https://doi.org/10.1093/nar/gkac1254

Detection of unintended on-target effects in CRISPR genome editing by DNA donors carrying diagnostic substitutions

Martin Lackner [1,*], Nelly Helmbrecht[1], Svante Pääbo[1,2] and Stephan Riesenberg [1,*]

[1]Max Planck Institute for Evolutionary Anthropology, Leipzig, Sachsen 04103, Germany and [2]Okinawa Institute of Science and Technology, Onna-son, Okinawa 904-0495, Japan

Received August 09, 2022; Revised November 18, 2022; Editorial Decision December 14, 2022; Accepted December 15, 2022

ABSTRACT

CRISPR nucleases can introduce double-stranded DNA breaks in genomes at positions specified by guide RNAs. When repaired by the cell, this may result in the introduction of insertions and deletions or nucleotide substitutions provided by exogenous DNA donors. However, cellular repair can also result in unintended on-target effects, primarily larger deletions and loss of heterozygosity due to gene conversion. Here we present a strategy that allows easy and reliable detection of unintended on-target effects as well as the generation of control cells that carry wild-type alleles but have demonstratively undergone genome editing at the target site. Our 'sequence-ascertained favorable editing' (SAFE) donor approach relies on the use of DNA donor mixtures containing the desired nucleotide substitutions or the wild-type alleles together with combinations of additional 'diagnostic' substitutions unlikely to have any effects. Sequencing of the target sites then results in that two different sequences are seen when both chromosomes are edited with 'SAFE' donors containing different sets of substitutions, while a single sequence indicates unintended effects such as deletions or gene conversion. We analyzed more than 850 human embryonic stem cell clones edited with 'SAFE' donors and detect all copy number changes and almost all clones with gene conversion.

INTRODUCTION

CRISPR genome editing is initiated by a CRISPR nuclease, such as Cas9, that is directed to cleave a DNA target sequence by a guide RNA (gRNA) (1). Double-strand breaks (DSBs) introduced by the nuclease are then repaired by cellular repair pathways. One of the two dominant pathways is non-homologous end joining that tends to cause small insertions or deletions (indels) (2). It is therefore used to produce gene knockouts (3,4). Another is homology directed repair (HDR) that generates 3' single-stranded DNA overhangs at the break and uses a template for repair (5). The HDR pathway can be harnessed to introduce precise changes such as single nucleotide variants into genomic DNA by providing cells with exogenous donor DNA containing the desired mutation (6).

However, CRISPR genome editing frequently produces unwanted effects at the target sites, for example (i) insertions and deletions that can involve large chromosomal regions (7–15); (ii) gene conversion of one chromosome to match the sequence of the other chromosome (12,15–19) and (iii) chromosomal rearrangements including translocations, inversions and aneuploidy (8,20–25). DSB repair might also have lasting epigenetic consequences (26–28).

Many of these unintended effects at and around the editing site cannot be detected by simply sequencing a short region around the edited site. For example, if monoallelic deletions, inversions or translocations include one or both primer binding sites used to amplify and sequence the region around the edited site, only the intact allele will be amplified and sequenced, causing the cell line to appear homozygously edited (7,12,13,15). Similarly, heterozygous insertions that increase the size of the region between the sequencing primers beyond the amplifiable range and most cases of editing-induced gene conversion are also not detectable by sequencing of the target site. However, such unintended genome alterations need to be detected as they may have biological effects that may erroneously be ascribed to the intended mutation (13).

Current methods to detect such unintended effects are cumbersome. In addition to sequencing of the edited site, they include quantification of the copy number of the edited site and genotyping of heterozygous positions on both sides of the edited site to detect loss of heterozygosity through gene conversion (Supplementary Figure S1) (29). Here, we present an approach that allows unintended events affect-

*To whom correspondence should be addressed. Tel: +49 341 3550519; Email: stephan_riesenberg@eva.mpg.de
Correspondence may also be addressed to Martin Lackner. Email: martin_lackner@eva.mpg.de

© The Author(s) 2023. Published by Oxford University Press on behalf of Nucleic Acids Research.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

ing editing target sites to be easily detected when the target site is sequenced to determine the genome editing outcome in cellular clones. It also generates cells that have undergone CRISPR-Cas9-mediated genome editing but carry the wild-type allele at the site of interest. Such cells are useful as controls to study biological effects of the introduced allele as they allow to control for effects that are due to DNA cleavage and repair rather than the introduced allele.

MATERIALS AND METHODS

Cell culture

H9 human embryonic stem cells (female, WiCell Research Institute, Ethics permit AZ 3.04.02/0118) were grown on Matrigel Matrix (Corning, 354277) in mTeSR1 medium (StemCell Technologies, 85850) supplemented with MycoZap Plus-CL (Lonza, VZA-2012) that was replaced daily in a humidified incubator at 37°C with 5% CO_2. At approximately 80% confluency, cells were dissociated using EDTA (Gibco, 12605010) and split 1:5 to 1:20 in medium supplemented with 10 M Rho-associated protein kinase (ROCK) inhibitor Y-27632 (StemCell Technologies, 72307) for 1 day after replating.

Oligonucleotide and ribonucleoprotein electroporation

'SAFE' donors were ordered as mixtures of 90 nt long synthetic single-stranded DNA oligodeoxynucleotides from IDT (Supplementary Table S1). During DNA synthesis, all respective bases were provided for random incorporation at the diagnostic positions. One million cells were mixed with 244 pmol recombinant Alt-R *Streptococcus pyogenes* Cas9 nuclease V3 (IDT, 1081059), 160 pmol Alt-R CRISPR-Cas9 crRNA (IDT) (Supplementary Table S1), 160 pmol Alt-R CRISPR-Cas9 tracrRNA (IDT, 1072534), 200 pmol single-stranded Donor DNA (IDT), 200 pmol Alt-R Cas9 Electroporation enhancer (IDT, 1075916), resuspended in 100 μl Nucleofector Solution for Human Stem Cells (Lonza, VVPH-5022) and electroporated in a 100 μl cuvette (Lonza) with program B-016 in the Nucleofector 2b Device (Lonza, AAB-1001). Electroporations were done in triplicates for each donor mixture. 30% of the electroporated cells were plated for bulk genotype analysis in a 12-well with media supplemented with ROCK inhibitor Y-27632. The rest of the cells electroporated with 'SAFE' donors containing the wild-type or missense allele were pooled (three independent electroporations each) and plated on a 12-well together for subsequent generation of clonal cell lines. After one passage, a single cell dilution of the pooled cells was plated onto a CytoSort Array (Cell Microsystems, CS-200Q-5) with 1:2 conditioned StemFlex media (Gibco, A3349401) supplemented with CloneR cloning supplement (StemCell Technologies, 05888). Five days after plating, single cell-derived colonies were sorted into 96-wells with the CellRaft AIR system (Cell Microsystems).

