nature communications

Article

https://doi.org/10.1038/s41467-024-53735-y

Cytosine base editors with increased PAM and deaminase motif flexibility for gene editing in zebrafish

Received: 8 November 2023

Accepted: 21 October 2024

Published online: 04 November 2024

Check for updates

Yu Zhang[1,2,3,7], Yang Liu[1,2,7], Wei Qin[3,7], Shaohui Zheng[1,2,7], Jiawang Xiao[1,2], Xinxin Xia[1,2], Xuanyao Yuan[1,2], Jingjing Zeng[1,2], Yu Shi[1,2], Yan Zhang[1,2], Hui Ma[4,5], Gaurav K. Varshney[3] ✉, Ji-Feng Fei[4,5,6] ✉ & Yanmei Liu[1,2] ✉

Cytosine base editing is a powerful tool for making precise single nucleotide changes in cells and model organisms like zebrafish, which are valuable for studying human diseases. However, current base editors struggle to edit cytosines in certain DNA contexts, particularly those with GC and CC pairs, limiting their use in modelling disease-related mutations. Here we show the development of zevoCDA1, an optimized cytosine base editor for zebrafish that improves editing efficiency across various DNA contexts and reduces restrictions imposed by the protospacer adjacent motif. We also create zevoCDA1-198, a more precise editor with a narrower editing window of five nucleotides, minimizing off-target effects. Using these advanced tools, we successfully generate zebrafish models of diseases that were previously challenging to create due to sequence limitations. This work enhances the ability to introduce human pathogenic mutations in zebrafish, broadening the scope for genomic research with improved precision and efficiency.

Single nucleotide variants (SNVs) account for more than 96% of the observed genetic variation between humans[1]. About half of all SNVs are non-synonymous, encoding protein variants with potential alterations in function or stability that may cause disease[2]. Precise animal models carrying specific SNVs are critical to determine whether and how SNVs cause pathogenic effects and to explore potential treatments. In particular, zebrafish have emerged as an ideal model system for studying genetic diseases due to their small size, high reproduction, in vitro development, transparency, and high genetic conservation with humans.

The CRISPR-Cas9 system has revolutionized the speed, ease, and efficiency with which genome editing can be achieved[3]. Cas9 targeting of a genetic locus requires complementarity between its single guide RNA (sgRNA) and target DNA, as well as recognition of the protospacer adjacent motif (PAM) immediately downstream of the target site (e.g. NGG for *Streptococcus pyogenes* Cas9). Following the resultant Cas9-mediated double-stranded DNA break (DSB), cellular DNA repair processes typically cause random insertions or deletions (indels) at DNA break sites, often resulting in gene disruption. More precise installation of specific point mutations at the target locus through homology-directed repair is very inefficient[4].

Base editing combines Cas9-mediated programmable genome targeting with other enzymes to chemically convert one nucleotide to another within a Cas9-defined "editing window" of approximately 3–9 nucleotides[5–8]. Cytosine base editors (CBE), which convert a C•G base pair into a T•A pair, and adenine base editors (ABE), which convert an

[1]Key Laboratory of Brain, Cognition and Education Sciences, Ministry of Education, South China Normal University, Guangzhou, China. [2]Institute for Brain Research and Rehabilitation, and Guangdong Key Laboratory of Mental Health and Cognitive Science, South China Normal University, Guangzhou, China. [3]Genes & Human Disease Research Program, Oklahoma Medical Research Foundation, Oklahoma City, OK, USA. [4]Department of Pathology, Guangdong Provincial People's Hospital (Guangdong Academy of Medical Sciences), Southern Medical University, Guangzhou, Guangdong, China. [5]School of Basic Medical Sciences, Southern Medical University, Guangzhou, China. [6]The Innovation Centre of Ministry of Education for Development and Diseases, School of Medicine, South China University of Technology, Guangzhou, China. [7]These authors contributed equally: Yu Zhang, Yang Liu, Wei Qin, Shaohui Zheng. ✉e-mail: gaurav-varshney@omrf.org; jifengfei@gdph.org.cn; yanmeiliu@m.scnu.edu.cn

A•T base pair into a G•C pair, are widely used in plant and animal systems[9–12]. C•G to T•A transitions, mainly caused by the spontaneous deamination of cytosine, account for about half of all known human pathogenic SNVs[5], and CBEs should, in theory, be able to generate model organisms bearing these changes. In practice, however, even the most advanced and efficient CBEs (e.g. BE3, BE4, BE4max to AncBE4max, all containing a rat APOBEC1 deaminase domain[5–7,13–15]) suffer from sequence context preferences that limit which sites they can target. APOBEC1 has a strong preference to edit the C in TC rather than GC motifs, and these biases are more pronounced in zebrafish compared to human cell lines. To overcome these APOBEC1 biases, researchers have developed CBEs incorporating the CDA1 deaminase domain instead[16], which is known to edit GC targets more efficiently[17]. However, they possess quite low editing efficiency in zebrafish[18]. A variant known as evoCDA1-BE4max, generated through phage-assisted continuous evolution[19], exhibits improved editing efficiency in human cells and drosophila[19,20], but at the expense of precision, owing to its expanded editing window, higher indels, and increased off-targets.

Here, we develop a set of engineered CBEs to enable efficient and precise on-target cytosine base editing for hard-to-edit sites in zebrafish and demonstrate their ability to model human diseases caused by SNVs. First, our codon-optimized zevoCDA1-BE4max editor shows high editing efficiency on GC and CC sites that were inaccessible using the current state-of-the-art zebrafish CBE, zAncBE4max or original evoCDA1-BE4max. Incorporating a PAM-flexible Cas9 variant, SpRYCas9[21], we develop zevoCDA1-SpRY-BE4max, which enables broad targeting capabilities using any PAM sequence paired with higher editing efficiency than the previous SpRY-CBE4max[22]. To improve precision, we engineer two variants of this, zevoCDA1-NL or zevoCDA1-198 with narrowed editing windows to pinpoint CBE activity to only 7 or 5 nucleotides at the PAM-distal end of the Cas9 target site, respectively. Using these tools, we successfully generate a precise disease model of Axenfeld–Rieger syndrome (ARS) and a previously unavailable zebrafish disease model of oculocutaneous albinism (OCA). Compared to zevoCDA1-SpRY-BE4max and zevoCDA1-NL, zevoCDA1-198 exhibits a lower indel rate and off-target effect, demonstrating promising potential as a complementary option to the context sequence-biased mainstream CBEs in zebrafish.

