Electronic Supplementary Material (ESI) for RSC Advances.

Supplementary Information

Target DNA mutagenesis-based fluorescence assessment of off-target activity of CRISPR-Cas9 system

Dan Wang [a], Cuili Niu [a], Jingxin Han [a], Dejun Ma [a], Zhen Xi [a,b] *

a. Department of Chemical Biology, State Key Laboratory of Elemento-Organic Chemistry, National Engineering Research Center of Pesticide (Tianjin), College of Chemistry, Nankai University, Tianjin 300071, China.

b. Collaborative Innovation Center of Chemical Science and Engineering (Tianjin), Tianjin 300071, China.

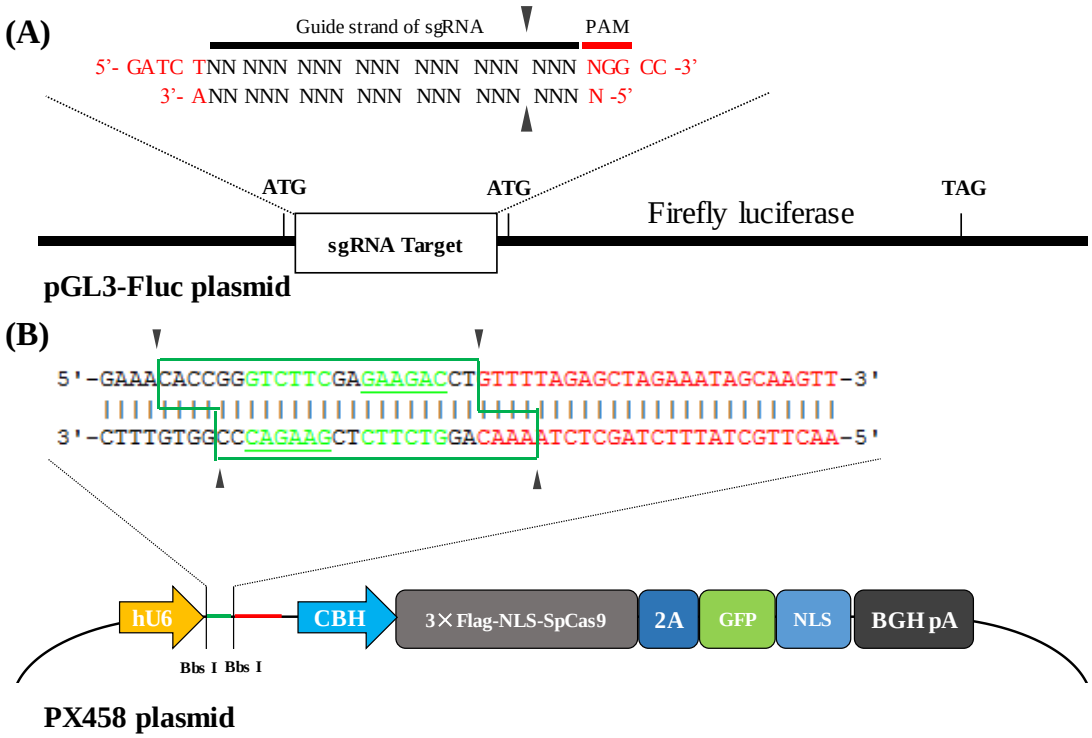

Figure S1. The illustration of an engineered dual-luciferase reporter toolkit for determine the off-target cleavage. (A) pGL3-Fluc plasmids were modified to insert a target region followed by the start codon ATG without interfering with the firefly luciferases activity. (B) PX458 plasmids could be digested with *Bbs* I enzymes and ligated with guide sequence cassettes to express sgRNAs for gene cleavage.

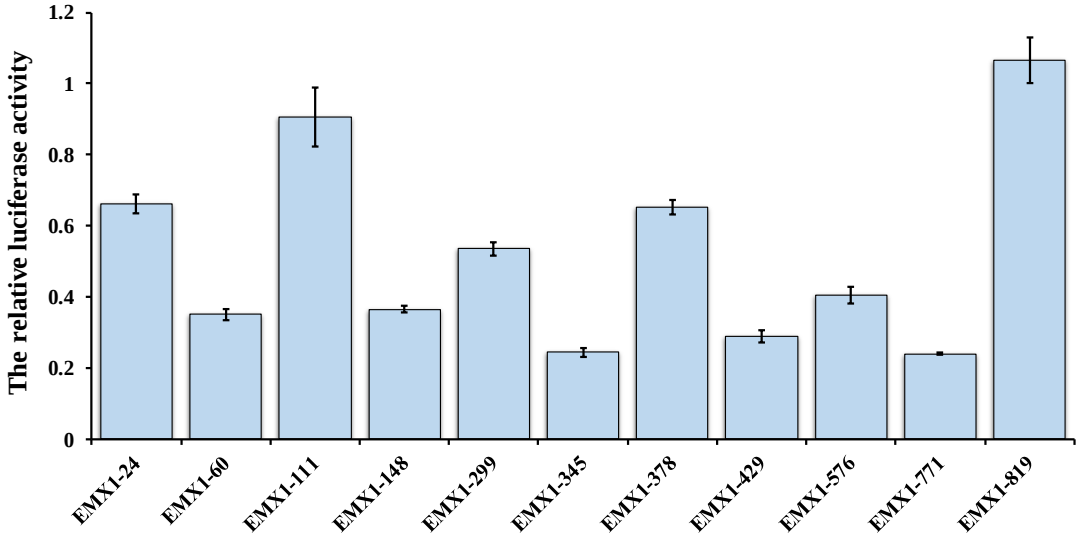

Figure S2. Evaluation of different sgRNAs targeting different sites of EMX1 gene. Cas9 and sgRNA co-expressing plasmids (400 ng) were co-transfected with the firefly luciferase-expressing plasmids pGL3-Fluc and the internal control plasmid pRL-TK expressing the Renilla luciferase. The cells were cultured for 48 h at 37℃ until the relative luciferase unit (RLU) was determined.

