

(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
2020 年 9 月 3 日 (03.09.2020)



(10) 国际公布号
WO 2020/173150 A1

(51) 国际专利分类号:
C12N 15/90 (2006.01) *C12Q 1/34* (2006.01)
C12N 9/78 (2006.01)

(21) 国际申请号: PCT/CN2019/119842

(22) 国际申请日: 2019 年 11 月 21 日 (21.11.2019)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201910153546.3 2019 年 2 月 28 日 (28.02.2019) CN
201910494323.3 2019 年 6 月 9 日 (09.06.2019) CN

(71) 申请人: 中国科学院脑科学与智能技术卓越创新中心 (CENTER FOR EXCELLENCE IN BRAIN SCIENCE AND INTELLIGENCE TECHNOLOGY, CHINESE ACADEMY OF SCIENCES) [CN/CN]; 中国上海市岳阳路 320 号, Shanghai 200031 (CN)。

(72) 发明人: 杨辉 (YANG, Hui); 中国上海市徐汇区岳阳路 320 号新生命科学大楼上海生命科学研究院神经科学研究所 A2316 室, Shanghai 200031 (CN)。
左二伟 (ZUO, Erwei); 中国上海市徐汇区岳阳路 320 号新生命科学大楼上海生命科学研究院神经科学研究所 A2316 室, Shanghai 200031 (CN)。

(74) 代理人: 上海专利商标事务所有限公司 (SHANGHAI PATENT & TRADEMARK LAW OFFICE, LLC); 中国上海市桂平路 435 号, Shanghai 200233 (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, 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, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, 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))。

(54) **Title:** OFF-TARGET SINGLE NUCLEOTIDE VARIANTS CAUSED BY SINGLE-BASE EDITING AND HIGH-SPECIFICITY OFF-TARGET-FREE SINGLE-BASE GENE EDITING TOOL

(54) **发明名称:** 单碱基编辑导致非靶向单核苷酸变异及避免该变异的高特异性无脱靶单碱基基因编辑工具

(57) **Abstract:** Provided are a method for reducing the off-target effect of a single-base editor, and a method (GOTI) for analyzing the targeting effect of a gene editing tool or a gene editing operation.

(57) **摘要:** 提供了一种降低单碱基编辑器的脱靶效应的方法和一种分析基因编辑工具或基因编辑操作的靶向效应的方法 (GOTI)。



WO 2020/173150 A1

单碱基编辑导致非靶向单核苷酸变异及避免该变异的高特异性无脱靶单碱基基因编辑工具

技术领域

5 本发明属于基因编辑技术领域，更具体地，本发明涉及单碱基编辑导致非靶向单核苷酸变异的现象，本发明还涉及用于避免该变异的高特异性无脱靶单碱基基因编辑工具。

背景技术

10 基因组编辑技术自问世以来受到人们的高度重视。CRISPR 是规律间隔性成簇短回文重复序列的简称，Cas 是 CRISPR 相关蛋白的简称。CRISPR/Cas 最初是在细菌体内发现的，是细菌用来识别和摧毁抗噬菌体和其他病原体入侵的防御系统。在 CRISPR/Cas9 系统中，酶 Cas9 在 DNA 靶位点进行切割。Cas9 与 sgRNA 一起被称作 Cas9-sgRNA 系统。CRISPR/Cas9 技术已被应用于疾病模型建立、药物靶点筛选、
15 并正在成为新一代基因治疗手段。

但是，CRISPR/Cas9 和 base editors 介导的基因编辑方法已经开发出来，并为治疗致病性突变引起的遗传疾病带来了很大的希望。基于 CRISPR/Cas9 基因编辑或者碱基编辑的临床应用需要综合性分析脱靶效应，减少有害突变的风险。尽管本领域中开发了多种方法来检测全基因组的基因编辑细胞的离靶活性，包括高通量全基因组易位测序 (HTGTS)，全基因组，双链断裂的无偏差识别(GUIDE-seq)和通过循环测序在体外报告切割效率(CIRCLE-seq)。然而，这些方法都不能有效地检测单核苷酸变异(SNV)。至今本
20 领域还没有一种有效的方法来检测 SNV。

并且，CRISPR/Cas9 在应用时，缺陷在于同源介导修复的低编辑效率。本领域人员利用 16 个碱基长的 XTEN 接头(XTEN linker)将胞苷脱氨酶 APOBEC1 与 dCas9 连
25 接在一起，从而构建出第一代碱基编辑器(BE1)。为了增加体内编辑效率，第二代碱基编辑器(BE2)系统除了将胞苷脱氨酶与 dCas9 连接在一起之外，还将碱基切除修复抑制剂 UGI 与 dCas9 融合在一起，从而将编辑效率提高三倍，最高达到 20%左右。

为了将碱基编辑效率提高，本领域人员将 dCas9 换为 Cas9n 来模拟错配修复，从而构建出第三代碱基编辑器(BE3)。BE3 在非互补 DNA 链上产生切口，细胞使用含尿嘧啶(U)的 DNA 链作为模板来进行修复，从而复制这种碱基编辑。在人细胞系的多种靶基因中，这种 BE3 系统使得碱基编辑效率发生显著性地提高，它的平均 indel(insertion-deletion)发生率仅为 1.1%。对这些测试的靶基因而言，这些数字是对 Cas9 介导的 HDR 的巨大改进；平均的 HDR 介导的编辑频率仅为 0.5%，并且相比于之前的单碱基编辑，更多的 indel 被观察到。在多次细胞分裂中，CRISPR 碱基编辑持
30 续存在，说明这种方法产生稳定的碱基编辑。但是，这种 BE3 系统也会遭受脱靶编

辑的影响。

基因组编辑具有治疗由致病突变诱导的遗传疾病的巨大潜力，对基因编辑的脱靶效应进行综合分析对其实用性是非常必需的。同时，本领域还需要进一步地找到解决脱靶问题的手段。

5

发明内容

本发明的目的在于研究单碱基编辑导致非靶向单核苷酸变异的现象，以及提供一种高特异性无脱靶单碱基基因编辑工具。

在本发明的第一方面，提供一种降低单碱基编辑器的脱靶效应的方法，包括：在
10 单碱基编辑器系统中，对其中的胞嘧啶脱氨酶(cytidine deaminase enzyme(s))进行改造，弱化其与 DNA 的结合能力。

在一个优选例中，所述的改造是针对胞嘧啶脱氨酶的 DNA 结合区域的改造；较佳地，所述的 DNA 结合区域为其与 DNA(如 ssDNA)结合的结构域。

在另一优选例中，所述的改造包括但不限于：基因突变，靶向性封闭或阻滞(如
15 以结合蛋白或抗体进行封闭或阻滞，或以竞争性结合分子进行封闭或阻滞)，干扰。

在另一优选例中，所述的单碱基编辑器系统为 BE3 基因编辑器系统。

在另一优选例中，所述的 DNA 为单链 DNA (ssDNA)或双链 DNA(dsDNA)。

在另一优选例中，所述的胞嘧啶脱氨酶包括但不限于选自下组的酶：AID (如：
20 人 AID)，APOBEC3G (如：人 APOBEC3G)，APOBEC1，APOBECA3A，CDA1 (如：
lamprey CDA1)。

在另一优选例中，所述弱化为显著性的弱化，例如该弱化使得胞嘧啶脱氨酶与 DNA(较佳地为 ssDNA)的结合能力降低 10%、20%、30%、40%、50%、60%、70%、80%、90%或以上，或降低 100%。

在另一优选例中，所述的胞嘧啶脱氨酶为 APOBEC1；较佳地，所述的改造为对
25 该酶的第 126 位的氨基酸进行改造；更佳地，所述的改造为将其第 126 位的 R 突变为 E。

在另一优选例中，所述的改造还包括：对 APOBEC1 酶的第 132 位的氨基酸进行改造；较佳地，所述的改造为将其第 132 位突变为 E。

在另一优选例中，所述的改造还包括：对 APOBEC1 酶的第 90 位的氨基酸进行
30 改造；较佳地，所述的改造为将其第 90 位突变为 Y。

在另一优选例中，所述的改造还包括：对 APOBEC1 酶的第 90 位的氨基酸进行改造；较佳地，所述的改造为将其第 90 位突变为 F。

在另一优选例中，所述的改造还包括：对 APOBEC1 酶第 90 位氨基酸和第 126 位氨基酸改造 W90Y 合并 R126E。

35 在另一优选例中，所述的改造还包括：对 APOBEC1 酶第 90 位氨基酸和第 126

位氨基酸改造 W90F 合并 R126E。

在另一优选例中，所述的胞嘧啶脱氨酶为 APOBECA3A，所述的改造为对该酶的第 130 位的氨基酸进行改造；较佳地，所述的改造为将其第 130 位的 Y 突变为 F。

在本发明的另一方面，提供一种胞嘧啶脱氨酶的突变体，其 DNA 结合区域发生
5 改造，该改造使得其与 DNA 的结合被弱化，所述的 DNA 如单链 DNA。

在一个优选例中，所述的胞嘧啶脱氨酶包括但不限于选自下组的酶：AID，APOBEC3G，APOBEC1，APOBECA3A，CDA1。

在另一优选例中，所述的酶为 APOBEC1；较佳地，所述改造发生在该结构域的第 126 位上或其邻近的位置上；更佳地，所述的改造为将其第 126 位的 R 突变为 E。

10 在另一优选例中，所述的酶为 APOBEC1；较佳地，所述改造发生在该结构域的第 132 位上或其邻近的位置上；更佳地，所述的改造为将其第 132 位的 R 突变为 E。

在另一优选例中，所述的改造还发生在 APOBEC1 酶的第 90 位氨基酸上；较佳地，所述的改造为将其第 90 位突变为 Y。

15 在另一优选例中，所述的酶为 APOBECA3A，所述改造发生在该酶的第 130 位的氨基酸上或其邻近的位置上；较佳地，所述的改造为将其第 130 位的 Y 突变为 F。

在本发明的另一方面，提供分离的多核苷酸，所述的多核苷酸编码所述的突变体。

在本发明的另一方面，提供一种载体，它含有所述的多核苷酸。

在本发明的另一方面，提供一种遗传工程化的宿主细胞，它含有所述的载体，或基因组中整合有所述的多核苷酸。

20 在本发明的另一方面，提供一种单碱基编辑器，其包含所述的胞嘧啶脱氨酶的突变体；较佳地所述的编辑器为 BE3 单碱基编辑器。

在本发明的另一方面，提供一种生产所述的胞嘧啶脱氨酶的突变体的方法，包括步骤：(1) 培养所述的宿主细胞，获得培养物；和(2) 从培养物中分离所述的胞嘧啶脱氨酶突变体。

25 在本发明的另一方面，提供所述的胞嘧啶脱氨酶突变体的用途，用于参与基于单碱基编辑器系统的基因编辑，以降低基因编辑器的脱靶效应。

在一个优选例中，所述的胞嘧啶脱氨酶突变体的用途可以是非治疗性的用途。

在本发明的另一方面，提供一种筛选对于降低单碱基编辑器的脱靶效应有用的物质的方法，包括：(1) 用候选物质处理胞嘧啶脱氨酶或其 DNA 结合结构域以及 DNA(如 ssDNA)相互作用(结合)的体系；和(2) 检测所述体系中胞嘧啶脱氨酶 DNA 结合结构域与 DNA 相互作用情况；其中，若所述候选物质抑制、阻滞或下调胞嘧啶脱氨酶或其 DNA 结合结构域与 DNA 的相互作用，则表明该候选物质是对于降低基因编辑器的脱靶效应有用的物质。

35 在一个优选例中，所述的候选物质包括(但不限于)：小分子化合物，针对胞嘧啶脱氨酶或其 DNA 结合结构域或其编码核酸设计的结合分子(如抗体或配体)、阻滞分子(如

基于氨基酸修饰的阻滞剂)、干扰分子、基因编辑试剂、核酸抑制物; 和/或

在另一优选例中, 所述的体系包括(但不限于): 细胞体系(如表达胞嘧啶脱氨酶或其 DNA 结合结构域以及含有 DNA(如 ssDNA)的细胞)(或细胞培养物体系)、亚细胞体系、溶液体系、组织体系、器官体系或动物体系。

5 在本发明的另一方面, 提供一种分析单碱基基因编辑工具的靶向效应的方法(GOTI), 述方法包括: (1) 获取 n -细胞阶段的胚胎, 在其中 1~n-1 个细胞中进行基因编辑操作; 保留至少一或几个细胞则不进行基因编辑操作; 其中所述 n 为 2~10 的正整数; (2) 在胚胎发育的下游阶段, 观测发生基因编辑的情况。

10 在一个优选例中, 步骤(1)中, 所述的 n 为 2~8, 2~6 或 2~4 的正整数; 较佳地, 所述 n 为 2。

在另一优选例中, 所述的方法为体外培育方法或体内培育方法。

另一优选例中, 在发育的卵裂期, 可以将同一个胚胎中基因编辑的卵裂球和没有编辑的卵裂球拆开, 然后移植到受体(如小鼠)体内, 可以分别发育成成体。

15 在另一优选例中, 步骤(2)中, 所述胚胎发育的下游阶段为胚胎发育到原肠胚阶段至出生前阶段, 或在体内状态下为胚胎植入子宫至出生前的阶段。

在另一优选例中, 所述的胚胎为小鼠的胚胎, 所述胚胎发育的下游阶段为胚胎发育的第 8~20 天(E8-E20 阶段), 较佳地为胚胎发育的第 9.5~18.5 天(E9.5-E18.5 阶段), 更佳地为胚胎发育的第 12~16(E11-E16 阶段, 如 E14.5)。

20 在另一优选例中, 所述的基因编辑操作包括(但不限于): CRISPR 介导的基因编辑, Base Editor(碱基编辑器)介导的基因编辑, Cre/loxP 介导的基因编辑, 腺嘌呤碱基编辑器介导的基因编辑。

在另一优选例中, 所述的 CRISPR 介导的基因编辑包括(但不限于): CRISPR/Cas9 介导的基因编辑, CRISPR/Cas9n 介导的基因编辑, CRISPR/Cas13(如 CRISPR/Cas13a, CRISPR/Cas13d)介导的基因编辑, CRISPR/CasRx 介导的基因编辑。

25 在另一优选例中, 所述的 Base Editor 包括: BE1, BE2, BE3, BE4, BE4-Max。

在另一优选例中, 所述的腺嘌呤碱基编辑器包括: ABE7.10, ABE 6.3, ABE 7.8, ABE 7.9, Prime Editing。

30 在一个优选例中, 步骤(1)中, 包括: 将用于切割核酸(如 DNA)靶位点的酶的(如 Cas mRNA, Cre mRNA), 以及相应导向序列(如 sgRNA)引入到所述的其中一个细胞中, 进行基因编辑。

在另一优选例中, 所述的用于切割核酸(如 DNA)靶位点的酶选自但不限于下组的酶: Cas9, Cas9n, Cas13a, CasRx, BE1, BE2, BE3, BE4, BE4-Max, ABE7.10, ABE 6.3, ABE 7.8, ABE 7.9, Prime Editing。

35 在另一优选例中, 步骤(1)中, 采用可检测标记物来标记所述的基因编辑操作, 在其中 1~n-1 个细胞中进行基因编辑操作并设置可检测标记物。

在另一优选例中，所述的可检测标记物包括但不限于：染色标记物、荧光信号分子、报告基因；更佳地，所述的可检测标记物为(但不限于)tdTomato，EGFP，mCherry，GFP，dsred。

在另一优选例中，步骤(2)中，所述观测发生基因编辑的情况包括：

- 5 分选发生基因编辑的细胞(如 tdTomato 阳性细胞)以及未发生基因编辑的细胞(如 tdTomato 阴性细胞)；

另一优选例中，在发育的卵裂期，可以将同一个胚胎中基因编辑的卵裂球和没有编辑的卵裂球拆开，然后移植到受体(如小鼠)体内，可以分别发育成成体，不使用流式细胞术分选。

- 10 进行测序分析(如 WGS 分析)；

利用单核苷酸变异(SNV)分析工具和/或 indel 分析工具进行分析；

比较发生编辑的细胞和未发生编辑的细胞来鉴定靶向效应或脱靶效应，包括检测 SNV 和 indel。

- 15 在另一优选例中，所述的 SNV 分析工具包括但不限于：Mutect2，Lofreq 和 Strelka 或其组合；或，所述的 indel 分析工具包括但不限于：Mutect2，Scalpel，Strelka 或其组合。

在另一优选例中，采用流式细胞术进行发生基因编辑的细胞(如 tdTomato 阳性细胞)以及未发生基因编辑的细胞(如 tdTomato 阴性细胞)的分选。

- 20 在另一优选例中，所述的分析单碱基基因编辑工具的靶向效应的方法(GOTI)可以为非治疗性的方法。

在另一优选例中，所述的分析单碱基基因编辑工具的靶向效应的方法可以为离体的方法。

在另一优选例中，所述的胚胎来自于哺乳动物，包括但不限于非人哺乳动物，如鼠、兔、羊、牛、猴等。

- 25 本发明的其它方面由于本文的公开内容，对本领域的技术人员而言是显而易见的。

附图说明

- 30 图 1、在一个 2-cell 阶段胚胎的一个卵裂球中注射 CRISPR-Cas9, BE3 或 ABE7.10 介导的基因编辑。(A)实验设计：将 Cre，Cas9/BE3 和 sgRNA 混合物注射入 2 细胞胚胎中的一个卵裂球，胚胎来自 Ai9 雄鼠与野生型雌鼠交配而来。Cre 的作用是有望产生一个嵌合胚胎，其中一半的细胞标记为 tdTomato(红色)。流式细胞仪将 E14.5 嵌合体胚胎中分离 tdTomato+和 tdTomato-细胞并用于全基因组测序。通过比较 tdTomato+细胞和 tdTomato-细胞，使用三种不同的调用算法来识别脱靶的 SNV 和 indel(SNV 使用 Mutect2、Lofreq 和 Strelka，indel 使用 Mutect2、Scalpel 和 Strelka)。SNV 和 indel 标记为彩点
- 35

和叉。(B)指定胚胎中的 FACS 分析。(C)基于 WGS 的 tdTomato +和 tdTomato-细胞的靶向效率。在 tdTomato +细胞中 Cas9, BE3 和 ABE7.10 的靶向效率为 66% $\pm$ 12%插入或者缺失(SEM, n = 5), 83% $\pm$ 10%(SEM, n = 4)和 47% $\pm$ 分别为 $\sim$ 18%(SEM, n = 2)单碱基编辑。(D)Cas9-Tyr-A、Cas9-Tyr-B、BE3-Tyr-C、BE3-Tyr-D 处理 E14.5 胚胎的流式细胞仪分析; 未注射胚胎的流式细胞分析如图 S1b 所示。(E)基于全基因组测序 tdTomato+ 细胞(左)和 tdTomato-细胞(右)的靶向效率; 同样处理的细胞用同样的颜色表示。(F)Cas9-Tyr-A、Cas9-Tyr-B、BE3-Tyr-C、BE3-Tyr-D 处理的 E14.5 胚胎, TA 克隆测序 On-target 分析; 每一列上的数字代表分析克隆的总数。

图 2、在 BE3 处理的小鼠胚胎中产生的大量脱靶 SNV。(A)检测到的脱靶 SNV 的总数的比较。 Cre-, Cas9-, BE3-和 ABE7.10 处理的胚胎的 SNV 数量为 14 $\pm$ 12(SEM, n=2), 12 $\pm$ 4(SEM, n = 11), 283 $\pm$ /分别为-32(SEM, n = 6)和 10 $\pm$ 5 SNV(SEM, n = 4)。(B)突变类型的分布。每个细胞中的数字表示所有突变中某种突变的比例。(C)C> T 和 G 的比例> Cre, Cas9 和 BE3 的 A 突变, 以及 ABE7.10 的 T> C 和 A> G 突变。分析了两个 Cre, 11 个 Cas9,6 个 BE3 和 4 个 ABE7.10 样品。A)和(B)中的 P 值通过 two-sided Wilcoxon。

图 3、BE3 诱导的脱靶 SNV 的特征。(A)与随机排列相比, 脱靶基因 SNV 在基因组的转录区域中富集。(B)含有脱靶 SNV 的基因表达显著高于 4 细胞胚胎中的随机模拟基因。(C)从每个胚胎中鉴定的 SNV 是非重叠的。(D)由 GOTI 检测到的 SNV 与 Cas-OFFinder 和 CRISPOR 预测的脱靶之间的重叠。(A)和(B)中的 P 值通过 two-sided Wilcoxon。

图 4、BE3 脱靶评估所构建的 BE2 系统。(a)质粒图谱。(b)来自 WGS 数据的 R126E 突变 BE3 的靶向效率。(c)检测到的脱靶 SNV 的总数的比较。

图 5、Apobec1 点突变可以消除 BE3 的 DNA 和 RNA 脱靶。(a) BE3 体系中的 APOBEC1; (b) BE3 用量(BE3 显微注射浓度)与靶向效率的关联性; (c)测序确认靶向效率; (d) 比较不同的突变体脱靶情况。显示 BE3^{R126E}, BE3^{R126E+W90Y} DNA 脱靶明显降低; (e)突变体与脱靶效应的关联性分析。

图 6、不同混合处理的 E14.5 胚胎流式细胞仪分析。(a)Cas9-Tyr-A 基因靶向胚胎的代表图。上部: 四细胞胚胎, 下部: 一个分散的 8 细胞胚胎。红色箭头表示 tdTomato+ 卵裂球。比例尺: 100 μ m。(b)从左到右, 从上到下依次排列: 未注射, Cre-#1, Cre-#2, Cre+Cas9-#1, Cre+Cas9-#2, Cre+Cas9+LacZ-#1, Cre+Cas9+LacZ-#2, Cre+Cas9+Pde6b-#1 和 Cre+Cas9+Pde6b-#2。(c)靶向 Tyr 基因的 8 细胞胚胎的基因型。从 4 个 Cas9-Tyr-A 和 4 个 Cas9-Tyr-B 基因靶向的囊胚中分离到单个 tdTomato+和 tdTomato-细胞。数字: 囊胚分析总数。WT: 野生型等位基因。Mutant: Tyr 等位基因突变。

