

Published in final edited form as:

Nat Biotechnol. 2018 September ; 36(8): 765–771. doi:10.1038/nbt.4192.

Repair of double-strand breaks induced by CRISPR-Cas9 leads to large deletions and complex rearrangements

Michael Kosicki, Kärt Tomberg, and Allan Bradley*

Wellcome Sanger Institute, Hinxton, United Kingdom

Abstract

CRISPR/Cas9 is poised to become the gene editing tool of choice in clinical contexts. Thus far, exploration of Cas9-induced genetic alterations has been limited to the immediate vicinity of the target site and distal off-target sequences, leading to the conclusion that CRISPR/Cas9 was reasonably specific. Here we report significant on-target mutagenesis such as large deletions and more complex genomic rearrangements at the targeted sites in mouse embryonic stem cells, mouse hematopoietic progenitors and a human differentiated cell line. Using long-read sequencing and long-range PCR genotyping, we show that DNA breaks introduced by single-guide RNA/Cas9 frequently resolved into deletions extending over many kilobases. Furthermore, lesions distal to the cut site and cross-over events were identified. The observed genomic damage in mitotically active cells caused by CRISPR/Cas9 editing may have pathogenic consequences.

The utility of the CRISPR/Cas9 system for gene therapy in humans has been recognized and extensively investigated 1. Initial concerns about the off-target activity have been addressed by the development of sensitive detection methods, as well as modified Cas9 enzymes and improved delivery protocols that limit this type of damage 2–12. The vast majority of on-target DNA repair outcomes after Cas9 cutting in a variety of cell types are thought to be insertions and deletions (indels) of less than 20bp 13–15. Although indels a few hundred nucleotides in size were also observed in experiments using Cas9 or other nucleases, they were reported to be rare 16–18. Consequently, Cas9 has been assumed to be reasonably specific and the first approved clinical trials using Cas9 edited cells are underway (clinicaltrials.gov: NCT03081715, NCT03398967, NCT03166878, NCT02793856, NCT03044743, NCT03164135).

Studies using paired gRNAs to induce localized deletions also reported generation of more complex genotypes, such as inversions, endogenous and exogenous DNA insertions and larger-than-expected deletions 19–23. Single gRNAs were shown to induce deletions of up to 600bp in mouse zygotes 24. Deletions of up to 1.5kb in a haploid cancer cell line potentially induced by single gRNAs have been described, but since the guides were directed

*Correspondence and requests for materials, additional data and code, should be addressed to A.B. (abradley@sanger.ac.uk).

Data Availability and Accession Code Availability Statements: Pacbio sequencing data is accessible at the European Nucleotide Archive under accession numbers [ERS2396492](https://www.ebi.ac.uk/ena/record/ERS2396492) (PigA) and [ERS2396493](https://www.ebi.ac.uk/ena/record/ERS2396493) (Cd9). Barcoding information is in Supplementary Data 2.

Author contributions: M.K. performed most of the experiments and analysed the data. K.T. performed the primary cell work. A.B. supervised the project. All authors contributed to writing of the manuscript.

The authors declare no competing interests.

to a small part of the genome and provided as a pool, the possibility of rare double-cutting events cannot be excluded²⁵. Furthermore, the analysis of the alleles generated using both single and paired gRNAs has in most studies relied on amplification of short regions (<1kb) around the target and potential off-target sites, limiting the scope of assessment. Lesions non-contiguous with the cleavage site, such as those reported in yeast upon *I-SceI* nuclease cutting, would also be missed by such short range assessments^{26–28}. Finally, cancer cell lines, whose genome and DNA repair mechanisms are abnormal, were often used in the context of studying Cas9-induced lesions, making extrapolations to normal tissues and cells problematic.

We speculate that current assessments may have missed a substantial proportion of potential genotypes generated by on-target Cas9 cutting and repair, some of which may have potential pathogenic consequences following somatic editing of large populations of mitotically active cells.

We first comprehensively explored allelic diversity induced by Cas9 at the X-linked *PigA* locus, which is hemizygous in male embryonic stem (ES) cells. In contrast to cancer-derived cell lines, ES cells have a normal karyotype and intact DNA repair mechanisms, which makes them more representative of a normal somatic cell. Although mouse ES cells and embryonic fibroblast differ in their use of DNA repair pathways, it is not known how they compare to other somatic cells²⁹. We introduced Cas9 and gRNA constructs targeting intronic and exonic sites of *PigA* into JM8 mouse ES cells using *PiggyBac* transposition. Cells with both constructs were selected and subsequently stained with FLAER reagent to quantify the proportion of PigA-deficient cells (Fig. 1, A and B). Single gRNAs targeting exons 2 to 4 yielded very high rates of PigA loss (59-97%). Notably, single gRNAs targeting intronic sites also yielded PigA-deficient cells at significant frequencies. Ten different guides located between 263 and 520 bp from the nearest exon caused 8-20% PigA loss whereas two guides greater than 2kb away induced 5-7% loss (Fig. 1C and Table S1). We obtained similar results with transient expression using electroporation or lipofection of ribonucleoprotein complexes (RNP), proving that these observations were not a consequence of *PiggyBac* transposition, delivery method, antibiotic selection or cellular response to transfected plasmid DNA (Fig. S1). Lower knock-out efficiency using exonic guides correlated with slower editing dynamics when delivered by *PiggyBac* transposition (data not shown).

To understand what genetic changes underlie the generation of PigA-deficient cells, we amplified a 5.7kb region around exon 2 from pools of cells edited with three selected gRNAs introduced by *PiggyBac* transposition and sequenced the PCR products using the PacBio platform. We observed a depletion of read coverage on a kilobase-scale around the cut sites, consistent with presence of large deletions (Fig. 2A). Cells edited with intronic guides and sorted for loss of PigA generally exhibited loss of the adjacent exon. If intronic regulatory sequences were present around the exon, the DNA of cells sorted for retention of PigA expression would be wild-type or contain small indels around the cut site. However, the most frequent lesions in these cells were deletions extending many kilobases up or downstream, away from the exon. We conclude that, in most cases, loss of PigA expression was likely caused by loss of the exon, rather than damage to intronic regulatory elements.

