

Supplementary Information

Different Effects of sgRNA Length on CRISPR-mediated Gene Knockout Efficiency

Jian-Ping Zhang^{1,2*}, Xiao-Lan Li^{1,2*}, Amanda Neises⁷, Wanqiu Chen⁷, Lin-Ping Hu^{1,2},
Guang-Zhen Ji^{1,2}, Jun-Yao Yu^{1,2}, Jing Xu^{1,2}, Wei-Ping Yuan^{1,2}, Tao Cheng^{1,2,3,4,5,6}, Xiao-Bing
Zhang^{1,2,3,4,7}

1. State Key Laboratory of Experimental Hematology;
2. Institute of Hematology and Blood Disease Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Tianjin, China;
3. Center for Stem Cell Medicine, Chinese Academy of Medical Sciences, Tianjin, China;
4. Department of Stem Cell & Regenerative Medicine, Peking Union Medical College, Tianjin, China;
5. Collaborative Innovation Center for Cancer Medicine, Tianjin, China;
6. Tianjin Key Laboratory of Blood Cell Therapy and Technology, Tianjin, China;
7. Department of Medicine, Loma Linda University, Loma Linda, CA 92350, USA.

*These authors contributed equally to this work.

Correspondences

Xiao-Bing Zhang, PhD. Division of Regenerative Medicine MC1528B, Department of Medicine, Loma Linda University. 11234 Anderson Street, Loma Linda, CA 92350. E-mail: xzhang@llu.edu. Phone: 909-651-5886. Fax: 909-558-0428.

Tao Cheng, MD. State Key Laboratory of Experimental Hematology, Institute of Hematology & Blood Diseases Hospital, Collaborative Innovation Center for Cancer Medicine, Tianjin, China; E-mail: chengtao@ihcams.ac.cn.

Supplementary Figure S1A

293T cells

sgGFP42- 17nt sgRNA

TGAACTTGTGGCCGTTTACGTCGCCGTCACCA¹GGA¹TGGGCACCACCCCGGTGAACAGCTC wild-type
TGAACTTGTGGCCGTTTACGTCGCCGTCAGCTC-----ATGGGCACCACCCCGGTGAACAGCTC-7
TGAACTTGTGGCCGTTTACGTCGCCGTCAGCTCGACC----GGGCACCACCCCGGTGAACAGCTC-5
TGAACTTGTGGCCGTTTACGTCGCCGTCAGCTCGAC--GATGGGCACCACCCCGGTGAACAGCTC-3
TGAACTTGTGGCCGTTTACGTCGCCGTCAGCTCGACC-GGATGGGCACCACCCCGGTGAACAGCTC-1
TGAACTTGTGGCCGTTTACGTCGCCGTCAGCTCGACC-GGATGGGCACCACCCCGGTGAACAGCTC-1
TGAACTTGTGGCCGTTTACGTCGCCGTCAGCTCGACCA¹GGATGGGCACCACCCCGGTGAACAGCT +1
TGAACTTGTGGCCGTTTACGTCGCCGTCAGCTCGAC¹gacgtCAGGATGGGCACCACCCCGGTGAAC +5
TGAACTTGTGGCCGTTTACGTCGCCGTCAGCTCGACC¹ggcaccggAGGATGGGCACCACCCCGGTG +8