Results

Optimized zevoCDA1-BE4max overcomes sequence context restriction

To develop a CBE capable of editing GC and CC targets in zebrafish, we initially employed evoCDA1-BE4max to edit 12 genes using 12 sgRNAs-targeting specific loci, with zAncBE4max as a control. Following the injection of evoCDA1-BE4max or zAncBE4max mRNA and relevant 2′-O-methyl-3′-phosphorothioate (MS)-modified gRNAs, into one-cell stage zebrafish embryos, we extracted genomic DNA at 48 h post-fertilization (hpf) to analyze base editing outcomes. While evoCDA1-BE4max demonstrated higher editing efficiency at some GC or CC sites than zAncBE4max, it was less effective at some TC or AC sites (Fig. 1a, b and Supplementary Fig. 1). We then optimized evoCDA1-BE4max according to the zebrafish codon preference to create zevoCDA1-BE4max, which includes N-terminal and C-terminal bipartite nuclear localization signal (bpNLS), evoCDA1 cytidine deaminase (with three amino acid mutations compared to CDA1), spCas9n (D10A) nickase, and two uracil glycosylase inhibitors (UGIs) in tandem (Fig. 1a). Notably, zAncBE4max produced over 10% C-to-T base substitutions at 2 out of 6 TC sites, 8 out of 10 AC sites, 4 out of 9 CC sites, and 1 out of 8 GC sites among the 12 target loci examined. In contrast, zevoCDA1-BE4max achieved significant C-to-T conversions at all TC, AC, CC, and GC motif-containing sites (Fig. 1b and Supplementary Fig. 1). In cases where both CBEs edited, the editing efficiencies of zevoCDA1-BE4max largely exceeded those of zAncBE4max and evoCDA1-BE4max. Analysing more site edits (a total of 20 genes and 21 sites) revealed that zevoCDA1-BE4max had a main working window ranging from positions 1 to 9 at the PAM-distal end of the Cas9 target site (Fig. 1c). These results indicate that zevoCDA1-BE4max breaks the editing limitations of zAncBE4max for CC and GC sites and can target cytosines in any sequence context in zebrafish.

Despite this significant progress in overcoming sequence context preferences, previous studies noted that evoCDA1-BE4max generated 6.8–20% indels in HEK293T cells[19]. Similarly, in addition to the 21 targeting sites discussed previously, we identified approximately one third (12/33) of the zevoCDA1-BE4max target regions displayed distinct overlapping peaks around the target sites in the Sanger sequencing chromatograms (Fig. 1c–f), suggesting the indels alongside cytosine base editing. The relatively high indel frequency necessitates further engineering to enhance the precise editing of zevoCDA1-BE4max as a powerful tool for zebrafish base editing.

zevoCDA1-SpRY-BE4max exhibits high activities at non-canonical PAM sites in the zebrafish genome

The most flexible SpCas9 variant SpRY and its related base editors SpRY-CBE4max have been reported to target almost all PAM sequences in the genomes of cultured cells, plants, and zebrafish[21–27]. Recently, we and Rosello et al. reported independently that SpRY-CBE4max recognizes almost all NRN PAM sequences and expands the potential to target previously inaccessible bases in zebrafish for base editing[22,24]. However, SpRY-CBE4max also contains the limitation of sequence context preference and hardly edits GC. Building on our above success in overcoming sequence context preference with zevoCDA1-BE4max, we next developed zevoCDA1-SpRY-BE4max by replacing the SpCas9n moiety with SpRYCas9n (Fig. 2a). Using 30 sgRNAs targeting non-canonical PAMs in 18 genes, including *raf1b*, *trappc10*, *prpf4*, *mek11* and *gars1*, we assessed the base editing activity of zevoCDA1-SpRY-BE4max at every cytosine site regardless of sequence context. Our results showed C-to-T base conversions at nearly all NC sites within the primary editing window (positions 1–9, at the PAM-distal end of the Cas9 target site), with significantly higher editing efficiency for both NRN and NYN PAMs compared to SpRY-CBE4max (Fig. 2b–f). Notably, SpRY-CBE4max exhibited minimal activity against most targets with NYN PAMs – effectively editing only 2 out of 23 cytosines with an efficiency over 25%. In contrast, zevoCDA1-SpRY-BE4max successfully edited 10 of these same sites with efficiencies ranging from 25% to 90% (Fig. 2d, e).

We further investigated the indel rate occurring during zevoCDA1-SpRY-BE4max editing. We selected six sgRNAs targeting sites where zevoCDA1-BE4max produced significant indel peaks in Sanger sequencing chromatograms. Using next-generation sequencing (NGS), we observed that zevoCDA1-SpRY-BE4max induced markedly lower rates of indels than zevoCDA1-BE4max at these six NGG PAM sites (Supplementary Fig. 2a and Supplementary Table 1), while maintaining similar editing efficiency (Supplementary Fig. 2b and Supplementary Table 1). We also examined six non-NGG PAM sites and found that the indel ratios caused by zevoCDA1-SpRY-BE4max at these six sites are also relatively low (Supplementary Fig. 3 and Supplementary Table 2). Taken together, the combination of PAM-flexible targeting by SpRY-Cas9 with the zevoCDA1 editor provides high efficiency and fidelity while alleviating the previously restrictive sequence context preference of most base editors.

To put these capabilities into practice, we designed a sgRNA to direct zevoCDA1-SpRY-BE4max to create a zebrafish SNV disease model that could not be generated with previous editors. *SLC24A5* (*solute carrier family 24, member 5*) encodes the NCKX5 protein, a potassium-dependent calcium, potassium: sodium antiporter involved in pigmentation in melanocytes. The c.184 C > T (p.Q62*) mutation in SLC24A5 causes oculocutaneous albinism (OCA), a hypopigmentation disorder accompanied by impaired visual acuity in humans[28]. The amino acid sequences of SLC24A5 are conserved between human and

