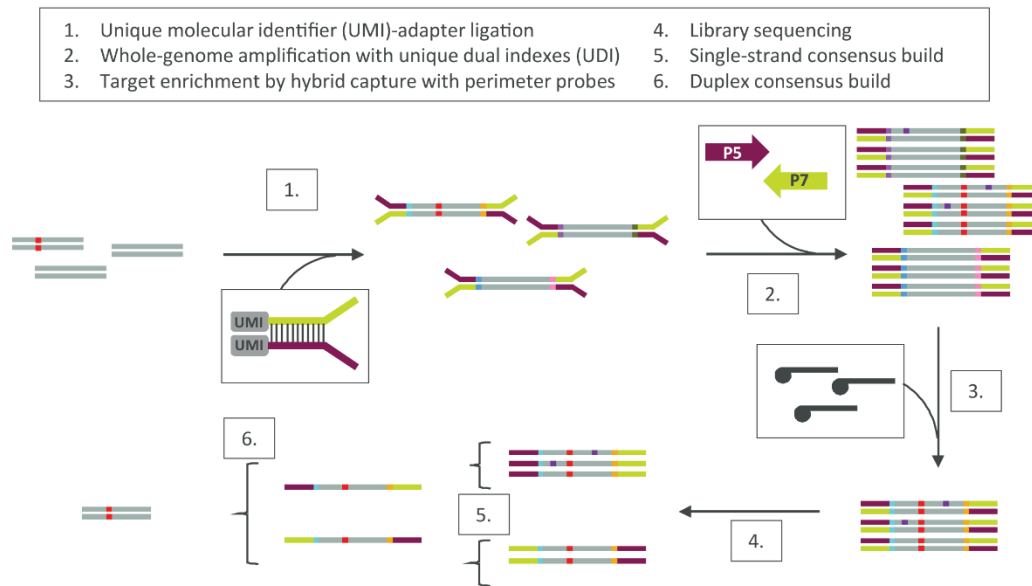


A Type II-B Cas9 nuclease with minimized off-targets and reduced chromosomal translocations *in vivo*.

Bestas et al. Supplementary Figure 1



Supplementary Figure 1. Overview Duplex-Seq workflow.

Schematic depiction of Duplex-Seq workflow. After shearing of genomic DNA, UMI-labelled adapters are ligated to the DNA fragments (1), followed by whole-genome amplification with UDIs (2). Enrichment of target sites from this unbiased and unenriched library is performed by hybrid capture with four perimeter probes per site (3). After library sequencing (4), a single-strand consensus sequence is built from replicate reads of each strand (5), followed by duplex-consensus sequence generation (6). The double consensus build facilitates superior sequencing error elimination (purple squares) compared to standard NGS analysis, enabling a more sensitive identification of actual mutations (red squares). Furthermore, probe-based target enrichment enables analysis of translocations. UMIs are depicted as differently coloured boxes at the end of each adapter. UMI=unique molecular identifier; UDIs=unique dual indexes.

