Chapter 4

Detection and quantification of complex lesions

4.1 Introduction

4.1.1 Mutation reporters

Basic research in genetics, genome engineering and DNA damage repair uses simple assays, which allow isolation and quantification of cells with particular mutagenic events. The Ames test used to measure the mutagenicity of chemicals is one example (Ames, 1979). In this assay, bacteria harboring a mutation that makes them histidine dependent are exposed to potentially mutagenic compounds, which causes some of them to become histidine independent. The frequency of this reversion can be used as a measure of the mutagenicity of the compound. Furthermore, surviving bacteria can be analysed to establish the exact genetic cause of the reversion, which can be the restoration of the original genotype or a compensating mutation elsewhere in the gene (e.g. restoration of the reading frame). Variations on this assay continue to be used in genetics and toxicology research.

Guided by a similar, selective principle, the first systematic investigation of gene targeting used the endogenous *Hprt* gene and an exogenous neomycin resistance cassette (neoR) to isolate targeted cells (Thomas and Capecchi, 1987). *Hprt* plays a central role in the purine salvage pathway and is dispensable for viability of cultured cells under normal conditions. Hprt-proficient cells can be isolated by using hypoxanthine-aminopterin-thymidine (HAT) medium, which kills cells unable to salvage the necessary nucleotides. Conversely, 6-thioguanine (6-TG) kills Hprt-proficient cells, which convert it to a toxic product. The researchers transfected cells with an Hrpt-targeting construct containing a neoR cassette and isolated correctly targeted cells by selecting for loss of Hprt expression and gain of neomycin resistance. Varying the concentration of the reagents, the transfection conditions and the parameters of the vectors itself (like the length of homology arms) allowed optimization of the experimental protocol. In addition to basic research on gene targeting, *Hprt* is commonly used as a safe locus for insertion of transgenes and as cassette for positive and negative selection (Conway et al., 2014; van der Lugt et al., 1991). NeoR split into two parts, which can recombine following I-SceI induced DSB to form a functional unit, has been used to study using HR activity, isolate intra and interchromosomal repair events as well as to quantify the relative length of repair tracts (Brenneman et al., 2002, 2000; Johnson and Jasin, 2000).

Methods outlined above rely on drug resistance and colony formation for isolation and quantification of mutagenic outcomes. This is potentially problematic, if drug selection interferes with the repair processes, if non-mutagenized cells need to be analyzed or if cells of interest do not form colonies. Furthermore, the need for colony formation limits the throughput of the procedure, due to time needed for colony outgrowth and low density at which cells need to be plated in order to recover pure clones. Discovery and development of fluorescent proteins eliminated these problems, allowing a simple flow cytometric efficiency readout and FACS isolation of both posi-

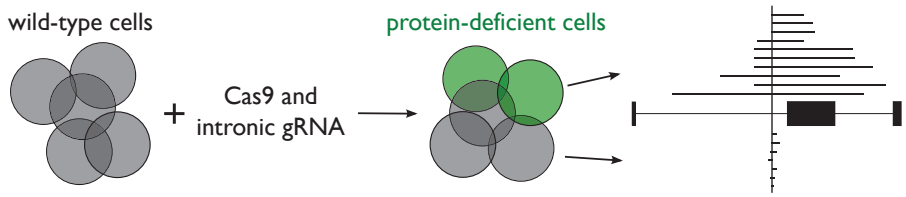


Figure 4.1: Assay design. Targeting introns allows quantification and isolation of complex lesions – i.e. lesions that are not small indels. The position of the gRNA is shown as a vertical line intersecting with the gene structure. Horizontal lines indicate indels and large deletions.

tive and negative cells (or intermediate states, if present; Julius et al., 1972; Prasher et al., 1992; Shimomura et al., 1962). Assays combining I-SceI induced DSB and fluorescent protein readout contributed substantially to research on DDR, allowing study of genetic requirements of HR (Pierce et al., 1999), translocations (Richardson and Jasin, 2000), SSA (Stark et al., 2004) and MMEJ (Bennardo et al., 2008). A principle similar to that in the split-neomycin assay was used, with split fluorescent proteins being placed at different loci and with different amounts of shared homology. Constructs combining multiple fluorescent proteins were designed to simultaneously quantify relative contributions of HR and NHEJ (Certo et al., 2011). An assay in which repair of Cas9-induced DSB using a ssDNA donor converts BFP to GFP, and mutagenic repair abolishes fluorescence altogether, was used to define optimal conditions for ssDNA donor integration (Richardson et al., 2016).

Most of the described DDR assays were designed to capture a specific type of mutation using a positive selection paradigm and often ignoring the negative population. I speculated that an assay based on negative selection against small indels, the most common lesion caused by Cas9 mutagenesis, will reveal repair outcomes that have been overlooked so far. Here, I describe the development of this assay.

4.2 Results

4.2.1 Assay design

I sought to establish an assay to detect and quantify cells with Cas9-induced lesions that are not small indels ("complex lesions"). I reasoned such an assay could be based on targeting intronic sites close to an exon (within 500 bp). Small intronic lesions are normally not expected to affect gene expression. Conversely, any other large intronic lesion, such as translocation, inversion or large deletions may affect gene expression (Fig. 4.1).

