



High-throughput genotyping of high-homology mutant mouse strains by next-generation sequencing

Diane Gleeson, Debarati Sethi, Radka Platte, Jonathan Burvill, Daniel Barrett, Shaheen Akhtar, Michaela Bruntraeger, Joanna Bottomley, Sanger Mouse Genetics Project, James Bussell¹, Edward Ryder^{*,2}

Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge CB10 1SA, UK

ARTICLE INFO

Keywords:

Genotyping
Mouse
NGS
CRISPR
Mutant
QC

ABSTRACT

Genotyping of knockout alleles in mice is commonly performed by end-point PCR or gene-specific/universal cassette qPCR. Both have advantages and limitations in terms of assay design and interpretation of results. As an alternative method for high-throughput genotyping, we investigated next generation sequencing (NGS) of PCR amplicons, with a focus on CRISPR-mediated exon deletions where antibiotic selection markers are not present. By multiplexing the wild type and mutant-specific PCR reactions, the genotype can be called by the relative sequence counts of each product. The system is highly scalable and can be applied to a variety of different allele types, including those produced by the International Mouse Phenotyping Consortium and associated projects.

One potential challenge with any assay design is locating unique areas of the genome, especially when working with gene families or regions of high homology. These can result in misleading or ambiguous genotypes for either qPCR or end-point assays. Here, we show that genotyping by NGS can negate these issues by simple, automated filtering of undesired sequences. Analysis and genotype calls can also be fully automated, using FASTQ or FASTA input files and an in-house Perl script and SQL database.

1. Introduction

The mouse remains a vitally important model in studying gene function [1] and modelling human disease [2]. A firm commitment to the 3Rs [3] and ARRIVE guidelines [4] from mouse production centers to ensure validity and reproducibility in experiments [5], dictate that over-production of breeding cohorts or repeats of phenotyping should be avoided whenever possible. Even after standardization of variables such as genetic background, husbandry regimes, and data analysis [6], possible sources of error include misidentification of mice, incorrectly targeted or poorly characterized mutations, and incorrect genotyping strategies or assay design. The longer these issues take to resolve, the more costly they are in terms of animal use, lost research time and housing. Similarly, delays in genotyping turnaround time can have

downstream effects on breeding decisions and cage use. Accurate, timely, and cost-effective genotyping of mice is therefore of utmost importance to facilitate correct cohort breeding and phenotyping analysis.

1.1. Large scale mouse production initiatives.

In the last ten years there has been a move away from individual laboratories creating knockout mouse models to a more centralized approach, highlighted by the Sanger Institute Mouse Genetics Project (MGP) [7], EUCOMM/KOMP [8], EUMODIC [9], and the International Mouse Phenotyping Consortium (IMPC) [10]. These projects have been highly successful, forming the basis of high-throughput screens studying developmental phenotypes [11], cancer metastasis [12], sexual

Abbreviations: CRISPR, clustered regularly interspaced short palindromic repeat; het, heterozygous; hom, homozygous; NGS, next generation sequencing; NHEJ, non-homologous end joining; PCR, polymerase chain reaction; QC, quality control; sgRNA, single guide RNA; WT, wild-type.

* Corresponding author.

E-mail address: edward.ryder@lscgroup.com (E. Ryder).

¹ Present address: Biomedical and Veterinary Services Department, University of Oxford, Oxford, OX1 3TA, UK.

² Present address: LGC, Sport and Specialised Analytical Services, Newmarket Road, Fordham, Ely CB7 5WW, UK.

<https://doi.org/10.1016/j.ymeth.2020.10.011>

Received 25 July 2020; Received in revised form 25 September 2020; Accepted 18 October 2020

Available online 20 October 2020

1046-2023/© 2020 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

dimorphism [13], hearing loss [14], metabolism [15], placental defects [16], and transcriptome signatures during development [17]. The mouse strains generated are freely available to the research community through repositories such as Infrafrontier [18] and MMRRC [19].

The advantages of this approach are cost-saving at scale, reduced numbers of mice used due to standardized and optimal breeding, and a unified approach to mouse quality control. As the IMPC moves into its third phase of research, focusing on precision models [20] and biological function of all genes [21], accurate quality control and genotyping of possibly complex mutation outcomes [22–24] will be paramount.

1.2. Allele QC and genotyping techniques

The mouse production initiatives discussed above historically used the LacZ gene-trapping tm1a cassette and derivative alleles created by EUCOMM/KOMP-CSD [25], with a switch to CRISPR/Cas9 mutations [26] (primarily exon deletions) as the technology became available and optimized for large-scale mutagenesis [27]. Whereas tm1a alleles could be genotyped in a high-throughput manner using a qPCR assay designed to the LacZ or Neo markers within the cassette [7], exon deletions carry no such signatures (Fig. 1). Genotyping deletions therefore relies predominantly on either end point PCR [28] or loss of wild type allele qPCR (LoA) [29,30] which uses an assay designed to the wild type locus and compares the amplification of an unknown sample against a calibration control.

Alternative forms of routine genotyping include Kompetitive allele

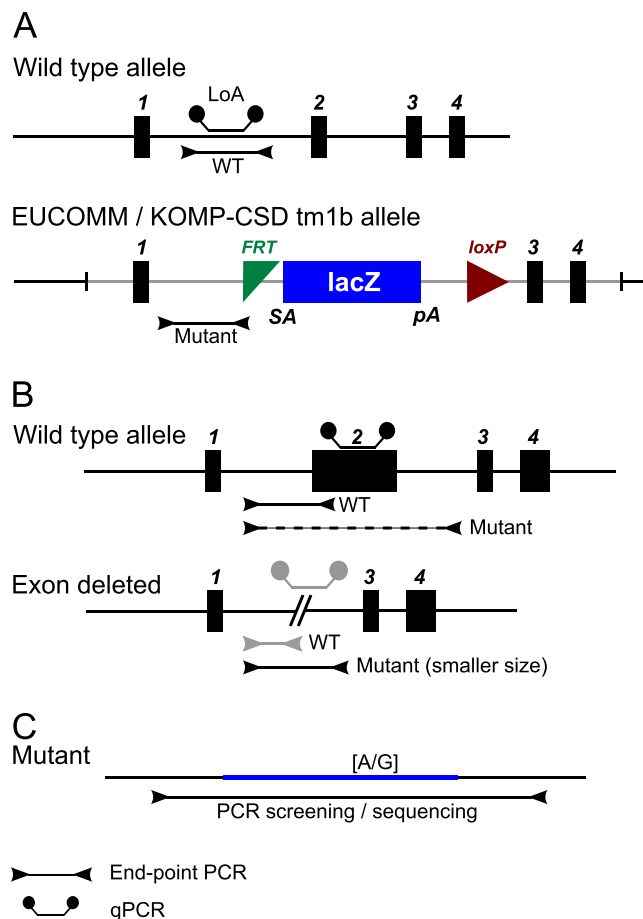


Fig. 1. Common knockout designs. End-point PCR assays for EUCOMM/KOMP-CSD mutants are designed to amplify either the wild type or mutant alleles and fail if the allele is not present. Exon deletions follow this same premise, although the mutant assay may amplify a larger size if the exon is sufficiently small. For SNP detection, a single assay spans the region of interest.

specific PCR (KASP) assays [31], High Resolution Melt Analysis (HRMA) [32] and fluorescent PCR-capillary gel electrophoresis [33]. Although KASP assays have been used to genotype SNPs and small indels from NHEJ CRISPR experiments [34], their use in exon deletions has not yet been explored.

Quality control of mouse mutant strains often takes a multi-assay and multi-technique approach, which is vital for situations where results may be confusing or conflicting [35]. Southern blot has been accepted as the gold standard of characterizing mutations based on homologous recombination [36] but is time consuming and difficult to adapt to a high-throughput environment. Alternatively, digital droplet PCR (ddPCR) has been used as an alternative to pre-microinjection QC for karyotyping [37] and evaluation of CRISPR mutants [38]. Although excellent for evaluation of a mutation, limitations in throughput and cost per sample makes digital droplet PCR less desirable for routine genotyping.

The clonal amplification technique used in NGS makes it a valuable tool for assessing the outcome of small, indel-based knockout CRISPR/Cas9 experiments [39], where compound alleles can make interpretation of capillary sequencing, and thus which mice to breed, very challenging in the G0 generation. Third-generation, long-read sequencing has also been used to determine allele structure in gene editing outcomes of more complex alleles [24].

1.3. Quality control and confirmation of homozygotes

One potential drawback of end point PCR and LoA qPCR genotyping, where an assay is designed to fail in the presence of the wild type allele, is that non-specific amplification may give a false result. Thus, strains which are actually homozygous viable may appear as homozygous lethal when genotyped. Since allele QC is performed initially on heterozygous founders, these problems may not be detected until non-Mendelian ratios of genotypes start appearing in heterozygote cohort breeding. This issue is particularly acute in gene families, where high homology to other regions of the genome can severely limit assay design.

1.4. Genotyping challenges in regions of high homology and gene families

Analysis of mouse gene paralogues from Ensembl v101 data [40] (not including olfactory receptors, vomeronasal genes, or genes on the Y chromosome) revealed that 2829 genes have over 80% identity to one or more genes within the genome (12.5% total). Raising the threshold to 90% identity reduces this to 1942 genes, but still represents approximately 9.5% of the genome. This not only has implications for selecting unique guide RNA design in creating mutant strains, but also designing specific genotyping assays. Two examples of mouse strains produced by the Sanger Institute Mouse Pipelines teams are derived from the Psg and Sirpb1 gene families.

