSLAEKV IPATDLSEQISTAGTEASGTGNMKFMLNGALTIGTMDGANVE MAEEAGEENLFIFGMRIDDVAALDKKGYEAKEYYEALPELKL VIDQIDNGFFSPKQPDLFKDIINMLFYHDRFKVFADYEAYVKC QDKVSQLYMNPKAWNTMVLKNIAASGKFSSDRTIKEYAQNI WNVEPSDLKISLSNESNKVNGN	
	193	MTEDKVTGTLVFTVITAVLGSFQFGYDIGVINAPQQVIISHYR HVLGVPLDDRKAINNYVINSTDELPTISYSMNPKPTPWAEEET VAAAQLITMLWSLSVSSFAVGGMTASFFGGWLGDTLGRIKA MLVANILSLVGALLMGFSKLGPSHILIIAGRSISGLYCGLISGL VPMYIGEIAPTALRGALGTFHQLAIVTGILISQIIGLEFILGNYD LWHILLGLSGVRAILQSLLLFFCPESPRYLYIKLDEEVKAKQSL KRLRGYDDVTKDINEMRKEREEASSEQKVSIIQLFTNSSYRQP ILVALMLHVAQQFSGINGIFYYSTSIFQTAGISKPVYATIGVGA VNMVFTAVSVFLVEKAGRRSLFLIGMSGMFVCAIFMSVGLVL LNKFSWMSYVSMIAIFLFVSFFEIGPGPIPWFMVAEFFSQGPRP AALAIAAFSNWTCNFIVALCFQYIADFCGPYVFFLFAGVLLAF TLFTFFKVPETKGKSFEEIAAEFQKKSGSAHRPKAAVEMKFLG ATETV	SLC2A2
[0862]	194	MAASCLVLLALCLLLPLLLLGGWKRWRRGRAARHVVAVVL GDVGRSPRMQYHALSLAMHGFSVTLLGFCNSKPHDELLQNN RIQIVGLTELQSLAVGPRVFQYGVKVVLQAMYLLWKLMWRE PGAYIFLQNPPGLPSIAVCWFVGCLCGSKLVIDWHNYGYSIM GLVHGPNHPLVLLAKWYEKFFGRLSHLNLCVTNAMREDLAD NWHIRAVTVYDKPASFFKETPLDLQHRLFMKLGSMHSPFRAR SEPEDPVTERSAFTERDAGSGLVTRLRERPALLVSSTSWTEDE DFSILLAALEKFEQLTLDGHNLPSLVCVITGKGPLREYYSRLIH QKHFQHIQVCTPWLEAEDYPLLLGSADLGVCLHTSSSGLDLP MKVVDMFGCCLPVCAVNFKCLHELVKHEENGLVFEDSEELA AQLQMLFSNFPDPAGKLNQFRKNLRESQQLRWDESWVQTVL PLVMDT	ALG1
	195	MAEEQGRERDSVPKPSVLFLHPDLGVGGAERLVLDAALALQ ARGCSVKIWTAHYDPGHCFAESRELPVRCAGDWLPRGLGWG GRGAAVCAYVRMVFLALYVLFLADEEFDVVVCDQVSACIPV FRLARRRKKILFYCHFPDLLLTKRDSFLKRLYRAPIDWIEEYTT GMADCILVNSQFTAAVFKETFKSLSHIDPDVLYPSLNVTSFDS VVPEKLDDLVPKGKKFLLLSINRYERKKNLTLALEALVQLRG RLTSQDWERVHLIVAGGYDERVLENVEHYQELKKMVQQSDL GQYVTFLRSFSDKQKISLLHSCTCVLYTPSNEHFGIVPLEAMY MQCPVIAVNSGGPLESIDHSVTGFLCEPDPVHFSEAIEKFIREPS LKATMGLAGRARVKEKFSPEAFTEQLYRYVTKLLV	ALG2
	196	MAAGLRKRGRSGSAAQAEGLCKQWLQRAWQERRLLLREPR YTLLVAACLCLAEVGITFWVIHRVAYTEIDWKAYMAEVEGVI NGTYDYTQLQGDTGPLVYPAGFVYIFMGLYYATSRGTDIRM AQNIFAVLYLATLLLVFLIYHQTCKVPPFVFFFMCCASYRVHS IFVLRLFNDPVAMVLLFLSINLLLAQRWGWGCCFFSLAVSVK MNVLLFAPGLLFLLLTQFGFRGALPKLGICAGLQVVLGLPFLL	ALG3

[0863]

	ENPSGYLSRSFDLGRQFLFHWTVNWRFLPEALFLHRAFHLAL LTAHLTL LLLFALCRWHRTGESILSLLRDPSKRKVPPQPLTPNQIVSTLFT SNFIGICFSRSLHYQFYVWYFHTLPYLLWAMPARWLTHLLRL LVLGLIELSWNTYPSTSCSSAALHICHAVILLQLWLGPQPFPKS TQHSKKAH	
197	MEKWYLMTVVVLIGLTVRWTVSLNSYSGAGKPPMFGDYEA QRHWQEITFNLPVKQWYFNSSDNNLQYWGLDYPPLTAYHSL LCAYVAKFINPDWIALHTSRGYESQAHKLFMRTTVLIADLLIY IPAVVLYCCCLKEISTKKKIANALCILLYPGLILIDYGHFQYNS VSLGFALWGVLGISCDCDLLGSLAFCLAINYKQMELYHALPF FCFLLGKCFKKGLKGKGFVLLVKLACIVVASFVLCWLPFFTE REQTLQVLRRLFPVDRGLFEDKVANIWCSFNVFLKIKDILPRH IQLIMSFCSTFLSLLPACIKLILQPSSKGFKFTLVSCALSFFLFSF QVHEKSILLVSLPVCLVLSEIPFMSTWFLLVSTFSMLPLLLKDE LLMPSVVTTMAFFIACVTSFSIFEKTSEEELQLKSFSISVRKYLP CFTFLSRIIQYLFLISVITMVLLTLMTVTLDPPQKLPDLFSVLVC FVSCLNFLFFLVYFNIIIMWDSKSGRNQKKIS	ALG6
198	MAALTIATGTGNWFSALALGVTLLKCLLIPTYHSTDFEVHRN WLAITHSLPISQWYYEATSEWTLDYPPFFAWFEYILSHVAKYF DQEMLNVHNLNYSSSRTLLFQRFSVIFMDVLFVYAVRECCKC IDGKKVGKELTEKPKFILSVLLLWNFGLLIVDHIHFQYNGFLF GLMLLSIARLFQKRHMEGAFLFAVLLHFKHIYLYVAPAYGVY LLRSYCFTANKPDGSIRWKSFSFVRVISLGLVVFLVSALSLGPF LALNQLPQVFSRLFPFKRGLCHAYWAPNFWALYNALDKVLS VIGLKLKFLDPNNIPKASMTSGLVQQFQHTVLPSVTPLATLICT LIAILPSIFCLWFKPQGPRGFLRCLTLCALSSFMFGWHVHEKAI LLAILPMSLLSVGKAGDASIFLILTTTGHYSLFPLLFTAPELPIKI LLMLLFTIYSISSLKTLFRKEKPLFNWMETFYLLGLGPLEVCCE FVFPFTSWKVKYPFIPLLLTSVYCAVGITYAWFKLYVSVLIDS AIGKTKKQ	ALG8
199	MASRGARQRLKGSGASSGDTAPAADKLRELLGSREAGGAEH RTELSGNKAGQVWAPEGSTAFKCLLSARLCAALLSNISDCDE TFNYWEPTHYLIYGEGFQTWEYSPAYAIRSYAYLLLHAWPAA FHARILQTNKILVFYFLRCLLAFVSCICELYFYKAVCKKFGLH VSRMMLAFLVLSTGMFCSSSAFLPSSFCMYTTLIAMTGWYM DKTSIAVLGVAAGAILGWPFSAALGLPIAFDLLVMKHRWKSF FHWSLMALILFLVPVVVIDSYYYGKLVIAPLNIVLYNVFTPHG PDLYGT EPWYFYLINGFLNFNVAFALALLVLPLTSLMEYLLQRFHVQN LGHPYWLTLAPMYIWFIIFFIQPHKEERFLFPVYPLICLCGAVA LSALQKCYHFVFQRYRLEHYTVTSNWLALGTVFLFGLLSFSR SVALFRGYHGPLDLYPEFYRIATDPTIHTVPEGRPVNVCVGKE WYRFPSSFLLPDNWQLQFIPSEFRGQLPKPFAEGPLATRIVPTD MNDQNLEEPSRYIDISKCHYLVDLDTMRETPREPKYSSNKEE WISLAYRPFLDASRSSKLLRAFYVPFLSDQYTVYVNYTILKPR KAKQIRKKSGG	ALG9
200	MAAGERSWCLCKLLRFFYSLFFPGLIVCGTLCVCLVIVLWGIR LLLQRKKKLVSTSKNGKNQMVIAFFHPYCNAGGGGERVLWC ALRALQKKYPEAVYVVYTGDVNVNGQQILEGAFRRFNIRLIH PVQFVFLRKRYLVEDSLYPHFTLLGQSLGSIFLGWEALMQCV PDVYIDSMGYAFTLPLFKYIGGCQVGSYVHYPTISTDMLSVV KNQNIGFNNAAFITRNPFLSKVKLIYYYLFAFIYGLVGSCSDV VMVNSSWTLNHILSLWKVGNCTNIVYPPCDVQTFLDIPLHEK	ALG11

[0864]

	KMTPGHLLVSVGQFRPEKNHPLQIRAFAKLLNKKMVESPPSL KLVLIGGCRNKDDELRVNQLRRLSEDLGVQEYVEFKINIPFDE LKNYLSEATIGLHTMWNEHFGIGVVECMAAGTIILAHNSGGP KLDIVVPHEGDITGFLAESEEDYAETIAHILSMSAEKRLQIRKS ARASVSRFSDQEFEVTFLSSVEKLFK	
201	MAGKGSSGRRPLLLGLLVAVATVHLVICPYTKVEESFNLQAT HDLLYHWQDLEQYDHLEFPGVVPRTFLGPVVIAVFSSPAVYV LSLLEMSKFYSQLIVRGVLGLGVIFGLWTLQKEVRRHFGAMV ATMFCWVTAMQFHLMFYCTRTLPNVLALPVVLLALAAWLR HEWARFIWLSAFAIIVFRVELCLFLGLLLLLALGNRKVSVVRA LRHAVPAGILCLGLTVAVDSYFWRQLTWPEGKVLWYNTVLN KSSNWGTSPLLWYFYSALPRGLGCSLLFIPLGLVDRRTHAPTV LALGFMALYSLLPHKELRFIIYAFPMLNITAARGCSYLLNNYK KSWLYKAGSLLVIGHLVVNAAYSATALYVSHFNYPGGVAM QRLHQLVPPQTDVLLHIDVAAAQTGVSRFLQVNSAWRYDKR EDVQPGTGMLAYTHILMEAAPGLLALYRDTHRVLASVVGTT GVSLNLTQLPPFNVHLQTKLVLLERLPRPS	ALG12
202	MKCVFVTVGTTSFDDLIACVSAPDSLQKIESLGYNRLILQIGR GTVVPEPFSTESFTLDVYRYKDSLKEDIQKADLVISHAGAGSC LETLEKGKPLVVVINEKLMNNHQLELAKQLHKEGHLFYCTC RVLTCPGQAKSIASAPGKCQDSAALTSTAFSGLDFGLLSGYLH KQALVTATHPTCTLLFPSCHAFFPLPLTPTLYKMHKGWKNYC SQKSLNEASMDEYLGSLGLFRKLTAKDASCLFRAISEQLFCSQ VHHLEIRKACVSYMRENQQTFESYVEGSFEKYLERLGDPKES AGQLEIRALSLIYNRDFILYRFPGKPPTYVTDNGYEDKILLCYS SSGHYDSVYSKQFQSSAAVCQAVLYEILYKDVFVVDEEELKT AIKLFRSGSKKNRNNAVTGSEDAHTDYKSSNQNRMEEWGAC YNAENIPEGYNKGTEETKSPENPSKMPFPYKVLKALDPEIYRN VEFDVWLDSRKELQKSDYMEYAGRQYYLGDKCQVCLESEG RYYNAHIQEVGNENNSVTVFIEELAEKHVVPLANLKPVTQVM SVPAWNAMPSRKGRGYQKMPGGYVPEIVISEMDIKQQKKMF KKIRGKEVYM TMAYGKGDPLLPPRLQHSMHYGHDPPMHYSQTAGNVMSNE HFHPQHPSPRQGRGYGMPRNSSRFINRHNMPGPKVDFYPGPG KRCCQSYDNFSYRSRSFRRSHRQMSCVNKESQYGFTPGNGQ MPRGLEETITFYEVEEGDETAYPTLPNHGGPSTMVPATSGYC VGRRGHSSGKQTLNLEEGNGQSENGRYHEEYLYRAEPDYET SGVYSTTASTANLSLQDRKSCSMSPQDTVTSYNYPQKMMGNI AAVAASCANNVPAPVLSNGAAANQAISTTSVSSQNAIQPLFV SPPTHGRPVIASPSYPCHSAIPHAGASLPPPPPPPPPPPPPPPPP PPPPPPPPALDVGETSNLQPPPPLPPPPYSCDPSGSDLPQDTKVL QYYFNLGLQCYYHSYWHSMVYVPQMQQQLHVENYPVYTEP PLVDQTVPQCYSEVRREDGIQAEASANDTFPNADSSSVPHGA VYYPVMSDPYGQPPLPGFDSCLPVVPDYSCVPPWHPVGTAY GGSSQIHGAINPGPIGCIAPSPPASHYVPQGM	ALG13
203	MGSLFRSETMCLAQLFLQSGTAYECLSALGEKGLVQFRDLNQ NVSSFQRKFVGEVKRCEELERILVYLVQEINRADIPLPEGEASP PAPPLKQVLEMQEQLQKLEVELREVTKNKEKLRKNLLELIEY THMLRVTKTFVKRNVEFEPTYEEFPSLESDSLLDYSCMQRLG AKLGFVSGLINQGKVEAFEKMLWRVCKGYTIVSYAELDESLE DPETGEVIKWYVFLISFWGEQIGHKVKKICDCYHCHVYPYPN TAEERREIQEGLNTRIQDLYTVLHKTEDYLRQVLCKAAESVY SRVIQVKKMKAIYHMLNMCSFDVTNKCLIAEVWCPEADLQD LRRALEEGSRESGATIPSFMNIIPTKETPPTRIRTNKFTEGFQNI	ATP6V0A2

		VDAYGVGSYREVNPALFTIITFPFLFAVMFGDFGHGFVMFLFA LLLVLNENHPRLNQSQEIMRMFFNGRYILLLMGLFSVYTGLIY NDCFSKSVNLFGSGWNVSAMYSSSHPPAEHKKMVLWNDSV VRHNSILQLDPSIPGVFRGPYPLGIDPIWNLATNRLTFLNSFKM KMSVILGIIHMTFGVILGIFNHLHFRKKFNIYLVSIPELLFMLCI FGYLIFMIFYKWLVFSAETSRVAPSILIEFINMFLFPASKTSGLY TGQEYVQRVLLVVTALSVPVLFLGKPLFLLWLHNGRSCFGVN RSGYTLIRKDSEEEVSLLGSQDIEEGNHQVEDGCREMACEEFN FGEILMTQVIHSIEYCLGCISNTASYLRLWALSLAHAQLSDVL WAMLMRVGLRVDTTYGVLLLLPVIALFAVLTIFILLIMEGLSA FLHAIRLHWVEFQNKFYVGAGTKFVPF SFSLLSSKFNNDDSVA	
	204	MRPPACWWLLAPPALLALLTCSLAFGLASEDTKKEVKQSQD LEKSGISRKNDIDLKGIVFVIQSQSNSFHAKRAEQLKKSILKQA ADLTQELPSVLLLHQLAKQEGAWTILPLLPHFSVTYSRNSSWI FFCEEETRIQIPKLLETLRRYDPSKEWFLGKALHDEEATIIHHY AFSENPTVFKYPDFAAGWALSIPLVNKLTKRLKSESLKSDFTI DLKHEIALYIWDKGGGPPLTPVPEF CTNDVDFYCATTFHSFLPLCRKPVKKKDIFVAVKTCKKFHGD RIPIVKQTWESQASLIEYYSDYTENSIPTVDLGIPNTDRGHCGK TFAILERFLNRSQDKTAWLVIVDDDTLISISRLQHLLSCYDSGE PVFLGERYGYGLGTGGYSYITGGGGMVFSREAVRRLLASKCR CYSNDAPDDMVLGMCFSGLGIPVTHSPLFHQARPVDYPKDYL SHQVPISFHKHWNIDPVKVYFTWLAPSDEDKARQETQKGFRE EL	B3GLCT
[0865]	205	MFPRPLTPLAAPNGAEPLGRALRRAPLGRARAGLGGPPLLLPS MLMFAVIVASSGLLLMIERGILAEMKPLPLHPPGREGTAWRG KAPKPGGLSLRAGDADLQVRQDVRNRTLRAVCGQPGMPRDP WDLPVGQRRTLLRHILVSDRYRFLYCYVPKVACSNWKRVMK VLAGVLDSVDVRLKMDHRSDLVFLADLRPEEIRYRLQHYFKF LFVREPLERLLSAYRNKFGEIREYQQRYGAEIVRRYRAGAGPS PAGDDVTFPEFLRYLVDEDPERMNEHWMPVYHLCQPCAVHY DFVGSYERLEADANQVLEWVRAPPHVRFPARQAWYRPASPE SLHYHLCSAPRALLQDVLPKYILDFSLFAYPLPNVTKEACQQ	CHST14
	206	MATAATSPALKRLDLRDPAALFETHGAEEIRGLERQVRAEIE HKKEELRQMVGERYRDLIEAADTIGQMRRCAVGLVDAVKAT DQYCARLRQAGSAAPRPPRAQQPQQPSQEKFYSMAAQIKLLL EIPEKIWSSMEASQCLHATQLYLLCCHLHSLLQLDSSSSRYSP VLSRFPILIRQVAAASHFRSTILHESKMLLKCQGVSDQAVAEA LCSIMLLEESSPRQALTDFLLARKATIQKLLNQPHHGAGIKAQI CSLVELLATTLKQAHALFYTLPEGLLPDPALPCGLLFSTLETIT GQHPAGKGTGVLQEEMKLCSWFKHLPASIVEFQPTLRTLAHP ISQEYLKDTLQKWIHMCNEDIKNGITNLLMYVKSMKGLAGIR DAMWELLTNESTNHSWDVLCRRLLEKPLLFWEDMMQQLFL DRLQTLTKEGFDSISSSSKELLVSALQELESSTSNSPSNKHIHFE YNMSLFLWSESPNDLPSDAAWVSVANRGQFASSGLSMKAQA ISPCVQNFCSALDSKLKVKLDDLLAYLPSDD SSLPKDVSPTQAKSSAFDRYADAGTVQEMLRTQSVACIKHIV DCIRAELQSIEEGVQGQQDALNSAKLHSVLFMARLCQSLGEL CPHLKQCILGKSESSEKPAREFRALRKQGKVKTQEIIPTQAKW QEVKEVLLQQSVMGYQVWSSAVVKVLIHGFTQSLLLDDAGS VLATATSWDELEIQEEAESGSSVTSKIRLPAQPSWYVQSFLFS LCQEINRVGGHALPKVTLQEMLKSCMVQVVAAYEKLSEEKQ IKKEGAFPVTQNRALQLLYDLRYLNIVLTAKGDEVKSGRSKP	COG1

	DSRIEK VTDHLEALIDPFDLDVFTPHLNSNLHRLVQRTSVLFGLVTGTE NQLAPRSSTFNSQEPHNILPLASSQIRFGLLPLSMTSTRKAKST RNIETKAQVVPPARSTAGDPTVPGSLFRQLVSEEDNTSAPSLF KLGWLSSMTK	
207	MEKSRMNLPKGPDTLCFDKDEFMKEDFDVDHFVSDCRKRVQ LEELRDDLELYYKLLKTAMVELINKDYADFVNLSTNLVGMD KALNQLSVPLGQLREEVLSLRSSVSEGIRAVDERMSKQEDIRK KKMCVLRLIQVIRSVEKIEKILNSQSSKETSALEASSPLLTGQIL ERIATEFNQLQFHAVQSKGMPLLDKVRPRIAGITAMLQQSLE GLLLEGLQTSDVDIIRHCLRTYATIDKTRDAEALVGQVLVKPY IDEVIIEQFVESHPNGLQVMYNKLLEFVPHHCRLLREVTGGAI SSEKGNTVPGYDFLVNSVWPQIVQGLEEKLPSLFNPGNPDAF HEKYTISMDFVRRLERQCGSQASVKRLRAHPAYHSFNKKWN LPVYFQIRFREIAGSLEAALTDVLEDAPAESPYCLLASHRTWS SLRRCWSDEMFLPLLVHRLWRLTLQILARYSVFVNELSLRPIS NESPKEIKKPLVTGSKEPSITQGNTEDQGSGPSETKPVVSISRT QLVYVVADLDKLQEQLPELLEIIKPKLEMIGFKNFSSISAALED SQSSFSACVPSLSSKIIQDLSDSCFGFLKSALEVPRLYRRTNKE VPTTASSYVDSALKPLFQLQSGHKDKLKQAIIQQWLEGTLSES THKYYETVSDVLNSVKKMEESLKRLKQARKTTPANPVGPSG GMSDDDKIRLQLALDVEYLGEQIQKLGLQASDIKSFSALAEL VAAAKDQATAEQP	COG2
208	MADLDSPPKLSGVQQPSEGVGGGRCSEISAELIRSLTELQELE AVYERLCGEEKVVERELDALLEQQNTIESKMVTLHRMGPNL QLIEGDAKQLAGMITFTCNLAENVSSKVRQLDLAKNRLYQAI QRADDILDLKFCMDGVQTALRSEDYEQAAAHTHRYLCLDKS VIELSRQGKEGSMIDANLKLLQEAEQRLKAIVAEKFAIATKEG DLPQVERFFKIFPLLGLHEEGLRKFSEYLCKQVASKAEENLLM VLGTDMSDRRAAVIFADTLTLLFEGIARIVETHQPIVETYYGP GRLYTLIKYLQVECDRQVEKVVDKFIKQRDYHQQFRHVQNN LMRNSTTEKIEPRELDPILTEVTLMNARSELYLRFLKKRISSDF EVGDSMASEEVKQEHQKCLDKLLNNCLLSCTMQELIGLYVT MEEYFMRETVNKAVALDTYEKGQLTSSMVDDVFYIVKKCIG RALSSSSIDCLCAMINLATTELESDFRDVLCNKLRMGFPATTF QDIQRGVTSAVNIMHSSLQQGKFDTKGIESTDEAKMSFLVTL NNVEVCSENISTLKKTLESDCTKLFSQGIGGEQAQAKFDSCLS DLAAVSNKFRDLLQEGLTELNSTAIKPQVQPWINSFFSVSHNI EEEEFNDYEANDPWVQQFILNLEQQMAEFKASLSPVIYDSLT GLMTSLVAVELEKVVLKSTFNRLGGLQFDKELRSLIAYLTTV TTWTIRDKFARLSQMATILNLERVTEILDYWGPNSGPLTWRL TPAEVRQVLALRIDFRSEDIKRLRL	COG4
209	MGWVGGRRRDSASPPGRSRSAADDINPAPANMEGGGGSVAV AGLGARGSGAAAATVRELLQDGCYSDFLNEDFDVKTYTSQSI HQAVIAEQLAKLAQGISQLDRELHLQVVARHEDLLAQATGIE SLEGVLQMMQTRIGALQGAVDRIKAKIVEPYNKIVARTAQLA RLQVACDLLRRIIRILNLSKRLQGQLQGGSREITKAAQSLNEL DYLSQGIDLSGIEVIENDLLFIARARLEVENQAKRLLEQGLETQ NPTQVGTALQVFYNLGTLKDTITSVVDGYCATLEENINSALDI KVLTQPSQSAVRGGPGRSTMPTPGNTAALRASFWTNMEKLM DHIYAVCGQVQHLQKVLAKKRDPVSHICFIEEIVKDGQPEIFY TFWNSVTQALSSQFHMATNSSMFLKQAFEGEYPKLLRLYND LWKRLQQYSQHIQGNFNASGTTDLYVDLQHMEDDAQDIFIPK KPDYDPEKALKDSLQPYEAAYLSKSLSRLFDPINLVFPPGGRN	COG5

[0866]

		PPSSDELDGIIKTIASELNVAAVDTNLTLAVSKNVAKTIQLYSV KSEQLLSTQGDASQVIGPLTEGQRRNVAVVNSLYKLHQSVTK AIHALMENAVQPLLTSVGDAIEAIIITMHQEDFSGSLSSSGKPD VPCSLYMKELQGFIARVMSDYFKHFECLDFVFDNTEAIAQRA VELFIRHASLIRPLGEGGKMRLAADFAQMELAVGPFCRRVSD LGKSYRMLRSFRPLLFQASEHVASSPALGDVIPFSIIIQFLFTRA PAELKSPFQRAEWSHTRFSQWLDDHPSEKDRLLLIRGALEAY VQSVRSREGKEFAPVYPIMVQLLQKAMSALQ	
	210	MAEGSGEVVAVSATGAANGLNNGAGGTSATTCNPLSRKLHK ILETRLDNDKEMLEALKALSTFFVENSLRTRRNLRGDIERKSL AINEEFVSIFKEVKEELESISEDVQAMSNCCQDMTSRLQAAKE QTQDLIVKTTKLQSESQKLEIRAQVADAFLSKFQLTSDEMSLL RGTREGPITEDFFKALGRVKQIHNDVKVLLRTNQQTAGLEIM EQMALLQETAYERLYRWAQSECRTLTQESCDVSPVLTQAME ALQDRPVLYKYTLDEFGTARRSTVVRGFIDALTRGGPGGTPR PIEMHSHDPLRYVGDMLAWLHQATASEKEHLEALLKHVTTQ GVEENIQEVVGHITEGVCRPLKVRIEQVIVAEPGAVLLYKISN LLKFYHHTISGIVGNSATALLTTIEEMHLLSKKIFFNSLSLHAS KLMDKVELPPPDLGPSSALNQTLMLLREVLASHDSSVVPLDA RQADFVQVLSCVLDPLLQMCTVSASNLGTADMATFMVNSLY MMKTTLALFEFTDRRLEMLQFQIEAHLDTLINEQASYVLTRV GLSYIYNTVQQHKPEQGSLANMPNLDSVTLKAAMVQFDRYL SAPDNLLIPQLNFLLSATVKEQIVKQSTELVCRAYGEVYAAV MNPINEYKDPENILHRSPQQVQTLLS	COG6
[0867]	211	MDFSKFLADDFDVKEWINAAFRAGSKEAASGKADGHAATLV MKLQLFIQEVNHAVEETSHQALQNMPKVLRDVEALKQEASF LKEQMILVKEDIKKFEQDTSQSMQVLVEIDQVKSRMQLAAES LQEADKWSTLSADIEETFKTQDIAVISAKLTGMQNSLMMLVD TPDYSEKCVHLEALKNRLEALASPQIVAAFTSQAVDQSKVFV KVFTEIDRMPQLLAYYYKCHKVQLLAAWQELCQSDLSLDRQ LTGLYDALLGAWHTQIQWATQVFQKPHEVVMVLLIQTLGAL MPSLPSCLSNGVERAGPEQELTRLLEFYDATAHFAKGLEMAL LPHLHEHNLVKVTELVDAVYDPYKPYQLKYGDMEESNLLIQ MSAVPLEHGEVIDCVQELSHSVNKLFGLASAAVDRCVRFTNG LGTCGLLSALKSLFAKYVSDFTSTLQSIRKKCKLDHIPPNSLFQ EDWTAFQNSIRIIATCGELLRHCGDFEQQLANRILSTAGKYLS DSCSPRSLAGFQESILTDKKNSAKNPWQEYNYLQKDNPAEYA SLMEILYTLKEKGSSNHNLLAAPRAALTRLNQQAHQLAFDSV FLRIKQQLLLISKMDSWNTAGIGETLTDELPAFSLTPLEYISNIG QYIMSLPLNLEPFVTQEDSALELALHAGKLPFPPEQGDELPEL DNMADNWLGSIARATMQTYCDAILQIPELSPHSAKQLATDID YLINVMDALGLQPSRTLQHIVTLLKTRPEDYRQVSKGLPRRL ATTVATMRSVNY	COG7
	212	MATAATIPSVATATAAALGEVEDEGLLASLFRDRFPEAQWRE RPDVGRYLRELSGSGLERLRREPERLAEERAQLLQQTRDLAF ANYKTFIRGAECTERIHRLFGDVEASLGRLLDRLPSFQQSCRN FVKEAEEISSNRRMNSLTLNRHTEILEILEIPQLMDTCVRNSYY EEALELAAYVRRLERKYSSIPVIQGIVNEVRQSMQLMLSQLIQ QLRTNIQLPACLRVIGYLRRMDVFTEAELRVKFLQARDAWLR SILTAIPNDDPYFHITKTIEASRVHLFDIITQYRAIFSDEDPLLPP AMGEHTVNESAIFHGWVLQKVSQFLQVLETDLYRGIGGHLD SLLGQCMYFGLSFSRVGADFRGQLAPVFQRVAISTFQKAIQET VEKFQEEMNSYMLISAPAILGTSNMPAAVPATQPGTLQPPMV LLDFPPLACFLNNILVAFNDLRLCCPVALAQDVTGALEDALA	COG8

[0868]

	KVTKIILAFHRAEEAAFSSGEQELFVQFCTVFLEDLVPYLNRC LQVLFPPAQIAQTLGIPPTQLSKYGNLGHVNIGAIQEPLAFILP KRETLFTLDDQALGPELTAPAPEPPAEEPRLEPAGPACPEGGR AETQAEPPSVGP	
213	DRLLQQGSAVFQFRMSANSGLLPASMVMPLLGLVMKERCQT AGNPFFERFGIVVAATGMAVALFSSVLALGITRPVPTNTCVIL GLAGGVIIYIMKHSLSVGEVIEVLEVLLIFVYLNMILLYLLPRC FTPGEALLVLGGISFVLNQLIKRSLTLVESQGDPVDFFLLVVV VGMVLMGIFFSTLFVFMDSGTWASSIFFHLMTCVLSLGVVLP WLHRLIRRNPLLWLLQFLFQTDTRIYLLAYWSLLATLACLVV LYQNAKRSSSESKKHQAPTIARKYFHLIVVATYIPGIIFDRPLL YVAATVCLAVFIFLEYVRYFRIKPLGHTLRSFLSLFLDERDSGP LILTHIYLLLGMSLPIWLIPRPCTQKGSLGGARALVPYAGVLA VGVGDTVASIFGSTMGEIRWPGTKKTFEGTMTSIFAQIISVALI LIFDSGVDLNYSYAWILGSISTVSLLEAYTTQIDNLLLPLYLLIL LMA	DOLK
214	MSWIKEGELSLWERFCANIIKAGPMPKHIAFIMDGNRRYAKK CQVERQEGHSQGFNKLAETLRWCLNLGILEVTVYAFSIENFK RSKSEVDGLMDLARQKFSRLMEEKEKLQKHGVCIRVLGDLH LLPLDLQELIAQAVQATKNYNKCFLNVCFAYTSRHEISNAVR EMAWGVEQGLLDPSDISESLLDKCLYTNRSPHPDILIRTSGEV RLSDFLLWQTSHSCLVFQPVLWPEYTFWNLFEAILQFQMNHS VLQKARDMYAEERKRQQLERDQATVTEQLLREGLQASGDA QLRRTRLHKLSARREERVQGFLQALELKRADWLARLGTASA	DHDDS
215	MWAFSELPMPLLINLIVSLLGFVATVTLIPAFRGHFIAARLCGQ DLNKTSRQQIPESQGVISGAVFLIILFCFIPFPFLNCFVKEQCKA FPHHEFVALIGALLAICCMIFLGFADDVLNLRWRHKLLLPTAA SLPLLMVYFTNFGNTTIVVPKPFRPILGLHLDLGILYYVYMGL LAVFCTNAINILAGINGLEAGQSLVISASIIVFNLVELEGDCRD DHVFSLYFMIPFFFTTLGLLYHNWYPSRVFVGDTFCYFAGMT FAVVGILGHFSKTMLLFFMPQVFNFLYSLPQLLHIIPCPRHRIP RLNIKTGKLEMSYSKFKTKSLSFLGTFILKVAESLQLVTVHQS ETEDGEFTECNNMTLINLLLKVLGPIHERNLTLLLLLLQILGSA ITFSIRYQLVRLFYDV	DPAGT1
216	MASLEVSRSPRRSRRELEVRSPRQNKYSVLLPTYNERENLPLI VWLLVKSFSESGINYEIIIIDDGSPDGTRDVAEQLEKIYGSDRIL LRPREKKLGLGTAYIHGMKHATGNYIIIMDADLSHHPKFIPEFI RKQKEGNFDIVSGTRYKGNGGVYGWDLKRKIISRGANFLTQI LLRPGASDLTGSFRLYRKEVLEKLIEKCVSKGYVFQMEMIVR ARQLNYTIGEVPISFVDRVYGESK LGGNEIVSFLKGLLTLFATT	DPM1
217	MATGTDQVVGLGLVAVSLIIFTYYTAWVILLPFIDSQHVIHKY FLPRAYAVAIPLAAGLLLLLFVGLFISYVMLKTKRVTKKAQ	DPM2
218	MTKLAQWLWGLAILGSTWVALTTGALGLELPLSCQEVLWPL PAYLLVSAGCYALGTVGYRVATFHDCEDAARELQSQIQEAR ADLARRGLRF	DPM3
219	MESTLGAGIVIAEALQNQLAWLENVWLWITFLGDPKILFLFYF PAAYYASRRVGIAVLWISLITEWLNLIFKWFLFGDRPFWWVH ESGYYSQAPAQVHQFPSSCETGPGSPSGHCMITGAALWPIMT ALSSQVATRARSRWVRVMPSLAYCTFLLAVGLSRIFILAHFPH QVLAGLITGAVLGWLMTPRVPMERELSFYGLTALALMLGTS LIYWTLFTLGLDLSWSISLAFKWCERPEWIHVDSRPFASLSRD SGAALGLGIALHSPCYAQVRRAQLGNGQKIACLVLAMGLLG	G6PC3

[0869]

	PLDWLGHPPQISLFYIFNFLKYTLWPCLVLALVPWAVHMFSA QEAPPIHSS	
220	MCGIFAYLNYHVPRTRREILETLIKGLQRLEYRGYDSAGVGF DGGNDKDWEANACKIQLIKKKGKVKALDEEVHKQQDMDLD IEFDVHLGIAHTRWATHGEPSPVNSHPQRSDKNNEFIVIHNGII TNYKDLKKFLESKGYDFESETDTETIAKLVKYMYDNRESQDT SFTTLVERVIQQLEGAFALVFKSVHFPGQAVGTRRGSPLLIGV RSEHKLSTDHIPILYRTARTQIGSKFTRWGSQGERGKDKKGSC NLSRVDSTTCLFPVEEKAVEYYFASDASAVIEHTNRVIFLEDD DVAAVVDGRLSIHRIKRTAGDHPGRAVQTLQMELQQIMKGN FSSFMQKEIFEQPESVVNTMRGRVNFDDYTVNLGGLKDHIKEI QRCRRLILIACGTSYHAGVATRQVLEELTELPVMVELASDFLD RNTPVFRDDVCFFLSQSGETADTLMGLRYCKERGALTVGITN TVGSSISRETDCGVHINAGPEIGVASTKAYTSQFVSLVMFALM MCDDRISMQERRKEIMLGLKRLPDLIKEVLSMDDEIQKLATE LYHQKSVLIMGRGYHYATCLEGALKIKEITYMHSEGILAGEL KHGPLALVDKLMPVIMIIMRDHTYAKCQNALQQVVARQGRP VVICDKEDTETIKNTKRTIKVPHSVDCLQGILSVIPLQLLAFHL AVLRGYDVDFPRNLAKSVTVE	GFPT1
221	MLKAVILIGGPQKGTRFRPLSFEVPKPLFPVAGVPMIQHHIEA CAQVPGMQEILLIGFYQPDEPLTQFLEAAQQEFNLPVRYLQEF APLGTGGGLYHFRDQILAGSPEAFFVLNADVCSDFPLSAMLE AHRRQRHPFLLLGTTANRTQSLNYGCIVENPQTHEVLHYVEK PSTFISDIINCGIYLFSPEALKPLRDVFQRNQQDGQLEDSPGLW PGAGTIRLEQDVFSALAGQGQIYVHL TDGIWSQIKSAGSALYASRLYLSRYQDTHPERLAKHTPGGPW IRGNVYIHPTAKVAPSAVLGPNVSIGKGVTVGEGVRLRESIVL HGATLQEHTCVLHSIVGWGSTVGRWARVEGTPSDPNPNDPR ARMDSESLFKDGKLLPAITILGCRVRIPAEVLILNSIVLPHKELS RSFTNQIIL	GMPPA
222	MKALILVGGYGTRLRPLTLSTPKPLVDFCNKPILLHQVEALAA AGVDHVILAVSYMSQVLEKEMKAQEQRLGIRISMSHEEEPLG TAGPLALARDLLSETADPFFVLNSDVICDFPFQAMVQFHRHH GQEGSILVTKVEEPSKYGVVVCEADTGRIHRFVEKPQVFVSN KINAGMYILSPAVLQRIQLQPTSIEKEVFPIMAKEGQLYAMEL QGFWMDIGQPKDFLTGMCLFLQSLRQKQPERLCSGPGIVGNV LVDPSARIGQNCSIGPNVSLGPGVVVEDGVCIRRCTVLRDARI RSHSWLESCIVGWRCRVGQWVRMENVTVLGEDVIVNDELYL NGASVLPHKSIGESVPEPRIIM	GMPPB
223	MAARWRFWCVSVTMVVALLIVCDVPSASAQRKKEMVLSEK VSQLMEWTNKRPVIRMNGDKFRRLVKAPPRNYSVIVMFTAL QLHRQCVVCKQADEEFQILANSWRYSSAFTNRIFFAMVDFDE GSDVFQMLNMNSAPTFINFPAKGKPKRGDTYELQVRGFSAEQ IARWIADRTDVNIRVIRPPNYAGPLMLGLLLAVIGGLVYLRRS NMEFLFNKTGWAFAALCFVLAMTSGQMWNHIRGPPYAHKN PHTGHVNYIHGSSQAQFVAETHIVLLFNGGVTLGMVLLCEAA TSDMDIGKRKIMCVAGIGLVVLFFSWMLSIFRSKYHGYPYSFL MS	MAGT1
224	MAACEGRRSGALGSSQSDFLTPPVGGAPWAVATTVVMYPPP PPPPHRDFISVTLSFGENYDNSKSWRRRSCWRKWKQLSRLQR NMILFLLAFLLFCGLLFYINLADHWKALAFRLEEEQKMRPEIA GLKPANPPVLPAPQKADTDPENLPEISSQKTQRHIQRGPPHLQI RPPSQDLKDGTQEEATKRQEAPVDPRPEGDPQRTVISWRGAV	MAN1B1

		IEPEQGTELPSRRAEVPTKPPLPPARTQGTPVHLNYRQKGVID VFLHAWKGYRKFAWGHDELKPVSRSFSEWFGLGLTLIDALD TMWILGLRKEFEEARKWVSKKLHFEKDVDVNLFESTIRILGG LLSAYHLSGDSLFLRKAEDFGNRLMPAFRTPSKIPYSDVNIGT GVAHPPRWTSDSTVAEVTSIQLEFRELSRLTGDKKFQEAVEK VTQHIHGLSGKKDGLVPMFINTHSGLFTHLGVFTLGARADSY YEYLLKQWIQGGKQETQLLEDYVEAIEGVRTHLLRHSEPSKL TFVGELAHGRFSAKMDHLVCFLPGTLALGVYHGLPASHMEL AQELMETCYQMNRQMETGLSPEIVHFNLYPQPGRRDVEVKP ADRHNLLRPETVESLFYLYRVTGDRKYQDWGWEILQSFSRFT RVPSGGYSSINNVQDPQKPEPRDKMESFFLGETLKYLFLLFSD DPNLLSLDAYVFNTEAHPLPIWTPA	
[0870]	225	MRFRIYKRKVLILTLVVAACGFVLWSSNGRQRKNEALAPPLL DAEPARGAGGRGGDHPSVAVGIRRVSNVSAASLVPAVPQPEA DNLTLRYRSLVYQLNFDQTLRNVDKAGTWAPRELVLVVQVH NRPEYLRLLLDSLRKAQGIDNVLVIFSHDFWSTEINQLIAGVN FCPVLQVFFPFSIQLYPNEFPGSDPRDCPRDLPKNAALKLGCIN AEYPDSFGHYREAKFSQTKHHWWWKLHFVWERVKILRDYA GLILFLEEDHYLAPDFYHVFKKMWKLKQQECPECDVLSLGTY SASRSF YGMADKVDVKTWKSTEHNMGLALTRNAYQKLIECTDTFCT YDDYNWDWTLQYLTVSCLPKFWKVLVPQIPRIFHAGDCGMH HKKTCRPSTQSAQIESLLNNNKQYMFPETLTISEKFTVVAISPP RKNGGWGDIRDHELCKSYRRLQ	MGAT2
	226	MARGERRRRAVPAEGVRTAERAARGGPGRRDGRGGGPRSTA GGVALAVVVLSLALGMSGRWVLAWYRARRAVTLHSAPPVL PADSSSPAVAPDLFWGTYRPHVYFGMKTRSPKPLLTGLMWA QQGTTPGTPKLRHTCEQGDGVGPYGWEFHDGLSFGRQHIQD GALRLTTEFVKRPGGQHGGDWSWRVTVEPQDSGTSALPLVS LFFYVVTDGKEVLLPEVGAKGQLKFISGHTSELGDFRFTLLPP TSPGDTAPKYGSYNVFWTSNPGLPLLTEMVKSRLNSWFQHRP PGAPPERYLGLPGSLKWEDRGPSGQGQGQFLIQQVTLKIPISIE FVFESGSAQAGGNQALPRLAGSLLTQALESHAEGFRERFEKTF QLKEKGLSSGEQVLGQAALSGLLGGIGYFYGQGLVLPDIGVE GSEQKVDPALFPPVPLFTAVPSRSFFPRGFLWDEGFHQLVVQR WDPSLTREALGHWLGLLNADGWIGREQILGDEARARVPPEFL VQRAVHANPPTLLLPVAHMLEVGDPDDLAFLRKALPRLHAW FSWLHQSQAGPLPLSYRWRGRDPALPTLLNPKTLPSGLDDYP RASHPSVTERHLDLRCWVALGARVLTRLAEHLGEAEVAAEL GPLAASLEAAESLDELHWAPELGVFADFGNHTKAVQLKPRPP QGLVRVVGRPQPQLQYVDALGYVSLFPLLLRLLDPTSSRLGP LLDILADSRHLWSPFGLRSLAASSSFYGQRNSEHDPPYWRGA VWLNVNYLALGALHHYGHLEGPHQARAAKLHGELRANVVG NVWRQYQATGFLWEQYSDRDGRGMGCRPFHGWTSLVLLA MAEDY	MOGS
	227	MAAEADGPLKRLLVPILLPEKCYDQLFVQWDLLHVPCLKILL SKGLGLGIVAGSLLVKLPQVFKILGAKSAEGLSLQSVMLELV ALTGTMVYSITNNFPFSSWGEALFLMLQTITICFLVMHYRGQT VKGVAFLACYGLVLLVLLSPLTPLTVVTLLQASNVPAVVVGR LLQAATNYHNGHTGQLSAITVFLLFGGSLARIFTSIQETGDPL MAGTFVVSSLCNGLIAAQLLFYWNAKPPHKQKKAQ	MPDU1
	228	MAAPRVFPLSCAVQQYAWGKMGSNSEVARLLASSDPLAQIA EDKPYAELWMGTHPRGDAKILDNRISQKTLSQWIAENQDSLG SKVKDTFNGNLPFLFKVLSVETPLSIQAHPNKELAEKLHLQAP	MPI