图 7、利用 DNA 体外切割法测定 sgRNAs 的切割效率。琼脂糖凝胶电泳从左到右分别表示 Cas9-Tyr-A, Cas9-Tyr-B, Cas9-LacZ 和 Cas9-Pde6b。对每个基因的 sgRNA 靶

位点两侧的基因组区域或结构进行 PCR 扩增, PCR 产物与 Cas9 核糖核蛋白及 sgRNA 孵育 3h。

图 8、CRISPR/Cas9 和 BE3 处理的嵌合胚胎的发育和基因型。(a)注射不同 mix 后 tdTomato+囊胚的百分比。数字: 囊胚或胚胎总数。(b)注射不同 mix 后 E14.5 胚胎存活率。数字: 胚胎计算总数或转移总数。(c)tdTomato+囊胚百分比及 E14.5 胚胎中 Tyr 突变百分比。数字: 囊胚或胚胎分析总数。

图 9、tdTomato+和 tdTomato-细胞全基因组测序得到的 On-target 序列。WT 和突变序列的全基因组测序结果显示为 WT 和 MUT, 斜杠前的数字表示 WT 或突变序列, 总数在斜杠后显示。突变序列以下划线突出显示, 末三位 TGG、CGG、GGG 为 PAM。

图 10、使用指定的软件工具, 在每个胚胎中, WGS 数据中检测到 SNV 的韦恩图。(a)在 Cre 或 CRISPR/Cas9 处理的样本中检测到的 SNV。(b)在 BE3 处理的胚胎中鉴定的 SNV。从接下来的分析中去除等位基因频率小于 10%的重复 SNV。

图 11、代表性的 Sanger 测序峰图显示全基因组测序在 Cre 或 CRISPR/Cas9 处理胚胎中检测突变的情况。(a)对所示样品进行 PCR 扩增, 并进行 Sanger 测序。绿色箭头: 野生型; 红色箭头: 插入核苷酸。红点线, 缺失核苷酸。(b)通过 Sanger 测序对所示样品的 SNV 进行了验证。绿色箭头: 野生型核苷酸; 红色箭头: 突变核苷酸。引物见表 S16。

图 12、Cre 和 CRISPR/Cas9 处理的胚胎中, 从 WGS 数据中检测到的 SNV 数量。同一组胚胎用同样的颜色表示。右边柱图模拟-自发突变数量的分布。

图 13、从 Cre 和 CRISPR/Cas9 处理的胚胎中鉴定出的脱靶 SNV 和 indel。(a)注射 Cre 或 CRISPR/Cas9 的每一个胚胎中所鉴定出的 snv 是相互排斥的。(b)从 CRISPR/ cas9 处理的胚胎中检测到的 SNV 与 Cas-OFFinder 和 CRISPOR 预测的靶外位点的重合。(c)前 10 位预测的 Cas9-Tyr-A 和 Cas9-Tyr-B Off-target 序列比对, 以及从 WGS 数据中检测到的 Cas9-Tyr-A 和 Cas9-Tyr-B 突变。

图 14、通过比较来自相同胚胎的 tdTomato-和 tdTomato+细胞, 总结从 WGS 数据反向调用的变异。(a)从 Cre-和 CRISPR/Cas9 处理的样本中调用相反的变量。(b)从 BE3 处理的样本中调用相反结果。

图 15、在此研究中鉴定的每一个胚胎的突变类型。每个小格的数量表示某种突变类型的比例, 颜色越深表示突变类型的比例越高。

图 16、从四组 Cistrome 数据集比较有无识别的 off-target 峰区。每个柱形图顶部的数字表示应用数据集的 GEO accession。P 值由 Wilcoxon rank sum test 计算得出。

具体实施方式

基因组编辑有望纠正致病突变。但是由于不同个体间的单核苷酸多态性, 基因编辑的脱靶效应很难确定。为了研究此类脱靶效应, 本发明人开发了一种通过二或多细胞(优选二细胞)胚胎注射进行全基因组脱靶分析的方法, 命名为 GOTI。本发明的方法适用于 CRISPR 介导的基因编辑, Base Editor 介导的基因编辑, Cre/loxP 介导的基

因编辑，腺嘌呤碱基编辑器介导的基因编辑后的追踪分析，检测靶向效应/效率。

本发明提供了一种分析单碱基基因编辑工具的靶向效应的方法(GOTI)，所述方法包括：(1)获取 n-细胞阶段的胚胎，在其中 1~n-1 个个细胞中进行基因编辑操作；其中所述 n 为 2~10 的正整数；(2)在胚胎发育的下游阶段，观测发生基因编辑的情况。

5 在一些优选方式中，所述的 n 为 2~8，2~6 或 2~4 的正整数。在本发明的优选实施例中，所述 n 较佳地为 2。

本发明的方法适用于在体外状态下的胚胎培养，例如试管中或其它胚胎培养容器中进行胚胎的培育。本发明的方法也适用于在体内状态下的胚胎培育，例如在体外进行如本发明的方法的处理后，将发育的细胞移植到体内，例如移植到动物的输卵管中，
10 由胚胎自行游动到子宫中；或移植到动物的子宫中。

本发明的方法适用于在体外状态下的胚胎培养，例如试管中或其它胚胎培养容器中进行胚胎的培育。

本发明的方法适用于在体外状态下的胚胎培养，胚胎培养容器中进行胚胎的培育，建立胚胎干细胞系。

15 本发明的方法适用于在体外状态下的胚胎培养，胚胎培养容器中进行胚胎的培育，编辑后的卵裂球和没有编辑的卵裂球分别建立胚胎干细胞系。

本发明的方法适用于同一胚胎将编辑后卵裂球和没有编辑的链裂球拆开，形成两个胚胎分别移植到受体不同小鼠体内或者体外建立胚胎干细胞系。

20 本发明的方法适用于同一胚胎将编辑后卵裂球和没有编辑的卵裂球拆开，形成两个胚胎同时移植到受体同一小鼠体内或者体外建立胚胎干细胞系。

本发明的方法也适用于在体内状态下的胚胎培育，例如在体外进行如本发明的方法的处理后，将发育的细胞移植到体内，例如移植到动物的输卵管中，由胚胎自行游动到子宫中；或移植到动物的子宫中。

在本发明的优选方式中，所述胚胎发育的下游阶段为胚胎发育到原肠胚阶段至出生前阶段，或在体内状态下为胚胎植入子宫至出生前的阶段。本发明人发现，在胚胎发育的“适当时间”进行细胞的分选以及基因编辑效应的测定是较为理想的。一般地，所述的“适当时间”为胚胎生长至适于被酶分解为单细胞的阶段。例如，所述的 n-细胞阶段的胚胎是小鼠的胚胎，那么所述胚胎发育的下游阶段为胚胎发育的第 8~20 天(E8-E20 阶段)，较佳地为胚胎发育的第 9.5~18.5 天(E9.5-E18.5 阶段)，更佳地为胚胎发育的第 12~16(E11-E16 阶段，如 E14.5)。
25 30

本发明的方法适用于多种单碱基基因编辑的方法中。在多种切割 DNA 靶位点的酶参与的基因编辑中，均可采取本发明的方法。所述的切割 DNA 靶位点的酶可以是本领域技术人员熟悉的多种参与此过程的酶，例如可以是但不限于下组的酶：Cas9，Cas9n，Cas13a，CasRx，BE1，BE2，BE3，BE4，ABE7.10，ABE 6.3，ABE 7.8，ABE
35 7.9，Prime Editing。

在所述的 GOTI 方法中，可以采用可检测标记物来标记所述的基因编辑操作，所述的可检测标记物包括但不限于：染色标记物、荧光信号分子、报告基因。本发明的实施例中，采用了 tdTomato，这是一种优选的方案，但是其它的标记物也可被应用于本发明中。

5

作为本发明的优选方式，所述观测发生基因编辑的情况包括：分选发生基因编辑的细胞(如 tdTomato 阳性细胞)以及未发生基因编辑的细胞(如 tdTomato 阴性细胞)；进行测序分析(如 WGS 分析)；利用 SNV 分析工具和/或 indel 分析工具进行分析；比较发生编辑的细胞和未发生编辑的细胞来鉴定脱靶 SNV 和 indel。应理解，测序工具以及分析工具并不限于上述以及本发明实施例中列举的，其它的一些测序工具以及分析工具也可被应用于本发明中。可以应用本领域中已知的多种方法进行细胞分选，例如但不限于磁珠法、流式细胞术等。

本发明中，所述“动物”是指哺乳纲的动物，包含人、非人灵长类动物(猴、猩猩)、家畜与农畜(例如，猪、绵羊、牛)，鼠(小鼠)，以及啮齿动物(例如，小鼠、大鼠、兔)等。所述的动物是不包含人的动物；在有限的或特殊的情况下，所述的动物也可以是人，但这仅被应用于并非涉及“人胚胎的商业应用”的范畴内。

在本发明的具体实施例中，编辑与未编辑的卵裂球的子代细胞在 E14.5 的全基因组序列的比较显示，在 CRISPR-Cas9 或腺嘌呤单碱基编辑的胚胎中，单核苷酸突变(SNV)脱靶是罕见的，频率接近于自发突变率。相比之下，胞嘧啶单碱基编辑诱导产生了超过 20 倍的脱靶单核苷酸突变。

在临床应用前，哺乳细胞中要求没有全基因组脱靶，但是由于个体中核苷酸多态性，因此很难确定脱靶效应的程度。本发明开发的 GOTI (双细胞胚胎注射全基因组脱靶分析)的方法改变了这一现状，在不干扰 SNPs 的情况下检测脱靶突变，能准确有效地实现基因组靶向效应的分析。

25

本发明人进一步研究了导致单碱基编辑发生脱靶效应(如单核苷酸脱靶变异)的原因。在观测到单碱基编辑工具 BE3 会导致大量的单核苷酸脱靶变异(SNV)后，本发明人进行了大量的研究工作，最终确定这些脱靶变异是由 APOBEC1 的过表达及其与 DNA(如 ssDNA)的结合引起的。在具体的实施例中，本发明揭示一种解决方案，通过在 APOBEC1 上进行突变，如进行 R126E, R132E, W90F, W90Y W90F/R126E, W90Y/R126E 突变，来解决 BE3 诱导的脱靶效应。

根据上述，本发明中确定了一个对于降低单碱基编辑器的脱靶效应有用的方法，包括：在单碱基编辑器系统中，对其中的胞嘧啶脱氨酶进行改造，弱化其与 DNA(如 ssDNA)的结合。较佳地，所述的改造是针对胞嘧啶脱氨酶的 DNA 结合区域的改造；更佳地，所述的 DNA 结合区域为其与 DNA 结合的结构域。所述的单碱基编辑器例

35

如是 BE3 基因编辑器。

只要能发挥弱化的作用，针对胞嘧啶脱氨酶的多种改造方法都是可用的。作为一些可选的方式，所述的改造可以是基因突变，靶向性封闭或阻滞(如以结合蛋白或抗体进行封闭或阻滞，或以竞争性结合分子进行封闭或阻滞)，干扰等。

5 可应用于单碱基编辑器系统的多种胞嘧啶脱氨酶或与其同功能的酶都可被本发明的方法改造，以降低其所在的单碱基编辑器系统的脱靶效应。例如，所述的胞嘧啶脱氨酶包括但不限于选自下组的酶：AID (如：人 AID)，APOBEC3G (如：人 APOBEC3G)，APOBEC1，CDA1 (如：lamprey CDA1)。

10 本发明中，所述的“弱化”是指胞嘧啶脱氨酶与 DNA 的相互作用(结合)能力被下调或消除。例如该弱化使得胞嘧啶脱氨酶与 DNA 的结合能力降低 10%、20%、30%、40%、50%、60%、70%、80%、90%或以上，或降低 100%。

作为本发明的优选方式，提供了一种具体的胞嘧啶脱氨酶 APOBEC1(野生型序列见 SEQ ID NO: 1，其一种突变体见 SEQ ID NO: 4)，对其 DNA 结合区域的进行改造后，其所参与的单碱基编辑器系统的编辑结果发生了极为显著的变化，脱靶效应显著
15 性降低。较佳地，这种改造为对该酶的第 126 位的氨基酸进行改造；更佳地，所述的改造为将其第 126 位的 R 突变为 E。

在更为优选的方式中，针对 APOBEC1 的改造还发生在 APOBEC1 酶的第 90 位氨基酸上；较佳地，所述的改造为将其第 90 位突变为 Y。

20 在更为优选的方式中，针对 APOBEC1 的改造还发生在 APOBEC1 酶的第 90 位氨基酸上；较佳地，所述的改造为将其第 90 位突变为 Y。

作为本发明的另一优选方式，提供了一种具体的胞嘧啶脱氨酶 APOBECA3A(SEQ ID NO: 37)，针对 APOBECA3A 的改造发生在该酶的第 130 位的氨基酸上或其邻近的位置上；较佳地，所述的改造为将其(SEQ ID NO: 37)第 130 位的 Y 突变为 F。

25 根据本发明人的新发现，进一步还提供一种筛选对于降低 BE3 基因编辑器的脱靶效应有用的物质的方法，包括：(1) 用候选物质处理胞嘧啶脱氨酶或其 DNA 结合结构域以及 DNA 相互作用(结合)的体系；和(2) 检测所述体系中胞嘧啶脱氨酶 DNA 结合结构域与 DNA 相互作用情况；其中，若所述候选物质抑制、阻滞或下调胞嘧啶脱氨酶或其 DNA 结合结构域与 DNA 的相互作用，则表明该候选物质是对于降低 BE3 基因编辑器的脱靶效应有用的物质。

35 在本发明的优选方式中，在进行筛选时，为了更易于观察到胞嘧啶脱氨酶或其 DNA 结合结构域以及 DNA 相互作用(结合)的改变，还可设置对照组，所述的对照组可以是不添加所述候选物质的胞嘧啶脱氨酶或其 DNA 结合结构域以及 DNA 相互作用(结合)的体系。

作为本发明的优选方式，所述的方法还包括：对获得的潜在物质进行进一步的细胞实验和/或动物试验，以进一步选择和确定对于调节两者的相互作用(结合)真正有用的物质。

5 下面结合具体实施例，进一步阐述本发明。应理解，这些实施例仅用于说明本发明而不适用于限制本发明的范围。下列实施例中未注明具体条件的实验方法，通常按照常规条件如 J.萨姆布鲁克等编著，分子克隆实验指南，第三版，科学出版社，2002中所述的条件，或按照制造厂商所建议的条件。

10 材料和方法

1、包含 GOTI 方法的实验设计

分别将 Cre, Cas9/BE3/ABE7.10 的 mRNA 和 sgRNA 的混合物注射到 2 细胞阶段胚胎的一个卵裂球中，该胚胎来自于野生型雌性小鼠 X Ai9 雄性小鼠。Cre 的作用产生嵌合胚胎，其中注射的细胞被 tdTomato(红色)标记，tdTomato 阳性表示发生了编辑，
15 tdTomato 阴性表示未编辑。通过 FACS 在 E14.5 从嵌合胚胎中分离 tdTomato 阳性细胞和 tdTomato 阴性细胞，并分别用于 WGS 分析。通过使用三种算法(用于 SNV 分析的 Mutect2, Lofreq 和 Strelka，以及用于 indel 分析的 Mutect2, Scalpel 和 Strelka)比较 tdTomato+细胞和 tdTomato-细胞来鉴定脱靶 SNV 和 indel。SNV 和插入缺失在图 1A 中表示为有色点和交叉。Cre 蛋白序列如 SEQ ID NO: 2 所示，Cas9 蛋白序列 SEQ
20 ID NO: 3 所示。

2、动物及护理

雌性 C57BL/6 小鼠(4 周龄)和杂合子 Ai9(B6.Cg-Gt(ROSA)26Sortm9 (CAG-td-Tomato)Hze/J; JAX strain 007909)雄性小鼠用于胚胎收集。ICR 雌性小鼠作为受体。动物的使用和护理符合中国科学院上海生命科学研究院生物医学研究伦理委员会的指导原则。
25

3、Cas9 mRNA, BE3 mRNA, ABE7.10 mRNA, Cre mRNA 和 sgRNA

使用引物 Cas9F 和 R 从 px260 质粒扩增出 Cas9 蛋白编码区域。纯化 T7-Cas9 PCR 产物，并且使用 mMESSAGE mMACHINE T7 ULTRA 转录 mRNA。从 px330 质粒扩增出 T7-sgRNA PCR，并使用 MEGA Shortscript T7 试剂盒(Life Technologies)将其体外
30 转录成 RNA。通过 PCR 扩增将 T7 启动子添加至 Cre 模板，纯化 T7-Cre PCR 产物，并使用 mMESSAGE mMACHINE T7 ULTRA 试剂盒(Life Technologies)将其体外转录成 mRNA。使用 MEGA clear kit (Life Technologies) 纯化 Cas9 mRNA, Cre mRNA 和 sgRNA，并在无 RNase 的水中洗脱。

sgRNA 序列(序列编号从上到下依次为：SEQ ID NO: 5-11)

Locus	Sequence (5'-3')
Tyr-A (22)	GCGAAGGCACCGCCCTCTTTTGG
Tyr-B (22)	CCAGAAGCCAATGCACCTATCGG
LacZ (23)	TGCGAATACGCCACGCGATGGG
Pde6b (24)	CCAACCTAAGTAGCAGAAAGTGG
Tyr-C (11)	GACCTCAGTTCCTTCAAAGGG
Tyr-D	CTGTGCCAAGGCAGAAACCCTGG
Tyr-E	CCATAACAGAGACTCTTACATGG

引物序列(序列编号从上到下依次为: SEQ ID NO: 12-26)

Name	Sequence (5'-3')
Cre IVT F	TAATACGACTCACTATAGGGAGACAGATCACCTTTCCTATCAACC
Cre IVT R	TCGGTATTTCCAGCACACTGGA
BE3 IVT F	TCCGCGGCCGCTAATACGACT
BE3 IVT R	TGGTTCCTTCCGCCTCAGAAGCC
Cas9 IVT F	TAATACGACTCACTATAGGGAGATTTCAGGTTGGACCGGTG
Cas9 IVT R	GACGTCAGCGTTCGAATIGC
ABE7.10 IVT F	GAGGTCTATATAAGCAGAGCTC
ABE7.10 IVT R	ATTAATAACTAGCGGCCGCTCCC
Tyr-A IVT F	TAATACGACTCACTATAGGGGCGAAGGCACCGCCCTCTTTGTTTTAGAGCTAGAAATAG
Tyr-B IVT F	TAATACGACTCACTATAGGGGCCAGAAGCCAATGCACCTATGTTTTAGAGCTAGAAATAG
Tyr-C IVT F	TAATACGACTCACTATAGGGGACCTCAGTTCCTTCAAAGTTTTAGAGCTAGAAATAG
Tyr-D IVT F	TAATACGACTCACTATAGGGCTGTGCCAAGGCAGAAACCCGTTTTAGAGCTAGAAATAG
Tyr-E IVT F	TAATACGACTCACTATAGGGCCATAACAGAGACTCTTACAGTTTTAGAGCTAGAAATAG
LacZ-IVT F	TAATACGACTCACTATAGGGTGCGAATACGCCACGCGATGTTTTAGAGCTAGAAATAG
sgRNA IVT R	AAAAGCACCGACTCGGTGCC

5 4、2-细胞注射，胚胎培养和胚胎移植

将超排 C57BL/6 雌性 (4 周龄) 与杂合的 Ai9 B6.Cg-Gt(ROSA)26Sortm9(CAG-td-Tomato)Hze/J; JAX strain 007909) 雄性交配，hCG 注射后 23 小时，从输卵管受精卵。对于 2-cell 编辑，Cas9 mRNA(50 ng/ μ l)，BE3 mRNA(50 ng/ μ l) 或 ABE7.10 mRNA(50 ng/ μ l)，sgRNA(50 ng/ μ l) 和 Cre mRNA(2 ng/ μ l) 的混合物，使用具有恒定流量的 FemtoJet 显微注射器(Eppendorf)，在 hCG 注射后 48 小时，在含有 5 μ g/ml 细胞松弛素 B(CB) 的 HEPES-CZB 培养基的液滴中将 RNA 混合物注射到 2-cell 胚胎的一个卵裂球的细胞质中。注射后的胚胎在含有氨基酸的 KSOM 培养基中于 37°C，5%CO₂ 的空气中培养 2 小时，然后移植到假孕 ICR 雌性的输卵管中。

15 5、单细胞 PCR 分析

解剖显微镜下使用自制玻璃毛细管，用 acid Tyrode solution 溶液消化 8-cell 阶段

小鼠胚胎去除透明带，然后将胚转移到 0.25%胰蛋白酶中并轻轻移液以分离单个卵裂球。最后，将卵裂球 KSOM 中洗涤 7 至 10 次，并转移到 PCR 管中。然后将含有 0.1%吐温 20，0.1%Triton X-100 和 4μg/ml 蛋白酶 K 的 1.5μl 裂解缓冲液移液到管中。将每个管离心以促进混合。将裂解液在 56℃ 的温度下孵育 30 分钟，然后在 95℃ 下孵育 5 分钟。裂解程序的产物用作巢式 PCR 分析中的模板。所有操作中避免污染样品。

Nest PCR 引物 (序列编号从上到下依次为: SEQ ID NO: 27-30)

Tyr	Outer F: GTTATCCTCACACTACTTCTG
	Outer R: GTAATCCTACCAAGAGTCTCA
	Inner F: TCCTCACACTACTTCTGATG
	Inner R: GTCTCAAGATGGAAGATCAC

