

(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
2024 年 10 月 3 日 (03.10.2024)



(10) 国际公布号
WO 2024/197857 A1

(51) 国际专利分类号:

C12N 15/113 (2010.01) *C12N 15/10* (2006.01)

C12N 15/09 (2006.01) *C12N 15/63* (2006.01)

C12N 15/11 (2006.01) *A61K 48/00* (2006.01)

(21) 国际申请号: PCT/CN2023/085605

(22) 国际申请日: 2023 年 3 月 31 日 (31.03.2023)

(25) 申请语言: 中文

(26) 公布语言: 中文

(71) 申请人: 时夕(广州)生物科技有限公司(RECORNA (GUANGZHOU) BIOTECHNOLOGY CO., LTD) [CN/CN]; 中国广东省广州市黄埔区广州国际生物岛星汉二路31号百济神州广州生物岛创新中心, Guangdong 510320 (CN)。

(72) 发明人: 杨文兵(YANG, Wenbing); 中国广东省广州市黄埔区广州国际生物岛星汉二路31号百济神州广州生物岛创新中心, Guangdong 510320 (CN)。

张锐(ZHANG, Rui); 中国广东省广州市黄埔区广州国际生物岛星汉二路31号百济神州广州生物岛创新中心, Guangdong 510320 (CN)。

(74) 代理人: 北京市万慧达律师事务所(BEIJING WANHUIDA LAW FIRM); 中国北京市海淀区中关村南大街1号友谊宾馆颐园写字楼2楼, Beijing 100873 (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, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD,

(54) Title: SCREENING METHOD FOR GUIDE RNA

(54) 发明名称: 一种筛选向导RNA的方法

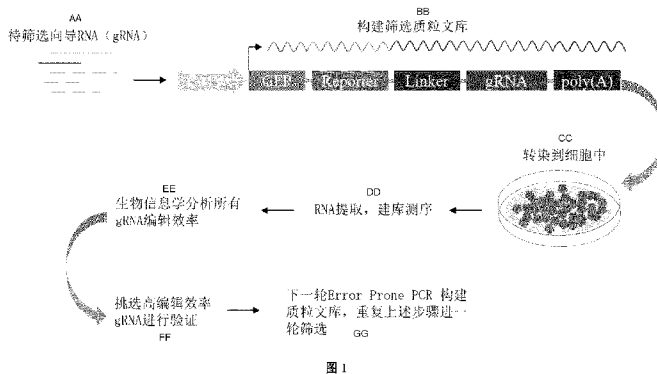


图 1

AA Guide RNAs (gRNAs) to be screened
BB Construct a screening plasmid library
CC Transfection into cells
DD RNA extraction, and construction of a library for sequencing
EE Bioinformatics analysis of the editing efficiency of all the gRNAs
FF Select a gRNA with high editing efficiency for validation
GG Perform the next round of Error-prone PCR to construct a plasmid library, and repeat the described steps for another round of screening

(57) Abstract: A screening method for a guide RNA for editing an RNA, comprising: (1) linking, on the same transcript, the guide RNA to a target RNA by means of a linker; (2) transfecting the transcript into cells by means of a vector; and (3) extracting RNAs of the cells, analyzing the editing levels of the guide RNAs for a target site, and screening for a guide RNA of interest. According to the screening method, a substrate sequence and a guide RNA sequence are designed to be on the same transcript, target RNA editing is directly completed in cells, and the sequence of each guide RNA and the editing efficiency of a target site corresponding to the guide RNA are sequenced by means of high-throughput sequencing, so that a guide RNA with high editing efficiency is rapidly screened.



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, CV, GH, GM, KE, LR, LS, MW, MZ,
NA, RW, SC, 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, ME, 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))。

(57) 摘要: 一种编辑RNA的向导RNA的筛选方法, 包括: (1) 将所述向导RNA通过连接子分别与靶RNA连接在同一个转录本上; (2) 将所述的转录本通过载体转入细胞中; (3) 提取所述细胞的RNA, 分析所述向导RNA对靶位点的编辑水平, 筛选出目的向导RNA。所述筛选方法将底物序列和向导RNA序列设计为同一个转录本, 直接在细胞内完成靶RNA编辑, 通过高通量测序将每个向导RNA的序列和其对应的靶位点的编辑效率测序出来, 从而快速筛选出高编辑效率的向导RNA。

说明书

发明名称：一种筛选向导RNA的方法

技术领域

[0001] 本发明属于生物医药领域，涉及一种筛选向导RNA的方法。

背景技术

[0002] 近年来随着三代测序技术的快速发展，被誉为生命科学“登月计划”的人类基因组测序最后一块拼图已经完成，新增绘制了近2亿个碱基对的完整无间隙基因组图谱，这将意味着越来越多与疾病相关的基因突变位点可被更彻底的挖掘出来。在此之前已经发现超过3000多个基因突变位点与疾病的发生发展有密切关联，而很多突变位点却很难成为合适的成药靶点，除此之外同一个疾病可能对应多个突变位点，且在人群中分布不均，让传统的大分子小分子药物研发变得更加艰难。此时，基因疗法慢慢展现出其巨大的临床应用价值，直接对突变位点进行修复达到疾病治疗的目的。

[0003] CRISPR(Clustered regularly interspaced short palindromic repeats)最早发现于细菌免疫系统中，可结合cas蛋白抵抗外源病毒的入侵。自2012年被报道该系统也可在哺乳动物细胞中进行基因组的剪切及修复后对CRISPR/Cas的研究如同雨后春笋，其编辑的特性机制、可应用的物种组织等被一一攻克，目前CRISPR/Cas技术已经广泛应用到生物医学及农业学研究中。CRISPR/Cas系统可被设计在基因组中插入完整的基因、敲低基因的表达、敲除基因或者对单个突变位点进行修复等功能。但该系统也有其致命的缺陷，该系统需要同时表达Cas蛋白及向导RNA(gRNA)及提供同源修复DNA序列，且在多篇研究报道中都检测出其在DNA水平或RNA水平有明显的脱靶效应，CRISPR/Cas蛋白切割基因组DNA造成DSB(double-strand break)，脱靶会造成基因组的不稳定性，此外，CRISPR/Cas系统还可能会引发细胞固有免疫反应，这些都极大的限制了其在临床上的应用。

[0004] RNA水平的A-to-I编辑则是在ADAR家族蛋白的介导下发生，I碱基在翻译的时候会被识别为G碱基，故在基因组上G-to-A的突变就可以通过ADAR的编辑在RNA水

平进行修复。ADAR家族蛋白主要有三个亚型，ADAR1, ADAR2和ADAR3，其中ADAR3不具有脱氨酶活性，编辑事件主要由ADAR1和ADAR2完成，且ADAR1蛋白在人体组织中广谱表达可确保其应用到多种组织或器官中。目前已经开发出多种利用过表达ADAR家族蛋白或者内源表达的ADAR家族蛋白结合向导RNA在RNA水平进行特定定位点的A-to-I编辑。从临床应用的角度分析最具应用价值的方法则是利用内源表达的ADAR家族蛋白对突变位点进行修复，因为只需对gRNA进行设计，而目前比较成熟的技术则是如下两种：第一种gRNA由两段组成，一段是54nt长度的带有结构特性的ADAR底物加上另外一段可定制的靶向序列(RESTORE方法)，该方法优化版本可减少脱靶编辑(CLUSTER方法)；第二种gRNA只有一段与靶标序列互补的长度大于100nt的靶向序列(LEAPER方法)。ADAR家族蛋白偏好于编辑双链RNA，但在编辑区域范围内同时也会存在多个突变形形成一定的二级结构，这就导致在设计gRNA的时候需对gRNA进行筛选，而目前还未有报道出比较高效的筛选方法。

[0005] 发明内容

[0006] 一些实施方案中，本发明提供了一种编辑RNA的向导RNA的筛选方法，包括步骤：(1)待筛选的每条向导RNA连有连接子，并与靶RNA连接在能够形成同一个转录本的核酸链上，形成多个所述的核酸链；(2)将所述的核酸链转入表达ADAR蛋白的细胞中；(3)提取和测定所述细胞中的RNA，确定靶RNA的被编辑水平，通过条形码序列确定对应的向导RNA的编辑水平。

[0007] 一些实施方案中，所述的连接子位于靶RNA和向导RNA之间。

[0008] 一些实施方案中，所述的条形码序列包含于所述连接子中。

[0009] 一些实施方案中，所述核酸链从上游到下游依次连接有报告基因核酸片段、靶RNA核酸片段、连接子核酸片段、向导RNA核酸片段。

[0010] 连接子序列在本发明中起到标签的作用。连接子优选地可以位于靶RNA和向导RNA之间，此时连接子同时起到在一定程度隔开靶RNA和向导RNA的作用，有利于两者折叠成二级结构。并且，此种连接方式还有利于在二代测序中应用，从而实现高效高通量的检测。为了适应二代测序的序列长度限制，测序时可以

切割成靶RNA+连接子、连接子+向导RNA两部分测序结果，通过连接子的标签作用，可获悉相应向导RNA作用下靶RNA的编辑水平。

- [0011] 一些实施方案中，所述转录本从上游到下游依次连接有报告基因核酸片段、靶RNA核酸片段、连接子核酸片段、向导RNA核酸片段和polyA尾核酸片段。
- [0012] 一些实施方案中，所述报告基因选自yfp基因、gus基因、rfp基因、gfp基因、cfp基因、bfp基因或卡那霉素抗性基因，但又不限于此。
- [0013] 一些实施方案中，所述连接子的长度为26nt-2121nt。
- [0014] 一些实施方案中，所述连接子包括SEQ ID NO:549或SEQ ID NO:557。
- [0015] 一些实施方案中，筛选方法中加入的是长连接子序列，目的是为了模拟分子间RNA之间形成双链进行编辑。
- [0016] 一些实施方案中，在短连接子的情况下不同长度的gRNA混合的文库也可以通过本发明的方法筛选出不同编辑效率的gRNA。
- [0017] 一些实施方案中，步骤(3)包括建立靶向测序文库进行二代测序。
- [0018] 一些实施方案中，所述核酸链接触ADAR蛋白通过将所述核酸链转入表达ADAR蛋白的细胞中实现。
- [0019] 一些实施方案中，所述细胞表达ADAR蛋白。
- [0020] 一些实施方案中，所述ADAR蛋白包括ADAR1或ADAR2。
- [0021] 一些实施方案中，所述向导RNA能够在所述细胞中招募ADAR蛋白对靶RNA进行编辑。
- [0022] 一些实施方案中，所述的细胞过表达ADAR蛋白。
- [0023] 一些实施方案中，所述编辑为A-to-I编辑。
- [0024] 一些实施方案中，所述向导RNA为以A碱基为中心上下游延伸得到的向导RNA。
- [0025] 一些实施方案中，所述向导RNA为以A碱基为中心上下游延伸15nt~75nt得到的30nt~151nt长度的向导RNA。
- [0026] 一些实施方案中，所述向导RNA为以A碱基为中心上下游延伸75nt得到的151nt的向导RNA。
- [0027] 一些实施方案中，所述向导RNA为以A碱基为中心上下游延伸得到的28nt~89nt的向导RNA。

- [0028] 一些实施方案中，所述筛选方法包括把编辑位点上下游20-200nt序列加入报告基因的下游并插入到载体中获得报告基因载体，再把报告基因载体转入细胞中。
- [0029] 一些实施方案中，所述筛选方法为高通量筛选方法。
- [0030] 一些实施方案中，本发明提供了一种可招募ADAR蛋白并对特定定位点进行A-to-I编辑的高通量筛选向导RNA(gRNA)序列的方法，填补了现有技术的空缺。
- [0031] 一些实施方案中，本发明提供了上述方法对基于ADAR的任何RNA编辑系统建立的gRNA筛选方法。
- [0032] 一些实施方案中，所述靶RNA序列长度大于向导RNA序列。
- [0033] 一些实施方案中，所述载体包括质粒或病毒载体。
- [0034] 一些实施方案中，所述载体是用于在高等真核细胞或原核细胞中表达的质粒或病毒载体。一些实施方案中，所述载体骨架为pcDNA3.1载体。
- [0035] 一些实施方案中，所述筛选方法包括：将多条不同的所述向导RNA分别与所述靶RNA连接组成转录本体系，通过所述载体转入细胞中。
- [0036] 一些实施方案中，所述目的向导RNA为靶向位点进行高效编辑的向导RNA。
- [0037] 一些实施方案中，所述方法还包括：从所述步骤(3)中获得所述向导RNA的编辑水平后，挑选编辑水平高的向导RNA合成序列相同的单链DNA后设计易错PCR引物对所述向导RNA的靶向区域进行随机引入突变，再重复所述步骤(1)~(3)。
- [0038] 一些实施方案中，本发明提供了一种编辑RNA的向导RNA的筛选方法，包括：
(a) 向导RNA文库的设计；(b) 把需要编辑的基因片段与报告基因插入到载体中获得报告基因载体；(c) 把所述向导RNA文库插入所述报告基因载体获得待筛选质粒文库；(d) 将所述待筛选质粒文库转入到细胞，转染后，提取所述细胞中的RNA，测序；(e) 使用生物信息学分析所述向导RNA的编辑水平；其中(a)和(b)的顺序可以调换。
- [0039] 一些实施方案中，步骤(b)包括：把编辑位点上下游序列加入报告基因的下游插入到载体中获得报告基因载体。一些实施方案中，步骤(b)包括：把编辑位点上下游20-200nt序列加入报告基因的下游插入到载体中获得报告基因载体。

