

1 **SUPPLEMENTAL MATERIAL**

2 **Cas12a mediates efficient and precise endogenous gene tagging via MITI: microhomology-**
3 **dependent targeted integrations**

4 Pan Li^{1,2} · Lijun Zhang^{1,2} · Zhifang Li^{1,2} · Chunlong Xu³ · Xuguang Du^{1,2} · Sen Wu^{1,2}

5

6 ¹ Beijing Advanced Innovation Center for Food Nutrition and Human Health, China Agricultural
7 University, No 2 Yuanmingyuan West Road, Beijing 100193, China

8 ² State Key Laboratory of Agrobiotechnology, College of Biological Sciences, China Agricultural
9 University, No 2 Yuanmingyuan West Road, Beijing 100193, China

10 ³ Institute of Neuroscience, State Key Laboratory of Neuroscience, Key Laboratory of Primate
11 Neurobiology, CAS Center for Excellence in Brain Science and Intelligence Technology, Shanghai
12 Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China

13

14 Pan Li and Lijun Zhang contributed equally to this work

15 To whom correspondence may be addressed. Tel: +1-86-10-6273-4163, Email: swu@cau.edu.cn

16 Correspondence may also be addressed to Xuguang Du. Tel: +1-86-10-6273-3075, Email:
17 xuguangdu@cau.edu.cn

18

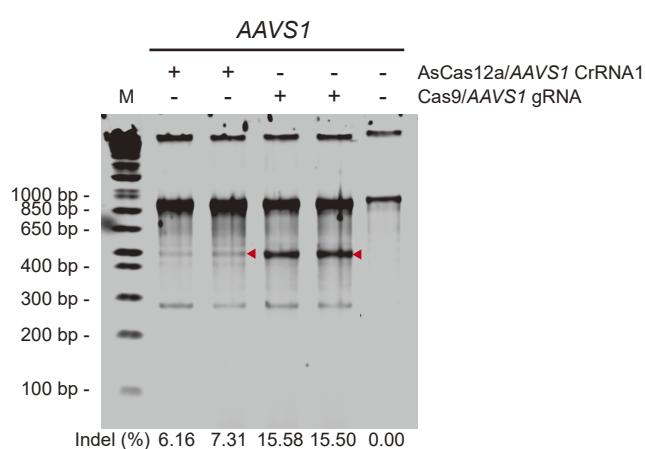
19

20

21

22

23

A

24

Supplemental Figure S1. The verification and comparison of the target efficiency between the Cas9 and Cas12a at the *AAVS1* locus. T7E1 analyses showing that the gRNA of Cas9 is more efficient than that of Cas12a when the same *AAVS1* site was targeted. The result was presented as mean \pm SD, n=2, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, unpaired Student's t-test.

29

30

31

32

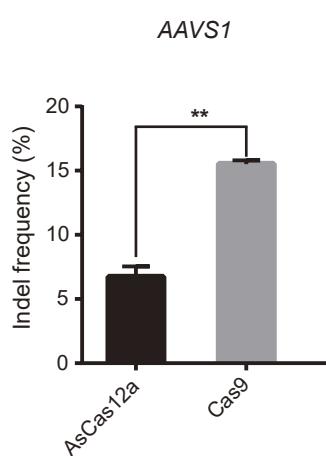
33

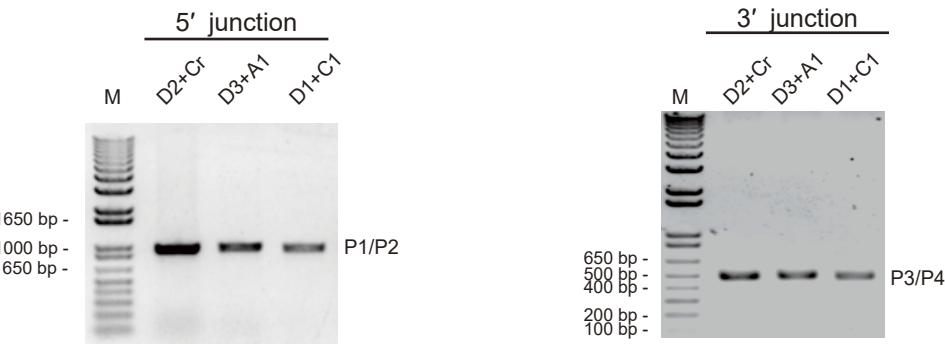
34

35

36

37

B

A**B**

5' junction (Cas12a HITI)

