Efficient *in vivo* editing of OTC-deficient patient-derived primary human hepatocytes

Samantha L Ginn, Anais K Amaya, Sophia H Y Liao, Erhua Zhu, Sharon C Cunningham, Michael Lee, Claus V Hallwirth, Grant J Logan, Szun S Tay, Anthony J Cesare, Hilda A Pickett, Markus Grompe, Kimberley Dilworth, Leszek Lisowski, Ian E. Alexander

Table of contents

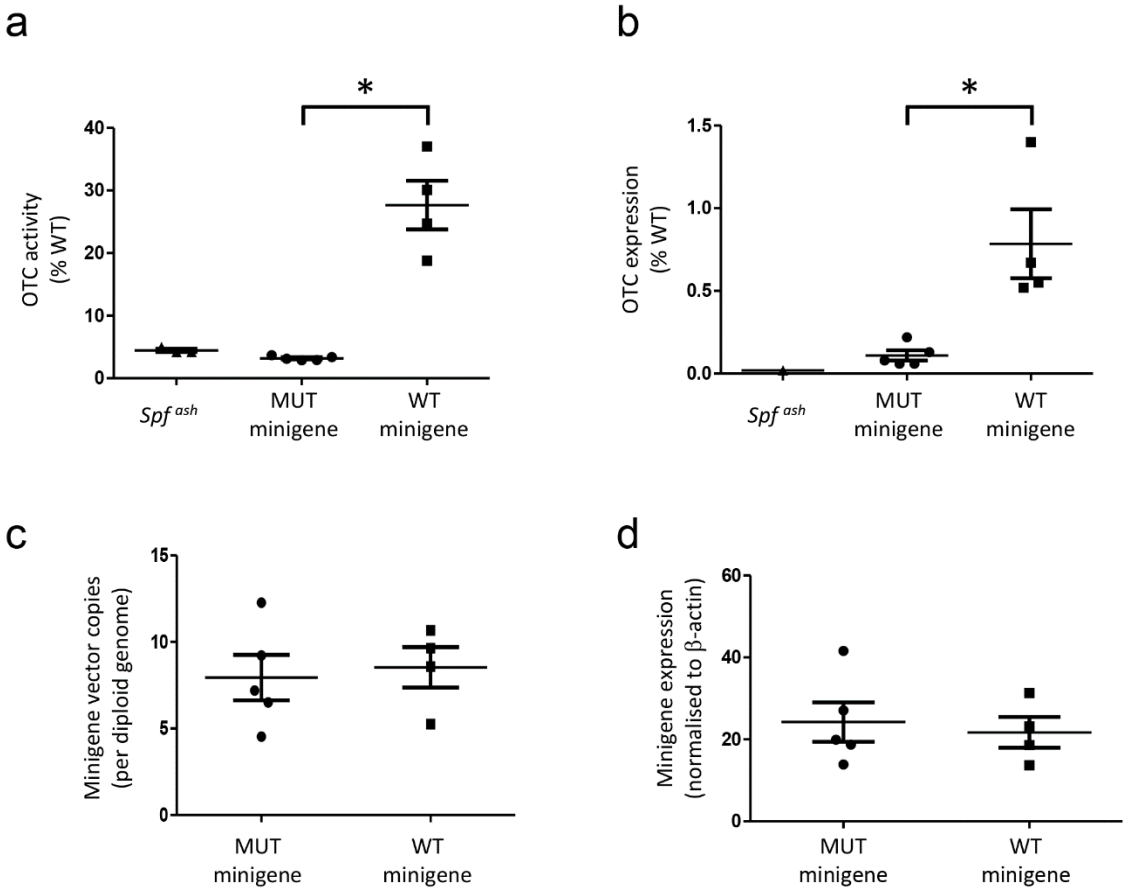


Fig. S1. Functional analysis of mutant and wild-type human *OTC* minigenes. Minigenes were delivered to newborn *spf*^{ash} mice using a hybrid AAV/*piggyBac* transposase vector system by i.p. injection. Animals received 1.5×10^{10} vg transposase and 1×10^{11} vg minigene vectors (both packaged in the rAAV2/8 capsid) and were harvested four weeks later (n = 5 for the mutant and n = 4 for the wild-type minigenes, respectively). OTC activity was determined **(a)** in whole liver lysates (expressed as the percentage of OTC activity of a wild-type animal; P = 0.0159). **(b)** Quantification of Western blot analysis showed significantly lower OTC protein levels in mice receiving the mutant minigene, consistent with the severe phenotype conferred by this mutation in humans (P = 0.0159). Equivalent delivery of mutant and wild-type minigenes was confirmed by using quantitative PCR on whole liver lysates for both **(c)** vector copies and **(d)** mRNA expression (cDNA). Data are plotted as mean ± s.e.m. and significance evaluated using the Mann-Whitney non-parametric test.