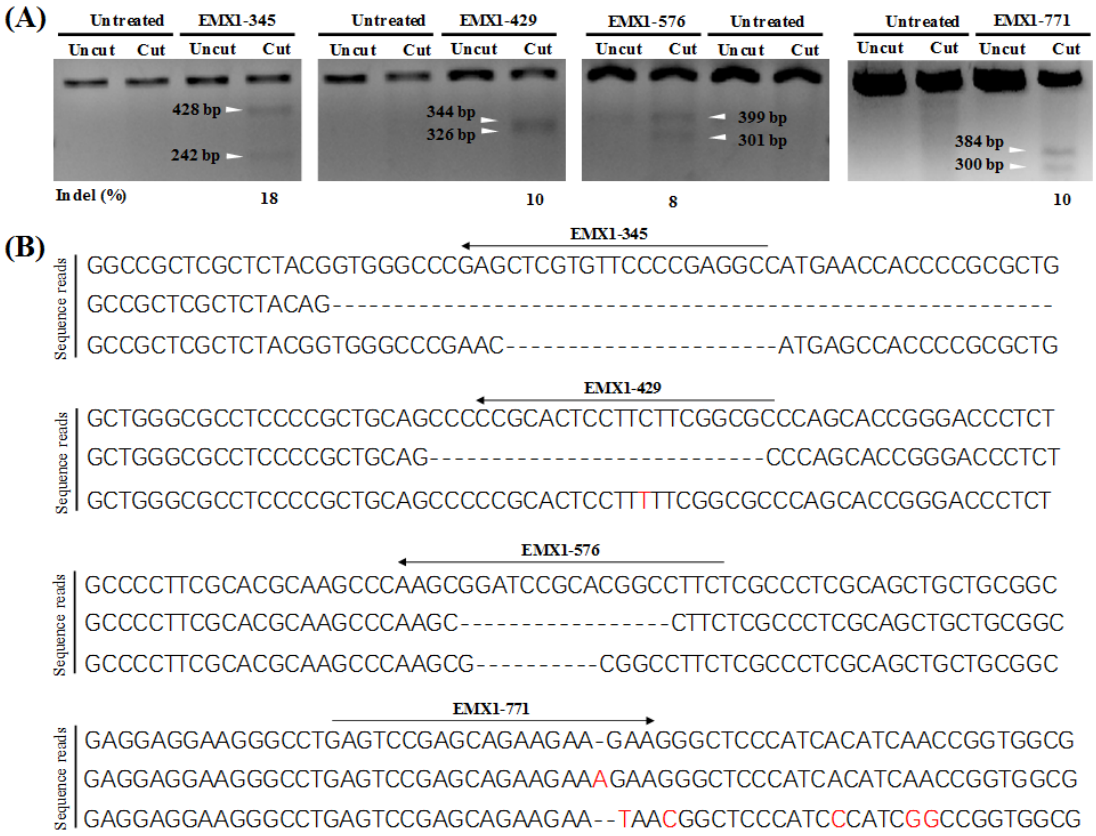

Figure S3. EMX1 gene modification was verified by T7EI assay (A) and sequencing (B). 400 ng PX458 plasmids expressing highly-efficient sgRNAs targeting EMX1 gene were transfected into HEK293T cells. After genomic DNA purification, gene-specific primers flanking the cleavage site for each coding region were used for polymerase chain reaction (PCR). The PCR products were subsequently annealed and digested with T7 endonuclease I (T7EI). The digested products were analyzed with 2% agarose. The remaining PCR products were ligated into TA vector for sequencing.

