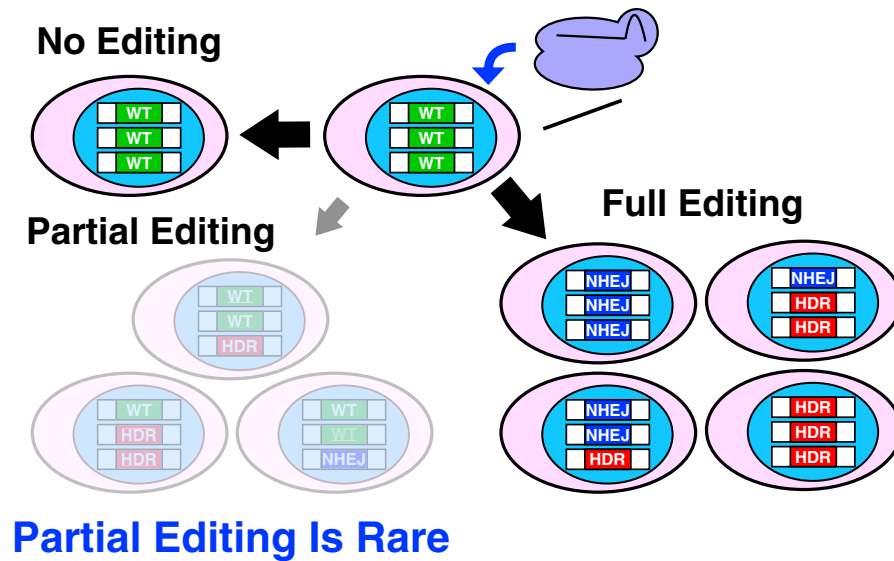


Article

Genome editing is induced in a binary manner in single human cells

Introduction of CRISPR Components Most Often Results in Either No Editing or Full Editing

Gou Takahashi,
Daiki Kondo,
Minato Maeda,
Yuji Morishita,
Yuichiro Miyaoka

miyaoka-yi@igakuken.or.jp

Highlights

Over 2,600 genome-edited human cell clones were isolated by using SPIS

Cas9 exhibited either full editing of target alleles or no editing in single cells

HDR was more often accompanied by NHEJ than WT in single cells

Article

Genome editing is induced in a binary manner in single human cells

Gou Takahashi,¹ Daiki Kondo,^{1,2} Minato Maeda,^{1,2} Yuji Morishita,³ and Yuichiro Miyaoka^{1,2,4,5,*}

SUMMARY

Even when precise nucleotide manipulations are intended, the outcomes of genome editing can be diverse, often including random insertions and deletions. The combinations and frequencies of these different outcomes in single cells are critical not only in the generation of genetically modified cell lines but also in the evaluation of the clinical effects of genome editing therapies. However, current methods only analyze cell populations, not single cells. Here, we utilized the Single Particle isolation System (SPiS) for the efficient isolation of single cells to systematically analyze genome editing results in individual human cultured cells. As a result, we discovered that genome editing induction has a binary nature, that is, the target alleles of cells tend to be all edited or not edited at all. This study enhances our understanding of the induction pattern of genome editing and provides a new strategy to analyze genome editing outcomes in single cells.

INTRODUCTION

Genome editing allows us to manipulate genetic information in basically any type of cell, and has been revolutionary in basic science, agriculture, and medicine.^{1–3} Genome editing tools were originally designed to cleave target sequences in the genome DNA in the cell, so that genetic manipulations can be introduced into the genome via activated DNA repair pathways at the target sites.⁴ These DNA repair pathways include non-homologous end-joining (NHEJ) and homology-directed repair (HDR).⁵ Each pathway gives distinct genome editing outcomes. Although base editing and prime editing technologies do not require DNA double-strand breaks to manipulate the genome DNA sequence, they can still never produce only one type of editing.^{6–9} Despite significant progress in the prediction of genome editing outcomes,^{10–13} it has been impossible to produce a sole genetic manipulation. In other words, genome editing outcomes are always mixtures of different modifications of DNA sequences, such as insertions or deletions of different sizes and targeted recombination events. Therefore, for the application of genome editing, it is important to precisely measure its outcomes.

Various types of techniques have been used to analyze genome editing outcomes,¹⁴ including sequencing-based methods (e.g., amplicon sequencing and TIDE¹⁵), denaturation-based methods (e.g., T7E1¹⁶ and single-stranded conformational polymorphism [SSCP] assays¹⁷), and digital PCR-based methods.^{18,19} However, all of these methods are designed to analyze cell populations, not single cells. The combination of fluorescent reporter systems, such as traffic light reporter²⁰ and flow cytometry, can visualize genome editing results in individual cells, but this setting cannot be applied to endogenous genes. To fully exploit the potential of genome editing, it is critical to grasp the editing outcomes in individual cells. For example, in a hypothetical situation where 50% of a population of diploid cells are WT/HDR heterozygotes and the other 50% are NHEJ/NHEJ homozygotes (Population 1 in Figure 1A), and the other population consists of 50% of WT/NHEJ and 50% of HDR/NHEJ heterozygotes (Population 2 in Figure 1A), the two populations would—as a whole—have identical allelic frequencies of WT, HDR, and NHEJ, whereas the cells would show a totally different composition (Figure 1A). Therefore, there is a strong demand for an efficient strategy to investigate genome editing outcomes in single cells.

The main reason why analyzing genome editing results in single cells has been so difficult is the limitation in the number of target molecules (i.e., a diploid cell has only two copies of genomic DNA). Even karyotypically abnormal cell lines only have several copies of target DNA sequences per cell at most. This is in clear contrast to the recent advancement of the single cell transcriptome,²¹ where a typical mammalian cell has ~10⁵ mRNA molecules to analyze.²² Recently, the Tapestry system to deep-sequence genomic

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

³On-chip Biotechnologies Co., Ltd., Tokyo, Japan

⁴Graduate School of Humanities and Sciences, Ochanomizu University, Tokyo, Japan

⁵Lead contact

*Correspondence: miyaoka-yi@igakuken.or.jp
<https://doi.org/10.1016/j.isci.2022.105619>



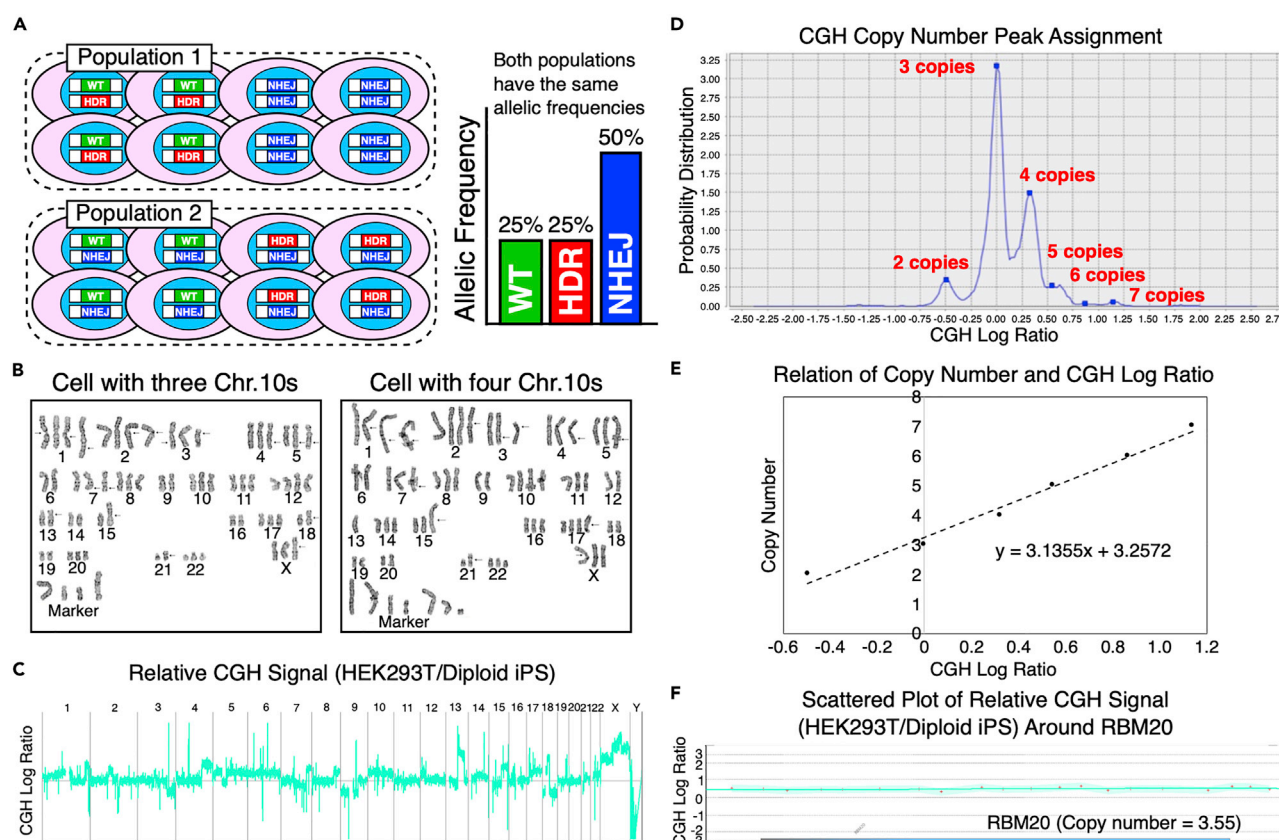


Figure 1. Measurement of the copy number of RBM20 in HEK293T cells

(A) A hypothetical situation that emphasizes the importance of the analysis of genome editing outcomes at the single cell level. Cell populations 1 and 2 consist of cells with totally different genotypes individually. However, the total allelic frequencies of WT, HDR, and NHEJ are exactly the same for both populations.

(B) Representative karyotypes of HEK293T cells with three and four chromosome 10s.

(C) A CGH analysis of HEK293T cells in comparison to diploid human iPS cells. The relative CGH signal of HEK293T cells normalized to that of diploid iPS cells is shown throughout the genome.

(D) CGH copy number peak assignment. In the CGH analysis, there were several peaks of the CGH signal ratio between HEK293T cells and diploid iPS cells based on the chromosomal numbers. The highest peak corresponded to three copies per cell, and was set as the baseline of the CGH log ratio between HEK293T cells and iPS cells.

(E) Line of fit between the copy number and the CGH log ratio based on the peak assignment shown in (D). The copy number and the CGH log ratio showed a clear linear correlation.

(F) Scattered plot of 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 that of iPS cells are represented by +. No microduplications or microdeletions were detected around the RBM20 locus.

