
Supplementary information

Precise DNA base editing using AlphaFold3-based contact modelling

In the format provided by the
authors and unedited

Supplementary Information

Precise DNA base editing via AlphaFold3-based contact modeling

Haowei Meng, Zhixin Lei, Yongchang Yan, Liren Wang, Sihan Zhang, Xichen Rao, Chuyun Shao, Xiaoting Zhang,
Ke Chen, Lei Yang, Rongrong Liu, Gaohui Yang, Ruoyu Shen, Ruichu Gu, Xinyan Wang, Yiya Wang, Suiru Lu,
Zhicong Lv, Bo He, Han Wen, Dali Li, and Chengqi Yi

Supplementary Information	1
Supplementary Discussion	6
Benchmarking of CP and ΔCP	6
Benchmarking of the ContactSeek framework	6
Potential specificity-improving mechanism of K1020D variants	7
Validation of AF3 prediction for the Cas12a complex	8
Limitations of AF3 prediction for protein-DNA-RNA complexes	8
Potential directions for future engineering	9
Supplementary Tables	10
Supplementary Table 1 Primers and guide RNA sequences used in this study.	10
Supplementary Table 2 List of off-target sites identified by dI profiling or Detect-seq.	10
Supplementary Table 3 Targeted amplicon sequencing results for on- and off-target editing.	10
Supplementary Table 4 Benchmark CP against experimental determined protein-DNA/RNA interaction.....	10
Supplementary Table 5 Consensus contact regions identified by ContactSeek.....	10
Supplementary Table 6 Public ATAC-seq and ChIP-seq data used in this study.....	11
Supplementary Figures	12
Supplementary Fig. 1 Flowchart of the dI-profiling method used for ABE off-target detection.....	12
Supplementary Fig. 2 The dI-profiling method efficiently and specifically enriches dI-containing model sequences.....	14
Supplementary Fig. 3 Integrative Genomic Viewer (IGV) views at representative on-target sites.....	18

Supplementary Fig. 4 Integrative Genomic Viewer (IGV) views at representative off-target sites.	20
Supplementary Fig. 5 Enrichment of sequencing reads at ABE off-target sites identified by dI-profiling.	23
Supplementary Fig. 6 Effects of post-transfection time, transfection dose and ABE8e expression level on dI-profiling signals.	25
Supplementary Fig. 7 dI-profiling results are highly reproducible across biological replicates and sequencing platforms.	26
Supplementary Fig. 8 Analysis of mismatches for off-target sequences of Cas9-based ABE.....	29
Supplementary Fig. 9 Distribution and features of ABE-induced off-targets.	31
Supplementary Fig. 10 ABE8e generates a superset of off-target sites compared to ABEmax, ABE7.10 and ABE8.8.....	32
Supplementary Fig. 11 Comparison of dI-profiling-identified ABE8e off-targets with canonical Cas9 nuclease off- targets.	33
Supplementary Fig. 12 Validation of additional off-target sites nominated by dI-profiling and other off-target detection methods.....	36
Supplementary Fig. 13 On-target and off-target complex schematics for Cas9-based ABE.	37
Supplementary Fig. 14 AF3 prediction of the full-length ABE8e complex.	38
Supplementary Fig. 15 Alignment of AF3-predicted Cas9-DNA-sgRNA complex structure to experimental determined structures.	39
Supplementary Fig. 16 High reproducibility and structural convergence of AF3 predictions for Cas9-DNA- sgRNA complex.....	40
Supplementary Fig. 17 AF3 inputs and outputs for ternary complex prediction.	42
Supplementary Fig. 18 Analysis of AF3 embedding differences against off-target mismatch position or signal strength.....	43
Supplementary Fig. 19 Correlation analysis of the two outputs of AF3 prediction.	45
Supplementary Fig. 20 Structural, contact probability and AF3 embedding differences increase with off-target mismatch number.....	46
Supplementary Fig. 21 Position-specific multi-nucleotide mismatches induce localized CP changes within different Cas9 domains.	48

Supplementary Fig. 22 Differential contact probability (Δ CP) analysis for key residues in SuperFi-Cas9.	49
Supplementary Fig. 23 CP-based analysis nominates residues for MaeB cofactor specificity engineering.	51
Supplementary Fig. 24 Schematic illustration of contact probability calculation steps.	52
Supplementary Fig. 25 The majority of CP differences are localized on the protein-NA interfaces.	53
Supplementary Fig. 26 Coverage of contact probability extraction using Top-N entries.	54
Supplementary Fig. 27 Systematic evaluation of the impact of thresholds on identifying amino acid residues that contact nucleic acid in four complexes.	55
Supplementary Fig. 28 CP alterations in off-target complexes.	57
Supplementary Fig. 29 CCRs of Cas9-based ABE across residues.	58
Supplementary Fig. 30 Spatial distribution of Cas9 CCRs from Cas9-based ABEs mapped onto the Cas9-DNA-sgRNA ternary off-target complexes.	59
Supplementary Fig. 31 ContactSeek framework identifies specificity-enhancing residues for Cas9 nuclease.	60
Supplementary Fig. 32 CCRs of Cas9 nuclease across Cas9 residues.	62
Supplementary Fig. 33 Applying ContactSeek to AI-designed DNA binding protein (DBP).	63
Supplementary Fig. 34 Specificity improvement of ABE8e variants based on Cas9 mutations.	65
Supplementary Fig. 35 Experimental determined structures suggest a potential off-target-specific K1020-nucleic acid contact.	67
Supplementary Fig. 36 GUIDE-seq results for Cas9 and Cas9(K1020D) nucleases.	69
Supplementary Fig. 37 ContactSeek framework identifies specificity-enhancing positions for Cas9-based CBE (BE4max).	72
Supplementary Fig. 38 CCRs of Cas9-based CBE across Cas9 residues.	73
Supplementary Fig. 39 The editing specificity of ABE8e variants nominated by using only CP without integrating the dI-profiling signals.	74
Supplementary Fig. 40 Token of AF3 predictions for TadA8e-RNA binary complex.	76
Supplementary Fig. 41 CP-based filtering of AF3-predicted TadA8e-RNA and APOBEC3A-RNA complexes prior to input into the ContactSeek framework.	78
Supplementary Fig. 42 Three positions altering TadA8e editing specificity identified by ContactSeek.	80
Supplementary Fig. 43 Schematic workflow for assessing the on-target activity of ABEs using a large-scale	

sgRNA library.	81
Supplementary Fig. 44 Comparison of ABE8e-DD with several leading high-fidelity ABE variants on additional genomic loci.....	83
Supplementary Fig. 45 Comparison of the most reduced off-target sites and the least reduced off-target sites.	84
Supplementary Fig. 46 On-target editing of SuperFi-Cas9-TadA8e compared to ABE8e.	86
Supplementary Fig. 47 Structural features and AF3 representation for Cas12a-based CBE.....	87
Supplementary Fig. 48 Limited template sequence availability in AF3 database for Cas12a.	88
Supplementary Fig. 49 AF3 prediction of Cas12a-based CBE complexes.....	89
Supplementary Fig. 50 High reproducibility and structural convergence of AF3 predictions for Cas12a-DNA-gRNA complex.	90
Supplementary Fig. 51 Analysis of off-target mismatch number, mismatch position, and signal strength for Cas12a.	91
Supplementary Fig. 52 Position-specific single-nucleotide mismatches induce localized CP changes within different Cas12a domains.	93
Supplementary Fig. 53 Validation of DNA off-target sites identified by Detect-seq for Cas12a-based CBE.	95
Supplementary Fig. 54 Sequence logos for off-target sequences of Cas12a-based CBE identified by Detect-seq.	97
Supplementary Fig. 55 Comparison of Detect-seq-identified Cas12a-based CBE off-targets with in silico predicted Cas12a off-target profiles.	98
Supplementary Fig. 56 CCRs of Cas12a-based CBE across Cas12a residues.....	99
Supplementary Fig. 57 Spatial distribution of Cas12a CCRs from Cas12a-based CBEs mapped onto the Cas12a-DNA-gRNA ternary off-target complexes.....	100
Supplementary Fig. 58 Specificity improvement observed for Cas12a-CBE variants.	101
Supplementary Fig. 59 Token of AF3 predictions for APOBEC3A-RNA binary complex.....	102
Supplementary Fig. 60 Specificity improvement of Cas12a(R284E)-A3A and Cas12a(R284E)-A3A(H29D) on genomic sites that were not used for ContactSeek modeling.....	103
Supplementary Fig. 61 On-target editing rates for Cas12a-A3A CBEs in HCT116 and U2OS cell lines.....	104
Supplementary Fig. 62 FACS gating strategy for assessing the impact of ABE expression on dI-profiling signals.	105

Supplementary Fig. 63 Gating strategy of FACS for RNA-seq.....	108
Supplementary Fig. 64 Representative FACS gating strategy for HCT116 and U2OS cell lines.	109
Supplementary Notes.....	111
Supplementary Reference	126

Supplementary Discussion

Benchmarking of CP and Δ CP

To test if Δ CP reflects known Cas9-DNA-sgRNA interactions, we generated *in silico* DNA sequences containing mismatches from protospacer index positions 1 to 20 and analyzed the corresponding CP changes of the ternary complex (Extended Data Fig. 2a). We found that position-specific mismatches elicited Δ CPs consistent with localized Cas9 conformational changes (Extended Data Fig. 2b and Supplementary Fig. 21); for example, PAM-distal and PAM-proximal mismatches specifically affected the REC3/RuvC-III and REC2/PID domains, respectively^{40,41}.

To further examine if CP analysis could capture key amino acid residues important for Cas9-tsDNA-sgRNA interactions, we attempted to identify, in an unbiased manner, key residues that interact with tsDNA and/or spacer region of sgRNA. CP analysis reported 106 residues, in which 53 overlapped with literature reported residues (56 in total)^{41,83,84}; among the 56 residues, 21 were believed to be key sites and they were fully captured by our unbiased CP analysis (Supplementary Table 4 and Supplementary Fig. 22). This observation demonstrated the ability of CP modeling to nominate key residues for protein-nucleic acid interactions. Notably, 45 of the 50 non-overlapping sites flank the 53 residues, indicating concordant CP changes of nearby residues with the key sites.

Encouraged by these findings, we benchmarked Δ CP against established interactions in known high-fidelity Cas9 variants. HiFi Cas9, HypaCas9 and SuperFi-Cas9 were developed by weakening protein-nucleic acid interactions in off-target complexes^{24,41,85}; and we indeed observed decreased CP at the mutation sites (Extended Data Fig. 2c, d). We also used CP to nominate key residues in a metabolic enzyme (MaeB), demonstrating its broad applicability (Supplementary Fig. 23)⁷⁸.

Benchmarking of the ContactSeek framework

To benchmark the ContactSeek framework, we first assessed whether or not ContactSeek could identify specificity-enhancing residues reported from several high-fidelity Cas9 nuclease variants.

Because Cas9 nuclease and Cas9-based ABE have different off-target landscape⁸⁶, we utilized matched Cas9 nuclease off-target lists, previously determined by GUIDE-seq³¹, as the input for ContactSeek (Supplementary Figs. 31 and 32). Four high-fidelity Cas9 variants (HypaCas9, SpCas9-HF1, LZ3 Cas9, and HiFi Cas9) are believed to exhibit balanced activity and specificity^{14,87}, and they contain 12 distinct specificity-improving residues. Within the top 5 CCRs by ContactSeek, we were able to find 7 out of the 12 residues; moreover, all 7 were identified as key contact residues (Supplementary Fig. 31d, e and Supplementary Table 5). These results validate the capability of ContactSeek to identify key residues for Cas9 nuclease specificity enhancement.

AI-designed DNA-binding proteins (DBPs) are de novo sequences that can bind defined DNA targets and serve as modular tools for gene regulation and editing⁷⁷. Here, we applied AF3-derived CP and ContactSeek to these DBPs. We first confirmed AF3 can correctly predict the designed architectures (Supplementary Fig. 33a). We then predicted DBP-DNA complexes containing single mismatches at experimentally examined specificity-determining bases, and computed CP changes relative to the cognate complex. By comparing the average Δ CPs, we observed that mismatched DNA caused a distinct drop in CP specifically within the designed helix, consistent with the DNA-binding strength measurements in the original study (Supplementary Fig. 33b)⁷⁷. To test whether ContactSeek can identify specificity-determining residues, we focused on DBP35, using AF3-predicted single-mismatch complexes together with the corresponding normalized binding strengths as inputs. ContactSeek identified CCRs and nominated eight key contact residues, including K18, R33 and P42, which together form the experimentally validated K18V/R33N/P42Q variant, as well as additional positions (such as A44) that also could enhance specificity in the site-saturation mutagenesis (SSM) data from the original study (Supplementary Fig. 33c-f). Taken together, the application of ContactSeek to AI-designed DBPs supports its utility in evaluating protein-nucleic acid interactions beyond genome editors.

Potential specificity-improving mechanism of K1020D variants

To look into the mechanism, we performed molecular dynamics (MD) simulations and found that K1020, located on the RuvC loop (Extended Data Fig. 3d and Supplementary Fig. 35), makes

off-target-specific contacts with the 5' segment of the ntsDNA, whereas the K1020D mutation disrupts such interactions (Extended Data Fig. 3e-i). Supporting this notion, we further showed that K1020D likewise improved the specificity of the Cas9 nuclease, CBE, and PE (Extended Data Fig. 3j-p and Supplementary Figs. 36-38). Notably, we demonstrated that sequencing-derived off-target editing strength is important for ranking and identifying key contact residues; using only CP without integrating the empirical strength failed to obtain specificity-enhancing residues (Supplementary Fig. 39 and Supplementary Table 5).

Validation of AF3 prediction for the Cas12a complex

The full-length Cas12a-based CBE consists of the Cas12a protein, a deaminase, and UGI components. However, AF3 initially produced mis-positioned predictions for this complete assembly; hence we omitted the deaminase and UGI components (Supplementary Fig. 49). We then found high reproducibility of AF3-mediated predictions for the Cas12a-DNA-gRNA ternary complex and consistent pLDDT scores (median >78) across independent initializations (Supplementary Fig. 50). In addition, cosine distance analyses of the Pairformer embeddings showed evident differences between on- and off-target predictions compared to the random background (Supplementary Fig. 51), indicating that AF3 captured interaction changes for off-target complex in Cas12a. We also used *in silico* DNA sequence analysis to confirm localized Δ CP responses in the corresponding Cas12a domains (Supplementary Fig. 52). We concluded that AF3 outputs for Cas12a could be used for subsequent ContactSeek analysis.

Limitations of AF3 prediction for protein-DNA-RNA complexes

When predicting the complete ABE8e ternary complex, we observed that TadA8e deaminase was incorrectly positioned onto the sgRNA tetraloop (possibly resembling its inherent substrate tRNA); hence we had to exclude the deaminase domain from our subsequent Cas9-DNA-sgRNA predictions. The deaminase was also mis-positioned during Cas12a-CBE predictions. We suspect the primary reason is the severe scarcity of high-quality training data for complexes, combined with their inherently complicated and diverse interaction patterns that current deep learning models have not been adequately optimized to capture. Moreover, previous studies show that Cas9 undergoes multiple

conformational rearrangements during DNA recognition, R-loop formation and nuclease activation, whereas current AF3 provides essentially static single-state models and do not explicitly sample these functionally important dynamics.

Potential directions for future engineering

We did not perform a second round of ContactSeek in this study; however, our framework supports iteration in principle, which could be utilized to further reduce mismatch tolerance in the PAM-distal region of Cas12a and Cas9. In this study, by analyzing residual off-target editing of ABE8e-DD at HEK4 and Cas12a(R284E)-A3A(H29D) at RUNX1 and DYRK1A (Supplementary Fig. 45), we found that the most-reduced off-target sites carried more mismatches and exhibited lower CP at the guide RNA-tsDNA interface than the least-reduced sites. While seed-region mismatches were generally not tolerated, PAM-distal mismatches for Cas12a-based CBEs and mismatches within the protospacer 7-9 region of Cas9 remained relatively tolerated, with the latter being consistent with the previously reported Cas9 “blind spot”²⁴. These mismatch-tolerant regions therefore represent potential targets for subsequent ContactSeek iterations aimed at further enhancing editor specificity.

Supplementary Tables

Supplementary Table 1 | Primers and guide RNA sequences used in this study.

This table includes the primers used for targeted amplicon sequencing, the guide RNA sequences, the pegRNA sequences, the sgRNA library sequences, and the spike-in model sequences that were used in dI-profiling method development. All genomic coordinates are based on the hg38 reference genome.

Supplementary Table 2 | List of off-target sites identified by dI profiling or Detect-seq.

This table details the off-target sites identified by dI-profiling or Detect-seq in this study. For each off-target site, the table specifies the unique identifier, the genomic coordinates of the guide RNA binding site, and the alignment sequence. For seven representative loci, the positions of ABE-edited adenines (As) and the corresponding normalized dI-profiling signal strengths are also included.

Supplementary Table 3 | Targeted amplicon sequencing results for on- and off-target editing.

This table reports the validation of edited As at off-target sites identified by dI-profiling, and the assessment of on- and off-target editing for ABEs and CBEs.

Supplementary Table 4 | Benchmark CP against experimental determined protein-DNA/RNA interaction.

This table compares AF3 contact probability with experimentally determined interactions at the sgRNA-tsDNA interface of SpCas9 complex. For each residue, the maximum CP values for sgRNA and tsDNA are provided.

Supplementary Table 5 | Consensus contact regions identified by ContactSeek.

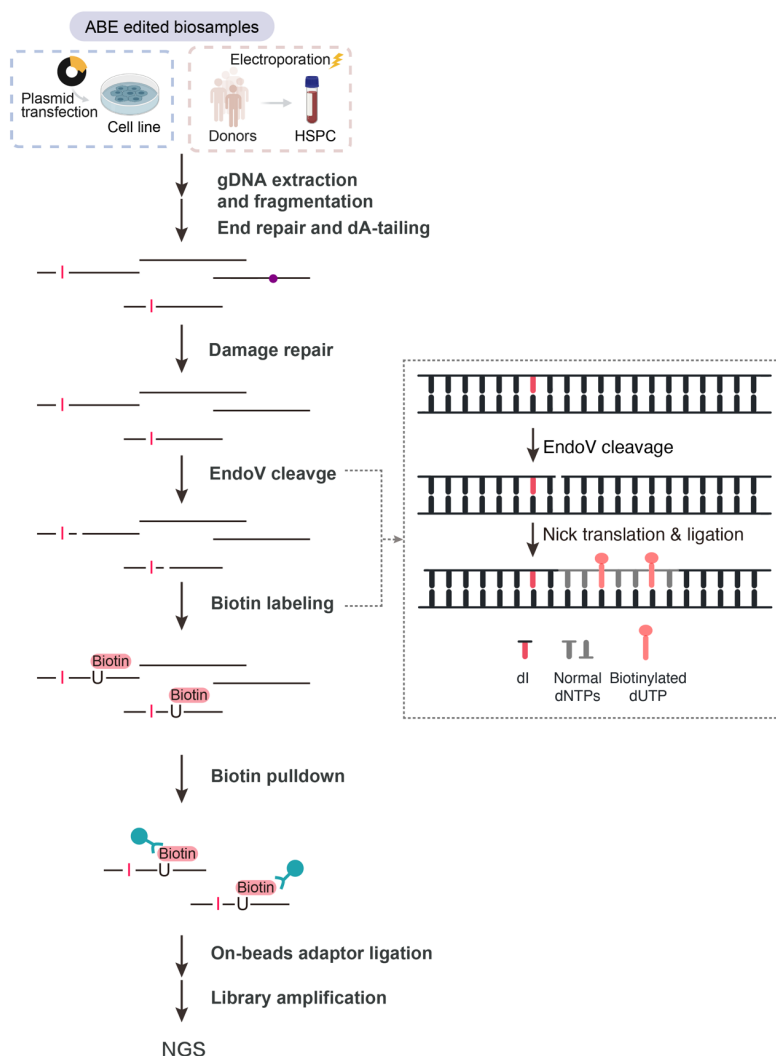
This table lists the consensus contact regions (CCRs) and key contact residues in SpCas9 (based on ABE, CBE or nuclease dataset), LbCas12a (CBE), TadA8e, and APOBEC3A identified by ContactSeek. For Cas9 nuclease, TadA8e, and APOBEC3A, overlaps of the top CCRs and key contact

residues with literature-reported mutation sites are provided. In addition, residue ranks based on Manhattan distance are also included in this table.

Supplementary Table 6 | Public ATAC-seq and ChIP-seq data used in this study.

This table lists the accession numbers of the public ATAC-seq and ChIP-seq datasets analyzed in Supplementary Fig. 9.

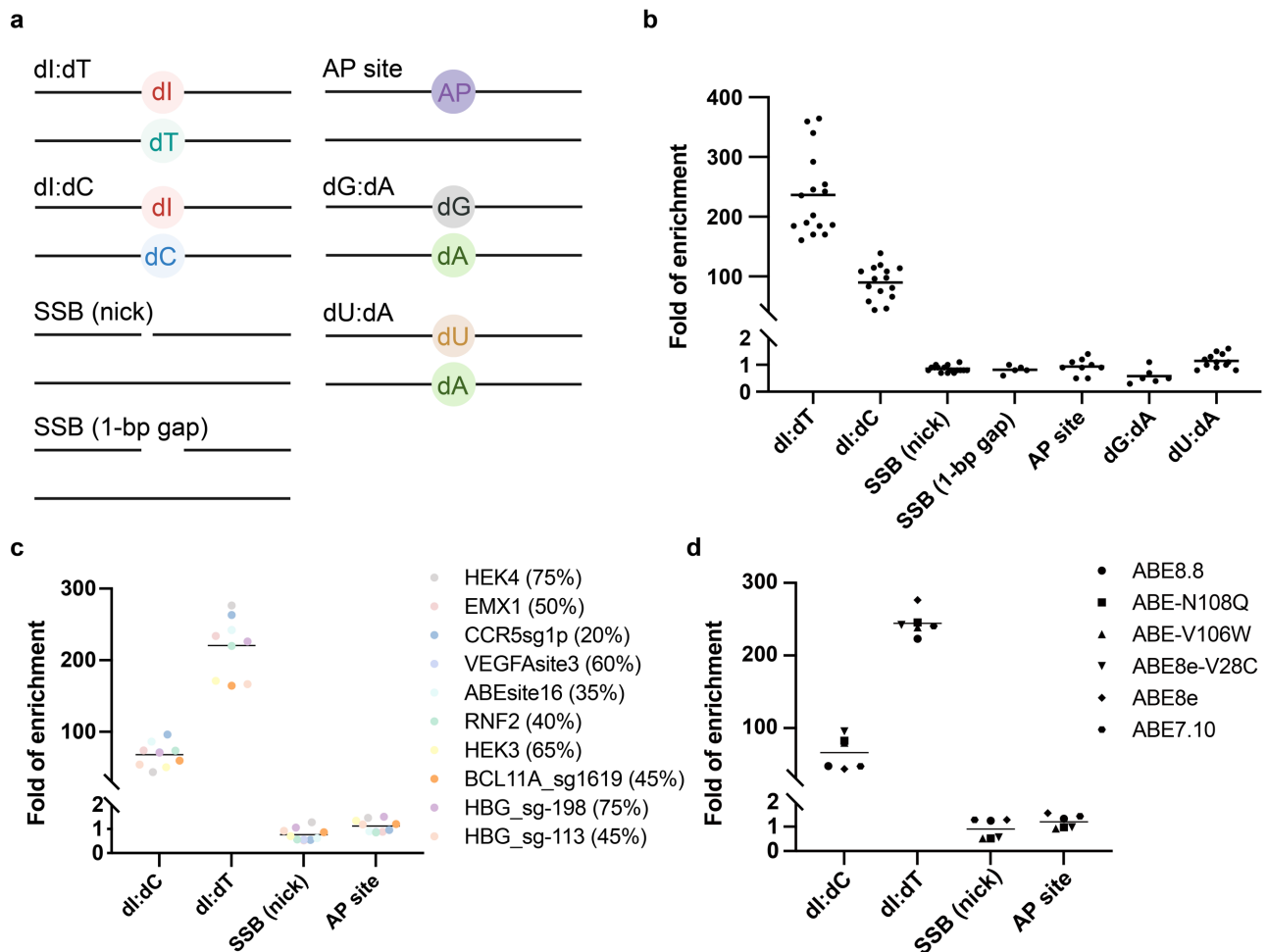
Supplementary Figures



Supplementary Fig. 1 | Flowchart of the dI-profiling method used for ABE off-target detection.

Genomic DNA was extracted from ABE-edited biosamples and sonicated into fragments, followed by end repair and dA-tailing to generate ligation-compatible ends. The DNA fragments were then subjected to a “Damage repair” step to remove native DNA lesions (such as single-stranded breaks) that may interfere with downstream labeling. EndoV was subsequently used to specifically cleave DNA fragments containing dI, introducing a nick downstream of the dI site. During nick translation, nucleotides downstream of dI were labeled with biotinylated dUTPs. The dI-containing fragments were then enriched via biotin pull-down and ligated with Illumina sequencing adaptors on beads.

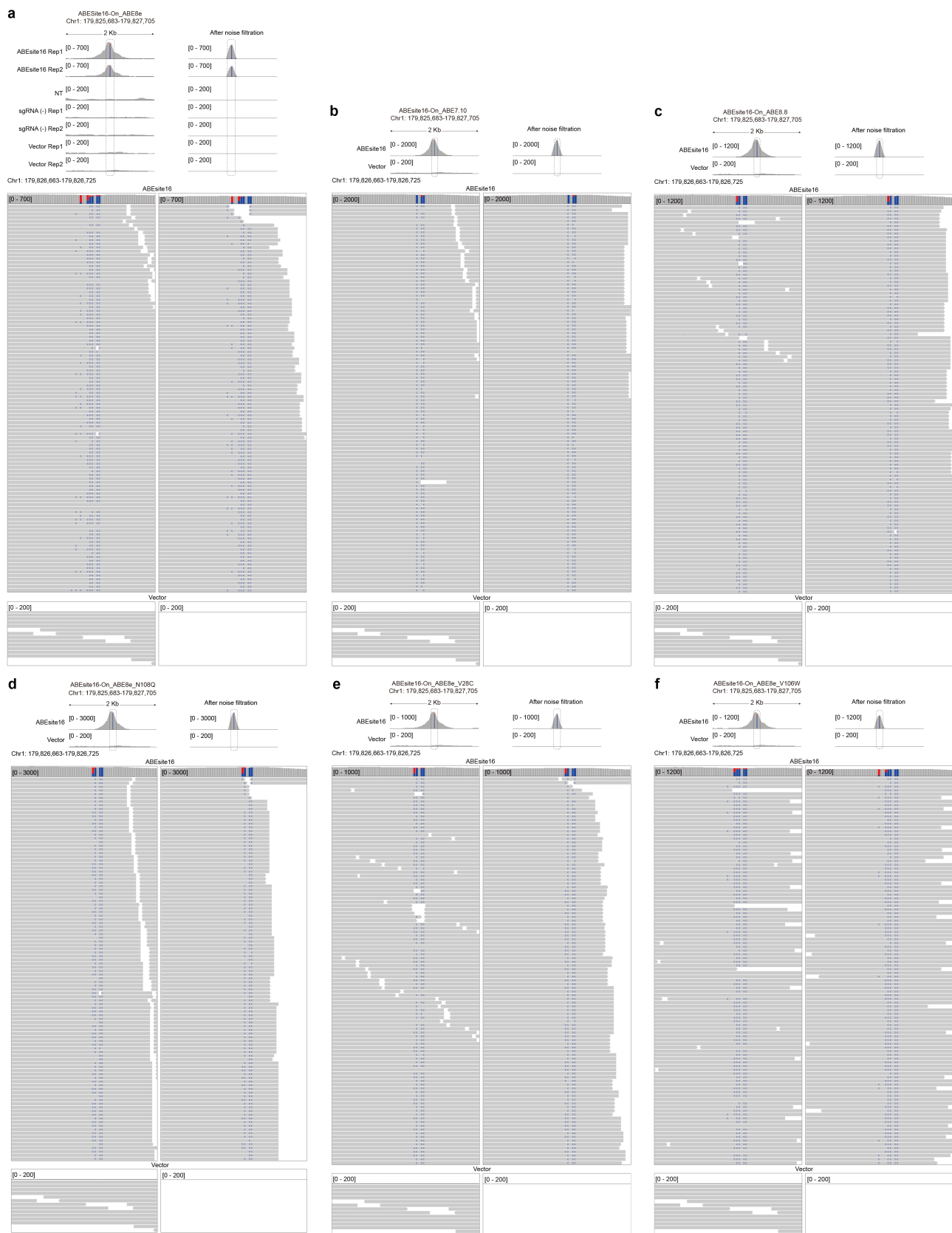
Libraries were constructed by PCR amplification of the captured fragments and then subjected to next-generation sequencing (NGS).

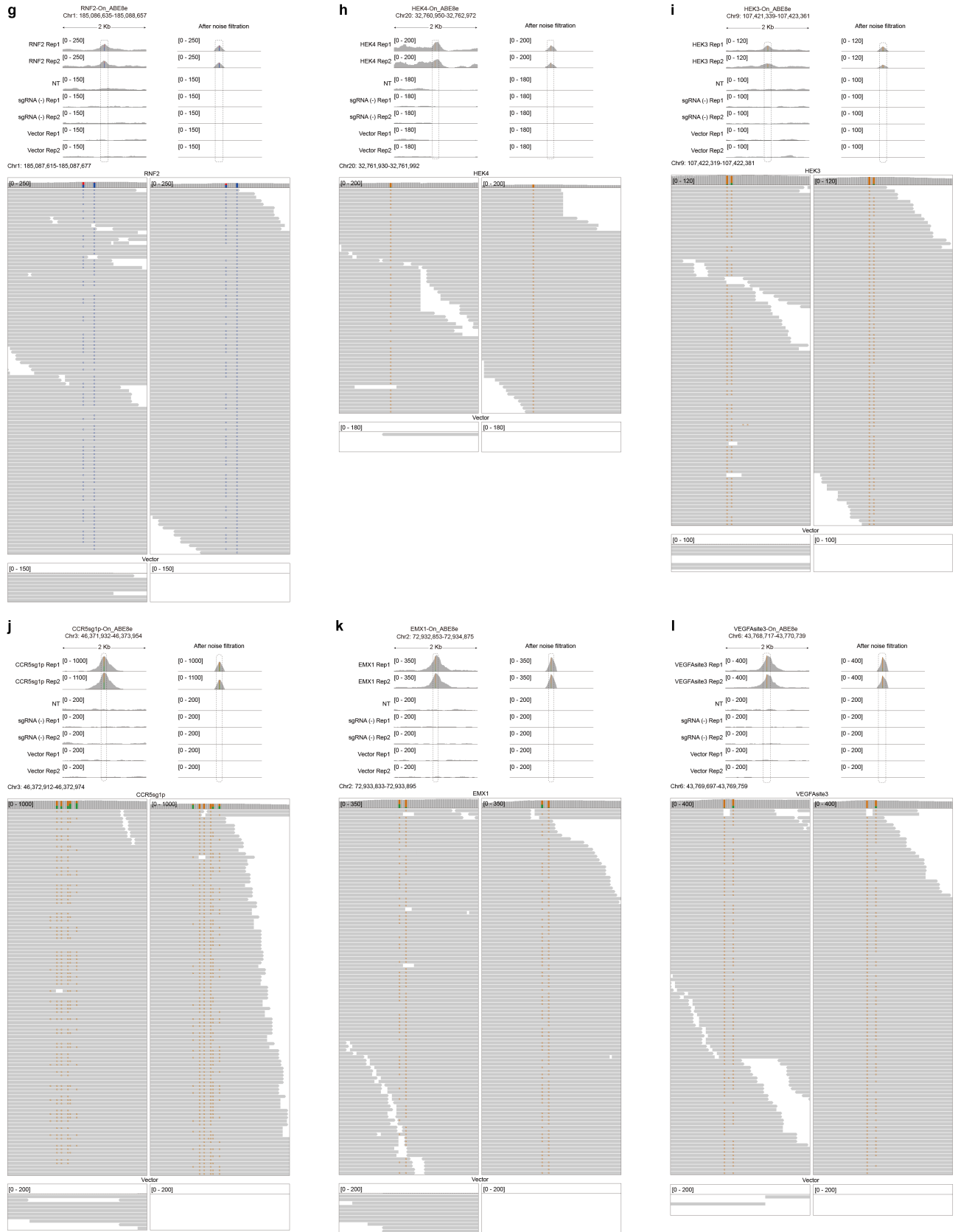


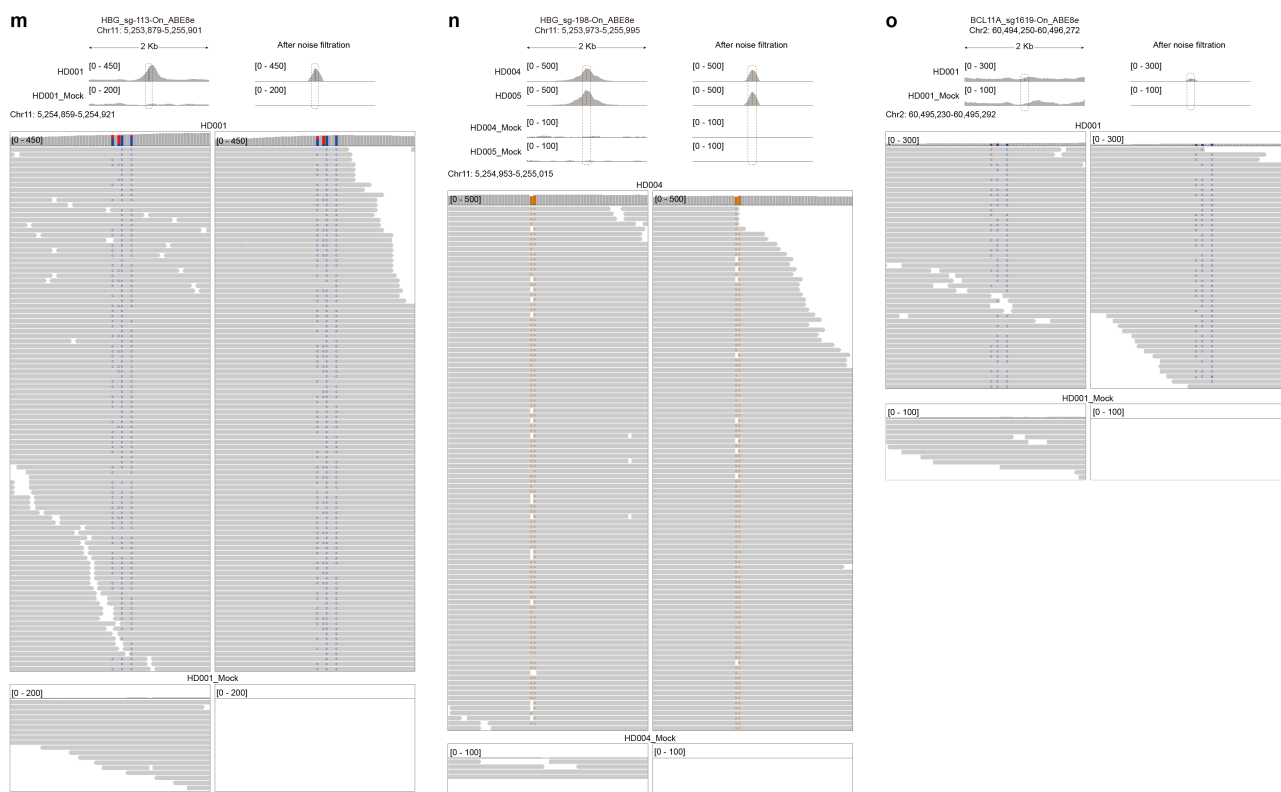
Supplementary Fig. 2 | The dI-profiling method efficiently and specifically enriches dI-containing model sequences.

a, Schematics of spike-in model sequences used in this study. Model DNA sequences containing a dl:dC or dl:dT base pair, a single-stranded break (SSB; including a nick or a 1-bp gap), an abasic (AP) site, a dG:dA mismatch, or a dU:dA base pair were spiked into genomic DNA for dI-profiling (see Supplementary Table 1 for sequences). **b**, Enrichment of each model sequence after biotin-pulldown. Enrichment values were calculated by quantitative PCR using the $2^{-\Delta\Delta C_t}$ method, comparing pull-down samples to input samples and normalizing to the control sequence containing only canonical bases. Each dot denotes an independent sample ($n = 16, 15, 16, 5, 9, 6$, and 12 respectively); the mean values are indicated with black horizontal lines. **c**, Enrichment of model sequences in edited samples using sgRNAs with varying spacer G:C contents ($n = 10$). Different colors represent individual target sites, and the percentage denote the G:C content of the spacer region for each sgRNA. **d**, Enrichment of

model sequences in samples edited by various ABE tools. Different point shapes represent distinct ABE tools ($n = 6$). These results showed that only dI-containing spike-ins were efficiently enriched, but not single-stranded breaks, abasic sites or mismatches, which are endogenous DNA damages that may compromise the accuracy of off-target detection.



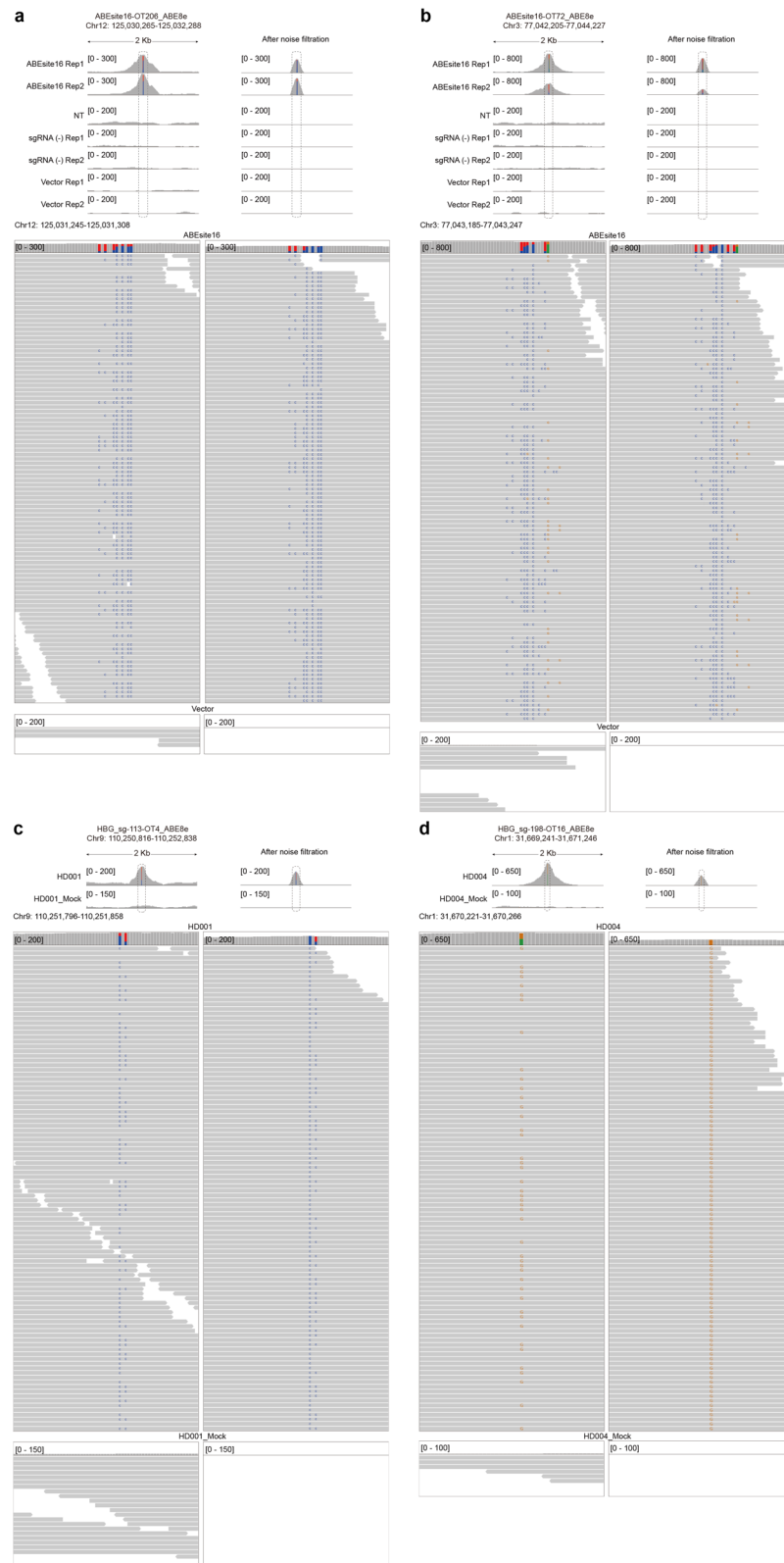




Supplementary Fig. 3 | Integrative Genomic Viewer (IGV) views at representative on-target sites.

