Chapter 2

CRISPR/Cas9 screening of isogenic iPSCs

2.1 Introduction

The strategy of exploiting synthetic lethality to selectively kill cancer cells was first suggested over 20 years ago,[33] yet few interactions have been well established. This may be in part because a comprehensive interrogation of SLIs has not been carried out for many genes. The development of CRISPR/Cas9 technology has made it possible to screen for interactions on a genome-wide scale.[104,105] We chose to screen a panel of candidate TSGs using an isogenic system i.e. comparing a parental line with a derivative that has a single gene knockout. As cancer cell lines often have many genetic aberrations, we decided to perform these experiments using a human iPSC line. Our rationale was that this clean background would reduce the confounding factors that may affect an interaction between two genes. This chapter describes the establishment of an isogenic iPSC line panel, construction of a genome-wide library and the various steps taken to perform CRISPR/Cas9 screening in these cells.

2.1.1 Aims of this chapter

- To engineer isogenic iPSC lines and confirm gene knockout.
- To design and construct a gRNA library backbone with a neomycin selectable marker.
- To generate stable Cas9-expressing iPSC lines.
- To optimise a CRISPR/Cas9 screening protocol in iPSCs.
- To understand how the screening process affects iPSCs.

2.2 Selection of genes for an isogenic cell line panel

For our findings to be therapeutically relevant, we decided to screen for synthetic lethal partners of TSGs that are already known to be lost in primary tumours. A recent analysis was performed on genomic sequences from 7651 tumours, covering 28 tumour types, to identify driver genes characterised by LOF mutations.[242] This study identified 53 genes that had a significant enrichment of nonsense mutations (Fig. 2.1); these included known drivers and novel candidate TSGs, likely missed previously due to their low mutation frequency. We screened 15 genes from this panel to identify synthetic lethal interactions (dark blue in Fig. 2.1). Specific focus was placed on *ARID1A, ARID1B, ARID2* and *PBRM1* as these are all subunits of the PBAF/BAF complexes.[140]

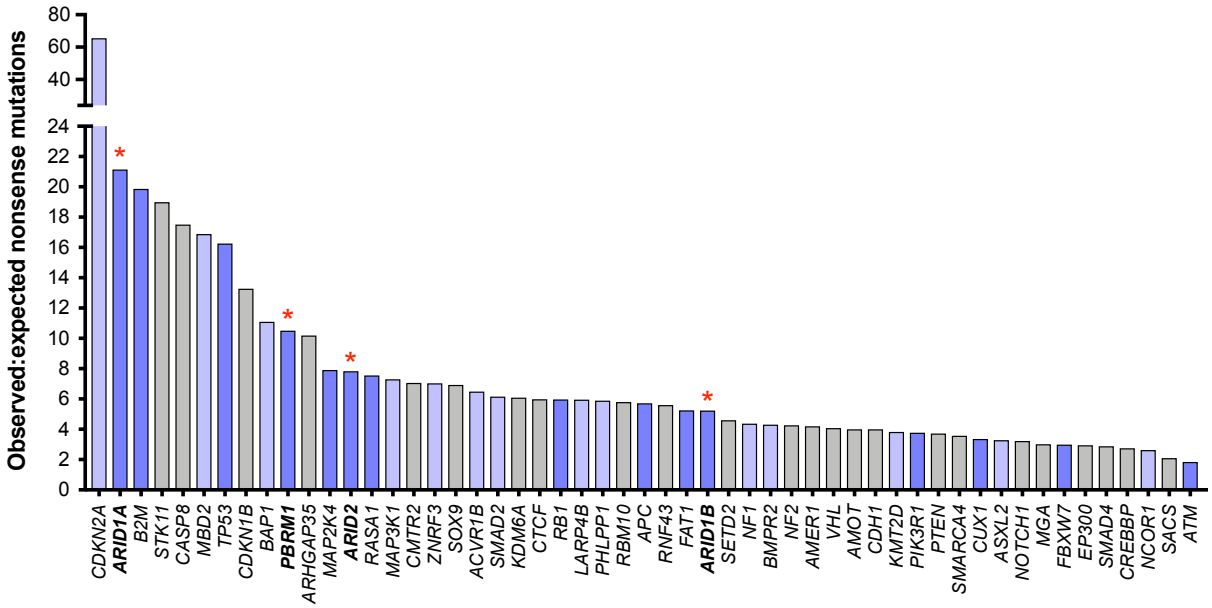

Figure 2.1. Candidate tumour suppressor genes. Wong *et al.* performed analysis on 7,651 genome sequences (352 whole genomes, 7,299 exomes) across 28 tumour types.[242] The genes shown had a high ratio of observed to expected nonsense mutations (Benjamini-Hochberg's false discovery rate adjusted p-value, $q < 0.01$). Isogenic knockout lines were engineered for 28 genes (blue) with two independent clones for each (*BAP1* and *ACVR1B* had only one clone). Of these, 15 genes were screened (darker blue) and 4 genes belonging to the PBAF/BAF complex were prioritised for further study (*). Figure adapted from Wong *et al.* (2014).[242]

2.3 Engineering isogenic iPSC lines

The Gene Editing facility at the Wellcome Sanger Institute (WSI) produced all of the KO iPSC lines described in this thesis, using a human iPSC line called 'BOB' (experimental details in Section 7.3). BOB was derived from fibroblasts donated by a 65-year old male with an autosomal recessive disorder, α1-antitrypsin deficiency.[243] This results from a mutation in the *A1AT* gene, which leads to accumulation of the protein in hepatocytes and can cause liver damage.[244] The BOB cells originally carried a single point mutation in *A1AT* but this was corrected to wildtype by Yusa *et al.* (2011) before the line was used in this project.[245] The iPSCs were reprogrammed using an integration-free method with Sendaiviral vectors that expressed the pluripotency markers *OCT4*, *SOX2*, *KLF4* and *MYC*.[245]

Two gRNAs targeting an early exonic region were designed for each gene. A template was designed for each target, with 5' and 3' homology arms flanking a puromycin-resistance cassette (Fig. 2.2). In the presence of Cas9, the gRNAs induce DSBs and the puromycin-resistance cassette can be inserted via HDR. As HDR occurs at a low frequency, it was expected that most edited cells would contain only one allele carrying the cassette and the other allele would be repaired by NHEJ. NHEJ often results in frameshift indels and, in combination with insertion of the puromycin cassette on the other allele, this would induce a homozygous KO of the targeted gene. Cells were transfected with both gRNAs, Cas9 and the homology template. Puromycin was then used to select for successfully edited clones. PCR and Sanger sequencing were performed to identify single-cell clones containing a frameshift indel. Two homozygous KO clones were expanded for each gene and transferred to our lab, where all lines were maintained in puromycin. We maintained two KO clones for each gene to allow us to cross-validate any observed effects and ensure that they were not specific to one clone. The knockout genotypes were confirmed by Sanger sequencing at various stages of culture (Fig. 2.3, Table 2.1, Section 7.4).

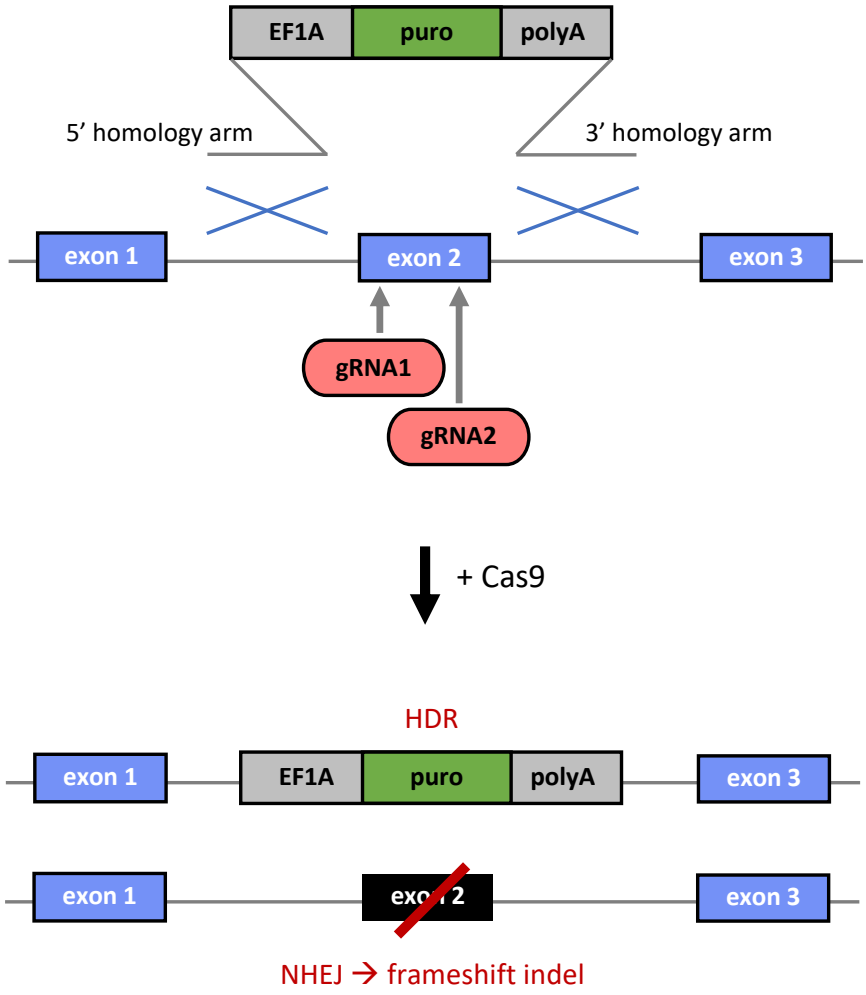

Figure 2.2. Engineering isogenic KO iPSCs. Schematic showing the strategy used to engineer isogenic KO iPSCs. A template containing a puromycin resistance cassette flanked by 5' and 3' homology arms, two gRNAs targeting an exon, and Cas9 were transfected into BOB cells. Edited clones had the puromycin cassette in one allele (inserted via HDR), and a frameshift indel in the other allele (caused by NHEJ).

