

METHOD

Open Access



sgBE: a structure-guided design of sgRNA architecture specifies base editing window and enables simultaneous conversion of cytosine and adenosine

Yanhong Wang[†], Lifang Zhou[†], Rui Tao[†], Nan Liu, Jie Long, Fengming Qin, Wenling Tang, Yang Yang, Qiang Chen and Shaohua Yao^{*}

* Correspondence: shaohuayao@scu.edu.cn

[†]Yanhong Wang, Lifang Zhou and Rui Tao contributed equally to this work.

Laboratory of Biotherapy, National Key Laboratory of Biotherapy, Cancer Center, West China Hospital, Sichuan University, Renmin Nanlu 17, Chengdu 610041, Sichuan, China

Abstract

We present a base editing system, in which base editors are attached to different sites of sgRNA scaffold (sgBE). Each independent sgBE has its own specific editing pattern for a given target site. Among tested sgBEs, sgBE-SL4, in which deaminase is attached to the last stem-loop of sgRNA, yields the highest editing efficiency in the window several nucleotides next to the one edited by BE3. sgBE enables the simultaneous editing of adenine and cytosine. Finally, in order to facilitate in vivo base editing, we extend our sgBE system to an AAV-compatible Cas9, SaCas9 (*Staphylococcus aureus*), and observe robust base editing.

Introduction

The clustered regularly interspaced short palindromic repeats (CRISPR)-based genome editing tools have shown significant successes in basic biomedical research and also provided great promise in clinical translation [1–3]. Recently, the development of base editing technique has enabled the conversion of one target DNA base pairs to another (C/G→T/A or A/T→G/C) in an efficient and irreversible way without causing robust double-strand breaks (DSBs), which significantly reduced off-target effects [4, 5] and made it practicable to correct genetic lesions of point mutation in inherited diseases [3].

Basically, base editors consist of a catalytically impaired Cas9 protein tethered with cytidine or adenosine deaminases that are active on ssDNA substrates. Cas9 binds a genomic locus of interest through the guidance of single sgRNA to form a protein-RNA-DNA ternary “R-loop” complex, in which the nontarget strand of sgRNA (NTS) is partially detached from the complex [6]. The exposed NTS provided a feasible substrate for deaminases to action. The most frequent way to tether deaminases is fusing them, through various linkers, to the N-terminus of Cas9 protein [4, 7]. Base editors



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

(BEs) with such architecture, including BE3, BE4, and ABE7.10, usually catalyzes the conversion of bases in a special region within NTS, called editing window, which is ~ 5 nt in width and ~ 15 nt upstream the protospacer adjacent motif (PAM) [2, 4, 5].

Although base editing technology has achieved significant success in a variety of basic biomedical researches [8–11], there are several limitations of current base editors standing in the way to clinical translation. In particular, the fixed editing windows of current base editors and definite PAMs of Cas9s employed in these editors significantly limited their targetable sites [12]. Besides, current base editors are too large to be in vivo delivered by adeno-associated virus (AAV), one of the most efficient gene therapy vectors [13, 14]. Therefore, there is a growing need for diversifying base editing tools capable of making designed changes in any given position and compatible with AAV delivery system.

Here, we sought to construct a base editing system in which deaminases were attached to the Cas9 complex through the interaction of MS2 RNA and its binding protein, MS2 coating protein (MCP) [15–17]. Previous studies had revealed that sgRNAs tagged with MS2 RNA were able to recruit various modulators to the Cas9 complex, such as transcriptional regulators, epigenetic modifiers, and cytosine deaminases [18–20]. In this platform, MS2 RNA tags are fused to sgRNA on its 1st and 3rd stem-loops, whereby they recruit these modulators to the Cas9 complex through fusing them to MCP. While this platform is efficient in a wide range of actions, including transcriptional regulation and visualization of local chromatin, it was less efficient in base editing [19]. This is possibly due to the fact that unlike other systems, the base editing system requires direct interaction between deaminases and their substrate, the NTS. Therefore, the relative location of deaminases to NTS is a key factor determining the scope and efficiency of base editing. According to a recently published Cryo-EM structure of the Cas9/DNA/sgRNA complex (PDB 5y36) [21], which for the first time presented the scenario with detached NTS included, each stem-loop of sgRNA scaffold had distinct positions relative to NTS. We designed a series of base editors in which deaminases were tethered to different sites of sgRNA scaffold and thereby were placed to different locations relative to NTS. In a set of target sites, these base editors generated distinct editing patterns. Importantly, one of these editors, sgCBE-SL4, in which deaminase was attached to the last stem-loop of sgRNA, yielded comparable editing efficiency to BE3, the most popular base editor, with an editing window several nucleotides next to the one edited by BE3s. By extending the sgBE with other Cas9 protein and cytosine deaminases, we further expanded the sgBE toolbox with AAV compatibility and diversified editing window.

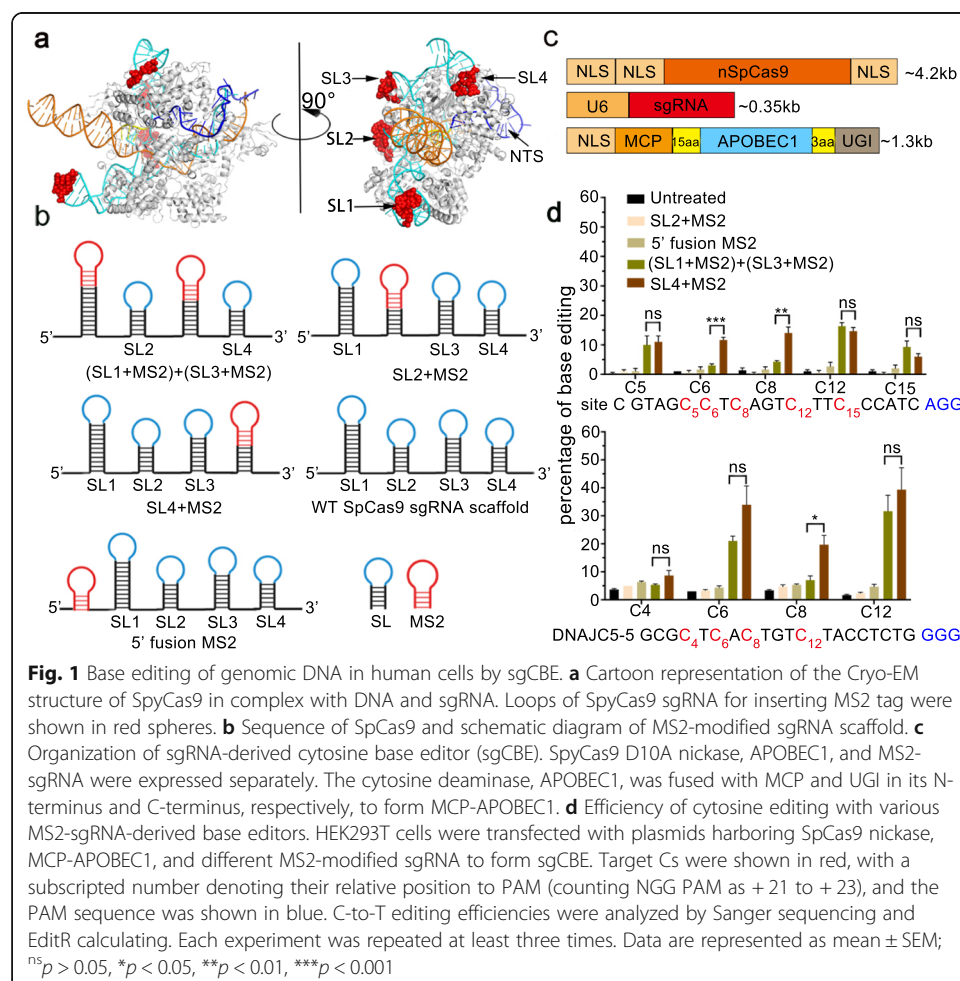
Results

Design and characterization of cytosine base editor with MS2-MCP-tethered APOBEC1 cytosine deaminase (sgCBE)

To test our hypothesis that tethering of deaminase onto sites close to the NTS would increase the editing efficiency and expand the window, we searched for candidate tethering sites by analyzing the 3D structure of the Cryo-EM structure of the Cas9/DNA/sgRNA complex [21]. We found that the 5' spacer region and the 4th stem-loop were much closer to the NTS than the 1st or the 3rd stem-loop. The 2nd stem-loop located

on the other side opposite to NTS was even more distant to the NTS than the 1st or the 3rd stem-loop (Fig. 1a).

