

Supplementary Information

Additional file 2: Supplementary figures

Small extrachromosomal circular DNA harboring targeted tumor suppressor gene mutations supports intratumor heterogeneity in mouse liver cancer induced by multiplexed CRISPR/Cas9

Tao Guo^{1,2,†}, Guo-Qiao Chen^{1,2,†}, Xu-Fan Li^{1,2,†}, Meng Wang^{1,2}, Kun-Ming Liu^{1,2}, Xiao-Ying Yang^{1,2}, Si-Cheng Liu^{1,2}, Yi-Li Feng^{1,2}, Peng-Yuan Liu^{1,2,*}, Hui Lin^{1,*}, and An-Yong Xie^{1,2,*}

¹ Innovation Center for Minimally Invasive Technique and Device, Department of General Surgery, Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, Hangzhou, Zhejiang 310019, P. R. China

² Institute of Translational Medicine, Zhejiang University School of Medicine and Zhejiang University Cancer Center, Hangzhou, Zhejiang 310029, P. R. China

* Corresponding authors:

An-Yong Xie, Ph.D., Institute of Translational Medicine, Zhejiang University School of Medicine, 268 Kai Xuan Rd, Hangzhou, Zhejiang 310029, China. Tel: +86 0571 86971680; Fax: +86 0571 88981576; Email: anyongxie@zju.edu.cn (lead contact).

Hui Lin, M.D., Department of General Surgery, Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, 3 East Qingchun Rd, Hangzhou, Zhejiang 310016, China. Email: 369369@zju.edu.cn.

Peng-Yuan Liu, Ph.D., Institute of Translational Medicine, Zhejiang University School of Medicine, 268 Kai Xuan Rd, Hangzhou, Zhejiang 310029, China. Email: pyliu@zju.edu.cn.

† These authors contributed equally to this work

Additional file 2: Supplement figures

Figure S1. Chromosomal locations of 34 targeted TSGs and the negative control target *Setd5* in mouse genome. Locations of TSGs and *Setd5* are indicated on mouse chromosomes. Color denotes cancer signaling pathways and TSGs associated.

Figure S2. Screening for single sgRNAs targeting 34 TSGs.

(A) Schematic of the gene editing reporter. The reporter was modified from the BGN reporter previously established and contains a *BsdR-GFP* fusion gene in which GFP is out of frame. *BsdR* encodes blasticidin S deaminase. The I-SceI and EcoRI sites in the *BsdR* portion are designed for insertion of target sequence used to test the efficiency of Cas9-sgRNA. After transfection of mammalian cells, expression of Cas9-sgRNA induces a site-specific DSB at its target in the reporter, and repair of this DSB by mutagenic NHEJ could reframe GFP, generating in-frame GFP and making cells GFP⁺. Thus, the frequency of Cas9-induced GFP⁺ cells reflects the efficiency of sgRNA.

(B) Percentages of Cas9-induced GFP⁺ cells with 3 different sgRNAs for each TSG target.

Figure S3. The information on CRISPR/Cas9-induced tumor nodules and single-cell clones derived from tumor modules in mice. Dosage per sgRNA used in 34-sgRNA library and days for tumor induction are indicated.

Figure S4. Representative target mutation oscillation of *Rb1* (A), *Lkb1* (B), *Arid1a* (C) and *Smad4* (D) between parental clone 6C7 and its subclones.

Figure S5. No stable expression of SpCas9 detectable in single-cell clones derived from CRISPR/Cas9-induced mouse primary liver tumors.

(A) Western blot of Flag-tagged *SpCas9* proteins stably expressed in single-cell clones derived from CRISPR/Cas9-induced mouse primary liver tumors. Flag-tagged *SpCas9* proteins in the control NIH3T3 cells transiently transfected with expression plasmids was used as a positive control.

(B,C) Detection of stable *SpCas9* activity in single-cell clones 1C3-1 (B) and several other single-cell clones (C) by deep sequencing. Single-cell clones were transfected with the expression plasmid for *Col1a1* sgRNA or *Rosa26* sgRNA together with (+)

or without (–) the *SpCas9* expression plasmid. gDNA was isolated from cells at 72 h post transfection and the editing frequency was determined by targeted PCR amplicon deep sequencing.

Figure S6. Oscillation of target site mutations during proliferation of single-cell subclone 1C3-1, 1C3-2, 1C3-3 and 1C3-4 derived from 1C3. Changes in target site mutation profiles of 1C3-1, 1C3-2, 1C3-3 and 1C3-4 at indicated time points of cell proliferation (i.e., Day 0, 15 and 30). Gene targets and their mutation types are shown on top and at bottom, respectively. The numbers flanking Del or Ins indicated at bottom are the numbers of deleted or inserted nucleotides. $\sqrt{\text{Frequency}}$ representing square root of mutation frequencies is indicated from the highest 1 in dark green to the lowest 0 in white. The top line chart and the boxplot under show the number of single-cell clones with a specific mutation type, i.e., Number of clones, and the percentage distribution of a specific mutation type in all single-cell clones, i.e., Proportion (%), respectively.

Figure S7. Purification of eccDNAs from 1C3-1 and 6C7.

(A) Experimental outline of eccDNA purification and identification by Circle-Seq.

(B) Validation of linear gDNA removal from eccDNAs by targeted PCR amplification of *Actb* in cell lines indicated. After gDNA was isolated, Plasmid-Safe ATP-Dependent DNase was added to digest linear gDNA and eccDNA was subsequently purified. gDNA and eccDNA purified for PCR were indicated by “–” and “+” DNase, respectively. PCR products were stained with ethidium bromide after DNA gel electrophoresis. M: DNA marker.

(C,D) Removal of linear gDNA and mitochondrial DNA from eccDNAs validated by targeted PCR amplification of the nuclear gene *Cox5b* (C) and the mitochondrial gene *mt-Co1* (D) in the control NIH-3T3 cells, 6C7 and 1C3-1 cells as indicated. After crude eccDNAs were isolated by column chromatography, PacI and Plasmid-Safe ATP-Dependent DNase were added to digest mitochondrial DNA and linear gDNA, respectively, allowing further purification of eccDNAs. PCR products of either gDNA, crude eccDNA or eccDNA indicated for nuclear gene *Cox5b* and mitochondrial gene *mt-Co1* were stained with SYBR Gold stain after DNA gel electrophoresis. M: DNA marker.

Figure S8. Identification and mapping of eccDNAs by Circle-Seq in 1C3-1 and 6C7.

(A) Chromosomal distribution of eccDNAs from NIH-3T3, 1C3-1 and 6C7 cells. The target gene sites by 35 sgRNAs for 34 TSGs and *Setd6* are depicted in red lines on 21 chromosomes. Mapping of eccDNAs from each cell line to 21 chromosomes is also indicated.

(B-D) Identification of eccDNAs carrying genes targeted by 35 sgRNAs in NIH-3T3 (B), 1C3-1 (C) and 6C7 cells (D). The overlap of gene regions between eccDNAs and genes targeted by 35 sgRNAs were determined by Venn diagrams.

Figure S9. Analysis of *Apob*-2119 eccDNA and its circularization junction.

(A) Schematic for analysis of *Apob*-2119 eccDNA and its circularization junction by Sanger sequencing. PCR amplification of *Apob*-2119 eccDNA and its circularization junction was performed with two pairs of outward primers (i.e., F1/R1 and F1/R2) and one pair of inward primers (i.e., F2/R2) as indicated, respectively. PCR products were cloned into the pUC19 vector for Sanger sequencing.

(B) PCR products of *Apob*-2119 eccDNA from NIH-3T3, 1C3-1 and 6C7 were analyzed by DNA gel electrophoresis. M: DNA marker.

Figure S10. Detection of eccDNA containing target site mutations in single-cell clones.

(A) Experimental outline of eccDNA purification and detection. gDNA was extracted from single-cell clones and digested with Plasmid-Safe ATP-Dependent DNase to enrich eccDNA. Targeted PCR amplicon deep sequencing was performed on gDNA and eccDNA respectively to identify target site mutations in gDNA or eccDNA.

(B) Read counts of 34 TSG targets and the *Setd5* target in eccDNA by targeted PCR amplicon deep sequencing from six single-cell clones.

(C) Ratio of total reads of each target in gDNA or in eccDNA to all target reads. After PCR amplification, the equal amounts of PCR products of each locus with two different templates were prepared for deep sequencing. All target reads were counted firstly, ratio of reads of each target to all target reads were calculated.

(D) Relative read ratio of eccDNA to gDNA. In order to exhibit clearly, we used relative read ratio of eccDNA to gDNA to describe the change of deep sequencing results those targets that over 1 indicated eccDNA existence.