On-target signals captured by the dI-profiling method were visualized in the Integrative Genomic Viewer (IGV). **a-f**, dI-profiling signals at the ABEsite16 on-target site in HEK293T cells edited by different ABE variants: ABE8e (**a**), ABE7.10 (**b**), ABE8.8 (**c**), ABE8e-N108Q (**d**), ABE8e-V28C (**e**), ABE8e-V106W (**f**). **g-l**, dI-profiling signals of ABE8e in HEK293T cells at 6 different on-target loci: RNF2 (**g**), HEK4 (**h**), HEK3 (**i**), CCR5sg1p (**j**), EMX1 (**k**), VEGFAsite3 (**l**). **m-o**, dI-profiling signals of ABE8e in CD34⁺ HSPCs at 3 different on-target loci: HBG_sg-113 (**m**), HBG_sg-198 (**n**) and BCL11A_sg1619 (**o**). HD001, HD004, and HD005 denote healthy donor indices 001, 004, or 005. CD34⁺ HSPCs from the denoted donors without electroporation were used as Mock controls. In each panel of **a-o**, dI-profiling signals for different ABE variants or sgRNAs are shown together with the indicated control samples. The dashed box highlights the region that is shown as zoomed-in IGV views of the dI-profiling signal; in both the full and zoomed views, the left column shows signals before noise filtering and the right column shows signals after noise filtering (i.e., after removing reads that lack A-to-G/T-to-C substitutions). A•T-to-G•C mutations are indicated in orange, with orange bars representing the A•T-to-G•C mutation rates. T•A-to-C•G mutations are indicated in blue, with blue

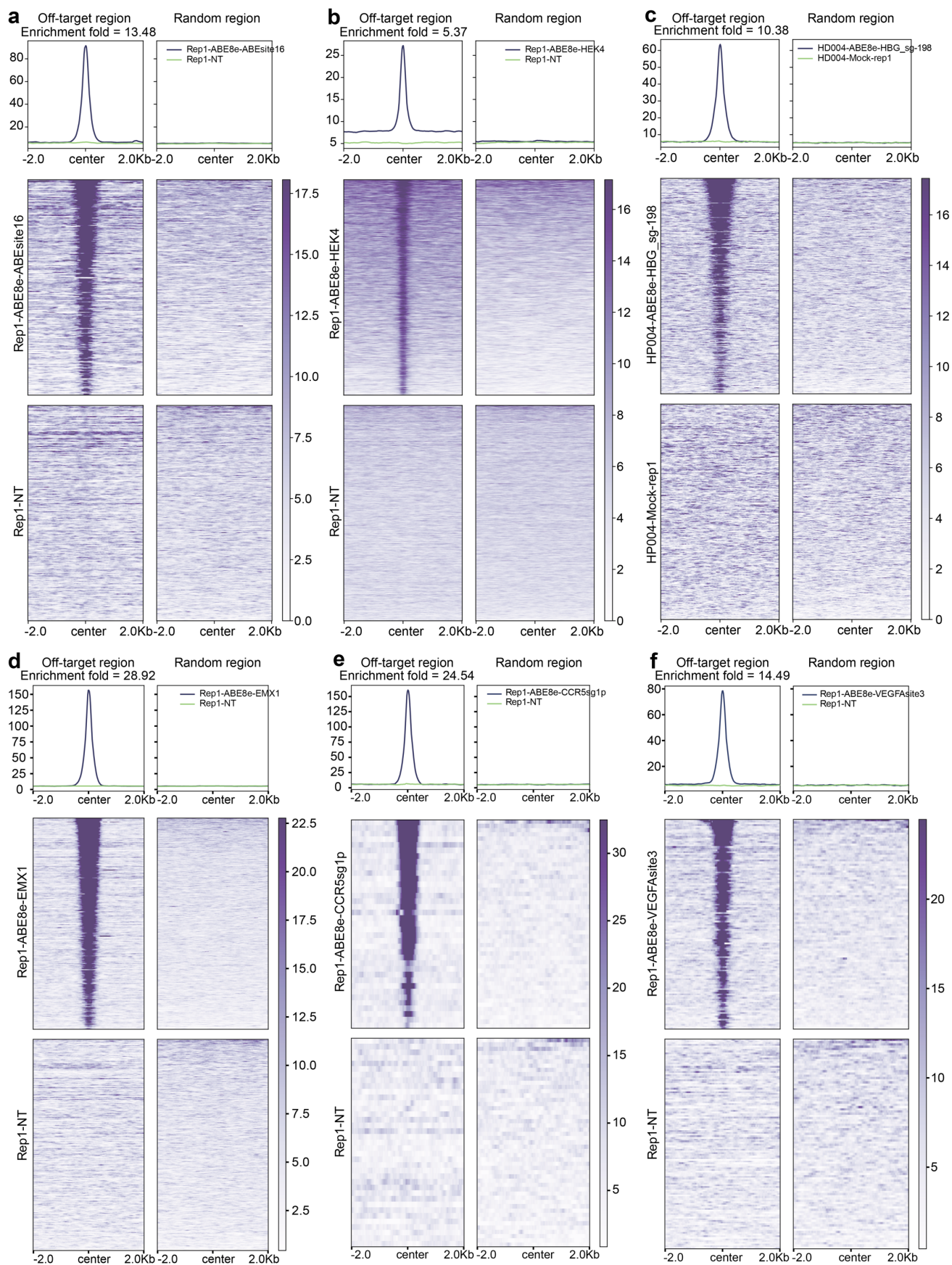
bars representing the T•A-to-C•G mutation rates. ABEsite16, RNF2, HEK4, HEK3, CCR5, EMX1, VEGFAsite3: samples transfected with the indicated ABE variant using the corresponding sgRNA. NT: samples transfected with ABE8e and a non-targeting sgRNA. sgRNA (-): samples transfected with ABE8e alone (without sgRNA). Vector: samples transfected with a plasmid containing only mCherry (empty vector control). For panel **a** to **f**, the same genomic region of the vector sample is presented for clarity.

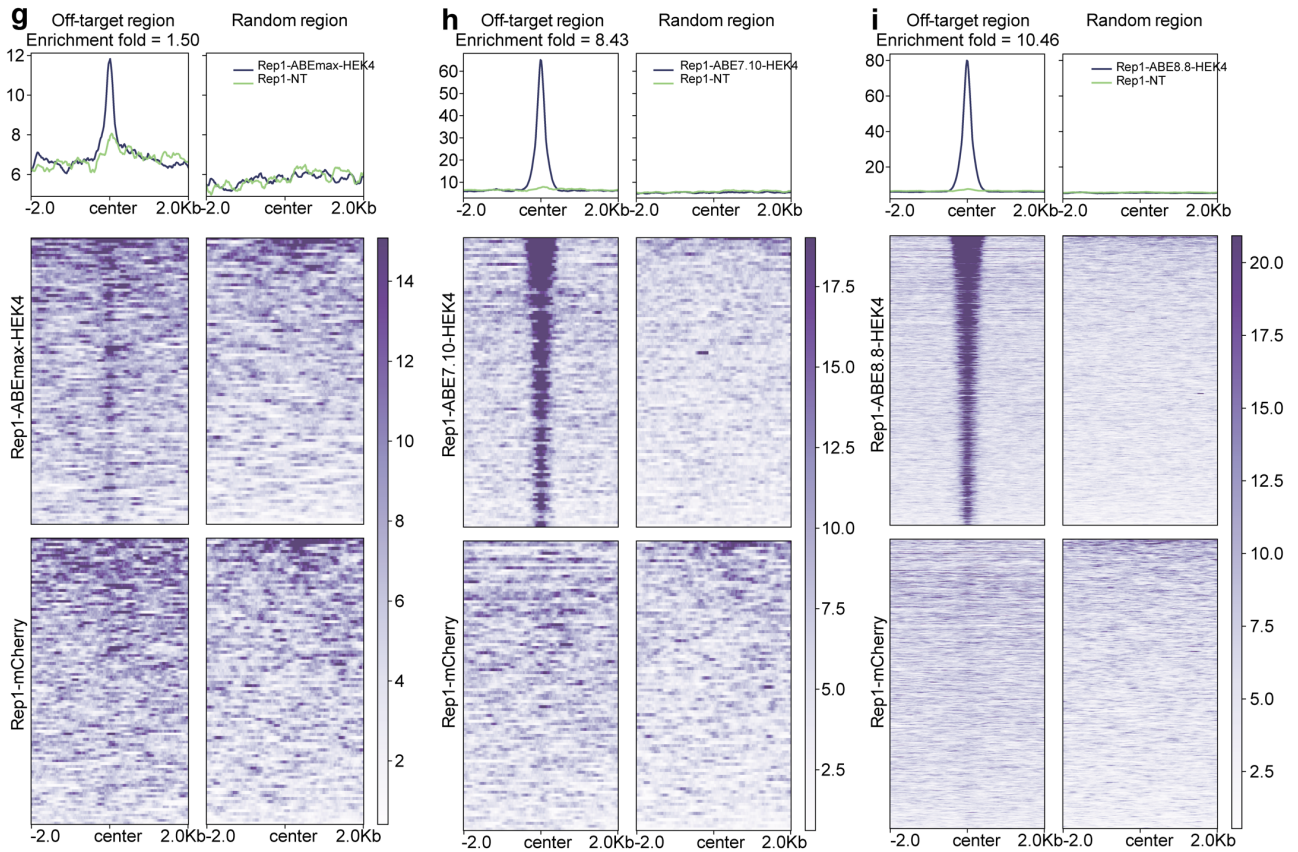


Supplementary Fig. 4 | Integrative Genomic Viewer (IGV) views at representative off-target sites.

Off-target signals captured by the dI-profiling method were visualized in the Integrative Genomic

Viewer (IGV). **a, b**, dI-profiling signals at the ABESite16-OT206 (**a**) and ABESite16-OT72 (**b**) off-target loci in HEK293T cells. NT: samples transfected with ABE8e and a non-targeting sgRNA. sgRNA (-): samples transfected with ABE8e alone (without sgRNA). Vector: samples transfected with a plasmid containing only mCherry (empty vector control). **c, d**, dI-profiling signals at the HBG_sgRNA-113-OT4 (**c**, the same off-target site illustrated in Fig. 1c) and HBG_sgRNA-198-OT16 (**d**) off-target loci in CD34⁺ HSPCs from healthy donors. HD001 and HD004 denote healthy donor indices 001 and 004. CD34⁺ HSPCs from the denoted donors without electroporation were used as mock controls. In each panel of **a-d**, dI-profiling signals for ABE8e-edited samples are shown together with the indicated control samples. The dashed box highlights the region that is shown as a zoomed-in IGV view of the dI-profiling signal; in both the full and zoomed views, the left column shows signals before noise filtering and the right column shows signals after noise filtering (i.e., after removing reads that lack A-to-G/T-to-C substitutions). A•T-to-G•C mutations are indicated in orange, with orange bars representing the A•T-to-G•C mutation rates. T•A-to-C•G mutations are indicated in blue, with blue bars representing the T•A-to-C•G mutation rates. ABESite16, HD001, HD004: samples transfected with ABE8e together with the sgRNA targeting the corresponding on-target locus (ABESite16 in HEK293T cells; HBG_sgRNA-113 in HD001 and HBG_sgRNA-198 in HD004 CD34⁺ HSPCs).

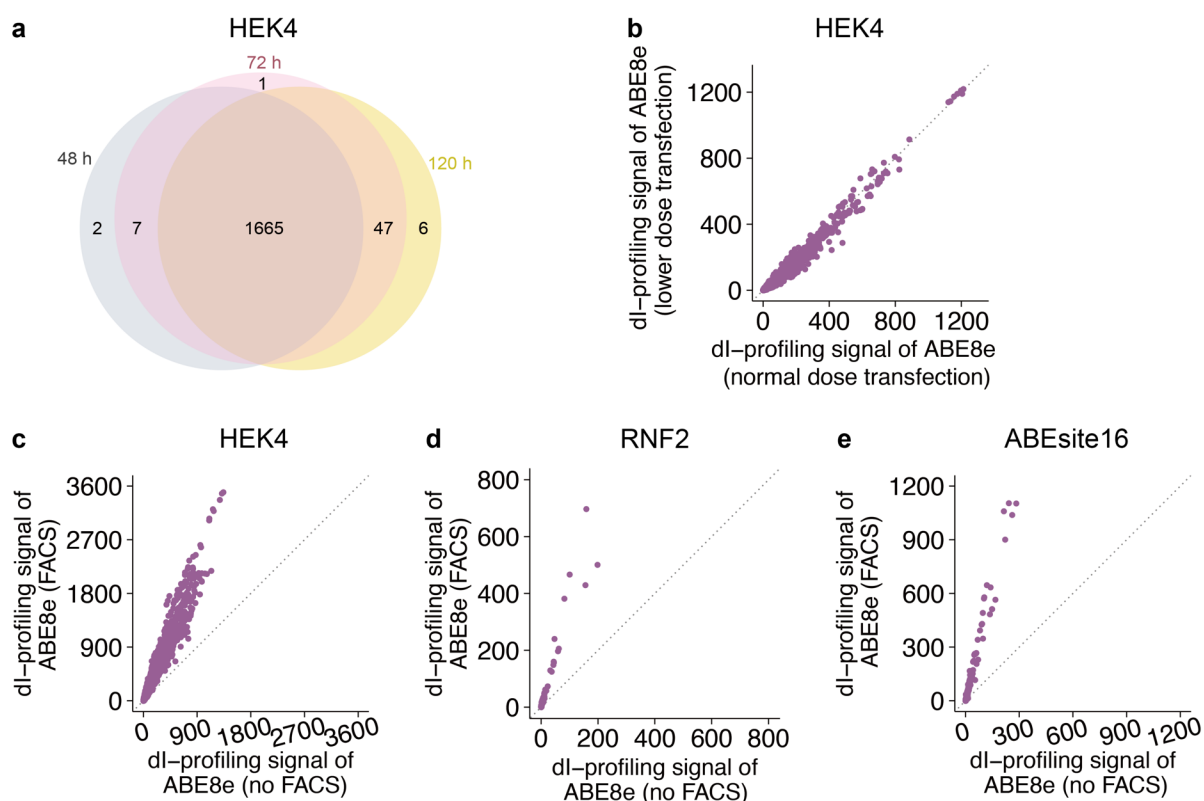




Supplementary Fig. 5 | Enrichment of sequencing reads at ABE off-target sites identified by dI-profiling.

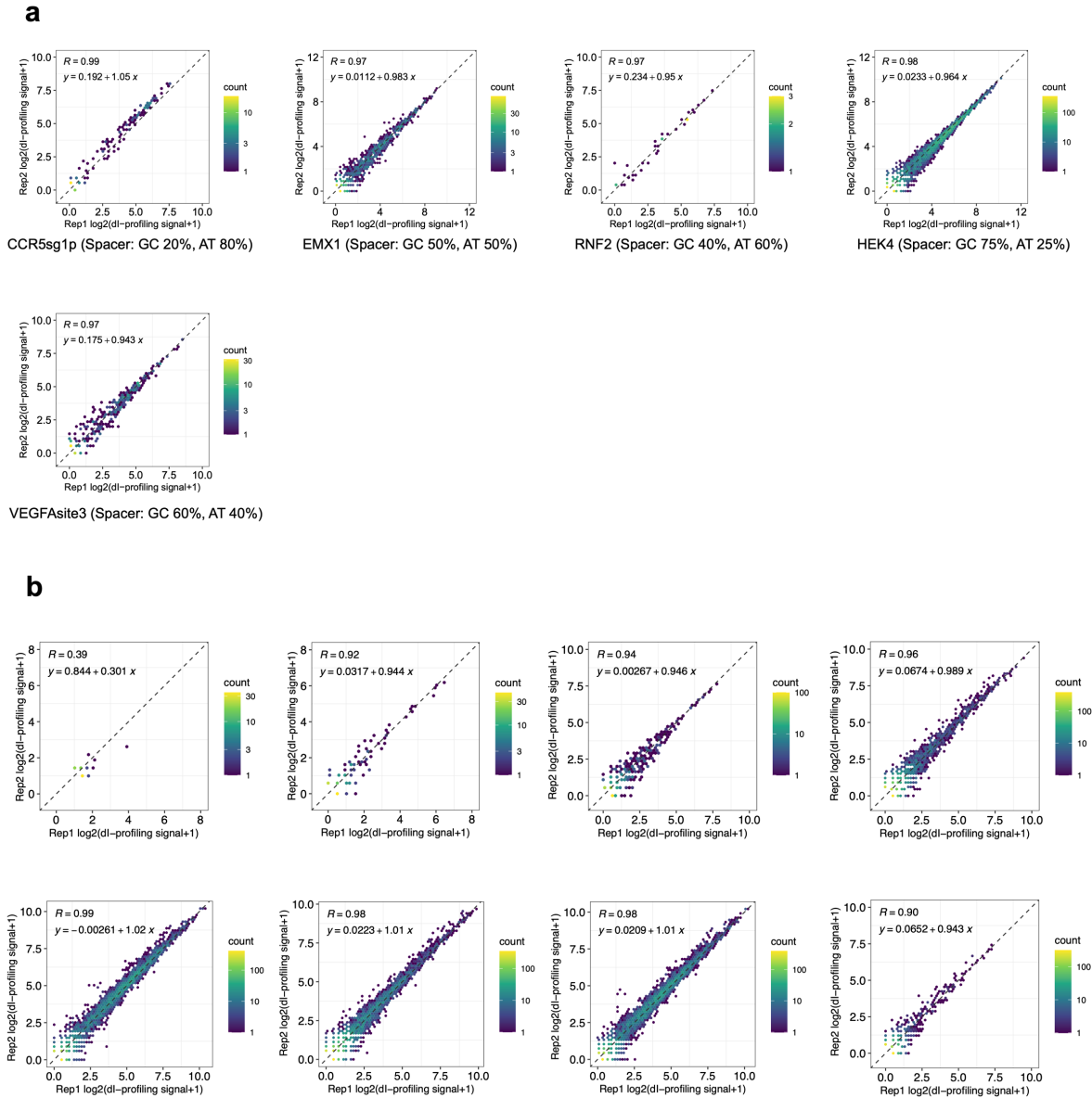
a-f, dI-profiling results from HEK293T cells transfected with ABE8e and the indicated sgRNAs: ABESite16 (**a**), HEK4 (**b**), EMX1 (**d**), CCR5sg1p (**e**), and VEGFASite3 (**f**); and from HSPCs of donor HD004 electroporated with ABE8e and HBG_sg-198 sgRNA (**c**). **g-i**, Enrichment of dI-profiling reads at off-target sites of HEK4 in HEK293T cells edited by ABEmax (**g**), ABE7.10 (**h**), or ABE8.8 (**i**). In each panel, the upper plots show mean RPKM values for genomic regions spanning ± 2.0 kb centered on either off-target sites or randomly selected regions. The middle (ABE-treated samples) and lower (control samples) heatmaps display RPKM values within the same ± 2.0 kb region for individual off-target sites or random regions. Note that distinct enrichment peaks are seen at off-target regions but not at random regions or in control samples. Corresponding control samples are shown in each panel. The enrichment fold of the off-target region relative to the control is indicated in each panel. NT: HEK293T cells treated with ABE8e and a non-targeting sgRNA. HD004 Mock: CD34⁺ HSPCs from the donor HD004 without electroporation. mCherry: HEK293T cells transfected with a mCherry-

expressing plasmid.



Supplementary Fig. 6 | Effects of post-transfection time, transfection dose and ABE8e expression level on dI-profiling signals.

a, Venn diagram illustrating the overlap of dI-profiling off-target sites in HEK293T cells collected 48 h (grey), 72 h (pink), and 120 h (yellow) after transfection with ABE8e and HEK4 sgRNA. **b**, Scatter plot comparing dI-profiling signals for the HEK4 sgRNA (n = 1,717 sites) in HEK293T cells edited by ABE8e under a lower transfection dose (two-thirds of the standard dose) versus the common transfection dose. **c-e**, Scatter plots comparing dI-profiling signals in unsorted versus FACS-enriched HEK293T cells edited by ABE8e and the HEK4 (n = 1,785 sites; **c**), RNF2 (n = 19 sites; **d**), and ABEsite16 (n = 279 sites; **e**) sgRNA. Each dot in **b** to **e** represents an off-target A. ABE8e was fused to EGFP via a P2A linker (ABE8e-P2A-EGFP), and EGFP fluorescence intensity was used as a proxy for ABE8e expression level. EGFP-high cells (top 20% of cells by EGFP fluorescence) were enriched by FACS and showed ~4.2-fold higher mean fluorescence than unsorted cells (Supplementary Fig. 62). Overall, dI is sufficiently stable in cells over the time window relevant to the dI-profiling measurements. While off-target signal strength can be influenced by editor expression level, the global off-target landscape remains largely similar (Supplementary Table 2).

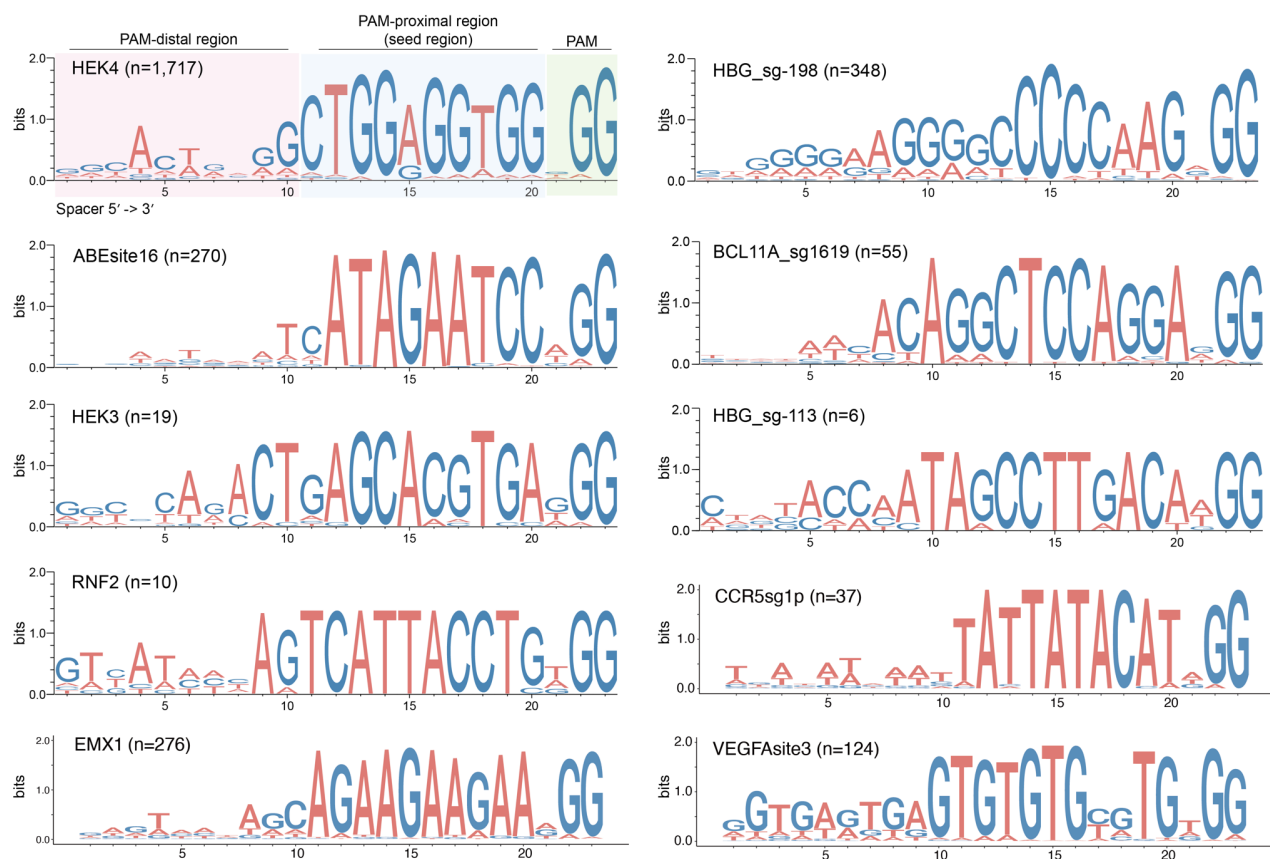


Supplementary Fig. 7 | dI-profiling results are highly reproducible across biological replicates and sequencing platforms.

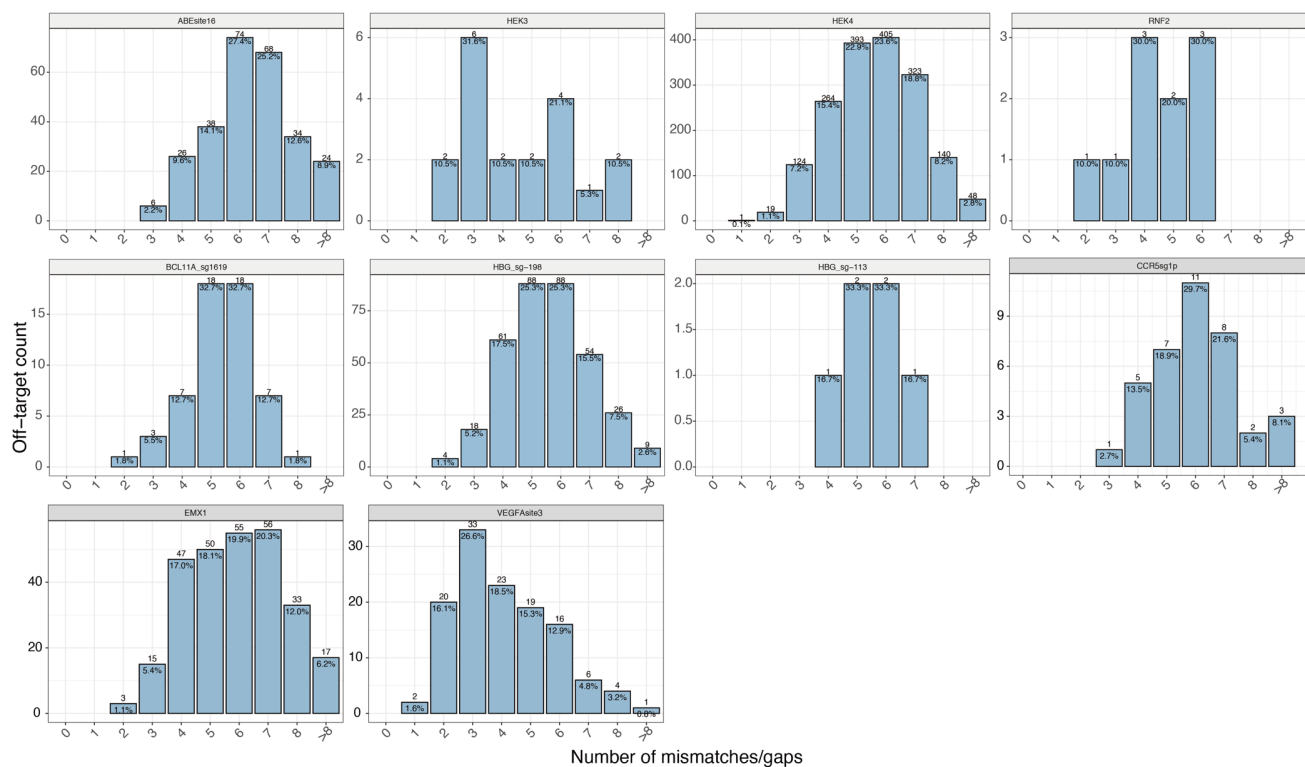
a, Correlation of dI-profiling signals between independent biological replicates for the CCR5sg1p, EMX1, RNF2, HEK4 and VEGFAsite3 sgRNAs. The x- and y-axes show $\log_2(\text{dI-profiling signal} + 1)$ for replicate 1 (Rep1) and replicate 2 (Rep2), respectively. GC and AT contents were calculated from the protospacer sequence excluding the 3-nt NGG PAM. **b**, Correlation of dI-profiling signals between technical replicates sequenced on the MGI DNBSEQ T7 and Salus EVO platforms for the ABEsite16 and HEK4 sgRNAs under the indicated editor conditions. The x- and y-axes show $\log_2(\text{dI-profiling signal} + 1)$ derived from the MGI DNBSEQ T7 and Salus EVO platforms, respectively. Dashed lines

indicate $y = x$. Pearson correlation coefficients (R) and fitted linear regression equations are shown in each panel. Color indicates the number of observations per bin.

a

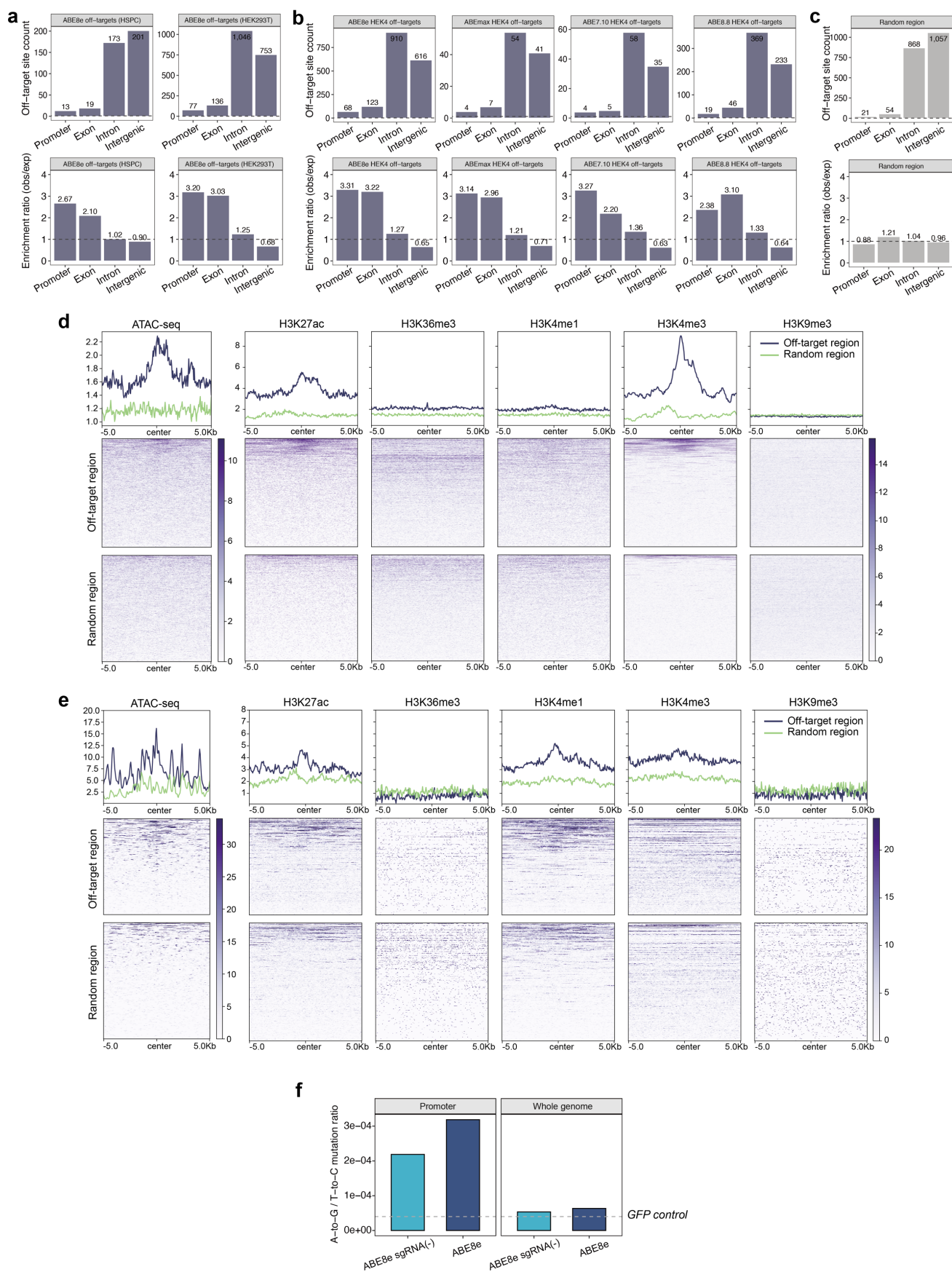


b



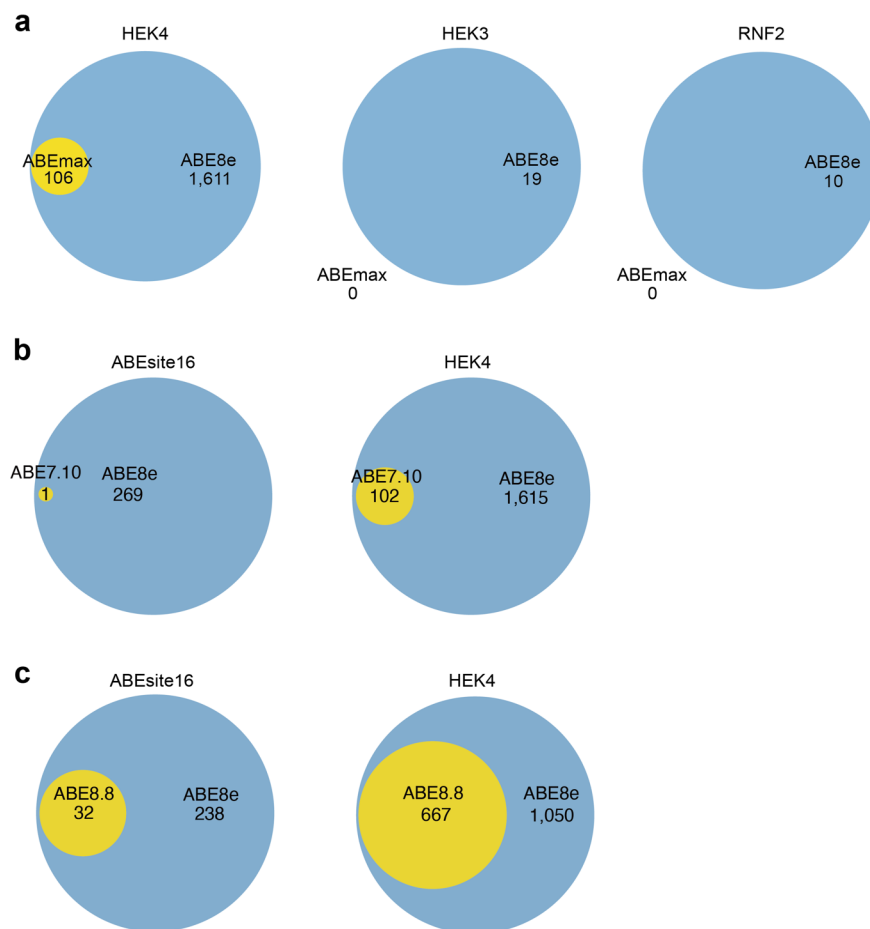
Supplementary Fig. 8 | Analysis of mismatches for off-target sequences of Cas9-based ABE.

a, Sequence logo analysis illustrates the nucleotide conservation of off-target editing regions identified in HEK293T cells transfected with ABE8e and either HEK4, ABESite16, HEK3, RNF2, CCR5sg1p, EMX1, or VEGFASite3 sgRNA, as well as in HSPCs electroporated with ABE8e and either HBG_sg-198, BCL11A_sg1619, or HBG_sg-113. **b**, Histograms depict the count distribution of mismatches and gaps in the off-target editing regions aligned to their respective on-target sequences, with the counts and proportions indicated above each group.