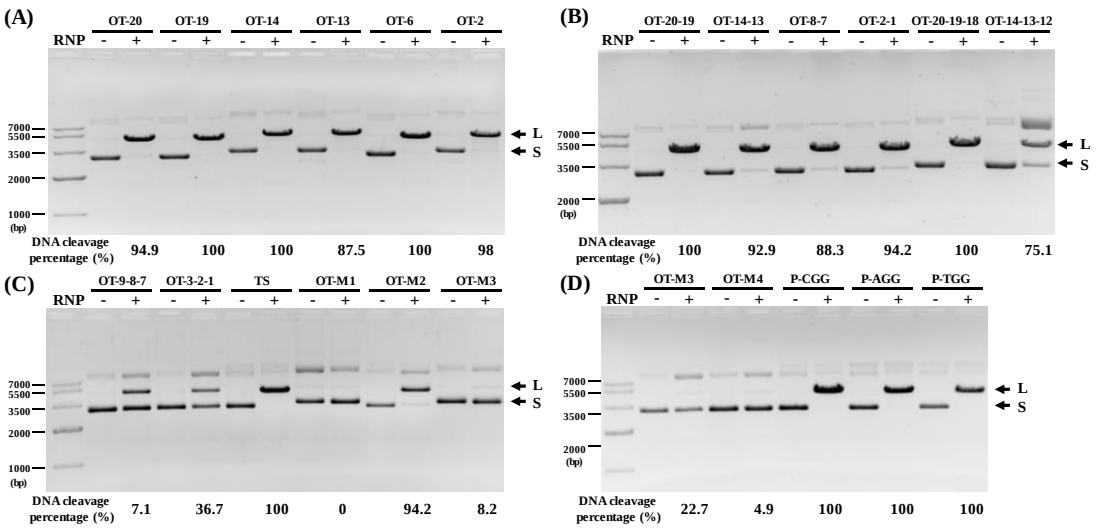

Figure S4. In vitro plasmid DNA cleavage assay of off-target sites containing mono-nucleotide mismatches (A), di-nucleotide mismatches (B), tri-nucleotide mismatches (B, C), multi-nucleotide mismatches (C, D) and PAM mutations (D). pGL3-FLuc plasmids containing the mono-nucleotide mismatch with guide RNA sequence was chosen as the target DNA. Reactions were performed at a molar ratio of 10:10:1 (Cas9/sgRNA/target DNA) in 20 μL reactions. After the pre-incubation of Cas9 and sgRNA at 37℃ for 10 min to form the active ribonucleoprotein (RNP), pGL3-FLuc plasmid (600 ng) was then added and incubated for 60 min. The reaction products were analyzed on 0.7% agarose gel. S: The supercoiled state of plasmid DNA. L: the linearized plasmid DNA by RNA-guided Cas9 nuclease.

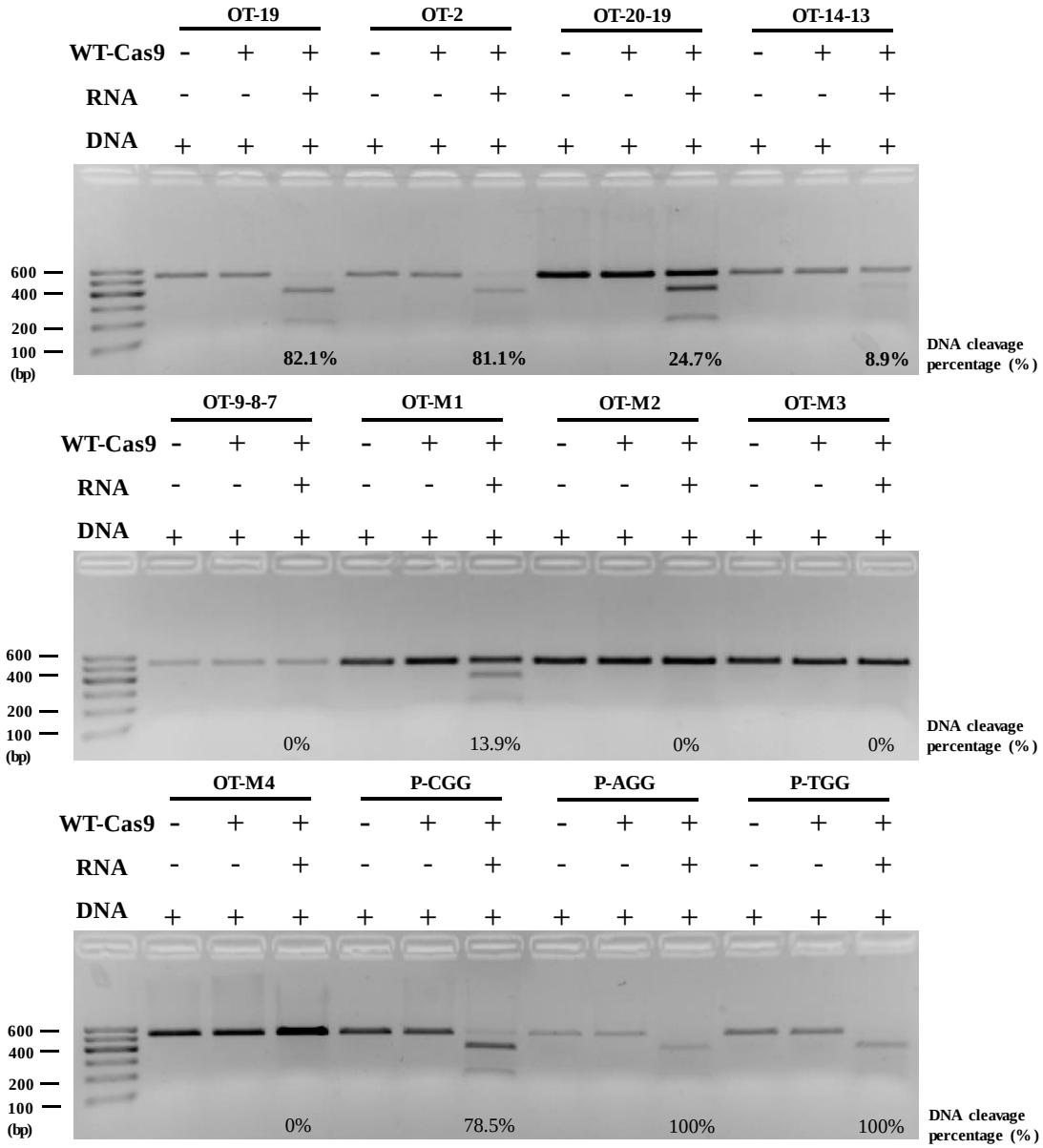

Figure S5. In vitro linear DNA cleavage assay of off-target sites by WT-Cas9 nuclease. Linear DNA template amplified from pGL3-FLuc plasmids containing the different nucleotide mismatches with guide RNA sequence was chosen as the target DNA. Reactions were performed at a molar ratio of 10:10:1 (Cas9/sgRNA/target DNA) in 20 μL reactions. After the pre-incubation of Cas9 and sgRNA at 37℃ for 10 min, linear DNA template (100 ng) was then added and incubated for 12 h. The reaction products were analyzed on 2% agarose gel.