Figure S11. Circular barplot for target site mutations of 35 targets in gDNA and eccDNA in representative single-cell clone 1C3-2 (A), 1C3-3 (B), 6C7-2 (C), 6C7-4 (D). The column in the inner circle for each target gene were relative read ratios of eccDNA to gDNA for each target site from deep sequencing data. The stacked barplot in the outer circle represents the mutation pattern at indicated target sites in eccDNA and gDNA. Mutation types are shown in color. Total read counts of eccDNA and gDNA are listed. *, invalid read counts at indicated sites, which are defined as <0.1% of total read counts from eccDNA or gDNA.

Figure S12. Identification of circle junction for target eccDNA.

(A) Schematic of target eccDNA formation and circularization junction analysis.

Extrachromosomal linear DNA fragments were generated and circularized to form eccDNA. PCR amplification of circularization junction was performed with a pair of outward primers and PCR products cloned into the pUC19 vector for Sanger sequencing.

(B) Original target sequence for eccDNA formation containing *Atm*, *Rb1* or *Kmt2c* target sites. > and < indicate the site for circularization. Exons are shown in red with PAM in bold. Sequences underlined are outward primer pair for PCR of circularization junctions.

Figure S1

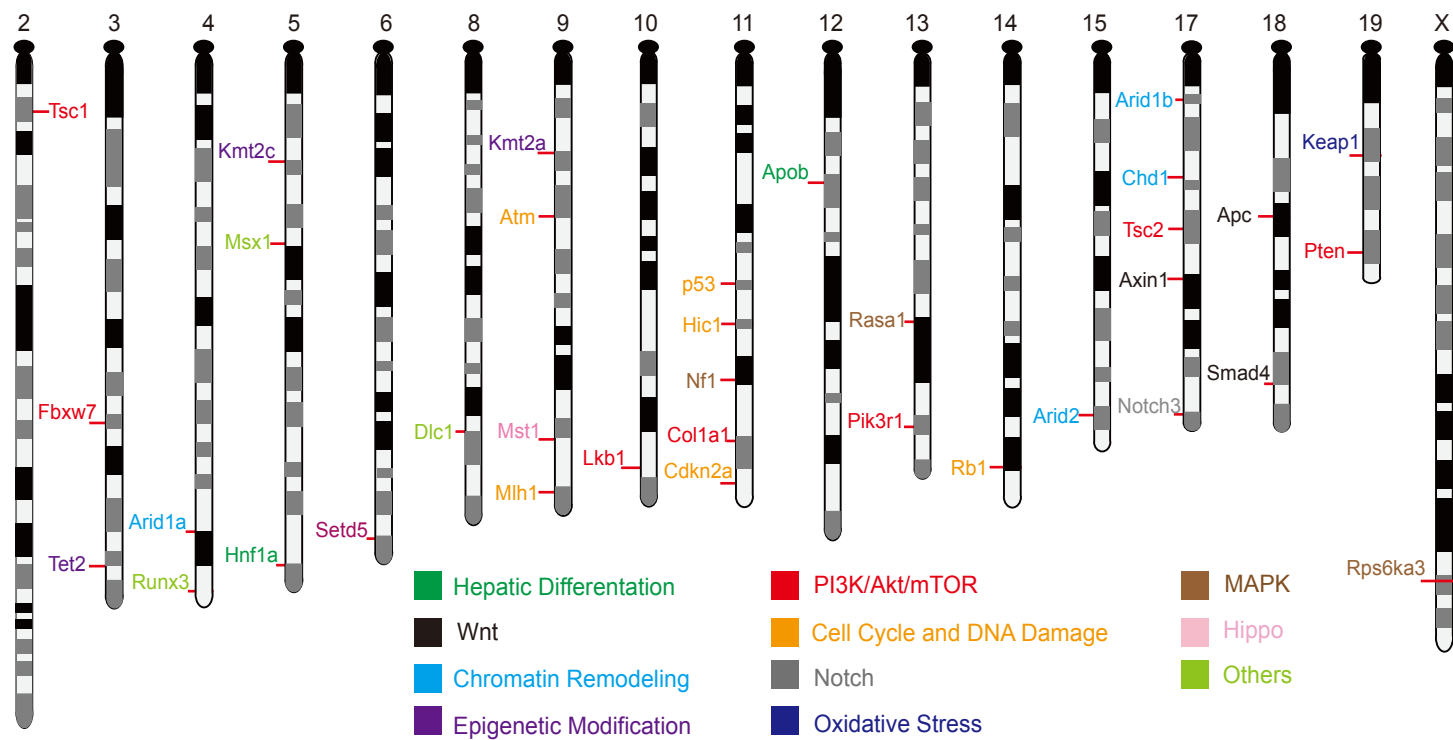
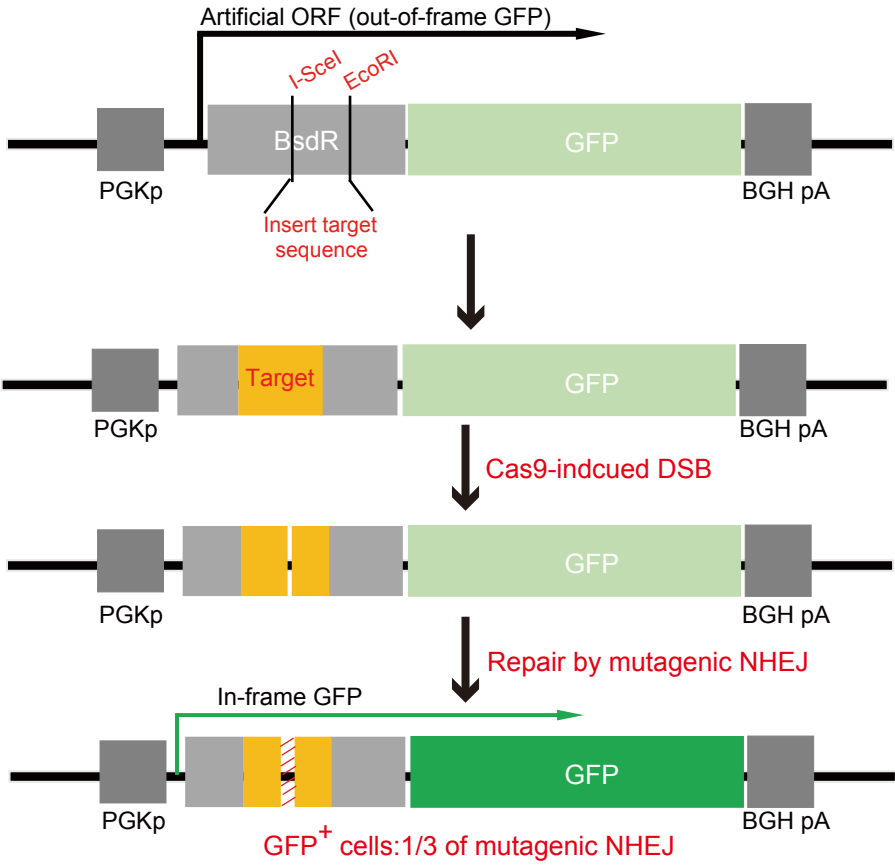


Figure S2

A



B

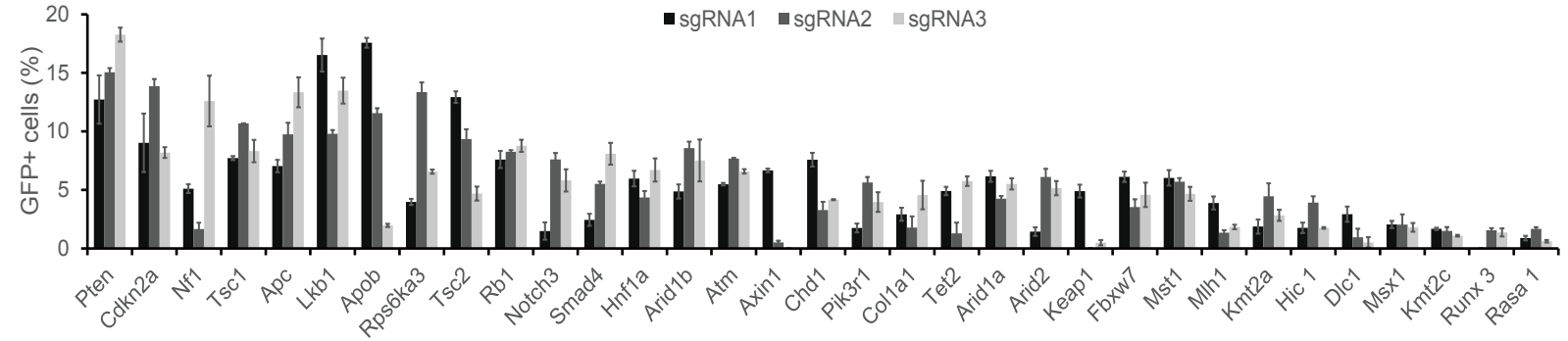


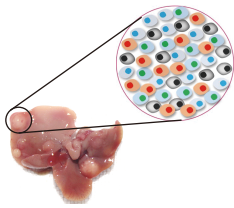
Figure S3

Tumor induction:

Dosage per
sgRNA/Days



Mouse



Tumor nodules



Single-cell clones

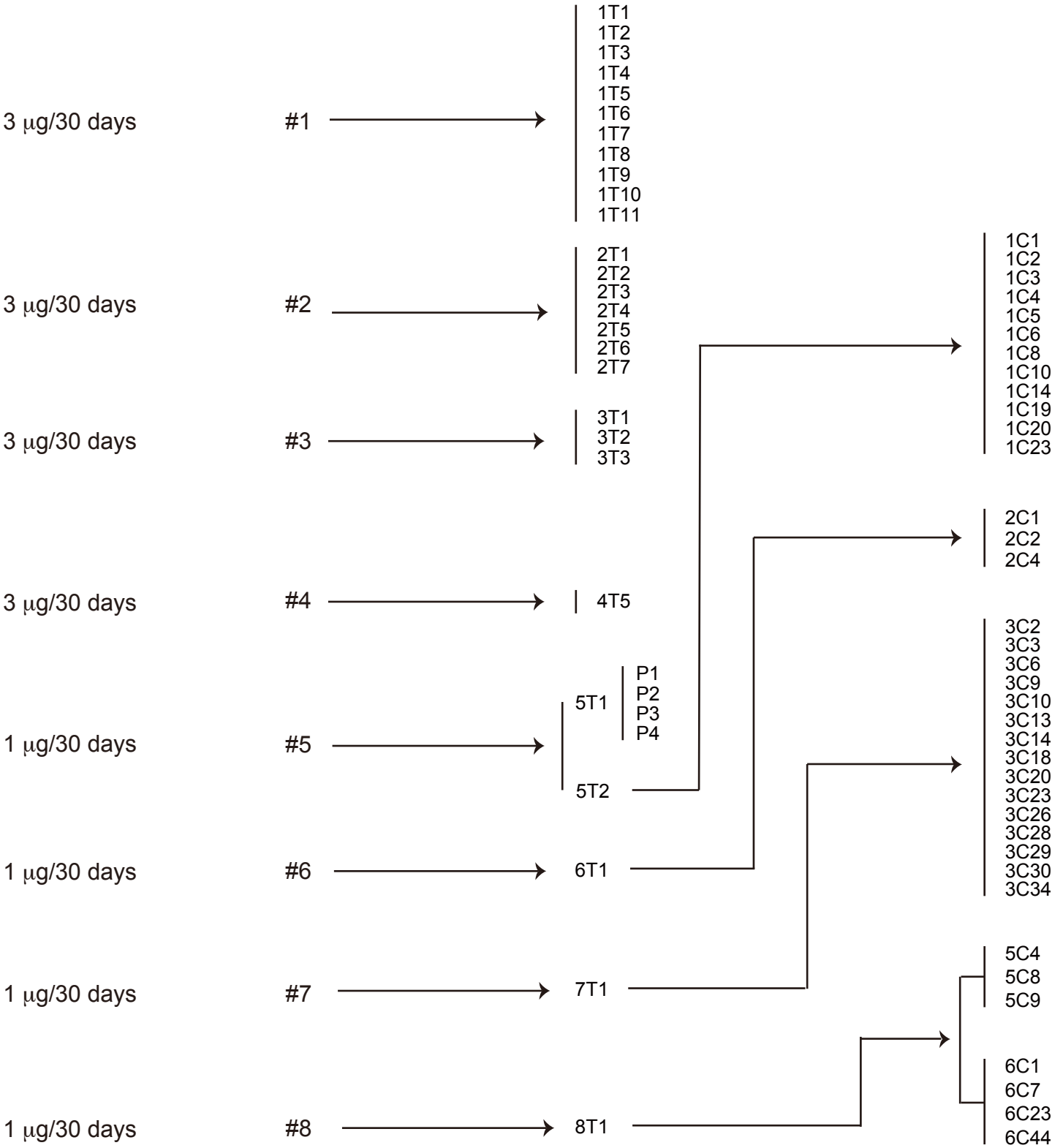


Figure S4

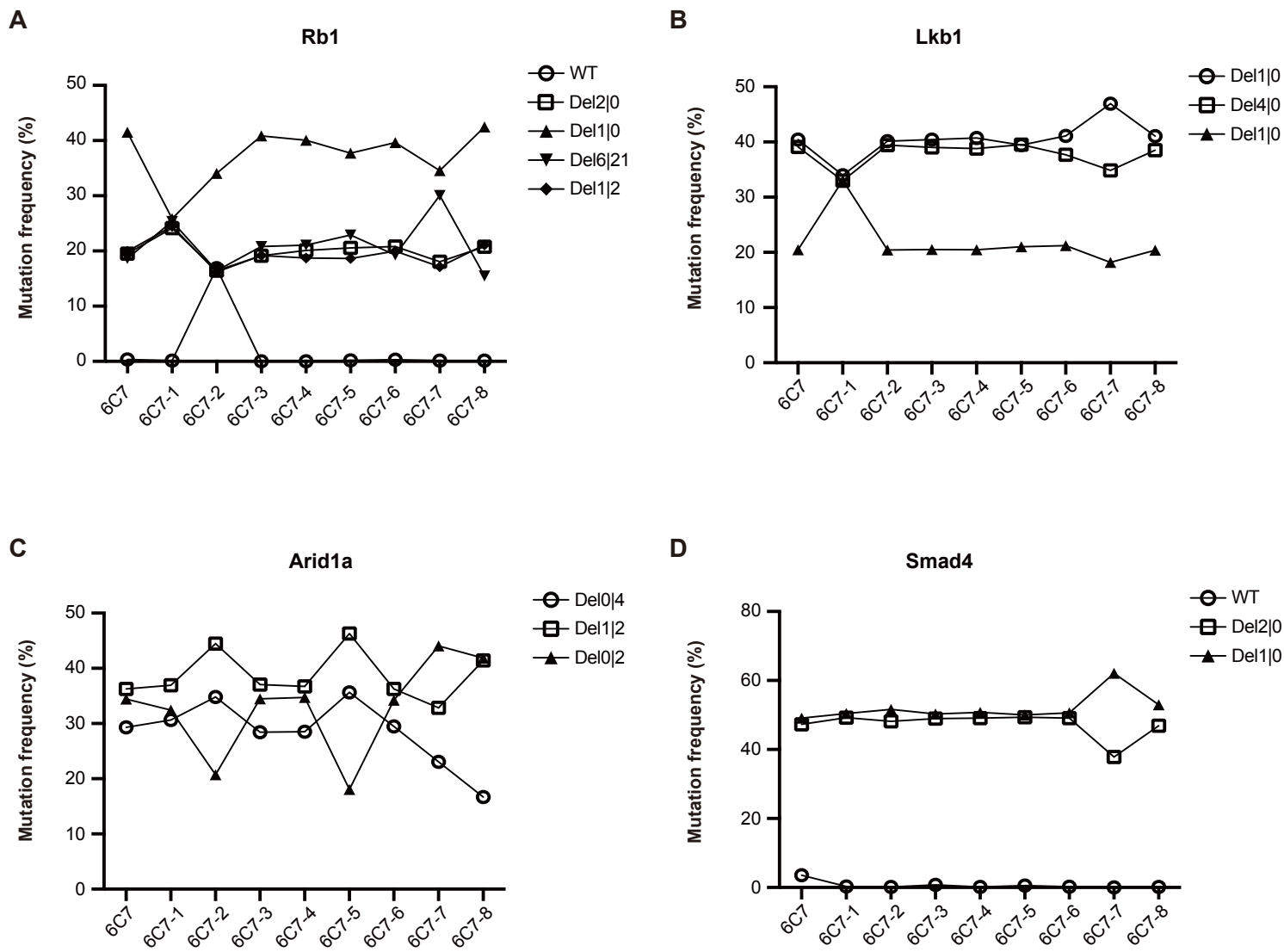
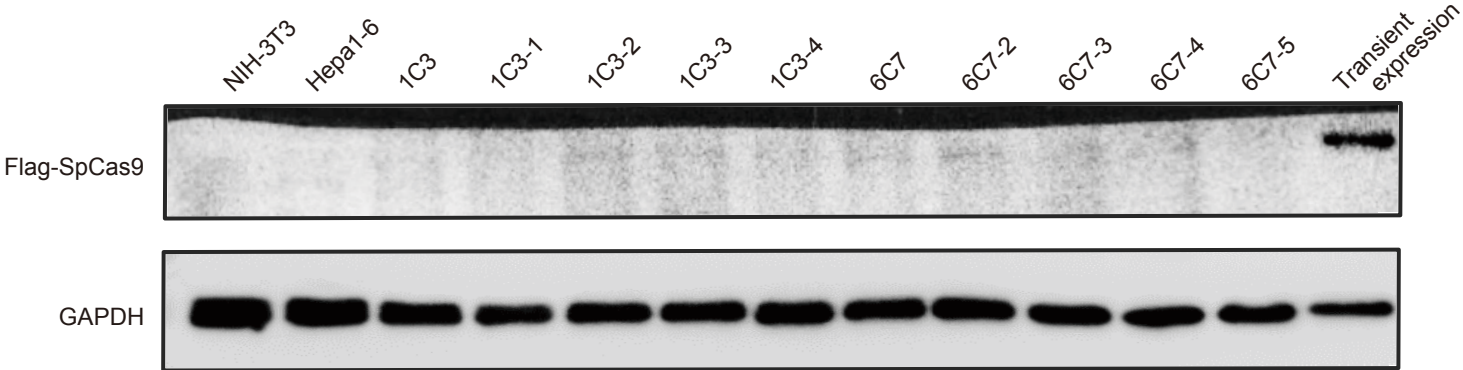
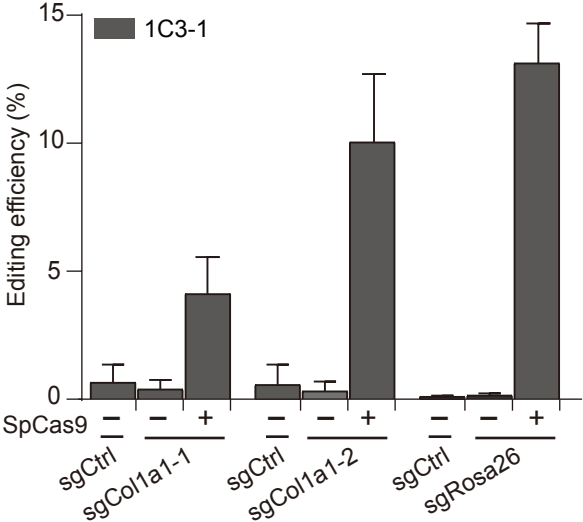


