

Application of long-read sequencing for robust identification of correct alleles in genome edited animals

Christopher V. McCabe^{1,†}, Gemma F. Codner^{1,†}, Alasdair J. Allan¹, Adam Caulder¹, Skevoulla Christou¹, Jorik Loeffler¹, Matthew Mackenzie¹, Elke Malzer¹, Joffrey Mianné¹, Fran J. Pike¹, Marie Hutchison¹, Michelle E. Stewart¹, Hilary Gates³, Sara Wells¹, Nicholas D. Sanderson^{2,*} and Lydia Teboul^{1,*}

Abstract

Recent developments in CRISPR/Cas9 genome editing tools have facilitated the introduction of more complex alleles, often spanning genetic intervals of several kilobases, directly into the embryo. These techniques often produce mosaic founder animals and the introduction of donor templates, via homologous directed repair, can be erroneous or incomplete. Newly generated alleles must be verified at the sequence level across the targeted locus. Screening for the presence of the desired mutant allele using traditional sequencing methods can be challenging due to the size of the desired edit(s) together with founder mosaicism. In order to help disentangle the genetic complexity of these animals, we tested the application of Oxford Nanopore long read sequencing of the targeted locus. Taking advantage of sequencing the entire length of the segment in each single read, we were able to determine whether the entire intended mutant sequence was present in both mosaic founders and their offspring.

Key words: Allele validation, floxed, CRISPR/Cas9, homologous recombination, long single-stranded DNA, long-read sequencing, mouse, mutant, Nanopore.

¹The Mary Lyon Centre and ³Mammalian Genetics Unit, MRC Harwell Institute, Didcot, Oxon, OX11 0RD, UK.

²Nuffield Department of Clinical Medicine, University of Oxford, John Radcliffe Hospital, Oxford, UK.

[†]These authors equally contributed.

^{*}Corresponding author

Corresponding authors:

Nicholas D. Sanderson, Nuffield Department of Clinical Medicine, University of Oxford, John Radcliffe Hospital, Oxford, UK.

nicholas.sanderson@ndm.ox.ac.uk

Lydia Teboul, The Mary Lyon Centre, MRC Harwell Institute, Didcot, Oxon, OX11 0RD, UK.

Telephone: +44 1235 841097

l.teboul@har.mrc.ac.uk

Background

Genome editing tools in conjunction with a single-stranded oligonucleotide (ssODN) donor are an effective method for the introduction of specific mutations in early embryos (Wang et al., 2013; Yang et al., 2013). However, this strategy often produces complex mosaic animals in the founder (G_0) generation (Singh et al., 2015; Mianné et al., 2017). Each of the founder alleles with evidence of the desired edits must be fully sequenced in order to detect unwanted mutations *in cis* (Mianné et al., 2016; Renaud et al., 2016; Birling et al., 2017).

Previously, Sanger sequencing has been sufficient to characterise CRISPR/Cas9 mutagenised loci using ssODN donors. These templates were a maximum of 200 bases in length (Mianné et al., 2017), which can be easily covered within a Sanger sequencing read. However, long single-stranded DNA (IssDNA) donors (Quadros et al., 2017; Codner et al., 2018) or multiple ssODNs (Yang et al., 2013; Lanza et al., 2018) can be used for the generation of complex alleles directly in one-cell embryos. Targeted edits spanning several kilobases (kb) are now being produced with increasing regularity. In order to cover intervals of this size, and piece together each of the many allele variants generated in a mosaic founder, several 500- to 800 bp Sanger sequencing reads are

required and subsequently combined *in silico* (Figure 1 and Supplemental Figure 1, process highlighted in orange). As such, the characterisation of mutant alleles in mosaic founders is particularly challenging. Events that are not captured by the chosen assays can be omitted (Shin et al., 2017; Kosicki et al., 2018; Owens et al., 2019) and screening can sometimes fail to distinguish rearranged from correct alleles in these complex animals (Codner et al., 2018). This suggests that the screen based on Sanger sequencing produced some false positives.

Oxford Nanopore Technologies (ONT) sequencing produces far longer reads, which can easily cover the entire length of the mutagenised interval in one molecule (Jain et al., 2016). We piloted the use of ONT sequencing as an alternative method for identifying the presence of the correctly mutated allele in mosaic founders, derived from the microinjection of CRISPR/Cas9 reagents and IssDNA donors, and their progeny (G_1 ; Figure 1 and process highlighted in blue in Supplemental Figure 1). We showed that the high error rate inherent to ONT sequencing can be offset by very deep sequencing coverage. We assembled a workflow for the analysis of sequencing data to identify the genomic DNA samples that include the correct allele. We found that ONT sequencing provides an accurate screen of these animals.

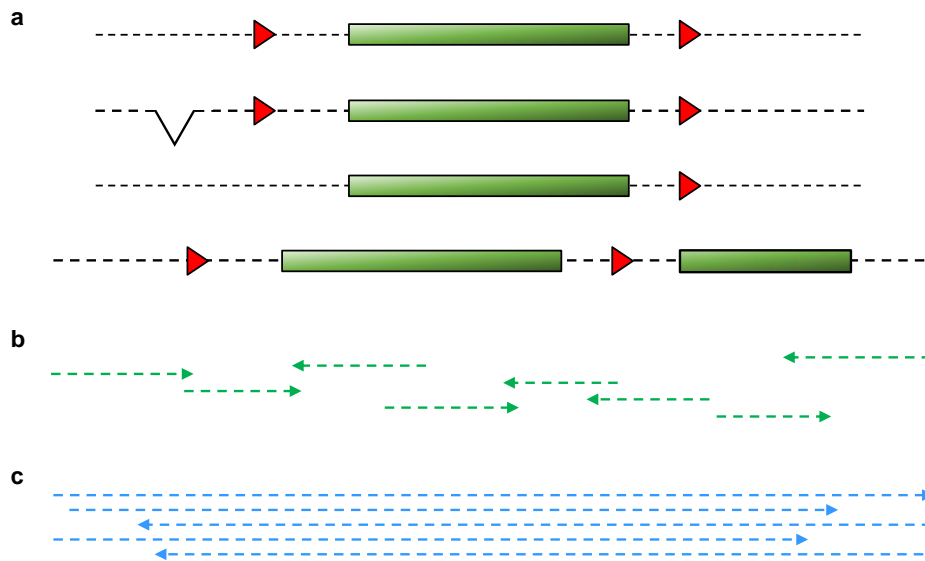


Figure 1: Disentangling sequences from a mosaic animal. (a) illustrates the complex genetic make up at a targeted locus in a mosaic (four alleles). Red triangles correspond to loxP sites, green boxes indicate the floxed exon. The mosaic animal contains: correct floxed allele, floxed allele with a deletion upstream, a partial integration of the donor (3' LoxP only) and a donor integration with a duplicated segment. (b) illustrates how several Sanger sequencing reads are required to span the whole interval interest. Note that it is impossible to ascertain how reads should be assembled to span each specific allele. (c) illustrates how each long read spans the whole interval of interest sequenced within a mosaic.

Importantly, long reads allow for the earlier exclusion of founder animals that were identified as positive for the presence of a correct integration by Sanger sequencing based screen but only transmitted incorrectly mutated alleles. The application of ONT sequencing for the screening of founders obtained with IssDNA donors represents an advance for ethical animal use, as it prevents breeding of some false-positive founders.

Results

Establishing an accurate ONT-based targeted sequencing screening process

As ONT has a higher error rate than other next generation sequencing technologies (Jain et al., 2016), we

first assessed the feasibility of unequivocally recognising known sequences and defined which quality thresholds can be used for such analysis. We analysed sequencing data from six PCR amplicons amplified from WT animal biopsies with tailed primers flanking genomic intervals ranging from 0.9 to 2 kb in size (Experiment A, Supplemental Table ST1). PCR amplicons were barcoded, assembled in sequencing libraries and sequenced with a MinION.

Using the Porechop tool, reads that showed both barcoded ends were selected and demultiplexed for each barcode. We then filtered reads according to a range of read quality filters (q84 to q96) with Filtlong. Each group of reads of a given

quality threshold was aligned against the genome reference sequence using Minimap2 (Li, 2018). We then evaluated targeted sequencing performance by its ability to make a call for each base in the interval (coverage breadth) and the sequencing accuracy of these calls at the base level.

