Engineering of CRISPR-Cas12b for human genome editing

Strecker et al.
Supplementary Figure 1. PAM discovery of Cas12b orthologs

(a) 

(b) 

1. Transform PAM library with E. coli with Cas12b or control vector

2. Identify depleted motifs to determine PAM

(c) 

AkCas12b  
BhCas12b  
EbCas12b  
LsCas12b

PAM depletion (-log2 ratio) vs. rank
Supplementary Figure 1. PAM discovery of Cas12b orthologs

a) Phylogenetic tree of the subtype V-B effector Cas12b proteins. Sequences are denoted by Genbank protein accession number and species name. The proteins that were experimentally studied in this work are highlighted in blue. b) Schematic of the PAM discovery assay in *E. coli*. c) Depleted PAMs were detected in only 4 out of 14 Cas12b systems in *E. coli*. A depletion threshold was set at a -log₂ ratio of 3.32 (dotted line) except for EbCas12b, which had a threshold set at 2.32. Depleted PAMs are shown as sequence motifs as well as PAM wheels starting in the middle of the wheel for the first 5′ base exhibiting sequence information.
Supplementary Figure 2. Cas12b RNA-Seq and in vitro reconstitution

(a) 

(b) 

(c) 

(d) 

(e) 

(f) 

| Cas12b | AkCas12b | BhCas12b | EbCas12b | LsCas12b |
|--------|----------|----------|----------|----------|
| sgRNA  | -        | -        | +        | -        |
| tracrRNA | -        | +        | -        | +        |
| crRNA  | +        | +        | -        | +        |

bp: 
- 250  
- 150  
- 100  
- 75   
- 50   
- 25   

90 min, 37°C
Supplementary Figure 2. Cas12b RNA-Seq and in vitro reconstitution

a-d) Alignment of small RNA-Seq reads for AkCas12b, BhCas12b, EbCas12b, and LsCas12b. The location of the tracrRNA used in cleavage reactions is highlighted in yellow. e) Coomassie stained SDS-PAGE gel of purified Cas12b proteins used in this study and commercially produced AsCas12a (IDT). f) In vitro cleavage reactions with AkCas12b and BhCas12b comparing tracrRNA and crRNA to v1 sgRNA scaffolds.
Supplementary Figure 3. Cas12b sgRNA optimization in mammalian cells

a

1. Transient transfection

2. Targeted deep sequencing

HEK293T cells

b

da

e

AkCas12b

BhCas12b

DNMT1 (5) VEGFA (7)

spacer

length 19 20 21 22 23 24 25 19 20 21 22 23 24 25

1. Transient transfection

2. Targeted deep sequencing

HEK293T cells

tracrRNA crRNA

U6

spacer

sgRNA

Cas12bCMV

NLS

3xHANLS

CMV

NLS

3xHA

sgRNA

tracrRNA crRNA

tracrRNA crRNA

tracrRNA crRNA

tracrRNA crRNA

tracrRNA crRNA

tracrRNA crRNA

tracrRNA crRNA
Supplementary Figure 3. Cas12b sgRNA optimization in mammalian cells

a) Schematic of expression constructs and assay for indel activity in mammalian cells.  
b, c) Sequence of AkCas12b and BhCas12b sgRNA variants. The location of deletions is denoted with a red line.  
d) Schematic of AkCas12b sgRNA design 1 with grey shading highlighting the location of changes.  
e) Indel activity in 293T cells with BhCas12b and varying spacer lengths. Error bars represent s.d. from n=2 replicates. Source data are provided as a Source Data file.
Supplementary Figure 4. Rational engineering of BhCas12b

(a) Indels (%) for Cas12b variants: control, WT, K641R, S893R, L915V, F905R/K846R, L915V/S893R, K846R/S893R, E837G/K641R, S893R/E837G, K641R/E837G, S893R/K846R/E837G, L915V/K846R/E837G, F905L/K788N/E837G/R749D.

(b) Indels (%) for RuvC active site mutants: control, WT, K641R, K749D, K846R, S893R, F905L, F905Y, L915V.

(c) Indels (%) for DNMT1 and VEGFA proteins: control, WT, K641R, K846R, S893R, F905L, F905Y, L915V.

(d) Indels (%) for DNMT1 and VEGFA proteins: control, WT, K641R, K846R, S893R, F905L, F905Y, L915V.

(f) RuvC active site and target strand structure.

(g) SDS-PAGE gel showing WT, v4 banding patterns.

(h) Gel electrophoresis of WT and v4 Cas12b over 15 minutes at 100 nM, 37°C.

(i) Cleavage efficiency of WT and v4 Cas12b over time in seconds.