The mouse pregnancy-specific glycoprotein (Psg) family consists of 17 members [41]. The mouse locus is approximately 1.74 Mb and contains six Psg genes in the A2 chromosome band and eleven genes in the A3 band on Chr 7 [42]. The signal-regulatory protein beta Sirpb1a-c family resides on a 0.5 Mb portion of the A1 region on Chr 3. Genes in this family have over 90% sequence homology to each other and 48% homology with the nearby gene Gm5150 [40].

1.5. Genotyping by next generation sequencing

In addition to mouse whole-genome [43] and exome sequencing [44], massively parallel amplicon sequencing has been used to classify the mouse gut microbiome [45] and gene editing outcomes [46]. Although the concept of using the extensive capacity of NGS to sequence single PCR products for mouse genotyping seems counter-intuitive, careful multiplexing of libraries can provide hundreds of samples in one sequencing run, vastly reducing the sequencing cost per sample. Coupled with the additional information sequencing provides,

decreasing costs of reagents and cheaper, smaller capacity instruments to reduce potential batching times, NGS becomes a more feasible proposition.

To investigate the utility of Illumina-based sequencing for mouse genotyping, especially in challenging regions of high homology and gene families, we performed a pilot study on twenty-six mutant strains and compared the results to our existing qPCR assays.

2. Materials and methods

The care and use of all mice in this study were in accordance with the UK Home Office regulations, UK Animals (Scientific Procedures) Act of 1986, and were approved by the Wellcome Trust Sanger Institute Ethical Review Committee.

2.1. Gene choice

Genes for this study were chosen based on sample availability, existing genotyping issues, or homology to other genes. In addition to the CRISPR/Cas9 exon deletions, we chose one CRISPR/Cas9-derived SNP and two EUComm/KOMP tm1b strains [47] to test the versatility of the system (Table 1).

2.2. Assay design

The rationale of assay designs for the different mutations are outlined in Fig. 1. A combination of end point and LoA qPCR is used to call and confirm the genotype. Primer sequences are shown in Supplemental Table ST1.

2.3. DNA extraction

DNA was isolated from ear punches taken as part of the mouse identification process, using the TaqMan® Sample-to-SNP™ kit (Thermo Scientific). Ear clips were heated to 95 °C for 3 min in 50 µl of lysis buffer, centrifuged briefly and neutralized with the addition of 50 µl of stabilizing solution. Extracted DNA was stored at –20 °C until required.

2.4. Two-round PCR

Amplicon-derived sequencing libraries were constructed using a 2-step PCR method (Fig. 2), similar to that used in metagenomics studies [48] and determining CRISPR/Cas9 editing outcomes [49,50]. Gene-specific primers (Table ST1) were tagged at the 5' end with the following sequences: PE_forward primer_F1 ACACTCTTCCCTACACGAGCTC TTCCGATCT, PE_reverse primer_R1 CGGTCTCGGCATTCCTGCT-GAACCGCTCTCCGATCT. Round 1 amplification was performed in a multiplex reaction of all three genotyping primers using Platinum Taq (Thermo Scientific) and an annealing temperature of 58 °C (Supplemental Data S1). Where possible, existing primers from our mutation QC steps were used as the basis of the PCR reaction. If the forward primer was greater than 150 bp from the discriminatory sequence motif, or the amplicon larger than 250 bp, then new primers were designed. Reactions were visualized using a

Table 1

Genes and allele types used in the study which were successfully amplified in the NGS pipeline. Genes highlighted in bold are characterized as high-homology (greater than 90% homology over the assay design area).

Mutation type	Genes
EUComm/KOMP tm1b	Afap111, Id2
SNP	Pcdh15
Exon deletion	Ceacam15, Clic3, Crip1, Etv3, Exo5, Gm17750, Itgam, Msl1, Pilrb2 , Psg18 , Psg19 , Psg21 , Psg26 , Sirpb1a , Sirpb1b , Sirpb1c , Sqrld, Stom, Vmn2r27 , Zfp748

Qiagel Advanced capillary electrophoresis instrument (Qiagen) and DNA Screening Kit, and genotypes called on the amplification pattern of wild type and mutant alleles.

The second round of PCR was performed as described in Bruntraeger et al. [49]. Briefly, reactions from round 1 are amplified in the presence of a high-fidelity polymerase and primers with P5 or P7 flow cell attachment sequences, 8 bp barcodes, and sequences identical to the round 1 PCR tag. Each 96 well plate contains a constant i5 index and a unique i7 index which allows for many samples and plates to be multiplexed on one MiSeq run, increasing throughput and decreasing costs per sample per run. Aliquots of each second round reaction were pooled and purified by use of Ampure XP bead (Beckman Coulter) size selection [51]. Sequencing was performed by the Sanger Institute DNA Pipelines teams, using a MiSeq instrument (Illumina) and a MiSeq Reagent Kit v2 (300-cycles) (Illumina).

2.5. Real time qPCR

FAM-labelled TaqMan™ assays were designed using Primer Express 3 (Thermo Scientific) and obtained from Thermo Scientific. Reactions were performed on a Vii7 instrument using GTXpress master mix in the presence of an endogenous control VIC-labelled Tfrc assay (Thermo Scientific). Reaction and cycling conditions are shown in Supplemental Data S2. Genotypes were called using the $2^{-\Delta\Delta C_t}$ method [52], comparing against known heterozygous and wild type calibration controls.

2.6. Sequence analysis

Sequence data analysis and genotype calling is performed by a custom Perl script, which accepts directories of sequence files in either FASTQ or FASTA format (Fig. 3). The script connects to a MySQL database containing mouse and gene ID, and compares each sequence to a wild type and mutant-specific 50 bp motif (Supplemental Table ST2). By this stage in the colony expansion the mutations and breakpoints have been characterized, so re-alignment of sequences to the genome using software such as Crispresso2 [53] is not necessary, and may prove problematic due to the deletion size. A motif-based approach therefore offers a simpler method of analysis and potential increases in computational speed.

Matches are assigned to bins, the ratio of wild type to mutant sequences calculated, and the genotype reported in a tab-delimited file. Genotypes are set to failed if the maximum number of sequence reads in both bins were below 500. In the case of CRISPR exon deletions where the mutant-specific assay is amplified in the wild type allele (albeit a different size), a filter motif removes the sequence from the analysis. The analysis presented here was performed primarily using the R1 sequence, but the Perl script can also process the R2 end if required. The Perl scripts and database schema are available from GitHub at (<https://github.com/EdRyder/ngs-genotyping>).

3. Results.

3.1. PCR amplification

In total, 23/26 genes tested (88%) amplified a product without any additional reaction optimization beyond our existing workflows (Table 1). Between twelve and twenty-four samples of each mouse strain were genotyped, chosen on the basis of a range of existing genotypes and litter sizes to facilitate sample processing. An example of results for the CRISPR-derived Ceacam15 exon deletion strain is shown in Table 2.

To demonstrate the high-throughput potential of next-generation sequencing for this application, one of the experiments included the pooling of 768 mice from 25 strains within one sequencing run.

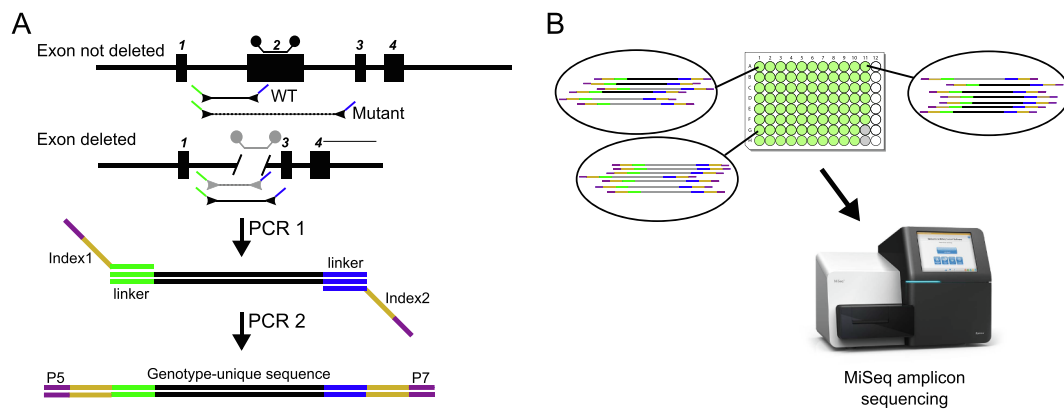


Fig. 2. NGS genotyping strategy for CRISPR-derived exon deletions. A) In PCR 1, tailed primers containing a linker sequence are amplified in a WT and mutant-specific multiplex reaction. A second round of PCR is then performed using additional tailed primers containing linker flow cell adaptors and an 8 bp barcode index sequence at each end. B) By using plate-specific and well-specific barcodes, many plates can be multiplexed into one sequencing run. After the sequencing by synthesis is complete, the MiSeq then deconvolutes the clusters and assigns the results to FASTQ files per sample, based on the barcode sequence.