As expected, higher quality filters retained fewer reads (Supplemental Figure 2a), with the number of reads dropping sharply from q95. At the highest quality thresholds, although the reads were of superior quality, there were not enough of them to achieve complete coverage over the target interval (Supplemental Figure 2b). This is in contrast to the larger numbers of reads retained at less stringent quality filters, that did achieve complete coverage over the target interval. Intermediate read depths (100X) identify very high proportions of the genome reference

sequences. With a quality filter set between q86 and q94, when very high depth of reads (1,000X-10,000X) remained after filtering, the process was able to cover all bases within the sequenced segments.

We then analysed sequencing accuracy with thresholds for base-calling (proportion of calls at each base position that has to be reached to declare a consensus call for the base) ranging from 50% to 100%. Figure 2 shows the percentage of reference sequence accurately identified with consensus thresholds ranging from 50% to 100%, for each of the six loci. Figure 2a shows that selecting the highest quality read did not achieve the highest sequencing accuracy. Figure 2b focuses on the parameter ranges that ensure a very high proportion of accurate sequence identification. The graph shows that in our

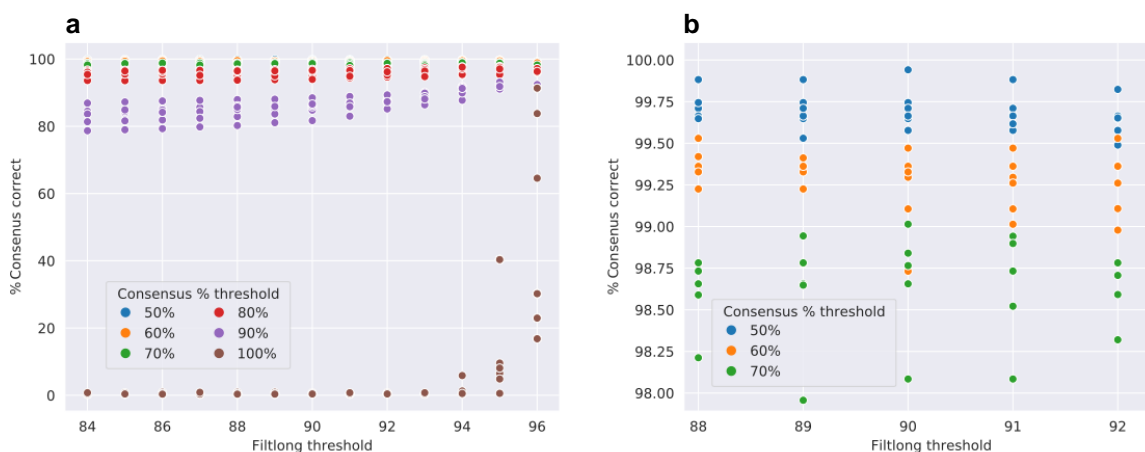


Figure 2: The graphs show the percentage of bases accurately sequenced by ONT sequencing for each WT segment analysed, for Filtlong read quality filters from q90 to q96 and with base calling threshold between 50% to 100% with increments of 10%. (a) Large numbers of calls achieved full coverage and highest accuracy for each base. (b) shows a zoomed in portion of (a). Very high percentages of sequence identification (>99.5%) are achieved with intermediate Filtlong quality threshold (q90) and lower threshold for consensus base calling (50%).

experimental conditions q90 Filtlong read quality threshold and a 50% consensus base calling threshold supports >99.5% sequence accuracy. We therefore selected these parameters for the remainder of the study.

ONT-based sequencing analysis of mutants generated with CRISPR/Cas9 and IssDNA donors

In the next run we analysed founder animals and their offspring

(Experiment B, Supplemental Table ST1) for two projects: a Cre KI into the *Mpeg1* gene, (Figure 3a) and a floxed *Cx3cl1* allele (Figure 3b). The sequence of donor IssDNAs and primers used in this article is shown in Supplemental Table ST2 and the animals produced in this study are summarised in Supplemental Table ST3. All of these animals had been previously identified as potentially bearing the desired allele change by Sanger sequencing (Supplemental Figures S3 and S4). PCR products

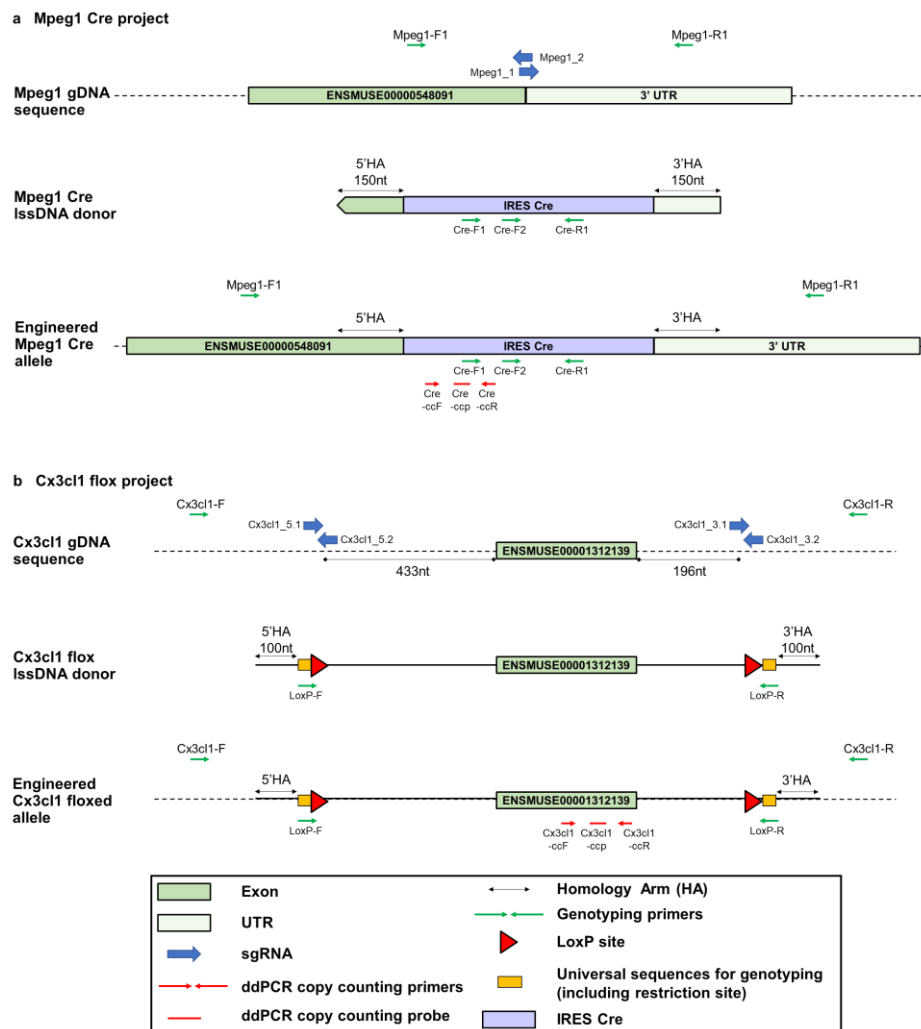


Figure 3: The figure details the design of the (a) *Mpeg1*-cre and the (b) *Cx3cl1*-flox (Codner et al., 2018) allele, respectively. The position of the primers used for analysis (sequences detailed in Supplemental Table 2) are shown, together with that of the ddPCR copy counting assays.

amplified from both G_0 and G_1 animals with external primers were sequenced with ONT sequencing. Q90 reads were aligned to the intended mutant sequence. Alignments were visualised with IGV (Thorvaldsdóttir et al., 2013; Figure 4). The sequence of the Mpeg1-cre allele was fully confirmed in both the positive founder Mpeg1-cre-80 and their offspring (Figure 4a).