(j) Nicking efficiency of WT and v4 Cas12b over time in seconds.
Supplementary Figure 4. Rational engineering of BhCas12b

a) Comparison of indel activity between BhCas12b and the highly similar BthCas12b in 293T cells. Error bars represent s.d. from n=2 replicates. b-e) Indel activity of BhCas12b mutant combinations at DNMT1 (target 5) and VEGFA (target 7). Error bars represent s.d. from a minimum of n=2 replicates. f) BhCas12b v4 mutations modeled into the structure of BthCas12b (PDB:5wti [10.2210/pdb5WTI/pdb]) using Pymol (Schrodinger). g) Coomassie stained SDS-PAGE gel of purified BhCas12b WT and v4 protein. h) In vitro cleavage time-course with BhCas12b WT and v4 variant. Gel is representative image from n=3 experiments. i,j) Quantitation of dsDNA cleavage products (i) and upper nicked product (j) from the reactions shown in panel h. Error bars represent s.d. from n=3 experiments. Source data are provided as a Source Data file.
Supplementary Figure 5. BhCas12b v4 mediates genome editing in human cell lines

(a) Indels (%)

(b) Distance to nearest cut site (bp)

Avg bp/cut site: 32 35 7

(c) VEGFA locus (10)

(d) Indels (%)

(e) HDR (%)
Supplementary Figure 5. BhCas12b v4 mediates genome editing in human cells lines

a) BhCas12b v4 indel activity in 293T cells at 56 targets. Each dot represents a single target site, averaged from n=4 replicates. b) Analysis of PAM prevalence for Class 2 CRISPR-Cas nucleases. Probability mass function for the distance from each base within non-masked human coding sequences to the nearest Cas9 or Cas12 cleavage site. c) Schematic of a VEGFA (target 10) site targetable by SpCas9 and Cas12b nucleases and a 120-nt ssODN donor containing a TC to CA mutation and PAM disrupting mutations d) Indel activity of each nuclease at the locus. Error bars represent s.d. from n=3 replicates. e) Frequency of homology-directed repair (HDR) using a target strand (T) or non-target strand (NT) donor. Grey bars indicate the frequency of TC to CA mutation, while blue bars indicate perfect edits with no detectable mutations in the 36-nt sequence shown in panel c. Error bars represent s.d. from n=3 replicates. Source data are provided as a Source Data file.
Supplementary Figure 6. Specificity analysis of matched CRISPR-Cas nuclease targets

**EMX1 (14)**
- SpCas9
- BhCas12b v4

**CXCR4 (18)**
- SpCas9
- BhCas12b v4

**EMX1 (15)**
- SpCas9
- BhCas12b v4

**CXCR4 (19)**
- SpCas9
- BhCas12b v4

**DNMT1 (16)**
- SpCas9
- BhCas12b v4

**CXCR4 (20)**
- SpCas9
- BhCas12b v4

**CXCR4 (17)**
- SpCas9
- BhCas12b v4

**CXCR4 (21)**
- SpCas9
- BhCas12b v4

**HPRT1 (22)**
- SpCas9
- BhCas12b v4

**AsCas12a**
Supplementary Figure 6. Specificity analysis of matched CRISPR-Cas nuclease targets

Full Guide-Seq analysis of detected off-targets in Fig. 4b. A list of detected cleavage sites (up to 20 per target) is presented for each nuclease with the on-target site denoted with a small box. Mismatches to the guide sequence are highlighted.
### Supplementary Figure 7. Specificity analysis of unmatched CRISPR-Cas nuclease targets