Mouse ES cells and hTERT immortalized, p53-deficient human retinoid pigment epithelial cells (RPE1) were used to establish the assay. In contrast to cancer-derived cell lines, both cell lines have a normal karyotype and intact DNA repair mechanisms, which makes them more representative of a normal somatic cell. Although mouse ES cells and embryonic fibroblasts differ in their use of DNA repair pathways, it is not known how they compare to other somatic cells (Tichy et al., 2010). P53 deficiency in the RPE1 cell line enabled easy characterization, as most Cas9-mutagenized p53-proficient RPE1 cells undergo apoptosis (Haapaniemi et al., 2018). Both ES and RPE1 cell lines can be single cell cloned, which allows creation of pure, Cas9 expressing lines as well as clonal genotypic analysis following mutagenesis.

Following criteria were used to pick targets for the assay:

- High surface expression or easily detectable function of the gene, which can be used as a readout and means of selection.
- Availability of flow cytometric reagents for detection of gene expression or function.
- "Isolated" exons flanked by more than 2 kb of intronic sequence in both directions, so that genotyping can be focused on one exon

only. Such exons could also be targeted on both flanks as a control.

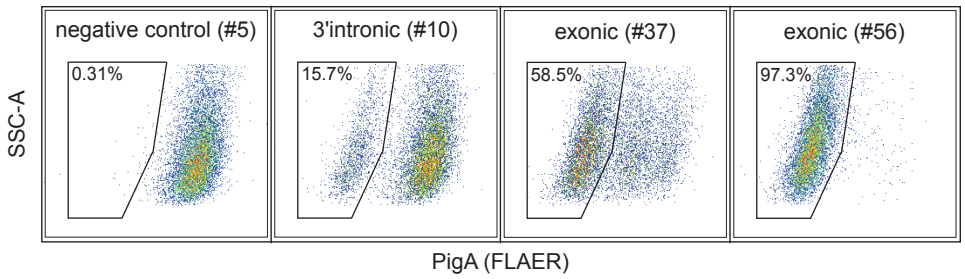
- Exons whose complete loss would change the reading frame of the transcript, so that no fully functional protein could be produced without them.
- Exons close to the 5' end of the transcript, as frameshifting mutations in these exons are more likely to trigger NMD.
- Genes which produce a single protein isoform (i.e. no alternative splicing and transcription start sites), as multiple ones could confound the readout.

For the initial study, I considered X-linked genes, which are present in only one copy in male mouse ES cells and are functionally hemizygous in female RPE1 cells, due to X inactivation. This makes the phenotypic readout stronger and easier to interpret, since only one copy of the gene needs to be inactivated to ablate the protein function. Additionally, in male ES cells I expected to detect exactly one allele per single cell clone, which would substantially simplify the genotyping strategy. Loss of chromosome X is lethal in male ES cells and they rarely maintain two X chromosomes. Therefore, detection of a single allele on chromosome X in male cells cannot be mistaken for detection of two identical alleles or monosomy, as is the case in female cells or at an autosomal locus.

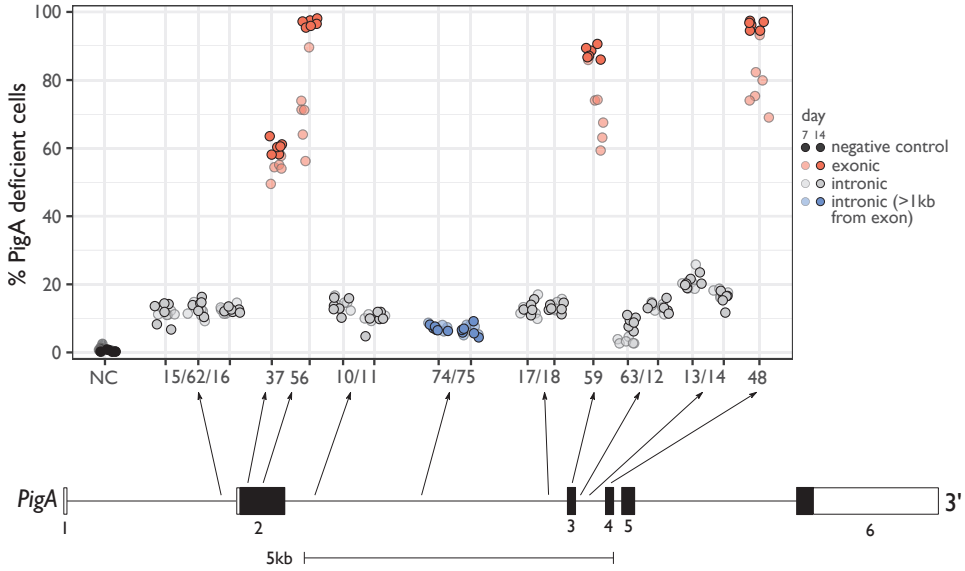
I considered *Hprt*, *Lamp2* and *PigA* as potential X-linked targets in my assay. **Hprt** is commonly used to enable gene targeting. However, my previous research indicated that the repeat rich regions around exons 2 and 3 make genotyping and Sanger sequencing particularly problematic. Furthermore, *Hprt* mutants can only be detected by a colony counting assay under 6-TG selection, which is time consuming. It may also be unreliable, if cells are plated too densely. Under such conditions some Hprt-deficient cells may be killed due to high local concentration of toxic products of 6-TG metabolism created by Hprt-proficient cells. **Lamp2** is a glycoprotein present at the lysosomal membrane that can also be found on the surface of mouse ES cells (unpublished data). However, exons 2-5 of *Lamp2* are not completely isolated, with both intron 2 and 4 being shorter than 1 kb. Moreover, available data implied Lamp2 may not be expressed highly enough to allow clear separation between positive and negative cells. **PigA** is one of the first elements of a biochemical pathway, which produces glycosylphosphatidylinositol (GPI) anchors necessary for attachment of some proteins to the surface of the cell (Miyata et al., 1993). The activity of the pathway can be assayed by flow cytometry using a fluorescent reagent, which binds to N-glycan on GPI-anchored proteins. This reagent (FLAER) is a fusion of the FITC molecule (FLuorescin isocyanate) and pro-aerolysin (AER), which is an inactive form of a bacterial toxin aerolysin (Sutherland et al., 2007). FLAER is routinely used in clinical practice to diagnose patients suspected to have paroxysmal nocturnal hemoglobinuria (PNH), a disease caused by deficient GPI-anchor production (Takeda et al., 1993). Genetic inactivation of *PigA* by CRISPR/Cas9 leads to complete loss of FLAER staining. It may result in slower cell growth, but the effect is modest (Koike-Yusa et al., 2014). Exon 2 is more than 2 kb away from exons 1 and 3, and its loss is an frameshifting mutation. Loss of exon 3 is also an out-of-frame mutation, but its proximity to exon 4 (500 bp) makes it less useful as a target. I chose to develop my assay based on the *PigA* gene.

