BEAR reveals that increased fidelity variants can successfully reduce the mismatch-tolerance of adenine but not cytosine base editors

András Tálas
Institute of Enzymology, Research Centre for Natural Sciences of the Hungarian Academy of Sciences
Dorottya Simon
Institute of Enzymology, Research Centre for Natural Sciences of the Hungarian Academy of Sciences
Péter Kulcsár
Institute of Enzymology, Research Centre for Natural Sciences of the Hungarian Academy of Sciences
Éva Varga
Institute of Enzymology, Research Centre for Natural Sciences of the Hungarian Academy of Sciences
Ervin Welker (welker.ervin@ttk.hu)
Institute of Enzymology, Research Centre for Natural Sciences of the Hungarian Academy of Sciences

Article

Keywords: Single Base Mutations, Undesired Indel Mutations, Dead Base Editors, Plasmid-based Fluorescent Tool

Posted Date: November 16th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-105846/v1

Version of Record: A version of this preprint was published at Nature Communications on November 3rd, 2021. See the published version at https://doi.org/10.1038/s41467-021-26461-y.

Abstract

Adenine and cytosine base editors (ABE, CBE) are designed to generate single base mutations in genes without necessarily generating DNA double-strand breaks and undesired indel mutations. However, the activity of base editors employing an inactive (dead) SpCas9 is generally low, which may be increased only at the expense of generating undesired indels by using a nickase SpCas9. We have increased the efficiency of dead base editors to a level comparable to that of nickase base editors by enriching cells labelled for efficient base editing using Base Editor Activity Reporter (BEAR), a plasmid-based, fluorescent tool. Furthermore, by exploiting the semi-high-throughput potential of BEAR, we have examined the applicability of increased fidelity variants to decrease Cas9-dependent off-target effects that revealed that CBE remains active on off-targets where increased fidelity mutations and/or mismatches decrease the activity of ABE, making the strategy of applying increased fidelity variants more beneficial for ABE than for CBE.

Introduction

Cas9 nucleases recognize DNA sequences that are located immediately upstream of their respective protospacer adjacent motif (PAM) sequences and are complementary to the ~20 nucleotide-long 5' part of their single guide RNAs (sgRNAs) [1, 2, 3, 4]. These nucleases facilitate effective genome modifications by introducing site specific double-strand breaks; however, unwanted insertions or deletions (indels) also frequently fleck the modified genome [5, 6]. In contrast, base editors originally have been designed to perform genome modifications, without introducing DNA double-strand breaks. They comprise an RNA guided nuclease (Cas9 or Cas12a) fused to a deaminase enzyme and are capable of inducing transition mutations. In theory, these modifications allow the correction of the majority of pathogenic single nucleotide variations found in the human genome [7, 8, 9]. Cytosine base editors (nCBEs or dCBEs, referring to a nickase or an inactive (dead) nuclease partner, respectively) contain a cytidine deaminase that converts cytosines of the non-targeted DNA strand into uracils to be replaced with thymines during DNA replication [7]. Similarly, adenine base editors (nABEs and dABEs) contain an adenosine deaminase that converts adenines of the non-targeted DNA strand into inosines that are read as guanines during DNA replication [8].

Unfortunately, the genome modification efficiency of nuclease inactive base editors, dCBEs and dABEs, lags behind that of SpCas9 nuclease and could be increased only by employing a nickase Cas9 [7]. By nicking the non-modified DNA strand, the DNA repair systems are biased to use the uridine- or inosine-containing strand as a template, which greatly improves the efficiency of base editing [7, 8]. However, repairing the nicks generated by nickases yields various amounts of unwanted indels in a target- and cell type-dependent manner, reported to be as high as 10–20% in some cases [7, 10, 11]. One solution for increasing the efficiency of base editing without generating nick-induced indels would be the enrichment of those cells that contain dABE- and dCBE-edited bases, using a marker. Unfortunately, the markers that have been developed so far to enrich CBE- or ABE-edited cells exclusively employ nickase SpCas9 [12, 13, 14,

[15, 16, 17]. One of the objectives of our study has been to develop a marker that is sensitive enough to enable high efficiency base editing by enriching dABE- and dCBE-edited cells, without intentionally nicking the DNA.

Several base editors have been developed to eliminate some of the limitations of current base editing techniques. The variations include using mutant and deletion variants of deaminases, modifying the length of the linker between the nuclease and the deaminase, using different orthologs and variants of Cas9 or Cas12a, fusing additional copies of uracil glycosylase inhibitors (UGI), as well as changing the architecture of base editors [9, 11, 18, 19, 20, 21, 22, 23]. In fact, these variants have successfully increased the activity of base editors and altered their editing windows or their specificities towards the edited bases, as well as they have partially decreased associated Cas9-dependent or Cas9-independent off-target effects [21, 24, 25, 26]. However, while approaches to monitor genome-wide Cas9-dependent off-target modifications have been developed [27, 28, 29], methodologies to diminish them are less well established. We especially miss a thorough assessment of the applicability of increased fidelity variants to decrease the mismatch tolerance of ABE and CBE. Thus, it has been designated as the second objective of our study.

Although a great number of base editor variants have been developed, it is rather difficult to get an overview of their features and the benefits these variants can offer relative to one another. The absence of simple and effective means to compare the performance of these base editors on various sequences and in any cells of choice hampers the exploration of their full potential. To monitor the activity of base editors usually Sanger or next generation sequencing is applied [7, 8, 21, 30]. Recently, a few approaches have been reported that allow the employment of fluorescence-based assays. These assays are based on the installation or alteration of a start or stop codon [16, 17], or they rescue a disruptive amino acid and concomitantly recover a fluorescent signal [14]. Alternatively, a non-synonymous mutation in the chromophore of a fluorescent protein that induces fluorescence spectral change has also been explored as an option to monitor base editing activity [12, 15]. Although these assays exploit clever strategies, they are limited to only a few target sequences or exhibit high background signal and/or have relatively low sensitivity [12, 13, 14, 15, 16].

Here we report on developing Base Editor Activity Reporter (BEAR) and employ it to better understand whether and how increased fidelity SpCas9 variants can decrease the off-target effects of ABEs and CBEs.

Results

We aimed to develop an easy-to-perform and quick, gain-of-signal fluorescent assay to monitor base editing activity with a plasmid-based format that allows using a number of sequences and can be easily adapted to various types of cells. The assay should report exclusively on the efficiency of base editing without being sensitive to potential indels generated by base editors. BEAR, the assay we designed in response to this demand, is based on a split GFP protein separated with the last intron of the mouse *Vim*

gene. The sequence of the functional 5’ splice site (5’ss) is altered in such a way, that it abrogates splicing and thus GFP fluorescence, but splicing and GFP fluorescence can be restored by applying base editors (**Fig. 1**).

This rationale could not be used by disrupting the canonical ‘GT’ 5’ splice site either in the first position from ‘G’ to ‘A’ to be compatible with ABEs, or in the second position from ‘T’ to ‘C’ for CBEs. Both ‘AT’ and ‘GC’ splice sites are known to be functional as very rare, non-canonical splice sites in the human genome [31], as we have also demonstrated it by transfecting the plasmids with these canonical (‘GT’) and non-canonical (‘AT’ or ‘GC’) 5’ splice sites into both N2a and HEK293T cells, and measuring the number of GFP positive cells afterwards (**Supplementary Fig. 1**).

The flanking sequences of 5’ss modulate the efficiency of the splicing process [32]. This exon-intron junction contains the 5’ NNG and 3’ RAGT flanking consensus sequences (**Fig. 2c, Supplementary Fig. 2a**), which have been reported to best enhance splicing [32]. In order to find appropriate disrupted and edited sequence pairs which fully diminish and support splicing, respectively, we have systematically modified the other, non-targeted nucleotide of 5’ss to ‘AN’ and ‘GN’ for ABEs (**Fig. 2a**) and to ‘NC’ and ‘NT’ for CBEs (**Fig. 2b**) and/or the 5’ or 3’ flanking sequences in both the disrupted and in the pre-edited plasmids (here and throughout the manuscript pre-edited plasmids are the positive controls generated by molecular cloning to represent the maximum fluorescence that can be reached by editing). Constructs were transfected into both HEK293T and N2a cells, and the cells were analyzed by flow cytometry (**Fig. 2a, b; Supplementary Fig. 2b, c**; respectively). Altering only one of the bases of the 5’ss to any of the three other bases while keeping the flanking region intact was found to preserve fluorescence. Altering both bases of 5’ss or one base and either the 3’ or the 5’ flanking consensus sequence generally abrogated the fluorescent signals. When both flanking regions were altered, even the canonical ‘GT’ 5’ss sequence could be insufficient to efficiently splice and to recover GFP fluorescence. These experiments have revealed a few candidate combinations for which no detectable fluorescent signal is apparent with the disrupted splice site sequence, but it is present in case of the corresponding pre-edited sequences (**Fig. 2**).

Next, we have tested whether base editors can indeed recover fluorescence exploring some of the best candidate constructs identified in **Figure 2** and in **Supplementary Figure 2**. Throughout the study, the adenine and cytosine base editors used are the codon optimized ABERA (shortened as ABE) and FNLS-CBE (shortened as CBE) variants, respectively, described by Zafra *et al.* [23], unless indicated otherwise. The five selected plasmids (p1, p9, p14, p15 and p24 in **Fig. 2a, b**) were co-transfected with ABE or CBE into both HEK293T and N2a cells, and the number of GFP positive cells were measured. In case of all selected constructs ABE and CBE could successfully recover fluorescence from 31% to 91% in HEK293T (**Fig. 2c**) and from 45% to 75% in N2a cells (**Fig. 2d**). Interestingly, both ABE and CBE can correct the disrupted 5’ss in P9 and restore GFP fluorescence, converting ‘AC’ to either ’GC’ or ‘AT’, respectively. Since both ABE and CBE reach the same levels on P9 as detected on the other best constructs (**Fig. 2c, d**), we could further examine both of these base editors on this common disrupted plasmid named BEAR-GFP (**Fig. 1**).

Figure 2e shows that fluorescence is not recovered when this construct is targeted by a single nickase or a nuclease SpCas9, supporting that the method exclusively informs about base editing. We have also found that the nuclease inactive (dead) base editor variants dABE and dCBE are also capable of correcting the 5'ss, however, with lower efficiency, as indicated by the recovered fluorescence signals of 36% and 18% for dABE and dCBE, respectively (**Fig. 2e**).

As an advantage, our method is not restricted to a few target sequences only. The intronic sequence between the PAM and the 3' flanking consensus site can be varied without restrictions. This also allows to move the PAM sequence, and thus, the editing window, with respect to the base position to be edited (**Supplementary Fig. 3a**). Furthermore, the exonic part of the target sequence can also be altered by applying different fluorescent proteins with BEAR (**Supplementary Fig. 3b**) or by moving the intron's position in the coding sequence for the protein (**Supplementary Fig. 3c** and **d**). Thus, even when the seven nucleotide-long consensus flanking sequence part of the target sequence is preserved unaltered, some tens of millions of possible different target sequences can be examined using BEAR. Since either the non-edited nucleotides of 5'ss or one of the flanking consensus sequences may also be varied (**Fig. 2a, b**), our method allows the targeted base to be examined in almost any sequence contexts.

To see whether the efficiency of base editing of target sequences in a plasmid or in a genomic context is governed by the same factors, we have generated stable HEK293T cell lines harboring either a disrupted GFP or a disrupted mScarlet protein, containing exactly the same exons, introns and target sequences as the BEAR plasmids have. When these cell lines were targeted by ABE and the corresponding sgRNA, fluorescence was efficiently recovered (**Fig. 3a**). We have compared the BEAR-GFP plasmid with the BEAR-GFP cell line, regarding their effects on the extent of fluorescence recovery, using 32 sgRNAs containing no, one or two consecutive mismatching nucleotides at different positions (**Fig. 3b**). The assays on the cell line and on the plasmid yielded highly similar outcomes (r=0.89), indicating that the plasmid-based assay properly mirrors the activities of ABEs on sequences in a genomic context.

To examine whether fluorescence recovery definitely results from successful base editing, we have employed one matched and one, two or three base-mismatched sgRNAs with ABE (**Fig. 3c**) or CBE (**Fig. 3d**) on the BEAR-GFP cell line, and monitored base editing activity by measuring the number of GFP positive cells, as well as by quantifying editing using EditR [33]. The measured fluorescence intensity has been found to be proportional to the level of actual base editing (r=0.98). Sequencing has also revealed that in case of ABE not only the 5'ss sequence, but also a bystander adenine has been edited to a certain extent (**Fig. 3e**). Constructing and testing the corresponding disrupted and pre-edited plasmids has proved that editing the second, bystander 'A' with or without the adenine of the 5'ss sequence does not decrease or increase GFP fluorescence, respectively (**Fig. 3f**). In case of CBE, no bystander nucleotides have been edited, but the targeted cytosine has been converted to guanine, although to a smaller extent (**Fig. 3g**), as it has also been reported in case of several target sequences [22, 30, 34]. By constructing the corresponding pre-edited plasmids, we have verified that the increase in fluorescence is derived from the intended editing of 'AC' to 'AT' only, without a contribution from 'AC' to 'AG' editing of 5'ss (**Fig. 3h**). Taken together, these data support that the BEAR method gives a faithful account of the activities of a base editor.