- [0040] 一些实施方案中，步骤(b)包括：将荧光基因全长以及需要编辑的基因片段通过PCR扩增出来同时两端加上同源重组的同源序列或引入酶切位点，同时对空载体进行NheI和HindIII双酶切并回收线性载体，将PCR扩增的多片段插入到线性载体中获得报告基因的载体。
- [0041] 一些实施方案中，步骤(c)包括：扩增步骤(a)中的所述向导RNA同时加上同源臂或酶切位点获得向导RNA序列池，步骤(b)中的报告基因载体用BamHI及EcoRI进行双酶切线性化；然后把所述的向导RNA序列池插入到线性化后的报告基因载体中，转化后获得待筛选质粒文库。
- [0042] 一些实施方案中，所述筛选方法还包括：(f)从所述向导RNA中挑选出编辑效率高的向导RNA直接合成序列相同的单链DNA后，设计易错PCR引物对所述向导RNA靶向区域进行随机引入突变，再重复步骤(b)～(e)。
- [0043] 一些实施方案中，本发明提供了一种可招募ADAR蛋白并对特定定位点进行A-to-I编辑的高通量筛选向导RNA序列的方法，包括如下步骤：S1. 对需要筛选的gRNA文库进行接头序列设计并合成，具体地，gRNA文库一端为芯片合成所需的固定序列，另一端为可区分不同文库的20nt左右的接头序列；S2. 筛选载体的设计及构建，具体地，本发明设计了在绿色荧光蛋白GFP编码区下游添加靶标RNA和gRNA序列，三者在同一条转录本的策略进行实施，三者之间可以直接连接也可以通过不同长度的接头序列进行间隔；S3. 待筛选的gRNA质粒文库的构建，具体地，通过对应不同文库的接头序列对混合的文库进行PCR扩增并进一步插入到S2的载体系统中；S4. 通过转染的方式把质粒文库转染到特定的细胞中；S5. RNA提取及靶向测序文库构建并进行二代测序；S6. 生物信息学方法分析每条gRNA对应靶标RNA的编辑效率；S7. 挑选高编辑水平的gRNA序列合成序列相同的单链DNA并进行易错PCR构建质粒文库并按照S3至S6的方法进行第二轮的筛选。
- [0044] 一些实施方案中，本发明提供了一种可招募ADAR蛋白并对特定定位点进行A-to-I编辑的高通量筛选向导RNA序列的方法。本发明通过开发实验及生物信息学方法填补了高通量筛选基于ADAR的靶标RNA编辑的gRNA序列方法上的空缺，通过一系列优选的实验设计方法完成筛选。

- [0045] 一些实施方案中，本发明所述的方法包括1)采用模拟基因不同区域的方式把靶标RNA及gRNA设计到同一个转录本上，荧光基因的引入便于观测转染效率且可替换成其他荧光基因；2)构建靶向测序文库及分析方法可以精准的把每条gRNA与其编辑效率一一对应；3)在筛选出的高编辑水平的gRNA序列基础上进行随机引入突变进一步筛选出更高编辑水平的gRNA。
- [0046] 一些实施方案中，本发明提供了任一所述的筛选方法获得的向导RNA。
- [0047] 一些实施方案中，本发明提供了一种向导RNA，所述向导RNA的序列如SEQ ID NO:557、SEQ ID NO:559、SEQ ID NO:561、SEQ ID NO:563、SEQ ID NO:565、SEQ ID NO:567、SEQ ID NO:569、SEQ ID NO:571、SEQ ID NO:573、SEQ ID NO:575、SEQ ID NO:577、SEQ ID NO:579、SEQ ID NO:583～SEQ ID NO:602、SEQ ID NO:603～SEQ ID NO:622、或SEQ ID NO:623～SEQ ID NO:642任一所示。
- [0048] 一些实施方案中，本发明提供了获得所述的向导RNA的方法或者获得的向导RNA在制备治疗疾病的药物中的用途。
- [0049] 一些实施方案中，所述药物用于在真核细胞中定点编辑靶基因中的核苷酸从而达到治疗疾病的目的。
- [0050] 一些实施方案中，所述疾病为包括GAPDH基因、ATP7B基因或FGFR基因突变引起的疾病。
- [0051] 一些实施方案中，所述GAPDH基因特定位点基序为AAC、AAU、AAG、UAA或UAG，但又不限于此。
- [0052] 一些实施方案中，所述ATP7B基因特定位点基序为UAG，但又不限于此。
- [0053] 一些实施方案中，所述FGFR基因特定位点基序为CAG，但又不限于此。
- [0054] 一些实施方案中，所述疾病包括肝豆状核变性、软骨发育不全、致死性骨发育不良或矮小症，但又限于此。
- [0055] 一些实施方案中，所述用途是通过天然存在于所述细胞中并且能够进行所述核苷酸的编辑的RNA编辑实体的作用。

- [0056] 一些实施方案中，本发明提供了一种工程化RNA，其结构包括：从上游到下游依次连接有报告基因核酸片段、靶RNA核酸片段、连接子核酸片段、向导RNA核酸片段。
- [0057] 一些实施方案中，本发明提供了一种工程化RNA，其结构包括：从上游到下游依次连接有报告基因核酸片段、靶RNA核酸片段、连接子核酸片段、向导RNA核酸片段和polyA尾核酸片段。
- [0058] 一些实施方案中，所述报告基因选自yfp基因、gus基因、rfp基因、gfp基因、cfp基因、bfp基因、或卡那霉素抗性基因。
- [0059] 一些实施方案中，所述连接子的长度为26nt到2121nt。
- [0060] 一些实施方案中，所述连接子包括SEQ ID NO:549或SEQ ID NO:557。
- [0061] 一些实施方案中，所述向导RNA为以A碱基为中心上下游延伸得到的向导RNA。
- [0062] 一些实施方案中，所述向导RNA为以A碱基为中心上下游延伸75nt得到的151nt的向导RNA。
- [0063] 一些实施方案中，本发明提供了一种工程化RNA结构文库，包含所述的多个不同的工程化RNA。
- [0064] 一些实施方案中，所述不同的工程化RNA是通过将靶RNA通过不同的连接子与不同的向导RNA连接获得的。
- [0065] 一些实施方案中，所述工程化RNA结构文库中，所述文库中，所述的工程化RNA所含有的靶RNA核酸片段是相同的，向导RNA核酸片段序列是不同的。
- [0066] 一些实施方案中，本发明提供了一种载体，包含所述的工程化RNA或所述的工程化RNA结构文库。
- [0067] 一些实施方案中，所述载体通过把靶RNA序列加入到报告基因序列的下游插入到载体骨架中，获得报告基因载体，再把扩增好的向导RNA文库插入入报告基因载体中获得的。
- [0068] 一些实施方案中，本发明提供了一种细胞，包含所述的工程化RNA或所述的工程化RNA结构文库或所述的载体。

附图说明

- [0069] 图1为高通量筛选方法示意图。

- [0070] 图1显示高通量筛选向导RNA的整体流程，待筛选序列的设计；合成好的待筛选序列构建到筛选载体；导入到细胞中进行编辑；RNA提取、二代测序文库构建及高通量测序；利用生物信息学方法对测序数据进行分析获取每条gRNA的编辑效率；挑选高编辑效率的gRNA进行验证；从高编辑gRNA中挑选多条进行Error prone随机引入突变再次重复筛选流程得到更高编辑效率的gRNA。
- [0071] 图2为筛选载体转录本信息示意图。
- [0072] 图2显示筛选载体具体的转录本组成信息，以GFP作为示例构建一个可以同时检测到编辑水平和gRNA的转录本，该转录本模拟正常基因的表达，GFP为编码区，待编辑的靶向序列及连接子及gRNA模拟3' UTR区域，其中靶RNA (Reporter) 和gRNA中会有一段可预设长度的连接子(Linker)序列，一方面可以间隔开二者模拟分子间形成二级结构，另一方面在Linker区域添加Barcode (用于区分不同的gRNA)序列可以在后续对高通量测序数据分析时对不同的gRNA及对应的编辑水平做区分。
- [0073] 图3筛选载体转录本进行RNA编辑模式图。
- [0074] 图3显示筛选载体表达出RNA后会进行自我折叠形成Reporter和gRNA的双链RNA区域，该区域会被ADAR蛋白识别结合并对靶位点进行编辑。
- [0075] 图4为GAPDH基因不同位点gRNA筛选结果。(图4a)编辑位点1为AAC基序的结果。(图4b)编辑位点2为AAU基序的结果。(图4c)编辑位点3为AAG基序的结果。(图4d)编辑位点4为UAA基序的结果。
- [0076] 图4显示GAPDH基因上选择的四个不同基序的编辑位点进行高通量筛选得到的每条gRNA的编辑效果，其中横纵坐标表示的为两次生物学重复。图中黑色的点为编辑位点对位为C碱基其他位置为互补配对的gRNA，非黑色的点则为其他的gRNA在该序列的基础上引入不同的突变等得到的序列，每个位点对应的编辑水平均为高通量测序获得。
- [0077] 图5为GAPDH基因不同位点gRNA编辑及脱靶编辑比较。
- [0078] 图5显示GAPDH基因不同位点gRNA在靶位点的编辑及上下游的脱靶编辑热图比较。其中横坐标的0表示靶位点被编辑的A碱基的位置，0~-25则为编辑位点上游的区间，0~25为编辑位点下游的区间，纵坐标为每一条gRNA在不同A碱基

上的编辑水平的情况，按照靶位点最高编辑序列往下排列，蓝色到红色表示A碱基位点的编辑水平从低到高的变化趋势。

[0079] 图6为不同长度gRNA靶向人GAPDH基因的第787位点的编辑及脱靶编辑效率。

[0080] 图6显示GAPDH基因的第787位所有gRNA在靶位点编辑及上下游脱靶编辑热图。

其中横坐标的-1与1之间为靶位点，-1~-25为编辑位点上游的区间，1~25为编辑位点下游的区间，纵坐标为每一条gRNA在不同A碱基上的编辑水平情况，按照靶位点最高编辑序列往下排列，蓝色到红色表示A碱基位点的编辑水平从低到高的变化趋势。

[0081] 图7高通量筛选gRNA验证分析。

[0082] 图7显示高通量筛选出的gRNA用低通量的方式进行验证。横坐标表示每条gRNA在高通量筛选的方法中测量得到的编辑水平，纵坐标为每条gRNA单独合成修饰后的RNA并转染到细胞内测量得到的编辑水平。相关性系数 R^2 值越高表示横纵坐标相关性越高。

[0083] 图8为单条gRNA在GAPDH基因上的编辑水平一代测序原始结果。

[0084] 图8显示单条修饰后的gRNA转染到细胞中通过一代测序获得的编辑水平的原始峰图。如图所示，不同的碱基对应不同颜色的峰，A碱基对应绿色的峰，G碱基对应黑色的峰，当一个位点出现不同颜色的峰表示该位点存在不同的碱基组成，在本结果指示A碱基的编辑水平，编辑水平的计算方法为G碱基的峰面积除以该位点A碱基和G碱基的峰面积之和。

[0085] 图9为人ATP7B基因gRNA筛选结果。

[0086] 图9显示ATP7B基因上选择的编辑位点进行高通量筛选得到的每条gRNA的编辑效果，其中横纵坐标表示的为两次生物学重复。图中黑色的点为编辑位点对位为C碱基其他位置为互补配对的gRNA，非黑色的点则为其他的gRNA在该序列的基础上引入不同的突变等得到的序列，每个点对应的编辑水平均为高通量测序获得，len41表示靶编辑位点及上下游41nt长度对应设计出的带有不同突变的gRNA，len31表示靶编辑位点上下游31nt长度对应设计出的带有不同突变的gRNA。