Clustering of PacBio reads yielded 183 unique, edited, high quality alleles derived from three different gRNAs. These alleles ranged from simple deletions and insertions to complex rearrangements (Fig. S2, B, C and D, Table S2 and Supplementary Data 1). One of the alleles contained an insertion with a perfect match to four consecutive exons derived from the *Hmgn1* gene (Fig. 2B). We speculate this represents a *de novo* insertion from the spliced and reverse transcribed RNA, rather than from one of the pseudogenized forms of *Hmgn1*, as the pseudogenes diverge in sequence from the observed insertion.

To fully characterize a variety of edited *PigA* loci, single cell clones were isolated. The *PigA* loci around the gRNA target site were amplified using PCR primer pairs positioned progressively further apart (up to 16kb), until an amplicon was generated. These were sequenced using conventional Sanger sequencing technology (Fig. S2A). This strategy allowed us to recover an allele in most cases (133/141, 94%; Table S2 and S3, Supplementary Data 1).

Simple deletions overlapping both the cut site and the exon were found in almost three quarters (69/93) of *PigA*-deficient alleles generated by single, intronic gRNAs (Fig. S2, B and C). The deletions varied in size, the largest spanning 9.5kb. The remaining events were deletions combined with large insertions or more complex, multiple-lesion alleles. To assess the frequency of large deletions without strong selection for that outcome, we used an exonic gRNA causing 97% *PigA* loss. Although two thirds of alleles (32/48) from *PigA*-deficient cells had indels smaller than 50bp, as expected, more than 20% (10/48) exhibited deletions greater than 250bp, extending up to 6kb (Fig. S2D). Because the deletions generated with the exonic gRNA were bi-directional, this is consistent with the average frequency of generating *PigA*-deficient cells with intronic guides positioned 263-520bp from an exon (~12%). We obtained similar results using electroporation of RNP (Fig. S1B).

Notably, 23 of 133 recovered alleles contained additional lesions (SNPs, indels, large deletions and insertions) that were non-contiguous with the lesion at the cut site. In 13 out of 23 cases, the only exonic lesion detected was non-contiguous with the cut site (Fig. 2C). Furthermore, we observed alleles in which the intronic gRNA caused an inversion of a region containing the exon (Fig. 2D). Had the assessment been limited to the immediate vicinity of the cleavage site, such alleles would have been misclassified as wild-type and their phenotypic consequences would have been underestimated.

Insertions were present in 35 out of 133 recovered alleles. We could not find convincing local mapping for insertions shorter than 7bp (13 alleles), which we speculate to be mostly non-templated nucleotides. The large majority of other insertions were constituted from sequence, which mapped to the *PigA* locus and encompassed inversions and duplications ranging from 11bp to 2.5kb (17 alleles; Fig. 2C, 2D, S2C). The remaining 5 alleles contained DNA sequences that mapped to other parts of the mouse genome, such as interspersed repeats or to exogenous, transfected sequences.

Six alleles did not contain lesions overlapping the nearest exon. Three of these were also wild-type around the cut sites and are likely to contain lesions in other exons or larger rearrangements. The remaining three alleles only contained intronic lesions, which may

interfere with splicing. In eight cases, it was not possible to recover any product with exon spanning primers (Fig. S3A, black primer pairs). To understand this class of events, we performed additional PCRs targeting each end of the *PigA* locus (Fig. S3A, grey primer pairs). In three cases, just one end or neither end of the locus could be amplified, suggesting a larger deletion. In the remaining five cases, both ends amplified. Since no product connecting the two ends could be obtained, these are likely to be translocations, inversions or large insertions (Table S4).

To understand the diversity of potential deletion outcomes, we have repeated our original experiment in biological quadruplicate using the 5' intronic gRNA. Cells with large deletions were enriched by sorting for *PigA* negative cells and deletion fingerprints were generated by PCR. Each biological replicate differed substantially, despite a large number of unique deletion events sampled, indicating that the diversity of potential deletion outcomes is vast (Fig. S4 and Supplementary Note).

Given that *PigA* is mono-allelic in the XY ES cells used in this study we wished to exclude the possibility that the observations reflect some peculiarity of the lack of a homologue. The autosomal *Cd9* locus was selected for this purpose as it is non-essential in ES cells and its protein product can be readily detected by cell surface staining. An exonic guide yielded 88% *Cd9* loss, while 5' and 3' intronic guides generated 4.2% and 5.4% *Cd9* loss, respectively (Fig. 3, A and B, Table S1A). Taking into account a 1.6% background of *Cd9*^{low} cells in the untransfected condition we estimate the true proportion of *Cd9* loss due to intronic cutting to be between 2.6-3.8%. This is consistent with results at the *PigA* locus, assuming haplosufficiency of *Cd9* and thus a requirement to edit both alleles to destroy *Cd9*.

To describe the genetic events underlying *Cd9* loss, we isolated single cell clones edited with the 3' intronic guide, ascertained their expression status by flow cytometry and sequenced the area around the cut site using PacBio and Sanger technologies. The largest deletion spanned 5.5kb. A pileup of 185 resolved alleles derived from 93 single cell clones shows a clear enrichment for deletions overlapping the exon in clones negative for *Cd9* compared to positive clones and ones exhibiting a mixture of *Cd9*⁺ and *Cd9*⁻ cells (Fig. 3C). The bimodal expression pattern of some of the clones may be the result of a mixed clone or a protracted repair event that was resolved during clone outgrowth. The haplosufficient nature of the *Cd9* gene is demonstrated by the fact we could detect at least one allele with an intact exon in all but one of the 66 *Cd9* positive and mixed clones. Similarly, only one of the 27 *Cd9* negative clones had an intact exon, this exception presumably harbouring other undetected lesions. We have further confirmed by PCR genotyping that large deletions are a common outcome of single cell clones edited at the *Cd9* locus using additional intronic and exonic guides (Table S5, Supplementary Note).