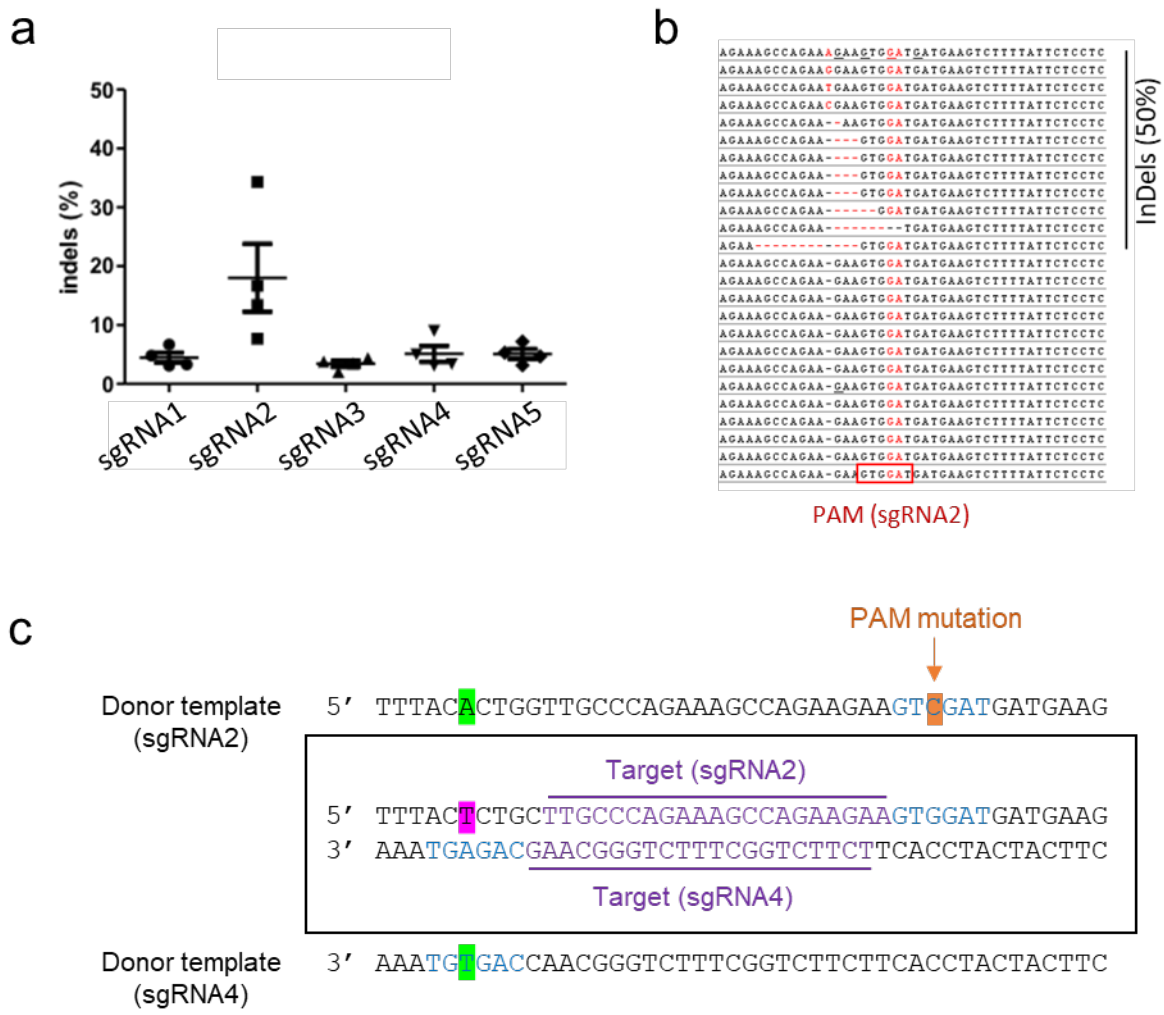


Fig. S2. Functional analysis of sgRNAs used to disrupt the human *OTC* locus. The presence of small insertions and deletions (InDels) was determined at 5 weeks post-minigene delivery by using **(a)** Tracking of InDels by Decomposition (TIDE) analyses.[1] Data are plotted as mean ± s.e.m.. This was further confirmed by sequencing cloned amplicons across the cleavage site, with **(b)** those for one representative mouse receiving the sgRNA2 vector shown. **(c)** Region of *OTC* exon 9 showing the location of the guide sequences for sgRNA2 and sgRNA4; sgRNA2 contains a mutation in the PAM sequence to prevent re-cleavage of repaired alleles, and for sgRNA4, upon correction, the PAM sequence is abolished. Targets for sgRNA2 and sgRNA4 are indicated by the purple line; mutant and wild-type nucleotides are highlighted in pink and green, respectively; PAM mutation is indicated in orange

Fig. S3. Amplification across the transposed *OTC* locus reveals large insertions containing ITR sequences. (a) PCR amplification strategy to detect editing events at the transposed *OTC* locus. The first round of amplification used primers (purple) that bind the minigene outside the region of homology with the donor template (indicated by dotted black line). Nested PCR was then performed with primers (green) to produce an amplicon of a suitable size for next generation Illumina® sequencing (318 bp) and analyzed using CRISPResso.[2] Insert sequences were extracted from CRISPResso output and aligned to the vector sequences using blastn (version 2.6.0+).[3] **(b)** Next generation sequencing revealed that, in addition to the expected

InDels at the target site, large insertions of up to 209 bp were identified. Notably, the majority of these large insertions contained ITR sequences. Alignment of representative reads that contain insertions of ITR sequence, identified in two treated *Spf*^{ash} mice (m218 and m219), are shown; yellow shading indicates exon 9 sequence adjacent to the predicted SaCas9 cleavage site and the A region of the AAV2 ITR is highlighted in blue.