https://doi.org/10.1038/s41467-024-53735-y

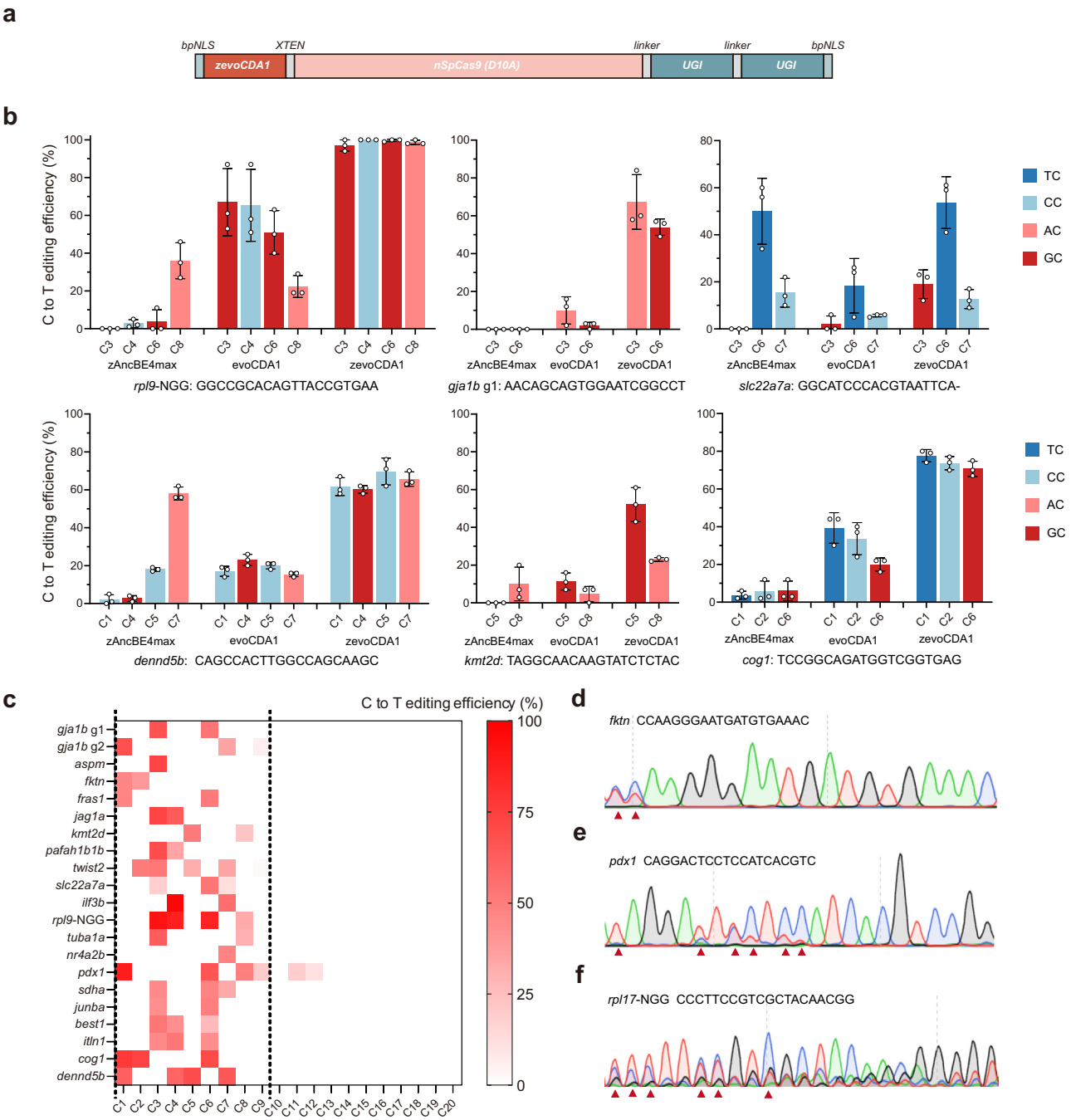
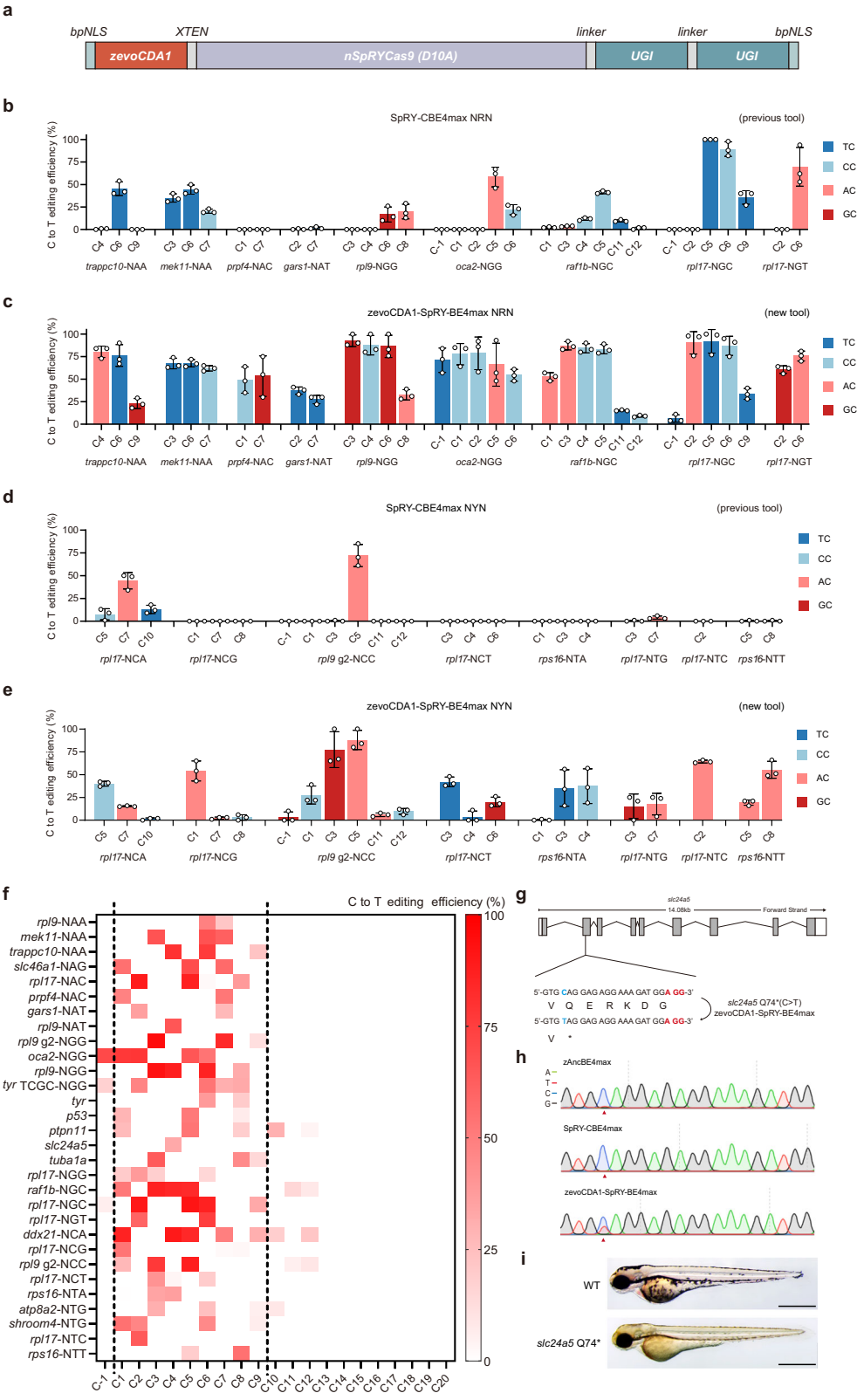


Fig. 1 | Cytosine base editing in all sequence contexts mediated by zevoCDA1-BE4max. a Schematic of the mRNA construct for zevoCDA1-BE4max. bpNLS: bipartite nuclear localization, zevoCDA1: cytosine deaminase, XTEN: a 32aa flexible linker, nSpCas9: SpCas9 nickase, linker: SGGSSGGS amino acid, UGI: Uracil glycosylase inhibitor. **b** Comparison of the editing efficiency of zAncBE4max, evoCDA1-BE4max, and zevoCDA1-BE4max, targeting six loci with NGG PAM. The data represent the aggregate result of three independently replicated experiments, and the error bars indicate the standard deviation of the mean values. Target sequence information is displayed below the data, respectively. **c** Summary of editing efficiency of zevoCDA1-BE4max at 21 NGG PAM sites on 20 genes. **d**–**f** Sanger sequencing results of zevoCDA1-BE4max at 3 loci. The sequence and name of gRNAs are labelled above the Sanger results. Source data are provided as a Source Data file.