6、T 载体克隆和基因型检测

纯化 PCR 产物并连接到 pMD18-T 载体上并转化到感受态大肠杆菌菌株 DH5α 中。在 37℃ 过夜培养后，通过 Sanger 方法对随机选择的克隆进行测序。通过从细胞中提取的基因组 DNA 的 PCR 确定突变体 E14.5 胚胎的基因型。ExTaq 在 95℃ 下活化 3 分钟；PCR 进行 34 个循环：95℃ 下 30 秒，62℃ 下 30 秒，72℃ 下 1 分钟；最终在 72℃ 下延伸 5 分钟。对于胚胎，用 KSOM 洗涤 6 次后，将单个胚胎直接转移到含有 1.5μl 胚胎裂解缓冲液(0.1%吐温 20,0.1%Triton X-100 和 4μg/ml 蛋白酶 K)的 PCR 管中并孵育 30 分钟。在 56℃ 时，95℃ 下灭活 10 分钟。使用巢引物组进行 PCR 扩增。ExTaq 在 95℃ 下活化 3 分钟；PCR 进行 34 个循环：95℃ 下 30 秒，62℃ 下 30 秒，72℃ 下 1 分钟；最终在 72℃ 下延伸 5 分钟。使用 0.5μg 第一轮 PCR 产物和内侧引物进行二次 PCR。PCR 在相同的反应混合物中进行。使用 pMD-19t 克隆试剂盒(Takara)按照制造商的说明凝胶纯化和克隆 PCR 产物。从每次转化中挑选菌落，然后进行 Sanger 测序以检测突变。

引物序列(序列编号从上到下依次为: SEQ ID NO: 31-36)

Name	Sequence (5'-3')
Tyr F	GTTATCCTCACACTACTTCTG
Tyr R	GTAATCCTACCAAGAGTCTCA
Tyr-OF	GTCTGTGACACTCATTAACC
Tyr-OR	CATAGGAGGTGCTAACAATAC
Tyr-IF	GTATTGCCTTCTGTGGAGTT
Tyr-IR	TGAACCAATCAGTCCTTGTT

7、流式分选(FACS)

为了分离细胞，将剪碎的组织在 5mL 胰蛋白酶-EDTA(0.05%)的溶液在 37℃ 下酶解 30 分钟。通过加入 5ml 含 10%胎牛血清(FBS)的 DMEM 培养基终止消化。然后

使用 1ml 枪头反复吹打 30-40 次使胎儿组织均质化。将细胞悬浮液离心 6 分钟 (800rpm)，将沉淀重悬于含 10%FBS 的 DMEM 培养基中。最后，将细胞悬浮液通过 40- μ m 细胞过滤器过滤，并通过 FACS 分离 tdtomato + / tdtomato- 细胞。使用第二轮流式细胞术和荧光显微镜分析评估，样品纯度 > 95% 判定合格。

5 8、全基因组测序和数据分析

根据制造商的说明，使用 DNeasy 血液和组织试剂盒(目录号 69504, Qiagen)从细胞中提取基因组 DNA。通过 Illumina HiSeq X Ten 以 50 倍的平均覆盖率进行 WGS。BWA(v0.7.12)用于将合格的测序读数映射到参考基因组(mm10)。然后应用 Picard 工具(v2.3.0)对映射的 BAM 文件的重复项进行排序和标记。为了高可信度地鉴定全基因组从头变异，本发明人分别使用三种算法 Mutect2(v3.5)，Lofreq(v2.1.2)和 Strelka(v2.7.1)进行单核苷酸变异(25-27)分析。同时，分别运行 Mutect2(v3.5)，Scalpel(v0.5.3)和 Strelka(v2.7.1)用于检测全基因组序列。三种 SNV 或 indel 算法的重叠被认为是真正的变体。在 tdTomato+样品的定位 BAM 文件中鉴定变体，其中 tdTomato-样品在与对照相同的胚胎中，仅可鉴定 tdTomato+样品中突变的变体。例如，如果 WT 等位基因在某个坐标中是 G，则 tdTomato+细胞携带 A，并且 tdTomato-细胞携带 G，然后突变体 A 将被称为从头突变。然而，如果 tdTomato-细胞携带 A，则无法鉴定突变体。为了进一步验证脱靶 SNV 仅在 tdTomato+样品中鉴定，本发明人还称 tdTomato-样品中的变体与同一胚胎中的 tdTomato+样品作为对照，其中只有变体在 tdTomato-细胞中突变但是在 WT 中可以识别 tdTomato+细胞。

WGS 分析显示 Cas9-Tyr-A 和 Cas9-Tyr-B 组的 tdTomato-细胞中的低水平靶向编辑范围为 0-6.3%，这可能是由假阴性 FACS 分选引起的(已知发生在低水平)。因此，本发明人只考虑等位基因频率超过 10%的变体在后续分析中是可靠的。本发明人还标记了与 UCSC 重复区域和微卫星序列重叠的变体，或存在于 dbSNP(v138)和 MGP(v3)数据库中。所有测序数据都保存在 NCBI (SRA)。

为了验证目标效率，本发明人将 BAM 文件与 e-value 设置为 0.0001 的 on-target 进行了对比。利用已有的两种算法对 on-target 的潜在 off-target 进行预测(Cas-OFFinder (<http://www.rgenome.net/cas-offinder/>)和 CRISPOR (<http://crispor.tefor.net/>))。

使用 RefSeq 数据库用 annovar (version 2016-02-01)注释 SNV 和 indel。原癌基因和抑癌基因从 UniprotKB/Swiss-Prot 数据库中检索(2018-09)。本发明人从 CistromeDB 数据库下载了 5 个 ATAC-seq 文件，生物来源为胚胎并通过所有质量控制。检索到的五个数据集包括 CistromeDB IDs “79877” (GSM2551659)， “79976” (GSM2551677)， “80493” (GSM2535470)， “81049” (GSM2551664) 和 “81052” (GSM2551667)。基于染色体位置将脱靶位点定位到每个文件中的峰值区域，然后通过双侧 Wilcoxon 秩和检验将具有或不具有脱靶的峰区域相互比较。

35 9、模拟胚胎发育过程中的自发突变

为了估计从 2 细胞阶段到 E14.5 阶段的自发突变的数量, 考虑到 40 的平均测序覆盖率和 10% 的等位基因频率阈值, 在计算机模拟中发现单核苷酸突变。对于每一轮模拟, 鉴于 1.8×10^{-10} 的突变率和小鼠核基因组的大小(2,785,490,220bp), 本发明人考虑了从 2 细胞阶段到 16 细胞阶段的复制过程, 对于细胞 16 阶段后发生的突变, 考虑等位基因频率, 不会检测到。在每次复制期间, 每个细胞可以突变或不突变, 并且一旦发生突变, 分裂的细胞将继承突变。然后随机选择累积突变及其野生型等位基因进行测序, 深度为 40, 将所选突变总计为每轮的自发突变数, 并重复相同的过程 10,000 次。

10、Digenome-seq 分析

如前所述(32)进行了多重 Digenome-seq, 包括 Cas9-LacZ, Cas9-Pde6b, Cas9-Tyr-A 和 Cas9-Tyr-B。具体地, 根据生产商说明书用 TIANamp Genomic DNA Kit(Tiagen)从小鼠尾部纯化基因组 DNA。PCR 扩增每个基因的 sgRNA 靶位点包括侧翼的基因组区域, 并根据制造商的说明用 Universal DNA Purification Kit (Tiagen) 纯化 PCR 扩增产物。将 Cas9 蛋白(1 μ g)和 sgRNA(1 μ g)在室温下预孵育 10 分钟以形成 RNP 复合物。将 DNA(4 μ g)与 RNP 复合物在反应缓冲液中于 37 $^{\circ}$ C 温育 3 小时。随后加入 RNase A(100 μ g/ml)以除去 sgRNA 后, 再次用 Universal DNA 纯化试剂盒(Tiagen)纯化消化的 DNA。

使用 Illumina HiSeq X Ten 测序仪对文库进行 WGS, 并且在 30x 至 40x 的测序深度下进行 WGS。使用 Digenome-seq2 (<https://github.com/chizksh/digenome-toolkit2>) 计算鉴定 DNA 切割位点。通过编辑距离用 R 包“Biostrings”分类鉴定的体外切割位点并列。

11、统计分析

本研究采用 R version 3.5.1 (<http://www.R-project.org/>)进行所有统计分析。所有检验均为双侧检验, $P < 0.05$ 为差异有统计学意义。

实施例 1、三种基因编辑工具的评估

利用 GOT1 的方法, 用于评估三种常用的基因编辑工具, CRISPR-Cas9, 胞嘧啶碱基编辑器 3(BE3, rAPOBEC1-nCas9-UGI)和腺嘌呤碱基编辑器 7.10(ABE7.10, TadA-TadA*-nCas9)诱导的脱靶效应(参考文献 6-8)。

本发明人将 CRISPR-Cas9, BE3 或 ABE7.10 与 Cre mRNA 再加上相应的 sgRNA 一起注射到来自 Ai9(CAG-LoxP-Stop-LoxP-tdTomato)小鼠的二细胞胚胎的一个卵裂球中(参考文献 9-10)(图 1A), 使用 CRISPR/Cas9 或 BE3 基因编辑, 并结合 Cre mRNA 进行编辑。根据 tdTomato 在基因编辑细胞中的表达情况, 用荧光激活细胞分选法(FACS)将子细胞群中编辑与未编辑的细胞进行分选。tdTomato+和 tdTomato-分别进行全基因组测序。

利用流式细胞分选(FACS)分离 E14.5 天的胚胎, 根据细胞中 tdTomato 进行分选, 此时整个胚胎可以容易地被消化以获得足够的单细胞(图 1B)。Cas9-Tyr-A、Cas9-Tyr-B、BE3-Tyr-C、BE3-Tyr-D 处理 E14.5 胚胎的流式细胞仪分析如图 1D; 未注射胚胎的流式细胞分析如图 6b 所示。Cas9-Tyr-A、Cas9-Tyr-B、BE3-Tyr-C、BE3-Tyr-D 处理的 E14.5 胚胎, TA 克隆测序 On-target 分析如图 1E。本发明中基于全基因组测序 tdTomato+ 细胞(左)和 tdTomato-细胞(右)的靶向效率如图 1F。

本发明人进一步论证分别用 Cre 和 Cas9/BE3 系统处理的编辑细胞可以有效地与未编辑的细胞分离。在 Cre 介导的重组过程中, 约有 50%的胚胎细胞表达 tdTomato。荧光显微镜下观察 4 细胞期或 8 细胞期或流式细胞仪分析 E14.5 天细胞验证了这一点, 如图 6a-b。此外, 本发明人还发现用 CRISPR/Cas9 编辑毛色基因酪氨酸酶, 方法是 2 细胞胚胎中的一个卵裂球注射任意一个 sgRNAs(Cas9-Tyr-A, Cas9-Tyr-B), 实现了高效的靶向编辑。对 Tyr 基因测序显示, 从 4 个分散的 8 细胞胚胎中收集到 13 个(Cas9-Tyr-A)和 15 个(Cas9-Tyr-B)tdTomato+细胞, 85%和 80%细胞存在 Tyr 等位基因突变。相反, 收集到的 16 个(Cas9-Tyr-A)和 15 个(Cas9-Tyr-B)tdTomato-细胞没有 Tyr 等位基因突变, 如图 6c。

对分离的 tdTomato+和 tdTomato-细胞分别进行全基因组测序(WGS), 并通过三种算法对 tdTomato+样本鉴定 SNV 和 indel, 同时将来自于同一个胚胎的 tdTomato-样品作为参照。

本发明人还验证了本方法在靶向 Tyr 基因时的编辑效率。对胚胎注射的方法进行全基因组测序的探索, 设计了四个 sgRNAs 用于 CRISPR/Cas9 编辑, Cas9-Tyr-A, Cas9-Tyr-B 靶向 Tyr; 一个对照 sgRNAs 靶向在 C57 小鼠基因组中没有切割位点的 LacZ; 一个 sgRNA 靶向 Pde6b, 它与 C57 小鼠基因组存在一个错配, 且之前报道中可以产生大量的 SNV。通过 DNA 切割实验, 本发明人在体外验证了这些 sgRNAs 的切割效率。结果如图 7, 表明发生了有效的切割。

本发明人还通过 BE3 介导, 测试了两个靶向 Tyr 基因的 sgRNAs。作为对照组, 也包括三组胚胎注射 Cre only, Cre 和 Cas9, Cre 和 BE3。将 CRISPR/Cas9 或 BE3, Cre mRNAs 和 sgRNAs 的混合物注射入一个卵裂球中, 发现胚胎发育没有受到损伤, 如正常的囊胚率(图 8a)和存活率(图 8b)所示。Sanger 测序显示, 全部检测到的来自 Cas9-Tyr-A, Cas9-Tyr-B, BE3-Tyr-C 和 BE3-Tyr-D 的囊胚和 E14.5 胎儿存在 Tyr 突变(图 8c)。为了验证靶向 Tyr 基因的编辑效率, 本发明人对 Cas9-Tyr-A 或 Cas9-Tyr-B 处理的 E14.5 的胎儿进行 FACS 分选, 收集约 400k tdTomato+和 400k tdTomato-细胞, 并进行 TA 克隆测序。实验结果表明, 用 Cas9-Tyr-A 和 Cas9-Tyr-B 编辑的 tdTomato+细胞存在 50%和 100%的等位基因突变。第二个重复实验中, 用 Cas9-Tyr-A 和 Cas9-Tyr-B 编辑的 tdTomato+细胞存在 58%和 50%的等位基因突变(31 个和 14 个克隆)。相反, 从 E14.5 天胚胎中收集的 tdTomato-细胞没有 Tyr 等位基因突变。相似的, BE3 存在很高的编辑效

率，在 BE3-Tyr-C 有 71% 的等位基因突变以及在 BE3-Tyr-D 有 100% 的等位基因突变，而相应的 tdTomato-细胞仅有大约 3% Tyr 突变。这些结果证明，CRISPR/Cas9 和 BE3 编辑在 tdTomato+产生高效的目标编辑效率，但是在 tdTomato-细胞中基本上没有目标编辑效率。

5 为了进一步探索编辑效率和潜在的全基因组脱靶效应，本发明人对来自 18 个 E14.5 胚胎 9 种处理的 36 个样本进行平均深度为 47(47x)的全基因组测序：Cre only，Cre 和 Cas9，Cre 和 Cas9-LacZ，Cre 和 Cas9-Pde6b，Cre 和 Cas9-Tyr-A，Cre 和 Cas9-Tyr-B，Cre 和 BE3，Cre 和 BE3-Tyr-C，Cre 和 BE3-Tyr-D，其中只有 Cas9-Tyr-A，Cas9-Tyr-B，BE3-Tyr-C 和 BE3-Tyr-D 在 C57 基因组中存再编辑位点。对 Cas9-Tyr-A，Cas9-Tyr-B 进行 On-target 分析显示，在 tdTomato+细胞中分别有 56% 和 72% 的 Tyr 等位基因突变，表明在 Tyr 基因上有高效的 On-target 效率；同样地，BE3-Tyr-C 和 BE3-Tyr-D 在 tdTomato+细胞中均表现出高效编辑(分别为平均 75% 和 92% 的 Tyr 等位基因突变)，如图 9。本发明人也分析了其它 tdTomato+胚胎 On-target 效率，所有对照胚胎全都没有 On-target 效率。然而，全基因组测序分析显示 Cas9-Tyr-A 和 Cas9-Tyr-B 处理的 tdTomato-细胞有 15 0-6.3% 的低水平靶向编辑，这可能是假阴性流式细胞仪分选造成的(已知发生在较低水平)。因此，在接下来的分析中，只考虑等位基因频率超过 10% 的变异是可靠的。

为了评估脱靶效应，通过每个胚胎中同时使用三种不同的变异调用算法比较 tdTomato+细胞和 tdTomato-细胞，本发明人分析了全基因组的从头到尾变异，三种算法定义的变量均为真正的变量。在全部 9 组中发现只有 0-4 个 indel(图 10)，这个结果进一步由 Sanger 20 测序验证(图 11)。同时，在 Cre only 的胚胎中，观察到平均 14 个 SNV，用 CRISPR/Cas9 处理的胚胎(Cas9，Cas9-LacZ，Cas9-Pde6b，Cas9-Tyr-A 和 Cas9-Tyr-B)，各组平均 SNV 分别为 12.5、5、0、16、19，与“Cre-only”组比较，差异无统计学意义。此外，观察到 Cas9-Pde6b 编辑胚胎中 SNV 没有增加，这与之前的许多研究一致(图 12)。所有在 CRISPR/Cas9 编辑胚胎中检测到的脱靶 SNV 均经 Sanger 测序证实。在 Cre-或者 25 CRISPR/Cas9 处理的样品中检测出 SNV 可能是因为在基因复制过程中自发突变引起的，但是变异数量在自发突变范围内。此外，本发明人的研究中没有出现同一变异(图 13a)，与 Cas-OFFinder，CRISPOR 和 Digenome-seq 推测的脱靶位点没有重叠(图 13b)。本发明人也观察到，所识别的 SNV 邻近序列与靶向位点没有序列相似性(图 13c)。

另外，通过相反调用变量，比较每个胚胎的 tdTomato-和 tdTomato+样本，发现 SNV 的 30 数量相似，表明 CRISPR/Cas9 编辑没有产生脱靶，本发明人观察到的 SNV 来自自发突变(图 14)。

本发明人进一步共设计了 12 个组：一个 Cre 组(仅 Cre)，六个包括或者不包括 sgRNA 的 Cas9 组(Cas9，Cas9-LacZ，Cas9-Pde6b，Cas9-Tyr-A，Cas9-Tyr-B 和 Cas9-Tyr-C)，具有或不具有 sgRNA 的三个 BE3 组(BE3，BE3-Tyr-C，BE3-Tyr-D)(参 35 考文献 11)和具有或不具有 sgRNA 的两个 ABE 组(ABE7.10，ABE7.10-Tyr-E)。

本发明人通过 Sanger 测序验证了在 8 细胞和 E14.5 阶段胚胎的靶向效率。为了进一步探索目的位点的编辑效率和潜在的全基因组脱靶效应，本发明人对来自 23 个 E14.5 胚胎的 46 个样品进行了平均深度为 47x 的 WGS(表 1)。

表 1、HiSeq X Ten sequencing 的汇总

Sample	Group	Accession	Mapped bases (Gbp)	Coverage
Cre-#1	tdTomato+	SRS2549042	127.72	45.85
	tdTomato-	SRS2549043	130.14	46.72
Cre-#2	tdTomato+	SRS2549040	131.92	47.36
	tdTomato-	SRS2549031	131.91	47.36
Cas9-#1	tdTomato+	SRS2549032	124.06	44.54
	tdTomato-	SRS2549035	135.23	48.55
Cas9-#2	tdTomato+	SRS2604284	132.48	47.56
	tdTomato-	SRS2604285	132.56	47.59
Cas9-LacZ-#1	tdTomato+	SRS2549038	128.44	46.11
	tdTomato-	SRS2549039	119.22	42.80
Cas9-LacZ-#2	tdTomato+	SRS2604286	127.49	45.77
	tdTomato-	SRS2604287	140.58	50.47
Cas9-Pde6b-#1	tdTomato+	SRS3024198	127.20	45.67
	tdTomato-	SRS3024199	122.52	43.98
Cas9-Pde6b-#2	tdTomato+	SRS3024196	131.14	47.08
	tdTomato-	SRS3024197	135.07	48.49
Cas9-Tyr-A-#1	tdTomato+	SRS2549029	133.97	48.10
	tdTomato-	SRS2549030	130.04	46.69
Cas9-Tyr-A-#2	tdTomato+	SRS2549033	135.19	48.53
	tdTomato-	SRS2549034	116.56	41.85
Cas9-Tyr-B-#1	tdTomato+	SRS2549037	129.74	46.58
	tdTomato-	SRS2549036	132.69	47.64
Cas9-Tyr-B-#2	tdTomato+	SRS2549041	139.00	49.90
	tdTomato-	SRS2549028	134.09	48.14
Cas9-Tyr-C	tdTomato+		147.33	52.89
	tdTomato-		147.45	52.94
BE3-#1	tdTomato+	SRR8169137	123.84	44.69
	tdTomato-	SRR8169136	128.06	46.21
BE3-#2	tdTomato+	SRR8169139	117.97	42.58
	tdTomato-	SRR8169138	128.93	46.53
BE3-Tyr-C-#1	tdTomato+	SRR8169133	150.20	54.21
	tdTomato-	SRR8169132	149.12	53.81
BE3-Tyr-C-#2	tdTomato+	SRR8169135	150.08	54.16
	tdTomato-	SRR8169134	148.89	53.73
BE3-Tyr-D-#1	tdTomato+	SRR8169131	151.27	54.59
	tdTomato-	SRR8169130	151.62	54.72
BE3-Tyr-D-#2	tdTomato+	SRR8169141	143.01	51.61
	tdTomato-	SRR8169140	143.54	51.80
ABE7.10-#1	tdTomato+		133.22	47.83
	tdTomato-		115.20	41.36
ABE7.10-#2	tdTomato+		144.31	51.81
	tdTomato-		143.07	51.36
ABE7.10-Tyr-E-#1	tdTomato+		130.09	46.70
	tdTomato-		148.97	53.48
ABE7.10-Tyr-E-#2	tdTomato+		148.18	53.20
	tdTomato-		133.12	47.79

Cas9, BE3 和 ABE7.10 在 tdTomato+细胞中的活性分别通过靶向位点的高 indel 和高 SNV 比例得到了证实(图 1C; 表 2-3)。

表 2、每个胚胎中 SNV 和 indel 的 WGS 鉴定

Variants	Cre+		Cas9+										BE3+						ABE7.10+				
	- #1	- #2	- #1	- #2	LacZ- #1	LacZ- #2	Pde6b- #1	Pde6b- #2	Tyr-A- #1	Tyr-A- #2	Tyr-B- #1	Tyr-B- #2	Tyr-C	- #1	- #2	Tyr-C- #1	Tyr-C- #2	Tyr-D- #1	Tyr-D- #2	- #1	- #2	Tyr-E- #1	Tyr-E- #2
On-target mutations	0	0	0	0	0	0	0	0	1	1	1	1	2	0	0	1	1	1	1	0	0	1	3
Off-target SNVs	2	26	22	3	8	2	0	0	22	10	5	33	31	277	137	320	355	332	277	1	1	17	21
Off-target Indels	0	3	0	1	0	1	0	0	0	0	2	0	4	1	0	1	4	1	0	1	1	3	2
Exon off-target SNVs	0	0	0	0	0	0	0	0	1	0	0	4	1	3	4	3	6	6	4	0	0	0	0
Exon off-target Indels	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
Non-synonymous off-target SNVs	0	0	0	0	0	0	0	0	0	0	0	2	1	2	2	0	4	4	2	0	0	0	0
Frameshift off-target Indels	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

*The sgRNA for *Pde6b* has one mismatch with the C57 genome (3), so there was no on-target sites.