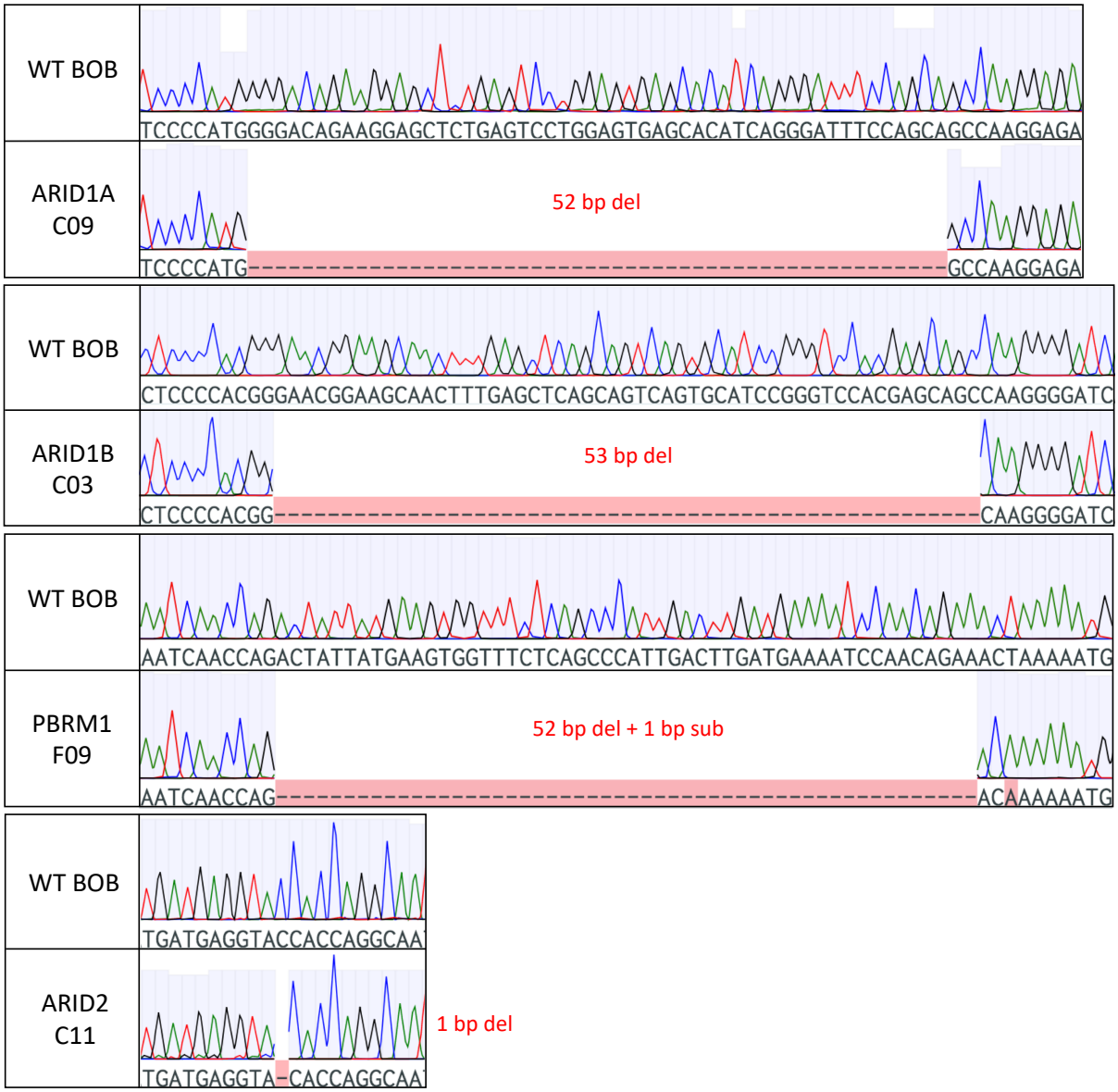

Figure 2.3. Sequencing of PBAF/BAF mutations in KO iPSCs. PCR and Sanger sequencing were performed to confirm the frameshift deletions (shown in red) in *ARID1A, ARID1B, PBRM1* and *ARID2*. The WT sequence in the parental was also sequenced for comparison. Traces are shown for one clonal line per gene. All KO lines were sequenced in the same way. Primers used for genotyping are detailed in Appendix A.1.

Table 2.1. Genotypes of screened KO iPSC lines.

Cell line	Mutation	Variant genomic position
APC_F10	8 bp insertion	5:112767214-112767236
ARID1A_B08	53 bp deletion	1:26760883-26760905
ARID1A_C09	52 bp deletion	1:26760883-26760905
ARID1B_C03	53 bp deletion	6:157084690-157084712
ARID1B_G01	52 bp deletion	6:157084690-157084712
ARID2_A11	1 bp insertion	12:45811458-45811480
ARID2_C11	1 bp deletion	12:45811458-45811480
ATM_A12	1 bp insertion	11:108244867-108244889
ATM_B11	5 bp deletion	11:108244867-108244889
B2M_B08	50 bp deletion	15:44715481-44715503
CUX1_H10	22 bp deletion	7:102104394-102104416
FAT1_A12	53 bp deletion	4:186636051-186636073
FBXW7_C01	53 bp deletion	4:152346958-152346980
MAP2K4_B01	2 bp deletion	17:12110414-12110436
PBRM1_F08	55 bp deletion	3:52668572-52668594
PBRM1_F09	52 bp deletion + 1 bp sub	3:52668572-52668594
PIK3R1_B08	7 bp deletion	5:68293402-68293424
RASA1_H02	56 bp deletion	5:87349327-87349349
RB1_C04	8 bp deletion	13:48362829-48362851
TP53_C05	86 bp deletion	17:7675126-7675148

Liquid chromatography-mass spectrometry (LC-MS) was performed on the parental BOB line and the KO derivatives chosen for screening (Fig. 2.4, Appendix A.2, Section 7.6). This was primarily to confirm loss of the targeted proteins, but could be a useful dataset to understand the global effect of knocking out these genes. Clara Alsinet (a postdoctoral fellow in the Adams' lab) prepared the cells for these experiments, and they were processed and analysed by Jyoti Choudhary and Theodoros Roumeliotis (previously in the Proteomics Facility at WSI).

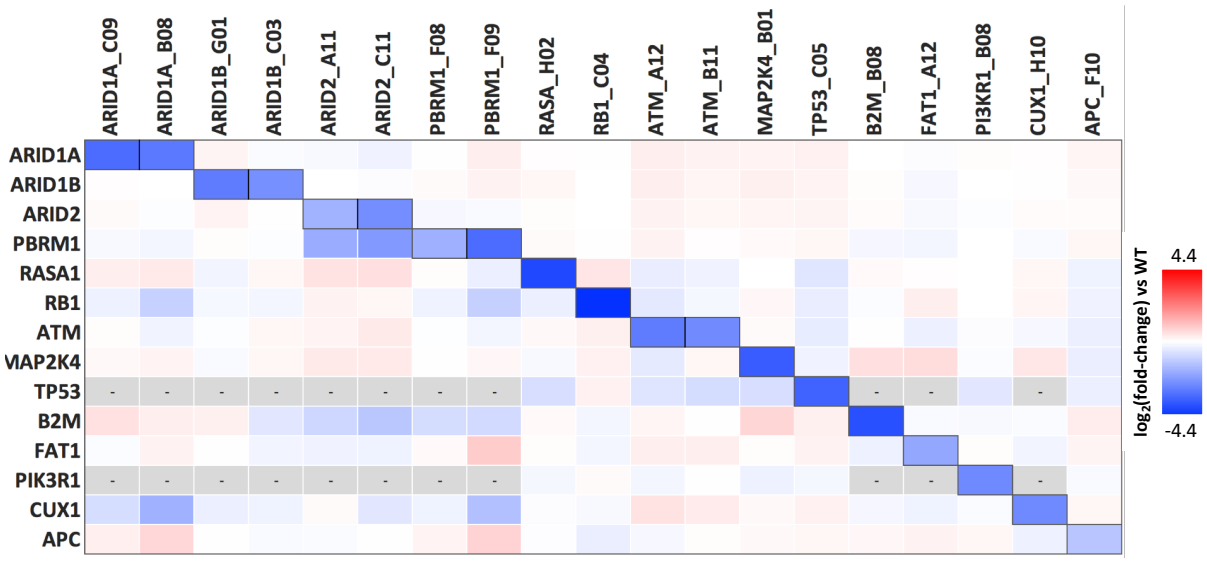

Figure 2.4. Proteomics of knockout iPSC lines. Log_2 (fold-change) of protein abundance in KO lines relative to the parental BOB line, as measured by LC-MS. The parental was always analysed in the same run as the KO line that it was compared to. (–) indicates that a protein was not detected. FBXW7 is not shown as the protein was not detected in any run. Values are provided in Appendix A.2.

Western blotting was also performed on the *ARID1A, ARID1B,* and *ARID2* KO lines for additional confirmation of protein loss (Fig. 2.5, Section 7.5). We do not have data available for the *PBRM1* knockout lines due to antibody issues. The quality of these blots was low and they should ideally be repeated. However, loss of ARID1A, ARID2 and ARID1B in their respective knockout clones was observed in comparison to the parental BOB line.

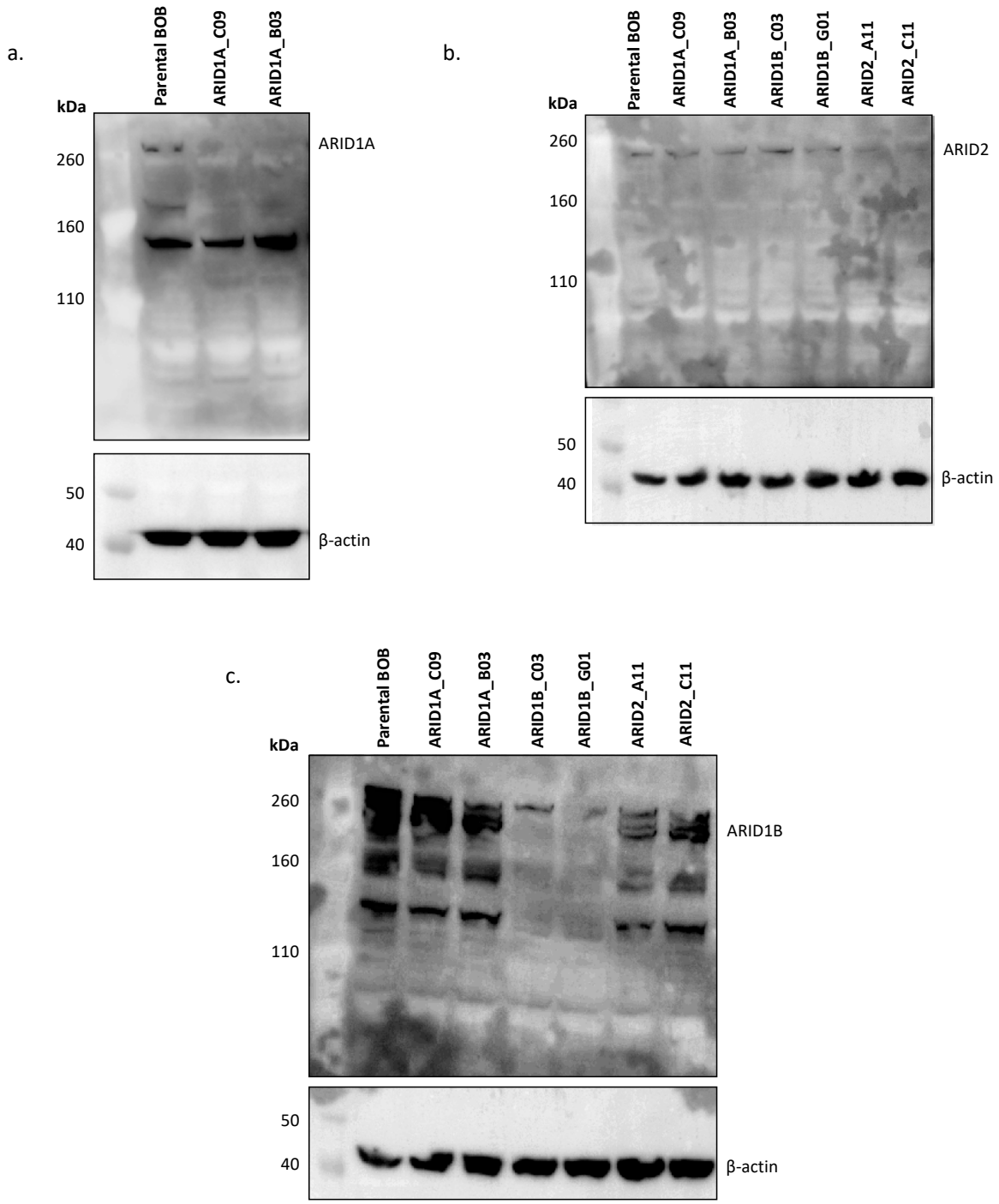

Figure 2.5. Western blotting of ARID1A, ARID2 and ARID1B in iPSCs. Protein lysates were extracted from parental BOB cells and both knockout clones for *ARID1A, ARID1B* and *ARID2*. Western blotting was performed to measure expression of ARID1A (a), ARID2 (b) and ARID1B (c) proteins. In each case, expression was measured in the parental BOB lysate for comparison. β-actin (42 kDa) was used as a loading control. The ARID1A antibody datasheet predicted the molecular weight within the range of 165-320 kDa. ARID1B has a predicted molecular weight of 236 kDa. The ARID2 antibody datasheet stated that the observed molecular weight should be 217 kDa.

