(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国 际 局

(43) 国际公布日
2021 年 6 月 24 日 (24.06.2021)

WIPO | PCT

(10) 国际公布号
WO 2021/121321 A1

(51) 国际专利分类号:
C07K 19/00 (2006.01) *A61K 48/00* (2006.01)
C07H 21/04 (2006.01) *C12N 15/87* (2006.01)

(21) 国际申请号: PCT/CN2020/137239

(22) 国际申请日: 2020 年 12 月 17 日 (17.12.2020)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201911310969.8 2019年12月18日 (18.12.2019) CN
201911312544.0 2019年12月18日 (18.12.2019) CN
201911312537.0 2019年12月18日 (18.12.2019) CN

(71) 申请人: 华东师范大学 (**EAST CHINA NORMAL UNIVERSITY**) [CN/CN]; 中国上海市闵行区东川路500号, Shanghai 200241 (CN)。 上海邦耀生物科技有限公司 (**BIORAY LABORATORIES INC.**) [CN/CN]; 中国上海市闵行区元江路3699号12幢, Shanghai 201109 (CN)。

(72) 发明人: 李大力 (**LI, Dali**); 中国上海市闵行区元江路3699号12幢, Shanghai (CN)。 张晓辉 (**ZHANG, Xiaohui**); 中国上海市闵行区元江路3699号12幢, Shanghai 201109 (CN)。 刘明耀 (**LIU, Mingyao**); 中国上海市闵行区元江路3699号12幢, Shanghai 201109 (CN)。 朱碧云 (**ZHU, Biyun**); 中国上海市闵行区元江路3699号12幢, Shanghai 201109 (CN)。 陈亮 (**CHEN, Liang**); 中国上海市闵行区元江路3699号12幢, Shanghai 201109 (CN)。

(54) **Title:** FUSION PROTEIN THAT IMPROVES GENE EDITING EFFICIENCY AND APPLICATION THEREOF

(54) **发明名称**: 一种提高基因编辑效率的融合蛋白及其应用

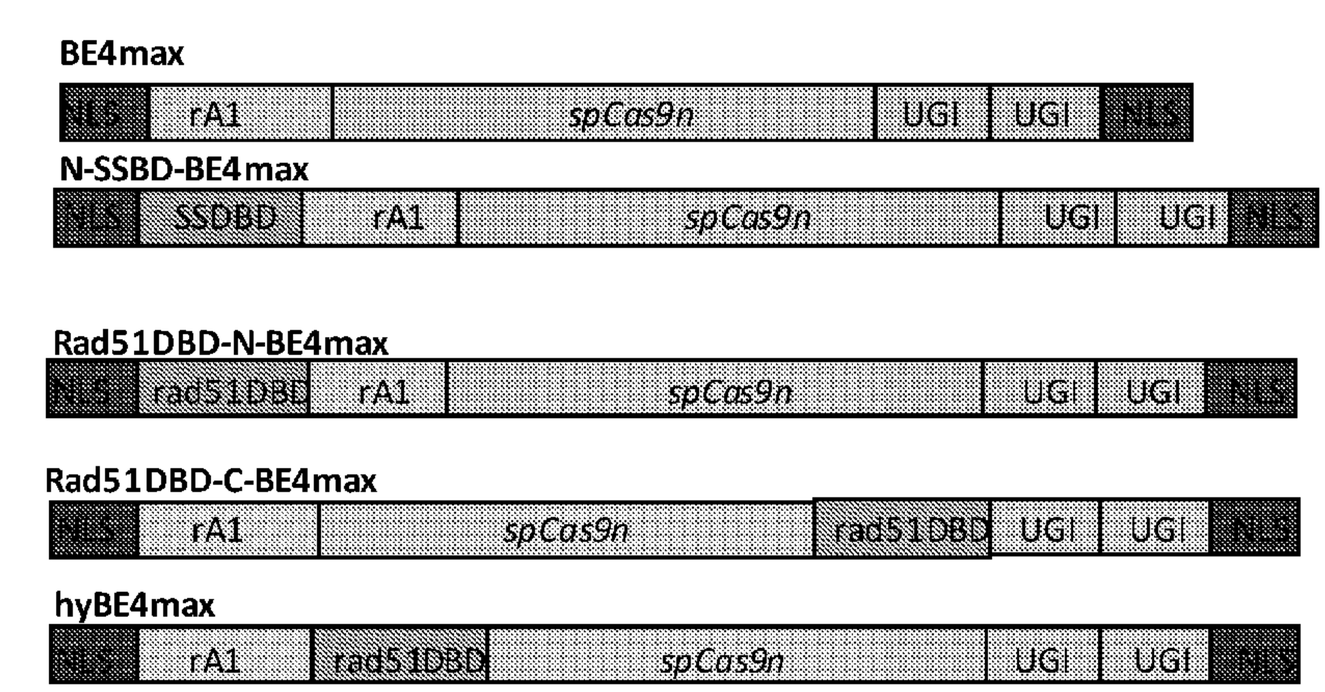


图 1

(57) **Abstract:** Provided are a fusion protein that improves gene editing efficiency and an application thereof. The fusion protein comprises a single-stranded DNA binding protein functional domain, nucleoside deaminase and nuclease. According to CBEs, when carrying out base conversion from C-G to TA, nucleoside deaminase such as cytosine deaminase carries out deamination by using single-stranded DNA as a substrate, and by re-fusing the single-stranded DNA binding protein functional domain on the fusion protein of the nucleoside deaminase and nuclease, the chance of single-stranded DNA being exposed to the nucleoside deaminase is greatly increased, thereby significantly improving base editing efficiency. The present invention provides a breakthrough improvement of single-base gene editing technology and can greatly promote the application thereof in aspects such as gene editing, gene therapy, cell therapy, animal model making, and crop genetic breeding.

(57) **摘要**: 提供一种提高基因编辑效率的融合蛋白及其应用。所述融合蛋白包括单链DNA结合蛋白功能域、核苷脱氨酶和核酸酶。根据CBEs行使C-G到T-A碱基转换过程中，核苷脱氨酶如胞嘧啶脱氨酶以单链DNA为底物进行脱氨，通过将核苷脱氨酶和核酸酶的融合蛋白上再融合单链DNA结合蛋白功能域，大大增加了单链DNA暴露于核苷脱氨酶的机会，从而明显提高了碱基编辑效率。对单碱基基因编辑技术进行了突破性的改进，可以极大地促进其在基因编辑、基因治疗、细胞治疗、动物模型制作、作物遗传育种等方面的应用。

WO 2021/121321 A1

[见续页]

(74) 代理人:北京君慧知识产权代理事务所（普通合伙）(BEIJING JUNHUI INTELLECTUAL PROPERTY AGENT OFFICE (ORDINARY PARTNERSHIP));中国北京市海淀区北四环西路66号中国技术交易大厦A座1228-1230室, Beijing 100080 (CN)。

(81) 指定国(除另有指明，要求每一种可提供的国家保护)：AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW。

(84) 指定国(除另有指明，要求每一种可提供的地区保护)：ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:
— 包括国际检索报告(条约第21条(3))。
— 包括说明书序列表部分(细则5.2(a))。

一种提高基因编辑效率的融合蛋白及其应用

技术领域

本发明涉及生物技术领域，具体说是一种提高基因编辑效率的融合蛋白及其应用。

背景技术

自 2013 年以来，以 CRISPR/Cas9 为代表的新一代基因编辑技术进入生物学领域的各个实验，改变了传统的基因操作手段。2016 年 4 月，David Liu 实验室首次报导单碱基基因编辑技术，之后，基于胞嘧啶脱氨酶原理的其它类型的单碱基基因编辑技术（如来源于七鳃鳗和人的胞嘧啶脱氨酶以不同方式与 dCas9 或 Cas9n 融合）也相继被报道。它以 CRISPR/Cas9 中来源于酿脓链球菌（*Streptococcus pyogenes*）spCas9 以 NGG 作为 PAM（间隔序列前体临近基序）并识别和特异结合 DNA 在 NGG 上游实现 C 到 T 或 G 到 A 的单碱基突变。

单碱基基因编辑技术，已被报导可用于高效地进行基因组的基因突变或修复、疾病动物模型制作、基因治疗。目前，已发现的单碱基基因编辑工具中，以 BE3（碱基编辑器 3）应用最为广泛。BE3 以最高达 37%的碱基替换效率，远远高于利用同源重组实现的效率，同时保持较低的脱靶效应，展现出其在用于基因组的单碱基突变修饰或单碱基突变治疗的巨大潜力。随着研究的深入，发现引入额外两个或者更多拷贝的 UGI（尿嘧啶糖苷酶抑制剂）到 BE3 可进一步增强其编辑效率及产物纯度。引入双分型 NLS（核定位信号）和密码子化即 BE4max，其编辑效率被一步提高。这些方法均一定程度提高其效率，但都比较有限。

β-血红蛋白病如 β 地中海贫血和镰状细胞病（SCD），是由编码 β 血红蛋白基因 HBB 突变引起的。在极少数情况下，遗传性胎儿血红蛋白病（HPFH）是一种良性遗传性疾病，成年病人体内高表达 γ-珠蛋白可以缓解因 β 血红蛋白突变而产生的疾病表型。已报导用 CRISPR / Cas9 删除 HBG1/2 的启动子区 13bp 可以激活 γ-球蛋白表达，进而缓解或治疗地中海贫血疾病，是一种有效地治疗策略。据报道，在其中一位 HPFH 患者中，HBG1/2 启动子区-117 位点的杂合点突变（G> A）产生 10-20%的胎儿血红蛋白（HbF）表达，其机制是-117G>A 的突变破坏了转录抑制因子 BCL11A 的结合位点。

现有基因编辑技术利用 CRISPR/Cas9 介导的同源重组制备因碱基替代而产生动物模型的效率还比较低。新型的单碱基基因编辑技术以 100%的效率产生疾病动物模型而备受关注，然而现有单碱基基因编辑技术通常是 C3-C8，靶向靠近 PAM 区 C 不是很有效。

发明内容

针对现有技术中存在的缺陷，本发明的目的在于提供一种提高基因编辑效率的融合蛋白及其应用。

一方面，本发明提供了一种提高基因编辑效率的融合蛋白，包括单链 DNA 结合蛋白功能域、核苷脱氨酶和核酸酶。

具体的，所述融合蛋白的连接顺序为：所述核苷脱氨酶位于所述核酸酶的 N 端或 C 端，所述单链 DNA 结合蛋白功能域位于所述核苷脱氨酶和所述核酸酶的 N 端、C 端和/或所述核苷脱氨酶和所述核酸酶之间；

优选的，所述核苷脱氨酶位于所述核酸酶的 N 端；

更优选的，所述单链 DNA 结合蛋白功能域位于所述核苷脱氨酶和所述核酸酶之间。

在上述融合蛋白中，所述单链 DNA 结合蛋白包括序列特异性单链 DNA 结合蛋白、和/或非序列特异性单链 DNA 结合蛋白，优选，非序列特异性单链 DNA 结合蛋白，

优选的，所述非序列特异性单链 DNA 结合蛋白选自 RPA70（人复制蛋白 A 的 70 亚基）、

RPA32（人复制蛋白 A 的 32 亚基）、BRCA2（乳腺癌 2 号基因）、hnRNPK(异质核核糖核蛋白 K)、PUF60（聚-U 结合剪接因子 60 KDa）和 Rad51（一种同源重组修复蛋白）中的任一种或任几种；

优选的，所述序列特异性单链 DNA 结合蛋白选自 TEBP（端粒结合蛋白）、Teb1（一种端粒酶的组成蛋白）和 POT1（人端粒保护蛋白 1）中的任一种或任几种；

优选的，所述单链 DNA 结合蛋白功能域包括如下四种结构域中的至少一种（任一种、任两种、任三种或全部四种）或如下四种结构域中具有与单链 DNA 结合功能的部分多肽片段及其任意组合：所述单链 DNA 结合蛋白的 OB 折叠（寡核苷酸/寡糖/寡肽结合折叠）、KH 结构域(K 同源结构域)、RRMS(RNA 识别基序)、涡状结构域（whirly domains）；

更优选的，所述单链 DNA 结合蛋白功能域包括 Rad51 的 DNA 结合结构域（DBD），更优选的，所述 Rad51 的 DNA 结合结构域的氨基酸序列包括 SEQ ID No.1 所示序列，更优选的，所述 Rad51 的 DNA 结合结构域的编码序列包括 SEQ ID No.2 所示序列；

更优选的，所述 RPA70 的 DNA 结合结构域的氨基酸序列包括 SEQ ID No.11 所示序列，更优选的，所述 RPA70 的 DNA 结合结构域的编码序列包括 SEQ ID No.12 所示序列。

在上述融合蛋白中，所述脱氨酶包括胞嘧啶脱氨酶（APOBEC）和/或腺苷脱氨酶，优选，胞嘧啶脱氨酶，所述胞嘧啶脱氨酶可以来源于不同的生物体，

在一些实施方式中，所述胞嘧啶脱氨酶包括来源于大鼠的胞嘧啶脱氨酶，优选的，所述来源于大鼠的胞嘧啶脱氨酶氨基酸序列包括 SEQ ID No.3 所示序列，更优选的，所述来源于大鼠的胞嘧啶脱氨酶的编码序列包括 SEQ ID No.4 所示序列；

在一些实施方式中，所述胞嘧啶脱氨酶包括来源于人的胞嘧啶脱氨酶 APOBEC3A，优选的，所述来源于人的胞嘧啶脱氨酶 APOBEC3A 的氨基酸序列包括 SEQ ID No.13 所示序列，更优选的，所述胞嘧啶脱氨酶 APOBEC3A 的编码序列包括 SEQ ID No.14 所示序列；

在一些实施方式中，所述胞嘧啶脱氨酶包括胞嘧啶脱氨酶 APOBEC3A 突变体，所述胞嘧啶脱氨酶 APOBEC3A 突变体是将所述胞嘧啶脱氨酶 APOBEC3A 第 57 位(自起始密码子)的天冬酰胺（N 或 Asn）突变为甘氨酸（G 或 Gly），优选的，该胞嘧啶脱氨酶 APOBEC3A 来源于人，更优选的，该胞嘧啶脱氨酶 APOBEC3A 的氨基酸序列包括 SEQ ID No.13 所示序列，更优选的，该胞嘧啶脱氨酶 APOBEC3A 的编码序列包括 SEQ ID No.14 所示序列；所述胞嘧啶脱氨酶 APOBEC3A 突变体的氨基酸序列包括 SEQ ID No.15 所示序列，更优选的，所述胞嘧啶脱氨酶 APOBEC3A 的编码序列包括 SEQ ID No.16 所示序列；

在上述融合蛋白中，所述核酸酶选自 Cas9、Cas3、Cas8a、Cas8b、Cas10d、Cse1、Csy1、Csn2、 Cas4、Cas10、Csm2、Cmr5、Fok1、Cpf1 中的一种或任意几种；优选的，所述核酸酶为 Cas9；更优选的，所述 Cas9 选自来源于肺炎链球菌、金黄色葡萄球菌、酿脓链球菌或嗜热链球菌的 Cas9，更优选的，所述 Cas9 选自 Cas9 突变体 VQR-spCas9、VRER-spCas9、spCas9n，更优选，spCas9n，更优选的，所述 spCas9n 的氨基酸序列包括 SEQ ID No.5 所示序列，更优选的，所述 spCas9n 的编码序列包括 SEQ ID No.6 所示序列。

在上述融合蛋白中，还包括 NLS（核定位信号），优选的，所述 NLS 位于所述融合蛋白的至少一端（C 端和/或 N 端）；更优选的，所述 NLS 的氨基酸序列包括 SEQ ID No.7 所示序列，更优选的，所述 NLS 的编码序列包括 SEQ ID No.8 所示序列；

所述融合蛋白还包括 UGI（尿嘧啶糖苷酶抑制剂），优选的，所述 UGI 位于所述融合蛋白的至少一端（C 端和/或 N 端）；更优选的，所述 UGI 的氨基酸序列包括 SEQ ID No.9 所示序

列，更优选的，所述 UGI 的编码序列包括 SEQ ID No.10 所示序列，更优选的，所述 UGI 为两个以上拷贝。

另一方面，本发明还提供了如下 A）-C）生物材料中的任一种：

A）一种基因，编码以上任一所述的融合蛋白；所述基因为 DNA 或 RNA（如 mRNA）；

B）一种重组载体，含有 A）所述基因；所述重组载体包括病毒载体、和/或非病毒载体；所述病毒载体包括腺相关病毒载体、腺病毒载体、慢病毒载体、逆转录病毒载体、和/或溶瘤病毒载体，所述非病毒载体包括阳离子高分子聚合物、质粒载体和/或脂质体；

C）一种重组细胞或重组菌，含有以上任一所述的融合蛋白，或含有 A）所述基因，所述重组菌可为工程菌，所述重组细胞可为待编辑的目的细胞，如免疫细胞（如 T 细胞）、造血干细胞、骨髓细胞、血红细胞，优选，红细胞前体细胞或造血干细胞。