[0087] 图10为人ATP7B基因gRNA靶向编辑及脱靶编辑效率。

- [0088] 图10显示ATP7B基因所有gRNA在靶位点编辑及上下游脱靶编辑热图。其中横坐标的0位置为靶位点，0~-20为编辑位点上游的区间，0~20为编辑位点下游的区间，纵坐标为每一条gRNA在不同A碱基上的编辑水平情况，按照靶位点最高编辑序列往下排列，蓝色到红色表示A碱基位点的编辑水平从低到高的变化趋势。
- [0089] 图11为ATP7B基因UAG基序优选gRNA的第二轮筛选结果。
- [0090] 图11显示ATP7B基因通过第一轮筛选得到的不同长度的高编辑效率前20条序列进行Error prone引入随机突变并进行第二轮筛选，筛选载体转染细胞后用高通量测序方法测量新产生的gRNA和其对应的编辑效率，图中横纵坐标表示的为两次生物学重复。len41表示选择41nt长度组的gRNA进行第二轮筛选的结果，len31表示选择31nt长度组的gRNA进行第二轮筛选的结果。
- [0091] 图12人FGFR基因gRNA筛选结果。
- [0092] 图12显示FGFR基因上选择的编辑位点进行高通量筛选得到的每条gRNA的编辑效果，其中横纵坐标表示的为两次生物学重复。图中黑色的点为编辑位点对位为C碱基其他位置为互补配对的gRNA，非黑色的点则为其他的gRNA在该序列的基础上引入不同的突变等得到的序列，每个点对应的编辑水平均为高通量测序获得，len41表示靶编辑位点及上下游41nt长度对应设计出的带有不同突变的gRNA，len31表示靶编辑位点上下游31nt长度对应设计出的带有不同突变的gRNA。
- [0093] 图13为人FGFR基因gRNA靶向编辑及脱靶编辑效率。
- [0094] 图13显示FGFR基因所有gRNA在靶位点编辑及上下游脱靶编辑热图。其中横坐标的0位置为靶位点，0~-20为编辑位点上游的区间，0~20为编辑位点下游的区间，纵坐标为每一条gRNA在不同A碱基上的编辑水平情况，按照靶位点最高编辑序列往下排列，蓝色到红色表示A碱基位点的编辑水平从低到高的变化趋势。
- [0095] 图14为人FGFR基因CAG基序优选gRNA的第二轮筛选结果。
- [0096] 图14显示FGFR基因通过第一轮筛选得到的41nt长度高编辑效率前20条序列进行Error prone引入随机突变并进行第二轮筛选，筛选载体转染细胞后用高通量测序方法测量新产生的gRNA和其对应的编辑效率，图中横纵坐标表示的为两次生物学重复。len41表示选择41nt长度组的gRNA进行第二轮筛选的结果。

具体实施方式

- [0097] 以下通过具体的实施例进一步说明本发明的技术方案，具体实施例不代表对本发明保护范围的限制。其他人根据本发明理念所做出的一些非本质的修改和调整仍属于本发明的保护范围。
- [0098] 除非特别说明，以下实施例所用试剂和材料均为市场上可以购买获得的。
- [0099] 本文中向导RNA高通量筛选方法流程如图1所示。
- [0100] 如本文和所附权利要求书所用，单数形式“一个/一种”、“一个/一种”和“所述”包括复数指代物，除非上下文另外清楚地规定。因此，例如，提及“一种方法”包括多个此类方法，并且提及“所述片段”包括提及一个或多个片段及其本领域技术人员已知的等同物，等等。
- [0101] 此外，除非另外说明，否则“或”的使用意指“和/或”。类似地，“包含”、“包括”是可互换的，并不旨在是限制性的。
- [0102] 应进一步理解，在使用术语“包含”描述各个实施方案的情况下，本领域技术人员将理解，在一些特定情形下，可以可替代地使用语言“基本上由……组成”或“由……组成”来描述实施方案。
- [0103] 除非另外定义，否则本文所用的所有技术和科学术语都具有与本公开文本所属领域的普通技术人员通常所理解的相同的含义。尽管许多方法和试剂与本文所述的那些类似或等同，但本文公开了示例性方法和材料。
- [0104] 应当理解，本公开文本不限于本文所述的特定方法、方案和试剂等，并且本身可以变化。本文所用的术语仅用于描述特定实施方案或方面的目的，并不旨在限制本公开文本的范围。
- [0105] “工程化的多核苷酸”或“工程化的向导RNA”可与环状向导RNA互换地使用。经工程化的多核苷酸可以包括DNA或RNA或杂交DNA/RNA构建体的重组多核苷酸。经工程化的多核苷酸可以产生向导RNA，更具体地可以产生环状向导RNA。
- [0106] 如本文所用，术语“突变”可以指代编码蛋白质的核酸序列相对于所述蛋白质的共有序列的改变。“错义”突变导致一个密码子取代为另一个密码子；“无义”突变将密码子从编码特定氨基酸的密码子改变为终止密码子。无义突变通常导致蛋白质的截短翻译。“沉默”突变是对所得蛋白质没有影响的突变。

如本文所用，术语“点突变”可以指代只影响基因序列中的一个核苷酸的突变。“剪接位点突变”是存在于前体mRNA(在加工以除去内含子之前)中的那些突变，其由于剪接位点的不正确描绘而导致蛋白质的错译并且通常导致截短。突变可以包括单核苷酸变异(SNV)。突变可以包括序列变体、序列变异、序列改变或等位基因变体。参考DNA序列可以从参考数据库获得。突变可能影响功能。突变可能不影响功能。突变可以在DNA水平上在一个或多个核苷酸中发生，在核糖核酸(RNA)水平上在一个或多个核苷酸中发生，在蛋白质水平上在一个或多个氨基酸中发生，或其任何组合。参考序列可以从诸如NCBI参考序列数据库(RefSEQ ID NO:)数据库等的数据库获得。可以构成突变的特定改变可以包括在一个或多个核苷酸或者一个或多个氨基酸中的取代、缺失、插入、倒位或转换。突变可以是点突变。

[0107] 术语“RNA编辑”是指，在基因组编码的RNA序列中引入变化，从而导致RNA突变的共转录或转录后修饰过程。双链RNA(dsRNA)中的腺苷编辑为肌苷(A-到-I)，由作用于RNA(ADAR)酶家族的腺苷脱氨酶催化，是哺乳动物中常见的RNA编辑类型。在脊椎动物中，先前已对三种ADAR蛋白ADAR1、ADAR2和ADAR3的家族进行了表征。ADAR1和ADAR2(ADAR)催化所有当前已知的A-到-I编辑位点。ADAR3没有已知的脱氨酶活性。肌苷(I)模拟鸟苷(G)，因此ADAR蛋白在转录物中引入了虚拟的A到G取代。这种变化可以导致特定的氨基酸取代、可变剪接、微RNA介导的基因沉默或转录物定位和稳定性的变化。

[0108] 在脱氨基之后，根据靶标RNA中的靶标腺苷的位置，可以使用不同方法确定靶标RNA和/或靶标RNA编码的蛋白质的修饰。例如，为了确定在靶标RNA中是否已将“A”编辑为“I”，可以使用本领域已知的RNA测序方法来检测RNA序列的修饰。当靶标腺苷位于mRNA的编码区中时，RNA编辑可以引起mRNA编码的氨基酸序列的改变。例如，可以将点突变引入mRNA，或者mRNA中的先天的或获得性的点突变被回复以产生野生型基因产物，因为“A”转化为了“I”。通过本领域已知的方法进行氨基酸测序可用于发现编码蛋白质中氨基酸残基的任何变化。终止密码子的修饰可以通过评估功能性的、伸长的、截短的、全长的和/或野生型的蛋白质的存在来确定。例如，当靶标腺苷位于UGA，UAG或UAA终止密码子

中时，靶标A (UGA或UAG)或多个A (UAA)的修饰可创造出通读突变和/或延长的蛋白质，或者，由靶标RNA编码的截短蛋白质可以被回复以产生有功能的、全长的和/或野生型的蛋白质。靶标RNA的编辑还可以在靶标RNA中产生异常剪接位点和/或可变的剪接位点，从而导致延长的，截短的或错误折叠的蛋白质，或者靶标RNA中经编码的异常剪接或可变剪接位点经回复以产生有功能的、正确折叠的、全长的和/或野生型的蛋白质。在一些实施方案中，本申请考虑编辑先天的和获得性的基因变化，例如，错义突变，过早的终止密码子，异常剪接或由靶标RNA编码的可变剪接位点。使用已知方法评估由靶标RNA编码的蛋白质的功能，可以发现RNA编辑是否实现了期望的效果。因为腺苷(A)对肌苷(I)的脱氨基可以纠正编码蛋白质的突变RNA中的靶标位置处的突变A，所以对脱氨基为肌苷的鉴定可以提供功能蛋白质是否存在的评估，或者突变的腺苷引起的、与疾病或药物抗性相关的RNA是否已回复或部分回复的评估。类似地，因为腺苷(A)对肌苷(I)的脱氨基作用可能在所得蛋白质中引入点突变，所以对脱氨基为肌苷的鉴定可以找出疾病原因或疾病相关因素的功能性指征。

[0109] 在一些实施方案中，靶标RNA是调节RNA。在一些实施方案中，待编辑的靶标RNA是核糖体RNA，转移RNA，长非编码RNA或小RNA (例如，miRNA, pri-miRNA, pre-miRNA, piRNA, siRNA, snoRNA, snRNA, exRNA或scaRNA)。靶标腺苷的脱氨基作用包括例如核糖体RNA，转移RNA，长非编码RNA或小RNA (例如，miRNA)，包括了三维结构和/或功能损失或功能获得的变化。在一些实施方案中，靶标RNA中靶标A的脱氨基作用改变了靶标RNA的一种或多种下游分子 (例如，蛋白质，RNA和/或代谢物)的表达水平。所述下游分子的表达水平的变化可以是表达水平的增加或减少。

[0110] 术语“RNA编辑实体”是指可以引起核苷酸的化学修饰以将核苷酸改变为不同核苷酸的生物分子。在一些实施方案中，可以将RNA编辑实体募集到多核苷酸中的特定位点以引起所需位点处核酸序列的改变。RNA编辑实体的例子包括APOBEC蛋白 (例如，APOBEC1、APOBEC2、APOBEC3A、APOBEC3B、APOBEC3C、APOBEC3E、APOBEC3F、APOBEC3G、APOBEC3H或APOBEC4蛋白)或ADAR蛋白 (例如，ADAR1、ADAR2或ADAR3蛋白)。

- [0111] 在本申请的上下文中，“靶RNA”是指将脱氨酶招募RNA序列设计为与其具有完全互补性或基本互补性的RNA序列，并且靶标序列与dRNA之间杂交形成含有靶标腺苷的双链RNA(dsRNA)区域，其招募作用于RNA的腺苷脱氨酶(ADAR)，该酶使靶标腺苷脱氨基。在一些实施方案中，ADAR天然地存在于宿主细胞中，例如真核细胞(优选哺乳动物细胞，更优选人类细胞)。在一些实施方案中，将所述ADAR引入宿主细胞中。
- [0112] 如本文所用的术语“条形码序列”是指将唯一性核苷酸序列用于在反应中鉴定和/或跟踪多核苷酸的来源。条形码序列的大小和组成可以差异巨大。在特定实施方案中，条形码序列的长度范围可为4至36个核苷酸、或6至30个核苷酸或8至20个核苷酸。
- [0113] 术语“载体”是指能够转运与其连接的另一种核酸的核酸分子。所述载体包括但不限于单链，双链或部分双链的核酸分子；包含一个或多个游离末端的，没有游离末端的(例如环状)核酸分子；包含DNA，RNA或两者的核酸分子；以及本领域已知的其他多种多核苷酸。一种类型的载体是“质粒”，它是指环状的双链DNA环，其中可以例如通过标准分子克隆技术插入额外的DNA区段。某些载体能够在引入它们的宿主细胞中自主复制(例如，具有细菌复制起点的细菌载体和附加型哺乳动物载体)。其他载体(例如，非附加型哺乳动物载体)在引入宿主细胞后整合到宿主细胞的基因组中，由此与宿主基因组一起复制。此外，某些载体能够指导它们可操作地连接的编码核苷酸序列的转录或表达。此类载体在本文中称为“表达载体”。
- [0114] 编辑实体通常在性质上是蛋白质的，比如在后生动物包括哺乳动物中发现的ADAR酶。编辑实体还可以包括核酸和蛋白质或肽的复合物，比如核糖核蛋白。编辑酶可以仅包括核酸或由核酸组成，比如核酶。所有这些编辑实体都包括在本发明中，只要它们是由根据本发明的寡核苷酸构建体募集的。优选地，编辑实体是酶，更优选为腺苷脱氨酶或胞苷脱氨酶，还更优选为腺苷脱氨酶。当编辑实体是腺苷脱氨酶时，Y优选为胞苷或尿苷、最优选胞苷。最感兴趣的之一是人ADAR，hADAR1和hADAR2，包括其任何亚型，比如hADAR1p110和p150。