Aiding the identification of mutant alleles

For floxed alleles, the WT and mutant sequences only differ by a small proportion of their overall

length (typically 120 bases out of 1.5 kb which is less than ONT sequencing raw read error frequency). Consequently, a high stringency parameter for alignment is not sufficient to prevent WT reads from aligning against the mutant reference along with mutant reads, potentially producing an ambiguous alignment file. To refine the analysis, we filtered reads for the presence of segments exclusive to the mutant sequence (determinant) prior to generating alignments. Typically two determinants for floxed alleles, or one determinant for cassette KIs can be used. This

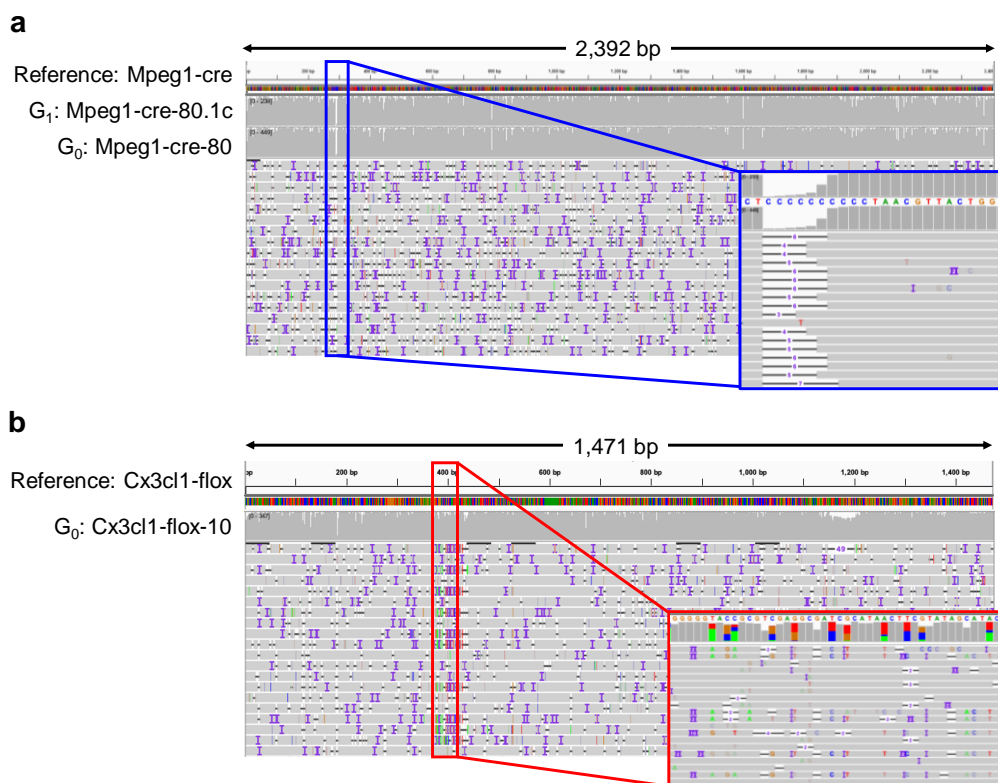


Figure 4: The figure summarises the outcome of ONT sequencing of animals for (a) the Mpeg1-cre and (b) Cx3cl1-flox project visualised with IGV. The alignment reflects the noisy nature of the method with errors distributed across the length of the sequenced segment. Note the complete alignment of reads to designed mutant sequences for both mosaic founders (G_0 s) and G_1 (grey histograms). Note that the different alleles in mosaic founder. Some reads align to the reference across the whole interval (red frame). The dip in sequence coverage coincides with at homopolymer repeat (Mpeg1-cre-80 and Mpeg1-cre-80.1c, blue frames).

yielded unequivocal sequence alignments and the correct mutated sequence was detected at the founder stage (Figure 4b).

Founder animals from four more floxed allele projects were tested in this run using PCR amplicons generated using tailed-primers external to the donor templates (generic design shown in Supplemental Figure S5, Project Prdm8-flox, Pam-flox, Hnf1a-flox and Inpp5k-flox; summary of samples in Experiment B, Supplementary Table ST1 and Sanger sequencing-based characterisation of mice in Supplemental Figures S6, S7, S8 and S9 respectively). ONT sequencing showed that the PCR amplicons amplified from founders Prdm8-flox-31, Pam-flox-3 and Hnf1a-flox-66 bear the correct sequences (Figure 5a and Supplemental Figure S10, respectively). Prdm8-flox-31 and Hnf1a-flox-66 matings had yet to produce G₁ animals but PCR amplicons amplified from Pam-flox-3 offspring were sequenced and confirmed as correct (Supplemental Figure S7).

Sanger sequencing of Inpp5k-flox allele founders showed the presence of loxP sites but did not lead to a conclusive outcome due to the mosaic nature of the template (Supplemental Figure S9). ONT sequencing of a PCR amplicon obtained from founder Inpp5k-flox-7 without preliminary filtering of

reads for the presence of both loxP revealed the genetic complexity of the animals (Supplemental Figure S11). Importantly, filtering for the presence of both loxP determinants resulted in no reads aligning to the designed mutant sequence. This absence of the fully conforming allele is in keeping with the result of Sanger sequencing the offspring of founder Inpp5k-flox-7 (Supplemental Figure S12).

Unwanted Single Nucleotide Polymorphisms (SNPs) can be visualised in ONT alignments

Having passed filtering with determinants, PCR amplicons amplified from Prdm8-flox-7 showed the mutant allele contained an unintended point mutation associated with the flox sequences (Figure 5a). The point mutation is in the synthetic interval flanking the 3' loxP and will not affect future use of this new allele. The SNP was also identified in Sanger sequencing of this individual (Supplemental Figure S6), and in the subsequent generation (data not shown).

In a third ONT sequencing run (Experiment C, Supplemental Table ST1), visualisation of the alignment file highlighted that mosaic founder 6430573F11Rik-flox-11 showed a G to A change at position 616 present at high representation (Figure 5b). This unintended point mutation was systematically associated *in cis* with alleles that contained loxP sites. These mutations are also seen in

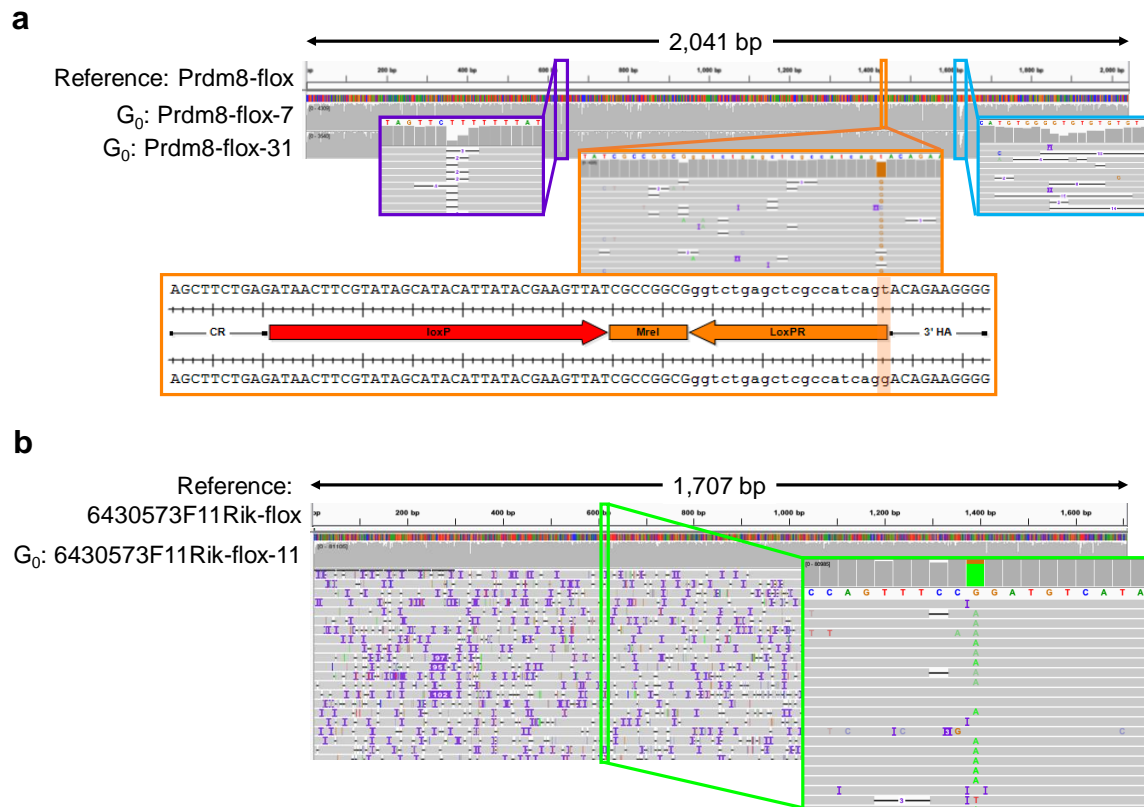


Figure 5: The figure shows alignments of sequencing reads obtained from founders Prdm8-flox-7 and Prdm8-flox-31 (a) and 6430573F11Rik-flox-11 (b) against the designed respective floxed sequence. ONT sequencing reveals an *in cis* point mutation, not present in the donor, associated with alleles that contained loxP sites in both Prdm8-flox-7 (a, orange frames) and 6430573F11Rik-flox-11 (b, green frames). The same sequences were confirmed with Sanger reads in these animals (Codner et al., 2018 and this article, Supplemental Figures S7 and S10). Note that dips in sequence coverage coincide with repeated sequences (Prdm8-flox-7 and Prdm8-flox-31, purple and blue frames, Figure 5a).

