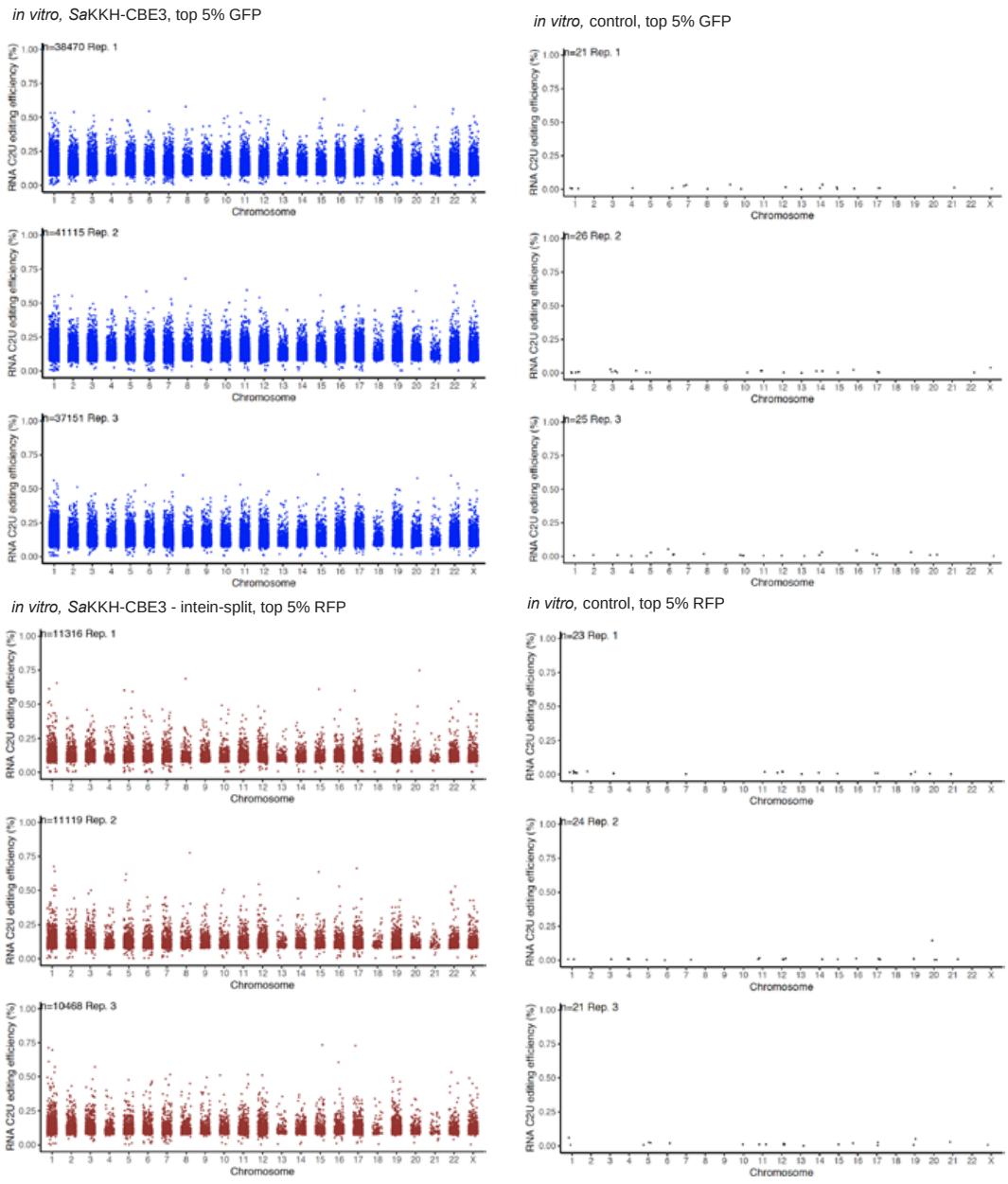
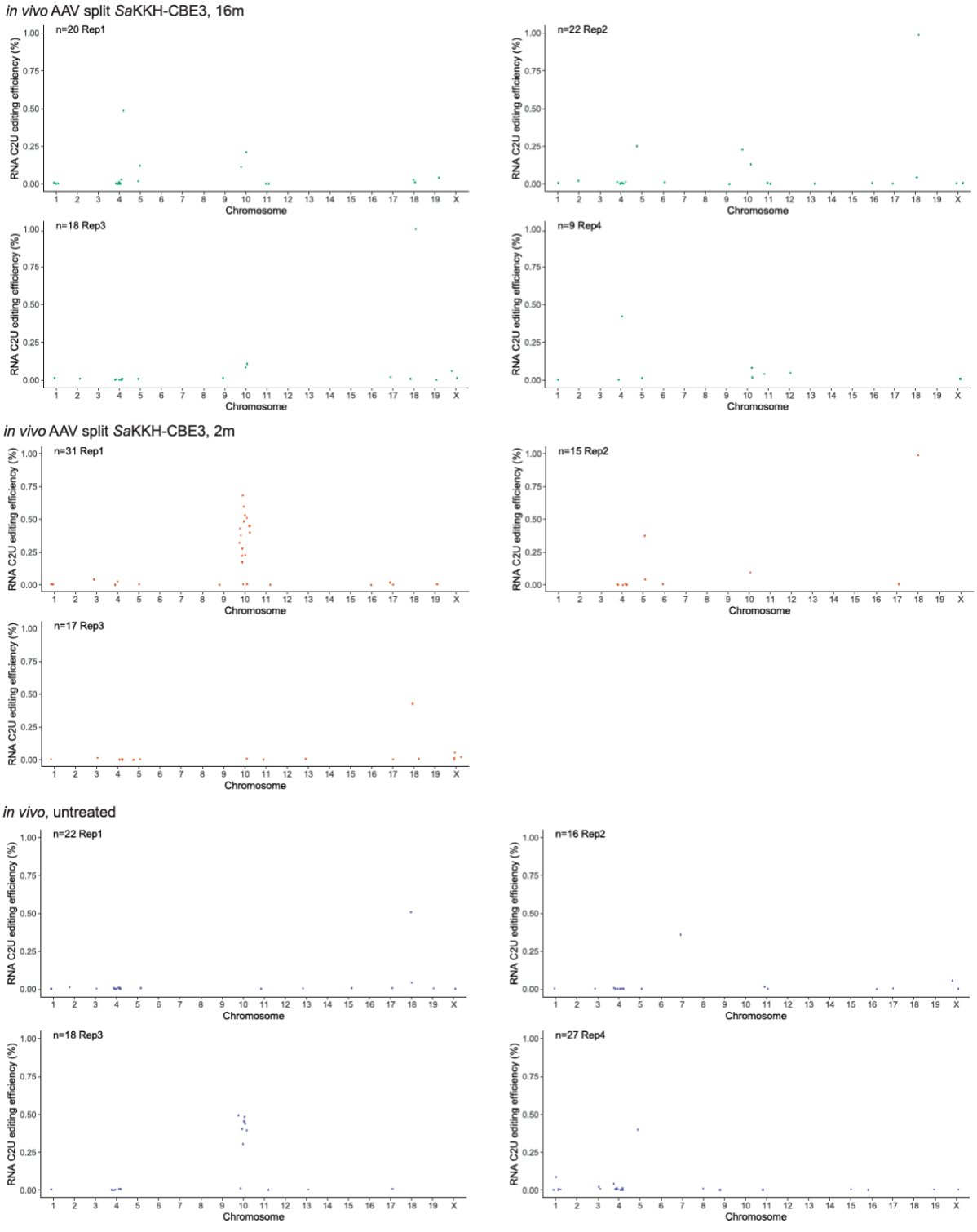
Nat Biomed Eng. 2021 Feb;5(2):179-189. doi: 10.1038/s41551-020-00671-z.

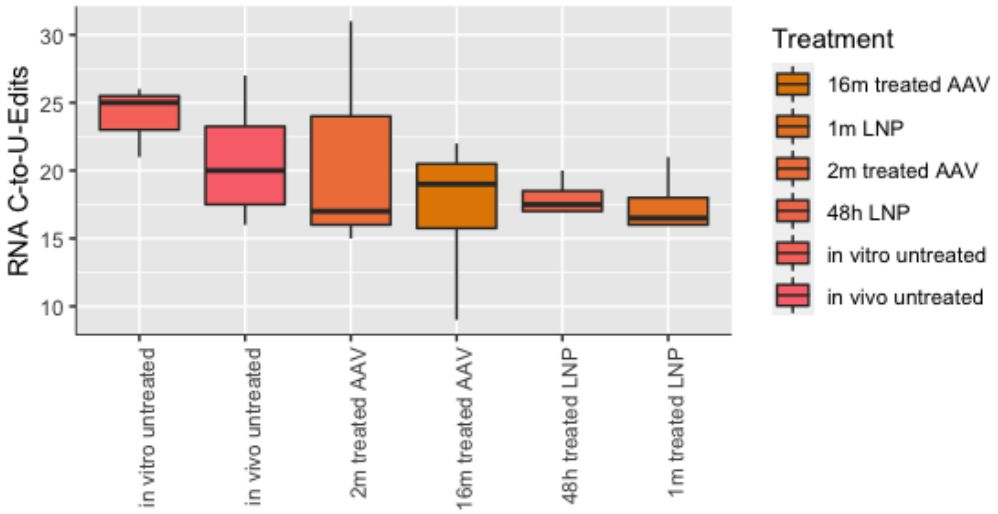
Supplementary Information: In vivo cytidine base editing of hepatocytes without detectable off-target mutations in RNA and DNA



Supplementary Figure 1 | Manhattan plots depicting transcriptome-wide C-to-U off-target editing events of base edited *in vitro* samples from HEK293T cells. *In vitro* samples were derived from top 5% GFP- or RFP-sorted HEK293T cells transfected with the indicated BE constructs, *Pah*^{enu2}-targeting sgRNA, and either GFP- or RFP-expressing plasmids. Control samples were transfected with GFP- or RFP-expressing plasmids only. Each sample is plotted individually. Each dot represents a C-to-U editing event. Counts per sample and replicates are indicated in the top left corners of each plot.

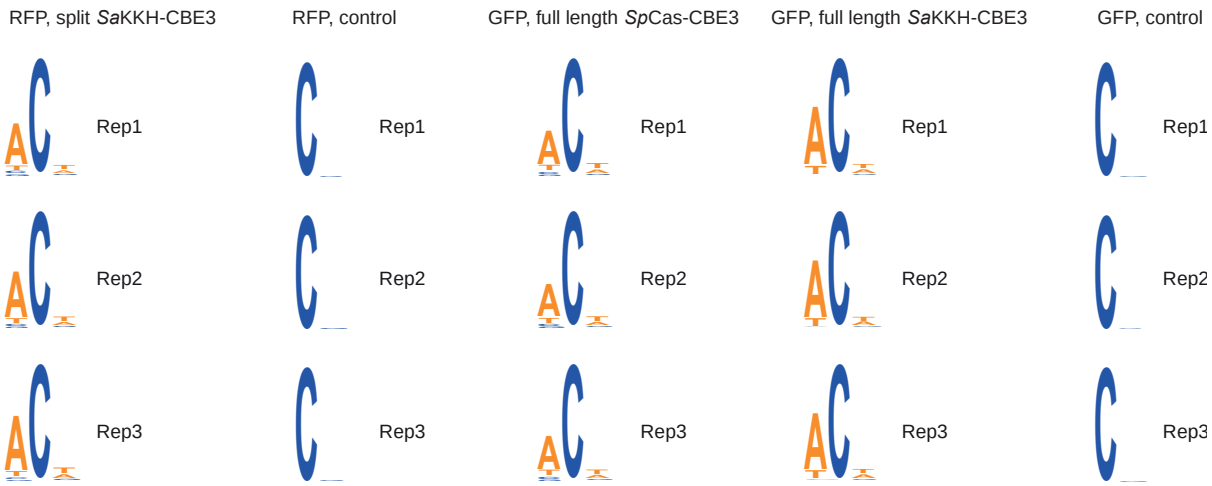


Supplementary Figure 2 | Manhattan plots depicting transcriptome-wide C-to-U off-target editing events of base edited *in vivo* samples from mouse livers. *In vivo* samples were derived from mouse livers treated with 5 x 10^{11} vg per AAV vector to express split *Sa*KKH-CBE3 and the *Pah*^{enu2}-targeting sgRNA. Controls were untreated mouse livers. Each dot represents a C-to-U editing event. Counts per sample and replicates are indicated in the top left corners of each plot.

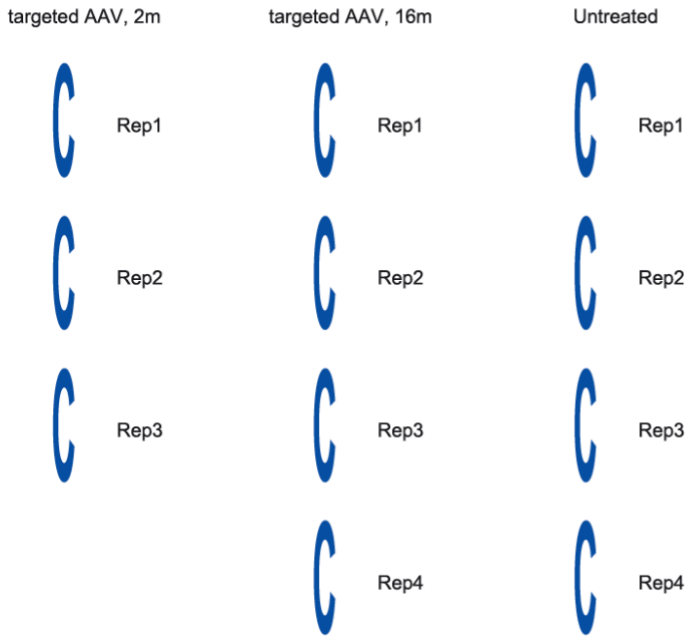


Supplementary Figure 3 | C-to-U conversions in untreated, AAV-treated, and LNP-treated samples. Box plot summarizing RNA C-to-U editing events of Fig 1b and 4b. Box plots are standard Tukey plots, where the centre line represents the median, the lower and upper hinges represent the first and third quartiles, and whiskers represent 1.5 times the interquartile range. n=3 biologically independent replicates for the '2 month treated AAV' group and for the 'in vitro untreated' group, n=4 biologically independent replicates for all other groups.