Increased base-editing without nicking the target DNA

Since BEAR is sensitive enough to detect the activity of nuclease inactive base editors (**Fig. 2e**), we have tested whether it could be used as a marker for those cells in which efficient base editing occurs, in order to increase the efficiency of base editing without intentionally nicking the DNA. We have co-transfected the BEAR-GFP plasmid with dABE and the corresponding sgRNAs into the BEAR-mScarlet cell line, and we have found that dABE has restored mScarlet fluorescence in 20% of the cells. Thirty-one percent of the cells in the transfected population exhibited mScarlet fluorescence, and 51% of the cells showed fluorescence for both mScarlet and GFP, indicating that the cells being active in processing the A-to-G base conversion on the plasmid are also efficient on the genomic DNA (**Fig. 4a**). We have also co-transfected the BEAR-mScarlet plasmid with dABE into the BEAR-GFP cell line. In this experiment BEAR-enrichment has increased the percentage of edited cells from 22% to 45%, highly exceeding the enrichment that we measured in the transfected population (30%; **Fig. 4a**).

Employing dABE and dCBE on endogenous genomic targets, namely on FANCF site 2 (**Fig. 4b, c**), VEGFA site 3 (**Fig. 4d**), and HEK site 4 (**Fig. 4e**), we have further tested the potential of BEAR to increase the efficiency of base editing without intentionally nicking the DNA. Using BEAR as a marker for increasing the efficiency of base editing (achieved by cell sorting of BEAR-positive cells) we have revealed an up to 12-fold enrichment for dCBE (FANCF site 2) and an up to 30-fold enrichment for dABE editing (VEGFA site 3) (**Fig. 4c, d**). In these experiments base editing activities have reached a maximum of 43% efficiency with dABE on VEGFA site 3 and 41% efficiency with dCBE on HEK site 4 (**Fig. 4d, e**). For comparison, nABE and nCBE (containing nickase Cas9) editing was monitored (without enrichment) on the same target sites (**Fig. 4b-e**). These experiments have indicated that BEAR facilitates base editing on genomic targets by dABE and dCBE without intentionally nicking the DNA, with efficiencies equal to or greater than that of nABE and nCBE.

On-target activity of base editors with increased fidelity SpCas9 variants

Several studies reported on CBE's showing higher or similar mismatch tolerance compared to ABE that results in various Cas9-dependent off-target effects [10, 29, 35, 36]. Applying increased fidelity variants may seem to be a plausible approach to decrease the Cas9-dependent off-target effects of base editors, however, only a few attempts of combining an increased fidelity variant with a base editor are reported in the literature [21, 29, 37, 38, 39, 40]. To get a more comprehensive understanding of these effects, exploiting BEAR we have compared the activity and mismatch-tolerance of CBE and ABE containing six increased fidelity SpCas9 variants: eSpCas9, SpCas9-HF1, HypaSpCas9, Hypa-R661ASpCas9 (i.e. HypaSpCas9 which also contains the R661A mutation) evoSpCas9 and HeFSpCas9 [41, 42, 43, 44, 45]. Regarding that the 'AC' 5'ss sequence can be edited by both ABEs and CBEs in the BEAR-GFP plasmid (**Fig. 1**), they can be compared on the same targets, by using the same sgRNAs. Accordingly, we have compared their on-

target base editing activities on 34 targets in which the 5'ss and flanking regions, as well as their distance from the PAM sequence was kept fixed and only the PAM proximal 10 nucleotides were varied. Thus, for both base editors, the sequences in their editing windows and the bases surrounding the edited bases were kept unaltered. Neighboring (+/-1) nucleotides can strongly influence the efficiency of base editing; 'G*A*C' and 'A*C*A' employed here for ABE and CBE, respectively, have been shown to be associated with medium level activities for both base editors [46]. Lacking data suggesting the opposite, we have expected that the differences in the 34 target sequences (in the PAM proximal 10 nucleotides) should primarily affect the interactions between the fused SpCas9 nuclease partner of the base editors and the targets, thus this experimental design was specifically suited to study how the binding and cleavage propensities of SpCas9 variants affect the base editors' activities.

The results illustrated in **Figure 5a** indicate that nABE is highly active on all 34 targets with 73% mean activity (its efficiency ranges between 62% and 89%). dABE was found to be less active with 24% mean activity. In theory, the activity profile of dABE is influenced by the sequence specificity of both the TadA deaminase and the binding of SpCas9. In contrast, the activity profile of nABE is also influenced by the nicking activity of SpCas9, which aims to bias the repair system in order to correct the mismatching bases of the unedited strand, and thus, to increase editing efficiency [7]. The activity profile of dABE and nABE shown in **Figure 5a** indicates a weak correlation (r=0.29; **Supplementary Fig. 4a**), suggesting that the nicking activity of SpCas9 in nABE substantially alters the relative efficiency of nABE compared to dABE on these target sequences.

Former studies of increased fidelity SpCas9 nuclease variants have shown that these nucleases have a trade-off between efficiency and fidelity, and can be ranked according to their average activities, with evo- and HeFSpCas9 showing much lower average activities than the rest of the increased fidelity variants [45, 47, 48]. Increased fidelity variants of ABE have been detected to exhibit gradually decreasing activity from ABE to HeF-ABE in our experiments, the latter showing minimal activity (4% in average), equal to the double of the background activity of nickase Cas9 (**Fig. 5a**), which parallels with the effect seen in case of increased fidelity nucleases. Furthermore, the activity profile of three nuclease variants, SpCas9-HF1, HypaSpCas9 and evoSpCas9 is reported to show low correlation with that of the WT-SpCas9 nuclease, while the activity profile of evoSpCas9 shows higher correlation with SpCas9-HF1 and Hypa-SpCas9 than with eSpCas9 or the WT nucleases [47]. The increased fidelity ABE variants demonstrate a similar pattern, as shown in **Figure 5a**. The activity profiles of HF1-, Hypa- and evo-ABE show weaker correlations with nABE (r=0.46–0.54), while evo-ABE shows higher correlations with HF1- and Hypa-ABE (r=0.86 and 0.93, respectively) than with e-ABE or ABE (r=0.66 and 0.54, respectively). These findings support that the activity profiles of increased fidelity ABE variants are primarily determined by the sequence specificities of the partner SpCas9 variants (**Supplementary Fig. 4a**). **Supplementary Figure 4c** shows that of the two codon-optimized adenine base editors, ABEmax [22] has higher activity than nABE (ABERA [23]), their average activities being 83% and 73%, respectively).

nCBE is less active on these 34 targets (its average editing activity is 50%), and shows higher sensitivity for sequence variations: its efficiency ranges between 26% and 69% (**Fig. 5b**). dCBE is considerably less active (its average activity is 12%) and its efficiency varies from 5% to 21% (**Fig. 5b**). Their activity profiles correlate (r=0.51, **Supplementary Fig. 4b**) more than the activity profiles of the ABEs, suggesting that the nicking activity has a weaker relative influence on CBE's sequence dependence than on that of ABEs.

A decreasing effect of increased fidelity mutations from e- to evo- and HeF-CBE variants on the average activities of CBE is also evident, although this decrease is much less prominent than it is in case of the ABE variants: their average activity decreases from 50% to 22% compared to the 73% to 4% decrease seen with ABEs (**Fig. 5b**). The activity profile of CBE strongly correlates with those of the increased fidelity HF1-, Hypa-, and evo-CBE variants (r=0.69–0.86, **Supplementary Fig. 4b**), and the correlations of evo-CBE with HF-, Hypa-, e-ABE and ABE (r=0.73–0.84) are more similar to one another, which is also in contrast with the activity profiles characterizing the increased fidelity ABE (**Supplementary Fig. 4a**) and nuclease [47] variants. These data suggest that in strong contrast to ABEs, the activity profiles of the CBE variants are more determined by factors other than the properties of the increased fidelity SpCas9 nucleases, presumably by the sequence specificity of the deaminase partner and the subsequent repair process. The differing activities of the CBE variants across the 34 targets also suggest that in case of CBE, rather than in case of ABE, these latter factors are affected more by the sequence features of the PAM proximal 10 nucleotides.

We have also examined the three CBEs whose sequences have been codon optimized by two independent research groups [22, 23], and found that CBE (FNLS-CBE) exhibits higher average activities than BE4max or AncBE4max (50% vs. 40% and 39% respectively) (**Supplementary Fig. 4d**).

Mismatch tolerance of ABE, CBE and their increased fidelity variants

We have compared the mismatch tolerance of ABE (**Fig. 6a**) and CBE (**Fig. 6b**) with their increased fidelity variants employing 50 mismatching sgRNAs (Target 1 from **Fig. 5**) in which the positions of one to five consecutive mismatches have systematically varied along the full length of each sgRNA. Examining ABE, we have found that it tolerates sgRNAs containing one or two mismatches in all the positions examined, with an average of 71% and 37% normalized activity, respectively (**Supplementary Fig. 5a**). Regarding dABE, it exhibits slightly higher fidelity, which is more apparent with the sgRNAs containing two mismatching positions (normalized average activity: 15%). The mismatching profiles of nABE and dABE show a strong correlation (r=0.88, **Supplementary Fig. 5b**), which is interesting since the off-target effects of the active and inactive forms of SpCas9 have been reported to differ considerably [49]. Based on this consideration, we expected a weaker correlation, similarly to the correlation between the on-target activities of nABE and dABE.

Regarding the increased fidelity ABE variants, five of them have been tested on the BEAR-GFP plasmid (Target 1 from **Fig. 5**), employing the same 50 mismatching sgRNAs. HeF-ABE was excluded from these experiments due to its low on-target activity. Increased fidelity mutations were found to decrease the mismatch tolerance of ABE (**Fig. 6a**). The fidelity of the same SpCas9 nuclease variants have been reported to increase in a great extent from eSpCas9 to evo- and HeFSpCas9 [45, 47, 48]. Remarkably, these fidelity increases are also evident in the mismatch tolerance of the ABE variants when sgRNAs mismatching in one position are employed (**Fig. 6a, Supplementary Fig. 5a**). In contrast, with almost all sgRNAs containing two or more mismatches, each increased fidelity ABE variant has been found to exhibit only background-level activities. Interestingly, increased fidelity ABE variants exhibit higher specificities on this target than dABE (**Fig. 6a**).

Regarding the mismatch tolerance of the CBE variants, tested using the same 50 mismatching sgRNAs (**Fig. 6b**), we have found that nCBE tolerates one or two mismatches in all the positions examined, with an average normalized activity of 100% and 61% when the sgRNAs include mismatches in one or two positions, respectively (**Supplementary Fig. 5c**). In turn, dCBE exhibits slightly higher fidelity, which is more apparent with the sgRNAs containing two mismatching positions (normalized average activity: 44%). The mismatching profiles of nCBE and dCBE show a strong correlation (r=0.87, **Supplementary Fig. 5d**) which is similar to that seen with nABE and dABE (**Supplementary Fig. 5b**).

Regarding the increased fidelity CBE variants, all the six of them have reached sufficiently high on-target activity on the BEAR-GFP plasmid, thus all six have been investigated with the previous set of 50 mismatching sgRNAs. Interestingly, although an overall increase in specificity towards the highest fidelity evo- and HeF-CBE has been evident (**Fig. 6b**), this effect is much less prominent than it is in case of the increased fidelity ABE variants (**Fig. 6a**). Compared to ABEs, increased fidelity CBE variants exhibit lower specificity. Specifically, while the ABE variants show target specificity resembling the background, characterized by 4–6% and 2–5% of normalized average activity with sgRNAs mismatching in two or three positions, respectively (**Supplementary Fig. 5a**), the CBE variants exhibit 16–27% and 6–12% of normalized average activity with the respective mismatching sgRNAs (**Supplementary Fig. 5c**)

Next, we have tested the mismatch tolerance of ABEmax and xABE (which contains the nickase version of xSpCas9 [50]) along with the previously used increased fidelity ABE variants on target 7 with the same set of 50 mismatching sgRNAs (**Supplementary Fig. 6a**). Compared to ABE, ABEmax has been found to show lower specificity, but their mismatch profiles are nearly identical (r=0.96, **Supplementary Fig. 6b**). Regarding xABE, which is effective on targets with loosened NG-like PAMs and is also reported to possess increased fidelity, it has been found to exhibit slightly higher specificities than nABE, but it is characterized by the lowest specificity among all the variants examined on this target, including dABE. Its mismatch profile seems to be different from that of the increased fidelity ABE variants (r=0.14-0.38, **Supplementary Fig. 6b**), while all four increased fidelity ABE variants show strong correlations with each other in their mismatch tolerance profile (r=0.84–0.93, **Supplementary Fig. 6b**). A similar difference between the activities of xSpCas9 and the other increased fidelity nuclease variants has recently been reported in two studies [47, 48].

To see whether these observations are specific to the target examined or are more general characteristics of these base editor variants, we have investigated the mismatch tolerance of ABE and CBE variants on another three targets (targets 2, 6 and 17; **Supplementary Fig. 7a-c**) using the same approach. These experiments have confirmed the conclusions drawn from our previous findings shown in **Figure 6** and **Supplementary Figure 6**. Namely, (i) CBE is more tolerant to mismatches than ABE is, although ABE also shows a considerable target-dependent mismatch tolerance. (ii) Their activity profiles investigated using the same set of mismatching sgRNAs show strong correlations (r=0.93–0.96), arguing that their mismatch tolerances are primarily influenced by the specificities of SpCas9 cleavage activity as seen also with target 1 in **Figure 6** earlier. (iii) The effect of variants from higher positions of fidelity ranking of increased fidelity SpCas9s is more prominent in case of the increased fidelity ABE variants than in case of the CBE variants, indicating that increased fidelity mutations decrease the mismatch tolerance of ABE more effectively than that of CBE.

Discussion

Base editors have been developed to edit the genome without intentionally causing DNA double-strand breaks. However, employing nuclease inactive base editor variants is associated with low editing efficiency which strongly limits their use [7]. Thus, instead of these nuclease inactive base editor variants, base editors are generally applied with concomitant nicking of the targeted strand which results in considerably more efficient genome modifications [7, 8, 11]. As an alternative, we suggest employing a plasmid-based, fluorescent tool named BEAR (Base Editor Activity Reporter), which makes it possible to achieve the high editing efficiencies attainable by nickase base editors, but without intentionally nicking the DNA, and thus, without generating considerable amounts of indels when using either dABE or dCBE. As such, BEAR provides a unique solution for base editor applications when DNA nicking and generation of indels are not well tolerated (**Fig. 4**).