4.2.2 Mouse *PigA* and human *PIGA* loci

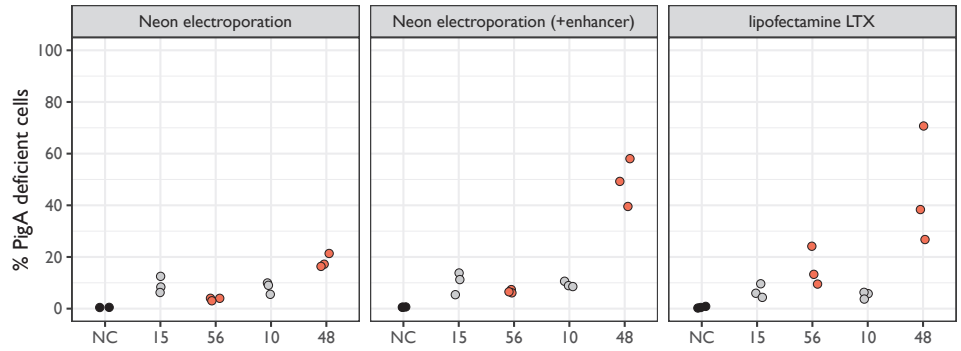
Cas9 and single gRNA constructs targeting intronic or exonic sites in chromosome X linked *PigA* gene were delivered into male JM8 mouse ES cells by PiggyBac transposition. Stable integration of both the Cas9 and gRNA expressing constructs was selected for using blasticidin and puromycin, respectively. This system allowed saturation mutagenesis of targeted loci, because even perfectly repaired targets would be recut until the site was destroyed. Staining with FLAER reagent was used to quantify the proportion of PigA-deficient cells 14 days post-delivery. Initial



(a) Examples of PigA mutagenesis revealed by FLAER staining, PiggyBac method.

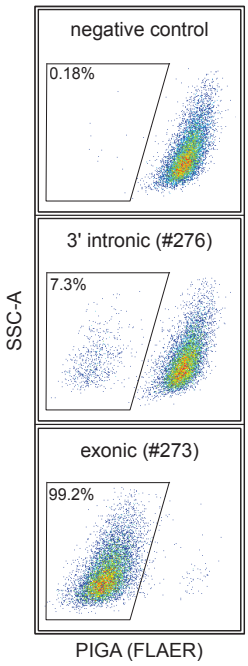


(b) Frequency of PigA loss, PiggyBac method. Each circle represents one independent cell culture (N = 6). Thick bars represent exons, hollow ones indicate UTRs. Dot transparency indicates time of sampling. NC: negative control, guide #5 targeting Cd9.

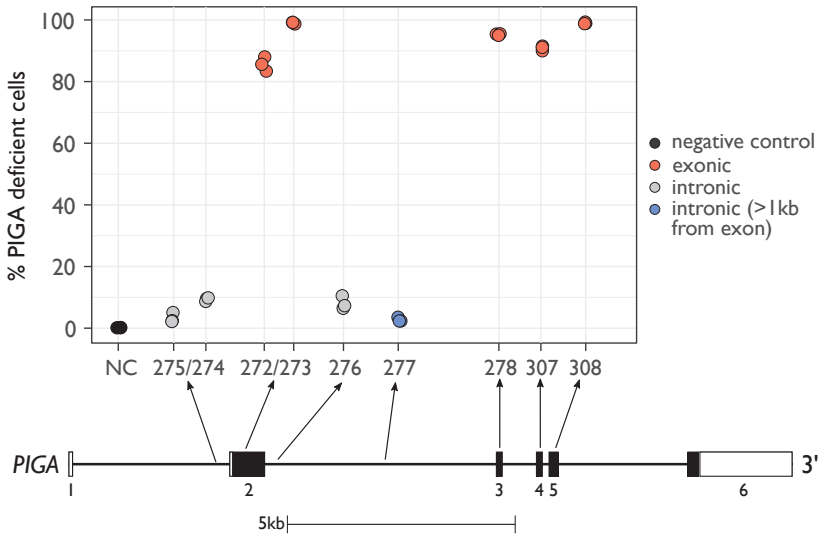


(c) Frequency of PigA loss, RNP method. Each circle represents one independent cell culture (N = 3).