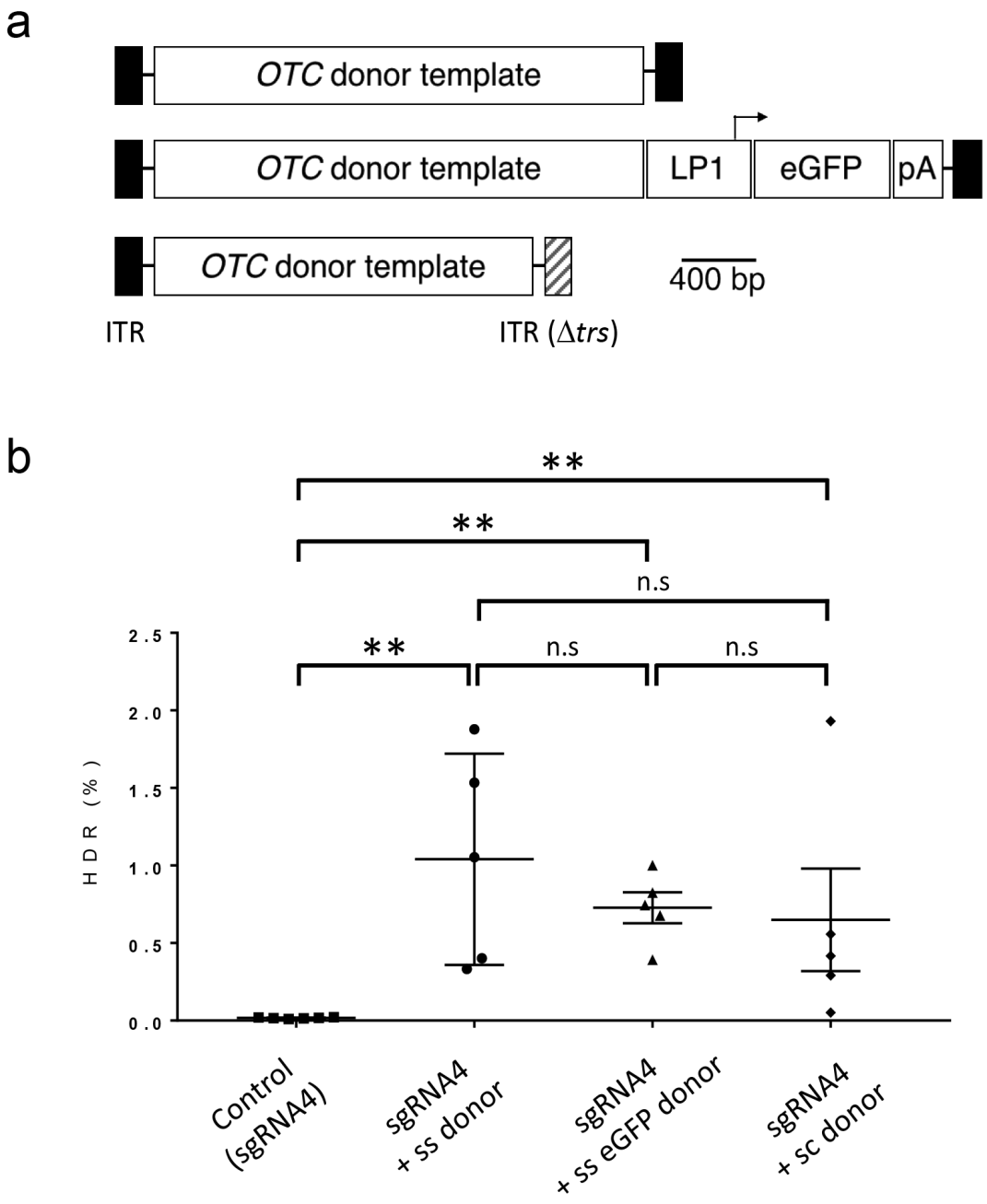


Fig. S4. Donor template vector configuration does not influence the rate of homology-directed repair. To determine the effect of donor template configuration on HDR rates in murine hepatocytes transposed with the *OTC* minigene, **(a)** each of three donor templates (5×10^{11} vg/mouse) were co-delivered with the saCas9/sgRNA4 vector (2×10^{11} vg/mouse). LP1, liver-specific promoter comprising an apolipoprotein E (ApoE) enhancer and short form of the human alpha 1-antitrypsin (hAAT) promoter;[4] eGFP, enhanced green fluorescent

protein, pA, bovine growth hormone polyadenylation signal sequence. Black boxes indicate ITR sequences with intact terminal resolution sites *(trs)* and the deleted 3′ *trs* in the scAAV vector shown by the striped box (ITR Δ*trs*).[5] **(b)** No significant difference in HDR rates were observed when animals receiving either the single-stranded (ss), single-stranded with eGFP expression cassette (ss eGFP) or self-complementary (sc) donor template vectors; $n = 6$ for sgRNA4 controls, $n = 5$ for mice receiving the single-stranded donor (with or without eGFP expression cassette) and $n = 4$ for self-complementary donor template vectors. Data are plotted as mean ± s.e.m. and significance evaluated using the Mann-Whitney non-parametric test (**<0.005).