- [0115] 靶RNA可以是任何细胞的或病毒的RNA序列，但更通常是具有蛋白质编码功能的mRNA前体或mRNA。
- [0116] 材料和方法
- [0117] 1、人源ADAR1表达质粒构建
- [0118] 利用PCR将人源ADAR1的CDS(人源ADAR1 CDS序列如SEQ:ID NO: 2所示)扩增出来并加上同源臂，利用NheI和EcoRI双酶切把pcDNA3.1载体(序列如SEQ ID NO: 1所示)线性化并用同源重组的方法把插入片段连接到载体。
- [0119] 2、报告基因质粒构建
- [0120] 首先将荧光基因GFP全长及需要编辑的基因片段通过PCR扩增出来同时两端加上同源重组的同源序列，同时对pcDNA3.1空载进行NheI和HindIII双酶切并回收线性载体，通过多片段重组对载体及插入片段进行同源重组完成报告基因的载体构建。
- [0121] 3、gRNA序列池质粒文库构建
- [0122] 通过PCR把gRNA混合序列池扩增出来同时加上同源臂，把报告基因载体用BamHI及EcoRI进行双酶切线性化并用同源重组的方式把gRNA序列池插入到报告基因载体，转化涂板后把所有单克隆收集到一起进行质粒提取，最终得到gRNA序列池质粒文库。
- [0123] 4、细胞培养
- [0124] 4.1细胞传代实验步骤：
- [0125] (1)紫外照射超净工作台消毒30分钟。
- [0126] (2)从培养箱中取出待传代的细胞，先用PBS清洗三遍，要轻轻地加入轻轻地吸弃，防止细胞脱落，再加入可淹没细胞表面量的0.25%胰酶，待细胞变圆后加入含有血清的培养基终止消化。
- [0127] (3)把消化后的细胞悬液转移到15ml离心管中，室温1000rpm离心4分钟。
- [0128] (4)小心吸弃上清液，加入新鲜的培养基并把细胞吹打成单悬浮。
- [0129] (5)取20u1细胞悬液加入到96孔细胞培养板并再加入20u1台盼蓝染色液，吹打混匀后吸取20u1加入到细胞计数板并在显微镜下进行细胞计数。

[0130] (6)根据接种的细胞培养板大小计算特定细胞数所需的细胞悬液体积，并加入对应体积的新鲜细胞培养基，放入细胞培养箱培养。

[0131] 4.2细胞转染实验步骤：

[0132] (1)根据实验需求细胞量设定所需要的细胞培养板，根据细胞传代的方法在对应的培养板中加入一定的培养基及细胞量，在细胞生长24小时可达到70%汇合度时则可进行细胞转染实验。

[0133] (2)紫外照射超净工作台，消毒30分钟。

[0134] (3)根据转染试剂lipofectamine 3000提供的说明书配制转染试剂混合液，室温静置5分钟。

[0135] (4)取出细胞并把转染液小心加入到细胞中，轻轻混合均匀后放入培养箱。

[0136] 4.3 ADAR1过表达HEK293细胞株构建

[0137] 按照上述的细胞传代及转染步骤把ADAR1表达载体转染到HEK293细胞，48小时后加入筛选药物G418，每两天换一次液，待未转染组已无活细胞完成稳转ADAR1的HEK293细胞株筛选。

[0138] 4.4靶向测序文库构建及质检

[0139] 细胞总RNA可以用常规RNA提取试剂盒进行提取，确保RNA无明显降解，用常规逆转录试剂盒进行逆转录反应，在第一轮PCR扩增时需在扩增引物两端加上测序接头序列，且控制第一轮扩增的循环数低于20个循环，琼脂糖凝胶电泳对第一轮PCR产物进行验证并回收对应大小的文库片段，纯化后再进行第二轮PCR反应，第二轮引物带有测序接头序列及特异性的碱基条形码序列，第二轮PCR一般控制在10个循环以内。再次对PCR产物进行琼脂糖凝胶电泳确认文库片段大小并切胶回收纯化文库。最后对靶向测序文库进行二代测序。

[0140] 4.4易错PCR(Error prone PCR)扩增

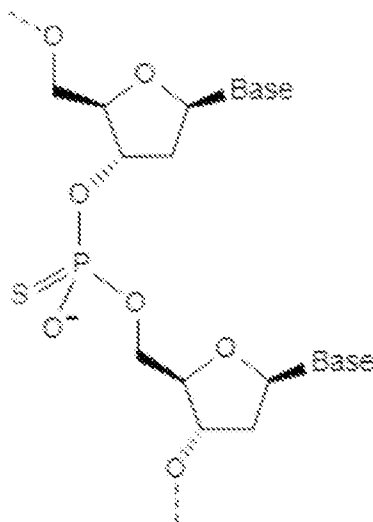
[0141] 利用Agilent公司的二代易错PCR试剂盒对特定区域进行扩增并随机引入突变，根据说明书进行样品配制，其中模板加入量为0.1ng，扩增循环数为30个循环。

[0142] 4.5靶向测序数据分析

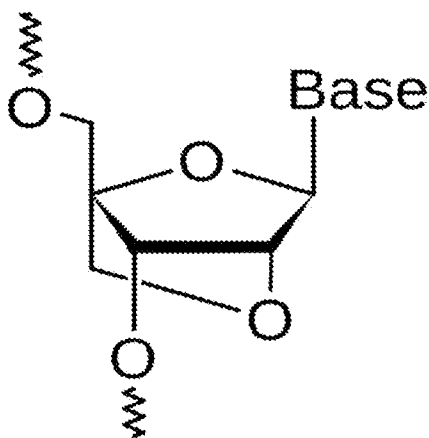
[0143] 首先对原始数据用cutadapt软件进行去接头处理，再把去接头后的原始读长比对到参考序列，根据连接子序列把报告基因的编辑与每条gRNA进行一一对应，再对报告基因的编辑水平进行计算，从而得出每条gRNA对靶位点的编辑水平。

[0144] 4.6化学修饰后的RNA合成

[0145] 修饰后的RNA合成主要由领域内的常规固相合成完成，其中碱基之间会有硫代磷酸化的骨架修饰和碱基上的2'位修饰。在一些实施方案中选择了硫代磷酸化修饰(下文用*代替)，在一些实施方案中选择了2'位LNA修饰(下文用L代替)。硫代磷酸化修饰和LNA修饰的结构式分别如下所示：



[0146] 硫代磷酸化修饰



[0147] LNA修饰

[0148] 实施例1在HEK293细胞中筛选GAPDH基因上的不同A碱基位点的gRNA序列

[0149] 本实施例以编辑人内源基因GAPDH(NM_002046.7)的多个A碱基位点的gRNA筛选方法为例,进一步说明本发明,具体地,所述方法包括如下步骤:

[0150] 1.1对人源GAPDH基因特定A碱基进行gRNA文库设计及接头序列设计

[0151] 以特定A碱基为中心上下游延伸75nt设计一条151nt的gRNA,在此gRNA上引入不同数量及位置的突变从而得到待筛选的文库,本实施例包含四个位点的A碱基,所在的三联基序分别为AAC,AAU,AAG和UAA,对该四个位点的多条gRNA文库进行筛选,本文举例列出了该四个位点中的AAC基序的文库序列信息,如SEQ ID NO:1~SEQ ID NO:540所示。

[0152] 序列合成可以通过设计接头序列加以区分,该接头序列在下游扩增步骤中不会造成不同组别间的交叉,四个文库对应的编辑位点信息及接头序列如下表1所示。

[0153] 表1编辑位点信息及接头序列

编辑位点	基因位置	编辑位点基序	3 端接头序列	5 端接头序列
编辑位点 1	816 号 A 碱基	AAC	ATCTTGACCCGTGGCGCCTG (SEQ ID NO:541)	GCAGTAGAACCAAGCAACCGG (SEQ ID NO:545)
编辑位点 2	1267 号 A 碱基	AAU	CGGTCGTGAGTGCAGACGTG (SEQ ID NO:542)	TCGAAGGTCGCTTAGACGGC (SEQ ID NO:546)
编辑位点 3	873 号 A 碱基	AAG	GACTCAGTTGCCACGCCTGT (SEQ ID NO:543)	CGGCGGGAGTTACAGCCTGT (SEQ ID NO:547)
编辑位点 4	1323 号 A 碱基	UAA	CATGGCTCGAGGGCAGCGA (SEQ ID NO:544)	TGTTCCGGGAATGTCCTGCTC (SEQ ID NO:548)

[0154] 1.2筛选所需载体的设计及构建

[0155] 如上述材料方法内容所述,在本实施例中,把编辑位点上下游100bp序列(共201个碱基)加入到GFP基因的下游重组到pcDNA3.1载体中,再在下游插入连接子1序列(序列如SEQ ID NO:549),在本实施例中使用的连接子1长度为2121bp,获得报告基因质粒载体,其中连接子1包含一段10bp的NNNYRNNNYR(Barcode序列,其中N表示A、C、G、T四种碱基中的一种,Y表示C或T碱基,R表示A或G碱基)简并碱基用于把每个gRNA的编辑效率区分开。得到插入了连接子1的载体后再把扩

增好的gRNA文库重组进报告基因质粒载体中，获得筛选载体如图2所示。所述载体表达出的转录本进行RNA编辑模式如图3所示。

[0156] 1.3接着将获得的筛选载体转入到过表达ADAR1的HEK293细胞中，48小时后提取RNA及靶向建库测序(建库引物如表2所示)并分析每条gRNA的靶位点编辑及脱靶编辑水平情况，如图4、图5所示。在本实施例中加入了长的连接子1序列，目的是为了模拟分子间RNA之间形成双链进行编辑，在建库时构建两种文库，一种是Reporter与Linker的区域文库构建每个Linker上的barcode序列与编辑水平的对应关系，所用的第一轮引物为cDNA-NGS-round1引物对，第二轮引物为NGS-round2引物对；另一种是Linker与gRNA库的区域，因为Linker与gRNA中间间隔超过2000bp无法直接建库测序，故先对Linker与gRNA的全长序列进行环化，环化后用plasmid-NGS-round1引物对进行第一轮扩增，再用NGS-round2引物对完成整个文库构建，此文库可建立Linker上的barcode与gRNA序列的对应关系，通过两个文库的barcode序列比对则可得出每条gRNA对应的编辑效率。

[0157] 结果显示，通过本发明的实验及数据分析方法可以筛选出每条gRNA对应的编辑效率，且生物学重复比较一致。编辑位点1为AAC基序(图4a)，黑色的点表示从编辑位点往两端延伸共151nt的gRNA(编辑位点对位碱基为C碱基)在两次重复实验中的编辑水平，其余黄色的点为引入突变后的不同gRNA在两次重复实验中的编辑水平，针对AAC基序的gRNA在50%以内的编辑效率有均匀的gRNA分布，在50%到75%的编辑效率区间仅有少些gRNA能够达到。通过热图(图5a)分析可知在本实施例设计的针对AAC基序的gRNA在多个位点有明显的脱靶编辑但靶位点的编辑水平整体高于脱靶位点。

[0158] 编辑位点2的AAU基序(图4b)gRNA整体表现出低于40%的编辑效果，且从编辑位点往上下游延伸的除了编辑位点处是AC的错配，其他都是互补配对的gRNA编辑效率也低于30%。

[0159] AAG基序(图4c)的gRNA编辑效率与AAC基序类似，与其他三个编辑位点不同的是UAA基序(图4d)的gRNA整体编辑水平位于25%到75%之间，但UAA基序的gRNA邻近的A碱基有比较严重的脱靶编辑，只有AAG基序筛选的gRNA整体脱靶编辑较低

(图5)。本实施例设计的gRNA长度为151nt左右，此外，不同的位点A对应的gRNA具有不同的编辑能力，表明通过本发明可筛选出编辑水平不同的gRNA序列。

[0160] 表2建库测序所用引物序列

编号	引物名称	引物序列
SEQ ID NO:550	cDNA-NGS-round1-F	ACGACGCTCTTCCGATCTAGCTAGCGTTTAAACTTAAGCTT
SEQ ID NO:551	cDNA-NGS-round1-R	AGACGTGTGCTCTTCCGATCTCCCTCACTCTGCGGCTTGACGTCGAATATCAC
SEQ ID NO:552	NGS-round2-F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT
SEQ ID NO:553	NGS-round2-R	CAAGCAGAAGACGGCATACGAGATATCACGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
SEQ ID NO:554	plasmid-NGS-round1-F	ACACGACGCTCTTCCGATCTTGATACAGGACAGCAACGGCGAGA
SEQ ID NO:555	plasmid-NGS-round1-R	AGACGTGTGCTCTTCCGATCTCCCTCACTCTGCGGCTTGACGTCGAATATCAC

[0161] 实施例2在HEK293细胞中筛选GAPDH基因特定位点基序为UAG的gRNA序列

[0162] 本实施例旨在验证在人源GAPDH基因上可以筛选出针对特定位点进行编辑的短gRNA及短连接子。选取GAPDH基因上787位点A碱基，所在的基序为TAG基序，针对该位点设计了1073条包含不同长度及不同位点突变的gRNA(长度范围28nt-89nt)。在本实施例选择用短连接子验证筛选体系，其中连接子2的序列为NNNNNNNNTGGGTTGAGGGTAGTGAG(序列号为SEQ ID NO:556；N为A, T, C, G碱基中的任一种，8个N为连接子的barcode序列)。

[0163] 具体的实施方案如下，首先直接合成gRNA文库，把连接子2与gRNA序列直接相连合成完成带有barcode序列的gRNA库；其次构建reporter载体，以编辑位点为中心上下游延伸共201bp的序列与GFP序列重组插入到pcDNA3.1载体；第三步把步骤一的gRNA文库插入到reporter载体完成筛选载体的构建；第四步把筛选载体转染到ADAR1蛋白过表达的HEK293细胞48小时后提取RNA及靶向建库测序并分析每条gRNA的靶位点编辑及脱靶编辑水平情况，结果如图6所示。结果表明在本实施例中在短连接子的情况下不同长度的gRNA混合的文库也可以通过此方法筛选出不同编辑效率的gRNA。