We fused MS2 RNA tag to these sites respectively to generate a series of MS2 tagged sgRNAs, including SL2+MS2, SL4+MS2, 5' fusion MS2, and (SL1+MS2)+(SL3+MS2) sgRNAs (Fig. 1b). Then, we fused cytosine deaminase, APOBEC1, and uracil-DNA glycosidase inhibitor (UGI) with MS2 RNA-interacting protein (MCP) (Fig. 1c). A previously developed base editing tool, CRISPR-X [19], used (SL1+MS2)+(SL3+MS2) sgRNA recruit AID to dead SpCas9 (dSpCas9) but showed very limited activity, which was possibly due to dSpCas9 lacking the activity to nick the target DNA strand. We compared the activities of dCas9 and nickase SpCas9 (nSpCas9) in the case of (SL1+MS2)+(SL3+MS2) sgRNA. As expected, nSpCas9 performed much better than dSpCas9, as detected by Sanger sequencing and subsequent EditR analysis (Additional file 1: Fig. S1) [22]. EditR is a calculating program developed to identify and quantify base editing from fluorescent Sanger sequencing paragraph. Due to the presence of base-line noise in Sanger sequencing results, the detection limit for EditR program is usually above 5%. Therefore, only efficiencies above 5% were considered to have been effectively edited.



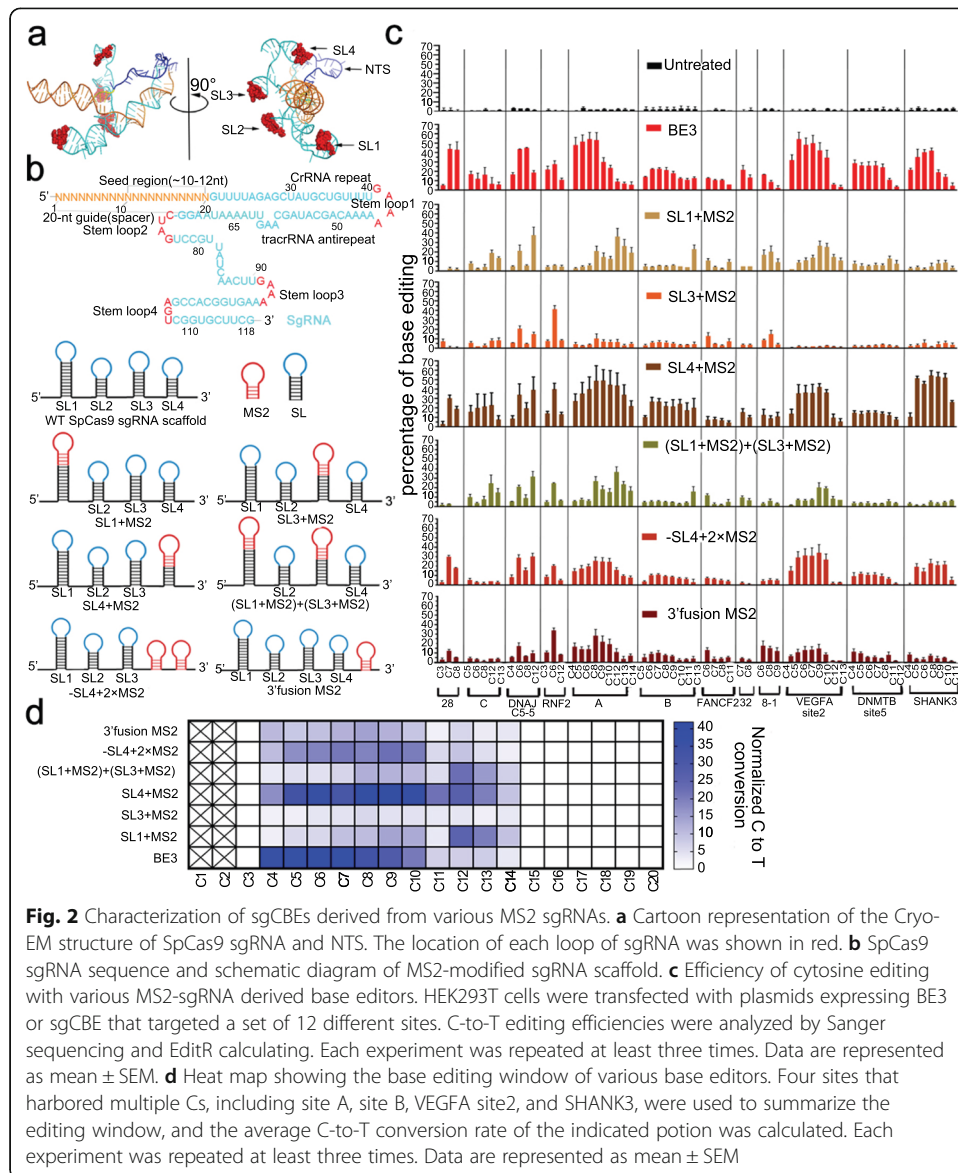
Next, we tested the performance of those newly designed sgRNAs. Co-transfection of these MS2-sgRNA, MCP-APOBEC1, and nickase SpCas9 constructs revealed that only SL4+MS2 and (SL1+MS2)+(SL3+MS2) sgRNAs produced obvious base editing. As shown in Fig. 1d, neither 5' fusion MS2 nor SL2+MS2 sgRNA produced detectable editing. Compared to the most popular MS2-sgRNA, (SL1+MS2)+(SL3+MS2) sgRNA and SL4+MS2 sgRNA produced much more robust editing in both sites tested. On average, editing efficiency in each targetable cytosines by SL4+MS2 sgRNA was ~40% higher than that by (SL1+MS2)+(SL3+MS2) sgRNA (16.14% vs 11.55%) (Fig. 1d and Additional file 1: Fig. S2), suggesting that reducing the distance between deaminase and the NTS would increase the editing efficiency.

In Fig. 1d, we also noticed that for (SL1+MS2)+(SL3+MS2) sgRNA, the editing efficiency in each different cytosines significantly varied. It seemed that (SL1+MS2)+(SL3+MS2) sgRNA had two editing peaks, one around position 5 and the other around position 12 (counting NGG PAM as 21–23), and cytosines in-between these two peaks were less efficiently edited. The variation in editing efficiency of each position was unlikely due to motif preference of the APOBEC1, because SL4sgRNA produced extensive editing in those positions that were poorly edited by (SL1+MS2)+(SL3+MS2) sgRNA.