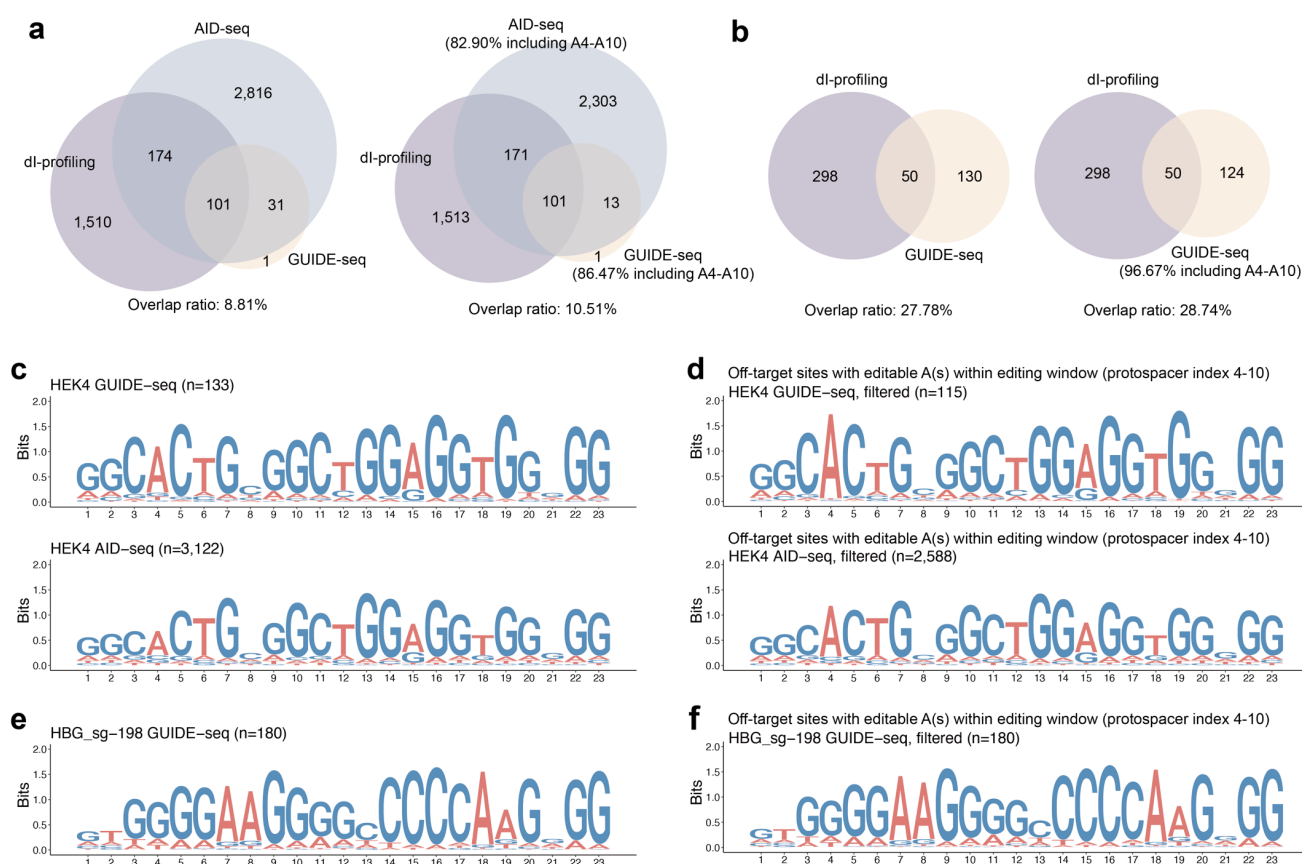
Supplementary Fig. 9 | Distribution and features of ABE-induced off-targets.

a-c, Bar plots showing the count and enrichment ratio of sgRNA-dependent off-target sites across different genomic elements. Data in **(a)** shows off-targets induced by ABE8e in HSPCs (sgRNA HBG_sg-198, BCL11A_sg1619, or HBG_sg-113) and HEK293T cells (sgRNA HEK4, ABESite16, HEK3, or RNF2); **(b)** shows off-targets induced by ABE8e, ABEmax, ABE7.10, or ABE8.8 with HEK4 sgRNA in HEK293T cells; **(c)** shows the distribution in randomly selected genomic regions used as a control. **d-e**, The histone marks and open chromatin signals at sgRNA-dependent off-target sites (dark blue lines) and random regions (green lines) in HEK293T cells (**d**) or HSPCs (**e**). The top panel shows the extracted line graph of the heatmap below; the x-axis represents the +/-5kb distance centered the off-target regions and y-axis shows the normalized RPKM. **f**, Rates of reads with A-to-G or T-to-C mutations within promoter region (defined as 2 kb upstream and 1 kb downstream of the TSS) or the whole genome, following transfection of ABE8e with or without sgRNA in HEK293T. For each bar, n = 2 biologically independent samples were used.



Supplementary Fig. 10 | ABE8e generates a superset of off-target sites compared to ABEmax, ABE7.10 and ABE8.8.

a, Venn diagrams illustrating the overlaps and inclusion relationships of the off-target sites generated by ABEmax and ABE8e with three sgRNA (HEK4, HEK3, and RNF2) in HEK293T cells. Notably, no off-target sites were detected in samples transfected with ABEmax using either the HEK3 or RNF2 sgRNA. **b, c**, Venn diagrams of off-target sites generated by ABE7.10 (**b**) or ABE8.8 (**c**) and ABE8e with sgRNA HEK4 or ABEsite16 in HEK293T cells. In all panels, blue circles represent off-target sites induced by ABE8e, while yellow circles represent those induced by other ABE versions.

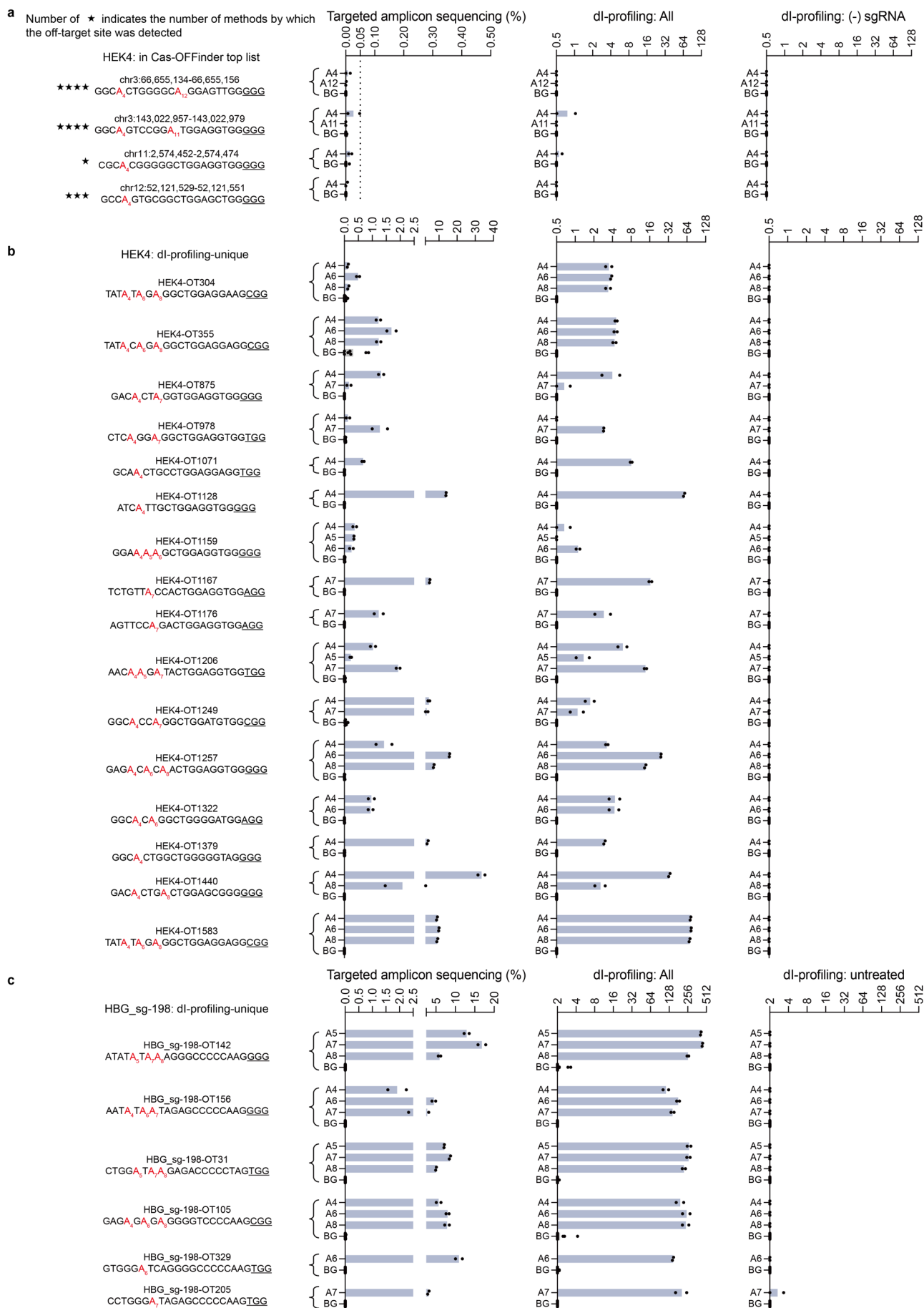


Supplementary Fig. 11 | Comparison of dI-profiling-identified ABE8e off-targets with canonical Cas9 nuclease off-targets.

a, b, Venn diagrams comparing off-target sites identified by dI-profiling for ABE8e with Cas9 nuclease off-target sites identified by GUIDE-seq or AID-seq for HEK4 (**a**) and HBG_sg-198 (**b**) sgRNAs. For HEK4, the dI-profiling off-target set used here includes the 1,717 sites identified in ABE8e-edited HEK293T cells plus 68 additional sites identified in FACS-sorted ABE8e-edited HEK293T cells (n = 1,785; see Supplementary Table 2). Left, overlap using all GUIDE-seq or AID-seq off-target sites; right, overlap restricted to GUIDE-seq or AID-seq off-target sites containing at least one editable adenine (A) within the canonical editing window (protospacer positions 4-10). The overlap ratios represent the fraction of GUIDE-seq or AID-seq off-target sites that are also captured by dI-profiling.

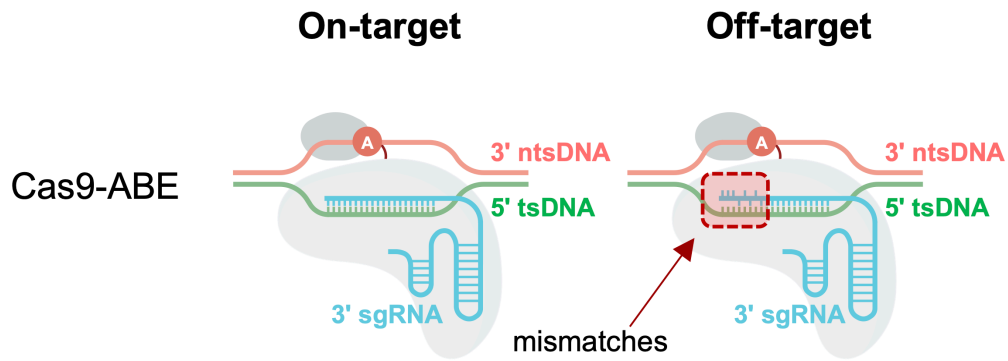
c-f, Sequence logos showing off-target sequences identified by GUIDE-seq or AID-seq before (**c, e**) and after (**d, f**) filtering for editable A(s) within the editing window. Although the filtering increased the overlap ratio of off-target sites among AID-seq, GUIDE-seq and dI-profiling, the resulting nuclease off-target landscapes still differ substantially from the Cas9-ABE off-targets observed in cellular

contexts (Supplementary Fig. 8a).



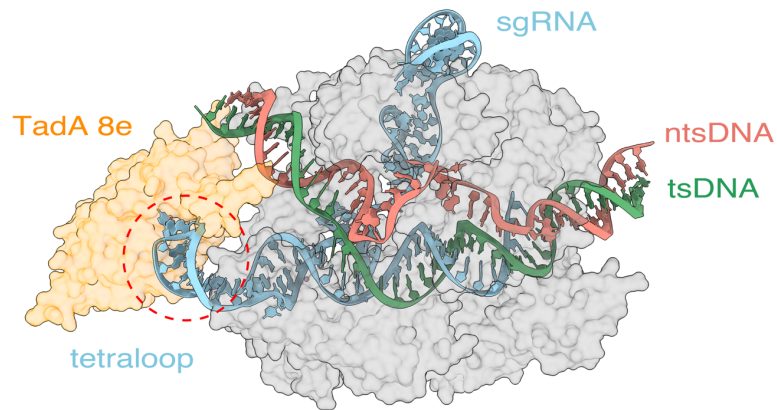
Supplementary Fig. 12 | Validation of additional off-target sites nominated by dI-profiling and other off-target detection methods.

a, Validation of HEK4 off-target sites in HEK293T cells among the top-ranked candidates from Cas-OFFinder that were not identified by dI-profiling. The number of asterisks indicates the number of methods by which the off-target site was detected. The dashed line indicates the threshold significantly higher than background noise. **b**, Validation of HEK4 off-target sites in HEK293T cells identified uniquely by dI-profiling. **c**, Validation of HBG_sg-198 off-target sites in HSPCs identified uniquely by dI-profiling. Left, A:T-to-G:C editing rates measured by targeted amplicon sequencing in samples treated with ABE8e and the corresponding sgRNA. Middle, dI-profiling signals in samples treated with ABE8e and the corresponding sgRNA. Right, dI-profiling signals in samples treated with ABE8e alone (without sgRNA) or in untreated samples. The subscripted red A denotes its position within the protospacer and the PAM sequence is underlined. Bars indicate the mean of independent biological samples (n = 2). BG, background.



Supplementary Fig. 13 | On-target and off-target complex schematics for Cas9-based ABE.

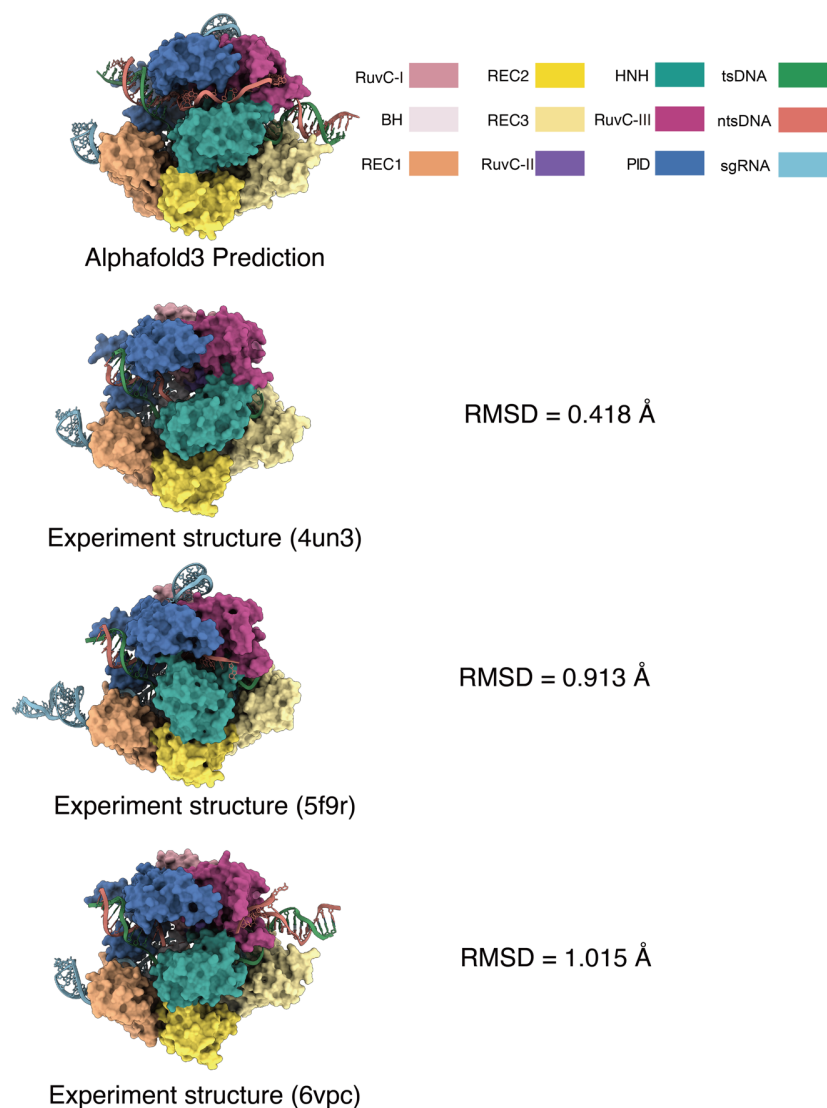
The off-target complex of Cas9-based ABE differs from the on-target complex by the presence of mismatches between the tsDNA and the sgRNA, highlighted by the red dashed box.



AF3-predicted ABE8e complex structure

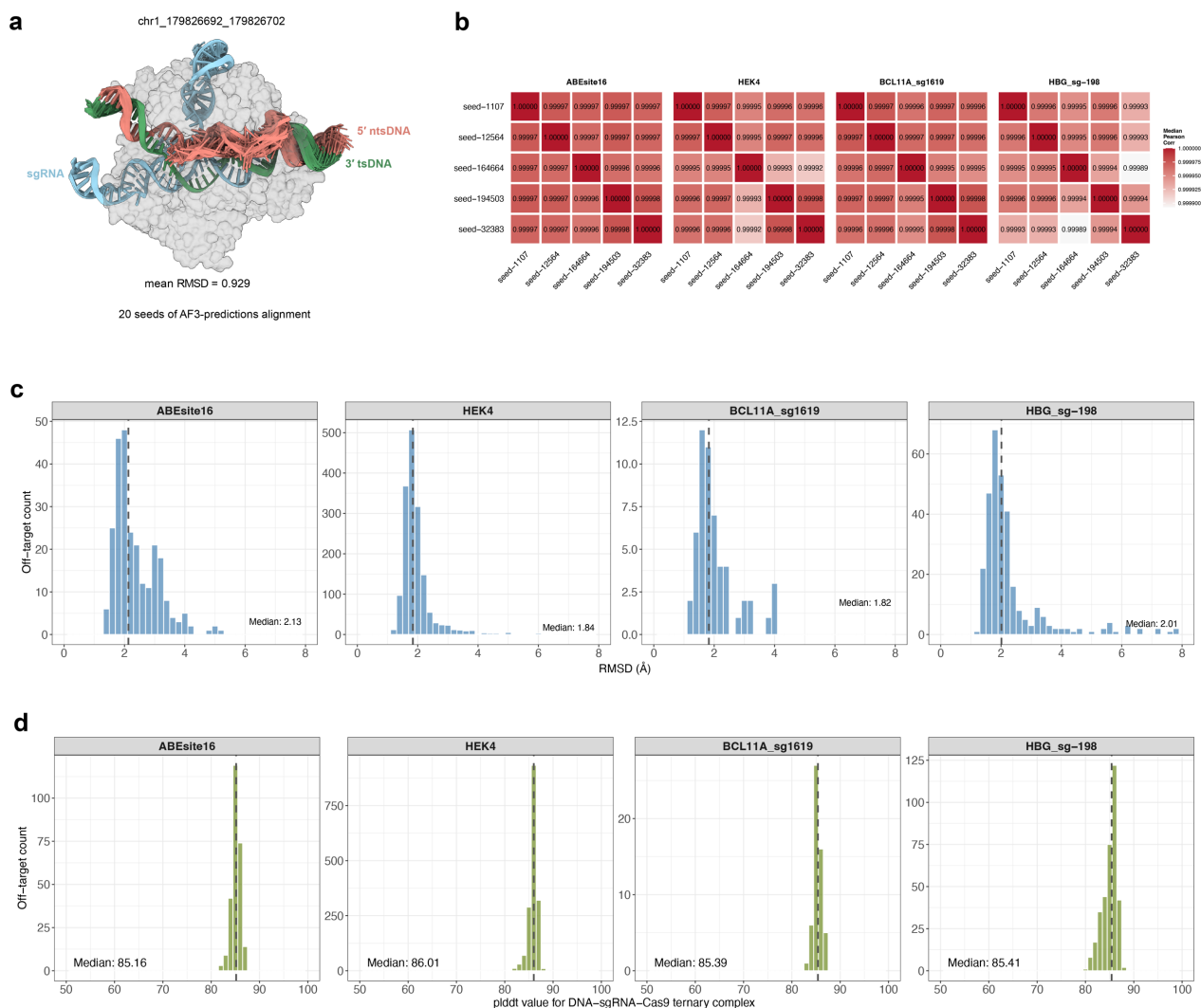
Supplementary Fig. 14 | AF3 prediction of the full-length ABE8e complex.

When predicting the complex structure of full-length ABE8e, AF3 erroneously inferred that TadA8e recognizes the tetraloop on the sgRNA (denoted by the red-dashed circle), an artifact likely arising from model hallucination in AF3.



Supplementary Fig. 15 | Alignment of AF3-predicted Cas9-DNA-sgRNA complex structure to experimental determined structures.

Comparisons of the AF3-predicted Cas9-DNA-sgRNA ternary complex structure with experimentally determined structures (PDB: 4un3, 5f9r and 6vpc, respectively). The root mean square deviation (RMSD) between the AF3-predicted structure and each experimentally determined structure were calculated accordingly (only paired residues are considered). Colors annotate the Cas9 domains and nucleotide components.



Supplementary Fig. 16 | High reproducibility and structural convergence of AF3 predictions for Cas9-DNA-sgRNA complex.

a, Overlay of 20 independently predicted Cas9-DNA-sgRNA ternary complex structures at the ABESite16 on-target site. These structures were generated by AF3 using identical inputs but with 20 different initialization seeds (Methods). The resulting complexes were superimposed, and the mean RMSD across the 20 predictions is shown below. **b**, Heatmap illustrating the median Pearson correlation coefficients of CPs from various Cas9 complex predictions. Predictions were generated using five different AF3 initial seed values (1107, 12564, 164664, 194503, and 32383) across four loci: ABESite16, HEK4, BCL11A_sg1619, and HBQ_sg-198. **c**, Histograms display the distribution of mean RMSD values for each off-target ternary complex, each predicted independently five times using five different initialization seeds. Median values are indicated by black dashed lines. The RMSD and

CP correlations in **(a)-(c)** were calculated based on the entire complex. **d**, Histograms show the distribution of pLDDT values for each AF3-predicted off-target ternary complex, with median values indicated by black dashed lines.

a

Constant sequence:

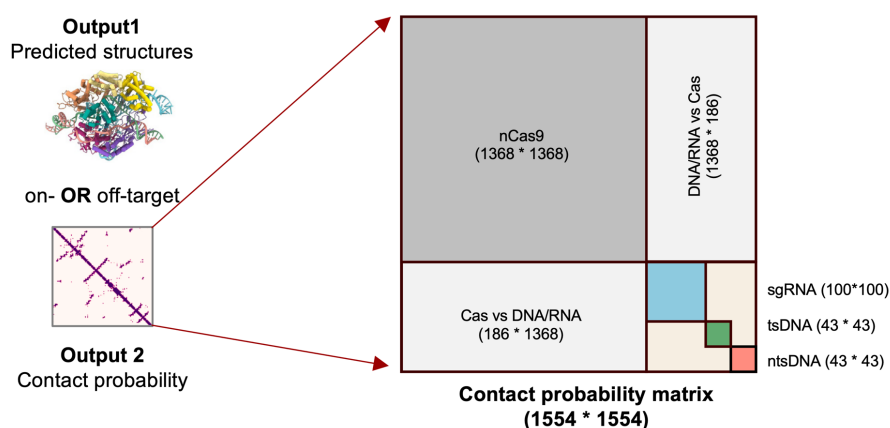
Chain A: nCas9 length = 1368
Chain B: sgRNA length = 100 (spacer 20nt + backbone 80nt)

Variable sequence:

Chain C: tsDNA length = 43 (sgRNA pairing region 20nt + 5' extend 13nt + 3' extend 10nt)
Chain D: ntsDNA length = 43 (reverse complementary of tsDNA)

Total token number = 1368 + 100 + 43 + 43 = 1554

b



Supplementary Fig. 17 | AF3 inputs and outputs for ternary complex prediction.

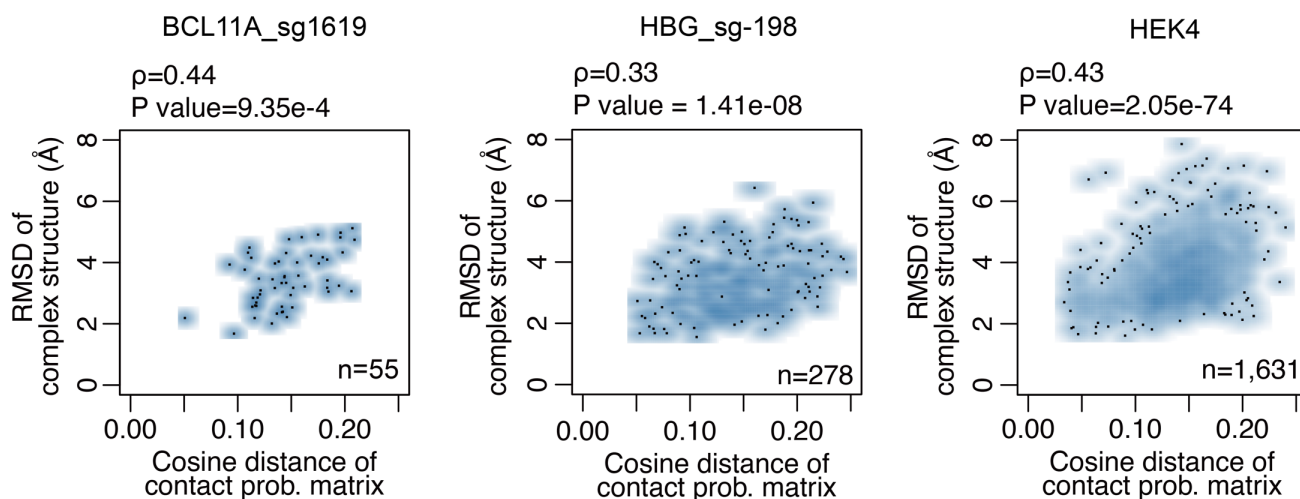
a, Illustration of the four inputs provided to AF3 for each ternary complex. The Cas9-D10A nickase (nCas9) protein sequence and the sgRNA sequence were kept constant between the on- and off-targets, whereas the sequences of the tsDNA and ntsDNA varied depending on the specific off-target site. The lengths of each input component are indicated. **b**, Two outputs generated by AF3: a predicted complex structure and a contact probability matrix. As illustrated in the detailed matrix representation, the contact probability information for both intra- and inter-macromolecular interactions can be derived from the matrix.



Supplementary Fig. 18 | Analysis of AF3 embedding differences against off-target mismatch position or signal strength.

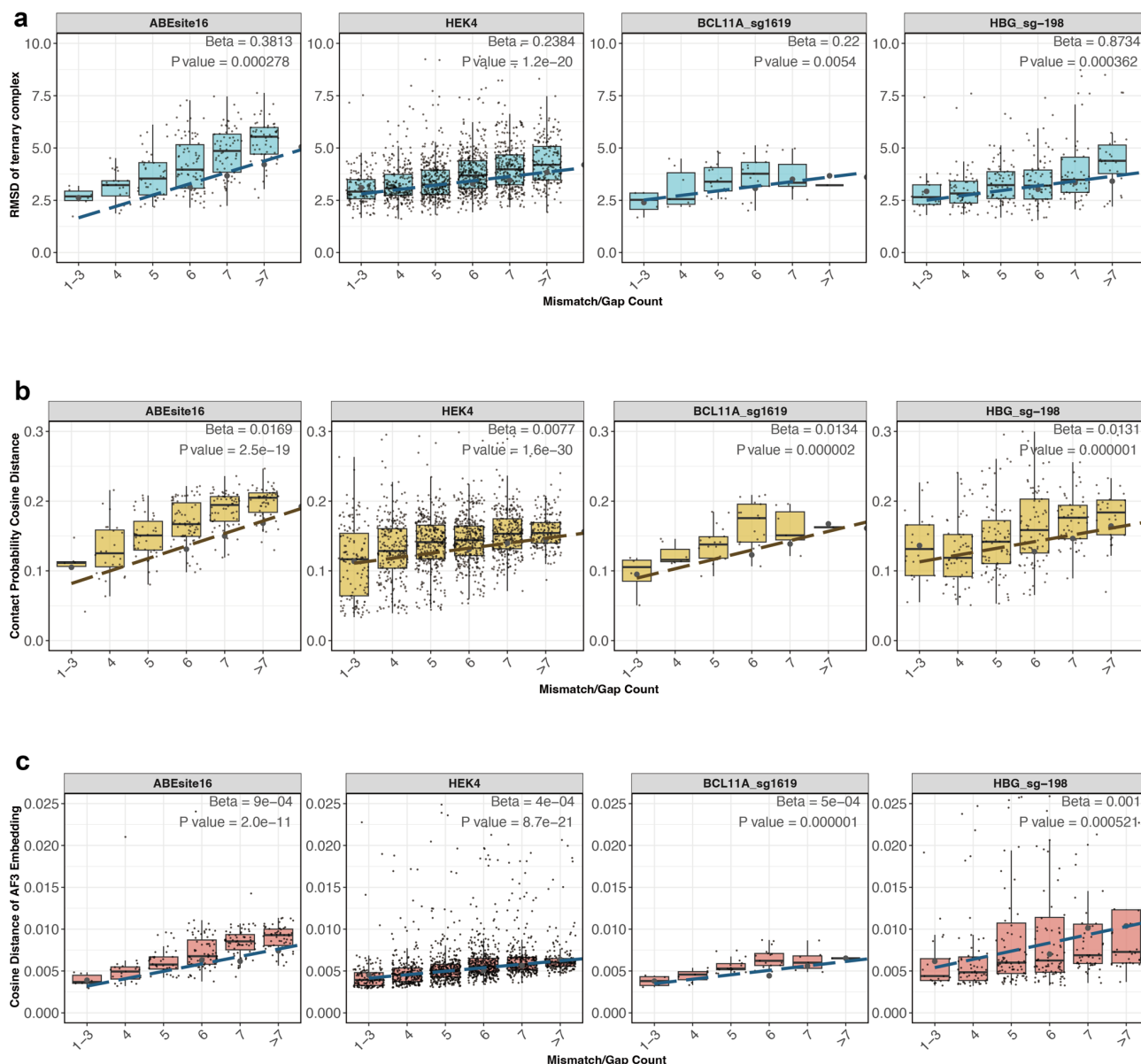
Histograms illustrate the relationship between off-targets and AF3 embedding differences (measured as $-10 \times \log_{10}(\text{cosine distance})$) across (a) two groups of mismatch positions (mismatches only in the PAM-distal region versus in both PAM-proximal and PAM-distal regions; no sites with mismatches only in the PAM-proximal region were observed) and (b) five classes of off-target signal strength (five quantile-based bins, Q1-Q5, represent off-target intensity ranges of 0-20%, 20-40%, 40-60%, 60-80%, and 80-100%, respectively). Overall, embedding differences between each off-target and corresponding on-target ternary complex show only minor variation between mismatch-location groups, whereas sites with stronger off-target signals tend to have smaller embedding differences (higher embedding similarity). Density curves were generated by Gaussian kernel density estimation, with bandwidth selected automatically according to Silverman's rule of thumb, and are plotted on the

same density scale as the histograms to enable direct visual comparison.



Supplementary Fig. 19 | Correlation analysis of the two outputs of AF3 prediction.

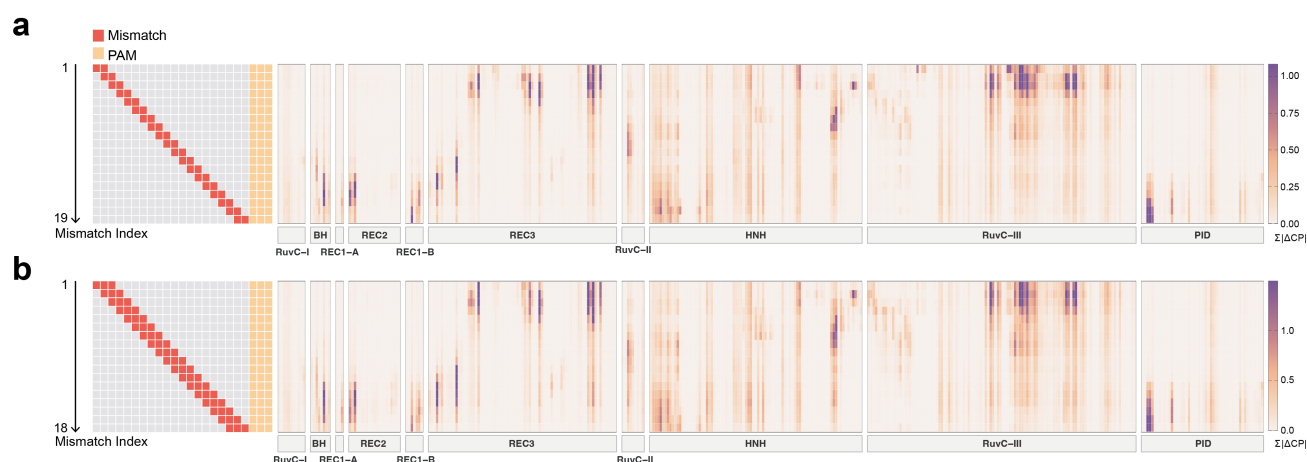
Smooth scatter plots depict alterations of AF3-predicted structures (RMSD) against differences in contact probability matrices (cosine distance) between each off-target and corresponding on-target ternary complex. The RMSD and CP difference were calculated based on the entire complex. Spearman correlation coefficient (ρ) and statistical significance P values were determined using a two-sided Spearman's rank correlation test.



Supplementary Fig. 20 | Structural, contact probability and AF3 embedding differences increase with off-target mismatch number.

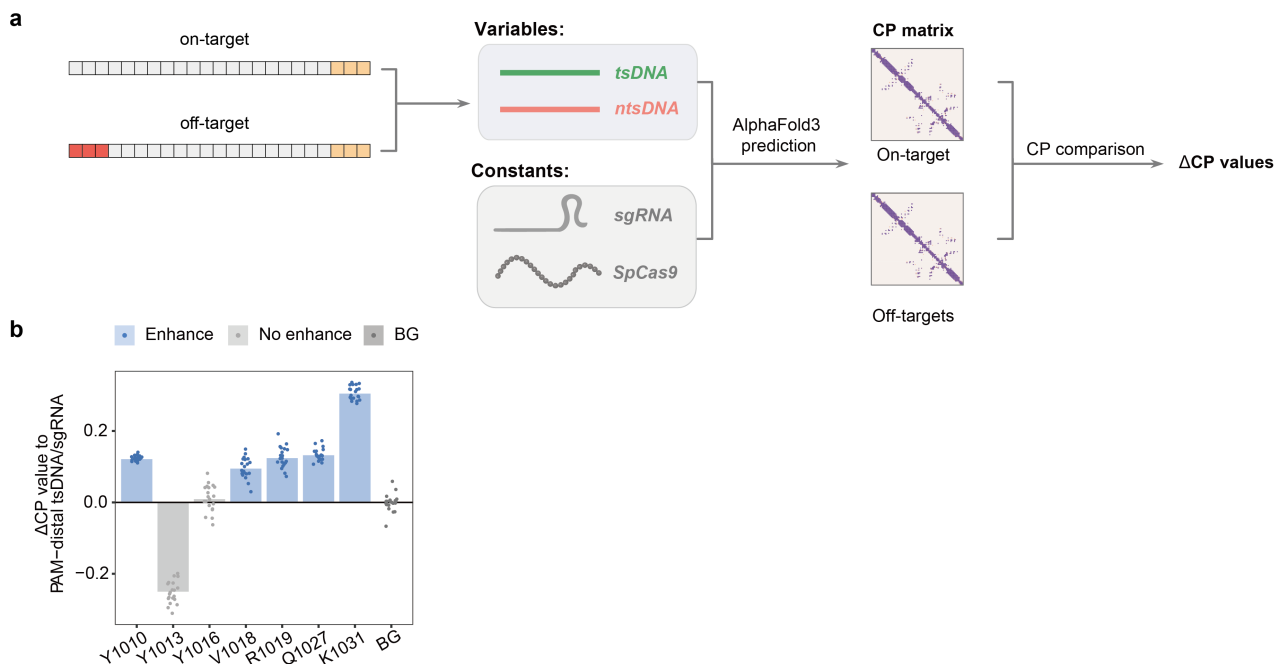
Boxplots illustrate the relationship between the number of mismatches/gaps and changes in AF3-predicted Cas9-DNA-sgRNA complexes, quantified as (a) structural deviations measured by RMSD, (b) contact-probability differences measured by cosine distance, and (c) AF3 embedding differences measured by cosine distance. Each value is computed for an off-target ternary complex relative to its corresponding on-target ternary complex. Dashed black lines represent the fitted regression models, with regression coefficients (Beta) and associated P values from the F-test indicating the statistical

significance. Dots number $n = 270, 1717, 55$ and 348 for ABESite16, HEK4, BCL11A_sg1619, and HBG_sg-198, respectively. All analyses were restricted to changes at the Cas9-tsDNA-sgRNA interface.



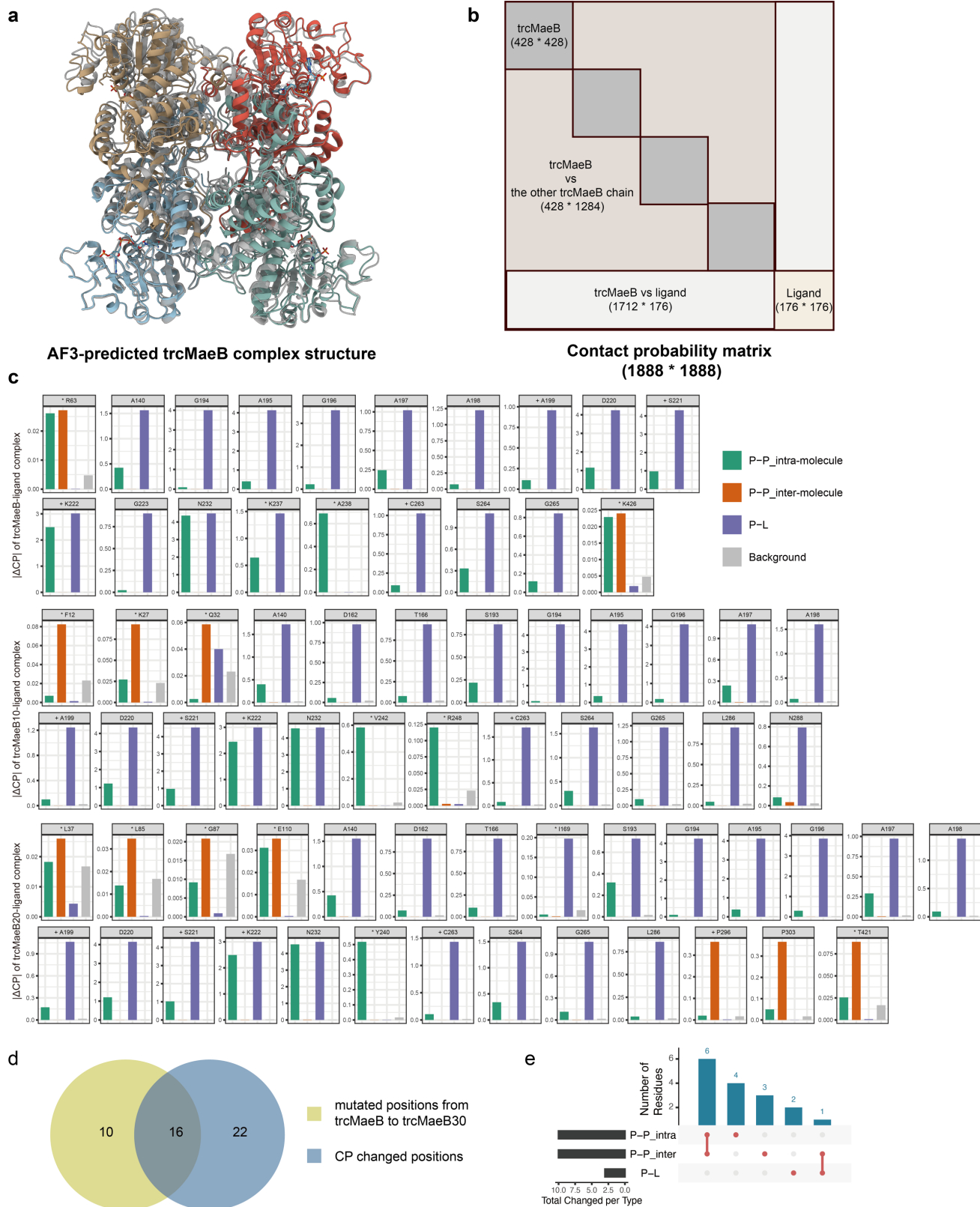
Supplementary Fig. 21 | Position-specific multi-nucleotide mismatches induce localized CP changes within different Cas9 domains.

a, b, Heatmaps showing CP changes induced by multi-nucleotide mismatches in the Cas9-DNA-sgRNA complex, computed using the same AF3-based workflow as in the single-mismatch tiling analysis (see Extended Data Fig. 2a). Target-strand (tsDNA) and non-target-strand (ntsDNA) sequences were systematically varied to introduce either double-nucleotide (2-bp; **a**) or triple-nucleotide (3-bp; **b**) mismatches within the protospacer region (positions 1-20; PAM positions 21-23). In both panels, the heatmap shows the mean cumulative $|\Delta CP|$ across 10 sgRNAs for SpCas9 (Supplementary Table 1). Each row represents a mismatch configuration (arrows indicate the tiling direction along the protospacer), and each column corresponds to a Cas9 residue. Only residues exhibiting contact with DNA/RNA are displayed. Cas9 domains are labeled below.