#Two types of on-target variants, shown in Fig. S4.

表 3

Indels	Mutant positions	Mutant	Mutect2 (Mut/Total reads)	Scalpel (Mut/Total reads)	Strelka (Mut/Total reads)	Manual realignment
Cas9-Tyr-A-#1	chr7:87438074	CCAAAAGAGGG (deletion)	16/36	11/29	15/34	15/33
Cas9-Tyr-A-#2	chr7:87438083	TCAT (insertion)	18/37	13/32	15/35	13/31
Cas9-Tyr-B-#1	chr7:87438048	GATAG (deletion)	14/43	12/32	11/29	12/34
Cas9-Tyr-B-#2	chr7:87438054	TGC (deletion)	13/23	10/23	11/22	11/22
Cas9-Tyr-C	chr7:87493149	CTTTGAAGGGGAA (deletion)	44/45	32/32	45/48	44/45
SNVs	Mutant positions	Mutant	Mutect2 (Mut/Total reads)	Lofreq (Mut/Total reads)	Strelka (Mut/Total reads)	Manual realignment
BE3-Tyr-C-#1	chr7:87493149	G>A	13/15	33/35	30/31	32/34
BE3-Tyr-C-#2	chr7:87493149	G>A	17/36	17/36	19/40	22/40
BE3-Tyr-D-#1	chr7:87492721;chr7:87492722	C>T; C>T	13/28; 29/29	15/28;	12/28; 28/28	15/34; 34/34
BE3-Tyr-D-#2	chr7:87492722	C>T	10/12	16/17	11/14	10/12
ABE7.10-Tyr-E-#1	chr7:87438041;chr7:87438042	G>A; G>A	7/34; 7/34	7/34; 7/34	7/34; 7/34	11/39; 11/39
ABE7.10-Tyr-E-#2	chr7:87438041;chr7:87438042; chr7:87438044; chr7:87438039	G>A; G>A; G>A; G>A	20/31; 20/31; 10/32; 9/31	19/29	14/29; 17/29; 10/32; 7/29	24/37; 21/37; 15/37; 9/37

5 对于脱靶效应，本发明人发现来自所有 12 个组的胚胎中仅有 0-4 个 indel(表 2 和 4)，并且它们都没有与预测的脱靶位点有重叠(表 5)。

表 4

Sample	Mutect2	Scalpel	Strelka	Mutect2 vs Scalpel			Overlap of 3 methods
				Mutect2	Strelka	Strelka	
Cre-#1	107	11490	4930	4	6	462	9
Cre-#2	118	8929	4665	6	4	379	3
Cas9-#1	98	10854	4378	6	0	357	0
Cas9-#2	64	10253	5703	6	2	434	1
Cas9-LacZ-#1	131	11941	4746	5	3	491	0
Cas9-LacZ-#2	57	9394	5338	2	3	398	1
Cas9-Pde6b-#1	137	12285	4687	3	4	443	0
Cas9-Pde6b-#2	125	12313	5397	7	4	505	0
Cas9-Tyr-A-#1	75	12348	5180	3	5	464	0
Cas9-Tyr-A-#2	81	11993	5480	3	4	471	0
Cas9-Tyr-B-#1	117	10659	4734	4	5	427	2
Cas9-Tyr-B-#2	70	9015	4791	2	0	447	0
Cas9-Tyr-C	287	21965	4539	13	14	828	4
BE3-#1	280	10654	3826	3	13	397	1
BE3-#2	269	10729	4176	1	10	432	0
BE3+Tyr-C-#1	289	14014	5502	9	9	607	1
BE3+Tyr-C-#2	259	14418	5111	7	9	606	4
BE3+Tyr-D-#1	273	14585	5510	4	13	590	1
BE3+Tyr-D-#2	268	12240	5240	4	7	518	0
ABE7.10-#1	284	53199	3662	25	5	1501	1
ABE7.10-#2	250	16468	3343	5	4	525	1
ABE7.10-Tyr-E-#1	283	90132	4684	30	7	2531	3
ABE7.10-Tyr-E-#2	238	32903	4378	16	5	1029	2

表 5

Sample-#1				
	Chr	Position	Digenome score	DNA sequence
Tyr-A_1 (On-target site)	chr7	87438083	295.585443	GCGAAGGCACCGCCCTCTTTTGG
Tyr-A_2	chr8	70679420	87.5901275	TGGTTCATGCACCCCCCTTAGG
Tyr-A_3	chr2	11906262	12.9829391	catgtatagcagtggtccagaag
Tyr-A_4	chr6	94012018	5.738594479	CTATGGGAGGAGGTAAC TAAGCG
Tyr-B_1	chr5	1.22E+08	39.98904854	AAGAGGGCGGTGCTAAGATGGGG
Tyr-B_2	chrX	1.12E+08	21.2247577	AGGTACATAGGCTTCATATCAGG
Tyr-B_3	chr11	1.14E+08	8.274157156	CCCATGGGGAACTCCTGGGGG
Tyr-B_4	chr11	3184621	8.198383346	ACAAGCAAGTGTGTGTCATAGG
Tyr-B_5	chr11	1.14E+08	6.681354696	CCCATGGGGAACTCCTGGGGG
Tyr-B_6	chrX	87640980	5.514465963	CAAAAGGAGCAATTTCCAATAGG
Tyr-B_7 (On-target site)	chr7	87438053	4.826363636	CCGATAGGTGCATTGGCTTCTGG
Tyr-B_8	chr1	23481074	4.209876693	ATATAAGTTAACATCCCCAAAAGG
Tyr-B_9	chr11	95292492	3.644424083	TATGGGTGTCTCTCTTTCTCC
Tyr-B_10	chr1	1.28E+08	3.544329556	CCCAAGACATGCACACCGATAGG
Tyr-B_11	chr6	68111031	2.614919835	caagaCATAAAACATACCTAAAg
LacZ_1	chr2	32395622	43.46541216	TTCGGCTTCGGGGCGGGGTCAAG
LacZ_2	chr13	54153138	37.98678846	TAATGGTGCTGACTGCTATGAGG
Pde6b_1	chr10	16088519	65.24995196	ATTACAATTAttatgceatag
Pde6b_2	chr1	88276189	5.989287063	CTACTGCATGTTAGGAAAGGCCG
Sample-#2				
	Chr	Position	Digenome score	DNA sequence
Tyr-A_1	chr8	70679420	100.166815	TGGTTCATGCACCCCCCTTAGG
Tyr-A_2 (On-target site)	chr7	87438083	80.04553734	GCGAAGGCACCGCCCTCTTTTGG
Tyr-A_3	chr2	32395622	52.38775481	AGAGGGCGGGGCCTTATAGTGGG
Tyr-A_4	chr10	16088519	48.26614325	catgaagccaaaacacatatag
Tyr-A_5	chr2	11906264	20.65930936	catgtatagcagtggtccagaag
Tyr-A_6	chr9	73142622	5.706386646	tcitcgggtgtgtctaaagacag
Tyr-A_7	chr6	94012018	4.735788874	CTATGGGAGGAGGTAAC TAAGCG
Tyr-B_1	chr5	1.22E+08	53.80789887	AAGAGGGCGGTGCTAAGATGGGG
Tyr-B_2 (On-target site)	chr7	87438053	48.19727891	CCGATAGGTGCATTGGCTTCTGG
Tyr-B_3	chr11	1.14E+08	12.7891659	CCCATGGGGAACTCCTGGGGG
Tyr-B_4	chrX	1.12E+08	8.13690641	AAAAAGGGTTTAGTTCTATAGG
Tyr-B_5	chr11	3184621	7.883665333	ACAAGCAAGTGTGTGTCATAGG
Tyr-B_6	chr16	24592641	7.196863075	CTATAGGCTTTGAACGTGCAGGG
Tyr-B_7	chr1	23481074	4.891318316	ATATAAGTTAACATCCCCAAAAGG
Tyr-B_8	chr15	88729863	2.849386317	ATTGGGCGACAGCACGCAATCCG
LacZ_1	chr13	54153138	18.86615566	TAATGGTGCTGACTGCTATGAGG
LacZ_2	chr17	57065755	4.891340168	AGAGGGTGTTCCTTCCCACGGG
Pde6b_1	chr4	69960267	7.637319157	ACCTTTGGGTCTTGGGAAGGATG

5 对于所有 Cas9 编辑的胚胎, 不同 Cas9 组之间没有显著的 SNV 差异(平均 12 个

SNV /胚胎), 并且与“Cre”组相比也没有明显的差异(平均 14 个 SNV /胚胎)(表 2)。

在 Cre 或 Cas9 处理的样品中检测到的 SNV 可能是由于发育期间基因组复制期间的自发突变引起的, 因为本发明人检测到的 SNV 的数目在模拟的自发突变的范围内, 并且相邻序列和靶向位点也没有序列相似性 (参考文献 12)。

5 令人惊讶的是, 本发明人在 BE3 编辑的胚胎中发现了平均 283 个 SNV /胚胎, 比在 Cre 或 Cas9 处理的胚胎中观察到的水平高至少 20 倍(图 2A 和表 2)。相比之下, ABE7.10 平均只产生 10 个 SNV /胚胎, 频率接近自发突变率(图 2A 和表 2)。本发明人进一步比较了“BE3 only”组中鉴定到的脱靶位点与 BE3-Tyr-C 或 BE3-Tyr-D 中的脱靶位点, 发现 sgRNA 的存在不会诱导产生更高的 SNV($P = 0.21$, Kruskal-Wallis 检
10 验)。此外, 这些突变是在 tdTomato+细胞而不是 tdTomato-细胞中特异性鉴定到的(参见方法, 表 6)。

表 6

SNVs	Mutect2	Lofreq	Strelka	Mutect2 vs Lofreq	Mutect2 vs Strelka	Lofreq vs Strelka	Overlap of 3 methods
Cre-#1	527	66	865	4	21	8	3
Cre-#2	379	109	1494	14	42	48	12
Cas9-#1	420	146	1161	7	29	48	7
Cas9-#2	416	107	1276	13	38	56	8
Cas9-LacZ-#1	634	80	1111	3	30	17	1
Cas9-LacZ-#2	604	68	1349	8	25	49	6
Cas9-Pde6b-#1	549	51	633	5	21	3	0
Cas9-Pde6b-#2	273	65	751	3	38	2	0
Cas9-Tyr-A-#1	3781	160	2057	47	374	104	36
Cas9-Tyr-A-#2	230	68	778	9	16	25	8
Cas9-Tyr-B-#1	549	91	1009	14	35	38	13
Cas9-Tyr-B-#2	1421	100	1391	16	106	51	14
BE3-#1	953	66	722	17	34	20	15
BE3-#2	968	75	807	23	43	24	19
BE3-Tyr-C-#1	602	106	1059	18	43	32	12
BE3-Tyr-C-#2	671	102	1019	24	42	35	19
BE3-Tyr-D-#1	667	136	1128	33	58	55	30
BE3-Tyr-D-#2	1261	64	1526	13	67	20	7

Indels	Mutect2	Scalpel	Strelka	Mutect2 vs Scalpel	Mutect2 vs Strelka	Scalpel vs Strelka	Overlap of 3 methods
Cre-#1	134	12372	4380	1	383	428	3
Cre-#2	125	9368	5162	2	7	6	0
Cas9-#1	177	10771	4342	14	4	393	1
Cas9-#2	83	9532	3975	11	6	394	2
Cas9-LacZ-#1	108	10849	4097	0	3	342	0
Cas9-LacZ-#2	68	10438	3686	3	5	317	1
Cas9-Pde6b-#1	255	4145	3335	8	7	256	0
Cas9-Pde6b-#2	215	3124	3079	7	6	255	0
Cas9-Tyr-A-#1	85	10913	4795	5	8	371	4
Cas9-Tyr-A-#2	78	8477	3953	4	2	459	1
Cas9-Tyr-B-#1	128	12457	4965	5	5	405	2
Cas9-Tyr-B-#2	79	10925	4751	4	5	387	1
BE3-#1	279	11847	4127	7	4	400	1
BE3-#2	280	12215	4434	4	2	440	1
BE3-Tyr-C-#1	240	14395	5223	4	10	545	1
BE3-Tyr-C-#2	264	15901	5518	5	7	617	0
BE3-Tyr-D-#1	291	14952	5487	2	8	606	1
BE3-Tyr-D-#2	263	12703	5431	4	6	517	1

15 E3 样本中检测到的脱靶在每一组中都没有重复, 并且在整个基因组中随机分布。然后本发明人将这些脱靶突变与 Cas-OFFinder 和 CRISPROR softwares 预测的所有潜在脱靶位点进行比较。不出意外的是, 这两种预测工具预测了大量的脱靶位点, 但是没有出

现在本发明人检测到的 SNV 中。除此之外，所鉴定的 SNV 邻近序列与 BE3 sgRNA 靶向位点之间没有序列相似性，而预测的脱靶点最多的位点与目标位点 BE3 序列相似。值得注意的是，尽管 BE3 编辑产生的 SNV 是独特的，但是突变类型与 APOBEC1 的突变类型是一致的。

5 值得注意的是，在 BE3 编辑的细胞中鉴定到的 SNV 超过 90% 都是从 G 突变为 A 或从 C 突变为 T，在 Cre-，Cas9-或 ABE7.10 处理的细胞中未观察到突变偏好性(图 2B 和 C，图 15)。这种突变偏好性与 APOBEC1 本身(参考文献 13)的偏好性相同，表明这些突变不是自发的，而是由 BE3 编辑诱导的。以前的研究表明，APOBEC 家族的几个成员(包括 APOBEC1)的作用需要单链 DNA 的存在(参考文献 14-16)。本发明人
10 的分析也显示 BE3 诱导的 SNV 在转录区域存在显著富集(图 3A)，尤其是在具有高表达的基因中(图 3B)。有趣的是，没有任何脱靶位点在不同的 BE3 编辑的胚胎中共享或与预测的脱靶位点重合(图 3C 和 D)。

据报道，DNA 的可结合性与基因编辑效率有关。因此本发明人通过评估 Cistrome 数据库中来自小鼠胚胎细胞的 ATAC-seq 数据集合，判断脱靶位点是否富集在开放染色
15 质区域。事实上，在具有混合 C57BL6/DBA2 背景的 E8.5 胚胎和 Cistrome 数据库中的四个高质量数据集中，脱靶位点显著富集于具有较高结合性的区域中(图 16)。

此外，在脱靶和目标位点之间未观察到序列相似性，而通过计算机预测的脱靶位点显示与 BE3 的靶向位点存在很高的序列相似性。因此，BE3 脱靶 SNV 是 sgRNA 非依赖性的并且可能由 APOBEC1 的过表达引起。

20 在 BE3 诱导产生的 1698 个 SNV 中，26 个位于外显子上，其中 14 个导致非同义变化。本发明人通过 PCR 成功扩增了 20 个，并通过 Sanger 测序证实了它们的存在(表 7)。

表 7

	Mutant	Type	Gene	Alt reads	Ref reads	Alt reads	Ref reads	Allele frequency	dbSNP	Repeats	PCR	Sanger sequence
BE3-#1												
chr2	p.V2987M/c.119964795G>A	exonic	Mga	11	20	0	39	35.48%			Y	Y
chr2	p.D419N/c.140158610C>T	exonic	Esfl	21	8	0	28	72.41%			Y	Y
chr4	p.L376L/c.128589747G>A	exonic	Zscan20	18	20	0	44	47.37%			Y	Y
BE3-#2												
chr15	p.F15F/c.80091438C>T	exonic	Syngri	13	19	0	36	40.63%			N	N
chr19	p.P184P/c.60756817G>C	exonic	Nanos1	6	30	0	40	16.67%			N	N
chr1	p.E488K/c.140507758G>A	exonic	Kcni2	8	31	0	28	29.51%			Y	Y
chr3	p.E59K/c.96708345C>T	exonic	Nudt17	14	23	0	41	37.84%			N	N
BE3-Tyr-C-#1												
chr11	p.F1507F/c.110030023G>A	exonic	Abca8a	12	23	0	35	34.29%			Y	Y
chr3	p.F314F/c.93826961C>T	exonic	Tdpoz3	24	22	0	48	52.17%		Y	Y	Y
chr7	p.Q21Q/c.127020229C>T	exonic	Prrt2	18	24	0	35	42.86%			Y	Y
BE3-Tyr-C-#2												
chr10	p.D627N/c.45158272G>A	exonic	Prep	22	21	0	49	51.16%			Y	Y
chr11	p.L29L/c.35833265C>T	exonic	Rars	17	23	0	40	42.50%			Y	Y
chr13	p.G230G/c.63545059C>T	exonic	Ptch1	27	17	0	33	61.36%			Y	Y
chr13	p.Q282X/c.104189738G>A	exonic	Trim23	21	23	0	41	47.73%			Y	Y
chr16	p.E3404K/c.15809689G>A	exonic	Prkdc	15	18	0	41	45.45%			Y	Y
chr1	p.Q202X/c.173462096C>T	exonic	Aim2	18	33	0	33	35.29%			Y	Y
BE3-Tyr-D-#1												
chr11	p.F311F/c.73354687C>T	exonic	Off20	25	15	0	27	62.50%			Y	Y
chr19	p.E33K/c.38396211G>A	exonic	Slc35g1	21	22	0	35	48.84%			N	N
chr1	p.F22F/c.60694502G>A	exonic	Wdr12	13	26	0	34	33.33%		Y	Y	Y
chr1	p.H346P/c.173683317A>C	exonic	Ifi208	7	58	1	48	10.77%	Y	Y	N	N
chr6	p.D401N/c.145862884C>T	exonic	Bhlhe41	17	9	0	38	65.38%			N	N
chr7	p.E421K/c.104265600C>T	exonic	Trim5	19	1	0	21	95.00%		Y	Y	Y
BE3-Tyr-D-#2												
chr14	p.L115L/c.73568707C>T	exonic	Sucla2	14	11	0	36	56.00%			Y	Y
chr2	p.E2105E/c.26460812C>T	exonic	Notch1	11	41	0	40	21.15%			Y	Y
chr2	p.E872K/c.28685723G>A	exonic	Tsc1	9	29	0	33	23.68%			Y	Y
chr8	p.E196K/c.11785830G>A	exonic	Arlgaf7	14	22	0	37	38.89%			Y	Y

26 个 SNVs 中, 14 个导致编码蛋白质的非同义变化, 2 个在 Trim23 和 Aim2 基因中导致提前终止。Trim23 编码一个 E3 泛素连接酶, 其功能障碍可导致肌营养不良。先前的研究报道了 Aim2 基因在先天免疫中起着重要作用, 是抵御病毒感染的基础。本发明人也发现 1 个 SNV 在原癌基因上, 13 个 SNVs 在抑癌基因上, 引起了对 BE3 编辑致癌风险的严重关注(图 16)。本发明人还发现 1 个 SNV 位于原癌基因, 13 个 SNV 位于抑癌基因上, 引起了对 BE3 编辑的致癌风险的担忧。本发明人猜测是否通过表达较低量的 BE3 可以降低这种风险, 然而, 使用较少量的 BE3 会逐渐降低目标位点编辑的效率(表 8)。

表 8

ID	Mutation	WT	Total	Frequency	Dose	sgRNA						
A1	8	7	15	53.33	50	Tyr-C	C17	0	9	9	0	2
A4	8	4	12	66.67	50	Tyr-C	C18	0	11	11	0	2
A6	11	4	15	73.33	50	Tyr-C	D2-1	11	2	13	84.62	50
A8	7	8	15	46.67	50	Tyr-C	D2-3	12	0	12	100	50
A9	10	1	11	90.91	50	Tyr-C	D2-6	10	4	14	71.43	50
A12	11	0	11	100	50	Tyr-C	D2-8	10	2	12	83.33	50
A13	12	3	15	80	50	Tyr-C	D2-9	15	0	15	100	50
A14	11	3	14	78.57	50	Tyr-C	D2-10	9	2	11	81.82	50
A16	6	8	14	42.86	50	Tyr-C	D2-11	7	5	12	58.33	50
A18	10	3	13	76.92	50	Tyr-C	D2-13	7	2	9	77.78	50
A19	9	5	14	64.29	50	Tyr-C	D10	8	2	10	80	50
G1	6	9	15	40	20	Tyr-C	H1	7	6	13	53.85	20
G2	1	13	14	7.14	20	Tyr-C	H2	9	5	14	64.29	20
G3	0	13	13	0	20	Tyr-C	H3	1	14	15	6.67	20
G4	2	12	14	14.29	20	Tyr-C	H4	3	12	15	20	20
G5	0	15	15	0	20	Tyr-C	H5	5	9	14	35.71	20
G6	5	10	15	33.33	20	Tyr-C	H6	4	10	14	28.57	20
G7	3	11	14	21.43	20	Tyr-C	H7	5	10	15	33.33	20
G8	4	9	13	30.77	20	Tyr-C	H8	4	10	14	28.57	20
G9	5	8	13	38.46	20	Tyr-C	H9	6	5	11	54.55	20
G10	4	9	13	30.77	20	Tyr-C	H10	11	4	15	73.33	20
G11	1	12	13	14.29	20	Tyr-C	E2-3	0	12	12	0	10
G12	3	9	12	25	20	Tyr-C	E2-5	2	10	12	16.67	10
B2	0	12	12	0	10	Tyr-C	E2-6	1	9	10	10	10
B3	4	9	13	30.77	10	Tyr-C	E2-7	8	2	10	80	10
B4	5	7	12	41.67	10	Tyr-C	E2-8	9	3	12	75	10
B7	0	13	13	0	10	Tyr-C	E2-9	6	6	12	50	10
B9	1	14	15	6.67	10	Tyr-C	E2-10	4	6	10	40	10
B10	0	12	12	0	10	Tyr-C	E2-11	1	10	11	9.09	10
B11	4	9	13	30.77	10	Tyr-C	E2-12	11	2	13	84.62	10
B12	1	12	13	7.69	10	Tyr-C	E2-13	1	11	12	8.33	10
B13	3	8	11	27.27	10	Tyr-C	E2-14	6	6	12	50	10
B14	0	12	12	0	10	Tyr-C	F2-9	2	9	11	18.18	2
C2	0	12	12	0	2	Tyr-C	F2-11	7	7	14	50	2
C3	1	8	9	11.11	2	Tyr-C	F3	2	11	13	15.38	2
C5	0	12	12	0	2	Tyr-C	F2-4	0	14	14	0	2
C7	1	13	14	7.14	2	Tyr-C	F2-5	4	8	12	33.33	2
C8	2	12	14	14.29	2	Tyr-C	F6	0	13	13	0	2
C10	0	13	13	0	2	Tyr-C	F8	0	13	13	0	2
C13	0	14	14	0	2	Tyr-C	F14	3	8	11	27.27	2
C14	0	15	15	0	2	Tyr-C	F15	1	10	11	9.09	2
C15	0	8	8	0	2	Tyr-C	F19	0	12	12	0	2
							F22	3	12	15	20	2
							F28	0	12	12	0	2

