参赛队员姓名：Arianna Xie

中学：上海中学国际部

省份：上海市

国家/地区：中国南部赛区

指导教师姓名：陈琳；冯依力

指导教师单位：上海中学国际部；

浙江大学转化医学研究院

论文题目：通过生成 DNA 间隙改良碱

基编辑器

Development of a Gap-Driven Base Editing System

Author: Arianna Xie

High School: Shanghai High School International Division

Province: Shanghai

Region: Southern China

Instructors: Lin Chen, Shanghai High School International Division,

Dr. Yili Feng, Institute of Translational Medicine, Zhejiang University

September 2024

Table of Contents

Abstract

Base editors hold great promise in the treatment of genetic diseases as nearly half of known pathogenic genetic variants are point mutations. However, despite recent progress, the base editors are limited by the low editing efficiency and a narrow editing window. In this study, we present a novel gap-driven base editing system in which paired Cas9 D10A-sgRNAs are designed to induce a gap instead of a single nick on the DNA target strand, thereby enhancing editing efficiency. Using a dye-labeled probe, we validated the ability of two Cas9 D10A-sgRNAs to generate a gap on the target strand *in vitro*. In mouse embryonic stem (ES) cells, we found gap-directed cytosine base editors (Gap-CBEs) demonstrated an approximately twofold increase in editing efficiency compared to single base editors. A sufficient increase is also detected for gap-directed adenine base editors (Gap-ABEs). The gap introduced by two Cas9 D10A could further expand the editing window, potentially offering a broader target region for deaminases. This novel strategy in base editing opens new avenues for further improvement of current base editors and advances base editing technology.

Keywords

Base editors, Point mutations, Gap, Editing efficiency, Editing window

论文摘要

考虑到将近一半的致病性遗传变异是由点突变引起,碱基编辑器正在成为治疗遗传疾病的重要工具。然而，尽管取得了振奋人心的进展，目前的碱基编辑器的编辑效率依旧很低，并且只能在有限的编辑窗口内工作。在本研究中，我们提出了一种通过间隙生成加强碱基编辑的策略。相比于诱导单链缺刻，该系统能够利用两个碱基编辑器里的 Cas9 D10A-sgRNA 在 DNA 靶标链上生成一个间隙，从而提高编辑效率。在体外实验中，我们利用染料标记的探针，验证了该系统能够在靶标链上生成间隙。通过转染细胞并进行流式细胞术分析，我们发现，与单一碱基编辑器相比，间隙引导的胞嘧啶碱基编辑器（Gap-CBE）将编辑效率提升了 1 倍，而间隙引导的腺嘌呤碱基编辑器（Gap-ABE）也能明显地提升编辑效率。由于间隙可以延长 DNA 的单链区域，间隙引导的碱基编辑器扩大了编辑窗口，提供了更广泛的脱氨酶活性目标区域。这一碱基编辑的创新工作模式为碱基编辑技术的改良开辟了新的途径，将推动碱基编辑技术的发展。

关键词

碱基编辑器，点突变，间隙，编辑效率，编辑窗口

1. Introduction

Single nucleotide mutation is considered to be the main cause of many life-threatening genetic diseases, such as Tay-Sachs, sickle cell disease, and Duchene muscular dystrophin. Thus, a gene editing technique that can efficiently correct these mutations holds the potential to achieve permanent cure with single treatment[1]. Over the decades, scientists have developed various gene editing systems for this purpose, but the low editing efficiency limits their application[2]. In 2012, Jinek et al demonstrated that the Clustered Regularly Interspersed Short Palindromic Repeats (CRISPR)/CRISPR associated protein 9 (Cas9), an adaptive immune system in bacteria and archaea, could be programmed to cleave a given DNA target[3]. In this programable system, Cas9 protein, guided by a single guide RNA (sgRNA), recognizes and binds its target. After initial DNA binding, Cas9 unwinds the double-stranded DNA (dsDNA), allowing base pairing between the spacer of sgRNA and the target strand and promoting formation of an R-loop. Subsequently, Cas9 uses two endonuclease domains (HNH and RuvC-like) to respectively cleave the target strand and the nontarget strand of unwound DNA, generating a DNA double strand break (DSB), repair of which generates editing products of interest[3]. CRISPR/Cas9 was thus developed into a powerful tool for targeted genome editing in cells and in organisms[4]. These attributes of CRISPR/Cas9 such as DNA targeting, DNA unwinding, target cleavage and target residence, separated or combined, confer great versality in the application of CRISPR/Cas9 genome editing[5].

While the DSB-based CRISPR/Cas9 genome editing could increase genome editing through homology-directed repair in correcting single nucleotide mutation, the efficiency remains unsatisfactory. Additionally, the generation of DSBs may also lead to generation of genotoxic byproducts (such as translocations and genome rearrangements). To address these issues, the base editors including cytosine base editors (CBEs) and adenine base editors (ABEs) were later developed to enhance correction of single nucleotide mutations without induction of DSBs[6,7]. By fusing a nickase of Cas9 (nCas9-D10A) with a cytidine deaminase and an uracil glycosylase inhibitor (UGI)

polypeptide, CBEs can convert Cytidine (C) to Uridine (U) in the single-stranded nontarget DNA strand in the nCas9-D10A-sgRNA-target complex by deamination. Similarly, ABEs were constructed by fusing a nickase of Cas9 (nCas9-D10A) with an adenine deaminase and could convert Adenosine (A) to Inosine (I) in the single-stranded nontarget DNA strand in the nCas9-D10A-sgRNA-target complex. Simultaneously, nCas9-D10A installs a nick on the target strand, and repair and replication could convert the original C:G into T:A or A:T into G:C. Recently, additional types of base editors were developed to allow the C-to-G and T-to-A transversion[8].

Despite the progress, application of base editors is still limited for several reasons. Firstly, the protospacer adjacent motif (PAM) recognition limits the sequence accessibility of base editors for robust editing. Secondly, the base editors only have a high deamination activity in a narrow editing window, ranging from positions 4 to 8 within the protospacer (from PAM-distal end to PAM-proximal end). Thirdly, although a nick is introduced on the opposite strand, it is still possible for the endogenous base excision repair pathway to remove the modified base in the nontarget strand and restore the original base. Several new versions of base editors have been developed to address these problems. For example, the use of Cas9 variants with relaxed or altered PAM specificities has broadened the targeting scope[9]. Evolution of deaminase has improved the activity and compatibility of base editors[10]. Altered architecture has resulted in higher efficiency and product purity of base editing[11].