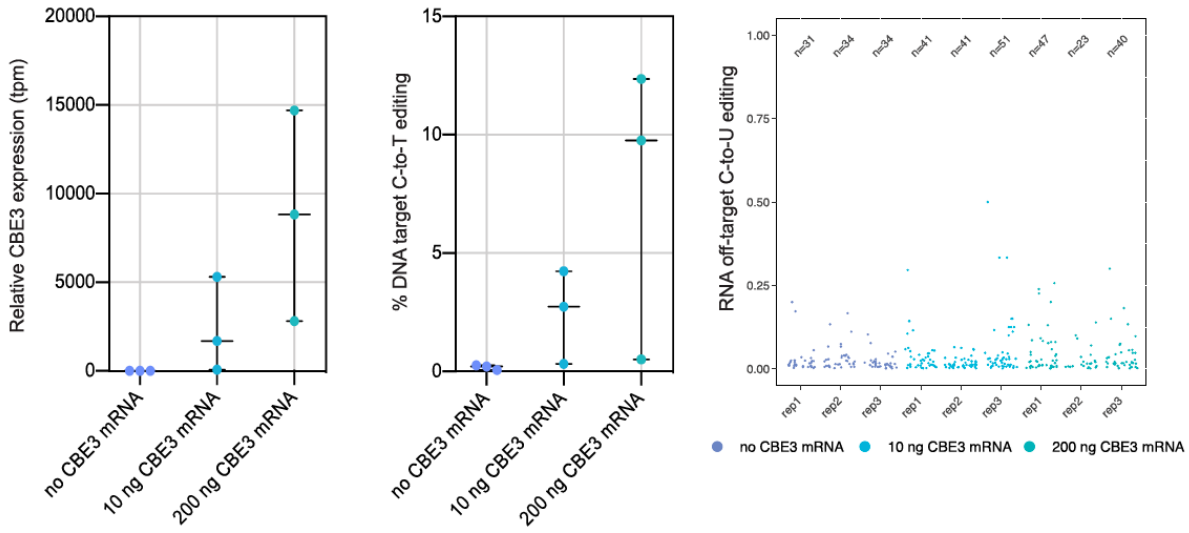
in vitro (HEK293T)



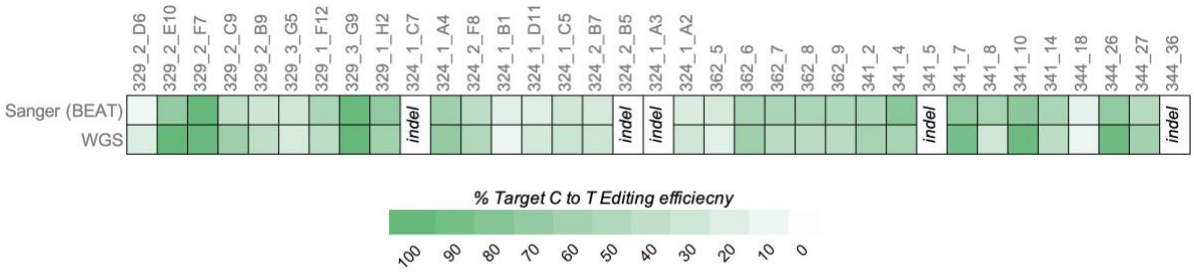
in vivo (mouse liver)



Supplementary Figure 4 | Sequence motifs of edited cytosines from RNA-seq data. Ts should be considered as Us. The typical APOBEC signature is ACW (W=A or U)[1]. Each logo consists of stacks of individual base-likelihoods at each position in the sequence. The overall height of the stack indicates the sequence conservation at that position, while the height of symbols within the stack indicates the relative frequency of each nucleic acid at that position.

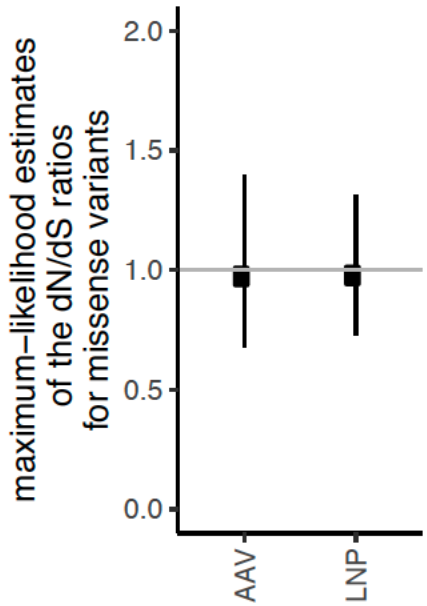


Supplementary Figure 5 | Titration of *Sa*KKH-CBE3 expression in mouse Hepa1-6 cells. *Sa*KKH-CBE3 expression was titrated by transfecting different *Sa*KKH-CBE3 mRNA and sgRNA concentrations into Hepa1-6 cells with exon 7 of *Pah*^{enu2} stably integrated. The first panel depicts relative expression of CBE3 in transcripts per million, the second panel shows corresponding C-to-T edits at the target locus. Values represent mean ± s.d. of 3 biologically independent replicates. Panel three shows RNA C-to-U editing events of the samples analysed in panel one and two. Each dot represents one editing event. Total counts are indicated above (n). Cells were harvested 48 hours after transfection and sorted for co-transfected mCherry expression.

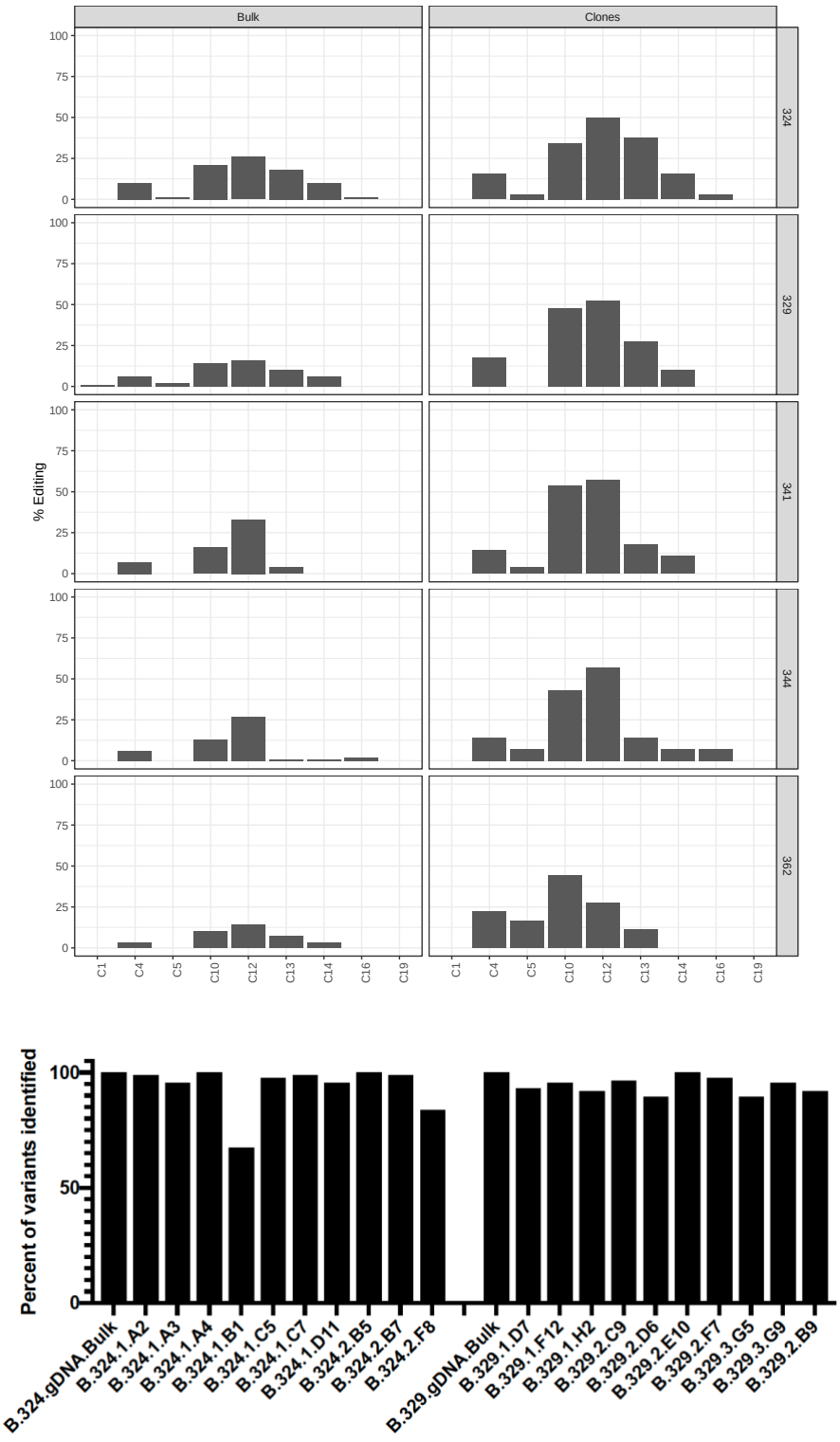


Treatment	Mouse No.	sample ID	Allele 1	Allele 2	Allele 3	Allele 4	Ploidy (BAF Plot)
-	28	28_1	cttccgagtctCccactgca	cttccgagtctCccactgca			2n
-	28	28_2	cttccgagtctCccactgca	cttccgagtctCccactgca	cttccgagtctCccactgca	cttccgagtctCccactgca	4n
-	28	28_3	cttccgagtctCccactgca	cttccgagtctCccactgca	cttccgagtctCccactgca	cttccgagtctCccactgca	4n
LNP	329	2_D6	cttacgagttttCccactgca	cttccgagtctCccactgca	cttccgagtctCccactgca	cttccgagtctCccactgca	4n
LNP	329	2_E10	cttccgagtttTtcactgca	cttccgagtttTtcactgca	cttccgagtttTtcactgca	ctttcgagtttTtcactgca	4n
LNP	329	2_F7	cttccgagtctTccactgca	cttccgagtctTccactgca	cttccgagtgtTccactgca	cttccgagtgtTccactgca	4n
LNP	329	2_C9	cttccgagttttTccactgca	cttccgagttttTccactgca	cttccgagttttTccactgca	cttccgagtctCccactgca	4n
LNP	329	2_B9	ctttcgagtttTttactgca	cttccgagtctCccactgca	cttccgagtctCccactgca	cttccgagtctCccactgca	4n
LNP	329	3_G5	ctttcgagtttTttactgca	cttccgagtctCccactgca	cttccgagtctCccactgca	cttccgagtctCccactgca	4n
LNP	329	1_F12	ctttcgagtttTtcactgca	cttccgagtttTtcactgca	cttccgagtctCccactgca	cttccgagtctCccactgca	4n
LNP	329	3_G9	ctttcgagtttTttactgca	cttccgagtctCccactgca			2n
LNP	329	1_H2	cttccgagtttTtcactgca	cttccgagtttTccactgca	cttccgagtttTccactgca	cttccgagtctCccactgca	4n
LNP	324	1_C7	cttccgagtttTtcactgca	cttccgagttCCCACtTtcactgca	cttccgagtctTccactgca	cttccgagtctCccactgca	4n
LNP	324	1_A4	cttccgagtttTtcactgca	cttccgagtctTccactgca	cttccgagtctTccactgca	cttccgagtctCccactgca	4n
LNP	324	2_F8	cttccgagtct--cctgca	cttccgagtttTtcactgca	cttttgagtttTtcactgca	cttccgagtctCccactgca	4n
LNP	324	1_B1	cttccgagtttTttactgca	cttccgagtctCccactgca	cttccgagtctCccactgca	cttccgagtctCccactgca	4n
LNP	324	1_D11	cttccgagtttTtcactgca	cttccgagtctCccactgca	cttccgagtctCccactgca	cttccgagtctCccactgca	4n
LNP	324	1_C5	ctttcgagtttTttactgca	cttccgagtctCccactgca	cttccgagtctCccactgca	cttccgagtctCccactgca	4n
LNP	324	2_B7	ctttcgagtttTttactgca	cttccgagtctCccactgca	cttccgagtctCccactgca	cttccgagtctCccactgca	4n
LNP	324	2_B5	cttccgagt------ctcca	cttccgagtct--cctgca	ctttcgagtttTttattgca	cttccgagtctCccactgca	4n
LNP	324	1_A3	ctttcgagtctCccactgca	ctttcgagtctCccactgca	gtctcgagattT---ctgca	ttactganctc-Tcganctgca	4n
LNP	324	1_A2	ctttcgagtttTtcactgca	cttccgagtttTccactgca	cttccgagtctGccactgca	cttccgagtctCccactgca	4n
LNP	362	362_5	ctttcgagtttCccactgca	cttccgagtctCccactgca	cttccgagtctCccactgca	cttccgagtctCccactgca	4n
LNP	362	362_6	cttttgagtttCccactgca	ctt-tgagtttCccactgca	cttccgagtctCccactgca	cttccgagtctCccactgca	4n
LNP	362	362_7	cttccgagtttTtcactgca	cttccgagtttTccactgca	cttccgagtctCccactgca	cttccgagtctCccactgca	4n
LNP	362	362_8	cttttgagtttTtcactgca	cttccgagtttTccactgca	cttccgagtctCccactgca	cttccgagtctCccactgca	4n
LNP	362	362_9	cttccgagtttTccactgca	cttccgagtctCccactgca			2n
AAV	341	341_2	cttccgagtttTccactgca	cttccgagtttCccactgca	cttccgagtttAccactgca	cttccgagtctCccactgca	4n
AAV	341	341_4	ctttcgagtctCccactgca	cttccgagtttTccactgca	cttccgagtttTtcactgca	cttacgagtttTccactgca	4n
AAV	341	341_5	cttc-------Ccactgca	cttccgagt--Ccactgca	cttccgagtctCccactgca	cttccgagtctCccactgca	4n
AAV	341	341_7	cttccgagtct---cctgca	cttccgagtttTtcactgca	ctttcgagtttTccactgca	cttccgagttttTtaactgca	4n
AAV	341	341_8	cttccgagtctTccactgca	cttccgagtctCccactgca	cttccgagtctCccactgca	cttccgagtctCccactgca	4n
AAV	341	341_10	cttccgag------actgca	ctttcgagtttTccactgca	cttccga-tttTccactgca	cttccgagtttTccactgca	4n
AAV	341	341_14	cttccgagtctTccactgca	cttccgagtctTccactgca	cttccgagtctCccactgca	cttccgagtctCccactgca	4n
AAV	344	344_18	cttccgagtttAccactgca	ctt-tgagtttCccactgca	cttccgagtctCccactgca	cttccgagtctCccactgca	4n
AAV	344	344_26	cttccgagtctTccactgca	cttccgagtctTccactgca	cttccgagtccCcc---gca	cttccga---tTccactgca	4n
AAV	344	344_27	cttccgagtctTccactgca	cttccgagtctCccactgca			2n
AAV	344	344_36	cttccgagtctTccactgca	cttccgagtctAccactgca	cttccga-tatTccactgca	cttccgagtctCccactgca	4n

