An engineered xCas12i with high activity, high specificity and broad PAM range

Hainan Zhang[1, 4], Xiangfeng Kong[1, 4], Mingxing Xue[1, 4], Zikang Wang[1, 4], Yinghui Wei[1, 4], Haoqiang Wang[1, 4], Jingxing Zhou[1], Weihong Zhang[1], Mengqiu Xu[1], Xiaowen Shen[1], Jinhui Li[1], Jing Hu[1], Na Zhong[1], Yingsi Zhou[1#], Hui Yang[1, 2, 3#]

[1] HUIEDIT Therapeutics Inc., Shanghai 200031, China.

[2] HUIGENE Therapeutics Inc., Shanghai 200131, China.

[3] Institute of Neuroscience, State Key Laboratory of Neuroscience, Key Laboratory of Primate Neurobiology, Center for Excellence in Brain Science and Intelligence Technology, Chinese Academy of Sciences, Shanghai 200031, China.

[4] These authors contributed equally to this work.

Correspondence to: ys@huiedit.com (Y.Z.), huiyang@huidagene.com (H.Y.)

Abstract

The type-V CRISPR effector Cas12i, with its smaller size, short crRNA guiding, and self-processing features, is a potentially versatile genome editing tool. By screening Cas12i proteins from a metagenomic database, we identified a natural variant with high activity in mammalian cells, named as xCas12i. We further engineered the PAM-interacting, REC, and RuvC domains for enhanced cleavage activity and specificity. This variant, named as high-fidelity Cas12Max, exhibited robust genome editing activity and minimal off-target activity with a broad 5'-TN recognition profile. With the fusion of deaminase TadA8e and further optimization of xCas12i, the base editor dCas12i-Tad8e also showed the high editing efficiency. This study provides highly efficient and specific tools for gene therapy.

Keywords: CRISPR, Cas12i, broad PAM recognition, dCas12i base editor

Introduction

The clustered regularly interspaced short palindromic repeats-Cas (CRISPR-Cas) systems, including type II Cas9 and type V Cas12 systems, which serve in the adaptive immunity of prokaryotes against viruses, have been developed into genome editing tools[1-3]. Compared with type II systems, the type V systems including V-A to V-K showed more functional diversity[4, 5]. Amongst them, Cas12i has a relatively smaller size (1033-1093 aa), compared to SpCas9 and Cas12a, and has a 5'-TTN protospacer adjacent motif (PAM) preference[4, 6, 7]. Cas12i is characterized by the capability of autonomously processing precursor crRNA (pre-crRNA) to form short mature crRNA. Cas12i mediates cleavage of dsDNA with a single RuvC domain, by preferentially nicking the non-target strand and then cutting the target strand[8-10]. These intrinsic features of Cas12i enable multiplex high-fidelity genome editing. However, the natural variants of Cas12i (Cas12i1 and Cas12i2), showed low editing efficiency which limits their utility for therapeutic gene editing.

To address these limitations, we screened ten natural Cas12i variants and found one, xCas12i, with robust high activity in HEK293T cells. Engineering of xCas12i by arginine substitutions at the PAM-interacting (PI), REC and RuvC domains led to the production of a variant, high-fidelity Cas12Max (hfCas12Max), with significantly elevated editing activity and minimal off-target cleavage efficiency. In addition, we assessed the base editing efficiency of xCas12i-based base editor, and thus expanded the genome-editing toolbox.

Results

Identification and characterization of type V-I systems xCas12i

In order to identify more Cas12i variants, we developed and employed a bioinformatics pipeline to annotate Cas12i proteins, CRISPR arrays and predicted PAM preferences, and found 10 new CRISPR/Cas12i systems. To evaluate the activity of these Cas12i variants in mammalian cells, we designed a fluorescent reporter system which detected the increased enhanced green fluorescent protein (EGFP) signal intensity activated by Cas-mediated dsDNA cleavage or double strand breaks (Fig.S1a). This system relied on

56 the co-transfection of a plasmid coding for mCherry, a nuclear localization signal (NLS)
57 -tagged Cas protein and its guide RNA (gRNA) or crRNA, and one coding BFP and
58 activatable EGxxFP cassette, which is EGxx-target site-xxFP[11]. EGFP activation was
59 carried out by Cas mediated DSB and single-strand annealing (SSA)-mediated repair.
60 Using this system, we observed that a variant, xCas12i, with targeted crRNA induced
61 significant activation of EGFP expression (Fig. 1a, Fig. S1b), and exhibited a higher
62 editing frequency than LbCas12a or SpCas9 as determined by Fluorescence Activated
63 Cell Sorter (FACS) analysis (Fig. 1a). The xCas12i variant was smaller in size compared
64 to SpCas9 and LbCas12a (Fig. S2a). We explored the effects of spacer length on
65 cleavage efficiency in xCas12i, and found that 17-22 nt were optimal length for their
66 activation (Fig. S2b). Considering the 5'-TTN PAM preference of Cas12i, we performed a
67 NTTN PAM identification assay using the reporter system. We found that xCas12i showed
68 a consistent high frequency of EGFP activation at target sites with 5'-NTTN PAM
69 sequences, while LbCas12a had comparable activity at 5'-TTTN PAM, respectively (Fig.
70 S2c).
71 To further confirm the dsDNA cleavage activity of xCas12i in mammalian cells, we
72 transfected an all-in-one plasmid encoding NLS tagged xCas12i with crRNAs targeting 37
73 sites from *TTR*[12], *PCSK9*[13] in HEK293T or *Ttr* in N2a cells. The editing efficiency, *i.e.* indel
74 (insertion and deletion) formation at these loci was measured 48 hours after transfection
75 using FACS and targeted deep sequencing (Fig. S2d). We found that xCas12i mediated a
76 high frequency, up to 90%, of indel formation at most sites from *Ttr*, *TTR* and *PCSK9*, with
77 a mean indel formation rate of over 50% (Fig. S2e-f). These data indicate that xCas12i
78 exhibits a robust genome editing efficiency in mammalian cells, suggesting it has
79 excellent potential for therapeutic genome editing applications.
80
81 **Engineered xCas12i mediates high-efficiency editing at 5'-TN PAM sites**
82 To enhance its activity and expand its scope of PAM site recognition, we sought to
83 engineer xCas12i protein via mutagenesis and screen for variants with higher efficiency
84 and broader PAM using a reporter system, similar to what is described above. Substitution
85 of an amino acid in the DNA-binding pocket with positively charged arginine (R) was

86 shown to enhance the activity of the type V system[14-16]. Combined with predictive
87 structural analysis of xCas12i, we performed an arginine scanning mutagenesis approach
88 in the PI, REC-I and RuvC-II domains, generating a library of over 500 mutants (Fig. 1b,
89 Fig. S3a). We then individually transfected these mutant variants with an activatable
90 EGFP reporter system in HEK293T cells and analyzed them by FACS (Fig. 1b). Based on
91 the fluorescence intensity of cells with activated EGFP, over 100 mutants showed an
92 increased frequency of activated cells relative to wild type (WT) xCas12i, and one mutant,
93 named as Cas12Max, containing N243R showed a 3.4-fold improvement (Fig. S3a). We
94 next targeted *DMD* or *Ttr* sites using the fluorescent reporter system, and found that
95 Cas12Max displayed a markedly increased frequency of EGFP activation, relative to WT
96 xCas12i (Fig. 1c, Fig. S4a-b). To further test the efficacy of Cas12Max in targeting
97 genomic loci, we designed a total of eight gRNAs to target sites *TTR* and *PCSK9* in
98 HEK293T cells and three more targeting *Ttr* in N2a cells. Consistent with our previous
99 result, Cas12Max exhibited a significantly increased frequency of indels compared to WT
100 xCas12i (Fig. 1d-e).
101 Additionally, to investigate Cas12Max's PAM preference, we performed a 5'-NNN
102 PAM recognition assay by designing reporter plasmids with the same target sequence but
103 different PAM. Besides showing a consistent or higher cleavage activity at sites with a
104 5'-TTN PAM, Cas12Max showed a similarly high cleavage activity for targets with TNN,
105 ATN, GTN and CTN PAM sites, compared with the commonly used Cas12[7, 17] (LbCas12a,
106 Ultra-AsCas12a) (Fig. 1f). Taken together, these results demonstrate that Cas12Max
107 exhibits high-efficiency editing activity with highly flexible 5'-TN or 5'-TNN PAM
108 recognition.
109
110 **hfCas12Max mediates high-efficient and specific genome editing**
111 To examine the specificity of Cas12Max, we transfected a construct designed to express it
112 with crRNA targeting *TTR*[12], and performed indel frequency analysis of on- and off-target
113 (OT) sites predicted by Cas-OFFinder[18]. Using reporter system or targeted deep
114 sequence analysis, we found that Cas12Max efficiently edited target sites and resulted in
115 significant indel formation at 2 of 3 predicted off-target sites (Fig. S5). To eliminate the

116 off-target activity of Cas12Max, we screened these mutants with mutations in the REC
117 and RuvC domains[19], which have undiminished on-target cleavage activity, for those with
118 no off-target activity, using two activatable reporter systems, each containing one OT site
119 (Fig. 1b). We found that four mutants (v4.1-V880R, v4.2-M923R, v4.3-D892R and
120 v4.4-G883R) maintained a high level of on-target editing activity and showed significantly
121 reduced off-target EGFP activation (Fig. S6a). We further combined these four amino acid
122 substitutions with N243R and/or E336R of Cas12Max and found that the variant v6.3
123 (N243R/D892R/G883R) showed the lowest off-target EGFP activation at OT.1 and OT.2
124 sites and high on-target at the ON.1 site (Fig. 1g, Fig. S6b-c). Targeted TIDE analysis of
125 endogenous TTR.2 site and its off-target sites in HEK293T showed that v6.3
126 (N243R/D892R/G883R) significantly reduced off-target indel frequencies at six OT sites
127 and retained on-target at ON site, compared to Cas12Max (Fig. 1h). In addition, relative to
128 Cas12Max (v1.1), v6.3 (N243R/D892R/G883R) retained comparable or even higher
129 on-target activity at DMD.1, DMD.2 and DMD.3 sties (Fig. S6d). Therefore, we named
130 v6.3 as high-fidelity Cas12Max (hfCas12Max).

