

Supplementary Information

Genome Editing Is Induced in a Binary Manner in Single Human Cells

¹Gou Takahashi, ^{1,2}Daiki Kondo, ^{1,3}Minato Maeda, ⁴Yuji Morishita, and ^{1,2}Yuichiro Miyaoka

¹Regenerative Medicine Project, Tokyo Metropolitan Institute of Medical Science, Tokyo, Japan

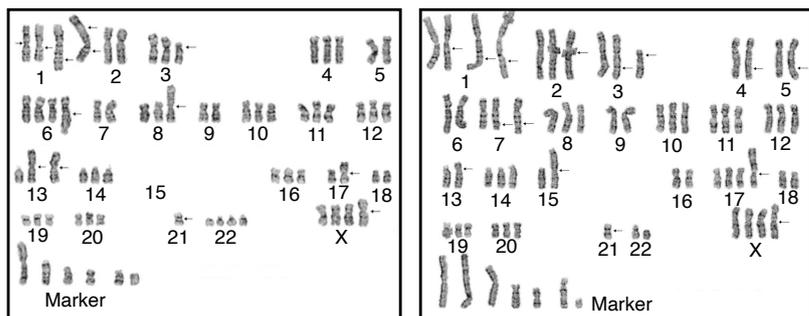
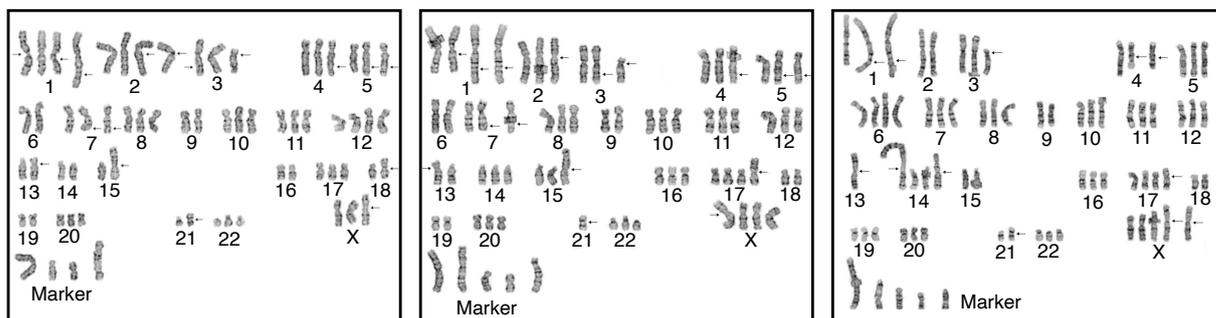
²Graduate School of Medical and Dental Sciences, Tokyo Medical and Dental University, Tokyo, Japan

³Department of Biological Sciences, Faculty of Science and Engineering, Chuo University, Tokyo, Japan.

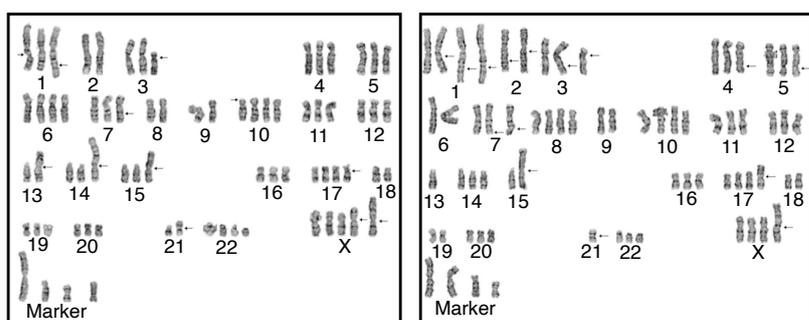
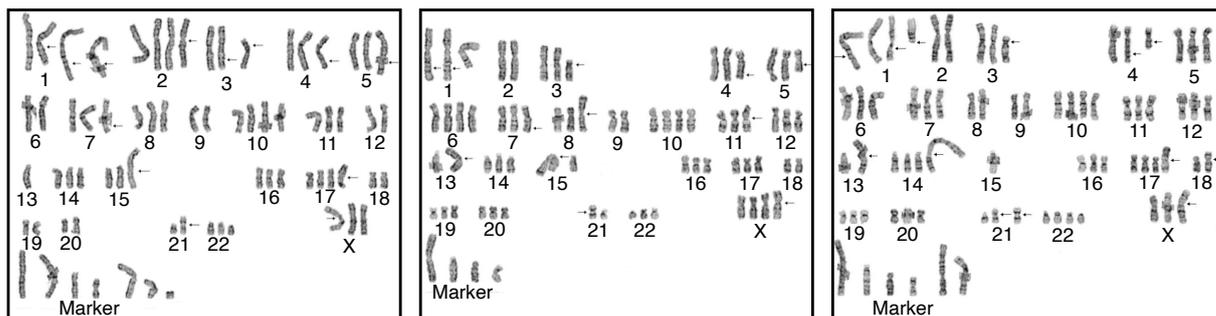
⁴On-chip Biotechnologies Co., Ltd., Tokyo, Japan.

Supplementary Figure. 1

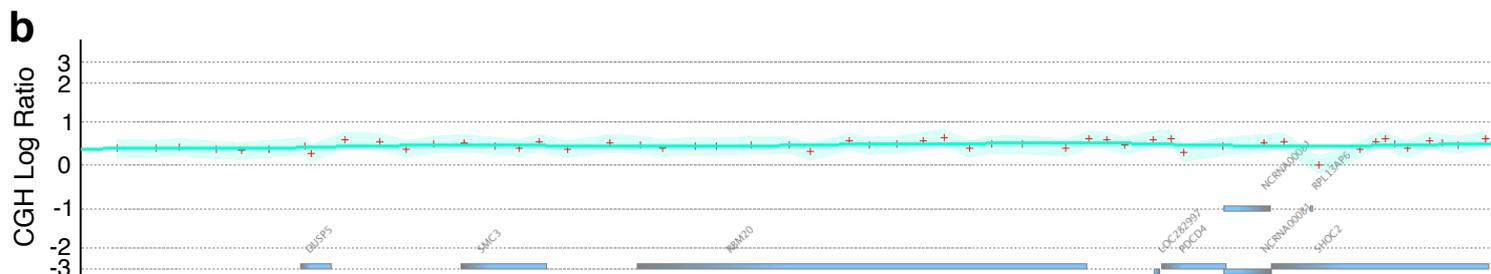
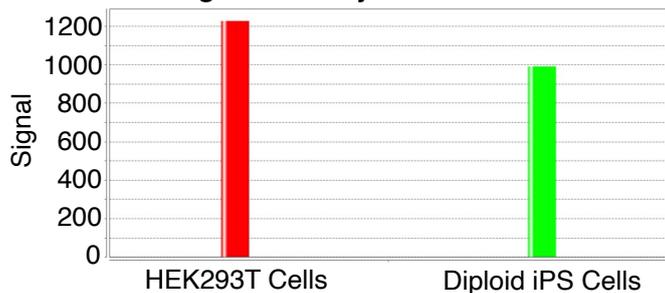
a Cells with three Chr.10



Cells with four Chr.10



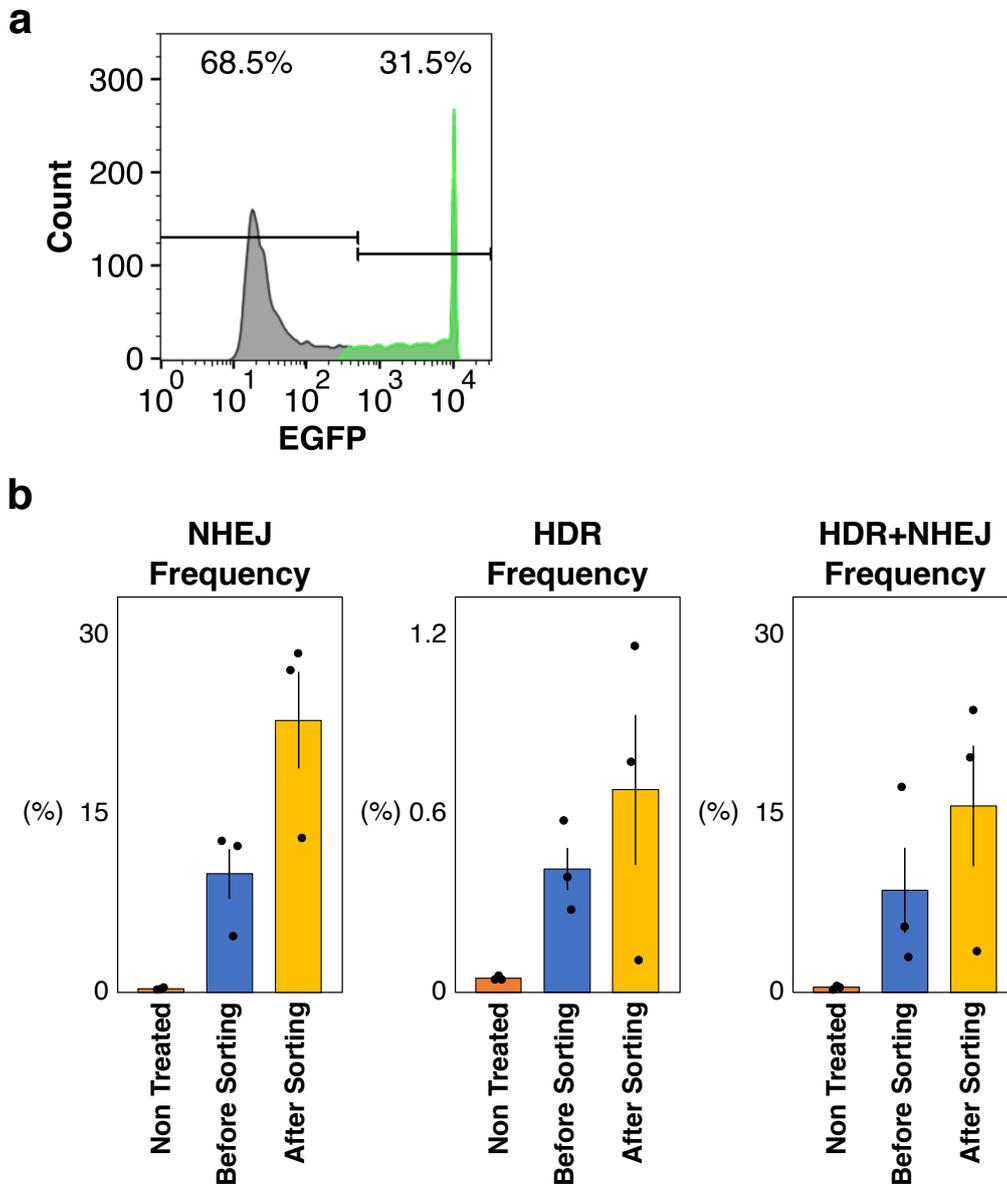
c Median Signal Intensity of CGH Around RBM20