另一方面，本发明还提供了一种对细胞内目的基因进行基因编辑的 sgRNA，所述 sgRNA 的靶序列包括 SEQ ID No.17-36 中的至少一种，

优选的，所述细胞为 T 细胞、造血干细胞、骨髓细胞或血红细胞，更优选，红细胞前体细胞或造血干细胞，

优选的，所述目的基因为 HBG1 和 HBG2 启动子区（具体为-117 位点的 G 编辑为 A，-117 位点的 G 即所述启动子区 CCAGCCTTGCCTTGACCAATAGCC 中左起第 14 位 G）。

另一方面，本发明还提供了一种单碱基基因编辑系统，包括以上任一所述的融合蛋白或所述的生物材料和 sgRNA，所述 sgRNA 引导所述融合蛋白对目的细胞中的目的基因进行单碱基基因编辑；

优选的，所述 sgRNA 的靶序列包括 SEQ ID No.17-36 中至少一种；

和/或，所述细胞为 T 细胞、造血干细胞、骨髓细胞、血红细胞、或红细胞前体细胞，

和/或，所述目的基因为 HBG1 和 HBG2 启动子区。

另一方面，本发明保护以上任一所述融合蛋白、所述的生物材料、所述的单碱基基因编辑系统在制备基因编辑产品、疾病治疗和/预防产品、动物模型、或植物新品种中的应用；

在一些实施例中，所述疾病为β血红蛋白病，所述β血红蛋白病包括β地中海贫血和/或镰刀型细胞贫血症。

另一方面，本发明还提供了一种提高单碱基基因编辑效率的方法，包括利用以上任一所述的融合蛋白和 sgRNA 引入细胞、对目的基因进行基因编辑的步骤，所述 sgRNA 引导所述融合蛋白对所述目的基因进行单碱基基因编辑。

在上述方法中，优选的，所述 sgRNA 的靶序列包括 SEQ ID No.17-36 中的至少一种；

和/或，所述细胞为 T 细胞、造血干细胞、骨髓细胞、血红细胞、或红细胞前体细胞；

和/或，所述目的基因为 HBG1 和 HBG2 启动子区。

另一方面，本发明还提供了一种疾病动物模型的构建方法，包括利用以上任一所述的融合蛋白和 sgRNA 引入动物细胞、对目的基因进行基因编辑的步骤；

优选的，所述 sgRNA 的靶序列包括 SEQ ID No.17-36 中的至少一种；更优选的，所述 sgRNA 的靶序列包括 SEQ ID No.36 所示序列，所述目的基因包括 DMD 基因；

优选的，所述动物为哺乳动物，更优选的，所述哺乳动物为大鼠或小鼠，更优选小鼠；

优选的，所述细胞为胚胎细胞；

优选的，所述引入的方式为载体转化、显微注射、转染、脂质转染、热休克、电穿孔、转导、基因枪、DEAE-葡聚糖介导的转移中的一种或任几种组合，更优选，显微注射；

优选的，所述引入使用以上任一所述的融合蛋白的 mRNA 和所述 sgRNA 进行，

更优选的，所述引入使用的以上任一所述的融合蛋白的 mRNA 的浓度为 1-1000ng/μL，更优选，10-600ng/μL，更优选，50-150ng/μL，更优选，100ng/μL，所述引入使用的所述 sgRNA 的浓度为 1-1000ng/μL，更优选，10-600ng/μL，更优选，150-250ng/μL，更优选，200ng/μL，

更优选的，所述引入使用的以上任一所述融合蛋白的 mRNA 与所述引入使用的所述 sgRNA 的浓度比为 1:（5-1），更优选，1:（4-1.5），更优选，1:（3-1.8），更优选，1:2。

本发明保护所述方法得到的动物模型在药物筛选、疾病治疗效果评价或疾病治病机理研究中的应用。

另一方面，本发明提供了一种用于治疗和/或预防β血红蛋白病的产品，包含：以上 A）中所述基因和 sgRNA 的递送载体，

所述 sgRNA 引导所述融合蛋白对目的细胞中 HBG1 和 HBG2 启动子区（具体为-117 位点的 G 编辑为 A，-117 位点的 G 即所述启动子区 CCAGCCTTGCCTTGACCAATAGCC 中左起第 14 位 G）进行单碱基基因编辑；

优选的，所述 sgRNA 的靶序列包括 SEQ ID No.35 所示序列；

优选的，所述β血红蛋白病包括β地中海贫血和/或镰刀型细胞贫血症；

优选的，所述细胞为 T 细胞、造血干细胞、骨髓细胞或血红细胞，更优选，红细胞前体细胞或造血干细胞，

优选的，所述目的基因为 HBG1 和 HBG2 启动子区（具体为-117 位点的 G 编辑为 A，-117 位点的 G 即所述启动子区 CCAGCCTTGCCTTGACCAATAGCC 中左起第 14 位 G）。

在上述产品中，所述递送载体包括病毒载体、和/或非病毒载体；所述病毒载体包括腺相关病毒载体、腺病毒载体、慢病毒载体、逆转录病毒载体、和/或溶瘤病毒载体，所述非病毒载体包括阳离子高分子聚合物、质粒载体和/或脂质体；

优选的，所述递送载体包括慢病毒载体。

以上所述“至少一种”为其限定的所有种类中的任一种、任两种组合、任三种组合、……、或所有种类组合，均在本申请的保护范围之内。

以上所述氨基酸序列或编码序列中，与本申请所述序列的同源性在 80%以上、85%以上、90%以上、95%以上、96%以上、97%以上、98%以上、或 99%以上的序列，和/或在本申请所述序列的基础上进行氨基酸残基或核苷酸的替换、删除或插入后的序列，且具有本申请所使用序列具有相同或相近功能的序列，均在本申请的保护范围之内。

本发明的有益效果如下：

本发明根据 CBEs（嘧啶碱基转换技术）行使 C-G 到 T-A 碱基转换过程中，核苷脱氨酶如胞嘧啶脱氨酶以单链 DNA 为底物进行脱氨，通过将核苷脱氨酶和核酸酶的融合蛋白上再融合单链 DNA 结合蛋白功能域，大大增加了单链 DNA 暴露于核苷脱氨酶的机会，从而明显提高了碱基编辑效率。

本发明通过筛选 10 种非序列偏好的单链 DNA 结合蛋白的功能域与 BE4max 进行融合，发现来源于人的 Rad51 的一个单链 DNA 结合功能域（1-114AA）融合在 Apobec1 与 Cas9n 的中间展示出最高效率提高，命名为 hyBE4max。相对于 BE4max，hyBE4max 的 C-G 到 T-A 编辑效率最大提高了 16 倍，尤其是靠近 PAM 区的位点效率提高更为明显，同时保持较低的 indels（插入或缺失）。

本发明对单碱基基因编辑技术进行了突破性的改进，可以极大地促进其在基因编辑、基因

治疗、细胞治疗、动物模型制作、作物遗传育种等方面的应用。

本发明在基因治疗中，以 β-血红蛋白病为例，相对于 eA3A-BE4max，A3A-BE4max，hyeA3A-BE4max 靶向 HBG1 和 HBG2（以下简称 HBG1/2）启动子区较靠近 PAM 区-117 位点，能更加精确高效地靶向-117 产生 G 到 A 的突变，从而激活 γ-珠蛋白的表达，为临床治疗 β-血红蛋白病提供了一种更加精准高效的治疗策略。

本发明将 hyA3A-BE4max 运用到制作小鼠疾病动物模型中，相对于 A3A-BE4max，hyA3A-BE4max 靶向靠近 PAM 区的 C 到 T 的突变产生疾病动物模型更加有效，因此本发明提供了的一个新型的高效产生疾病动物模型制作的平台，将极大地促进不同动物模型的生产进程。

附图说明

图 1 为不同单链 DNA 结合蛋白功能域与 BE4max 融合的结构示意图。其中，NLS 为核定位信号（其氨基酸序列如 SEQ ID No.7 所示，编码序列如 SEQ ID No.8 所示），rA1 为胞苷脱氨酶 APOBEC1（其氨基酸序列如 SEQ ID No.3 所示，编码序列如 SEQ ID No.4 所示），spCas9n 为源于酿脓链球菌（*Streptococcus pyogenes*）的 Cas9n（其氨基酸序列如 SEQ ID No.5 所示，编码序列如 SEQ ID No.6 所示），UGI 为尿嘧啶糖苷酶抑制剂（其氨基酸序列如 SEQ ID No.9 所示，编码序列如 SEQ ID No.10 所示），SSDBD 为一种单链 DNA 结合蛋白功能域。

图 2 为 hyBE4max 与 BE4max 在 293T 上 8 个靶点实现的 C 到 T 碱基编辑效率（即纵坐标，单位为%）对比。

图 3 为 hyBE4max 与 BE4max 在 293T 上 8 个靶点产生的平均 C 到 T 碱基编辑效率（即纵坐标，单位为%）对比。

图 4 为 hyBE4max 与 BE4max 在 293T 上 8 个靶点产生的 indels 的碱基编辑效率（即纵坐标，单位为%）对比。

图 5 为融合蛋白 A3A-BE4max 和 hyA3A-BE4max 的结构示意图。其中，hA3A 为来源于人的胞苷脱氨酶 APOBEC3A（其氨基酸序列如 SEQ ID No.13 所示，编码序列如 SEQ ID No.14 所示），NLS、spCas9n 和 UGI 同图 1。

图 6 为 hyA3A-BE4max 与 A3A-BE4max 在 293T 上 8 个内源性靶点实现的 C 到 T 碱基编辑效率（即纵坐标，单位为%）的对比。

图 7 为 hyA3A-BE4max 与 A3A-BE4max 在 293T 上 8 个内源性靶点实现的 C 到 T 碱基编辑效率（即纵坐标，单位为%）的平均值对比。

图 8 为 hyA3A-BE4max 与 A3A-BE4max 在 293T 上 8 个内源性靶点产生的 indels 的碱基编辑效率（即纵坐标，单位为%）对比。

图 9 为融合蛋白 eA3A-BE4max 和 hyeA3A-BE4max 的结构示意图。其中，A3A N57G 为图 5 中所用 hA3A 的 N57G 突变体（其氨基酸序列如 SEQ ID No.15 所示，编码序列如 SEQ ID No.16 所示），NLS、spCas9n 和 UGI 同图 1。

图 10 为 hyeA3A-BE4max 与 eA3A-BE4max 在 293T 上 11 个内源性靶点实现的 C 到 T 碱基编辑效率（即纵坐标，单位为%）的对比。

图 11 为 hyeA3A-BE4max 与 eA3A-BE4max 在 293T 上 11 个内源性靶点实现的 C 到 T 碱基编辑效率（即纵坐标，单位为%）的平均值对比。

图 12 为 hyeA3A-BE4max 与 eA3A-BE4max 在 293T 上 11 个内源性靶点产生的 indels 的碱基编辑效率（即纵坐标，单位为%）对比。

其中，图 2、3、5、6、11 中横坐标的 C 及其后面的数字代表被编辑为 T 的 C 在相应靶点序列上的位置，如 C5 代表自相应靶点序列 5'端的第 5 位 C 被编辑为 T 的效率。

图 13 为 hyeA3A-BE4max 靶向 HBG1 / 2 启动子区-117G 示意图。其中，-117 G> A 突变显示为红色，转录因子 BCL11A 结合位点的核心序列用方框表示，PAM 序列为蓝色，G> A 转化破坏了转录抑制因子 BCL11A 的结合位点并激活了 HUDEP-2($\Delta^{G}\gamma$)中 HBG1 / 2 的表达。

图 14 为 hyeA3A-BE4max、eA3A-BE4max、A3A-BE4max、hyA3A-BE4max 在 HEK293T 细胞中靶向 HBG-117G 靶点实现的 C 到 T 碱基编辑效率（即纵坐标，单位为%）对比。

图 15 为慢病毒载体 Lenti-117G-hyA3A-BE4max-P2A-GFP 和 Lenti-117G-hyeA3A-BE4max -P2A-GFP 构建示意图。其中，sgRNA 的靶点序列为 HBG-117G。

图 16 为 hyeA3A-BE4max 和 hyA3A-BE4max 在 HUDEP-2($\Delta^{G}\gamma$)细胞中靶向 HBG-117G 靶点实现的 C 到 T 碱基编辑效率（即纵坐标，单位为%）对比。

图 17 为感染慢病毒 Lenti-117G-hyeA3A-BE4max-P2A-GFP 和 Lenti-117G-hyA3A-BE4max -P2A -GFP 的 HUDEP-2($\Delta^{G}\gamma$)细胞分化后珠蛋白 mRNA 表达对比。其中，****表示差异显著性水平 P＜0.0001。

图 18 为 hyA3A-BE4max 靶向杜氏肌营养不良（DMD）基因的动物模型构建示意图。

图 19 为 A3A-BE4max 和 hyA3A-BE4max 显微注射后产生 F0 高通量测序结果的比对。

图 20 为 A3A-BE4max 和 hyA3A-BE4max 注射产生 F0 中的含有 TAA 终止密码子的 Reads 平均比率。

图 21 为将产生的 F0 小鼠进行免疫荧光染色检测抗肌萎缩蛋白（Dystrophin）的表达。

图 22 为 DMD 突变小鼠的生殖系遗传(F0→F1)。

图 23 为 hyA3A-BE4max 与 DMD-sg3 的预测脱靶位点组合在 F0 代上的脱靶分析。

具体实施方式

一、融合 Rad51DBD(1-114aa)单链 DNA 结合蛋白的功能域的 BE4max 编辑效率提高最为明显

1.1、质粒设计及构建

1.1.1、根据单碱基编辑技术中 CBEs 的 Apobec1 以单链 DNA 作为底物的特性，我们设计了 10 个来源于人的非序列偏好性的单链 DNA 结合蛋白不同的功能域（主要为 RPA70（630aa）-A、RPA70-B、RPA70-AB、RPA70-C、RPA32-D、BRCA2-OB2、BRCA2-OB3、HNRNPK KH 结构域、PUF60 RRM、Rad51 DBD）(表 1)，由于已报道的融合蛋白放于 BE4max（图 1 中由上至下的第一个图）的 C 端倾向于没有活性，因此，将上述这些功能域融合在 BE4max 的 N 端(图 1 中由上至下的第二个图)，同时设计了两个来自于人的内源性靶点 EMX1 site1、Tim3-sg1（序列如表 2 所示）。

1.1.2、人工合成表 1 所示 10 个来源于人的非序列偏好性的单链 DNA 结合蛋白不同的功能域的 DNA，之后进行无缝克隆组装至质粒 pCMV-BE4max (addgene，#112093)中 BE4max 的 N 端，分别构建了 10 个重组质粒（图 1）：pRPA70-A-BE4max、pRPA70-B-BE4max、pRPA70-AB-BE4max、pRPA70-C-BE4max、pRPA32-D-BE4max、pBRCA2-OB2-BE4max、pBRCA2-OB3-BE4max、pKH-BE4max、pRRM-BE4max、pRad51DBD-BE4max。

人工合成表 2 所示靶点 EMX1 site1 和 Tim3-sg1 的 DNA，分别连接至 sgRNA 表达质粒 U6-sgRNA-EF1α-GFP 的 Bbs Ⅰ位点处（用于表达相应靶点的 sgRNA），得到重组质粒 pE 和 pT。