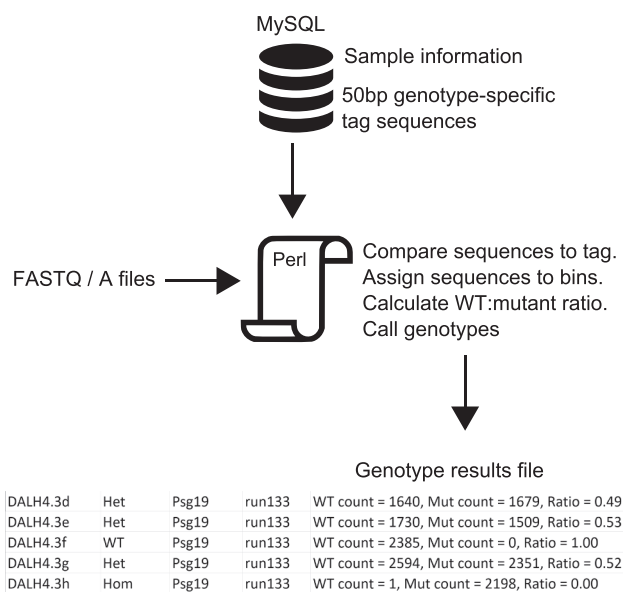


Fig. 3. Data analysis pipeline. Mouse ID and strain information is retrieved from the MySQL database by processing the FASTQ filename. Sequences are compared to a database of genotype-specific motifs, any matches reported, and the genotype called on the ratio. FASTQ files can be pre-filtered on the quality score if required.

3.2. NGS genotyping ratios

A graph of sequence ratios per gene is shown in Fig. 4. Three distinct clusters were formed in most cases, and average WT:mutant ratios for heterozygotes ranged from 0.39 to 0.729 (Supplemental Data S3), reflecting possible allele-specific bias in amplification of the two products. For the NGS assays, the average difference between WT and mutant assays was 16 bp, with a median of 10 bp (assays were originally designed to preserve a minimum size difference). No clear pattern was observed between the difference in WT and mutant amplicon sizes, and the heterozygote genotype ratio. The cause of the altered ratios may, therefore, be due to other factors such as secondary structure or GC content.

3.3. Comparisons with LoA qPCR results

Genotypes were compared with previous LoA qPCR results, splitting the genes into two groups based on whether they were highly

homologous (>90%) to other genes over the design area (Table 3).

For the ‘no high homology’ genes, agreement between the two methods was 90% (231/262), which was lower than expected. Upon further investigation, the majority of failed matches belonged to the genes Gm17750 and Crip1 (8 and 13 mismatches, respectively). Amplification of Gm17750 suppressed the mutant assay in the NGS PCR compared to the wild type (data not shown), lowering the ratio scores and causing heterozygotes to appear as wild types in the analysis.

Interestingly, Crip1 formed a fourth distinct cluster of 5 mice with WT:Mut count ratios between 0.15 and 0.23. The reason for this is unclear, but all the affected mice were from the same litter (Crip1_7.2) and genotyped as homozygous by LoA qPCR (due to the band sizes of 185 bp and 186 bp for the wild type and mutant assays, the genotype could not be called by observing the end point PCR in this study). A separate litter of 11 mice (Crip1_15.3) all genotyped as expected, and contained a mix of homozygotes, heterozygotes, and wild type mice. As all mice in the breeding cohorts were derived from a single exon-deleted heterozygous founder (a characterized G1 male), this is unlikely to be the result of multiple mutations within the colony and may be due to a poor DNA preparation in the Crip1_7.2 litter. Removing Crip1 and Gm17750 from the analysis increased the percentage match with the LoA qPCR results to 98% in ‘no homology’ genes where a result was obtained for both samples.

3.4. Genotyping and filtering of high-homology targets

The concordance between NGS and LoA qPCR dropped to 82% when analyzing ‘high-homology’ genes, indicating potential issues with our existing qPCR-based method. LoA qPCR assays were originally designed to minimize the amount of homology with other genomic regions (which can be iterative and labor-intensive), while staying within the parameters required for successful TaqMan-based amplification and PCR efficiency.

LoA qPCR Assays with greater than 10 base mismatches to other regions (e.g. exon deletion strains Psg18, Psg19, Psg26) agreed with the NGS results. The Pilbr2 mutants also showed no issues at all on the mice tested, with 100% agreement between the LoA qPCR, and NGS genotyping, despite only 5 base mismatches across the 58 bp LoA assay.

We found a greater likelihood of the LoA qPCR and NGS assays producing conflicting results when the number of mismatches within the LoA qPCR assay were below 5 bp. Exon deletion strains for Psg21, Sirbp1a, Sirbp1c, and Vmn2r27 all showed serious issues in genotyping concordance, although Sirbp1b did not, despite an only 2 bp difference in the assay sequence.

The advantage of the NGS approach over LoA qPCR is the precise sequence of the products can be determined, allowing the filtering out of

Table 2

Ceacam15 genotyping, comparing the Qiaxcel end-point PCR genotypes, LoA qPCR and NGS results. NGS ratios are consistent with previous LoA qPCR genotype calls. Mice are identified by colony prefix, mating, litter and individual. The LoA score is reported here as the relative quantification (RQ) value, where 1 = amplification identical to WT controls and 0 = no amplification (homozygote). Heterozygotes are called where the RQ value is in a defined range of 0.5. PCR gel sizes: WT = 141 bp, mutant = 162 bp.

Mouse ID	PCR	Gel genotype	LoA score	LoA genotype	NGS WT count	NGS Mutant count	NGS Ratio	NGS genotype	Match
DAEZ20.2f		Homozygous	0	Homozygous	8	15162	0	Homozygous	Yes
DAEZ20.2g		Heterozygous	0.488	Heterozygous	11253	7682	0.59	Heterozygous	Yes
DAEZ20.2h		Homozygous	0	Homozygous	8	16068	0	Homozygous	Yes
DAEZ20.2i		Wildtype	0.971	Wildtype	16026	417	0.97	Wildtype	Yes
DAEZ20.2j		Wildtype	0.977	Wildtype	13734	5	1	Wildtype	Yes
DAEZ19.3a		Heterozygous	0.349	Heterozygous	7449	5242	0.59	Heterozygous	Yes
DAEZ19.3b		Heterozygous	0.54	Heterozygous	9033	5778	0.61	Heterozygous	Yes
DAEZ21.3a		Heterozygous	0.507	Heterozygous	9484	7251	0.57	Heterozygous	Yes
DAEZ21.3b		Wildtype	0.98	Wildtype	11609	3	1	Wildtype	Yes

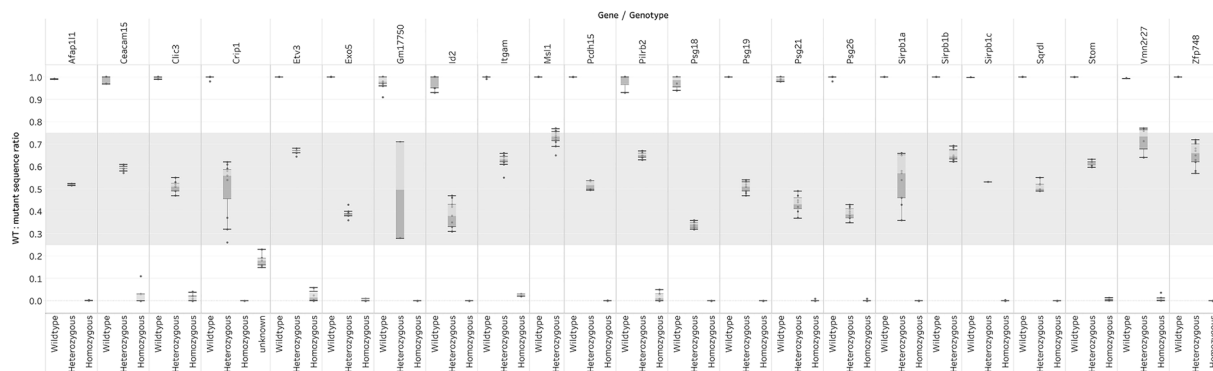


Fig. 4. Genotyping results per mouse strain using next generation sequencing. Ratios are calculated by comparing the number of wild type motif sequence hits against mutant motif hits. The shaded area represents the ideal range for heterozygous genotypes.

Table 3

Genotype comparisons between LoA qPCR and NGS sequence ratios. High homology genes have a lower genotype concordance than those with no homology, suggesting an issue with the specificity of the design in one or either assay.

NGS and LoA qPCR agree?	High homology gene	
	No	Yes
Yes	90%	82%
No	10%	18%

unintended amplification. Exon deletion Psg21 has a 97% sequence match to Psg23 across the entire 2.4 kb deletion and flanking areas and numerous matches to other Psg genes (e.g. Psg27) throughout the region (Supplemental Data S4). This makes designing specific qPCR assays extremely challenging, and in the case of Psg21 resulted in an assay differing in only 2 bp from Psg23 and Psg 27, albeit at different locations. While the end point, gel-based PCR results for Psg21 agreed with the LoA qPCR results, three mice from litter Psg21_3.2 (het × het mating) were called as clear homozygotes by our NGS pipeline (Table 4). To investigate further, we mapped the FASTQ file for Psg21_3.2a to the mouse genome (release mm10) using Bowtie2 [54]. From a total of 3702

Table 4

Psg21 genotyping results comparison. Two mice were genotyped as heterozygotes by LoA qPCR and from the PCR 1 reaction used in the NGS experiment. Further analysis, however, revealed that the WT band was spurious, and the true genotype was homozygous (shaded regions). PCR gel sizes: WT = 191 bp, mutant = 157 bp.