A systematic examination of the effects of tethering sites on base editing behaviors

Inspired by the observation that tethering deaminase to different stem-loops of SpCas9 sgRNA endowed base editors with different editing patterns (editing efficiency and window), we systematically designed a series of MS2-sgRNAs, in which MS2 was fused to different stem-loops to form additional MS2-sgRNAs, including SL1+MS2, SL3+MS2, -SL4+2×MS2, and 3' fusion MS2 (Fig. 2a, b). We then co-transfected those MS2-sgRNAs targeting a set of 12 sites together with SpCas9 nickase and MCP-APOBEC1 constructs into HEK293T cells and examined the editing patterns of each MS2-sgRNA by Sanger sequencing and EditR analysis. In consistence with previous observations, MS2 tagged to SL4 or SL4+MS2 sgRNA had the most efficient base editing that was close to the level of BE3.

Besides overall editing activity, we noticed that for a given target, those sgCBEs generated different editing patterns (editing window and editing efficiency in each cytosine), all of which are quite different from the ones generated by BE3, the most popular base editing tool (Fig. 2c and Additional file 1: Fig. S3). Generally, BE3 showed an editing window ranging from C3 to C8, with peak positions at C5 to C7 [4]. Compared to BE3, SL4+MS2 sgRNA produced a slightly wider editing window, with peak positions at C5 to C10. The editing windows of -SL4+2×MS2 and 3' fusion MS2 were similar to that of SL4+MS2 sgRNA, yet with lower efficiencies (Fig. 2c, d). SL1+MS2 and (SL1+MS2)+(SL3+MS2) showed irregular editing windows that were narrower than those of BE3 or SL4+MS2. The editing patterns of SL1+MS2 and (SL1+MS2)+(SL3+MS2) were similar on most target sites except for site 28. On this site, SL1+MS2 had very limited activity. Interestingly, on the same site, SL3+MS2 had similar editing activity and pattern to (SL1+MS2)+(SL3+MS2), suggesting that both SL1 and SL3 contributed to the action of (SL1+MS2)+(SL3+MS2). Together with the observations in Fig. 1d, these data confirmed the notion that different tethering sites of



deaminase produced different editing patterns, which was particularly meaningful in situations where the target cytosines were located in complicated circumstances.

To confirm the accuracy of the above data determined by EditR, we conducted a high-throughput sequencing analysis on a set of 5 sites, including sites 28, 30, C, DNAJC5-5, and RNF2. In agreement with a previous report [22], the results of HTS were consistent with those obtained from EditR analysis (Additional file 1: Fig. S4). Next, we sought to replace MS2-MCP with another RNA-protein interaction system, BoxB-λN22 [23]. We found that sgCBEs modified with BoxB had a similar editing pattern and activity to the ones modified with MS2 (Additional file 1: Fig. S5).

Base editing involves C to U conversion in the NTS and ssDNA strand break in the target strand (TS), and those genomic lesions are fixed by the DNA repair system,

occasionally resulting in indel formation and base conversions besides C to T (C to R, R=A or G). Next, we sought to study the purity of C to T conversions in the editing products of the above 5 sites that had been examined by HTS. The HTS analysis revealed that all base editors induced indel formations across 5 sites, and sgCBEs induced less indels than BE3 did in 4 out of 5 sites (Additional file 1: Fig. S6). The sgCBEs also induced less C to R conversions than BE3 (Additional file 1: Fig. S6). These results suggested that sgCBEs performed better than BE3 in terms of product purity.

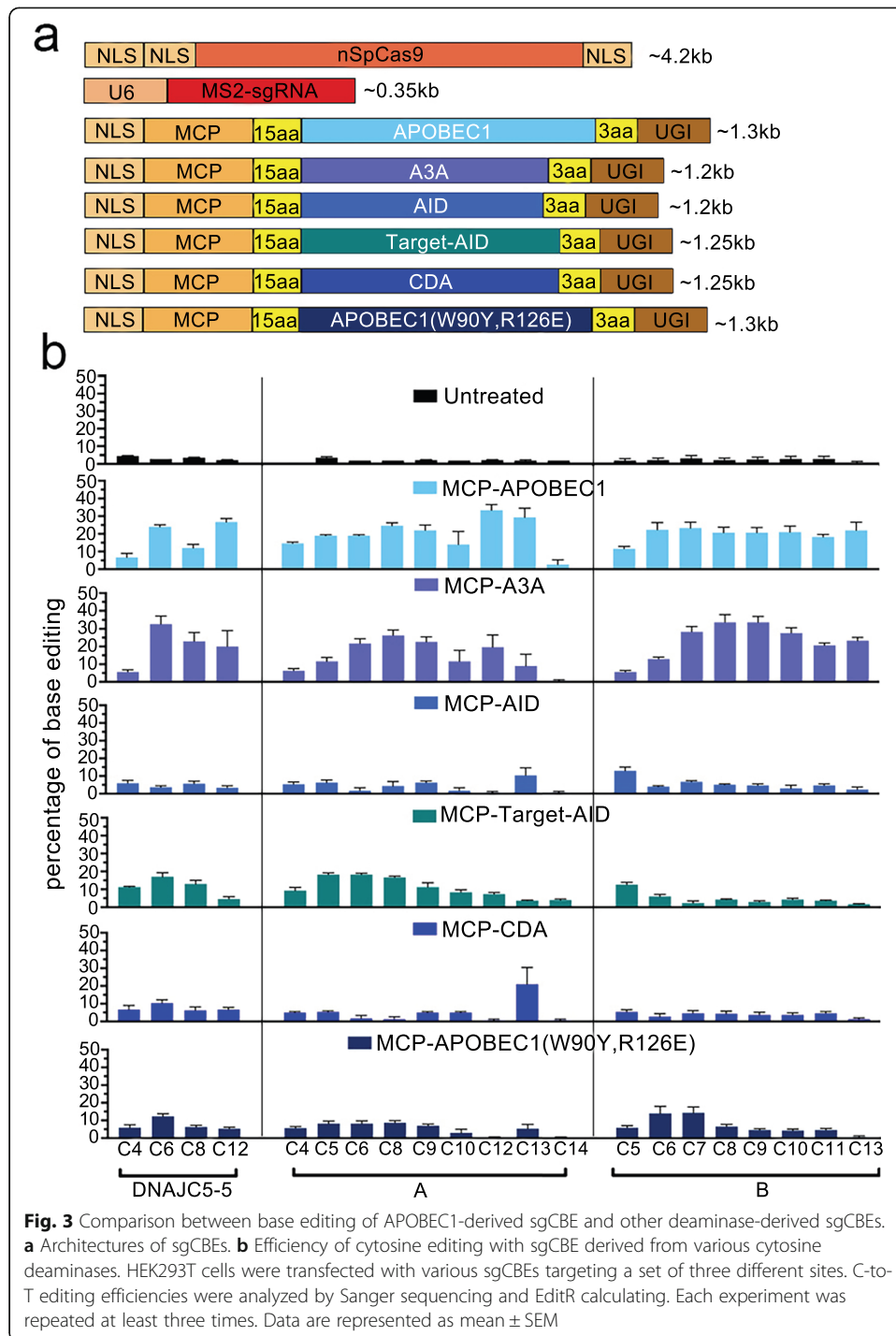
Previous studies have found that CBEs have off-target DNA editing in both sequence-dependent and sequence-independent manners. To define the specificity of sgCBEs, 3 on-target sites were selected for sequence-dependent off-target analysis. A total of 11 off-target sites (OTSs) were predicted with an online program (Table S5), all of which contained Cs within the editing window of BE3 or sgCBEs. HTS results showed that in the majority of OTSs, the mutation rates of OTS induced by sgCBEs were comparable to those induced by BE3 (Additional file 1: Fig. S7).

