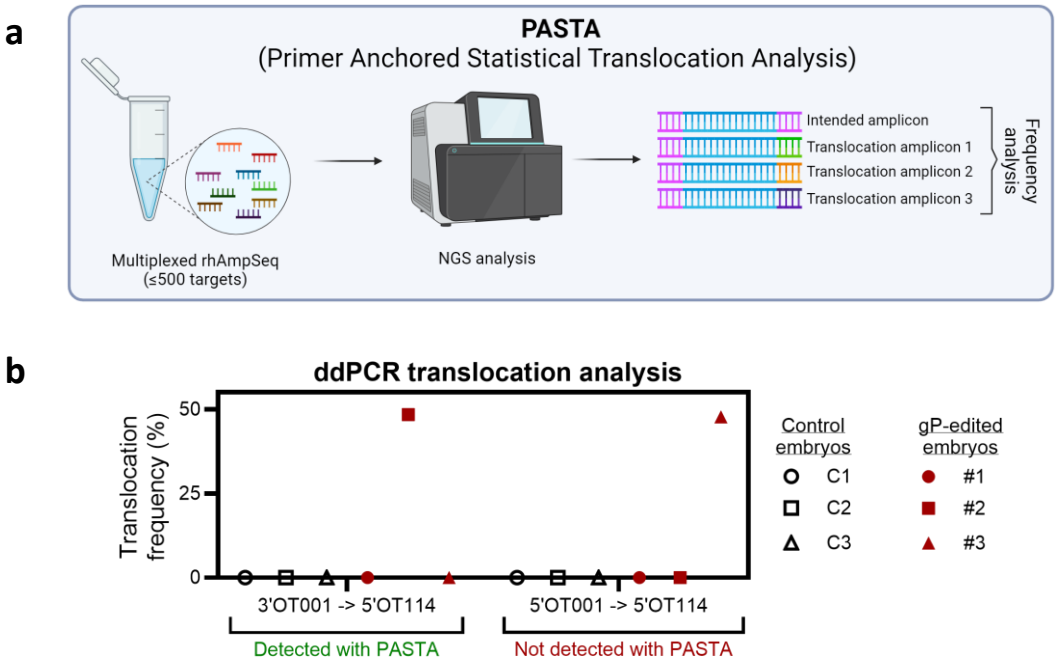
Supplementary information: Single-cell and in vivo profiling reveal heterogeneous and organ-specific CRISPR-Cas9 off-target and translocation outcomes