AAVS1 locus: 5'-CCCCCACAGTGGGGC**CACTA**GGGACAGGATTGGTGACAGAAA-3'
3'-GGGGTGTACCCCG**GAT**CCCTGTCTAACCACTGTCTTT-5'
D2 vector: 5'-**TTTC**TGTACCAATCCTGTCCC**TAGT**GCTCGAGGCTGGCAC-3'
3'-**AAAG**ACAGTGGTAGGACAGGG**ATCA**CAGAGCTCCGACCGTG-5'
6/13 (46.15%) 5'-CCCCCACAGTGGGGC**CAGT**GCTCGAGGCTGGCAC-3'
2/13 (15.38%) 5'-CCCCCACAGTGGGGC**CACT**GCTCGAGGCTGGCAC-3'
1/13 (7.69%) 5'-CCCCCACAGTGGGGC-----TGGCAC-3'
1/13 (7.69%) 5'-TCTTC ----- (del 174 bp) -----TGGCAC-3'
1/13 (7.69%) 5'-ATGGC ----- (del 201 bp) -----TCGAGG-3'
1/13 (7.69%) 5'-ACAGC ----- (del 128 bp) -----TCGAGG-3'

C

5' junction (Cas9 HITI)

AAVS1 locus: 5'-**CACAGT**GGGGCCACTAGGGACAGGATTGGTGACAGAAA-3'
3'-**GTC**ACCCCG**GAT**CCCTGTCTAACCACTGTCTTT-5'
D1 vector: 5'-**TAGT**TTCTGTACCAATCCTGTCCC**TAGT**GCTCGAGGCTGG-3'
3'-**ATCAA**ACAGTGGTAGGACAGGG**ATCA**CACCGAGCTCCGACC-5'
5'-**CACAGT**GGGGCCACTAGTAGGGCTCGAGGCTGG-3'
5'-**CACAGT**GGGGCCACTAGTAGGGCTCGAGGCTGG-3'
5'-**CACAGT**GGGGCCACT-**GT**-GGGCTCGAGGCTGG-3'
5'-**CACAGT**GGG-----CTCGAGGCTGG-3'
5'-**CACAGT**GGGGCCACT---**GT**GGGCTCGAGGCTGG-3'
5'-**CACAGT**GGGGCCACTAG-----AGGCTGG-3'
5'-**CACAGT**GGAGGCCACT-**GT**AGGGCTCGAGGCTGG-3'
5'-**CACAGT**GGGGCCACTAG-(del 132 bp)-**TTCAC**-3'
5'-**CACAGT**GGG-----**CGCA**CTGGAAAGC-3'
5'-**GGTACT**TTT-(del 36 bp)-**TGG**GCTCGAGGCTGG-3'
5'-**GGTCC**-(del 185 bp)-**AGT**GGGCTCGAGGCTGG-3'
5'-**CCAGGGCCGG**-(del 112 bp)-----**TGA**-3'

D

5' junction (Cas12a MITI)

AAVS1 locus: 5'-CCCCCACAGTGGGGC**CACTA**GGGACAGGATTGGTGACAGAAA-3'
3'-GGGGTGTACCCCG**GAT**CCCTGTCTAACCACTGTCTTT-5'
D3 vector: 5'-**TTTC**TGTACCAATCCTGTCCC**CACT**A**CTCG**AGGCTGGCACG-3'
3'-**AAAG**ACAGTGGTAGGACAGGG**GAT**GAGCTCCGACCGTGC-5'

E

3' junction (Cas12a MITI)