Library preparation and sequencing

At least 3 days after plating, cells were dissociated using TrypLE Express Enzyme (Gibco, 12605010), pelleted and resuspended in QuickExtract (Biozym, 101094). Incubation at 65°C for 10 min, 68°C for 5 min and finally 98°C for 5 min was performed to yield single-stranded DNA as a PCR template. Primers containing adapters for Illumina sequencing to amplify each targeted locus and positions up- and downstream of the targeted locus that are heterozygous in H9 were designed with Primer3Plus (30) and obtained from IDT (Coralville, USA) (Supplementary Table S1). PCR was done in a C1000 Touch thermal cycler (Bio-Rad, 1851197) using the KAPA2G Robust HotStart PCR-Kit (Roche, KK5525) with supplied buffer B and 3 μl of cell extract in a total volume of 25 μl. The thermal cycling profile of the PCR was: 95°C 3 min; 35× (95°C 15 s, 65°C 15 s, 72°C 15 s); 72°C 60 s. P5 and P7 Illumina adapters with sample-specific indices were added in a second PCR reaction (31) using Phusion HF MasterMix (Thermo Scientific, F-531L) and 0.5 μl of the first PCR product. The thermal cycling profile of the PCR was: 98°C 30 s; 25× (98°C 10 s, 58°C, 10 s, 72°C 20 s); 72°C 5 min. Amplifications were analyzed by agarose gel electrophoresis using 2% EX gels (Invitrogen, G402022). The indexed amplicons were purified using SPRIselect beads (Beckmann Coulter, B23317) in a 1:0.8 ratio of PCR solution to beads. Double-indexed libraries were sequenced on a MiSeq platform (Illumina) using paired-end sequences of 2 × 151 bp (+7 bp index). After base calling using Bustard (Illumina) adapters were trimmed using leeHom (32). For edited cell populations, 12 000–49 000 (average: 26 000) molecules were sequenced and 1000–33 000 (average: 16 000) molecules for cellular clones.

Amplicon sequence analysis

Bam-files were demultiplexed and converted into fastq files using SAMtools (33). Fastq-files were analyzed with CRISPResso2 (34) for the percentages of alleles and indels at the targeted positions and SNPs. Analyses were restricted to amplicons with a minimum of 70% similarity to the wild-type sequence and to a window of 40 bp centered at the Cas9 cleavage site. Sequence homology for an HDR occurrence was set to 95%. Unexpected substitutions were ignored as putative sequencing errors. An R script employing SAMtools was used to count sequencing reads with all expected diagnostic substitution motifs in edited cell bulks and clones with >2000 reads aligned to the expected amplicon by CRISPResso2 and <10% indels (Supplementary Code 1).

Copy number analysis

Copy number of edited sites were determined with the QX200 Droplet Digital PCR System (Bio-Rad) according to the manufacturer's instructions. In a 20 μl reaction, 200 nM primers and probes for the respective editing site and a reference site in the *FOXP2* gene (Supplementary Table S1) were mixed with 1 μl genomic DNA extract and ddPCR Supermix for Probes (Bio-Rad, 1863024). Droplets were generated in a QX200 Droplet Generator with Droplet Generation Oil for Probes (Bio-Rad, 1863005). The thermal cycling was done in a C1000 Touch Thermal Cycler with the following thermal profile: 95°C 10 min; 40 × (95°C 30 s [ramp: 2°C/s], 60°C 1 min [ramp: 1°C/s]), 98°C 5 min.

PCR-positive and negative droplets were counted with the QX200 Droplet Reader (Bio-Rad, 1864003). The copy number of the editing site resulting from the ratio of droplets that showed PCR amplification for the reference and editing site was adjusted with the method of least squares to approximate integers.

RESULTS

A mixed donor strategy

To allow the detection of unintended effects at the editing site by sequencing after editing, we provided cells with a mixture of donor DNA molecules, where all molecules contained the desired nucleotide substitution as well as different combinations of additional substitutions used to detect the editing of both chromosomes in a diploid target (Figure 1A). In the following, these additional substitutions are referred to as 'diagnostic' substitutions. As a result, in any one cell, Cas9-mediated DSBs on the two chromosomes are likely to be repaired with different donors, containing different sets of diagnostic substitutions. Thus, if diagnostic substitutions are observed in a heterozygous state when the target site is sequenced while the desired substitution is homozygous, this shows that both chromosomes have been successfully edited (Figure 1B). In contrast, the absence of heterozygous diagnostic substitutions makes it likely that unintended effects have affected one chromosome. Additionally, an abnormal number or ratio of diagnostic substitutions may indicate other errors, such as multiple insertions of the donor sequence, duplications, aneuploidy of the targeted chromosome or cell populations made up of several clones carrying different editing events. We thus dubbed the mixed donor approach 'sequence-ascertained favorable editing' (SAFE).

Design of DNA donor mixtures

To test the 'SAFE' donor approach, we chose two target sites in the genes *TEX2* (chr17:62290457) and *TTF1* (chr9:135277401) that in our hands had frequently produced deletions and loss of heterozygosity affecting the editing sites. We designed single-stranded DNA donors that were 90 nucleotides long, centered on the Cas9 cut site, and non-complementary to the gRNA. Each single-stranded DNA donor molecule in the mixture carried an intended missense substitution (*TEX2*: G374D; *TTF1*: T270A) and in addition a set of diagnostic substitutions at up to three positions (Figure 2A). At each of the three diagnostic positions donor molecules carried either the reference allele or a diagnostic substitution yielding 18 distinct sequence motifs that occur among the donor molecules in the mixture. Additionally, we designed 'SAFE' donors where all molecules carried the wild-type allele at the position under study in addition to diagnostic substitutions to produce wild-type allele control cell lines that have undergone CRISPR-mediated genome editing. The diagnostic substitutions were chosen to not change the amino acid sequences of the encoded proteins and result in codons whose usage in humans do not differ >1.5-fold from the wild-type ones (35). If all donors in the mixtures were equally likely to be used as HDR templates, this would result in a probability of 94.4% that two chromosomes would be edited by donors carrying different DNA sequences. Donors were ordered as mixtures of synthetic single-stranded oligodeoxynucleotides where the respective nucleotides were provided for random incorporation at the diagnostic positions during synthesis. Prior to editing, we assessed the distribution of individual donors in the mixtures by sequencing (Figure 2B). Based on the actual frequencies of individual donors in the mixtures, the expected fraction of edited cells that would carry heterozygous diagnostic substitutions was 93.7% and 92.9% for *TEX2* and *TTF1*, respectively.