DNA of single cells was applied to analyze genome editing outcomes.²³ Although this is an efficient strategy, sequencing with genomic DNA of single cells as a starting material is always at risk of losing alleles, as their copy number is so limited. Therefore, it is extremely challenging to directly analyze the genomic DNA sequences in single cells. However, there are two clear advantages in the analysis of genomic DNA: one is that cells can replicate their own genome DNA as long as they are alive and proliferate; the other is that its sequence does not change, in contrast to the repertoire of mRNA that changes dynamically depending on the cell state. Therefore, we decided to systematically isolate clones from a pool of genome edited cells. To accomplish this task, we utilized the Single Particle isolation System (SPiS) (On-chip Biotechnologies). The SPiS is an automated single cell dispenser that can gently and accurately dispense single cells in multi-well plates using image recognition technology to monitor the number of cells in aliquots.^{24,25} The SPiS enabled us to isolate clones derived from genome edited cells with unprecedented efficiency. Our new pipeline to analyze genome editing based on the SPiS and findings from it will greatly contribute to improving the understanding of how genome editing occurs in the cell.

RESULTS

Measurement of the copy number of RBM20 in HEK293T cells

First, we measured the copy number of our target gene, RBM20, per HEK293T cell to comprehend genome editing outcomes in single cells, because cell lines often have abnormal karyotypes. Therefore, we combined karyotyping and a comparative genomic hybridization (CGH) analysis to precisely estimate the copy number. We first karyotyped a total of 10 HEK293T cells, and found extensive chromosomal abnormalities (Figure 1B). The most frequent chromosomal number was three (Table S1). We also found that five cells each had three and four chromosome 10s where RBM20 is located, respectively (Figures 1B and S1A).

Next, we analyzed the genome DNA of HEK293T cells by a CGH analysis. We compared HEK293T cells with WTC11 iPS cells that had been confirmed to be diploid.¹⁹ The CGH analysis also demonstrated the chromosomal abnormalities of HEK293T cells, as we revealed by karyotyping (Figures 1B and 1C). We detected peaks of the CGH signal corresponding to the chromosome number in HEK293T cells (Figure 1D). Because the most frequent chromosome number was three (as determined by karyotyping), the most frequently observed peak of the CGH signal corresponded to three copies. We were able to incrementally assign copy numbers to these peaks to draw a line of fit between the relative CGH signal intensity and the copy number (Figures 1D and 1E). By applying the relative median signal intensity of HEK293T cells to WTC11 iPS cells around the RBM20 gene (Figures 1F, S1B, and S1C) to this line of fit, we estimated the average copy number of RBM20 to be 3.55 (Table S2), which matched the karyotyping results (Figure 1D).

Efficient isolation of HEK293T cell clones driven by the SPiS

To systematically analyze the genome editing outcomes in individual cells, we established an efficient pipeline to isolate cell clones that had gone through the genome editing process. Using CRISPR-Cas9 and a single-strand DNA donor in HEK293T cells, as we described previously,¹⁸ we introduced the RBM20 R636S mutation (c.1906C>A, chr.10), which causes inherited dilated cardiomyopathy²⁶ (Figure 2). We used pX458 (Addgene plasmid #48138) to express EGFP via the T2A peptide fused to Cas9 in combination with gRNA. Therefore, the expression of EGFP guaranteed the expression of the CRISPR components in the same cell (Figures 2B and S2A). To minimize damage in the sorting of these EGFP-positive cells, we used a microfluidic cell sorter, On-chip Sort.²⁷ We confirmed that the sorting of EGFP-positive cells successfully enriched genome-edited cells (Figure S2B). These sorted cells were then subjected to the single cell isolation process using the SPiS, in which 384 cells were individually plated into four 96-well plates. The SPiS aspirates diluted cell suspension into a microtip and analyzes the images (taken by a CCD camera) of the contents. The camera takes two images with a 1-s interval; thus, the contents in the microtip go down by gravity. The SPiS analyzes the two images to measure the velocity of precipitation and the size of the contents (Figure 2C). By doing this, the SPiS can discriminate cells from other objects (e.g., cell debris and dust). The SPiS dispenses contents of a microtip into a culture dish, only when the system recognizes one cell. After two to four weeks of culture, genomic DNA of HEK293T cell clones derived from single cells was harvested for amplicon sequencing to analyze the outcomes of genome editing (Figure 2B).

Successful analysis of genome editing outcomes in single HEK293T cells

We conducted single cell isolation processes using the SPiS three times (Cas9 No.1 to No. 3 experiments), and obtained >90 clones in each attempt (Figure 3A). We performed amplicon sequencing on the RBM20 target site and analyzed the outcomes in isolated clones using CRISPResso2²⁸ (Figure S3A). We defined the clean RBM20 R636S (c.1906C>A) substitution and insertions or deletions as “HDR” and “NHEJ” events, respectively, whereas combinations of both in the same allele were defined as “HDR + NHEJ” events (Figures 2A and S3B). HDR + NHEJ events occurred at a relatively low frequency in comparison to NHEJ, and the HDR frequency was even lower (Figure 3A). Based on these HEK293T cell clones with edited alleles, we were able to estimate the overall allelic frequencies of each genome editing event (i.e., WT, NHEJ, HDR, and HDR + NHEJ) (Figure 3B). Thus, our SPiS-based approach enables us to systematically analyze genome editing outcomes in individual cells.

Binary induction of genome editing in single cells

One notable feature in our results was that roughly half of the cells had all their target RBM20 alleles edited by NHEJ (47/128, 62/109, and 44/93, respectively). On the other hand, there were quite a few cells that remained completely unedited (49/128, 26/109, and 24/93, respectively), although these cells had expressed

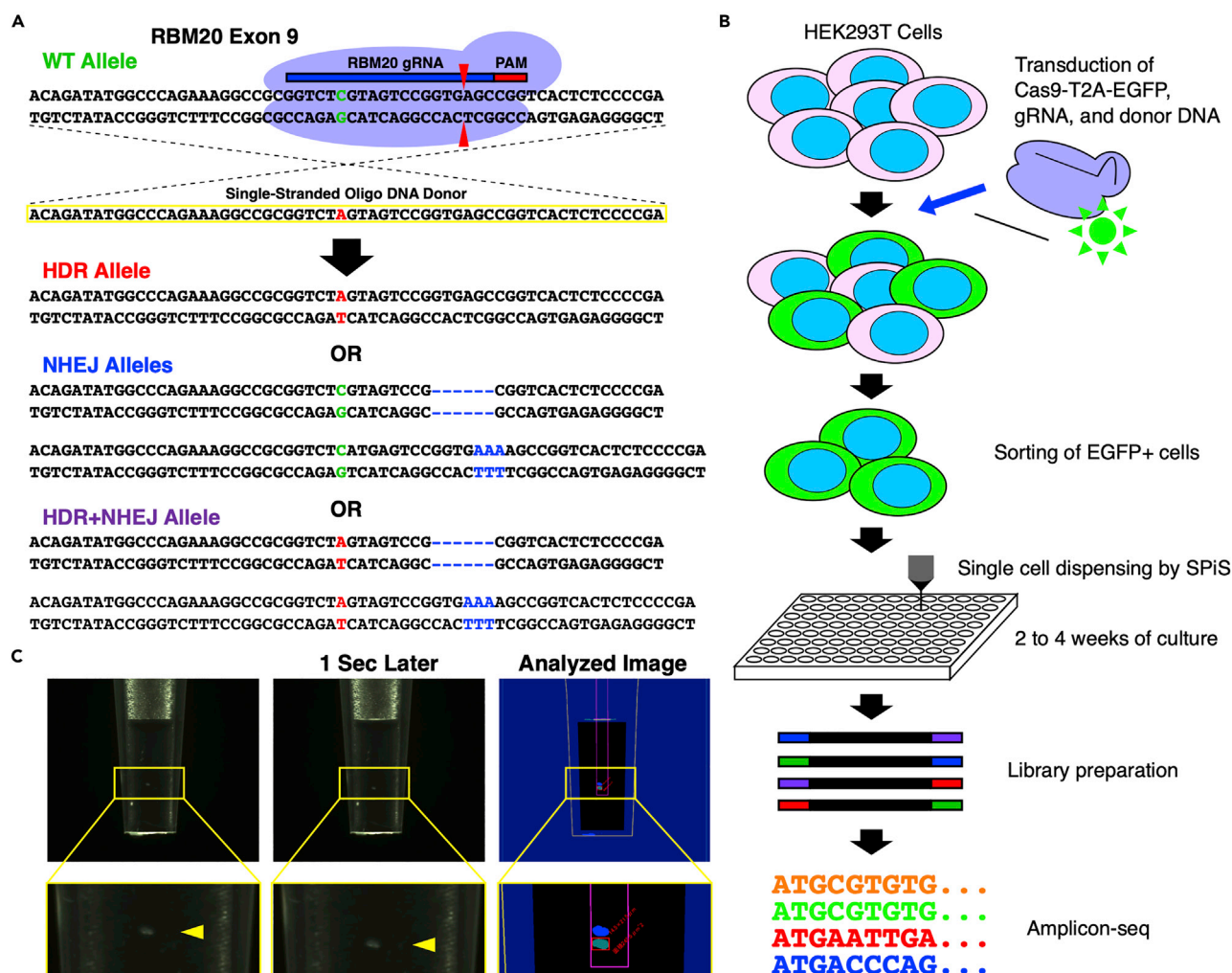


Figure 2. Schematic illustration of the experiments of this study

(A) Design of genome editing in RBM20 as a model case in this study. We introduced the R636S (c.1906C>A) mutation using CRISPR-Cas9 and a single-stranded oligonucleotide donor DNA. The resulting HDR allele has a C to A single nucleotide substitution, whereas the NHEJ alleles have various insertions and deletions.

(B) The experimental flow of this study. We expressed Cas9-T2A-EGFP together with the gRNA targeting RBM20 to label HEK293T cells in which Cas9 protein was expressed. Using a microfluidic cell sorter (On-chip Sort), we sorted EGFP+ cells. Then, the sorted cells were plated into four 96-well plates by the SPiS. Cells were cultured for 2 to 4 weeks and then the genome editing outcomes were analyzed by amplicon sequencing.

(C) An imaging analysis of single cells dispensed by the SPiS. The SPiS aspirates cell suspension into a specialized microtip and takes two images with a 1-s interval. Only when the SPiS recognized a single cell, the content of the tip was dispensed into a well of a 96-well plate. Otherwise, the SPiS would take another aliquot of the cell suspension to repeat the process.

Cas9 and gRNA (Figure 3A). These observations were quite surprising considering the fact that these cells have three or four copies of RBM20 on average (Figure 1).