[0164] 基于本发明的方法建立的高通量筛选是向细胞中转染混合的gRNA文库，可能会存在一定程度的gRNA之间的干扰情况，故本实施例在高通量筛选的基础上进

行单个gRNA的验证，本实施例挑选了高通量筛选出编辑水平从3.93%到21.66%的gRNA序列，序列信息如表3的GAPDH-TAG-S1至GAPDH-TAG-S12，化学修饰后的gRNA序列对应为GAPDH-TAG-M1至GAPDH-TAG-M12。针对高通量筛选出来的序列则用化学修饰后的单个RNA分别转染进入细胞验证编辑水平情况(如图8及表3所示)，修饰后的gRNA稳定性高于质粒表达出的gRNA，故编辑水平均高于高通量筛选的编辑水平，化学修饰的gRNA修饰策略为碱基之间用硫代磷酸修饰，两端各加3个LNA修饰。分析高通量筛选出的编辑水平与单个修饰gRNA的编辑水平可知二者呈现较强的相关性，相关系数达到0.83(如图7所示)，证实了本发明的方法的准确性。其中表3中的GAPDH-TAG-S1至GAPDH-TAG-S13的编辑水平为高通量的gRNA混合体系测定出来的gRNA编辑水平；GAPDH-TAG-M1至GAPDH-TAG-M13的编辑水平为单个gRNA体系测定出来的编辑水平。

[0165] 表3高通量筛选验证序列信息表

序列编号	序列信息	编辑水平 (%)
GAPDH-TAG-S1 (SEQ ID NO:557)	CAUGGGUGCCCUUGGAAAGUGGUGCCAAGAAGUUGGGGGUGCAGGAAGUAU	21.66
GAPDH-TAG-M1 (SEQ ID NO:558)	LC*LA*LT*G*G*U*G*C*C*U*U*G*G*A*A*A*G*U*G*U*G*C*C*A*A*G*A*A*G*U*U*G*G*G*G*U*G*C*A*G*G*A*A*G*LT*LA*LT	50
GAPDH-TAG-S2 (SEQ ID NO:559)	CGUGGAUGACAUUGGCGGUGUAGCCAAGCAUAAAAUUGGUGGGUGCAGGAGUAU	17.61
GAPDH-TAG-M2 (SEQ ID NO:560)	LC*LG*LT*G*G*A*U*G*A*C*A*U*U*G*G*C*C*G*G*U*G*U*A*G*C*C*A*A*G*C*A*U*A*A*A*A*U*U*G*G*U*G*G*G*U*G*C*A*G*G*A*G*LT*LA*LT	40
GAPDH-TAG-S3	AUGACCUUGGCCAUGUGUGCCAAGCGUGAGUGGUGCAGGA	10.64

[0166] 实施例3在HEK293细胞中筛选ATP7B基因特定位点基序为UAG的gRNA序列

[0167] 本实施例旨在验证在人源ATP7B(NM_000053.4)基因上可以筛选出针对特定位点进行编辑的短gRNA。ATP7B基因突变是引起肝豆状核变性疾病的主要原因,在本实施例选取ATP7B基因上1532位点A碱基,在该位点上游有一个碱基的突变C碱基突变为T碱基,故突变后编辑位点所在的基序为UAG基序,针对该位点设计了1011条不同长度及不同位点突变的gRNA,本实施例选取包含编辑位点及其上下游在

内的长度为31nt和41nt的gRNA，针对这两个长度引入突变或者缺失或者增加了凸起等会使得实际长度有一点的浮动，浮动范围为27nt-59nt。在本实施例选择用连接子2验证筛选体系。

[0168] 具体的实施方案如下，首先直接合成长度为31nt和41nt的gRNA文库，把连接子2与gRNA序列直接相连合成完成带有barcode序列的gRNA库；其次构建reporter载体，以编辑位点为中心上下游延伸共201bp的序列与GFP序列重组插入到pcDNA3.1载体；第三步把步骤一的gRNA文库插入到reporter载体完成筛选载体的构建；第四步把筛选载体转染到ADAR1蛋白过表达的HEK293细胞48小时后提取RNA及靶向建库测序并分析每条gRNA的靶位点编辑及脱靶编辑水平情况，结果如图9和图10所示。

[0169] 结果表明，通过本实施例可以从1011条不同长度的gRNA文库中筛选出高编辑效率的gRNA序列，整体上41nt的gRNA编辑能力高于31nt的gRNA(图9)，其中黑色的点表示除了编辑位点处有一个碱基的错配外，其他碱基都是互补配对的gRNA。

[0170] 把所有不同长度的gRNA排列在一起比较靶位点的编辑效率及脱靶效率(图10)可知，在ATP7B 1532位点的A碱基设计的gRNA有比较弱的脱靶编辑。整体比较来看基于41nt的靶序列设计的gRNA大部分序列的编辑水平低于20%，基于31nt靶序列设计的gRNA全部低于20%的编辑。

[0171] 本实施例从高通量筛选的两组结果中各挑选了20条编辑效率高且测序测的读长多的序列进行下一轮的error prone验证引入突变后的编辑效果，从41nt组的挑选出的序列如SEQ ID NO:583~SEQ ID NO:602所示，31nt组的挑选序列如SEQ ID NO:603~SEQ ID NO:622所示。在所有的挑选序列两端分别加上error prone PCR的引物序列5'端为TCGAAGGTCGCTTAGACGGC(SEQ ID NO:581)，3'端为CGGTCGTGAGTGCAGACGTG(SEQ ID NO:582)直接合成单链序列，在error prone PCR扩增时分两组进行扩增，41nt组则把所有序列等质量混合后作为模板扩增，31nt组同样处理，扩增后的产物重新插入到reporter载体中按照第一轮筛选进行第二轮筛选，结果如图11所示。

[0172] 结果表明无论是41nt组还是31nt组在经过第二轮的筛选后都达到了显著的提升，31nt组从第一轮的全部低于20%的编辑效率提升到20%至80%的编辑效

果，主要分布在40%的编辑效率以内，40%到80%的编辑占比较少，而41nt组20%到50%编辑效率的占比也是少的，在经过第二轮的筛选后明显可以看到40%到90%的编辑占比比较高，通过以上数据也证实了本发明的两轮筛选可以筛选出比较高的编辑效果的gRNA，尤其是第二轮error prone引入突变的方法。

[0173] 实施例4在HEK293细胞中筛选FGFR基因特定位点基序为CAG的gRNA序列

[0174] 本实施例旨在验证在人源FGFR(NM_015850.4)基因上可以筛选出针对特定位点进行编辑的短gRNA。FGFR基因突变是引起软骨发育不全、致死性骨发育不良及矮小症等疾病的病因，在本实施例选取FGFR基因上G358R位点产生了G到A的突变，故突变后编辑位点所在的基序为CAG基序，针对该位点设计了989条不同长度及不同位点突变的gRNA，本实施例选取包含编辑位点及其上下游在内的长度为31nt和41nt的gRNA，针对这两个长度引入突变或者缺失或者增加了凸起等会使得实际长度有一点的浮动，浮动范围为23nt-62nt。

[0175] 在本实施例选择用连接子2验证筛选体系。具体的实施方案如下，首先直接合成31nt和41nt的gRNA文库，把连接子2与gRNA序列直接相连合成完成带有barcode序列的gRNA库；其次构建reporter载体，以编辑位点为中心上下游延伸共201bp的序列与GFP序列重组插入到pcDNA3.1载体；第三步把步骤一的gRNA文库插入到reporter载体完成筛选载体的构建；第四步把筛选载体转染到ADAR1蛋白过表达的HEK293细胞48小时后提取RNA及靶向建库测序并分析每条gRNA的靶位点编辑及脱靶编辑水平情况，结果如图12和图13所示。

[0176] 结果表明，通过本实施例可以从989条不同长度的gRNA文库中筛选出高效编辑gRNA序列，整体上41nt的gRNA编辑能力高于31nt的gRNA(图12)，其中黑色的点表示除了编辑位点处有一个碱基的错配外，其他碱基都是互补配对的gRNA，把所有不同长度的gRNA排列在一起比较靶位点的编辑效率及脱靶效率(图13)可知，在所有的gRNA中在编辑位点上游有两个位点的明显脱靶，但在编辑位点下游完全没有脱靶编辑。整体比较来看基于41nt的靶序列设计的gRNA大部分序列的编辑水平低于20%，基于31nt靶序列设计的gRNA几乎都低于10%的编辑。

[0177] 鉴于本实施例31nt组整体编辑水平偏低，故仅从41nt的gRNA组高通量筛选结果挑选了20条编辑水平高且测序测的读长多的序列进行下一轮的error prone

验证引入突变后的编辑效果，41nt组的挑选序列如SEQ ID NO:623～SEQ ID NO:642所示。在所有的挑选序列两端分别加上error prone PCR的引物序列5端为TCGAAGGTCGCTTAGACGGC (SEQ ID NO:581)，3端为CGGTCGTGAGTGCAGACGTG (SEQ ID NO:582)直接合成单链序列，在error prone PCR扩增时把所有 序列等质量混合后作为模板扩增，扩增后的产物重新插入到reporter载体中按照第一轮筛选进行第二轮筛选，结果如图14所示。

[0178] 结果表明，在经过error prone引入随机突变的第二轮的筛选后都达到了显著的提升，从第一轮的20%以内提升到80%的编辑效果，通过以上数据再次证实了本发明的两轮筛选可以筛选出比较高的编辑效果的gRNA，尤其是第二轮error prone引入突变的方法。

权利要求书

- [权利要求 1] 一种编辑RNA的向导RNA的筛选方法，其特征在于，包括步骤：(1)待筛选的每条向导RNA连有连接子，并分别与靶RNA连接在能够形成一个转录本的核酸链上，形成多个所述的核酸链；所述的核酸链含有条形码序列；(2)将所述的多个核酸链转入表达ADAR蛋白的细胞中；(3)提取和测定所述细胞中的RNA，确定靶RNA的被编辑水平，通过条形码序列确定对应的向导RNA的编辑水平。
- [权利要求 2] 如权利要求1所述的筛选方法，其特征在于，所述的连接子位于靶RNA和向导RNA之间；
- 优选地，所述的条形码序列包含于所述连接子中；
- 所述核酸链从上游到下游依次连接有报告基因核酸片段、靶RNA核酸片段、连接子核酸片段、向导RNA核酸片段；
- 优选地，所述向导RNA核酸片段的下游还连接有polyA尾核酸片段；
- 优选地，所述报告基因选自yfp基因、gus基因、rfp基因、gfp基因、cfp基因、bfp基因、或卡那霉素抗性基因；
- 优选地，所述连接子的长度为26nt~2121nt。
- 优选地，所述连接子包括SEQ ID NO:549或SEQ ID NO:557；
- 优选地，步骤(3)包括建立靶向测序文库进行二代测序；
- 优选地，所述核酸链接触ADAR蛋白通过将所述核酸链转入表达ADAR蛋白的细胞中实现；
- 优选地，所述ADAR蛋白包括ADAR1或ADAR2；
- 优选地，所述向导RNA能够在所述细胞中招募ADAR蛋白对靶RNA进行编辑；
- 优选地，所述的细胞过表达ADAR蛋白；
- 优选地，所述编辑为A-to-I编辑；
- 优选地，所述向导RNA为以A碱基为中心上下游延伸得到的向导RNA；

优选地，所述向导RNA为以A碱基为中心上下游延伸15nt～75nt得到的30nt～151nt长度的向导RNA；

优选地，所述向导RNA为完全互补配对或不完全互补配对的序列；

优选地，所述向导RNA的长度为20nt～200nt；

优选地，所述筛选方法包括把编辑位点上下游序列加入报告基因的下流并插入到载体中获得报告基因载体，再把报告基因载体转入细胞中；

优选地，所述编辑位点上下游加入到报告基因的序列长度为20～200nt；

优选地，所述筛选方法为高通量筛选方法；

优选地，所述靶RNA序列长度大于向导RNA序列；

优选地，所述载体包括质粒或病毒载体；

优选地，所述载体是用于在高等真核细胞或原核细胞中表达的质粒或病毒载体；

优选地，所述载体骨架为pcDNA3.1载体；

优选地，所述目的向导RNA为靶向位点进行高效编辑的向导RNA；

优选地，所述方法还包括：从所述步骤(3)中获得所述向导RNA的编辑水平后，挑选编辑水平高的向导RNA合成序列相同的单链DNA后设计易错PCR引物对所述向导RNA的靶向区域进行随机引入突变，再重复所述步骤(1)～(3)；

优选地，所述易错PCR引物的序列如SEQ ID NO:581和SEQ ID NO:582所示。

[权利要求 3]