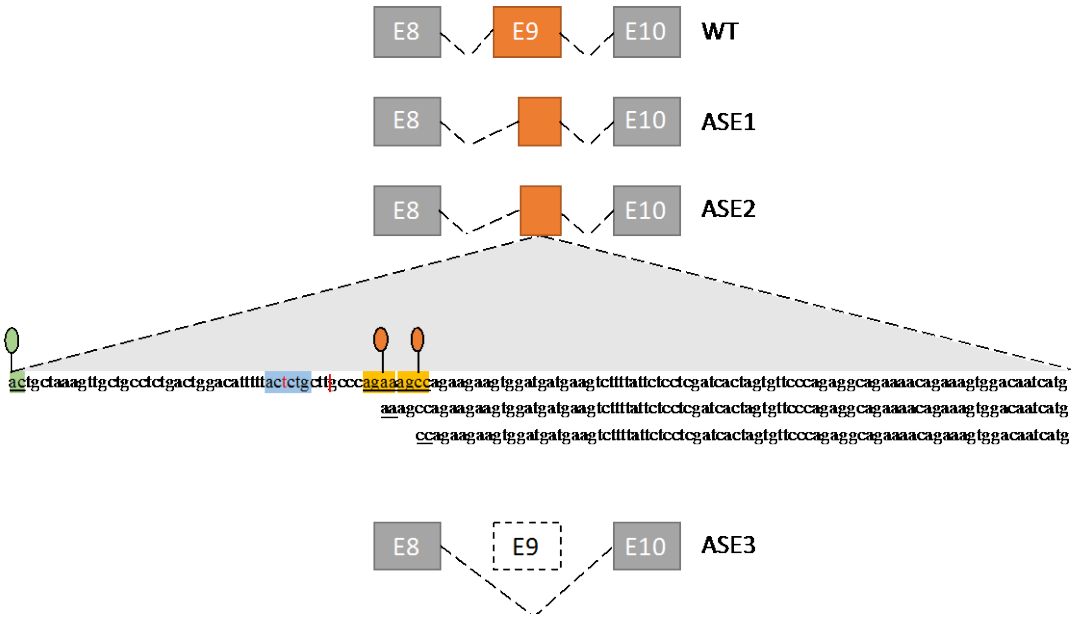


Fig. S5. Aberrant splicing of *OTC* exon 9 in patient-derived primary human hepatocytes. Illumina deep sequencing of amplicons produced from cDNA revealed alternative splicing events (ASE) in addition to constitutive splicing of exon 9. The two most common ASE are shown underneath the sequence for the OTC donor infant; ASE1 and ASE2 are missing the first 50 bp and 54 bp of exon 9, respectively. Two new acceptor sites were identified (underlined and shaded in yellow) at the 5' end of the alternatively spliced exon. These scored highly as new potential splice sites using Human Splicing Finder (http://www.umd.be/HSF3/) and