Diversify the sgCBE editing tools with additional cytosine deaminases

Several eukaryote-derived cytosine deaminases had been demonstrated efficient in base editing, including CDA, AID, Target-AID, A3A, and APOBEC1 [4, 7, 24–26]. Among them, APOBEC1 from rat was most frequently used. Although there were very limited data of comparing the editing efficiency and scope between APOBEC1 and the rest deaminases, two recent reports suggested that at least A3A had different editing patterns to APOBEC1 [25, 27], even under the same settings, such as BE3 architecture which included a 16aa XTEN linker between deaminase and Cas9 protein plus one copy of UGI fused to the C-terminus of Cas9.

We investigated if we could further obtain additional pattern-specified editing tools, simply by changing the cytosine deaminases in our system, including A3A, AID, CDA, and Target-AID. When coupled with these deaminases to sgRNAs, through the interaction between SL4+MS2 and MCP, only A3A displayed comparable editing activity to APOBEC1. AID, Target-AID, and CDA did not generate robust base editing. Interestingly, in all the 3 target sites tested, A3A produced distinct editing patterns from APOBEC1 (Fig. 3b and Additional file 1: Fig. S8). Although both of them displayed a wide editing window, from C4 to C13, A3A seemed to prefer the cytosines located in the 5' portions of this window, as compared to APOBEC1. Then, we compared the editing patterns between APOBEC1 and A3A in other MS2-sgRNA settings. In consistence with the observations in SL4+MS2, A3A displayed different editing patterns from APOBEC1 when they were coupled to those MS2-sgRNAs (Additional file 1: Fig. S9). Together, these results suggested that the substitution of APOBEC1 with A3A further diversified the editing pattern of sgCBE.

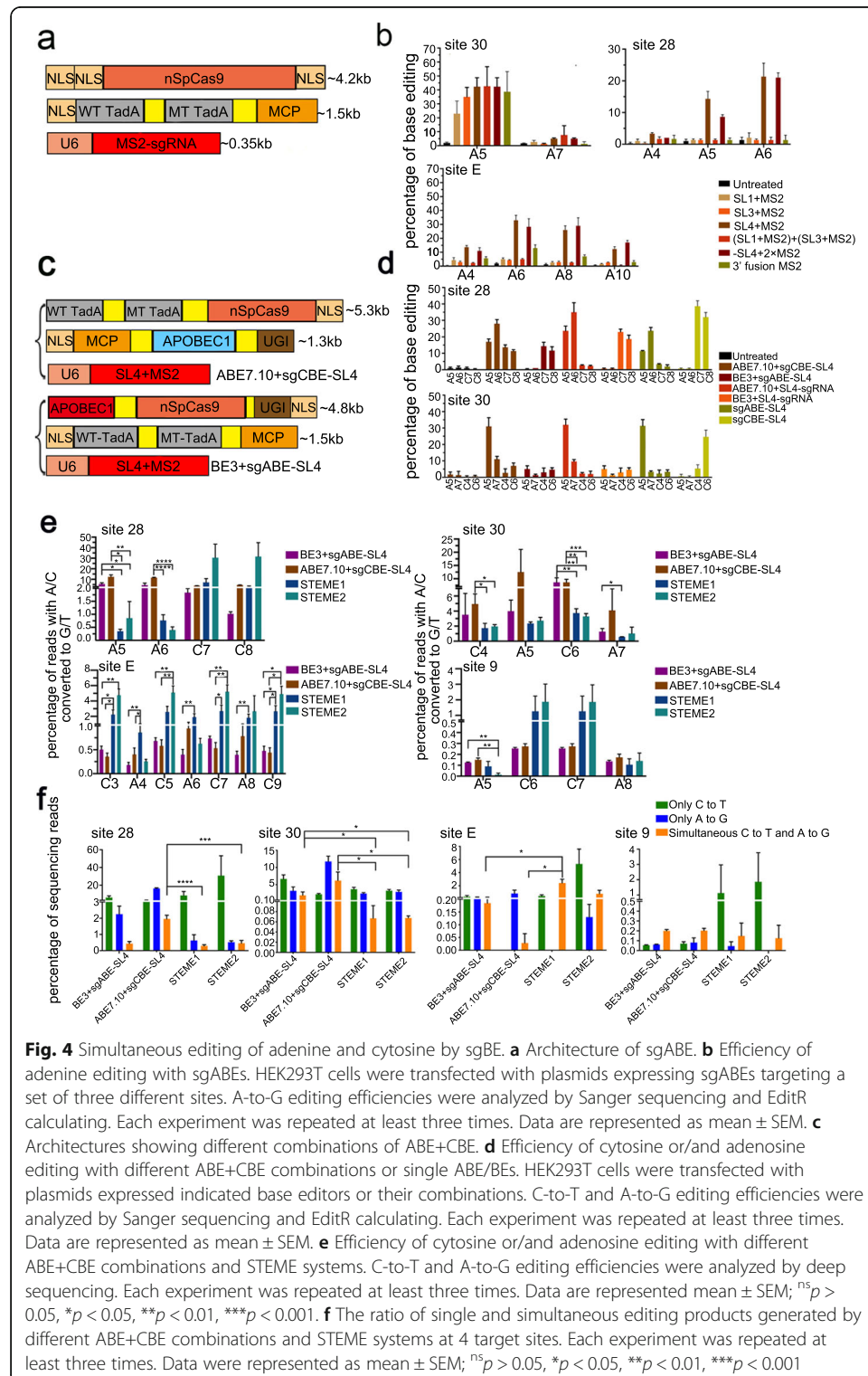
We also tested if we could narrow the editing window of each MS2-sgRNA by impairing the cytosine activity. We introduced 2 point mutations, W90Y and R126E, which had been previously demonstrated to be able to narrow the editing window of BE3-APOBEC1 [28, 29]. As shown in Fig. 3b, mutant APOBEC1 did mildly narrow the editing window in 2 out of 3 sites (site DNAJC5-5 and site B), albeit at the expense of lowering editing efficiency.



Design and characterization of adenosine base editor based on sgRNA-MS2/MCP system (sgABE)

After establishing the sgCBE system, we investigated if we could apply our MS2-sgRNA system to adenosine base editing (sgABE). We fused the ABE editor, the hetero-dimer of a wild-type TadA (WT TadA) and an evolved mutant TadA (MT TadA), to the N-terminus of MCP with a 32-aa linker to form TadA-MCP. Then, we co-transfected this construct together with a set of MS2-sgRNAs harboring varied MS2 fusion and different spacers. In all

three targets tested, SL4-derived sgABEs, including SL4+MS2 and -SL4+2×MS2, produced robust adenosine conversion (Fig. 4b and Additional file 1: Fig. S10). The rest of the sgABEs only produced robust adenosine editing in one target. These data were consistent with the observations in sgCBE, in which SL4-derived editors were more efficient than that derived from other SLs.