		QHYPDANHKPEMAIALTPFQGLCGFRPVEEIVTFLKKVPEFQF LIGDEAATHLKQTMSHDSQAVASSLQSCFSHLMKSEKKVVVE QLNLLVKRISQQAAAGNNMEDIFGELLLQLHQQYPGDIGCFAI YFLNLLTLKPGEAMFLEANVPHAYLKGDCVECMACSDNTVR AGLTP KFIDVPTLCEMLSYTPSSSKDRLFLPTRSQEDPYLSIYDPPVPD FTIMKTEVPGSVTEYKVLALDSASILLMVQGTVIASTPTTQTPI PLQRGGVLFIGANESVSLKLTEPKDLLIFRACCLL	
[0871]	229	MAAAALGSSSGSASPAVAELCQNTPETFLEASKLLLTYADNIL RNPNDEKYRSIRIGNTAFSTRLLPVRGAVECLFEMGFEEGETH LIFPKKASVEQLQKIRDLIAIERSSRLDGSNKSHKVKSSQQPAA STQLPTTPSSNPSGLNQHTRNRQGQSSDPPSASTVAADSAILE VLQSNIQHVLVYENPALQEKALACIPVQELKRKSQEKLSRAR KLDKGINISDEDFLLLELLHWFKEE FFHWVNNVLCSKCGGQTRSRDRSLLPSDDELKWGAKEVEDH YCDACQFSNRFPRYNNPEKLLETRCGRCGEWANCFTLCCRA VGFEARYVWDYTDHVWTEVYSPSQQRWLHCDACEDVCDKP LLYEIGWGKKLSYVIAFSKDEVVDVTWRYSCKHEEVIARRTK VKEALLRDTINGLNKQRQLFLSENRRKELLQRIIVELVEFISPK TPKPGELGGRISGSVAWRVARGEMGLQRKETLFIPCENEKISK QLHLCYNIVKDRYVRVSNNNQTISGWENGVWKMESIFRKVE TDWHMVYLARKEGSSFAYISWKFECGSVGLKVDSISIRTSSQT FQTGTVEWKLRSDTAQVELTGDNSLHSYADFSGATEVILEAE LSRGDGDVAWQHTQLFRQSLNDHEENCLEIIIKFSDL	NGLY1
	230	MVKIVTVKTQAYQDQKPGTSGLRKRVKVFQSSANYAENFIQS IISTVEPAQRQEATLVVGGDGRFYMKEAIQLIARIAAANGIGR LVIGQNGILSTPAVSCIIRKIKAIGGIILTASHNPGGPNGDFGIKF NISNGGPAPEAITDKIFQISKTIEEYAVCPDLKVDLGVLGKQQF DLENKFKPFTVEIVDSVEAYATMLRSIFDFSALKELLSGPNRL KIRIDAMHGVVGPYVKKILCEELGAPANSAVNCVPLEDFGGH HPDPNLTYAADLVETMKSGEHDFGAAFDGDGDRNMILGKHG FFVNPSDSVAVIAANIFSIPYFQQTGVRGFARSMPTSGALDRV ASATKIALYETPTGWKFFGNLMDASKLSLCGEESFGTGSDHIR EKDGLWAVLAWLSILATRKQSVEDILKDHWQKYGRNFFTRY DYEEVEAEGANKMMKDLEALMFDRSFVGKQFSANDKVYTV EKADNFEYSDPVDGSISRNQGLRLIFTDGSRIVFRLSGTGSAG ATIRLYIDSYEKDVAKINQDPQVMLAPLISIALKVSQLQERTG RTAPTVIT	PGM1
	231	MDLGAITKYSALHAKPNGLILQYGTAGFRTKAEHLDHVMFR MGLLAVLRSKQTKSTIGVMVTASHNPEEDNGVKLVDPLGEM LAPSWEEHATCLANAEEQDMQRVLIDISEKEAVNLQQDAFV VIGRDTRPSSEKLSQSVIDGVTVLGGQFHDYGLLTTPQLHYM VYCRNTGGRYGKATIEGYYQKLSKAFVELTKQASCSGDEYR SLKVDCANGIGALKLREMEHYFSQGLSVQLFNDGSKGKLNH LCGADFVKSHQKPPQGMEIKSNERCCSFDGDADRIVYYYHD ADGHFHLIDGDKIATLISSFLKELLVEIGESLNIGVVQTAYANG SSTRYLEEVMKVPVYCTKTGVKHLHHKAQEFDIGVYFEANG HGTALFSTAVEMKIKQSAEQLEDKKRKAAKMLENIIDLFNQA AGDAISDMLVIEAILALKGLTVQQWDALYTDLPNRQLKVQV ADRRVISTTDAERQAVTPPGLQEAINDLVKKYKLSRAFVRPS GTEDVVRVYAEADSQESADHLAHEVSLAVFQLAGGIGERPQP GF	PGM3
	232	MGSQEVLGHAARLASSGLLLQVLFRLITFVLNAFILRFLSKEIV GVVNVRLTLLYSTTLFLAREAFRRACLSGGTQRDWSQTLNLL	RFT1

[0872]

	WLTVPLGVFWSLFLGWIWLQLLEVPDPNVVPHYATGVVLFG LSAVVELLGEPFWVLAQAHMFVKLKVIAESLSVILKSVLTAFL VLWLPHWGLYIFSLAQLFYTTVLVLCYVIYFTKLLGSPESTKL QTLPVSRITDLLPNITRNGAFINWKEAKLTWSFFKQSFLKQILT EGERYVMTFLNVLNFGDQGVYDIVNNLGSLVARLIFQPIEESF YIFFAKVLERGKDATLQKQEDVAVAAAVLESLLKLALLAGLT ITVFGFAYSQLALDIYGGTMLSSGSGPVLLRSYCLYVLLLAIN GVTECFTFAAMSKEEVDRYNFVMLALSSSFLVLSYLLTRWCG SVGFILANCFNMGIRITQSLCFIHRYYRRSPHRPLAGLHLSPVL LGTFALSGGVTAVSEVFLCCEQGWPARLAHIAVGAFCLGATL GTAFLTETKLIHFLRTQLGVPRRTDKMT	
233	MATYLEFIQQNEERDGVRFSWNVWPSSRLEATRMVVPLACL LTPLKERPDLPPVQYEPVLCSRPTCKAVLNPLCQVDYRAKLW ACNFCFQRNQFPPAYGGISEVNQPAELMPQFSTIEYVIQRGAQ SPLIFLYVVDTCLEEDDLQALKESLQMSLSLLPPDALVGLITFG RMVQVHELSCEGISKSYVFRGTKDLTAKQIQDMLGLTKPAMP MQQARPAQPQEHPFASSRFLQPVHKIDMNLTDLLGELQRDP WPVTQGKRPLRSTGVALSIAVGLLEGTFPNTGARIMLFTGGPP TQGPGMVVGDELKIPIRSWHDIEKDNARFMKKATKHYEMLA NRTAANGHCIDIYACALDQTGLLEMKCCANLTGGYMVMGD SFNTSLFKQTFQRIFTKDFNGDFRMAFGATLDVKTSRELKIAG AIGPCVSLNVKGPCVSENELGVGGTSQWKICGLDPTSTLGIYF EVVNQHNTPIPQGGRGAIQFVTHYQHSSTQRRIRVTTIARNW ADVQSQLRHIEAAFDQEAAAVLMARLGVFRAESEEGPDVLR WLDRQLIRLCQKFGQYNKEDPTSFRLSDSFSLYPQFMFHLRRS PFLQVFNNSPDESSYYRHHFARQDLTQSLIMIQPILYSYSFHGP PEPVLLDSSSILADRILLMDTFFQIVIYLGETIAQWRKAGYQD MPEYENFKHLLQAPLDDAQEILQARFPMPRYINTEHGGSQAR FLLSKVNPSQTHNNLYAWGQETGAPILTDDVSLQVFMDHLK KLAVSSAC	SEC23B
234	MAAPRDNVTLLFKLYCLAVMTLMAAVYTIALRYTRTSDKEL YFSTTAVCITEVIKLLLSVGILAKETGSLGRFKASLRENVLGSP KELLKLSVPSLVYAVQNNMAFLALSNLDAAVYQVTYQLKIP CTALCTVLMLNRTLSKLQWVSVFMLCAGVTLVQWKPAQAT KVVVEQNPLLGFGAIAIAVLCSGFAGVYFEKVLKSSDTSLWV RNIQMYLSGIIVTLAGVYLSDGAEIKEKGFFYGYTYYVWFVIF LASVGGLYTSVVVKYTDNIMKGFSAAAAIVLSTIASVMLFGL QITLTFALGTLLVCVSIYLYGLPRQDTTSIQQGETASKERVIGV	SLC35A1
235	MAAVGAGGSTAAPGPGAVSAGALEPGTASAAHRRLKYISLA VLVVQNASLILSIRYARTLPGDRFFATTAVVMAEVLKGLTCL LLLFAQKRGNVKHLVLFLHEAVLVQYVDTLKLAVPSLIYTLQ NNLQYVAISNLPAATFQVTYQLKILTTALFSVLMLNRSLSRLQ WASLLLLFTGVAIVQAQQAGGGGPRPLDQNPGAGLAAVVAS CLSSGFAGVYFEKILKGSSGSVWLRNLQLGLFGTALGLVGLW WAEGTAVATRGFFFGYTPAVWGVVLNQAFGGLLVAVVVKY ADNILKGFATSLSIVLSTVASIRLFGFHVDPLFALGAGLVIGAV YLYSLPRGAAKAIASASASASGPCVHQQPPGQPPPPQLSSHRG DLITEPFLPKLLTKVKGS	SLC35A2
236	MNRAPLKRSRILHMALTGASDPSAEAEANGEKPFLLRALQIA LVVSLYWVTSISMVFLNKYLLDSPSLRLDTPIFVTFYQCLVTT LLCKGLSALAACCPGAVDFPSLRLDLRVARSVLPLSVVFIGMI TFNNLCLKYVGVAFYNVGRSLTTVFNVLLSYLLLKQTTSFYA LLTCGIIIGGFWLGVDQEGAEGTLSWLGTVFGVLASLCVSLN AIYTTKVLPAVDGSIWRLTFYNNVNACILFLPLLLLLGELQAL	SLC35C1

		RDFAQLGSAHFWGMMTLGGLFGFAIGYVTGLQIKFTSPLTHN VSG TAKACAQTVLAVLYYEETKSFLWWTSNMMVLGGSSAYTWV RGWEMKKTPEEPSPKDSEKSAMGV	
	237	MAAMASLGALALLLLSSLSRCSAEACLEPQITPSYYTTSDAVI STETVFIVEISLTCKNRVQNMALYADVGGKQFPVTRGQDVGR YQVSWSLDHKSAHAGTYEVRFFDEESYSLLRKAQRNNEDISII PPLFTVSVDHRGTWNGPWVSTEVLAAAIGLVIYYLAFSAKSHI QA	SSR4
	238	MAPWAEAEHSALNPLRAVWLTLTAAFLLTLLLQLLPPGLLPG CAIFQDLIRYGKTKCGEPSRPAACRAFDVPKRYFSHFYIISVL WNGFLLWCLTQSLFLGAPFPSWLHGLLRILGAAQFQGGELAL SAFLVLVFLWLHSLRRLFECLYVSVFSNVMIHVVQYCFGLVY YVLVGLTVLSQVPMDGRNAYITGKNLLMQARWFHILGMMM FIWSSAHQYKCHVILGNLRKNKAGVVIHCNHRIPFGDWFEYV SSPNYLAELMIYVSMAVTFGFHNLTWWLVVTNVFFNQALSA FLSHQFYKSKFVSYPKHRKAFLPFLF	SRD5A3
	239	MAAAAPGNGRASAPRLLLLFLVPLLWAPAAVRAGPDEDLSH RNKEPPAPAQQLQPQPVAVQGPEPARVEKIFTPAAPVHTNKE DPATQTNLGFIHAFVAAISVIIVSELGDKTFFIAAIMAMRYNRL TVLAGAMLALGLMTCLSVLFGYATTVIPRVYTYYVSTVLFAI FGIRMLREGLKMSPDEGQEELEEVQAELKKKDEEFQRTKLLN GPGDVETGTSITVPQKKWLHFISPIFVQALTLTFLAEWGDRSQ LTTIVLAAREDPYGVAVGGTVGHCLCTGLAVIGGRMIAQKIS VRTVTIIGGIVFLAFAFSALFISPDSGF	TMEM165
[0873]	240	MSSWLGGLGSGLGQSLGQVGGSLASLTGQISNFTKDMLMEG TEEVEAELPDSRTKEIEAIHAILRSENERLKKLCTDLEEKHEAS EIQIKQQSTSYRNQLQQKEVEISHLKARQIALQDQLLKLQSAA QSVPSGAGVPATTASSSFAYGISHHPSAFHDDDMDFGDIISSQ QEINRLSNEVSRLESEVGHWRHIAQTSKAQGTDNSDQSEICKL QNIIKELKQNRSQEIDDHQHEMSVLQNAHQQKLTEISRRHREE LSDYEERIEELENLLQQGGSGVIETDLSKIYEMQKTIQVLQIEK VESTKKMEQLEDKIKDINKKLSSAENDRDILRREQEQLNVEK RQIMEECENLKLECSKLQPSAVKQSDTMTEKERILAQSASVEE VFRLQQALSDAENEIMRLSSLNQDNSLAEDNLKLKMRIEVLE KEKSLLSQEKEELQMSLLKLNNEYEVIKSTATRDISLDSELHD LRLNLEAKEQELNQSISEKETLIAEIEELDRQNQEATKHMILIK DQLSKQQNEGDSIISKLKQDLNDEKKRVHQLEDDKMDITKEL DVQKEKLIQSEVALNDLHLTKQKLEDKVENLVDQLNKSQES NVSIQKENLELKEHIRQNEEELSRIRNELMQSLNQDSNSNFKD TLLKEREAEVRNLKQNLSELEQLNENLKKVAFDVKMENEKL VLACEDVRHQLEECLAGNNQLSLEKNTIVETLKMEKGEIEAE LCWAKKRLLEEANKYEKTIEELSNARNLNTSALQLEHEHLIK LNQKKDMEIAELKKNIEQMDTDHKETKDVLSSSLEEQKQLTQ LINKKEIFIEKLKERSSKLQEELDKYSQALRKNEILRQTIEEKD RSLGSMKEENNHLQEELERLREEQSRTAPVADPKTLDSVTEL ASEVSQLNTIKEHLEEEIKHHQKIIEDQNQSKMQLLQSLQEQK KEMDEFRYQHEQMNATHTQLFLEKDEEIKSLQKTIEQIKTQL HEERQDIQTDNSDIFQETKVQSLNIENGSEKHDLSKAETERLV KGIKERELEIKLLNEKNISLTKQIDQLSKDEVGKLTQIIQQKDL EIQALHARISSTSHTQDVVYLQQQLQAYAMEREKVFAVLNEK TRENSHLKTEYHKMMDIVAAKEAALIKLQDENKKLSTRFESS GQDMFRETIQNLSRIIREKDIEIDALSQKCQTLLAVLQTSSTGN EAGGVNSNQFEELLQERDKLKQQVKKMEEWKQQVMTTVQN	TRIP11

[0874]

	MQHESAQLQEELHQLQAQVLVDSDNNSKLQVDYTGLIQSYE QNETKLKNFGQELAQVQHSIGQLCNTKDLLLGKLDIISPQLSS ASLLTPQSAECLRASKSEVLSESSELLQQELEELRKSLQEKDA TIRTLQENNHRLSDSIAATSELERKEHEQTDSEIKQLKEKQDV LQKLLKEKDLLIKAKSDQLLSSNENFTNKVNENELLRQAVTN LKERILILEMDIGKLKGENEKIVETYRGKETEYQALQETNMKF SMMLREKEFECHSMKEKALAFEQLLKEKEQGKTGELNQLLN AVKSMQEKTVVFQQERDQVMLALKQKQMENTALQNEVQRL RDKEFRSNQELERLRNHLLESEDSYTREALAAEDREAKLRKK VTVLEEKLVSSSNAMENASHQASVQVESLQEQLNVVSKQRD ETALQLSVSQEQVKQYALSLANLQMVLEHFQQEEKAMYSAE LEKQKQLIAEWKKNAENLEGKVISLQECLDEANAALDSASRL TEQLDVKEEQIEELKRQNELRQEMLDDVQKKLMSLANSSEG KVDKVLMRNLFIGHFHTPKNQRHEVLRLMGSILGVRREEME QLFHDDQGGVTRWMTGWLGGGSKSVPNTPLRPNQQSVVNS SFSELFVKFLETESHPSIPPPKLSVHDMKPLDSPGRRKRDTNAP ESFKDTAESRSGRRTDVNPFLAPRSAAVPLINPAGLGPGGPGH LLLKPISDVLPTFTPLPALPDNSAGVVLKDLLKQ	
[0874]	241	MGARGAPSRRRQAGRRLRYLPTGSFPFLLLLLLLCIQLGGGQ KKKENLLAEKVEQLMEWSSRRSIFRMNGDKFRKFIKAPPRNY SMIVMFTALQPQRQCSVCRQANEEYQILANSWRYSSAFCNKL FFSMVDYDEGTDVFQQLNMNSAPTFMHFPPKGRPKRADTFD LQRIGFAAEQLAKWIADRTDVHIRVFRPPNYSGTIALALLVSL VGGLLYLRRNNLEFIYNKTGWAMVSLCIVFAMTSGQMWNHI RGPPYAHKNPHNGQVSYIHGSSQAQFVAESHIILVLNAAITMG MVLLNE AATSKGDVGKRRIICLVGLGLVVFFFSFLLSIFRSKYHGYPYSD LDFE	TUSC3
	242	MVCVLVLAAAAGAVAVFLILRIWVVLRSMDVTPRESLSILVV AGSGGHTTEILRLLGSLSNAYSPRHYVIADTDEMSANKINSFE LDRADRDPSNMYTKYYIHRIPRSREVQQSWPSTVFTTLHSMW LSFPLIHRVKPDLVLCNGPGTCVPICVSALLLGILGIKKVIIVYV ESICRVETLSMSGKILFHLSDYFIVQWPALKEKYPKSVYLGRI V	ALG14
	243	MRLREPLLSGSAAMPGASLQRACRLLVAVCALHLGVTLVYY LAGRDLSRLPQLVGVSTPLQGGSNSAAAIGQSSGELRTGGAR PPPPLGASSQPRPGGDSSPVVDSGPGPASNLTSVPVPHTTALSL PACPEESPLLVGPMLIEFNMPVDLELVAKQNPNVKMGGRYAP RDCVSPHKVAIIIPFRNRQEHLKYWLYYLHPVLQRQQLDYGI YVINQAGDTIFNRAKLLNVGFQEALKDYDYTCFVFSDVDLIP MNDHNAYRCFSQPRHISVAMDKFGFSLPYVQYFGGVSALSK QQFLTINGFPNNYWGWGGEDDDIFNRLVFRGMSISRPNAVVG RCRMIRHSRDKKNEPNPQRFDRIAHTKETMLSDGLNSLTYQV LDVQRYPLYTQITVDIGTPS	B4GALT1
	244	MGYFRCARAGSFGRRRKMEPSTAARAWALFWLLLPLLGAVC ASGPRTLVLLDNLNVRETHSLFFRSLKDRGFELTFKTADDPSL SLIKYGEFLYDNLIIFSPSVEDFGGNINVETISAFIDGGGSVLVA ASSDIGDPLRELGSECGIEFDEEKTAVIDHHNYDISDLGQHTLI VADTENLLKAPTIVGKSSLNPILFRGVGMVADPDNPLVLDILT GSSTSYSFFPDKPITQYPHAVGKNTLLIAGLQARNNARVIFSGS LDFFSDSFFNSAVQKAAPGSQRYSQTGNYELAVALSRWVFKE EGVLRVGPVSHHRVGETAPPNAYTVTDLVEYSIVIQQLSNGK WVPFDGDDIQLEFVRIDPFVRTFLKKKGGKYSVQFKLPDVYG VFQFKVDYNRLGYTHLYSSTQVSVRPLQHTQYERFIPSAYPY	DDOST

		YASAFSMMLGLFIFSIVFLHMKEKEKSD	
	245	MTGLYELVWRVLHALLCLHRTLTSWLRVRFGTWNWIWRRC CRAASAAVLAPLGFTLRKPPAVGRNRRHHRHPRGGSCLAAA HHRMRWRADGRSLEKLPVHMGLVITEVEQEPSFSDIASLVV WCMAVGISYISVYDHQGIFKRNNSRLMDEILKQQQELLGLDC SKYSPEFANSNDKDDQVLNCHLAVKVLSPEDGKADIVRAAQ DFCQLVAQKQKRPTDLDVDTLASLLSSNGCPDPDLVLKFGPV DSTLGFLPWHIRLTEIVSLPSHLNISYEDFFSALRQYAACEQRL GK	NUS1
	246	MAPPGSSTVFLLALTIIASTWALTPTHYLTKHDVERLKASLDR PFTNLESAFYSIVGLSSLGAQVPDAKKACTYIRSNLDPSNVDS LFYAAQASQALSGCEISISNETKDLLLAAVSEDSSVTQIYHAV AALSGFGLPLASQEALSALTARLSKEETVLATVQALQTASHLS QQADLRSIVEEIEDLVARLDELGGVYLQFEEGLETTALFVAAT YKLMDHVGTEPSIKEDQVIQLMNAIFSKKNFESLSEAFSVASA AAVLSHNRYHVPVVVVPEGSASDTHEQAILRLQVTNVLSQPL TQATVKLEHAKSVASRATVLQKTSFTPVGDVFELNFMNVKFS SGYYDFLVEVEGDNRYIANTVELRVKISTEVGITNVDLSTVDK DQSIAPKTTRVTYPAKAKGTFIADSHQNFALFFQLVDVNTGA ELTPHQTFVRLHNQKTGQEVVFVAEPDNKNVYKFELDTSERK IEFDSASGTYTLYLIIGDATLKNPILWNVADVVIKFPEEEAPST VLSQNLFTPKQEIQHLFREPEKRPPTV VSNTFTALILSPLLLLFALWIRIGANVSNFTFAPSTIIFHLGHAA MLGLMYVYWTQLNMFQTLKYLAILGSVTFLAGNRMLAQQA VKRTAH	RPN2
[0875]	247	MTTYLEFIQQNEERDGVRFSWNVWPSSRLEATRMVVPVAAL FTPLKERPDLPPIQYEPVLCSRTTCRAVLNPLCQVDYRAKLW ACNFCYQRNQFPPSYAGISELNQPAELLPQFSSIEYVVLRGPQ MPLIFLYVVDTCMEDEDLQALKESMQMSLSLLPPTALVGLITF GRMVQVHELGCEGISKSYVFRGTKDLSAKQLQEMLGLSKVP LTQATRGPQVQQPPPSNRFLQPVQKIDMNLTDLLGELQRDPW PVPQGKRPLRSSGVALSIAVGLLECTFPNTGARIMMFIGGPAT QGPGM VVGDELKTPIRSWHDIDKDNAKYVKKGTKHFEALANRAATT GHVIDIYACALDQTGLLEMKCCPNLTGGYMVMGDSFNTSLF KQTFQRVFTKDMHGQFKMGFGGTLEIKTSREIKISGAIGPCVS LNSKGPCVSENEIGTGGTCQWKICGLSPTTTLAIYFEVVNQHN APIPQGGRGAIQFVTQYQHSSGQRRIRVTTIARNWADAQTQIQ NIAASFDQEAAAILMARLAIYRAETEEGPDVLRWLDRQLIRLC QKFGEYHKDDPSSFRFSETFSLYPQFMFHLRRSSFLQVFNNSP DESSYYRHHFMRQDLTQSLIMIQPILYAYSFSGPPEPVLLDSSS ILADRILLMDTFFQILIYHGETIAQWRKSGYQDMPEYENFRHL LQAPVDDAQEILHSRFPMPRYIDTEHGGSQARFLLSKVNPSQT HNNMYAWGQESGAPILTDDVSLQVFMDHLKKLAVSSAA	SEC23A
	248	MFANLKYVSLGILVFQTTSLVLTMRYSRTLKEEGPRYLSSTA VVVAELLKIMACILLVYKDSKCSLRALNRVLHDEILNKPMET LKLAIPSGIYTLQNNLLYVALSNLDAATYQVTYQLKILTTALF SVSMLSKKLGVYQWLSLVILMTGVAFVQWPSDSQLDSKELS AGSQFVGLMAVLTACFSSGFAGVYFEKILKETKQSVWIRNIQ LGFFGSIFGLMGVYIYDGELVSKNGFFQGYNRLTWIVVVLQA LGGLVIAAVIKYADNILKGFATSLSIILSTLISYFWLQDFVPTSV FFLGAILVITATFLYGYDPKPAGNPTKA	SLC35A3
	249	MGLLVFVRNLLLALCLFLVLGFLYYSAWKLHLLQWEEDSNS	ST3GAL3

		VVLSFDSAGQTLGSEYDRLGFLLNLDSKLPAELATKYANFSE GACKPGYASALMTAIFPRFSKPAPMFLDDSFRKWARIREFVPP FGIKGQDNLIKAILSVTKEYRLTPALDSLRCRRCIIVGNGGVLA NKSLGSRIDDYDIVVRLNSAPVKGFEKDVGSKTTLRITYPEGA MQRPEQYERDSLFVLAGFKWQDFKWLKYIVYKERVSASDGF WKSVATRVPKEPPEIRILNPYFIQEAAFTLIGLPFNNGLMGRG NIPTLGSVAVTMALHGCDEVAVAGFGYDMSTPNAPLHYYET VRMAAIKESWTHNIQREKEFLRKLVKARVITDLSSGI	
	250	MTKFGFLRLSYEKQDTLLKLLILSMAAVLSFSTRLFAVLRFES VIHEFDPYFNYRTTRFLAEEGFYKFHNWFDDRAWYPLGRIIG GTIYPGLMITSAAIYHVLHFFHITIDIRNVCVFLAPLFSSFTTIVT YHLTKELKDAGAGLLAAAMIAVVPGYISRSVAGSYDNEGIAI FCMLLTYYMWIKAVKTGSICWAAKCALAYFYMVSSWGGYV FLINLIPLHVLVLMLTGRFSHRIYVAYCTVYCLGTILSMQISFV GFQPVLSSEHMAAFGVFGLCQIHAFVDYLRSKLNPQQFEVLF RSVISLVGFVLLTVGALLMLTGKISPWTGRFYSLLDPSYAKNN IPIIASVSEHQPTTWSSYYFDLQLLVFMFPVGLYYCFSNLSDAR IFIIMYGVTSMYFSAVMVRLMLVLAPVMCILSGIGVSQVLSTY MKNLDISRPDKKSKKQQDSTYPIKNEVASGMILVMAFFLITYT FHSTWVTSEAYSSPSIVLSARGGDGSRIIFDDFREAYYWLRHN TPEDAKVMSWWDYGYQITAMANRTILVDNNTWNNTHISRV GQAMASTEEKAYEIMRELDVSYVLVIFGGLTGYSSDDINKFL WMVRIGGSTDTGKHIKENDYYTPTGEFRVDREGSPVLLNCL MYKMCYYRFGQVYTEAKRPPGFDRVRNAEIGNKDFELDVLE EAYTTEHWLVRIYKVKDLDNRGLSRT	STT3A
[0876]	251	MAEPSAPESKHKSSLNSSPWSGLMALGNSRHGHHGPGAQCA HKAAGGAAPPKPAPAGLSGGLSQPAGWQSLLSFTILFLAWLA GFSSRLFAVIRFESIIHEFDPWFNYRSTHHLASHGFYEFLNWFD ERAWYPLGRIVGGTVYPGLMITAGLIHWILNTLNITVHIRDVC VFLAPTFSGLTSISTFLLTRELWNQGAGLLAACFIAIVPGYISRS VAGSFDNEGIAIFALQFTYYLWVKSVKTGSVFWTMCCCLSYF YMVSAWGGYVFIINLIPLHVFVLLLMQRYSKRVYIAYSTFYIV GLILSMQIPFVGFQPIRTSEHMAAAGVFALLQAYAFLQYLRDR LTKQEFQTLFFLGVSLAAGAVFLSVIYLTYTGYIAPWSGRFYS LWDTGYAKIHIPIIASVSEHQPTTWVSFFFDLHILVCTFPAGLW FCIKNINDERVFVALYAISAVYFAGVMVRLMLTLTPVVCMLS AIAFSNVFEHYLGDDMKRENPPVEDSSDEDDKRNQGNLYDK AGKVRKHATEQEKTEEGLGPNIKSIVTMLMLMLLMMFAVHC TWVTSNAYSSPSVVLASYNHDGTRNILDDFREAYFWLRQNT DEHARVMSWWDYGYQIAGMANRTTLVDNNTWNNSHIALV GKAMSSNETAAYKIMRTLDVDYVLVIFGGVIGYSGDDINKFL WMVRIAEGEHPKDIRESDYFTPQGEFRVDKAGSPTLLNCLMY KMSYYRFGEMQLDFRTPPGFDRTRNAEIGNKDIKFKHLEEAF TSEHWLVRIYKVKAPDNRETLDHKPRVTNIFPKQKYLSKKTT KRKRGYIKNKLVFKKGKKISKKTV	STT3B
	252	MARKSNLPVLLVPFLLCQALVRCSSPLPLVVNTWPFKNATEA AWRALASGGSALDAVESGCAMCEREQCDGSVGFGGSPDELG ETTLDAMIMDGTTMDVGAVGDLRRIKNAIGVARKVLEHTTH TLLVGESATTFAQSMGFINEDLSTTASQALHSDWLARNCQPN YWRNVIPDPSKYCGPYKPPGILKQDIPIHKETEDDRGHDTIGM VVIHKTGHIAAGTSTNGIKFKIHGRVGDSPIPGAGAYADDTAG AAAATGNGDILMRFLPSYQAVEYMRRGEDPTIACQKVISRIQ KHFPEF FGAVICANVTGSYGAACNKLSTFTQFSFMVYNSEKNQPTEEK	AGA

[0877]

	VDCI	
253	MGAPRSLLLALAAGLAVARPPNIVLIFADDLGYGDLGCYGHP SSTTPNLDQLAAGGLRFTDFYVPVSLCTPSRAALLTGRLPVR MGMYPGVLVPSSRGGLPLEEVTVAEVLAARGYLTGMAGKW HLGVGPEGAFLPPHQGFHRFLGIPYSHDQGPCQNLTCFPPATP CDGGCDQGLVPIPLLANLSVEAQPPWLPGLEARYMAFAHDL MADAQRQDRPFFLYYASHHTHYPQFSGQSFAERSGRGPFGDS LMELDAAVGTLMTAIGDLGLLEETLVIFTADNGPETMRMSRG GCSGLLRC GKGTTYEGGVREPALAFWPGHIAPGVTHELASSLDLLPTLAA LAGAPLPNVTLDGFDLSPLLLGTGKSPRQSLFFYPSYPDEVRG VFAVRTGKYKAHFFTQGSAHSDTTADPACHASSSLTAHEPPL LYDLSKDPGENYNLLGGVAGATPEVLQALKQLQLLKAQLDA AVTFGPSQVARGEDPALQICCHPGCTPRPACCHCPDPHA	ARSA
254	MGPRGAASLPRGPGPRRLLLPVVLPLLLLLLLAPPGSGAGASR PPHLVFLLADDLGWNDVGFHGSRIRTPHLDALAAGGVLLDN YYTQPLCTPSRSQLLTGRYQIRTGLQHQIIWPCQPSCVPLDEK LLPQLLKEAGYTTHMVGKWHLGMYRKECLPTRRGFDTYFG YLLGSEDYYSHERCTLIDALNVTRCALDFRDGEEVATGYKN MYSTNIFTKRAIALITNHPPEKPLFLYLALQSVHEPLQVPEEYL KPYDFIQDKNRHHYAGMVSLMDEAVGNVTAALKSSGLWNN TVFIFSTDNGGQTLAGGNNWPLRGRKWSLWEGGVRGVGFV ASPLLKQKGVKNRELIHISDWLPTLVKLARGHTNGTKPLDGF DVWKTISEGSPSPRIELLHNIDPNFVDSSPCPRNSMAPAKDDSS LPEYSAFNTSVHAAIRHGNWKLLTGYPGCGYWFPPPSQYNVS EIPSSDPPTKTLWLFDIDRDPEERHDLSREYPHIVTKLLSRLQF YHKHSVPVYFPAQDPRCDPKATGVWGPWM	ARSB
255	MPGRSCVALVLLAAAVSCAVAQHAPPWTEDCRKSTYPPSGP TYRGAVPWYTINLDLPPYKRWHELMLDKAPVLKVIVNSLKN MINTFVPSGKIMQVVDEKLPGLLGNFPGPFEEEMKGIAAVTDI PLGEIISFNIFYELFTICTSIVAEDKKGHLIHGRNMDFGVFLGW NINNDTWVITEQLKPLTVNLDFQRNNKTVFKASSFAGYVGML TGFKPGLFSLTLNERFSINGGYLGILEWILGKKDVMWIGFLTR TVLENSTSYEEAKNLLTKTKILAPAYFILGGNQSGEGCVITRD RKESLDVYELDAKQGRWYVVQTNYDRWKHPFFLDDRRTPA KMCLNRTSQENISFETMYDVLSTKPVLNKLTVYTTLIDVTKG QFETYLRDCPDPCIGW	ASAH1
256	MSADSSPLVGSTPTGYGTLTIGTSIDPLSSSVSSVRLSGYCGSP WRVIGYHVVVWMMAGIPLLLFRWKPLWGVRLRLRPCNLAH AETLVIEIRDKEDSSWQLFTVQVQTEAIGEGSLEPSPQSQAED GRSQAAVGAVPEGAWKDTAQLHKSEEAVSVGQKRVLRYYL FQGQRYIWIETQQAFYQVSLLDHGRSCDDVHRSRHGLSLQDQ MVRKAIYGPNVISIPVKSYPQLLVDEALNPYYGFQAFSIALWL ADHYYWYALCIFLISSISICLSLYKTRKQSQTLRDMVKLSMRV CVCRPGGEEEWVDSSELVPGDCLVLPQEGGLMPCDAALVAG ECMVNESSLTGESIPVLKTALPEGLGPYCAETHRRHTLFCGTL ILQARAYVGPHVLAVVTRTGFCTAKGGLVSSILHPRPINFKFY KHSMKFVAALSVLALLGTIYSIFILYRNRVPLNEIVIRALDLVT VVVPPALPAAMTVCTLYAQSRLRRQGIFCIHPLRINLGGKLQL VCFDKTGTLTEDGLDVMGVVPLKGQAFLPLV PEPRRLPVGPLLRALATCHALSRLQDTPVGDPMDLKMVESTG WVLEEEPAADSAFGTQVLAVMRPPLWEPQLQAMEEPPVPVS VLHRFPFSSALQRMSVVVAWPGATQPEAYVKGSPELVAGLC NPETVPTDFAQMLQSYTAAGYRVVALASKPLPTVPSLEAAQQ	ATP13A2

		LTRDTVEGDLSLLGLLVMRNLLKPQTTPVIQALRRTRIRAVM VTGDNLQTAVTVARGCGMVAPQEHLIIVHATHPERGQPASLE FLPMESPTAVNGVKDPDQAASYTVEPDPRSRHLALSGPTFGII VKHFPKL LPKVLVQGTVFARMAPEQKTELVCELQKLQYCVGMCGDGA NDCGALKAADVGISLSQAEASVVSPFTSSMASIECVPMVIREG RCSLDTSFSVFKYMALYSLTQFISVLILYTINTNLGDLQFLAID LVITTTVAVLMSRTGPALVLGRVRPPGALLSVPVLSSLLLQM VLVTGVQLGGYFLTLAQPWFVPLNRTVAAPDNLPNYENTVV FSLSSFQYLILAAAVSKGAPFRRPLYTNVPFLVALALLSSVLV GLVLVPGLLQGPLALRNITDTGFKLLLLGLVTLNFVGAFMLES VLDQCLPACLRRLRPKRASKKRFKQLERELAEQPWPPLPAGP LR	
[0878]	257	MGGCAGSRRRFSDSEGEETVPEPRLPLLDHQGAHWKNAVGF WLLGLCNNFSYVVMLSAAHDILSHKRTSGNQSHVDPGPTPIP HNSSSRFDCNSVSTAAVLLADILPTLVIKLLAPLGLHLLPYSPR VLVSGICAAGSFVLVAFSHSVGTSLCGVVFASISSGLGEVTFLS LTAFYPRAVISWWSSGTGGAGLLGALSYLGLTQAGLSPQQTL LSMLGIPALLLASYFLLLTSPEAQDPGGEEEAESAARQPLIRTE APESKPGSSSSLSLRERWTVFKGLLWYIVPLVVVYFAEYFINQ GLFELLFFWNTSLSHAQQYRWYQMLYQAGVFASRSSLRCCRI RFTWALALLQCLNLVFLLADVWFGFLPSIYLVFLIILYEGLLG GAAYVNTFHNIALETSDEHREFAMAATCISDTLGISLSGLLAL PLHDFLCQLS	CLN3
	258	MAQEVDTAQGAEMRRGAGAARGRASWCWALALLWLAVVP GWSRVSGIPSRRHWPVPYKRFDFRPKPDPYCQAKYTFCPTGS PIPVMEGDDDIEVFRLQAPVWEFKYGDLLGHLKIMHDAIGFR STLTGKNYTMEWYELFQLGNCTFPHLRPEMDAPFWCNQGAA CFFEGIDDVHWKENGTLVQVATISGNMFNQMAKWVKQDNE TGIYYETWNVKASPEKGAETWFDSYDCSKFVLRTFNKLAEFG AEFKNIETNYTRIFLYSGEPTYLGNETSVFGPTGNKTLGLAIKR FYYPFKPHLPTKEFLLSLLQIFDAVIVHKQFYLFYNFEYWFLP MKFPFIKITYEEIPLPIRNKTLSGL	CLN5
	259	MEATRRRQHLGATGGPGAQLGASFLQARHGSVSADEAARTA PFHLDLWFYFTLQNWVLDFGRPIAMLVFPLEWFPLNKPSVGD YFHMAYNVITPFLLLKLIERSPRTLPRSITYVSIIIFIMGASIHLV GDSVNHRLLFSGYQHHLSVRENPIIKNLKPETLIDSFELLYYY DEYLGHCMWYIPFFLILFMYFSGCFTASKAESLIPGPALLLVA PSGLYYWYLVTEGQIFILFIFTFFAMLALVLHQKRKRLFLDSN GLFLFSSFALTLLLVALWVAWLWNDPVLRKKYPGVIYVPEP WAFYTLHVSSRH	CLN6
	260	MNPASDGGTSESIFDLDYASWGIRSTLMVAGFVFYLGVFVVC HQLSSSLNATYRSLVAREKVFWDLAATRAVFGVQSTAAGLW ALLGDPVLHADKARGQQNWCWFHITTATGFFCFENVAVHLS NLIFRTFDLFLVIHHLFAFLGFLGCLVNLQAGHYLAMTTLLLE MSTPFTCVSWMLLKAGWSESLFWKLNQWLMIHMFHCRMVL TYHMWWVCFWHWDGLVSSLYLPHLTLFLVGLALLTLIINPY WTHKKTQQLLNPVDWNFAQPEAKSRPEGNGQLLRKKRP	CLN8
	261	MIRNWLTIFILFPLKLVEKCESSVSLTVPPVVKLENGSSTNVSL TLRPPLNATLVITFEITFRSKNITILELPDEVVVPPGVTNSSFQV TSQNVGQLTVYLHGNHSNQTGPRIRFLVIRSSAISIINQVIGWI YFVAWSISFYPQVIMNWRRKSVIGLSFDFVALNLTGFVAYSV FNIGLLWVPYIKEQFLLKYPNGVNPVNSNDVFFSLHAVVLTLI	CTNS

[0879]

	IIVQCCLYERGGQRVSWPAIGFLVLAWLFAFVTMIVAAVGVT TWLQFLFCFSYIKLAVTLVKYFPQAYMNFYYKSTEGWSIGNV LLDFTGGSFSLLQMFLQSYNNDQWTLIFGDPTKFGLGVFSIVF DVVFFIQHFCLYRKRPGYDQLN	
262	MIRAAPPPLFLLLLLLLLLVSWASRGEAAPDQDEIQRLPGLAK QPSFRQYSGYLKGSGSKHLHYWFVESQKDPENSPVVLWLNG GPGCSSLDGLLTEHGPFLVQPDGVTLEYNPYSWNLIANVLYL ESPAGVGFSYSDDKFYATNDTEVAQSNFEALQDFFRLFPEYK NNKLFLTGESYAGIYIPTLAVLVMQDPSMNLQGLAVGNGLSS YEQNDNSLVYFAYYHGLLGNRLWSSLQTHCCSQNKCNFYDN KDLECVTNLQEVARIVGNSGLNIYNLYAPCAGGVPSHFRYEK DTVVVQD LGNIFTRLPLKRMWHQALLRSGDKVRMDPPCTNTTAASTYL NNPYVRKALNIPEQLPQWDMCNFLVNLQYRRLYRSMNSQYL KLLSSQKYQILLYNGDVDMACNFMGDEWFVDSLNQKMEVQ RRPWLVKYGDSGEQIAGFVKEFSHIAFLTIKGAGHMVPTDKP LAAFTMFSRFLNKQPY	CTSA
263	MQPSSLLPLALCLLAAPASALVRIPLHKFTSIRRTMSEVGGSV EDLIAKGPVSKYSQAVPAVTEGPIPEVLKNYMDAQYYGEIGIG TPPQCFTVVFDTGSSNLWVPSIHCKLLDIACWIHHKYNSDKSS TYVKNGTSFDIHYGSGSLSGYLSQDTVSVPCQSASSASALGG VKVERQVFG EATKQPGITFIAAKFDGILGMAYPRISVNNVLPVFDNLMQQKL VDQNIFSFYLSRDPDAQPGGELMLGGTDSKYYKGSLSYLNVT RKAYWQVHLDQVEVASGLTLCKEGCEAIVDTGTSLMVGPVD EVRELQKAIGAVPLIQGEYMIPCEKVSTLPAITLKLGGKGYKL SPEDYTLKVSQAGKTLCLSGFMGMDIPPPSGPLWILGDVFIGR YYTVFDRDNNRVGFAEAARL	CTSD
264	MAPWLQLLSLLGLLPGAVAAPAQPRAASFQAWGPPSPELLAP TRFALEMFNRGRAAGTRAVLGLVRGRVRRAGQGSLYSLEAT LEEPPCNDPMVCRLPVSKKTLLCSFQVLDELGRHVLLRKDCG PVDTKVPGAGEPKSAFTQGSAMISSLSQNHPDNRNETFSSVIS LLNEDPLSQDLPVKMASIFKNFVITYNRTYESKEEARWRLSVF VNNMVRAQKIQALDRGTAQYGVTKFSDLTEEEFRTIYLNTLL RKEPGNKMKQAKSVGDLAPPEWDWRSKGAVTKVKDQGMC GSCWAFSVTGNVEGQWFLNQGTLLSLSEQELLDCDKMDKAC MGGLPSNAYSAIKNLGGLETEDDYSYQGHMQSCNFSAEKAK VYINDSVELSQNEQKLAAWLAKRGPISVAINAFGMQFYRHGI SRPLRPLCSPWLIDHAVLLVGYGNRSDVPFWAIKNSWGTDW GEKGYYYLHRGSGACGVNTMASSAVVD	CTSF
265	MWGLKVLLLPVVSFALYPEEILDTHWELWKKTHRKQYNNK VDEISRRLIWEKNLKYISIHNLEASLGVHTYELAMNHLGDMT SEEVVQKMTGLKVPLSHSRSNDTLYIPEWEGRAPDSVDYRKK GYVTPVKNQGQCGSCWAFSSVGALEGQLKKKTGKLLNLSPQ NLVDCVSENDGCGGGYMTNAFQYVQKNRGIDSEDAYPYVG QEESCMYNPTGKAAKCRGYREIPEGNEKALKRAVARVGPVS VAIDASLTSFQFYSKGVYYDESCNSDNLNHAVLAVGYGIQKG NKHWIIKNSWGENWGNKGYILMARNKNNACGIANLASFPK M	CTSK
266	MADQRQRSLSTSGESLYHVLGLDKNATSDDIKKSYRKLALK YHPDKNPDNPEAADKFKEINNAHAILTDATKRNIYDKYGSLG LYVAEQFGEENVNTYFVLSSWWAKALFVFCGLLTCCYCCCC LCCCFNCCCGKCKPKAPEGEETEFYVSPEDLEAQLQSDEREA	DNAJC5