To address whether this trend is significant, we mathematically calculated the number of HEK293T cell clones with different genome editing outcomes assuming that these events occurred randomly. Because the copy number of RBM20 was 3.55 (Figure 1), we built two models with three or four copies of RBM20 (Figures 3C and S4). In these models, we calculated how many cells with three or four copies of RBM20 should have specific genotypes based on the overall frequencies of WT, NHEJ, HDR, and HDR + NHEJ alleles. We compared the proportions of clones that are expected to be unedited to remain as WT or fully edited by NHEJ between the two models and the actual observation (Figures 3A and 3C). As we expected, the proportions of actually isolated clones were significantly higher than those based on the models for both WT and full NHEJ clones (Figure 3D). In addition, among 330 clones, we observed one clone in which

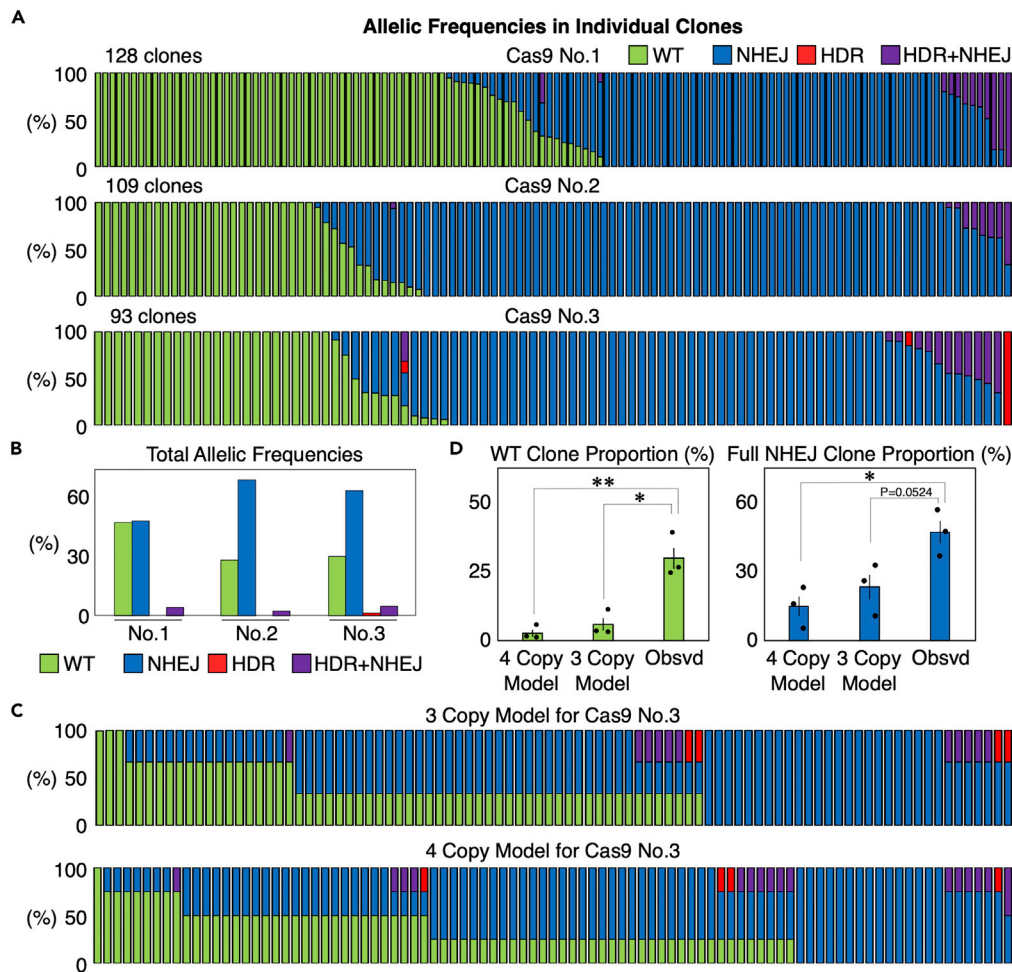


Figure 3. Genome editing outcomes in RBM20 in individual HEK293T cells edited by Cas9

(A) Genome editing outcomes in isolated clones derived from single HEK293T cells edited by Cas9. We repeated the experiment three times, and isolated more than 90 clones out of 384 cells plated in all three trials. Each bar represents one clone, and the proportions of WT (green), NHEJ (blue), HDR (red), and HDR + NHEJ (purple) in one clone are also shown in each bar.

(B) Total allelic frequencies in genome edited HEK293T cells in the three experiments shown in (A).

(C) Models of the distributions of HEK293T cell clones with different genome editing outcomes in the Cas9 No.3 experiment, if genome editing randomly occurred at the frequencies shown in (B) in HEK293T cells with three or four copies of RBM20.

(D) Comparison of the proportions of WT and full NHEJ clones between the mathematical models with three and four copies of RBM20 and the actually observed cells. Values \pm S.E. are shown ($n = 3$). Student's *t* test was used to evaluate differences. * $p < 0.05$ and ** $p < 0.01$.

all target alleles were modified by HDR (Figure 3A). No such clone was expected in the mathematical models (Figures 3C and S4). These results suggested that NHEJ is induced in a binary manner, in which cells show either no editing at all or full editing.

HypaCas9 induces HDR more efficiently than Cas9 in single cells

Because the frequency of HDR was so low with Cas9, we were not able to investigate how HDR is induced in single cells (Figure 3). HDR is often more desirable than NHEJ, as HDR can introduce precise manipulation of the DNA sequence. Therefore, we decided to apply our SPIS-based analysis to genome editing conditions that are more favorable for the induction of HDR. We found that Cas9 with improved proof-reading (e.g., HypaCas9,²⁹ Cas9-HF1,³⁰ and eSpCas9³¹) induced more HDR than Cas9 in our previous cell population-based assay;³² thus, we targeted the same RBM20 R636S mutation in HEK293T cells using HypaCas9,

and conducted the SPiS-based single cells assay three times. We confirmed that HypaCas9 produced more clones with HDR than Cas9 (Figure 4A). We calculated the allelic frequencies of WT, NHEJ, HDR, and HDR + NHEJ (Figure 4B), and compared these between Cas9 and HypaCas9. We found that the HDR allelic frequency was higher with HypaCas9 than with Cas9; however, the WT, NHEJ, and HDR + NHEJ allelic frequencies were comparable (Figure 4C). Based on these overall allelic frequencies, we built mathematical models of HEK293T cell clones with genome editing by HypaCas9 in the same way as for Cas9 (Figures 4D and S5). The binary manner of genome editing induction was also observed with HypaCas9, as the WT and full NHEJ clone proportions were significantly higher than the mathematical models (Figure 4E). Moreover, binary induction was also observed in HDR. Among 422 clones, we isolated a total of six clones with all alleles edited by HDR (Figure 4A). Because the overall frequency of HDR was 5.89% (Figure 4C), no such clone was expected to be isolated if genome editing events occurred randomly in single cells (1 out of 4894 cells would be expected to have having three HDR alleles and 1 out of 83,088 cells would be expected to have four HDR alleles). To exclude the possibility that different genome editing outcomes affected the growth or survival of the cells, we induced the RBM20 R636S mutation in HEK293T cells by HypaCas9 and investigated how the overall frequencies of the WT, NHEJ, and HDR alleles changed after 25 days of continuous passaging and culture. If NHEJ or HDR gave advantage or disadvantage in growth or survival to the cells, we would observe a significant increase or decrease in the frequency of the NHEJ or HDR alleles. However, we observed only slight increase in the WT allelic frequency (75.8%–81.6%, Figure S6). Both the NHEJ (21.8–17%) and HDR (2.3%–1.2%) frequencies slightly decreased, the difference was not statistically significant (Figure S6). Therefore, we concluded that the influence of different genome editing outcomes on cell growth or survival was marginal if any. Thus, we confirmed the binary induction of genome editing, even with HypaCas9, which induces more HDR than Cas9. The binary mode was observed in the induction of both HDR and NHEJ.

HDR is more often accompanied by NHEJ than WT

As described above, we identified more clones fully edited by HDR than mathematically expected if the induction of HDR had been random (Figure 4). However, we noticed that clones partially edited by HDR often had NHEJ alleles together. Therefore, we calculated the proportions of clone that would be expected to have only HDR and WT alleles, as these clones represent partial editing by HDR. Even though 2.5–3.5% of the cells were expected to have only HDR and WT alleles, we isolated no such clones (Figures 4A and 4F). We further investigated the expected and observed the proportions of clones with both HDR and NHEJ events, including any clones with the HDR + NHEJ alleles, as they represent HDR and NHEJ events induced in the same alleles. As we expected, the proportions of clones with both HDR and NHEJ events that were observed and those that were predicted by the models were comparable (Figure 4F). These results coincide with the fact that cells tend to have all target alleles edited when genome editing was induced, so partial editing is rare (Figure 4E). Moreover, HDR and NHEJ are often induced together in single cells (Figure 4F).

Genome editing is also induced in a binary manner in ATP7B and GRN

We next addressed whether the observed binary fashion of genome editing stands in other target sites. Therefore, we targeted ATP7B and GRN to introduce R778L (c.2333G>T, chr.13)³³ and R493X (c.1477C>T, chr.17)³⁴ mutations, respectively into HEK293T cells as described previously.¹⁸ HypaCas9 and single stranded donor DNA were used in the same way as for RBM20 R636S mutagenesis (Figures S7A and S7B). We isolated clones and performed amplicon sequencing to quantify the frequency of genome editing (Figures 5A and 5D). To build models for genome editing of ATP7B and GRN, we quantified the copy numbers of these genes, which were determined to be 2.88 for ATP7B and 3.54 for GRN (Figures 1, S7C, and S7D). We then calculated the expected clone numbers with the assumption that genome editing events occurred randomly in the cells to build models for genome editing of ATP7B and GRN (Figures S8A and S8B), and compared them to the observed results. Similar to the RBM20 R636S mutagenesis, the number of clones in which all target alleles were edited by NHEJ or not edited at all was significantly higher in comparison to the models (Figures 5B and 5E). We also isolated clones in which all target alleles were edited by HDR (8/280 and 2/276 for ATP7B and GRN, respectively) (Figures 5A and 5D). This was particularly noteworthy for GRN, as the overall efficiency of HDR induction was only 1.02% (Figure 5D). Collectively, these results demonstrate that the binary nature of genome editing is not restricted to a specific target site. Moreover, we did not isolate any clones with partial HDR editing in ATP7B or GRN; however, HDR and NHEJ were often induced together in single cells, similarly to in RBM20 (Figures 5A, 5C, and 5D).