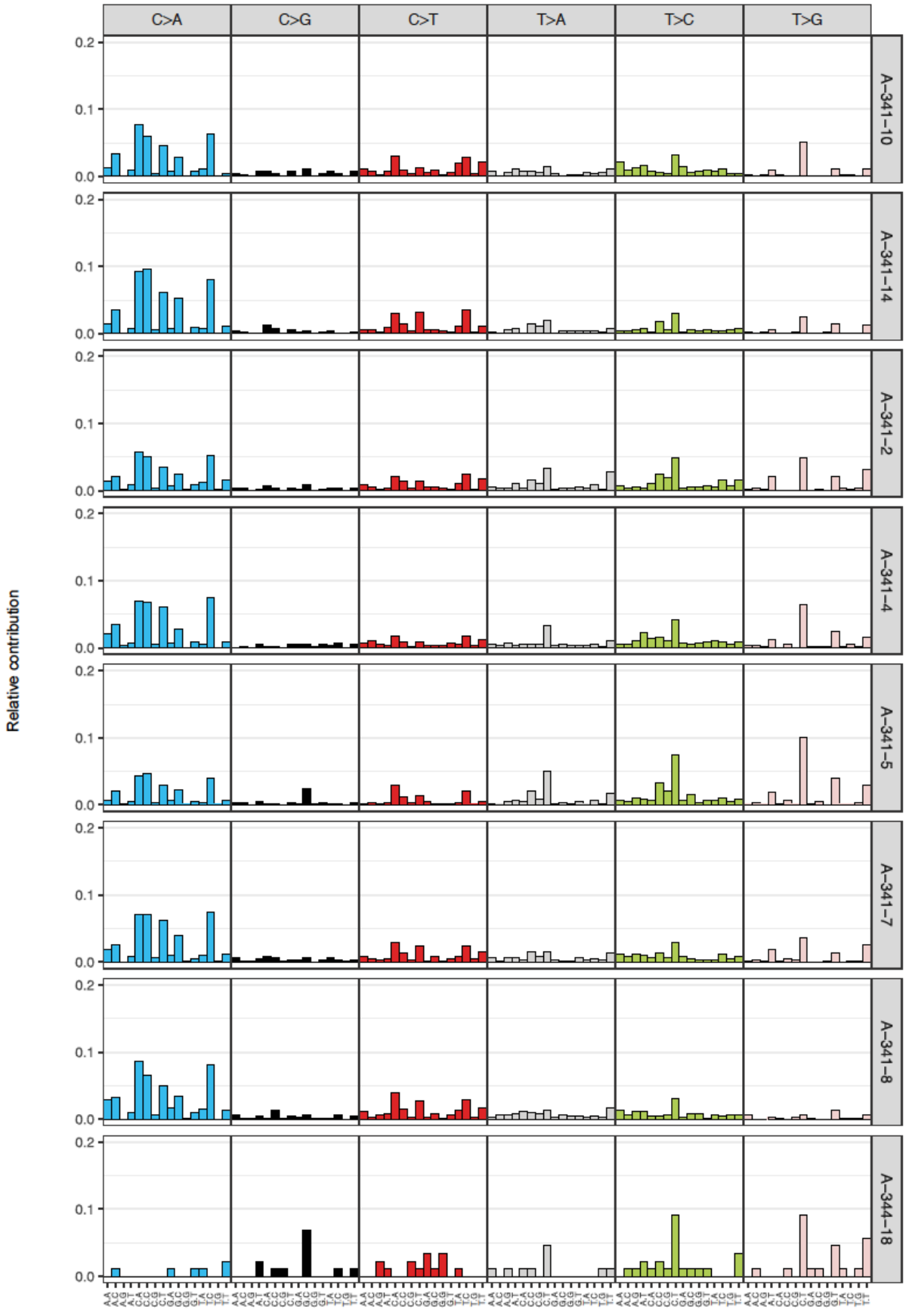
Supplementary Figure 6 | On-target editing of hepatocyte clones selected for WGS off-target analysis. To ensure that hepatocyte genomes were exposed to the base editor, only clones with on-target C-to-T editing were selected for WGS off-target analysis. Upper panel: Correlation of target C-to-T editing efficiency between Sanger sequencing and WGS analysis. Lower panel: Editing within the protospacer region of WGS-analysed clones. The majority of hepatocytes was tetraploid (89%) as previously described[2]. Ploidy was further confirmed by bi-allelic frequency plot analysis**.**

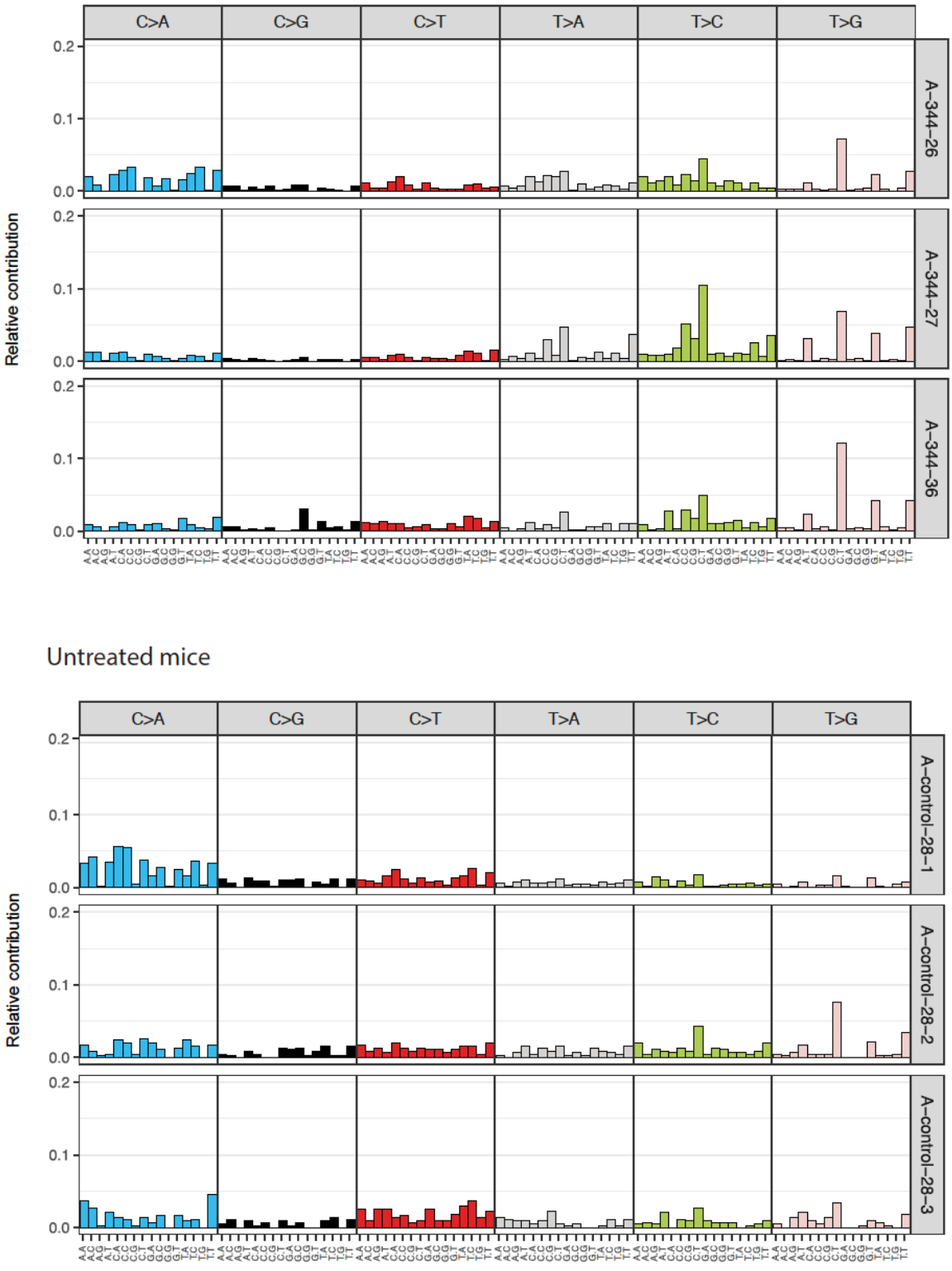


Supplementary Figure 7 | Evaluation of the d*N*/d*S* ratio in expanded hepatocytes. The d*N*/d*S* ratio for all protein-coding somatic point mutations observed in all AAV- and LNP-treated clones. dN/dS ratio=1 indicates no selection bias for non-synonymous or synonymous mutations. Values represent mean ± s.d. of n=11 (AAV) or n=24 (LNP) biologically independent replicates.

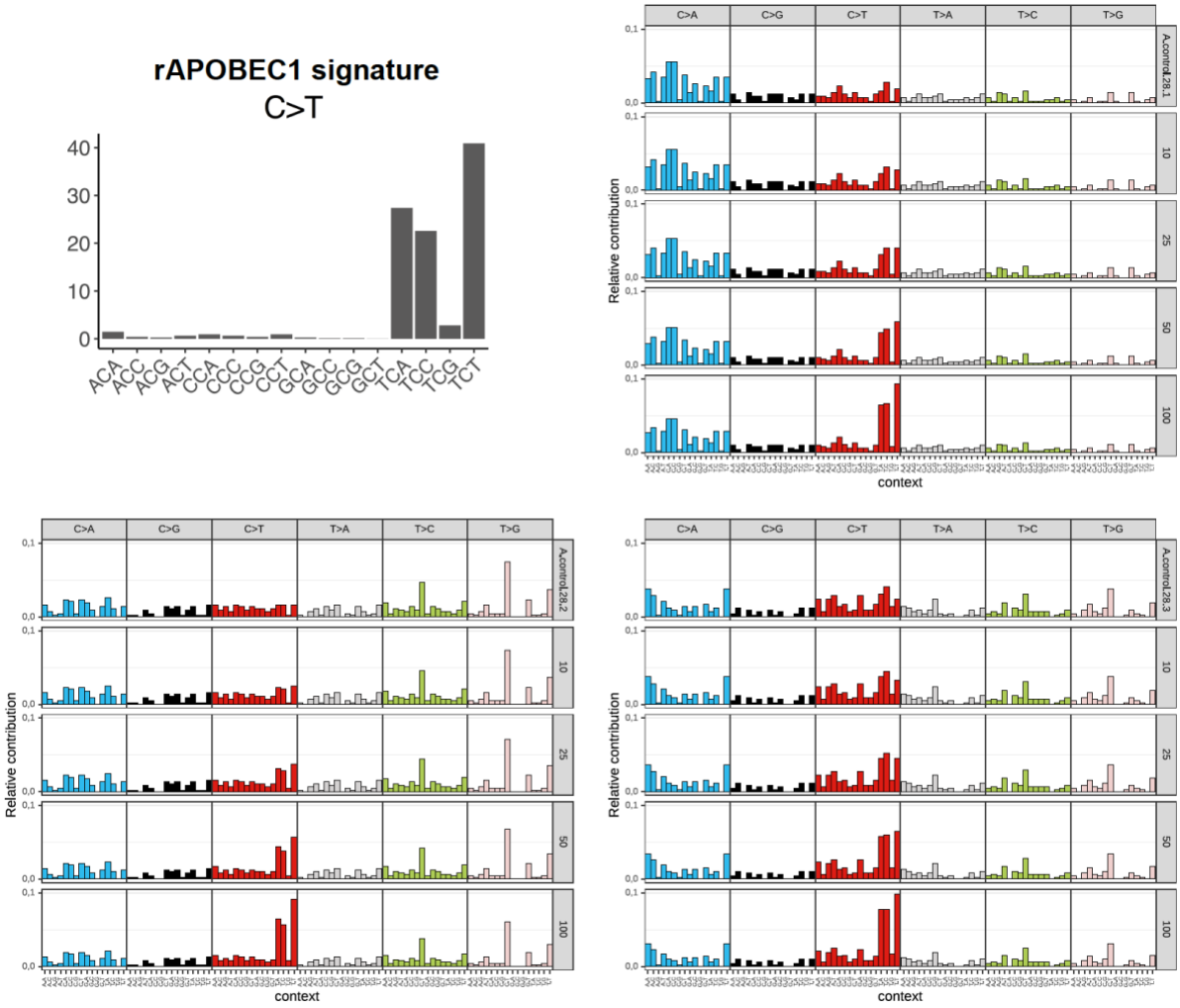


Supplementary Figure 8 | Sequence analysis of the target locus and of germline variants in bulk DNA and in clonal hepatocyte DNA. Upper panel: C-to-T editing efficiencies of all Cs in the protospacer assessed by sequencing of bulk liver DNA (left), and by sequencing of CLiP clones (values from all clones per mouse were cumulated) (right). For mouse 324 we analysed 10 clones, for mouse 329 we analysed 9 clones, for mouse 341 we analysed 7 clones, for mouse 344 we analysed 4 clones, and for mouse 362 we analysed 5 clones. Lower panel: Clones from LNP treated mice were analysed to assess if heterozygous germline variants can be detected. On average, 94% of the germline variants found in bulk DNA could be re-identified in clones (lower panel).

