Chapter 2

Materials and Methods

2.1 Cell lines and cell culture

JM8A3 mouse ES cell line was derived from a C57BL/6N blastocyst (Pettitt et al., 2009). CB9, BC2 and BC8 mouse ES cell lines were derived from F1 cross between C57BL/6N and CAST/EiJ mice (Strogantsev et al., 2015), a gift from A.F. Smith. JBG7 and CBA9 are Cas9-expressing single cell clones derived from JM8 or CB9 cells, respectively. Cas9 was introduced by stable transduction using a Cas9-2A-Blast lentiviral construct (in pKLV2 backbone, see Vectors section) at a low titre to ensure single copy integration (<0.1% transduction rate). Human HEK293 cell line and its subclone expressing Cas9 from the same lentiviral Cas9-2A-Blast construct were single-cell cloned and their karyotype was verified (a gift from E. Metzakopian). hTERT RPE1, *trp*53-/- cell line expressing Cas9 was obtained from Steve Jackson's group. AB2.2 mCherry/GFP reporter cells were a gift from Dr. Xiufei Gao and Prof. Pentao Liu. 293T cells for lentivirus production were obtained from Ao Zhou. Virus was obtained by lipofectamine LTX mediated transfection of 293T cells with ViraPower Lentiviral Packaging Mix (Thermo Scientific) and the Cas9-2A-Blast construct, following manufacturers' instructions.

All ES cell lines were cultured in M15 media (High-Glucose DMEM, with 15% FSC, β-mercaptoethanol and L-Glutamate, Gibco) on sublethally irradiated feeder cells. Feeders were derived from SNL76/6 cell line (expressing neomycin resistance and LIF cassette, Ramírez-Solis et al., 1993) by transgenic insertion of a resistance cassette (blasticidin or puromycin). HEK cells were cultured in M15 and RPE1 cell lines were cultured in M10 (High-Glucose DMEM with 10% FSC).

2.2 Vectors

Vectors for expression of gRNAs contained a U6 promoter with a „F+E" scaffold (reported to mediate higher levels of mutagenesis than the standard scaffold, Hsu et al., 2013a) and a Puro-2A-BFP cassette driven by PGK promoter. Constructs were flanked by PiggyBac repeats (PBCV backbone from Mathias Friedrich), lentiviral repeats (pKLV1 backbone from K.Yusa, Koike-Yusa et al., 2014) or both PiggyBac and lentiviral repeat elements (pKLV2 backbone from E. Metzakopian, Metzakopian et al., 2017). Cas9-expression vectors contained a truncated EF1α (EFS) promoter driving a Cas9-2A-Blast cassette in a pKLV2 backbone. Hyperactive PiggyBac transposase was driven by CMV promoter (Yusa et al., 2011b). See vector schematics in Fig. 2.1. Vectors were amplified in NEB10β E.coli strain (Thermo Scientific) under Ampicillin selection and purified using Machelerey-Nagel plasmid extraction kits. gRNAs were cloned into BbsI digested backbones using DNA Ligation Kit V.1 (Takara). Subcloned plasmids were Sanger sequenced at Eurofins or GATC.

2.3 Transfections, flow cytometry and sequencing

Transfections took place in 24W plates coated with gelatin. About 300,000 wild-type mouse ES cells were "reverse" transfected with 2.5 µl