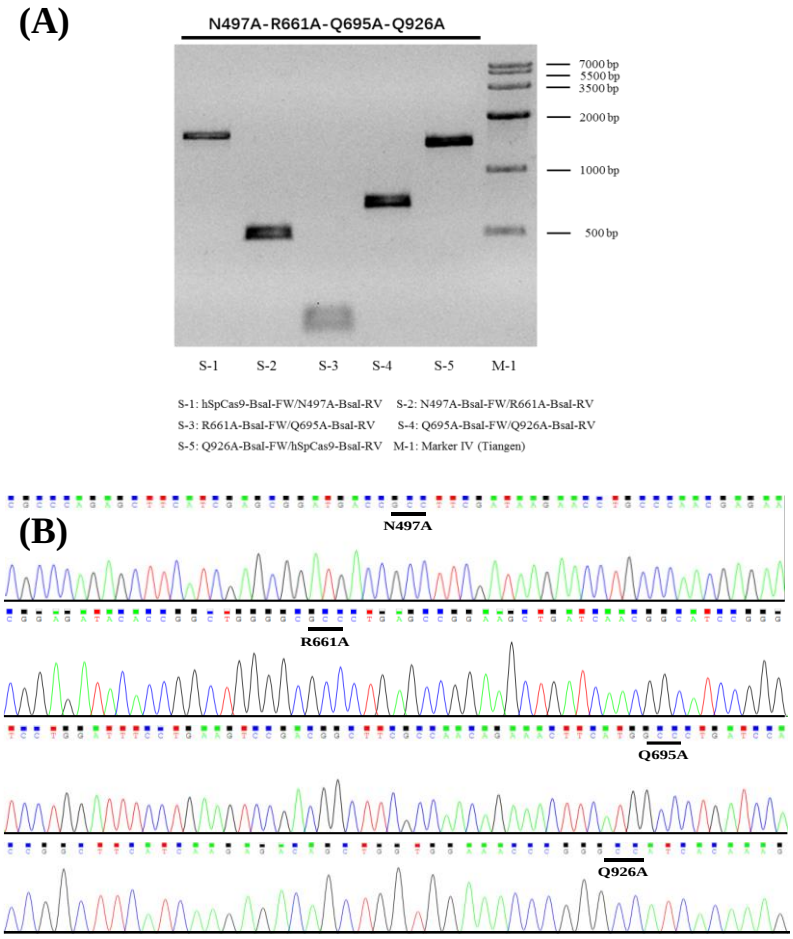

Figure S6. The construction of PX458-HF-Cas9 plasmid (N497A/R661A/Q695A/Q926A). (A) The PCR product in 1% agarose. (B) The sequencing validation of PX458-HF-Cas9 plasmid.

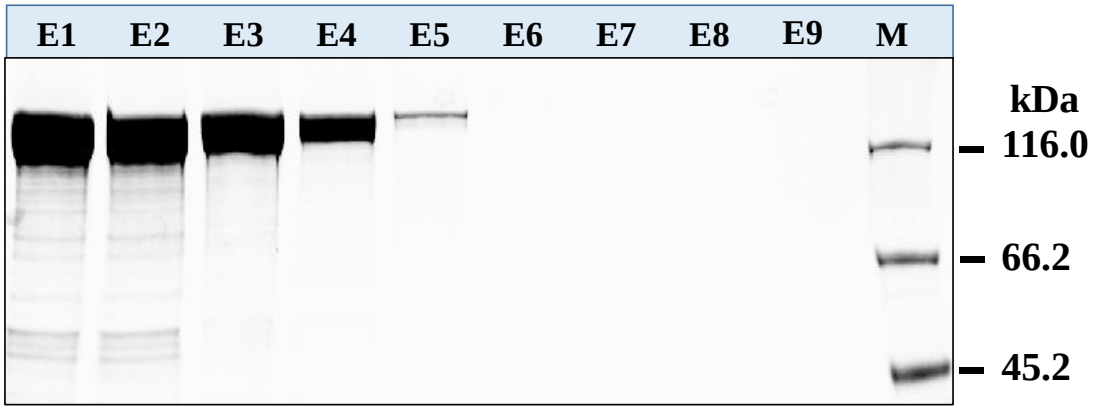

Figure S7. The purification of HF-Cas9 proteins. After 9 times of elution (E1, E2, … , E8, E9), each elution was validated by 12% SDS-PAGE.