Diagnostic motif distribution in edited cells

We electroporated the donor mixtures and Cas9-gRNA ribonucleoprotein complexes into human H9 embryonic stem cells and extracted genomic DNA from the populations of edited cells. The editing sites were then amplified by PCR and the amplification product sequenced to gauge the frequency of substitutions at each nucleotide position. Approximately 56% of sequenced DNA molecules carried the intended missense substitution in *TEX2* and 40% in *TTF1,* when 'SAFE' donors were used. When donors containing only the intended missense substitutions without diagnostic substitutions were used, the percentages of molecules with intended missense substitutions were similar (Supplementary Figure S2).

Among the target sequences carrying the missense substitutions, all combinations of diagnostic substitutions were present and their frequencies generally reflect those seen among the donor oligonucleotide mixtures (Figure 2B, C). However, some substitutions that were close to the Cas9 cleavage site, such as the G374D missense substitution for *TEX2* or the diagnostic substitution at position Z for *TTF1*, were more efficiently introduced than others. Given the distribution of diagnostic substitution motifs, 89% (*TEX2*) and 94% (*TTF1*) of cells with intended substitutions were expected to carry heterozygous diagnostic substitutions.

Detection of unintended on-target effects in cellular clones

To generate cellular clones, for the two genes *TEX2* and *TTF1*, we electroporated human stem cells separately with 'SAFE' donors carrying the wild-type or missense substitution of interest, subsequently pooled both for propagation, and isolated 437 and 435 clones from the *TEX2*- and *TTF1*-pools, respectively (Supplementary Figure S3). We extracted genomic DNA from each of the clones and amplified and sequenced the edited sites (Figure 3A). 18 *TEX2*-clones and 41 *TTF1*-clones had abnormal ratios of sequences, which may be indicative of non-clonality or aneuploidy. Many of the remaining clones contained insertions or deletions within the amplified sequence (*TEX2*: 265, *TTF1*: 193) and in a few cases unintended substitutions (*TEX2*: 0, *TTF1*: 4) (Figure 3A).

Among the remaining clones, 79 *TEX2*-clones and 107 *TTF1*-clones carried two sequences each making up 40–60% of aligned reads (deemed to be heterozygous), while 75 and 90, respectively, carried only one sequence that made up more than 90% of the reads. Among the latter, approximately half produced only the unmodified reference sequence while half carried substitutions at intended missense