一种编辑RNA的向导RNA的筛选方法，其特征在于，包括：

(a) 向导RNA文库的设计；

(b) 把需要编辑的基因片段与报告基因插入到载体中获得报告基因载体；

(c) 把所述向导RNA文库插入进所述报告基因载体获得待筛选质粒文库；

(d)将所述待筛选质粒文库转入到细胞，转染后，提取所述细胞中的RNA，测序；

(e)使用生物信息学分析所述向导RNA的编辑水平；

其中(a)和(b)的顺序可以调换；

优选地，步骤(b)包括：把编辑位点上下游序列加入报告基因的下游插入到载体中获得报告基因载体；

优选地，所述编辑位点上下游加入到报告基因的序列长度为20～200nt；

优选地，步骤(b)包括：将荧光基因全长以及需要编辑的基因片段通过PCR扩增出来同时两端加上同源重组的同源序列或引入酶切位点，同时对空载体进行NheI和HindIII双酶切并回收线性载体，将多片段插入到线性载体中获得报告基因的载体；

优选地，步骤(c)包括：扩增步骤(a)中的所述向导RNA同时加上同源臂或酶切位点获得向导RNA序列文库，步骤(b)中的报告基因载体用BamHI及EcoRI进行双酶切线性化；然后把所述的向导RNA序列文库插入到线性化后的报告基因载体中，转化后获得待筛选质粒文库；

优选地，所述筛选方法还包括：

(f)从所述向导RNA中挑选出编辑效率高的向导RNA直接合成序列相同的单链DNA后，设计易错PCR引物对所述向导RNA靶向区域进行随机引入突变，再重复步骤(b)～(e)。

[权利要求 4]

权利要求1-3任一所述的筛选方法获得的向导RNA。

[权利要求 5]

一种向导RNA，其特征在于，所述向导RNA的序列如SEQ ID NO:557、SEQ ID NO:559、SEQ ID NO:561、SEQ ID NO:563、SEQ ID NO:565、SEQ ID NO:567、SEQ ID NO:569、SEQ ID NO:571、SEQ ID NO:573、SEQ ID NO:575、SEQ ID NO:577、SEQ ID NO:579、SEQ ID NO:583～SEQ ID NO:602、SEQ ID NO:603～SEQ ID NO:622、或SEQ ID NO:623～SEQ ID NO:642任一所示。

[权利要求 6]

如权利要求1-3任一所述的筛选方法或权利要求4-5任一所述的向导RNA在制备治疗疾病的药物中的用途；

优选地，所述药物用于在真核细胞中定点编辑靶基因中的核苷酸从而达到治疗疾病的目的；

优选地，所述疾病为包括GAPDH基因、ATP7B基因或FGFR基因突变引起的疾病；

优选地，所述GAPDH基因特定位点基序为AAC、AAU、AAG、UAA或UAG；

优选地，所述ATP7B基因特定位点基序为UAG；

优选地，所述FGFR基因特定位点基序为CAG；

优选地，所述疾病包括肝豆状核变性、软骨发育不全、致死性骨发育不良或矮小症；

优选地，所述用途是通过天然存在于所述细胞中并且能够进行所述核苷酸的编辑的RNA编辑实体的作用。

[权利要求 7]

一种工程化RNA，其特征在于，包括：从上游到下游依次连接有报告基因核酸片段、靶RNA核酸片段、连接子核酸片段、向导RNA核酸片段；

优选地，所述向导RNA核酸片段的下游还连接有polyA尾核酸片段；

优选地，所述报告基因选自yfp基因、gus基因、rfp基因、gfp基因、cfp基因、bfp基因、或卡那霉素抗性基因；

优选地，所述连接子的长度为26nt至2121nt；

优选地，所述连接子包括SEQ ID NO:549或SEQ ID NO:557；

优选地，所述向导RNA为以A碱基为中心上下游延伸得到的向导RNA；

优选地，所述向导RNA为以A碱基为中心上下游延伸15nt~75nt得到的30nt~151nt长度的向导RNA。

[权利要求 8]

一种工程化RNA结构文库，其特征在于，包含如权利要求7所述的多个不同的工程化RNA；

优选地，所述不同的工程化RNA是通过将靶RNA通过不同的连接子与不同的向导RNA连接获得的；

优选地，所述工程化RNA结构文库中，所述文库中，所述的工程化RNA所含有的靶RNA核酸片段是相同的，向导RNA核酸片段序列是不同的。

[权利要求 9]

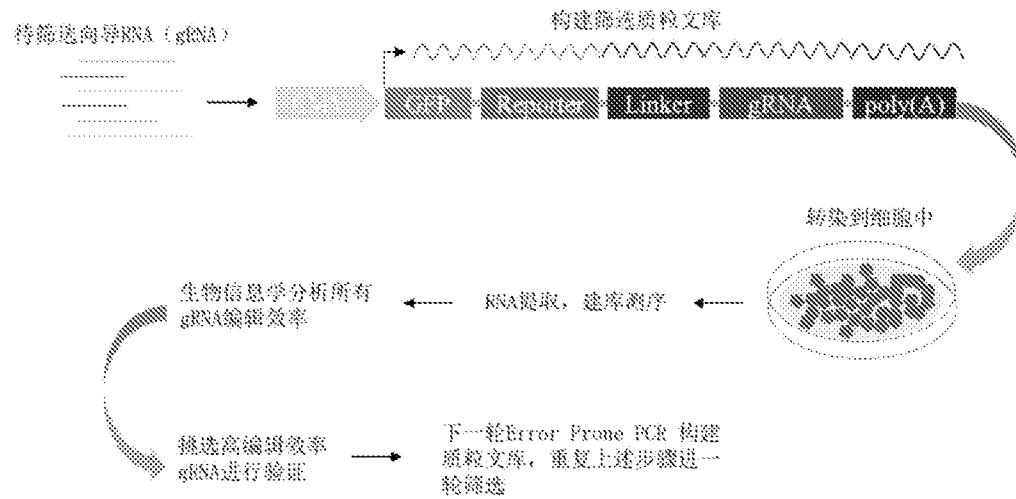
一种载体，其特征在于，包含权利要求7所述的工程化RNA或权利要求8所述的工程化RNA结构文库；

优选地，所述载体是通过把靶RNA序列加入到报告基因序列的下游插入到载体骨架中，获得报告基因载体，再把扩增好的向导RNA文库插入报告基因载体中获得的。

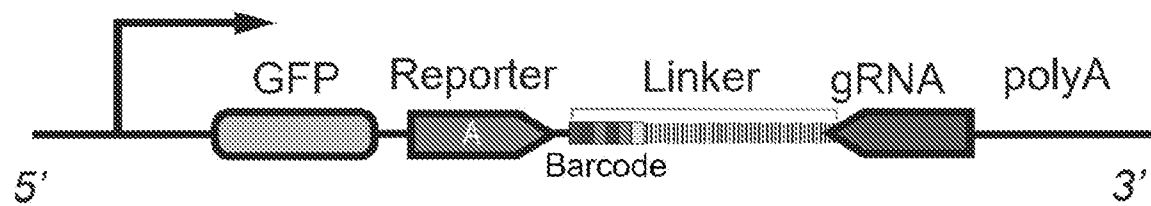
[权利要求 10]

一种细胞，其特征在于，包含权利要求7所述的工程化RNA或权利要求8所述的工程化RNA结构文库或权利要求9所述的载体。

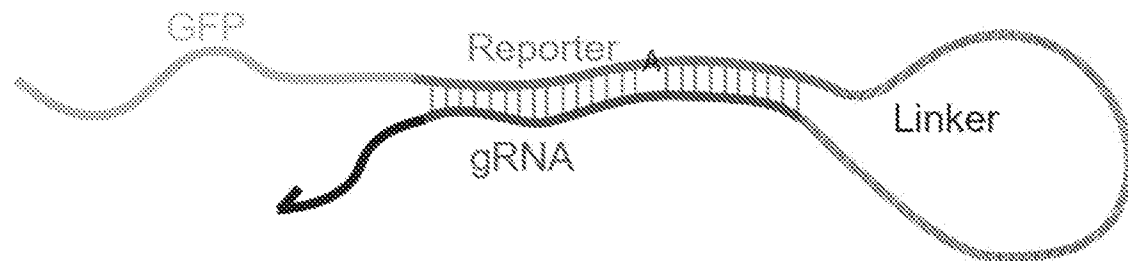
[图 1]



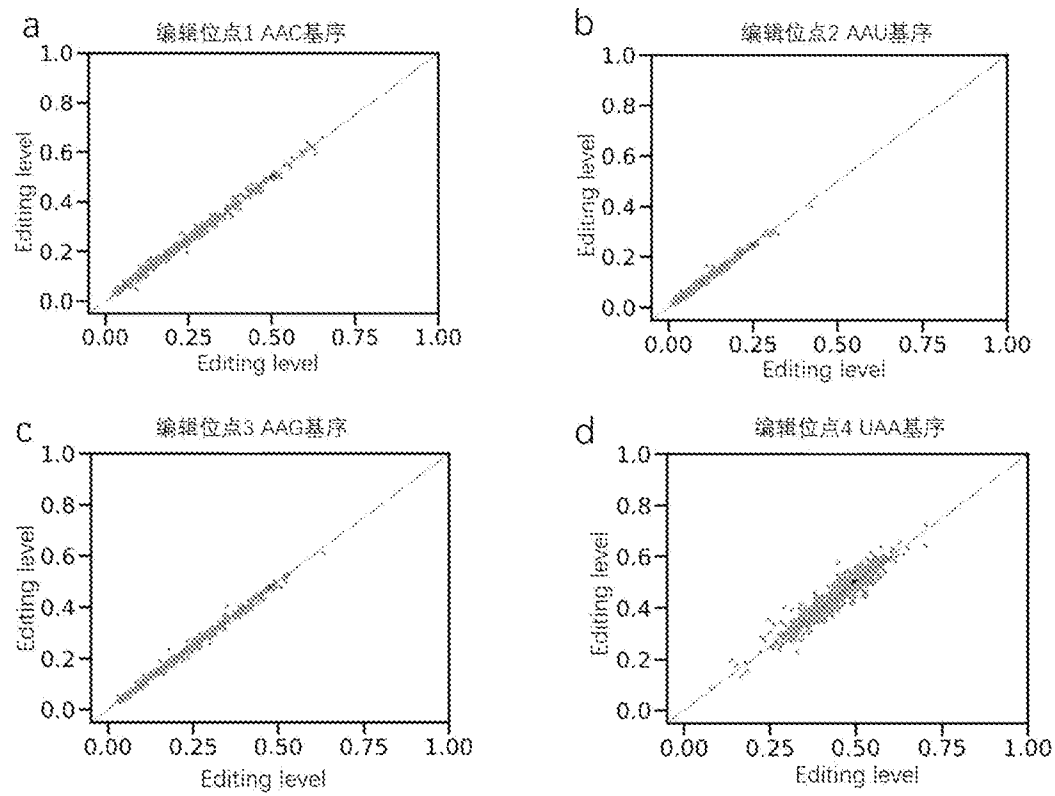
[图 2]



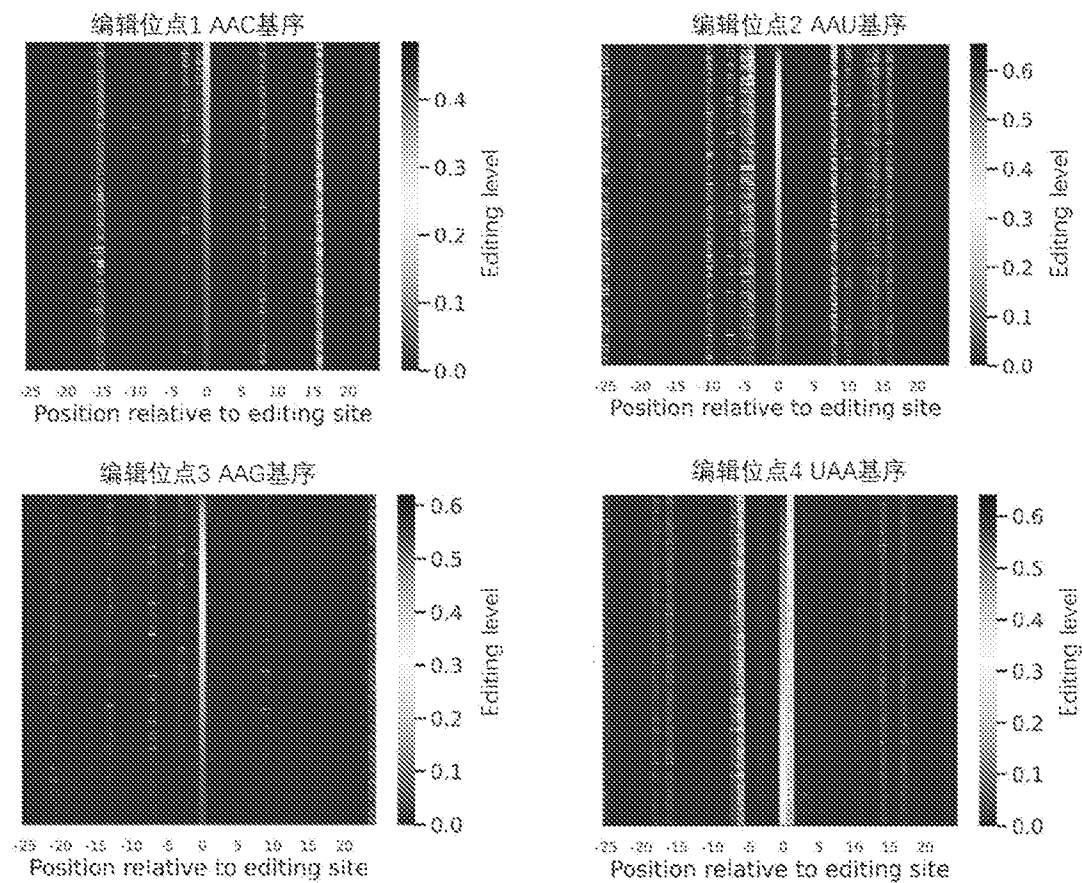
[图 3]