Figure 4.2: Frequency of PigA loss upon mutagenesis with exonic and intronic guides in mouse ES cells. Individual guides are identified by numbers (Table 4.2).



(a) Examples of PIGA mutagenesis.



(b) Frequency of PIGA loss. Each circle represents one independent cell culture (N = 3). NC: negative control. Thick bars represent exons, hollow ones indicate UTRs.

Figure 4.3: Frequency of PIGA loss upon mutagenesis with exonic and intronic guides in human RPE1 cells using PiggyBac vectors. Individual guides are identified by numbers (Table 4.2).

experiments indicated that cells become FLAER negative gradually starting at around day 5 and plateauing at day 10. This is likely because it takes time for all GPI-anchored proteins to be recycled from the cell surface. Furthermore, some guides reached their plateau faster than others, likely due to guide-specific cutting rate or target-specific mutagenic repair (compare day 7 and day 14, Fig. 4.2b).

At 14 days post-transfection, three individual guides targeting exons 2 to 4 yielded very high rates of PigA loss (80–97%; Fig. 4.2a and 4.2b, red dots), consistent with frequent out-of-frame indels. Guide #37 targeting the 5' end of exon 2 yielded only 59%, which may be due to creation of hypomorphic PigA forms with in-frame mutations, as evidenced by intermediate FLAER intensity in some of the transfected cells (Fig. 4.2a).

Notably, guides targeting intronic sites also yielded PigA-deficient cells at significant frequencies. Ten different guides located 263–520 bp from the nearest exon caused 8–20% PigA loss, whereas two guides greater than 2 kb away induced 5–7% loss (Fig. 4.2b, gray and blue dots; Table 4.2), consistent with the mutagenic effect being distance-dependent.

I obtained similar results with transient expression using electroporation or lipofection of ribonucleoprotein complexes (RNP), proving that these observations were not a consequence of PiggyBac transposition, delivery method, antibiotic selection or cellular response to transfected plasmid DNA (Fig. 4.2c). While rates of PigA loss induced by intronic gRNAs #10 and #15 were nearly identical to those obtained by PiggyBac method, exonic gRNAs #48 and #56 were much less efficient. The difference was likely caused by slower cutting and mutagenic repair dynamics of the chosen exonic gRNAs combined with the fact their time of action is limited when using RNP. Consistently, in PiggyBac experiments the fraction of PigA-deficient cells plateaued earlier (on day 7) when using most intronic compared to exonic gRNAs (Fig. 4.2b).

To investigate whether loss of PigA expression upon intronic mutagenesis is an intrinsic property of undifferentiated mouse ES cells, I repeated the experiments in a human female differentiated cell line, RPE1 (Fig. 4.3a). I expected similar results, since RPE1 is functionally hemizygous at the *PIGA* locus, due to X inactivation. Complete ablation of FLAER staining was observed only by day 17 in RPE1 cells, later than in mouse ES cells (data not shown). This may be due to slower proliferation or increased stability of GPI-anchored proteins in this cell line. On day 17, mutagenesis of *PIGA* with all exonic and two intronic gRNAs #274 and #276 (<400 bp away from nearest exon) delivered with PiggyBac vectors resulted in a loss of PIGA at frequencies comparable to those observed in mouse ES cells (86-99% and 8.1-9.4%, respectively; Fig. 4.3b). An intronic guide #277 (>2 kb away from nearest exon) and another intronic guide #275 (<400 bp away) were much less efficient (2.7-3.2%). The exon-proximal gRNA #275 might have been exceptionally inefficient at inducing on-target damage.

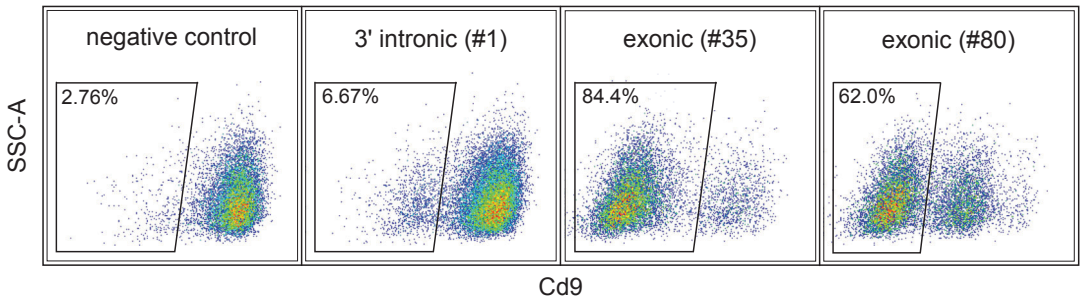
4.2.3 Autosomal *Cd9* locus

Given that only one copy of PigA is present in the male mouse ES cells I wished to exclude the possibility that the observations reflect some peculiarity of the lack of a homolog. I considered a number of autosomal genes, which are highly expressed on the surface of mouse ES cells (unpublished data) and whose exonic structure conforms to the conditions outlined above. To be able to distinguish the homologous chromosomes at the genotyping stage, I performed the experiments in mouse ES cells derived from an F1 cross between *Mus musculus* (BL6) and *Mus musculus castaneus* (CAST) mouse strains. Therefore, I was also looking for genes with high degree of divergence between the two mouse strains.