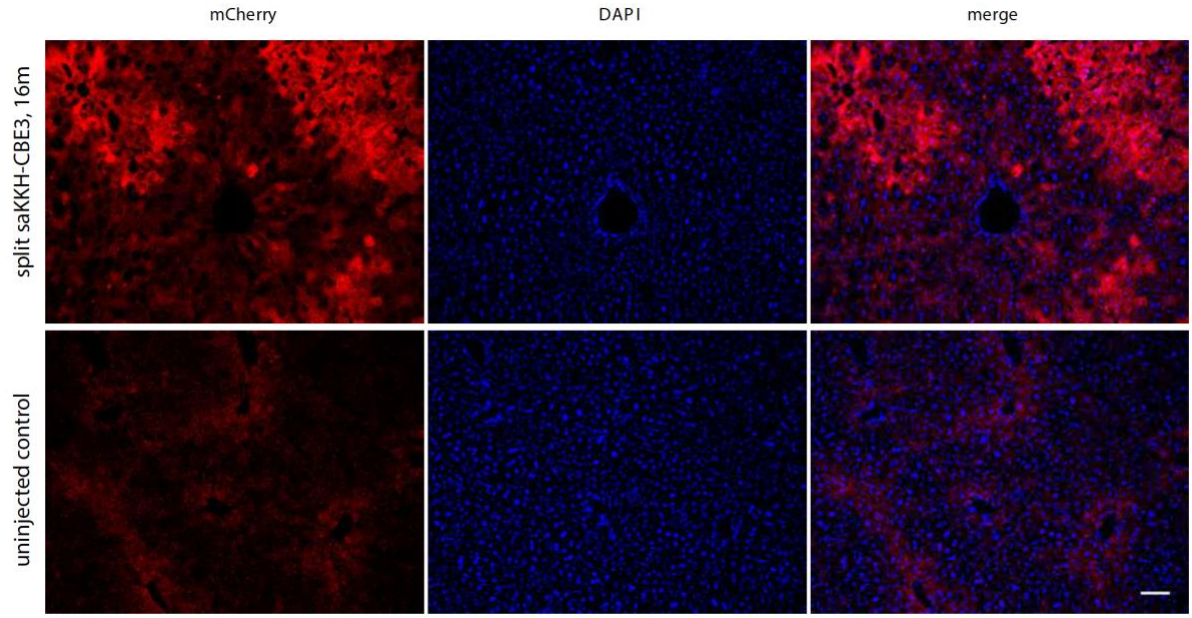

C>A
C>G
C>T
T>A
T>C
T>G
A–341–10
A–341–14
A–341–2
A–341–4
A–341–5
A–341–7
A–341–8
A–344–18
Relative contribution
0.0
0.1
0.2



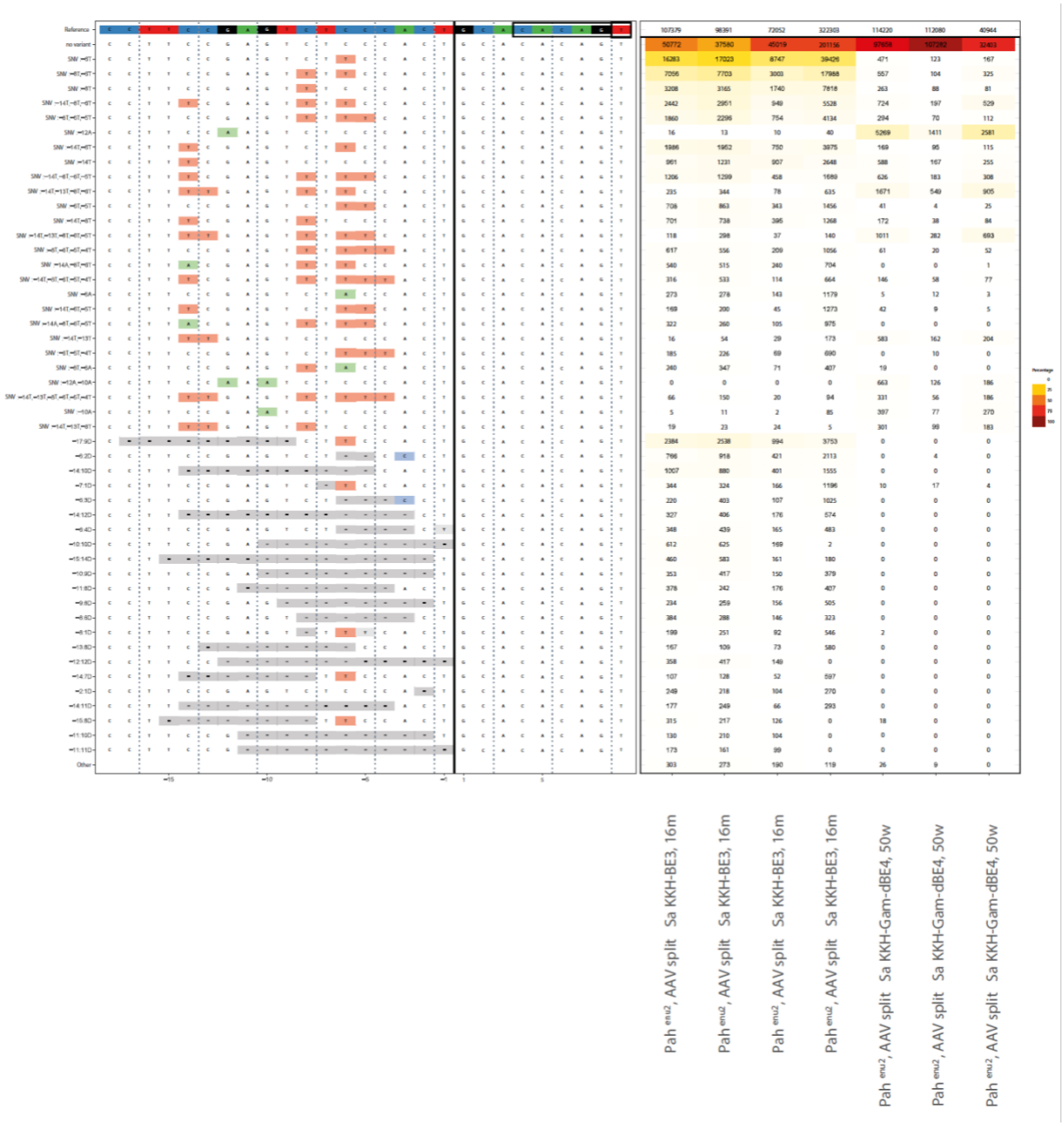
Supplementary Figure 9 | 96-nt profile plot of edited clones following AAV delivery. Primary hepatocytes were isolated and clonally expanded as chemically induced liver progenitor cells (CLiPs). On-target editing and therefore base editor exposure in clones was confirmed by Sanger sequencing before WGS at an average 30x coverage. The frequency (y-axis) for 96 mutational types (x-axis) is shown. Control clones were derived from untreated mice. Plots show motif contributions of individual clones presented in main figure 2c,d.



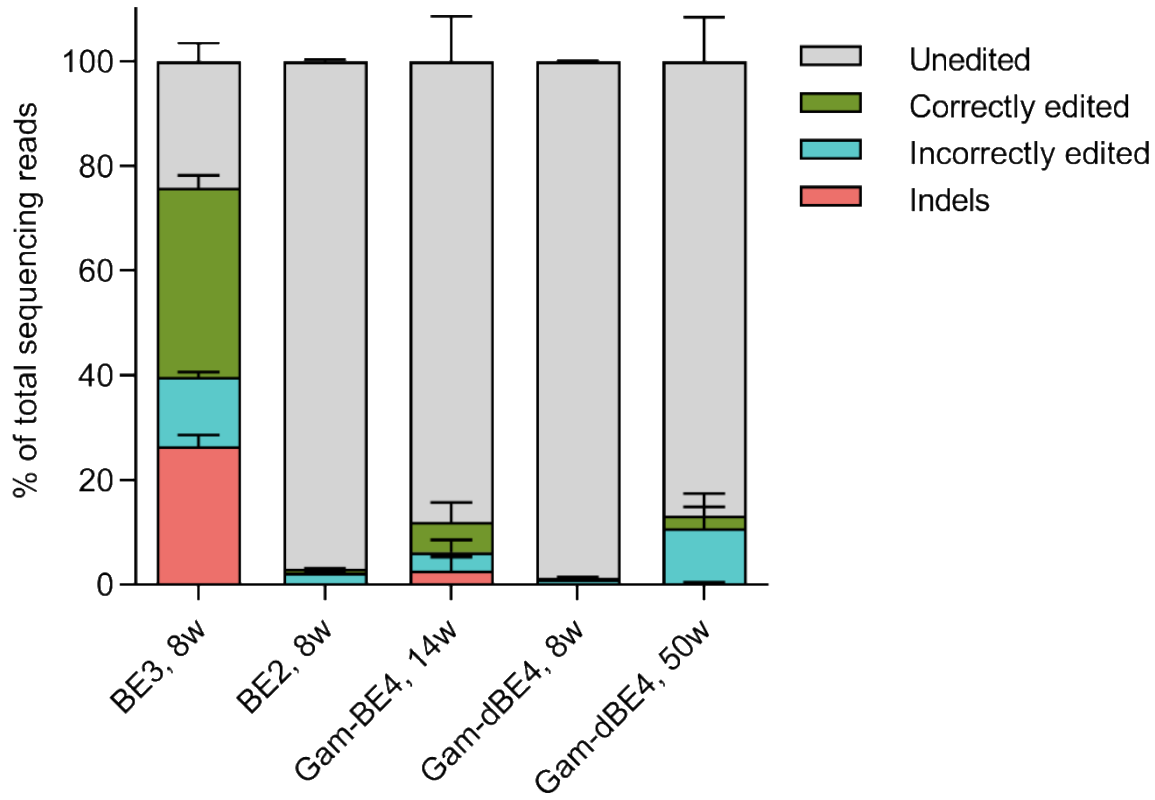
Supplementary Figure 10 | 96-nt profile plot of *in silico* added APOBEC signature. The upper left panel shows the proportion of the different tri-nucleotide motifs generated by CBE off-target editing in mouse zygotes. Data was obtained from Table S7 in Song et al.[3]. Upper right and lower panels: 96-nt profile plot of a control sample with *in silico* added rAPOBEC1 signature. The frequency (y-axis) for 96 mutational types (x-axis) is shown. The numbers on the right indicate the number of SNVs added to the original pattern of control clones.



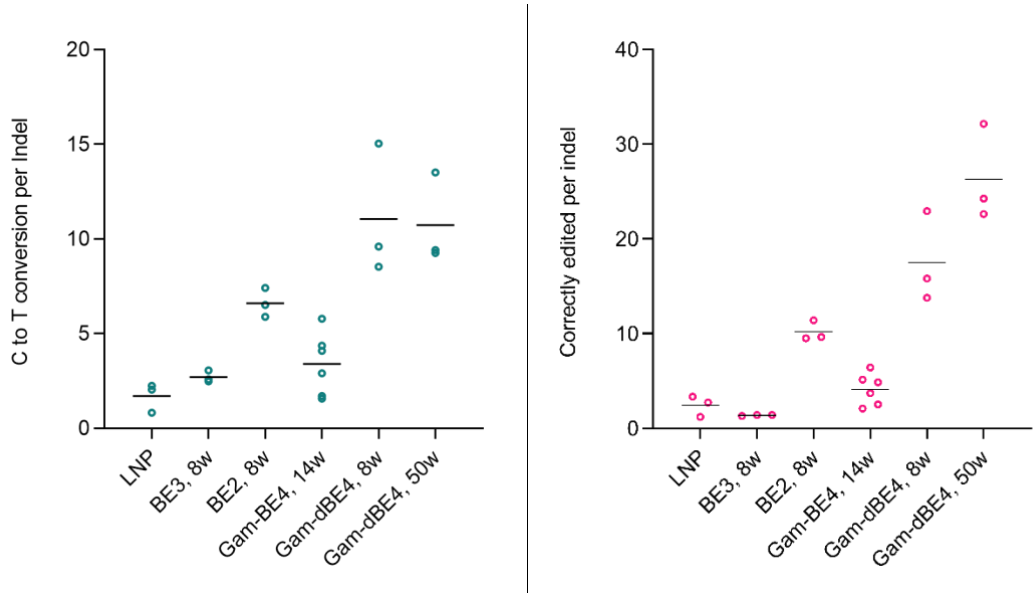
Supplementary Figure 11 | Cryosections of long-term AAV-treated mouse livers. Representative cryosections of 8 pictures taken per animals from 2 independent mice 16 months after AAV split *Sa*KKH-CBE3 (5×10^{11} vg per AAV) administration. Polycistronic tagRFP expression from one AAV. Controls are sections from an uninjected *Pah*^{enu2} mouse. Red channel, RFP. Blue channel, DAPI. Scale bar, 100 µm.



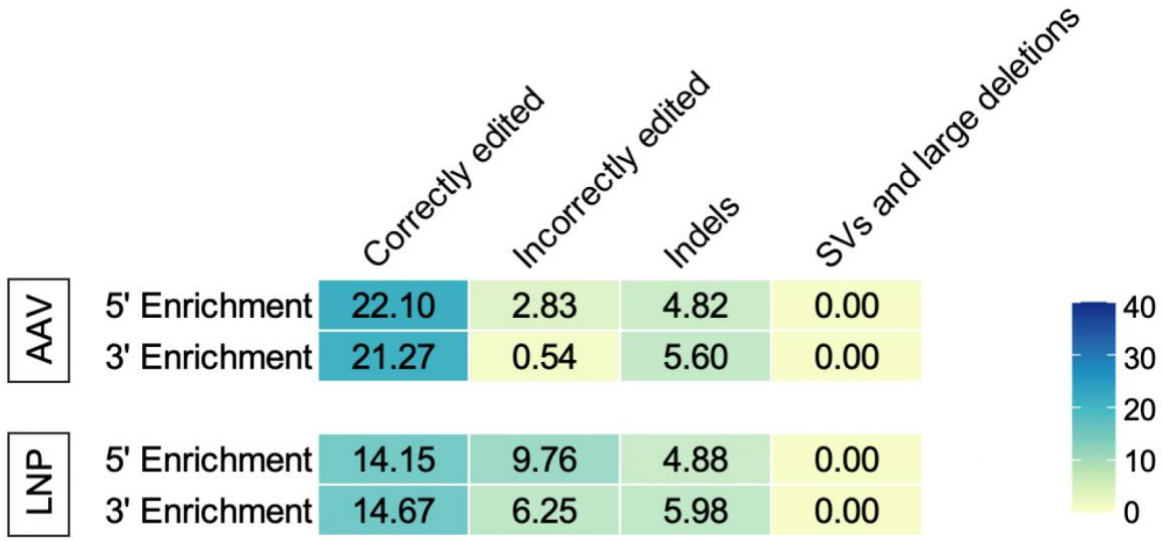
Supplementary Figure 12 | Allele plots of HTS data after AAV-mediated BE expression for 16 months *in vivo*. The left panel shows consensus sequences for the top 50 variant alleles according to the sum of their proportional contributions to each sample. Variants are numbered according to their leftmost position with respect to the cut site. Variants that did not contain insertions or deletions within the guide region are separated and labelled according to any single nucleotide variants (SNVs) present. The right panel shows variant counts, with the header showing the total number of reads per sample. Data are from *in vivo* experiments after expression of CBEs for 16 months and 50 weeks. All biological replicates are plotted individually.