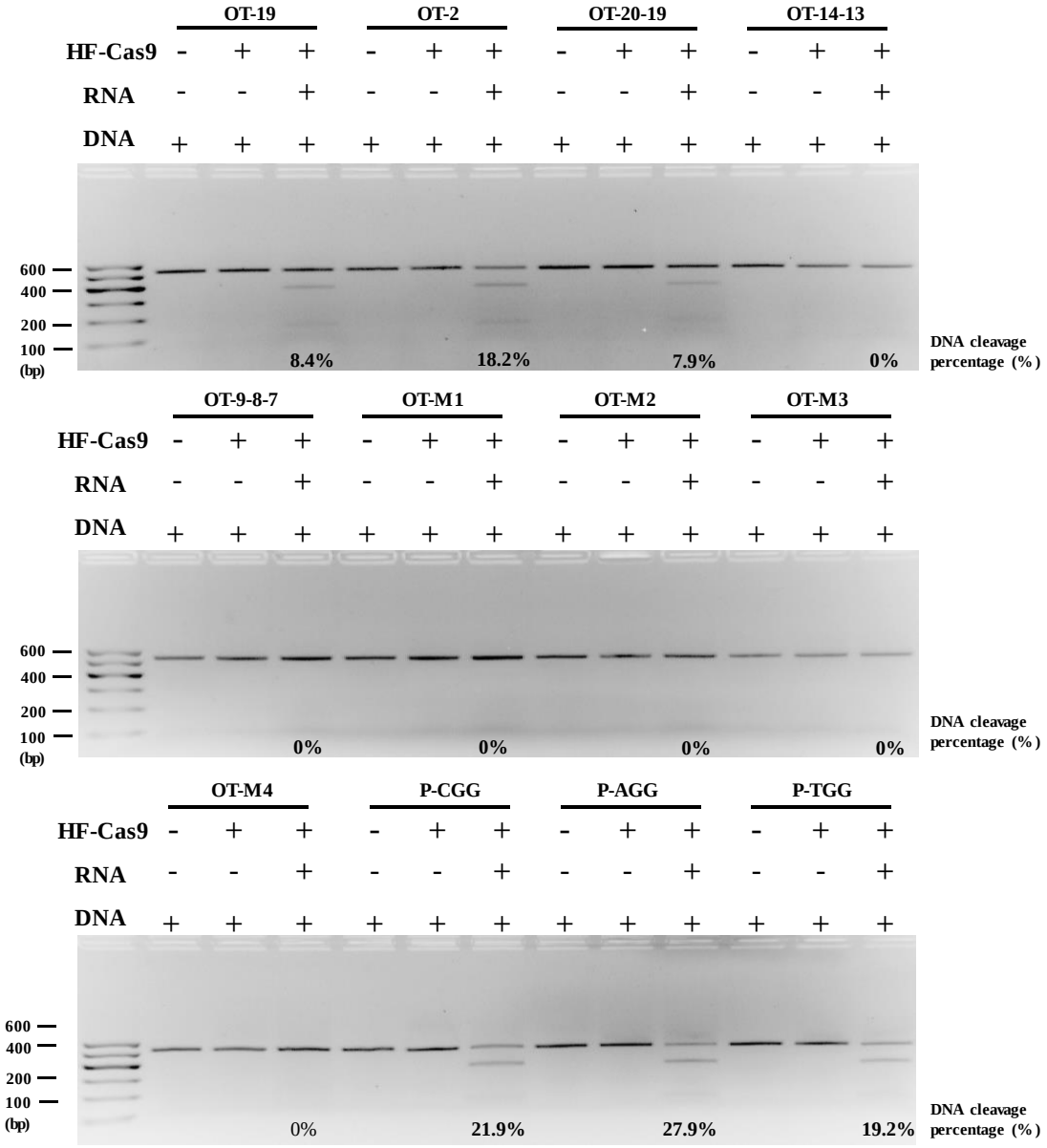

Figure S8. In vitro linear DNA cleavage assay of off-target sites by HF-Cas9 nuclease. Linear DNA template amplified from pGL3-FLuc plasmids containing the different nucleotide mismatches with guide RNA sequence was chosen as the target DNA. Reactions were performed at a molar ratio of 10:10:1 (Cas9/sgRNA/target DNA) in 20 μL reactions. After the pre-incubation of Cas9 and sgRNA at 37℃ for 10 min, linear DNA template (100 ng) was then added and incubated for 12 h. The reaction products were analyzed on 2% agarose gel.

Table S1 The comparison of CRISTA prediction scores and tested off-target scores for target DNA with various mismatches

Code	Mismatches	CRISTA Score [1]	Off-target Score [2]	CRISPR-off Score [3]
P-TGG	0	0.9247	0.8706	0.8051
P-AGG	0	0.9374	0.8044	0.7540
OT-20	1	0.8081	0.9204	0.7438
P-GGG	0	0.9145	0.8092	0.7400
OT-19	1	0.8103	0.9028	0.7316
OT-13	1	0.7740	0.8687	0.6724
P-CGG	0	0.9096	0.6797	0.6183
OT-2	1	0.8098	0.7319	0.5927
OT-14	1	0.7751	0.7642	0.5923
OT-20-19	2	0.7613	0.7652	0.5826
OT-20-19-18	3	0.5437	0.8800	0.4785
OT-8-7	2	0.7046	0.4314	0.3040
OT-6	1	0.7825	0.3720	0.2911
OT-14-13	2	0.7249	0.3527	0.2557
OT-15-14-13	3	0.5727	0.4457	0.2553
OT-M2	4	0.3972	0.4415	0.1754
OT-M3	4	0.3414	0.2599	0.0888
OT-3-2-1	3	0.6492	0.1036	0.0672
OT-9-8-7	3	0.4784	0.1121	0.0536

OT-M1	5	0.0515	0.9311	0.0480
OT-M4	6	0.1358	0.3006	0.0408
OT-2-1	2	0.8025	0.0098	0.0078

1. CRISTA Score was acquired from the CRISTA database (http://crista.tau.ac.il/findofftargets.html).
2. Off-target Score was calculated as the gene cleavage efficiency from the formulation (Off-target Score =1-the relative luciferase activity).
3. CRISPR-off Score was calculated according to the formulation (CRISPR-off Score= CRISTA Score×Off-target Score). The score scope ranges from 0 to 1, indicating that the lowest represents the inefficient cleavage on the DNA while the highest represents the highly-efficient cleavage.