lipofectamine LTX, 0.5 µl plus reagent (Thermo Scientific), 200 ng hyperactive PiggyBac transposase (Yusa et al., 2011b), 100 ng of the pKLV2-PiggyBac Cas9-Blast plasmid and 50 ng of the PBCV-gRNA-Puro plasmid in 50 µl OptiMEM following manufacturer's instructions. For Cas9-expressing mouse ES cells, 50 ng hyperactive PiggyBac transposase and 150 ng of the PiggyBac gRNA-Puro plasmid were used. A similar setup was used for lipofection of 20 pmol of hybridized crRNA:trRNA (Sigma) and 20 pmol of EnGen Cas9 NLS (NEB), except plus reagent was omitted. Hybridization was performed by warming up mixed crRNA and trRNA to 95_oC and letting it cool down at room temperature. Neon Transfection System (Thermo Fisher Scientific; 1600 v / 10 ms / 3 pulses) was used for electroporation of 150,000 mouse ES cells in buffer R with 6 pmol each of crRNA:trRNA, electroporation enhancer (IDT) and Cas9 protein or 9 pmol each of crRNA:trRNA and Cas9 protein. Cells were cultured in M15 media supplemented with LIF to maintain pluripotence (Williams et al., 1988). Stable integration of Cas9 and gRNA-expressing constructs was selected for using blasticidin (10 µg/ml) and puromycin (3 µg/ml), respectively. The drugs were added on day 2 and cells were maintained in selective media for the duration of the experiment. Cells were split 1:1 or 1:2 on day 5 (depending on confluency) after 10-30' incubation with trypsin, 1:4 on day 7 and 1:6 from then on, any time they were nearing confluency.

For flow cytometric analysis, around 300,000 cells (1/6 of a near-confluent well) were collected by trypsinization, transfered to a U-bottomed 96W plate, washed once and stained for 15-60' in 50 µl buffer with 1 µg/ml FLAER reagent (Cedarlane) or 1:200 anti-Cd9-PE antibody (cat 124805, Biolegend). After staining, cells were washed three times and analysed using a Cytoflex flow cytometer. All procedures were performed at room temperature. PBS+0.1% BSA buffer was used throughout. All centrifugations were performed for 1 min at 500 G.

FACS sorting was performed on day 14 using MoFlow XDP (Beckman Coulter) or SH800 (Sony). Cells were plated at a limiting dilution of 500-2000 cells per 10 cm feeder plate (yielding around 100-400 colonies) in M15 supplied with Penicilin/Streptomycin and colonies were picked 7-10 days later in 96W feeder plates. Genomic DNA was extracted from grown colonies by overnight digestion at 56°C using a lysis buffer supplied with 1 µg/ml proteinase K (100 mM Tris pH 8.5, 5 mM EDTA pH 8.0, 0.2% SDS, 200 mM NaCl) followed by precipitation using 100% ethanol with 75 mM NaCl and three washes with 70% ethanol. DNA was resuspended in 200 µl T0.1E buffer (10 mM Tris-HCl with 0.1 mM EDTA). PCR amplification was performed using LongAMP or Q5 polymerase (NEB) following manufacturer's instructions. The products were resolved on an agarose gel (2% for primer pairs spanning <1.5 kb, 0.8% otherwise) and stained using ethidium bromide. If multiple

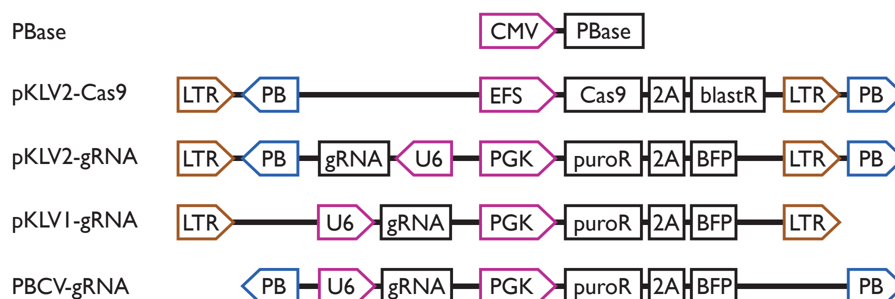

Figure 2.1: Vector schematics. LTR = Long Terminal Repeats, lentiviral elements; PB = PiggyBac repeats; blastR and puroR = blasticidin and puromycin resistance cassettes; LTRs, PBs and promoters (EFS, CMV and PGK PolII promoters and U6 PolIII promoter) are maked with colors.

products were present, they were individually cut out of the gel and purified using QIAquick Gel Extraction Kit (Qiagen). If only one product was obtained, the PCR reaction was purified using AMPure XP magnetic beads (1:1 ratio). Products were Sanger sequenced at Eurofins or GATC.