Mouse ID	PCR	Gel genotype	LoA score	LoA genotype	NGS WT count	NGS Mutant count	NGS Ratio	NGS genotype	Match
DAKM3.2a		Heterozygous	0.54	Heterozygous	0	1843	0	Homozygous	No
DAKM3.2b		Wildtype	0.87	Wildtype	1432	7	1	Wildtype	Yes
DAKM3.2c		Heterozygous	0.51	Heterozygous	1445	1782	0.45	Heterozygous	Yes
DAKM3.2d		Heterozygous	0.66	Heterozygous	1028	1398	0.42	Heterozygous	Yes
DAKM3.2e		Heterozygous	0.34	Heterozygous	18	3478	0.01	Homozygous	No
DAKM3.2f		Wildtype	0.88	Wildtype	1831	41	0.98	Wildtype	Yes
DAKM3.2g		Heterozygous	0.56	Heterozygous	2174	2972	0.42	Heterozygous	Yes

mapped sequence reads corresponding to the wild type assay (the mutant-specific assay reported as unmapped due to the mutation), 2406 (65%) mapped to a homologous region in Psg23, and 1268 (34%) to Psg27. Mice from litter Psg21_2.1 (het × WT mating) all agreed with the existing genotypes.

Three genes from the *Sirpb1* family were also processed as part of the study, and although LoA qPCR and NGS genotyping agreed for *Sirpb1b*, *Sirpb1c* failed to detect NGS homozygotes by LoA qPCR, and the LoA assay for *Sirpb1a* was highly variable between litters. The litter *Sirpb1a_9.1*, for example, was reported as all heterozygotes by LoA qPCR but genotyped as all homozygous by NGS. The FASTQ file for *Sirpb1a_9.1a*, a heterozygote by LoA qPCR, was aligned to the genome and showed only 8 alignments to the WT *Sirpb1a* allele (and 0 hits for the motif region), but 2517 hits to *Sirpb1c* and 371 to *Gm5150*. From this data, we conclude that the true genotypes for this litter are homozygotes and that the LoA qPCR assay is amplifying non-specific products at a signal high enough to produce a false result.

3.5. Cross-contamination of samples and the effects on NGS sequence ratios

One advantage of qPCR over gel-based end point PCR methods is that the analysis method is resilient to low levels of cross-contamination, which can occur at any stage throughout the process of ear clipping to PCR experiment. In negative controls, wild type (cassette qPCR), or homozygote (LoA qPCR) samples, any amplification in this type of event is usually associated with a much higher cycle threshold (Ct) value than typically seen for the assay. The comparative analysis reports these as low copy numbers which are then filtered out of the genotype calculation.

To investigate the robustness of the next-generation sequencing approach, we spiked wild type or homozygous mice with DNA from the corresponding opposite genotype at different levels. Results are shown in Supplemental Data S5. A 1:1 ratio of WT:hom DNA gave values indistinguishable from heterozygotes as expected, with a shift in sequence ratios observed with increasing levels of spiked DNA. Spiking ratios of 1:500 and above were out of the genotype calling window for heterozygote mice for all genes tested.

4. Discussion

We present here the results of a pilot project to evaluate next generation sequencing for both routine genotyping and additional quality control of mutant mouse strains. The workflow has been tested primarily on CRISPR/Cas9-derived exon deletions, but also on EUCOMM/KOMP alleles and a SNP mutation to assess the versatility of the workflow. Twenty-three of twenty-six (88%) genes tested amplified a product without any additional reaction optimization beyond our routine end point PCR pipeline, despite the addition of tails over 30 bp to the 5' end of each primer. No changes were required to the DNA isolation protocol, and primer design required minimal changes to be within 150 bp of the mutation boundary/motif region. Performing sequencing on the Illumina MiSeq with a larger cycle kit would relax these restrictions, but these were not available during the study.

In addition to the sequence-level confirmation of the assay tested, another advantage of next generation sequencing over other approaches is its scalability. By using a dual combinational indexing approach, hundreds of samples can be batched into one sequencing run (which typically takes fewer than 24 hrs on a MiSeq, less if only a single end is sequenced), greatly reducing sequencing costs per sample. Combining the 16 standard Illumina Nextera XT i5 indexes with the i7 indexes described in Bruntraeger et al. [49], for example, provides barcodes for 1536 samples within one sequencing run.

For high-throughput mouse programs such as the IMPC, any increase in turn-around time due to batching would, therefore, be marginal and within the weaning time (where mice are transferred into new cages) if

the identification clips are taken in a timely manner. Comparative costs of NGS and LoA qPCR will depend greatly on equipment available to the genotyping facility, and genotyping workflows. For example, the costs associated with the extra step required for the indexing round in NGS may be offset by not requiring the fluorescent hydrolysis probe needed for LoA qPCR. With pipetting automation, labor costs per sample can also be kept low for both methods, and any existing high-throughput setup for LoA qPCR or end-point PCR should be easily adaptable to the sequencing approach.

Our analysis script greatly simplifies the processing of results and can be used with minimal training. Additionally, the tab-delimited output format allows results to be easily processed downstream and uploaded into a mouse tracking database if required. Although the Bioperl module for FASTQ processing can take several minutes per file (depending on size), FASTA format is significantly faster at only a few seconds if speed of analysis is of greater importance than filtering for lower quality sequences. The tight clustering of motif ratios within different genes, coupled with the large number of sequence reads associated with PCR products, indicates that low levels of index hopping [55] is unlikely to pose any issues if the workflow is performed on patterned flow cell instruments. We found that low levels of cross contamination (for example, from clipping regimes or sample processing) does not cause a significant issue in genotyping, and could be filtered out in the analysis if the WT:mutant ratio is outside of normal limits for the gene in question.

4.1. Use of NGS for genotyping mutations in gene families

By harnessing the parallel nature of NGS, non-specific or closely related sequences can be easily filtered out, enhancing the accuracy of the genotyping result, especially for mutations in gene families. Sanger sequencing of the wild type allele of *Sirpb1a*, for example, differed from the other family members by fewer than 10 bp. As initial QC is performed on heterozygotes, slight mismatches in the primer sequence may result in amplification of the wild type amplicon being too strong for the analyst to confidently separate true from background signal. Certainly, from our own experience in attempting to genotype in this manner, the analysis is complicated and potentially error prone, especially when no exogenous sequences such as selection markers are present. The use of NGS and placement of unique motifs for filtering (Supplemental Data S6) removes this issue, giving a much higher confidence in the results. We therefore highly recommend when processing exon deletions with high homology, especially those with minimal mismatches within the assay, that alternative assays and sequencing are performed as an additional QC step during heterozygote × heterozygote cohort matings. This will ensure that potential homozygotes are not mis-genotyped, and the strain mistakenly assigned as homozygous-lethal. Non-Mendelian genotype ratios can also offer a vital clue that the assays are not performing as expected (for example, all heterozygote offspring from a het × het mating), especially later in the breeding scheme.

CRedit authorship contribution statement

Diane Gleeson: Investigation, Methodology, Writing - review & editing. **Debarati Sethi:** Investigation. **Radka Platte:** Investigation. **Jonathan Burvill:** Investigation. **Daniel Barrett:** Investigation. **Shaheen Akhtar:** Investigation. **Michaela Bruntraeger:** Resources. **Joanna Bottomley:** Project administration. **Sanger Mouse Genetics Project:** . **James Bussell:** Conceptualization, Supervision. **Edward Ryder:** Conceptualization, Investigation, Software, Supervision, Visualization, Writing - original draft, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank staff from the Sanger Institute's Research Support Facility, Mouse Pipelines, DNA Pipelines, and Mouse Informatics Group for their excellent support. We also thank Andrew Bassett for his valuable input on 2-step PCR and the manuscript, and Vivek Iyer for informatics advice. This work was supported by the Wellcome Trust under Grant No. WT206194.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ymeth.2020.10.011>.