1.1.3、将 1.1.1 与 1.1.2 中构建的质粒经 sanger 测序，确保完全正确。

表 1、所用的单链 DNA 结合蛋白不同的功能域序列

名称	序列
RPA70-A	编码序列（5’-3’）： Cagtccaaagtggtgcccattgccagcctcactccttaccagtccaagtggaccatttgtgctcgtgttaccaaca aaagtcagatccgtacctggagcaactcccgaggggaagggaagcttttctccctagaactggttgacgaaagt ggtgaaatccgagctacagctttcaatgagcaagtggacaagttctttcctcttattgaagtgaacaaggtgtattat ttctcgaaaggcaccctgaagattgctaacaagcagttcacagctgttaaaaatgactacgagatgaccttcaata acgagacttccgtcatgccctgtgag 氨基酸序列： QSKVVPIASLTPYQSKWTICARVTNKSQIRTWSNSRGEGKLFSLELVDES GEIRATAFNEQVDKFFPLIEVNKVYYFSKGTLKIANKQFTAVKNDYEMT FNNETSVMPCE
RPA70-B	编码序列（5’-3’）： Gatttcacggggattgatgacctcgagaacaagtcgaaagactcacttgtagacatcatcgggatctgcaagag ctatgaagacgccactaaaatcacagtgaggtctaacaacagagaagttgccaagaggaatatctacttgatgga cacatccgggaaggtggtgactgctacactgtgggggggaagatgctgataaatttgatggttctagacagcccgt gttggctatcaaaggagcccgagtctctgatttcggtggacggagcctctccgtgctgtcttcaagcactatcattg cgaatcctgacatcccagaggcctataagcttcgtggatggtttgacgcagaaggacaagcctta 氨基酸序列： DFTGIDDLENKSKDSLVDIIGICKSYEDATKITVRSNNREVAKRNIYLMD TSGKVVTATLWGEDADKFDGSRQPVLAIKGARVSDFGGRSLSVLSSSTI IANPDIPEAYKLRGWFDAEGQAL
RPA70-AB	编码序列（5’-3’）：RPA70-A+RPA70-B 氨基酸序列：RPA70-A+RPA70-B
RPA70-C	编码序列（5’-3’）： Ggagggagtaacaccaactggaaaaccttgtatgaggtcaaatccgagaacctgggccaaggcgacaagcc ggactactttagttctgtggccacagtggtgtatcttcgcaaagagaactgcatgtaccaagcctgcccgactcag gactgcaataagaaagtgattgatcaacagaatggattgtaccgctgtgagaagtgcgacaccgaatttcccaatt tcaagtaccgcatgatcctgtcagtaaatattgcagattttcaagagaatcagtgggtgacttgtttccaggagtctg ctgaagctatccttggacaaaatgctgcttatcttggggaattaaaagacaagaatgaacaggcatttgaagaagt tttccagaatgccaacttccgatctttcatattcagagtcagggtcaaagtggagacctacaacgacgagtctcga attaaggccactgtgatggacgtgaagcccgtggactacagagagtatggccgaaggctggtcatgagcatca ggagaagtgcattgatg（SEQ ID No.12） 氨基酸序列： GGSNTNWKTLYEVKSENLGQGDKPDYFSSVATVVYLRKENCMYQACP TQDCNKKVIDQQNGLYRCEKCDTEFPNFKYRMILSVNIADFQENQWVT CFQESAEAILGQNAAYLGELKDKNEQAFEEVFQNANFRSFIFRVRVKVE TYNDESRIKATVMDVKPVDYREYGRRLVMSIRRSALM（SEQ ID No.11）
RPA32-D	编码序列（5’-3’）：

	Gcccagcacattgtgccctgtactatatctcagctgctttctgccactttggttgatgaagtgttcagaattgggaat gttgagatttcacaggtcactattgtggggatcatcagacatgcagagaaggctccaaccaacattgtttacaaaat agatgacatgacagctgcacccatggacgttcgccagtgggttgacacagatgacaccagcagtgaaaacact gtggttcctccagaaacatatgtgaaagtggcaggccacctgagatcttttcagaacaaaaagagcctggtagcc tttaagatcatgcccctggaggatatgaatgagttcaccacacatattctggaagtgatcaatgcacacatggtact aagcaaa 氨基酸序列： AQHIVPCTISQLLSATLVDEVFRIGNVEISQVTIVGIIRHAEKAPTNIVYKI DDMTAAPMDVRQWVDTDDTSSENTVVPPETYVKVAGHLRSFQNKKSL VAFKIMPLEDMNEFTTHILEVINAHMVLSK
BRCA2-OB2	编码序列（5’-3’）： Ttatcatcgcttttcagtgatggaggaaatgttggttgtgttgatgtaattattcaaagagcataccctatacagtgga tggagaagacatcatctggattatacatatttcgcaatgaaagagaggaagaaaaggaagcagcaaaatatgtg gaggcccaacaaaagagactagaagccttattcactaaaattcaggaggaatttgaagaacatgaagaaaacac aacaaaaccatatttaccatcacgtgcactaacaagacagcaagttcgtgctttgcaagatggtgcagagctttat gaagcagtgaagaatgcagcagacccagcttaccttgagggttatttcagtgaagagcagttaagagccttgaat aatcacaggcaaatgttgaatgataagaaacaagctcagatccagttggaaattaggaaggccatggaatctgct gaacaaaaggaacaaggtttatcaagggatgtcacaaccgtgtggaagttgcgtattgtaagctattcaaaaaaa gaaaaagattcagttatactgagtatttggcgtccatcatcagatttatattctctgttaacagaaggaaagagatac agaatttatcatcttgcaacttcaaaatctaaaagtaaatctgaaagagctaacatacagttagcagcgacaaaaaa aactcagtatcaacaactaccggtttcagatgaaattttatttcagatttaccagccacgggagccc 氨基酸序列： LSSLFSDGGNVGCVDVIIQRAYPIQWMEKTSSGLYIFRNEREEEKEAAK YVEAQQKRLEALFTKIQEEFEEHEENTTKPYLPSRALTRQQVRALQDG AELYEAVKNAADPAYLEGYFSEEQLRALNNHRQMLNDKKQAQIQLEIR KAMESAEQKEQGLSRDVTTVWKLRIVSYSKKEKDSVILSIWRPSSDLYS LLTEGKRYRIYHLATSKSKSKSERANIQLAATKKTQYQQLPVSDEILFQI YQPREP
BRCA2-OB3	编码序列（5’-3’）： Gacctaataggatttgtcgtttctgttgtgaaaaaaacaggacttgccccttttcgtctatttgtcagacgaatgttaca atttactggcaataaagttttggatagaccttaatgaggacattattaagcctcatatgttaattgctgcaagcaacct ccagtggcgaccagaatccaaatcaggccttcttactttatttgctggagatttttctgtgttttctgctagtccaaaag agggccactttcaagagacattcaacaaaatgaaaaatactgttgag 氨基酸序列： DLIGFVVSVVKKTGLAPFVYLSDECYNLLAIKFWIDLNEDIIKPHMLIA ASNLQWRPESKSGLLTLFAGDFSVFSASPKEGHFQETFNKMKNTVE
hnRNPK KH 结构域	编码序列（5’-3’）： Aacactgatgagatggttgaattacgcattctgcttcagagcaagaatgctggggcagtgattggaaaaggagg caagaatattaaggctctccgtacagactacaatgccagtgtttcagtcccagacagcagtggccccgagcgcat attgagtatcagtgctgatattgaaacaattggagaaattctgaagaaaatcatccctaccttggaagag

	氨基酸序列： NTDEMVELRILLQSKNAGAVIGKGGKNIKALRTDYNASVSVPDSSGPE RILSISADIETIGEILKKIIPTLEE
PUF60 RRM	编码序列（5’-3’）： Tgccgcgtctacgtgggctctatctactatgagctgggggaggacaccatccgccaggcctttgccccctttgg ccccatcaagagcatcgacatgtcctgggactccgtcaccatgaagcacaagggctttgccttcgtggagtatga ggtccccgaagctgcacagctggccttggagcagatgaactcggtgatgctgggggggcaggaacatcaaggt gggcagacccagcaac 氨基酸序列： CRVYVGSIYYELGEDTIRQAFAPFGPIKSIDMSWDSVTMKHKGFAFVEY EVPEAAQLALEQMNSVMLGGRNIKVGRPSN
Rad51 DBD	编码序列（5’-3’）： Atggcaatgcagatgcagcttgaagcaaatgcagatacttcagtggaagaagaaagctttggcccacaacccat ttcacggttagagcagtgtggcataaatgccaacgatgtgaagaaattggaagaagctggattccatactgtgga ggctgttgcctatgcgccaaagaaggagctaataaatattaagggaattagtgaagccaaagctgataaaattctg gctgaggcagctaaattagttccaatgggtttcaccactgcaactgaattccaccaaaggcggtcagagatcata cagattactactggctccaaagagcttgacaaactacttcaa（SEQ ID No.2） 氨基酸序列： MAMQMQLEANADTSVEEESFGPQPISRLEQCGINANDVKKLEEAGFHT VEAVAYAPKKELINIKGISEAKADKILAEAAKLVPMGFTTATEFHQRRSE IIQITTGSKELDKLLQ（SEQ ID No.1）

表 2、所用靶点及序列

靶点名称	序列（5’-3’）	SEQ ID No.
EMX1 site1	GAGTCCGAGCAGAAGAAGAAGGG	17
Tim3-sg1	TTCTACACCCCAGCCGCCCCAGG	18
VEGFA site2	GACCCCCTCCACCCCGCCTCCGG	19
Lag3-sg2	CGCTACACGGTGCTGAGCGTGGG	20
HEK3	GGCCCAGACTGAGCACGTGATGG	21
HEK4	GGCACTGCGGCTGGAGGTGGGGG	22
EMX1-sg2p	GACATCGATGTCCTCCCCATTGG	23
Nme1-sg1	AGGGATCGTCTTTCAAGGCGAGG	24

1.2、细胞转染

将 HEK293T 5×10^5 细胞铺 24 孔板，待细胞长至 70%-80%时，按 pssDBD-BE4max:pE（或 pT）=750ng:250ng 进行质粒组合的转染，每种质粒组合转染设 3 孔重复，每孔 2×10^5 个细胞。同时设不转染任何质粒的空白对照。

pssDBD-BE4max 代表：质粒 pRPA70-A-BE4max、pRPA70-B-BE4max、pRPA70-AB-BE4max、pRPA70-C-BE4max、pRPA32-D-BE4max、pBRCA2-OB2-BE4max、

pBRCA2-OB3-BE4max、pKH-BE4max、pRRM-BE4max、pRad51DBD-BE4max 中任一种，以质粒 pCMV-BE4max 为阴性对照。

1.3、基因组提取及扩增子文库的准备

转染后 72h，用天根细胞基因组提取试剂盒（DP304）提取细胞基因组 DNA。之后用 Hitom 试剂盒的操作流程，设计相对应的鉴定引物（表 3），即在正向鉴定引物 5’端加上搭桥序列 5’-ggagtgagtacggtgtgc-3’，反向鉴定引物 5’端加上搭桥序列 5’-gagttggatgctggatgg-3’，即得到一轮 PCR 产物，之后利用一轮 PCR 产物作为模板，进行二轮 PCR，后混在一起进行切胶回收纯化后进行送公司进行深度测序。

表 3、所用靶点的鉴定引物

靶点名称	序列（5’-3’）
EMX1 site1	F:ggagtgagtacggtgtgcGTGGTTCCAGAACCGGAGGACAAAG
	R: gagttggatgctggatggGTTTGTGGTTGCCCACCCTAGTCAT
Tim3-sg1	F:ggagtgagtacggtgtgcCGCTTGAGTCTTGGCTCTCCTTCTC
	R:gagttggatgctggatggCACCACGTTGCCACATTCAAACACA
VEGFA site2	F: ggagtgagtacggtgtgcGACAGACAGACAGACACCGCCC
	R: gagttggatgctggatggACAGCCCAGAAGTTGGACGAAAAGT
Lag3-sg2	F: ggagtgagtacggtgtgcTTCCTACCCCTGGAGCTTCTCAACT
	R: gagttggatgctggatggCCTCCGGGACCCACGCTCAG
HEK3	F: ggagtgagtacggtgtgcAGGGAAACGCCCATGCAATTAGTCT
	R: gagttggatgctggatggCCCTGTCTAGGAAAAGCTGTCCTGC
HEK4	F: ggagtgagtacggtgtgcCAGAGGGTCCAAAGCAGGATGACAG
	R: gagttggatgctggatggCTTTCAACCCGAACGGAGACACACA
EMX1-sg2p	F: ggagtgagtacggtgtgcGTGGTTCCAGAACCGGAGGACAAAG
	R: gagttggatgctggatggGTTTGTGGTTGCCCACCCTAGTCAT
Nme1-sg1	F: ggagtgagtacggtgtgcGGGGAGGCAGACACACAAACAGAAA
	R: gagttggatgctggatggGCGCTCATGACCTACCCTGTATCAC

1.4、深度测序结果分析与统计

利用 BE-analyzer 网站分析步骤 1.3 的深度测序结果，统计 C 到 T、Indels 的比率，结果如表 4 和表 5 所示。

结果表明：与 BE4max 相比，融合 Rad51 单链 DNA 结合蛋白的功能域的 BE4max（Rad51DBD-N-BE4max 或 Rad51DBD-BE4max），对靶点上的 C 到 T 的编辑效率提高最为明显，其次为融合 RPA70-C 单链 DNA 结合蛋白的功能域的 BE4max。

二、hyBE4max 的编辑效率最优

为了进一步测试步骤一中对靶点上的 C 到 T 的编辑效率提高最高的 Rad51 单链 DNA 结合蛋白功能域的融合位置，将 Rad51 DBD 融合于 BE4max 的另两个不同位置，将融合了 Rad51 DBD 的 BE4max（图 1 中的由上至下的第三至五图）共三种重组质粒分别与重组质粒 pE 或 pT 按照步骤一中的 1.2 的方法进行细胞转染，并按照步骤一中的 1.3 和 1.4 的方法获得编辑效率结果（表 4 和表 5）。

图 1 中由上至下的第三至五图所示三种融合了 Rad51 DBD 的 BE4max 分别如下：

Rad51DBD-N-BE4max：在 BE4max 中 NLS 和 rA1 之间融合 Rad51 DBD，即 Rad51 DBD 位于 rA1 和 spCas9n 的 N 端；

Rad51DBD-C-BE4max：在 BE4max 中 spCas9n 和 UGI 之间融合 Rad51 DBD，即 Rad51 DBD

位于 rA1 和 spCas9n 的 C 端；

hyBE4max：在 BE4max 中 rA1 和 spCas9n 之间融合 Rad51 DBD。

表 4、靶点 EMX1 site1 的编辑效率结果（单位，%）

碱基编辑器	重复	C5	C6
BE4max	Rep.1	46.7	45.2
	Rep.2	44.9	42.9
	Rep.3	62.5	60.7
RPA70-A-BE4max	Rep.1	41	36.9
	Rep.2	41	36.9
	Rep.3	41.9	37.2
RPA70-B-BE4max	Rep.1	42.7	36.2
	Rep.2	42.6	36.3
	Rep.3	41	33.3
RPA70-AB-BE4max	Rep.1	48.4	47
	Rep.2	47.1	45.5
	Rep.3	45.7	44.1
RPA70-C-BE4max	Rep.1	65.2	63
	Rep.2	64.9	62.8
	Rep.3	64.7	62
RPA32-D-BE4max	Rep.1	41.8	37.3
	Rep.2	41.3	36.9
	Rep.3	41	36.7
BRCA2-OB2-BE4max	Rep.1	30	20.6
	Rep.2	30	20.5
	Rep.3	31	20.8
BRCA2-OB3-BE4max	Rep.1	32.5	25.8
	Rep.2	32.4	25.8
	Rep.3	32.1	25
KH-BE4max	Rep.1	34.8	21.7
	Rep.2	34.6	21.6
	Rep.3	31.9	19.3
RRM-BE4max	Rep.1	34.8	25.2
	Rep.2	35.2	25.3
	Rep.3	34.7	24.3
Rad51DBD-BE4max	Rep.1	68	66.4
	Rep.2	67.7	66.2
	Rep.3	65.9	60.6
hyBE4max	Rep.1	80.8	75.1

	Rep.2	81.3	75.5
	Rep.3	70	64.5
BE4max-C-Rad51	Rep.1	41.5	38.9
	Rep.2	42.3	39.8
	Rep.3	45.2	42.9
空白对照	Rep.1	0.1	0
	Rep.2	0	0
	Rep.3	0	0

表 5、靶点 Tim3-sg1 的编辑效率结果（单位，%）

碱基编辑器	重复	C3	C6	C8	C9	C10	C11
BE4max	Rep.1	24.1	25.2	21.3	17.9	5.6	2.5
	Rep.2	23.7	24.7	20.8	17.4	5.6	2.5
	Rep.3	24.3	25.9	21.4	17.9	5.2	2.5
RPA72-A-BE4max	Rep.1	17.7	17.7	11.9	10.5	7.1	4.7
	Rep.2	18	17.8	12	10.6	7.3	4.8
	Rep.3	17.9	17.6	12.3	10.7	7.5	5
RPA72-B-BE4max	Rep.1	15.4	14.2	9.9	8.3	5	2.9
	Rep.2	15.9	14.6	10.1	8.6	5.1	2.9
	Rep.3	15.1	14.3	9.4	7.8	4.7	2.7
RPA72-AB-BE4max	Rep.1	17.2	11.7	5.8	3.7	1.4	0.7
	Rep.2	17.3	11.6	5.8	3.7	1.4	0.6
	Rep.3	16.6	9.8	4.3	2.6	0.9	0.4
RPA70-C-BE4max	Rep.1	27.8	29.5	20.4	15.4	3.5	1.8
	Rep.2	28.1	29.4	20.9	15.7	3.8	1.9
	Rep.3	30.1	31.7	21.4	17	4.2	1.5
RPA32-D-BE4max	Rep.1	17.5	15.8	9.7	8.6	5.6	3.2
	Rep.2	17.9	16	9.9	8.8	5.8	3.3
	Rep.3	16.1	13.4	8.7	7.3	4.7	2.9
BRCA2-OB2-BE4max	Rep.1	9.2	9.3	7.9	7.3	6.2	5.6
	Rep.2	9.2	9.3	7.9	7.3	6.2	5.6
	Rep.3	8.8	9	7.4	7	5.6	4.7
BRCA2-OB3-BE4max	Rep.1	10.7	10.4	7.4	5.9	3.7	2.8
	Rep.2	13.1	10.1	8.2	7.7	4	2.6
	Rep.3	10.4	10	7.5	6.9	4.8	3.3
KH-BE4max	Rep.1	6.6	5.9	3.9	3.1	2.1	1.5
	Rep.2	6.5	5.9	4.1	3.1	2	1.5
	Rep.3	6.7	6.7	4.6	3.3	2.3	1.5
RRM-BE4max	Rep.1	8.1	6.9	4.8	3.8	2.5	1.6
	Rep.2	8	6.6	4.8	3.7	2.4	1.5