Nevertheless, the editing window for deamination remains narrow. It is difficult to efficiently correct the mutations outside of the editing window. Despite the current improvements, the editing efficiency is still too low for its application in clinics[12]. Based on these two concerns, this study here designed and tested a new base editing strategy, in which we used dual Cas9-D10A sgRNAs in base editors to induce two nicks on the same strand. Dissociation of base editors from the target removed the intervening sequence hybridized with sgRNA between two nicks on the target strand *in vitro*, generating a gap opposite the modified nontarget strand as expected. Theoretically, this gap

generation could expand the window for deamination and also enhance the base editing efficiency in cells. Indeed, using a green fluorescent protein (GFP)-based base editing reporter in mouse embryonic stem (ES) cells[13], we found that this strategy improved the base editors by broadening the window for deamination and increasing the base editing efficiency.

2. Materials and Methods

2.1. Plasmid preparation

The NG-ABE9e (ABE) plasmid was a gift from Dr. Feng Gu[14]. The original CBE plasmid, BE4max, was obtained from Addgene (Plasmid #138491)[15]. The wild-type GFP expression plasmid pcDNA3-EGFP (Plasmid #13031) and the CRISPR/Cas9 expression plasmid px330 (Cat #42230) were originally obtained from Addgene. The *Streptococcus pyogenes* Cas9 (SpCas9) cassette of px330 was cloned into a pcDNA3β-Hyg-based expression vector. The original vector pcDNA3β was described previously[16]. Plasmids expressing nCas9 D10A were generated by site-directed mutagenesis from pcDNA3β-Cas9 using the KOD-Plus-Neo Kit (TOYOBO)[17]. The U6-sgRNA vector (pU6-gRNA) was derived from px330 by removing the CBh-hSpCas9 cassette, and individual sgRNAs were cloned into pU6-gRNA using a standard protocol[13]. The sgRNA gBE1 and gBE2 were constructed previously, and the target sequences for gBE1 and gBE2 are GGACAATAGGGATTGGCGGCGGT and GGTACCTAATGGGACAATAGGGA, respectively. The non-homologous end joining (NHEJ) reporter sGEJ was constructed and described previsouly[18].

The cloning of plasmid was performed using *E. coli* (DH5α) with a standard protocol. Briefly, 10 μL of cells and 1 μL of plasmid were transferred to a 1.5 mL centrifuge tube. The tube was incubated on ice for 20 minutes then heated in the 42°C water bath for 45 seconds and incubated on ice for 3 minutes. The transformed cells were subsequently added to a LB agar plate containing 100 mg/L ampicillin and evenly spread with a flattened micropipette tip. The LB plate was reversed and incubated at 37°C overnight. Until colonies were formed, one colony was picked and transferred into a 50 mL

centrifuge tube containing 20 mL LB liquid medium and incubated in a 37°C shaker. In the next day, the tube was centrifuged at 5,000 rpm for 5 minutes (Thermo Fisher), and the supernatants were removed. The plasmid was isolated and purified using the FastPure Plasmid Mini Kit (Vazyme) according to the manufacturer's protocol. The concentration of the plasmid was measured by Nanodrop 2000 (Thermo Fisher) and varied from 0.137 mg/mL to 0.766 mg/mL.

2.2. Cells and cell culture

Mouse ES cells harboring single-copy non-homologous end joining (NHEJ) reporter sGEJ at the *ROSA26* locus were developed previosuly[18]. Mouse ES cells were grew on a plate coated with 0.5 mL of 0.1% gelatin solution (Sangon Biotech) in the medium prepared with 500 mL DMEM (High Glucose) (Gibco), 6 mL 100× L-glutamine (Gibco), 3 mL 100× 2-mercaptoethanol (Sigma), 6 mL 100× MEM non-essential amino acid (Gibco), 6 mL 100× sodium pyruvate (Gibco), 6 mL 100× penicillin/ streptomycin (Gibco), 40 μL leukemia inhibitory factor (Millipore), and 70 mL fetal bovine serum (Gibco). Cells recovered from cryopreservation were passaged twice prior to transfection. Cells were passaged after their proliferation reached 70% confluency. The culture medium was removed, and the cells were rinsed with 3 mL of phosphate buffered saline (PBS) (Biological Industries) to eliminate residual medium. Following the removal of PBS buffer, 1 mL of 0.1% trypsin-EDTA (trypsin) (Gibco) was added to the 10 cm culture plate (Biofil). The dish was then incubated at 37°C for 3 minutes to detach the cells. 1 mL of fresh medium was added to neutralize the trypsin, and the detached cells were gently resuspended by pipetting. The cell suspension was transferred to a 15 mL centrifuge tube and centrifuged at 800 rpm for 3 minutes. After the supernatants were removed, 1 mL of fresh medium was added to the tube, and the cells were pipetted to resuspended thoroughly. The cell suspension, along with 8 mL of fresh medium, was added to a new 10 cm dish coated with gelatin. The cells were incubated in a 37°C, 5% CO_2 incubator (Thermo Fisher). To coat a dish, 0.5

mL of 0.1% gelatin solution was added into a new 10 cm dish, was removed after 5-minute incubation, and air dried in a biosafety cabinet (Thermo Fisher).

2.3. Analysis of *in vitro* generation of gap by dual Cas9 D10A-sgRNAs

The sGEJ reporter was used as a template for PCR with primer pairs (forward primer: AGTCTCGTGCAGATGGACAGCA; reverse primer: TCAGGTAGTGGTTGTCGGGCAG) to generate a 984-bp DNA substrate for *in vitro* gap generation by dual Cas9 D10A-sgRNAs. PCR and purification of the PCR products were performed by standard protocol. Cas9 D10A proteins, sgRNAs, and oligonucleotide probes labelled with Cy5.5 (Sulfo-Cyanine5.5) dye were purchased from Tsingke Biotech, Novoprotein, and Genescript Biotech, respectively. Four pairs of sgRNAs included gb2/gdd2 for the PAM-containing target sequence ATCCATGGTGGCGGCGGTTAGGG and GGATCCTAGGGATAACAGG of AGGATGGATCCTAGGGATAACAGGGTAATCCATGGTGGCGGCGGTTAGGG, respectively, gX1/gX2 for the target sequence CAAGTTCAGCGTGTCCGGCGAGG and GGCGAGGGCGATGCCACCTACGG of CCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGG, respectively, gY1/gY2 for the target sequence GCGCTCCTGGACGTAGCCTTCGG and GCCGTCGTCCTTGAAGAAGATGG of GTTGCCGTCGTCCTTGAAGAAGATGGTGCGCTCCTGGACGTAGCCTTCGG, respectively, and gX3/gX4 for the target sequence CAGCCACAACGTCTATATCATGG and ATGGCCGACAAGCAGAAGAACGG of CAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGG, respectively. The four probes were fully complementary to the nontarget strand within the predicted gaps generated respectively by paired Cas9 D10A-sgRNAs above. To test the ability of paired Cas9 D10A-sgRNAs to induce a gap, 1 μL of 0.1 μg/μL Cas9 D10A nuclease was mixed with 1 uL of 1 μM sgRNAs and