131 To comprehensively evaluate the performance of hfCas12Max in human cells, we
132 designed large number of target sites in the exons of *TTR* for various Cas nucleases. In
133 total, editing activity was monitored over 30 sites for hfCas12Max with TTN PAMs, 40 sites
134 for SpCas9 with NGG PAMs, 4 sites for LbCas12a with TTTN PAMs, 4 sites for Ultra
135 AsCas12a, and 7 sites for KKH-saCas9 with NNNRRT PAMs. Indel analysis showed that
136 hfCas12Max exhibited an average efficiency of 70%, higher activity than other Cas
137 nucleases, comparable activity with Cas12Max (Fig. 1i, Fig. S7a). Using 5'-NNN PAM
138 reporter system, hfCas12Max maintained the broad 5'-TN and 5'-TNN PAM recognition
139 profile similarly to Cas12Max (Fig. 1f). To further evaluate the specificity of hfCas12Max in
140 human cells, we determined indel frequencies of P2RX5 and NLRC4 on-target and their
141 corresponding *in silico* predicted off-target sites[20]. TIDE analysis showed that
142 hfCas12Max had a higher on-target editing efficiency and similarly almost no indel activity
143 at potential off target sites, compared to Ultra AsCas12a and LbCas12a (Fig. S8a-b).
144 Overall, these results demonstrate that hfCas12Max has high efficiency and specificity
145 and is superior to SpCas9 and other Cas12 nucleases.

146

147 **dxCas12i-based base editors efficiently mediate base editing in mammalian cells**

148 We further explored the base editing of xCas12i by generating a nuclease-deactivation
149 xCas12i (dxCas12i). This was done by first introducing single mutations (D650A, D700A,
150 E875A, or D1049A) in the conserved active site of xCas12i based on alignment to
151 Cas12i1[8] and Cas12i2[10] (Fig. S9a-b). Then, these variants were fused with TadA8e^{V106W}
152 or human APOBEC3A (hA3A^{W104A}) to form the dxCas12i base editors TadA8e.1-dxCas12i
153 and hA3A.1-dxCas12i, respectively[21-23]. The initial versions of TadA8e.1-dxCas12i and
154 hA3A.1-dxCas12i showed low base editing activity with frequencies of 8% A-to-G and 2%
155 C-to-T, respectively (Fig. 2a-d). To address this, we introduced single and combined
156 mutations for high cleavage activity into the PI and Rec domains of dxCas12i, which
157 resulted in significantly increased A-to-G editing activity. Among the improved variants,
158 dxCas12i-TadA8e-v2.2 (N243R/E336R) achieved 50% activity at A9 and A11 sites of the
159 *KLF4* locus, markedly higher than the 30% activity of dLbCas12a-TadA8e (Fig. 2b, Fig.
160 S10a-b). At target sites within *PCSK9*, TadA8e-dxCas12i-v2.2 showed a similarly
161 increased efficiency to mediate A-to-G transitions, higher than dLbCas12a-TadA8e (Fig.
162 S11). We then further engineered the NLS, linker, TadA8e protein and the N-terminus to
163 produce dxCas12i-TadA8e-v4.3 which exhibited a nearly 80% A-to-G editing efficiency,
164 while the editing activities of other dxCas12i-ABE versions were unchanged (Fig. 2c, Fig.
165 S10c-d). In addition, the substitution of dxCas12i-v1.2 (N243R) or dxCas12i-v2.2
166 (N243R/E336R) showed consistently elevated C-to-T editing at C7 and C10 sites of
167 *DYRK1A*, even higher at C7 by dLbCas12a-hA3A (Fig. 2d). These results together
168 demonstrate that engineered dxCas12i-based editors exhibit the high base editing activity
169 in mammalian cells.

170

171 **Discussion**

172 In this study, we demonstrate that the Type V-I Cas12i system enables versatile and
173 efficient genome editing in mammalian cells. We found a natural Cas12i variant, xCas12i,
174 that shows high editing efficiency at TTN-PAM sites. By semi-rational design and protein

175 engineering of its PI, REC, RuvC domains, we obtained a high-efficiency, high-fidelity
176 variant, hfCas12Max, which contains N243R, E336R, and D892R substitutions. In
177 agreement with the hypothesis that introducing arginine at key sites could strengthen the
178 binding between Cas and DNA, the introduction of N243R in the PI domain and E336R at
179 REC domain significantly increased editing activity and expanded PAM recognition.
180 Interestingly, D892R or G883R substitutions in the RuvC domain reduced off-target and
181 retained on-target cleavage activity, whereas alanine substitutions[24, 25], which has been
182 used to reduce off-target activity, did not (Fig. S6c). The D892R substituted hfCas12Max
183 was obviously more sensitive to mismatch, which suggests that D892R or G883R
184 improved sgRNA binding specificity. Our data suggests that a semi-rational engineering
185 strategy with arginine substitutions based on the EGFP-activated reporter system could
186 be used as a general approach to improve the activity of CRISPR editing tools.

187 Through engineering, our Cas12i system has achieved high editing activity, high
188 specificity and a broad PAM range, comparable to SpCas9, and better than other Cas12
189 systems. Given its smaller size, short crRNA guide, and self-processing features[4, 8, 10], the
190 Type V-I Cas12i system is suitable for *in vivo* multiplexed gene editing applications,
191 including AAV[26] or LNP[12, 13].

192 In addition, we have confirmed that the Type V-I Cas12i system can be used in base
193 editing applications. For base editor, the dCas12i system shows high A-to-G editing at
194 A9-A11 sites even A19 of KLF locus, and C-to-T editing at A7-A10 sites, which is similar to
195 the dCas12a system but is distinct from the dCas9/nCas9 system. Comparable to
196 dCas12a, dCas12i-BE exhibited higher base editing activity at KLF4, PCSK9 and
197 DYRK1A loci (Fig. 4c-d), suggesting it may have more potential as a base editor. This
198 suggests that the dCas12i system is useful for broad genome engineering applications,
199 including epigenome editing, genome activation, and chromatin imaging[1, 27-30].

200 In summary, the Cas12i system described here, which has robust editing activity and
201 high specificity, is a versatile platform for genome editing or base editing in mammalian
202 cells and could be useful in the future for *in vivo* or *ex vivo* therapeutic applications.

203 Note that two groups have recently reported improved Cas12i2 using similar
204 strategies in this study[31, 32].

205

206 **Authors' Contributions**

207 H.Z., and Y.Z. jointly conceived and designed the project. H.Z., H.W., J.Z., W.Z., M.X., and
208 X.K., performed experiments with plasmid construction, cell culture and FACS and data
209 analysis. M.X., Z.W., Y.W., X.S., J.L., J.H., and N.Z. assisted with experiments. H.Y.
210 supervised the whole project. H.Z., Y.Z., and H.Y. wrote and revised the manuscript.

211

212 **Conflict of interest**

213 The authors disclose a patent application relating to aspects of this work. Y.Z. and H.Y. is
214 the founder of HUIEDIT Therapeutics Inc., and H.Y. is also the founder of HUIGENE
215 Therapeutics Inc.

216

217 **References**

218 1. Anzalone, A.V., Koblan, L.W. & Liu, D.R. Genome editing with CRISPR-Cas nucleases,
219 base editors, transposases and prime editors. *Nat Biotechnol* **38**, 824-844 (2020).
220 2. Doudna, J.A. The promise and challenge of therapeutic genome editing. *Nature* **578**,
221 229-236 (2020).
222 3. Makarova, K.S. et al. Evolutionary classification of CRISPR-Cas systems: a burst of
223 class 2 and derived variants. *Nat Rev Microbiol* **18**, 67-83 (2020).
224 4. Yan, W.X. et al. Functionally diverse type V CRISPR-Cas systems. *Science* **363**,
225 88-91 (2019).
226 5. Kleinstiver, B.P. et al. Genome-wide specificities of CRISPR-Cas Cpf1 nucleases in
227 human cells. *Nature Biotechnology* **34**, 869-+ (2016).
228 6. Cong, L. et al. Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science*
229 **339**, 819-823 (2013).
230 7. Zetsche, B. et al. Cpf1 is a single RNA-guided endonuclease of a class 2
231 CRISPR-Cas system. *Cell* **163**, 759-771 (2015).
232 8. Zhang, B. et al. Mechanistic insights into the R-loop formation and cleavage in
233 CRISPR-Cas12i1. *Nature communications* **12**, 3476 (2021).
234 9. Zhang, H., Li, Z., Xiao, R. & Chang, L. Mechanisms for target recognition and
235 cleavage by the Cas12i RNA-guided endonuclease. *Nat Struct Mol Biol* **27**, 1069-1076
236 (2020).
237 10. Huang, X. et al. Structural basis for two metal-ion catalysis of DNA cleavage by
238 Cas12i2. *Nature communications* **11**, 5241 (2020).
239 11. Yang, Y. et al. Highly Efficient and Rapid Detection of the Cleavage Activity of
240 Cas9/gRNA via a Fluorescent Reporter. *Appl Biochem Biotechnol* **180**, 655-667