Table S2. sgRNA oligos used in the construction of CRISPR-expressing plasmids

Code	Primer Name	Sequence (5'-3')
1	PX458-EMX1-24-FW	CACCGCCTGGCGCCGCCGCCTTGCG
	PX458-EMX1-24-RV	AAACCGCAAGGCGGCGGCGCCAGGC
2	PX458-EMX1-60-FW	CACCGTGCGAGGCAGCCGGGCTCT
	PX458-EMX1-60-RV	AAACAGAGCCCGGCTGCCTCGCAC
3	PX458-EMX1-111-FW	CACCGTAAAGCCGCGCTTGGCCGC
	PX458-EMX1-111-RV	AAACGCGGCCAAGCGCGGCTTTAC
4	PX458-EMX1-148-FW	CACCGCCAAGGACGGCGGCACCGGC
	PX458-EMX1-148-RV	AAACGCCGGTGCCGCCGTCCTTGGC
5	PX458-EMX1-299-FW	CACCGCGCCCGCGGCGGCCGCGGCA
	PX458-EMX1-299-RV	AAACTGCCGCGGCCGCCGCGGGCGC
6	PX458-EMX1-345-FW	CACCGCCTCGGGGAACACGAGCTC
	PX458-EMX1-345-RV	AAACGAGCTCGTGTTCCCCGAGGC
7	PX458-EMX1-378-FW	CACCGCCGGATGCACGGTCAGCGC
	PX458-EMX1-378-RV	AAACGCGCTGACCGTGCATCCGGC
8	PX458-EMX1-429-FW	CACCGCGCCGAAGAAGGAGTGCGG
	PX458-EMX1-429-RV	AAACCCGCACTCCTTCTTCGGCGC
9	PX458-EMX1-576-FW	CACCGAAGGCCGTGCGGATCCGCTT
	PX458-EMX1-576-RV	AAACAAGCGGATCCGCACGGCCTTC
10	PX458-EMX1-771-FW	CACCGAGTCCGAGCAGAAGAAGAA
	PX458-EMX1-771-RV	AAACTTCTTCTTCTGCTCGGACTC
11	PX458-EMX1-819-FW	CACCGCGCCACCGGTTGATGTGAT
	PX458-EMX1-819-RV	AAACATCACATCAACCGGTGGCGC

Table S3. Off-target DNA oligonucleotides used in the construction of pGL3-Fluc plasmids

Primer Name	Sequence (5'-3')
OT-20-F	GATCTCCCTCGGGGAACACGAGCTCGGGCC
OT-19-F	GATCTGGCTCGGGGAACACGAGCTCGGGCC
OT-14-F	GATCTGCCTCGCGGAACACGAGCTCGGGCC
OT-13-F	GATCTGCCTCGGCGAACACGAGCTCGGGCC
OT-6-F	GATCTGCCTCGGGGAACACCAGCTCGGGCC
OT-2-F	GATCTGCCTCGGGGAACACGAGCACGGGCC
OT-20-19-F	GATCTCGCTCGGGGAACACGAGCTCGGGCC
OT-14-13-F	GATCTGCCTCGCCGAACACGAGCTCGGGCC
OT-8-7-F	GATCTGCCTCGGGGAACTGGAGCTCGGGCC
OT-2-1-F	GATCTGCCTCGGGGAACACGAGCAGGGGCC
OT-20-19-18-F	GATCTCGGTCGGGGAACACGAGCTCGGGCC
OT-14-13-12-F	GATCTGCCTCCCCGAACACGAGCTCGGGCC
OT-9-8-7-F	GATCTGCCTCGGGGAAGTGGAGCTCGGGCC
OT-3-2-1-F	GATCTGCCTCGGGGAACACGAGGAGGGGCC
TS-F	GATCTGCCTCGGGGAACACGAGCTCGGGCC
OT-M1-F	GATCTGGCTCGCCGAACACCAGCACGGGCC
OT-M2-F	GATCTCGCTCGGGGAACACGAGCAGGGGCC
OT-M3-F	GATCTGCCTCGCCGAACTGGAGCTCGGGCC
OT-M4-F	GATCTGCCTCCCCGAAGTGGAGCTCGGGCC
P-CGG-F	GATCTGCCTCGGGGAACACGAGCTCCGGCC
P-AGG-F	GATCTGCCTCGGGGAACACGAGCTCAGGCC
P-TGG-F	GATCTGCCTCGGGGAACACGAGCTCTGGCC
OT-20-R	CGAGCTCGTGTTCCCCGAGGGA
OT-19-R	CGAGCTCGTGTTCCCCGAGCCA
OT-14-R	CGAGCTCGTGTTCCGCGAGGCA
OT-13-R	CGAGCTCGTGTTCGCCGAGGCA
OT-6-R	CGAGCTGGTGTTCCCCGAGGCA
OT-2-R	CGTGCTCGTGTTCCCCGAGGCA
OT-20-19-R	CGAGCTCGTGTTCCCCGAGCGA
OT-14-13-R	CGAGCTCGTGTTCGGCGAGGCA
OT-8-7-R	CGAGCTCCAGTTCCCCGAGGCA
OT-2-1-R	CCTGCTCGTGTTCCCCGAGGCA
OT-20-19-18-R	CGAGCTCGTGTTCCCCGACCGA
OT-14-13-12-R	CGAGCTCGTGTTCGGGGAGGCA
OT-9-8-7-R	CGAGCTCCACTTCCCCGAGGCA
OT-3-2-1-R	CCTCCTCGTGTTCCCCGAGGCA
TS-R	CGAGCTCGTGTTCCCCGAGGCA
OT-M1-R	CGTGCTGGTGTTCGGCGAGCCA
OT-M2-R	CCTGCTCGTGTTCCCCGAGCGA
OT-M3-R	CGAGCTCCAGTTCGGCGAGGCA
OT-M4-R	CGAGCTCCACTTCGGGGAGGCA

P-CGG-R	GGAGCTCGTGTTCCCCGAGGCA
P-AGG-R	TGAGCTCGTGTTCCCCGAGGCA
P-TGG-R	AGAGCTCGTGTTCCCCGAGGCA

Table S4. On-target DNA oligonucleotides used in the construction of pGL3-Fluc plasmids