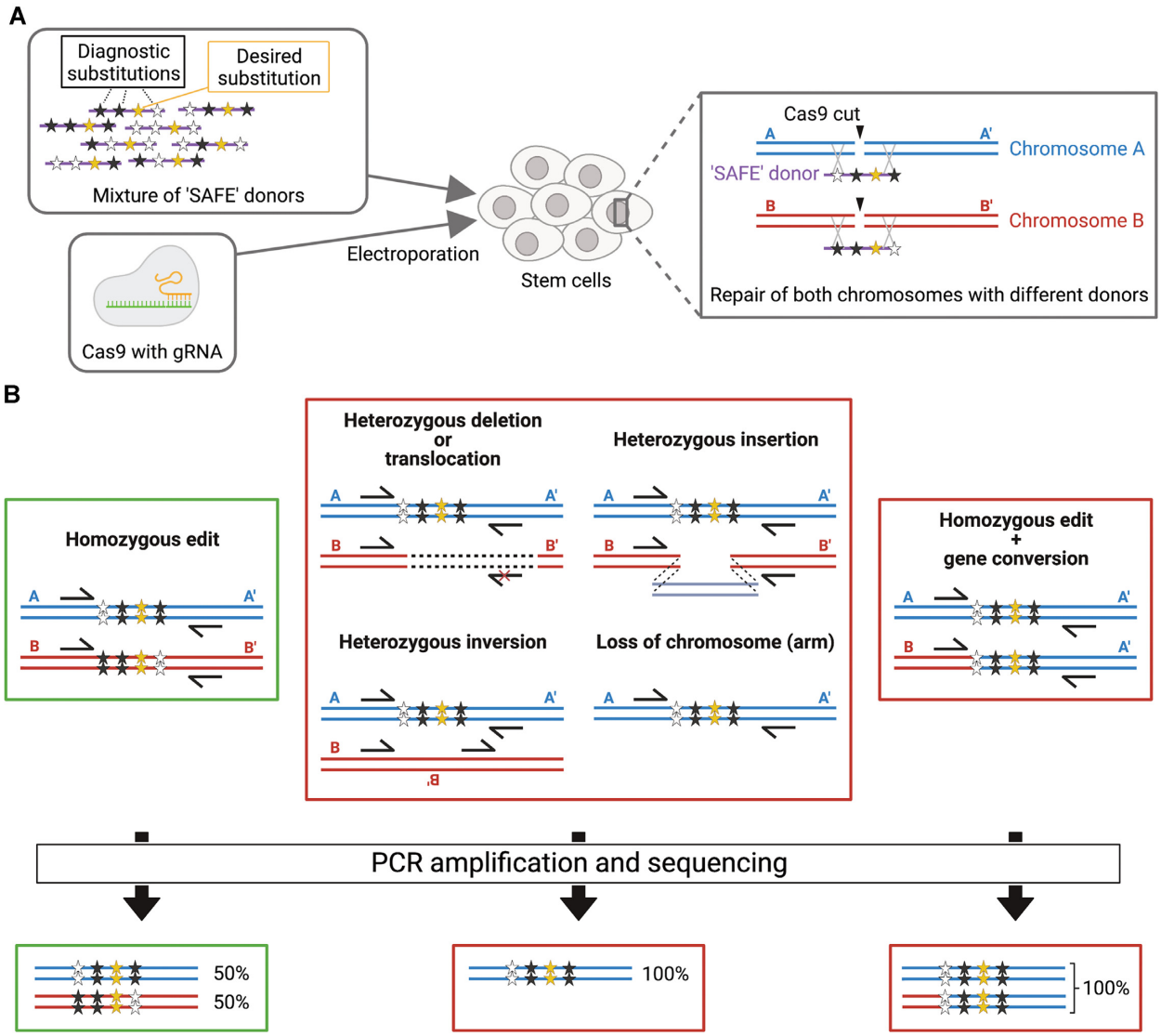

Figure 1. Detection of unwanted editing effects with 'SAFE' donors. (**A**) Cells are edited using a mixture of 'sequence-ascertained favorable editing' (SAFE) donor molecules carrying different combinations of diagnostic substitutions such that the two chromosomes are likely to be repaired with donors containing different sets of diagnostic substitutions. (**B**) Thus, if both chromosomes are edited, sequencing of the edited site produces two distinct sequences with similar amounts of reads (left) but only one sequence if deletions or other unwanted effects prevent PCR amplification and sequencing of the affected chromosome (middle). If the sequence on one chromosome is converted to become identical to the other chromosome, only one sequence is likely to result (right).

or diagnostic positions (Figure 3A). The copy number of the edited sites in genomic DNA extracts from the cellular clones carrying one or two sequences was determined with quantitative PCR. Furthermore, genomic DNA around single nucleotide polymorphisms (SNP) up- and downstream of the edited sites were amplified and sequenced to detect loss of heterozygosity in the chromosomal vicinity of the target site (Supplementary Figure S3).

Of the 39 *TEX2*- and 48 *TTF1*-clones that carried only the wild-type sequence, none had aberrant copy numbers of the edited site and 5 and 10%, respectively, had lost heterozygosity at one or both SNPs located up- and downstream of the edited site. Among the 36 *TEX2*-clones and 42 *TTF1*-clones that carried a single sequence containing substitutions from the donor oligonucleotides, 33 and 50%, respectively, had lost one copy of the target site. 22 and 29% of clones were affected by loss of heterozygosity at SNPs flanking the edited site, but carried two copies of the edited site. The remaining 16 *TEX2*-clones and 9 *TTF1*-clones carried the same set of diagnostic substitutions on both chromosomes. In contrast, out of the 79 *TEX2*-clones and 107 *TTF1*-clones that carried heterozygous diagnostic substitutions, none were affected by copy number loss of the edited site. A single *TEX2*-clone had lost heterozygosity at both the up- and downstream SNPs while a single *TTF1*-clone had lost heterozygosity at the upstream SNP (Figure 3B).

Thus, 56% of *TEX2*-clones ($n = 20$) and 79% of *TTF1*-clones ($n = 33$) that carried a single sequence with a substitution were affected by unintended on-target effects. In contrast, clones where the diagnostic substitutions indicated