2.4 Construction of a novel gRNA backbone

There are several human genome-wide gRNA libraries available for CRISPR/Cas9 KO screens; we opted to use Kosuke Yusa's library (referred to hereafter as Yusa v1.1) as it has been widely used at WSI. Therefore, colleagues had experimental/analytical expertise using this library and pipelines were in place for sample processing.[109] This library contains 101,090 gRNAs targeting 18,009 genes, with 1004 non-targeting control (NTC) gRNAs that do not match any sequence in the human genome.[109] In this vector, gRNAs and a tracrRNA scaffold are under the control of a U6 promoter, and a PGK promoter drives a mammalian puromycin-resistance cassette and BFP, separated by a T2A element (Fig. 2.6a).[110] Puromycin can be used to select for cells that have been successfully transduced with the library. BFP can also be used to select for transduced cells via fluorescence-activated cell sorting (FACS), or it can act as a marker of transduction efficiency and puromycin selection.

As discussed previously, the strategy used to engineer isogenic KO iPSC lines involved insertion of a puromycin cassette (Fig. 2.2). Therefore, it was not possible to utilise the puromycin selectable marker in the Yusa v1.1 library as our KO lines were already resistant. The other commercially available genome-wide libraries (Brunello[246], GeCKO[107], Toronto[247]) also have a puromycin selectable marker. We trialled sorting BOB cells based on BFP expression but changes in cell morphology during the process made it difficult to sort efficiently, and so large cell numbers would have been required to maintain good coverage.

2.4.1 Swapping puromycin resistance for neomycin resistance

We therefore decided to alter the Yusa v1.1 backbone to replace the puromycin resistance with a neomycin resistance gene, which allows mammalian cells to survive in an antibiotic known as G418.[248] All oligonucleotides and primers used during construction of the backbone are detailed in Appendix A.1. Experimental details for all steps described here are provided in Section 7.7. The backbone was digested with Kpn2I (Fig. 2.6b) to prepare for insertion of a 901 bp fragment containing the *neo* gene (fragment A); an additional 120 bp fragment (fragment B) was used to provide an overlap with the first Kpn2I restriction site (Fig. 2.6c). Gibson Assembly was performed, transformed into *E. coli* and bacterial colonies were screened by Sanger sequencing using primers tiled along the cloning region (results not shown). A single successfully edited clone containing *neo* was taken forward (Fig. 2.6d).

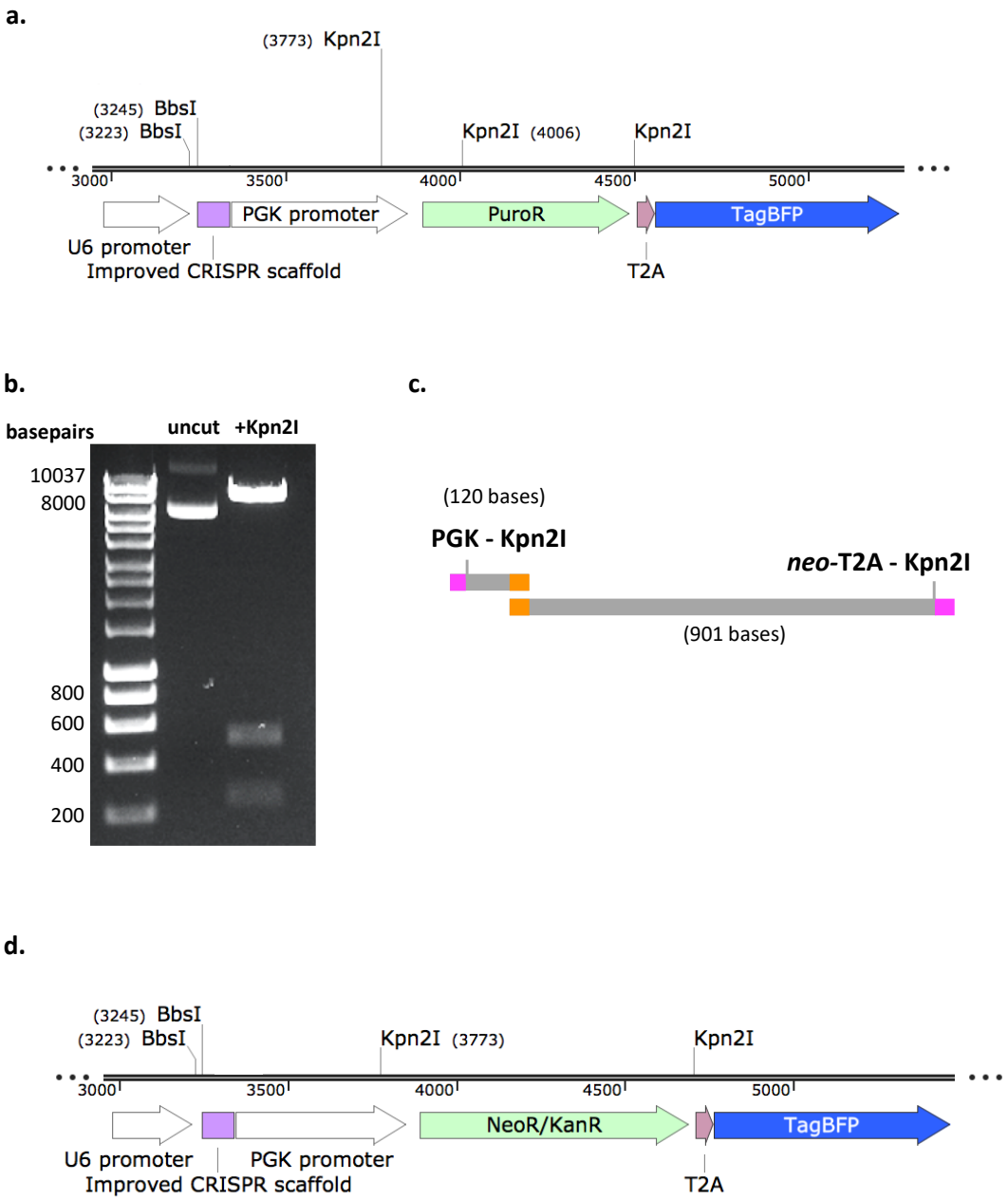

Figure 2.6. Cloning of a neomycin-resistant library backbone. a) Schematic showing key features of the Yusa v1.1 library backbone, **b)** agarose gel image showing Kpn2I digestion of the Yusa v1.1 backbone (marker sizes indicated on left), **c)** illustration of the oligonucleotides designed for Gibson Assembly to insert *neo* (vector overlaps shown in pink), **d)** schematic of the resulting neoR backbone.

To confirm that the neomycin resistance and BFP were functional in the new backbone (referred to hereafter as neoR), the plasmid was packaged into a lentivirus. BOB-Cas9 cells were transduced and BFP was measured by flow cytometry after 48 hours. Cells were then seeded in 6-well and 12-well plates; one well contained untransduced BOB-Cas9 cells, and all other wells contained transduced cells with 0, 0.1, 0.3, 0.6 or 1 mg/ml of G418. After five days in culture with G418, cells in the 12-well plate were fixed and BFP was measured by flow cytometry. Transduced cells did not survive G418 selection so no data was obtained for these conditions. The transduced cells cultured without drug did survive and BFP was stable during this period, suggesting that the backbone itself was not inducing lethality (Fig. 2.7a). Cells in the 6-well plate were fixed after five days in G418 and staining with crystal violet confirmed that there was no survival in the drug (Fig. 2.7b). We concluded that the neomycin resistance gene was not functional.

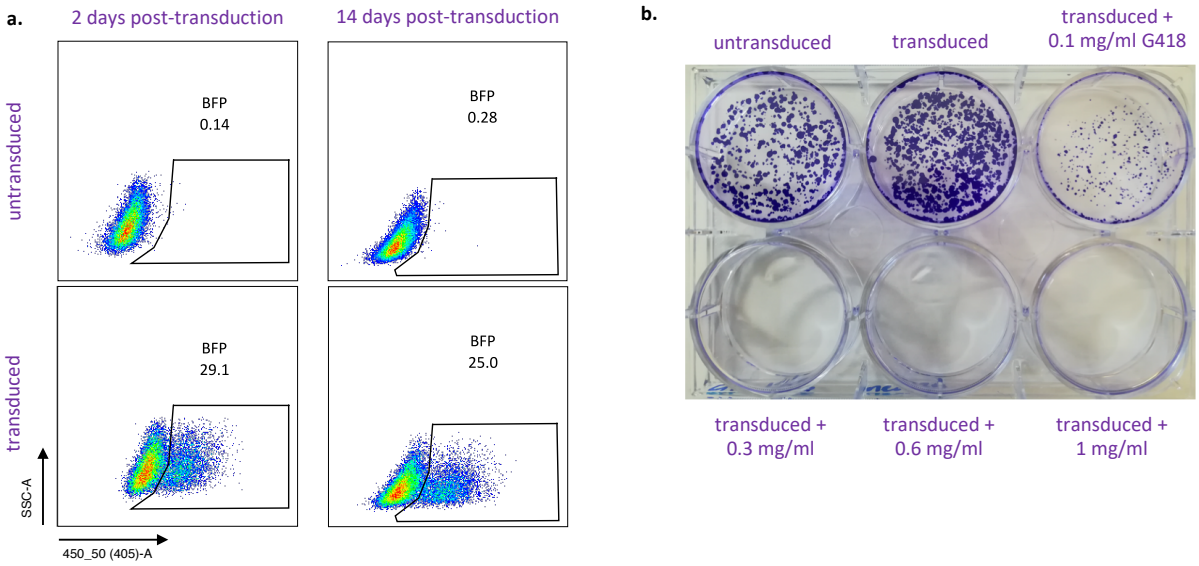

Figure 2.7. Testing neoR backbone resistance to G418. a) Flow cytometry plots showing BFP expression (450_50 (405)-A) in BOB-Cas9 cells at 2- and 14-days post-transduction with the neoR backbone, **b)** clonogenic assay showing BOB-Cas9 cells stained with crystal violet on day 14 post-transduction with the neoR backbone. Cells were selected with various doses of G418 from day 9-14. Untransduced cells were used as a control to show that transduction itself was not toxic.

2.4.2 Alteration of neomycin-resistant library backbone

We reasoned that there may be an issue with the T2A element that separated the *neo* and BFP coding regions; we were unable to find any commercially available plasmids that contained *neo* followed by a T2A. Coding regions separated by a T2A are transcribed as a single mRNA and translated as one protein, then cleaved at the T2A site, leaving a small additional sequence on the upstream protein.[249] It is possible that the additional sequence left by the T2A interfered with the function or folding of the neomycin resistance protein. A colleague at WSI, Luca Crepaldi, had successfully achieved G418 resistance using an IRES element downstream of the *neo* gene. Like the T2A, an IRES is used to separate two coding regions; it introduces a new translation site and the two proteins are translated separately.[250,251] We decided to replace the T2A with an IRES to test whether this would provide resistance using our backbone.