The experimental data provided here demonstrate the versatility of BEAR for comparing base editors with various features to select the best one for a particular task, as well as for facilitating the development of novel base editor variants with improved properties. Its versatility in terms of accepting target sequences allows the comparison of the efficiencies of base editors in different positions of the editing window, as well as the comparison of how neighboring nucleotides of the edited bases affect their editing efficiency. This feature of BEAR is demonstrated in **Figures 2** and **3**. In contrast, to assess how the nuclease activity of SpCas9 affects base editing, we kept the PAM distal region (the region surrounding the targeted nucleotide) constant, as these sequences are considered to have the most effect on deaminase function [46, 51].

An interesting finding from our experiments is the low activity of the increased fidelity variant HeF-ABE, but not HeF-CBE. Fidelity-increasing mutations have been shown to be associated with somewhat reduced nuclease activity: these nuclease variants are reported to pass through the quality checkpoints less efficiently [43, 52] than the wild type SpCas9 [53, 54, 55] during target cleavage, while their binding to the

target DNA is largely unaffected [43, 45]. This may be manifested in lower cleavage activity on off-targets when there are mismatches between the spacer and the target DNA sequences, and sometimes on on-target sequences as well [39, 41, 42, 43, 44, 45, 56]. With such non-cleavable targets, the variants separate the DNA strands in the PAM-proximal region upon binding to the target, however, the mutations inhibit the effective full-length separation of the two strands of the target DNA, and thus they also prevent the formation of the cleavage-competent conformation of SpCas9 [52, 55, 57]. HeF-SpCas9, which has the highest fidelity and lowest average cleavage activity among these variants, binds to the targets of WT-SpCas9 without being able to cleave most of them: in contrast to the wild type nuclease [53, 54, 55], it fails to effectively separate the strands of the target DNA at full length and to acquire the cleavage-competent conformation [43, 45, 52]. Based on these considerations, we expected that HeF-ABE, which lacks a nicking activity, would show a rather dABE-like activity on many targets and a dABE-like activity profile. Interestingly, our findings have not supported this expectation. The average activities of both evo- (the other highest fidelity variant) and HeF-ABE (15% and 4%, respectively, **Fig. 5a**) are less than that of dABE (24%), and their activity profiles hardly correlate with that of dABE (r=0.16 and 0.29, respectively, **Supplementary Fig. 4a**). In fact, dSpCas9 in dABE separates the strands of the target DNA at full length [57], facilitating the action of the deaminase, which is active only on single-strand DNA [8]. However, when increased fidelity nucleases are simply bound to the DNA without being able to cleave or nick the target, they separate the two strands of the target DNA at its seed region only, while the PAM distal region is separated less efficiently, rather transiently [52, 55, 57]. We speculate that in such cases, the deaminase is able to proceed in dABE, but not in HeF-ABE. Interestingly, the average activities of evo- and HeF-CBE do not decrease below that of dCBE (22% and 27% vs. 12%, respectively, **Fig. 5b**), suggesting that the effective, full-length separation of the two strands of target DNA is less critical for CBE than for ABE. Compared to ABE, CBE may have a greater activity on targets where the separation of the PAM-distal region is rather transient in nature, likely due to the higher deaminase activity of APOBEC in CBE than that of the evolved TadA in ABE [58]. In line with these interpretations, Liu and colleagues have developed a new version of ABE, ABE8e which exhibits an 1,000-fold higher deaminase activity and does not require the effective full-length separation of the two target DNA strands: it is active even on a transiently separated DNA strand, upon SpCas9 binding to a non-target PAM motif [59, 60].

This observation also provides an explanation for the higher off-target propensity of CBE. Mismatches between the spacer and the target sequences have been described to reduce the cleavage activities of SpCas9 via a mechanism that also involves less efficient full-length separation of the target DNA strands [52, 55, 57, 60]. Thus, ABEs are more sensitive to mismatches interfering with SpCas9 activity to an extent that inhibits the full-length separation of the strands of the target DNA. By contrast, CBE, which exhibits higher deaminase activity [58], can also act on some of the off-target sequences that are rather just transiently separated in the PAM distal region of the target DNA.

These interpretations are also in line with the following observations: (i) similar correlations are evident for the sequence specificity profiles of the activities of increased fidelity nuclease variants, as well as for that of increased fidelity ABE variants (e.g. for the Hypa- and HF-ABE variants). In fact, these similarities

are expected if we consider that ABE variants preferentially work on those sequences on which the corresponding nuclease variants show full-length strand separation, and thus, show cleavage activities. This apparent matching with the cleavage activity profiles of the increased fidelity SpCas9s is diminished when we use CBE variants, which are less sensitive to the effective separation of the target DNA strands. (ii) Increased fidelity variants decrease the off-target activity of ABE more efficiently than that of CBE. (**Fig. 6**, **Supplementary Fig. 6, Supplementary Fig. 7**). (iii) The sequence dependence of dABE's and nABE's activities on mismatching targets strongly correlate, despite the differing off-target binding and cleavage profiles of SpCas9. So, it works on a similar subset of the off-targets for dSpCas9 and SpCas9. This is rational if we consider that ABE preferentially works on sequences where the target DNA strands are fully separated, and it does not work properly on off-targets where the mismatches inhibit effective full-length separation, preventing cleavage, but not binding.

This understanding of the actions of ABE and CBE variants suggests that the Cas9-dependent off-target propensity of ABE editing can be effectively diminished by the application of increased fidelity variants. It also suggests a strategy for CBE editing: by fine tuning the deaminase activity of APOBEC to match that of the evolved TadA would make CBE combinations with increased fidelity SpCas9 variants more fruitful.

Materials And Methods

Materials

Restriction enzymes, T4 ligase, Dulbecco's modified Eagle medium (DMEM), fetal bovine serum, Turbofect and penicillin/streptomycin were purchased from Thermo Fischer Scientific Inc. DNA oligonucleotides and the GenElute HP Plasmid Miniprep and Midiprep kit used for plasmid purifications were acquired from Sigma-Aldrich. Q5 High-Fidelity DNA Polymerase, NEB5-alpha competent cells, HiFi Assembly Master Mix were purchased from New England Biolabs Inc.

Plasmid construction

Vectors were constructed using standard molecular biology techniques. All base editor coding sequences were cloned to the same plasmid backbone. For detailed cloning, primer and sequence information see Supplementary Materials. The sequences of all plasmid constructs were confirmed by Sanger sequencing (Microsynth AG).

Plasmids acquired from the non-profit plasmid distribution service Addgene were the following: pU6-pegRNA-GG-acceptor (#132777 [61]), pLenti-ABERA-P2A-Puro (#112675 [23]), pLenti-FNLS-P2A-Puro (#110841 [23]), pX330-Flag-wtSpCas9 (#92353 [45]), pX330-Flag-dSpCas9 (#92113 [45]), pX330-Flag-wtSpCas9-D10A (#80448), pcDNA3.1-mCherry (#128744), pCytERM_mScarlet_N1 (#85066 [62]), pSc1-puro (#80438 [63]), pmCherry-gRNA (#80457 [45]), pCMV_BE4max (#112093 [22]), pCMV_AncBE4max (#112100 [22]), pCMV_ABEmax (#112101 [22]), pX330-Flag-eSpCas9 (#126754), pX330-Flag-SpCas9-HF1 (#126755),

pX330-Flag-HypaSpCas9 (#126756), pX330-Flag-Hypa-A-SpCas9 (#126757), pX330-Flag-evoSpCas9 (#126758), pX330-Flag-HeFSpCas9 (#126759) [56].

The following plasmids developed in this study are available from Addgene: pAT9624-BEAR-cloning (#162986), pAT9658-sgRNA-mCherry (#162987), pAT9679-sgRNA-BFP (#162988), pAT9651-BEAR-GFP (#162989), pAT9750-BEAR-mCherry (#162990), pAT9752-BEAR-mScarlet (#162991), pAT9650-BEAR-GFP-preedited (#162992), pAT9751-BEAR-mCherry-preedited (#162993), pAT9753-BEAR-mScarlet-preedited (#162994), pAT15415-BEAR-GFP-target-mCherry (#162995), pAT15416-BEAR-mScarlet-target-BFP (#162996), pAT9676-ABE (#162997), pAT9749-dABE (#162998), pAT9991-eABE (#162999), pAT9992-HF-ABE (#163000), pAT9993-Hypa-ABE (#163001), pAT9994-HypaR661A-ABE (#163002), pAT9995-evoABE (#163003), pAT9996-HeF-ABE (#163004), pAT9784-xABE (#163005), pAT15002-ABEmax (#163006), pAT9675-CBE (#163007), pAT9748-dCBE (#163008), pAT15064-eCBE (#163009), pAT15065-HF-CBE (#163010), pAT15066-Hypa-CBE (#163011), pAT15067-HypaA-CBE (#163012), pAT15068-evoCBE (#163013), pAT15069-HeF-CBE (#163014), pAT15000-BE4max (#163015), pAT15001-AncBE4max (#163016).

Cell culturing and transfection

N2a (neuro-2a mouse neuroblastoma cells, ATCC, CCL-131) and HEK293T (ATCC, CRL-1573) cells were grown at 37 °C in a humidified atmosphere of 5% CO_2 in DMEM medium supplemented with 10% heat-inactivated fetal bovine serum with 100 units/mL penicillin and 100 µg/mL streptomycin.

N2a and HEK293T cells cultured on 48-well plates were seeded a day before transfection at a density of 3×10^4 and 5×10^4 cells/well, respectively. 250 ng total DNA: 66 ng of BEAR target plasmid, 56 ng of sgRNA-mCherry (or sgRNA-BFP in case of BEAR-mScarlet) and 127 ng of CBE or 124 ng of ABE coding plasmid was mixed with 1 µL Turbofect reagent in 50 µL serum-free DMEM and was incubated for 30 minutes prior to being added to the cells. Three parallel transfections were made from each sample. Cells were analysed by flow cytometry three days after transfection.

Preparation of BEAR stable cell lines

BEAR stable cell lines were prepared by modifying the self-cleaving plasmid method described earlier [63]. Disrupted BEAR-mScarlet and BEAR-GFP constructs were cloned into a plasmid which bears a puromycin expression cassette and a Cas9 target that was used for self-cleaving plasmids (pSc-BEAR-mScarlet, pSc-BEAR-GFP). A corresponding spacer (which targets the pSc-BEAR plasmids) and AAVS1 genomic targets (described in ref [64]) were cloned into an sgRNA-mCherry plasmid. When these three plus a Cas9 coding plasmids are co-transfected into a cell, the sgRNA linearizes the pSc-BEAR plasmid, and integrates into the targeted locus (AAVS1) via non-homologous end joining.

HEK293T cells cultured on 6-well plates were seeded a day before transfection at a density of 5×10^5 cells/well. 1100 ng of pSc-BEAR, 800 ng of pSc-gRNA-mCherry, 800 ng of AAVS1-gRNA-mCherry and 1300 ng of SpCas9-HF1-plus [56] was mixed with 6 µL Turbofect reagent in 400 µL serum-free DMEM and was incubated for 30 minutes prior to being added to the cells. Two days after transfection the cells were treated with 1 µg/µL puromycin for 15 days, then single cells were cloned in 96-well plates. Clones were checked negative for mCherry fluorescence via flow cytometry and for Cas9 via PCR analysis.

Stable cell lines were transfected with 250 ng total DNA: 76 ng sgRNA-mCherry (or sgRNA-BFP) and 174 ng of ABE coding plasmid was mixed with 1 µL Turbofect reagent in 50 µL serum-free DMEM and was incubated for 30 minutes prior to being added to the cells.

Flow cytometry and cell sorting

Flow cytometry analysis was carried out using an Attune NxT Acoustic Focusing Cytometer (Applied Biosystems by Life Technologies). In all experiments, a minimum of 10,000 viable single cells were acquired by gating based on side and forward light-scatter parameters. GFP, mCherry and mScarlet signals were detected using the 488, 561 and 561 nm diode laser for excitation and the 530/30, 620/15 and 585/16 nm filter for emission, respectively. For data analysis Attune Cytometric Software v.2.1.0 was used.

In enrichment experiments where fluorescence activated cell sorting was used, HEK293T cells cultured on T-25 flasks were seeded a day before transfection at a density of $1{,}3 \times 10^6$ cells/flask. 1754 (or 175) ng of BEAR-GFP, 1894 ng of BEAR-gRNA-mCherry, 1894 ng of genome targeting sgRNA and 2921 ng of dABE or 2859 dCBE coding plasmid was mixed with 16 µL Turbofect reagent in 800 µL serum-free DMEM and was incubated for 30 minutes prior to being added to the cells. One, two or three days after transfection the cells were trypsinized and sorted directly into genomic lysis buffer, then genomic DNA was extracted. Cell sorting was carried out on a FACSAria III cell sorter (BD Biosciences). The live single cell fraction was acquired by gating based on side and forward light-scatter parameters. GFP or mCherry signals were detected using the 488 or 561 nm diode laser for excitation and the 530/30 or 610/20 nm filter for emission, respectively. A minimum of 50,000 cells were sorted in all experiments.