#### SpCas9 unmatched: DNMT1 (23)

| Read | 6121 |
|------|------|

#### SpCas9 unmatched: EMX1 (24)

| Read | 23244 |
|------|------|

#### SpCas9 unmatched: VEGFA (25)

| Read | 4962 |
|------|------|

#### SpCas9 unmatched: VEGFA (26)

| Read | 13960 |
|------|------|

#### SpCas9 unmatched: VEGFA (27)

| Read | 4642 |
|------|------|

#### SpCas9 unmatched: VEGFA (28)

| Read | 5371 |
|------|------|

#### SpCas9 unmatched: VEGFA (29)

| Read | 5619 |
|------|------|

#### SpCas9 unmatched: TUBB (31)

| Read | 27867 |
|------|------|

#### SpCas9 unmatched: GRIN2B (28)

| Read | 2348 |
|------|------|

#### BhCas12b v4 unmatched: DNMT1 (37)

| Read | 1078 |
|------|------|

#### BhCas12b v4 unmatched: DNMT1 (38)

| Read | 5446 |
|------|------|

#### BhCas12b v4 unmatched: VEGFA (39)

| Read | 3514 |
|------|------|

#### BhCas12b v4 unmatched: EMX1 (40)

| Read | 2159 |
|------|------|

#### BhCas12b v4 unmatched: EMX1 (41)

| Read | 1675 |
|------|------|

#### BhCas12b v4 unmatched: DNMT1 (44)

| Read | 1327 |
|------|------|

#### BhCas12b v4 unmatched: GRIN2B (42)

| Read | 1636 |
|------|------|

#### BhCas12b v4 unmatched: EMX1 (43)

| Read | 1804 |
|------|------|

#### BhCas12b v4 unmatched: EMX1 (44)

| Read | 396 |
|------|------|

#### BhCas12b v4 unmatched: EMX1 (45)

| Read | 726 |
|------|------|

#### BhCas12b v4 unmatched: EMX1 (46)

| Read | 1164 |
|------|------|

#### BhCas12b v4 unmatched: EMX1 (47)

| Read | 1959 |
|------|------|

#### BhCas12b v4 unmatched: VEGFA (48)

| Read | 5962 |
|------|------|

#### BhCas12b v4 unmatched: EMX1 (49)

| Read | 1495 |
|------|------|
Supplementary Figure 7. Specificity analysis of unmatched CRISPR-Cas nuclease targets

Full Guide-Seq analysis of detected off-targets for unmatched targets. A list of detected cleavage sites (up to 20 per target) is presented for each nuclease with the on-target site denoted with a small box. Mismatches to the guide sequence are highlighted.
Supplementary Figure 8. Characterization of BvCas12b

(a) PAM depletion (log2 ratio) for BvCas12b.

(b) Alignment of crRNA, tracrRNA, and sgRNA designs.

(c) Heat map showing indels for different PAM sequences.

(d) Western blot analysis of BvCas12b.

(e) SDS-PAGE gels for BvCas12b.

(f) Sequence alignment for sgRNAs 1-6.

(g) Diagram of sgRNA and tracrRNA interactions.

(h) Bar graph showing indels for different sgRNA designs.

(i) Scatter plot showing indels vs. PAM depletion.

(j) Scatter plot showing indels for BhCas12b v4.
Supplementary Figure 8. Characterization of BvCas12b

a) PAM discovery as described in Supplementary Fig. 1b,c. b) Alignment of small RNA-Seq reads for BvCas12b. The location of the tracrRNA used in cleavage reactions is highlighted in yellow. c,d) In vitro reactions with 250 nM purified BvCas12 protein and synthesized RNA were carried out at the indicated temperatures for 90 min. e) Coomassie stained SDS-PAGE gel of purified BvCas12b. f) BvCas12b sgRNA variants. The location of deletions is denoted with a red line. g) Schematic of BvCas12b sgRNA design 1 with grey shading highlighting the location of changes to the guide design in variants 2 - 6. h) BvCas12b indel activity in 293T cells with sgRNA variants. Error bars represent s.d. from n=4 replicates. i) BvCas12b indel activity in 293T cells at 57 targets. Each dot represents a single target site, averaged from n=4 replicates. j) Correlation of BhCas12b v4 and BvCas12b activity at matched target sites. Source data are provided as a Source Data file.
Supplementary Figure 9. BvCas12b mismatch tolerance and specificity

(a) Graph showing the activity of BvCas12b with mismatch positions and ratios of DNMT1 and VEGFA.

(b) Examples of BvCas12b matched and unmatched sequences for EMX1, DNMT1, CXCR4, and GRIN2B.
Supplementary Figure 9. BvCas12b mismatch tolerance and specificity

a) BvCas12b indel activity in 293T cells when mismatches are present between the guide sgRNA and target DNA. Mismatches were inserted in the sgRNA to match the target strand (i.e., C to G, A to T). Error bars represent s.d. from n=4 replicates. b) Guide-Seq analysis of BvCas12b at 24 targets. A list of detected cleavage sites is presented with the on-target site denoted with a small box. Mismatches to the guide sequence are highlighted. Source data are provided as a Source Data file.
### Supplementary Table 1: Guide sequences used in this study