In the original vector, gRNAs were cloned into a region downstream of the U6 promoter after linearisation with BbsI.[110] The IRES sequence also had BbsI restriction sites; it was unclear whether altering this sequence would have a detrimental effect on IRES function so an alternative gRNA cloning strategy was planned instead. An insert was designed to replace the gRNA cloning region between the U6 promoter and tracrRNA scaffold with a sequence containing an AjuI recognition site. AjuI works in a similar way to BbsI and as the surrounding sequence would be unaltered, an established PCR protocol (provided by Kosuke Yusa) could still be used to amplify gRNAs from the original library for the transfer. The neoR backbone (Fig. 2.6d) was digested with BbsI (Fig. 2.8a) and Gibson Assembly was performed to clone in the AjuI insert (fragment C, Fig. 2.8b). Bacterial colonies were screened by Sanger sequencing (results not shown) and an edited clone was taken forward.

This vector was digested with RsrII and BsrGI (Fig. 2.8c) to prepare for insertion of two fragments. One contained a *neo*-IRES sequence amplified by PCR from a plasmid provided by Luca Crepaldi (fragment D, Fig. 2.8d); the other was synthesised to overlap with the IRES and the BFP coding region, extending across the BsrGI site (fragment E, Fig. 2.8e). Gibson Assembly was performed, transformed into *E. coli* and bacterial colonies were screened by Sanger sequencing (results not shown). A clone containing the *neo*-IRES-BFP sequence was obtained (Fig. 2.8f). Appendix A.3 shows the complete plasmid map and sequence for this plasmid backbone.

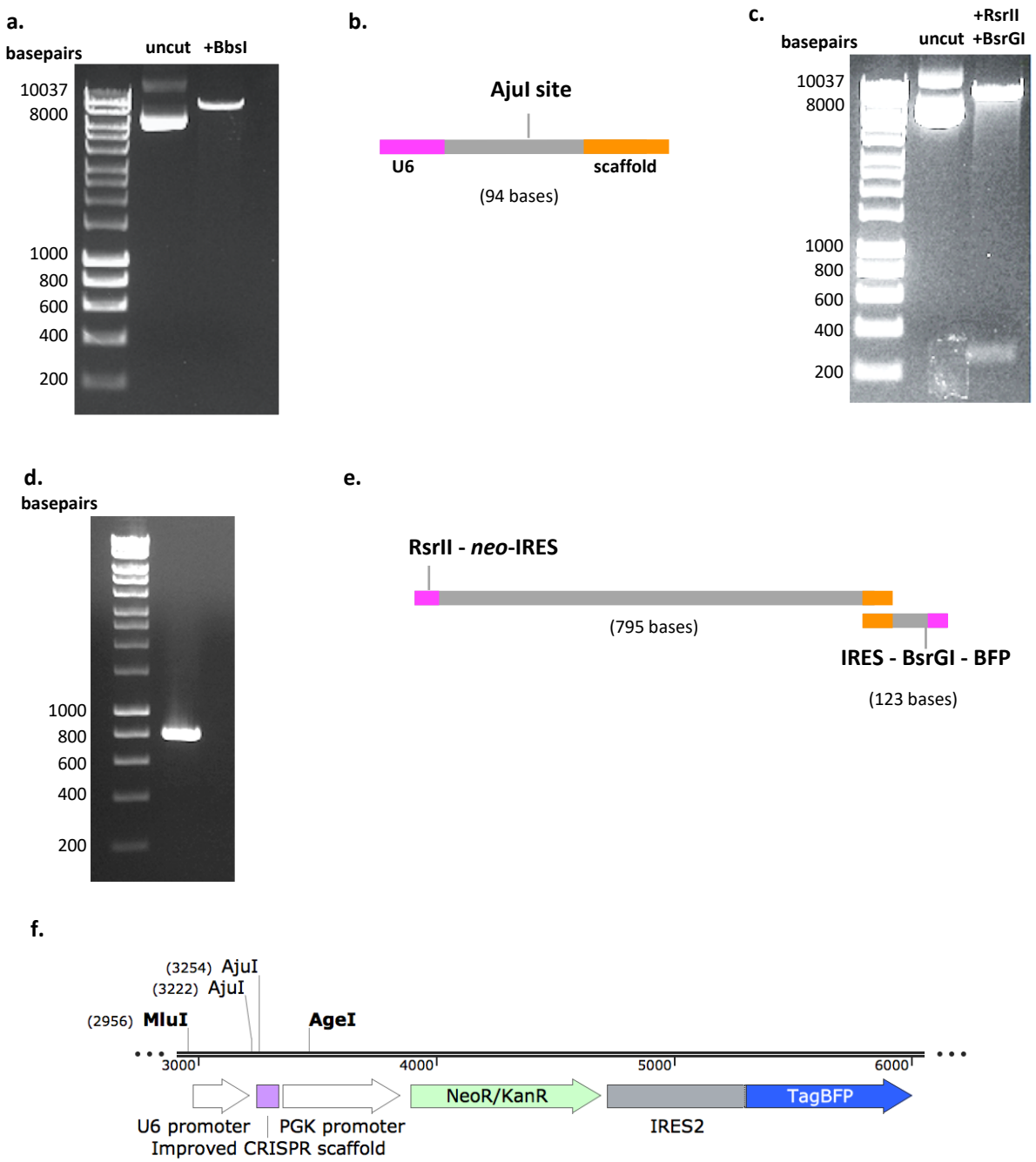

Figure 2.8. Cloning in AjuI recognition site and an IRES element. a) BbsI digest of neoR backbone, **b)** illustration of the oligonucleotide designed to insert AjuI recognition site via Gibson Assembly, **c)** RsrII/BsrGI double-digest of neoR (AjuI) backbone, **d)** PCR amplification of neoIRES sequence from a plasmid provided by Luca Crepaldi, **e)** illustration of the neoIRES PCR amplicon and the IRES-BFP oligonucleotide designed for Gibson Assembly, **f)** schematic of the final neoR-IRES backbone.

This neoR-IRES backbone was then packaged into a lentivirus and BOB-Cas9 cells were transduced. As described in Section 2.4.1, BFP was measured 48 hours post-transduction and a clonogenic assay and flow cytometry analysis were performed to test resistance to G418 (Fig. 2.9). For the clonogenic assay, cells were also transduced with the neoR backbone for comparison, to confirm that G418 was effective (Fig. 2.9b). Cells transduced with the neoR-IRES backbone survived selection with all doses of G418; 1 mg/ml caused the greatest increase in BFP positive cells. The neoR-IRES backbone was therefore deemed suitable for use in screening.

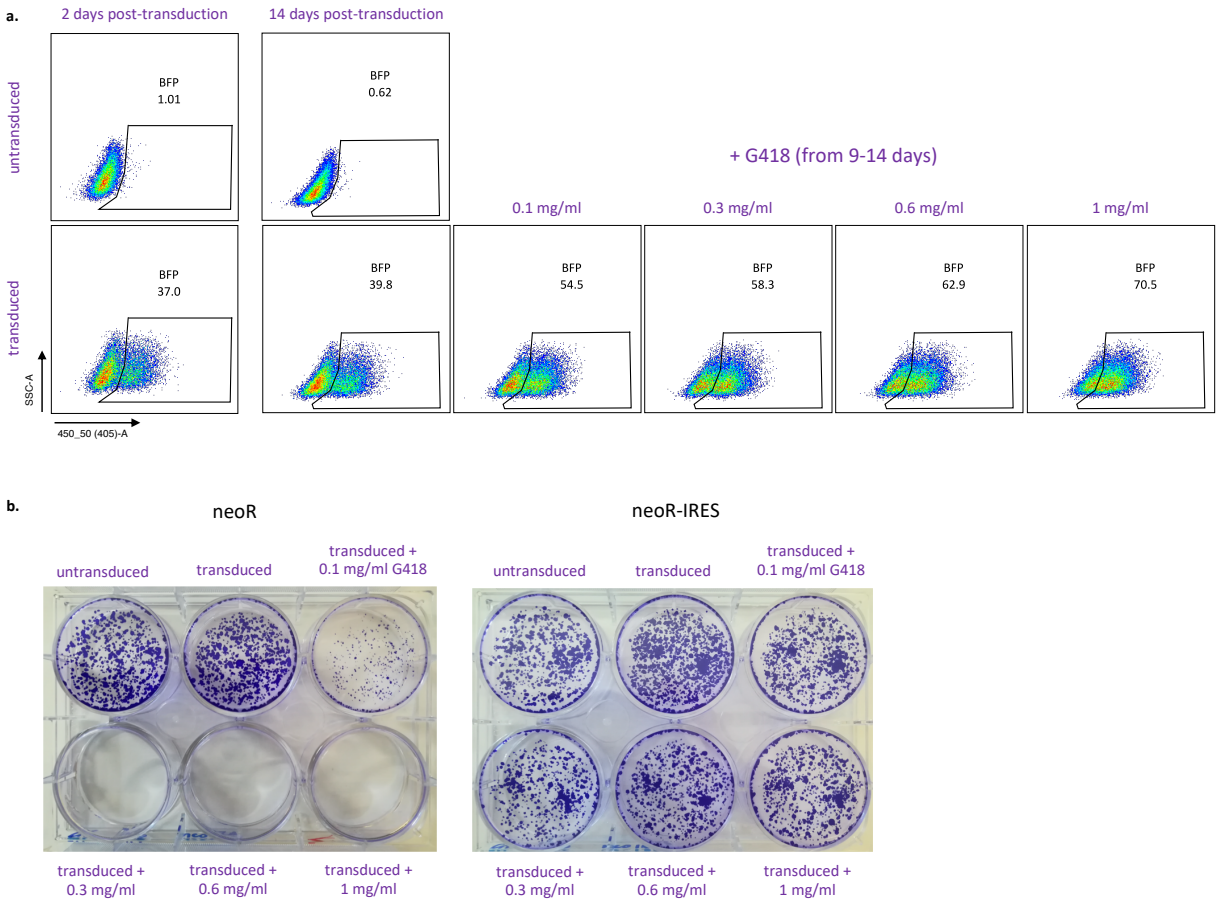

Figure 2.9. Testing neoR-IRES backbone resistance to G418. a) Flow cytometry plots showing BFP expression in BOB-Cas9 cells at 2- and 14- days post-transduction with the neoR-IRES backbone. Cells were selected with various doses of G418 from day 9-14, **b)** Clonogenic assay showing BOB-Cas9 cells stained with crystal violet on day 14 post-transduction with the neoR or neoR-IRES backbone. Cells were selected with various doses of G418 from day 9-14. Untransduced cells were used as a control to show that transduction itself was not toxic.

2.4.3 Alternative strategies for gRNA transfer

In preparation for the transfer of gRNAs from the Yusa v1.1 library, the neoR-IRES backbone was digested with AjuI. Several digestion protocols were tested, including varying the number of enzyme units, digestion volume, DNA:enzyme ratio and digestion time, but complete digestion could not be achieved (Fig. 2.10a). Linearised plasmid was extracted from an agarose gel and transformed into chemically-competent bacteria. As a control, uncut plasmid was also transformed. A large number of colonies were successfully transformed with the extracted linearised plasmid (at least 70% of the number with uncut plasmid). This suggested that there was incomplete separation on the gel and uncut plasmid was still present. It was vital to have a low level of uncut plasmid in the cloning reaction as there was no way to separate this from gRNA-containing plasmid. This would decrease the library coverage in the final DNA pool and essentially waste space as it would enter cells during a screen but have no function.