After establishing both sgCBE and sgABE systems, we sought to combine ABE7.10 or BE3 with our sgBEs to form dual base editors capable of simultaneously converting both cytosines and adenines in one target site (Fig. 4c). Co-transfection of ABE7.10 with SL4+MS2-derived sgCBE (sgCBE-SL4) generated detectable conversion of both bases in both target sites as determined by Sanger sequencing and EditR analysis, while the combination of BE3 and sgABE-SL4 did not (Fig. 4d and Additional file 1: Fig. S11). In comparison between the combinations (ABE7.10+sgCBE-SL4 or BE3+sgABE-SL4) and single-base editors (ABE7.10-SL4, BE3-SL4, sgABE-SL4, or sgCBE-SL4), we noticed that the combination reduced the editing efficiency of both types of editors. For example, in site 28, the combination of ABE7.10 and sgCBE-SL4 resulted in a 23.3% loss of adenosine editing compared with ABE7.10+SL4-sgRNA and 64.62% loss of cytosine editing compared with sgCBE-SL4. In the same site, the combination of BE3 and sgABE-SL4 also obviously reduced both cytosine and adenosine editing (37% loss for cytosine and nearly 100% loss for adenosine, as compared to BE3 and sgABE-SL4, respectively). These results suggested when targeting the same site, ABE and CBE might interrupt each other, which was possibly due to the fact that both editors could bind to ssDNA and thus competed with each other when they were put in the same target site. In consistence with our observations, a recent report using TadA and A3A fusion editors also observed dramatic inhibition of TadA activity [30].

In addition, we designed dual tag sgBEs using MS2 and BoxB RNA scaffolds to recruit cytosine and adenine deaminases (Additional file 1: Fig. S12). However, compared to the original strategy, dual tag sgBEs produced much lower editing (Fig. S12). Then, we focused on the original dual base editors for further characterization. To analyze the efficiency of simultaneous conversion, HTS was performed and additional targets were examined. The results showed that our dual base editors produced simultaneous conversion of A and C across all 4 targets, with an efficiency ranging from 0.2 to 8% (Fig. 4e, f). We also compared our dual base editors with previously published STEME in plants. In comparison, we found that all those dual base editors produced simultaneous A and C editing, yet the efficiency of which varied greatly dependent on target sites tested and the editors used (Fig. 4e, f). Overall, our dual base editors were less efficient than STEME in terms of C-to-T editing, but their activity of A-to-G and simultaneous editing was similar to or higher than STEME (Fig. 4e, f).

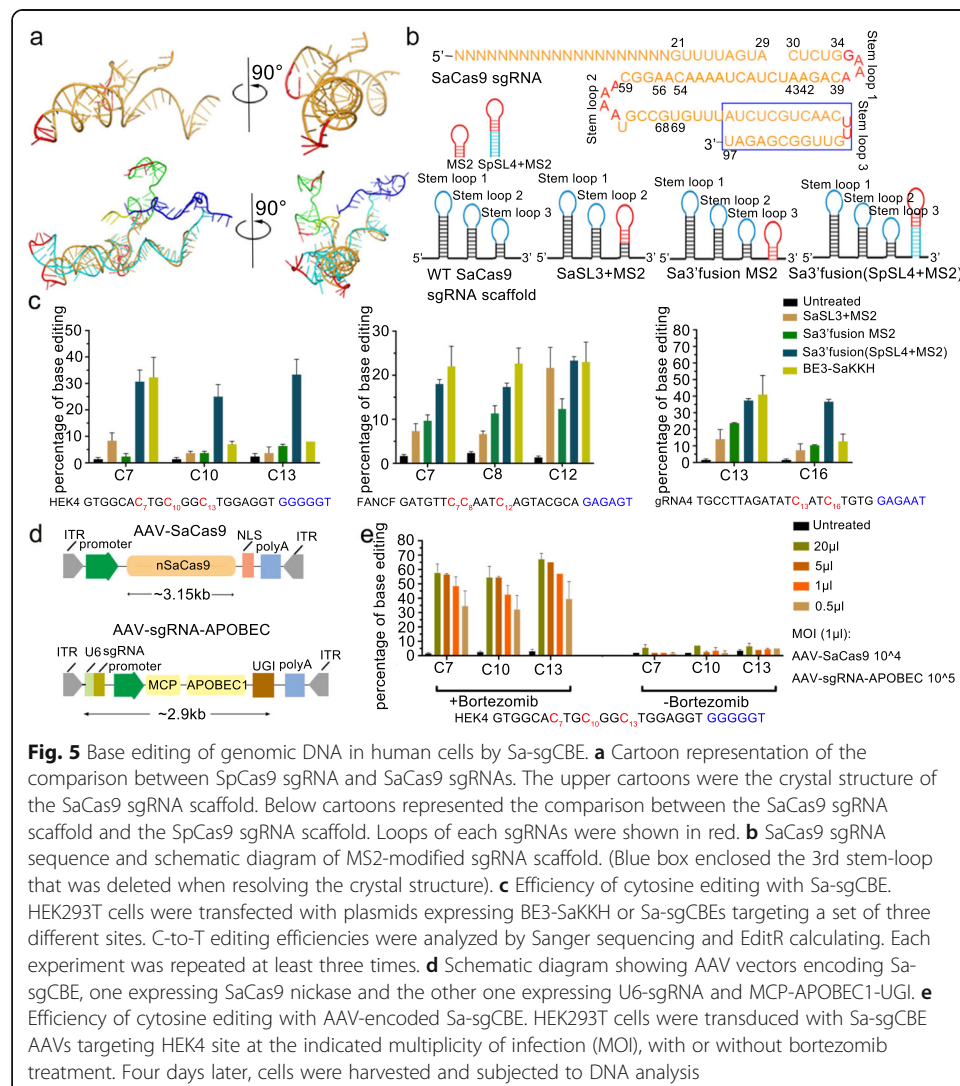
Expanding sgBE to SaCas9, an AAV-compatible CRISPR system

A critical obstacle towards the way of base editing to the therapeutic application is the large size of current base editors [14], especially SpCas9-derived base editors. Compared to SpCas9, SaCas9 has a much smaller size, ~ 3.2 kb, which makes it more commonly used than SpCas9 in in vivo delivery for therapeutic purposes, especially through the AAV delivery system [31, 32]. Therefore, developing a SaCas9-based sgBE system (Sa-sgCBE) is meaningful for translational research of base editing technology.

Because there is no available SaCas9 structure that included the NTS, we sought to design Sa-sgCBE following the parameters obtained from SpCas9. Firstly, we compared the structure of the SaCas9 complex (PDB:5axw) [33] with the one of SpCas9 with NTS included (PDB:5y36) [21]. In comparison, we found that although these Cas9s shared only 17% identical amino acids, their general structures were largely alike. The positions of sgRNA, the PAM region of the target DNA, the sgRNA/TS heteroduplex,

and various Cas9 domains in the SaCas9 complex were almost the same as those in SpCas9 (Additional file 1: Fig. S13). Importantly, the shape of SaCas9 sgRNA, especially the direction of its stem-loops, was also similar to that of SpCas9 sgRNA. However, the SaCas9 complex used to resolve the structure only included an sgRNA with 2 stem-loops, which lacked the 3rd stem-loop as compared to the ones that were frequently used in SaCas9-mediated gene editing experiments (Fig. 5a). A previous study in SpCas9 had revealed that sgRNAs with 3' stem-loop deleted would decrease gene editing activity [34]; therefore, we chose the 3-stem/loop SaCas9 sgRNA for subsequent designs.