[0880]

	TDTPIVIQPASATETTQLTADSHPSYHTDGFN	
267	MRAPGMRSRPAGPALLLLLLFLGAAESVRRAQPPRRYTPDWP SLDSRPLPAWFDEAKFGVFIHWGVFSVPAWGSEWFWWHWQ GEGRPQYQRFMRDNYPPGFSYADFGPQFTARFFHPEEWADLF QAAGAKYVVLTTKHHEGFTNWPSPVSWNWNSKDVGPHRDL VGELGTALRKRNIRYGLYHSLLEWFHPLYLLDKKNGFKTQHF VSAKTMPELYDLVNSYKPDLIWSDGEWECPDTYWNSTNFLS WLYNDSPVKDEVVVNDRWGQNCSCHHGGYYNCEDKFKPQS LPDHKWEMCTSIDKFSWGYRRDMALSDVTEESEIISELVQTV SLGGNYLLNIGPTKDGLIVPIFQERLLAVGK WLSINGEAIYASKPWRVQWEKNTTSVWYTSKGSAVYAIFLH WPENGVLNLESPITTSTTKITMLGIQGDLKWSTDPDKGLFISLP QLPPSAVPAEFAWTIKLTGVK	FUCA1
268	MGVRHPPCSHRLLAVCALVSLATAALLGHILLHDFLLVPREL SGSSPVLEETHPAHQQGASRPGPRDAQAHPGRPRAVPTQCDV PPNSRFDCAPDKAITQEQCEARGCCYIPAKQGLQGAQMGQP WCFFPPSYPSYKLENLSSSEMGYTATLTRTTPTFFPKDILTLRL DVMMETENRLHFTIKDPANRRYEVPLETPHVHSRAPSPLYSV EFSEEPFGVIVRRQLDGRVLLNTTVAPLFFADQFLQLSTSLPSQ YITGLAEHLSPLMLSTSWTRITLWNRDLAPTPGANLYGSHPFY LALEDGGSAHGVFLLNSNAMDVVLQPSPALSWRSTGGILDV YIFLGPEPKSVVQQYLDVVGYPFMPPYWGLGFHLCRWGYSS TAITRQVVENMTRAHFPLDVQWNDLDYMDSRRDFTFNKDGF RDFPAMVQELHQGGRRYMMIVDPAISSSGPAGSYRPYDEGLR RGVFITNETGQPLIGKVWPGSTAFPDFTNPTALAWWEDMVAE FHDQVPFDGMWIDMNEPSNFIRGSEDGCPNNELENPPYVPGV VGGTLQAATICASSHQFLSTHYNLHNLYGLTEAIASHRALVK ARGTRPFVISRSTFAGHGRYAGHWTGDVWSSWEQLASSVPEI LQFNLLGVPLVGADVCGFLGNTSEELCVRWTQLGAFYPFMR NHNSLLSLPQEPYSFSEPAQQAMRKALTLRYALLPHLYTLFH QAHVAGETVARPLFLEFPKDSSTWTVDHQLLWGEALLITPVL QAGKAEVTGYFPLGTWYDLQTVPVEALGSLPPPPAAPREPAI HSEGQWVTLPAPLDTINVHLRAGYIIPLQGPGLTTTESRQQPM ALAVALTKGGEARGELFWDDGESLEVLERGAYTQVIFLARN NTIVNELVRVTSEGAGLQLQKVTVLGVATAPQQVLSNGVPVS NFTYSPDTKVLDICVSLLMGEQFLVSWC	GAA
269	MAEWLLSASWQRRAKAMTAAAGSAGRAAVPLLLCALLAPG GAYVLDDSDGLGREFDGIGAVSGGGATSRLLVNYPEPYRSQI LDYLFKPNFGASLHILKVEIGGDGQTTDGTEPSHMHYALDEN YFRGYEWWLMKEAKKRNPNITLIGLPWSFPGWLGKGFDWPY VNLQLTAYYVVTWIVGAKRYHDLDIDYIGIWNERSYNANYIK ILRKMLNYQGLQRVKIIASDNLWESISASMLLDAELFKVVDVI GAHYPGTHSAKDAKLTGKKLWSSEDFSTLNSDMGAGCWGRI LNQNYINGYMTSTIAWNLVASYYEQLPYGRCGLMTAQEPWS GHYVVESPVWVSAHTTQFTQPGWYYLKTVGHLEKGGSYVA LTDGLGNLTIIIETMSHKHSKCIRPFLPYFNVSQQFATFVLKGS FSEIPELQVWYTKLGKTSERFLFKQLDSLWLLDSDGSFTLSLH EDELFTLTTLTTGRKGSYPLPPKSQPFPSTYKDDFNVDYPFFSE APNFADQTGVFEYFTNIEDPGEHHFTLRQVLNQRPITWAADA SNTISIIGDYNWTNLTIKCDVYIETPDTGGVFIAGRVNKGGILIR SARGIFFWIFANGSYRVTGDLAGWIIYALGRVEVTAKKWYTL TLTIKGHFTSGMLNDKSLWTDIPVNFPKNGWAAIGTHSFEFA QFDNFLVEATR	GALC
270	MAAVVAATRWWQLLLVLSAAGMGASGAPQPPNILLLLMDD	GALNS

		MGWGDLGVYGEPSRETPNLDRMAAEGLLFPNFYSANPLCSPS RAALLTGRLPIRNGFYTTNAHARNAYTPQEIVGGIPDSEQLLP ELLKKAGYVSKIVGKWHLGHRPQFHPLKHGFDEWFGSPNCH FGPYDNKARPNIPVYRDWEMVGRYYEEFPINLKTGEANLTQI YLQEALDFIKRQARHHPFFLYWAVDATHAPVYASKPFLGTSQ RGRYGDAVREIDDSIGKILELLQDLHVADNTFVFFTSDNGAAL ISAPEQGGSNGPFLCGKQTTFEGGMREPALAWWPGHVTAGQ VSHQLGSIMDLFTTSLALAGLTPPSDRAIDGLNLLPTLLQGRL MDRPIFYYRGDTLMAATLGQHKAHFWTWTNSWENFRQGIDF CPGQNVSGVTTHNLEDHTKLPLIFHLGRDPGERFPLSFASAEY QEALSRITSVVQQHQEALVPAQPQLNVCNWAVMNWAPPGCE KLGKCLTPPESIPKKCLWSH	
[0881]	271	MQLRNPELHLGCALALRFLALVSWDIPGARALDNGLARTPT MGWLHWERFMCNLDCQEEPDSCISEKLFMEMAELMVSEGW KDAGYEYLCIDDCWMAPQRDSEGRLQADPQRFPHGIRQLAN YVHSKGLKLGIYADVGNKTCAGFPGSFGYYDIDAQTFADWG VDLLKFDGCYCDSLENLADGYKHMSLALNRTGRSIVYSCEW PLYMWPFQKPNYTEIRQYCNHWRNFADIDDSWKSIKSILDWT SFNQERIVDVAGPGGWNDPDMLVIGNFGLSWNQQVTQMAL WAIMAAPLFMSNDLRHISPQAKALLQDKDVIAINQDPLGKQG YQLRQGDNFEVWERPLSGLAWAVAMINRQEIG GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRL RSHINPTGTVLLQLENTMQMSLKDLL	GLA
	272	MPGFLVRILPLLLVLLLLGPTRGLRNATQRMFEIDYSRDSFLK DGQPFRYISGSIHYSRVPRFYWKDRLLKMKMAGLNAIQTYVP WNFHEPWPGQYQFSEDHDVEYFLRLAHELGLLVILRPGPYIC AEWEMGGLPAWLLEKESILLRSSDPDYLAAVDKWLGVLLPK MKPLLYQNGGPVITVQVENEYGSYFACDFDYLRFLQKRFRH HLGDDVVLFTTDGAHKTFLKCGALQGLYTTVDFGTGSNITDA FLSQRKCEPKGPLINSEFYTGWLDHWGQPHSTIKTEAVASSLY DILARG ASVNLYMFIGGTNFAYWNGANSPYAAQPTSYDYDAPLSEAG DLTEKYFALRNIIQKFEKVPEGPIPPSTPKFAYGKVTLEKLKTV GAALDILCPSGPIKSLYPLTFIQVKQHYGFVLYRTTLPQDCSNP APLSSPLNGVHDRAYVAVDGIPQGVLERNNVITLNITGKAGA TLDLLVENMGRVNYGAYINDFKGLVSNLTLSSNILTDWTIFPL DTEDAVRSHLGGWGHRDSGHHDEAWAHNSSNYTLPAFYMG NFSIPSGIPDLPQDTFIQFPGWTKGQVWINGFNLGRYWPARGP QLTLFVPQHILMTSAPNTITVLELEWAPCSSDDPELCAVTFVD RPVIGSSVTYDHPSKPVEKRLMPPPPQKNKDSWLDHV	GLB1
	273	MQSLMQAPLLIALGLLLAAPAQAHLKKPSQLSSFSWDNCDEG KDPAVIRSLTLEPDPIIVPGNVTLSVMGSTSVPLSSPLKVDLVL EKEVAGLWIKIPCTDYIGSCTFEHFCDVLDMLIPTGEPCPEPLR TYGLPCHCPFKEGTYSLPKSEFVVPDLELPSWLTTGNYRIESV LSSSGKRLGCIKIAASLKGI	GM2A
	274	MLFKLLQRQTYTCLSHRYGLYVCFLGVVVTIVSAFQFGEVVL EWSRDQYHVLFDSYRDNIAGKSFQNRLCLPMPIDVVYTWVN GTDLELLKELQQVREQMEEEQKAMREILGKNTTEPTKKSEKQ LECLLTHCIKVPMLVLDPALPANITLKDLPSLYPSFHSASDIFN VAKPKNPSTNVSVVVFDSTKDVEDAHSGLLKGNSRQTVWRG YLTTDKEVPGLVLMQDLAFLSGFPPTFKETNQLKTKLPENLSS KVKLLQLYSEASVALLKLNNPKDFQELNKQTKKNMTIDGKE LTISPA YLLWDLSAISQSKQDEDISASRFEDNEELRYSLRSIERHAPWV	GNPTAB

[0882]

	RNIFIVTNGQIPSWLNLDNPRVTIVTHQDVFRNLSHLPTFSSPAI ESHIHRIEGLSQKFIYLNDDVMFGKDVWPDDFYSHSKGQKVY LTWPVPNCAEGCPGSWIKDGYCDKACNNSACDWDGGDCSG NSGGSRYIAGGGGTGSIGVGQPWQFGGGINSVSYCNQGCANS WLADKFCDQACNVLSCGFDAGDCGQDHFHELYKVILLPNQT HYIIPKGECLPYFSFAEVAKRGVEGAYSDNPIIRHASIANKWK TIHLIMHSGMNATTIHFNLTFQNTNDEEFKMQITVEVDTREGP KLNSTAQKGYENLVSPITLLPEAEILFEDIPKEKRFPKFKRHDV NSTRRAQEEVKIPLVNISLLPKDAQLSLNTLDLQLEHGDITLK GYNLSKSALLRSFLMNSQHAKIKNQAIITDETNDSLVAPQEKQ VHKSILPNSLGVSERLQRLTFPAVSVKVNGHDQGQNPPLDLE TTARFRVETHTQKTIGGNVTKEKPPSLIVPLESQMTKEKKITG KEKENSRMEENAENHIGVTEVLLGRKLQHYTDSYLGFLPWE KKKYFQDLLDEEESLKTQLAYFTDSKNTGRQLKDTFADSLRY VNKILNSKFGFTSRKVPAHMPHMIDRIVMQELQDMFPEEFDK TSFHKVRHSEDMQFAFSYFYYLMSAVQPLNISQVFDEVDTDQ SGVLSDREIRTLATRIHELPLSLQDLTGLEHMLINCSKMLPADI TQLNNIPPTQESYYDPNLPPVTKSLVTNCKPVTDKIHKAYKDK NKYRFEIMGEEEIAFKMIRTNVSHVVGQLDDIRKNPRKFVCL NDNIDHNHKDAQTVKAVLRDFYESMFPIPSQFELPREYRNRF LHMHELQEWRAYRDKLKFWTHCVLATLIMFTIFSFFAEQLIA LKRKIFPRRRIHKEASPNRIRV	
275	MAAGLARLLLLLGLSAGGPAPAGAAKMKVVEEPNAFGVNNP FLPQASRLQAKRDPSPVSGPVHLFRLSGKCFSLVESTYKYEFC PFHNVTQHEQTFRWNAYSGILGIWHEWEIANNTFTGMWMRD GDACRSRSRQSKVELACGKSNRLAHVSEPSTCVYALTFETPL VCHPHALLVYPTLPEALQRQWDQVEQDLADELITPQGHEKLL RTLFEDAGYLKTPEENEPTQLEGGPDSLGFETLENCRKAHKEL SKEIKRLKGLLTQHGIPYTRPTETSNLEHLGHETPRAKSPEQLR GDPG LRGSL	GNPTG
276	MRLLPLAPGRLRRGSPRHLPSCSPALLLLVLGGCLGVFGVAA GTRRPNVVLLLTDDQDEVLGGMTPLKKTKALIGEMGMTFSS AYVPSALCCPSRASILTGKYPHNHHVVNNTLEGNCSSKSWQK IQEPNTFPAILRSMCGYQTFFAGKYLNEYGAPDAGGLEHVPL GWSYWYALEKNSKYYNYTLSINGKARKHGENYSVDYLTDV LANVSLDFLDYKSNFEPFFMMIATPAPHSPWTAAPQYQKAFQ NVFAPRNKNFNIHGTNKHWLIRQAKTPMTNSSIQFLDNAFRK RWQTLLSVD DLVEKLVKRLEFTGELNNTYIFYTSDNGYHTGQFSLPIDKRQL YEFDIKVPLLVRGPGIKPNQTSKMLVANIDLGPTILDIAGYDL NKTQMDGMSLLPILRGASNLTWRSDVLVEYQGEGRNVTDPT CPSLSPGVSQCFPDCVCEDAYNNTYACVRTMSALWNLQYCE FDDQEVFVEVYNLTADPDQITNIAKTIDPELLGKMNYRLMML QSCSGPTCRTPGVFDPGYRFDPRLMFSNRGSVRTRRFSKHLL	GNS
277	MWTLVSWVALTAGLVAGTRCPDGQFCPVACCLDPGGASYS CCRPLLDKWPTTLSRHLGGPCQVDAHCSAGHSCIFTVSGTSSC CPFPEAVACGDGHHCCPRGFHCSADGRSCFQRSGNNSVGAIQ CPDSQFECPDFSTCCVMVDGSWGCCPMPQASCCEDRVHCCP HGAFCDLVHTRCITPTGTHPLAKKLPAQRTNRAVALSSSVMC PDARSRCPDGSTCCELPSGKYGCCPMPNATCCSDHLHCCPQD TVCDLIQSKCLSKENATTDLLTKLPAHTVGDVKCDMEVSCPD GYTCCRLQSGAWGCCPFTQAVCCEDHIHCCPAGFTCDTQKG TCEQGPHQVPWMEKAPAHLSLPDPQALKRDVPCDNVSSCPSS	GRN

		DTCCQLTSGEWGCCPIPEAVCCSDHQHCCPQGYTCVAEGQC QRGSEIVAGLEKMPARRASLSHPRDIGCDQHTSCPVGQTCCPS LGGSWACCQLPHAVCCEDRQHCCPAGYTCNVKARSCEKEV VSAQPATFLARSPHVGVKDVECGEGHFCHDNQTCCRDNRQG WACCPYRQGVCCADRRHCCPAGFRCAARGTKCLRREAPRW DAPLRDPALRQLL	
[0883]	278	MARGSAVAWAALGPLLWGCALGLQGGMLYPQESPSRECKE LDGLWSFRADFSDNRRRGFEEQWYRRPLWESGPTVDMPVPS SFNDISQDWRLRHFVGWVWYEREVILPERWTQDLRTRVVLRI GSAHSYAIVWVNGVDTLEHEGGYLPFEADISNLVQVGPLPSR LRITIAINNTLTPTTLPPGTIQYLTDTSKYPKGYFVQNTYFDFF NYAGLQRSVLLYTTPTTYIDDITVTTSVEQDSGLVNYQISVKG SNLFKLEVRLLDAENKVVANGTGTQGQLKVPGVSLWWPYL MHERPAYL YSLEVQLTAQTSLGPVSDFYTLPVGIRTVAVTKSQFLINGKPF YFHGVNKHEDADIRGKGFDWPLLVKDFNLLRWLGANAFRTS HYPYAEEVMQMCDRYGIVVIDECPGVGLALPQFFNNVSLHH HMQVMEEVVRRDKNHPAVVMWSVANEPASHLESAGYYLK MVIAHTKSLDPSRPVTFVSNSNYAADKGAPYVDVICLNSYYS WYHDYGHLELIQLQLATQFENWYKKYQKPIIQSEYGAETIAG FHQDPPLMFTEEYQKSLLEQYHLGLDQKRRKYVVGELIWNF ADFMTEQSPTRVLGNKKGIFTRQRQPKSAAFLLRERYWKIAN ETRYPHSVAKSQCLENSLFT	GUSB
	279	MTSSRLWFSLLLAAAFAGRATALWPWPQNFQTSDQRYVLYP NNFQFQYDVSSAAQPGCSVLDEAFQRYRDLLFGSGSWPRPYL TGKRHTLEKNVLVVSVVTPGCNQLPTLESVENYTLTINDDQC LLLSETVWGALRGLETFSQLVWKSAEGTFFINKTEIEDFPRFP HRGLLLDTSRHYLPLSSILDTLDVMAYNKLNVFHWHLVDDPS FPYESFTFPELMRKGSYNPVTHIYTAQDVKEVIEYARLRGIRV LAEFDTPGHTLSWGPGIPGLLTPCYSGSEPSGTFGPVNPSLNNT YEFMSTFFLEVSSVFPDFYLHLGGDEVDFTCWKSNPEIQDFM RKKGFGEDFKQLESFYIQTLLDIVSSYGKGYVVWQEVFDNKV KIQPDTIIQVWREDIPVNYMKELELVTKAGFRALLSAPWYLN RISYGPDWKDFYIVEPLAFEGTPEQKALVIGGEACMWGEYVD NTNLVPRLWPRAGAVAERLWSNKLTSDLTFAYERLSHFRCEL LRRGVQAQPLNVGFCEQEFEQT	HEXA
	280	MELCGLGLPRPPMLLALLLATLLAAMLALLTQVALVVQVAE AARAPSVSAKPGPALWPLPLSVKMTPNLLHLAPENFYISHSPN STAGPSCTLLEEAFRRYHGYIFGFYKWHHEPAEFQAKTQVQQ LLVSITLQSECDAFPNISSDESYTLLVKEPVAVLKANRVWGAL RGLETFSQLVYQDSYGTFTINESTIIDSPRFSHRGILIDTSRHYL PVKIILKTLDAMAFNKFNVLHWHIVDDQSFPYQSITFPELSNK GSYSLSHVYTPNDVRMVIEYARLRGIRVLPEFDTPGHTLSWG KGQKDLLTPCYSRQNKLDSFGPINPTLNTTYSFLTTFFKEISEV FPDQFIHLGGDEVEFKCWESNPKIQDFMRQKGFGTDFKKLES FYIQKVLDIIATINKGSIVWQEVFDDKAKLAPGTIVEVWKDSA YPEELSRVTASGFPVILSAPWYLDLISYGQDWRKYYKVEPLD FGGTQKQKQLFIGGEACLWGEYVDATNLTPRLWPRASAVGE RLWSSKDVRDMDDAYDRLTRHRCRMVERG IAAQPLYAGYCNHENM	HEXB
	281	MTGARASAAEQRRAGRSGQARAAERAAGMSGAGRALAALL LAASVLSAALLAPGGSSGRDAQAAPPRDLDKKRHAELKMDQ ALLLIHNELLWTNLTVYWKSECCYHCLFQVLVNVPQSPKAG KPSAAAASVSTQHGSILQLNDTLEEKEVCRLEYRFGEFGNYSL	HGSNAT

		LVKNIHNGVSEIACDLAVNEDPVDSNLPVSIAFLIGLAVIIVISF LRLLLSLDDFNNWISKAISSRETDRLINSELGSPSRTDPLDGDV QPATWRLSALPPRLRSVDTFRGIALILMVFVNYGGGKYWYFK HASWNGLTVADLVFPWFVFIMGSSIFLSMTSILQRGCSKFRLL GKIAWRSFLLICIGIIIVNPNYCLGPLSWDKVRIPGVLQRLGVT YFVVAVLELLFAKPVPEHCASERSCLSLRDITSSWPQWLLILV LEGLWLGLTFLLPVPGCPTGYLGPGGIGDFGKYPNCTGGAAG YIDRLLLGDDHLYQHPSSAVLYHTEVAYDPEGILGTINSIVMA FLGVQAGKILLYYKARTKDILIRFTAWCC ILGLISVALTKVSENEGFIPVNKNLWSLSYVTTLSSFAFFILLVL YPVVDVKGLWTGTPFFYPGMNSILVYVGHEVFENYFPFQWK LKDNQSHKEHLTQNIVATALWVLIAYILYRKKIFWKI	
	282	MAAHLLPICALFLTLLDMAQGFRGPLLPNRPFTTVWNANTQ WCLERHGVDVDVSVFDVVANPGQTFRGPDMTIFYSSQLGTY PYYTPTGEPVFGGLPQNASLIAHLARTFQDILAAIPAPDFSGLA VIDWEAWRPRWAFNWDTKDIYRQRSRALVQAQHPDWPAPQ VEAVAQDQFQGAARAWMAGTLQLGRALRPRGLWGFYGFPD CYNYDFLSPNYTGQCPSGIRAQNDQLGWLWGQSRALYPSIY MPAVLEGTGKSQMYVQHRVAEAFRVAVAAGDPNLPVLPYV QIFYDTTNHFLPLDELEHSLGESAAQGAAGVVLWVSWENTRT KESCQAIKEYMDTTLGPFILNVTSGALLCSQ ALCSGHGRCVRRTSHPKALLLLNPASFSIQLTPGGGPLSLRGA LSLEDQAQMAVEFKCRCYPGWQAPWCERKSMW	HYAL1
[0884]	283	MPPPRTGRGLLWLGLVLSSVCVALGSETQANSTTDALNVLLII VDDLRPSLGCYGDKLVRSPNIDQLASHSLLFQNAFAQQAVCA PSRVSFLTGRRPDTTRLYDFNSYWRVHAGNFSTIPQYFKENG YVTMSVGKVFHPGISSNHTDDSPYSWSFPPYHPSSEKYENTKT CRGPDGELHANLLCPVDVLDVPEGTLPDKQSTEQAIQLLEKM KTSASPFFLAVGYHKPHIPFRYPKEFQKLYPLENITLAPDPEVP DGLPPVAYNPWMDIRQREDVQALNISVPYGPIPVDFQRKIRQS YFASVSYLDTQVGRLLSALDDLQLANSTIIAFTSDHGWALGE HGEWAKYSNFDVATHVPLIFYVPGRTASLPEAGEKLFPYLDP FDSASQLMEPGRQSMDLVELVSLFPTLAGLAGLQVPPRCPVP SFHVELCREGKNLLKHFRFRDLEEDPYLPGNPRELIAYSQYPR PSDIPQWNSDKPSLKDIKIMGYSIRTIDYRYTVWVGFNPDEFL ANFSDIHAGELYFVDSDPLQDHNMYNDSQGGDLFQLLMP	IDS
	284	MRPLRPRAALLALLASLLAAPPVAPAEAPHLVHVDAARALW PLRRFWRSTGFCPPLPHSQADQYVLSWDQQLNLAYVGAVPH RGIKQVRTHWLLELVTTRGSTGRGLSYNFTHLDGYLDLLREN QLLPGFELMGSASGHFTDFEDKQQVFEWKDLVSSLARRYIGR YGLAHVSKWNFETWNEPDHHDFDNVSMTMQGFLNYYDACS EGLRAASPALRLGGPGDSFHTPPRSPLSWGLLRHCHDGTNFFT GEAGVRLDYISLHRKGARSSISILEQEKVVAQQIRQLFPKFAD TPIYNDEADPLVGWSLPQPWRADVTYAAMVVKVIAQHQNLL LANTTSAFPYALLSNDNAFLSYHPHPFAQRTLTARFQVNNTR PPHVQLLRKPVLTAMGLLALLDEEQLWAEVSQAGTVLDSNH TVGVLASAHRPQGPADAWRAAVLIYASDDTRAHPNRSVAVT LRLRGVPPGPGLVYVTRYLDNGLCSPDGEWRRLGRPVFPTAE QFRRMRAAEDPVAAAPRPLPAGGRLTLRPALRLPSLLLVHVC ARPEKPPGQVTRLRALPLTQGQLVLVWSDEHVGSKCLWTYEI QFSQDGKAYTPVSRKPSTFNLFVFSPDTGAVSGSYRVRALDY WARPGPFSDPVPYLEVPVPRGPPSPGNP	IDUA
	285	MVVVTGREPDSRRQDGAMSSSDAEDDFLEPATPTATQAGHA LPLLPQEFPEVVPLNIGGAHFTTRLSTLRCYEDTMLAAMFSGR	KCTD7

[0885]

	HYIPTDSEGRYFIDRDGTHFGDVLNFLRSGDLPPRERVRAVYK EAQYYAIGPLLEQLENMQPLKGEKVRQAFLGLMPYYKDHLE RIVEIARLRAVQRKARFAKLKVCVFKEEMPITPYECPLLNSLR FERSESDGQLFEHHCEVDVSFGPWEAVADVYDLLHCLVTDLS AQGLTVDHQCIGVCDKHLVNHYYCKRPIYEFKITWW	
286	MVCFRLFPVPGSGLVLVCLVLGAVRSYALELNLTDSENATCL YAKWQMNFTVRYETTNKTYKTVTISDHGTVTYNGSICGDDQ NGPKIAVQFGPGFSWIANFTKAASTYSIDSVSFSYNTGDNTTF PDAEDKGILTVDELLAIRIPLNDLFRCNSLSTLEKNDVVQHYW DVLVQAFVQNGTVSTNEFLCDKDKTSTVAPTIHTTVPSPTTTP TPKEKPEAGTYSVNNGNDTCLLATMGLQLNITQDKVASVINI NPNTTHSTGSCRSHTALLRLNSSTIKYLDFVFAVKNENRFYLK EVNISMYLVNGSVFSIANNNLSYWDAPLGSSYMCNKEQTVSV SGAFQINTFDLRVQPFNVTQGKYSTAQDCSADDDNFLVPIAV GAALAGVLILVLLAYFIGLKHHHAGYEQF	LAMP2
287	MGAYARASGVCARGCLDSAGPWTMSRALRPPLPPLCFFLLLL AAAGARAGGYETCPTVQPNMLNVHLLPHTHDDVGWLKTVD QYFYGIKNDIQHAGVQYILDSVISALLADPTRRFIYVEIAFFSR WWHQQTNATQEVVRDLVRQGRLEFANGGWVMNDEAATHY GAIVDQMTLGLRFLEDTFGNDGRPRVAWHIDPFGHSREQASL FAQMGFDGFFFGRLDYQDKWVRMQKLEMEQVWRASTSLKP PTADLFTGVLPNGYNPPRNLCWDVLCVDQPLVEDPRSPEYNA KELVDYFLNVATAQGRYYRTNHTVMTMGSDFQYENANMW FKNLDKLIRLVNAQQAKGSSVHVLYSTPACYLWELNKANLT WSVKHDDFFPYADGPHQFWTGYFSSRPALKRYERLSYNFLQ VCNQLEALVGLAANVGPYGSGDSAPLNEAMAVLQHHDAVS GTSRQHVANDYARQLAAGWGPCEVLLSNALARLRGFKDHFT FCQQLNISICPLSQTAARFQVIVYNPLGRKVNWMVRLPVSEG VFVVKDPNGRTVPSDVVIFPSSDSQAHPPELLFSASLPALGFST YSVAQVPRWKPQARAPQPIPRRSWSPALTIENEHIRATFDPDT GLLMEIMNMNQQLLLPVRQTFFWYNASIGDNESDQASGAYIF RPNQQKPLPVSRWAQIHLVKTPLVQEVHQNFSAWCSQVVRL YPGQRHLELEWSVGPIPVGDTWGKEVISRFDTPLETKGRFYT DSNGREILERRRDYRPTWKLNQTEPVAGNYYPVNTRIYITDG NMQLTVLTDRSQGGSSLRDGSLELMVHRRLLKDDGRGVSEP LMENGSGAWVRGRHLVLLDTAQAAAAGHRLLAEQEVLAPQ VVLAPGGGAAYNLGAPPRTQFSGLRRDLPPSVHLLTLASWGP EMVLLRLEHQFAVGEDSGRNLSAPVTLNLRDLFSTFTITRLQE TTLVANQLREAASRLKWTTNTGPTPHQTPYQLDPANITLEPM EIRTFLASVQWKEVDG	MAN2B1
288	MRLHLLLLLALCGAGTTAAELSYSLRGNWSICNGNGSLELPG AVPGCVHSALFQQGLIQDSYYRFNDLNYRWVSLDNWTYSKE FKIPFEISKWQKVNLILEGVDTVSKILFNEVTIGETDNMFNRYS FDITNVVRDVNSIELRFQSAVLYAAQQSKAHTRYQVPPDCPPL VQKGECHVNFVRKEQCSFSWDWGPSFPTQGIWKDVRIEAYNI CHLNYFTFSPIYDKSAQEWNLEIESTFDVVSSKPVGGQVIVAIP KLQTQQTYSIELQPGKRIVELFVNISKNITVETWWPHGHGNQT GYNMTVLFELDGGLNIEKSAKVYFRTVELIEEPIKGSPGLSFYF KINGFPIFLKGSNWIPADSFQDRVTSELLRLLLQSVVDANMNT LRVWGGGIYEQDEFYELCDELGIMVWQDFMFACALYPTDQG FLDSVTAEVAYQIKRLKSHPSIIIWSGNNENEEALMMNWYHIS FTDRPIYIKDYVTLYVKNIRELVLAGDKSRPFITSSPTNGAETV AEAWVSQNPNSNYFGDVHFYDYISDC WNWKVFPKARFASEYGYQSWPSFSTLEKVSSTEDWSFNSKFS	MANBA

[0886]

	LHRQHHEGGNKQMLYQAGLHFKLPQSTDPLRTFKDTIYLTQ VMQAQCVKTETEFYRRSRSEIVDQQGHTMGALYWQLNDIW QAPSWASLEYGGKWKMLHYFAQNFFAPLLPVGFENENTFYI YGVSDLHSDYSMTLSVRVHTWSSLEPVCSRVTERFVMKGGE AVCLYEEPVSELLRRCGNCTRESCVVSFYLSADHELLSPTNYH FLSSPKEAVGLCKAQITAIISQQGDIFVFDLETSAVAPFVWLDV GSIPGRFSDNGFLMTEKTRTILFYPWEPTSKNELEQSFHVTSLT DIY	
289	MTAPAGPRGSETERLLTPNPGYGTQAGPSPAPPTPPEEEDLRR RLKYFFMSPCDKFRAKGRKPCKLMLQVVKILVVTVQLILFGL SNQLAVTFREENTIAFRHLFLLGYSDGADDTFAAYTREQLYQ AIFHAVDQYLALPDVSLGRYAYVRGGGDPWTNGSGLALCQR YYHRGHVDPANDTFDIDPMVVTDCIQVDPPERPPPPSDDLTL LESSSSYKNLTLKFHKLVNVTIHFRLKTINLQSLINNEIPDCYTF SVLITFDNKAHSGRIPISLETQAHIQECKHPSVFQHGDNSFRLL FDVVVILTCSLSFLLCARSLLRGFLLQNEFVGFMWRQRGRVIS LWERLEFVNGWYILLVTSDVLTISGTIMKIGIEAKNLASYDVC SILLGTSTLLVWVGVIRYLTFFHNYNILIATLRVALPSVMRFCC CVAVIYLGYCFCGWIVLGPYHVKFRSLSMVSECLFSLINGDD MFVTFAAMQAQQGRSSLVWLFSQLYLYSFISLFIYMVLSLFIA LITGAYDTIKHPGGAGAEESELQAYIAQCQDSPTSGKFRRGSG SACSLLCCCGRDPSEEHSLLVN	MCOLN1
290	MAGLRNESEQEPLLGDTPGSREWDILETEEHYKSRWRSIRILY LTMFLSSVGFSVVMMSIWPYLQKIDPTADTSFLGWVIASYSL GQMVASPIFGLWSNYRPRKEPLIVSILISVAANCLYAYLHIPAS HNKYYMLVARGLLGIGAGNVAVVRSYTAGATSLQERTSSMA NISMCQALGFILGPVFQTCFTFLGEKGVTWDVIKLQINMYTTP VLLSAFLGILNIILILAILREHRVDDS GRQCKSINFEEASTDEAQVPQGNIDQVAVVAINVLFFVTLFIF ALFETIITPLTMDMYAWTQEQAVLYNGIILAALGVEAVVIFLG VKLLSKKIGERAILLGGLIVVWVGFFILLPWGNQFPKIQWEDL HNNSIPNTTFGEIIIGLWKSPMEDDNERPTGCSIEQAWCLYTPV IHLAQFLTSAVLIGLGYPVCNLMSYTLYSKILGPKPQGVYMG WLTASGSGARILGPMFISQVYAHWGPRWAFSLVCGIIVLTITL LGVVYKRLIALSVRYGRIQE	MFSD8
291	MLLKTVLLLGHVAQVLMLDNGLLQTPPMGWLAWERFRCNI NCDEDPKNCISEQLFMEMADRMAQDGWRDMGYTYLNIDDC WIGGRDASGRLMPDPKRFPHGIPFLADYVHSLGLKLGIYADM GNFTCMGYPGTTLDKVVQDAQTFAEWKVDMLKLDGCFSTP EERAQGYPKMAAALNATGRPIAFSCSWPAYEGGLPPRVNYSL LADICNLWRNYDDIQDSWWSVLSILNWFVEHQDILQPVAGPG HWNDPDMLLIGNFGLSLEQSRAQMALWTVLAAPLLMSTDLR TISAQNMDILQNPLMIKINQDPLGIQGRRIHKEKSLIEVYMRPL SNKASALVFFSCRTDMPYRYHSSLGQLNFTGSVIYEAQDVYS GDIISGLRDETNFTVIINPSGVVMWYLYPIKNLEMSQQ	NAGA
292	MEAVAVAAAVGVLLLAGAGGAAGDEAREAAAVRALVARLL GPGPAADFSVSVERALAAKPGLDTYSLGGGGAARVRVRGST GVAAAAGLHRYLRDFCGCHVAWSGSQLRLPRPLPAVPGELT EATPNRYRYYQNVCTQSYSFVWWDWARWEREIDWMALNGI NLALAWSGQEAIWQRVYLALGLTQAEINEFFTGPAFLAWGR MGNLHTWDGPLPPSWHIKQLYLQHRVLDQMRSFGMTPVLPA FAGHVPEAVTRVFPQVNVTKMGSWGHFNCSYSCSFLLAPEDP IFPIIGSLFLRELIKEFGTDHIYGADTFNEMQPPSSEPSYLAAAT TAVYEAMTAVDTEAVWLLQGWLFQHQPQF	NAGLU

[0887]

	WGPAQIRAVLGAVPRGRLLVLDLFAESQPVYTRTASFQGQPFI WCMLHNFGGNHGLFGALEAVNGGPEAARLFPNSTMVGTGM APEGISQNEVVYSLMAELGWRKDPVPDLAAWVTSFAARRYG VSHPDAGAAWRLLLRSVYNCSGEACRGHNRSPLVRRPSLQM NTSIWYNRSDVFEAWRLLLTSAPSLATSPAFRYDLLDLTRQA VQELVSLYYEEARSAYLSKELASLLRAGGVLAYELLPALDEV LASDSRFLLGSWLEQARAAAVSEAEADFYEQNSRYQLTLWG PEGNILDYANKQLAGLVANYYTPRWRLFLEALVDSVAQGIPF QQHQFDKNVFQLEQAFVLSKQRYPSQPRGDTVDLAKKIFLKY YPRWVAGSW	
293	MTGERPSTALPDRRWGPRILGFWGGCRVWVFAAIFLLLSLAA SWSKAENDFGLVQPLVTMEQLLWVSGRQIGSVDTFRIPLITAT PRGTLLAFAEARKMSSSDEGAKFIALRRSMDQGSTWSPTAFIV NDGDVPDGLNLGAVVSDVETGVVFLFYSLCAHKAGCQVAST MLVWSKDDGVSWSTPRNLSLDIGTEVFAPGPGSGIQKQREPR KGRLIVCGHGTLERDGVFCLLSDDHGASWRYGSGVSGIPYGQ PKQENDFNPDECQPYELPDGSVVINARNQNNYHCHCRIVLRS YDACDTLRPRDVTFDPELVDPVVAAGAVVTSSGIVFFSNPAH PEFRVNLTLRWSFSNGTSWRKET VQLWPGPSGYSSLATLEGSMDGEEQAPQLYVLYEKGRNHYT ESISVAKISVYGTL	NEU1
294	MTARGLALGLLLLLLCPAQVFSQSCVWYGECGIAYGDKRYN CEYSGPPKPLPKDGYDLVQELCPGFFFGNVSLCCDVRQLQTL KDNLQLPLQFLSRCPSCFYNLLNLFCELTCSPRQSQFLNVTAT EDYVDPVTNQTKTNVKELQYYVGQSFANAMYNACRDVEAP SSNDKALGLLCGKDADACNATNWIEYMFNKDNGQAPFTITP VFSDFPVHGMEPMNNATKGCDESVDEVTAPCSCQDCSIVCGP KPQPPPPPAPWTILGLDAMYVIMWITYMAFLLVFFGAFFAVW CYRKRYFVSEYTPIDSNIAFSVNASDKGEASCCDPVSAAFEGC LRRLFTRWGSFCVRNPGCVIFFSLVFITACSSGLVFVRVTTNPV DLWSAPSSQARLEKEYFDQHFGPFFRTEQLIIRAPLTDKHIYQP YPSGADVPFGPPLDIQILHQVLDLQIAIENITASYDNETVTLQDI CLAPLSPYNTNCTILSVLNYFQNSHSVLDHKKGDDFFVYADY HTHFLYCVRAPASLNDTSLLHDPCLGTFGGPVFPWLVLGGYD DQNYNNATALVITFPVNNYYNDTEKLQRAQAWEKEFINFVK NYKNPNLTISFTAERSIEDELNRESDSDVFTVVISYAIMFLYISL ALGHMKSCRRLLVDSKVSLGIAGILIVLSSVACSLGVFSYIGLP LTLIVIEVIPFLVLAVGVDNIFILVQAYQRDERLQGETLDQQLG RVLGEVAPSMFLSSFSETVAFFLGALSVMPAVHTFSLFAGLAV FIDFLLQITCFV SLLGLDIKRQEKNRLDIFCCVRGAEDGTSVQASESCLFRFFKN SYSPLLLKDWMRPIVIAIFVGVLSFSIAVLNKVDIGLDQSLSMP DDSYMVDYFKSISQYLHAGPPVYFVLEEGHDYTSSKGQNMV CGGMGCNNDSLVQQIFNAAQLDNYTRIGFAPSSWIDDYFDW VKPQSSCCRVDNITDQFCNASVVDPACVRCRPLTPEGKQRPQ GGDFMRFLPMFLSDNPNPKCGKGGHAAYSSAVNILLGHGTR VGATYFMTYHTVLQTSADFIDALKKARLIASNVTETMGINGS AYRVFPYSVFYVFYEQYLTIIDDTIFNLGVSLGAIFLVTMVLLG CELWSAVIMCATIAMVLVNMFGVMWLWGISLNAVSLVNLV MSCGISVEFCSHITRAFTVSMKGSRVERAEEALAHMGSSVFSG ITLTKFGGIVVLAFAKSQIFQIFYFRMYLAMVLLGATHGLIFLP VLLSYIGPSVNKAKSCATEERYKGTERERLLNF	NPC1
295	MRFLAATFLLLALSTAAQAEPVQFKDCGSVDGVIKEVNVSPC PTQPCQLSKGQSYSVNVTFTSNIQSKSSKAVVHGILMGVPVPF	NPC2