Figure S5

A



B



C

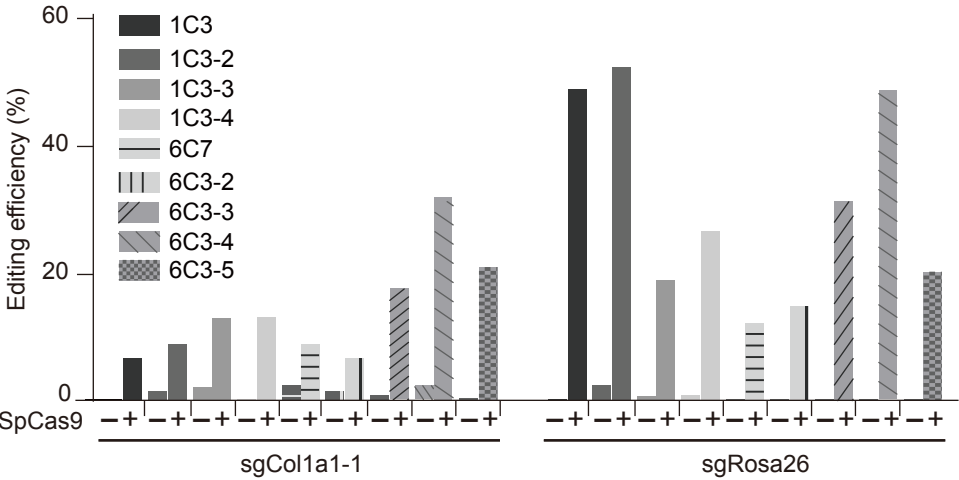
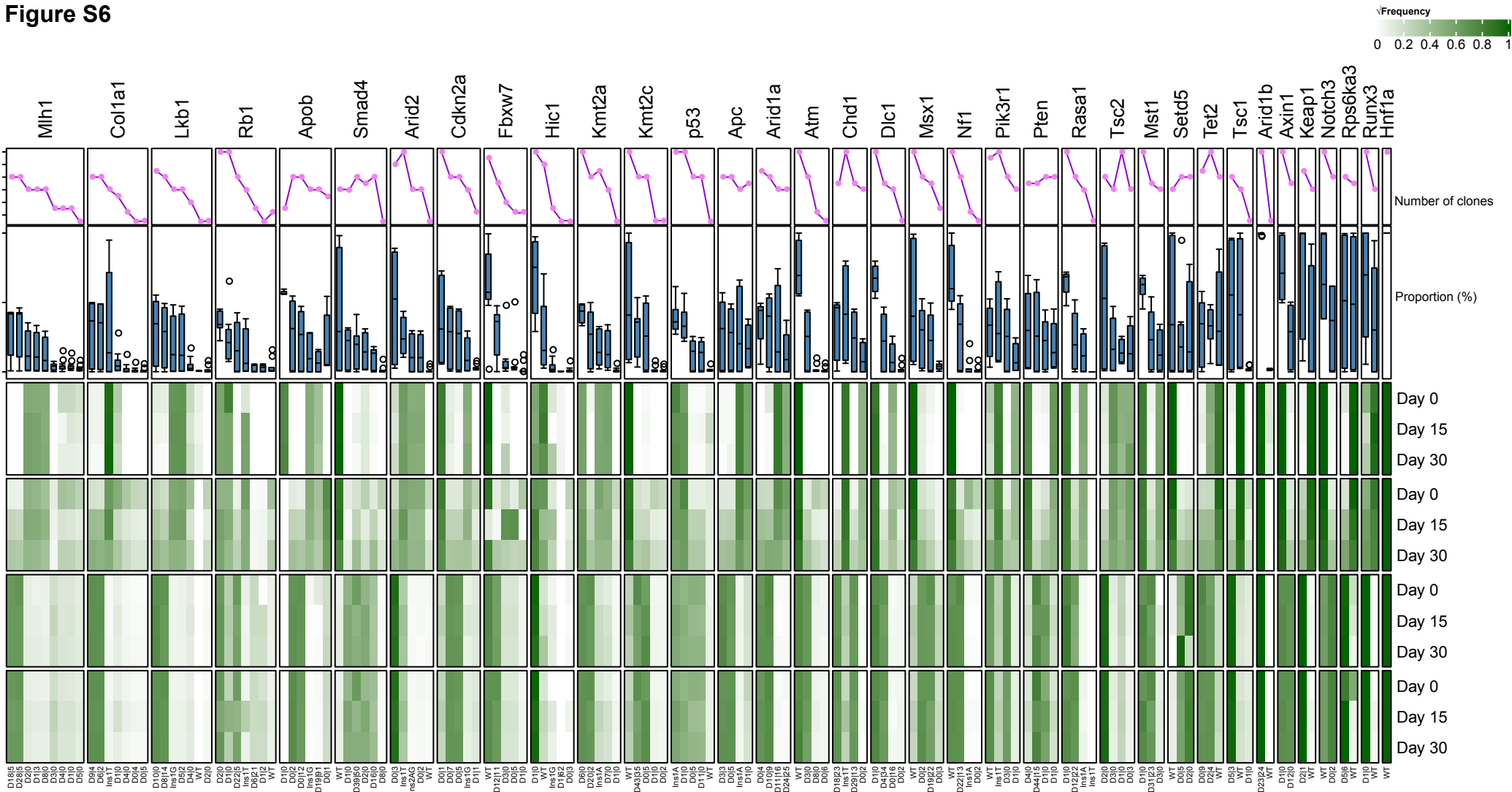
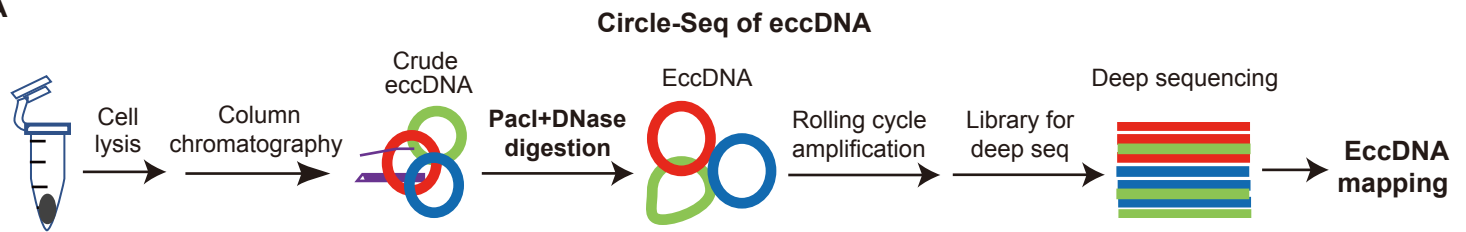