Following the guidance of the architecture of SpCas9 SL4+MS2 sgRNA, which is the most efficient one, we designed a series of MS2-Sa-sgRNAs, including SaSL3+MS2, Sa 3'fusion MS2, and Sa 3'fusion (SpSL4+MS2) (Fig. 5b). Transfection of these MS2-Sa-sgRNAs that harbored various spacers revealed that all these sgRNAs produced detectable base editing, with an activity order: Sa3'fusion (SpSL4+MS2) > Sa3'fusion MS2 > SaSL3+MS2 (Fig. 5c and Additional file 1: Fig. S14). Notably, Sa3'fusion (SpSL4+MS2) sgRNA-derived cytosine editor (Sa-sgBE-SpSL4) had a similar editing



activity to BE3-SaCas9, with a 3' expanded editing window, which was in consistence with the observations in SpCas9-derived sgCBE-SL4. Next, in order to prove the notion that Sa-sgBE was compatible for AAV delivery, we constructed its elements into two AAV vectors (Fig. 5d), one expressing SaCas9 nickase and the other expressing MCP-APOBEC1-UGI and sgRNA. The resulting AAVs were used to transduce 293T cells. To improve the transduction, cells were treated with bortezomib, a commonly utilized proteasome inhibitor that was known to improve AAV transduction in multiple cell types [35]. As shown in Fig. 5e, our Sa-sgBE could be efficiently expressed by AAVs and produced robust base editing in 293 cells in the presence of bortezomib. In the absence of bortezomib, editing activity was significantly reduced, suggesting the improvement of AAV transduction by proteasome inhibition.

Discussion

Current base editors, in which deaminases were fused to the N-terminus of Cas9, such as BE3 and ABE7.10 [4, 5], catalyze the conversion of bases in an ~ 5-nt window that is ~ 15 nt upstream the PAM. This editing window is a double-edged sword. On one hand, widening the editing window will increase the targetable range. However, the expansion of the editing window will also increase the probability to introduce unwanted editing, which sometimes discounted or even disabled the aimed editing. On the other hand, narrowing the window will increase the editing precision while also lead to the inability to the target sites without a proper neighboring PAM. An ideal way to resolve this obstacle is to expand base editing toolbox with editors possessing diversified editing windows, especially considering those targets that were located in a complicated circumstance. These diversified editing tools would make it possible to optimize the editing strategy for the complicated targets.

Previous studies had demonstrated that translocation of deaminases by directly inlaying them into various internal sites of Cas9 protein lead to varied editing window [36, 37]. Although these strategies more or less impaired the editing activity, possibly due to the interruption of Cas9 integrity, they proved the concept that positioning deaminases to different regions of the RNP complex would affect the editing window. As mentioned above, according to the Cryo-EM structure of the SpCas9 RNP complex, each stem-loop of sgRNA had different localizations relative to NTS, which provided a possibility to design editing tools possessing varied editing window while keeping Cas9 intact. A systematic attachment of cytosine deaminase to all stem-loops revealed that SL1, SL3, and SL4 generated robust base editing. Importantly, these CBEs also had different editing windows. Therefore, these editors provided a possibility for optimizing editing strategy for some complicated targets. When wide editing range is required, such as mutating gene enhancers, SL4 may be a good choice, since it has a wider editing window. And when a single base is aimed and this base is surrounded by other editable bases, these editors could be screened for improving the relative editing of target base to unwanted bases. Taken site B locating at *DNMT3B* locus as an example (Additional file 1: Fig. S15), if the 438th amino acid is required to be converted from glutamic acid to lysine while keeping other potions unchanged, SL1 and SL1+SL3 would be the best choice. These data also proved a notion that reengineering the appending architecture of sgRNA scaffold could generate base editors with distinct editing features. Considering that the scaffold modifications described in this study are only simply grafting MS2 to different stem-loops of sgRNAs,

there is still room for designing more complicated and precise architectures to improve this toolbox. For example, the appending architectures can be designed to place MS2 or other tags to the site close to a specific NTS fragment, further improving the precision of the editing window. In addition, sgRNA architectures can also be designed to carry multiple RNA tags, such as PP7 and S1 [38, 39], which in turn recruit multiple players to perform complicated reactions, such as modulating DNA repair programs following base deamination. It should be noted that off-target editing of base editors has been observed in both RNA and DNA [40, 41]. Adenosine base editors induced mainly RNA off-target mutations, and cytosine base editors induced both RNA and DNA off-target mutations. Although we did not examine RNA and sequence-independent DNA off-target editing, considering that deaminases were being separately expressed and would be leaked from the binding of the Cas9 complex, sgBEs might have higher off-target editing of both types than BE3. Recently, novel versions of cytosine and adenosine deaminases with lower off-target editing were developed [42, 43], which provide a possible solution for reducing sgBEs' off-target editing.

In addition, sgBE system provided a strategy to simultaneously convert both adenosines and cytosines in a given window of the genome, which is particularly useful in conditions where a specific mosaic mutant pattern was required, especially those aimed to change bases within a small window, say within a 3–5-bp window. Besides, simultaneous conversion of adenosines and cytosines also makes base editing more powerful for the unbiased screening that targets a relatively limited DNA fragment, for example, the evolution of the antibody or fluorescence protein.

Finally, the sgBE system also provided a solution for in vivo delivery of base editors by AAV vectors, the currently most efficient in vivo delivery tool. The major obstacle for the in vivo applications of current base editors is its big size (~ 5.3 kb for SpCas9 and ~ 4.2 kb for SaCas9-derived editors), which is larger than or very close to the package limitation of AAV. One possible way to deliver such base editors with AAV is to split the base editors into two parts. This strategy often involves the split of Cas9 proteins into two expression cassettes and reconstituted by intein-mediated protein splicing [44–48], which potentially attenuated the editing efficiency. Our sgBE system does not require the split of Cas9 proteins and thus generates a similar editing efficiency as that of corresponding full-length base editors. The sgBE system can be in vivo delivered by two AAV vectors, one harboring Cas9 nickase cassette and the other harboring MCP-deaminase plus sgRNA cassettes.

Conclusion

We engineered the sgRNA scaffold of SpCas9 and used RNA/RNP interaction system to recruit cytosine or adenosine deaminases. The resulting sgBEs had diversified editing patterns, all of which were distinct to that of BE3. When coupled with traditional CBE or ABE, sgBEs enabled simultaneously conversion of both cytosine and adenosine in a single target. In addition, we also expanded these findings to SaCas9 to generate an AAV-compatible sgBE. In summary, this study developed a series of base editors with diverse editing features, which offer an opportunity for the optimization of editing tools when the target bases were located in complicated circumstances, and also provided a base editing tool for in vivo gene corrections.

Materials and methods

Plasmid construction

Plasmids encoding BE3, BE3-SaKKH (SaCas9 KKH variant), and ABE7.10 editors were obtained from addgene (#73021, #85170, #102919). The scaffold of the sgRNA tagged with MS2 or BoxB was synthesized by General Biosystems (Anhui) Co. Ltd. Sequences for each MS2-sgRNAs or BoxB-sgRNAs were listed in the Note S1. The corresponding spacers were inserted into each sgRNAs digested with Bbs1. Oligos used to generate sgRNA plasmids were listed in Additional file 1: Table S1. The rest of the plasmids, including MCP-deaminases, λN22-deaminases,, pAAV-CMV-saKKH (D10A), and pAAV-U6sgRNA-CMV-MCP-APOBEC1-UGI were constructed through the seamless cloning method (ClonExpress II One Step Cloning Kit. Vazyme Biotech Co. Ltd). All plasmids were verified by Sanger sequencing.

Cell culture

HEK293T were cultured in Dulbecco's modified Eagle's medium (Thermo Fisher Scientific), supplemented with 10% (v/v) fetal bovine serum (Life Technologies) and 1% penicillin/streptomycin (Boster Biological Technology Co. Ltd.), and maintained at 37 °C with 5% CO₂.