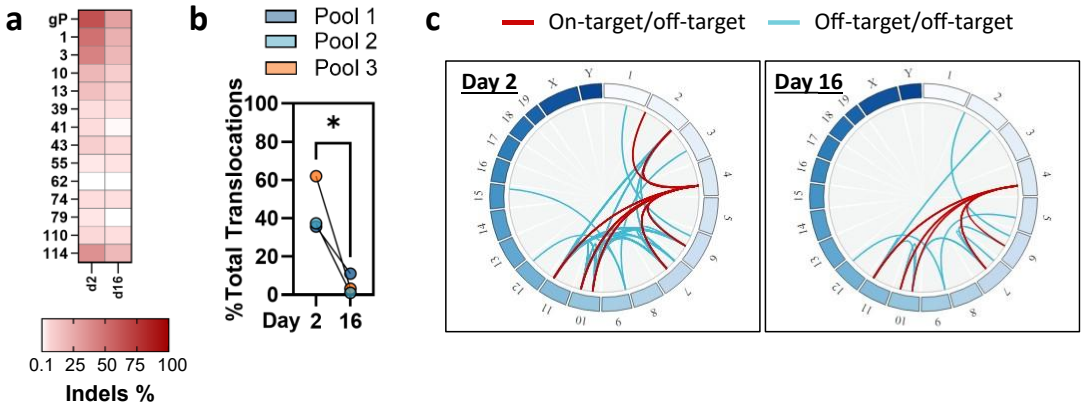
Supplementary Figures

Supplementary Figure 1



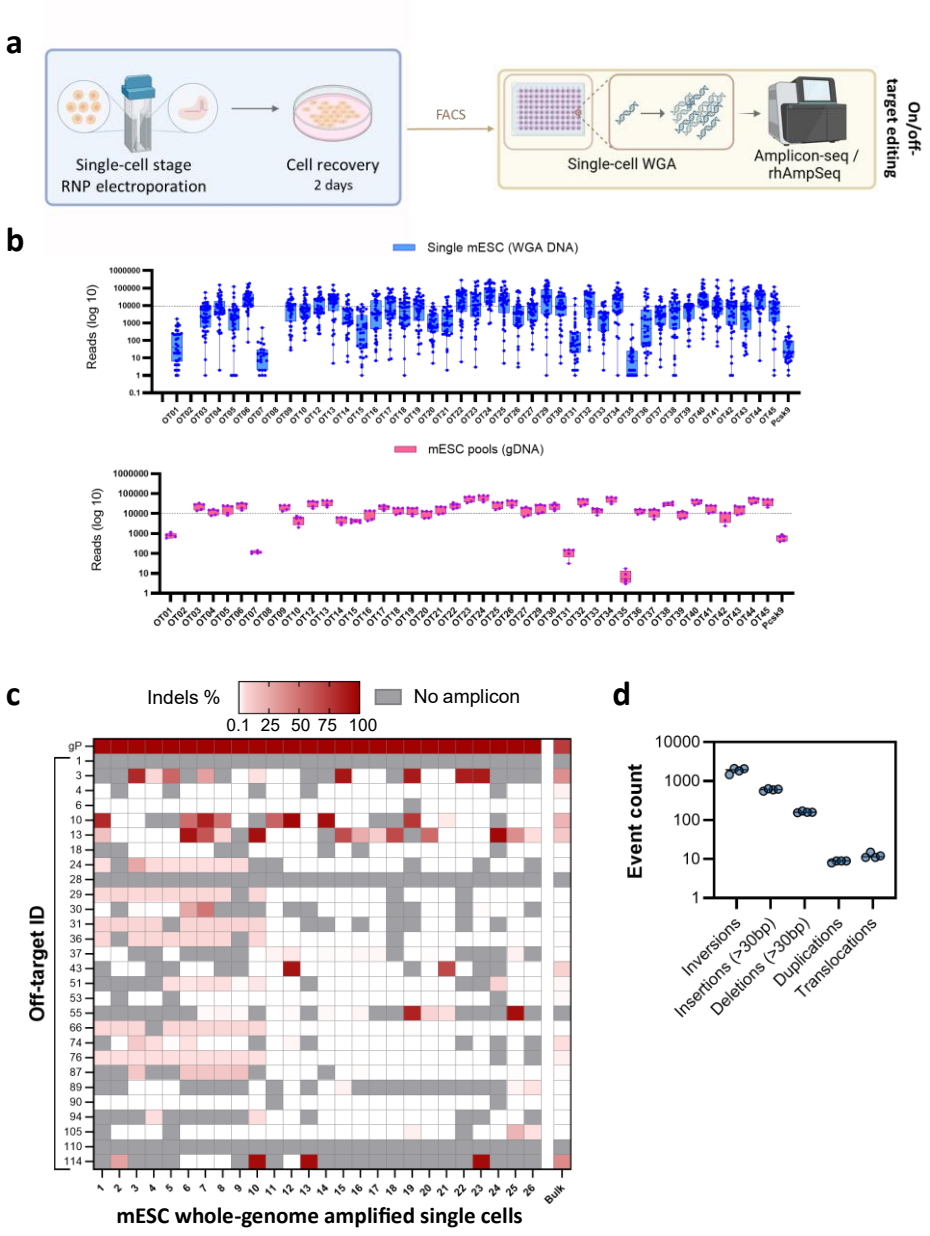
Supplementary Figure 1 – Translocation analysis with PASTA and ddPCR.
a) Schematic of PASTA translocation analysis. Created in BioRender. Madsen, A. (2026) https://BioRender.com/r8ngkyt. The multiplexed nature of rhAmpSeq gives rise to translocation product amplicons during library preparation in addition to the expected amplicons of off-target sites, which allows frequency analysis of the detected translocations using the rhAmpSeq sequencing data. b) Targeted translocation analysis in mouse embryos by ddPCR. Detectability of the indicated translocations by PASTA is indicated at the bottom of the graph. Source data are provided as a Source Data file.

Supplementary Figure 2



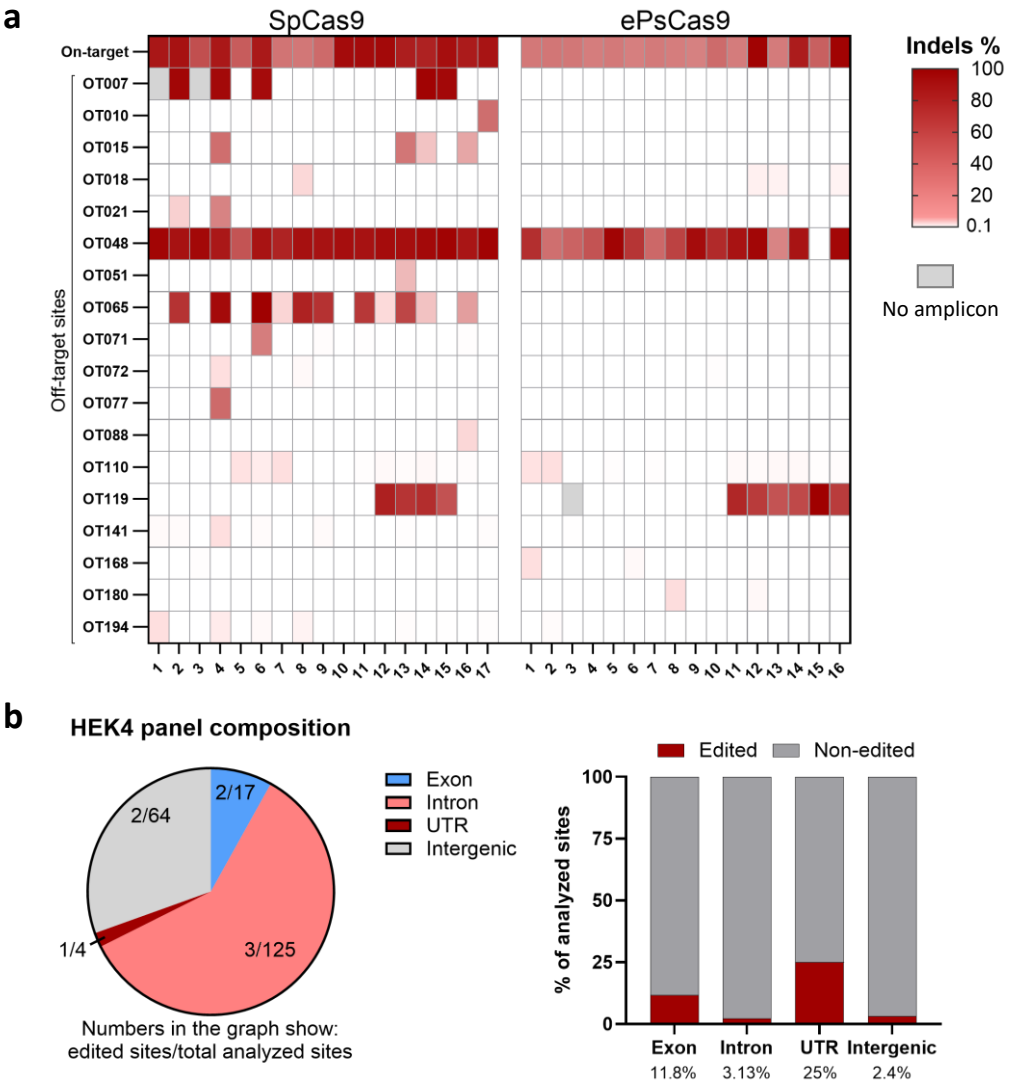
Supplementary Figure 2 – Changes in off-target editing and translocations in mESC pools over time. a) Heatmap showing the percentage of indels detected at the indicated promiscuous gP sgRNA on-/off-target sites in pools of bulk-edited mESCs. The pools show mean editing of n=3 replicate pools 2 days and 16 days after editing. All sites with >0.1% editing in at least one of the pools are shown. b) Dot plots showing total translocation burden in the gP-edited mESC pools over time, calculated as the sum of the frequencies of all detected translocations for each of the three individual cell pools on day 2 and day 16 after electroporation. % Total translocations is calculated as the sum of the frequencies of all detected translocations at a given time point. n=3 for each time point. One-tailed paired t-test, *p=0.0289. c) Circos plots showing translocations detected by PASTA in the mESC pools on day 2 and day 16, respectively. Red lines indicate translocations between the on-target and off-target sites, blue lines indicate translocations between off-target sites. Source data are provided as a Source Data file.