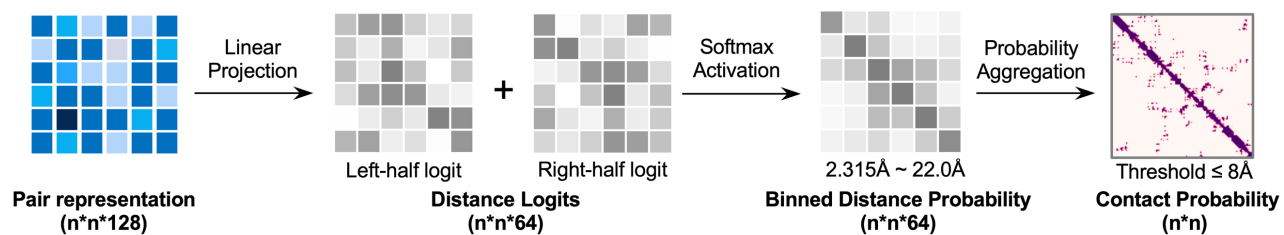
Supplementary Fig. 22 | Differential contact probability (ΔCP) analysis for key residues in SuperFi-Cas9.

a. AF3-based workflow for ΔCP analysis of key residues substituted in SuperFi-Cas9. AF3 predictions were performed using WT SpCas9 and the same nucleic acid contexts (both on-target and off-target) in the original SuperFi-Cas9 study²⁴. For each complex (on-target and off-target), 20 independent predictions were performed using different random seeds (including 1094, 1842, 7215, 9937, 11039, 29384, 59021, 64722, 258310, 427613, 448302, 682491, 906144, 3382715, 4529013, 5530182, 6291048, 7731904, 8102395, and 8815206). CP values represent the probability of residue side-chain atoms locating within ~ 5 Å of the nucleic acid strands⁴¹. For each residue, ΔCP is defined as the difference in the summed CP values for contacts with the PAM-distal sgRNA and tsDNA (protospacer positions 1-5) between the off-target and on-target complexes ($\Delta CP = CP_{\text{off-target}} - CP_{\text{on-target}}$), following the original design rationale. **b.** Bar plot showing ΔCP values between off- and on-target complexes for the seven key residues engineered in SuperFi-Cas9. Five of the seven residues exhibit increased CP to the sgRNA and tsDNA in the PAM-distal region of the off-target complex, consistent with their proposed role in stabilizing the off-target conformation. Each bar represents the mean ΔCP from independent predictions ($n = 20$). BG means background variation across AF3 predictions with random seeds.



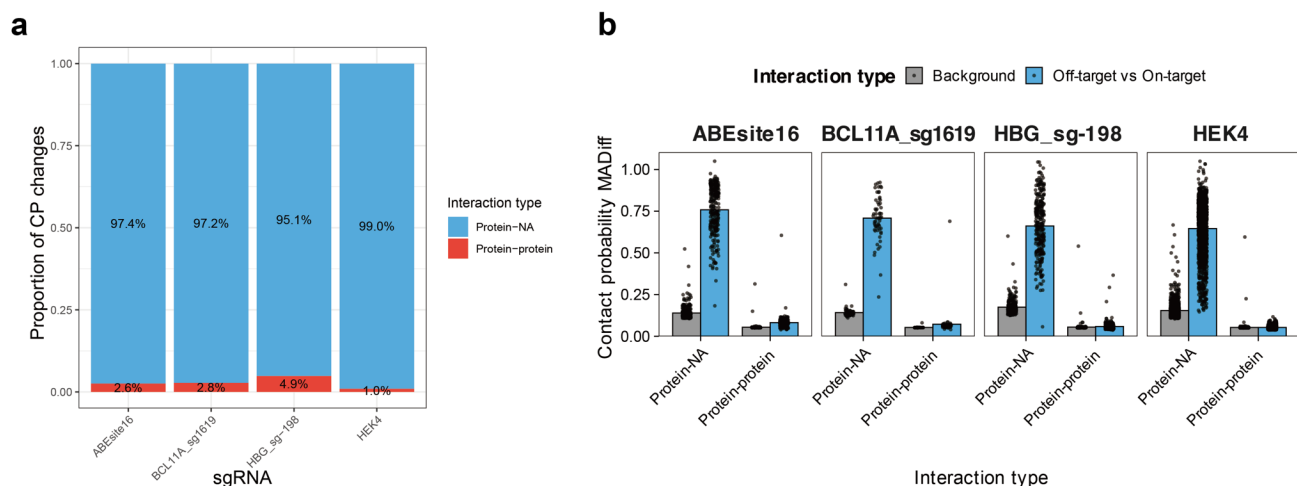
Supplementary Fig. 23 | CP-based analysis nominates residues for MaeB cofactor specificity engineering.

In the referenced MaeB engineering study⁷⁸, iterative mutagenesis from wild-type trcMaeB to trcMaeB30 shifted cofactor preference from NADP⁺ to NAD⁺. Here, we used AF3-derived contact probability (CP) to test whether CP-based analysis can recover these specificity-determining positions. **a**, AF3-predicted tetramer structure of trcMaeB enzyme, with cofactors NADP⁺ (colored) and NAD⁺ (gray). **b**, Illustration of the AF3-derived contact probability matrix for trcMaeB. From this matrix, protein-protein intra-molecular (P-P intra), protein-protein inter-molecular (P-P inter) and protein-ligand (P-L) contact probabilities can be extracted. **c**, Absolute value of ΔCP ($|\Delta\text{CP}|$) reflects the changes in interaction at corresponding mutation sites among different trcMaeB variants. The figure illustrates $|\Delta\text{CP}|$ analysis for trcMaeB variants (trcMaeB, trcMaeB10 and trcMaeB20) bound to different cofactors. For each variant, residue and class of contacts (P-P intra, P-P inter, P-L), we calculated the $|\Delta\text{CP}|$ between the NADP⁺-bound and NAD⁺-bound complexes and compared these values to background distributions generated using different AF3 seeds. The y-axis represents the $|\Delta\text{CP}|$ value for each residue, while the x-axis indicates the different types of contact probabilities. Asterisks (*) indicate mutations accumulated from the wild-type trcMaeB to the trcMaeB30 variant, while plus signs (+) denote additional mutations introduced in subsequent variants derived from trcMaeB30. **d**, Overlap between the mutation sites in trcMaeB30 (compared to wild-type trcMaeB) and the key residues identified by contact probability analysis. **e**, UpSet plot showing the types of contact changes at the overlapping mutation sites shown in **c**. Over half of these residues display altered P-P contacts but no significant changes in P-L contacts. This is consistent with the original study, which reported that more than half of mutations are located away from the substrate pocket and likely act by modulating multimer formation and inducing conformational changes. These results suggest the potential of CP-based analysis in identifying amino acid sites that alter enzyme's ligand specificity.



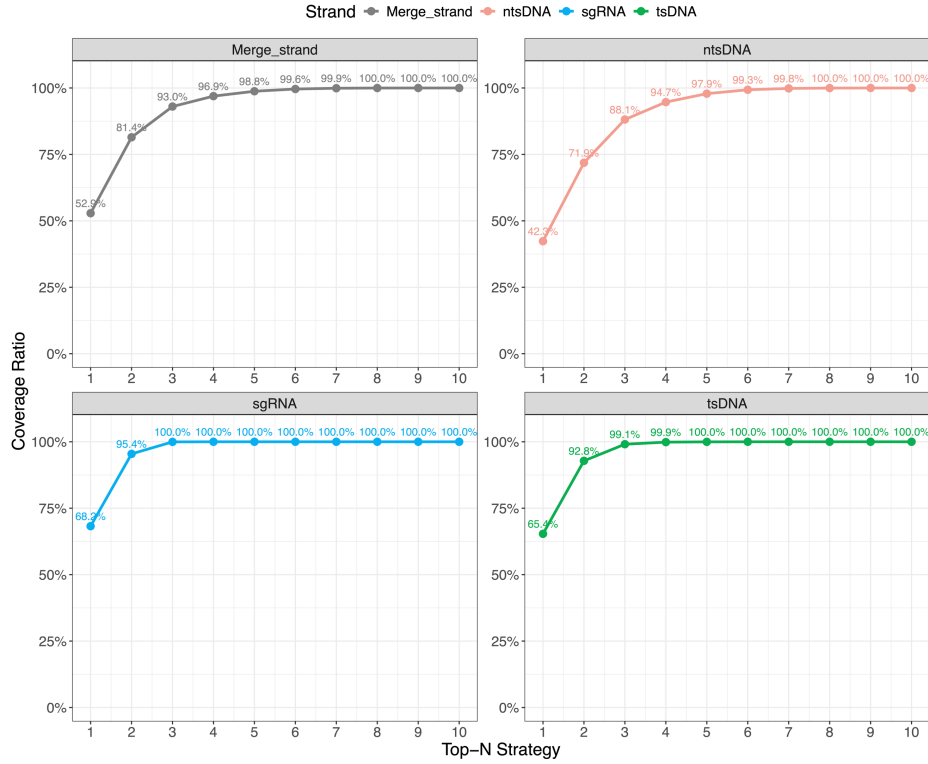
Supplementary Fig. 24 | Schematic illustration of contact probability calculation steps.

After input sequence processing and multiple sequence alignment (MSA), AF3 utilizes its Pairformer trunk to generate a pair representation, producing an $n \times n \times 128$ -dimensional embedding (where n denotes the number of tokens). This embedding is then subjected to a linear projection to predict spatial distances, which are subsequently normalized to yield the final distogram and contact probability values (defined using an 8 Å threshold). Because these downstream operations involve only linear transformations and symmetric normalization, the pair representation embedding itself is the primary determinant of the final contact probability output.



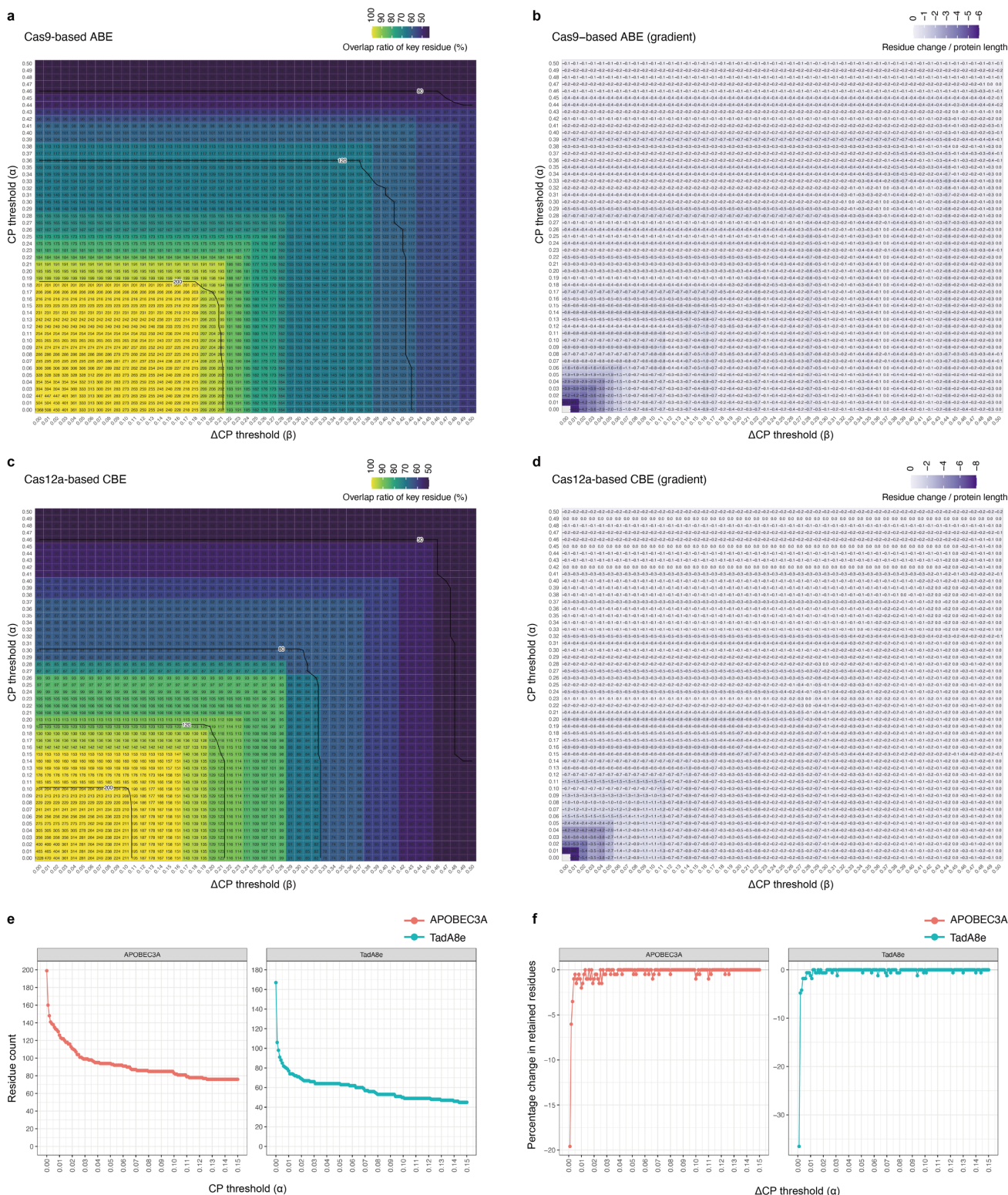
Supplementary Fig. 25 | The majority of CP differences are localized on the protein-NA interfaces.

a, The proportion of CP changes attributed to protein-protein versus protein-nucleic acid (protein-DNA/RNA, protein-NA) interactions. CP matrices from AF3-modeled on-target and off-target complexes for multiple sgRNAs were compared, and contact changes were partitioned by the interaction type. When excluding nucleic acid-nucleic acid (NA-NA) interactions, protein-nucleic acid contacts account for the vast majority of the observed CP changes. **b**, Comparison of CP changes across four individual sgRNA relative to background variations. Each bar represents the mean of CP MADiff value. The background variations were calculated based on the differences between AF3 predictions generated with different initial 5 seeds (Methods). Dots number $n = 267, 1631, 55$, and 283 for ABEsite16, HEK4, BCL11A_sg1619, and HBG_sg-198, respectively.



Supplementary Fig. 26 | Coverage of contact probability extraction using Top-N entries.

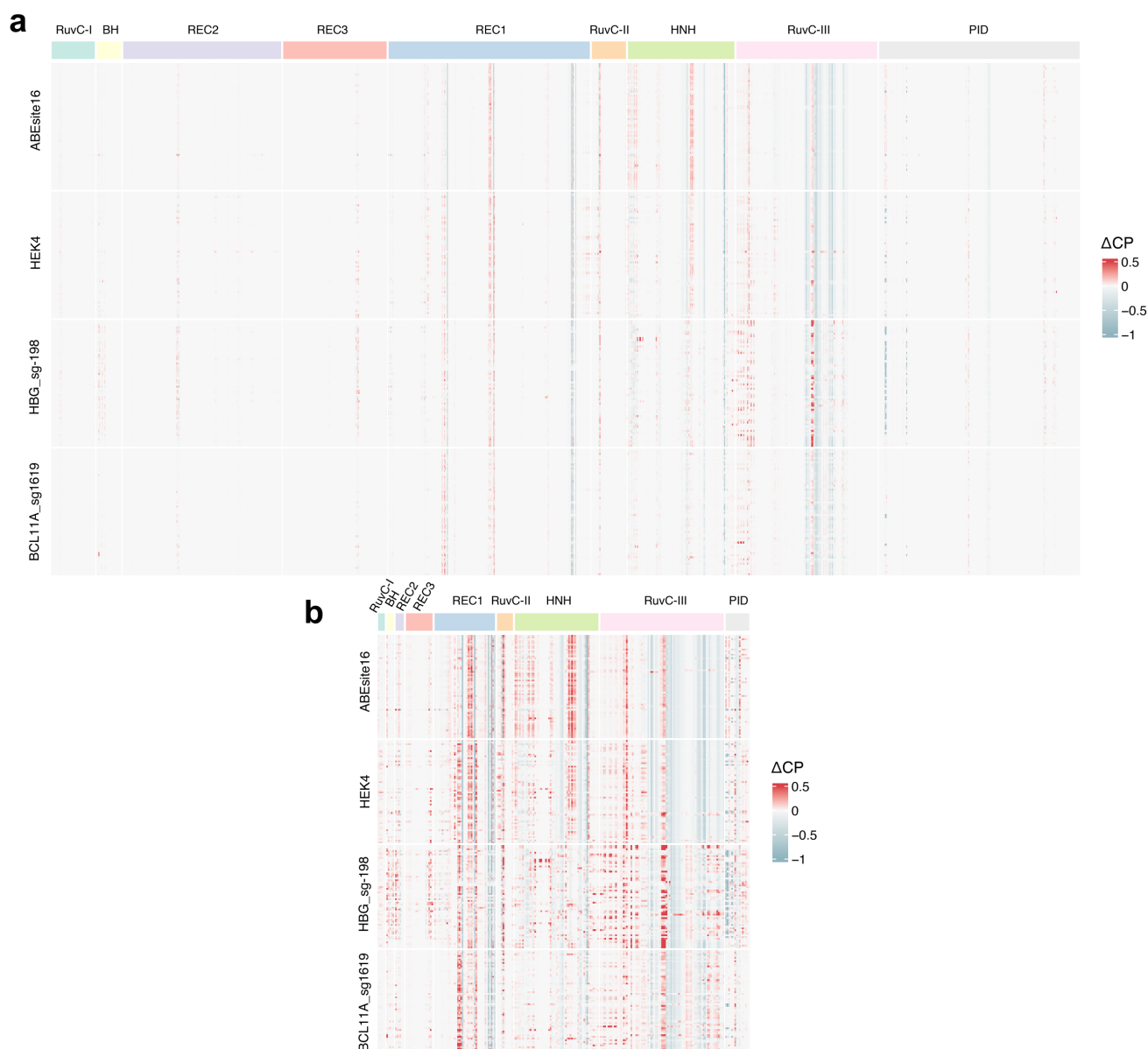
The Top-N contact probability values for each Cas9 residue with sgRNA, ntsDNA or tsDNA nucleotides were extracted separately. Coverage ratios for sgRNA, ntsDNA, tsDNA and their combined set (“merge_strand”) were defined as the proportion of the extracted contact probability to all contact probability with values greater than 0.1.



Supplementary Fig. 27 | Systematic evaluation of the impact of thresholds on identifying amino acid residues that contact nucleic acid in four complexes.

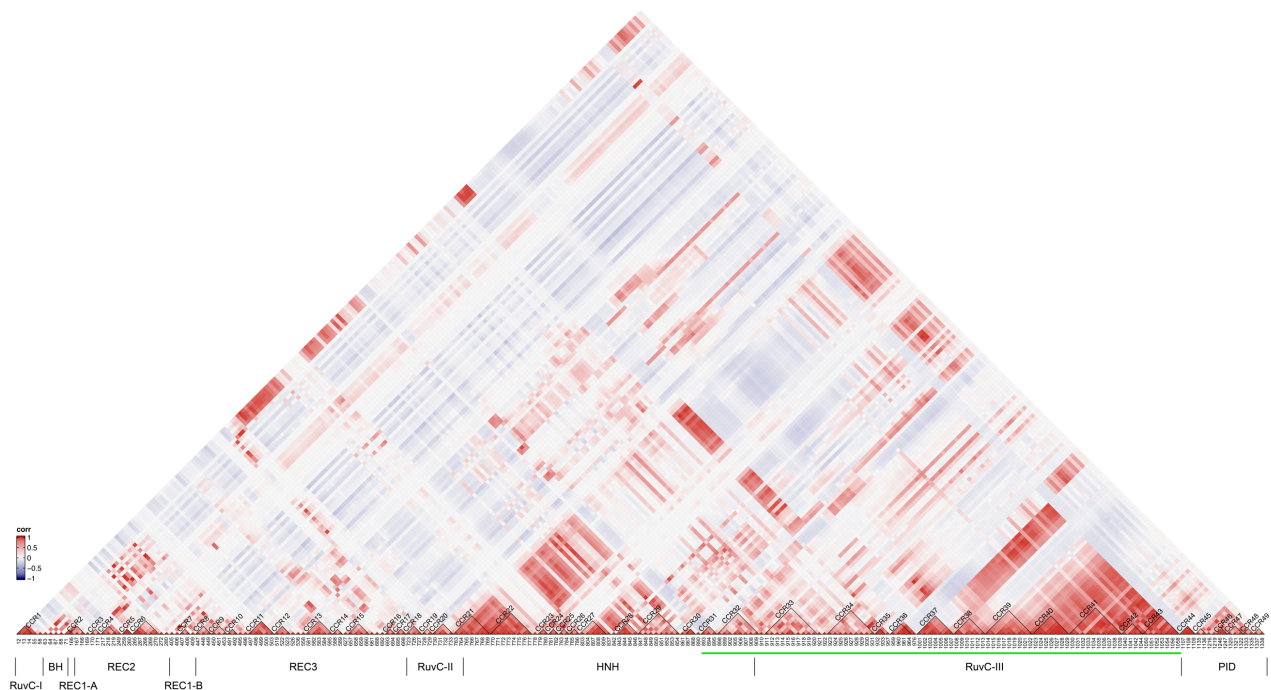
a, c, Two-dimensional heatmaps of the residue selection for (a) Cas9-based ABE and (c) Cas12a-based CBE as a function of the CP threshold (α ; y-axis) and Δ CP threshold (β ; x-axis), each swept from 0 to

0.5 in 0.01 step. Numbers inside each tile give the count of residues retained by the corresponding (α , β) pair. Tile color encodes the overlap ratio between the identified contact residues and the curated key-residue set (Cas9: $n = 21$ key residues; Cas12a: $n = 20$ key residues, as listed in Supplementary Table 4), expressed as a percentage. Black contour lines mark iso-count levels (20, 30, 50, 80, 120 and 200 residues) to facilitate comparison of residue retention across parameter space. **b**, **d**, Gradient heatmaps showing the changes of the retained residue count to threshold changes for **(b)** Cas9-based ABE and **(d)** Cas12a-based CBE on the same (α , β) grid as in **a** and **c**. For each tile, the change in retained residue count was computed relative to its lower-left neighbor (i.e., the tile at the next lower α and β); for edge tiles lacking a lower-left neighbor, the left or bottom neighbor was used instead (priority: lower-left $\rightarrow$ left $\rightarrow$ bottom). The (0, 0) tile has no reference and is shown in grey. Changes were normalized by the total protein length (Cas9: 1,368 aa; Cas12a: 1,228 aa) to express the gradient as a percentage of the full sequence. Tile color: darker purple indicates a steeper local decrease in the retained residue count and therefore a less stable region of parameter space; near-white tiles indicate stable regions where small threshold changes minimally alter the residue selection. Numbers inside each tile report the percentage change. **e**, Number of retained residues for APOBEC3A (left) and TadA8e (right) as a function of the CP threshold α (0 to 0.15 in 0.001 step). **f**, Corresponding gradient curves for APOBEC3A and TadA8e: the per-step change in retained residue count, normalized by protein length (APOBEC3A: 199 aa; TadA8e: 167 aa), is plotted against α . The grey horizontal reference line at 0 marks “no change”; values closer to zero indicate α intervals where the residue selection is stable.



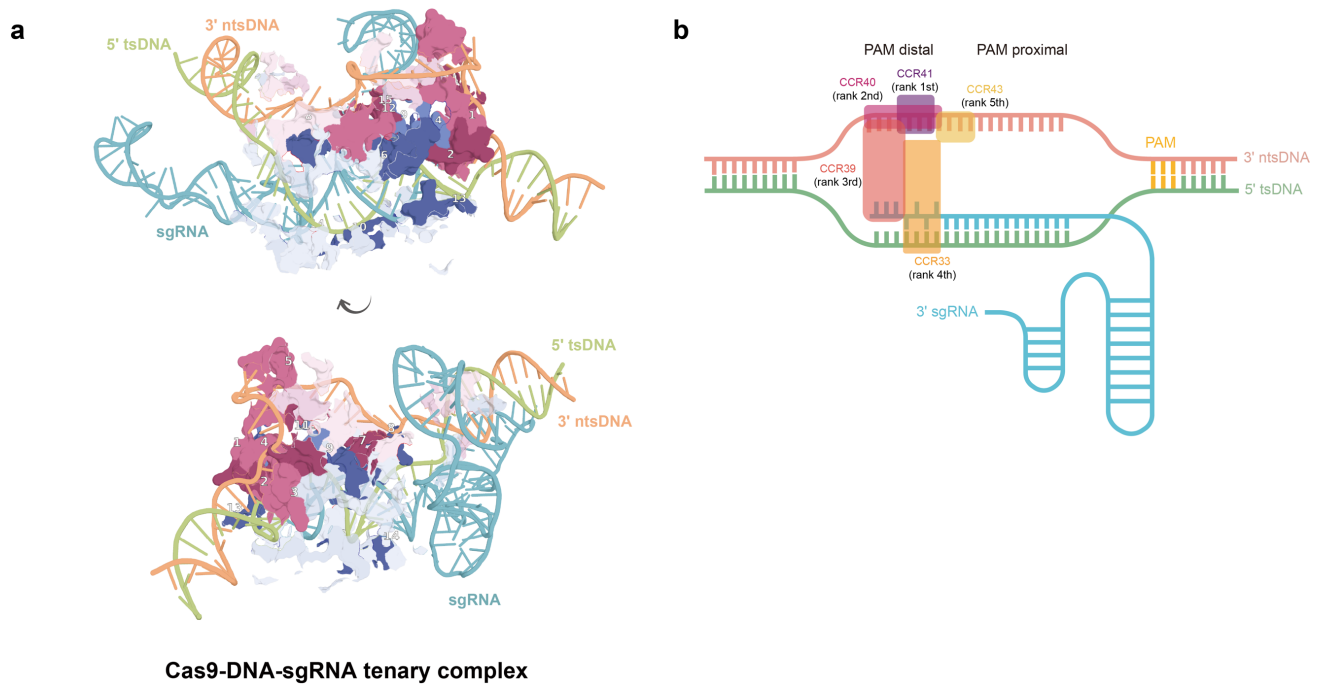
Supplementary Fig. 28 | CP alterations in off-target complexes.

a, The heatmap displays the maximum differences in protein-nucleic acid contact probabilities (ΔCP) between each off-target ternary complex and its corresponding on-target ternary complex. Each row represents a distinct off-target complex and each column represents a Cas9 residue. **b**, Cas9 residues showing evident ΔCP were extracted from panel **a**.



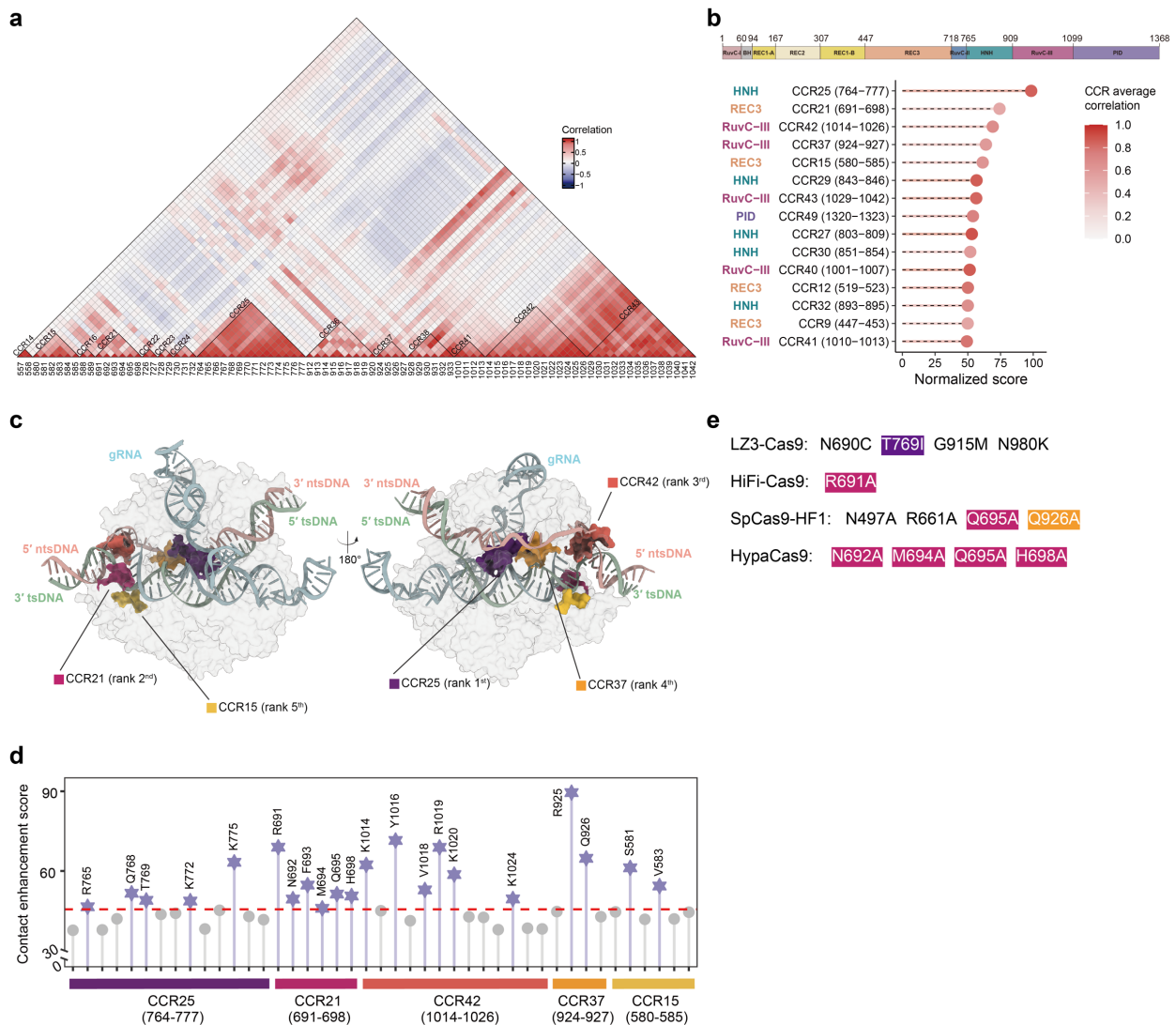
Supplementary Fig. 29 | CCRs of Cas9-based ABE across residues.

Heatmap of CP correlations among residues in contact with DNA/RNA across the Cas9-based ABE off-target training dataset (from sgRNA ABESite16, HEK4, HBG_sg-198, or BCL11A_sg1619; $n = 2,231$), revealing clustered CCRs. The green line under the residue indices delineates the zoomed-in regions shown in Fig. 4b, and Cas9 domain annotations are indicated below the heatmap.



Supplementary Fig. 30 | Spatial distribution of Cas9 CCRs from Cas9-based ABEs mapped onto the Cas9-DNA-sgRNA ternary off-target complexes.

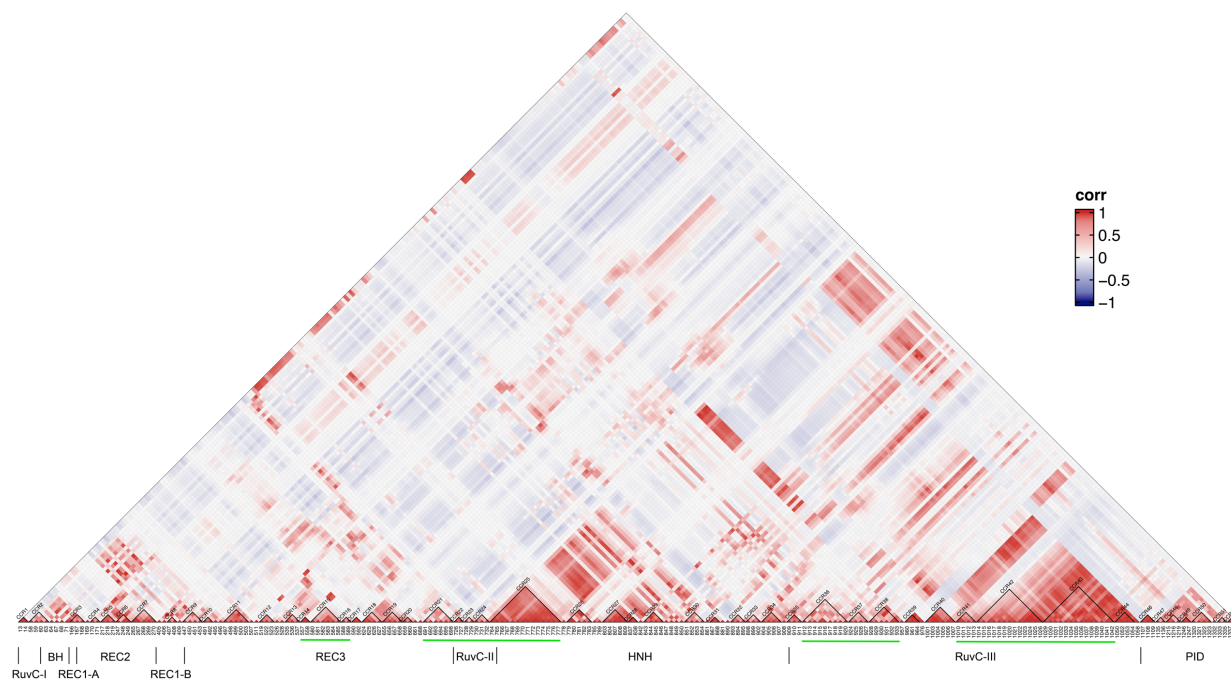
a, Structural mapping of the CCRs of Cas9-based ABE on the Cas9-DNA-sgRNA complexes. The guide RNA (sgRNA) is colored in light blue, the target strand DNA (tsDNA) in light green, and the non-target strand DNA (ntsDNA) in light orange. CCRs located at the Cas9-tsDNA-gRNA interfaces are colored in the blue spectrum, whereas those located at the Cas9-ntsDNA interfaces are colored in the red spectrum. The top 15 high-ranking CCRs are highlighted as opaque surfaces and numerically labeled (1-15) with dark-outlined white text, while the remaining lower-ranked CCRs are rendered with 25% transparency. The bottom panel show an alternative view rotated as indicated by the arrow, allowing optimal visualization of the hidden CCRs in the three-dimensional architecture. The spatial assignment of each CCR to the respective macromolecular interfaces was determined algorithmically: for each CCR, the atomic proximity to the ntsDNA and tsDNA-sgRNA heteroduplex was calculated using expanding distance thresholds (ranging from 4 Å to 100 Å) in PyMOL; a CCR was definitively assigned to the interface with which it shared the highest number of proximal atoms, ensuring an objective, structurally rigorous classification. **b**, Schematic representation of the positions and interactions of the top 5 CCRs relative to the R-loop in the Cas9-DNA-sgRNA complex.



Supplementary Fig. 31 | ContactSeek framework identifies specificity-enhancing residues for Cas9 nuclease.

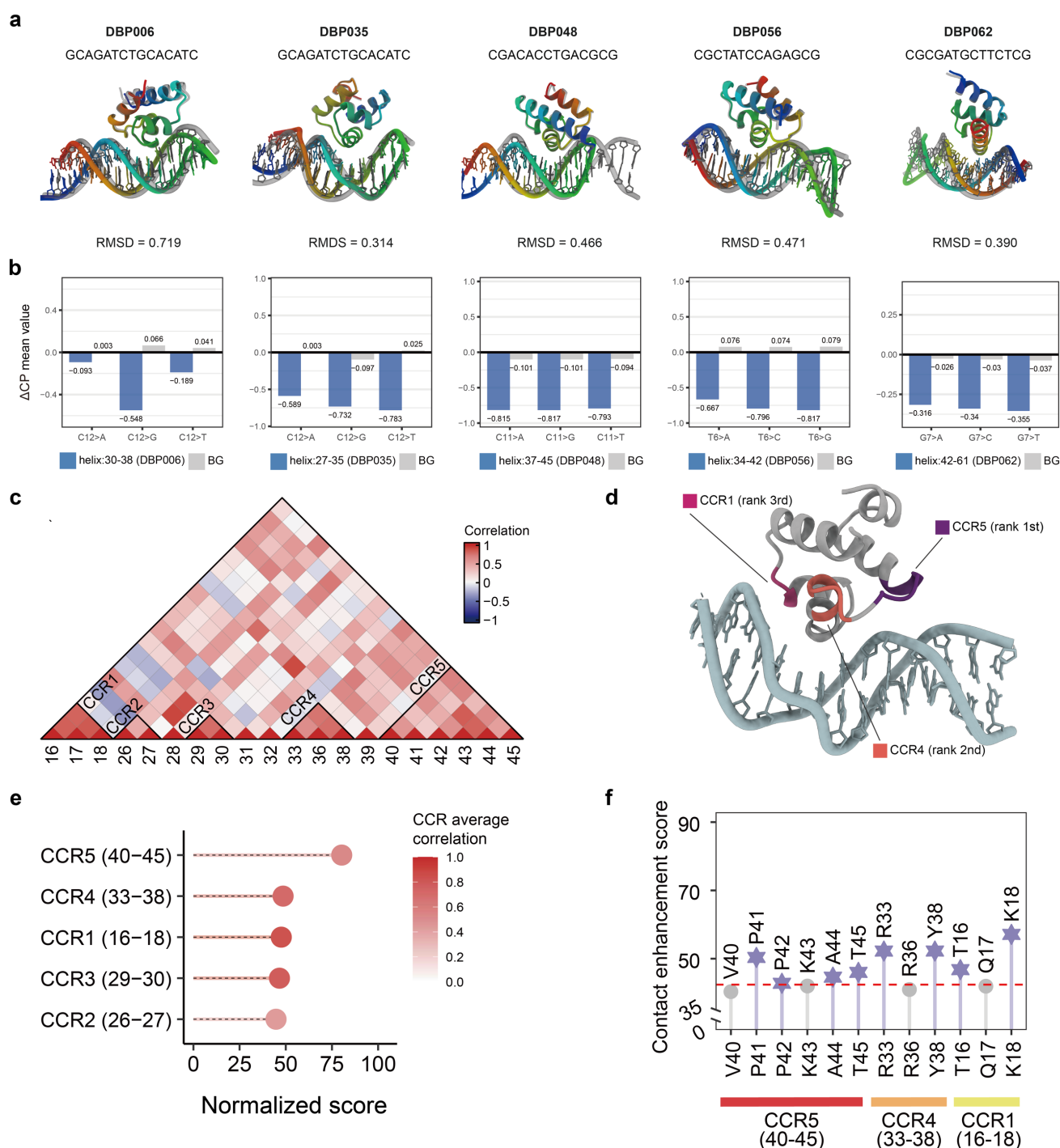
a, Heatmap of CP correlations among Cas9 residues 557-1042 across the Cas9 nuclease off-target training datasets, revealing clustered CCRs. For CP correlations across all Cas9 residues in contact with DNA/sgRNA, see Supplementary Fig. 32. **b**, CCRs ranked by ContactSeek output score (x-axis) and mapped onto Cas9 domains annotated at left. The color depth of the dot indicates the average Pearson correlation between residues within each CCR. **c**, 3D views of the top 5 CCRs (colored) in an AF3-predicted off-target complex. **d**, Contact enhancement score for each residue within the top 5 CCRs. Purple stars indicate residues with contact enhancement above the cutoff; red dashed line marks the cutoff. **e**, Overlap of key residues identified by ContactSeek with established high-fidelity Cas9

variants. Background shading under each residue label indicates mutation sites that fall within the top five CCRs, using the same color scheme as in **c** (each color corresponds to a distinct CCR).