incubated at 37°C for 10 minutes to assemble the Cas9 D10A-sgRNA ribonucleoprotein complex. In a 10 μL reaction volume made up by sterilized water, 2 μL of the Cas9 D10A-sgRNA ribonucleoprotein complex solution was incubated with 2 μL of 0.1 μg/μL 984-bp DNA substrate at 37°C for 1 hour. 1 μL of 10 μM probe with Cy5.5 dye was subsequently added into the reaction and incubated at 37°C for an additional 1 hour. The reaction samples were collected and analyzed by electrophoresis on a 1% agarose gel in TAE buffer (Solarbio). Cy5.5 is a near infrared absorbing dye whose excitation peak is around 680 nm and emission peak is around 710 nm. The DNA hybridized with a probe labelled with Cy5.5 was determined by Odyssey scanner with a 700 nm channel.

2.4. Transfection of mouse ES cells

Transfection is the process of introducing foreign nucleic acid to eucaryotic cells. Transfection of mouse ES cells was performed with LipofectAMINE 2000 (Invitrogen) as described previously but with some modifications[13,19]. Briefly, mouse ES cells cultured were passaged once a day and cultured in fresh medium. On the day for transfection, the cells were trypsinized with 0.1% Tripsin-EDTA, counted and prepared in fresh mouse ES medium in a 10-mL tube at the concentration of 4-5 $\times 10^5$ per mL. 200 μL of this mouse ES cell suspension was added into each gelatin-coated well of a 24-well plate prepared in advance. In the meantime, two sets of solutions were prepared using 1.5mL centrifuge tubes: 1) 0.25 μg expression plasmids for sgRNAs or U6, 0.25 μg expression plasmids for Cas9 D10A, ABE, or CBE, and 35 μL OptiMEM solution (Gibco) were mixed as Solution A; 2) 1.2 μL LipofectAMINE 2000 and 35 μL OptiMEM solution were mixed as Solution B. Both solutions were incubated at room temperature for 5 minutes. Solution B was then added to each Solution A, and the mixture was incubated at room temperature for 20 minutes. After 200 μL of the mouse ES cell suspension at the concentration of 4-5 x 10^5 per mL was seeded in each gelatin-coated well of a 24-well plate, all of the Solution A/B mixture (about 70 μL) was added into one of the wells containing mouse ES cells. The 24-well plate was placed in the cell culture incubator for 6 hours and 1 mL

medium per well was added afterwards. The medium was replaced with fresh medium the next day. In 48 more hours, the cells were analyzed by flow cytometry. Transfection efficiency was performed with 0.10 μg of the GFP expression plasmids among 0.5 μg total plasmids. Specifically, 0.5 μg sgRNA expression plasmids in transfection for each sample were replaced with 0.10 μg of the GFP expression plasmids; 0.20 μg of expression plasmids for D10A, CBE, or ABE; and 0.20 μg of pU6. Transfection efficiency was also determined at 72 hours post transfection by flow cytometry.

2.5. Determination of base editing efficiency by flow cytometry

At 72 hours post tarnsfection, mouse ES cells were washed once with the PBS buffer and trypsinized with 100 μL of 0.1% trypsin in a 37°C incubator for 3 minutes. Cells were resuspended with 100 μL of culture medium, and the resulting 200 μL cell suspension in each well was passed through a 300-mesh sieve into the corresponding flow cytometry tube (Biofil). The tubes were placed on ice. The cells were counted, and GFP^+ cells were identified by flow cytometry using the Beckman Coulter CytoFLEX S flow cytometer according to the manufacturer's protocol. The data from the cytometer were uploaded and was analyzed using the CytExpert 3.0 software to determine the percentage of GFP^+ cells induced spontaneously, by Cas9 D10A, or by base editors. The percentage of GFP^+ cells was calculated as the number of GFP^+ cells divided by total number of cells validly counted.

2.6. Statistical analysis

In this study, triplicates were performed for each independent experiment of base editing. The formula $E = \frac{x_r - \overline{x_d} - \overline{x_{EV}}}{\overline{x_{GFP}}}$ was applied to calculate the editing efficiency for each base editing system with the data obtained from flow cytometry. x_r is the frequency of spontaneous GFP^+ cells combined with GFP^+ cells induced by base editing of the base editors and Cas9 D10A-induced indels from the

base editors. $\overline{x_d}$ is the average percentage of GFP$^+$ cells of triplicates from the Cas9 D10A groups and represents the frequency of GFP$^+$ cells induced by Cas9 D10A-induced indels. $\overline{x_{EV}}$ is the average percentage of GFP$^+$ cells of triplicates from the sgRNA empty vector control group and represents the background frequency of GFP$^+$ cells spontaneously induced. $\overline{x_{GFP}}$ is the average number of GFP$^+$ cells of triplicates for transfection efficiency. The AVERAGE function and the STDEV function in Excel were used to assess the mean of triplicates and Standard Deviation (S.D.), respectively. S.D. was calculated by the formula $S = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\overline{x})^2}{n-1}}$ and Standard Error of the Mean (S.E.M) were further calculated by the formula $S.E.M = S/\sqrt{n-1}$ to inform the extent to which values deviate from the mean, in which n is the number of repetitions, x_i is the value of each repetition, and $\bar{x}$ is the average of the repetitions. The results were demonstrated through histograms generated by Excel. The mean and S.E.M. were represented by the columns and the error bars, respectively.

