Experimental design

1. Sample size
   Describe how sample size was determined.
   No statistical methods were used to predetermine sample size. Sample sizes were chosen after deep sequencing depending on the read number and quality. All sample sizes were sufficient for the following statistical test.

2. Data exclusions
   Describe any data exclusions.
   We excluded data according to the filtering conditions for the indel frequency analysis. The filtering condition is described in Supplementary Table 6. The criteria were not pre-established and were established after deep sequencing depending on the read number.

3. Replication
   Describe whether the experimental findings were reliably reproduced.
   Experimental replication was not attempted in this study

4. Randomization
   Describe how samples/organisms/participants were allocated into experimental groups.
   Data set HT 1 was split into data sets HT 1-1 and HT 1-2 by random sampling.

5. Blinding
   Describe whether the investigators were blinded to group allocation during data collection and/or analysis.
   The investigators were not blinded to group allocation. This study does not involve animals or human research participants.
   Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters
   For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).
   n/a □ Confirmed
   □ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
   □ A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
   □ A statement indicating how many times each experiment was replicated
   □ The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
   □ A description of any assumptions or corrections, such as an adjustment for multiple comparisons
   □ The test results (e.g. P values) given as exact values whenever possible and with confidence intervals noted
   □ A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
   □ Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.
Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.

Custom python scripts were used for indel frequency analysis, implementation of Seq-deepCpf1, and DeepCpf1. For obtaining DNase-seq narrow peak for each target site, bowtie (version 1.1.2) was used to align the reads; samtools (version 0.1.19) was used to convert the aligned files from sam format to bam format; bedtools (version 2.25.0) was used to retrieve peak information from the aligned files. Statistical significances were determined using SciPy (version 1.0.0) library.

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). Nature Methods guidance for providing algorithms and software for publication provides further information on this topic.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.

No restrictions on availability of the unique materials. Custom python scripts were deposited in a GitHub. Plasmids used for this study are available from the Addgene (#84750, #84752). Deep sequencing data from this study have been submitted to the NCBI Sequence Read Archive under accession number SRP107920. The data sets used in this study are provided as Supplementary Table 1. The data supporting the findings of this study are available within the paper and its supplementary information files including Supplementary Table 1.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

No antibodies were used.

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.

The source of cell line is American Type Culture Collection (ATCC).

b. Describe the method of cell line authentication used.

Not been authenticated.

c. Report whether the cell lines were tested for mycoplasma contamination.

Not been tested.

d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.

HEK293 cells were listed in the ICLAC. HEK293 cells were purchased from ATCC and we do not culture HeLa cells in our lab.

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.

No animals were used.

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

The study did not involve human research participants.