Supplementary Fig. 32 | CCRs of Cas9 nuclease across Cas9 residues.

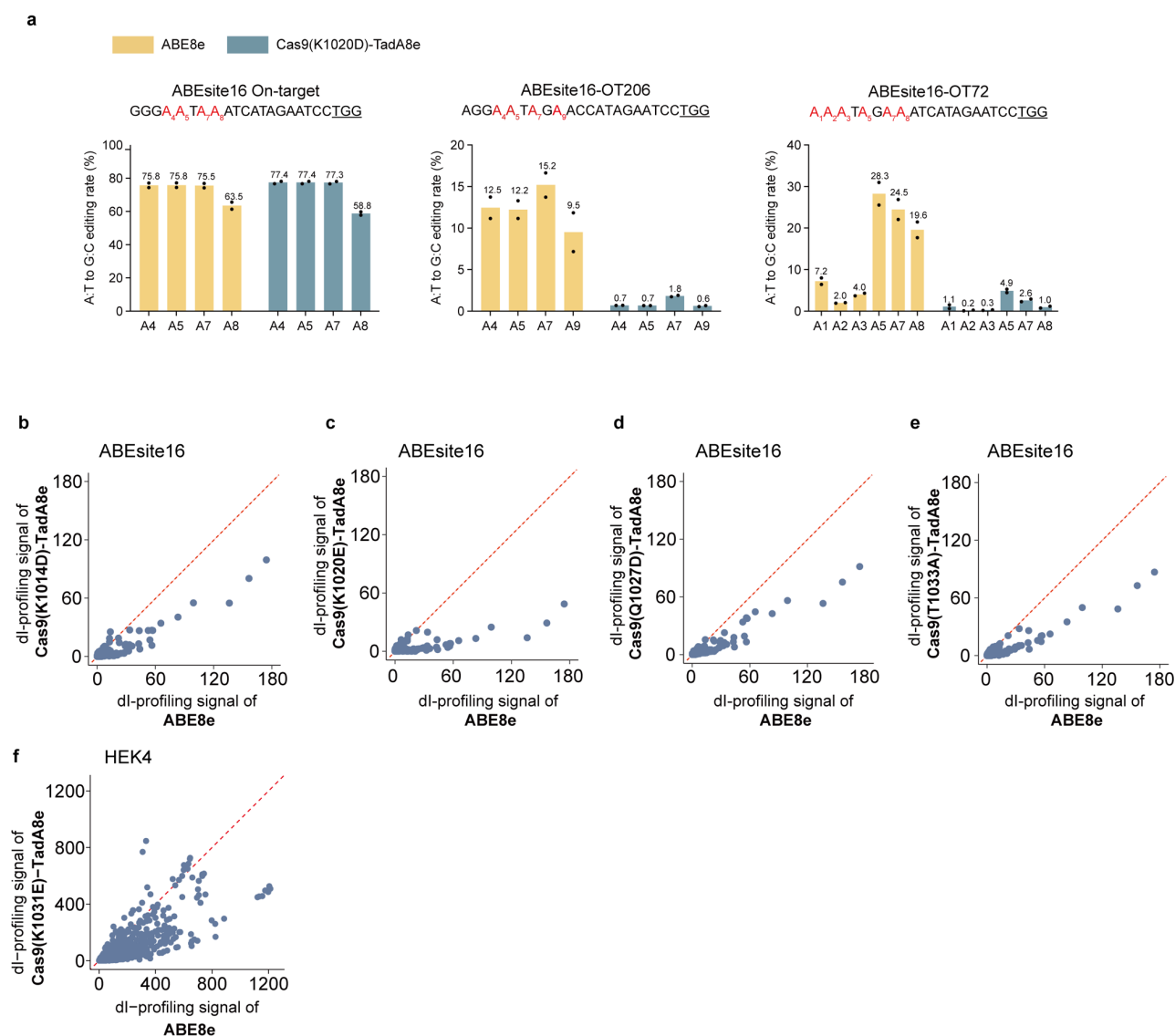
Heatmap of CP correlations among all Cas9 residues in contact with DNA/RNA across the off-target training data obtained from published GUIDE-seq datasets (from sgRNA AAVS1site13, AAVS1site14, AAVS1site2, AAVS1site4, CCR5site8, CTLA4site1, CTLA4site9, CXCR4site3, EMX1, FANCF, HEK1, HEK4, LAG3site9, PDCD1site10, PDCD1site11, PDCD1site7, PDCD1site8, PTPN6site1, TRACsite1, TRACsite2, TRACsite4, VEGFAsite1, VEGFAsite2, VEGFAsite3; $n = 1,652$), revealing clustered CCRs. The green lines under the residue indices delineate the zoomed-in regions shown in Supplementary Fig. 31a, and Cas9 domain annotations are indicated below the heatmap.



Supplementary Fig. 33 | Applying ContactSeek to AI-designed DNA binding protein (DBP).

a, The DBP-DNA complex structures predicted by AF3 (gray) shows high similarity to the AI-designed structures (colored). The figure shows the superpositions of the crystal structures of five DBPs with their corresponding AF3 predicted structures. The upper section lists the DBP names along with their intended DNA targets; middle section provides representation of the structure overlay results; the lower section presents the RMSD values between AF3 predictions and AI-designed structures (all residues

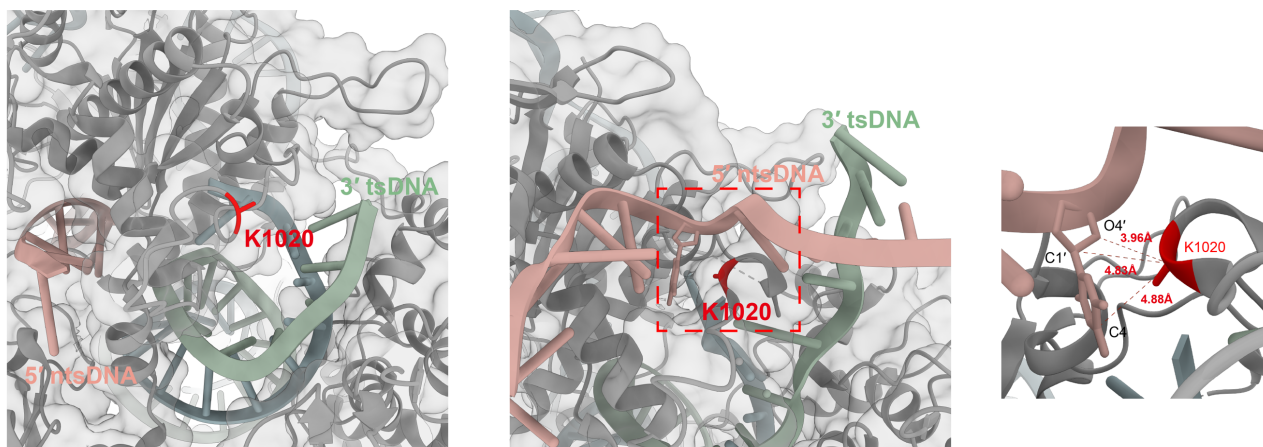
were considered). **b**, CP sensitively captures the reduction in the binding affinity of DBPs for non-target DNA sequences. The y-axis represents the mean Δ CP value (CP_mismatched - CP_target) for DBP-DNA contacts from the designed DNA-binding helix or non-designed region; the x-axis indicates the base substitutions at the designed positions. The background denotes the level of CP changes at non-designed helix region. **c**, Heatmap of CP correlations among DBP35 residues with non-zero CP value to DNA across the off-target training dataset (n = 42 single-mismatch complexes), revealing clustered CCRs. **d**, A 3D view of the top 3 CCRs (colored) identified by ContactSeek in an AF3-predicted DBP35-DNA complex. **e**, CCRs ranked by ContactSeek output score (x-axis). The color depth of the dot indicates the average Pearson correlation between residues within each CCR. **f**, Contact enhancement score for each residue within the top 3 CCRs. Purple stars indicate residues with contact enhancement above the cutoff; red dashed line marks the cutoff.



Supplementary Fig. 34 | Specificity improvement of ABE8e variants based on Cas9 mutations.

a, Editing rates of ABE8e and Cas9(K1020D)-TadA8e at the on-target site and two representative off-target sites of ABESite16 sgRNA in HEK293T cells. The subscripted red A denotes the edited A position within the protospacer and the PAM sequence is underlined. Each dot represents an independent biological replicate ($n = 2$); bars indicate the mean. **b-e**, Comparison of genome-wide sgRNA-dependent off-targets (evaluated by dl-profiling) for Cas9(K1014D)-TadA8e (**b**), Cas9(K1020E)-TadA8e (**c**), Cas9(Q1027D)-TadA8e (**d**) and Cas9(T1033A)-TadA8e (**e**) with those for ABE8e, using ABESite16 sgRNA ($n = 270$ sites) in HEK293T cells. The percentage reductions of off-target signals for Cas9(K1014D)-TadA8e (**b**), Cas9(K1020E)-TadA8e (**c**), Cas9(Q1027D)-TadA8e (**d**) and Cas9(T1033A)-TadA8e (**e**) relative to ABE8e are 52.9%, 73.7%, 55.1% and 59.0%, respectively.

f, Comparison of genome-wide sgRNA-dependent off-targets (evaluated by dI-profiling) for Cas9(K1031E)-TadA8e with those for ABE8e, using HEK4 sgRNA (n = 1,717 sites) in HEK293T cells. The percentage reduction of off-target signals for Cas9(K1031E)-TadA8e is 62.7%. Each dot represents an off-target A.



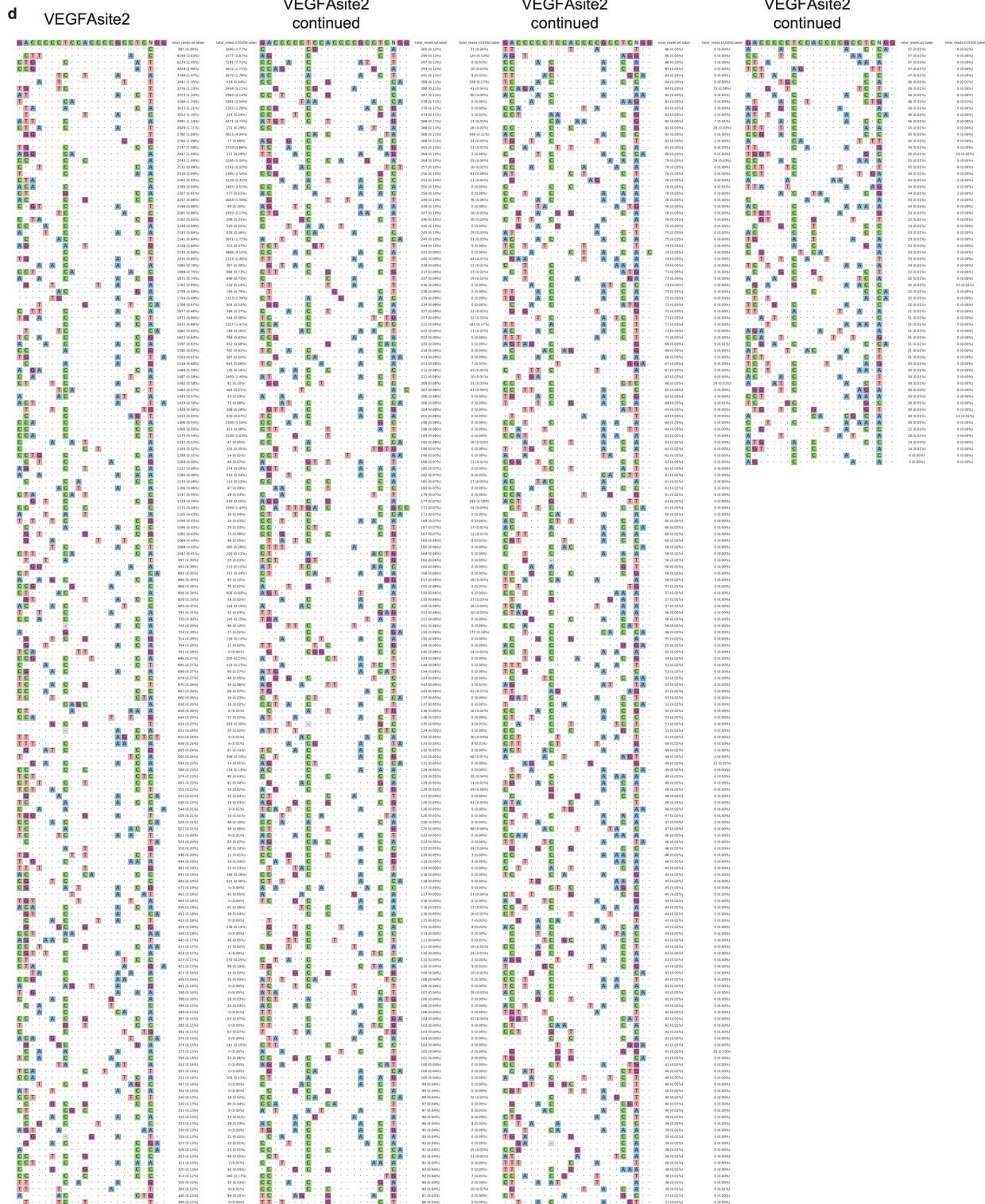
Cas9-gRNA-DNA on-target structure
(PDB ID: 5f9r)

Cas9-gRNA-DNA off-target structure
(PDB ID: 7s4x)

Supplementary Fig. 35 | Experimental determined structures suggest a potential off-target-specific K1020-nucleic acid contact.

In an off-target Cas9-sgRNA-DNA complex (PDB: 7s4x), K1020 is positioned close to the 5' segment of the ntsDNA, a contact that is not observed in the on-target structure (PDB: 5f9r). The rightmost panel shows the zoomed-in view of the dashed-box region in the middle panel.

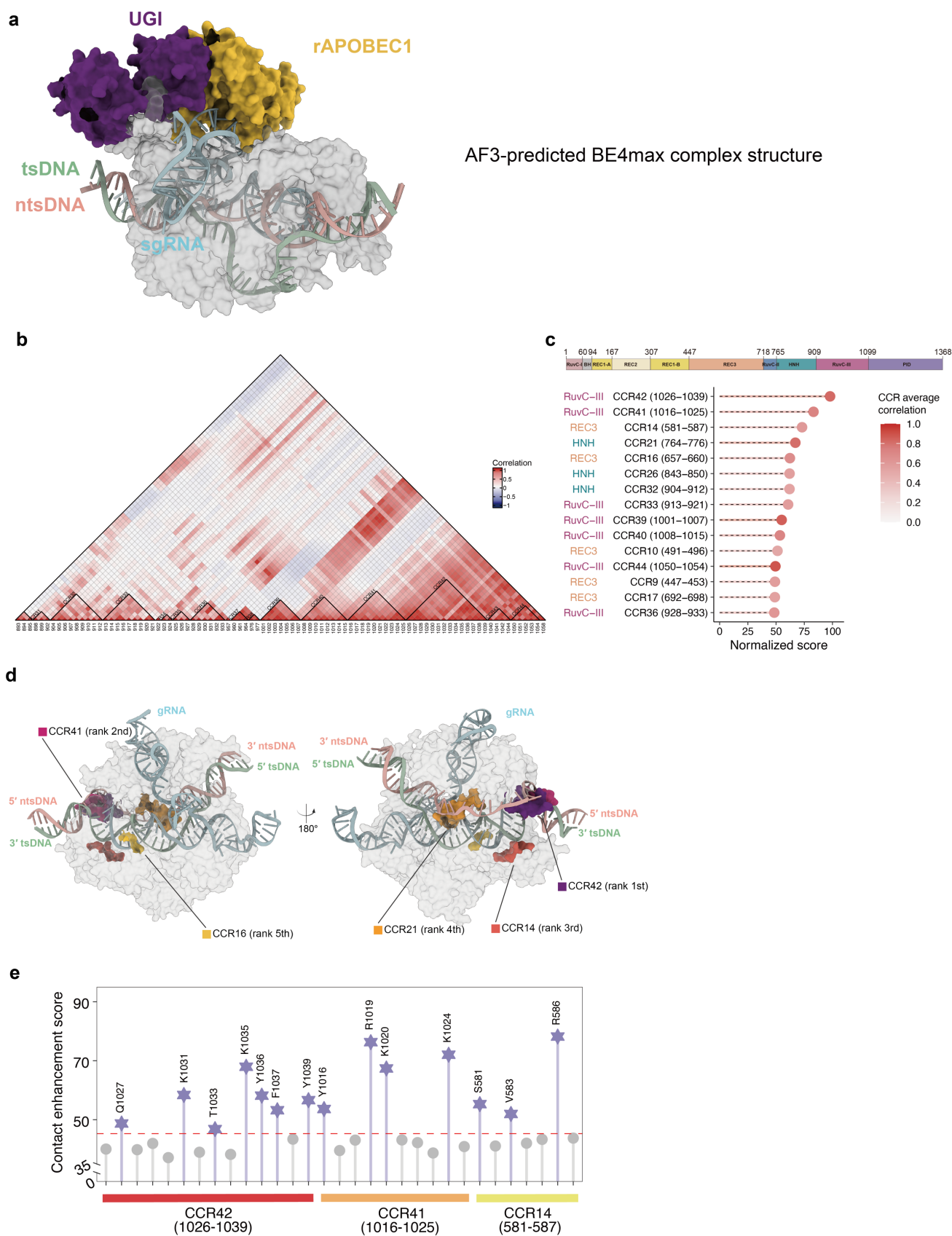
Continued



Supplementary Fig. 36 | GUIDE-seq results for Cas9 and Cas9(K1020D) nucleases.

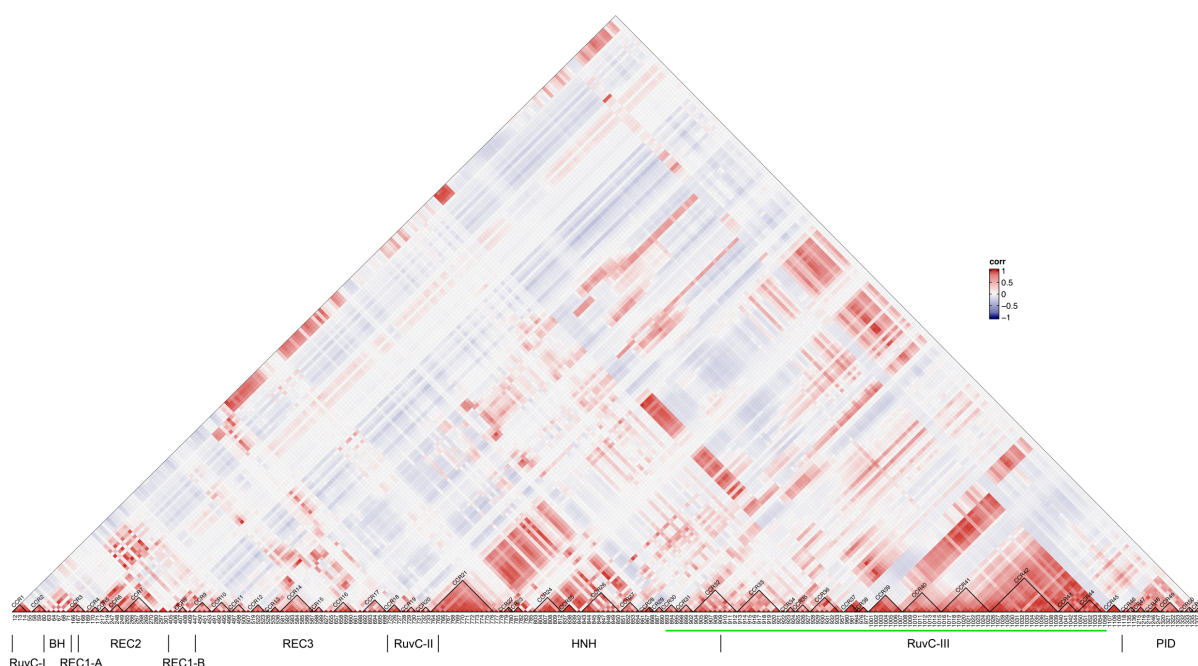
a, Number of off-target sites for Cas9 and Cas9(K1020D) nucleases identified by GUIDE-seq using HEK4 and VEGFA site2 sgRNAs in HEK293T cells. **b**, Ratios of GUIDE-seq off-target reads to on-

target reads for Cas9 and Cas9(K1020D) nucleases using HEK4 and VEGFAsite2 sgRNAs. **c, d**, Sequences of on-target and off-target sites identified by GUIDE-seq for Cas9 nuclease using the HEK4 sgRNA (**c**) and VEGFAsite2 sgRNA (**d**). The left column shows the alignment of each site to the on-target sequence. The middle and right columns show GUIDE-seq read counts for Cas9 nuclease and Cas9(K1020D) nuclease, respectively, with read count-to-total read ratios are indicated in parentheses.



Supplementary Fig. 37 | ContactSeek framework identifies specificity-enhancing positions for Cas9-based CBE (BE4max).

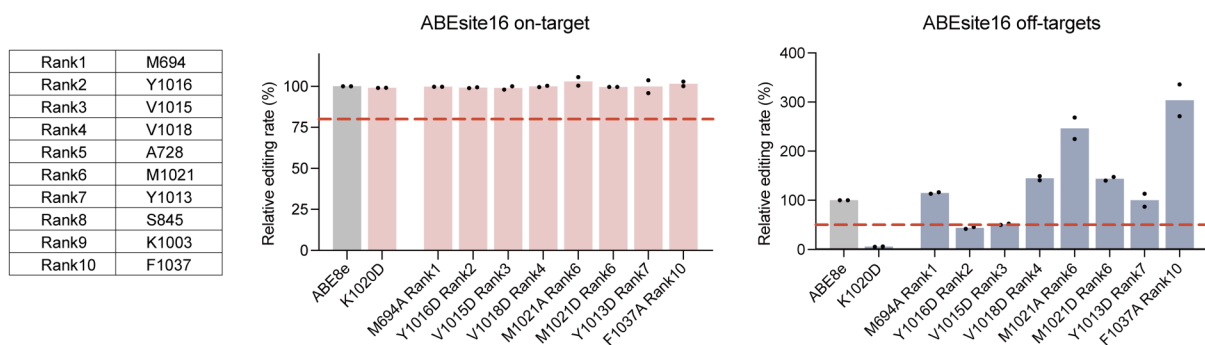
a, AF3 prediction of the full-length BE4max complex. When predicting the complex structure of BE4max, AF3 incorrectly positioned the deaminase onto the sgRNA scaffold. Therefore, as for Cas9-based ABEs, ContactSeek was applied to the core Cas-DNA-sgRNA complexes for BE4max. **b**, Heatmap of CP correlations among Cas9 residues 893-1058 across the BE4max off-target training dataset, revealing clustered CCRs. For CP correlations across all Cas9 residues in contact with DNA/sgRNA, see Supplementary Fig. 38. **c**, CCRs ranked by ContactSeek output score (x-axis) and mapped to Cas9 domains annotated at left. The color depth of the dot indicates the average Pearson correlation between residues within each CCR. **d**, 3D views of the top 5 CCRs (colored) in an AF3-predicted BE4max off-target complex. **e**, Contact enhancement score for each residue within the top 5 CCRs. Purple stars indicate residues with contact enhancement above the cutoff; red dashed line marks the cutoff.



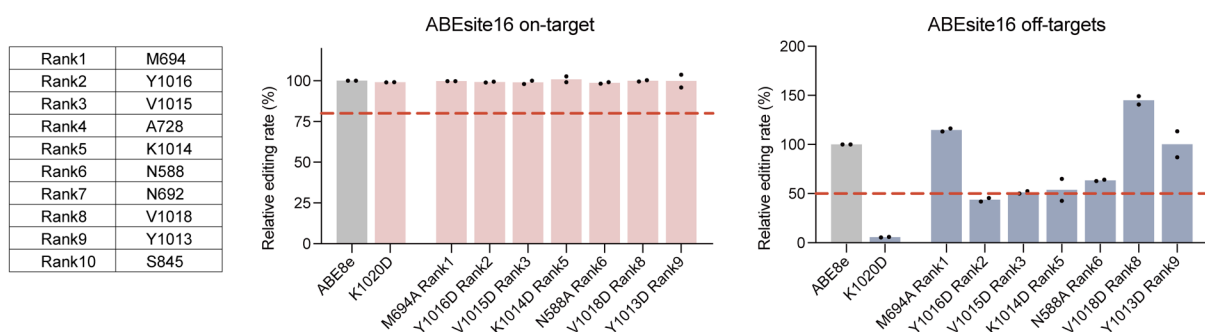
Supplementary Fig. 38 | CCRs of Cas9-based CBE across Cas9 residues.

Heatmap of CP correlations among all Cas9 residues in contact with DNA/RNA across the off-target training dataset of Cas9-based CBE (from sgRNA EMX1, HEK4 and VEGFAsite2; $n = 804$), revealing clustered CCRs. The green line under the residue indices delineates the zoomed-in regions shown in Supplementary Fig. 37b, and Cas9 domain annotations are indicated below the heatmap.

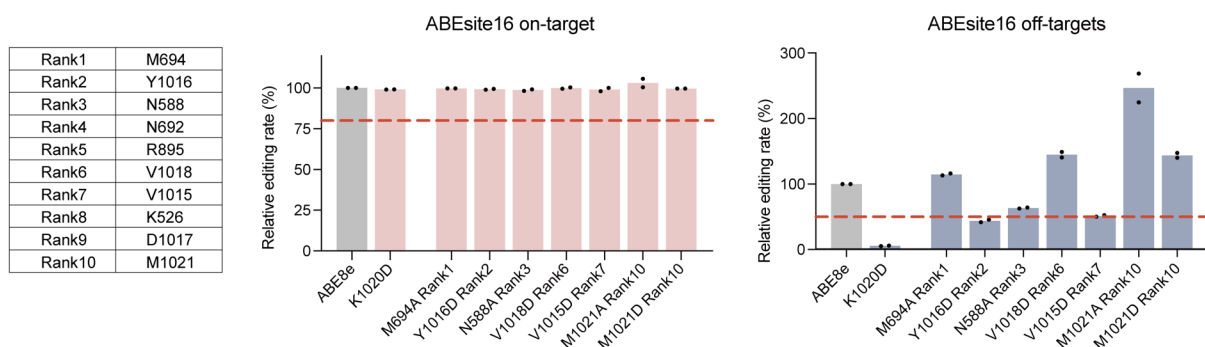
a Ranked by manhattan distance: tiling mismatch



b Ranked by manhattan distance: tiling mismatch + AID-seq + GUIDE-seq



c Ranked by manhattan distance: dI-profiling



Supplementary Fig. 39 | The editing specificity of ABE8e variants nominated by using only CP without integrating the dI-profiling signals.

a-c, Relative on-target and off-target editing rates of ABE8e variants based on Cas9 mutations, selected from the ranking list by the Manhattan distance of Δ CP between on and off-targets using three off-target sets: **(a)** in silico single-mismatch tiling sequences (protospacer positions 1-20); **(b)** a combined off-target set consisting of the tiling-mismatch sequences and off-target sequences identified by AID-seq and GUIDE-seq but not by dI-profiling; and **(c)** off-target sequences identified by dI-profiling. The measured on-target and off-target editing rates for each variant were normalized to those of ABE8e

and compared with Cas9(K1020D)-TadA8e. In each panel, the table on the left shows the ranking list; the middle plot shows relative on-target editing at ABESite16; and the right plot shows the mean off-target editing across ABESite16-OT72 and ABESite16-OT206. Each dot indicates an independent biological replicate ($n = 2$); bars indicate the mean. Across all three ranking schemes, the nominated mutations generally retain on-target activity but confer at most ~ 2 -fold improvements in specificity, indicating that CP-only Manhattan-distance scanning has limited power for engineering high-specificity variants compared with the typical ContactSeek framework that are guided by empirical sequencing signals.

a

Constant sequence:

Chain A: TadA8e length = 167

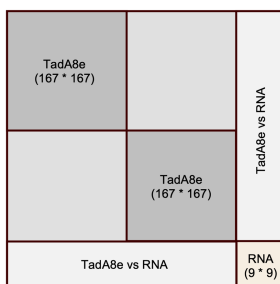
Chain B: TadA8e length = 167

Variable sequence:

Chain C: RNA length = 9 (edited A 1nt + 5' extend 4nt + 3' extend 4nt)

Total token number = $167 * 2 + 9 = 343$

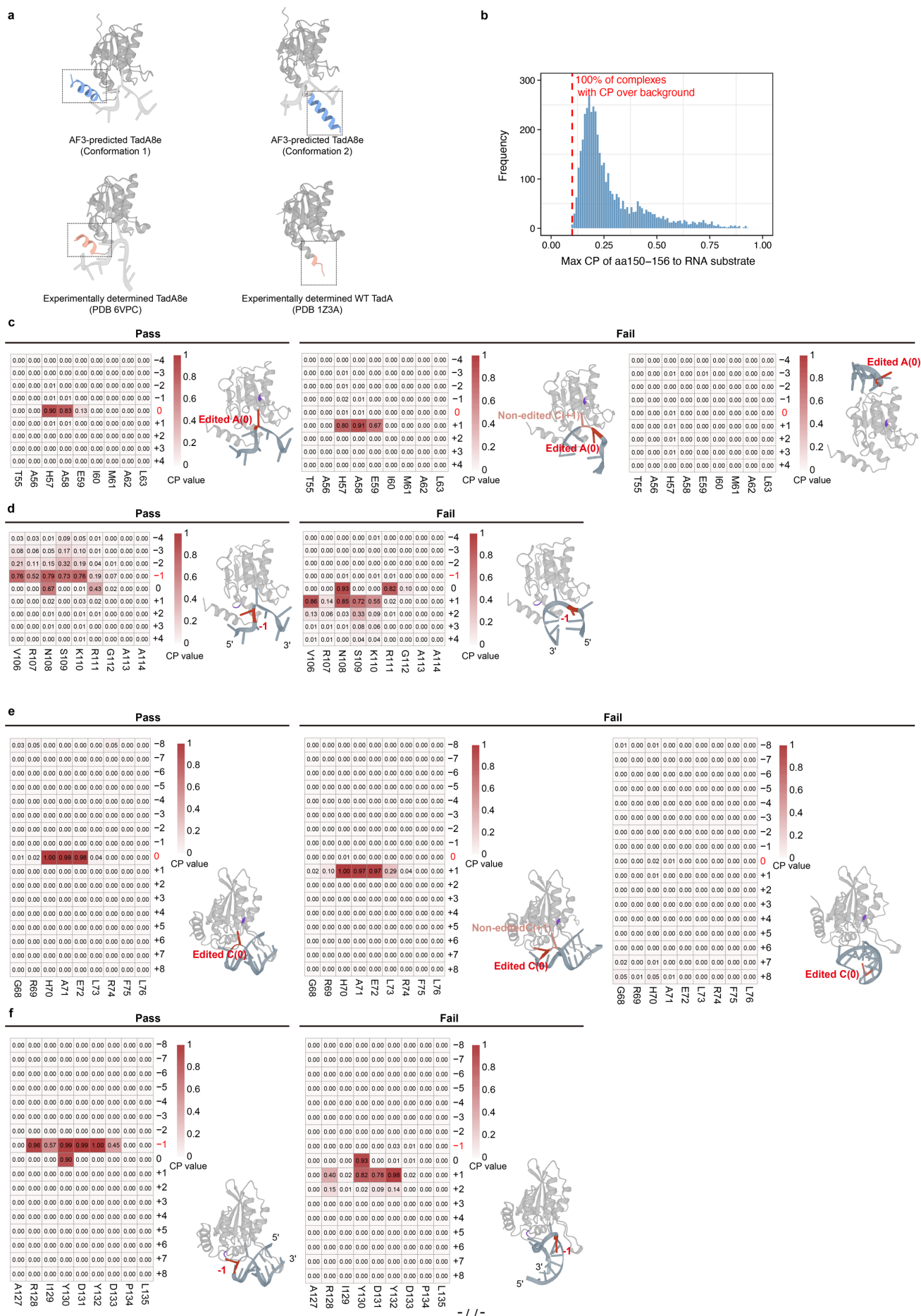
b



**Contact probability matrix
(343 * 343)**

Supplementary Fig. 40 | Token of AF3 predictions for TadA8e-RNA binary complex

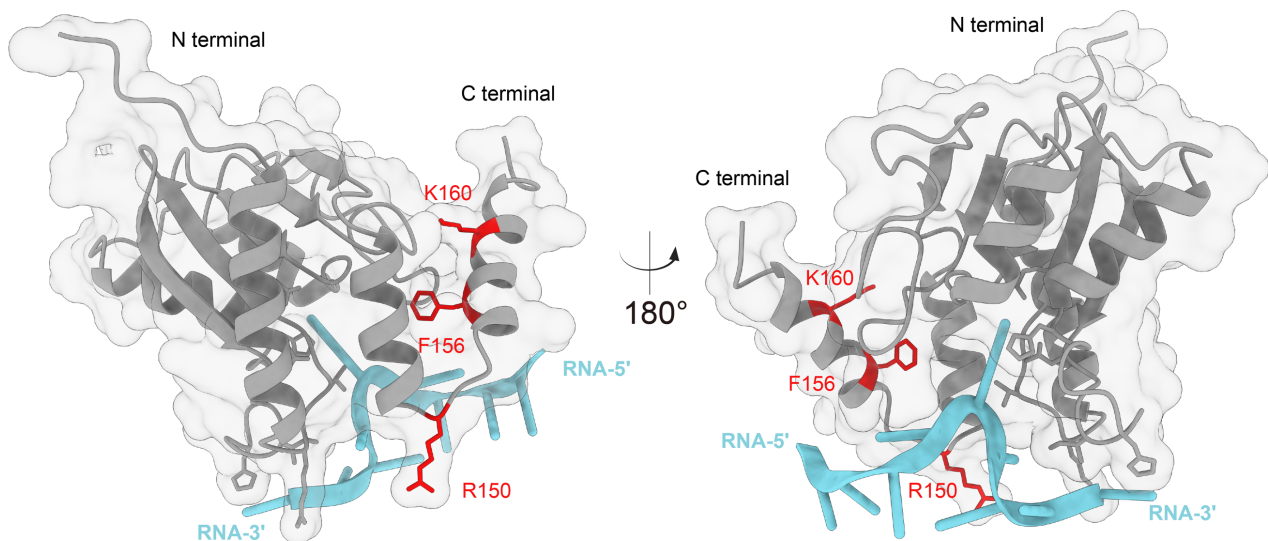
a, Illustration of the inputs provided to AF3 for TadA8e-RNA complex prediction. Because TadA8e functions as a dimer, two identical TadA8e protein chains (chain A and chain B) were included. The TadA8e protein sequence is constant across all complexes, whereas the RNA sequences varied depending on the specific off-target site. The lengths of each input component are indicated. **b**, Detailed representation of the contact probability matrix illustrating the inter- and intra-macromolecular contact probability values among the components of the TadA8e-RNA complex.



Supplementary Fig. 41 | CP-based filtering of AF3-predicted TadA8e-RNA and APOBEC3A-RNA complexes prior to input into the ContactSeek framework.

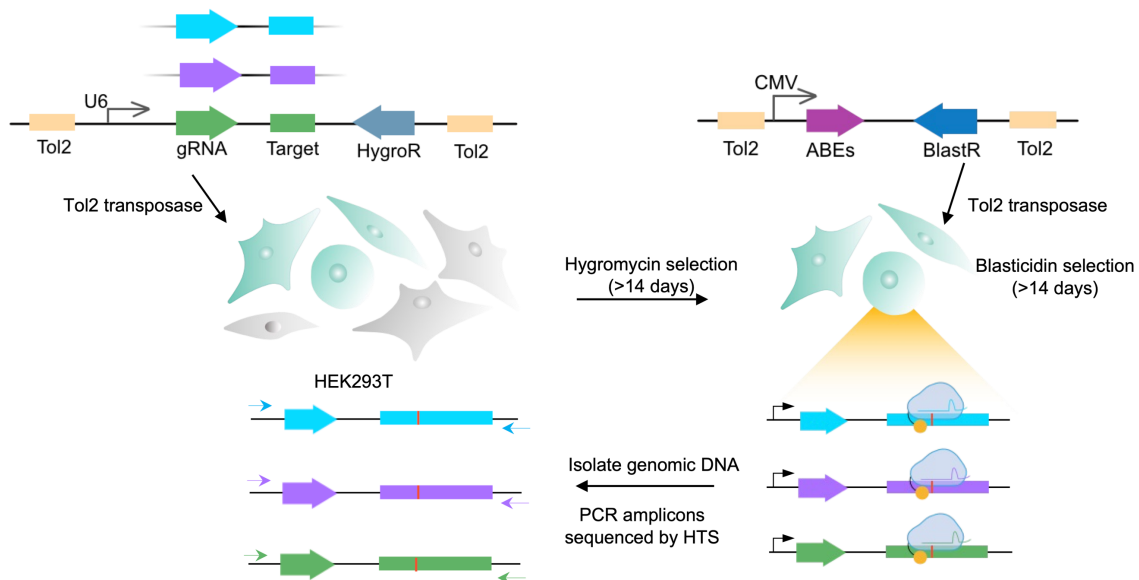
a, Comparison of AF3-predicted TadA8e structures with experimentally determined TadA8e and wild-type TadA structures. Top, two distinct conformations of TadA8e-RNA complex observed during AF3 prediction. Bottom left, cryo-EM structure of the TadA8e-DNA complex (PDB: 6vpc). Bottom right, crystal structure of wild-type TadA (PDB: 1z3a). In the TadA8e structure, residues 152-167 in the C-terminal α -helix adopt a sharply bent conformation toward the substrate, whereas in wild-type TadA the helix is fully extended and points away from the rest of the protein. AF3 occasionally hallucinates residues 152-167 of TadA8e (highlighted and marked by black dashed boxes) back into the wild-type TadA conformation. **b**, Histogram of the maximum CP value between TadA8e residues 150-156 and the RNA substrate across 1,426 AF3-predicted complexes. The red dashed line indicates the background CP threshold (CP = 0.05). Despite the conformational heterogeneity shown in panel **a**, this region exhibits CP values above background in almost all complexes. **c**, **d**, CP heatmaps and corresponding structures illustrating the CP-based filtering strategy for AF3-predicted TadA8e-RNA complexes. **c**, Filter step (i) ensures that the edited off-target adenine (A, indexed as 0) is correctly positioned in the catalytic pocket. We required evident CP (>0.5) between base 0 and residues near the catalytic residue E59, and that this CP value is the highest among all RNA bases. Shown are examples that pass this filter (left), fail because contacts form primarily with the non-edited base C (+1) rather than the edited A (middle), or fail because the RNA is positioned away from the catalytic site and shows little CP (right). The catalytic residue E59 is highlighted in purple. **d**, Filter step (ii) requires the correct 5'-to-3' orientation of the RNA substrate. Based on the known interaction between the -1 base and K110, we required a higher CP between the -1 base and residues centered on K110 than between the +1 base and the same residues. Complexes passing this filter show evident CP between the -1 base (red) and residues N108-R111 (centered on K110) (left), whereas contacts with the +1 base indicate an incorrect orientation (right). The residue K110 is highlighted in purple. **e**, **f**, CP heatmaps and corresponding structures illustrating the CP-based filtering strategy for AF3-predicted A3A-RNA complexes. **e**, Filter step (i) requires evident CP (>0.5) between the edited C (indexed as 0) and residues near the catalytic residue E72, and that this CP value is the highest among all RNA bases. Shown are

examples that pass this filter (left), fail because contacts are primarily formed with the non-edited C (+1) rather than the edited C (middle), or fail because the RNA is positioned away from the catalytic site and shows little CP (right). The catalytic residue E72 is highlighted in purple. **f**, Filter step (ii) requires the correct 5'-to-3' orientation of the RNA substrate. Based on the known interaction between the -1 base and D131, we required a higher CP between the -1 base and residues centered on D131 than between the +1 base and the same residues. Complexes passing this filter show evident CP between the -1 base and residues Y130-D133 (centered on D131) (left), whereas contacts with the +1 base indicate an incorrect orientation (right). The residue D131 is highlighted in purple. Color scales for CP values are shown beside each heatmap.