Supplementary Figure 1. Karyotyping and a CGH analysis of HEK293T cells to estimate the copy number of RBM20.

(a) Karyotypes of ten HEK293T cells. Five cells had three chromosome 10s, and the other five had four chromosome 10s. Please note that two of them are also shown in Fig. 1b. (b) Scattered plot of the relative CGH signal of HEK293T cells in comparison to diploid human iPS cells around the RBM20 gene. The relative CGH signals of HEK293T cells normalized by the CGH signals of iPS cells are represented by +. This panel shows a wider genomic region than Fig. 1f. There were no detectable microduplications or microdeletions around the RBM20 locus. (c) The median signal intensity of CGH around RBM20. The median signal intensities of HEK293T cells and iPS cells in the genomic region shown in (b). By applying these values to calculate the CGH log ratio and applying it to the formula shown in Fig. 1e, the copy number of RBM20 was calculated to be 3.54.

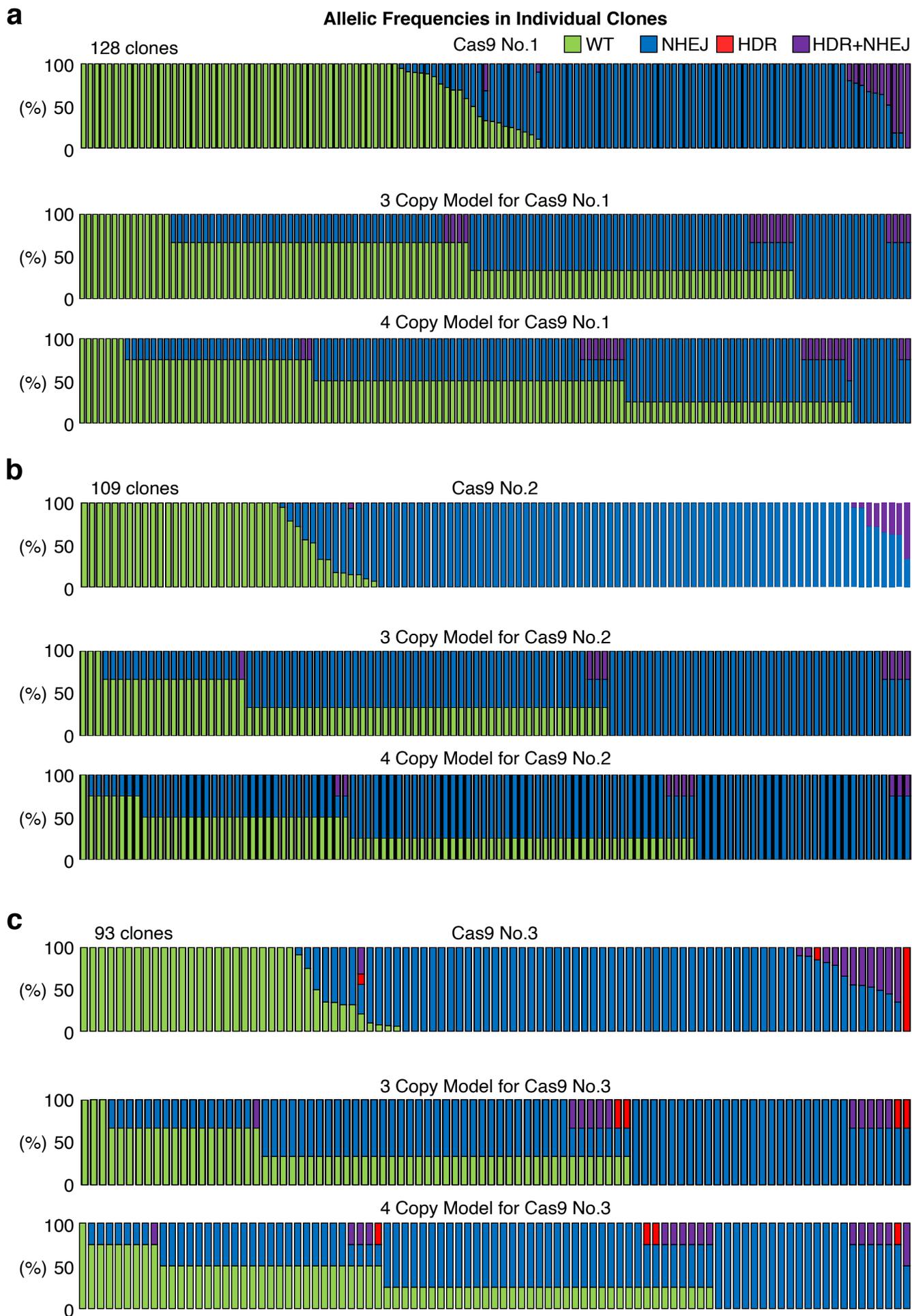
Supplementary Figure. 2



Supplementary Figure 2. Sorting of EGFP-positive HEK293T cells.

(a) Histogram showing the sorting gate of EGFP-positive HEK293T cells. (b) Frequencies of NHEJ, HDR, and HDR+NHEJ alleles before and after cell sorting. Cas9, gRNA targeting RBM20, and EGFP were co-expressed in HEK293T cells and EGFP-positive cells were sorted. The frequencies of NHEJ, HDR, and HDR+NHEJ alleles were analyzed by amplicon sequencing. Values \pm S.E. are shown (n=3).