	Rep.3	7.7	7.1	5.2	4.5	3	2
Rad51DBD-BE4max	Rep.1	39.4	40.6	33.1	31.2	24.4	16.8
	Rep.2	39.2	40.6	32.7	30.7	23.6	16.6
	Rep.3	34.3	34.7	27.8	25.9	19.7	13.2
hyBE4max	Rep.1	49.5	49.2	49	50.1	49.1	47
	Rep.2	48.8	48.6	48.1	49.7	48.6	46.8
	Rep.3	43.3	42.6	43.4	44.5	44.1	41.8
BE4max-C-Rad51	Rep.1	18.8	20.7	15.6	12.6	3.3	1.8
	Rep.2	18.9	20.4	15.6	12.7	3.5	1.9
	Rep.3	21.1	22.5	16.4	12.9	3.1	1.7
空白对照	Rep.1	0.1	0	0	0	0	
	Rep.2	0	0	0	0	0	0
	Rep.3	0	0	0	0	0	0

表 4 和表 5 的结果表明：与在 BE4max 中 NLS 和 rA1 之间融合 Rad51 DBD（即 Rad51DBD-N-BE4max）相比，在 BE4max 中 rA1 和 spCas9n 之间融合 Rad51 DBD（即 hyBE4max）对靶点上的 C 到 T 的编辑效率提高最为明显。

三、hyBE4max 的工作特性

为了进一步公平地描述 hyBE4max 的工作特性，设计另外 6 个额外的靶点 VEGFA site2、Lag3-sg2、HEK3、HEK4、EMX1-sg2p、Nme1-sg1（序列如表 2），连接至质粒 U6-sgRNA-EF1α-GFP 的 BbsI 位点处，得到重组质粒 pV、pL、pH3、pH4、pEP 和 pN。质粒经 sanger 测序，确保完全正确。

将步骤二中含有 hyBE4max 的重组质粒分别与重组质粒 pE、pT、pV、pL、pH3、pH4、pEP 或 pN 按照步骤一中的 1.2 的方法进行细胞转染，并按照步骤一中的 1.3 和 1.4 的方法获得编辑效率结果，并用 graphpad prism 8.0 进行统计作图。

结果如图 2 和图 3 所示，在编辑窗口 C3-C8，hyBE4max 的 C 到 T 编辑效率为 19-71%，对应的 BE4max 为 13-47%；在编辑窗口 C9-C12，hyBE4max 的 C 到 T 编辑效率为 19-55%，对应的 BE4max 为 1.4-17%。相对于 BE4max，在编辑窗口 C3-C8 内，hyBE4max 平均 C 到 T 编辑效率为 BE4max 的 1.6-2.2 倍；在编辑窗口 C9-C12 内，hyBE4max 平均 C 到 T 编辑效率为 BE4max 的 3.3-17 倍。同时 hyBE4max 保持较低的 indels 产生（图 4）。

四、含不同胞嘧啶脱氨酶的融合蛋白的效果

（一）、融合蛋白 hyA3A-BE4max 工作特性

4.1.1、Rad51-DBD 按照表 1 中编码序列进行合成，之后进行无缝克隆组装至表达蛋白 A3A-BE4max（图 5）的质粒 pCMV-A3A-BE4max 中 hA3A 和 spCas9n 之间，构建表达融合蛋白 hyA3A-BE4max（图 5）的重组质粒 pA。

4.1.2、依次合成 8 种人的内源性靶点：EMX1 site1、Tim3-sg1、VEGFA site2、EMX1-sg2p、Nme1-sg1 的靶序列如表 2 所示，FANCF site1、EGFR-sg5、EGFR-sg21 的靶序列如表 6 所示；分别连接至 sgRNA 表达质粒 pU6-sgRNA-EF1α-GFP 的 Bbs Ⅰ位点处，得到表达相应靶点的 sgRNA 的重组质粒 pB1、pB2、……pB8。

4.1.3、将 4.1.1 与 4.1.2 中构建的质粒经 sanger 测序，确保完全正确。

表 6、所用靶点及序列

靶点名称	序列（5’-3’）	SEQ ID No.
FANCF site1	GGAATCCCTTCTGCAGCACCTGG	25
EGFR-sg5	GTGCTGGGCTCCGGTGCGTTCGG	26
EGFR-sg21	CAAAGCAGAAACTCACATCGAGG	27

4.1.4、细胞转染

将 5×10^5 个 HEK293T 细胞铺 24 孔板，待细胞长至 70%-80%时，按 pA（或质粒 pCMV-A3A-BE4max）:pB1（或 pB2、pB3、……pB8）=750ng:250ng 进行质粒组合的转染，每种质粒组合转染设 3 孔重复，每孔 2×10^5 个细胞。同时设不转染任何质粒的空白对照。

4.1.5、基因组提取及扩增子文库的准备

按照步骤 1.3 的方法进行，其中，FANCF site1、EGFR-sg5、EGFR-sg21 的靶点的鉴定引物如表 7 所示，其余表达的鉴定引物如表 3 所示。

表 7、所用靶点的鉴定引物

靶点名称	序列（5’-3’）
FANCF site1	F:ggagtgagtacggtgtgcCAGAGAGGCGTATCATTTCGCGGAT
	R:gagttggatgctggatggCCAGGTGCTGACGTAGGTAGTGCTT
EGFR-sg5	F:ggagtgagtacggtgtgcCTTGTGGAGCCTCTTACACCCAGTG
	R:gagttggatgctggatggCTCCCCACCAGACCATGAGAGGC
EGFR-sg21	F:ggagtgagtacggtgtgcGCAGCATGTGGCACCATCTCA
	R:gagttggatgctggatggTGGACCCCCACACAGCAA

4.1.6、深度测序结果分析与统计

按照步骤 1.4 的方法进行。

结果表明：相对于蛋白 A3A-BE4max，融合蛋白 hyA3A-BE4max 对各靶点不同位置（C3-C15）单个碱基 C 到 T 的编辑效率均明显提高（图 6）。相对于 A3A-BE4max，hyA3A-BE4max 的高活性窗口从原来 C3-C11，拓展到 C3-C15；其中，在远离 PAM 区的 C3-C11，hyA3A-BE4max 对单个碱基 C 到 T 的编辑效率是 A3A-BE4max 的 1.1-2.3 倍，在靠近 PAM 区的 C12-C15，hyA3A-BE4max 对单个碱基 C 到 T 的编辑效率是 A3A-BE4max 的 3.1-4.1 倍，即在靠近 PAM 区的 C12-C15，hyA3A-BE4max 对单个碱基 C 到 T 的编辑效率提高更为明显（图 7）。且 hyA3A-BE4max 同时维持较低的 indels（图 8）。

（二）、融合蛋白 hyeA3A-BE4max 工作特性

4.2.1、工作系统质粒构建

Rad51-DBD 按照表 1 中编码序列进行合成，之后进行无缝克隆组装至表达蛋白 eA3A-BE4max（图 9）的质粒 pCMV-eA3A-BE4max 中 eA3A 与 spCas9n 之间，构建表达融合蛋白 hyeA3A-BE4max（图 9）的重组质粒 pAe。

4.2.2、靶点质粒的构建

同时设计并合成 11 个来自于人的内源性靶点：EMX1-sg2p、EMX1 site1、Nme1-sg1 的靶序列如表 2 所示，EGFR-sg21 的靶序列如表 6 所示，其余靶点序列如表 8 所示，分别连接至 sgRNA 表达质粒 U6-sgRNA-EF1α-GFP 的 BbsI 位点处，用于表达相应靶点的 sgRNA，得到重组质粒 pC1、pC2、……pC11。

4.2.3、将 4.2.1 与 4.2.2 中构建的质粒经 sanger 测序，确保完全正确。

表 8、所用靶点及序列

靶点名称	序列（5`-3`）	SEQ ID No.
CTLA-sg1	CTCCCTCAAGCAGGCCCCGCTGG	28
EGFR-sg5	GTGCTGGGCTCCGGTGCGTTCGG	29
CDK10-sg1	TTCTCGGAGGCTCAGGTGCGTGG	30
EMX1-sg1	GCTCCCATCACATCAACCGGTGG	31
HPRT1-sg6	GCCCTCTGTGTGCTCAAGGGGGG	32
EGFR-sg26	CATGCCCTTCGGCTGCCTCCTGG	33
CCR5-sg1	TAATAATTGATGTCATAGATTGG	34

4.2.4、细胞转染-验证 hyeA3A-BE4max 工作系统

将 5×10^5 个 HEK293T 细胞铺 24 孔板，待细胞长至 70%-80%时，按 pA（或质粒 pCMV-eA3A-BE4max）:pC1（或 pC2、pC3、……pC11）=750ng:250ng 进行质粒组合的转染，每种质粒组合转染设 3 孔重复，每孔 2×10^5 个细胞。同时设不转染任何质粒的空白对照。

4.2.5、基因组提取及扩增子文库的准备

按照步骤 1.3 的方法进行，其中，EMX1-sg2p、EMX1 site1、Nme1-sg1 的鉴定引物如表 3 所示，EGFR-sg21 的鉴定引物如表 7 所示，其余靶点序列如表 9 所示。

表 9、所用靶点的鉴定引物

鉴定引物	序列（5`-3`）
CTLA-sg1	F:GGAGTGAGTACGGTGTGCACTTTAACCCCAAGTCTAGCAAGC
	R:GAGTTGGATGCTGGATGGATAATTAATCCAATGACTGTGAAGAGC
EGFR-sg5	F:GGAGTGAGTACGGTGTGCCTTGTGGAGCCTCTTACACCCAGTG
	R:GAGTTGGATGCTGGATGGCTCCCCACCAGACCATGAGAGGC
CDK10-sg1	F:GGAGTGAGTACGGTGTGCGGTGATGGGTTACTGTGAGCAGGAC
	R:GAGTTGGATGCTGGATGGGTCCCCAAGTGCCCTGTTTCGTTAT
EMX1-sg1	F:GGAGTGAGTACGGTGTGCGTGGTTCCAGAACCGGAGGACAAAG
	R:GAGTTGGATGCTGGATGGGTTTGTGGTTGCCCACCCTAGTCAT
HPRT1-sg6	F:GGAGTGAGTACGGTGTGCTTCCTGATTTTATTTCTGTAGGACTG
	R:GAGTTGGATGCTGGATGGTCTACAGTCATAGGAATGGATCTATCA
EGFR-sg26	F:GGAGTGAGTACGGTGTGCACCTCCACCGTGCAGCTC
	R:GAGTTGGATGCTGGATGGAGCGCAGACCGCATGTGAG
CCR5-sg1	F:GGAGTGAGTACGGTGTGCTGCACAGGGTGGAACAAGATGGATT
	R:GAGTTGGATGCTGGATGGATGACCAGCATGTTGCCCACAAAAC

4.2.6 深度测序结果分析与统计

按照步骤 1.4 的方法进行。

结果表明：相对于蛋白 eA3A-BE4max，融合蛋白 hyeA3A-BE4max 对各靶点不同位置（C3-C15）单个碱基 C 到 T 的编辑效率大部分明显提高，且高活性窗口从原来 C3-C11 拓展到 C3-C15 位，能特异靶向 TC motif 中的单个碱基 C 实现 C 到 T 转换（图 10）；其中，在远离 PAM 区的 C3-C11，hyeA3A-BE4max 对单个碱基 C 到 T 的编辑效率是 eA3A-BE4max 的 1.6-2.8 倍，在靠近 PAM 区的 C12-C15，hyeA3A-BE4max 对单个碱基 C 到 T 的编辑效率是 eA3A-BE4max 的 4.5-31.9 倍，即在靠近 PAM 区的 C12-C15，hyeA3A-BE4max 对单个碱基 C 到 T 的编辑效率提高更为明显（图 11）。同时 hyeA3A-BE4max 保持较低的 indels(图 12)。

五、应用融合蛋白 hyeA3A-BE4max 进行基因治疗

5.1、hyeA3A-BE4max 靶向 HBG-117G 位点在 HEK293T 细胞上编辑效率的测试

按照 4.2.4 的细胞转染方法，将按 pA（或质粒 pCMV-A3A-BE4max、或 pAe、或 pCMV-eA3A-BE4max）:pC12=750ng:250ng 进行质粒组合的转染 HEK293T 细胞；按照 4.2.6 的方法进行深度测序及统计分析。

上述重组质粒 pC12 的构建方法如下：将 HBG-117G 的 sgRNA 靶点序列（GGCTATTGGTCAAGGCAAGGCTGG，SEQ ID No.35）连接至 sgRNA 表达质粒 U6-sgRNA-EF1α-GFP 的 BbsI 位点处，用于表达相应靶点的 sgRNA，得到重组质粒 pC12。

上述深度测序靶点 HBG-117G 所用的鉴定引物如下：

HBG-117G F:AGTGAGTACGGTGTGCTGGAATGACTGAATCGGAACAAGGC；

HBG-117G R:GTTGGATGCTGGATGGCTGGCCTCACTGGATACTCTAAGACT。

结果：如图 14 所示，A3A-BE4max 除能靶向 HBG1/2 启动子区的-117G>A（C11）突变外，还会产生-109G>A(C3)，-122G>A(C16)“旁观者”突变；对于 eA3A-BE4max 和 hyeA3A-BE4max 而言，eA3A-BE4max 精确地编辑了-117 上的 G 到 A 转换(即互补链 C 到 T 转换)而不引起“旁观者”突变，但效率非常低，与 eA3A-BE4max 相比，hyeA3A-BE4max 提高了 6.6 倍的 G 到 A 转换效率，并且在-109 和-122 处均未产生可检测到的“旁观者”突变。结果表明，hyeA3A-BE4max 展现出精准高效靶向 HBG1/2 启动子区-117G 的特性（作用机理总结如图 13 所示）。

5.2、慢病毒载体的构建与病毒包装

5.2.1、慢病毒载体的构建

以 pLenti-BE3-P2A-Puro(Addgene，#110838)为骨架载体，将 hyA3A-BE4max 的编码序列通过无缝克隆替换骨架载体上的 BE3 得到慢病毒载体 Lenti-hyA3A-BE4max-P2A-GFP。

以 pLenti-BE3-P2A-Puro(Addgene，#110838)为骨架载体，将 hyeA3A-BE4max 的编码序列通过无缝克隆替换骨架载体上的 BE3 得到慢病毒载体 Lenti- hyeA3A-BE4max-P2A-GFP。

将 5.2 中 HBG-117G 靶点序列分别连接至上述两个慢病毒载体的 hyA3A-BE4max 或 hyeA3A-BE4max 的上游（图 15 上图），得到重组质粒 Lenti-117G-hyA3A-BE4max-P2A-GFP 和重组质粒 Lenti-117G hyeA3A-BE4max-P2A-GFP（图 15 下图）。

5.2.2、慢病毒包装

5.2.2.1、转染

第 1 天，将生长状态良好的 HEK293T 细胞消化后铺 10cm 皿，每个病毒约 30 皿。 第 2 天 ，待汇合度 80%-90%时，按如下量进行质粒转染:Lenti-117G- hyA3A-BE4max-P2A-GFP（或 Lenti-117G-hyeA3A-BE4max-P2A-GFP）: PSPAX2:PMD2.G= 10ug:10ug:10ug。

5.2.2.2、病毒的收集和纯化

病毒上清收集于转染后 48 h(转染即可为 0 小时计起)、72h 的 HEK293T 细胞上清。于 4℃，4000 g 离心 10 min，除去细胞碎片，后以 0.45 μm 滤器过滤上清液于 40 mL 超滤离心管中，把慢病毒粗提液加入到过滤杯中，4℃、25000g 离心 2.5 小时。离心结束后，取出离心装置，将过滤杯和下面的滤过液收集杯分开。样品收集杯中的液体即为病毒浓缩液(含有慢病毒 Lenti-117G-hyA3A-BE4max-P2A-GFP 或慢病毒 Lenti-117G-hyeA3A-BE4max- P2A-GFP)。将病毒浓缩液移出，分装后保存在病毒管中，-80℃长期保存。

5.3、基因治疗

5.3.1、HUDEP-2($\Delta^{G}\gamma$)细胞感染病毒

在 96 孔板铺 5×10^4 的 HUDEP-2($\Delta^G\gamma$)细胞 3 个孔，总体积为 100ul 培养基，分别等滴度感染慢病毒 Lenti-117G-hyA3A-BE4max-P2A-GFP 和 Lenti -117G-hyeA3A-BE4max-P2A-GFP，感染体系如下：