3. Results

3.1. Design of *in vitro* gap generation by Cas9 D10A-sgRNAs in dual base editors

In base editing, the Cas9 D10A component of base editors is used to induce a DNA nick on the target strand, promoting removal of the target strand and enhancing base editing[7]. However, this removal is countered by efficient nick ligation, which suppresses base editing. It is reasonable to believe that base editing could be improved by increased removal of the target strand or inhibition of nick ligation. We thus designed a strategy employing paired base editors to help remove the target strand and improve base editing (**Figure 1a**). The targeted sites by dual base editors are close enough and even overlapped for a few nucleotides. Cas9 D10A of dual base editors targets the same strand for DNA nicking. After induction of the first nick by Cas9 D10A in one base editor, Cas9 D10A in the second base editor could unwind the targeted site near the DNA nick and cleave the target strand

(**Figure 1a**). Due to annealing to the sgRNA, dissociation of the second Cas9 D10A-sgRNA from the targeted site could take the cleaved target strand fragment away to generate a gap opposite to the nontarget strand with deaminated bases, promoting base editing (**Figure 1a**).

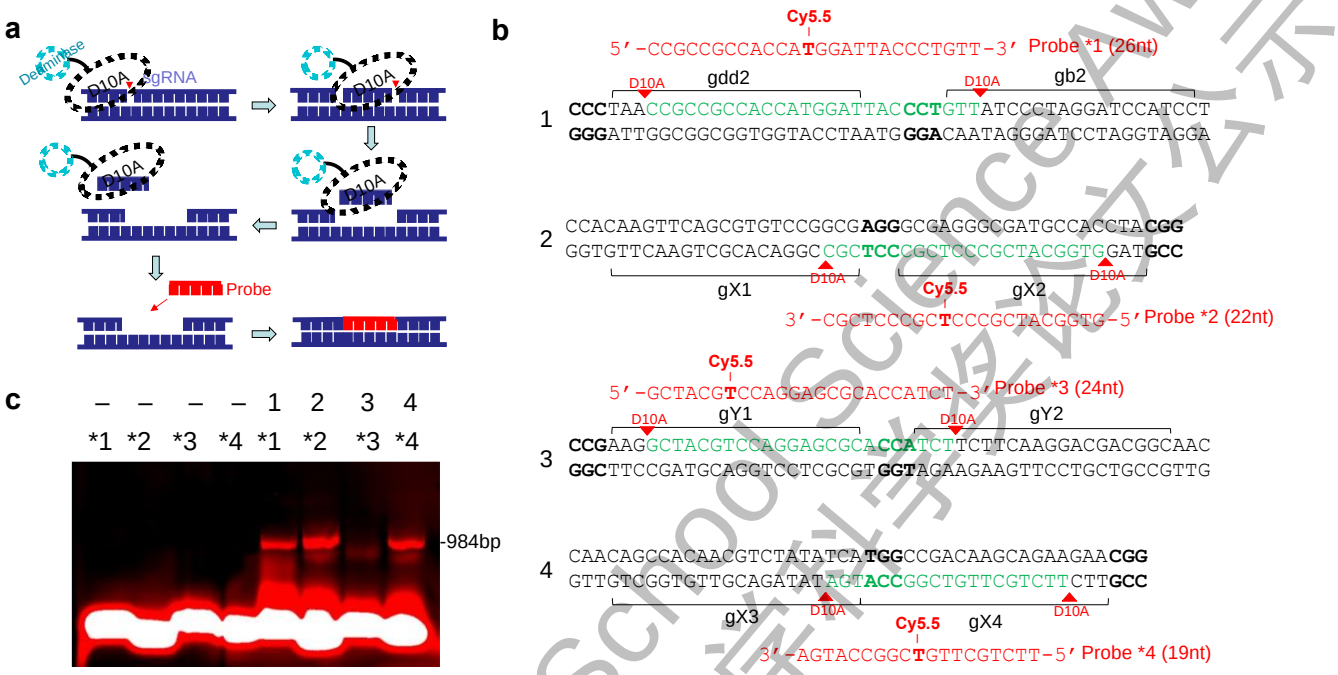

Figure 1. ***In vitro*** **generation of a gap by dual Cas9 D10A-sgRNAs.** (**a**) Design of *in vitro* generation of a gap by Cas9 D10A in dual base editors. A gap can be generated by paired nicking with dual Cas9 D10A-sgRNAs, allowing annealing of a complementary fluorophore-labelled probe with the nontarget strand. (**b**) Sequences of targeted sites by dual Cas9 D10A-sgRNAs and the probes perfectly complementary to the nontarget DNA strand in the gap. Paired sgRNAs are indicated as gdd2/gb2 (1), gX1/gX2 (2), gY1/gY2 (3) and gX3/gX4 (4) along with -NGG PAM in bold. The position expected for gaps is highlighted in green and the probes (i.e., *1 for gdd2/gb2; *2 for gX1/gX2; *3 for gY1/gY2, and *4 for gX3/gX4) in red. The length of the gaps and probes are also shown. The thymidine (T) attached with Cy5.5 is indicated in each of the probes. (**c**) *In vitro* detection of a gap by electrophoresis. The probes are indicated as with paired sgRNAs that guide Cas9 D10A to generate a gap. "–" represents no paired sgRNAs. The Length of the 984-bp substrate for gap generation is also shown.

3.2. *In vitro* generation of a gap by dual Cas9 D10A-sgRNAs in dual base editors

To determine whether dual base editors could generate a gap as predicted, we first tested the ability of dual Cas9 D10A-sgRNAs to generate a gap *in vitro*. We designed 4 pair of sgRNAs (i.e., 1: gdd2/gb2; 2: gX1/gX2; 3: gY1/gY2; and 4: gX3/gX4), each pair targeting a selected site in a 984-bp PCR product as a substrate, for Cas9 D10A-mediated gap generation (**Figure 1b**). We used a fluorophore-labeled single-stranded oligonucleotide as a probe to detect the gap generated by dual Cas9 D10A-sgRNAs. The gaps generated by Cas9 D10A with gdd2/gb2 (1), gX1/gX2 (2), gY1/gY2 (3) and gX3/gX4 (4) are 26 nts, 22 nts, 24 nts and 19 nts in length, respectively (**Figure 1b**). The probe (i.e., *1 for gdd2/gb2, *2 for gX1/gX2, *3 for gY1/gY2, or *4 for gX3/gX4) is fully complementary to the nontarget DNA strand in the gap, respectively (**Figure 1b**). The fluorophore Cy5.5, a near infrared absorbing dye whose excitation peak is around 680 nm and emission peak is around 710 nm, was attached to a thymine (T) of the probe during oligonucleotide synthesis. The emission of the dye can be captured by the Odyssey scanner with the 700 nm channel.