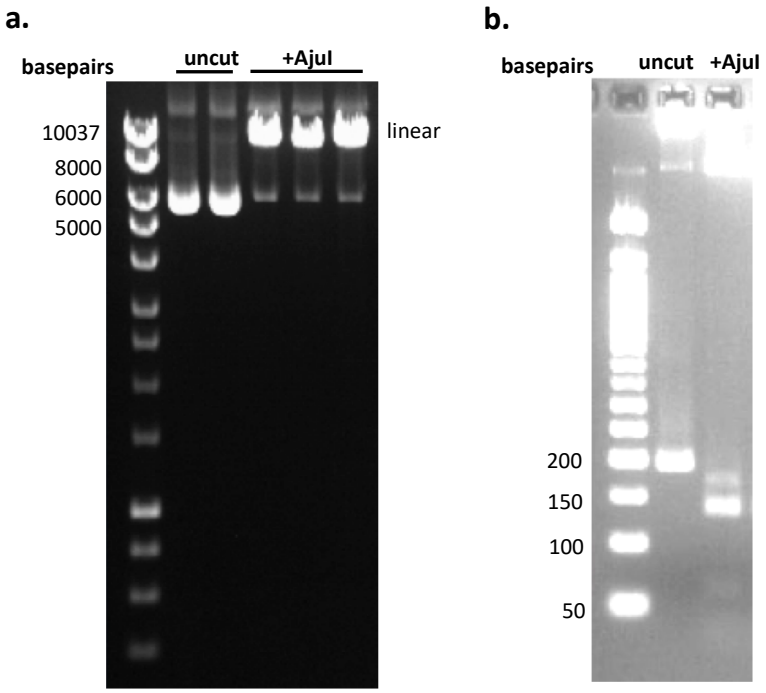

Figure 2.10. Digestion of neoR-IRES backbone with AjuI. a) neoR-IRES plasmid digested with AjuI for (left-right) 15 minutes, 1 hour and 2 hours, **b)** AjuI digestion of a PCR amplicon across the restriction sites. Uncut amplicon = 194 bp; expected bands with 2 cuts = 132 bp, 32 bp, 30 bp; expected bands with 1st cut only: 164 bp and 30 bp; expected bands with 2nd cut only: 132 bp and 62 bp.

Additionally, as AjuI cut at two sites very close to each other, we wanted to confirm that digestion occurred at both. As it was difficult to see such a small size difference from the plasmid digest, PCR was performed across the AjuI restriction sites and the amplicon was digested. A mixture of single and double cut products was evident (Fig. 2.10b), indicating that the enzyme was not efficiently cutting at both sites, which would be necessary for Gibson Assembly. High efficiency was desirable due to the scale of transformation required for cloning the library. Considering these factors, digesting the backbone with AjuI was unsuitable.

As an alternative, we tested a strategy involving double digestion with MluI and AgeI (sites indicated on Fig. 2.8f). The established protocol for transferring gRNAs from the Yusa v1.1 library involved PCR amplification of a region from the end of the U6 promoter to the end of the tracrRNA scaffold. Using AjuI, the same protocol could be used. However, MluI and AgeI digestion removed a larger fragment and so an amplicon of ~600 bp was required to clone gRNAs into the backbone. Digestion appeared to be complete (from gel visualisation, Fig. 2.11a) but extraction of the digested plasmid was carried out as a precaution. This product was transformed into bacteria and, in comparison to uncut plasmid, < 1% of colonies survived with the digested plasmid. This low level of background was ideal for gRNA library transfer.

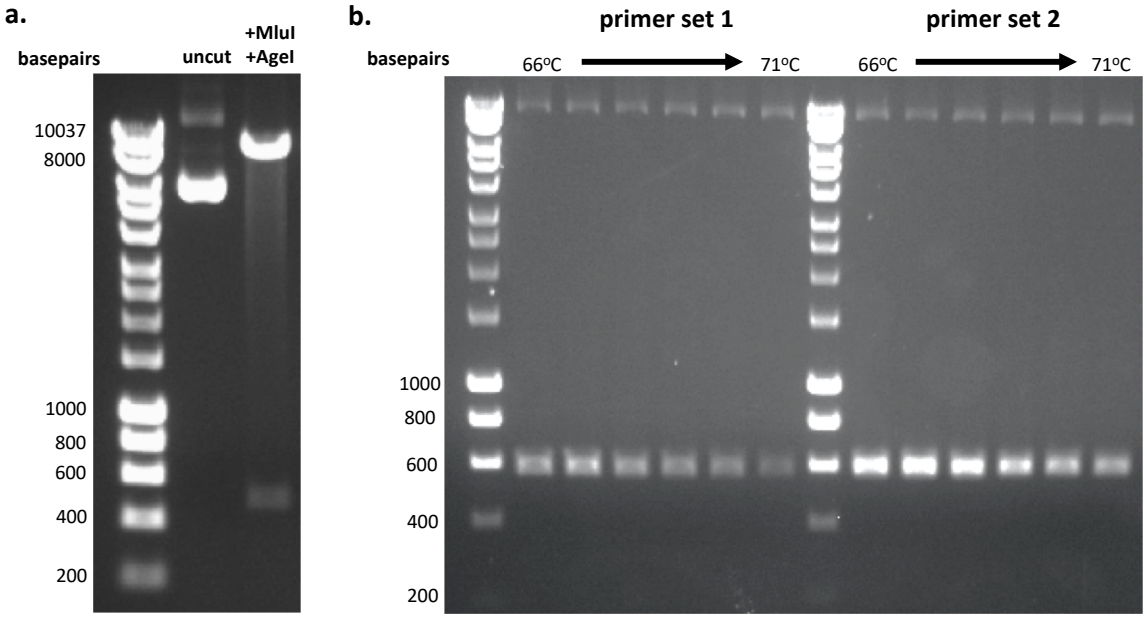

Figure 2.11. Cloning strategy for gRNA library transfer. a) MluI/AgeI double-digest of the neoR-IRES backbone, **b)** PCR amplification of gRNAs from the Yusa v1.1 library with two primer sets and a gradient of annealing temperatures.

PCR was performed to amplify gRNAs from the Yusa v1.1 library plasmid, using primers that extended across the MluI and AgeI sites. A protocol provided by Kosuke Yusa was used; the annealing temperature was optimised and two primer sets were tested. PCR optimisation was performed using the empty Yusa v1.1 backbone as library DNA was limited. Primer set 2 (599 bp amplicon) with an annealing temperature of 67ºC provided the best yield (Fig. 2.11b).

The optimised PCR reaction was performed on Yusa v1.1 library plasmid DNA. A high-fidelity polymerase was used for amplification to reduce the likelihood of sequence errors being introduced to the gRNAs. To ensure high coverage of all 101,090 gRNAs, ten PCR reactions were performed, the products were pooled and the 599 bp band was extracted. Twelve Gibson Assembly reactions were performed to insert the gRNA amplicon into the MluI/AgeI-digested neoR-IRES backbone. These were pooled, column purified and concentrated, then electroporated into electro-competent bacteria. Bacteria were expanded in liquid cultures, with a small fraction plated on agar to allow measurement of transformation efficiency. Based on the colony number, over 2000x coverage of the gRNA library was achieved.

2.5 Assessing gRNA library quality

As an initial quality control measure, 94 bacterial colonies from the library transformation were sequenced by Sanger sequencing (as described in Section 7.7.10). Of these, 10 were discarded due to low quality sequencing. The majority (83%) had the correct backbone sequence; 9% had a single basepair substitution (sub) upstream of the U6 promoter in a region of the plasmid with unknown function; 8% had a single basepair substitution in the PGK promoter (Fig. 2.12a). It was unclear whether these substitutions would affect plasmid function. As they occurred at a low frequency and would be randomly distributed in terms of the gRNAs affected, we concluded that they were unlikely to have a large impact. In terms of the gRNA, 75% of colonies had a correct sequence matching the Yusa v1.1 library and 20% were mixed colonies and so were uninformative (Fig. 2.12b). Only 5% had a sequence error but it was not possible to determine whether we introduced these errors or if they were present in the original library.

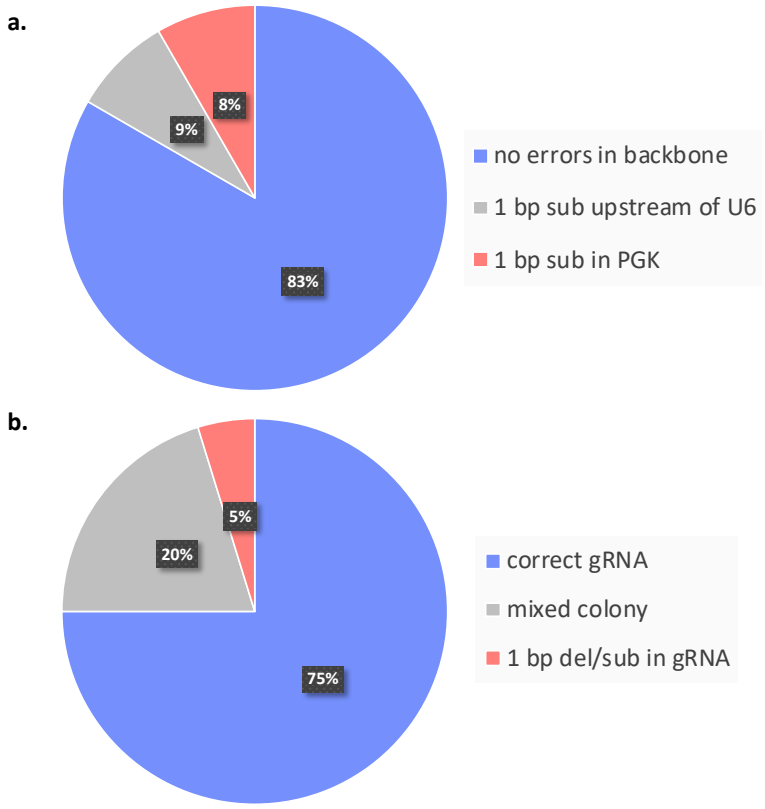

Figure 2.12. Sequencing of colonies from gRNA library transfer. Results from Sanger sequencing of 84 colonies cultured after the transfer of gRNAs into the neoR-IRES backbone. The error rate was calculated for **(a)** the plasmid backbone (in the area altered by cloning) and **(b)** the gRNA sequence (mixed colony indicates cases with mixed sequencing traces in the gRNA region).

Plasmid DNA was extracted from the bulk bacterial culture and PCR was performed to amplify the gRNA region (details provided in Section 7.12). The gRNAs were then sequenced on a HiSeq 2500. To assess the quality of the neoR-IRES library, the results were compared with independent sequencing of the Yusa v1.1 library and the Yusa v1 library, which has also been used successfully in published screens (Table 2.2, Fig. 2.13, Appendix A.4).[110]

Three key variables were considered:

1. Sequencing artefacts or errors introduced during cloning (% unmapped reads)
2. Missing gRNAs (% zero counts)
3. Distribution of gRNA representation (90th/10th percentile ratio)

Table 2.2. Comparison of gRNA library quality.

	Yusa v1	Yusa v1.1	neoR-IRES
% unmapped reads	10%	8%	10%
% zero counts	0.76%	0.2%	0.14%
90th/10th percentile ratio	7.7	3.2	3.1

Considering these factors, the neoR-IRES library appeared to be of a similar quality to Yusa v1.1, and was equal to or improved upon the Yusa v1 library. It should be noted that the neoR-IRES library was sequenced at a deeper coverage than the original libraries, with ~3600x library coverage based on mapped reads for the neoR-IRES library compared to ~1000x for the Yusa v1 and v1.1 libraries.