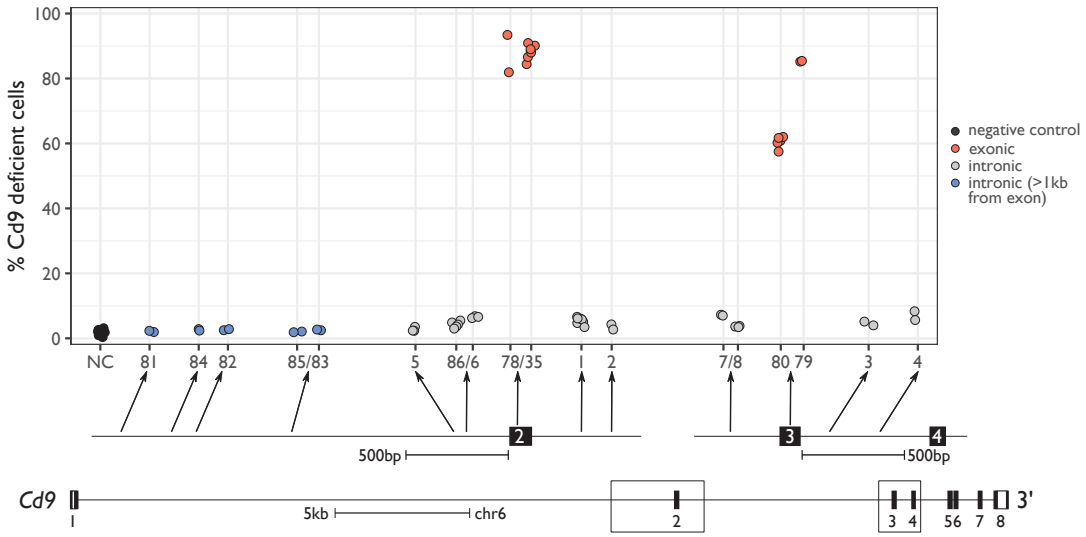
Genes fulfilling these criteria included *Cd9*, *Cd81*, *Itga6* and *Tfrc*. Initial flow cytometric tests confirmed high expression, but revealed sensitivity of Cd9 and Tfrc to differentiating conditions (plating on gelatin without LIF supplementation or dense plating on feeders). Furthermore, Cd81 and to some degree Tfrc proteins were sensitive to trypsinization, when compared to a milder Accutase treatment. Exonic guides abolished Itga6 and Cd9 expression, while Cd81 retained subpopulations with unaffected and intermediate expression. No viability phenotype was observed with these knock-outs, consistent with previous reports (Georges-Labouesse et al., 1996; Le Naour and Boucheix, 2000). Mutagenesis of *Tfrc* led to massive cell death, indicating it is an essential gene for cellular viability of mouse ES cells. This is consistent with evidence of depletion in knock-out CRISPR screens and essential role in mouse development (Blomen et al., 2015; Levy et al., 1999; Wang et al., 2015). I selected *Cd9* instead of *Itga6* for the assay, because of its higher expression and because I interpreted intermediate levels of staining with some guides as evidence that hemizygous populations can be isolated. This has proven to be misleading (see Discussion), but has not substantially influenced the results.

Most exonic guides against *Cd9* delivered with a PiggyBac vector yielded over 80% protein loss. Intronic guides 140-1900 bp away from the nearest exon generated 2.1-7.1% Cd9 loss (Fig. 4.4b; Table 4.2). Taking into account a 1.6% background of Cd9-deficient cells in the untransfected condition, I estimate the true proportion of Cd9 loss due to intronic cutting to be between 0.5–5.5%. This is consistent with results at the *PigA* locus, assuming both *Cd9* alleles have to be destroyed to prevent Cd9 expression. I confirmed that these results were not an artifact of a specific mouse ES cell line by using guides against *Cd9* locus in multiple independently derived lines (Fig. 4.5). Notably, different guides induced different levels of Cd9 loss (Fig. 4.4c and Discussion).

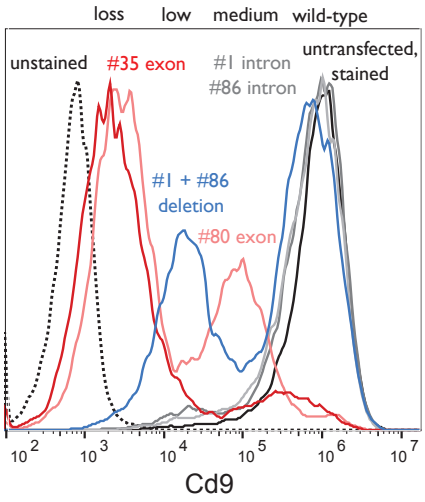
To understand the phenotypic outcomes of Cd9 editing, I isolated single-cell clones mutagenized with different gRNAs and ascertained their expression status by flow cytometry. Most clones retained the Cd9 expression status for which they were sorted. A few clones exhibited bimodal expression pattern (at 9-45% frequency), which may be the result of a mixed clone or



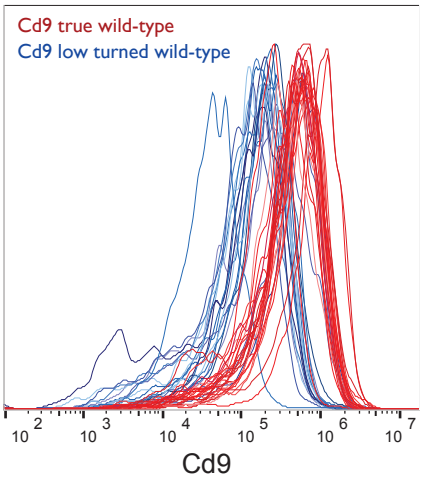
(a) Examples of Cd9 mutagenesis. A different gate was used for gRNA #80 (see Discussion).