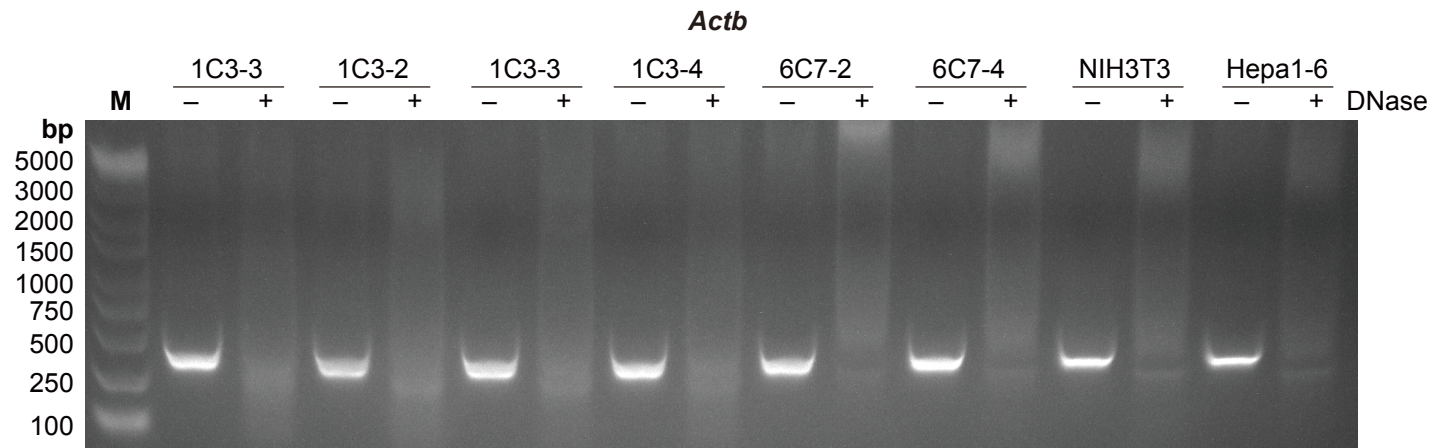
Figure S6



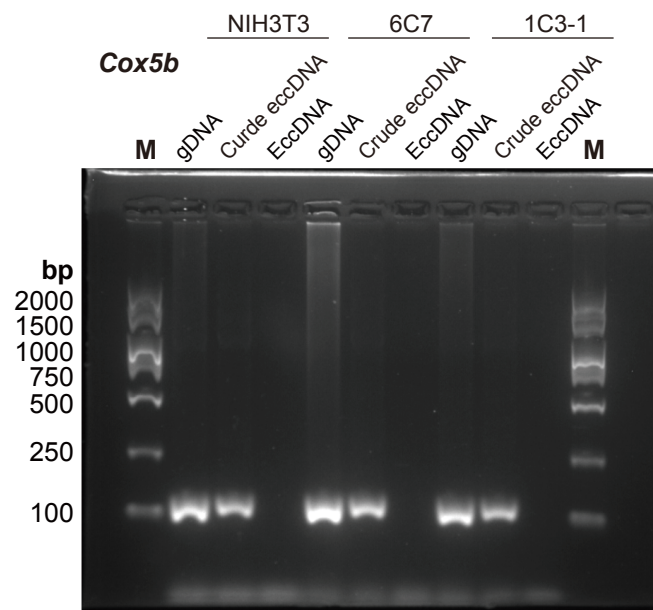
A



B



C



D

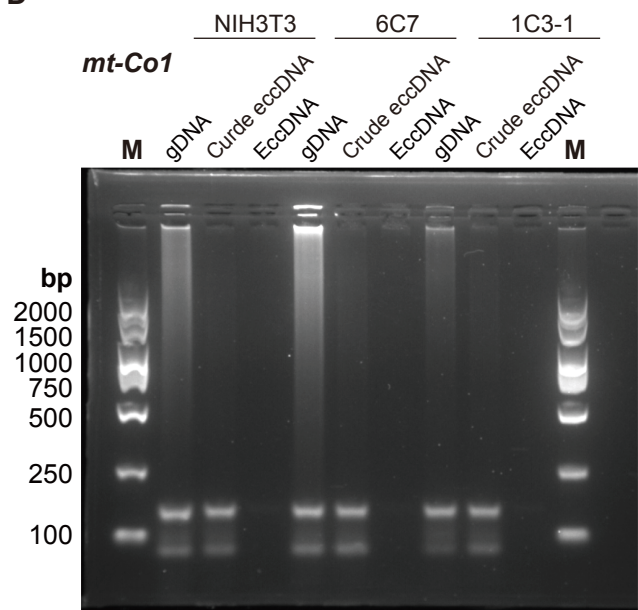
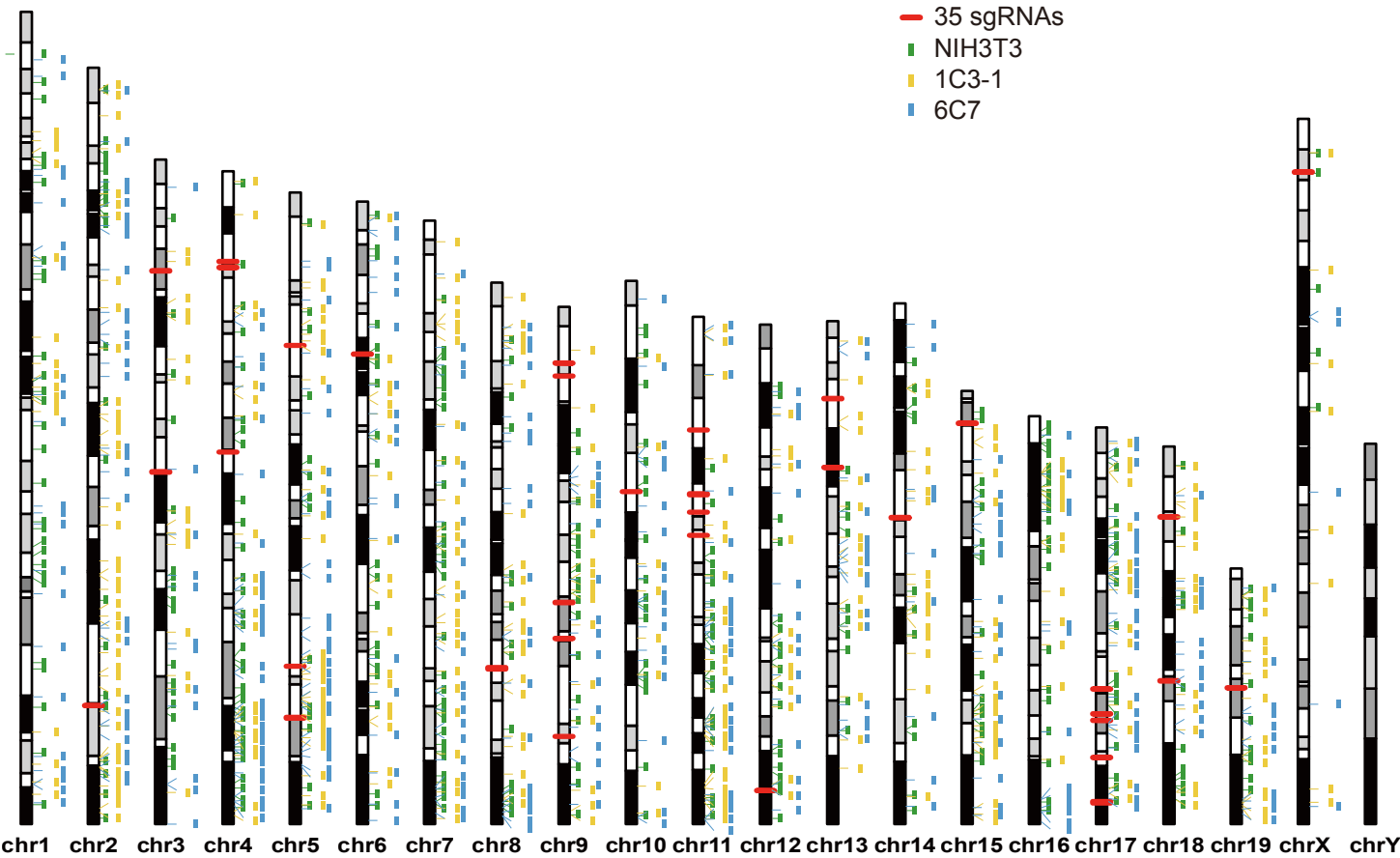
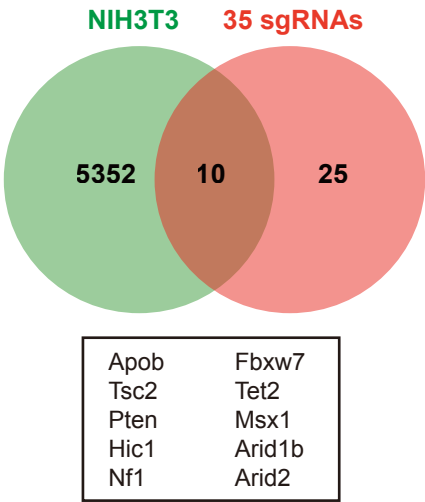