References

- [1] C.B. Gurumurthy, K.C. Kent Lloyd, Generating mouse models for biomedical research: Technological advances, *DMM Dis. Model. Mech.* 12 (2019), <https://doi.org/10.1242/dmm.029462>.
- [2] P. Cacheiro, M.A. Haendel, D. Smedley, T. Meehan, J. Mason, H.H. Mashhadi, V. Muñoz-Fuentes, G. Tocchini, K.K.C. Lloyd, C. McKerie, L. Bower, D. Clary, L.M. Nutter, A.M. Flenniken, L. Teboul, G. Codner, S. Wells, Y. Herault, T. Sorg, L. Vasseur, M. Selloum, M. Roux, H. Jacobs, H. Meziane, M.F. Champy, G. B. About, S. Murray, E. Chesler, V. Kumar, J. White, R.E. Braun, A.L. Beaudet, M. E. Dickinson, J.D. Heaney, I. Lorenzo, D.G. Lanza, C.L. Reynolds, C.S. Ward, R. C. Samaco, S. Veeraragavan, C.W. Hsu, A.E. Christianson, J.J. Gallegos, J. R. Seavitt, A. Gaspero, J.R. Green, A. Garza, R. Garza, R. Bohat, S.D.M.B. Sedlacek, New models for human disease from the International Mouse Phenotyping Consortium, *Mamm. Genome* 30 (2019) 143–150, <https://doi.org/10.1007/s00335-019-09804-5>.
- [3] A. Greenfield, Editing mammalian genomes: ethical considerations, *Mamm. Genome* 28 (2017) 388–393, <https://doi.org/10.1007/s00335-017-9702-y>.
- [4] C. Kilkenny, W.J. Browne, I.C. Cuthill, M. Emerson, D.G. Altman, Improving bioscience research reporting: the ARRIVE guidelines for reporting animal research, *PLoS Biol.* 8 (2010), e1000412, <https://doi.org/10.1371/journal.pbio.1000412>.
- [5] M.J. Justice, P. Dhillon, Using the mouse to model human disease: Increasing validity and reproducibility, *DMM Dis. Model. Mech.* 9 (2016) 101–103, <https://doi.org/10.1242/dmm.024547>.
- [6] N.A. Karp, D. Melvin, R.F. Mott, Robust and sensitive analysis of mouse knockout phenotypes, *PLoS One* 7 (2012), e52410, <https://doi.org/10.1371/journal.pone.0052410>.
- [7] J.K. White, A.-K. Gerdin, N.A. Karp, E. Ryder, M. Buljan, J.N. Bussell, J. Salisbury, S. Clare, N.J. Ingham, C. Podrini, R. Houghton, J. Estabel, J.R. Bottomley, D. G. Melvin, D. Sunter, N.C. Adams, D. Tannahill, D.W. Logan, D.G. MacArthur, J. Flint, V.B. Mahajan, S.H. Tsang, I. Smyth, F.M. Watt, W.C. Skarnes, G. Dougan, D.J. Adams, R. Ramirez-Solis, A. Bradley, K.P. Steel, Genome-wide generation and systematic phenotyping of knockout mice reveals new roles for many genes, *Cell* 154 (2013) 452–464, <https://doi.org/10.1016/j.cell.2013.06.022>.
- [8] A. Ayadi, M.-C. Birling, J. Bottomley, J. Bussell, H. Fuchs, M. Fray, V. Gailus-Durner, S. Greenaway, R. Houghton, N. Karp, S. Leblanc, C. Lengger, H. Maier, A.-M. Mallon, S. Marschall, D. Melvin, H. Morgan, G. Pavlovic, E. Ryder, W. C. Skarnes, M. Selloum, R. Ramirez-Solis, T. Sorg, L. Teboul, L. Vasseur, A. Walling, T. Weaver, S. Wells, J.K. White, A. Bradley, D.J. Adams, K.P. Steel, M. Hrabě de Angelis, S.D. Brown, Y. Herault, Mouse large-scale phenotyping initiatives: overview of the European Mouse Disease Clinic (EUMODIC) and of the Wellcome Trust Sanger Institute Mouse Genetics Project, *Mamm. Genome* (2012), <https://doi.org/10.1007/s00335-012-9418-y>.
- [9] M.H. de Angelis, G. Nicholson, M. Selloum, J.K. White, H. Morgan, R. Ramirez-Solis, T. Sorg, S. Wells, H. Fuchs, M. Fray, D.J. Adams, N.C. Adams, T. Adler, A. Aguilar-Pimentel, D. Ali-Hadjji, G. Amann, P. André, S. Atkins, A. Auburtin, A. Ayadi, J. Becker, L. Becker, E. Bedu, R. Bekeredjian, M.-C. Birling, A. Blake, J. Bottomley, M.R. Bowl, V. Braut, D.H. Busch, J.N. Bussell, J. Calzada-Wack, H. Cater, M.-F. Champy, P. Charles, C. Chevalier, F. Chiani, G.F. Codner, R. Combe, R. Cox, E. Dalloneau, A. Dierich, A. Di Fenza, B. Doe, A. Duchon, O. Eickelberg, C. T. Esapa, L. El Fertak, T. Feigel, I. Emelyanova, J. Estabel, J. Favor, A. Flenniken, A. Gambadoro, L. Garrett, H. Gates, A.-K. Gerdin, G. Gkoutos, S. Greenaway, L. Glasl, P. Goetz, I.G. Da Cruz, A. Götz, J. Graw, A. Guimond, W. Hans, G. Hicks, S. M. Hölter, H. Höfler, J.M. Hancock, R. Hoehndorf, T. Hough, R. Houghton, A. Hurt, B. Ivandic, H. Jacobs, S. Jacquot, N. Jones, N.A. Karp, H.A. Katus, S. Kitchen, T. Klein-Rodewald, M. Klingenspor, T. Klopstock, V. Lalanne, S. Leblanc, C. Lengger, E. le Marchand, T. Ludwig, A. Lux, C. McKerie, H. Maier, J.-L. Mandel, S. Marschall, M. Mark, D.G. Melvin, H. Meziane, K. Micklich, C. Mittelhauser, L. Monassier, D. Moulaert, S. Muller, F. Naton, F. Neff, P.M. Nolan, L.M.J. Nutter, M. Ollert, G. Pavlovic, N.S. Pellegata, E. Peter, B. Petit-Demoulière, A. Pickard, C. Podrini, P. Potter, L. Pouilly, O. Puk, D. Richardson, S. Rousseau, L. Quintanilla-Fend, M.M. Quwaillid, I. Racz, B. Rathkolb, F. Riet, J. Rossant, M. Roux, J. Rozman, E. Ryder, J. Salisbury, L. Santos, K.-H. Schable, E. Schiller, A. Schrewe, H. Schulz, R. Steinkamp, M. Simon, M. Stewart, C. Stöger, T. Stöger, M. Sun, D. Sunter, L. Teboul, I. Tilly, G.P. Tocchini-Valentini, M. Tost, I. Treise, L. Vasseur, E. Velot, D. Vogt-Weisenhorn, C. Wagner, A. Walling, M. Wattenhofer-Donze, B. Weber, O. Wendling, H. Westerberg, M. Willershäuser, E. Wolf, A. Wolter, J. Wood, W. Wurst, A.Ö. Yildirim, R. Zeh, A. Zimmer, A. Zimprich, C. Holmes, K.P. Steel, Y. Herault, V. Gailus-Durner, A.-M. Mallon, S.D.M. Brown, Analysis of mammalian gene function through broad-based phenotypic screens across a consortium of mouse clinics, *Nat. Genet. Adv.* (2015), <https://doi.org/10.1038/ng.3360>.
- [10] V. Muñoz-Fuentes P. Cacheiro T.F. Meehan J.A. Aguilar-Pimentel S.D.M. Brown A. M. Flenniken P. Flicek A. Galli H.H. Mashhadi M. Hrabě de Angelis J.K. Kim K.C.K. Lloyd C. McKerie H. Morgan S.A. Murray L.M.J. Nutter P.T. Reilly J.R. Seavitt J.K. Seong M. Simon H. Wardle-Jones A.-M. Mallon D. Smedley H.E. Parkinson the I. consortium, The International Mouse Phenotyping Consortium (IMPC): a functional catalogue of the mammalian genome that informs conservation *Conserv. Genet.* 2018 1 11 10.1007/s10592-018-1072-9.
- [11] M.E. Dickinson, A.M. Flenniken, X. Ji, L. Teboul, M.D. Wong, J.K. White, T. F. Meehan, W.J. Weninger, H. Westerberg, H. Adissu, C.N. Baker, L. Bower, J. M. Brown, L.B. Caddle, F. Chiani, D. Clary, J. Cleak, M.J. Daly, J.M. Denegre, B. Doe, M.E. Dolan, S.M. Edie, H. Fuchs, V. Gailus-Durner, A. Galli, A. Gambadoro, J. Gallegos, S. Guo, N.R. Horner, C.-W. Hsu, S.J. Johnson, S. Kalaga, L.C. Keith, L. Lanoue, T.N. Lawson, M. Lek, M. Mark, S. Marschall, J. Mason, M.L. McElwee, S. Newbigging, L.M.J. Nutter, K.A. Peterson, R. Ramirez-Solis, D.J. Rowland, E. Ryder, K.E. Samochoa, J.R. Seavitt, M. Selloum, Z. Szoke-Kovacs, M. Tamura, A. G. Trainor, I. Tudose, S. Wakana, J. Warren, O. Wendling, D.B. West, L. Wong, A. Yoshiki, M. McKay, B. Urban, C. Lund, E. Froeter, T. LaCasse, A. Mehalow, E. Gordon, L.R. Donahue, R. Taft, P. Kutney, S. Dion, L. Goodwin, S. Kales, R. Urban, K. Palmer, F. Pertuy, D. Bitz, B. Weber, P. Goetz-Reiner, H. Jacobs, E. Le Marchand, A. El Amri, L. El Fertak, H. Ennah, D. Ali-Hadjji, A. Ayadi, M. Wattenhofer-Donze, S. Jacquot, P. André, M.-C. Birling, G. Pavlovic, T. Sorg, I. Morse, F. Benou, M.E. Stewart, C. Copley, J. Harrison, S. Joynson, R. Guo, D. Qu, S. Spring, L. Yu, J. Ellegood, L. Morikawa, X. Shang, P. Feugas, A. Creighton, P. Castellanos-Penton, O. Danisment, N. Griggs, C.L. Tudor, A.L. Green, C. Icoresi Mazzeo, E. Siragher, C. Lillistone, E. Tuck, D. Gleeson, D. Sethi, T. Bayzatinova, J. Burvill, B. Habib, L. Weavers, R. Maswood, E. Miklejewska, M. Woods, E. Grau, S. Newman, C. Sinclair, E. Brown, S. Ayabe, M. Iwama, A. Murakami, D. G. MacArthur, G.P. Tocchini-Valentini, X. Gao, P. Flicek, A. Bradley, W.C. Skarnes, M.J. Justice, H.E. Parkinson, M. Moore, S. Wells, R.E. Braun, K.L. Svenson, M.H. de Angelis, Y. Herault, T. Mohun, A.-M. Mallon, R.M. Henkelman, S.D.M. Brown, D. J. Adams, K.C.K. Lloyd, C. McKerie, A.L. Beaudet, M. Bučan, S.A. Murray, High-throughput discovery of novel developmental phenotypes, *Nature* 537 (2016) 508–514, <https://doi.org/10.1038/nature19356>.
- [12] L. Van Der Weyden, M.J. Arends, A.D. Campbell, T. Bald, H. Wardle-Jones, N. Griggs, M.D.C. Velasco-Herrera, T. Tütting, O.J. Sansom, N.A. Karp, S. Clare, D. Gleeson, E. Ryder, A. Galli, E. Tuck, E.L. Cambridge, T. Voet, I.C. MacAulay, K. Wong, S. Spiegel, A.O. Speak, D.J. Adams, Genome-wide in vivo screen identifies novel host regulators of metastatic colonization, *Nature* 541 (2017), <https://doi.org/10.1038/nature20792>.
- [13] N.A. Karp, J. Mason, A.L. Beaudet, Y. Benjamini, L. Bower, R.E. Braun, S.D.M. Brown, E.J. Chesler, M.E. Dickinson, A.M. Flenniken, H. Fuchs, M.H. De Angelis, X. Gao, S. Guo, S. Greenaway, R. Heller, Y. Herault, M.J. Justice, N. Kurbatova, C.J. Lelliott, K.C.K. Lloyd, A.M. Mallon, J.E. Mank, H. Masuya, C. McKerie, T.F. Meehan, R.F. Mott, S.A. Murray, H. Parkinson, R. Ramirez-Solis, L. Santos, J.R. Seavitt, D. Smedley, T. Sorg, A.O. Speak, K.P. Steel, K.L. Svenson, S. Wakana, D. West, S. Wells, H. Westerberg, S. Yaacoby, J.K. White, Prevalence of sexual dimorphism in mammalian phenotypic traits, *Nat. Commun.* 8 (2017) 1–12. <https://doi.org/10.1038/ncomms15475>.
- [14] M.R. Bowl M.M. Simon N.J. Ingham S. Greenaway L. Santos H. Cater S. Taylor J. Mason N. Kurbatova S. Pearson L.R. Bower D.A. Clary H. Meziane P. Reilly O. Minowa L. Kelsey S. Allen S. Clementson-Mobbs G. Codner M. Fray W. Gardiner R. Joynson J. Kenyon J. Loeffler B. Nell A. Parker D. Quwaillid M. Stewart A. Walling R. Zaman C.K. Chen N. Conte P. Matthews M. Relac I. Tudose J. Warren E. Le Marchand A. El Amri L. El Fertak H. Ennah D. Ali-Hadjji A. Ayadi M. Wattenhofer-Donze D. Moulaert S. Jacquot P. André M.C. Birling G. Pavlovic V. Lalanne A. Lux F. Riet C. Mittelhaeuser R. Bour A. Guimond C. Bam Hamed, S. Leblanc, L. Vasseur, M. Selloum, T. Sorg, S. Ayabe, T. Furuse, H. Kaneda, K. Kobayashi, H. Masuya, I. Miura, Y. Obata, T. Suzuki, M. Tamura, N. Tanaka, I. Yamada, A. Yoshiki, Z. Berberovic, M. Bubshait, J. Cabezas, T. Carroll, G. Clark, S. Clarke, A. Creighton, O. Danisment, M. Eskandarian, P. Feugas, M. Gertsenstein, R. Guo, J. Hunter, E. Jacob, Q. Lan, V. Laurin, N. Law, S. MacMaster, D. Miller, L. Morikawa, S. Newbigging, C. Owen, P. Penton, M. Pereira, D. Qu, X. Shang, G. Sleep, K. Sohel, S. Tondat, Y. Wang, I. Vukobradovic, Y. Zhu, F. Chiani, C. Di Pietro, G. Di Segni, O. Ermakova, F. Ferrara, P. Fruscoloni, A. Gambadoro, S. Gastaldi, E. Golini, G. La Sala, S. Mandillo, D. Marazziti, M. Massimi, R. Matteoni, T. Orsini, M. Pasquini, M. Raspa, A. Rauch, G. Rossi, N. Rossi, S. Putti, F. Scavizzi, G.D. Tocchini-Valentini, J. Beig, A. Bürger, F. Giesert, J. Graw, R. Kühn, O. Ortiz, J. Schick, C. Seisenberger, O. Amarie, L. Garrett, S.M. Hölter, A. Zimprich, A. Aguilar-Pimentel, J. Beckers, R. Brommager, J. Calzada-Wack, H. Fuchs, V. Gailus-Durner, C. Lengger, S. Leuchtberger, H. Maier, S. Marschall, K. Moreth, F. Neff, M.A. Östereicher, J. Rozman, R. Steinkamp, C. Stoeger, I. Treise, T. Stoeger, A.Ö. Yildirim, O. Eickelberg, L. Becker, T. Klopstock, M. Ollert, D.H. Busch, C. Schmidt-Weber, R. Bekeredjian, A. Zimmer, B. Rathkolb, E. Wolf, M. Klingenspor, G.P. Tocchini-Valentini, X. Gao, A. Bradley, W.C. Skarnes, M. Moore, A.L. Beaudet, M.J. Justice, J. Seavitt, M.E. Dickinson, W. Wurst, M.H. De Angelis, Y. Herault, S. Wakana, L.M. J. Nutter, A.M. Flenniken, C. McKerie, S.A. Murray, K.L. Svenson, R.E. Braun, D.B. West, K.C.K. Lloyd, D.J. Adams, J. White, N. Karp, P. Flicek, D. Smedley, T.F. Meehan, H.E. Parkinson, L.M. Teboul, S. Wells, K.P. Steel, A.M. Mallon, S.D.M. Brown, A large scale hearing loss screen reveals an extensive unexplored genetic landscape for auditory dysfunction *Nat. Commun.* 8 2017 10.1038/s41467-017-00595-4.