their respective locations have been marked with an orange oval. The cutting location for the wild-type acceptor site (green oval) used for constitutive splicing at the start of exon 9 is also indicated (underlined and shaded in green). The sgRNA4 PAM, that contains the mutation of the donor infant used in this study (red nucleotide), is shaded in blue and the location of the SaCas9 cleavage site indicated by the red line. In addition to ASE1 and ASE2, a third population of aberrantly spliced message (ASE3), with skipping of exon 9, was also identified.

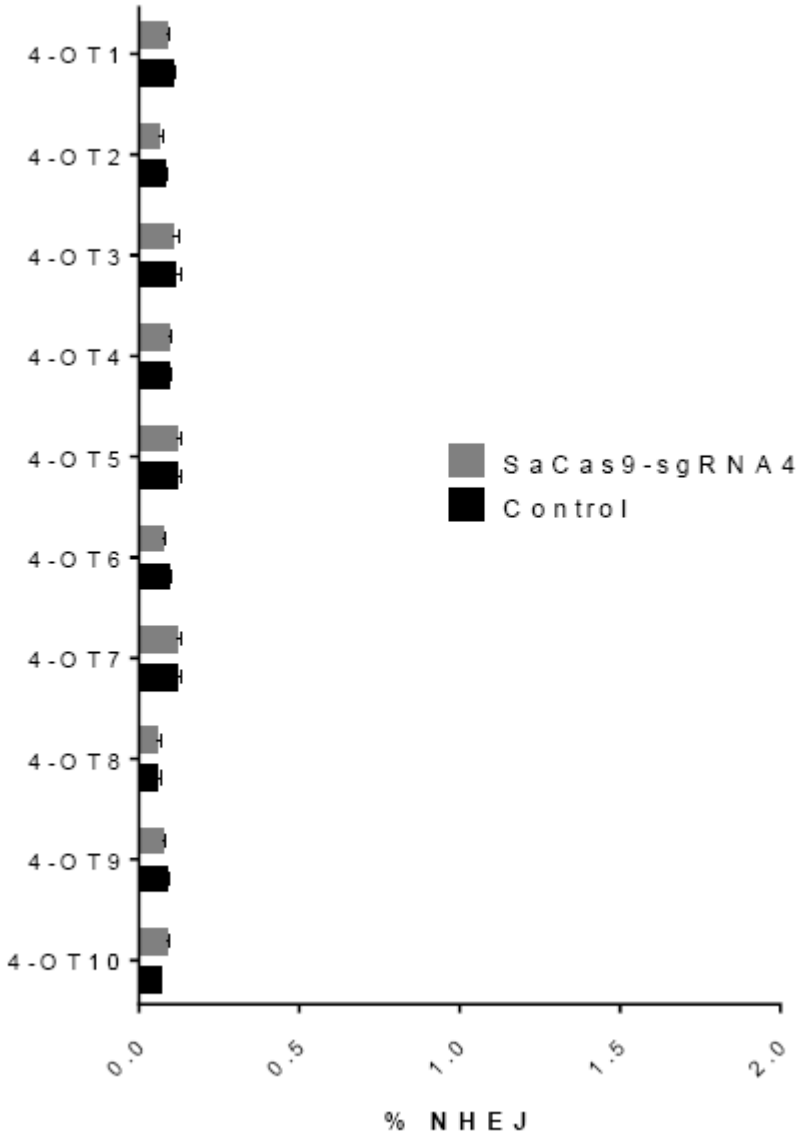
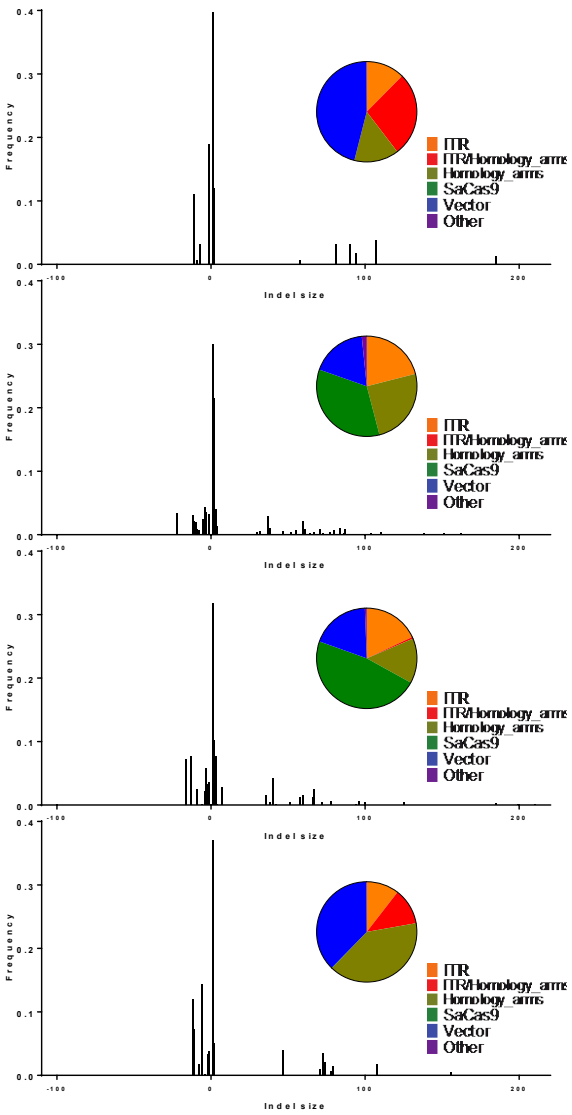