Indeed, after electrophoresis in the *in vitro* gap detection experiment, no bands were detected at the expected size with the probe in the absence of dual Cas9 D10A-sgRNAs in the first four lanes, indicating no gaps were generated without dual Cas9 D10A-sgRNAs (**Figure 1c**). In contrast, in the last four lanes, a 984-bp band was captured with the respective probe in the presence of paired sgRNAs gdd2/gb2 (1), gX1/gX2 (2), and gX3/gX4 (4), suggesting generation of a gap (**Figure 1c**). Upon the second nicking, only PAM-distal 17 nts of the nicked target strand is paired with the spacer, and 9 bp, 5 bp and 3 bp of the first nicked target strand remain annealed to the first nontarget strand for paired sgRNAs gdd2/gb2 (1), gX1/gX2 (2) and gX3/gX4 (4), respectively. This also indicated that dissociation of Cas9 D10A-sgRNA from the targeted site could pull the nicked target strand annealed to the nontarget strand from the targeted site with the 17-bp target DNA strand/sgRNA hybrid. No 984-bp band was found in the presence of gY1/gY2 (3), possibly due to the poor nicking efficiency of

Cas9 D10A with paired gY1/gY2 (**Figure 1c**). Taken together, these data demonstrated that dual Cas9 D10A-sgRNAs are capable of generating a gap in targeted sites under *in vitro* conditions, providing basis for the development of gap-directed base editors to improve base editing.

3.3. Design of the gap-driven base editing system with dual base editors

As described above, the nick induced on the target strand by the Cas9 D10A-sgRNA component of a base editor (e.g., the CBE BE4max or the ABE ABE9e) can potentially facilitate base editing by removing the target strand and retaining the nontarget strand with deaminated bases (e.g., U from C or I from A) as the repair substrate[6,7,14] (**Figure 2a**). Using the nontarget strand with deaminated bases as the repair substrate, DNA synthesis over deaminated bases (e.g., U from C or I from A) could generate new base pairs (e.g., C-G to U-A or A-T to I-C) of interest, achieving base editing intended (e.g., C-G to T-A or A-T to G-C) (**Figure 2a**). However, as the most frequent type of DNA lesions in mammalian cells, DNA nicks or single-stranded DNA breaks could be repaired efficiently and quickly by the repair pathway for single-stranded DNA breaks in mammalian cells. In this case, the nick induced by Cas9 D10A-sgRNA in the base editor would be readily ligated, and the nick ligation thus prevents the removal of the target strand, suppressing base editing (**Figure 2a**). Given the result that dual Cas9 D10A-sgRNAs are able to generate a gap opposite to the nontarget strand of the targeted site in *in vitro* assays, the Cas9 D10A-sgRNA component of dual base editors could directly generate a gap opposite to the nontarget strand of the targeted site with deaminated bases in cells, thus preventing nick ligation and promoting base editing with the nontarget strand as the repair substrate (**Figure 2b**). Compared to the original base editor composed of only one Cas9 complex (**Figure 2a**), our base editing strategy with dual base editors works by using two separate Cas9 D10A-sgRNAs in dual base editors to directly create a gap opposite the nontarget strand with deaminated bases and is expected to enhance the base editing efficiency and expand the base editing window (**Figure 2b**).

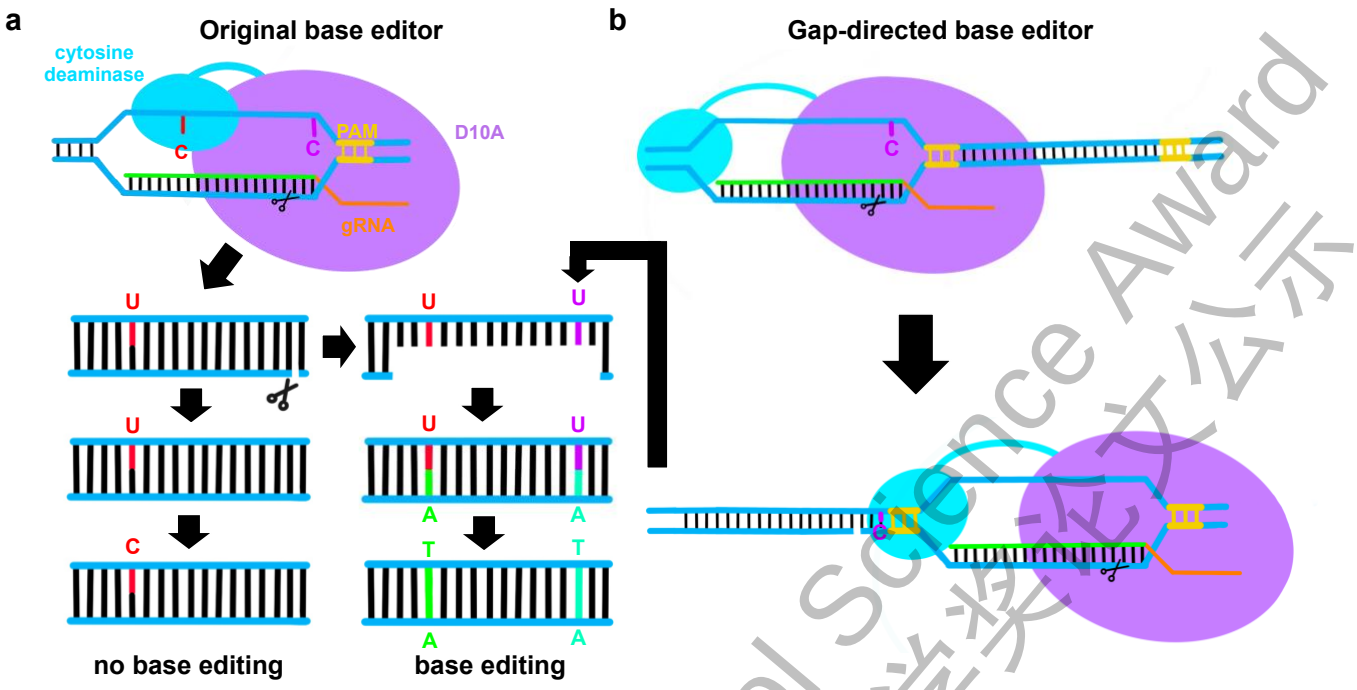

Figure 2. Design of gap-driven base editing with dual base editors. Schematic of the original base editor (**a**) and the gap-driven base editing strategy with dual base editors (**b**) are shown. The CBE is used as an example. The purple C is a cytosine base outside of the editing window of the original base editor.