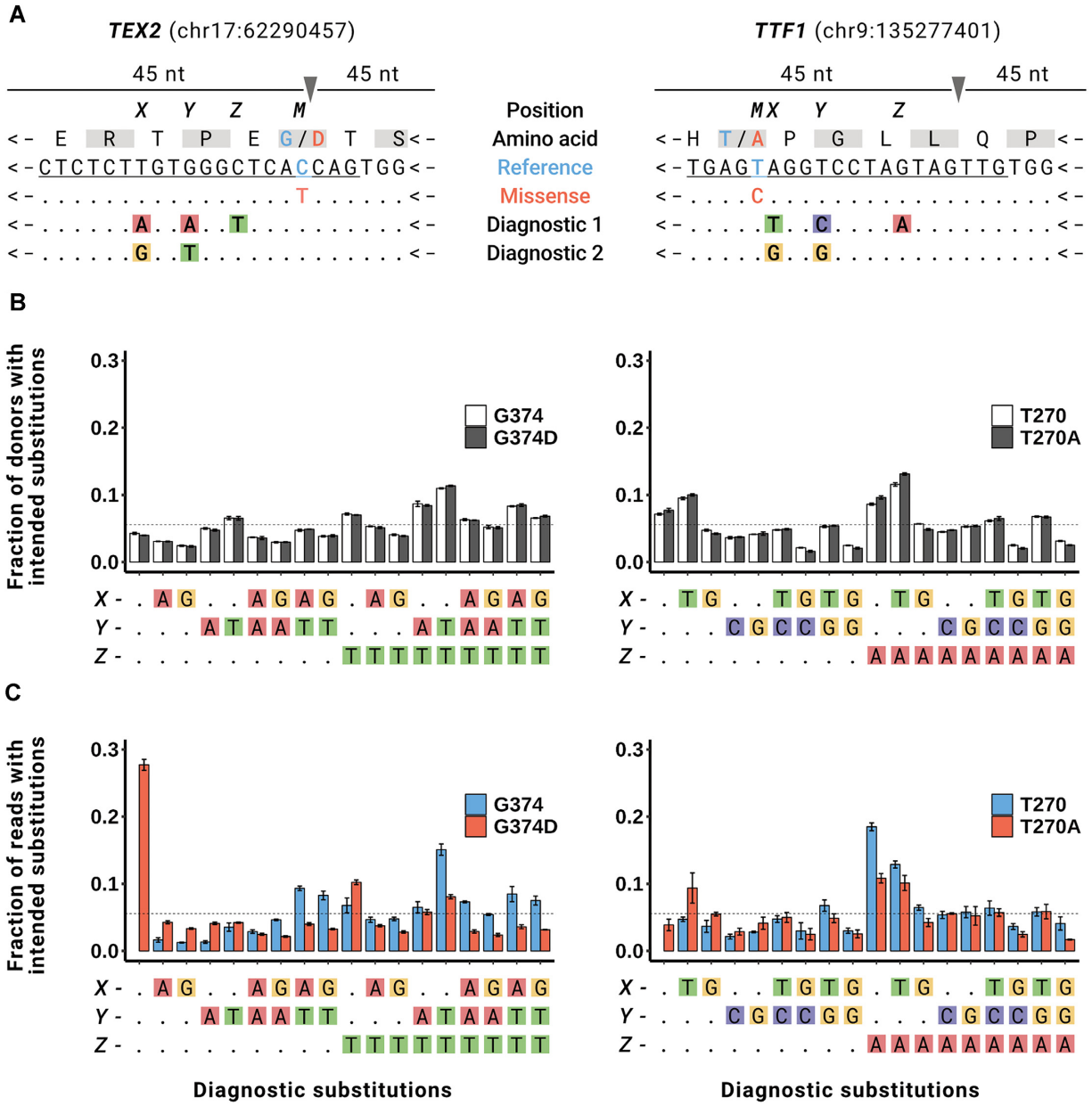

Figure 2. Substitution distributions in 'SAFE' DNA donors and edited cells. (**A**) Design of 'SAFE' donors for editing positions of interest in the human *TEX2* (left) and *TTF1* genes (right). The donors are a mixture of 90 nt long single-stranded DNA molecules centered at the Cas9 cleavage site (grey triangle), that carry combinations of diagnostic substitutions at positions X, Y and Z. In the mixtures for generating wild-type control cell lines, all molecules contain the reference allele (blue) at the position under study (M), all donors for generating mutated cell lines contain the G374D (TEX2) or T270A (TTF1) substitutions (red), respectively. Each of the diagnostic positions can have the reference allele or one of the synonymous codon substitutions (diagnostic 1, diagnostic 2), resulting in 18 different sequence motifs. Arrows indicate the direction of translation. The gRNA recognition sequence is underlined. (**B**) Frequencies of motifs containing all combinations of diagnostic substitutions alongside the wild-type (white) or missense allele (grey) at the position of interest in 'SAFE' donors used for editing *TEX2* (left) and *TTF1* (right). The dashed line represents the expected frequency for each of the 18 motifs. (**C**) Frequencies of diagnostic motifs in genomic DNA extracts from H9 cells edited with 'SAFE' donors containing the wild-type (blue) or missense allele (red) at the position of interest for *TEX2* (left) and *TTF1* (right). No reads are attributed to the wild-type motif without diagnostic substitutions, as it is indistinguishable from unedited sequences. Error bars give the standard deviation from three independent transfections per donor mixture.