zebrafish, providing a clear opportunity to faithfully study the pathogenesis of this disease in the zebrafish context (Supplementary Fig. 4). To generate an OCA zebrafish model mimicking the human SLC24A5 (p.Q62*) mutation, we targeted the homologous site in zebrafish to install a c.220 C > T (p.Q74*) mutation (Fig. 2g). Notably, the target cytosine is located in a GC sequence context, meaning that any APOBEC1 deaminase-based CBEs including zAncBE4max would be poorly suited to edit this site. As anticipated, neither zAncBE4max nor SpRY-CBE4max exhibited activity at this site (Fig. 2h). Notably, zevoCDA1-SpRY-BE4max demonstrated an editing efficiency of 42.67% ± 10.69% at this site (Fig. 2h). Unsurprisingly, the pigmentation of homozygous juveniles of the F1 generation was significantly lighter, confirming the pathogenicity of this human SNV in the zebrafish context (Fig. 2i). These results demonstrate a powerful application of zevoCDA1-SpRY-BE4max as a flexible and accurate base editor that overcomes sequence context preference for disease modelling in zebrafish.

https://doi.org/10.1038/s41467-024-53735-y



Truncated zevoCDA1-198 narrows the target editing window and reduces the off-target effect

As shown above, zevoCDA1-SpRY-BE4max can efficiently target all NC sites independent of the PAM, making it possible to edit any cytosine in the zebrafish genome. However, the rather wide editing window of zevoCDA1-SpRY-BE4max (positions 1 to 9, at the PAM-distal end of the Cas9 target site) raises a different problem: in addition to the desired C-to-T base conversion, unwanted base conversions of other cytosines falling within this window may also occur. To improve editing precision, we engineered zevoCDA1-SpRY-BE4max using established strategies for narrowing the editing window[29]. First, we deleted the linker sequence between zevoCDA1 and SpRYCas9 to generate zevoCDA1-NL (Fig. 3a and Supplementary Data 1). We further removed the nuclear export signal (NES) sequence from zevoCDA1-NL to produce

Fig. 2 | Efficient cytosine base editing at non-canonical PAM sites by zevoCDA1-SpRY-BE4max. a Schematic representation of the mRNA construct for zevoCDA1-SpRY-BE4max. bpNLS: bipartite nuclear localization, zevoCDA1: cytosine deaminase, XTEN: a 32aa flexible linker, nSpRYCas9: SpCas9 nickase varient, linker: SGGSSGGS amino acid, UGI: Uracil glycosylase inhibitor. **b**–**e** Comparison of editing efficiencies between SpRY-CBE4max (**b**, **d**) and zevoCDA1-SpRY-BE4max (**c**, **e**) using 17 gRNAs targeting NRN PAMs (**b**, **c**) and NYN PAMs (**d**, **e**). The data represent the aggregate results of three independently replicated experiments, and the error bars indicate the standard deviation of the mean values. **f** Editing efficiency of zevoCDA1-SpRY-BE4max at 30 NNN PAM sites across 18 genes. **g** Schematic diagram of the *slc24a5* target locus. The targeted sequence is shown with the PAM highlighted in red. The targeted cytosine nucleotide and expected changes are highlighted in blue. **h** Sanger sequencing results comparing zAncBE4mx, SpRY-CBE4max and zevoCDA1-SpRY-BE4max at the *slc24a5* Q74* target locus. The red arrowhead indicates the expected nucleotide substitutions. **i** Lateral view of 3 dpf F1 homozygous embryos with the *slc24a5* Q74* mutation (bottom) showing pigmentation defects compared with wild-type (top). Scale bar: 500 μm. Three independent experiments were repeated with similar results. Source data are provided as a Source Data file.

zevoCDA1-198 (Fig. 3a and Supplementary Data 1). We selected ten gRNAs with different PAMs, each targeting sites containing multiple cytosines within the 1 to 9 editing window, to assess the cytosine base editing (CBE) activity of these variants at each site. Notably, the main editing windows of zevoCDA1-NL and zevoCDA1-198 have been narrowed to positions 1 to 7 and 1 to 5, respectively (Fig. 3b–e). Furthermore, the editing efficiencies of zevoCDA1-NL and zevoCDA1-198 for the most active cytosine within the primary editing window at each site are comparable to those of zevoCDA1-SpRY-BE4max.

Next, we applied these narrow-window variants to create a precise disease model in zebrafish. We selected Axenfeld–Rieger syndrome (ARS), a autosomal-dominant clinically heterogeneous disorder comprising anterior segment abnormalities of the eye, craniofacial and dental malformations, cardiovascular malformations, and additional periaqueductal skin[30]. The c.271 C > T/p.R91W mutation in PITX2 – a gene that is highly conserved between humans and zebrafish (Supplementary Fig. 5) – has been reported to cause ARS[31]. However, the precise C-to-T conversion at the corresponding site of *pitx2* necessary to generate an accurate disease model is challenging to install without editing other surrounding cytosines. We designed a gRNA targeting the corresponding site c.265 C/p.R89 of pitx2 editing the target site at C2 in the editing window with a potential bystander site at C6. Unlike zevoCDA1-SpRY-BE4max and zevoCDA1-NL, which converted both cytosines on C2 and C6, zevoCDA1-198 achieved specific editing of the C2 site, creating a zebrafish ARS disease model without significant C6 site editing (Fig. 3f, g, Table 1). Homozygous *pitx2* ^{R89W} mutant zebrafish exhibited underdeveloped anterior chambers (Fig. 3h, i) and craniofacial deformities at 5 days post-fertilization (dpf) (Fig. 3j, k), resembling the phenotypes characterized in human ARS disease. Together, these data demonstrate that zevoCDA1-198 is a valuable tool for precise, efficient, and PAM-flexible base editing without sequence context bias, opening up possibilities for human genetic disease modelling in zebrafish.

To assess the indel ratio and product purity of the tools, we PCR amplified the edited sequences of the three tools at the *pitx2* targeting site for NGS analysis. Additionally, we PCR amplified the sequences surrounding the three most probable off-target sites for the *pitx2* sgRNA based on CRISPOR prediction for NGS analysis to investigate the potential off-target effects of these three tools. We detected a relatively low level (3.69-8.43%) of indels and undesired C-to-A or C-to-G conversions at 0.05-0.23% for these three tools (Table 1). At the three most likely off-target sites, zevoCDA1-SpRY-BE4max and zevoCDA1-NL exhibited negligible editing at the first two sites, but showed 36.07% and 43.72% off-target editing at the third site, respectively (Table 1). In contrast, zevoCDA1-198 demonstrated significantly reduced off-target effects at the same sites (0.04%, 0.02%, and 6.04%, respectively) (Table 1). This positions zevoCDA1-198 as a valuable CBE, complementing the context sequence-biased mainstream CBEs in zebrafish.

We also show that the three variants of zevoCDA1-SpRY-BE4max developed here exhibit high germline targeting efficiency and germline transmission rate (Supplementary Table 3), supporting their strong ability to generate accurate base edits with high efficiency.