Fig. S6. Analysis of predicted off-target sites. InDels were quantified using CRISPRessoPooled[2] from Illumina® sequencing data of *in silico* predicted off-target loci identified using Benchling ($n = 3$ samples per treatment group). FRG mice engrafted with OTC-deficient patient cells were treated with sgRNA4 and PCR was performed across the predicted off-target loci. Control samples represent engrafted FRG mice that did not receive a vector containing SaCas9 and sgRNA.

a


Frequency
Indel size
ITR
ITR/Homology_arms
Homology_arms
SaCas9
Vector
Other

b

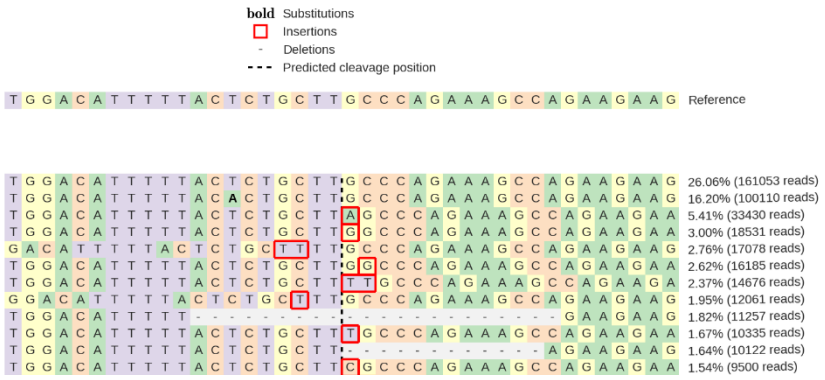

bold Substitutions
Insertions
Deletions
Predicted cleavage position
Reference
26.06% (161053 reads)
16.20% (100110 reads)
5.41% (33430 reads)
3.00% (18531 reads)
2.76% (17078 reads)
2.62% (16185 reads)
2.37% (14676 reads)
1.95% (12061 reads)
1.82% (11257 reads)
1.67% (10335 reads)
1.64% (10122 reads)
1.54% (9500 reads)