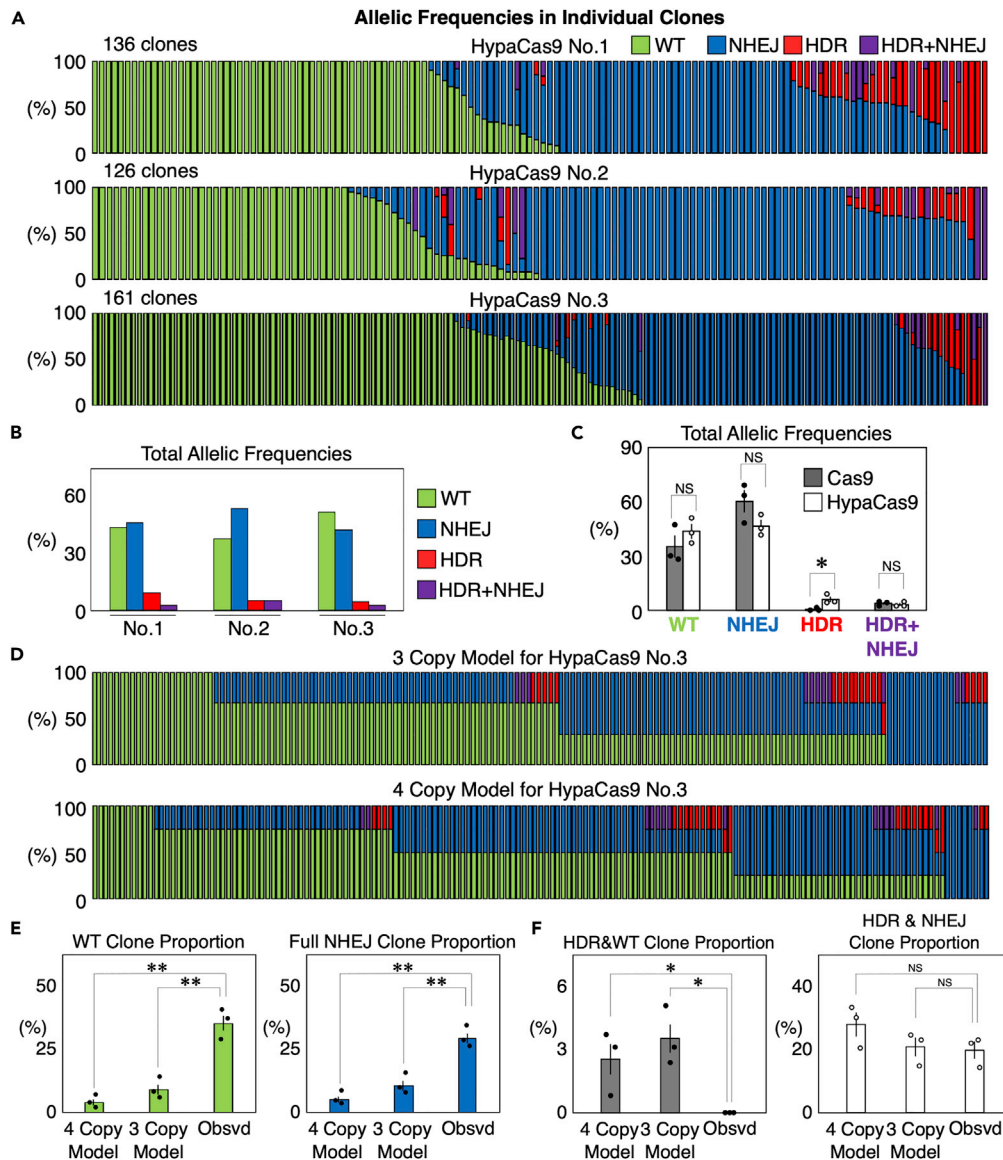


Figure 4. Genome editing outcomes in RBM20 in individual HEK293T cells edited by HypaCas9

(A) Genome editing outcomes in isolated clones derived from single HEK293T cells edited by HypaCas9 and the single-stranded donor DNA. We repeated the same experiment three times. Each bar represents one clone, and the proportions of WT (green), NHEJ (blue), HDR (red), and HDR + NHEJ (purple) in one clone are also shown in each bar.

(B) Total frequencies of WT (green), NHEJ (blue), HDR (red), and HDR + NHEJ (purple) allelic frequencies in genome edited HEK293T cells in the three experiments shown in (A).

(C) Comparison of total allelic frequencies of WT, NHEJ, HDR, and HDR + NHEJ between Cas9 and HypaCas9.

Values \pm S.E. are shown ($n = 3$). Student's t test was used to evaluate differences. * $p < 0.05$ and NS: not significantly different ($p > 0.1$).

(D) Models of the distributions of HEK293T cell clones with different genome editing outcomes in the HypaCas9 No.3 experiment, if genome editing randomly occurred at the frequencies shown in (B) in HEK293T cells with three or four copies of RBM20.

(E) Comparison of the proportions of WT and full NHEJ clones between the mathematical models with three and four copies of RBM20 and the actually observed cells. Values \pm S.E. are shown ($n = 3$). Student's t test was used to evaluate differences. ** $p < 0.01$.

(F) Comparison of the proportions of clones with partial editing by HDR and clones with HDR accompanied by NHEJ events between the models with three and four copies of RBM20 and the actually observed cells. Values \pm S.E. are shown ($n = 3$). Student's t test was used to evaluate differences. * $p < 0.05$ and NS: not significantly different ($p > 0.1$).

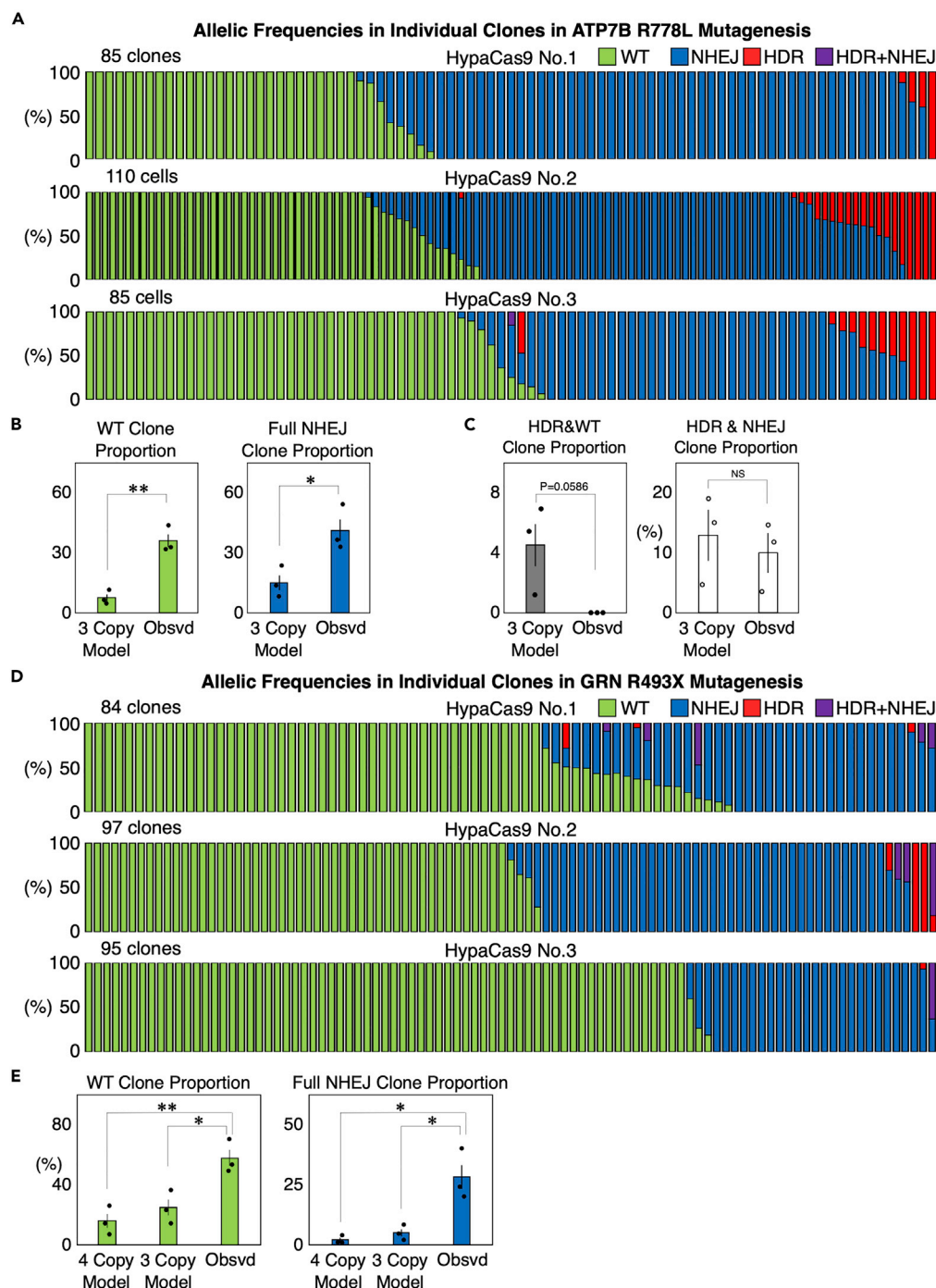


Figure 5. Genome editing outcomes in ATP7B and GRN in individual HEK293T cells edited by HypaCas9

(A) Genome editing outcomes in isolated clones derived from single HEK293T cells edited by HypaCas9 and the single-stranded donor DNA targeting ATP7B. We repeated the same experiment three times. Each bar represents one clone, and the proportions of WT (green), NHEJ (blue), HDR (red), and HDR + NHEJ (purple) in one clone are also shown in each bar. (B) Comparison of the proportions of WT and full NHEJ clones between the mathematical models with three copies of ATP7B and the actually observed cells. Values \pm S.E. are shown (n = 3). Student's t test was used to evaluate differences. *p < 0.05 and **p < 0.01.

(C) Comparison of the proportions of clones with partial editing by HDR and clones with HDR accompanied by NHEJ events between the models with three copies of ATP7B and the actually observed cells. Values \pm S.E. are shown (n = 3). Student's t test was used to evaluate differences. NS: not significantly different (p > 0.1).

Figure 5. Continued

(D) Genome editing outcomes in isolated clones derived from single HEK293T cells edited by HypaCas9 and the single-stranded donor DNA targeting GRN. We repeated the same experiment three times. Each bar represents one clone, and proportions of WT (green), NHEJ (blue), HDR (red), and HDR + NHEJ (purple) in one clone are also shown in each bar. (E) Comparison of the proportions of WT and full NHEJ clones between the mathematical models with three and four copies of GRN and the actually observed cells. Values \pm S.E. are shown ($n = 3$). Student's *t* test was used to evaluate differences. * $p < 0.05$ and ** $p < 0.01$.

Binary genome editing is also prominent in HeLa cells

We further addressed whether the binary induction of genome editing occurs in other cell types. For this purpose, we introduced the RBM20 R636S (c.1906C>A), ATP7B R778L (c.2333G>T), and GRN R493X (c.1477C>T) mutations into HeLa cells in the same way as for HEK293T cells. We were able to apply our SPiS-based system to isolated HeLa cell clones that had gone through the genome editing process, although HeLa cells exhibited lower cell survival in comparison to HEK293T cells (Figures 6A–6G). We measured the copy numbers of these target genes by combining the karyotyping and CGH analyses in HeLa cells (Figure S9 and Table S3). Based on these copy numbers, mathematical models of clones with different genotypes were built in the same way as for HEK293T cells (Figure S10). We found that the observed proportions of clones that were unedited or fully edited by NHEJ were significantly higher than in the calculated models based on the overall allelic frequencies in all three genes (Figures 6B, 6D, and 6G). The HDR frequency was higher in ATP7B than in RBM20 or GRN in HeLa cells, and we isolated 10 clones fully edited by HDR out of 234 clones (Figure 6C). In ATP7B, the proportion of clones with partial editing by HDR was significantly lower in comparison to the model, whereas that of clones with HDR accompanied by NHEJ was comparable to the model (Figure 6E). These results indicate that the binary nature of genome editing induction is shared between HeLa cells and HEK293T cells.