(b) Frequency of Cd9 loss. Circles represent independent cell cultures (N = 2-8). NC: negative control. Boxed exons in the bottom diagram are magnified above.



(c) Histogram comparing possible Cd9 mutagenic outcomes. Main outcomes are named on top.



(d) Comparison of "wild-type" Cd9 expression between clones sorted for low and wild-type expression.

Figure 4.4: Frequency of Cd9 loss upon mutagenesis with exonic and intronic guides in mouse ES cells using PiggyBac vectors. Individual guides are identified by numbers (Table 4.2).

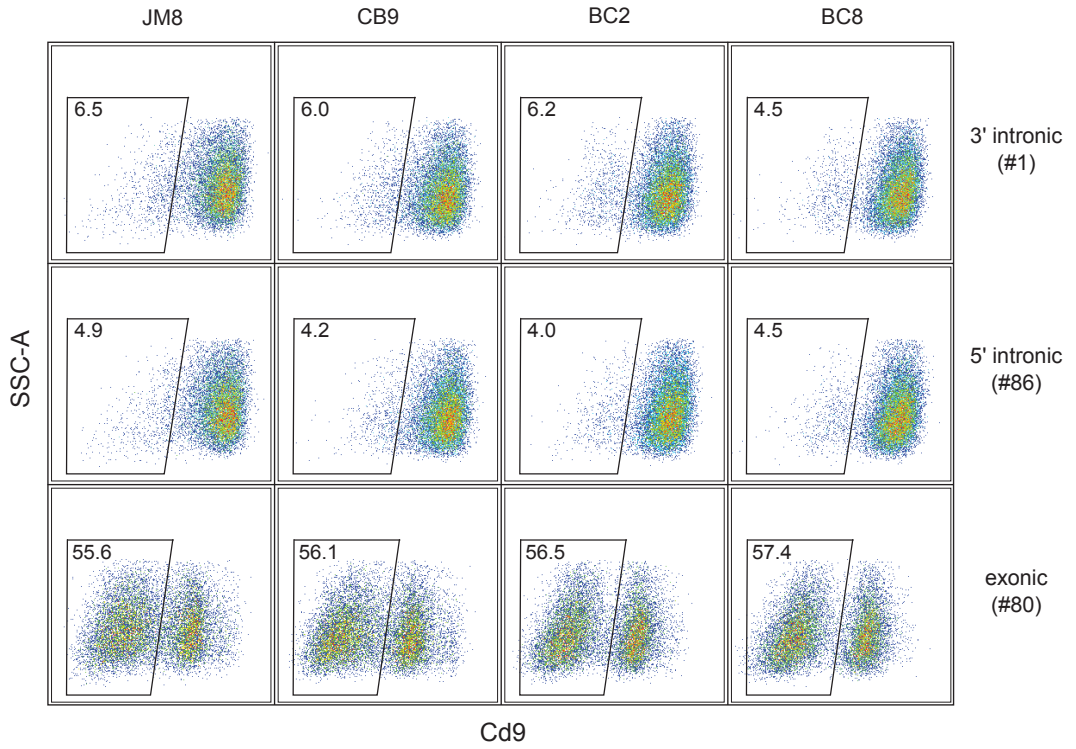


Figure 4.5: Mutagenesis in independently derived mouse ES cells lines. Name of the line indicated on top. Individual guides are identified by numbers (Table 4.2). Results shown are representative of three biological replicates. A different gate was used for gRNA #80.

mis-segregation of mutagenized chromatids during clone outgrowth (i.e. Cd9-deficient and Cd9-proficient chromatids segregating into separate cells). Notably, some of the clones derived from the Cd9-deficient population induced by intronic guides were later found to retain, on average, around 50% of wild-type levels of expression (Fig. 4.4d). They likely represent a distinct population found at the high end of the Cd9-deficient sorting gate (see Discussion). No cells mutagenized with the exonic gRNA #35 and sorted for "medium" expression of Cd9 retained that status, ending up as either "loss" or "wild-type" clones. This confirms that "medium" status is a unique population induced only by specific gRNAs (e.g. #80; data not shown).

4.3 Discussion

I have set up a simple flow cytometric assay for detection and quantification of complex Cas9-induced genomic lesions. It detected substantial levels of mutagenesis when targeting intronic sites at a hemizygous *PigA* and *PIGA* loci and an autosomal *Cd9* locus. This could be caused by either lesions destroying the nearby exon or ubiquitous presence of strong intronic regulatory elements at all intronic loci tested. The latter seems unlikely, as enhancers are neither ubiquitous, nor do they often have strong phenotypic effects. I investigate these hypotheses by directly genotyping PigA and Cd9-deficient cells in chapter 5.