Genomic DNA purification and EditR analysis

Genomic DNA from FACS or other experiments was extracted according to the Puregene DNA Purification protocol (Gentra Systems Inc.). The purified genomic DNA was executed to PCR analysis, conducted with Q5 polymerase and locus specific primers (see Supplementary Materials). PCR products were gel purified via NucleoSpin Gel and PCR Clean-up kit (Macherey-Nagel) and were Sanger sequenced. Editing efficiencies were analysed by EditR 1.0.9 (https://moriaritylab.shinyapps.io/editr_v10/, [33]).

Data availability

Expression vectors developed in this study are available from Addgene: pAT9624-BEAR-cloning (#162986), pAT9658-sgRNA-mCherry (#162987), pAT9679-sgRNA-BFP (#162988), pAT9651-BEAR-GFP (#162989), pAT9750-BEAR-mCherry (#162990), pAT9752-BEAR-mScarlet (#162991), pAT9650-BEAR-GFP-preedited (#162992), pAT9751-BEAR-mCherry-preedited (#162993), pAT9753-BEAR-mScarlet-preedited (#162994), pAT15415-BEAR-GFP-target-mCherry (#162995), pAT15416-BEAR-mScarlet-target-BFP (#162996), pAT9676-ABE (#162997), pAT9749-dABE (#162998), pAT9991-eABE (#162999), pAT9992-HF-ABE (#163000), pAT9993-Hypa-ABE (#163001), pAT9994-HypaR661A-ABE (#163002), pAT9995-evoABE (#163003), pAT9996-HeF-ABE (#163004), pAT9784-xABE (#163005), pAT15002-ABEmax (#163006), pAT9675-CBE (#163007), pAT9748-dCBE (#163008), pAT15064-eCBE (#163009), pAT15065-HF-CBE (#163010), pAT15066-Hypa-CBE (#163011), pAT15067-HypaA-CBE (#163012), pAT15068-evoCBE (#163013), pAT15069-HeF-CBE (#163014), pAT15000-BE4max (#163015), pAT15001-AncBE4max (#163016).

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* **337**, 816-821 (2012).

2. Mali P*, et al.* RNA-guided human genome engineering via Cas9. *Science* **339**, 823-826 (2013).

3. Cong L*, et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819-823 (2013).

4. Gasiunas G, Barrangou R, Horvath P, Siksnys V. Cas9–crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proceedings of the National Academy of Sciences* **109**, E2579-E2586 (2012).

5. Cullot G*, et al.* CRISPR-Cas9 genome editing induces megabase-scale chromosomal truncations. *Nature communications* **10**, 1-14 (2019).

6. Tsai SQ*, et al.* GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nature biotechnology* **33**, 187-197 (2015).

7. 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* **533**, 420-424 (2016).

8. Gaudelli NM*, et al.* Programmable base editing of A• T to G• C in genomic DNA without DNA cleavage. *Nature* **551**, 464-471 (2017).

9. Li X*, et al.* Base editing with a Cpf1–cytidine deaminase fusion. *Nature biotechnology* **36**, 324-327 (2018).

10. Lee HK*, et al.* Targeting fidelity of adenine and cytosine base editors in mouse embryos. *Nature communications* **9**, 1-6 (2018).

11. Komor AC*, 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. *Science advances* **3**, eaao4774 (2017).

12. Coelho MA*, et al.* BE-FLARE: a fluorescent reporter of base editing activity reveals editing characteristics of APOBEC3A and APOBEC3B. *BMC biology* **16**, 150 (2018).

13. St. Martin A*, et al.* A fluorescent reporter for quantification and enrichment of DNA editing by APOBEC–Cas9 or cleavage by Cas9 in living cells. *Nucleic acids research* **46**, e84-e84 (2018).

14. Martin AS, Salamango D, Serebrenik AA, Shaban NM, Brown WL, Harris R. A panel of eGFP reporters for single base editing by APOBEC-Cas9 editosome complexes. *Scientific reports* **9**, 1-8 (2019).

15. Standage-Beier K*, et al.* A transient reporter for editing enrichment (TREE) in human cells. *Nucleic acids research* **47**, e120-e120 (2019).

16. Katti A*, et al.* GO: a functional reporter system to identify and enrich base editing activity. *Nucleic acids research* **48**, 2841-2852 (2020).

17. Wang P, Xu L, Gao Y, Han R. BEON: A functional fluorescence reporter for quantification and enrichment of adenine base editing activity. *Molecular Therapy*, (2020).

18. 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. *Nature biotechnology* **35**, 371-376 (2017).

19. Kleinstiver BP*, et al.* Engineered CRISPR–Cas12a variants with increased activities and improved targeting ranges for gene, epigenetic and base editing. *Nature biotechnology* **37**, 276-282 (2019).

20. Tóth E*, et al.* Improved LbCas12a variants with altered PAM specificities further broaden the genome targeting range of Cas12a nucleases. *Nucleic acids research* **48**, 3722-3733 (2020).

21. Gehrke JM*, et al.* An APOBEC3A-Cas9 base editor with minimized bystander and off-target activities. *Nature biotechnology* **36**, 977 (2018).

22. Koblan LW*, et al.* Improving cytidine and adenine base editors by expression optimization and ancestral reconstruction. *Nature biotechnology* **36**, 843-846 (2018).

23. Zafra MP*, et al.* Optimized base editors enable efficient editing in cells, organoids and mice. *Nature biotechnology* **36**, 888-893 (2018).

24. Grünewald J*, et al.* Transcriptome-wide off-target RNA editing induced by CRISPR-guided DNA base editors. *Nature* **569**, 433-437 (2019).

25. Huang TP*, et al.* Circularly permuted and PAM-modified Cas9 variants broaden the targeting scope of base editors. *Nature biotechnology* **37**, 626-631 (2019).

26. Tan J, Zhang F, Karcher D, Bock R. Engineering of high-precision base editors for site-specific single nucleotide replacement. *Nature communications* **10**, 1-10 (2019).

27. Liang P*, et al.* Genome-wide profiling of adenine base editor specificity by EndoV-seq. *Nature communications* **10**, 1-9 (2019).

28. Kim D*, et al.* Genome-wide target specificities of CRISPR RNA-guided programmable deaminases. *Nature biotechnology* **35**, 475-480 (2017).

29. Kim D, Kim D-e, Lee G, Cho S-I, Kim J-S. Genome-wide target specificity of CRISPR RNA-guided adenine base editors. *Nature biotechnology* **37**, 430-435 (2019).

30. Kurt IC*, et al.* CRISPR C-to-G base editors for inducing targeted DNA transversions in human cells. *Nature Biotechnology*, 1-6 (2020).

31. Burset M, Seledtsov I, Solovyev V. Analysis of canonical and non-canonical splice sites in mammalian genomes. *Nucleic acids research* **28**, 4364-4375 (2000).

32. Senapathy P, Shapiro MB, Harris NL. [16] Splice junctions, branch point sites, and exons: Sequence statistics, identification, and applications to genome project. (1990).

33. Kluesner MG*, et al.* EditR: a method to quantify base editing from Sanger sequencing. *The CRISPR journal* **1**, 239-250 (2018).

34. Zhao D*, et al.* Glycosylase base editors enable C-to-A and C-to-G base changes. *Nature biotechnology*, 1-6 (2020).

35. Jin S*, et al.* Cytosine, but not adenine, base editors induce genome-wide off-target mutations in rice. *Science* **364**, 292-295 (2019).

36. Zuo E*, et al.* Cytosine base editor generates substantial off-target single-nucleotide variants in mouse embryos. *Science* **364**, 289-292 (2019).

37. Rees HA*, et al.* Improving the DNA specificity and applicability of base editing through protein engineering and protein delivery. *Nature communications* **8**, 1-10 (2017).

38. Liang P*, et al.* Effective gene editing by high-fidelity base editor 2 in mouse zygotes. *Protein & cell* **8**, 601-611 (2017).

39. Lee JK*, et al.* Directed evolution of CRISPR-Cas9 to increase its specificity. *Nature communications* **9**, 1-10 (2018).

40. Xu W*, et al.* Multiplex nucleotide editing by high-fidelity Cas9 variants with improved efficiency in rice. *BMC plant biology* **19**, 1-10 (2019).

41. Slaymaker IM, Gao L, Zetsche B, Scott DA, Yan WX, Zhang F. Rationally engineered Cas9 nucleases with improved specificity. *Science* **351**, 84-88 (2016).

42. Kleinstiver BP*, et al.* High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* **529**, 490-495 (2016).

43. Chen JS*, et al.* Enhanced proofreading governs CRISPR–Cas9 targeting accuracy. *Nature* **550**, 407-410 (2017).

44. Casini A*, et al.* A highly specific SpCas9 variant is identified by in vivo screening in yeast. *Nature biotechnology* **36**, 265-271 (2018).

45. Kulcsár PI*, et al.* Crossing enhanced and high fidelity SpCas9 nucleases to optimize specificity and cleavage. *Genome biology* **18**, 1-17 (2017).

46. Song M*, et al.* Sequence-specific prediction of the efficiencies of adenine and cytosine base editors. *Nature Biotechnology*, 1-7 (2020).

47. Kim N*, et al.* Prediction of the sequence-specific cleavage activity of Cas9 variants. *Nature Biotechnology*, 1-9 (2020).

48. Schmid-Burgk JL*, et al.* Highly parallel profiling of Cas9 variant specificity. *Molecular Cell*, (2020).

49. Wu X*, et al.* Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nature biotechnology* **32**, 670-676 (2014).

50. Hu JH*, et al.* Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature* **556**, 57-63 (2018).

51. Arbab M*, et al.* Determinants of Base Editing Outcomes from Target Library Analysis and Machine Learning. *Cell*, (2020).

52. Okafor IC*, et al.* Single molecule analysis of effects of non-canonical guide RNAs and specificity-enhancing mutations on Cas9-induced DNA unwinding. *Nucleic acids research* **47**, 11880-11888 (2019).

53. Sternberg SH, LaFrance B, Kaplan M, Doudna JA. Conformational control of DNA target cleavage by CRISPR–Cas9. *Nature* **527**, 110-113 (2015).

54. Sternberg SH, Redding S, Jinek M, Greene EC, Doudna JA. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* **507**, 62-67 (2014).

55. Dagdas YS, Chen JS, Sternberg SH, Doudna JA, Yildiz A. A conformational checkpoint between DNA binding and cleavage by CRISPR-Cas9. *Science advances* **3**, eaao0027 (2017).

56. Kulcsár PI*, et al.* Blackjack mutations improve the on-target activities of increased fidelity variants of SpCas9 with 5' G-extended sgRNAs. *Nature communications* **11**, 1-14 (2020).

57. Sung K, Park J, Kim Y, Lee NK, Kim SK. Target specificity of Cas9 nuclease via DNA rearrangement regulated by the REC2 domain. *Journal of the American Chemical Society* **140**, 7778-7781 (2018).

58. Anzalone AV, Koblan LW, Liu DR. Genome editing with CRISPR–Cas nucleases, base editors, transposases and prime editors. *Nature biotechnology*, 1-21 (2020).

59. Richter MF*, et al.* Phage-assisted evolution of an adenine base editor with improved Cas domain compatibility and activity. *Nature Biotechnology*, 1-9 (2020).

60. Lapinaite A*, et al.* DNA capture by a CRISPR-Cas9–guided adenine base editor. *Science* **369**, 566-571 (2020).

61. Anzalone AV*, et al.* Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576**, 149-157 (2019).

62. Bindels DS*, et al.* mScarlet: a bright monomeric red fluorescent protein for cellular imaging. *Nature methods* **14**, 53-56 (2017).

63. Tálas A*, et al.* A convenient method to pre-screen candidate guide RNAs for CRISPR/Cas9 gene editing by NHEJ-mediated integration of a 'self-cleaving'GFP-expression plasmid. *DNA Research* **24**, 609-621 (2017).

64. Richardson CD, Ray GJ, DeWitt MA, Curie GL, Corn JE. Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. *Nature biotechnology* **34**, 339-344 (2016).

Declarations

Acknowledgements

We thank Dr. György Váradi and the FACS core facility for their valuable help with cell sorting. We thank Vivien Karl, Ildikó Szűcsné Pulinka, Judit Szűcs, Lilla Burkus and Judit Kálmán for their excellent technical assistance. We thank Dr. Dora Bokor for proofreading the manuscript.

The project was supported by grants K128188 and K134968 to E.W and PD134858 to P.I.K. from the Hungarian Scientific Research Fund (OTKA) and P.I.K. by 2018-1.1.1-MKI-2018-00167 and by ÚNKP-20-5-SE-20, and E.V. by VEKOP-2.1.7-15-2016-00393 from the National Research, Development and Innovation Office. P. I. K. is a recipient of the János Bolyai Research Scholarship of the Hungarian Academy of Sciences (BO/764/20).

Contributions

A.T. conceived the idea of BEAR, designed and performed the experiments and interpreted the results. D.A.S contributed in molecular cloning and in mismatch screens. P.I.K contributed in creating stable BEAR cell lines. É.V. contributed in molecular cloning. E.W. designed the experiments, interpreted the results and supervised the research. A.T. and E.W. wrote the manuscript with input from all authors.

Competing interests

None declared.