241 (2016).
242 12. Gillmore, J.D. et al. CRISPR-Cas9 In Vivo Gene Editing for Transthyretin Amyloidosis.
243 *The New England journal of medicine* **385**, 493-502 (2021).
244 13. Musunuru, K. et al. In vivo CRISPR base editing of PCSK9 durably lowers cholesterol
245 in primates. *Nature* **593**, 429-434 (2021).
246 14. Strecker, J. et al. Engineering of CRISPR-Cas12b for human genome editing. *Nature*
247 *communications* **10**, 212 (2019).
248 15. Kleinstiver, B.P. et al. Engineered CRISPR-Cas12a variants with increased activities
249 and improved targeting ranges for gene, epigenetic and base editing. *Nat Biotechnol*
250 **37**, 276-282 (2019).
251 16. Xu, X. et al. Engineered miniature CRISPR-Cas system for mammalian genome
252 regulation and editing. *Mol Cell* **81**, 4333-4345 e4334 (2021).
253 17. Zhang, L. et al. AsCas12a ultra nuclease facilitates the rapid generation of therapeutic
254 cell medicines. *Nature communications* **12**, 3908 (2021).
255 18. Bae, S., Park, J. & Kim, J.S. Cas-OFFinder: a fast and versatile algorithm that
256 searches for potential off-target sites of Cas9 RNA-guided endonucleases.
257 *Bioinformatics* **30**, 1473-1475 (2014).
258 19. Yuen, C.T.L. et al. High-fidelity KKH variant of Staphylococcus aureus Cas9 nucleases
259 with improved base mismatch discrimination. *Nucleic Acids Res* **50**, 1650-1660
260 (2022).
261 20. Kim, D.Y. et al. Efficient CRISPR editing with a hypercompact Cas12f1 and
262 engineered guide RNAs delivered by adeno-associated virus. *Nat Biotechnol* **40**,
263 94-102 (2022).
264 21. Wang, X. et al. Cas12a Base Editors Induce Efficient and Specific Editing with Low
265 DNA Damage Response. *Cell Rep* **31**, 107723 (2020).
266 22. Richter, M.F. et al. Phage-assisted evolution of an adenine base editor with improved
267 Cas domain compatibility and activity. *Nat Biotechnol* **38**, 883-891 (2020).
268 23. Li, X. et al. Base editing with a Cpf1-cytidine deaminase fusion. *Nat Biotechnol* **36**,
269 324-327 (2018).
270 24. Bravo, J.P.K. et al. Structural basis for mismatch surveillance by CRISPR-Cas9.
271 *Nature* **603**, 343-347 (2022).
272 25. Kleinstiver, B.P. et al. High-fidelity CRISPR-Cas9 nucleases with no detectable
273 genome-wide off-target effects. *Nature* **529**, 490-495 (2016).
274 26. Wang, D., Zhang, F. & Gao, G. CRISPR-Based Therapeutic Genome Editing:
275 Strategies and In Vivo Delivery by AAV Vectors. *Cell* **181**, 136-150 (2020).
276 27. Wang, H. et al. CRISPR-Mediated Programmable 3D Genome Positioning and
277 Nuclear Organization. *Cell* **175**, 1405-1417 e1414 (2018).
278 28. Konermann, S. et al. Genome-scale transcriptional activation by an engineered
279 CRISPR-Cas9 complex. *Nature* **517**, 583-588 (2015).
280 29. Nakamura, M., Gao, Y., Dominguez, A.A. & Qi, L.S. CRISPR technologies for precise
281 epigenome editing. *Nat Cell Biol* **23**, 11-22 (2021).
282 30. Fellmann, C., Gowen, B.G., Lin, P.C., Doudna, J.A. & Corn, J.E. Cornerstones of
283 CRISPR-Cas in drug discovery and therapy. *Nat Rev Drug Discov* **16**, 89-100 (2017).
284 31. Chen, Y. et al. Synergistic engineering of CRISPR-Cas nucleases enables robust

285 mammalian genome editing. *Innovation (Camb)* **3**, 100264 (2022).
286 32. McGaw, C. et al. Engineered Cas12i2 is a versatile high-efficiency platform for
287 therapeutic genome editing. *Nature communications* **13**, 2833 (2022).
288

289 **Material and methods**

290 **Plasmid vector construction.**

291 Human codon-optimized *Cas12i*, *TadA8e* and human *APOBEC3A* genes were
292 synthesized by the GenScript Co., Ltd., and cloned to generate
293 pCAG_NLS-Cas12i-NLS_pA_pU6_BpiI_pCMV_mCherry_pA by Gibson Assembly. crRNA
294 oligos were synthesized by HuaGene Co., Ltd., annealed and ligated into *Bpi*I site to
295 produce the pCAG_NLS-Cas12i-NLS_pA_pU6_crRNA_pCMV_mCherry_pA.
296

297 **Cell culture, transfection and flow cytometry analysis.**

298 The mammalian cell lines used in this study were HEK293T and N2A. Cells were cultured
299 in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% FBS,
300 penicillin/streptomycin and GlutMAX. Transfections were performed using Polyetherimide
301 (PEI). For variant screening, HEK293T cells were cultured in 24-well plates, and after 12
302 hours 2 μg of the plasmids (1 μg Cas12i of a mutant plasmid and 1 μg of the reporter
303 plasmid) were transfected into these cells with 4 μL PEI. 48 hours after transfection,
304 mCherry and EGFP fluorescence were analyzed using a Beckman CytoFlex
305 flow-cytometer. For assay of mutations in target sites of endogenous genes, 1 μg plasmid
306 with Cas12i targeting crRNA was transfected into HEK293T or N2A cells, which were then
307 sorted using a BD FACS Aria III, BD LSRFortessa X-20 flow cytometer, 48 hours after
308 transfection.
309

310 **Detection of gene editing frequency.**

311 Six thousand sorted cells were lysed in 20 μl of lysis buffer (Vazyme). Targeted sequence
312 primers were synthesized and used in nested PCR amplification by Phanta Max
313 Super-Fidelity DNA Polymerase (Vazyme). TIDE and targeted deep sequence analysis
314 was used to determine indel frequencies. Sanger sequencing and EditR were used for

315 quantification of base substitutions (A-to-G or C-to-T).

316

317 **Figure Legends**

318 **Fig.1 hfCas12Max, an engineered natural variant xCas12i, mediated high-efficient**
319 **and -specificity genome editing in mammalian cells. a**, Transfection of plasmids
320 coding Cas12i and sgRNA activates EGFP. **b**, A natural variant xCas12i mediated EGFP
321 activation efficiency determined by flow cytometry. **c**, Flow diagram for detection of
322 genome editing efficiency by transfection of an all-in-one plasmid containing xCas12i and
323 targeted gRNA into HEK293T cells, followed by FACS and NGS analysis. **d**, xCas12i
324 mediated robust genome editing at the *Ttr* locus in N2a cells and *TTR* and *PCSK9* in
325 HEK293T cells. **e-g**, Cas12Max exhibited significantly increased cleavage activity than
326 xCas12i at various genomic or reporter plasmids target sites. **h**, Both Cas12Max and
327 hfCas12Max exhibited a broader PAM recognition profile than other Cas proteins,
328 including 5'-TN and 5'-TNN PAM. **i**, v6.3 reduced off-target at OT.1 and OT.2 sites and
329 retained indel activity at *TTR.2*-ON targets, compared to v1.1-Cas12Max. **j**, TIDE analysis
330 showed that v6.3-hfCas12Max retained comparable activity at TTR.2-ON targets and
331 almost no at 6 OT sites, to Cas12Max. **k**, Comparison of indel activity from Cas12Max,
332 hfCas12Max, LbCas12a, Ultra AsCas12a, SpCas9 and KKH-saCas9 at TTR locus.
333 HfCas12Max retained the comparable activity of Cas12Max, and higher gene-editing
334 efficiency than other Cas proteins. Each dot represents one of three repeats of single
335 target site.

336

337 **Fig.2 dCas12Max-TadA8e and dCas12Max-hA3A.1 mediated high base editing**
338 **activity. a**, Schematics of different versions of dCas12Max base editors. **b**, The A-to-G
339 editing activity of TadA8e.1-dxCas12i-v2.2 was largely increased by combining v1.2 and
340 v1.3 variant. **c**, Comparison of A-to-G editing frequency at the KLKF4 site from
341 TadA8e.1-dxCas12i-v1.2, v2.2 and v4.3, ABE-dCas12Max (TadA8e-dxCas12i-v4.3)
342 showed a high editing activity of 80%. **d**, Comparison of C-to-T editing frequency at the
343 DYRK1A site from hA3A.1-dxCas12i, -v1.2 and v2.2, CBE-dCas12Max

344 (hA3A.1-dxCas12i-v2.2) showed a high editing activity of 50%.

345 **Supplementary Figure Legends**

346 **Supplementary Figure 1. Screen for functional Cas12i in HEK293T cells. a,** Five of
347 ten natural Cas12i nuclease mediated EGFP-activated efficiency in HEK293T cells.
348

349 **Supplementary Figure 2. Identification and characterization of type V-I systems. a**,
350 Nuclease domain organization of SpCas9, LbCas12a, and xCas12i. **b**, Optimal spacer
351 length for xCas12i. **c**, PAM scope comparison of LbCas12a, and xCas12i. xCas12i
352 exhibited a higher editing efficiency at 5'-TTN PAM than Cas12a. **d**, xCas12i mediated
353 robust genome editing (up to 90%) at the *Ttr* locus in N2a cells and *TTR* and *PCSK9* in
354 HEK293T cells.
355

356 **Supplementary Figure 3. Screen for engineered xCas12i mutants with**
357 **high-efficiency editing activity. a**, Schematics of protein engineering strategy using an
358 activatable EGFP reporter screening system. **b**, The relative editing frequencies of over
359 500 rationally engineered xCas12i mutants.
360

361 **Supplementary Figure 4. Other mutants mediated high-efficiency editing. a-b**,
362 xCas12i mutant with N243R increased 1.2, 5, 20-fold activity at DMD.1, DMD.2 and
363 DMD.3 locus. **c**, Both Cas12Max (xCas12i-N243R) and Cas12Max-E336R elevated
364 EGFP-activated fluorescent at different PAM recognition sites.
365

366 **Supplementary Figure 5. Cas12Max induced off-target editing efficiency at sites**
367 **with mismatches using the reporter system (a) and targeted deep sequence (b).**
368

369 **Supplementary Figure 6. hfCas12Max mediates high-efficiency and -specificity**
370 **editing. a**, Schematics of protein engineering strategy using EGFP activatable reporter
371 screening system with on-targeted and off-targeted crRNA. **b**, Rational protein
372 engineering screen of over 200 mutants for highly-fidelity Cas12Max.Four mutants show