Supplementary Figure 3



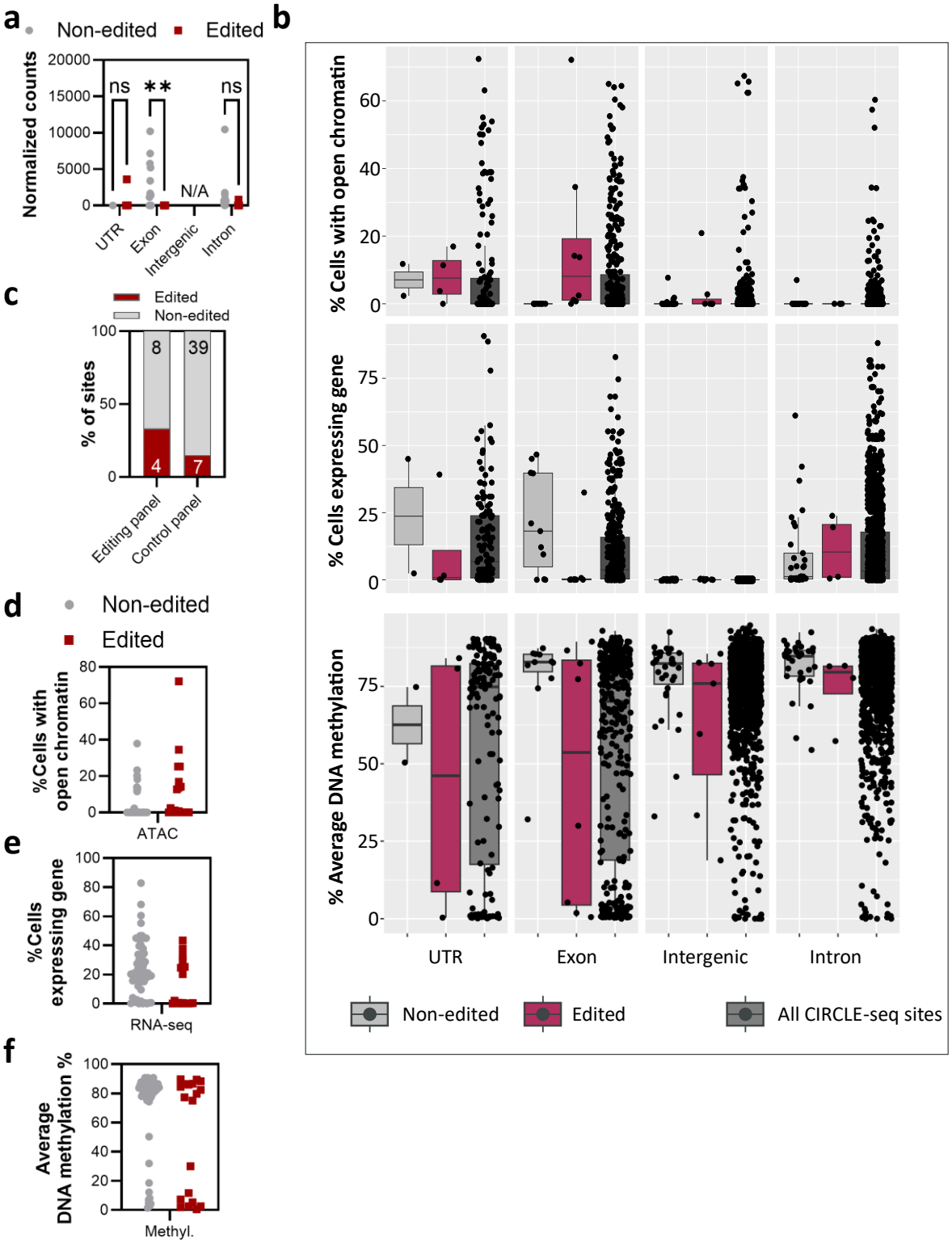
Supplementary Figure 3 – Whole-genome amplification of edited single cells.
a) Overview of the experimental workflow for single-cell editing analyses in mESCs using single-cell whole-genome amplification instead of clonal expansion. Created in BioRender. Madsen, A. (2026) https://BioRender.com/2zywinr. b) Inter-sample variation of site coverage after rhAmpSeq using WGA DNA as input (upper panel) vs. using genomic DNA (bottom panel). The dotted line indicates a coverage of 10,000x, which is needed to reliably detect mutations with a sensitivity of 0.1% of editing. Box limits show 25th-75th percentile, error bars show min to max, the line depicts the median. All individual data points are plotted. c) Heatmap showing the percentage of indels detected at the indicated promiscuous gP sgRNA on-/off-target sites in whole-genome amplified single-cell edited mESCs and a pool of bulk-edited mESCs. d) Large structural variants detected by whole-genome long-read sequencing of whole-genome amplified mESCs. Each data point represents a biological replicate. Source data are provided as a Source Data file.