试剂	Lenti -117G-hyA3A-BE4max-P2A -GFP	Lenti-117G-hyeA3A-BE4max-P2A-GFP	空白对照
病毒	10MOI	10MOI	0
聚凝胺 Polybrene(100×)	1ul	1ul	1ul
培养基	补齐至 100ul		

5.3.2、编辑效率检测

将感染慢病毒 Lenti-117G-hyA3A-BE4max-P2A-GFP 或 Lenti -117G-hyeA3A-BE4max-P2A-GFP 的 HUDEP-2(ΔGγ)细胞经流式分选 GFP 阳性细胞，培养到细胞数量大于 5×10^4 后，收取细胞，提取基因组 DNA，按照步骤 5.1 的方法进行深度测序和分析。

结果：与 hyA3A-BE4max 相比，hyeA3A-BE4max 有效靶向 HUDEP-2（$\Delta^G\gamma$）细胞中精确的–117G> A 突变，并表现出更高的活性（图 16）。

5.3.3 分化及γ珠蛋白表达检测

将感染慢病毒 Lenti-117G-hyA3A-BE4max-P2A-GFP 或 Lenti -117G-hyeA3A-BE4max-P2A-GFP 的 HUDEP-2(ΔGγ)细胞经流式分选 GFP 阳性细胞，培养到细胞数量大于 5×10^4，约 5-7 天后收集 HUDEP-2($\Delta^G\gamma$)细胞进行分化表达。分化过程如下：

计数后，1×10^5 个 HUDEP-2($\Delta^G\gamma$)细胞分化在类红细胞分化培养基（IMDM）中进行，同时补充 2%人血 AB 型血浆（血清）(Gemini，100-512)，1%L-谷氨酰胺，2 IU / mL 肝素，促红细胞生成素（EPO，3 IU / ml，PeproTech），330μg/ mL Holo-人类转铁蛋白（Sigma-Aldrich），人干细胞因子（SCF，50 ng / ml，PeproTech），2%的 Pen / Strep（Gibco），10μg/ ml 重组人胰岛素，分化 8 天。

γ珠蛋白表达检测：分化 8 天后，收细胞用酚氯仿抽提法抽提总的 mRNA。使用 HiScript II Q RT SuperMix（Vazyme）对分离的 mRNA 进行逆转录；在 QuantiStudio 3 实时 PCR 系统（ABI）上进行 qPCR，并通过 SYBRGreen qPCR 对 HBG 和 HBB mRNA 进行定量。引物 (5'-3')如下：

HBG-QPCR-F：GGTTATCAATAAGCTCCTAGTCC；

HBG-QPCR-R：ACAACCAGGAGCCTTCCCA；

HBB-QPCR-F：TGAGGAGAAGTCTGCCGTTAC；

HBB-QPCR-F：ACCACCAGCAGCCTGCCCA。

结果：与 HUDEP-2（$\Delta^G\gamma$）的 WT 细胞相比，hyA3A-BE4max 和 hyeA3A-BE4max 可以显著提高 HUDEP-2（$\Delta^G\gamma$）细胞中 γ-珠蛋白 mRNA 的水平；hyeA3A-BE4max 对 HUDEP-2（$\Delta^G\gamma$）细胞中 γ-珠蛋白 mRNA 水平提高程度是 hyA3A-BE4max 的 3 倍（图 17）。

六、应用融合蛋白 hyA3A-BE4max 构建 DMD 疾病动物模型

以下所用小鼠为 C57/BL6 小鼠。

6.1、工作系统 mRNA 与靶点 sgRNA 转录模板的构建

在 NCBI 上下载小鼠相关基因序列，如图 18 所示，在目标位点（抗肌萎缩蛋白基因即 DMD 基因第 12 个外显子中矩形框处的位点）设计 sgRNA（DMD-sg3 靶点序列：ACATCTCATCAAGGACTTGTTGG，SEQ ID No.36），并订购 Oligo 引物，将通过退火形成的

sgRNA 克隆到 T7 载体骨架中,使用引物对 IVT-PCR-F 和 IVT-PCR-R(表 10)体外转录法(IVT)扩增含有 T7 启动子的 DMD-sg3 模板，使用引物对 IVT-T7-hyA3A-BE4max-F 和 IVT-T7-hyA3A-BE4max-R（表 10）将 T7 启动子通过 PCR 引入 hyA3A-BE4max 或者 A3A-BE4max 的 mRNA 模板。

表 10、IVT 所用 PCR 引物

引物名称	引物序列 (5’-3’)
IVT-PCF-F	GCGGCTTTGTTGAATAAATCGCATTCG
IVT-PCF-R	AAAAGCACCGACTCGGTGCC
IVT-T7-hyA3A-BE4max-F	TTAATACGACTCACTATAGGGAGAATGAAGAGGACCGCCGATGGCTC
IVT-T7-hyA3A-BE4max-R	CTAGTCACCTCCCAGCTGAGACAGGTC

6.2、sgRNA (DMD-sg3)体外转录

利用普通 DNA 产物纯化试剂盒对 6.1 中 PCR 产物进行纯化，以纯化后的 PCR 产物为线性化 DNA 模板，利用 T7 体外转录试剂盒（MEGAshortscriptTM Kit）进行体外转录，利用氯化锂沉淀法对转录的 sgRNA 进行纯化。

6.3、工作系统 mRNA(A3A-BE4max 和 hyA3A-BE4max)的转录

利用体外 RNA 转录试剂盒（mMESSAGE mMACHINE®T7 Ultra Kit）对 A3A-BE4max 和 hyA3A-BE4max 的 T7 模板分别进行体外转录得到工作系统 mRNA，并对其进行纯化。

6.4、显微注射混合物的制备

用无核酸酶的水配置注射混合物，获得总体积为 20μL，终浓度为 100ng/μL 的工作系统 mRNA（含有 A3A-BE4max 的 mRNA 或含有 hyA3A-BE4max 的 mRNA）与终浓度为 200ng/μL 的 sgRNA (DMD-sg3)的混合液。

6.5、一细胞期胚胎的收集

（1）第一天：于下午 1-2 点间向 6-8 周龄的供体母鼠腹腔注射 100μL（5IU）的 PMSG 工作液。

（2）第三天：于下午 2-4 点间向注射过 PMSG 的母鼠腹腔注射 100μL（5IU）的 hCG 工作液，注射后，将经激素处理后的母鼠与 10-14 周龄的公鼠一对一合笼。同时，将处于发情期未经激素处理的雌鼠与输卵管结扎的雄鼠在下午 4 点左右交配用于假孕母鼠的准备。

（3）第四天：早上 9 点前，检查与结扎公鼠合笼的受体母鼠是否有孕栓，将有孕栓的母鼠集中于新的笼子内用于下午的胚胎移植实验。

（4）通过二氧化碳窒息法将超排卵的供体母鼠处死取出输卵管，放置于平皿中，平皿中加入预热的 M2 培养基。

（5）将输卵管放置于另一个新的平皿中，平皿内加入预热的 M2 培养基与透明质酸，M2 培养基与透明质酸的体积比为 9:1。在体式显微镜下，用摄子拉扯输卵管壶腹部，使胚胎从输卵管释放至平皿中。把胚胎一直孵育在有透明质酸的 M2 培养基中直到卵丘细胞掉落。在去除卵丘细胞后，将胚胎转移至一个新的平皿中，平皿内加入无透明质酸的 M2 培养基，反复用 M2 培养基冲洗胚胎使透明质酸和卵丘细胞均被冲洗干净。

（6）将冲洗干净的胚胎转移至一个新的平皿中，平皿内先分散加入几滴 KSOM 培养基，然后再缓慢向平皿加入矿物油，使 KSOM 培养基被矿物油隔开与覆盖。一般来说，在一个 35 毫米的平皿中，可加 6 个点的 KSOM 培养基，每个点为 50μL。将 50 个胚胎作为一组首先放于中间的 KSOM 培养基点进行漂洗，然后转移至一个新的培养基点。在显微注射前，把取出的胚胎孵育在细胞培养箱的 M2 培养基中。

6.6、显微注射与胚胎移植

（1）准备固定针，注射针以及硅化的玻璃载玻片，往载玻片中间滴一滴被矿物油覆盖的 M2 培养基。

（2）使注射针依靠毛细管作用自动吸入并充满步骤 2.4 制备的显微注射混合液，将注射针装载至显微注射仪的固定柄上。

（3）转移 50 个胚胎至载玻片的 M2 培养基内，移动固定针靠近胚胎，使胚胎依靠负压作用被固定在固定针上。胚胎被固定后在高倍镜下找到细胞质，推动注射针的尖端穿过透明带，细胞膜，将混合液注射到胚胎的细胞质中。

（4）将注射过的胚胎细胞转移至新的 M2 培养基点中。重复步骤（3）和（4）直至所有的胚胎被注射完。注射完一组实验组后，将胚胎转移至新的 KSOM 培养基中，将胚胎放置于细胞培养箱中培养 1-2 小时或过夜。待所有胚胎均被注射后，排除因机械力损伤致死的胚胎，将健康的胚胎转移至新的 KSOM 培养基中。

（5）向假孕母鼠腹腔注射 600 μL 阿佛丁，将假孕母鼠麻醉。使用剃毛器将母鼠背部的毛剔除。使用 70%乙醇擦拭剃剃毛后的皮肤。

（6）在卵巢位置处剪开一个小口，使用钝头镊子牵拉着卵巢的脂肪垫从而将卵巢拉出，并同时用止血钳将卵巢固定于外侧，使用钝头镊子找到位于卵巢囊下侧的输卵管漏斗状口。

（7）使转移针依次吸入 M2 培养基，两个小气泡，约 15 个胚胎，气泡是为了便于观察胚胎在转移针中的位置。

（8）轻轻地剖开卵巢的囊，用辕子定位输卵管的漏斗状口，将转移针伸向卵巢的开口处，然后将转移针内的胚胎打出，轻轻地将转移针撤回。

（9）释放固定卵巢脂肪垫的止血钳，将卵巢放回原来的腔洞，用缝合线分别缝合肌肉开口处与皮肤开口处。

（10）将手术后的小鼠放置于恒温 37 度的保温台上，待小鼠恢复知觉后，将其转移至饲养笼具中饲养，等待胚胎发育直至分娩。一般来说，移植成功的母鼠在 3 周后生出小鼠。

6.7、小鼠基因组鉴定

取步骤 6.6 小鼠出生 10-15 天左右，剪其脚趾进行基因组的鉴定，具体步骤如下：

6.7.1 基因组提取

①剪其脚趾装入 1.5mL 离心管中，每管加入 500μL 按蛋白酶 K：组织裂解液=1：500 的比例配制的脚趾消化液，55°C 水浴过夜；

②取出过夜消化的脚趾，室温放置 10-15 分钟，充分颠倒混匀，13000rpm 离心 15 分钟。

③每管吸出 400 μL 上清，加入等体积氯仿，充分混匀待 DNA 析出后，12000 rpm 离心 10 分钟。

④每管加入提前在-20℃冰箱预冷的 75%酒精 200μL，轻柔地混匀，12000 rpm 4°C 离心 5 min，弃上清，在洁净工作台中晾干。

⑤根据 DNA 量加入 50-100 μL 去离子超纯水，55℃溶解 2 小时即可作为 PCR 模板。

6.7.2 基因组的鉴定

使用靶点 DMA-sg3 的引物对 F/R（表 11）获得含有靶点的 DNA 片段，先通过一代测序确认有双峰出现，后续再进行高通量深度测序得出编辑效率。根据高通量结果显示，hyA3A-BE4max 处理组产生突变的 10 只 F0 中，有 6 只纯合的从 CAA 到 TAA 的终止无义突变(图 19 下图中编号为#BD03、#BD05、#BD07、#BD12、#BD15、#BD16 的小鼠)，而 A3A-BE4max 处理组 F0 中未发现从 CAA 到 TAA 的终止无义突变(图 19 上图)。与 A3A-BE4max 处理组相比，hyA3A-BE4max 处理组小鼠经过纯合的突变中含有终止密码子 TAA 的 Reads（高通量测序片段）数目占总 Reads 的比率高(图 20)。

表 11

靶点名称	序列（5'-3'）
DMD-sg3	F:ggagtgagtacggtgtgcTCAAACTCCCAATGATTTCCTCAAT
	R:gagttggatgctggatggTGCACTTCAGCTTCTTCATCTTCTG

6.8、DMD 基因的表型鉴定

取来自 5 周大的野生型小鼠（空白对照）以及 2.7.2 鉴定的 DMD 基因突变小鼠，按如下方法进行免疫组织化学检测：

取小鼠的胫骨前肌，用酒精、PBS 漂洗干净。将其放入铺有 OTC 胶的小正方体盒中，置于异戊烷烧杯，在液氮中冷冻，约 30s 取出，放入-20℃保存，之后进行冰冻切片。将制备好的切片用 PBST 洗 3 次/5min，对准组织位置画油圈，加入封闭液进行封闭，封闭 1h 后，分别用层粘连蛋白（Laminin）一抗或抗肌萎缩蛋白（Dystrophin）一抗（用 1：500 稀释后的一抗兔多克隆抗体(Abcam，ab11575)或一抗兔多克隆抗体 (Abcam，ab15277)）进行过夜孵育。用 PBST 洗 3 次/5min，然后以 1：1000 的抗兔稀释液二抗孵育 2h，PBST 洗 3 次/5min 后，加入以 1:100 稀释的 DAPI 孵育 10min，PBST 洗 3 次/5min，滴加防猝灭剂，加盖玻片，指甲油封片，最后用荧光显微镜观察切片的荧光。

结果表明：与 WT(+/+)和 A3A-BE4max 处理组小鼠(如#AD26)相比，只有 hyA3A-BE4max 处理组小鼠中引起 DMA-sg3 靶点序列第 10 位 C 到 T 纯合无义突变的 6 只 F0 代小鼠（如图 21 中的#BD03）中的 DMD 未得到蛋白表达(图 21)，这也证明 DMD 动物疾病模型构建成功。

根据荧光观察结果，同时获得 hyA3A-BE4max 和 A3A-BE4max 编辑后产生的 F0 代小鼠总结，如表 12 所示。

表 12、hyA3A-BE4max 和 A3A-BE4max 编辑后产生的 F0 突变结果对比

处理	突变频率	
	突变小鼠数量比例	纯合无义突变比例
A3A-BE4max	10/11(91%)	0/11（0）
hyA3A-BE4max	10/10(100%)	6/10（60%）

6.9、DMD 突变小鼠的生殖系遗传分析(F0→F1)

取雌性纯合的具有 DMD 表型的#BD12（F0）与雄性野生型小鼠进行交配，获得 8 只纯合的 F1，对出生的 F1 进行基因型鉴定，sanger 测序结果发现，每只 F1 产生无义突变的 Reads 频率均超过 96%，即这种无义突变可以稳定遗传到 F1 代(图 22)。

6.10、脱靶检测

利用 CRISPR RGEN Tools 网站（http://www.rgenome.net）的 Cas-OFFinder 功能设计脱靶引物(表 13)：首先选择所测试工具的 PAM 的类型以及被测试的物种类型（如小鼠为 Mus

musculus (mm10) - Mouse)，然后将所设计的sgRNA序列除去PAM部分填写在Query Sequences的方框内，选择错配碱基在3个以内，形成的DNA Bulge Size为在1以内，然后提交后获得相应的脱靶引物。