373 significantly decreased activity at both OT (off-target) sites and retains at ON.1 (on-target)
374 site. **c**, Different versions of xCas12i mutants. **d**, v6.3-hfCas12Max reduced off-target at
375 OT.1 and OT.2 sites and retained indel activity at *TTR.2*-ON targets, compared to
376 v1.1-Cas12Max. **e**, v6.3-hfCas12Max exhibited comparable indel activity at DMD.1,
377 DMD.2, and higher at DMD.3 locus, than v1.1-Cas12Max.
378
379 **Supplementary Figure 7. hfCas12Max exhibited the high editing activity.** **a**, Indel
380 frenquencies from Cas12Max, hfCas12Max, LbCas12a, Ultra AsCas12a, SpCas9 and
381 KKH-saCas9 at *TTR* locus.
382
383 **Supplementary Figure 8. hfCas12Max mediates high-efficient and -specific editing.**
384 **a-b**, Off-target efficiency of hfCas12Max, LbCas12a, and UltraAsCas12a at in-silico
385 predicted off-target sites, determined by targeted deep sequencing. Sequences of
386 on-target and predicted off-target sites are shown, PAM sequences are in blue and
387 mismatched bases are in red.
388
389 **Supplementary Figure 9. Conserved cleavage sites of Cas12i. a**, Sequence alignment
390 of xCas12i, Cas12i1 and Cas12i2 shows that D650, D700, E875 and D1049 are
391 conserved cleavage sites at RuvC domain. **b**, Introducing point mutations off D650A,
392 E875A, and D1049A results in abolished activity of xCas12i.
393
394 **Supplementary Figure 10. Other strategies for high-efficiency ABE-dxCas12i. a**,
395 TadA8e.1-dxCas12i-v1.2 and v1.3 exhibits significantly increased A-to-G editing activity
396 among various variants at KLKF4 site of genome. **b**, Unchanged or even decreased
397 editing activity from various dCas12-ABEs carrying different NLS at N-terminal. **c**,
398 Increased A-to-G editing activity of TadA8e-dxCas12i-v4.3 by combining v2.2,
399 changed-NLS linker and high-activity Tade8e. **d**, dxCas12i-ABE-N by TadA at the
400 C-terminus of dCas12 slightly increased editing activity.
401
402 **Supplementary Figure 11. Comparison of editing frequencies induced by various**

403 **dCas12-BEs at different genomic target sites. a**, Comparison of A-to-G editing
404 frequencies induced by indicated TadA8e.1-dxCas12i, v1.2, v2.2, and TadA8e.1-dCas12a
405 at *PCSK9* genomic locus.

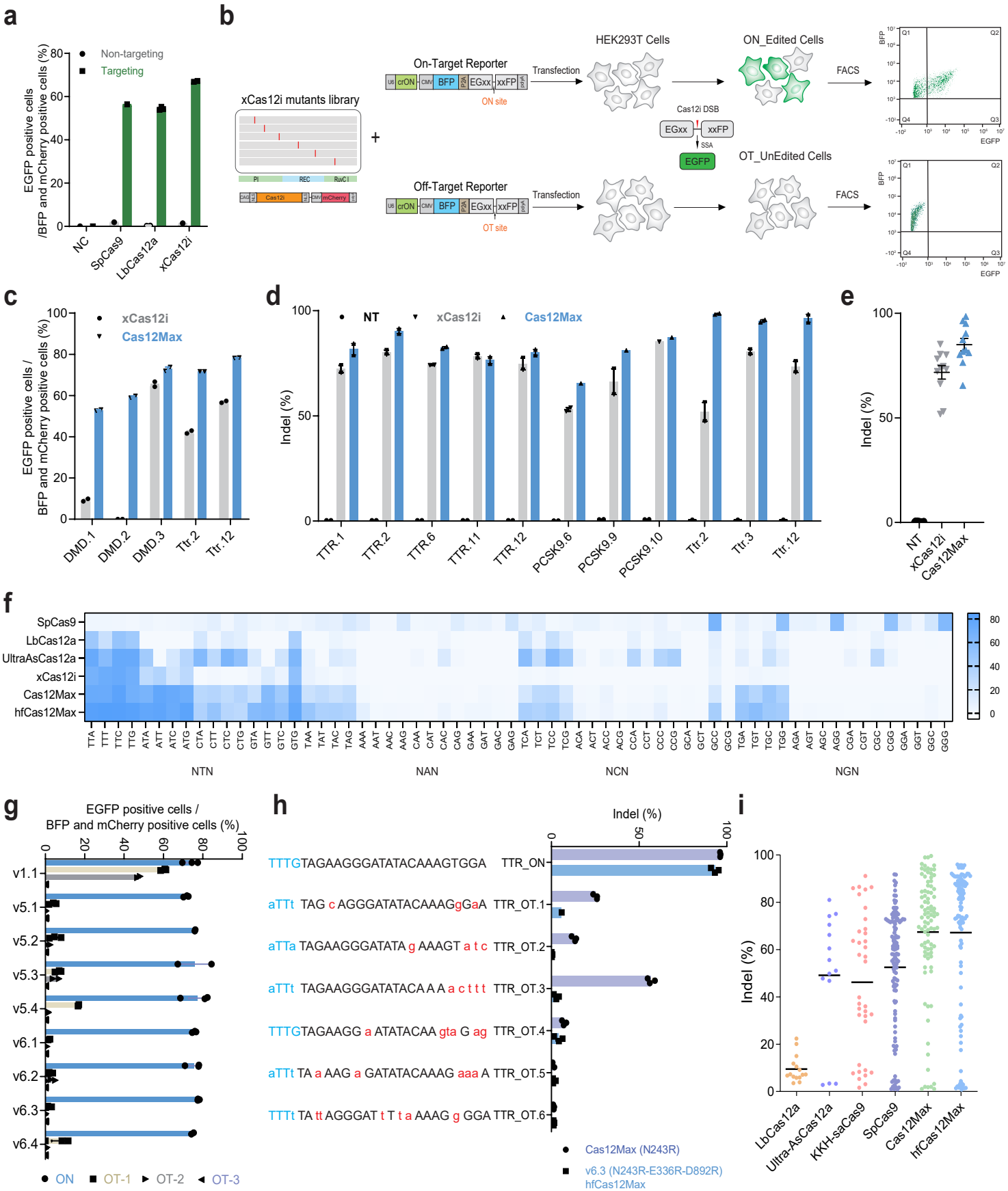

a
EGFP positive cells /BFP and mCherry positive cells (%)
Non-targeting
Targeting
NC
SpCas9
LbCas12a
xCas12i
b
xCas12i mutants library
PI
REC
RuvC I
Cas12i
mCherry
On-Target Reporter
ON site
Off-Target Reporter
OT site
Transfection
HEK293T Cells
ON_Edited Cells
OT_UnEdited Cells
Cas12i DSB
SSA
EGFP
FACS
BFP
EGFP
c
EGFP positive cells / BFP and mCherry positive cells (%)
xCas12i
Cas12Max
DMD.1
DMD.2
DMD.3
Ttr.2
Ttr.12
d
Indel (%)
NT
xCas12i
Cas12Max
TTR.1
TTR.2
TTR.6
TTR.11
TTR.12
PCSK9.6
PCSK9.9
PCSK9.10
Ttr.2
Ttr.3
Ttr.12
e
Indel (%)
NT
xCas12i
Cas12Max
f
SpCas9
LbCas12a
UltraAsCas12a
xCas12i
Cas12Max
hfCas12Max
NTN
NAN
NCN
NGN
g
EGFP positive cells / BFP and mCherry positive cells (%)
v1.1
v5.1
v5.2
v5.3
v5.4
v6.1
v6.2
v6.3
v6.4
ON
OT-1
OT-2
OT-3
h
Indel (%)
TTTGTAGAAGGGATATACAAAGTGGA TTR_ON
aTTt TAG c AGGGATATACAAAGgGaA TTR_OT.1
aTTa TAGAAGGGATATA g AAAGT a t c TTR_OT.2
aTTt TAGAAGGGATATACA A A a c t t t TTR_OT.3
TTTGTAGAAGG a ATATACAA gta G ag TTR_OT.4
aTTt TA a AAG a GATATACAAAG aaa A TTR_OT.5
TTTt TA tt AGGGAT t T t a AAAG g GGA TTR_OT.6
Cas12Max (N243R)
v6.3 (N243R-E336R-D892R) hfCas12Max
i
Indel (%)
LbCas12a
Ultra-AsCas12a
KKH-saCas9
SpCas9
Cas12Max
hfCas12Max

Fig.1 hfCas12Max, an engineered natural variant xCas12i, mediated high-efficient and -specificity genome editing in mammalian cells. a, A natural variant xCas12i mediated EGFP activation efficiency determined by flow cytometry. **b**, Schematics of protein engineering strategy for mutants with high efficiency and high fidelity using an activatable EGFP reporter screening system with on-targeted and off-targeted crRNA. **c-e**, Cas12Max exhibited significantly increased cleavage activity than xCas12i at various genomic or reporter plasmids target sites. **f**, Both Cas12Max and hfCas12Max exhibited a broader PAM recognition profile than other Cas proteins, including 5'-TN and 5'-TNN PAM. **g**, v6.3 reduced off-target at OT.1 and OT.2 sites and retained indel activity at TTR-ON targets, compared to v1.1-Cas12Max. **h**, TIDE analysis showed that v6.3-hfCas12Max retained comparable activity at TTR.2-ON targets and almost no at 6 OT sites, to Cas12Max. **i**, Comparison of indel activity from Cas12Max, hfCas12Max, LbCas12a, Ultra AsCas12a, SpCas9 and KKH-saCas9 at TTR locus. hfCas12Max retained the comparable activity of Cas12Max, and higher gene-editing efficiency than other Cas proteins. Each dot represents one of three repeats of single target site.