Discussion

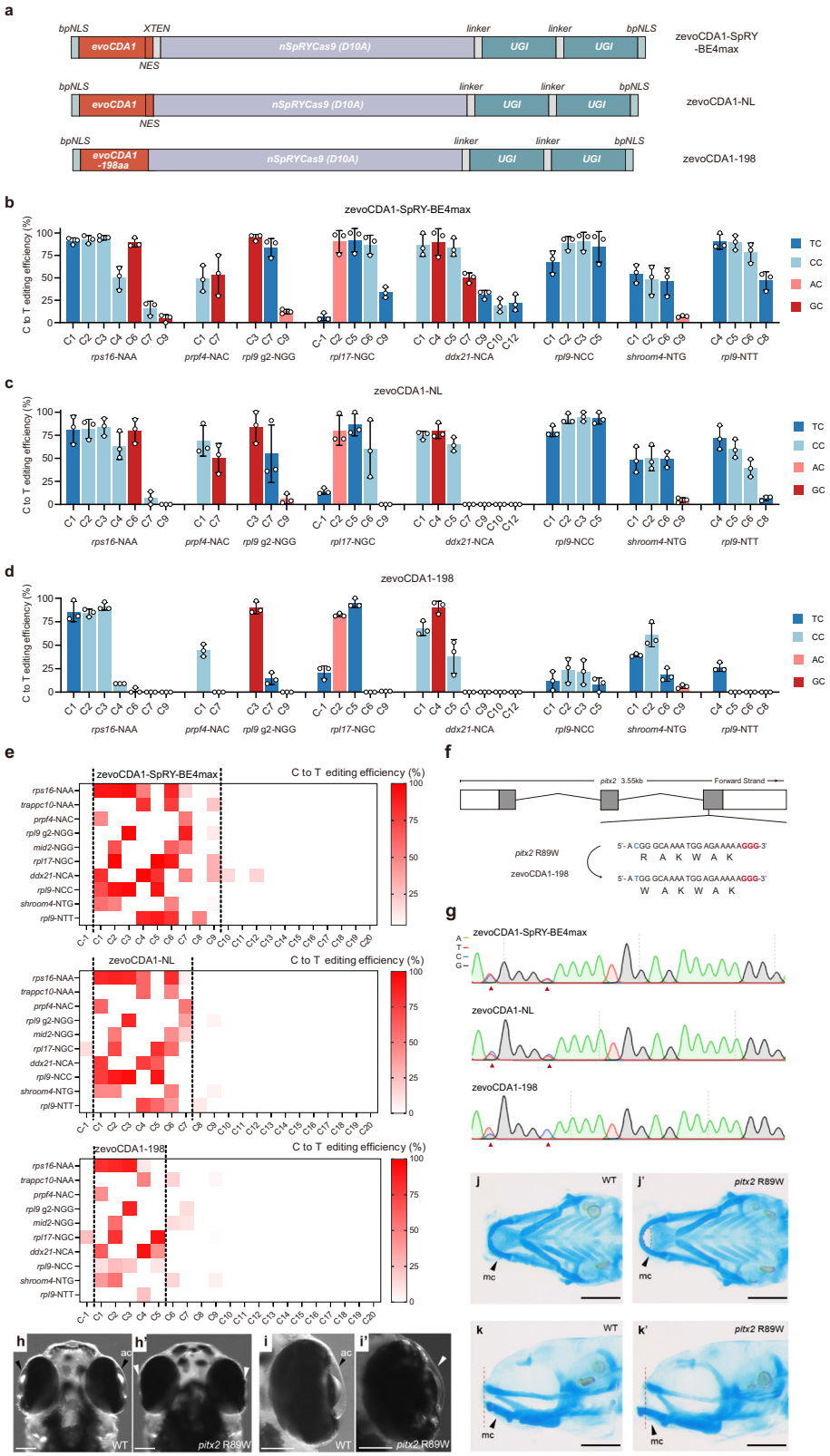
The application of base editing in zebrafish has greatly enhanced its potential to study disease pathogenesis caused by SNVs and screen potential drugs. However, major limitations of existing technologies, including disfavoured targeting of cytosines in GC and CC contexts, PAM restrictions, and high levels of indels and bystander edits have limited the application of these tools in zebrafish. Here, we engineered efficient and precise base editors for flexible targeting of any cytosine of interest in the zebrafish genome, using a combination of codon optimization, a PAM-less Cas9 variant, and narrowed base editing windows.

The base editing context bias of APOBEC1 deaminase-based CBEs (preference following the order TC, AC ≥ CC > GC in our hands) hinders the construction of models for certain mutation types and the application of base editing techniques in biomedical study. CDA1-BE3 and AID-BE3 can achieve higher GC editing than BE3 at certain sites due to their alternative deaminase domains. The evolved evoCDA1-BE4max is able to produce higher efficiency base edits at GC sites in mammalian cells, but suffers from higher indels (6.8-20%), wider editing window (from 1st to 13th) and higher off-target editing (0.3-40%)[19]. To the best of our knowledge, zevoCDA1-BE4max is the first CBE that breaks the GC context restriction in zebrafish and will enable precise base editing of many difficult sites for establishment of previously unavailable disease models.

To overcome the rather restrictive PAM preference of the SpCas9 moiety in most base editors, we updated zevoCDA1-BE4max to zevoCDA1-SpRY-BE4max which not only opens up all potential PAM sequences for efficient targeting, but also maintains very high editing activity and significantly reduces indel formation. Another group recently reported SpRY-CBE4max, which also produced fewer indels than AncBE4max in zebrafish[24], suggesting that the weaker cleavage activity of SpRYCas9 compared to traditional SpCas9 is beneficial for reducing unwanted indels. However, although the indel rate of zevoCDA1-SpRY-BE4max is much lower than that of zevoCDA1-BE4max (Supplementary Fig. 2), it is crucial to acknowledge that indels may still arise at certain loci at a significant frequency, emphasizing the importance of screening different guide sequences for favourable product purity. Importantly, the expansion of targeting sequence space afforded by the SpRYCas9 module would allow for multiple guide RNAs to be designed to target any desired cytosine in the genome.

It is noteworthy that similar to the preference observed with SpRY-CBE4max, zevoCDA1-SpRY-BE4max demonstrates generally higher editing efficiency for NRN PAM sites compared to NYN PAM sites. While our experiments indicate that zevoCDA1-SpRY-BE4max outperforms SpRY-CBE4max at the eight NYN PAM sites tested, it is important to recognize that these sites are all derived from ribosomal protein subunit genes. Therefore, we cannot rule out the possibility that the selection of these specific genes may introduce a bias in the observed editing efficiencies. The local chromatin environment, accessibility of the target site, and the context of adjacent sequences can vary significantly between different genes, potentially affecting editing outcomes. Thus, the relatively favourable performance of

https://doi.org/10.1038/s41467-024-53735-y



zevoCDA1-SpRY-BE4max on ribosomal protein subunit genes does not necessarily imply its broad applicability across diverse gene targets.

zevoCDA1-SpRY-BE4max has a wide editing window spanning positions 1 to 9 of the sgRNA target sequence – and when more than one cytosine appears in this window, non-target cytosines are also edited. This bystander effect often generates non-synonymous mutations, precluding the direct assessment of the consequences of a single human SNV in zebrafish disease models. Previous studies with other base editors have demonstrated that engineering the linkers between the deaminase domain and the Cas domain of CBEs can shorten the editing window to achieve high editing precision[29,32]. Inspired by these studies, we removed the linkers between the zevoCDA1 domain and the SpRYCas9 domain and further truncated the NES sequence at the zevoCDA1 C-terminus. By doing this, we successfully narrowed the