Plasmid transfection and AAV transduction

HEK293T cells were seeded on 96-well plates (BIOFIL). Cells at a confluence of ~70–80% were transfected with Transeasy™ (Forgene) according to the manufacturer's instruction. Seventy-two hours post-transfection, genomic DNA was extracted by the addition of 30 μl freshly prepared lysis buffer. The mixture was incubated at 55 °C for 10 min and then was heat-inactivated at 95 °C for another 10 min. The resulting DNA lysate was subjected to PCR amplification and subsequent analysis.

For packaging serotype 2 AAVs, pAAV-CMV-SaKKH (D10A) or pAAV-CMV-MCP-APOBEC1-U6sgRNA were co-transfected with pHelper and AAV2 Rep-Cap plasmids into HEK293 cells. The resulting AAVs were extracted from host cells by three cycles of freeze-thawing, then were purified by ultra-centrifuge. AAV titers were determined by qPCR with a pair of primers targeting CMV promoter, CMV forward: 5' CTGACCGCCCAACGACCC, CMV reverse: 5'CTGACCGCCCAACGACCC. For AAV transduction assay, cells at a confluence of ~70–80% were transduced with AAV2-CMV-SaKKH(D10A) and pAAV2-CMV-MCP-APOBEC1-U6sgRNA. Sixty-five nanomole bortezomib was added to the culture medium to increase AAV transduction as previously described [35]. Four days later, AAV-transduced cells were harvested for DNA analysis.

Base editing analysis with Sanger sequencing and EditR software

On-target genomic regions of interest were amplified by PCR and then were analyzed with Sanger sequencing. Then, the sequencing graphs were further quantified by EditR software (baseditr.com), according to the author's description. Primers used for amplifying each target loci are listed in Table S2.

High-throughput sequencing of genomic DNA samples

Cells were harvested at 72 h post-transfection, and the genomic DNA was extracted with freshly prepared lysis buffer. Genomic regions of interest were amplified by PCR with primers flanked with different barcodes (Table S3 and S4). The products were purified with DNA gel-extraction kit and quantified with NanoDrop (Thermo Fisher). Samples were sequenced commercially using the Illumina Novaseq6000 platform (Tsingke, Beijing and Personal Biotechnology, Shanghai, China).

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s13059-020-02137-6>.

Additional file 1. Integrated supplementary Figures, Tables and Note. Contains figures from S1 to S15, tables from S1 to S5 and Note S1.

Additional file 2. Review history.

Acknowledgements

We thank Li Mo and Yu Liu for their help in AAV production. We thank Dr. Lu Chen and Dr. Chao Tang for their help in deep-sequencing data analysis. We also thank the reviewers and the editor for their time and valuable feedback.

Peer review information

Yixin Yao was the primary editor of this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

Review history

The review history is available as Additional file 2.

Authors' contributions

Y. W., L. Z., R. T., N. L., J. L., F. Q., W. T., and Y. Y. performed the experiments. Y. W., L. Z., and Q. C. analyzed the data. Y. W. and S.Y. designed the experiments and wrote the manuscript. The author(s) read and approved the final manuscript.

Funding

This work was supported by the Hi-Tech Research & Development (863) Program of China (No. 2015AA020309), National Natural Science Foundation of China (No. U19A2002, No. 81771220, and No. 81974238), and Salubris Academician Workstation for Innovative Biopharmaceuticals (No. 2017B090904017).

Availability of data and materials

Deep-sequencing data are available under BioProject ID PRJNA655177 (<https://www.ncbi.nlm.nih.gov/bioproject/PRJNA655177>) [49].

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no conflicts of interest.

Received: 17 January 2020 Accepted: 7 August 2020

Published online: 28 August 2020

References

1. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*. 2012;337:816–21.
2. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, Zhang F. Multiplex genome engineering using CRISPR/Cas systems. *Science*. 2013;339:819–23.
3. Rees HA, Liu DR. Base editing: precision chemistry on the genome and transcriptome of living cells. *Nat Rev Genet*. 2018;19:770–88.
4. Komor AC, Kim YB, Packer MS, Zuris JA, Liu DR. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature*. 2016;533:420–4.
5. Gaudelli NM, Komor AC, Rees HA, Packer MS, Badran AH, Bryson DI, Liu DR. Programmable base editing of A*T to G*C in genomic DNA without DNA cleavage. *Nature*. 2017;551:464–71.
6. Anders C, Niewoehner O, Duerst A, Jinek M. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature*. 2014;513:569–73.