		PIPEPDGCKSGINCPIQKDKTYSYLNKLPVKSEYPSIKLVVEWQ LQDDKNQSLFCWEIPVQIVSHL	
	296	MSCPVPACCALLLVLGLCRARPRNALLLLADDGGFESGAYN NSAIATPHLDALARRSLLFRNAFTSVSSCSPSRASLLTGLPQHQ NGMYGLHQDVHHFNSFDKVRSLPLLLSQAGVRTGIIGKKHV GPETVYPFDFAYTEENGSVLQVGRNITRIKLLVRKFLQTQDDR PFFLYVAFHDPHRCGHSQPQYGTFCEKFGNGESGMGRIPDWT PQAYDPLDVLVPYFVPNTPAARADLAAQYTTVGRMDQGVGL VLQELRDAGVLNDTLVIFTSDNGIPFPSGRTNLYWPGTAEPLL VSSPE HPKRWGQVSEAYVSLLDLTPTILDWFSIPYPSYAIFGSKTIHLT GRSLLPALEAEPLWATVFGSQSHHEVTMSYPMRSVQHRHFRL VHNLNFKMPFPIDQDFYVSPTFQDLLNRTTAGQPTGWYKDLR HYYYRARWELYDRSRDPHETQNLATDPRFAQLLEMLRDQLA KWQWETHDPWVCAPDGVLEEKLSPQCQPLHNEL	SGSH
	297	MASPGCLWLLAVALLPWTCASRALQHLDPPAPLPLVIWHGM GDSCCNPLSMGAIKKMVEKKIPGIYVLSLEIGKTLMEDVENSF FLNVNSQVTTVCQALAKDPKLQQGYNAMGFSQGGQFLRAV AQRCPSPPMINLISVGGQHQGVFGLPRCPGESSHICDFIRKTLN AGAYSKVVQERLVQAEYWHDPIKEDVYRNHSIFLADINQERG INESYKKNLMALKKFVMVKFLNDSIVDPVDSEWFGFYRSGQ AKETIPLQETSLYTQDRLGLKEMDNAGQLVFLATEGDHLQLS EEWFYAHIIPFLG	PPT1
[0888]	298	MYALFLLASLLGAALAGPVLGLKECTRGSAVWCQNVKTASD CGAVKHCLQTVWNKPTVKSLPCDICKDVVTAAGDMLKDNA TEEEILVYLEKTCDWLPKPNMSASCKEIVDSYLPVILDIIKGEM SRPGEVCSALNLCESLQKHLAELNHQKQLESNKIPELDMTEV VAPFMANIPLLLYPQDGPRSKPQPKDNGDVCQDCIQMVTDIQ TAVRTNSTFVQALVEHVKEECDRLGPGMADICKNYISQYSEI AIQMMMHMQPKEICALVGFCDEVKEMPMQTLVPAKVASKN VIPALELVEPIKKHEVPAKSDVYCEVCEFLVKEVTKLIDNNKT EKEILDAFDKMCSKLPKSLSEECQEVVDTYGSSILSILLEEVSP ELVCSMLHLCSGTRLPALTVHVTQPKDGGFCEVCKKLVGYL DRNLEKNSTKQEILAALEKGCSFLPDPYQKQCDQFVAEYEPV LIEILVEVMDPSFVCLKIGACPSAHKPLLGTEKCIWGPSYWCQ NTETAAQCNAVEHCKRHVWN	PSAP
	299	MRSPVRDLARNDGEESTDRTPLLPGAPRAEAAPVCCSARYNL AILAFFGFFIVYALRVNLSVALVDMVDSNTTLEDNRTSKACPE HSAPIKVHHNQTGKKYQWDAETQGWILGSFFYGYIITQIPGG YVASKIGGKMLLGFGILGTAVLTLFTPIAADLGVGPLIVLRAL EGLGEGVTFPAMHAMWSSWAPPLERSKLLSISYAGAQLGTVI SLPLSGIICYYMNWTYVFYFFGTIGIFWFLLWIWLVSDTPQKH KRISHYEKEYILSSLRNQLSSQKSVPWVPILKSLPLWAIVVAHF SYNWTFYTLLTLLPTYMKEILRFNVQENGFLSSLPYLGSWLC MILSGQAADNLRAKWNFSTLCVRRIFSLIGMIGPAVFLVAAGF IGCDYSLAVAFLTISTTLGGFCSSGFSINHLDIAPSYAGILLGIT NTFATIPGMVGPVIAKSLTPDNTVGEWQTVFYIAAAINVFGAI FFTLFAKGEVQNWALNDHHGHRH	SLC17A5
	300	MPRYGASLRQSCPRSGREQGQDGTAGAPGLLWMGLVLALAL ALALALALSDSRVLWAPAEAHPLSPQGHPARLHRIVPRLRDV FGWGNLTCPICKGLFTAINLGLKKEPNVARVGSVAIKLCNLL KIAPPAVCQSIVHLFEDDMVEVWRRSVLSPSEACGLLLGSTCG HWDIFSSWNISLPTVPKPPPKPPSPPAPGAPVSRILFLTDLHWD	SMPD1

		HDYLEGTDPDCADPLCCRRGSGLPPASRPGAGYWGEYSKCD LPLRTLESLLSGLGPAGPFDMVYWTGDIPAHDVWHQTRQDQ LRALTTVTALVRKFLGPVPVYPAVGNHESTPVNSFPPPFIEGN HSSRWLYEAMAKAWEPWLPAEALRTLRIGGFYALSPYPGLR LISLNMNFCSRENFWLLINSTDPAGQLQWLVGELQAAEDRGD KVHIIGHIPPGHCLKSWSWNYYRIVARYENTLAAQFFGHTHV DEFEVFYDEETLSRPLAVAFLAPSATTYIGLNPGYRVYQIDGN YSGSSHVVLDHETYILNLTQANIPGAIPHWQLLYRARETYGLP NTLPTAWHNLVYRMRGDMQLFQTFWFLYHKGHPPSEPCGTP CRLATLCAQLSARADSPALCRHLMPDGSLPEAQSLWPRPLFC	
	301	MAAPALGLVCGRCPELGLVLLLLLLSLLCGAAGSQEAGTGA GAGSLAGSCGCGTPQRPGAHGSSAAAHRYSREANAPGPVPG ERQLAHSKMVPIPAGVFTMGTDDPQIKQDGEAPARRVTIDAF YMDAYEVSNTEFEKFVNSTGYLTEAEKFGDSFVFEGMLSEQV KTNIQQAVAAAPWWLPVKGANWRHPEGPDSTILHRPDHPVL HVSWNDAVAYCTWAGKRLPTEAEWEYSCRGGLHNRLFPWG NKLQPKGQHYANIWQGEFPVTNTGEDGFQGTAPVDAFPPNG YGLYNIVGNAWEWTSDWWTVHHSVEETLNPKGPPSGKDRV KKGGSYMCHRSYCYRYRCAARSQNTPDSSASN LGFRCAADRLPTMD	SUMF1
[0889]	302	MGLQACLLGLFALILSGKCSYSPEPDQRRTLPPGWVSLGRAD PEEELSLTFALRQQNVERLSELVQAVSDPSSPQYGKYLTLENV ADLVRPSPLTLHTVQKWLLAAGAQKCHSVITQDFLTCWLSIR QAELLLPGAEFHHYVGGPTETHVVRSPHPYQLPQALAPHVDF VGGLHRFPPTSSLRQRPEPQVTGTVGLHLGVTPSVIRKRYNLT SQDVGSGTSNNSQACAQFLEQYFHDSDLAQFMRLFGGNFAH QASVARVVGQQGRGRAGIEASLDVQYLMSAGANISTWVYSS PGRHEG QEPFLQWLMLLSNESALPHVHTVSYGDDEDSLSSAYIQRVNT ELMKAAARGLTLLFASGDSGAGCWSVSGRHQFRPTFPASSPY VTTVGGTSFQEPFLITNEIVDYISGGGFSNVFPRPSYQEEAVTK FLSSSPHLPPSSYFNASGRAYPDVAALSDGYWVVSNRVPIPW VSGTSASTPVFGGILSLINEHRILSGRPPLGFLNPRLYQQHGAG LFDVTRGCHESCLDEEVEGQGFCSGPGWDPVTGWGTPNFPA LLKTLLNP	TPP1
	303	MSDKLPYKVADIGLAAWGRKALDIAENEMPGLMRMRERYS ASKPLKGARIAGCLHMTVETAVLIETLVTLGAEVQWSSCNIFS TQDHAAAAIAKAGIPVYAWKGETDEEYLWCIEQTLYFKDGP LNMILDDGGDLTNLIHTKYPQLLPGIRGISEETTTGVHNLYKM MANGILKVPAINVNDSVTKSKFDNLYGCRESLIDGIKRATDV MIAGKVAVVAGYGDVGKGCAQALRGFGARVIITEIDPINALQ AAMEGYEVTTMDEACQEGNIFVTTTGCIDIILGRHFEQMKDD AIVCNIG HFDVEIDVKWLNENAVEKVNIKPQVDRYRLKNGRRIILLAEG RLVNLGCAMGHPSFVMSNSFTNQVMAQIELWTHPDKYPVGV HFLPKKLDEAVAEAHLGKLNVKLTKLTEKQAQYLGMSCDGP FKPDHYRY	AHCY
	304	MVDSVYRTRSLGVAAEGLPDQYADGEAARVWQLYIGDTRSR TAEYKAWLLGLLRQHGCQRVLDVACGTGVDSIMLVEEGFSV TSVDASDKMLKYALKERWNRRHEPAFDKWVIEEANWMTLD KDVPQSAEGGFDAVICLGNSFAHLPDCKGDQSEHRLALKNIA SMVRAGGLLVIDHRNYDHILSTGCAPPGKNIYYKSDLTKDVT TSVLIVNNKAHMVTLDYTVQVPGAGQDGSPGLSKFRLSYYP HCLASFTELLQAAFGGKCQHSVLGDFKPYKPGQTYIPCYFIHV	GNMT

	LKRTD	
305	MNGPVDGLCDHSLSEGVFMFTSESVGEGHPDKICDQISDAVL DAHLKQDPNAKVACETVCKTGMVLLCGEITSMAMVDYQRV VRDTIKHIGYDDSAKGFDFKTCNVLVALEQQSPDIAQCVHLD RNEEDVGAGDQGLMFGYATDETEECMPLTIILAHKLNARMA DLRRSGLLPWLRPDSKTQVTVQYMQDNGAVIPVRIHTIVISVQ HNEDITLEEMRRALKEQVIRAVVPAKYLDEDTVYHLQPSGRF VIGGPQGDAGVTGRKIIVDTYGGWGAHGGGAFSGKDYTKVD RSAAYAARWVAKSLVKAGLCRRVLVQVSYAIGVAEPLSISIF TYGTSQKTERELLDVVHKNFDLRPGVIVRDLDLKKPIYQKTA CYGHFGRSEFPWEVPRKLVF	MAT1A
306	MEKGPVRAPAEKPRGARCSNGFPERDPPRPGPSRPAEKPPRPE AKSAQPADGWKGERPRSEEDNELNLPNLAAAYSSILSSLGEN PQRQGLLKTPWRAASAMQFFTKGYQETISDVLNDAIFDEDHD EMVIVKDIDMFSMCEHHLVPFVGKVHIGYLPNKQVLGLSKLA RIVEIYSRRLQVQERLTKQIAVAITEALRPAGVGVVVEATHMC MVMRGVQKMNSKTVTSTMLGVFREDPKTREEFLTLIRS	GCH1
307	MAGKAHRLSAEERDQLLPNLRAVGWNELEGRDAIFKQFHFK DFNRAFGFMTRVALQAEKLDHHPEWFNVYNKVHITLSTHEC AGLSERDINLASFIEQVAVSMT	PCBD1
308	MSTEGGGRRCQAQVSRRISFSASHRLYSKFLSDEENLKLFGKC NNPNGHGHNYKVVVTVHGEIDPATGMVMNLADLKKYMEEA IMQPLDHKNLDMDVPYFADVVSTTENVAVYIWDNLQKVLPV GVLYKVKVYETDNNIVVYKGE	PTS
309	MAAAAAAGEARRVLVYGGRGALGSRCVQAFRARNWWVAS VDVVENEEASASIIVKMTDSFTEQADQVTAEVGKLLGEEKVD AILCVAGGWAGGNAKSKSLFKNCDLMWKQSIWTSTISSHLA TKHLKEGGLLTLAGAKAALDGTPGMIGYGMAKGAVHQLCQ SLAGKNSGMPPGAAAIAVLPVTLDTPMNRKSMPEADFSSWTP LEFLVETFHDWITGKNRPSSGSLIQVVTTEGRTELTPAYF	QDPR
310	MEGGLGRAVCLLTGASRGFGRTLAPLLASLLSPGSVLVLSAR NDEALRQLEAELGAERSGLRVVRVPADLGAEAGLQQLLGAL RELPRPKGLQRLLLINNAGSLGDVSKGFVDLSDSTQVNNYWA LNLTSMLCLTSSVLKAFPDSPGLNRTVVNISSLCALQPFKGWA LYCAGKAARDMLFQVLALEEPNVRVLNYAPGPLDTDMQQL ARETSVDPDMRKGLQELKAKGKLVDCKVSAQKLLSLLEKDE FKSGAHVDFYDK	SPR
311	MDAILNYRSEDTEDYYTLLGCDELSSVEQILAEFKVRALECHP DKHPENPKAVETFQKLQKAKEILTNEESRARYDHWRRSQMS MPFQQWEALNDSVKTSMHWVVRGKKDLMLEESDKTHTTK MENEECNEQRERKKEELASTAEKTEQKEPKPLEKSVSPQNSD SSGFADVNGWHLRFRWSKDAPSELLRKFRNYEI	DNAJC12
312	MLLPAPALRRALLSRPWTGAGLRWKHTSSLKVANEPVLAFT QGSPERDALQKALKDLKGRMEAIPCVVGDEEVWTSDVQYQV SPFNHGHKVAKFCYADKSLLNKAIEAALAARKEWDLKPIAD RAQIFLKAADMLSGPRRAEILAKTMVGQGKTVIQAEIDAAAE LIDFFRFNAKYAVELEGQQPISVPPSTNSTVYRGLEGFVAAISP FNFTAIGGNLAGAPALMGNVVLWKPSDTAMLASYAVYRILR EAGLPPNIIQFVPADGPLFGDTVTSSEHLCGINFTGSVPTFKHL WKQVAQ NLDRFHTFPRLAGECGGKNFHFVHRSADVESVVSGTLRSAFE YGGQKCSACSRLYVPHSLWPQIKGRLLEEHSRIKVGDPAEDF GTFFSAVIDAKSFARIKKWLEHARSSPSLTILAGGKCDDSVGY	ALDH4A1

[0890]

[0891]

	FVEPCIVESKDPQEPIMKEEIFGPVLSVYVYPDDKYKETLQLV DSTTSYGLTGAVFSQDKDVVQEATKVLRNAAGNFYINDKST GSIVGQQPFGGARASGTNDKPGGPHYILRWTSPQVIKETHKPL GDWSYAYMQ	
313	MALRRALPALRPCIPRFVQLSTAPASREQPAAGPAAVPGGGS ATAVRPPVPAVDFGNAQEAYRSRRTWELARSLLVLRLCAWP ALLARHEQLLYVSRKLLGQRLFNKLMKMTFYGHFVAGEDQE SIQPLLRHYRAFGVSAILDYGVEEDLSPEEAEHKEMESCTSAA ERDGSGTNKRDKQYQAHRAFGDRRNGVISARTYFYANEAKC DSHMETFLRCIEASGRVSDDGFIAIKLTALGRPQFLLQFSEVLA KWRCFFHQMAVEQGQAGLAAMDTKLEVAVLQESVAKLGIA SRAEIEDW FTAETLGVSGTMDLLDWSSLIDSRTKLSKHLVVPNAQTGQLE PLLSRFTEEEELQMTRMLQRMDVLAKKATEMGVRLMVDAE QTYFQPAISRLTLEMQRKFNVEKPLIFNTYQCYLKDAYDNVT LDVELARREGWCFGAKLVRGAYLAQERARAAEIGYEDPINPT YEATNAMYHRCLDYVLEELKHNAKAKVMVASHNEDTVRFA LRRMEELGLHPADHQVYFGQLLGMCDQISFPLGQAGYPVYK YVPYGPVMEVLPYLSRRALENSSLMKGTHRERQLLWLELLR RLRTGNLFHRPA	PRODH
314	MTTYSDKGAKPERGRFLHFHSVTFWVGNAKQAASFYCSKM GFEPLAYRGLETGSREVVSHVIKQGKIVFVLSSALNPWNKEM GDHLVKHGDGVKDIAFEVEDCDYIVQKARERGAKIMREPWV EQDKFGKVKFAVLQTYGDTTHTLVEKMNYIGQFLPGYEAPA FMDPLLPKLPKCSLEMIDHIVGNQPDQEMVSASEWYLKNLQF HRFWSVDDTQVHTEYSSLRSIVVANYEESIKMPINEPAPGKKK SQIQEYVDYNGGAGVQHIALKTEDIITAIRHLRERGLEFLSVPS TYYKQLREKLKTAKIKVKENIDALEELKILVDYDEKGYLLQIF TKPVQDRPTLFLEVIQRHNHQGFGAGNFNSLFKAFEEEQNLR GNLTNMETNGVVPGM	HPD
315	MEFSSPSREECPKPLSRVSIMAGSLTGLLLLQAVSWASGARPC IPKSFGYSSVVCVCNATYCDSFDPPTFPALGTFSRYESTRSGRR MELSMGPIQANHTGTGLLLTLQPEQKFQKVKGFGGAMTDAA ALNILALSPPAQNLLLKSYFSEEGIGYNIIRVPMASCDFSIRTYT YADTPDDFQLHNFSLPEEDTKLKIPLIHRALQLAQRPVSLLASP WTSPTWLKTNGAVNGKGSLKGQP GDIYHQTWARYFVKFLDAYAEHKLQFWAVTAENEPSAGLLS GYPFQCLGFTPEHQRDFIARDLGPTLANSTHHNVRLLMLDDQ RLLLPHWAKVVLTDPEAAKYVHGIAVHWYLDFLAPAKATLG ETHRLFPNTMLFASEACVGSKFWEQSVRLGSWDRGMQYSHS IITNLLYHVVGWTDWNLALNPEGGPNWVRNFVDSPIIVDITK DTFYKQPMFYHLGHFSKFIPEGSQRVGLVASQKNDLDAVAL MHPDGSAVVVVLNRSSKDVPLTIKDPAVGFLETISPGYSIHTY LWRRQ	GBA
316	MAELKYISGFGNECSSEDPRCPGSLPEGQNNPQVCPYNLYAE QLSGSAFTCPRSTNKRSWLYRILPSVSHKPFESIDEGQVTHNW DEVDPDPNQLRWKPFEIPKASQKKVDFVSGLHTLCGAGDIKS NNGLAIHIFLCNTSMENRCFYNSDGDFLIVPQKGNLLIYTEFG KMLVQPNEICVIQRGMRFSIDVFEETRGYILEVYGVHFELPDL GPIGANGLANPRDFLIPIAWYEDRQVPGGYTVINKYQGKLFA AKQDVSPFNVVAWHGNYTPYKYNLKNFMVINSVAFDHADPS IFTVLTAKSVRPGVAIADFVIFPPRWGVADKTFRPPYYHRNCM SEFMGLIRGHYEAKQGGFLPGGGSLHSTMTPHGPDADCFEKA SKVKLAPERIADGTMAFMFESSLSLAVTKWGLKASRCLDENY	HGD

[0892]

	HKCWEPLKSHFTPNSRNPAEPN	
317	MGVLGRVLLWLQLCALTQAVSKLWVPNTDFDVAANWSQNR TPCAGGAVEFPADKMVSVLVQEGHAVSDMLLPLDGELVLAS GAGFGVSDVGSHLDCGAGEPAVFRDSDRFSWHDPHLWRSGD EAPGLFFVDAERVPCRHDDVFFPPSASFRVGLGPGASPVRVRS ISALGRTFTRDEDLAVFLASRAGRLRFHGPGALSVGPEDCADP SGCVCGNAEAQPWICAALLQPLGGRCPQAACHSALRPQGQC CDLCGAVVLLTHGPAFDLERYRARILDTFLGLPQYHGLQVAV SKVPRSSRLREADTEIQVVLVENGPETGGAGRLARALLADVA ENGEALGVLEATMRESGAHVWGSSAAGLAGGVAAAVLLAL LVLLVAPPLLRRAGRLRWRRHEAAAPAGAPLGFRNPVFDVT ASEELPLPRRLSLVPKAAADSTSHSYFVNPLFAGAEAEA	AMN
318	MSGGWMAQVGAWRTGALGLALLLLLGLGLGLEAAASPLST PTSAQAAGPSSGSCPPTKFQCRTSGLCVPLTWRCDRDLDCSD GSDEEECRIEPCTQKGQCPPPPGLPCPCTGVSDCSGGTDKKLR NCSRLACLAGELRCTLSDDCIPLTWRCDGHPDCPDSSDELGC GTNEILPEGDATTMGPPVTLESVTSLRNATTMGPPVTLESVPS VGNATSSSAGDQSGSPTAYGVIAAAAVLSASLVTATLLLLSW LRAQERLRPLGLLVAMKESLLLSEQKTSLP	CD320
319	MMNMSLPFLWSLLTLLIFAEVNGEAGELELQRQKRSINLQQP RMATERGNLVFLTGSAQNIEFRTGSLGKIKLNDEDLSECLHQI QKNKEDIIELKGSAIGLPQNISSQIYQLNSKLVDLERKFQGLQQ TVDKKVCSSNPCQNGGTCLNLHDSFFCICPPQWKGPLCSADV NECEIYSGTPLSCQNGGTCVNTMGSYSCHCPPETYGPQCASK YDDCEGGSVARCVHGICEDLMREQAGEPKYSCVCDAGWMF SPNSPACTLDRDECSFQPGPCSTLVQCFNTQGSFYCGACPTG WQGNGYICEDINECEINNGGCSVAPPVECVNTPGSSHCQACPP GYQGDGRVCTLTDICSVSNGGCHPDASCSSTLGSLPLCTCLPG YTGNGYGPNGCVQLSNICLSHPCLNGQCIDTVSGYFCKCDSG WTGVNCTENINECLSNPCLNGGTCVDGVDSFSCECTRLWTG ALCQVPQQVCGESLSGINGSFSYRSPDVGYVHDVNCFWVIKT EMGKVLRITFTFFRLESMDNCPHEFLQVYDGDSSSAFQLGRFC GSSLPHELLSSDNALYFHLYSEHLRNGRGFTVRWETQQPECG GILTGPYGSIKSPGYPGNYPPGRDCVWIVVTSPDLLVTFTFGTL SLEHHDDCNKDYLEIRDGPLYQDPLLGKFCTTFSVPPLQTTGP FARIHFHSDSQISDQGFHITYLTSPSDLRCGGNYTDPEGELFLP ELSGPFTHTRQCVYMMKQPQGEQIQINFTHVELQCQSDSSQN YIEVRDGETLLGKVCGNGTISHIKSITNSVWIRFKIDASVEKAS FRAVYQVACGDELTGEGVIRSPFFPNVYPGERTCRWTIHQPQS QVILLNFTVFEIGSSAHCETDYVEIGSSSILGSPENKKYCGTDIP SFITSVYNFLYVTFVKSSSTENHGFMAKFSAEDLACGEILTEST GTIQSPGHPNVYPHGINCTWHILVQPNHLIHLMFETFHLEFHY NCTNDYLEVYDTDSETSLGRYCGKSIPPSLTSSGNSL MLVFVTDSDLAYEGFLINYEAISAATACLQDYTDDLGTFTSPN FPNNYPNNWECIYRITVRTGQLIAVHFTNFSLEEAIGNYYTDF LEIRDGGYEKSPLLGIFYGSNLPPTIISHSNKLWLKFKSDQIDTR SGFSAYWDGSSTGCGGNLTTSSGTFISPNYPMPYYHSSECYW WLKSSHGSAFELEFKDFHLEHHPNCTLDYLAVYDGPSSNSHL LTQLCGDEKPPLIRSSGDSMFIKLR TDEGQQGRGFKAEYRQTCENVVIVNQTYGILESIGYPNPYSEN QHCNWTIRATTGNTVNYTFLAFDLEHHINCSTDYLELYDGPR QMGRYCGVDLPPPGSTTSSKLQVLLLTDGVGRREKGFQMQW FVYGCGGELSGATGSFSSPGFPNRYPPNKECIWYIRTDPGSSIQ LTIHDFDVEYHSRCNFDVLEIYGGPDFHSPRIAQLCTQRSPENP	CUBN

[0893]

	MQVSSTGNELAIRFKTDLSINGRGFNASWQAVTGGCGGIFQA PSGEIHSPNYPSPYRSNTDCSWVIRVDRNHRVLLNFTDFDLEP QDSCIMAYDGLSSTMSRLARTCGREQLANPIVSSGNSLFLRFQ SGPSRQNRGFRAQFRQACGGHILTSSFDTVSSPRFPANYPNNQ NCSWIIQAQPPLNHITLSFTHFELERSTTCARDFVEILDGGHED APLRGRYCGTDMPHPITSFSSALTLRFVSDSSISAGGFHTTVTA SVSACGGTFYMAEGIFNSPGYPDIYPPNVECVWNIVSSPGNRL QLSFISFQLEDSQDCSRDFVEIREGNATGHLVGRYCGNSFPLN YSSIVGHTLWVRFISDGSGSGTGFQATFMKIFGNDNIVGTHGK VASPFWPENYPHNSNYQWTVNVNASHVVHGRILEMDIEEIQN CYYDKLRIYDGPSIHARLIGAYCGTQTESFSSTGNSLTFHFYSD SSISGKGFLLEWFAVDAPDGVLPTIAPGACGGFLRTGDAPVFL FSPGWPDSYSNRVDCTWLIQAPDSTVELNILSLDIESHRTCAY DSLVIRDGDNNLAQQLAVLCGREIPGPIRSTGEYMFIRFTSDSS VTRAGFNASFHKSCGGYLHADRGIITSPKYPETYPSNLNCSW HVLVQSGLTIAVHFEQPFQIPNGDSSCNQGDYLVLRNGPDICS PPLGPPGGNGHFCGSHASSTLFTSDNQMFVQFISDHSNEGQGF KIKYEAKSLACGGNVYIHDADSAGYVTSPNHPHNYPPHADCI WILAAPPETRIQLQFEDRFDIEVTPNCTSNYLELRDGVDSDAPI LSKFCGTSLPSSQWSSGEVMYLRFRSDNSPTHVGFKAKYSIAQ CGGRVPGQSGVVESIGHPTLPYRDNLFCEWHLQGLSGHYLTI SFEDFNLQNSSGCEKDFVEIWDNHTSGNILGRYCGNTIPDSIDT SSNTAVVRFVTDGSVTASGFRLRFESSMEECGGDLQGSIGTFT SPNYPNPNPHGRICEWRITAPEGRRITLMFNNLRLATHPSCNN EHVIVFNGIRSNSPQLEKLCSSVNVSNEIKSSGNTMKVIFFTDG SRPYGGFTASYTSSEDAVCGGSLPNTPEGNFTSPGYDGVRNY SRNLNCEWTLSNPNQGNSSISIHFEDFYLESHQDCQFDVLEFR VGDADGPLMWRLCGPSKPTLPLVIPYSQVWIHFVTNERVEHI GFHAKYSFTDCGGIQIGDSGVITSPNYPNAYDSLTHCSSLLEAP QGHTITLTFSDFDIEPHTTCAWDSVTVRNGGSPESPIIGQYCGN SNPRTIQSGSNQLVVTFNSDHSLQGGGFYATWNTQTLGCGGI FHSDNGTIRSPHWPQNFPENSRCSWTAITHKSKHLEISFDNNF LIPSGDGQCQNSFVKVWAGTEEVDKALLATGCGNVAPGPVIT PSNTFTAVFQSQEAPAQGFSASFVSRCGSNFTGPSGYIISPNYP KQYDNNMNCTYVIEANPLSVVLLTFVSFHLEARSAVTGSCVN DGVHIIRGYSVMSTPFATVCG DEMPAPLTIAGPVLLNFYSNEQITDFGFKFSYRIISCGGVFNFSS GIITSPAYSYADYPNDMHCLYTITVSDDKVIELKFSDFDVVPST SCSHDYLAIYDGANTSDPLLGKFCGSKRPPNVKSSNNSMLLV FKTDSFQTAKGWKMSFRQTLGPQQGCGGYLTGSNNTFASPD SDSNGMYDKNLNCVWIIIAPVNKVIHLTFNTFALEAASTRQRC LYDYVKLYDGDSENANLAGTFCGSTVPAPFISSGNFLTVQFIS DLTLEREGFNATYTIMDMPCGGTYNATWTPQNISSPNSSDPD VPFSICTWVIDSPPHQQVKITVWALQLTSQDCTQNYLQLQDSP QGHGNSRFQFCGRNASAVPVFYSSMSTAMVIFKSGVVNRNSR MSFTYQIADCNRDYHKAFGNLRSPGWPDNYDNDKDCTVTLT APQNHTISLFFHSLGIENSVECRNDFLEVRNGSNSNSPLLGKY CGTLLPNPVFSQNNELYLRFKSDSVTSDRGYEIIWTSSPSGCG GTLYGDRGSFTSPGYPGTYPNNTYCEWVLVAPAGRLVTINFY FISIDDPGDCVQNYLTLYDGPNASSPSSGPYCGGDTSIAPFVAS SNQVFIKFHADYARRPSAFRLTWDS	
320	MAWFALYLLSLLWATAGTSTQTQSSCSVPSAQEPLVNGIQVL MENSVTSSAYPNPSILIAMNLAGAYNLKAQKLLTYQLMSSDN NDLTIGQLGLTIMALTSSCRDPGDKVSILQRQMENWAPSSPN	GIF

[0894]

	AEASAFYGPSLAILALCQKNSEATLPIAVRFAKTLLANSSPFN VDTGAMATLALTCMYNKIPVGSEEGYRSLFGQVLKDIVEKIS MKIKDNGIIGDIYSTGLAMQALSVTPEPSKKEWNCKKTTDMI LNEIKQGKFHNPMSIAQILPSLKGKTYLDVPQVTCSPDHEVQP TLPSNPGPGPTSASNITVIYTINNQLRGVELLFNETINVSVKSGS VLLVVLEEAQRKNPMFKFETTMTSWGLVVSSINNIAENVNHK TYWQFLSGVTPLNEGVADYIPFNHEHITANFTQY	
321	MRQSHQLPLVGLLLFSFIPSQLCEICEVSEENYIRLKPLLNTMI QSNYNRGTSAVNVVLSLKLVGIQIQTLMQKMIQQIKYNVKSR LSDVSSGELALIILALGVCRNAEENLIYDYHLIDKLENKFQAEI ENMEAHNGTPLTNYYQLSLDVLALCLFNGNYSTAEVVNHFT PENKNYYFGSQFSVDTGAMAVLALTCVKKSLINGQIKADEGS LKNISIYTKSLVEKILSEKKENGLIGN TFSTGEAMQALFVSSDYYNENDWNCQQTLNTVLTEISQGAFS NPNAAAQVLPALMGKTFLDINKDSSCVSASGNFNISADEPITV TPPDSQSYISVNYSVRINETYFTNVTVLNGSVFLSVMEKAQK MNDTIFGFTMEERSWGPYITCIQGLCANNNDRTYWELLSGGE PLSQGAGSYVVRNGENLEVRWSKY	TCN1
322	MRHLGAFLFLLGVLGALTEMCEIPEMDSHLVEKLGQHLLPW MDRLSLEHLNPSIYVGLRLSSLQAGTKEDLYLHSLKLGYQQC LLGSAFSEDDGDCQGKPSMGQLALYLLALRANCEFVRGHKG DRLVSQLKWFLEDEKRAIGHDHKGHPHTSYYQYGLGILALCL HQKRVHDSVVDKLLYAVEPFHQGHHSVDTAAMAGLAFTCL KRSNFNPGRRQRITMAIRTVREEILKAQTPEGHFGNVYSTPLA LQFLMTSPMRGAELGTACLKARVALLASLQDGAFQNALMIS QLLPVLNHKTYIDLIFPDCLAPRVMLEPAAETIPQTQEIISVTLQ VLSLLPPYRQSISVLAGSTVEDVLKKAHELGGFTYETQASLSG PYLTSVMGKAAGEREFWQLLRDPNTPLLQGIADYRPKDGETI ELRLVSW	TCN2
323	MQQKTKLFLQALKYSIPHLGKCMQKQHLNHYNFADHCYNRI KLKKYHLTKCLQNKPKISELARNIPSRSFSCKDLQPVKQENEK PLPENMDAFEKVRTKLETQPQEEYEIINVEVKHGGFVYYQEG CCLVRSKDEEADNDNYEVLFNLEELKLDQPFIDCIRVAPDEK YVAAKIRTEDSEASTCVIIKLSDQPVMEASFPNVSSFEWVKDE EDEDVLFYTFQRNLRCHDVYRATFGDNKRNERFYTEKDPSYF VFLYLTKDSRFLTINIMNKTTSEVWLIDGLSPWDPPVLIQKRIH GVLYYVEHRDDELYILTNVGEPTEFKLMRTAADTPAIMNWD LFFTMKRNTKVIDLDMFKDHCVLFLKHSNLLYVNVIGLADDS VRSLKLPPWACGFIMDTNSDPKNCPFQLCSPIRPPKYYTYKFA EGKLFEETGHEDPITKTSRVLRLEAKSKDGKLVPMTVFHKTD SEDLQKKPLLVHVYGAYGMDLKMNFRPERRVLVDDGWILA YCHVRGGGELGLQWHADGRLTKKLNGLADLEACIKTLHGQ GFSQPSLTTLTAFSAGGVLAGALCNSNPELVRAVTLEAPFLDV LNTMMDTTLPLT LEELEEWGNPSSDEKHKNYIKRYCPYQNIKPQHYPSIHITAYE NDERVPLKGIVSYTEKLKEAIAEHAKDTGEGYQTPNIILDIQPG GNHVIEDSHKKITAQIKFLYEELGLDSTSVFEDLKKYLKF	PREPL
324	MAFANLRKVLISDSLDPCCRKILQDGGLQVVEKQNLSKEELIA ELQDCEGLIVRSATKVTADVINAAEKLQVVGRAGTGVDNVD LEAATRKGILVMNTPNGNSLSAAELTCGMIMCLARQIPQATA SMKDGKWERKKFMGTELNGKTLGILGLGRIGREVATRMQSF GMKTIGYDPIISPEVSASFGVQQLPLEEIWPLCDFITVHTPLLPS TTGLLNDNTFAQCKKGVRVVNCARGGIVDEGALLRALQSGQ CAGAALDVFTEEPPRDRALVDHENVISCPHLGASTKEAQSRC	PHGDH

[0895]

	GEEIA VQFVDMVKGKSLTGVVNAQALTSAFSPHTKPWIGLAEALGT LMRAWAGSPKGTIQVITQGTSLKNAGNCLSPAVIVGLLKEAS KQADVNLVNAKLLVKEAGLNVTTSHSPAAPGEQGFGECLLA VALAGAPYQAVGLVQGTTPVLQGLNGAVFRPEVPLRRDLPL LLFRTQTSDPAMLPTMIGLLAEAGVRLLSYQTSLVSDGETWH VMGISSLLPSLEAWKQHVTEAFQFHF	
325	MDAPRQVVNFGPGPAKLPHSVLLEIQKELLDYKGVGISVLEM SHRSSDFAKIINNTENLVRELLAVPDNYKVIFLQGGGCGQFSA VPLNLIGLKAGRCADYVVTGAWSAKAAEEAKKFGTINIVHPK LGSYTKIPDPSTWNLNPDASYVYYCANETVHGVEFDFIPDVK GAVLVCDMSSNFLSKPVDVSKFGVIFAGAQKNVGSAGVTVVI VRDDLLGFALRECPSVLEYKVQAGNSSLYNTPPCFSIYVMGL VLEWIKNNGGAAAMEKLSSIKSQTIYEIIDNSQGFYVCPVEPQ NRSKMNIPFRIGNAKGDDALEKRFLDKALELNMLSLKGHRSV GGIRASLYNAVTIEDVQKLAAFMKKFLEMHQL	PSAT1
326	MVSHSELRKLFYSADAVCFDVDSTVIREEGIDELAKICGVEDA VSEMTRRAMGGAVPFKAALTERLALIQPSREQVQRLIAEQPP HLTPGIRELVSRLQERNVQVFLISGGFRSIVEHVASKLNIPATN VFANRLKFYFNGEYAGFDETQPTAESGGKGKVIKLLKEKFHF KKIIMIGDGATDMEACPPADAFIGFGGNVIRQQVKDNAKWYI TDFVELLGELEE	PSPH
327	MQRAVSVVARLGFRLQAFPPALCRPLSCAQEVLRRTPLYDFH LAHGGKMVAFAGWSLPVQYRDSHTDSHLHTRQHCSLFDVSH MLQTKILGSDRVKLMESLVVGDIAELRPNQGTLSLFTNEAGGI LDDLIVTNTSEGHLYVVSNAGCWEKDLALMQDKVRELQNQ GRDVGLEVLDNALLALQGPTAAQVLQAGVADDLRKLPFMTS AVMEVFGVSGCRVTRCGYTGEDGVEISVPVAGAVHLATAIL KNPEVKLAGLAARDSLRLEAGLCLYGNDIDEHTTPVEGSLSW TLGKRRRAAMDFPGAKVIVPQLKGRVQRRRVGLMCEGAPM RAHSPILNMEGTKIGTVTSGCPSPSLKKNVAMGYVPCEYSRP GTMLLVEVRRKQQMAVVSKMPFVPTNYYTLK	AMT
328	MALRVVRSVRALLCTLRAVPSPAAPCPPRPWQLGVGAVRTL RTGPALLSVRKFTEKHEWVTTENGIGTVGISNFAQEALGDVV YCSLPEVGTKLNKQDEFGALESVKAASELYSPLSGEVTEINEA LAENPGLVNKSCYEDGWLIKMTLSNPSELDELMSEEAYEKYI KSIEE	GCSH
329	MQSCARAWGLRLGRGVGGGRRLAGGSGPCWAPRSRDSSSG GGDSAAAGASRLLERLLPRHDDFARRHIGPGDKDQREMLQT LGLASIDELIEKTVPANIRLKRPLKMEDPVCENEILATLHAISS KNQIWRSYIGMGYYNCSVPQTILRNLLENSGWITQYTPYQPE VSQGRLESLLNYQTMVCDITGLDMANASLLDEGTAAAEALQ LCYRHNKRRKFLVDPRCHPQTIAVVQTRAKYTGVLTELKLPC EMDFSGKDVSGVLFQYPDTEGKVEDFTELVERAHQSGSLACC ATDLLALC ILRPPGEFGVDIALGSSQRFGVPLGYGGPHAAFFAVRESLVRM MPGRMVGVTRDATGKEVYRLALQTREQHIRRDKATSNICTA QALLANMAAMFAIYHGSHGLEHIARRVHNATLILSEGLKRAG HQLQHDLFFDTLKIQCGCSVKEVLGRAAQRQINFRLFEDGTL GISLDETVNEKDLDDLLWIFGCESSAELVAESMGEECRGIPGS VFKRTSPFLTHQVFNSYHSETNIVRYMKKLENKDISLVHSMIP LGSCTMKLNSSSELAPITWKEFANIHPFVPLDQAQGYQQLFRE LEKDLCELTGYDQVCFQPNSGAQGEYAGLATIRAYLNQKGE	GLDC

		GHRTVCLIPKSAHGTNPASAHMAGMKIQPVEVDKYGNIDAV HLKAMVDKHKENLAAIMITYPSTNGVFEENISDVCDLIHQHG GQVYLDGANMNAQVGICRPGDFGSDVSHLNLHKTFCIPHGG GGPGMGPIGVKKHLAPFLPNHPVISLKRNEDACPVGTVSAAP WGSSSILPISWAYIKMMGGKGLKQATETAILNANYMAKRLET HYRILFRGARGYVGHEFILDTRPFKKSANIEAVDVAKRLQDY GFHAPTMSWPVAGTLMVEPTESEDKAELDRFCDAMISIRQEI ADIEEGRIDPRVNPLKMSPHSLTCVTSSHWDRPYSREVAAFPL PFVKPENKFWPTIARIDDIYGDQHLVCTCPPMEVYESPFSEQK RASS	
	330	MSLRCGDAARTLGPRVFGRYFCSPVRPLSSLPDKKKELLQNG PDLQDFVSGDLADRSTWDEYKGNLKRQKGERLRLPPWLKTE IPMGKNYNKLKNTLRNLNLHTVCEEARCPNIGECWGGGEYA TATATIMLMGDTCTRGCRFCSVKTARNPPPLDASEPYNTAKA IAEWGLDYVVLTSVDRDDMPDGGAEHIAKTVSYLKERNPKIL VECLTPDFRGDLKAIEKVALSGLDVYAHNVETVPELQSKVRD PRANFDQSLRVLKHAKKVQPDVISKTSIMLGLGENDEQVYAT MKALREADVDCLTLGQYMQPTRRHLKVEEYITPEKFKYWEK VGNELGFHYTASGPLVRSSYKAGEFFL KNLVAKRKTKDL	LIAS
	331	MAATARRGWGAAAVAAGLRRRFCHMLKNPYTIKKQPLHQF VQRPLFPLPAAFYHPVRYMFIQTQDTPNPNSLKFIPGKPVLET RTMDFPTPAAAFRSPLARQLFRIEGVKSVFFGPDFITVTKENEE LDWNLLKPDIYATIMDFFASGLPLVTEETPSGEAGSEEDDEVV AMIKELLDTRIRPTVQEDGGDVIYKGFEDGIVQLKLQGSCTSC PSSIITLKNGIQNMLQFYIPEVEGVEQVMDDESDEKEANSP	NFU1
[0896]	332	MSGGDTRAAIARPRMAAAHGPVAPSSPEQVTLLPVQRSFFLP PFSGATPSTSLAESVLKVWHGAYNSGLLPQLMAQHSLAMAQ NGAVPSEATKRDQNLKRGNWGNQIEFVLTSVGYAVGLGNV WRFPYLCYRNGGGAFMFPYFIMLIFCGIPLFFMELSFGQFASQ GCLGVWRISPMFKGVGYGMMVVSTYIGIYYNVVICIAFYYFF SSMTHVLPWAYCNNPWNTHDCAGVLDASNLTNGSRPAALPS NLSHLLNHSLQRTSPSEEYWRLYVLKLSDDIGNFGEVRLPLLG CLGVSWLVVFLCLIRGVKSSGKVVYFTATFPYVVLTILFVRG VTLEGAFDGIMYYLTPQWDKILEAKVWGDAASQIFYSLGCA WGGLITMASYNKFHNNCYRDSVIISITNCATSVYAGFVIFSILG FMANHLGVDVSRVADHGPGLAFVAYPEALTLLPISPLWSLLF FFMLILLGLGTQFCLLETLVTAIVDEVGNEWILQKKTYVTLGV AVAGFLLGIPLTSQAGIYWLLLMDNYAASFSLVVISCIMCVAI MYIYGHRNYFQDIQMMLGFPPPLFFQICWRFVSPAIIFFILVFT VIQYQPITYNHYQYPGWAVAIGFLMALSSVLCIPLYAMFRLC RTDGDTLLQRLKNATKPSRDWGPALLEHRTGRYAPTIAPSPE DGFEVQPLHPDKAQIPIVGSNGSSRLQDSRI	SLC6A9
	333	MEPSSKKLTGRLMLAVGGAVLGSLQFGYNTGVINAPQKVIEE FYNQTWVHRYGESILPTTLTTLWSLSVAIFSVGGMIGSFSVGL FVNRFGRRNSMLMMNLLAFVSAVLMGFSKLGKSFEMLILGR FIIGVYCGLTTGFVPMYVGEVSPTALRGALGTLHQLGIVVGILI AQVFGLDSIMGNKDLWPLLLSIIFIPALLQCIVLPFCPESPRFLLI NRNEENRAKSVLKKLRGTADVTHDLQEMKEESRQMMREKK VTILELFRSPAYRQPILIAVVLQLSQQLSGINAVFYYSTSIFEKA GVQQPVYATIGSGIVNTAFTVVSLFVVERAGRRTLHLIGLAG MAGCAILMTIALALLEQLPWMSYLSIVAIFGFVAFFEVGPGPIP WFIVAELFSQGPRPAAIAVAGFSNWTSNFIVGMCFQYVEQLC GPYVFIIFTVLLVLFFIFTYFKVPETKGRTFDEIASGFRQGGASQ	SLC2A1

[0897]