Fig. 3 | Precise cytosine base editing by zevoCDA1-NL and zevoCDA1-198. a The mRNA construct of zevoCDA1-NL and zevoCDA1-198 for precise cytosine base editing. The XTEN linker was deleted from zevoCDA1-SpRY-BE4max to generate zevoCDA1-NL, and the nuclear export signal (NES) located at the C-terminus of the evoCDA1 deaminase was subsequently removed to generate zevoCDA1-198. bpNLS: bipartite nuclear localization, zevoCDA1: cytosine deaminase, nSpRYCas9: SpCas9 nickase varient, linker: SGGSSGGS amino acid, UGI: Uracil glycosylase inhibitor. **b**–**d** Assessment of the editing efficiency and targeting window of zevoCDA1-SpRY-BE4max (**b**), zevoCDA1-NL (**c**) and zevoCDA1-198 (**d**) using 8 gRNAs targeting NNN PAMs. The data represent the sum of three independently replicated experiments, and the error bars represent the standard deviation of the mean values. **e** Summary of editing efficiency of zevoCDA1-SpRY-BE4max, zevoCDA1-NL and zevoCDA1-198 at 10 NNN PAM sites across 8 genes. The dotted range indicates the editing window. **f** Schematic diagram of the *pitx2* target locus. The target sequence is displayed with the PAM highlighted in red and the target nucleotide and expected nucleotide changes are highlighted in blue. **g** Comparison of Sanger sequencing results for zevoCDA1-SpRY-BE4max, zevoCDA1-NL, and zevoCDA1-198 at the *pitx2* R89W target locus. The red arrowhead indicates the expected nucleotide substitutions. **h**–**i′** Dorsal view of 5 dpf F1 homozygous embryos with the *pitx2* R89W mutation showing absence of anterior chamber (ac). WT anterior chambers are highlighted with black arrows in the close-up view of the head (**h**) and with a dashed outline and black arrow in the close-up view of the eye (**h′**). The absence of the anterior chamber in mutant zebrafish is highlighted by a white arrow in the close-up view of the head (**i**) and eye (**i′**). Scale bars: 100 μm. **j**–**k′** Alcian blue staining of 5-dpf wild-type and *pitx2* R89W mutant embryos. The ventral view (**j**, **j′**) and lateral view of (**k**, **k′**) wild-type AB zebrafish (**j**, **k**) and *pitx2* R89W (**j′**, **k′**) embryos at 5 dpf. *pitx2* R89W mutant (**j′**, **k′**) shows severe structural malformations in Meckel's cartilage (mc) highlighted with arrow and red dotted line compared with WT AB (**j**, **k**). Scale bars: 200 μm. Three independent experiments were repeated with similar results. Source data are provided as a Source Data file.

Table 1 | Assessment of on-target and off-target editing by zevoCDA1-SpRY-BE4max, zevoCDA1-NL and zevoCDA1-198 using *pitx2* R89W sgRNA and NGS

gRNA	Mismatch location	Sequence	MitOfftarget Score	Tools	Efficiency	Indel	Incorrect editing
pitx2 R89W	.....	ACGGGCAAAATGGAGAAAAA**GGG**	-	**zevoCDA1-SpRY-BE4max**	C2T, 38.26% C6T, 35.47%	8.43%	0.05%
	.*..*.....	AAGGTCAAAATGGAGAAAAA**ATT**	5.7229		C6T, 0.10%	1.01%	0.02%
	.*.*.....	AAGAGCAAAATGGAGAAAAA**CAT**	5.4598		C6T, 0.01%	0.09%	0.02%
	.*.....*...	AAGGGCAAAATGGAGAGAAA**AGT**	5.2250		C6T, 36.07%	2.35%	2.65%
	.....	ACGGGCAAAATGGAGAAAAA**GGG**	-	**zevoCDA1-NL**	C2T, 52.53% C6T, 35.99%	4.63%	0.23%
	.*..*.....	AAGGTCAAAATGGAGAAAAA**ATT**	5.7229		C6T, 0.10%	0.05%	0.01%
	.*.*.....	AAGAGCAAAATGGAGAAAAA**CAT**	5.4598		C6T, 0.02%	0.05%	0.00%
	.*.....*...	AAGGGCAAAATGGAGAGAAA**AGT**	5.2250		C6T, 43.72%	3.00%	2.80%
	.....	ACGGGCAAAATGGAGAAAAA**GGG**	-	**zevoCDA1-198**	C2T, 51.69% C6T, 4.73%	3.69%	0.23%
	.*..*.....	AAGGTCAAAATGGAGAAAAA**ATT**	5.7229		C6T, 0.04%	0.31%	0.00%
	.*.*.....	AAGAGCAAAATGGAGAAAAA**CAT**	5.4598		C6T, 0.02%	0.00%	0.00%
	.*.....*...	AAGGGCAAAATGGAGAGAAA**AGT**	5.2250		C6T, 6.04%	0.02%	0.03%

PAM sequences and editing tools are highlighted in bold.

editing window to positions 1 to 5, greatly improving the editing precision. It is worth noting that at certain loci, the precision of zevoCDA1-198 may be accompanied by a reduction in editing efficiency (Fig. 3b–d). Taking advantage of the narrow window of zevoCDA1-198, we created the ARS disease model mimicking a disease-causing SNV in PITX2 without any detectable bystander effects and with minimal off-targets (Table 1). Considering the extensive range of measured off-target rates for evoCDA1-BE4max (ranging from 0.3% to >40% at certain off-target sites) in mammalian cells[19], it is unsurprising to witness off-target effects of 36.07% and 43.72% for zevoCDA1-SpRY-BE4max and zevoCDA1-NL at a specific site. While the off-target editing activity rate is likely closely tied to the sgRNA sequence itself, zevoCDA1-198 demonstrated significantly lower effects at the same sgRNA off-target site compared to the other variants. This leads us to believe that zevoCDA1-198, with a more precise editing window, will also exhibit lower off-target rates for other sgRNAs. Although potential off-target effects with other sgRNAs may persist with zevoCDA1-198, any erroneous phenotypes that arise in zebrafish disease models due to off-target effects can be corrected through outcrossing.

In conclusion, our work provides a set of CBE editors that overcome the major limitations of current base editing tools, including overcoming sequence context bias and PAM limitations, and reducing all forms of unwanted editing, and it is likely to work in other organisms besides zebrafish as well. Together, these editors provide a powerful toolbox to enrich the capabilities of zebrafish to model human genetic diseases.

Methods

Ethical Statement

All animal experiments complied with the relevant regulations and were approved by the University Animal Care and Use Committee of South China Normal University(SCNU-BRR-2021-021), and as per protocol 20-07 approved by the Institutional Animal Care Committee (IACUC) of Oklahoma Medical Research Foundation, Oklahoma City, USA.

Zebrafish maintenance

Wild-type zebrafish strain AB eggs were incubated at a temperature of 28.5 °C. Pairs were randomly selected from the AB male and female fish lines at an aquaculture density of 30 fish per 3-liter tank, with an age range of 6-15 months.