The experiment at the *Cd9* locus was performed in mouse ES cells derived from an F1 cross between BL6 and CAST mouse strains, which allowed us to distinguish the homologous chromosomes. In no case was the repair outcome identical between homologous within a clone, despite 15 alleles reoccurring between clones. This result is consistent with the great diversity of outcomes at the *PigA* locus. Just over half of the edited clones (52 out of 93) contained precisely one CAST and one BL6 allele, as expected. Notably, in 18 clones only

one allele was detected, potentially due to translocations, very large deletions, insertions or inversions, monosomy or loss of heterozygosity (LOH) either local or chromosome-wide. 21 clones contained an abnormal number of alleles, which could have resulted from a mixed clone, large duplication, repair events happening during clone outgrowth or aneuploidy induced by Cas9 cutting. Finally, two clones contained recombinant BL6-CAST alleles (Fig. 3D). In one case, a LOH event distal to the breakpoints converted part of the CAST allele to BL6. In another case, the BL6-CAST cross-over boundary did not coincide with the breakpoint. We conclude that the creation of these alleles likely involved inter-homologue strand invasion as they cannot be explained by a simple re-joining of the resected ends of two broken chromosomes.

To investigate whether the observed on-target extensive DNA repair-associated damage is an intrinsic property of undifferentiated mouse ES cells, we examined the consequence of editing in a human differentiated cell line. An immortalized human female pigmented retinal epithelial cell line (RPE1) was used. Although this is a female cell line, X-inactivation renders it functionally hemizygous at the *PIGA* locus. Editing *PIGA* with single exonic and intronic gRNAs delivered with *PiggyBac* vectors, resulted in a loss of *PIGA* at frequencies comparable to those observed in mouse ES cells (Fig. 4 A-B). PCR genotyping and Sanger sequencing of 41 *PIGA*-deficient single cell clones edited with intronic gRNAs revealed large deletions, insertions, inversions and non-contiguous lesions overlapping the exon (Fig. 4 C-E). In some clones only one small, intronic indel allele was detected, which we interpret as an inconsequential edit of the inactive chromosome coupled with a loss-of-function lesion on the active X-chromosome that inactivates one or both primer binding sites.

Similar results were obtained in lineage negative cells from the bone marrow of mice homozygous for a *Cas9-GFP* cassette at the *Rosa26* locus. Stem cells enriched by removal of differentiated cells on magnetic columns were electroporated with a crRNA:trRNA complex against the GFP locus, GFP-negative single cell clones were isolated and genotyped around the cut site with three different primer pairs spanning up to 3.6kb. At least one large deletion product between 100bp and ~3kb in size was detected in 35 out of 96 clones (Fig. S5A-B). We verified eight deletion products by Sanger sequencing across the deletion junction (Fig. S6A). Only wild-type size products were detected in the remaining clones and none of the 96 control clones exhibited any deletion bands (Table S5, “progenitor” experiment).

The editing in this study was conducted at actively transcribed loci in normal embryonic stem cells and progenitor cells both with intact DNA repair processes, as well as in immortalized, differentiated human cell line, surrogates for various clinical editing applications. We show that extensive on-target genomic damage is a common outcome at all loci and in all cell lines tested. Moreover, the genetic consequences observed are not limited to the target locus, as events such as loss-of-heterozygosity will uncover recessive alleles, whereas translocations, inversions and deletions will elicit long range transcriptional consequences. Given that a target locus would presumably be transcriptionally active, mutations that juxtapose this to one of the hundreds of cancer-driver genes may initiate neoplasia. In the clinical context of editing many billions of cells the multitude of different mutations generated make it likely that one or more edited cells in each protocol would be

endowed with an important pathogenic lesion. Such lesions may constitute a first carcinogenic “hit” in stem cells and progenitors, which have a long replicative lifespan and may become neoplastic with time. Such a circumstance would be similar to the activation of *LMO2* by pro-viral insertion in some of the early gene therapy trials, which caused cancer in these patients 30. Results reported here also illustrate a need to thoroughly examine the genome when editing is conducted *ex vivo*. As genetic damage is frequent, extensive and undetectable by the short-range PCR assays that are commonly used, comprehensive genomic analysis is warranted to identify cells with normal genomes prior to patient administration.

Online Methods

Mouse Embryonic Stem Cell Culture and Transfection

gRNA-expression vectors contain a U6 promoter with a „F+E” scaffold 31 and a *PGK-Puro-2A-BFP* cassette, flanked by *PiggyBac* repeats. The Cas9-expression vector contains a *Cas9-Blast* cassette expressed from a short EF1 α promoter in a pKLV backbone 13,32. CAST/BL6 (CB9; a gift from Prof. A. Fergusson-Smith), AB2.2 mCherry/GFP reporter (a gift from Dr. Xiufei Gao and Prof. P. Liu) or JM8.A3 mouse embryonic stem cells 33,34 were cultured in M15 media (High-Glucose DMEM, with 15% FCS, beta-mercaptoethanol and L-Glutamate) on SNLP feeder cells.

Complexes of lipofectamine LTX (2.5 μ l), plus reagent (0.5 μ l), 200ng hyperactive *PiggyBac* transposase 35, 100ng of the *PiggyBac Cas9-Blast* plasmid and 50ng of the *PiggyBac gRNA-Puro* plasmid were prepared in 50 μ l OptiMEM following manufacturer's instructions. Cells were trypsinized, washed in M15, re-suspended in M15+LIF and seeded onto a gelatinized 24 well plate, containing the lipofectamine DNA complexes, at 3×10^5 cells per well. From day 2, M15+LIF media containing puromycin (3 μ g/ml) and blasticidin (10 μ g/ml) was used. The same setup was used for RPE1 cell line, except the *Cas9-Blast* plasmid was omitted. A similar setup was used for lipofection of RNP complexes with 20pmol of both hybridized crRNA:trRNA (Sigma) and EnGen Cas9 NLS (NEB). Neon Transfection System (Thermo Fisher Scientific; 1600v / 10ms / 3pulses) was used for electroporation of 1.5×10^5 cells in buffer R with 6pmol each of crRNA:trRNA, electroporation enhancer (IDT) and Cas9 protein or 9pmol each of crRNA:trRNA and Cas9 protein. Around 3×10^5 cells were collected on day 14 (or day 17, in case of the RPE1 cells), stained in PBS+0.1% BSA for 30' at room temperature with 1 μ g/ml FLAER reagent (Cedarlane) or anti-Cd9-PE antibody (cat 124805, Biolegend), washed twice and analysed using a Cytoflex flow cytometer. For single cell cloning and PacBio experiments, cells were transfected in 6 well plates with five times more cells and reagents, expanded onto 10cm dishes and FACS sorted for loss of FLAER or Cd9 staining on day 14 using MoFlow XDP (Beckman Coulter). Single cell clones were isolated and grown in 96 well plates. DNA was extracted by proteinase K digestion followed by ethanol precipitation. PCR reaction were conducted using primers in Table S6 and LongAMP polymerase (NEB) following manufacturer's instructions.