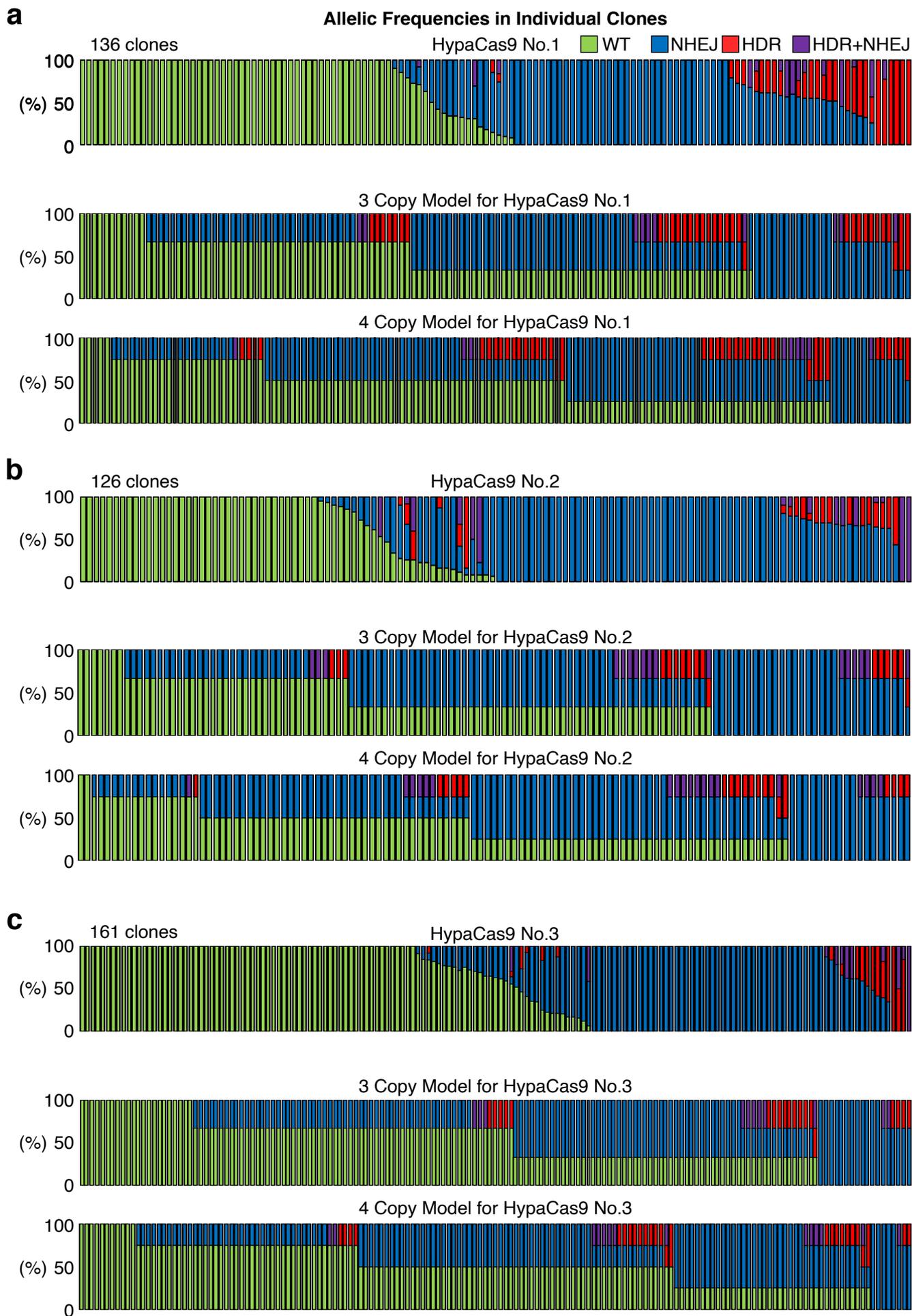
Supplementary Figure. 4



Supplementary Figure 4. Models assuming random RBM20 R636S mutagenesis by Cas9 in HEK293T cells.

(a-c) Models of the distribution of HEK293T cell clones with different genome editing outcomes by Cas9, if genome editing randomly occurred at the overall frequencies in HEK293T cells with three or four copies of RBM20 in experiment No. 1 (a), No. 2 (b), and No. 3 (c). The top row in each panel shows the observed distribution of the isolated HEK293T cell clones. Note that the original data and the 3 and 4 copy models for Cas9 No. 3 are also shown in Fig. 3.

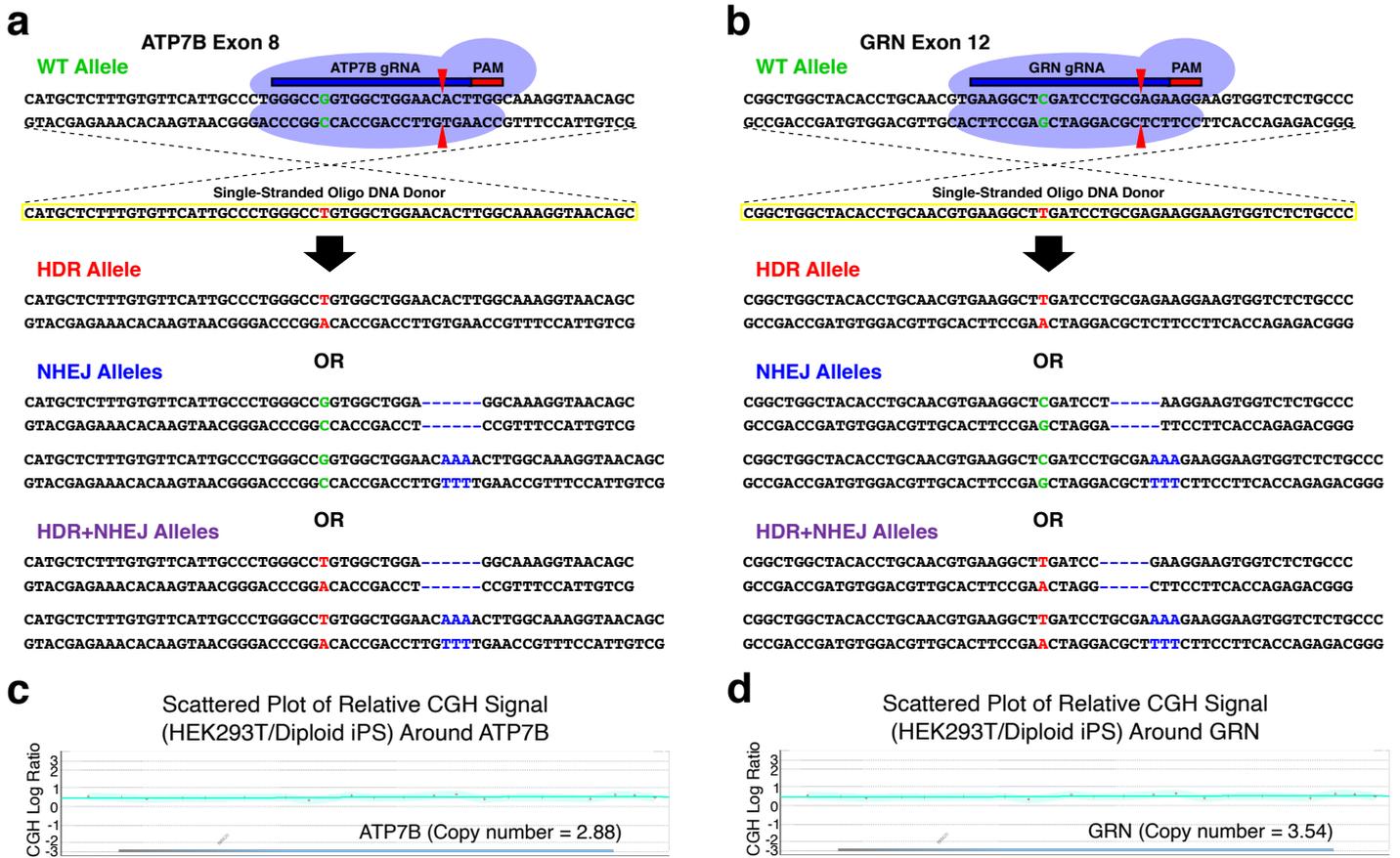
Supplementary Figure. 5



Supplementary Figure 5. Models assuming random RBM20 R636S mutagenesis by HypaCas9 in HEK293T cells.

(a-c) Models of the distribution of HEK293T cell clones with different genome editing outcomes by HypaCas9, if genome editing randomly occurred at the overall frequencies in HEK293T cells with three or four copies of RBM20 in experiment No. 1 (a), No. 2 (b), and No. 3 (c). The top row in each panel shows the observed distribution of the isolated HEK293T cell clones. Note that the original data and the 3 and 4 copy models for HypaCas9 No. 3 are also shown in Fig. 4.

Supplemental Figure 6



Supplementary Figure 6. Design of ATP7B R778L and GRN R493X targeting, and estimation of copy numbers of ATP7B and GRN in HEK293T cells.

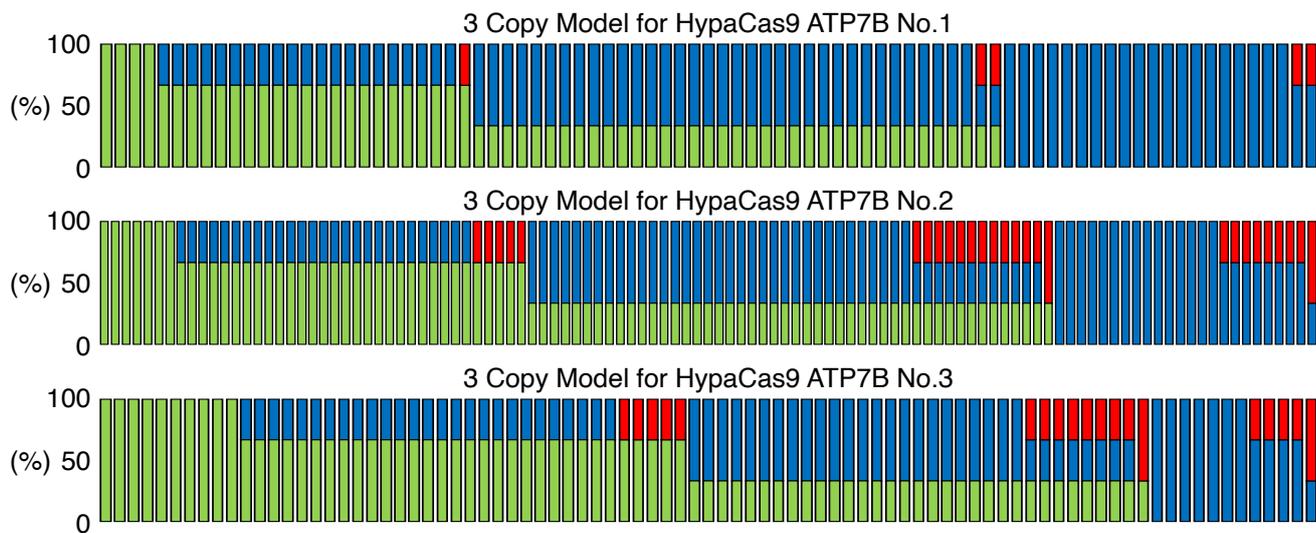
(a, b) Design of genome editing in ATP7B (a) and GRN (b), respectively. We introduced the R778L (c.2333C>G) (a) and R493X (c.1477C>T) (b) mutations using HypaCas9 and a single-stranded oligonucleotide donor DNA. The resulting HDR alleles have a single nucleotide substitution from C to G (a) and C to T (b), whereas the NHEJ alleles have various insertions and deletions. (c, d) Scattered plots of the relative CGH signal of HEK293T cells in comparison to diploid human iPS cells around the ATP7B (c) and GRN (d) genes, respectively. The relative CGH signals of HEK293T cells in comparison to diploid iPS cells are represented by +. No microduplications or microdeletions were detected around the ATP7B (c) or GRN (d) loci. The median CGH log ratios in these genomic regions shown here were applied to the formula shown in Fig. 1e to calculate the copy numbers of ATP7B (c) and GRN (d) were 2.88 (c) and 3.54 (d), respectively.