Binary genome editing is less evident in PC9 cells

Finally, we investigated the binary nature of genome editing in another cell line, PC9 cells. We introduced the same three mutations into PC9 cells and conducted our SPiS-based analysis (Figures 7A–7F). We noticed that the genome editing efficiency in PC9 cells was generally lower than that in HEK293T cells or HeLa cells, and that HDR was barely induced (Figures 7A, 7C and 7E). We measured the copy numbers of the target genes in PC9 cells (Figure S11 and Table S4). Based on these copy numbers, the mathematical models of clones with different genotypes were built for PC9 cells (Figure S12). We compared the proportion of clones that were unedited or fully edited by NHEJ between the observations and the models. We found that the trend regarding the binary induction of genome editing was noticeable but less evident, as the proportions of clones fully edited by NHEJ in RBM20 and clones unedited in ATP7B were the only significant differences between the observed clones and the models (Figures 7B, 7D, and 7F). Therefore, binary induction is a general feature of genome editing; however, the extent of binary induction can vary in different contexts.

DISCUSSION

Most of the current methods used to analyze genome editing outcomes deal with cell populations. In this study, we were able to isolate single cell clones with an unprecedented efficiency by the SPiS. The SPiS not only increased the number of clones that we were able to isolate, but also ensured that the isolated clones were derived from single cells based on the image analysis, which avoided plating multiple cells together. This original system allowed us to investigate genome editing outcomes in single cells.

We were surprised to observe that most HEK293T cells were either unedited WT or had all targeted alleles fully manipulated by Cas9. The number of partially edited cells that had both unedited WT alleles and edited alleles was limited (Figure 3). We previously reported that Cas9 variants with enhanced proof-reading are more efficient at inducing HDR than regular Cas9.³² We confirmed that HypaCas9 produced more HDR in single cells than Cas9, and found that HypaCas9 also induced genome editing in a binary fashion (Figure 4D). This trend was shared among all RBM20, ATP7B, and GRN genes in both HEK293T cells and HeLa cells, and to a lesser extent in PC9 cells (Figures 4, 5, 6, and 7). Because we sorted cells that expressed EGFP together with Cas9 and gRNA, these WT clones escaped from DNA cleavage by Cas9, or precisely repaired the genomic DNA after DNA cleavage. We speculate that each cell has an intracellular environment that is favorable for a particular genome editing outcome,

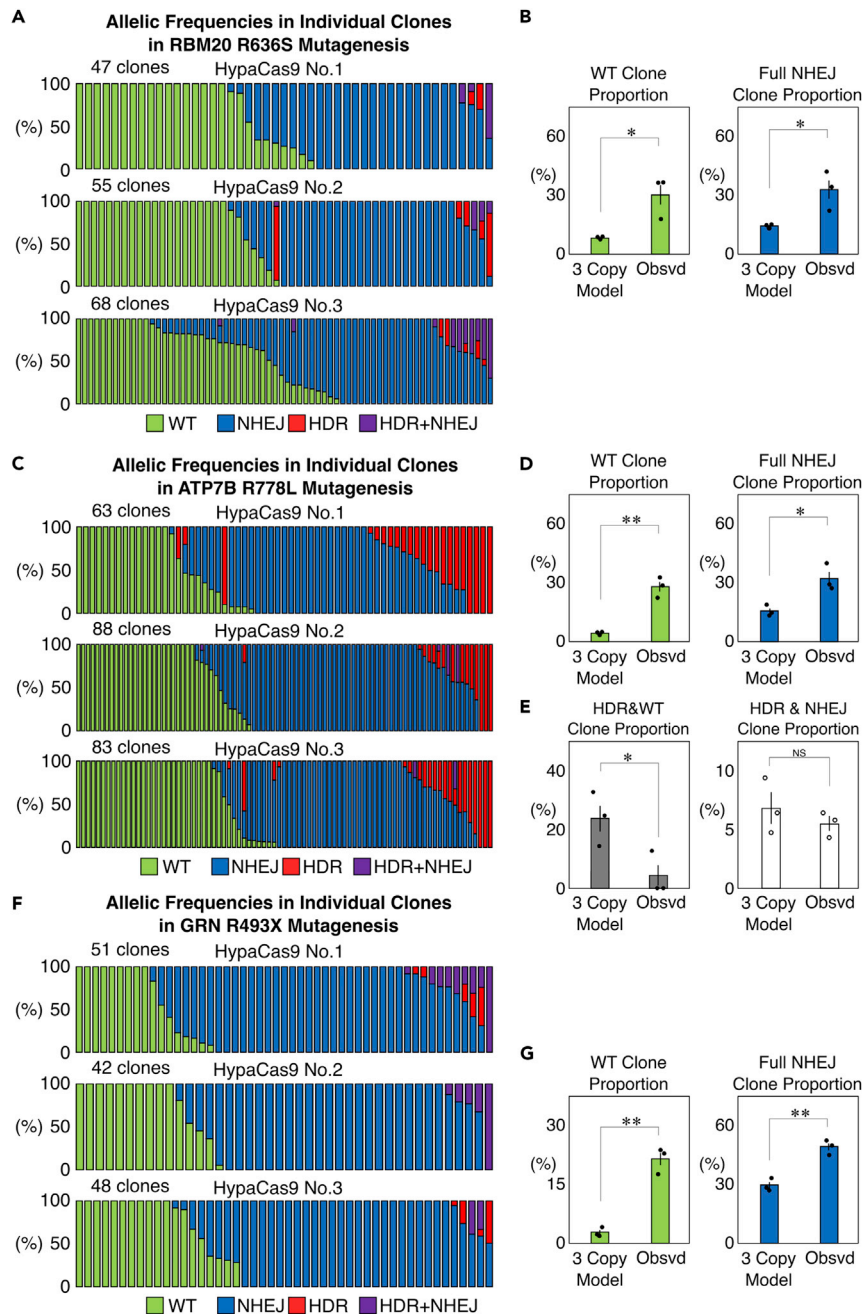


Figure 6. Genome editing outcomes in individual HeLa cells

(A) Genome editing outcomes in isolated clones derived from single HeLa cells edited by HypaCas9 and the single-stranded donor DNA targeting RBM20. Each bar represents one clone, and the proportions of WT (green), NHEJ (blue), HDR (red), and HDR + NHEJ (purple) in one clone are also shown in each bar.

(B) Comparison of the proportions of WT and full NHEJ clones between the mathematical models with three copies of RBM20 and the actually observed cells. Values \pm S.E. are shown ($n = 3$). Student's t test was used to evaluate differences. * $p < 0.05$.

(C) Genome editing outcomes in isolated clones derived from single HeLa cells edited by HypaCas9 and the single-stranded donor DNA targeting ATP7B. Each bar represents one clone, and the proportions of WT (green), NHEJ (blue), HDR (red), and HDR + NHEJ (purple) in one clone are also shown in each bar.

(D) Comparison of the proportions of WT and full NHEJ clones between the mathematical models with three copies of ATP7B and the actually observed cells. Values \pm S.E. are shown ($n = 3$). Student's t test was used to evaluate differences. * $p < 0.05$ and ** $p < 0.01$.

Figure 6. Continued

(E) Comparison of the proportions of clones with partial editing by HDR and clones with HDR accompanied by NHEJ events between the models with three copies of ATP7B and the actually observed cells. Values \pm S.E. are shown ($n = 3$). Student's t test was used to evaluate differences. * $p < 0.05$. NS: not significantly different ($p > 0.1$).

(F) Genome editing outcomes in isolated clones derived from single HeLa cells edited by HypaCas9 and the single-stranded donor DNA targeting GRN. Each bar represents one clone, and the proportions of WT (green), NHEJ (blue), HDR (red), and HDR + NHEJ (purple) in one clone are also shown in each bar.

(G) Comparison of the proportions of WT and full NHEJ clones between the mathematical models with three copies of GRN and the actually observed cells. Values \pm S.E. are shown ($n = 3$). Student's t test was used to evaluate differences. ** $p < 0.01$.

so that all target alleles in one cell tend to have the same fate. The average copy numbers of RBM20, ATP7B, and GRN in HEK293T cells, HeLa cells, and PC9 cells were 2.88–3.71 (Figures 1, S1, S9, and S11). The fact that some cells have full HDR or HDR + NHEJ events in roughly 3 to 4 copies despite the much lower overall frequencies also suggests that target alleles in one cell tend to have the same fate (Figures 4, 5, and 6). This binary nature of genome editing can be beneficial if homozygous mutagenesis is necessary. Indeed, we previously reported the isolation of a human iPS cell line with a homozygous PHOX2B Y14Xmutation generated by TALENs, even though the overall induction efficiency of HDR was just approximately 1%.¹⁹ At the same time, however, it makes heterozygous mutagenesis extremely difficult, as we observed that HDR was more often accompanied by NHEJ rather than unedited WT alleles (Figures 4, 5, and 6). A similar trend was reported in mouse Ba/F3 cells,²³ so the binary manner of genome editing induction may be shared among different species, at least among mammals.

At this point, however, the factors that determine which DNA repair pathway a single cell takes remain unknown. One possibility is that the cells respond differently to Cas9 cleavage at different points in the cell cycle. It is known that the frequency of HDR and NHEJ increases in the S/G2 and G1 phases, respectively.^{35,36} Therefore, it would be interesting to address whether the cell cycle also influences this binary choice between full editing and no editing in single cells.

PC9 cells showed a much lower efficiency of overall induction of genome editing than HEK293T cells and HeLa cells (Figure 7). The binary fashion of induction of genome editing was also less evident in PC9 cells than in HEK293T cells and HeLa cells. This could be due to the difference in the expression of HypaCas9 protein. We observed a relatively low gene transduction efficiency and low expression level in PC9 cells in comparison to the other two cell lines based on the expression level of EGFP, which was co-expressed with Cas9 via the T2-peptide (Figure S13). Therefore, it is possible that certain levels of Cas9 expression and activity are necessary for the binary induction of genome editing. Another possibility is that different DNA repair pathways are active in different cell types to yield different genome editing outcomes. In PC9 cells, HDR was barely induced even with HypaCas9, which could be because the HDR pathway is not very active in PC9 cells (Figures 7A, 7C, and 7E). An investigation of the active DNA repair pathways in these cell lines would be an interesting way to address the observed differences in genome editing outcomes.