- [15] J. Rozman, B. Rathkolb, M.A. Oestereicher, C. Schütt, A.C. Ravindranath, S. Leuchtenberger, S. Sharma, M. Kistler, M. Willershäuser, R. Brommage, T. F. Meehan, J. Mason, H. Haselimirshadi, A. Aguilar-Pimentel, L. Becker, I. Treise, K. Moreth, L. Garrett, S.M. Höller, A. Zimprich, S. Marschall, O.V. Amarie, J. Calzada-Wack, F. Neff, L. Brachthäuser, C. Lengger, C. Stoeger, L. Zapf, Y.-L. Cho, P. Da Silva-Buttkus, M.J. Kraiger, P. Mayer-Kuckuk, K.K. Gampe, M. Wu, N. Conte, J. Warren, C.-K. Chen, I. Tudose, M. Relac, P. Matthews, H.L. Cater, H. P. Natukunda, J. Cleak, L.M. Teboul, S. Clementson-Mobbs, Z. Szoke-Kovacs, A. P. Walling, S.J. Johnson, G.F. Codner, T. Fiegel, N. Ring, H. Westerberg, S. Greenaway, D. Sneddon, H. Morgan, J. Loeffler, M.E. Stewart, R. Ramirez-Solis, A. Bradley, W.C. Skarnes, K.P. Steel, S.A. Maguire, J. Dench, D. Lafont, V. E. Vancollie, S.A. Pearson, A.S. Gates, M. Sanderson, C. Shannon, L.F.E. Anthony, M.T. Sumowski, R.S.B. McLaren, B. Doe, H. Wardle-Jones, M.N.D. Griffiths, A. Galli, A. Swiatkowska, C.M. Isherwood, A.O. Speak, E.L. Cambridge, H. M. Wilson, S.S. Caetano, A.K.B. Maguire, D.J. Adams, J. Bottomley, E. Ryder, D. Gleeson, L. Pouilly, S. Rousseau, A. Auburtin, P. Reilly, A. Ayadi, M. Selloum, J. A. Wood, D. Clary, P. Havel, T. Tolentino, H. Tolentino, M. Schuchbauer, S. Pedroia, A. Trainor, E. Djan, M. Pham, A. Huynh, V. De Vera, J. Seavitt, J. Gallegos, A. Garza, E. Mangin, J. Senderstrom, I. Lazo, K. Mowrey, R. Bohat, R. Samaco, S. Veeraragavan, C. Beeton, S. Kalaga, L. Kelsey, I. Vukobradovic, Z. Berberovic, C. Owen, D. Qu, R. Guo, S. Newbigging, L. Morikawa, N. Law, X. Shang, P. Feugas, Y. Wang, M. Eskandarian, Y. Zhu, P. Penton, V. Laurin, S. Clarke, Q. Lan, G. Sleep, A. Creighton, E. Jacob, O. Danisment, M. Gertsenstein, M. Pereira, S. MacMaster, S. Tondat, T. Carroll, J. Cabezas, J. Hunter, G. Clark, M. Bubshait, D. Miller, K. Sohel, H. Adissu, M. Ganguly, A. Bezginov, F. Chiani, C. Di Pietro, G. Di Segni, O. Ermakova, F. Ferrara, P. Fruscoloni, A. Gambadoro, S. Gastaldi, E. Golini, G. La Sala, S. Mandillo, D. Marazziti, M. Massimi, R. Matteoni, T. Orsini, M. Pasquini, M. Raspa, A. Rauch, G. Rossi, N. Rossi, S. Putti, F. Scavizzi, G.D. Tocchini-Valentini, S. Wakana, T. Suzuki, M. Tamura, H. Kaneda, T. Furuse, K. Kobayashi, I. Miura, I. Yamada, Y. Obata, A. Yoshiki, S. Ayabe, J. N. Chambers, K. Chalupsky, C. Seisenberger, A. Bürger, J. Beig, R. Kühn, A. Hörlein, J. Schick, O. Ortiz, F. Giesert, J. Graw, M. Ollert, C. Schmidt-Weber, T. Stoeger, A. Önder Yildirim, O. Eickelberg, T. Klopstock, D.H. Busch, R. Bekerdjian, A. Zimmer, J.O. Jacobsen, D. Smedley, M.E. Dickinson, F. Benso, I. Morse, H.-C. Kim, H. Lee, S.Y. Cho, T. Hough, A.-M. Mallon, S. Wells, L. Santos, C.J. Lelliott, J.K. White, T. Sorg, M.-F. Champy, L.R. Bower, C.L. Reynolds, A. M. Flenniken, S.A. Murray, L.M.J. Nutter, K.L. Svenson, D. West, G.P. Tocchini-Valentini, A.L. Beaudet, F. Bosch, R.B. Braun, M.S. Dobbie, X. Gao, Y. Hérault, A. Moshiri, B.A. Moore, K.C. Kent Lloyd, C. McKerlie, H. Masuya, N. Tanaka, P. Flicek, H.E. Parkinson, R. Sedlacek, J.K. Seong, C.-K.L. Wang, M. Moore, S. D. Brown, M.H. Tschöp, W. Wurst, M. Klingenspor, E. Wolf, J. Beckers, F. MacHicao, A. Peter, H. Staiger, H.-U. Häring, H. Grallert, M. Campillos, H. Maier, H. Fuchs, V. Gailus-Durner, T. Werner, M.H. De Angelis, Identification of genetic elements in metabolism by high-throughput mouse phenotyping, *Nat. Commun.* 9 (2018), <https://doi.org/10.1038/s41467-017-01995-2>.
- [16] V. Perez-Garcia, E. Fineberg, R. Wilson, A. Murray, C.I. Mazzeo, C. Tudor, A. Sienrath, J.K. White, E. Tuck, E.J. Ryder, D. Gleeson, E. Siragher, H. Wardle-Jones, N. Staudt, N. Wali, J. Collins, S. Geyer, E.M. Busch-Nentwich, A. Galli, J. C. Smith, E. Robertson, D.J. Adams, W.J. Weninger, T. Mohun, M. Hemberger, Placentation defects are highly prevalent in embryonic lethal mouse mutants, *Nature*. 555 (2018), <https://doi.org/10.1038/nature26002>.
- [17] J.E. Collins, R.J. White, N. Staudt, I.M. Sealy, I. Packham, N. Wali, C. Tudor, C. Mazzeo, A. Green, E. Siragher, E. Ryder, J.K. White, I. Papatheodorou, A. Tang, A. Füllgrabe, K. Billis, S.H. Geyer, W.J. Weninger, A. Galli, M. Hemberger, D. L. Stemple, E. Robertson, J.C. Smith, T. Mohun, D.J. Adams, E.M. Busch-Nentwich, Common and distinct transcriptional signatures of mammalian embryonic lethality, *Nat. Commun.* 10 (2019) 2792, <https://doi.org/10.1038/s41467-019-10642-x>.
- [18] M. Raess A.A. de Castro V. Gailus-Durner S. Fessele M. Hrabě de Angelis, the I. Consortium, INFRAFRONTIER: a European resource for studying the functional basis of human disease *Mamm. Genome*. 2016 1 6 10.1007/s00335-016-9642-y.
- [19] K. Lloyd, C. Franklin, C. Lutz, T. Magnuson, Reproducibility: Use mouse biobanks or lose them, *Nature*. 522 (2015) 151–153, <https://doi.org/10.1038/522151a>.
- [20] K.C. Kent Lloyd, P.N. Robinson, C.A. MacRae, Animal-based studies will be essential for precision medicine, *Sci. Transl. Med.* 8 (2016).
- [21] K.C.K. Lloyd, D.J. Adams, G. Baynam, A.L. Beaudet, F. Bosch, K.M. Boycott, R. E. Braun, M. Caulfield, R. Cohn, M.E. Dickinson, M.S. Dobbie, A.M. Flenniken, P. Flicek, S. Galande, X. Gao, A. Grobler, J.D. Heaney, Y. Hérault, M.H. de Angelis, J.R. Lupski, S. Lyonnet, A.-M. Mallon, F. Mammiano, C.A. MacRae, R. McInnes, C. McKerlie, T.F. Meehan, S.A. Murray, L.M.J. Nutter, Y. Obata, H. Parkinson, M. S. Pepper, R. Sedlacek, J.K. Seong, T. Shiroishi, D. Smedley, G. Tocchini-Valentini, D. Valle, C.-K.L. Wang, S. Wells, J. White, W. Wurst, Y. Xu, S.D.M. Brown, The deep genome project, *Genome Biol.* 21 (2020) 18, <https://doi.org/10.1186/s13059-020-1931-9>.
- [22] K. Boroviak, B. Doe, R. Banerjee, F. Yang, A. Bradley, Chromosome engineering in zygotes with CRISPR/Cas9, *Genesis*. (2016), <https://doi.org/10.1002/dvg.22915>.
- [23] M.-C. Birling, L. Schaeffer, P. André, L. Lindner, D. Maréchal, A. Ayadi, T. Sorg, G. Pavlovic, Y. Hérault, Efficient and rapid generation of large genomic variants in rats and mice using CRISMER, *Sci. Rep.* 7 (2017) 43331, <https://doi.org/10.1038/srep43331>.
- [24] M. Kosicki, K. Tomberg, A. Bradley, Repair of double-strand breaks induced by CRISPR–Cas9 leads to large deletions and complex rearrangements, *Nat. Biotechnol.* (2018), <https://doi.org/10.1038/nbt.4192>.