Bioinformatics

Primers were designed using Primer3-BLAST (Table S6). Guide RNAs were designed using Benchling and CRISPRscan 36. Alignment of Sanger sequenced PCR products from *PigA* and RPE1 experiments was performed using BLAT (v 36) and converted into BAM format using a customized script from Tobias Marschall (<https://github.com/ALLBio/allbiotc2/tree/master/synthetic-benchmark>). Mixed traces were resolved using PolyPeakParser 37. Analysis of PacBio data was performed using command line version of SMRT-Link software (pbtranscript 1.0.1.TAG-1470). For *PigA* locus pileup, circular consensus sequences were called with at least one full pass and minimum predicted accuracy of 0.9. Individual *PigA* and *Cd9* alleles were reconstructed by following “Running Iso Seq using SMRTLink” tutorial on github, except “--targeted_isoseq” option was used at the clustering step. Resulting alleles were mapped to the reference genome using bwa mem (v 0.7.17-r1188). In case of the *PigA* locus, mapped reads were clustered further using a custom script. Genome coverage was calculated with “bedtools genomecov -dz” (v 2.27.1) using circular consensus sequences (*PigA* locus) or reconstructed alleles (*Cd9* locus). All downstream analysis was performed using custom R (v 3.3.2) and bash scripts and visualized with ggplot2 package. Flow cytometric data were processed with FlowJo (v 10.4.1).

Mouse Bone Marrow Cell Culture and Transfection

Bone marrow cells from a homozygous C57BL/6 CAS9-EGFP knock-in mouse 38 were isolated by flushing tibias and femurs in HBSS (Life Technologies) supplemented with 2% FBS and 10mM HEPES (Sigma). Lineage negative cells were isolated using Direct Lineage Cell Depletion Kit Mouse (Miltenyl Biotec) and cultured in X-Vivo (Lonza) with 2% FBS, 50ng/ml stem cell factor, 50ng/ml thrombopoietin, 10ng/ml IL-6 (Peprotech). After culturing for 3h, 1×10^5 cells were electroporated (1550v / 20 ms / 1 pulse) in buffer T with 44pmols of preassembled crRNA:trRNA duplex (guide #311, Table S1; IDT) using the Neon Transfection System. GFP-negative cells were sorted 4 days after the electroporation and plated into Methocult M3434 media (6000 cells per 3ml, StemCell Technologies). Seven days later, single colonies were picked into 25 μ l of direct PCR lysis buffer (Peqlab) with 1 μ g/ml proteinase K and analysed by PCR (Fig. S5, Table S5).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We wish to thank M. Friedrich for sharing his gRNA expression construct and technical advice, E. Metzakopian for technical advice and critical reading of the manuscript, G. Rutledge for critical reading of the early manuscript, A. Ferguson-Smith for the CAST/B6 hybrid ES cells, P. Liu and X. Gao for mCherry/GFP reporter cells, S. Jackson’s group for the Cas9-expressing RPE1 cell line and the Cytometry Core Facility for assistance with cell sorting. This work was supported by the Wellcome Trust Grant number 098051.

References

1. Cornu TI, Mussolino C, Cathomen T. Refining strategies to translate genome editing to the clinic. *Nat Med.* 2017; 23:415–423. [PubMed: 28388605]

2. Kim S, Kim D, Cho SW, Jungeun K, Kim J-S. Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome Res.* 2014; 128:1–32.
3. Komor AC, Kim YB, Packer MS, Zuris JA, Liu DR. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature.* 2016; 61:5985–91.
4. Frock RL, et al. Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases. *Nat Biotechnol.* 2015; 33:179–86. [PubMed: 25503383]
5. Xie F, et al. Seamless gene correction of α -thalassemia mutations in patient-specific iPSCs using CRISPR/Cas9 and piggyBac. *Genome Res.* 2014; 24:1526–1533. [PubMed: 25096406]
6. Guilinger JP, Thompson DB, Liu DR. Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. *Nat Biotechnol.* 2014; 32:577–582. [PubMed: 24770324]
7. Kleinstiver BP, et al. High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects. *Nature.* 2016; 529:490–495. [PubMed: 26735016]
8. Ran FA, et al. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell.* 2013; 154:1380–9. [PubMed: 23992846]
9. Slaymaker IM, et al. Rationally engineered Cas9 nucleases with improved specificity. *Science* (80-.). 2016; 351:84–88.
10. Tsai SQ, et al. Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. *Nat Biotechnol.* 2014; 32:569–576. [PubMed: 24770325]
11. Fu Y, et al. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotechnol.* 2013; 31:822–6. [PubMed: 23792628]
12. Tsai SQ, et al. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat Biotechnol.* 2015; 33:187–197. [PubMed: 25513782]
13. Koike-Yusa H, Li Y, Tan E-P, Velasco-Herrera MDC, Yusa K. Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. *Nat Biotechnol.* 2014; 32:267–73. [PubMed: 24535568]
14. Van Overbeek M, et al. DNA Repair Profiling Reveals Nonrandom Outcomes at Cas9-Mediated Breaks. *Mol Cell.* 2016; :1–14. DOI: 10.1016/j.molcel.2016.06.037
15. Tan E, Li Y, Velasco-herrera MDC, Yusa K, Bradley A. Off-Target Assessment of CRISPR-Cas9 Guiding RNAs in Human iPS and Mouse ES Cells. 2015; 236:225–236.
16. Hendel A, et al. Chemically modified guide RNAs enhance CRISPR-Cas genome editing in human primary cells. *Nat Biotechnol.* 2015; 33:985–989. [PubMed: 26121415]
17. Ghezraoui H, et al. Chromosomal Translocations in Human Cells Are Generated by Canonical Nonhomologous End-Joining. *Mol Cell.* 2014; 55:829–842. [PubMed: 25201414]
18. Weinstock DM, Elliott B, Jasin M. A model of oncogenic rearrangements: Differences between chromosomal translocation mechanisms and simple double-strand break repair. *Blood.* 2006; 107:777–780. [PubMed: 16195334]
19. Canver MC, et al. Characterization of Genomic Deletion Efficiency Mediated by CRISPR/Cas9 in Mammalian Cells. *J Biol Chem.* 2014; doi: 10.1074/jbc.M114.564625
20. Kraft K, et al. Deletions, inversions, duplications: Engineering of structural variants using CRISPR/Cas in mice. *Cell Rep.* 2015; 10:833–839. [PubMed: 25660031]
21. Boroviak K, Doe B, Banerjee R, Yang F, Bradley A. Chromosome engineering in zygotes with CRISPR/Cas9. *Genesis.* 2016; 54:78–85. [PubMed: 26742453]
22. Boroviak K, Fu B, Yang F, Doe B, Bradley A. Revealing hidden complexities of genomic rearrangements generated with Cas9. *Sci Rep.* 2017; :1–8. DOI: 10.1038/s41598-017-12740-6 [PubMed: 28127051]
23. Parikh BA, Beckman DL, Patel SJ, White JM. Detailed Phenotypic and Molecular Analyses of Genetically Modified Mice Generated by CRISPR-Cas9-Mediated Editing. *PLoS One.* 2015; :1–28. DOI: 10.1371/journal.pone.0116484
24. Shin HY, et al. CRISPR/Cas9 targeting events cause complex deletions and insertions at 17 sites in the mouse genome. *Nat Commun.* 2017; 8:1–10. [PubMed: 28232747]