We are currently applying the same analytic procedure to other cell types to further address the generality of our findings. Our target cell types include human iPS cells, because we are highly interested in the genome editing outcomes of normal diploid human cells. However, the biggest challenge is the low survival rate of iPS cells, especially in single cell culture;³⁷ thus, we are optimizing the culture conditions. We can also apply different transduction methods, including lipofection, electroporation of plasmids and RNPs, and viral infection. We will test these various parameters by our SPiS-based strategy to figure out general characteristics of genome editing outcomes in single cells, and also best conditions to induce specific types of genome editing.

Our study reveals the previously unknown but fundamental features of genome editing in single cells including the “binary nature” of genome editing induction. This was only possible with the SPiS. Our findings contribute to the better understanding of the underlying mechanism of induction of genome editing. Moreover, NHEJ often results in gene disruption; thus, the gene functions in a cell may be lost in clones fully edited by NHEJ. This can be very important in genome editing therapy that is dependent on HDR, as by-products of NHEJ might hamper the therapy if the function of a certain gene is completely lost in some cells. Therefore, the analysis of genome editing results in single cells is critical in precise evaluation of

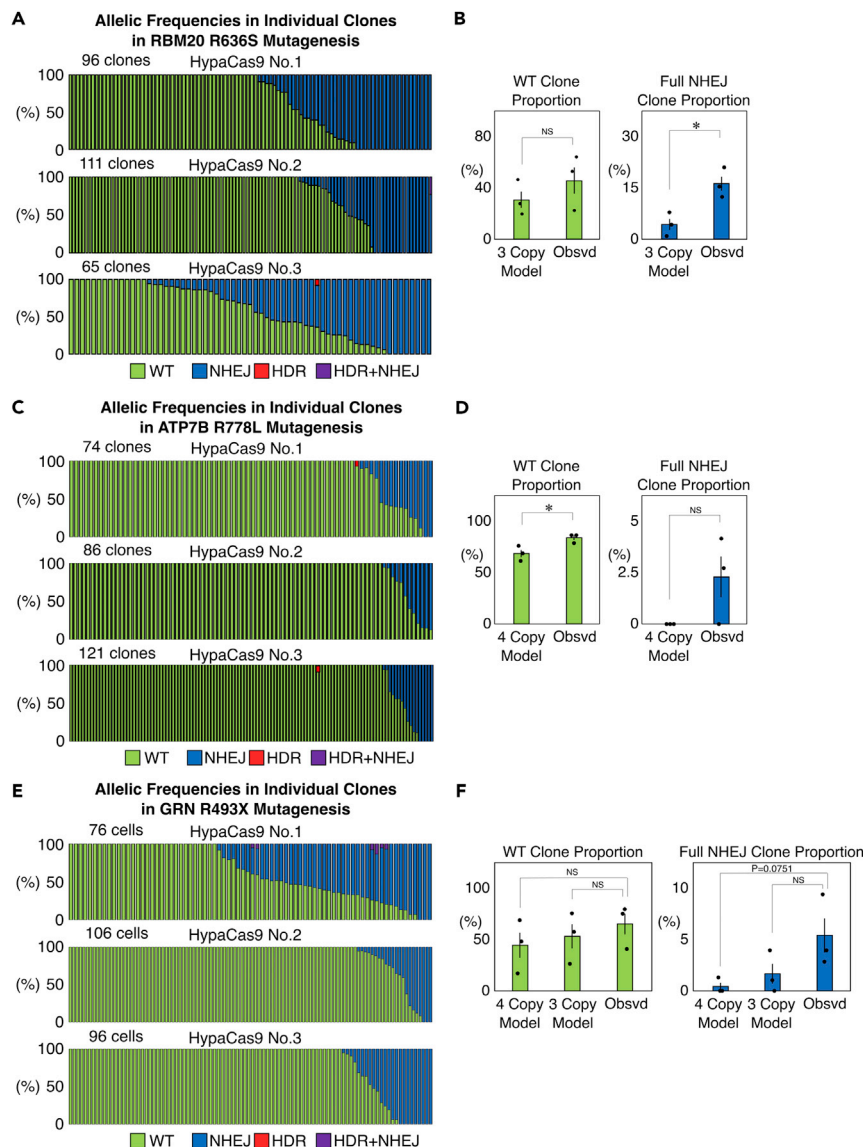


Figure 7. Genome editing outcomes in individual PC9 cells

(A) Genome editing outcomes in isolated clones derived from single PC9 cells edited by HypaCas9 and the single-stranded donor DNA targeting RBM20. Each bar represents one clone, and the proportions of WT (green), NHEJ (blue), HDR (red), and HDR + NHEJ (purple) in one clone are also shown in each bar.

(B) Comparison of the proportions of WT and full NHEJ clones between the mathematical models with three copies of RBM20 and the actually observed cells. Values \pm S.E. are shown ($n = 3$). Student's t test was used to evaluate the difference. * $p < 0.05$. NS: not significantly different ($p > 0.1$).

(C) Genome editing outcomes in isolated clones derived from single PC9 cells edited by HypaCas9 and the single-stranded donor DNA targeting ATP7B. Each bar represents one clone, and the proportions of WT (green), NHEJ (blue), HDR (red), and HDR + NHEJ (purple) in one clone are also shown in each bar.

(D) Comparison of the proportions of WT and full NHEJ clones between the mathematical models with three copies of ATP7B and the actually observed cells. Values \pm S.E. are shown ($n = 3$). Student's t test was used to evaluate differences. * $p < 0.05$. NS: not significantly different ($p > 0.1$).

(E) Genome editing outcomes in isolated clones derived from single PC9 cells edited by HypaCas9 and the single-stranded donor DNA targeting GRN. Each bar represents one clone, and the proportions of WT (green), NHEJ (blue), HDR (red), and HDR + NHEJ (purple) in one clone are also shown in each bar.

(F) Comparison of the proportions of WT and full NHEJ clones between the mathematical models with three copies of GRN and the actually observed cells. Values \pm S.E. are shown ($n = 3$). Student's t test was used to evaluate differences. NS: not significantly different ($p > 0.1$).

therapeutic effects. Our SPiS also provides researchers with a versatile platform to study genome editing in single cells.

Limitations of the study

In this study, we found that genome editing is induced in a binary manner in single human cells. However, further study is required to clarify whether normal diploid cells (e.g., iPS cells and human primary cultured cells) have the same binary trend.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
 - Lead contact
 - Materials availability
 - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
 - Cell lines
- METHOD DETAILS
 - Plasmids and oligonucleotides for transfection
 - Calculation of the copy number of target genes by karyotyping and CGH analysis
 - Transfection and cell sorting
 - Preparation of multiplexed amplicon sequencing library
 - Amplicon sequencing data analysis
 - Time course analysis of allelic frequencies
 - Creating a model in which alleles were evenly distributed
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.isci.2022.105619>.

ACKNOWLEDGMENTS

This work was supported by Japan Society for the Promotion of Science (JSPS) Grant-in-Aid for Challenging Research (Pioneering) 20K21409, Grant-in-Aid for Scientific Research (B) 20H03442, Interstellar Initiative from Japan Agency for Medical Research and development (AMED) (20jm0610032h0001), Takeda Science Foundation Medical Research Grant, Sumitomo Foundation Grant For Basic Science Research Projects, Ichiro Kanehara Foundation Grant to Y.M., and a JSPS Grant-in-Aid for Early-Career Scientists (18K15054 and 22K15386) to G.T. We thank Dr. Y. Nishito (Center for Basic Technology Research) for his technical help in the CGH analysis. We thank Dr. N. Yachie (University of British Columbia) for his technical help in the amplicon sequencing analysis. We also thank Drs. Y. Fujimura and K. Takeda (On-chip Biotechnologies) and all the lab members for their helpful discussions.

AUTHOR CONTRIBUTIONS

G.T. and Y.M. conceived the idea of this study and designed the experiments. G.T., D.K., and M.M. isolated genome edited HEK293T cell clones with technical help from Y.M. D.K. and M.M. prepared libraries for amplicon sequencing. G.T., D.K., M.M., and Y.M. analyzed the results. Y.M. supervised the project. G.T. and Y.M. wrote the manuscript with help from all of the authors.

DECLARATION OF INTERESTS

Y.M. is an employee of On-chip Biotechnologies Co., Ltd.