Supplementary Figure 4



Supplementary Figure 4 – Off-target analysis in single cell-edited HSPCs.
a) Heatmap showing off-target editing in expanded single cell-edited HSPCs after editing with SpCas9 or ePsCas9 and the HEK4 sgRNA. Each column represents a single-cell clone. b) Overview of the HEK4 rhAmpSeq panel composition regarding coverage of exonic, intronic, UTR and intergenic regions, as well as the fraction of SpCas9-edited sites for each sub-group. Source data are provided as a Source Data file.

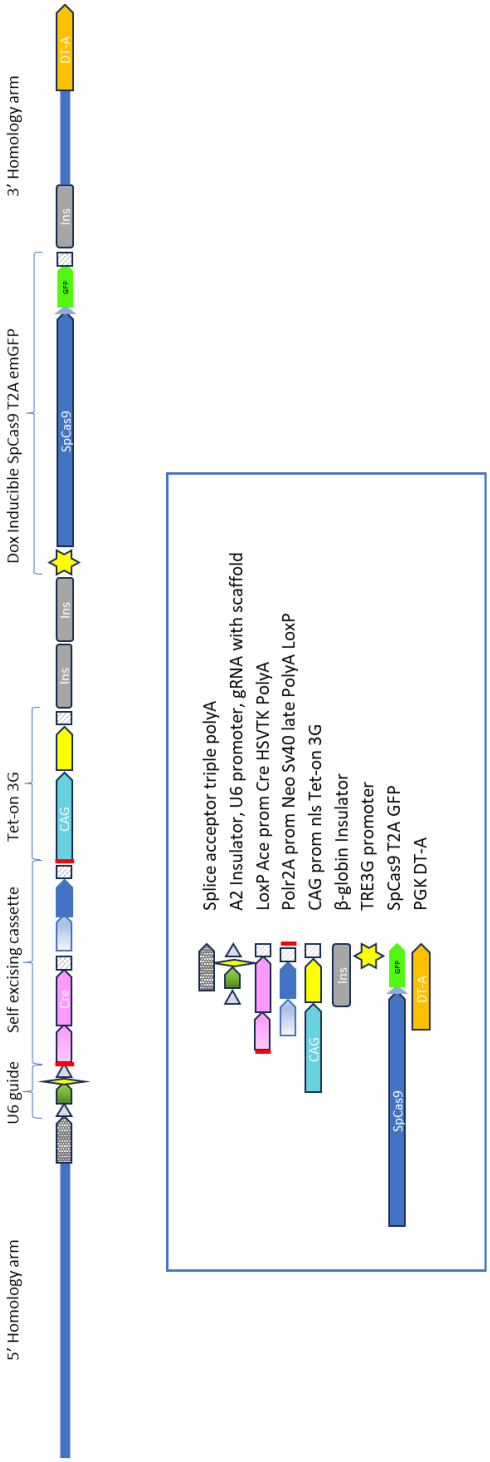
Supplementary Figure 5



Supplementary Figure 5 – Validation of observations from scATAC-seq/scRNA-seq experiments.
a) Bulk RNA-sequencing of mESCs to validate observed correlations between editing and gene expression for exonic sites from scRNA-seq results. n=5 (UTR), n=20 (exon), n=42 (intergenic), n=37 (intron). Two-way ANOVA plus Sidak's multiple comparisons test, **p=0.0079, ns = non-significant. b) Comparison of distribution of open chromatin (upper panel), gene expression (middle panel) and DNA

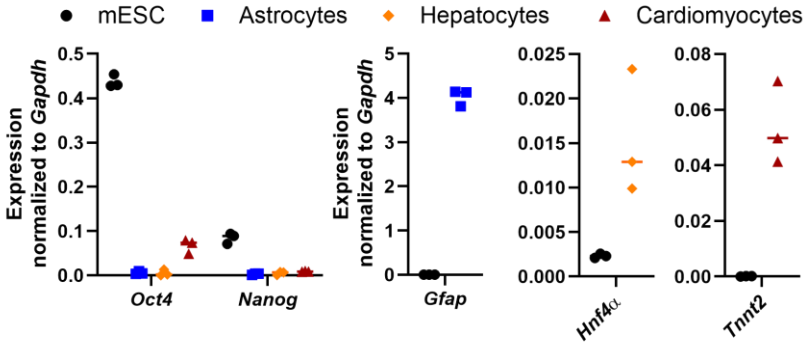
methylation (lower panel) in the analyzed sites vs. all CIRCLE-seq-identified sites. Box limits show 25th-75th percentile, error bars show 1.5x IQR range, the line depicts the median. All individual data points are plotted. c) Overview of gP validation rhAmpSeq panels. The “Editing panel” included sites that were considered likely to be edited according to the proposed prediction features, whereas the “Control panel” included sites unlikely to be edited. The graph shows the number of sites for both panels that showed editing or remained unedited. d) Analysis of scATAC-seq data from mock-electroporated mESCs based on functional region of the respective gP off-target site from the replication panels. Plot showing percentage of cells with open chromatin at sites showing editing vs. sites that remain non-edited. e) Analysis of scRNA-seq data from mock-electroporated mESCs based on functional region of the respective gP off-target site from the replication panels. Plot showing percentage of cells expressing the off-target-associated gene at sites showing editing vs. sites that remain non-edited. f) DNA methylation analysis from whole-genome bisulfite sequencing data[1, 2] at gP off-target sites from the replication panels showing editing vs. sites that remain non-edited. Plot showing the average methylation percentage of a region ±500 bp around each sgRNA off-target site. Source data are provided as a Source Data file.