25. Gasperini M, et al. CRISPR/Cas9-Mediated Scanning for Regulatory Elements Required for HPRT1 Expression via Thousands of Large, Programmed Genomic Deletions. *Am J Hum Genet.* 2017; 101:192–205. [PubMed: 28712454]
26. Roberts SA, et al. Clustered Mutations in Yeast and in Human Cancers Can Arise from Damaged Long Single-Strand DNA Regions. *Mol Cell.* 2012; 46:424–435. [PubMed: 22607975]
27. Sinha S, et al. Microhomology-mediated end joining induces hypermutagenesis at breakpoint junctions. *PLOS Genet.* 2017; 13:e1006714. [PubMed: 28419093]
28. Yang Y, Sterling J, Storici F, Resnick MA, Gordenin DA. Hypermutability of damaged single-strand DNA formed at double-strand breaks and uncapped telomeres in yeast *Saccharomyces cerevisiae*. *PLoS Genet.* 2008; 4
29. Tichy ED, et al. Mouse Embryonic Stem Cells, but Not Somatic Cells, Predominantly Use Homologous Recombination to Repair Double-Strand DNA Breaks. *Stem Cells Dev.* 2010; 19:1699–1711. [PubMed: 20446816]
30. Hacein-Bey-Abina S, et al. A Serious Adverse Event after Successful Gene Therapy for X-Linked Severe Combined Immunodeficiency. *N Engl J Med.* 2003; 348:255–256. [PubMed: 12529469]
31. Chen B, et al. Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell.* 2013; 155:1479–1491. [PubMed: 24360272]
32. Hsu PD, et al. Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells. *Nat Biotechnol.* 2013; 31:827–32. [PubMed: 23873081]
33. Strogantsev R, et al. Allele-specific binding of ZFP57 in the epigenetic regulation of imprinted and non-imprinted monoallelic expression. *Genome Biol.* 2015; :1–18. DOI: 10.1186/s13059-015-0672-7 [PubMed: 25583448]
34. Pettitt SJ, et al. Agouti C57BL/6N embryonic stem cells for mouse genetic resources. *Nat Methods.* 2009; 6:493–495. [PubMed: 19525957]
35. Yusa K, Zhou L, Li MA, Bradley A, Craig NL. A hyperactive piggyBac transposase for mammalian applications. *Proc Natl Acad Sci U S A.* 2011; 108:1531–6. [PubMed: 21205896]
36. Moreno-Mateos, Ma; , et al. CRISPRscan: designing highly efficient sgRNAs for CRISPR-Cas9 targeting in vivo. *Nat Meth.* 2015; 12:982–988.
37. Hill JT, et al. Poly peak parser: Method and software for identification of unknown indels using sanger sequencing of polymerase chain reaction products. *Dev Dyn.* 2014; 243:1632–1636. [PubMed: 25160973]
38. Platt RJ, et al. CRISPR-Cas9 Knockin Mice for Genome Editing and Cancer Modeling. *Cell.* 2014; 159:440–455. [PubMed: 25263330]