	SDKTPE ELFHPLGADSQV	
334	MDPSMGVNSVTISVEGMTCNSCVWTIEQQIGKVNGVHHIKVS LEEKNATIIYDPKLQTPKTLQEAIDDMGFDAVIHNPDPLPVLT DTLFLTVTASLTLPWDHIQSTLLKTKGVTDIKIYPQKRTVAVTI IPSIVNANQIKELVPELSLDTGTLEKKSGACEDHSMAQAGEVV LKMKVEGMTCHSCTSTIEGKIGKLQGVQRIKVSLDNQEATIV YQPHLISVEEMKKQIEAMGFPAFVKKQPKYLKLGAIDVERLK NTPVKSSEGSQQRSPSYTNDSTATFIIDGMHCKSCVSNIESTLS ALQYVSSIVVSLENRSAIVKYNASSVTPESLRKAIEAVSPGLYR VSITSEVESTSNSPSSSSLQKIPLNVVSQPLTQETVINIDGMTCN SCVQSIEGVISKKPGVKSIRVSLANSNGTVEYDPLLTSPETLRG AIEDMGFDATLSDTNEPLVVIAQPSSEMPLLTSTNEFYTKGMT PVQD KEEGKNSSKCYIQVTGMTCASCVANIERNLRREEGIYSILVAL MAGKAEVRYNPAVIQPPMIAEFIRELGFGATVIENADEGDGV LELVVRGMTCASCVHKIESSLTKHRGILYCSVALATNKAHIK YDPEIIGPRDIIHTIESLGFEASLVKKDRSASHLDHKREIRQWR RSFLVSLFFCIPVMGLMIYMMVMDHHFATLHHNQNMSKEEM INLHSSMFLERQILPGLSVMNLLSFLLC VPVQFFGGWYFYIQAYKALKHKTANMDVLIVLATTIAFAYSL IILLVAMYERAKVNPITFFDTPPMLFVFIALGRWLEHIAKGKTS EALAKLISLQATEATIVTLDSDNILLSEEQVDVELVQRGDIIKV VPGGKFPVDGRVIEGHSMVDESLITGEAMPVAKKPGSTVIAG SINQNGSLLICATHVGADTTLSQIVKLVEEAQTSKAPIQQFAD KLSGYFVPFIVFVSIATLLVWIVIG FLNFEIVETYFPGYNRSISRTETIIRFAFQASITVLCIACPCSLGL ATPTAVMVGTGVGAQNGILIKGGEPLEMAHKVKVVVFDKTG TITHGTPVVNQVKVLTESNRISHHKILAIVGTAESNSEHPLGTA ITKYCKQELDTETLGTCIDFQVVPGCGISCKVTNIEGLLHKNN WNIEDNNIKNASLVQIDASNEQSSTSSSMIIDAQISNALNAQQ YKVLIGNREWMIRNGLVINNDVN DFMTEHERKGRTAVLVAVDDELCGLIAIADTVKPEAELAIHIL KSMGLEVVLMTGDNSKTARSIASQVGITKVFAEVLPSHKVAK VKQLQEEGKRVAMVGDGINDSPALAMANVGIAIGTGTDVAI EAADVVLIRNDLLDVVASIDLSRKTVKRIRINFVFALIYNLVGI PIAAGVFMPIGLVLQPWMGSAAMAASSVSVVLSSLFLKLYRK PTYESYELPARSQIGQKSPSEISVHVGIDDTSRNSPKLGLLDRI VNYSRASINSLLSDKRSLNSVVTSEPDKHSLLVGDFREDDDTA L	ATP7A
335	MMRFMLLFSRQGKLRLQKWYLATSDKERKKMVRELMQVVL ARKPKMCSFLEWRDLKVVYKRYASLYFCCAIEGQDNELITLE LIHRYVELLDKYFGSVCELDIIFNFEKAYFILDEFLMGGDVQD TSKKSVLKAIEQADLLQEEDESPRSVLEEMGLA	AP1S1
336	MKILILGIFLFLCSTPAWAKEKHYYIGIIETTWDYASDHGEKK LISVDTEHSNIYLQNGPDRIGRLYKKALYLQYTDETFRTTIEKP VWLGFLGPIIKAETGDKVYVHLKNLASRPYTFHSHGITYYKE HEGAIYPDNTTDFQRADDKVYPGEQYTYMLLATEEQSPGEG DGNCVTRIYHSHIDAPKDIASGLIGPLIICKKDSLDKEKEKHID REFVVMFSVVDENFSWYLEDNIKTYC SEPEKVDKDNEDFQESNRMYSVNGYTFGSLPGLSMCAEDRV KWYLFGMGNEVDVHAAFFHGQALTNKNYRIDTINLFPATLF DAYMVAQNPGEWMLSCQNLNHLKAGLQAFFQVQECNKSSS KDNIRGKHVRHYYIAAEEIIWNYAPSGIDIFTKENLTAPGSDSA	CP

[0898]

	VFFEQGTTRIGGSYKKLVYREYTDASFTNRKERGPEEEHLGIL GPVIWAEVGDTIRVTFHNKGAYPLSIEPIGVRFNKNNEGTYYS PNYNPQSRSVPPSASHVAPTETFTYEWTVPKEVGPTNADPVC LAKMYY SAVDPTKDIFTGLIGPMKICKKGSLHANGRQKDVDKEFYLFPT VFDENESLLLEDNIRMFTTAPDQVDKEDEDFQESNKMHSMN GFMYGNQPGLTMCKGDSVVWYLFSAGNEADVHGIYFSGNT YLWRGERRDTANLFPQTSLTLHMWPDTEGTFNVECLTTDHY TGGMKQKYTVNQCRRQSEDSTFYLGERTYYIAAVEVEWDYS PQREWEKELHHLQEQNVSNAFLDKGEFYIGSKYKKVVYRQY TDSTFRVPVERKAEEEHLGILGPQLHADVGDKVKIIFKNMAT RPYSIHAHGVQTESSTVTPTLPGETLTYVWKIPERSGAGTEDS ACIPWAYYSTVDQVKDLYSGLIGPLIVCRRPYLKVFNPRRKLE FALLFLVFDENESWYLDDNIKTYSDHPEKVNKDDEEFIESNK MHAINGRMFGNLQGLTMHVGDEVNWYLMGMGNEIDLHTV HFHGHSFQYKHRGVYSSDVFDIFPGTYQTLEMFPRTPGIWLL HCHVTDHIHAGMETTYTVLQNEDTKSG	
337	MSPTISHKDSSRQRRPGNFSHSLDMKSGPLPPGGWDDSHLDS AGREGDREALLGDTGTGDFLKAPQSFRAELSSILLLLFLYVLQ GIPLGLAGSIPLILQSKNVSYTDQAFFSFVFWPFSLKLLWAPLV DAVYVKNFGRRKSWLVPTQYILGLFMIYLSTQVDRLLGNTD DRTPDVIALTVAFFLFEFLAATQDIAVDGWALTMLSRENVGY ASTCNSVGQTAGYFLGNVLFLALESADFCNKYLRFQPQPRGI VTLSDFLFFWGTVFLITTTLVALLKKENEVSVVKEETQGITDT YKL LFAIIKMPAVLTFCLLILTAKIGFSAADAVTGLKLVEEGVPKEH LALLAVPMVPLQIILPLIISKYTAGPQPLNTFYKAMPYRLLLGL EYALLVWWTPKVEHQGGFPIYYYIVVLLSYALHQVTVYSMY VSIMAFNAKVSDPLIGGTYMTLLNTVSNLGGNWPSTVALWL VDPLTVKECVGASNQNCRTPDAVELCKKLGGSCVTALDGYY VESIICVFIGFGWWFFLGPKFKKLQDEGSSSWKCKRNN	SLC33A1
338	MSAVCGGAARMLRTPGRHGYAAEFSPYLPGRLACATAQHY GIAGCGTLLILDPDEAGLRLFRSFDWNDGLFDVTWSENNEHV LITCSGDGSLQLWDTAKAAGPLQVYKEHAQEVYSVDWSQTR GEQLVVSGSWDQTVKLWDPTVGKSLCTFRGHESIIYSTIWSP HIPGCFASASGDQTLRIWDVKAAGVRIVIPAHQAEILSCDWCK YNENLLVTGAVDCSLRGWDLRNVRQPVFELLGHTYAIRRVK FSPFHASVLASCSYDFTVRFWNFSKPDSLLETVEHHTEFTCGL DFSLQSPTQVADCSWDETIKIYDPACLTIPA	PEX7
339	MEQLRAAARLQIVLGHLGRPSAGAVVAHPTSGTISSASFHPQ QFQYTLDNNVLTLEQRKFYEENGFLVIKNLVPDADIQRFRNEF EKICRKEVKPLGLTVMRDVTISKSEYAPSEKMITKVQDFQED KELFRYCTLPEILKYVECFTGPNIMAMHTMLINKPPDSGKKTS RHPLHQDLHYFPFRPSDLIVCAWTAMEHISRNNGCLVVLPGT HKGSLKPHDYPKWEGGVNKMFHGIQDYEENKARVHLVMEK GDTVFFHPLLIHGSGQNKTQGFRKAISCHFASADCHYIDVKGT SQENIEKEVVGIAHKFFGAENSVNLKDIWMFRARLVKGERTN L	PHYH
340	MAEAAAAAGGTGLGAGASYGSAADRDRDPDPDRAGRRLRV LSGHLLGRPREALSTNECKARRAASAATAAPTATPAAQESGTI PKKRQEVMKWNGWGYNDSKFIFNKKGQIELTGKRYPLSGM GLPTFKEWIQNTLGVNVEHKTTSKASLNPSDTPPSVVNEDFLH DLKETNISYSQEADDRVFRAHGHCLHEIFLLREGMFERIPDIVL WPTCHDDVVKIVNLACKYNLCIIPIGGGTSVSYGLMCPADET	AGPS

[0899]

	RTIISLDTSQMNRILWVDENNLTAHVEAGITGQELERQLKESG YCTGH EPDSLEFSTVGGWVSTRASGMKKNIYGNIEDLVVHIKMVTPR GIIEKSCQGPRMSTGPDIHHFIMGSEGTLGVITEATIKIRPVPEY QKYGSVAFPNFEQGVACLREIAKQRCAPASIRLMDNKQFQFG HALKPQVSSIFTSFLDGLKKFYITKFKGFDPNQLSVATLLFEGD REKVLQHEKQVYDIAAKFGGLAAGEDNGQRGYLLTYVIAYI RDLALEYYVLGESFETSAPWDRVVDLCRNVKERITRECKEKG VQFAPFSTCRVTQTYDAGACIYFYFAFNYRGISDPLTVFEQTE AAAREEILANGGSLSHHHGVGKLRKQWLKESISDVGFGMLK SVKEYVDPNNIFGNRNLL	
341	MESSSSSNSYFSVGPTSPSAVVLLYSKELKKWDEFEDILEERR HVSDLKFAMKCYTPLVYKGITPCKPIDIKCSVLNSEEIHYVIKQ LSKESLQSVDVLREEVSEILDEMSHKLRLGAIRFCAFTLSKVF KQIFSKVCVNEEGIQKLQRAIQEHPVVLLPSHRSYIDFLMLSFL LYNYDLPVPVIAAGMDFLGMKMVGELLRMSGAFFMRRTFG GNKLYWAVFSEYVKTMLRNGYAPVEFFLEGTRSRSAKTLTP KFGLLNIVMEPFFKREVFDTYLVPISISYDKILEETLYVYELLG VPKPKESTTGLLKARKILSENFGSIHVYFGDPVSLRSLAAGRM SRSSYNLVPRYIPQKQSEDMHAFVTEVAYKMELLQIENMVLS PWTLIVAVLLQNRPSMDFDALVEKTLWLKGLTQAFGGFLIWP DNKPAEEVVPASILLHSNIASLVKDQVILKVDSGDSEVVDGL MLQHITLLMCSAYRNQLLNIFVRPSLVAVALQMTPGFRKEDV YSCFRFLRDVFADEFIFLPGNTLKDFEEGCYLLCKSEAIQVTTK DILVTEKGNTVLEFLVGLFKPFVESYQIICKYLLSEEEDHFSEE QYLAAVRKFTSQLLDQGTSQCYDVLSSDVQKNALAACVRLG VVEKKKINNNCIFNVNEPATTKLEEMLGCKTPIGKPATAKL	GNPAT
342	MPVLSRPRPWRGNTLKRTAVLLALAAYGAHKVYPLVRQCLA PARGLQAPAGEPTQEASGVAAAKAGMNRVFLQRLLWLLRLL FPRVLCRETGLLALHSAALVSRTFLSVYVARLDGRLARCIVR KDPRAFGWQLLQWLLIALPATFVNSAIRYLEGQLALSFRSRL VAHAYRLYFSQQTYYRVSNMDGRLRNPDQSLTEDVVAFAAS VAHLYSNLTKPLLDVAVTSYTLLRAARSRGAGTAWPSAIAGL VVFLTANVLRAFSPKFGELVAEEARRKGELRYMHSRVVANS EEIAFYGGHEVELALLQRSYQDLASQINLILLERLWYVMLEQF LMKYVWSASGLLMVAVPIITATGYSESDAEAVKKAALEKKE EELVSERTEAFTIARNLLTAAADAIERIMSSYKEVTELAGYTA RVHEMFQVFEDVQRCHFKRPRELEDAQAGSGTIGRSGVRVE GPLKIRGQVVDVEQGIICENIPIVTPSGEVVVASLNIRVEEGMH LLITGPNGCGKSSLFRILGGLWPTYGGVLYKPPPQRMFYIPQR PYMSVGSLRDQVIYPDSVEDMQRKGYSEQDLEAILDVVHLH HILQREGGWEAMCD WKDVLSGGEKQRIGMARMFYHRPKYALLDECTSAVSIDVEG KIFQAAKDAGIALLSITHRPSLWKYHTHLLQFDGEGGWKFEK LDSAARLSLTEEKQRLEQQLAGIPKMQRRLQELCQILGEAVA PAHVPAPSPQGPGGLQGAST	ABCD1
343	MNPDLRRERDSASFNPELLTHILDGSPEKTRRRREIENMILNDP DFQHEDLNFLTRSQRYEVAVRKSAIMVKKMREFGIADPDEIM WFKKLHLVNFVEPVGLNYSMFIPTLLNQGTTAQKEKWLLSSK GLQIIGTYAQTEMGHGTHLRGLETTATYDPETQEFILNSPTVT SIKWWPGGLGKTSNHAIVLAQLITKGKCYGLHAFIVPIREIGT HKPLPGITVGDIGPKFGYDEIDNGYLKMDNHRIPRENMLMKY AQVKPDGTYVKPLSNKLTYGTMVFVRSFLVGEAARALSKAC TIAIRYSAVRHQSEIKPGEPEPQILDFQTQQYKLFPLLATAYAF	ACOX1

[0900]

	QFVGAYMKETYHRINEGIGQGDLSELPELHALTAGLKAFTSW TANTGIEACRMACGGHGYSHCSGLPNIYVNFTPSCTFEGENT VMMLQTARFLMKSYDQVHSGKLVCGMVSYLNDLPSQRIQPQ QVAVWPTMVDINSPESLTEAYKLRAARLVEIAAKNLQKEVIH RKSKEVAWNLTSVDLVRASEAHCHYVVVKLFSEKLLKIQDK AIQAVLRSLCLLYSLYGISQNAGDFLQGSIMTEPQITQVNQRV KELLTLIRSDAVALVDAFDFQDVTLGSVLGRYDGNVYENLFE WAKNSPLNKAEVHESYKHLKSLQSKL	
344	MWGSDRLAGAGGGGAAVTVAFTNARDCFLHLPRRLVAQLH LLQNQAIEVVWSHQPAFLSWVEGRHFSDQGENVAEINRQVG QKLGLSNGGQVFLKPCSHVVSCQQVEVEPLSADDWEILELHA VSLEQHLLDQIRIVFPKAIFPVWVDQQTYIFIQIVALIPAASYGR LETDTKLLIQPKTRRAKENTFSKADAEYKKLHSYGRDQKGM MKELQTKQLQSNTVGITESNENESEIPVDSSSVASLWTMIGSIF SFQSEKKQETSWGLTEINAFKNMQSKVVPLDNIFRVCKSQPPS IYNASATSVFHKHCAIHVFPWDQEYFDVEPSFTVTYGKLVKL LSPKQQQSKTKQNVLSPEKEKQMSEPLDQKKIRSDHNEEDEK ACVLQVVWNGLEELNNAIKYTKNVEVLHLGKVWIPDDLRKR LNIEMHAVVRITPVEVTPKIPRSLKLQPRENLPKDISEEDIKTVF YSWLQQSTTTMLPLVISEEEFIKLETKDGLKEFSLSIVHSWEKE KDKNIFLLSPNLLQKTTIQVLLDPMVKEEN SEEIDFILPFLKLSSLGGVNSLGVSSLEHITHSLLGRPLSRQLMS LVAGLRNGALLLTGGKGSGKSTLAKAICKEAFDKLDAHVER VDCKALRGKRLENIQKTLEVAFSEAVWMQPSVVLLDDLDLIA GLPAVPEHEHSPDAVQSQRLAHALNDMIKEFISMGSLVALIAT SQSQQSLHPLLVSAQGVHIFQCVQHIQPPNQEQRCEILCNVIK NKLDCDINKFTDLDLQHVAKETGGFVARDFTVLVDRAIHSRL SRQSISTREKLVLTTLDFQKALRGFLPASLRSVNLHKPRDLGW DKIGGLHEVRQILMDTIQLPAKYPELFANLPIRQRTGILLYGPP GTGKTLLAGVIARESRMNFISVKGPELLSKYIGASEQAVRDIFI RAQAAKPCILFFDEFESIAPRRGHDNTGVTDRVVNQLLTQLD GVEGLQGVYVLAATSRPDLIDPALLRPGRLDKCVYCPPPDQV SRLEILNVLSDSLPLADDVDLQHVASVTDSFTGADLKALLYN AQLEALHGMLLSSGLQDGSSSSDSDLSLSSMVFLNHSSGSDDS AGDGECGLDQSLVSLEMSEILPDESKFNMYRLYFGSSYESELG NGTSSDLSSQCLSAPSSMTQDLPGVPGKDQLFSQPPVLRTASQ EGCQELTQEQRDQLRADISIIKGRYRSQSGEDESMNQPGPIKT RLAISQSHLMTALGHTRPSISEDDWKNFAELYESFQNPKRRK NQSGTMFRPGQKVTLA	PEX1
345	MASRKENAKSANRVLRISQLDALELNKALEQLVWSQFTQCF HGFKPGLLARFEPEVKACLWVFLWRFTIYSKNATVGQSVLNI KYKNDFSPNLRYQPPSKNQKIWYAVCTIGGRWLEERCYDLFR NHHLASFGKVKQCVNFVIGLLKLGGLINFLIFLQRGKFATLTE RLLGIHSVFCKPQNICEVGFEYMNRELLWHGFAEFLIFLLPLIN VQKLKAKLSSWCIPLTGAPNSDNTLATSGKECALCGEWPTMP HTIGCEHIFCYFCAKSSFLFDVYFTCPKCGTEVHSLQPLKSGIE MSEVNAL	PEX2
346	MLRSVWNFLKRHKKKCIFLGTVLGGVYILGKYGQKKIREIQE REAAEYIAQARRQYHFESNQRTCNMTVLSMLPTLREALMQQ LNSESLTALLKNRPSNKLEIWEDLKIISFTRSTVAVYSTCMLV VLLRVQLNIIGGYIYLDNAAVGKNGTTILAPPDVQQQYLSSIQ HLLGDGLTELITVIKQAVQKVLGSVSLKHSLSLLDLEQKLKEI RNLVEQHKSSSWINKDGSKPLLCHYMMPDEETPLAVQACGL SPRDITTIKLLNETRDMLESPDFSTVLNTCLNRGFSRLLDNMA	PEX3

		EFFRPTEQDLQHGNSMNSLSSVSLPLAKIIPIVNGQIHSVCSETP SHFVQDLLTMEQVKDFAANVYEAFSTPQQLEK	
[0901]	347	MAMRELVEAECGGANPLMKLAGHFTQDKALRQEGLRPGPW PPGAPASEAASKPLGVASEDELVAEFLQDQNAPLVSRAPQTF KMDDLLAEMQQIEQSNFRQAPQRAPGVADLALSENWAQEFL AAGDAVDVTQDYNETDWSQEFISEVTDPLSVSPARWAEEYL EQSEEKLWLGEPEGTATDRWYDEYHPEEDLQHTASDFVAKV DDPKLANSEFLKFVRQIGEGQVSLESGAGSGRAQAEQWAAEF IQQQGTSDAWVDQFTRPVNTSALDMEFERAKSAIESDVDFW DKLQAELEEMAKRDAEAHPWLSDYDDLTSATYDKGYQFEEE NPLRDHPQPFEEGLRRLQEGDLPNAVLLFEAAVQQDPKHME AWQYLGTTQAENEQELLAISALRRCLELKPDNQTALMALAV SFTNESLQRQACETLRDWLRYTPAYAHLVTPAEEGAGGAGL GPSKRILGSLLSDSLFLEVKELFLAAVRLDPTSIDPDVQCGLGV LFNLSGEYDKAVDCFTAALSVRPNDYLLWNKLGATLANGNQ SEEAVAAYRRALELQPGYIRSRYNLGISCINLGAHREAVEHFL EALNMQRKSRGPRGEGGAMSENIWSTLRLALSMLGQSDAYG AADARDLSTLLTMFGLPQ	PEX5
	348	MALAVLRVLEPFPTETPPLAVLLPPGGPWPAAELGLVLALRP AGESPAGPALLVAALEGPDAGTEEQGPGPPQLLVSRALLRLL ALGSGAWVRARAVRRPPALGWALLGTSLGPGLGPRVGPLLV RRGETLPVPGPRVLETRPALQGLLGPGTRLAVTELRGRARLCP ESGDSSRPPPPPVVSSFAVSGTVRRLQGVLGGTGDSLGVSRSC LRGLGLFQGEWVWVAQARESSNTSQPHLARVQVLEPRWDLS DRLGPGSGPLGEPLADGLALVPATLAFNLGCDPLEMGELRIQ RYLEGS IAPEDKGSCSLLPGPPFARELHIEIVSSPHYSTNGNYDGVLYRH FQIPRVVQEGDVLCVPTIGQVEILEGSPEKLPRWREMFFKVKK TVGEAPDGPASAYLADTTHTSLYMVGSTLSPVPWLPSEESTL WSSLSPPGLEALVSELCAVLKPRLQPGGALLTGTSSVLLRGPP GCGKTTVVAAACSHLGLHLLKVPCSSLCAESSGAVETKLQAI FSRARRCRPAVLLLTAVDLLGRDRDGLGEDARVMAVLRHLL LNEDPLNSCPPLMVVATTSRAQDLPADVQTAFPHELEVPALS EGQRLSILRALTAHLPLGQEVNLAQLARRCAGFVVGDLYALL THSSRAACTRIKNSGLAGGLTEEDEGELCAAGFPLLAEDFGQ ALEQLQTAHSQAVGAPKIPSVSWHDVGGLQEVKKEILETIQLP LEHPELLSLGLRRSGLLLHGPPGTGKTLLAKAVATECSLTFLS VKGPELINMYVGQSEENVREVFARARAAAPCIIFFDELDSLAP SRGRSGDSGGVMDRVVSQLLAELDGLHSTQ DVFVIGATNRPDLLDPALLRPGRFDKLVFVGANEDRASQLRV LSAITRKFKLEPSVSLVNVLDCCPPQLTGADLYSLCSDAMTAA LKRRVHDLEEGLEPGSSALMLTMEDLLQAAARLQPSVSEQEL LRYKRIQRKFAAC	PEX6
	349	MAPAAASPPEVIRAAQKDEYYRGGLRSAAGGALHSLAGARK WLEWRKEVELLSDVAYFGLTTLAGYQTLGEEYVSIIQVDPSRI HVPSSLRRGVLVTLHAVLPYLLDKALLPLEQELQADPDSGRP LQGSLGPGGRGCSGARRWMRHHTATLTEQQRRALLRAVFVL RQGLACLQRLHVAWFYIHGVFYHLAKRLTGITYLRVRSLPGE DLRARVSYRLLGVISLLHLVLSMGLQLYGFRQRQRARKEWR LHRGLSHRRASLEERAVSRNPLCTLCLEERRHPTATPCGHLFC WECITAW CSSKAECPLCREKFPPQKLIYLRHYR	PEX10
	350	MAEHGAHFTAASVADDQPSIFEVVAQDSLMTAVRPALQHVV KVLAESNPTHYGFLWRWFDEIFTLLDLLLQQHYLSRTSASFSE	PEX12

		NFYGLKRIVMGDTHKSQRLASAGLPKQQLWKSIMFLVLLPYL KVKLEKLVSSLREEDEYSIHPPSSRWKRFYRAFLAAYPFVNM AWEGWFLVQQLRYILGKAQHHSPLLRLAGVQLGRLTVQDIQ ALEHKPAKASMMQQPARSVSEKINSALKKAVGGVALSLSTG LSVGVFFLQFLDWWYSSENQETIKSLTALPTPPPPVHLDYNSD SPLLPKMKTVCPLCRKTRVNDTVLATSGYVFCYRCVFHYVRS HQACPITGYPTEVQHLIKLYSPEN	
	351	MASQPPPPPKPWETRRIPGAGPGPGPGPTFQSADLGPTLMTRP GQPALTRVPPPILPRPSQQTGSSSVNTFRPAYSSFSSGYGAYGN SFYGGYSPYSYGYNGLGYNRLRVDDLPPSRFVQQAEESSRGA FQSIESIVHAFASVSMMMDATFSAVYNSFRAVLDVANHFSRL KIHFTKVFSAFALVRTIRYLYRRLQRMLGLRRGSENEDLWAE SEGTVACLGAEDRAATSAKSWPIFLFFAVILGGPYLIWKLLST HSDEVTDSINWASGEDDHVVARAEYDFAAVSEEEISFRAGDM LNLALKEQQPKVRGWLLASLDGQTTGLIPANYVKILGKRKGR KTVESSKVSKQQQSFTNPTLTKGATVADSLDEQEAAFESVFV ETNKVPVAPDSIGKDGEKQDL	PEX13
	352	MASSEQAEQPSQPSSTPGSENVLPREPLIATAVKFLQNSRVRQ SPLATRRAFLKKKGLTDEEIDMAFQQSGTAADEPSSLGPATQ VVPVQPPHLISQPYSPAGSRWRDYGALAIIMAGIAFGFHQLYK KYLLPLILGGREDRKQLERMEAGLSELSGSVAQTVTQLQTTL ASVQELLIQQQQKIQELAHELAAAKATTSTNWILESQNINELK SEINSLKGLLLNRRQFPPSPSAPKIPSWQIPVKSPSPSSPAAVNH HSSSDISPVSNESTSSSPGKEGHSPEGSTVTYHLLGPQEEGEGV VDVKGQVRMEVQGEEEKREDKEDEEDEEDDDVSHVDEEDC LGVQREDRRGGDGQINEQVEKLRRPEGASNESERD	PEX14
[0902]	353	MEKLRLLGLRYQEYVTRHPAATAQLETAVRGFSYLLAGRFA DSHELSELVYSASNLLVLLNDGILRKELRKKLPVSLSQQKLLT WLSVLECVEVFMEMGAAKVWGEVGRWLVIALVQLAKAVLR MLLLLWFKAGLQTSPPIVPLDRETQAQPPDGDHSPGNHEQSY VGKRSNRVVRTLQNTPSLHSRHWGAPQQREGRQQQHHEELS ATPTPLGLQETIAEFLYIARPLLHLLSLGLWGQRSWKPWLLAG VVDVTSLSLLSDRKGLTRRERRELRRRTILLLYYLLRSPFYDR FSEARIL FLLQLLADHVPGVGLVTRPLMDYLPTWQKIYFYSWG	PEX16
	354	MAAAEEGCSVGAEADRELEELLESALDDFDKAKPSPAPPSTT TAPDASGPQKRSPGDTAKDALFASQEKFFQELFDSELASQAT AEFEKAMKELAEEEPHLVEQFQKLSEAAGRVGSDMTSQQEFT SCLKETLSGLAKNATDLQNSSMSEEELTKAMEGLGMDEGDG EGNILPIMQSIMQNLLSKDVLYPSLKEITEKYPEWLQSHRESLP PEQFEKYQEQHSVMCKICEQFEAETPTDSETTQKARFEMVLD LMQQLQDLGHPPKELAGEMPPGLNFDLDALNLSGPPGASGE QCLIM	PEX19
	355	MKSDSSTSAAPLRGLGGPLRSSEPVRAVPARAPAVDLLEEAA DLLVVHLDFRAALETCERAWQSLANHAVAEEPAGTSLEVKC SLCVVGIQALAEMDRWQEVLSWVLQYYQVPEKLPPKVLELC ILLYSKMQEPGAVLDVVGAWLQDPANQNLPEYGALAEFHVQ RVLLPLGCLSEAEELVVGSAAFGEERRLDVLQAIHTARQQQK QEHSGSEEAQKPNLEGSVSHKFLSLPMLVRQLWDSAVSHFFS LPFKKSLLAALILCLLVVRFDPASPSSLHFLYKLAQLFRWIRK AAFSRLYQ LRIRD	PEX26
	356	MALQGISVVELSGLAPGPFCAMVLADFGARVVRVDRPGSRY	AMACR

[0903]

	DVSRLGRGKRSLVLDLKQPRGAAVLRRLCKRSDVLLEPFRRG VMEKLQLGPEILQRENPRLIYARLSGFGQSGSFCRLAGHDINY LALSGVLSKIGRSGENPYAPLNLLADFAGGGLMCALGIIMALF DRTRTGKGQVIDANMVEGTAYLSSFLWKTQKLSLWEAPRGQ NMLDGGAPFYTTYRTADGEFMAVGAIEPQFYELLIKGLGLKS DELPNQMSMDDWPEMKKKFADVFAEKTKAEWCQIFDGTDA CVTPVLTFEEVVHHDHNKERGSFITSEEQDVSPRPAPLLLNTP AIPSFKRDPFIGEHTEEILEEFGFSREEIYQLNSDKIIESNKVKAS L	
357	MAQTPAFDKPKVELHVHLDGSIKPETILYYGRRRGIALPANTA EGLLNVIGMDKPLTLPDFLAKFDYYMPAIAGCREAIKRIAYEF VEMKAKEGVVYVEVRYSPHLLANSKVEPIPWNQAEGDLTPD EVVALVGQGLQEGERDFGVKARSILCCMRHQPNWSPKVVEL CKKYQQQTVVAIDLAGDETIPGSSLLPGHVQAYQEAVKSGIH RTVHAGEVGSAEVVKEAVDILKTERLGHGYHTLEDQALYNR LRQENMHFEICPWSSYLTGAWKPDTEHAVIRLKNDQANYSL NTDDPLIF KSTLDTDYQMTKRDMGFTEEEFKRLNINAAKSSFLPEDEKRE LLDLLYKAYGMPPSASAGQNL	ADA
358	MAAGGDHGSPDSYRSPLASRYASPEMCFVFSDRYKFRTWRQ LWLWLAEAEQTLGLPITDEQIQEMKSNLENIDFKMAAEEEKR LRHDVMAHVHTFGHCCPKAAGIIHLGATSCYVGDNTDLIILR NALDLLLPKLARVISRLADFAKERASLPTLGFTHFQPAQLTTV GKRCCLWIQDLCMDLQNLKRVRDDLRFRGVKGTTGTQASFL QLFEGDDHKVEQLDKMVTEKAGFKRAFIITGQTYTRKVDIEV LSVLASLGASVHKICTDIRLLANLKEMEEPFEKQQIGSSAMPY KRNPMRSERCCSLARHLMTLVMDPLQTASVQWFERTLDDSA NRRICLAEAFLTADTILNTLQNISEGLVVYPKVIERRIRQELPF MATENIIMAMVKAGGSRQDCHEKIRVLSQQAASVVKQEGGD NDLIERIQVDAYFSPIHSQLDHLLDPSSFTGRASQQVQRFLEEE VYPLLKPYESVMKVKAELCL	ADSL
359	MNVRIFYSVSQSPHSLLSLLFYCAILESRISATMPLFKLPAEEK QIDDAMRNFAEKVFASEVKDEGGRQEISPFDVDEICPISHHEM QAHIFHLETLSTSTEARRKKRFQGRKTVNLSIPLSETSSTKLSHI DEYISSSPTYQTVPDFQRVQITGDYASGVTVEDFEIVCKGLYR ALCIREKYMQKSFQRFPKTPSKYLRNIDGEAWVANESFYPVF TPPVKKGEDPFRTDNLPENLGYHLKMKDGVVYVYPNEAAVS KDEPKPLPYPNLDTFLDDMNFLLALIAQGPVKTYTHRRLKFLS SKFQVHQMLNEMDELKELKNNPHRDFYNCRKVDTHIHAAAC MNQKHLLRFIKKSYQIDADRVVYSTKEKNLTLKELFAKLKM HPYDLTVDSLDVHAGRQTFQRFDKFNDKYNPVGASELRDLY LKTDNYINGEYFATIIKEVGADLVEAKYQHAEPRLSIYGRSPD EWSKLSSWFVCNRIHCPNMTWMIQVPRIYDVFRSKNFLPHFG KMLENIFMPVFEATINPQADPELSVFLKHIT GFDSVDDESKHSGHMFSSKSPKPQEWTLEKNPSYTYYAYYM YANIMVLNSLRKERGMNTFLFRPHCGEAGALTHLMTAFMIA DDISHGLNLKKSPVLQYLFFLAQIPIAMSPLSNNSLFLEYAKNP FLDFLQKGLMISLSTDDPMQFHFTKEPLMEEYAIAAQVFKLST CDMCEVARNSVLQCGISHEEKVKFLGDNYLEEGPAGNDIRRT NVAQIRMAYRYETWCYELNLIAEGLKSTE	AMPD1
360	MATEGMILTNHDHQIRVGVLTVSDSCFRNLAEDRSGINLKDL VQDPSLLGGTISAYKIVPDEIEEIKETLIDWCDEKELNLILTTGG TGFAPRDVTPEATKEVIEREAPGMALAMLMGSLNVTPLGMLS RPVCGIRGKTLIINLPGSKKGSQECFQFILPALPHAIDLLRDAIV	GPHN

[0904]

	KVKEVHDELEDLPSPPPPLSPPPTTSPHKQTEDKGVQCEEEEE EKKDSGVASTEDSSSSHITAAAIAAKIPDSIISRGVQVLPRDTA SLSTTPSESPRAQATSRLSTASCPTPKVQSRCSSKENILRASHS AVDITKVARRHRMSPFPLTSMDKAFITVLEMTPVLGTEIINYR DGMGRVLAQDVYAKDNLPPFPASVKDGYAVRAADGPGDRFI IGESQAGEQPTQTVMPGQVMRVTTGAPIPCGADAVVQVEDT ELIRESDDGTEELEVRILVQARPGQDIRPIGHDIKRGECVLAKG THMGPS EIGLLATVGVTEVEVNKFPVVAVMSTGNELLNPEDDLLPGKI RDSNRSTLLATIQEHGYPTINLGIVGDNPDDLLNALNEGISRA DVIITSGGVSMGEKDYLKQVLDIDLHAQIHFGRVFMKPGLPTT FATLDIDGVRKIIFALPGNPVSAVVTCNLFVVPALRKMQGILD PRPTIIKARLSCDVKLDPRPEYHRCILTWHHQEPLPWAQSTGN QMSSRLMSMRSANGLLMLPPKTEQYVELHKGEVVDVMVIGRL	
361	MAGAAAESGRELWTFAGSRDPSAPRLAYGYGPGSLRELRAR EFSRLAGTVYLDHAGATLFSQSQLESFTSDLMENTYGNPHSQ NISSKLTHDTVEQVRYRILAHFHTTAEDYTVIFTAGSTAALKL VAEAFPWVSQGPESSGSRFCYLTDSHTSVVGMRNVTMAINVI STPVRPEDLWSAEERSASASNPDCQLPHLFCYPAQSNFSGVRY PLSWIEEVKSGRLHPVSTPGKWFVLLDAASYVSTSPLDLSAH QADFVPISFYKIFGFPTGLGALLVHNRAAPLLRKTYFGGGTAS AYLAGEDFYIPRQSVAQRFEDGTISFLDVIALKHGFDTLERLT GGMENIKQHTFTLAQYTYVALSSLQYPNGAPVVRIYSDSEFSS PEVQGPIINFNVLDDKGNIIGYSQVDKMASLYNIHLRTGCFCN TGACQRHLGISNEMVRKHFQAGHVCGDNMDLIDGQPTGSVR ISFGYMSTLDDVQAFLRFIIDTRLHSSGDWPVPQAHADTGETG APSADSQADVIPAVMGRRSLSPQEDALTGSRVWNNSSTVNA VPVAPPVCDVARTQPTPSEKAAGVLEGALGPHVVTNLYLYPI KSCAAFEVTRWPVGNQGLLYDRSWMVVNHNGVCLSQKQEP RLCLIQPFIDLRQRIMVIKAKGMEPIEVPLEENSERTQIRQSRV CADRVSTYDCGEKISSWLSTFFGRPCHLIKQSSNSQRNAKKK HGKDQLPGTMATLSLVNEAQYLLINTSSILELHRQLNTSDENG KEELFSLKDLSLRFRANIIINGKRAFEEEKWDEISIGSLRFQVLG PCHRCQMICIDQQTGQRNQHVFQKLSESRETKVNFGMYLMH ASLDLSSPCFLSVGSQVLPVLKENVEGHDLPASEKHQDVTS	MOCOS
362	MAARPLSRMLRRLLRSSARSCSSGAPVTQPCPGESARAASEE VSRRRQFLREHAAPFSAFLTDSFGRQHSYLRISLTEKCNLRCQ YCMPEEGVPLTPKANLLTTEEILTLARLFVKEGIDKIRLTGGEP LIRPDVVDIVAQLQRLEGLRTIGVTTNGINLARLLPQLQKAGL SAINISLDTLVPAKFEFIVRRKGFHKVMEGIHKAIELGYNPVK VNCVVMRGLNEDELLDFAALTEGLPLDVRFIEYMPFDGNKW NFKKMVSYKEMLDTVRQQWPELEKVPEEESSTAKAFKIPGFQ GQISFITSMSEHFCGTCNRLRITADGNLKVCLFGNSEVSLRDH LRAGASEQELLRIIGAAVGRKKRQHAGMFSISQMKNRPMILIE LFLMFPNSPPANPSIFSWDPLHVQGLRPRMSFSSQVATLWKG CRVPQTPPLAQQRLGSGSFQRHYTSRADSDANSKCLSPGSWA SAAPSGPQLTSEQLTHVDSEGRAAMVDVGRKPDTERVAVAS AVVLLGPVAFKLVQQNQLKKGDALVVAQLAGVQAAKVTSQ LIPLCHHVALSHIQVQLELDSTRHAVKIQASCRARGPTGVEME ALTSAAVAALTLYDMCKAVSRDIVLEEIKLISKTGGQRGDFH RA	MOCS1
363	MENGYTYEDYKNTAEWLLSHTKHRPQVAIICGSGLGGLTDK LTQAQIFDYGEIPNFPRSTVPGHAGRLVFGFLNGRACVMMQG	PNP

[0905]

	RFHMYEGYPLWKVTFPVRVFHLLGVDTLVVTNAAGGLNPKF EVGDIMLIRDHINLPGFSGQNPLRGPNDERFGDRFPAMSDAY DRTMRQRALSTWKQMGEQRELQEGTYVMVAGPSFETVAEC RVLQKLGADAVGMSTVPEVIVARHCGLRVFGFSLITNKVIMD YESLEKANHEEVLAAGKQAAQKLEQFVSILMASIPLPDKAS	
364	MTADKLVFFVNGRKVVEKNADPETTLLAYLRRKLGLSGTKL GCGEGGCGACTVMLSKYDRLQNKIVHFSANACLAPICSLHHV AVTTVEGIGSTKTRLHPVQERIAKSHGSQCGFCTPGIVMSMYT LLRNQPEPTMEEIENAFQGNLCRCTGYRPILQGFRTFARDGGC CGGDGNNPNCCMNQKKDHSVSLSPSLFKPEEFTPLDPTQEPIF PPELLRLKDTPRKQLRFEGERVTWIQASTLKELLDLKAQHPD AKLVVGNTEIGIEMKFKNMLFPMIVCPAWIPELNSVEHGPDGI SFGAACPLSIVEKTLVDAVAKLPAQKTEVFRGVLEQLRWFAG KQVKSVASVGGNIITASPISDLNPVFMASGAKLTLVSRGTRRT VQMDHTFFPGYRKTLLSPEEILLSIEIPYSREGEYFSAFKQASR REDDIAKVTSGMRVLFKPGTTEVQELALCYGGMANRTISALK TTQRQLSKLWKEELLQDVCAGLAEELHLPPDAPGGMVDFRC TLTLSFFFKFYLTVLQKLGQENLEDKCGKLDPTFASATLLFQK DPPADVQLFQEVPKGQSEEDMVGRPLPHLAADMQASGEAVY CDDIPRYENELSLRLVTSTRAHAKIKSIDTSEAKKVPGFVCFIS ADDVPGSNITGICNDETVFAKDKVTCVGHIIGAVVADTPEHT QRAAQGVKITYEELPAIITIEDAIKNNSFYGPELKIEKGDLKKG FSEADNVVSGEIYIGGQEHFYLETHCTIAVPKGEAGEMELFVS TQNTMKTQSFVAKMLGVPANRIVVRVKRMGGGFGGKETRST VVSTAVALAAYKTGRPVRCMLDRDEDMLITGGR HPFLARYKVGFMKTGTVVALEVDHFSNVGNTQDLSQSIMER ALFHMDNCYKIPNIRGTGRLCKTNLPSNTAFRGFGGPQGMLI AECWMSEVAVTCGMPAEEVRRKNLYKEGDLTHFNQKLEGF TLPRCWEECLASSQYHARKSEVDKFNKENCWKKRGLCIIPTK FGISFTVPFLNQAGALLHVYTDGSVLLTHGGTEMGQGLHTK MVQVASRALKIPTSKIYISETSTNTVPNTSPTAASVSADLNGQ AVYAACQTILKRLEPYKKKNPSGSWEDWVTAAYMDTVSLSA TGFYRTPNLGYSFETNSGNPFHYFSYGVACSEVEIDCLTGDHK NLRTDIVMDVGSSLNPAIDIGQVEGAFVQGLGLFTLEELHYSP EGSLHTRGPSTYKIPAFGSIPIEFRVSLLRDCPNKKAIYASKAV GEPPLFLAASIFFAIKDAIRAARAQHTGNNVKELFRLDSPATPE KIRNACVDKFTTLCVTGVPENCKPWSVRV	XDH
365	MLLLHRAVVLRLQQACRLKSIPSRICIQACSTNDSFQPQRPSLT FSGDNSSTQGWRVMGTLLGLGAVLAYQDHRCRAAQESTHIY TKEEVSSHTSPETGIWVTLGSEVFDVTEFVDLHPGGPSKLMLA AGGPLEPFWALYAVHNQSHVRELLAQYKIGELNPEDKVAPT VETSDPYADDPVRHPALKVNSQRPFNAEPPPELLTENYITPNPI FFTRNHLPVPNLDPDTYRLHVVGAPGGQSLSLSLDDLHNFPR YEITVTLQCAGNRRSEMTQVKEVKGLEWRTGAISTARWAGA RLCDVLAQAGHQLCETEAHVCFEGLDSDPTGTAYGASIPLAR AMDPEAEVLLAYEMNGQPLPRDHGFPVRVVVPGVVGARHV KWLGRVSVQPEESYSHWQRRDYKGFSPSVDWETVDFDSAPSI QELPVQSAITEPRDGETVESGEVTIKGYAWSGGGRAVIRVDV SLDGGLTWQVAKLDGEEQRPRKAWAWRLWQLKAPVPAGQ KELNIVCKAVDDGYNVQPDTVAPIWNLRGVLSNAWHRVHV YVSP	SUOX
366	MFHLRTCAAKLRPLTASQTVKTFSQNRPAAARTFQQIRCYSA PVAAEPFLSGTSSNYVEEMYCAWLENPKSVHKSWDIFFRNTN AGAPPGTAYQSPLPLSRGSLAAVAHAQSLVEAQPNVDKLVED	OGDH

[0906]