7. Ma Y, Zhang J, Yin W, Zhang Z, Song Y, Chang X. Targeted AID-mediated mutagenesis (TAM) enables efficient genomic diversification in mammalian cells. *Nat Methods*. 2016;13:1029–35.
8. Koblan LW, Doman JL, Wilson C, Levy JM, Tay T, Newby GA, Maianti JP, Raguram A, Liu DR. Improving cytidine and adenine base editors by expression optimization and ancestral reconstruction. *Nat Biotechnol*. 2018;36:843–6.
9. Zafra MP, Schatoff EM, Katti A, Foronda M, Breinig M, Schweitzer AY, Simon A, Han T, Goswami S, Montgomery E, et al. Optimized base editors enable efficient editing in cells, organoids and mice. *Nat Biotechnol*. 2018;36:888–93.
10. Chadwick AC, Wang X, Musunuru K. In vivo base editing of PCSK9 (proprotein convertase subtilisin/kexin type 9) as a therapeutic alternative to genome editing. *Arterioscler Thromb Vasc Biol*. 2017;37:1741–7.
11. Zhang Y, Qin W, Lu X, Xu J, Huang H, Bai H, Li S, Lin S. Programmable base editing of zebrafish genome using a modified CRISPR-Cas9 system. *Nat Commun*. 2017;8:118.
12. Zong Y, Gao CX. Progress on base editing systems. *Yi Chuan*. 2019;41:777–800.
13. Gao G, Vandenberghe LH, Wilson JM. New recombinant serotypes of AAV vectors. *Curr Gene Ther*. 2005;5:285–97.
14. Senis E, Fatouros C, Grosse S, Wiedtke E, Niopek D, Mueller AK, Borner K, Grimm D. CRISPR/Cas9-mediated genome engineering: an adeno-associated viral (AAV) vector toolbox. *Biotechnol J*. 2014;9:1402–12.
15. Uhlenbeck OC, Carey J, Romaniuk PJ, Lowary PT, Beckett D. Interaction of R17 coat protein with its RNA binding site for translational repression. *J Biomol Struct Dyn*. 1983;1:539–52.
16. Laprade H, Lalonde M, Guerit D, Chartrand P. Live-cell imaging of budding yeast telomerase RNA and TERRA. *Methods*. 2017;114:46–53.
17. Valegard K, Murray JB, Stonehouse NJ, van den Worm S, Stockley PG, Liljas L. The three-dimensional structures of two complexes between recombinant MS2 capsids and RNA operator fragments reveal sequence-specific protein-RNA interactions. *J Mol Biol*. 1997;270:724–38.
18. Konermann S, Brigham MD, Trevino AE, Joung J, Abudayyeh OO, Barcena C, Hsu PD, Habib N, Gootenberg JS, Nishimasu H, et al. Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature*. 2015;517:583–8.
19. Hess GT, Fresard L, Han K, Lee CH, Li A, Cimprich KA, Montgomery SB, Bassik MC. Directed evolution using dCas9-targeted somatic hypermutation in mammalian cells. *Nat Methods*. 2016;13:1036–42.
20. Pankert T, Jegou T, Caudron-Herger M, Rippe K. Tethering RNA to chromatin for fluorescence microscopy based analysis of nuclear organization. *Methods*. 2017;123:89–101.
21. Huai C, Li G, Yao R, Zhang Y, Cao M, Kong L, Jia C, Yuan H, Chen H, Lu D, Huang Q. Structural insights into DNA cleavage activation of CRISPR-Cas9 system. *Nat Commun*. 2017;8:1375.
22. Kluesner MG, Nedveck DA, Lahr WS, Garbe JR, Abrahamte JE, Webber BR, Moriarity BS. EditR: a method to quantify base editing from Sanger sequencing. *CRISPR J*. 2018;1:239–50.
23. Chattopadhyay S, Garcia-Mena J, DeVito J, Wolska K, Das A. Bipartite function of a small RNA hairpin in transcription antitermination in bacteriophage lambda. *Proc Natl Acad Sci U S A*. 1995;92:4061–5.
24. Komor AC, Zhao KT, Packer MS, Gaudelli NM, Waterbury AL, Koblan LW, Kim YB, Badran AH, Liu DR. 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*. 2017;3:eaa04774.
25. Zong Y, Song Q, Li C, Jin S, Zhang D, Wang Y, Qiu JL, Gao C. Efficient C-to-T base editing in plants using a fusion of nCas9 and human APOBEC3A. *Nat Biotechnol*. 2018;36:950–3.
26. Ren B, Yan F, Kuang Y, Li N, Zhang D, Zhou X, Lin H, Zhou H. Improved base editor for efficiently inducing genetic variations in rice with CRISPR/Cas9-guided hyperactive hAID mutant. *Mol Plant*. 2018;11:623–6.
27. Gehrke JM, Cervantes O, Clement MK, Wu Y, Zeng J, Bauer DE, Pinello L, Joung JK. An APOBEC3A-Cas9 base editor with minimized bystander and off-target activities. *Nat Biotechnol*. 2018;36:977–82.
28. Holden LG, Prochnow C, Chang YP, Bransteitter R, Chelico L, Sen U, Stevens RC, Goodman MF, Chen XS. Crystal structure of the anti-viral APOBEC3G catalytic domain and functional implications. *Nature*. 2008;456:121–4.
29. Kim YB, Komor AC, Levy JM, Packer MS, Zhao KT, Liu DR. Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. *Nat Biotechnol*. 2017;35:371–6.
30. Li C, Zhang R, Meng X, Chen S, Zong Y, Lu C, Qiu JL, Chen YH, Li J, Gao C. Targeted, random mutagenesis of plant genes with dual cytosine and adenine base editors. *Nat Biotechnol*. 2020;38:875–82.
31. Naso MF, Tomkowicz B, Perry WL 3rd, Strohl WR. Adeno-associated virus (AAV) as a vector for gene therapy. *BioDrugs*. 2017;31:317–34.
32. Lee JH, Kim Y, Yoon YE, Kim YJ, Oh SG, Jang JH, Kim E. Development of efficient adeno-associated virus (AAV)-mediated gene delivery system with a phytoactive material for targeting human melanoma cells. *New Biotechnol*. 2017;37:194–9.
33. Nishimasu H, Cong L, Yan WX, Ran FA, Zetsche B, Li Y, Kurabayashi A, Ishitani R, Zhang F, Nureki O. Crystal structure of *Staphylococcus aureus* Cas9. *Cell*. 2015;162:1113–26.
34. Wright AV, Sternberg SH, Taylor DW, Staahl BT, Bardales JA, Kornfeld JE, Doudna JA. Rational design of a split-Cas9 enzyme complex. *Proc Natl Acad Sci U S A*. 2015;112:2984–9.
35. Douar AM, Poulard K, Stockholm D, Danos O. Intracellular trafficking of adeno-associated virus vectors: routing to the late endosomal compartment and proteasome degradation. *J Virol*. 2001;75:1824–33.
36. Wang Y, Zhou L, Liu N, Yao S. BE-PIGS: a base-editing tool with deaminases inlaid into Cas9 PI domain significantly expanded the editing scope. *Signal Transduct Target Ther*. 2019;4:36.
37. Huang TP, Zhao KT, Miller SM, Gaudelli NM, Oakes BL, Fellmann C, Savage DF, Liu DR. Circularly permuted and PAM-modified Cas9 variants broaden the targeting scope of base editors. *Nat Biotechnol*. 2019;37:626–31.
38. Lange S, Katayama Y, Schmid M, Burkacky O, Brauchle C, Lamb DC, Jansen RP. Simultaneous transport of different localized mRNA species revealed by live-cell imaging. *Traffic*. 2008;9:1256–67.
39. Leppke K, Stoecklin G. An optimized streptavidin-binding RNA aptamer for purification of ribonucleoprotein complexes identifies novel ARE-binding proteins. *Nucleic Acids Res*. 2014;42:e13.
40. Zuo E, Sun Y, Wei W, Yuan T, Ying W, Sun H, Yuan L, Steinmetz LM, Li Y, Yang H. Cytosine base editor generates substantial off-target single-nucleotide variants in mouse embryos. *Science*. 2019;364:289–92.
41. Jin S, Zong Y, Gao Q, Zhu Z, Wang Y, Qin P, Liang C, Wang D, Qiu JL, Zhang F, Gao C. Cytosine, but not adenine, base editors induce genome-wide off-target mutations in rice. *Science*. 2019;364:292–5.

42. Zuo E, Sun Y, Yuan T, He B, Zhou C, Ying W, Liu J, Wei W, Zeng R, Li Y, Yang H. A rationally engineered cytosine base editor retains high on-target activity while reducing both DNA and RNA off-target effects. *Nat Methods*. 2020;17:600–4.
43. Grunewald J, Zhou R, Garcia SP, Iyer S, Lareau CA, Aryee MJ, Joung JK. Transcriptome-wide off-target RNA editing induced by CRISPR-guided DNA base editors. *Nature*. 2019;569:433–7.
44. Truong DJ, Kuhner K, Kuhn R, Werfel S, Engelhardt S, Wurst W, Ortiz O. Development of an intein-mediated split-Cas9 system for gene therapy. *Nucleic Acids Res*. 2015;43:6450–8.
45. Nunez JK, Harrington LB, Doudna JA. Chemical and biophysical modulation of Cas9 for tunable genome engineering. *ACS Chem Biol*. 2016;11:681–8.
46. Topilina NI, Mills KV. Recent advances in in vivo applications of intein-mediated protein splicing. *Mob DNA*. 2014;5:5.
47. Lim CKW, Gapinske M, Brooks AK, Woods WS, Powell JE, Zeballos CM, Winter J, Perez-Pinera P, Gaj T. Treatment of a mouse model of ALS by in vivo base editing. *Mol Ther*. 2020;28:1177–89.
48. Levy JM, Yeh WH, Pendse N, Davis JR, Hennessey E, Butcher R, Koblan LW, Comander J, Liu Q, Liu DR. Cytosine and adenine base editing of the brain, liver, retina, heart and skeletal muscle of mice via adeno-associated viruses. *Nat Biomed Eng*. 2020;4:97–110.
49. Wang Y, Zhou L, Tao R, Liu N, Long J, Qin F, Tang W, Yang Y, Chen Q, Yao S: sgBE: a structure-guided design of sgRNA architecture specifies base editing window and enables simultaneous conversion of cytosine and adenosine. Dataset. 2020. NCBI <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA655177>. Accessed 5 Aug 2020.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