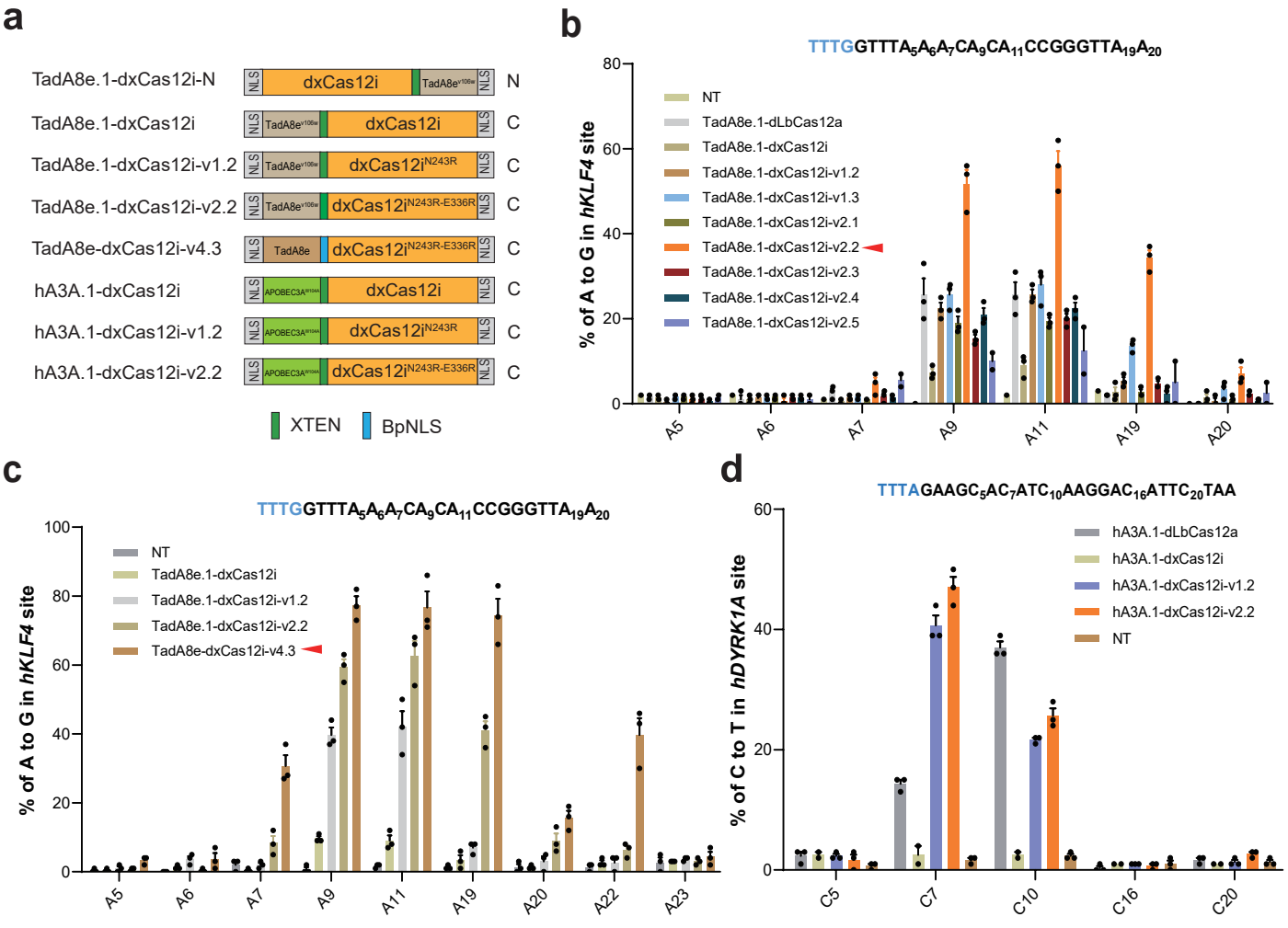
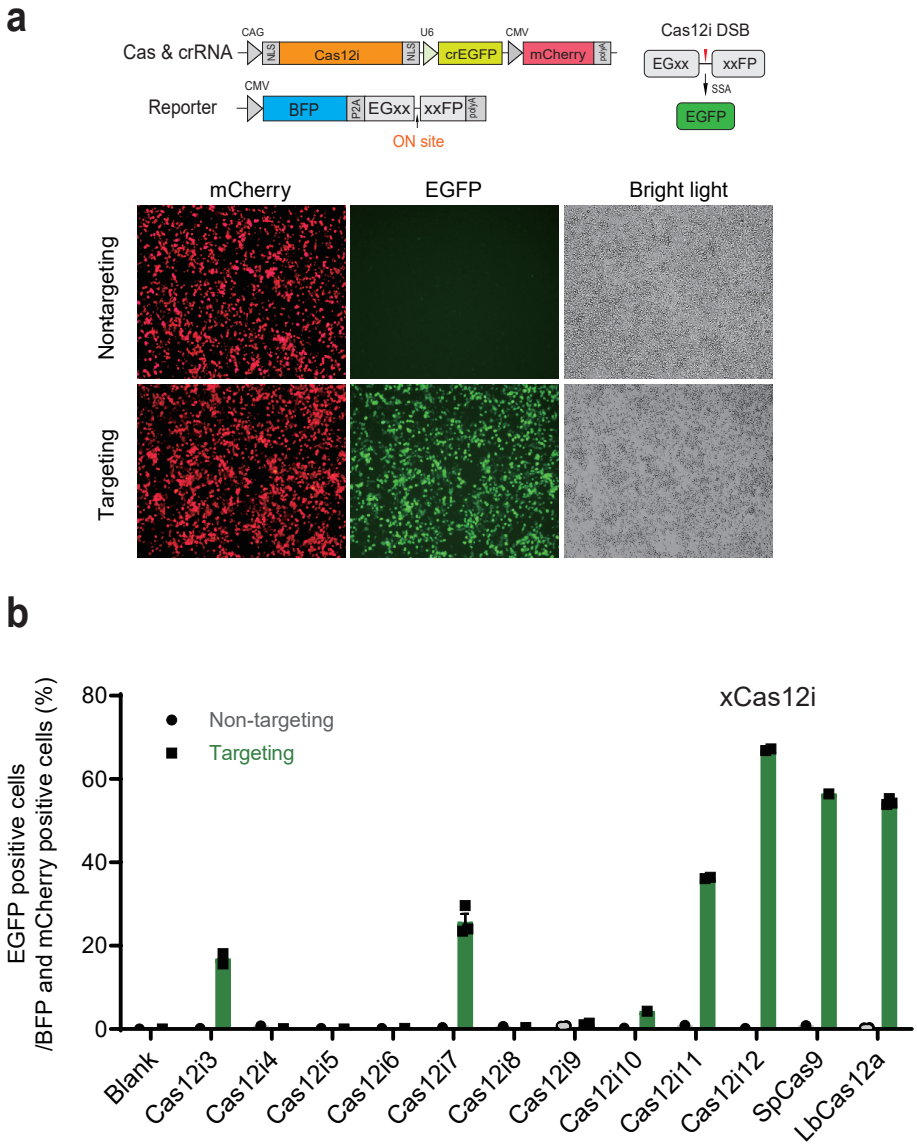
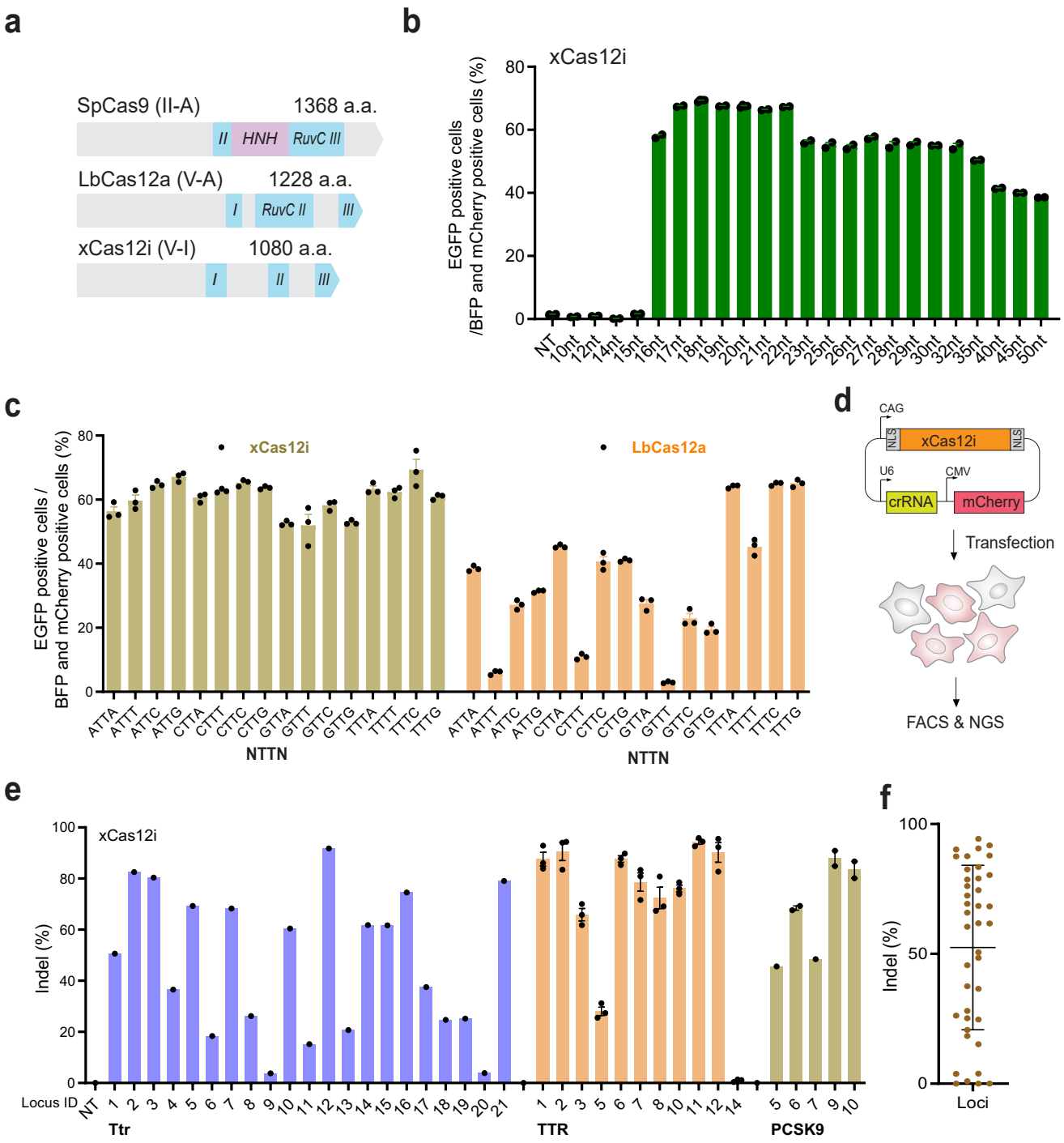


Fig.2 dCas12Max-TadA8e and dCas12Max-hA3A.1 mediated high base editing activity. **a**, Schematics of different versions of dCas12Max base editors. **b**, The A-to-G editing activity of TadA8e.1-dxCas12i-v2.2 was largely increased by combining v1.2 and v1.3 variant. **c**, Comparison of A-to-G editing frequency at the KLKF4 site from TadA8e.1-dxCas12i-v1.2, v2.2 and v4.3, ABE-dCas12Max (TadA8e-dxCas12i-v4.3) showed a high editing activity of 80%. **d**, Comparison of C-to-T editing frequency at the DYRK1A site from hA3A.1-dxCas12i, -v1.2 and v2.2, CBE-dCas12Max (hA3A.1-dxCas12i-v2.2) showed a high editing activity of 50%.