Figures

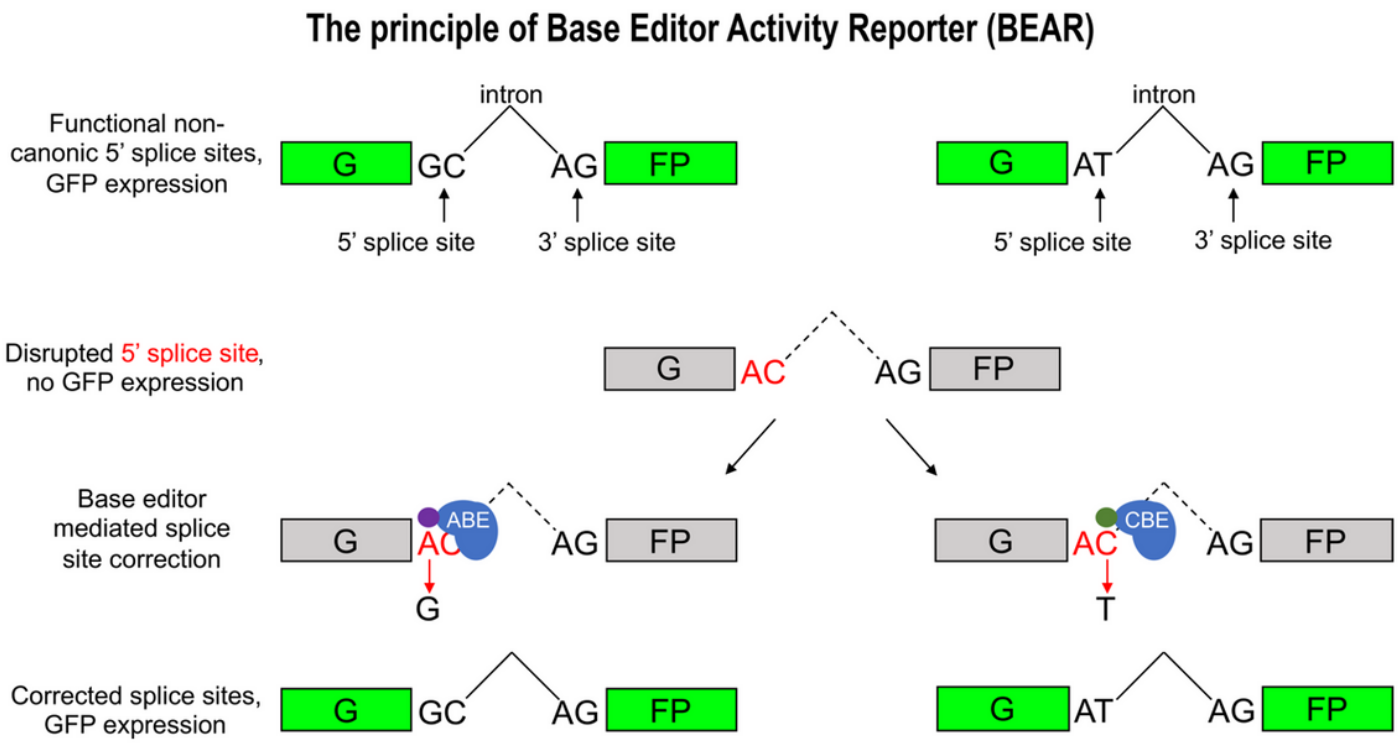

Figure 1

Principle of the Base Editor Activity Reporter (BEAR) assay BEAR consist of a split GFP coding sequence (green) separated with an intron of which the 'GC' or 'AT' 5' splice donor site (5'ss) is disrupted to 'AC', resulting in a dysfunctional protein (grey). These splice sites are known to be functional non-canonical splice sites in the human genome. The disrupted 5'ss is rescued either by ABEs reverting the 'AC' 5'ss to 'AT' or by CBEs reverting the 'AC' 5'ss to 'GC', respectively.

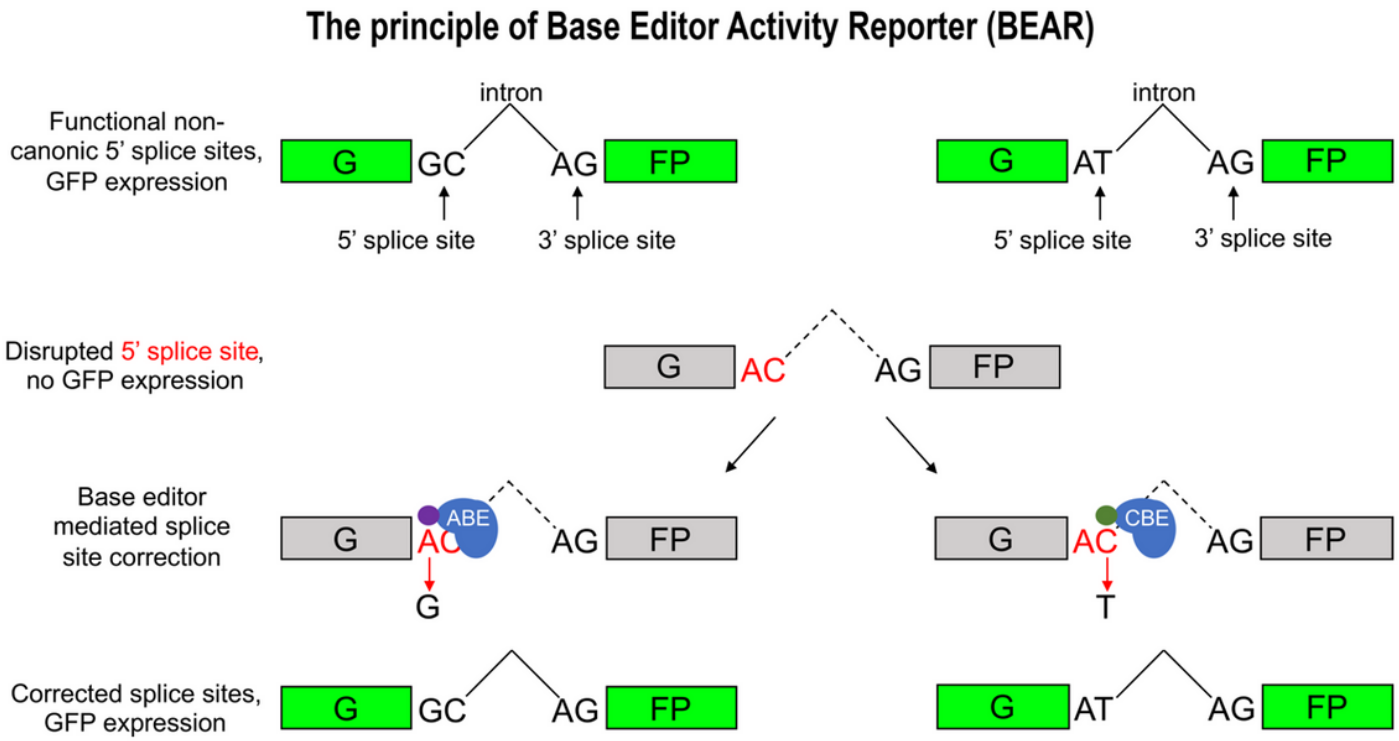

Figure 1

Principle of the Base Editor Activity Reporter (BEAR) assay BEAR consist of a split GFP coding sequence (green) separated with an intron of which the 'GC' or 'AT' 5' splice donor site (5'ss) is disrupted to 'AC', resulting in a dysfunctional protein (grey). These splice sites are known to be functional non-canonical splice sites in the human genome. The disrupted 5'ss is rescued either by ABEs reverting the 'AC' 5'ss to 'AT' or by CBEs reverting the 'AC' 5'ss to 'GC', respectively.

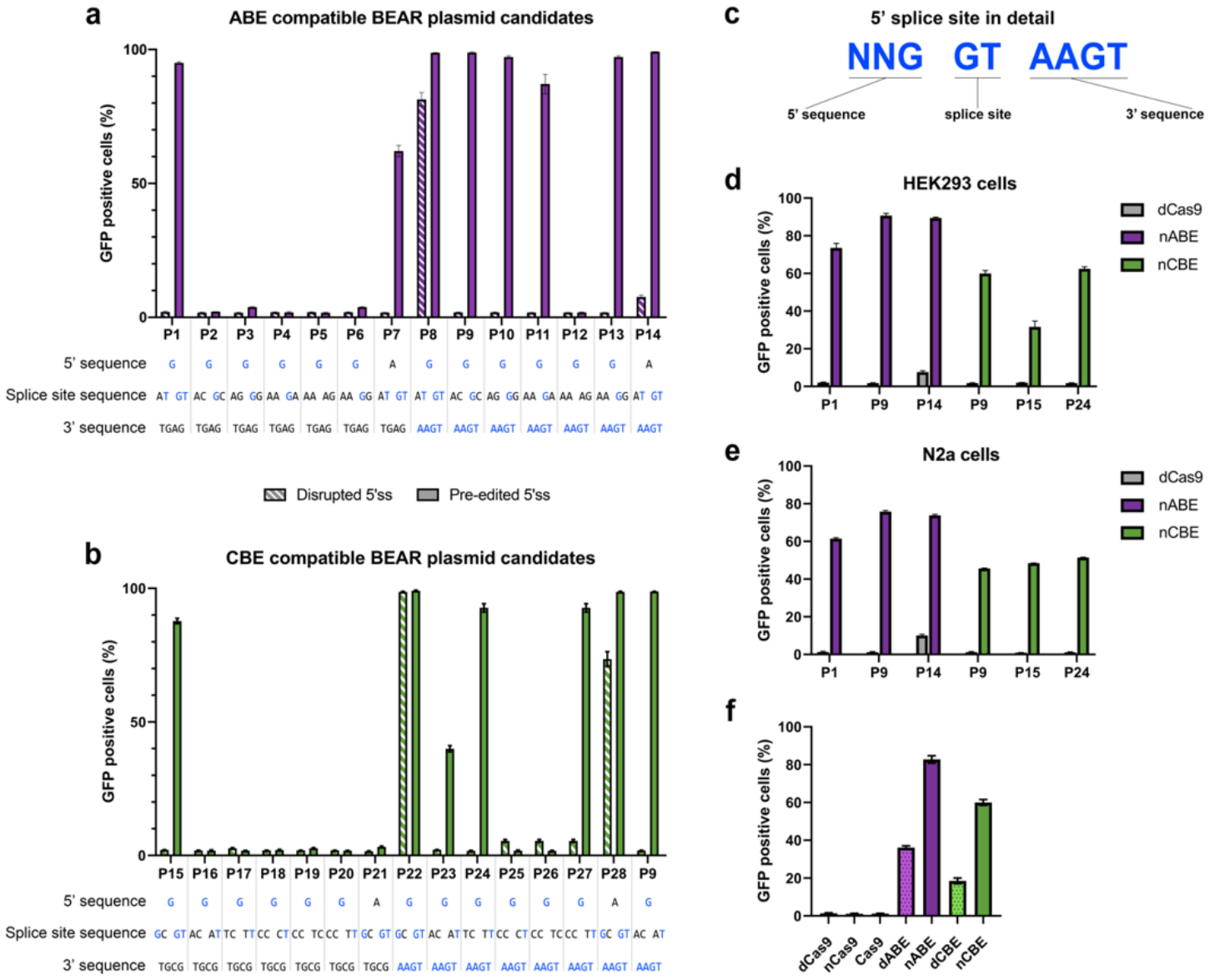

Figure 2

Splice site variants for identifying candidate BEAR sequences Flow cytometry measurements of GFP positive HEK293T cells, transfected with ABE (a) or CBE (b) editing compatible reporter plasmids harboring systematically altered splice sites. The letters beneath the column charts represent the intended disrupted or pre-edited splice site sequences. Letters highlighted in blue indicate the bases that correspond to the canonical 5' - G GT AAGT - 3' consensus splice site sequence (c), and sequence alterations are shown in black. Good candidate sequence pairs with minimal fluorescence for the disrupted and maximal fluorescence for the pre-edited 5'ss were selected for further analyses as detailed in Fig. 2d-f. Flow cytometry measurements of GFP positive HEK293T (d and f) and N2a (e) cells co-transfected with a selected reporter plasmid (d, e) or plasmid 9 (f) harboring a disrupted splice site and a base editor or control nuclease constructs as indicated in the figure. Columns represent means, +/- SD of three parallel transfections (a, b, d-f).