| Gene   | Target | 5' PAM | Sequence                      |
|--------|--------|--------|-------------------------------|
| DNMT1  | 1      | GTTCT  | AGACCCAGAGGCTCAAGTGAGCA       |
| DNMT1  | 2      | ATTTT  | AGCTGAAGGAAATAAAAAGAAAA       |
| VEGFA  | 3      | ATTTCT | TCTCCCTGGGAAGCATCCCTG        |
| EMX1   | 4      | ATTTTT | TCATGGGAAATAATTCAGAATC       |
| DNMT1  | 5      | TTTTC  | CCTCAGCTCTGCTCGGTGAATT       |
| DNMT1  | 6      | TTTTG  | AGAGGTGTCAGTCTCCGTGAAC       |
| VEGFA  | 7      | TTTG   | GGAGGTAGAAATAGGGGTCCA        |
| VEGFA  | 8      | TTTT   | CAAAGCCCATTCCCTTTAGACC       |
| DNMT1  | 9      | ATTT   | CCTTCAGCTAAATAAAGGAGG        |
| VEGFA  | 10     | ATTC   | TTCTCCCTGGGAAGCATCCCTG       |
| GRIN2B | 11     | ATTC   | TGCAGAGCAAATACCAGAGATAA      |
| PDCD1  | 12     | ATTG   | CGCCGGCCCTGACCACTCCCCAT      |
| CXCR4  | 13     | ATTC   | CCGACTTCATCTTTGCAACGTC       |
| EMX1   | 14     | ATTT   | TAGAGCAGCTGGAAGCATGGGATG     |
| EMX1   | 15     | ATTC   | TTGCTCCAGAGGCCCTTCTGGGATG   |
| DNMT1  | 16     | ATTC   | CTGGTGCCAGAAACAGGGGTGAC      |
| CXCR4  | 17     | ATTC   | TGGGCTTACAGCACTTTAGTG        |
| CXCR4  | 18     | ATTT   | TGAATTGTGTCTACCAAGAAG        |
| CXCR4  | 19     | ATTT   | AGAGGCAGAGGGGCGGCTGGAAC     |
| GRIN2B | 20     | TTTT   | CTTCAGCCCCAAGAACAGTACAGA    |
| CXCR4  | 21     | TTTT   | TCTGGTGCAAGGAAACAGCACGAC     |
| HPRT1  | 22     | TTTT   | CTTCAGCTCTGCTCGGTGAATT      |
| DNMT1  | 23     | -      | TACCTCCTGCTCGGTGAATT        |
| EMX1   | 24     | -      | GCTACAGGGCAGAGCACCAAGG       |
| VEGFA  | 25     | -      | AGGTCAGAAATAGGGGGGTCC       |
| VEGFA  | 26     | -      | CAGGCTGTGAAACCTTTGGTG      |
| VEGFA  | 27     | -      | GACCCCTCCACCCCCGCTC         |
| GRIN2B | 28     | -      | GTATCTAGGGTCCTTCTAAGAC      |
| VEGFA  | 29     | -      | TCTCCCTGGGAAGCATCCCTC       |
| EMX1   | 30     | -      | GAGTCCAGGCAAGAAGAGAA        |
| TUBB   | 31     | -      | TTTTGGGAGTAGAAGAAAGGT       |
| VEGFA  | 32     | -      | AGTTGTCAGGGGATGTCTCC        |
| DNMT1  | 33     | TTTT   | CCTCAGCTCGGCTCGGTGAATTT     |
| VEGFA  | 34     | TTTT   | GGAGGTGCAAAATAGGGGGGTCCA    |
| EMX1   | 35     | TTTT   | GATGGGCACTTCAGGCAACAGAT     |
| EMX1   | 36     | TTTT   | GGAAGTGTCAGGGGATGCTCC       |
| DNMT1  | 37     | ATTT   | CCCTCAGCTAAATAAAGGAGG       |
| Gene     | Position | Motif | Sequence                        |
|----------|----------|-------|---------------------------------|
| DNMT1    | 38       | ATTT  | GGCTCAGCAGGCACCTGCCTCAG         |
| VEGFA    | 39       | ATTT  | GGGACTGGAGTTGTCTTCATGTAC        |
| EMX1     | 40       | ATTT  | TCTCCATGAAAAATACTGGGGTC         |
| EMX1     | 41       | ATTT  | TTCATGGAGAAAAATATTCAGAAT        |
| GRIN2B   | 42       | ATTG  | GCAGCTACAGGCAGGACAAAGG          |
| EMX1     | 43       | ATTT  | CCTGGAAACCATCCAGGCCTTG          |
| DNMT1    | 44       | ATTG  | GGTCACTGTTAACATCAGTACG          |
| CXCR4    | 45       | ATTT  | TCTTCACGGAACAGGGGTTCCTT         |
| EMX1     | 46       | TTTG  | TGGTTGCCCACCCCTAGTCATTGG        |
| EMX1     | 47       | TTTG  | GATGGCGACTTCAGGCACAGGAT         |
| DNMT1    | 49       | TTTC  | CCTCACTTCTGCTCGTGAAATT          |
| VEGFA    | 50       | TTTG  | GGAAGTGTCAGGGATGCTCCC           |
| DNMT1    | 51       | ATTT  | GGCTCAGCAGGCACCTGCCTCAG         |
| EMX1     | 52       | ATTT  | TTCATGGAGAAAAATATTCAGAAT        |
| VEGFA    | 53       | ATTT  | CTGACCTCCCCAAACAGCTACATA        |