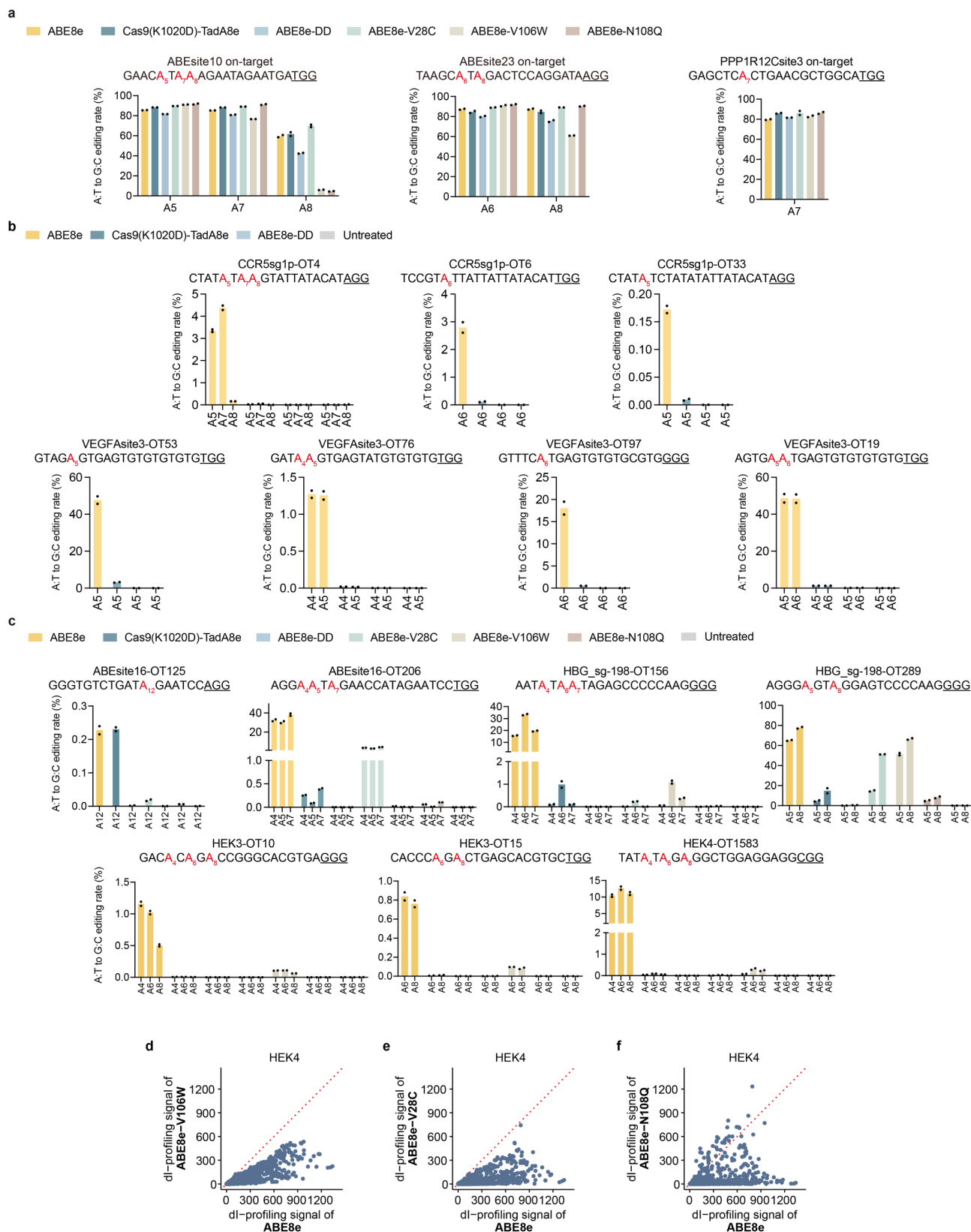
Supplementary Fig. 42 | Three positions altering TadA8e editing specificity identified by ContactSeek.

An AF3-predicted structure of a TadA8e-RNA complex, highlighting the three specificity-altering residues (R150, F156, and K160). These residues, identified by ContactSeek, are located in the C-terminal region of TadA8e, far from the catalytic active core where the target adenosine is flipped into the deaminase active pocket.



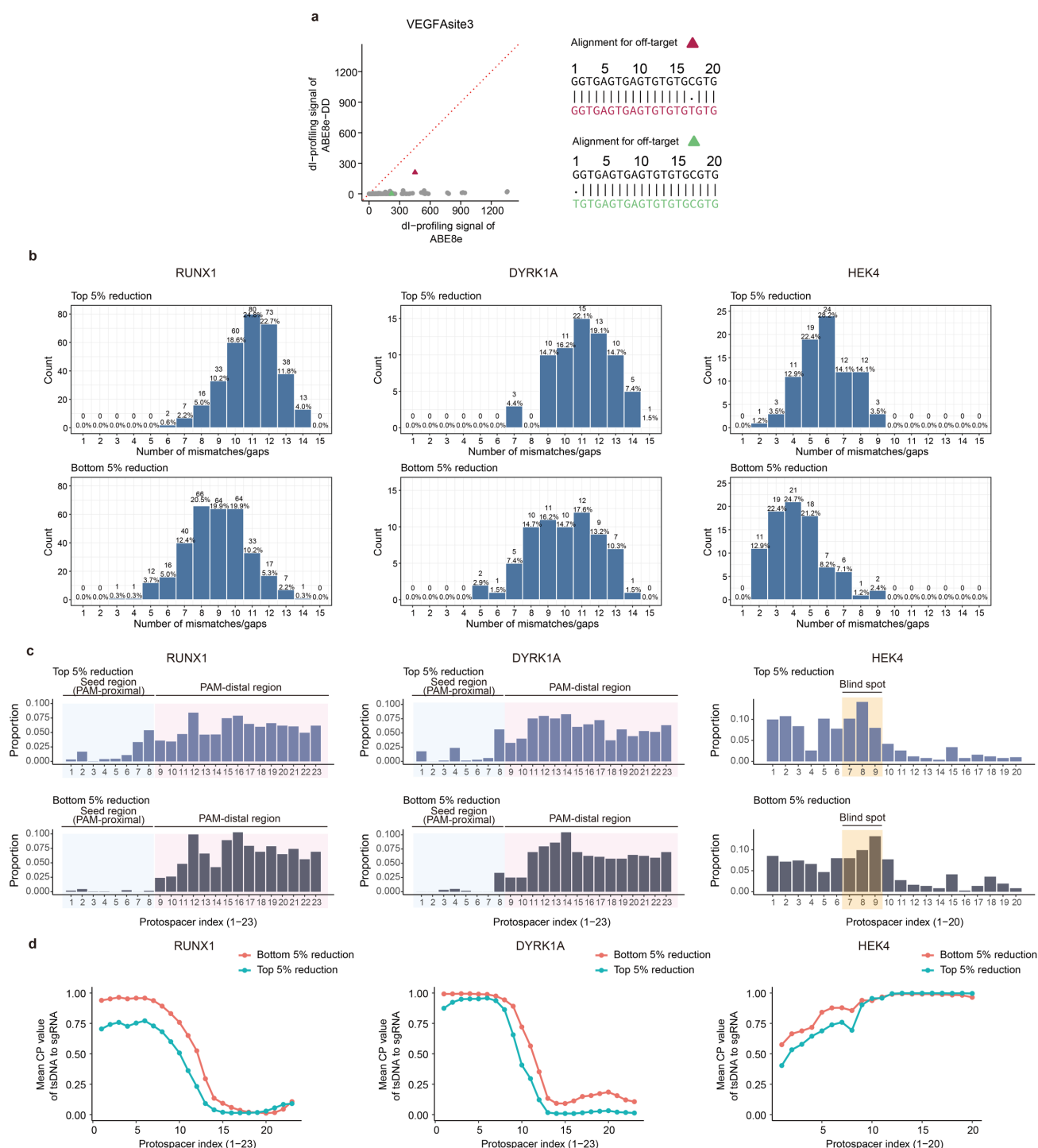
Supplementary Fig. 43 | Schematic workflow for assessing the on-target activity of ABEs using a large-scale sgRNA library.

A synthetic sgRNA library, in which each sgRNA is linked to its corresponding target sequence, was cloned into a Tol2 transposon vector containing a hygromycin resistance cassette (HygroR) and co-transfected with Tol2 transposase into HEK293T cells. Cells were then subjected to hygromycin selection for more than 14 days to enable stable genomic integration of the sgRNA-target library. Next, ABE8e variants cloned into a Tol2 transposon vector containing a blasticidin resistance cassette (BlastR) were co-transfected with Tol2 transposase into the library-integrated cells, followed by blasticidin selection for more than 14 days. Genomic DNA was then isolated, and the integrated target sites were PCR-amplified and analyzed by high-throughput sequencing.



Supplementary Fig. 44 | Comparison of ABE8e-DD with several leading high-fidelity ABE variants on additional genomic loci.

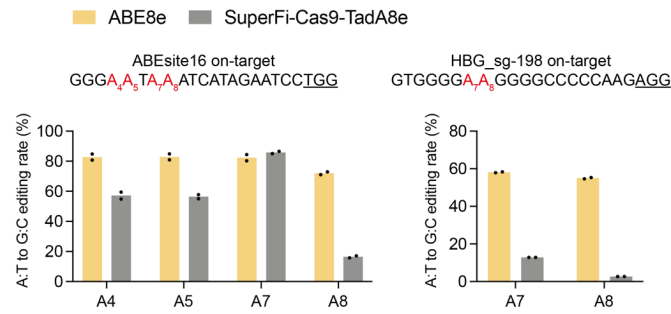
a, On-target editing rates of ABE8e, Cas9(K1020D)-TadA8e, ABE8e-DD, ABE8e-V28C, ABE8e-V106W and ABE8e-N108Q at three endogenous genomic loci in HEK293T cells. **b**, Off-target editing rates of ABE8e, Cas9(K1020D)-TadA8e and ABE8e-DD at additional representative off-target sites for sgRNAs (CCR5sg1p or VEGFA site3) that were not used for ContactSeek modeling. **c**, Off-target editing rates of ABE8e, Cas9(K1020D)-TadA8e, ABE8e-DD, ABE8e-V28C, ABE8e-V106W and ABE8e-N108Q at additional representative off-target loci of 4 sgRNAs. In **a-c**, editing efficiencies were measured in HEK293T cells by targeted amplicon sequencing. The subscripted red A denotes the edited A position within the protospacer and the PAM sequence is underlined. Each dot indicates an independent biological replicate ($n = 2$); bars indicate the mean. **d-f**, Comparison of genome-wide sgRNA-dependent off-targets (evaluated by dI-profiling) for ABE8e-V106W (**d**) ABE8e-V28C (**e**) and ABE8e-N108Q (**f**) with those for ABE8e, using HEK4 sgRNA ($n = 1,717$ sites) in HEK293T cells. The percentage reductions of off-target signals for ABE8e-V106W (**d**) ABE8e-V28C (**e**) and ABE8e-N108Q (**f**) are 70.5%, 88.2% and 86.4%, respectively. Each dot represents an off-target A.



Supplementary Fig. 45 | Comparison of the most reduced off-target sites and the least reduced off-target sites.

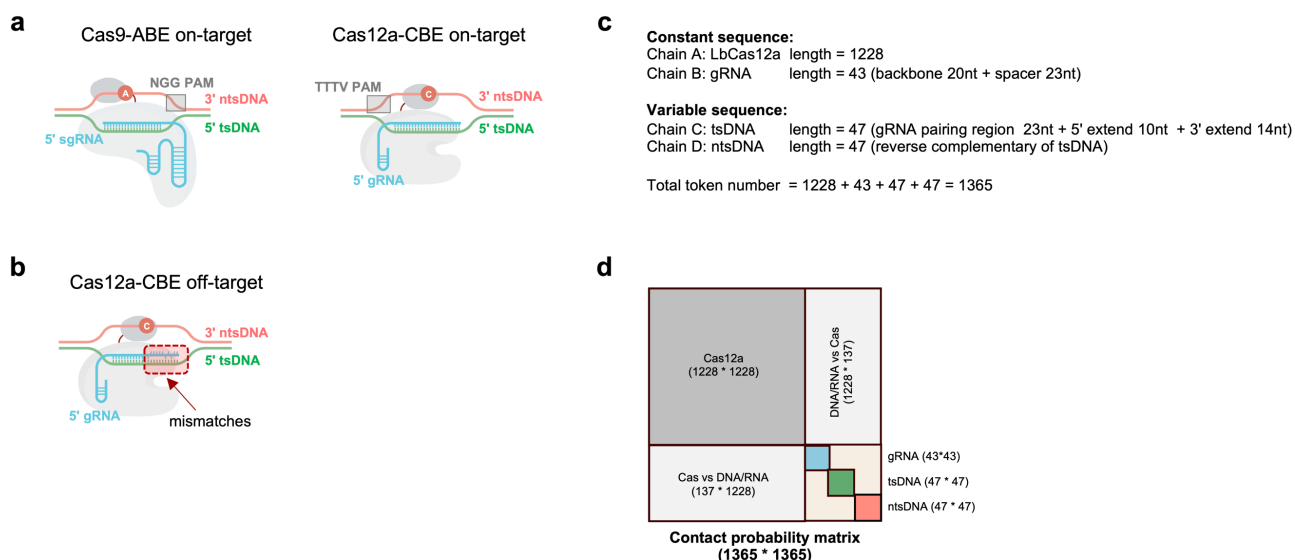
a, Left, comparison of genome-wide sgRNA-dependent off-target editing between ABE8e-DD and ABE8e at the VEGFAsite3, evaluated by dI-profiling (using the same dataset in Extended Data Fig. 5g, $n = 1,717$ sites). Right, sequence alignments between the on-target site (black) and representative

single-mismatch off-target sites showing strong (green) versus modest (red) signal reduction. Black dashes indicate matched positions and black dots indicate mismatches. **b**, **c**, Distributions of the number of mismatches/gaps (**b**) and mismatch/gap positions within the protospacer (**c**) of the top 5% (most reduced) and bottom 5% (least reduced) of off-target sites, ranked by the degree of off-target editing reduction. **d**, Mean CP value between tsDNA and sgRNA across the protospacer for the top 5% and bottom 5% of off-target sites, ranked by the degree of off-target editing reduction. In panels **b-d**, off-target editing reduction levels were calculated based on Detect-seq signals for Cas12a(R284E)-A3A(H29D) relative to Cas12a-A3A at RUNX1 (left column) or DYRK1A loci (middle column), or based on dI-profiling signals for ABE8e-DD relative to ABE8e at HEK4 locus (right column).



Supplementary Fig. 46 | On-target editing of SuperFi-Cas9-TadA8e compared to ABE8e.

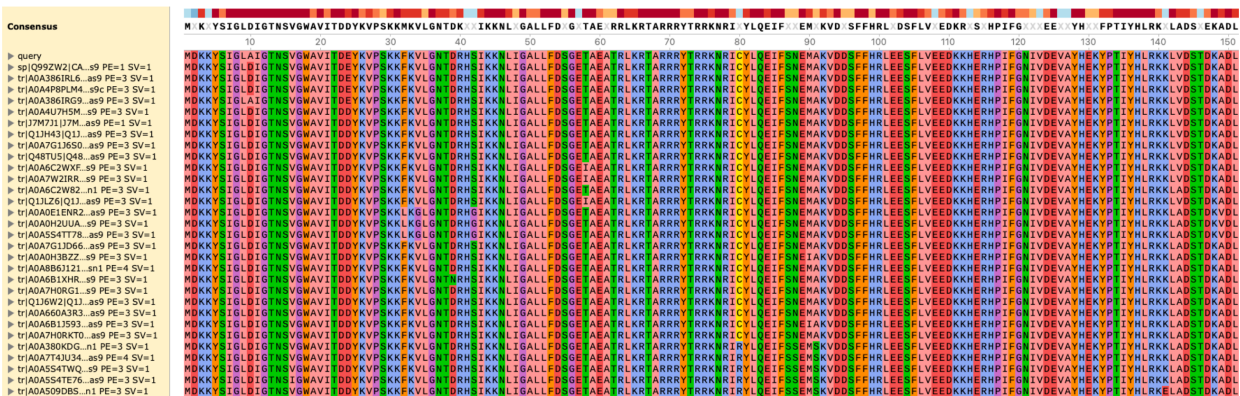
On-target editing rates of ABE8e and SuperFi-Cas9-TadA8e at four endogenous genomic sites. The subscripted red A denotes the edited A position within the protospacer and the PAM sequence is underlined. Each dot represents an independent biological replicate (n = 2); bars indicate the mean.



Supplementary Fig. 47 | Structural features and AF3 representation for Cas12a-based CBE.

a, On-target ternary complex schematics for Cas9-ABE and Cas12a-CBE. The Cas12a-CBE ternary complex differs markedly from the Cas9-ABE ternary complex: first, Cas12a recognizes its PAM sequence at the 5' end of the protospacer, in contrast to Cas9's recognition of PAM at the 3' end. This reversal in PAM positioning results in opposite R-loop formation directions and consequently different mismatch locations in the tsDNA-sgRNA hybrid; second, as a Class 2 Type V effector, Cas12a utilizes a longer protospacer sequence than Class 2 Type II Cas9, resulting in an expanded R-loop; third, Cas12a uses a much shorter guide RNA than Cas9 (~43 nt versus ~100 nt), thereby reducing the contactable sites between protein and RNA. **b**, Off-target ternary complex schematic for Cas12a-CBE. The off-target Cas12a-CBE complex differs from its on-target counterpart by the presence of mismatches between the tsDNA and gRNA, as highlighted by the red dashed box. **c**, For the prediction of each Cas12a-DNA-gRNA ternary complex, the LbCas12a protein sequence and the gRNA sequence are constant inputs, whereas the tsDNA and the ntsDNA are variable inputs. The lengths of each input component are indicated. **d**, A detailed representation of the contact probability matrix illustrating the inter- and intra-macromolecular contact probability values among the four components of the Cas12a-DNA-gRNA complex.

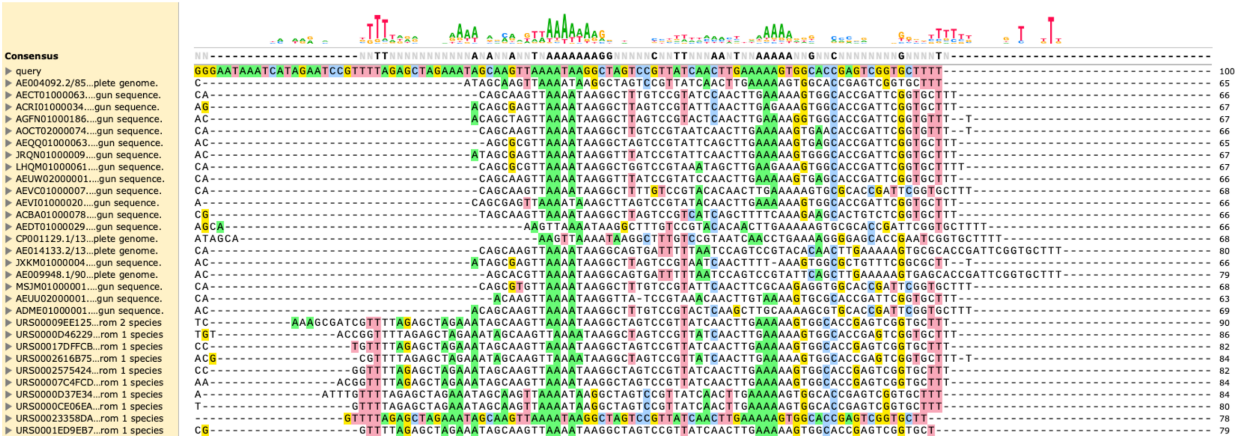
a



b



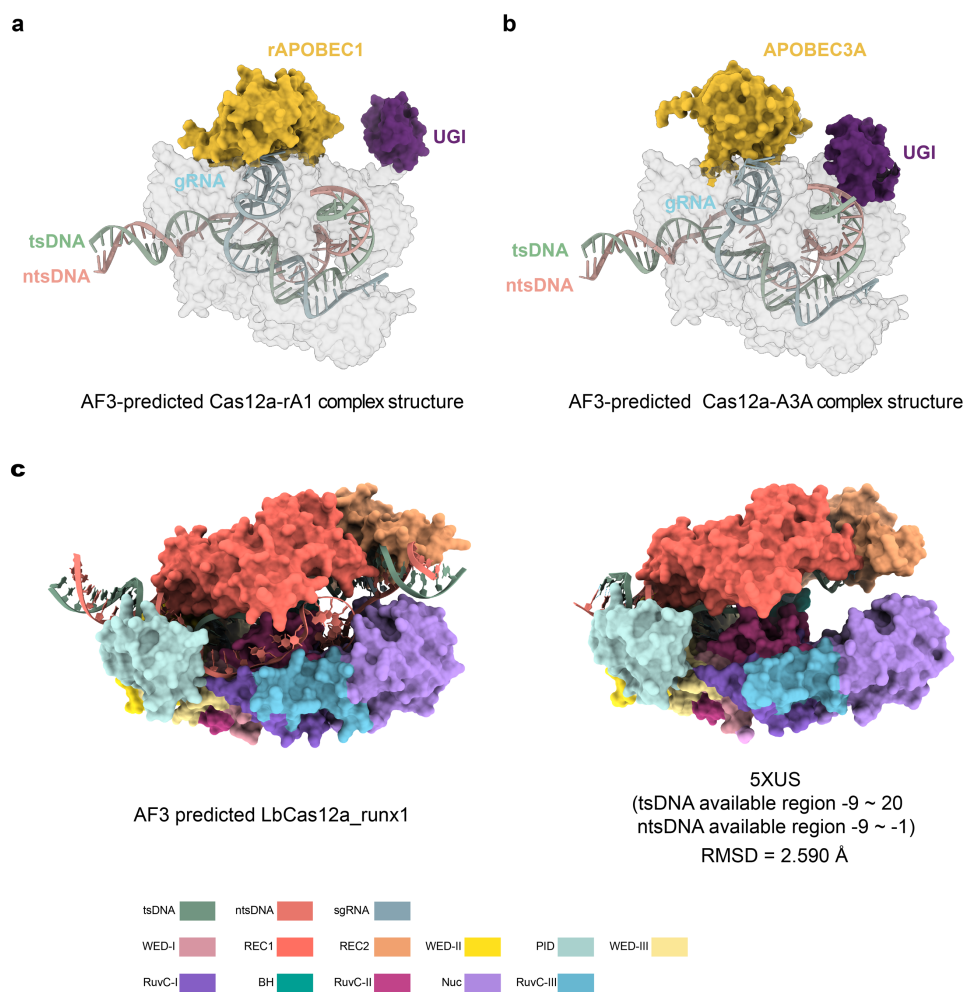
c



AF3 MSA results for SpCas9-sgRNA, Top-30 shown, number of query template = 78
No RNA template in AF3 database for LbCas12a-gRNA, number of query template = 0

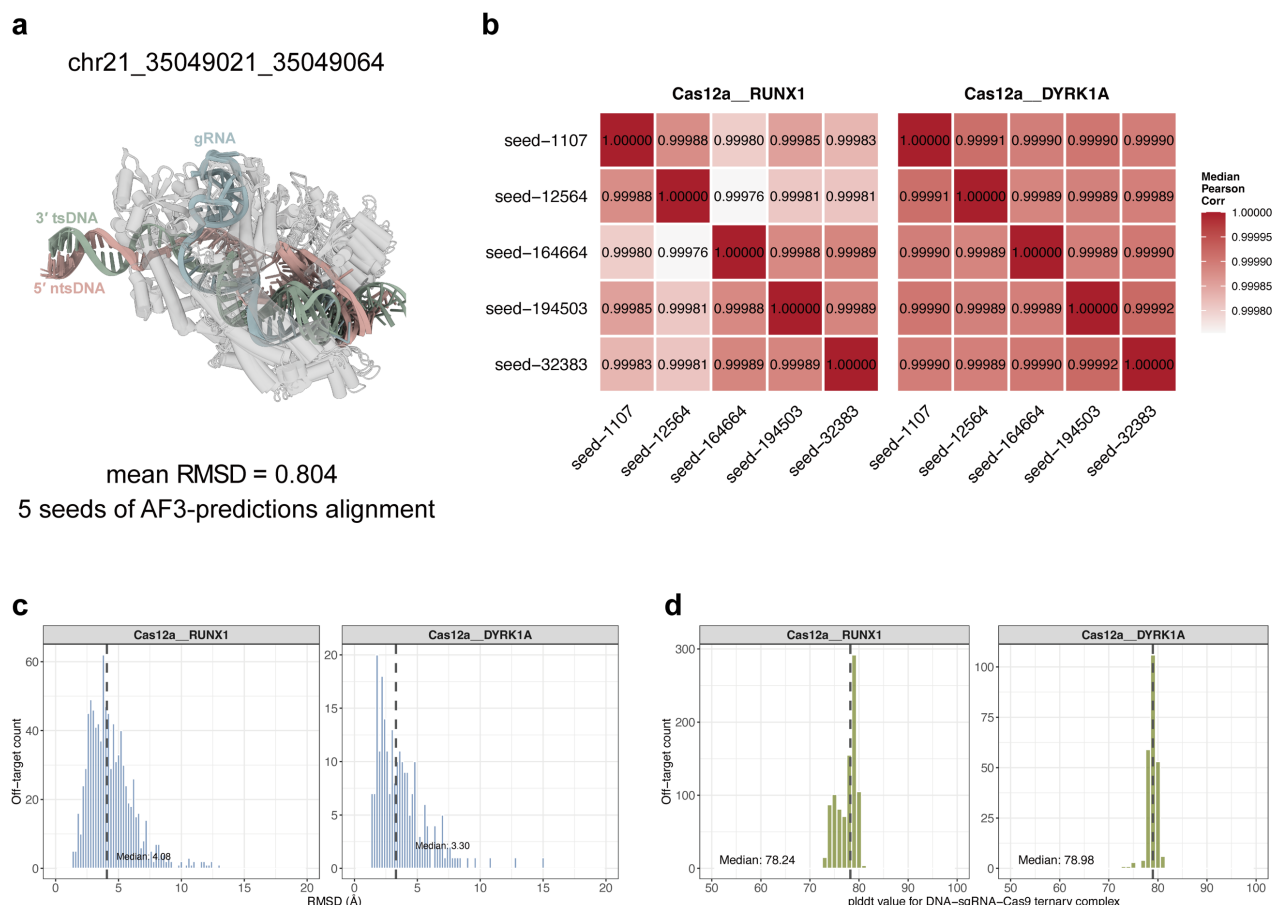
Supplementary Fig. 48 | Limited template sequence availability in AF3 database for Cas12a.

Multiple sequence alignment (MSA) results of the template sequences during AF3 predictions for the SpCas9 protein (a), the LbCas12a protein (b), and the SpCas9 sgRNA (c). Only the top 30 alignments are presented. Notably, there is no RNA template for the LbCas12a gRNA in AF3 database.



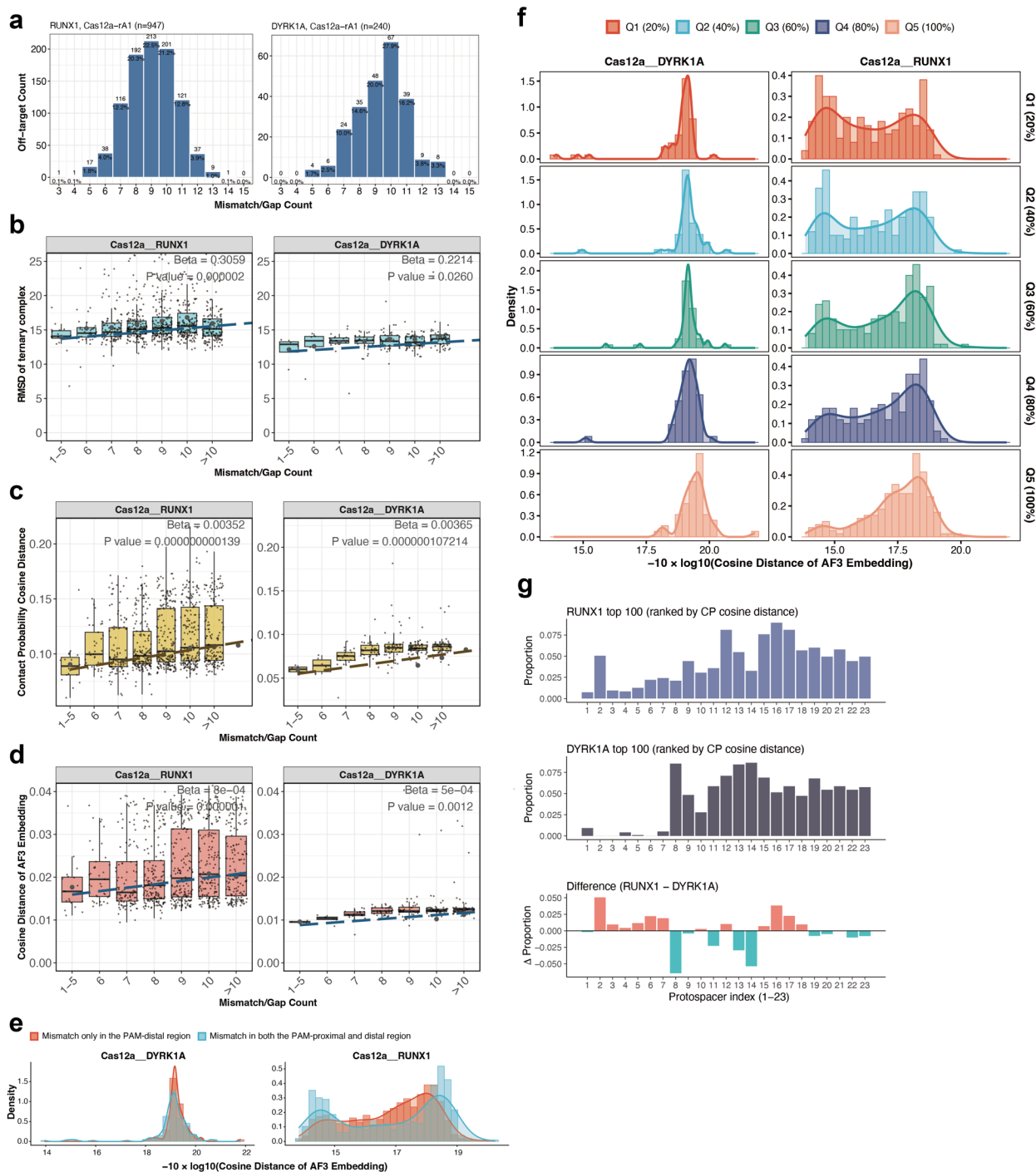
Supplementary Fig. 49 | AF3 prediction of Cas12a-based CBE complexes.

a, b, AF3 predictions of the full-length Cas12a-rAPOBEC1 (Cas12a-rA1, **a**) and Cas12a-APOBEC3A (Cas12a-A3A, **b**) ternary complex. When predicting the on-target complex structure (at RUNX1 loci) of full-length Cas12a-rA1/A3A CBE tools, AF3 incorrectly positioned the deaminases onto the sgRNA scaffold. Therefore, ContactSeek was applied to the core Cas-DNA-RNA complexes for Cas12a-based CBEs. **c**, Comparisons of the AF3-predicted LbCas12a-DNA-gRNA complex with the experimentally determined structure (PDB: 5xus). Pruned and all-paired RMSD values are shown, and major structural domains and nucleic acid components are labelled.



Supplementary Fig. 50 | High reproducibility and structural convergence of AF3 predictions for Cas12a-DNA-gRNA complex.

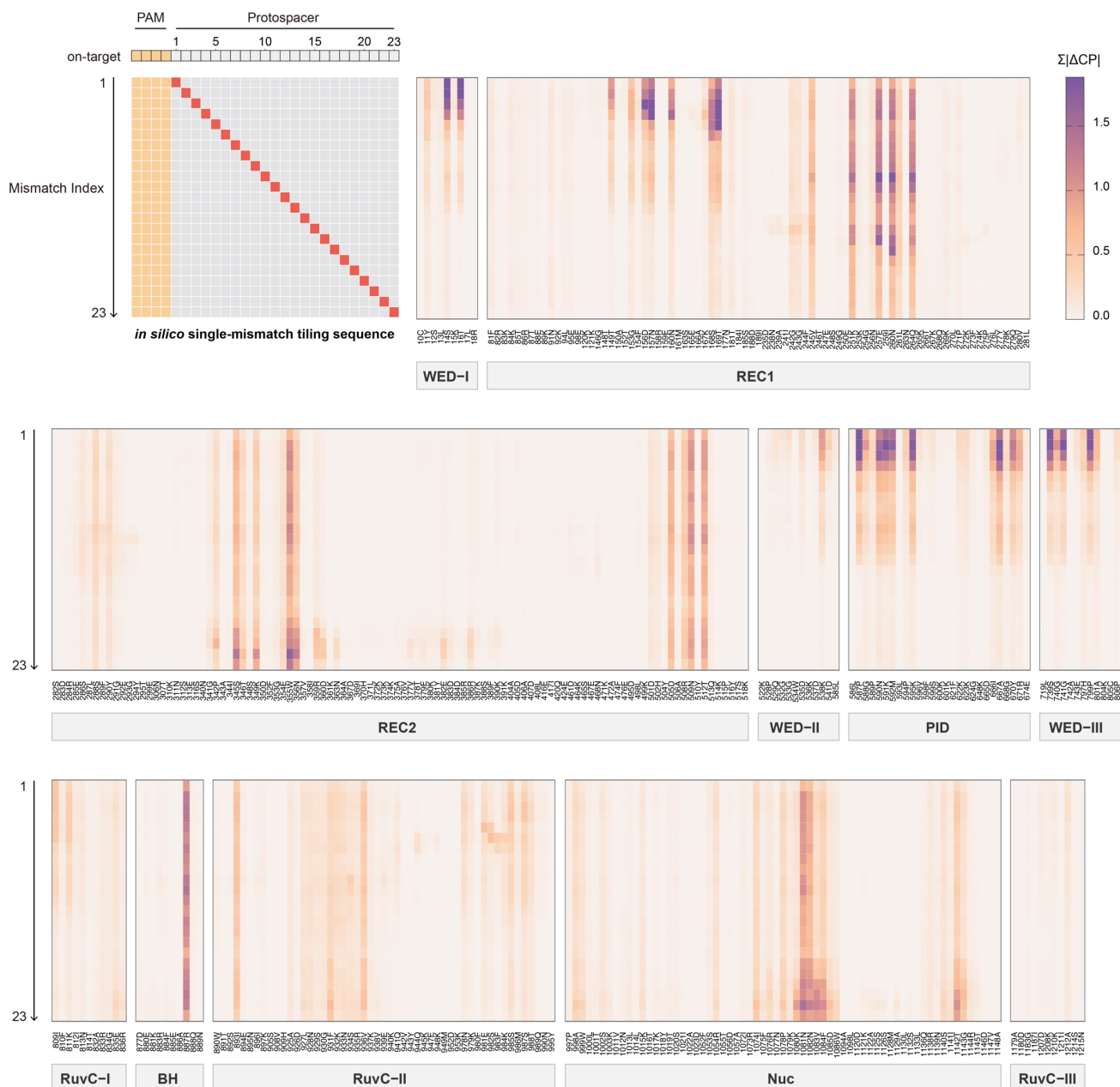
a, Overlay of 5 independently predicted Cas12a-DNA-gRNA ternary complex structures at the RUNX1 on-target site. These structures were generated by AF3 using identical inputs but with 5 different initialization seeds. The resulting complexes were superimposed, and the mean RMSD across the 5 predictions is shown below. **b**, Heatmap illustrates the median Pearson correlation coefficients of CPs from various Cas12a complex predictions. Predictions used five different AF3 initial seed values (1107, 12564, 164664, 194503, and 32383) across two loci: RUNX1 and DYRK1A. **c**, Histograms display the distribution of mean RMSD values for each off-target ternary complex, each predicted independently five times using five different initialization seeds. Median values are indicated by dark gray dashed lines. The RMSD and CP correlations in **(a)-(c)** were calculated based on the entire complex. **d**, Histograms show the distribution of pLDDT values for each AF3-predicted off-target ternary complex, with median values indicated by dark gray dashed lines.