sgGFP42-20nt sgRNA

TGAACTTGTGGCCGTTTACGTCGCCGTCACCA¹GGA¹TGGGCACCACCCCGGTGAACAGCTC wild-type
TGAACTTGTGGCCG-----GGGCACCACCCCGGTGAACAGCTC -29
TGAACTTGTGGCCGTTTACGTCGCCGTCAGCTCG-----ACCCCGGTGAACAGCTC -15
TGAACTTGTGGCCGTTTACGTCGCCGTCAGCT-----CACCACCCCGGTGAACAGCTC -13
TGAACTTGTGGCCGTTTACGTCGCCGTCAGCT-----CACCACCCCGGTGAACAGCTC -13
TGAACTTGTGGCCGTTTACGTCGCCGTCAGCTCG-----ACCACCCCGGTGAACAGCTC -12
TGAACTTGTGGCCGTTTACGTCGCCGTCAGCTC----GATGGGCACCACCCCGGTGAACAGCTC-6
TGAACTTGTGGCCGTTTACGTCGCCGTCAGCTC----GATGGGCACCACCCCGGTGAACAGCTC-6
TGAACTTGTGGCCGTTTACGTCGCCGTCAGCTC----GATGGGCACCACCCCGGTGAACAGCTC-6
TGAACTTGTGGCCGTTTACGTCGCCGTCAGCTCGACCA¹AGGATGGGCACCACCCCGGTGAACAGCT +1
TGAACTTGTGGCCGTTTACGTCGCCGTCAGCTCGACCA¹gacgtTGGGCACCACCCCGGTGAACA +4
TGAACTTGTGGCCGTTTACGTCGCCGTCAGCTCGAC¹gacgtCAGGATGGGCACCACCCCGGTGAAC +5
TGAACTTGTGGCCGTTTACGTCGCCGTCAGCTCGACCA¹ccccgatgaacagccGGATGGGCACCAC +15
TGAACTTGTGGCCGTTTACGTCGCCGTCAGCTCGAC¹gtccatccccgttacgtCAGGATGGGCACC +17

sgGFP101- 17nt sgRNA

CAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCAC¹CTA¹CGGCAAGCTGACCCTGAAGTTCATC wild-type
<-----CACCTACGGCAAGCTGACCCTGAAGTTCATC -83
CAAGTTCAGCGTGTCCG-----> -60
<-----ACGGCAAGCTGACCCTGAAGTTCATC -55
CAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGAGGGC-----AAGCTGACCCTGAAGTTCATC-15
CAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGC-----TGACCCTGAAGTTCATC -14
CAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCAC-----AAGCTGACCCTGAAGTTCATC-7
CAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATA-----CGGCAAGCTGACCCTGAAGTTCATC-7
CAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGC---TACGGCAAGCTGACCCTGAAGTTCATC-4
CAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCT-CGGCAA-CTGACCCTGAAGTTCATC-2
CAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCAC-TACGGCAAGCTGACCCTGAAGTTCATC-1

sgGFP101-20nt sgRNA

CAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCAC[↓]CTACGGCAAGCTGACCCTGAAGTTCATC wild-type
CAAGTTCAGCGTGTCCGGCGAGGGCGA----->-189
<-----CTACGGCAAGCTGACCCTGAAGTTCATC-47
CAAGTTCAGCGTGTCCGGCG-----AGGGC-----GACCCTGAAGTTCATC-26(-6,-20)
CAAGTTCAGCGTGTCCGG-----CGAGGGCGATG-----ACCCTGAAGTTCATC -23(-6,-17)
CAAGTTCAGCGTGTCCGGCGAGGGC-----AAGCTGACCCTGAAGTTCATC -21
CAAGTTCAGCGTGTCCGGCGAGGGCGAGGGC-----AAGCTGACCCTGAAGTTCATC-15
CAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATG-----GCAAGCTGACCCTGAAGTTCATC-9
CAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGAT-----ACGGCAAGCTGACCCTGAAGTTCATC -7
CAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGAT-----ACGGCAAGCTGACCCTGAAGTTCATC-7
CAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCAC-TACGGCAAGCTGACCCTGAAGTTCATC-1