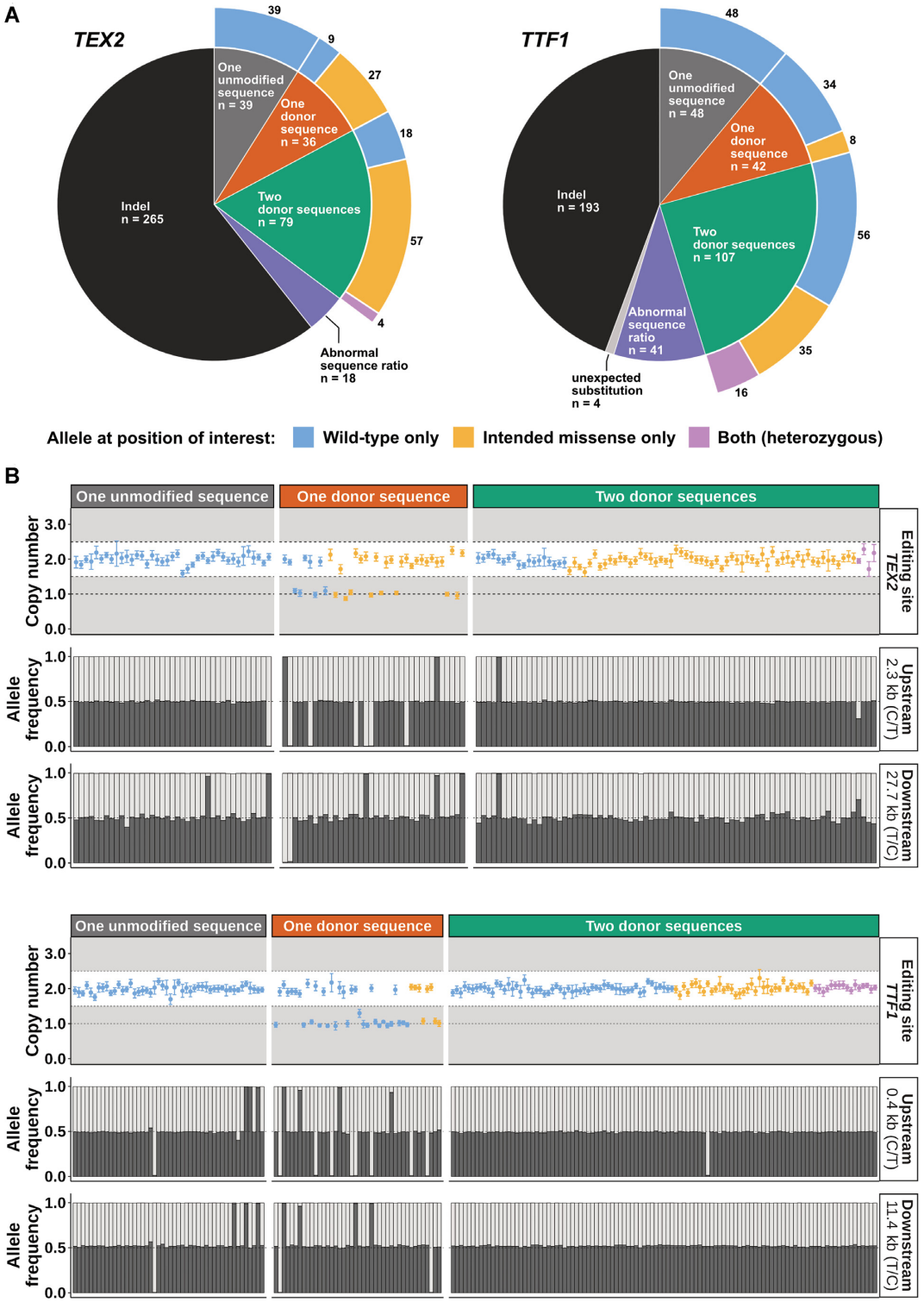

Figure 3. Editing outcomes in cellular clones using 'SAFE' donors. (**A**) Clones were categorized by the type and number of sequences obtained from next-generation sequencing of the target site. The number of clones showing only the unmodified sequence (dark grey), only one sequence containing substitutions from the donor mixtures (orange), two sequences in similar proportions (green), two or more sequences in unequal proportions (purple), unexpected substitutions (light grey), or indels (black), are given for *TEX2* (left) and *TTF1* (right). For each relevant category, cellular clones are further color coded in the outer circle by the allele at the single position of interest: wildtype only (blue), intended missense only (yellow), or both heterozygously (magenta). (**B**) Copy numbers of the edited site in clones carrying only the unmodified sequence (dark grey), only one donor sequence (orange) or two donor sequences in equal proportions (green) (*TEX2*: upper panel; *TTF1*: lower panel). A copy number of two is expected, while a copy number of one indicates an on-target effect such as e.g. a large deletion, translocation, inversion or chromosomal truncation. SNPs located 2.3 kb upstream (rs17563098) and 27.7 kb downstream (rs11869645) of the edited site for *TEX2* and 0.4 kb upstream (rs3739916) and 11.4 kb (rs2885495) downstream for *TTF1* are expected to show an allele frequency of 0.5 for their two alleles (black and white). An allele frequency of one indicates loss of heterozygosity at that position.

heterozygosity at the edited sites carried two copies of the edited site and the frequency of loss of heterozygosity at up- and downstream SNPs was 1% (2 out of 186) rather than 8% (29 out of 351).

DISCUSSION

When CRISPR-Cas9 editing is used to modify a position in the genome, loss of one allelic copy of the target sequence is a frequent problem (7–15) which has to be excluded, generally by performing quantitative PCR of DNA extracted from the edited cells (29). Loss of heterozygosity where one chromosome has been converted to the sequence of the other one in one or both directions from the edited site is another common problem (12,15–19) that requires that polymorphisms flanking the edited site are identified and sequenced or genotyped in the edited cells to ensure that heterozygosity has not been lost (29). The 'SAFE' donor approach described here allows such unintended effects to be easily detected when the target site is sequenced. It avoids the need for additional validation of heterozygosity at flanking polymorphisms and determination of copy numbers of the target site.

Nevertheless, two clones that carried heterozygous diagnostic substitutions had lost heterozygosity at positions flanking the target site. We hypothesize that this is caused by an initial Cas9-mediated DSB that was repaired using the other chromosome as a template resulting in gene conversion of flanking sequences. Since no substitutions prevent repeated Cas9-cleavage at the target site, subsequent DSBs can then be repaired with different DNA donors on both chromosomes, thus producing clones with heterozygous editing sites but homozygous flanking sequences. However, this sequence of events seems rare as it occurred in only two out of 186 clones carrying heterozygous diagnostic substitutions.

Interestingly, more clones than expected carried two identical sets of diagnostic substitutions without being affected by deleterious on-target effects. This is likely due to that substitutions close to the DSB introduced by Cas9, such as the G374D substitution in the *TEX2* gene, are introduced at higher efficiencies than more distant substitutions (36–38). This likely also causes some cells to only partially incorporate substitutions from donors omitting more distant substitutions (Supplementary Figure S4).

A potential drawback with the 'SAFE' donor approach is that one or more additional changes are introduced close to the site of interest, which might have unintended effects on gene expression, splicing, RNA folding and stability or translation (39–42). However, additional substitutions are commonly introduced in genome editing as 'blocking mutations' to prevent repeated cleavage of the target site (38). The 'SAFE' donors thus combine the benefits of blocking and diagnostic substitutions. In addition, diagnostic substitutions can be chosen to minimize the risk that they have a biological impact. For example, in protein-coding sequences synonymous substitutions will not change the amino acid sequence of the encoded protein. Furthermore, if codons with similar frequencies as the wild-type codon are used, it will minimize the risk that the substitutions affect the translation of the protein.

To generate control cell lines for biological experiments, donors carrying the wild-type nucleotide at the site of interest in combination with the same diagnostic substitutions as the donors carrying the nucleotide to be introduced at the site of interest can be used. Such cellular clones are excellent controls because they have demonstratively been edited in the region of interest and are therefore likely to be equally affected by any potential effects caused by Cas9-induced DSBs and their repair, including epigenetic effects or selection of cells carrying oncogenic mutations (20,43,44). For *TEX2* and *TTF1*, 24 and 33% of cellular clones ($n = 18$ and 30) can be matched with clones carrying the alternative allele at the position of interest but identical substitutions at diagnostic positions (Supplementary Figure S5).

If high HDR editing efficiency of the target of interest can be achieved and thus fewer cellular clones need to be isolated, it is an advantage to use fewer diagnostic positions, especially to obtain more clones that only differ in the position of interest. Easily applicable tools to increase HDR efficiency are small molecule inhibition of the competing non-homologous end-joining (18) or chemical modification of the DNA donors (45,46).

In summary, the genome editing approach presented allows detection of copy number variants of the target site as well as loss of heterozygosity when the target site is sequenced, thus greatly reducing time and cost of genome editing experiments. In addition, it facilitates the identification of non-clonal cellular colonies and the generation of control cell lines that have demonstratively undergone editing but carry the wild-type state at the position of interest.

DATA AVAILABILITY

The NGS data generated in this study have been deposited in the Dryad database under accession code dryad.905qfttp6. Data are also available on request from authors.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Antje Weihmann and Barbara Schellbach for DNA sequencing, Tomislav Maricic for helpful input and the Max Planck Society and the NOMIS foundation for financial support. Schematics were designed with biorender.com.

Author contributions: S.R. and N.H. conceived the idea. M.L. performed the experiments and analyzed the data. S.R., N.H. and S.P. gave input for data analysis. M.L. and S.R. wrote the manuscript with input from all authors.

FUNDING

Max-Planck Society; NOMIS Foundation. Funding for open access charge: Max-Planck-Institute for Evolutionary Anthropology, Department of Evolutionary Genetics.

Conflict of interest statement. None declared.