Figure S8

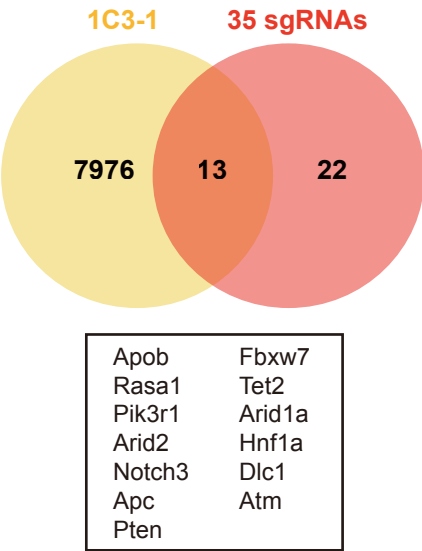
A



B



C



D

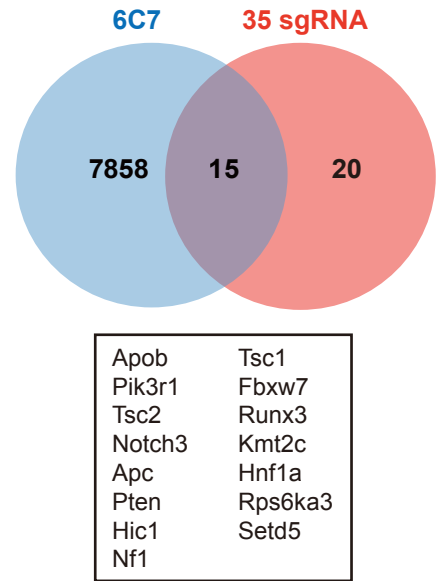
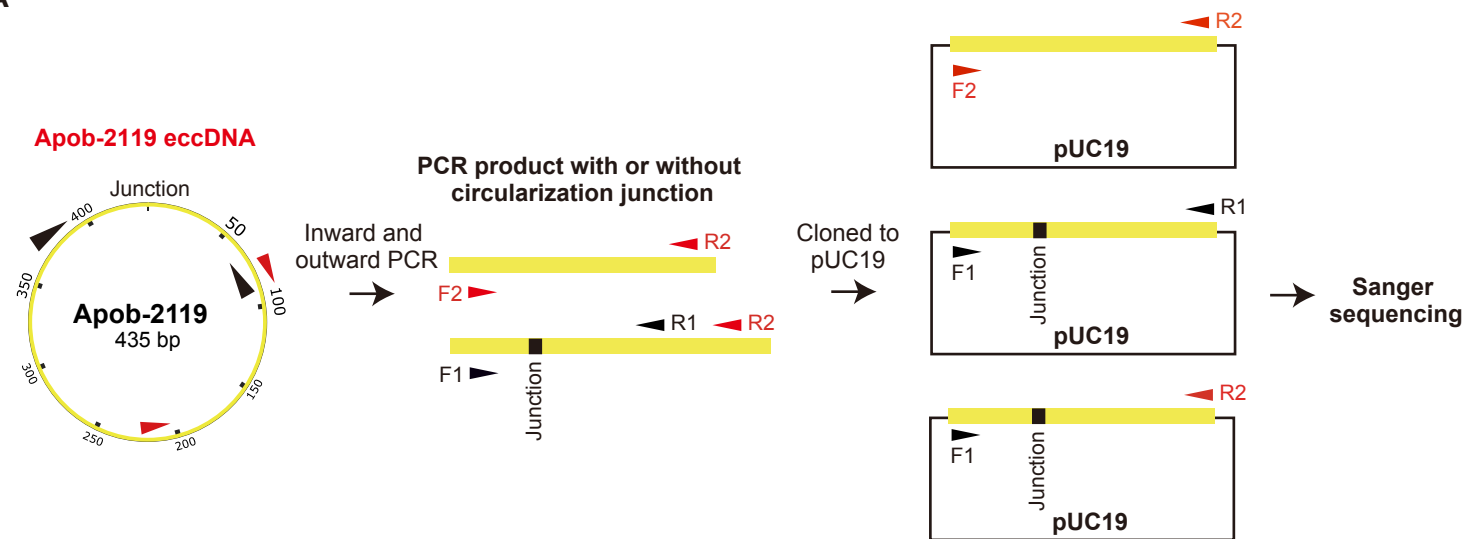


Figure S9

A



B

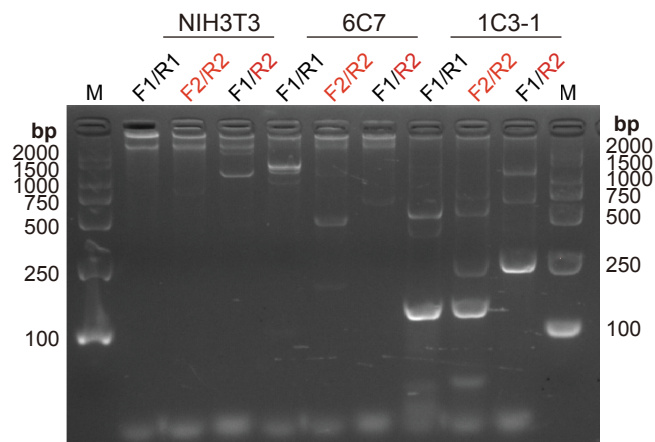
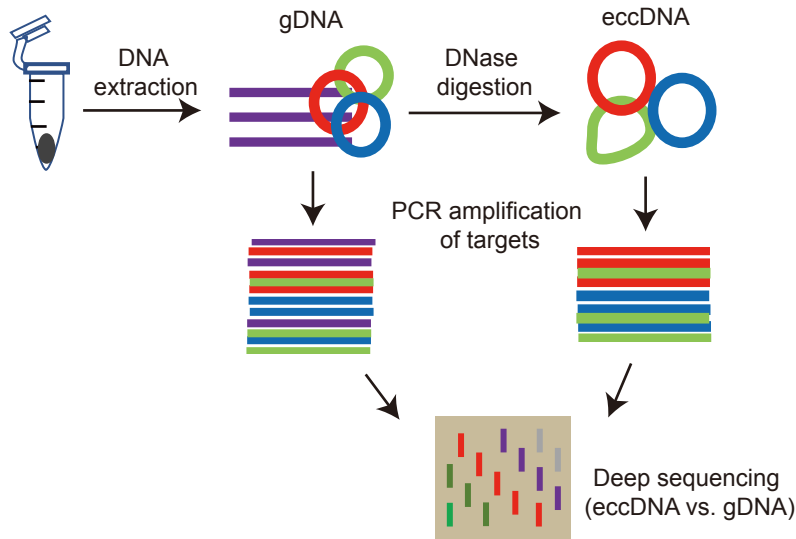
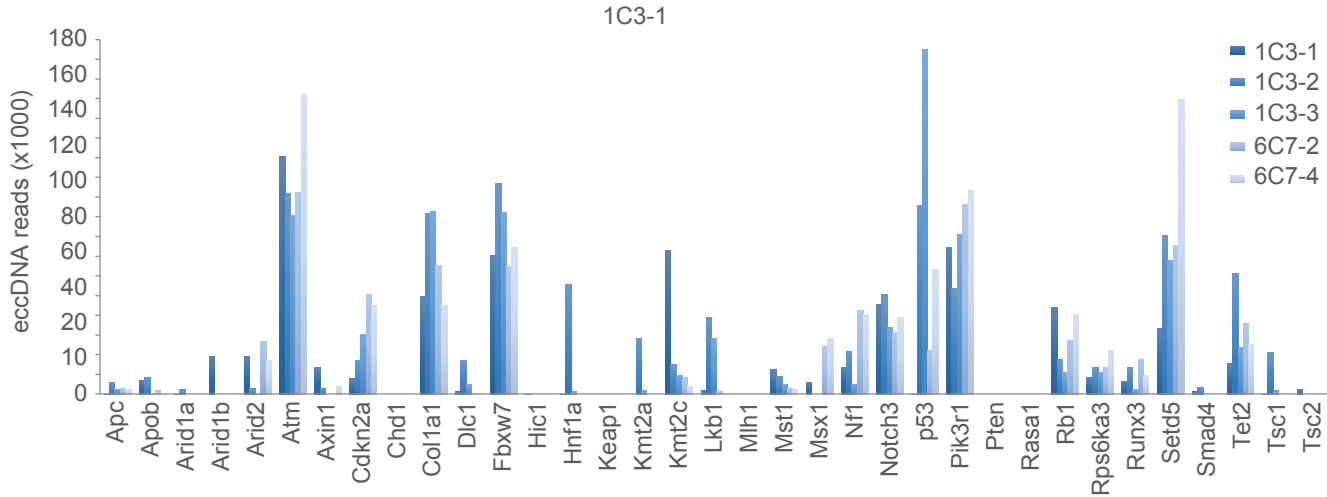