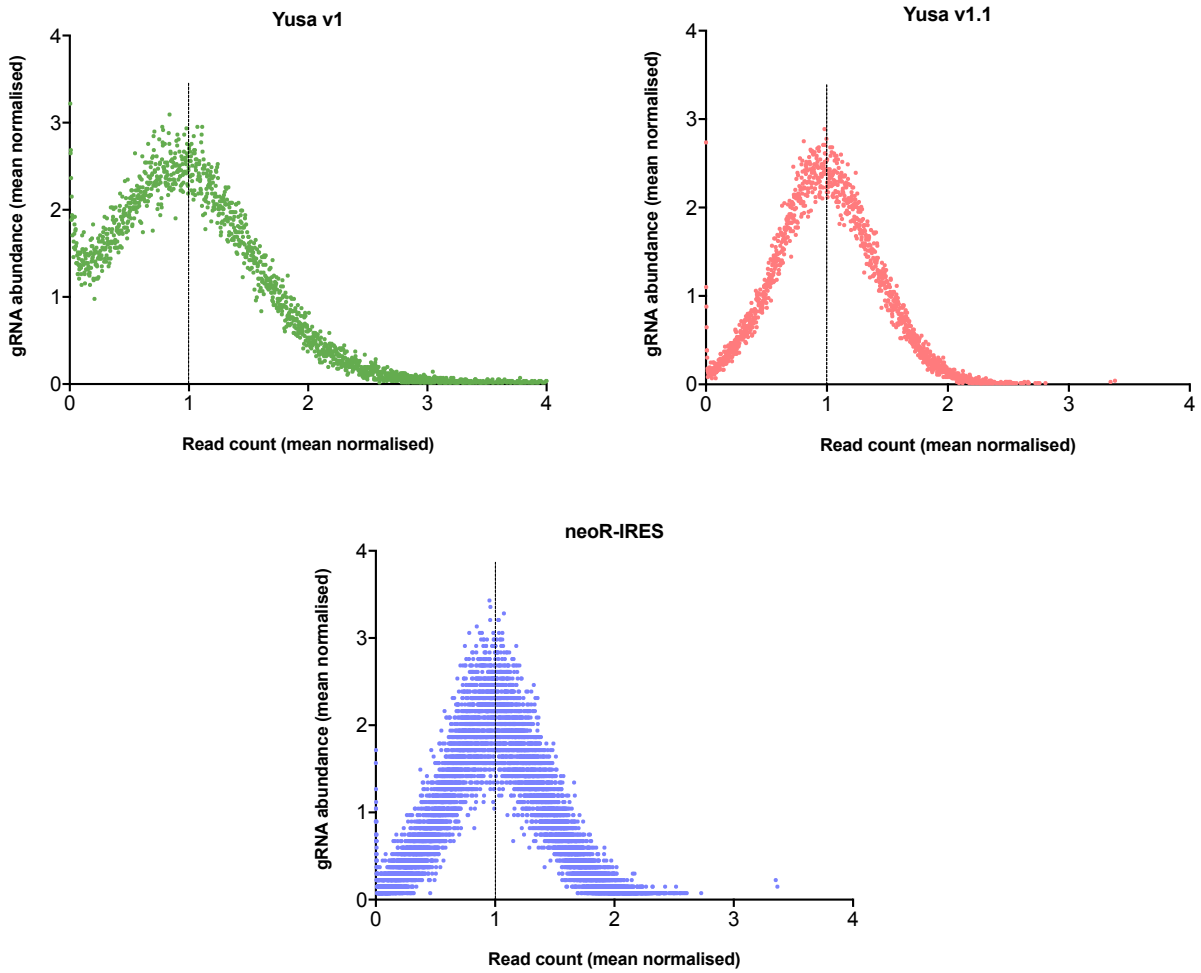

Figure 2.13. Comparison of gRNA distribution between libraries. Abundance of gRNAs (normalised against the mean) with each read count (normalised against the mean) is shown for the Yusa v1, Yusa v1.1 and neoR-IRES libraries. Read counts were obtained from sequencing each plasmid on a HiSeq 2500.

2.6 Engineering stable Cas9 lines

As the genome-wide library encoded gRNAs but not Cas9, we generated iPSCs that had stable expression of Cas9 to perform screening (experimental details provided in Section 7.10). Cells were transduced with a lentiviral plasmid containing Cas9 and blasticidin was used for selection.[110] All iPSC Cas9 lines used in this project were engineered by myself or Clara Alsinet (a postdoctoral fellow in the Adams' lab). After complete selection, Cas9 activity was measured using a fluorescent reporter assay (Fig. 2.14). Cells were transduced with a control BFP/GFP-expressing lentivirus or a reporter lentivirus that expressed BFP/GFP and a gRNA targeting GFP. In the presence of active Cas9, the gRNA targeted the GFP and produced a BFP-only positive population. Cas9 activity was calculated as the % of total transduced cells that were positive only for BFP but not GFP, as measured by flow cytometry. All cell lines displayed a high level of Cas9 activity (> 70%).

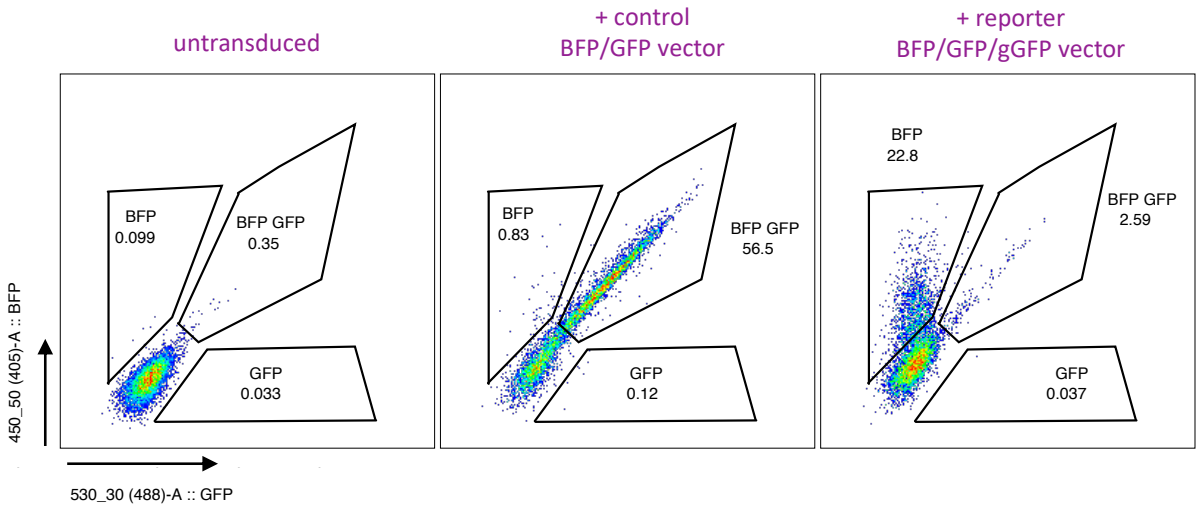

Figure 2.14. Cas9 activity in iPSCs. BOB-Cas9 cells were transduced with a control BFP/GFP vector (middle), a reporter BFP/GFP/gGFP vector (right) or untransduced (left). BFP and GFP expression were measured by flow cytometry at 3 days post-transduction. Untransduced and control cells were used for gating. Cas9 activity was calculated as the percentage of BFP positive cells divided by the percentage of total cells transduced with the reporter vector. In this case 90% activity was measured. These plots are for the parental BOB-Cas9 line and are representative of the other stable Cas9 derivatives generated for the KO lines.

2.7 Genome-wide screening in iPSCs

2.7.1 NeoR-IRES library lentivirus titration

For screening, the library plasmid was packaged into a lentivirus (as described in Section 7.8.1). To ensure that each cell was infected with a single copy of the virus (i.e. cells did not carry multiple gRNAs), we aimed to transduce cells with an MOI of 0.3. To this end, each batch of library lentivirus was titrated in every cell line to be screened, using BFP as a marker (as described in Section 7.8.2). Cells were transduced with various volumes of lentivirus and seeded in 6-well plates. After 48 hours, cells were fixed and BFP expression was analysed by flow cytometry. The volume of lentivirus required to obtain 30% BFP-positive cells was calculated and scaled up for the screens.

A dose-dependent effect on cell viability was observed during the titration, suggesting that the library induced lethality. Significant cell death had not been observed with any other lentiviral transductions using these cells. We hypothesised that double-strand breaks created upon addition of the gRNA library may have induced toxicity. BOB and BOB-Cas9 cells were transduced with various doses of neoR-IRES library lentivirus and BFP expression was measured after 3 days (as described in Section 7.13.1). At every dose, BOB cells had a larger BFP-positive population than BOB-Cas9 cells, despite being transduced in the same conditions (Fig. 2.15). Thus, cells expressing the gRNA (i.e. BFP positive) were depleted in the population when Cas9 was active, relative to when Cas9 was absent. This suggested that the activity of Cas9 in the presence of targeting gRNAs may have had a negative impact on cell fitness.

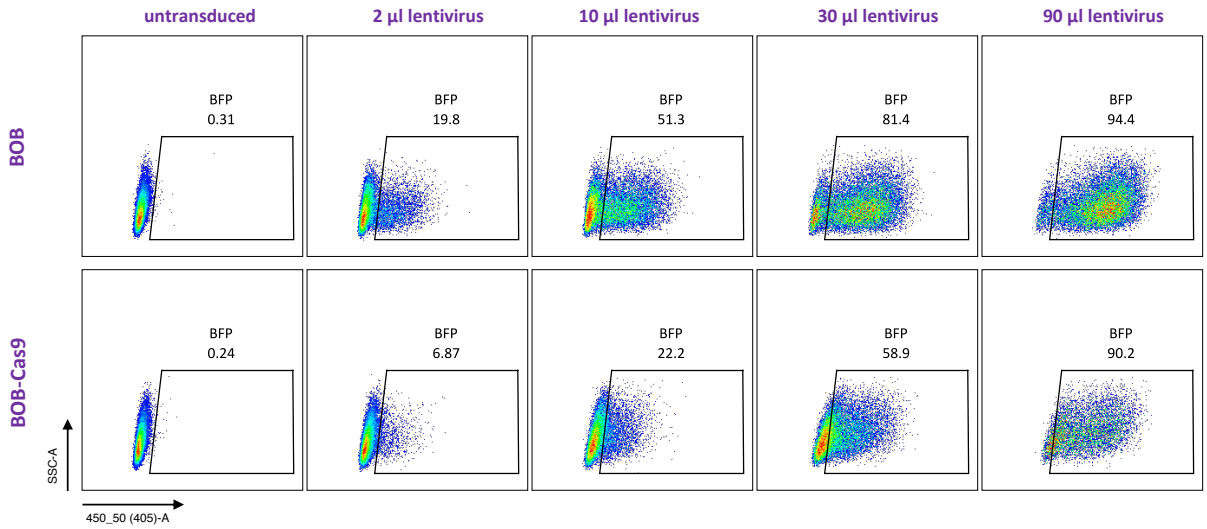

Figure 2.15. Library titration in BOB vs BOB-Cas9. Flow cytometry plots showing BFP expression (450_50 (405)-A) in BOB and BOB-Cas9 cells 3 days post-transduction with various volumes of the neoR-IRES library lentivirus. Untransduced cells were used as a control for both lines.