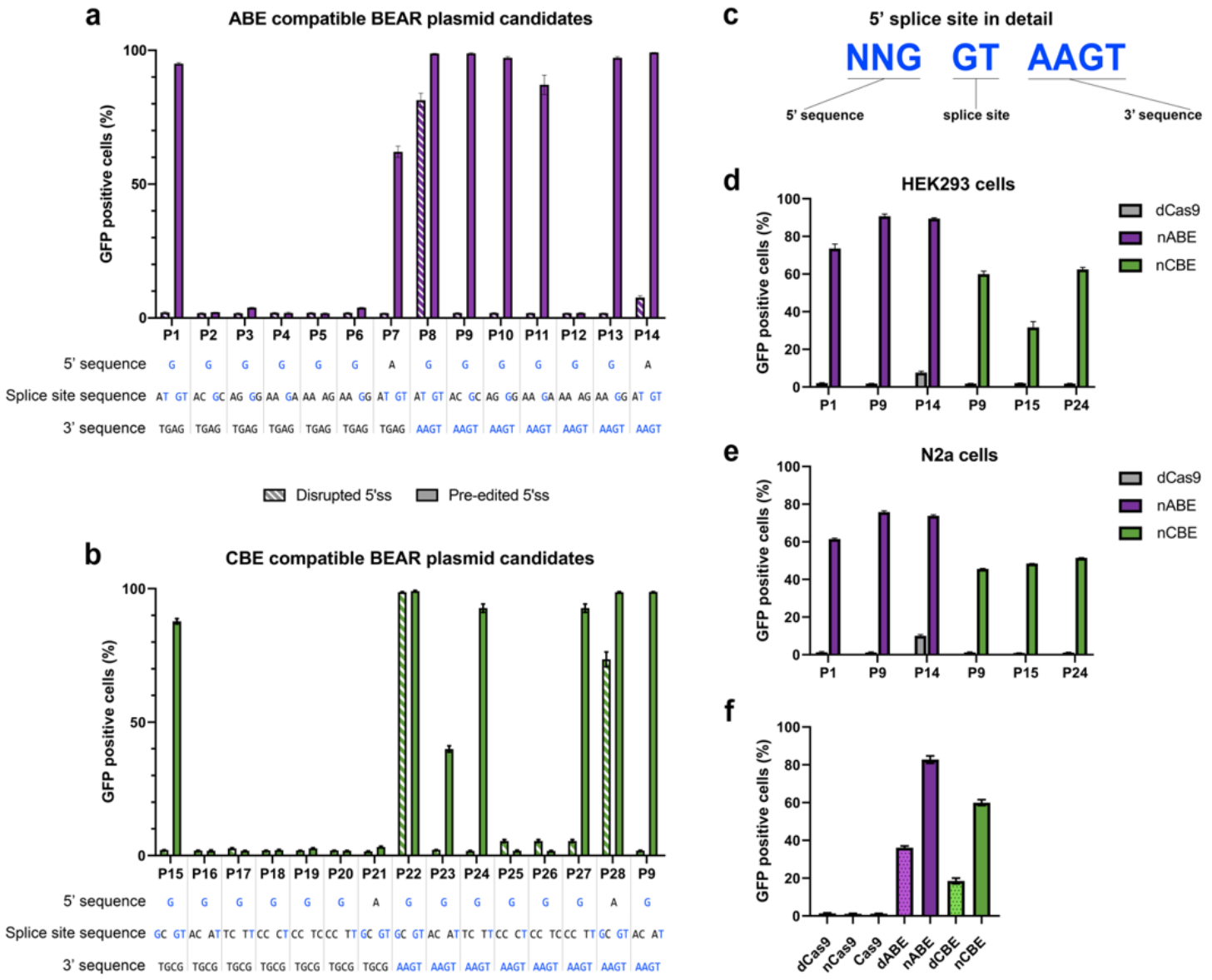

Figure 2

Splice site variants for identifying candidate BEAR sequences Flow cytometry measurements of GFP positive HEK293T cells, transfected with ABE (a) or CBE (b) editing compatible reporter plasmids harboring systematically altered splice sites. The letters beneath the column charts represent the intended disrupted or pre-edited splice site sequences. Letters highlighted in blue indicate the bases that correspond to the canonical 5' - G GT AAGT - 3' consensus splice site sequence (c), and sequence alterations are shown in black. Good candidate sequence pairs with minimal fluorescence for the disrupted and maximal fluorescence for the pre-edited 5'ss were selected for further analyses as detailed in Fig. 2d-f. Flow cytometry measurements of GFP positive HEK293T (d and f) and N2a (e) cells co-transfected with a selected reporter plasmid (d, e) or plasmid 9 (f) harboring a disrupted splice site and a base editor or control nuclease constructs as indicated in the figure. Columns represent means, +/- SD of three parallel transfections (a, b, d-f).

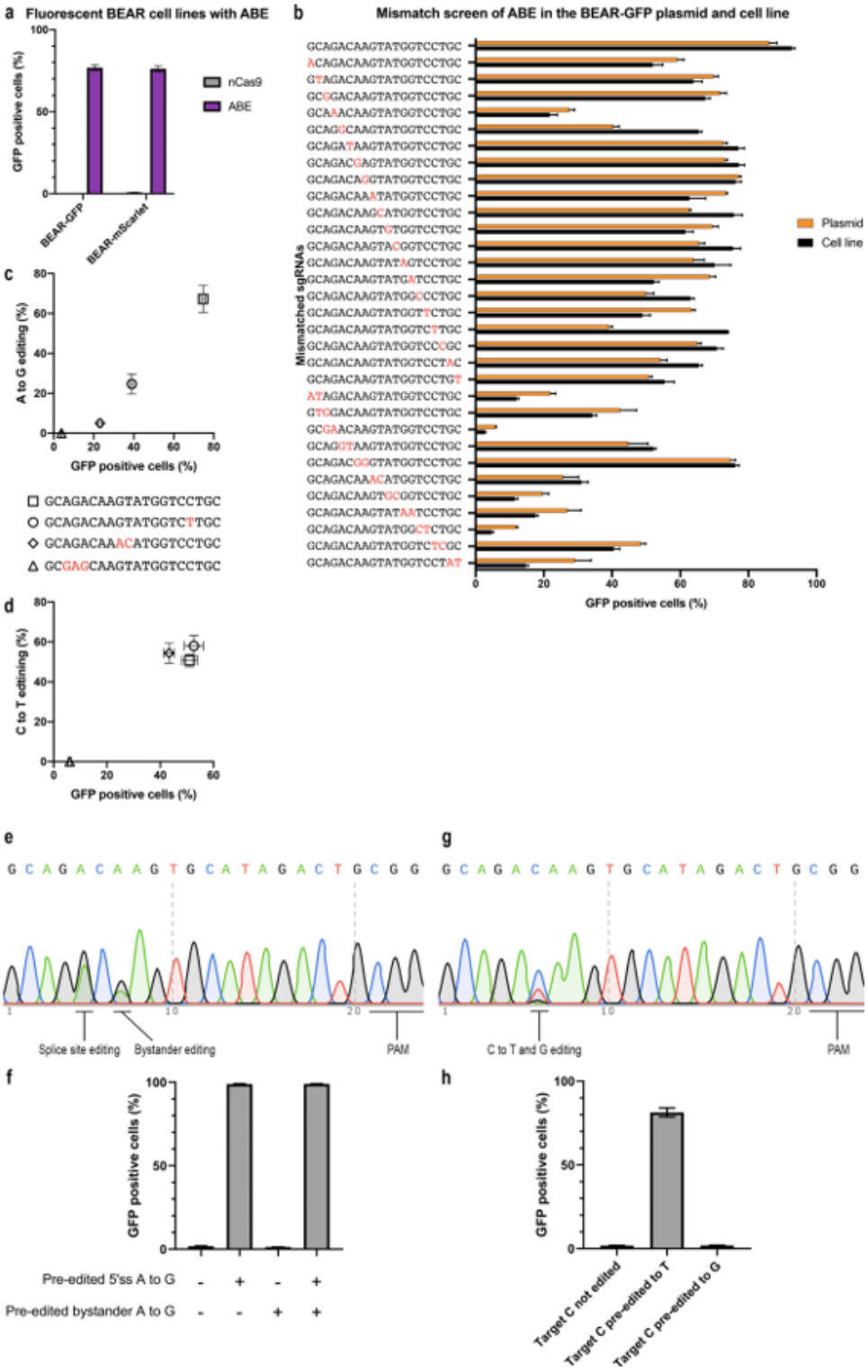

Figure 3

Validation of BEAR in a genomic context (a) Split GFP and mScarlet BEAR reporter constructs were integrated into HEK293T genomes and the resulting BEAR-GFP and BEAR-mScarlet cell lines were targeted by ABE and by nCas9 as a negative control. Columns represent means +/- SD of three parallel transfections. (b) The BEAR-GFP cell line (black bars) and the BEAR plasmid (orange bars) were targeted by ABE using sgRNAs, one matching and 31 mismatching in one or two positions. Recovery of the GFP

signal measured by flow cytometry was compared, and a correlation of r=0.89 was found. Columns represent means +/- SD of three parallel transfections. (c,d) The BEAR-GFP cell line was targeted by ABE (c) or CBE (d) using sgRNAs, one matching and three mismatching in one, two or three bases. GFP positive cells were quantified by flow cytometry and the resulting base editing was quantified by Sanger sequencing followed by EditR. A correlation of 0.98 and 0.98 was found for ABE and CBE, respectively. Scatter plots represent means, error bars represent +/- SD of three parallel transfections. (e,g) Chromatograms of Sanger sequencing of the target region of the BEAR-GFP cell line. (e) Sequencing of the ABE-edited BEAR-GFP cell line revealed that a bystander adenine was edited in the consensus splice site sequence. (f) Testing the corresponding pre-edited plasmids proved that editing the second, bystander adenine alone or together with the adenine of the 5'ss sequence does not influence GFP fluorescence. (g) In case of the CBE-edited cell line, no bystander edits were seen. However, a portion of the edited C was converted into G instead of T. (h) Testing the corresponding pre-edited plasmids proved that changing the edited C to G does not influence GFP fluorescence.

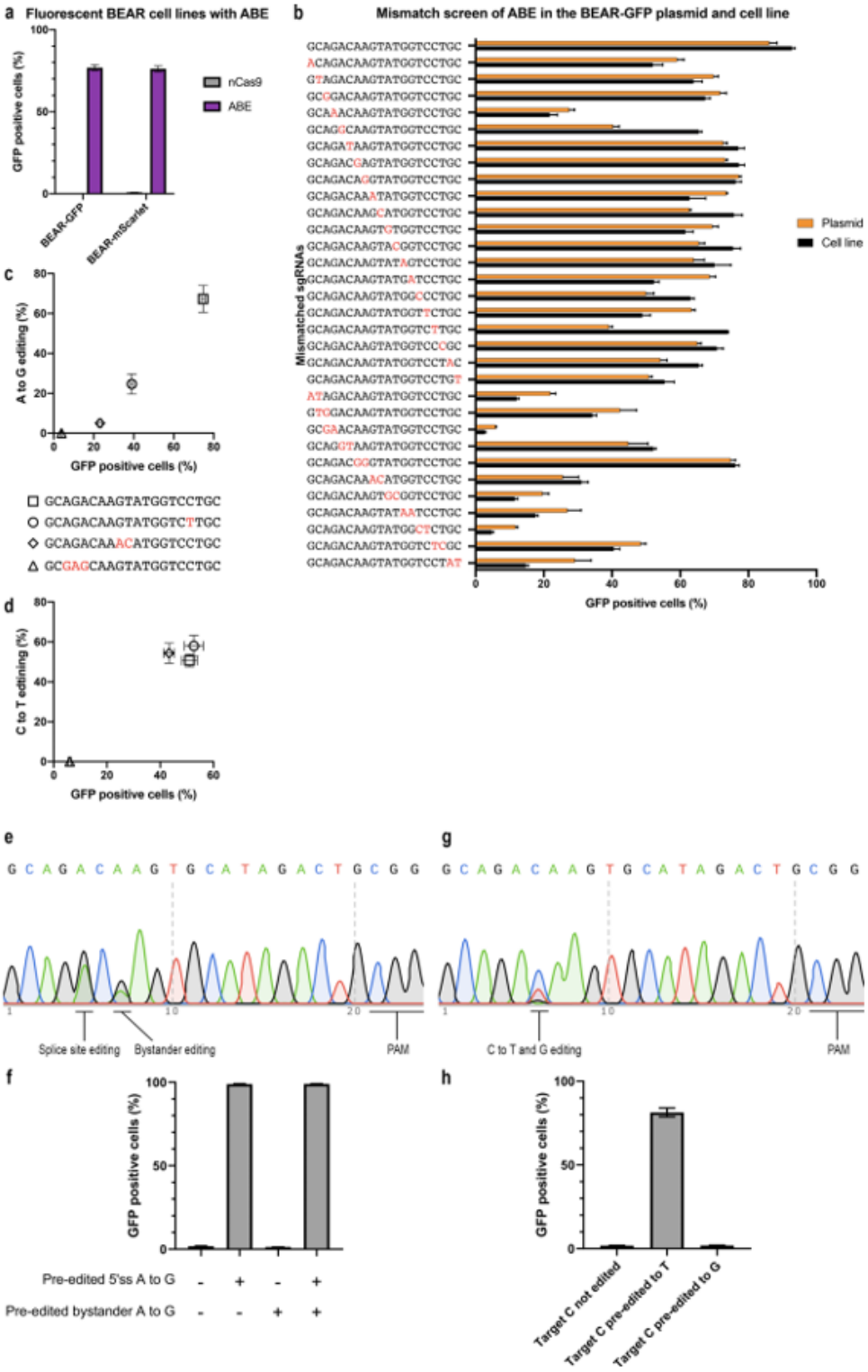

Figure 3

Validation of BEAR in a genomic context (a) Split GFP and mScarlet BEAR reporter constructs were integrated into HEK293T genomes and the resulting BEAR-GFP and BEAR-mScarlet cell lines were targeted by ABE and by nCas9 as a negative control. Columns represent means +/- SD of three parallel transfections. (b) The BEAR-GFP cell line (black bars) and the BEAR plasmid (orange bars) were targeted by ABE using sgRNAs, one matching and 31 mismatching in one or two positions. Recovery of the GFP

signal measured by flow cytometry was compared, and a correlation of r=0.89 was found. Columns represent means +/- SD of three parallel transfections. (c,d) The BEAR-GFP cell line was targeted by ABE (c) or CBE (d) using sgRNAs, one matching and three mismatching in one, two or three bases. GFP positive cells were quantified by flow cytometry and the resulting base editing was quantified by Sanger sequencing followed by EditR. A correlation of 0.98 and 0.98 was found for ABE and CBE, respectively. Scatter plots represent means, error bars represent +/- SD of three parallel transfections. (e,g) Chromatograms of Sanger sequencing of the target region of the BEAR-GFP cell line. (e) Sequencing of the ABE-edited BEAR-GFP cell line revealed that a bystander adenine was edited in the consensus splice site sequence. (f) Testing the corresponding pre-edited plasmids proved that editing the second, bystander adenine alone or together with the adenine of the 5'ss sequence does not influence GFP fluorescence. (g) In case of the CBE-edited cell line, no bystander edits were seen. However, a portion of the edited C was converted into G instead of T. (h) Testing the corresponding pre-edited plasmids proved that changing the edited C to G does not influence GFP fluorescence.