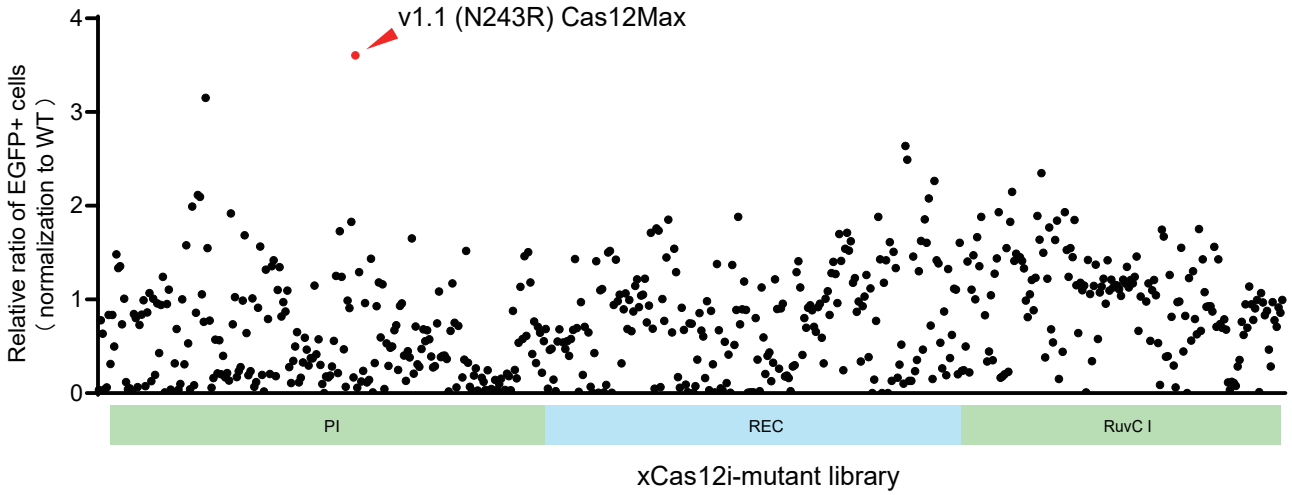


Supplementary Figure 1. Screen for functional Cas12i in HEK293T cells. a, Transfection of plasmids coding Cas12i and crRNA mediated EGFP activation. **b,** Five of ten natural Cas12i nuclease mediated EGFP-activated efficiency in HEK293T cells.

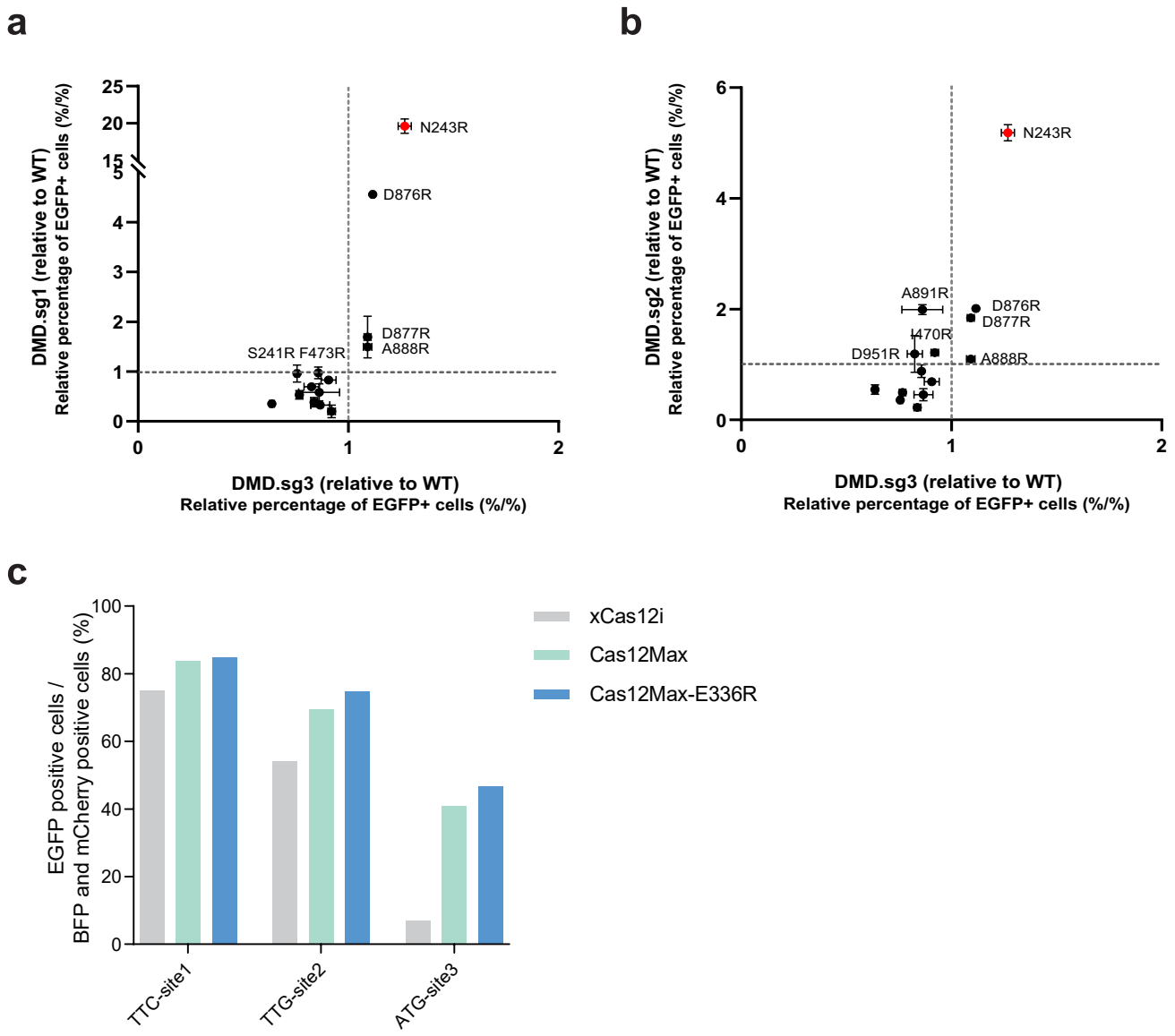


Supplementary Figure 2. Identification and characterization of type V-I systems. a, Nuclease domain organization of SpCas9, LbCas12a, and xCas12i. **b,** Optimal spacer length for xCas12i. **c**, PAM scope comparison of LbCas12a, and xCas12i. xCas12i exhibited a higher editing efficiency at 5'-TTN PAM than Cas12a. **d**, Flow diagram for detection of genome editing efficiency by transfection of an all-in-one plasmid containing xCas12i and targeted gRNA into HEK293T cells, followed by FACS and NGS analysis. **e-f**, xCas12i mediated robust genome editing (up to 90%) at the Ttr locus in N2a cells and TTR and PCSK9 in HEK293T cells.