本发明的方法的一大优势是可在一只动物内对比编辑与未编辑的细胞，排除了遗传背景的区别。先前的研究中比较编辑与未编辑的动物时，由于遗传背景的不同混淆了。事实上，本发明人也将该方法用于一个已发表的数据集中，并发现在 CRISPR/Cas9 编辑的和未编辑的小鼠中平均有~1000 个 SNVs 和 ~100 indels。基于这一发现，本发明人认为同胞之间的差异是由于遗传变异而不是 CRISPR/Cas9 编辑的结果。此外，对任意两个不同胚胎之间的序列进行比较时，发现更多的 SNV(3706 ± 5232)和 indel (583 ± 762) ($n = 18$ pairs)，因为所使用的胚胎并非来自同一亲本。这些结果表明，即使它们有相同的亲本，由于小鼠之间存在大量的遗传变异，也很难在脱靶分析中找到一个完全空白对照对编辑过的小鼠与未编辑过的小鼠进行比较。

总而言之，本发明证明 GOT1 在研究基因编辑引起的脱靶效应方面的优势，即利用同一个胚胎的子细胞进行全基因组测序。本发明人也发现由 CRISPR/cas9 介导的基因编辑引起的不希望发生的 off-target 突变在小鼠胚胎中很少见，这支持以前的研究结果，基于 CRISPR/Cas9 的体内编辑不会导致显著的 SNVs 和 indels。然而，不能排除其他研究报道的大部分缺失或者大部分染色体易位。相比之下，本发明发现许多由 BE3 编辑引起的新 SNV，这提高了在治疗应用中的基础编辑方法的安全性。

本发明人发现 BE3 诱导产生了许多新的 SNV，这在以前的研究中没有报道。可能的解释是，本发明中 GOTI 可以检测来自单个基因编辑的卵裂球的细胞群，而之前的研究使用了大量的细胞池，其中编辑是不一样的，由于种群平均导致随机脱靶的信号丢失。与 BE3 不同，ABE7.10 诱导的 SNV 没有增加，可能是由于 TadA 缺乏 DNA 结合能力(参考文献 17)。BE3 的脱靶效应可能可以通过降低 APOBEC1 的 DNA 结合能力或使用不同形式的胞嘧啶脱氨酶得到解决。总之，GOTI 免于受不同个体之间的 SNP 干扰而用于检查各种基因编辑工具的脱靶效应。

实施例 2、APOBEC1 酶对于脱靶效应的影响作用

如上，单碱基编辑工具 BE3 会导致大量的单核苷酸脱靶变异(SNV)。本发明人预期这些脱靶变异是由 APOBEC1 的过表达及其与单链 DNA(ssDNA)的结合引起的。然而，单碱基基因编辑工具(BEs)已被广泛用于单个碱基的突变研究，并且有成为纠正病原突变的潜力。本实施例中，本发明人测试解决 BE3 的脱靶问题的可能性，从而使疾病相关的靶 Cs 能够被特异性地矫正。野生型 APOBEC1 蛋白序列如 SEQ ID NO: 1 所示。

为了进行 BE3 脱靶评估所构建的 BE2 系统如图 4a，其中包括 Apobec1 以及 SpnCas9 酶、UGI 酶，通过 16AA(序列为 SGSETPGTSESATPES(SEQ ID NO: 38))和 4AA(序列为 SGGs(SEQ ID NO: 39))两段肽进行连接。

本发明人首先减少注入胚胎的 BE3 mRNA 的量，并应用 GOTI 来检测脱靶变体。随着 BE3 的注射减少，相应地降低了靶向位点基因编辑的效率(图 4b)。但是脱靶 SNV 的数量没有显著降低(图 4c)。

作为替代方法，本发明人尝试对 Apobec1 蛋白上的 ssDNA 结合域进行突变，观测是否可降低 APOBEC1 的脱靶活性。本发明人根据先前的研究将相应的 BE3 相应氨基酸位点进行突变，并使用 GOTI 方法评估它们对靶向效率和脱靶效应的影响(图 4a)。

本发明人评估了基因编辑 Tyr-C 和 Tyr-D 靶位点不同突变的效率。首先使用 sgRNA-C 和 D 评估突变体 BE3，BE3-W90A(发生于 Apobec1 蛋白氨基酸序列第 90 位上)，BE3-W90F，BE3-R132E(发生于 Apobec1 蛋白氨基酸序列第 132 位上)，BE3-R126E(发生于 Apobec1 蛋白氨基酸序列第 126 位上)和 BE3-E63A(发生于 Apobec1 蛋白氨基酸序列第 63 位上)编辑活性。结果显示，与 BE3 相比 BE3-R126E 突变在两个靶位点上编辑效率相差不大。突变体 BE3-R126E 的活性也通过来自 WGS 的高靶向效率证实(图 4b)。但是，值得注意的是，与 BE3 相比，R126E 突变胚胎中的脱靶 SNV 的数量显著减少，并且与“仅 Cre”相比没有发现显著差异(图 4c)。此外，在 R126E 处理的两个胚胎中相差不大。检测到的 SNV 数量接近自发突变率，并且没有 SNV 与预测潜在的脱靶位点重叠，表明通过突变 Apobec1 126 精氨酸成为谷氨酸可显著降低 BE3 诱导的脱靶 SNV。

基于上述，本发明人首次揭示一种解决方案，通过在 APOBEC1 上进行突变，如进行 R126E 突变，来解决 BE3 诱导的脱靶效应。

本发明中建立的模块性表明 GOTI 适用于 APOBEC1 的其他突变版本或胞苷脱氨酶的新工程化版本的进一步解决方案。

5

实施例 3、突变优化研究

首先，本发明人将不同量的 BE3 mRNA(50 ng/ μ l 和 10 ng/ μ l)与 sgRNA-Tyr-C 或 sgRNA-Tyr-D 一起注入胚胎，并通过单细胞 Sanger 测序评估靶向效率。

10 本发明人发现，使用较少量的 BE3 可显著降低靶向效率($72.6 \pm 5.3\%$, 50 ng/ μ l; $12.6 \pm 2.9\%$, 10 ng/ μ l)。

然后本发明人应用 GOTI 方法进行全基因组脱靶评估，通过双细胞胚胎注射 (GOTI)进行全基因组脱靶分析，以检测 BE3-Tyr-D 处理的胚胎上的脱靶变体，并且发现两种不同水平(分别注射 50ng/nl 和 10ng/nl)的 BE3 mRNA 的脱靶 SNV 的数量没有变化。

15 本发明人接下来检查了 APOBEC1 的 DNA 结合域处的点突变是否会降低 BE3 的脱靶率。基于先前研究提出的 DNA 结合域，本发明人将各种点突变引入 BE3 体系中的 APOBEC1 的推定 DNA 结合域，并评估它们对靶向效率和脱靶率的影响(图 5a)。对于点突变 E63A, R126E 和 R132E, BE3 的碱基编辑效率在 Tyr 基因上的两个位点的靶向碱基编辑中评估，其中在 2 细胞小鼠胚胎中具有相应的 sgRNA Tyr-C 和
20 Tyr-D(图 5b)。本发明人发现，与野生型 BE3 相比，BE3-E63A 或 BE3-R132E 在 Tyr 上的编辑效率都显著降低，而 BE3-R126E 在两个靶位点都保持了高编辑效率。本发明人进一步证实，BE3-R126E 的 DNA 靶向活性与 HEK293T 细胞中另外三个位点上的 BE3 相似。有趣的是，BE3-R126E 的编辑窗口缩小了。

25 本发明人接下来进行 GOTI 以评估 BE3-R126E 在具有或不具有 sgRNA(BE3-R126E, BE3-R126E-Tyr-C 和 BE3-R126E-Tyr-D) 的三组，BE3-W90Y+R126E(YE1)-Tyr-C 和 BE3-W90F+R126E(FE1)-Tyr-C 中的靶向效率和脱靶频率。通过全基因组测序确认了靶向效率(图 5c)。值得注意的是，用 BE3-R126E 和 BE3-W90Y+R126E(YE1)处理的胚胎中脱靶 SNV 的数量从野生型 BE3 处理的胚胎中的 283 ± 2 (n=6)显著降低至 24 ± 8 ，接近于自发突变(图 5d)。此外，本发明人没有观察到
30 突变偏差和没有与预测的脱靶位点重叠的 SNV，表明 BE3R126E 诱导的脱靶 SNV 不存在。

本发明人进一步在 293T 细胞上检测 BE3-W90Y+R126E(YE1), BE3-R126E 在转录组水平脱靶情况。结果发现 BE3-R126E 可以显著降低 RNA 的脱靶。BE3-W90Y+R126E(YE1)可以完全消除 RNA 的脱靶(图 5e)。

35 图 5d-e 中，BE3(hA3A) 是使用人的 APOBECA3A(human APOBECA3A)取代了

BE3 上的 apobec1, 构建的新的 BE3 编辑工具。BE3(hA3AY130F)是在 human APOBECA3A 突变 Y130 转变为 F, 可以观测到这一突变显著减少脱靶 SNV 的数量。

总之, 通过应用 GOTI 方法评估脱靶 SNV 的量, 能够证明通过突变基础编辑器的脱氨酶的假定 ssDNA 结合域可以消除胞嘧啶碱基编辑器的 DNA 和 RNA 水平脱靶效应。

该结果表明, 基础编辑器可以被设计成用于基因编辑和治疗应用的有效且安全的工具。

在本发明提及的所有文献都在本申请中引用作为参考, 就如同每一篇文献被单独引用作为参考那样。此外应理解, 在阅读了本发明的上述讲授内容之后, 本领域技术人员可以对本发明作各种改动或修改, 这些等价形式同样落于本申请所附权利要求书所限定的范围。

参考文献:

1. G. J. Knott, J. A. Doudna, CRISPR-Cas guides the future of genetic engineering. *Science* **361**, 866-869 (2018).
2. S. Q. Tsai, J. K. Joung, Defining and improving the genome-wide specificities of CRISPR-Cas9 nucleases. *Nat Rev Genet* **17**, 300-312 (2016).
3. C. R. Lazzarotto *et al.*, Defining CRISPR-Cas9 genome-wide nuclease activities with CIRCLE-seq. *Nat Protoc* **13**, 2615-2642 (2018).
4. K. R. Anderson *et al.*, CRISPR off-target analysis in genetically engineered rats and mice. *Nat Methods* **15**, 512-514 (2018).
5. D. Kim *et al.*, Genome-wide target specificities of CRISPR RNA-guided programmable deaminases. *Nat Biotechnol* **35**, 475-480 (2017).
6. T. I. Cornu, C. Mussolino, T. Cathomen, Refining strategies to translate genome editing to the clinic. *Nature Medicine* **23**, 415-423 (2017).
7. H. A. Rees, D. R. Liu, Base editing: precision chemistry on the genome and transcriptome of living cells. *Nat Rev Genet*, (2018).
8. N. M. Gaudelli *et al.*, Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature* **551**, 464-471 (2017).
9. L. Madisen *et al.*, A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. *Nat Neurosci* **13**, 133-140 (2010).
10. L. Wang *et al.*, CRISPR-Cas9-mediated genome editing in one blastomere of two-cell embryos reveals a novel Tet3 function in regulating neocortical development. *Cell Res* **27**, 815-829 (2017).
11. K. Kim *et al.*, Highly efficient RNA-guided base editing in mouse embryos. *Nat Biotechnol* **35**, 435-437 (2017).
12. J. W. Drake, B. Charlesworth, D. Charlesworth, J. F. Crow, Rates of spontaneous mutation. *Genetics* **148**, 1667-1686 (1998).
13. A. C. Komor, Y. B. Kim, M. S. Packer, J. A. Zuris, D. R. Liu, Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420-424 (2016).
14. R. S. Harris, S. K. Petersen-Mahrt, M. S. Neuberger, RNA editing enzyme APOBEC1 and some of its homologs can act as DNA mutators. *Mol Cell* **10**, 1247-1253 (2002).
15. S. Rebhandl, M. Huemer, R. Greil, R. Geisberger, AID/APOBEC deaminases and cancer. *Oncoscience* **2**, 320-333 (2015).
16. L. B. Alexandrov *et al.*, Signatures of mutational processes in human cancer. *Nature* **500**, 415-421 (2013).
17. H. C. Losey, A. J. Ruthenburg, G. L. Verdine, Crystal structure of *Staphylococcus aureus* tRNA adenosine deaminase TadA in complex with RNA. *Nat Struct Mol Biol* **13**, 153-159 (2006).
18. S. Jin *et al.*, Cytosine, but not adenine, base editors induce genome-wide off-target mutations in rice. *Science*, in press (2019).
19. Y. B. Kim *et al.*, Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. *Nat Biotechnol* **35**, 371-376 (2017).
20. X. Wang *et al.*, Efficient base editing in methylated regions with a human APOBEC3A-Cas9 fusion. *Nat Biotechnol* **36**, 946-949 (2018).
21. J. M. Gehrke *et al.*, An APOBEC3A-Cas9 base editor with minimized bystander and off-target activities. *Nat Biotechnol* **36**, 977-982 (2018).

权 利 要 求

1. 一种降低单碱基编辑器的脱靶效应的方法，包括：在单碱基编辑器系统中，对其中的胞嘧啶脱氨酶进行改造，弱化其与 DNA 的结合能力。

2. 如权利要求 1 所述的方法，其特征在于，所述的改造是针对胞嘧啶脱氨酶的 DNA 结合区域的改造；较佳地，所述的 DNA 结合区域为其与 DNA 结合的结构域。

3. 如权利要求 2 所述的方法，其特征在于，所述的改造包括但不限于：基因突变，靶向性封闭或阻滞，干扰。

4. 如权利要求 1 所述的方法，其特征在于，所述的单碱基编辑器系统为 BE3 基因编辑器系统；或

所述的 DNA 为单链 DNA 或双链 DNA。

5. 如权利要求 1 所述的方法，其特征在于，所述的胞嘧啶脱氨酶包括但不限于选自下组的酶：AID，APOBEC3G，APOBEC1，APOBECA3A，CDA1。

6. 如权利要求 5 所述的方法，其特征在于，所述的胞嘧啶脱氨酶为 APOBEC1；较佳地，所述的改造为对该酶的第 126 位的氨基酸进行改造；更佳地，所述的改造为将其第 126 位的 R 突变为 E。

7. 如权利要求 6 所述的方法，其特征在于，APOBEC1 的改造还包括：对 APOBEC1 酶的第 90 位的氨基酸进行改造；更佳地，所述的改造为将其第 90 位突变为 Y。

8. 如权利要求 5 所述的方法，其特征在于，所述的胞嘧啶脱氨酶为 APOBECA3A，所述的改造为对该酶的第 130 位的氨基酸进行改造；较佳地，所述的改造为将其第 130 位的 Y 突变为 F。

9. 一种胞嘧啶脱氨酶的突变体，其 DNA 结合区域发生改造，该改造使得其与 DNA 的结合被弱化，所述的 DNA 如单链 DNA；较佳地，所述的胞嘧啶脱氨酶包括但不限于选自下组的酶：AID，APOBEC3G，APOBEC1，APOBECA3A，CDA1。

10. 如权利要求 9 所述的突变体，其特征在于，所述的酶为 APOBEC1；较佳地，所述改造发生在该结构域的第 126 位氨基酸上或其邻近的位置上；更佳地，所述的改造为将其第 126 位的 R 突变为 E。

11. 如权利要求 9 所述的突变体，其特征在于，所述的改造还发生在 APOBEC1 酶的第 90 位氨基酸上，更佳地，所述的改造为将其第 90 位突变为 Y。

5 12. 如权利要求 9 所述的突变体，其特征在于，所述的酶为 APOBECA3A，所述改造发生在该酶的第 130 位的氨基酸上或其邻近的位置上；更佳地，所述的改造为将其第 130 位的 Y 突变为 F。

10 13. 分离的多核苷酸，其特征在于，所述的多核苷酸编码权利要求 9-12 任一所述的突变体。

14. 一种单碱基编辑器，其包含权利要求 9-12 任一所述的胞嘧啶脱氨酶的突变体；更佳地所述的编辑器为 BE3 单碱基编辑器。

15 15. 权利要求 9-13 任一所述的胞嘧啶脱氨酶突变体的用途，用于参与基于单碱基编辑器系统的基因编辑，以降低基因编辑器的脱靶效应。

16. 一种筛选对于降低单碱基编辑器的脱靶效应有用的物质的方法，包括：

20 (1) 用候选物质处理胞嘧啶脱氨酶或其 DNA 结合结构域以及 DNA 相互作用的体系；和

(2) 检测所述体系中胞嘧啶脱氨酶 DNA 结合结构域与 DNA 相互作用情况；其中，若所述候选物质抑制、阻滞或下调胞嘧啶脱氨酶或其 DNA 结合结构域与 DNA 的相互作用，则表明该候选物质是对于降低基因编辑器的脱靶效应有用的物质。

25 17. 一种分析单碱基基因编辑工具或基因编辑操作的靶向效应的方法，其特征在于，所述方法包括：

(1) 获取 n-细胞阶段的胚胎，在其中 1~n-1 个细胞中进行基因编辑操作；其中所述 n 为 2~10 的正整数；

(2) 在胚胎发育的下游阶段，观测发生基因编辑的情况。

30 18. 如权利要求 17 所述的方法，其特征在于，步骤(1)中，所述的 n 为 2~8，2~6 或 2~4 的正整数；更佳地，所述 n 为 2。

35 19. 如权利要求 17 所述的方法，其特征在于，所述的方法为体外培育方法或体内培育方法。

20. 如权利要求 17 所述的方法，其特征在于，步骤(2)中，所述胚胎发育的下游阶段为胚胎发育到原肠胚阶段至出生前阶段，或在体内状态下为胚胎植入子宫至出生前的阶段。

5

21. 如权利要求 17 所述的方法，其特征在于，所述的胚胎为小鼠的胚胎，所述胚胎发育的下游阶段为胚胎发育的第 8~20 天(E8-E20 阶段)，较佳地为胚胎发育的第 9.5~18.5 天(E9.5-E18.5 阶段)，更佳地为胚胎发育的第 12~16(E11-E16 阶段，如 E14.5)。

10

22. 如权利要求 17 所述的方法，其特征在于，在胚胎发育的卵裂期，将胚胎中基因编辑的卵裂球和没有编辑的卵裂球进行拆分，移植到受体中，发育成成体。

23. 如权利要求 22 所述的方法，其特征在于，基因编辑的卵裂球和没有编辑的卵裂球形成分别的胚胎，移植到不同的受体或同一受体中，或体外建立胚胎干细胞系。

15

24. 如权利要求 17 所述的方法，其特征在于，所述的基因编辑操作包括(但不限于)：CRISPR 介导的基因编辑，Base Editor(碱基编辑器)介导的基因编辑，Cre/loxP 介导的基因编辑，Prime editor。

20

25. 如权利要求 24 所述的方法，其特征在于，所述的 CRISPR 介导的基因编辑包括(但不限于)：CRISPR/Cas9 介导的基因编辑，CRISPR/Cas9n 介导的基因编辑，CRISPR/Cas13(如 CRISPR/Cas13a，CRISPR/Cas13d)介导的基因编辑，CRISPR/CasRx 介导的基因编辑。

25

26. 如权利要求 24 所述的方法，其特征在于，所述的 Base Editor 包括：BE1，BE2，BE3，BE4，BE4-Max。

27. 如权利要求 24 所述的方法，其特征在于，所述的腺嘌呤碱基编辑器包括：ABE7.10，ABE 6.3，ABE 7.8，ABE 7.9，Prime Editing。

30

28. 如权利要求 17 所述的方法，其特征在于，步骤(1)中，包括：将用于切割核酸(如 DNA)靶位点的酶(如 Cas mRNA，Cre mRNA)，以及相应导向序列(如 sgRNA)引入到所述的其中一个细胞中，进行基因编辑。

35

29. 如权利要求 28 所述的方法，其特征在于，所述的用于切割核酸(如 DNA)靶位点的酶选自但不限于下组的酶：Cas9，Cas9n，Cas13a，CasRx，BE1，BE2，BE3，BE4，ABE7.10，ABE 6.3，ABE 7.8，ABE 7.9。

5 30. 如权利要求 17 所述的方法，其特征在于，步骤(1)中，采用可检测标记物来标记所述的基因编辑操作，在其中 1~n-1 个细胞中进行基因编辑操作并设置可检测标记物。