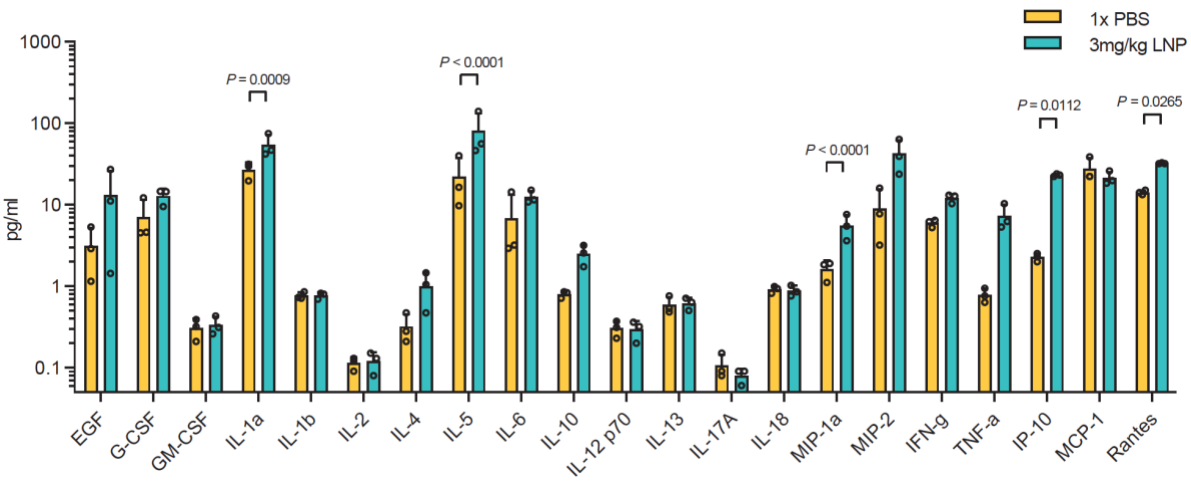
Supplementary Figure 13 | HTS data from *in vivo* experiments following AAV-mediated delivery of intein split base editor constructs. *Sa*KKH-BE2 and Gam- *Sa*KKH-dBE4 contain nuclease-dead Cas9 and have no nickase activity. Editing efficiencies are analysed after 8, 14, and 50 weeks. Correctly edited reads restore the wildtype amino acid sequence of the PAH enzyme. Incorrectly edited reads include nonsynonymous mutations, non-C-to-T conversions and C- to-T conversions at positions other than the target base. Values represent mean ± s.d. of n=3 mice per treatment or n=6 mice for Gam-BE4 treatment for 14 weeks



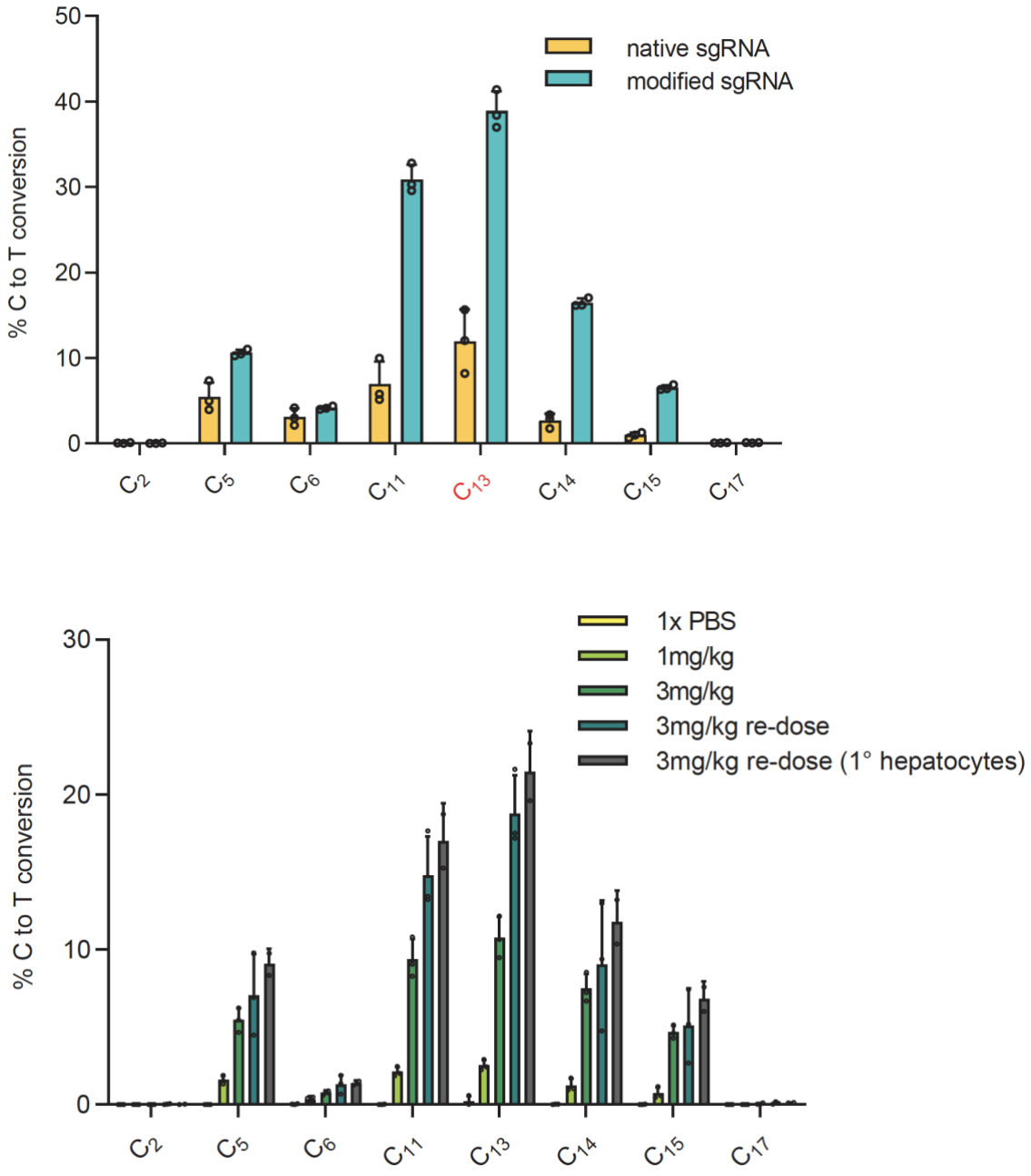
Supplementary Figure 14 | Indel formation after expression of different BE constructs *in vivo.* Indels per C-to-T edit (left panel) and indels per correctly edited read (right panel) are depicted. Bars represent mean of n=3 mice per treatment or n=6 mice for Gam-BE4 treatment for 14 weeks



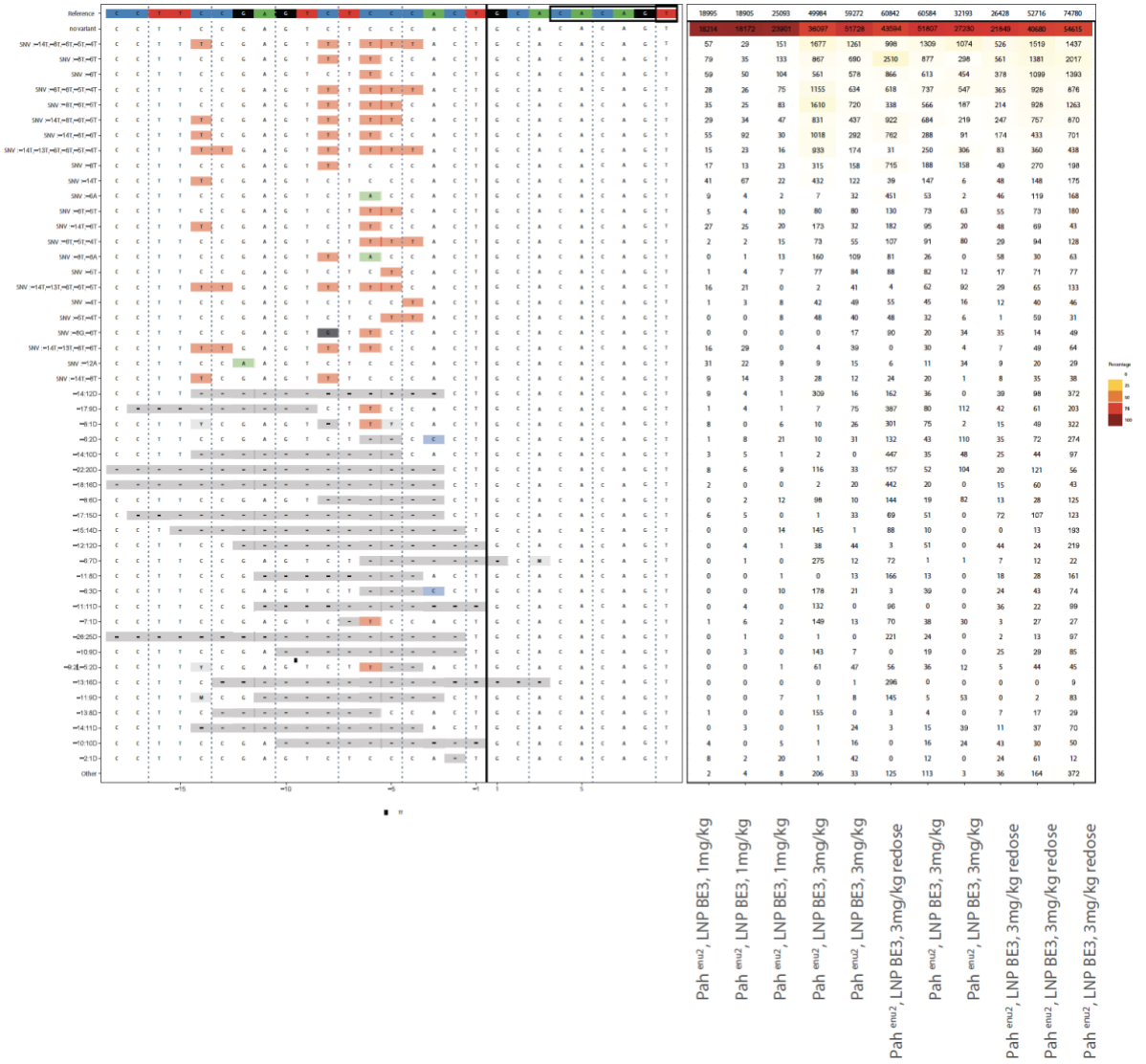
Supplementary Figure 15 | Unidirectional amplification of the target locus for unbiased evaluation of structural variants and large deletions. Correctly edited reads translate the wildtype PAH amino acid sequence, while incorrectly edited reads are nonsynonymous to the wildtype PAH amino acid sequence. 5' Enrichment data was derived by the use of a primer binding upstream of the cut site, while 3' enrichment was facilitated by the use of a primer binding downstream of the cut site. No large deletions or structural variants (SVs) were detected. For a detailed description of the experiment see methods section.



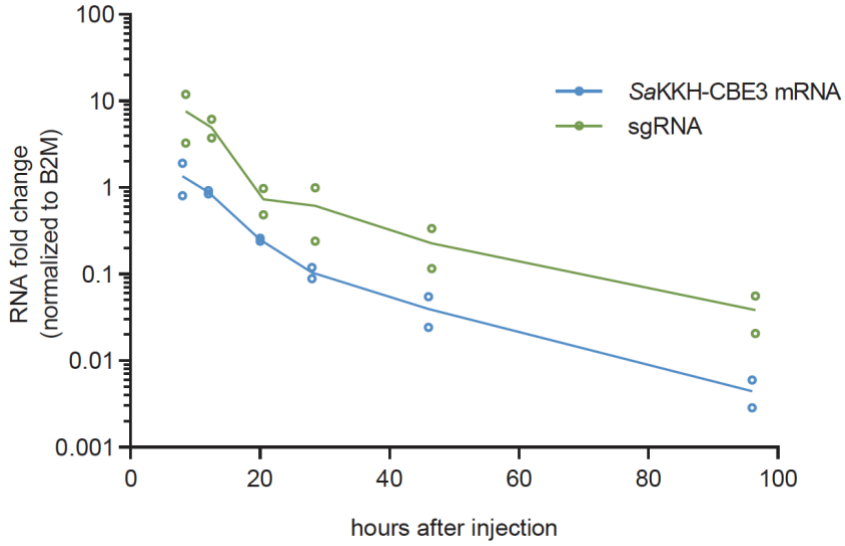
Supplementary Figure 16 | Tolerability of LNP. Cytokine profile from blood serum 4 h after administration of 3 mg/kg LNPs or 1x PBS via the tail vein. Statistical analysis was performed using two-sided Fisher's least statistical significance. Values represent mean ± s.d. of n=3 individual biological replicates.



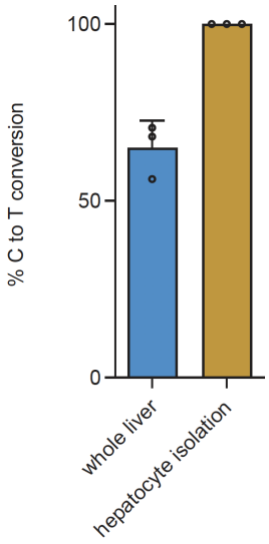
Supplementary Figure 17 | Editing efficiencies at different C positions. Editing efficiencies at different C positions within the protospacer region were determined *in vitro* using chemically modified and unmodified sgRNAs in HEK293T reporter cells (upper panel). Editing profiles *in vivo* were determined at different doses from whole liver lysates and isolated primary hepatocytes (lower panel). Values represent mean ± s.d. of n=3 individual biological replicates.