Fig. S7. Next generation sequence analysis of OTC-deficient patient-derived primary human hepatocytes following delivery of dual AAV editing reagents. FRG mice were engrafted with OTC-deficient patient-derived primary human hepatocytes and 11 weeks later, when the mice were on day 11 of a 21-day water cycle, received 2.5×10^{10} vg/mouse SaCas9-sgRNA4 and 1×10^{11} vg/mouse donor vectors packaged in the NP59 capsid via i.p. delivery (n = 4). Livers were analysed 5 weeks following vector delivery. **(a)** Analysis of insert sequences extracted from CRISPResso output revealed the presence of, in addition to the predicted InDels, large insertions (≥10 bp) that mapped to the introduced vector sequence at the cleavage site. **(b)** Representative CRISPResso output showing the most common reads for one mouse (FRG384) which received the dual editing vectors.

Table S1. PCR primers sequences used in this study

Primer name	**Sequence (5' to 3')**	**Purpose**
sgRNA1_top	CACCGACTTTAGCAGTCTAAAGAGAT	sgRNA1
sgRNA1_botom	AAACATCTCTTTAGACTGCTAAAGTC	
sgRNA2_top	CACCGTTGCCCAGAAAGCCAGAAGAA	sgRNA2
sgRNA2_botom	AAACTTCTTCTGGCTTTCTGGGCAAC	
sgRNA3_top	CACCGCTGGGAACACTAGTGATCGAG	sgRNA3
sgRNA3_botom	AAACCTCGATCACTAGTGTTCCCAGC	
sgRNA4_top	CACCGTCTTCTGGCTTTCTGGGCAAG	sgRNA4
sgRNA4-botom	AAACCTTGCCCAGAAAGCCAGAAGAC	
sgRNA5_top	CACCGATCATGGTAAGCAAGAAACA	sgRNA5
sgRNA5_botom	AAACTGTTTCTTGCTTACCATGATC	
Minigene_74R	TGGAGCTGAGGTGAGTAATCTGT	PCR to amplify the
Minigene_75F	ACCATGCTGTTTAATCTGAGGA	minigene[a]
hOTC intron8_F2	AAAATGGCACCTCATCTTTGTT	PCR to amplify the
hOTC intron9_R2	CCCCTGCTTATTTTATAGGGTCT	*OTC* locus FRG mice[a]
Nested_82F	TTGCATTGATGTCTGACTACTGG	Nested PCR for on-
Nested_83R	TCCCCAAAGGCATGAATGACCAA	target deep sequencing
Minigene_F1	TTGCACTTCTGGGAGGACAT	qPCR for minigene
Minigene_R1	TAGTGTTCCTGGAGCGTGAG	vectors
D571	CTACGAGGCCAGAGTGAAGG	qPCR for SaCas9
D572	GGCCACGTATTTCTCTTCCA	vectors
Repair_F1	TCCCAAATGTCACCTGCTCT	qPCR for donor vectors
Repair_R1	CCAAAAGATTCTGGGAGCCG	
D504	ACGGCAAATTCAACGGCAC	Murine *Gapdh*
D505	TAGTGGGGTCTCGCTCCTGG	
D278	CTAGACTTCGAGCAGGAGATGGC	β-actin
D279	TCTGCATCCTGTCAGCAATGCC	
Sur_F1	TCAGCCGAGATTTTGCCACT	TIDE analyses
Sur_R1	GGCGGATCAAGGGTGGTAAG	
Sur_F2	TGACTACTGGCCATGTGTGT	Surveyor analyses
Sur_R2	AGAGCAGGTGACATTTGGGA	

[a]PCR primer binding site are located outside the region of homology with the donor template

Table S2. Predicted off-target sites for sgRNA4 in human hepatocytes identified and scored by Benchling