10 31. 如权利要求 30 所述的方法，其特征在于，所述的可检测标记物包括但不限于：染色标记物、荧光信号分子、报告基因；更佳地，所述的可检测标记物为 tdTomato，EGFP，mCherry，GFP，dsred。

15 32. 如权利要求 17 所述的方法，其特征在于，步骤(2)中，所述观测发生基因编辑的情况包括：

分选发生基因编辑的细胞以及未发生基因编辑的细胞；

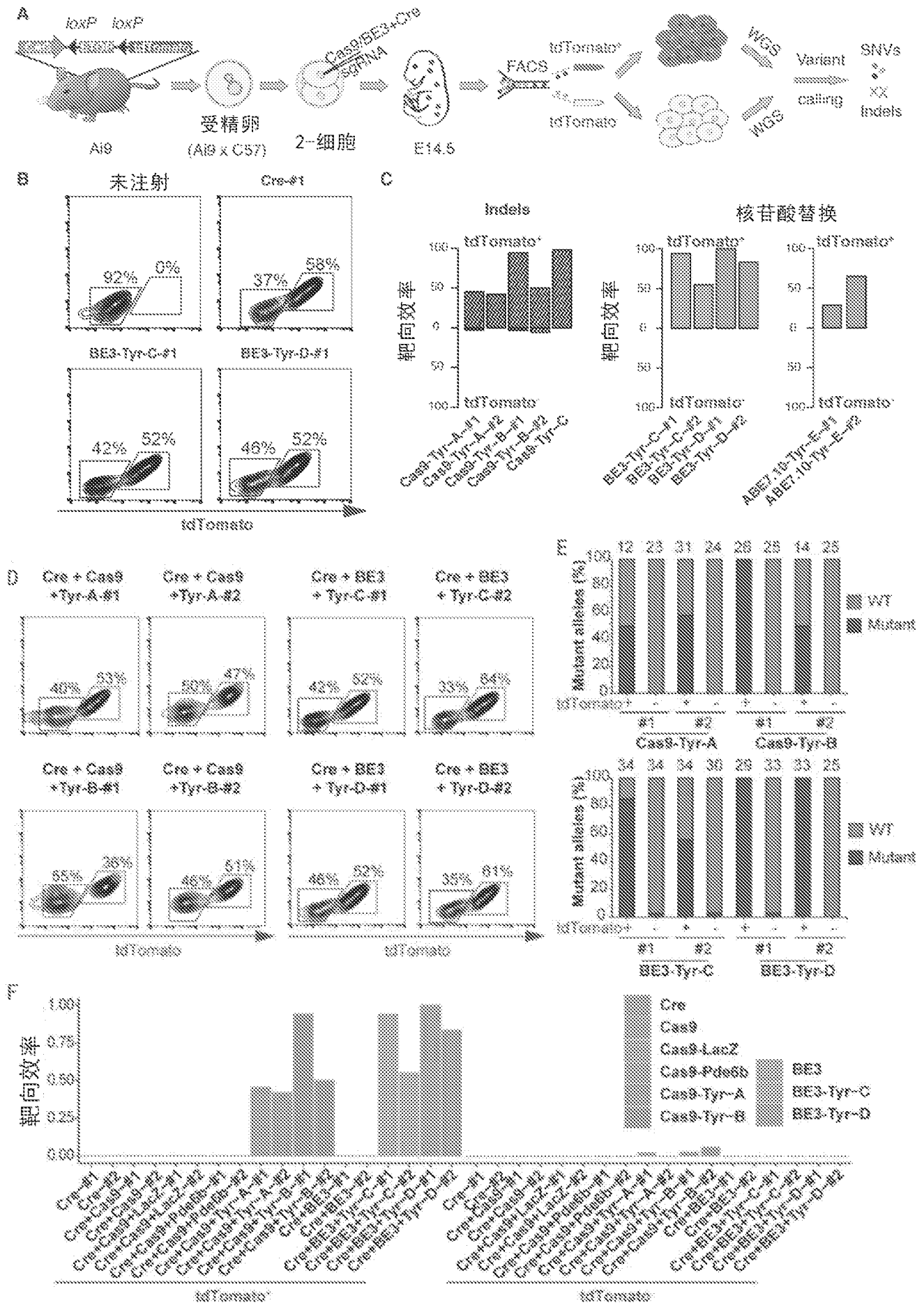
进行测序分析；

利用单核苷酸变异分析工具和/或 indel 分析工具进行分析；

20 比较发生编辑的细胞和未发生编辑的细胞来鉴定靶向效应或脱靶效应，包括检测 SNV 和 indel。

33. 如权利要求 32 所述的方法，其特征在于，所述的单核苷酸变异分析工具包括但不限于：Mutect2，Lofreq 和 Strelka 或其组合；或，所述的 indel 分析工具包括但不限于：Mutect2，Scalpel，Strelka 或其组合。

25 34. 如权利要求 17 所述的方法，其特征在于，所述的胚胎来自于哺乳动物，包括非人哺乳动物。



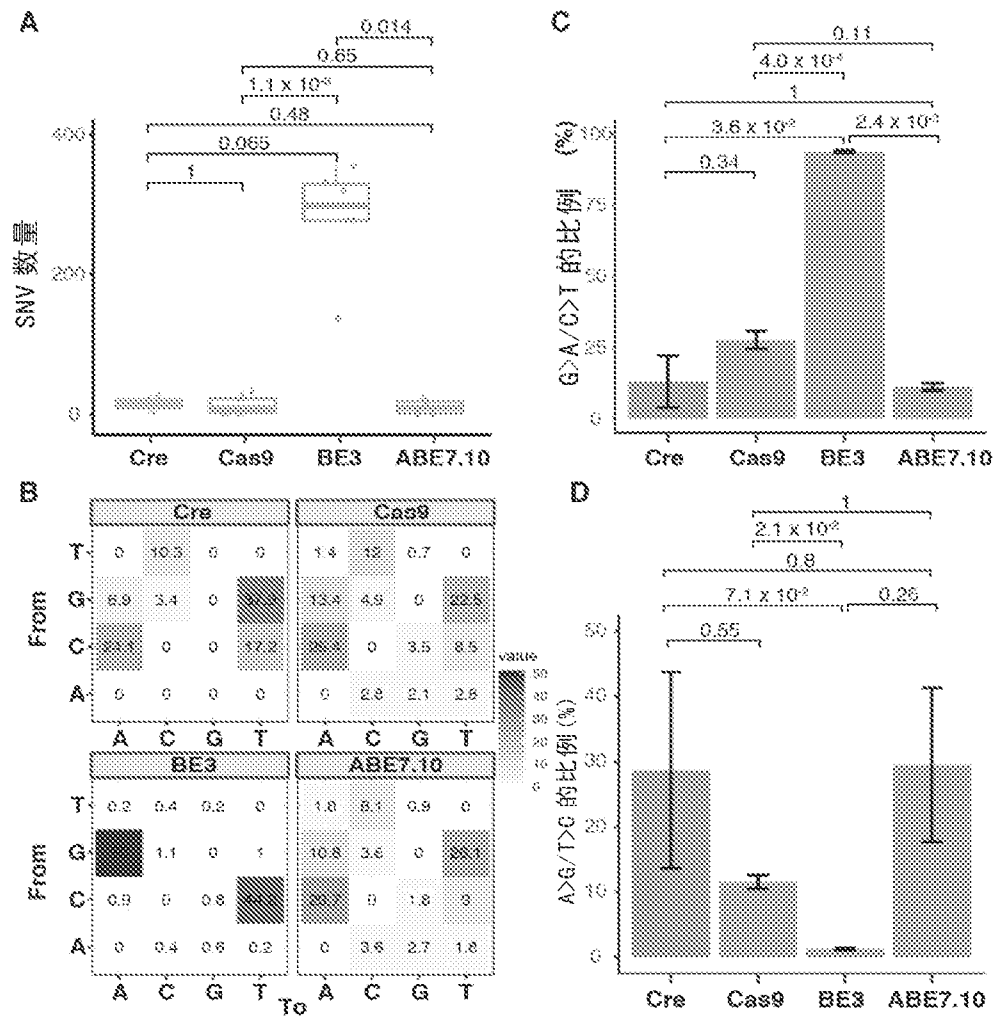


图 2

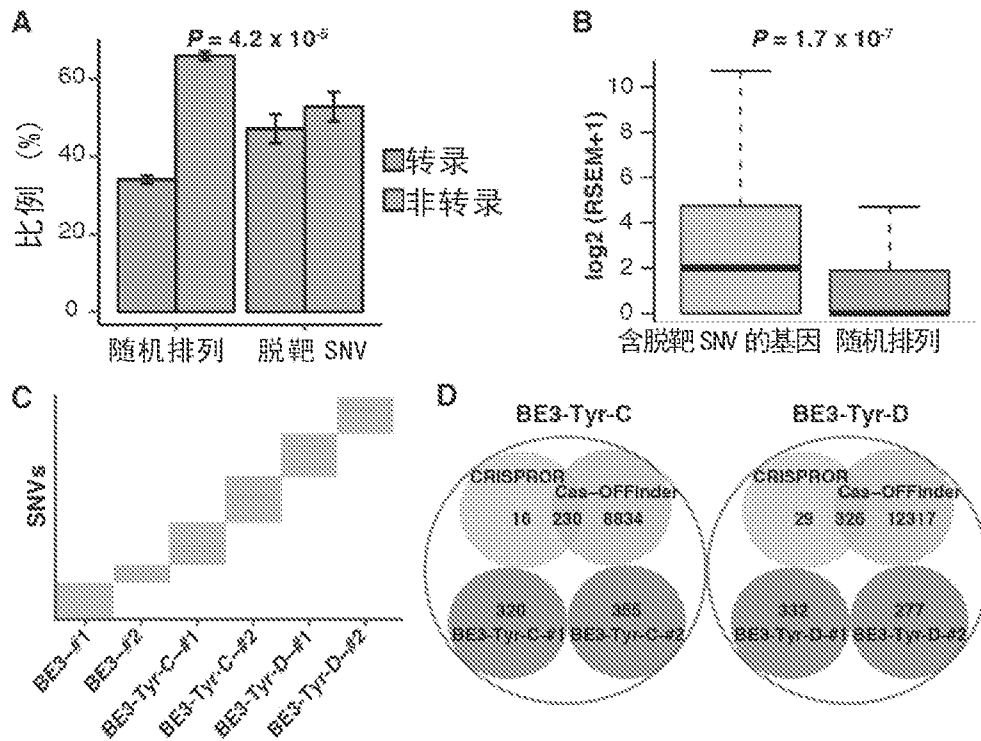
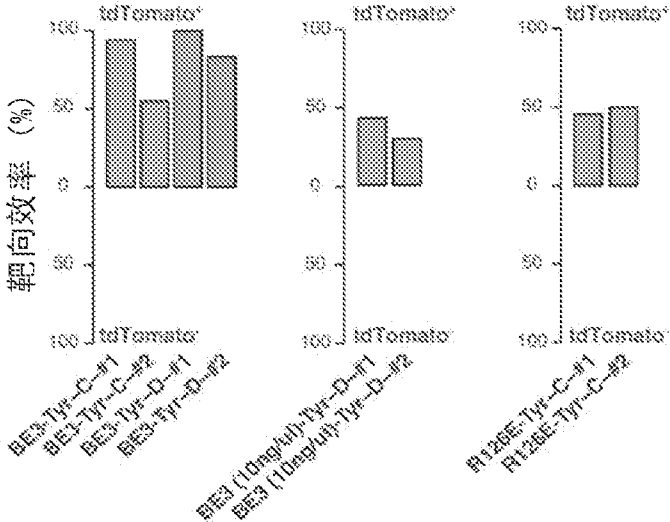


图 3

a

Proportion (%)	100%	50.0%	50.0%
Proportion (%)	100%	50.0%	50.0%
Proportion (%)	100%	50.0%	50.0%
Proportion (%)	100%	50.0%	50.0%
Proportion (%)	100%	50.0%	50.0%
Proportion (%)	100%	50.0%	50.0%

b



c

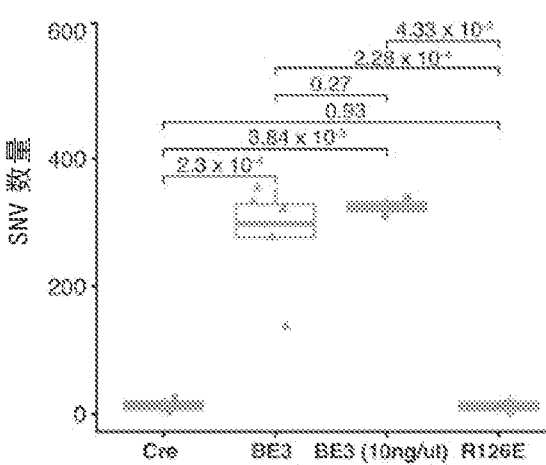
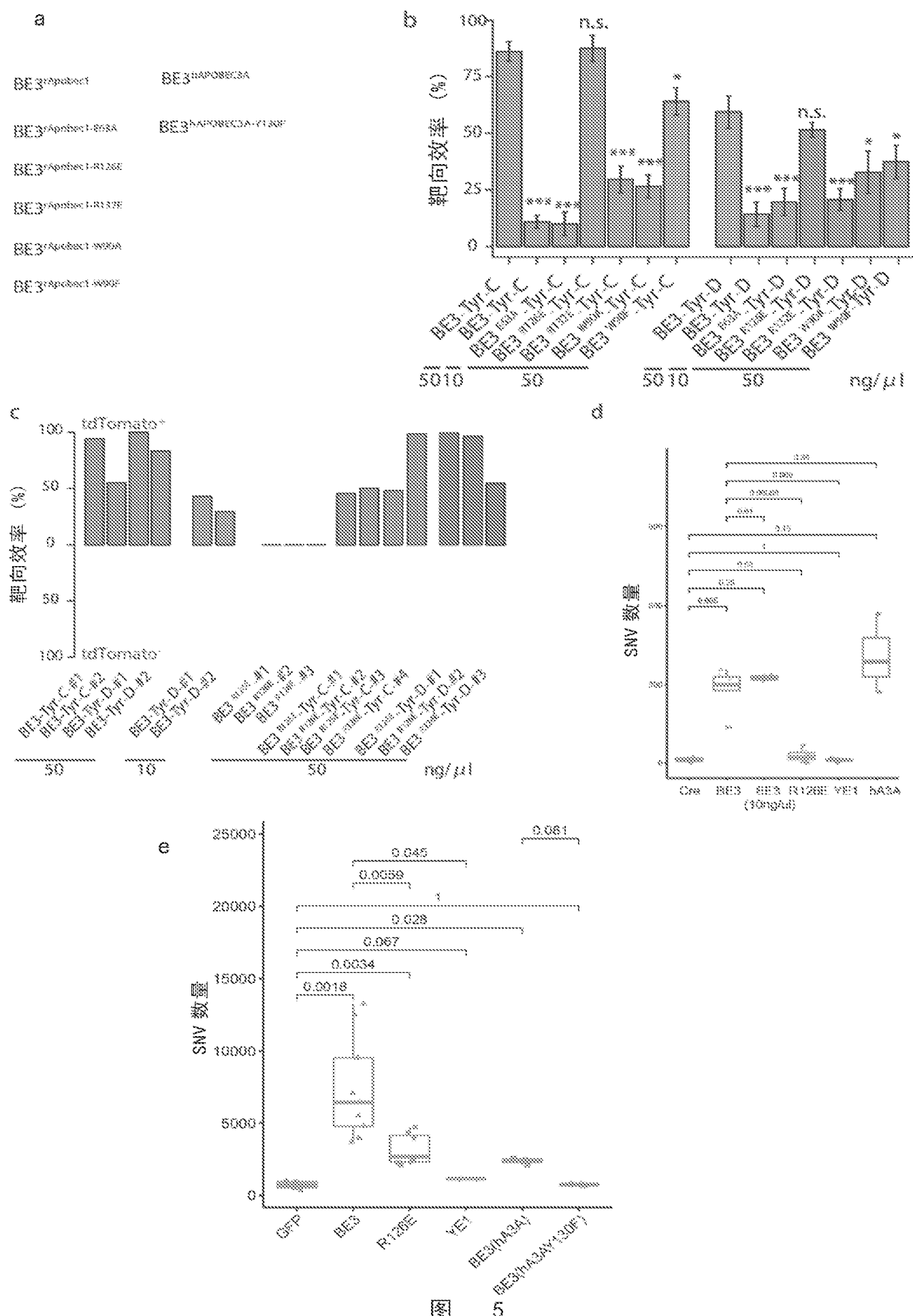