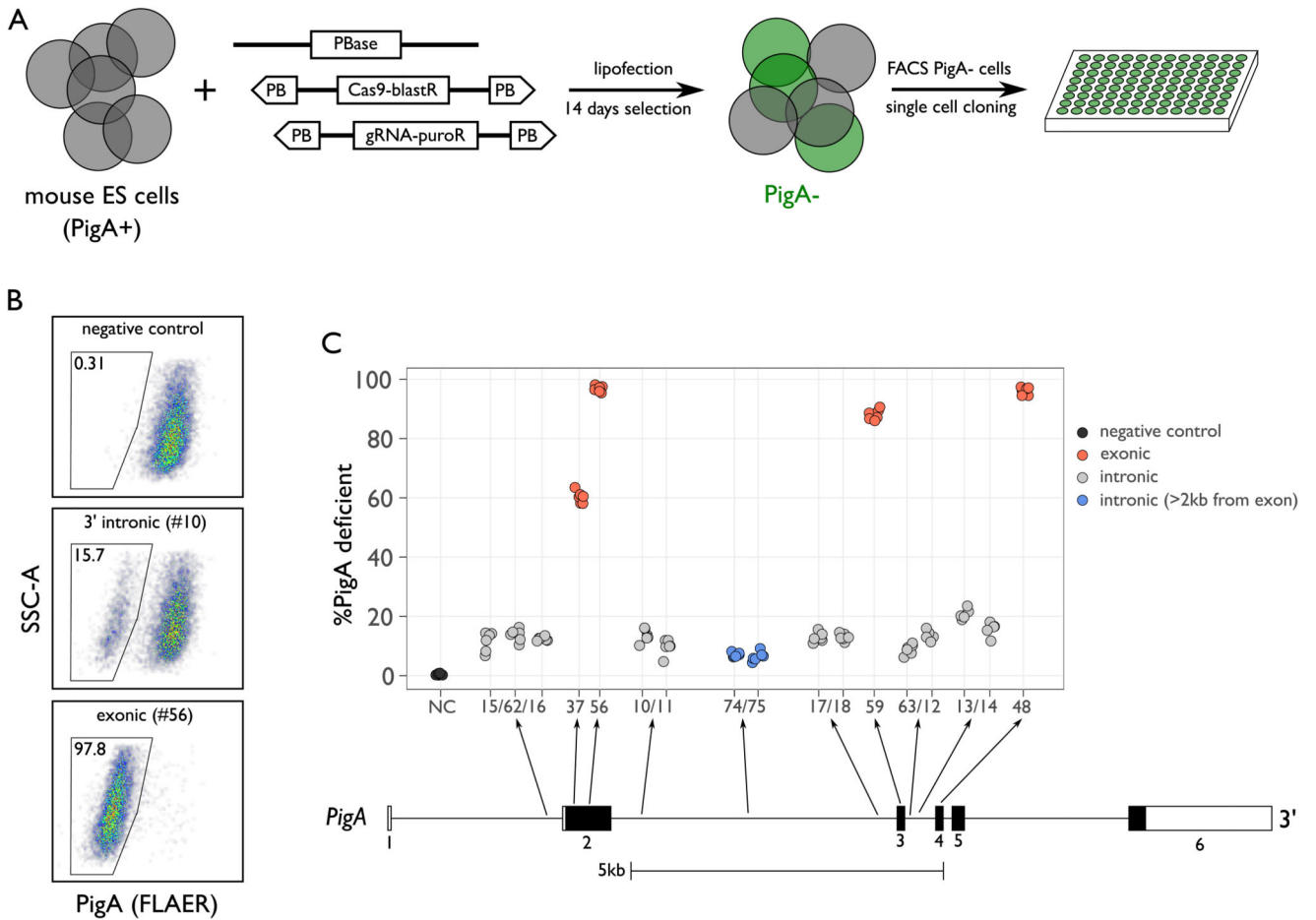


Figure 1. Frequency of *PigA* loss upon editing with exonic and intronic gRNAs in mouse ES cells. (A) Experimental design. Cells were transfected with separate *PiggyBac* transposons carrying gRNA and *Cas9* genes and selected for stable transposition. *PigA* negative cells (green) were sorted, single cell clones isolated, the region around the cut site amplified, sequenced and mapped to the reference genome. (B) Examples of *PigA* editing revealed by FLAER staining, for two gRNAs and one control. (C) Frequency of *PigA* loss caused by *Cas9* with intronic and exonic gRNAs (Table S1; N=6 biologically independent cell cultures). NC = negative control, a guide targeting the *Cd9* gene. Thick bars represent exons, hollow ones indicate UTRs.