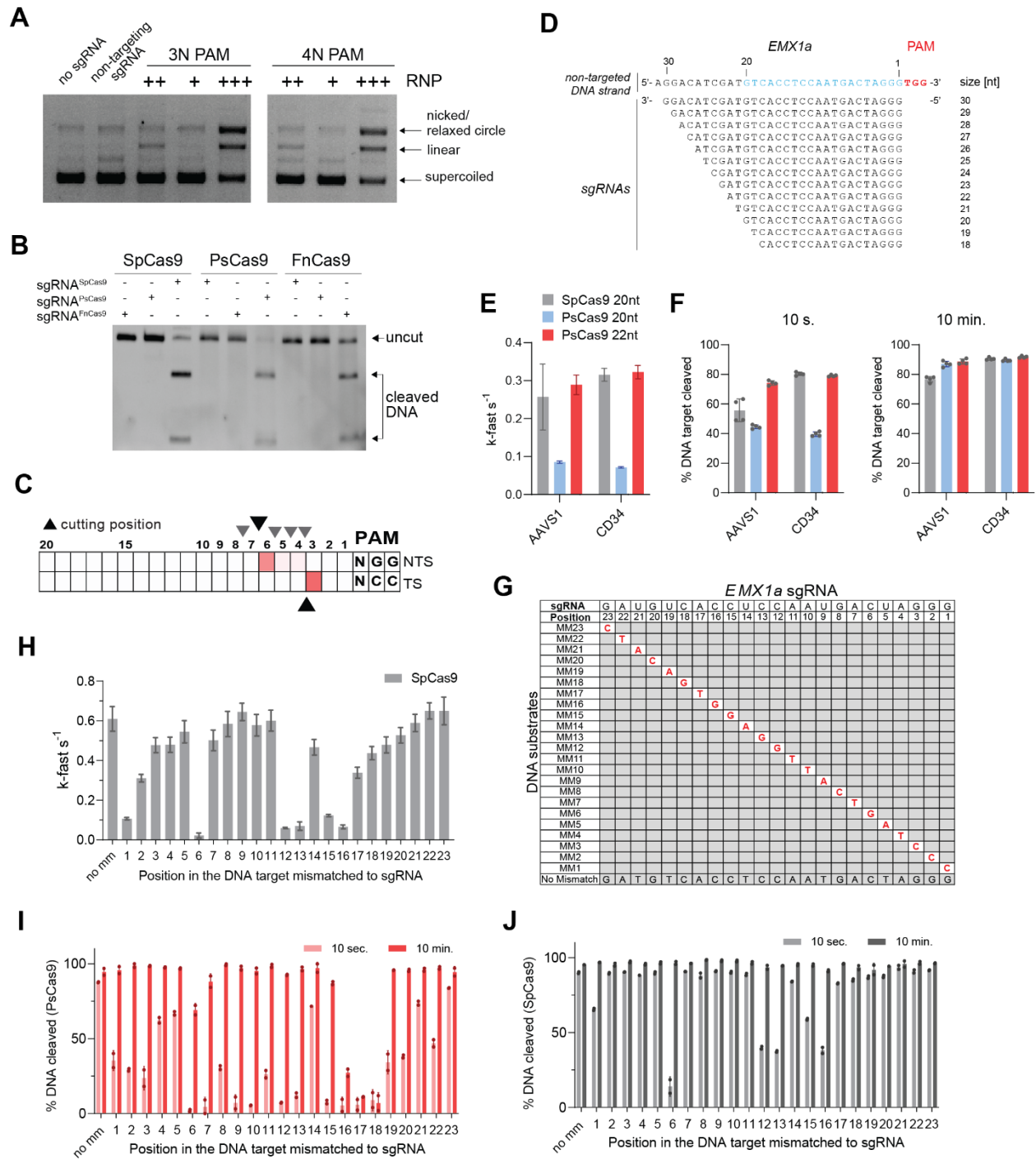
Supplementary Figure 2. Extended table *in vivo* amplicon sequencing data.

Extended table from Figure 1C showing gene editing efficiency for all target sites assayed by multiplexed amplicon sequencing followed by CRISPResso2 analysis. Heatmaps present editing percentages for the on-target sites (highlighted in green) and 78 off-target sites (highlighted in magenta). Amplicon sequencing was performed with n=3 biologically independent mice per group (1, 2, 3) using genomic DNA isolated from the liver of mice treated with adenoviral vector expressing either SpCas9 plus gMH sgRNA or SpCas9-GFP as a control. Values show non-normalized editing data as percent indel-containing reads. All values < 0.1% fall below the detection limit of amplicon sequencing and are considered as non-specific/background.

Supplementary Figure 3. Extended table *in vivo* Duplex-Seq data.

Extended table from Figures 1B and 4C showing gene editing efficiency for all on/off-target sites assayed by Duplex-Seq followed by CRISPResso2 analysis. Heatmaps present editing percentages for the on-target sites (highlighted in green) and 75 off-target sites (highlighted in magenta). Duplex-Seq was performed with n=3 biologically independent mice per group (1, 2, 3) using genomic DNA isolated from the liver of mice treated with adenoviral vector expressing sgRNA gMH together with SpCas9 or PsCas9. Adenoviral vector encoding SpCas9 without a sgRNA was used as a control. Values show non-normalized editing data as percent indel-containing reads. The detection limit reached in this experimental set-up was 0.01%. All Duplex-Seq experiments were performed in parallel in the same experiment.