Sanger read data from this individual (Supplemental Figure S13).

sequence by standard sequencing, such as regions downstream of homopolymer repeats (Figure 4).

Discussion

Unequivocal identification of positive founders

Here we demonstrate that ONT, in spite of a higher per-base error rate, can be employed to efficiently identify correctly targeted alleles when screening mosaic G_0 s (Figures 4 and 5). This can be extended to G_1 to validate the transmitted allele and to confirm segments difficult to

Depth of sequencing offsets sequencing error rate

We have optimised our strategy for sequencing and data analysis workflow by sequencing targeted regions in WT animals. Importantly, confidence in sequencing data was achieved from the extensive depth of coverage, rather than through setting the most stringent quality filters for sequencing data (Figure

2). With sufficient sequencing read depth, reads were mapped to WT reference sequences concordantly across the entire genomic interval (Experiment A). The depth of coverage required depended on the complexity of the sequence with repeated sequences requiring deeper sequencing (Figures 4a and 5a). We then interrogated the sequencing data obtained from mutant animals for multiple projects for previously verified mutant alleles and found they could be unequivocally detected (Experiment B; Projects Mpeg1-cre, Cx3cl1-flox, Prdm8-flox and Pam-flox). In all cases where WT and mutant sequences differ by a proportion smaller than or close to the sequencing error rate, it was preferable to filter reads for the presence of determinants specific to mutant reference sequences prior to alignment. This prevents reads that correspond to WT alleles or partial integrations of mutant donors from being included in the alignments. This produced an unambiguous readout for the presence of correct mutant alleles. Homopolymer repeats were the exception to this conclusion as such segments remain a challenge for any sequencing techniques.

Exclusion of Sanger sequencing-based false positive animals

We applied ONT sequencing to projects where the presence of desired sequences had been shown

in some founder G_0 animals employing Sanger sequencing but only imperfectly mutated alleles were found transmitted to the subsequent generation (Projects 6430573F11Rik-flox, Figure 5b and Inpp5k-flox, Supplemental Figure S13). Crucially, application of ONT sequencing to these same founders showed that the desired mutation was systematically associated with additional base-pair changes or deletion *in cis*. Therefore, a long-read sequencing-based screening strategy can support the identification of undesired mutant alleles at an earlier stage of the mutagenesis process (G_0 screening). By contrast, with a screening strategy based on Sanger sequencing, the definitive sequence of each allele can only be ascertained in the simpler, non-mosaic, G_{1s} . Importantly, the ONT approach is also applicable when a pair of ssODN donors is employed instead of a lssDNA (Yang et al., 2013), allowing the identification of animals where the two sequences have been correctly integrated *in cis*.

Run capacity

We have used targeted sequencing of relatively short PCR amplicons (up to approximately 3 kb) permitting the generation of ultra-deep (1,000X to 10,000X) coverage datasets. This uses a fraction of the sequencing data production capacity of a MinION run. We used a kit that

supports twelve barcodes so that twelve animals can be sequenced in parallel analysing the same genomic segment. It would be possible to further multiplex samples by designing alternative primer pairs to amplify the same core region of interest but which include different flanking region lengths. Differentiation between individual animals for the same locus/project within a run can then be achieved using the genomic context flanking the region as an internal barcode. With the alternative format of 96 barcodes, a conservative set-up of 96 individuals (each under one barcode), if they were all mosaic, would require in the order of twenty gigabases sequenced to achieve 10,000X coverage of a 3 kb segment (96 animals for a given project containing up to eight genetic identities), which can be produced in a single MinION run.

We note that close to full identification of the target region was obtained with much lower coverage, offering the possibility of screening many more samples within a run with very high reliability. Finally, we have used an additional dimension for multiplexing, as animals corresponding to two different projects can be analysed in parallel under the same barcode. More samples could have been multiplexed using this strategy.

Large numbers of reads, rather than read quality, underpin the

accurate recognition of the desired mutant sequences. However, it is noteworthy that all sequencing runs we employed for the study were interrupted within twenty four hours, well before the standard forty eight hours recommended by the manufacturer, to reduce the amount of sequencing data excessive to our purpose being generated.

Currently, the main limiting factors of this process are the reliance on PCR, which may introduce sequence errors and the length of PCR product that can be amplified from genomic DNA extracted from a tissue biopsy. The recently proposed nanopore Cas9 Targeted-sequencing (nCATS) facilitates targeted sequencing by ONT without PCR and may increase the size of the genomic segments that can be surveyed for more extensive validation of animals (Gilpatrick et al., 2019).

A simple process

Although the approach appears a step change at first glance, the use of long-read sequencing turns out to involve a fairly simple and accessible process. It requires only minimal investment in sequencing equipment. The generation of sufficient sequencing data utilises a small fraction of the system capacity and the datasets that are produced do not represent a challenge for existing analysis tools. We further facilitate access to the process by packaging the analysis in a simple

workflow (see Methods). The timeline from genomic DNA extraction from a potential founder to a fully analysed dataset informing on the presence of the desired mutant allele fits within one week. ONT sequencing is a simple and efficient tool for screening genome-edited founders obtained with lssDNA donors. This is in contrast to traditional Sanger sequencing methods which rely on the amplification of multiple PCR products, which must then be individually sequenced and assembled into contigs. Assembly of these contigs is liable to mis-associate *trans* reads in mosaics animals, resulting in false positive calls.

However, the extensive characterisation of the target loci that ONT sequencing supports, in both founder and subsequent generations, does not suffice to validate newly mutated lines. Indeed, checks of off-target events (Anderson et al., 2018; Iyer et al., 2018), in particular, those physically linked to the locus of interest) and copy counting of the donor sequence by ddPCR to eliminate additional integrations remain essential and complementary steps to fully validate G₁ animals (Supplemental Figures S3 and S4 and Codner et al., 2018).

Finally, this simple sequencing process is also applicable to any other circumstance where a sequence of a specific locus must be

validated, for example in cultured cells following gene targeting by homologous recombination.

A more accurate screening tool for more ethical animal management

Here we have illustrated how long-read sequencing can be employed to exclude founders previously misidentified as positive using Sanger sequencing-based methods, because they only contained targeted mutations that were associated with unwanted base-pair changes or sequence rearrangements. The method also allows for the analysis of mosaic animals where the genetic make-up is too complex to be easily disentangled by standard Sanger sequencing. This constitutes a refinement in terms of the use of animals for the generation of targeted mutations, as it reduces the number of false positive founders carried forward for breeding. It also serves to shorten the timeline of mutagenesis projects, as founder characterisation is greatly facilitated and no animals nor time are unnecessarily used testing misidentified positive founders for germline transmission of the desired mutant allele. This is particularly useful as founders can present a broad range of welfare issues, as part or all of their body may contain mutations that may affect both alleles of modified loci.

Conclusion

CRISPR/Cas9 with lssDNA donors delivered into one-cell embryos generates complex mosaic founders that are challenging to analyse by classical Sanger sequencing. We showed that targeted sequencing with ONT technology is a simple and powerful method to faithfully identify the animals that bear a correct integration on target. This represents progress in ethical animal use, as it prevents breeding of false-positive founders.

Materials and methods

Sequences of reagents

The sequences of the sgRNAs, templates for lssDNA generation, primers and probes are shown in Supplemental Table ST2.

sgRNAs

Guide sequence selection was carried out using the following online tools: CRISPOR (Haeussler et al., 2016) and WTSI Genome Editing (WGE) (Hodgkins et al., 2015). sgRNA sequences were selected with as few predicted off-target events as possible, particularly on the same chromosome as the intended modification. sgRNAs used in this study are shown in Table ST2. sgRNAs were synthesised directly from gBlock[®] (IDT) templates containing the T7 promoter using the HiScribe[™] T7 high yield RNA synthesis kit (New England BioLabs[®]) following manufacturer's instructions. RNAs were purified

using the MEGAclean kit (Ambion). RNA quality was assessed using a NanoDrop (ThermoScientific) and by electrophoresis on 2% agarose gel containing ethidium bromide (Fisher Scientific).

Templates for lssDNA synthesis

Templates for lssDNA synthesis were either assembled by cloning in a plasmid or, when possible, were obtained from IDT as a single gBlock[®].