Plasmid construction and mRNA generation

To construct the pT3TS-evoCDA1-BE4max plasmid, evoCDA1-SpCas9-N synthesized by Tsingke Biotechnology and SpCas9-C-2X UGI synthesized by GenScript were integrated to the pT3TS vector derived from the pT3TS-AncBE4max-nCas9 plasmid (a gift from Professor Rongjia Zhou). To construct the pT3TS-zevoCDA1-BE4max plasmid, the zebrafish codon-optimized evoCDA1 cytidine deaminase gene segment, synthesized by GenScript, replaced the cytidine deaminase ancAPOBEC1 in the pT3TS-AncBE4max-nCas9 plasmid. The plasmid pT3TS-zevoCDA1-SpRY-BE4max was derived from the pT3TS-zevoCDA1-BE4max plasmid by substituting the SpCas9 codon

sequence with a SpRYCas9 DNA fragment. The plasmid pT3TS-zevoCDA1- SpRY- BE4max-NL was built by removing the linker between evoCDA1 and SpRYCas9 from the pT3TS-zevoCDA1-SpRY-BE4max plasmid. The pT3TS-zevoCDA1-SpRY-BE4max-198 was constructed by removing 30 nucleotides at the 3' end of the evoCDA1 cytidine deaminase (nuclear export signal region) from the pT3TS-zevoCDA1-SpRY-BE4max-NL plasmid. The primers for plasmid construction are listed in Supplementary Data 2.

The above cloning steps were performed using the Vazyme High-Fidelity DNA Polymerase 2x Phanta Max Master Mix for PCR amplification. Additionally, the Vazyme Mut Express II Fast Mutagenesis Kit V2 was used for infusion cloning to introduce desired mutations. Sufficient plasmid clones were obtained through transformation into Trans10 Chemically Competent Cells. The XbaI restriction enzyme was utilized to linearize the plasmid, and subsequently, the T3 mMESSAGE mMACHINE kit from Ambion was employed for in vitro transcription. Ultimately, purification was carried out employing the RNA Clean Kit provided by TianGen Company.

gRNA generation

Every gRNA was chemically synthesized by GenScript, with MS modifications present at both ends. These synthesized gRNAs were dissolved in a stock solution at a concentration of 1000 ng/μl and stored at −80 °C. The target sequences can be found in Supplementary Data 3.

Microinjection of CBE mRNA and gRNA and image acquisition in zebrafish

During the one-cell stage, zebrafish embryos were injected with 2 nl of a solution containing 400 ng/μl CBE mRNA and 200 ng/μl gRNA. At 3 dpf or 5 dpf, the embryos were anesthetized using 0.03% Tricaine (Sigma-Aldrich) and carefully mounted in 4% methylcellulose. Imaging was conducted using either an XM10 digital camera (OLYMPUS) or AxioCam MRc5 digital camera (Leica) on SZX2-FOF microscope (OLYMPUS). Post-capture adjustments and enhancements were made using Adobe Illustrator software.

Base editing analysis

For the base editing experiments, we acquired three pools of embryos, each pool containing 6 embryos that were randomly selected. Alkaline lysis was performed to extract the genomic DNA for PCR amplification with primers approximately 100 bp upstream and downstream of each sgRNA site. After Sanger sequencing the PCR products, the data was analyzed using the EditR (1.0.10) program[33]. The primers for PCR amplification and Sanger sequencing are listed in Supplementary Data 4.

Next-generation sequencing (NGS) and analysis

Genomic DNA extraction from both wild-type and injected embryos was conducted following standard protocols[34]. To construct the NGS library, we PCR amplified the regions of genomic DNA at targeted on/off-target sites, covering sequences ranging from 50 to 280 bp in length. 4-6 amplified products were then combined to generate the sequencing samples. For sequencing, an Illumina MiSeq instrument was utilized with PE150 sequencing mode by a commercial sequencing service from Biomarker Technologies. The resulting sequencing data was subjected to CRISPResso2 analysis[35] to determine the efficiency of genome editing and assess formation of indels. NGS data, along with gRNA sequence and the amplicon sequence, were analyzed using the CRISPResso2 local program or online platform to generate the "Alleles_frequency_table_around_sgRNA". The analysis results present Aligned Sequences (actual sequencing results), alongside the Reference Sequence. A "-" in the Aligned Sequence denotes a deleted base, while a "-" in the Reference Sequence indicates an inserted base within the Aligned Sequence. The count and proportion of indel reads were determined by summing instances where "-" appeared in either sequence. After excluding sequences with indels, the count and proportion of various single-base edits at the target site were calculated. C-to-T alterations at the target site were considered effective edits, whereas C-to-G/A modifications were classified as erroneous edits. The primers used for NGS are listed in Supplementary Data 5.

Alcian blue cartilage staining

Zebrafish embryos were collected at 5 dpf and fixed overnight in 4% paraformaldehyde in 1× diethyl pyrocarbonate (DEPC)-phosphate-buffered saline (PBS) at 4 °C. Embryos were stained overnight in 0.15% Alcian Blue solution comprised of Alcian Blue (Shanghai Sangon) dissolved in 75% acidic ethanol. Stained embryos were washed thoroughly with PBS, digested in 0.25% trypsin overnight at 37 °C and bleached in 1 ml 3% hydrogen peroxide supplemented with 50 μl of 2 M KOH twice on a rotating platform. Progressively dehydrated with ethanol, and stored in 70% glycerol at 4 °C. Craniofacial structures were identified as presented by Hendee et al. [31]. Images were obtained on an AxioCam MRc5 digital camera (Leica).

Off-target analysis

For each gRNA, we utilized CRISPOR (Version 4.99) to predict off-target sites[36]. Based on the specificity scores calculated by CRISPOR (Version 4.99) (Supplementary Data 6), we selected the top three sites with the highest scores as the most likely off-target sites, and evaluated their effects through NGS analysis.

Statistics and reproducibility

The sample size was not predetermined by any statistical method and acquisition of samples was random. No data were excluded from the analyses. The experiments were repeated independently three times to ensure robustness. Statistical analysis was conducted using GraphPad Prism 9 software. The results are presented as the mean value ± standard deviation (SD). To assess significant differences between different groups, a two-tailed unpaired t-test was performed, with a significance level set at P value < 0.05. The significance levels are denoted by *, **, ***, and ****, representing P values less than 0.05, 0.01, 0.001, and 0.0001, respectively. P values for all figures are listed in Supplementary Data 7.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

NGS data are available on the National Centre for Biotechnology Information Sequencing Read Archive (SRA) database under project numbers PRJNA1149283 and PRJNA1151843. All data supporting the findings of this study are available within the article and Supplementary Information files. Source data are provided with this paper.

References

1. Genomes Project, C. et al. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
2. Bamshad, M. J. et al. Exome sequencing as a tool for Mendelian disease gene discovery. *Nat. Rev. Genet* **12**, 745–755 (2011).
3. Anzalone, A. V., Koblan, L. W. & Liu, D. R. Genome editing with CRISPR-Cas nucleases, base editors, transposases and prime editors. *Nat. Biotechnol.* **38**, 824–844 (2020).
4. Zu, Y. et al. TALEN-mediated precise genome modification by homologous recombination in zebrafish. *Nat. Methods* **10**, 329–331 (2013).
5. Koblan, L. W. et al. Improving cytidine and adenine base editors by expression optimization and ancestral reconstruction. *Nat. Biotechnol.* **36**, 843–846 (2018).