使用上述脱靶引物分别对经hyA3A-BE4max编辑的F0小鼠基因组DNA（以WT为空白对照）进行PCR后对产物进行高通量深度测序，获得脱靶效率结果(图23)。

表13、DMD-sg3脱靶位点及PCR引物

脱靶位点名称	靶点序列(5'-3')	引物序列(5'-3')
DMD-sg3-OT1	acatttaatcaaggccttgttgg	F: ggagtgagtacggtgtgcaaaagcctacaacacaatgagaaac
		R: gagttggatgctggatggttgataaggctctaccaatgtgaac
DMD-sg3-OT2	acctgtcatcaagcacttgttgg	F: ggagtgagtacggtgtgcaaggaatgaccttgcagagactgcc
		R: gagttggatgctggatggtgctcagtcccatggatgactgtga
DMD-sg3-OT3	tcctgtcatcaaggacttgttgg	F: ggagtgagtacggtgtgcggtgcagagactgaagaaaaggcca
		R: gagttggatgctggatggtctatctctgagtatgtcaggcactgg
DMD-sg3-OT4	tcctctcatcaagcacttgttgg	F: ggagtgagtacggtgtgcggaatgaccatccagagacttcccc
		R: gagttggatgctggatggttggggaccctatgctcagtccaat
DMD-sg3-OT5	acatctcacccaggacctgtggg	F: ggagtgagtacggtgtgcggcttccttttctgtcactgtggga
		R: gagttggatgctggatgggggggactgccactcctaactttcat
DMD-sg3-OT6	acttctcatcaaggacctctggg	F: ggagtgagtacggtgtgctccctatccgtcttagtctggagcc
		R: gagttggatgctggatggacaagaaaccatgcatcagcttgtga
DMD-sg3-OT7	acttctcatcaaggacctctggg	F: ggagtgagtacggtgtgcgcagaaagaaagtacagacacttaca
		R: gagttggatgctggatggaatgggtaaatagttaaaagccatga
DMD-sg3-OT8	acaactcaaaaaggacttgttgg	F: ggagtgagtacggtgtgcggctaccataggccattttctgcat
		R: gagttggatgctggatggtgagaatgaatgttaggtcaggtgga
DMD-sg3-OT9	acatctcatctatgatttgtagg	F: ggagtgagtacggtgtgcggcttgcccaagtttattaggggagt
		R: gagttggatgctggatggccagccccttggcctcctttaaaat
DMD-sg3-OT10	acaactcctcatggacttgtggg	F: ggagtgagtacggtgtgcacatcccagagaatcctgagtttaaga
		R: gagttggatgctggatggtgctaggagagatggttgtatcaagga
DMD-sg3-OT11	gcatctcatctaggacttggtgg	F: ggagtgagtacggtgtgctgtgataggtgatttctcatgcacca
		R: gagttggatgctggatggatcacattcacttcccttcaccccg
DMD-sg3-OT12	aaatctcatcaattacttgtagg	F: ggagtgagtacggtgtgccagccttatccatttacttgtgaattt
		R: gagttggatgctggatggtcaagaacaataaaacaaatgacagc
DMD-sg3-OT13	tcatgtcagcaaggacttgttgg	F: ggagtgagtacggtgtgcatgaccatccagagactgacccact
		R: gagttggatgctggatggccacttctgtattagtctggcactggc
DMD-sg3-OT14	acatctgaccaagggcttgtagg	F: ggagtgagtacggtgtgccagttcctgtcggcagttcctcaat
		R: gagttggatgctggatggtgtgccctggaaaagatcacacaga
DMD-sg3-OT15	acttctcagcaagcacttgtagg	F: ggagtgagtacggtgtgccaacggtagcctaaggactgtccac
		R: gagttggatgctggatggcaggcttcctgactttcaaaaccgc

图23结果表明，与空白对照相比，融合蛋白hyA3A-BE4max及DMD-sg3对表13中15个与DMD-sg3靶点相似的位点（即脱靶位点）基本不发生C到T的编辑，即本申请的融合蛋

白 hyA3A-BE4max 及 DMD-sg3 基本不产生脱靶效应。

1、一种提高基因编辑效率的融合蛋白，其特征在于，所述融合蛋白包括单链 DNA 结合蛋白功能域、核苷脱氨酶和核酸酶。

2、根据权利要求 1 所述的融合蛋白，其特征在于，所述融合蛋白的连接顺序为：所述核苷脱氨酶位于所述核酸酶的 N 端或 C 端，所述单链 DNA 结合蛋白功能域位于所述核苷脱氨酶和所述核酸酶的 N 端、C 端和/或所述核苷脱氨酶和所述核酸酶之间。

3、根据权利要求 2 所述的融合蛋白，其特征在于，所述核苷脱氨酶位于所述核酸酶的 N 端。

4、根据权利要求 3 所述的融合蛋白，其特征在于，所述单链 DNA 结合蛋白功能域位于所述核苷脱氨酶和所述核酸酶之间。

5、根据权利要求 1 所述的融合蛋白，其特征在于，所述单链 DNA 结合蛋白包括序列特异性单链 DNA 结合蛋白、和/或非序列特异性单链 DNA 结合蛋白。

6、根据权利要求 5 所述的融合蛋白，其特征在于，所述非序列特异性单链 DNA 结合蛋白选自 RPA70、RPA32、BRCA2、hnRNPK、PUF60 和 Rad51 中的任一种或任几种；

所述序列特异性单链 DNA 结合蛋白选自 TEBP、Teb1 和 POT1 中的任一种或任几种。

7、根据权利要求 5 所述的融合蛋白，其特征在于，所述单链 DNA 结合蛋白功能域包括如下四种结构域中的至少一种或如下四种结构域中具有与单链 DNA 结合功能的部分多肽片段及其任意组合：所述单链 DNA 结合蛋白的 OB 折叠、KH 结构域、RRMS、涡状结构域。

8、根据权利要求 5 所述的融合蛋白，其特征在于，所述单链 DNA 结合蛋白功能域包括 Rad51 的 DNA 结合结构和/或 RPA70 的 DNA 结合结构域。

9、根据权利要求 8 所述的融合蛋白，其特征在于，所述 Rad51 的 DNA 结合结构域的氨基酸序列包括 SEQ ID No.1 所示序列；

和/或，所述 Rad51 的 DNA 结合结构域的编码序列包括 SEQ ID No.2 所示序列。

10、根据权利要求 8 所述的融合蛋白，其特征在于，所述 RPA70 的 DNA 结合结构域的氨基酸序列包括 SEQ ID No.11 所示序列；

和/或，所述 RPA70 的 DNA 结合结构域的编码序列包括 SEQ ID No.12 所示序列。

12、根据权利要求 1 所述的融合蛋白，其特征在于，所述脱氨酶包括胞嘧啶脱氨酶和/或腺苷脱氨酶。

13、根据权利要求 12 所述的融合蛋白，其特征在于，所述胞嘧啶脱氨酶包括来源于大鼠的胞嘧啶脱氨酶。

14、根据权利要求 13 所述的融合蛋白，其特征在于，所述来源于大鼠的胞嘧啶脱氨酶氨基酸序列包括 SEQ ID No.3 所示序列；

和/或，所述来源于大鼠的胞嘧啶脱氨酶的编码序列包括 SEQ ID No.4 所示序列。

15、根据权利要求 12 所述的融合蛋白，其特征在于，所述胞嘧啶脱氨酶包

括来源于人的胞嘧啶脱氨酶 APOBEC3A。

16、根据权利要求 15 所述的融合蛋白，其特征在于，所述来源于人的胞嘧啶脱氨酶 APOBEC3A 的氨基酸序列包括 SEQ ID No.13 所示序列；

和/或，所述来源于人的胞嘧啶脱氨酶 APOBEC3A 的编码序列包括 SEQ ID No.14 所示序列。

17、根据权利要求 12 所述的融合蛋白，其特征在于，所述胞嘧啶脱氨酶包括胞嘧啶脱氨酶 APOBEC3A 的突变体，所述突变体是将所述胞嘧啶脱氨酶 APOBEC3A 第 57 位的天冬酰胺突变为甘氨酸。

18、根据权利要求 17 所述的融合蛋白，其特征在于，所述胞嘧啶脱氨酶 APOBEC3A 来源于人。

19、根据权利要求 18 所述的融合蛋白，其特征在于，所述胞嘧啶脱氨酶 APOBEC3A 的氨基酸序列包括 SEQ ID No.13 所示序列；

和/或，所述胞嘧啶脱氨酶 APOBEC3A 的编码序列包括 SEQ ID No.14 所示序列。

20、根据权利要求 17 所述的融合蛋白，其特征在于，所述胞嘧啶脱氨酶 APOBEC3A 突变体的氨基酸序列包括 SEQ ID No.15 所示序列；

和/或，所述胞嘧啶脱氨酶 APOBEC3A 的编码序列包括 SEQ ID No.16 所示序列。

21、根据权利要求 1 所述的融合蛋白，其特征在于，所述核酸酶选自 Cas9、Cas3、Cas8a、Cas8b、Cas10d、Cse1、Csy1、Csn2、 Cas4、Cas10、Csm2、Cmr5、Fok1、Cpf1 中的一种或任意几种。

22、根据权利要求 21 所述的融合蛋白，其特征在于，所述核酸酶为 Cas9。

23、根据权利要求 22 所述的融合蛋白，其特征在于，所述 Cas9 选自来源于肺炎链球菌、金黄色葡萄球菌、酿脓链球菌或嗜热链球菌的 Cas9。

24、根据权利要求 22 所述的融合蛋白，其特征在于，所述 Cas9 选自 Cas9 突变体 VQR-spCas9、VRER-spCas9、或 spCas9n。

25、根据权利要求 24 所述的融合蛋白，其特征在于，所述 spCas9n 的氨基酸序列包括 SEQ ID No.5 所示序列；

和/或，所述 spCas9n 的编码序列包括 SEQ ID No.6 所示序列。

26、根据权利要求 1 所述的融合蛋白，其特征在于，所述融合蛋白还包括 NLS。

27、根据权利要求 26 所述的融合蛋白，其特征在于，所述 NLS 位于所述融合蛋白的至少一端。

28、根据权利要求 26 所述的融合蛋白，其特征在于，所述 NLS 的氨基酸序列包括 SEQ ID No.7 所示序列；

和/或，所述 NLS 的编码序列包括 SEQ ID No.8 所示序列。

29、根据权利要求 1 所述的融合蛋白，其特征在于，所述融合蛋白还包括两拷贝以上的 UGI。

30、根据权利要求 29 所述的融合蛋白，其特征在于，所述 UGI 位于所述融合蛋白的至少一端。

31、根据权利要求 29 所述的融合蛋白，其特征在于，所述 UGI 的氨基酸序列包括 SEQ ID No.9 所示序列；

和/或，所述 UGI 的编码序列包括 SEQ ID No.10 所示序列。

32、如下 A）-C）生物材料中的任一种：

A）一种基因，编码权利要求 1-31 中任一所述融合蛋白；

B）一种重组载体，含有 A）所述基因；

C）一种重组细胞或重组菌，含有权利要求 1-31 中任一所述的融合蛋白，或含有 A）所述基因。

33、根据权利要求 32 所述的生物材料，其特征在于，所述细胞为 T 细胞、造血干细胞、骨髓细胞、血红细胞、或红细胞前体细胞。

34、一种对细胞内目的基因进行基因编辑的 sgRNA，其特征在于，所述 sgRNA 的靶序列包括 SEQ ID No.17-36 中至少一种。

35、根据权利要求 34 所述的 sgRNA，其特征在于，所述细胞为 T 细胞、造血干细胞、骨髓细胞、血红细胞、或红细胞前体细胞。

36、根据权利要求 34 所述的 sgRNA，其特征在于，所述目的基因为 HBG1 和 HBG2 启动子区。

37、一种单碱基基因编辑系统，其特征在于，所述系统包括权利要求 1-31 中任一所述的融合蛋白、和/或权利要求 32 或 33 所述的生物材料、和 sgRNA，所述 sgRNA 引导所述融合蛋白对目的细胞中的目的基因进行单碱基基因编辑。

38、根据权利要求 37 所述的单碱基基因编辑系统，其特征在于，所述 sgRNA 的靶序列包括 SEQ ID No.17-36 中至少一种；

和/或，所述细胞为 T 细胞、造血干细胞、骨髓细胞、血红细胞、或红细胞前体细胞，

和/或，所述目的基因为 HBG1 和 HBG2 启动子区。

39、权利要求 1-31 中任一所述融合蛋白、权利要求 32 或 33 所述的生物材料、权利要求 34-36 中任一所述的 sgRNA、或权利要求 37 或 38 所述的单碱基基因编辑系统在制备基因编辑产品、疾病治疗和/或预防产品、动物模型或植物新品种中的应用。

40、根据权利要求 39 所述的应用，其特征在于，所述疾病为 β 血红蛋白病，所述 β 血红蛋白病包括 β 地中海贫血和/或镰刀型细胞贫血症。

41、一种提高单碱基基因编辑效率的方法，其特征在于，所述方法包括利用权利要求 1-31 中任一所述的融合蛋白和 sgRNA 引入细胞、对目的基因进行基因编辑的步骤，所述 sgRNA 引导所述融合蛋白对所述目的基因进行单碱基基因编辑。

42、根据权利要求 41 所述的方法，其特征在于，所述 sgRNA 的靶序列包括 SEQ ID No.17-36 中的至少一种；

和/或，所述细胞为 T 细胞、造血干细胞、骨髓细胞、血红细胞、或红细胞前体细胞；

和/或，所述目的基因为 HBG1 和 HBG2 启动子区。

43、一种疾病动物模型的构建方法，其特征在于：所述方法包括利用权利要

求 1-31 中任一所述的融合蛋白和 sgRNA 引入动物细胞、对目的基因进行基因编辑的步骤。

44、根据权利要求 43 所述的方法，其特征在于，所述 sgRNA 的靶序列包括 SEQ ID No.17-36 中的至少一种。

45、根据权利要求 44 所述的方法，其特征在于，所述 sgRNA 的靶序列包括 SEQ ID No.36 所示序列，所述目的基因包括 DMD 基因。

46、根据权利要求 43 所述的方法，其特征在于，所述动物为哺乳动物，

和/或，所述细胞为胚胎细胞，

和/或，所述引入的方式为载体转化、显微注射、转染、脂质转染、热休克、电穿孔、转导、基因枪、DEAE-葡聚糖介导的转移中的一种或任几种组合，

和/或，所述引入使用权利要求 1-31 中任一所述的融合蛋白的 mRNA 和所述 sgRNA 进行。

47、根据权利要求 46 所述的方法，其特征在于，所述哺乳动物为大鼠或小鼠；

和/或，当所述引入的方式为显微注射时，所述引入使用的权利要求 1-31 中任一所述的融合蛋白的 mRNA 的浓度为 1-1000ng/μL。

48、根据权利要求 47 所述的方法，其特征在于，所述引入使用的权利要求 1-31 中任一所述融合蛋白的 mRNA 与所述引入使用的所述 sgRNA 的浓度比为 1:（5-1）。

49、权利要求 43-48 中任一所述方法得到的动物模型在药物筛选、疾病治疗效果评价或疾病治病机理研究中的应用。

50、一种用于治疗和/或预防 β 血红蛋白病的产品，其特征在于，所述产品中包含：权利要求 32 中 A）中所述基因和 sgRNA 的递送载体，

所述 sgRNA 引导所述融合蛋白对目的细胞中目的基因进行单碱基基因编辑；所述目的基因为 HBG1 和 HBG2 启动子区。

51、根据权利要求 50 所述的产品，其特征在于，所述 sgRNA 的靶序列包括 SEQ ID No.35 所示序列；

和/或，所述 β 血红蛋白病包括 β 地中海贫血和/或镰刀型细胞贫血症；

和/或，所述细胞为 T 细胞、造血干细胞、骨髓细胞、血红细胞、或红细胞前体细胞。

52、根据权利要求 50 所述的产品，其特征在于，所述递送载体包括病毒载体、和/或非病毒载体；所述病毒载体包括腺相关病毒载体、腺病毒载体、慢病毒载体、逆转录病毒载体、和/或溶瘤病毒载体，所述非病毒载体包括阳离子高分子聚合物、和/或脂质体。

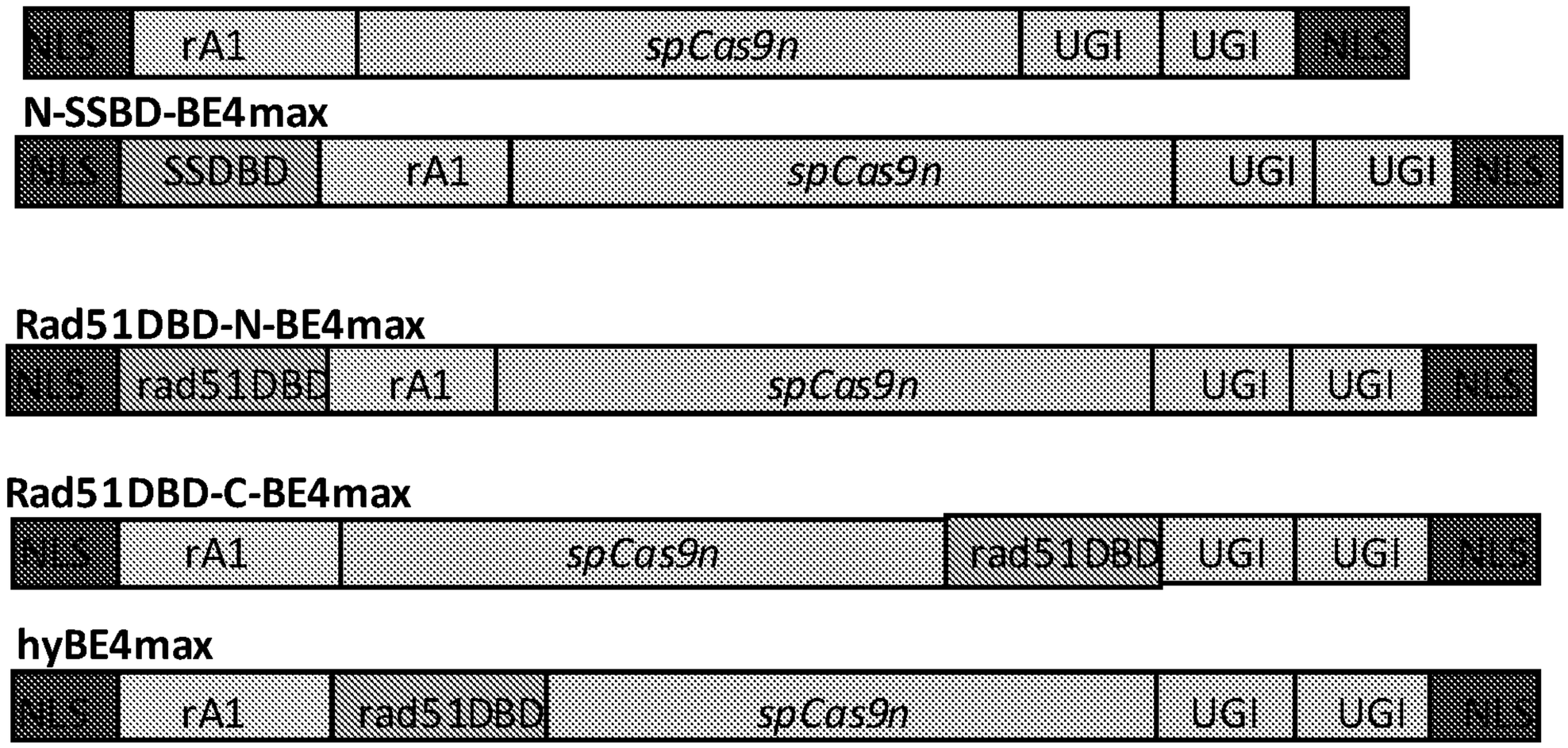

BE4max
NLS
rA1
spCas9n
UGI
UGI
NLS
N-SSBD-BE4max
NLS
SSDBD
rA1
spCas9n
UGI
UGI
NLS
Rad51DBD-N-BE4max
NLS
rad51DBD
rA1
spCas9n
UGI
UGI
NLS
Rad51DBD-C-BE4max
NLS
rA1
spCas9n
rad51DBD
UGI
UGI
NLS
hyBE4max
NLS
rA1
rad51DBD
spCas9n
UGI
UGI
NLS