Donor templates

Donor lssDNAs were generated following a method adapted from Miura et al., 2015. Briefly, templates for *in vitro* transcription (donor sequence flanked by the T7 promoter) were obtained as a gBlock[®] (IDT) or cloned in a plasmid that was subsequently linearised. Typically, 150 ng of double stranded gBlock[®] template or 2 µg of plasmid template was transcribed using the HiScribe T7 High Yield RNA Synthesis Kit (New England BioLabs[®]). At the end of the reaction, DNase I was added to remove the DNA template. RNA was purified employing the MEGAclean Transcription Clean-Up kit (Ambion). Single-stranded DNA was synthesised by reverse transcription from 20 µg of RNA template employing SuperScript III Reverse Transcriptase (Invitrogen), treated with RNase H (Ambion) and purified employing the QIAquick Gel Extraction Kit (Qiagen) or, for higher

yields, employing the RNA Clean & Concentrator™ kit (Zymogen). Alternatively, lssDNAs were synthesised with the Guide-it™ Long ssDNA Strandase Kit according to the manufacturer instruction. Donor concentration was quantified using a NanoDrop (Thermo Scientific) and integrity was checked on 1.5% agarose gel containing ethidium bromide (Fisher Scientific).

Mixes for microinjection

Microinjection buffer (10 mM Tris-HCl, 0.1 mM EDTA, 100 mM NaCl, pH7.5) was prepared and filtered through a 2 nm filter and autoclaved. Mixes containing 100 ng/μl Cas9 mRNA (5meC,Ψ) (TriLink BioTechnologies), 50 ng/μl sgRNAs and 50 ng/μl ssODN or 50 ng/μl lssDNA were prepared in microinjection buffer, filtered through Costar® SpinX® Centrifuge Tube Filters (Corning) and stored at -80°C until microinjection.

Mice

All animals were housed and maintained in the Mary Lyon Centre, MRC Harwell Institute under specific pathogen-free (SPF) conditions, in individually ventilated cages adhering to environmental conditions as outlined in the Home Office Code of Practice. Mice were euthanised by Home Office Schedule 1 methods. Animals used for transgenesis projects are detailed in Supplemental Table ST3. Colonies established during the course of this

study are available for distribution and are detailed in Supplemental Table ST4.

Pronuclear microinjection of zygotes

All embryos were obtained by superovulation. Pronuclear microinjection was performed as per Gardiner and Teboul, 2009, employing a FemtoJet (Eppendorf) and C57BL/6NTac embryos. Specifically, injection pressure (Pi) was set between 100 and 700 hPa, depending on needle opening; injection time (Ti) was set at 0.5 seconds and the compensation pressure (PC) was set at 10 hPa. Mixes were centrifuged at high speed for one minute prior to microinjection. Injected embryos were re-implanted in CD-1 pseudo-pregnant females. Host females were allowed to litter and rear G₀s.

Breeding for germline transmission

G₀ animals where the presence of a desired allele was detected were mated to WT isogenic animals to obtain G₁ animals to assess the germline transmission of the allele of interest and permit the definitive validation of its integrity.

Genomic DNA extraction from ear biopsies

Genomic DNA from G₀ and G₁ animals was extracted from ear clip biopsies using the DNA Extract All Reagents Kit (Applied Biosystems)

according to manufacturer's instructions. The crude lysate was stored at -20°C.

PCR amplification and Sanger sequencing

New primer pairs were set up in a PCR reaction containing 500 ng genomic DNA extracted from a wild type mouse, 1 x Expand Long Range Buffer with 12.5 mM MgCl₂ (Roche), 500 μM PCR Nucleotide Mix (dATP, dCTP, dGTP, dTTP at 10 mM, Roche), 0.3 μM of each primer, 3% DMSO, and 1.8 U Expand Long Range Enzyme mix (Roche) in a total volume of 25 μl. Using a T100 thermocycler (Bio-Rad), PCRs were subject to the following thermal conditions; 92°C for 2 minutes followed by 40 cycles of 92°C for 10 seconds, a gradient of annealing temperatures between 55-65°C for 15 seconds and 68°C for 1 minute/kb and a final elongation step for 10 minutes at 68°C. PCR outcome was analysed on a 1.5 to 2% agarose gel, depending on the amplicon size and the highest efficient annealing temperature was identified for the primer pair. If no temperature allowed for an efficient and/or specific PCR amplification the assay was repeated with an increased DMSO concentration (up to 12%). Using optimised conditions, as defined above, PCRs for each project were run and an aliquot analysed on agarose gel. PCR products were purified employing QIAquick Gel Extraction

Kit (Qiagen) and sent for Sanger sequencing (Source Bioscience, Oxford). Genotyping primers were chosen at least at 200 bp away from the extremity of donors, depending on available sequences for design.

Analysis of Sanger sequencing data

Sequencing data were analysed differently depending on whether they were obtained from G₀s or G₁s (as per Gardiner and Teboul, 2009). At the G₀ stage, animals were screened for evidence of the expected change i.e. the presence of loxP sites for conditional allele projects or presence of the cre knock-in sequence for Mpeg-cre allele. G₀ animals should be considered mosaic animals. All G₁ animals are heterozygous containing one WT allele and one allele to be determined as they are obtained from mating G₀ animals with desired gene edits to WT animals. The G₁ stage enables definitive characterisation of the new mutant.

Preparation of libraries for ONT sequencing

DNA LoBind tubes (Eppendorf) were used. PCR was performed with tailed-end primers using the same conditions as for amplicons produced for Sanger sequencing, to generate amplicons for ONT sequencing. PCR amplicons were barcoded employing LongAmp Taq (New England BioLabs®). The ends

of pooled DNA fragments were repaired employing the NEBNext End repair / dA-tailing Module (New England BioLabs®). Sequencing adaptors were added using the 1D-Ligation Sequencing Kit (Oxford ONT technology). All reactions were performed according to the manufacturer's instruction. DNA was purified at all steps using AMPure XP beads (Agencourt) employing a 0.8X to 1X beads to sample ratio. DNA was quantified with a Qubit fluorometer at all steps. Sequencing libraries were loaded on primed SpotON Flow Cell (R9.4) (ONT). Runs were performed employing the MinKNOW GUI at defaults settings for up to 24 hours (ONT).

Analysis of ONT sequencing data

A nextflow (Di Tommaso et al., 2017) workflow for the bioinformatics processes was assembled and is available on gitlab (<https://gitlab.com/nick297/cas9point4>). In brief, reads were demultiplexed with Porechop (Wick et al., 2017), requiring the recognition of two barcodes (both extremities of the PCR amplicon sequenced). Reads were then filtered for a minimum quality (q score) using Filtlong (<https://github.com/rrwick/Filtlong>). Filtered reads were aligned against the relevant reference sequence using minimap2 (Li, 2018) and filtered using samtools for a mapping quality score of q90 (Li et

al., 2009). Alignments were then visualised using IGV (<http://software.broadinstitute.org/software/igv/>; Thorvaldsdóttir et al., 2013). After alignment, corresponding reads were filtered for the presence of mutant specific sequence determinants using BLASTn (Altschul et al., 1990) to simplify the readout of the experiment.

Copy counting of the donor by ddPCR

Copy Number Variation experiments were performed as duplex reactions. A FAM-labelled assay was used to amplify a region contained within the ssDNA donor (sourced from Biosearch Technologies), in parallel with a VIC-labelled reference gene assay (*Dot1l*, sourced from ThermoFisher) set at 2 copies (CNV2) on the Bio-Rad QX200 ddPCR system (Bio-Rad, CA) as per Codner and colleagues (Codner et al., 2016). Reaction mixes (22 µl) contained 2 µl crude DNA lysate or 50 ng of phenol/chloroform purified genomic DNA, 1x ddPCR Supermix for probes (Bio-Rad, CA, USA), 225 nM of each primer (two primers per assay) and 50 nM of each probe (one VIC-labelled probe for the reference gene assay and one FAM-labelled for the ssODN sequence assay). These reaction mixes were loaded either into DG8 cartridges together with 70 µl droplet oil per sample and droplets generated using the QX100 Droplet Generator

or loaded in plate format into the Bio-Rad QX200 AutoDG and droplets generated as per the manufacturer's instructions. Post droplet generation, the oil/reagent emulsion was transferred to a 96 well semi-skirted plate (Eppendorf AG, Hamburg, Germany) and the samples were amplified on the Bio-Rad C1000 Touch thermocycler (95°C for 10 min, followed by 40 cycles of 94°C for 30 s and 58°C for 60 s, with a final elongation step of 98 °C for 10 min, all temperature ramping set to 2.5°C/second). The plate containing the droplet amplicons was subsequently loaded into the QX200 Droplet Reader (Bio-Rad, CA, USA). Standard reagents and consumables supplied by Bio-Rad were used, including cartridges and gaskets, droplet generation oil and droplet reader oil. Copy number was assessed using the Quantasoft software using at least 10,000 accepted droplets per sample. Copy numbers were calculated by applying Poisson statistics to the fraction of end-point positive reactions and the 95% confidence interval of this measurement is shown.