PigA, *PIGA* and *Cd9* were actively transcribed. Outcomes could be different at inactive loci, if transcription or chromatin structure interferes with Cas9 activity or DNA repair. Low chromatin accessibility has been shown to impede Cas9 binding and lower editing efficiency (Horlbeck et al., 2016; Uusi-Mäkelä et al., 2018). Since in my assay both Cas9 and gRNA are constitutively active and since SpCas9 can functionally open the chromatin (Barkal et al., 2016; Polstein

Table 4.1: Non wild-type Cd9 expression levels.

Level	gRNAs	Putative cause
loss	all exonic	out-of-frame mutation and NMD
low	single and paired intronic	alternative splicing due to exon skipping
medium	exonic #80 and #53	in-frame mutation of exon 3
l-wt	all intronic	monoallelic mutation

l-wt: low turned wild-type. #53 is a Cpf1 guide targeting exon 3 (data not shown).

et al., 2015), its structure at inactive loci should not make much difference. When Cas9 is bound to the non-transcribed strand, it blocks DNA damage repair proteins from accessing the break and prevents formation of indels at 48 hours post-delivery (Clarke et al., 2018). It is not clear how the damage in those cells is eventually resolved. There does not seem to be any clear difference between intronic gRNAs targeting the template strand and non-template strand in term of frequency of PigA loss, but the dataset is not well balanced with respect to strandedness (Table 4.2). A direct experiment at a non-transcribed or temporarily silenced locus may be the only way to resolve this issue.

Lesions at the *PigA* locus resulted in either complete ablation of PigA expression or left the PigA expression unaffected (except those induced by gRNA #37, as Discussed in results section). In contrast, mutagenesis of *Cd9* locus had three distinct, non wild-type phenotypic outcomes separated by at least an order of magnitude fluorescent intensity, termed "loss", "low" and "medium" in Fig. 4.4c and Table 4.1. This variation in expression level suggests different underlying genotypes. A "negative" population was induced by three different gRNAs against exons 2 and 3 and may have resulted from out-of-frame indels triggering NMD and complete loss of protein expression. "Low" was only seen with single intronic gRNAs and deletions induced by a pair of intronic gRNAs flanking an exon. It may represent an alternative TSS or splice form, which "buffers" against complete loss of an "out-of-frame" exon 2 or 3. "Medium" expression was only observed with two specific gRNAs targeting exon 3 (incl. one Cpf1 gRNA, data not shown), which also induced a "negative" population. This "medium" state may result from an in-frame mutations that decreases the protein affinity for the antibody. This set of hypotheses can be tested by profiling local indels and RNA transcripts in cell populations sorted for their Cd9 expression level. If it is true, targeting different parts of *Cd9* would allow quantification and isolation of specific classes of genomic lesions.

Some cells edited with intronic gRNAs and sorted for low Cd9 expression were found to express near wild-type levels of Cd9 after clone outgrowth. These "low turned wild-type" clones could stem from the "background" $Cd9^{low}$ population observed in the negative control (Fig. 4.4b). Such population would be partially differentiated due to prolonged culture on gelatin with LIF supplementation. However, in a control experiment using gRNA against an irrelevant locus only about 3% of the expected number of such $Cd9^{low}$ cells formed colonies (all retaining wild-type expression; data not shown). Therefore, they would not have contributed substantially to the "low turned wild-type" population observed here. Observation that these clones express on average 50% less Cd9 than "true wild-type" clones indicates that they may represent a hemizygous population (see chapter 5) If this is the case, then exonic gRNAs should also induce a similar population.

Table 4.2: Flow cytometry results and gRNA sequences.