Bestas et al. Supplementary Figure 4

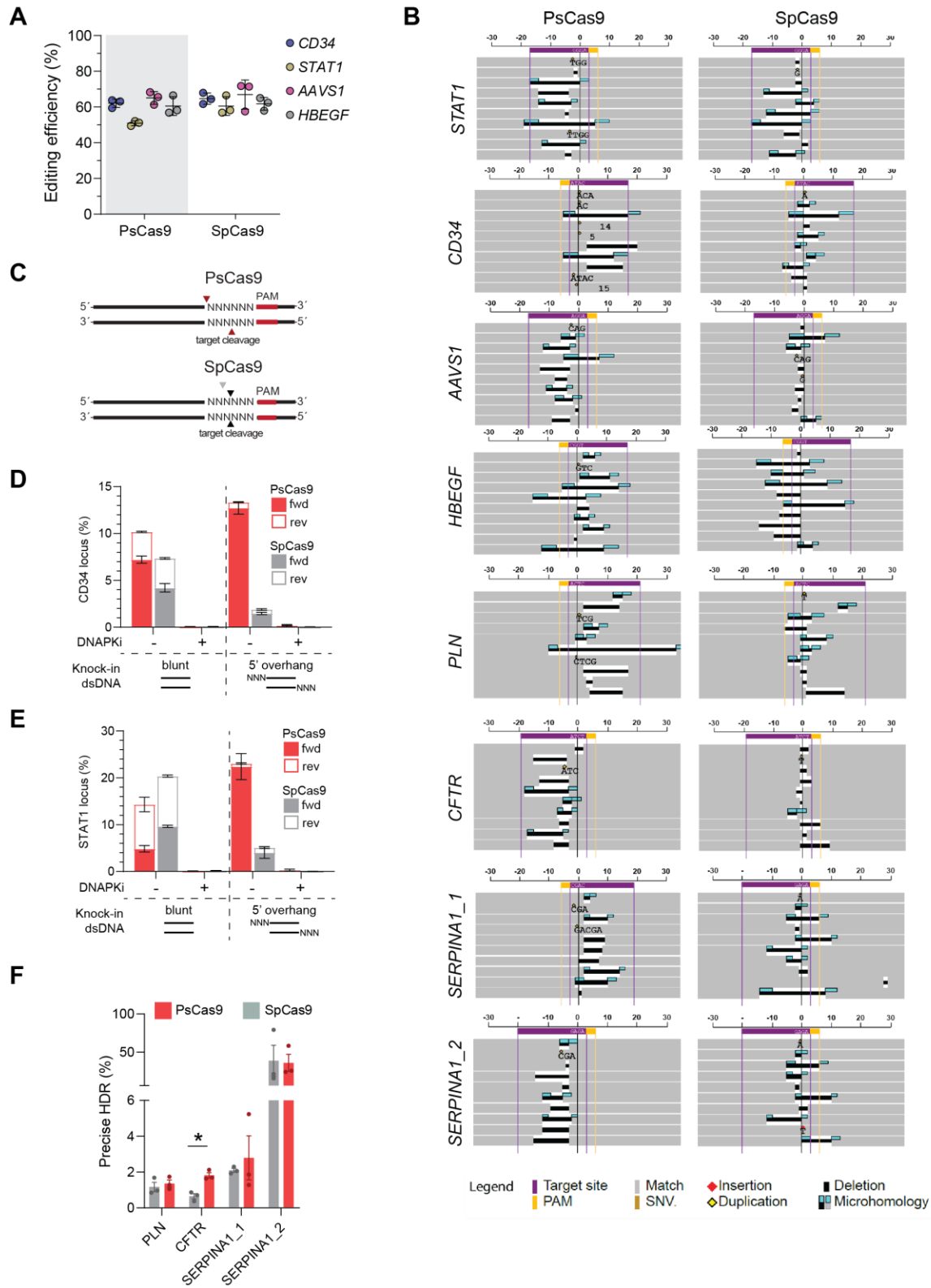


Supplementary Figure 4. Extended characterization of PsCas9, an RNA-guided DNA nuclease.

a. Gel electrophoresis analysis of 3 and 4 base pair randomized PAM sequence containing plasmid libraries (3N, 4N) cut with three different doses of RNP complex: 3 nM(+), 30 nM(++), 300 nM(+++). The first two lanes show plasmid libraries without RNP treatment as a control. Representative gel of at least two independent experiments.