Supplementary Figure. 7

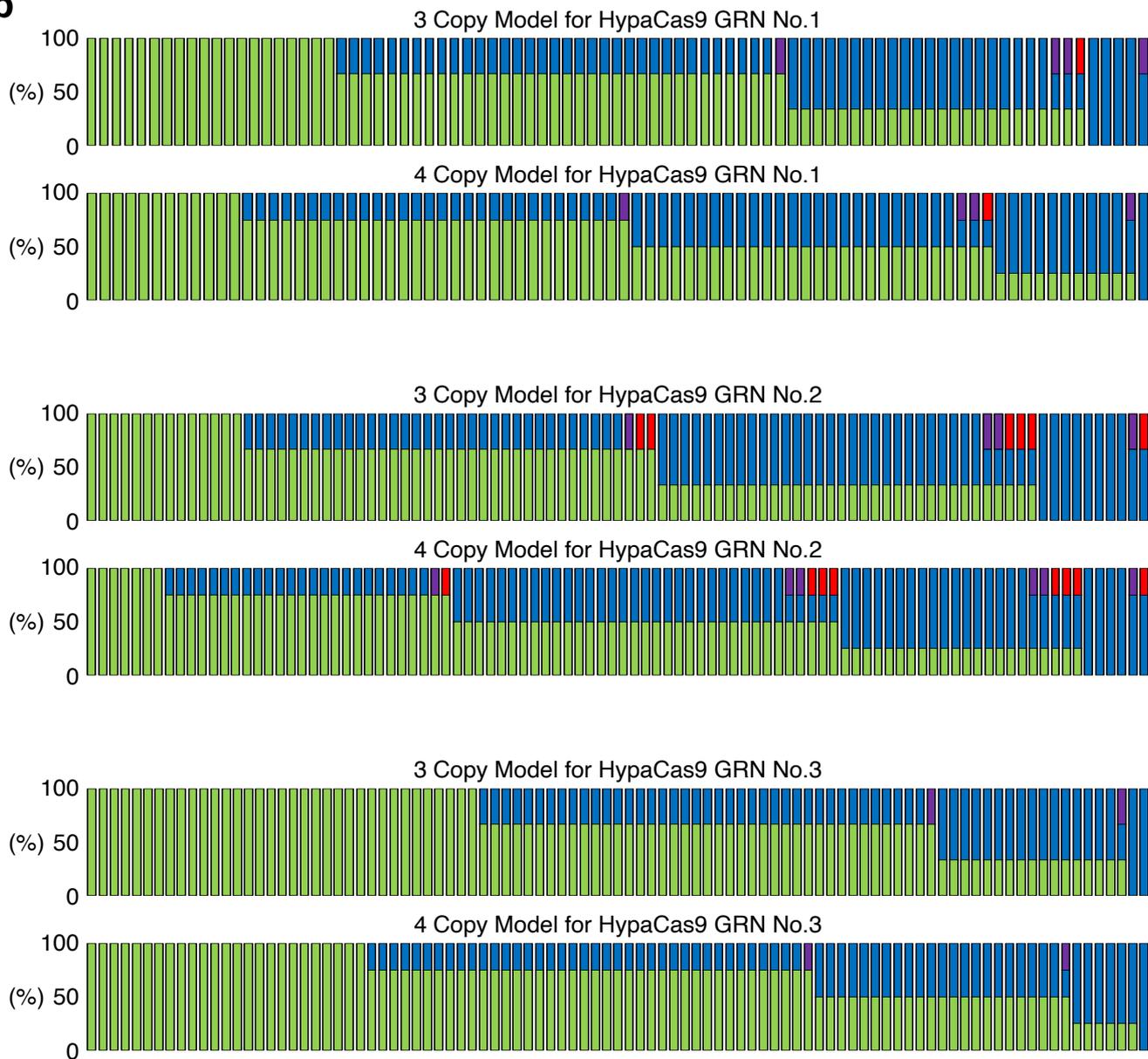
a

Expected Allelic Frequencies in Individual Clones

WT NHEJ HDR HDR+NHEJ



b

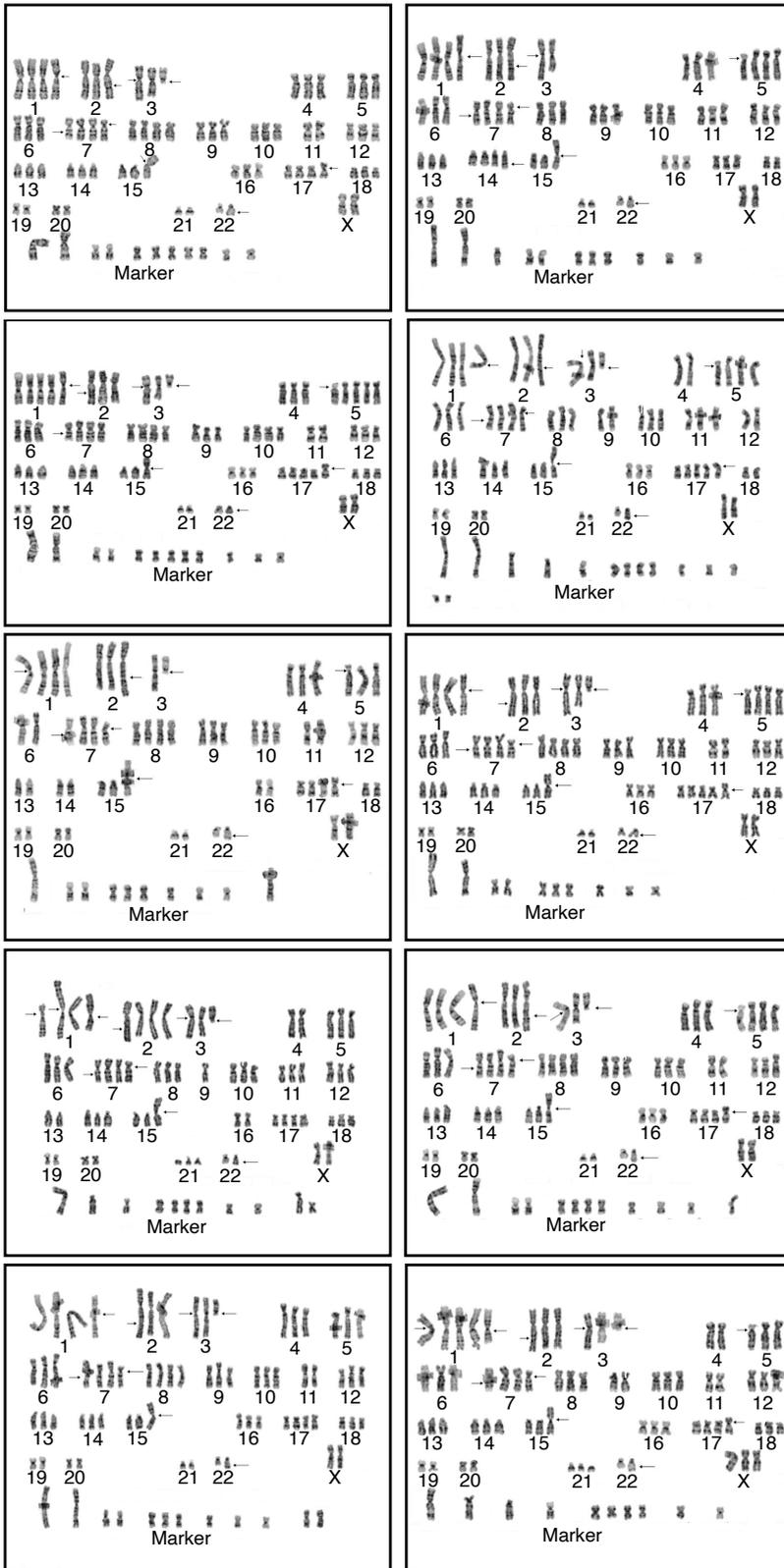


Supplementary Figure 7. Models assuming random ATP7B R778L and GRN R493X mutagenesis by HypaCas9 in HEK293T cells.