- [25] W.C. Skarnes, B. Rosen, A.P. West, M. Koutsourakis, W. Bushell, V. Iyer, A. O. Mujica, M. Thomas, J. Harrow, T. Cox, D. Jackson, J. Severin, P. Biggs, J. Fu, M. Nefedov, P.J. de Jong, A.F. Stewart, A. Bradley, A conditional knockout resource for the genome-wide study of mouse gene function, *Nature* 474 (2011) 337–342, <https://doi.org/10.1038/nature10163>.
- [26] H. Wang, H. Yang, C.S. Shivalila, M.M. Dawlaty, A.W. Cheng, F. Zhang, R. Jaenisch, One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering, *Cell* 153 (2013) 910–918, <https://doi.org/10.1016/j.cell.2013.04.025>.
- [27] K.A. Peterson, G.L. Beane, L.O. Goodwin, P.M. Kutny, L.G. Reinholdt, S.A. Murray, CRISPRtools: a flexible computational platform for performing CRISPR/Cas9 experiments in the mouse, *Mamm. Genome* (2017) 1–8, <https://doi.org/10.1007/s00335-017-9681-z>.
- [28] S. Jacquot, N. Chartoire, F. Piguet, Y. Hérault, G. Pavlovic, Optimizing PCR for mouse genotyping: recommendations for reliable, rapid, cost effective, robust and adaptable to high-throughput genotyping protocol for any type of mutation, *Curr. Protoc. Mouse Biol.* 9 (2019), e65, <https://doi.org/10.1002/cpmo.65>.
- [29] D.M. Valenzuela, A.J. Murphy, D. Frenthewy, N.W. Gale, A.N. Economides, W. Auerbach, W.T. Poueymirou, N.C. Adams, J. Rojas, J. Yasenchak, R. Chernomorsky, M. Boucher, A.L. Elsassler, L. Esau, J. Zheng, J.A. Griffiths, X. Wang, H. Su, Y. Xue, M.G. Dominguez, I. Noguera, R. Torres, L.E. Macdonald, A. F. Stewart, T.M. DeChiara, G.D. Yancopoulos, High-throughput engineering of the mouse genome coupled with high-resolution expression analysis, *Nat. Biotechnol.* 21 (2003) 652–659, <https://doi.org/10.1038/nbt822>.
- [30] E. Ryder, D. Gleeson, D. Sethi, S. Vyas, E. Miklejewska, P. Dalvi, B. Habib, R. Cook, M. Hardy, K. Jhaveri, J. Bottomley, H. Wardle-Jones, J.N. Bussell, R. Houghton, J. Salisbury, W.C. Skarnes, R. Ramirez-Solis, Molecular characterization of mutant mouse strains generated from the EUCOMM/KOMP-CSD ES cell resource, *Mamm. Genome* 24 (2013) 286–294, <https://doi.org/10.1007/s00335-013-9467-x>.
- [31] U. Majeed, E. Darwish, S.U. Rehman, X. Zhang, Competitive allele specific PCR (KASP): a singleplex genotyping platform and its application, *J. Agric. Sci.* 11 (2018) 11, <https://doi.org/10.5539/jas.v11n1p11>.
- [32] N. Thomsen, R.G. Ali, J.N. Ahmed, R.M. Arkel, High resolution melt analysis (HRMA): a viable alternative to agarose gel electrophoresis for mouse genotyping, *PLoS One* 7 (2012), <https://doi.org/10.1371/journal.pone.0045252>.
- [33] M.K. Ramlee, T. Yan, A.M.S. Cheung, C.T.H. Chuah, S. Li, High-throughput genotyping of CRISPR/Cas9-mediated mutants using fluorescent PCR-capillary gel electrophoresis, *Sci. Rep.* 5 (2015) 15587, <https://doi.org/10.1038/srep15587>.
- [34] I. Brocal, R.J. White, C.M. Dooley, S.N. Carruthers, R. Clark, A. Hall, E.M. Busch-Nentwich, D.L. Stemple, R.N.W. Kettleborough, Efficient identification of CRISPR/Cas9-induced insertions/deletions by direct germline screening in zebrafish, *BMC Genomics* 17 (2016) 259, <https://doi.org/10.1186/s12864-016-2563-z>.
- [35] J. Kreuger, P. O'Callaghan, Failure to genotype: a cautionary note on an elusive loop sequence, *PLoS One* 11 (2016), e0165012, <https://doi.org/10.1371/journal.pone.0165012>.
- [36] M.-C. Birling, A. Dierich, S. Jacquot, Y. Hérault, G. Pavlovic, Highly-efficient, fluorescent, locus directed Cre and flopo deleter mice on a pure C57BL/6N genetic background, *Genesis* 50 (2011) 482–489, <https://doi.org/10.1002/dvg.20826>.
- [37] G.F. Codner, L. Lindner, A. Caulder, M. Wattenhofer-Donzé, A. Radage, A. Mertz, B. Eisenmann, J. Mianné, E.P. Evans, C.V. Beechey, M.D. Fray, M.C. Birling, Y. Hérault, G. Pavlovic, L. Teboul, Aneuploidy screening of embryonic stem cell clones by metaphase karyotyping and droplet digital polymerase chain reaction, *BMC Cell Biol.* 17 (2016) 30, <https://doi.org/10.1186/s12860-016-0108-6>.
- [38] J. Mianné, G.F. Codner, A. Caulder, R. Fell, M. Hutchison, R. King, M.E. Stewart, S. Wells, L. Teboul, Analysing the outcome of CRISPR-aided genome editing in embryos: Screening, genotyping and quality control, *Methods* 121–122 (2017) 68–76, <https://doi.org/10.1016/j.jymeth.2017.03.016>.
- [39] S.-T. Yen, M. Zhang, J.M. Deng, S.J. Usman, C.N. Smith, J. Parker-Thornburg, P. G. Swinton, J.F. Martin, R.R. Behringer, Somatic mosaicism and allele complexity induced by CRISPR/Cas9 RNA injections in mouse zygotes, *Dev. Biol.* 393 (2014) 3–9, <https://doi.org/10.1016/j.ydbio.2014.06.017>.
- [40] A.D. Yates, P. Achuthan, W. Akanni, J. Allen, J. Allen, J. Alvarez-Jarreta, M.R. Amodé, I.M. Armean, A.G. Azov, R. Bennett, J. Bhai, K. Billis, S. Boddu, J.C. Marugán, C. Cummins, C. Davidson, K. Dodiya, R. Fatima, A. Gall, C.G. Giron, L. Gil, T. Grego, L. Haggerty, E. Haskell, T. Hourlier, O.G. Izuogu, S.H. Janacek, T. Juettemann, M. Kay, I. Lavidas, T. Le, D. Lemos, J.G. Martuzio, T. Maurel, A. McDowall, A. McMahon, S. Mohanan, B. Moore, M. Nuhn, D.N. Oheh, A. Parker, A. Parton, M. Patricio, M.P. Sakhthivel, A.I. Abdul Salam, B.M. Schmitt, H. Schuilenburg, D. Sheppard, M. Sycheva, M. Szuba, K. Taylor, A. Thormann, G. Threadgold, A. Vullo, B. Waits, A. Winterbottom, A. Zadissa, M. Chakiachvili, B. Flint, A. Frankish, S.E. Hunt, G. Isles, M. Kostadima, N. Langridge, J.E. Loveland, F.J. Martin, J. Morales, J.M. Mudge, M. Muffato, E. Perry, M. Ruffier, S.J. Trevanion, F. Cunningham, K.L. Howe, D.R. Zerbino, P. Flicek, *Ensembl* 2020, *Nucleic Acids Res.* 48 (2019) D682–D688, <https://doi.org/10.1093/nar/gkz966>.
- [41] A.S. McLellan, B. Fischer, G. Dveksler, T. Hori, F. Wynne, M. Ball, K. Okumura, T. Moore, W. Zimmermann, Structure and evolution of the mouse pregnancy-specific glycoprotein (Psg) gene locus, *BMC Genomics* 6 (2005) 4, <https://doi.org/10.1186/1471-2164-6-4>.
- [42] T. Moore, G.S. Dveksler, Pregnancy-specific glycoproteins: complex gene families regulating maternal-fetal interactions, *Int. J. Dev. Biol.* 58 (2014) 273–280, <https://doi.org/10.1387/jidb.130329gd>.
- [43] B. Yalcin, D.J. Adams, J. Flint, T.M. Keane, Next-generation sequencing of experimental mouse strains, *Mamm. Genome* 23 (2012) 490–498, <https://doi.org/10.1007/s00335-012-9402-6>.
- [44] H. Fairfield, A. Srivastava, G. Ananda, R. Liu, M. Kircher, A. Lakshminarayana, B. S. Harris, S.Y. Karst, L.A. Dionne, C.C. Kane, M. Curtain, M.L. Berry, P.F. Ward-Bailey, I. Greenstein, C. Byers, A. Czechanski, J. Sharp, K. Palmer, P. Gudis, W. Martin, A. Tadenev, L. Bogdanik, C.H. Pratt, B. Chang, D.G. Schroeder, G. A. Cox, P. Cliften, J. Milbrandt, S. Murray, R. Burgess, D.E. Bergstrom, L.