- b. Gel electrophoresis of *in vitro* cleaved or uncut DNA products after incubation with a nuclease-specific crRNA:tracrRNA duplex. Each combination for SpCas9, PsCas9 and FnCas9 were assessed individually. Representative gel of at least two independent experiments.
- c. PsCas9 cleavage pattern extracted from the sequenced PAM plasmid library (4N). Heat map shows the calculated frequent (black arrows) and less-frequent (grey arrows) cleavage sites on target-strand (TS) and non-target strand (NTS).
- d. Spacer targeting *EMX1a* with varying length from 18 to 30 nucleotides which were used in *in vitro* cleavage assays to determine the ideal configuration for PsCas9. The preferred spacer length of SpCas9 (20 bp) is highlighted in blue.
- e, f. Plots showing the reaction rate (e) and the fraction (f) of *AAVSI* and *CD34* DNA substrate digested by SpCas9 and PsCas9 over time (10 sec and 10min) in the *in vitro* DNA cleavage assay. The amount of substrate digested was calculated using local integral fluorescent of capillary electrophoresis traces and fitted to the bi-exponential decay function. Reaction rate is presented as k_{fast} (kf). While SpCas9 was loaded with 20-nt spacer, PsCas9 was loaded with either 20-nt or 22-nt long spacers targeting depicted sites. Data are presented as mean \pm SD, n=4 (all experimental timepoints in Source Data).
- g. The *EMX1a* sgRNA sequence and DNA substrates that were used in Figure 2H and Supplementary Figure 4H-J to determine DNA cleavage rates.
- h. In vitro DNA cleavage rates by wild-type SpCas9 with a mismatch at every single base position along the protospacer (see Supplementary Figure 4G). Mismatch positions are labelled in 3' to 5' direction, with the position 1 being directly upstream PAM. Reaction rate is presented as k_{fast} (kf). Data are presented as mean \pm SD, n=2 (all experimental timepoints in Source Data).
- i-j. In vitro the fraction of cleaved DNA of *EMX1a* by PsCas9 (i) and SpCas9 (j) with a mismatch at every single base position along the protospacer (see Figure 2H and Supplementary Figure 4G). Mismatch positions are labelled in 3' to 5' direction, with Position 1 being directly upstream PAM. Data are presented as mean \pm SD, n=2 (all experimental timepoints in Source Data).

Bestas et al. Supplementary Figure 5



Supplementary Figure 5. Extended analysis of PsCas9 edits in the human genome.

a. Editing efficiency of the human genome by PsCas9 and SpCas9 on depicted loci upon plasmid transfection of HEK293T cells. The editing efficiency was assessed by CRISPResso2 analysis of NGS data and shown as the percentage modified reads in each sample. Data are presented as mean \pm SD, n=3.

b. Visualization of the top ten most frequent mutations of PsCas9 and SpCas9 using the same sgRNA at selected target sites (*STAT1*, *CD34*, *AAVS1*, *EMX1a*, *HBEGF*, *PLN*, *CFTR* and *SERPINA1* with two different sgRNAs) analyzed by RIMA⁵⁵.