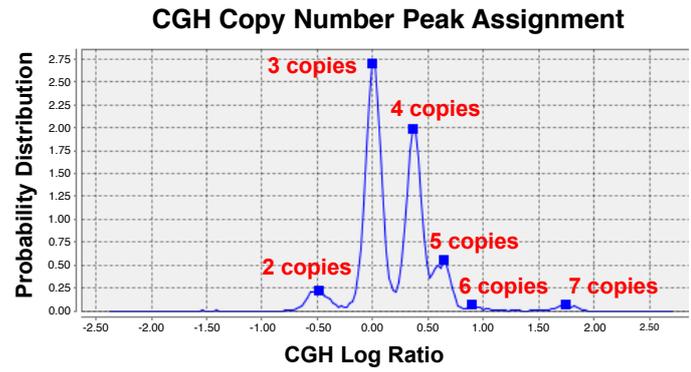
(a-b) Models of the distribution of HEK293T cell clones with different genome editing outcomes by HypaCas9, if genome editing randomly occurred at the overall frequencies in HEK293T cells with three copies of ATP7B (a) and three or four copies of GRN (b) in experiment Nos. 1-3.

Supplementary Figure. 8

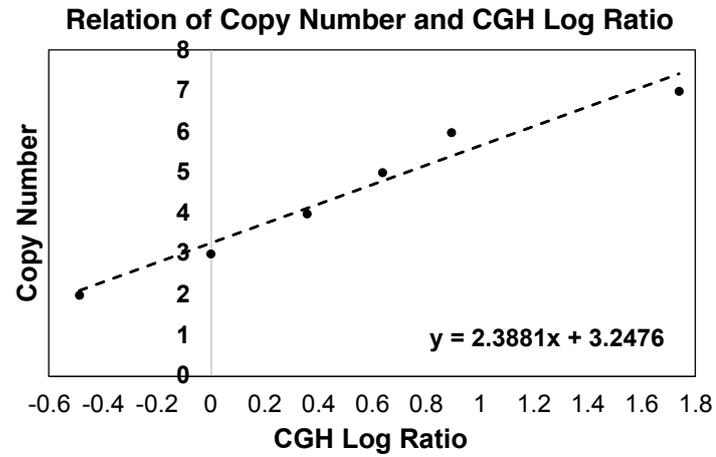
a



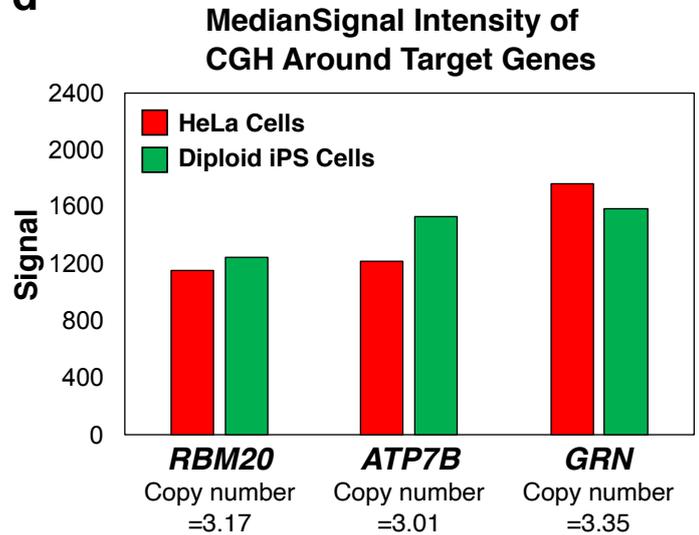
b



c



d

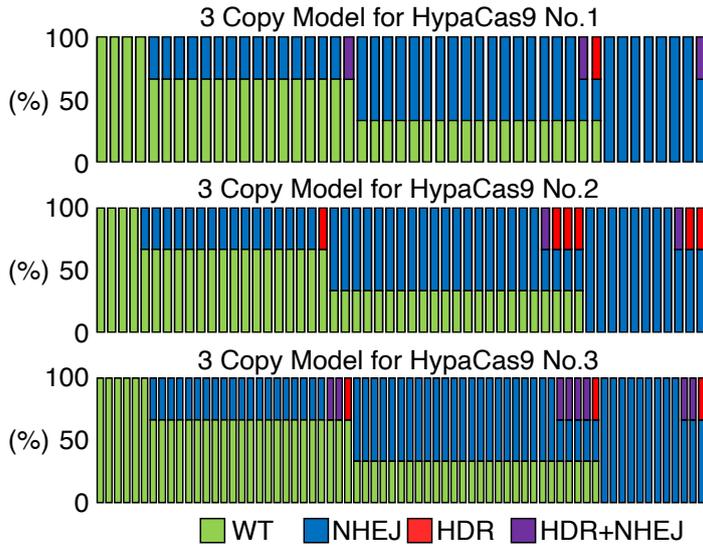


Supplementary Figure 8. Karyotyping and a CGH analysis to estimate the copy number of RBM20, ATP7B, and GRN in HeLa cells.

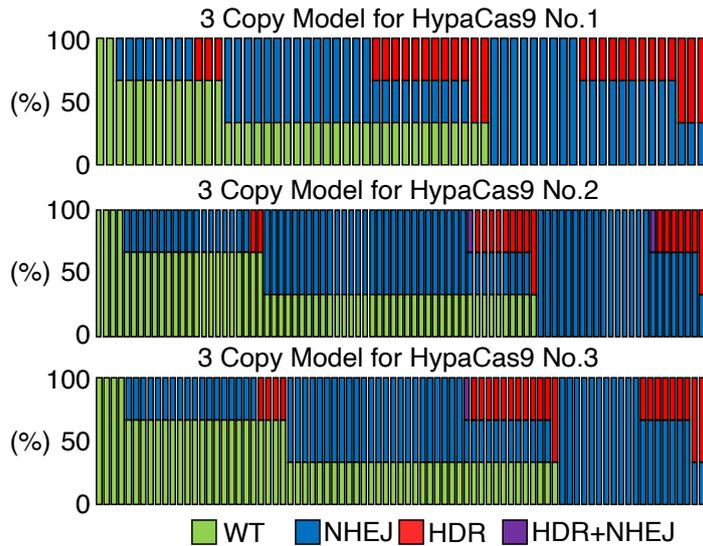
(a) Karyotypes of ten HeLa cells. (b) CGH copy number peak assignment. In the CGH analysis, there were several peaks of the CGH signal ratio between HeLa cells and diploid iPS cells based on the chromosomal numbers. The highest peak corresponded to three copies per cell, and the other peaks were incrementally assigned to the other copy numbers. (c) Line of fit between the copy number and the CGH log ratio based on the peak assignment shown in (b). (d) The median signal intensities of CGH around the RBM20, ATP7B, and GRN genes. By applying these values to calculate the CGH log ratios, which were then applied to the formula shown in (c), the copy numbers of RBM20, ATP7B, and GRN, in HeLa cells were 3.17, 3.01, and 3.35, respectively.

Supplementary Figure. 9

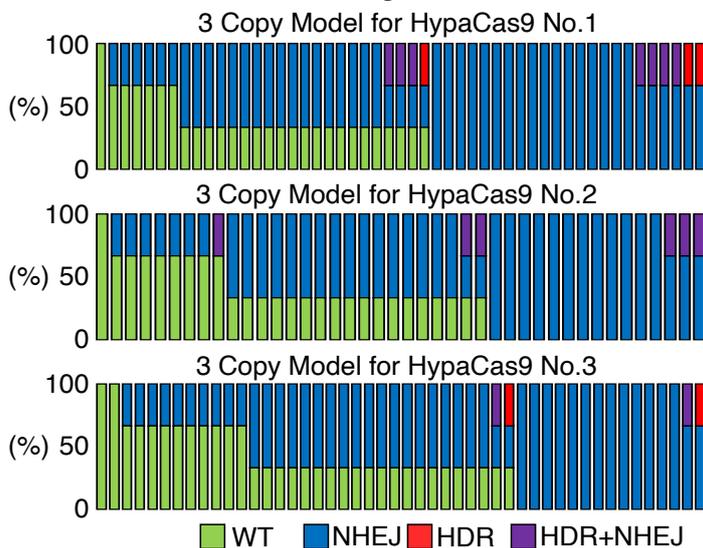
a Expected Allelic Frequencies in Individual Clones in RBM20 R636S Mutagenesis in HeLa Cells



b Expected Allelic Frequencies in Individual Clones in ATP7B R778L Mutagenesis in HeLa Cells



c Expected Allelic Frequencies in Individual Clones in GRN R493X Mutagenesis in HeLa Cells