Supplementary Fig. 51 | Analysis of off-target mismatch number, mismatch position, and signal strength for Cas12a.

a, Histograms show the count distribution of mismatches and gaps in Cas12a-rA1 CBE off-target editing regions in HEK293T cells, aligned to the on-target sequences of RUNX1 or DYRK1A. The number and the proportion of mismatches and gaps are shown above each group. **b-d**, Boxplots illustrate the relationship between the number of mismatches/gaps and changes in AF3-predicted

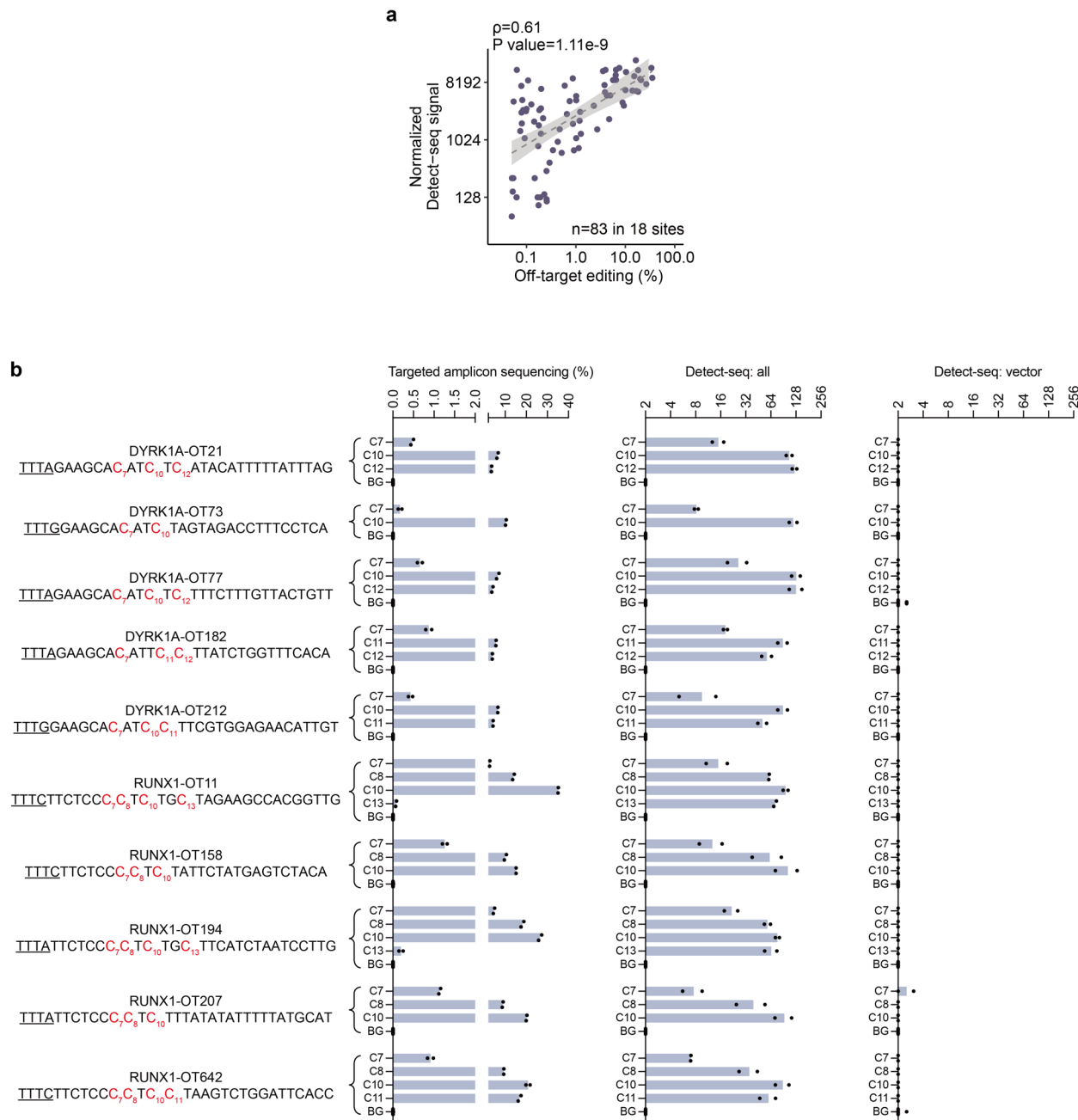
Cas12a-DNA-gRNA complexes, quantified as **(b)** structural deviations measured by RMSD, **(c)** contact-probability differences measured by cosine distance, and **(d)** AF3 embedding differences measured by cosine distance. Each value is computed for an off-target ternary complex relative to its corresponding on-target ternary complex. Dashed black lines represent the fitted regression models, with regression coefficients (Beta) and associated P values from the F-test indicating statistical significance. Dots number $n = 232$ and 911 for Cas12a__DYRK1A and Cas12a__RUNX1, respectively. **e-f**, Histograms illustrate the relationship between off-target density and AF3 embedding differences (measured as $-10 \times \log_{10}$ cosine distance) across **(e)** two groups of mismatch positions (PAM-distal only vs. both PAM-proximal and distal), or **(f)** five classes of off-target signal strength (five quantile-based bins, Q1-Q5, represent off-target intensity ranges of 0-20%, 20-40%, 40-60%, 60-80%, and 80-100%, respectively). In **e** and **f**, density curves were generated by Gaussian kernel density estimation, with bandwidth selected automatically according to Silverman's rule of thumb, and are plotted on the same density scale as the histograms to enable direct visual comparison. **g**, Distributions of the mismatch and gap positions along the protospacer among the top 100 off-target sites (ranked by CP cosine distance relative to the corresponding on-target site) for RUNX1 (upper panel) and DYRK1A (middle panel) respectively. The difference between the two distributions (RUNX1 – DYRK1A) along the protospacer is shown in the bottom panel.



Supplementary Fig. 52 | Position-specific single-nucleotide mismatches induce localized CP changes within different Cas12a domains.

Heatmaps showing CP changes induced by single-nucleotide mismatches in the Cas12a-DNA-gRNA complex, computed using the same AF3-based workflow as in the single-mismatch tiling analysis for Cas9-DNA-sgRNA (see Extended Data Fig. 2a). Target-strand (tsDNA) and non-target-strand (ntsDNA) sequences were systematically varied to introduce single mismatches within the protospacer region (positions 1-23; PAM positions -4 to -1). The color scale indicates the mean cumulative $|\Delta CP|$ across 10 gRNAs for Cas12a (Supplementary Table 1). Each row represents a mismatch configuration

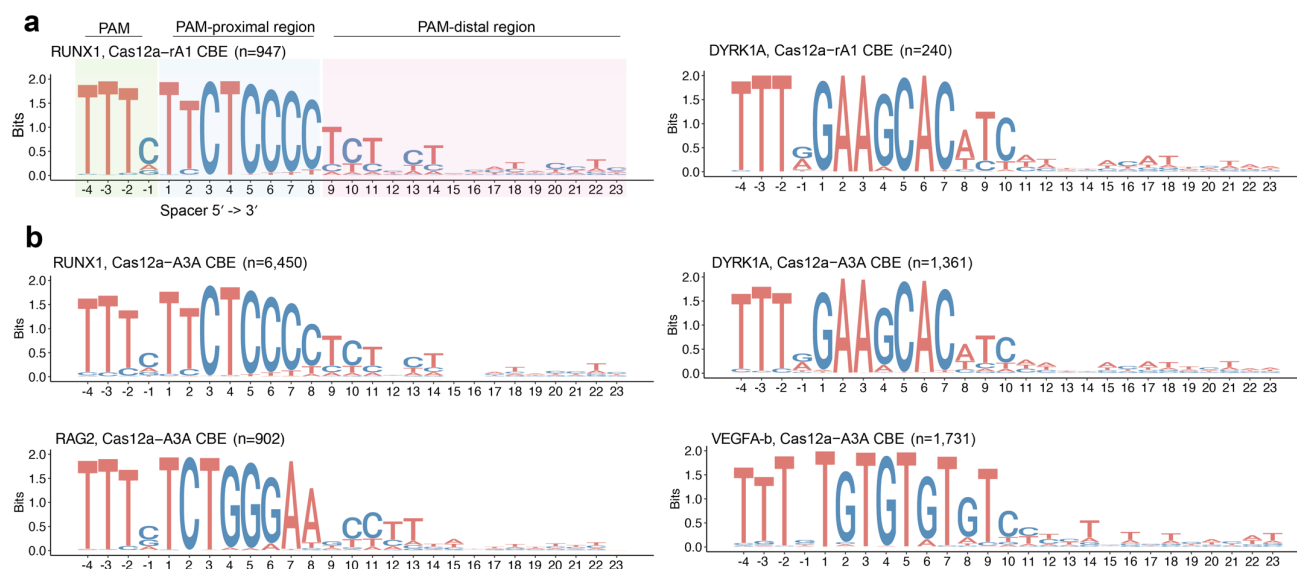
(arrows indicate the tiling direction along the protospacer), and each column corresponds to a Cas12a residue. Only residues exhibiting contact with DNA/RNA are displayed. Cas12a domains are labeled below. Mismatches in different protospacer regions elicit distinct, domain-localized Δ CP responses in LbCas12a; for example, PAM-proximal, but not PAM-distal, mismatches strongly affect Δ CP in REC1/WED/PI, consistent with previous observations.



Supplementary Fig. 53 | Validation of DNA off-target sites identified by Detect-seq for Cas12a-based CBE.

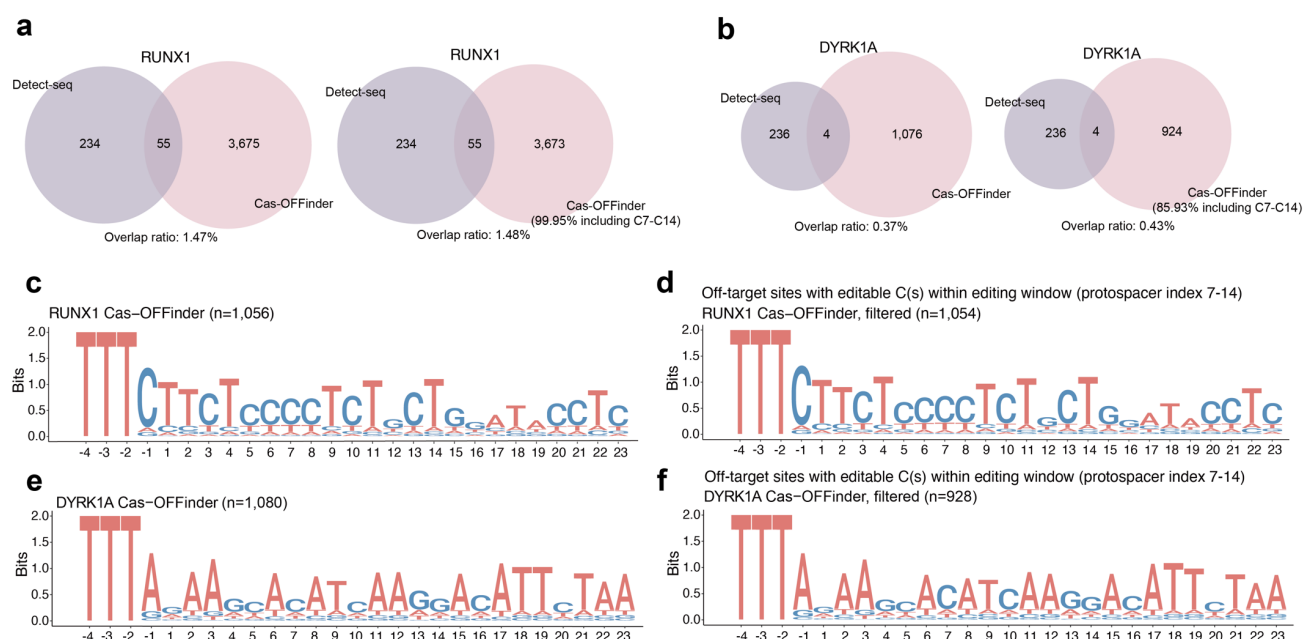
a, Correlation plots of Detect-seq signals and validated editing rates for RUNX1 and DYRK1A off-target sites induced by Cas12a-rA1 CBE in HEK293T cells. Each data point represents an individual edited cytosine (C). The Spearman correlation coefficient is indicated. P value were determined using a two-sided Spearman's rank correlation test. Shaded regions of the linear fit represent 95% confidence intervals. For each point, two biologically independent samples were used (n = 2). **b**, Validation of representative off-target sites identified by Detect-seq. Left, C:G-to-T:A editing rates measured by

targeted amplicon sequencing in HEK293T cells treated with Cas12a-rA1 and RUNX1 or DYRK1A gRNA. Middle, Detect-seq signals in HEK293T cells treated with Cas12a-rA1 and RUNX1 or DYRK1A gRNA. Right, Detect-seq signals in HEK293T cells treated with the GFP control. Bars indicate the mean of two independent biological samples ($n = 2$). BG, background. Detect-seq signals in **a** and **b**, as well as a subset of targeted amplicon-sequencing results in **a**, were taken from our previously published datasets²⁹; all other targeted amplicon-sequencing measurements were generated in this study.



Supplementary Fig. 54 | Sequence logos for off-target sequences of Cas12a-based CBE identified by Detect-seq.

a. Sequence logos illustrating the nucleotide conservation of off-target sites identified in HEK293T cells transfected with Cas12a-rA1 CBE and either RUNX1 or DYRK1A gRNA. **b.** Sequence logos of off-target sequences induced by Cas12a-A3A CBE with either RUNX1, DYRK1A, RAG2, or VEGFA-b gRNA.



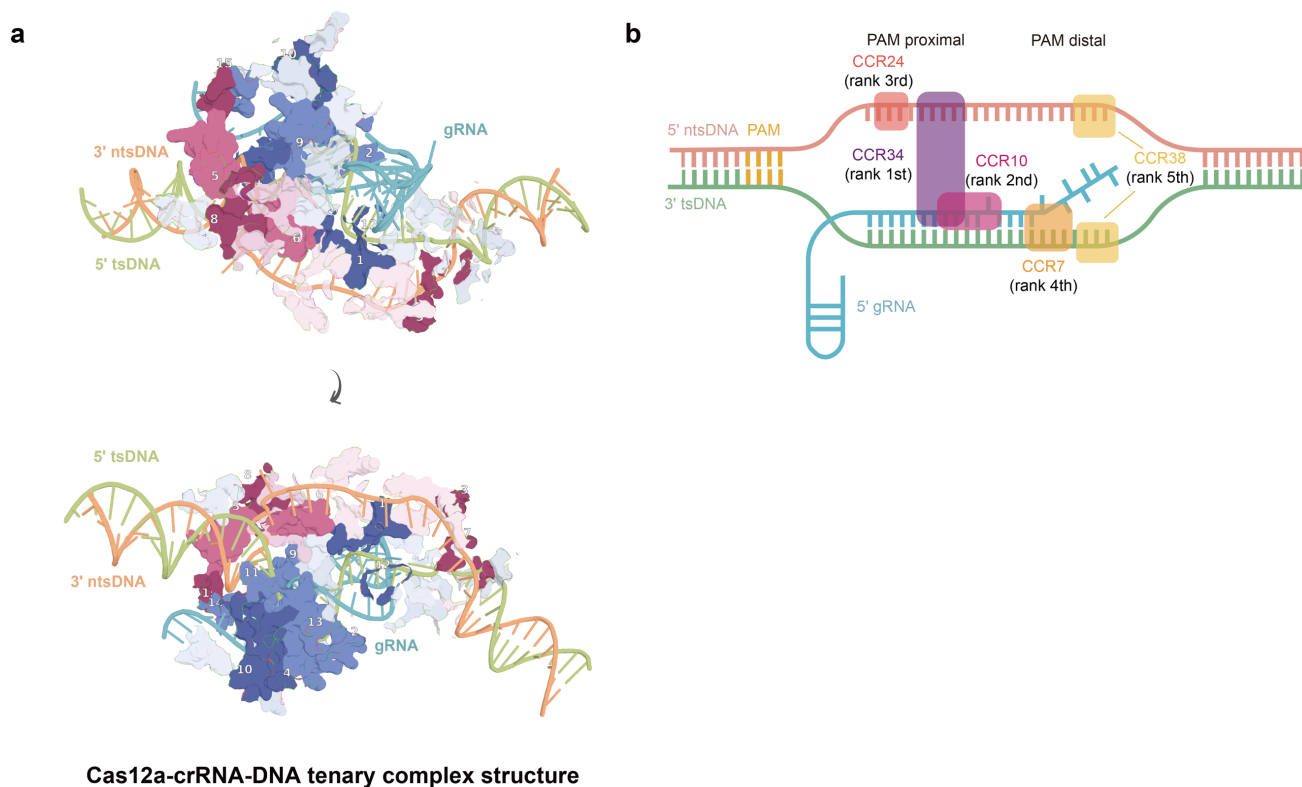
Supplementary Fig. 55 | Comparison of Detect-seq-identified Cas12a-based CBE off-targets with in silico predicted Cas12a off-target profiles.

a, b, Venn diagrams showing the comparison of Cas12a-rA1 CBE off-targets identified by Detect-seq with those reported by Cas-OFFFinder for RUNX1 (**a**) and DYRK1A (**b**) gRNAs. Left, overlap using all GUIDE-seq or AID-seq off-target sites; right, overlap restricted to Cas-OFFFinder-identified off-target sites containing at least one editable cytosine (C) within the canonical editing window (spacer positions 7-14). The overlap ratios represent the proportion of Cas-OFFFinder off-target sites that are also identified by Detect-seq. **c-f**, Sequence logos showing off-target sequences reported by Cas-OFFFinder before (**c, e**) and after (**d, f**) filtering for editable C(s) within the editing window. Although the filtering increased the overlap ratio of off-target sites among Cas-OFFFinder and Detect-seq, the resulting off-target landscapes still differ substantially from the Cas12a-CBE off-targets observed in cellular contexts (Supplementary Fig. 54).



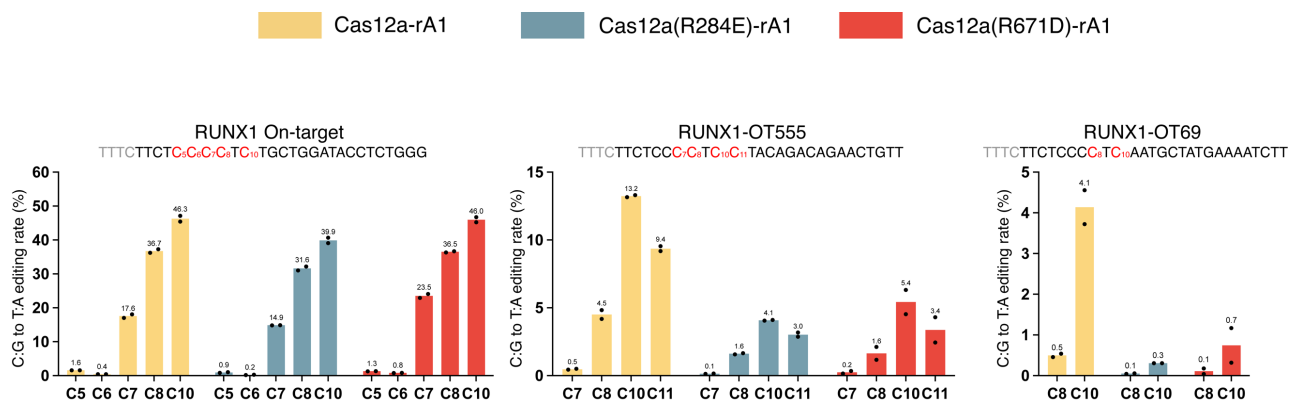
Supplementary Fig. 56 | CCRs of Cas12a-based CBE across Cas12a residues.

Heatmap of CP correlations among all Cas12a residues in contact with DNA/RNA across the Cas12a-rA1 CBE off-target training dataset (from gRNA RUNX1 and DYRK1A; $n = 1,185$), revealing clustered CCRs. The green lines under the residue indices delineate the zoomed-in regions shown in Fig. 5b, and Cas12a domain annotations are indicated below the heatmap.



Supplementary Fig. 57 | Spatial distribution of Cas12a CCRs from Cas12a-based CBEs mapped onto the Cas12a-DNA-gRNA ternary off-target complexes.

a, Structural mapping of the identified Cas12a CCRs on a Cas12a-DNA-gRNA off-target complex. The guide RNA (gRNA) is colored light blue, the target strand DNA light green, and the non-target strand DNA light orange. CCRs located at the Cas12a-tsDNA-gRNA interfaces are depicted in the blue spectrum, whereas those located at the Cas12a-ntsDNA interfaces are colored in the red spectrum. The top 15 high-ranking CCRs are highlighted as opaque surfaces and numerically labeled (1-15) with dark-outlined white text, while the remaining lower-ranked CCRs are rendered with 25% transparency. The bottom panel show an alternative view rotated as indicated by the arrows, allowing optimal visualization of the hidden CCRs in the three-dimensional architecture. The spatial assignment of CCRs was performed as detailed in Supplementary Fig. 30a. **b**, Schematic representation of the positions and interactions of the top 5 CCRs relative to the R-loop in the Cas12a-DNA-gRNA complex.



Supplementary Fig. 58 | Specificity improvement observed for Cas12a-CBE variants.

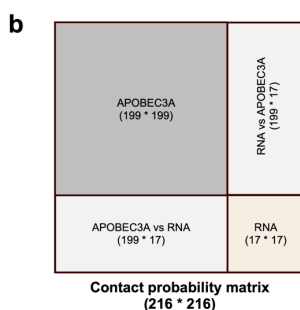
The editing efficiencies of Cas12a-rA1 CBE, Cas12a(R284E)-rA1 and Cas12a(R671D)-rA1 CBE in HEK293T cells at the on-target site and two representative off-target sites for the RUNX1 gRNA, measured by targeted amplicon sequencing. The corresponding reference sequences for each locus is shown. PAM sequences are indicated in faint grey, while the edited cytosines are marked in red with subscript numbers denoting their positions in the protospacer. Each dot represents an independent biological replicate ($n = 2$), and mean editing efficiencies are indicated above each bar.

a

Constant sequence:
Chain A: APOBEC3A length = 199

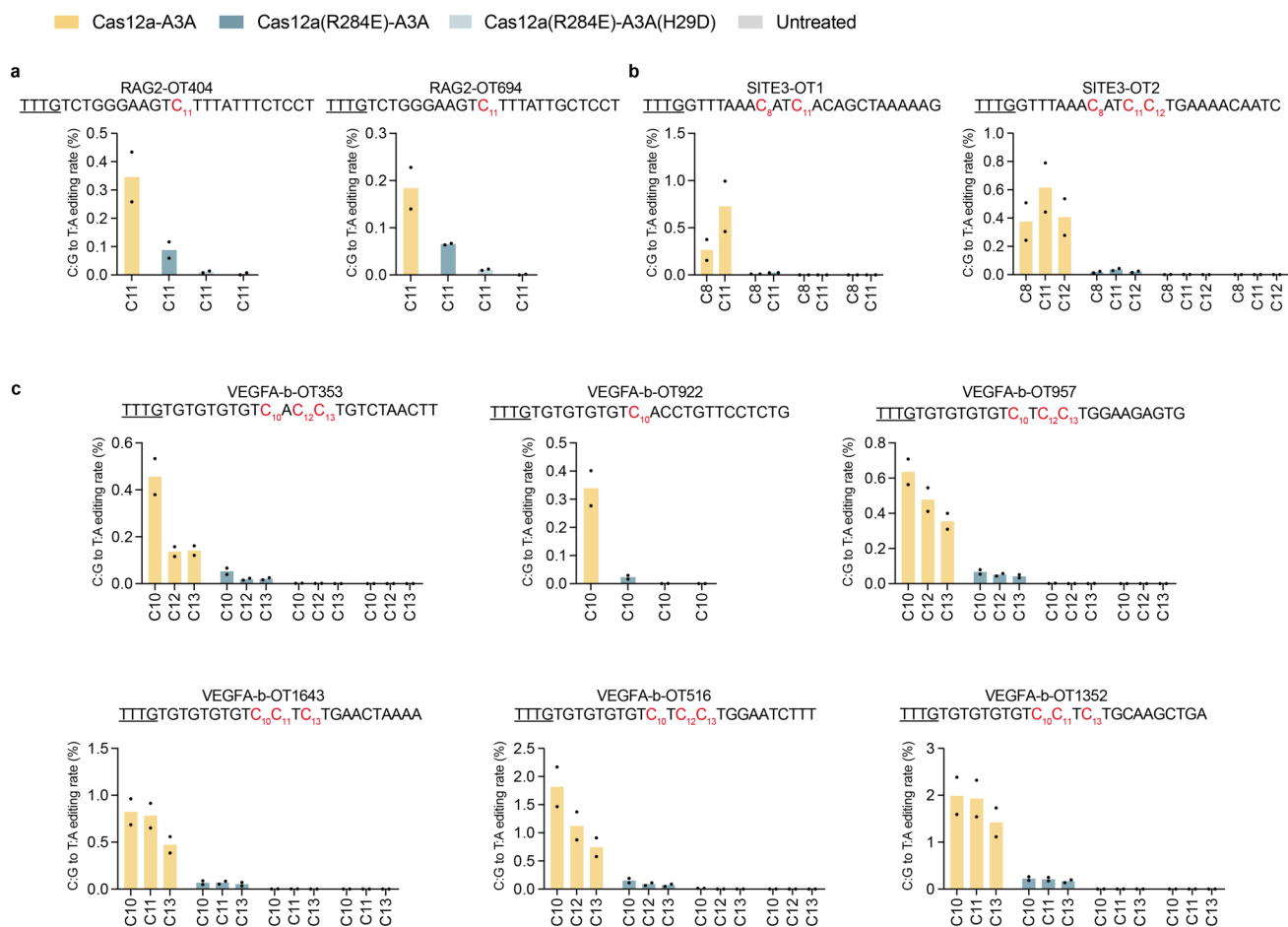
Variable sequence:
Chain B: RNA length = 17 (edited C 1nt + 5' extend 8nt + 3' extend 8nt)

Total token number = 199 + 17 = 216



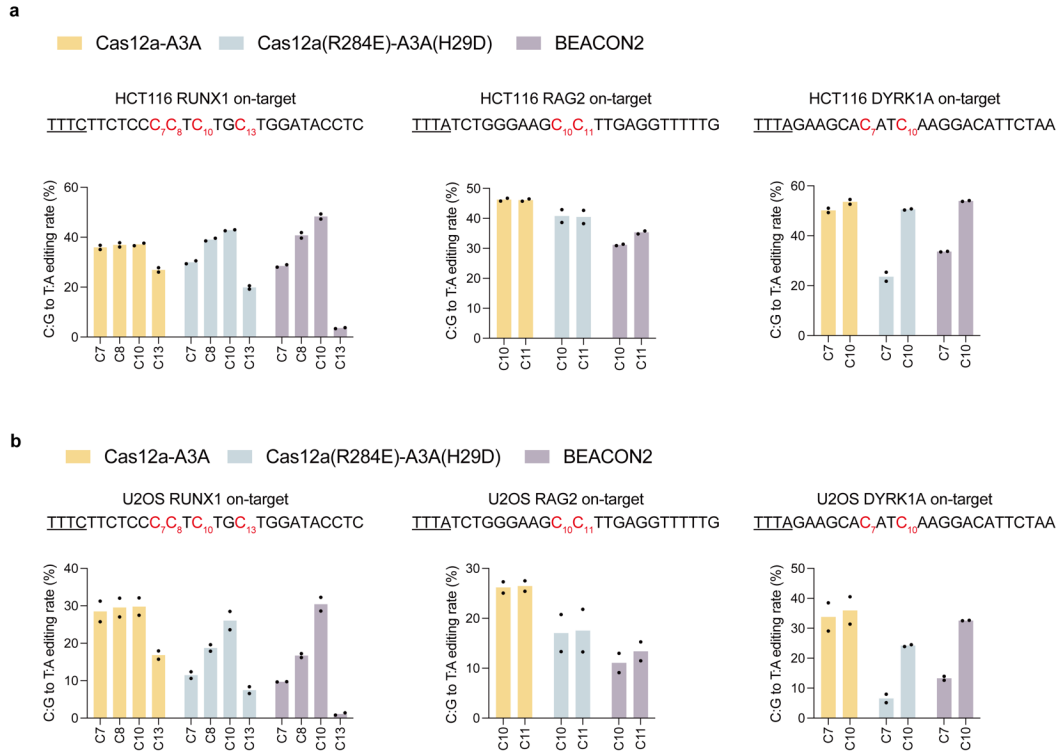
Supplementary Fig. 59 | Token of AF3 predictions for APOBEC3A-RNA binary complex.

a, Illustration of the inputs provided to AF3 for APOBEC3A-RNA complex prediction. The A3A protein sequence is constant in all complexes, whereas the RNA sequences varied depending on the specific off-target site. The lengths of each input component are indicated. **b**, Detailed representation of the contact probability matrix illustrating the inter- and intra-macromolecular contact probability values among the two components of the A3A-RNA complex.



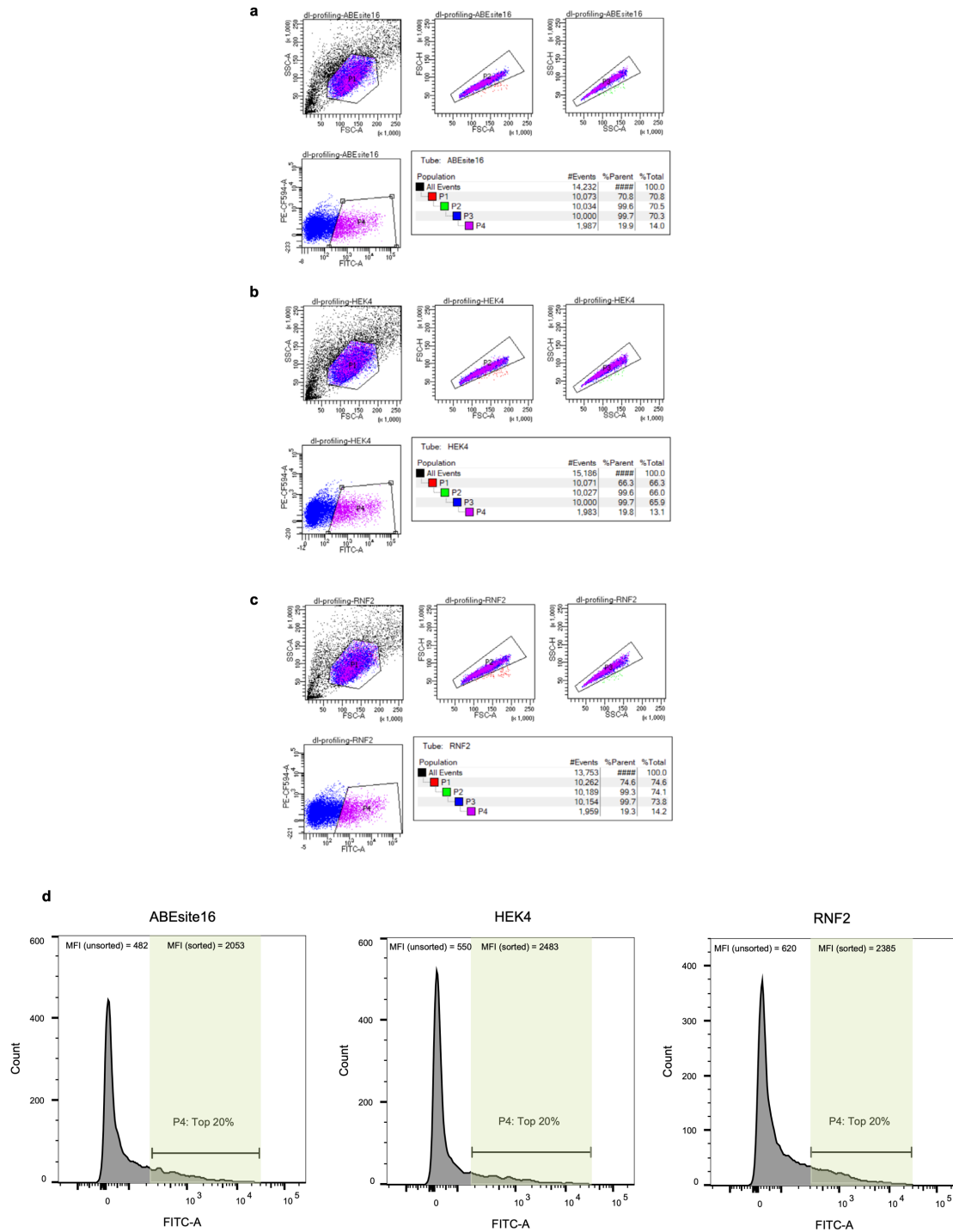
Supplementary Fig. 60 | Specificity improvement of Cas12a(R284E)-A3A and Cas12a(R284E)-A3A(H29D) on genomic sites that were not used for ContactSeek modeling.

a-c, Off-target editing rates of Cas12a-A3A, Cas12a(R284E)-A3A and Cas12a(R284E)-A3A(H29D) at representative DNA off-target sites of RAG2 (**a**), SITE3 (**b**) and VEGFA-b (**c**) gRNAs, which were not used for ContactSeek modeling. The subscripted red C denotes the edited C position within the protospacer and the PAM sequence is underlined. Each dot represents an independent biological replicate (n = 2); bars indicate the mean.



Supplementary Fig. 61 | On-target editing rates for Cas12a-A3A CBEs in HCT116 and U2OS cell lines.

a-b, On-target editing rates of Cas12a-A3A, Cas12a(R284E)-A3A(H29D) and BEACON2 at three endogenous genomic loci in HCT116 (**a**) and U2OS (**b**) cell lines. The subscripted red C denotes the edited C position within the protospacer and the PAM sequence is underlined. Each dot represents an independent biological replicate (n = 2); bars indicate the mean.

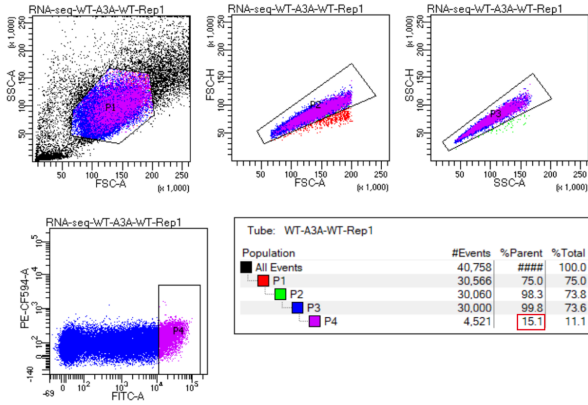


Supplementary Fig. 62 | FACS gating strategy for assessing the impact of ABE expression on di-profiling signals.

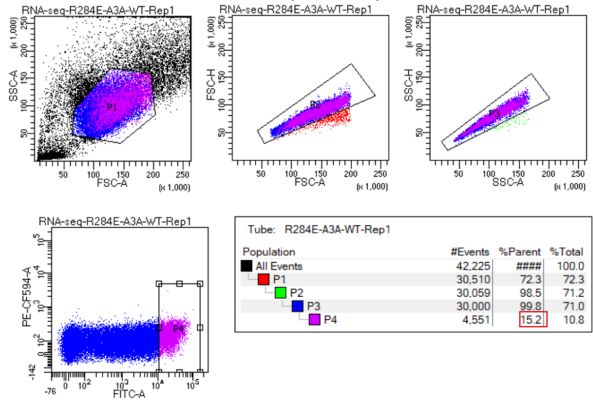
a-c, HEK293T cells transfected with ABE8e-P2A-EGFP together with the ABEsite16 (**a**), HEK4 (**b**) or RNF2 (**c**) sgRNA were dissociated into single-cell suspensions by enzymatic detachment and

analyzed by flow cytometry. Cells were sequentially gated to exclude debris and doublets, and the P4 population, corresponding to the top 20% of cells by EGFP fluorescence intensity (FITC-A), was collected for dI-profiling. Numbers in the tables indicate the count and proportion of events in each gate. **d**, Histograms of EGFP (FITC-A) fluorescence intensity for the corresponding P3 populations before sorting, using the ABEsite16, HEK4 or RNF2 sgRNA, respectively. Shaded regions indicate the P4 gate used for sorting, corresponding to the top 20% of cells by EGFP fluorescence intensity. Mean fluorescence intensity (MFI) values for the unsorted P3 populations and sorted P4 populations are indicated in each panel.

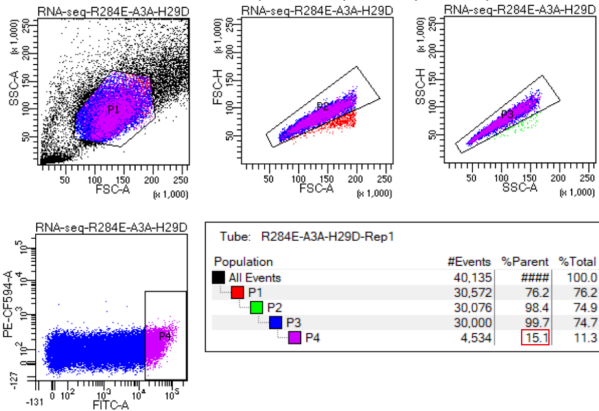
Cas12a-A3A



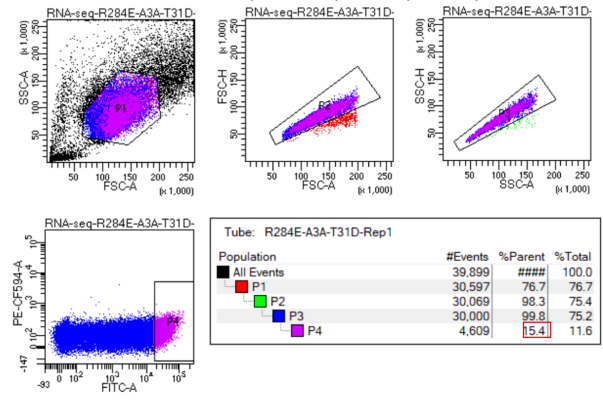
Cas12a(R284E)-A3A



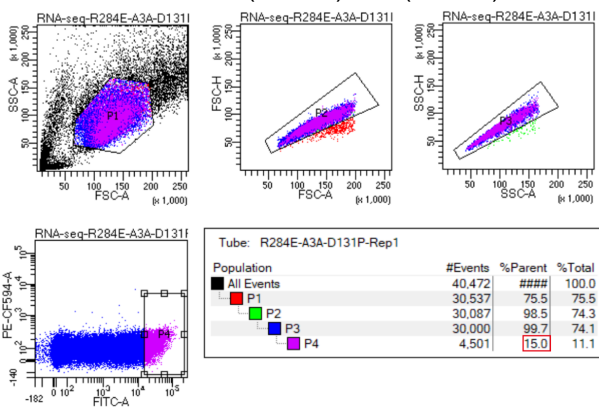
Cas12a(R284E)-A3A(H29D)



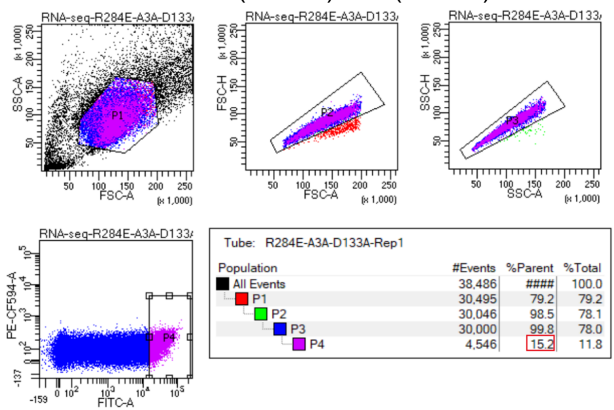
Cas12a(R284E)-A3A(T31D)



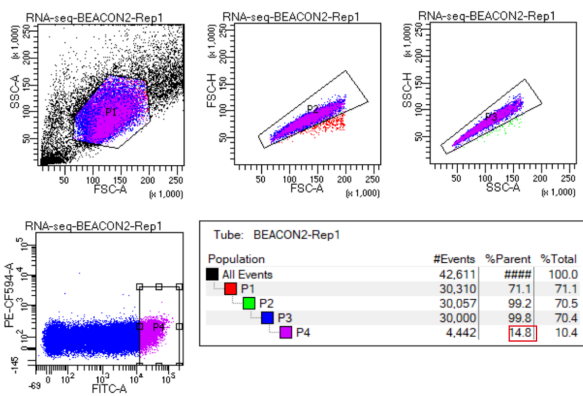
Cas12a(R284E)-A3A(D131P)



Cas12a(R284E)-A3A(D133A)



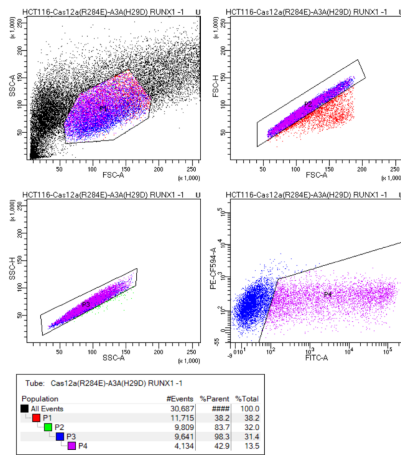
BEACON2



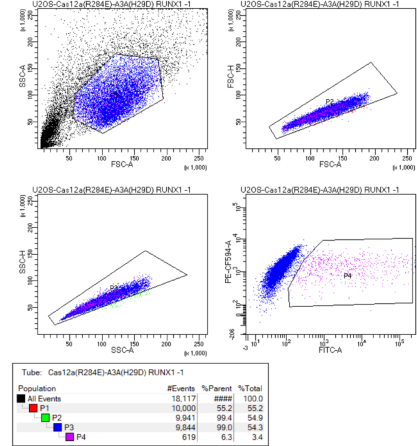
Supplementary Fig. 63 | Gating strategy of FACS for RNA-seq.