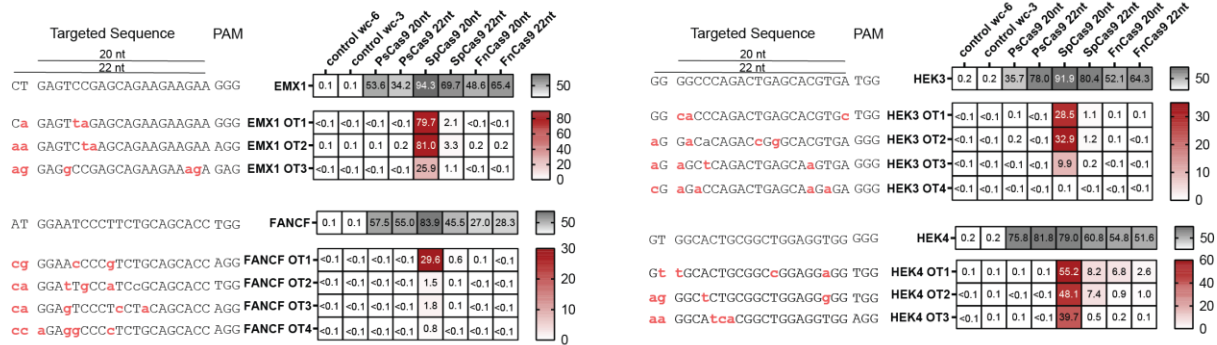
c. Schematic depicting the genomic DNA cutting pattern induced by PsCas9 (red) or SpCas9 (black/grey).

d-e. Genome insertion efficiency (knock-in, %) at *CD34* (d) or *STAT1* (e) by PsCas9 (red) and SpCas9 (grey) in HEK293T cells. Double-stranded DNA substrates used for the insertions are depicted below. Experiments were performed with or without DNA-PK inhibitor (DNA-PKi, M9831). The knock-in efficiency (reverse (empty) or forward (filled) direction) was calculated using NGS and shown as percentage of knock-in from total modified reads in each sample. Data are presented as mean \pm SD, n=3.

f. Precise HDR-mediated genome insertion efficiency (%) at *PLN*, *CFTR*, *SERPINA1* (two independent sgRNAs) by PsCas9 (red) and SpCas9 (grey) in HEK293T cells. The precise HDR efficiency was calculated using NGS and shown as percentage of knock-in from total reads in each sample. A two-sided Student's t-test was performed to evaluate statistical significance (*P-value = 00.5317). Data are presented as mean \pm SD, n=3.