Identification	Sequence	PAM	Score	Strand	Chromosome	Position	Mismatches
sgRNA4 (on-target)	TCTTCTGGCTTTCTGGGCAAG	CAGAG	100	-	X	38411802	0
sgRNA4_OT1	CATTCTGGCTCTCTGGGAAAG	CAGAG	1.4	+	7	20842747	4
sgRNA4_OT2	TTTTCTGGTTTTCTGGGCATG	AAGGG	1.4	-	15	64944504	3
sgRNA4_OT3	ACACATGGCATTCTGGGCAAG	AGGAG	0.9	+	5	158684961	5
sgRNA4_OT4	GCTGCAGGCTTTCTGCGCAAG	CAGGA	0.9	+	15	29270341	4
sgRNA4_OT5	CCTTTTGGTTTTCTGAGCAAG	CTGGA	0.9	-	11	62123112	4
sgRNA4_OT6	CCTTTTGGTTTTCTGAGCAAG	CTGAA	0.9	-	12	95514726	4
sgRNA4_OT7	GCCCCTGGCATGCTGGGCAAG	GAGAG	0.5	-	10	110831202	5
sgRNA4_OT8	TCCTCTAGATTTCTGGGTAAG	AGGGA	0.4	+	1	150733060	4
sgRNA4_OT9	GCTTTTGGCAGTTTGGGCAAG	TGGAG	0.4	+	12	120095728	5
sgRNA4_OT10	CCGTTTGGCTTTCTGGGGATG	CAGAG	0.3	+	19	17542253	5

Table S3. PCR primers sequences used for analysis of predicted off-target sites for sgRNA4 in patient-derived human hepatocytes by deep sequencing

Primer name	**Sequence (5' to 3')**	**Off-target site**
sgRNA4_OT1F	GGCCTGATGCTGTGAGAAAC	sgRNA4_OT1
sgRNA4_OT1R	GCTTCAGCACAATTCAGCAAGGT	
sgRNA4_OT2F	ACCCCTATTTCACAGGAGATGTTG	sgRNA4_OT2
sgRNA4_OT2R	ACAGGAGGGGTACACAAGGG	
sgRNA4_OT3F	TCAAGCCTTACGAGGTCTGATG	sgRNA4_OT3
sgRNA4_OT3R	TAGAGAGTAAAACTACCAGCTTCC	
sgRNA4_OT4F	GGGCTCTTAGTTTTCGTCCCTCCA	sgRNA4_OT4
sgRNA4_OT4R	CCAAAAGATAAACCCATGTGACCAA	
sgRNA4_OT5F	GACTTAGTAAAAACATTGTGCCAAC	sgRNA4_OT5
sgRNA4_OT5R	CAAAGGCCAGGGTGTTACGG	
sgRNA4_OT6F	TAGGAACATTGTGCCAACTCAAA	sgRNA4_OT6
sgRNA4_OT6R	AATCCAAACCTGTCAAAGGGGAAGG	
sgRNA4_OT7F	GCCATCCTAACCCTGCGTGTCT	sgRNA4_OT7
sgRNA4_OT7R	AAGCAAGGCCAAGGACGCCA	
sgRNA4_OT8F	TTCTGGATACAGCAGGAAAAA	sgRNA4_OT8
sgRNA4_OT8R	CACTTATAGGAGAAAAACAATGAGA	
sgRNA4_OT9F	CCCCCGACCAGATGAGGCTG	sgRNA4_OT9
sgRNA4_OT9R	CCAGGGAGAGCCCTGATTGGA	
sgRNA4_OT10F	GGGGAGATGCCGTGGGACTT	sgRNA4_OT10
sgRNA4_OT10R	ATGAGACTGGCAGCCCCGC	

REFERENCES

1. Brinkman, EK, Chen, T, Amendola, M and van Steensel, B (2014). Easy quantitative assessment of genome editing by sequence trace decomposition *Nucleic Acids Res* **42:** e168
2. Pinello, L, Canver, MC, Hoban, MD, Orkin, SH, Kohn, DB, Bauer, DE*, et al.* (2016). Analyzing CRISPR genome-editing experiments with CRISPResso *Nat Biotechnol* **34:** 695-697
3. Camacho, C, Coulouris, G, Avagyan, V, Ma, N, Papadopoulos, J, Bealer, K*, et al.* (2009). BLAST+: architecture and applications *BMC Bioinformatics* **10:** 421
4. Nathwani, AC, Gray, JT, Ng, CY, Zhou, J, Spence, Y, Waddington, SN*, et al.* (2006). Self-complementary adeno-associated virus vectors containing a novel liver-specific human factor IX expression cassette enable highly efficient transduction of murine and nonhuman primate liver *Blood* **107:** 2653-2661
5. McCarty, DM, Fu, H, Monahan, PE, Toulson, CE, Naik, P and Samulski, RJ (2003). Adeno-associated virus terminal repeat (TR) mutant generates self-complementary vectors to overcome the rate-limiting step to transduction in vivo *Gene Therapy* **10:** 2112-2118