图 4



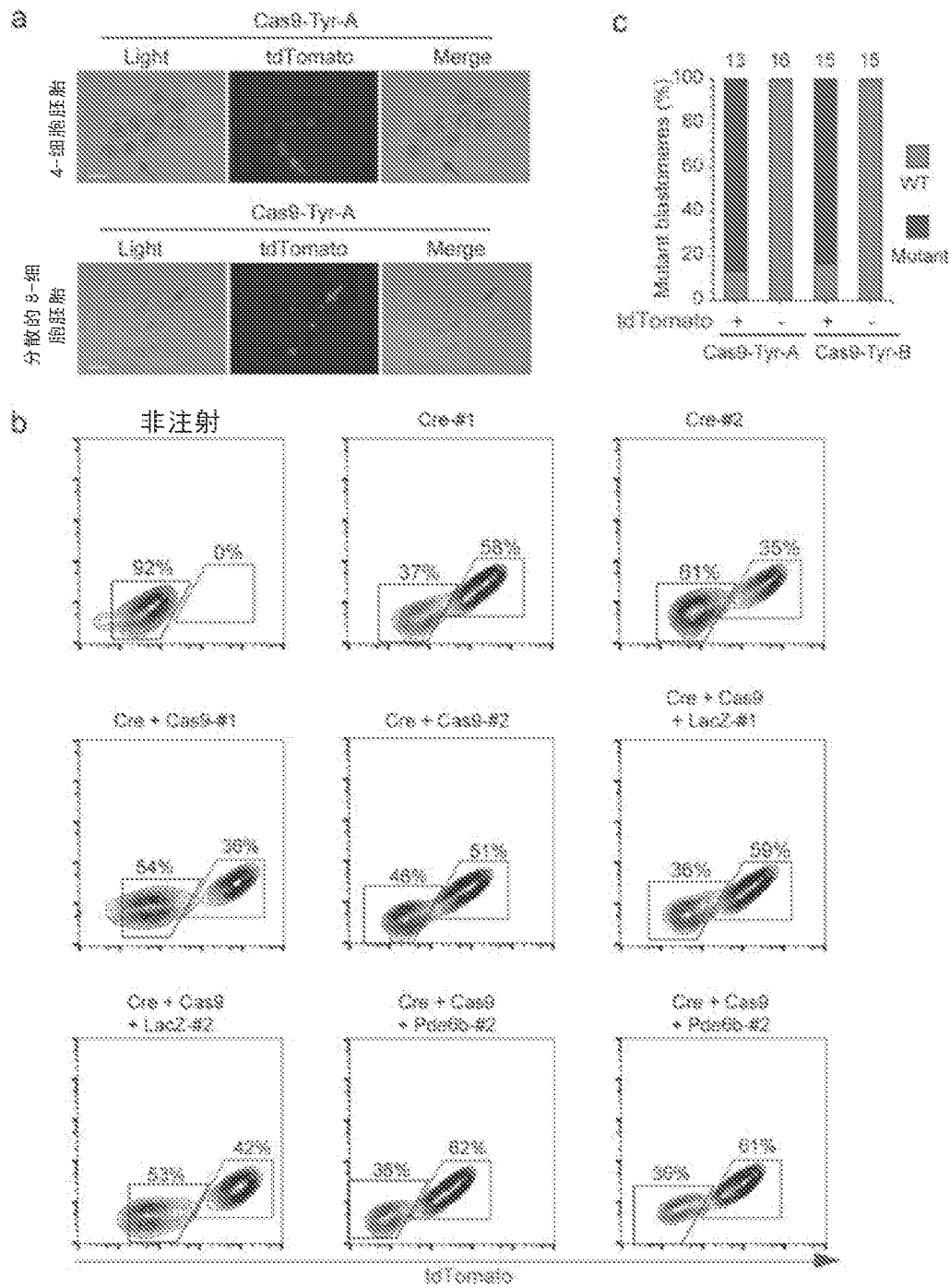


图 6

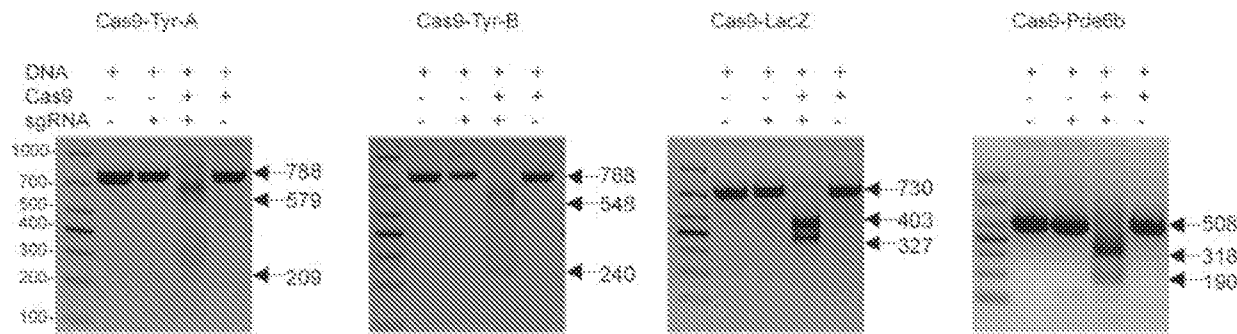


图 7

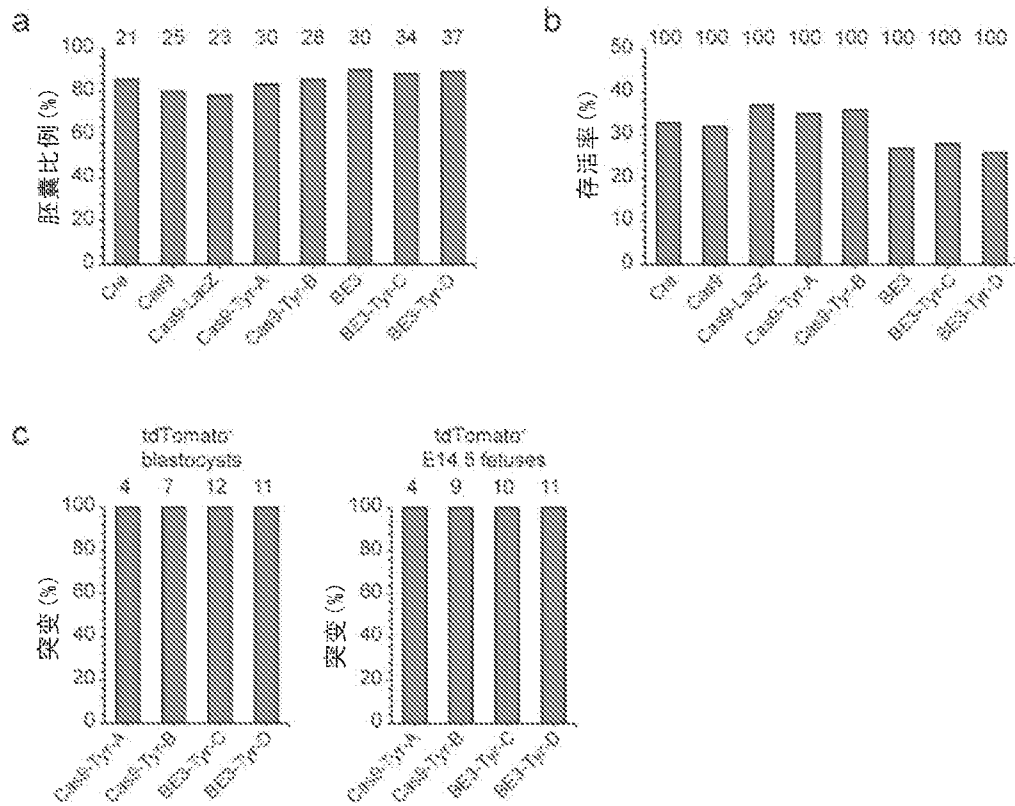


图 8

		tdTomato ⁺				tdTomato ⁺	
Cas9-Tyr-A	#1	GC GAAGGCACCGCCCTCTTTTGG	WT (18/38)	Cas9-Tyr-A	#1	GC GAAGGCACCGCCCTCTTTTGG	WT (27/38)
	#2	GC GAAGGCACCG.....	-14bp (18/38)		#2	GC GAAGGCACCG.....	-14bp (1/38)
Cas9-Tyr-B	#1	CCCAGAAGCCAATGCACCTATCGG	WT (3/34)	Cas9-Tyr-B	#1	CCCAGAAGCCAATGCACCTATCGG	WT (31/32)
	#2	CCCAGAAGCCAATGCACCTATCGG	-5bp (12/34)		#2	CCCAGAAGCCAATGCACCTATCGG	-5bp (1/32)
BES-Tyr-C	#1	GACCTCAGTTCCCTTCAAAGGG	WT (2/34)	BES-Tyr-C	#1	GACCTCAGTTCCCTTCAAAGGG	WT (40/40)
	#2	GACCTCAGTTCCCTTCAAAGGG	MUT (22/34)		#2	GACCTCAGTTCCCTTCAAAGGG	WT (50/50)
BES-Tyr-D	#1	CTGTGCCAAGGCAGAAACCCCTGG	WT (3/34)	BES-Tyr-D	#1	CTGTGCCAAGGCAGAAACCCCTGG	WT (47/47)
	#2	CTGTGCCAAGGCAGAAACCCCTGG	MUT (19/34)		#2	CTGTGCCAAGGCAGAAACCCCTGG	WT (32/32)

图 9

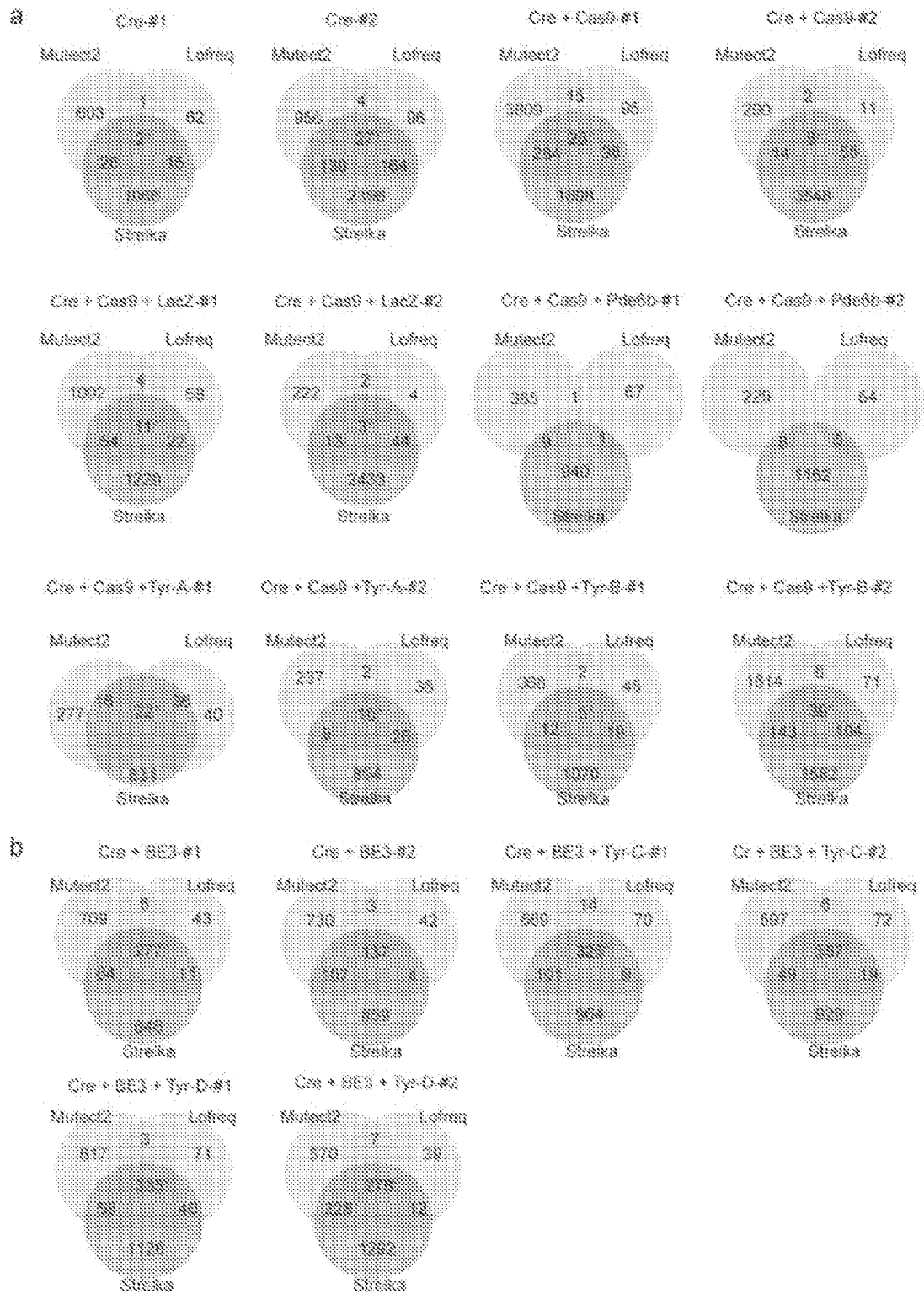
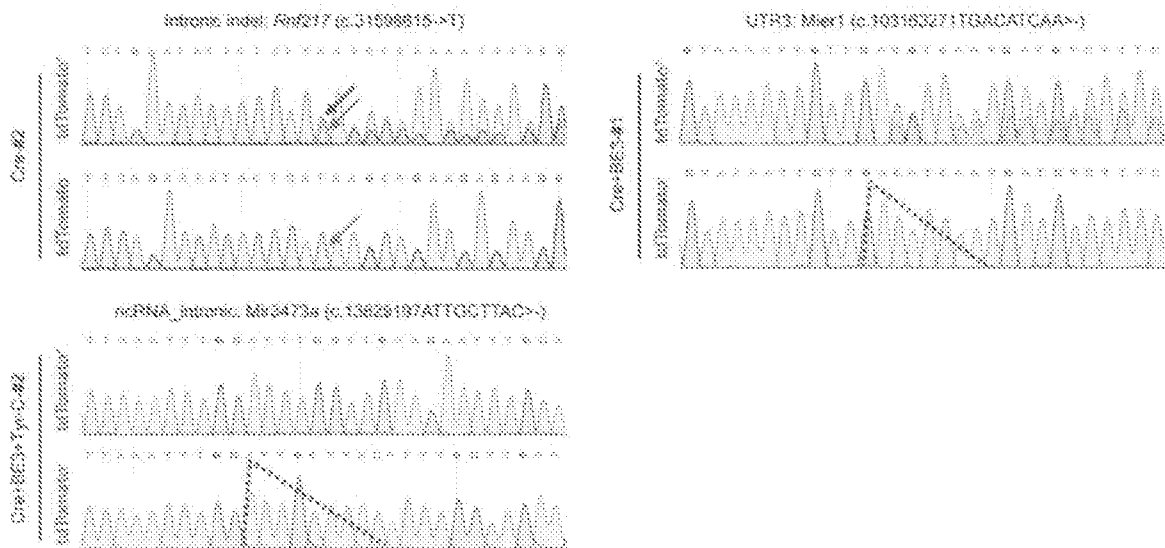


图 10

a



b

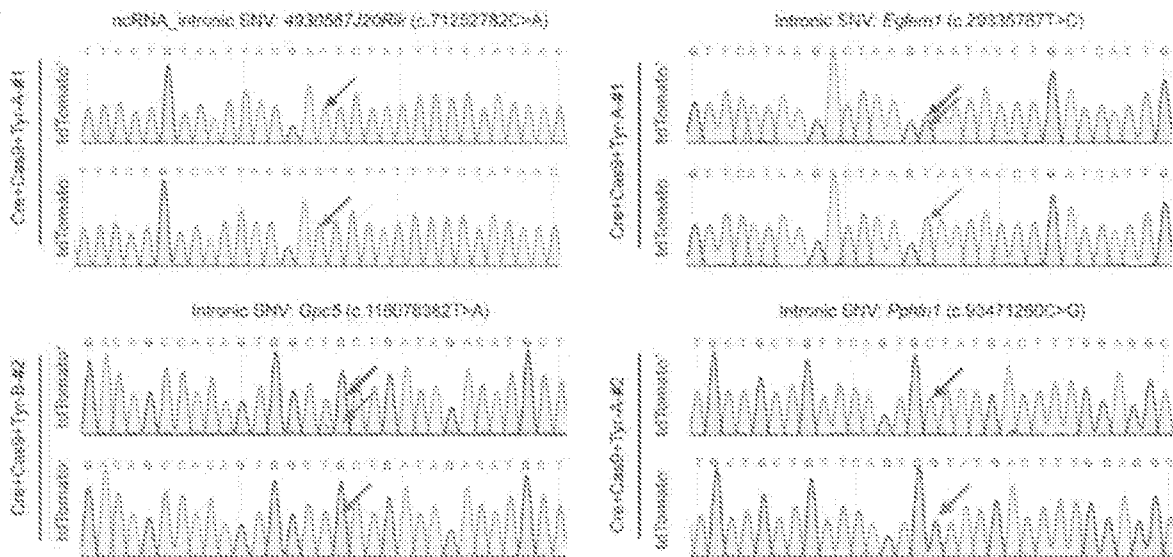


图 11

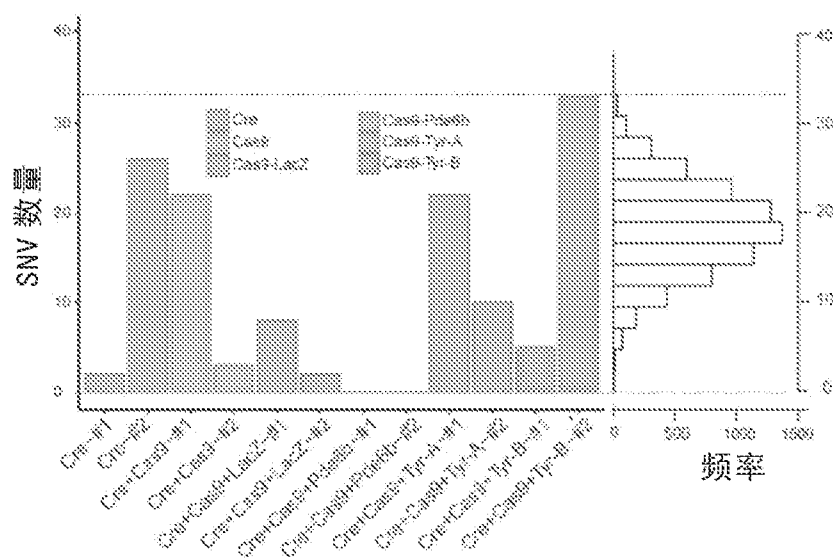


图 12

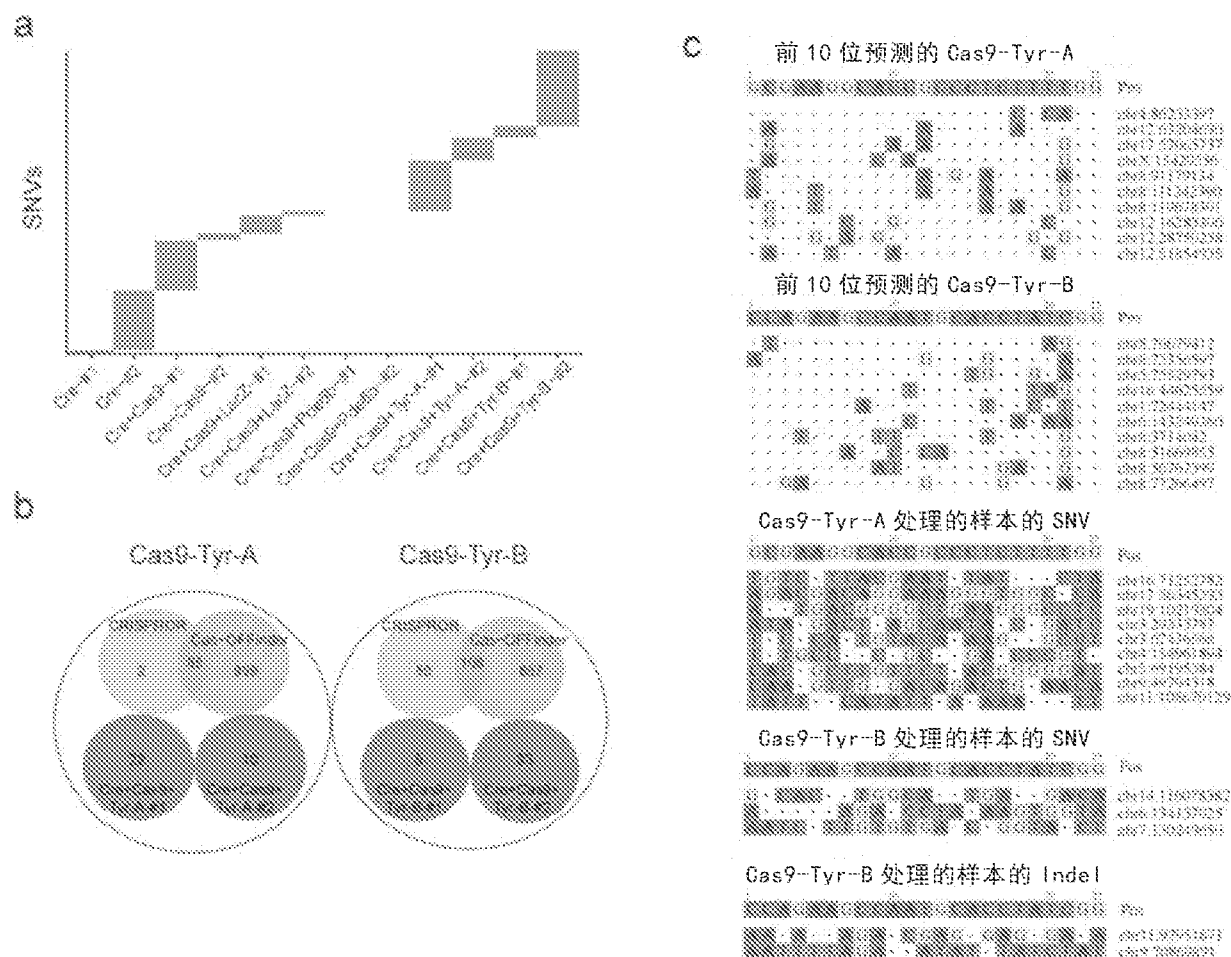


图 13

a

Variants	Cre + Cas9 +											
	Cre-#1	Cre-#2	-#1	-#2	LacZ-#1	LacZ-#2	Poab-#1	Poab-#2	Tyr-A-#1	Tyr-A-#2	Tyr-B-#1	Tyr-B-#2
WGS SNVs	3	12	7	8	1	6	0	0	38	8	13	14
WGS Indels	3	0	1	2	0	1	0	0	4	1	2	1
Exon SNVs	0	0	0	0	0	0	0	0	1	0	2	2
Exon Indels	0	0	0	0	0	0	0	0	0	0	0	0
Nonsynonymous SNVs	0	0	0	0	0	0	0	0	1	0	1	2
Frameshift Indels	0	0	0	0	0	0	0	0	0	0	0	0
Predicted offtarget sites							0	0	0	0	0	0
Digizome-seq putative offtarget sites							0	0	0	0	0	0

b

Variants	Cre + BE3 +					
	Cre+BE3-#1	Cre+BE3-#2	Tyr-C-#1	Tyr-C-#2	Tyr-D-#1	Tyr-D-#2
WGS SNVs	15	19	12	19	30	7
WGS Indels	1	1	1	0	1	1
Exon SNVs	0	0	0	0	0	0
Exon Indels	0	0	0	0	0	0
Nonsynonymous SNVs	0	0	0	0	0	0
Frameshift Indels	0	0	0	0	0	0

图 14

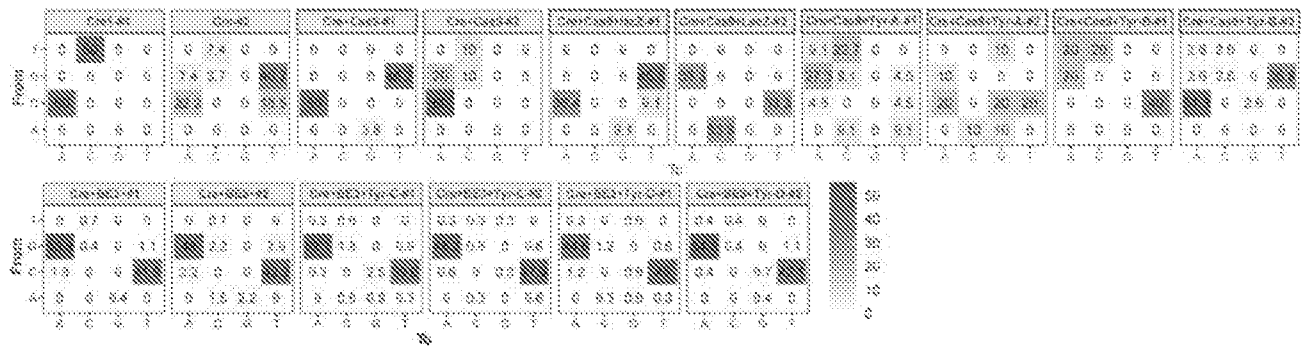


图 15

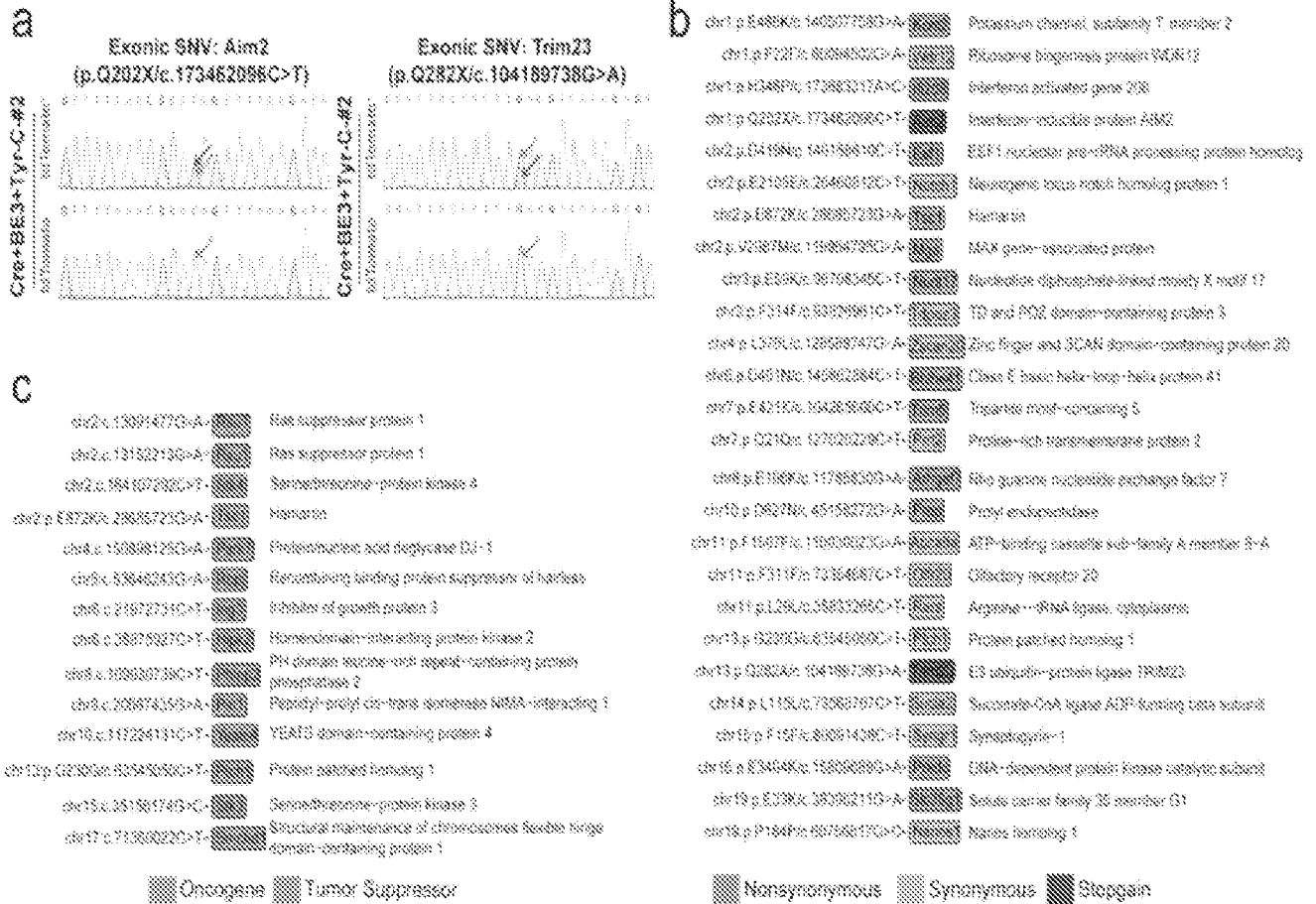


图 16

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2019/119842

A. CLASSIFICATION OF SUBJECT MATTER

C12N 15/90(2006.01)i; C12N 9/78(2006.01)i; C12Q 1/34(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)