图 1

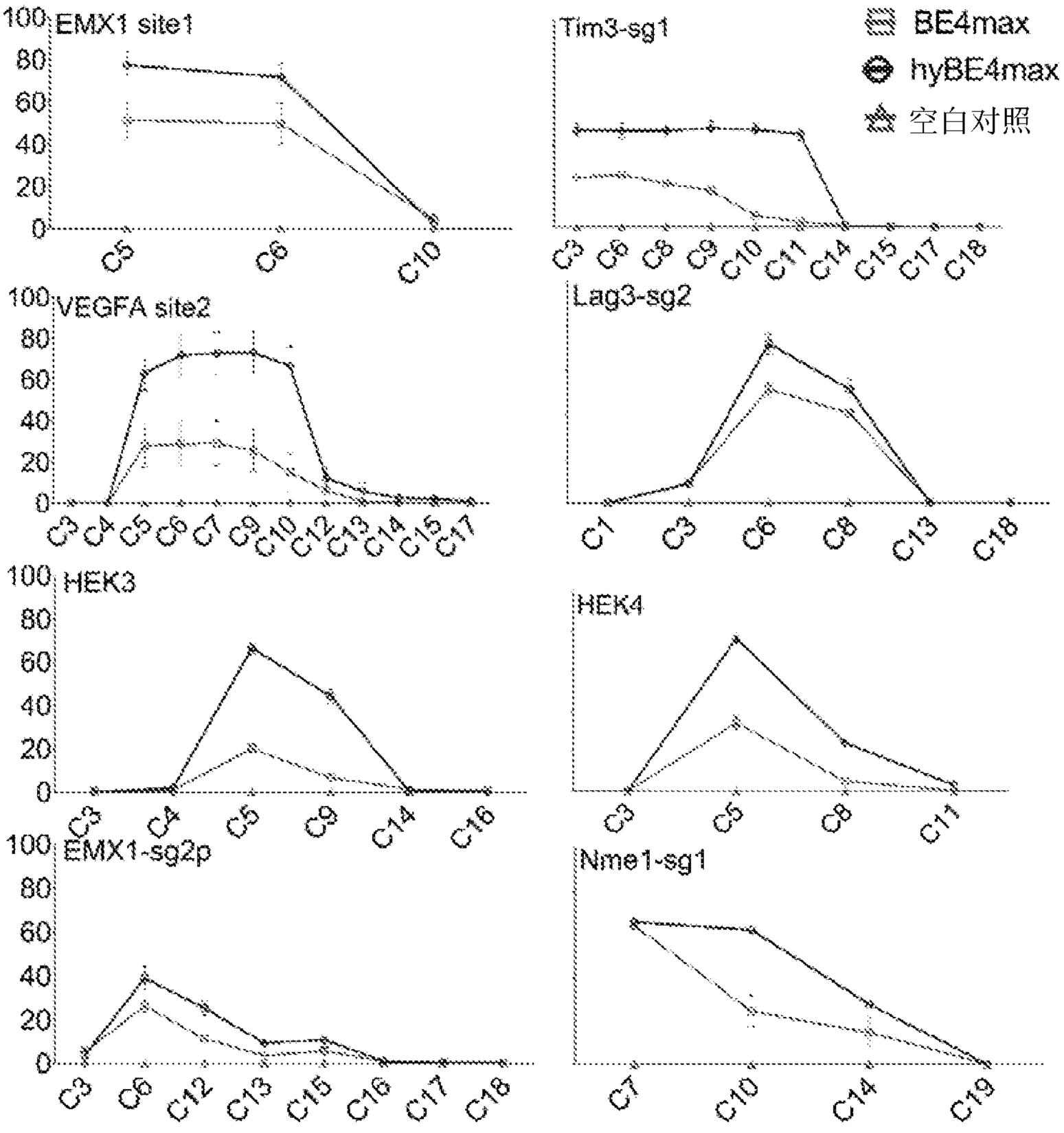

EMX1 site1
Tim3-sg1
BE4max
hyBE4max
空白对照
VEGFA site2
Lag3-sg2
HEK3
HEK4
EMX1-sg2p
Nme1-sg1

图 2

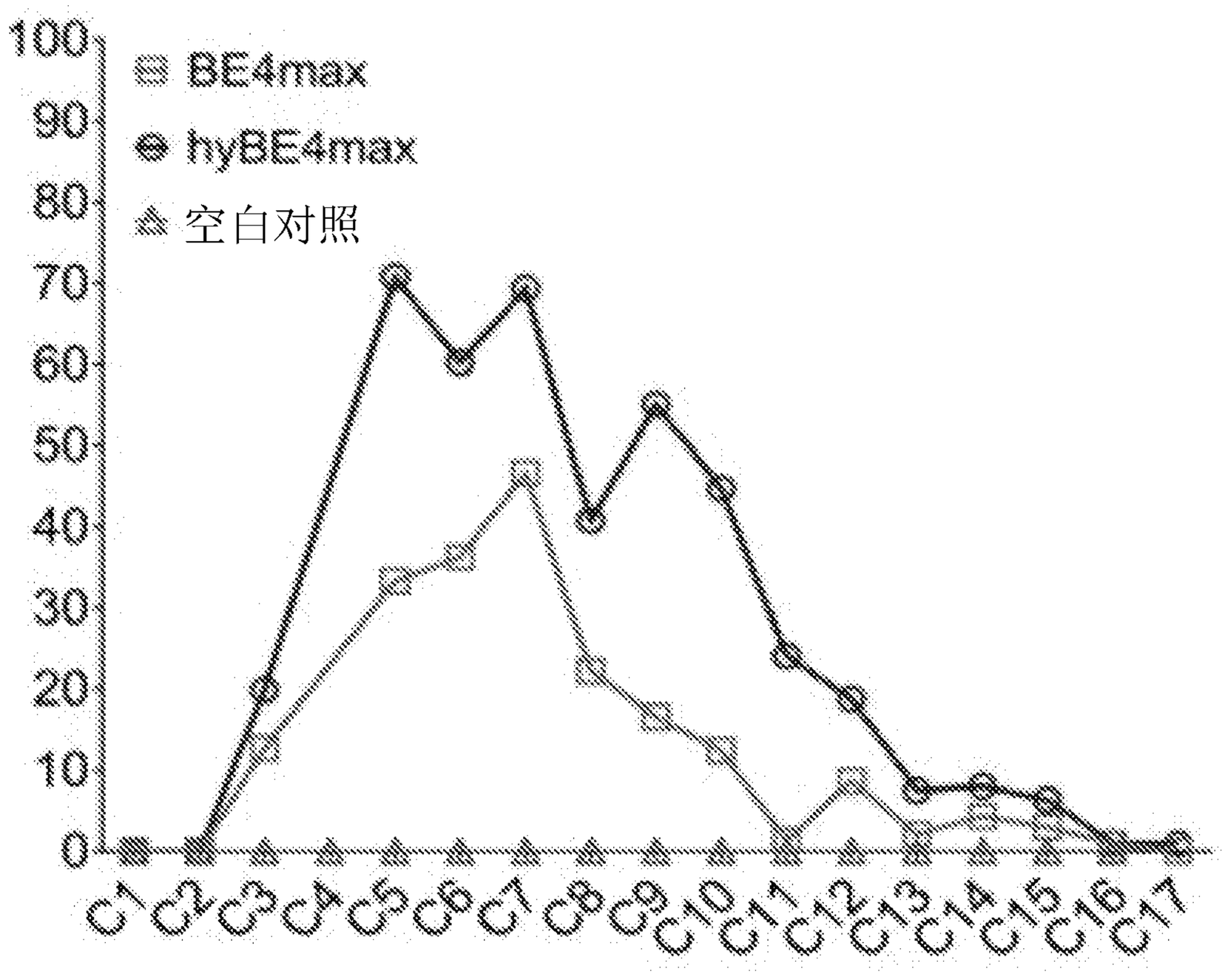
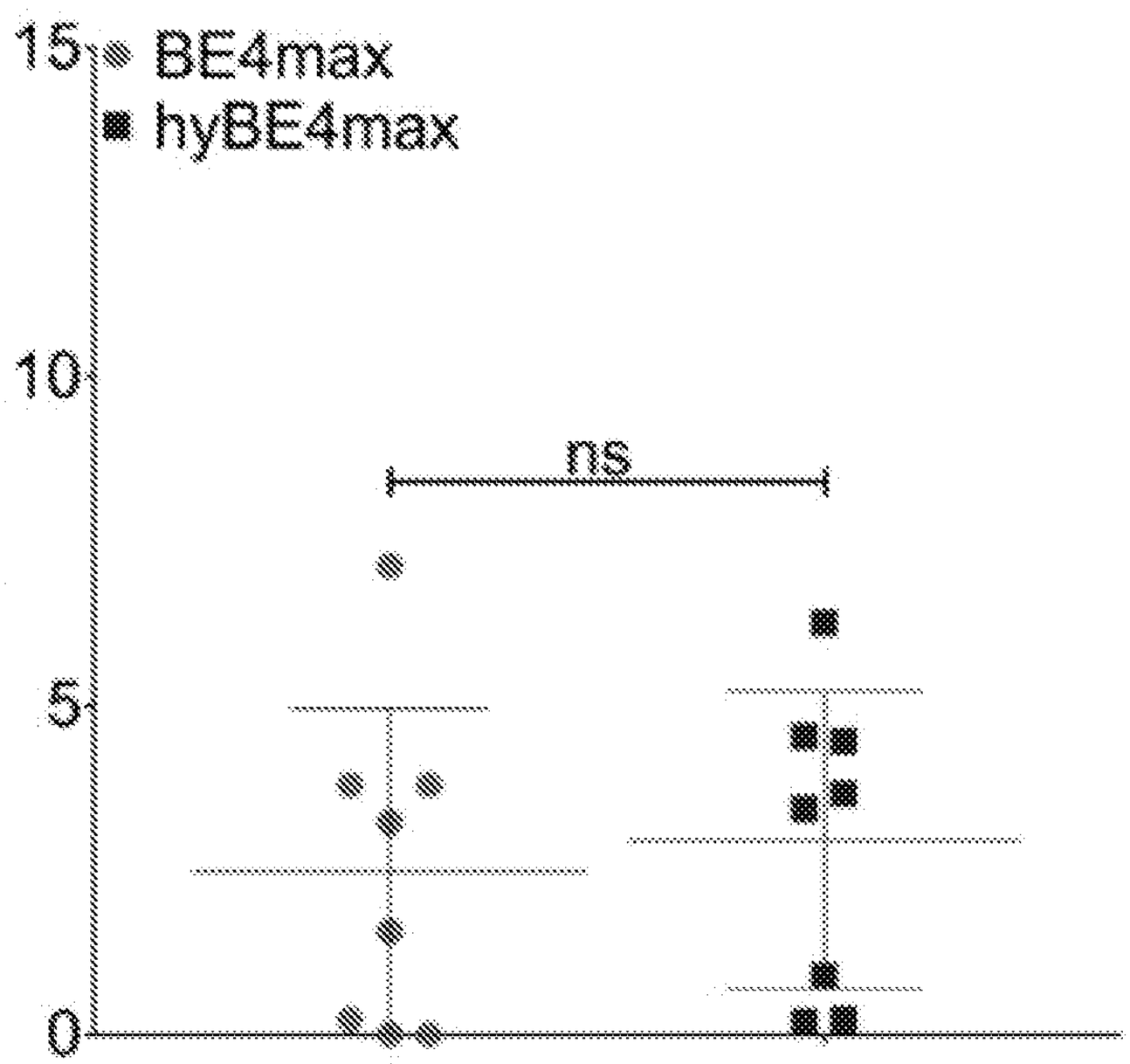

BE4max
hyBE4max
空白对照
100
90
80
70
60
50
40
30
20
10
0
C1 C2 C3 C4 C5 C6 C7 C8 C9 C10 C11 C12 C13 C14 C15 C16 C17