Supplementary Figure 6



Supplementary Figure 6 – Schematic of the targeting vector "R26-U6-gP/gMH-Tet-on-SpCas9" that was used for generation of editing-inducible mESC lines.
The position and orientation of all functional elements within the vector are depicted. The U6 guide cassette contains the gP or gMH gRNA sequence, respectively.

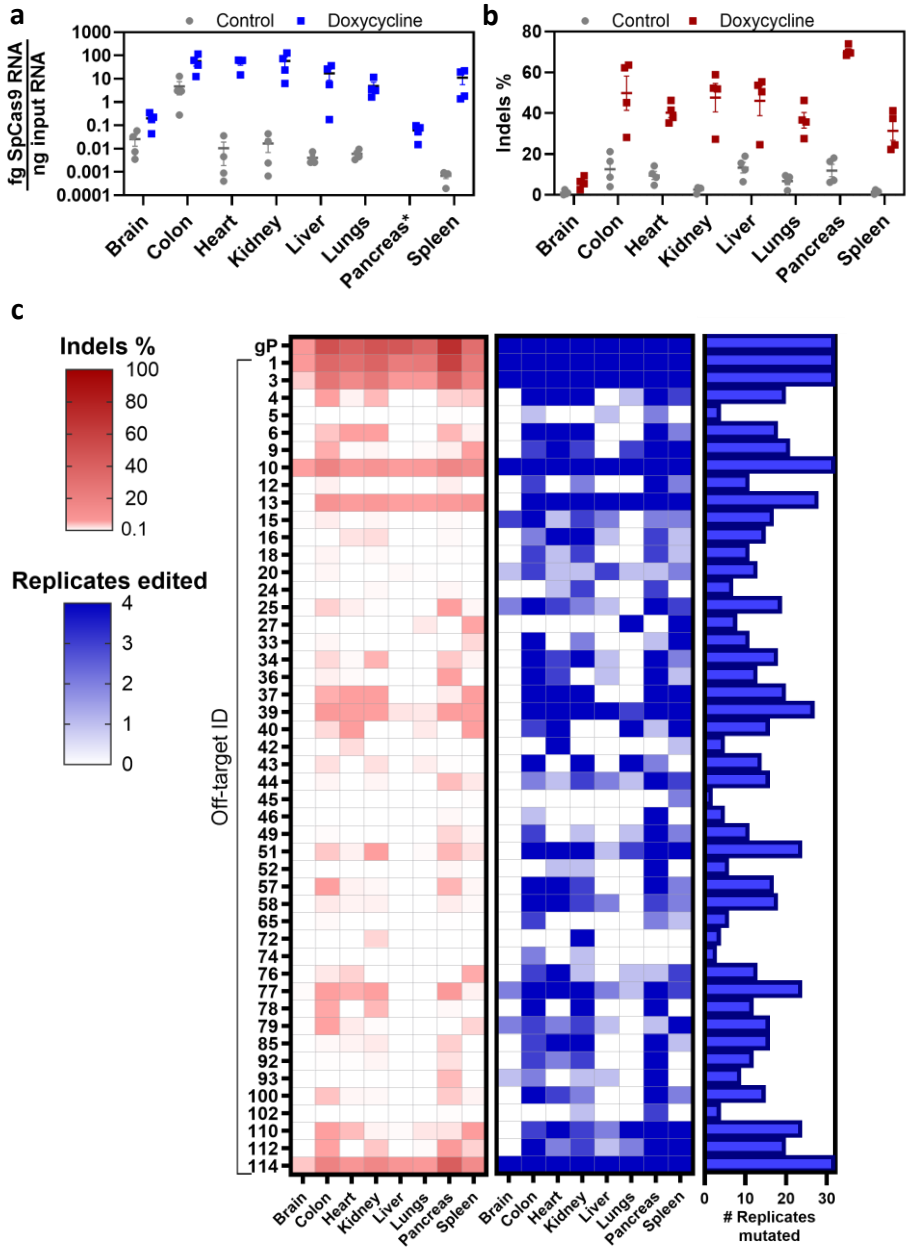
Supplementary Figure 7



Supplementary Figure 7 – Validation of successful mESC differentiation by cell marker expression analysis.

Differentiated mESCs showed a reduced expression of the pluripotency markers *Oct4* and *Nanog*, as well as increased expression of the respective cell type-specific markers, i.e. *Gfap* for astrocytes, *Hnf4a* for hepatocytes and *Tnnt2* for cardiomyocytes. Expression of all marker genes was normalized to the endogenous control gene *Gapdh*. Source data are provided as a Source Data file.

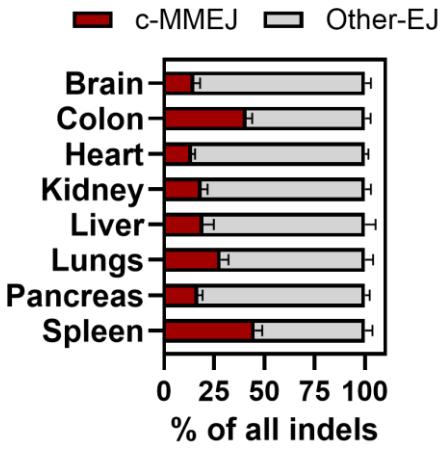
Supplementary Figure 8



Supplementary Figure 8 – *In vivo* gP editing.