Received: May 15, 2022

Revised: October 19, 2022

Accepted: November 15, 2022

Published: December 22, 2022

REFERENCES

- Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science* 339, 819–823. <https://doi.org/10.1126/science.1231143>.
- Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337, 816–821. <https://doi.org/10.1126/science.1225829>.
- Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. (2013). RNA-guided human genome engineering via Cas9. *Science* 339, 823–826. <https://doi.org/10.1126/science.1232033>.
- Gaj, T., Gersbach, C.A., and Barbas, C.F. (2013). ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol.* 31, 397–405. <https://doi.org/10.1016/j.tibtech.2013.04.004>.
- Yeh, C.D., Richardson, C.D., and Corn, J.E. (2019). Advances in genome editing through control of DNA repair pathways. *Nat. Cell Biol.* 21, 1468–1478. <https://doi.org/10.1038/s41556-019-0425-z>.
- Nishida, K., Arazoe, T., Yachie, N., Banno, S., Kakimoto, M., Tabata, M., Mochizuki, M., Miyabe, A., Araki, M., Hara, K.Y., et al. (2016). Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science* 353, aaf8729. <https://doi.org/10.1126/science.aaf8729>.
- Gaudelli, N.M., Komor, A.C., Rees, H.A., Packer, M.S., Badran, A.H., Bryson, D.I., and Liu, D.R. (2017). Programmable base editing of A·T to G·C in genomic DNA without DNA cleavage. *Nature* 551, 464–471. <https://doi.org/10.1038/nature24644>.
- Komor, A.C., Kim, Y.B., Packer, M.S., Zuris, J.A., and Liu, D.R. (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* 533, 420–424. <https://doi.org/10.1038/nature17946>.
- Anzalone, A.V., Randolph, P.B., Davis, J.R., Sousa, A.A., Koblán, L.W., Levy, J.M., Chen, P.J., Wilson, C., Newby, G.A., Raguram, A., et al. (2019). Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* 576, 149–157. <https://doi.org/10.1038/s41586-019-1711-4>.
- Abadi, S., Yan, W.X., Amar, D., and Mayrose, I. (2017). A machine learning approach for predicting CRISPR-Cas9 cleavage efficiencies and patterns underlying its mechanism of action. *PLoS Comput. Biol.* 13, e1005807. <https://doi.org/10.1371/journal.pcbi.1005807>.
- Arbab, M., Shen, M.W., Mok, B., Wilson, C., Matuszek, Z., Cassa, C.A., and Liu, D.R. (2020). Determinants of base editing outcomes from target library analysis and machine learning. *Cell* 182, 463–480.e30. <https://doi.org/10.1016/j.cell.2020.05.037>.
- Marquart, K.F., Allam, A., Janjuha, S., Sintsova, A., Villiger, L., Frey, N., Krauthammer, M., and Schwank, G. (2021). Predicting base editing outcomes with an attention-based deep learning algorithm trained on high-throughput target library screens. *Nat. Commun.* 12, 5114. <https://doi.org/10.1038/s41467-021-25375-z>.
- Song, M., Kim, H.K., Lee, S., Kim, Y., Seo, S.-Y., Park, J., Choi, J.W., Jang, H., Shin, J.H., Min, S., et al. (2020). Sequence-specific prediction of the efficiencies of adenine and cytosine base. *Nat. Biotechnol.* 38, 1037–1043. <https://doi.org/10.1038/s41587-020-0573-5>.
- Germini, D., Tsfasman, T., Zakharova, V.V., Sjakste, N., Lipinski, M., and Vassetzky, Y. (2018). A comparison of techniques to evaluate the effectiveness of genome editing. *Trends Biotechnol.* 36, 147–159. <https://doi.org/10.1016/j.tibtech.2017.10.008>.
- Brinkman, E.K., Chen, T., Amendola, M., and van Steensel, B. (2014). Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res.* 42, e168. <https://doi.org/10.1093/nar/gku936>.
- Mashal, R.D., Koontz, J., and Sklar, J. (1995). Detection of mutations by cleavage of DNA heteroduplexes with bacteriophage resolvases. *Nat. Genet.* 9, 177–183. <https://doi.org/10.1038/ng0295-177>.
- Inazuka, M., Wenz, H.M., Sakabe, M., Tahira, T., and Hayashi, K. (1997). A streamlined mutation detection system: multicolor post-PCR fluorescence labeling and single-strand conformational polymorphism analysis by capillary electrophoresis. *Genome Res.* 7, 1094–1103. <https://doi.org/10.1101/gr.7.11.1094>.
- Miyaoka, Y., Berman, J.R., Cooper, S.B., Mayerl, S.J., Chan, A.H., Zhang, B., Karlin-Neumann, G.A., and Conklin, B.R. (2016). Systematic quantification of HDR and NHEJ reveals effects of locus, nuclease, and cell type on genome-editing. *Sci. Rep.* 6, 23549. <https://doi.org/10.1038/srep23549>.
- Miyaoka, Y., Chan, A.H., Judge, L.M., Yoo, J., Huang, M., Nguyen, T.D., Lizarraga, P.P., So, P.-L., and Conklin, B.R. (2014). Isolation of single-base genome-edited human iPS cells without antibiotic selection. *Nat. Methods* 11, 291–293. <https://doi.org/10.1038/nmeth.2840>.
- Certo, M.T., Ryu, B.Y., Annis, J.E., Garibov, M., Jarjour, J., Rawlings, D.J., and Scharenberg, A.M. (2011). Tracking genome engineering outcome at individual DNA breakpoints. *Nat. Methods* 8, 671–676. <https://doi.org/10.1038/nmeth.1648>.
- Aldridge, S., and Teichmann, S.A. (2020). Single cell transcriptomics comes of age. *Nat. Commun.* 11, 4307. <https://doi.org/10.1038/s41467-020-18158-5>.
- Islam, S., Kjällquist, U., Moliner, A., Zajac, P., Fan, J.-B., Lönnberg, P., and Linnarsson, S. (2011). Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. *Genome Res.* 21, 1160–1167. <https://doi.org/10.1101/gr.110882.110>.
- Ten Hacken, E., Clement, K., Li, S., Hernández-Sánchez, M., Redd, R., Wang, S., Ruff, D., Gruber, M., Baranowski, K., Jacob, J., et al. (2020). High throughput single-cell detection of multiplex CRISPR-edited gene modifications. *Genome Biol.* 21, 266. <https://doi.org/10.1186/s13059-020-02174-1>.
- Koike, S. (2020). The risk of unintentional propagation of poliovirus can be minimized by using human cell lines lacking the functional CD155 gene. *Microbiol. Immunol.* 64, 835–839. <https://doi.org/10.1111/1348-0421>.
- Saito, K., Ota, Y., Tourlousse, D.M., Matsukura, S., Fujitani, H., Morita, M., Tsuneda, S., and Noda, N. (2021). Microdroplet-based system for culturing of environmental microorganisms using FNAP-sort. *Sci. Rep.* 11, 9506. <https://doi.org/10.1038/s41598-021-88974-2>.
- Brauch, K.M., Karst, M.L., Herron, K.J., de Andrade, M., Pellikka, P.A., Rodeheffer, R.J., Michels, V.V., and Olson, T.M. (2009). Mutations in ribonucleic acid binding protein gene cause familial dilated cardiomyopathy. *J. Am. Coll. Cardiol.* 54, 930–941. <https://doi.org/10.1016/j.jacc.2009.05.038>.
- Takeda, K., Fujimura, Y., and Jimma, F. (2011). Development of a new compact cytometer using a disposable microfluidic chip for contamination-free and biosafety measurement. *Cytometry Res.* 21, 43–50.
- Clement, K., Rees, H., Canver, M.C., Gehrke, J.M., Farouni, R., Hsu, J.Y., Cole, M.A., Liu, D.R., Joung, J.K., Bauer, D.E., et al. (2019). CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nat. Biotechnol.* 37, 224–226. <https://doi.org/10.1038/s41587-019-0032-3>.
- Chen, J.S., Dagdas, Y.S., Kleinstiver, B.P., Welch, M.M., Sousa, A.A., Harrington, L.B., Sternberg, S.H., Joung, J.K., Yildiz, A., and Doudna, J.A. (2017). Enhanced proofreading governs CRISPR-Cas9 targeting accuracy. *Nature* 550, 407–410. <https://doi.org/10.1038/nature24268>.
- Kleinstiver, B.P., Pattanayak, V., Prew, M.S., Tsai, S.Q., Nguyen, N.T., Zheng, Z., and Joung, J.K. (2016). High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* 529, 490–495. <https://doi.org/10.1038/nature16526>.
- Slaymaker, I.M., Gao, L., Zetsche, B., Scott, D.A., Yan, W.X., and Zhang, F. (2016). Rationally engineered Cas9 nucleases with improved specificity. *Science* 351, 84–88. <https://doi.org/10.1126/science.1232757>.
- Kato-Inui, T., Takahashi, G., Hsu, S., and Miyaoka, Y. (2018). Clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated protein 9 with improved proof-reading enhances homology-directed repair. *Nucleic Acids Res.* 46, 4677–4688. <https://doi.org/10.1093/nar/gky264>.

33. Ferenci, P. (2006). Regional distribution of mutations of the ATP7B gene in patients with Wilson disease: impact on genetic testing. *Hum. Genet.* 120, 151–159. <https://doi.org/10.1007/s00439-006-0202-5>.
34. Chen-Plotkin, A.S., Martinez-Lage, M., Sleiman, P.M.A., Hu, W., Greene, R., Wood, E.M., Bing, S., Grossman, M., Schellenberg, G.D., Hatanpaa, K.J., et al. (2011). Genetic and clinical features of progranulin-associated frontotemporal lobar degeneration. *Arch. Neurol.* 68, 488–497. <https://doi.org/10.1001/archneurol.2011.53>.
35. Lin, S., Staahl, B.T., Alla, R.K., and Doudna, J.A. (2014). Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *Elife* 3, e04766. <https://doi.org/10.1038/nature17946>.
36. Wienert, B., Nguyen, D.N., Guenther, A., Feng, S.J., Locke, M.N., Wyman, S.K., Shin, J., Kazane, K.R., Gregory, G.L., Carter, M.A.M., et al. (2020). Timed inhibition of CDC7 increases CRISPR-Cas9 mediated templated repair. *Nat. Commun.* 11, 2109. <https://doi.org/10.1038/s41467-020-15845-1>.
37. Watanabe, K., Ueno, M., Kamiya, D., Nishiyama, A., Matsumura, M., Wataya, T., Takahashi, J.B., Nishikawa, S.I., Nishikawa, S., Muguruma, K., et al. (2007). A ROCK inhibitor permits survival of dissociated human embryonic stem cells. *Nat. Biotechnol.* 25, 681–686. <https://doi.org/10.1038/nbt1310>.
38. Sakata, R.C., Ishiguro, S., Mori, H., Tanaka, M., Tatsuno, K., Ueda, H., Yamamoto, S., Seki, M., Masuyama, N., Nishida, K., et al. (2020). Base editors for simultaneous introduction of C-to-T and A-to-G mutations. *Nat. Biotechnol.* 38, 865–869. <https://doi.org/10.1038/s41587-020-0509-0>.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, peptides, and recombinant proteins		
Betaine Solution	FUJIFILM Wako Pure Chemical Corporation	Cat#028-16294
Critical commercial assays		
Lipofectamine™ 2000 Transfection Reagent	Thermo Fisher Scientific	Cat#11668027
Lipofectamine™ 3000 Transfection Reagent	Thermo Fisher Scientific	Cat#L3000001
KOD FX	Toyobo	Cat#KFX-101
GenNext NGS Library Quantification Kit	Toyobo	Cat#NLQ-101
DNeasy Blood and Tissue Kit	Qiagen	Cat#69504
SurePrint G3 Human CGH Microarray Kit	Agilent	Cat#G4449A
NucleoSpin Gel and PCR Clean-up Midi kit	MACHEREY-NAGEL	Cat#740986.20
PhiX Control v3	Illumina	Cat#FC-110-3001
Miseq reagent kit v2	Illumina	Cat#MS-102-2003
Miseq Reagent kit v2 Nano	Illumina	Cat#MS-103-1003
ddPCR™ Supermix for Probes (No dUTP)	Bio-Rad	Cat#1863024
Deposited data		
CGH array data	https://www.ncbi.nlm.nih.gov/geo/	GSE217534
Amplicon sequencing raw data	https://www.ddbj.nig.ac.jp/dra/index-e.html	DRA013570
pX458-HypaCas9	https://www.addgene.org/Yuichiro_Miyaoka/	Addgene #190604
Experimental models: Cell lines		
HEK293T cells	RIKEN BioResource Research Center	RCB2202
HeLa cells	RIKEN BioResource Research Center	RCB0007
PC9 cells	RIKEN BioResource Research Center	RCB4455
WTC11 iPS cells	Coriell Institute	GM25256
Oligonucleotides		
Single strand DNA donors for genome editing, see Table S4	Miyaoka et al. ¹⁸ Kato-Inui et al. ³²	N/A
Oligonucleotides for gRNA cloning, see Table S5	Miyaoka et al. ¹⁸ Kato-Inui et al. ³²	N/A
Primers in first PCR for Amplicon sequencing, see Table S6	This paper	N/A
Primers in second PCR for Amplicon sequencing, see Table S7	Sakata et al. ³⁸	N/A
Recombinant DNA		
pX458	Addgene	#48138
pX458-HypaCas9	This paper	#190604
Software and algorithms		
Agilent Genomic Workbench software	Agilent	RRID: SCR_010918
CLC Genomics Workbench	Qiagen	RRID: SCR_011853
CRISPResso2	Clement et al. ²⁸	https://github.com/pinellolab/CRISPResso2
Other		
On-chip SPIS	On-chip Biotechnologies	70001
On-chip Sort	On-chip Biotechnologies	N/A

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Microfluidic chip for cell sorting	On-chip Biotechnologies	2D Chip-Z1001
QX100 Droplet Digital PCR System	Bio-Rad	186-3001

RESOURCE AVAILABILITY**Lead contact**

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Yuichiro Miyaoka (miyaoka-yi@igakuken.or.jp).