Code	Primer Name	Sequence (5'-3')
1	pGL3-EMX1-24-FW	GATCTCCTGGCGCCGCCGCCTTGCGGGGCC
	pGL3-EMX1-24-RV	CCGCAAGGCGGCGGCGCCAGGA
2	pGL3-EMX1-60-FW	GATCTGTGCGAGGCAGCCGGGCTCTGGGCC
	pGL3-EMX1-60-RV	CAGAGCCCGGCTGCCTCGCACA
3	pGL3-EMX1-111-FW	GATCTGTAAAGCCGCGCTTGGCCGCGGGCC
	pGL3-EMX1-111-RV	CGCGGCCAAGCGCGGCTTTACA
4	pGL3-EMX1-148-FW	GATCTCCAAGGACGGCGGCACCGGCGGGCC
	pGL3-EMX1-148-RV	CGCCGGTGCCGCCGTCCTTGGA
5	pGL3-EMX1-299-FW	GATCTCGCCCGCGGCGGCCGCGGCAGGGCC
	pGL3-EMX1-299-RV	CTGCCGCGGCCGCCGCGGGCGA
6	pGL3-EMX1-345-FW	GATCTGCCTCGGGGAACACGAGCTCGGGCC
	pGL3-EMX1-345-RV	CGAGCTCGTGTTCCCCGAGGCA
7	pGL3-EMX1-378-FW	GATCTGCCGGATGCACGGTCAGCGCGGGCC
	pGL3-EMX1-378-RV	CGCGCTGACCGTGCATCCGGCA
8	pGL3-EMX1-429-FW	GATCTGCGCCGAAGAAGGAGTGCGGGGGCC
	pGL3-EMX1-429-RV	CCCGCACTCCTTCTTCGGCGCA
9	pGL3-EMX1-576-FW	GATCTAAGGCCGTGCGGATCCGCTTGGGCC
	pGL3-EMX1-576-RV	CAAGCGGATCCGCACGGCCTTA
10	pGL3-EMX1-771-FW	GATCTGAGTCCGAGCAGAAGAAGAAGGGCC
	pGL3-EMX1-771-RV	CTTCTTCTTCTGCTCGGACTCA
11	pGL3-EMX1-819-FW	GATCTGCGCCACCGGTTGATGTGATGGGCC
	pGL3-EMX1-819-RV	CATCACATCAACCGGTGGCGCA

Table S5. Primers for constructing PX458-HF-Cas9 plasmids by Golden gate method

Name	Sequence (5'-3')	Length (bp)
hSpCas9-BsaI-FW	ATGGTCTCACCGGTGCCACCATGGACTATAAG	1637
N497A-BsaI-RV	ATGGTCTCAGGCGGTCATCCGCTCGATGAAGCTCTG	
N497A-BsaI-FW	ATGGTCTCACGCCTTCGATAAGAACCTGCCCAACGAGA	515
R661A-BsaI-RV	ATGGTCTCAGGGCGCCCCAGCCGGTGTATCTCCG	
R661A-BsaI-FW	ATGGTCTCAGCCCTGAGCCGGAAGCTGATCAACGGCA	124
Q695A-BsaI-RV	ATGGTCTCAGGGCCATGAAGTTTCTGTTGGCGAA	
Q695A-BsaI-FW	ATGGTCTCAGCCCTGATCCACGACGACAGCC	715
Q926A-BsaI-RV	ATGGTCTCATGGCCCGGGTTTCCACCAGCTGTC	
Q926A-BsaI-FW	ATGGTCTCAGCCATCACAAAGCACGTGGCACAGATCC	1400
hSpCas9-BsaI-RV	ATGGTCTCAAATTCCTTTTTCTTTTTTGCCTGG	

Table S6. Primers for constructing pET21-HF-Cas9 by Gibson assembly method

Code	Name	Sequence (5'-3')	Length (bp)
1	PX458-spCas9-HR-FW	TAACTTTAAGAAGGAGATATACATATGG ACTATAAGGACCACGACGGA	4269
	PX458-spCas9-HR-RV	TGGTGGTGGTGGTGGTGCTCGAGCTTT TTCTTTTTTGCCTGGCCGGC	
2	pET21b-HR-FW	CTCGAGCACCACCACCACCACCA	5000
	pET21b-HR-RV	ATGTATATCTCCTTCTTAAAGTTA	

The underline indicated the homology arm sequence of pET21b vectors.

Table S7. Primers for amplifying EMX1 gene in T7EI assay

Name	Sequence (5'-3')	Length (bp)
PX458-EMX1-sg345-F	GGCTTTACCATAGAGTCCTTG	650
PX458-EMX1-sg345-R	GCTTGCGTCCGAACTGGTATA	
PX458-EMX1-sg429-F	GGCTTTACCATAGAGTCCTTG	650
PX458-EMX1-sg429-R	GCTTGCGTCCGAACTGGTATA	
PX458-EMX1-sg576-F	CAGGCATTGTGAATGTGGGT	679
PX458-EMX1-sg576-R	CTCGAGGAAGAGGCTCTAAAGA	
PX458-EMX1-sg771-F	AAAACCACCCTTCTCTCTGGC	684
PX458-EMX1-sg771-R	GGAGATTGGAGACACGGAGAG	

Table S8. Oligonucleotides for the transcription template of EMX1-345 sgRNA

Name	Sequence (5'-3')
sgRNA-BB	GCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTA ACTTGCTATTTCTAGCTCTAAAAC
FS-EMX1-345	TAATACGACTCACTATAGGGCCTCGGGGAACACGAGCTCGTTTTAGAGCTA

Note: the underline indicated the T7 promoter.