Bestas et al. Supplementary Figure 6

A



B

CHANGE-seq off-targets present in all replicates

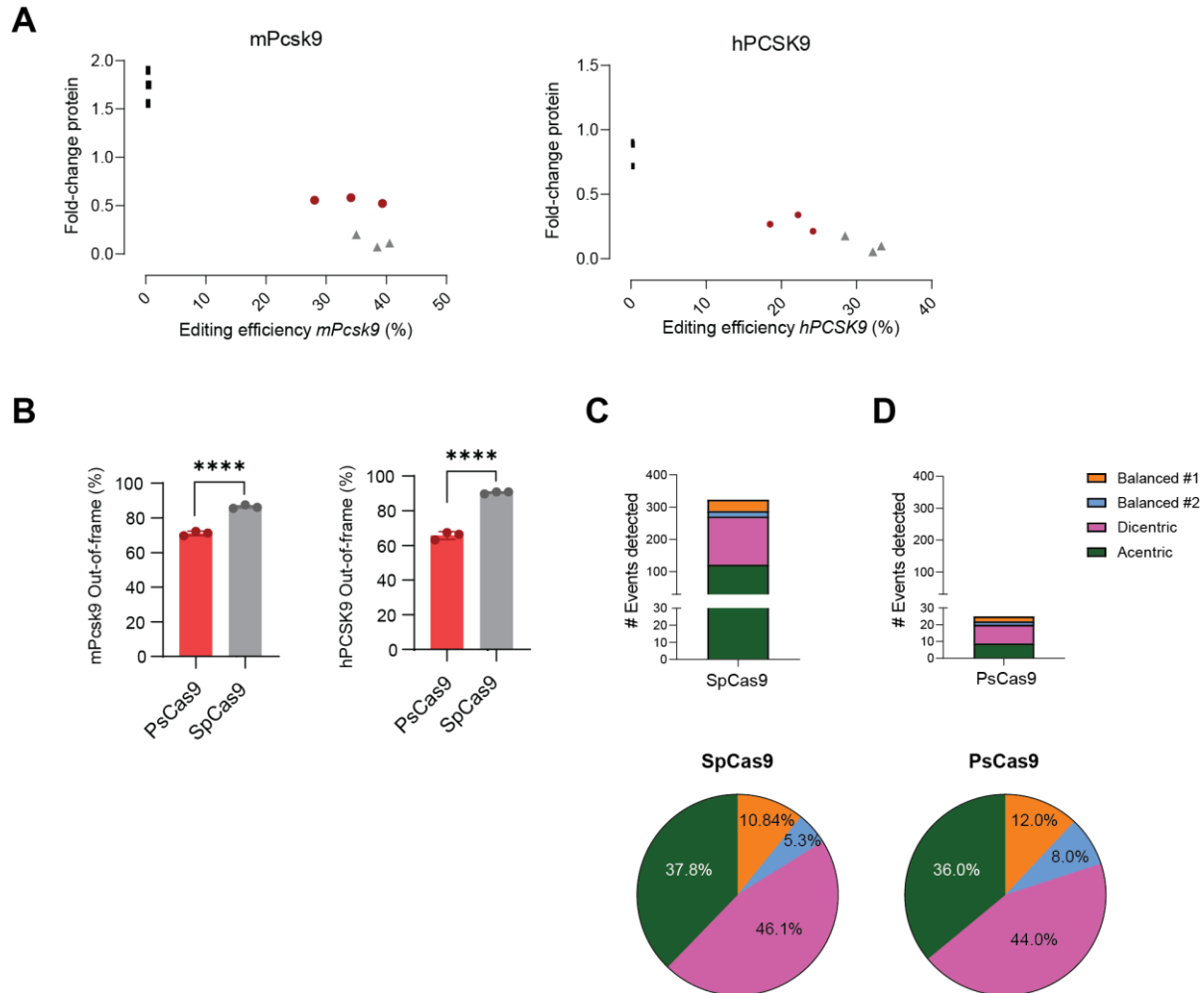


Supplementary Figure 6. Extended analysis of PsCas9 edits in the human genome.

a. Heatmap showing *EMX1*, *FANCF*, *HEK3* and *HEK4* editing and off-target editing by PsCas9, SpCas9 and FnCas9 with 20 nt and 22 nt spacer (right). The DNA on-target site and three to four off-target sites are shown on the left with mismatches highlighted in red. Control samples were untransfected and the editing efficiency was determined using NGS. Sequencing data was analyzed by CRISPResso2 and presented as percentages of modified reads in each sample. Data are presented as mean ± SD, n=3.

b. Chart showing the HEK4 CHANGE-seq off-target hits overlap between SpCas9 (20-nt spacer) and PsCas9 (22-nt spacer). Only hits that are present in all three replicates were considered (Supplementary Data Table 7).

Bestas et al. Supplementary Figure 7



Supplementary Figure 7. Extended *in vivo* data.

a. Correlation of editing efficiency and reduction of plasma protein levels *in vivo*. Editing efficiencies at the mouse *Pcsk9* (left) and human *PCSK9* (right) loci were determined by targeted deep amplicon sequencing. Control samples are presented in black, PsCas9 in red and SpCas9 in grey. Changes in plasma protein levels were calculated from ELISA measurements of plasma samples taken prior to virus injection (baseline) and during termination. Note that control virus-treated animals showed an increase in mouse *Pcsk9* plasma levels over the course of the study (represented as negative values of reduction).

b. Editing outcome at the on-target sites *in vivo*. The percentage of out-of-frame mutations induced at the mouse (left panel) and human *PCSK9* sites (right panel) as calculated from amplicon-sequencing results is shown. Graphs show

individual values for each mouse (n=3 independent animals), as well as mean \pm SEM. Unpaired, two-tailed t-test was performed to evaluate statistical significance (****P-value \leq 0.0001).

c-d. Bar plots (above) and pie charts (below) presenting color-coded types of translocation events detected in Duplex-seq data in the animals treated with SpCas9 (c) and PsCas9 (d).