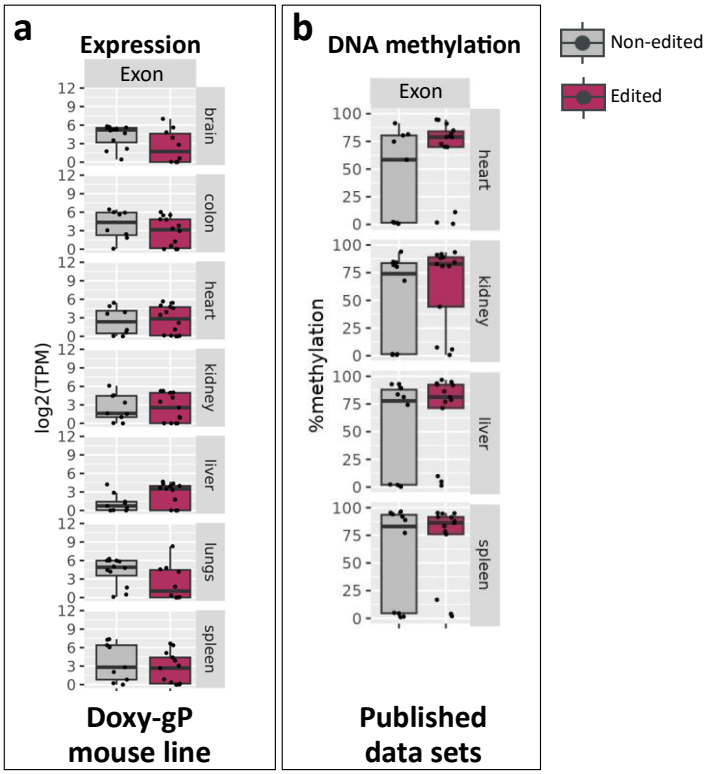
a) SpCas9 RNA expression in the different organs with or without doxycycline treatment. Graph shows mean±SEM SpCas9 RNA (fg) per ng input RNA for n=4 organ replicates. *Note that RNA quality for pancreas samples was very low compared to the other organs due to the high levels of endogenous RNases. b) Percentage of indels detected at the *Pcsk9* on-target site in the different organs with or without doxycycline treatment. Graph shows mean±SEM of editing % for n=4 organ replicates. c) Reproducibility of editing results within organ replicates in vivo. The first heat map shows the mean percentage of indels detected in n=4 organ replicates for sites that exhibit >0.1 % indels in at least one organ. The second heat map shows number of replicates (0-4) that exhibit >0.1% indels at the indicated sites within each organ group. The Waterfall plot on the right shows the total number of replicates edited across all groups. Source data are provided as a Source Data file.

Supplementary Figure 9



Supplementary Figure 9 – Comparison of DNA repair mechanisms in the different organs.
All sites that exhibit >0.1% indels in the respective organs upon doxycycline-induced SpCas9 expression with gP are shown. n=4 organ replicates. Bars show mean±SEM. c-MMEJ = micro-homology mediated end-joining; Other-EJ = other end-joining. Source data are provided as a Source Data file.