2.7.2 iPSC screening protocol

After the initial titration, a small-scale version of the screen was carried out in BOB-Cas9 cells to optimise seeding density and passaging methods. We tested seeding $1x10^6$, $2x10^6$ and $3x10^6$ cells per 15 cm dish, using 1 mg/ml G418 for selection based on the backbone assay (discussed previously, Fig. 2.9). A density of $2x10^6$ cells per 15 cm dish was chosen as this allowed cells to grow for 13 days with only one passage required. A brief timeline of the screen is shown in Fig. 2.16 and full details are provided in Section 7.11. I used this protocol to screen the parental BOB line and four KO lines: ARID1A_C09, ARID1B_C03, ARID2_C11 and PBRM1_F09. A further 19 screens, in the parental and additional KO lines, were performed by the Cellular Genotyping and Phenotyping (CGaP) facility at WSI. After transfer to CGaP, this protocol was adapted to include two passages (on day 6 & 8/9) as cells were growing faster, which may have been due to a manufacturer change to the medium.

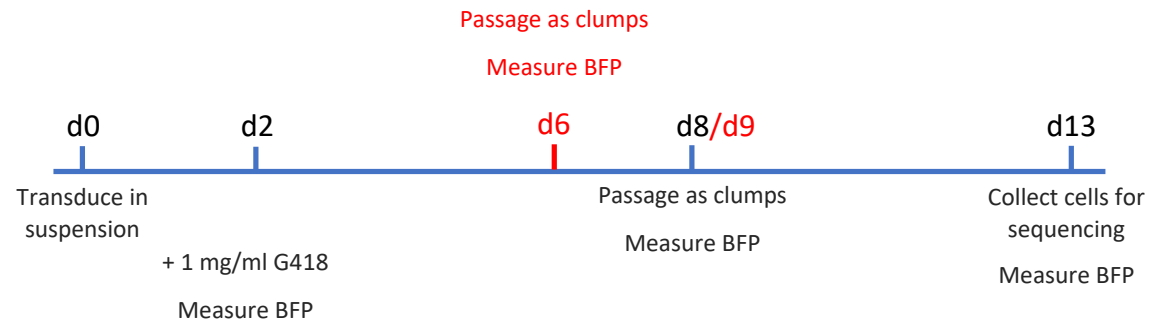

Figure 2.16. iPSC screening protocol. Stable Cas9 cells were transduced in suspension with the neoR-IRES library lentivirus then seeded. After 48 hours, 1 mg/ml of G418 was added. A sample of cells (from a small dish seeded in parallel at the point of transduction) was taken for flow cytometry analysis to measure BFP expression. On day 8, cells were passaged as clumps and BFP expression was measured. On day 13, cells were collected, pelleted and frozen, and BFP expression was measured. Additions highlighted in red indicate changes made by CGaP, with an additional passage on day 6.

2.7.3 BFP expression as a transduction/selection marker

In all screens, BFP expression was measured on day 2 to confirm that an acceptable MOI was achieved, and on day 8 and day 13 to confirm that G418 selection worked (Fig. 2.17) (as described in Section 7.11). On average, there was a ~30% BFP positive population on day 2, but some screens had slightly higher or lower levels. G418 appeared to be working as an increase in BFP was observed at day 8, but an unexpected plateau occurred in the initial screens. The BFP positive populations generally remained stable or dropped after the passage rather than increasing with further selection. Considering that cells still survived in G418, we hypothesised two explanations for this plateau of BFP expression:

a) BFP was being silenced or expressed at a low level

b) G418 selection stopped working > 50% due to cross-resistance

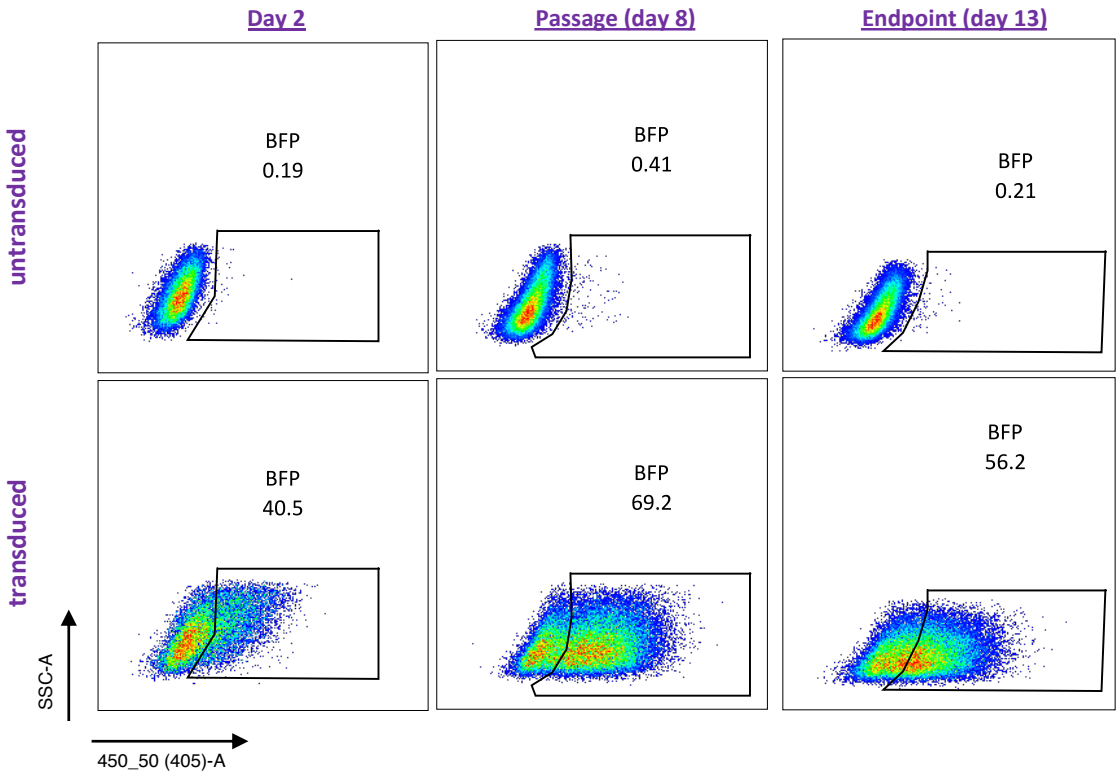

Figure 2.17. BFP expression throughout BOB-Cas9 screen. Flow cytometry plots showing BFP expression (450_50 (405)-A) in BOB-Cas9 cells on day 2, day 8 and day 13 post-transduction with the neoR-IRES library. Untransduced cells were used as a control for gating. This analysis was performed for each screen.

To investigate this, BOB-Cas9 cells were maintained after screening and seeded at a low density. Twenty individual colonies were picked and expanded. Two colonies were picked from untransduced BOB-Cas9 cells and 18 were picked from the cell population that had undergone screening. All clones were cultured with or without G418 for one week and BFP expression was measured by flow cytometry (as described in Section 7.13.2). Of the 18 colonies, three appeared to be mixed with two separate populations evident and only five had complete BFP expression (i.e. almost 100% BFP positive based on gating of untransduced cells, Fig. 2.18: clone C). The remaining colonies had lower levels of BFP expression and were gated as only partially BFP positive despite being a single population (Fig. 2.18: clones A&B). It appeared that many cells were expressing the plasmid at a level sufficient to provide G418 resistance but with a BFP level that could not be accurately detected. This may be due to varying promoter strength at different lentiviral integration sites, but could also be related to the IRES element. It is known that the second gene in a polycistronic mRNA separated by an IRES (BFP) can have lower expression than the first gene (neomycin resistance).[252]

To determine whether cross-resistance was an issue for G418 selection, an untransduced clone was mixed in various ratios with a clone with complete BFP expression (as described in Section 7.13.3). Flow cytometry was performed at the point of mixing the cells and after 5 days in culture with G418 (Fig. 2.19). As a control, cells were also maintained without G418 and BFP expression was stable. Selection led to an increase in the % of BFP positive cells in all conditions, albeit with a slightly lower increase in those that started with a higher % of BFP positive cells.

Our conclusion from this analysis was that it was more likely the BFP expression was not accurately representing the plasmid expression, rather than there being an issue with G418 selection. During the screens performed by CGaP, the expression of BFP was found to vary even further. In some screens, the % of BFP positive cells increased and stabilised above 80%. In others, the increase was much lower and a drop in % BFP was observed in some cases (similar to that shown in Fig. 2.17) despite continued G418 selection. This could be due to the cells silencing BFP expression. In the Cas9 activity test, we also observed that BFP expression was low and did not result in a clearly distinct population (Fig. 2.14). The plasmids used for this test encoded BFP as the first protein rather than the second (as in our backbone), but it was still under the control of a PGK promoter.[110] Thus, BFP expression in these iPSC lines may be a general issue, not specific to our library plasmid. In future, an antibody against BFP could be used to improve detection of expression. These findings indicate that BFP was not a reliable marker of library plasmid expression and we may have underestimated the transduction levels.

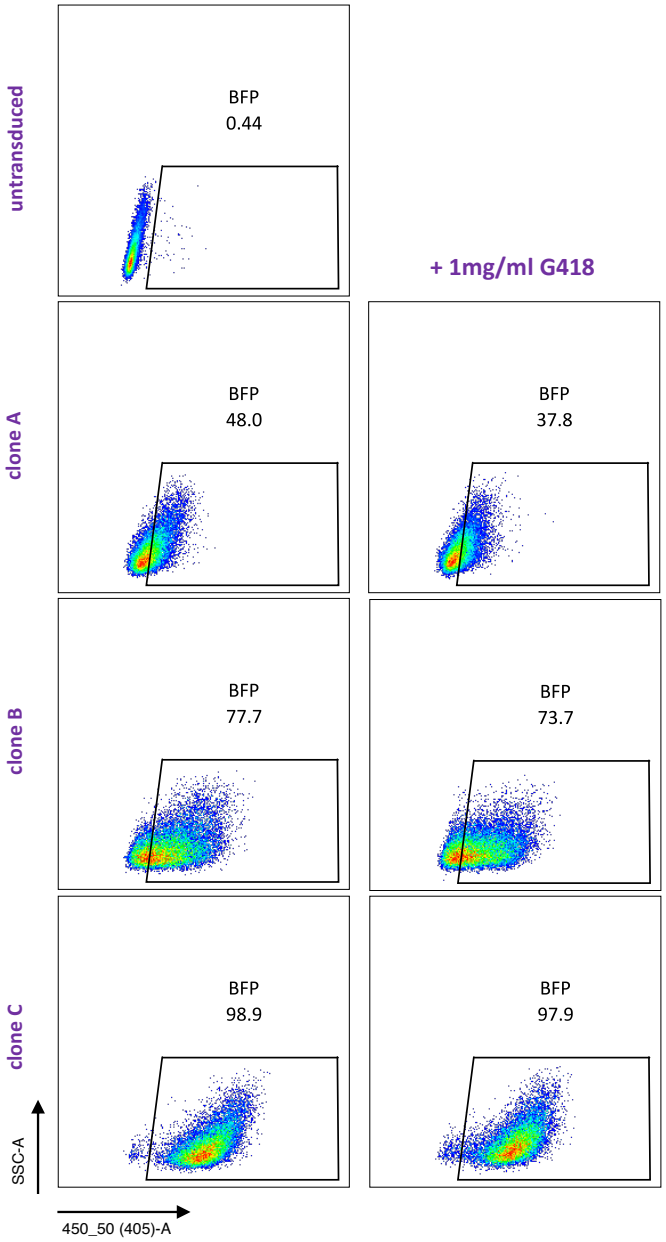

Figure 2.18. Variable BFP expression in clonal transduced lines. Single-cell colonies were picked from populations of BOB-Cas9 cells that were either untransduced or had been screened with the neoR-IRES library. All transduced cells had been selected with 1 mg/ml G418 throughout the screen. Each transduced clonal line was maintained with or without 1 mg/ml G418 for 7 days. Cells were fixed and BFP expression (450_50 (405)-A) was measured by flow cytometry. Representative plots are shown for three colonies.