D3 vector: 5'-**TTTC**TGTACCAATCCTGTCCC**CACT**A**CTCG**AGGCTGGCACG-3'
3'-**AAAG**ACAGTGGTAGGACAGGG**GAT**GAGCTCCGACCGTGC-5'
AAVS1 locus: 5'-CCCCCACAGTGGGGC**CACTA**GGGACAGGATTGGTGACAGAAA-3'
3'-GGGGTGTACCCCG**GAT**CCCTGTCTAACCACTGTCTTT-5'
5'-**TTTC**TGTACCAATCCTG-----GG----TTGGTGACAGAAA-3'
5'-**TTTC**TGTACCAATCCTGTC-----ACAGGATTGGTGACAGAAA-3'
5'-**TTTC**TGTACCAATCCTGTC-----A-AGGATTGGTGACAGAAA-3'
5'-**TTTC**TGTACCAATCCT----- (del about 157 bp) -3'
5'-**TTTC**TGTACCAATCCT-----AGGATTGGTGACAGAAA-3'
5'-**TTTC**TGTCA-----AA-3'

38

Supplemental Figure S2. Analyze the validity and accuracy of two junctions. **(A)** PCR analysis of 5' and 3' junctions of cells with Cas12a HITI, Cas12a MITI, and Cas9 HITI targeted integrations at the *AAVS1* locus in HeLa cells. Genomic PCR products amplified from pooled HeLa cells transfected with three groups of plasmids. The first group contained the Cas12a HITI donor (D2), *AAVS1* CrRNA1 and the Cas12a targeting vector (Cr); the second group included the Cas12a MITI donor (D3), *AAVS1* CrRNA1, *AAVS1* CrRNA1.1, and the Cas12a targeting vector (A1); and the third group had the Cas9 HITI donor (D1), *AAVS1* gRNA1 and Cas9 targeting vector (C1). M is the 1kb plus ladder maker. **(B)** The representative TA cloning sequence analysis at the 5' target junction of *AAVS1* targeted integration events mediated by Cas12a HITI strategy after PCR-based amplification. **(C)** The representative TA cloning sequence analysis at the 5' target junction of *AAVS1* targeted integration events mediated by Cas9 HITI strategy. **(D and E)** The representative TA cloning sequence analysis at the 5' and 3' target junction of *AAVS1* targeted integration events mediated by Cas12a MITI strategy.

39

40

41

42

43

44

45

46

47

48

49

50

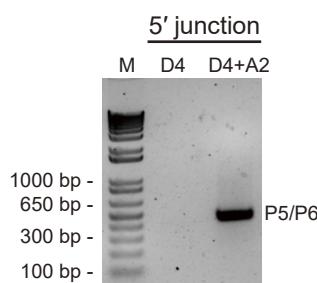
51

52

53

54

Cas12a MITI



5' junction (Cas12a MITI)

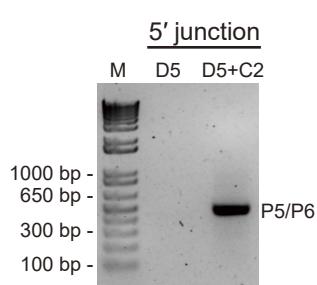
CLTA locus: 5'-GGCCCCGCTG**GTGCA**TGAAGAGCCACCCCTGTGG**AAA**-3'
3'-CCGGGGCGAC**CACGT**GACTTCTCGGTGGACACC**TTT**-5'

D4 vector: 5'-**TTTC**CACAGGGTGGCTCT**TCAG****GTGCA**CAGGCCGGCC-3'
3'-**AAAG**GTGTCCCACCGAGA**AGTCCACGT**GTCCGGCCGG-5'

14/19 (73.68%) 5'-GGCCCCGCTG**GTGCA**CAGGCCGGCC-3'
1/19 (5.26%) 5'-GGCCCCGCTG**GCGCA**CAGGCCGGCC-3'
1/19 (5.26%) 5'-GCC-----**GTGCA**CAGGCCGGCC-3'
1/19 (5.26%) 5'-GGCCCCGCTG-T-----GGC-3'
1/19 (5.26%) 5'-GGCCCCGCTG**GTGCA**-----CGGCC-3'
1/19 (5.26%) 5'-GCC-----GGC-3'

55

Cas9 HITI



5' junction(Cas9 HITI)