sgGFP261- 17nt sgRNA

TGAAGAAGATGGTGCCTCCTGGACGTAGCCTTCGGGCA[↓]TGGCGGACTTGAAGAAGTCGTGCTGCTT wild-type
TGAAGAAGATGGTGCCTCCTGGACGTAGCCTTCG-----GGCGGACTTGAAGAAGTCGTGCTGCTT-5
TGAAGAAGATGGTGCCTCCTGGACGTAGCCTT-----CATGGCGGACTTGAAGAAGTCGTGCTGCTT-4
TGAAGAAGATGGTGCCTCCTGGACGTAGCCTTCGGGCAATGGCGGACTTGAAGAAGTCGTGCTGCT +1
TGAAGAAGATGGTGCCTCCTGGACGTAGCCTTCGGGCAATGGCGGACTTGAAGAAGTCGTGCTGCT +1
TGAAGAAGATGGTGCCTCCTGGACGTAGCCTTCGGGCAATGGCGGACTTGAAGAAGTCGTGCTGCT +1
TGAAGAAGATGGTGCCTCCTGGACGTAGCCTTCGGGCAATGGCGGACTTGAAGAAGTCGTGCTGCT +1
TGAAGAAGATGGTGCCTCCTGGACGTAGCCTTCGGGCAATGGCGGACTTGAAGAAGTCGTGCTGCT +1
TGAAGAAGATGGTGCCTCCTGGACGTAGCCTTCGGGCA[↓]tactggaATGGCGGACTTGAAGAAGTCGT +7

sgGFP261-20nt sgRNA

TGAAGAAGATGGTGCCTCCTGGACGTAGCCTTCGGGCA[↓]TGGCGGACTTGAAGAAGTCGTGCTGCTT wild-type
<-----GCGGACTTGAAGAAGTCGTGCTGCTT-106
TGAAGAAGATGGTGCCTCCTGGACGTAGCCTTCGGGCA-----TT-27
TGAAGAAGATGGTGCCTCCTGGACGTAGCCTTCGGGCA-----GCGGACTTGAAGAAGTCGTGCTGCTT -13
TGAAGAAGATGGTGCCTCCTGGACGTAGCCTTCGGG-----CTTGAAGAAGTCGTGCTGCTT -9
TGAAGAAGATGGTGCCTCCTGGACGTAGCCTTCG-----GGCGGACTTGAAGAAGTCGTGCTGCTT-5
TGAAGAAGATGGTGCCTCCTGGACGTAGCCTTCG-----GGCGGACTTGAAGAAGTCGTGCTGCTT-5
TGAAGAAGATGGTGCCTCCTGGACGTAGCCTTCGGGCA-GGCGGACTTGAAGAAGTCGTGCTGCTT-1
TGAAGAAGATGGTGCCTCCTGGACGTAGCCTTCGGGCAATGGCGGACTTGAAGAAGTCGTGCTGCT +1
TGAAGAAGATGGTGCCTCCTGGACGTAGCCTTCGGGCAATGGCGGACTTGAAGAAGTCGTGCTGCT +1
TGAAGAAGATGGTGCCTCCTGGACGTAGCCTTCGGGCAATGGCGGACTTGAAGAAGTCGTGCTGCT +1

sgGFP379- 17nt sgRNA

CCCTGGTGAACCGCATCGAGCTGAAGGGGCA[↓]AGGAGGACGGCAACATCCTGGGGCACA wild-type
CCCTGGTGAACCGCATCGAGCTGAAGGGCA-----ACATCCTGGGGCACA-21
CCCTGGTGAACCGCATCG-----AGGAGGACGGCAACATCCTGGGGCACA-21
CCCTGGTGAACCGCATCGAGCTGAAGGGGATCGACTTCAAGGA-----ATCCTGGGGCACA-10
CCCTGGTGAACCGCATCGAGCTGAAGGGGATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACA +1

CCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACA +1
 CCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACA +1
 CCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAAAGGAGGACGGCAACATCCTGGGGCACA +1
 CCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAggagggatcgacttcAGGAGGACGGCA +16

sgGFP379-20nt sgRNA

CCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCA¹AGGAGGACGGCAACATCCTGGGGCACAA wild-type
 <----->-291
 CCCTGGTGAACCGCATCGAGCTGAAGGGCATCGAC----->-70
 CCCTGGTGAACCGCATCGAGCTG-----AAGGAGGACG-----TCCTGGGGCACAA -21 (-15,-6)
 CCCTGGTGAACCGCATCGAGCTGA-----AGGACGGCAACATCCTGGGGCACAA-18
 CCCTGGTGAACCGCATCGAGCTG-----AAGGAGGACGGCAACATCCTGGGGCACAA-15
 CCCTGGTGAACCGCATCGAGCTG-----AAGGAGGACGGCAACATCCTGGGGCACAA-15
 CCCTGGTGAACCGCATCGAGCTG-----AAGGAGGACGGCAACATCCTGGGGCACAA-15
 CCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTC---AGGACGGCAACATCCTGGGGCACAA-4

Supplementary Figure S1A. DNA sequences of indel mutations induced by 17nt vs. 20nt sgGFPs that target GFP sites 42, 101, 261 and 379 in 293T GFP reporter cells.

Supplementary Figure S1B

iPSC

sgGFP101- 17nt sgRNA

CAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCAC[↓]CTACGGCAAGCTGACCCTGAAGTTCATC wild-type
CAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGC-----TGACCCTGAAGTTCATC -14
CAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGC-----TGACCCTGAAGTTCATC -14
CAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGC-----AAGCTGACCCTGAAGTTCATC -10
CAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCAC--ACGGCAAGCTGACCCTGAAGTTCATC -2
CAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCAC-TACGGCAAGCTGACCCTGAAGTTCATC -1
CAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACC[↓]TACGGCAAGCTGACCCTGAAGTTCAT +1

sgGFP101-20nt sgRNA

CAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCAC[↓]CTACGGCAAGCTGACCCTGAAGTTCATC wild-type
-----AAGCTGACCCTGAAGTTCATC -60
CAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCC-----CTGAAGTTCATC -18
CAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGC-----TGACCCTGAAGTTCATC -14
CAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACC-----AAGCTGACCCTGAAGTTCATC -6
CAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCAC----GGCAAGCTGACCCTGAAGTTCATC -4
CAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCAC-TACGGCAAGCTGACCCTGAAGTTCATC -1
CAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACC[↓]gatCTACGGCAAGCTGACCCTGAAGTT +4

sgGFP379- 17nt sgRNA

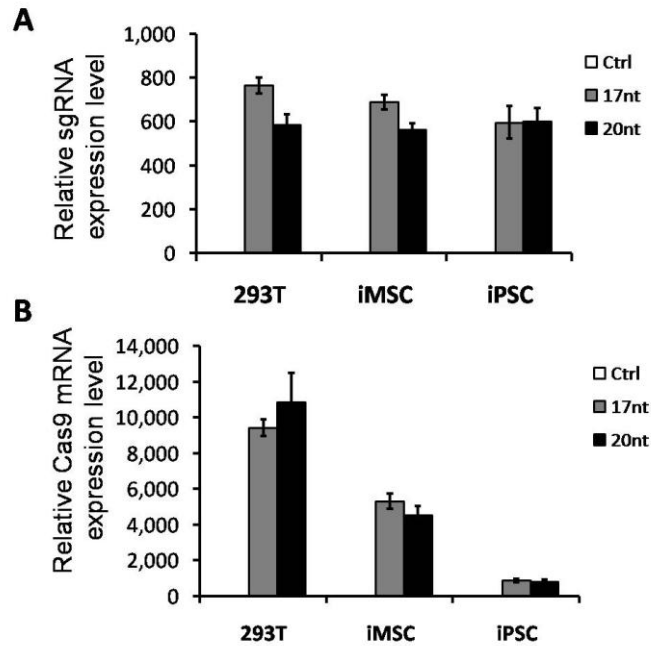
CCCTGGTGAACCGCATCGAGCTGAAGGGGCATCGACTTCA[↓]AGGAGGACGGCAACATCCTGGGGCACAA wild-type
CCCTGGTGAACCGCATCGAGCTGAAGGGGCATCGACTGAgcTgAAGGAGGACGGCAACATCCTGGGGCACAA mutation
CCCTGGTGAACCGCATCGAGCTGAAGGGGCATCGACTTcaAAGGAGGACGGCAACATCCTGGGGCACAA +1
CCCTGGTGAACCGCATCGAGCTGAAGGGGCATCGACTTcaAAGGAGGACGGCAACATCCTGGGGCACAA +1
CCCTGGTGAACCGCATCGAGCTGAAGGGGCATCGACTTcaAAGGAGGACGGCAACATCCTGGGGCACAA +1
CCCTGGTGAACCGCATCGAGCTGAAGGGGCATCGACTTcaAAGGAGGACGGCAACATCCTGGGGCACAA +1