HEK293T cells transfected with Cas12a-based CBEs fused to P2A-EGFP were dissociated into single-cell suspensions by enzymatic detachment and analyzed by flow cytometry. Cells were sequentially gated to exclude debris and doublets, and the P4 population, corresponding to the top 15% of cells by EGFP intensity (FITC-A), was collected for RNA-seq. Numbers in the tables indicate the count and proportion of events in each gate. Red boxes indicate the percentage of the P4 population within the P3 gate.

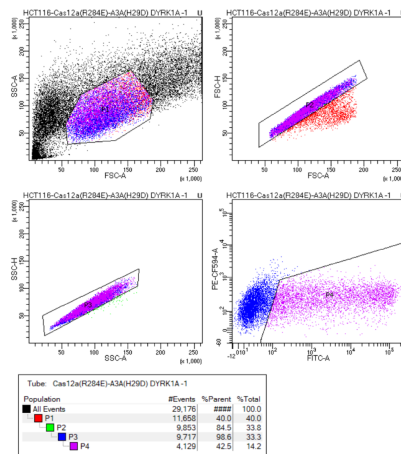
HCT116 Cas12a(R284E)-A3A(H29D) RUNX1



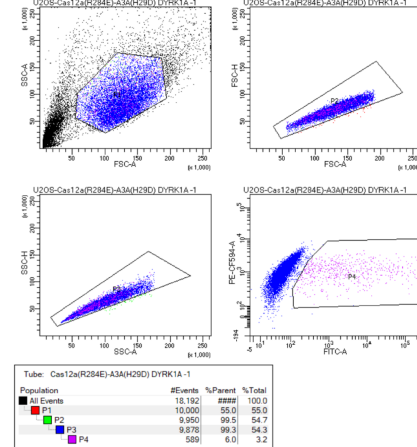
U2OS Cas12a(R284E)-A3A(H29D) RUNX1



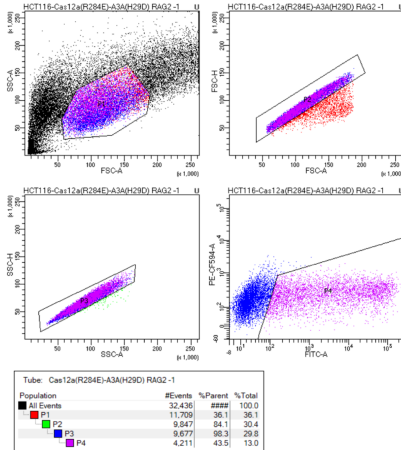
HCT116 Cas12a(R284E)-A3A(H29D) DYRK1A



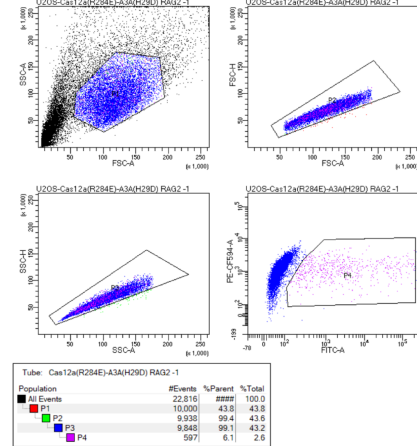
U2OS Cas12a(R284E)-A3A(H29D) DYRK1A



HCT116 Cas12a(R284E)-A3A(H29D) RAG2



U2OS Cas12a(R284E)-A3A(H29D) RAG2



Supplementary Fig. 64 | Representative FACS gating strategy for HCT116 and U2OS cell lines. HCT116 and U2OS cells were transfected with Cas12a-based CBEs fused to P2A-EGFP and cultured for 72 h. Adherent cells were then dissociated into single-cell suspensions by enzymatic detachment and analyzed by flow cytometry. Cells were sequentially gated to exclude debris and doublets, and the P4 population, corresponding to EGFP-positive cells, was collected for targeted amplicon sequencing.

Numbers in the tables indicate the count and proportion of events in each gate.

Supplementary Notes

Protein sequences and guide RNA or pegRNA constructs

SpCas9 nuclease:

Grey: SpCas9 Purple: P2A Green: EGFP Yellow: NLS

K1020 residue in SpCas9 indicated in red

MKRTADGSEFESP~~KKR~~KVDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDS
GETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRL EESFLVEEDKKHERHPIFGNIVDEVAY
HEKYPTIYHLRKKLV DSTDKADLR LIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPI
NASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKS NFDLAEDAKLQLSKD TYDDD
LDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLP EKY
KEIFFDQSKNGYAGYIDGGASQE EYKF I KPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELH
AILRRQEDFY PFLKDNREKIEKILTFRIPYYVGPLARGNSRF AWMTRKSEETITPWNFE EVVDKGASAQSFIE
RMTNFDKNLPNEKVLPHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKE
DYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDI EDIVLTLT LFEDREMIEERL KTY
AHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNF MQLIHDDSLTFKEDIQKAQ
VSGQGDSLHEHIANLAGSPAIKKGILQTVKVVD ELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIE
EGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLT
RSDKNRGKSDNVPSEEVVKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHV
AQILDSRMNTKYDENDKLIREVKVITLKS KLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVG TALIKKYPKL
ESEFVYGDYKVYDVR~~K~~MIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKG
RDFATVRKVL SMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWD PKKYGGFDSPTVAYSVLV VAKVE
KGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKG
NELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKR VILADANLDKVL SAYNK
HRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQ SITGLYETRIDLSQLGGDS
GGSKRTADGSEFEP~~KKR~~KVGS~~G~~ATNFSLLKQAGDVEENPGPMVSKGEELFTGVVPILVELDGDVNGHKFSVS
GEGEGDATYGKLT LKFICTTGKLPVPWPTLVTTLT YGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDD

GNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNNSHNVYIMADKQNGIKVNFKIRHNIEDGSV
QLADHYQQNTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLLEFVTAAGITLGMDELYKSGGS P K K K R K V

*

BE4max:

Grey: SpCas9 Blue: rAPOBEC1 Yellow: NLS Brown: UGI

K1020 residue in SpCas9 indicated in red

MKRTADGSEFES P K K K R K V SSETGPVAVDPTLRRRIEPHEFEVFFDPREL RKETCLLYEINWGRHSIWRHTS
QNTNKHVEVNFIEKF T T E R Y F C P N T R C S I T W F L S W S P C G E C S R A I T E F L S R Y P H V T L F I Y I A R L Y H H A D P R N R
QGLRDLISSGV TI Q I M T E Q E S G Y C W R N F V N Y S P S N E A H W P R Y P H L W V R L Y V L E L Y C I I L G L P P C L N I L R R K Q P
Q L T F F T I A L Q S C H Y Q R L P P H I L W A T G L K S G G S S G G S S G S E T P G T S E S A T P E S S G G S S G G S D K K Y S I G L A I G T N
S V G W A V I T D E Y K V P S K K F K V L G N T D R H S I K K N L I G A L L F D S G E T A E A T R L K R T A R R R Y T R R K N R I C Y L Q E I F S
N E M A K V D D S F F H R L E E S F L V E E D K K H E R H P I F G N I V D E V A Y H E K Y P T I Y H L R K K L V D S T D K A D L R L I Y L A L A H
M I K F R G H F L I E G D L N P D N S D V D K L F I Q L V Q T Y N Q L F E E N P I N A S G V D A K A I L S A R L S K S R R L E N L I A Q L P G E K
K N G L F G N L I A L S L G L T P N F K S N F D L A E D A K L Q L S K D T Y D D D L D N L L A Q I G D Q Y A D L F L A A K N L S D A I L L S D I L
R V N T E I T K A P L S A M I K R Y D E H H Q D L T L L K A L V R Q Q L P E K Y K E I F F D Q S K N G Y A G Y I D G G A S Q E E F Y K F I K P I
L E K M D G T E E L L V K L N R E D L L R K Q R T F D N G S I P H Q I H L G E L H A I L R R Q E D F Y P F L K D N R E K I E K I L T F R I P Y Y V
G P L A R G N S R F A W M T R K S E E T I T P W N F E E V D K G A S A Q S F I E R M T N F D K N L P N E K V L P K H S L L Y E Y F T V Y N E L T
K V K Y V T E G M R K P A F L S G E Q K K A I V D L L F K T N R K V T V K Q L K E D Y F K K I E C F D S V E I S G V E D R F N A S L G T Y H D L L
K I I K D K D F L D N E E N E D I L E D I V L T L T L F E D R E M I E E R L K T Y A H L F D D K V M K Q L K R R R Y T G W G R L S R K L I N G I R
D K Q S G K T I L D F L K S D G F A N R N F M Q L I H D D S L T F K E D I Q K A Q V S G Q G D S L H E H I A N L A G S P A I K K G I L Q T V K V V
D E L V K V M G R H K P E N I V I E M A R E N Q T T Q K G Q K N S R E R M K R I E E G I K E L G S Q I L K E H P V E N T Q L Q N E K L Y L Y Y L Q
N G R D M Y V D Q E L D I N R L S D Y D V D H I V P Q S F L K D D S I D N K V L T R S D K N R G K S D N V P S E E V V K K M K N Y W R Q L L N A K
L I T Q R K F D N L T K A E R G G L S E L D K A G F I K R Q L V E T R Q I T K H V A Q I L D S R M N T K Y D E N D K L I R E V K V I T L K S K L V
S D F R K D F Q F Y K V R E I N N Y H H A H D A Y L N A V V G T A L I K K Y P K L E S E F V Y G D Y K V Y D V R K M I A K S E Q E I G K A T A K Y
F F Y S N I M N F F K T E I T L A N G E I R K R P L I E T N G E T G E I V W D K G R D F A T V R K V L S M P Q V N I V K K T E V Q T G G F S K E S

ILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSEKNPIDFL
 EAKGYKEVKKDLIIKLPKYSLEFENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQK
 QLFVEQHKHYLDEIIEQISEFSKRVLADANLDKVL SAYNKH RDKPIREQAENIIHLFTLTNLGAPAAFKYFD
 TTIDRKRYTSTKEVL DATLIHQ SITGLYETRIDLSQLGGDSGGSGGSGGSGGSTNLSDIIEKETGKQLVIQESILM
 LPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSGGSGGSGTNL
 SDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNG
 ENKIKMLSGGSKRTADGSEFEPKKKRKV

ABE8e:

Grey: SpCas9 Blue: TadA8e Purple: P2A Green: EGFP Yellow: NLS

F156 residue in TadA8e indicated in red

K1020 residue in SpCas9 indicated in red

MKRTADGSEFESPKKKRKVSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVNNRVIGEGWNRAIGLHDPT
 AHAEIMALRQGGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNSKRGGAAGSLMNVLNYPGMNH
 RVEITEGILADECAALLCDFYRMPRQVFNAQKKAQSSINSGGSSGGSSGSETPGTSESATPESSGGSSGGSDK
 KYSIGLAIGTNSVGWAVITDEYKVP SKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRK
 NRICYLQEIFS NEMAKVDDSF FHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLV DSTDKA
 DLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLF EENPINASGVDAKAILSARLSKSRR
 ENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKN
 LSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGAS
 QEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIE
 KILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEVVDKGASAQSFIERMTNFDKNLPNEKVL PKHSL
 YEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRF
 NASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTTLTFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG
 RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA
 KKGILQTVKVVDLVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIE EGIKELGSQILKEHPVENTQL

QNEKLYLYYLQNGRDMYVDQEELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKM
 KNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIRE
 VKVITLKSCLVSDFRKDFQFYKVVREINNYHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVYDVR^KMIAKS
 EQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIKKKT
 EVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVAKVEKGKSKKLKSVKELLGITIMER
 SSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLELENGRKRMLASAGELQKGNELALPSKYVNFYLYLASHYEK
 LKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVLADANLDKVL SAYNKH RDKPIREQAENIIHLFTLTN
 LGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQ SITGLYETRIDLSQLGGD^{SGGS}^{KRTADGSEFEPKKKRKV}^G
^{SGATNFSLLKQAGDVEENPGPMVSKGEELFTGVVPI}^{LV}^{ELDGDVNGHKFSVSGEGEGDATYGKLT}^{TLKFI}^{CTTG}
^{KLPVPWPTLVTT}^{LT}^{YGVQCFSRYPDHMKQHDF}^{FKS}^{AMPEGYVQERTIFFKDD}^{GN}^{YKTRA}^{EV}^{KFE}^{GDTLVN}^{RIE}
^{LKGIDFKEDGNILGHKLEYNYN}^{SHNVYIMADKQKNGIKVNF}^{KIRHNI}^{EDGSVQLADHYQQNTPI}^{GDGPVLLPD}
^{NHYLSTQSALS}^{KDPNEKRDHMLLEFVTAAGITL}^{GMDELYK}^{SGGS}^{PKKKRKV}^{*}

ABE7.10:

Grey: SpCas9 Dark blue: wildtype TadA Blue: TadA7.10 Purple: P2A Green: EGFP Yellow: NLS

^{MKRTADGSEFES}^{PKKKRKMSEVEFSHEYWMRHAL}^{TLAKRAWDERE}^{VPVGAVLVHNNRVIGEGWNRPIGRHDP}
^{TAHAEIMALRQGG}^{LMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMN}
^{HRVEITEGILADECAALLSDFFRMR}^{QEIKAQKKAQS}^{TD}^{SGGSSGGSSGSETPGTSESATPESSGGSSGGSS}
^{EVEFSHEYWMRHAL}^{TLAKRARDERE}^{VPVGAVLVHNNRVIGEGWNRRAIGLHDPTAHAEIMALRQGG}^{LMQNYRL}
^{IDATLYVTLEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYF}
^{FRMPRQVFNAQKKAQS}^{TD}^{SGGSSGGSSGSETPGTSESATPESSGGSSGGSSDKKYSIGLAIGTNSVGWAVITD}
 EYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDS
 FFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFL
 IEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLI
 ALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKA
 PLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEE

LLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPYYVGPLARGNSR
FAWMTRKSEETITPWNFEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGM
RKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFL
DNEENEDILEDIVLTTLTFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTIIL
DFLKSDGFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIAKKGILQTVKVVDELVKVMGR
HKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQ
ELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKMKMKNYWRQLLNAKLITQRKFDN
LTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQF
YKVVREINNYHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVYDVRKMKIAKSEQEIGKATAKYFFYSNIMNF
FKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIKKTEVQTGGFSKESILPKRNSDK
LIARKKDWDPKKYGGFDSPTVAYSVLVVAKEVGKSKKLKSVKELLGITIMERSSEKPNIDFLEAGYKEVK
KDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFYLYLASHYEKLGKSPEDNEQQLFVEQHKH
YLDEIIEQISEFSKRVILADANLDKVL SAYNKH RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYT
STKEVL DATLIHQ SITGLYETRIDL SQLGGD **SGGS** **KRTADGSEFEPKKRKY** **SGSATNFSLLKQAGDVEENPG**
PMVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLT LKFICTTGKLPVPWPTLVTTLT YGVQCF
SRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYN
YN SHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDH
MVLLFEVTAAGITLGMDELYKSGGS **PKKKRKY***

ABEmax:

Grey: SpCas9 Blue: TadA7.10max Purple: P2A Green: EGFP Yellow: NLS

MKRTADGSEFESPKKKRKY **SEVEFSHEYWMRHALTLAKRARDEREVPVGAVLV LNNRVIGEGWNRAIGLHDPT**
AHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNH
RVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSD **SGGSSGGSSGSETPGTSESATPESSGGSSGGSDK**
KYSIGLAIGTNSVGWAVITDEYKVP SKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRK
NRICYLQEIFS NEMAKVDDSF FHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLV DSTDKA
DLRLIYLALAHMIKFRGHFLIEGDLNPDNSDV DKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKS RRL

ENLIAQLPGEKKNGLFGNLIASLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKN
 LSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGAS
 QEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIE
 KILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEVVDKGASQSFIERMTNFDKNLPNEKVLPKHSL
 YEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRF
 NASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTTLTFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG
 RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAI
 KKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL
 QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKM
 KNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIRE
 VKVITLKSCLVSDFRKDFQFYKVBREINNYHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVYDVRKMIAS
 EQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIKK
 EVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVAKVEKGKSKKLKSVKELLGITIMER
 SSFEKNPIDFLEAGYKEVKKDLIIKLPKYSLEFENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEK
 LKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVLADANLDKVL SAYNKHARDKPIREQAENIIHLFTLTN
 LGAPAAFKYFDTTIDRKRYTSTKEVLDTLIHQSI TGLYETRIDLSQLGGD**SGGS****KRTADGSEFEPKKKRKV**
SGATNFSLLKQAGDVEENPGPMVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTCLKFICTTG
KLPVPWPTLVTTLTYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIE
LKGIDFKEDGNILGHKLEYNYN SHNVYIMADKQKNGIKVNFKIRHNIEDG SVQLADHYQQNTPIGDGPVLLPD
NHYLSTQSALSKDPNEKRDHMLLEFVTAAGITLGMDELYKSGGS**PKKKRKV***

ABE8.8:

Grey: SpCas9 **Blue:** TadA8.8 **Purple:** P2A **Green:** EGFP **Yellow:** NLS

MKRTADGSEFESPKKKRKVSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLN RVIGEGWNRAIGLHDPT
AHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNH
RVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTD**SGGSSGGSSGSETPGTSESATPESSGGSSGGSDK**
 KYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRK

NRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLV DSTDKA
DLRLIYLALAHMIKFRGHFLIEGDLNPDNSDV DKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRL
ENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKN
LSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGAS
QEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPLKDNREKIE
KILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPHKSL
YEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRF
NASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTTLTFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG
RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAI
KKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL
QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKM
KNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIRE
VKVITLKSCLVSDFRKDFQFYKVVREINNYHHAHDAYLNAVVG TALIKKYPKLESEFVYGDYKVYDVRKMIAS
EQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKT
EVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSP TVAYSVLVAKVEKGKSKKLKSVKELLGITIMER
SSFEKNPIDFLEAGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEK
LKGSPEDNEQKQLFVEQHKHYLDEIIIEQISEFSKRVLADANLDKVL SAYNKH RDKPIREQAENIIHLFTLTN
LGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQ SITGLYETRIDLSQLGGD SGGSKRTADGSEFEPKKKRKV
SGATNFSLLKQAGDVEENPGPMVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLT LKFICTTG
KLPVPWPTLVTTLT YGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRA EVKFEGDTLVNRIE
LKGIDFKEDGNILGHKLEYNYN SHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPD
NHYLSTQSALSKDPNEKRDHMLLEFVTAAGITLGMDELYKSGGSPKKKRKV*

Cas12a-rA1:

Grey: LbCas12a Blue: rAPOBEC1 Yellow: NLS Brown: UGI

R284 in LbCas12a is indicated in red

MPKKKRKVSSETGPVAVDPTLRRRIEPHEFEVFFDPRELKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNF
IEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGV
TIQIMTEQESGYCWRNFMVNYSPSNEAHWPYPHLLWRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQS
CHYQRLPPHILWATGLKSGSETPGTSESATPESMSKLEKFTNCYLSKTLRFKAIPVGKTQENIDNKRLLED
EKRAEDYKGVKKLLDRYYLSFINDVLHSLKLNLYISLFRKKTRTEKENKELENLEINLRKEIAKAFKGNE
GYKSLFKKDIETILPEFLDDKDEIALVNSFNGFTTAFTGFFDNRENMFSEEAKSTSIAFRGINENLTRYISN
MDIFEKVDAIFDKHEVQEIKEILNSDYDVEDFFEGEFFNFVLTQEGIDVYNAIIGGFVTESGEKIKGLNEYI
NLYNQKTKQKLPKFKPLYKQVLSDR^{ESLSFYGEGYTSDEEVLEVFRNTLNKNSEIFSSIKKLEKLFKNFDEYS}
SAGIFVKNGPAISTISKDIFGEWNVIRDKWNAEYDDIHLKKKAVVTEKYEDDRKSFKKIGSFSLEQLQEYAD
ADLSVVEKLKEIIQKVDEIYKVYGSSEKLFDAFVLEKSLKKNDAVVAIMKDLLDSVKSFENYIKAFFGEGK
ETNRDESFYGDFVLAYDILLKVDHIYDAIRNYVTQKPYSKDKFKLYFQNPQFMGGWDKDKETDYRATILRYGS
KYYLAIMDKKYAKCLQKIDKDDVNGNYEKINYKLLPGPNKMLPKVFFSKKWMAYYNPSEDIQKIYKNGTFKKG
DMFNLNDCHKLIDFFKDSISRYPKWSNAYDFNFSETEKYKDIAGFYREVEEQGYKVSFESASKKEVDKLVEEG
KLYMFQIYNKDFSDKSHGTPNLHTMYFKLLFDENNHGQIRLSGGAELFMRRASLKKEELVHPANSPIANKNP
DNPKKTTTTLSYDVYKDKRFSEDQYELHIPIAINKCPKNIFKINTEVRVLLKHDDNPYVIGIARGERNLLYIV
VDGKGNIVEQYSLNEIINNFMGIRIKTDYHSLLDKKEKERFEARQNWTSIENIKELKAGYISQVVKICELVE
KYDAVIALADLNSGFKNRVRKVEKQVYQKFEKMLIDKLNMYVDKKSNPCATGGALKGYQITNKFESFKSMSTQ
NGFIFYIPAWLTSKIDPSTGFVNLLKTKYTSIADSKKFISFDRIMYVPEEDLFEFALDYKNFSRTDADYIKK
WKLYSYGNRIRIFRNPKNVFDWEEVCLTSAYKELFNKYGINYQQGDIRALLCEQSDKAFYSSFMALMSML
QMRNSITGRTDVAFLISPVKNSDGIFYDSRNYEAQENAILPKNADANGAYNIARKVLWAIGQFKKAEDEKLDK
VKIAISNKEWLEYAQTSVKHGS^{PKKKRKV}SGGSTNLS^{SDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDI}
^{LVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGS}^{PKKKRKV}*

Cas12a-A3A:

Grey: LbCas12a Blue: APOBEC3A Yellow: NLS Brown: UGI Purple: P2A Green: EGFP

^{H29} in APOBEC3A is indicated in red

R284 in LbCas12a is indicated in red

MPKKRKVEASPASGPRHLMDPHIFTSNFNNGIGRHKTYLCYEVERLDNGTSVKMDQHRGFLHNQAKNLLCGF
YGRHAELRFLDLVPSLQLDPAQIYRVTFISWSPCFSWGCAEVRAFLQENTHVRLRIFAARIYDYDPLYKEA
LQMLRDAGAQVSIMTYDEFKHCWDTFVDHQGCPFQPWDGLDEHSQALSGRLRAILQNQNSGSETPGTSESAT
PESMSKLEKFTNCYLSKTLRFKAIPVGKTQENIDNKRLLEVEDEKRAEDYKGVKKLLDRYYLSFINDVLHSIK
LKNLNNYISLFRKKTRTEKENKELENLEINLRKEIAKAFKGNEGYKSLFKKDIETILPEFLDDKDEIALVNS
FNGFTTAFTGFFDNRENMFSEEAKSTSIAFRCINENLTRYISNMDIFEKVDAIFDKHEVQEIKEKILNSDYDV
EDFFEGEFFNFVLTQEGIDVYNAIIGGFVTESGEKIKGLNEYINLYNQKTKQKLPKFKPLYKQVLSDRESLSF
YGEGYTSDEEVLEVFRNTLNKNSEIFSSIKKLEKLFKNFDEYSSAGIFVKNGPAISTISKDIFGEWNVIRDKW
NAEYDDIHLKKKAVVTEKYEDDRRSFKKIGSFSLEQLQEYADADLSVVEKLKEIIIQKVDEIYKVYGSSEKL
FDADFVLEKSLKKNDAVVAIMKDLLDSVKSFENYIKAFFGEGKETNRDESFGDFVLAYDILLKVDHIYDAIR
NYVTQKPYSKDKFKLYFQNPQFMGGWDKDKETDYRATILRYGSKYYLAIMDKKYAKCLQKIDKDDVNGNYEKI
NYKLLPGPNKMLPKVFFSKKWMAYNPSEDIQKIYKNGTFKKGDMFNLNDCHKLIDFFKDSISRYPKWSNAYD
FNFSETEKYKDIAGFYREVEEQGYKVSFESASKKEVDKLVEEGKLYMFQIYNKDFSDKSHGTPNLHTMYFKLL
FDENNHGQIRLSGGAELFMRRASLKKEELVHPANSPIANKNPDNPKKTTLSDVYKDKRFSEDQYELHIPI
AINKCPKNIFKINTEVRVLLKHDDNPYVIGIARGERNLLYIVVVDGKGNIVEQYSLNEIINNFNIGIRIKTDYH
SLLDKKEKERFEARQNWTSIENIKELKAGYISQVVKICELVEKYDAVIALADLNSGFKNSRVKVEKQVYQKF
EKMLIDKLNMYVDKKSNPCATGGALKGYQITNKFESFKSMSTQNGFIFYIPAWLTSKIDPSTGFVNLLKTKYT
SIADSKKFISFDRIMYVPEEDLFEFALDYKNFSRTDADYIKKWKLYSYGNRIRIFRNPKNNVFDWEEVCLT
SAYKELFNKYGINYQQGDIRALLCEQSDKAFYSSFMALMSLMLQMRNSITGRTDVAFLISPVKNSDGIFYDSR
NYEAQENAILPKNADANGAYNIARKVLWAIGQFKKAEDEKLDKVKIAISNKEWLEYAQTSVKHGSPPKKRKVS
GGSTNLSDIIEKETGKQLVIQESILMLPEEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALV
IQDSNGENKIKMLS GGSPPKKRKVGSGATNFSLLKQAGDVEENPGPMVSKGEELFTGVVPILVELDGDVNGHK
FSVSGEGEGDATYGKLT LKFICTTGKLPVPWPTLVTTLTYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIF
FKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKVNFKIRHNIE
DGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLLEFVTAAGITLGMDELYKSGGSPPKK
RKV*

PE2 (used for PE2 and PE3 editing):

Grey: SpCas9 Blue: MMLV Yellow: NLS Purple: P2A Magenta: mCherry

K1020 residue in SpCas9 indicated in red

MKRTADGSEFESPKKKRKVDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDS
GETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNIVDEVAY
HEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPI
NASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDITYDD
LDNLLAQIGDQYADFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKY
KEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELH
AILRRQEDFYFPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEVVDKGASAQSFIE
RMTNFDKNLPNEKVLPKHSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKE
DYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTTLTFEDREMIEERLKTY
AHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFQMQLIHDDSLTFKEDIQKAQ
VSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIAMARENQTTQKGQKNSRERMKRIE
EGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLT
RSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHV
AQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVVREINNYHHAHDAYLNAVVGTAIIKKYPKL
ESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKG
RDFATVRKVLSPQVNIIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVAKVE
KGKSKKLKSVKELLGITIMERSSEFKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKG
NELALPSKYVNFYLYLASHYEKLKGSPEDEQKQLFVEQHKHYLDEIIIEQISEFSKRVILADANLDKVL SAYNK
HRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQSI TGLYETRIDLSQLGGDS
GGSSGGSSGSETPGTSESATPESSGGSSGGSSTLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGL
AVRQAPLIIPLKATSTPVSIKQYPMSQEARLGKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQD
LREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTR
LPQGFKNSTPLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCCQQTALLQTLGNLGYRASAKKAQ

ICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTL
FNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTKQLGPWRRPVAYLSKKLDPVAAGW
PPCLRMVAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALN
PATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETETEWAKAL
PAGTSAQRAELIALTQALKMAEGKKLVYTDSDRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALF
LPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSPSSGSKRTADGSEFEPKKKRKV
GSGATNFSLLKQAGDVEENPGPMVSKGEEDNMAIIKEFMRFKVHMEGSVNGHEFEIEGEGEGRPYEGTQTAKL
KVTKGGPLPFAWDILSPQFMYGSKAYVKHPADIPDYLLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGEFI
YKVKLRGTNFPDGPVMQKKTMGWEASSERMYPEDGALKGEIKQRLKLDGGHYDAEVKTTYKAKKPVQLPGA
YNNVIKLDITSHNEDYTIVEQYERAEGRHSTGGMDELYKSSGSPKKKRKV*

PEmax (used for PE2max and PE3max editing):

Grey: SpCas9 Blue: MMLV Yellow: NLS Purple: P2A Magenta: mCherry

K1020 residue in SpCas9 indicated in red

MKRTADGSEFEPKKKRKVDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDS
GETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNIVDEVAY
HEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPI
NASGVDAKAILSARLSKSRKLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDD
LDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKY
KEIFFDQSKNGYAGYIDGGASQEEFYKFIPKILEKMDGTEELLVKLKREDLLRKQRTFDNGSIPHQIHLGELH
AILRRQEDFYFPLKDNREKIEKILTRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEVVDKGASAQSFIE
RMTNFDKNLPNEKVLPKHSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKE
DYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTTLTFEDREMIEERLKTY
AHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFQMQLIHDDSLTFKEDIQKAQ
VSGQGDSLHEHIANLAGSPAIIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIE
EGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLT

RSDKNRGKSDNVPSEEVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGELSELDKAGFIKRQLVETRQITKHV
 AQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKREINNYHHAHDAYLNAVVGTAIIKKYPKL
 ESEFVYGDYKVYDVR**K**MIASEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKG
 RDFATVRKVLSPQVNIIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVAKVE
 KGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAGYKEVKKDLIIKLPKYSLELENGRKRMLASAGELQKG
 NELALPSKYVNFYLYLASHYEKLKGSPEDEQKQLFVEQHKHYLDEIIIEQISEFSKRVLADANLDKVL SAYNK
 HRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQSI TGLYETRIDLSQLGGDS
GGSSGGSKRTADGSEFESP**KKR****KV**SGSSGGST**LN**IEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMG
 LAVRQAPLIIPKATSTPVSIKQYPMSEARLGIKPHIQRLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQ
 DLREVNRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWT
 RLPQGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLAATSELDCQQGTRALLQTLGNLGYRASAKKA
 QICQKQVKYLGILLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGT
 LFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVL TQKLGWRRPVAYLSKKLDPVAAG
 WPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLDTRVQFGPVVAL
 NPATLLPLPEEGLQHNCILDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETETEWAKA
 LPAGTSAQRAELIALTQALKMAEGKKLNVTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKAL
 FLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSP**SGGSKRTADGSEFESP****KKR**
KVGS**GPA****AKR****V****K****L**DGSGATNFSLLKQAGDVEENPGPM**V****S****K****G****E****E****D****N****M****A****I****K****E****F****M****R****F****K****V****H****M****E****G****S****V****N****G****H****E****F****E****I****E****G****E**
 GEGRPYEGTQAKLKVTKGGLPFAWDILSPQFMYGSKAYVKHPADIPDYLLKLSFPEGFKWERVMNFEDGGVV
 TVTQDSSLQDGEFIYVKVLRGTNFSPDGPVMQKKTMGWEASSERMYPEDGALKGEIKQRLKLDGGHYDAEVK
 TTYKAKKPVQLPGAYNVNIKLDITSHNEDYTIVEQYERAEGRHSTGGMDELYK**SGGS****PKKR****KV***

SaCas9(D10A) nickase:

Grey:SaCas9 Yellow: NLS Purple: P2A Magenta: mCherry

MKRTADGSEFESP**KKR****KV**SGSGKRNILGLAIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSK
 RGARRLKRRRRHRIQRVKLLFDYNLLTDHSELGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNE
 VEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSYVKEAKQLLKVKQAYHQLDQS

FIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVITR
DENEKLEYEYEFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEII
ENAELLDQIAKILTIYQSSEDIQEELTNLNSLTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQI
AIFNRLKLVPPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIELAREKNSKDAQ
KMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNPNFNYEVDHIIPR
SVSFDNSFNKVLVKQEENSKGNRTPFQYLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINR
FSVQKDFINRNLVDTRYATRGLMNLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHAEDALI
IANADFIKWKLDKAKKVMENQMFEEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPN
RELINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKLKKLINKSPEKLLMYHHPQTYQKLKLIMEQYGDEKNP
LYKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLCLKPYRFDVYLDNGVYKFV
TVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQAEFIASFYNNDLIKINGELYRVIGVNNDLLNRIEVMIDI
TYREYLENMNDKRPPRIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKGSGGSKRTADGSEFEPPKKKR
KVGSYPYDVPDYAGSGATNFSLLKQAGDVEENPGPMVSKGEEDNMAIIKEFMRFKVHMEGSVNGHEFEIEGEG
EGRPYEGTQTAKLKVTKGGPLPFAWDILSPQFMYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVT
VTQDSSLQDGEFIYKVKLRGTNFPDGPVMQKKTMGWEASSERMYPEDGALKGEIKQRLKLDGGHYDAEVKT
TYKAKKPVQLPGAYNVNIKLDITSHNEDYTIVEQYERAEGRHSTGGMDELYKSGGSPKKKRKV*

SpCas9 sgRNA:

Yellow: U6 promoter Brown: spacer Green: scaffold

The red **G** is the transcription start site.

GAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGCTGTTAGAGAGATAATTGGAATTA
ATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTGACGTAGAAAGTAATAATTTCTTGGGTAGTTTGC
AGTTTTAAATTATGTTTTAAATGGACTATCATATGCTTACCGTAACTTGAAAGTATTTGATTTCTTGGCT
TTATATATCTTGTGGAAGGACGAAACACCGNNNNNNNNNNNNNNNNNNNNGTTTAAGAGCTATGCTGGAAC
AGCATAGCAAGTTTAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCAGTCGGTGCTTTTTTTTT

SaCas9 sgRNA:

Yellow: U6 promoter Brown: spacer Green: scaffold

The red **G** is the transcription start site.

GAGGGCCTATTTCCCATGATTTCCTTCATATTTGCATATACGATACAAGGCTGTTAGAGAGATAATTGGAATTA
ATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTGACGTAGAAAGTAATAATTTCTTGGGTAGTTTG
AGTTTTAAAATTATGTTTTAAATGGACTATCATATGCTTACCCTAAGTAAAGTATTTGATTTCTTGGCT
TTATATATCTTGTGGAAGGACGAAACACCNNNNNNNNNNNNNNNNNNNNNNNGTTTTAGTACTCTGGAACA
GAATCTACTAAAACAAGGCCAAATGCCGTGTTTATCTCGTCAACTTGTTGGCGAGATTTTTTTT

PE pegRNA:

Yellow: U6 promoter Brown: spacer Green: scaffold Blue: 3' PBS

The red G is the transcription start site.

[illegible]

LbCas12a gRNA:

Yellow: U6 promoter Brown: spacer Green: scaffold

The red **G** is the transcription start site.

GAGGGCCTATTTCCCATGATTCTTCATATTTGCATATACGATACAAGGCTGTTAGAGAGATAATTGGAATTA
ATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTGACGTAGAAAGTAATAATTTCTTGGGTAGTTTGC
AGTTTTAAAATTATGTTTTAAATGGACTATCATATGCTTACCGTAACTTGAAAGTATTTTCGATTTCTTGGCT

TTATATATCTTGTGGAAAGGACGAAACACCGAATTTCTACTAAGTGTAGATNNNNNNNNNNNNNNNNNNNNNN
NTTTTTTT

Supplementary Reference

- 14 Kim, N. *et al.* Prediction of the sequence-specific cleavage activity of Cas9 variants. *Nat Biotechnol* **38**, 1328-1336 (2020). <https://doi.org/10.1038/s41587-020-0537-9>
- 24 Bravo, J. P. K. *et al.* Structural basis for mismatch surveillance by CRISPR-Cas9. *Nature* **603**, 343-347 (2022). <https://doi.org/10.1038/s41586-022-04470-1>
- 29 Lei, Z. *et al.* Detect-seq reveals out-of-protospacer editing and target-strand editing by cytosine base editors. *Nat Methods* **18**, 643-651 (2021). <https://doi.org/10.1038/s41592-021-01172-w>
- 31 Tsai, S. Q. *et al.* GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat Biotechnol* **33**, 187-197 (2015). <https://doi.org/10.1038/nbt.3117>
- 40 Pacesa, M. *et al.* R-loop formation and conformational activation mechanisms of Cas9. *Nature* **609**, 191-196 (2022). <https://doi.org/10.1038/s41586-022-05114-0>
- 41 Chen, J. S. *et al.* Enhanced proofreading governs CRISPR-Cas9 targeting accuracy. *Nature* **550**, 407-410 (2017). <https://doi.org/10.1038/nature24268>
- 77 Glasscock, C. J. *et al.* Computational design of sequence-specific DNA-binding proteins. *Nat Struct Mol Biol* **32**, 2252-2261 (2025). <https://doi.org/10.1038/s41594-025-01669-4>
- 78 Sugiki, S., Niide, T., Toya, Y. & Shimizu, H. Logistic Regression-Guided Identification of Cofactor Specificity-Contributing Residues in Enzyme with Sequence Datasets Partitioned by Catalytic Properties. *ACS Synth Biol* **11**, 3973-3985 (2022). <https://doi.org/10.1021/acssynbio.2c00315>
- 83 Nishimasu, H. *et al.* Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* **156**, 935-949 (2014). <https://doi.org/10.1016/j.cell.2014.02.001>
- 84 Kleinstiver, B. P. *et al.* High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* **529**, 490-495 (2016). <https://doi.org/10.1038/nature16526>
- 85 Vakulskas, C. A. *et al.* A high-fidelity Cas9 mutant delivered as a ribonucleoprotein complex enables efficient gene editing in human hematopoietic stem and progenitor cells. *Nat Med* **24**, 1216-1224 (2018). <https://doi.org/10.1038/s41591-018-0137-0>
- 86 Kim, D., Kim, D. E., Lee, G., Cho, S. I. & Kim, J. S. Genome-wide target specificity of CRISPR RNA-guided adenine base editors. *Nat Biotechnol* **37**, 430-435 (2019). <https://doi.org/10.1038/s41587-019-0050-1>
- 87 Schmid-Burgk, J. L. *et al.* Highly Parallel Profiling of Cas9 Variant Specificity. *Mol Cell* **78**, 794-800 e798 (2020). <https://doi.org/10.1016/j.molcel.2020.02.023>