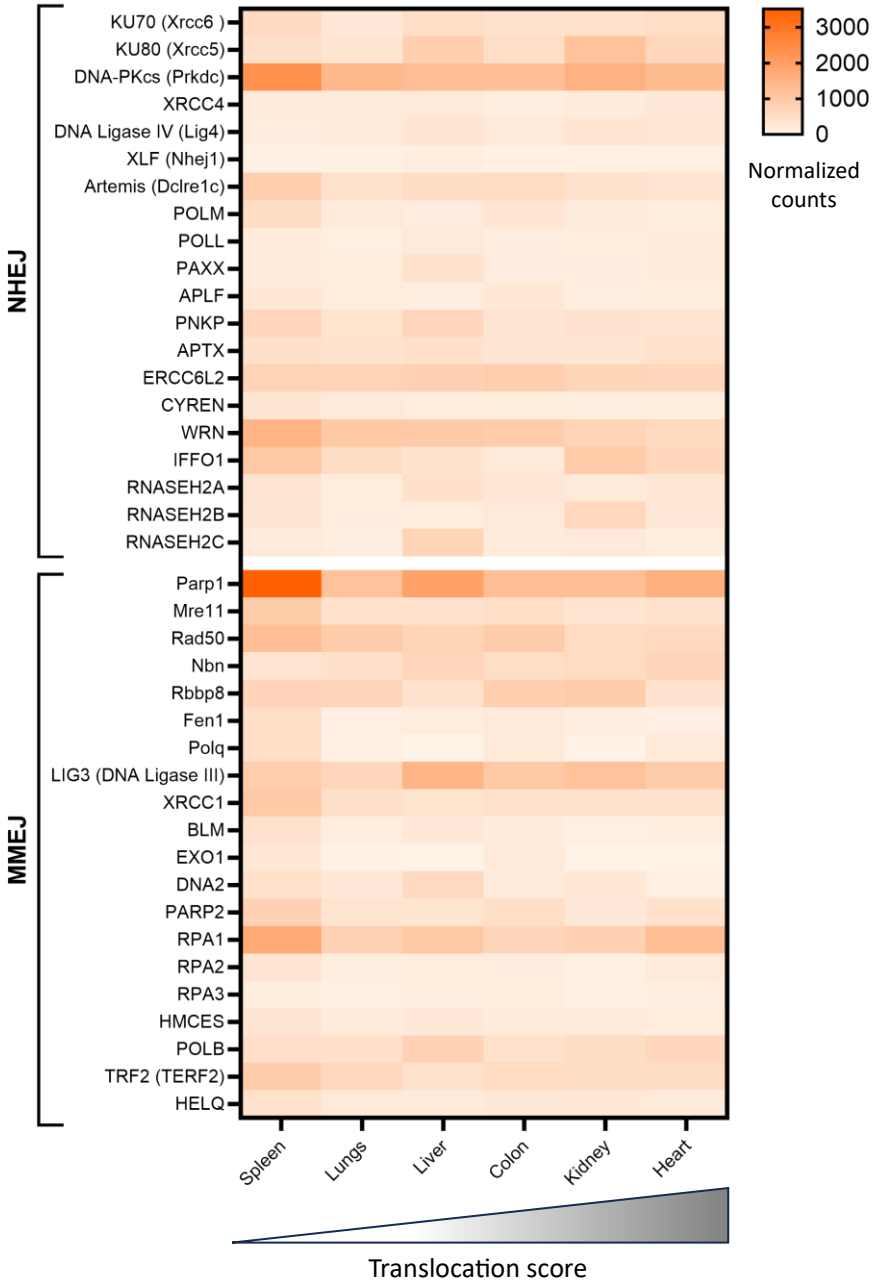
Supplementary Figure 10



Supplementary Figure 10 – Correlation of gene expression and DNA methylation in mouse tissues with the observed gP off-target editing after SpCas9 induction.

a) RNA-sequencing analysis of gene expression in the doxy-gP mouse line. Box limits show 25th-75th percentile, error bars show 1.5x IQR range, the line depicts the median. All individual data points are plotted. b) Analysis of publicly available DNA methylation data sets from an unrelated mouse line that were downloaded from the ENCODE portal (https://www.encodeproject.org). Box limits show 25th-75th percentile, error bars show 1.5x IQR range, the line depicts the median. All individual data points are plotted. Source data are provided as a Source Data file.

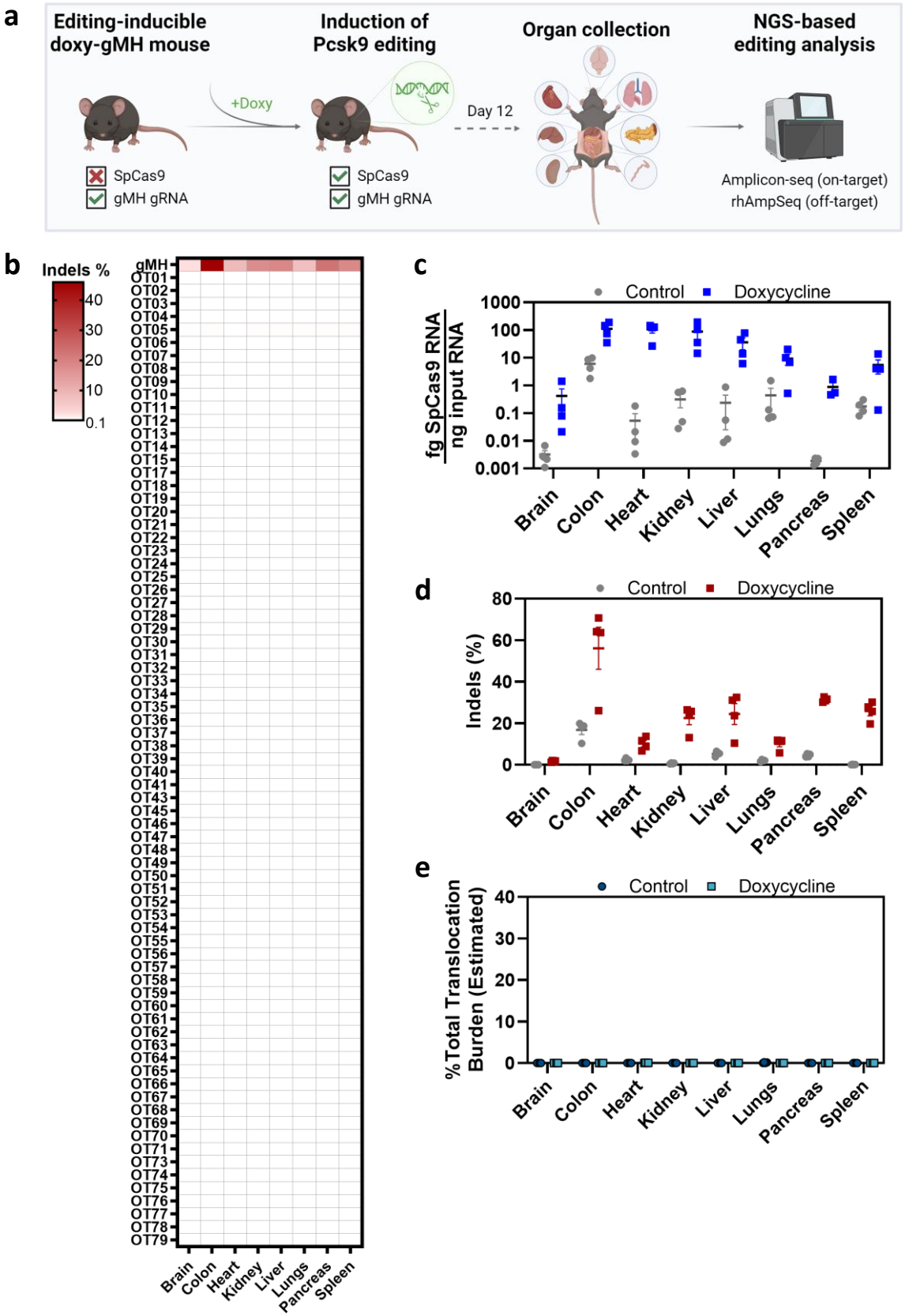
Supplementary Figure 11



Supplementary Figure 11 – Heatmap showing an overview of DNA repair gene expression in the different mouse organs.