sgGFP379-20nt sgRNA

CCCTGGTGAACCGCATCGAGCTGAAGGGGCATCGACTTCA[↓]AGGAGGACGGCAACATCCTGGGGCACAA wild-type
CCCTGGTGAACCGCATCGAGCT-----GAGGACGGCAACATCCTGGGGCACAA -19
CCCTGGTGAACCGCATCGAGCTGAAGGGGCATCGACT-----TCCTGGGGCACAA -18
CCCTGGTGAACCGCATCGAGCTG-----AAGGAGGACGGCAACATCCTGGGGCACAA -15
CCCTGGTGAACCGCATCGAGCTGAAGGGCA----TCAAGGAGGACGGCAACATCCTGGGGCACAA -6
CCCTGGTGAACCGCATCGAGCTGAAGGGGCATCGACTTC-AGGAGGACGGCAACATCCTGGGGCACAA -1
CCCTGGTGAACCGCATCGAGCTGAAGGGGCATCGACTTcaAAGGAGGACGGCAACATCCTGGGGCACAA +1
CCCTGGTGAACCGCATCGAGCTGAAGGGGCATCGACTTcagcgcaaggacggcaagGACTTCAAGGAGGACGG +18

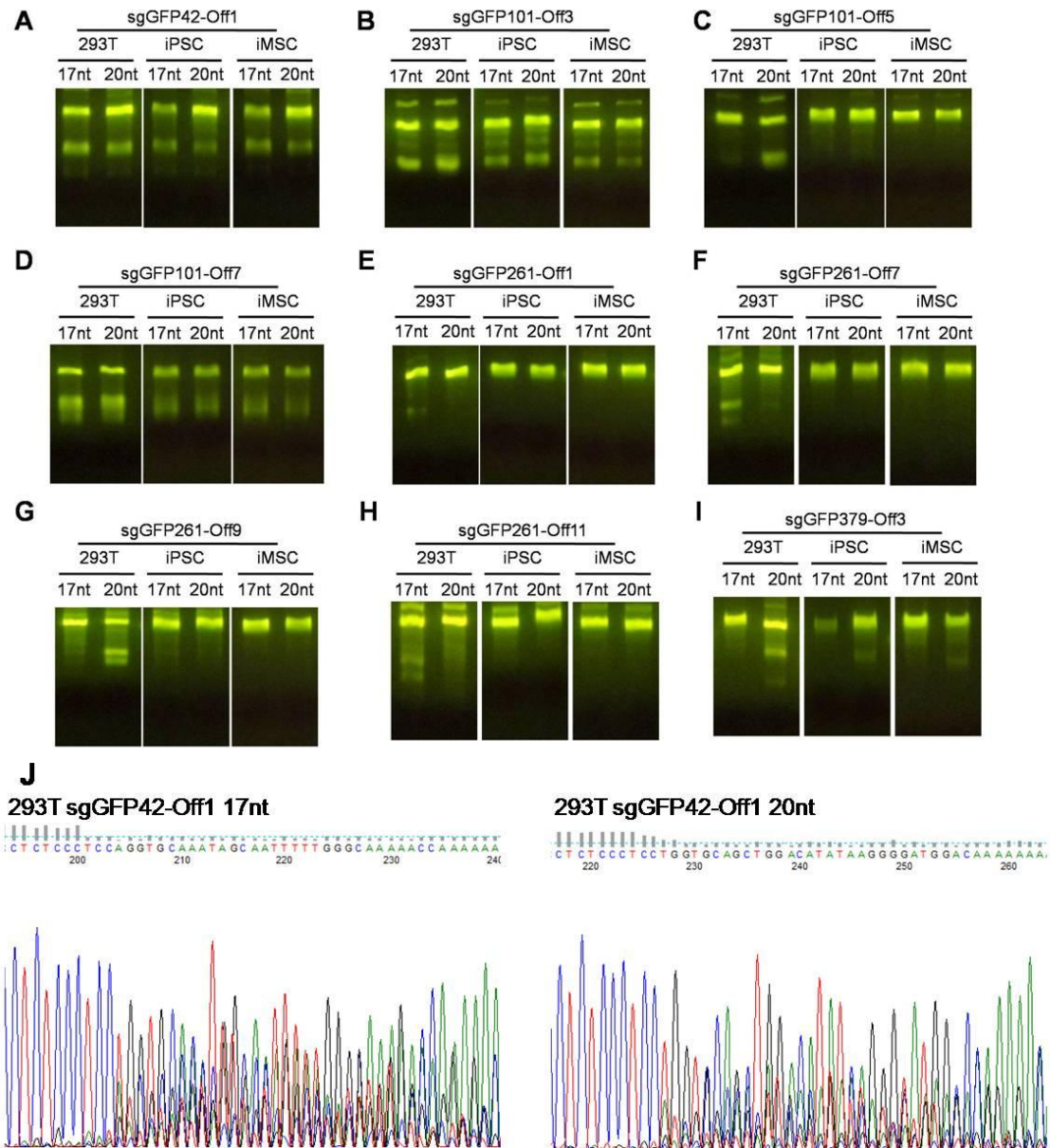
Supplementary Figure S1B. DNA sequences of indel mutations induced by 17nt vs. 20nt sgGFPs that target GFP sites 42, 101, 261 and 379 in iPSCGFP reporter cells.

Supplementary Figure S2



Supplementary Figure S2. Relative expression of 17nt vs. 20nt sgRNA and Cas9 in transduced 293T cells, iMSCs and iPSC. 293T, iPSC and iMSC cells were transduced with Lenti-U6-sgCD73a-SFFV-Cas9-2A-Puro vectors at an MOI of 1 in the presence of 8 $\mu\text{g/ml}$ protamine sulfate. Two days after transduction, cells were treated with 0.5-1 $\mu\text{g/ml}$ puromycin. At 10 days following puromycin selection, cells were harvested by treating with Accutase. Total RNA was extracted using miRCURY RNA Isolation Kit (EXIQON). Reverse transcription was performed using the EasyScript Plus cDNA Synthesis Kit (ABM), following the manufacturer's instructions.

Supplementary Figure S3



Supplementary Figure S3. Detection of off-targets by T7E1 mismatch nucleotide cleavage assay and DNA sequencing. Gel pictures of off-target amplicons after T7E1 digestion (A-I) and representative histograms of off-target amplicon sequencing (J) are shown.