Abbreviations: bp, base-pair; Cas9, CRISPR associated protein 9; CRISPR, clustered regularly interspaced short palindromic repeat; ddPCR, droplet digital PCR; DNA, deoxyribonucleic acid; HA, homology arm; kb, kilobases; KI, knock-in; lssDNA, long single-

stranded DNA; ONT, Oxford Nanopore Technologies; sgRNA, single guide RNA; SNP, single nucleotide polymorphism; ssODN, single-stranded oligo-deoxynucleotide; WT, wild type.

Authors' contributions:

CM, GC, AA, AC, JL, MM, EM, JM, FP and LT designed protocols, sgRNA and oligonucleotides, synthesised reagents and genotyped progeny mice. CM performed MinION sequencing. CM, NDS and LT analysed long reads sequencing data. MH, MES and SW undertook mouse colony management. GC, HG, NDS and LT conceived the study. CM, GC, EM, MES and LT analysed the data. All authors contributed to the writing of the manuscript. All authors read and approved the final manuscript.

Acknowledgements:

The authors would like to thank the staff of the Mary Lyon Centre for providing excellent animal husbandry and microinjection services, Michael Micorescu (Oxford Nanopore Technologies, New York) and Duncan Sneddon for expert support with data mining, Dr Rosie Bunton-Stasyshyn for helpful discussions and Dr Louise Tinsley for expert assistance with the preparation of this manuscript.

Ethics approval and consent to participate:

All animal studies were licensed by the Home Office under the Animals (Scientific Procedures) Act 1986 Amendment Regulations 2012 (SI 4 2012/3039), UK, and additionally approved by the Institutional Ethical Review Committee.

Availability of data and materials: ONT sequencing data deposited at the European Nucleotide Archive (<https://www.ebi.ac.uk/ena>). Identifiers are shown in Supplemental Table 1. All other data and their analyses are included in

this published article and its supplementary information files.

Competing interests:

LT was the recipient of a travel award from ONT.

Funding:

This work was supported by a Medical Research Council IMPC Strategic Award (53658 to SW and LT), the National Institute for Health (Supplement to Grant U42OD011174 to SW) and the Science and Technology Facilities Council (Proof of Concept Grant to LT and ONT).

Supplemental Tables

Supplemental Table ST1: ONT sequencing experiments.

The table summarises the barcode, project, animal employed in each Nanopore sequencing experiment and the corresponding references to access datasets in the ENA repository.

Supplemental Table ST2: The table details the sequences of sgRNAs, for lssDNA templates, primers and probes employed in this study.

Supplemental Table ST3: The table details the mice that were obtained from each microinjection session during the course of this study.

Supplemental Table ST4: The table details the mouse colonies that were established during the course of this study.

Supplemental Figures

Supplemental Figure S1: Experimental plan: Genome-edited loci can be characterised by Sanger sequencing (process highlighted in orange boxes), producing partial reads that must be assembled to reconstitute the whole region of interest. Alternatively, ONT sequencing (process highlighted in blue boxes) produces longer sequence reads spanning the whole region of interest.

Supplemental Figure S2: Required sequencing coverage for full identification of a known sequence. (a) shows the number of reads retained for by filters for each experiment. Note that the number of reads retained drops sharply for quality greater than q90. (b) shows the percentage of the sequenced interval identified by ONT sequencing for each WT segment analysed in Experiment A. Note that the large numbers of reads in lower quality range identify more reliably the reference sequences than fewer higher quality reads.

Supplemental Figure S3: Details of the analysis of the microinjection session containing animals interrogated by ONT sequencing for the Mpeg1 Cre project. The figure shows the PCR amplification of the genomic region of interest with (a) Mpeg1-F1 and Mpeg1-R1 (WT yields 873 bp amplicon, Cre KI yields 2412 bp amplicon) and (b) CreF and CreR primers (Cre KI yields 472 bp amplicon) from biopsies taken from the G₀ animals. Animals yielding amplicons with CreF and CreR primers were subject to PCR to assess whether Cre is on target with primer combinations (c) Mpeg1-F1 and CreR (Cre KI yields 1440 bp amplicon) and (d) CreF and Mpeg1-R1

(Cre KI yields 1182 bp amplicon). (e) The panels show the sequencing of PCR amplicons using Mpeg1_F1/CreR and CreF/Mpeg1_R1 respectively obtained from animal Mpeg1-80. (f) The table details the analysis of G₀ animals analysed: Animal ID, outcome of PCR analysis of the region of interest and the overall conclusion for each individual are shown. Panel of PCR amplicons with four different primer combinations (Mpeg1-F1 and Mpeg1-R1; CreF and CreR; Mpeg1-F1 and CreR; CreF and Mpeg1-R1) obtained for G₁ animals derived from founder Mpeg1-75 crossed to WT (g) and founder Mpeg1-80 mated with WT (h). The table (i) details the ID, outcome of sequencing the region of interest, copy counting of the region of interest and the conclusion for each G₁ individual. Sanger sequencing traces obtained from PCR amplification using Mpeg1_F1/CreR and CreF/Mpeg1_R1 respectively obtained from animal Mpeg1-75.1d. + is positive control amplified from an unrelated (a) WT, (b) Cre-KI animal. L1 = 1 kb DNA molecular weight ladder (thick band is 3 kb). Animal(s) interrogated by ONT sequence analysis are highlighted in green. * denotes animals with evidence of Cre KI on target but yielding multiple possible KI-related bands.

Supplemental Figure S4: Details of the analysis of the microinjection session and F1 colony containing animals interrogated by ONT sequencing for the Cx3cl1-flox project. The figure shows the PCR amplification of the genomic region of interest with (a) Cx3cl1-F1 and Cx3cl1-R1 primers (WT yields 1488 bp amplicon, floxed allele yields 1483 bp amplicon) and (b) LoxPF and LoxPR primers (floxed allele yields 835 bp amplicon) from biopsies taken from the G₀ animals. (c) The panels show the sequencing of PCR amplicon obtained from animal Cx3cl1-10 with Cx3cl1-F1 and Cx3cl1-R1. LoxP site sequences are highlighted in blue. (d) The table details the G₀ animals obtained from the microinjection. The ID and outcome of PCR analysis of the region of interest, as well as the conclusion for each founder are shown. Founder Cx3cl1-10 was mated for floxed allele transmission (LoxP PCR positive and sequence of complex mosaic). PCR amplification of region of interest with (e) Cx3cl1-F1 and Cx3cl1-R1 primers (1483 bp amplicon) and LoxPF and LoxPR primers (835 bp amplicon) from biopsies taken from Cx3cl1-10's offspring. (f) The table details the first litter obtained by mating Cx3cl1-10 with a WT mouse. The ID, outcome of sequencing the region of interest and copy counting of the region of interest as well as the conclusion for each individual are shown. (g) Sanger sequence traces of the Cx3cl1 PCR product from G₁ animal Cx3cl1-10.1c illustrating insertion of each LoxP site and associated genotyping handles (primer sequence and restriction

enzyme site) highlighted in blue on target. + is positive control amplified from an unrelated (a) WT, (b) floxed animal. L1 = 1 kb DNA molecular weight ladder (thick band is 3 kb). Animal(s) interrogated by ONT sequence analysis are highlighted in green.

Supplemental Figure S5: General design strategy and generation of lssDNA donors for the generation of floxed alleles and cre KIs.

Supplemental Figure S6: Details of the analysis of the microinjection session containing animals interrogated by ONT sequencing for the Prdm8-flox project. The figure shows the PCR amplification of the genomic region of interest with (a) Prdm8-F1 and Prdm8-R1 primers (WT yields 1984 bp amplicon, floxed yields 2054 bp amplicon) and (b) LoxPF and LoxPR primers (floxed yields 1025 bp amplicon) from biopsies taken from the G₀ animals. (c) The panels show the sequencing of PCR amplicon obtained from animal Prdm8-7 with Prdm8-F1 and LoxPF. LoxP site sequences are highlighted in blue. The SNP in the designed mutant sequence at 5' end of the LoxPR primer sequence is highlighted in red. (d) The table details the G₀ animals obtained from the microinjection analysed by ONT. The ID and outcome of PCR analysis of the region of interest, as well as the conclusion for each individual are shown. Animal(s) interrogated by ONT sequence analysis are highlighted in green. + is positive control amplified from an unrelated (a) WT, (b) floxed animal. L1 = 1 kb DNA molecular weight ladder (thick band is 3 kb), L2 = 100 bp DNA molecular weight ladder (thick bands are 1 kb and 500 bp).