CLTA locus: 5'-GCC**CCGCTG**|GTGCACT**TGA**AGAGCCACCCCTGTGG**AAA**-3'
3'-CCGG**GGCGAC**|CACGTG**A**CTTCTCGGTGGACACC**TTT**-5'

D5 vector: 5'-**TAAGTGGCTT**TCAGTG**AC**CAGGGAGGCCGG**GAC**-3'
3'-**ATTCACCGAGA****AGT**CACGTG|GTCCCTCGGG**GGCTG**-5'

6/16 (37.50%) 5'-GGCCCCGCTGCAGCGGAGGCCGG**GAC**-3'
3/16 (18.75%) 5'-GGCCCCGCTGCCAGCGGAGGCCGG**CGA**-3'
2/16 (12.50%) 5'-GGCCCCGCTGG**CCAGCGGAGGCCGGCG**-3'
1/16 (6.25%) 5'-GGCCCCGCT-----GGAGGCCGG**GAC**-3'
1/16 (6.25%) 5'-GGCCCCGCTG**AAGCGG**-GCCGG**GAC**-3'
1/16 (6.25%) 5'-GGCCCCGCTG**AAAG**-----GCCGG**GAC**-3'
1/16 (6.25%) 5'-TCATCTCCCT--(del 23 bp)--**CGGCCGAC**-3'
1/16 (6.25%) 5'-GGCGGCTGTG-----**A**-GCCGG**GAC**-3'

56

57

58

59

Supplemental Figure S3. Detection of target integration at the 5' junction at the *CLTA* locus. Genomic PCR products amplified from pooled HEK293T cells transfected with *CLTA* donor and the corresponding Cas12a or Cas9 targeting vector and the TA cloning sequence analysis at the 5' junction of *CLTA* targeted integration events after PCR-based amplification. A2 represents the Cas12a co-expression plasmid with the *CLTA* CrRNA array. C2 represents the *CLTA* gRNA and Cas9 co-expression plasmid. P5 and P6 primers are utilized to amplify the 5' junction. The right panel is the representative TA cloning sequence results of the 5' target junction of *CLTA* integration using the Cas12a MITI or Cas9 HITI strategy.