	HLAVQSLIRAYQIRGHHVAQLDPLGILDADLDSSVPADIISSTD KLGFYGLDESDLDKVFHLPTTTFIGGQESALPLREIIRRLEMAY CQHIGVEFMFINDLEQCQWIRQKFETPGIMQFTNEEKRTLLAR LVRSTRFEEFLQRKWSSEKRFGLEGCEVLIPALKTIIDKSSENG VDYVIMGMPHRGRLNVLANVIRKELEQIFCQFDSKLEAADEG SGDVKYHLGMYHRRINRVTDRNITLSLVANPSHLEAADPVV MGKTKAEQFYCGDTEGKKVMSILLHGDAAFAGQGIVYETFH LSDLPSYTTHGTVHVVVNNQIGFTTDPRMARSSPYPTDVARV VNAPIFHVNSDDPEAVMYVCKVAAEWRSTFHKDVVVDLVC YRRNGHNEMDEPMFTQPLMYKQIRKQKPVLQKYAELLVSQG VVNQPEYEEEISKYDKICEEAFARSKDEKILHIKHWLDSPWPG FFTLDGQPRSMSCPSTGLTEDILTHIGNVASSVPVENFTIHGGL SRILKTRGEMVKNRTVDWALAEYMAFGSLLKEGIHIRLSGQD VERGTFSHRHHVLHDQNVDKRTCIPMNHLWPNQAPYTVCNS SLSEYGVLGFELGFAMASPNALVLWEAQFGDFHNTAQCIIDQ FICPGQAKWVRQNGIVLLLPHGMEGMGPEHSSARPERFLQMC NDDPDVLPDLKEANFDINQLYDCNWVVVNCSTPGNFFHVLR RQILLPFRKPLIIFTPKSLLRHPEARSSFDEMLPGTHFQRVIPED GPAAQNPENVKRLLFCTGKVYYDLTRERKARDMVGQVAITR IEQLSPFPFDLLLKEVQKYPNAELAWCQEEHKNQGYYDYVKP RLRTTISRAKPVWYAGRDPAAAPATGNKKTHLTELQRLLDTA FDLDVFKNFS	
367	MVGYDPKPDGRNNTKFQVAVAGSVSGLVTRALISPFDVIKIR FQLQHERLSRSDPSAKYHGILQASRQILQEEGPTAFWKGHVP AQILSIGYGAVQFLSFEMLTELVHRGSVYDAREFSVHFVCGG LAACMATLTVHPVDVLRTRFAAQGEPKVYNTLRHAVGTMY RSEGPQVFYKGLAPTLIAIFPYAGLQFSCYSSLKHLYKWAIPA EGKKNENLQNLLCGSGAGVISKTLTYPLDLFKKRLQVGGFEH ARAAFGQVRRYKGLMDCAKQVLQKEGALGFFKGLSPSLLKA ALSTGFMFFSYEFFCNVFHCMNRTASQR	SLC25A19
368	MASATAAAARRGLGRALPLFWRGYQTERGVYGYRPRKPESR EPQGALERPPVDHGLARLVTVYCEHGHKAAKINPLFTGQALL ENVPEIQALVQTLQGPFHTAGLLNMGKEEASLEEVLVYLNQI YCGQISIETSQLQSQDEKDWFAKRFEELQKETFTTEERKHLSK LMLESQEFDHFLATKFSTVKRYGGEGAESMMGFFHELLKMS AYSGITDVIIGMPHRGRLNLLTGLLQFPPELMFRKMRGLSEFP ENFSATGDVLSHLTSSVDLYFGAHHPLHVTMLPNPSHLEAVN PVAVGK TRGRQQSRQDGDYSPDNSAQPGDRVICLQVHGDASFCGQGIV PETFTLSNLPHFRIGGSVHLIVNNQLGYTTPAERGRSSLYCSDI GKLVGCAIIHVNGDSPEEVVRATRLAFEYQRQFRKDVIIDLLC YRQWGHNELDEPFYTNPIMYKIIRARKSIPDTYAEHLIAGGLM TQEEVSEIKSSYYAKLNDHLNNMAHYRPPALNLQAHWQGLA QPEAQITTWSTGVPLDLLRFVGMKSVEVPRELQMHSHLLKTH VQSRMEKMMDGIKLDWATAEALALGSLLAQGFNVRLSGQD VGRGT FSQRHAIVVCQETDDTYIPLNHMDPNQKGFLEVSNSPLSEEAV LGFEYGMSIESPKLLPLWEAQFGDFFNGAQIIFDTFISGGEAK WLLQSGIVILLPHGYDGAGPDHSSCRIERFLQMCDSAEEGVD GDTVNMFVVHPTTPAQYFHLLRRQMVRNFRKPLIVASPKML LRLPAAVSTLQEMAPGTTFNPVIGDSSVDPKKVKTLVFCSGK HFYSLVKQRESLGAKKHDFAIIRVEELCPFPLDSLQQEMSKYK HVKDHIWSQEEPQNMGPWSFVSPRFEKQLACKLRLVGRPPLP VPAV	DHTKD1

[0907]

	GIGTVHLHQHEDILAKTFA	
369	MASALSYVSKFKSFVILFVTPLLLLPLVILMPAKFVRCAYVIIL MAIYWCTEVIPLAVTSLMPVLLFPLFQILDSRQVCVQYMKDT NMLFLGGLIVAVAVERWNLHKRIALRTLLWVGAKPARLMLG FMGVTALLSMWISNTATTAMMVPIVEAILQQMEATSAATEA GLELVDKGKAKELPGSQVIFEGPTLGQQEDQERKRLCKAMTL CICYAASIGGTATLTGTGPNVVLLGQMNELFPDSKDLVNFAS WFAFAFPNMLVMLLFAWLWLQFVYMRFNFKKSWGCGLESK KNEKAALKVLQEEYRKLGPLSFAEINVLICFFLLVILWFSRDP GFMPGWLTVAWVEGETKYVSDATVAIFVATLLFIVPSQKPKF NFRSQTEEERKTPFYPPPLLDWKVTQEKVPWGIVLLLGGGFA LAKGSEASGLSVWMGKQMEPLHAVPPAAITLILSLLVAVFTE CTSNVATTTLFLPIFASMSRSIGLNPLYIMLPCTLSASFAFMLP VATPPNAIVFTYGHLKVADMVKTGVIMNIIGVFCVFLAVNTW GRAIFDLDHFPDWANVTHIET	SLC13A5
370	MYRALRLLARSRPLVRAPAAALASAPGLGGAAVPSFWPPNA ARMASQNSFRIEYDTFGELKVPNDKYYGAQTVRSTMNFKIGG VTERMPTPVIKAFGILKRAAAEVNQDYGLDPKIANAIMKAAD EVAEGKLNDHFPLVVWQTGSGTQTNMNVNEVISNRAIEMLG GELGSKIPVHPNDHVNKSQSSNDTFPTAMHIAAAIEVHEVLLP GLQKLHDALDAKSKEFAQIIKIGRTHTQDAVPLTLGQEFSGYV QQVKYAMTRIKAAMPRIYELAAGGTAVGTGLNTRIGFAEKV AAKVAALTGLPFVTAPNKFEALAAHDALVELSGAMNTTACS LMKIANDIRFLGSGPRSGLGELILPENEPGSSIMPGKVNPTQCE AMTMVAAQVMGNHVAVTVGGSNGHFELNVFKPMMIKNVL HSARLLGDASVSFTENCVVGIQANTERINKLMNESLMLVTAL NPHIGYDKAAKIAKTAHKNGSTLKETAIELGYLTAEQFDEWV KPKDMLGPK	FH
371	MWRVCARRAQNVAPWAGLEARWTALQEVPGTPRVTSRSGP APARRNSVTTGYGGVRALCGWTPSSGATPRNRLLLQLLGSPG RRYYSLPPHQKVPLPSLSPTMQAGTIARWEKKEGDKINEGDLI AEVETDKATVGFESLEECYMAKILVAEGTRDVPIGAIICITVG KPEDIEAFKNYTLDSSAAPTPQAAPAPTPAATASPPTPSAQAP GSSYPPHMQVLLPALSPTMTMGTVQRWEKKVGEKLSEGDLL AEIETDKATIGFEVQEEGYLAKILVPEGTRDVPLGTPLCIIVEK EADISAFADYRPTEVTDLKPQVPPPTPPPVAAVPPTPQPLAPTP SAPCPATPAGPKGRVFVSPLAKKLAVEKGIDLTQVKGTGPDG RITKKDIDSFVPSKVAPAPAAVVPPTGPGMAPVPTGVFTDIPIS NIRRVIAQRLMQSKQTIPHYYLSIDVNMGEVLLVRKELNKILE GRSKISVNDFIIKASALACLKVPEANSSWMDTVIRQNHVVDV SVAVSTPAGLITPIVFNAHIKGVETIANDVVSLATKAREGKLQ PHEFQGGTFTISNLGMFGIKNFSAIINPPQACILAIGASEDKLVP ADNEKGFDVASMMSVTLSCDHRVVDGAVGAQWLAEFRKYL EKPITMLL	DLAT
372	MAGALVRKAADYVRSKDFRDYLMSTHFWGPVANWGLPIAA INDMKKSPEIISGRMTFALCCYSLTFMRFAYKVQPRNWLLFA CHATNEVAQLIQGGRLIKHEMTKTASA	MPC1
373	MRKMLAAVSRVLSGASQKPASRVLVASRNFANDATFEIKKC DLHRLEEGPPVTTVLTREDGLKYYRMMQTVRRMELKADQL YKQKIIRGFCHLCDGQEACCVGLEAGINPTDHLITAYRAHGFT FTRGLSVREILAELTGRKGGCAKGKGGSMHMYAKNFYGGNG IVGAQVPLGAGIALACKYNGKDEVCLTLYGDGAANQGQIFE AYNMAALWKLPCIFICENNRYGMGTSVERAAASTDYYKRGD	PDHA1

		FIPGLRVDGMDILCVREATRFAAAYCRSGKGPILMELQTYRY HGHSMSDPGVSYRTREEIQEVRSKSDPIMLLKDRMVNSNLAS VEELKEIDVEVRKEIEDAAQFATADPEPPLEELGYHIYSSDPPF EVRGANQWIKFKSVS	
	374	MAAVSGLVRRPLREVSGLLKRRFHWTAPAALQVTVRDAINQ GMDEELERDEKVFLLGEEVAQYDGAYKVSRGLWKKYGDKR IIDTPISEMGFAGIAVGAAMAGLRPICEFMTFNFSMQAIDQVIN SAAKTYYMSGGLQPVPIVFRGPNGASAGVAAQHSQCFAAWY GHCPGLKVVSPWNSEDAKGLIKSAIRDNNPVVVLENELMYG VPFEFPPEAQSKDFLIPIGKAKIERQGTHITVVSHSRPVGHCLE AAAVLSKEGVECEVINMRTIRPMDMETIEASVMKTNHLVTVE GGWPQFG VGAEICARIMEGPAFNFLDAPAVRVTGADVPMPYAKILEDNSI PQVKDIIFAIKKTLNI	PDHB
	375	MAASWRLGCDPRLLRYLVGFPGRRSVGLVKGALGWSVSRG ANWRWFHSTQWLRGDPIKILMPSLSPTMEEGNIVKWLKKEG EAVSAGDALCEIETDKAVVTLDASDDGILAKIVVEEGSKNIRL GSLIGLIVEEGEDWKHVEIPKDVGPPPPVSKPSEPRPSPEPQISIP VKKEHIPGTLRFRLSPAARNILEKHSLDASQGTATGPRGIFTKE DALKLVQLKQTGKITESRPTPAPTATPTAPSPLQATAGPSYPR PVIPPVSTPGQPNAVGTFTEIPASNIRRVIAKRLTESKSTVPHAY ATADCDLGAVLKVRQDLVKDDIKVSVNDFIIKAAAVTLKQM PDVNVSWDGEGPKQLPFIDISVAVATDKGLLTPIIKDAAAKGI QEIADSVKALSKKARDGKLLPEEYQGGSFSISNLGMFGIDEFT AVINPPQACILAVGRFRPVLKLTEDEEGNAKLQQRQLITVTMS SDSRVVDDELATRFLKSFKANLENPIRLA	PDHX
[0908]	376	MPAPTQLFFPLIRNCELSRIYGTACYCHHKHLCCSSSYIPQSRL RYTPHPAYATFCRPKENWWQYTQGRRYASTPQKFYLTPPQV NSILKANEYSFKVPEFDGKNVSSILGFDSNQLPANAPIEDRRSA ATCLQTRGMLLGVFDGHAGCACSQAVSERLFYYIAVSLLPHE TLLEIENAVESGRALLPILQWHKHPNDYFSKEASKLYFNSLRT YWQELIDLNTGESTDIDVKEALINAFKRLDNDISLEAQVGDPN SFLNYLVLRVAFSGATACVAHVDGVDLHVANTGDSRAMLG VQEEDGSWSAVTLSNDHNAQNERELERLKLEHPKSEAKSVV KQDRLLGLLMPFRAFGDVKFKWSIDLQKRVIESGPDQLNDNE YTKFIPPNYHTPPYLTAEPEVTYHRLRPQDKFLVLATDGLWET MHRQDVVRIVGEYLTGMHHQQPIAVGGYKVTLGQMHGLLT ERRTKMSSVFEDQNAATHLIRHAVGNNEFGTVDHERLSKML SLPEELARMYRDDITIIVVQFNSHVVGAYQNQE	PDP1
	377	MLEKFCNSTFWNSSFLDSPEADLPLCFEQTVLVWIPLGYLWL LAPWQLLHVYKSRTKRSSTTKLYLAKQVFVGFLLILAAIELAL VLTEDSGQATVPAVRYTNPSLYLGTWLLVLLIQYSRQWCVQ KNSWFLSLFWILSILCGTFQFQTLIRTLLQGDNSNLAYSCLFFIS YGFQILILIFSAFSENNESSNNPSSIASFLSSITYSWYDSIILKGY KRPLTLEDVWEVDEEMKTKTLVS KFETHMKRELQKARRALQRRQEKSSQQNSGARLPGLNKNQS QSQDALVLEDVEKKKKKSGTKKDVPKSWLMKALFKTFYMV LLKSFLLKLVNDIFTFVSPQLLKLLISFASDRDTYLWIGYLCAI LLFTAALIQSFCLQCYFQLCFKLGVKVRTAIMASVYKKALTLS NLARKEYTVGETVNLMSVDAQKLMDVTNFMHMLWSSVLQI VLSIFFLWRELGPSVLAGVGVMVLVIPINAILSTKSKTIQVKN MKNKDKRLKIMNEILSGIKILKYFAWEPSFRDQVQNLRKKEL KNLLAFS QLQCVVIFVFQLTPVLVSVVTFSVYVLVDSNNILDAQKAFTSI	ABCC2

		TLFNILRFPLSMLPMMISSMLQASVSTERLEKYLGGDDLDTSA IRHDCNFDKAMQFSEASFTWEHDSEATVRDVNLDIMAGQLV AVIGPVGSGKSSLISAMLGEMENVHGHITIKGTTAYVPQQSWI QNGTIKDNILFGTEFNEKRYQQVLEACALLPDLEMLPGGDLA EIGEKGINLSGGQKQRISLARATYQNLDIYLLDDPLSAVDAHV GKHIFNKVLGPNGLLKGKTRLLVTHSMHFLPQVDEIVVLGNG TIV EKGSYSALLAKKGEFAKNLKTFLRHTGPEEEATVHDGSEEED DDYGLISSVEEIPEDAASITMRRENSFRRTLSRSSRSNGRHLKS LRNSLKTRNVNSLKEDEELVKGQKLIKKEFIETGKVKFSIYLE YLQAIGLFSIFFIILAFVMNSVAFIGSNLWLSAWTSDSKIFNSTD YPASQRDMRVGVYGALGLAQGIFVFIAHFWSAFGFVHASNIL HKQLLNNILRAPMRFFDTTPTGRI VNRFAGDISTVDDTLPQSLRSWITCFLGIISTLVMICMATPVFTI IVIPLGIIYVSVQMFYVSTSRQLRRLDSVTRSPIYSHFSETVSGL PVIRAFEHQQRFLKHNEVRIDTNQKCVFSWITSNRWLAIRLEL VGNLTVFFSALMMVIYRDTLSGDTVGFVLSNALNITQTLNWL VRMTSEIETNIVAVERITEYTKVENEAPWVTDKRPPPDWPSK GKIQFNNYQVRYRPELDLVLRGI TCDIGSMEKIGVVGRTGAGKSSLTNCLFRILEAAGGQIIIDGVD IASIGLHDLREKLTIIPQDPILFSGSLRMNLDPFNNYSDEEIWKA LELAHLKSFVASLQLGLSHEVTEAGGNLSIGQRQLLCLGRALL RKSKILVLDEATAAVDLETDNLIQTTIQNEFAHCTVITIAHRLH TIMDSDKVMVLDNGKIIECGSPEELLQIPGPFYFMAKEAGIEN VNSTKF	
[0909]	378	MDQNQHLNKTAEAQPSENKKTRYCNGLKMFLAALSLSFIAK TLGAIIMKSSIIHIERRFEISSSLVGFIDGSFEIGNLLVIVFVSYFG SKLHRPKLIGIGCFIMGIGGVLTALPHFFMGYYRYSKETNINSS ENSTSTLSTCLINQILSLNRASPEIVGKGCLKESGSYMWIYVFM GNMLRGIGETPIVPLGLSYIDDFAKEGHSSLYLGILNAIAMIGPI IGFTLGSLFSKMYVDIGYVDLSTIRITPTDSRWVGAWWLNFLV SGLFSIISSIPFFFLPQTPNKPQKERKASLSLHVLETNDEKDQTA NLTNQGKNITKNVTGFFQSFKSILTNPLYVMFVLLTLLQVSSY IGAFTYVFKYVEQQYGQPSSKANILLGVITIPIFASGMFLGGYII KKFKLNTVGIAKFSCFTAVMSLSFYLLYFFILCENKSVAGLTM TYDGNNPVTSHRDVPLSYCNSDCNCDESQWEPVCGNNGITYI SPCLAGCKSSSGNKKPIVFYNCSCLEVTGLQNRNYSAHLGECP RDDACTRKFYFFVAIQVLNLFFSALGGTSHVMLIVKIVQPELK SLALGFHSMVIRALGGILAPIYFGALIDTTCIKWSTNNCGTRGS CRTYNSTSFSRVYLGLSSMLRVSSLVLYIILIYAMKKKYQEKD INASENGSVMDEANLESLNKNKHFVPSAGADSETHC	SLCO1B1
	379	MDQHQHLNKTAESASSEKKKTRRCNGFKMFLAALSFSYIAK ALGGIIMKISITQIERRFDISSSLAGLIDGSFEIGNLLVIVFVSYF GSKLHRPKLIGIGCLLMGTGSILTSLPHFFMGYYRYSKETHINP SENSTSSLSTCLINQTLSFNGTSPEIVEKDCVKESGSHMWIYVF MGNMLRGIGETPIVPLGISYIDDFAKEGHSSLYLGSLNAIGMIG PVIGFALGSLFAKMYVDIGYVDLSTIRITPKDSRWVGAWWLG FLVSGLFSIISSIPFFFLPKNPNKPQKERKISLSLHVLKTNDDRN QTANLTNQGKNVTKNVTGFFQSLKSILTNPLYVIFLLLTLLQV SSFIGSFTYVFKYMEQQYGQSASHANFLLGIITIPTVATGMFLG GFIIKKFKLSLVGIAKFSFLTSMISFLFQLLYFPLICESKSVAGLT LTYDGNNSVASHVDVPLSYCNSECNCDESQWEPVCGNNGIT YLSPCLAGCKSSSGIKKHTVFYNCSCVEVTGLQNRNYSAHLG ECPRDNTCTRKFFIYVAIQVINSLFSATGGTTFILLTVKIVQPEL	SLCO1B3

[0910]

	KALAMGFQSMVIRTLGGILAPIYFGALIDKTCMKWSTNSCGA QGACRIYNSVFFGRVYLGLSIALRFPALVLYIVFIFAMKKKFQ GKDTKASDNERKVMDEANLEFLNNGEHFVPSAGTDSKTCNL DMQDNAAAN	
380	MGEPGQSPSPRSSHGSPPTLSTLTLLLLLCGHAHSQCKILRCN AEYVSSTLSLRGGGSSGALRGGGGGGRGGGVGSGGLCRALR SYALCTRRTARTCRGDLAFHSAVHGIEDLMIQHNCSRQGPTA PPPPRGPALPGAGSGLPAPDPCDYEGRFSRLHGRPPGFLHCAS FGDPHVRSFHHHFHTCRVQGAWPLLDNDFLFVQATSSPMAL GANATATRKLTIIFKNMQECIDQKVYQAEVDNLPVAFEDGSI NGGDRPGGSSLSIQTANPGNHVEIQAAYIGTTIIIRQTAGQLSF SIKVAEDVAMAFSAEQDLQLCVGGCPPSQRLSRSERNRRGAI TIDTARRLCKEGLPVEDAYFHSCVFDVLISGDPNFTVAAQAAL EDARAFLPDLEKLHLFPSDAGVPLSSATLLAPLLSGLFVLWLC IQ	HFE2
381	MHQRHPRARCPPLCVAGILACGFLLGCWGPSHFQQSCLQALE PQAVSSYLSPGAPLKGRPPSPGFQRQRQRRAAGGILHLELL VAVGPDVFQAHQEDTERYVLTNLNIGAELLRDPSLGAQFRVH LVKMVILTEPEGAPNITANLTSSLLSVCGWSQTINPEDDTDPG HADLVLYITRFDLELPDGNRQVRGVTQLGGACSPTWSCLITE DTGFDLGVTIAHEIGHSFGLEHDGAPGSGCGPSGHVMASDGA APRAGLAWSPCSRRQLLSLLSAGRARCVWDPPRPQPGSAGHP PDAQPGLYYSANEQCRVAFGPKAVACTFAREHLDMCQALSC HTDPLDQSSCSRLLVPLLDGTECGVEKWCSKGRCRSLVELTPI AAVHGRWSSWGPRSPCSRSCGGGVVTRRRQCNNPRPAFGGR ACVGADLQAEMCNTQACEKTQLEFMSQQCARTDGQPLRSSP GGASFYHWGAAVPHSQGDALCRHMCRAIGESFIMKRGDSFL DGTRCMPSGPREDGTLSLCVSGSCRTFGCDGRMDSQQVWDR CQVCGGDNSTCSPRKGSFTAGRAREYVTFLTVTPNLTSVYIA NHRPLFTHLAVRIGGRYVVAGKMSISPNTTYPSLLEDGRVEY RVALTEDRLPRLEEIRIWGPLQEDADIQVYRRYGEEYGNLTRP DITFTYFQPKPRQAWVWAAVRGPCSVSCGAGLRWVNYSCLD QARKELVETVQCQGSQQPPAWPEACVLEPCPPYWAVGDFGP CSASCGGGLRERPVRCVEAQGSLLKTLPPARCRAGAQQPAVA LETCNPQPCPARWEVSEPSSCTSAGGAGLALENETCVPGADG LEAPVTEGPGSVDEKLPAPEPCVGMSCPPGWGHLDATSAGEK APSPWGSIRTGAQAAHVWTPAAGSCSVSCGRGLMELRFLCM DSALRVPVQEELCGLASKPGSRREVCQAVPCPARWQYKLAA CSVSCGRGVVRRILYCARAHGEDDGEEILLDTQCQGLPRPEP QEACSLEPCPPRWKVMSLGPCSASCGLGTARRSVACVQLDQ GQDVEVDEAACAALVRPEASVPCLIADCTYRWHVGTWMEC SVSCGDGIQRRRDTCLGPQAQAPVPADFCQHLPKPVTVRGC WAGPCVGQGTPSLVPHEEAAAPGRTTATPAGASLEWSQARG LLFSPAPQPRRLLPGPQENSVQSSACGRQHLEPTGTIDMRGPG QADCAVAIGRPLGEVVTLRVLESSLNCSAGDMLLLWGRLTW RKMCRKLLDMTFSSKTNTLVVRQRCGRPGGGVLLRYGSQLA PETFYRECDMQLFGPWGEIVSPSLSPATSNAGGCRLFINVAPH ARIAIHALATNMGAGTEGANASYILIRDTHSLRTTAFHGQQVL YWESESSQAEMEFSEGFLKAQASLRGQYWTLQSWVPEMQDP QSWKGKEGT	ADAMTS13
382	MSRPLSDQEKRKQISVRGLAGVENVTELKKNFNRHLHFTLVK DRNVATPRDYYFALAHTVRDHLVGRWIRTQQHYYEKDPKRI YYLSLEFYMGRTLQNTMVNLALENACDEATYQLGLDMEELE EIEEDAGLGNGGLGRLAACFLDSMATLGLAAYGYGIRYEFGI	PYGM

		FNQKISGGWQMEEADDWLRYGNPWEKARPEFTLPVHFYGH VEHTSQGAKWVDTQVVLAMPYDTPVPGYRNNVVNTMRLW SAKAPNDFNLKDFNVGGYIQAVLDRNLAENISRVLYPNDNFF EGKELRLKQEYFVVAATLQDIIRRFKSSKFGCRDPVRTNFDAF PDKVAIQLNDTHPSLAIPELMRILVDLERMDWDKAWDVTVR TCAYTNHTVLPEALERWPVHLLETLLPRHLQIIYEINQRFLNR VAAAFPGDVDRLRRMSLVEEGAVKRINMAHLCIAGSHAVNG VARIHSEILKKTIFKDFYELEPHKFQNKTNGITPRRWLVLCNP GLAEVIAERIGEDFISDLDQLRKLLSFVDDEAFIRDVAKVKQE NKLKFAAYLEREYKVHINPNSLFDIQVKRIHEYKRQLLNCLH VITLYNRIKREPNKFFVPRTVMIGGKAAPGYHMAKMIIRLVTA IGDVVNHDPAVGDRLRVIFLENYRVSLAEKVIPAADLSEQIST AGTEASGTGNMKFMLNGALTIGTMDGANVEMAEEAGEENFF IFGMRVEDVDKLDQRGYNAQEYYDRIPELRQVIEQLSSGFFSP KQPDLFKDIVNMLMHHDRFKVFADYEDYIKCQEKVSALYKN PREWTRMVIRNIATSGKFSSDRTIAQYAREIWGVEPSRQRLPA PDEAI	
[0911]	383	MLSFVDTRTLLLLAVTLCLATCQSLQEETVRKGPAGDRGPRG ERGPPGPPGRDGEDGPTGPPGPPGPPGPPGLGGNFAAQYDGK GVGLGPGPMGLMGPRGPPGAAGAPGPQGFQGPAGEPGEPGQ TGPAGARGPAGPPGKAGEDGHPGKPGRPGERGVVGPQGARG FPGTPGLPGFKGIRGHNGLDGLKGQPGAPGVKGEPGAPGENG TPGQTGARGLPGERGRVGAPGPAGARGSDGSVGPVGPAGPIG SAGPPGFPGAPGPKGEIGAVGNAGPAGPAGPRGEVGLPGLSG PVGPPGNP GANGLTGAKGAAGLPGVAGAPGLPGPRGIPGPVGAAGATGA RGLVGEPGPAGSKGESGNKGEPGSAGPQGPPGPSGEEGKRGP NGEAGSAGPPGPPGLRGSPGSRGLPGADGRAGVMGPPGSRG ASGPAGVRGPNGDAGRPGEPGLMGPRGLPGSPGNIGPAGKEG PVGLPGIDGRPGPIGPAGARGEPGNIGFPGPKGPTGDPGKNGD KGHAGLAGARGAPGPDGNNGAQGPPGPQGVQGGKGEQGPP GPPGFQGLPGPSGPAGEVGKPGERGLHGEFGLPGPAGPRGER GPPGESGAA GPTGPIGSRGPSGPPGPDGNKGEPGVVGAVGTAGPSGPSGLPG ERGAAGIPGGKGEKGEPGLRGEIGNPGRDGARGAPGAVGAP GPAGATGDRGEAGAAGPAGPAGPRGSPGERGEVGPAGPNGF AGPAGAAGQPGAKGERGAKGPKGENGVVGPTGPVGAAGPA GPNGPPGPAGSRGDGGPPGMTGFPGAAGRTGPPGPSGISGPPG PPGPAGKEGLRGPRGDQGPVGRTGEVGAVGPPGFAGEKGPS GEAGTAGPPGTPGPQGLLGAPGILGLPGSRGERGLPGVAGAV GEPGPLGIAGPPGARGPPGAVGSPGVNGAPGEAGRDGNPGND GPPGRDGQPGHKGERGYPGNIGPVGAAGAPGPHGPVGPAGK HGNRGETGPSGPVGPAGAVGPRGPSGPQGIRGDKGEPGEKGP RGLPGLKGHNGLQGLPGIAGHHGDQGAPGSVGPAGPRGPAG PSGPAGKDGRTGHPGTVGPAGIRGPQGHQGPAGPPGPPGPPG PPGVSGGGYDFGYDGDFYRADQPRSAPSLRPKDYEVDATLKS LNNQIETLLTPEGSRKNPARTCRDLRLSHPEWSSGYYWIDPNQ GCTMDAIKVYCDFSTGETCIRAQPENIPAKNWYRSSKDKKHV WLGETINAGSQFEYNVEGVTSKEMATQLAFMRLLANYASQN ITYHCKNSIAYMDEETGNLKKAVILQGSNDVELVAEGNSRFT YTVLVDGCSKKTNEWGKTIIEYKTNKPSRLPFLDIAPLDIGGA DQEFFVDIGPVCFK	COL1A2
	384	MNNLLCCALVFLDISIKWTTQETFPPKYLHYDEETSHQLLCD KCPPGTYLKQHCTAKWKTVCAPCPDHYYTDSWHTSDECLYC	TNFRSF11B

[0912]

	SPVCKELQYVKQECNRTHNRVCECKEGRYLEIEFCLKHRSCP PGFGVVQAGTPERNTVCKRCPDGFFSNETSSKAPCRKHTNCS VFGLLLTQKGNATHDNICSGNSESTQKCGIDVTLCEEAFFRFA VPTKFTPNWLSVLVDNLPGTKVNAESVERIKRQHSSQEQTFQ LLKLWKHQNKDQDIVKKIIQDIDLCENSVQRHIGHANLTFEQL RSLMESLPGKKVGAEDIEKTIKACKPSDQILKLLSLWRIKNGD QDTLKGLMHALKHSKTYHFPKTVTQSLKKTIRFLHSFTMYKL YQKLFLEMIGNQVQSVKISCL	
385	MAQQANVGELLAMLDSPMLGVRDDVTAVFKENLNSDRGPM LVNTLVDYYLETSSQPALHILTTLQEPHDKHLLDRINEYVGKA ATRLSILSLLGHVIRLQPSWKHKLSQAPLLPSLLKCLKMDTDV VVLTTGVLVLITMLPMIPQSGKQHLLDFFDIFGRLSSWCLKKP GHVAEVYLVHLHASVYALFHRLYGMYPCNFVSFLRSHYSMK ENLETFEEVVKPMMEHVRIHPELVTGSKDHELDPRRWKRLET HDVVIECAKISLDPTEASYEDGYSVSHQISARFPHRSADVTTSP YADT QNSYGCATSTPYSTSRLMLLNMPGQLPQTLSSPSTRLITEPPQ ATLWSPSMVCGMTTPPTSPGNVPPDLSHPYSKVFGTTAGGKG TPLGTPATSPPPAPLCHSDDYVHISLPQATVTPPRKEERMDSA RPCLHRQHHLLNDRGSEEPPGSKGSVTLSDLPGFLGDLASEED SIEKDKEEAAISRELSEITTAEAEPVVPRGGFDSPFYRDSLPGS QRKTHSAASSSQGASVNPEPLHSSL DKLGPDTPKQAFTPIDLPCGSADESPAGDRECQTSLETSIFTPS PCKIPPPTRVGFGSGQPPPYDHLFEVALPKTAHHFVIRKTEELL KKAKGNTEEDGVPSTSPMEVLDRLIQQGADAHSKELNKLPLP SKSVDWTHFGGSPPSDEIRTLRDQLLLLHNQLLYERFKRQQH ALRNRRLLRKVIKAAALEEHNAAMKDQLKLQEKDIQMWKV SLQKEQARYNQLQEQRDTMVTKLHSQIRQLQHDREEFYNQS QELQTKLEDCRNMIAELRIELKKANNKVCHTELLLSQVSQKL SNSESVQQQMEFLNRQLLVLGEVNELYLEQLQNKHSDTTKE VEMMKAAYRKELEKNRSHVLQQTQRLDTSQKRILELESHLA KKDHLLLEQKKYLEDVKLQARGQLQAAESRYEAQKRITQVF ELEILDLYGRLEKDGLLKKLEEEKAEAAEAAEERLDCCNDGC SDSMVGHNEEASGHNGETKTPRPSSARGSSGSRGGGGSSSSSS ELSTPEKPPHQRAGPFSSRWETTMGEASASIPTTVGSLPSSKSF LGMKARELFRNKSESQCDEDGMTSSLSESLKTELGKDLGVEA KIPLNLDGPHPSPPTPDSVGQLHIMDYNETHHEHS	TSC1
386	MAKPTSKDSGLKEKFKILLGLGTPRPNPRSAEGKQTEFIITAEI LRELSMECGLNNRIRMIGQICEVAKTKKFEEHAVEALWKAVA DLLQPERPLEARHAVLALLKAIVQGQGERLGVLRALFFKVIK DYPSNEDLHERLEVFKALTDNGRHITYLEEELADFVLQWMD VGLSSEFLLVLVNLVKFNSCYLDEYIARMVQMICLLCVRTAS SVDIEVSLQVLDAVVCYNCLPAESLPLFIVTLCRTINVKELCEP CWKLMRNLLGTHLGHSAIYNMCHLMEDRAYMEDAPLLRGA VFFVGMALWGAHRLYSLRNSPTSVLPSFYQAMACPNEVVSY EIVLSITRLIKKYRKELQVVAWDILLNIIERLLQQLQTLDSPELR TIVHDLLTTVEELCDQNEFHGSQERYFELVERCADQRPESSLL NLISYRAQSIHPAKDGWIQNLQALMERFFRSESRGAVRIKVLD VLSFVLLINRQFYEEELINSVVISQLSHIPEDKDHQVRKLATQL LVDLAEGCHTHHFNSLLDIIEKVMARSLSPPPELEERDVAAYS ASLEDVKTAVLGLLVILQTKLYTLPASHATRVYEMLVSHIQL HYKHSYTLPIASSIRLQAFDFLLLLRADSLHRLGLPNKDGVVR FSPYCVCDYMEPERGSEKKTSGPLSPPTGPPGPAPAGPAVRLG SVPYSLLFRVLLQCLKQESDWKVLKLVLGRLPESLRYKVLIFT	TSC2

		SPCSVDQLCSALCSMLSGPKTLERLRGAPEGFSRTDLHLAVVP VLTALISYHNYLDKTKQREMVYCLEQGLIHRCASQCVVALSI CSVEMPDIIIKALPVLVVKLTHISATASMAVPLLEFLSTLARLP HLYRNFAAEQYASVFAISLPYTNPSKFNQYIVCLAHHVIAMW FIRCRLPFRKDFVPFITKGLRSNVLLSFDDTPEKDSFRARSTSL NERPKSLRIARPPKQGLNNSPPVKEFKESSAAEAFRCRSISVSE HVVRSRIQTSLTSASLGSADENSVAQADDSLKNLHLELTETCL DMMARYVFSNFTAVPKRSPVGEFLLAGGRTKTWLVGNKLVT VTTSVGTGTRSLLGLDSGELQSGPESSSSPGVHVRQTKEAPAK LESQAGQQVSRGARDRVRSMSGGHGLRVGALDVPASQFLGS ATSPGPRTAPAAKPEKASAGTRVPVQEKTNLAAYVPLLTQG WAEILVRRPTGNTSWLMSLENPLSPFSSDINNMPLQELSNAL MAAERFKEHRDTALYKSLSVPAASTAKPPPLPRSNTVASFSSL YQSSCQGQLHRSVSWADSAVVMEEGSPGEVPVLVEPPGLED V EAALGMDRRTDAYSRSSSVSSQEEKSLHAEELVGRGIPIERVV SSEGGRPSVDLSFQPSQPLSKSSSSPELQTLQDILGDPGDKADV GRLSPEVKARSQSGTLDGESAAWSASGEDSRGQPEGPLPSSSP RSPSGLRPRGYTISDSAPSRRGKRVERDALKSRATASNAEKVP GINPSFVFLQLYHSPFFGDESNKPILLPNESQSFERSVQLLDQIP SYDTHKIAVLYVGEGQSNSELA ILSNEHGSYRYTEFLTGLGRLIELKDCQPDKVYLGGLDVCGE DGQFTYCWHDDIMQAVFHIATLMPTKDVDKHRCDKKRHLG NDFVSIVYNDSGEDFKLGTIKGQFNFVHVIVTPLDYECNLVSL QCRKDMEGLVDTSVAKIVSDRNLPFVARQMALHANMASQV HHSRSNPTDIYPSKWIARLRHIKRLRQRICEEAAYSNPSLPLVH PPSHSKAPAQTPAEPTPGYEVGQRKRLISSVEDFTEFV	
[0913]	387	MAAKSQPNIPKAKSLDGVTNDRTASQGQWGRAWEVDWFSL ASVIFLLLFAPFIVYYFIMACDQYSCALTGPVVDIVTGHARLS DIWAKTPPITRKAAQLYTLWVTFQVLLYTSLPDFCHKFLPGY VGGIQEGAVTPAGVVNKYQINGLQAWLLTHLLWFANAHLLS WFSPTIIFDNWIPLLWCANILGYAVSTFAMVKGYFFPTSARDC KFTGNFFYNYMMGIEFNPRIGKWFDFKLFFNGRPGIVAWTLI NLSFAAKQRELHSHVTNAMVLVNVLQAIYVIDFFWNETWYL KTIDICHDHFGWYLGWGDCVWLPYLYTLQGLYLVYHPVQLS TPHAVGVLLLGLVGYYIFRVANHQKDLFRRTDGRCLIWGRKP KVIECSYTSADGQRHHSKLLVSGFWGVARHFNYVGDLMGSL AYCLACGGGHLLPYFYIIYMAILLTHRCLRDEHRCASKYGRD WERYTAAVPYRLLPGIF	DHCR7
	388	MSLSNKLTLDKLDVKGKRVVMRVDFNVPMKNNQITNNQRIK AAVPSIKFCLDNGAKSVVLMSHLGRPDGVPMPDKYSLEPVA VELKSLLGKDVLFLKDCVGPEVEKACANPAAGSVILLENLRF HVEEEGKGKDASGNKVKAEPAKIEAFRASLSKLGDVYVNDA FGTAHRAHSSMVGVNLPQKAGGFLMKKELNYFAKALESPER PFLAILGGAKVADKIQLINNMLDKVNEMIIGGGMAFTFLKVL NNMEIGTSLFDEEGAKIVKDLMSKAEKNGVKITLPVDFVTAD KFDENAKTGQATVASGIPAGWMGLDCGPESSKKYAEAVTRA KQIVWNGPVGVFEWEAFARGTKALMDEVVKATSRGCITIIGG GDTATCCAKWNTEDKVSHVSTGGGASLELLEGKVLPGVDAL SNI	PGK1
	389	MGTSALWALWLLLALCWAPRESGATGTGRKAKCEPSQFQCT NGRCITLLWKCDGDEDCVDGSDEKNCVKKTCAESDFVCNNG QCVPSRWKCDGDPDCEDGSDESPEQCHMRTCRIHEISCGAHS TQCIPVSWRCDGENDCDSGEDEENCGNITCSPDEFTCSSGRCI	VLDLR

[0914]

	SRNFVCNGQDDCSDGSDELDCAPPTCGAHEFQCSTSSCIPISW VCDDDADCSDQSDESLEQCGRQPVIHTKCPASEIQCGSGECIH KKWRCDGDPDCKDGSDEVNCPSRTCRPDQFECEDGSCIHGSR QCNGI RDCVDGSDEVNCKNVNQCLGPGKFKCRSGECIDISKVCNQEQ DCRDWSDEPLKECHINECLVNNGGCSHICKDLVIGYECDCAA GFELIDRKTCGDIDECQNPGICSQICINLKGGYKCECSRGYQM DLATGVCKAVGKEPSLIFTNRRDIRKIGLERKEYIQLVEQLRN TVALDADIAAQKLFWADLSQKAIFSASIDDKVGRHVKMIDNV YNPAAIAVDWVYKTIYWTDAASKTISVATLDGTKRKFLFNSD LREPASIAVDPLSGFVYWSDWGEPAKIEKAGMNGFDRRPLVT ADIQ WPNGITLDLIKSRLYWLDSKLHMLSSVDLNGQDRRIVLKSLE FLAHPLALTIFEDRVYWIDGENEAVYGANKFTGSELATLVNN LNDAQDIIVYHELVQPSGKNWCEEDMENGGCEYLCLPAPQIN DHSPKYTCSCPSGYNVEENGRDCQSTATTVTYSETKDTNTTEI SATSGLVPGGINVTTAVSEVSVPPKGTSAAWAILPLLLLVMA AVGGYLMWRNWQHKNMKSMNFDNPVYLKTTEEDLSIDIGR HSASVGHTYPAISVVSTDDDLA	
390	MEPSSLELPADTVQRIAAELKCHPTDERVALHLDEEDKLRHF RECFYIPKIQDLPPVDLSLVNKDENAIYFLGNSLGLQPKMVKT YLEEELDKWAKIAAYGHEVGKRPWITGDESIVGLMKDIVGA NEKEIALMNALTVNLHLLMLSFFKPTPKRYKILLEAKAFPSDH YAIESQLQLHGLNIEESMRMIKPREGEETLRIEDILEVIEKEGDS IAVILFSGVHFYTGQHFNIPAITKAGQAKGCYVGFDLAHAVG NVELYLHDWGVDFACWCSYKYLNAGAGGIAGAFIHEKHAH TIKPALVGWFGHELSTRFKMDNKLQLIPGVCGFRISNPPILLVC SLHASLEIFKQATMKALRKKSVLLTGYLEYLIKHNYGKDKAA TKKPVVNIITPSHVEERGCQLTITFSVPNKDVFQELEKRGVVC DKRNPNGIRVAPVPLYNSFHDVYKFTNLLTSILDSAETKN	KYNU
391	MFPGCPRLWVLVVLGTSWVGWGSQGTEAAQLRQFYVAAQG ISWSYRPEPTNSSLNLSVTSFKKIVYREYEPYFKKEKPQSTISG LLGPTLYAEVGDIIKVHFKNKADKPLSIHPQGIRYSKLSEGAS YLDHTFPAEKMDDAVAPGREYTYEWSISEDSGPTHDDPPCLT HIYYSHENLIEDFNSGLIGPLLICKKGTLTEGGTQKTFDKQIVL LFAVFDESKSWSQSSSLMYTVNGYVNGTMPDITVCAHDHIS WHLLGMSSGPELFSIHFNGQVLEQNHHKVSAITLVSATSTTA NMTVGPEGKWIISSLTPKHLQAGMQAYIDIKNCPKKTRNLKK ITREQRRHMKRWEYFIAAEEVIWDYAPVIPANMDKKYRSQH LDNFSNQIGKHYKKVMYTQYEDESFTKHTVNPNMKEDGILG PIIRAQVRDTLKIVFKNMASRPYSIYPHGVTFSPYEDEVNSSFT SGRNNTMIRAVQPGETYTYKWNILEFDEPTENDAQCLTRPYY SDVDIMRDIASGLIGLLLICKSRSLDRRGIQRAADIEQQAVFAV FDENKSWYLEDNINKFCENPDEVKRDDPKFYESNIMSTINGY VPESITTLGFCFDDTVQWHFCSVGTQNEILTIHFTGHSFIYGKR HEDTLTLFPMRGESVTVTMDNVGTWMLTSMNSSPRSKKLRL KFRDVKCIPDDDEDSYEIFEPPESTVMATRKMHDRLEPEDEES DADYDYQNRLAAALGIRSFRNSSLNQEEEEFNLTALALENGT EFVSSNTDIIVGSNYSSPSNISKFTVNNLAEPQKAPSHQQATTA GSPLRHLIGKNSVLNSSTAEHSSPYSEDPIEDPLQPDVTGIRLLS LGAGEFKSQEHAKHKGPKVERDQAAKHRFSWMKLLAHKVG RHLSQDTGSPSGMRPWEDLPSQDTGSPSRMRPWKDPPSDLLL LKQSNSSKILVGRWHLASEKGSYEIIQDTDEDTAVNNWLISPQ NASRAWGESTPLANKPGKQSGHPKFPRVRHKSLQVRQDGGK	F5