3.4. Improved editing efficiency by the gap-driven base editing system

To test whether gap-driven base editing with dual base editors could increase the efficiency of base editing, we used the GFP-based sGEJ reporter in mouse ES cells for analysis of base editing. This reporter was originally developed for NHEJ analysis and contains a copy of the GFP gene with two translation start sites, "Koz-ATG" and "ATG"[18] (**Figure 3a**). Translation of GFP is suppressed by "Koz-ATG", an upstream and out-of-frame translation start site. Inactivation of "Koz-ATG" by base editing with either ABEs or the CBEs would convert T to C or G to A via deamination of A to I or deamination of C to U, allowing translation of GFP in the correct frame by the second translation start site ATG. Thus, if our gap-driven base editing with dual base editors successfully changes one base of

"Koz-ATG", GFP would be translated to generate GFP^+ cells, which can be measured by flow cytometry (**Figure 3a**). Specifically, ABE and CBE with either gBE1 or gBE2 alone could not efficiently edit the base in "TAC" paired with "ATG" of "Koz-ATG" because the "TAC" is neither in the range of base editing by the base editor with gBE1 nor in a good window for base editing by the base editor with gBE2. However, dual base editors would generate a gap, marking the I- or U-containing nontarget strand as the template for gap filling by DNA polymerases. Consequently, A to G or C to T transition would be enhanced, increasing the frequency of GFP^+ cells as compared to single base editors with either gBE1 or gBE2 (**Figure 3a**).

After we transfected mouse ES cells harboring single-copy sGEJ reporter at the *ROSA26* locus with expression plasmids for Cas9 D10A-sgRNAs, ABE-sgRNAs, or CBE-sgRNAs, we found that Cas9 D10A-gBE1, Cas9 D10A-gBE2 and Cas9 D10A-gBE1/gBE2 induced minimal levels of GFP^+ cells at 0.056%, 0.260%, and 0.190%, respectively (**Figure 3b**). As reported previously, Cas9 D10A-sgRNA could cause a low level of indels due to conversion of DNA nicks to DSBs by DNA replication and correction of the GFP reading frame by additional frame-shift[17]. We also found that ABE guided by gBE1, gBE2 and gBE1/gBE2 performed base editing at the efficiency of 0.077%, 0.280%, and 0.586%, respectively, after indel-induced GFP^+ cells were subtracted (**Figure 3c**). CBE with gBE1, gBE2 and gBE1/gBE2 had a base editing efficiency at 0.524%, 0.798% and 1.792%, respectively, after indel-induced GFP^+ cells were subtracted (**Figure 3d**). Importantly, ABE or CBE with gBE1/gBE2 had a higher efficiency of base editing than the base editor with either gBE1 or gBE2 alone (**Figure 3c,d**). ABE with gBE1/gBE2 increased the base editing efficiency by nearly 8-fold over ABE with gBE1 and 2-fold over ABE with gBE2; CBE with gBE1/gBE2 increased the base editing efficiency by over 3-fold as compared to CBE with gBE1 and over 2-fold as compared to CBE with gBE2 (**Figure 3c,d**). These results suggested that our gap-driven base editing with dual base editors could significantly elevate the efficiency of base editing. While dual base editors improve base

editing by gap generation, the level of base editing remains low at this targeted site likely due to the initial poor position for base editing. Latest results from the team members revealed that gap-driven base editing with dual CBEs increased the efficiency of C to T transition to 20% from maximal 10% by single CBE when the base editing window was proper (personal communication with Yao Rui and Feng Yili).