60

61

62

63

64

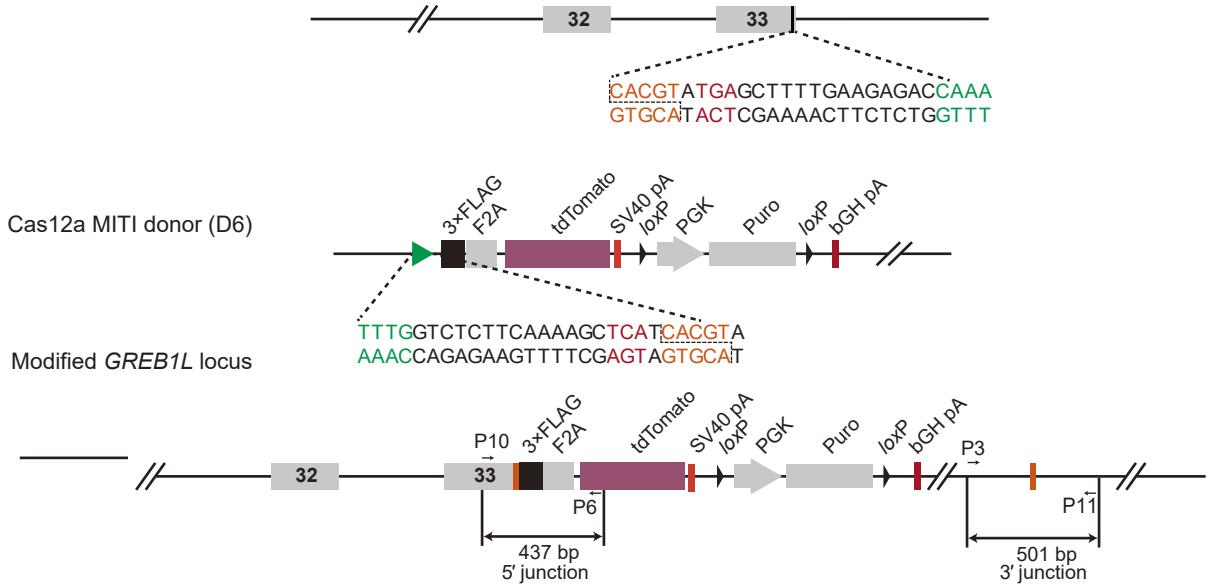
65

66

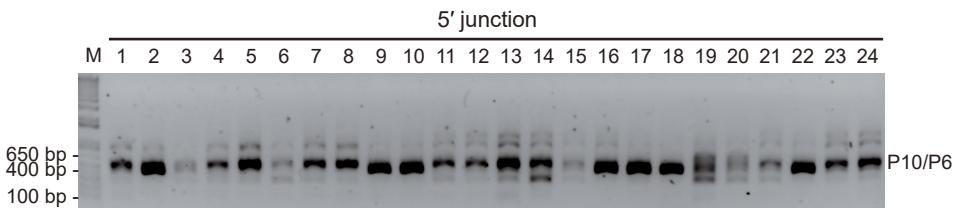
67

68

A Pig GREB1L locus



B



C

5' junction (Cas12a MITI)

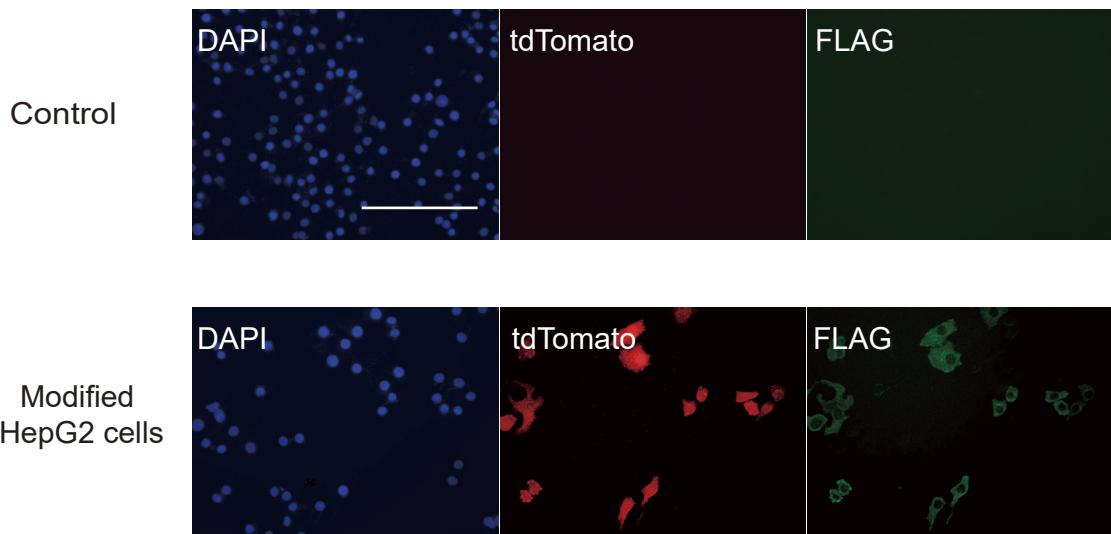
GREB1L locus: 5'-CTTTCTTACTGGACGT**CACGT**ATGA GCTTTGAAGAGAC**CAAA**ACTAG -3'
3'-GAAAGAACATGACCTTG**CA****GTGCA**TACTCGAAA**ACTTC**CTG**GTTTG**ATC -5'

D6 vector: 5'-TTAATT**TTG**TCTCTCAAAG**TCA****T****CACGT**AAGGCCGGCCGACTATA-3'
3'-AATT**AAAC**CAGAGAAGT**TTTG****AGT****AGTGC**A**T****TCCGGCCGG**CTGATAT-5'