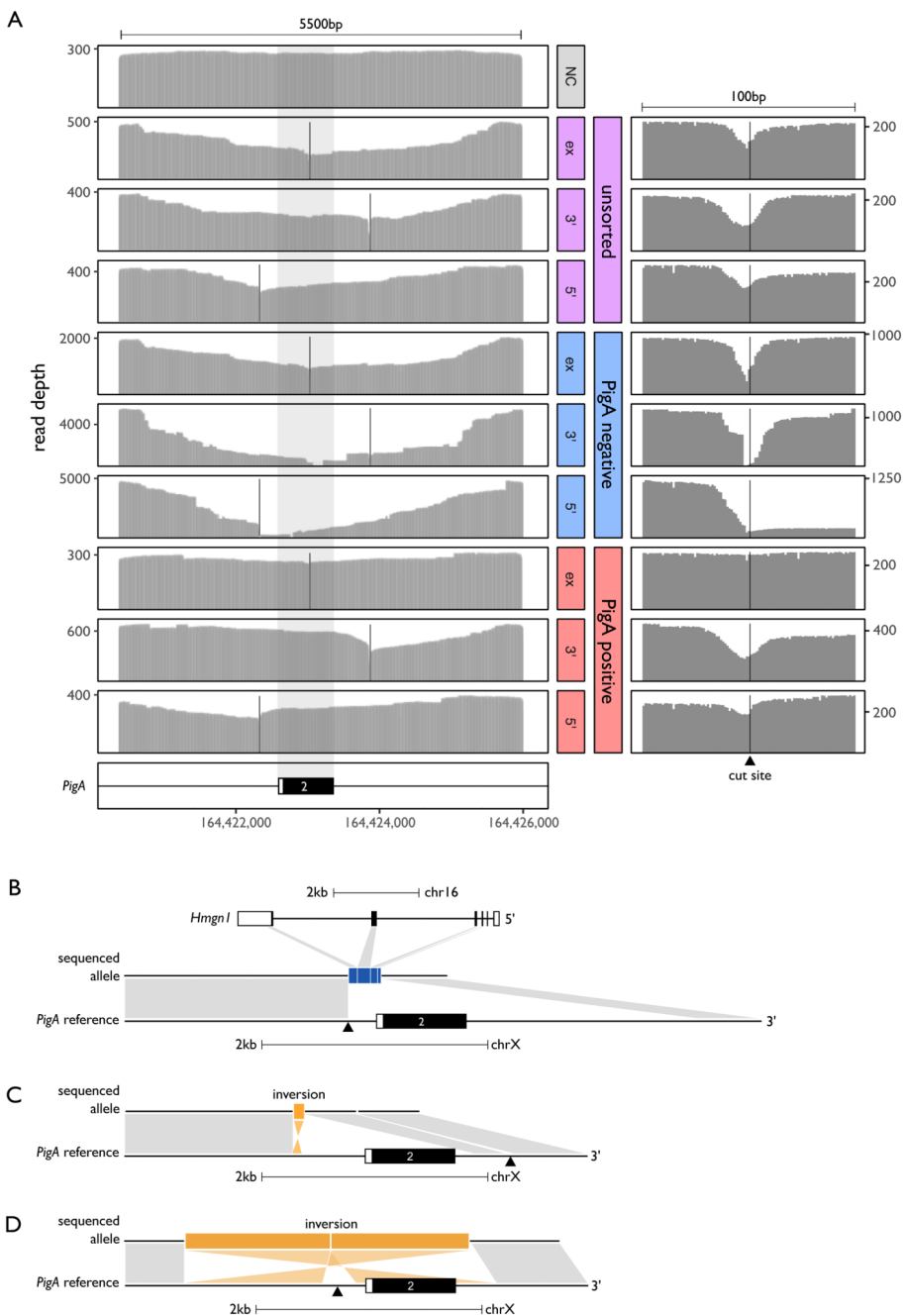


Figure 2. Analysis of the *PigA* locus edited with selected gRNAs.

(A) Coverage of PacBio reads at the *PigA* locus. The locus was PCR amplified from a pool of cells sorted for *PigA* expression (or from the unsorted population) and the resulting products were sequenced using the PacBio platform. The right panel depicts a 100bp region centered at the cut site. NC: negative control gRNA, ex: exonic gRNA (#56), 5': 5' intronic gRNA (#15), 3': 3' intronic gRNA (#10). The cut site of the gRNA (between 3rd and 4th nucleotide from the PAM sequence) is indicated with a vertical black bar. Genomic position is given with respect to the GRCm38 reference genome. N=1. (B-D) Examples of alleles.

The bottom diagram represents the *PigA* reference allele around exon 2, the diagram immediately above shows the structure of the sequenced allele. The top diagram in **B** shows the genomic *Hmgn1* gene structure, note the different scale. Black horizontal line – direct reference match, orange bar – inversion, blue bar – insertion from another part of the genome, black arrowhead – gRNA target site. Grey and orange shadows represent, respectively, direct and inverted match between the reference and the sequenced allele. Lack of shadow at the reference locus represents a deletion in the sequenced allele.

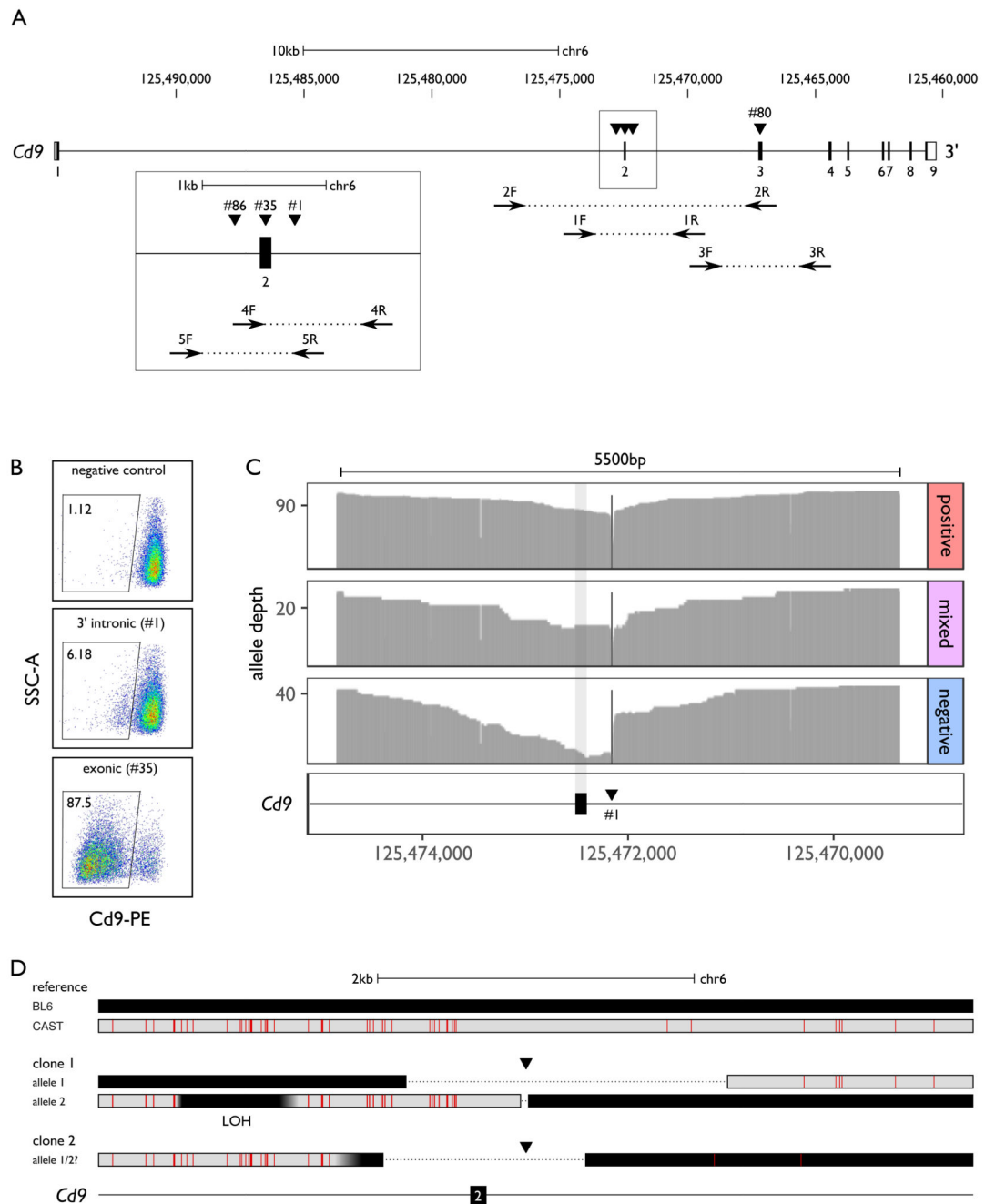


Figure 3. Analysis of Cas9 editing at the autosomal *Cd9* locus in mouse ES cells.