REFERENCES

1. 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 (New York, N.Y.)*, **337**, 816–821.
2. Heidenreich,E., Novotny,R., Kneidinger,B., Holzmann,V. and Wintersberger,U. (2003) Non-homologous end joining as an important mutagenic process in cell cycle-arrested cells. *EMBO J.*, **22**, 2274–2283.
3. Carbery,I.D., Ji,D., Harrington,A., Brown,V., Weinstein,E.J., Liaw,L. and Cui,X. (2010) Targeted genome modification in mice using zinc-finger nucleases. *Genetics*, **186**, 451–459.
4. Wang,H., Yang,H., Shivalila,C.S., Dawlaty,M.M., Cheng,A.W., Zhang,F. and Jaenisch,R. (2013) One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell*, **153**, 910–918.
5. Moynahan,M.E. and Jasin,M. (2010) Mitotic homologous recombination maintains genomic stability and suppresses tumorigenesis. *Nat. Rev. Mol. Cell Biol.*, **11**, 196–207.
6. Chen,F., Pruett-Miller,S.M., Huang,Y., Gjoka,M., Duda,K., Taunton,J., Collingwood,T.N., Frodin,M. and Davis,G.D. (2011) High-frequency genome editing using ssDNA oligonucleotides with zinc-finger nucleases. *Nat. Methods*, **8**, 753–755.
7. Adikusuma,F., Piltz,S., Corbett,M.A., Turvey,M., McColl,S.R., Helbig,K.J., Beard,M.R., Hughes,J., Pomerantz,R.T. and Thomas,P.Q. (2018) Large deletions induced by Cas9 cleavage. *Nature*, **560**, E8–E9.
8. Kosicki,M., Tomberg,K. and Bradley,A. (2018) Repair of double-strand breaks induced by CRISPR-Cas9 leads to large deletions and complex rearrangements. *Nat. Biotechnol.*, **36**, 765–771.
9. Cullot,G., Boutin,J., Toutain,J., Prat,F., Pennamen,P., Rooryck,C., Teichmann,M., Rousseau,E., Lamrissi-Garcia,I., Guyonnet-Duperat,V. *et al.* (2019) CRISPR-Cas9 genome editing induces megabase-scale chromosomal truncations. *Nat. Commun.*, **10**, 1136.
10. Owens,D.D.G., Caulder,A., Frontera,V., Harman,J.R., Allan,A.J., Bucakci,A., Greder,L., Codner,G.F., Hublitz,P., McHugh,P.J. *et al.* (2019) Microhomologies are prevalent at Cas9-induced larger deletions. *Nucleic Acids Res.*, **47**, 7402–7417.
11. Przewrocka,J., Rowan,A., Rosenthal,R., Kanu,N. and Swanton,C. (2020) Unintended on-target chromosomal instability following CRISPR/Cas9 single gene targeting. *Ann. Oncol.*, **31**, 1270–1273.
12. Weisheit,I., Kroeger,J.A., Malik,R., Klimmt,J., Crusius,D., Dannert,A., Dichgans,M. and Paquet,D. (2020) Detection of deleterious on-target effects after HDR-mediated CRISPR editing. *Cell Rep.*, **31**, 107689.
13. Maricic,T., Helmbrecht,N., Riesenberg,S., Macak,D., Kanis,P., Lackner,M., Pugach-Matveeva,A.D. and Pääbo,S. (2021) Comment on "Reintroduction of the archaic variant of NOVA1 in cortical organoids alters neurodevelopment". *Science*, **374**, eabi6060.
14. Boutin,J., Cappellen,D., Rosier,J., Amintas,S., Dabernat,S., Bedel,A. and Moreau-Gaudry,F. (2022) ON-target adverse events of CRISPR-Cas9 nuclease: more chaotic than expected. *CRISPR J.*, **5**, 19–30.
15. Simkin,D., Papakis,V., Bustos,B.I., Ambrosi,C.M., Ryan,S.J., Baru,V., Williams,L.A., Dempsey,G.T., McManus,O.B., Landers,J.E. *et al.* (2022) Homozygous might be hemizygous: cRISPR/Cas9 editing in iPSCs results in detrimental on-target defects that escape standard quality controls. *Stem Cell Rep.*, **17**, 993–1008.
16. Ikeda,K., Uchida,N., Nishimura,T., White,J., Martin,R.M., Nakauchi,H., Sebastiano,V., Weinberg,K.I. and Porteus,M.H. (2018) Efficient scarless genome editing in human pluripotent stem cells. *Nat. Methods*, **15**, 1045–1047.
17. Gorter de Vries,A.R., Couwenberg,L.G.F., van den Broek,M., La Torre Cortés,P.d., Ter Horst,J., Pronk,J.T. and Daran,J.-M.G. (2019) Allele-specific genome editing using CRISPR-Cas9 is associated with loss of heterozygosity in diploid yeast. *Nucleic Acids Res.*, **47**, 1362–1372.
18. Riesenberg,S., Chintalapati,M., Macak,D., Kanis,P., Maricic,T. and Pääbo,S. (2019) Simultaneous precise editing of multiple genes in human cells. *Nucleic Acids Res.*, **47**, e116.
19. Alanis-Lobato,G., Zohren,J., McCarthy,A., Fogarty,N.M.E., Kubikova,N., Hardman,E., Greco,M., Wells,D., Turner,J.M.A. and Niakan,K.K. (2021) Frequent loss of heterozygosity in CRISPR-Cas9-edited early human embryos. *Proc. Natl. Acad. Sci. U.S.A.*, **118**, e2004832117.
20. Cho,S.W., Kim,S., Kim,Y., Kweon,J., Kim,H.S., Bae,S. and Kim,J.-S. (2014) Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Res.*, **24**, 132–141.
21. Shin,H.Y., Wang,C., Lee,H.K., Yoo,K.H., Zeng,X., Kuhns,T., Yang,C.M., Mohr,T., Liu,C. and Hennighausen,L. (2017) CRISPR/Cas9 targeting events cause complex deletions and insertions at 17 sites in the mouse genome. *Nat. Commun.*, **8**, 15464.
22. Rayner,E., Durin,M.-A., Thomas,R., Moralli,D., O'Cathail,S.M., Tomlinson,I., Green,C.M. and Lewis,A. (2019) CRISPR-Cas9 causes chromosomal instability and rearrangements in cancer cell lines, detectable by cytogenetic methods. *CRISPR J.*, **2**, 406–416.
23. Zuccaro,M.V., Xu,J., Mitchell,C., Marin,D., Zimmerman,R., Rana,B., Weinstein,E., King,R.T., Palmerola,K.L., Smith,M.E. *et al.* (2020) Allele-specific chromosome removal after Cas9 cleavage in human embryos. *Cell*, **183**, 1650–1664.
24. Leibowitz,M.L., Papathanasiou,S., Doerfler,P.A., Blaine,L.J., Sun,L., Yao,Y., Zhang,C.-Z., Weiss,M.J. and Pellman,D. (2021) Chromothripsis as an on-target consequence of CRISPR-Cas9 genome editing. *Nat. Genet.*, **53**, 895–905.
25. Nahmad,A.D., Reuveni,E., Goldschmidt,E., Tenne,T., Liberman,M., Horovitz-Fried,M., Khosravi,R., Kobo,H., Reinstein,E., Madi,A. *et al.* (2022) Frequent aneuploidy in primary human T cells after CRISPR-Cas9 cleavage. *Nat. Biotechnol.*, **40**, 1807–1813.
26. Cuozzo,C., Porcellini,A., Angrisano,T., Morano,A., Lee,B., Di Pardo,A., Messina,S., Iuliano,R., Fusco,A., Santillo,M.R. *et al.* (2007) DNA damage, homology-directed repair, and DNA methylation. *PLos Genet.*, **3**, e110.
27. Russo,G., Landi,R., Pezone,A., Morano,A., Zuchegna,C., Romano,A., Muller,M.T., Gottesman,M.E., Porcellini,A. and Avvedimento,E.V. (2016) DNA damage and repair modify DNA methylation and chromatin domain of the targeted locus: mechanism of allele methylation polymorphism. *Sci. Rep.*, **6**, 33222.
28. Caron,P., Pobega,E. and Polo,S.E. (2021) DNA double-strand break repair: all roads lead to HeterochROMAtin marks. *Front. Genet.*, **12**, 730696.
29. Weisheit,I., Kroeger,J.A., Malik,R., Wefers,B., Lichtner,P., Wurst,W., Dichgans,M. and Paquet,D. (2021) Simple and reliable detection of CRISPR-induced on-target effects by qgPCR and SNP genotyping. *Nat. Protoc.*, **16**, 1714–1739.
30. Untergasser,A., Cutcutache,I., Koressaar,T., Ye,J., Faircloth,B.C., Remm,M. and Rozen,S.G. (2012) Primer3—new capabilities and interfaces. *Nucleic Acids Res.*, **40**, e115.
31. Kircher,M., Sawyer,S. and Meyer,M. (2012) Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res.*, **40**, e3.
32. Renaud,G., Stenzel,U. and Kelso,J. (2014) leeHom: adaptor trimming and merging for Illumina sequencing reads. *Nucleic Acids Res.*, **42**, e141.
33. Li,H., Handsaker,B., Wysoker,A., Fennell,T., Ruan,J., Homer,N., Marth,G., Abecasis,G. and Durbin,R. (2009) The sequence alignment/map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.
34. 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.
35. Athey,J., Alexaki,A., Osipova,E., Rostovtsev,A., Santana-Quintero,L.V., Katneni,U., Simonyan,V. and Kimchi-Sarfaty,C. (2017) A new and updated resource for codon usage tables. *BMC Bioinf.*, **18**, 391.
36. Inui,M., Miyado,M., Igarashi,M., Tamano,M., Kubo,A., Yamashita,S., Asahara,H., Fukami,M. and Takada,S. (2014) Rapid generation of mouse models with defined point mutations by the CRISPR/Cas9 system. *Sci. Rep.*, **4**, 5396.
37. Bialk,P., Rivera-Torres,N., Strouse,B. and Kmiec,E.B. (2015) Regulation of gene editing activity directed by single-stranded oligonucleotides and CRISPR/Cas9 systems. *PLoS One*, **10**, e0129308.
38. Paquet,D., Kwart,D., Chen,A., Sproul,A., Jacob,S., Teo,S., Olsen,K.M., Gregg,A., Noggle,S. and Tessier-Lavigne,M. (2016)