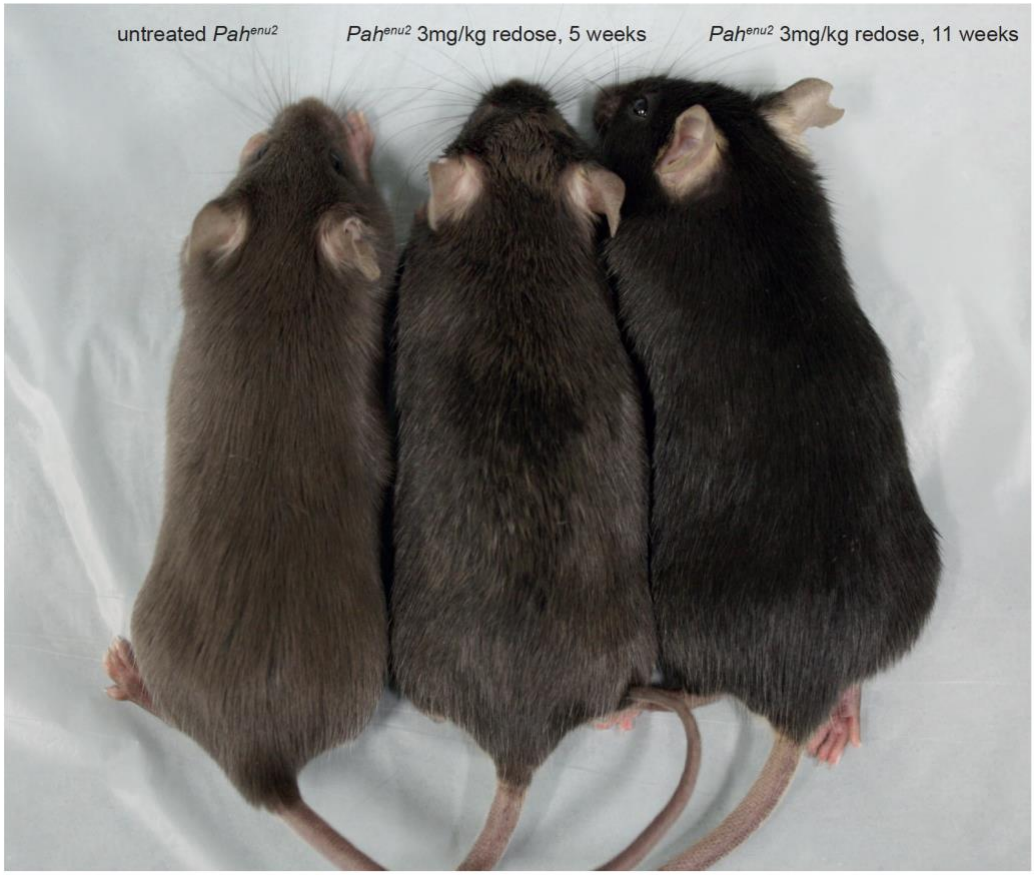
Supplementary Figure 18 | Allele plots of HTS data after LNP-mediated delivery of *Sa*KKH-BE3 *in vivo.* The left panel shows consensus sequences for the top 50 variant alleles. Variants that did not contain insertions or deletions within the guide region are separated and labelled according to any single nucleotide variants (SNVs) present. The right panel shows variant counts, with the header showing the total number of reads per sample. Data are from *in vivo* experiments after LNP-mediated delivery of SakKH-CBE3 at different doses, with and without re-dosing. All biological replicates are plotted individually.



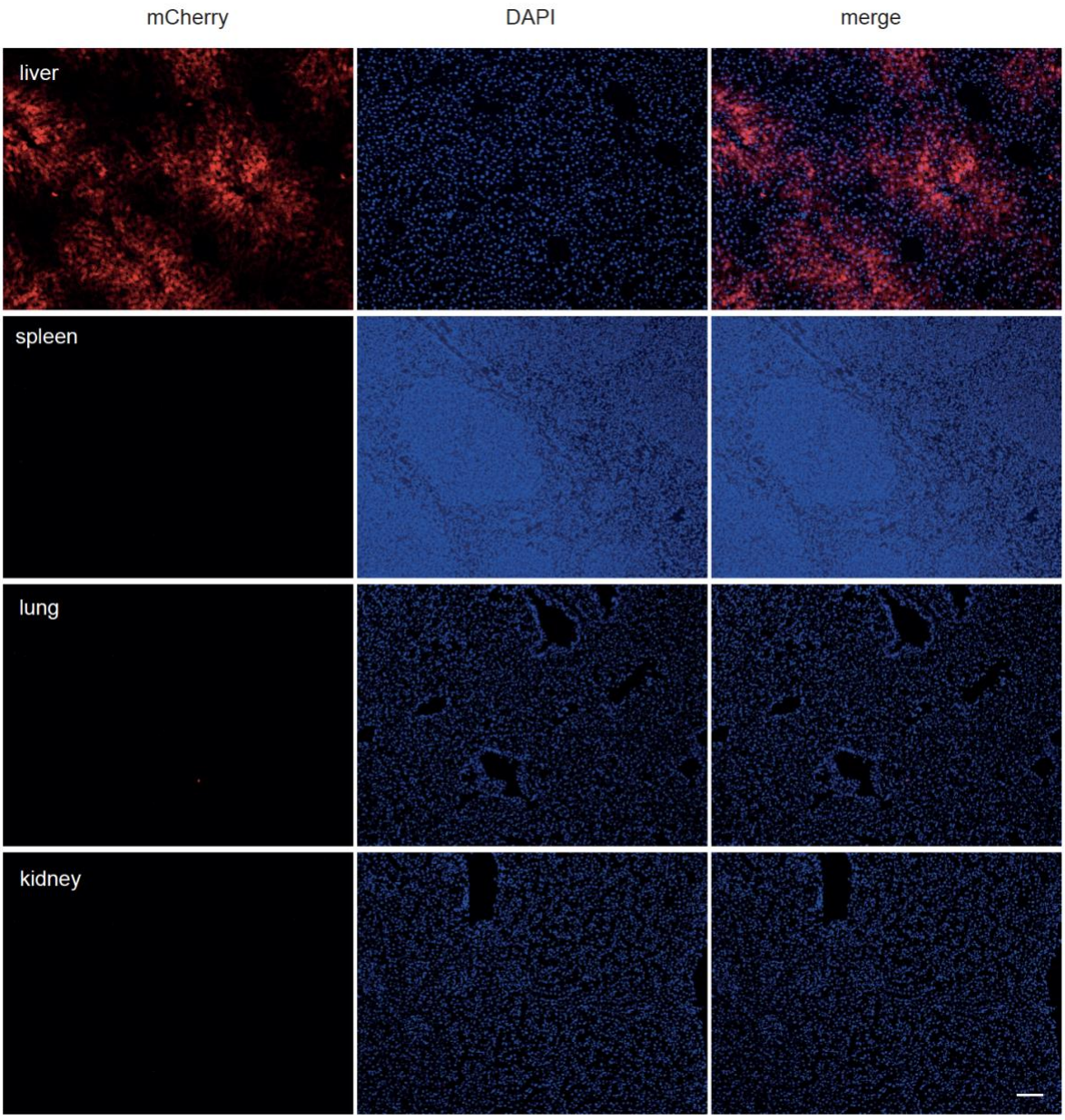
Supplementary Figure 19 | Stability of mRNA and sgRNA following LNP-mediated delivery *in vivo*. Guide RNA and mRNA fold-change in primary hepatocytes isolated 8 h after administration of 1 mg/kg LNP-encapsulated *Sa*KKH-CBE3 mRNA and sgRNA. n=3 biologically independent replicates.



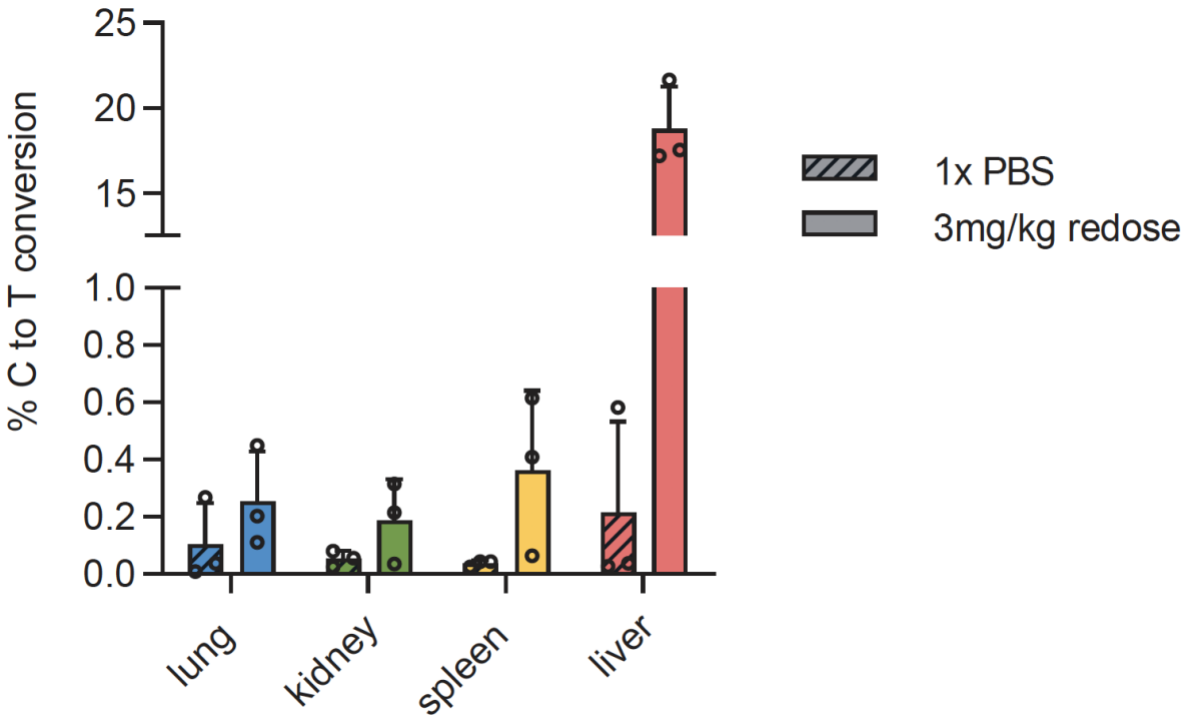
Supplementary Figure 20 | Editing efficiencies in whole liver extracts and in isolated primary hepatocyte. C to T conversion at C_{13} of isolated hepatocytes compared to genomic DNA from whole liver lysates. Values represent mean ± s.d. of n=3 individual biological replicates.



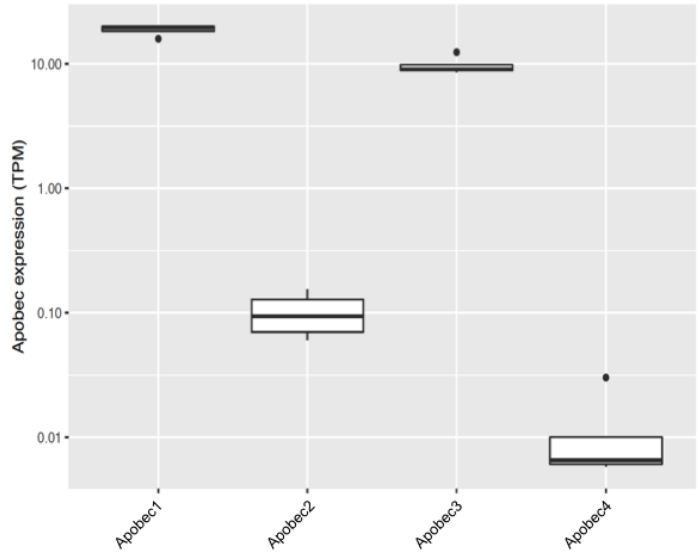
Supplementary Figure 21 | Fur phenotype in *Pah*^{enu2} after LNP-mediated genome editing. Mice after administration of PBS (left) and 3m/kg LNP encapsulating *Sa*KKH-CBE3 mRNA and sgRNA after 5 weeks (middle), and after 11 weeks (right).

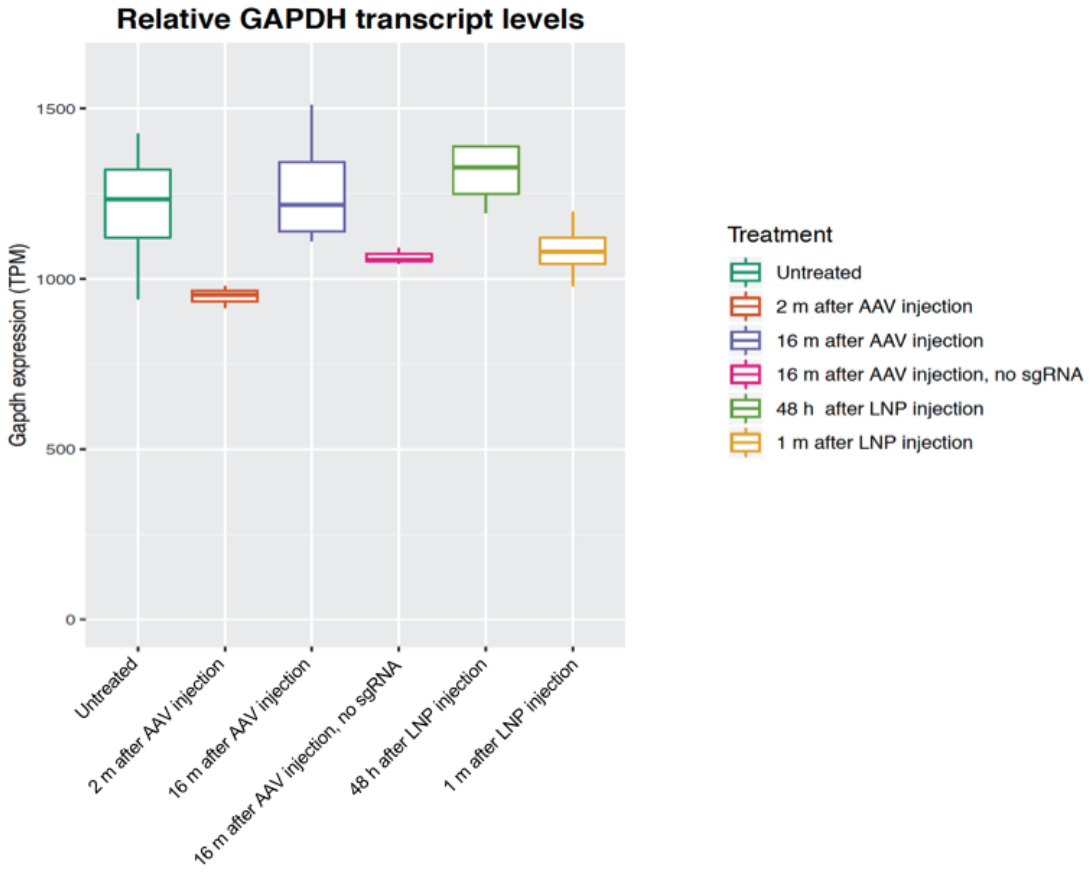


Supplementary Figure 22 | Hepatotropism of LNPs encapsulating mCherry mRNA. Representative cryosections of murine liver, spleen, lung and kidney after administration of 1mg/kg LNP-encapsulated mCherry mRNA. Representative pictures were selected from 4 pictures taken from 2 individual animals. Red channel, RFP, blue channel, DAPI. Scale bar, 100 µm.