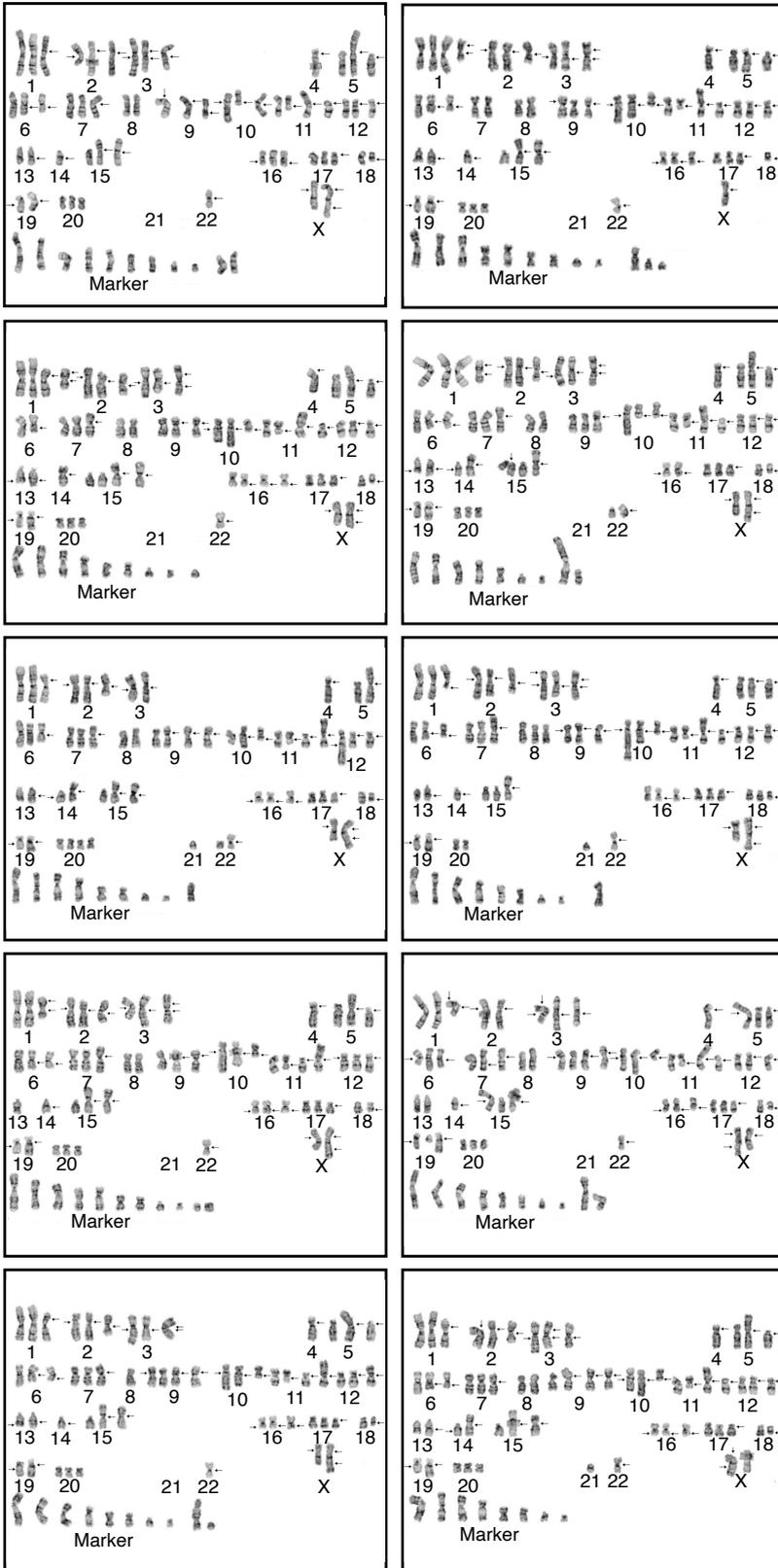


Supplementary Figure 9. Models assuming random RBM20 R636S, ATP7B R778L, and GRN R493X mutagenesis by HypaCas9 in HeLa cells.

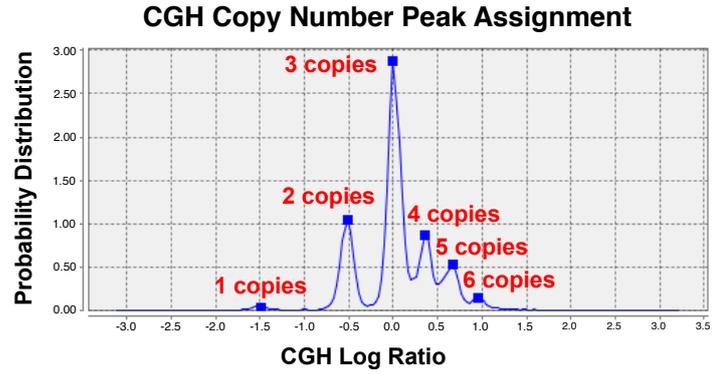
(a-c) Models of the distribution of HeLa cell clones with different genome editing outcomes by HypaCas9, if genome editing randomly occurred at the overall frequencies in HeLa cells with three copies of RBM20 (a), ATP7B (b), and GRN (c) in experiment Nos. 1-3.

Supplementary Figure. 10

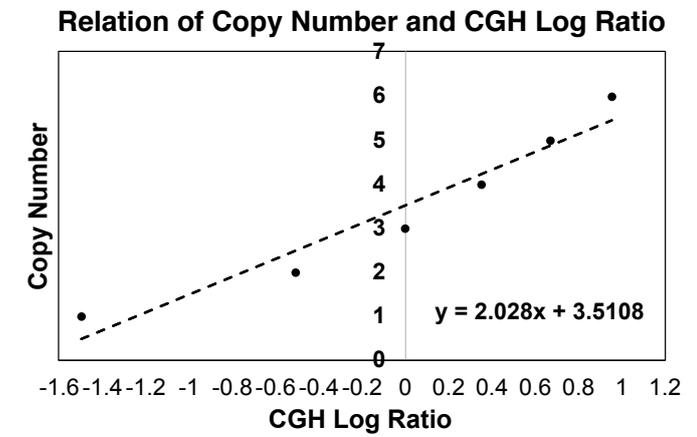
a



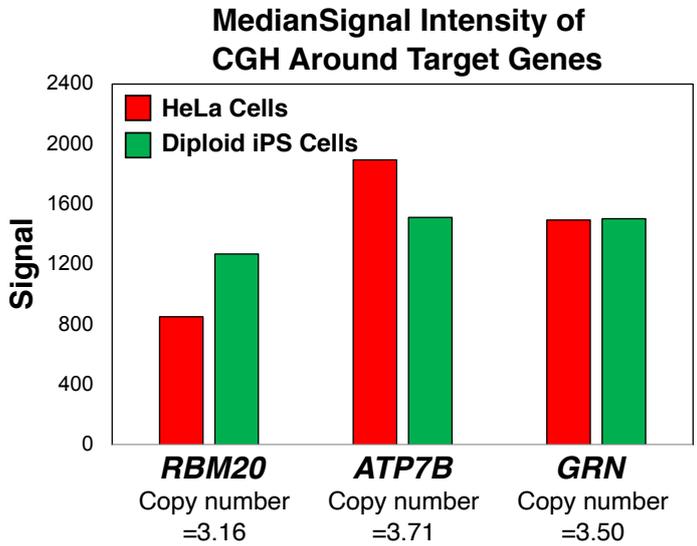
b



c



d

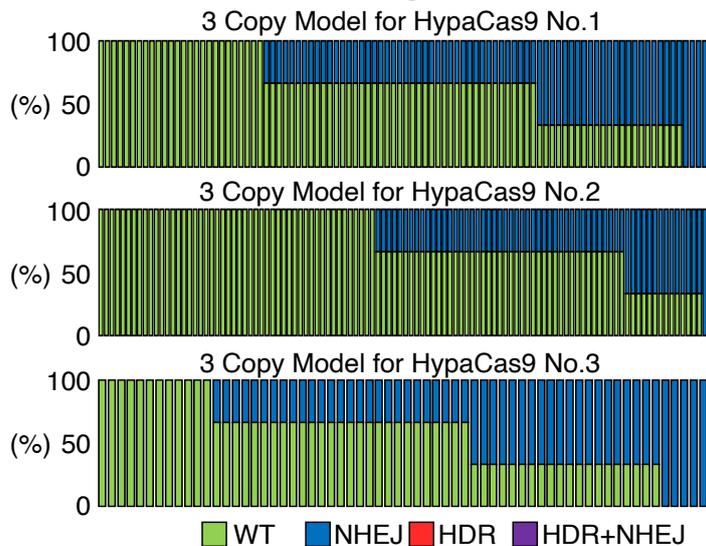


Supplementary Figure 10. Karyotyping and a CGH analysis to estimate the copy number of RBM20, ATP7B, and GRN in PC9 cells.

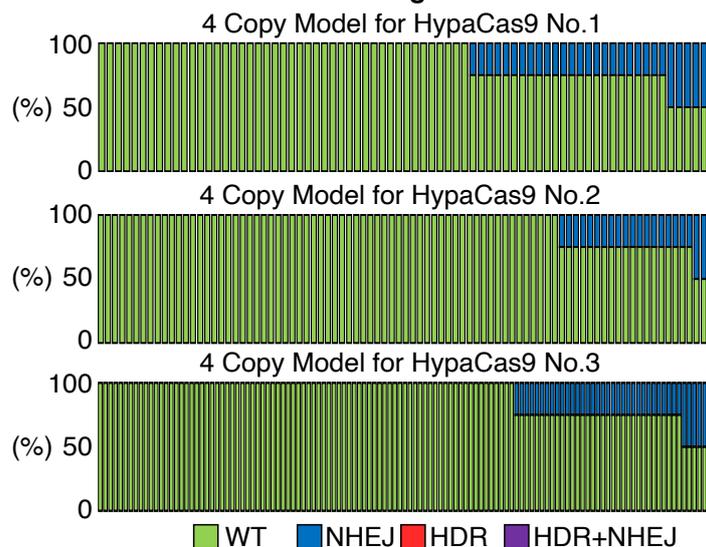
(a) Karyotypes of ten PC9 cells. (b) CGH copy number peak assignment. In the CGH analysis, there were several peaks of the CGH signal ratio between PC9 cells and diploid iPS cells based on the chromosomal numbers. The highest peak corresponded to three copies per cell, and the other peaks were incrementally assigned to the other copy numbers. (c) Line of fit between the copy number and the CGH log ratio based on the peak assignment shown in (b). (d) The median signal intensities of CGH around the RBM20, ATP7B, and GRN genes. By applying these values to calculate the CGH log ratios, which were then applied to the formula shown in (c), the copy numbers of RBM20, ATP7B, and GRN in PC9 cells were 3.16, 3.71, and 3.50, respectively.