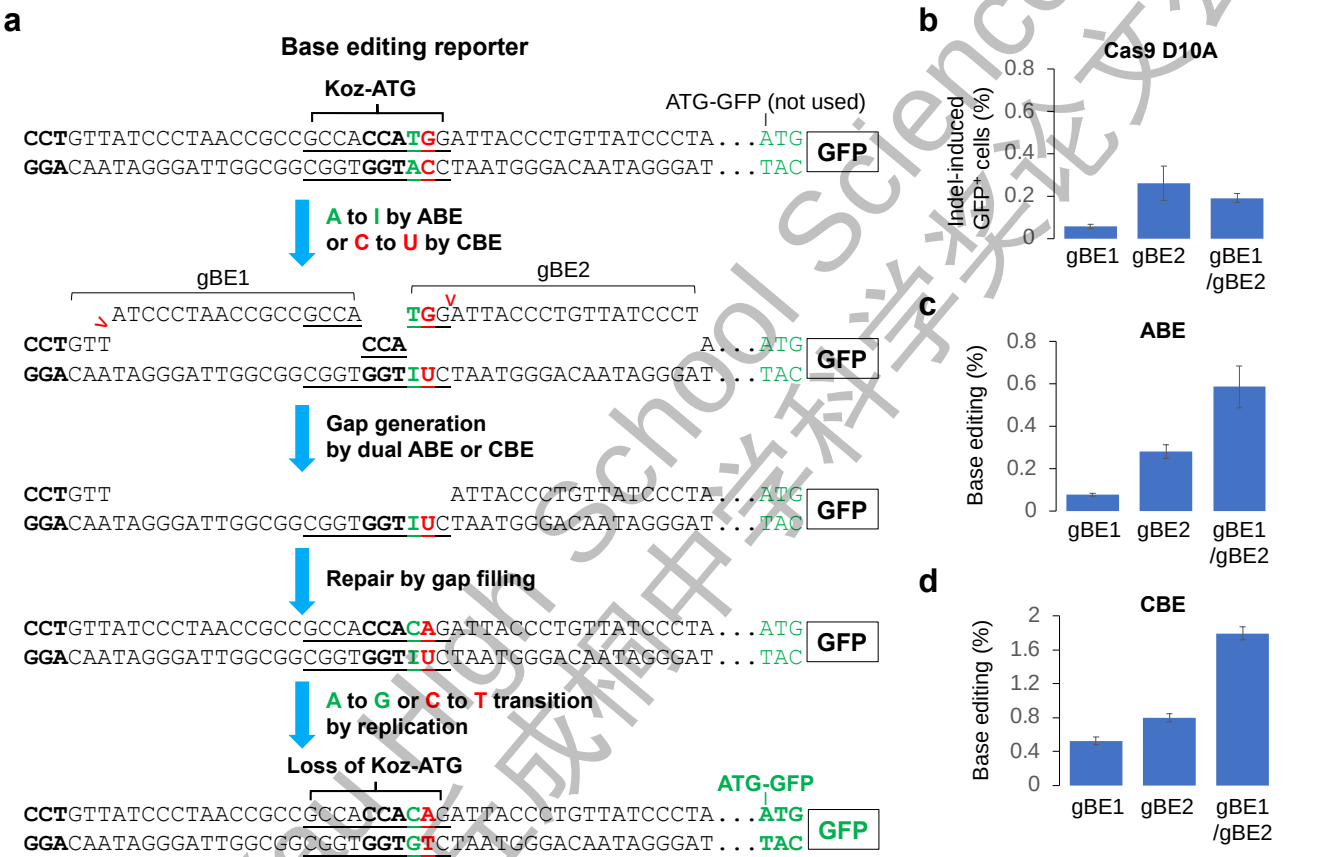

Figure 3 The gap-driven base editing system improves efficiency of base editing. (**a**) Schematic of the base editing reporter for gap-driven base editing with dual base editors. Base editing inactivates Koz-ATG, allowing ATG-GFP to work normally. (**b**) Indel-induced GFP$^+$ cells by Cas9 D10A-gBE1, Cas9 D10A-gBE2, and Cas9 D10A-gBE1/gBE2. (**c**) The efficiency of base editing by ABE-gBE1, ABE-gBE2, and ABE-gBE1/gBE2. (**d**) The efficiency of base editing by CBE-gBE1, CBE-gBE2 and CBE-gBE1/gBE2. Columns and error bars represent the average of triplicates and S.E.M. in an independent experiment, respectively, in **b-d**.

4. Discussion

In this study, we designed a gap-driven base editing system using a pair of base editors. The *in vitro* experiment demonstrated the feasibility of this approach, as evidenced by the detection of a 984-bp substrate through electrophoresis in groups containing Cas9 D10A guided by paired sgRNAs. A dye-labeled probe can fully complement with the nontarget strand within the gap induced by paired Cas9 D10A-sgRNAs, allowing the detection of the substrate. This finding indicates that paired Cas9 D10A-sgRNAs effectively induce a gap on the target strand, bypassing the need for additional nucleotide removal on the nicked strand. Consequently, this gap generation increases the likelihood of recognizing the deaminated strand in base editing as the repair template, leading to a significant improvement of base editing efficiency.

The predicted increase of base editing efficiency was further detected in our cell-based assays. We found a marked increase in editing efficiency for both gap-directed adenine base editors (Gap-ABEs) and gap-directed cytosine base editors (Gap-CBEs). These results underscore the substantial potential of gap-directed base editors, which offer several distinct advantages. First and foremost, the increased editing efficiency addresses a critical need in current base editing technologies. The application of base editing in disease treatment has been hindered by its low efficiency. By improving base editing efficiency, our study not only establishes a more powerful base editing strategy but also paves the way for further advancements in the field. Notably, our work extends the practice beyond the traditional single Cas9 D10A protein, opening up numerous possibilities to reprogramme the architecture of base editors for specific purposes. The concept of introducing a gap could be further explored in various fields, including the tracking of Cas9 D10A behavior. Moreover, the gap induced by paired base editors leads to an expansion of single-stranded regions, effectively broadening the base editing window. This expanded window holds promise for treating genetic diseases caused by point mutations beyond the editing scope of existing base editors[8].

However, this study also has limitations. During transfection, we encountered methodological challenges. We used three GFP control groups to estimate the transfection efficiency of different experimental groups (Cas9 D10A-sgRNA, ABE-sgRNA, and CBE-sgRNA). Since the GFP control group was separate, the calculated transfection efficiency serves as an approximation rather than an exact value. Additionally, using only three groups to infer the transfection efficiency across nine groups lacks specificity, which could affect the accuracy of absolute indel occurrence calculations in the D10A groups.

Despite demonstrating the feasibility and effectiveness of gap-directed base editors, many questions remain unclear due to the technological constraints. For instance, how do paired Cas9 D10A-sgRNAs communicate before and after gap generation? Can we further significantly improve base editing using paired based editors if the base editing efficiency by single base editor is already high? We also need more evidence to prove the validity and significance of the gap-driven base editing system in improving base editing technology. Additionally, our study was performed in a single mammalian cell line at one site. This restricts the generalizability of our findings. Furthermore, the study was based on a single independent experiment with triplicates. We need to perform at least three independent experiments in order to determine the statistical significance of our strategy and our findings.

Nevertheless, this study provides a proof of concept and potential of the gap-driven base editing system. Moving forward, Dr. Feng and other lab members will expand the scope of this research, and I will continue to communicate with them. If this technique proves successful in broader applications, the promising advancements highlighted in this study could become a reality, as our findings suggest a high promise of success.

References

1. Genomics Education Programme. Base editing: What is it and what does it mean for healthcare? https://www.genomicseducation.hee.nhs.uk/blog/base-editing-what-is-it-and-what-does-it-mean-for-healthcare/. (2023).
2. Lander, E. S. The heroes of CRISPR. *Cell* **164**, 18–28 (2016).
3. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).
4. Hsu, P. D., Lander, E. S. & Zhang, F. Development and applications of CRISPR-Cas9 for genome engineering. *Cell* **157**, 1262–1278 (2014).
5. 冯依力，陈若丹 & 谢安勇. “藕断丝连”的 CRISPR/Cas：基因编辑中靶点滞留的作用与挑战. 生物化学与生物物理进展 (2024). 电子版提前出版. doi: 10.16476/j.pibb.2024.0274.
6. Gaudelli, N. M. *et al.* Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature* **551**, 464–471 (2017).
7. Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A. & Liu, D. R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420–424 (2016).
8. Fan, X., Lei, Y., Wang, L., Wu, X. & Li, D. Advancing CRISPR base editing technology through innovative strategies and ideas. *Sci China Life Sci* (2024). Advance online publication. doi:10.1007/s11427-024-2699-5.
9. Nishimasu, H. *et al.* Engineered CRISPR-Cas9 nuclease with expanded targeting space. *Science* **361**, 1259–1262 (2018).
10. Thuronyi, B. W. *et al.* Continuous evolution of base editors with expanded target compatibility and improved activity. *Nat Biotechnol* **37**, 1070–1079 (2019).