Supplemental Figure S7: Details of the analysis of the microinjection session containing animals interrogated by ONT sequencing for the Pam-flox project. The figure shows the PCR amplification of the genomic region of interest with (a) Pam-F1 and Pam-R1 primers (WT yields 1426 bp amplicon, floxed yields 1431 b amplicon) and (b) LoxPF and LoxPR primers (floxed yields 801 bp amplicon) from biopsies taken from the G₀ animals. (c) The panels show the sequencing of PCR amplicon obtained from animal Pam-3 with Pam-F1 and Pam-R1. LoxP site sequences are highlighted in blue. (d) The table details the G₀ animals analysed: ID, outcome of PCR analysis of the region of interest and conclusion for each individual are shown. PCR amplification of region of interest with (e) Pam-F1 and Pam-R1 primers (1431 bp amplicon) and (f) LoxPF and LoxPR primers (801 bp amplicon) from biopsies taken from Pam-3's offspring. (g) The table details the first litter obtained by mating Pam-flox-3 with a WT mouse. The ID, outcome of PCR amplification of the regions of

interest as well as the initial conclusion for each individual are shown. NB. No Sanger sequencing was performed on the Pam floxed G₁ generation prior to analysis with ONT. Animal(s) interrogated by ONT sequence analysis are highlighted in green. + is positive control amplified from an unrelated (a) WT, (b) floxed animal. L1 = 1 kb DNA molecular weight ladder (thick band is 3 kb).

Supplemental Figure S8: Details of the analysis of the microinjection session containing animals interrogated by ONT sequencing for the Hnf1a-flox project. The figure shows the PCR amplification of the genomic region of interest with (a) Hnf1a-F1 and Hnf1a-R1 primers (WT yields 1221 bp amplicon, floxed allele yields 1191 bp amplicon) and (b) LoxPF and LoxPR primers (floxed allele yields 691 bp amplicon) from biopsies taken from the G₀ animals. (c) The panels show the sequencing of PCR amplicon obtained from animal Hnf1a-66 with Hnf1a-F1 and Hnf1a-R1 and sequenced with LoxPF and LoxPR. LoxP site sequences are highlighted in blue. (d) The table details the G₀ animals obtained; the ID and outcome of PCR analysis of the region of interest, as well as the conclusion for each individual are shown. Animal(s) interrogated by ONT sequence analysis are highlighted in green. + is positive control amplified from an unrelated (a) WT, (b) floxed animal. L1 = 1 kb DNA molecular weight ladder (thick band is 3 kb), L2 = 100 bp DNA molecular weight ladder (thick bands are 1 kb and 500 bp).

Supplemental Figure S9: Details of the analysis of the microinjection sessions containing animals interrogated by ONT sequencing for the Inpp5k-flox project. The figure shows the PCR amplification of the genomic region of interest with (a) Inpp5k-F1 and Inpp5k-R1 primers (WT yields 1701 bp amplicon, floxed allele yields 1705 bp amplicon) and (b) LoxPF and LoxPR primers (floxed allele yields 1194 bp amplicon) from biopsies taken from the G₀ animals. (c) The panels show the sequencing of PCR amplicon obtained from animal Inpp5k-7 with Inpp5k-F1 and LoxPR, and with LoxPF and Inpp5k-R1 respectively. LoxP site sequences are highlighted in blue. (d) The table details the G₀ animals analysed: the ID, outcome of PCR analysis of the region of interest and the conclusion for each individual are shown. Animal(s) interrogated by ONT sequence analysis are highlighted in green. + is positive control amplified from an unrelated (a) WT, (b) floxed animal. L1 = 1 kb DNA molecular weight ladder (thick band is 3 kb).

Supplemental Figure S10: Confirmation of potential animals for the Pam-flox and Hnf1a-floxed projects. (a) shows the outcome of ONT sequencing of the founder animal Pam-flox-3 aligned against the mutant Pam-flox reference visualised with IGV. (b) shows the outcome of ONT sequencing of the founder animal Hnf1a-flox-66 aligned against the mutant Hnf1a-flox reference visualised with IGV. The alignments reflect the noisy nature of the method with errors distributed across the length of the sequenced segment. Note the complete alignment of reads to designed mutant sequence (grey histograms).

Supplemental Figure S11: Alignment of sequencing reads from G₀ Inpp5k-7 and -33 against the Inpp5k-flox reference without filtering for determinants. The yellow frame highlights the presence of segments that are different to the reference in G₀ Inpp5k-7. The blue and red frames highlight that although G₀ Inpp5k-36 contains sequences that are overall similar to the mutant sequence reference, these alleles also contain point mutations.

Supplemental Figure S12: Details of the analysis of the G₁ generation containing animals interrogated by ONT sequencing for the Inpp5k-flox project. The figure shows the PCR amplification of the genomic region of interest with (a) Inpp5k-F1 and Inpp5k-R1 primers (WT yields 1701 bp amplicon, floxed allele yields 1705 bp amplicon) and (b) LoxPF and LoxPR primers (floxed allele yields 1194 bp amplicon) from biopsies taken from the G₁ animals derived from crossing founder animals Inpp5k-7 and Inpp5k-8 to WT. (c) The table details the G₁ animals obtained from the two lines. The ID and outcome of PCR analysis of the region of interest, as well as the conclusion for each individual are shown. (d) The panels show the sequencing of PCR amplicon obtained from animal Inpp5k-7.1b with Inpp5k-F1 and LoxPR, and with LoxPF and Inpp5k-R1 respectively. Deviations from the intended mutant sequence are highlighted in blue. Animal(s) interrogated by ONT sequence analysis are highlighted in green. + is positive control amplified from an unrelated (a) WT, (b) floxed animal. L1 = 1 kb DNA molecular weight ladder (thick band is 3 kb).

Supplemental Figure S13: Details of the analysis microinjection session containing animals interrogated by ONT sequencing for the 6430573F11Rik project. The figure shows the PCR amplification of the genomic region of interest with (a) 6430573F11Rik-F3 and 6430573F11Rik-R2 primers (WT yields 1724 bp amplicon, floxed yields 1721 bp amplicon) and (b) LoxPF and LoxPR primers (floxed yields 999 bp

amplicon) from biopsies taken from the G0 animals. (c) The panels show the sequencing of PCR amplicon obtained from animal 6430573F11Rik-11 with 6430573F11Rik-F2 and 6430573F11Rik-R3. LoxP site sequences are highlighted in blue. The SNP in the critical region is also highlighted in blue (d) The table details the G0 animals analysed: The ID, outcome of PCR analysis of the region of interest and the conclusion for each individual are shown. Animal(s) interrogated by ONT sequence analysis are highlighted in green. + is positive control amplified from an unrelated (a) WT, (b) floxed animal. L1 = 1 kb DNA molecular weight ladder (thick band is 3 kb).

References:

- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. *Journal of Molecular Biology* 215, 403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2)
- Anderson, K.R., Haeussler, M., Watanabe, C., Janakiraman, V., Lund, J., Modrusan, Z., Stinson, J., Bei, Q., Buechler, A., Yu, C., Thamminana, S.R., Tam, L., Sowick, M.-A., Alcantar, T., O’Neil, N., Li, J., Ta, L., Lima, L., Roose-Girma, M., Rairdan, X., Durinck, S., Warming, S., 2018. CRISPR off-target analysis in genetically engineered rats and mice. *Nat Methods* 15, 512–514. <https://doi.org/10.1038/s41592-018-0011-5>
- Birling, M.-C., Schaeffer, L., André, P., Lindner, L., Maréchal, D., Ayadi, A., Sorg, T., Pavlovic, G., Héroult, Y., 2017. Efficient and rapid generation of large genomic variants in rats and mice using CRISMERE. *Sci Rep* 7, 43331–43331. <https://doi.org/10.1038/srep43331>
- Codner, G.F., Lindner, L., Caulder, A., Wattenhofer-Donzé, M., Radage, A., Mertz, A., Eisenmann, B., Mianné, J., Evans, E.P., Beechey, C.V., Fray, M.D., Birling, M.-C., Héroult, Y., Pavlovic, G., Teboul, L., 2016. Aneuploidy screening of embryonic stem cell clones by metaphase karyotyping and droplet digital polymerase chain reaction. *BMC Cell Biol* 17, 30–30. <https://doi.org/10.1186/s12860-016-0108-6>
- Codner, G.F., Mianné, J., Caulder, A., Loeffler, J., Fell, R., King, R., Allan, A.J., Mackenzie, M., Pike, F.J., McCabe, C.V., Christou, S., Joynson, S., Hutchison, M., Stewart, M.E., Kumar, S., Simon, M.M., Agius, L., Anstee, Q.M., Volynski, K.E., Kullmann, D.M., Wells, S., Teboul, L., 2018. Application of long single-stranded DNA donors in genome editing: generation and validation of mouse mutants. *BMC Biology* 16, 70. <https://doi.org/10.1186/s12915-018-0530-7>
- Di Tommaso, P., Chatzou, M., Floden, E.W., Barja, P.P., Palumbo, E., Notredame, C., 2017. Nextflow enables reproducible computational workflows. *Nature Biotechnology* 35, 316–319. <https://doi.org/10.1038/nbt.3820>
- Gardiner, W.J.K., Teboul, L., 2009. Overexpression Transgenesis in Mouse: Pronuclear Injection, in: Cartwright, E.J. (Ed.), *Transgenesis Techniques: Principles and Protocols*. Humana Press, Totowa, NJ, pp. 111–126. https://doi.org/10.1007/978-1-60327-019-9_8
- Gilpatrick, T., Lee, I., Graham, J.E., Raimondeau, E., Bowen, R., Heron, A., Sedlazeck, F.J., Timp, W., 2019. Targeted Nanopore Sequencing

- with Cas9 for studies of methylation, structural variants and mutations. *bioRxiv* 604173. <https://doi.org/10.1101/604173>
- Haeussler, M., Schönig, K., Eckert, H., Eschstruth, A., Mianné, J., Renaud, J.-B., Schneider-Maunoury, S., Shkumatava, A., Teboul, L., Kent, J., Joly, J.-S., Concordet, J.-P., 2016. Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. *Genome Biol* 17, 148–148. <https://doi.org/10.1186/s13059-016-1012-2>
- Hodgkins, A., Farne, A., Perera, S., Grego, T., Parry-Smith, D.J., Skarnes, W.C., Iyer, V., 2015. WGE: a CRISPR database for genome engineering. *Bioinformatics* 31, 3078–3080. <https://doi.org/10.1093/bioinformatics/btv308>
- Iyer, V., Boroviak, K., Thomas, M., Doe, B., Riva, L., Ryder, E., Adams, D.J., 2018. No unexpected CRISPR-Cas9 off-target activity revealed by trio sequencing of gene-edited mice. *PLoS Genet* 14, e1007503–e1007503. <https://doi.org/10.1371/journal.pgen.1007503>
- Jain, M., Olsen, H.E., Paten, B., Akeson, M., 2016. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biol* 17, 239–239. <https://doi.org/10.1186/s13059-016-1103-0>
- Kosicki, M., Tomberg, K., Bradley, A., 2018. Repair of double-strand breaks induced by CRISPR-Cas9 leads to large deletions and complex rearrangements. *Nature Biotechnology* 36, 765.
- Lanza, D.G., Gaspero, A., Lorenzo, I., Liao, L., Zheng, P., Wang, Y., Deng, Y., Cheng, C., Zhang, C., Seavitt, J.R., DeMayo, F.J., Xu, J., Dickinson, M.E., Beaudet, A.L., Heaney, J.D., 2018. Comparative analysis of single-stranded DNA donors to generate conditional null mouse alleles. *BMC Biol* 16, 69–69. <https://doi.org/10.1186/s12915-018-0529-0>
- Li, H., 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34, 3094–3100. <https://doi.org/10.1093/bioinformatics/bty191>
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 1000 Genome Project Data Processing Subgroup, 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
- Mianné, J., Chessum, L., Kumar, S., Aguilar, C., Codner, G., Hutchison, M., Parker, A., Mallon, A.-M., Wells, S., Simon, M.M., Teboul, L., Brown, S.D.M., Bowl, M.R., 2016. Correction of the auditory phenotype in C57BL/6N mice via CRISPR/Cas9-mediated homology

- directed repair. *Genome Med* 8, 16–16.
<https://doi.org/10.1186/s13073-016-0273-4>
- Mianné, J., Codner, G.F., Caulder, A., Fell, R., Hutchison, M., King, R., Stewart, M.E., Wells, S., Teboul, L., 2017a. Analysing the outcome of CRISPR-aided genome editing in embryos: Screening, genotyping and quality control. *Methods* 121–122, 68–76.
<https://doi.org/10.1016/j.ymeth.2017.03.016>
- Miura, H., Gurumurthy, C.B., Sato, T., Sato, M., Ohtsuka, M., 2015. CRISPR/Cas9-based generation of knockdown mice by intronic insertion of artificial microRNA using longer single-stranded DNA. *Scientific reports* 5, 12799. <https://doi.org/10.1038/srep12799>
- 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., Teboul, L., de Bruijn, M.F.T.R., 2019. Microhomologies are prevalent at Cas9-induced larger deletions. *Nucleic Acids Res* 47, 7402–7417. <https://doi.org/10.1093/nar/gkz459>
- Quadros, R.M., Miura, H., Harms, D.W., Akatsuka, H., Sato, T., Aida, T., Redder, R., Richardson, G.P., Inagaki, Y., Sakai, D., Buckley, S.M., Seshacharyulu, P., Batra, S.K., Behlke, M.A., Zeiner, S.A., Jacobi, A.M., Izu, Y., Thoreson, W.B., Urness, L.D., Mansour, S.L., Ohtsuka, M., Gurumurthy, C.B., 2017. Easi-CRISPR: a robust method for one-step generation of mice carrying conditional and insertion alleles using long ssDNA donors and CRISPR ribonucleoproteins. *Genome Biol* 18, 92–92. <https://doi.org/10.1186/s13059-017-1220-4>
- Renaud, J.-B., Boix, C., Charpentier, M., De Cian, A., Cochenne, J., Duvernois-Berthet, E., Perrouault, L., Tesson, L., Edouard, J., Thinard, R., Cherifi, Y., Menoret, S., Fontanière, S., de Crozé, N., Fraichard, A., Sohm, F., Anegon, I., Concordet, J.-P., Giovannangeli, C., 2016. Improved Genome Editing Efficiency and Flexibility Using Modified Oligonucleotides with TALEN and CRISPR-Cas9 Nucleases. *Cell Reports* 14, 2263–2272.
<https://doi.org/10.1016/j.celrep.2016.02.018>
- Shin, H.Y., Wang, C., Lee, H.K., Yoo, K.H., Zeng, X., Kuhns, T., Yang, C.M., Mohr, T., Liu, C., Hennighausen, L., 2017. CRISPR/Cas9 targeting events cause complex deletions and insertions at 17 sites in the mouse genome. *Nat Commun* 8, 15464–15464.
<https://doi.org/10.1038/ncomms15464>
- Singh, P., Schimenti, J.C., Bolcun-Filas, E., 2015. A mouse geneticist's practical guide to CRISPR applications. *Genetics* 199, 1–15.
<https://doi.org/10.1534/genetics.114.169771>

- Thorvaldsdóttir, H., Robinson, J.T., Mesirov, J.P., 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* 14, 178–192. <https://doi.org/10.1093/bib/bbs017>
- Wang, H., Yang, H., Shivalila, C.S., Dawlaty, M.M., Cheng, A.W., Zhang, F., Jaenisch, R., 2013. One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell* 153, 910–918. <https://doi.org/10.1016/j.cell.2013.04.025>
- Wick, R.R., Judd, L.M., Gorrie, C.L., Holt, K.E., 2017. Completing bacterial genome assemblies with multiplex MinION sequencing. *Microb Genom* 3, e000132–e000132. <https://doi.org/10.1099/mgen.0.000132>
- Wick, R.R., Judd, L.M., Holt, K.E., 2019. Performance of neural network basecalling tools for Oxford Nanopore sequencing. *Genome Biol* 20, 129–129. <https://doi.org/10.1186/s13059-019-1727-y>
- Yang, H., Wang, H., Shivalila, C.S., Cheng, A.W., Shi, L., Jaenisch, R., 2013. One-step generation of mice carrying reporter and conditional alleles by CRISPR/Cas-mediated genome engineering. *Cell* 154, 1370–1379. <https://doi.org/10.1016/j.cell.2013.08.022>