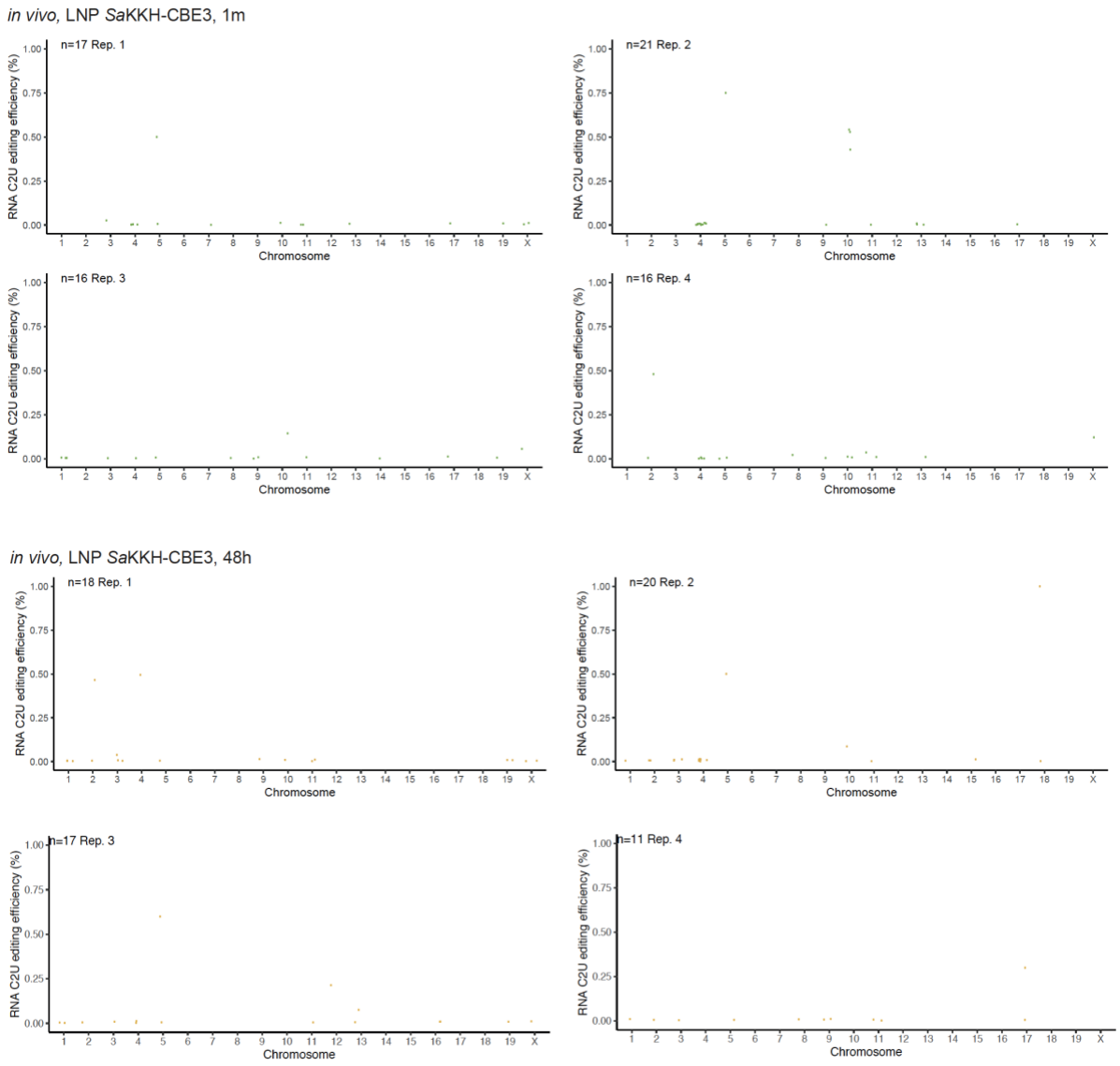


Supplementary Figure 23 | Hepatotropism of LNPs. Lung, kidney, and spleen were analysed after mice (n=3) were systemically dosed twice at 3mg/kg and compared to mice administered 1x PBS by HTS of the target locus from genomic DNA. Values represent mean ± s.d. of n=3 biological replicates.

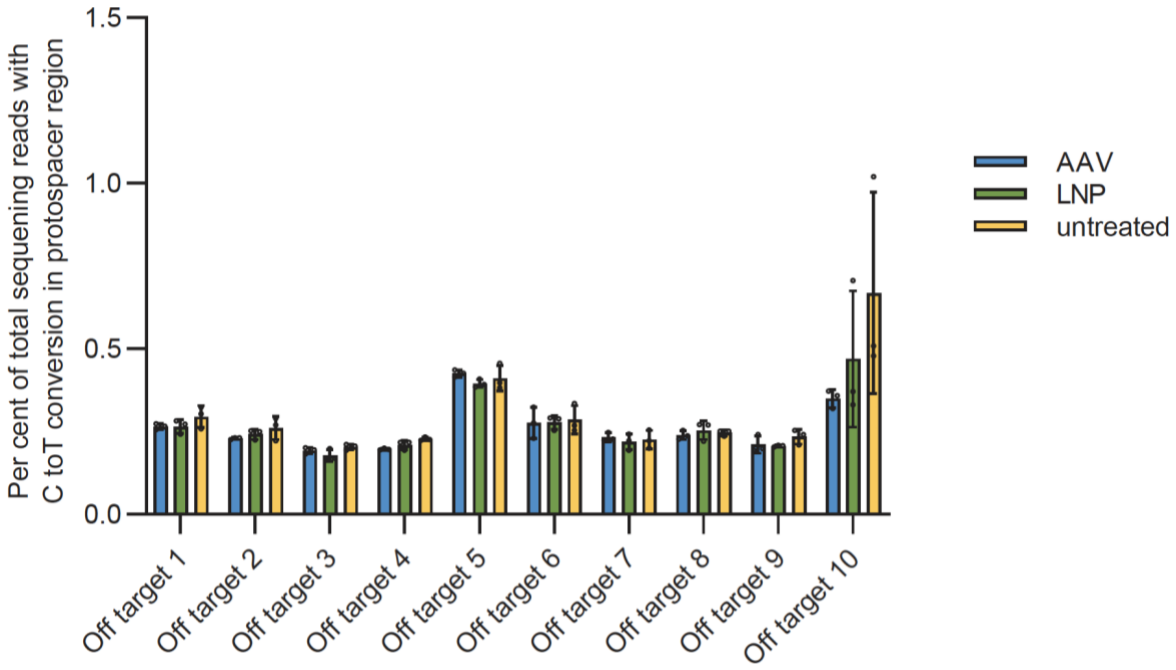




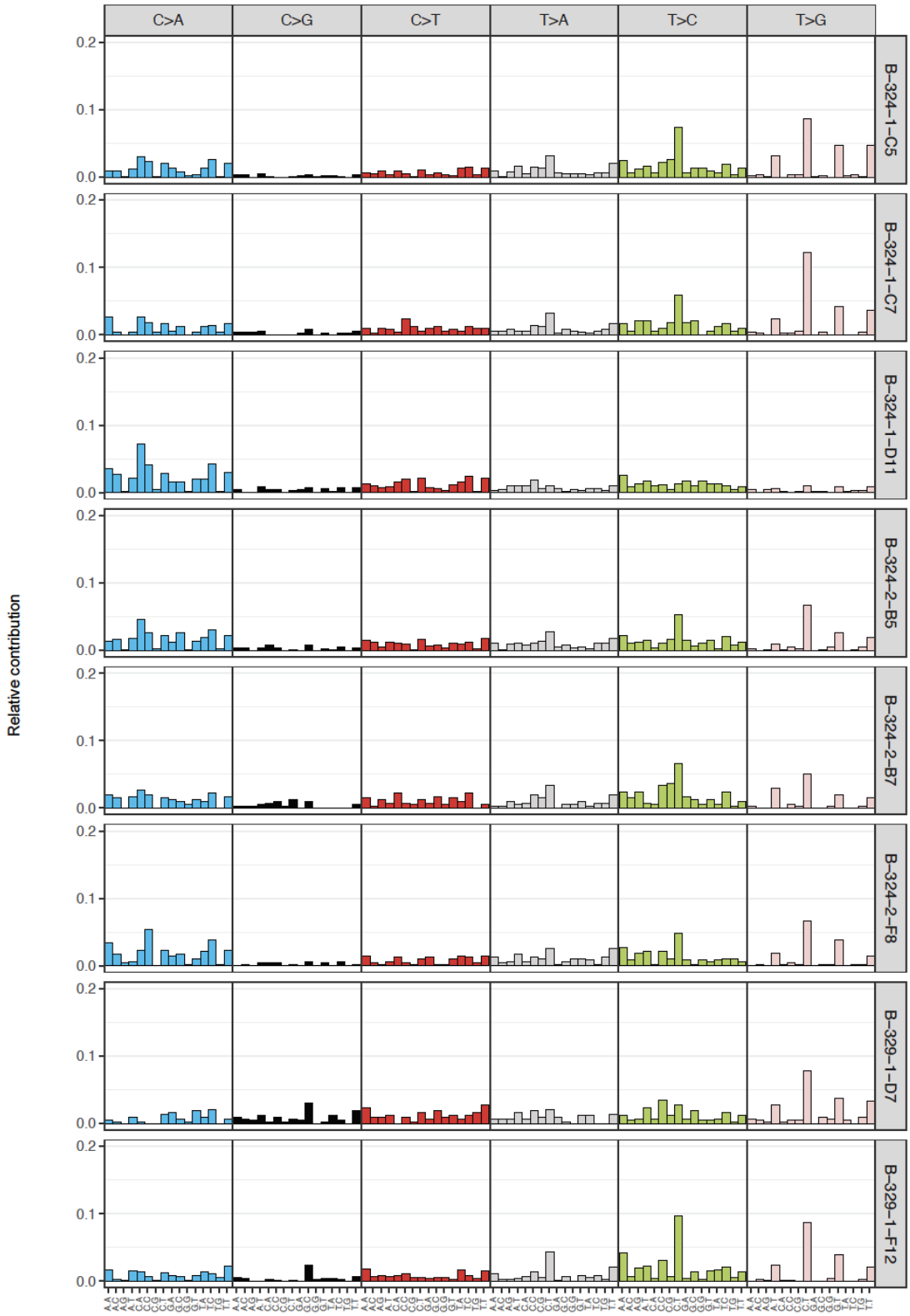
Supplementary Figure 24 | Endogenous mAPOBEC expression in the liver. Relative expression values for endogenous mouse APOBEC variants in the mouse liver are depicted in the top panel. n=3 untreated mice (upper panel). Relative GAPDH expression levels of *in vivo* samples shown in Fig1c and 4a. (tpm, transcripts per million). Box plots are standard Tukey plots, where the centre line represents the median, the lower and upper hinges represent the first and third quartiles, and whiskers represent + 1.5 the interquartile range. n=3 mice for 2 months after AAV injection, n=4 mice for all other samples (lower panel).