11. Komor, A. C. *et al.* Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity. *Sci Adv* **3**, eaao4774 (2017).
12. Jeong, Y. K., Song, B. & Bae, S. Current status and challenges of DNA base editing tools. *Mol Ther* **28**, 1938–1952 (2020).
13. Feng, Y.-L. *et al.* H2AX facilitates classical non-homologous end joining at the expense of limited nucleotide loss at repair junctions. *Nucleic Acids Res.* **45**, 10614–10633 (2017).
14. Tu, T. *et al.* A precise and efficient adenine base editor. *Mol Ther* **30**, 2933–2941 (2022).
15. Koblan, L. W. *et al.* Improving cytidine and adenine base editors by expression optimization and ancestral reconstruction. *Nat Biotechnol* **36**, 843–846 (2018).
16. Puget, N., Knowlton, M. & Scully, R. Molecular analysis of sister chromatid recombination in mammalian cells. *DNA Repair (Amst)* **4**, 149–161 (2005).
17. Feng, Y.-L. *et al.* DNA nicks induce mutational signatures associated with BRCA1 deficiency. *Nat Commun* **13**, 4285 (2022).
18. Xie, A., Kwok, A. & Scully, R. Role of mammalian Mre11 in classical and alternative nonhomologous end joining. *Nat. Struct. Mol. Biol.* **16**, 814–818 (2009).
19. Willis, N. A. & Scully, R. Measurement of homologous recombination at stalled mammalian replication forks. *Methods Mol Biol* **2153**, 329–353 (2021).

Acknowledgements

On my birthday last year, I received the book *A Crack in Creation: Gene Editing and the Unthinkable Power to Control Evolution*, a book written by the 2020 Nobel Laureate in Chemistry Jennifer Doudna. The book introduces CRISPR/Cas9 gene-editing technology and discusses the implications of this invention through clear and concise demonstration. Professor Doudna mentioned that this technology is a breakthrough in the treatment of diseases, but it also raises many ethics-related debate and thought-provoking issues. After reading the book, despite having not yet learned biology in school, I started to learn more about this incredible invention. I was especially attracted by the content presented in documentary *Unnatural Selection*. In this documentary, people explored the potential and risks of gene editing, and I profoundly felt the significance of this technology for humanity. In recent years, as CRISPR/Cas9 has been expanded and applied in multiple gene editing technologies including base editing and prime editing, people have been seeking ways to improve base editors. One advancement is to create a nick on the target strand while the deaminase is working on the nontarget strand, thereby increasing editing efficiency. I wondered: Could two D10A complexes be used to create a fragment loss on the target strand, thus improving editing efficiency? If successful, this technique could also expand the editing window, providing more effective treatments for specific genetic diseases. I have described the significance of this research in the introduction, and this is what motivated me to study base editing systems among many interesting topics.

To clarify my contributions, I was responsible for the writing of this paper and the majority of the experimental work. Except for the *in vitro* experiment that is closely related to the testing of the gap-driven base editing system, I independently conducted all mentioned experimental procedures and data analyses. The final experimental data and results are the demonstration of my own work. The *in vitro* experiments were primarily conducted by Mr. Liu Sicheng, a member of our research team.

In this research, I received tremendous help from my two instructors and all the staff in the lab. Ms. Chen Lin, a chemistry and science competition instructor at Shanghai High School International

Division, was my project guidance teacher. She helped me understand the competition-related requirements and assisted me in planning the research and paper writing process. Mr. Feng Yili, a professor at the Institute of Translational Medicine at Zhejiang University, was my research instructor. He provided meticulous guidance on the research topic, theoretical knowledge, experimental design, data analysis, and paper writing. Meanwhile, I am also very grateful to the team member Yao Rui for his guidance in theoretical knowledge and experimental operations, and to Mr. Liu Sicheng for his assistance in the *in vitro* experiment. All the tutorial work mentioned above are unpaid but the study is supported by the National Natural Science Foundation of China (32071439 and 32370774 to Prof. Feng Yili) and I would like to express my heartfelt thanks to all the team members once again. Under their guidance, I have learned a lot about the attitude and techniques required for scientific research, and I will carry this spirit with me as I continue to explore life and science in the future.

致谢

在我去年生日的时候我收到了一份特殊的礼物，是2020年诺贝尔化学奖获得者Jennifer Doudna所写的《A Crack in Creation: Gene Editing and the Unthinkable Power to Control Evolution》一书，里面讲述了她对于CRISPR/Cas9基因编辑技术的介绍，以及她对于这一项发明的讨论。她提到，这项技术对于疾病的救治是突破性的发现，但这项技术也带来了许多值得思考的问题。看完这本书后，还未学习生物的我就会去寻找各种有趣的纪录片观看，令我印象最深刻的就是《Unnatural Selection》。人们在这一纪录片中探讨基因编辑的潜力与隐患，而我也深刻地感受到了这项技术对于人类的意义。近些年，人们一直在寻找改善碱基编辑器的方式。其中一项进展是在脱氨酶工作的同时在靶标链引导缺口，从而使编辑效率增高。所以，能不能用两个D10A蛋白在靶标链制造一个间隙，以此来提高编辑效率？如果成功的话，该技术还可以扩大编辑的窗口，为治疗特定的基因疾病提供更有效的治疗手段。我在引言的部分已经介绍了该研究的潜质，而这也驱使着我在众多有趣的课题中选择了碱基编辑器。

本次研究的论文全篇由本人独立撰写，所提到的实验也均有参与。除体外实验外，所有实验操作和数据分析均由本人独立完成，最终的实验数据和结果也均为本人成果。体外实验主要由实验室工作人员和课题成员刘嗣诚老师进行操作。

在本次研究中，我的两位指导老师和实验室的所有工作人员都给予了我莫大的帮助。上海中学国际部化学和科创竞赛指导老师陈琳老师是我的科创活动指导老师，她帮助我了解了比赛相关事项并帮助我规划比赛进程和最终的论文撰写。浙江大学转化医学研究院的研究生导师冯依力老师是我的学术指导老师。他对本次研究的选题，知识理论，实验设计，数据分析，和论文撰写方面进行了悉心指导。与此同时，我也非常感谢课题成员和实验室技术人员姚睿在生物理论和实验操作上的指导和实验室工作人员刘嗣诚老师在体外实验上的帮助。以上所提到的均为无偿指导，但课题研究得到国家自然科学基金面上项目（项目编号：32071439和32370774；负责人：冯依力老师）的支持。我也再次向所有的工作人员表示衷心的感谢。在各位的指导下我学到了很多关于科研的态度和技巧，之后我也会带着这份精神继续探索下去。