Similar procedures as for mouse ES cells were used with RPE1-Cas9 cell line, with following exceptions. For flow cytometry experiments, cells were transfected with 50 ng pKLV2-gRNA-Puro and 200 ng hyperactive PiggyBac transposase and selected with puromycin (3 µg/ml) from day 2 till the end of the experiment. For FACS, cells were transfected transiently with 500 ng pKLV1-gRNA-Puro plasmid and selected with puromycin on days 1-3. Instead of limiting dilution plating, RPE1 cells were single cell sorted on day 17 into a 96W plate with M10 media supplied with Penicilin/Streptomycin. Plating efficiency was around 10-20% on day 17 after sorting.

Bone marrow cells from a homozygous C57BL/6N CAS9-EGFP knock-in mouse (Platt et al., 2014) were isolated by flushing tibias and femurs in Hank's Balanced Salt Solution (Life Technologies) supplemented with 2% Fetal Bovine Serum (FBS) and 10mM HEPES (Sigma). Lineage negative cells were isolated using Direct Lineage Cell Depletion Kit Mouse (Miltenyl Biotec). After isolation and before sorting, cells were cultured in X-Vivo (Lonza) with 2% FBS, 50 ng/ml stem cell factor, 50 ng/ml thrombopoietin, 10 ng/ml IL-6 (Peprotech). Following a 3 h initial culture, 100'000 cells were electroporated (1550 v / 20 ms / 1 pulse) in buffer T with 44 pmols of crRNA:trRNA (IDT). On day 4 they were stained and sorted as described above. Single cell cloning was performed in Methocult M3434 media (6000 cells per 3 ml, StemCell Technologies) and colonies were picked 7-10 days later into 25 µl of direct PCR lysis buffer (Peqlab).

2.4 Comparison of delivery methods in HEK cells

For the RNP-electroporation condition, 150,000 HEK cells were electroporated (1700 V, 20 ms, 1 pulse) with 10 pmol Cas9 protein and 10 pmol hybridized crRNA:trRNA. For other methods, cells were seeded in 24W plates the day before the transfection so as to achieve 50%–70% confluency. For PiggyBac and transient plasmid conditions, cells were then transfected with 150 ng gRNA plasmid, 150 ng Cas9 plasmid, and 50 ng of either hyperactive PiggyBac transposase or carrier plasmid (pBluescript II SK+). For the protein + plasmid (P&P) and protein + plasmid + carrier (P&P-carrier) conditions, cells were transfected with two separately prepared mixes: (1) 3 pmol Cas9 protein, (2) 150 ng gRNA plasmid with or without 200 ng carrier plasmid (pBluescript II SK+). For RNP-lipofectamine conditions, cells were transfected with 3 pmol Cas9 protein, 3 pmol hybridized crRNA:trRNA (regular or stabilized). Plus reagent was added at 1 µL per 1 µg plasmid and Lipofectamine 3000 at double that volume. Cas9 protein was mixed with 1.5 µL Lipofectamine 3000. Cells were collected at indicated timepoints using trypsin, assessed for transfection effciency by flow cytometry and cell pellets were frozen for genomic DNA extraction.

Profiling of indels using IDAA procedure was performed according to the published protocol (Lonowski et al., 2017). In short, genomic DNA was extracted, a ~350 bp region around the cut site was amplified and tagged with a fluorescent dye using TEMPase Hot Start DNA polymerase (Ampliqon) and the products were resolved using a sequenator, yielding the indel profile.