C12N; C12Q

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, DWPI, SIPOABS, CNTXT, WOTXT, USTXT, 百度学术, BADU SCHOLAR, ISI WEB OF SCIENCE: 基因编辑, 碱基编辑器, 胞嘧啶脱氨酶, 脱靶, 检测, 胚胎, gene edit+, base edit+, BE, APOBEC1, APOBEC3A, OFF TARGET, DETECT+, ANALYS+, EMBRYO.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 108513575 A (PRESIDENT AND FELLOWS OF HARVARD COLLEGE) 07 September 2018 (2018-09-07) see description, paragraphs 252-256 and 968-970, and claims 1, 10, and 297	1-7, 9-11, 13-16
X	CN 109021111 A (SHANGHAITECH UNIVERSITY) 18 December 2018 (2018-12-18) see description, paragraphs 84-88, and claims 1, 5, and 19	1-5, 8, 9, 12-15
X	CN 108822217 A (SHANGHAITECH UNIVERSITY) 16 November 2018 (2018-11-16) see claims 1, 4, and 5	9, 12-15
X	WO 2019005886 A1 (BROAD INST INC. et al.) 03 January 2019 (2019-01-03) see the abstract	9-11, 13-15
PX	WO 2018176009 A1 (HARVARD COLLEGE) 27 September 2018 (2018-09-27) see the abstract	9-11, 13-15
PX	WO 2019161783 A1 (SHANGHAITECH UNIV) 29 August 2019 (2019-08-29) see the abstract	9, 12-15
PX	WO 2019126762 A2 (THE BROAD INSTITUTE, INC. et al.) 27 June 2019 (2019-06-27) see the abstract	9-11 and 13-15

☒ 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

17 March 2020

Date of mailing of the international search report

24 March 2020

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/CN2019/119842

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	WO 2019126709 A1 (THE BROAD INSTITUTE, INC. et al.) 27 June 2019 (2019-06-27) see the abstract	9-11 and 13-15
PX	WO 2019042284 A1 (SHANGHAITECH UNIV) 07 March 2019 (2019-03-07) see the abstract	9-11, 13-15
A	CN 109295186 A (SUN YAT-SEN UNIVERSITY) 01 February 2019 (2019-02-01) see entire document	1-34
A	CN 109136351 A (HUAZHONG AGRICULTURAL UNIVERSITY) 04 January 2019 (2019-01-04) see entire document	1-34
A	WO 2018135838 A2 (INST BASIC SCIENCE) 26 July 2018 (2018-07-26) see entire document	1-34
A	CN 107557394 A (NANJING DRUM TOWER HOSPITAL) 09 January 2018 (2018-01-09) see entire document	1-34
PA	CN 109868283 A (UNIV ZHEJIANG A&F) 11 June 2019 (2019-06-11) see entire document	1-34
X	GEHRKE, J. M. et al. "An APOBEC3A-Cas9 Base Editor With Minimized Bystander and Off-Target Activities" <i>Nature biotechnology</i> , Vol. 36, No. 10, 30 November 2018 (2018-11-30), ISSN: 1546-1696, pages 977-982, see abstract, and page 978, last paragraph	1-5, 8, 9, 12-15
X	WANG, X. et al. "Efficient Base Editing in Methylated Regions With a Human APOBEC3A-Cas9 Fusion" <i>Nature Biotechnology</i> , Vol. 36, No. 10, 30 November 2018 (2018-11-30), ISSN: 1546-1696, pages 1-8, see the abstract, and figure 2	9, 12-15
PX	ZUO, E. et al. "Cytosine base editor generates substantial off-target single-nucleotide variants in mouse embryos" <i>Science</i> , Vol. 364, No. 6437, 28 February 2019 (2019-02-28), ISSN: 1095-9203, pages 289-292, see abstract	17-34
A	MARTIN AS. et al. "A Panel of eGFP Reporters for Single Base Editing by APOBEC-Cas9 Editosome Complexes" <i>Scientific Reports</i> , Vol. 9, No. 1, 24 January 2019 (2019-01-24), ISSN: 2045-2322, pages 1-8, see entire document	1-34

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

- [1] Invention 1: Claims 1-7, 9-11, and 13-15 (all partially), relate to cytosine deaminase APOBEC1 mutant, an encoding nucleic acid thereof, an application thereof in gene editing, a single base editor containing the mutant, and a mutation method for the cytosine deaminase APOBEC1.
- [2] Invention 2: Claims 1-5, 8, 9, 12, and 13-15 (all partially), relate to cytosine deaminase APOBEC1 mutant, an encoding nucleic acid thereof, an application thereof in gene editing, a single base editor containing the mutant, and a mutation method for the cytosine deaminase APOBEC1.
- [3] Invention 3: Claim 16, relates to a method for screening substances for reducing off-target effect of the single base editor by detecting interaction between cytosine deaminase and DNA.
- [4] Invention 4: Claims 17-34, relate to a method for analyzing the targeted effects of gene editing.
- [5] The cytosine deaminase is a known enzyme in the prior art (for example, editor in chief: Cao Guangwen and Du Ping, "Modern Cancer Biotherapy", Beijing: People's Military Medical Press, November 1995, see pages 364-368), cytosine deaminase APOBEC1 mutants and APOBECA3A mutants do not share important structural units, and inventions 1-4 do not have the same or corresponding technical features, so inventions 1-4 are not so linked as to form a single general inventive concept and therefore do not comply with the requirement of unity of invention as defined in PCT Rule 13.1

- 1. ☒ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
- 2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.
- 3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
- 4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- ☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- ☒ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2019/119842

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	108513575	A	07 September 2018	EP	3365357	A2	29 August 2018
				SG	11201803173V	A	30 May 2018
				US	2019225955	A1	25 July 2019
				US	2018312825	A1	01 November 2018
				JP	2018537963	A	27 December 2018
				US	10167457	B2	01 January 2019
				CA	3002827	A1	27 April 2017
				WO	2017070633	A3	13 July 2017
				WO	2017070632	A3	08 June 2017
				US	2017121693	A1	04 May 2017
				WO	2017070633	A2	27 April 2017
				JP	2019509012	A	04 April 2019
				AU	2016342380	A1	10 May 2018
				WO	2017070632	A2	27 April 2017
				KR	20180069898	A	25 June 2018
				IL	258821	D0	28 June 2018
				EP	3365356	A2	29 August 2018
				CN	108699116	A	23 October 2018
CN	109021111	A	18 December 2018	CN	108822217	A	16 November 2018
CN	108822217	A	16 November 2018	CN	109021111	A	18 December 2018
WO	2019005886	A1	03 January 2019	None			
WO	2018176009	A1	27 September 2018	KR	20190130613	A	22 November 2019
				EP	3601562	A1	05 February 2020
				SG	11201908658T	A	30 October 2019
				GB	2575930	A	29 January 2020
				CA	3057192	A1	27 September 2018
				US	2018312828	A1	01 November 2018
				GB	201915338	D0	04 December 2019
				AU	2018240571	A1	17 October 2019
				IL	269458	D0	28 November 2019
WO	2019161783	A1	29 August 2019	None			
WO	2019126762	A2	27 June 2019	WO	2019126762	A3	08 August 2019
WO	2019126709	A1	27 June 2019	None			
WO	2019042284	A1	07 March 2019	WO	2019041296	A1	07 March 2019
CN	109295186	A	01 February 2019	None			
CN	109136351	A	04 January 2019	None			
WO	2018135838	A2	26 July 2018	KR	20180084671	A	25 July 2018
				CN	110234770	A	13 September 2019
				EP	3572525	A2	27 November 2019
				WO	2018135838	A3	06 December 2018
				US	2018258418	A1	13 September 2018
CN	107557394	A	09 January 2018	None			
CN	109868283	A	11 June 2019	None			

国际检索报告

国际申请号

PCT/CN2019/119842

A. 主题的分类 C12N 15/90(2006.01)i; C12N 9/78(2006.01)i; C12Q 1/34(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) C12N; C12Q 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, DWPI, SIPOABS, CNTXT, WOTXT, USTXT, 百度学术, ISI WEB OF SCIENCE:基因编辑, 碱基编辑器, 胞嘧啶脱氨酶, 脱靶, 检测, 胚胎, gene edit+, base edit+, BE, APOBEC1, APOBEC3A, OFF TARGET, DETECT+, ANALYS+, EMBRYO。																													
C. 相关文件 <table border="1"> <thead> <tr> <th>类 型*</th> <th>引用文件, 必要时, 指明相关段落</th> <th>相关的权利要求</th> </tr> </thead> <tbody> <tr> <td>X</td> <td>CN 108513575 A (哈佛大学的校长及成员们) 2018年 9月 7日 (2018 - 09 - 07) 参见说明书第252-256、968-970段和权利要求1、10、297</td> <td>1-7、9-11、13-16</td> </tr> <tr> <td>X</td> <td>CN 109021111 A (上海科技大学) 2018年 12月 18日 (2018 - 12 - 18) 参见说明书第84-88段和权利要求1、5、19</td> <td>1-5、8-9、12-15</td> </tr> <tr> <td>X</td> <td>CN 108822217 A (上海科技大学) 2018年 11月 16日 (2018 - 11 - 16) 参见权利要求1、4、5</td> <td>9、12-15</td> </tr> <tr> <td>X</td> <td>WO 2019005886 A1 (BROAD INST INC等) 2019年 1月 3日 (2019 - 01 - 03) 参见摘要</td> <td>9-11、13-15</td> </tr> <tr> <td>PX</td> <td>WO 2018176009 A1 (HARVARD COLLEGE) 2018年 9月 27日 (2018 - 09 - 27) 参见摘要</td> <td>9-11、13-15</td> </tr> <tr> <td>PX</td> <td>WO 2019161783 A1 (SHANGHAITECH UNIV) 2019年 8月 29日 (2019 - 08 - 29) 参见摘要</td> <td>9、12-15</td> </tr> <tr> <td>PX</td> <td>WO 2019126762 A2 (BROAD INST INC等) 2019年 6月 27日 (2019 - 06 - 27) 参见摘要</td> <td>9-11、13-15</td> </tr> <tr> <td>PX</td> <td>WO 2019126709 A1 (BROAD INST INC等) 2019年 6月 27日 (2019 - 06 - 27) 参见摘要</td> <td>9-11、13-15</td> </tr> </tbody> </table>			类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求	X	CN 108513575 A (哈佛大学的校长及成员们) 2018年 9月 7日 (2018 - 09 - 07) 参见说明书第252-256、968-970段和权利要求1、10、297	1-7、9-11、13-16	X	CN 109021111 A (上海科技大学) 2018年 12月 18日 (2018 - 12 - 18) 参见说明书第84-88段和权利要求1、5、19	1-5、8-9、12-15	X	CN 108822217 A (上海科技大学) 2018年 11月 16日 (2018 - 11 - 16) 参见权利要求1、4、5	9、12-15	X	WO 2019005886 A1 (BROAD INST INC等) 2019年 1月 3日 (2019 - 01 - 03) 参见摘要	9-11、13-15	PX	WO 2018176009 A1 (HARVARD COLLEGE) 2018年 9月 27日 (2018 - 09 - 27) 参见摘要	9-11、13-15	PX	WO 2019161783 A1 (SHANGHAITECH UNIV) 2019年 8月 29日 (2019 - 08 - 29) 参见摘要	9、12-15	PX	WO 2019126762 A2 (BROAD INST INC等) 2019年 6月 27日 (2019 - 06 - 27) 参见摘要	9-11、13-15	PX	WO 2019126709 A1 (BROAD INST INC等) 2019年 6月 27日 (2019 - 06 - 27) 参见摘要	9-11、13-15
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求																											
X	CN 108513575 A (哈佛大学的校长及成员们) 2018年 9月 7日 (2018 - 09 - 07) 参见说明书第252-256、968-970段和权利要求1、10、297	1-7、9-11、13-16																											
X	CN 109021111 A (上海科技大学) 2018年 12月 18日 (2018 - 12 - 18) 参见说明书第84-88段和权利要求1、5、19	1-5、8-9、12-15																											
X	CN 108822217 A (上海科技大学) 2018年 11月 16日 (2018 - 11 - 16) 参见权利要求1、4、5	9、12-15																											
X	WO 2019005886 A1 (BROAD INST INC等) 2019年 1月 3日 (2019 - 01 - 03) 参见摘要	9-11、13-15																											
PX	WO 2018176009 A1 (HARVARD COLLEGE) 2018年 9月 27日 (2018 - 09 - 27) 参见摘要	9-11、13-15																											
PX	WO 2019161783 A1 (SHANGHAITECH UNIV) 2019年 8月 29日 (2019 - 08 - 29) 参见摘要	9、12-15																											
PX	WO 2019126762 A2 (BROAD INST INC等) 2019年 6月 27日 (2019 - 06 - 27) 参见摘要	9-11、13-15																											
PX	WO 2019126709 A1 (BROAD INST INC等) 2019年 6月 27日 (2019 - 06 - 27) 参见摘要	9-11、13-15																											
☒ 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 2020年 3月 17日		国际检索报告邮寄日期 2020年 3月 24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		授权官员 张丹丹 电话号码 62411308																											

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
PX	WO 2019042284 A1 (SHANGHAITECH UNIV) 2019年 3月 7日 (2019 - 03 - 07) 参见摘要	9-11、13-15
A	CN 109295186 A (中山大学) 2019年 2月 1日 (2019 - 02 - 01) 参见全文	1-34
A	CN 109136351 A (华中农业大学) 2019年 1月 4日 (2019 - 01 - 04) 参见全文	1-34
A	WO 2018135838 A2 (INST BASIC SCIENCE) 2018年 7月 26日 (2018 - 07 - 26) 参见全文	1-34
A	CN 107557394 A (南京鼓楼医院) 2018年 1月 9日 (2018 - 01 - 09) 参见全文	1-34
PA	CN 109868283 A (UNIV ZHEJIANG A&F) 2019年 6月 11日 (2019 - 06 - 11) 参见全文	1-34
X	GEHRKE JM. et al. "An APOBEC3A-Cas9 Base Editor With Minimized Bystander and Off-Target Activities" 《Nature biotechnology》, 第36卷, 第10期, 2018年 11月 30日 (2018 - 11 - 30), ISSN: 1546-1696, 第977-982页, 参见摘要及第978页末段	1-5、8-9、12-15
X	WANG X. et al. "Efficient Base Editing in Methylated Regions With a Human APOBEC3A-Cas9 Fusion" 《Nature Biotechnology》, 第36卷, 第10期, 2018年 11月 30日 (2018 - 11 - 30), ISSN: 1546-1696, 第1-8页, 参见摘要及图2	9、12-15
PX	ZUO, E. et al. "Cytosine base editor generates substantial off-target single-nucleotide variants in mouse embryos" 《Science》, 第364卷, 第6437期, 2019年 2月 28日 (2019 - 02 - 28), ISSN: 1095-9203, 第289-292页, 参见摘要	17-34
A	MARTIN AS. et al. "A Panel of eGFP Reporters for Single Base Editing by APOBEC-Cas9 Editosome Complexes" 《Scientific Reports》, 第9卷, 第1期, 2019年 1月 24日 (2019 - 01 - 24), ISSN: 2045-2322, 第1-8页, 参见全文	1-34

第III栏 缺乏发明单一性的意见(续第1页第3项)

本国际检索单位在该国际申请中发现多项发明, 即:

- [1] 发明1: 权利要求1-7、9-11、13-15 (均为部分), 涉及胞嘧啶脱氨酶APOBEC1突变体、其编码核酸、其在基因编辑中的用途以及含有所述突变体的单碱基编辑器和对胞嘧啶脱氨酶APOBEC1进行突变的方法。
- [2] 发明2: 权利要求1-5、8-9、12、13-15 (均为部分), 涉及胞嘧啶脱氨酶APOBECA3A突变体、其编码核酸、其在基因编辑中的用途以及含有所述突变体的单碱基编辑器和对胞嘧啶脱氨酶APOBECA3A进行突变的方法。
- [3] 发明3: 权利要求16, 通过检测胞嘧啶脱氨酶与DNA相互作用筛选降低单碱基编辑器的脱靶效应的物质的方法。
- [4] 发明4: 权利要求17-34, 一种分析基因编辑靶向效应的方法。
- [5] 胞嘧啶脱氨酶为现有技术中已知酶(例如, 曹广文和杜平主编, 《现代癌症生物治疗学》, 北京: 人民军医出版社, 1995年11月, 参见第364-368页), 胞嘧啶脱氨酶APOBEC1突变体与APOBECA3A突变体之间并不共享具有重要意义的结构单元, 发明1-4之间没有相同或相应的技术特征, 因此, 发明1-4之间不存在技术关联, 不属于一个总的发明构思, 因而不满足发明单一性的要求, 不符合细则13.1的规定。

1. ☒ 由于申请人按时缴纳了被要求缴纳的全部附加检索费, 本国际检索报告涉及全部可作检索的权利要求。
2. ☐ 由于无需付出有理由要求附加费的劳动即能对全部可检索的权利要求进行检索, 本单位未通知缴纳任何加费。
3. ☐ 由于申请人仅按时缴纳了部分被要求缴纳的附加检索费, 本国际检索报告仅涉及已缴费的那些权利要求, 具体地说, 是权利要求:
4. ☐ 申请人未按时缴纳被要求缴纳的附加检索费。因此, 本国际检索报告仅涉及权利要求书中首先提及的发明; 包含该发明的权利要求是:

对异议的意见

- ☐ 申请人缴纳了附加检索费, 同时提交了异议书, 适用时, 缴纳了异议费。
- ☐ 申请人缴纳了附加检索费, 同时提交了异议书, 但未在通知书规定的时间期限内缴纳异议费。
- ☒ 缴纳附加检索费时未提交异议书。

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2019/119842

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	108513575	A	2018年 9月 7日	EP 3365357 A2	2018年 8月 29日
				SG 11201803173V A	2018年 5月 30日
				US 2019225955 A1	2019年 7月 25日
				US 2018312825 A1	2018年 11月 1日
				JP 2018537963 A	2018年 12月 27日
				US 10167457 B2	2019年 1月 1日
				CA 3002827 A1	2017年 4月 27日
				WO 2017070633 A3	2017年 7月 13日
				WO 2017070632 A3	2017年 6月 8日
				US 2017121693 A1	2017年 5月 4日
				WO 2017070633 A2	2017年 4月 27日
				JP 2019509012 A	2019年 4月 4日
				AU 2016342380 A1	2018年 5月 10日
				WO 2017070632 A2	2017年 4月 27日
				KR 20180069898 A	2018年 6月 25日
				IL 258821 D0	2018年 6月 28日
				EP 3365356 A2	2018年 8月 29日
				CN 108699116 A	2018年 10月 23日
CN	109021111	A	2018年 12月 18日	CN 108822217 A	2018年 11月 16日
CN	108822217	A	2018年 11月 16日	CN 109021111 A	2018年 12月 18日
WO	2019005886	A1	2019年 1月 3日	无	
WO	2018176009	A1	2018年 9月 27日	KR 20190130613 A	2019年 11月 22日
				EP 3601562 A1	2020年 2月 5日
				SG 11201908658T A	2019年 10月 30日
				GB 2575930 A	2020年 1月 29日
				CA 3057192 A1	2018年 9月 27日
				US 2018312828 A1	2018年 11月 1日
				GB 201915338 D0	2019年 12月 4日
				AU 2018240571 A1	2019年 10月 17日
				IL 269458 D0	2019年 11月 28日
WO	2019161783	A1	2019年 8月 29日	无	
WO	2019126762	A2	2019年 6月 27日	WO 2019126762 A3	2019年 8月 8日
WO	2019126709	A1	2019年 6月 27日	无	
WO	2019042284	A1	2019年 3月 7日	WO 2019041296 A1	2019年 3月 7日
CN	109295186	A	2019年 2月 1日	无	
CN	109136351	A	2019年 1月 4日	无	
WO	2018135838	A2	2018年 7月 26日	KR 20180084671 A	2018年 7月 25日
				CN 110234770 A	2019年 9月 13日
				EP 3572525 A2	2019年 11月 27日
				WO 2018135838 A3	2018年 12月 6日
				US 2018258418 A1	2018年 9月 13日
CN	107557394	A	2018年 1月 9日	无	
CN	109868283	A	2019年 6月 11日	无	