The color scale shows the number of normalized read counts from RNA-sequencing. The mean of n=4 animals is shown for each organ. Organs on the x-axis are sorted according to their translocation score from lowest to highest. Source data are provided as a Source Data file.

Supplementary Figure 12



Supplementary Figure 12 – **Analysis of gMH-mediated editing outcome in different organs *in vivo*.** a) Overview of the *in vivo* editing experiments using a doxycycline-inducible SpCas9-gMH (doxy-gMH) mouse model. Created in BioRender. Madsen, A. (2026) https://BioRender.com/71p3urb. b) Heat map showing the mean percentage of indels detected in different organs upon editing induction by doxycycline treatment. n=4 organ replicates. c) SpCas9 RNA expression in the different organs with or without doxycycline treatment. Graph shows mean±SEM SpCas9 RNA (fg) per ng input RNA for n=4 organ replicates. d) Percentage of indels detected at the *Pcsk9* on-target site in the different organs with or without doxycycline treatment. Graph shows mean±SEM of indels % for n=4 organ replicates. e) Comparison of translocation burden in the different organs. Shown is the total translocation burden (i.e. sum of frequency of all detected translocations per organ). No significant occurrence of translocations was detected in the different organs after gMH editing. Source data are provided as a Source Data file.

Supplementary Tables

Supplementary Table 1 – sgRNA sequence information. All sgRNAs were purchased as synthetic sgRNAs from IDT.

Guide ID	Guide sequence 5'-3'
gMH	CAGGTTCCATGGGATGCTCT
gP	AGCAGCAGCGGCGGCAACAG
FANCF - SpCas9	GGAATCCCTTCTGCAGCACC
FANCF - ePsCas9	ATGGAATCCCTTCTGCAGCACC
CASGEVY	CTAACAGTTGCTTTTATCAC
HEK4 - SpCas9	GGCACTGCGGCTGGAGGTGG
HEK4 - ePsCas9	GTGGCACTGCGGCTGGAGGTGG

Supplementary Table 2 – Sex and age information of HSPC donors for HEK4, FANCF, and CASGEVY editing experiments.

Experiment_Donor #	Sex	Age
HEK4_D1	female	44 years
HEK4_D2	male	32 years
HEK4_D3	male	38 years
FANCF_D1	female	44 years
FANCF_D2	male	38 years
CASGEVY_D1	male	57 years
CASGEVY_D2	male	48 years
CASGEVY_D3	male	52 years

Supplementary Table 3 – Primer information for amplicon sequencing. All primers were ordered from IDT.

Target	Forward primer 5'-3'	Reverse primer 5'-3'
Pcsk9 - gP	TCATCAGCCAGGCCATCCTCCT	GTCCCAGGCGTCCATGTCCTTC
Pcsk9 - gMH	CTGTAGGCCCTGAAGTTGCCCC	AGTGCAGACTCTGGAGCCCTGA
HEK4	TGGAAGGAAGGGAGGAAGGG	CCTTTCAACCCGAACGGAGA

Supplementary Table 4 – ddPCR primer/probe information. Primers and probes were ordered from IDT.

Translocation	Forward primer 5'-3'	Reverse primer 5'-3'	Probe
3'OT001-5'OT114	GATCTCGGCTTGGTGTGA	CAGGTGGAGATGACACAAG	CCCAGCGGCCGACAGCGTTGTG
5'OT001-5'OT114	CACGTCGATTGGTACTCG	TAGCAAGTGGTTTGGCAG	CCGGGCCCGACACGGTGCTC

Supplementary Table 5 – Information on animal number, sex, and age for the *in vivo* editing experiments.

Genotype	Treatment	# Males	# Females	Age
R26-U6-gP-Tet-on-SpCas9	Control	2	2	8-9 weeks
R26-U6-gP-Tet-on-SpCas9	Doxycycline-treated	2	2	8-9 weeks

R26-U6-gMH-Tet-on-SpCas9	Control	2	2	8-9 weeks
R26-U6-gMH-Tet-on-SpCas9	Doxycycline-treated	2	2	8-9 weeks
Wild-type mice	Control	4	0	8-9 weeks
Wild-type mice	Doxycycline-treated	2	2	8-9 weeks

Supplementary Table 6 – Primer information for SYBR qPCR. Primers were ordered from IDT.

Gene	Forward primer 5'-3'	Reverse primer 5'-3'	Marker
mOct4	CAGCAGATCACTCACATCGCCA	GCCTCATACTCTTCTCGTTGGG	Pluripotency marker
mNanog	GAACGCCTCATCAATGCCTGCA	GAATCAGGGCTGCCTTGAAGAG	Pluripotency marker
mHnf4a	GGTCAAGCTACGAGGACAGC	ATGTACTTGGCCCACTCGAC	Hepatocytes
mGfap	ATCGAGATCGCCACCTACAG	CTCACATCACCACGTCCTTG	Astrocytes
mTnnt2	CAAGGAGCTGTGGCAGAGTA	TTCTGGTTGTCATTGATCCG	Cardiomyocytes
mGapdh	ACCACAGTCCATGCCATCAC	TCCACCACCCTGTTGCTGTA	Endogenous control
SpCas9	CGGAGATCAGTACGCAGACC	CATGCTAGCTGACAGAGGGG	Tet-ON induction

Supplementary information references

1. Lauria, A. et al. DNMT3B supports meso-endoderm differentiation from mouse embryonic stem cells. *Nat Commun* **14**, 367 (2023).
2. Liu, Y. et al. DNA methyltransferases are complementary in maintaining DNA methylation in embryonic stem cells. *iScience* **25**, 105003 (2022).