- R. Donahue, H. Hamamy, A. Masri, F.A. Santoni, P. Makrythanasis, S. E. Antonarakis, J. Shendure, L.G. Reinholdt, Exome sequencing reveals pathogenic mutations in 91 strains of mice with Mendelian disorders, *Genome Res.* 25 (2015) 948–957, <https://doi.org/10.1101/gr.186882.114>.
- [45] J. Shin, S. Lee, M.J. Go, S.Y. Lee, S.C. Kim, C.H. Lee, B.K. Cho, Analysis of the mouse gut microbiome using full-length 16S rRNA amplicon sequencing, *Sci. Rep.* 6 (2016) 1–10, <https://doi.org/10.1038/srep29681>.
- [46] Y. Shinmyo, S. Tanaka, S. Tsunoda, K. Hosomichi, A. Tajima, H. Kawasaki, CRISPR/Cas9-mediated gene knockout in the mouse brain using in utero electroporation, *Sci. Rep.* 6 (2016) 1–13, <https://doi.org/10.1038/srep20611>.
- [47] E. Ryder, B. Doe, D. Gleeson, R. Houghton, P. Dalvi, E. Grau, B. Habib, E. Miklejewska, S. Newman, D. Sethi, C. Sinclair, S. Vyas, H. Wardle-Jones, J. Bottomley, J. Bussell, A. Galli, J. Salisbury, R. Ramirez-Solis, Rapid conversion of EUCOMM/KOMP-CSD alleles in mouse embryos using a cell-permeable Cre recombinase, *Transgenic Res.* 23 (2013) 177–185, <https://doi.org/10.1007/s11248-013-9764-x>.
- [48] P. Cruaud, J.-Y. Rasplus, L.J. Rodriguez, A. Cruaud, High-throughput sequencing of multiple amplicons for barcoding and integrative taxonomy, *Sci. Rep.* 7 (2017) 41948, <https://doi.org/10.1038/srep41948>.
- [49] M. Bruntraeger, M. Byrne, K. Long, A.R. Bassett, Editing the genome of human induced pluripotent stem cells using CRISPR/Cas9 ribonucleoprotein complexes, *Methods Mol. Biol.* (2019) 153–183, https://doi.org/10.1007/978-1-4939-9170-9_11.
- [50] V. Iyer, K. Boroviak, M. Thomas, B. Doe, L. Riva, E. Ryder, D.J. Adams, No unexpected CRISPR-Cas9 off-target activity revealed by trio sequencing of gene-edited mice, *PLoS Genet.* 14 (2018), e1007503, <https://doi.org/10.1371/journal.pgen.1007503>.
- [51] M.A. Quail, H. Swerdlow, D.J. Turner, H. Swerdlow, Improved protocols for the illumina genome analyzer sequencing system., *Curr. Protoc. Hum. Genet.* Chapter 18 (2009) Unit 18.2. <https://doi.org/10.1002/0471142905.hg1802s62>.
- [52] K.J. Livak, T.D. Schmittgen, Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method, *Methods* 25 (2001) 402–408, <https://doi.org/10.1006/meth.2001.1262>.
- [53] K. Clement, H. Rees, M.C. Canver, J.M. Gehrke, R. Farouni, J.Y. Hsu, M.A. Cole, D. R. Liu, J.K. Joung, D.E. Bauer, L. Pinello, CRISPResso2 provides accurate and rapid genome editing sequence analysis, *Nat. Biotechnol.* 37 (2019) 224–226, <https://doi.org/10.1038/s41587-019-0032-3>.
- [54] B. Langmead, S.L. Salzberg, Fast gapped-read alignment with Bowtie 2, *Nat. Methods* 9 (2012) 357–359, <https://doi.org/10.1038/nmeth.1923>.
- [55] M. Costello, M. Fleharty, J. Abreu, Y. Farjoun, S. Ferreira, L. Holmes, B. Granger, L. Green, T. Howd, T. Mason, G. Vicente, M. Dasilva, W. Brodeur, T. DeSmet, S. Dodge, N.J. Lennon, S. Gabriel, Characterization and remediation of sample index swaps by non-redundant dual indexing on massively parallel sequencing platforms, *BMC Genomics* 19 (2018) 332, <https://doi.org/10.1186/s12864-018-4703-0>.