图 3


BE4max
hyBE4max
15
10
5
0
ns

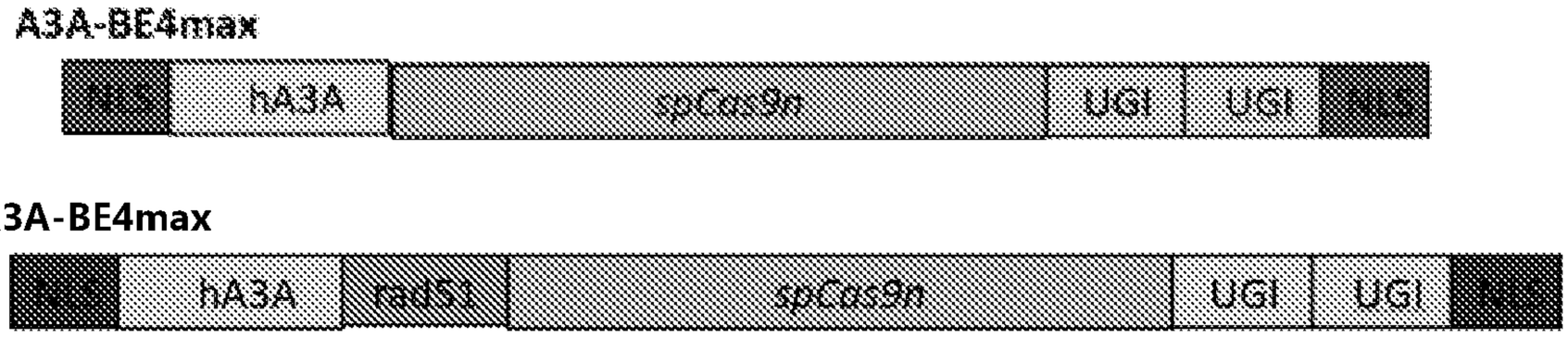

A3A-BE4max
hA3A
spCas9n
UGI
UGI
hyA3A-BE4max
hA3A
rad51
spCas9n
UGI
UGI

图 5

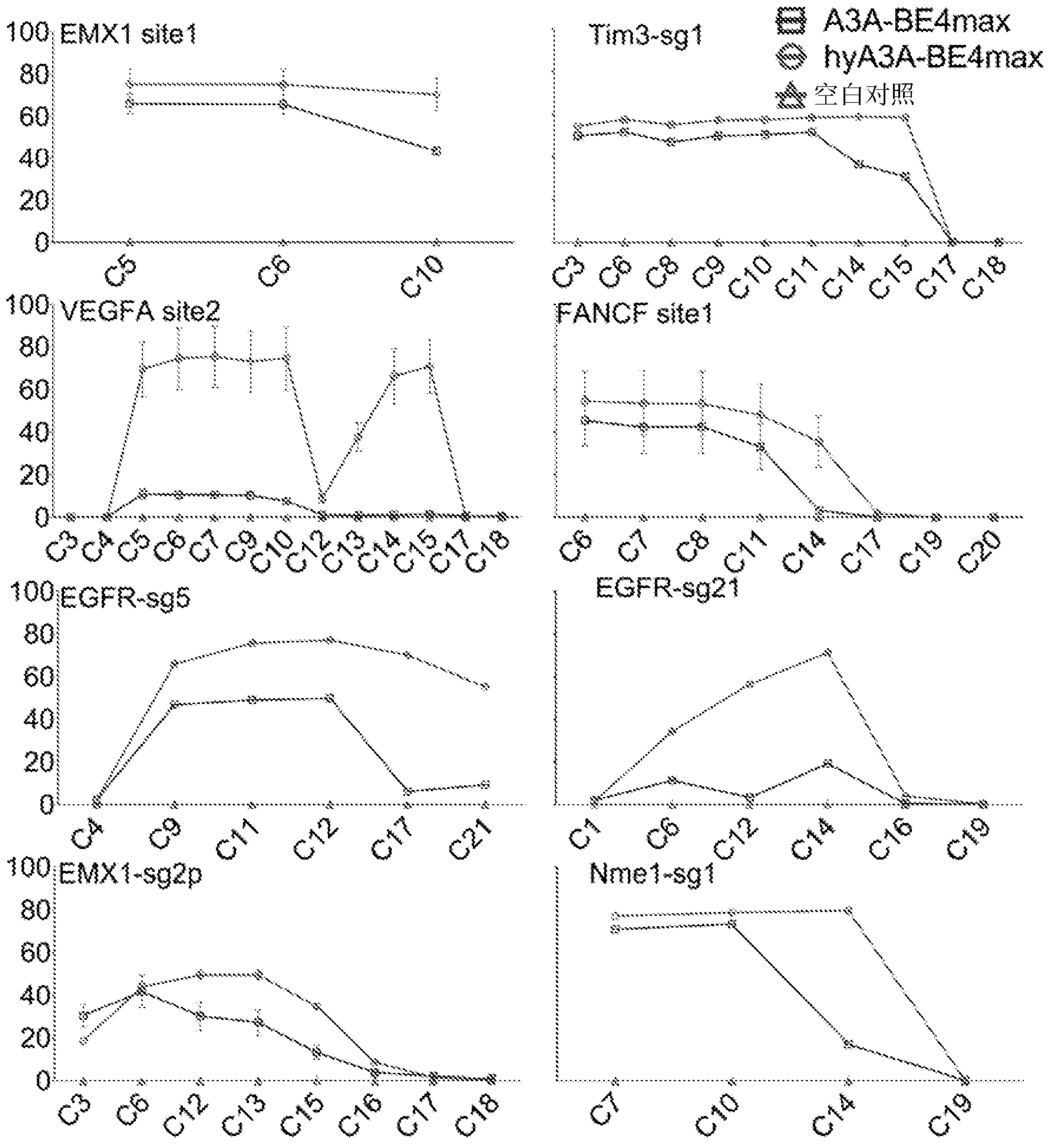

EMX1 site1
Tim3-sg1
A3A-BE4max
hyA3A-BE4max
空白对照
VEGFA site2
FANCF site1
EGFR-sg5
EGFR-sg21
EMX1-sg2p
Nme1-sg1

图 6

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2020/137239

A. CLASSIFICATION OF SUBJECT MATTER

C07K 19/00(2006.01)i; C07H 21/04(2006.01)i; A61K 48/00(2006.01)i; C12N 15/87(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

C07K; C07H; A61K; C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

TWTXT; GBTXT; SGTXT; EPTXT; USTXT; CNTXT; VEN; ISI WEB OF SCIENCES, PUBMED: 单链结合蛋白, 核苷脱氨酶, 融合蛋白, 核酸酶, 碱基编辑, single stranded binding protein, nucleotide deaminase, nuclease, APOBEC, rad51, cas9, fusion protein, base editor, SEQ ID NO:17-36

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 109880851 A (NORTHWEST AGRICULTURE & FORESTRY UNIVERSITY) 14 June 2019 (2019-06-14) SEQ ID NO: 14	34-36
X	CN 108350449 A (MASSACHUSETTS GEN HOSPITAL) 31 July 2018 (2018-07-31) table 2	34-36
X	CN 108707635 A (EAST CHINA NORMAL UNIVERSITY; SHANGHAI BRL MEDICINE CO., LTD.) 26 October 2018 (2018-10-26) embodiment 1	34-36
X	CN 108588182 A (SHENZHEN INSTITUTES OF ADVANCED TECHNOLOGY, CHINESE ACADEMY OF SCIENCES) 28 September 2018 (2018-09-28) SEQ ID NO: 2	34-36
X	CN 105647968 A (ZHEJIANG UNIVERSITY) 08 June 2016 (2016-06-08) embodiment 3	34-36
X	CHUAI, G. et al. "DeepCRISPR: Optimized CRISPR Guide RNA Design by Deep Learning" *Genome Biology,* Vol. 19, 31 December 2018 (2018-12-31), Article no. 80	34-36

☑ Further documents are listed in the continuation of Box C. ☑ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
24 February 2021	**16 March 2021**
Name and mailing address of the ISA/CN **China National Intellectual Property Administration (ISA/CN)** **No. 6, Xitucheng Road, Jimenqiao, Haidian District, Beijing 100088** **China** Facsimile No. **(86-10)62019451**	Authorized officer Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2020/137239

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	HAVLICEK, S. et al. "Re-engineered RNA-Guided FokI-Nucleases for Improved Genome Editing in Human Cells" *Molecular Therapy,* Vol. 25, No. 2, 28 February 2017 (2017-02-28), pp. 342-355	34-36
Y	KOBLAN, L. W. et al. "Improving Cytidine and Adenine Base Editors by Expression Optimization and Ancestral Reconstruction" *Nature Biotechnology,* Vol. 36, No. 9, 29 May 2018 (2018-05-29), pp. 843-848	1-33, 37-52
Y	WO 2017142923 A1 (EMENDOBIO INC.) 24 August 2017 (2017-08-24) claims 1-50	1-33, 37-52
Y	CN 109880851 A (NORTHWEST AGRICULTURE & FORESTRY UNIVERSITY) 14 June 2019 (2019-06-14) SEQ ID NO: 14	38-40, 42, 44, 45, 49
Y	CN 108350449 A (MASSACHUSETTS GEN HOSPITAL) 31 July 2018 (2018-07-31) table 2	38-40, 42, 44, 45, 49
Y	CN 108707635 A (EAST CHINA NORMAL UNIVERSITY; SHANGHAI BRL MEDICINE CO., LTD.) 26 October 2018 (2018-10-26) embodiment 1	38-40, 42, 44, 45, 49
Y	CN 108588182 A (SHENZHEN INSTITUTES OF ADVANCED TECHNOLOGY, CHINESE ACADEMY OF SCIENCES) 28 September 2018 (2018-09-28) SEQ ID NO: 2	38-40, 42, 44, 45, 49
Y	CN 105647968 A (ZHEJIANG UNIVERSITY) 08 June 2016 (2016-06-08) embodiment 3	38-40, 42, 44, 45, 49
Y	CHUAI, G. et al. "DeepCRISPR: Optimized CRISPR Guide RNA Design by Deep Learning" *Genome Biology,* Vol. 19, 31 December 2018 (2018-12-31), Article no. 80	38-40, 42, 44, 45, 49
Y	HAVLICEK, S. et al. "Re-engineered RNA-Guided FokI-Nucleases for Improved Genome Editing in Human Cells" *Molecular Therapy,* Vol. 25, No. 2, 28 February 2017 (2017-02-28), pp. 342-355	38-40, 42, 44, 45, 49
PX	ZHANG, X. H. et al. "Increasing the Efficiency and Targeting Range of Cytidine Base Editors through Fusion of a Single-Stranded DNA-binding Protein Domain" *Nature Cell Biology,* Vol. 22, 30 June 2020 (2020-06-30), pp. 740-750	1-52
PX	张晓辉 (ZHANG, Xiaohui). "碱基编辑器的优化与应用 (The Optimization and Application of Base Editing Technology)" *中国博士学位论文全文数据库基础科学辑 (China Doctoral Dissertations Full-Text Database (Basic Sciences)),* No. 8, 15 August 2020 (2020-08-15), entire document	1-52
A	GEHRKE, J. M. et al. "An APOBEC3A-Cas9 Base Editor with Minimized Bystander and Off-Target Activities" *Nature Biotechnology,* Vol. 36, No. 10, 31 October 2018 (2018-10-31), pp. 977-984	1-52

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2020/137239

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:
   a. ☑ forming part of the international application as filed:
      ☑ in the form of an Annex C/ST.25 text file.
      ☐ on paper or in the form of an image file.
   b. ☐ furnished together with the international application under PCT Rule 13*ter*.1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.
   c. ☐ furnished subsequent to the international filing date for the purposes of international search only:
      ☐ in the form of an Annex C/ST.25 text file (Rule 13*ter*.1(a)).
      ☐ on paper or in the form of an image file (Rule 13*ter*.1(b) and Administrative Instructions, Section 713).

2. ☐ In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

3. Additional comments:

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2020/137239

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	109880851	A	14 June 2019	None			
CN	108350449	A	31 July 2018	JP	2020010695	A	23 January 2020
				EP	3341477	A4	10 July 2019
				JP	2018525019	A	06 September 2018
				EP	3341477	A1	04 July 2018
				AU	2016316845	A1	22 March 2018
				CA	2996888	A1	09 March 2017
				JP	6745391	B2	26 August 2020
				KR	20180042394	A	25 April 2018
				WO	2017040348	A1	09 March 2017
				HK	1257676	A1	25 October 2019
CN	108707635	A	26 October 2018	None			
CN	108588182	A	28 September 2018	None			
CN	105647968	A	08 June 2016	CN	105647968	B	23 July 2019
WO	2017142923	A1	24 August 2017	None			

国际检索报告

国际申请号
PCT/CN2020/137239

A. 主题的分类

C07K 19/00(2006.01)i; C07H 21/04(2006.01)i; A61K 48/00(2006.01)i; C12N 15/87(2006.01)i

按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

C07K; C07H; A61K; C12N

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称，和使用的检索词（如使用））

TWTXT;GBTXT;SGTXT;EPTXT;USTXT;CNTXT;VEN;ISI WEB OF SCIENCES,PUBMED:单链结合蛋白，核苷脱氨酶，融合蛋白，核酸酶，碱基编辑，single stranded binding protein，nucleotide deaminase，nuclease，APOBEC，rad51，cas9，fusion protein，base editor，SEQ ID NO:17-36

C. 相关文件

类 型*	引用文件，必要时，指明相关段落	相关的权利要求
X	CN 109880851 A (西北农林科技大学) 2019年 6月 14日 (2019-06-14) SEQ ID NO:14	34-36
X	CN 108350449 A (通用医疗公司) 2018年 7月 31日 (2018-07-31) 表2	34-36
X	CN 108707635 A (华东师范大学 上海邦耀生物科技有限公司) 2018年 10月 26日 (2018-10-26) 实施例1	34-36
X	CN 108588182 A (中国科学院深圳先进技术研究院) 2018年 9月 28日 (2018-09-28) SEQ ID NO: 2	34-36
X	CN 105647968 A (浙江大学) 2016年 6月 8日 (2016-06-08) 实施例3	34-36
X	CHUAI, G.等. "DeepCRISPR: optimized CRISPR guide RNA design by deep learning" Genome Biology, 第19卷，2018年 12月 31日 (2018-12-31), 第80号文章	34-36

☑ 其余文件在C栏的续页中列出。

☑ 见同族专利附件。

* 引用文件的具体类型：

"A" 认为不特别相关的表示了现有技术一般状态的文件

"E" 在国际申请日的当天或之后公布的在先申请或专利

"L" 可能对优先权要求构成怀疑的文件，或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件（如具体说明的）

"O" 涉及口头公开、使用、展览或其他方式公开的文件

"P" 公布日先于国际申请日但迟于所要求的优先权日的文件

"T" 在申请日或优先权日之后公布，与申请不相抵触，但为了理解发明之理论或原理的在后文件

"X" 特别相关的文件，单独考虑该文件，认定要求保护的发明不是新颖的或不具有创造性

"Y" 特别相关的文件，当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时，要求保护的发明不具有创造性

"&" 同族专利的文件

国际检索实际完成的日期	国际检索报告邮寄日期
2021年 2月 24日	2021年 3月 16日
ISA/CN的名称和邮寄地址 中国国家知识产权局(ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088 传真号 (86-10)62019451	受权官员 李影 电话号码 86-(10)-53961973

国际检索报告

国际申请号
PCT/CN2020/137239

C. 相关文件

类 型*	引用文件，必要时，指明相关段落	相关的权利要求
X	HAVLICEK, S.等. "Re-engineered RNA-Guided FokI-Nucleases for Improved Genome Editing in Human Cells" Molecular Therapy, 第25卷，第2期，2017年 2月 28日 (2017-02-28)， 第342-355页	34-36
Y	KOBLAN, L.W.等. "Improving cytidine and adenine base editors by expression optimization and ancestral reconstruction" nature biotechnology, 第36卷，第9期，2018年 5月 29日 (2018-05-29)， 第843-848页	1-33, 37-52
Y	WO 2017142923 A1 (EMENDOBIO INC.) 2017年 8月 24日 (2017-08-24) 权利要求1-50	1-33, 37-52
Y	CN 109880851 A (西北农林科技大学) 2019年 6月 14日 (2019-06-14) SEQ ID NO:14	38-40, 42, 44-45, 49
Y	CN 108350449 A (通用医疗公司) 2018年 7月 31日 (2018-07-31) 表2	38-40, 42, 44-45, 49
Y	CN 108707635 A (华东师范大学 上海邦耀生物科技有限公司) 2018年 10月 26日 (2018-10-26) 实施例1	38-40, 42, 44-45, 49
Y	CN 108588182 A (中国科学院深圳先进技术研究院) 2018年 9月 28日 (2018-09-28) SEQ ID NO: 2	38-40, 42, 44-45, 49
Y	CN 105647968 A (浙江大学) 2016年 6月 8日 (2016-06-08) 实施例3	38-40, 42, 44-45, 49
Y	CHUAI, G.等. "DeepCRISPR: optimized CRISPR guide RNA design by deep learning" Genome Biology, 第19卷，2018年 12月 31日 (2018-12-31)， 第80号文章	38-40, 42, 44-45, 49
Y	HAVLICEK, S.等. "Re-engineered RNA-Guided FokI-Nucleases for Improved Genome Editing in Human Cells" Molecular Therapy, 第25卷，第2期，2017年 2月 28日 (2017-02-28)， 第342-355页	38-40, 42, 44-45, 49
PX	ZHANG, X.H.等. "Increasing the efficiency and targeting range of cytidine base editors through fusion of a single-stranded DNA-binding protein domain" Nature Cell Biology, 第22卷，2020年 6月 30日 (2020-06-30)， 第740-750页	1-52
PX	张晓辉. "碱基编辑器的优化与应用" 中国博士学位论文全文数据库基础科学辑, 第8期，2020年 8月 15日 (2020-08-15)， 全文	1-52
A	GEHRKE, J.M.等. "An APOBEC3A-Cas9 base editor with minimized bystander and off-target activities" nature biotechnology, 第36卷，第10期，2018年 10月 31日 (2018-10-31)， 第977-984页	1-52

国际检索报告

国际申请号
PCT/CN2020/137239

第I栏　核苷酸和/或氨基酸序列(续第1页第1.c项)

1. 关于国际申请中所公开的任何核苷酸和/或氨基酸序列,国际检索是基于下列序列表进行的:

a. ☑ 作为国际申请的一部分提交的:
   - ☑ 附件C/ST.25文本文件形式
   - ☐ 纸件或图形文件形式

b. ☐ 根据细则13之三.1(a)仅为国际检索目的以附件C/ST.25文本文件形式与国际申请同时提交的:

c. ☐ 仅为国际检索目的在国际申请日之后提交的:
   - ☐ 附件C/ST.25文本文件形式（细则13之三.1(a)）
   - ☐ 纸件或图形文件形式（细则13之三.1(b)和行政规程第713段）

2. ☐ 另外，在提交/提供了多个版本或副本的序列表的情况下，提供了关于随后提交的或附加的副本中的信息与申请时提交的作为申请一部分的序列表的信息相同或未超出申请时提交的申请中的信息范围（如适用）的所需声明。

3. 补充意见:

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2020/137239

检索报告引用的专利文件			公布日 (年/月/日)	同族专利			公布日 (年/月/日)
CN	109880851	A	2019年 6月 14日	无			
CN	108350449	A	2018年 7月 31日	JP	2020010695	A	2020年 1月 23日
				EP	3341477	A4	2019年 7月 10日
				JP	2018525019	A	2018年 9月 6日
				EP	3341477	A1	2018年 7月 4日
				AU	2016316845	A1	2018年 3月 22日
				CA	2996888	A1	2017年 3月 9日
				JP	6745391	B2	2020年 8月 26日
				KR	20180042394	A	2018年 4月 25日
				WO	2017040348	A1	2017年 3月 9日
				HK	1257676	A1	2019年 10月 25日
CN	108707635	A	2018年 10月 26日	无			
CN	108588182	A	2018年 9月 28日	无			
CN	105647968	A	2016年 6月 8日	CN	105647968	B	2019年 7月 23日
WO	2017142923	A1	2017年 8月 24日	无			