Experiment	Guide	Sequence with PAM	Chr	Cutting position	Strand	Type	Distance from the nearest exon	%Expression deficient (mean)	%Expression deficient (sd)	N	%Expression deficient (adj.mean)
PigA	5	GCAGTGAAGATAAATCACAAGGG	6	125472779	T	NC	-	0.4	0.3	6	-
PigA	15	CGTTGTGTCACAGTGCATAATGG	X	164422321	NT	intronic	260	12	3.3	6	-
PigA	62	TGTGACACAACGTTTAAAAGTGG	X	164422349	T	intronic	238	14	2.1	6	-
PigA	16	GAACATCTACTTGCTTAGCAGGG	X	164422416	NT	intronic	165	12	0.7	6	-
PigA	37	AAGGTTTCCAGAGCTACCCGGGG	X	164422701	NT	exonic	-	60	2.0	6	-
PigA	56	GCAGAGAAAGAACTGTGGGAATGG	X	164423023	NT	exonic	-	97	1.0	6	-
PigA	10	AGGAAGCCATAAGATAGCCACGG	X	164423864	NT	intronic	503	14	2.2	6	-
PigA	11	GCATAAGAGTGGATAAAACCAGG	X	164423884	NT	intronic	523	9.7	2.6	6	-
PigA	74	TGAGGTACTGTACCATGCACAGG	X	164425741	NT	intronic	2188	7.1	0.7	6	-
PigA	75	GAGGGTAAGTAACTCGCCCAAGG	X	164425844	T	intronic	2091	6.4	1.6	6	-
PigA	17	ACTTGTTCATACAGCCTACGTGG	X	164427667	NT	intronic	262	13	1.7	6	-
PigA	18	GATATGGGTATGTGGCAGTAGCGG	X	164427749	T	intronic	186	13	1.2	6	-
PigA	59	GGGACCAAAGAGAATCATTTTGG	X	164428028	T	exonic	-	88	1.8	6	-
PigA	63	TGCCTCTTATAAATTGAAGCAGG	X	164428148	NT	intronic	86	8.9	1.8	6	-
PigA	12	ATAAGAGGCATGCAAATAGAAGG	X	164428178	T	intronic	110	13	1.6	6	-
PigA	13	CATACGAGCTGTGACACAACAGG	X	164428347	T	intronic	196	21	1.6	6	-
PigA	14	AAGTTGTGTCTATTACACTGGGG	X	164428376	T	intronic	167	16	2.2	6	-
PigA	48	ATGCAGAACGCTTCAGTGAGGG	X	164428620	NT	exonic	-	96	1.3	6	-
Cd9	NC	[untransfected or edited at *PigA*]	-	-	-	NC	-	1.6	1.0	8	0
Cd9	81	GTGCAGAGCACGCCTTCACGGGG	6	125474376	NT	intronic	1865	2.1	-	2	0.5
Cd9	84	GCAGTGTCCTTATCTAAGAGGGG	6	125474156	T	intronic	1645	2.6	-	2	1.0
Cd9	82	ATGTAAGCCCCCTTAGTCCCCGG	6	125474028	T	intronic	1517	2.7	-	2	1.1
Cd9	85	CAGCCAGCCACTACACTGGAGGG	6	125473582	T	intronic	1071	2.0	-	2	0.4
Cd9	83	ACCTCTTATCACTGGTACCAGGG	6	125473547	NT	intronic	1036	2.6	-	2	1.0
Cd9	5	GCAGTGAAGATAAATCACAAGGG	6	125472775	NT	intronic	264	2.7	0.7	3	1.1
Cd9	86	CAACTGCAGCACTTCCGGCAGGG	6	125472720	T	intronic	209	4.2	1.0	5	2.6
Cd9	6	GATTCACACACAGTTCCCTGCCGG	6	125472717	NT	intronic	206	6.5	0.3	3	4.9
Cd9	78	CAGTGCTTGCTATTGGACTATGG	6	125472481	T	exonic	-	88	-	2	86
Cd9	35	TCTTGGTCTGAGAGTCGAATCGG	6	125472467	NT	exonic	-	88	2.2	7	87
Cd9	1	AAGGATGCCACCACTCCTGAGGG	6	125472162	T	intronic	246	5.4	1.1	7	3.8
Cd9	2	ATTCAGGAAGCCGGTCTGGAGGG	6	125472028	NT	intronic	380	3.5	-	2	1.9
Cd9	7	GGTTGTCCCCTAAGCATCAAGGG	6	125464747	T	intronic	260	7.1	-	2	5.5
Cd9	8	TCAACACTCTACCTCATCCTCGG	6	125464703	NT	intronic	216	3.6	0.2	3	2.0
Cd9	80	AGCCGGGGCCCTCATGATGCTGG	6	125464449	T	exonic	-	60	1.8	5	59
Cd9	79	GTACAGCTCCACAGCAGCCCAGG	6	125464432	NT	exonic	-	85	-	2	84
Cd9	3	GCCTGAAGTAAGGATGGTGAGAGG	6	125464252	NT	intronic	137	4.6	-	2	3.0
Cd9	4	CTTTGTTTCCCCGATCTCGGTGG	6	125464003	T	intronic	386	7.0	-	2	5.4
progenitor	311	GGGCGAGGAGCTGTTCACCGGGG	-	-	T	exonic	-	-	-	-	-
cherry/gfp	33	GAAGTTCGAGGGCGACACCCTGG	-	-	T	exonic	-	-	-	-	-
cherry/gfp	34	GGAACAGTACGAACGCGCCGAGG	-	-	T	exonic	-	-	-	-	-
RPE1	231	AGGCTTCCCGCATTCAAAATCGG	3	46371987	NT	NC	-	0.2	0.0	3	-
RPE1	308	GTTGTAAGTACCAGAGTTGGTGG	X	15324855	T	exonic	-	99	0.3	3	-
RPE1	307	TTTGGGAGCTTTAGAACACAAGG	X	15325127	T	exonic	-	91	0.8	3	-
RPE1	278	GGATATTTCTGACAGAGTTCAGG	X	15326014	NT	exonic	-	95	0.3	3	-
RPE1	277	GAATGTCTTAAGTGAGAGAGAGG	X	15328488	T	intronic	2278	2.7	0.7	3	-
RPE1	276	AGAGGGCAGGCCGTGTACGGTGG	X	15330896	T	intronic	323	8.1	2.1	3	-
RPE1	273	TGCTCAGGTACATATTTGTTCGG	X	15331580	T	exonic	-	99	0.3	3	-
RPE1	272	GTAATAGACTTTGAGGCCACTGG	X	15331674	NT	exonic	-	86	2.3	3	-
RPE1	274	TGGTAAACCATGATATGCTGTGG	X	15332254	T	intronic	261	9.4	0.6	3	-
RPE1	275	GGTAAAGTATAAGAGTAAAGGGG	X	15332346	T	intronic	353	3.2	1.6	3	-

Genomic position is given with respect to the GRCm38 or GRCh38 (RPE1 experiment) reference genome. Last column contains negative control subtracted mean (Cd9 experiment). SD = standard deviation. Strand: T = transcribed, NT = non-transcribed.