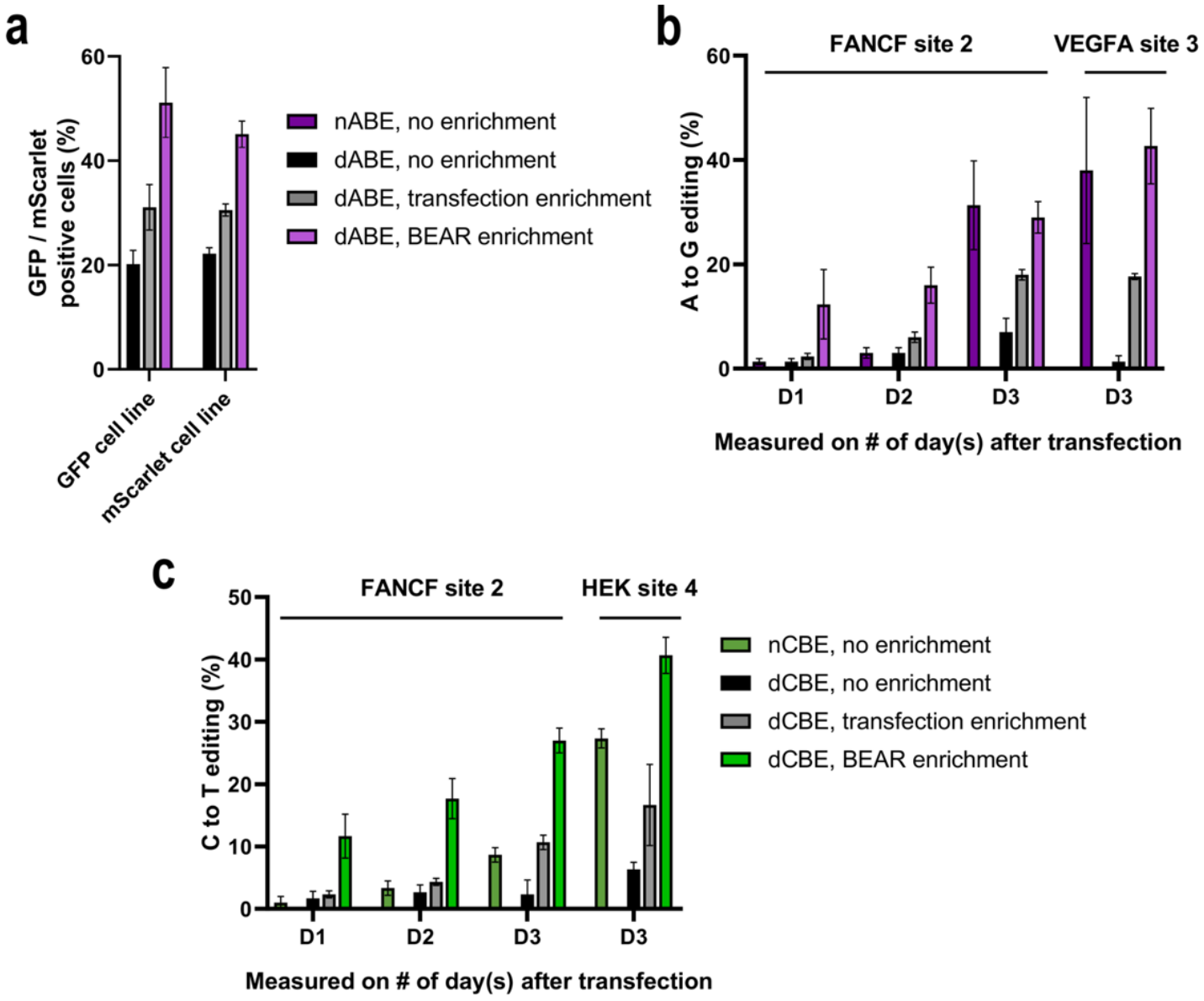

Figure 4

Enrichment of base edited cells with dead ABE and CBE (a) The BEAR-GFP plasmid and dABE were co-transfected into the BEAR-mScarlet cell line, and the cells were analysed by flow cytometry. To monitor cells with no enrichment (no enrichment, black bars), mScarlet positive cell count was determined among all live single cells gated via flow cytometry. To monitor transfection enrichment (transfection enrichment, gray bars), mScarlet positive cell count was determined among the BFP positive cells which were the reporter for transfection efficiency. BEAR enriched cells were measured by counting the amount of mScarlet positive cells among the GFP positive cells (BEAR enrichment, purple bars). Enrichment was also monitored when the BEAR-mScarlet plasmid was used as the enrichment reporter in BEAR-GFP cell lines. (b-c) Sanger sequencing of the PCR product amplified from the indicated genomic regions and EditR were used to assess genome editing efficiency. The BEAR-GFP plasmid and endogenous genomic targets (FANCF site 2 (b, c), VEGFA site 3 (b), and HEK site 4 (c)) were edited by dABE (b) and dCBE (c). Edited cells were sorted to 3 fractions: all cells (no enrichment, black), mCherry positive cells (transfection enrichment, grey), and cells with mCherry plus GFP positivity representing base editing enriched cells (dCBE – light green, dABE – light purple). In all experiments nABE (purple) and nCBE (green) edited cells were monitored without enrichment as controls. Columns represent means +/- SD of three parallel transfections.

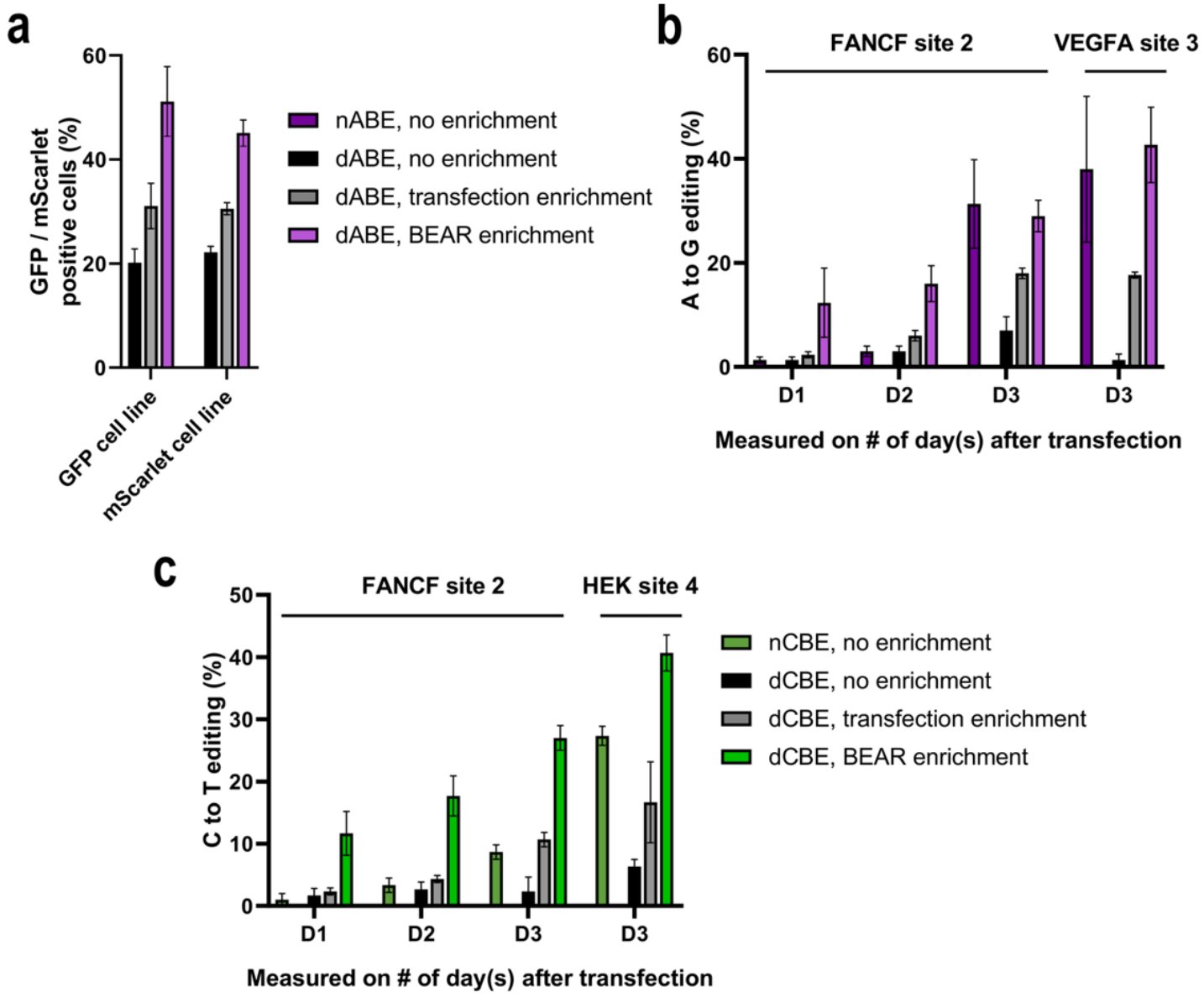

Figure 4

Enrichment of base edited cells with dead ABE and CBE (a) The BEAR-GFP plasmid and dABE were co-transfected into the BEAR-mScarlet cell line, and the cells were analysed by flow cytometry. To monitor cells with no enrichment (no enrichment, black bars), mScarlet positive cell count was determined among all live single cells gated via flow cytometry. To monitor transfection enrichment (transfection enrichment, gray bars), mScarlet positive cell count was determined among the BFP positive cells which were the reporter for transfection efficiency. BEAR enriched cells were measured by counting the amount of mScarlet positive cells among the GFP positive cells (BEAR enrichment, purple bars). Enrichment was also monitored when the BEAR-mScarlet plasmid was used as the enrichment reporter in BEAR-GFP cell lines. (b-c) Sanger sequencing of the PCR product amplified from the indicated genomic regions and EditR were used to assess genome editing efficiency. The BEAR-GFP plasmid and endogenous genomic targets (FANCF site 2 (b, c), VEGFA site 3 (b), and HEK site 4 (c)) were edited by dABE (b) and dCBE (c). Edited cells were sorted to 3 fractions: all cells (no enrichment, black), mCherry positive cells (transfection

enrichment, grey), and cells with mCherry plus GFP positivity representing base editing enriched cells (dCBE – light green, dABE – light purple). In all experiments nABE (purple) and nCBE (green) edited cells were monitored without enrichment as controls. Columns represent means +/- SD of three parallel transfections.

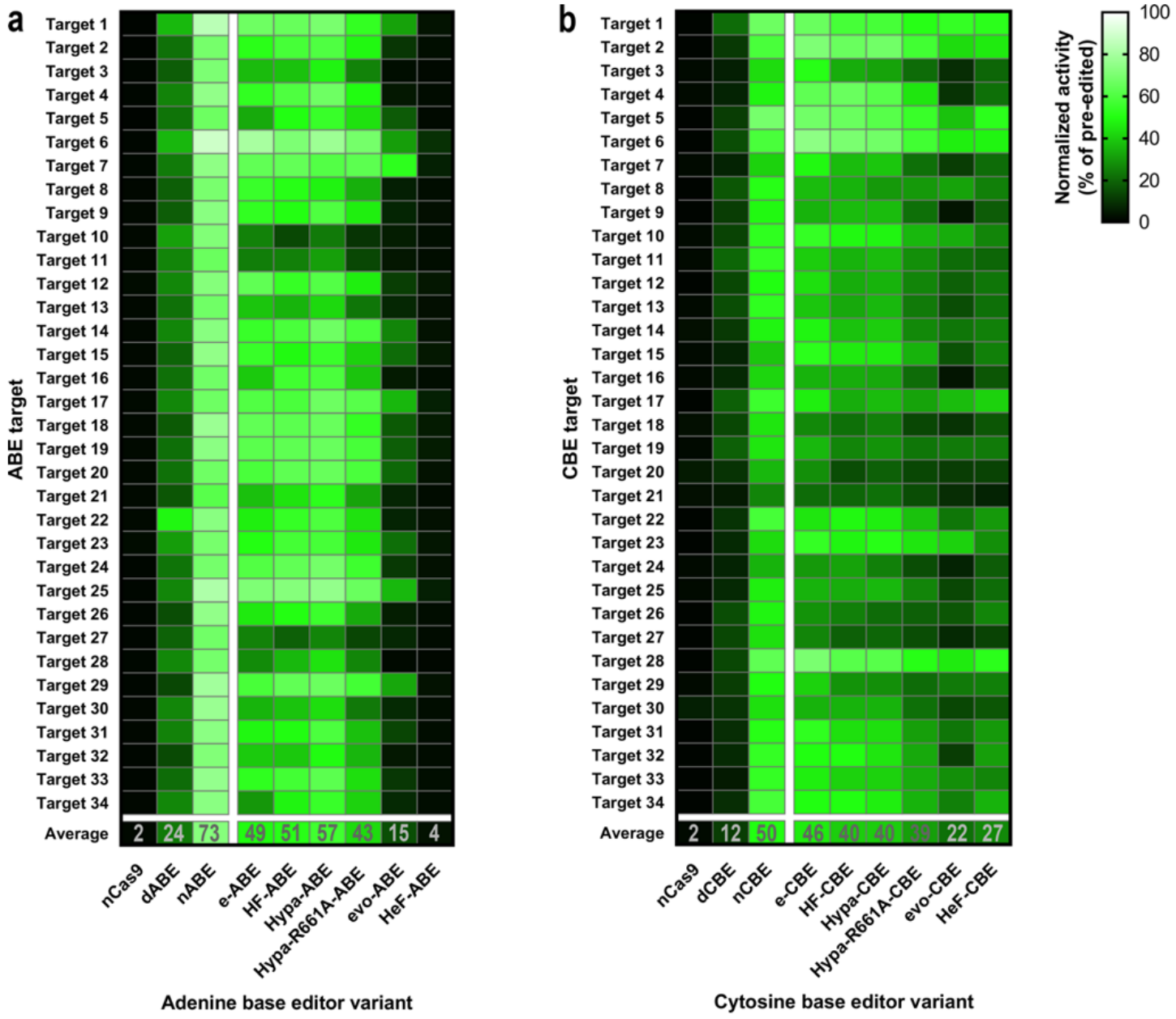

Figure 5

On-target activities of different increased fidelity ABE and CBE variants on 34 target sites (a, b) The heatmap shows normalized on-target activity (measured/pre-edited) derived from three parallel transfections. A total of 34 target sequences, differing in their PAM proximal 10 nucleotides, were constructed. dABE, nABE and 6 increased fidelity ABEs (a) as well as dCBE, nCBE and 6 increased fidelity CBEs (b) were used to target the same 34 target sites as indicated in the figure. As a negative control, nCas9 was used with all 34 targets.