6. Komor, A. C. et al. Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity. *Sci. Adv.* **3**, eaao4774 (2017).
7. Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A. & Liu, D. R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420–424 (2016).
8. Gaudelli, N. M. et al. Programmable base editing of A*T to G*C in genomic DNA without DNA cleavage. *Nature* **551**, 464–471 (2017).
9. Zhang, Y. H. et al. Programmable base editing of zebrafish genome using a modified CRISPR-Cas9 system. *Nat. Commun.* **8**, 118 (2017).
10. Hua, K., Tao, X., Yuan, F., Wang, D. & Zhu, J. K. Precise A.T to G.C Base Editing in the Rice Genome. *Mol. Plant* **11**, 627–630 (2018).
11. Qin, W. et al. Precise A*T to G*C base editing in the zebrafish genome. *BMC Biol.* **16**, 139 (2018).
12. Zong, Y. et al. Efficient C-to-T base editing in plants using a fusion of nCas9 and human APOBEC3A. *Nat. Biotechnol.* **36**, 950–953 (2018).
13. Cornean, A. et al. Precise in vivo functional analysis of DNA variants with base editing using ACEofBASEs target prediction. *Elife* **11**, e72124 (2022).
14. Zhao, Y., Shang, D., Ying, R., Cheng, H. & Zhou, R. An optimized base editor with efficient C-to-T base editing in zebrafish. *BMC Biol.* **18**, 190 (2020).
15. Carrington, B., Weinstein, R. N. & Sood, R. BE4max and AncBE4max are efficient in germline conversion of C:G to T:A base pairs in zebrafish. *Cells* **9**, 1690 (2020).
16. Nishida, K. et al. Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science* **353**, aaf8729 (2016).
17. Kohli, R. M. et al. Local sequence targeting in the AID/APOBEC family differentially impacts retroviral restriction and antibody diversification. *J. Biol. Chem.* **285**, 40956–40964 (2010).
18. Lu, X. et al. Optimized Target-AID system efficiently induces single base changes in zebrafish. *J. Genet Genomics* **45**, 215–217 (2018).
19. Thuronyi, B. W. et al. Continuous evolution of base editors with expanded target compatibility and improved activity. *Nat. Biotechnol.* **37**, 1070–1079 (2019).
20. Doll, R. M., Boutros, M. & Port, F. A temperature-tolerant CRISPR base editor mediates highly efficient and precise gene editing in Drosophila. *Sci. Adv.* **9**, eadj1568 (2023).
21. Walton, R. T., Christie, K. A., Whittaker, M. N. & Kleinstiver, B. P. Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants. *Science* **368**, 290–296 (2020).
22. Liang, F. et al. SpG and SpRY variants expand the CRISPR toolbox for genome editing in zebrafish. *Nat. Commun.* **13**, 3421 (2022).
23. Vicencio, J. et al. Genome editing in animals with minimal PAM CRISPR-Cas9 enzymes. *Nat. Commun.* **13**, 2601 (2022).
24. Rosello, M. et al. Disease modeling by efficient genome editing using a near PAM-less base editor in vivo. *Nat. Commun.* **13**, 3435 (2022).
25. Xu, Z. et al. SpRY greatly expands the genome editing scope in rice with highly flexible PAM recognition. *Genome Biol.* **22**, 6 (2021).
26. Ren, Q. et al. PAM-less plant genome editing using a CRISPR-SpRY toolbox. *Nat. Plants* **7**, 25–33 (2021).
27. Li, J. et al. Genome editing mediated by SpCas9 variants with broad non-canonical PAM compatibility in plants. *Mol. Plant* **14**, 352–360 (2021).
28. Lasseaux, E. et al. Molecular characterization of a series of 990 index patients with albinism. *Pigment Cell Melanoma Res* **31**, 466–474 (2018).
29. Tan, J., Zhang, F., Karcher, D. & Bock, R. Engineering of high-precision base editors for site-specific single nucleotide replacement. *Nat. Commun.* **10**, 439 (2019).
30. Tümer, Z. & Bach-Holm, D. Axenfeld-Rieger syndrome and spectrum of and mutations. *Eur. J. Hum. Genet.* **17**, 1527–1539 (2009).
31. Hendee, K. E. et al. PITX2 deficiency and associated human disease: insights from the zebrafish model. *Hum. Mol. Genet* **27**, 1675–1695 (2018).
32. Tan, J. J., Zhang, F., Karcher, D. & Bock, R. Expanding the genome-targeting scope and the site selectivity of high-precision base editors. *Nat. Commun.* **11**, 629 (2020).
33. Kluesner, M. G. et al. EditR: A method to quantify base editing from sanger sequencing. *Crispr J.* **1**, 239–250 (2018).
34. Zheng, S. H. et al. Efficient PAM-less base editing for zebrafish modeling of human genetic disease with zSpRY-ABE8e. *Jove-J. Visual. Exp.* e64977 (2023).
35. Clement, K. et al. CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nat. Biotechnol.* **37**, 224–226 (2019).
36. Concordet, J. P. H.M. CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. *Nucleic Acids Res.* **46**, 242–W245 (2018).

Acknowledgements

We thank Lin Li, and Guifang Yang for zebrafish husbandry and thank Dr. Fang Liang for technical support. We are grateful to Professor Rongjia Zhou (Wuhan University) for sharing zAncBE4max with us. This work was supported by the National Key R&D Program of China 2021YFA0805000 (JF.F.), 2023YFA1800600 (JF.F.); the National Natural Science Foundation of China 32070819 (Yanmei L.), 92268114 (JF.F.), 31970782 (JF.F.); the High-level Hospital Construction Project of Guangdong Provincial People's Hospital DFJHBF202103 (JF.F.), KJ012021012 (JF.F.) and Presbyterian Health Foundation, Oklahoma City, OK, USA (G.K.V.).

Author contributions

Yanmei L., JF.F. and W.Q. conceived the project and designed the experiments. Yu Z., Yang L. and S.Z. did most of the experiments and analysed the data. W.Q., J.X., X.X., X.Y., J.Z., Y.S., Yan Z., and H.M. contributed to the experimental works and analysis. Yanmei L., JF.F. and G.K.V. were responsible for the funding acquisition. Yanmei L., JF.F. and G.K.V wrote the manuscript and supervised the work. All authors read and approved the manuscript.

Competing interests

Yanmei L., JF.F., Yu Z. and S.Z. have been granted a patent (Patent No. ZL 2023 1 0872424.6) in China pertaining to the development and application of zevoCDA1-SpRY-BE4max, zevoCDA1-NL, and zevoCDA1-198. The remaining authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-024-53735-y.

Correspondence and requests for materials should be addressed to Gaurav K. Varshney, Ji-Feng Fei or Yanmei Liu.

Peer review information *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. A peer review file is available.