#1 5'-CTTTCTTACTGGACGT**CACGT**AAGGCCGGCCGACTATA -3'
#2 5'-CTTTCTTACTGGACGT**CACGT**AAGGCCGGCCGACTATA -3'
#3 5'-CTTTCTTACTGGACGT**CACGT**AAGGCCGGCCGACTATA -3'
#4 5'-CTTTCTTACTGGACGT**CACGT**AAGGCCGGCCGACTATA -3'
#5 5'-CTTTCTTACTGGACGT**CACGT**AAGGCCGGCCGACTATA -3'
#6 5'-CTTTCTTACTGGACGT**CACGT**AAGGCCGGCCGACTATA -3'
#7 5'-CTTTCTTACTGGACGT**CACGT**AAGGCCGGCCGACTATA -3'
#8 5'-CTTTCTTACTGGACGT**CACGT**AAGGCCGGCCGACTATA -3'
#9 5'-CTTTCTTACTGGACGT**CACGT**AAGGCCGGCCGACTATA -3'
#10 5'-CTTTCTTACTGGACGT**CACGT**AAGGCCGGCCGACTATA -3'
#11 5'-CTTTCTTACTGGACGT**CACGT**AAGGCCGGCCGACTATA -3'
#12 5'-CTTTCTTACTGGACGT**CACACGT**AAGGCCGGCCGACTATAA -3'
#13 5'-CTTTCTTACTGGACGT**CACGT**AAGGCCGGCCGACTATA -3'
#14 5'-CTTTCTTACTGGACGT**CACGT**AAGGCCGGCCGACTATA -3'
#15 5'-CTTTCTTACTGGACGT**CACGT**AAGGCCGGCCGACTATA -3'
#16 5'-CTTTCTTACTGGACGT**CACGT**AAGGCCGGCCGACTATA -3'
#17 5'-CTTTCTTACTGGACGT**CACGT**AAGGCCGGCCGACTATA -3'
#18 5'-CTTTCTTACTGGACGT**CACGTGGT**AAGGCCGGCCGACTATAA -3'
#21 5'-CTTTCTTACTGGACGT**CACGT**AAGGCCGGCCGACTATA -3'
#22 5'-CTTTCTTACTGGACGT**CACGT**AAGGCCGGCCGACTATA -3'
#23 5'-CTTTCTTACTGGACGT**CACGT**AAGGCCGGCCGACTATA -3'
#24 5'-CTTTCTTACTGGACGT**CACGT**AAGGCCGGCCGACTATA -3'

69
70

71 **Supplemental Figure S4.** Tagging the *GREB1L* gene in pig fetal fibroblasts (PFFs) cells using the
72 MITI approach. (A) Strategy for targeting *GREB1L* locus in PFF cells. (B) PCR identification of
73 positive PFF clones bearing predicted integration of 3×FLAG-F2A-tdTomato. (C) The 5' junction
74 sequences of positive PFF clones.



75 **Supplemental Figure S5.** Immunostaining results of tdTomato positive HepG2 cells. The tdTomato
76 positive HepG2 cells bearing 3×FLAG-2A-tdTomato integration in *CLTA* were fixed, stained with anti-
77 FLAG antibody and examined by fluorescence microscopy. Scale bar, 200 μ m
78
79
80

81

82

83

84

85

86

87

88

Supplementary Tables

Supplementary Table S1. The primers for donor construction. All sequences are in the 5' to 3' direction

Primer name	Sequence
D1/D3-linker-F	CTAGTTCTGTCACCAATCCTGTCCCTAGTGGC
D1/D3-linker-R	TCGAGCCACTAGGGACAGGATTGGTGACAGAAA
D2-linker-F	CTAGTTCTGTCACCAATCCTGTCCCCACTAC
D2-linker-R	TCGAGTAGTGGGGACAGGATTGGTGACAGAAA
D4-linker-F	TATGtaattaaTTTCCACAGGGTGGCTCTTCAGGTGCAcAGGCCGG
D4-linker-R	CCTgTGCACCTGAAGAGCCACCCTGTGGAAAttaattaaCA
D5-linker-F	TATGtaattaaGTGGCTCTTCAGTGCACCAGCGGAGGCCGG
D5-linker-R	CCTCCGCTGGTGCACTGAAGAGCCACtttaattaaCA
D6-linker-F	TATGtaattaaTTTGGTCTCTCAAAAGCTCATCACGTaAGGCCGG
D6-linker-R	CCTtACGTGATGAGCTTGAAGAGACCAAAttaattaaCA
D7-linker-F	CGCGCCAGGATCTCTGGCTCCATCGTAAGCAAAACCGGTGTCGACA
D7-linker-R	GATCTGTCGACACCGCGTTGCTTACGATGGAGCCAGAGATCCTGG

89

Supplementary Table S2. The primers for CrRNAs construction. All oligos are in the 5' to 3' direction.