Figure S10

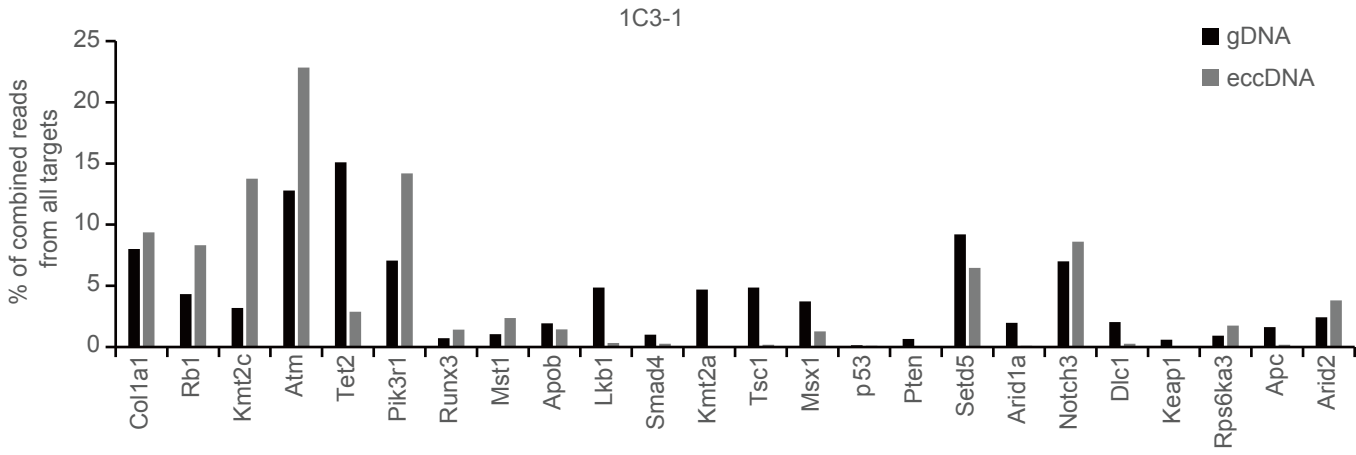
A



B



C



D

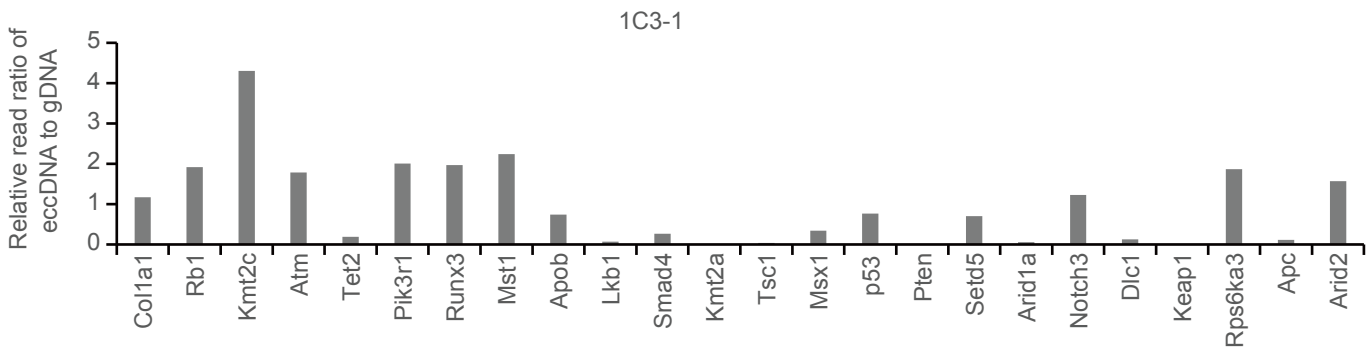
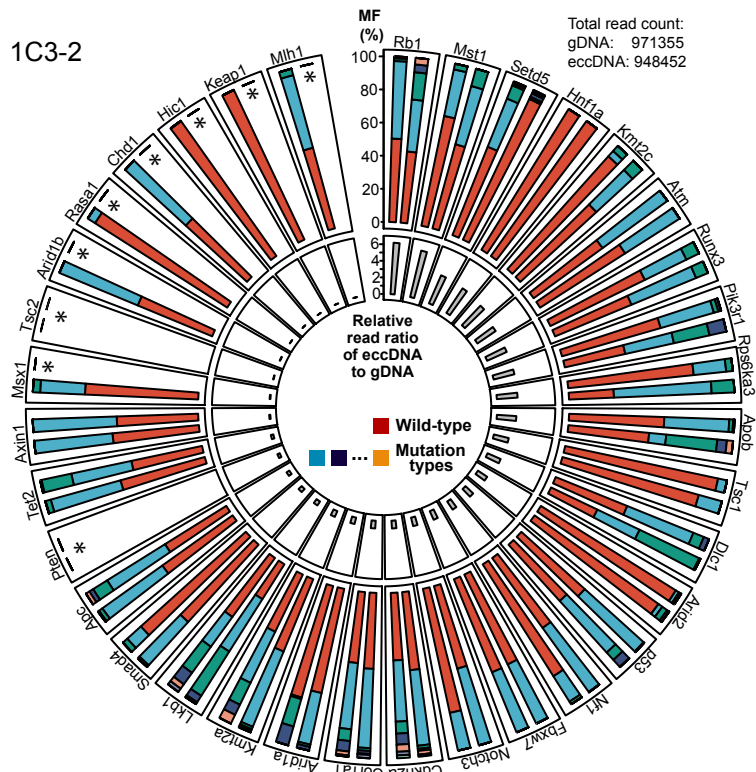
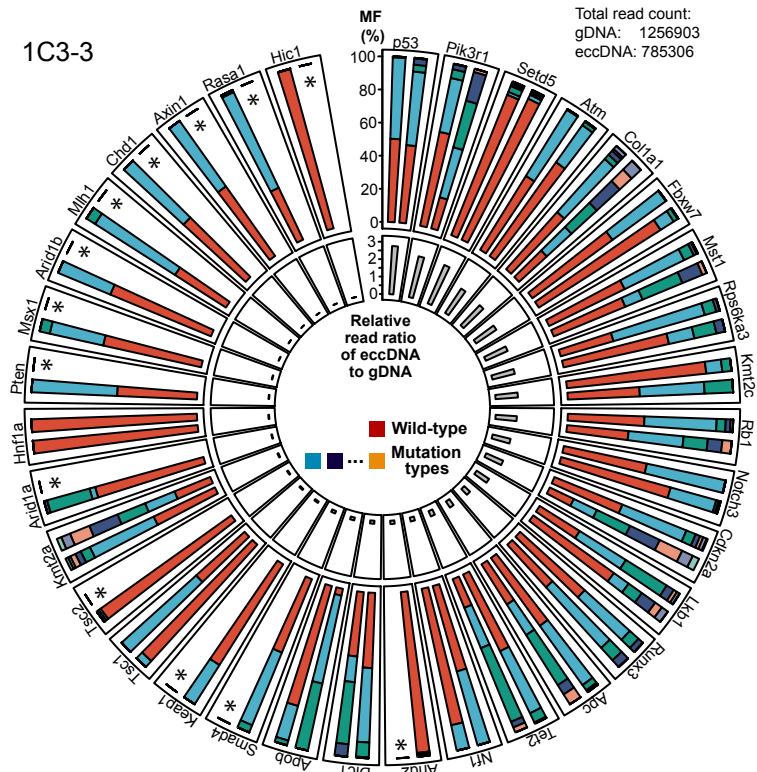