Supplementary Figure. 11

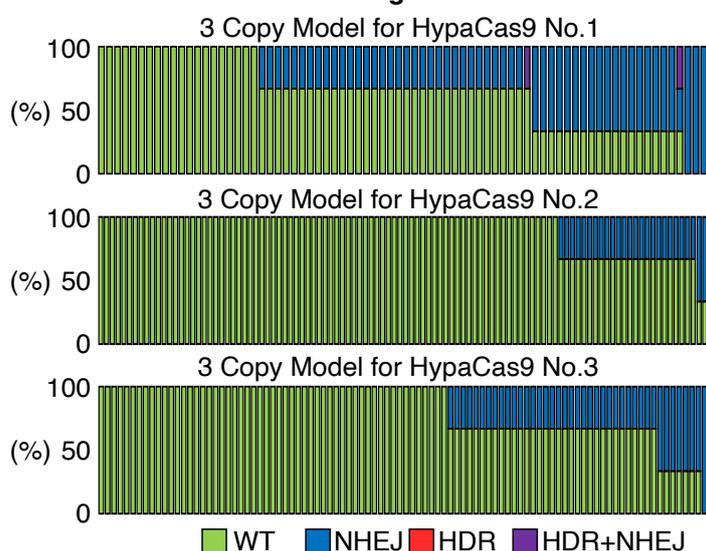
a Expected Allelic Frequencies in Individual Clones in RBM20 R636S Mutagenesis in PC9 Cells



b Expected Allelic Frequencies in Individual Clones in ATP7B R778L Mutagenesis in PC9 Cells

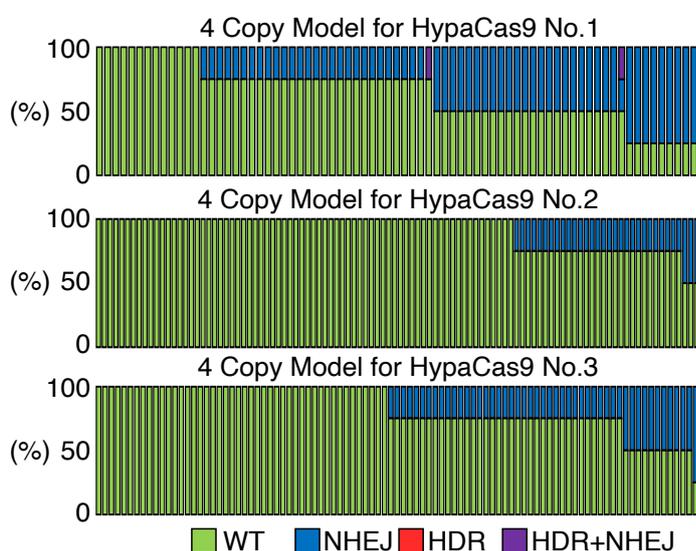


c Expected Allelic Frequencies in Individual Clones in GRN R493X Mutagenesis in PC9 Cells

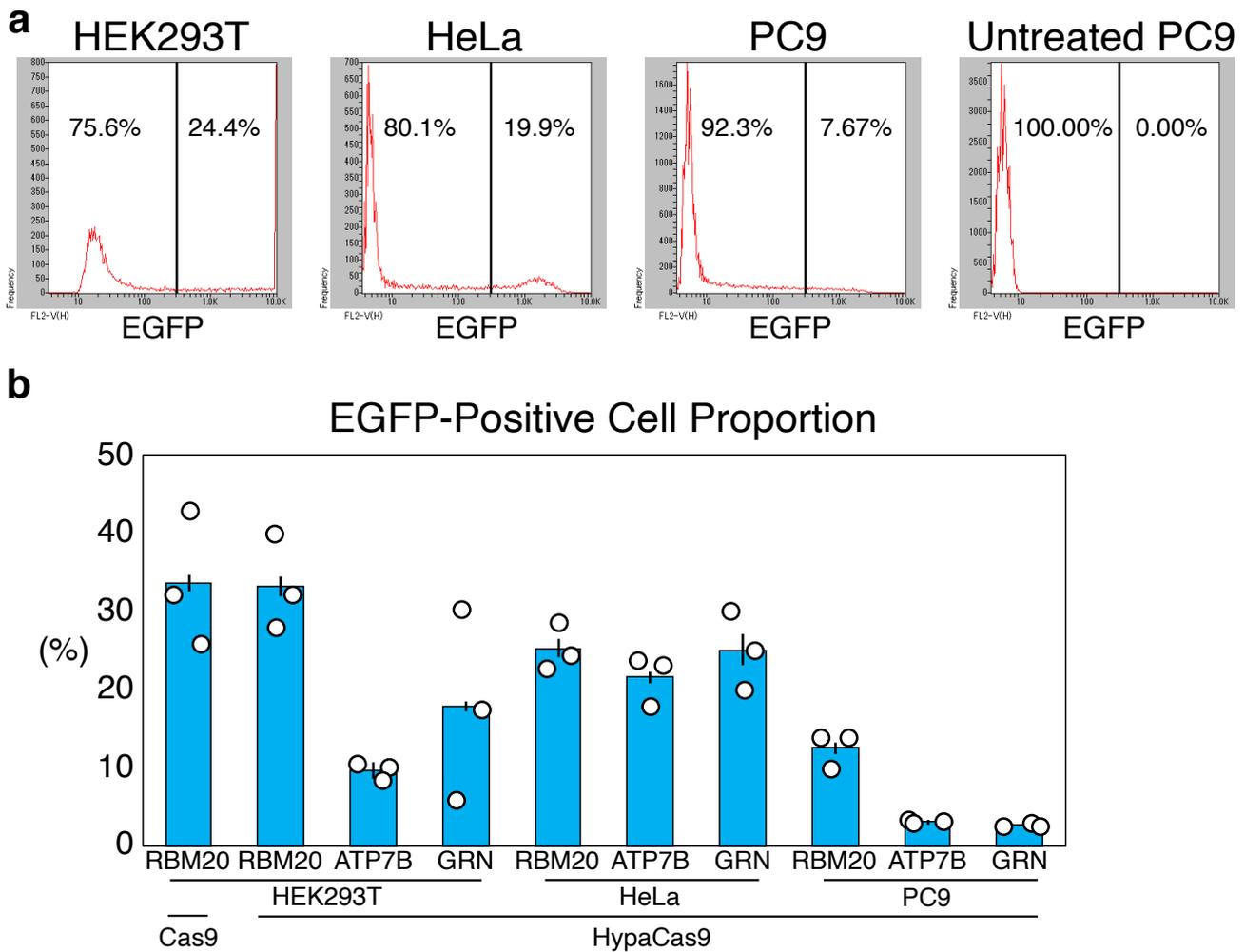


Supplementary Figure 11. Models assuming random RBM20 R636S, ATP7B R778L, and GRN R493X mutagenesis by HypaCas9 in PC9 cells.

(a-c) Models of the distribution of PC9 cell clones with different genome editing outcomes by HypaCas9, if genome editing randomly occurred at the overall frequencies in PC9 cells with three copies of RBM20 (a), four copies of ATP7B (b), and three or four copies of GRN (c) in experiment Nos. 1-3.



Supplementary Figure. 12



Supplementary Figure 12. EGFP-positive populations after co-expression of Cas9/HypaCas9, gRNA, and EGFP in HEK293T cells, HeLa cells, and PC9 cells.

(a) Representative histograms of EGFP-positive populations in HEK293T cells, HeLa cells, and PC9 cells 2 days after transfection of the plasmids to target RBM20. PC9 cells without any treatment are shown as a negative cell population. The expression level of EGFP in HEK293T cells was saturated in this histogram. (b) EGFP-positive cell populations determined by flow cytometry. Values \pm S.E. are shown (n=3).

Supplementary Table 4. Oligonucleotide donor DNAs used in this study

Name	Sequence
RBM20 R636S	ACAGATATGGCCCAGAAAGGCCGCGGTCTAGTAGTCCGGTGAGCCGGTCACTCTCCCCGA
ATP7B R778L	CATGCTCTTTGTGTTCAATTGCCCTGGGCCTGTGGCTGGAACACTTGGCAAAGGTAACAGC
GRN R493X	CGGCTGGCTACACCTGCAACGTGAAGGCTTGATCCTGCGAGAAGGAAGTGGTCTCTGCC

Supplementary Table 5. gRNAs used in this study.