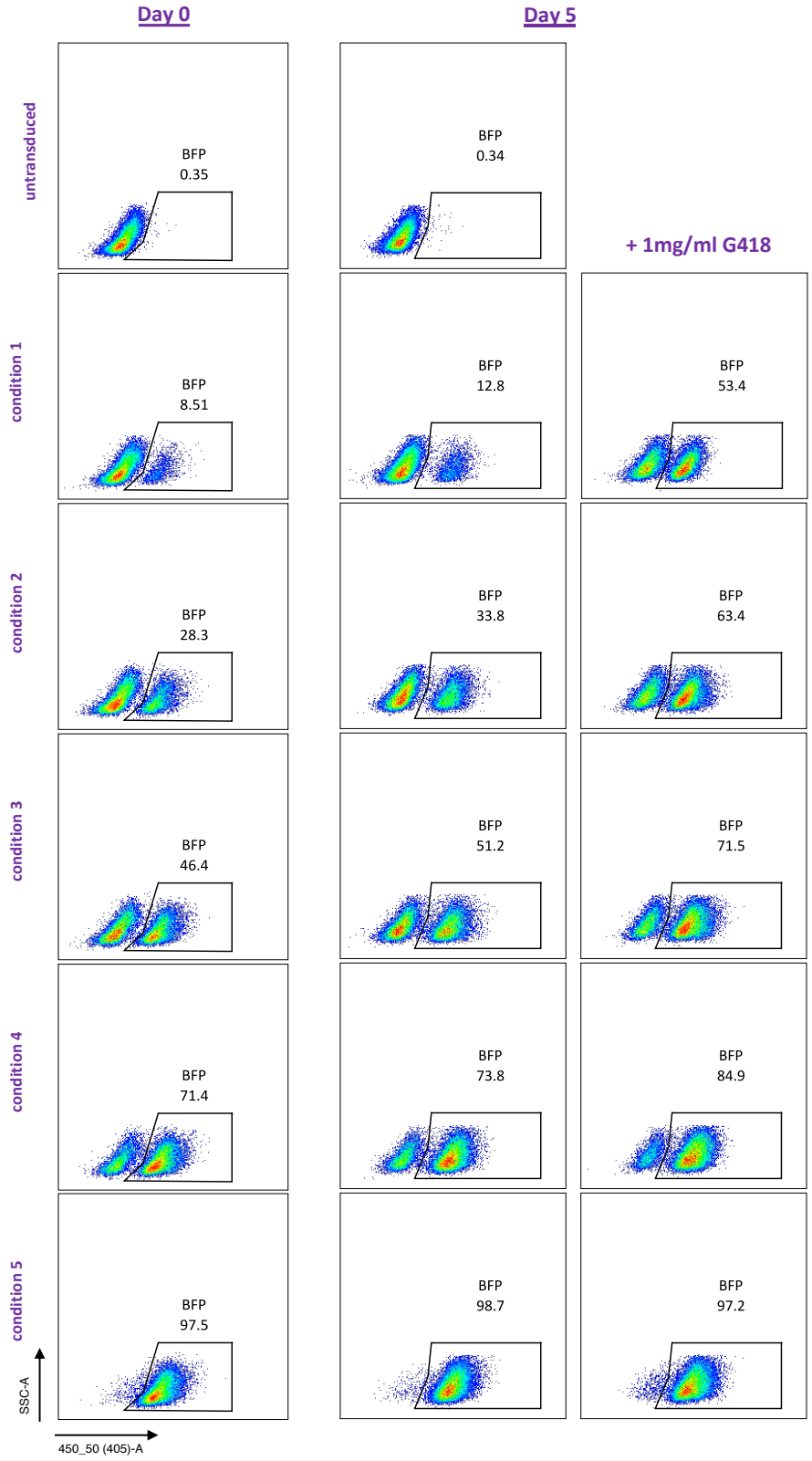

Figure 2.19. Efficiency of G418 selection. Untransduced BOB-Cas9 cells were mixed with a clonal transduced BOB-Cas9 line that was previously confirmed to have ~100% BFP expression. Populations containing 10%, 30%, 50%, 70% and 100% of the transduced line were prepared. A fraction of each population was fixed at the point of mixing, and the remainder were cultured for 5 days with or without 1 mg/ml G418. Cells from day 0 and day 5 were analysed by flow cytometry to measure BFP expression (450_50 (405)-A).

2.8 Effect of CRISPR/Cas9 on iPSC karyotype

Prior to screening, stable Cas9 derivatives of BOB and four PBAF/BAF gene KO lines (ARID1A_C09, ARID1B_C03, ARID2_C11, PBRM1_F09) were karyotyped to confirm that they were stable over time, and were unaffected by Cas9 expression. To visualise the karyotypes, multiplex fluorescence *in situ* hybridisation (M-FISH) was performed by the FISH facility at WSI (as described in Section 7.14). BOB-Cas9 cells were also karyotyped post-screen, in parallel with untransduced cells cultured for the same time (Fig. 2.20). Karyotypes for the KO lines are shown in Fig. B.1, Appendix B. All cells had a balanced translocation t(6;8)(p21.1;q24.1), which was previously identified in the parental BOB line by other groups at WSI and was likely present in the donor cells used to derive the line. No other chromosomal abnormalities were identified. The t(6;8)(p21;q24) translocation has been reported in cases of a rare haematological malignancy, blastic plasmacytoid dendritic cell neoplasm.[253] These cases involved translocation of *MYC*, which lies on 8q24, but always occurred alongside other chromosomal aberrations. Further analysis carried out on the BOB line by others at WSI (unpublished) excluded gene disruption and smaller copy number variations at the breakpoint. This line was used as it had better survival and a greater tolerance of the editing process than more sensitive iPSC lines. However, as BOB was used as a normal cell model, potential effects of this chromosomal abnormality should be kept in mind. Results must always be validated in an independent cell line to rule out the possibility of cell line-specific effects.

Karyotyping using M-FISH gives a high-level view, limited to a resolution of 3-5 Mb, thus smaller chromosomal changes would not be detected using this assay. The technique is more sensitive at detecting small insertions than it is for small translocations involving segments from the end of chromosomes. M-FISH is not ideal for detecting small deletions or duplications. Microarray-based comparative genomic hybridisation (array CGH) is a technique which offers greater resolution than M-FISH.[254] Array CGH uses DNA, rather than cells in culture, to identify changes in ploidy compared to a reference sample. However, unlike M-FISH, this technique cannot detect balanced chromosomal abnormalities. The most thorough approach to identifying abnormalities induced by the CRISPR/Cas9 screening process would be to perform whole-genome sequencing on the cell lines, but this is an expensive option. It may be appropriate to first sequence the parental line after Cas9 transduction/screening to get an initial indication of any substantial effects on the genome of these cells.

Figure 2.20. Karyotype of BOB-Cas9 cells. M-FISH of BOB-Cas9 cells that were untransduced or had undergone screening with the neoR-IRES library. Ten randomly selected metaphases were analysed for each line; a representative karyotype is shown for both. A balanced translocation between chromosomes 6 and 8 is indicated.

2.9 Effect of CRISPR/Cas9 on iPSC pluripotency

Transient changes in morphology were observed when cells were transduced with lentivirus containing the Cas9 plasmid or the library plasmid. Cells returned to a normal morphology after passaging. However, to ensure that the screening process was not affecting the pluripotency of the cells, expression of 3 pluripotency markers (OCT4, SOX2, NANOG) was measured in BOB-Cas9 cells post-screen and compared with untransduced cells that had been cultured for the same period of time. This assay was carried out with the assistance of Mary Goodwin (CGaP, WSI) as described in Section 7.15. DAPI was used to stain the nuclei of individual cells. Cells were also stained with antibodies to detect OCT4, SOX2 and NANOG expression. The percentage of DAPI-positive cells expressing each marker was then calculated. No significant difference was observed for any of the markers when comparing untransduced and screened cells (Fig. 2.21).

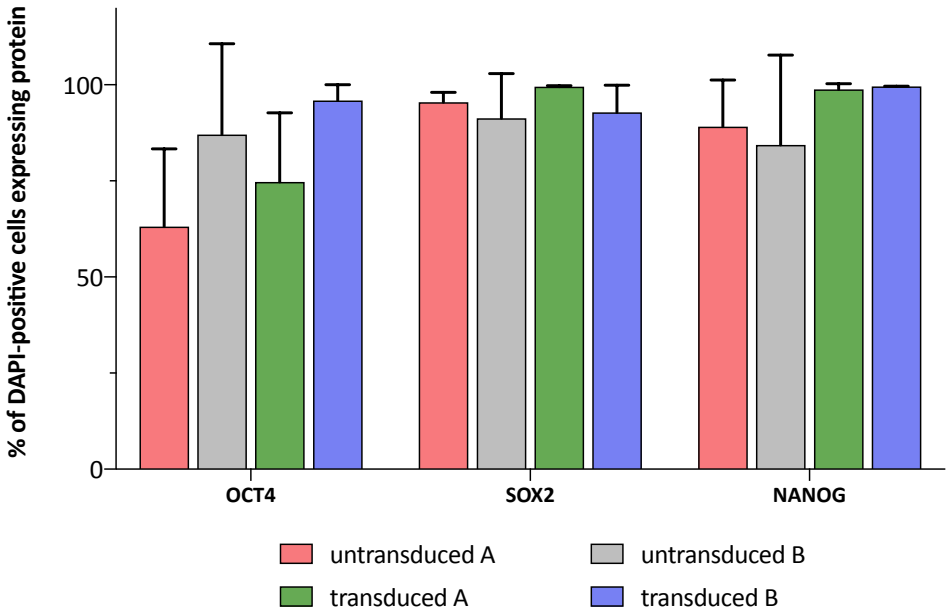

Figure 2.21. Expression of pluripotency markers in BOB-Cas9 cells. Cells from two replicates of the BOB-Cas9 neoR-IRES screen were maintained separately after the screen endpoint (transduced A/B). Untransduced cells from each replicate population were cultured in parallel (untransduced A/B). Cells from all conditions were seeded in a 96-well plate and cultured for 72 hours. These were fixed and stained for OCT4, SOX2 and NANOG expression; DAPI was used for nuclear quantification. Expression of each protein was measured using an ArrayScan VTI HCS Reader. The percentage of DAPI-positive cells expressing each marker was calculated. Background expression was measured in control wells and was subtracted from each condition. The mean expression values (from four technical replicates) are shown for each protein in each cell line. Error bars represent standard deviation.

2.10 Summary

We engineered isogenic derivatives of a human iPSC line, BOB, each carrying LOF of a single candidate TSG (29 genes in total). Knockout of each gene was confirmed at the DNA and protein level by Sanger sequencing and mass spectrometry, respectively. As the KO lines were designed to be resistant to puromycin (the antibiotic resistance marker used in many gRNA libraries), we constructed a gRNA library backbone with a neomycin resistance gene. After confirming that the backbone was functional, an existing genome-wide gRNA library was transferred. Sequencing confirmed that the library was of a comparable quality to the original. Each iPSC line was then engineered to constitutively express Cas9. The new gRNA library was packaged into a lentivirus and titred in each line. A screening protocol was optimised in the parental BOB-Cas9 line before being applied in a further 20 KO lines. Various assays indicated that BFP expression from the backbone (used as a marker of transduction and selection) was weak and may have caused underestimations. We also found that the screening process had a large impact on cell viability, but did not affect karyotype stability or expression of key pluripotency markers.