Figure S11

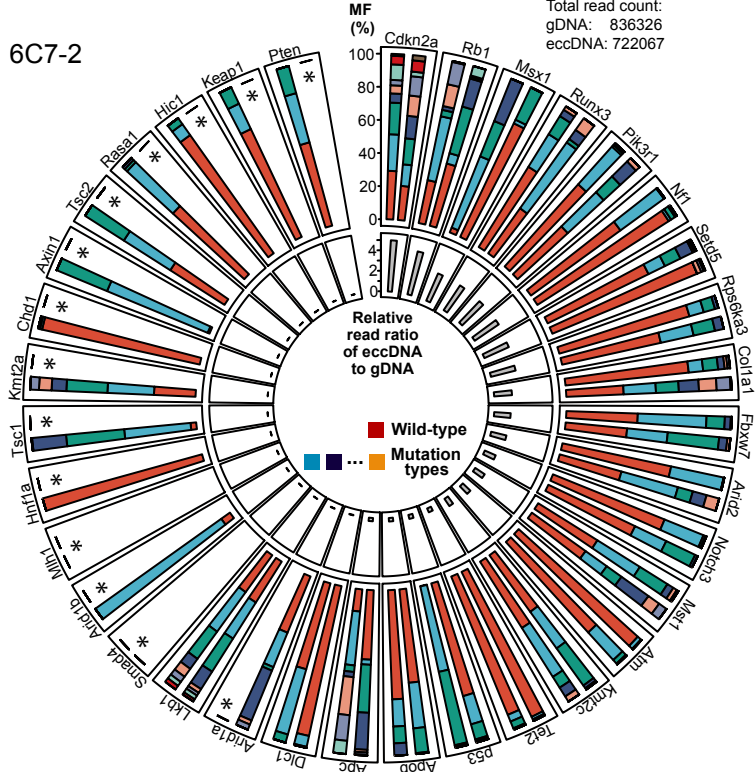
A 1C3-2



B 1C3-3



C 6C7-2



D 6C7-4

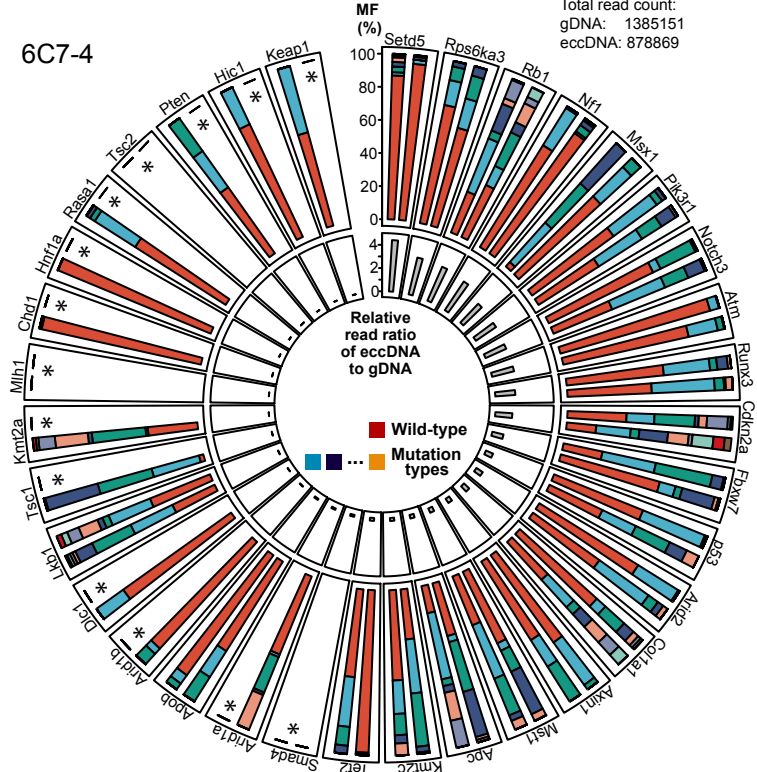
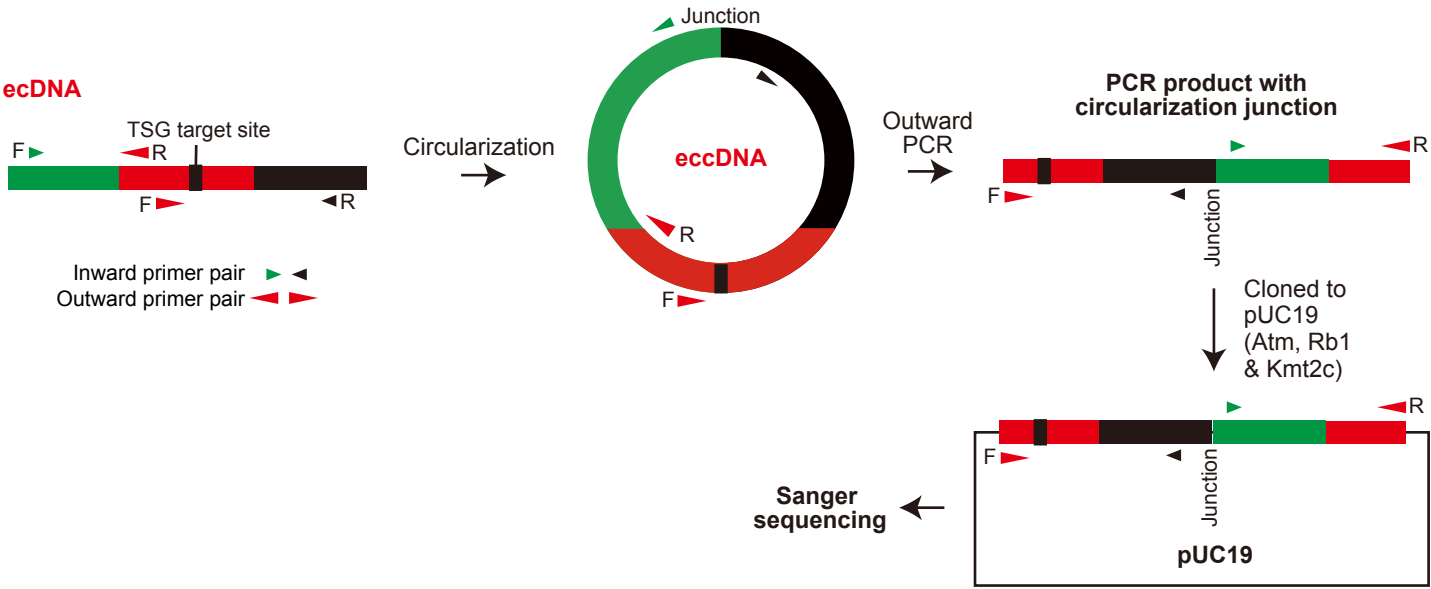


Figure S12

A



B

Target sequences for eccDNA generation

Note: Exon sequences in red; >< for circularization position and direction; Outward primer pairs for PCR underlined; CRISPR/Cas9 target sites indicated by PAM in bold.

Atm exon 25

ATGTTTATAATGTGTTTTCTATGTGCTCTATGCTATACCAGTTATTAAATACATTTTCCATTTTCTCATT
TGCTATCTT >TCTGTCATTCATTGGTGAAGAAATCAAAGTAGAAAAACATTGAGCGACTGCTCTAAAGTCC
TATTATTCCCATTCTTAAACATCATGTTTTCTCTTCTGCAGGTTTTAGAGAAAGTCTCCGAATCGTT
TGG ATGTAGAAGTTTAGAAGACTTCATGATTTCTCACCTAGACTACCTGGTTTTGGAATGGCTGAACCTT
CAAGATACTG< AATATAGCTTATCTTCTTTTCTTTTATGTTATTAACTACACAAGCATTGAGGATTTCT
ATCGGTAGGCTGCAGCAGTGTTCGACAACCCCTCCTCCCTAGCTTCAAGTAGGAAGTAGCCTAGTTTT

Rb1 exon 8

ATTGTTTTTATTCTAATTTACCATTTT >TACAGAAACAGCTGCAATCCCCATTAATGGTTCACCTCGAAC
ACCCAGAAGAGGTCAGAACAGGAGCGCTCGGATAGCAAAACAACTAGAAAATGATACGAGG ATTATCGAGG
TTCTCTGTAAAGAACACGAGTGTAATATAGATGAGGTAATTTCTCTTACACTTTATTTAAACAGCTGAA
GTAGATATATATGCTAAGTCTGCCTGTCA< ATATATACTTCTTGTGAGCATACTTTCTTTTTTTAAGTAGG

Kmt2c exon 36

TTCTCAGCCACCAGCAGTACCAAGGCCAAGGACTTCAGAGGGTTTTACTAGGCCCTCCAGTGCAAGACCA
GCC >CTCATGCCAAACCAGGATCCTTTTTTGCAAGCAGCACAAaCCGAGTACCAGGTTTACCTGGCCCTT
TGATAAGGCCACCTGATACATGCTCCCAGACTCCCAGGCCACCTGGGCCT GGCCGTATAGACACATTAC
TCATGCTTCTCATCTGCTGTTTCGTGATCCATATGATCAGCCTCCAGTGACTCCCAGGCCTCATTCTGAG
TCTTTTCG< GAACTAGTCAAGTTGTTTCACGATCTTGTTGACCGTCCAGTTCCTGGGTCAGAGGGAACTTTA