Table S9. Primers for amplifying linear target DNA from pGL3-Fluc plasmids and for sequencing

Name	Sequence (5'-3')	Product length (bp)
pGL3-F529	CCACGCTGTTTTGACCTCCATA	531
pGL3-R1029	GCGAAATGCCCATACTGTTGAG	
siQuantRev primer	GCTGCGAAATGCCCATACTGTTG	

The core pGL3-Fluc sequence:

```
1       AATGTATTTA GAAAAATAAA CAAATAGGGG TTCCGCGCAC ATTTCCCCGA AAAGTGCCAC
61      CTGACGTCTA AGAAACCATT ATTATCATGA CATTAACCTA TAAAAATAGG CGTATCACGA
121     GGCCCTTTCG TCTTCACTCG AGTTTACCAC TCCCTATCAG TGATAGAGAA AAGTGAAAGT
181     CGAGTTTACC ACTCCCTATC AGTGATAGAG AAAAGTGAAA GTCGAGTTTA CCACTCCCTA
241     TCAGTGATAG AGAAAAGTGA AAGTCGAGTT TACCACTCCC TATCAGTGAT AGAGAAAAGT
301     GAAAGTCGAG TTTACCACTC CCTATCAGTG ATAGAGAAAA GTGAAAGTCG AGTTTACCAC
361     TCCCTATCAG TGATAGAGAA AAGTGAAAGT CGAGTTTACC ACTCCCTATC AGTGATAGAG
421     AAAAGTGAAA GTCGAGCTCG GTACCCGGGT CGAGGTAGGC GTGTACGGTG GGAGGCCTAT
481     ATAAGCAGAG CTCGTTTAGT GAACCGTCAG ATCGCCTGGA GACGCCATCC ACGCTGTTTT
541     GACCTCCATA GAAGACACCG GGACCGATCC AGCCTCCGCG GCCCCGAATT CTGCAGTCGA
601     CGGTACCGAG CTCTTACGCG TGCTAGCCCG GGCTCGAGAT CCGCACCATG GGCTGTGAGA
661     TCTNNNNNNN NNNNNNNNNN NNNNNNGGGC CCGGCGCCAT GGAAGACGCC AAAAACATAA
721     AGAAAGGCCC GGCGCCATTC TATCCGCTGG AAGATGGAAC CGCTGGAGAG CAACTGCATA
781     AGGCTATGAA GAGATACGCC CTGGTTCCTG GAACAATTGC TTTTACAGAT GCACATATCG
841     AGGTGGACAT CACTTACGCT GAGTACTTCG AAATGTCCGT TCGGTTGGCA GAAGCTATGA
901     AACGATATGG GCTGAATACA AATCACAGAA TCGTCGTATG CAGTGAAAAC TCTCTTCAAT
961     TCTTTATGCC GGTGTTGGGC GCGTTATTTA TCGGAGTTGC AGTTGCGCCC GCGAACGACA
1021    TTTATAATGA ACGTGAATTG CTCAACAGTA TGGGCATTTC GCAGCCTACC GTGGTGTTCG
1081    TTTCCAAAAA GGGGTTGCAA AAAATTTTGA ACGTGCAAAA AAAGCTCCCA ATCATCCAAA
1141    AAATTATTAT CATGGATTCT AAAACGGATT ACCAGGGATT TCAGTCGATG TACACGTTCG
1201    TCACATCTCA TCTACCTCCC GGTTTTAATG AATACGATTT TGTGCCAGAG TCCTTCGATA
1261    GGGACAAGAC AATTGCACTG ATCATGAACT CCTCTGGATC TACTGGTCTG CCTAAAGGTG
1321    TCGCTCTGCC TCATAGAACT GCCTGCGTGA GATTCTCGCA TGCCAGAGAT CCTATTTTTG
1381    GCAATCAAAT CATTCCGGAT ACTGCGATTT TAAGTGTTGT TCCATTCCAT CACGGTTTTG
1441    GAATGTTTAC TACACTCGGA TATTTGATAT GTGGATTTCG AGTCGTCTTA ATGTATAGAT
1501    TTGAAGAAGA GCTGTTTCTG AGGAGCCTTC AGGATTACAA GATTCAAAGT GCGCTGCTGG
1561    TGCCAACCCT ATTCTCCTTC TTCGCCAAAA GCACTCTGAT TGACAAATAC GATTTATCTA
1621    ATTTACACGA AATTGCTTCT GGTGGCGCTC CCCTCTCTAA GGAAGTCGGG GAAGCGGTTG
1681    CCAAGAGGTT CCATCTGCCA GGTATCAGGC AAGGATATGG GCTCACTGAG ACTACATCAG
1741    CTATTCTGAT TACACCCGAG GGGGATGATA AACCGGGCGC GGTCGGTAAA GTTGTTCCAT
1801    TTTTTGAAGC GAAGGTTGTG GATCTGGATA CCGGGAAAAC GCTGGGCGTT AATCAAAGAG
1861    GCGAACTGTG TGTGAGAGGT CCTATGATTA TGTCCGGTTA TGTAAACAAT CCGGAAGCGA
1921    CCAACGCCTT GATTGACAAG GATGGATGGC TACATTCTGG AGACATAGCT TACTGGGACG
1981    AAGACGAACA CTTCTTCATC GTTGACCGCC TGAAGTCTCT GATTAAGTAC AAAGGCTATC
2041    AGGTGGCTCC CGCTGAATTG GAATCCATCT TGCTCCAACA CCCCAACATC TTCGACGCAG
2101    GTGTCGCAGG TCTTCCCGAC GATGACGCCG GTGAACTTCC CGCCGCCGTT GTTGTTTTGG
2161    AGCACGGAAA GACGATGACG GAAAAAGAGA TCGTGGATTA CGTCGCCAGT CAAGTAACAA
2221    CCGCGAAAAA GTTGCGCGGA GGAGTTGTGT TTGTGGACGA AGTACCGAAA GGTCTTACCG
2281    GAAAACTCGA CGCAAGAAAA ATCAGAGAGA TCCTCATAAA GGCCAAGAAG GGCGGAAAGA
2341    TCGCCGTGTA A
```

There are two restriction endonuclease sites including *Bgl* II-*Apa* I. The underline represents the

binding region by CRISPR-Cas9 system. The highlighted sequence in gray represents the sequence of firefly luciferase.