[0915]

	SRLKKSQFLIKTRKKKKEKHTHHAPLSPRTFHPLRSEAYNTFS ERRLKHSLVLHKSNETSLPTDLNQTLPSMDFGWIASLPDHNQ NSSNDTGQASCPPGLYQTVPPEEHYQTFPIQDPDQMHSTSDPS HRSSSPELSEMLEYDRSHKSFPTDISQMSPSSEHEVWQTVISPD LSQVTLSPELSQTNLSPDLSHTTLSPELIQRNLSPALGQMPISPD LSHTTLSPDLSHTTLSLDLSQTNLSPELSQTNLSPALGQMPLSP DLSHTTLSLDFSQTNLSPELSHMTLSPELSQTNLSPALGQMPIS PDLSHTTLSLDFSQTNLSPELSQTNLSPALGQMPLSPDPSHTTL SLDLSQTNLSPELSQTNLSPDLSEMPLFADLSQIPLTPDLDQMT LSPDLGETDLSPNFGQMSLSPDLSQVTLSPDISDTTLLPDLSQIS PPPDLDQIFYPSESSQSLLLQEFNESFPYPDLGQMPSPSSPTLND TFLSKEFNPLVIVGLSKDGTDYIEIIPKEEVQSSEDDYAEIDYVP YDDPYKTDVRTNINSSRDPDNIAAWYLRSNNGNRRNYYIAAE EISWDYSEFVQRETDIEDSDDIPEDTTYKKVVFRKYLDSTFTK RDPRGEYEEHLGILGPIIRAEVDDVIQVRFKNLASRPYSLHAH GLSYEKSSEGKTYEDDSPEWFKEDNAVQPNSSYTYVWHATE RSGPESPGSACRAWAYYSAVNPEKDIHSGLIGPLLICQKGILH KDSNMPMDMREFVLLFMTFDEKKSWYYEKKSRSSWRLTSSE MKKSHEFHAINGMIYSLPGLKMYEQEWVRLHLLNIGGSQDIH VVHFHGQTLLENGNKQHQLGVWPLLPGSFKTLEMKASKPG WWLLNTEVGENQRAGMQTPFLIMDRDCRMPMGLSTGIISDS QIKASEFLGYWEPRLARLNNGGSYNAWSVEKLAAEFASKPWI QVDMQKEVIITGIQTQGAKHYLKSCYTTEFYVAYSSNQINWQ IFKGNSTRNVMYFNGNSDASTIKENQFDPPIVARYIRISPTRAY NRPTLRLELQGCEVNGCSTPLGMENGKIENKQITASSFKKSW WGDYWEPFRARLNAQGRVNAWQAKANNNKQWLEIDLLKIK KITAIITQGCKSLSSEMYVKSYTIHYSEQGVEWKPYRLKSSMV DKIFEGNTNTKGHVKNFFNPPIISRFIRVIPKTWNQSIALRLELF GCDIY	
392	MGPTSGPSLLLLLLTHLPLALGSPMYSIITPNILRLESEETMVLE AHDAQGDVPVTVTVHDFPGKKLVLSSEKTVLTPATNHMGNV TFTIPANREFKSEKGRNKFVTVQATFGTQVVEKVVLVSLQSG YLFIQTDKTIYTPGSTVLYRIFTVNHKLLPVGRTVMVNIENPE GIPVKQDSLSSQNQLGVLPLSWDIPELVNMGQWKIRAYYENS PQQVFSTEFEVKEYVLPSFEVIVEPTEKFYYIYNEKGLEVTITA RFLYGKKVEGTAFVIFGIQDGEQRISLPESLKRIPIEDGSGEVV LSRKVLLDGVQNPRAEDLVGKSLYVSATVILHSGSDMVQAE RSGIPIVTSPYQIHFTKTPKYFKPGMPFDLMVFVTNPDGSPAY RVPVAVQGEDTVQSLTQGDGVAKLSINTHPSQKPLSITVRTK KQELSEAEQATRTMQALPYSTVGNSNNYLHLSVLRTELRPGE TLNVNFLLRMDRAHEAKIRYYTYLIMNKGRLLKAGRQVREP GQDLVVLPLSITTDFIPSFRLVAYYTLIGASGQREVVADSVWV DVKDSCVGSLVVKSGQSEDRQPVPGQQMTLKIEGDHGARVV LVAVDKGVFVLNKKNKLTQSKIWDVVEKADIGCTPGSGKDY AGVFSDAGLTFTSSSGQQTAQRAELQCPQPAARRRRSVQLTE KRMDKVGKYPKELRKCCEDGMRENPMRFSCQRRTRFISLGE ACKKVFLDCCNYITELRRQHARASHLGLARSNLDEDIIAEENI VSRSEFPESWLWNVEDLKEPPKNGISTKLMNIFLKDSITTWEIL AVSMSDKKGICVADPFEVTVMQDFFIDLRLPYSVVRNEQVEI RAVLYNYRQNQELKVRVELLHNPAFCSLATTKRRHQQTVTIP PKSSLSVPYVIVPLKTGLQEVEVKAAVYHHFISDGVRKSLKV VPEGIRMNKTVAVRTLDPERLGREGVQKEDIPPADLSDQVPD TESETRILLQGTPVAQMTEDAVDAERLKHLIVTPSGCGEQNMI GMTPTVIAVHYLDETEQWEKFGLEKRQGALELIKKGYTQQL	C3

	AFRQPSSAFAAFVKRAPSTWLTAYVVKVFSLAVNLIAIDSQVL CGAVKWLILEKQKPDGVFQEDAPVIHQEMIGGLRNNNEKDM ALTAFVLISLQEAKDICEEQVNSLPGSITKAGDFLEANYMNLQ RSYTVAIAGYALAQMGRLKGPLLNKFLTTAKDKNRWEDPGK QLYNVEATSYALLALLQLKDFDFVPPVVRWLNEQRYYGGGY GSTQATFMVFQALAQYQKDAPDHQELNLDVSLQLPSRSSKIT HRIHWESASLLRSEETKENEGFTVTAEGKGQGTLSVVTMYHA KAKDQLTCNKFDLKVTIKPAPETEKRPQDAKNTMILEICTRYR GDQDATMSILDISMMTGFAPDTDDLKQLANGVDRYISKYELD KAFSDRNTLIIYLDKVSHSEDDCLAFKVHQYFNVELIQPGAVK VYAYYNLEESCTRFYHPEKEDGKLNKLCRDELCRCAEENCFI QKSDDKVTLEERLDKACEPGVDYVYKTRLVKVQLSNDFDEY IMAIEQTIKSGSDEVQVGQQRTFISPIKCREALKLEEKKHYLM WGLSSDFWGEKPNLSYIIGKDTWVEHWPEEDECQDEENQKQ CQDLGAFTESMVVFGCPN	
[0916]	393	MGPRLSVWLLLLPAALLLHEEHSRAAAKGGCAGSGCGKCDC HGVKGQKGERGLPGLQGVIGFPGMQGPEGPQGPPGQKGDTG EPGLPGTKGTRGPPGASGYPGNPGLPGIPGQDGPPGPPGIPGC NGTKGERGPLGPPGLPGFAGNPGPPGLPGMKGDPGEILGHVP GMLLKGERGFPGIPGTPGPPGLPGLQGPVGPPGFTGPPGPPGPP GPPGEKGQMGLSFQGPKGDKGDQGVSGPPGVPGQAQVQEK GDFATKGEKGQKGEPGFQGMPGVGEKGEPGKPGPRGKPGKD GDKGEKGSPGFPGEPGYPGLIGRQGPQGEKGEAGPPGPPGIVI GTGPLGEKGERGYPGTPGPRGEPGPKGFPGLPGQPGPPGLPVP GQAGAPGFPGERGEKGDRGFPGTSLPGPSGRDGLPGPPGSPGP PGQPGYTNGIVECQPGPPGDQGPPGIPGQPGFIGEIGEKGQKG ESCLICDIDGYRGPPGPQGPPGEIGFPGQPGAKGDRGLPGRDG VAGVPGPQGTPGLIGQPGAKGEPGEFYFDLRLKGDKGDPGFP GQPGMPGRAGSPGRDGHPGLPGPKGSPGSVGLKGERGPPGG VGFPGSRGDTGPPGPPGYGPAGPIGDKGQAGFPGGPGSPGLPG PKGEPGKIVPLPGPPGAEGLPGSPGFPGPQGDRGFPGTPGRPG LPGEKGAVGQPGIGFPGPPGPKGVDGLPGDMGPPGTPGRPGF NGLPGNPGVQGQKGEPGVGLPGLKGLPGLPGIPGTPGEKGSIG VPGVPGEHGAIGPPGLQGIRGEPGPPGLPGSVGSPGVPGIGPPG ARGPPGGQGPPGLSGPPGIKGEKGFPGFPGLDMPGPKGDKGA QGLPGITGQSGLPGLPGQQGAPGIPGFPGSKGEMGVMGTPGQ PGSPGPVGAPGLPGEKGDHGFPGSSGPRGDPGLKGDKGDVGL PGKPGSMDKVDMGSMKGQKGDQGEKGQIGPIGEKGSRGDPG TPGVPGKDGQAGQPGQPGPKGDPGISGTPGAPGLPGPKGSVG GMGLPGTPGEKGVPGIPGPQGSPGLPGDKGAKGEKGQAGPPG IGIPGLRGEKGDQGIAGFPGSPGEKGEKGSIGIPGMPGSPGLKG SPGSVGYPGSPGLPGEKGDKGLPGLDGIPGVKGEAGLPGTPGP TGPAGQKGEPGSDGIPGSAGEKGEPGLPGRGFPGFPGAKGDK GSKGEVGFPGLAGSPGIPGSKGEQGFMGPPGPQGQPGLPGSPG HATEGPKGDRGPQGQPGLPGLPGPMGPPGLPGIDGVKGDKG NPGWPGAPGVPGPKGDPGFQGMPGIGGSPGITGSKGDMGPPG VPGFQGPKGLPGLQGIKGDQGDQGVPGAKGLPGPPGPPGPYD IIKGEPGLPGPEGPPGLKGLQGLPGPKGQQGVTGLVGIPGPPGI PGFDGAPGQKGEMGPAGPTGPRGFPGPPGPDGLPGSMGPPGT PSVDHGFLVTRHSQTIDDPQCPSGTKILYHGYSLLYVQGNERA HGQDLGTAGSCLRKFSTMPFLFCNINNVCNFASRNDYSYWLS TPEPMPMSMAPITGENIRPFISRCAVCEAPAMVMAVHSQTIQI PPCPSGWSSLWIGYSFVMHTSAGAEGSGQALASPGSCLEEFRS APFIECHGRGTCNYYANAYSFWLATIERSEMFKKPTPSTLKA	COL4A1

[0917]

	GELRTHVSRCQVCMRRT	
394	MRLLAKIICLMLWAICVAEDCNELPPRRNTEILTGSWSDQTYP EGTQAIYKCRPGYRSLGNVIMVCRKGEWVALNPLRKCQKRP CGHPGDTPFGTFTLTGGNVFEYGVKAVYTCNEGYQLLGEINY RECDTDGWTNDIPICEVVKCLPVTAPENGKIVSSAMEPDREY HFGQAVRFVCNSGYKIEGDEEMHCSDDGFWSKEKPKCVEISC KSPDVINGSPISQKIIYKENERFQYKCNMGYEYSERGDAVCTE SGWRPLPSCEEKSCDNPYIPNGDYSPLRIKHRTGDEITYQCRN GFYPATRGNTAKCTSTGWIPAPRCTLKPCDYPDIKHGGLYHE NMRRPYFPVAVGKYYSYYCDEHFETPSGSYWDHIHCTQDGW SPAVPCLRKCYFPYLENGYNQNYGRKFVQGKSIDVACHPGY ALPKAQTTVTCMENGWSPTPRCIRVKTCSKSSIDIENGFISESQ YTYALKEKAKYQCKLGYVTADGETSGSITCGKDGWSAQPTC IKSCDIPVFMNARTKNDFTWFKLNDTLDYECHDGYESNTGST TGSIVCGYNGWSDLPICYERECELPKIDVHLVPDRKKDQYKV GEVLKFSCKPGFTIVGPNSVQCYHFGLSPDLPICKEQVQSCGP PPELLNGNVKEKTKEEYGHSEVVEYYCNPRFLMKGPNKIQCV DGEWTTLPVCIVEESTCGDIPELEHGWAQLSSPPYYYGDSVEF NCSESFTMIGHRSITCIHGVWTQLPQCVAIDKLKKCKSSNLIIL EEHLKNKKEFDHNSNIRYRCRGKEGWIHTVCINGRWDPEVN CSMAQIQLCPPPPQIPNSHNMTTTLNYRDGEKVSVLCQENYLI QEGEEITCKDGRWQSIPLCVEKIPCSQPPQIEHGTINSSRSSQES YAHGTKLSYTCEGGFRISEENETTCYMGKWSSPPQCEGLPCK SPPEISHGVVAHMSDSYQYGEEVTYKCFEGFGIDGPAIAKCLG EKWSHPPSCIKTDCLSLPSFENAIPMGEKKDVYKAGEQVTYT CATYYKMDGASNVTCINSRWTGRPTCRDTSCVNPPTVQNAYI VSRQMSKYPSGERVRYQCRSPYEMFGDEEVMCLNGNWTEPP QCKDSTGKCGPPPPIDNGDITSFPLSVYAPASSVEYQCQNLYQ LEGNKRITCRNGQWSEPPKCLHPCVISREIMENYNIALRWTAK QKLYSRTGESVEFVCKRGYRLSSRSHTLRTTCWDGKLEYPTC AKR	CFH
395	MEPRPTAPSSGAPGLAGVGETPSAAALAAARVELPGTAVPSV PEDAAPASRDGGGVRDEGPAAAGDGLGRPLGPTPSQSRFQVD LVSENAGRAAAAAAAAAAAAAAAGAGAGAKQTPADGEASG ESEPAKGSEEAKGRFRVNFVDPAASSSAEDSLSDAAGVGVDG PNVSFQNGGDTVLSEGSSLHSGGGGGSGHHQHYYYDTHTNT YYLRTFGHNTMDAVPRIDHYRHTAAQLGEKLLRPSLAELHDE LEKEPFEDGFANGEESTPTRDAVVTYTAESKGVVKFGWIKGV LVRCMLNIWGVMLFIRLSWIVGQAGIGLSVLVIMMATVVTTI TGLSTSAIATNGFVRGGGAYYLISRSLGPEFGGAIGLIFAFANA VAVAMYVVGFAETVVELLKEHSILMIDEINDIRIIGAITVVILL GISVAGMEWEAKAQIVLLVILLLAIGDFVIGTFIPLESKKPKGF FGYKSEIFNENFGPDFREEETFFSVFAIFFPAATGILAGANISGD LADPQSAIPKGTLLAILITTLVYVGIAVSVGSCVVRDATGNVN DTIVTELTNCTSAACKLNFDFSSCESSPCSYGLMNNFQVMSM VSGFTPLISAGIFSATLSSALASLVSAPKIFQALCKDNIYPAFQ MFAKGYGKNNEPLRGYILTFLIALGFILIAELNVIAPIISNFFLA SYALINFSVFHASLAKSPGWRPAFKYYNMWISLLGAILCCIVM FVINWWAALLTYVIVLGLYIYVTYKKPDVNWGSSTQALTYL NALQHSIRLSGVEDHVKNFRPQCLVMTGAPNSRPALLHLVHD FTKNVGLMICGHVHMGPRRQAMKEMSIDQAKYQRWLIKNK MKAFYAPVHADDLREGAQYLMQAAGLGRMKPNTLVLGFKK DWLQADMRDVDMYINLFHDAFDIQYGVVVIRLKEGLDISHL QGQEELLSSQEKSPGTKDVVVSVEYSKKSDLDTSKPLSEKPIT	SLC12A2

		HKVEEEDGKTATQPLLKKESKGPIVPLNVADQKLLEASTQFQ KKQGKNTIDVWWLFDDGGLTLLIPYLLTTKKKWKDCKIRVFI GGKINRIDHDRRAMATLLSKFRIDFSDIMVLGDINTKPKKENII AFEEIIEPYRLHEDDKEQDIADKMKEDEPWRITDNELELYKTK TYRQIRLNELLKEHSSTANIIVMSLPVARKGAVSSALYMAWL EALSKDLPPILLVRGNHQSVLTFYS	
[0918]	396	MAASKKAVLGPLVGAVDQGTSSTRFLVFNSKTAELLSHHQV EIKQEFPREGWVEQDPKEILHSVYECIEKTCEKLGQLNIDISNI KAIGVSNQRETTVVWDKITGEPLYNAVVWLDLRTQSTVESLS KRIPGNNNFVKSKTGLPLSTYFSAVKLRWLLDNVRKVQKAV EEKRALFGTIDSWLIWSLTGGVNGGVHCTDVTNASRTMLFNI HSLEWDKQLCEFFGIPMEILPNVRSSSEIYGLMKISHSVKAGA LEGVPISGCLGDQSAALVGQMCFQIGQAKNTYGTGCFLLCNT GHKCVFSDHGLLTTVAYKLGRDKPVYYALEGSVAIAGAVIR WLRDNLGIIKTSEEIEKLAKEVGTSYGCYFVPAFSGLYAPYWE PSARGIICGLTQFTNKCHIAFAALEAVCFQTREILDAMNRDCGI PLSHLQVDGGMTSNKILMQLQADILYIPVVKPSMPETTALGA AMAAGAAEGVGVWSLEPEDLSAVTMERFEPQINAEESEIRYS TWKKAVMKSMGWVTTQSPESGDPSIFCSLPLGFFIVSSMVML IGARYISGIP	GK
	397	MDVGSKEVLMESPPDYSAAPRGRFGIPCCPVHLKRLLIVVVV VVLIVVVIVGALLMGLHMSQKHTEMVLEMSIGAPEAQQRLA LSEHLVTTATFSIGSTGLVVYDYQQLLIAYKPAPGTCCYIMKI APESIPSLEALNRKVHNFQMECSLQAKPAVPTSKLGQAEGRD AGSAPSGGDPAFLGMAVNTLCGEVPLYYI	SFTPC
	398	MEPGRRGAAALLALLCVACALRAGRAQYERYSFRSFPRDEL MPLESAYRHALDKYSGEHWAESVGYLEISLRLHRLLRDSEAF CHRNCSAAPQPEPAAGLASYPELRLFGGLLRRAHCLKRCKQG LPAFRQSQPSREVLADFQRREPYKFLQFAYFKANNLPKAIAA AHTFLLKHPDDEMMKRNMAYYKSLPGAEDYIKDLETKSYES LFIRAVRAYNGENWRTSITDMELALPDFFKAFYECLAACEGS REIKDFKDFYLSIADHYVEVLECKIQCEENLTPVIGGYPVEKF VATMYHYLQFAYYKLNDLKNAAPCAVSYLLFDQNDKVMQQ NLVYYQYHRDTWGLSDEHFQPRPEAVQFFNVTTLQKELYDF AKENIMDDDEGEVVEYVDDLLELEETS	CRTAP
	399	MAVRALKLLTTLLAVVAAASQAEVESEAGWGMVTPDLLFAE GTAAYARGDWPGVVLSMERALRSRAALRALRLRCRTQCAA DFPWELDPDWSPSPAQASGAAALRDLSFFGGLLRRAACLRRC LGPPAAHSLSEEMELEFRKRSPYNYLQVAYFKINKLEKAVAA AHTFFVGNPEHMEMQQNLDYYQTMSGVKEADFKDLETQPH MQEFRLGVRLYSEEQPQEAVPHLEAALQEYFVAYEECRALCE GPYDYDGYNYLEYNADLFQAITDHYIQVLNCKQNCVTELAS HPSREKPFEDFLPSHYNYLQFAYYNIGNYTQAVECAKTYLLFF PNDEVMNQNLAYYAAMLGEEHTRSIGPRESAKEYRQRSLLE KELLFFAYDVFGIPFVDPDSWTPEEVIPKRLQEKQKSERETAV RISQEIGNLMKEIETLVEEKTKESLDVSRLTREGGPLLYEGISL TMNSKLLNGSQRVVMDGVISDHECQELQRLTNVAATSGDGY RGQTSPHTPNEKFYGVTVFKALKLGQEGKVPLQSAHLYYNV TEKVRRIMESYFRLDTPLYFSYSHLVCRTAIEEVQAERKDDSH PVHVDNCILNAETLVCVKEPPAYTFRDYSAILYLNGDFDGGN FYFTELDAKTVTAEVQPQCGRAVGFSSGTENPHGVKAVTRG QRCAIALWFTLDPRHSERDRVQADDLVKMLFSPEEMDLSQE QPLDAQQGPPEPAQESLSGSESKPKDEL	P3H1

[0919]

	400	MTLRLLVAALCAGILAEAPRVRAQHRERVTCTRLYAADIVFL LDGSSSIGRSNFREVRSFLEGLVLPFSGAASAQGVRFATVQYS DDPRTEFGLDALGSGGDVIRAIRELSYKGGNTRTGAAILHVA DHVFLPQLARPGVPKVCILITDGKSQDLVDTAAQRLKGQGVK LFAVGIKNADPEELKRVASQPTSDFFFFVNDFSILRTLLPLVSR RVCTTAGGVPVTRPPDDSTSAPRDLVLSEPSSQSLRVQWTAA SGPVTGYKVQYTPLTGLGQPLPSERQEVNVPAGETSVRLRGL RPLTEYQVTVIALYANSIGEAVSGTARTTALEGPELTIQNTTA HSLLVAWRSVPGATGYRVTWRVLSGGPTQQQELGPGQGSVL LRDLEPGTDYEVTVSTLFGRSVGPATSLMARTDASVEQTLRP VILGPTSILLSWNLVPEARGYRLEWRRETGLEPPQKVVLPSDV TRYQLDGLQPGTEYRLTLYTLLEGHEVATPATVVPTGPELPV SPVTDLQATELPGQRVRVSWSPVPGATQYRIIVRSTQGVERTL VLPGSQTAFDLDDVQAGLSYTVRVSARVGPREGSASVLTVRR EPETPLAVPGLRVVVSDATRVRVAWGPVPGASGFRISWSTGS GPESSQTLPPDSTATDITGLQPGTTYQVAVSVLRGREEGPAAV IVARTDPLGPVRTVHVTQASSSSVTITWTRVPGATGYRVSWH SAHGPEKSQLVSGEATVAELDGLEPDTEYTVHVRAHVAGVD GPPASVVVRTAPEPVGRVSRLQILNASSDVLRITWVGVTGAT AYRLAWGRSEGGPMRHQILPGNTDSAEIRGLEGGVSYSVRVT ALVGDREGTPVSIVVTTPPEAPPALGTLHVVQRGEHSLRLRW EPVPRAQGFLLHWQPEGGQEQSRVLGPELSSYHLDGLEPATQ YRVRLSVLGPAGEGPSAEVTARTESPRVPSIELRVVDTSIDSVT LAWTPVSRASSYILSWRPLRGPGQEVPGSPQTLPGISSSQRVT GLEPGVSYIFSLTPVLDGVRGPEASVTQTPVCPRGLADVVFLP HATQDNAHRAEATRRVLERLVLALGPLGPQAVQVGLLSYSH RPSPLFPLNGSHDLGIILQRIRDMPYMDPSGNNLGTAVVTAHR YMLAPDAPGRRQHVPGVMVLLVDEPLRGDIFSPIREAQASGL NVVMLGMAGADPEQLRRLAPGMDSVQTFFAVDDGPSLDQA VSGLATALCQASFTTQPRPEPCPVYCPKGQKGEPGEMGLRGQ VGPPGDPGLPGRTGAPGPQGPPGSATAKGERGFPGADGRPGS PGRAGNPGTPGAPGLKGSPGLPGPRGDPGERGPRGPKGEPGA PGQVIGGEGPGLPGRKGDPGPSGPPGPRGPLGDPGPRGPPGLP GTAMKGDKGDRGERGPPGPGEGGIAPGEPGLPGLPGSPGPQG PVGPPGKKGEKGDSEDGAPGLPGQPGSPGEQGPRGPPGAIGP KGDRGFPGPLGEAGEKGERGPPGPAGSRGLPGVAGRPGAKGP EGPPGPTGRQGEKGEPGRPGDPAVVGPAVAGPKGEKGDVGP AGPRGATGVQGERGPPGLVLPGDPGPKGDPGDRGPIGLTGRA GPPGDSGPPGEKGDPGRPGPPGPVGPRGRDGEVGEKGDEGPP GDPGLPGKAGERGLRGAPGVRGPVGEKGDQGDPGEDGRNGS PGSSGPKGDRGEPGPPGPPGRLVDTGPGAREKGEPGDRGQEG PRGPKGDPGLPGAPGERGIEGFRGPPGPQGDPGVRGPAGEKG DRGPPGLDGRSGLDGKPGAAGPSGPNGAAGKAGDPGRDGLP GLRGEQGLPGPSGPPGLPGKPGEDGKPGLNGKNGEPGDPGED GRKGEKGDSGASGREGRDGPKGERGAPGILGPQGPPGLPGPV GPPGQGFPGVPGGTGPKGDRGETGSKGEQGLPGERGLRGEPG SVPNVDRLLETAGIKASALREIVETWDESSGSFLPVPERRRGP KGDSGEQGPPGKEGPIGFPGERGLKGDRGDPGPQGPPGLALG ERGPPGPSGLAGEPGKPGIPGLPGRAGGVGEAGRPGERGERG EKGERGEQGRDGPPGLPGTPGPPGPPGPKVSVDEPGPGLSGEQ GPPGLKGAKGEPGSNGDQGPKGDRGVPGIKGDRGEPGPRGQ DGNPGLPGERGMAGPEGKPGLQGPRGPPGPVGGHGDPGPPG APGLAGPAGPQGPSGLKGEPGETGPPGRGLTGPTGAVGLPGP PGPSGLVGPQGSPGLPGQVGETGKPGAPGRDGASGKDGDRG	COL7A1

[0920]

	SPGVPGSPGLPGPVGPKGEPGPTGAPGQAVVGLPGAKGEKGA PGGLAGDLVGEPGAKGDRGLPGPRGEKGEAGRAGEPGDPGE DGQKGAPGPKGFKGDPGVGVPGSPGPPGPPGVKGDLGLPGLP GAPGVVGFPGQTGPRGEMGQPGPSGERGLAGPPGREGIPGPL GPPGPPGSVGPPGASGLKGDKGDPGVGLPGPRGERGEPGIRG EDGRPGQEGPRGLTGPPGSRGERGEKGDVGSAGLKGDKGDS AVILGPPGPRGAKGDMGERGPRGLDGDKGPRGDNGDPGDKG SKGEPGDKGSAGLPGLRGLLGPQGQPGAAGIPGDPGSPGKDG VPGIRGEKGDVGFMGPRGLKGERGVKGACGLDGEKGDKGE AGPPGRPGLAGHKGEMGEPGVPGQSGAPGKEGLIGPKGDRG FDGQPGPKGDQGEKGERGTPGIGGFPGPSGNDGSAGPPGPPG SVGPRGPEGLQGQKGERGPPGERVVGAPGVPGAPGERGEQG RPGPAGPRGEKGEAALTEDDIRGFVRQEMSQHCACQGQFIAS GSRPLPSYAADTAGSQLHAVPVLRVSHAEEEERVPPEDDEYS EYSEYSVEEYQDPEAPWDSDDPCSLPLDEGSCTAYTLRWYHR AVTGSTEACHPFVYGGCGGNANRFGTREACERRCPPRVVQS QGTGTAQD	
401	MSIQENISSLQLRSWVSKSQRDLAKSILIGAPGGPAGYLRRAS VAQLTQELGTAFFQQQQLPAAMADTFLEHLCLLDIDSEPVAA RSTSIIATIGPASRSVERLKEMIKAGMNIARLNFSHGSHEYHAE SIANVREAVESFAGSPLSYRPVAIALDTKGPEIRTGILQGGPES EVELVKGSQVLVTVDPAFRTRGNANTVWVDYPNIVRVVPVG GRIYIDDGLISLVVQKIGPEGLVTQVENGGVLGSRKGVNLPGA QVDLPGLSEQDVRDLRFGVEHGVDIVFASFVRKASDVAAVR AALGPEGHGIKIISKIENHEGVKRFDEILEVSDGIMVARGDLGI EIPAEKVFLAQKMMIGRCNLAGKPVVCATQMLESMITKPRPT RAETSDVANAVLDGADCIMLSGETAKGNFPVEAVKMQHAIA REAEAAVYHRQLFEELRRAAPLSRDPTEVTAIGAVEAAFKCC AAAIIVLTTTGRSAQLLSRYRPRAAVIAVTRSAQAARQVHLCR GVFPLLYREPPEAIWADDVDRRVQFGIESGKLRGFLRVGDLVI VVTGWRPGSGYTNIMRVLSIS	PKLR
402	MSSSPVKRQRMESALDQLKQFTTVVADTGDFHAIDEYKPQD ATTNPSLILAAAQMPAYQELVEEAIAYGRKLGGSQEDQIKNAI DKLFVLFGAEILKKIPGRVSTEVDARLSFDKDAMVARARRLIE LYKEAGISKDRILIKLSSTWEGIQAGKELEEQHGIHCNMTLLFS FAQAVACAEAGVTLISPFVGRILDWHVANTDKKSYEPLEDPG VKSVTKIYNYYKKFSYKTIVMGASFRNTGEIKALAGCDFLTIS PKLLGELLQDNAKLVPVLSAKAAQASDLEKIHLDEKSFRWLH NEDQMAVEKLSDGIRKFAADAVKLERMLTERMFNAENGK	TALDO1
403	MRLAVGALLVCAVLGLCLAVPDKTVRWCAVSEHEATKCQSF RDHMKSVIPSDGPSVACVKKASYLDCIRAIAANEADAVTLDA GLVYDAYLAPNNLKPVVAEFYGSKEDPQTFYYAVAVVKKDS GFQMNQLRGKKSCHTGLGRSAGWNIPIGLLYCDLPEPRKPLE KAVANFFSGSCAPCADGTDFPQLCQLCPGCGCSTLNQYFGYS GAFKCLKDGAGDVAFVKHSTIFENLANKADRDQYELLCLDN TRKPVDEYKDCHLAQVPSHTVVARSMGGKEDLIWELLNQAQ EHFGKDKSKEFQLFSSPHGKDLLFKDSAHGFLKVPPRMDAK MYLGYEYVTAIRNLREGTCPEAPTDECKP VKWCALSHHERLKCDEWSVNSVGKIECVSAETTEDCIAKIMN GEADAMSLDGGFVYIAGKCGLVPVLAENYNKSDNCEDTPEA GYFAIAVVKKSASDLTWDNLKGKKSCHTAVGRTAGWNIPM GLLYNKINHCRFDEFFSEGCAPGSKKDSSLCKLCMGSGLNLC EPNNKEGYYGYTGAFRCLVEKGDVAFVKHQTVPQNTGGKNP DPWAKNLNEKDYELLCLDGTRKPVEEYANCHLARAPNHAV	TF

[0921]

	VTRKDKEACVHKILRQQQHLFGSNVTDCSGNFCLFRSETKDL LFRDDTVCLAKLHDRNTYEKYLGEEYVKAVGNLRKCSTSSL LEACTFRRP	
404	MAPPQVLAFGLLLAAATATFAAAQEECVCENYKLAVNCFVN NNRQCQCTSVGAQNTVICSKLAAKCLVMKAEMNGSKLGRR AKPEGALQNNDGLYDPDCDESGLFKAKQCNGTSMCWCVNT AGVRRTDKDTEITCSERVRTYWIIIELKHKAREKPYDSKSLRT ALQKEITTRYQLDPKFITSILYENNVITIDLVQNSSQKTQNDVD IADVAYYFEKDVKGESLFHSKKMDLTVNGEQLDLDPGQTLIY YVDEKAPEFSMQGLKAGVIAVIVVVVIAVVAGIVVLVISRKK RMAKYEKAEIKEMGEMHRELNA	EPCAM
405	MPRRAENWDEAEVGAEEAGVEEYGPEEDGGEESGAEESGPE ESGPEELGAEEEMEAGRPRPVLRSVNSREPSQVIFCNRSPRVV LPVWLNFDGEPQPYPTLPPGTGRRIHSYRGHLWLFRDAGTHD GLLVNQTELFVPSLNVDGQPIFANITLPVYTLKERCLQVVRSL VKPENYRRLDIVRSLYEDLEDHPNVQKDLERLTQERIAHQRM GD	VHL
406	MKRVLVLLLAVAFGHALERGRDYEKNKVCKEFSHLGKEDFT SLSLVLYSRKFPSGTFEQVSQLVKEVVSLTEACCAEGADPDC YDTRTSALSAKSCESNSPFPVHPGTAECCTKEGLERKLCMAA LKHQPQEFPTYVEPTNDEICEAFRKDPKEYANQFMWEYSTNY GQAPLSLLVSYTKSYLSMVGSCCTSASPTVCFLKERLQLKHLS LLTTLSNRVCSQYAAYGEKKSRLSNLIKLAQKVPTADLEDVL PLAEDITNILSKCCESASEDCMAKELPEHTVKLCDNLSTKNSK FEDCCQEKTAMDVFVCTYFMPAAQLPELPDVELPTNKDVCD PGNTKVMDKYTFELSRRTHLPEVFLSKVLEPTLKSLGECCDV EDSTTCFNAKGPLLKKELSSFIDKGQELCADYSENTFTEYKKK LAERLKAKLPDATPTELAKLVNKHSDFASNCCSINSPPLYCDS EIDAELKNIL	GC
407	MPSSVSWGILLLAGLCCLVPVSLAEDPQGDAAQKTDTSHHDQ DHPTFNKITPNLAEFAFSLYRQLAHQSNSTNIFFSPVSIATAFA MLSLGTKADTHDEILEGLNFNLTEIPEAQIHEGFQELLRTLNQP DSQLQLTTGNGLFLSEGLKLVDKFLEDVKKLYHSEAFTVNFG DTEEAKKQINDYVEKGTQGKIVDLVKELDRDTVFALVNYIFF KGKWERPFEVKDTEEEDFHVDQVTTVKVPMMKRLGMFNIQ HCKKLSSWVLLMKYLGNATAIFFLPDEGKLQHLENELTHDIIT KFLENEDRRSASLHLPKLSITGTYDLKSVLGQLGITKVFSNGA DLSGVTEEAPLKLSKAVHKAVLTIDEKGTEAAGAMFLEAIPM SIPPEVKFNKPFVFLMIEQNTKSPLFMGKVVNPTQK	SERPINA1
408	MAAPAEPCAGQGVWNQTEPEPAATSLLSLCFLRTAGVWVPP MYLWVLGPIYLLFIHHHGRGYLRMSPLFKAKMVLGFALIVLC TSSVAVALWKIQQGTPEAPEFLIHPTVWLTTMSFAVFLIHTER KKGVQSSGVLFGYWLLCFVLPATNAAQQASGAGFQSDPVRH LSTYLCLSLVVAQFVLSCLADQPPFFPEDPQQSNPCPETGAAF PSKATFWWVSGLVWRGYRRPLRPKDLWSLGRENSSEELVSR LEKEWMRNRSAARRHNKAIAFKRKGGSGMKAPETEPFLRQE GSQWRPLLKAIWQVFHSTFLLGTLSLIISDVFRFTVPKLLSLFL EFIGDPKPPAWKGYLLAVLMFLSACLQTLFEQQNMYRLKVL QMRLRSAITGLVYRKVLALSSGSRKASAVGDVVNLVSVDVQ RLTESVLYLNGLWLPLVWIVVCFVYLWQLLGPSALTAIAVFL SLLPLNFFISKKRNHHQEEQMRQKDSRARLTSSILRNSKTIKFH GWEGAFLDRVLGIRGQELGALRTSGLLFSVSLVSFQVSTFLVA LVVFAVHTLVAENAMNAEKAFVTLTVLNILNKAQAFLPFSIH	ABCC6

		SLVQARVSFDRLVTFLCLEEVDPGVVDSSSSGSAAGKDCITIH SATFAWSQESPPCLHRINLTVPQGCLLAVVGPVGAGKSSLLSA LLGELSKVEGFVSIEGAVAYVPQEAWVQNTSVVENVCFGQEL DPPWLERVLEACALQPDVDSFPEGIHTSIGEQGMNLSGGQKQ RLSLARAVYRKAAVYLLDDPLAALDAHVGQHVFNQVIGPGG LLQGTTRILVTHALHILPQADWIIVLANGAIAEMGSYQELLQR KGALMCLLDQARQPGDRGEGETEPGTSTKDPRGTSAGRRPEL RRERSIKSVPEKDRTTSEAQTEVPLDDPDRAGWPAGKDSIQY GRVKATVHLAYLRAVGTPLCLYALFLFLCQQVASFCRGYWL SLWADDPAVGGQQTQAALRGGIFGLLGCLQAIGLFASMAAV LLGGARASRLLFQRLLWDVVRSPISFFERTPIGHLLNRFSKETD TVDVDIPDKLRSLLMYAFGLLEVSLVVAVATPLATVAILPLFL LYAGFQSLYVVSSCQLRRLESASYSSVCSHMAETFQGSTVVR AFRTQAPFVAQNNARVDESQRISFPRLVADRWLAANVELLG NGLVFAAATCAVLSKAHLSAGLVGFSVSAALQVTQTLQWVV RNWTDLENSIVSVERMQDYAWTPKEAPWRLPTCAAQPPWPQ GGQIEFRDFGLRYRPELPLAVQGVSFKIHAGEKVGIVGRTGAG KSSLASGLLRLQEAAEGGIWIDGVPIAHVGLHTLRSRISIIPQDP ILFPGSLRMNLDLLQEHSDEAIWAALETVQLKALVASLPGQL QYKCADRGEDLSVGQKQLLCLARALLRKTQILILDEATAAVD PGTELQMQAMLGSWFAQCTVLLIAHRLRSVMDCARVLVMD KGQVAESGSPAQLLAQKGLFYRLAQESGLV	
[0922]	409	MQIELSTCFFLCLLRFCFSATRRYYLGAVELSWDYMQSDLGE LPVDARFPPRVPKSFPFNTSVVYKKTLFVEFTDHLFNIAKPRPP WMGLLGPTIQAEVYDTVVITLKNMASHPVSLHAVGVSYWKA SEGAEYDDQTSQREKEDDKVFPGGSHTYVWQVLKENGPMAS DPLCLTYSYLSHVDLVKDLNSGLIGALLVCREGSLAKEKTQT LHKFILLFAVFDEGKSWHSETKNSLMQDRDAASARAWPKMH TVNGYVNRSLPGLIGCHRKSVYWHVIGMGTTPEVHSIFLEGH TFLVRNHRQASLEISPITFLTAQTLLMDLGQFLLFCHISSHQHD GMEAYVKVDSCPEEPQLRMKNNEEAEDYDDDLTDSEMDVV RFDDDNSPSFIQIRSVAKKHPKTWVHYIAAEEEDWDYAPLVL APDDRSYKSQYLNNGPQRIGRKYKKVRFMAYTDETFKTREAI QHESGILGPLLYGEVGDTLLIIFKNQASRPYNIYPHGITDVRPL YSRRLPKGVKHLKDFPILPGEIFKYKWTVTVEDGPTKSDPRCL TRYYSSFVNMERDLASGLIGPLLICYKESVDQRGNQIMSDKR NVILFSVFDENRSWYLTENIQRFLPNPAGVQLEDPEFQASNIM HSINGYVFDSLQLSVCLHEVAYWYILSIGAQTDFLSVFFSGYT FKHKMVYEDTLTLFPFSGETVFMSMENPGLWILGCHNSDFRN RGMTALLKVSSCDKNTGDYYEDSYEDISAYLLSKNNAIEPRS FSQNSRHPSTRQKQFNATTIPENDIEKTDPWFAHRTPMPKIQN VSSSDLLMLLRQSPTPHGLSLSDLQEAKYETFSDDPSPGAIDS NNSLSEMTHFRPQLHHSGDMVFTPESGLQLRLNEKLGTTAAT ELKKLDFKVSSTSNNLISTIPSDNLAAGTDNTSSLGPPSMPVHY DSQLDTTLFGKKSSPLTESGGPLSLSEENNDSKLLESGLMNSQ ESSWGKNVSSTESGRLFKGKRAHGPALLTKDNALFKVSISLL KTNKTSNNSATNRKTHIDGPSLLIENSPSVWQNILESDTEFKK VTPLIHDRMLMDKNATALRLNHMSNKTTSSKNMEMVQQKK EGPIPPDAQNPDMSFFKMLFLPESARWIQRTHGKNSLNSGQGP SPKQLVSLGPEKSVEGQNFLSEKNKVVVGKGEFTKDVGLKE MVFPSSRNLFLTNLDNLHENNTHNQEKKIQEEIEKKETLIQEN VVLPQIHTVTGTKNFMKNLFLLSTRQNVEGSYDGAYAPVLQ DFRSLNDSTNRTKKHTAHFSKKGEEENLEGLGNQTKQIVEKY ACTTRISPNTSQQNFVTQRSKRALKQFRLPLEETELEKRIIVDD	F8

[0923]