Efficient introduction of specific homozygous and heterozygous mutations using CRISPR/Cas9. *Nature*, **533**, 125–129.
39. Cartegni,L., Chew,S.L. and Krainer,A.R. (2002) Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nat. Rev. Genet.*, **3**, 285–298.
40. Presnyak,V., Alhusaini,N., Chen,Y.-H., Martin,S., Morris,N., Kline,N., Olson,S., Weinberg,D., Baker,E., Kristian,Graveley *et al.* (2015) Codon optimality is a major determinant of mRNA stability. *Cell*, **160**, 1111–1124.
41. Hanson,G. and Coller,J. (2018) Codon optimality, bias and usage in translation and mRNA decay. *Nat. Rev. Mol. Cell Biol.*, **19**, 20–30.
42. Bae,H. and Coller,J. (2022) Codon optimality-mediated mRNA degradation: linking translational elongation to mRNA stability. *Mol. Cell*, **82**, 1467–1476.
43. Enache,O.M., Rendo,V., Abdusamad,M., Lam,D., Davison,D., Pal,S., Currimjee,N., Hess,J., Pantel,S., Nag,A. *et al.* (2020) Cas9 activates the p53 pathway and selects for p53-inactivating mutations. *Nat. Genet.*, **52**, 662–668.
44. Sinha,S., Barbosa,K., Cheng,K., Leiserson,M.D.M., Jain,P., Deshpande,A., Wilson,D.M., Ryan,B.M., Luo,J., Ronai,Z.e.A. *et al.* (2021) A systematic genome-wide mapping of oncogenic mutation selection during CRISPR-Cas9 genome editing. *Nat. Commun.*, **12**, 6512.
45. Renaud,J.-B., Boix,C., Charpentier,M., Cian,A.d., Cochennec,J., Duvernois-Berthet,E., Perrouault,L., Tesson,L., Edouard,J., Thinard,R. *et al.* (2016) Improved genome editing efficiency and flexibility using modified oligonucleotides with TALEN and CRISPR-Cas9 nucleases. *Cell Rep.*, **14**, 2263–2272.
46. Ghanta,K.S., Chen,Z., Mir,A., Dokshin,G.A., Krishnamurthy,P.M., Yoon,Y., Gallant,J., Xu,P., Zhang,X.-O., Ozturk,A.R. *et al.* (2021) 5′-Modifications improve potency and efficacy of DNA donors for precision genome editing. *Elife*, **10**, e72216.