Primer name	Sequence
<i>AAVS1</i> -CrRNA1-F	agatTGTCACCAATCCTGTCCCTAGTG
<i>AAVS1</i> -CrRNA1-R	aaaaCACTAGGGACAGGATTGGTGACA

<i>AAVS1</i> -CrRNA1.1-F	agatTGTCACCAATCCTGTCCCCACTA
<i>AAVS1</i> -CrRNA1.1-R	aaaaTAGTGGGGACAGGATTGGTGACA
<i>AAVS1</i> -sgRNA1-F	CACCGTCACCAATCCTGTCCCTAG
<i>AAVS1</i> -sgRNA1-R	AAACCTAGGGACAGGATTGGTGAC
<i>CLTA</i> -CrRNA-array-F	agatCACAGGGTGGCTCTCAGTGCACaaatttctactttgttagatCACA GGGTGGCTCTCAGGTGCA
<i>CLTA</i> -CrRNA-array-R	aaaaTGCACCTGAAGAGGCCACCCTGTGatctacaagagttagaaattGT GCACTGAAGAGGCCACCCTGTG
<i>CLTA</i> -sgRNA-F	CACC GTGGCTCTCAGTGCACCAG
<i>CLTA</i> -sgRNA-R	AAAC CTGGTGCACTGAAGAGCCAC
<i>GREB1L</i> -CrRNA-array-F	agatGTCTCTTCAAAAGCTCATACGTGaaatttctactttgttagatGTCT CTTCAAAAGCTCATCACGT
<i>GREB1L</i> -CrRNA-array-R	aaaaACGTGATGAGCTTTGAAGAGACatctacaagagttagaaattCA CGTATGAGCTTTGAAGAGAC
<i>AAVS1</i> -CrRNA-array-F1	agatTGTCACCAATCCTGTCCCTAGTGAATTCTACTCTGT AGATTGTCACCAATCCTGTCC
<i>AAVS1</i> -CrRNA-array-F2	CCACTAaatttctactttgttagatCTTACGATGGAGGCCAGAGAGGAT aatttctactttgttagatCTTACGATGGAGGCCAGAGATCCT
<i>AAVS1</i> -CrRNA-array-R1	aaaaAGGATCTCTGGCTCCATCGTAAGAtctacaagagttagaaattATC CTCTCTGGCTCCAT

	CGTAAGatctacaagagttagaaattTAGTGGGGACAGGATTGGTGAC
<i>AAVS1</i> -CrRNA-array-R2	AATCTACAAGAGTAGAAATTCACTAGGGACAGGATTGGT
	GACA

90

Supplementary Table S3. Primers used for Knock-in detection

Primer name	Sequence
P1	TGCCATCTCTCGTTCTAGGATG
P2	cagaTcgataaaacacatgcgtcaattt
P3	GCGTTTCGGTGATGACGGTG
P4	CTGCCAAGCTCTCCCTCCCAG
P5	GGGACAAATAGGCAGTTGCT
P6	tcctcgcccttgctaccat
P7	CTCTGAATGCCAGGGAGAAC
P8	TCTGTTCCACATACACTTCATT
P9	CCCGGTGCCTGAGATAaacG
P10	CGGCTGTCACATCTGGTTT
P11	TCCAAGCATCTCCTCAGGC
P12	CAGGACGGGGCTGGCTACTG

91

Supplementary Table S4. Primers used for T7E1 assay

Primer name	Sequence
<i>AAVS1</i> -T7E1-sur-F	TGCCATCTCTCGTTCTAGGATG

AAVS1-T7E1-sur-R

CTGCCAAGCTCTCCCTCCCAG

92