	TSTQWSKNMKHLTPSTLTQIDYNEKEKGAITQSPLSDCLTRSH SIPQANRSPLPIAKVSSFPSIRPIYLTRVLFQDNSSHLPAASYRK KDSGVQESSHFLQGAKKNNLSLAILTLEMTGDQREVGSLGTS ATNSVTYKKVENTVLPKPDLPKTSGKVELLPKVHIYQKDLFP TETSNGSPGHLDLVEGSLLQGTEGAIKWNEANRPGKVPFLRV ATESSAKTPSKLLDPLAWDNHYGTQIPKEEWKSQEKSPEKTA FKKKDTILSLNACESNHAIAAINEGQNKPEIEVTWAKQGRTER LCSQNPPVLKRHQREITRTTLQSDQEEIDYDDTISVEMKKEDF DIYDEDENQSPRSFQKKTRHYFIAAVERLWDYGMSSSPHVLR NRAQSGSVPQFKKVVFQEFTDGSFTQPLYRGELNEHLGLLGP YIRAEVEDNIMVTFRNQASRPYSFYSSLISYEEDQRQGAEPRK NFVKPNETKTYFWKVQHHMAPTKDEFDCKAWAYFSDVDLE KDVHSGLIGPLLVCHTNTLNPAHGRQVTVQEFALFFTIFDETK SWYFTENMERNCRAPCNIQMEDPTFKENYRFHAINGYIMDTL PGLVMAQDQRIRWYLLSMGSNENIHSIHFSGHVFTVRKKEEY KMALYNLYPGVFETVEMLPSKAGIWRVECLIGEHLHAGMST LFLVYSNKCQTPLGMASGHIRDFQITASGQYGQWAPKLARLH YSGSINAWSTKEPFSWIKVDLLAPMIIHGIKTQGARQKFSSLYI SQFIIMYSLDGKKWQTYRGNSTGTLMVFFGNVDSSGIKHNIF NPPIIARYIRLHPTHYSIRSTLRMELMGCDLNSCSMPLGMESK AISDAQITASSYFTNMFATWSPSKARLHLQGRSNAWRPQVNN PKEWLQVDFQKTMKVTGVTTQGVKSLLTSMYVKEFLISSSQ DGHQWTLFFQNGKVKVFQGNQDSFTPVVNSLDPPLLTRYLRI HPQSWVHQIALRMEVLGCEAQDLY	
410	MQRVNMIMAESPGLITICLLGYLLSAECTVFLDHENANKILNR PKRYNSGKLEEFVQGNLERECMEEKCSFEEAREVFENTERTT EFWKQYVDGDQCESNPCLNGGSCKDDINSYECWCPFGFEGK NCELDVTCNIKNGRCEQFCKNSADNKVVCSCTEGYRLAENQ KSCEPAVPFPCGRVSVSQTSKLTRAETVFPDVDYVNSTEAETI LDNITQSTQSFNDFTRVVGGEDAKPGQFPWQVVLNGKVDAF CGGSIVNEKWIVTAAHCVETGVKITVVAGEHNIEETEHTEQK RNVIRII PHHNYNAAINKYNHDIALLELDEPLVLNSYVTPICIADKEYTN IFLKFGSGYVSGWGRVFHKGRSALVLQYLRVPLVDRATCLRS TKFTIYNNMFCAGFHEGGRDSCQGDSGGPHVTEVEGTSFLTG IISWGEECAMKGKYGIYTKVSRYVNWIKEKTKLT	F9
411	MDPPRPALLALLALPALLLLLLAGARAEEEMLENVSLVCPKD ATRFKHLRKYTYNYEAESSSGVPGTADSRSATRINCKVELEV PQLCSFILKTSQCTLKEVYGFNPEGKALLKKTKNSEEFAAAMS RYELKLAIPEGKQVFLYPEKDEPTYILNIKRGIISALLVPPETEE AKQVLFLDTVYGNCSTHFTVKTRKGNVATEISTERDLGQCDR FKPIRTGISPLALIKGMTRPLSTLISSSQSCQYTLDAKRKHVAE AICKEQHLFLPFSYKNKYGMVAQVTQTLKLEDTPKINSRFFGE GTKKMGLAFESTKSTSPPKQAEAVLKTLQELKKLTISEQNIQR ANLFNKLVTELRGLSDEAVTSLLPQLIEVSSPITLQALVQCGQP QCSTHILQWLKRVHANPLLIDVVTYLVALIPEPSAQQLREIFN MARDQRSRATLYALSHAVNNYHKTNPTGTQELLDIANYLME QIQDDCTGDEDYTYLILRVIGNMGQTMEQLTPELKSSILKCVQ STKPSLMIQKAAIQALRKMEPKDKDQEVLLQTFLDDASPGDK RLAAYLMLMRSPSQAINKIVQILPWEQNEQVKNFVASHIANIL NSEELDIQDLKKLVKEALKESQLPTVMDFRKFSRNYQLYKSV SLPSLDPASAKIEGNLIFDPNNYLPKESMLKTTLTAFGFASADL IEIGLEGKGFEPTLEALFGKQGFFPDSVNKALYWVNGQVPDG VSKVLVDHFGYTKDDKHEQDMVNGIMLSVEKLIKDLKSKEV	ApoB

[0924]

PEARAYLRILGEELGFASLHDLQLLGKLLLMGARTLQGIPQMI
GEVIRKGSKNDFFLHYIFMENAFELPTGAGLQLQISSSGVIAPG
AKAGVKLEVANMQAELVAKPSVSVEFVTNMGIIIPDFARSGV
QMNTNFFHESGLEAHVALKAGKLKFIIPSPKRPVKLLSGGNTL
HLVSTTKTEVIPPLIENRQSWSVCKQVFPGLNYCTSGAYSNAS
STDSASYYPLTGDTRLELELRPTGEIEQYSVSATYELQREDRA
LVDTLKFVTQAEGAKQTEATMTFKYNRQSMTLSSEVQIPDFD
VDLGTILRVNDESTEGKTSYRLTLDIQNKKITEVALMGHLSCD
TKEERKIKGVISIPRLQAEARSEILAHWSPAKLLLQMDSSATA
YGSTVSKRVAWHYDEEKIEFEWNTGTNVDTKKMTSNFPVDL
SDYPKSLHMYANRLLDHRVPQTDMTFRHVGSKLIVAMSSWL
QKASGSLPYTQTLQDHLNSLKEFNLQNMGLPDFHIPENLFLKS
DGRVKYTLNKNSLKIEIPLPFGGKSSRDLKMLETVRTPALHFK
SVGFHLPSREFQVPTFTIPKLYQLQVPLLGVLDLSTNVYSNLY
NWSASYSGGNTSTDHFSLRARYHMKADSVVDLLSYNVQGSG
ETTYDHKNTFTLSYDGSLRHKFLDSNIKFSHVEKLGNNPVSK
GLLIFDASSSWGPQMSASVHLDSKKKQHLFVKEVKIDGQFRV
SSFYAKGTYGLSCQRDPNTGRLNGESNLRFNSSYLQGTNQIT
GRYEDGTLSLTSTSDLQSGIIKNTASLKYENYELTLKSDTNGK
YKNFATSNKMDMTFSKQNALLRSEYQADYESLRFFSLLSGSL
NSHGLELNADILGTDKINSGAHKATLRIGQDGISTSATTNLKC
SLLVLENELNAELGLSGASMKLTTNGRFREHNAKFSLDGKAA
LTELSLGSAYQAMILGVDSKNIFNFKVSQEGLKLSNDMMGSY
AEMKFDHTNSLNIAGLSLDFSSKLDNIYSSDKFYKQTVNLQL
QPYSLVTTLNSDLKYNALDLTNNGKLRLEPLKLHVAGNLKG
AYQNNEIKHIYAISSAALSASYKADTVAKVQGVEFSHRLNTDI
AGLASAIDMSTNYNSDSLHFSNVFRSVMAPFTMTIDAHTNGN
GKLALWGEHTGQLYSKFLLKAEPLAFTFSHDYKGSTSHHLVS
RKSISAALEHKVSALLTPAEQTGTWKLKTQFNNNEYSQDLDA
YNTKDKIGVELTGRTLADLTLLDSPIKVPLLLSEPINIIDALEM
RDAVEKPQEFTIVAFVKYDKNQDVHSINLPFFETLQEYFERNR
QTIIVVLENVQRNLKHINIDQFVRKYRAALGKLPQQANDYLN
SFNWERQVSHAKEKLTALTKKYRITENDIQIALDDAKINFNEK
LSQLQTYMIQFDQYIKDSYDLHDLKIAIANIIDEIIEKLKSLDEH
YHIRVNLVKTIHDLHLFIENIDFNKSGSSTASWIQNVDTKYQIR
IQIQEKLQQLKRHIQNIDIQHLAGKLKQHIEAIDVRVLLDQLGT
TISFERINDILEHVKHFVINLIGDFEVAEKINAFRAKVHELIERY
EVDQQIQVLMDKLVELAHQYKLKETIQKLSNVLQQVKIKDYF
EKLVGFIDDAVKKLNELSFKTFIEDVNKFLDMLIKKLKSFDYH
QFVDETNDKIREVTQRLNGEIQALELPQKAEALKLFLEETKAT
VAVYLESLQDTKITLIINWLQEALSSASLAHMKAKFRETLEDT
RDRMYQMDIQQELQRYLSLVGQVYSTLVTYISDWWTLAAK
NLTDFAEQYSIQDWAKRMKALVEQGFTVPEIKTILGTMPAFE
VSLQALQKATFQTPDFIVPLTDLRIPSVQINFKDLKNIKIPSRFS
TPEFTILNTFHIPSFTIDFVEMKVKIIRTIDQMLNSELQWPVPDI
YLRDLKVEDIPLARITLPDFRLPEIAIPEFIIPTLNLNDFQVPDLH
IPEFQLPHISHTIEVPTFGKLYSILKIQSPLFTLDANADIGNGTTS
ANEAGIAASITAKGESKLEVLNFDFQANAQLSNPKINPLALKE
SVKFSSKYLRTEHGSEMLFFGNAIEGKSNTVASLHTEKNTLEL
SNGVIVKINNQLTLDSNTKYFHKLNIPKLDFSSQADLRNEIKTL
LKAGHIAWTSSGKGSWKWACPRFSDEGTHESQISFTIEGPLTS
FGLSNKINSKHLRVNQNLVYESGSLNFSKLEIQSQVDSQHVG
HSVLTAKGMALFGEGKAEFTGRHDAHLNGKVIGTLKNSLFFS
AQPFEITASTNNEGNLKVRFPLRLTGKIDFLNNYALFLSPSAQ

[0925]

	QASWQVSARFNQYKYNQNFSAGNNENIMEAHVGINGEANLD FLNIPLTIPEMRLPYTIITTPPLKDFSLWEKTGLKEFLKTTKQSF DLSVKAQYKKNKHRHSITNPLAVLCEFISQSIKSFDRHFEKNR NNALDFVTKSYNETKIKFDKYKAEKSHDELPRTFQIPGYTVPV VNVEVSPFTIEMSAFGYVFPKAVSMPSFSILGSDVRVPSYTLIL PSLELPVLHVPRNLKLSLPDFKELCTISHIFIPAMGNITYDFSFK SSVITLNTNAELFNQSDIVAHLLSSSSSVIDALQYKLEGTTRLT RKRGLKLATALSLSNKFVEGSHNSTVSLTTKNMEVSVATTTK AQIPILRMNFKQELNGNTKSKPTVSSSMEFKYDFNSSMLYSTA KGAVDHKLSLESLTSYFSIESSTKGDVKGSVLSREYSGTIASEA NTYLNSKSTRSSVKLQGTSKIDDIWNLEVKENFAGEATLQRIY SLWEHSTKNHLQLEGLFFTNGEHTSKATLELSPWQMSALVQ VHASQPSSFHDFPDLGQEVALNANTKNQKIRWKNEVRIHSGS FQSQVELSNDQEKAHLDIAGSLEGHLRFLKNIILPVYDKSLWD FLKLDVTTSIGRRQHLRVSTAFVYTKNPNGYSFSIPVKVLADK FIIPGLKLNDLNSVLVMPTFHVPFTDLQVPSCKLDFREIQIYKK LRTSSFALNLPTLPEVKFPEVDVLTKYSQPEDSLIPFFEITVPES QLTVSQFTLPKSVSDGIAALDLNAVANKIADFELPTIIVPEQTIE IPSIKFSVPAGIVIPSFQALTARFEVDSPVYNATWSASLKNKAD YVETVLDSTCSSTVQFLEYELNVLGTHKIEDGTLASKTKGTFA HRDFSAEYEEDGKYEGLQEWEGKAHLNIKSPAFTDLHLRYQ KDKKGISTSAASPAVGTVGMDMDEDDDFSKWNFYYSPQSSP DKKLTIFKTELRVRESDEETQIKVNWEEEAASGLLTSLKDNVP KATGVLYDYVNKYHWEHTGLTLREVSSKLRRNLQNNAEWV YQGAIRQIDDIDVRFQKAASGTTGTYQEWKDKAQNLYQELL TQEGQASFQGLKDNVFDGLVRVTQEFHMKVKHLIDSLIDFLN FPRFQFPGKPGIYTREELCTMFIREVGTVLSQVYSKVHNGSEIL FSYFQDLVITLPFELRKHKLIDVISMYRELLKDLSKEAQEVFK AIQSLKTTEVLRNLQDLLQFIFQLIEDNIKQLKEMKFTYLINYI QDEINTIFSDYIPYVFKLLKENLCLNLHKFNEFIQNELQEASQE LQQIHQYIMALREEYFDPSIVGWTVKYYELEEKIVSLIKNLLV ALKDFHSEYIVSASNFTSQLSSQVEQFLHRNIQEYLSILTDPDG KGKEKIAELSATAQEIIKSQAIATKKIISDYHQQFRYKLQDFSD QLSDYYEKFIAESKRLIDLSIQNYHTFLIYITELLKKLQSTTVM NPYMKLAPGELTIIL	
412	MGTVSSRRSWWPLPLLLLLLLLLGPAGARAQEDEDGDYEEL VLALRSEEDGLAEAPEHGTTATFHRCAKDPWRLPGTYVVVL KEETHLSQSERTARRLQAQAARRGYLTKILHVFHGLLPGFLV KMSGDLLELALKLPHVDYIEEDSSVFAQSIPWNLERITPPRYR ADEYQPPDGGSLVEVYLLDTSIQSDHREIEGRVMVTDFENVP EEDGTRFHRQASKCDSHGTHLAGVVSGRDAGVAKGASMRSL RVLNCQGKGTVSGTLIGLEFIRKSQLVQPVGPLVVLLPLAGGY SRVLNAACQRLARAGVVLVTAAGNFRDDACLYSPASAPEVIT VGATNAQDQPVTLGTLGTNFGRCVDLFAPGEDIIGASSDCSTC FVSQSGTSQAAAHVAGIAAMMLSAEPELTLAELRQRLIHFSA KDVINEAWFPEDQRVLTPNLVAALPPSTHGAGWQLFCRTVW SAHSGPTRMATAVARCAPDEELLSCSSFSRSGKRRGERMEAQ GGKLVCRAHNAFGGEGVYAIARCCLLPQANCSVHTAPPAEA SMGTRVHCHQQGHVLTGCSSHWEVEDLGTHKPPVLRPRGQP NQCVGHREASIHASCCHAPGLECKVKEHGIPAPQEQVTVACE EGWTLTGCSALPGTSHVLGAYAVDNTCVVRSRDVSTTGSTSE GAVTAVAICCRSRHLAQASQELQ	PCSK9
413	MDALKSAGRALIRSPSLAKQSWGGGGRHRKLPENWTDTRET LLEGMLFSLKYLGMTLVEQPKGEELSAAAIKRIVATAKASGK	LDLRAP1

		KLQKVTLKVSPRGIILTDNLTNQLIENVSIYRISYCTADKMHD KVFAYIAQSQHNQSLECHAFLCTKRKMAQAVTLTVAQAFKV AFEFWQVSKEEKEKRDKASQEGGDVLGARQDCTPSLKSLVA TGNLLDLEETAKAPLSTVSANTTNMDEVPRPQALSGSSVVWE LDDGLDEAFSRLAQSRTNPQVLDTGLTAQDMHYAQCLSPVD WDKPDSSGTEQDDLFSF	
[0926]	414	MGDLSSLTPGGSMGLQVNRGSQSSLEGAPATAPEPHSLGILH ASYSVSHRVRPWWDITSCRQQWTRQILKDVSLYVESGQIMCI LGSSGSGKTTLLDAMSGRLGRAGTFLGEVYVNGRALRREQF QDCFSYVLQSDTLLSSLTVRETLHYTALLAIRRGNPGSFQKKV EAVMAELSLSHVADRLIGNYSLGGISTGERRRVSIAAQLLQDP KVMLFDEPTTGLDCMTANQIVVLLVELARRNRIVVLTIHQPR SELFQLFDKIAILSFGELIFCGTPAEMLDFFNDCGYPCPEHSNPF DFYMDLTSVDTQSKEREIETSKRVQMIESAYKKSAICHKTLK NIERMKHLKTLPMVPFKTKDSPGVFSKLGVLLRRVTRNLVRN KLAVITRLLQNLIMGLFLLFFVLRVRSNVLKGAIQDRVGLLYQ FVGATPYTGMLNAVNLFPVLRAVSDQESQDGLYQKWQMML AYALHVLPFSVVATMIFSSVCYWTLGLHPEVARFGYFSAALL APHLIGEFLTLVLLGIVQNPNIVNSVVALLSIAGVLVGSGFLRN IQEMPIPFKIISYFTFQKYCSEILVVNEFYGLNFTCGSSNVSVTT NPMCAFTQGIQFIEKTCPGATSRFTMNFLILYSFIPALVILGIVV FKIRDHLISR	ABCG5
	415	MAGKAAEERGLPKGATPQDTSGLQDRLFSSESDNSLYFTYSG QPNTLEVRDLNYQVDLASQVPWFEQLAQFKMPWTSPSCQNS CELGIQNLSFKVRSGQMLAIIGSSGCGRASLLDVITGRGHGGK IKSGQIWINGQPSSPQLVRKCVAHVRQHNQLLPNLTVRETLAF IAQMRLPRTFSQAQRDKRVEDVIAELRLRQCADTRVGNMYV RGLSGGERRRVSIGVQLLWNPGILILDEPTSGLDSFTAHNLVK TLSRLAKGNRLVLISLHQPRSDIFRLFDLVLLMTSGTPIYLGAA QHMVQYFTAIGYPCPRYSNPADFYVDLTSIDRRSREQELATRE KAQSLAALFLEKVRDLDDFLWKAETKDLDEDTCVESSVTPLD TNCLPSPTKMPGAVQQFTTLIRRQISNDFRDLPTLLIHGAEACL MSMTIGFLYFGHGSIQLSFMDTAALLFMIGALIPFNVILDVISK CYSERAMLYYELEDGLYTTGPYFFAKILGELPEHCAYIIIYGM PTYWLANLRPGLQPFLLHFLLVWLVVFCCRIMALAAAALLPT FHMASFFSNALYNSFYLAGGFMINLSSLWTVPAWISKVSFLR WCFEGLMKIQFSRRTYKMPLGNLTIAVSGDKILSVMELDSYP LYAIYLIVIGLSGGFMVLYYVSLRFIKQKPSQDW	ABCG8
	416	MGPPGSPWQWVTLLLGLLLPPAAPFWLLNVLFPPHTTPKAEL SNHTRPVILVPGCLGNQLEAKLDKPDVVNWMCYRKTEDFFTI WLDLNMFLPLGVDCWIDNTRVVYNRSSGLVSNAPGVQIRVP GFGKTYSVEYLDSSKLAGYLHTLVQNLVNNGYVRDETVRAA PYDWRLEPGQQEEYYRKLAGLVEEMHAAYGKPVFLIGHSLG CLHLLYFLLRQPQAWKDRFIDGFISLGAPWGGSIKPMLVLAS GDNQGIPIMSSIKLKEEQRITTTSPWMFPSRMAWPEDHVFISTP SFNYTGRDFQRFFADLHFEEGWYMWLQSRDLLAGLPAPGVE VYCLYGVGLPTPRTYIYDHGFPYTDPVGVLYEDGDDTVATRS TELCGLWQGRQPQPVHLLPLHGIQHLNMVFSNLTLEHINAILL GAYRQGPPASPTASPEPPPPE	LCAT
	417	MKIATVSVLLPLALCLIQDAASKNEDQEMCHEFQAFMKNGK LFCPQDKKFFQSLDGIMFINKCATCKMILEKEAKSQKRARHL ARAPKATAPTELNCDDFKKGERDGDFICPDYYEAVCGTDGK TYDNRCALCAENAKTGSQIGVKSEGECKSSNPEQDVCSAFRP FVRDGRLGCTRENDPVLGPDGKTHGNKCAMCAELFLKEAEN	SPINK5

[0927]

	AKREGETRIRRNAEKDFCKEYEKQVRNGRLFCTRESDPVRGP DGRMHGNKCALCAEIFKQRFSEENSKTDQNLGKAEEKTKVK REIVKLCSQYQNQAKNGILFCTRENDPIRGPDGKMHGNLCSM CQAYFQAENEEKKKAEARARNKRESGKATSYAELCSEYRKL VRNGKLACTRENDPIQGPDGKVHGNTCSMCEVFFQAEEEEK KKKEGKSRNKRQSKSTASFEELCSEYRKSRKNGRLFCTREND PIQGPDGKMHGNTCSMCEAFFQQEERARAKAKREAAKEICSE FRDQVRNGTLICTREHNPVRGPDGKMHGNKCAMCASVFKLE EEEKKNDKEEKGKVEAEKVKREAVQELCSEYRHYVRNGRLP CTRENDPIEGLDGKIHGNTCSMCEAFFQQEAKEKERAEPRAK VKREAEKETCDEFRRLLQNGKLFCTRENDPVRGPDGKTHGN KCAMCKAVFQKENEERKRKEEEDQRNAAGHGSSGGGGGNT QDECAEYREQMKNGRLSCTRESDPVRDADGKSYNNQCTMC KAKLEREAERKNEYSRSRSNGTGSESGKDTCDEFRSQMKNG KLICTRESDPVRGPDGKTHGNKCTMCKEKLEREAAEKKKKE DEDRSNTGERSNTGERSNDKEDLCREFRSMQRNGKLICTREN NPVRGPYGKMHINKCAMCQSIFDREANERKKKDEEKSSSKPS NNAKDECSEFRNYIRNNELICPRENDPVHGADGKFYTNKCYM CRAVFLTEALERAKLQEKPSHVRASQEEDSPDSFSSLDSEMCK DYRVLPRIGYLCPKDLKPVCGDDGQTYNNPCMLCHENLIRQT NTHIRSTGKCEESSTPGTTAASMPPSDE	
418	MEKNGNNRKLRVCVATCNRADYSKLAPIMFGIKTEPEFFELD VVVLGSHLIDDYGNTYRMIEQDDFDINTRLHTIVRGEDEAAM VESVGLALVKLPDVLNRLKPDIMIVHGDRFDALALATSAALM NIRILHIEGGEVSGTIDDSIRHAITKLAHYHVCCTRSAEQHLIS MCEDHDRILLAGCPSYDKLLSAKNKDYMSIIRMWLGDDVKS KDYIVALQHPVTTDIKHSIKMFELTLDALISFNKRTLVLFPNID AGSKEMVRVMRKKGIEHHPNFRAVKHVPFDQFIQLVAHAGC MIGNSSCGVREVGAFGTPVINLGTRQIGRETGENVLHVRDAD TQDKILQALHLQFGKQYPCSKIYGDGNAVPRILKFLKSIDLQE PLQKKFCFPPVKENISQDIDHILETLSALAVDLGGTNLRVAIVS MKGEIVKKYTQFNPKTYEERINLILQMCVEAAAEAVKLNCRI LGVGISTGGRVNPREGIVLHSTKLIQEWNSVDLRTPLSDTLHL PVWVDNDGNCAALAERKFGQGKGLENFVTLITGTGIGGGIIH QHELIHGSSFCAAELGHLVVSLDGPDCSCGSHGCIEAYASGM ALQREAKKLHDEDLLLVEGMSVPKDEAVGALHLIQAAKLGN AKAQSILRTAGTALGLGVVNILHTMNPSLVILSGVLASHYIHI VKDVIRQQALSSVQDVDVVVSDLVDPALLGAASMVLDYTTR RIY	GNE
419	DIQMTQTTSSLSASLGDRVTISCRASQDISKYLNWYQQKPDGT VKLLIYHTSRLHSGVPSRFSGSGSGTDYSLTISNLEQEDIATYF CQQGNTLPYTFGGGTKLEITGSTSGSGKPGSGEGSTKGEVKL QESGPGLVAPSQSLSVTCTVSGVSLPDYGVSWIRQPPRKGLE WLGVIWGSETTYYNSALKSRLTIIKDNSKSQVFLKMNSLQTD DTAIYYCAKHYYYGGSYAMDYWGQGTSVTVSS	抗 CD19 scFv (FMC63)
420	DIQMTQTTSSLSASLGDRVTISCRASQDISKYLNWYQQKPDGT VKLLIYHTSRLHSGVPSRFSGSGSGTDYSLTISNLEQEDIATYF CQQGNTLPYTFGGGTKLEITGGGGSGGGGSGGGGSEVKLQES GPGLVAPSQSLSVTCTVSGVSLPDYGVSWIRQPPRKGLEWLG VIWGSETTYYNSALKSRLTIIKDNSKSQVFLKMNSLQTDDTAI YYCAKHYYYGGSYAMDYWGQGTSVTVSS	抗 CD19 scFv (FMC63)
421	ESKYGPPCPPCP	IgG4 铰链
422	TTTPAPRPPTPAPTIASQPLSLRPE	CD8 铰链

[0928]

423	IEVMYPPPYLDNEKSNGTIIHVKGKHLCPSPLFPGPSKP	CD28
424	ACRPAAGGAVHTRGLDFACDIYIWAPLAGTCGVLLLSLVITL YC	CD8
425	FWVLVVVGGVLACYSLLVTVAFIIFWV	CD28
426	FWVLVVVGGVLACYSLLVTVAFIIFWV	CD28
427	RSKRSRLLHSDYMNMTPRRPGPTRKHYQPYAPPRDFAAYRS	CD28
428	KRGRKKLLYIFKQPFMRPVQTTQEEDGCSCRFPEEEEGGCEL	4-1BB
429	RVKFSRSADAPAYQQGQNQLYNELNLGRREEYDVLDKRRGR DPEMGGKPRRKNPQEGLYNELQKDKMAEAYSEIGMKGERRR GKGHDGLYQGLSTATKDTYDALHMQALPPR	CD3ζ
430	RVKFSRSADAPAYKQGQNQLYNELNLGRREEYDVLDKRRGR DPEMGGKPRRKNPQEGLYNELQKDKMAEAYSEIGMKGERRR GKGHDGLYQGLSTATKDTYDALHMQALPPR	CD3ζ

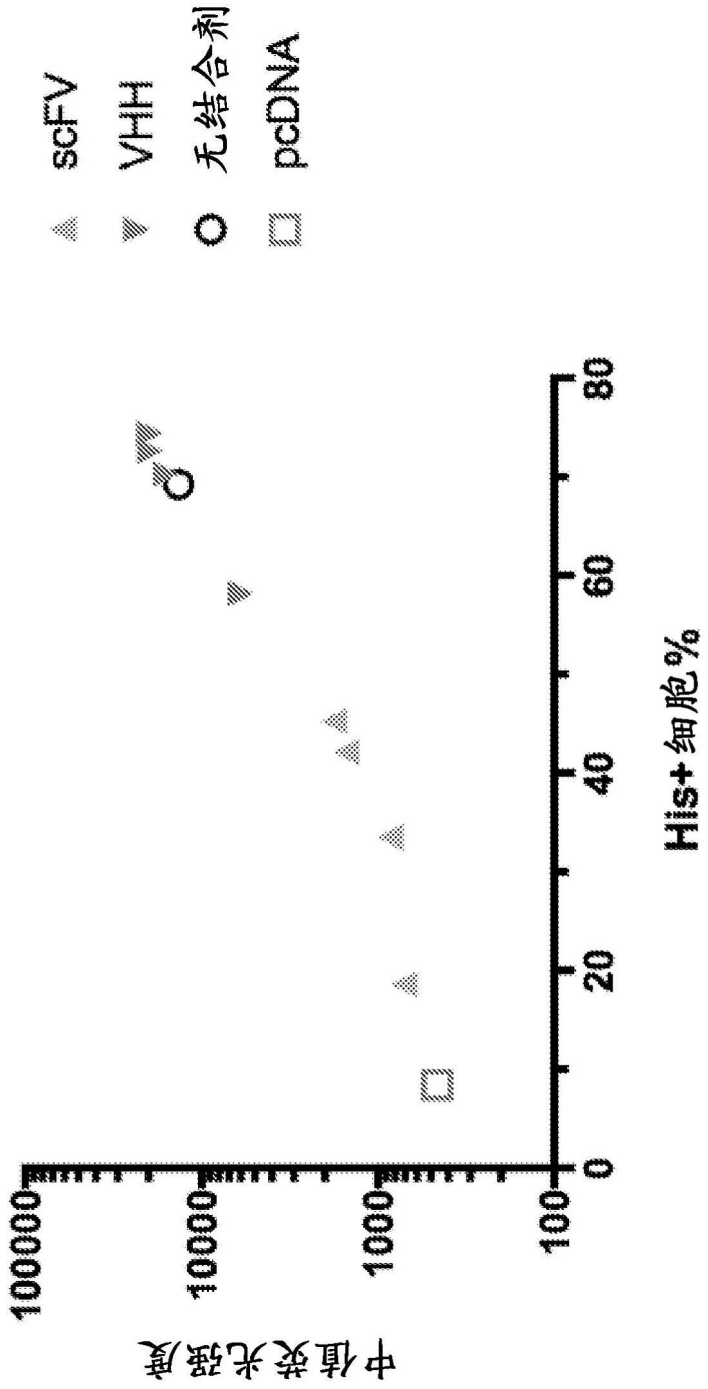

scFV
VHH
无结合剂
pcDNA
中值荧光强度
100000
10000
1000
100
0
20
40
60
80
His+细胞%

图1A

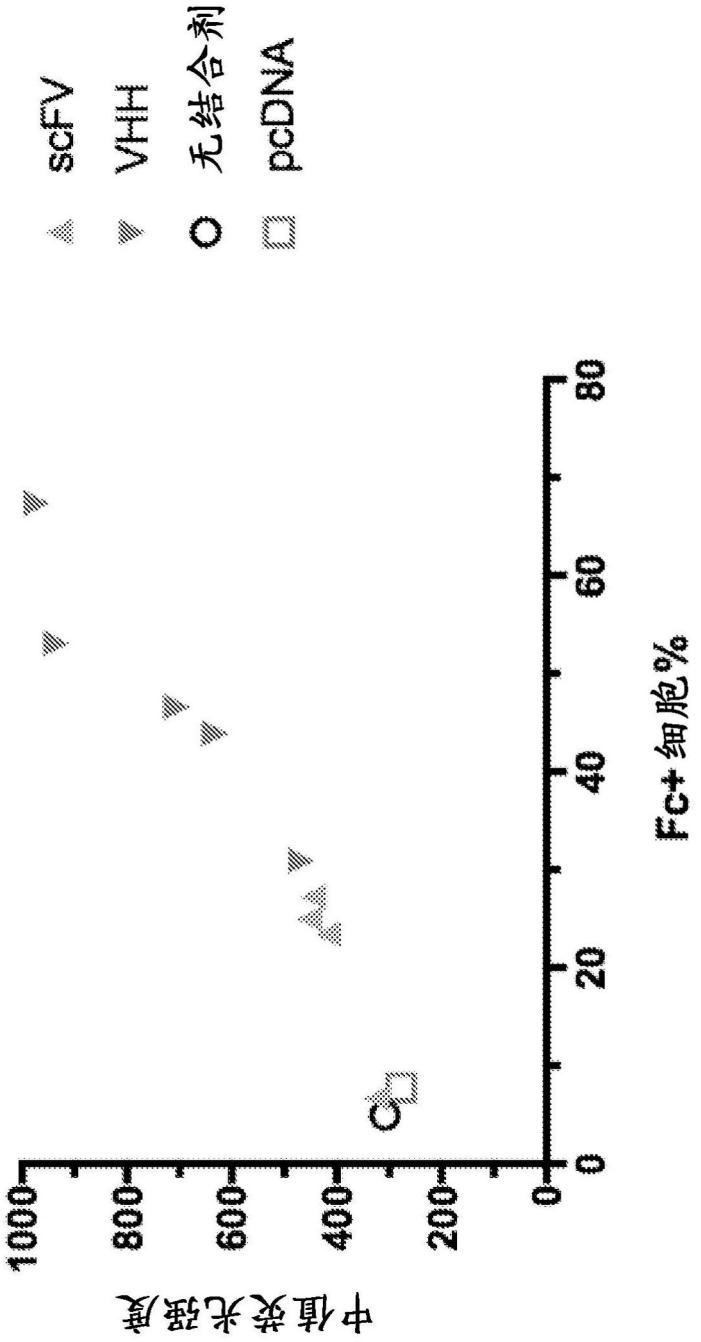

scFV
VHH
无结合剂
pcDNA
中值荧光强度
1000
800
600
400
200
0
0
20
40
60
80
Fc+ 细胞%

图1B

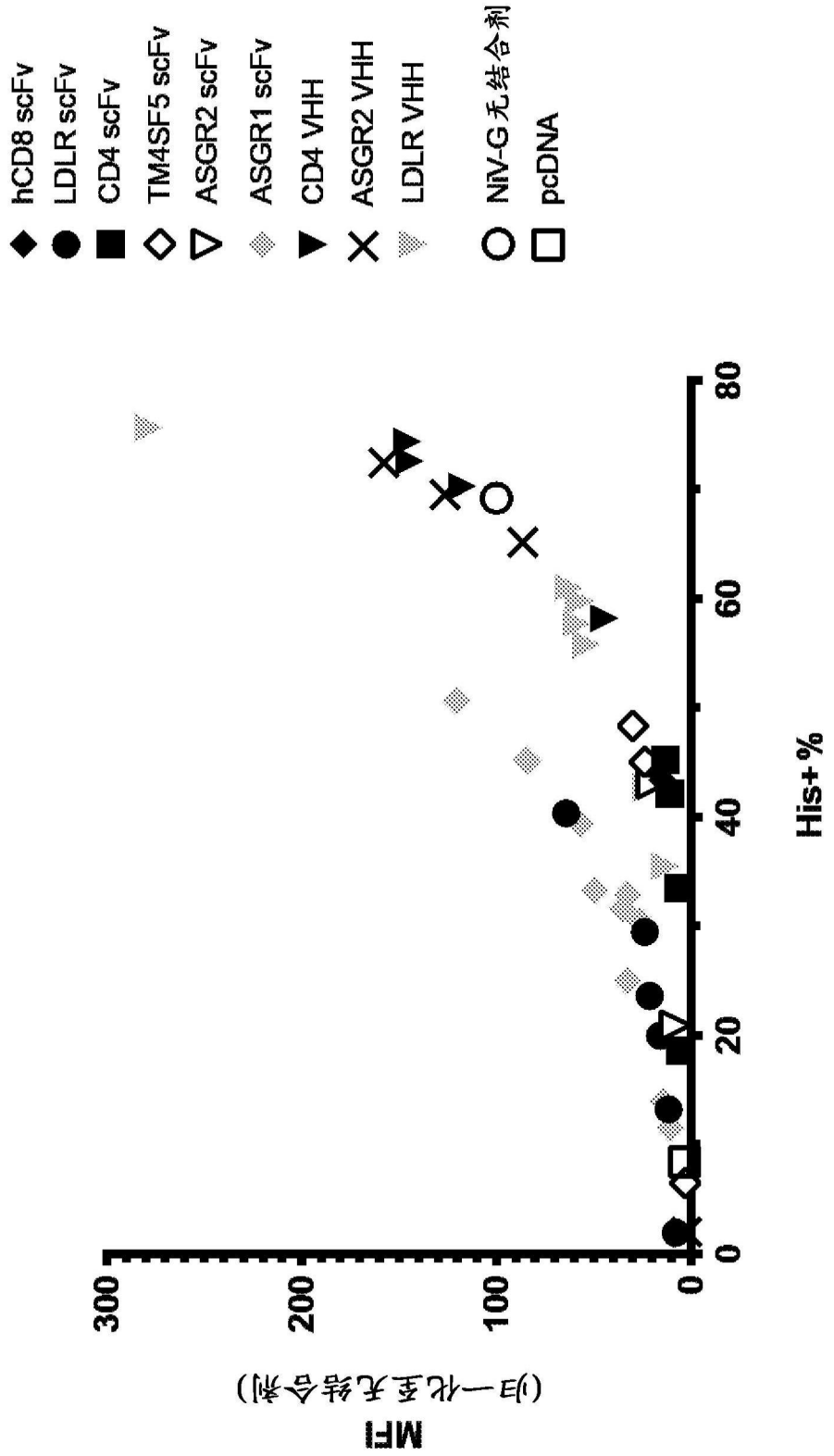

hCD8 scFv
LDLR scFv
CD4 scFv
TM4SF5 scFv
ASGR2 scFv
ASGR1 scFv
CD4 VHH
ASGR2 VHH
LDLR VHH
NiV-G无结合剂
pcDNA
MFI
(归一化至无结合剂)
300
200
100
0
0
20
40
60
80
His+ %

图1C

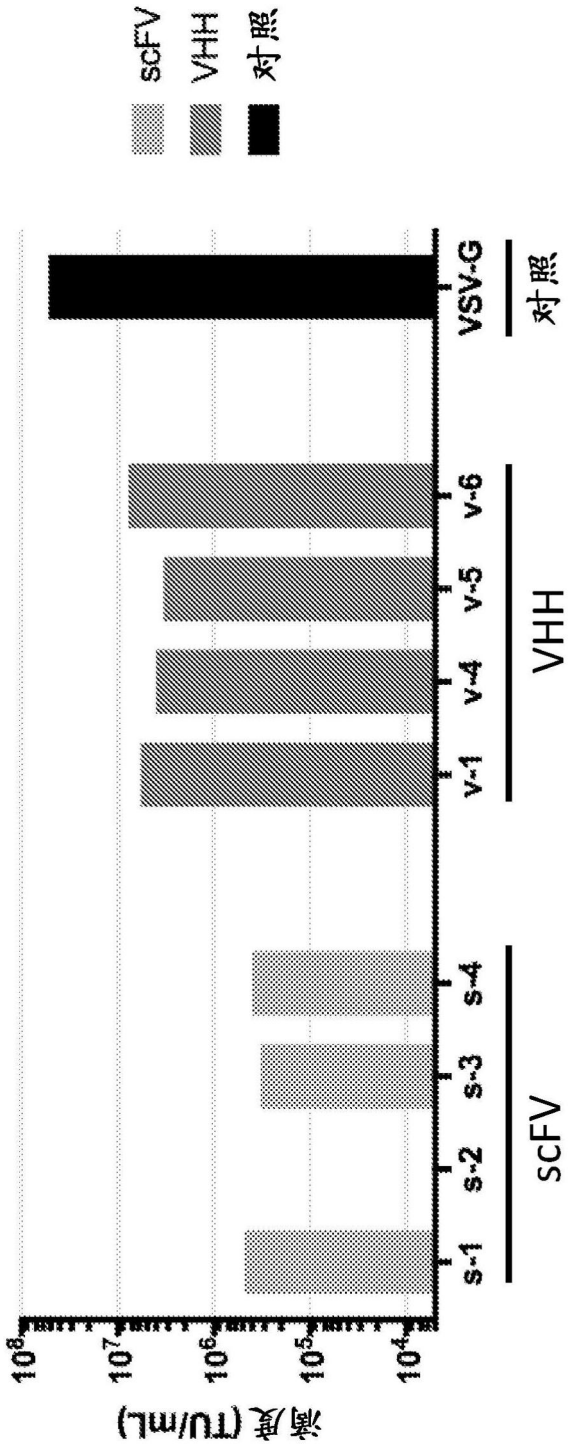

scFV
VHH
对照
滴度 (TU/mL)
10^8
10^7
10^6
10^5
10^4
VSV-G
v-6
v-5
v-4
v-1
s-4
s-3
s-2
s-1
对照
VHH
scFV

图2

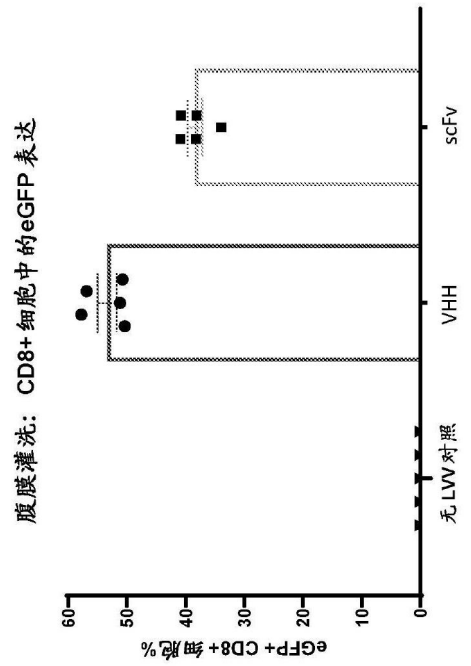

腹膜灌洗：CD8+细胞中的eGFP表达
eGFP+ CD8+细胞%
60
50
40
30
20
10
0
无LVV对照
VHH
scFv

图3A

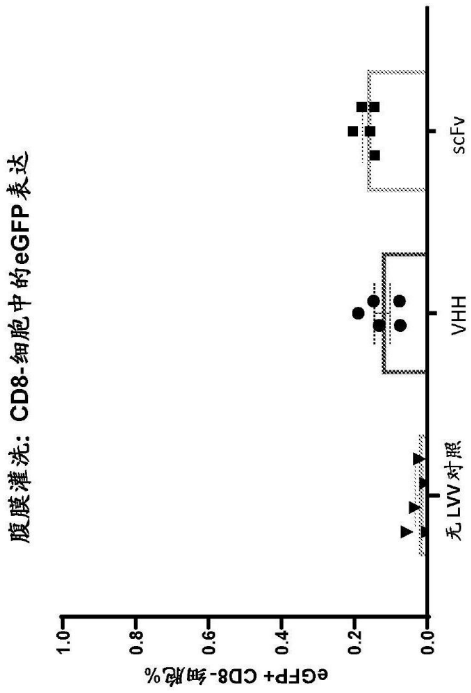

腹膜灌洗：CD8-细胞中的eGFP表达
eGFP+ CD8-细胞%
1.0
0.8
0.6
0.4
0.2
0.0
无LVV对照
VHH
scFv

图3B

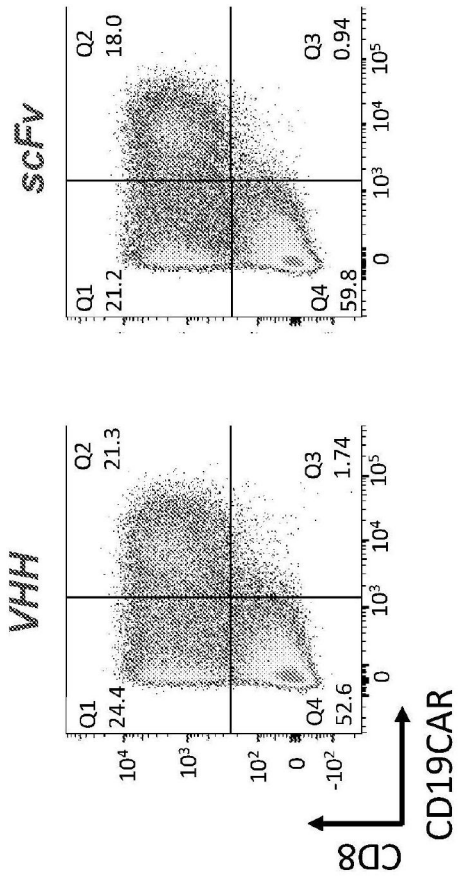

scFv
Q1 21.2
Q2 18.0
Q3 0.94
Q4 59.8
VHH
Q1 24.4
Q2 21.3
Q3 1.74
Q4 52.6
CD8
CD19CAR

图4A

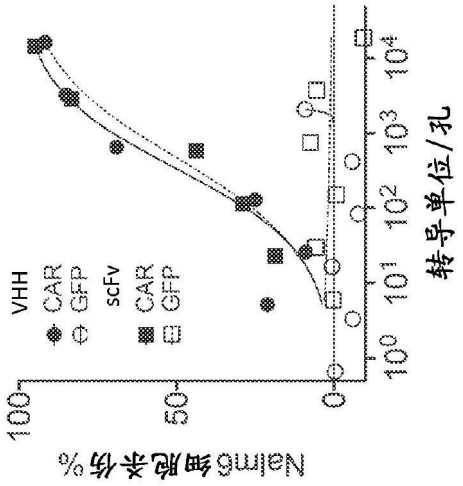

Nalm6细胞杀伤%
转导单位/孔
VHH
CAR
GFP
scFv
CAR
GFP

图4B