Table S1. PCR primers used in this study

Name	Forward Primer	Reverse Primer
GFP	CAGGTGTCGTGAGCGATCGCC	GAACTCCAGCAGGACCATGT
sgGFP42-Off1	CCCAGTGACATTGCCCTCTT	CCAGGCTGCTGCTCATCATA
sgGFP42-Off2	ATGGCCAGGCTTCAGTACAC	AGGGTGGGAAGGAAAGGAGA
sgGFP42-Off4	GGGAGTTCGTCCTTTCTGCA	GGTCCCAGGGGAGTTTTTCC
sgGFP42-Off5	AAGCCTTCCAGAACAGCCTC	TGTTCTCAGACGTGGACAGC
sgGFP42-Off6	CCTTTCTCCTGCCCATCTGG	CCACTGTACCACGCTGTCTT
sgGFP42-Off7	CCCAGGCTGCAACATTTGAC	TGACTTGCTAAGAGGGCAGG
sgGFP42-Off9	CTCCAGGACCTGCCAAAGAG	GCCTCCAGGGTTAAGAGCAG
sgGFP42-Off10	ACAGCCCTTGTGTCATCCTC	CTCACTCTGTGCCAGTCCTG
sgGFP42-Off11	CAGTGTGGAACCTCGCCTTCT	GAAGCTGAGCCAAGCCCTAA
sgGFP42-Off12	CACCTTGGGGGAAGACTGAC	GGCCCTTCTTCTTTTGCTT
sgGFP42-Off13	CAGAACCCCTGGACACTGAC	CTCTATGACCACTTGCAGCA
sgGFP42-Off14	CCTGATAGGGAAGCAAGGCC	GCAAATAAGGGGTGGAGGCT
sgGFP42-Off15	GGGTCAGGCCAGGTCTAAAC	TCCTGCCTCACAGCCTATCT
sgGFP42-Off16	TGGAACCTGGGGCTCAAATCC	CTGGGGACAATGAGGCTGAG
sgGFP101-Off1	GGGTGGAAGAAAGCACGGTA	CCCCCATTCATCCAGGTCAG
sgGFP101-Off2	AAGTCTTGGGGCTTTGAGG	AGAACACCCGATTCAGCCTG
sgGFP101-Off3	ACTAAGGTACCAGCCAAACACA	GGGTGAGAGCCAAGGAAACA
sgGFP101-Off4	ACACATGTTCTTGCACGTGC	CTGCCTCACTCTTCTGGCTC
sgGFP101-Off5	TCGGCACTTTGATGAGCACT	GGTTCACCAAGCTCAGGGTT
sgGFP101-Off6	GGATCGCTACAACCCAGGAG	TGCGCTCAAACCATCTTCTT
sgGFP101-Off7	GCCAACCTGGAGAAGTGGA	ACTCCCATCCTGACTAGTGCT
sgGFP261-Off1	AGACCACAATTCCCGGCATT	CGAACGGGTTCCAGAGTTCT
sgGFP261-Off3	TGTGAGCCCCTCAGAGAGAG	ATGGATCCTCCCCCTCTCAC
sgGFP261-Off4	CAGCTCCATGACCTACTGCC	GTGCCCTACCTCACATTCC
sgGFP261-Off5	CAAGTCCCCGCTCATCTCAG	AGACCCGGCTCCCTCTATTT
sgGFP261-Off6	GCCTTCTTCGGACCTTACACA	CCTGCAGTAGACTCCCTTGC
sgGFP261-Off7	AGGCTCCCTACTCTGTCTG	ACCGTTTGTTGCTTGTATGCA
sgGFP261-Off9	CAGCACCTCCTTCCCTTCTG	CCTCAGCCTCCTCCCTTAGT
sgGFP261-Off10	GGCCACGCGTAGGAATTAGA	TCTTCCACAGGGGCAACATC
sgGFP261-Off11	GCACACGGTGTACATCAAC	CCCAGAAATCCCACCCCATC
sgGFP379-Off3	CGGCCAAGTGGTCAGTTTTT	AGATGGGCCAACTTCTCAGC
sgRNA-qPCR	AGCTAGAAATAGCAAGTTAAATAAGG	GACTCGGTGCCACTTTTTTCA
Cas9-qPCR	CCGAAGAGGTCGTGAAGAAG	GCCTTATCCAGTTCGCTCAG
GAPDH-qPCR	GTGGACCTGACCTGCCGTCT	GGAGGAGTGGGTGTCGCTGT