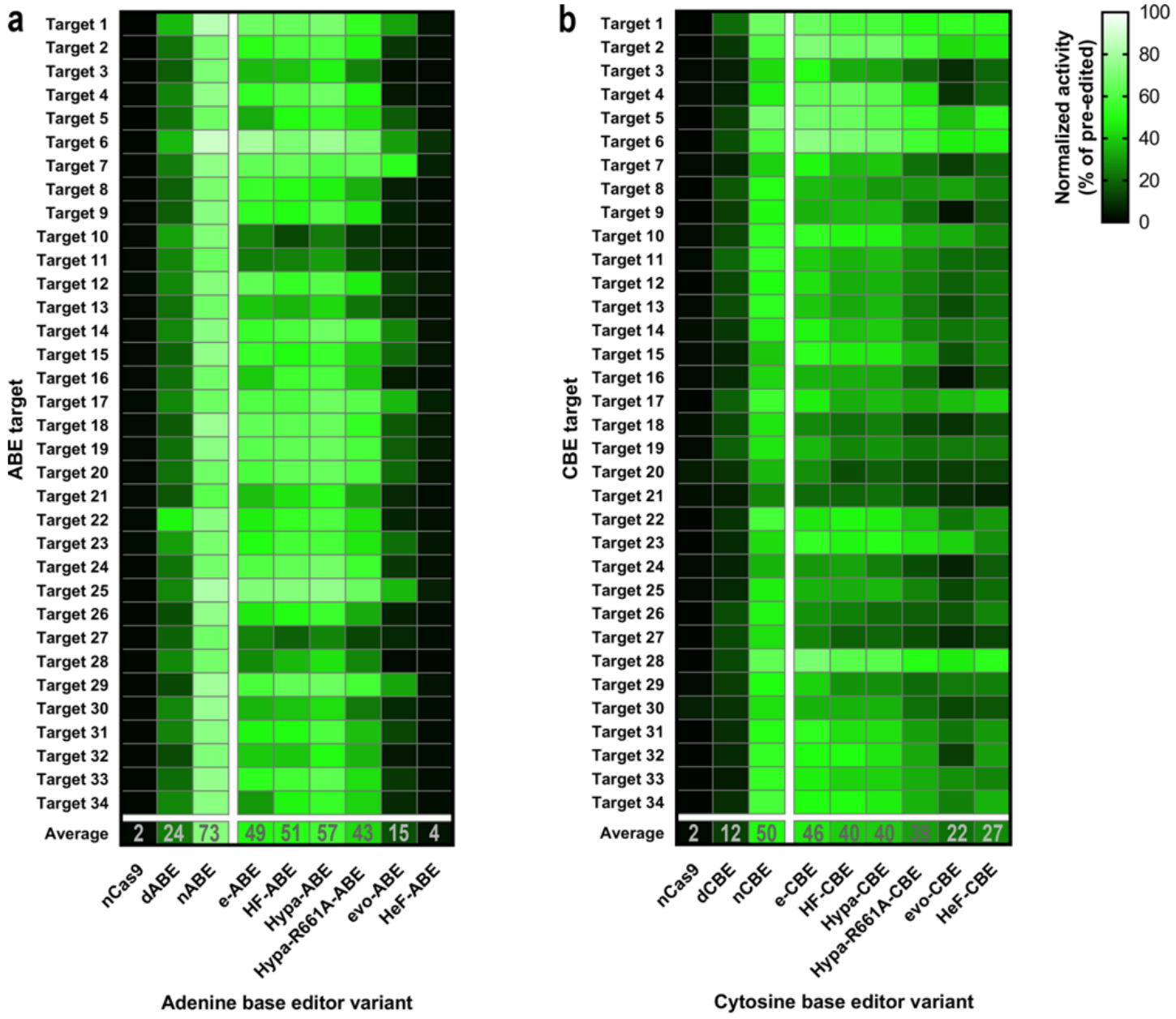

Figure 5

On-target activities of different increased fidelity ABE and CBE variants on 34 target sites (a, b) The heatmap shows normalized on-target activity (measured/pre-edited) derived from three parallel transfections. A total of 34 target sequences, differing in their PAM proximal 10 nucleotides, were constructed. dABE, nABE and 6 increased fidelity ABEs (a) as well as dCBE, nCBE and 6 increased fidelity CBEs (b) were used to target the same 34 target sites as indicated in the figure. As a negative control, nCas9 was used with all 34 targets.

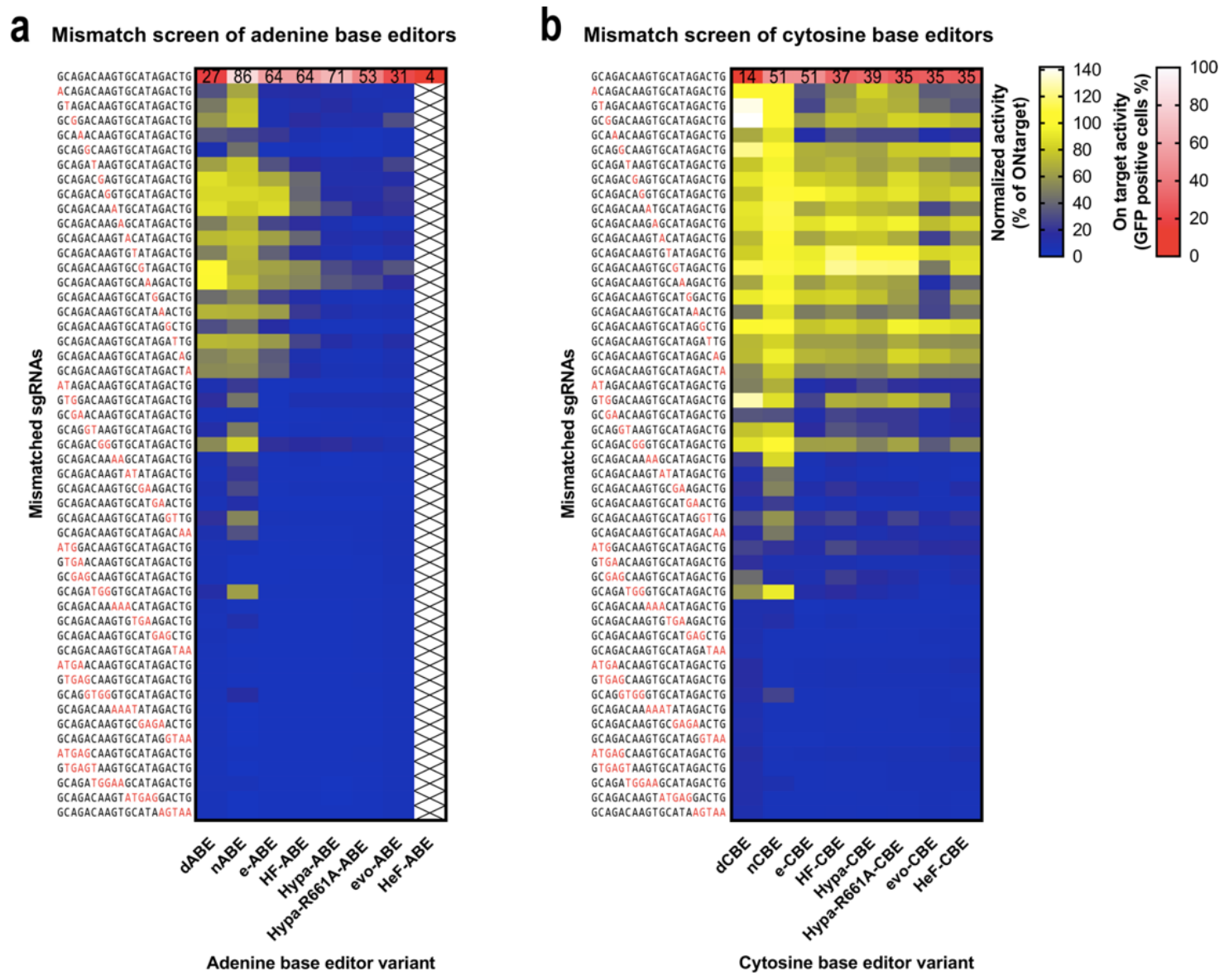

Figure 6

Off-target activities of different ABE and CBE variants with 50 mismatching sgRNAs Mismatch tolerance of ABE (a) and CBE (b) and their increased fidelity variants were compared utilizing exactly the same matching sgRNA (target 1 in Fig. 5) and 50 sgRNAs mismatching in one, two, three, four or five positions. Blue-yellow heatmaps show the mean normalized activity (off-target/on-target) derived from three parallel transfections. White-red heatmaps show the on-target activity (mean rates of GFP positive cells) derived from three parallel transfections.

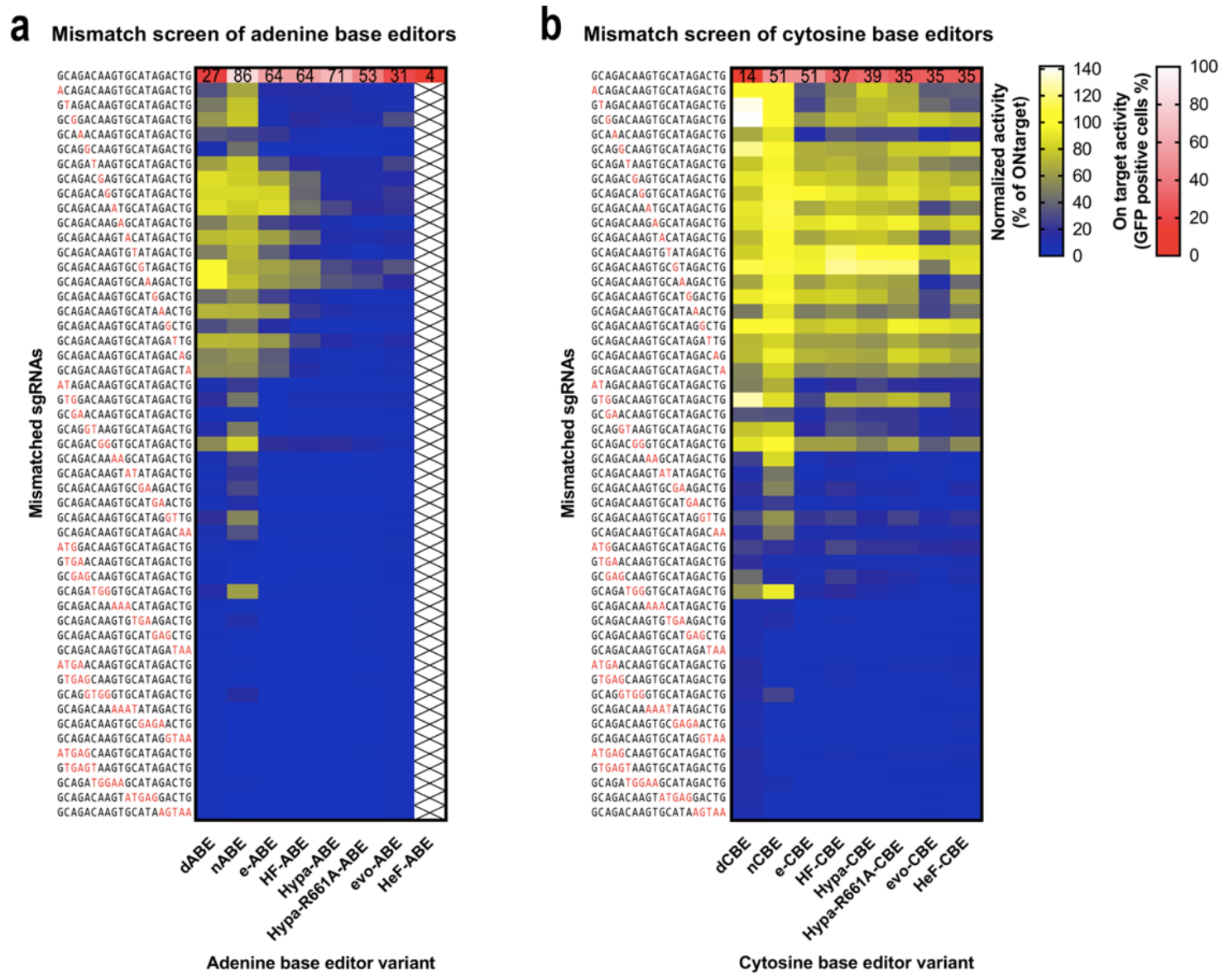

Figure 6

Off-target activities of different ABE and CBE variants with 50 mismatching sgRNAs Mismatch tolerance of ABE (a) and CBE (b) and their increased fidelity variants were compared utilizing exactly the same matching sgRNA (target 1 in Fig. 5) and 50 sgRNAs mismatching in one, two, three, four or five positions. Blue-yellow heatmaps show the mean normalized activity (off-target/on-target) derived from three parallel transfections. White-red heatmaps show the on-target activity (mean rates of GFP positive cells) derived from three parallel transfections.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryInformation.docx
- SupplementaryInformation.docx