细则 26,
08. 06. 2023细则 26,
08. 06. 2023细则 26,
08. 06. 2023

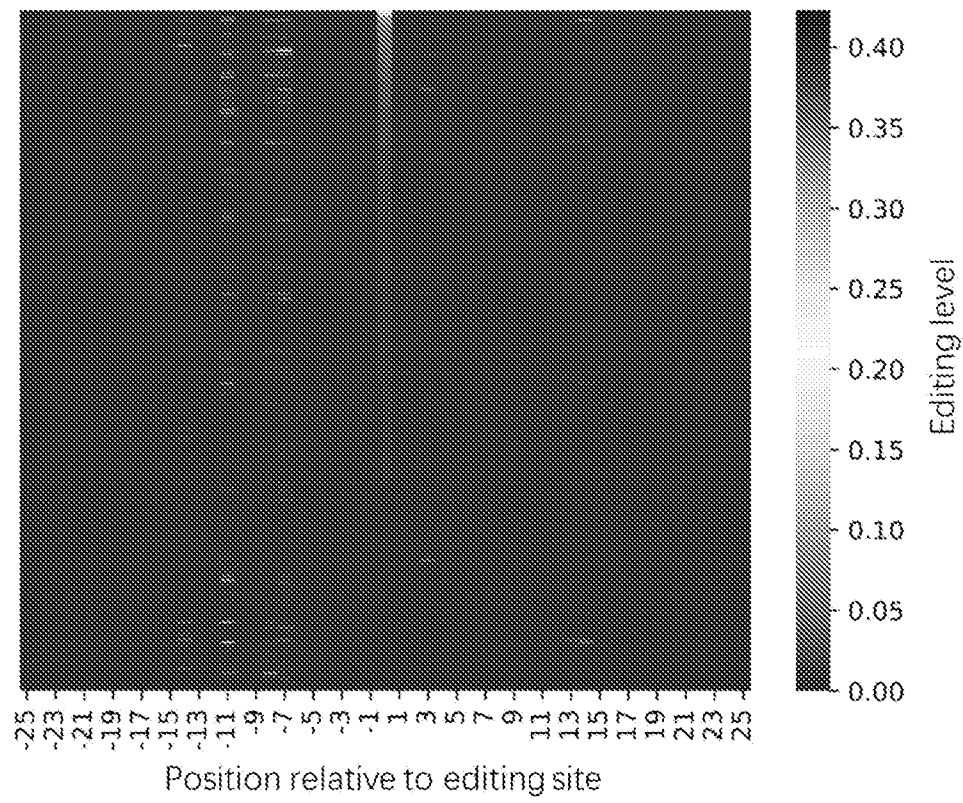
[图 4]



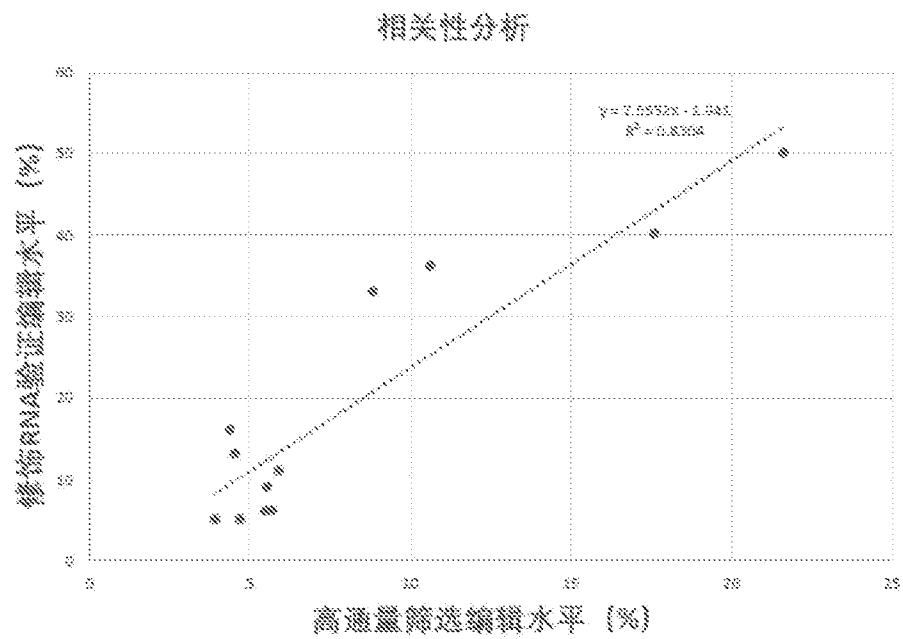
[图 5]

细则 26,
08. 06. 2023细则 26,
08. 06. 2023

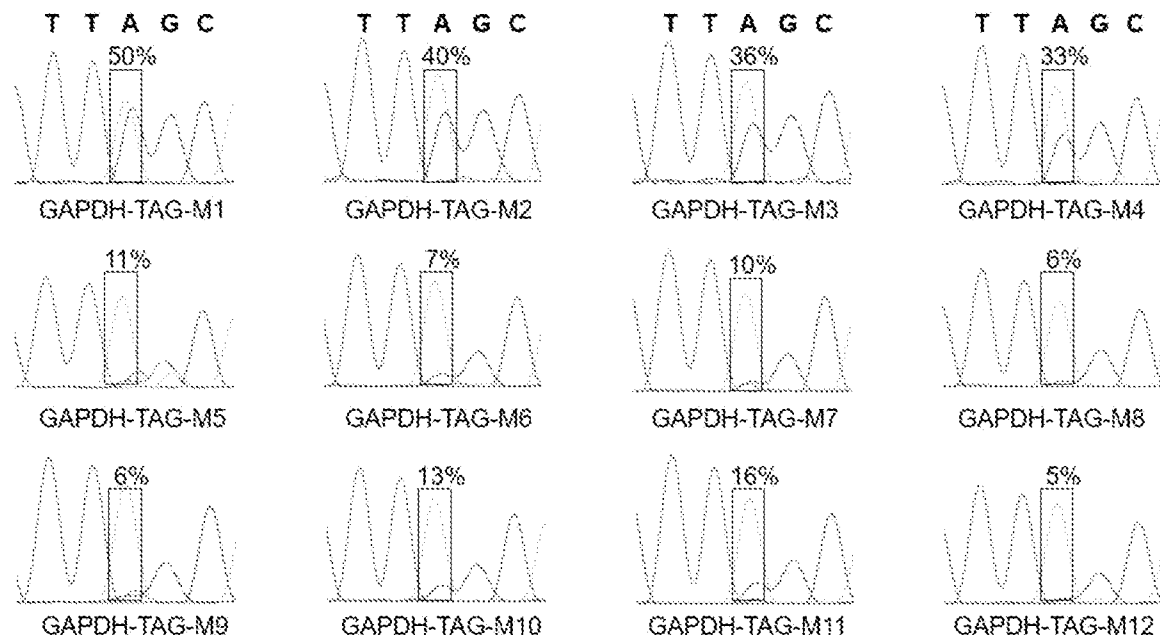
[图 6]



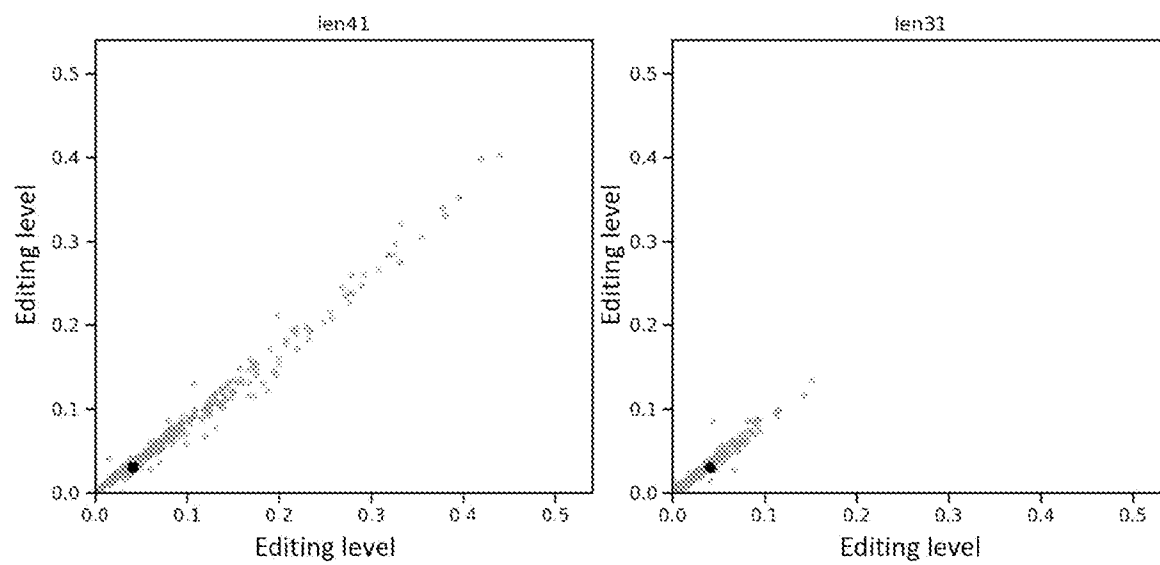
[图 7]



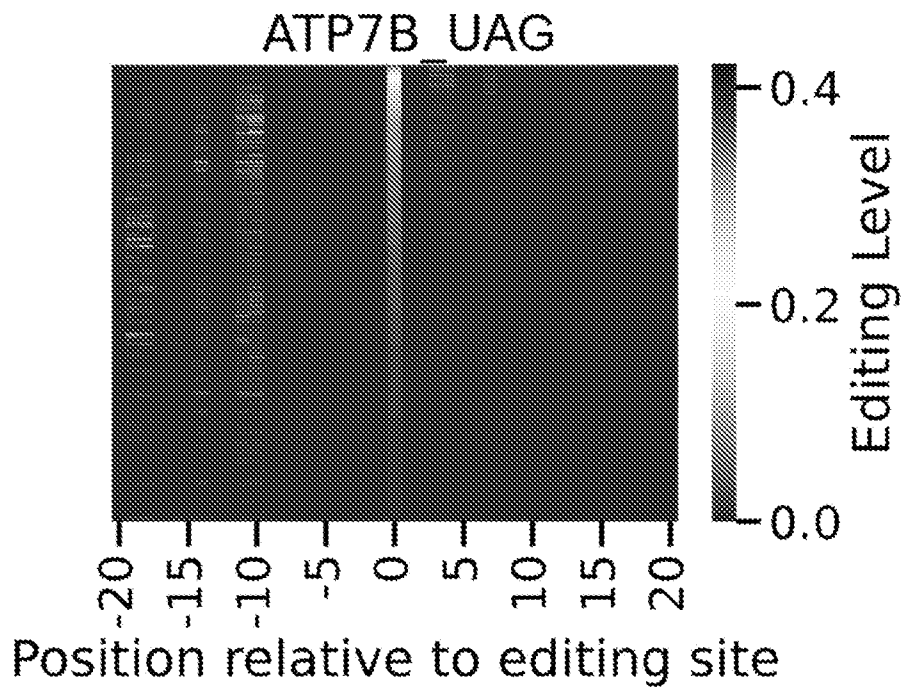
[图 8]



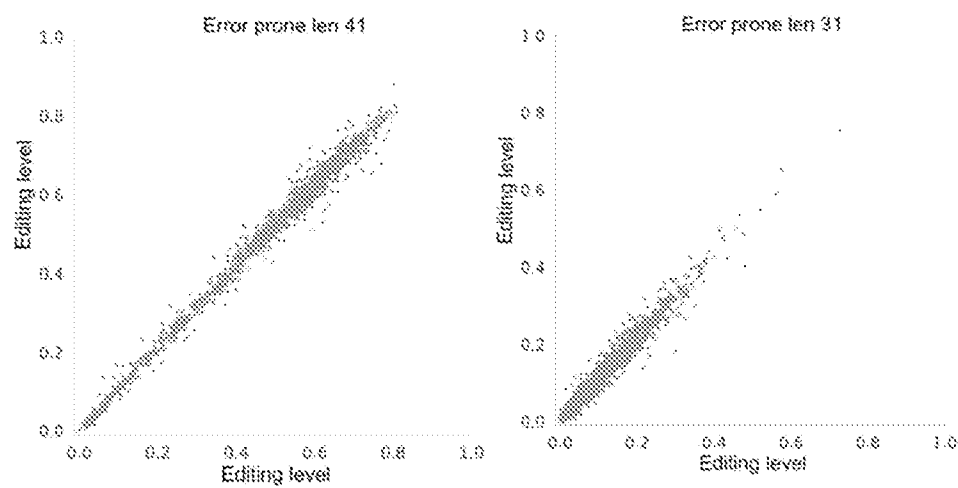
[图 9]