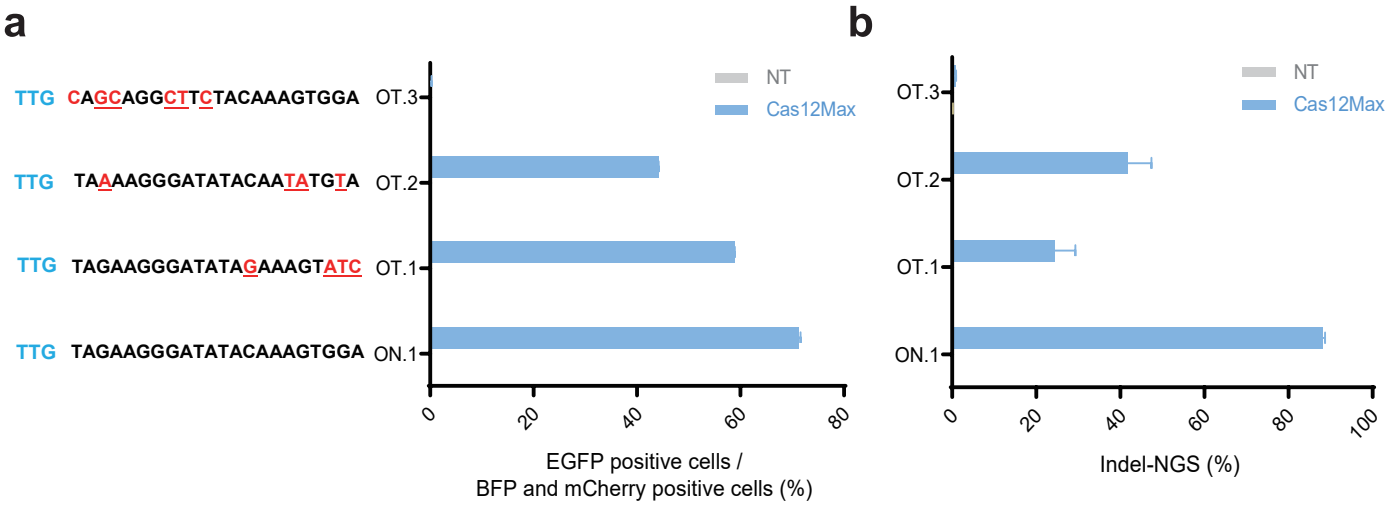
a



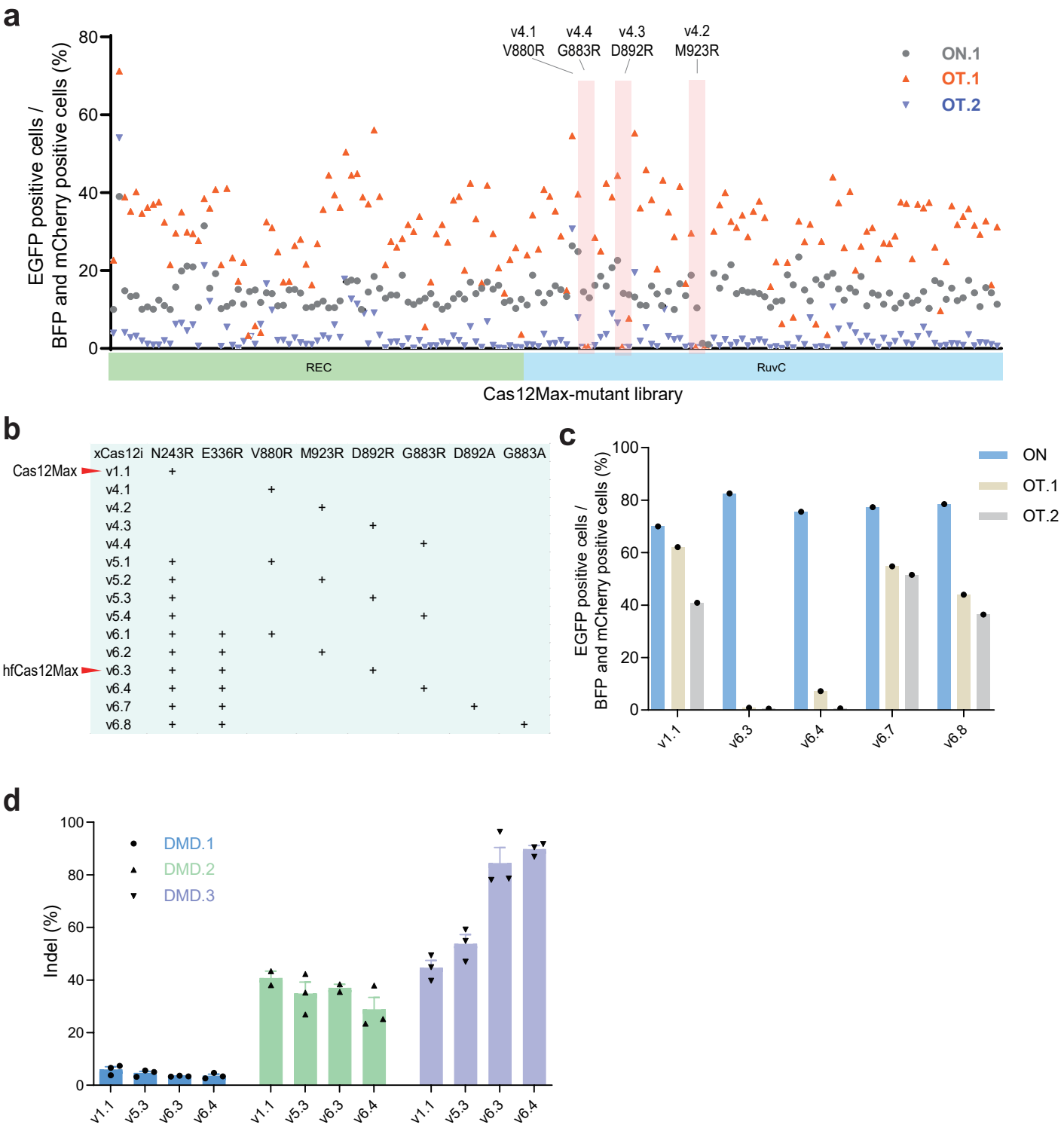
Supplementary Figure 3. Screen for engineered xCas12i mutants with high-efficiency editing activity. a, The relative editing frequencies of over 500 rationally engineered xCas12i mutants.



Supplementary Figure 4. Other mutants mediated high-efficiency editing. a-b, xCas12i mutant with N243R increased 1.2, 5, 20-fold activity at DMD.1, DMD.2 and DMD.3 locus. **c,** Both Cas12-Max (xCas12i-N243R) and Cas12Max-E336R elevated EGFP-activated fluorescent at different PAM recognition sites.

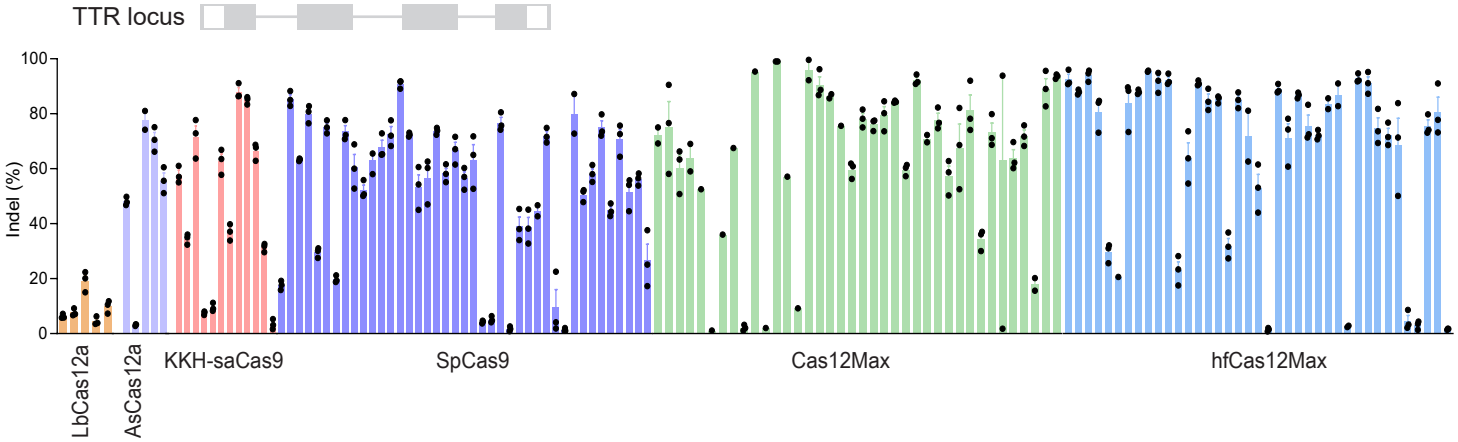


Supplementary Figure 5. Cas12Max induced off-target editing efficiency at sites with mismatches using the reporter system (a) and targeted deep sequence (b).

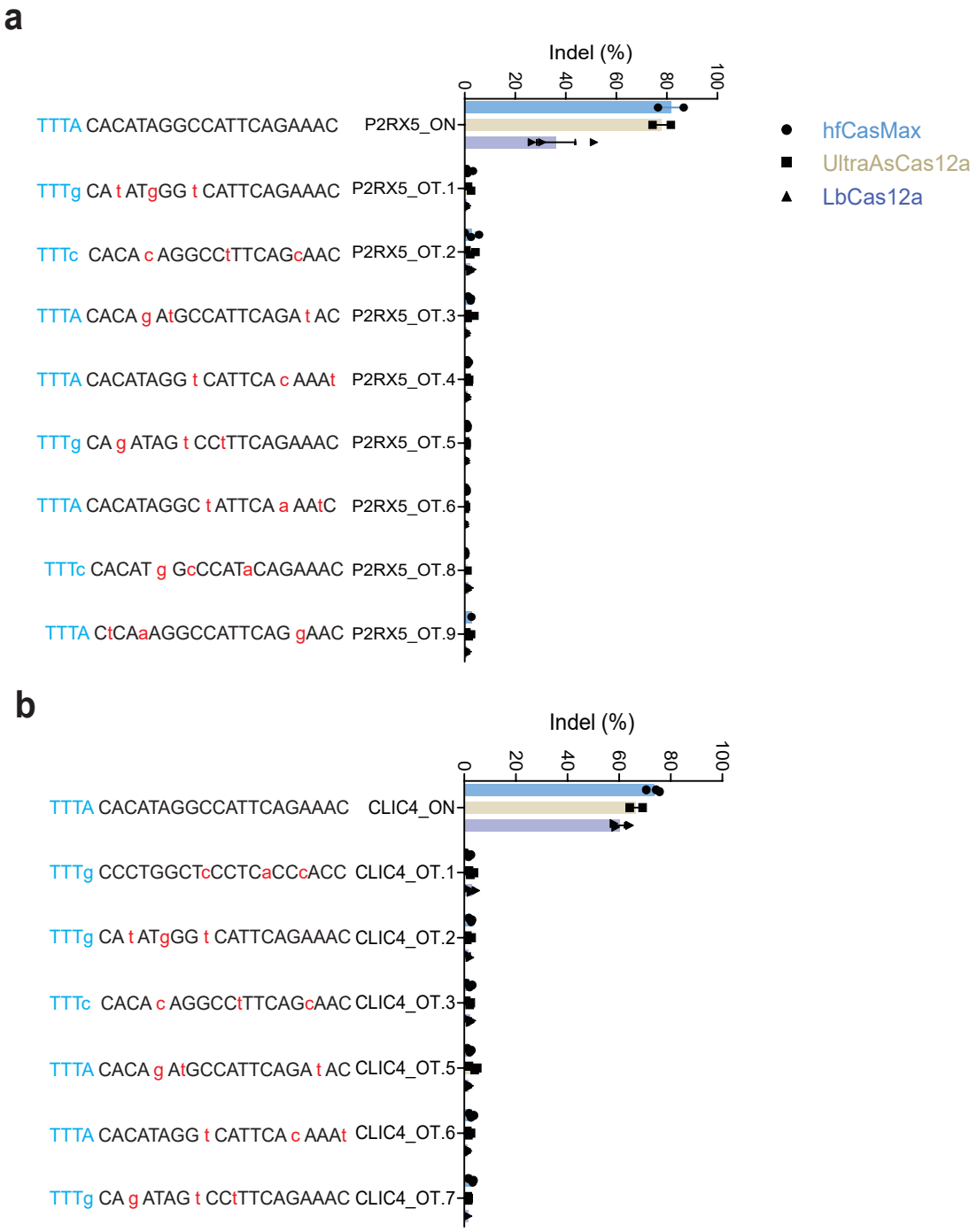


Supplementary Figure 6. hfCas12Max mediates high-efficiency and -specificity editing. a, Rational protein engineering screen of over 200 mutants for highly-fidelity Cas12Max.Four mutants show significantly decreased activity at both OT (off-target) sites and retains at ON.1 (on-target) site. **b**, Different versions of xCas12i mutants. **c**, v6.3-hfCas12Max reduced off-target at OT.1 and OT.2 sites and retained indel activity at TTR-ON targets, compared to v1.1-Cas12-Max. **d**, v6.3-hfCas12Max exhibited comparable indel activity at DMD.1, DMD.2, and higher at DMD.3 locus, than v1.1-Cas12Max.