Materials availability

All reagents and materials used in this manuscript are available upon request or prepared to be available from commercial sources. HEK293T cells (RCB2202), HeLa cells (RCB0007) and PC9 cells (RCB4455) are available at RIKEN BioResource Research Center (<https://cell.brc.riken.jp/en/>) and WTC11 iPS cells (GM25256) are available at Coriell Institute for Medical Research (<https://www.coriell.org/>).

Data and code availability

- All data reported in this study will be shared by the [lead contact](#) upon request. CGH array and amplicon sequencing raw data are available at the GEO database and the DDBJ Sequence Read Archive, respectively (see [key resources table](#)).
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS**Cell lines***HEK293T cells (RRID: CVCL_0063)*

Culture conditions: HEK293T cells were maintained in Dulbecco's modified Eagle medium (DMEM) with high glucose, sodium pyruvate, and L-glutamine (Thermo Fisher Scientific, 11965118) supplemented with 10% FBS and 1% Penicillin-Streptomycin (P/S, Sigma-Aldrich, P4333).

HeLa cells (RRID: CVCL_0030)

Culture conditions: HeLa cells were maintained in Dulbecco's modified Eagle medium (DMEM) with high glucose, sodium pyruvate, and L-glutamine (Thermo Fisher Scientific, 11965118) supplemented with 10% FBS and 1% Penicillin-Streptomycin (P/S, Sigma-Aldrich, P4333).

PC9 cells (RRID: CVCL_B260)

Culture conditions: PC9 cells were maintained in Roswell Park Memorial Institute (RPMI) 1640 medium with L-glutamine (Nacalai Tesque, 3026456) supplemented with 10% FBS and 1% Penicillin-Streptomycin (P/S, Sigma-Aldrich, P4333).

WTC11 iPS cells (RRID: CVCL_Y803)

Culture conditions: WTC11 iPS cells were maintained on Matrigel Matrix (Corning, 356231) in mTeSR Plus medium (STEMCELL Technologies, ST-100-0276) supplemented with 1% Penicillin-Streptomycin (P/S, Sigma-Aldrich, P4333). To passaged wells, we added 10 μ M Y-27632, a Rho-associated kinase (ROCK) inhibitor (FUJIFILM Wako Pure Chemical, 034-24024), to promote cell survival.

METHOD DETAILS**Plasmids and oligonucleotides for transfection**

We generated a vector that co-expresses EGFP and HypaCas9 (pX458-HypaCas9, Addgene #190604), based on pX458 (Addgene #48138). The single strand DNA donors were all 60 nt and had point mutations

in the middle of them. The sequences of the single strand DNA donors and gRNAs that were used in this study are summarized in [Tables S6](#) and [S7](#), respectively. All the single strand DNA donors and oligonucleotides for gRNA cloning, which were purified by standard desalting, were purchased from FASMAC, Japan.

Calculation of the copy number of target genes by karyotyping and CGH analysis

The karyotyping analysis of the cell lines was performed by Nihon Gene Research Laboratories. Karyotypes were analyzed in 10 cells for each cell line. The CGH assays were performed using the SurePrint G3 Human CGH Microarray Kit (Agilent) and the genomic DNA of WTC11 cells as a reference of the diploid cells. The genomic DNA of the cell lines used in this study for the CGH analysis was extracted using the DNeasy Blood and Tissue Kit (QIAGEN) according to the manufacturers' instructions. The CGH analysis gave peaks of the frequently observed signal intensity, which correspond to the number of chromosomes. Therefore, the chromosomal number most frequently observed in the karyotype analysis was assigned to the most frequent peak in the CGH array (for example, because the most frequently observed chromosomal number was 3 by karyotyping of HEK293T cells, the most frequent peak of the signal intensity in the CGH assay corresponded to 3 copies). Then, we were able to incrementally assign copy numbers to the peaks in the CGH assay around the highest peak to draw a line of fit between the copy number and the CGH peak values. The precise copy number of the target gene was calculated by this line of fit and the signal intensity around the target genes measured by the CGH assay (see also [Table S2](#)).

Transfection and cell sorting

For transfection, 2×10^5 cells were plated in a well of a 12-well plate coated with 80 $\mu\text{g}/\text{mL}$ Growth Factor Reduced Matrigel Basement Membrane Matrix (BD Biosciences). One day later, the cells were transfected with 720 ng of pX458 or pX458-HypaCas9 with a gRNA targeting RBM20, ATP7B, or GRN, and 80 ng of oligonucleotide donor DNA using 2.4 μL of Lipofectamine 2000 (Thermo Fisher Scientific) for HEK293T cells and HeLa cells, and 2.0 μL of Lipofectamine 3000 (Thermo Fisher Scientific) for PC9 cells, respectively, per well, according to the manufacturers' instructions. After 24 hours, cells were detached by 0.25% Trypsin-EDTA (Thermo Fisher Scientific) and EGFP-positive cells were sorted using On-chip Sort (On-chip Biotechnologies). The isolated EGFP-positive cells were dispensed one by one into Matrigel-coated 96-well plates using the On-chip Single Particle isolation System (SPiS). Subsequently, 100 $\mu\text{L}/\text{well}$ of conditioned medium was added to the plates with dispensed cells. One week after single cell dispensing by SPiS, the medium was changed. Two weeks after dispensing, for HEK293T cells and HeLa cells, the surviving colonies were first detached by Trypsin-EDTA, then the cells were resuspended in DMEM supplemented with 20% FBS and 1% P/S to evenly re-distribute the cells within the wells of a 96-well plate. The cells were cultured in this media until they reached confluence.

Preparation of multiplexed amplicon sequencing library

Genomic DNA was extracted from the cells in 96-well plates as described previously.³² The DNA was resuspended in 30 $\mu\text{L}/\text{well}$ of water. Targeted sites were amplified by the first PCR round using primers with homology to the region of interest and the Illumina forward and reverse adapters ([Table S8](#)). The first PCR round consisted of 0.3 μL each of 100 μM forward and reverse primer, 1.0 μL of genomic DNA, 2.0 μL of 2 mM dNTPs, 5.0 μL of 2 \times PCR Buffer KOD FX (Toyobo), 0.1 μL of KOD FX enzyme (Toyobo), 2 μL of betaine (FUJIFILM Wako Pure Chemical), and 0.3 μL of water. The thermal cycling settings were as follows: 95°C for 2 min, then 30 cycles of (95°C for 30 sec, 60°C for 30 sec, and 72°C for 1 min), followed by a final 72°C extension for 3 min. The first PCR products were diluted by adding 90 μL of water. Then, the DNA barcodes and illumina adaptors were added to the amplicons in the second PCR round, which consisted of 0.3 μL each of 100 μM unique forward and reverse index barcoding primers³⁸ ([Table S9](#)), 0.5 μL of the diluted first PCR product, 2 μL of 2 mM dNTPs, 5.0 μL of 2 \times PCR Buffer KOD FX, 0.1 μL of KOD FX enzyme, and 1.8 μL of water. The thermal cycling settings were as follows: 98°C for 3 min, then 30 cycles of (98°C for 30 sec, 57°C for 30 sec, and 72°C for 1 min), followed by a final 72°C extension for 3 min. The PCR products were electrophoresed in 2% agarose gel. NucleoSpin Gel and PCR Clean-up Midi kit (MACHEREY-NAGEL) were used to extract the pooled PCR products from 24 samples in 200 μL of water. The DNA concentrations of these library mixtures of 24 samples were quantified using the GenNext NGS Library Quantification Kit (Toyobo). All libraries were then mixed in equimolar amounts, 20% PhiX Control v3 (illumina) was added for amplicon sequencing. Sequencing was performed with MiSeq (Illumina) using the Miseq v2 reagent kit (Illumina) or Miseq Reagent kit v2 Nano (illumina) according to the manufacturer's instructions.

Amplicon sequencing data analysis

Fastq files generated by MiSeq were imported into the CLC Genomics Workbench (QIAGEN). Adapter sequences were removed and demultiplexed using the DNA Index. The data were then analyzed by CRISPResso2²⁸ (<https://github.com/pinellolab/CRISPResso2>) in the CRISPResso Batch mode. CRISPResso2 was installed as recommended using a Docker containerization system. The commands for the CRISPResso2 analysis used in this study are listed in Table S5. Alleles with $\leq 5\%$ frequency were excluded from the analysis as they were expected to be either sequence errors or very minor populations generated during cell proliferation from single cells. In the CRISPResso2 analysis, any alleles with deletions spanning the nucleotides for single nucleotide substitutions were characterized as "ambiguous". Therefore, we classified these "ambiguous" alleles as NHEJ alleles in this study.

Time course analysis of allelic frequencies

The RBM20 gene in HEK293T cells was edited induce the R636S mutation by HDR in the method described above. The edited cell population was passaged every 4-5 days at 1:30 ratio, and the WT, NHEJ, and HDR allelic frequencies were analyzed by droplet digital PCR (ddPCR) on day 0 (D0) and day 25 (D25). The ddPCR assay for RBM20 used in this study was described in our previous publications in detail.^{18,19}

Creating a model in which alleles were evenly distributed

Models assuming that the WT, NHEJ, HDR, and HDR+NHEJ alleles were randomly induced in all target alleles were built by distributing these edited alleles at their observed overall frequencies. The number of target alleles per cell was calculated by karyotyping and a CGH analysis, as described above. For example, when ATP7B was targeted in HEK293T cells, one cell has three ATP7B alleles. If the overall frequencies of WT, NHEJ, HDR, and HDR+NHEJ were 40%, 30%, 20%, and 10%, respectively, and 100 HEK293T cell clones were isolated, the expected number of clones with WT, NHEJ, and HDR alleles would be two ($100 \times (0.4 \times 0.3 \times 0.2) = 2.4$, rounded up to 2). We calculated the expected clone numbers of all the possible combinations of WT, NHEJ, HDR, and HDR+NHEJ alleles, and combined them to build the models.

QUANTIFICATION AND STATISTICAL ANALYSIS

The transfection and single cell dispensing experiments were performed in triplicate (three biological replicates). Clones isolated by the SPiS that showed no amplification of target sites by PCR or no successful alignments of amplicon sequencing reads by CRISPResso2²⁸ were excluded from the analyses. Values are displayed as Mean \pm Standard Error (S.E.). Statistical significance between two groups was assessed by a non-paired two-tailed Student's *t*-test and displayed on the figures with asterisks as follows: **p* < 0.05; ***p* < 0.01. NS: not significantly different (*p* > 0.1).