PAM Sequences are underlined. Bold letters indicate the sites of single nucleotide substitutions.

Name	Sequence
RBM20	GGTCT CG TAGTCCGGTGAGCCGG
ATP7B	GGGCC GG TGGCTGGAACACTTGG
GRN	GAAGGCT CG ATCCTGCGAGAAGG

Supplementary Table 6. First PCR primers for amplicon sequencing

The 3' side of the primer sequences, represented by lowercase letters, are gene-specific. The 5' side of the primers, represented by the uppercase letters, are the common adaptor sequences.

Primer Name	Sequence (5' -> 3')
RBM20_FW-2	TAAC TT ACGGAGTCGCTCTACGgtgtctgtgtgtgggtggggtggg
RBM20_RV-2	GGATGGGAT TT CTTTAGGTCCTGagagctgcaggaggtgaagctgggag
ATP7B_NGS_FW-1	TAAC TT ACGGAGTCGCTCTACGttattctctggtcatcctggtg
ATP7B_NGS_RV-2	GGATGGGAT TT CTTTAGGTCCTGagcagctcttttctgaacctga
GRN-NGS_FW1	TAAC TT ACGGAGTCGCTCTACGataatgccattctgtgctcccttc
GRN-NGS_RV1	GGATGGGAT TT CTTTAGGTCCTGcactccacgtccttcacacc

Supplementary Table 7. Second PCR primer for amplicon sequencing

Custom P5 index	Custom P7 index	Custom P5 index sequence	Custom P7 index sequence	Custom P5 adapter sequence (5'→3')
P01 PE1.0	P01 PE2.0	CAAGTGTTTC	CAAGTGTTTC	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNCAAGTGTCTTAACCTACGGAGTCGCTCTACG
P02 PE1.0	P02 PE2.0	AGGACATTTC	AGGACATTTC	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNAGGACATTCTAACCTACGGAGTCGCTCTACG
P03 PE1.0	P03 PE2.0	CACATAATGG	CACATAATGG	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNCACTATGGTAACCTACGGAGTCGCTCTACG
P04 PE1.0	P04 PE2.0	AGCCTGATG	AGCCTGATG	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNAGCGGATTAACCTACGGAGTCGCTCTACG
P05 PE1.0	P05 PE2.0	TTACGCTAA	TTACGCTAA	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNTACGCTAATAACCTACGGAGTCGCTCTACG
P06 PE1.0	P06 PE2.0	ACTCTCCGT	ACTCTCCGT	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNACTCTCCGTTAACTACGGAGTCGCTCTACG
P07 PE1.0	P07 PE2.0	GTGATGCA	GTGATGCA	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNGTCGATGCACTAACCTACGGAGTCGCTCTACG
P08 PE1.0	P08 PE2.0	ACGGGAATT	ACGGGAATT	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNAGGGGAATTAACCTACGGAGTCGCTCTACG
P09 PE1.0	P09 PE2.0	CGGCCCCAG	CGGCCCCAG	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNCGGCCCCAGTAACCTACGGAGTCGCTCTACG
P10 PE1.0	P10 PE2.0	ACTAGTTTG	ACTAGTTTG	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNACTAGTTTGTAACTACGGAGTCGCTCTACG
P11 PE1.0	P11 PE2.0	AGTATTACA	AGTATTACA	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNAGTATTACATAACCTACGGAGTCGCTCTACG
P12 PE1.0	P12 PE2.0	AGGTTGGGT	AGGTTGGGT	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNAGGTTGAACTACGGAGTCGCTCTACG
P13 PE1.0	P13 PE2.0	GTGAACCGA	GTGAACCGA	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNGTGAACCGATAACCTACGGAGTCGCTCTACG
P14 PE1.0	P14 PE2.0	GCACAAAAC	GCACAAAAC	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNGCACAAAACCTAACCTACGGAGTCGCTCTACG
P15 PE1.0	P15 PE2.0	CTGTCTTCG	CTGTCTTCG	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNCTGTCTTCGTAACCTACGGAGTCGCTCTACG
P16 PE1.0	P16 PE2.0	TAAGCGACT	TAAGCGACT	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNAGCGGACTTAACCTACGGAGTCGCTCTACG
P17 PE1.0	P17 PE2.0	CAGCCCATATA	CAGCCCATATA	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNACAGCCCATATAACCTACGGAGTCGCTCTACG
P18 PE1.0	P18 PE2.0	AGATATCTG	AGATATCTG	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNAGATATCTGTAACCTACGGAGTCGCTCTACG
P19 PE1.0	P19 PE2.0	AATAGCAC	AATAGCAC	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNAATAGCACCTAACCTACGGAGTCGCTCTACG
P20 PE1.0	P20 PE2.0	TATCGTGCC	TATCGTGCC	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNTATCGTGCCCTAACCTACGGAGTCGCTCTACG
P21 PE1.0	P21 PE2.0	TCCTGGTAT	TCCTGGTAT	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNCTCTGGTATTAACCTACGGAGTCGCTCTACG
P22 PE1.0	P22 PE2.0	GGCAGAGGA	GGCAGAGGA	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNGGCAGAGGATAACCTACGGAGTCGCTCTACG
P23 PE1.0	P23 PE2.0	ACAATGGAG	ACAATGGAG	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNACAATGGAGTAACCTACGGAGTCGCTCTACG
P24 PE1.0	P24 PE2.0	TAGTGCCCA	TAGTGCCCA	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNTAGTGCCCATAACTACGGAGTCGCTCTACG
P25 PE1.0	P25 PE2.0	ATCCTATCT	ATCCTATCT	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNATCCTATCTTAACCTACGGAGTCGCTCTACG
P26 PE1.0	P26 PE2.0	CTGCTTAGC	CTGCTTAGC	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNCTGCTTAGCTAACCTACGGAGTCGCTCTACG
P27 PE1.0	P27 PE2.0	CGTGGCTGC	CGTGGCTGC	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNCGTGGCTGCTAACCTACGGAGTCGCTCTACG
P28 PE1.0	P28 PE2.0	GTAACATAAT	GTAACATAAT	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNGTAACATAATTAACCTACGGAGTCGCTCTACG
P29 PE1.0	P29 PE2.0	GCTTAGCAA	GCTTAGCAA	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNGCTTAGCAATAACCTACGGAGTCGCTCTACG
P30 PE1.0	P30 PE2.0	GATGCAGTG	GATGCAGTG	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNGATGCAGTGAACCTACGGAGTCGCTCTACG
P31 PE1.0	P31 PE2.0	GGCTACAA	GGCTACAA	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNGGCTACAACTAACCTACGGAGTCGCTCTACG
P32 PE1.0	P32 PE2.0	ATAGCACGC	ATAGCACGC	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNATAGCACGCTAACCTACGGAGTCGCTCTACG
P33 PE1.0	P33 PE2.0	TATAACTCT	TATAACTCT	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNTATAACTCTTAACCTACGGAGTCGCTCTACG
P34 PE1.0	P34 PE2.0	AAATGCGTT	AAATGCGTT	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNAAATGCGTTAACCTACGGAGTCGCTCTACG
P35 PE1.0	P35 PE2.0	CCATCCAGG	CCATCCAGG	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNCCATCCAGGTAACCTACGGAGTCGCTCTACG
P36 PE1.0	P36 PE2.0	GTCTGCACC	GTCTGCACC	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNGTCTGCACCTAACCTACGGAGTCGCTCTACG

Supplementary Table 8. Commands for the CRISPResso2 analysis used in this study.

CRISPResso version (CRISPResso Batch mode): 2.0.45

Commands used:

```
/opt/conda/bin/CRISPResso -o /DATA/CRISPRessoBatch_on_batch --name (File Name 1) --  
needleman_wunsch_gap_extend -2 --aln_seed_count 5 --amplicon_name Reference --  
amplicon_seq (Sequence 1) --max_rows_alleles_around_cut_to_plot 50 --  
prime_editing_pegRNA_extension_quantification_window_size 5 --fastq_r1 (File Name 2) --  
quantification_window_size 1 --quantification_window_center -3 --trimmomatic_command  
trimmomatic --conversion_nuc_from C --min_bp_quality_or_N 30 --default_min_aln_score 30 --  
needleman_wunsch_gap_incentive 1 --min_paired_end_reads_overlap 10 --plot_window_size 20 -  
-prime_editing_pegRNA_scaffold_min_match_length 1 --aln_seed_min 2 --aln_seed_len 10 --  
expected_hdr_amplicon_seq (Sequence 2) --needleman_wunsch_gap_open -20 --  
max_paired_end_reads_overlap 100 --guide_seq (Sequence 3) --dsODN (Sequence 4) --  
conversion_nuc_to T --flexiguide_homology 80 --flash_command flash --min_single_bp_quality 30  
--exclude_bp_from_left 15 --needleman_wunsch_aln_matrix_loc EDNAFULL --  
min_average_read_quality 30 --min_frequency_alleles_around_cut_to_plot 0.2 --  
exclude_bp_from_right 15
```

File Name 1_ Experiment Name

File Name 2_ "File name".fastq.gz

Sequence 1_ Reference Sequence

Sequence 2_ Excepted HDR Sequence

Sequence 3_ gRNA Sequence

Sequence 4_ Donor Sequence