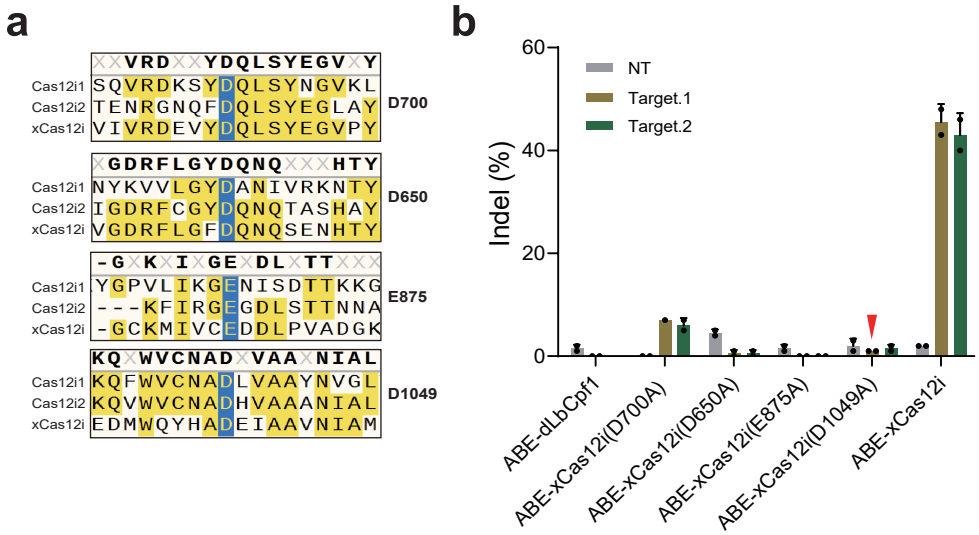
a



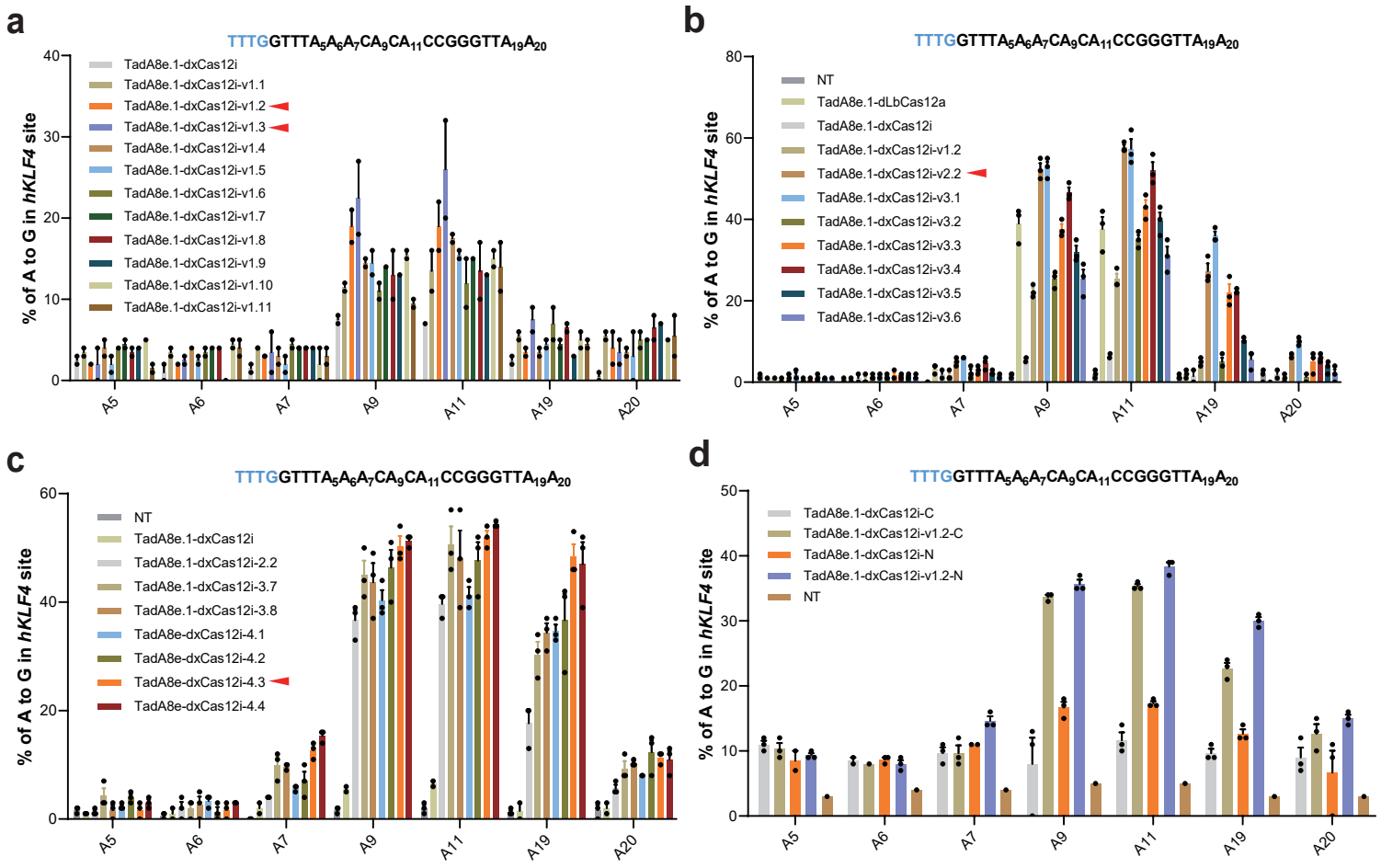
Supplementary Figure 7. hfCas12Max exhibited the high editing activity. **a**, Indel frenquencies from Cas12Max, hfCas12Max, LbCas12a, Ultra AsCas12a, SpCas9 and KKH-saCas9 at TTR locus.



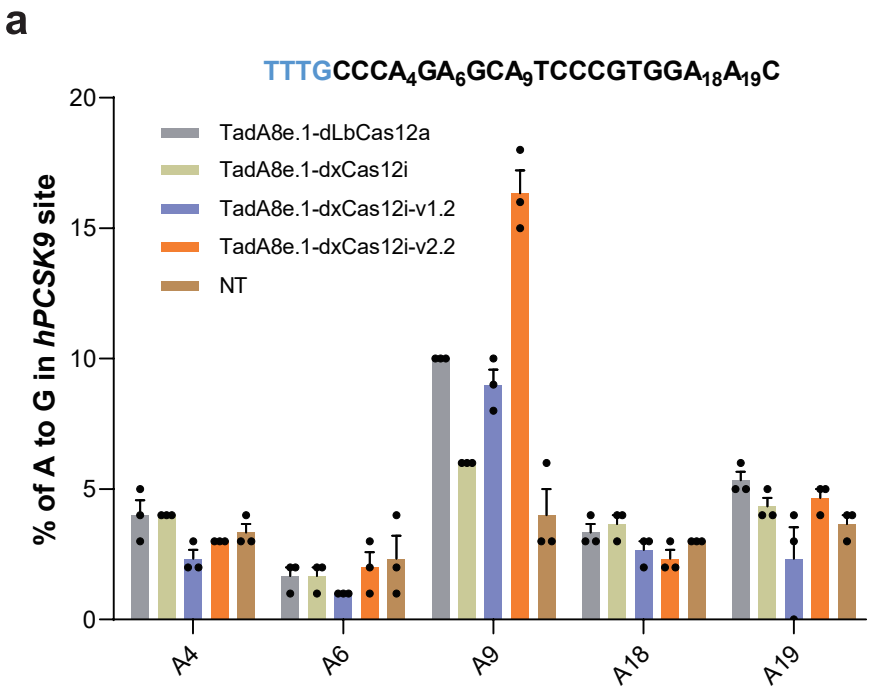
Supplementary Figure 8. hfCas12Max mediates high-efficient and -specific editing. a-b, Off-target efficiency of hfCas12Max, LbCas12a, and UltraAsCas12a at *in-silico* predicted off-target sites, determined by targeted deep sequencing. Sequences of on-target and predicted off-target sites are shown, PAM sequences are in blue and mismatched bases are in red.



Supplementary Figure 9. Conserved cleavage sites of Cas12i. a, Sequence alignment of xCas12i, Cas12i1 and Cas12i2 shows that D650, D700, E875 and D1049 are conserved cleavage sites at RuvC domain. **b**, Introducing point mutations off D650A, E875A, and D1049A results in abolished activity of xCas12i.



Supplementary Figure 10. Other strategies for high-efficiency ABE-dxCas12i. a, TadA8e.1-dx-Cas12i-v1.2 and v1.3 exhibits significantly increased A-to-G editing activity among various variants at KLKF4 site of genome. **b**, Unchanged or even decreased editing activity from various dCas12-ABEs carrying different NLS at N-terminal. **c**, Increased A-to-G editing activity of TadA8e-dxCas12i-v4.3 by combining v2.2, changed-NLS linker and high-activity Tade8e. **d**, dxCas12i-ABE-N by TadA at the C-terminus of dCas12 slightly increased editing activity.



Supplementary Figure 11. Comparison of editing frequencies induced by various dCas12-BEs at different genomic target sites. a, Comparison of A-to-G editing frequencies induced by indicated TadA8e.1-dxCas12i, v1.2, v2.2, and TadA8e.1-dCas12a at PCSK9 genomic locus.