Experimental setup is analogous to the PigA experiment in Fig. 1A. A mouse ES cell line derived from an F1 cross between *Mus musculus castaneus* (CAST) and *Mus musculus* (BL6) was used. (A) Positions of primer pairs and gRNAs (Tables S1 and S6). Genomic position is given with respect to the GRCm38 reference genome. (B) Examples of *Cd9* editing revealed by antibody staining, for two gRNAs and one control (Table S1; N=7 biologically independent cell cultures). (C) PacBio alleles derived from *Cd9* positive, mixed (bimodal) and negative, individually sequenced single cell clones, displayed as a pileup. (D) Genomic map of the *Cd9* locus on chromosome 6, showing LOH (Loss of Heterozygosity) for allele 1/2 in clone 1 and clone 2.

Display conventions as in Fig. 2. N=1. **(D)** Recombinant alleles. Two of the sequenced single cell clones contained alleles indicative of a cross over event between the homologous chromosomes. Red vertical bars in CAST allele (grey bar) indicate positions of sequence divergence from the BL6 reference genome (black bar), dotted black line indicates missing sequence (deletion), thin black line indicates an intron. LOH – loss of heterozygosity.

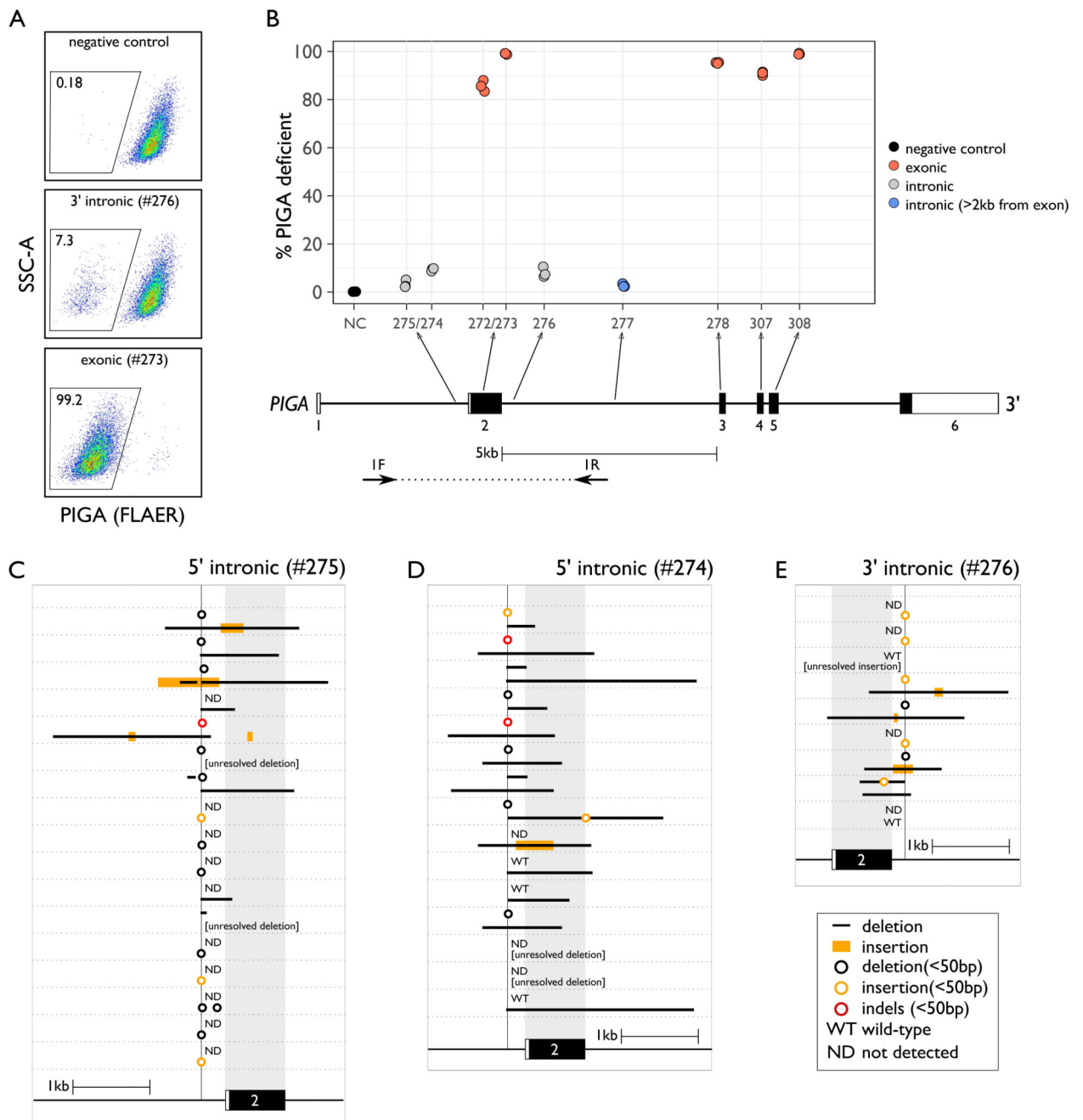


Figure 4. Frequency of PIGA loss upon editing with exonic and intronic gRNAs and structure of the recovered alleles in human RPE1 cells.

Cas9-expressing cells were transfected with *PiggyBac* transposons carrying a gRNA and selected for stable transposition. PIGA negative cells were sorted, single cell clones isolated, the region around the cut site amplified, sequenced and mapped to the reference genome.

(A) Examples of *PIGA* editing revealed by FLAER staining, for two gRNAs and one control. (B) Frequency of PIGA loss caused by Cas9 with intronic and exonic gRNAs (Table S1; N=3 biologically independent cell cultures). Position of the primers with the largest span

(6kb) is indicated. **(C-E)** Recovered alleles. **(C)** 5' intronic guide #275, **(C)** 5' intronic guide #274, **(D)** 3' intronic guide #276. The position of the gRNA is shown as a vertical line intersecting with the *PIGA* gene structure. Pure insertions and deletions of <50bp are indicated with orange and black circles, respectively. Combined insertion/deletion events of <50bp and SNPs ('indels') are indicated with a red circle. Black lines represent deletions >50bp. Orange bars indicate size of the >50bp insertions (but not their map position). They are centred on the insertion locus or on the associated deletion. Thin, horizontal, dashed line separates clones.