2.5 PacBio sequencing and analysis

PCR amplification was performed using Q5 (NEB) or HiFi Hotstart ReadyMix (Kapa Biosciences). First 25 cycles used genomic primers with an adapter overhang (forward: GATGTACAGAGTGATATTATTGACACGCCC, reverse: CCAGGGGGATCACCATCCGTCGCCC or forward and reverse: CGACTCGCTACCAATGAAGACAGC). Products were purified using AMPure XP magnetic beads. One tenth of the eluate was used in a secondary 6 cycle PCR reaction to add recommended PacBio barcodes. For

PigA, these corresponded to different gRNAs and protein expression levels, whereas for *Cd9* they corresponded to single cell clones. Products were pooled equimolarly, prepared for sequencing by ligation of "SMRTbell" adapters by the Bespoke Sequencing Team (Wellcome Sanger Institute) and sequenced on the RSII instrument.

Analysis of PacBio data was performed using command line version of SMRT-Link software (pbtranscript 1.0.1.TAG-1470). For the purpose of calculating *PigA* locus coverage, a circular consensus sequences (CCS) were derived from multiple read-throughs of the same DNA molecule using "ccs --minPasses=1 --minPredictedAccuracy=0.9". Genome coverage was calculated with "bedtools genomecov –dz" (v 2.27.1) using CCS and visualized using ggplot2.

Individual *PigA* and *Cd9* alleles were reconstructed using Iso Seq workflow. In short, CCS were called using "ccs --minPasses=0 --minPredictedAccuracy=0.8" and classified into full length non chimeric ("FLNC", with both primer binding sites detected) and non full length ("NFL") reads using "classify" command. FL reads were also split by barcode, separating single cell clones (*Cd9*) or split into bins of 1 kb size using "separate_flnc" command (*PigA*). Iterative Clustering and Error correction (ICE) was peformed on each group (clone or size bin) individually using "cluster --targeted_isoseq" command. A custom script rebuilding the mapping index on each iteration of the clustering was used to fix a programming bug. Resulting "high quality" alleles (as classified by the clustering script) were mapped to the reference genome using "bwa mem" (v 0.7.17-r1188). Downstream analysis was performed using custom R (v 3.3.2) and bash scripts. For the *PigA* locus, reads were clustered furthered based on mapping and alleles with less than four FL reads support were filtered out. For the *Cd9* locus, additional filters based on FL to NFL ratio and within clone abundance were applied. Remaining *Cd9* alleles were visually inspected and ambiguities were resolved by Sanger sequencing. Additional alleles were discovered by custom PCRs (to detect larger deletions, large insertions and small indels) and Sanger sequenced. Lesions both smaller than 6 bp and farther than 20 bp from the cut site, as well as lesions in low complexity regions were removed from *Cd9* alleles.

2.6 Bioinformatics

Analysis of IDAA experiments was performed in R using binner package (https://github.com/plantarum/binner). Efficiency score was calculated as 1–(wild-type peak intensity/sum of wild-type and prominent peaks intensities). Spurious "–1 bp" signal present in wild-type samples was estimated to be around 10% of wild-type peak. This intensity was subtracted from "–1 bp" peak and added to the wild-type peak in all samples. Prominent peaks were defined using an arbitrary cutoff on the sum of intensities over many experiments.

Approximately 25 bp long primers with melting temperature of 60°C were designed using Primer3 or Primer3-BLAST. Guide RNAs were designed using Benchling and CRISPRscan (Moreno-Mateos et al., 2015), each guide being mismatched on at least two positions to any predicted off-target site. Flow cytometric data were processed with FlowJo (v 10.4.1). Mixed Sanger traces were resolved using the online tool PolyPeakParser (Hill et al., 2014). For visualization purposes, alignment of alleles (whether derived from PacBio or Sanger sequencing) was performed using BLAT (v 35, with settings -tileSize=6 -minScore=50 -minIdentity=90) and converted into BAM format using a customized script from Tobias Marschall (https://github.com/ALLBio/allbiotc2/tree/master/synthetic-benchmark). All visualizations were made in R using ggplot2 package.