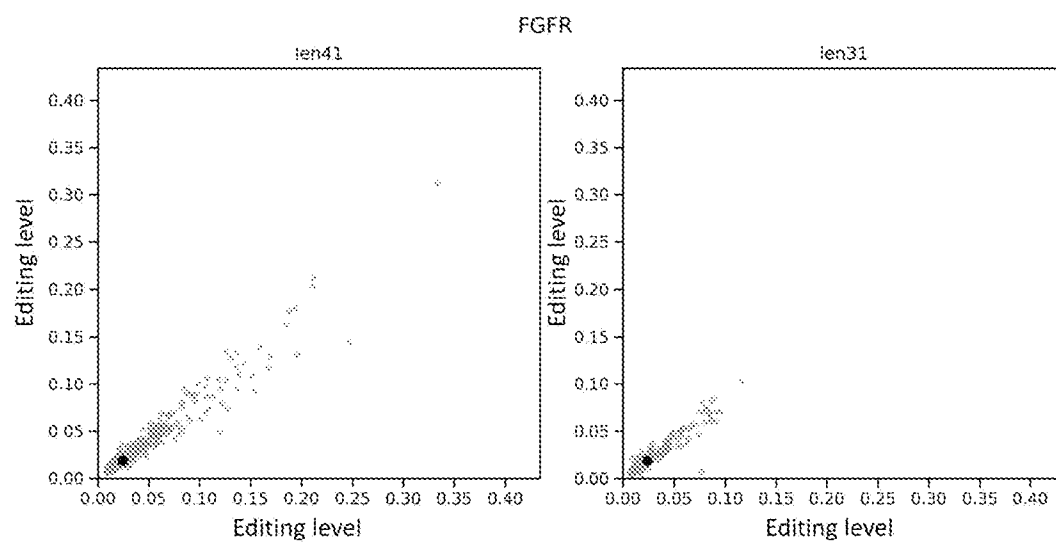
[图 10]



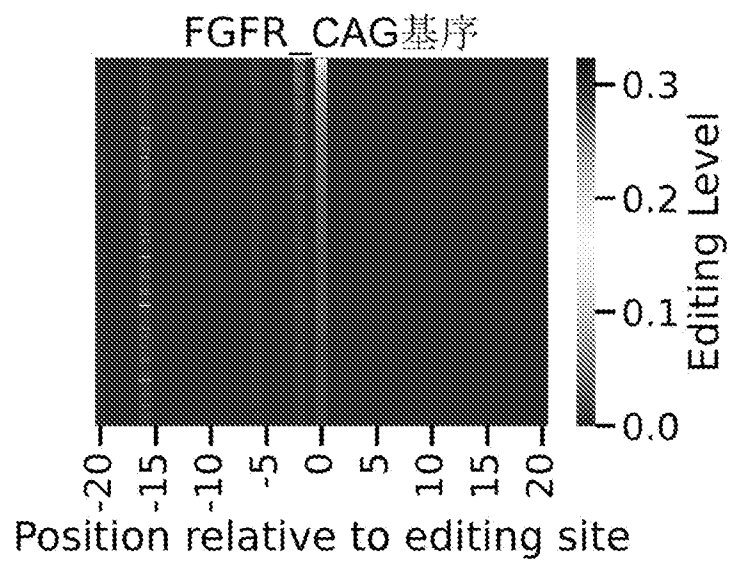
[图 11]



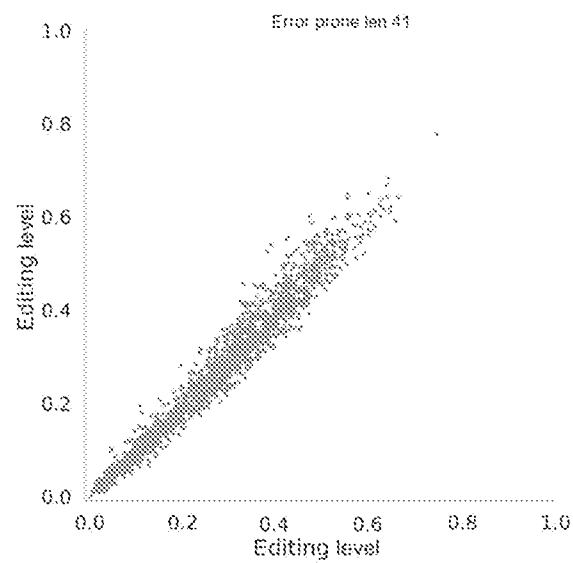
[图 12]

细则 26,
08. 06. 2023细则 26,
08. 06. 2023细则 26,
08. 06. 2023

[图 13]



[图 14]

细则 26,
08. 06. 2023细则 26,
08. 06. 2023

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2023/085605

A. CLASSIFICATION OF SUBJECT MATTER C12N15/113(2010.01)i; C12N15/09(2006.01)i; C12N15/11(2006.01)i; C12N15/10(2006.01)i; C12N15/63(2006.01)i; A61K48/00(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) IPC:C12N A61K48 C12Q G16B 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) CNABS; CNTXT; CJFD; WPABS; WPABSC; DWPI; ENTXT; ENTXTC; CNKI; 万方, WANFANG; ISI Web of Science; NCBI; EMBL; STN; 中国专利生物序列检索数据库, China Patents Biological Sequence Search Database: 时夕生物, 杨文兵, 向导RNA, 指导RNA, 连接子, 接头, 环序列, 茎环, 条形码, 分子标识符, 筛选, 靶序列, 靶RNA, 靶标, RNA 编辑, 编辑RNA, 测序, gRNA, guide RNA, linker, barcode, UMI, screen, RNA editing, sequencing, ADAR, 序列557, 559, 561, 563, 565, 567, 569, 571, 573, 575, 577, 579, 583-602, 603-622, 623-642																	
C. DOCUMENTS CONSIDERED TO BE RELEVANT <table border="1"> <thead> <tr> <th>Category*</th> <th>Citation of document, with indication, where appropriate, of the relevant passages</th> <th>Relevant to claim No.</th> </tr> </thead> <tbody> <tr> <td>X</td> <td>WO 2022087272 A1 (LELAND STANFORD JUNIOR UNIVERSITY) 28 April 2022 (2022-04-28) claims 1-8, 24, 25 and 42-46, and description, paragraphs 0037 and 00165</td> <td>4</td> </tr> <tr> <td>Y</td> <td>WO 2022087272 A1 (LELAND STANFORD JUNIOR UNIVERSITY) 28 April 2022 (2022-04-28) claims 1-8, 24, 25 and 42-46, and description, paragraphs 0037 and 00165</td> <td>1-3, 5-10</td> </tr> <tr> <td>Y</td> <td>CN 111613272 A (WESTLAKE UNIVERSITY) 01 September 2020 (2020-09-01) description, paragraph 0006</td> <td>1-3, 5-10</td> </tr> <tr> <td>A</td> <td>WO 2022147573 A1 (UNIVERSITY OF CALIFORNIA) 07 July 2022 (2022-07-07) claims 1-137</td> <td>1-10</td> </tr> </tbody> </table>			Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.	X	WO 2022087272 A1 (LELAND STANFORD JUNIOR UNIVERSITY) 28 April 2022 (2022-04-28) claims 1-8, 24, 25 and 42-46, and description, paragraphs 0037 and 00165	4	Y	WO 2022087272 A1 (LELAND STANFORD JUNIOR UNIVERSITY) 28 April 2022 (2022-04-28) claims 1-8, 24, 25 and 42-46, and description, paragraphs 0037 and 00165	1-3, 5-10	Y	CN 111613272 A (WESTLAKE UNIVERSITY) 01 September 2020 (2020-09-01) description, paragraph 0006	1-3, 5-10	A	WO 2022147573 A1 (UNIVERSITY OF CALIFORNIA) 07 July 2022 (2022-07-07) claims 1-137	1-10
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.															
X	WO 2022087272 A1 (LELAND STANFORD JUNIOR UNIVERSITY) 28 April 2022 (2022-04-28) claims 1-8, 24, 25 and 42-46, and description, paragraphs 0037 and 00165	4															
Y	WO 2022087272 A1 (LELAND STANFORD JUNIOR UNIVERSITY) 28 April 2022 (2022-04-28) claims 1-8, 24, 25 and 42-46, and description, paragraphs 0037 and 00165	1-3, 5-10															
Y	CN 111613272 A (WESTLAKE UNIVERSITY) 01 September 2020 (2020-09-01) description, paragraph 0006	1-3, 5-10															
A	WO 2022147573 A1 (UNIVERSITY OF CALIFORNIA) 07 July 2022 (2022-07-07) claims 1-137	1-10															
☐ 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 “D” document cited by the applicant in the international application “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 12 December 2023		Date of mailing of the international search report 15 December 2023															
Name and mailing address of the ISA/CN China National Intellectual Property Administration (ISA/CN) China No. 6, Xitucheng Road, Jimenqiao, Haidian District, Beijing 100088		Authorized officer Telephone No.															

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2023/085605

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.
 - b. ☐ furnished subsequent to the international filing date for the purposes of international search (Rule 13ter.1(a)),
☐ accompanied by a statement to the effect that the sequence listing does not go beyond the disclosure in the international application as filed.
2. ☐ With regard to any nucleotide and/or amino acid sequence disclosed in the international application, this report has been established to the extent that a meaningful search could be carried out without a WIPO Standard ST.26 compliant sequence listing.
3. Additional comments:

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2023/085605

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
WO	2022087272	A1	28 April 2022	JP	2023546681	A	07 November 2023
				CA	3196425	A1	28 April 2022
				EP	4232584	A1	30 August 2023
CN	111613272	A	01 September 2020	CN	111613272	B	13 October 2023
WO	2022147573	A1	07 July 2022	EP	4240338	A1	13 September 2023

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:
- b. ☐ 为国际检索的目的在国际申请日之后提交(细则13之三.1(a)),
☐ 附有说明序列表不超出所提交国际申请公开范围的声明。

2. ☐ 本报告是在没有收到符合WIPO ST.26标准的序列表的情况下,考虑了国际申请中披露的任何核苷酸和/或氨基酸序列,在可进行有意义检索的范围内做出的。

3. 补充意见:

A. 主题的分类		
C12N15/113(2010.01)i; C12N15/09(2006.01)i; C12N15/11(2006.01)i; C12N15/10(2006.01)i; C12N15/63(2006.01)i; A61K48/00(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)		
IPC:C12N A61K48 C12Q G16B		
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))		
CNABS;CNTXT;CJFD;WPABS;WPABSC;DWPI;ENTXT;ENTXTC;CNKI;万方;ISI Web of Science;NCBI;EMBL;STN;中国专利生物序列检索数据库: 时夕生物, 杨文兵, 向导RNA, 指导RNA, 连接子, 接头, 环序列, 茎环, 条形码, 分子标识符, 筛选, 靶序列, 靶RNA, 靶标, RNA 编辑, 编辑RNA, 测序, gRNA, guide RNA, linker, barcode, UMI, screen, RNA editing, sequencing, ADAR, 序列557、559、561、563、565、567、569、571、573、575、577、579、583-602、603-622、623-642		
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	WO 2022087272 A1 (UNIV LELAND STANFORD JUNIOR) 2022年4月28日 (2022 - 04 - 28) 权利要求1-8、24、25、42-46, 说明书第0037段、第00165段	4
Y	WO 2022087272 A1 (UNIV LELAND STANFORD JUNIOR) 2022年4月28日 (2022 - 04 - 28) 权利要求1-8、24、25、42-46, 说明书第0037段、第00165段	1-3、5-10
Y	CN 111613272 A (西湖大学) 2020年9月1日 (2020 - 09 - 01) 说明书第0006段	1-3、5-10
A	WO 2022147573 A1 (UNIV CALIFORNIA) 2022年7月7日 (2022 - 07 - 07) 权利要求1-137	1-10
☐ 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 “D” 申请人在国际申请中引证的文件 “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
2023年12月12日		2023年12月15日
ISA/CN的名称和邮寄地址		授权官员
中国国家知识产权局 中国北京市海淀区蓟门桥西土城路6号 100088		夏文静
		电话号码 (+86) 0512-88996514

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2023/085605

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)
WO	2022087272	A1	2022年4月28日	JP	2023546681	A	2023年11月7日
				CA	3196425	A1	2022年4月28日
				EP	4232584	A1	2023年8月30日
CN	111613272	A	2020年9月1日	CN	111613272	B	2023年10月13日
WO	2022147573	A1	2022年7月7日	EP	4240338	A1	2023年9月13日