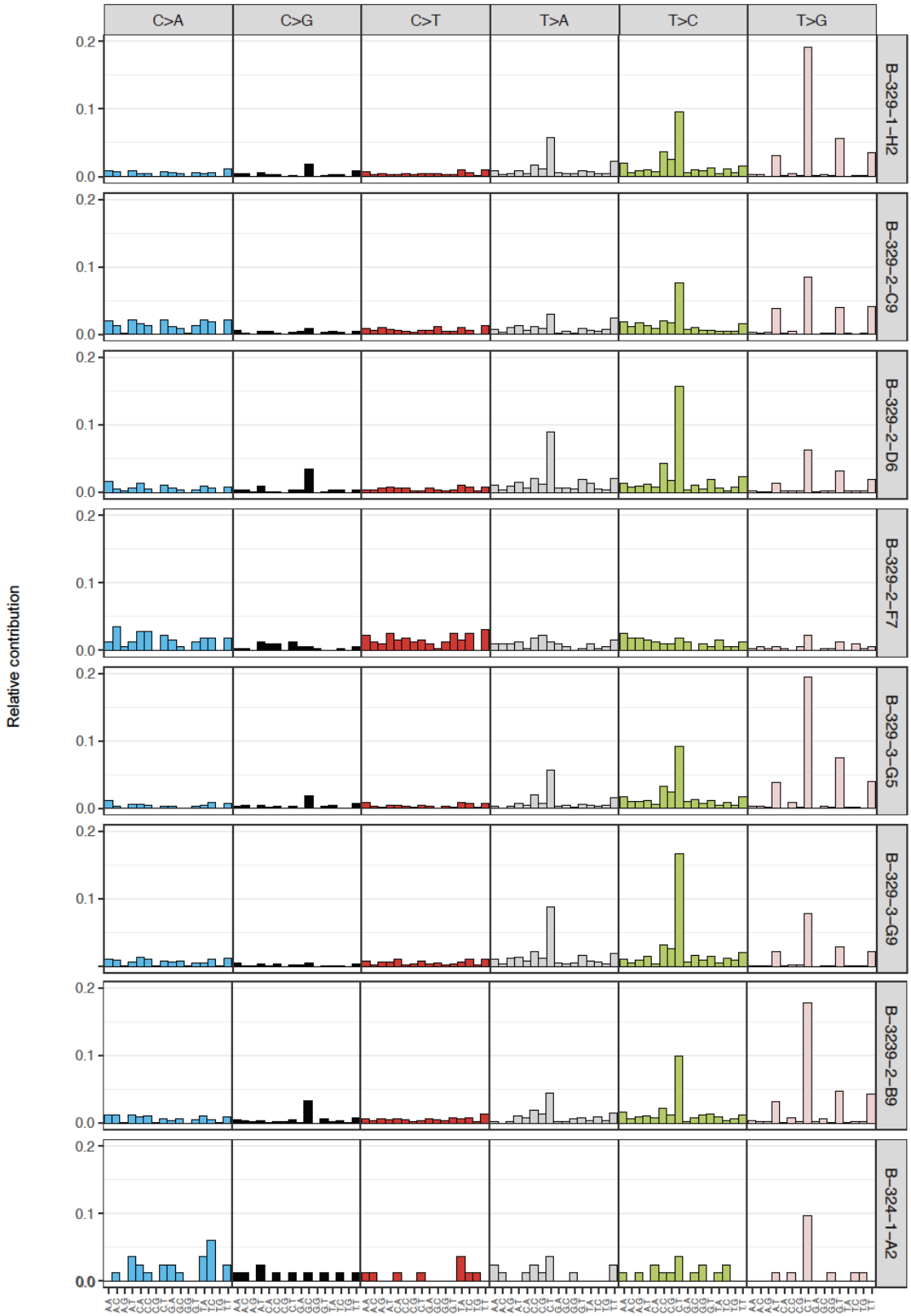


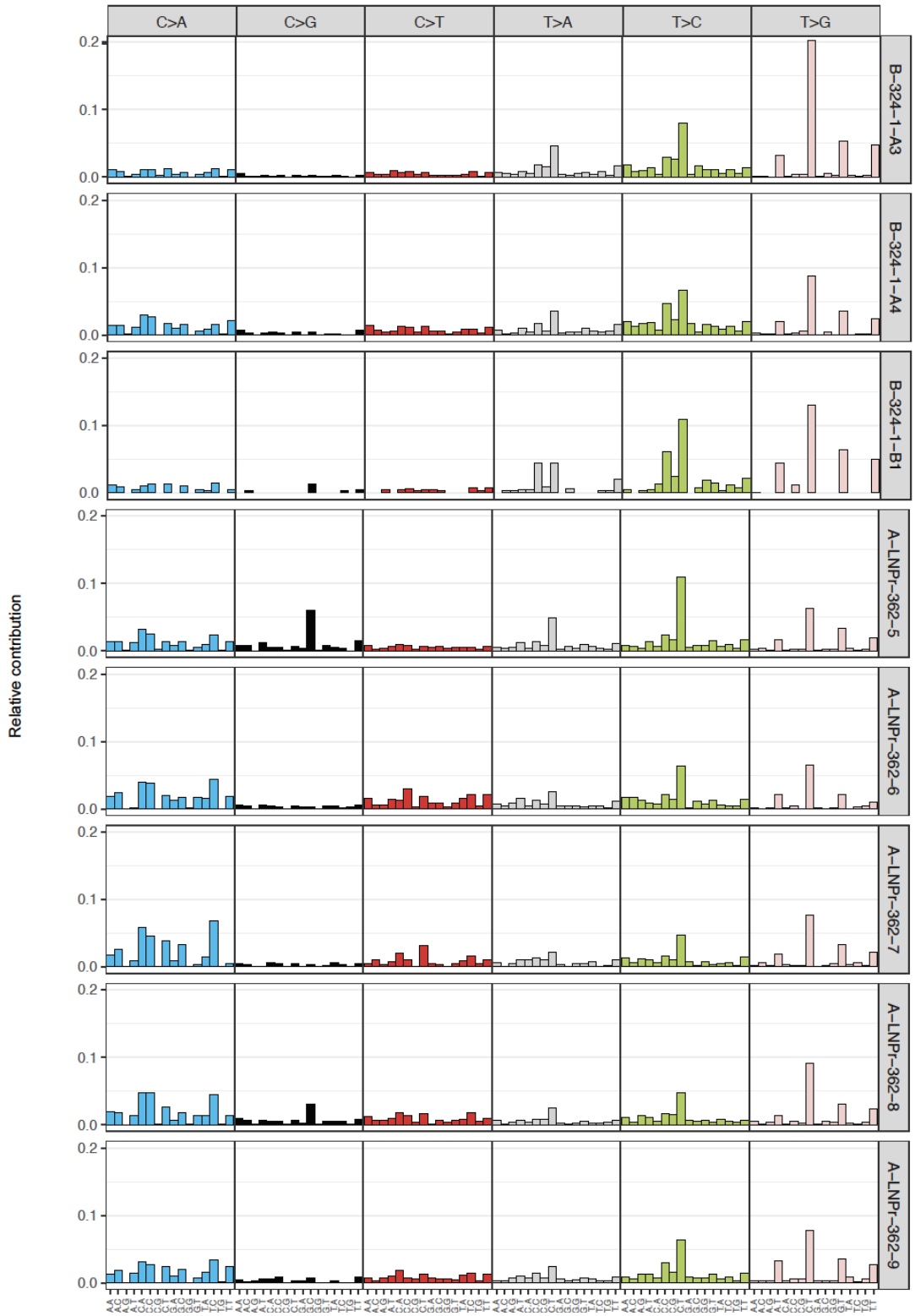
Supplementary Figure 25 | Manhattan plots depicting transcriptome-wide C-to-U editing events of base edited *in vivo* samples from mouse livers treated with LNP. Mouse livers treated with LNP delivery the *Sa*KKH-CBE3 mRNA and the *Pah*^{enu2}-targeting modified sgRNA. Each dot represents a C-to-U editing event. Counts per sample and replicates are indicated in the top left corners of each plot.



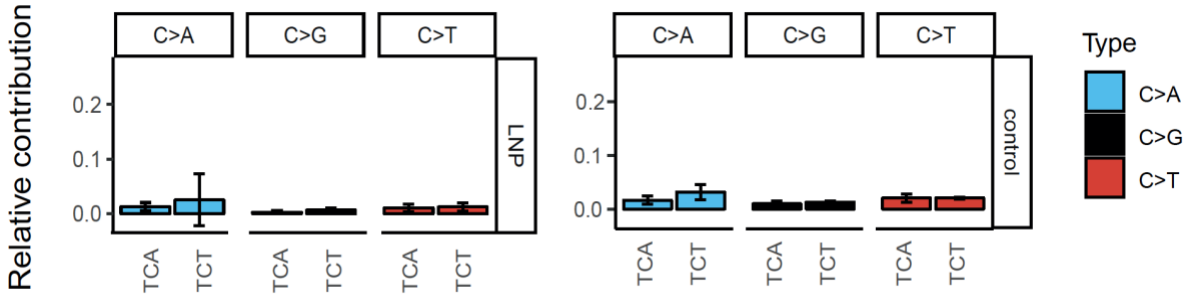
Supplementary Figure 26 | sgRNA-dependent off-targets. Percent of total sequencing reads with C-to-T conversion in protospacer region are depicted for AAV-, LNP-treated and untreated samples. Values represent mean ± s.d. of n=3 mice per group.


C>A
C>G
C>T
T>A
T>C
T>G
B–324–1–C5
B–324–1–C7
B–324–1–D11
B–324–2–B5
B–324–2–B7
B–324–2–F8
B–329–1–D7
B–329–1–F12
Relative contribution
0.0
0.1
0.2

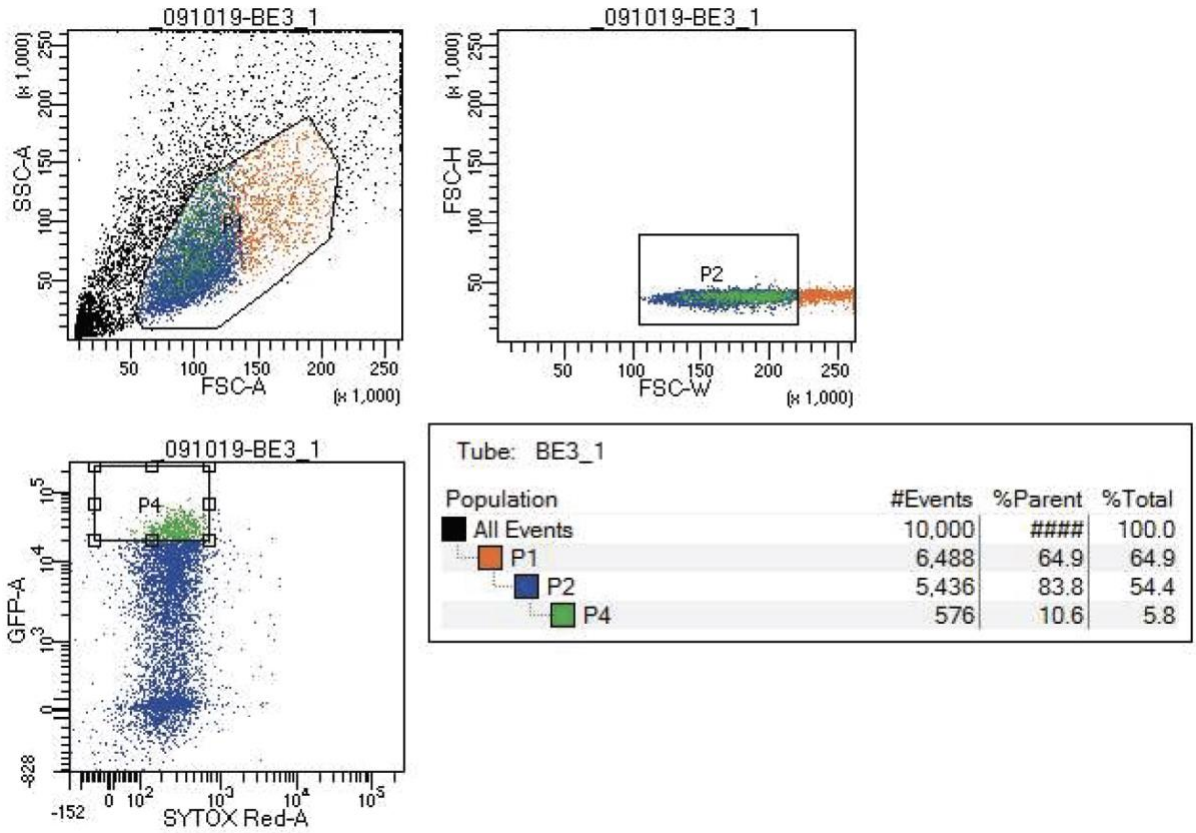

C>A
C>G
C>T
T>A
T>C
T>G
B-329-1-H2
B-329-2-C9
B-329-2-D6
B-329-2-F7
B-329-3-G5
B-329-3-G9
B-3239-2-B9
B-324-1-A2
Relative contribution
0.0
0.1
0.2



Supplementary Figure 27 | 96-nt profile plot of individual base edited clones following LNP delivery. Primary hepatocytes were isolated and clonally expanded as chemically induced liver progenitor cells (CLiPs). On-target editing and therefore base editor exposure in clones was confirmed by Sanger sequencing before WGS at an average 30x coverage. The frequency (y-axis) for 96 mutational types (x-axis) is shown. Control clones were derived from untreated mice. Plots present motif contributions of each clone represented in main figure 4d.

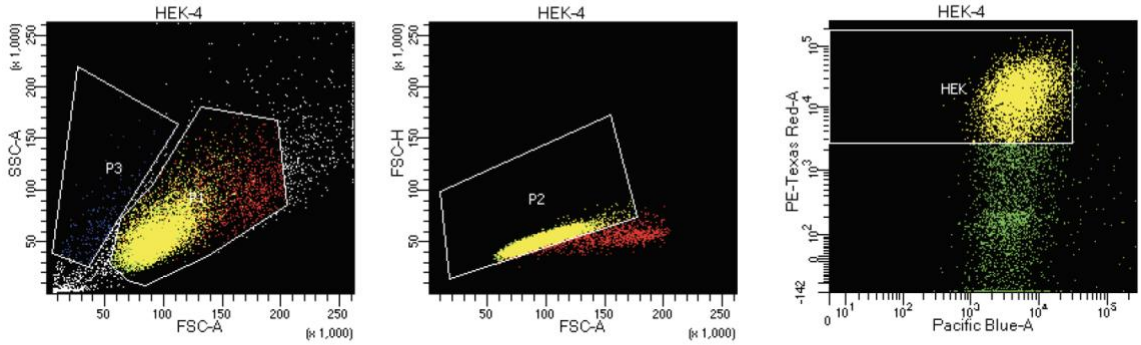


Supplementary Figure 28 | Detailed motif-dependent analysis of identified variants. Relative contributions of the APOBEC-relatable sequence motif of C>A (G>T), C>G (G>C) and C>T (G>A) conversions of LNP-treated hepatocytes versus unedited controls. Values represent mean ± s.d. of n=24 (for LNP) or n=3 (for control) clones per group.

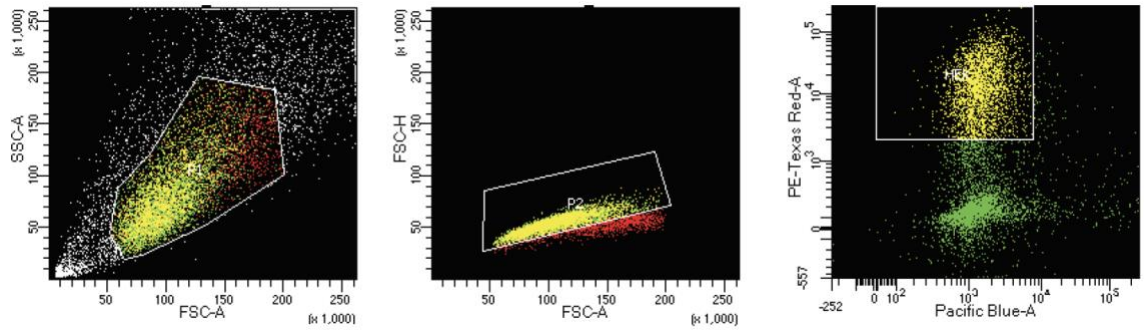


Supplementary Figure 29 | FACS gating strategy for sorting the top 5% GFP positive HEK293T cells. Cells were co-transfected with the *Sa*KKH-CBE3 plasmid and the *Pah*^{enu2}-targeting sgRNA plasmid to determine transcriptome-wide off-target effects depicted in Figure 1b.

HEK293T



Hepa 1-6



Supplementary Figure 30 | FACS gating strategy for sorting all RFP positive HEK293T cells (upper panel) or Hepa1-6 cells (lower panel). Cells were co-transfected with *Sa*KKH-CBE3 mRNA and the *Pah*^{enu2}-targeting sgRNA to determine transcriptome-wide off-target effects at lower dose depicted in Figure 1e and Supplementary Figure 5.

Supplementary References

1. Grünewald, J. *et al.* Transcriptome-wide off-target RNA editing induced by CRISPR-guided DNA base editors. *Nature* (2019). doi:10.1038/s41586-019-1161-z

2. Katsuda, T. *et al.* Conversion of Terminally Committed Hepatocytes to Culturable Bipotent Progenitor Cells with Regenerative Capacity. (2017). doi:10.1016/j.stem.2016.10.007

3. Zuo, E. *et al.* Cytosine base editor generates substantial off-target single-nucleotide variants in mouse embryos. *Science (80-.).* **126**, 21 (2019).